ENCYCLOPEDIA OF MODERN OPTICS

EDITED BY
B.D. GUENTHER
A. MILLER
L. BAYVEL
J.E. MIDWINTER

VOLUME ONE

AP

EDITOR-IN-CHIEF

Robert D. Guenther
Duke University
Durham, NC, USA

EDITORS

Duncan G. Steel
University of Michigan
Ann Arbor, MI, USA

Leopold Bayvel
London, UK

CONSULTING EDITORS

John E. Midwinter
University College London, London, UK

Alan Miller
University of St Andrews, Fife, UK

EDITORIAL ADVISORY BOARD

R Alferness
Lucent Technologies Bell Laboratories
Holmdel, NJ, USA

D J Brady
Duke University
Durham, NC, USA

J Caulfield
Diversified Research Corporation
Cornersville, TN, USA

R L Donofrio
Display Device Consultants LLC
Ann Arbor, MI, USA

J Dudley
Université de Franche-Comté
Besançon, France

J G Eden
University of Illinois
Urbana, IL, USA

H O Everitt
U.S. Army Research Office
Research Triangle Park, NC, USA

E L Hecht
Adelphi University
Garden City, NY, USA

H S Hinton
Utah State University
Logan, UT, USA

T N Hurst
Hewlett-Packard Laboratories
Tucson, AZ, USA

D Killinger
University of South Florida
Tampa, FL, USA

T A King
University of Manchester
UK

E Leith
University of Michigan
Ann Arbor, MI, USA

A Macleod
Thin Film Center, Inc.
Tucson, AZ, USA

P F McManamon
Air Force Research Laboratory
Wright-Paterson AFB, OH, USA

J N Mait
U.S. Army Research Laboratory
Adelphi, MD, USA

D Phillips
Imperial College
London, UK

C R Pidgeon
Heriot-Watt University
Edinburgh, UK

R Pike
King's College London
UK

E Van Stryland
University of Central Florida
Orlando, FL, USA

Preface

We live in a world powered by light, but much of the understanding of light was developed around the time of the French Revolution. Before the 1950s, optics technology was viewed as well established and there was little expectation of growth either in scientific understanding or in technological exploitation. The end of the Second World War brought about huge growth in scientific exploration, and the field of optics benefited from that growth. The key event was the discovery of methods for producing a source of coherent radiation, with the key milestone being the demonstration of the first laser by Ted Maiman in 1960. Other lasers, nonlinear optical phenomena, and technologies such as holography and optical signal processing, followed in the early 1960s. In the 1970s the foundations of fiber optical communications were laid, with the development of low-loss glass fibers and sources that operated in the wavelength region of low loss. The 1980s saw the most significant technological accomplishment: the development of efficient optical systems and resulting useful devices. Now, some forty years after the demonstration of the first coherent light source, we find that optics has become the enabling technology in areas such as:

- information technology and telecommunications;
- health care and the life sciences;
- sensing, lighting, and energy;
- manufacturing;
- national defense;
- manufacturing of precision optical components and systems; and
- optics research and education.

We find ourselves depending on CDs for data storage, on digital cameras and printers to produce our family photographs, on high speed internet connections based on optical fibers, on optical based DNA sequencing systems; our physicians are making use of new therapies and diagnostic techniques founded on optics.

To contribute to such a wide range of applications requires a truly multidisciplinary effort drawing together knowledge spanning many of the traditional academic boundaries. To exploit the accomplishments of the past forty years and to enable a revolution in world fiber-optic communications, new modalities in the practice of medicine, a more effective national defense, exploration of the frontiers of science, and much more, a resource to provide access to the foundations of this field is needed. The purpose of this Encyclopedia is to provide a resource for introducing optical fundamentals and technologies to the general technical audience for whom optics is a key capability in exploring their field of interest.

Some 25 internationally recognized scientists and engineers served as editors. They helped in selecting the topical coverage and choosing the over 260 authors who prepared the individual articles. The authors form an international group who are expert in their discipline and come from every part of the technological community spanning academia, government and industry. The editors and authors of this Encyclopedia hope that the reader finds in these pages the information needed to provide guidance in exploring and utilizing optics.

As Editor-in-Chief I would like to thank all of the topical editors, authors and the staff of Elsevier for each of their contributions. Special thanks should go Dr Martin Ruck of Elsevier who provided not only organizational skills but also technological knowledge which allowed all of the numerous loose ends to be tied.

B D Guenther
Editor-in-Chief

A

ALL-OPTICAL SIGNAL REGENERATION

O Leclerc, Alcatel Research & Innovation, Marcoussis, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The breakthrough of optical amplification, combined with the techniques of wavelength division multiplexing (WDM) and dispersion management, have made it possible to exploit a sizeable fraction of the optical fiber bandwidth (several terahertz). Systems based on 10 Gbit/s per channel bit-rate and showing capacities of several terabit/s, with transmission capabilities of hundreds or even thousands of kilometers, have reached the commercial area.

While greater capacities and spectral efficiencies are likely to be reached with current technologies, there is potential economic interest in reducing the number of wavelength channels by increasing the channel rate (e.g., 40 Gbit/s). However, such fourfold increase in the channel bit-rate clearly results in a significant increase in propagation impairments, stemming from the combined effects of noise accumulation, fiber dispersion, fiber nonlinearities, and inter-channel interactions and contributing to two main forms of signal degradation. The first one is related to the amplitude domain; power levels of marks and spaces can suffer from random deviations arising from interaction between signal and amplified spontaneous emission (ASE) noise or with signals from other channels through cross-phase modulation (XPM) from distortions induced by chromatic dispersion. The second type of signal degradations occurs in the time domain; time position of pulses can also suffer from random deviations arising from interactions between signal and ASE noise through fiber dispersion. Preservation of high power contrast between '1' and '0', and of both amplitude fluctuations and timing jitter below some acceptable levels, are mandatory for high transmission quality, evaluated through bit-error-rate (BER) measurements or estimated by Q-factors. Moreover, in future optical networks, it appears mandatory to ensure similar but high optical signal quality at the output of whatever nodes in the networks, as to enable successful transmission of the data over arbitrary distance.

Among possible solutions to overcome such systems limitations is the implementation of Optical Signal Regeneration, either in-line for long-haul transmission applications or at the output of network nodes. Such Signal Regeneration performs, or should be able to perform, three basic signal-processing functions that are Re-amplifying, Re-shaping, and Re-timing, hence the generic acronym '3R' (Figure 1). When Re-timing is absent, one usually refers to the regenerator as '2R' device, which has only re-amplifying and re-shaping capabilities. Thus, full 3R regeneration with retiming capability requires clock extraction.

Given system impairments after some transmission distance, two solutions remain for extending the actual reach of an optical transmission system or the scalability of an optical network. The first consists in segmenting the system into independent trunks, with full electronic repeater/transceivers at interfaces (we shall refer to this as 'opto-electronic regeneration' or O/E Regeneration forthwith). The second solution, i.e., all-optical Regeneration, is not the optical version of the first which would have higher bandwidth capability but still performs the same signal-restoring functions with far reduced complexity. At this point, it should be noted that Optical 3R techniques are not necessarily void of any electronic functions (e.g., when using electronic clock recovery and O/E modulation), but the main feature is that these electronic functions are narrowband (as opposed to broadband in the case of electronic regeneration).

Some key issues have to be considered when comparing such Signal Regeneration approaches. The first is that today's and future optical transmission systems or/and networks are WDM networks.

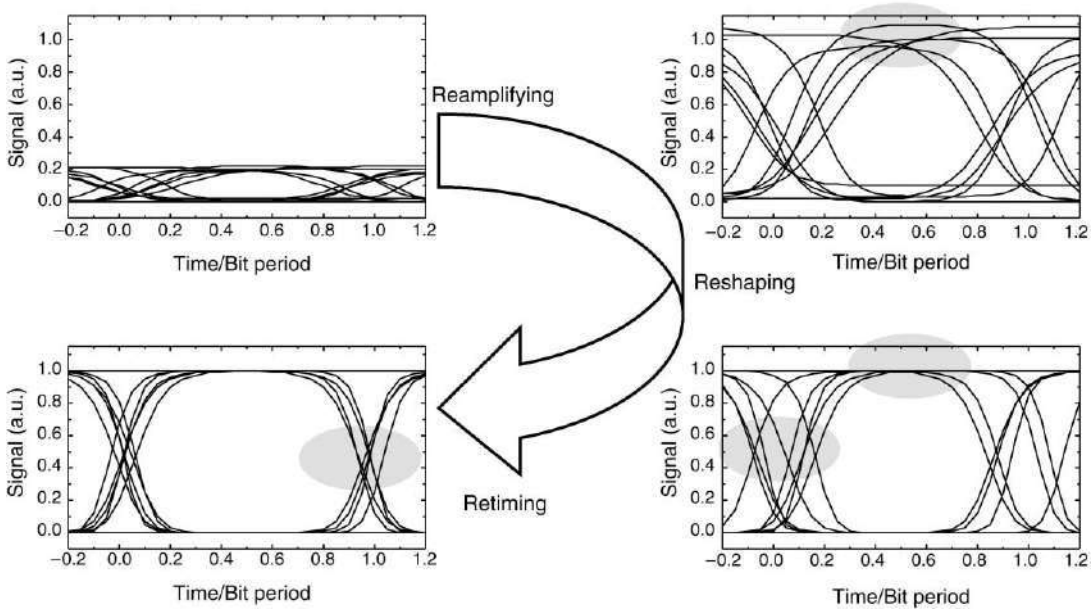

Figure 1 Principle of 3R regeneration, as applied to NRZ signals. (a) Re-Amplifying; (b) Re-Shaping; (c) Re-Timing. NB: Such eye diagrams can be either optical or electrical eye diagrams.

Under this condition, the WDM compatibility – or the fact that any Regeneration solution can simultaneously process several WDM channels – represents a key advantage. The maturity of the technology – either purely optical or opto-electronic – also plays an important role in the potential (pre-)development of such solutions. But the main parameter that will decide the actual technology (and also technique) relies on the tradeoff between actual performance of the regeneration solutions and their costs (device and implementation), depending on the targeted applications (long-haul system, medium haul transport, wide area optical network, etc.).

In this article, we review the current alternatives for all-optical Signal Regeneration, considering both theoretical and experimental performance and practical implementation issues. Key advantages and possible drawbacks of each solutions are discussed, to sketch the picture in this field. However, first we must focus on some generalities about Signal Regeneration and the way to define (and qualify) such regenerator performance. In a second part, we will detail the currently-investigated optical solutions for Signal Regeneration with a specific highlight for semiconductor-based solutions using either semiconductor optical amplifiers (SOA) technology or newly-developed saturable absorbers. Optical Regeneration techniques based on synchronous modulation will also be discussed in a third section. The conclusion will summarize the key features of each solution, so as to underline the demanding challenge optical components are facing in this application.

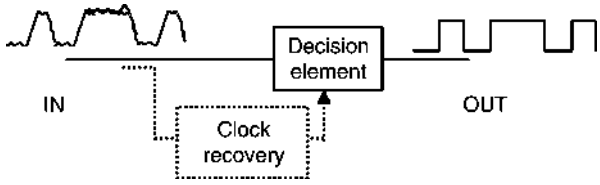

Figure 2 Generic structure of Signal 2R/3R Regenerator based on Decision Element (2R) and Decision Element and Clock Recovery (3R).

Generalities on Signal Regeneration

Principles

In the general case, Signal Regeneration is performed using a decision element exhibiting a nonlinear transfer function. Provided with a threshold level and when associated with an amplifier, such an element then performs the actual Re-shaping of the incoming data (either in the electrical or optical domain) and completes a 2R Signal Regenerator. Figure 2 shows the generic structure of such a Signal Regenerator in the general case as applied to non return to zero (NRZ) data. A clock recovery block can be added (dotted lines) to provide the decision element with time reference and hence perform the third R (Re-timing) of full Signal 3R Regeneration. At this point, it should be mentioned that the decision

element can operate either on electrical signals (standard electrical DFF) provided that optical → electrical and electrical → optical signal conversion stages are added or directed onto optical signals using the different techniques described below. The clock signal can be of an electrical nature, as for electrical decision element in O/E regenerator – or either an electrical or a purely optical signal in all-optical regenerators.

Prior to reviewing and describing the various current technology alternatives for such Optical Signal Regeneration, the issue of the actual characterization of regenerator performance needs to be explained and clarified. As previously mentioned, the core element of any Signal Regenerator is the decision element showing a nonlinear transfer function that can be of varying steepness. As will be seen in Figure 3, the actual regenerative performance of the regenerator will indeed depend upon the degree of nonlinearity of the decision element transfer function.

Figure 3 shows the principle of operation of a regenerator incorporating a decision element with two steepnesses of the nonlinear transfer function. In any case, the '1' and '0' symbols amplitude probability densities (PD) are squeezed after passing through the decision element. However, depending upon the addition of a clock reference for triggering the decision element, the symbol arrival time PD will be also squeezed (clocked decision = 3R regeneration) or enlarged (no clock reference = 2R regeneration) resulting in conversion of amplitude fluctuations to time position fluctuations.

As for system performance – expressed through BER – regenerative capabilities of any regenerator simultaneously depend upon both the output amplitude and arrival time PD of the '1' and '0' symbols. In the unusual case of 2R regeneration (no clocked decision), a tradeoff has then to be derived, considering both the reduction of amplitude PD and the enlarged arrival time PD induced by the regenerator, to ensure sufficient signal improvement. In Figure 3a, we consider a step function for the transfer function of the decision element. In this case, amplitude PD are squeezed to Dirac PD after the decision element, and depending upon addition or not of a clock reference, arrival time PD is reduced (3R) or dramatically enlarged (2R). In Figure 3b, the decision element exhibits a moderately nonlinear transfer function. This results in an asymmetric and less-pronounced squeezing of the amplitude PD compared to the previous case, but in turn results in a significantly less enlarged arrival time PD when no clock reference is added (2R regeneration). Comparison of these two decision element of different nonlinear transfer function indicates that for 3R regeneration applications, the more nonlinear the transfer function of the decision element the better performance, the ideal case being the step function. In the case of 2R regeneration applications, a tradeoff between the actual reduction of the amplitude PD and enlargement of timing PD is to be derived and clearly depends upon the actual shape of the nonlinear transfer function of the decision element.

Qualification of Signal Regenerator Performance

To further illustrate the impact of the actual shape of the nonlinear transfer function of the decision

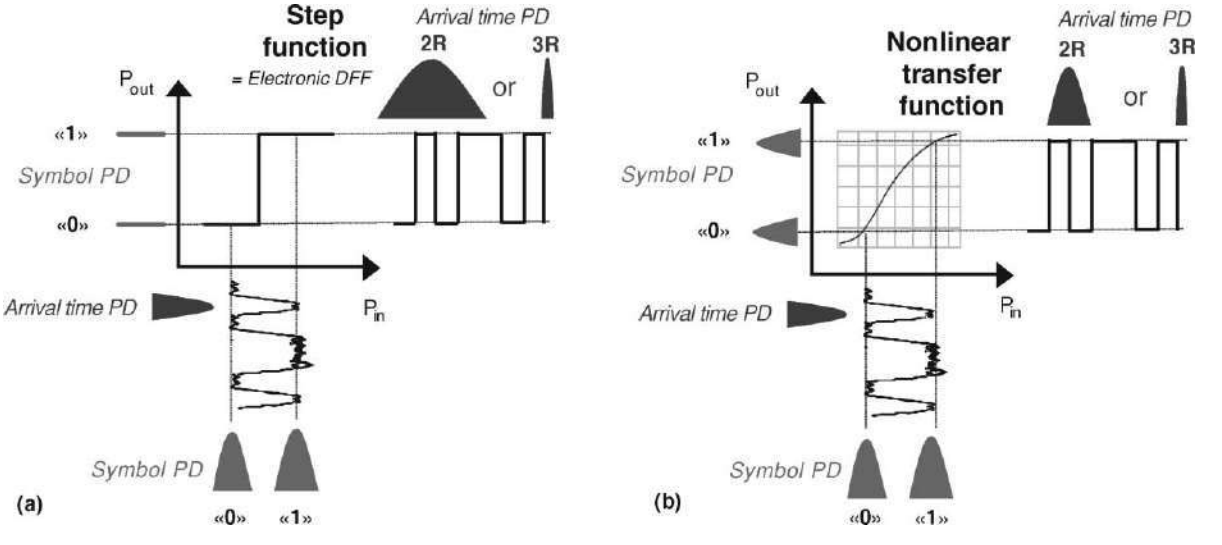

Figure 3 Signal Regeneration process using Nonlinear Gates. (a) Step transfer function (= Electronic DFF); (b) 'moderately' nonlinear transfer function. As an illustration of the Regenerator operation '1' and '0' symbols amplitude probability density (PD) and arrival time probability density (PD) are shown in light gray and dark gray, respectively.

element in 3R application, the theoretical evolution of BER with number of concatenated regenerators have been plotted for regenerators having different nonlinear responses. Figure 4 shows the numerically calculated evolution of the BER of a 10 Gbit/s NRZ signal with fixed optical signal-to-noise ratio (OSNR) at the input of the 3R regenerator, as a function of the cascaded regenerator incorporating nonlinear gates with nonlinear transfer function of different depths. From Figure 4, can be seen different behaviors depending on the nonlinear function shape. As previously stated, the best regeneration performance is obtained with an ideal step function (case a), which is actually the case for O/E regenerator using electronic decision flip-flop (DFF). In that case, BER linearly increase (i.e., more errors) in the cascade. Conversely, when nonlinearity is reduced (cases (b) and (c)), both BER and noise accumulate, until the concatenation of nonlinear functions reach some steady-state pattern, from which BER linearly increases. Concatenation of nonlinear devices thus magnifies shape differences in their nonlinear response, and hence their regenerative capabilities.

Moreover, as can be seen in Figure 4, all curves standing for different regeneration efficiencies pass through a common point defined after the first device. This clearly indicates that it is not possible to qualify the regenerative capability of any regenerator when considering the output signal after only one regenerator. Indeed, the BER is the same for either a 3R regenerator or a mere amplifier if only measured after a single element. This originates from the initial overlap between noise distributions associated with marks and spaces, that cannot be suppressed but only minimized by a single decision element through threshold adjustment.

As a general result, the actual characterization of the regenerative performance of any regenerator should *in fine* be conducted considering a cascade of regenerators. In practice this can easily be done with the experimental implementation of the regenerator under study in a recirculating loop. Moreover, such an investigation tool will also enable access to the regenerator performance with respect to the transmission capabilities of the regenerated signal, which should not be overlooked.

Let us now consider the physical implementation of such all-optical Signal Regenerators, along with the key features offered by the different alternatives.

All-Optical 2R/3R Regeneration Using Optical Nonlinear Gates

Prior to describing the different solutions for all-optical in-line Optical Signal Regeneration, it should be mentioned that since the polarization states of the optical data signals cannot be preserved during propagation, it is required that the regenerator exhibits an extremely low polarization sensitivity. This clearly translates to a careful optimization of all the different optical components making up the 2R/3R regenerator. It should be noted that this also applies to the O/E solutions but is of limited impact, since only the photodiode has to be polarization insensitive.

Figure 5 illustrates the generic principle of operation of an all-optical 3R Regenerator using optical nonlinear gates. Contrary to what occurs in O/E, regenerator where the extracted clock signal drives the electrical decision element, the incoming and distorted optical data signal triggers the nonlinear gate, hence generating a switching window which is

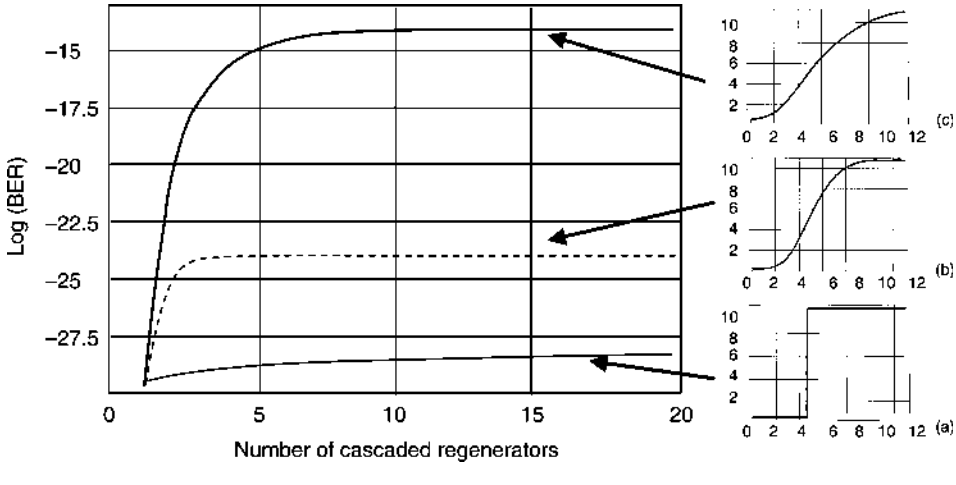

Figure 4 Evolution of the BER with concatenated regenerators for nonlinear gates with nonlinear transfer function of decreasing depths from case (a)–(c) (step function).

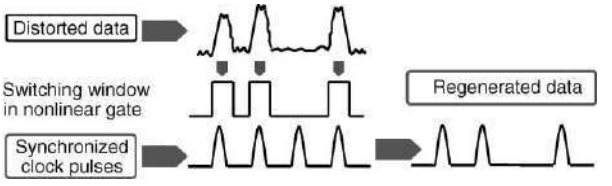

Figure 5 Principle of operation of an all-Optical Signal 3R Regenerator using nonlinear gates.

applied to a newly generated optical clock signal so as to reproduce the initial data stream on the new optical carrier.

In the case of 2R Optical Signal Regeneration, a continuous wave (CW) signal is substituted for the synchronized optical clock pulses. As previously mentioned, the actual regeneration performance of the 2R/3R devices will mainly depend upon the nonlinearity of the transfer function of the decision element but in 3R applications the quality of the optical clock pulses has also to be considered. In the following, we describe current solutions to explain the two main building blocks of all-optical Signal Regenerators: the decision element (i.e., nonlinear gate) and the clock recovery (CR) elements.

Optical Decision Element

In the physical domain, optical decision elements with ideal step response – as for electrical DFF – do not exist. Different nonlinear optical transfer functions, approaching more or less the ideal case, can be realized in various media such as fiber, SOA, electro-absorption modulators (EAM), and lasers. Generally, as described below, the actual response (hence the regenerative properties of the device) of such optical gates directly depends upon the incoming signal instantaneous power. Under these conditions, it appears essential to add an adaptation stage so as to reduce intensity fluctuations (as caused by propagation or crossing routing/switching node) and provide the decision element with fixed power conditions. In practice, this results in the addition of a control circuitry (either optical or electrical) in the Re-amplification block, whose complexity directly depends on actual system environment (ultra-fast power equalization for packet-switching applications and compensation of slow power fluctuations in transmission applications).

As previously described the decision gate performs Re-shaping (and Re-timing when clock pulses are added) of the incoming distorted optical signal, and represent the regenerator's core element. Ideally, it should also act as a transmitter with characteristics ensuring the actual propagation of the regenerated data stream. In that respect, the chirp possibly induced by the optical decision gate onto the regenerated signal – and the initial quality of the optical clock pulses in 3R applications – should be carefully considered (ideally by means of loop transmission) as to adequately match line transmission requirements.

Different solutions for the actual realization of the optical decision element have been proposed and extensively investigated using, for example, cross gain modulation in semiconductor optical amplifier (SOA) devices but the most promising and flexible devices probably are interferometers, for which, descriptions of the generic principle of operation follows. Consider a CW signal (probe) at λ_2 wavelength injected into an optical interferometer, in which one arm incorporates a nonlinear medium in which an input signal carried by λ_1 wavelength (command) is, in turn, injected. Such a signal, at λ_1 wavelength, induces a phase shift through cross-phase modulation (XPM) in this arm of the interferometer, the amount depending upon power level $P_{in,\lambda 1}$. In turn, such phase modulation (PM) induces amplitude modulation (AM) on the signal at λ_2 wavelength when recombined at the output of the interferometer and translates the information carried by wavelength λ_1 onto λ_2. Under these conditions, such optical gates clearly act as wavelength converters (it should be mentioned that Wavelength Conversion is not necessarily equivalent to Regeneration; i.e., a linear transfer function performs suitable Wavelength Conversion but by no means Signal Regeneration).

Optical interferometers can be classified according to the nature of the nonlinearity exploited to achieve a π phase shift. In the case of fiber-based devices, such as the nonlinear optical loop mirror (NOLM), the phase shift is induced through the Kerr effect in an optical fiber. The key advantage of fiber-based devices such as NOLM lies in the near-instantaneous (fs) response of the Kerr nonlinearity, making them very attractive for ultra-high bit-rate operation ($\geq$160 Gbit/s). Polarization-insensitive NOLMs have been realized, although with the same drawbacks concerning integrability. With recent developments in highly nonlinear (HNL) fibers, however, the required NOLM fiber length could be significantly reduced, hence dramatically reducing environmental instability.

A second type of device is the integrated SOA-based Mach-Zehnder interferometers (MZI). In MZIs, the phase shift is due to the effect of photo-induced carrier depletion in the gain saturation regime of one of the SOAs. The control and probe can be launched in counter- or co-directional ways. In the first case, no optical filter is required at the output

of the device for rejecting the signal at λ_1 wavelength but operation of the MZI is limited by its speed. At this point, one should mention that the photo-induced modulation effects in SOAs are intrinsically limited in speed by the gain recovery time, which is a function of the carrier lifetime and the injection current. An approach referred to as differential operation mode (DOM) and illustrated on Figure 6, which takes advantage of the MZI's interferometric properties, makes it possible to artificially increase the operation speed of such 'slow' devices up to 40 Gbit/s.

As discussed earlier, the nonlinear response is a key parameter for regeneration efficiency. Combining two interferometers is a straightforward means to improve the nonlinearity of the decision element transfer function, and hence regeneration efficiency. This approach was validated at 40 Gbit/s using a cascade of two SOA-MZI, (see Figure 7 (**left**)). Such a scheme offers the advantage of restoring data polarity and wavelength, hence making the regenerator inherently transparent. Finally, the second conversion stage can be used as an adaptation interface to the transmission link achieved through chirp tuning in this second device.

Such an Optical 3R Regenerator was upgraded to 40 Gbit/s, using DOM in both SOA-MZIs with validation in a 40 Gbit/s loop RZ transmission. The 40 Gbit/s eye diagram monitored at the regenerator output after 1, 10, and 100 circulations are shown in Figure 7 (**right**) and remain unaltered by distance. With this all-optical regenerator structure, the minimum OSNR tolerated by the regenerator (1 dB sensitivity penalty at 10^{-10} BER) was found to be as low as 25 dB/0.1 nm. Such results clearly illustrate the high performance of this SOA-based regenerator structure for 40 Gbit/s optical data signals.

Such a complex mode of operation for addressing 40 Git/s bit-rates will probably be discarded when we consider the recent demonstration of standard-mode wavelength conversion at 40 Gbit/s, which uses a newly-designed active-passive SOA-MZI incorporating evanescent-coupling SOAs. The device architecture is flexible in the number of SOAs, thus enabling easier operation optimization and reduced power consumption, leading to simplified architectures and operation for 40 Gbit/s optical 3R regeneration.

Based on the same concept of wavelength conversion for Optical 3R Regeneration, it should be noted many devices have been proposed and experimentally validated as wavelength converters at rates up to 84 Gbit/s, but with cascadability issues still to be demonstrated to assess their actual regenerative properties.

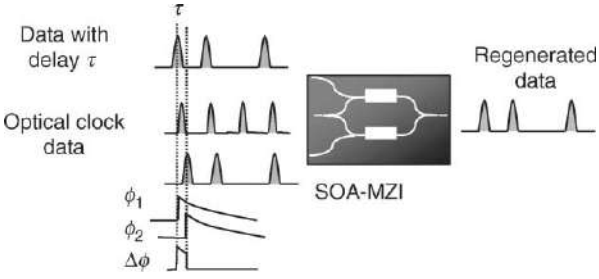

Figure 6 Schematic and principle of operation of SOA-MZI in differential mode.

Optical Clock Recovery (CR)

Next to the decision, the CR is a second key function in 3R regenerators. One possible approach for CR uses electronics while another only uses optics. The former goes with OE conversion by means of a photodiode and subsequent EO conversion through a modulator. This conversion becomes more complex and power-hungry as the data-rate increases. It is clear that the maturity of electronics gives a current advantage to this approach. But considering the pros and cons of electronic CR for cost-effective implementation, the all-optical approach seems more promising, since full regenerator integration is potentially possible with reduced power consumption. In this view, we shall focus here on the optical approach and more specifically on the self-pulsating effect in three-sections distributed feedback (DFB) lasers or more recently in distributed Bragg reflector

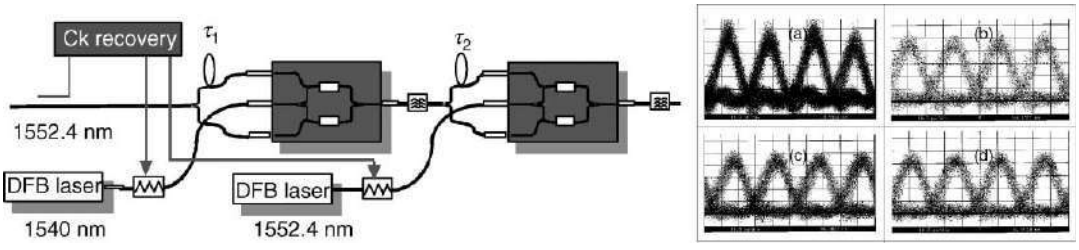

Figure 7 Optimized structure of a 40 Gbit/s SOA-based 3R regenerator. 40 Gbit/s eye diagram evolution: (a) B-to-B; (b) 1 lap; (c) 10 laps; (d) 100 laps.

(DBR) lasers. Recent experimental results illustrate the potentials of such devices for high bit rates (up to 160 Gbit/s), broad dynamic range, broad frequency tuning, polarization insensitivity, and relatively short locking times (1 ns). This last feature makes these devices good candidates for operation in asynchronous-packet regimes.

Optical Regeneration by Saturable Absorbers

We next consider saturable absorbers (SA) as nonlinear elements for optical regeneration. Figure 8 (**left**) shows a typical SA transfer function and illustrates the principle of operation. When illuminated with an optical signal with peak power below some threshold (P_{sat}), the photonic absorption of the SA is high and the device is opaque to the signal (low transmittance). Above P_{sat}, the SA transmittance rapidly increases and asymptotically saturates to transparency (passive loss being overlooked). Such a nonlinear transfer function only applies to 2R optical regeneration.

Different technologies for implementing SAs are available, but the most promising approach uses semiconductors. In this case, SA relies upon the control of carrier dynamics through the material's recombination centers. Parameters such as on–off contrast (ratio of transmittance at high and low incident powers), recovery time (1/e) and saturation energy, are key to device optimization. In the following, we consider a newly-developed ion-irradiated MQW-based device incorporated in a micro-cavity and shown on Figure 8 (**right**). The device operates as a reflection-mode vertical cavity, providing both a high on/off extinction ratio by canceling reflection at low intensity and a low saturation energy of 2 pJ. It is also intrinsically polarization-insensitive. Heavy ion-irradiation of the SA ensures recovery times (at 1/*e*) shorter than 5 ps (hence compatible with bit-rate above 40 Gbit/s), while maintaining a dynamic contrast in excess of 2.5 dB at 40 GHz repetition rate.

The regenerative properties of SA make it possible to reduce cumulated amplified spontaneous emission (ASE) in the '0' bits, resulting in a higher contrast between mark and space, hence increasing system performance. Yet SAs do not suppress intensity noise in the marks, which makes the regenerator incomplete. A solution for this noise suppression is optical filtering with nonlinear (soliton) pulses. The principle is as follows. In absence of chirp, the soliton temporal width scales in the same way as the reciprocal of its spectral width (Fourier-transform limit) times its intensity (fundamental soliton relation). Thus, an increase in pulse intensity corresponds to both time narrowing and spectral broadening. Conversely, a decrease in pulse intensity corresponds to time broadening and spectral narrowing. Thus, the filter causes higher loss when intensity increases, and lower loss when intensity decreases. The filter thus acts as an automatic power control (APC) in feed-forward mode, which causes power stabilization. The resulting 2R regenerator (composed by the SA and the optical filter) is fully passive, which is of high interest for submarine systems where the power consumption must be minimal, but it does not include any control in the time domain (no Re-timing).

System demonstrations of such passive SA-based Optical Regeneration have been reported with a 20 Gbit/s single-channel loop experiment. Implementation of the SA-based 2R Regenerator with 160 km-loop periodicity made it possible to double the error-free distance (Q = 15.6 dB or 10^{-9} BER) of a 20 Gbit/s RZ signal. So as to extend the capability of passive 2R regeneration to 40 Gbit/s systems, an improved configuration was derived from numerical optimization and experimentally demonstrated in a 40 Gbit/s WDM-like, dispersion-managed loop transmission, showing more than a fourfold increase in the WDM transmission distance at 10^{-4} BER (1650 km without the SA-based regenerator and 7600 km when implementing the 2R regenerator with 240 km periodicity).

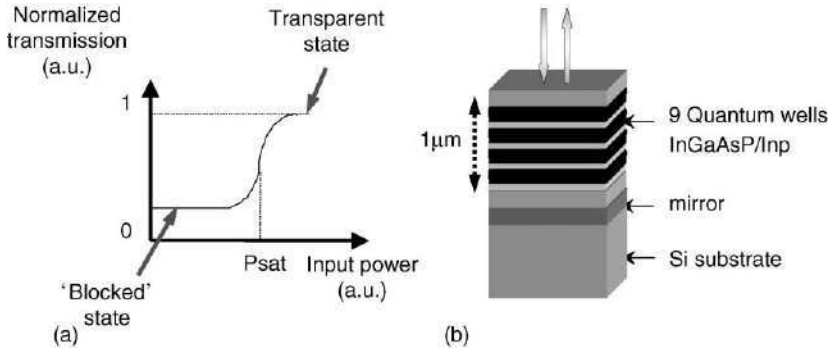

Figure 8 (a) Saturable Absorber (SA) ideal transfer function. (b) Structure of Multi-Quantum Well SA.

Such a result illustrates the potential high interest of such passive optical 2R regeneration in long-haul transmission (typically in noise-limited systems) since the implementation of SA increases the system's robustness to OSNR degradation without any extra power consumption. Reducing both saturation energy and insertion loss along with increasing dynamic contrast represent key future device improvements. Regeneration of WDM signals from the same device, such as one SA chip with multiple fibers implemented between Mux/DMux stages, should also be thoroughly investigated. In this respect, SA wavelength selectivity in quantum dots could possibly be advantageously exploited.

Synchronous Modulation Technique

All-optical 3R regeneration can be also achieved through in-line synchronous modulation (SM) associated with narrowband filtering (NF). Figure 9 shows the basic layout of such an Optical 3R Regenerator. It is composed of an optical filter followed by an Intensity and Phase Modulator (IM/PM) driven by a recovered clock. Periodic insertion of SM-based modulators along the transmission link provides efficient jitter reduction and asymptotically controls ASE noise level, resulting in virtually unlimited transmission distances. Re-shaping and Re-timing provided by IM/PM intrinsically requires nonlinear (soliton) propagation in the trunk fiber following the SM block. Therefore, one can refer to the approach as distributed optical regeneration. This contrasts with lumped regeneration, where 3R is completed within the regenerator (see above with Optical Regeneration using nonlinear gates), and is independent of line transmission characteristics.

However, when using a new approach referred to 'black box' optical regeneration (BBOR), it is possible to make the SM regeneration function and transmission work independently in such a way that any type of RZ signals (soliton or non-soliton) can be transmitted through the system. The BBOR technique includes an adaptation stage for incoming RZ pulses in the SM-based regenerator, which ensures high regeneration efficiency regardless of RZ signal format (linear RZ, DM-soliton, C-RZ, etc.). This is achieved using a local and periodic soliton conversion of RZ pulses by means of launching an adequate power into some length of fiber with anomalous dispersion. The actual experimental demonstration of the BBOR approach and its superiority over the 'classical' SM-based scheme for DM transmission was experimentally investigated in 40 Gbit/s DM loop transmission. Under these conditions, one can then independently exploit dispersion management (DM) techniques for increasing spectral efficiency in long-haul transmission, while ensuring high transmission quality through BBOR.

One of the key properties of the SM-based all-optical Regeneration technique relies on its WDM compatibility. The first (Figure 10, **left**) and straightforward solution to apply Signal Regeneration to WDM channels amounts to allocating a regenerator to each WDM channel. The second consists in sharing a single modulator, thus processing the WDM channels at once in serial fashion. This approach requires WDM synchronicity, meaning that all bits be synchronous with the modulation, that can be achieved either by use of appropriate time-delay lines located within a DMux/Mux apparatus (Figure 10, **upper right**), or by making the WDM channels inherently time-coincident at specific regenerator locations (Figure 10, **bottom right**). Clearly, the serial regeneration scheme is far simpler and cost-effective than the parallel version; however, optimized re-synchronization schemes still remain to be

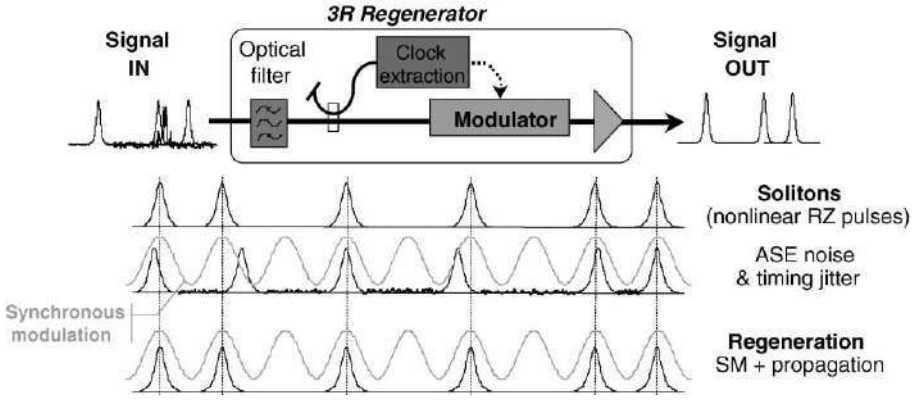

Figure 9 Basic layout of the all-optical Regenerator by Synchronous Modulation and Narrowband Filtering and illustration of the principle of operation.

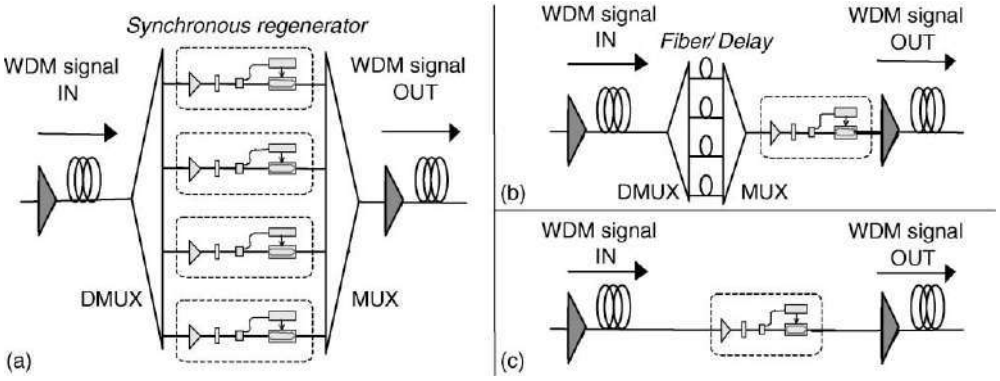

Figure 10 Basic implementation schemes for WDM all-Optical Regeneration. (a) parallel asynchronous; (b) serial re-synchronized; (c) serial self-synchronized.

developed for realistic applications. Experimental demonstration of this concept was assessed by means of a 4×40 Gbit/s dispersion-managed transmission over 10 000 km (BER $< 5.10^{-8}$) in which a single modulator was used for the simultaneous regeneration of the 4 WDM channels.

Considering next all-optical regeneration schemes with ultra-high speed potential, a compact and loss-free 40 Gbit/s Synchronous Modulator, based on optically-controlled SOA-MZI, was proposed and loop-demonstrated at 40 Gbit/s with an error-free transmission distance in excess of 10 000 km. Moreover, potential ultra-high operation of this improved BBOR scheme was recently experimentally demonstrated by means of a 80 GHz clock conversion with appropriate characteristics through the SOA-MZI. One should finally mention all fiber-based devices such as NOLM and NALM for addressing ultra-high speed SM-based Optical Regeneration, although no successful experimental demonstrations have been reported so far in this field.

Conclusion

In summary, optical solutions for Signal Regeneration present many key advantages. These are the only advantages to date to possibly ensure WDM compatibility of the regeneration function (mostly 2R related). Such optical devices clearly exhibits the best 2R regeneration performance (wrt to O/E solutions) as a result of the moderately nonlinear transfer function (which in turn can be considered as a drawback in 3R applications), but the optimum configuration is still to be clearly derived and identified depending upon the system application. Optics also allow to foresee and possibly target ultrafast applications above 40G, for signal regeneration if needed. Among the current drawbacks, one should mention the relative lack of wavelength/formats flexibility of these solutions (compared to O/E solutions). It is complex or difficult to restore the input wavelength or address any C-band wavelength at the output of the device or to successfully regenerate modulation formats other than RZ. In that respect, investigations should be conducted to derive new optical solutions capable of processing more advanced modulation formats at 40G. Finally, the fact that the nonlinear transfer function of the optical is in general triggered by the input signal instantaneous power also turns out to be a drawback since it requires control circuitry. The issue of cost (footprint, power consumption, etc.) of these solutions, compared to O/E ones, is still open. In this respect, purely optical solutions incorporating all-optical clock recovery, the performance of which is still to be technically assessed, are of high interest for reducing costs. Complete integration of an all-optical 2R/3R regenerator or such parallel regenerators onto a single semiconductor chip should also contribute to make all-optical solutions cost-attractive even though acceptable performance of such fully integrated devices is still to be demonstrated.

From today's status concerning the two alternative approaches for in-line regeneration (O/E or all-optical), it is safe to say that the choice between either solution will be primarily dictated by both engineering and economical considerations. It will result from a tradeoff between overall system performance, system complexity and reliability, availability, time-to-market, and rapid returns from the technology investment.

See also

Interferometry: Overview. **Optical Amplifiers:** Semiconductor Optical Amplifiers. **Optical Communication Systems:** Wavelength Division Multiplexing.

Further Reading

Bigo S, Leclerc O and Desurvire E (1997) All-optical fiber signal processing for solitons communications. *IEEE Journal of Selected Topics on Quantum Electronics* 3(5): 1208–1223.

Dagens B, Labrousse A, Fabre S, *et al.* (2002) New modular SOA-based active–passive integrated Mach–Zehnder interferometer and first standard mode 40 Gb/s all-optical wavelength conversion on the C-band. paper PD 3.1, Proceedings of ECOCOL 02, Copenhague.

Durhuus T, Mikkelsen B, Joergensen C, Danielsen SL and Stubkjaer KE (1996) All-optical wavelength conversion by semiconductor optical amplifiers. *Journal of Lightwave Technology* 14(6): 942–954.

Lavigne B, Guerber P, Brindel P, Balmefrezol E and Dagens B (2001) Cascade of 100 optical 3R regenerators at 40 Gbit/s based on all-active Mach–Zehnder interferometers. paper We.F.2.6, Proceedings of ECOC 2001, Amsterdam.

Leclerc O, Lavigne B, Chiaroni D and Desurvire E (2002) All-optical regeneration: Principles and WDM implementation. In: Kaminow IP and Li T (eds) *Optical Fiber Technologies IVA*, chap. 15, pp. 732–784. San Diego: Academic Press.

Leuthold J, Raybon G, Su Y, *et al.* (2002) 40 Gbit/s transmission and cascaded all-optical wavelength conversion over 1,000,000 km. *Electronic Letters* 38(16): 890–892.

Öhlén P and Berglind E (1997) Noise accumulation and BER estimates in concatenated nonlinear optoelectronic repeaters. *IEEE Photon Technol. Letters* 9(7): 1011–1013.

Otani T, Suzuki M and Yamamoto S (2001) 40 Gbit/s optical 3R regenerator for all-optical networks. We.F.2.1, Proceedings of ECOC 2001, Amsterdam.

Oudar J-L, Aubin G, Mangeney J, *et al.* (2004) Ultra-fast quantum-well saturable absorber devices and their application to all-optical regeneration of telecommunication optical signals. *Annales des Télécommunications* 58(11–12): 1667–1707.

Sartorius B, Mohrle M, Reichenbacher S, *et al.* (1997) Dispersive self-Q-switching in self-pulsating DFB lasers. *IEEE Journal of Quantum Electronics* 33(2): 211–218.

Ueno Y, Nakamura S, Sasaki J, *et al.* (2001) Ultrahigh-speed all-optical data regeneration and wavelength conversion for OTDM systems. Th.F.2.1, Proceedings of ECOC 2001, Amsterdam.

B

BABINET'S PRINCIPLE

B D Guenther, Duke University, Durham, NC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

In the development of the theory of diffraction, the diffraction field is due to a surface integral, the Fresnel integral, but no restrictions are imposed on the choice of the surface over which the integration must be performed. This fact leads to a very useful property called Babinet's principle, first stated by Jacques Babinet (1794–1872) in 1837 for scalar waves. We will discuss only the scalar Babinet's principle; discussion of the rigorous vector formulation can be found in a number of books on electromagnetic theory.

To introduce Babinet's principle, we label a plane, separating the source, S, and observation point, P, as Σ. If no obstructions are present, a surface integration over Σ yields the light distribution at P. If we place an opaque obstruction in this plane with a clear aperture, Σ_1, then the field at P is given by integrating over only Σ_1; contributions from Σ, outside of Σ_1, are zero since the obstruction is opaque.

We may define an aperture, Σ_2, as complementary to Σ_1 if the obstruction is constructed by replacing the transparent regions of Σ, i.e., Σ_1, by opaque surfaces and the opaque regions of Σ by clear apertures. Figure 1 shows two complementary obstructions, where the shaded region indicates an opaque surface.

The surface integral over Σ generates the field, E, in the absence of any obstruction. If obstruction Σ_1 is present then the diffraction field is E_1, obtained by integrating over Σ_1. According to Babinet's principle, the diffraction field due to obstruction Σ_2, must be

$$\tilde{E}_2 = \tilde{E} - \tilde{E}_1$$

We will look at examples of the application of the principle for both Fraunhofer and Fresnel diffraction.

Fraunhofer Diffraction

The electric field due to Fraunhofer diffraction is given by

$$E_{\mathrm{P}}(\omega_x, \omega_y) = \frac{\mathrm{i}\alpha \mathrm{e}^{-\mathrm{i}kR_0}}{\lambda R_0} \iint_{\Sigma} f(x, y) \mathrm{e}^{-\mathrm{i}(\omega_x x + \omega_y y)} \, \mathrm{d}x \, \mathrm{d}y$$

where α is the amplitude of the incident plane wave and R_0 is the distance from the obstruction to the observation point. We have defined the spatial frequencies, ω_x and ω_y by the equations

$$\omega_x = kL = -\frac{2\pi\xi}{\lambda R_0} \qquad \omega_y = kM = -\frac{2\pi\eta}{\lambda R_0}$$

This surface integral is an approximation of the Hygens–Fresnel integral and can be identified as a Fourier transform of the aperture transmission function, $f(x, y)$ in two dimensions. For our discussion we will consider only one dimension and ignore the parameters in front of the integral.

Assume that $a(x)$ is the amplitude transmission of an aperture in an obstruction, $b(x)$ is the amplitude transmission function of the complementary obstruction and g is the amplitude of the wave when no

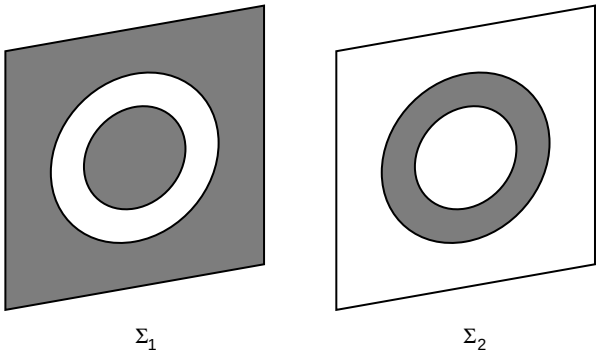

Figure 1 An aperture Σ_1 shown by the unshaded region and its complement Σ_2. Reproduced with permission from Guenther R (1990) *Modern Optics*. New York: Wiley. Copyright © 1990 John Wiley & Sons.

obstruction is present. The Fourier transforms of the one-dimensional amplitude transmission functions of the two apertures are equal to the Fraunhofer diffraction patterns that will be generated by the two apertures

$$A(k) = \int_{-\infty}^{\infty} a(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x$$

$$B(k) = \int_{-\infty}^{\infty} b(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x$$

With no aperture present, the far field amplitude is

$$G(k) = g\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x$$

Babinet's principle states that

$$B(x) = G - A(x)$$

must be the Fraunhofer diffraction field of the complementary aperture. We may rewrite this equation for the diffraction field as

$$B(k) = G\delta(k) + A(k)\mathrm{e}^{\mathrm{i}\pi} \qquad [1]$$

The first term on the right of the Fraunhofer diffraction field for the complementary obstruction [1] is located at the origin of the observation plane and is proportional to the amplitude of the unobstructed wave. The second term in the equation for the Fraunhofer diffraction pattern [1] is identical to the Fraunhofer diffraction pattern of the original aperture, except for a constant phase. Thus the Fraunhofer diffraction from the two complementary apertures are equal, except for a constant phase and the bias term at the origin. Physically this means that the diffraction intensity distributions of complementary apertures will be identical but their brightness will differ!

Fresnel Diffraction

We can calculate the Fresnel diffraction from an opaque disk by applying Babinet's principle to the diffraction pattern calculated for a circular aperture.

We assume that a circular aperture of radius a is illuminated by a point source a distance Z' from the aperture. We observe the transmitted wave at the point P, located a distance Z from the aperture (see Figure 2).

The Fresnel diffraction integral is

$$\tilde{E}_{\mathrm{P}} = \frac{\mathrm{i}\alpha}{\lambda\rho D}\mathrm{e}^{-\mathrm{i}kD} \times \int\int f(x,y)\mathrm{e}^{-\frac{\mathrm{i}k}{2\rho}\left[(x-x_0)^2+(y-y_0)^2\right]}\,\mathrm{d}x\,\mathrm{d}y \qquad [2]$$

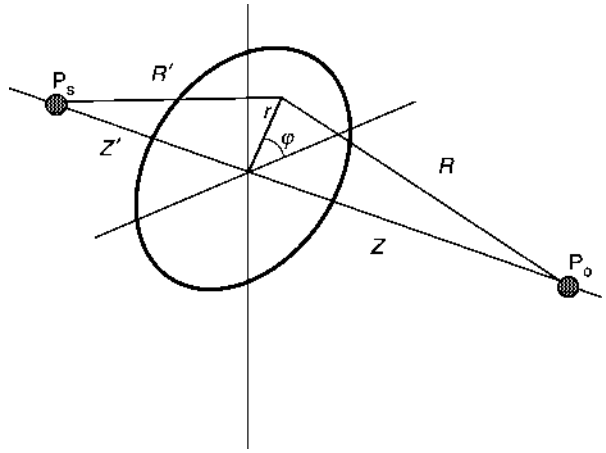

Figure 2 Geometry for the analysis of Fresnel diffraction of a circular aperture. Reproduced with permission from Guenther R (1990) *Modern Optics*. New York: Wiley. Copyright © 1990 John Wiley & Sons.

where the distance between the source and observation point is

$$D = Z + Z' + \frac{(\xi - x_s)^2 + (\eta - y_s)^2}{2(Z+Z')} \qquad \frac{1}{\rho} = \frac{1}{Z} + \frac{1}{Z'}$$

and the coordinates of the stationary point are

$$x_0 = \frac{Z'\xi + Zx_s}{Z+Z'} \qquad y_0 = \frac{Z'\eta + Zy_s}{Z+Z'}$$

The parameters of the integral are simplified by the geometry selected for Figure 2. Both the observation point and source are located on the axis of symmetry of the aperture which lies along the z-axis of a cylindrical coordinate system thus

$$x_0 = y_0 = \xi = \eta = 0 \qquad \text{and} \qquad D = Z' + Z$$

The approximation for the distance from the source and observation point to the aperture yields

$$R + R' \approx Z + Z' + \frac{r^2}{2\rho} = D + \frac{r^2}{2\rho}$$

where $r^2 = x^2 + y^2$. The cylindrical coordinate version of [2] is then

$$\tilde{E}_{\mathrm{P}} = \frac{\mathrm{i}\alpha}{\lambda\rho D}\mathrm{e}^{-\mathrm{i}kD}\int_0^a\int_0^{2\pi} f(r,\varphi)\mathrm{e}^{-\mathrm{i}\frac{kr^2}{2\rho}}\,r\,\mathrm{d}r\,\mathrm{d}\phi \qquad [3]$$

We assume that the transmission function of the aperture is a constant, $f(r,\phi) = 1$, to make the integral as simple as possible. After performing the integration over the angle ϕ, [3] can be

rewritten as

$$\tilde{E}_{\mathrm{Pa}} = \frac{\mathrm{i}\alpha\pi}{D}\mathrm{e}^{-\mathrm{i}kD}\int_0^{\frac{a^2}{\rho\lambda}} \mathrm{e}^{-\mathrm{i}\pi r^2/\rho\lambda}\,\mathrm{d}\left(\frac{r^2}{\rho\lambda}\right) \qquad [4]$$

Performing the integration of [4], results in the Fresnel diffraction amplitude

$$\tilde{E}_{\mathrm{Pa}} = \frac{\alpha}{D}\mathrm{e}^{-\mathrm{i}kD}\left[1 - \mathrm{e}^{-\mathrm{i}\pi a^2/\rho\lambda}\right] \qquad [5]$$

The intensity of the diffraction field is

$$\begin{aligned} I_{\mathrm{Pa}} &= 2I_0\left(1 - \cos\frac{\pi a^2}{\rho\lambda}\right) \\ &= 4I_0\sin^2\frac{\pi a^2}{2\rho\lambda} \end{aligned} \qquad [6]$$

[6] predicts a sinusoidal variation in intensity as the observation point moves toward the aperture.

Babinet's principle can now be used to evaluate Fresnel diffraction by an opaque disk, the same size as the circular aperture, i.e., of radius a. The field for the disk is obtained by subtracting the field, diffracted by a circular aperture, from the field of an unobstructed spherical wave, E_{P}:

$$\begin{aligned} \tilde{E}_{\mathrm{Pd}} &= \tilde{E}_{\mathrm{P}} - \tilde{E}_{\mathrm{Pa}} \\ \tilde{E}_{\mathrm{Pd}} &= \frac{\alpha}{D}\mathrm{e}^{-\mathrm{i}kD} - \tilde{E}_{\mathrm{Pa}} \\ &= \frac{\alpha}{D}\mathrm{e}^{-\mathrm{i}kD}\mathrm{e}^{-\mathrm{i}\pi a^2/\rho\lambda} \end{aligned} \qquad [7]$$

The intensity is

$$I_{\mathrm{Pd}} = I_0$$

The result of the application of Babinet's principle is the conclusion that, at the center of the geometrical shadow cast by the disk, there is a bright spot with the same intensity as would exist if no disk were present. This is Poisson's spot. The intensity of this spot is independent of the choice of the observation point along the z-axis.

See also

Diffraction: Fraunhofer Diffraction; Fresnel Diffraction.

Further Reading

Abramowitz M and Stegun IA (1964) *Handbook of Mathematical Functions*. NBS Applied Mathematics Series-55 Washington, DC: National Bureau of Standards.

Born M and Wolf E (1959) *Principles of Optics*. New York: Pergamon Press.

Guenther R (1990) *Modern Optics*. New York: Wiley.

Haus HA (1984) *Waves and Fields in Optoelectronics*. Englewood Cliffs, NJ: Prentice-Hall.

Hecht E (1998) *Optics*, 3rd edn. Reading, MA: Addison-Wesley.

Klein MV (1970) *Optics*. New York: Wiley.

Rossi B (1957) *Optics*. Reading, MA: Addison-Wesley.

Sommerfeld A (1964) *Optics*. New York: Academic Press.

C

CHAOS IN NONLINEAR OPTICS

R G Harrison and W Lu, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Instabilities in laser emission, notably in the form of spontaneous coherent pulsations, have been observed almost since the first demonstration of laser action. Indeed, the first laser operated in 1960 by Maiman produced noisy spiked output even under conditions of quasi-steady excitation and provided the early impetus for studies of such effects. Subsequent theoretical efforts towards understanding these phenomena occurred up to the 1980s at a modest level, due in part to the wide scope of alternative areas of fertile investigation provided by lasers during this period. However, since the 1990s, there has been a major resurgence of interest in this area. This has been due to profound mathematical discoveries over this period which have revolutionized our understanding of classical dynamical systems. It is now clear that many systems containing some form of nonlinearity are dynamically unstable and even chaotic and that such behavior is deterministic. Further, the discovery that chaos evolves through particular routes with well-defined scenarios and that such routes are universal, has stimulated experimentalists to search for physical systems that exhibit these properties. These phenomena have now been observed in areas as diverse as fluid flow, chemical reactions, population ecology, and superconductor devices. The laser, perhaps the most recent newcomer to the field, is a paradigm for such investigation, owing to its simplicity both in construction and in the mathematics that describe it, and already a wide range of these phenomena have been observed. Along with a proliferation of new observations, many of the instabilities known to occur in these systems are now being reinterpreted in lieu of our new insight in this area. From this, fascinating new concepts of control and synchronization of chaos have emerged and spawned new fields of applications, not least in secure communication. In this article we introduce the general principles of dynamical instabilities, chaos, and control in lasers. For in-depth study, the reader is referred to texts and review articles cited in the Further Reading section at the end of this article.

Laser Physics

Our conventional understanding of laser physics is concerned with how cooperative order evolves from randomness. This transition is explained by first considering the ensemble of lasing atoms in thermal equilibrium. Every atom executes a small motion, giving rise to an oscillating electric dipole, described by linear dynamics. For larger motions, the atomic dipoles interfere with each other's motion and beyond a particular threshold, the motion becomes 'co-operative' or ordered over a long range. The role of the electromagnetic field in the laser cavity in ordering the induced atomic dipoles, is collectively described as giving rise to a macroscopic polarization, the magnitude of which depends on the number of dipoles (excited atoms). The dynamical interplay between the cavity field amplitude (E) as one variable and the atomic material variables of polarization (P) and population (D) of excited atoms, all of which are in general complex quantities, provide a full description of lasing action. For the simplest laser, a single mode two-level system with a homogeneously broadened gain lasing at resonance, these reduce to just three real variables, the Maxwell–Bloch equations:

$$E = -\kappa E + \kappa P \qquad [1]$$

$$P = \gamma_{\perp} ED - \gamma_{\perp} P \qquad [2]$$

$$D = \gamma_{II}(\lambda + 1) - \gamma_{II} D - \gamma_{II} \lambda EP \qquad [3]$$

where κ is the cavity decay rate, $\gamma_\perp$ is the decay rate of atomic polarization, γ_{II} is the decay rate of population inversion, and λ the pumping parameter.

Described in this way the laser acts as a single oscillator in much the same way as a mechanical nonlinear oscillator, where laser output represents damping of the oscillator and excitation of the atoms is the driving mechanism to sustain lasing (oscillation). For ordered or coherent emission, it is necessary for one or both of the material variables (which are responsible for further lasing emission) to respond sufficiently fast to ensure a phase correlation with the existing cavity field. This is readily obtained in many typical lasers with output mirrors of relatively high reflectivity, since the field amplitude within the laser cavity will then vary slowly compared to the fast material variables which may then be considered through their equilibrium values. This situation, commonly referred to as adiabatic elimination of fast variables, reduces the dynamics of lasing action to that of one (or two) variables, the field (and population), all others being forced to adapt constantly to the slowly varying field variable. Our familiar understanding of DC laser action presupposes such conditions to hold. However when, for example, the level of excitation of the atoms is increased to beyond a certain value, i.e., the second laser threshold, all three dynamical variables may have to be considered, which satisfies the minimum requirement of three degrees of freedom (variables) for a system to display chaos. So this simple laser is capable of chaotic motion, for which the emission is aperiodic in time and has a broadband spectrum. Prediction of such behavior was initially made by Haken in 1975 through establishing the mathematical equivalence of the Maxwell–Bloch equations to those derived earlier by Lorentz describing chaotic motion in fluids.

Nonlinear Dynamics and Chaos in Lasers

In general there is no general analytical approach for solving nonlinear systems such as the Maxwell–Bloch equations. Instead solutions are obtained by numerical means and analyzed through geometric methods originally developed by Poincaré as early as 1892. This involves the study of topological structures of the dynamical trajectories in phase space of the variables describing the system. If an initial condition of a dissipative nonlinear dynamical system, such as a laser, is allowed to evolve for a long time, the system, after all the transients have died out, will eventually approach a restricted region of the phase space called an attractor. A dynamical system can have more than one attractor, in which case different initial conditions lead to different types of long-time behavior. The simplest attractor in phase space is a fixed point which is a solution with just one set of values for its dynamical variables; the nonlinear system is attracted towards this point and stays there, giving a DC solution. For other control conditions the system may end up making a periodic motion; the attractor of this motion is called a limit cycle. However, when the operating conditions exceed a certain critical value, the periodic motion of the system breaks down into a more complex dynamical trajectory, which never repeats. This motion represents a third kind of attractor in phase space called a chaotic or strange attractor.

Figure 1 shows a sequence of trajectories of the Lorentz equations in the phase space of (x, y, z) on increasing the value of one of the control parameters; this corresponds to (E, P, D) in the Maxwell–Block equations on increasing the laser pump. For a control setting near zero, all trajectories approach stable equilibrium at the origin, the topological structure of the basin of attraction being hyperbolic about zero (Figure 1a). For the laser equations, this corresponds to operation below lasing threshold for which the magnitude of the control parameter (laser gain) is insufficient to produce lasing. As the control parameter is increased, the basin lifts at its zero point to create an unstable fixed point here but with now two additional fixed points located in the newly formed troughs on either side of the zero point, which is now a saddle point. This is illustrated in Figure 1b where the system eventually settles to one or other of the two new and symmetrical fixed points, depending on the initial conditions. This corresponds to DC or constant lasing as defined by the parameter values of one or other of these points which are indistinguishable. Pictorially, think of a conventional saddle shape comprising a hyperbolic and inverted hyperbolic form in mutually perpendicular planes and connected tangentially at the origin. With the curve of the inverted hyperbola turned up at its extremity and filling in the volume with similar profiles which allow the two hyperbolic curves to merge into a topological volume, one sees that a ball placed at the origin is constrained to move most readily down either side of the inverted hyperbolas into one or other of the troughs formed by this volume. For the Lorentz system, chaotic behavior occurs at larger values of the control parameter when all three fixed points become saddles. Since none of the equilibrium points are now attracting, the behavior of the system cannot be a steady motion. Although perhaps difficult to visualize topologically, it is then possible to find a region in this surface enclosing all three points and

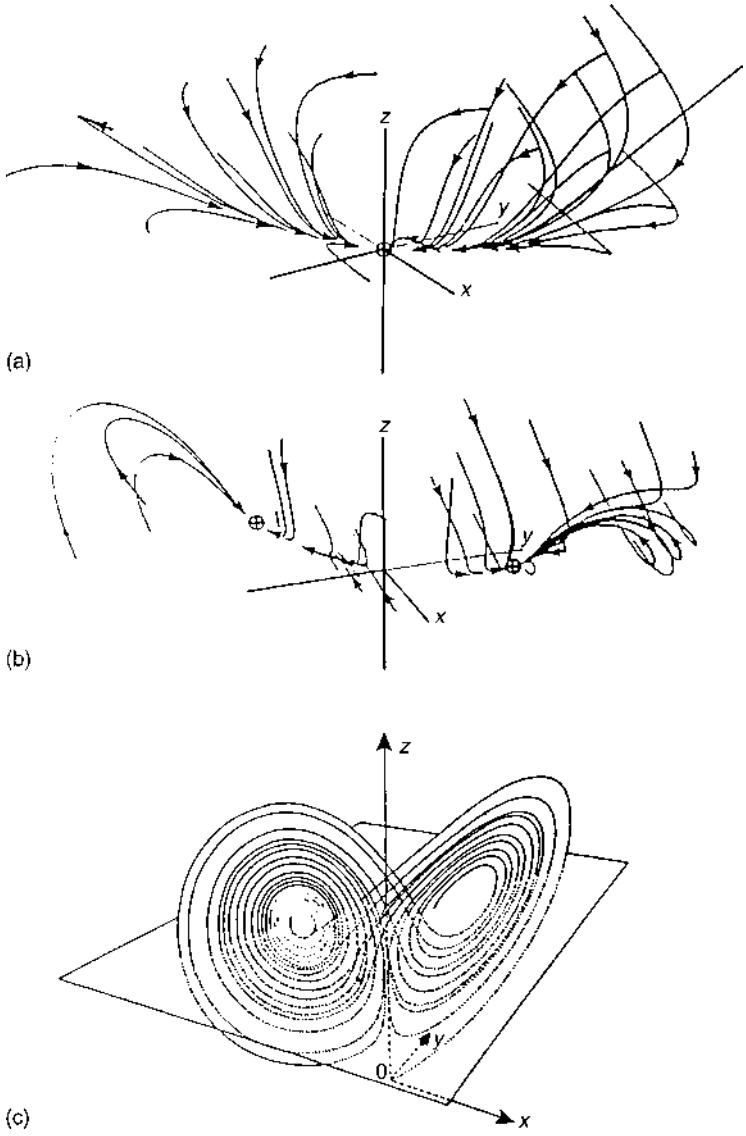

Figure 1 Trajectories in the three-dimensional phase space of the Lorentz attractor on increasing one of the control parameters. (Reproduced with permission from Thompson JMT and Stewart HB (1987) *Nonlinear Dynamics and Chaos*. New York: John Wiley.)

large enough so that no trajectories leave the region. Thus, all initial conditions outside the region evolve into the region and remain inside from then on. A corresponding chaotic trajectory is shown in Figure 1c. A point outwardly spirals from the proximity of one of the new saddle points until the motion brings it under the influence of the symmetrically placed saddle, the trajectory then being towards the center of this region from where outward spiraling again occurs. The spiraling out and switching over continues forever though the trajectory never intersects. In this case, arbitrarily close initial conditions lead to trajectories, which, after a sufficiently long time, diverge widely. Since a truly exact assignment of the initial condition is never possible, even numerically, a solution comprising several such trajectories therefore evolves and, as a consequence, long-term predictability is impossible. This is in marked contrast to that of the fixed point and limit cycle attractors, which settle down to the same solutions. A recording of one of these variables in time, say for a laser the output signal amplitude (in practice recorded as signal intensity, proportional to the square of the field amplitude), then gives oscillatory emission of increasing intensity with sudden discontinuities (resulting from flipping from one saddle to the other in Figure 1c) as expected.

While the Lorentz/Haken model is attractive for its relative simplicity, many practical lasers cannot be reduced to this description. Nevertheless, the

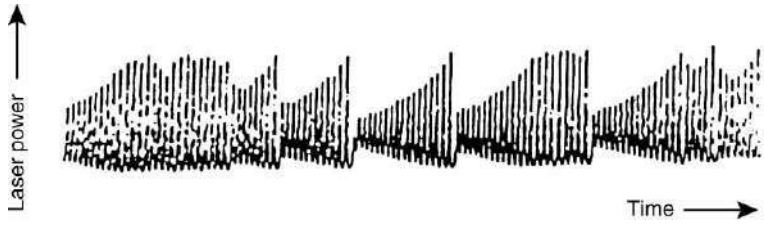

Figure 2 Lorentz-type chaos in the NH_3 laser (emitting at 81 μm optically pumped by a N_2O laser.

underlying topology of the trajectory of solutions in phase space for these systems is often found to be relatively simple and quite similar, as for three-level models descriptive of optically pumped lasers. An experimental example of such behavior for a single mode far infared molecular laser is shown in Figure 2. In other lasers, for which there is adiabatic elimination of one or other of the fast variables, chaotic behavior is precluded if, as a consequence, the number of variables of the system is reduced to less than three. Indeed this is the case for many practical laser systems. For these systems the addition of an independent external control parameter to the system, such as cavity length or loss modulation, have been extensively used as a means to provide the extra degrees of freedom. Examples include gas lasers with saturable absorbers, solid state lasers with loss modulation, and semiconductor lasers with external cavities, to name but a few. In contrast, for multimode rather than single mode lasers, intrinsic modulation of inversion (or photon flux) by multimode parametric interaction ensures the additional degrees of freedom. Furthermore, when the field is detuned from gain center, the dynamical variables E, P, D are complex, providing five equations for single mode systems, which is more than sufficient to yield deterministic chaos for suitable parameter values. Also, of significance is the remarkably low threshold found for the generation of instabilities and chaos in single-mode laser systems in which the gain is inhomogeneously broadened, an example being the familiar He–Ne laser.

Erratic and aperiodic temporal behavior of any of the system's variables implies a corresponding continuous spectrum for its Fourier transform, which is thus a further signature of chaotic motion. Time series, power spectra, and routes to chaos collectively provide evidence of deterministic behavior. Of the wide range of possible routes to chaos, three have emerged as particularly common and are frequently observed in lasers. These are period doubling, intermittency, and two-frequency scenarios. In the first, a solution, which is initially stable is found to oscillate, the period of which successively doubles at distinct values of the control parameter. This continues until the number of fixed points becomes infinite at a finite parameter value, where the variation in time of the solution becomes irregular. For the intermittency route, a signal that behaves regularly in time becomes interrupted by statistically distributed periods of irregular motion, the average number of which increases with the external control parameter until the condition becomes chaotic. The two-frequency route is more readily identified with early concepts of turbulence, considered to be the limit of an infinite sequence of instabilities (Hopf bifurcation) evolving from an initial stable solution each of which creates a new basic frequency. It is now known that only two or three instabilities (frequencies) are sufficient for the subsequent generation of chaotic motion.

Applications of Chaos in Lasers

It has been accepted as axiomatic since the discovery of chaos that chaotic motion is in general neither predictable nor controllable. It is unpredictable because a small disturbance will produce exponentially growing perturbation of the motion. It is uncontrollable because small disturbances lead to other forms of chaotic motion and not to any other stable and predictable alternative. It is, however, this very sensitivity to small perturbations that has been more recently used to stabilize and control chaos, essentially using chaos to control itself. Among the many methods proposed in the late 1980s, a feedback control approach was proposed by Ott, Grebogi, and Yorke (OGY) in which tiny feedback was used to stabilize unstable periodic orbits or fixed points of chaotic attractors. This control strategy can be best understood by a schematic of the OGY control algorithm for stabilizing a saddle point P^*, as shown in Figure 3. Curved trajectories follow a stable (unstable) manifold towards (away from) the saddle point. Without perturbation of a parameter, the starting state, s_1, would evolve to the state s_2.

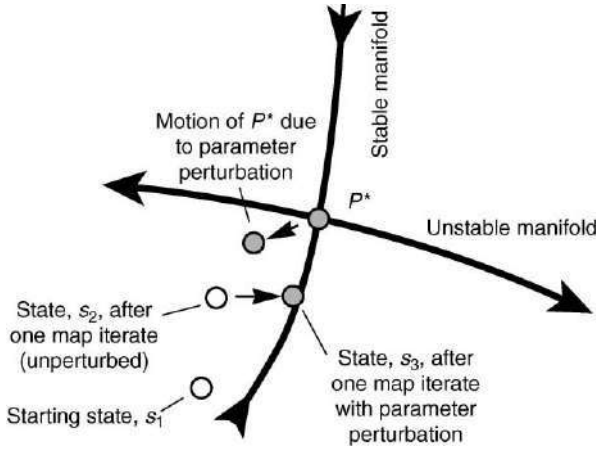

Figure 3 The OGY control strategy for stabilizing an orbit in a chaotic attractor.

The effect of changing a parameter of the system is depicted as shifting states near P^* along the solid black arrows, whereas the combination of the unperturbed trajectory and the effect of the perturbation is to bring the state to a point, s_3, on the stable manifold. Once on the stable manifold, the trajectory naturally tends towards the desired point. This algorithm has been successfully applied in numerous experimental systems without *a priori* modeling of these systems, examples being in cardiology, electronics, and lasers. Based on the concept of feedback control, various other approaches have been developed where emphasis has been given to algorithms which are more readily implemented in practical systems, in particular that utilizing occasional proportional feedback (OPF). These pioneering studies have since inspired prolific activities, both in theory and experiment, of control of chaos across many disciplines, opening up possibilities of utilizing chaos in many diverse systems. In optics, Roy *et al.* first demonstrated dynamical control of an autonomously chaotic and high-dimensional laser system on microsecond time-scales. The laser, a diode pumped solid-state Nd:YAIG system with an intracavity KTP doubling crystal, exhibited chaotic behavior from coupling of the longitudinal modes through nonlinear sum-frequency generation. To control the system, the total laser output was sampled within a window of selected offset and width. A signal proportional the deviation of the sampled intensity from the center of the window was generated and applied to perturb the driving current of the diode laser. The sampling frequency is related to the relaxation oscillation frequency of the system. This control signal repeatedly attempts to bring the system closer to a periodic unstable orbit that is embedded in the chaotic attractor, resulting in a realization of the periodic orbit. By adjusting the frequency of the feedback signal they observed the basic (relaxation oscillation) and many higher-order periodic waveforms of the laser intensity. In a subsequent experiment, Gills *et al.* showed that an unstable steady state of this laser could also be stabilized by the OPF control technique and tracked with excellent stability to higher output power with increase of the pump excitation.

Another area to receive considerable attention in the last few years is that of synchronization of chaos. It may be expected that chaotic systems will defy synchronization because two such identical systems started at nearly the same initial conditions have dynamical trajectories that quickly becomes uncorrelated. However, it has now been widely demonstrated that when these systems are linked, their chaotic trajectories converge to be the same and remain in step with each other. Further, such synchronization is found to be structurally stable and does not require the systems to be precisely identical. Not surprisingly, these findings have attracted interest from the telecommunication community; the natural masking of information by chaotic fluctuations offering a means to a certain degree of security. In this new approach, a chaotic carrier of information can be considered as a generalization of the more traditional sinusoidal carrier. In communication systems that use chaotic waveforms, information can be recovered from the carrier using a receiver synchronized, or tuned, to the dynamics of the transmitter. Optical systems are particularly attractive since they display fast dynamics, offering the possibility of communication at bandwidths of a hundred megahertz or higher. Van Wiggeren and Roy first demonstrated data transmission rates of 10 Mbites per second using erbium-doped fiber ring lasers. These lasers are particularly well-suited for communication purposes since their lasing wavelength is close to the minimum-loss wavelength in optical fiber. Figure 3 shows a schematic of this system. The tiny message, in order of one thousandth of the carrier intensity, is decoded in the transmitter. The receiver tends to synchronize its behavior to the chaotic part of the transmitted wave (but not the message). Subtracting the waveform created in the receiver from the transmitted signal recovers the tiny message. In principle, it is possible to communicate information at ultrahigh data rate with the use of this scheme because the chaotic dynamics in the ring laser has a very large spectral width. In a later experiment they showed that a receiver can recover information at 126 Mbits/sec from the chaotic carrier (Figure 4).

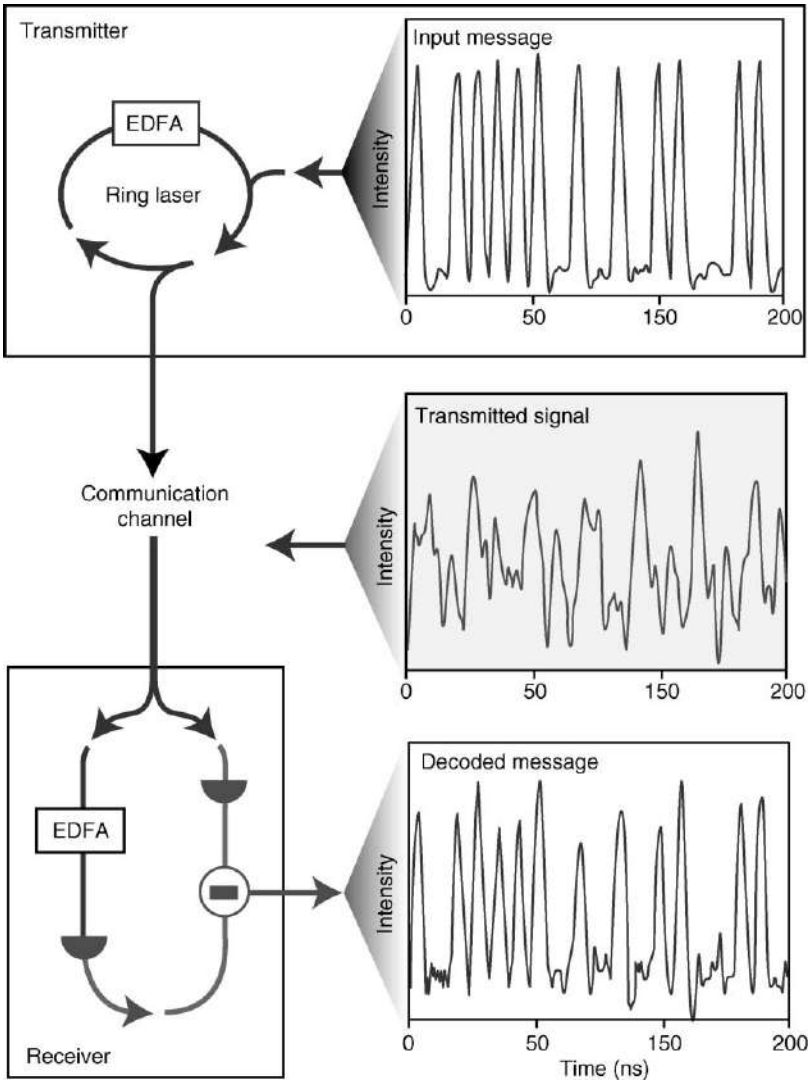

Figure 4 Schematic of communication with chaotic erbium-doped fiber amplifiers (EDFAs). Injecting a message into the transmitter laser folds the data into the chaotic frequency fluctuations. The receiver reverses this process, thereby recovering a high-fidelity copy of the message. (Reproduced with permission from Gauthier DJ (1998) Chaos has come again. *Nature* 279: 1156–1157.)

See also

Fourier Optics. **Optical Amplifiers:** Erbrium Doped Fiber Amplifiers for Lightwave Systems. **Polarization:** Introduction.

Further Reading

Abraham NB, Lugiato LA and Narducci LM (eds) (1985) Instabilities in active optical media. *Journal of the Optical Society of America B* 2: 1–272.

Arecchi FT and Harrison RG (1987) *Instabilities and Chaos in Quantum Optics*. Synergetics Series 34, 1–253. Berlin: Springer Verlag.

Arecchi FT and Harrison RG (eds) (1994) *Selected Papers on Optical Chaos*. SPIE Milestone Series, vol. MS 75, SPIE Optical Engineering Press.

Gauthier DJ (1998) Chaos has come again. *Nature* 279: 1156–1157.

Gills Z, Iwata C and Roy R (1992) Tracking unstable steady states: extending the stability regime of a multimode laser system. *Physical Review Letters* 69: 3169–3172.

Haken H (1985) Light. In: *Laser Light Dynamics*, vol. 2. Amsterdam: North-Holland.

Harrison RG (1988) Dynamical instabilities and chaos in lasers. *Contemporary Physics* 29: 341–371.

Harrison RG and Biswas DJ (1985) Pulsating instabilities and chaos in lasers. *Progress in Quantum Electronics* 10: 147–228.

Hunt ER (1991) Stabilizing high-period orbits in a chaotic system – the diode resonator. *Physical Review Letters* 67: 1953–1955.

Ott E, Grebogi C and Yorke JA (1990) Controlling chaos. *Physical Review Letters* 64: 1196–1199.

Pecore LM and Carrol TL (1990) Synchronization of chaotic systems. *Physical Review Letters* 64: 821–824.
Roy R, *et al.* (1992) Dynamical control of a chaotic laser: experimental stabilization of a globally coupled system. *Physical Review Letters* 68: 1259–1262.
VanWiggeren GD and Roy R (1998) Communication with chaotic lasers. *Science* 279: 1198–1200.
VanWiggeren GD and Roy R (1998) Optical communication with chaotic waveforms. *Physical Review Letters* 279: 1198–1200.

CHEMICAL APPLICATIONS OF LASERS

Contents

Detection of Single Molecules in Liquids
Diffuse-Reflectance Laser Flash Photolysis
Laser Manipulation in Polymer Science
Nonlinear Spectroscopies
Photodynamic Therapy of Cancer
Pump and Probe Studies of Femtosecond Kinetics
Time-Correlated Single-Photon Counting
Transient Holographic Grating Techniques in Chemical Dynamics

Detection of Single Molecules in Liquids

A J de Mello, J B Edel and E K Hill, Imperial College of Science, Technology and Medicine, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

A significant challenge facing experimentalists in the physical and biological sciences is the detection and identification of single molecules. The ability to perform such sensitive and selective measurements is extremely valuable in applications such as DNA analysis, immunoassays, environmental monitoring, and forensics, where small sample volumes and low analyte concentrations are the norm. More generally, most experimental observations of physical systems provide a measurement of ensemble averages, and yield information only on mean properties. In contrast, single molecule measurements permit observation of the interactions and behavior of a heterogeneous population in real time.

Over the past few years a number of techniques with sufficient sensitivity have been developed to detect single molecules. Scanning probe microscopies (most notably scanning tunneling microscopy and atomic force microscopy) have been used to great effect in the analysis of surface bound species, but for the detection of single molecules in liquids, optical methods incorporating the measurement of absorption or emission processes, have proved most successful.

The Absorption–Emission Cycle

The key concept underlying most emissive approaches to single molecule detection is that a single molecule can be cycled repeatedly between its ground state and an excited electronic state to yield multiple photons. The process can be understood by reference to Figure 1. Fluorescence emission in the condensed phase can be described using a four-step cycle. Excitation from a ground electronic state to an excited state is followed by rapid (internal) vibrational relaxation. Subsequently, radiative decay to the ground state is observed as fluorescence emission and is governed by the excited state lifetime. The final stage is internal relaxation back to the original ground state. Under saturating illumination, the rate-limiting step for this cycle is governed by the fluorescence lifetime (τ_f), which is typically of the order of a few nanoseconds. If a single molecule diffuses through an illuminated zone (e.g., the focus of a laser beam) it may reside in that

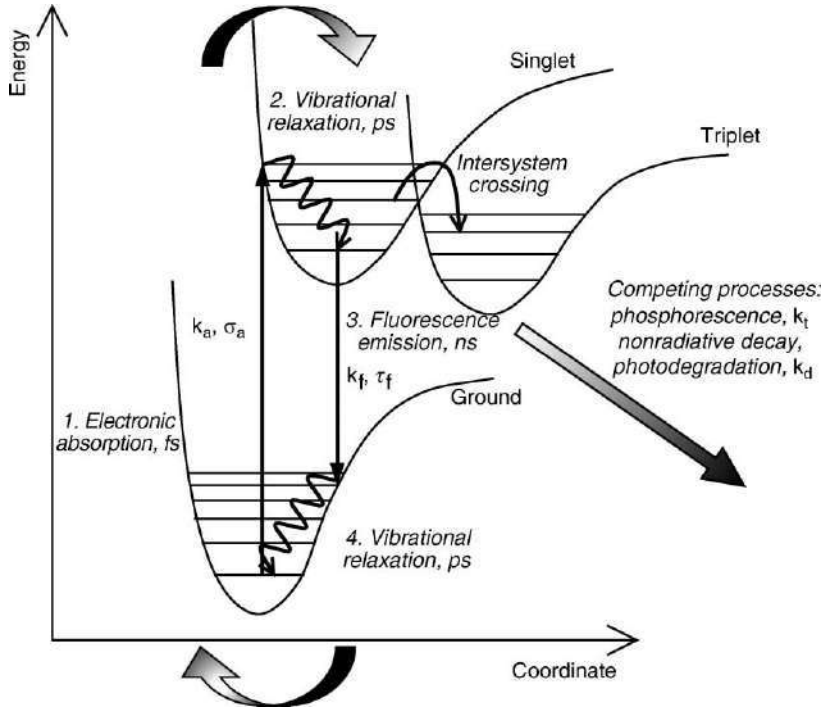

Figure 1 Schematic illustration of the molecular absorption–emission cycle and timescales for the component processes. Competing processes that may reduce the ultimate photon yield are also shown.

region for several milliseconds. The rapid photon absorption-emission cycle may therefore be repeated many times during the residence period, resulting in a burst of fluorescence photons as the molecule transits the beam. The burst size is limited theoretically by the ratio of the beam transit time (τ_r) and the fluorescence lifetime:

$$N_{\text{photons}} = \frac{\tau_r}{\tau_f} \qquad [1]$$

For typical values of τ_r (5 ms) and τ_f (5 ns) up to one million photons may be emitted by the single molecule. In practice, photobleaching and photodegradation processes limit this yield to about 10^5 photons. Furthermore, advances in optical collection and detection technologies enable registration of about 1–5% of all photons emitted. This results in a fluorescence burst signature of up to a few thousand photons or photoelectrons.

Successful single molecule detection (SMD) depends critically on the optimization of the fluorescence burst size and the reduction of background interference from the bulk solvent and impurities. Specifically a molecule is well suited for SMD if it is efficiently excited by an optical source (i.e., possesses a large molecular absorption cross-section at the wavelength of interest), has a high fluorescence quantum efficiency (favoring radiative deactivation of the excited state), has a short fluorescence lifetime, and is exceptionally photostable.

Ionic dyes are often well suited to SMD as fluorescence quantum efficiencies can be close to unity and fluorescence lifetimes below 10 nanoseconds. For example, xanthene dyes such as Rhodamine 6G and tetramethyl-rhodamine isothiocyanate are commonly used in SMD studies. However, other highly fluorescent dyes such as fluorescein are unsuitable for such applications due to unacceptably high photodegradation rate coefficients. Furthermore, some solvent systems may enhance nonradiative processes, such as intersystem crossing, and yield significant reduction in the photon output. Structures of three common dyes suitable for SMD are shown in Figure 2.

Signal vs Background

The primary challenge in SMD is to ensure sufficient reduction in background levels to enable discrimination between signal and noise. As an example, in a 1 nM aqueous dye solution each solute molecule occupies a volume of approximately 1 fL. However, this same volume also contains in excess of 10^{10} solvent molecules. Despite the relatively small scattering cross-section for an individual water molecule ($\sim 10^{-28}$ cm^2 at 488 nm), the cumulative scattering

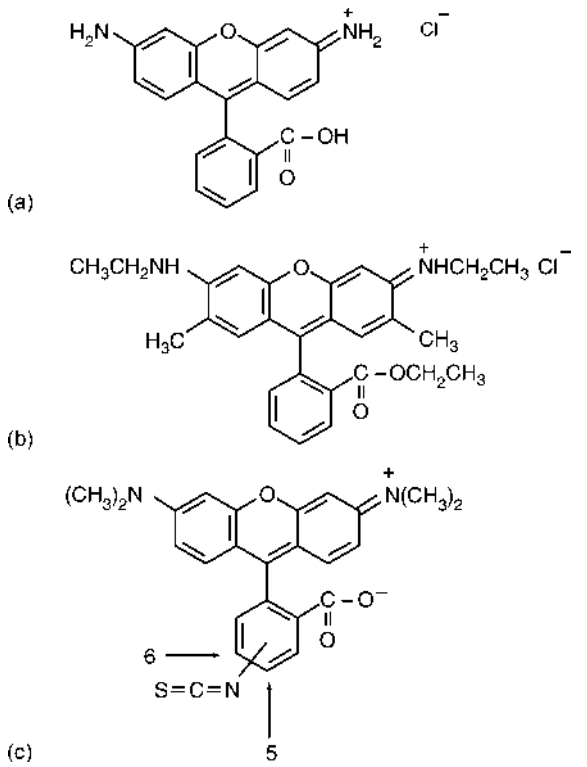

Figure 2 Structures of common dye molecules suitable for SMD in solution: (a) 3,6-diamino-9-(2-carboxyphenyl)-chloride (rhodamine 110); (b) 9-(2-(ethoxycarbonyl)phenyl)-3,6-bis (ethylamino)-2,7-dimethyl chloride (rhodamine 6G); (c) 9-(2-carboxy-isothiocyanatophenyl)-3,6-bis(dimethylamino)-inner salt (tetra methylrhodamine-5-(and-6)-isothiocyanate).

signal from the solvent may swamp the desired fluorescence signal. The principal method of reducing the solvent background is to minimize the optical detection volume: the signal from a single molecule is independent of probe volume dimensions, but the background scales proportionally with the size of the detection region. Although there are several experimental approaches to SMD in solution, several factors hold common:

1. Tiny detection volumes (10^{-12}–10^{-15} L) are used to reduce background signals.
2. A low analyte concentration combined with the small observation volume, ensures that less than one analyte molecule is present in the probe volume on average.
3. High-efficiency photon collection (optics) and detection maximize the proportion of the isotropic fluorescence burst that is registered.
4. Background reduction methods are employed to improve signal-to-noise ratios. These include: optical rejection of Raman and Rayleigh scatter, time-gated discrimination between prompt scatter and delayed emission, and photobleaching of the solvent immediately before detection.

The minute volumes within which single molecules are detected can be generated in a variety of ways. Picoliter volumes can be defined by mutually orthogonal excitation and detection optics focused in a flowing stream. Much smaller, femtoliter probe volumes are generated using confocal microscopes. At this level, background emission is significantly reduced and high signal-to-noise ratios can be achieved. Confocal detection techniques are versatile and have been widely adopted for SMD in freely diffusing systems. Consequently, confocal methods will be discussed in detail in this article. The other general approach to performing SMD in solution involves the physical restriction of single molecules within defined volumes. Of particular note are techniques where single molecules are confined within a stream of levitated microdroplets. Droplet volumes are typically less than 1 fL and imaging of the entire microdroplet enables single molecule fluorescence to be contrasted against droplet 'blanks' with good signal-to-noise. Furthermore, since molecules are confined within discrete volumes, the technique can be utilized for high-efficiency molecular counting applications. More recently, spatial confinement of molecules in capillaries and microfabricated channels (with submicron dimensions) has been used to create probe volumes between 1 fL and 1 pL, and immobilized molecules on surfaces have been individually probed using wide-field microscopy with epi-illumination or evanescent wave excitation.

Single Molecule Detection using Confocal Microscopy

As previously stated, the confocal fluorescence microscope is an adaptable and versatile tool for SMD. In its simplest form, a confocal microscope is one in which a point light source, a point focus in the object plane, and a pinhole detector are all confocal with each other. This optical superposition generates superior imaging properties and permits definition of ultra-small probe volumes. The concepts behind a basic confocal microscope and its use in SMD are schematically illustrated in Figure 3. Coherent light (typically from a laser and tuned to an optical transition of the molecule under investigation) behaves as a point light source and is focused into a sample chamber using a high-numerical aperture objective lens. As a single molecule traverses the laser beam it is continuously cycled between the ground and an excited electronic state, emitting a burst of fluorescence photons. Fluorescence emission is isotropic (spontaneous emission), so photons are emitted in all directions (4π steradians).

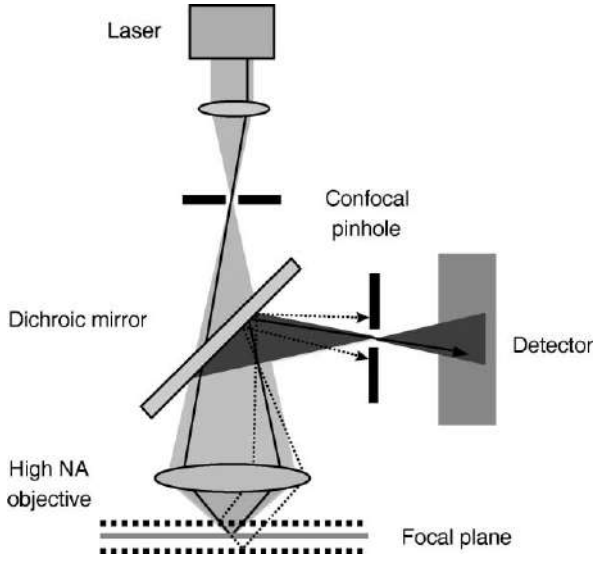

Figure 3 Principle of confocal detection. A confocal pinhole only selects light that emanates from the focal region. Dashed lines indicate paths of light sampled above and below the focal plane that are rejected by the pinhole. The solid ray derives from the focal point, and is transmitted through the pinhole to a detector.

Consequently, the high numerical aperture is used to collect as large a fraction of photons emitted from the focal plane as possible. Designing the objective to be used with an immersion medium, such as oil, glycerin, or water, can dramatically increase the objective numerical aperture, and thus the number of collected photons. Light collected by the objective is then transmitted towards a dichroic beam splitter. In the example shown, fluorescence photons (of lower energy) are reflected towards the confocal detector pinhole, whereas scattered radiation (of higher energy) is transmitted through the dichroic towards the light source. Creation of a precise optical probe volume is effected through the definition of the confocal pinhole. The detector is positioned such that only photons that pass through the pinhole are detected. Consequently, light emanating from the focal plane in the sample is transmitted through the pinhole and detected, whereas light not deriving from the focal plane is rejected by the aperture, and therefore not detected (Figure 3).

To ensure that the maximum number of photons are detected by the system, high efficiency detectors must be used. Photomultiplier tubes (the most common detectors for light-sensing applications) are robust and versatile but have poor detection efficiencies (approximately 5% of all photons that fall on the photocathode yield an electrical signal). Consequently, the most useful detectors for SMD (or low light level) applications are single photon-counting avalanche photodiodes (SPADs). A SPAD is essentially a p–n junction reverse biased above the breakdown voltage, that sustains an avalanche current when triggered by a photon-generated carrier. Detection efficiencies for typical SPADs are normally between 60–70% and are thus ideal for SMD in solution. An approximation of the overall detection efficiency of a confocal system for SMD can be made using eqn [2], which incorporates an estimation of photon losses at all stages of the collection/detection process. Typical transmission efficiencies for each step are also shown.

overall detection efficiency	≈	objective collection efficiency	×	dichroic transmission coefficient	×	additional optical losses	×	detector efficiency
0.06		0.24		0.9		0.5		0.6

[2]

Optical Probe Volumes

The precise nature of the probe volume is determined by the image of the pinhole in the sample and the spherical aberration of the microscope objective. Routinely, confocal probe volumes are approximated as resembling a cylinder with a radius defined by the diffraction-limited waist of a Gaussian beam. This approximation is useful when the incident beam is narrow or not tightly focused. However, when the radius of the incident beam is large, the corresponding diffraction limited focus is narrowed, and the probe volume more closely resembles a pair of truncated cones. Figure 4a illustrates the dependence of the curvature of the $1/e^2$ intensity contour on the collimated beam radius.

Consequently, it is clear that a simple cylindrical approximation for the probe volume breaks down for wide, tightly focused beams. If a broad incident beam (diameter >1.5 mm) is used, a large noncylindrical contribution to the probe volume is anticipated and a more appropriate model is required. An alternative and more accurate model for the confocal probe volume considers the Gaussian profile of the focused beam. The $1/e^2$ intensity contour radius of a Gaussian waveform with wavelength λ, at some distance z from the beam waist radius w_0, is given by eqn [3]:

$$w(z) = w_0\sqrt{1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2} \qquad [3]$$

In this case, the probe volume V is given by the volume of rotation of $w(z)$ around the z-axis between Z' and $-Z'$. The volume of rotation can therefore be

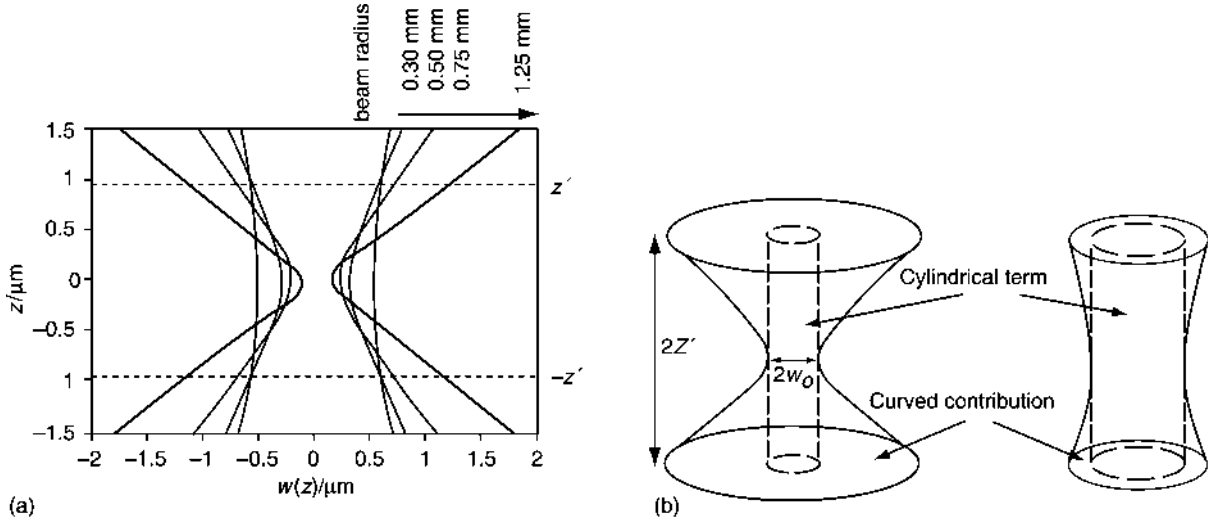

Figure 4 (a) $1/e^2$ Gaussian intensity contours plotted for a series of laser beam radii ($\lambda = 488$ nm, $f = 1.6$ mm, $n = 1.52$). (b) Cylindrical and curved components of the Gaussian probe volume. The curved contribution is more significant for larger beam radii and correspondingly tight beam waists.

simply defined according to,

$$V = \int_{-Z'}^{Z'} \pi w(z)^2 dz \qquad [4]$$

Solution of eqn [4] yields

$$V = 2\pi w_0^2 Z' + \frac{2\lambda^2}{3\pi w_0^2} Z'^3 \qquad [5]$$

The Gaussian volume expression contains two terms. The first term, $2\pi w_0^2 Z'$, corresponds to a central cylindrical volume; the second term has a more complex form that describes the extra curved volume (Figure 4b). The diffraction-limited beam waist radius w_0 can be defined in terms of the focusing objective focal length f, the refractive index n, and the collimated beam radius R according to

$$w_0 = \frac{\lambda f}{n\pi R} \qquad [6]$$

Substitution in eqn [5] yields,

$$V = 2\pi\left(\frac{\lambda f}{n\pi R}\right)^2 Z' + \frac{2\lambda^2}{3\pi}\left(\frac{n\pi R}{\lambda f}\right)^2 Z'^3$$

$$= \frac{2\lambda^2 f^2}{\pi n^2 R^2} Z' + \frac{2\pi n^2 R^2}{3 f^2} Z'^3 \qquad [7]$$

The volume is now expressed in terms of identifiable experimental variables and constants. Once again, the first term may be correlated with the cylindrical contribution to the volume, and the second term is the additional volume due to the curved contour. It is clear from Figure 4a that, for a given focal length, the noncylindrical contribution to the probe volume increases with incident beam diameter, when the diffraction limited focus is correspondingly sharp and narrow. Furthermore, it can also be seen that the second term in eqn [7] is inversely proportional to f^2, and thus the extent to which the probe volume is underestimated by the cylindrical approximation increases with decreasing focal length. This fact is significant when performing confocal measurements, since high numerical aperture objectives with short focal lengths are typical. Some realistic experimental parameters give an indication of typical dimensions for the probe volume in confocal SMD systems. For example, if $\lambda = 488$ nm, $f = 1.6$ mm, $Z' = 1.0$ μm and $n = 1.52$, a minimum optical probe volume of 1.1 fL is achievable with a collimated beam diameter of 1.1 mm.

Intensity Fluctuations: Photon Burst Statistics

When sampling a small volume within a system that may freely exchange particles with a large surrounding analyte bath, a Poisson distribution of particles is predicted. A Poisson distribution is a discrete series that is defined by a single parameter μ equating to the mean and variance of the distribution:

$$P(n = x) = \frac{\mu^x e^{-\mu}}{x!} \qquad [8]$$

Common Poisson processes include radioactive disintegration, random walks and Brownian motion. Although particle number fluctuations in the excitation volume are Poissonian in nature, the corresponding fluorescence intensity modulation induces a stronger correlation between photon counts. For a single molecular species the model is described by two parameters: an intrinsic molecular brightness and the average occupancy of the observation volume. A super-Poissonian distribution has a width or variance that is greater than its mean; in a Poisson distribution the mean value and the variance are equal. The fractional deviation Q is defined as the scaled difference between the variance and the expectation value of the photon counts, and gives a measure of the broadening of the photon counting histogram (PCH). Q is directly proportional to the molecular brightness factor ε and the shape factor γ of the optical point spread function. (γ is constant for a given experimental geometry.)

$$Q = \frac{\langle \Delta n^2 \rangle - \langle n \rangle}{\langle n \rangle} = \gamma\varepsilon \qquad [9]$$

A pure Poisson distribution has $Q = 0$; for super-Poissonian statistics $Q > 0$. Deviation from the Poisson function is maximized at low number density and high molecular brightness.

In a typical SMD experiment raw data are generally collected with a multichannel scalar and photons are registered in binned intervals. Figure 5 illustrates typical photon burst scans demonstrating the detection of single molecules (R-phycoerythrin) in solution. Fluorescence photon bursts, due to single molecule events, are clearly distinguished above a low background baseline (top panel) of less than 5 counts per channel in the raw data. It is noticeable that bursts vary in both height and size. This is in part due to the range of possible molecular trajectories through the probe volume, photobleaching kinetics, and the nonuniform illumination intensity in the probe region. In addition, it can be seen that the burst frequency decreases as bulk solution concentration is reduced. This effect is expected since the properties of any given single-molecule event are determined by molecular parameters alone (e.g., photophysical and diffusion constants) and concentration merely controls the frequency/number of events.

Although many fluorescence bursts are clearly distinguishable from the background, it is necessary to set a count threshold for peak discrimination in order to correctly identify fluorescence bursts above the background. A photocount distribution can be used as the starting point for determining an appropriate threshold for a given data set. The overlap between signal and background photocount distributions affects the efficiency of molecular detection. Figure 6 shows typical signal and background photocount probability distributions, with a threshold set at approximately 2 photocounts per channel. The probability of spurious (or false)

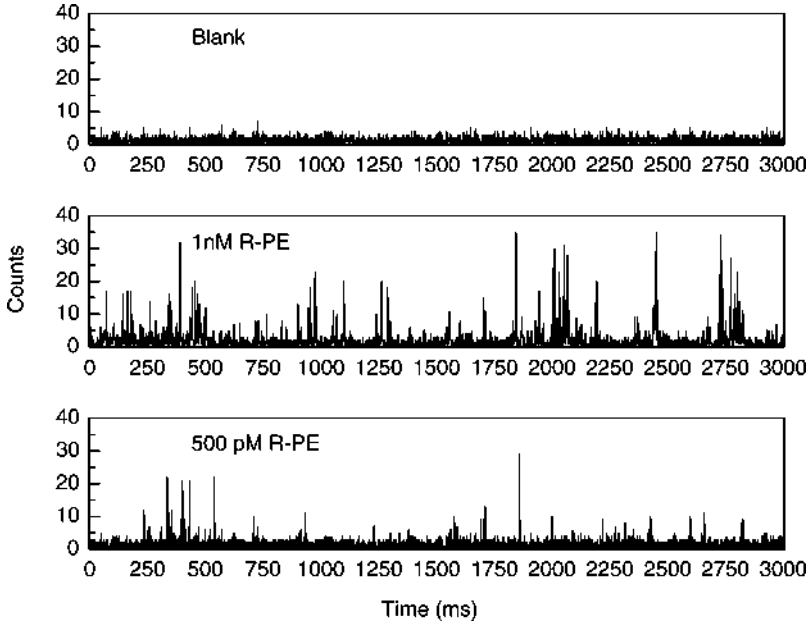

Figure 5 Photon burst scans originating from 1 nM and 500 pM R-phycoerythrin buffered solutions. Sample is contained within a 50 μm square fused silica capillary. Laser illumination = 5 μW, channel width = 1 ms. The top panel shows a similar burst scan originating from a deionized water sample measured under identical conditions.

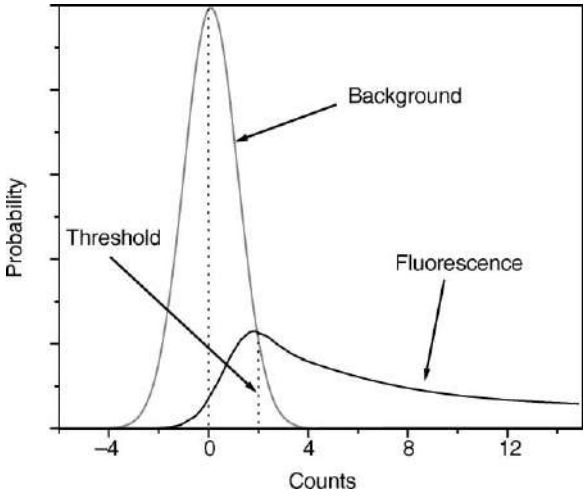

Figure 6 Simulated fluorescence and background photocount probability distributions. The vertical dashed line at 2 counts represents an arbitrarily defined threshold value for peak determination.

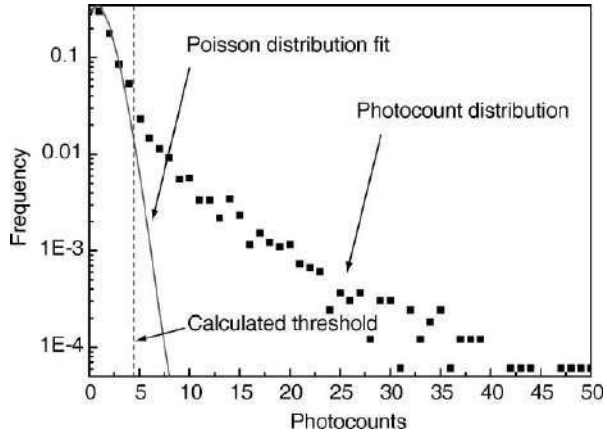

Figure 7 A photon counting histogram generated from a 16 second photon burst scan originating from a 10 μg/mL solution of 1000 nm fluorescent microbeads. The dotted curve shows a least-squares fit of early channels to a Poisson distribution, and the dashed vertical line marks the peak threshold (defined as $\mu + 3\sqrt{\mu} = 4.47$ counts).

detection resulting from statistical fluctuations in the background can be quantified by the area under the 'background' curve at photocount values above the threshold. Similarly, the probability that 'true' single molecule events are neglected can be estimated by the area under the 'fluorescence' curve at photocount values below the threshold. Choice of a high threshold value will ensure a negligible probability of calling a false positive, but will also exclude a number of true single molecule events that lie below the threshold value. Conversely, a low threshold value will generate an unacceptably high number of false positives. Consequently, choice of an appropriate threshold is key in efficient SMD.

Since the background shot noise is expected to exhibit Poisson statistics, the early part of the photocount distribution (i.e., the portion that is dominated by low, background counts) can be modeled with a Poisson distribution, to set a statistical limit for the threshold. Photon counting events above this threshold can be defined as photon bursts associated with the presence of single molecules. In an analogy with Gaussian systems the selected peak discrimination threshold can be defined as three standard deviations from the mean count rate:

$$n_{\text{threshold}} = \mu + 3\sqrt{\mu} \qquad [10]$$

Adoption of a threshold that lies 3 standard deviations above the mean yields a confidence limit that is typically greater than 99%. Figure 7 illustrates a sample photocount distribution, a least-squares fit to an appropriate Poisson distribution, and the calculated threshold that results. Once the threshold has been calculated, its value is subtracted from all channel counts and a peak search utility used to identify burst peaks in the resulting data set.

Data Filtering

As stated, the primary challenge in detecting single molecules in solution is not the maximization of the detected signal, but the maximization of the signal-to-noise ratio (or the reduction of background interferences). Improving the signal-to-noise ratio in such experiments is important, as background levels can often be extremely high.

Several approaches have been used to smooth SMD data with a view to improving signal-to-noise ratios. However, the efficacy of these methods is highly dependent on the quality of the raw data obtained in experiment. As examples, three common methods are briefly discussed. The first method involves the use of a weighted quadratic sum (WQS) smoothing filter. The WQS function creates a weighted sum of adjacent terms according to

$$\tilde{n}_{k,WQS} = \sum_{j=0}^{m-1} w_j (n_{k+j})^2 \qquad [11]$$

The range of summation m is the same order as the burst width, and the weighting factors w_j are chosen

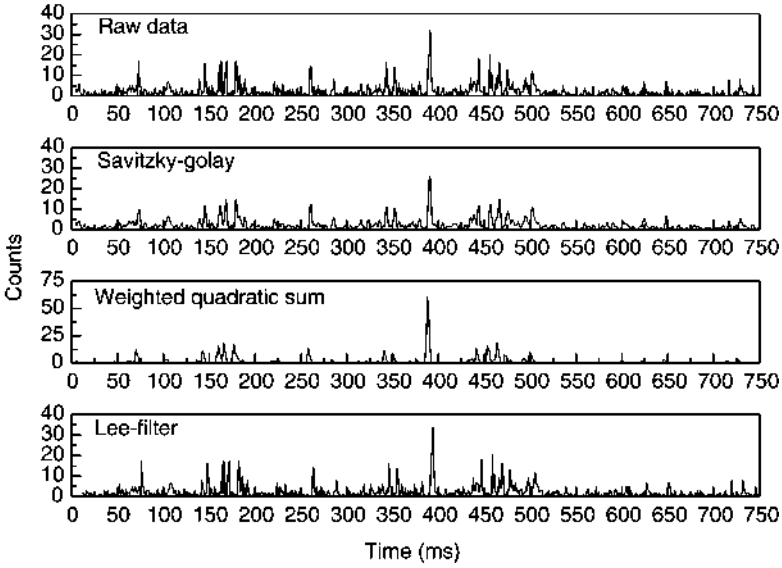

Figure 8 Effect of various smoothing filters on a 750 ms photon burst scan originating from a 1 nM R-phycoerythrin buffered solution. Raw data are shown in the top panel.

to best discriminate between single-molecule signals and random background fluctuations. This method proves most useful for noisy systems, in which the raw signal is weak. There is a practical drawback, in that peak positions are shifted by the smoothing function, and subsequent burst analysis is therefore hampered.

Another popular smoothing filter is the Lee-filtering algorithm. The Lee filter preferentially smoothes background photon shot noise and is defined according to

$$\tilde{n}_k = \bar{n}_k + (n_k - \bar{n}_k)\frac{\sigma_k^2}{\sigma_k^2 + \sigma_0^2} \qquad [12]$$

where the running mean (n_k) and running variance (σ_k^2) are defined by

$$\bar{n}_k = \frac{1}{(2m+1)}\sum_{j=-m}^{m} n_{k+j} \quad m < k \leq N - m \qquad [13]$$

$$\sigma_k^2 = \frac{1}{(2m+1)}\sum_{j=-m}^{m}(n_{k+j} - \bar{n}_{k+j})^2 \quad 2m < k \leq N - 2m \qquad [14]$$

for a filter $(2m+1)$ channels wide. Here, n_k is the number of detected photons stored in a channel k, σ_0 is a constant filter parameter, and N is the total number of data points.

A final smoothing technique worth mentioning is the Savitsky-Golay smoothing filter. This filter uses a least-squares method to fit an underlying polynomial function (typically a quadratic or quartic function) within a moving window. This approach works well for smooth line profiles of a similar width to the filter window and tends to preserve features such as peak height, width and position, which may be lost by simple adjacent averaging techniques. Figure 8 shows the effects of using each approach to improve signal-to-noise for raw burst data.

Photon Burst Statistics

A valuable quantitative analysis method for analysis of fluorescence bursts utilizes the analysis of Poisson statistics. Burst interval distributions are predicted to follow a Poissonian model, in which peak separation frequencies adopt an exponential form. The probability of a single molecule (or particle) event occurring after an interval Δt is given by eqn [15]:

$$N(\Delta t) = \lambda \exp(-\beta t) \qquad [15]$$

where λ is a proportionality constant and β is a characteristic frequency at which single molecule events occur. The recurrence time τ_R can then be simply defined as

$$\tau_R = \frac{1}{\beta} \qquad [16]$$

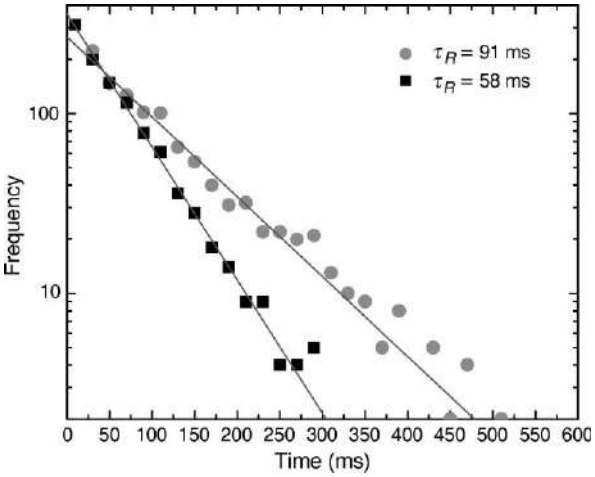

Figure 9 Burst interval distribution analysis of photon burst scans. Data originate from 1 μm fluorescent beads moving through 150 μm wide microchannels at flow rates of 200 nL min^{-1} (circles) and 1000 nL min^{-1} (squares). Least squares fits to a single exponential function are shown by the solid lines.

Equation [6] simply states that longer intervals between photon bursts are less probable than shorter intervals. Furthermore, the recurrence time reflects a combination of factors that control mobility, probe volume occupancy, or other parameters in the single molecule regime. Consequently, it is expected that τ_R should be inversely proportional to concentration, flow rate or solvent viscosity in a range of systems. Figure 9 shows an example of frequency $N(\Delta t)$ versus time plots for two identical particle systems moving at different velocities through the probe volume. A least-squares fit to a single exponential function yields values of $\tau_R = 91$ ms for a volumetric flow rate of 200 nL/min and $\tau_R = 58$ ms for a volumetric flow rate of 1000 nL/min.

Temporal Fluctuations: Autocorrelation Analysis

Autocorrelation analysis is an extremely sensitive method for detecting the presence of fluorescence bursts in single molecule experiments. This approach essentially measures the average of a fluctuating signal as opposed to the mean spectral intensity. As previously discussed, the number of molecules contained within a probe volume at any given time is governed by Poisson statistics. Consequently, the root mean square fluctuation can be defined according to eqn [17],

$$\frac{\sqrt{\langle(\delta N)^2\rangle}}{\langle N\rangle} = \frac{\sqrt{\langle(N-\langle N\rangle)^2\rangle}}{\langle N\rangle} = \frac{1}{\sqrt{\langle N\rangle}} \qquad [17]$$

where N is the number of molecules. It is observed that the relative fluctuation diminishes as the number of particles measured is increased. Hence, it is important to minimize the number of molecules present in the probe volume. It should be noted, however, that if there are too few molecules in the solution there will be long dark periods were no single molecule bursts are observed. If the probe volume is bathed in radiation of constant intensity, fluctuation of the resulting fluorescence signal can simply be defined as deviations from the temporal signal average:

$$\langle F(t)\rangle = \frac{1}{T}\int_0^T F(t)\mathrm{d}t \qquad [18]$$

Here, t is defined as the total measurement time, $F(t)$ is the fluorescence signal at time t, and $\langle F(t)\rangle$ is the temporal signal average. Fluctuations in the fluorescence intensity, $\delta F(t)(\delta F(t) = F(t) - \langle F(t)\rangle)$, with time t, about an equilibrium value $\langle F\rangle$, can be statistically investigated by calculating the normalized autocorrelation function, $G(\tau)$, where

$$G(\tau) = \frac{\langle F(t+\tau)F(t)\rangle}{\langle F\rangle^2} = \frac{\langle \delta F(t+\tau)\delta F(t)\rangle}{\langle F\rangle^2} + 1 \qquad [19]$$

In dedicated fluorescence correlation spectroscopy experiments, the autocorrelation curve is usually generated in real time in a high-speed digital correlator. Post data acquisition calculation is also possible using the following expression

$$G(\tau) = \sum_{t=0}^{N-1} g(t)g(t+\tau) \qquad [20]$$

Here $g(t)$ is the total number of counts during the time interval $(t, t+\Delta t)$, $g(t+\tau)$ is the number of counts detected in an interval of Δt at a later time $t+\tau$, and N is the total number of time intervals in the dataset. In a diffusion controlled system with a single fluorescent molecule that is irradiated with a three dimensional Gaussian intensity profile, the autocorrelation curve is governed by the mean probe volume occupancy N and the characteristic diffusion time (τ_D). The laser beam waist radius ω and the probe depth $2z$ describe the Gaussian profile:

$$G(\tau) = 1 + \frac{1}{N}\left(1+\frac{\tau}{\tau_d}\right)^{-1}\left(1+\left(\frac{\omega}{z}\right)^2\frac{\tau}{\tau_d}\right)^{-1/2} \qquad [21]$$

The diffusion time is a characteristic molecular residence time in the probe volume and inversely

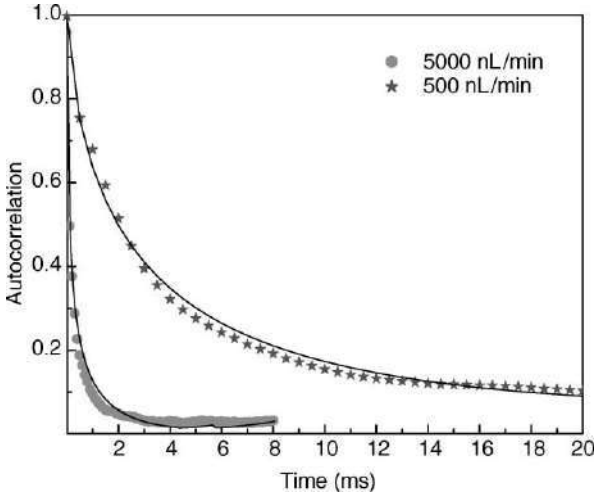

Figure 10 Autocorrelation analysis of photon burst scans of 1 μm fluorescent beads moving through 150 μm wide microchannels at flow rates of 500 nL min^{-1} (stars) and 1000 nL min^{-1} (circles). Solid lines represent fits to the data according to eqn [23].

related to the translational diffusion coefficient for the molecule:

$$D = \frac{w^2}{4\tau_D} \qquad [22]$$

In a flowing system, the autocorrelation function depends on the average flow time through the probe volume τ_{flow}. A theoretical fit to the function can be described according to

$$G(\tau) = 1 + \frac{1}{N} A \exp\left\{\left(\frac{\tau}{\tau_{\text{flow}}}\right)^2 A\right\}$$
$$A = \left(1 + \frac{\tau}{\tau_d}\right)^{-1}\left(1 + \left(\frac{\omega}{z}\right)^2 \frac{\tau}{\tau_d}\right) \qquad [23]$$

where N is the mean probe volume occupancy; the flow velocity v can then be extracted from the characteristic flow time according to

$$v = \frac{w}{\tau_{\text{flow}}} \qquad [24]$$

It should be noted that in the case that directed flow is negligible or defined to be zero, the autocorrelation function simplifies to eqn [21].

Figure 10 illustrates experimentally determined autocorrelation curves for two identical particle systems moving at different velocities through the probe volume. As particle flow velocity is increased, the width of the autocorrelation curves is seen to narrow as a result of the reduced residence time in the probe volume. A plot of the reciprocal of the full width half maximum of the autocorrelation curve as a function of volumetric flow rate is linear, and provides a simple way of calculating particle/molecule velocities within flowing systems.

Applications

The basic tools and methods outlined in this chapter have been used to perform SMD in a variety of chemically and biologically relevant systems, and indeed there is a large body of work describing the motion, conformational dynamics and interactions of individual molecules (see Further Reading). A primary application area has been in the field of DNA analysis, where SMD methods have been used in DNA fragment sizing, single-molecule DNA sequencing, high-throughput DNA screening, single-molecule immunoassays, and DNA sequence analysis. SMD methods have also proved highly useful in studying protein structure, protein folding, protein-molecule interactions, and enzyme activity.

More generally, SMD methods may prove to be highly important as a diagnostic tool in systems where an abundance of similar molecules masks the presence of distinct molecular anomalies that are markers in the early stages of disease or cancer.

See also

Microscopy: Confocal Microscopy.

Further Reading

Ambrose WP, Goodwin PM, Jett JH, Van Orden A, Werner JH and Keller RA (1999) Single molecule fluorescence spectroscopy at ambient temperature. *Chem. Rev.* 99: 2929–2956.

Barnes MD, Whitten WB and Ramsey JM (1995) Detecting single molecules in liquids. *Anal. Chem.* 67: 418A–423A.

Basche T, Orrit M and Rigler R (2002) *Single Molecule Spectroscopy in Physics, Chemistry, and Biology.* New York: Springer-Verlag.

Zander CJ, Enderlein J and Keller RA (2002) *Single Molecule Detection in Solution: Methods and Applications*. New York: John Wiley & Sons.

Diffuse-Reflectance Laser Flash Photolysis

D R Worrall and S L Williams, Loughborough University, Loughborough, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Flash photolysis was developed as a technique to study short-lived intermediates in photoinduced reactions by George Porter (later Lord Porter of Luddenham) and Ronald Norrish in 1949, drawing on Porter's wartime experience with radar techniques. Such was the impact of this development that it earned the Nobel prize jointly for Porter, Norrish, and Eigen in 1967. In this article we will describe the application of flash photolysis to opaque, scattering samples, detailing how light propagation in such samples can be treated theoretically, and will discuss methods by which the data obtained from diffuse reflectance flash photolysis experiments may be analyzed.

The technique of flash photolysis was originally based on using an intense flash of light (the photolysis flash) from a xenon tube to excite the sample, followed a certain time delay later by a spectroscopic flash of lower intensity from a second flash tube, the light from the latter being detected using a photographic plate. The photolysis flash is of sufficient intensity to produce a large population of intermediates (radicals, ions, excited states, isomers) in the sample, which then absorb light from the spectroscopic flash depending on the concentration of intermediates according to the Beer–Lambert law:

$$A = \varepsilon c l \qquad [1]$$

where A is the sample absorbance, ε the molar absorption coefficient, c the concentration and l the pathlength. The absorbance is related to the incident and transmitted intensities as

$$A = \log_{10} I_0 / I \qquad [2]$$

With I_0 the incident and I the transmitted intensities, there is therefore an exponential fall-off of intensity with pathlength for a homogeneous absorber. Hence, by monitoring the evolution of the absorption spectra, the changes in concentration of the photoproduced intermediates and hence the kinetics of the processes in which they are involved are elucidated.

Flash photolysis has evolved subsequently to make use of laser sources and sophisticated electronic detection apparatus to push the limits of time resolution to the femtosecond regime. Indeed, recently the Nobel prize for chemistry was awarded to Ahmed Zewail for his work with ultrafast pump-probe techniques. However, in its conventional geometry, flash photolysis is limited to transparent samples, since it is necessary to be able to probe the excited species by monitoring the absorption spectra.

Many biological systems and industrially important samples are opaque or highly scattering, and hence the attenuation of light through the sample is no longer described by the Beer–Lambert law. In 1981, Frank Wilkinson and Rudolph Kessler had the idea of using diffusely reflected light to interrogate the changes in concentration within a scattering sample subjected to a high-intensity excitation pulse. When photons enter a sample, they may be absorbed or scattered. Those which are scattered may re-emerge from the irradiated surface as diffusely reflected light. The intensity of diffusely reflected light emerging from the surface at a particular wavelength is a unique function of the ratio of scattering to absorption. The more scattering events occurring before absorption, the more likely the photon is to escape from the sample as diffusely reflected light. Hence the probability of escape decreases as the absorption probability increases, and the diffusely reflected light is deficient in those wavelengths where the absorption is strongest, i.e., the ratio of the incident to absorbed light intensity at a given wavelength is related to the absorption of the sample at that wavelength.

Kubelka–Munk Theory of Reflectance

The theory which describes the relationship between incident and scattered light intensity, absorption and scatter, and concentration which is widely applied in this context, is the Kubelka–Munk theory of reflectance. The theory was originally developed to describe the reflectance characteristics of paint films but it works quite well for many samples containing a homogeneous distribution of scatterers and absorbers. The limiting assumption in this theory is that the scatterers from which the scattering layer is composed are very much smaller than the total layer thickness. Additionally, the layer should be optically thick such that all of the light entering the layer should be either absorbed or reflected, with a negligible fraction transmitted.

For a layer of thickness X diffusely irradiated with monochromatic light, the diagram shown in Figure 1

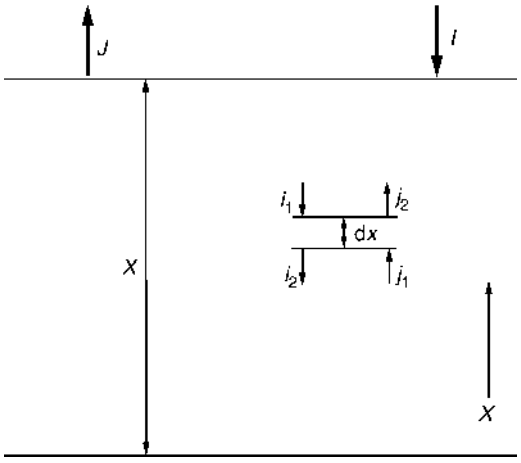

Figure 1 Schematic of counterpropagating fluxes in a diffusely irradiated scattering medium.

can be constructed, with I the incident flux and J the diffusely reflected flux, and i and j the fluxes traveling upwards and downwards through an infinitesimally small thickness element dx. Two further parameters may be defined which are characteristic of the medium.

K The absorption coefficient. Expresses the attenuation of light due to absorption per unit thickness.
S The scattering coefficient. Expresses the attenuation of light due to scattering per unit thickness.

Both of these parameters can be thought of as arising due to the particle (or chromophore) acting as a sphere of characteristic size, which casts a shadow either due to the prevention of on-axis transmission or due to absorption. The scattering or absorption coefficient then depends on the effective size of this sphere, and the number density of spheres in the medium. In each case, the probability P of a photon being transmitted through a particular thickness X of a medium is related exponentially to the absorption or scattering coefficient:

$$P = \exp(-KX) \qquad [3]$$

$$P = \exp(-SX) \qquad [4]$$

The scattering coefficient S depends on the refractive index difference between the particle and the dispersion medium. The scattering coefficient is also dependent upon the particle size, and shows an inverse correlation, i.e., the scattering coefficient increases as the particle size decreases. This effect is a function of the size of the particle relative to the wavelength of light impinging on it, with the scattering coefficient increasing as the wavelength decreases. However, this change with wavelength is small provided the particle size is large relative to the wavelength of the light.

The effect of the material in the element dx on the counterpropagating fluxes i and j depends on the absorption and scattering coefficients. Both i and j will be attenuated by both absorption and scattering:

$$i_2 = i_1 - i_1(S + K)\mathrm{d}x \qquad [5]$$

$$j_2 = j_1 - j_1(S + K)\mathrm{d}x \qquad [6]$$

Both i and j are reinforced by backscattering from the other flux:

$$i_2 = i_1 + j_1 S\mathrm{d}x \qquad [7]$$

$$j_2 = j_1 + i_1 S\mathrm{d}x \qquad [8]$$

The net effect of this in the attenuation of the flux propagating into the sample (i) and the flux backscattered from the sample (j) can be expressed as the following differential equations:

$$\mathrm{d}i = (S + K)i_1\mathrm{d}x - j_1 S\mathrm{d}x \qquad [9]$$

$$\mathrm{d}j = -(S + K)j_1\mathrm{d}x + i_1 S\mathrm{d}x \qquad [10]$$

For a sample of infinite thickness, these equations can be solved to give an analytical solution for the observed reflectance of the sample in terms of the absorption and scattering coefficients:

$$\frac{K}{S} = \frac{(1 - R_\infty)^2}{2R_\infty} \qquad [11]$$

where R_∞ is the reflectance from a sample of such thickness that an increase in sample thickness has no effect on the observed reflectance. The absorption coefficient K is dependent upon the concentration of absorbers in the sample through

$$K = 2\varepsilon c \qquad [12]$$

with ε the naperian absorption coefficient, c the concentration of absorbers and the factor 2 which is the geometrical factor for an isotropic scatterer. For multiple absorbers, K is simply the linear sum of absorption coefficients and concentrations. Hence for a diffusely scattering medium, an expression analogous to the Beer–Lambert law can be derived to relate concentration to a physically measurable parameter, in this case the sample reflectance. The ratio K/S is usually referred to as the Kubelka–Munk remission function, and is the parameter usually quoted in this context. It is important to note that the relationship

with concentration is only valid for a homogeneous distribution of absorbers within a sample.

Transient Concentration Profiles

As has been discussed in the previous section, a light flux propagating through a scattering medium is attenuated by both scattering and absorption events, whilst in a nonturbid medium attenuation is by absorption only. Hence even in a sample with a very low concentration of absorbers, the light flux is rapidly attenuated as it penetrates into the sample by scattering events. This leads to a significant flux gradient through the sample, resulting in a reduction in the concentration of photo-induced species as penetration depth into the sample increases. There are three distinct concentration profiles which may be identified within a scattering sample which are dependent on the scattering and absorption coefficients. These concentration profiles are interrogated by a beam of analyzing light, and hence an understanding of the effect of these differing profiles on the diffusely reflected intensity is vital in interpreting transient absorption data. The transient depth profiles are illustrated in Figure 2.

Kubelka–Munk Plug

This occurs when the photolysis flash, i.e., laser fluence is high and the absorption coefficient is low at the laser wavelength. If we assume a simple photophysical model involving simply the ground state S_0, the first excited singlet state S_1 and first excited triplet state T_1, and we assume that either the quantum yield of triplet state production is high or the S_1 lifetime is very much shorter than the laser pulse allowing re-population of the ground state, then it is possible at high enough fluence to completely convert all of the S_0 states to T_1 states within the sample. Provided the fluence is high enough, this complete conversion will penetrate some way into the sample (Figure 2). Under circumstances where the T_1 state absorbs strongly at some wavelength other than the photolysis wavelength, probing at this wavelength will result in the probe beam being attenuated significantly within a short distance of the surface of the sample, and thus will only interrogate regions where there is a homogeneous excited state concentration. Under these circumstances the reflectance of the sample as a function of the concentration of excited states is described by the Kubelka–Munk equation, and the change in remission function can be used to probe the change in excited state concentration.

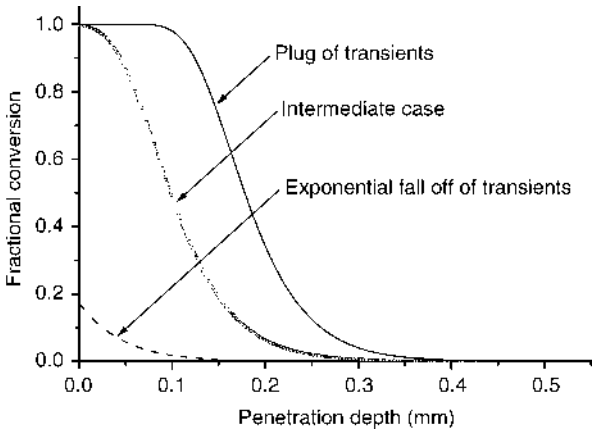

Figure 2 Transient concentration profiles following pulsed laser excitation of a scattering and absorbing sample.

Exponential Fall-Off of Concentration

This occurs when the laser fluence is low and the absorption coefficient at the laser wavelength high. Again considering the simple photophysical model, most of the laser flux will be absorbed by the ground state in the first few layers of sample, and little will penetrate deeply into the sample. In the limiting case this results in an exponential fall-off of transient concentration with penetration depth into the sample (Figure 2). Here the distribution of absorbers is not random and the limiting Kubelka–Munk equation is no longer applicable since the mean absorption coefficient varies with sample penetration depth. Lin and Kan solved eqs [9] and [10] with the absorption coefficient K varying exponentially with penetration depth and showed that the series solution converges for changes in reflectance of less than 10% such that the reflectance change is a linear function of the number of absorbing species.

Intermediate Case

Between the two extremes described above is a case where significant transient conversion takes place at the front surface, but with little penetration into the sample. This can occur, for example, with high laser fluences and large ground state absorption coefficients at the laser wavelength. Under these circumstances, illustrated in Figure 2, the analyzing light interrogates not only the transient concentration profile, but also a significant amount of the analyzing light may penetrate through the transient layer into the unconverted sample behind, if the transient absorption at the analyzing wavelength is low. This creates a more complex problem for analysis since effectively the sample is irradiated from both front and back faces, with consequent effects on the diffusely reflected intensity. It is possible to numerically model the reflectance change as a function of

transient concentration under these circumstances but a precise knowledge of the absorption and scattering coefficients is required. Under most circumstances, this case is avoided and diffuse reflectance flash photolysis experiments are arranged such that one of the two limiting cases above, generally the exponential fall-off (usually achieved by attenuation of the laser beam), prevails in a particular experiment.

Sample Geometry

In the case of conventional nanosecond laser flash photolysis, it is generally the case that right-angled geometry is employed for the photolysis and probe beams (Figure 3a). This geometry has a number of advantages over alternatives. The photolysis beam and analyzing beam are spatially well separated, such that the analyzing beam intensity is largely unaffected by scattered photolysis light. Also, the fluorescence generated by the sample will be emitted in all directions, whilst the analyzing beam is usually collimated allowing spatial separation of fluorescence and analyzing light; sometimes an iris is used to aid this spatial discrimination. This geometry is appropriate for quite large beam diameters and fluences; where smaller beams are used collinear geometry may be more appropriate in order to achieve long interaction pathlengths.

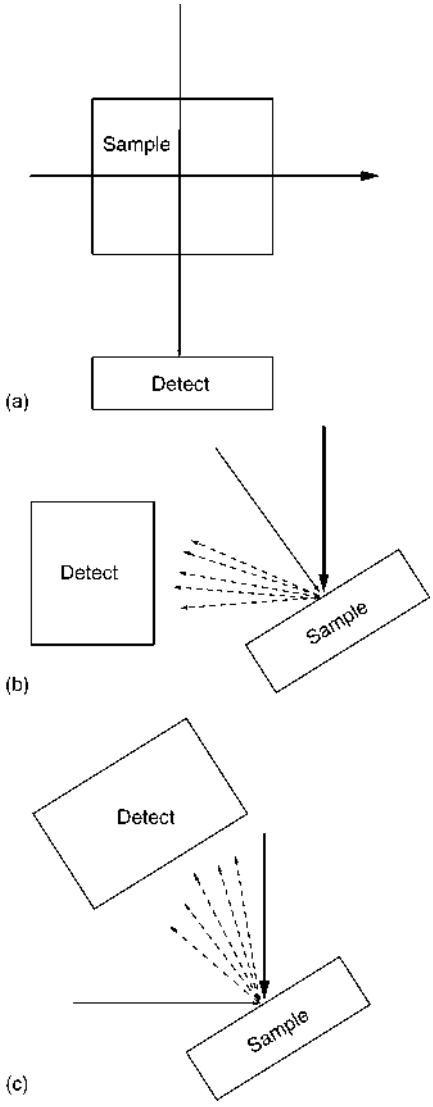

Figure 3 Sample geometries for (a) conventional and (b,c) diffuse reflectance flash photolysis. ⟶, photolysis beam; ⟶, analyzing beam.

In the case of nanosecond diffuse reflectance flash photolysis, the geometry required is quite different (Figure 3b,c). Here, the photolysis beam and analyzing beam must be incident on the same sample surface, and the diffusely reflected analyzing light is collected from the same surface. The geometry is often as shown in Figure 3b, where the analyzing light is incident almost perpendicularly on the sample surface, with the photolysis beam incident at an angle such that the specular reflection of the photolysis beam passes between detector and analyzing beam (not shown). Alternatively, the geometry shown in Figure 3c may be employed, where diffusely reflected light is detected emerging perpendicular to the sample surface. In both cases, the geometry is chosen such that specular (mirror) reflection of either exciting or analyzing light from the sample is not detected, since specular reflection is light which has not penetrated the sample and therefore contains no information regarding the concentrations of species present. A requirement, as in conventional flash photolysis, is that the analyzing beam probes only those areas which are excited by the photolysis beam, requiring the latter to be larger than the former. The nature of the scattering described previously means that the required geometry does not give spatial discrimination at the detector between photolysis and analyzing light, and fluorescence. This is since both analyzing and photolysis light undergo scattering and absorption processes (although with wavelength-dependent absorption and scattering coefficients) and emerge with the same spatial profiles. Fluorescence, principally that stimulated by the photolysis beam since this is of greatest intensity, originates from within the sample but again undergoes absorption and scattering and emerges with the same spatial distribution as the exciting light. In diffuse reflectance flash photolysis, separation of the analyzing light from the excitation or fluorescence must be achieved using spectral (filters and/or monochromators) or temporal, rather than spatial discrimination. Time-gated charge coupled devices

(CCD) or photodiode array detectors can be effectively used in nanosecond laser flash photolysis to exclude excitation light and fluorescence, since these occur on time-scales usually much shorter than the transient absorption of species of interest. The usual analyzing light source used in nanosecond diffuse reflectance flash photolysis is a xenon arc lamp, due to its good spectral coverage and high intensity. High probe intensity is particularly important in diffuse reflectance flash photolysis since the scattered intensities are often low, and the scattered light emerges in a large solid angle, only part of which can be effectively collected and detected. Also, high intensities allow for the light from the analyzing source to dominate over fluorescence if the latter is relatively weak.

In femtosecond diffuse reflectance laser flash photolysis, sample geometry considerations are also important. Such experiments are performed using the pump-probe technique, with the probe often being a femtosecond white-light continuum. The sample geometry employed is illustrated in Figure 4.

Here the pump and probe are almost colinear, and are incident on the same sample area; again, the requirement is for the pumped area to be larger than the probed area. Diffusely reflected light is then collected and analyzed, time resolution being achieved by varying the delay between pump and probe beams. It should be noted that the temporal resolution is worse than in conventional pump-probe techniques. In conventional femtosecond flash photolysis, the time resolution is generally limited by the widths of pump and probe pulses; in diffuse reflectance mode, the pulses undergo numerous refractions, reflections, and diffractions such that the pulse is temporally broadened during its transit through the material. The extent of this broadening is a sensitive function of the optical properties of the individual sample.

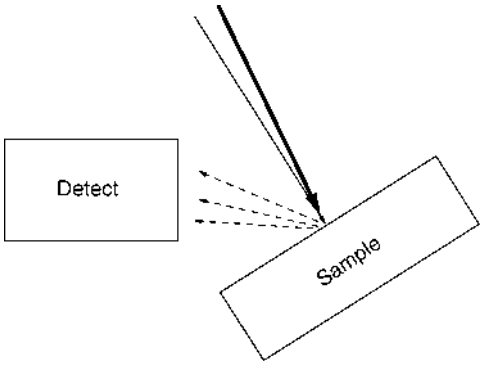

Figure 4 Sample geometry for femtosecond diffuse reflectance flash photolysis. ⟶, photolysis beam; ⟶, analyzing beam.

Kinetic Analysis

Kinetic analysis, and of course time-resolved spectroscopic analysis, require a quantitative treatment of the concentration changes within a sample following an excitation pulse as a function of time. When studying transient absorption phenomena in opaque samples, it is usual to define the reflectance change in terms of the sample reflectance before and after excitation, such that in spectrally resolved data a transient difference spectrum rather than an absolute reflectance spectrum of the transient species is obtained. The latter can, however, be reconstructed from a knowledge of the absorption coefficients and concentration of the species involved. It is possible to define the reflectance change as

$$\frac{\Delta J_t}{J_0} = 1 - \frac{R_t}{R_0} \qquad [13]$$

where R_0 and R_t represent the intensity of probe light diffusely reflected from the sample before excitation and at a time t after excitation, respectively. Frequently reflectance change is expressed as '% absorption', which is defined in eqn [14].

$$\%\ \text{absorption} = 100 \times \left(1 - \frac{R_t}{R_0}\right) \qquad [14]$$

These parameters are often used as being proportional to transient concentration, subject to satisfying the criteria for an exponential fall-off of transient concentration with penetration depth as discussed previously, and are used to replace transient concentration in kinetic equations used in data analysis.

It is generally the case that the samples studied using diffuse reflectance laser flash photolysis have either some degree of heterogeneity, for example paper, microcrystalline cellulose, silica gel or alumina, or have well-defined porous structures such as zeolites. Molecules adsorbed to these supports may be present on the surface, within micro- or mesopores, or intimately included within the structure. Hence each molecule may experience its own unique environment, and this will obviously influence its observed photophysics. It is therefore the case that in these systems, even very simple photo-induced reactions such as unimolecular photoisomerizations do not follow first-order kinetics; rather, a distribution of rates is observed which reflect the differing environments experienced by the molecules, and hence the molecules act as probes of this heterogeneity.

There are a number of approaches to kinetic analysis in opaque, often heterogeneous systems, as described below.

Rate Constant Distributions

Here the sample heterogeneity is treated as producing a series of micro-environments, and the reaction studied will have its own unique rate constant in each of these environments. The width of the observed rate constant distribution therefore describes the number of these possible environments, with the distribution amplitude at any particular rate constant, reflecting the probability of the molecule existing in the corresponding environment.

Exponential series lifetime distribution analysis is the simplest approach to distributed lifetime analysis. This analysis has the inherent advantage that there are no presuppositions regarding the kinetic model describing the data. Rather, a large number of exponentials with fixed rate constants and with amplitudes as adjustable parameters are applied to the data, and a least-squares procedure used to optimize the amplitude of each exponential. Generally there is a relationship between the rate constants, they being equally spaced on either a linear or logarithmic scale. Physically this can be interpreted as a large number of first-order or pseudo-first-order (in the case of bimolecular reactions) rates arising from the sample heterogeneity. Hence the rate constant distribution emerges naturally from the fitting procedure, with no preimposed constraints. Since there are a large number of adjustable parameters, these models are applied most successfully to data with good signal-to-noise.

Alternative approaches involve imposing a distribution shape onto a set of exponentials, and optimizing this distribution to the data. This approach has the advantage that by assuming a rate constant distribution, the relative amplitudes of the exponentials are fixed and hence the number of fitting parameters greatly reduced. One of the more successful models applied in this context is that developed by Albery *et al*. In the development of their model it is assumed that, for the reaction in question, the free energy of activation, $\Delta G^{\neq}$, is distributed normally around a mean value $\Delta\bar{G}^{\neq}$ according to eqn [15]:

$$\Delta G^{\neq} = \Delta\bar{G}^{\neq} - \gamma x RT \qquad [15]$$

with γ being the width of the distribution for $\infty \geq x \geq -\infty$. This assumed normal distribution of the free energy of activation leads to a log-normal distribution of the decay rate constants distributed around some average rate constant $\bar{k}$. The dispersion in the first-order rate constant is then

$$\ln(k) = \ln(\bar{k}) + \gamma x \qquad [16]$$

The equation used to describe the data is given as:

$$\frac{c}{c_0} = \frac{\int_{-\infty}^{\infty} \exp(-t^2)\exp\left[-\bar{k}t\exp(\gamma t)\right]\mathrm{d}t}{\int_{-\infty}^{\infty} \exp(-t^2)\mathrm{d}t} \qquad [17]$$

with c and c_0 being concentration at time $t = t$ and $t = 0$ relative to the excitation flash, respectively. Where reflectance change is proportional to transient concentration, reflectances can be used directly in place of concentrations. This equation can be solved by making the following substitutions:

$$\int_{-\infty}^{\infty} \exp(-t^2)\mathrm{d}t = \pi^{1/2} \qquad [18]$$

$$\int_{-\infty}^{\infty} \exp(-t^2)\exp\left[-\bar{k}t\exp(\gamma t)\right]\mathrm{d}t = \int_0^1 \lambda^{-1}\exp\left(-(\ln\lambda^2)\right)\left(\exp(-\bar{k}t\lambda^{\gamma}) + \exp(-\bar{k}t\lambda^{-\gamma})\right)\mathrm{d}\lambda \qquad [19]$$

Hence here there are only two fitting parameters, namely the width of the distribution γ and the distribution center, $\bar{k}$. Note that for $\gamma = 0$, eqn [17] reduces to a single exponential function.

A further model which constrains the rate constant distribution, which has been successfully applied to describe the rates of photo-induced processes on silica gel surfaces, is a Lévy stable distribution of rate constants, described as:

$$\mathrm{P}(\bar{k}, \alpha, \gamma) = \frac{1}{\pi}\int_0^{\infty} \exp(-\gamma q^{\alpha})\cos(\bar{k}q)\mathrm{d}q \qquad [20]$$

where α is the characteristic power law exponent ($0 < \alpha \leq 2$) and γ is the distribution width ($\gamma > 0$). Special cases of the Lévy stable distribution occur for $\alpha = 1$ and $\alpha = 2$, where the distribution shape becomes Lorentzian and Gaussian, respectively. The Lévy distribution gives an increasing weighting to the tails of the distribution as the distribution width decreases, and can be described as a random walk consisting of long jumps followed by several short walks. It has been shown that this type of foraging behavior is more efficient at seeking out randomly distributed targets than a simple random walk. The Lévy stable distribution has three adjustable parameters which allows greater flexibility in the distribution of rate constants than does the Gaussian model, but still constrains the distribution to be symmetrical about some mean value.

Physical Models

A number of these have been applied to very specific data sets, where parameters of the systems are accurately known. These include random walk models in zeolites, and time-dependent fractal-dimensional rate constants to describe kinetics on silica gel surfaces. The advantage of these methods over rate constant distributions is that since they are based on the physical description of the system, they yield physically meaningful parameters such as diffusion coefficients from analysis of kinetic data. However, for many systems the parameters are known in insufficient detail for accurate models to be developed.

Examples

An example of bimolecular quenching data is shown in Figure 5. Here, anthracene at a concentration of 1 μmol g^{-1} is co-adsorbed to silica gel from acetonitrile solution together with azulene at a concentration of 0.8 μmol g^{-1}. Laser excitation at 355 nm from an Nd:YAG laser produces the excited triplet state of the anthracene, which undergoes energy transfer to the co-adsorbed azulene molecule as a result of the rapid diffusion of the latter.

The data shown in Figure 5 are recorded monitoring at 420 nm, and the laser energy (approximately 5 mJ per pulse) is such that an exponential fall-off of transient concentration with penetration depth is expected such that reflectance change is proportional to transient concentration (see section on Transient Concentration Profiles above). The data are shown on a logarithmic time axis for clarity. The fitted line is obtained by applying the model of Albery *et al.*

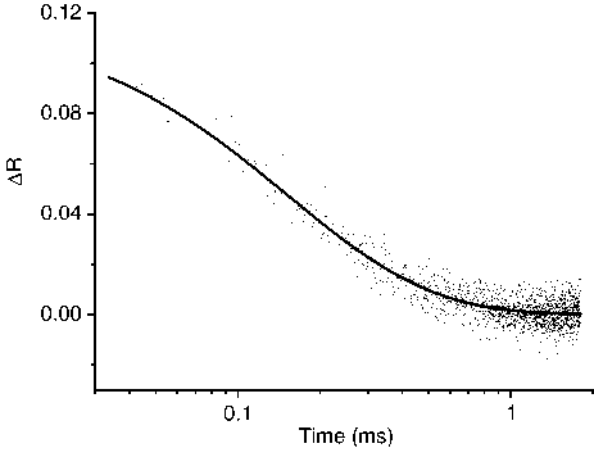

Figure 5 Transient absorption decay for anthracene (1 μmol g^{-1}) co-adsorbed with azulene (0.8 μmol g^{-1}) to silica gel monitored at 420 nm. Fitted using the model of Albery *et al.*

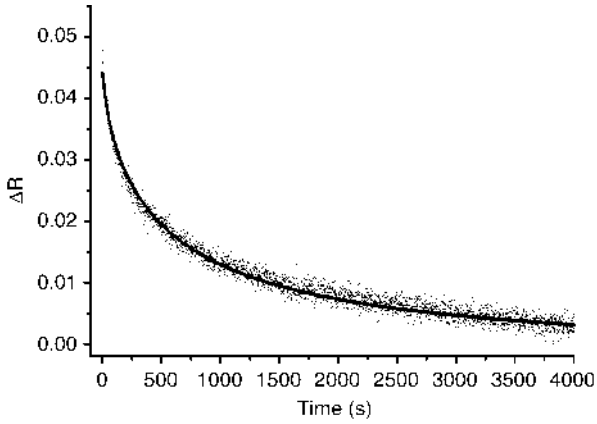

Figure 6 Transient absorption decay for the naphthalene (1 μmol g^{-1}) radical cation monitored at 680 nm on silica gel. Fitted using a Lévy stable distribution.

(see section on Rate Constant Distributions above) with fitting parameters $\bar{k} = 1.01 \times 10^4$ s^{-1} and $\gamma = 0.73$.

Where ion–electron recombination is concerned, the Albery model often fails to adequately describe the data obtained since it does not allow sufficient small rate constants relative to the value of $\bar{k}$, given the constraints of the Gaussian distribution. The Lévy stable distribution is ideal in this application since it allows greater flexibility in the shape of the distribution.

Figure 6 shows example data for naphthalene adsorbed to silica gel (1 μmol g^{-1}). The laser pulse energy at 266 nm (approximately 40 mJ per pulse) is such that photo-ionization of the naphthalene occurs producing the radical cation. The subsequent decay of the radical cation via ion–electron recombination can be monitored at 680 nm. Note that decay is observed on a time-scale of several thousand seconds. The fitted line is according to a Lévy stable distribution with parameters $\bar{k} = 8.2 \times 10^{-4}$ s^{-1}, $\gamma = 0.5$ and $\alpha = 1.7$ (see section on Rate Constant Distributions above). This model allows the shape of the distribution to deviate from a Gaussian, and can be more successful than the model of Albery *et al.* in modeling the complex kinetics which arise on surfaces such as silica gel. Note that where the model of Albery *et al.* can successfully model the data, the data can also be described by a Lévy stable distribution with $\alpha = 2$.

List of Units and Nomenclature

A	Sample absorbance
c	Concentration
I	Incident flux
J	Diffusely reflected flux

k	Rate constant
$\bar{k}$	Mean rate constant
K	Absorption coefficient (absorption per unit thickness) [cm^{-1}]
l	Pathlength [cm]
P	Probability of transmission of a photon through a sample of defined thickness
R_∞	Reflectance of an infinitely thick sample
S	Scattering coefficient (scattering per unit thickness) [cm^{-1}]
t	Time
X	Thickness of sample layer [cm]
α	Characteristic Lévy power law exponent
γ	Width of rate constant distribution
$\Delta G^{\neq}$	Free energy of activation [kJ mol^{-1}]
$\Delta \bar{G}^{\neq}$	Mean free energy of activation [kJ mol^{-1}]
ε	Molar decadic or naperian absorption coefficient

See also

Chemical Applications of Lasers: Pump and Probe Studies of Femtosecond Kinetics. **Optical Materials**: Measurement of Optical Properties of Solids. **Scattering**: Scattering from Surfaces and Thin Films; Scattering Theory.

Further Reading

Albery WJ, Bartlett PN, Wilde CP and Darwent JR (1985) A general-model for dispersed kinetics in heterogeneous systems. *Journal of the American Chemical Society* 107: 1854–1858.

Anpo M (ed.) (1996) Surface photochemistry. *Wiley Series in Photoscience and Photoengineering*, vol. 1. Chichester, UK: John Wiley and Sons.

Anpo M and Matsuura T (eds) (1989) Photochemistry on solid surfaces. *Elsevier Series in Studies in Surface Science and Catalysis*, vol. 47. Amsterdam, The Netherlands: Elsevier Science Publishers B.V.

Asahi T, Furube A, Fukumura H, Ichikawa M and Masuhara H (1998) Development of a femtosecond diffuse reflectance spectroscopic system, evaluation of its temporal resolution, and applications to organic powder systems. *Review of Scientific Instruments* 69: 361–371.

Bertoin J (1998) *Lévy Processes*. Cambridge, UK: Cambridge University Press.

Hapke B (1993) *Introduction to the Theory of Reflectance and Emittance Spectroscopy*. Cambridge, UK: Cambridge University Press.

Kamat PV (1993) Photochemistry on nonreactive and reactive (semiconductor) surfaces. *Chemical Reviews* 93: 287–300.

Kan HKA and Lin TP (1970) Calculation of reflectance of a light diffuser with non-uniform absorption. *Journal of the Optical Society of America* 60: 1252–1256.

Kessler RW and Wilkinson F (1981) Diffuse reflectance triplet–triplet absorption spectroscopy of aromatic hydrocarbons chemisorbed on γ-alumina. *Journal of the Chemical Society – Faraday Transactions I* 77: 309–320.

Kortum G (1969) *Reflectance Spectroscopy*. Berlin: Springer-Verlag.

Kubelka P and Munk F (1931) Ein beitrag zur optik der farbanstriche. *zeitschrift fuer Technische Physik* 12: 593–601.

Ramamurthy V (ed.) (1991) *Photochemistry in Organised & Constrained Media*. New York: VCH Publishers.

Thomas JK (1993) Physical aspects of photochemistry and radiation chemistry of molecules adsorbed on SiO_2, γ-Al_2O_3, zeolites and clays. *Chemical Reviews* 93: 301–320.

Thomas JK and Ellison EH (2001) Various aspects of the constraints imposed on the photochemistry of systems in porous silica. *Advances in Colloid and Interface Science* 89–90: 195–238.

Wendlandt WW and Hecht HG (1966) Reflectance Spectroscopy. *Chemical Analysis Series*, vol. 21. New York: Interscience Publishers.

Wilkinson F and Kelly G (1989) Diffuse reflectance flash photolysis. In: Scaiano JC (ed.) *Handbook of Organic Photochemistry*, vol. 1, pp. 293–314. Florida: CRC Press.

Laser Manipulation in Polymer Science

S Ito, Y Hosokawa and H Masuhara, Osaka University, Osaka, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Laser manipulation is a method for manipulating single particles with a size of less than a few tens of microns in which optical pressure of a focused laser beam is used to trap and control the particles without any mechanical contact. As infrared lasers, as a trapping light source, have become more user-friendly, even operators who are not familiar with lasers and microscopes can perform laser manipulation. Those laser manipulators already on sale have attracted significant attention as a new tool. It is especially interesting to combine this method with nanotechnology and biotechnology, which have progressed rapidly during the 1990s. In this article,

the principle and method of laser manipulation is described and then its applications and possibilities in nanotechnology and biotechnology are summarized.

Principle and Method

When light is reflected or refracted at the interface between two media with different refractive indices, the momentum of photons is changed, which leads to the generation of photon force as a reaction to the momentum change. For example, when a laser beam is reflected by a mirror, the momentum of photons is changed by Δp (Figure 1a). As a result, a photon force, F_{phot}, acts on the mirror to deflect it vertically away from the reflected beam. In refraction, the momentum of photons is also changed, so that the photon force acts on the interface, as shown in Figure 1b. Thus light does not only give its energy to materials via absorption but also applies a mechanical force to them. However, when we are exposed to light, such as from a halogen lamp, we are never aware of the photon force because its magnitude is less than an order of pN. However, the photon force of the pN order acquired from a focused laser beam is sufficient to manipulate nm–μm-sized particles in solution, under an optical microscope.

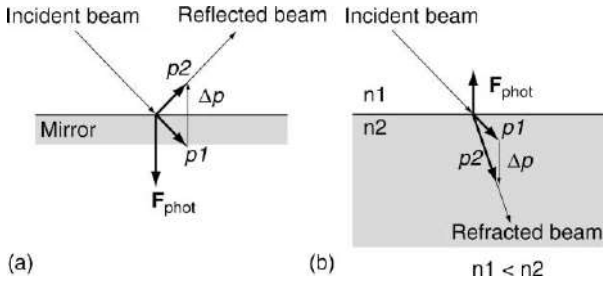

Figure 1 Optical pressure originating from the momentum change of photons.

Principle of Laser Trapping

If the size of a particle is larger than the wavelength of a trapping laser beam (μm-sized particles), the principle of single-beam gradient force optical trapping can be explained in terms of geometrical optics. When a tightly focused laser beam is irradiated onto a transparent dielectric particle, an incident laser beam is refracted at the interface between the particle and medium, as represented by Figure 2a. The propagation direction of the beam is changed, i.e., the momentum of the photon is changed, and consequently, a photon force is generated. As the laser beam leaves the particle and enters the surrounding medium, this refraction causes a photon force to be exerted again on that interface. Summing up the force contributions of all rays, if the refractive index of the particle is higher than that of the medium, a resultant force exerted on the particle is directed toward the focal point as an attractive force. However, reflection at the surface of the particle is negligible if that particle is transparent and the refractive index ratio of the particle to medium is close to unity. Therefore, the incident beam is reflected at two surfaces by a small amount, and as a result, the particle is directed slightly to the propagation direction of the incident light. Where the particle absorbs the incident beam, the photon force is also generated to push it in the propagation direction. The negligible effects of the

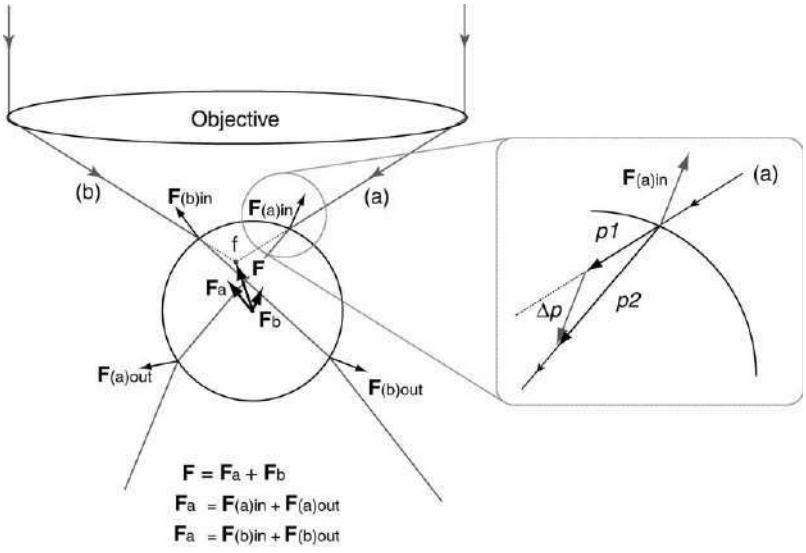

Figure 2 Principle of single-beam gradient force optical trapping explained by ray optics.

reflection and absorption occur in trapping experiment of transparent particles, such as polymer particles, silica microspheres, etc. However, particles with high reflectance and absorption coefficients such as metallic particles, exert a more dominant force at the surface where absorption and reflection occurs. This force is a repulsive force away from the focal point; consequently, metallic microparticles cannot be optically trapped.

If a dielectric particle is much smaller than the wavelength of a trapping light (nm-sized), it can be regarded as a point dipole (Rayleigh approximation) and the photon force ($\mathbf{F}_{\text{phot}}$) acted on it is given by

$$\mathbf{F}_{\text{phot}} = \mathbf{F}_{\text{grad}} + \mathbf{F}_{\text{scat}} \qquad [1]$$

Here, $\mathbf{F}_{\text{grad}}$ and $\mathbf{F}_{\text{scat}}$ are called the gradient force and scattering force, respectively. The scattering force is caused by the scattering of light, and it pushes the particle toward the direction of light propagation. On the other hand, the gradient force is generated when a particle is placed in a heterogeneous electric field of light. If the dielectric constant of the particle is higher than that of the surrounding medium, the gradient force acts on the particle to push it toward the higher intensity region of the beam. In the case of laser trapping of dielectric nanoparticles, the magnitude of gradient force is much larger than the scattering force. Consequently, the particle is trapped at the focal point of the trapping laser beam, where the beam intensity (electric field intensity) is maximum. The photon force can be expressed as follows:

$$\mathbf{F}_{\text{phot}} \cong \mathbf{F}_{\text{grad}} = \frac{1}{2}\varepsilon_{\text{m}}\alpha\nabla|\mathbf{E}|^2 \qquad [2]$$

$$\alpha = 3V\frac{\varepsilon_{\text{p}} - \varepsilon_{\text{m}}}{\varepsilon_{\text{p}} + 2\varepsilon_{\text{m}}} \qquad [3]$$

where, $\mathbf{E}$ is electric field of the light, V is volume of the particle, and ε_{p} and ε_{m} are dielectric constants of the particle and medium. In trapping nanoparticles that absorb the laser beam, such as gold nanoparticles, the complex dielectric constant and attenuation of the electric field in the particle need to be taken into consideration. Although the forces due to light absorption and scattering are both propelling the particle toward the direction of light propagation, the magnitude of gradient force is much larger than these forces. This is in contrast to the μm-sized metallic particle, which cannot be optically trapped.

Laser Manipulation System

An example of a laser micromanipulation system, with a pulsed laser to induce photoreactions, is shown in Figure 3. A linearly polarized laser beam from a CW Nd^{3+}:YAG laser is modulated to the circularly polarized light by a λ/4 plate and then split into horizontally and vertically polarized beams by a

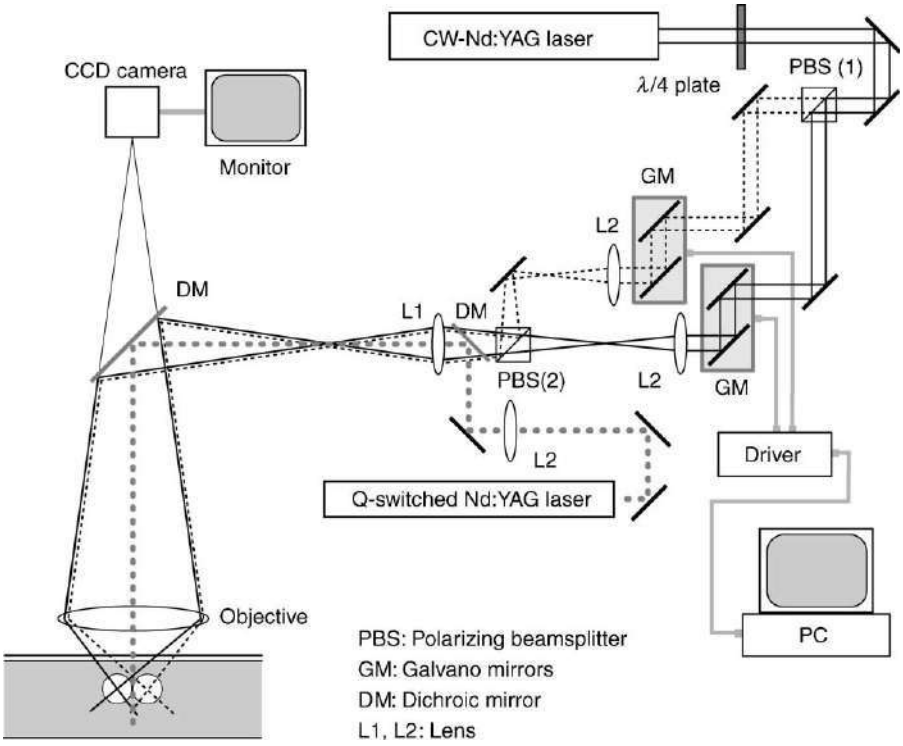

Figure 3 A block diagram of a dual beam laser manipulation-reaction system.

polarizing beamsplitter (PBS1). Both laser beams are combined by another (PBS2), introduced coaxially into an optical microscope, and then focused onto a sample through a microscope objective. The focal spots of both beams in the sample solution can be scanned independently by two sets of computer-controlled galvano mirrors. Even if the beams are spatially overlapping, interference does not take place because of their orthogonal polarization relation. Pulsed lasers, such as a Q-switched Nd^{3+}:YAG laser, are used to induce photoreactions, such as photopolymerization, photothermal reaction, laser ablation, etc.

Laser Manipulation and Patterning of Nanoparticles

Laser manipulation techniques enable us to capture and mobilize fine particles in solution. Most studies using this technique have been conducted on μm-sized objects such as polymer particles, microcrystals, living cells, etc. Because it is difficult to identify individual nanoparticles with an optical microscope, laser manipulation techniques have rarely been applied to nanotechnology and nanoscience. However, the laser trapping technique can be a powerful tool for manipulation of nanoparticles in solution where individual particles can be observed. This is achieved by detection of fluorescence emission from labeled dyes or scattered light. Now, even single molecule fluorescence spectroscopy has been achieved by the use of a highly sensitive photodetector, and a single metallic nanoparticle can be examined by detecting the light scattered from it. There have been several reports on the application of laser manipulation techniques for patterning of nm-sized particles. Here, fixation methods of nanoparticles, using the laser manipulation technique and local photoreactions, are introduced.

Patterning of Polymer Nanoparticles

Patterning of individual polymer nanoparticles onto a substrate can be achieved by using local photopolymerization. The following example shows the strength of this method. Polystyrene nanoparticles of 220 nm with fluorescent dye were dispersed in ethylene glycol solution containing polymerizable vinyl monomer (acrylamide, 31 wt%), crosslinker (*N*,*N*′-methylenebisacrylamide, 2.2 wt%), and radical photoinitiator (Irgacure2959, Ciba Specialty Chemicals, 1.1 wt%). When the sample was irradiated by blue light from a high-pressure mercury lamp, green fluorescence from dye molecules within each nanoparticle was observed. A nanoparticle that entered the region, irradiated by a near-infrared laser beam (1064 nm), was trapped at the focal point, and moved onto the surface of the glass substrate by handling the 3D stage of the microscope. An additional fixation laser beam (355 nm, 0.03 μJ, pulse duration ~6 ns, repetition rate ~5 Hz) was then focused to the same point for ~10 s, which led to the generation of acrylamide gel around the trapped nanoparticle. By repeating the procedure, patterning of single polymer nanoparticles on a glass substrate was achieved, and a fluorescence image of single nanoparticles as a letter 'H', is shown in Figure 4. A magnified atomic force microscope (AFM) image of one of the fixed nanoparticles is also shown, which confirms that only one polymer nanoparticle was contained in the polymerized gel.

Multiple polymer nanoparticles can also be gathered, patterned, and fixed on a glass substrate by scanning both trapping and fixation laser beams with the use of two pairs of galvano mirrors. The optical transmission and fluorescence images of the 'H' patterned nanoparticles on a glass substrate are shown in Figure 5a and b, respectively. The letter 'H' consists of three straight lines of patterned and fixed nanoparticles. The trapping laser beam (180 mW) was scanned at 30 Hz along each line with a length of ~10 μm on a glass substrate for 300 s. Nanoparticles were gathered and patterned along the locus of the focal point on the substrate. Then the fixation laser beam (0.097 μJ) was scanned for another 35 s. As a result, the straight lines of

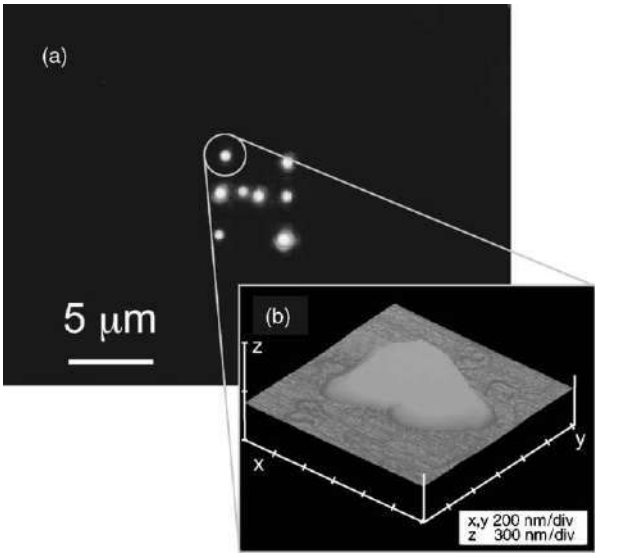

Figure 4 (a) A fluorescence image of spatially patterned individual polymer nanoparticles as the letter 'H' in distilled water. (b) A magnified AFM image of one of the produced acrylamide gels containing only one polymer nanoparticle on a glass substrate in the air.

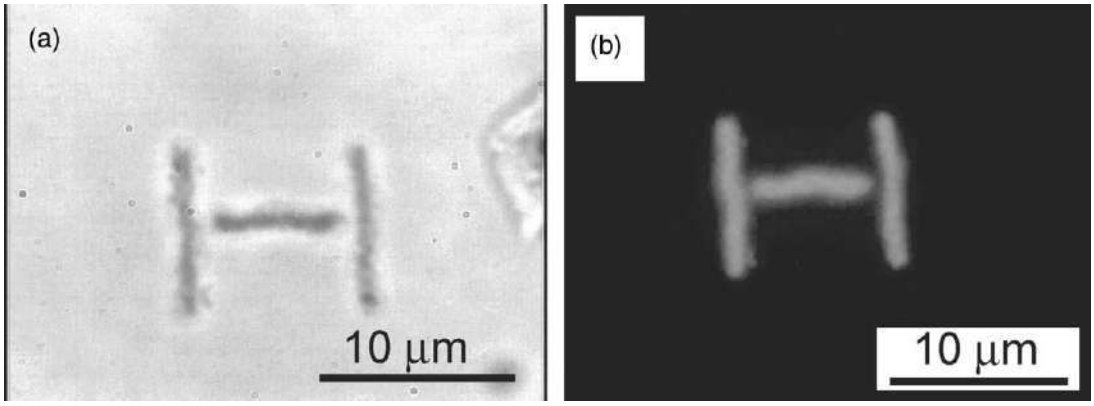

Figure 5 Optical transmission (a) and fluorescence (b) images of patterned and fixed polymer nanoparticles on the glass substrate, as the letter 'H'.

patterned nanoparticles were fixed in generated acrylamide gel on the substrate. Combining several simple fixed patterns, a more complex arrangement can be created with use of the present manipulation and fixation techniques.

Fixation of Individual Gold Nanoparticles

Manipulation and fixation of single metallic nanoparticles in solution has been achieved by means of photo-induced transient melting of nanoparticles. As an example, the following method describes this technique. Gold nanoparticles (diameter ~80 nm) were dispersed in ethylene glycol. In order to identify a single gold nanoparticle, extinction of light from the halogen lamp of the optical microscope by the particle was utilized. Thus, the trapped particle was observed as a black spot in transmission image. The optically trapped single gold nanoparticle was then transferred to a precise position on the surface of glass substrate in the sample solution. The focused 355 nm pulse was additionally irradiated onto the pressed nanoparticle, which led to a transient temperature elevation of the gold nanoparticle to enable its fixation. It was confirmed by AFM observation that, at suitable laser fluence (32–64 mJ cm^{-2}), a gold nanoparticle was fixed on the glass substrate without fragmentation. Repeating the same manipulation and fixation procedure, single gold nanoparticles could be patterned on a glass substrate. Figure 6 shows an AFM image of successive spatial patterning of single 80 nm gold particles.

The significance of the laser manipulation-fixation technique is that we can trap, manipulate, and fix single and/or many nanoparticles in solution at room temperature. We have already demonstrated assembling entangled polymer chains of 10–20 nm mean radius by laser trapping, and formation of a single μm-sized particle. Thus, these various kinds of nanoscopic materials can be well manipulated using these techniques and it is believed that the present nanomanipulation-fixation method will be useful for future nanoscience and nanotechnology.

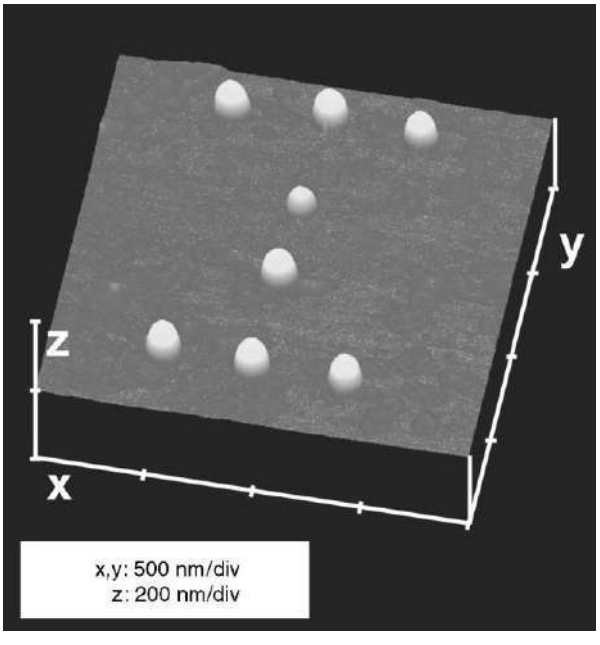

Figure 6 An AFM image of fixed and patterned gold nanoparticles, as the letter I.

Application to Biotechnology

Recent progress in biotechnology, using single cell manipulation by laser and microscope, has been attracting significant attention. In conventional methods, cell manipulation has been performed by mechanical manipulation using microneedles and micropipettes. However, laser manipulation can be applied as microtweezers. In comparison with the former, this form of laser manipulation has the

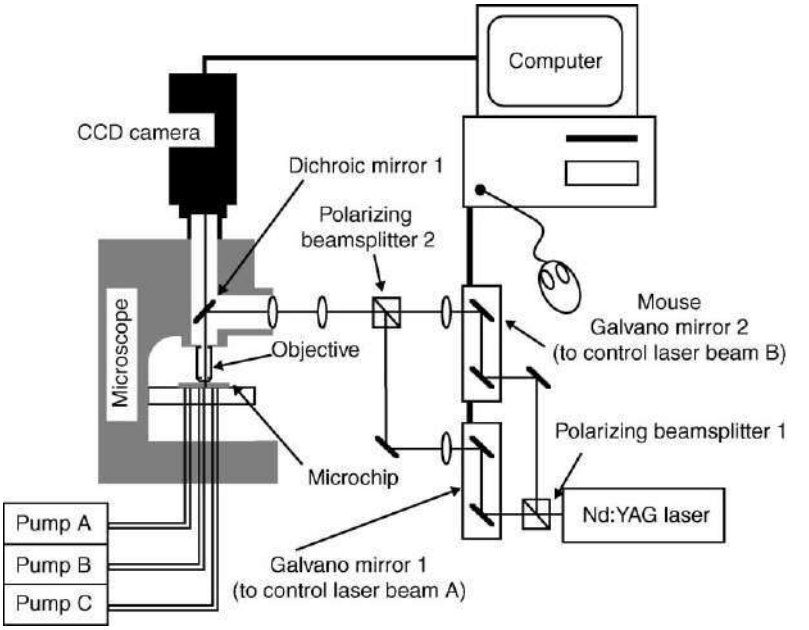

Figure 7 An experimental setup of dual-beam laser manipulation and microfluidic devices.

advantage of no contact with cells and can perform some complex and characteristic movement, for example, rotation, separation, accumulation, etc. Furthermore, by combining it with a microcutter using laser ablation, cell fusion and extraction/injection of organs from/into the cells can be achieved. In this section, single-cell manipulation achieved by using the laser manipulation, is introduced.

Noncontact Rotation of Cells Using Dual Beam Laser Manipulation

As a special feature of laser-cell manipulation, it is noticeable that multiple laser tweezers can be operated independently without any contact. Here, noncontact rotation of the fission yeast, *Schizosaccharomyces pombe* (h^-), demonstrates this method. The yeast has the shape of an ellipse of length 8 and 3 μm for major and minor axes, and is shown set on a dual-beam laser manipulation system in Figure 7. One trapping beam A was focused at the end of the cell to anchor it, and another beam, B, was at the other end and scanned circularly around beam A by controlling the Galvano mirrors. The rotation of the yeast cell was realized at a frequency of less than 2 Hz, as shown in Figure 8. Such a cell manipulation is impossible by mechanical manipulation and indicates the superior performance of laser manipulation.

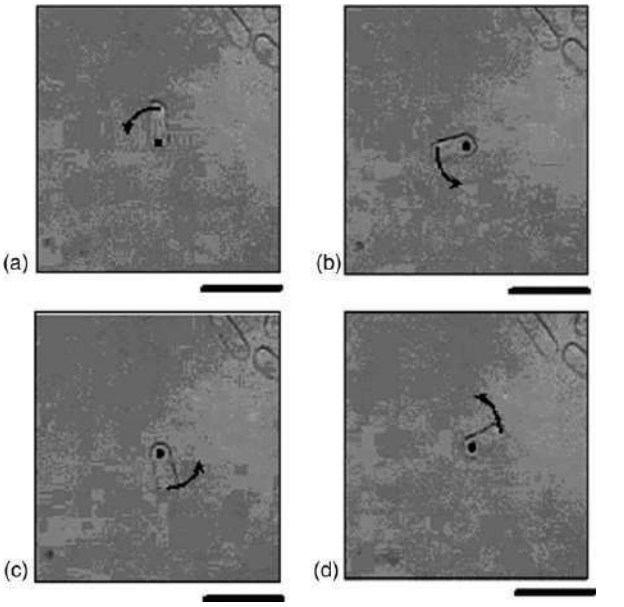

Figure 8 Rotation of fission yeast cells using dual-beam laser manipulation. The bars are 10 μm.

Transfer of Cells in Microchannel

A flow cytometry to physically separate and identify specific types of cells from heterogeneous populations by fluorescence, which is called fluorescence-activated cell sorter (FACS), has attracted significant attention as an important technique in biotechnology. In the separation process, charged single droplets containing single, fluorescence labeled cells are selected by passing them between two high-voltage deflection plates. Since it is a sequential process, which does not use microscopy, flexible and high-purity cell separation is limited. To overcome this problem, a selective cell separation system,

combining the laser trapping with the microchannel, is used. The method using microscopy is useful and flexible, because cells are not simply detected by fluorescence but also identified by their shape, size, surface morphology, absorption, etc. However, cell sorting under a microscope is not processed efficiently by laser trapping. A more effective method is described by the following.

The microchannel was prepared with laser microfabrication and photopolymerization, whose top view is shown in Figure 9a, and was then set on the microscope stage. Three syringes were equipped with homemade pumps A, B, and C, and connected with a microchip. There were two wing-like chambers in the microchip, that were connected with pumps A and C. Between chambers, there was a microchannel 100 μm wide, which joined these chambers and crossed a long microchannel connected with the pump B, thus forming a drain. Thickness of these chambers was 100 μm. By controlling these pumps, the left and right chambers were filled by the culture medium including the yeast cells and the neat culture medium, respectively, as shown in Figure 9b.

Individual cells were transferred from left to right chambers by using a single laser beam and handling an electrically movable microscope stage. The trapping laser irradiated a yeast cell in the left chamber and the trapped cell was transferred to the right chamber. A representative demonstration is given in Figure 9c. The position of the trapped yeast cell was controlled by the microscope stage with the velocity of 20 μm/s. The single yeast cells were successfully transferred from left to right chambers at the rate of 26 s. By combining laser trapping with the microchannel, single separation of cells was achieved under a microscope.

Collection and Alignment of Cells

When cell sorting using laser trapping is applied to biotechnology, the time to transfer cells should be short compared to the above-mentioned times. One method is to transfer multiple cells simultaneously by developing dual-beam laser manipulation, which is shown schematically in Figure 10. One trapping beam A was scanned on a line given as an arrow in Figure 10a at a rate faster than motions of cells.

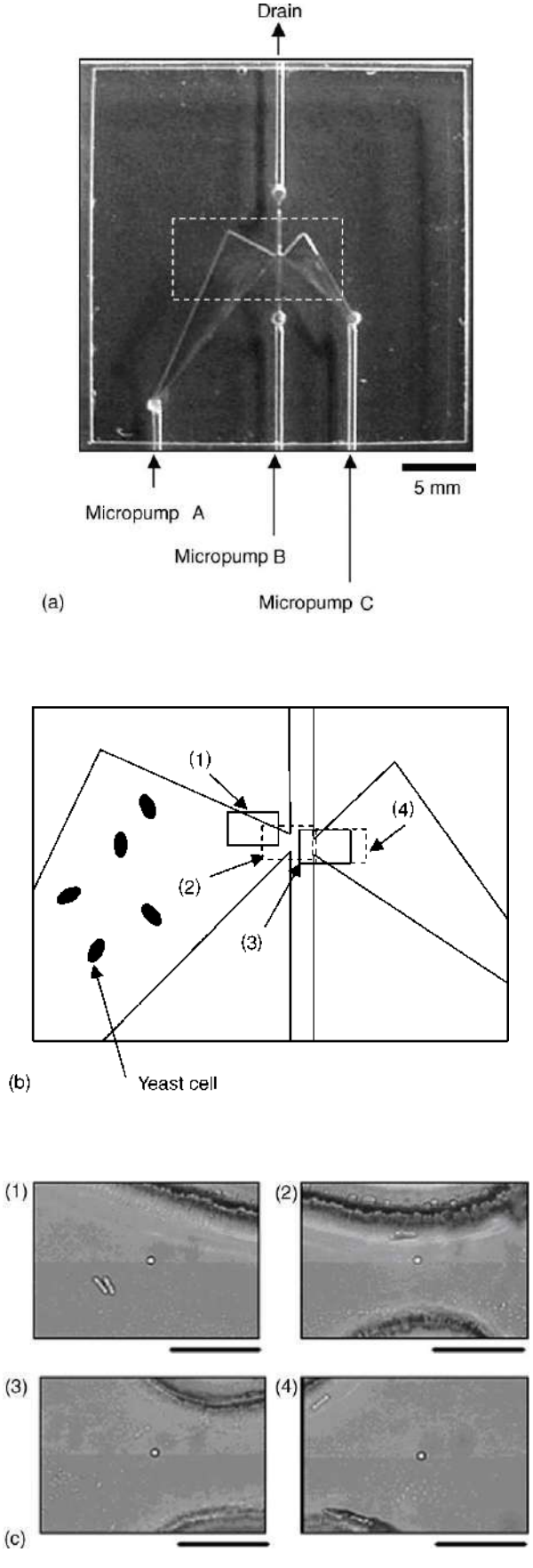

Figure 9 (a) A microchip fabricated by laser polymerization technique and (b) its schematic representation corresponding to dashed area in the microchip. (c) Cell transfer in microchip using laser trapping. The bars are 100 μm. A trapping laser beam is fixed at the center of each picture, by which a particle is trapped and transferred from left to right chambers. The position of each picture is shown in (b).

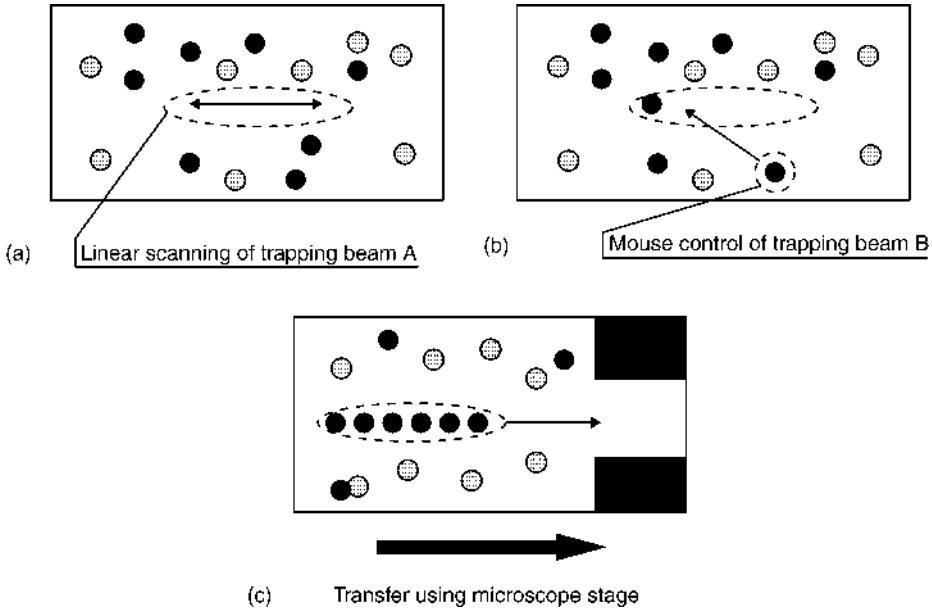

Figure 10 A schematic illustration of efficient collection and transfer of cells based by dual-beam laser manipulation.

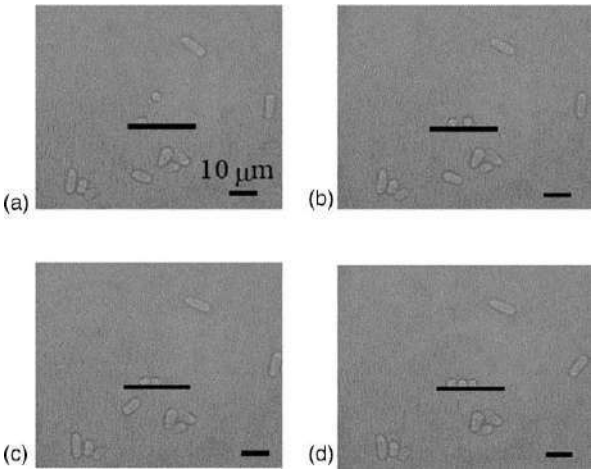

Figure 11 Collection and alignment of fission yeast cells using dual-beam laser manipulation.

The other trapping beam B was used to move individual cells to the line sequentially (Figure 10b). Finally, by moving the microscope stage, many particles trapped at the scanning line can be transferred to another chamber (Figure 10c).

An example is shown in Figure 11. The trapping beam A was scanned on a line with the rate of 15 Hz whose length was 30 μm. The trapping beam B was used to trap a yeast cell and move it to the line, which was controlled by a computer mouse pointer. The cells stored on the line were successfully isolated and with this method three cells can be collected within 15 s. Furthermore, the cells could be transferred with the velocity of 20 μm/s by driving the microscope stage. By improving the presented cell-sorting system, it is expected to realize flexible and high-purity cell separation.

List of Units and Nomenclature

Laser fluence (light energy density per pulse)	[mJ cm^{-2}]

See also

Time-Resolved Fluorescence: Measurements in Polymer Science.

Further Reading

Ashkin A, Dziedzic MJ, Bjorkholm JE and Chu S (1986) Observation of a single-beam gradient force optical trap for dielectric particles. *Optical Letters* 11: 288–290.

Ashkin A, Dziedzic MJ and Yamane T (1987) Optical trapping and manipulation of single cells using infrared laser beams. *Nature* 330: 769–771.

Berns WM, Aist J, Edwards J, *et al.* (1981) Laser microsurgery in cell and developmental biology. *Science* 213: 505–513.

Hoffmann F (1996) Laser microbeams for the manipulation of plant cells and subcellular structures. *Plant Science* 113: 1–11.

Hosokawa Y, Masuhara H, Matsumoto Y and Sato S (2002) Dual-beam laser micromanipulation for sorting

biological cells and its device application. *Proceedings of SPIE* 4622: 138–142.

Ito S, Yoshikawa H and Masuhara H (2001) Optical patterning and photochemical fixation of polymer nanoparticles on glass substrates. *Applied Physics Letters* 78: 2566–2568.

Ito S, Yoshikawa H and Masuhara H (2002) Laser manipulation and fixation of single gold nanoparticles in solution at room temperature. *Applied Physics Letters* 80: 482–484.

Kamentsky AL, Melamed RM and Derman H (1965) Spectrophotometer: new instrument for ultrarapid cell analysis. *Science* 150: 630–631.

Kim BH, Kogi O and Kitamura N (1999) Single-microparticle measurements: laser trapping-absorption microscopy under solution-flow conditions. *Analytical Chemistry* 71: 4338–4343.

Masuhara H, De Schryver FC, Kitamura N and Tamai N (eds) (1994) *Microchemistry: Spectroscopy and Chemistry in Small Domains*. Amsterdam: Elsevier-Science.

Sasaki K, Koshioka M, Misawa H, Kitamura N and Masuhara H (1991) Laser-scanning micromanipulation and spatial patterning of fine particles. *Japanese Journal of Applied Physics* 30: L907–L909.

Schütze K, Clement-Segelwald A and Ashkin A (1994) Zona drilling and sperm insertion with combined laser microbeam and optical tweezers. *Fertility and Sterility* 61: 783–786.

Shikano S, Horio K, Ohtsuka Y and Eto Y (1999) Separation of a single cell by red-laser manipulation. *Applied Physics Letters* 75: 2671–2673.

Svoboda K and Block SM (1994) Optical trapping of metallic Rayleigh particles. *Optical Letters* 19: 930–932.

Nonlinear Spectroscopies

S R Meech, University of East Anglia, Norwich, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Chemistry is concerned with the induction and observation of changes in matter, where the changes are to be understood at the molecular level. Spectroscopy is the principal experimental tool for connecting the macroscopic world of matter with the microscopic world of the molecule, and is, therefore, of central importance in chemistry. Since its invention, the laser has greatly expanded the capabilities of the spectroscopist. In linear spectroscopy, the monochromaticity, coherence, high intensity, and high degree of polarization of laser radiation are ideally suited to high-resolution spectroscopic investigations of even the weakest transitions. The same properties allowed, for the first time, the investigation of the nonlinear optical response of a medium to intense radiation. Shortly after the foundations of nonlinear optics were laid, it became apparent that these nonlinear optical signals could be exploited in molecular spectroscopy, and since then a considerable number of nonlinear optical spectroscopies have been developed. This short article is not a comprehensive review of all these methods. Rather, it is a discussion of some of the key areas in the development of the subject, and indicates some current directions in this increasingly diverse area.

Nonlinear Optics for Spectroscopy

The foundations of nonlinear optics are described in detail elsewhere in the encyclopedia, and in some of the texts listed in the Further Reading section at the end of this article. The starting point is usually the nonlinearity in the polarization, P_i, induced in the sample when the applied electric field, E, is large:

$$P_i = \varepsilon_0\left[\chi_{ij}^{(1)}E_j + \frac{1}{2}\chi_{ijk}^{(2)}E_jE_k + \frac{1}{4}\chi_{ijkl}^{(3)}E_jE_kE_l + \cdots\right] \quad [1]$$

where ε_0 is the vacuum permittivity, $\chi^{(n)}$ is the nth order nonlinear susceptibility, the indices represent directions in space, and the implied summation over repeated indices convention is used. The signal field, resulting from the nonlinear polarization, is calculated by substituting it as the source polarization in Maxwell's equations and converting the resulting field to the observable, which is the optical intensity.

The nonlinear polarization itself arises from the anharmonic motion of electrons under the influence of the oscillating electric field of the radiation. Thus, there is a microscopic analog of eqn [1] for the induced molecular dipole moment, μ_i:

$$\mu_i = [\varepsilon_0\alpha_{ij}d_j + \varepsilon_0^{-2}\beta_{ijk}d_jd_k + \varepsilon_0^{-3}\gamma_{ijkl}d_jd_kd_l + \cdots] \quad [2]$$

in which α is the polarizability, β the first molecular hyperpolarizability, γ the second, etc. The power series is expressed in terms of the displacement field **d** rather than **E** to account for local field effects. This somewhat complicates the relationship between the molecular hyperpolarizabilities, for example, γ_{ijkl} and

the corresponding macroscopic susceptibility, $\chi^{(3)}_{ijkl}$, but it is nevertheless generally true that a molecule exhibiting a high value of γ_{ijkl} will also yield a large third-order nonlinear polarization. This relationship between the macroscopic and microscopic para meters is the basis for an area of chemistry which has influenced nonlinear optics (rather than the inverse). The synthesis of molecules which have giant molecular hyperpolarizabilities has been an active area, because of their potential applications in lasers and electro-optics technology. Such molecules commonly exhibit a number of properties, including an extended linear π electron conjugation and a large dipole moment, which changes between the ground and excited electronic states. These properties are in accord with the predictions of theory and of quantum chemical calculations of molecular hyperpolarizabilities.

Nonlinear optical spectroscopy in the frequency domain is carried out by measuring the nonlinear signal intensity as a function of the frequency (or frequencies) of the incident radiation. Spectroscopic information is accessible because the molecular hyperpolarizability, and therefore the nonlinear susceptibility, exhibits resonances: the signal is enhanced when one or more of the incident or generated frequencies are resonant with the frequency of a molecular transition. The rich array of nonlinear optical spectroscopies arises in part from the fact that with more input fields there are more accessible resonances than is the case with linear spectroscopy. As an example we consider the third-order susceptibility, $\chi^{(3)}_{ijkl}E_jE_kE_l$, in the practically important case of two incident fields at the same frequency, ω_1, and a third at ω_2. The difference between the two frequencies is close to the frequency of a Raman active vibrational mode, Ω_{rg} (Figure 1). The resulting susceptibility can be calculated to have the form:

$$\chi^{(3)}_{ijkl}(-2\omega_1+\omega_2;\omega_1,\omega_1,-\omega_2) = \frac{N\Delta\rho_{rg}\Omega_{rg}(\alpha^{\mathrm{R}}_{ij}\alpha^{\mathrm{R}}_{kl}+\alpha^{\mathrm{R}}_{ik}\alpha^{\mathrm{R}}_{jl})}{12\hbar[\Omega^2_{rg}-(\omega_1-\omega_2)^2+\Gamma^2-2\mathrm{i}(\omega_1-\omega_2)\Gamma]} \quad [3]$$

in which the α^{R} are elements of the Raman susceptibility tensor, Γ is the homogeneous linewidth of the Raman transition, $\Delta\rho_{rg}$ a population difference, and N the number density. A diagram illustrating this process is shown in Figure 1, where the signal field is at the anti-Stokes Raman frequency $(2\omega_1-\omega_2)$. The spectroscopic method which employs this scheme is called coherent anti-Stokes Raman spectroscopy (CARS) and is one of the most widely applied nonlinear optical spectroscopies (see below). Clearly, from eqn [3] we can see that the signal will be enhanced when the difference frequency is resonant with the Raman transition frequency.

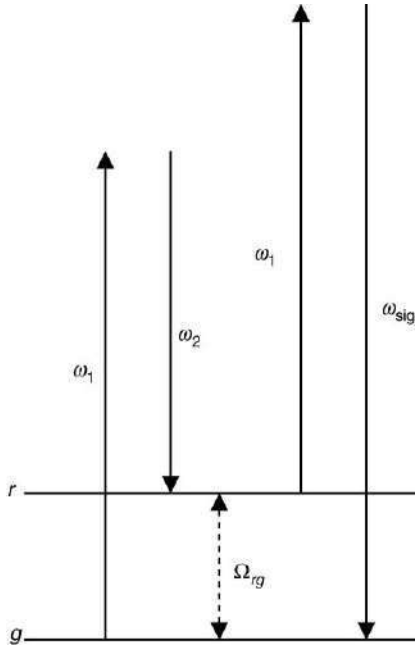

Figure 1 An illustration of the resonance enhancement of the CARS signal, $\omega_{\mathrm{sig}} = 2\omega_1 - \omega_2$ when $\omega_1 - \omega_2 = \Omega_{rg}$.

With essentially the same experimental geometry there will also be nonlinear signals generated at both the Stokes frequency $(2\omega_2-\omega_1)$ and at the lower of the incident frequencies, ω_2. These signals have distinct phase matching conditions (see below), so they can easily be discriminated from one another, both spatially and energetically. Additional resonance enhancements are possible if either of the individual frequencies is resonant with an electronic transition of the molecule, in which case information on Raman active modes in the excited state is also accessible.

It is worthwhile noting here that there is an equivalent representation of the nth order nonlinear susceptibility tensor $\boldsymbol{\chi}^{(n)}(\omega_{\mathrm{sig}}:\omega_1,\ldots,\omega_n)$ as a time domain response function, $\mathbf{R}^{(n)}(\tau_1,\ldots,\tau_n)$. While it is possible to freely transform between them, the frequency domain representation is the more commonly used. However, the response function approach is increasingly applied in time domain nonlinear optical spectroscopy when optical pulses shorter than the homogeneous relaxation time are used. In that case, the time ordering of the incident fields, as well as their frequencies, is of importance in defining an experiment.

An important property of many nonlinear optical spectroscopies is the directional nature of the signal, illustrated in Figure 2. The directional nonlinear

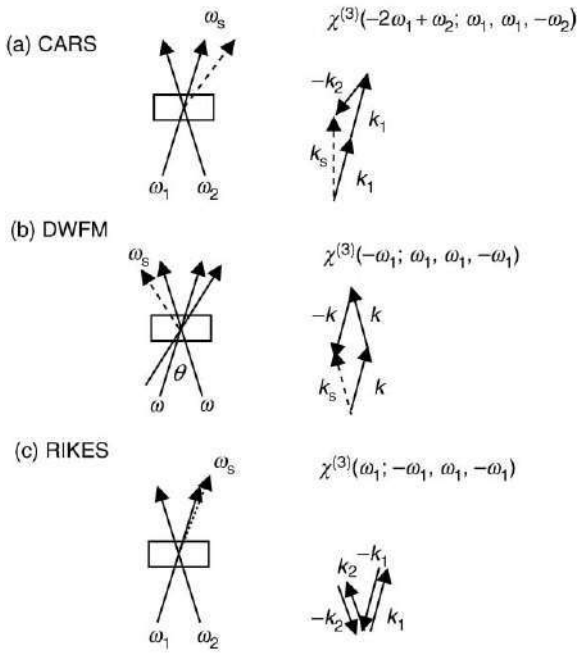

Figure 2 Experimental geometries and corresponding phase matching diagrams for (a) CARS (b) DFWM (c) RIKES. Note that only the RIKES geometry is fully phase matched.

signal arises from the coherent oscillation of induced dipoles in the sample. Constructive interference can lead to a large enhancement of the signal strength. For this to occur the individual induced dipole moments must oscillate in phase – the signal must be phase matched. This requires the incident and the generated frequencies to travel in the sample with the same phase velocity, $k_\omega/\omega = c/n_\omega$ where n_ω is the index of refraction and k_ω is the wave propagation constant at frequency ω. For the simplest case of second-harmonic generation (SHG), in which the incident field oscillates at ω and the generated field at 2ω, phase matching requires $2k_\omega = k_{2\omega}$. The phase-matching condition for the most efficient generation of the second harmonic is when the phase mismatch, $\Delta k = 2k_\omega - k_{2\omega} = 0$, but this is not generally fulfilled due to the dispersion of the medium, $n_\omega < n_{2\omega}$. In the case of SHG, a coherence length L can be defined as the distance traveled before the two waves are 180° out of phase, $L = |\pi/\Delta k|$. For cases in which the input frequencies also have different directions, as is often the case when laser beams at two frequencies are combined (e.g., CARS), the phase-matching condition must be expressed as a vectorial relationship, hence, for CARS, $\Delta\mathbf{k} = \mathbf{k}_{AS} - 2\mathbf{k}_1 + \mathbf{k}_2 \approx 0$, where $\mathbf{k}_{AS}$ is the wavevector of the signal at the anti-Stokes frequency (Figure 2). Thus for known input wavevectors one can easily calculate the expected direction of the output signal, $\mathbf{k}_s$. This is illustrated for a number of important cases in Figure 2.

Nonlinear Optical Spectroscopy in Chemistry

As already noted, there is a vast array of nonlinear optical spectroscopies, so it is clear some means of classification will be required. For coarse graining the order of the nonlinear process is very helpful, and that is the scheme we will follow here.

Second-Order Spectroscopies

Inspection of eqn [1] shows that in gases, isotropic liquids, and solids where the symmetry point group contains a center of inversion, $\boldsymbol{\chi}^{(2)}$ is necessarily zero. This is required to satisfy the symmetry requirement that polarization must change sign when the direction of the field is inverted, yet for a quadratic, or any even order dependence on the field strength, it must remain positive. Thus second-order nonlinearities might not appear very promising for spectroscopy. However, there are two cases in which second-order nonlinear optical phenomena are of very great significance in molecular spectroscopy, harmonic generation and the spectroscopy of interfaces.

Harmonic conversion

Almost every laser spectroscopist will have made use of second-harmonic or sum frequency generation for frequency conversion of laser radiation. Insertion of an oscillating field of frequency ω into eqn [1] yields a second-order polarization oscillating of 2ω. If two different frequencies are input, the second-order polarization oscillates at their sum and difference frequencies. In either case, the second-order polarization acts as a source for the second-harmonic (or sum, or difference frequency) emission, provided $\boldsymbol{\chi}^{(2)}$ is nonzero. The latter can be arranged by selecting a noncentrosymmetric medium for the interaction, the growth of such media being an important area of materials science. Optically robust and transparent materials with large values of $\boldsymbol{\chi}^{(2)}$ are available for the generation of wavelengths shorter than 200 nm to longer than 5 μm. Since such media are birefringent by design, a judicious choice of angle and orientation of the crystal with respect to the input beams allows a degree of control over the refractive indices experienced by each beam. Under the correct phase matching conditions $n_\omega \approx n_{2\omega}$, and very long interaction lengths result, so the efficiency of signal generation is high.

Higher-order harmonic generation in gases is an area of growing importance for spectroscopy in the

deep UV and X-ray region. The generation of such short wavelengths depends on the ability of amplified ultrafast solid state lasers to generate extremely high instantaneous intensities. The mechanism is somewhat different to the one outlined above. The intense pulse is injected into a capillary containing an appropriate gas. The high electric field of the laser causes ionization of atoms. The electrons generated begin to oscillate in the applied laser field. The driven recombination of the electrons with the atoms results in the generation of the high harmonic emission. Although the mechanism differs from the SHG case, questions of phase matching are still important. By containing the gas in a corrugated waveguide phase matching is achieved, considerably enhancing the intensity of the high harmonic. Photon energies of hundreds of electronvolts are possible using this technique. The generation of such high energies is not yet routine, but a number of potential applications are already apparent. A powerful coherent source of X-ray and vacuum UV pulses will certainly aid surface analysis techniques such as UV photoemission and X-ray photoelectron spectroscopy. Much excitement is currently being generated by the possibility of using intense ultrashort X-ray pulses to record structural dynamics on an ultrafast timescale.

Surface second-order spectroscopy

At an interface inversion symmetry is absent by definition, so the second-order nonlinear susceptibility is finite. If the two bulk phase media are themselves isotropic, then even a weak second-order signal necessarily arises from the interface. This surface-specific all-optical signal is unique, because it can be used to probe the interface between two condensed phases. This represents a great advantage over every other form of surface spectroscopy. In linear optical spectroscopy, the signal due to species at the interface are usually swamped by contributions from the bulk phase. Other surface-specific signals do exist, but they rely on the scattering of heavy particles (electrons, atoms) and so can only be applied to the solid vacuum interface. For this reason the techniques of surface SHG and SFG are widely applied in interface spectroscopy.

The most widely used method is sum frequency generation (SFG) between temporally overlapped tuneable infrared and fixed frequency visible lasers, to yield a sum frequency signal in the visible region of the spectrum. The principle of the method is shown in Figure 3. The surface nonlinear susceptibility exhibits resonances at vibrational frequencies, which are detected as enhancements in the visible SFG intensity. Although the signal is weak, it is directional and

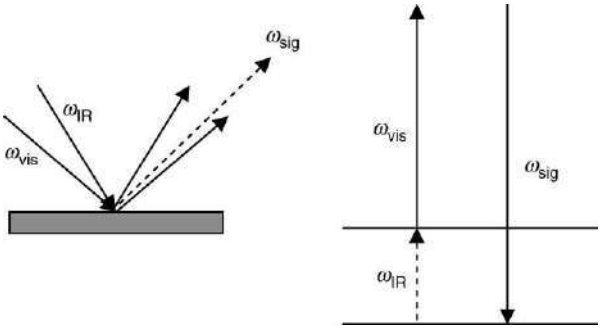

Figure 3 The experimental geometry for SFG, and an illustration of the resonance enhancement of $\omega_{sig} = \omega_{IR} + \omega_{vis}$ at a Raman and IR allowed vibrational transition.

background free, so relatively easily measured by photon counting techniques. Thus, SFG is used to measure the vibrational spectra of essentially any optically accessible interface. There are, however, some limitations on the method. The surface must exhibit a degree of order – if the distribution of adsorbate orientation is isotropic the signal again becomes zero by symmetry. Second, a significant enhancement at the vibrational frequency requires the transition to be both IR and Raman allowed, as suggested by the energy level diagram (Figure 3).

The SHG signal can also be measured as a function of the frequency of the incident laser, to recover the electronic spectrum of the interface. This method has been used, particularly in the study of semiconductor surfaces, but generally the electronic spectra of adsorbates contain less information than their vibrational spectra. However, by measuring the SHG intensity as a function of time, information on adsorbate kinetics is obtained, provided some assumptions connecting the surface susceptibility to the molecular hyperpolarizability are made. Finally, using similar assumptions, it is possible to extract the orientational distribution of the adsorbate, by measuring the SHG intensity as a function of polarization of the input and output beams. For these reasons, SHG has been widely applied to analyze the structure and dynamics of interfaces.

Third-Order Spectroscopies

The third-order coherent Raman spectroscopies were introduced above. One great advantage of these methods over conventional Raman is that the signal is generated in a coherent beam, according to the appropriate phase matching relationship (Figure 2). Thus, the coherent Raman signal can easily be distinguished from background radiation by spatial filtering. This has led to CARS finding widespread application in measuring spectra in (experimentally)

hostile environments. CARS spectroscopy has been widely applied to record the vibrational spectra of flames. Such measurements would obviously be very challenging for linear Raman or IR, due to the strong emission from the flame itself. The directional CARS signal in contrast can be spatially filtered, minimizing this problem. CARS has been used both to identify unstable species formed in flames, and to probe the temperature of the flame (e.g., from measured population differences eqn [3]). A second advantage of the technique is that the signal is only generated when the two input beams overlap in space. Thus, small volumes of the sample can be probed. By moving the overlap position around in the sample, spatially resolved information is recovered. Thus, it is possible to map the population of a particular transient species in a flame.

CARS is probably the most widely used of the coherent Raman methods, but it does have some disadvantages, particularly in solution phase studies. In that case the resonant $\boldsymbol{\chi}^{(3)}$ signal (eqn [3]) is accompanied by a nonresonant third-order background. The interference between these two components may result in unusual and difficult to interpret lineshapes. In this case, some other coherent Raman methods are more useful. The phase matching scheme for Raman-induced Kerr effect spectroscopy (RIKES) was shown in Figure 2. The RIKES signal is always phase matched, which leads to a long interaction length. However, the signal is at ω_2 and in the direction of ω_2, which would appear to be a severe disadvantage in terms of detection. Fortunately, if the input polarizations are correctly chosen the signal can be isolated by its polarization. In an important geometry, the signal (ω_2) is isolated by transmission through a polarizer oriented at 45° to a linearly polarized pump (ω_1). The pump is overlapped in the sample with the probe (ω_2) linearly polarized at $-45°$. Thus the probe is blocked by the polarizer, but the signal is transmitted. This geometry may be viewed as pump-induced polarization of the isotropic medium to render it birefringent, thus inducing ellipticity in the transmitted probe, such that the signal leaks through the analyzing polarizer (hence the alternative name optical Kerr effect).

In this geometry, it is possible to greatly enhance signal-to-noise ratios by exploiting interference between the probe beam and the signal. Placing a quarterwave plate in the probe beam with its fast axis aligned with the probe polarization, and reorienting it slightly ($<1°$) yields a slightly elliptically polarized probe before the sample. A fraction of the probe beam, the local oscillator (LO), then also leaks through the analyzing polarizer, temporally and spatially overlapped with the signal. Thus, the signal and LO fields are seen by the detector, which measures the intensity as:

$$I(t) = \frac{nc}{8\pi}|\mathbf{E}_{\mathrm{LO}}(t) + \mathbf{E}_{\mathrm{S}}(t)|^2$$
$$= I_{\mathrm{LO}}(t) + I_{\mathrm{S}}(t) + \frac{nc}{4\pi}\mathrm{Re}[\mathbf{E}_{\mathrm{S}}^*(t)\cdot\mathbf{E}_{\mathrm{LO}}(t)] \qquad [4]$$

where the final term may be very much larger than the original signal, and is linear in $\boldsymbol{\chi}^{(3)}$. This term is usually isolated from the strong I_{LO} by lock-in detection. This method is called optical heterodyne detection (OHD), and generally leads to excellent signal to noise. It can be employed with other coherent signals by artificially adding the LO to the signal, provided great care is taken to ensure a fixed phase relationship between LO and signal. In the RIKES experiment, however, the phase relationship is automatic. The arrangement described yields an out-of-phase LO, and measures the real part of $\boldsymbol{\chi}^{(3)}$, the birefringence. Alternatively, the quarterwave plate is omitted, and the analyzing polarizer is slightly reoriented, to introduce an in-phase LO, which measures the imaginary part of $\boldsymbol{\chi}^{(3)}$, the dichroism. This is particularly useful for absorbing media. The OHD-RIKES method has been applied to measure the spectroscopy of the condensed phase, and has found particularly widespread application in transient studies (below).

Degenerate four wave mixing (DFWM) spectroscopy is a simple and widely used third-order spectroscopic method. As the name implies, only a single frequency is required. The laser beam is split into three, and recombined in the sample, in the geometry shown in Figure 2. The technique is also known as laser-induced grating scattering. The first two beams can be thought of as interfering in the sample to write a spatial grating, with fringe spacing dependent on the angle between them. The signal is then scattered from the third beam in the direction expected for diffraction from that grating. The DFWM experiment has been used to measure electronic spectra in hostile environments, by exploiting resonances with electronic transitions. It has also been popular in the determination of the numerical value of $\boldsymbol{\chi}^{(3)}$, partly because it is an economical technique, requiring only a single laser, but also because different polarization combinations make it possible to access different elements of $\boldsymbol{\chi}^{(3)}$. The technique has also been used in time resolved experiments, where the decay of the grating is monitored by diffraction intensity from a time delayed third pulse.

Two-photon or, more generally, multiphoton excitation has applications in both fundamental spectroscopy and analytical chemistry. Two relevant level

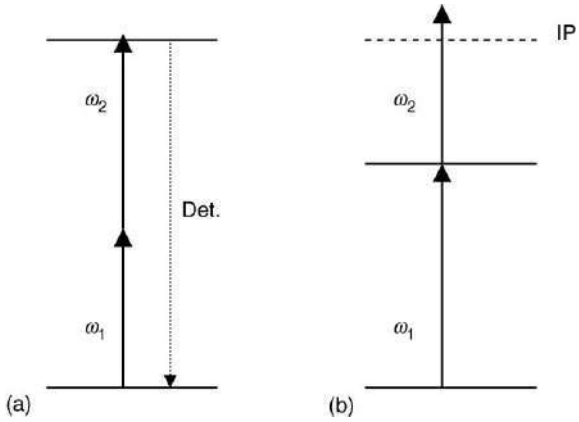

Figure 4 Two cases of resonant two photon absorption. In (a) the excited state is two-photon resonant, and the process is detected by the emission of a photon. In (b) the intermediate state is resonant, and the final energy is above the ionization potential (IP) so that photocurrent or mass detection can be used.

schemes are shown in Figure 4. Some property associated with the final state permits detection of the multiphoton absorption, for example, fluorescence in (a) and photocurrent in (b).

Excitation of two-photon transitions, as in Figure 4a, is useful in spectroscopy because the selection rules are different to those for the corresponding one-photon transition. For example the change in angular momentum quantum number, ΔL, in a two-photon transition is 0, ± 2, so, for example, an atomic S to D transition can be observed. High spectroscopic resolution may be attained using Doppler free two-photon absorption spectroscopy. In this method, the excitation beams are arranged to be counter-propagating, so that the Doppler broadening is cancelled out in transitions where the two excitation photons arise from beams with opposing wavevectors. In this case, the spectroscopic linewidth is governed only by the homogeneous dephasing time.

The level scheme in Figure 4b is also widely used in spectroscopy, but in this case the spectrum of the intermediate state is obtained by monitoring the photocurrent as a function of ω_1. The general technique is known as resonance enhanced multiphoton ionization (REMPI) and yields high-quality spectra of intermediate states which are not detectable by standard methods, such as fluorescence. The sensitivity of the method is high, and it is the basis of a number of analytical applications, often in combination with mass spectrometry.

Ultrafast Time Resolved Spectroscopy

The frequency and linewidth of a Raman transition may be extracted from the CARS measurement, typically by combining two narrow bandwidth pulsed lasers, and tuning one through the resonance while measuring the nonlinear signal intensity. The time resolved analogue requires two pulses, typically of a few picoseconds duration (and therefore a few wavenumbers bandwidth) at ω_1 and ω_2 to be incident on the sample. This pair coherently excites the Raman mode. A third pulse at ω_1 is incident on the sample a time t later, and stimulates the CARS signal at $2\omega_1 - \omega_2$ in the phase-matched direction. The decay rate of the vibrational coherence is measured from the CARS intensity as a function of the delay time. Thus, the frequency of the mode is measured in the frequency domain, but the linewidth is measured in the time domain. If very short pulses are used (such that the pulsewidth is shorter than the inverse frequency of the Raman active mode) the Raman transition is said to be impulsively excited, and the CARS signal scattered by the time delayed pulse reveals an oscillatory response at the frequency of the Raman active mode, superimposed on its decay. Thus, in this case, spectroscopic information is measured exclusively in the time domain. In the case of nonlinear Raman spectroscopy, similar information is available from the frequency and the time domain measurements, and the choice between them is essentially one of experimental convenience. For example, time domain CARS, RIKES, and DFWM spectroscopy have turned out to be particularly powerful routes to extracting low-frequency vibrational and orientational modes of liquids and solutions, thus providing detailed insights into molecular interactions and reaction dynamics in the condensed phase.

Other time domain experiments contain information that is not accessible in the frequency domain. This is particularly true of photon echo methods. The name suggests a close analogy with nuclear magnetic resonance (NMR) spectroscopy, and the (optical) Bloch vector approach may be used to describe both measurements, although the transition frequencies and time-scales involved differ by many orders of magnitude. In the photon echo experiment, two or three ultrafast pulses with carefully controlled interpulse delay times are resonant with an electronic transition of the solute. In the two-pulse echo, the echo signal is emitted in the phase match direction at twice the interpulse delay, and its intensity as a function of time yields the homogeneous dephasing time associated with the transition. In the three-pulse experiment the pulses are separated by two time delays. By measuring the intensity of the stimulated echo as a function of both delay times it is possible to separately determine the dephasing time and the population relaxation time associated with the resonant transition. Such information is not accessible from linear spectroscopy, and can be extracted

only with difficulty in the frequency domain. The understanding of photon echo spectroscopy has expanded well beyond the simple description given here, and it now provides unprecedented insights into optical dynamics in solution, and thus informs greatly our understanding of chemistry in the condensed phase. The methods have recently been extended to the infra red, to study vibrational transitions.

Higher-Order and Multidimensional Spectroscopies

The characteristic feature of this family of spectroscopies is the excitation of multiple resonances, which may or may not require measurements at $\chi^{(n)}$ with $n > 3$. Such experiments require multiple frequencies, and may yield weak signals, so they only became experimentally viable upon the availability of stable and reliable solid state lasers and optical parametric generators. Measurements are made in either the time or the frequency domain, but in either case benefit from heterodyne detection.

One of the earliest examples was two-dimensional Raman spectroscopy, where multiple Raman active modes are successively excited by temporally delayed pulse pairs, to yield a fifth-order nonlinear signal. The signal intensity measured as a function of both delay times (corresponding to the two dimensions) allows separation of homogeneous and inhomogeneous contributions to the line shape. This prodigiously difficult $\chi^{(5)}$ experiment has been completed in a few cases, but is plagued by interference from third-order signals.

More widely applicable are multidimensional spectroscopies using infrared pulses or combinations of them with visible pulses. The level scheme for one such experiment is shown in Figure 5 (which is one of many possibilities). From the scheme, one can see that the nonlinear signal in the visible depends on two resonances, so both can be detected. This can be regarded as a multiply resonant nondegenerate four-wave mixing (FWM) experiment. In addition, if the two resonant transitions are coupled, optical excitation of one affects the other. Thus, by measuring the signal as a function of both frequencies, the couplings between transitions are observed. These appear as cross peaks when the intensity is plotted in the two frequency dimensions, very much as with 2D NMR. This technique is already providing novel information on molecular structure and structural dynamics in liquids, solutions, and proteins.

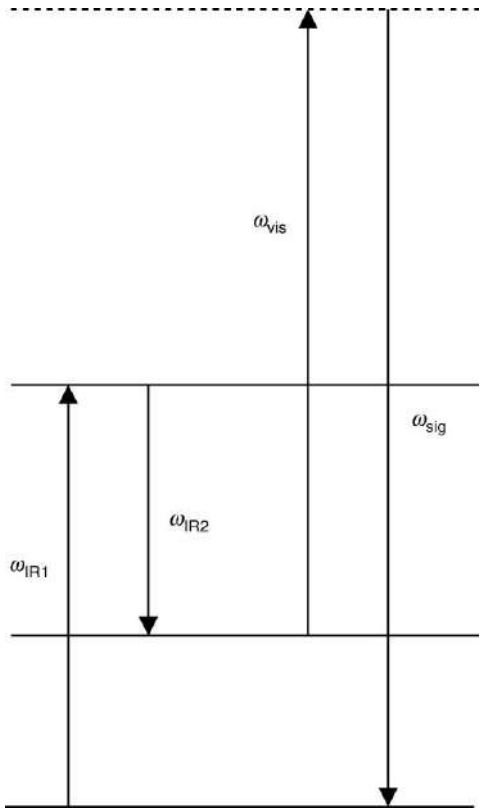

Figure 5 Illustration of multiple resonance enhancements in a FWM geometry, from which 2D spectra may be generated.

Spatially Resolved Spectroscopy

A recent innovation is nonlinear optical microscopy. The nonlinear dependence of signal strength on intensity means that nonlinear processes are localized at the focal point of a lens. When focusing is strong, such as in a microscope objective, spatial localization of the nonlinear signal can be dramatic. This is the basis of the two-photon fluorescence microscopy method, where a high repetition rate source of low-energy ultrafast pulses is focused by a microscope objective into a sample labeled with a fluorescent molecule, which has absorption at half the wavelength of the incident photons (Figure 4). The fluorescence is necessarily localized at the focal point because of its dependence on the square of the incident intensity. By measuring intensity while scanning the position of the focal point in space, a 3D image of the distribution of the fluorophore is constructed. This technique turns out to have a number of advantages over one photon fluorescence microscopy, most notably in terms of ease of implementation, minimization of sample damage, and depth resolution. The technique is widely employed in cell biology.

Stimulated by the success of two-photon microscopy, further nonlinear microscopies have been developed, all relying on the spatial localization of the signal. CARS microscopy has been

demonstrated to yield 3D images of the distribution of vibrations in living cells. It would be difficult to recover such data by linear optical microscopy. SHG has been applied in microscopy. In this case, by virtue of the symmetry selection rule referred to above, a 3D image of orientational order is recovered. Both these and other nonlinear signals provide important new information on complex heterogeneous samples, most especially living cells.

See also

Spectroscopy: Nonlinear Laser Spectroscopy; Raman Spectroscopy.

Further Reading

Andrews DL and Allcock P (2002) *Optical Harmonics in Molecular Systems*. Weinheim, Germany: Wiley-VCH.

Andrews DL and Demidov AA (eds) (2002) *An Introduction to Laser Spectroscopy*. New York: Kluwer Academic.

Butcher PN and Cotter D (1990) *The Elements of Nonlinear Optics*. Cambridge, UK: Cambridge University Press.

Cheng JX and Xie XS (2004) Coherent anti-Stokes Raman scattering microscopy: Instrumentation, theory, and applications. *Journal of Physical Chemistry B* 108: 827.

de Boeij WP, Pshenichnikov MS and Wiersma DA (1998) Ultrafast solvation dynamics explored by femtosecond photon echo spectroscopies. *Annual Review of Physical Chemistry* 49: 99.

Eisenthal KB (1992) Equilibrium and dynamic processes at interfaces by 2nd harmonic and sum frequency generation. *Annual Review of Physical Chemistry* 43: 627.

Fleming GR (1986) *Chemical Applications of Ultrafast Spectroscopy*. Oxford, UK: Oxford University Press.

Fleming GR and Cho M (1996) Chromophore-solvent dynamics. *Annual Review of Physical Chemistry* 47: 109.

Fourkas JT (2002) Higher-order optical correlation spectroscopy in liquids. *Annual Review of Physical Chemistry* 53: 17.

Hall G and Whitaker BJ (1994) Laser-induced grating spectroscopy. *Journal of Chemistry Society Faraday Transactions* 90: 1.

Heinz TF (1991) Second order nonlinear optical effects at surfaces and interfaces. In: Ponath H-E and Stegeman GI (eds) *Nonlinear Surface Electromagnetic Phenomena*, pp. 353.

Hesselink WH and Wiersma DA (1983) Theory and experimental aspects of photon echoes in molecular solids. In: Hochstrasser RM and Agranovich VM (eds) *Spectroscopy and Excitation Dynamics of Condensed Molecular Systems*, chap. 6. Amsterdam: North Holland.

Levenson MD and Kano S (1988) *Introduction to Nonlinear Laser Spectroscopy*. San Diego, CA: Academic Press.

Meech SR (1993) Kinetic application of surface nonlinear optical signals. In: Lin SH, Fujimura Y, and Villaeys A (eds) *Advances in Multiphoton Processes and Spectroscopy*, vol. 8, pp. 281. Singapore: World Scientific.

Mukamel S (1995) *Principles of Nonlinear Optical Spectroscopy*. Oxford: Oxford University Press.

Rector KD and Fayer MD (1998) Vibrational echoes: a new approach to condensed-matter vibrational spectroscopy. *International Review of Physical Chemistry* 17: 261.

Richmond GL (2001) Structure and bonding of molecules at aqueous surfaces. *Annual Review of Physical Chemistry* 52: 357.

Shen YR (1984) *The Principles of Nonlinear Optics*. New York: Wiley.

Smith NA and Meech SR (2002) Optically heterodyne detected optical Kerr effect – Applications in condensed phase dynamics. *International Review of Physical Chemistry* 21: 75.

Tolles WM, Nibler JW, McDonald JR and Harvey AB (1977) Review of theory and application of coherent anti-Stokes Raman spectroscopy (CARS). *Applied Spectroscopy* 31: 253.

Wright JC (2002) Coherent multidimensional vibrational spectroscopy. *International Review of Physical Chemistry* 21: 185.

Zipfel WR, Williams RM and Webb WW (2003) Nonlinear magic multiphoton microscopy in the biosciences. *Nature Biotechnology* 21: 1368.

Zyss J (ed.) (1994) *Molecular Nonlinear Optics*. San Diego: Academic Press.

Photodynamic Therapy of Cancer

A J MacRobert and T Theodossiou, University College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

One of the most active areas of photomedical research, in recent years, has been the exploration of the use of light-activated drugs known as photosensitizers. These compounds may be activated using light, usually provided by a laser via an optical fiber which is placed at the site of the target lesion. This treatment is known as photodynamic therapy (PDT) and is being applied to the local destruction of malignant tumors and certain nonmalignancies. Activation of the photosensitizer results in the

generation of reactive oxidizing intermediates which are toxic to cells, and this process ultimately leads to tumor destruction. PDT is a relatively low-power, nonthermal, photochemical technique that uses fluence rates not exceeding 200 mW/cm^2 and total light doses or fluences of typically 100 J/cm^2. Generally red or near-infrared light is used, since tissue is relatively transparent at these wavelengths. It is a promising alternative approach to the local destruction of tumors for several reasons. First, it is a minimally invasive treatment since laser light can be delivered with great accuracy to almost any site in the body via thin flexible optical fibers with minimal damage to overlying normal tissues. Second, the nature of PDT damage to tissues is such that healing is safer and more complete than after most other forms of local tissue destruction (e.g., radiotherapy). PDT is also capable of some degree of selectivity for tumors when the sensitizer levels, light doses, and irradiation geometry are carefully controlled. This selectivity is based on the higher sensitizer retention in tumors after administration relative to the adjacent normal tissues in which the tumor arose (generally 3:1 for extracranial tumors but up to 50:1 for brain tumors).

The photosensitizer is administered intravenously to the patient and time allowed (3–96 hours depending on the sensitizer) for it to equilibrate in the body before the light treatment (Figure 1). This time is called the drug–light interval. PDT may also be useful for treating certain nonmalignant conditions, in particular, psoriasis and dysfunctional menorrhagia (a disorder of the uterus), and local treatment of infections; such as genital papillomas, infections in the mouth and the upper gastrointestinal tract. In certain cases, the photosensitizer may be applied directly to the lesions, particularly for treatment of skin tumors, as discussed later. The main side-effect of PDT is skin photosensitivity, owing to retention of the drug in the skin, so patients must avoid exposure to sunlight in particular for a short period following treatment. Retreatment is then possible once the photosensitizer has cleared from the skin since these drugs have little intrinsic toxicity, unlike many conventional chemotherapy agents.

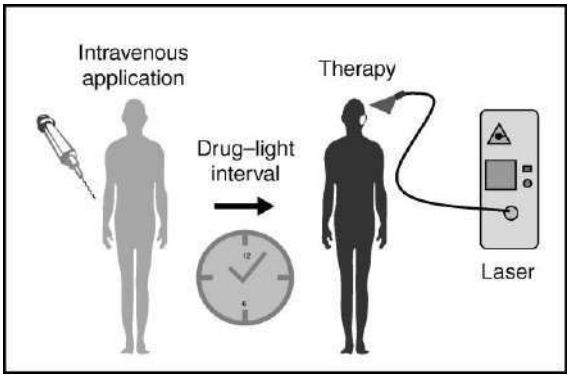

Figure 1 Photodynamic therapy: from sensitization to treatment.

Photoproperties of Photosensitizers

By definition, there are three fundamental requirements for obtaining a photodynamic effect: (a) light of the appropriate wavelength matched to the photosensitizer absorption; (b) a photosensitizer; and (c) molecular oxygen. The ideal photochemical and biological properties of a photosensitizer may be easily summarized, although the assessment of a sensitizer in these terms is not as straightforward as might be supposed because the heterogeneous nature of biological systems can sometimes profoundly affect these properties. Ideally though, a sensitizer should possess the following attributes: (a) red or near infrared light absorption; (b) nontoxic, and with low skin photosensitizing potency; (c) selective retention in tumors relative to normal adjacent tissue; (d) an efficient generator of cytotoxic species, usually singlet oxygen; (e) fluorescence, for visualization; (f) a defined chemical composition, and (g) preferably water soluble. A list of several photosensitizers possessing the majority of the above-mentioned attributes is given in Table 1.

The reasons for these requirements are partly self-evident, but worth amplifying. Strong absorption is desirable in the red and near-infrared spectral region, where tissue transmittance is optimum enabling penetration of the therapeutic light within the tumor (Figure 2). To minimize skin photosensitization by solar radiation, the sensitizer absorption spectrum should ideally consist of a narrow red wavelength band, with little absorption at other wavelengths down to 400 nm, below which solar irradiation falls off steeply. Another advantage of red wavelength irradiation is that the potential mutagenic effects encountered with UV-excited sensitisers (e.g., psoralens) are avoided. Since the object of the treatment is the selective destruction of tumor tissue, leaving surrounding normal tissue undamaged, some degree of selective retention of the dye in tumor tissue is desirable. Unfortunately, the significance of this aspect has been exaggerated in the literature and in many cases treatment selectivity owes more to careful light irradiation geometry. Nevertheless, many normal tissues have the capacity to heal safely following PDT damage.

The key photochemical property of photosensitizers is to mediate production of some active

Table 1 Properties of several photosensitizers

Compound	*λ/nm (ε/M⁻¹ cm⁻¹)*	*Drug dose/mg kg⁻¹*	*Light dose/J cm⁻²*	*Diseases treated*
Hematoporphyrin (HpD–Photofrin–Photosan)	628 (3.0×10^3)	1.5–5	75–250	Early stage esophagus, bladder, lung, cervix, stomach, and mouth cancers. Palliative in later stages
ALA (converted to protoporphyrin IX)	635 (5×10^3)	60	50–150	Skin, stomach, colon, bladder, mouth cancers. Esophageal dysplasia. Various nonmalignant conditions
Benzoporphyrin derivative (BpD)	690 (3.5×10^4)	4	150	Age-related macular degeneration (AMD)
Tin etiopurpurin (SnET2–Purlytin)	665 (3.0×10^4)	1.2	150–200	Breast and skin cancers, AMD
Monoaspartyl chlorin e_6 (MACE)	660 (4.0×10^4)	1.0	25–200	Skin cancers
Lutetium texaphyrin (Lu–Tex)	732 (4.2×10^4)	1.0	150	Metastatic brain tumors, breast cancers, atherosclerotic plaques
Aluminum disulfonated phthalocyanine	675 (2.0×10^5)	1.0	50–200	Brain, colon, bladder, and pancreatic cancers. Head and neck cancers in animal studies only
Metatetrahydroxychlorin (mTHPC–temoporphin–Foscan)	652 (3.5×10^4)	0.15	5–20	Head, neck, prostate, pancreas, lung, brain, biliary tract, and mouth cancers. Superior to HpD and ALA in mouth cancers
Palladium pheophorbide (Tookad)	763 (8.6×10^4)	2.0	–	Prostate cancer
Hypericin	590 (3×10^4)	–	–	Psoriasis

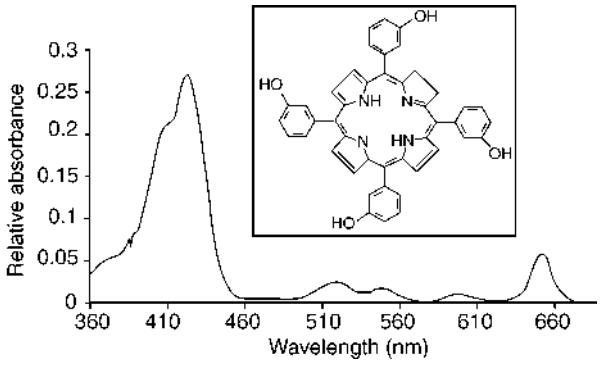

Figure 2 The absorption spectrum of m-THPC. The structure of m-THPC is shown in the inset.

molecule which is cytotoxic, that is, will destroy cells. The first electronically excited state of molecular oxygen, so-called singlet oxygen, fulfills this role very well and may be produced via the interaction of an excited electronic state of the sensitizer with oxygen present in the tissue. Thus, in summary, to achieve effective photosensitization, a sensitizer should exhibit appreciable absorption at red to near-infrared wavelengths and generate cytotoxic species via oxygen-dependent photochemical reactions. The first clinical photosensitizer was hematoporphyrin derivative (HpD), which is derived synthetically from hematoporphyrin by reaction with acetic and sulfuric acids to give a complex mixture of porphyrins. A purified fraction of these (Photofrin) is available commercially, and this has been used most widely in clinical applications to date. Second-generation (photosensitizers are now becoming available, including phthalocyanine and chlorin compounds as shown in Table 1.

A new approach to PDT has recently emerged involving the administration of a natural porphyrin precursor, 5-aminolaevulinic acid (ALA), which is metabolized within cells to produce protoporphyrin IX (see Table 1). This porphyrin is known to be a powerful photosensitizing agent but suffers from the drawback of being a poor tumor localizer when used directly. In contrast, administration of ALA induces protoporphyrin biosynthesis, particularly in rapidly proliferating cells, which may then be destroyed using irradiation at 630 nm. Therefore, this new therapy may offer enhanced treatment selectivity with little risk of skin photosensitivity owing to the rapid clearance after 24 h of the protoporphyrin. Investigation of this new approach has already proved successful in clinical treatment of

skin tumors using topical application of ALA in a thick emulsion.

Considerable effort is also being expended on exploiting the retention of sensitizers in tumors for diagnostic purposes, although the results are rather mixed to date. The prospects with ALA-induced protoporphyrin IX are, however, more promising, as illustrated in Figure 3 where the fluorescence is selectively confined to the skin tumor (basal cell carcinoma).

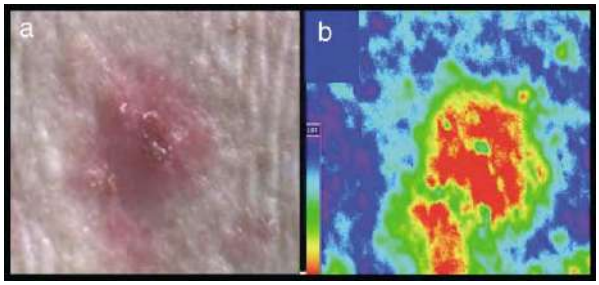

Figure 3 (a) Image of a basal cell carcinoma; (b) fluorescence imaging of the same lesion after ALA sensitization and 405 nm light excitation.

Mechanisms of Photodynamic Therapy

The main principles of the photosensitization mechanism are now well established with the initial step being excitation of the sensitizer from its electronic ground state to the short-lived fluorescent singlet state. The lifetime of the singlet state is generally only a few nanoseconds and the main role of this state in the photosensitization mechanism is to act as a precursor of the metastable triplet state through intersystem crossing. Efficient formation of this metastable state is required because it is the interaction of the triplet state with tissue components that generates cytotoxic species such as singlet oxygen. Thus the triplet state quantum yield (i.e., probability of triplet state formation per photon absorbed) of photosensitizers should ideally approach unity. Interaction of the metastable triplet state (which in de-aerated solutions has a lifetime extending to the millisecond range) with tissue components may proceed via either a type I or II mechanism or a combination (see Figure 4). A type I process can involve hydrogen abstraction from the sensitizer to

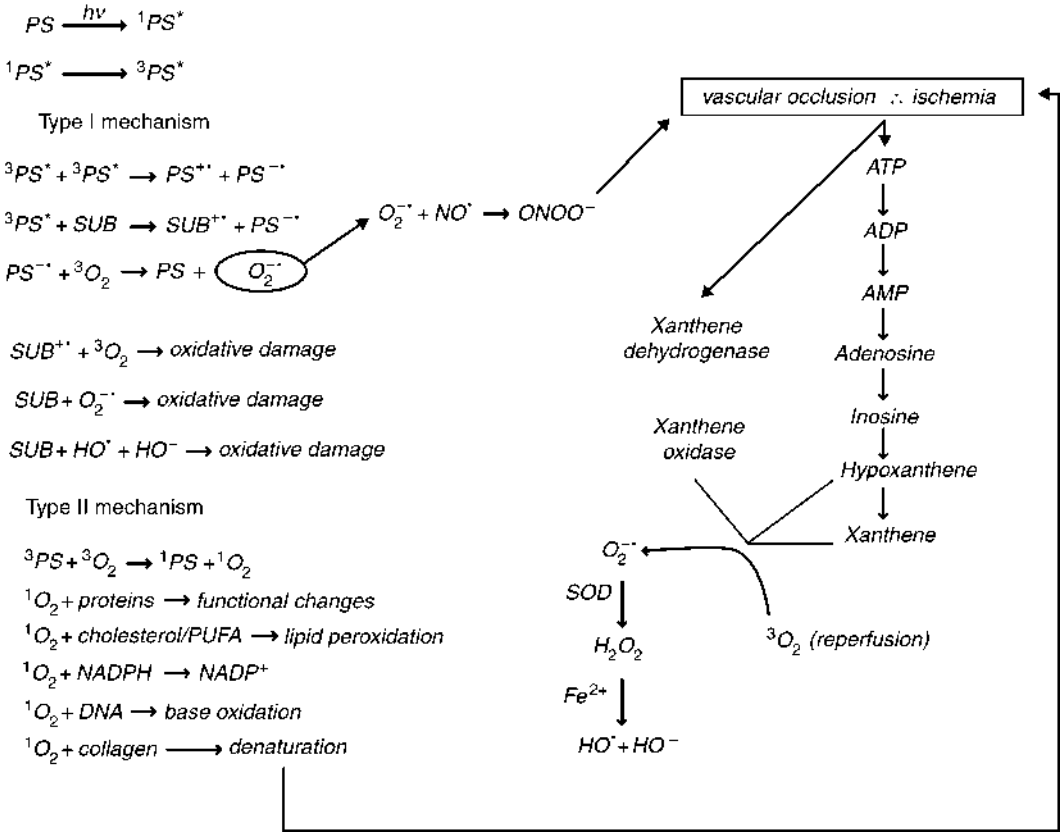

Figure 4 The photophysical and photochemical mechanisms involved in PDT. PS denotes the photosensitizer, SUB denotes the substrate, either biological, a solvent or another photosensitizer; * denotes the excited state and • denotes a radical.

produce free radicals or electron transfer, resulting in ion formation. The type II mechanism, in contrast, exclusively involves interaction between molecular oxygen and the triplet state to form singlet oxygen which is highly reactive in biological systems. The near-resonant energy transfer from the triplet state to O_2 can be highly efficient and the singlet oxygen yield can approach the triplet state yield provided that the triplet state energy exceeds 94 kJ/mol, the singlet oxygen excitation energy. It is widely accepted that the type II mechanism underlies the oxygen-dependent phototoxicity of sensitizers used for photodynamic therapy. Both proteins and lipids (the main constituents of membranes) are susceptible to photooxidative damage induced via a type II mechanism which generally results in the formation of unstable peroxide species. For example, unsaturated lipids may be oxidized by the 'ene' reaction where the singlet oxygen reacts with a double bond. Other targets include such important biomolecules as cholesterol, certain amino acid residues, collagen, and the coenzyme, NADPH. Another synergistic mode of action involves the occlusion of the microcirculation to tumors and reperfusion injury. Reperfusion injury relates to the formation of xanthene oxidase from xanthene dehydrogenase in anoxic conditions which reacts with xanthene or hypoxanthene products of ATP dephosphorylation to convert the restored oxygen into superoxide anion which directly or via the Fenton reaction causes cellular damage.

An important feature of the type II mechanism, which is sometimes overlooked, is that when the sensitizer transfers electronic energy to O_2 it returns to its ground state. Thus the cytotoxic singlet oxygen species is generated without chemical transformation of the sensitizer which may then absorb another photon and repeat the cycle. Effectively, a singlet photosensitizer molecule is capable of generating many times its own concentration of singlet oxygen, which is clearly a very efficient means of photosensitization provided the oxygen supply is adequate.

Lasers in PDT

Lasers are the most popular light source in PDT since they have several key characteristics that differentiate them from conventional light sources, namely coherence, monochromaticity, and collimated output. The two main practical features that make them so useful in PDT are their monochromaticity and their combination with fiber-optic delivery. The monochromaticity is important since the laser can be tuned to a specific absorption peak of a photosensitizer, thus ensuring that all the energy delivered is utilized for the excitation and photodynamic activation of the photosensitizer. This is not true for a conventional light source (as for example a tungsten lamp) where the output power is divided over several hundred nanometers throughout the UV, visible, and IR regions and only a fraction of the power lies within the absorption band of the photosensitizer. To numerically illustrate this, let us simplify things by representing the lamp output as a square profile and also consider a photosensitizer absorption band about 30 nm full width half maximum. If a laser with a frequency span of about 2 nm and a lamp with the same power but with a frequency span of about 600 nm are used, then the useful portion of it is about 0.05 and if we consider a Gaussian profile for the sensitizer absorption band, the above fraction drops even lower, perhaps even to 0.01. That means that the lamp would achieve 100 times less the excitation rate of the laser, or in other words, we would require a lamp with an output power of 100 times that of the laser to achieve the same rate of excitation and consequently the same time of treatment, provided we use a low enough laser power to remain within the linear regime of the photosensitizer excitation. The other major disadvantage of a lamp source is the lack of collimation which results in low efficiency for fiber-optic delivery.

We now review the different lasers that have found application in PDT. Laser technology has significantly advanced in recent years and there are now a range of options in PDT laser sources for closely matching the absorption profiles of the various photosensitizers. Moreover, since PDT is becoming more widely used clinically, practical considerations, such as portability and turn-key operation are increasingly important. In Figure 5 the optical spectrum is shown from about 380 nm (violet–blue) to 800 nm (near-infrared, NIR). We have superimposed on this spectrum, the light penetration depth of tissue to roughly illustrate how deeply light can penetrate (and consequently activate photosensitizer molecules) into lower-lying tissue layers. The term 'penetration depth' is a measure of the light attenuation by the tissue so the figure of 2 mm corresponds to $1/e$ attenuation of the incident intensity at a tissue depth of 2 mm. It is obvious that light at the blue end of the spectrum has only a superficial effect whereas the more we shift further into the red and NIR regions the penetration depth increases. This is due to two main reasons: absorption and scattering of light by various tissue component molecules. For example, in the blue/green region for the absorption of melanin and hemoglobin is relatively high. But there is a noticeable increase in penetration depth beyond about 600 nm leading to an optimal wavelength region, or

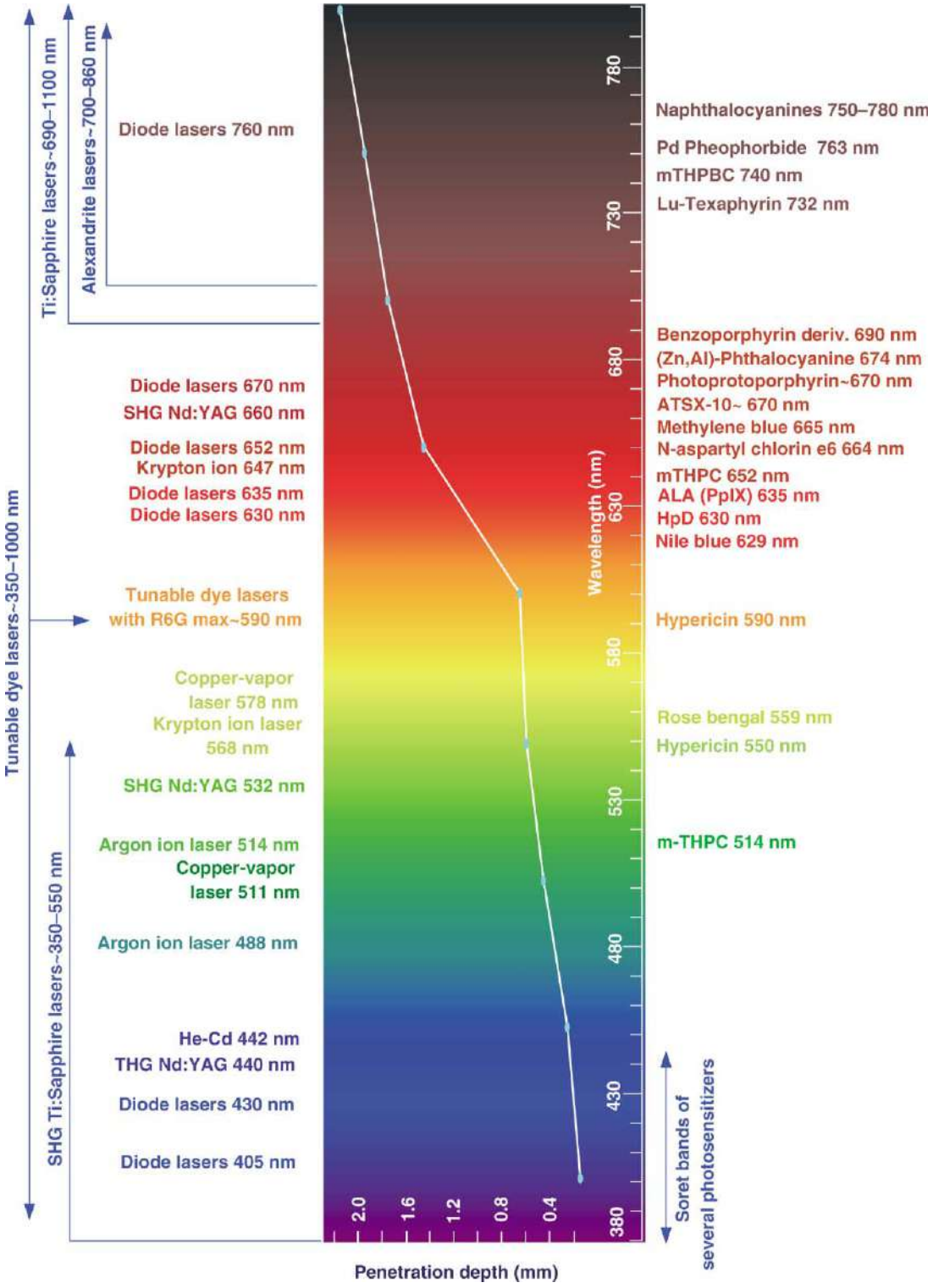

Figure 5 Photosensitizers and laser light sources available in the visible and near-infrared spectral regions.

'therapeutic window' for laser therapy of around 600–1100 nm (Figure 5).

The fundamental wavelength of the Nd:YAG laser lies within this therapeutic window at 1064 nm and this laser is now widely used in thermal laser therapy. The Nd:YAG lasers can be operated either in cw (output power ~ 200 W multimode), long-pulse (~500 W average power at 50 Hz), or Q-switched

(50 MW peak power at around 10 ns pulse duration) modes. Nd:YAG is a solid-state laser with a yttrium aluminum garnet crystal doped with about 1% of trivalent Nd ions as the active medium of the laser; and with another transition of the Nd ion, this laser (with the choice of different optics) can also operate at 1320 nm. Although dyes are available which absorb from 800–1100 nm, the generation of singlet oxygen via the type II mechanism is energetically unfavorable because the triplet state energies of these dyes are too low. However, the fundamental frequency of the laser can be doubled (second-harmonic generation, SHG) or tripled (third-harmonic generation, THG) with the use of nonlinear crystals, to upconvert the output radiation to the visible range, thus rendering this laser suitable for PDT: i.e., frequency doubling to 532 nm and tripling to 355 nm. Note that the 532 nm output is suitable for activation of the absorption band of hypericin with a maximum at 550 nm even though it is not tuned to the maximum of this absorption.

In the early days of PDT and other laser therapies, ion lasers were widely used. The argon ion laser uses ionized argon plasma as gain medium, and produces two main wavelengths at 488 nm and 514 nm. The second of the two lies exactly at the maximum of the 514 nm m-THPC absorption. Argon ion lasers are operated in cw mode and usually have output powers in the region of 5–10 W at 514 nm (the most powerful argon ion line). Krypton ion lasers, which emit at 568 or 647 nm, are similar to their argon ion counterparts; however, they utilize ionized krypton plasma as gain medium. The 568 nm output can be used to activate the Rose Bengal peak at 559 nm, whereas the 647 nm most powerful line of krypton ion lasers, has been used for activation of m-THPC (652 nm).

Ideally, for optimal excitation, it is best to exactly tune the laser wavelength to the maxima of photosensitizer absorption bands. For this reason tunable organic dye lasers have been widely used for PDT. In these lasers the gain medium is an organic dye with a high quantum yield of fluorescence. Due to the broad nature of their fluorescence gain profile, tuning elements within the laser cavity can be used for selection of the lasing frequency within the gain band. Tunable dye lasers with the currently available dyes can quite easily cover the range from about 350–1000 nm. However, their operation is not of a 'turn key' nature since they require frequent replacement of their active material either for tuning to a different spectral region or for a better performance when the dye degrades. Tunable dye lasers also need to be pumped by some other light source like an argon ion laser, an excimer laser, a solid state laser (e.g., Nd:YAG 532 or 355 nm), a copper vapor laser, or lamps. Dye lasers have been used for the excitation of HpD at 630 nm, protoporphyrin IX at 635 nm, photoprotoporphyrin at 670 nm, and phthalocyanines around 675 nm. A further example is hypericin which, apart from the band at 550 nm, has a second absorption band at 590 nm. This is the optimum operation wavelength of dye lasers with rhodamine 590, better known as rhodamine 6G.

Dye lasers can operate in either pulsed or cw mode. However, there is a potential disadvantage in using a pulsed laser for PDT which becomes apparent when using lower repetition rates and higher pulse energies (>1 mJ per pulse) laser excitation. High pulse energies can induce saturation or transient bleaching of the sensitizer during the laser pulse and consequently much of the energy supplied is not efficiently absorbed by the sensitizer. It has been suggested that this effect accounts for the lack of photosensitized damage in tissue sensitized with a phthalocyanine and irradiated by a 5 Hz, 25 mJ-per-pulse flash lamp pumped dye laser. However, using a low pulse energy, high repetition rate copper vapor pumped dye laser (see below), the results were indistinguishable from cw irradiation with an Argon ion dye laser.

Another laser that has found clinical application in PDT is the copper vapor laser. In this laser the active medium is copper vapor at high temperature maintained in the tube by a repetitively pulsed discharge current. The copper vapor lasers are pulsed with typical pulse duration of about 50 ns and repetition rates reaching 20 kHz. Copper vapor lasers have been operated at quite high average powers up to about 40 W. The output radiation is produced at two distinct wavelengths, namely 511 and 578 nm both of which have found clinical use. Copper vapor pumped dye lasers have also been widely used for activating red-absorbing photosensitizers, and the analogous gold vapor lasers operating at 628 nm have been used to activate HpD.

A relatively new class of tunable lasers is the solid state Ti:Sapphire laser. The gain medium in this laser is a $\sim$1%Ti doped sapphire (Al_2O_3) crystal and its output may be tuned throughout the 690–1100 nm region, covering many photosensitizer absorption peaks. For example, most of the bacteriochlorin type photosensitizers have their absorption bands in that spectral region, e.g., m-THPBC with an absorption maximum at 740 nm. Also, Texaphyrin sensitizers absorb highly in this spectral region: lutetium texaphyrin has an absorption maximum at 732 nm.

Although tunable dye or solid-state lasers are widely used in laboratory studies for PDT, fixed-wavelength semiconductor diode lasers are gradually supplanting tunable lasers, owing to their practical convenience. It is now generally the case that when

a photosensitizer enters clinical trials or enters standard clinical use, laser manufacturers provide dedicated diode lasers with outputs matched to the chosen sensitizer. In this context there are diode lasers available at 630 nm for use with HpD, at 635 nm for use with ALA-induced protoporphyrin IX, 670 nm for ATSX-10 or phthalocyanines, or even in the region of 760 nm, for use with bacteriochlorins. The advantage of these diode lasers is that they are tailormade for use with a particular photosensitizer, they are highly portable, and they have a relatively easy turn-key operation.

So far we have concentrated on red wavelengths but most of the porphyrin and chlorin family of photosensitizers exhibit quite strong absorption bands in the blue region, known as 'Soret' bands. Despite the fact that tissue penetration in this spectral region is minimal, these bands are far more intense than the corresponding red absorption bands of the sensitizer. In this respect it is possible to activate these sensitizers at blue wavelengths, especially in the case of treatment of superficial (e.g., skin) malignant lesions. There are now blue diode lasers (and light-emitting diodes) for the activation of these bands, in particular at 405 nm where the Soret band of protoporphyrin IX lies or at 430 nm for selective excitation of the photoprotoporphyrin species. For very superficial lesions (e.g., in the retina) it may even be possible to use multiphoton excitation provided by femtosecond diode lasers operating at about 800 nm.

Clinical Applications of Lasers in PDT

In the previous section we reviewed the various types of lasers being used for PDT but their combination with fiber-optic delivery is also of key importance for their clinical application. The laser light may be delivered to the target lesion either externally using surface irradiation, or internally within the lesion which is denoted as interstitial irradiation, as depicted in Figure 6. For example, in case of superficial cutaneous lesions the laser irradiation is applied externally (Figure 6a). The area of the lesion is marked prior to treatment to determine the diameter of the laser spot. For a given fluence rate (100–200 W/cm^2) and the area of the spot, the required laser output power is then calculated. A multimode optical fiber, terminated with a microlens to ensure uniform irradiation, is then positioned at a distance from the lesion yielding the desired beam waste, and the surrounding normal tissue is shielded with dark material. However, if the lesion is a deeper-lying solid tumor, interstitial PDT is employed

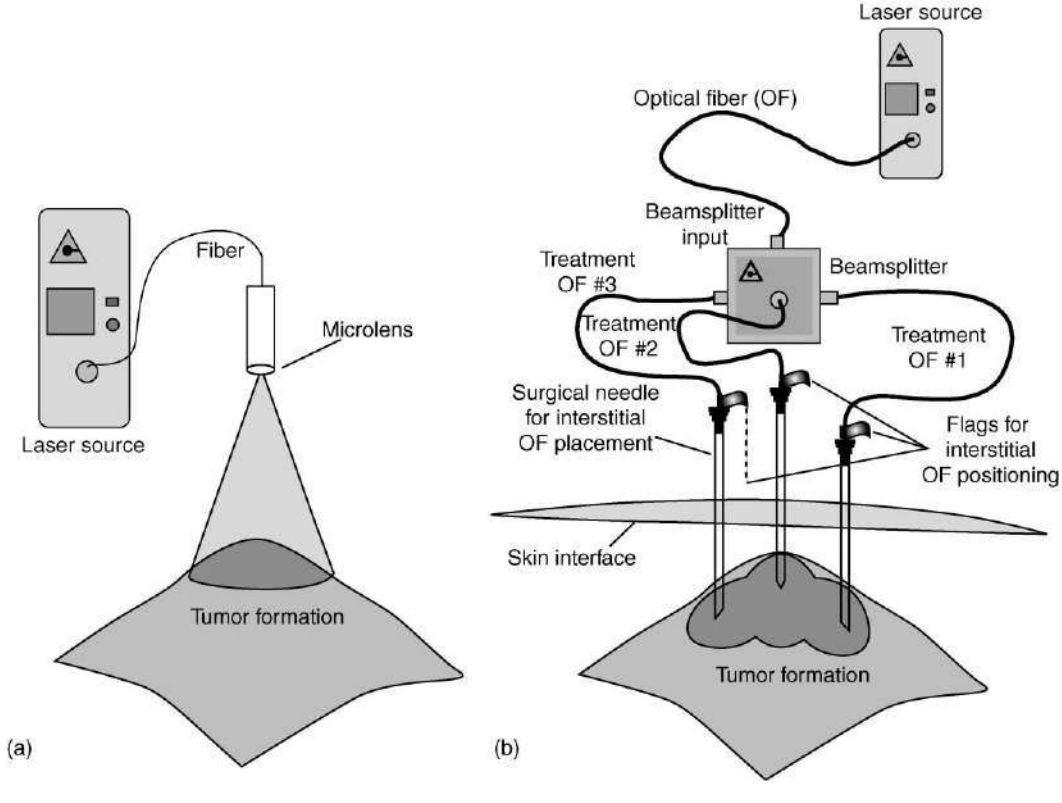

Figure 6 Clinical application of PDT. (a) Surface treatment. (b) Interstitial treatment.

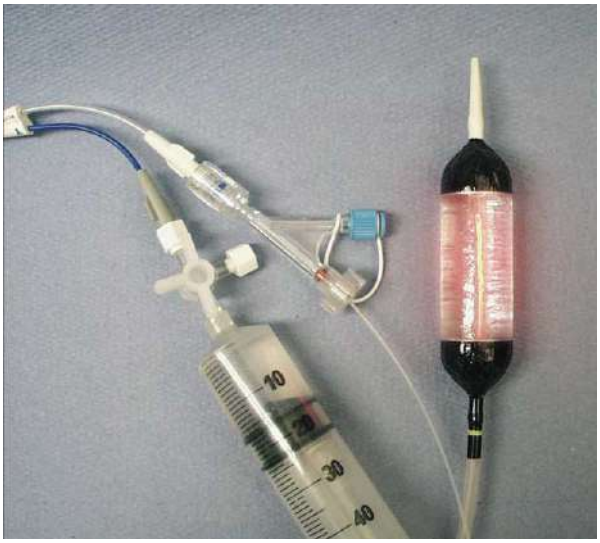

Figure 7 Balloon applicator and diffuser fiber used for PDT in Barrett's esophagus.

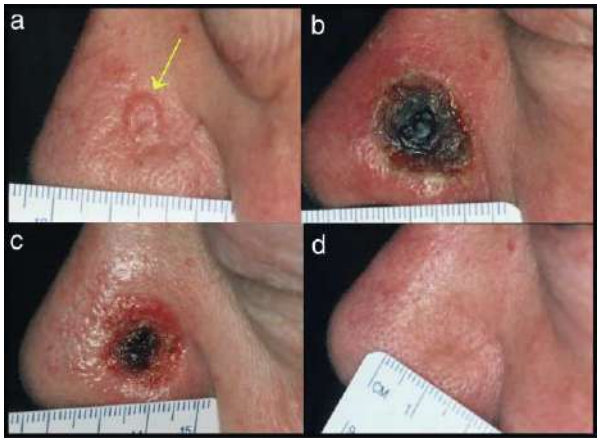

Figure 8 PDT treatment of a nasal basal cell carcinoma with m-THPC (Foscan®): (a) prior to PDT; (b) tissue necrosis a week after treatment; (c) four weeks after treatment; gradual recession of the necrosis and inflammation; (d) healing two months after treatment. Courtesy of Dr Alex Kuebler.

(Figure 6b). Surgical needles are inserted in an equidistant parallel pattern within the tumor, either freehand or under image (MRI/CT) guidance. Bare tip optical fibers are guided through the surgical needles to the lesion and flagged for position. A beamsplitter is used to divide the laser output into 2–4 components so that two to four optical fibers can be simultaneously employed. The laser power is adjusted so that the output fluence of each of the optical fibers is 100–200 mW. Even light distribution in the form of overlapping spheres approximately 1 cm diameter ensures the treatment of an equivalent volume of tissue around each fiber tip. Once each 'station' is likewise treated the fibers are repositioned accordingly, employing a pull-back technique in order for another station to be treated. Once the volume of the tumor is scanned in this manner the treatment is complete. Interstitial light delivery allows PDT to be used in the treatment of large, buried tumors and is particularly suitable for those in which surgery would involve extensive resection.

The treatment of tumors of internal hollow organs is also possible with PDT. The most representative example of this is the treatment of precancerous lesions in the esophagus, known medically as 'Barrett's esophagus'. In this case, a balloon applicator is used to house a special fiber with a light-diffusing tip or 'diffuser' (Figure 7). These diffusers are variable in length and transmit light uniformly along the whole of their length in order to facilitate treatment of circumferential lesions. The balloon applicator is endoscopically inserted into the patient. The balloon is inflated to stretch the esophageal walls and the flagged diffuser fiber is placed in position, central to the balloon optical window. In this way the whole of the lesion may be treated circumferentially at the same time.

Finally, the clinical response to PDT treatment is shown in Figure 8. The treated area becomes necrotic and inflamed within two to four days following PDT. This necrotic tissue usually sloughs off or is mechanically removed. The area eventually heals with minimal scarring (Figure 8).

Future Prospects

A major factor in the future development of PDT will be the application of relatively cheap and portable semiconductor diode lasers. High-power systems consisting of visible diode arrays coupled into multimode fibers with an output of several Watts, have recently become available which will greatly simplify the technical difficulties which have held back PDT. The use of new photosensitizers, in the treatment of certain tumors by PDT, is making steady progress, and although the toxicology testing and clinical evaluation are lengthy and costly processes, the expectation is that these compounds will become more widely available for patient use during the next five years. Regarding the current clinical status of PDT, Photofrin has recently been approved for treatment of lung and esophageal cancers in Europe, USA, and Japan. Clinical trials are in progress for several of the second-generation photosensitizers which offer significant improvements over Photofrin in terms of chemical purity, photoproperties, and skin

clearance. With the increasing clinical application of compact visible diode lasers the prospects for photodynamic therapy are therefore encouraging.

See also

Lasers: Dye Lasers; Free Electron Lasers; Metal Vapor Lasers. **Ultrafast Laser Techniques:** Generation of Femtosecond Pulses.

Further Reading

Bonnet R (2000) *Chemical Aspects of Photodynamic Therapy*. Amsterdam: Gordon and Breach Science Publishers.

Dolmans DEJGJ, Fukumura D and Jain RK (2003) Photodynamic therapy for cancer. *Nature Reviews (Cancer)* 3: 380–387.

Hopper C (2000) Photodynamic therapy: a clinical reality in the treatment of cancer. *Lancet Oncology* 1: 212–219.

Milgrom L and MacRobert AJ (1998) Light years ahead. *Chemistry in Britain* 34: 45–50.

Rosen GM, Britigan BE, Halpern HJ and Pou S (1999) *Free Radicals Biology and Detection by Spin Trapping*. New York: Oxford University Press.

Svelto O (1989) *Principles of Lasers*, 3rd edn. New York: Plenum Press.

Vo-Dinh T (2003) *Biomedical Photonics Handbook*. Boca Raton, FL: CRC Press.

Pump and Probe Studies of Femtosecond Kinetics

G D Scholes, University of Toronto, Toronto, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Pump-probe spectroscopy is a ubiquitous time-resolved optical spectroscopy that has found application to the study of all manner of ultrafast chemical processes that involve excited states. The principle of the measurement is that a 'pump pulse' – usually an intense, short laser pulse – impulsively excites a sample, thus defining the start time for the ensuing photophysical dynamics. A probe pulse interrogates the system at later times in order to capture 'snap shots' of the state of the system. Many kinds of pump-probe experiments have been devised. The present article will mainly focus on transient absorption measurements.

Femtosecond pump-probe experiments have been employed to reveal the kinetics of excited state chemical processes such as solvation, isomerization reactions, proton transfer, electron transfer, energy transfer, or photochromism. It is becoming more common to use pump-probe techniques to study systems of increasing complexity, such as photosynthetic proteins, photoreceptor proteins, molecular aggregates, nanostructured materials, conjugated polymers, semiconductors, or assemblies for artificial light harvesting. However, the underlying chemical kinetics are not always easily revealed by pump-probe signals.

The principle of the pump-probe measurement is related to the seminal flash-photolysis experiment devised by Norrish and Porter. However, in order to achieve ultrafast time resolution – beyond that attainable by electronic detection – the probe pulse is temporally short and is controlled to arrive at the sample at variable time delays after the pump pulse has excited the sample. The change in intensity of the probe pulse, as it is transmitted through the sample, is monitored at each pump-probe time delay using a 'slow' detector that integrates over the probe pulse duration. One possible xprocess that can diminish the probe transmission is transient absorption. After formation of an excited state S_1 by the pump pulse, the probe pulse can monitor a resonance with a higher excited state $S_1 \rightarrow S_n$ absorption. Figure 1 shows transient absorption spectra, corresponding to excited state absorption at various time delays, that has been probed by a white light continuum probe and monitored by a multichannel CCD detector. The transient spectrum is seen to blue-shift with

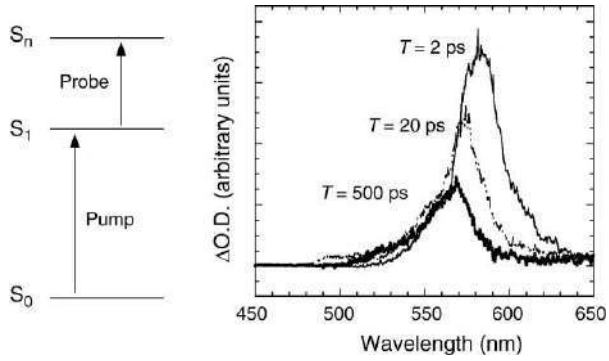

Figure 1 Excited state absorption spectra of 4′-*n*-pentyl-4-cyanoterphenyl in octanol solvent. The pump wavelength is 277 nm (100 nJ, 150 fs, 40 kHz) and the probe is a white light continuum (500 to 700 nm) generated in a sapphire crystal. The pump-probe delay *T* for each spectrum is indicated on the plot.

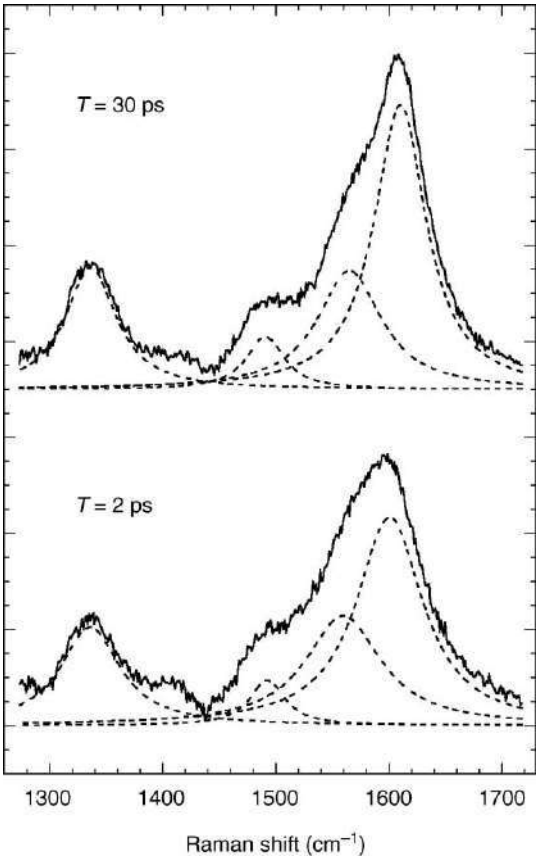

Figure 2 TR^3 spectra of 4′-*n*-pentyl-4-cyanoterphenyl in octanol solvent, recorded using the same set-up as summarized in Figure 1. The dashed lines are deconvolutions of the data to show the Raman bands.

pump-probe delay owing to solvation of the S_1 state of the molecule – the dynamic Stokes' shift. The dynamic Stokes' shift is the shift to lower energy of the emission spectrum as the solvent relaxes in response to the change in dipole moment of the excited electronic state compared to the ground state. The overall intensity of the induced absorption decays according to the lifetime of the excited state.

Although they will not be discussed in detail in the present article, various other pump-probe experiments are possible. For example, when in resonance with this transient absorption, the probe pulse can induce resonance Raman scattering, which reveals the evolution of the vibrational spectrum of the excited state. In this kind of pump-probe experiment we do not monitor the change in probe transmission through the sample. Instead, we monitor the difference between probe-induced resonance Raman scatter from the excited and ground states of the system. The resonance Raman (TR^3) scatter, time-resolved as a function of pump-probe delay, can be detected provided that it is not overwhelmed by fluorescence emission. Typical TR^3 data for different pump-probe delays are shown in Figure 2. The dashed lines show a deconvolution of the spectra to reveal the underlying vibrational bands. These bands are rather broad owing to the spectral width of the 1 ps laser pulses used in this experiment. It is clear that the time-resolution of a TR^3 experiment is limited to the picosecond regime, otherwise most of the vibrational band information is lost. TR^3 and complementary pump-probe infrared transient absorption experiments, utilizing an infrared probe beam, have been useful for observing the evolution of structural information, such as geometry changes, subsequent to photoexcitation.

In the present article we will describe the foundation of experiment and theory necessary to understand resonant ultrafast pump-probe spectroscopy applied to chemical processes in the condensed phase. By resonant, it is meant that the pump and probe pulses have frequencies (energies) resonant with electronic transitions of the molecules being studied. The implications of condensed phase are that the experiment interrogates an ensemble of molecules and the electronic transitions of these molecules are coupled to the random motions and environments of a bath, for example the solvent.

Experimental Measurement of Pump-Probe Data

The experimental setup for a one-color pump-probe experiment (i.e. pump and probe of the same color) is shown in Figure 3. The setup is easily adapted for a two-color experiment. Most of the laser intensity (70%) is transmitted through the first beam splitter so that the pump is more intense than the probe. A second beam splitter is used to split off a small amount of probe light for the reference beam. Note that the arrangement of these beam splitters results in both the pump and probe passing through the same amount of dispersive material on their way to the sample. The pump and probe beams each propagate through half-wave plates in order to control polarization. Usually the probe polarization is set to the magic angle (54.7°) relative to the pump in order to remove polarization bias. The beams are sent towards the sample by way of retroreflectors mounted on $x-y-z$ translation stages. The pump retroreflector is mounted on a precision computer-controlled translation stage such that it can be scanned in the x-direction to arrive at variable delays before and after the probe. The pump and probe beams are aligned using the $y-z$ controls on the retroreflector translation stages together with the pick-off mirrors

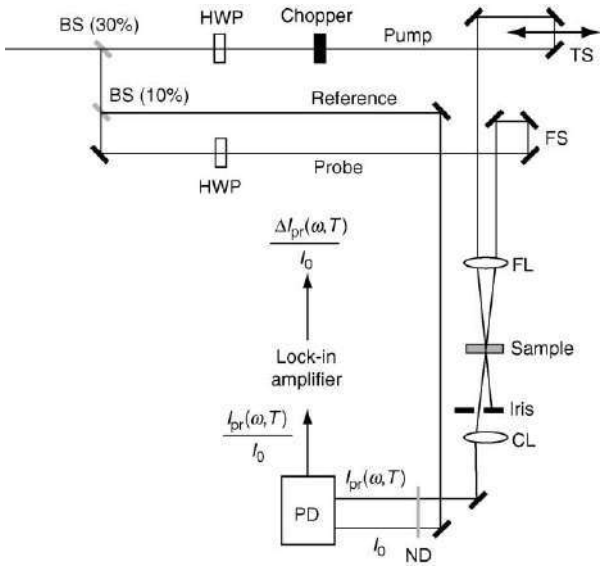

Figure 3 An experimental layout for femtosecond pump-probe spectroscopy. The reference beam ideally passes through the sample (not shown as such in aid of clarity). See text for a description. HWP, half-wave plate; BS, beamsplitter (% reflection); FL, focusing lens; CL, collimating lens; ND, neutral density filter; PD, photodetectors.

such that they travel in parallel towards the sample. The overall setup of the beams is in a box geometry so that pump and probe arms have the same path length.

The pump and probe beams are focused into the sample using a transmissive (i.e. lens) or reflective optic FL. The focal length is typically 20 to 30 cm, providing a small crossing angle to improve phase matching. The pump beam should have a larger spot size at its focal point in the sample than the probe beam, in order to avoid artifacts arising from the wandering of the pump and probe spatial overlap as a function of delay time – a potential problem, particularly for long delays. Good spatial overlap of the beams in the sample is important. It can be checked using a 50 μm pinhole. Fine adjustment of the pump-probe delay, to establish temporal overlap of the pump and probe pulses, can be dictated by autocorrelating the beams in a second harmonic generating crystal mounted at the sample position. At the same time the pulse compression is tweaked to ensure that the pulse dispersion in the experiment has been pre-compensated.

The sample is usually flowed or mounted in a spinning cell if it is liquid, in order to avoid thermal effects or photochemical bleaching. The path length of the sample is typically ≤ 1 mm. The transmitted probe beam is spatially isolated from the pump beam using an iris, is collimated, and then directed onto the photodetector.

The reference beam, which provides the relative reference intensity I_0 should preferably pass through the sample (away from the pump spot). This provides the probe-only transmission – that is, the reference beam intensity is attenuated according to sample ground state optical density. The photodetector outputs are used to ratio the pump-probe arm intensity transmitted through the sample with the reference intensity, $I_{pr}(\omega, T)/I_0$. Since the pump beam is being chopped at the lock-in reference frequency, the lock-in amplifier outputs the pump-induced fractional change in transmission intensity, $\Delta I_{pr}(\omega, T)/I_0$, usually simply written as $\Delta T/T$. When the pump-induced signal I_{pr} is small compared to the reference intensity I_0, then the detected $\Delta I_{pr}(\omega, T)/I_0$ signal is approximately equal to the change in optical density, ΔO.D.:

$$\Delta\text{O.D.} = \log\left(\frac{I_0 - \Delta I_{pr}}{I_0}\right) \approx \frac{-\Delta I_{pr}}{I_0} \equiv \frac{-\Delta T}{T} \quad [1]$$

Here we have used a Taylor expansion $\log(1 + x) \approx x$ for small x.

The power dependence of the signal in the $\chi^{(3)}$ limit should be linear in both the pump and probe intensities. It is also possible to pump the sample using two-photon absorption, then probe an excited state absorption. This is a $\chi^{(5)}$ experiment, so that the signal depends quadratically on the pump intensity.

What is Measured in a Pump-Probe Experiment?

Origin of the Signal

The pump-probe measurement is a third-order nonlinear spectroscopy. A complete calculation of the signal requires computation of the third-order polarization induced by the interaction of the pump and probe with the material sample, together with an account for the kinetic evolution of populations (e.g., excited state reactant and product states, etc.).

The pump pulse, wavevector $\mathbf{k}_1$, interacts twice with the sample, thereby creating excited state population density $|e\rangle\langle e|$ and a hole-in-the-ground state population density $|g\rangle\langle g|$. These population densities propagate until the probe pulse, with wavevector $\mathbf{k}_2$, interacts with the system to induce a polarization that depends on the state of the system at time T after the pump. This induced polarization $P_s(t)$ is radiated in the $\mathbf{k}_s = \mathbf{k}_1 - \mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_2$ direction. Because the signal is radiated in the probe direction, the probe pulse acts as a local oscillator to heterodyne the signal. The change in probe-pulse spectral intensity after transmission through an

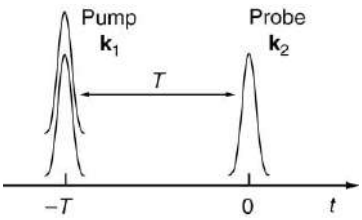

Figure 4 Labeling and time variable for the description of a pump-probe experiment. T defines the pump-probe delay.

optically thin sample of path length l, with concentration of absorbers c and refractive index n, is given by eqn [2]

$$\Delta I_{\mathrm{pr}}(\omega,T) \propto \frac{cl\omega_{\mathrm{pr}}}{n}\mathrm{Im}\left[\int dt E^{*}_{\mathrm{pr}}(t)P^{(3)}(0,T,t)e^{i\omega t}\right] \quad [2]$$

In eqn [2] $E_{\mathrm{pr}}(t)$ is the probe pulse electric field, ω is the center frequency of the probe, and * indicates complex conjugate. The time variables correspond to those indicated in Figure 4, and the signal is resolved as a function of frequency by Fourier transformation of $t \rightarrow \omega$. Experimentally this is achieved by dispersing the transmitted probe using a spectrograph. The first-time variable, the time interval between the two pump pulse interactions, is set to zero because we are interested here in the limit where the pulse duration is much shorter than T. In this regime – impulsive pump-probe – the measurement is sensitive to formation and decay of excited state species.

The induced polarization is given by eqn [3]

$$P^{(3)}(0,T,t) = \int_0^{\infty} \mathrm{d}t \int_0^{\infty} \mathrm{d}T[R^{\mathrm{SE}} - R^{\mathrm{ESA}} + R^{\mathrm{GSR}}] \times E_1^{*}(t-T)E_1(t-T)E_2(t) \quad [3]$$

where the response functions that contain information about the lineshape functions and kinetics are given by eqn [4]:

$$R^{\mathrm{SE}}(0,T,t) = |\mu_{eg}|^2|\mu_{eg}|^2 \exp(-i\omega_{eg}t) \times \exp[-g^{*}(t) + 2ig''(T+t) - 2ig''(T)] \times K^{\mathrm{SE}}(T) \quad [4a]$$

$$R^{\mathrm{ESA}}(0,T,t) = |\mu_{eg}|^2|\mu_{fe}|^2 \exp(-i\omega_{fe}t) \times \exp[-g^{*}(t) - 2ig''(T+t) + 2ig''(T)] \times K^{\mathrm{ESA}}(T) \quad [4b]$$

$$R^{\mathrm{GSR}}(0,T,t) = |\mu_{eg}|^2|\mu_{eg}|^2 \exp(-i\omega_{eg}t)\exp[-g(t)] \times K^{\mathrm{GSR}}(T) \quad [4c]$$

The overall intensity of each contribution to the signal is scaled by the transition dipole moment that connects the ground and excited state μ_{eg} or excited state and a higher excited state μ_{fe}. The intensity is further influenced by resonance of the probe-pulse spectrum with the transition frequencies ω_{eg} and ω_{fe}. These transition energies have a time-dependence owing to the relaxation of the excited state density in the condensed phase environment – the dynamic Stokes' shift. The details of the dynamic Stokes' shift depend on the bath, and are therefore contained in the imaginary part of the lineshape function $g(t) = g'(t) + ig''(t)$. The evolution of the Stokes' shift with pump-probe delay T is clearly seen in the excited state absorption spectrum shown in Figure 1. The lineshape function contains all the details of the time-scales of fluctuations of the bath and the coupling of the electronic transition of the molecule to these fluctuations. The lineshape function can be cast in the simple form given by eqn [5] using the Brownian oscillator model in the high temperature, high friction limit:

$$g(t) \approx (2\lambda k_B T_B/\hbar\Lambda)[\Lambda t - 1 + \exp(-\Lambda t)] + i(\lambda/\Lambda)[1 - \exp(-\Lambda t)] \quad [5]$$

The Brownian oscillator model attributes fluctuations of the transition frequencies to coupling between electronic states and an ensemble of low-frequency bath motions. In eqn [5]; λ is the solvent reorganization energy (half the Stokes' shift) and Λ is the modulation frequency of the solvent oscillator (τ_{L}^{-1} for a Debye solvent, where τ_{L} is the longitudinal dielectric relaxation time); k_{B} is the Boltzmann constant, and T_{B} is the temperature.

Information about the kinetics of the evolution of the system initiated by the pump pulse is contained in the terms K^{SE}, K^{ESA}, and K^{GSR}. These terms will be discussed in the next section.

The evolution of the excited state density is mapped onto the response functions R^{SE} and R^{ESA}. R^{SE} is the contribution arising from the probe pulse stimulating emission from the excited state e in the probe direction, which increases the probe intensity. Thus the spectrum of the signal arising from R^{SE} resembles the fluorescence emission of the sample, as shown in Figure 5. The excited state absorption contribution to the signal R^{ESA} depletes the probe intensity, and corresponds to an e $\rightarrow$ f electronic resonance, such as that shown in Figure 1. Note that the ESA signal contributes to the overall $\Delta I_{\mathrm{pr}}(\omega, T)$ with an opposite sign to the SE and GSR contributions. The ground-state recovery term R^{GSR} arises from depletion of the ground-state population by the pump pulse, thereby increasing the transparency of the sample over the

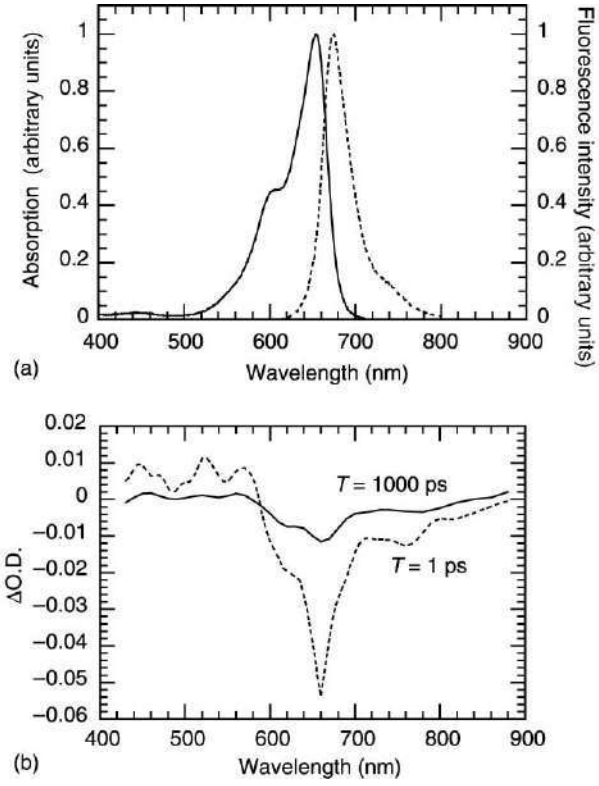

Figure 5 (a) shows the absorption and emission spectrum of a dilute solution of oxazine-1 in water. (b) shows the transient absorption spectrum of this system at various delays after pumping at 600 nm using picosecond pulses. The transient spectrum, ground state recovery and stimulated emission resemble the sum of the absorption and emission spectra.

spectrum of the ground-state absorption (see Figure 5). Thus the transmitted probe intensity is increased relative to that without the pump pulse. The GSR contribution to $\Delta I_{pr}(\omega, T)$ decreases with T according to the excited state lifetime of the photo-excited species.

The Coherent Spike

To calculate the pump-probe signal over delay times that are of the order of the pulse duration, one must calculate all possible ways that a signal can be generated with respect to all possible time orderings of the two pump interactions and the probe interactions (Liouville space pathways). Such a calculation reveals coherent as well as sequential contributions to the response of the system owing to the entanglement of time orderings that arise from the overlap of pulses of finite duration. The nonclassical, coherent, contributions dominate the signal for the time period during which pump and probe pulses overlap, leading to the coherent spike (or coherent 'artifact') (Figure 6).

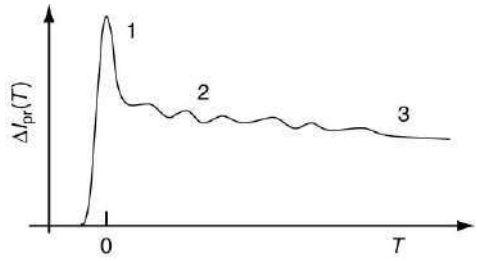

Figure 6 A schematic depiction of a pump-probe signal. The coherent spike is labeled 1. In region 2 we can see quantum beats as a sum of sinusoidal modulations of the signal. The label 3 denotes the long time decay of the signal, as determined by the population dynamics, K^{SE}, K^{GSR}, and K^{ESA}.

Nuclear Wavepackets

The electronic absorption spectrum represents a sum over all possible vibronic transitions, each weighted by a corresponding nuclear overlap factor according to the Franck–Condon principle. It was recognized by Heller and co-workers that in the time-domain representation of the electronic absorption spectrum, the sum over vibrational overlaps is replaced by the propagation of an initially excited nuclear wavepacket on the excited state according to $|i(t)\rangle = \exp(-Ht/\hbar)\,|i(0)\rangle$. Thus the absorption spectrum is written as

$$\sigma_i(\omega) = \frac{2\pi\omega}{3c}\int_{-\infty}^{\infty} \mathrm{d}t\, \exp[-i(\omega - \omega_{eg}^{i})t - g(t)] \times \langle i(0)\,|i(t)\rangle \qquad [6]$$

for mode i. The Hamiltonian H is determined by the displacement of mode i, Δ_i. A large displacement results in a strong coupling between the vibration and the electronic transition from state g to e. In the frequency domain this is seen as an intense vibronic progression. In the time domain this translates to an oscillation in the time-dependent overlap $\langle i(0)\,|i(t)\rangle$, with frequency ω_i, thereby modulating the intensity of the linear response as a function of t.

It follows that an impulsively excited coherent superposition of vibrational modes, created by an optical pulse that is much shorter than the vibrational period of each mode, propagates on the excited state as a wavepacket. The oscillations of this wavepacket then modulate the intensity of the transmitted probe pulse as a function of T to introduce features into the pump-probe signal known as quantum beats (see Figure 6). In fact, nuclear wavepackets propagate on both the ground and excited states in a pump-probe experiment. The short probe pulse interrogates the evolution of the nuclear wavepackets by projecting them onto the manifold of vibrational wavefunctions in an excited or ground state.

Femtosecond Kinetics

According to eqn [4] the kinetics of the evolution of the system initiated by the pump pulse are governed by the terms K^{SE}, K^{ESA}, and K^{GSR}. Each of these terms describes the kinetics of decay, production, or recovery of excited state and ground state populations. In the simplest conceivable case, each might be a single exponential function. The stimulated emission term K^{SE} contains information about the depopulation of the initially excited state, which for example, may occur through a chemical reaction involving the excited state species or energy transfer from that state. The excited state absorption term K^{ESA} contains information on the formation and decay of all excited state populations that give a transient absorption signal. The ground-state recovery term K^{GSR} contains information on the time-scales for return of all population to the ground state, which will occur by radiative and nonradiative processes. Usually the kinetics are assumed to follow a multi-exponential law such that

$$K^m(T, \lambda_{exc}, \omega) = \sum_{j=1}^{N} A_j(\lambda_{exc}, \omega) \exp(-t/\tau_j) \qquad [7]$$

where the amplitude A_j, but not the decay time coefficient τ_j, of each contribution in the coupled N-component system depends on the excitation wavelength λ_{exc} and the signal frequency ω. The index m denotes SE, ESA, or GSR.

In order to extract meaningful physical parameters from analysis of pump-probe data, a kinetic scheme in the form of eqn [7] needs to be constructed, based on a physical model to connect the decay of initially excited state population to formation of a product excited state and subsequent recovery of the ground-state population. An example of such a model is depicted in Figure 7. Here the stimulated emission contribution to the signal is attenuated according to the rate that state 1 goes to 2. The transient absorption appears at that same rate and decays according to the radiationless process 2 to 3. The rate of ground state recovery depends on 1 to 2 to 3 to 5 and possibly pathways involving the isomer 4.

The kinetic scheme is not easily extracted from pump-probe data because contributions from each of K^{SE}, K^{ESA}, and K^{GSR} contribute additively at each detection wavelength ω, as shown in eqns [2]–[4]. Moreover, additional, wavelength-dependent, decay or rise components may appear in the data on ultrafast time-scales owing to the time-dependent Stokes' shift (the $g''(t)$ terms indicated in eqns [4]). A spectral shift can resemble either a decay or rise component. In addition, the very fastest dynamics may be hidden beneath the coherent spike. For these reasons, simple fitting of pump-probe data usually cannot reveal the underlying kinetic model. Instead the data should be simulated according to eqns [3] and [4].

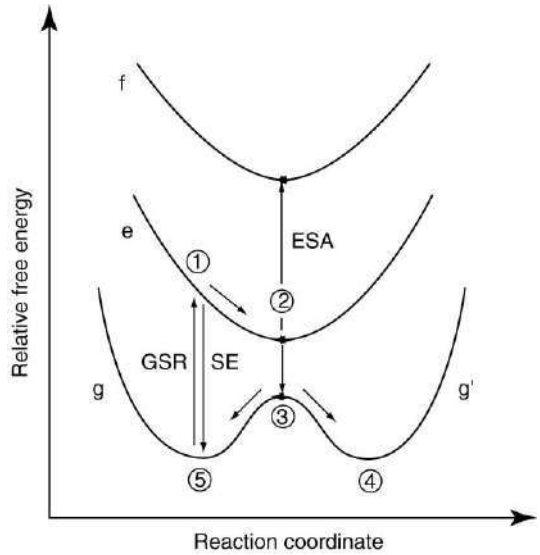

Figure 7 Free energy curves corresponding to a model for an excited state isomerization reaction, with various contributions to a pump-probe measurement indicated. See text for a description. (Courtesy of Jordanides XJ and Fleming GR.)

Alternative strategies involve global analysis of wavelength-resolved data to extract the kinetic evolution of spectral features, known as 'species associated decay spectra'. Global analysis of data effectively reduces the number of adjustable parameters in the fitting procedure and allows one to obtain physically meaningful information by associating species with their spectra. These methods are described in detail by Holzwarth.

List of Units and Nomenclature

ESA	excited state absorption
GSR	ground state recovery
SE	stimulated emission
S_n	nth singlet excited state
TR^3	time-resolved resonance Raman
$\chi^{(3)}$	nth-order nonlinear optical susceptibility

See also

Coherent Lightwave Systems. **Coherent Transients:** Ultrafast Studies of Semiconductors. **Nonlinear Optics, Applications:** Raman Lasers. **Optical Amplifiers:** Raman, Brillouin and Parametric Amplifiers. **Scattering:** Raman Scattering. **Spectroscopy:** Raman Spectroscopy.

Further Reading

Fleming GR (1986) *Chemical Applications of Ultrafast Spectroscopy*. New York: Oxford University Press.

Heller EJ (1981) The semi-classical way to molecular spectroscopy. *Accounts of Chemical Research* 14: 368–375.

Holzwarth AR (1996) Data analysis of time-resolved measurements. In: Amesz J and Hoff AJ (eds) *Biophysical Techniques in Photosynthesis*, pp. 75–92. Dordrecht, Germany: Kluwer.

Ippen EP, Shank CV, Wiensenfeld JM and Migus A (1980) Subpicosecond pulse techniques. *Philosophical Transactions of the Royal Society of London A* 298: 225–232.

Jonas DM and Fleming GR (1995) Vibrationally abrupt pulses in pump-probe spectroscopy. In: El-Sayed MA, Tanaka I and Molin Y (eds) *Ultrafast Processes in Chemistry and Photobiology*, pp. 225–256. Oxford, UK: Blackwell Scientific.

Klimov VI and McBranch DW (1998) Femtosecond high-sensitivity, chirp-free transient absorption spectroscopy using kilohertz lasers. *Optics Letters* 23: 277–279.

Lee SY (1995) Wave-packet model of dynamic dispersed and integrated pump-probe signals in femtosecond transition-state spectroscopy. In: Manz J and Wöste L (eds) *Femtosecond Chemistry*, vol. 1, pp. 273–298. Weinheim, Germany: VCH.

Lin SH, Alden R, Islampour R, Ma H and Villaeys AA (1991) *Density Matrix Method and Femtosecond Processes*. Singapore: World Scientific.

Mukamel S (1995) *Principles of Nonlinear Optical Spectroscopy*. New York: Oxford University Press.

Norrish RGW and Porter G (1949) Chemical reactions produced by very high light intensities. *Nature* 164: 658.

Towrie M, Parker AW, Shaikh W and Matousek P (1998) Tunable picosecond optical parametric generator-amplifier system for time resolved Raman spectroscopy. *Measurement Science and Technology* 9: 816–823.

Yan YJ and Mukamel S (1990) Femtosecond pump-probe spectroscopy of polyatomic molecules in condensed phases. *Physical Reviews A* 41: 6485–6504.

Time-Correlated Single-Photon Counting

A Beeby, University of Durham, Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Time-correlated single-photon counting, TCSPC, is a well-established technique for the determination of fluorescence lifetimes and related time-resolved fluorescence properties. Since the 1980s, there have been dramatic changes in the laser sources used for this work, which in turn have opened up the technique to a wide user base.

TCSPC: The Basics

In the TCSPC experiment, the sample is repeatedly excited by a high-repetition rate, low-intensity light source. Fluorescence from the sample is collected and may be passed though a cut-off filter or monochromator to remove scattered excitation light and to select the emission wavelength. The fluorescence is then focused onto a detector, typically a photomultiplier (PMT) or single-photon avalanche diode (SPAD) which detects single emitted photons. The 'raw' signal from the detector fluctuates considerably from event to event. In order to remove any timing error that could be induced by this, the signal is processed first by a constant fraction discriminator (CFD) in order to provide a stable and consistent timing pulse. This signal is used as an input for the time-to-amplitude converter (TAC) which determines the time interval between the excitation of the sample and the emission of the photon. The output from the TAC is a relatively slow pulse whose intensity is proportional to the time interval between the start and stop signals. A pulse height analyzer (PHA) is used to process the output from the TAC and increments a channel in its memory corresponding to a time window during which the photon was detected. As the measurement is repeated many times over, a histogram is constructed in the memory of the PHA, showing the number of photons detected as a function of the time interval. This histogram corresponds to the fluorescence decay.

As outlined below, it is desirable to have a light source with a moderately high repetition rate, typically of the order of MHz. Because it is not practical to start the TAC each time the light source 'fires', it is more usual to operate the TAC in the so-called 'reversed mode', whereby the start signal is derived from the detected photon and the stop is derived from the light source. The electronic circuitry is illustrated in Figure 1. In practice, many researchers still use modular components that are wired together manually, although there is increasingly a tendency towards the use of single PC-based cards which contain all the necessary discriminators, timing circuits, and data acquisition electronics. There are a number of commercial PC

cards that contain all the timing electronics, etc., required for TCSPC, for example, Becker and Hickl (Germany).

This approach to obtaining the fluorescence decay profile can only work if the detected photons are 'randomly selected' from the total emission from the sample. In practice, this means that the rate of detection of photons should be of the order of 1% of the excitation rate: that is, only one in a hundred excitation pulses gives rise to a detected photon. This histogram mirrors the fluorescence decay of the sample convolved with the instrument response function (IRF). The IRF is itself a convolution of the temporal response of the optical light pulse, the optical path in the collection optics, and the response of the detector and ancillary electronics. Experimentally, the IRF is usually recorded by placing a scattering material in the spectrometer.

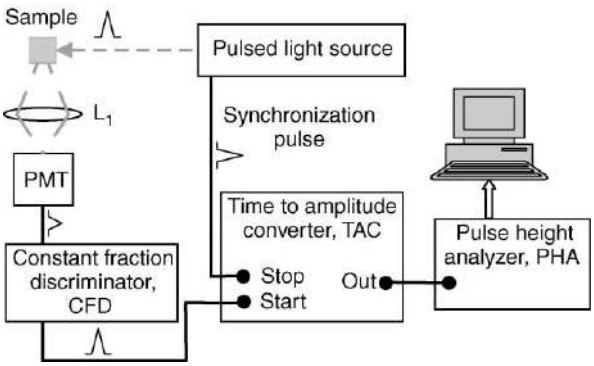

Figure 1 Schematic diagram of a TSCPC spectrometer. Fluorescence from the sample is collected by L_1, which may also include a monochromator/wavelength selection filter. A synchronization pulse is obtained either from the laser driver electronics directly, or via a photodiode and second constant fraction discriminator. The TAC is shown operating in a 'reversed mode' as is normal with a high repetition rate laser source.

In principle, the decay function of the sample can be extracted by deconvolution of the IRF from the decay function, although in practice it is more common to use the method of iterative re-convolution using a model decay function(s). Thus, the experimentally determined IRF is convolved with a model decay function and parameters within the function are systematically varied to obtain the best fit for a specific model. The most commonly used model assumes the decaying species to follow one or more exponential functions (sum of exponentials) or a distribution of exponentials. Further details regarding the mechanics of data analysis and fitting procedures can be found in the literature. A typical fit is illustrated in Figure 2.

The main aim of this review is to illustrate how recent advances in laser technology have changed the face of TCSPC and to discuss the types of laser light sources that are in use. The easiest way of approaching this topic is to define the criteria for an ideal light source for this experiment, and then to discuss the ways in which new technologies are meeting these criteria.

High Repetition Rate

A typical decay data set may contain something of the order of 0.1–5 million 'counts': it is desirable to

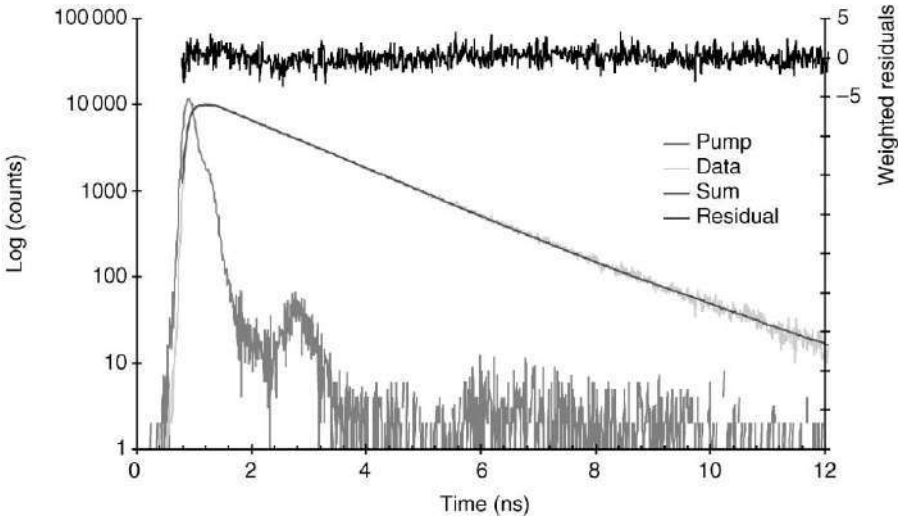

Figure 2 The fluorescence decay of Rhodamine B in water. The sample was excited using a 397 nm pulsed laser diode (1 MHz, <200 ps FWHM), and the emission was observed at 575 nm. Deconvolution of the IRF (dark gray trace) and observed decay (light gray trace) with a single exponential function gave a lifetime of 1.50 ns, with a $\chi^2 R = 1.07$, Durbin–Watson parameter = 1.78. The fitted curve is shown overlaid in gray, and the weighted residuals are shown offset in black.

acquire such a number of events in order that we can reliably employ statistical parameters to judge the quality of the fitted decay function(s). As stated above, the rate of photon acquisition should be of the order of 1% of the excitation rate in order to obtain reliable, good-quality data. It is clear then that a high repetition rate source is desirable in order to reduce the overall data acquisition time. However, it is important that the repetition rate of the source is not too high, otherwise the data show 'overlap' effects, caused by the fluorescence intensity from one excitation pulse not completely decaying before the next laser pulse arrives at the sample. Under these conditions the observed fluorescence decay is perturbed and extraction of complex kinetic behavior is more difficult. As a general rule of thumb, it is recommended that the inter-pulse separation is greater than five times the lifetime of the longest component of the fluorescence decay. Fluorescence lifetimes can be as long as 100 ns, indicating a maximum repetition rate of 2 MHz, although when investigating fluorophores with a shorter lifetime correspondingly higher frequencies can be selected. Clearly the ability to vary the repetition rate according to the type of sample under investigation is an advantage.

Short Pulse Duration

Typically the minimum lifetime that can be determined by a TCSPC spectrometer is limited by the instrument response function: as a general rule it is possible to extract fluorescence lifetimes of about one tenth of the full width at half maximum (FWHM) of the IRF by reconvolution methods. As discussed above, the IRF is the product of a number of factors, including the optical pulse duration, differences in the optical path traversed by the fluorescence, the response time of the detector, and the temporal response of the associated timing electronics. Ideally the excitation source should be as short as possible, and with modern laser-based TCSPC systems, it is usually the case that the time-response of the detector is the limiting factor. The introduction of high-speed microchannel plate (MCP) photomultipliers have lead to very short IRFs, which can be <50 ps FWHM. When using these detectors is it essential to keep the optical transit-time spread as low as possible: that is, the path taken by light through a monochromator should be independent of wavelength and trajectory taken through the optics. Many research groups advocate the use of a subtractive dispersive double monochromator for this purpose.

Tunability

Early TCSPC experiments were carried out using hydrogen- or nitrogen-filled flashlamps which produced broadband ns-duration pulses with low repetition rates (<100 kHz). Variable wavelength excitation is possible using these sources, but their low intensity and low repetition rates mean that data acquisition periods can be very long. The early 1980s saw the introduction of cavity-dumped ion lasers and synchronously pumped mode-locked cavity-dumped dye lasers, which provided significantly higher repetition rates, typically up to 4 MHz, and greater intensities. However, the TCSPC experiments that could be carried out using these laser-based sources were somewhat limited by the wavelengths that could be generated by the argon ion and synchronously pumped dye lasers, and their second harmonics. The dye lasers were restricted to the use of Rhodamine 6G as the gain medium, which has a limited tuning range ($\sim$580–620 nm). It is often desirable to vary the excitation wavelength as part of a photophysical study, either as a systematic study of the effects of excitation energy upon the excited state behavior, or in order to optimize excitation of a particular chromophore in a system. To some extent in the past, the wavelengths available from the laser systems dictated the science that they could be used to address. The mid- to late-1980s saw the introduction of mode-locked continuous wave (CW) Nd:YAG lasers as pump lasers and increased number of dyes that could be synchronously pumped, providing a broader range of excitation wavelengths as well as increased stability. One drawback of these complex and often fickle laser systems was the level of expertise and attention they required to keep their performance at its peak.

Since the 1990s, the introduction of two important new types of laser have had a significant impact on the use of TCSPC as an investigative method. The first is based upon the mode-locked titanium sapphire laser which offers tuneable radiation from the deep UV to the NIR, whilst the second series of laser are based upon (relatively) low-cost, turn-key solid state diode lasers. The developments and merits of these new laser sources, which now dominate the market for TCSPC sources, are discussed below.

Ti-Sapphire

The discovery of the titanium doped sapphire laser has led to a revolution in ultrafast laser sources. Coupled with efficient diode-pumped solid state

Nd:YAG pump lasers, this medium currently dominates the entire ultrafast laser market, from high repetition, low pulse energy systems through to systems generating very high pulse energies at kHz repetition rates.

Ti-sapphire is, in many respects, an ideal laser gain medium: it has a high energy storage capacity, it has a very broad-gain bandwidth, and it has good thermal properties. Following the first demonstration of laser action in a Ti-sapphire crystal, the CW Ti-sapphire laser was quickly commercialized. It proved to be a remarkable new laser medium, providing greater efficiency, with outputs of the order of 1 W. The laser also provided an unprecedented broad tuning range, spanning from just below 700 nm to 1000 nm. The most important breakthrough for time-resolved users came in 1991, when self-mode-locking in the medium was demonstrated. This self-mode-locking, sometimes referred to as Kerr-lens mode-locking, arises from the nonlinear refractive index of the material and is implemented by simply placing an aperture within the cavity at an appropriate point. The broad-gain bandwidth of the medium results in very short pulses: pulse durations of the order of tens of femtoseconds can be routinely achieved with commercial laser systems.

The mode-locked Ti-sapphire laser was quickly seized upon and commercialized. The first generation of systems used multiwatt CW argon ion lasers to pump them, making the systems unwieldy and expensive to run, but in recent years these have been replaced by diode-pumped Nd:YAG lasers, providing an all solid-state, ultrafast laser source. Equally significantly, the current generation of lasers are very efficient and do not require significant electrical power or cooling capacity. There are several commercial suppliers of pulsed Ti-sapphire lasers, including Coherent and Spectra Physics.

A typical mode-locked Ti-sapphire laser provides an 80 MHz train of pulses with a FWHM of ~100 fs, with an average power of up to 1.5 W and are tuneable from 700 nm to almost 1000 nm. The relatively high peak powers associated with these pulses means that second- or even third-harmonic generation is relatively efficient, providing radiation in the ranges 350–480 nm and 240–320 nm. The output powers of the second and third harmonics far exceeds that required for TCSPC experiments, making them an attractive source for the study of fluorescence lifetime measurements. The simple mode-locked Ti-sapphire laser has some drawbacks: the very short pulse duration provides a broad-bandwidth excitation, the repetition rate is often too high with an inter-pulse separation of ~12.5 ns, and the laser does not provide continuous tuning from the deep-UV though to the visible. Incrementally, all of these points have been addressed by the laser manufacturers.

As well as operating as a source of femtosecond pulses, with their associated bandwidths of many tens of nm, some modern commercial Ti-sapphire lasers can, by adjustment of the optical path and components, be operated in a mode that provides picosecond pulses whilst retaining their broad tuning curve. Under these conditions, the bandwidth of the output is greatly reduced, making the laser resemble a synchronously pumped mode-locked dye laser.

As stated above, a very high repetition rate can be undesirable if the system under study has a long fluorescence lifetime due to the fluorescence decay from an earlier pulse not decaying completely before the next excitation pulse. Variation of the repetition of output frequency can be addressed by two means. The first is to use a pulse-picker. This is an opto-accoustic device which is placed outside the laser cavity. An electrical signal is applied to a device, synchronized to the frequency of the mode-locked laser to provide pulses at frequencies from ~4 MHz down to single-shot. The operation of a pulse picker is somewhat inefficient, with less than half of the original pulse energy being obtained in the output, although output power is not usually an issue when using these sources. More importantly, when using pulse-pickers, it is essential to ensure that there is no pre- or post-pulse which can lead to artifacts in the observed fluorescence decays. The second means of reducing the pulse duration is cavity dumping. This is achieved by extending the laser cavity and placing an opto-acoustic device, known as a Bragg cell, at a beam-waist within the cavity. When a signal is applied to the intracavity Bragg cell, synchronized with the passage of a mode-locked pulse though the device, a pulse of light is extracted from the cavity. The great advantage of cavity-dumping is that it builds up energy within the optical cavity and when pulses are extracted these are of a higher pulse energy than the mode-locked laser alone. This in turn leads to more efficient harmonic generation. This addition of the cavity dumper system extends the laser cavity and hence reduces the overall frequency of the mode-locked laser to ~50 MHz, and can provide pulses at 9 MHz down to single shot.

The extension of the tuning range of the Ti-sapphire laser has been facilitated by the introduction of the optical parametric oscillator (OPO). In the OPO, a short wavelength pump beam passes through a nonlinear optical crystal and generates two tuneable

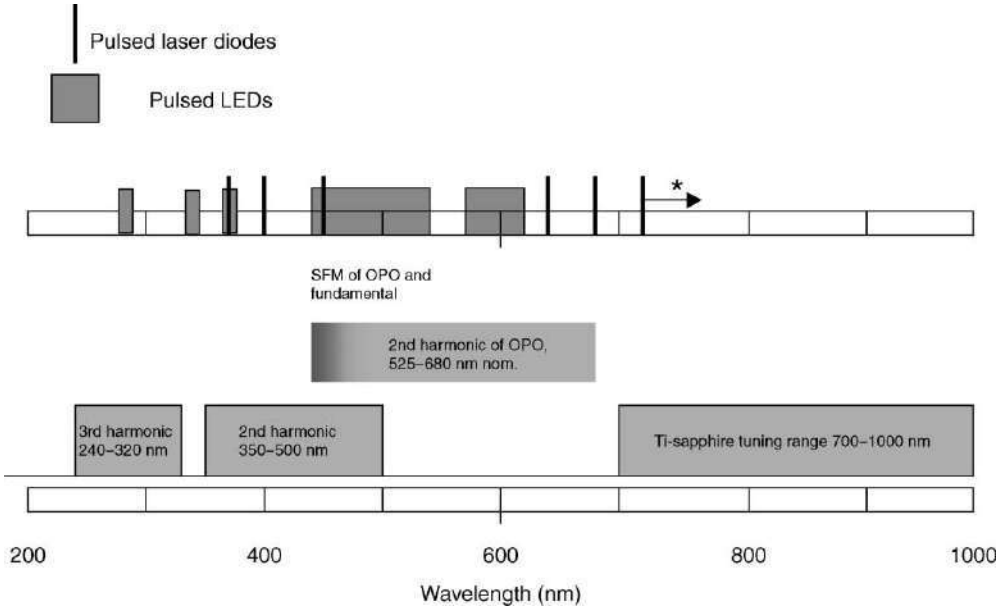

Figure 3 An illustration of the spectral coverage provided by currently available pulsed semiconductor lasers and LEDs, and a modern Ti-sapphire laser and OPO. Note that the OPO's output is frequency doubled, and the wavelengths in the range 525– ~ 440 nm may be obtained by sum frequency mixing of the Ti-sapphire's fundamental output with the output of the OPO. * denotes that there are many long wavelength pulsed laser diodes falling in the range $\lambda > 720$ nm.

beams of longer wavelengths. Synchronous pumping of the OPOs with either a fs or ps Ti-sapphire laser ensures reasonably high efficiencies. The outputs from the OPO may be frequency doubled, with the option of intracavity doubling enhancing the efficiency still further. Using an 800 nm pump wavelength, an OPO provides typical signal and idler tuning ranges of 1050– >1350 nm and 2.1–3.0 μm, respectively. The signal can be doubled to give 525–680 nm, effectively filling the visible spectrum. The average powers obtainable from Ti-sapphire pumped OPOs makes them eminently suitable for TCSPC experiments. A summary of the spectral coverage of the Ti-sapphire laser and OPO, and low-cost solid-state sources are illustrated in Figure 3.

An additional advantage of the Ti-sapphire laser is the possibility of multiphoton excitation. The combination of high average power and very short pulse duration of a mode-locked Ti-sapphire laser means that the peak-powers of the pulses can be very high. Focusing the near-IR output of the laser gives rise to exceedingly high photon densities within a sample which can lead to the simultaneous absorption of two or more photons. This is of particular significance when recording fluorescence lifetimes in conjunction with a microscope, particularly when the microscope is operated in a confocal configuration. Such systems are now being used for a derivative of TCSPC, known as fluorescence lifetime imaging (FLM).

In summary, the Ti-sapphire laser is a very versatile and flexible source that is well suited to the TCSPC experiment. Off-the-shelf systems provide broad spectral coverage and variable repetition rates. Undoubtedly this medium will be the mainstay of TCSPC for some time to come.

Diode Lasers

Low-power diode lasers, with outputs in the range 630–1000 nm, have been commercially available for some time. These devices, which are normally operated as CW sources, can be operated in a pulsed mode, providing a high frequency train of sub-ns pulses. However, the relatively long output wavelengths produced by this first generation of diode lasers severely restricted the range of chromophores that could be excited and this proved to be a limiting factor for their use as TCSPC sources. Attempts to generate shorter wavelengths from these NIR laser diodes by second harmonic generation (SHG), were never realized: the SHG process is extremely inefficient due to the very low peak output power of these lasers. Shorter wavelengths could be obtained by pulsing conventional light-emitting diodes (LEDs), which were commercially available with emission wavelengths covering the visible spectrum down to ~450 nm. However, these had the disadvantage of exhibiting longer pulse durations than those achievable from laser diodes and have

much broader emission bandwidths than a laser source.

The introduction of violet and blue diode lasers and UV-LEDs by the Nichia Corporation (Japan) in the late 1990s resulted in a renewed interest in sources utilizing these small, solid-state devices. At least three manufacturers produce ranges of commercial TCSPC light sources based upon pulsed LEDs and laser diodes and can now boast a number of pulsed laser sources in their ranges including 375, 400, 450, 473, 635, 670, and 830 nm. Currently commercial pulsed laser diodes and LEDs are manufactured by Jobin-Yvon IBH (UK), Picoquant (Germany), and Hamamatso (Japan). These small, relatively low-cost devices offer pulse durations of <100 ps and excitation repetition rates of up to tens of MHz making them ideally suited to 'routine' lifetime measurements. Although these diode laser sources are not tunable and do not span the full spectrum, they are complemented by pulsed LEDs which fill the gap in the visible spectrum. Pulsed LEDs generally have a longer pulse duration and broader, lower intensity spectral output, but are useful for some applications. At the time of writing there are no diode laser sources providing wavelengths of <350 nm, although Jobin-Yvon IBH (UK) have recently announced pulsed LEDs operating at 280 nm and 340 nm. As LEDs, and possibly diode lasers, emitting at these shorter wavelengths become more widely available, the prospects of low running costs and ease of use makes these semiconductor devices an extremely attractive option for many TCSPC experiments. Commercial TCSPC spectrometers based around these systems offer a comparatively low-cost, turnkey approach to fluorescence lifetime based experiments, opening up what was a highly specialized field to a much broader range of users.

See also

Lasers: Noble Gas Ion Lasers. **Light Emitting Diodes**. **Ultrafast Laser Techniques:** Pulse Characterization Techniques.

Further Reading

Lakowicz JR (ed.) (2001) *Principles of Fluorescence Spectroscopy*. Plenum Press.

O'Connor DV and Phillips D (1984) *Time-Correlated Single Photon Counting*.

Spence DE, Kean PN and Sibbett W (1991) *Optics Letters* 16: 42–44.

Transient Holographic Grating Techniques in Chemical Dynamics

E Vauthey, University of Geneva, Geneva, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Over the past two decades, holographic techniques have proved to be valuable tools for investigating the dynamics of chemical processes. The aim of this chapter is to give an overview of the main applications of these techniques for chemical dynamics. The basic principle underlying the formation and the detection of elementary transient holograms, also called transient gratings, is first presented. This is followed by a brief description of a typical experimental setup. The main applications of these techniques to solve chemical problems are then discussed.

Basic Principle

The basic principle of the transient holographic technique is illustrated in Figure 1. The sample material is excited by two laser pulses at the same wavelength and crossed at an angle $2\theta_{pu}$. If the two pump pulses have the same intensity, I_{pu}, the intensity distribution at the interference region, assuming plane waves, is

$$I(x) = 2I_{pu}\left[1 + \cos\left(\frac{2\pi x}{\Lambda}\right)\right] \qquad [1]$$

where $\Lambda = \lambda_{pu}/(2\sin\theta_{pu})$ is the fringe spacing, λ_{pu} is the pump wavelength, and I_{pu} is the intensity of one pump pulse.

As discussed below, there can be many types of light–matter interactions that lead to a change in the optical properties of the material. For a dielectric material, they result in a spatial modulation of the optical susceptibility and thus of the complex refractive index, $\tilde{n}$. The latter distribution can be described as a Fourier cosine series:

$$\tilde{n}(x) = \tilde{n}_0 + \sum_{m=1}^{\infty} \tilde{n}_m \cos\left(\frac{m2\pi x}{\Lambda}\right) \qquad [2]$$

where $\tilde{n}_0$ is the average value of $\tilde{n}$. In the absence of saturation effects, the spatial modulation of $\tilde{n}$ is harmonic and the Fourier coefficients with $m > 1$ vanish. In this case, the peak to null variation of the complex refractive index, $\Delta\tilde{n}$, is equal to the Fourier coefficient $\tilde{n}_1$. The complex refractive index can be split into its real and imaginary components:

$$\tilde{n} = n + iK \qquad [3]$$

where n is the refractive index and K is the attenuation constant.

The hologram created by the interaction of the crossed pump pulses consists in periodic, one-dimensional spatial modulations of n and K. Such distributions are nothing but phase and amplitude gratings, respectively. A third laser beam at the probe wavelength, λ_{pr}, striking these gratings at Bragg angle, $\theta_B = \arcsin(\lambda_{pr}/2\Lambda)$, will thus be partially diffracted (Figure 1b). The diffraction efficiency, η, depends on the modulation amplitude of the optical properties. In the limit of small diffraction efficiency ($\eta < 0.01$), this relationship is given by

$$\eta = \frac{I_{dif}}{I_{pr}} \cong \left[\left(\frac{\ln 10\Delta A}{4\cos\theta_B}\right)^2 + \left(\frac{\pi d\Delta n}{\lambda_{pr}\cos\theta_B}\right)^2\right] \times \exp\left(-\frac{\ln 10A}{\cos\theta_B}\right) \qquad [4]$$

where I_{dif} and I_{pr} are the diffracted and the probe intensity respectively, d is the sample thickness, and $A = 4\pi dK/(\lambda \ln 10)$ is the average absorbance.

The first and the second terms in the square bracket describe the contribution of the amplitude and phase gratings respectively, and the exponential term accounts for the reabsorption of the diffracted beam by the sample.

The main processes responsible for the variation of the optical properties of an isotropic dielectric material are summarized in Figure 2.

The modulation of the absorbance, ΔA, is essentially due to the photoinduced concentration change,

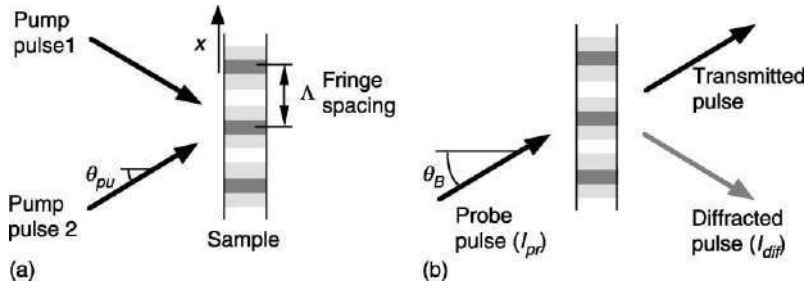

Figure 1 Principle of the transient grating technique: (a) grating formation (pumping), (b) grating detection (probing).

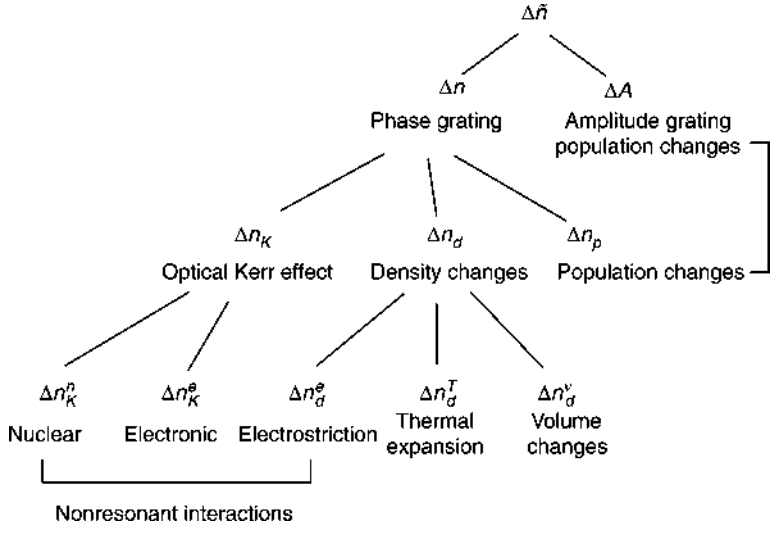

Figure 2 Classification of the possible contributions to a transient grating signal.

ΔC, of the different chemical species i (excited state, photoproduct, …):

$$\Delta A(\lambda_{pr}) = \sum_i \varepsilon_i(\lambda_{pr})\Delta C_i \qquad [5]$$

where ε_i is the absorption coefficient of the species i.

The variation of the refractive index, Δn, has several origins and can be expressed as

$$\Delta n = \Delta n_K + \Delta n_p + \Delta n_d \qquad [6]$$

Δn_K is the variation of refractive index due to the optical Kerr effect (OKE). This nonresonant interaction results in an electronic polarization (electronic OKE) and/or in a nuclear reorientation of the molecules (nuclear OKE) along the direction of the electric field associated with the pump pulses. As a consequence, a transient birefringence is created in the material. This effect is usually discussed within the framework of nonlinear optics in terms of intensity dependent refractive index or third order nonlinear susceptibility. Electronic OKE occurs in any dielectric material under sufficiently high light intensity. On the other hand, nuclear OKE is mostly observed in liquids and gases and depends strongly on the molecular shape.

Δn_p is the change of refractive index related to population changes. Its magnitude and wavelength dependence can be obtained by Kramers–Kronig transformation of $\Delta A(\lambda)$ or $\Delta K(\lambda)$:

$$\Delta n_p(\lambda) = \frac{1}{2\pi^2}\int_0^\infty \frac{\Delta K(\lambda')}{1-(\lambda'/\lambda)^2}\,\mathrm{d}\lambda' \qquad [7]$$

Δn_d is the change of refractive index associated with density changes. Density phase gratings can have essentially three origins:

$$\Delta n_d = \Delta n_d^t + \Delta n_d^v + \Delta n_d^e \qquad [8]$$

Δn_d^t is related to the temperature-induced change of density. If a fraction of the excitation energy is converted into heat, through a nonradiative transition or an exothermic process, the temperature becomes spatially modulated. This results in a variation of density, hence to a modulation of refractive index with amplitude Δn_d^t. Most of the temperature dependence of n originates from the density. The temperature-induced variation of n at constant density is much smaller than Δn_d^t.

Δn_d^v is related to the variation of volume upon population changes. This volume comprises not only the reactant and product molecules but also their environment. For example, in the case of a photodissociation, the volume of the product is larger than that of the reactant and a positive volume change can be expected. This will lead to a decrease of the density and to a negative Δn_d^v.

Finally, Δn_d^e is related to electrostriction in the sample by the electric field of the pump pulses. Like OKE, this is a nonresonant process that also contributes to the intensity dependent refractive index. Electrostriction leads to material compression in the regions of high electric field strength. The periodic compression is accompanied by the generation of two counterpropagating acoustic waves with wave vectors, $\vec{k}_{ac} = \pm(2\pi/\Lambda)\vec{i}$, where $\vec{i}$ is the unit vector along the modulation axis. The interference of these acoustic waves leads to a temporal modulation of Δn_d^e at the acoustic frequency υ_{ac}, with $2\pi\upsilon_{ac} = k_{ac}\upsilon_s$, υ_s being the speed of sound. As Δn_d^e oscillates between negative and positive values, the diffracted intensity, which is proportional to $(\Delta n_d^e)^2$, shows at temporal oscillation at twice the acoustic frequency. In most cases, Δn_d^e is weak and can be neglected if the pump pulses are within an absorption band of the sample.

The modulation amplitudes of absorbance and refractive index are not constant in time and their temporal behavior depends on various dynamic processes in the sample. The whole point of the transient grating techniques is precisely the measurement of the diffracted intensity as a function of time after excitation to deduce dynamic information on the system.

In the following, we will show that the various processes shown in Figure 2, that give rise to a diffracted signal, can in principle be separated by choosing appropriate experimental parameters, such as the timescale, the probe wavelength, the polarization of the four beams, or the crossing angle.

Experimental Setup

A typical experimental arrangement for pump–probe transient grating measurements is shown in Figure 3. The laser output pulses are split in three parts. Two parts of equal intensity are used as

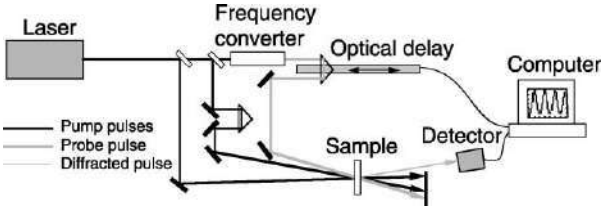

Figure 3 Schematic of a transient grating setup with pump–probe detection.

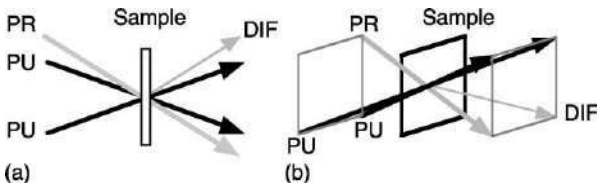

Figure 4 Beam geometry for transient grating: (a) in plane; (b) boxcars.

pump pulses and are crossed in the sample. In order to ensure time coincidence, one pump pulse travels along an adjustable optical delay line. The third part, which is used for probing, can be frequency converted using a dye laser, a nonlinear crystal, a Raman shifter, or white light continuum generation. The probe pulse is sent along a motorized optical delay line before striking the sample at the Bragg angle. There are several possible beam configurations for transient grating and the two most used are illustrated in Figure 4. When the probe and pump pulses are at different wavelengths, they can be in the same plane of incidence as shown in Figure 4a. However, if the pump and probe wavelengths are the same, the folded boxcars geometry shown in Figure 4b has to be used. The transient grating technique is background free and the diffracted signal propagates in a well-defined direction. In a pump–probe experiment, the diffracted signal intensity is measured as a function of the time delay between the pump and probe pulses. Such a setup can be used to probe dynamic processes occurring in timescales going from a few fs to a few ns, the time resolution depending essentially on the duration of the pump and probe pulses. For slower processes, the grating dynamics can be probed in real time with a *cw* laser beam and a fast photodetector.

Applications

The Transient Density Phase Grating Technique

If the grating is probed at a wavelength far from any absorption band, the variation of absorbance, $\Delta A(\lambda_{pr})$, is zero and the corresponding change of refractive index, $\Delta n_p(\lambda_{pr})$, is negligibly small. In this case, eqn [4] simplifies to

$$\eta \cong \left(\frac{\pi d}{\lambda_{pr}\cos\theta_B}\right)^2 \times \Delta n_d^2 \qquad [9]$$

In principle, the diffracted signal may also contain contributions from the optical Kerr effect, Δn_K, but we will assume here that this non-resonant and ultrafast response is negligibly small. The density change can originate from both heat releasing processes and volume differences between the products and reactants, the former contribution usually being much larger than the latter. Even if the heat releasing process is instantaneous, the risetime of the density grating is limited by thermal expansion. This expansion is accompanied by the generation of two counterpropagating acoustic waves with wave vectors, $\vec{k}_{ac} = \pm(2\pi/\Lambda)\vec{i}$. One can distinguish two density gratings: 1) a diffusive density grating, which reproduces the spatial distribution of temperature and which decays by thermal diffusion; and 2) an acoustic density grating originating from the standing acoustic wave and whose amplitude oscillates at the acoustic frequency υ_{ac}.

The amplitudes of these two gratings are equal but of opposite sign. Consequently, the time dependence of the modulation amplitude of the density phase grating is given by

$$\Delta n_d(t) = \left(\frac{\beta Q}{\rho C_v} + \Delta V\right)\rho\left(\frac{\partial n}{\partial \rho}\right)R(t) \qquad [10a]$$

with

$$R(t) = 1 - \cos(2\pi\upsilon_{ac}t) \times \exp(-\alpha_{ac}\upsilon_s t) \qquad [10b]$$

where ρ, β, C_v, and α_{ac} are the density, the volume expansion coefficient, the heat capacity, and the acoustic attenuation constant of the medium, respectively, Q is the amount of heat deposited during the photoinduced process, and ΔV is the corresponding volume change. As the standing acoustic wave oscillates, its corresponding density grating interferes with the diffusive density grating. Therefore, the total modulation amplitude of the density and thus Δn_d exhibits the oscillation at υ_{ac}. Figure 5 shows the time profile of the diffracted

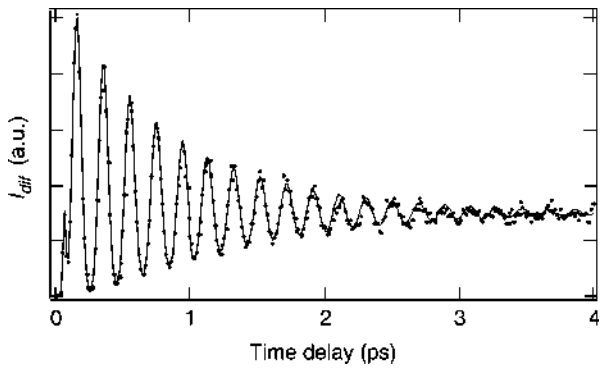

Figure 5 Time profile of the diffracted intensity measured with a solution of malachite green after excitation with two pulses with close to counterpropagating geometry.

intensity measured with a solution of malachite green. After excitation to the S_1 state, this dye relaxes nonradiatively to the ground state in a few ps. For this measurement, the sample solution was excited with two 30 ps laser pulses at 532 nm crossed with an angle close to 180°. The continuous line is the best fit of eqns [9] and [10]. The damping of the oscillation is due to acoustic attenuation. After complete damping, the remaining diffracted signal is due to the diffusive density grating only.

$R(t)$ can be considered as the response function of the sample to a prompt heat release and/or volume change. If these processes are not instantaneous compared to an acoustic period ($\tau_{ac} = \upsilon_{ac}^{-1}$), the acoustic waves are not created impulsively and the time dependence of Δn_d is

$$\Delta n_d(t) = \left(\frac{\beta Q}{\rho C_v} + \Delta V\right)\rho\left(\frac{\partial n}{\partial \rho}\right)F(t) \qquad [11a]$$

with

$$F(t) = \int_{-\infty}^{t} R(t - t') \cdot f(t')\mathrm{d}t' \qquad [11b]$$

where $f(t)$ is a normalized function describing the time evolution of the temperature and/or volume change. In many cases, $f(t) = \exp(-k_r t)$, k_r being the rate constant of the process responsible for the change. Figure 6 shows the time profile of the diffracted intensity calculated with eqns [9] and [11] for different values of k_r.

If several processes take place, the total change of refractive index is the sum of the changes due to the individual processes. In this case, Δn_d should be expressed as

$$\Delta n_d(t) = \sum_i \left(\frac{\beta Q_i}{\rho C_v} + \Delta V_i\right)\rho\left(\frac{\partial n}{\partial \rho}\right)F_i(t) \qquad [12]$$

The separation of the thermal and volume contributions to the diffracted signal is problematic. Several approaches have been proposed, the most used being the measurement in a series of solvents with different expansion coefficient β. This method requires that the other solvent properties like refractive index, dielectric constant, or viscosity, are constant or that the energetics of the system investigated does not depend on them. This separation is easier when working with water because its β value vanishes at 4 °C. At this temperature, the density variations are due to the volume changes only.

The above equations describe the growth of the density phase grating. However, this grating is not permanent and decays through diffusive processes. The phase grating originating from thermal expansion decays via thermal diffusion with a rate constant k_{th} given by

$$k_{th} = D_{th}\left(\frac{2\pi}{\Lambda}\right)^2 \qquad [13]$$

where D_{th} is the thermal diffusivity. Table 1 shows k_{th} values in acetonitrile for different crossing angles.

The decay of the phase grating originating from volume changes depends on the dynamics of the population responsible for ΔV (vide infra).

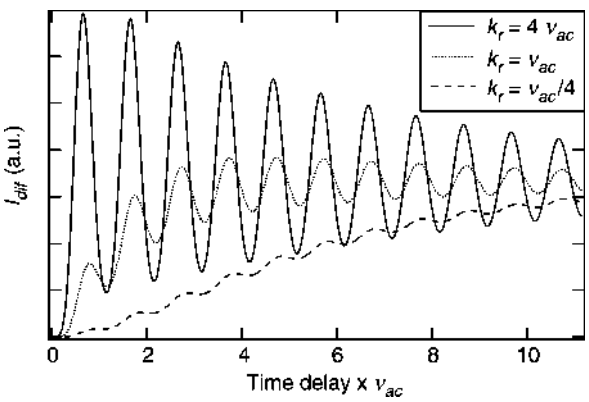

Figure 6 Time profiles of the diffracted intensity calculated using eqns [9] and [11] with various k_r values.

Time resolved optical calorimetry

A major application of the density phase grating in chemistry is the investigation of the energetics of photo-induced processes. The great advantage of this technique over other optical calorimetric methods, like the thermal lens or the photoacoustic spectroscopy, is its superior time resolution. The time constant of the fastest heat releasing process that can be time resolved with this technique is of the order of the acoustic period. The shortest acoustic period is achieved when forming the grating with two counterpropagating pump pulses ($2\theta_{pu} = 180°$). In this case the fringe spacing is

Table 1 Fringe spacing, Λ, acoustic frequency, ν_{ac}, thermal diffusion rate constant, k_{tr}, for various crossing angles of the pump pulses, 2 θ_{pu}, at 355 nm and in acetonitrile

$2\,\theta_{pu}$	Λ (μm)	υ_{ac} (s^{-1})	k_{th} (s^{-1})
0.5°	40.7	3.2×10^7	4.7×10^3
50°	0.42	3.1×10^9	4.5×10^7
180°	0.13	9.7×10^9	4.7×10^8

$\Lambda = \lambda_{pu}/2n$ and, with UV pump pulses, an acoustic period of the order of 100 ps can be obtained in a typical organic solvent.

For example, Figure 7a shows the energy diagram for a photoinduced electron transfer (ET) reaction between benzophenone (BP) and an electron donor (D) in a polar solvent. Upon excitation at 355 nm, $^1BP^*$ undergoes intersystem-crossing (ISC) to $^3BP^*$ with a time constant of about 10 ps.

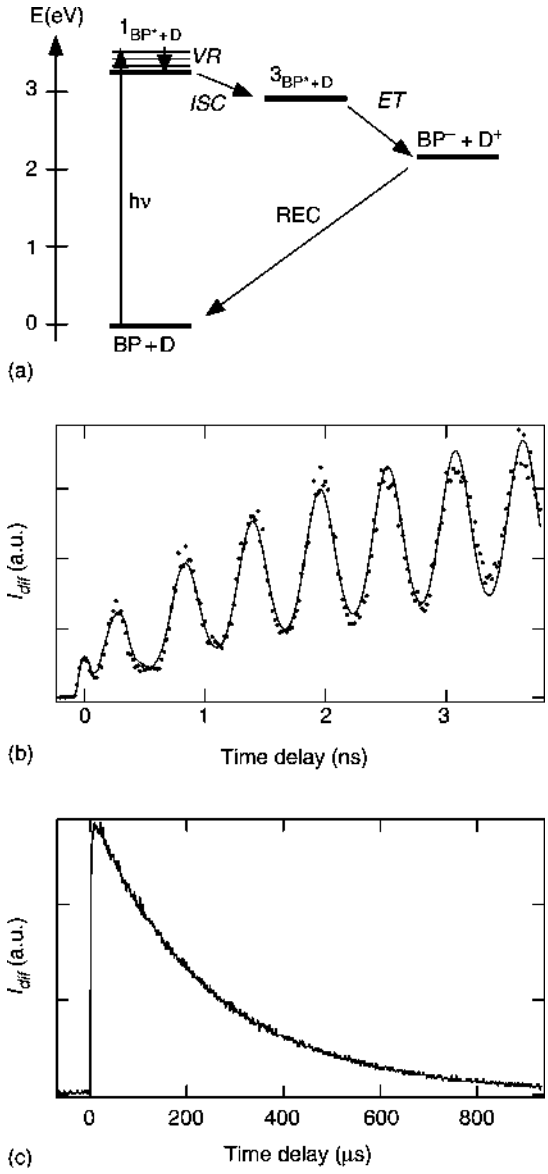

Figure 7 (a) Energy diagram of the states involved in the photoinduced electron transfer reaction between benzophenone (BP) and an electron donor (D) (VR: vibrational relaxation, ISC: intersystem crossing; ET: electron transfer; REC: charge recombination). (b) Time profile of the diffracted intensity measured at 590 nm after excitation at 355 nm of a solution of BP and 0.05 M D. (c) Same as (b) but measured at 1064 nm with a *cw* laser.

After diffusional encounter with the electron donor, ET takes place and a pair of ions is generated. With this system, the whole energy of the 355 nm photon ($E = 3.49$ eV) is converted into heat. The different heat releasing processes can be differentiated according to their timescale. The vibrational relaxation to $^1BP^*$ and the ensuing ISC to $^3BP^*$ induces an ultrafast release of 0.49 eV as heat. With donor concentrations of the order of 0.1 M, the heat deposition process due to the electron transfer is typically in the ns range. Finally, the recombination of the ions produces a heat release in the microsecond timescale. Figure 7b shows the time profile of the diffracted intensity measured after excitation at 355 nm of BP with 0.05 M D in acetonitrile. The 30 ps pump pulses were crossed at 27° ($\tau_{ac} = 565$ ps) and the 10 ps probe pulses were at 590 nm. The oscillatory behavior is due to the ultrafast heat released upon formation of $^3BP^*$, while the slow dc rise is caused by the heat dissipated upon ET. As the amount of energy released in the ultrafast process is known, the energy released upon ET can be determined by comparing the amplitudes of the fast and slow components of the time profile. The energetics as well as the dynamics of the photoinduced ET process can thus be determined. Figure 7c shows the decay of the diffracted intensity due to the washing out of the grating by thermal diffusion. As both the acoustic frequency and the thermal diffusion depends on the grating wavevector, the experimental time window in which a heat releasing process can be measured, depends mainly on the crossing angle of the pump pulses as shown in Table 1.

Determination of material properties

Another important application of the transient density phase grating technique is the investigation of material properties. Acoustics waves of various frequencies, depending on the pump wavelength and crossing angle, can be generated without physical contact with the sample. Therefore, acoustic properties, such as the speed of sound and the acoustic attenuation of the material, can be easily obtained from the frequency and the damping of the oscillation of the transient grating signal.

Similarly, the optoelastic constant of a material, $\rho\partial n/\partial\rho$, can be determined from the amplitude of the signal (see eqn [11]). This is done by comparing the signal amplitude of the material under investigation with that obtained with a known standard.

Finally, the thermal diffusivity, D_{th}, can be easily obtained from the decay of the thermal density

phase grating, such as that shown in Figure 7c. This technique can be used with a large variety of bulk materials as well as films, surfaces and interfaces.

Investigation of Population Dynamics

The dynamic properties of a photogenerated species can be investigated by using a probe wavelength within its absorption or dispersion spectrum. As shown by eqn [4], either ΔA or Δn_p has to be different from zero. For this application, Δn_K and Δn_d should ideally be equal to zero. Unless working with ultrashort pulses (< 1 ps) and a weakly absorbing sample, Δn_K can be neglected. Practically, it is almost impossible to find a sample system for which some fraction of the energy absorbed as light is not released as heat. However, by using a sufficiently small crossing angle of the pump pulses, the formation of the density phase grating, which depends on the acoustic period, can take as much as 30 to 40 ns. In this case, Δn_d is negligible during the first few ns after excitation and the diffracted intensity is due to the population grating only:

$$I_{dif}(t) \propto \sum_i \Delta C_i^2(t) \qquad [14]$$

where ΔC_i is the modulation amplitude of the concentration of every species i whose absorption and/or dispersion spectrum overlaps with the probe wavelength.

The temporal variation of ΔC_i can be due either to the dynamics of the species i in the illuminated grating fringes or to processes taking place between the fringes, such as translational diffusion, excitation transport, and charge diffusion. In liquids, the decay of the population grating by translational diffusion is slow and occurs in the microsecond to millisecond timescale, depending on the fringe spacing. As thermal diffusion is typically hundred times faster, eqn [14] is again valid in this long timescale. Therefore if the population dynamics is very slow, the translational diffusion coefficient of a chemical species can be obtained by measuring the decay of the diffracted intensity as a function of the fringe spacing. This procedure has also been used to determine the temperature of flames. In this case however, the decay of the population grating by translational diffusion occurs typically in the sub-ns timescale.

In the condensed phase, these interfringe processes are of minor importance when measuring the diffracted intensity in the short timescale, i.e., before the formation of the density phase grating. In this case, the transient grating technique is similar to transient absorption, and it thus allows the measurement of population dynamics. However, because holographic detection is background free, it is at least a hundred times more sensitive than transient absorption.

The population gratings are usually probed with monochromatic laser pulses. This procedure is well suited for simple photoinduced processes, such as the decay of an excited state population to the ground state. For example, Figure 8 shows the decay of the diffracted intensity at 532 nm measured after excitation of a cyanine dye at the same wavelength. These dynamics correspond to the ground state recovery of the dye by non-radiative deactivation of the first singlet excited state. Because the diffracted intensity is proportional to the square of the concentration changes (see eqn [14]), its decay time is twice as short as the ground state recovery time. If the reaction is more complex, or if several intermediates are involved, the population grating has to be probed at several wavelengths. Instead of repeating many single wavelength experiments, it is preferable to perform multiplex transient grating. In this case, the grating is probed by white light pulses generated by focusing high intensity fs or ps laser pulses in a dispersive medium. If the crossing angle of the pump pulses is small enough (< 1°), the Bragg angle for probing is almost independent on the wavelength. A transient grating spectrum is obtained by dispersing the diffracted signal in a spectrograph. This spectrum consists in the sum of the square of the transient absorption and transient dispersion spectra. Practically, it is very similar to a transient absorption spectrum, but with a much superior signal to noise ratio. Figure 9 shows the transient grating spectra measured at various time delays after excitation of a solution of chloranil (CA) and

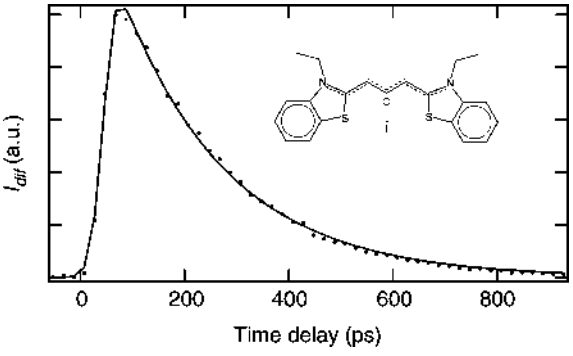

Figure 8 Time profile of the diffracted intensity at 532 nm measured after excitation at the same wavelength of a cyanine dye (inset) in solution and best fit assuming exponential ground state recovery.

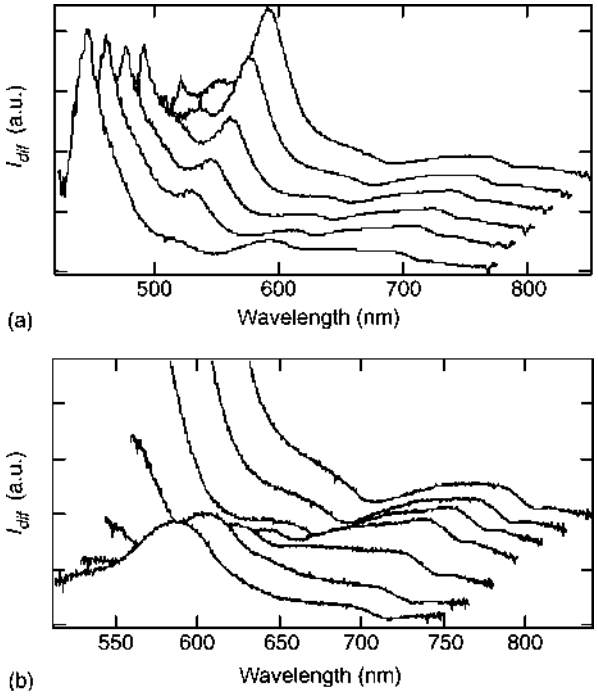

Figure 9 Transient grating spectra obtained at various time delays after excitation at 355 nm of a solution of chloranil and 0.25 M methylnaphthalene (a): from top to bottom: 60, 100, 180, 260, 330 and 600 ps; (b): from top to bottom: 100, 400, 750, 1100, 1500 ps).

methylnaphthalene (MNA) in acetonitrile. The reaction that takes place is

$$^3\mathrm{CA}^* \cdot \mathrm{MNA} \xrightarrow{\text{diffusion}} (^3\mathrm{CA}^* \cdot \mathrm{MNA}) \xrightarrow{\text{ET}}$$
$$^3(\mathrm{CA}^- \cdot \mathrm{MNA}^+) \xrightarrow{\text{MNA}} {}^3(\mathrm{CA}^- \cdot \mathrm{MNA}^+)$$

After its formation upon ultrafast ISC, $^3\mathrm{CA}^*$, which absorbs around 510 nm, decays upon ET with MNA to generate CA^-(450 nm) and MNA^+ (690 nm). The ion pair reacts further with a second MNA molecule to form the dimer cation (580 nm).

The time profile of the diffracted intensity reflects the population dynamics as long as these populations follow first- or pseudo-first order kinetics. Higher order kinetics leads to inharmonic gratings and in this case the time dependence of the diffracted intensity is no longer a direct measure of the population dynamics.

Polarization Selective Transient Grating

In the above applications, the selection between the different contributions to the diffracted signal was essentially made by choosing the probe wavelength and the crossing angle of the pump pulses. However, this approach is not always sufficient. Another important parameter is the polarization of the four waves involved in a transient grating experiment. For example, when measuring population dynamics, the polarization of the probe beam has to be at magic angle (54.7°) relatively to that of the pump pulses. This ensures that the observed time profiles are not distorted by the decay of the orientational anisotropy of the species created by the polarized pump pulses.

The magnitude of ΔA and Δn depends on the pump pulses intensity, and therefore the diffracted intensity can be expressed by using the formalism of nonlinear optics:

$$I_{dif}(t) = C \int_{-\infty}^{+\infty} \mathrm{d}t\, I_{pr}(t - t'') \times \left[\int_{-\infty}^{t} \mathrm{d}t' |R_{ijkl}(t'' - t')| I_{pu}(t') \right]^2 \quad [15]$$

where C is a constant and R_{ijkl} is an element of the fourth rank tensor $\mathbf{R}$ describing the nonlinear response of the material to the applied optical fields. In isotropic media, this tensor has only 21 nonzero elements, which are related as follows:

$$R_{1111} = R_{1122} + R_{1212} + R_{1221} \quad [16]$$

where the subscripts are the Cartesian coordinates. Going from right to left, they design the direction of polarization of the pump, probe, and signal pulses. The remaining elements can be obtained by permutation of these indices ($R_{1122} = R_{1133} = R_{2211} = \ldots$). In a conventional transient grating experiment, the two pump pulses are at the same frequency and are time coincident and therefore their indices can be interchanged. In this case, $R_{1212} = R_{1221}$ and the number of independent tensor elements is further reduced:

$$R_{1111} = R_{1122} + 2R_{1212} \quad [17]$$

The tensor $\mathbf{R}$ can be decomposed into four tensors according to the origin of the sample response: population and density changes, electronic and nuclear optical Kerr effects:

$$\mathbf{R} = \mathbf{R}(p) + \mathbf{R}(d) + \mathbf{R}(K, e) + \mathbf{R}(K, n) \quad [18]$$

As they describe different phenomena, these tensors do not have the same symmetry properties. Therefore, the various contributions to the diffracted intensity can, in some cases, be measured selectively by choosing the appropriate polarization of the four waves.

Table 2 shows the relative amplitude of the most important elements of these four tensors.

The tensor $\mathbf{R}(p)$ depends on the polarization anisotropy of the sample, r, created by excitation with polarized pump pulses:

$$r(t) = \frac{N_{\parallel}(t) - N_{\perp}(t)}{N_{\parallel}(t) + 2N_{\perp}(t)} = \frac{2}{5}\mathrm{P}_2[\cos(\gamma)]\exp(-k_{re}t) \qquad [19]$$

where $N_{\parallel}$ and $N_{\perp}$ are the number of molecules with the transition dipole oriented parallel and perpendicular to the polarization of the probe pulse, respectively, γ is the angle between the transition dipoles involved in the pump and probe processes, P_2 is the second Legendre polynomial, and k_{re} is the rate constant for the reorientation of the transition dipole, for example, by rotational diffusion of the molecule or by energy hopping.

The table shows that $R_{1212}(d) = 0$, i.e., the contribution of the density grating can be eliminated with the set of polarization $(0°, 90°, 0°, 90°)$, the so-called crossed grating geometry. In this geometry, the two pump pulses have orthogonal polarization, the polarization of the probe pulse is parallel to that of one pump pulse and the polarization component of the signal that is orthogonal to the polarization of the probe pulse is measured. In this geometry, $R_{1212}(p)$ is nonzero as long as there is some polarization anisotropy, ($r \neq 0$). In this case, the diffracted intensity is

$$I_{dif}(t) \propto |R_{1212}(p)|^2 \propto [\Delta C(t) \times r(t)]^2 \qquad [20]$$

The crossed grating technique can thus be used to investigate the reorientational dynamics of molecules, through $r(t)$, especially when the dynamics of r is faster than that of ΔC. For example, Figure 10 shows the time profile of the diffracted intensity measured in the crossed grating geometry with rhodamine 6G in ethanol. The decay is due to the reorientation of the molecule by rotational diffusion, the excited lifetime of rhodamine being about 4 ns.

On the other hand, if the decay of r is slower than that of ΔC, the crossed grating geometry can be used to measure the population dynamics without any interference from the density phase grating. For example, Figure 11 shows the time profile of the diffracted intensity after excitation of a suspension of TiO_2 particles in water. The upper profile was measured with the set of polarization $(0°, 0°, 0°, 0°)$ and thus reflects the time dependence of $R_{1111}(p)$ and $R_{1111}(d)$. $R_{1111}(p)$ is due to the trapped electron population, which decays by charge recombination, and $R_{1111}(d)$ is due to the heat dissipated upon both charge separation and recombination. The lower time profile was measured in the crossed grating geometry and thus reflects the time dependence of R_{1212} and in particular that of $R_{1212}(p)$.

Each contribution to the signal can be selectively eliminated by using the set of geometry $(\zeta, 45°, 0°, 0°)$ where the value of the 'magic angle' ζ for each

Table 2 Relative value of the most important elements of the response tensors $\mathbf{R}(i)$ and polarization angle ζ of the signal beam, where the contribution of the corresponding process vanishes for the set of polarization $(\zeta, 45°, 0°, 0°)$.

Process	R_{1111}	R_{1122}	R_{1212}	ζ
Electronic OKE	1	1/3	1/3	−71.6°
Nuclear OKE	1	−1/2	3/4	63.4°
Density	1	1	0	−45°
Population:				
$\gamma = 0°$	1	1/3	1/3	−71.6°
$\gamma = 90°$	1	2	−1/2	−26.6°
No correlation: $r = 0$	1	1	0	−45°

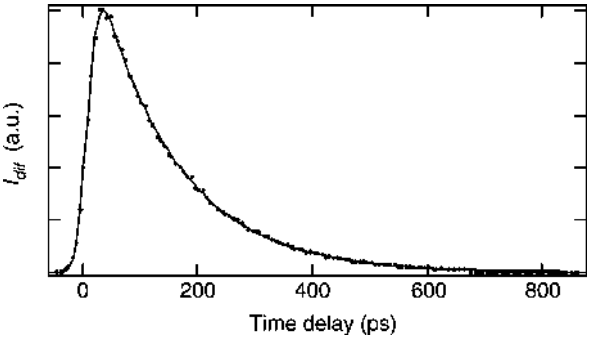

Figure 10 Time profile of the diffracted intensity measured with a solution of rhodamine 6G with crossed grating geometry.

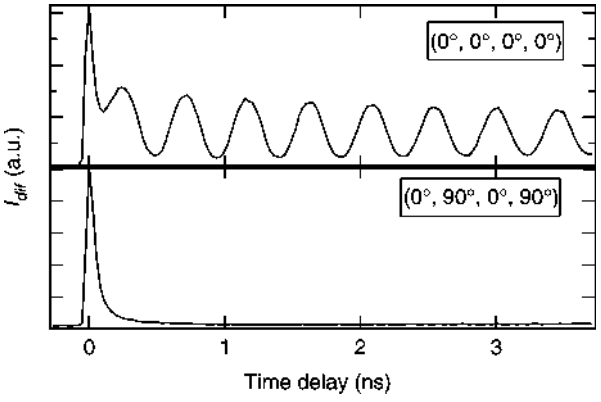

Figure 11 Time profiles of the diffracted intensity after excitation at 355 nm of a suspension of TiO_2 particles in water and using different set of polarization of the four beams.

contribution is listed in Table 2. This approach allows for example the nuclear and electronic contributions to the optical Kerr effect to be measured separately.

Concluding Remarks

The transient grating techniques offer a large variety of applications for investigating the dynamics of chemical processes. We have only discussed the cases where the pump pulses are time coincident and at the same wavelength. Excitation with pump pulses at different wavelengths results to a moving grating. The well-known CARS spectroscopy is such a moving grating technique. Finally, the three pulse photon echo can be considered as a special case of transient grating where the sample is excited by two pump pulses, which are at the same wavelength but are not time coincident.

List of Units and Nomenclature

C	concentration [mol L^{-1}]
C_v	heat capacity [J K^{-1} kg^{-1}]
d	sample thickness [m]
D_{th}	thermal diffusivity [m^2 s^{-1}]
I_{dif}	diffracted intensity [W m^{-2}]
I_{pr}	intensity of the probe pulse [W m^{-2}]
I_{pu}	intensity of a pump pulse [W m^{-2}]
k_{ac}	acoustic wavevector [m^{-1}]
k_r	rate constant of a heat releasing process [s^{-1}]
k_{re}	rate constant of reorientation [s^{-1}]
k_{th}	rate constant of thermal diffusion [s^{-1}]
K	attenuation constant
n	refractive index
$\tilde{n}$	complex refractive index
N	number of molecule per unit volume [m^{-3}]
r	polarization anisotropy
$\mathbf{R}$	fourth rank response tensor [m^2 V^{-2}]
v_s	speed of sound [m s^{-1}]
V	volume [m^3]
α_{ac}	acoustic attenuation constant [m^{-1}]
β	cubic volume expansion coefficient [K^{-1}]
γ	angle between transition dipoles
Λ	fringe spacing [m]
Δn_d	variation of refractive index due to density changes
Δn_d^e	Δn_d due to electrostriction
Δn_K	variation of refractive index due to optical Kerr effect
Δn_d^p	Δn_d due to volume changes
Δn_p	variation of refractive index due to population changes
Δn_d^t	Δn_d due to temperature changes
Δx	peak to null variation of the parameter x
ε	molar decadic absorption coefficient [cm L mol^{-1}]
ζ	angle of polarization of the diffracted signal
η	diffraction efficiency
θ_B	Bragg angle (angle of incidence of the probe pulse)
θ_{pu}	angle of incidence of a pump pulse
λ_{pr}	probe wavelength [m]
λ_{pu}	pump wavelength [m]
ρ	density [kg m^{-3}]
τ_{ac}	acoustic period [s]
υ_{ac}	acoustic frequency [s^{-1}]

See also

Chemical Applications of Lasers: Nonlinear Spectroscopies; Pump and Probe Studies of Femtosecond Kinetics. **Holography, Applications:** Holographic Recording Materials and their Processing. **Holography, Techniques:** Holographic Interferometry. **Materials Characterization Techniques:** $\chi^{(3)}$. **Nonlinear Optics Basics:** Four-Wave Mixing. **Ultrafast Technology:** Femtosecond Chemical Dynamics, Gas Phase; Femtosecond Condensed Phase Spectroscopy – Structural Dynamics.

Further Reading

Bräuchle C and Burland DM (1983) Holographic methods for the investigation of photochemical and photophysical properties of molecules. *Angewandte Chemie International Edition Engl* 22: 582–598.

Eichler HJ, Günter P and Pohl DW (1986) *Laser-Induced Dynamics Gratings*. Berlin: Springer.

Fleming GR (1986) *Chemical Applications of Ultrafast Spectroscopy*. Oxford: Oxford University Press.

Fourkas JT and Fayer MD (1992) The transient grating: a holographic window to dynamic processes. *Accounts of Chemical Research* 25: 227–233.

Hall G and Whitaker BJ (1994) Laser-induced grating spectroscopy. *Journal of the Chemistry Society Faraday Trans* 90: 1–16.

Levenson MD and Kano SS (1987) *Introduction to Nonlinear Laser Spectroscopy*, revised edn. Boston: Academic Press.

Mukamel S (1995) *Nonlinear Optical Spectroscopy*. Oxford: Oxford University Press.

Rullière C (ed.) (1998) *Femtosecond Laser Pulses*. Berlin: Springer.

Terazima M (1998) Photothermal studies of photophysical and photochemical processes by the transient grating method. *Advances in Photochemistry* 24: 255–338.

cross-sectional area 10^4 times and dispersed the beam across it. The CPA technique makes it possible to use conventional laser amplifiers and to stay below the onset of nonlinear effects.

See also

Diffraction: Diffraction Gratings. **Ultrafast Laser Techniques:** Pulse Characterization Techniques.

COHERENCE

Contents

Overview
Coherence and Imaging
Speckle and Coherence

Overview

A Sharma and A K Ghatak, Indian Institute of Technology, New Delhi, India

H C Kandpal, National Physical Laboratory, New Delhi, India

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Coherence refers to the characteristic of a wave that indicates whether different parts of the wave oscillate in tandem, or in a definite phase relationship. In other words, it refers to the degree of confidence by which one can predict the amplitude and phase of a wave at a point, from the knowledge of these at another point at the same or a different instant of time. Emission from a thermal source, such as a light bulb, is a highly disordered process and the emitted light is incoherent. A well-stabilized laser source, on the other hand, generates light in a highly ordered manner and the emitted light is highly coherent. Incoherent and coherent light represent two extreme cases. While describing the phenomena of physical optics and diffraction theory, light is assumed to be perfectly coherent in both spatial as well as temporal senses, whereas in radiometry it is generally assumed to be incoherent. However, practical light sources and the fields generated by them are in between the two extremes and are termed as partially coherent sources and fields. The degree of order that exists in an optical field produced by a source of any kind may be described in terms of various correlation functions. These correlation functions are the basic theoretical tools for the analyses of statistical properties of partially coherent light fields.

Light fields generated from real physical sources fluctuate randomly to some extent. On a microscopic level quantum mechanical fluctuations produce randomness and on macroscopic level the randomness occurs as a consequence of these microscopic fluctuations, even in free space. In real physical sources, spontaneous emission causes random fluctuations and even in the case of lasers, spontaneous emission cannot be suppressed completely. In addition to spontaneous emission, there are many other processes that give rise to random fluctuations of light fields. Optical coherence theory was developed to describe the random nature of light and it deals with the statistical similarity between light fluctuations at two (or more) space–time points.

As mentioned earlier, in developing the theory of interference or diffraction, light is assumed to be perfectly coherent, or, in other words, it is taken to be monochromatic and sinusoidal for all times. This is, however, an idealization since the wave is obviously generated at some point of time by an atomic or molecular transition. Furthermore, a wavetrain generated by such a transition is of a finite duration, which is related to the finite lifetime of the atomic or molecular levels involved in the transition. Thus, any wave emanating from a source is an ensemble of a large number of such wavetrains of finite duration, say τ_c. A simplified visualization of such an ensemble is shown in Figure 1 where a wave is shown as a series of wavetrains of duration τ_c. It is evident from the figure that the fields at time t and $t + \Delta t$ will have a definite phase relationship if $\Delta t \ll \tau_c$ and will have no or negligible phase relationship when $\Delta t \gg \tau_c$. The time τ_c is known as the coherence time of the

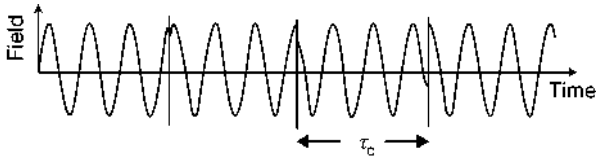

Figure 1 Typical variation of the radiation field with time. The coherence time $\sim \tau_c$.

radiation and the field is said to remain coherent for time $\sim \tau_c$. This property of waves is referred to as the time coherence or the temporal coherence and is related to the spectral purity of the radiation. If one obtains the spectrum of the wave shown in Figure 1 by taking the Fourier transform of the time variation, it would have a width of $\Delta\nu$ around ν_0 which is the frequency of the sinusoidal variation of individual wavetrains. The spectral width $\Delta\nu$ is related to the coherence time as

$$\Delta\nu \sim 1/\tau_c \qquad [1]$$

For thermal sources such as a sodium lamp, $\tau_c \sim 100$ ps, whereas for a laser beam it could be as large as a few milliseconds. A related quantity is the coherence length l_c, which is the distance covered by the wave in time τ_c,

$$l_c = c\tau_c \sim \frac{c}{\Delta\nu} = \frac{\lambda_0^2}{\Delta\lambda} \qquad [2]$$

where λ_0 is the central wavelength ($\lambda_0 = c/\nu_0$) and $\Delta\lambda$ is the wavelength spread corresponding to the spectral width $\Delta\nu$. In a two-beam interference experiment (e.g., Michelson interferometer, Figure 2), the interfering beam derived from the same source will produce good interference fringes if the path difference between the two interfering beams is less than the coherence length of the radiation given out by the source.

It must be added here that for the real fields, generated by innumerable atoms or molecules, the individual wavetrains have different lengths around an average value τ_c. Furthermore, several wavetrains in general are propagating simultaneously, overlapping in space and time, to produce an ensemble whose properties are best understood in a statistical sense as we shall see in later sections.

Another type of coherence associated with the fields is the space coherence or the spatial coherence, which is related to the size of the source of radiation. It is evident that when the source is an ideal point source, the field at any two points (within the coherence length) would have definite phase relationship. However, the field from a thermal source of finite area S can be thought as resultant of the fields

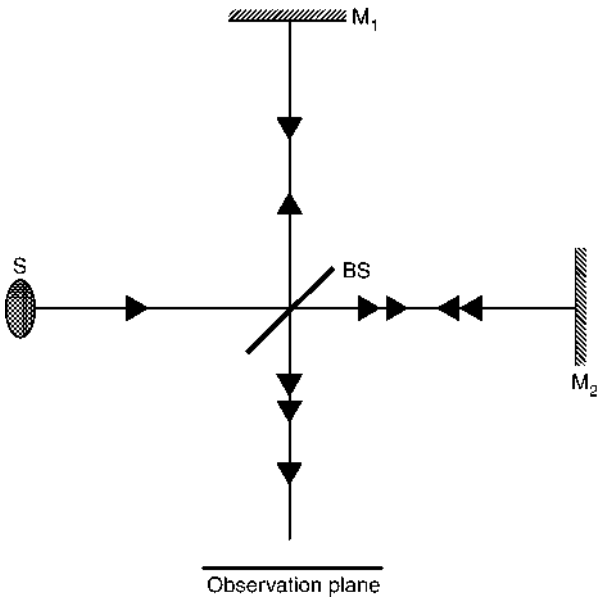

Figure 2 Michelson interferometer to study the temporal coherence of radiation from source S; M_1 and M_2 are mirrors and BS is a 50%–50% beamsplitter.

from each point on the source. Since each point source would usually be independent of the other, the phase relationship between the fields at any two points would depend on their position and size of the source. It can be shown that two points will have a strong phase relationship if they lie within the solid angle $\Delta\Omega$ from the source (Figure 3) such that

$$\Delta\Omega \sim \frac{\lambda_0^2}{S} \qquad [3]$$

Thus, on a plane R distance away from the source, one can define an area $A_c = R^2\Delta\Omega$ over which the field remains spatially coherent. This area A_c is called the coherence area of the radiation and its square root is sometimes referred to as the transverse coherence length. It is trivial to show that

$$A_c \sim \frac{\lambda_0^2}{\Delta\Omega_S} \qquad [4]$$

where $\Delta\Omega_S$ is the solid angle subtended by the source on the plane at which the coherence area is defined. In Young's double-hole experiment, if the two pinholes lie within the coherence area of the radiation from the primary source, good contrast in fringes would be observed.

Combining the concepts of coherence length and the coherence area, one can define a coherence volume as $V_c = A_c l_c$. For the wavefield from a thermal source, this volume represents that portion of the space in which the field is coherent and any

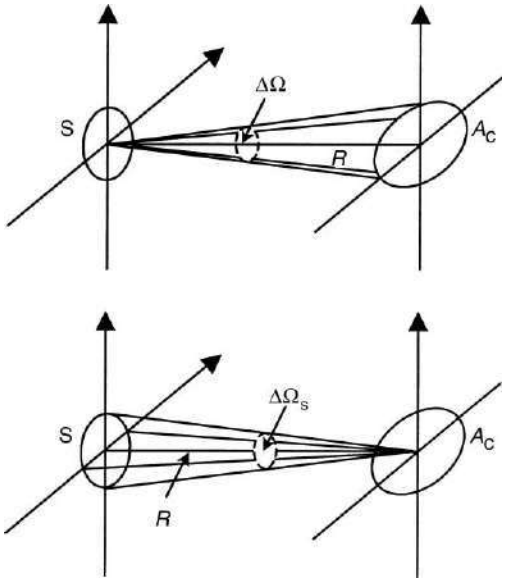

Figure 3 Spatial coherence and coherence area.

interference produced using the radiation from points within this volume will produce fringes of good contrast.

Mathematical Description of Coherence

Analytical Field Representation

Coherence properties associated with fields are best analyzed in terms of complex representation for optical fields. Let the real function $V^{(r)}(\mathbf{r}, t)$ represent the scalar field, which could be one of the transverse Cartesian components of the electric field associated with the electromagnetic wave. One can then define a complex analytical signal $V(\mathbf{r}, t)$ such that

$$V^{(r)}(\mathbf{r}, t) = \mathrm{Re}[V(\mathbf{r}, t)], \qquad V(\mathbf{r}, t) = \int_0^\infty \tilde{V}^{(r)}(\mathbf{r}, \nu)\mathrm{e}^{-2\pi i\nu t}\,\mathrm{d}\nu \qquad [5]$$

where the spectrum $\tilde{V}^{(r)}(\mathbf{r}, \nu)$ is the Fourier transform of the scalar field $V^{(r)}(\mathbf{r}, t)$ and the spectrum for negative frequencies has been suppressed as it does not give any new information since $\tilde{V}^{(r)^*}(\mathbf{r}, \nu) = \tilde{V}^{(r)}(\mathbf{r}, -\nu)$.

In general, the radiation from a quasi-monochromatic thermal source fluctuates randomly as it is made of a large number of mutually independent contributions from individual atoms or molecules in the source. The field from such a source can be regarded as an ensemble of a large number of randomly different analytical signals such as $V(\mathbf{r}, t)$. In other words, $V(\mathbf{r}, t)$ is a typical member of an ensemble, which is the result of a random process representing the radiation from a quasi-monochromatic source. This process is assumed to be stationary and ergodic so that the ensemble average is equal to the time average of a typical member of the ensemble and that the origin of time is unimportant. Thus, the quantities of interest in the theory of coherence are defined as time averages:

$$\langle f(t)\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)\mathrm{d}t \qquad [6]$$

Mutual Coherence

In order to define the mutual coherence, the key concept in the theory of coherence, we consider Young's interference experiment, as shown in Figure 4, where the radiation from a broad source of size S is illuminating a screen with two pinholes P_1 and P_2. The light emerging from the two pinholes produces interference, which is observed on another screen at a distance R from the first screen. The field at point P, due to the pinholes, would be $K_1 V(\mathbf{r}_1, t - t_1)$ and $K_2 V(\mathbf{r}_2, t - t_2)$ respectively, where $\mathbf{r}_1$ and $\mathbf{r}_2$ define the positions of P_1 and P_2, t_1 and t_2 are the times taken by the light to travel to P from P_1 and P_2, and K_1 and K_2 are imaginary constants that depend on the geometry and size of the respective pinhole and its distance from the point P. Thus, the resultant field at point P would be given by

$$V(\mathbf{r}, t) = K_1 V(\mathbf{r}_1, t - t_1) + K_2 V(\mathbf{r}_2, t - t_2) \qquad [7]$$

Since the optical periods are extremely small as compared to the response time of a detector, the detector placed at point P will record only the time-averaged intensity:

$$I(P) = \langle V^*(\mathbf{r}, t)V(\mathbf{r}, t)\rangle \qquad [8]$$

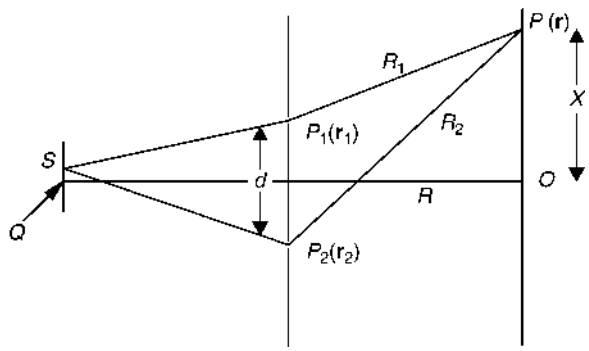

Figure 4 Schematic of Young's double-hole experiment.

where some constants have been ignored. With eqn [7], the intensity at point P would be given by

$$I(P) = I_1 + I_2 + 2\,\mathrm{Re}\{K_1 K_2^* \Gamma(\mathbf{r}_1, \mathbf{r}_2, t - t_1, t - t_2)\} \quad [9]$$

where I_1 and I_2 are the intensities at point P, respectively, due to radiations from pinholes P_1 and P_2 independently (defined as $I_i = \langle |K_i V(\mathbf{r}_i, t)|^2 \rangle$) and

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, t - t_1, t - t_2) \equiv \Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle V^*(\mathbf{r}_1, t) V(\mathbf{r}_2, t + \tau) \rangle \quad [10]$$

is the mutual coherence function of the fields at P_1 and P_2, and it depends on the time difference $\tau = t_2 - t_1$, since the random process is assumed to be stationary. This function is also sometimes denoted by $\Gamma_{12}(\tau)$. The function $\Gamma_{ii}(\tau) \equiv \Gamma(\mathbf{r}_i, \mathbf{r}_i, \tau)$ defines the self-coherence of the light from the pinhole at P_i, and $|K_i|^2 \Gamma_{ii}(0)$ defines the intensity I_i at point P, due to the light from pinhole P_i.

Degree of Coherence and the Visibility of Interference Fringes

One can define a normalized form of the mutual coherence function, namely:

$$\gamma_{12}(\tau) \equiv \gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}} = \frac{\langle V^*(\mathbf{r}_1, t) V(\mathbf{r}_2, t + \tau) \rangle}{\left[\langle |V(\mathbf{r}_1, t)|^2 \rangle \langle |V(\mathbf{r}_2, t)|^2 \rangle\right]^{1/2}} \quad [11]$$

which is called the complex degree of coherence. It can be shown that $0 \le |\gamma_{12}(\tau)| \le 1$. The intensity at point P, given by eqn [9], can now be written as

$$I(P) = I_1 + I_2 + 2\sqrt{I_1 I_2}\,\mathrm{Re}[\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)] \quad [12]$$

Expressing $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ as

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = |\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \exp[\mathrm{i}\alpha(\mathbf{r}_1, \mathbf{r}_2, \tau) - 2\pi \mathrm{i} \nu_0 \tau] \quad [13]$$

where $\alpha(\vec{r}_1, \vec{r}_2, \tau) = \arg[\gamma(\vec{r}_1, \vec{r}_2, \tau)] + 2\pi\nu_0\tau$, and ν_0 is the mean frequency of the light, the intensity in eqn [12] can be written as

$$I(P) = I_1 + I_2 + 2\sqrt{I_1 I_2}|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \times \cos[\alpha(\mathbf{r}_1, \mathbf{r}_2, \tau) - 2\pi\nu_0\tau] \quad [14]$$

Now, if we assume that the source is quasi-monochromatic, i.e., its spectral width $\Delta\nu \ll \nu_0$, the quantities $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ and $\alpha(\mathbf{r}_1, \mathbf{r}_2, \tau)$ vary slowly on the observation screen, and the interference fringes are mainly obtained due to the cosine term. Thus, defining the visibility of fringes as $\upsilon = (I_{\max} - I_{\min})/(I_{\max} + I_{\min})$, we obtain

$$\upsilon = \frac{2(I_1 I_2)^{1/2}}{I_1 + I_2}|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \quad [15]$$

which shows that for maximum visibility, the two interfering fields must be completely coherent ($|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| = 1$). On the other hand, if the fields are completely incoherent ($|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| = 0$), no fringes are observed ($I_{\max} = I_{\min}$). The fields are said to be partially coherent when $0 < |\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| < 1$. When $I_1 = I_2$, the visibility is the same as $|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)|$. The relation in eqn [15] shows that in an interference experiment, one can obtain the modulus of the complex degree of coherence by measuring I_1, I_2, and the visibility. Similarly, eqn [14] shows that from the measurement the positions of maxima, one can obtain the phase of the complex degree of coherence, $\alpha(\mathbf{r}_1, \mathbf{r}_2, \tau)$.

Temporal and Spatial Coherence

If the source illuminating the pinholes is a point source of finite spectral width situated at point Q, the fields at point P_1 and P_2 (Figure 4) at any given instant are the same and the mutual coherence function becomes

$$\Gamma_{11}(\tau) = \Gamma(\mathbf{r}_1, \mathbf{r}_1, \tau) = \langle V^*(\mathbf{r}_1, t) V(\mathbf{r}_1, t + \tau) \rangle = \langle V^*(\mathbf{r}_2, t) V(\mathbf{r}_2, t + \tau) \rangle = \Gamma_{22}(\tau) \quad [16]$$

The self-coherence, $\Gamma_{11}(\tau)$, of the light from pinhole P_1, is a direct measure of the temporal coherence of the source. On the other hand, if the source is of finite size and we observe the interference at point O which corresponds to $\tau = 0$, the mutual coherence function would be

$$\Gamma_{12}(0) = \Gamma(\mathbf{r}_1, \mathbf{r}_2, 0) = \langle V^*(\mathbf{r}_1, t) V(\mathbf{r}_2, t) \rangle \equiv J_{12} \quad [17]$$

which is called the mutual intensity and is a direct measure of the spatial coherence of the source. In general, however, the function $\Gamma_{12}(\tau)$ measures, for a source of finite size and spectral width, a combination of the temporal and spatial coherence, and in only some limiting cases, the two types of coherence can be separated.

Spectral Representation of Mutual Coherence

One can also analyze the correlation between two fields in the spectral domain. In particular, one can define the cross-spectral density function $W(\mathbf{r}_1, \mathbf{r}_2, \nu)$ which defines the correlation between the amplitudes of the spectral components of frequency ν of the light

at the points P_1 and P_2. Thus:

$$W(\mathbf{r}_1, \mathbf{r}_2, \nu)\delta(\nu - \nu') = \langle V^*(\mathbf{r}_1, \nu)V(\mathbf{r}_2, \nu')\rangle \qquad [18]$$

Using the generalized Wiener–Khintchine theorem, the cross-spectral density function can be shown to be the Fourier transform of the mutual coherence function:

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \int_0^\infty W(\mathbf{r}_1, \mathbf{r}_2, \nu)\mathrm{e}^{-2\pi \mathrm{i}\nu t}\,\mathrm{d}\nu \qquad [19]$$

$$W(\mathbf{r}_1, \mathbf{r}_2, \nu) = \int_{-\infty}^\infty \Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)\mathrm{e}^{2\pi \mathrm{i}\nu t}\,\mathrm{d}\tau \qquad [20]$$

If the two points P_1 and P_2 coincide (i.e., there is only one pinhole), the cross-spectral density function reduces to the spectral density function of the light, which we denote by $S(\mathbf{r}, \nu)[\equiv W(\mathbf{r}, \mathbf{r}, \nu)]$. Thus, it follows from eqn [20] that the spectral density of light is the inverse Fourier transform of the self-coherence function. This leads to the Fourier transform spectroscopy, as we shall see later. One can also define spectral degree of coherence at frequency ν as

$$\mu(\mathbf{r}_1, \mathbf{r}_2, \nu) = \frac{W(\mathbf{r}_1, \mathbf{r}_2, \nu)}{\sqrt{W(\mathbf{r}_1, \mathbf{r}_1, \nu)W(\mathbf{r}_2, \mathbf{r}_2, \nu)}} = \frac{W(\mathbf{r}_1, \mathbf{r}_2, \nu)}{\sqrt{S(\mathbf{r}_1, \nu)S(\mathbf{r}_2, \nu)}} \qquad [21]$$

It is easy to see that $0 \le |\mu(\mathbf{r}_1, \mathbf{r}_2, \nu)| \le 1$. It is also sometimes referred to as the complex degree of coherence at frequency ν. It may be noted here that in the literature the notation $G^{(1,1)}(\nu) \equiv G(\mathbf{r}_1, \mathbf{r}_2, \gamma)$ has also been used for $W(\mathbf{r}_1, \mathbf{r}_2, \nu)$.

Propagation of Coherence

The van Cittert–Zernike Theorem

Perfectly coherent waves propagate through diffraction formulae, which have been discussed in this volume elsewhere. However, incoherent and partially coherent waves would evidently propagate somewhat differently. The earliest treatment of propagation noncoherent light is due to van Cittert, which was later generalized by Zernike to obtain what is now the van Cittert–Zernike theorem. The theorem deals with the correlations developed between fields at two points, which have been generated by a quasi-monochromatic and spatially incoherent planar source. Thus, as shown in Figure 5, we consider a quasi-monochromatic ($\Delta\nu \ll \nu_0$) planar source σ, which has an intensity distribution $I(\mathbf{r}')$ on its plane and is spatially incoherent, i.e., there is no correlation between the fields at any two points on the source.

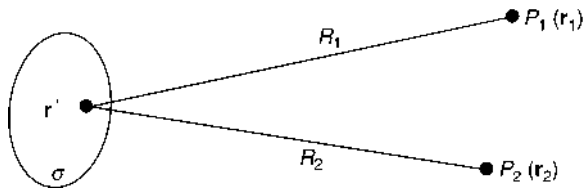

Figure 5 Geometry for the van Cittert–Zernike theorem.

The field, due to this source, would develop finite correlations after propagation, and the theorem states that

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, 0) = \frac{\Gamma_{12}(0)}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{12}(0)}} = \frac{1}{\sqrt{I(\mathbf{r}_1)I(\mathbf{r}_2)}}\int\int_\sigma I(\mathbf{r}')\frac{\mathrm{e}^{2\pi \mathrm{i}\nu_0(R_2-R_1)/c}}{R_1R_2}\mathrm{d}^2r' \qquad [22]$$

where $R_i = |\mathbf{r}_i - \mathbf{r}'|$ and $I(\mathbf{r}_i)$ is the intensity at point P_i. This relation is similar to the diffraction pattern produced at point P_1 due to a wave, with a spherical wavefront converging towards P_2 and with an amplitude distribution $I(\mathbf{r}')$ when it is diffracted by an aperture of the same shape, size, and the intensity distribution as those of the incoherent source σ. Thus, the theorem shows that a radiation, which was incoherent at the source, becomes partially coherent as it propagates.

Generalized Propagation

The expression $\mathrm{e}^{-2\pi \mathrm{i}\nu_0 R_1/c}/R_1$ can be interpreted as the field obtained at P_1 due to a point source located at the point $\mathbf{r}'$ on the planar source. Thus, this expression is simply the point spread function of the homogeneous space between the source and the observation plane containing the points P_1 and P_2. Hopkins generalized this to include any linear optical system characterized by a point spread function $h(\mathbf{r}_j - \mathbf{r}')$ and obtained the formula for the complex degree of coherence of a radiation emerging from an incoherent, quasi-monochromatic planar source after it has propagated through such a linear optical system:

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, 0) = \frac{1}{\sqrt{I(\mathbf{r}_1)I(\mathbf{r}_2)}}\int\int_\sigma I(\mathbf{r}')h(\mathbf{r}_1 - \mathbf{r}')h^*(\mathbf{r}_2 - \mathbf{r}')\mathrm{d}^2r' \qquad [23]$$

It would thus seem that the correlations propagate in much the same way, as does the field. Indeed, Wolf showed that the mutual correlation function

$\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)$ satisfies the wave equations:

$$\nabla_1^2\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=\frac{1}{c^2}\frac{\partial^2}{\partial\tau^2}\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)$$
$$\nabla_2^2\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=\frac{1}{c^2}\frac{\partial^2}{\partial\tau^2}\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau) \qquad [24]$$

where ∇_i^2 is the Laplacian with respect to the point $\mathbf{r}_i$. Here the first of eqns [24], for instance, gives the variation of the mutual coherence function with respect to $\mathbf{r}_1$ and τ for fixed $\mathbf{r}_2$. Further, the variable τ is the time difference defined through the path difference, since $\tau=(R_2-R_1)/c$, and the actual time does not affect the mutual coherence function (as the fields are assumed to be stationary).

Using the relation in eqn [20], one can also obtain from eqn [24], the propagation equations for the cross-spectral density function $W(\mathbf{r}_1,\mathbf{r}_2,\nu)$:

$$\nabla_1^2 W(\mathbf{r}_1,\mathbf{r}_2,\nu)+k^2W(\mathbf{r}_1,\mathbf{r}_2,\nu)=0$$
$$\nabla_2^2 W(\mathbf{r}_1,\mathbf{r}_2,\nu)+k^2W(\mathbf{r}_1,\mathbf{r}_2,\nu)=0 \qquad [25]$$

where $k=2\pi\nu/c$.

Thompson and Wolf Experiment

One of the most elegant experiments for studying various aspects of coherence theory was carried out by Thompson and Wolf by making slight modifications in the Young's double-hole experiment. The experimental set-up shown in Figure 6 consists of a quasi-monochromatic broad incoherent source S of diameter $2a$. This was obtained by focusing filtered narrow band light from a mercury lamp (not shown in the figure) on to a hole of size $2a$ in an opaque screen A. A mask consisting of two pinholes, each of diameter $2b$ with their axes separated by a distance d, was placed symmetrically about the optical axis of the experimental setup at plane B between two lenses L_1 and L_2, each having focal length f. The source was at the front focal plane of the lens L_1 and the observations were made at the plane C at the back focus of the lens L_2. The separation d was varied to study the spatial coherence function on the mask plane by measuring the visibility and the phase of the fringes formed in plane C.

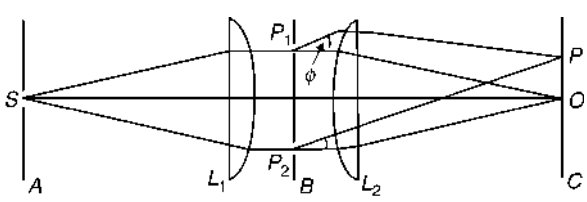

Figure 6 Schematic of the Thompson and Wolf experiment.

Using the van Cittert–Zernike theorem, the complex degree of coherence at the pinholes P_1 and P_2 on the mask after the lens L_1 is

$$\gamma_{12}=|\gamma_{12}|e^{i\beta_{12}}=\frac{2J_1(v)}{v}\ \text{with}\ v=\left(\frac{2\pi\nu ad}{cf}\right) \qquad [26]$$

for symmetrically placed pinholes about the optical axis. Here β_{12} is the phase of the complex degree of coherence and in this special case where the two holes are equidistant from the axis, β_{12} is either zero or π, respectively, for positive or negative values of $2J_1(v)/v$. Let us assume that the intensities at two pinholes P_1 and P_2 are equal. The interference pattern observed at the back focal plane of lens L_2 is due to the superposition of the light diffracted from the pinholes. The beams are partially coherent with degree of coherence given by eqn [26]. Since the pinholes are symmetrically placed, the intensity due to either of the pinholes at a point P at the back focal plane of the lens L_2 is the same and is given by the Fraunhofer formula for diffraction from circular apertures, i.e.:

$$I_1(P)=I_2(P)=\left|\frac{2J_1(u)}{u}\right|^2,\ \text{with}\ u=\frac{2\pi\nu}{c}b\sin\phi \qquad [27]$$

and ϕ is the angle that the diffracted beam makes from normal to the plane of the pinholes. The intensity of the interference pattern produced at the back focal plane of the lens L_2 is:

$$I(P)=2I_1(P)\left[1+\left|\frac{2J_1(v)}{v}\right|^2\cos(\delta+\beta_{12})\right] \qquad [28]$$

where $\delta=d\sin\phi$ is the phase difference between the two beams reaching P from P_1 and P_2. For the on-axis point O, the quantity δ is zero.

The maximum and minimum values of $I(P)$ are given by

$$I_{\max}(P)=2I_1(P)\left[1+|2J_1(v)/v|^2\right] \qquad [29(a)]$$

$$I_{\max}(P)=2I_1(P)\left[1-|2J_1(v)/v|^2\right] \qquad [29(b)]$$

Figure 7 shows an example of the observed fringe patterns obtained by Thompson (1958) in a subsequent experiment.

Types of Fields

As mentioned above, $\gamma_{12}(\tau)$ is a measure of the correlation of the complex field at any two points P_1

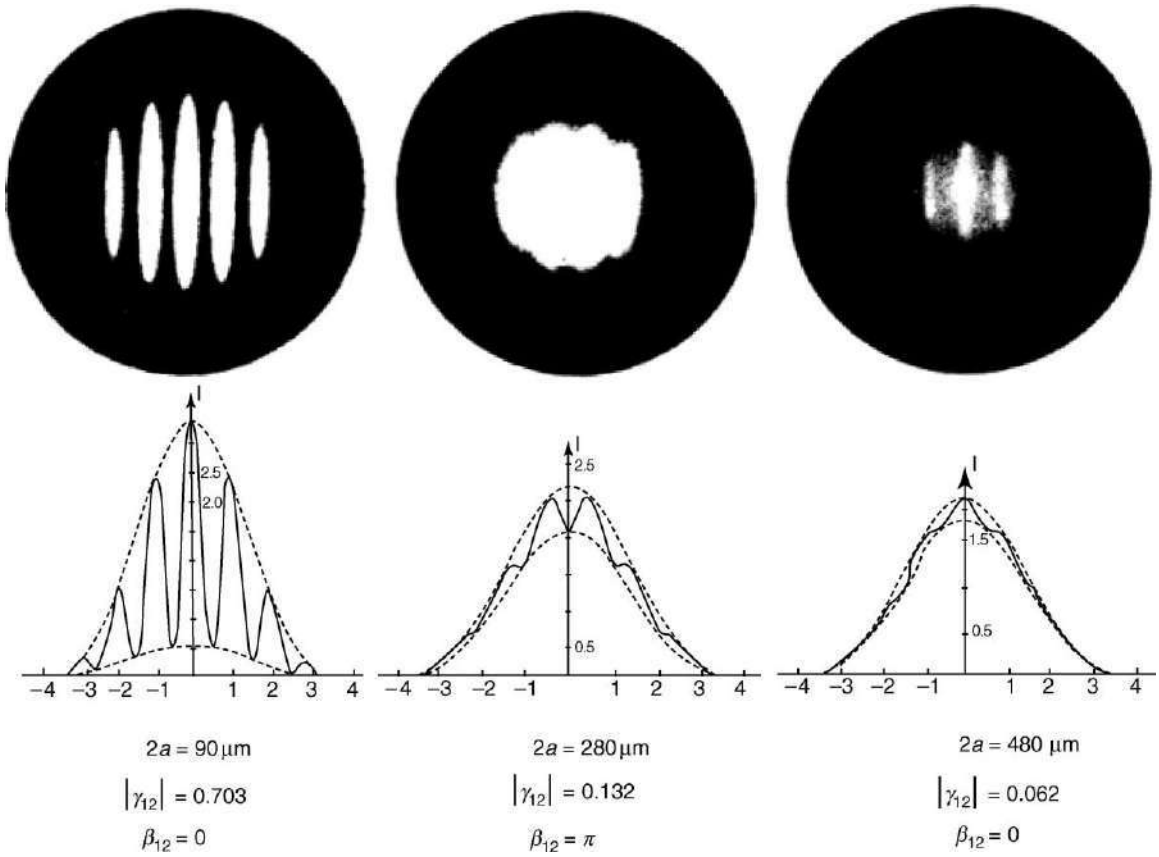

Figure 7 Two beam interference patterns obtained by using partially coherent light. The wavelength of light used was 0.579 μm and the separation *d* of the pinholes in the screen at B was 0.5 cm. The figure shows the observed fringe pattern and the calculated intensity variation for three sizes of the secondary source. The dashed lines show the maximum and minimum intensity. The values of the diameter, 2*a*, of the source and the corresponding values of the magnitude, $|\gamma_{12}|$, and the phase, β_{12}, are also shown in each case. Reproduced with permission from Thompson BJ (1958) Illustration of phase change in two-beam interference with partially coherent light. *Journal of the Optical Society of America* 48: 95–97.

and P_2 at specific time delay τ. Extending this definition, an optical field may be coherent or incoherent, if $|\gamma_{12}(\tau)| = 1$ or $|\gamma_{12}(\tau)| = 0$, respectively, for all pairs of points in the field and for all time delays. In the following, we consider some specific cases of fields and their properties.

Perfectly Coherent Fields

A field would be termed as perfectly coherent or self-coherent at a fixed point, if it has the property that $|\gamma(\mathbf{R}, \mathbf{R}, \tau)| = 1$ at some specific point $\mathbf{R}$ in the domain of the field for all values of time delay τ. It can also be shown that $\gamma(\mathbf{R}, \mathbf{R}, \tau)$ is periodic in time, i.e.:

$$\gamma(\mathbf{R}, \mathbf{R}, \tau) = \mathrm{e}^{-2\pi \mathrm{i} \nu_0 \tau} \quad \text{for} \quad \nu_0 > 0 \qquad [30]$$

For any other point $\mathbf{r}$ in the domain of the field, $\gamma(\mathbf{R}, \mathbf{r}, \tau)$ and $\gamma(\mathbf{r}, \mathbf{R}, \tau)$ are also unimodular and also periodic in time.

A perfectly coherent optical field at two fixed points has the property that $|\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau)| = 1$ for any two fixed points $\mathbf{R}_1$ and $\mathbf{R}_2$ in the domain of the field for all values of time delay τ. For such a field, it can be shown that $\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau) = \exp[\mathrm{i}(\beta - 2\pi\nu_0\tau)]$ with ν_0 (> 0) and β are real constants. Further, as a corollary, $|\gamma(\mathbf{R}_1, \mathbf{R}_1, \tau)| = 1$ and $|\gamma(\mathbf{R}_2, \mathbf{R}_2, \tau)| = 1$ for all τ, i.e., the field is self-coherent at each of the field points $\mathbf{R}_1$ and $\mathbf{R}_2$, and

$$\gamma(\mathbf{R}_1, \mathbf{R}_1, \tau) = \gamma(\mathbf{R}_2, \mathbf{R}_2, \tau) = \exp(-2\pi \mathrm{i} \nu_0 \tau) \text{ for } \nu_0 > 0$$

It can be shown that for any other point $\mathbf{r}'$ within the field

$$\begin{aligned} \gamma(\mathbf{r}, \mathbf{R}_2, \tau) &= \gamma(\mathbf{R}_1, \mathbf{R}_2, 0)\gamma(\mathbf{r}, \mathbf{R}_1, \tau) = \gamma(\mathbf{r}, \mathbf{R}_1, \tau)\mathrm{e}^{\mathrm{i}\beta} \\ \gamma(\mathbf{R}_1, \mathbf{R}_2, \tau) &= \gamma(\mathbf{R}_1, \mathbf{R}_2, 0) \exp[-2\pi \mathrm{i} \nu_0 \tau] \end{aligned} \qquad [31]$$

A perfectly coherent optical field for all points in the domain of the field has the property that $|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| = 1$ for all pairs of points $\mathbf{r}_1$ and $\mathbf{r}_2$ in the domain of the field, for all values of time delay τ.

It can then be shown that $\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=\exp\{i[\alpha(\mathbf{r}_1)-\alpha(\mathbf{r}_2)]-2\pi i\nu_0\tau\}$, where $\alpha(\mathbf{r}')$ is real function of a single position $\mathbf{r}'$ in the domain of the field. Further, the mutual coherence function $\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)$ for such a field has a factorized periodic form as

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=U(\mathbf{r}_1)U^*(\mathbf{r}_2)\ \exp(-2\pi i\nu_0\tau) \qquad [32]$$

which means the field is monochromatic with frequency ν_0 and $U(\mathbf{r})$ is any solution of the Helmholtz equation:

$$\nabla^2 U(\mathbf{r})+k_0^2U(\mathbf{r})=0, \quad k_0=2\pi\nu_0/c \qquad [33]$$

The spectral and cross-spectral densities of such fields are represented by Dirac δ-functions with their singularities at some positive frequency ν_0. Such fields, however, can never occur in nature.

Quasi-Monochromatic Fields

Optical fields are found in practice for which spectral spread $\Delta\nu$ is much smaller than the mean frequency $\bar{\nu}$. These are termed as quasi-monochromatic fields. The cross-spectral density $W(\mathbf{r}_1,\mathbf{r}_2,\nu)$ of the quasi-monochromatic fields attains an appreciable value only in the small region $\Delta\nu$, and falls to zero outside this region.

$$W(\mathbf{r}_1,\mathbf{r}_2,\nu)=0, \quad |\nu-\bar{\nu}|>\Delta\nu \ \text{ and } \ \Delta\nu\ll\bar{\nu} \qquad [34]$$

The mutual coherence function $\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)$ can be written as the Fourier transform of cross-spectral density function as

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=e^{-2\pi i\bar{\nu}\tau}\int_0^\infty W(\mathbf{r}_1,\mathbf{r}_2,\nu)e^{-2\pi i(\nu-\bar{\nu})\tau}d\nu \qquad [35]$$

If we limit our attention to small τ such that $\Delta\nu|\tau|\ll 1$ holds, the exponential factor inside the integral is approximately equal to unity and eqn [35] reduces to

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=\exp(-2\pi i\bar{\nu}\tau)\int_0^\infty [W(\mathbf{r}_1,\mathbf{r}_2,\nu)]d\nu \qquad [36]$$

which gives

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=\Gamma(\mathbf{r}_1,\mathbf{r}_2,0)\exp(-2\pi i\bar{\nu}\tau) \qquad [37]$$

Equation [37] describes the behavior of $\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)$ for a limited range of τ values for quasi-monochromatic fields and in this range it behaves as a monochromatic field of frequency $\bar{\nu}$. However, due to the factor $\Gamma(\mathbf{r}_1,\mathbf{r}_2,0)$, the quasi-monochromatic field may be coherent, partially coherent, or even incoherent.

Cross-Spectrally Pure Fields

The complex degree of coherence, if it can be factored into a product of a component dependent on spatial coordinates and a component dependent on time delay, is called reducible. In the case of perfectly coherent light, the complex degree of coherence is reducible, as we have seen above, e.g., in eqn [32], and in the case of quasi-monochromatic fields, this is reducible approximately as shown in eqn [37]. There also exists a very special case of a cross-spectrally pure field for which the complex degree of coherence is reducible. A field is called a cross-spectrally pure field if the normalized spectrum of the superimposed light is equal to the normalized spectrum of the component beams, a concept introduced by Mandel. In the space-frequency domain, the intensity interference law is expressed as the so-called spectral interference law:

$$S(\mathbf{r},\nu)=S^{(1)}(\mathbf{r},\nu)+S^{(2)}(\mathbf{r},\nu)+2\left[\sqrt{S^{(1)}(\mathbf{r},\nu)}\sqrt{S^{(2)}(\mathbf{r},\nu)}\right]\mathrm{Re}[\mu(\mathbf{r}_1,\mathbf{r}_2,\nu)e^{-2\pi i\nu\tau}] \qquad [38]$$

where $\mu(\mathbf{r}_1,\mathbf{r}_2,\nu)$ is the spectral degree of coherence, defined in eqn [21] and τ is the relative time delay that is needed by the light from the pinholes to reach any point on the screen; $S^{(1)}(\mathbf{r},\nu)$ and $S^{(2)}(\mathbf{r},\nu)$ are the spectral densities of the light reaching P from the pinholes P_1 and P_2, respectively (see Figure 4) and it is assumed that the spectral densities of the field at pinholes P_1 and P_2 are the same $[S(\mathbf{r}_1,\nu)=CS(\mathbf{r}_2,\nu)]$. Now, if we consider a point for which the time delay is τ_0, then it can be seen that the last term in eqn [38] would be independent of frequency, provided that

$$\mu(\mathbf{r}_1,\mathbf{r}_2,\nu)\ \exp(-2\pi i\nu\tau_0)=f(\mathbf{r}_1,\mathbf{r}_2,\tau_0) \qquad [39]$$

where $f(\mathbf{r}_1,\mathbf{r}_2,\tau_0)$ is a function of $\mathbf{r}_1$, $\mathbf{r}_2$ and τ_0 only and the light at this point would have the same spectral density as that at the pinholes. If a region exists around the specified point on the observation plane, such that the spectral distribution of the light in this region is of the same form as the spectral distribution of the light at the pinholes, the light at the pinholes is cross-spectrally pure light.

In terms of the spectral distribution of the light $S(\mathbf{r}_1,\nu)$ at pinhole P_1 and $S(\mathbf{r}_2,\nu)[=CS(\mathbf{r}_1,\nu)]$ at pinhole P_2, the mutual coherence function at the pinholes can be written as

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=\sqrt{C}\int S(\mathbf{r}_1,\nu)\mu(\mathbf{r}_1,\mathbf{r}_2,\nu)\ \exp(-2\pi i\nu\tau)d\nu \qquad [40]$$

and using eqn [39], we get the very important condition for the field to be cross-spectrally pure, i.e.:

$$\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)=\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau_0)\gamma(\mathbf{r}_1,\mathbf{r}_1,\tau-\tau_0) \qquad [41]$$

The complex degree of coherence $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ is thus expressible as the product of two factors: one factor characterizes the spatial coherence at the two pinholes at time delay τ_0 and the other characterizes the temporal coherence at one of the pinholes. Equation [41] is known as the reduction formula for cross-spectrally pure light. It can further be shown that

$$\mu(\mathbf{r}_1, \mathbf{r}_2, \nu) = \gamma(\mathbf{r}_1, \mathbf{r}_2, \tau_0)\ \exp(2\pi i \nu \tau_0) \qquad [42]$$

Thus, the absolute value of the spectral degree of coherence is the same for all frequencies and is equal to the absolute value of the degree of coherence for the point specified by τ_0. It has been shown that cross-spectrally pure light can be generated, for example, by linear filtering of light that emerges from the two pinholes in Young's interference experiments.

Types of Sources

Primary Sources

A primary source is a set of radiating atoms or molecules. In a primary source the randomness comes from true source fluctuations, i.e., from the spatial distributions of fluctuating charges and currents. Such a source gives rise to a fluctuating field. Let $Q(\mathbf{r}, t)$ represent the fluctuating source variable at any point $\mathbf{r}$ at time t, then the field generated by the source is represented by fluctuating field variable $V(\mathbf{r}, t)$. The source is assumed to be localized in some finite domain such that $Q(\mathbf{r}, t) = 0$ at any time $t > 0$ outside the domain. Assuming that field variable $V(\mathbf{r}, t)$ and the source variable $Q(\mathbf{r}, t)$ are scalar quantities, they are related by an inhomogeneous equation as

$$\nabla^2 V(\mathbf{r}, t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} V(\mathbf{r}, t) = -4\pi Q(\mathbf{r}, t) \qquad [43]$$

The mutual coherence functions of the source $\Gamma_Q(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle Q^*(\mathbf{r}_1, t) Q(\mathbf{r}_2, t + \tau)\rangle$ and of the field $\Gamma_V(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle V^*(\mathbf{r}_1, t) V(\mathbf{r}_2, t + \tau)\rangle$ characterize the statistical similarity of the fluctuating quantities at the points $\mathbf{r}_1$ and $\mathbf{r}_2$. Following eqn [20], one can define, respectively, the cross-spectral density function of the source and the field as

$$W_Q(\mathbf{r}_1, \mathbf{r}_2, \nu) = \int_{-\infty}^{\infty} \Gamma_Q(\mathbf{r}_1, \mathbf{r}_2, \tau) e^{-2\pi i \nu t}\, d\tau$$

$$W_V(\mathbf{r}_1, \mathbf{r}_2, \nu) = \int_{-\infty}^{\infty} \Gamma_V(\mathbf{r}_1, \mathbf{r}_2, \tau) e^{-2\pi i \nu t}\, d\tau \qquad [44]$$

The cross-spectral density functions of the source and of the field are related as

$$(\nabla_2^2 + k^2)(\nabla_1^2 + k^2) W_V(\mathbf{r}_1, \mathbf{r}_2, \nu) = 4\pi^2 W_Q(\mathbf{r}_1, \mathbf{r}_2, \nu) \qquad [45]$$

The solution of eqn [45] is represented as

$$W_V(\mathbf{r}_1, \mathbf{r}_2, \nu) = \int_S \int_S W_Q(\mathbf{r}'_1, \mathbf{r}'_2, \nu) \frac{e^{ik(R_2 - R_1)}}{R_1 R_2} d^3 r'_1\, d^3 r'_2 \qquad [46]$$

where $R_1 = |\mathbf{r}_1 - \mathbf{r}'_1|$ and $R_2 = |\mathbf{r}_2 - \mathbf{r}'_2|$ (see Figure 8). Using eqn [46], one can then obtain an expression for the spectrum at a point ($\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r} = r\mathbf{u}$) in the far field ($r \gg r'_1, r'_2$) of a source as

$$S^\infty(r\mathbf{u}, \nu) = \frac{1}{r^2} \int_S \int_S W_Q(\mathbf{r}'_1, \mathbf{r}'_2, \nu) e^{-ik\mathbf{u}\cdot(\mathbf{r}'_1 - \mathbf{r}'_2)} d^3 r'_1\, d^3 r'_2 \qquad [47]$$

where $\mathbf{u}$ is the unit vector along $\mathbf{r}$. The integral in eqn [47], i.e., the quantity defined by $r^2 S^\infty(r\mathbf{u}, \nu)$, is also defined as the radiant intensity, which represents the rate of energy radiated at the frequency ν from the source per unit solid angle in the direction of $\mathbf{u}$.

Secondary Sources

Sources used in a laboratory are usually secondary planar sources. A secondary source is a field, which arises from the primary source in the region outside the domain of the primary sources. This kind of source is an aperture on an opaque screen illuminated either directly or via an optical system by primary sources. Let $V(\boldsymbol{\rho}, t)$ represent the fluctuating field in a secondary source plane σ at $z = 0$ (Figure 9) and $W_0(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \nu)$ represent its cross-spectral density (the subscript 0 refers to $z = 0$). One can then solve eqn [25] to obtain the propagation of the cross-spectral density from this planar source. For two points P_1 and P_2 located at distances which are large compared to wavelength, the cross-spectral density is

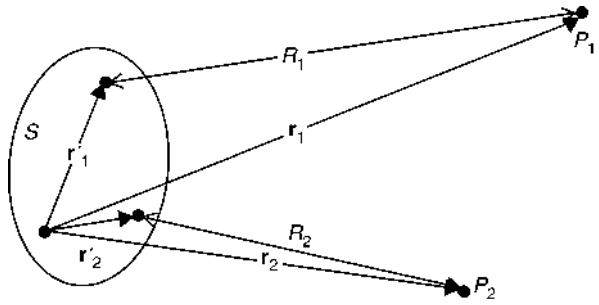

Figure 8 Geometry of a 3-dimensional primary source S and the radiation from it.

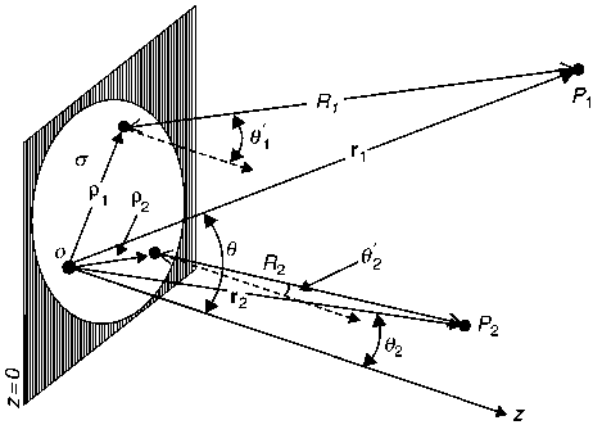

Figure 9 Geometry of a planar source σ and the radiation from it.

then given by

$$W(\mathbf{r}_1, \mathbf{r}_2, \nu) = \left(\frac{k}{2\pi}\right)^2 \int_\sigma \int_\sigma W_0(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \nu) \times \frac{e^{ik(R_2 - R_1)}}{R_1 R_2} \cos\theta'_1 \cos\theta'_2 \, d^2\rho_1 \, d^2\rho_2 \quad [48]$$

where $R_j = |\mathbf{r}_j - \boldsymbol{\rho}_j|$, $(j = 1, 2)$, θ'_1 and θ'_2 are the angles that R_1 and R_2 directions make with the z-axis (Figure 9). Using eqn [48], one can then obtain an expression for the spectrum at a point ($\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r} = r\mathbf{u}$) in the far field ($r \gg \rho_1, \rho_2$) of a planar source as

$$S^\infty(r\mathbf{u}, \nu) = \frac{k^2}{2\pi^2} \frac{\cos^2\theta}{r^2} \int_\sigma \int_\sigma W_0(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \nu) \times e^{-ik\mathbf{u}_\perp \cdot (\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)} d^2\rho_1 \, d^2\rho_2 \quad [49]$$

where $\mathbf{u}_\perp$ is the projection of the unit vector $\mathbf{u}$ on the plane σ of the source.

Schell-Model Sources

In the framework of coherence theory in space–time domain, two-dimensional planar model sources of this kind were first discussed by Schell. Later, the model was adopted for formulation of coherence theory in space–frequency domain. Schell-model sources are the sources whose degree of spectral coherence $\mu_A(\mathbf{r}_1, \mathbf{r}_2, \nu)$ (for either primary or secondary source) is stationary in space. It means that $\mu_A(\mathbf{r}_1, \mathbf{r}_2, \nu)$ depends on $\mathbf{r}_1$ and $\mathbf{r}_2$ only through the difference $\mathbf{r}_2 - \mathbf{r}_1$, i.e., of the form:

$$\mu_A(\mathbf{r}_1, \mathbf{r}_2, \nu) \equiv \mu_A(\mathbf{r}_2 - \mathbf{r}_1, \nu) \quad [50]$$

for each frequency ν present in the source spectrum. Here A stands for field variables V, in the case of a Schell-model secondary source and for source variables Q, in the case of a Schell-model primary source. The cross-spectral density function of a Schell-model source is of the form:

$$W_A(\mathbf{r}_1, \mathbf{r}_2, \nu) = [S_A(\mathbf{r}_1, \nu)]^{1/2} [S_A(\mathbf{r}_2, \nu)]^{1/2} \mu_A(\mathbf{r}_2 - \mathbf{r}_1, \nu) \quad [51]$$

where $S_A(\mathbf{r}, \nu)$ is the spectral density of the light at a typical point in the primary source or on the plane of a secondary source. Schell model sources do not assume low coherence, and therefore, can be applied to spatially stationary light fields of any state of coherence. The Schell-model of the form given in eqn [51] has been used to represent both three-dimensional primary sources and two-dimensional secondary sources.

Quasi-Homogeneous Sources

Useful models of partially coherent sources that are frequently encountered in nature or in the laboratory are the so-called quasi-homogeneous sources. These are an important sub-class of Schell-model sources. A Schell-model source is called quasi-homogeneous if the intensity of a Schell model source is essentially constant over any coherence area. Under these approximations, the cross-spectral density function for a quasi-homogeneous source is given by

$$W_A(\mathbf{r}_1, \mathbf{r}_2, \nu) = S_A[\tfrac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2), \nu] \mu_A(\mathbf{r}_2 - \mathbf{r}_1, \nu) = S_A[\mathbf{r}, \nu] \mu_A(\mathbf{r}', \nu) \quad [52]$$

where $\mathbf{r} = (\mathbf{r}_1 + \mathbf{r}_2)/2$, and $\mathbf{r}' = \mathbf{r}_2 - \mathbf{r}_1$. The subscript A stands for either V or Q for the field variable or a source variable, respectively. It is clear that for a quasi-homogeneous source the spectral density $S_A(\mathbf{r}, \nu)$ varies so slowly with position that it is approximately constant over distances across the sources that are of the order of the correlation length Λ, which is a measure of the effective width of $|\mu_A(\mathbf{r}', \nu)|$. Therefore, $S_A(\mathbf{r}, \nu)$ is a slowly varying function of $\mathbf{r}$ (Figure 10b) and $|\mu_A(\mathbf{r}', \nu)|$ is a fast varying function of $\mathbf{r}'$ (Figure 10a). In addition, the linear dimensions of the source are large compared with the wavelength of light and with the correlation length Δ (Figure 10c).

Quasi-homogeneous sources are always spatially incoherent in the 'global' sense, because their linear dimensions are large compared with the correlation length. This model is very good for representing two-dimensional secondary sources with sufficiently low coherence such that the intensity does not vary over the coherence area on the input plane. It has also been applied to three-dimensional primary sources,

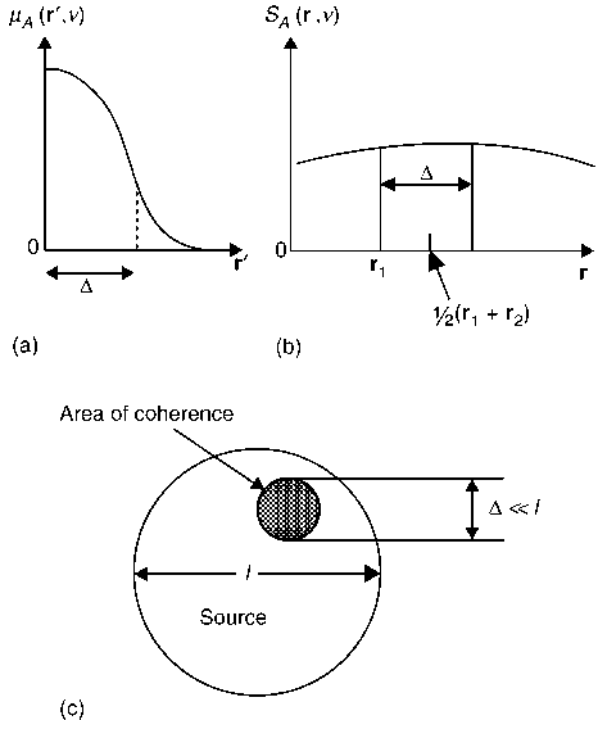

Figure 10 Concept of quasi-homogeneous sources.

three-dimensional scattering potentials, and two-dimensional primary and secondary sources.

Equivalence Theorems

The study of partially coherent sources led to the formulations of a number of equivalence theorems, which show that sources of any state of coherence can produce the same distribution of the radiant intensity as a fully spatially coherent laser source. These theorems provide conditions under which sources of different spatial distribution of spectral density and of different state of coherence will generate fields, which have the same radiant intensity. It has been shown, by taking examples of Gaussian–Schell model sources, that sources of completely different coherence properties and different spectral distributions across the source generate identical distribution of radiant intensity. Experimental verifications of the results of these theorems have also been carried out. For further details on this subject, the reader is referred to Mandel and Wolf (1995).

Correlation-Induced Spectral Changes

It was assumed that spectrum is an intrinsic property of light that does not change as the radiation propagates in free-space, until studies on partially coherent sources and radiations from them in 1980s, revealed that this was true only for specific type of sources. It was discovered on general grounds that the spectrum of light, which originates in an extended source, either a primary source or a secondary source, depends not only on the source spectrum but also on the spatial coherence properties of the source. It was also predicted theoretically by Wolf that the spectrum of light would, in general, be different from the spectrum of the source, and be different at different points in space on propagation in free space.

For a quasi-homogeneous planar secondary source defined by eqn [52], whose normalized spectrum is the same at each source point, one can write the spectral density as

$$S_0(\boldsymbol{\rho}, \nu) = I_0(\boldsymbol{\rho}) g_0(\nu) \quad \text{with} \quad \int_0^\infty g_0(\nu)\mathrm{d}\nu = 1 \qquad [53]$$

where $I_0(\boldsymbol{\rho})$ is the intensity of light at point $\boldsymbol{\rho}$ on the plane of the source, $g_0(\nu)$ is the normalized spectrum of the source and the subscript 0 refers to the quantities of the source plane. Using eqn [49], one can obtain an expression for the far-field spectrum due to this source as

$$S^\infty(r\mathbf{u}, \nu) = \frac{k^2}{2\pi^2}\frac{\cos^2\theta}{r^2}\int_\sigma\int_\sigma I_0(\boldsymbol{\rho}) g_0(\nu)\mu_0(\boldsymbol{\rho}', \nu) \times \mathrm{e}^{-\mathrm{i}k\mathbf{u}_\perp\cdot(\boldsymbol{\rho}_1-\boldsymbol{\rho}_2)}\,\mathrm{d}^2\rho_1\,\mathrm{d}^2\rho_2 \qquad [54]$$

Noting that $\boldsymbol{\rho} = (\boldsymbol{\rho}_1 + \boldsymbol{\rho}_2)/2$ and $\boldsymbol{\rho}' = \boldsymbol{\rho}_2 - \boldsymbol{\rho}_1$, one can transform the variables of the integration and obtain after some manipulation:

$$S^{(\infty)}(r\mathbf{u}, \nu) = \frac{k^2\cos^2\theta}{(2\pi r)^2}\tilde{I}_0\tilde{\mu}_0(k\mathbf{u}_\perp, \nu) g_0(\nu) \qquad [55]$$

where

$$\tilde{\mu}_0(k\mathbf{u}_\perp, \nu) = \int_\sigma \mu_0(\boldsymbol{\rho}', \nu)\mathrm{e}^{-\mathrm{i}k\mathbf{u}_\perp\cdot\boldsymbol{\rho}'}\,\mathrm{d}^2\rho' \qquad [56]$$

and

$$\tilde{I}_0 = \int_\sigma I_0(\boldsymbol{\rho})\mathrm{d}^2\rho \qquad [57]$$

Equation [55] shows that the spectrum of the field in the far-zone depends on the coherence properties of the source through its spectral degree of coherence $\mu_0(\boldsymbol{\rho}', \nu)$ and on the normalized source spectrum $g_0(\nu)$.

Scaling Law

The reason why coherence-induced spectral changes were not observed until recently is that the usual thermal sources employed in laboratories or

commonly encountered in nature have special coherence properties and the spectral degree of coherence has the function form:

$$\mu^{(0)}(\boldsymbol{\rho}_2-\boldsymbol{\rho}_1,\nu)=f[k(\boldsymbol{\rho}_2-\boldsymbol{\rho}_1)] \quad \text{with} \quad k=\frac{2\pi\nu}{c} \qquad [58]$$

which shows that the spectral degree of coherence depends only on the product of the frequency and space coordinates. This formula expresses the so-called scaling law, which was enunciated by Wolf. Commonly used sources satisfy this property. For example, the spectral degree of coherence of Lambertian sources and black-body sources can be shown to be

$$\mu_0(\boldsymbol{\rho}_2-\boldsymbol{\rho}_1,\nu)=\frac{\sin(k|\boldsymbol{\rho}_2-\boldsymbol{\rho}_1|)}{k|\boldsymbol{\rho}_2-\boldsymbol{\rho}_1|} \qquad [59]$$

This expression evidently satisfies the scaling law. If the spectral degree of coherence does not satisfy the scaling law, the normalized far-field spectrum will, in general, vary in different directions in the far-zone and will differ from the source spectrum.

Spectral Changes in Young's Interference Experiment

Spectral changes in Young's interference experiment with broadband light are not as well understood as in experiments with quasi-monochromatic, probably because in such experiments no interference fringes are formed. However, if one were to analyze the spectrum of the light in the region of superposition, one would observe changes in the spectrum of light in the region of superposition in the form of a shift in the spectrum for narrowband spectrum and spectral modulation for broadband light. One can readily derive an expression for the spectrum of light in the region of superposition. Let $S^{(1)}(P,\nu)$ be the spectral density of the light at P which would be obtained if the small aperture at P_1 alone was open; $S^{(2)}(P,\nu)$ has a similar meaning if only the aperture at P_2 was open (Figure 4).

Let us assume, as is usually the case, that $S^{(2)}(P,\nu)\approx S^{(1)}(P,\nu)$ and let d be the distance between the two pinholes. Consider the spectral density at the point P, at distance x from the axis of symmetry in an observation plane located at distance of R from the plane containing the pinholes. Assuming that $x/R \ll 1$, one can make the approximation $R_2-R_1\approx xd/R$. The spectral interference law (eqn [38]) can then be written as

$$S(P,\nu)\approx 2S^{(1)}(P,\nu)\{1+|\mu(P_1,P_2,\nu)| \times\cos[\beta(P_1,P_2,\nu)+2\pi\nu xd/cR]\} \qquad [60]$$

where $\beta(P_1,P_2,\nu)$ denotes the phase of the spectral degree of coherence. Equation [60] implies the two results:

(i) at any fixed frequency ν, the spectral density varies sinusoidally with the distance x of the point from the axis, with the amplitude and the phase of the variation depending on the (generally complex) spectral degree of coherence $\mu(P_1,P_2,\nu)$; and
(ii) at any fixed point P in the observation plane the spectrum $S(P,\nu)$ will, in general, differ from the spectrum $S^{(1)}(P,\nu)$, the change also depending on the spectral degree of coherence $\mu(P_1,P_2,\nu)$ of the light at the two pinholes.

Experimental Confirmations

Experimental tests of the theoretical prediction of spectral invariance and noninvariance due to correlation of fluctuations across the source were performed just after the theoretical predictions. Figure 11 shows results of one such experiment in which spectrum changes in the Young's experiment were studied. Several other experiments also reported confirmation of the source correlation-dependent spectral changes. One of the important applications of these observations has been to explain the discrepancies in the maintenance of the spectroradiometric scales by national laboratories in different countries. These studies also have potential application in determining experimentally the spectral degree of coherence of partially coherent fields. The knowledge of spectral degree of coherence is often important in remote sensing, e.g., for determining angular diameters of stars.

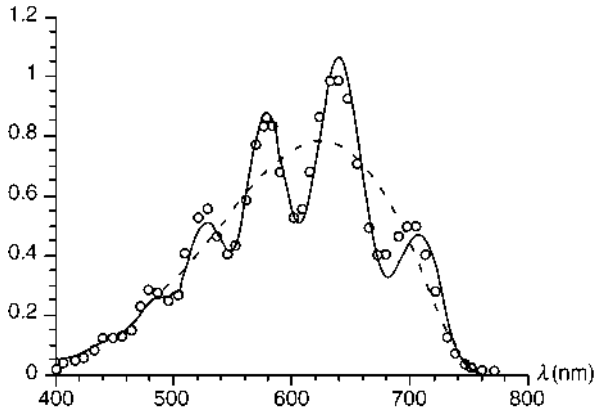

Figure 11 Correlation-induced changes in spectrum in Young's interference. Dashed line represents the spectrum when only one of the slits is open, the continuous curve shows the spectrum when both the slits are open and the circles are the measured values in the latter case. Reproduced with permission from Santarsiero M and Gori F (1992) Spectral changes in Young interference pattern. *Physics Letters* 167: 123–128.

Applications of Optical Coherence

Stellar Interferometry

The Michelson stellar interferometer, named after Albert Michelson, was used to determine the angular diameters of stars and also the intensity distribution across the star. The method was devised by Michelson without using any concept of coherence, although subsequently the full theory of the method was developed on the basis of propagation of correlations. A schematic of the experiment is shown in Figure 12. The interferometer is mounted in front of a telescope, a reflecting telescope in this case. The light from a star is reflected from mirrors M_1 and M_2 and is directed towards the primary mirror (or the objective lens) of the telescope. The two beams thus collected superpose in the focal plane F of telescope where an image crossed with fringes is formed. The outer mirrors M_1 and M_2 can be moved along the axis defined as $M_1M_3M_4M_2$ while the inner mirrors M_3 and M_4 remain fixed. The fringe spacing depends on the position of mirrors M_3 and M_4 and hence is fixed, while the visibility of the fringes depends on the separation of the mirrors M_1 and M_2 and hence, can be varied. Michelson showed that from the measurement of the variation of the visibility with the separation of the two mirrors, one could obtain information about the intensity distribution of the stars, which are rotationally symmetric. He also showed that if the stellar disk is circular and uniform, the visibility curve as a function of the separation d of the mirrors M_1 and M_2 will have zeros for certain values of d, and that the smallest of these d values for which zero occurs is given by $d_0 = 0.61\lambda_a/\alpha$, where α is the semi-angular diameter of the star and λ_a is the mean wavelength of the filtered quasi-monochromatic light from the star. Angular diameters of several stars down to 0.02 second of an arc were determined.

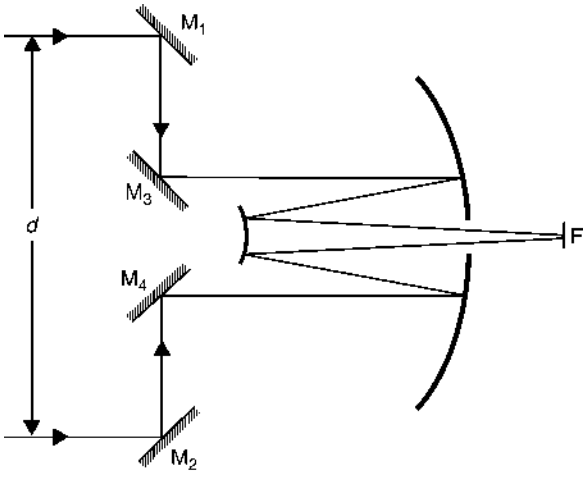

Figure 12 Schematic of the Michelson stellar interferometer.

From the standpoint of second-order coherence theory the principles of the method can be readily understood. The star is considered an incoherent source and according to the van Cittert–Zernike theorem, the light reaching the outer mirrors M_1 and M_2 of the interferometer will be partially coherent. This coherence would depend on the size of and the intensity distribution across the star. Let (x_1, y_1) and (x_2, y_2) be the coordinates of the positions of the mirrors M_1 and M_2, respectively, and (ξ, η) the coordinates of a point on the surface plane of the star which is assumed to be at a very large (astronomical) distance R from the mirrors. The complex degree of coherence at the mirrors would then be given by eqn [22] which can now be written as

$$\gamma(\Delta x, \Delta y, 0) = \frac{\int_\sigma I(u,v)\mathrm{e}^{-\mathrm{i}k_a(u\Delta x + v\Delta y)}\,\mathrm{d}u\,\mathrm{d}v}{\int_\sigma I(u,v)\mathrm{d}u\,\mathrm{d}v} \qquad [61]$$

where $I(u, v)$ is the intensity distribution across the star disk σ as a function of the angular coordinates $u = \xi/R$, $v = \eta/R$, $\Delta x = x_1 - x_2$, $\Delta y = y_1 - y_2$, and $k_a = 2\pi/\lambda_a$, λ_a being the mean wavelength of the light from the star. Equation [61] shows that the equal-time ($\tau = 0$) complex degree of coherence of the light incident at the outer mirrors of the interferometer is the normalized Fourier-transform of the intensity distribution across the stellar disk. Further, eqn [15] shows that the visibility of the interference fringes is the absolute value of γ, if the intensity of the two interfering beams is equal, as in the present case. The phase of γ can be determined by the position of the intensity maxima of the fringe pattern (eqn [14]). If one is interested in determining only the angular size of the star and the star is assumed to be a circularly symmetric disk of angular diameter 2α and of uniform intensity [$I(u, v)$ is constant across the disk], then eqn [61] reduces to

$$\gamma(\Delta x, \Delta y) = \frac{2J_1(v)}{v}, \quad v = \frac{2\pi\alpha}{\lambda_a}d, \quad d = \sqrt{(\Delta x)^2 + (\Delta y)^2} \qquad [62]$$

The smallest separation of the mirrors for which the visibility γ vanishes corresponds to $v = 3.832$, i.e., $d_0 = 0.61\lambda_a/\alpha$, which is in agreement with Michelson's result.

Interference Spectroscopy

Another contribution of Michelson, which was subsequently identified as an application of the

coherence theory, was the use of his interferometer (Figure 2) to determine the energy distribution in the spectral lines. The method he developed is capable to resolving spectral lines that are too narrow to be analyzed by the spectrometer. The visibility of the interference fringes depends on the energy distribution in the spectrum of the light and its measurement can give information about the spectral lines. In particular, if the energy distribution in a spectral line is symmetric about some frequency ν_0, its profile is simply the Fourier transform of the visibility variation as a function of the path difference between the two interfering beams. This method is the basis of interference spectroscopy or the Fourier transform spectroscopy.

Within the framework of second-order coherence theory, if the mean intensity of the two beams in a Michelson's interferometer is the same, then the visibility of the interference fringes in the observation plane is related to the complex degree of coherence of light at a point at the beamsplitter where the two beams superimpose (eqn [15]). The two quantities are related as $v(\tau) = |\gamma(\tau)|$ where $\gamma(\tau) \equiv \gamma(\mathbf{r}_1, \mathbf{r}_1, \tau)$ is the complex degree of self-coherence at the point $\mathbf{r}_1$ on the beamsplitter. Following eqn [19], $\gamma(\tau)$ can be represented by a Fourier integral as

$$\gamma(\tau) = \int_0^\infty s(\nu)\exp(-\mathrm{i}2\pi\nu\tau)\mathrm{d}\nu \qquad [63]$$

where $s(\nu)$ is the normalized spectral density of the light defined as $s(\nu) = S(\nu)/\int_0^\infty S(\nu)\mathrm{d}\nu$ and $S(\nu) \equiv S(\mathbf{r}_1, \nu) = W(\mathbf{r}_1, \mathbf{r}_1, \nu)$ is the spectral density of the beam at the point $\mathbf{r}_1$. The method is usually applied to very narrow spectral lines for which one can assume that the peak that occurs at ν_0, and $\gamma(\tau)$ can be represented as

$$\begin{aligned} \gamma(\tau) &= \bar{\gamma}(\tau)\exp(-2\pi\mathrm{i}\nu_0\tau) \quad \text{with} \\ \bar{\gamma}(\tau) &= \int_{-\infty}^\infty \bar{s}(\mu)\exp(-\mathrm{i}2\pi\mu\tau)\mathrm{d}\mu \end{aligned} \qquad [64]$$

where $\bar{s}(\mu)$ is the shifted spectrum such that

$$\begin{aligned} \bar{s}(\mu) &= s(\nu_0 + \mu) \quad & \mu \geq -\nu_0 \\ &= 0 & \mu < -\nu_0 \end{aligned}$$

From the above one readily gets

$$v(\tau) = |\gamma(\tau)| = \left| \int_{-\infty}^\infty \bar{s}(\mu)\,\exp(-\mathrm{i}2\pi\mu\tau)\mathrm{d}\mu \right| \qquad [65]$$

If the spectrum is symmetric about ν_0, then $v(\tau)$ would be an even function of τ and the Fourier inversion would give:

$$\bar{s}(\mu) = s(\nu_0 + \mu) = 2\int_0^\infty v(\tau)\cos(2\pi\mu\tau)\mathrm{d}\tau \qquad [66]$$

which can be used to calculate the spectral energy distribution for a symmetric spectrum about ν_0 from the visibility curve. However, for an asymmetric spectral distribution, the visibility and the phase of the complex degree of coherence must be determined as the Fourier transform of the shifted spectrum is no longer real everywhere.

Higher-Order Coherence

So far we have considered correlations of the fluctuating field variables at two space–time $(\mathbf{r}, t)$ points, as defined in eqn [10]. These are termed as second-order correlations. One can extend the concept of correlations to more than two space–time points, which will involve higher-order correlations. For example, one can define the space–time cross correlation function of order (M, N) of the random field $V(\mathbf{r}, t)$, represented by $\Gamma^{(M,N)}$, as an ensemble average of the product of the field $V(\mathbf{r}, t)$ values at N space–time points and $V^*(\mathbf{r}, t)$ at other M points. In this notation, the mutual coherence function as defined in eqn [10] would now be $\Gamma^{(1,1)}$. Among higher-order correlations, the one with $M = N = 2$, is of practical significance and is called the fourth-order correlation function, $\Gamma^{(2,2)}(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2; \mathbf{r}_3, t_3, \mathbf{r}_4, t_4)$. The theory of Gaussian random variables tells us that any higher correlation can be written in terms of second-order correlations over all permutations of pairs of points. In addition, if we assume that $(\mathbf{r}_3, t_3) = (\mathbf{r}_1, t_1)$ and $(\mathbf{r}_4, t_4) = (\mathbf{r}_2, t_2)$, and that the field is stationary, then $\Gamma^{(2,2)}$ is called the intensity–intensity correlation and is given as

$$\begin{aligned} &\Gamma^{(2,2)}(\mathbf{r}_1, \mathbf{r}_2, t_2 - t_1) \\ &= \langle V(\mathbf{r}_1, t_1)V(\mathbf{r}_2, t_2)V^*(\mathbf{r}_1, t_1)V^*(\mathbf{r}_2, t_2)\rangle \\ &= \langle I(\mathbf{r}_1, t_1)I(\mathbf{r}_2, t_2)\rangle \\ &= \langle I(\mathbf{r}_1, t_1)\rangle\langle I(\mathbf{r}_2, t_2)\rangle(1 + |\gamma^{(1,1)}(\mathbf{r}_1, \mathbf{r}_2, t_2 - t_1)|^2) \end{aligned} \qquad [67]$$

where

$$\gamma^{(1,1)}(\mathbf{r}_1, \mathbf{r}_2, t_2 - t_1) = \frac{\Gamma^{(1,1)}(\mathbf{r}_1, \mathbf{r}_2, t_2 - t_1)}{[\langle I(\mathbf{r}_1, t_1)\rangle]^{1/2}[\langle I(\mathbf{r}_2, t_2)\rangle]^{1/2}} \qquad [68]$$

We now define fluctuations in intensity at $(\mathbf{r}_j, t_j)$ as

$$\Delta I_j = I(\mathbf{r}_j, t_j) - \langle I(\mathbf{r}_j, t_j)\rangle$$

and then the correlation of intensity fluctuations becomes

$$\langle \Delta I_1 \Delta I_2 \rangle = \langle I(\mathbf{r}_1,t_1) I(\mathbf{r}_2,t_2) \rangle - \langle I(\mathbf{r}_1,t_1) \rangle \langle I(\mathbf{r}_2,t_2) \rangle$$
$$= \langle I(\mathbf{r}_1,t_1) \rangle \langle I(\mathbf{r}_2,t_2) \rangle |\gamma^{(1,1)}(\mathbf{r}_1,\mathbf{r}_2,t_2-t_1)|^2 \qquad [69]$$

where we have used eqn [67]. Equation [69] forms the basis for intensity interferometry.

Hanbury-Brown and Twiss Experiment

In this landmark experiment conducted both on the laboratory scale and astronomical scale, Hanbury-Brown and Twiss demonstrated the existence of intensity–intensity correlations in terms of the correlations in the photocurrents in the two detectors and thus measured the squared modulus of complex degree of coherence. In the laboratory experiment, the arc of a mercury lamp was focused onto a circular hole to produce a secondary source. The light from this source was then divided equally into two parts through a beamsplitter. Each part was, respectively, detected by two photomultiplier tubes, which had identical square apertures in front. One of the tubes could be translated normally to the direction of propagation of light and was so positioned that its image through the splitter could be made to coincide with the other tube. Thus, by suitable translation a measured separation d between the two square apertures could be introduced. The output currents from the photomultiplier tubes were taken by cables of equal length to a correlator. In the path of each cable a high-pass filter was inserted, so that only the current fluctuations could be transmitted to the correlator. Thus, the normalized correlations between the two current fluctuations:

$$C(d) = \frac{\langle \Delta J_1(t) \Delta J_2(t) \rangle}{\langle [\Delta J_1(t)]^2 \rangle^{1/2} \langle [\Delta J_2(t)]^2 \rangle^{1/2}} \qquad [70]$$

as a function of detector separation d could be measured. Now when the detector response time is much larger than the time-scale of the fluctuations in intensity, then it can be shown that the correlations in the fluctuations of the photocurrent are proportional to the correlations of the fluctuations in intensity of the light being detected. Thus, we would have

$$C(d) \approx \delta |\gamma^{(1,1)}(\mathbf{r}_1, \mathbf{r}_2, 0)|^2 \qquad [71]$$

where δ is the average number of photocounts of the light of one polarization during the time-scale of the fluctuations (for general thermal sources, this is much less than one). Equation [71] represents the Hanbury-Brown–Twiss effect.

Stellar Intensity Interferometry

Michelson stellar interferometry can resolve stars which have angular sizes of the order of $0.01''$, since for smaller stars, the separation between the primary mirrors runs into several meters and maintaining stability of mirrors such that the optical paths do not change, even by fraction of a wavelength, is extremely difficult. The atmospheric turbulence further adds to this problem and obtaining stable fringe pattern becomes next to impossible for very small stars. Hanbury-Brown and Twiss applied the intensity interferometry based on their photoelectric correlation technique for determining the angular sizes of such stars. Two separate parabolic mirrors collected light from the star and the output of the photodetectors placed at the focus of each mirror was sent to a correlator. The cable lengths were made unequal so as to compensate for the time difference of the light arrival at the two mirrors. The normalized correlation of the fluctuations of the photocurrents was determined as described above. This would give the variation of the modulus-squared degree of coherence as a function of the mirror separation d from which the angular size of the stars can be estimated. The advantage of the stellar intensity interferometer over the stellar (amplitude) interferometer is that the light need not interfere as in the latter, since the photodetectors are mounted directly at the focus of the primary mirrors of the telescope. Thus, the constraint on the large path difference between the two beams is removed and large values of d can now be used. Moreover, the atmosphere turbulence and the mirror movements have very small effect. Stellar angular diameters as small as $0.0004''$ of arc with resolution of $0.00003''$ could be measured by such interferometers.

See also

Coherence: Coherence and Imaging. **Coherent Lightwave Systems**. **Coherent Transients:** Coherent Transient Spectroscopy in Atomic and Molecular Vapours. **Coherent Control:** Applications in Semiconductors; Experimental; Theory. **Interferometry:** Overview. **Information Processing:** Coherent Analogue Optical Processors. **Terahertz Technology:** Coherent Terahertz Sources.

Further Reading

Beran MJ and Parrent GB (1964) *Theory of Partial Coherence*. Englewood Cliffs, NJ: Prentice-Hall.

Born M and Wolf E (1999) *Principles of Optics*. New York: Pergamon Press.
Carter WH (1996) Coherence theory. In: Bass M (ed.) *Hand Book of Optics*. New York: McGraw-Hill.
Davenport WB and Root WL (1960) *An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill.
Goodman JW (1985) *Statistical Optics*. Chichester: Wiley.
Hanbury-Brown R and Twiss RQ (1957) Interferometry of intensity fluctuations in light: I basic theory: the correlation between photons in coherent beams of radiation. *Proceedings of the Royal Society* 242: 300–324.
Hanbury-Brown R and Twiss RQ (1957) Interferometry of intensity fluctuations in light: II an experimental test of the theory for partially coherent light. *Proceedings of the Royal Society* 243: 291–319.
Kandpal HC, Vaishya JS and Joshi KC (1994) Correlation-induced spectral shifts in optical measurements. *Optical Engineering* 33: 1996–2012.
Mandel L and Wolf E (1965) Coherence properties of optical fields. *Review of Modern Physics* 37: 231–287.
Mandel L and Wolf E (1995) *Optical Coherence and Quantum Optics*. Cambridge, UK: Cambridge University Press.
Marathay AS (1982) *Elements of Optical Coherence*. New York: John Wiley.
Perina J (1985) *Coherence of Light*. Boston, MA: M.A. Reidel.
Santarsiero M and Gori F (1992) Spectral changes in Young interference pattern. *Physics Letters* 167: 123–128.
Schell AC (1967) A technique for the determination of the radiation pattern of a partially coherent aperture. *IEEE Transactions on Antennas and Propagation* AP-15: 187.
Thompson BJ (1958) Illustration of phase change in two-beam interference with partially coherent light. *Journal of the Optical Society of America* 48: 95–97.
Thompson BJ and Wolf E (1957) Two beam interference with partially coherent light. *Journal of the Optical Society of America* 47: 895–902.
Wolf E and James DFV (1996) Correlation-induced spectral changes. *Report on Progress in Physics* 59: 771–818.

Coherence and Imaging

J van der Gracht, HoloSpex, Inc., Columbia, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical imaging systems are strongly affected by the coherence of the light that illuminates the object of interest. In many cases, the light is approximately coherent or incoherent. These approximations lead to simple mathematical models for the image formation process and allow straightforward analysis and design of such systems. When the light is partially coherent, the mathematical models are more complicated and system analysis and design is more difficult. Partially coherent illumination is often used in microscopy, machine vision, and optical lithography. The intent of this article is to provide the reader with a basic understanding of the effects of coherence on imaging. The information should enable the reader to recognize when coherence effects are present in an imaging system and give insight into when coherence can be modified to improve imaging performance. The material relies mostly on concepts drawn from the Fourier optics perspective of imaging and a rigorous coherence theory treatment is avoided. We encourage the reader to consult the Further Reading list at the end of this article, for more complete definitions of terms for coherence theory. A number of experimental results highlight the effects of spatial coherence of the illumination on image formation. Discussions of the role of coherence in such key applications are interspersed throughout the article.

Image Formation – Ideal and Optimal

An image is typically defined as the reproduction or likeness of the form of an object. An image that is indistinguishable from the original object is generally considered to be ideal. In a general context, the sound of a voice coming from a loudspeaker can be thought of as the image of the sound coming directly from the original speaker's mouth. In optical imaging, the ideal image replicates the light emanating from the object. Taken to the extreme, the ideal image replicates the light leaving the object in terms of intensity, wavelength, polarization, and even coherence. When the final image is viewed by the eye, the ideal image only needs to replicate the spatial distribution of the light leaving the object in terms of color and relative intensity at each point on the object. (In this article, intensity is defined as optical power per unit area (watts per meter squared).)

A general diagram of a direct view image formation system is shown in Figure 1. The condenser optics gathers light from a primary source and illuminates a transmissive object having a complex wave amplitude

transmission of $O(x, y)$. The imaging optics produce an optical image that is viewed directly by the viewer. A growing number of image formation systems now include a solid-state image detector as shown in Figure 2. The raw intensity optical image is converted to an electronic signal that can be digitally processed and then displayed on a monitor. In this system, the spatial intensity distribution emanating from the monitor is the final image. The task of the optical system to the left of the detector is to gather spatial information about the light properties of the object. In fact, the information gathered by the detector includes information about the object, the illumination system, and the image formation optics. If the observer is only interested in the light transmission properties of the object, the effects of the illumination and image formation optics must be well understood. When the light illuminating the object is known to be coherent or incoherent, reasonably simple models for the overall image formation process can be used. More general, partially coherent illumination can produce optically formed images that differ greatly from the intensity pattern leaving the object. Seen from this perspective, partially coherent illumination is an undesirable property that creates nonideal images and complicates the image analysis.

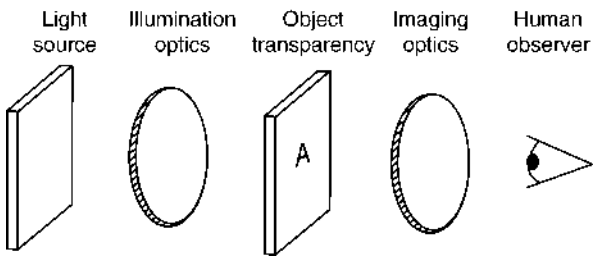

Figure 1 Direct view optical system. Imaging optics conveys the light from an illuminated object directly to the human viewer.

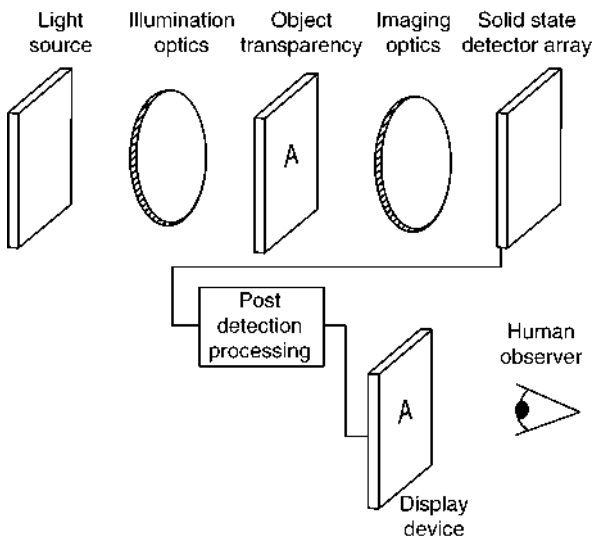

Figure 2 Indirect view optical system. The raw intensity optical image is detected electronically, processed and a final image is presented to the human observer on a display device.

The ideal image as described above is not necessarily the optimum image for a particular task. Consider the case of object recognition when the object has low optical contrast. Using the above definition, an ideal image would mimic the low contrast and render recognition difficult. An image formation system that alters the contrast to improve the recognition task would be better than the so-called ideal image. The image formation system that maximized the appropriate performance metric for a particular recognition task would be considered optimal. In optimizing the indirect imaging system of Figure 2, the designer can adjust the illumination, the imaging optics, and the post-detection processing. In fact, research microscopes often include an adjustment that modifies illumination coherence and often alters the image contrast. Darkfield and phase contrast imaging microscopes usually employ partially coherent illumination combined with pupil modification to view otherwise invisible objects. Seen from this perspective, partially coherent illumination is a desirable property that provides more degrees of freedom to the imaging system designer. Quite often, partially coherent imaging systems provide a compromise between the performance of coherent and incoherent systems.

Elementary Coherence Concepts

Most readers are somewhat familiar with the concept of temporal coherence. In Figure 3, a Michelson interferometer splits light from a point source into two paths and recombines the beams to form interference fringes. The presence of interference fringes indicates that the wave amplitude fluctuations of the two beams are highly correlated so the light adds in wave amplitude. If the optical path difference between the two paths can be made large without reducing the fringe contrast, the light is said to be

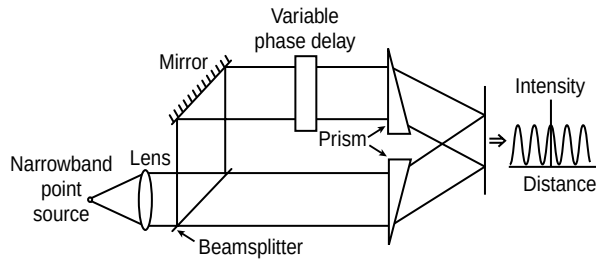

Figure 3 The presence of high contrast fringes in a Michelson interferometer indicates high temporal coherence.

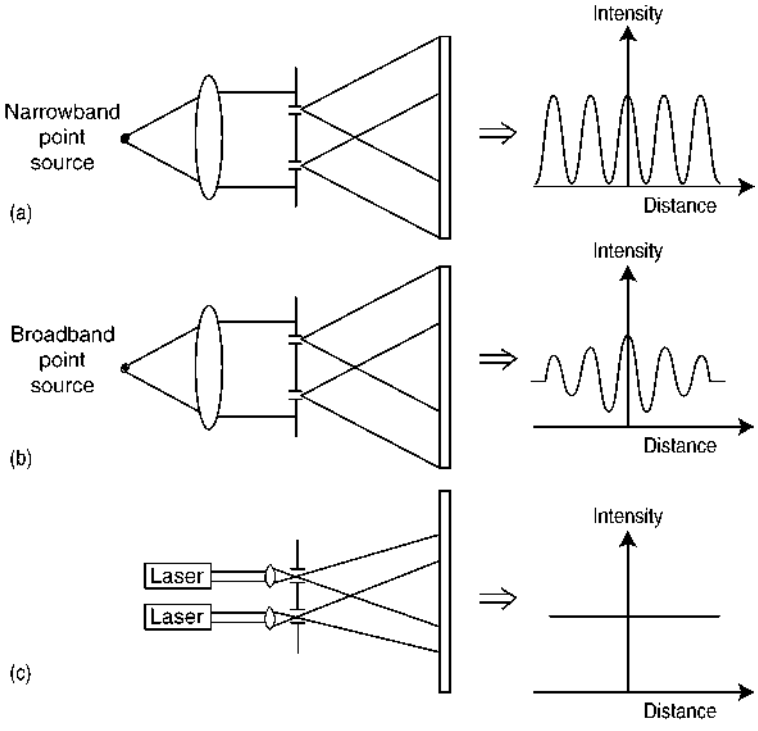

Figure 4 A Young's two pinhole interferometer produces: (a) a uniform high contrast fringe pattern for light that is highly spatially and temporally coherent; (b) high contrast fringes only near the axis for low temporal coherence and high spatial coherence light; and (c) no fringe pattern for high temporal coherence, low spatial coherence light.

highly temporally coherent. Light from very narrowband lasers can maintain coherence over very large optical path differences. Light from broadband light sources requires very small optical path differences to add in wave amplitude.

Spatial coherence is a measure of the ability of two separate points in a field to interfere. The Young's two pinhole experiment in Figure 4 measures the coherence between two points sampled by the pinhole mask. Only a one-dimensional pinhole mask is shown for simplicity. Figure 4a shows an expanded laser beam illuminating two spatially separated pinholes and recombining to form an intensity fringe pattern. The high contrast of the fringe indicates that the wave amplitude of the light from the two pinholes is highly correlated. Figure 4b shows a broadband point source expanded in a similar fashion. The fringe contrast is high near the axis because the optical path difference between the two beams is zero on the axis and relatively low near the axis. For points far from the axes, the fringe pattern disappears because the low temporal coherence from the broadband source results in a loss in correlation of the wave amplitudes.

A final Young's experiment example in Figure 4c shows that highly temporally coherent light can be spatially incoherent. In the figure, the light illuminating the two pinholes comes from two separate and highly temporally coherent lasers that are designed to have the same center frequency. Since the light from the lasers is not synchronized in phase, any fringes that might form for an instant will move rapidly and average to a uniform intensity pattern over the integration time of a typical detector. Since the fringe contrast is zero over a practical integration time, the light at the two pinholes is effectively spatially incoherent.

Two-point Imaging

In a typical partially coherent imaging experiment we need to know how light from two pinholes adds at the optical image plane, as shown in Figure 5. Diffraction and system aberrations cause the image of a single point to spread so that the images of two spatially separated object points overlap in the image plane. The wave amplitude image of a single pinhole is called the coherent point spread function (CPSF). Since the CPSF is compact, two CPSFs will only overlap when the corresponding object points are closely spaced. When the two points are sufficiently close, the relevant optical path differences will be small so full temporal coherence can be assumed. Spatial coherence will be the critical factor in determining how to add the responses to pairs of

point images. We assume full temporal coherence in the subsequent analyses and the light is said to be *quasimonochromatic*.

Coherent Two-Point Imaging

Consider imaging a two-point object illuminated by a spatially coherent plane wave produced from a point source, as depicted in Figure 6a. In the figure, the imaging system is assumed to be space-invariant. This means that the spatial distribution of the CPSF is the same regardless of position of the input pinhole. The CPSF in the figure is broad enough such that the image plane point responses overlap for this particular pinhole spacing. Since the wave amplitudes from the two pinholes are correlated, the point responses add in wave amplitude, resulting in a two-point image intensity given by

$$I_{2\mathrm{coh}}(x) = I_0|h(x - x_1)|^2 + I_0|h(x - x_2)|^2 \qquad [1]$$

where $h(x)$ is the normalized CPSF and I_0 is a scaling factor that determines the absolute image intensity value. Since the CPSF has units of wave amplitude, the magnitude squaring operation accounts for the square law response of the image plane detector.

More generally, the image intensity for an arbitrary object distribution can be found by breaking the object into a number of points and adding the CPSFs due to each point on the object. The resultant image plane intensity is the spatial convolution of the object amplitude transmittance with the CPSF and is given by

$$I_{\mathrm{coh}}(x) = I_0\left|\int O(\xi)h(\xi - x)\mathrm{d}\xi\right|^2 \qquad [2]$$

where $O(x)$ is the object amplitude transmittance.

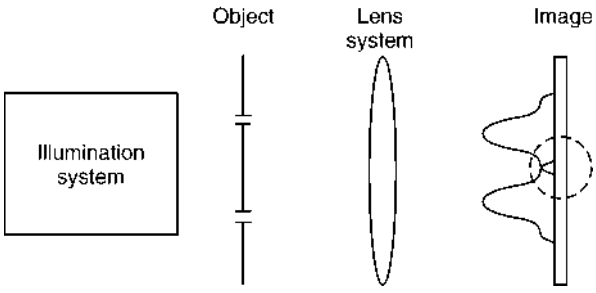

Figure 5 Imaging of two points. The object plane illumination coherence determines how the light adds in the region of overlap of the two image plane points.

Incoherent Two-Point Imaging

Two pinhole imaging with spatially incoherent light is shown in Figure 6b, where a spatially extended blackbody radiator is placed directly behind the pinholes. Once again the imaging optics are assumed to be space-invariant and have the same CPSF as the system in Figure 6a. Since the radiation is originating from two completely different physical points on the source, no correlation is expected between the wave amplitude of the light leaving the pinholes and the light is said to be spatially incoherent. The resultant image intensity corresponding to the two pinhole object with equal amplitude transmittance values is calculated by adding the individual intensity responses to give

$$I_{2\mathrm{inc}}(x) = I_0|h(x - x_1)|^2 + I_0|h(x - x_2)|^2 \qquad [3]$$

For more general object distributions, the image intensity is the convolution of the object intensity transmittance with the incoherent PSF, and is given by

$$I_{\mathrm{inc}}(x) = I_0\int |o(\xi)|^2|h(x - \xi)|^2\mathrm{d}\xi \qquad [4]$$

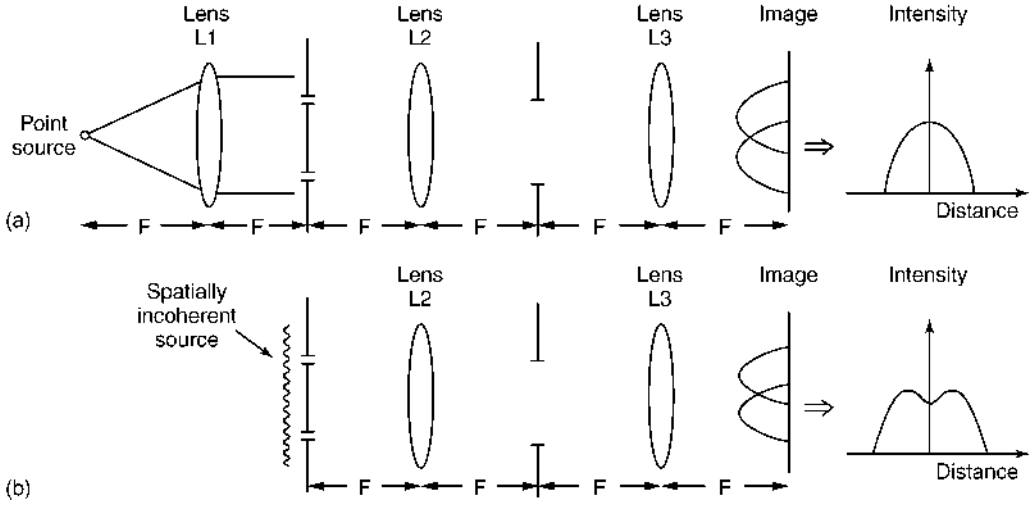

Figure 6 Imaging of two points for (a) spatially coherent illumination and (b) spatially incoherent imaging.

where the intensity transmittance is the squared magnitude of the amplitude transmittance and the incoherent PSF is the squared magnitude of the CPSF.

Partially Coherent Two-Point Imaging

Finally we consider two pinhole imaging illuminated by partially spatially coherent light. Now the two-point responses do not add simply in amplitude or intensity. Rather, the image intensity is given by the more general equation:

$$I_{2\mathrm{pc}}(x) = I_0[|h(x-x_1)^2 + |h(x-x_2)|^2 + 2\mathrm{Re}\{\mu(x_1;x_2)h(x-x_1)h^*(x-x_2)\}] \quad [5]$$

where the $*$ denotes complex conjugations and $\mu(x_1;x_2)$ is the normalized form of the mutual intensity of the object illumination evaluated at the object points in question. The mutual intensity function is often denoted as $J_0(x_1;x_2)$ and is a measure of the cross correlation of the wave amplitude distributions leaving the two pinholes. Rather than providing a rigorous definition, we note that the magnitude of $J_0(x_1;x_2)$ corresponds to the fringe contrast that would be produced if the two object pinholes were placed at the input of a Young's interference experiment. The phase is related to the relative spatial shift of the fringe pattern. When the light is uncorrelated, $J_0(x_1;x_2)=0$ and eqn [5] collapses to the incoherent limit of eqn [4]. When the light is coherent, $J_0(x_1;x_2)=1$ and eqn [5] reduces to the coherent form of eqn [2].

Image formation for a general object distribution is given by the bilinear equation:

$$I_{\mathrm{pc}}(x) = I_0 \int o(\xi_1)o^*(\xi_2)h(x-\xi_1)h^*(x-\xi_2) \times J_o(\xi_1-\xi_2)\mathrm{d}\xi_1\,\mathrm{d}\xi_2 \quad [6]$$

Note that, in general, $J_0(x_1;x_2)$ must be evaluated for all pairs of object points. Close examination of this equation reveals that just as a linear system can be evaluated by considering all possible single point responses, a bilinear system requires consideration of all possible pairs of points. This behavior is much more complicated and does not allow the application of the well-developed linear system theory.

Source Distribution and Object Illumination Coherence

According to eqn [6], the mutual intensity of all pairs of points at the object plane must be known, to calculate a partially coherent image. Consider the telecentric Kohler illumination imaging system shown in Figure 7. The primary source is considered to be spatially incoherent and illuminates the object after passing through lens L1 located one focal distance away from the source and one focal distance away from the object. Even though the primary source is spatially incoherent, the illumination at the object plane is partially coherent. The explicit mutual intensity function corresponding to the object plane illumination is given by applying the van Cittert Zernike theorem:

$$J_0(\Delta x) = \int S(\xi)\exp[\mathrm{j}2\pi\Delta x\xi/\lambda F]\mathrm{d}\xi \quad [7]$$

where $S(x)$ is the intensity distribution of the spatially incoherent primary source, F is the focal length of the lens, λ is the illumination wavelength, and $\Delta x = x_1 - x_2$. The van Cittert Zernike theorem reflects a Fourier transform relationship between the source image intensity distribution and the mutual intensity at the object plane. When the source plane is effectively spatially incoherent, the object plane mutual intensity is only a function of separation distance. For a two-dimensional object, the mutual intensity needs to be characterized for all pairs of unique spacings in x and y. In Figure 7, the lens arrangement ensures that the object plane is located in the far field of the primary source plane. In fact, the van Cittert Zernike theorem applies more generally, even in the Fresnel propagation regime, as long as the standard paraxial approximation for optics is valid.

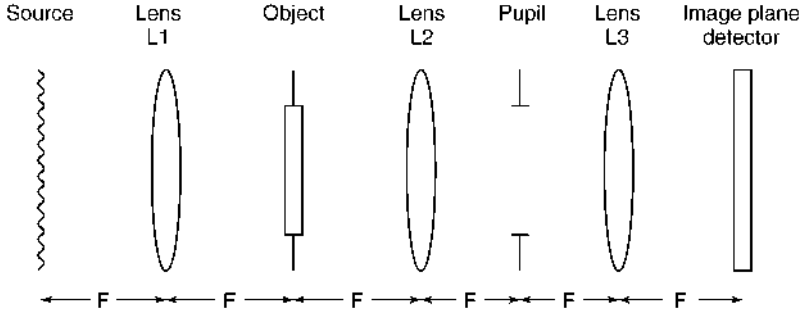

Figure 7 A telecentric Kohler illumination system with a spatially incoherent primary source imaged onto the pupil.

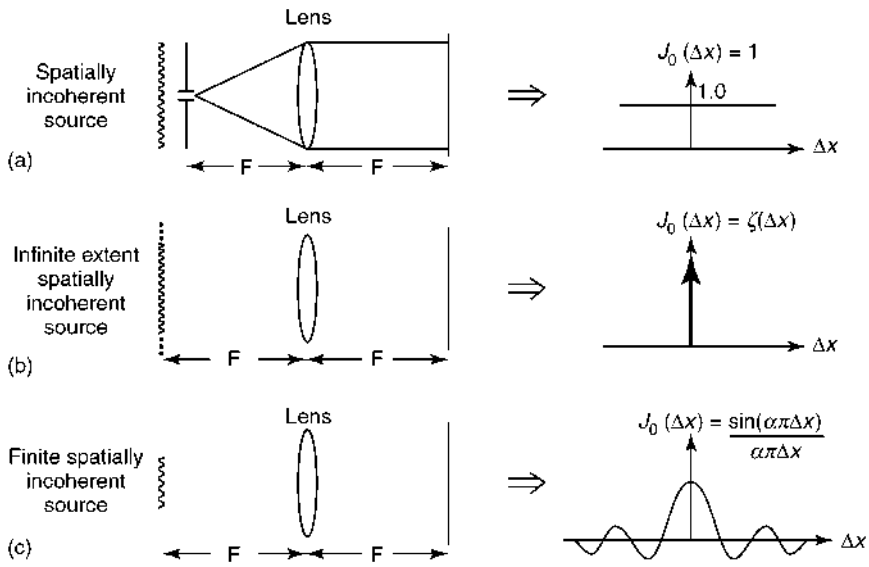

Figure 8 Light propages from the source plane to the object plane and produces (a) coherent illumination from a single source point; (b) incoherent illumination from an infinite extent source; and (c) partially coherent illumination from a finite extent primary source.

Figure 8 shows some simple examples of primary source distributions and the corresponding object plane mutual intensity function. Figure 8a assumes an infinitesimal point source and the corresponding mutual intensity is 1.0 for all possible pairs of object points. Figure 8b assumes an infinite extent primary source and results in a dirac delta function for the mutual intensity function. This means that there is no correlation between any two object points having a separation greater than zero so the object plane illumination is spatially incoherent. In Figure 5c, a finite extent uniform source gives a mutual intensity function of the form $\sin(\alpha\pi\Delta x)/(\alpha\Delta x)$. The finite-sized source corresponds to partially coherent imaging and shows that the response to pairs of points is affected by the spatial distribution of the source in a complicated way. Note that a large primary source corresponds to a large range of angular illumination at the object plane.

Varying the size of the source in the imaging system of Figure 8 will affect the spatial coherence of the illumination and hence the optical image intensity. Many textbook treatments discuss the imaging of two points separated by the Rayleigh resolution criterion which corresponds to the case where the first minimum of one point image coincides with the maximum of the adjacent point image. With a large source that provides effectively incoherent light, the two point image has a modest dip in intensity between the two points, as shown in Figure 9a. Fully coherent illumination of two points separated by the Rayleigh resolution produces a single large spot with no dip in intensity, as shown in Figure 9b. Varying the source size and hence the illumination spatial coherence produces a dip that is less than the incoherent intensity dip. This result is often used to suggest that coherent imaging gives poorer resolution than incoherent imaging. In fact, generalizations about two-point resolution can be misleading.

Recent developments in optical lithography have shown that coherence can be used to effectively increase two-point resolution beyond traditional diffraction limits. In Figure 9c, one of the pinholes in a coherent two-point imaging experiment has been modified with a phase shift corresponding to one half of a wavelength of the illuminating light. The two images add in wave amplitude and the phase shift creates a distinct null at the image plane and effectively enhances the two-point resolution. This approach is termed phase screen lithography and has been exploited to produce finer features in lithography by purposely introducing small phase shift masks at the object mask. In practice, the temporal and spatial coherence of the light is engineered to give sufficient coherence to take advantage of the two-point enhancement while maintaining sufficient incoherence to avoid speckle-like artifacts associated with coherent light.

Spatial Frequency Modeling of Imaging

Spatial frequency models of image formation are also useful in understanding how coherence affects image formation. A spatially coherent imaging system has a particularly simple spatial frequency model. In the

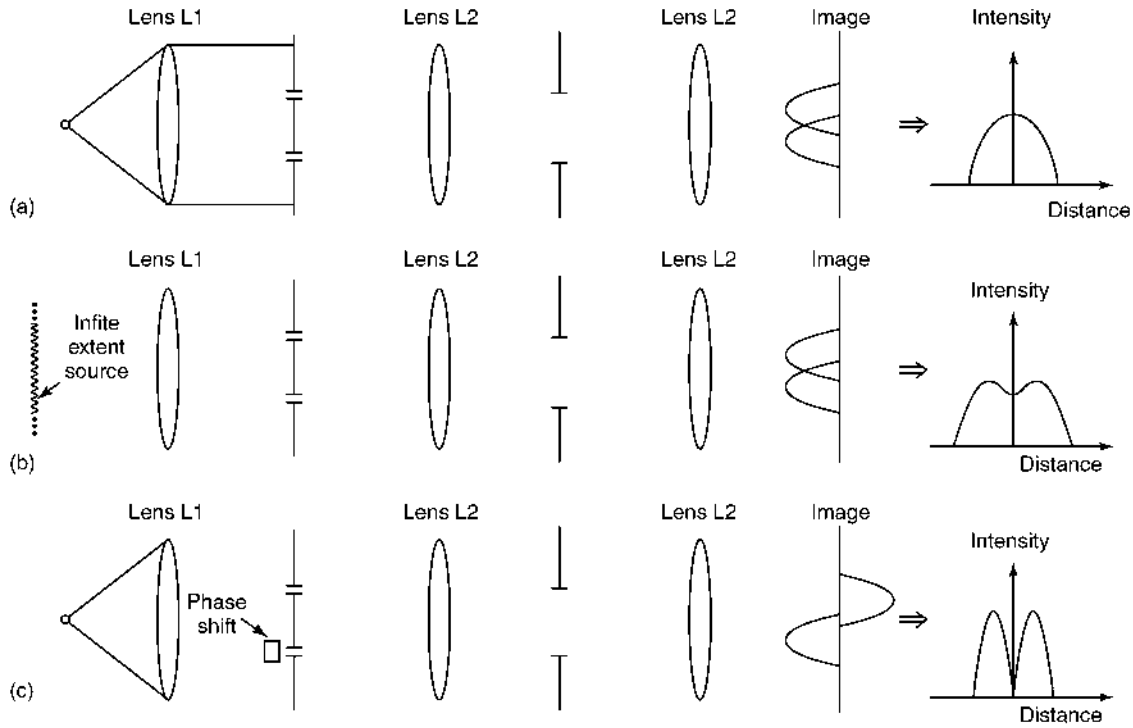

Figure 9 Comparison of coherent and two point imaging of two pinholes separated by the Rayleigh distance. (a) Coherent image cannot resolve the two points. (b) Incoherent image barely resolves the two points. (c) Coherent illumination with a phase shifting plate at one pinhole produces a null between the two image points.

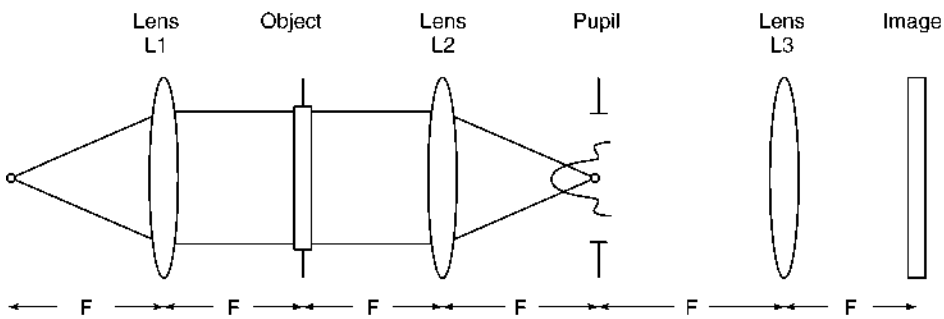

Figure 10 The object Fourier transform is filtered by the pupil function in a coherent imaging system.

spatially coherent imaging system shown in Figure 10, the on-axis plane wave illumination projects the spatial Fourier transform of the object, or Fraunhofer pattern, onto the pupil plane of the imaging system. The pupil acts as a frequency domain filter that can be modified to perform spatial frequency filtering. The complex transmittance of the pupil is the wave amplitude spatial frequency transfer function of the imaging system. It follows that the image plane wave amplitude frequency spectrum, $\tilde{U}_c(f_x)$, is given by

$$\tilde{U}_c(f_x) = \tilde{O}(f_x)H(f_x) \qquad [8]$$

where $\tilde{O}(f_x)$ is the spatial Fourier transform of the object amplitude transmittance function and $H(f_x)$ is proportional to the pupil plane amplitude transmittance. Equation [8] is the frequency domain version of the convolution representation given by eqn [2], but does not account for the squared magnitude response of the detector.

In the previous section, we learned that an infinite extent source is necessary to achieve fully incoherent imaging for the imaging system of Figure 7. As a thought experiment, one can start with a single point source on the axis and keep adding mutually incoherent source points to build up from a coherent imaging system to an incoherent imaging system. Since the individual source points are assumed to be incoherent with each other, the images from each point source can be added in intensity. In Figure 11,

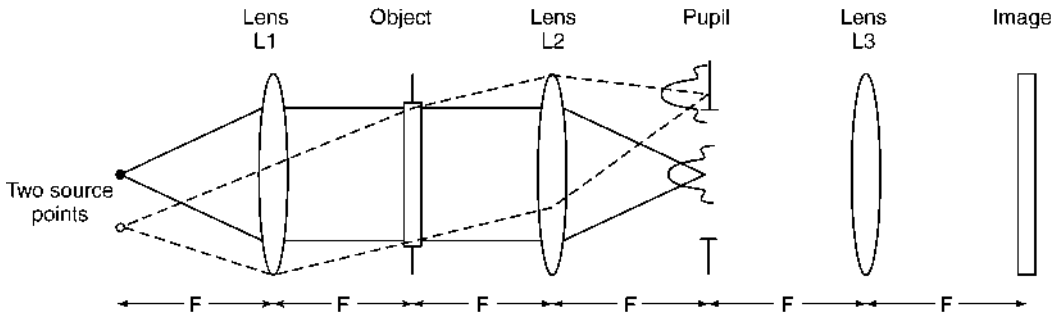

Figure 11 A second source point projects a second object Fourier transform onto the pupil plane. The images produced by the two source points add in intensity at the image plane.

only two point sources are shown. Each point source projects an object Fraunhofer pattern onto the pupil plane. The centered point source will result in the familiar coherent image. The off-axis point source projects a displaced object Fraunhofer pattern that is also filtered by the pupil plane before forming an image of the object. The image formed from this second point source can be modeled with the same coherent imaging model with a shifted pupil plane filter. Since the light from the two source points is uncorrelated, the final image is calculated by adding the intensities of the two coherently formed images.

The two source point model of Figure 11 can be generalized to an arbitrary number of source points. The final image is an intensity superposition of a number of coherent images. This suggests that partially coherent imaging systems behave as a number of redundant coherent imaging systems, each having a slightly different amplitude spatial frequency transfer function due to the relative shift of the pupil filter with respect to the object Fraunhofer pattern. As the number of point sources is increased to infinity, the primary source approaches an infinite extent spatially incoherent source. In practice, the source need not be infinite. When the source is large enough to effectively produce linear-in-intensity imaging, the imaging system is effectively spatially incoherent. The corresponding image intensity in the spatial frequency domain, $\tilde{I}_{\mathrm{inc}}(f_x)$, is given by

$$\tilde{I}_{\mathrm{inc}}(f_x) = \tilde{I}_{\mathrm{obj}}(f_x)\mathrm{OTF}(f_x) \qquad [9]$$

where $\tilde{I}_{\mathrm{obj}}(f_x)$ is the spatial Fourier transform of the object intensity transmittance and $\mathrm{OTF}(f_x)$ is the incoherent optical transfer function which is proportional to the spatial autocorrelation of the pupil function.

Many texts include detailed discussions comparing the coherent transfer function and the OTF and attempt to make comparisons about the relative performance of coherent and incoherent systems. These comparisons are often misleading since one transfer function describes the wave amplitude spatial frequency transfer function and the other describes the intensity spatial frequency transfer function. Here we note that incoherent imaging systems do indeed allow higher object amplitude spatial frequencies to participate in the image formation process. This argument is often used to support the claim that incoherent systems have higher resolution. However, both systems have the same image intensity spatial frequency cutoff. Furthermore, the nature of the coherent transfer function tends to produce high contrast images that are typically interpreted as higher-resolution images than their incoherent counterparts. Perhaps the real conclusion is that the term resolution is not well defined and direct comparisons between coherent and incoherent imaging must be treated carefully.

The frequency domain treatment for partially coherent imaging of two-dimensional objects involves a four-dimensional spatial frequency transfer function that is sometimes called the bilinear transfer function or the transmission cross-coefficient model. This model describes how constituent object wave amplitude spatial frequencies interact to form image plane intensity frequencies. The utility of this approach to analysis is limited for someone new to the field, but is often used in numerical simulations of partially coherent imaging systems used in optical lithography.

Experimental Examples of Important Coherence Imaging Phenomena

Perhaps the best way to gain an understanding of coherence phenomena in imaging is to examine experimental results. In the following section we use experimental data to see how coherence affects edge response, noise immunity, and depth of field. Several experimental configurations were used to collect

the image data, but all of the systems can be represented generically by the Kohler illumination system shown in Figure 12. Kohler illumination is often employed in many projection illumination systems. In order to obtain Kohler illumination, the primary spatially incoherent source is imaged onto the pupil plane of the imaging portion of the system and the object is placed at the pupil plane of the condenser optics.

Primary Source Generation

Figure 13 shows a highly coherent illumination system produced by focusing a laser beam to a point and imaging the focused spot onto the pupil of the imaging system to produce spatially coherent illumination. The use of a laser produces highly temporally coherent light.

Two methods were used to create an extended spatially incoherent primary source with control over the spatial intensity distribution. Figure 14a shows a collimated laser beam with a 633 nm center wavelength illuminating a rotating diffuser. A photographically produced mask defines the spatial shape of the primary source. The laser provides highly temporally coherent light and the diffuser destroys the spatial coherence of the light. Consider the thought experiment of two pinholes placed immediately to the right of the moving diffuser. Without the diffuser, the two wave amplitudes would be highly correlated. Assuming that the diffuser can be modeled as a spatially random phase plate, a fixed diffuser would only introduce a fixed phase difference between the amplitudes leaving the two pinholes and would not destroy the coherence. When the diffuser is rotated, the light from each pinhole encounters a different phase modulation that is changing over time. This random modulation destroys the effective correlation between the wave amplitude of the light leaving the two pinholes provided that the rotation speed is sufficiently fast.

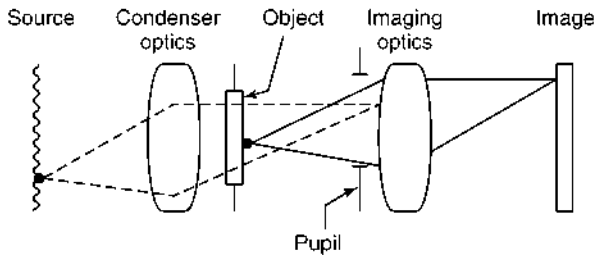

Figure 12 General representation of the Kohler illumination imaging systems used in the experimental result section. The primary incoherent source is imaged onto the imaging system pupil plane and the object resides in the pupil of the condenser optics.

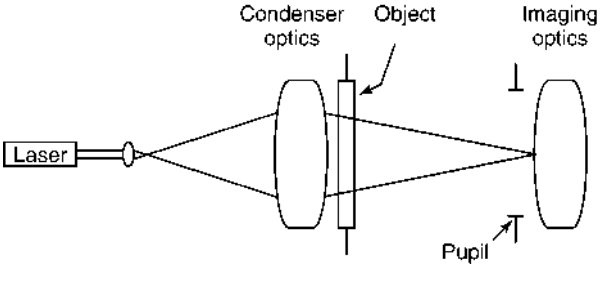

Figure 13 Highly spatially and temporally coherent illumination produced by imaging a focused laser beam into the imaging system pupil.

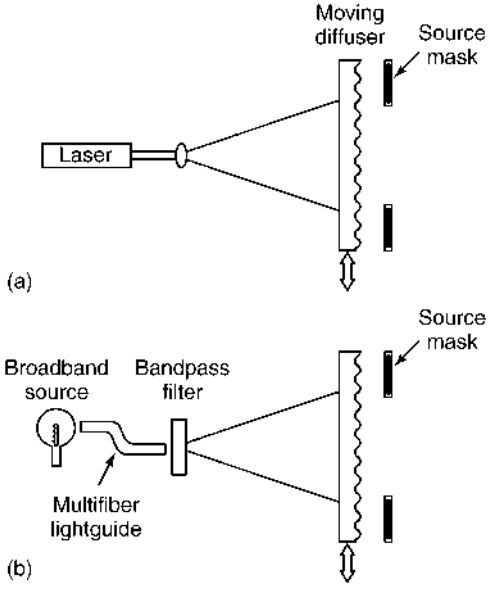

Figure 14 Two methods for generating a spatially incoherent primary source. (a) An expanded laser beam passes through a moving diffuser followed by a mask to define the extent and shape of the source. (b) A broadband source exits a multifiber lightguide and propagates to a moving diffuser followed by a source mask.

The moving diffuser method is light inefficient but is a practical way of exploring coherence in imaging in a laboratory environment. The choice of the diffuser is critical. The diffuser should spread light out uniformly over an angular subtense that overfills the object of interest. Many commercially available diffusers tend to pass too much light in the straight through direction. Engineered diffusers can be purchased to produce an optimally diffused beam. When quick and inexpensive results are required, thin plastic sheets used in day-to-day packaging often serve as excellent diffusers. When the diffusion angles are not high enough, a number of these plastic sheets can be layered on top of each other to achieve the appropriate angular spread.

The second method for producing a spatially incoherent source is shown in Figure 14b. Broadband

light is delivered through a large multifiber lightguide and illuminates a moving diffuser. The main purpose of the diffuser is to ensure that the entire object is illuminated uniformly. The low temporal coherence of the source and the long propagation distances are usually enough to destroy any spatial coherence at the plane of the diffuser. The rotation further ensures that no residual spatial coherence exists at the primary source plane. A chromatic filter centered around 600 nm, with a spectral width of approximately 100 nm, is shown in the light path. The filter was used to minimize the effects of chromatic aberrations in the system. The wide spectral width certainly qualifies as a broadband source in relation to a laser. High-contrast photographically defined apertures determine the spatial intensity distribution at the plane of the primary source. In the following results, the primary source distributions were restricted to circular sources of varying sizes as well as annular sources. These shapes are representative of most of the sources employed in microscopy, machine vision, and optical lithography.

Noise Immunity

Coherent imaging systems are notorious for introducing speckle-like noise artifacts at the image. Dust and optically rough surfaces within an optical system result in a complicated textured image plane intensity distribution that is often called speckle noise. We refer to such effects as coherent artifact noise. Figure 15a shows the image of a standard binary target as imaged by a benchtop imaging system, with the highly coherent illumination method shown in Figure 13. Some care was taken to clean individual lenses and optical surfaces within the system and no dust was purposely introduced. The image is corrupted by a complicated texture that is due in

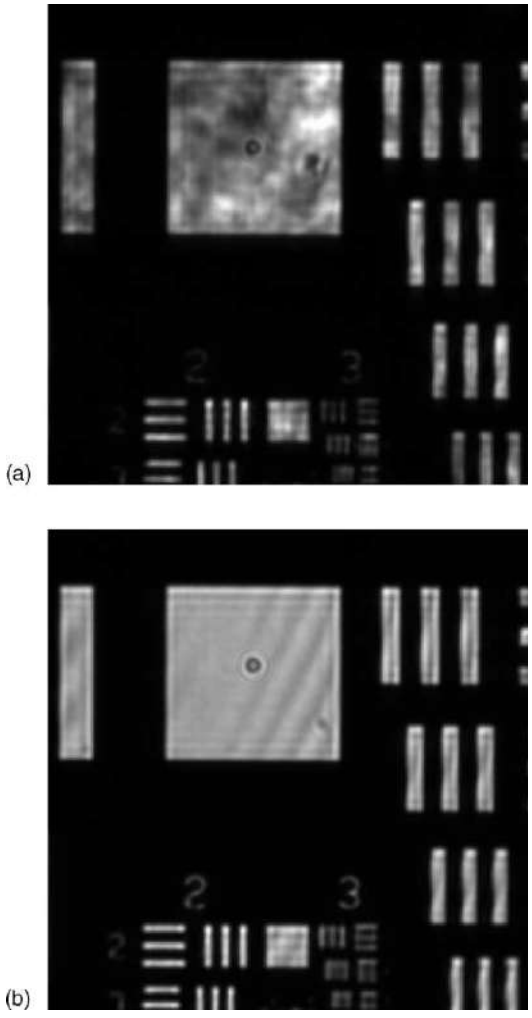

Figure 15 High temporal coherence imaging with disk sources corresponding to (a) extremely high spatial coherence ($K = 0.05$) and (b) slightly reduced but high spatial coherence ($K = 0.1$).

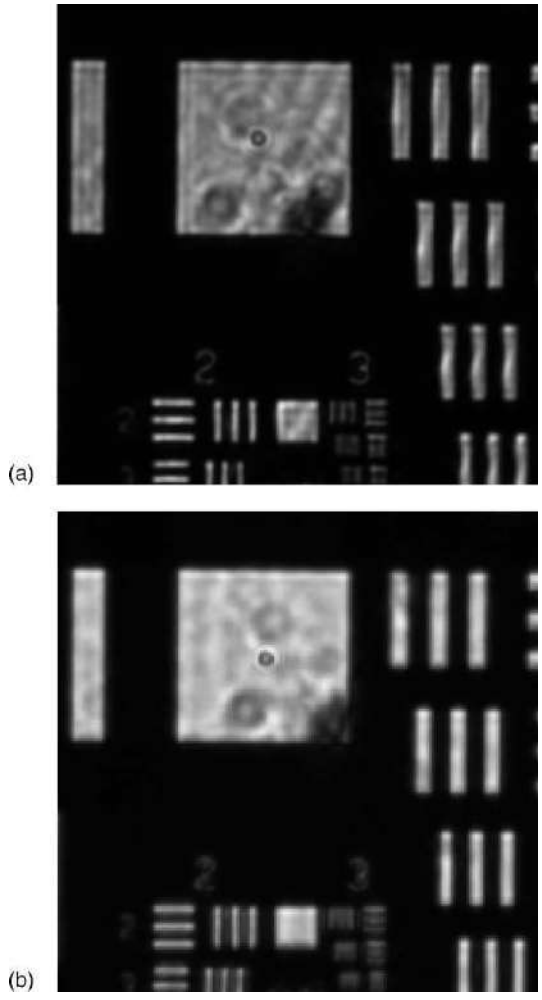

Figure 16 High spatial coherence disk illumination ($K = 0.1$) imaging through a dust covered surface with (a) narrowband laser illumination and (b) with broadband illumination.

part to imperfections in the laser beam itself and in part to unwanted dust and optically rough surfaces. Certainly, better imaging performance can be obtained with more attention to surface cleanliness and laser beam filtering, but the result shows that it can be difficult to produce high-quality coherent images in a laboratory setting.

Lower temporal coherence and lower spatial coherence will reduce the effect of speckle noise. The image of Figure 15b was obtained with laser illumination according to Figure 14a with a source mask corresponding to a source to pupil diameter ratio of $K = 0.1$. The object plane illumination is still highly temporally coherent and the spatial coherence has been reduced but is still very high. The artifact noise has been nearly eliminated by the modest reduction of spatial coherence. The presence of diagonal fringes in some regions is a result of multiple reflections produced by the cover glass in front of the CCD detector. The reflections produce a weak secondary image that is slightly displaced from the main image and the high temporal coherence allows the two images to interfere.

The noise performance was intentionally perturbed in the image of Figure 16a by inserting a slide with a modest sprinkling of dust at an optical surface in between the object and the pupil plane. The illumination conditions are the same as for the image of Figure 15b. The introduction of the dust has further degraded the image. The image of Figure 16b maintains the same spatial coherence ($K = 0.1$) but employs the broadband source. The lower temporal coherence eliminates the unwanted diagonal fringes but the speckle noise produced by the dust pattern does not improve significantly relative to the laser illumination $K = 0.1$ system. Figures 17a–c show that increasing the source size (and hence the range of angular illumination) reduces

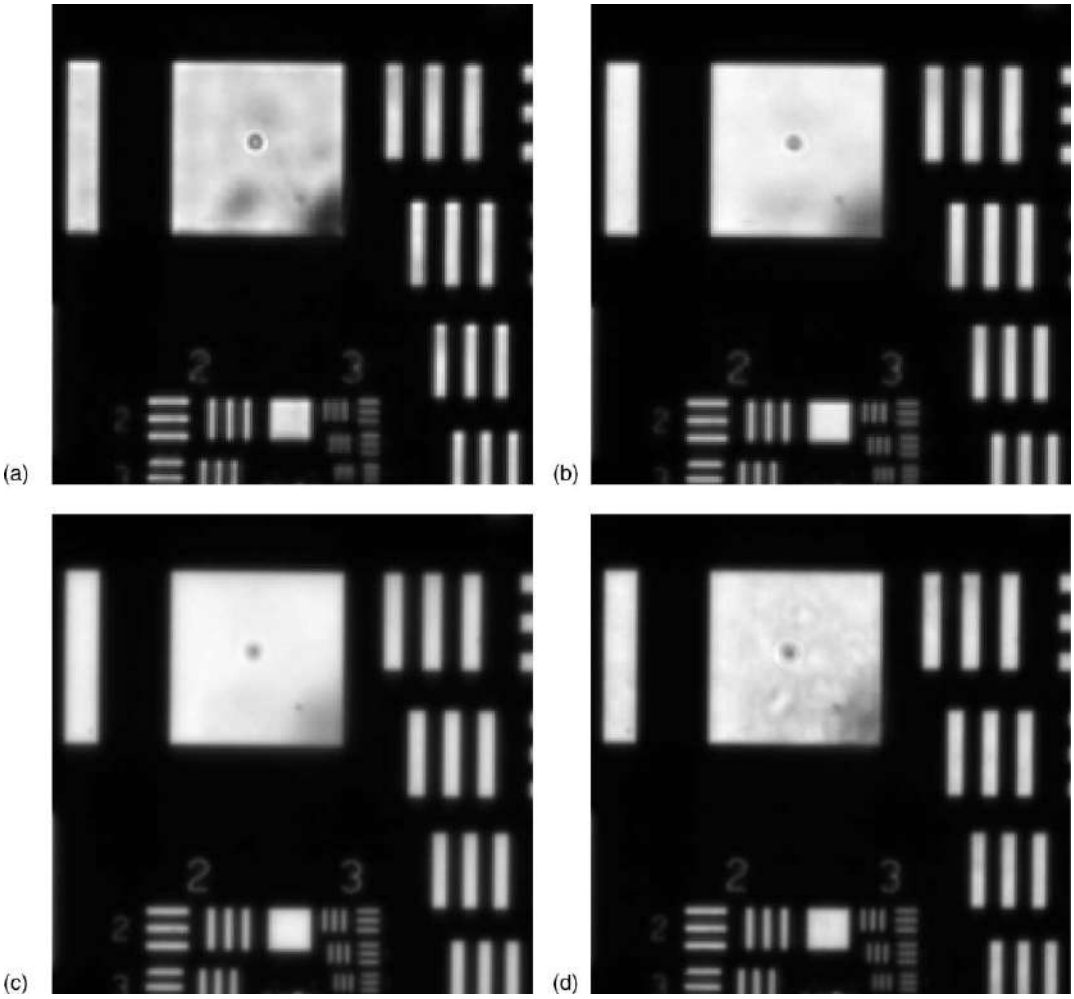

Figure 17 Low temporal coherence imaging through a dust-covered surface with various source sizes and shapes: (a) disk source with $K = 0.3$; (b) disk source with $K = 0.7$; (c) disk source with $K = 2.0$; and (d) a thin annular source with an outer diameter corresponding to $K = 0.5$ and inner diameter corresponding to $K = 0.45$.

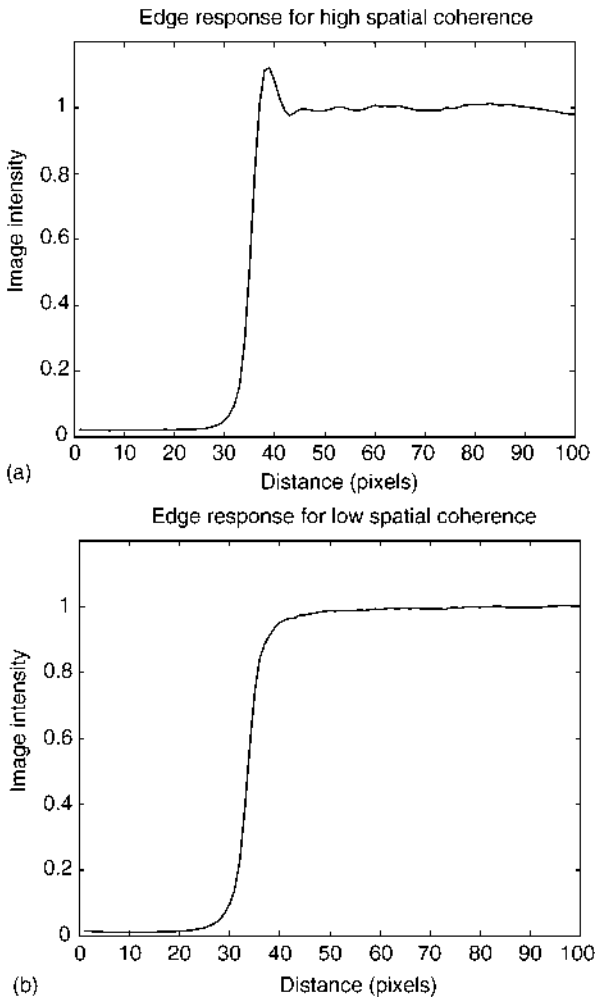

Figure 18 Slices of edge intensity images of high contrast edges taken by a laboratory system with low temporal coherence and (a) high spatial coherence (disk source with $K = 0.1$) and (b) effectively incoherent (disk source with $K = 2$).

the spatial coherence and virtually eliminates the unwanted speckle pattern. Finally, Figure 17d shows that an annular source can also provide some noise immunity. The amount of noise immunity is related to the total area of the source rather than the outer diameter of annular source. Seen from the spatial frequency model, each source point produces a noise-corrupted coherent image. Since the effect of the noise is different for each coherent image, the image plane noise averages out as the images add incoherently. This perspective suggests that an extended source provides redundancy in the transfer of object information to the image plane.

Edge Response

Spatial coherence has a strong impact on the images of edges. The sharp spatial frequency cutoff of spatially coherent imaging systems creates oscillations at the images of an edge. Figure 18 shows slices of experimentally gathered edge imagery as a function of the ratio of source diameter to pupil diameter. Higher coherence systems produce sharper edges, but tend to have overshoots. The sharper edges contribute to the sense that high-coherence systems produce higher-resolution images. One advantage to an incoherent imaging system is that the exact location of the edge corresponds to the 50% intensity location. The exact location of an edge in a partially coherent imaging system is not as easily determined. The presence or lack of edge ringing can be used to assess whether a given imaging system can be modeled as spatially incoherent or partially coherent.

Depth of Field

Coherent imaging systems exhibit an apparent increase in depth-of-field compared to spatially incoherent systems. Figure 19 shows spatially incoherent imagery for four different focus positions. Figure 20 shows imagery with the same amount of defocus produced with highly spatially coherent light. Finally, Figure 21 gives an example of how defocused imagery depends on the illumination coherence. The images of a spoke target were all gathered with a fixed amount of defocus and the source size was varied to control the illumination coherence. The images of Figure 21 differ greatly, even though the CPSF was the same for all the cases.

Digital Post-detection Processing and Partial Coherence

The model of Figure 2 suggests that post-detection image processing can be considered as part of the image formation system. Such a general view can result in imaging that might be otherwise unobtainable by classical means. In fact, microscopists routinely use complicated deblurring methods to reconstruct out-of-focus imagery and build up three-dimensional images from multiple image slices.

The coherence of the illumination should be considered when undertaking such image restoration. Rigorous restoration of partially coherent imagery is computationally intensive and requires precise knowledge of the coherence. In practice, a linear-in-intensity model for the image formation is almost always used in developing image restoration algorithms. Even nonlinear restoration algorithms have built-in linear assumptions about the image formation models which imply spatially incoherent illumination.

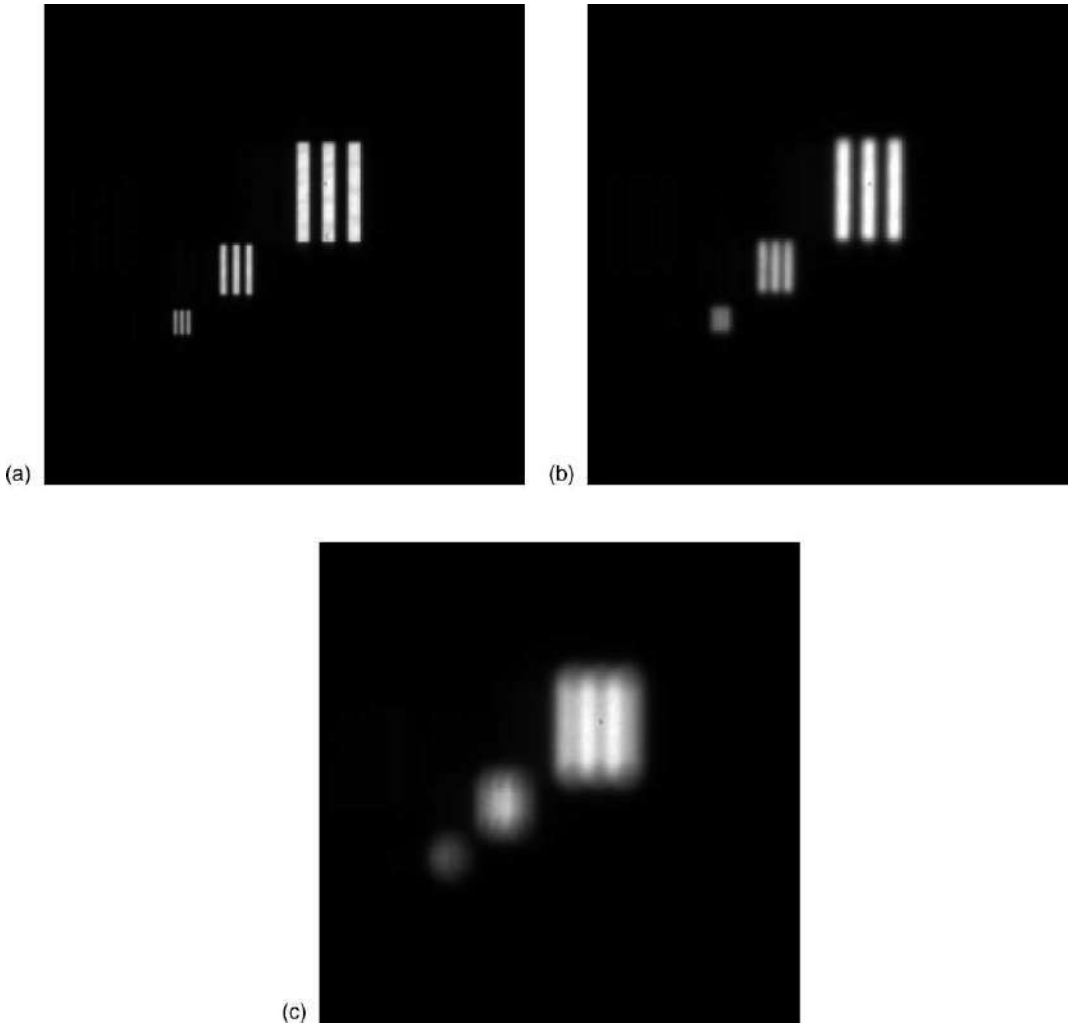

Figure 19 Spatially incoherent imaging for (a) best focus, (b) moderate misfocus, and (c) high misfocus.

The images of Figures 22a and b are digitally restored versions of the images of Figure 19b and Figure 20b. The point spread function was directly measured and used to create a linear restoration filter that presumed spatially incoherent illumination. The restored image in Figure 22a is faithful since the assumption of spatially incoherent light was reasonable. The restored image of Figure 22b suffers from a loss in fidelity since the actual illumination coherence was relatively high. The image is visually pleasing and correctly conveys the presence of three bar targets. However, the width of the bars is not faithful to the original target which had spacings equal to the widths of the bars.

As discussed earlier, the ideal image is not always the optimal image for a given task. In general, restoration of partially coherent imaging with an implicit spatially incoherent imaging model will produce visually pleasing images that are not necessarily faithful to the object plane intensity. While they are not faithful, they often preserve and even enhance edge information and the overall image may appear sharper than the incoherent image. When end task performance is improved, the image may be considered to be more optimal than the ideal image. It is important to keep in mind that some of the spatial information may be misleading and a more complete understanding of the coherence may be needed for precise image plane measurements. This warning is relevant in microscopy where the user is often encouraged to increase image contrast by reducing illumination spatial coherence. Annular sources can produce highly complicated spatial coherence functions that will strongly impact the restoration of such images.

Summary and Discussion

The coherence of the illumination at the object plane is important in understanding image formation.

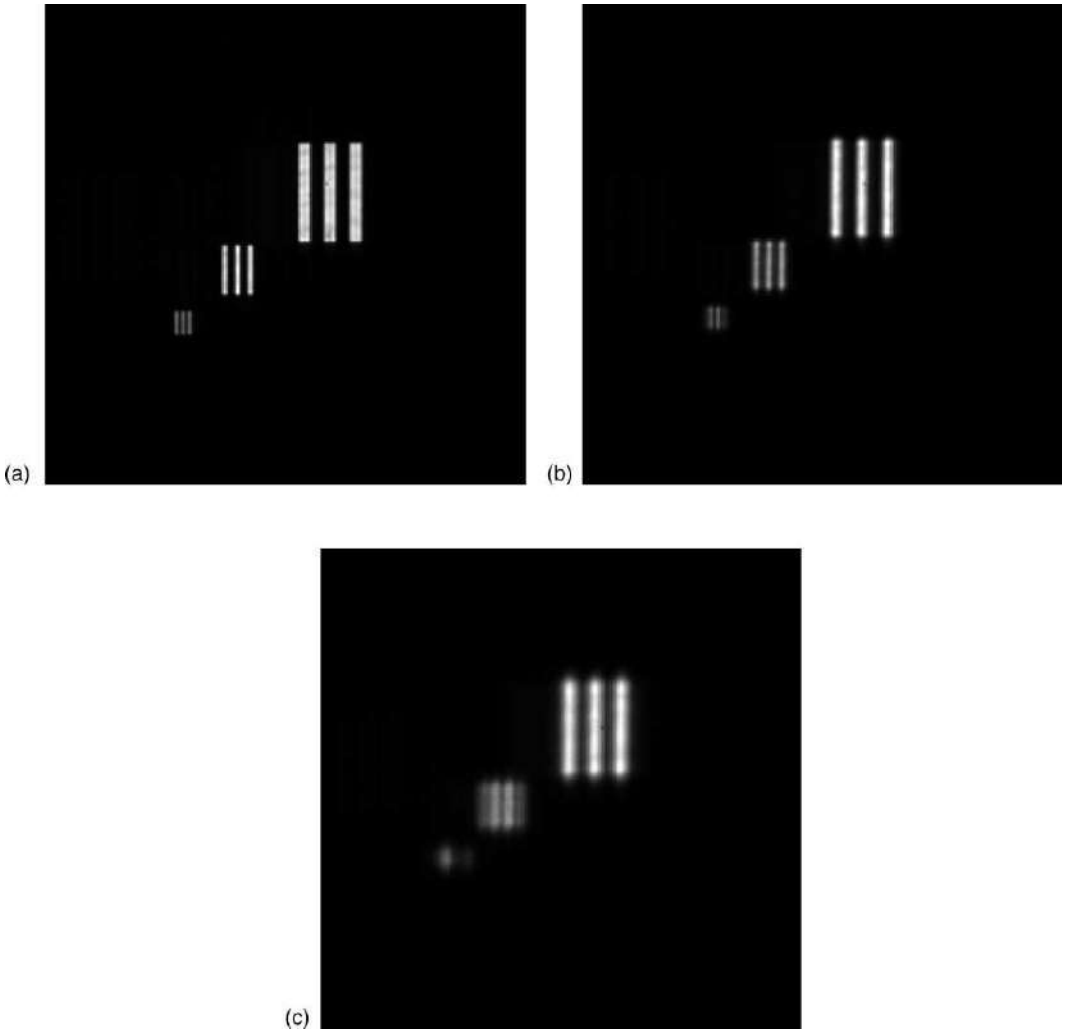

Figure 20 High spatial coherence disk source ($K = 0.2$) imaging for (a) best focus, (b) moderate misfocus, and (c) high misfocus.

Highly coherent imaging systems produce high-contrast images with high depth of field and provide the opportunity for sophisticated manipulation of the image with frequency plane filtering. Darkfield imaging and phase contrast imaging are examples of frequency plane filtering. Unfortunately, coherent systems are sensitive to optical noise and are generally avoided in practical system design. Relaxing the temporal coherence of the illumination can provide some improvement, but reducing the spatial coherence is more powerful in combating noise artifacts. Research microscopes, machine vision systems, and optical lithography systems are the most prominent examples of partially coherent imaging systems. These systems typically employ adjustable spatially incoherent extended sources in the shapes of disks or annuli. The exact shape and size of the primary source, shapes the angular extent of the illumination at the object plane and determines the spatial coherence at the object plane. Spatial coherence effects can be significant, even for broadband light. Control over the object plane spatial coherence allows the designer to find a tradeoff between the various strengths and weaknesses of coherent and incoherent imaging systems.

As more imaging systems employ post-detection processing, there is an opportunity to design fundamentally different systems that effectively split the image formation process into a physical portion and a post-detection portion. The simple example of image deblurring cited in this article shows that object plane coherence can affect the nature of the digitally restored image. The final image can be best understood when the illumination coherence effects are well understood. The spatially incoherent case results in the most straightforward model for image interpretation, but is not always the best choice since the coherence can often be manipulated to increase the contrast and hence the amount of useful information in the raw image. A completely general approach to

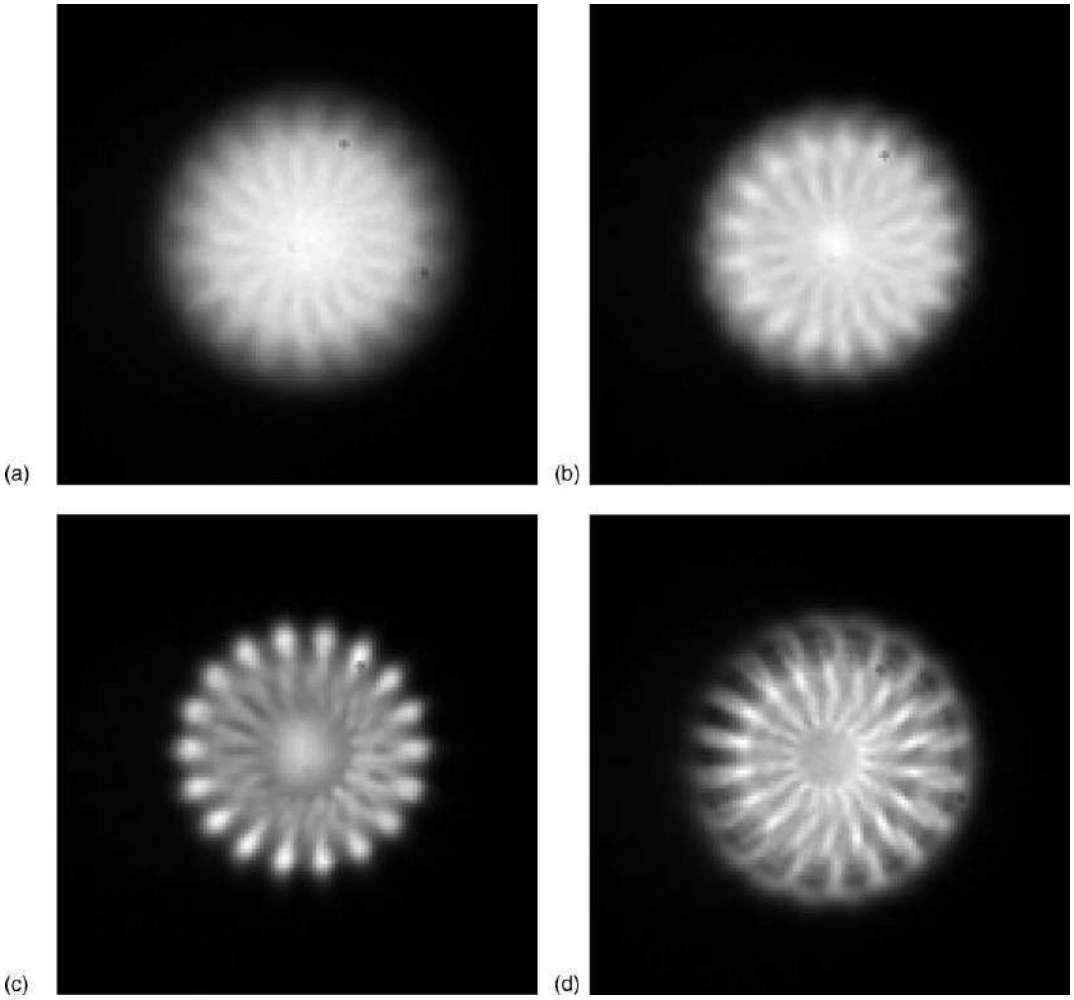

Figure 21 Imaging with a fixed amount of misfocus and varying object spatial coherence produced by: (a) highly incoherent disk source illumination ($K = 2$); (b) moderate coherence disk source ($K = 0.5$); (c) high coherence disk source ($K = 0.2$); and (d) annular source with outer diameter corresponding to $K = 0.5$ and inner diameter corresponding to $K = 0.45$.

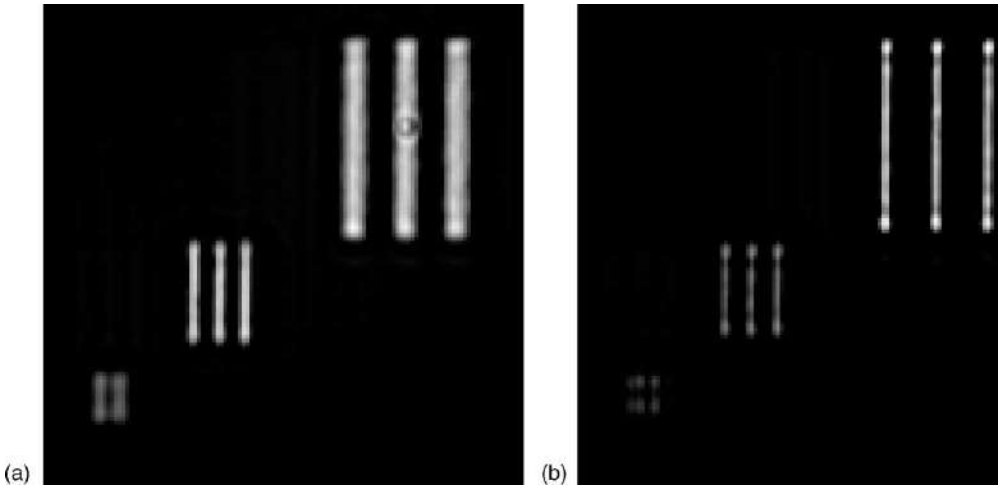

Figure 22 Digital restoration of blurred imagery with an inherent assumption of linear-in-intensity imaging for (a) highly spatially incoherent imaging ($K = 2$) and (b) high spatial coherence ($K = 0.2$).

imaging would treat the coherence of the source, the imaging optics, and the post-detection restoration all as free variables that can be manipulated to produce the optimal imaging system for a given task.

See also

Coherent Lightwave Systems. **Information Processing:** Coherent Analogue Optical Processors. **Terahertz Technology:** Coherent Terahertz Sources.

Further Reading

Bartelt H, Case SK and Hauck R (1982) Incoherent Optical Processing. In: Stark H (ed.) *Applications of Optical Fourier Transforms*, pp. 499–535. London: Academic Press.

Born M and Wolf E (1985) *Principles of Optics*, 6th edn. (corrected), pp. 491–554. Oxford, UK: Pergamon Press.

Bracewell RN (1978) *The Fourier Transform and its Applications*. New York: McGraw Hill.

Gonzales RC and Woods EW (1992) *Digital Image Processing*, 3rd edn., pp. 253–304. Reading, MA: Addison-Wesley.

Goodman JW (1985) *Statistical Optics*, 1st edn. New York: John Wiley & Sons.

Goodman JW (1996) *Introduction to Fourier Optics*, 2nd edn. New York: McGraw Hill.

Hecht E and Zajac A (1979) *Optics*, 1st edn., pp. 397–435, Reading, MA: Addison-Wesley.

Inoue S and Spring KR (1997) *Video Microscopy*, 2nd edn. New York: Plenum.

Reynolds GO, DeVelis JB, Parrent GB and Thompson JB (1989) *Physical Optics Notebook: Tutorials in Fourier Optics*, 1st edn. New York: SPIE, Bellingham and AIP.

Saleh BEA and Teich MC (1991) *Fundamentals of Photonics*, pp. 343–381. New York: John Wiley & Sons.

Speckle and Coherence

G Häusler, University of Erlangen-Nürnberg, Erlangen, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Opticists are aware that the amount of coherence plays a significant role in imaging systems: laser speckles are known to add significant noise to the image, as well as parasitic interferences from dusty lenses. Optical systems are often called coherent, if a laser is used (right), and incoherent if other sources come into play (wrong). Many users of optical systems are unaware that it is not the high temporal coherence but the spatial coherence that commonly afflicts the image quality, and that this parasitic spatial coherence is ubiquitous, even though not obvious. Coherent artifacts can occur without the use of lasers, although speckle noise is more prominent with lasers. Even opticists sometimes underestimate the damage that residual coherent noise can cause, and as laser oriented sensor funding programs are 'en vogue', nonexperts are disappointed if some metrology device does not include a laser.

This encyclopedia addresses the many uses of lasers. In this article, we will discuss the costs of coherence. The commonly pretended incoherent approach of everyday optics may lead to significant quantitative measuring errors of illumination or reflectivity, 3d shape, distance or size. Spatial coherence is the dominant source of noise. We will give some rules of thumb to estimate these errors and a few tricks to reduce coherent noise. These rules will help to minimize coherent noise, however, it turns out that as spatial coherence is ubiquitous, there are only limited options available to clear it. One of the options to build good sensors that measure shape, reflectivity, etc. is avoiding the use of lasers!

To become familiar with some basics of the theory of coherence we refer the reader the Further Reading section at the end of this article.

Coherence can be boon or disaster for the opticist, as is explained in other articles of this encyclopedia about interferometry, diffraction, and holography. A specific topic is information acquisition from coherently scattered light. An enlightening example is where speckles in white light interferometry at rough surfaces and in speckle interferometry are exploited. We will briefly discuss white light interferometry at rough surfaces in the section on speckles as carriers of information below.

Practical Coherence Theory

A major issue of this chapter will be corrupting properties of coherence in the daily life of an optical metrologist. We will demonstrate that 'speckle' noise is ever present, and essentially unavoidable, in the images of (diffusively reflecting) objects. Its influence on the quality of optical measurements leads to a lower limit of the physical measuring uncertainty.

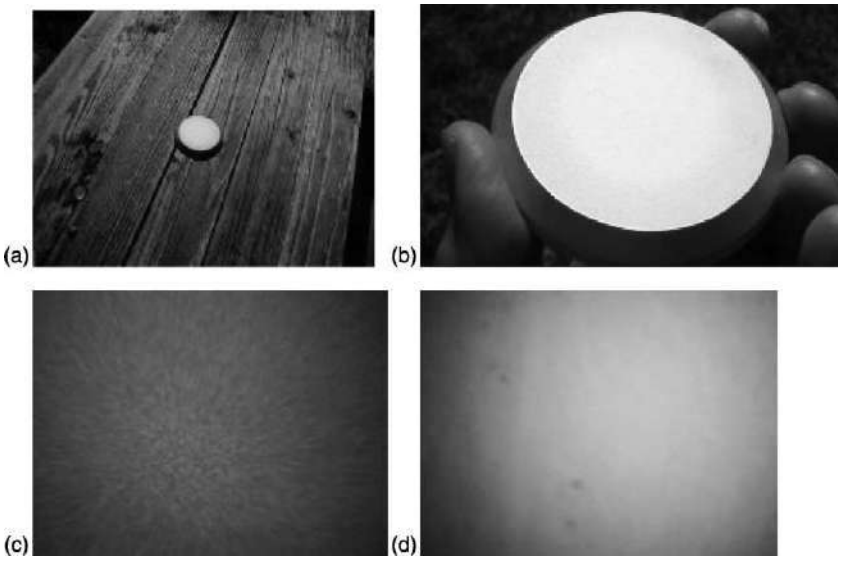

Figure 1 Ground glass in sunlight: (a) the image of a ground glass, illuminated by the sun, observed with small aperture (from large distance), displays no visible speckles; (b) a medium observation aperture (smaller distance) displays weak visible speckles; (c) observation with very high aperture displays strong contrast speckles; and (d) the image of the ground glass on a cloudy day does not display speckles, due to the very big aperture of illumination. Reproduced with permission from Häusler G (1999) Optical sensors and algorithms for reverse engineering. In: Donah S (ed.) Proceedings of the 3rd Topical Meeting on Optoelectronic Distance/Displacement Measurements and Applications. IEEE-LEOS.

This 'coherent noise' limit is, not surprisingly, identical to Heisenberg's limit. We will start by summarizing the important results, and will give some simple explanations later.

Some Basic Observations

(i) The major source of noise in optical measurements is spatial coherence, the temporal coherence status, generally, does not play a significant role.
(ii) An optically rough surface displays subjective speckles with contrast $C = 1$, if the observation aperture $\sin u_o$ is larger than the illumination aperture $\sin u_q$.
(iii) The speckle contrast is $C \sim \sin u_q / \sin u_o$, if the observation aperture is smaller than the illumination aperture.

Figure 1 illustrates the situation by a simple experiment.

The results of Figure 1 can be explained simply by summarizing some 80 pages of coherence theory in a nutshell, by useful simplifications and approximations (Figure 2).

If the source is at an infinite distance, the coherence function no longer depends on the two variables but just on the slit distance $d = |x_1 - x_2|$. In this case, the coherence function $\Gamma(d)$ can be easily calculated as the Fourier transform of the spatial intensity

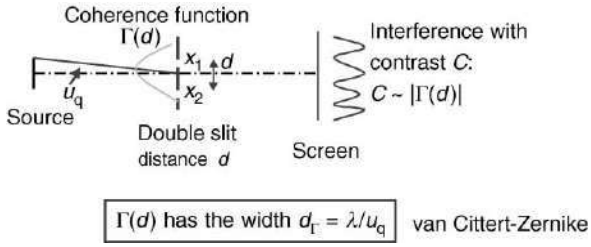

Figure 2 Basic experiment for spatial coherence. An extended source, such as the sun or some incandescent lamp illuminates a double slit, from a large distance. On a screen behind the double slit we can observe 'double slit interference', if the waves coming from x_1 and x_2 display coherence. The contrast of the interference fringes is given by the magnitude of the coherence function Γ (neglecting some normalization factor). Γ is a function of the two slit locations x_1 and x_2.

distribution of the source (van Cittert-Zernike theorem). Let us assume a one dimensional source with an angular size of $2u_q$, where u_q is commonly called the aperture of the illumination. In this case, the coherence function will be:

$$\Gamma(d) \sim \sin c(2u_q d/\lambda) \qquad [1]$$

The width of the coherence function which gives the size of the coherently illuminated area (coherence area) can be approximated from eqn [1]:

$$d_\Gamma = \lambda / u_q \qquad [2]$$

Equation [1] includes some approximations: for circular sources such as the sun, we get an additional

factor 1.22 for the width of the coherence function – the 'sinc-function' has to be replaced by the Bessel function $J_1(r)/(r)$. There is another approximation: we should replace the aperture angle u_q by $\sin u_q$, for larger angles. We choose the wavelength of the maximum sensitivity of the eye, at $\lambda = 0.55\ \mu m$, which is the wavelength of the maximum emission of the sun also.

In conclusion, if two points at the object are closer than d_Γ these points will be illuminated with a high degree of coherence. The light waves, scattered from these object-locations can display interference contrast. Specifically, they may cancel each other out, causing a low signal-to-noise ratio of only 1:1. This does not happen if the points have a distance larger than d_Γ.

From Figure 3 we learn that the width of the spatial coherence function from sunlight illumination at the earth's surface is ~110 μm. (Stars more distant than the sun appear smaller, and hence have a wider coherence function, at the earth's surface. Michelson was able to measure the angular size of some close stars, by measuring the width of this coherence function, which was about $d_\Gamma \sim 10$ m.)

Figure 4 again displays an easy to perform experiment. We can see speckles at sunlight illumination. We can observe speckles as well in shops because they often use small halogen spot illumination.

So far we have only discussed the first stage (from the source to the object). We still have to discuss the way from the object to the retina (or CCD-chip of a video camera). This is illustrated by Figure 5.

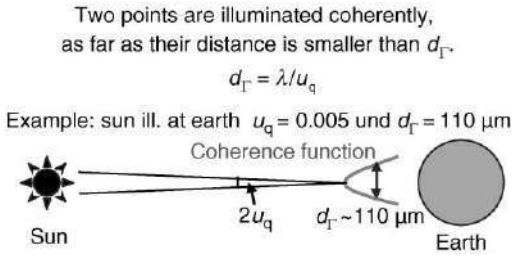

Figure 3 Coherence from sunlight illumination. With the van Cittert–Zernike theorem and an approximate illumination aperture of the sun $u_q^{sun} \sim 0.005$ (0.25°), we get a width of the coherence function at the Earth's surface of about $d_\Gamma \sim 110\ \mu m$.

Figure 4 Finger nail in sunlight, with speckles.

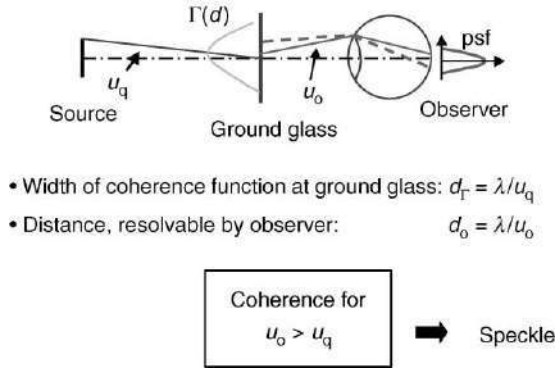

Figure 5 Coherence in a nutshell: spatial coherence and observer.

As an object, we choose some diffusely reflecting or transmitting surface (such as a ground glass). The coherence function at the ground glass has again a width of d_Γ. What happens if we image this ground glass onto our retina? Let us assume our eye to be diffraction limited (which, in fact, it is at sunlight conditions, where the eye pupil has only a diameter $\Phi_{pupil} = 2.5$ mm diameter or less). Then a point x_1 at the ground glass will cause a diffraction spot at the retina, with a size of:

$$d'_{diffr} = \lambda/u'_o = 8\ \mu m \qquad [3]$$

The image sided aperture of observation u'_0 of the eye is calculated from Φ_{pupil} and from its focal length $f_{eye} = 18$ mm as $u'_o = \Phi_{pupil}/2f_{eye} = 0.07$. If we project the diffraction spot at the retina back onto the object, we get its size d from:

$$d_{diffr} = \lambda/u_o = \lambda/(\Phi_{pupil}/2z_o) \qquad [4]$$

with z_o = observation distance of the object from the eye. We call $u_o = \Phi_{pupil}/2z_o$ the object sided observation aperture. Let us calculate the laterally resolvable distance d_{diffr} at the object, with a distance $z_o = 250$ mm, which is called the 'comfortable viewing distance':

$$d_{diffr}(z_o = 250\ mm) = 110\ \mu m \qquad [5]$$

After these preparations we come to the crucial issue: how does the image at the retina appear if we cannot resolve distances at an object smaller than d_Γ, and how does it appear if the resolution of the eye is sufficient to resolve distances smaller than d_Γ?

Let us start with the first assumption: $d_{diffr} \gg d_\Gamma$. Now the images of points at the object, over an area of the diffraction spot are more or less incoherently averaged at the retina, so we will see little interference contrast or 'speckles'. From eqn [4] we see that at

this incoherent averaging starts for distances larger than 250 mm, which is the comfortable viewing distance. Adults cannot see objects from much closer distances. We generally do not see speckles, not just at the limit of maximum accommodation. For larger distances, if we take $z_o = 2500$ mm, the resolvable distance at the object will be 1.1 mm, which is 10 times larger than the diameter of the coherence area. Averaging over such a large area will drop the interference contrast by a factor of 10. Note that such a small interference contrast might not be visible, but it is not at zero!

Coming to the second assumption we can laterally resolve distances smaller than d_Γ. In order to understand this, we first have to learn what is an 'optically rough' object surface. Figure 6 illustrates the problem.

A surface commonly is called 'rough' if the local height variations are larger than λ. However, a surface appears only rough, if the height variation within the distance that can be resolved by the observer, is larger than λ (reflected light will travel the distance twice, hence a roughness of $\lambda/2$ will be sufficient). Then the scattered different phasors, or 'Huygens elementary waves':

$$u_s \sim \exp[i2kz(x, y)] \qquad [6]$$

scattered from the object at x, y may have big phase differences so that destructive interferences (and speckle) can occur in the area of the diffraction image. So the attribute 'rough' depends on the object as well as on the observation aperture. With a high observation aperture (microscope), the diffraction image is small and within that area the phase differences might be small as well. So a ground glass may then look locally like a mirror, while with a small observation aperture it appears 'rough'.

With the 'rough' observation mode assumption, we can summarize what we have assumed: with resolving distances smaller than the coherence width d_Γ we will see high interference contrast; in fact we see a speckle contrast of $C = 1$, if the roughness of the object is smaller than the coherence length of the source. This is the case for most metal surfaces, for ground glasses, or for worked plastic surfaces. The assumption is not true for 'translucent' surfaces, such as skin, paper, or wood. This will be discussed below.

The observation that coherent imaging is achieved, if we can resolve distances smaller than the coherence width d_Γ, is identical to the simple rule that fully coherent imaging occurs if the observation aperture u_o is larger than the illumination aperture u_q.

As mentioned, we will incoherently average in the image plane, over some object area, determined by the size of the diffraction spot. According to the rules of speckle-averaging, the speckle contrast C decreases with the inverse square root of the number N of incoherently averaged speckle patterns. This number N is equal to the ratio of the area A_{diff} of the diffraction spot divided by the coherence area $A_\Gamma = d_\Gamma^2$. So we obtain for the speckle contrast C:

$$C = 1 \quad \text{for } u_q < u_o \qquad [7a]$$

$$C = 1/\sqrt{(N)} = u_o/u_q, \quad \text{for } u_q >= u_o \qquad [7b]$$

We summarize the results in Figure 7.

Equation [7b] has an interesting consequence: we never get rid of speckle noise, even for large illumination apertures and small observation apertures. In many practical instruments, such as a slide projector, the illumination 'aperture stop' cannot be greater than the imaging lens. Fortunately, the observer's eye commonly has a pupil smaller than that of the projector, and/or looks from a distance at the projection screen. Laser projection devices however cause strong and disturbing speckle effects for the user and significant effort is invested to cope with this effect.

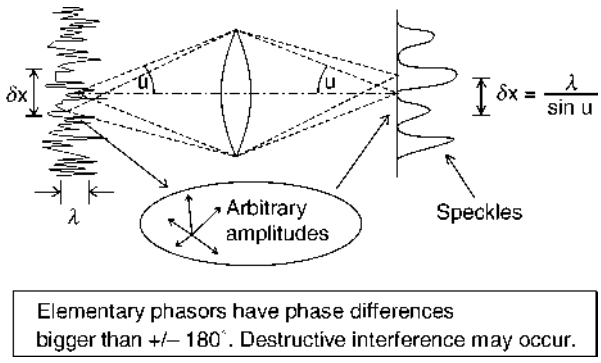

Figure 6 What is a rough surface?

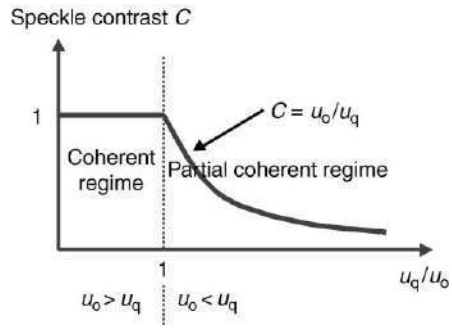

Figure 7 Coherence in a nutshell: what are the conditions for incoherent and coherent imaging?

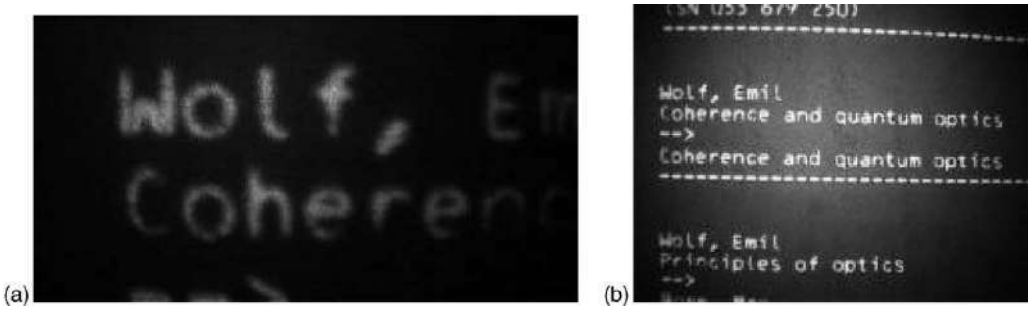

Figure 8 Microfiche projection. For high observation aperture (a) strong speckles occur, and for small observation aperture (b), the speckle contrast is low.

In Figure 8, an image is depicted from a microfiche reading projector. Figure 8a displays the close-up look (high observation aperture) and Figure 8b depicts a more distant view (small observation aperture). The amount of noise is much less in the second case, as it should be according to the rules of eqn [7].

Let us finish this section with some speculation. How would a fly see the sunny world, with a human type of eye? With high observation aperture and at short distance, the world is full of speckle noise. Fortunately, nature invented the facet-eye, for insects, as depicted in Figure 9.

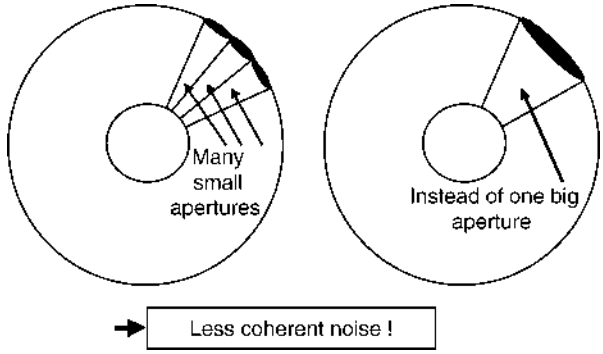

Figure 9 Facet eye and 'human' eye. The facet eye consists of many low aperture lenses, in contrast to the human eye.

Speckle Limits of Metrology

The consequences of the effects discussed above are often underestimated. We should be suspicious, if some effect in nature that disturbs us – like our coherent noise – is ubiquitous, that there might be some deep underlying principle that does not allow us to know everything about the object under observation. Indeed, it turns out that Heisenberg's uncertainty principle is strongly connected with coherent noise (Figure 10).

We can see this from the following experiment: a laser spot is projected onto a ground glass, and imaged with high magnification by a video camera, with an aperture sin u_o. The ground glass is macroscopically planar. When the ground glass is laterally shifted, we find that the observed spot is 'dancing' at the video target. Its observed position is not constant, although its projected position is. It turns out that the standard deviation of the observed position is equal to the uncertainty calculated from the aperture by Heisenberg's principle.

We can calculate the limit for the distance measuring uncertainty, from speckle theory, as well from Heisenberg's principle (within some factor of order 1):

$$\delta z \cdot \delta p_z \gg h/4\pi \qquad [8]$$

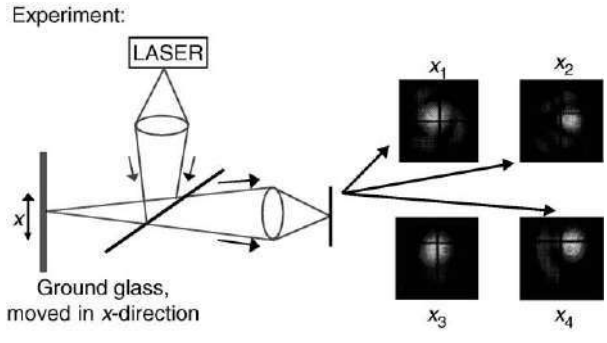

Figure 10 Coherence makes it impossible to localize objects with very high accuracy. Laser spots projected onto a ground glass cannot be imaged without some uncertainty of the lateral position. The four pictures of the spot images, taken with different lateral positions of the ground glass, display the cross bar (true position) and the apparent position of the spot images. The position uncertainty is equal to the uncertainty calculated from Heisenberg's uncertainty principle.

where δp_z is the uncertainty of the photon impulse h/λ in the z-direction (along the optical axis of the measuring system) and h is Planck's constant. For a small measurement uncertainty of the distance, we should allow a big uncertainty δp_z. This can be achieved by a large aperture of the observation system, giving the photons a wide range of possible directions to the lens. We can also allow different

wavelengths, and come to white light interferometry, see section on speckles as carriers of information below. The result is – not surprising – the same as Rayleigh scattering found for the depth of field:

$$\delta z = \lambda/\sin^2 u_o \qquad [9]$$

Coherent noise is the source of the fundamental limit of the distance measuring uncertainty δz of triangulation based sensors.

$$\delta z = C \cdot \lambda/(\sin u_o \sin \theta) \qquad [10]$$

where C is the speckle contrast, λ is the wavelength, $\sin u_o$ is the aperture of observation, and θ is the angle of triangulation (between the direction of the projection and the direction of observation). For a commercially available laser triangulation sensor, (see Figure 11), with $\sin u_o \sim 0.01, C = 1$ and $\theta = 30°$, the measuring uncertainty δz will be larger than 100 μm. We may add, that for $\sin u_o = \sin \theta$, which is valid for auto-focus sensors (such as the confocal scanning microscope), eqn [10] degenerates to the well known Rayleigh depth of field (eqn [9]).

The above results are remarkable in that we cannot know the accurate position of the projected spot or about the position of an intrinsic reflectivity feature; we also cannot know the accurate local reflectivity of a coherently illuminated object. A further consequence is that we do not know the accurate shape of an object in 3D-space. We can calculate this 'physical measuring uncertainty' from the considerations above.

These consequences hold for optically rough surfaces, and for measuring principles that exploit the local intensity of some image, such as with all triangulation type of sensors. The consequences of these considerations are depressing: triangulation with a strong laser is not better than triangulation with only one single photon. The deep reason is, that all coherent photons stem from the same quantum mechanical phase cell and are indistinguishable.

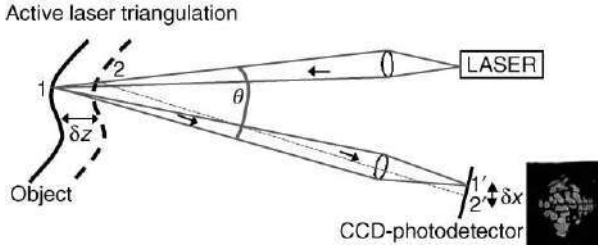

Figure 11 Principle of laser triangulation. The distance of the projected spot is calculated from the location of the speckled spot image and from the angle of triangulation. The speckle noise introduces an ultimate limit of the achievable measuring uncertainty. Reproduced with permission from *Physikalische Plätter* (May 1997): 419.

Hence, many photons do not supply more information than one single photon.

Why Are We Not Aware of Coherence

Generally, coherent noise is not well visible, even in sophisticated technical systems the visibility might be low. There are two main reasons; first, the observation aperture might be much smaller than the aperture of illumination. This is often true, for large distance observation, even with small apertures of illumination. The second reason holds for technical systems, if the observation is implemented via pixelized video targets. If the pixel size is much larger than a single speckle, which is commonly so, then, by averaging over many speckles, the noise is greatly reduced. However, we have to take into account that we pay for this by loss of lateral resolution $1/\delta x$. We can formulate another uncertainty relation:

$$\delta x > \lambda/(C \cdot \sin u_o) \qquad [11]$$

which says that if we want to reduce the speckle contrast C (noise), by lateral averaging, we can do this but lose lateral resolution $1/\delta x$.

Can We Overcome Coherence Limits?

Since the daily interaction of light with matter is coherent scattering, we can overcome the limit of eqn [10] only by looking for measuring principles that are not based on the exploitation of local reflectivity, i.e., on 'conventional imaging'. Concerning optical 3D-measurements, the principle of triangulation uses the position of some image detail, for the calculation of the shape of objects, and has to cope with coherent noise.

Are there different mechanisms of photon–matter interaction, with noncoherent scattering? Fluorescence and thermal excitation are incoherent mechanisms. It can be shown (see Figure 12) that triangulation utilizing fluorescent light, displays much less noise than given by eqn [10]. This is exploited in fluorescent confocal microscopy.

Thermally excited matter emits perfectly incoherent radiation as well. We use this incoherence to measure the material wear in laser processing (see Figure 13), with, again, much better accuracy than given by eqns [9,10].

The sensor is based on triangulation. Nevertheless, by its virtually zero coherent noise, it allows a measuring uncertainty which is limited only by camera noise and other technical imperfections. The uncertainty of the depth measurement through the narrow aperture of the laser nozzle is only a few

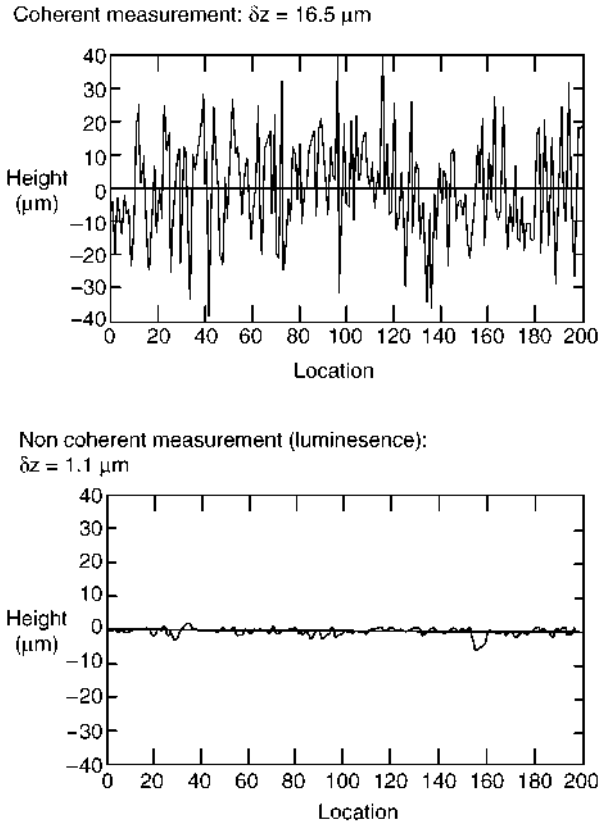

Figure 12 Reduction of speckle noise by triangulation with fluorescent light. A flat metal surface is measured by laser triangulation. Top: surface measured by laser triangulation with full speckle contrast. Bottom: the surface is covered with a very thin fluorescent layer, and illuminated by the same laser, however, the triangulation is done after suppressing the scattered laser light, and utilizing the (incoherent!) fluorescent light. This experiment proves that it is the coherent noise that causes the measuring uncertainty.

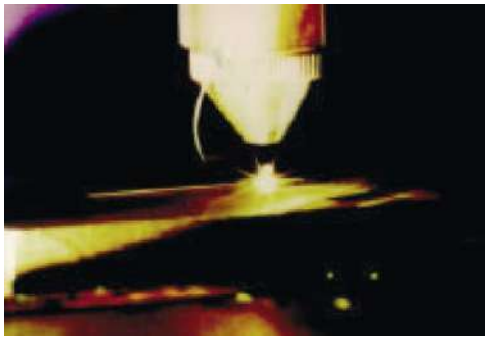

Figure 13 A 100 W laser generates a plasma that emits perfectly incoherent radiation at the object surface. The emitted light can be used to measure on-line the distance of the object surface. A triangulation sensor, exploiting the incoherent plasma emission, controls surface ablation with an accuracy of only a few microns. Reproduced with permission from F & M, *Feinwerktechnik Mikrotechnik Masstechnik* (1995) Issue 9.

microns, beating the limitation of eqn [10] by a factor of about 5, even in the presence of a turbulent plasma spot with a temperature of 3000 K.

Broadband Illumination

Speckle noise can hardly be avoided by broadband illumination. The speckle contrast C is related to the surface roughness σ_z and the coherence length l_c by eqn [12]:

$$C^2 \sim l_c/2\sigma_z \text{ for } \sigma_z \gg l_c \qquad [12]$$

So, only for very rough surfaces, and white light illumination, C can be reduced. However, the majority of technical objects is smooth, with a roughness of a few micrometers, hence speckles can be observed even with white light illumination ($l_c \sim 3$ μm). This situation is different for 'translucent' objects such as skin, paper, or some plastic material.

Speckles as a Carrier of Information

So far we discussed the problems of speckles when measuring rough surfaces. We can take advantage of speckles, if we do not stick to triangulation as the basic mechanism of distance measurement. We may utilize the fact that although the phase within each speckle has a random value, this phase is constant over some area. This observation can be used to build an interferometer that works even at rough surfaces. So, with proper choice of the illumination aperture and observation aperture, and with a surface roughness less than the coherence length, we can measure distances by the localization of the temporal coherence function. This function can easily be measured by moving the object under test along the optical axis and measuring the interference intensity (correlogram) at each pixel of the camera. The signal generating mechanism of this 'white light interferometer' distance measurement is not triangulation but 'time of flight'. Here we do not suffer from the limitations of eqn [10]. The limits of white light rough surface interferometry are discussed in the Further Reading section at the end of this article.

The 'coherence radar' and the correlograms of different speckles are depicted in Figures 14 and 15.

The physical mechanism of the signal formation in white light interferometry at rough surfaces is different from that of white light interferometry at smooth surfaces. That is why the method is sometimes referred as 'coherence radar'. We will briefly summarize some advantageous, and – probably unexpected – features of the coherence radar.

We call the real shape of the object, or the surface profile, $z(x, y)$, where z is the local distance at the position (x, y). We give the measured data an index 'm^2':

(i) The physical measuring uncertainty of the measured data is independent from the observation aperture. This is quite a remarkable property, as it enables us to take accurate measurements at the bottom of deep boreholes.

(ii) The standard deviation σ_z of the object can be calculated from the measured data z_m: $\sigma_z = <|z_m|>$. According to German standards, σ_z corresponds to the roughness measure R_q. This measure R_q was calculated from measurements at different roughness gauges, with different observation apertures. Two measurements are depicted in Figure 16.

The experiments shown in Figure 16 display the correct roughness measure, even if the higher frequencies of the object spectrum $Z(\nu, \mu)$ (Figure 17), are not optically resolved. The solution of this paradox might be explained as follows.

From the coherently illuminated object each point scatters a spherical wave. The waves scattered from different object points have, in general, a different phase. At the image plane, we see the laterally

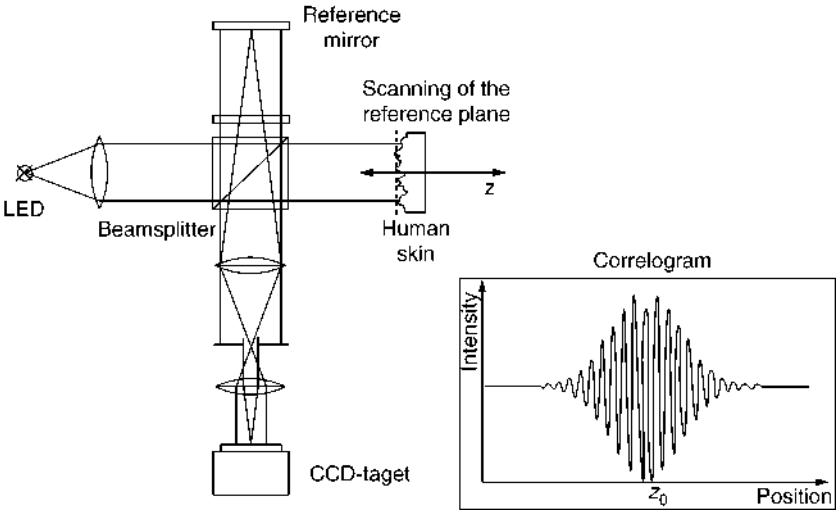

Figure 14 'Coherence radar', white light interferometry at rough surfaces. The surface under test is illuminated with high spatial coherence and low temporal coherence. We acquire the temporal coherence function (correlogram) within each speckle, by scanning the object under test in depth.

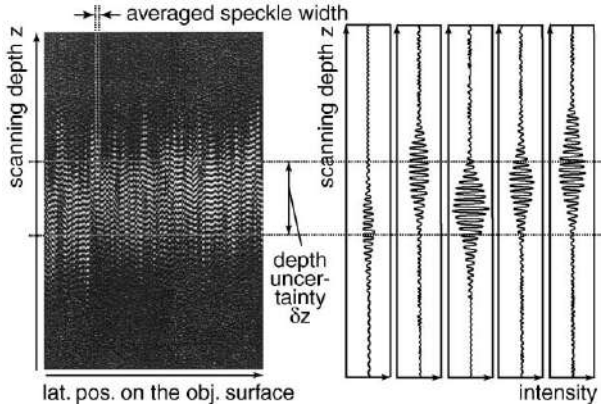

Figure 15 White light speckles in the $x-z$ plane. The left side displays the acquired correlograms, for a couple of speckles, the right side shows the graphs of some of the correlograms. We see that the correlograms are not located at a constant distance, but display some 'distance uncertainty'. This uncertainty does not originate from the instrument, but from the roughness of the object.

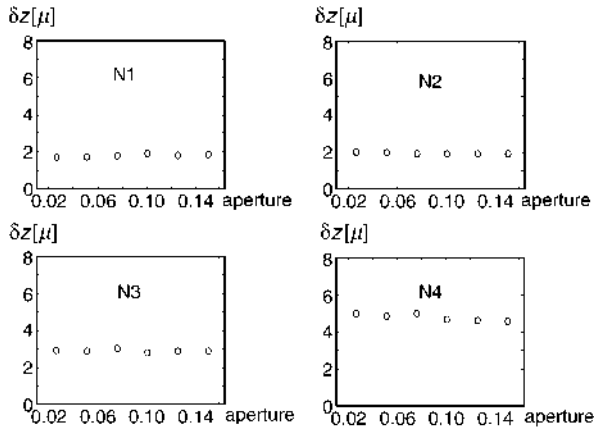

Figure 16 Optical measurement of the roughness beyond the Abbe resolution limit. The roughness is measured for 4 roughness standards N1–N4, with different observation apertures. The roughness is measured correctly, although the microtopology is not resolved by the observing optics.

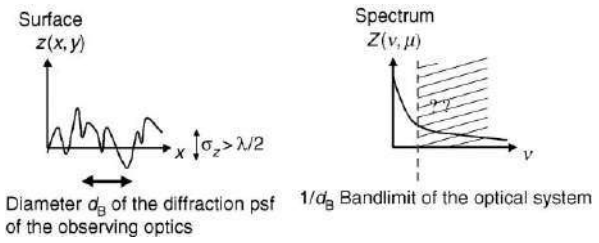

Figure 17 Signal generation beyond the Abbe limit. The microtopology of the object $z(x, y)$ is not resolved, due to the large diffraction image with diameter d_B. Yet we get information from beyond the bandlimit $1/d_B$ of the observing optics.

averaged complex amplitude $<u>$, according to eqn [6]. The averaging is due to the diffraction limited resolution.

Equation [6] reveals that the (spatially) averaged field amplitude $<u>$ is a nonmonotonic, nonlinear function of the surface profile $z(x, y)$, if the surface is rough (i.e., $z \gg \lambda$). We do not average over the profile z but over the complex amplitude. As a consequence of this nonlinearity, the limited observation aperture does not only collect the spatial frequencies ν, μ of the spatial Fourier spectrum $Z(\nu, \mu)$ within the bandlimit $1/d_B$, but also acquires information beyond this bandlimit. The reason is the 'down conversion' of higher spatial frequencies by nonlinear mixing. Simulations and experiments confirm the consequences.

Thus we can evaluate the surface roughness, even without laterally resolving the microtopology. However, there is a principal uncertainty on the measured data. We do not see the true surface but a surface with a 'noise' equivalent to the roughness of the real surface.

Nonlinear nonmonotonic nonlinearities in a sequence of operations may cause 'chaotic' behavior, in other words, small parameter changes such as vibrations, humidity on the surface, etc. may cause large differences of the outcome. In fact, we observe a significant variation of the measured data $z_m(x, y)$ within a sequence of repeated measurements. Our conjecture is that the complex signal formation may be involved in this irregular behavior. The hypothesis may be supported by the observation that much better repeatability can be achieved at specular surfaces, where eqn [6] degenerates to a linear averaging over the surface profile:

$$<u> \sim 1 + ik <z(x, y)>, \quad \text{for } z \ll \lambda \qquad [13]$$

Summary

Spatial coherence is ubiquitous and unavoidable. Spatial coherence disturbs most optical measurements, where optically rough surfaces exist. Then, coherence gives the ultimate limit of the achievable measuring uncertainty. There are simple rules to estimate the measuring uncertainty, by calculating the width of the coherence function and the resolution of the observation system. Spatial coherence and Heisenberg's uncertainty principle lead to the same results of measuring uncertainty.

On the other hand – the formation of signals from coherent light that is scattered at rough surfaces is quite complex and a strongly nonlinear process, which sometimes might encode information which is not otherwise available. One example appears to be the measurement of the roughness with white light interferometry – which is possible, even if the microtopology of the surface is not optically resolved.

See also

Coherence: Overview; Coherence and Imaging.

Further Reading

Born M and Wolf E (1999) *Principles of Optics*, 7th edn. Cambridge: Cambridge University Press.

Brooker G (2003) *Modern Classical Optics*. Oxford: Oxford University Press.

Dainty JC (ed.) (1984) *Laser Speckle and Related Phenomena*. Berlin, Heidelberg, New York: Springer Verlag.

Dorsch R, Herrmann J and Häusler G (1994) Laser triangulation: fundamental uncertainty in distance measurement. *Applied Optics* 33: 1306–1314.

Dresel T, Venzke H and Häusler G (1992) 3D-sensing of rough surfaces by coherence radar. *Applied Optics* 31: 919–925.

Ettl P (2002) *Über die Signalenstehung bei Weisslichtinterferometrie*. PhD Dissertation, University of Erlangen-Nürnberg.

Häusler G, Ettl P, Schenk M, Bohn G and Laszlo I (1999) Limits of optical range sensors and how exploit them. In: Asakura T (ed.) *Trends in Optics and Phototonics*. Ico IV, Springer Series in Optical Sciences, vol. 74, pp. 328–342. Berlin, Heidelberg, New York: Springer Verlag.

Häusler G and Herrmann J (1992) Physical limits of 3D-sensing. Proceedings of the SPIE Conf. 1822. In: Svetkov D (ed.) *Optics, Illumination, and Image Sensing for Machine Vision VII*, pp. 150–158. Boston, MA: SPIE.

Klein M and Furtak TE (1986) *Optics*. New York: John Wiley & Sons.

Klinger P, Spellenberg B, Herrmann JM and Häusler G (2001) *In Process 3D-Sensing for Laser Material Processing*. Proceedings of Third Int. Conf. on 3-D Digital Imaging and Modeling, Quebec City, Canada. Los Alamitos: IEEE Computer Society, 38–41.

Pedrotti F and Pedrotti L (1993) *Introduction to Optics*, 2nd edn. New Jersey: Prentice Hall.
Pérez J-Ph (1996) *Optik, Spektrum Akademischer*. Heidelberg: Verlag.
Sharma A, Kandpal HC and Ghatak AK (2004) *Optical Coherence*. See this Encyclopedia.
van de Gracht (2004) *Coherence and Imaging*. See this Encyclopedia.
Young M (1993) *Optics and Lasers, Including Fibers and Optical Wave Guides*, 4th revised edn. Berlin, Heidelberg: Springer Verlag.

COHERENT CONTROL

Contents

Theory
Experimental
Applications in Semiconductors

Theory

H Rabitz, Princeton University, Princeton, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Since the inception of quantum mechanics almost a century ago, a prime activity has been the observation of quantum phenomena in virtually all areas of chemistry and physics. However, the natural evolution of science leads to the desire to go beyond passive observation to active manipulation of quantum mechanical processes. Achieving control over quantum phenomena could be viewed as engineering at the atomic scale guided by the principles of quantum mechanics, for the alteration of system properties or dynamic behavior. From this perspective, the construction of quantum mechanically operating solid-state devices through selective material growth would fall into this category. The focus of this article is principally on the manipulation of quantum phenomena through tailored laser pulses. The suggestion of using coherent radiation for the active alteration of microworld processes may be traced to the early 1960s, almost immediately after the discovery of lasers. Since then, the subject has grown enormously to encompass the manipulation of (1) chemical reactions, (2) quantum electron transport in semiconductors, (3) excitons in solids, (4) quantum information systems, (5) atom lasers, and (6) high harmonic generation, amongst other topics. Perhaps the most significant use of these techniques may be their provision of refined tools to ultimately better understand the basic physical interactions operative at the atomic scale.

Regardless of the particular application of laser control over quantum phenomena, there is one basic operating principle involved: active manipulation of constructive and destructive quantum wave interferences. This process is depicted in Figure 1, showing the evolution of a quantum system from the initial state $|\psi_i\rangle$ to the desired final state $|\psi_f\rangle$ along three of possibly many interfering pathways. In general, there may be many possible final accessible states $|\psi_f\rangle$, and often the goal is to achieve a high amplitude in one of these states and low amplitude in all the others. The target state might actually be a superposition of states, and an analogous picture to

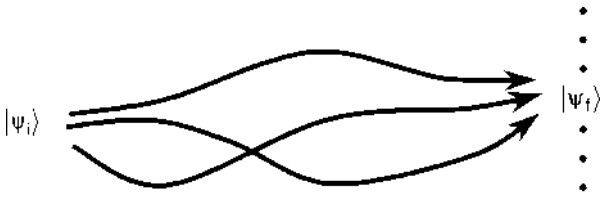

Figure 1 The evolution of a quantum system under laser control from the initial state $|\psi_i\rangle$ to the final state $|\psi_f\rangle$. The latter state is chosen to have desirable physical properties, and three of possibly many pathways between the states are depicted. Successful control of the process $|\psi_i\rangle \rightarrow |\psi_f\rangle$ by a tailored laser pulse generally requires creating constructive quantum wave interferences in the state $|\psi_f\rangle$ from many pathways, and destructive interferences in all other accessible final states $|\psi_{f'}\rangle \neq |\psi_f\rangle$.

that in Figure 1 would also apply to steering about the density matrix. The process depicted in Figure 1 may be thought of as a microscale analog of the classic double-slit experiment for light waves. As the goal is often high-finesse focusing into a particular target quantum state, success may call for the manipulation of many quantum pathways (i.e., the notion of a 'many-slit' experiment). Most applications are concerned with achieving a desirable outcome for the expectation value $\langle\psi_f|O|\psi_f\rangle$ associated with some observable Hermitian operator O.

The practical realization of the task above becomes a control problem when the system is expressed through its Hamiltonian $H = H_0 + V_c$, where H_0 is the free Hamiltonian describing the dynamics without explicit control; it is assumed that the free evolutionary dynamics under H_0 will not satisfactorily achieve the physical objective. Thus, a laboratory-accessible control term V_c is introduced in the Hamiltonian to achieve the desired manipulation. Even just considering radiative interactions, the form of V_c could be quite diverse depending on the nature of the system (e.g., nuclear spins, electrons, atoms, molecules, etc.) and the intensity of the radiation field. Many problems may be treated through an electric dipole interaction $V_c = -\boldsymbol{\mu}\cdot\boldsymbol{\varepsilon}(t)$ where $\boldsymbol{\mu}$ is a system dipole moment and $\boldsymbol{\varepsilon}(t)$ is the laser electric field. Where appropriate, this article will consider the interaction in this form, but other suitable radiative control interactions may be equally well treated. Thus, the laser field $\boldsymbol{\varepsilon}(t)$ as a function of time (or frequency) is at our disposal for attempted manipulation of quantum systems.

Before considering any practical issues associated with the identification of shaped laser control fields, a fundamental question concerns whether it is, in principle, possible to steer about any particular quantum system from an arbitrary initial state $|\psi_i\rangle$ to an arbitrary final state $|\psi_f\rangle$. Questions of this sort are addressed by a controllability analysis of Schrödinger's equation

$$i\hbar\frac{\partial|\psi(t)\rangle}{\partial t} = [H_0 - \boldsymbol{\mu}\cdot\boldsymbol{\varepsilon}(t)]|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_i\rangle \qquad [1]$$

Controllability concerns whether, in principle, some field $\boldsymbol{\varepsilon}(t)$ exists, such that the quantum system described by $H_0 = H_0 - \boldsymbol{\mu}\cdot\boldsymbol{\varepsilon}(t)$ permits arbitrary degrees of control. For finite-dimensional quantum systems (i.e., those described by evolution amongst a discrete set of quantum states), the formal tools for such a controllability analysis exist both for evaluating the controllability of the wave function, as well as the more general time evolution operator $U(t)$, which satisfies

$$i\hbar\frac{\partial U}{\partial t} = [H_0 - \boldsymbol{\mu}\cdot\boldsymbol{\varepsilon}(t)]U, \quad U(0) = 1 \qquad [2]$$

Analyses of this type can be quite insightful, but they require detailed knowledge about H_0 and $\boldsymbol{\mu}$. Cases involving quantum information science applications are perhaps the most demanding with regard to achieving total control. Most other physical applications would likely accept much more modest levels of control and still be categorized as excellent achievements.

Theoretical tools and concepts have a number of roles in considering the control of quantum systems, including (1) an exploration of physical/chemical phenomena under active control, (2) the design of viable control fields, (3) the development of algorithms to actively guide laboratory control experiments towards achieving their dynamical objectives, and (4) the introduction of special algorithms to reveal the physical mechanisms operative in the control of quantum phenomena. Activities (1)–(3) have thus far been the primary focus of theoretical studies in this area, and it is anticipated that item (4) will grow in importance in the future.

A few comments on the history of laser control over quantum systems are relevant, as they speak to the special nature of the currently employed successful closed-loop quantum control experiments. Starting in the 1960s and spanning roughly 20 years, it was thought that the design of lasers to manipulate molecular motion could be achieved by the application of simple physical logic and intuition. In particular, the thinking at the time focused on using cw laser fields resonant with one or more local modes of the molecule, as state or energy localization was believed to be the key to successful control. Since quantum dynamics phenomena typically occur on ultrafast time-scales, possibly involving spatially dispersed wave packets, expecting to achieve high quality control with a light source operating at one or two resonant frequencies is generally wishful thinking. Quantum dynamics phenomena occur in a multifrequency domain, and controls with a few frequencies will not suffice. Over approximately the last decade, it became clear that successful control often calls for manipulating multiple interfering quantum pathways (cf., Figure 1). In turn, this recognition led to the need for broad-bandwidth laser sources (i.e., tailored laser pulses). Fortunately, the necessary laser pulse-shaping technologies have become available, and these sources continue to expand into new frequency ranges with enhanced bandwidth capabilities. Many chemical and physical

applications of control over quantum phenomena involve performing a significant amount of work (i.e., quantum mechanical action), and sufficient laser field intensity is required. A commonly quoted adage is 'no field, no yield', and at the other extreme, there was much speculation that operating nonlinearly at high field intensities would lead to a loss of control because the quantum system would effectively act as an amplifier of even weak laser field noise. Fortunately, the latter outcome has not occurred, as is evident from a number of successful high field laser control experiments.

Quantum control may be expressed as an inverse problem with a prescribed chemical or physical objective, and the task being the discovery of a suitable control field $\boldsymbol{\varepsilon}(t)$ to meet the objective. Normally, Schrödinger's equation [1] is viewed as linear, and this perspective is valid if the Hamiltonian is known a priori, leaving the equation to be solved for the wave function. However, in the case of quantum control, neither the wave function nor the control field is known a priori, and the two enter bilinearly on the right-hand side of eqn [1]. Thus, quantum control is mathematically a nonlinear inverse problem. As such, one can anticipate possibly diverse behavior in the process of seeking successful controls, as well as in the ensuing quantum dynamical control behavior. The control must take into account the evolving system dynamics, thereby resulting in the control field being a function(al) of the current state of the system $\boldsymbol{\varepsilon}(\psi)$ in the case of quantum control design, or dependent on the system observations $\boldsymbol{\varepsilon}(\langle O\rangle)$ in the case of laboratory closed-loop field discovery efforts where $\langle O\rangle = \langle\psi|O|\psi\rangle$. Furthermore, the control field at the current time t will depend in some implicit fashion on the future state of the evolving system at the final target time T. It is evident that control field design and laboratory discovery present highly nontrivial tasks.

Although the emphasis in this review is on the theoretical aspects of quantum control theory, the subject exists for its laboratory realizations, and those realizations, in turn, intimately depend on the capabilities of theoretical and algorithmic techniques for their successful implementations. Accordingly, theoretical laser control field design techniques will be discussed, along with algorithmic aspects of current laboratory practice. This material respectively will focus on optimal control theory (OCT) and optimal control experiments (OCEs), as seeking optimality provides the best means to achieve any posed objective. Finally, the last part of the article will present some general conclusions on the state of the quantum control field.

Quantum Optimal Control Theory for Designing Laser Fields

When considering control in any domain of application, a reasonable approach is to computationally design the control for subsequent implementation in the laboratory. Quantum mechanical laser field design has taken on a number of realizations. At one limit is the application of simple intuition for design purposes, and in some special cases (e.g., the weak field perturbation theory regime), this approach may be applicable. Physical insights will always play a central role in laser field design, but to be especially useful, they need to be channeled into the proper mathematical framework. Many of the interesting applications operate in the strong-field nonperturbative regime. In this domain, serious questions arise regarding whether sufficient information is available about the Hamiltonian to execute reliable designs. Regardless of whether the Hamiltonian is known accurately or is an acknowledged model, the theoretical study of quantum control can provide physical insight into the phenomena involved, as well as possibly yielding trial laser pulses for further refinement in the laboratory.

Achieving control over quantum mechanical phenomena often involves a balance of competing dynamical processes. For example, in the case of aiming to create a particular excitation in a molecule or material, there will always be the concomitant need to minimize other unwanted excitations. Often, there are also limitations on the form, intensity, or other characteristics of the laser controls that must be adhered to. Another goal is for the control outcome to be as robust as possible to the laser field fluctuations and Hamiltonian uncertainties. Overriding all of these issues is merely the desire to achieve the best possible physical result for the posed quantum mechanical objective. In summary, all of these desired extrema conditions translate over to posing the design of control laser fields as an optimization problem. Thus, optimal control theory forms a basic foundation for quantum control field design.

Control field design starts by specification of the Hamiltonian components H_0 and $\boldsymbol{\mu}$ in eqn [1], with the goal of finding the best control electric field $\boldsymbol{\varepsilon}(t)$ to balance the competing objectives. Schrödinger's equation must be solved as part of this process, but this effort is not merely a forward propagation task, as the control field is not known a priori. As an optimal design is the goal, a cost functional $J = J(\text{objectives, penalties, } \boldsymbol{\varepsilon}(t))$ is prescribed, which contains the information on the physical objectives, competing penalties, and any costs or constraints associated with the structure of the laser field. The functional J could

contain many terms if the competing physical goals are highly complex. As a specific simple illustration, consider the common goal of steering the system to achieve an expectation value $\langle\psi(T)|O|\psi(T)\rangle$ as close as possible to the target value O_{target} associated with the observable operator O at time T. An additional cost is imposed to minimize the laser fluence. These criteria may be embodied in a cost functional of the form

$$J = \left(\langle\psi(T)|O|\psi(T)\rangle - O_{\text{target}}\right)^2 + \omega\int_0^T \varepsilon^2(t)\,\mathrm{d}t - 2\Im\int_0^T \mathrm{d}t\left\langle\lambda(t)|\mathrm{i}\hbar\frac{\partial}{\partial t} - H_0 + \boldsymbol{\mu}\cdot\boldsymbol{\varepsilon}(t)|\psi(t)\right\rangle \quad [3]$$

Here, the parameter $\omega \geq 0$ is a weight to balance the significance of the fluence term relative to achieving the target expectation value, and $|\lambda(t)\rangle$ serves as a Lagrange multiplier, assuring that Schrödinger's equation is satisfied during the variational minimization of J with respect to the control field. Carrying out the latter minimization will lead to eqn [1], along with the additional relations

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\lambda(t)\rangle = [H_0 - \boldsymbol{\mu}\cdot\boldsymbol{\varepsilon}(t)]|\lambda(t)\rangle, \quad |\lambda(T)\rangle = 2\sigma O|\lambda(T)\rangle \quad [4]$$

$$\boldsymbol{\varepsilon}(t) = \frac{1}{\omega}\Im\langle\lambda(t)|\boldsymbol{\mu}|\psi(t)\rangle \quad [5]$$

$$\sigma = \langle\psi(T)|O|\psi(T)\rangle - O_{\text{target}} \quad [6]$$

Equations [1] and [4]–[6] embody the OCT design equations that must be solved to yield the optimal field $\boldsymbol{\varepsilon}(t)$ given in eqn [5]. These design equations have an unusual mathematical structure within quantum mechanics. First, insertion of the field expression in eqn [5] into eqns [1] and [4] produces two cubically nonlinear coupled Schrödinger-like equations. These equations are in the same family as the standard nonlinear Schrödinger equation, and as such, one may expect that unusual dynamical behavior could arise. Although eqn [4] for $|\lambda(t)\rangle$ is identical in form to the Schrödinger equation [1], only $|\psi(t)\rangle$ is the true wavefunction, with $|\lambda(t)\rangle$ serving to guide the controlled quantum evolution towards the physical objective. Importantly, eqn [1] is an initial value problem, while eqn [4] is a final value problem. Thus, the two equations together form a two-point boundary value problem in time, which is an inherent feature of temporal engineering optimal control as well. The boundary value nature of these equations typically leads to the existence of multiple solutions, where σ in eqn [6] plays the role of a discrete eigenvalue, specifying the quality of the particular achieved control field design. The fact that there are generally multiple solutions to the laser design equations can be attractive from a physical perspective, as the designs may be sorted through to identify those being most attractive for laboratory implementation.

The overall mathematical structure of eqns [1] and [4]–[6] can be melded together into a single design equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = [H_0 - \boldsymbol{\mu}\cdot\boldsymbol{\varepsilon}(\psi,\sigma,t)]|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_{\mathrm{i}}\rangle \quad [7]$$

The structure of this equation embodies the comments in the introduction that control field design is inherently a nonlinear process. The main source of complexity arising in eqn [7] is through the control field $\boldsymbol{\varepsilon}(\psi,\sigma,t)$ depending not only on the wave function at the present time t, but also on the future value of the wavefunction at the target time T contained in σ. This structure again reflects the two-point boundary value nature of the control equations.

Given the typical multiplicity of solutions to eqns [1] and [4]–[6], or equivalently, eqn [7], it is attractive to consider approximations to these equations (except perhaps for the inviolate Schrödinger equation [1], and much work continues to be done along these lines. One case involves what is referred to as tracking control, whereby a path is specified for $\langle\psi(t)|O|\psi(t)\rangle$ evolving from $t = 0$ out to $t = T$. Tracking control eliminates σ to produce a field of the form $\boldsymbol{\varepsilon}(\psi,t)$, thereby permitting the design equations to be explicitly integrated as a forward-marching problem toward the target; the tracking equations are still nonlinear with respect to the evolving state. Many variations on these concepts can be envisioned, and other approximations may also be introduced to deal with special circumstances. One technique appropriate for at least few-level systems is stimulated Raman adiabatic passage (STIRAP), which seeks robust adiabatic passage from an initial state to a particular final state, typically with nearly total destructive interference occurring in the intermediate states. This design technique can have special attractive robustness characteristics with respect to field errors. STIRAP, tracking, and various perturbation theory-based control design techniques can be expressed as special cases of OCT, with suitable cost functionals and constraints. Further approximation methods will surely be developed, with the rigorous OCT concepts forming the general foundation for control field design.

As the OCT equations are inherently nonlinear, their solution typically requires numerical iteration,

and a variety of procedures may be utilized for this purpose. Almost all of the methods employ local search techniques (e.g., gradient methods), and some also show monotonic convergence with respect to each new iteration step. Local techniques typically evolve to the nearest solution in the space of possible control fields. Such optimizations can be numerically efficient, although the quality of the attained result can depend on the initial trial for the control field and the details of the algorithm involved. In contrast, global search techniques, such as simulated annealing or genetic algorithms, can search more broadly for the best solution in the control field space. These more expansive searches are attractive, but at the price of typically requiring more intensive computations. Effort has also gone into seeking global or semiglobal input $\rightarrow$ output maps relating the electric field $\boldsymbol{\varepsilon}(t)$ structure to the observable $O[\boldsymbol{\varepsilon}(t)]$. Such maps may be learned, hopefully, from a modest number of computations with a selected set of fields, and then utilized as high-speed interpolators over the control field space to permit the efficient use of global search algorithms to attain the best control solution possible. At the foundation of all OCT design procedures is the need to solve the Schrödinger equation. Thus, computational technology to improve this basic task is of fundamental importance for designing laser fields.

Many computations have been carried out with OCT to attain control field designs for manipulating a host of phenomena, including rotational, vibrational, electronic, and reactive dynamics of molecules, as well as electron motion in semiconductors. Every system has its own rich details, which in turn, are compounded by the fact that multiple control designs will typically exist in many applications producing comparable physical outcomes. Collectively, these control design studies confirm the manipulation of constructive and destructive quantum wave interferences as the general mechanism for achieving successful control over quantum phenomena. This conclusion may be expressed as the following principle:

> **Control field–system cooperativity principle:** Successful quantum control requires that the field must have the proper structure to take full advantage of all of the dynamical opportunities offered by the system, to best satisfy the physical objective.

This simple statement of system–field cooperativity embodies the richness, as well as the complexity, of seeking control over quantum phenomena, and also speaks to why simple intuition alone has not proved to be a generally viable design technique. Quantum systems can often exhibit highly complex dynamical behavior, including broad dispersion of the wave packet over spatial domains or multitudes of quantum states. Handling such complexity can require fields with subtle structure to interact with the quantum system in a global fashion to manage all of the motions involved. Thus, we may expect that successful control pulses will often have broad bandwidth, including amplitude and phase modulation. Until relatively recently, laser sources with these characteristics were not available, but the technology is now in hand and rapidly evolving (see the discussion later).

The cooperativity principle above is of fundamental importance in the control of all quantum phenomena, and a simple illustration of this principle is shown in Figure 2 for the control of wave packet motion on an excited state of the NO molecule. The target for the control is a narrow wave packet located over the well of the excited B state. The optimal control field consists of two coordinated pulses, at early and late times, with both features having internal structure. The figure indicates that the ensuing excited state wave packet has two components, one of which actually passes through the target region during the initial evolution, only to return again and meet the second component at just the right place and time, to achieve the target objective as best as possible. Similar cooperativity interpretations can be found in virtually all implementations with OCT.

The main reason for performing quantum field design is to ultimately implement the designs in the laboratory. The design procedures must face laboratory realities, which include the fact that most Hamiltonians are not known to high accuracy (especially for polyatomic molecules and complex solid-state structures), and secondly, a variety of laboratory field imperfections may unwittingly be present. Notwithstanding these comments, OCT has been fundamental to the development of quantum control, including laying out the logic for how to perform the analogous optimal control experiments. Design implementations and further OCT development will continue to play a basic role in the quantum control field. At present, perhaps the most important contribution of OCT has been to (1) highlight the basic control cooperativity principle above, and (2) provide the basis for developing algorithms to successfully guide optimal control experiments, as discussed below.

Algorithms for Implementing Optimal Control Experiments

The ultimate purpose of considering control theory within quantum mechanics is to take the matter into

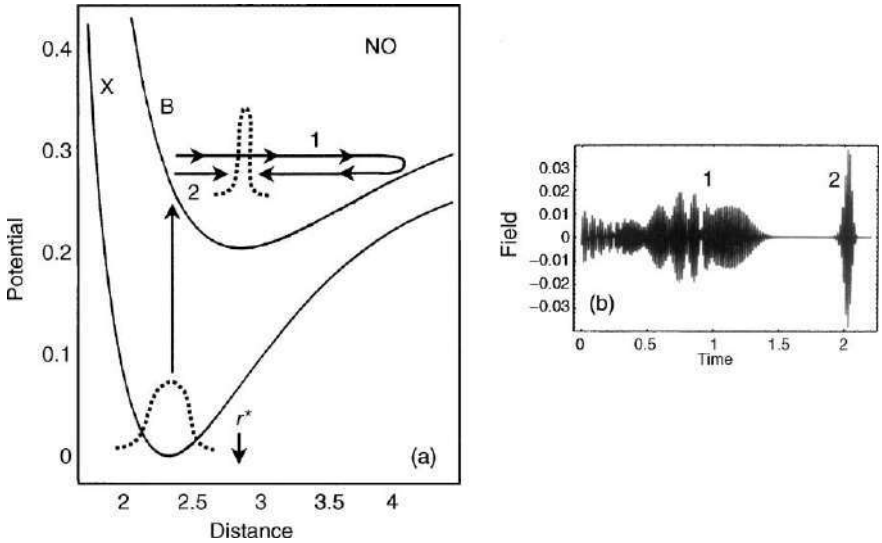

Figure 2 Control of wave packet evolution on the electronically excited B state of NO, with the goal of creating a narrow final packet over the excited state well. (a) indicates that the ground-state packet is brought up in two pieces using the coordinated dual pulse in (b). The piece of the packet from the first pulse actually passes through the target region, then bounces off the right side of the potential, to finally meet the second piece of the packet at just the right time over the target location r^* for successful control. This behavior is an illustration of the control field–system cooperativity principle stated in the text.

the laboratory for implementation and exploitation of its capabilities. Ideally, theoretical control field designs may be attained using the techniques discussed above, followed by the achievement of successful control upon execution of the designs in the laboratory. This appealing approach is burdened with three difficulties: (1) Hamiltonians are often imprecisely known, (2) accurately solving the design equations can be a significant task, and (3) realization of any given design will likely be imperfect due to laboratory noise or other unaccounted-for systematic errors. Perhaps the most serious of these difficulties is point (1), especially considering that the best quality control will be achieved by maximally drawing on subtle constructive and destructive quantum wave interference effects. Exploiting such subtleties will generally require high-quality control designs that, in turn, depend on having reliable Hamiltonians. Although various designs have been carried out seeking robustness with respect to Hamiltonian uncertainties, the issue in point (1) should remain of significance in the foreseeable future, especially for the most complex (and often, the most interesting!) chemical/physical applications. Mitigating this serious problem is the ability to create shaped laser pulse controls and apply them to a quantum system, followed by a probe of their effects at an unprecedented rate of thousands or more independent trials per minute. This unique capability led to the suggestion of partially, if not totally, sidestepping the design process by performing closed-loop experiments to let the quantum system teach the laser how to achieve its control in the laboratory. Figure 3 schematically shows this closed loop process, drawing on the following logic:

1. **The molecular view.** Although there may be theoretical uncertainty about the system Hamiltonian, the actual chemical/physical system under study 'knows' its own Hamiltonian precisely! This knowledge would also include any unusual exigencies, perhaps associated with structural or other defects in the particular sample. Furthermore, upon exposure to a control field, the system 'solves' its own Schrödinger equation impeccably accurately and as fast as possible in real time. Considering these points, the aim is to replace the offline arduous digital computer control field design effort with the actual quantum system under study acting as a precise analog computer, solving the true equations of motion.
2. **Control laser technology.** Pulse shaping under full computer control may be carried out using even hundreds of discrete elements in the frequency domain controlling the phase and amplitude structure of the pulse. This technology is readily available and expanding in terms of pulse center frequency flexibility and bandwidth capabilities.

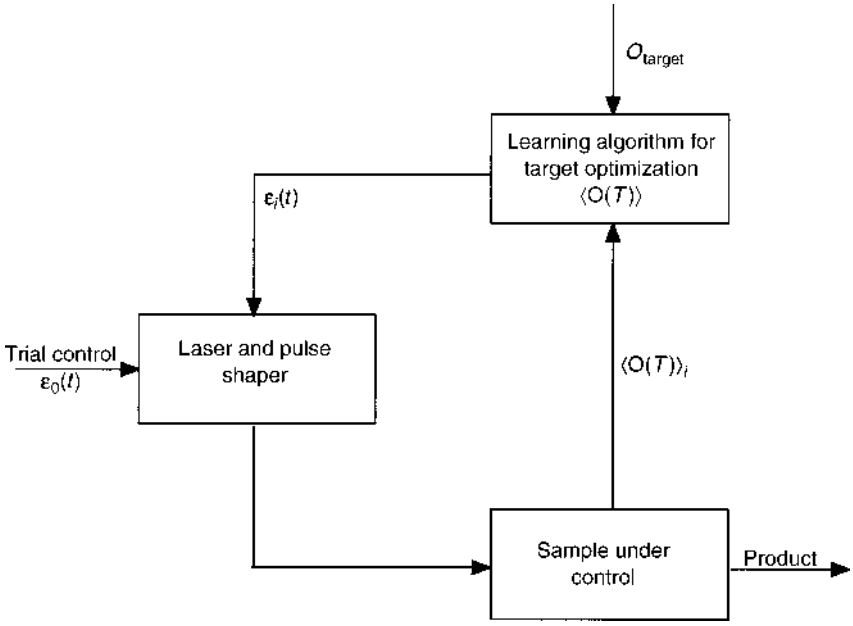

Figure 3 A schematic of the closed-loop concept for allowing the quantum system to teach the laser how to achieve its control. The actual quantum system under control is in a loop with a laser and pulse shaper all slaved together with a pattern recognition algorithm to guide the excursions around the loop. The success of this concept relies on the ability to perform very large numbers of control experiments in a short period of time. In principle, no knowledge of the system Hamiltonian is required to steer the system to the desired final objective, although a good trial design $\varepsilon_0(t)$ may accelerate the process. The ith cycle around the loop attempts to find a better control field $\varepsilon_i(t)$ such that the system response $\langle O(T)\rangle_i$ for the observable operator O comes closer to the desired value O_{target}.

3. **Quantum mechanical objectives.** Many chemical/physical objectives may be simply expressed as the desire to steer a quantum mechanical flux out one clearly defined channel versus another.
4. **Detection of the control action.** Detection of the control outcome can often be carried out by a second laser pulse or any other suitable high duty cycle detection means, such as laser-induced mass spectrometry. The typical circumstance in point (3) implies that little, if any, time-consuming offline data analysis will be necessary beyond that of simple signal averaging.
5. **Fast learning algorithms.** All of the four points above may be slaved together using pattern recognition learning algorithms to identify those control fields which are producing better results, and bias the next sequence of experiments in their favor for further improvement. Although many algorithms may be employed for this purpose, the global search capabilities of genetic algorithms are quite attractive, as they may take full advantage of the high throughput nature of the experiments.

In linking together all of these components into a closed-loop OCE, it is important that no single operational step significantly lags behind any other, for efficiency reasons. Fortunately, this economy is coincident with all the tasks and technologies involved. For achieving control alone, there is no requirement that the actual laser pulse structure be identified on each cycle of the loop. Rather, the laser control 'knob' settings are adequate information for the learning algorithm, as the only criterion is to suitably adjust the knobs to achieve an acceptable value for the chemical/physical objective. The learning algorithm in point (5) operates with a cost functional J, analogous to that used for computational design of fields, but now the cost functional can only depend on those quantities directly observable in the laboratory (i.e., minimally, the current achieved target expectation value). In cases where a physical understanding is sought about the control mechanism, the laser field structure must also be measured. This additional information is typically also not a burden to attain, as it often is only required for the single best control field at the end of the closed-loop learning experiments.

The OCE process in Figure 3 and steps (1)–(5) above constitute the laboratory process for achieving optimal quantum control, and exactly the same reasons put forth for OCT motivate the desire for attaining optimal performance: principally, seeking

the best control result that can possibly be obtained. Seeking optimal performance also ensures some degree of inherent robustness to field noise, as fleeting observations would not likely survive the signal averaging carried out in the laboratory. Additionally, robustness as a specific criterion may also be included in the laboratory cost functional in item (5). When a detailed physical interpretation of the controlled dynamics is desired, it is essential to remove any extraneous control field features that have little impact on the system manipulations. Without the latter clean-up carried out during the experiments, the final field may be contaminated by structures that have little physical impact. Although quantum dynamics typically occurs on femtosecond or picosecond time-scales, the loop excursions in Figure 3 need not be carried out on the latter time-scales. Rather, the loop may be traversed as rapidly as convenient, consistent with the capabilities of the apparatus. A new system sample is introduced on each cycle of the loop. This process is referred learning control, to distinguish it from real-time feedback control.

The closed-loop quantum learning control algorithm can be self-starting, requiring no prior control field designs, under favorable circumstances. Virtually all of the current experiments were carried out in this fashion. It remains to be seen if this procedure of 'going in blind' will be generally applicable in highly complex situations where the initial state is far from the target state. Learning control can only proceed if at least a minimal signal is observed in the target state. Presently, this requirement has been met by drawing on the overwhelmingly large number of exploratory experiments that can be carried out, even in a brief few minutes, under full computer control. However, in some situations, the performance of intermediate observations along the way to the target goal may be necessary in order to at least partially guide the quantum dynamics towards the ultimate objective. Beneficial use could also be made from prior theoretical laser control designs $\boldsymbol{\varepsilon}_0(t)$ capable of at least yielding a minimal target signal.

The OCE closed-loop procedure was based on the growing number of successful theoretical OCT design calculations, even with all of their foibles, especially including less than perfect Hamiltonians. The OCT and OCE processes are analogous, with their primary distinction involving precisely what appears in their corresponding cost functionals. Theoretical guidance can aid in identifying the appropriate cost functionals and learning algorithms for OCE, especially for addressing the needs associated with attaining a physical understanding about the mechanisms governing controlled quantum dynamics phenomena.

A central issue is the rate that learning control occurs upon excursions around the loop in Figure 3. A number of factors control this rate of learning, including the nature of the physical system, the choice of objective, the presence of field uncertainties and measurement errors, the number of control variables, and the capabilities of the learning algorithm employed. Achieving control is a matter of discrimination, and some of these factors, whether inherent to the system or introduced by choice, may work against attaining good-quality discrimination. At this juncture, little is known quantitatively about the limitations associated with any of these factors.

The latter issues have not prevented the successful performance of a broad variety of quantum control experiments, and at present the ability to carry out massive numbers of closed-loop excursions has overcome any evident difficulties. The number of examples of successful closed-loop quantum control is rapidly growing, with illustrations involving the manipulation of laser dye fluorescence, chemical reactivity, high harmonic generation, semiconductor optical switching, fiber optic pulse transmission, and dynamical discrimination of similar chemical species, amongst others. Perhaps the most interesting cases are those carried out at high field intensities, where prior speculation suggested that such experiments would fail due to even modest laser field noise being amplified by the quantum dynamics. Fortunately, this outcome did not occur, and theoretical studies have indicated that surprising degrees of robustness to field noise may accompany the manipulation of quantum dynamics phenomena. Although the presence of field noise may diminish the beneficial influence of constructive and destructive interferences, evidence shows that field noise does not appear to kill the actual control process, at least in terms of obtaining reasonable values for the control objectives. One illustration of control in the high-field regime is shown in Figure 4, demonstrating the learning process for the dissociative rearrangement of acetophenone to form toluene. Interestingly, this case involves breaking two bonds, with the formation of a third, indicating that complex dynamics can be managed with suitably tailored laser pulses.

Operating in the strong-field regime, especially for chemical manipulations, also has the important benefit of producing a generic laser tool to control a broad variety of systems by overcoming the long-standing problem of having sufficient bandwidth. For example, the result in Figure 4 uses a Ti:sapphire laser in the near-infrared regime. Starting with a bandwidth-limited pulse of intensity near

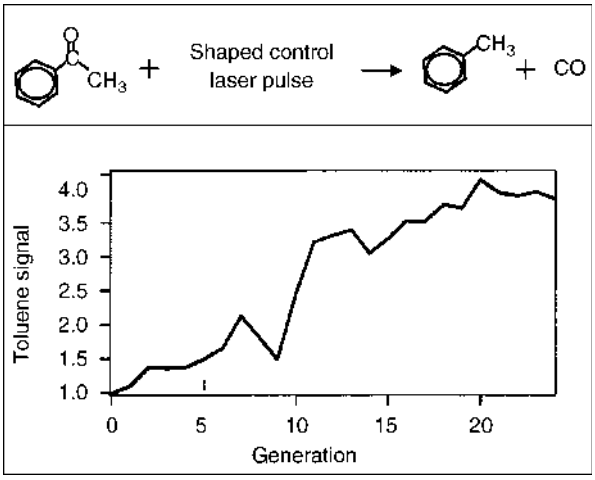

Figure 4 An illustration of dissociative rearrangement achieved by closed-loop learning control in Figure 3. The optimally deduced laser pulse broke two bonds in the parent acetophenone molecule and formed a new one to yield the toluene product. The laboratory learning curve is shown for the toluene product signal as a function of the generations in the genetic algorithm guiding the experiments. The fluctuations in the learning curve for optimizing the toluene yield corresponds to the algorithm searching for the optimal yield as the experiments proceed.

$\sim 10^{14}$ W cm^{-2}, in this regime, the dynamic power broadening can easily be on the order of ~ 1 eV or more, thereby effectively converting the otherwise discrete spectrum of the molecule into an effective continuum for ready multiphoton matching by the control laser pulse. Operating under these physical conditions is very attractive, as the apparatus is generic, permitting a single shaped-pulse laser source (e.g., a Ti:sapphire system with phase and amplitude modulation) to be utilized with virtually any chemical system where manipulations are desired. Although every desired physical/chemical goal may not be satisfactorily met (i.e., the goals must be physically attainable!), the means are now available to explore large numbers of systems. Operating under closed loop in the strong-field tailored-pulse regime eliminates the prior serious limitation of first finding a physical system to meet the laser capabilities; now, the structure of the laser pulse can be shaped to meet the molecule's characteristics. Strong field operations may be attractive for these reasons, but other broadband laser sources, possibly working at weaker field intensities, might be essential in some applications (e.g., controlling electron dynamics in semiconductors, where material damage is to be avoided). It is anticipated that additional broadband laser sources will become available for these purposes.

The coming years should be a period of rapid expansion, including a thorough exploration of closed-loop quantum control capabilities. As with the use of any laboratory tool, certain applications may be more amenable than others to attaining successful control. The particular physical/chemical questions need to be well posed, and controls need to have sufficient flexibility to meet the objectives. The experiments ahead should be able to reveal the degree to which drawing on optimality in OCE, combined with the performance of massive numbers of experiments can lead to broad-scale successful control of quantum phenomena. One issue of concern is the richness associated with the large numbers of phase and amplitude control knobs that may be adjusted in the laboratory. Some experiments have already operated with hundreds of knobs, while others have restricted their number in a variety of ways, to simplify the search process. Additional technological and algorithmic advances may be required to manage the high-dimensional control space searches. Fortunately, for typical applications, the search does not reduce to seeking a needle in a haystack, as generally there are multiple control solutions, possibly all of very good quality.

As a final comment on OCE, it is useful to appreciate the subtle distinction of the procedure from closed-loop feedback experiments. This distinction is illustrated in Figure 5, pointing to three types of closed-loop quantum control laboratory experiments. The OCE procedure in Figure 5a produces a generic laser tool capable of controlling a broad variety of systems with an emphasis in the figure placed on the point that each cycle around the loop starts with a new sample for control. The replacement of the sample on each cycle eliminates a number of difficulties, principally including concerns about sample damage, avoidance of operating the loop at the ultrafast speed of the quantum mechanical processes, and elimination of the effect that 'to observe is to disturb' in quantum mechanics. Thus, learning control provides a generic practical procedure, regardless of the nature of the quantum system. Figure 5b has the same structure as that of Figure 5a, except the loop is now closed around the same single quantum system, which is followed throughout its evolution. All of the issues mentioned above that are circumvented by operating through laboratory learning control in Figure 5a must now be directly faced in the setup of Figure 5b. The procedure in Figure 5b will likely only be applicable in special circumstances, at least for the reason that many quantum systems operate on time-scales far too fast for opto-electronics and computers to keep up with. However, there are certain quantum mechanical processes that are sufficiently slow to meet the criteria. Furthermore, a period of free evolutionary

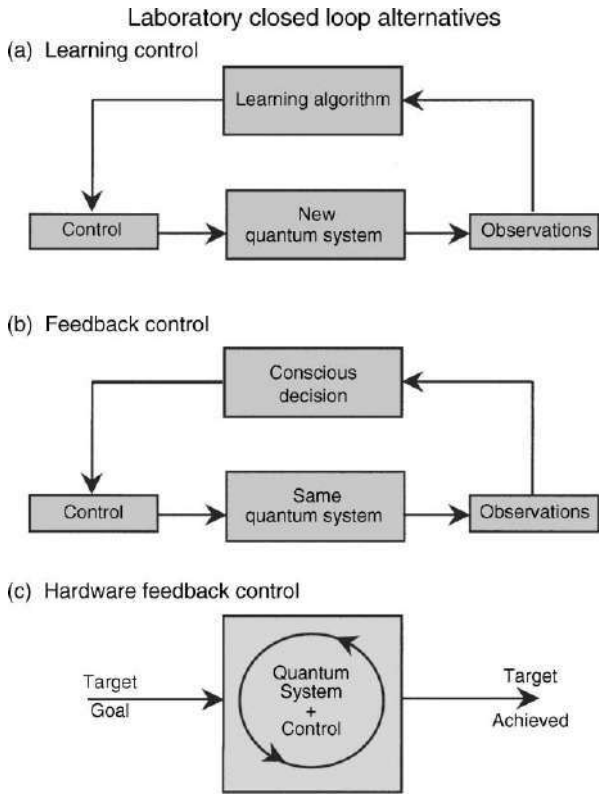

Figure 5 Three possible closed-loop formulations for attaining laser control over quantum phenomena. (a) is a learning control process where a new system is introduced on each excursion around the loop. (b) utilizes feedback control with the loop being traversed with the same sample, generally calling for an accurate model of the system to adjust the controls on each loop excursion. (c) is, at present, a dream machine where the laser and the sample under control are one unit operating without external algorithmic guidance. The closed-loop learning control procedure in (a) appears to form the most practical means to achieve laser control over quantum phenomena, especially for complex systems.

quantum dynamics may be permitted to occur between one control pulse and the next, to make the time issue manageable. While the learning control process in Figures 3 and 5a can be performed model-free, the feedback algorithm in Figure 5b generally must operate with a sufficiently reliable system Hamiltonian to carry out fast design corrections to the control field, based on the previous control outcome. These are very severe demands, but they may be met under certain circumstances. One reason for considering closed-loop feedback control in Figure 5b is to explore the basic physical issue of quantum mechanical limitations inherent in the statement 'to observe is to disturb'. It is suggestive that control over many types of quantum phenomena may fall into the semiclassical regime, lying somewhere between classical engineering behavior and the hard limitations of quantum mechanics. Experiments of the type in Figure 5b will be most interesting for exploring this matter. Finally, Figure 5c introduces a *gedanken* experiment in the sense that the closed-loop process is literally built into the hardware. That is, the laser control and quantum system act as a single functioning unit operating in a stable fashion, so as to automatically steer the system to the desired target. This process may involve engineering the quantum mechanical system prior to control in cases where that freedom exists, as well as engineering the laser components involved. The meaning of such a device in Figure 5c can be understood by considering an analogy with airplane flight, where the aircraft is constructed to have an inherent degree of aerodynamic stability and will essentially fly (glide) on its own accord when pushed forward. It is an open question whether closing the loop in the hardware can be attained for quantum control, and an exploration of this concept may require additional laser technologies.

Conclusions

This article presented an overview of theoretical concepts and algorithmic considerations associated with the control of quantum phenomena. Theory has played a central role in this area by revealing the fundamental principles underlying quantum control, as well as by providing algorithms for designing controls and guiding experiments to discover successful controls in the laboratory. The challenge of controlling quantum phenomena, in one sense, is an old subject, going back at least 40 years. However, roughly the first 30 of these years would best be described as a period of frustration, due to a lack of full understanding of the principles involved and the nature of the lasers needed to achieve success. Thus, from another perspective, the subject is quite young, with perhaps the most notable development being the introduction of closed-loop laboratory learning procedures. At the time of writing this article, these procedures are just beginning to be explored for their full capabilities. To appreciate the young nature of this subject, it is useful to note that the analogous engineering control disciplines presently occupy many thousands of engineers worldwide, both theoreticians and practitioners, and have done so for many years. Yet, engineering control is far from considered a mature subject. Armed now with the basic concepts and proper laboratory tools, one may anticipate a thorough exploration of control over quantum phenomena, including its many possible applications.

Acknowledgments

Support for this work has come from the National Science Foundation and the Department of Defense.

Further Reading

Brixner T, Damrauer NH and Gerber G (2001) Femtosecond quantum control. *Advances in Atomic, Molecular, and Optical Physics* 44: 1–54.

Rabitz H and Zhu W (2000) Optimal control of molecular motion: Design, implementation and inversion. *Accounts of Chemical Research* 33: 572–578.

Rabitz H, de Vivie-Riedle R, Motzkus M and Kompa K (2000) Whither the future of controlling quantum phenomena? *Science* 288: 824–828.

Rice SA and Zhao M (2000) Optical Control of Molecular Processes. NewYork: John Wiley.

Experimental

R J Levis, Temple University, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The experimental control of quantum phenomena in atoms, molecules, and condensed phase materials has been a long-standing challenge in the physical sciences. For example, the idea that a laser beam could be used to selectively cleave a chemical bond has been pursued since the demonstration of the laser in the early 1960s. However, it was not until very recent times that one could realize selective bond cleavage. One key to the rapid acceleration of such experiments is the capability of manipulating the relative phases between two or more paths leading to the same final state. This is done in practice by controlling the phase of two (or two thousand!) laser frequencies coupling the initial state to some desired final state. The present interest in phase-related control stems in part from research focusing on quantum information sciences, controlling chemical reactivity, and developing new optical technologies, such as in biophotonics for imaging cellular materials. Ultimately, one would like to identify coherent control applications having impact similar to linear regime applications like compact disk reading (and writing), integrated circuit fabrication, and photodynamic therapy. Note that in each of these the phase of the electromagnetic field is not an important parameter and only the brightness of the laser at a particular frequency is used. While phase is of negligible importance in any one photon process, for excitations involving two or more photons, phase not only becomes a useful characteristic, it can modulate the yield of a desired process between zero and one hundred percent.

The use of optical phase to manipulate atomic and molecular systems rose to the forefront of laser science in the 1980s. Prior to this there was considerable effort directed toward using the intensity and high spectral resolution features of the laser for various applications. For example, the high spectral resolution character was employed in purifying nuclear isotopes. Highly intense beams were also used in the unsuccessful attempt to pump enough energy into a single vibrational resonance to selectively dissociate a molecule. The current revolution in coherent control has been driven in part by the realization that controlling energy deposition into a single degree of freedom (such as selectively breaking a bond in a polyatomic molecule) can not be accomplished by simply driving a resonance with an ever more intense laser source. In the case of laser-induced photo dissociation, such excitation ultimately results in statistical dissociation throughout the molecule due to vibrational mode coupling. In complex systems, controlling energy deposition into a desired quantum state requires a more sophisticated approach and one scheme involves coherent control.

The two distinct regimes for excitation in coherent control, the time and frequency domain experiments, are formally equivalent. While the experimental approach and level of success for each domain is vastly different, both rely on precisely specifying the phase relations for a series of excitation frequencies. The basic idea behind phase control is simple. One emulates the interference effects that can be easily visualized in a water tank experiment with a wave propagating through two slits. In the coherent control experiment, two or more indistinguishable pathways must be excited that connect one quantum state to another. In either water waves, or coherent control, the probability for observing constructive or destructive interference at a target state depends on the phase difference between the paths connecting the two locations in the water tank, or the initial and final states in the quantum system. In the case of coherent control, the final state may have some technological

value, for instance, a desirable chemical reaction product or an intermediate in the implementation of quantum computation.

Frequency domain methods concentrate on interfering two excitation routes in a system by controlling the phase difference between two distinct frequencies coupling an initial state to a final state. Perhaps the first mention of the concept of phase control in the optical regime can be traced to investigations in Russia in the early 1960s, where the interference between a one- and three-photon path was theoretically proposed as a means to modulate a two-photon absorption. This two-path interference idea lay dormant until the reality of controlling the relative phase of two distinct frequencies was achieved in 1990 using gas phase pressure modulation for frequency domain experiments. The idea was theoretically extended to the control of nuclear motion in 1988. The first experimental example of two-color phase control over nuclear motion was demonstrated in 1991. It is interesting that the first demonstrations of coherent control did not, however, originate in the frequency domain approach, but came from time domain investigations in the mid-1980s. This was work that grew directly out of the chemical dynamics community and will be described below.

While frequency domain methods were developed exclusively for the observation of quantum interference, the experimental implementation of time-domain methods was first developed for observing nuclear motion in real time. In the time domain, a superposition of multiple states is prepared in a system using a short duration (~ 50 fs) laser pulse. The superposition state will subsequently evolve according to the phase relationships of the prepared states and the Hamiltonian of the system. At some later time the evolution must be probed using a second, short-time-duration laser pulse. The time-domain methods for coherent control were first proposed in 1985 and were based on wavepacket propagation methods developed in the 1970s. The first experiments were reported in the mid-1980s.

To determine whether the time or frequency domain implementation is more applicable for a particular control context, one should compare the relevant energy level spacing of the quantum system to the bandwidth of the excitation laser. For instance, control of atomic systems is more suited to frequency domain methods because the characteristic energy level spacing in an atom (several electron volts) is large compared to the bandwidth of femtosecond laser pulses (millielectron volts). In this case, preparation of a superposition of two or more states is impractical with a single laser pulse at the present time, dictating that the control scheme must employ the initial state and a single excited eigenstate. In practice, quasi-CW nanosecond duration laser pulses are employed for such experiments because current ultrafast (fs duration) laser sources cannot typically prepare a superposition of electronic states. One instructive exception involves the excitation of Rydberg states in atoms. Here the electronic energy level spacing can be small enough that the bandwidth of a femtosecond excitation laser may span several electronic states and thus can create a superposition state. One should also note that the next generation of laser sources in the attosecond regime may permit the preparation of a wavepacket from low-lying electronic states in an atom. In the case of molecules, avoiding a superposition state is nearly impossible. Such experiments employ pico- to femtosecond duration laser pulses, having a bandwidth sufficient to excite many vibrational levels, either in the ground or excited electronic state manifold of a molecule. To complete the time-dependent measurement, the propagation of the wavepacket to the final state of interest must be observed (probed) using a second ultrafast laser pulse that can produce fluorescence, ionization, or stimulated emission in a time-resolved manner. The multiple paths that link the initial and final state via the pump and the probe pulses are the key to the equivalence of the time and frequency domain methods.

There is another useful way to classify experiments that have been performed recently in coherent control, that is into either open-loop or closed-loop techniques. Open-loop signifies that a calculation may be performed to specify the required pulse shape for control before the experiment is attempted. All of the frequency-based, and most of the time-based experiments fall into this category. Closed-loop, on the other hand, signifies that the results from experimental measurements must be used to assist in determining the correct parameters for the next experiment, this pioneering approach was suggested in 1992. It is interesting that in the context of control experiments performed to date, closed-loop experiments represent the only route to determining the optimal laser pulse shape for controlling even moderately complex systems. In this approach, no input from theoretical models is required for coherent control of complex systems. This is unlike typical control engineering environments, where closed-loop signifies the use of experiments to refine parameters used in a theoretical model.

There is one additional distinction that is of value for understanding the state of coherent control at the present time. Experiments may be classified into systems having Hamiltonians that allow calculation of frequencies required for the desired interference

patterns (thus enabling open-loop experiments) and those having ill-defined Hamiltonians (requiring closed-loop experiments). The open-loop experiments have demonstrated the utility of various control schemes, but are not amenable to systems having even moderate complexity. The closed-loop experiments are capable of controlling systems of moderate complexity but are not amenable to precalculation of the required control fields. Systems requiring closed-loop methods include propagation of short pulses in an optical fiber, control of high harmonic generation for soft X-ray generation, control of chemical reactions and control of biological systems. In the experimental regime, the value of learning control for dealing with complex systems has been well documented.

The remainder of this article represents an overview of the experimental implementation of coherent control. The earliest phase control experiments involved time-dependent probing of molecular wavepackets, and were followed by two-path interference in an atomic system. The most recent, and general, implementation of control involves tailoring a time-dependent electromagnetic field for a desired objective using closed-loop techniques. Since the use of tailored laser pulses appears to be rather general, a more complete description of the experiment is provided.

Time-Dependent Methods

The first experiments demonstrating the possibility of coherent control were performed to detect nuclear motion in molecules in real time. These investigations were the experimental realization of the pump-dump or pulse-timing control methods as originally described in 1985. The experimental applications have expanded to numerous systems including diatomic and triatomic molecules, small clusters, and even biological molecules. For such experiments, a vibrational coherence is prepared by exciting a superposition of states. These measurements implicitly used phase control both in the generation of the ultrashort pulses and in the coherent probing of the superposition state by varying the time delay between the pump and probe pulses, in effect modulating the outcome of a quantum transition. Since the earliest optical experiment in the mid-1980s, there have been many hundreds of such experiments reported in the literature. The motivation for the very first experiments was not explicitly phase control, but rather the observation of nuclear motion of molecules in real time. In such later measurements, oscillations in the superposition states were observable for many cycles, up to many tens of picoseconds in favorable cases, i.e., diatomic molecules in the gas phase. The loss of signal in the coherent oscillations is ultimately due to passage of the superposition state in to surrounding bath states. This may be due to collision with another molecule, dissociation of the system or redistribution of the excitation energy into other modes of the system not originally excited in the superposition. It is notable that coherence can be maintained in solution phase systems for tens of picoseconds and for the case of resonances having weak coupling to other modes, coherence can even be modulated in large biological systems on the picosecond time scale.

Decoherence represents loss of phase information about a system. In this context, phase information represents our detailed knowledge of the electronic and nuclear coordinates of the system. In the case of vibrational coherence, decoherence may be represented by intramolecular vibrational energy transfer to other modes of the molecule. This can be thought of as the propensity of vibrational modes of a molecule to couple together in a manner that randomizes deposited energy throughout the molecule. Decoherence in a condensed phase system includes transfer of energy to the solvent modes surrounding the system of interest. For an electronically excited molecule or atom, decoherence can involve spontaneous emission of radiation (fluorescence), or dissociation in the case of molecules. Certain excited states may have high probability for fluorescence and would represent a significant decoherence pathway and an obstacle for coherent control. These states may be called 'lossy' and are often used in optical pumping schemes as well as for stimulated emission pumping. An interesting question arises as to whether one can employ such states in a pathway leading to a desired final state, or must such lossy states be avoided at all costs in coherent control. This question has been answered to some degree by the method of rapid adiabatic passage. In this experiment, an initial state is coupled to a final state by way of a lossy state. The coupling is a coherent process and the preparation involves dressed states that are eigenstates of the system. In this coherent superposition the lossy state is employed, but no population is allowed to build up and thus decoherence is circumvented.

Two-Path Interference Methods

Interference methods form perhaps the most intuitive example of coherent control. The first experimental demonstration was reported in 1990 and involved the modulation of ionization probability in Hg by interfering a one-photon and three-photon excitation

pathway to an excited electronic state (followed by two-photon ionization). The long delay between the introduction of the two-path interference concept in 1960 and the demonstration in 1990 was the difficulty in controlling the relative phase of the one- and three-photon paths. The solution involved generating the third harmonic of the fundamental in a nonlinear crystal, and copropagating the two beams through a low-density gas. Changing the pressure of the gas changes the optical phase retardance at different rates for different frequencies, thus allowing relative phase control between two colors. The ionization yield of Hg was clearly modulated as a function of pressure. Such schemes have been extended to diatomic molecular systems and the concept of determining molecular phase has been investigated. While changing the phase of more than two frequencies has not been demonstrated using pressure tuning (because of experimental difficulties), a more flexible method has been developed to alter relative phase of up to 1000 frequency bands, as will be described next.

Closed-Loop Control Methods

Perhaps the most exciting new aspect of control in recent years concerns the prospect for performing closed-loop experiments. In this approach a more complex level of control in matter is achieved in comparison to the two-path interference and pump-probe control methods. In the closed-loop method, an arbitrarily complex time-dependent electromagnetic field is prepared that may have multiple subpulses and up to thousands of interfering frequency bands. The realization of closed-loop control over complex processes relies on the confluence of three technologies: (i) production of large bandwidth ultrafast laser pulses; (ii) spatial modulation methods for manipulating the relative phases and amplitudes of component frequencies of the ultrafast laser pulse; and (iii) closed-loop learning control methods for sorting through the vast number of pulse shapes available as potential control fields. In the closed-loop method, the result of a laser molecule interaction is used to tailor an optimized pulse.

In almost all of the experimental systems used for closed-loop control today, Kerr lens mode-locking in the Ti:sapphire crystal is used to phase lock the wide bandwidth of available frequencies (750–950 nm) to create the ultrafast pulse. In this phenomenon, the intensity variation in the short laser pulse creates a transient lens in the Ti:sapphire lasing medium that discriminates between free lasing and mode-locked pulses. As a result, all of the population inversion available in the lasing medium may be coerced into enhancing the ultrashort traveling wave in the cavity in this way. The wide bandwidth available may then be amplified and tailored into a shaped electromagnetic field. Spatial light modulation is one such method for modulating the relative phases and amplitudes of the component frequencies in the ultrafast laser pulse to tailor the shaped laser pulse. Pulse shaping first involves dispersing the phase-locked frequencies on an optical grating, the dispersed radiation is then collimated using a cylindrical lens and the individual frequencies are focused to a line forming the Fourier plane. At the Fourier plane, the time-dependent laser pulse is transformed into a series of phase-locked continuous wave (CW) laser frequencies, and thus Fourier transformed from time to frequency space. The relative phases and amplitudes may be modulated in the Fourier plane using an array of liquid crystal pixels. After altering the spectral phase and amplitude profile, the frequencies are recombined on a second grating to form the shaped laser pulse. With one degree of phase control and 10 pixels there are an astronomic number (360^{10}) of pulse shapes that may be generated in this manner. (At the present time, pulse shapers have up to 2000 independent elements.) Evolutionary algorithms are employed to manage the available phase space. Realizing that the pulse shape is nothing more than the summation of a series of sine waves, each having an associated phase and amplitude, we find that a certain pulse shape can be represented by a genome consisting of an array of these frequency-dependent phases and amplitudes. This immediately suggests that the methods of evolutionary search strategies may be useful for determining the optimal pulse shape for a desired photoexcitation process.

The method of closed-loop optimal control for experiments was first proposed in 1992 and the first experiments were reported in 1997 regarding optimization of the efficiency of a laser dye molecule in solution. Since that time a number of experiments have been performed in both the weak and strong field regime. In the weak field regime, the intensity of the laser does not alter the field free structure of the excited states of the system under investigation. In the strong field, the field free states of the system are altered by the electric field of the laser. Whether one is in the weak or strong field regime depends on parameters including the characteristic level spacing of the states of the system to be controlled and the intensity of the laser coupling into those states. Examples of weak field closed-loop control include optimization of laser dye efficiency, compression of an ultrashort laser pulse, dissociation of alkali clusters, optimization of coherent anti-Stokes

Raman excitation, and optical pulse propagation in fibers. In the strong field, the laser field is used to manipulate the excited states of the molecule as well as induce population transfer among these states. In this sense, the laser is both creating and exciting resonances, in effect allowing universal excitation with a limited frequency bandwidth of 750–850 nm. For chemical applications, it is worth noting that most molecules do not absorb in this wavelength region in the weak field. Strong field control has been used for controlling bond dissociation, chemical rearrangement, photonic reagents, X-ray generation, molecular centrifuges and the manipulation of mass spectral fragmentation intensities. In the latter case, a new sensing technology has emerged wherein each individual pulse shape represents an independent sensor for a molecule. Given the fact that thousands of pulse shapes can be tested per minute, this represents a new paradigm for molecular analysis.

At the time of this writing, the closed-loop method for adaptively tailoring control fields, has far outstripped our theoretical understanding of the process (particularly in the strong field regime). Adaptive control is presently an active field of investigation, encompassing the fields of physics, engineering, optics, and chemistry. In the coming years, there will be as much work done on the mechanism of control as in the application of the methods. Due to small energy level spacings and complexity of the systems, we anticipate the majority of applications to be in the chemical sciences.

See also

Coherence: Coherence and Imaging; Overview. **Interferometry:** Overview. **Microscopy:** Interference Microscopy.

Further Reading

Brumer PW and Shapiro M (2003) *Principles of the Quantum Control of Molecular Processes*. New York: Wiley Interscience.

Chen C and Elliott DS (1990) Measurement of optical phase variations using interfering multiphoton ionization processes. *Physics Review Letters* 65: 1737.

Levis RJ and Rabitz H (2002) Closing the loop on bond selective chemistry using tailored strong field laser pulses. *Journal of Physical Chemistry A* 106: 6427–6444.

Levis RJ, Menkir GM and Rabitz H (2001) Selective bond dissociation and rearrangement with optimally tailored, strong-field laser pulses. *Science* 292(5517): 709–713.

Rice SA and Zhao M (2000) *Optical Control of Molecular Dynamics*. New York: Wiley Interscience.

Stolow A (2003) Femtosecond time-resolved photoelectron spectroscopy of polyatomic molecules. *Annual Review of Physical Chemistry* 54: 89–119.

Zewail AH, Casati G, Rice SA, *et al.* (1997) Femtochemistry: Chemical reaction dynamics and their control. *Chemical Reactions and Their Control on the Femtosecond Time Scale, Solvay Conference on Chemistry* 101: 3–45.

Applications in Semiconductors

H M van Driel and J E Sipe, University of Toronto, Toronto, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Interference phenomena are well-known in classical optics. In the early 19th century, Thomas Young showed conclusively that light has wave properties, with the first interference experiment ever performed. He passed quasi-monochromatic light from a single source through a pair of double slits using the configuration of Figure 1a and observed an interference pattern consisting of a series of bright and dark fringes on a distant screen. The intensity distribution can be explained only if it is assumed that light has wave or phase properties. In modern terminology, if E_1 and E_2 are the complex electric fields of the two light beams arriving at the screen, by the superposition principle of field addition the intensity at a particular point on the screen can be written as

$$\begin{aligned} I &\propto |E_1 + E_2|^2 \\ &= |E_1|^2 + |E_2|^2 + |E_1||E_2| \cos(\phi_1 - \phi_2) \end{aligned} \qquad [1]$$

where $\phi_1 - \phi_2$ is the phase difference of the beams arriving at the screen. While this simple experiment was used originally to demonstrate the wave properties of light, it has subsequently been used for many other purposes, such as measuring the wavelength of light. However, for our purposes, here the Young's double slit apparatus can also be viewed as a device that redistributes light, or

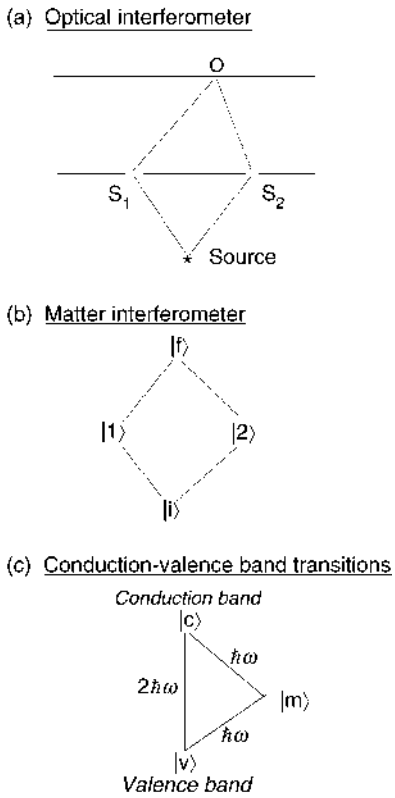

Figure 1 (a) Interference effects in the Young's double slit experiment; (b) illustration of the general concept of coherence control via multiple quantum mechanical pathways; (c) interference of single- and two-photon transitions connecting the same valence and conduction band states in a semiconductor.

controls its intensity at a particular location. For example, let's consider one of the slits to be a source, with the other slit taken to be a gate, with both capable of letting through the same amount of light. If the gate is closed, the distant screen is nearly uniformly illuminated. But if the gate is open, at a particular point on the distant screen there might be zero intensity or as much as four times the intensity emerging from one slit because of interference effects, with all the light merely being spatially redistributed. In this sense the double slit system can be viewed as a device to redistribute the incident light intensity. The key to all this is the superposition principle and the properties of optical phase.

The superposition principle and interference effects also lie at the heart of quantum mechanics. For example, let's now consider a system under the influence of a perturbation with an associated Hamiltonian that has phase properties. In general the system can evolve from one quantum state $|i\rangle$ to another $|f\rangle$ via multiple pathways involving intermediate states $|m\rangle$. Because of the phased perturbation, interference between those pathways can influence the system's final state. If a_m is the (complex) amplitude associated with a transition from the initial to the final state, via intermediate virtual state $|m\rangle$, then for all possible intermediate states the overall transition probability, W, can be written as

$$W = \left|\sum_m a_m\right|^2 \qquad [2]$$

In the case of only two pathways, as illustrated in Figure 1b, W becomes

$$\begin{aligned} W &= |a_1 + a_2|^2 \\ &= |a_1|^2 + |a_2|^2 + |a_1||a_2|\cos(\phi_1 - \phi_2) \end{aligned} \qquad [3]$$

where ϕ_1 and ϕ_2 are now the phases of the two transition amplitudes. While this expression strongly resembles that of eqn [1], here ϕ_1 and ϕ_2 are influenced by the phase properties of the perturbation Hamiltonian that governs the transition process; these phase factors arise, for example, if light fields constitute the perturbation. But overall it is clear that the change in the state of the system is affected by both the amplitude and phase properties of the perturbation. Analogous with the classical interferometer, equality of the transition amplitudes leads to maximum contrast in the transition rate. Although it has been known since the early days of quantum mechanics that the phase of a perturbation can influence the evolution of a system, generally phase has only been discussed in a passive role. Only in recent years has phase been used as a control parameter for a system, on the same level as the strength of the perturbation itself.

Coherence Control

Coherence control, or quantum control as it is sometimes called, refers to the active process through which one can use phase-dependent perturbations originating with, for example, coherent light waves, to control one or more properties of a quantum system, such as state population, momentum, or spin. This more general perspective of interference leads to a picture of interference of matter waves, rather than simply light waves, and one can now speak of an effective 'matter interferometer'. Laser beams, in particular, are in a unique position to play a role in such processes, since they offer a macroscopic 'phase handle' with which to create such effects. This was recognized by the community of atomic and molecular scientists who emphasized the active manifestations of quantum interference effects, and proposed that branching ratios in photochemical reactions

might be controlled through laser-induced interference processes. The use of phase as a control parameter also represents a novel horizon of applications for the laser since most previous applications had involved only amplitude (intensity).

Of course, just as the 'visibility' of the screen pattern for a classical Young's double slit interferometer is governed by the coherence properties of the two slit source, the evolution of a quantum system reflects the coherence properties of the perturbations, and the degree to which one can exercise phase control via external perturbations is also influenced by the interaction of the system of interest with a reservoir – essentially any other degrees of freedom in the system – which can cause decoherence and reduce the effectiveness of the 'matter interferometer'. Since the influence of a reservoir will generally increase with the complexity of a system, one might think that coherence control can only be effectively achieved in simple atomic and molecular systems. Indeed, some of the earliest suggestions for coherence control of a system involved using interference between single and three photon absorption processes, connecting the same initial and final states with photons of frequency 3ω and ω, respectively, and controlling the population of excited states in atomic or diatomic systems. However, it has been difficult to extend this 'two color' paradigm to more complex molecular systems. Since the reservoir leads to decoherence of the system overall, shorter and shorter pulses must be used to overcome decoherence effects. However, short pulses possess a large bandwidth with the result that, for complex systems, selectivity of the final state can be lost. One must therefore consider using interference between multiple pathways in order to control the system, as dictated by the details of the unperturbed Hamiltonian of the system.

For a polyatomic molecule or a solid, the complete Hamiltonian is virtually impossible to determine exactly and it is therefore equally difficult to prescribe the optimal spectral (amplitude and phase) content of the pulses that should be used for control purposes. An alternative approach has therefore emerged, in which one foregoes knowledge of the eigenstates of the Hamiltonian and details of the different possible interfering transition paths in order to achieve control of the end state of a system. This branch of coherence control has come to be known as optimal control. In optimal control one employs an optical source for which one can (ideally) have complete control over the spectral and phase properties. One then uses a feedback process in which experiments are carried out, the effectiveness of achieving a certain result is determined, and the pulse characteristics are then altered to obtain a new result. A key component is the use of an algorithm to select the new pulse properties as part of the feedback system. In this approach the molecule teaches the external control system what it 'requires' for a certain result to be optimally achieved. While the 'best' pulse properties may not directly reveal details of the multiple interference process required to achieve the optimal result, this technique can nonetheless be used to gain some insight into the properties of the unperturbed Hamiltonian and, regardless of such understanding, achieve a desirable result. It has been used to control chemical reaction rates involving several polyatomic molecules, with considerable enhancement in achieving a certain product relative to what can be done using simple thermodynamics.

Coherence Control in Semiconductors

Since coherence control of chemical reactions involving large molecules has represented a significant challenge, it was generally felt that ultrafast decoherence processes would also make control in solids, in general, and semiconductors, in particular, difficult. Early efforts therefore focused on atomic-like situations in which the electrons are bound and associated with discrete states and long coherence times. In semiconductors, the obvious choice is the excitonic system, defect states, or discrete states offered by quantum wells. Population control of excitons and directional ionization of electrons from quantum wells has been clearly demonstrated, similar to related population control of and directional ionization from atoms. Other manifestations of coherence control of semiconductors include control of electron–phonon interactions and intersub-band transitions in quantum wells.

Among the different types of coherence control in semiconductors is the remarkable result that it is possible to coherently control the properties of free electrons associated with continuum states. Although optimal control might be used for some of these processes, we have achieved clear illustrations of control phenomena based on the use of harmonically related beams and interference of single- and two-photon absorption processes connecting the continuum valence and conduction band states as generically illustrated in Figure 1c. Details of the various processes observed can be found elsewhere. Of course, the momentum relaxation time of electrons or holes in continuum states is typically of the order of 100 fs at room temperature, but this time is sufficiently long that it can permit phase-controlled processes. Indeed, with respect to conventional carrier transport, this 'long time' lapse is responsible

for the typically high electrical mobilities in crystalline semiconductors such as Si and GaAs. In essence, a crystalline semiconductor with translational symmetry has crystal momentum as a good quantum number, and selection rules for scattering prevent the momentum relaxation time from being prohibitively small. Since electrons or holes in any continuum state can participate in such control processes, one need not be concerned about pulsewidth or bandwidth, unless the pulses were to be so short that carriers of both positive and negative effective mass were generated within one band.

Coherent Control of Electrical Current Using Two Color Beams

We now illustrate the basic principles that describe how the interference process involving valence and conduction band states can be used to control properties of a bulk semiconductor. We do so in the case of using one or two beams to generate and control carrier population, electrical current, and spin current. In pointing out the underlying ideas behind coherence control of semiconductors, we will skip or suppress many of the mathematical details which are required for a full understanding but which might obscure the essential physics. In particular we consider how phase-related optical beams with frequencies ω and 2ω interact with a direct gap semiconductor such that $\hbar\omega < E_g < 2\hbar\omega$ where E_g is the electronic bandgap. For simplicity we consider exciting electrons from a single valence band via single-photon absorption at 2ω and two-photon absorption at ω. As is well known, within an independent particle approximation, the states of electrons and holes in semiconductors can be labeled by their vector crystal momentum $\mathbf{k}$; the energy of states near the conduction or valence bandedges varies quadratically with $|\mathbf{k}|$. For one-photon absorption the transition amplitude can be derived using a perturbation Hamiltonian of the form $H = e/mc\,\mathbf{A}^{2\omega}\cdot\mathbf{p}$ where $\mathbf{A}^{2\omega}$ is the vector potential associated with the light field and $\mathbf{p}$ is the momentum operator. The transition amplitude is therefore of the form

$$\mathrm{a}^{2\omega} \propto E^{2\omega} e^{i\phi_{2\omega}} p_{cv} \qquad [4]$$

where p_{cv} is the interband matrix element of $\mathbf{p}$ along the field ($E^{2\omega}$) direction; for illustration purposes the field is taken to be linearly polarized. The overall transition rate between two particular states of the same $\mathbf{k}$ can be expressed as $W_1 \propto \mathrm{a}^{2\omega}(\mathrm{a}^{2\omega})^* \propto I^{2\omega}|p_{cv}|^2$ where $I^{2\omega}$ is the intensity of the beam. This rate is independent of the phase of the light beam as well as the sign of $\mathbf{k}$. Hence the absorption of light via single-photon transitions generally populates states of equal and opposite momentum with equal probability or, equivalently, establishes a standing electron wave with zero crystal momentum. This is not surprising since photons possess very little momentum and, in the approximation of a uniform electric field, do not give any momentum to an excited electron. The particular states that are excited depends on the light polarization and crystal orientation. However, the main point is that while single-photon absorption can lead to anisotropic filling of electron states, the distribution in momentum space is not polar (dependent on sign of $\mathbf{k}$). Similar considerations apply to the excited holes, but to avoid repetition we will focus on the electrons only.

For two-photon absorption involving the ω photons and connecting states similar to those connected with single-photon absorption, one must employ the 2nd order perturbation theory using the Hamiltonian $H = e/mc\mathbf{A}^{\omega}\cdot\mathbf{p}$. For two-photon absorption there is a transition from the valence band to an (energy nonallowed) intermediate state followed by a transition from the intermediate state to the final state. To determine the total transition rate one must sum over all possible intermediate states. For semiconductors, by far the dominant intermediate state is the final state itself, so that the associated transition amplitude has the form

$$\mathrm{a}^{\omega} \propto (E^{\omega} e^{i\phi_{\omega}} p_{cv})(E^{2\omega} e^{i\phi_{\omega}} p_{cc}) \propto (E^{\omega})^2 e^{2i\phi_{\omega}} p_{cv}\hbar k \qquad [5]$$

where the matrix element p_{cc} is simply the momentum of the conduction band state ($\hbar k$) along the direction of the field. Note that unlike an atomic system, p_{cc} is nonzero, since Bloch states in a crystalline solid do not possess inversion symmetry. If two-photon absorption acts alone, the overall transition rate between two particular $\mathbf{k}$ states would be $W_2 \propto (I^{\omega})^2|p_{cv}|^2k^2$. As with single-photon absorption, because this transition rate is independent of the sign of k, two-photon absorption leads to production of electrons with no net momentum.

When both single- and two-photon transitions are present simultaneously, the transition amplitude is the sum of the transition amplitudes expressed in eqn [3]. The overall transition rate is then found using eqn [1] and yields $W = W_1 + W_2 + Int$ where the interference term Int is given by

$$Int \propto E^{2\omega}E^{\omega}E^{\omega}\sin(\phi_{2\omega} - 2\phi_{\omega})k \qquad [6]$$

Note, however, that the interference effect depends on the sign of k and hence can be constructive for one part of the conduction band but destructive for other parts of that band, depending on the value of the

relative phase, $\Delta\phi = \phi_{2\omega} - 2\phi_{\omega}$. In principle one can largely eliminate transitions with $+\mathbf{k}$ and enhance those with $-\mathbf{k}$. This effectively generates a net momentum for electrons or holes and hence, at least temporarily, leads to an electrical current in the absence of any external bias. The net momentum of the carriers, which is absent in the individual processes, must come from the lattice. Because the carriers are created with net momentum during the pulses one has a form of current injection that can be written as

$$dJ/dt \propto E^{2\omega}E^{\omega}E^{\omega}\sin(\Delta\phi) \qquad [7]$$

where J is the current density. This type of current injection is allowed in both centrosymmetric and noncentrosymmetric materials. The physics and concepts behind the quantum interference leading to this form of current injection are analogous with the Young's double slit experiment. Note as well that if this 'interferometer' is balanced (single- and two-photon transition rates are similar), then it is possible to control the overall transition rate with great contrast. Roughly speaking, one balances the two arms of the 'effective interferometer' involved in the interference process, one 'arm' corresponding to the one-photon process and the other to the two-photon process. As an example, let's consider the excitation of GaAs, which has a room-temperature bandgap of 1.42 eV (equivalent to 870 nm). For excitation of GaAs using 1550 and 775 nm light under balanced conditions, the electron cloud is injected with a speed close to 500 km s^{-1}. Under 'balanced' conditions, nearly all the electrons are moving in the same direction. Assuming that the irradiance of the 1550 nm beam is 100 MW cm^{-2}, while that of the second harmonic beam is only 15 kW cm^{-2} (satisfying the 'balance' condition), with Gaussian pulse widths of 100 fs, one obtains a surprisingly large peak current of about 1 kA cm^{-2} for a carrier density of only 10^{14} cm^{-3} if scattering effects are ignored. When scattering is taken into account, the value of the peak current is reduced and the transient current decays on a time-scale of the momentum relaxation time. Figure 2a shows an experimental setup which can be used to demonstrate coherence control of electrical current using the above parameters. Figure 2a shows the region between a pair of electrodes on GaAs being illuminated by a train of harmonically related pulses. Figure 2b illustrates how the steady-state experimental voltage across the capacitor changes as the phase parameter $\Delta\phi$ is varied. Transient electrical currents, generated though incident femtosecond optical pulses, have also been detected through the emission of the associated Terahertz radiation.

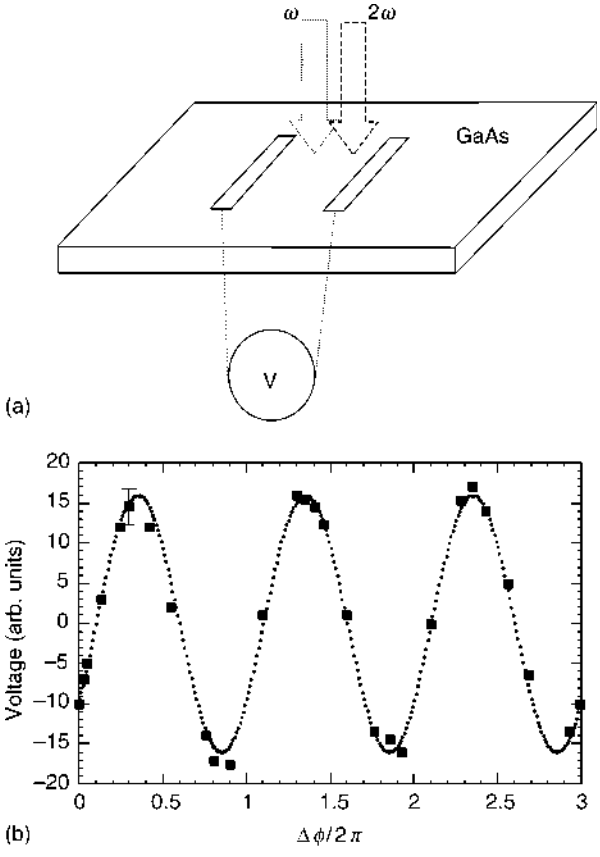

Figure 2 (a) Experimental setup to measure steady-state voltage (V) across a pair of electrodes on GaAs with the intervening region illuminated by phased radiation at ω and 2ω; (b) induced voltage across a pair of electrodes on a GaAs semiconductor as a function of the phase parameter associated with two harmonically related incident beams.

The current injection via coherence control and conventional cases (i.e., under a DC bias) differs with respect to their evolution. The current injected via coherent control has an onset determined by the rise time of the optical pulses. In the case of normal current production, existing carriers are accelerated by an electric field, and the momentum distribution is never far from isotropic. For a carrier density of 10^{14} cm^{-3} a DC field $\sim$80 kV cm^{-1} is required to produce a current density of 1 kA cm^{-2}. In GaAs, with an electron mobility of 8000 cm^2 V^{-1} s^{-1} this current would occur about 1/2 ps after the field is 'instantaneously' turned on. This illustrates that the coherently controlled phenomenon efficiently and quickly produces a larger current than can be achieved with the redirecting of statistically distributed electrons.

A more detailed analysis must take into account the actual light polarization, crystal symmetry, crystal face, and orientation relative to the optical polarization. For given optical intensities of the two beams,

the maximum electrical current injection in the case of GaAs occurs for linearly polarized beams both oriented along the (111) or equivalent direction. However, large currents can also be observed with the light beams polarized along other high symmetry directions.

Coherent Control of Carrier Density, Spin Population, and Spin Current Using Two Color Beams

The processes described above do not exhaust the coherent control effects that can be observed in bulk semiconductors using harmonic beams. Indeed, for noncentrosymmetric materials, certain light polarizations and crystal orientations allow one to coherently control the total carrier generation rate with or without generating an electrical current. In the case of the cubic material GaAs, provided the ω beam has electric field components along two of the three principal crystal axes, with the 2ω beam having a component along the third direction, one can coherently control the total carrier density. However, the overall degree of control here is determined by how 'noncentrosymmetric' the material is.

The spin degrees of freedom of a semiconductor can also be controlled using two color beams. Due to the spin–orbit interaction, the upper valence bands of a typical semiconductor have certain spin characteristics. To date, optical manipulation of electron spin has been largely based on the fact that partially spin-polarized carriers can be injected in a semiconductor via one-photon absorption of circularly polarized light from these upper valence bands. In such carrier injection – where in fact two-photon absorption could be used as well – spins with no net velocity are injected, and then are typically dragged by a bias voltage to produce a spin-polarized current. However, given the protocols discussed above it should not come as a surprise that the two color coherence scheme, when used with certain light polarizations, can coherently control the spin polarization, making it dependent on $\Delta\phi$. Furthermore, for certain polarization combinations it is also possible to generate a spin current with or without an electrical current. Given the excitement surrounding the field of spintronics, where the goal is the use of the spin degree of freedom for data storage and processing, the control of quantities involving the intrinsic angular momentum of the electron is of particular interest.

Various polarization and crystal geometries can be examined for generating spin currents with or without electrical current. For example, with reference to Figure 3a, for both beams propagating in the **z** (*normal*) direction of a crystal and possessing the same circular polarization, an electrical current can be injected in the (xy) plane, at an angle from the crystallographic **x** direction dependent on the relative phase parameter, $\Delta\phi$; this current is spin-polarized in the **z** direction. As well, the injected carriers have a $\pm$**z** component of their velocity, as many with one component as with the other; but those going in one direction are preferentially spin-polarized in one direction in the (xy) plane, while those in the other direction are preferentially spin-polarized in the opposite direction. This is an example of a spin current in the absence of an electrical current.

Such a pure spin current that is perhaps more striking is observable with the two beams cross linearly polarized, for example with the fundamental beam in the **x** direction and the second harmonic beam in the **y** direction as shown in Figure 3b. Then there is no net spin injection; the average spin in any

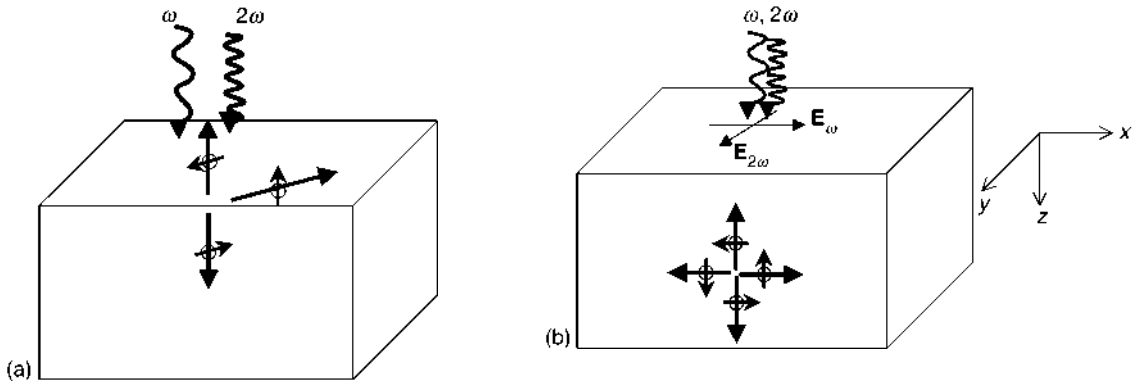

Figure 3 (a) Excitation of a semiconductor by co-circularly polarized ω and 2ω pulses. Arrows indicate that a spin polarized electrical current is generated in the $x-y$ plane in a direction dependent on $\Delta\phi$ while a pure spin current is generated in the beam propagation (z) direction. (b) Excitation of a semiconductor by orthogonally, linearly polarized ω and 2ω pulses. Arrows indicate that a spin polarized electrical current is generated in the direction of the fundamental beam polarization as well as along the beam propagation (z) direction.

direction is zero. And for a vanishing phase parameter $\Delta\phi$ there is no net electrical current injection. Yet, for example, the electrons injected with a $+\mathbf{x}$ velocity component will have one spin polarization with respect to the $\mathbf{z}$ direction, while those injected with a $-\mathbf{x}$ component to their velocity will have the opposite spin polarization with respect to the $\mathbf{z}$ direction.

The examples given above are the spin current analog of the two-color electrical current injection discussed above. There should also be the possibility of injecting a spin current with a single beam into crystals lacking center-of-inversion symmetry.

The simple analysis presented above relies on simple quantum mechanical ideas and calculations using essentially nothing more than Fermi's Golden Rule. Yet the process involves a mixing of two frequencies, and can therefore be thought of as a nonlinear optical effect. Indeed, one of the key ideas to emerge from the theoretical study of such phenomena is that an interpretation of coherence control effects alternate to that provided by the simple quantum interference picture that is provided by the usual susceptibilities of nonlinear optics. These susceptibilities are, of course, based on quantum mechanics, but the macroscopic viewpoint allows for the identification and classification of the effects in terms of 2nd order nonlinear optical effects, 3rd order optical effects, etc. Indeed, one can generalize many of the processes we have discussed above to a hierarchy of frequency mixing effects or high-order nonlinear processes involving multiple beams with frequency $n\omega, m\omega, p\omega\ldots$ with n, m, p being integers. Many of theses schemes require high intensity of one or more of the beams, but can occur in a simple semiconductor.

Coherent Control of Electrical Current Using Single Color Beams

It is also possible to generate coherence control effects using beams at a single frequency (ω), if one focuses on the two orthogonal components (e.g., x and y) polarization states and uses a noncentrosymmetric semiconductor of reduced symmetry such as a strained cubic semiconductor or a wurtzite material such as CdS or CdSe. In this case, interference between absorption pathways associated with the orthogonal components can lead to electrical current injection given by

$$dJ/dt \propto E^{\omega}E^{\omega}\sin(\phi_{\omega}^{x} - \phi_{\omega}^{y}) \qquad [8]$$

Since in this process current injection is linear in the beam's intensity, the high intensities necessary for two-photon absorption are not necessary. Nonetheless, the efficacy of this process is limited by the fact that it relies on the breaking of center-of-inversion symmetry of the underlying crystal. It is also clear that the maximum current occurs for circularly polarized light and that right and left circularly polarized light lead to a difference in sign of the current injection. Finally, for this particular single beam scheme, when circularly polarized light is used the electrical current is partially spin polarized.

Conclusions

Through the phase of optical pulses it is possible to control electrical and spin currents, as well as carrier density, in bulk semiconductors on a time-scale that is limited only by the rise time of the optical pulse and the intrinsic response of the semiconductor. A few optically based processes that allow one to achieve these types of control have been illustrated here. Although at one level one can understand these processes in terms of quantum mechanical interference effects, at a macroscopic level one can understand these control phenomena as manifestations of nonlinear optical phenomena. For fundamental as well as applied reasons, our discussion has focused on coherence control effects using the continuum states in bulk semiconductors, although related effects can also occur in quantum dot, quantum well, and superlattice semiconductors. Applications of these control effects will undoubtedly exploit the all-optical nature of the process, including the speed at which the effects can be turned on or off. The turn-off effects, although not discussed extensively here, are related to transport phenomena as well as momentum scattering and related dephasing processes.

See also

Semiconductor Physics: Spin Transport and Relaxation in Semiconductors; Spintronics. **Interferometry**: Overview. **Coherent Control:** Experimental; Theory.

Further Reading

Brumer P (1995) Laser control of chemical reactions. *Scientific American* 272: 56.

Rabitz H, de Vivie-Riedle R, Motzkus M and Kompa K (2000) Whither the future of controlling quantum phenomena? *Science* 288: 824.

Shah J (1999) *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*. New York: Springer.

van Driel HM and Sipe JE (2001) Coherent control of photocurrents in semiconductors. In: Tsen T (ed.) *Ultrafast Phenomena in Semiconductors*, p. 261. New York: Springer Verlag.

COHERENT LIGHTWAVE SYSTEMS

M J Connelly, University of Limerick, Limerick, Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Conventional optical communication systems are based on the intensity modulation (IM) of an optical carrier by an electrical data signal and direct detection (DD) of the received light. The simplest scheme employs on/off keying (OOK), whereby turning on or off, an optical source transmits a binary 1 or 0. The signal light is transmitted down an optical fiber and at the receiver is detected by a photodetector, as shown in Figure 1. The resulting photocurrent is then processed to determine if a 1 or 0 was received. In the ideal case, where a monochromatic light source is used and where no receiver noise (detector dark current and thermal noise) is present, the probability of error is determined by the quantum nature (shot noise) of the received light. In this case, the number of photons/bit required (sensitivity) at the receiver, to achieve a bit-error-rate (BER) of 10^{-9}, is 10. This is called the quantum-limit. In practical receivers without an optical preamplifier, operating in the 1.3 μm and 1.55 μm optical fiber communication windows, the actual sensitivity is typically 10–30 dB less than the theoretical sensitivity. This is due to receiver noise, in particular thermal noise. This means that the actual sensitivity is between 100 to 10 000 photons/bit. The capacity of IM/DD fiber links can be increased using wavelength division multiplexing (WDM). At the receiver the desired channel is selected using a narrowband optical filter. Some advanced commercial IM/DD systems utilize dense WDM with 50 GHz spacing and single-channel bit rates as high as 40 Gb/s. Indeed, IM/DD systems with terabit (10^{12} bit/s) capacity are now possible.

Compared to IM/DD systems, coherent optical communication systems have greater sensitivity and selectivity. In the context of coherent lightwave systems, the term coherent refers to any technique employing nonlinear mixing between two optical waves on a photodetector, as shown in Figure 2. Typically, one of these is an information-bearing signal and the other is a locally generated wave (local oscillator). The result of this heterodyne process is a modulation of the photodetector photocurrent at a frequency equal to the difference between the signal and local oscillator frequencies. This intermediate frequency (IF) electronic signal contains the information in the form of amplitude, frequency, or phase that was present in the original optical signal. The IF signal is filtered and demodulated to retrieve the transmitted information. An automatic frequency control circuit is required to keep the local-oscillator frequency stable.

It is possible for the information signal to contain a number of subcarriers (typically at microwave frequencies), each of which can be modulated by a separate data channel. It is a simple matter to select the desired channel at the receiver by employing electronic heterodyning and low pass filtering, as shown in Figure 3. In IM/DD systems, channel selection can only be carried out using narrowband optical filters.

Coherent optical communications utilize techniques that were first investigated in radio communications. Most of the basic research work on coherent optical communications was carried out in the 1980s and early 1990s, and was primarily motivated by the need for longer fiber links using no repeaters. Improving receiver sensitivity using coherent techniques made it possible to increase fiber link spans. In the mid-1990s, reliable optical fiber amplifiers became available. This made it possible to construct fiber links using optical amplifiers spaced at appropriate intervals to compensate for fiber and other transmission losses.

The sensitivity of an IM/DD receiver can be greatly improved by the use of an optical preamplifier. Indeed the improvement is so marked that coherent techniques are not a viable alternative to IM/DD in the vast majority of commercial optical communication systems due to their complexity and higher cost. Moreover, because semiconductor lasers with very precise wavelengths and narrowband optical filters can be fabricated, the selectivity advantage of coherent optical communications has become less important. This has meant that the research work currently being carried out on coherent optical communications is very limited, and in particular, very few field trials have been carried out since the early 1990s.

We will review the basic principles underlying coherent optical communications, modulation schemes, detection schemes, and coherent receiver sensitivity. The effects of phase noise and polarization on coherent receiver performance will be outlined. A brief summary of pertinent experimental results will also be presented.

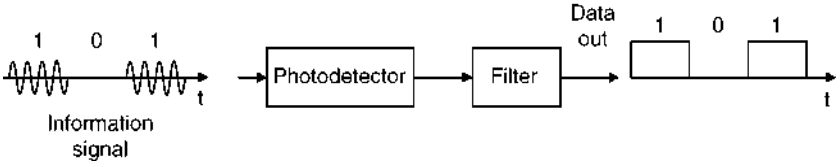

Figure 1 IM/DD optical receiver.

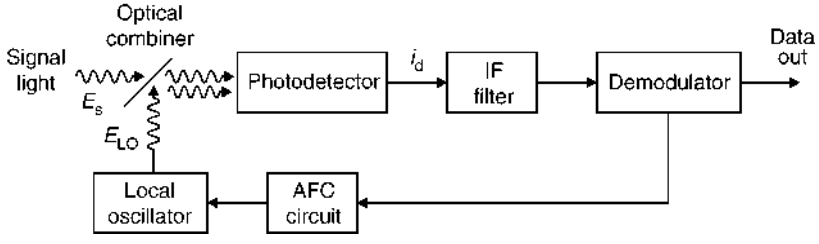

Figure 2 Basic coherent optical receiver. AFC: automatic frequency control.

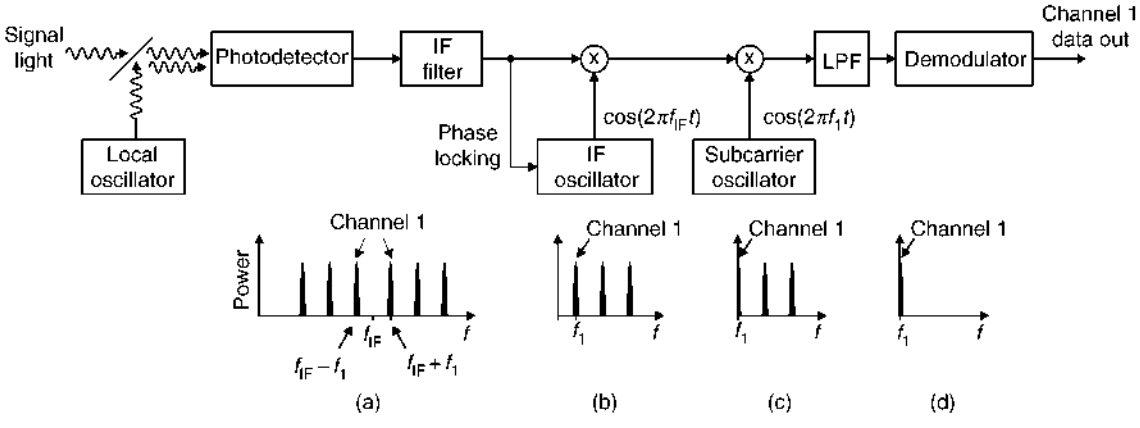

Figure 3 Coherent detection and processing of a subcarrier modulated signal lightwave. The power spectrum of the detected signal is shown: (a) after IF filtering; (b) after multiplication by the IF oscillator; (c) after multiplication with the subcarrier oscillator (in this case channel 1's carrier frequency); and (d) after low pass filtering prior to demodulation and data detection. LPF: lowpass filter.

Basic Principles

In coherent detection, a low-level optical signal field E_s is combined with a much larger power optical signal field E_{LO} from a local oscillator laser. E_s and E_{LO} can be written as

$$E_s(t) = \sqrt{2P_s}\cos[2\pi f_s t + \phi_s(t)] \qquad [1]$$

$$E_{LO}(t) = \sqrt{2P_{LO}}\cos[2\pi f_{LO} t + \phi_{LO}(t)] \qquad [2]$$

where P_s, f_s, and ϕ_s are the signal power, optical frequency, and phase (including phase noise), respectively. P_{LO}, f_{LO}, and ϕ_{LO} are the equivalent quantities for the local oscillator. For ideal coherent detection, the polarization states of the signal and local oscillator must be equal. If this is the case, the photocurrent is given by

$$i_d(t) = R[E_s(t) + E_{LO}(t)]^2 \qquad [3]$$

where R is the detector responsivity. Because the detector cannot respond to optical frequencies, we have

$$i_d(t) = R\left\{P_s + P_{LO} + 2\sqrt{P_s P_{LO}} \times \cos[2\pi f_{IF} t + \phi_s(t) - \phi_{LO}(t)]\right\} \qquad [4]$$

The IF signal is centered at frequency $f_{IF} = f_s - f_{LO}$. In homodyne detection $f_s = f_{LO}$ and the homodyne signal is centered at baseband. In heterodyne detection $f_s \neq f_{LO}$ and the signal generated by the photodetector is centered around f_{IF} (typically three to six times the bit rate). The IF photocurrent is proportional to $\sqrt{P_s}$, rather than P_s as is the case in

direct detection, and is effectively amplified by a factor proportional to $\sqrt{P_{LO}}$. If P_{LO} is large enough, the signal power can be raised above the receiver noise, thereby leading to a greater sensitivity than possible with conventional IM/DD receivers using p–i–n or avalanche photodiodes. Theoretically it is possible to reach shot noise limited receiver sensitivity although this is not possible in practice. Homodyne optical receivers usually require less bandwidth and are more sensitive than heterodyne receivers but require that the local oscillator is phase locked to the information signal. This requires an optical phase locked loop (OPLL), which is very difficult to design.

Modulation Schemes

The three principal modulation schemes used in coherent optical communications; amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK), are shown in Figure 4. ASK modulation is essentially the same as the OOK scheme used in IM/DD systems. The optical signal is given by

$$E_s(t) = 2a(t)\sqrt{2P_s}\cos[2\pi f_s t + \phi_s(t)] \qquad [5]$$

where $a(t) = 0$ or 1 for transmission of a 0 or 1, respectively. P_s is the average signal power, assuming that it is equally likely that a 0 or 1 is transmitted. ASK modulation can be achieved by modulating the output of a semiconductor laser using an external modulator. The most common type is based on $LiNbO_3$ waveguides in a Mach–Zehnder interferometer configuration.

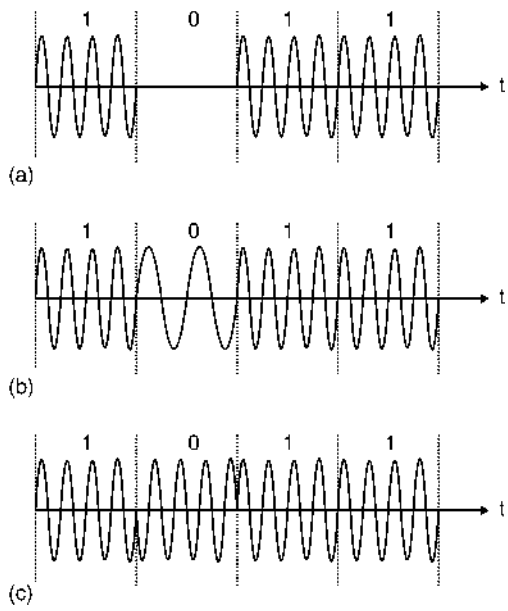

Figure 4 Modulation schemes for coherent optical communications: (a) ASK; (b) FSK; and (c) PSK.

In FSK modulation, the optical frequency of the signal lightwave is changed slightly by the digital data stream. In binary FSK, the optical signal is given by

$$E_s(t) = \sqrt{2P_s}\cos[2\pi(f_s + a(t)\Delta f)t + \phi_s(t)] \qquad [6]$$

where $a(t) = -1$ or 1 for transmission of a 0 or 1, respectively. The choice of the frequency deviation Δf depends on the available bandwidth, the bit rate B_T and the demodulation scheme used in the receiver. The modulation index m is given by

$$m = \frac{2\Delta f}{B_T} \qquad [7]$$

The minimum value of m that produces orthogonality (for independent detection of the 0s and 1s) is 0.5. If $m = 0.5$, the signal is referred to as minimum FSK (MFSK). When $m \geq 2$, the signal is referred to as wide-deviation FSK, and when $m < 2$, as narrow-deviation FSK. Wide-deviation FSK receivers are more resilient to phase noise than narrow-deviation FSK receivers. FSK modulation can be achieved using $LiNbO_3$ external modulators, acoustic optic modulators and distributed feedback semiconductor lasers. The phase of the lightwave does not change between bit transitions.

In PSK modulation, the digital data stream changes the phase of the signal lightwave. In binary PSK the optical signal is given by

$$E_s(t) = \sqrt{2P_s}\cos[2\pi f_s t + a(t)\theta + \frac{\pi}{2} + \phi_s(t)] \qquad [8]$$

where $a(t) = -1$ or 1 for transmission of a 0 or 1, respectively. The phase shift between a 0 and 1 is $\theta + \pi/2$. Most PSK schemes use $\theta = \pi/2$. PSK modulation can be achieved by using an external phase modulator or a multiple quantum well electro-absorption modulator.

Detection Schemes

There are two classes of demodulation schemes: synchronous and asynchronous. The former exploits the frequency and phase of the carrier signal to perform the detection. The latter only uses the envelope of the carrier to perform the detection. In the following we review the principal demodulation schemes.

PSK Homodyne Detection

In an ideal PSK homodyne receiver, shown in Figure 5, the signal light and local oscillator have identical

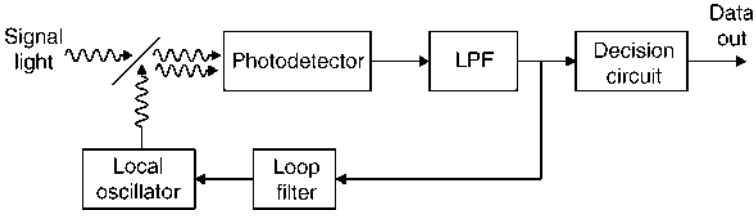

Figure 5 PSK homodyne receiver.

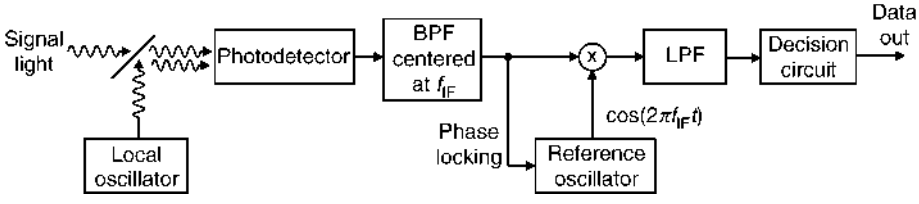

Figure 6 PSK synchronous heterodyne receiver. BPF: bandpass filter.

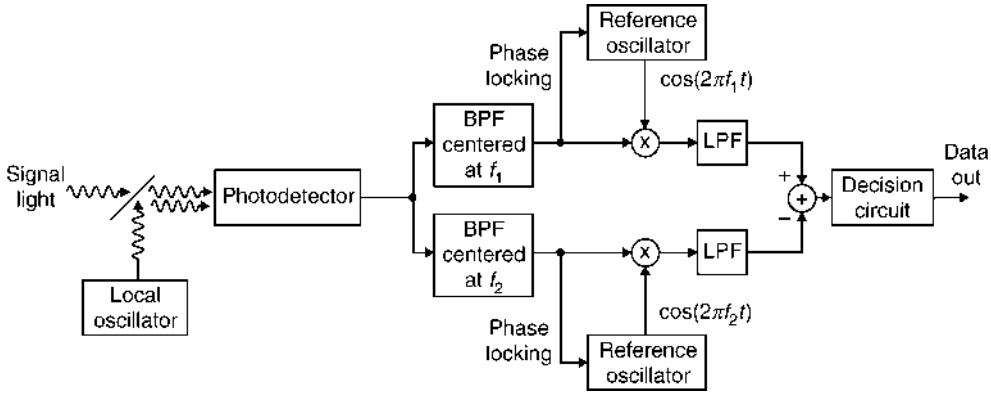

Figure 7 FSK heterodyne synchronous receiver.

optical frequency and phase. If the resulting baseband information signal is positive, then a 1 has been received. If the baseband information signal is negative, then a 0 bit has been received. This scheme has the highest theoretical sensitivity but requires an OPLL and is also highly sensitive to phase noise.

ASK Homodyne Detection

An ideal ASK homodyne receiver is identical to an ideal PSK homodyne receiver. Because the baseband information signal is either zero (0 bit) or nonzero (1 bit), the receiver sensitivity is 3 dB less than the PSK homodyne receiver.

PSK Heterodyne Synchronous Detection

Figure 6 shows an ideal PSK synchronous heterodyne receiver. The IF signal is filtered by a bandpass filter, multiplied by the phase locked reference oscillator to move the signal to baseband, low pass filtered and sent to a decision circuit that decides if the received bit is a 0 or 1. The receiver requires synchronization of the reference oscillator with the IF signal. This is not easy to achieve in practice because of the large amount of semiconductor laser phase noise.

ASK Heterodyne Synchronous Detection

This scheme is basically the same as the PSK heterodyne detection and is similarly difficult to design.

FSK Heterodyne Synchronous Detection

This scheme is shown in Figure 7 for a binary FSK signal consisting of two possible frequencies $f_1 = f_s - \Delta f$ and $f_2 = f_s + \Delta f$. There are two separate branches in which correlation with the two possible IF signals at f_1 and f_2 is performed. The two resulting signals are subtracted from each other and a decision taken on the sign of the difference. Because of the requirement for two electrical phase locked loops, the scheme is not very practical.

ASK Heterodyne Envelope Detection

In this scheme, shown in Figure 8, the IF signal is envelope detected to produce a baseband signal proportional to the original data signal. The output of the envelope detector is relatively insensitive to laser phase noise.

FSK Heterodyne Dual-Filter Detection

In this scheme, shown in Figure 9, two envelope detectors are used to demodulate the two possible IF signals at f_1 and f_2 and a decision taken based on the outputs. The scheme is tolerant to phase noise and can achieve high sensitivity. However, it needs high-bandwidth receiver electronics because a large frequency deviation is required.

FSK Heterodyne Single-Filter Detection

This scheme omits one branch of the FSK heterodyne dual-filter receiver and performs a decision in the same way as for the ASK heterodyne envelope detection. Because half the power of the detected signal is not used, the sensitivity of the scheme is 3 dB worse than the dual-filter scheme.

CPFSK Heterodyne Differential Detection

This scheme, shown in Figure 10, uses the continuous phase (CP) characteristic of an FSK signal, i.e., there is no phase discontinuity at bit transition times. After bandpass filtering, the IF photocurrent is

$$i_{\mathrm{IF}}(t) = A\cos\{2\pi[(f_{\mathrm{IF}} + a(t)\Delta f)]t\} \qquad [9]$$

where $A = R\sqrt{P_s P_{\mathrm{LO}}}$. The output from the delay-line demodulator, after low pass filtering to remove the double IF frequency term, is

$$x(t) = \frac{A^2}{2}\cos\{2\pi[f_{\mathrm{IF}} + a(t)\Delta f]\tau\} \qquad [10]$$

The function of the delay-line demodulator is to act as a frequency discriminator. We require that $x(t)$ is a maximum when $a(t) = 1$ and a minimum when $a(t) = -1$. This is the case if the following relations are satisfied:

$$2\pi(f_{\mathrm{IF}} + \Delta f)\tau = 2\pi k \qquad [11]$$

$$2\pi(f_{\mathrm{IF}} - \Delta f)\tau = (2k-1)\pi, \qquad k \text{ an integer}$$

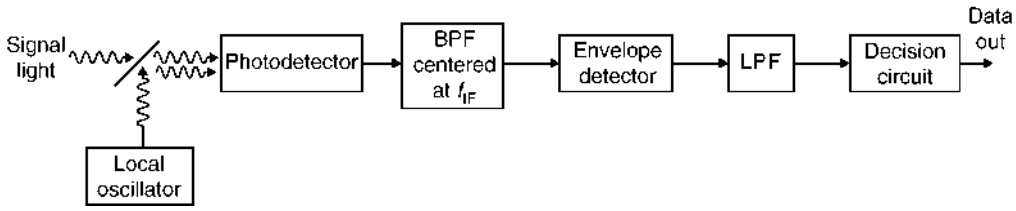

Figure 8 ASK heterodyne envelope receiver.

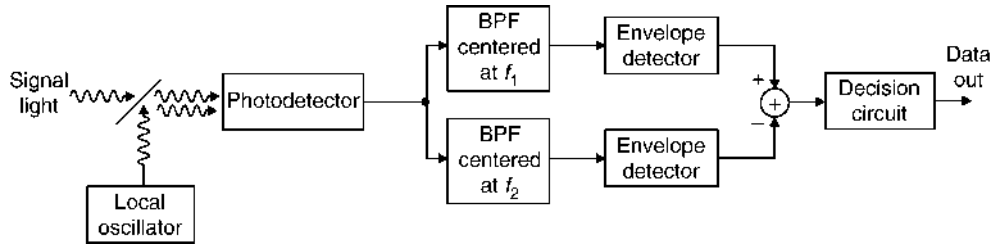

Figure 9 FSK heterodyne dual-filter receiver.

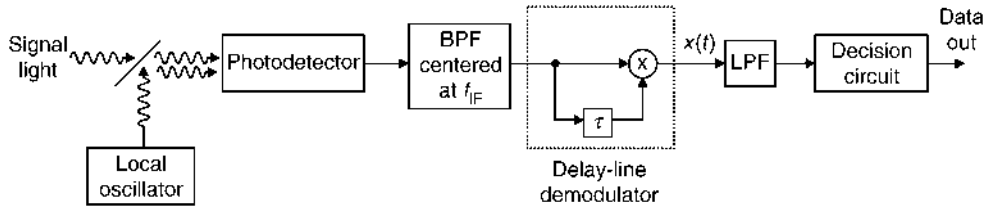

Figure 10 CPFSK heterodyne differential detection receiver.

Hence, we require that

$$\Delta f \tau = \frac{1}{4} \qquad [12]$$

$$f_{\mathrm{IF}} \tau = \left[k - \frac{1}{4} \right], \qquad k \text{ integer}$$

In terms of the modulation index we have

$$\tau = \frac{1}{2mB_{\mathrm{T}}} \qquad [13]$$

The above equation shows that it is possible to decrease the value of τ required by increasing m, thereby reducing the sensitivity of the receiver to phase noise. The minimum value of m possible is 0.5. This scheme can operate with a smaller modulation index compared to the dual-filter and single-filter schemes for a given bit rate, thereby relaxing the bandwidth requirements of the receiver electronics.

DPSK Heterodyne Differential Detection

In differential phase shift key (DPSK) modulation, a 0 is sent by changing the phase of the carrier by π with respect to the previous signal. A 1 is sent by not changing the phase. The receiver configuration is identical to the CPFSK heterodyne differential detection receiver, except that the time delay τ is equal to the inverse of the data rate. Compared to other PSK detection schemes, DPSK differential detection is relatively simple to implement and is relatively immune to phase noise.

Coherent Receiver Sensitivity

The signal-to-noise ratio γ_{c} of an ideal PSK homodyne receiver, in the shot-noise limit, is given by

$$\gamma_{\mathrm{c}} = \frac{2RP_{\mathrm{s}}}{eB_{\mathrm{c}}} \qquad [14]$$

where B_{c} is the receiver bandwidth. The detector responsivity is given by

$$R = \frac{\eta e}{hf_{\mathrm{s}}} \qquad [15]$$

where η is the detector quantum efficiency (≤ 1). The average signal power $P_{\mathrm{s}} = N_{\mathrm{s}} h f_{\mathrm{s}} B_{\mathrm{T}}$, where N_{s} is the average number of photons per bit. If it is assumed that $B_{\mathrm{c}} = B_{\mathrm{T}}$, then

$$\gamma_{\mathrm{c}} = 2\eta N_{\mathrm{s}} \qquad [16]$$

The actual detected number of photons per bit $N_{\mathrm{R}} = \eta N_{\mathrm{s}}$, so

$$\gamma_{\mathrm{c}} = 2N_{\mathrm{R}} \qquad [17]$$

If the noise is assumed to have a Gaussian distribution, then the bit error rate (BER) is given by

$$\mathrm{BER} = \frac{1}{2} erfc\left(\sqrt{2N_{\mathrm{R}}}\right) \qquad [18]$$

where *erfc* is the complementary error function; $erfc(x) \approx e^{-x^2}/(x\sqrt{\pi})$ for $x > 5$. The receiver sensitivity for a BER $= 10^{-9}$ is 9 photons/bit.

BER expressions and sensitivity for the demodulation schemes discussed above are listed in Table 1. The sensitivity versus N_{R} is plotted for some of the demodulation schemes in Figure 11. The received signal power required for a given BER is proportional to B_{T}. This dependency is shown in Figure 12 for some of the demodulation schemes.

Table 1 Shot-noise limited BER expressions and sensitivity for coherent demodulation schemes

Modulation scheme	*BER*	*Sensitivity (photons/bit)*
Homodyne		
PSK	$\frac{1}{2} erfc\left(\sqrt{2N_{\mathrm{R}}}\right)$	9
ASK	$\frac{1}{2} erfc\left(\sqrt{N_{\mathrm{R}}}\right)$	18 (peak 36)
Synchronous heterodyne		
PSK	$\frac{1}{2} erfc\left(\sqrt{N_{\mathrm{R}}}\right)$	18
ASK	$\frac{1}{2} erfc\left(\sqrt{\frac{N_{\mathrm{R}}}{2}}\right)$	36 (peak 72)
FSK	$\frac{1}{2} erfc\left(\sqrt{\frac{N_{\mathrm{R}}}{2}}\right)$	36
Asynchronous heterodyne		
ASK (envelope)	$\frac{1}{2}\exp\left(-\frac{N_{\mathrm{R}}}{2}\right)$	40 (peak 80)
FSK (dual-filter)	$\frac{1}{2}\exp\left(-\frac{N_{\mathrm{R}}}{2}\right)$	40
FSK (single-filter)	$\frac{1}{2}\exp\left(-\frac{N_{\mathrm{R}}}{4}\right)$	80
Differential detection		
CPFSK	$\frac{1}{2}\exp(-N_{\mathrm{R}})$	20
DPSK	$\frac{1}{2}\exp(-N_{\mathrm{R}})$	20

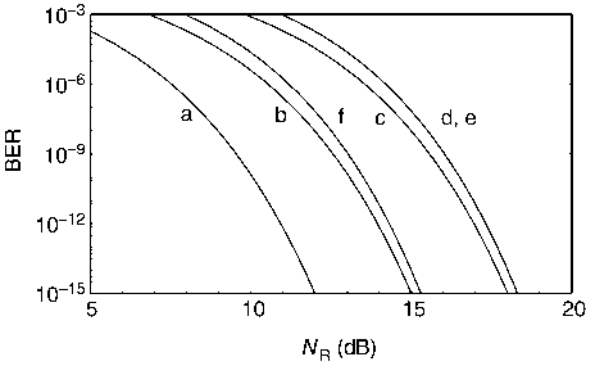

Figure 11 Shot-noise limited BER versus number of received photons/bit for various demodulation schemes: (a) PSK homodyne; (b) synchronous PSK; (c) synchronous FSK; (d) ASK envelope; (e) FSK dual-filter; and (f) CPFSK and DPSK.

Phase Noise

A major influence on the sensitivity possible with practical coherent receivers is laser phase noise, which is responsible for laser linewidth (3 dB bandwidth of the laser power spectrum). The effect of phase noise on system performance depends on the modulation and demodulation scheme used. In general, the effects of phase noise are more severe in homodyne and synchronous heterodyne schemes than in asynchronous schemes. If $\Delta\nu_{IF}$ is the beat linewidth (without modulation present) between the signal and local oscillator lasers, the normalized beat linewidth $\Delta\nu$ is defined as

$$\Delta\nu = \frac{\Delta\nu_{IF}}{B_T} \qquad [19]$$

where $\Delta\nu_{IF}$ is the IF phase noise bandwidth. Semiconductor lasers usually have a Lorentzian spectrum in which case $\Delta\nu_{IF}$ is equal to the sum of the signal and local-oscillator laser linewidths. The requirements placed on $\Delta\nu$ become less severe as the bit rate increases. Table 2 compares the $\Delta\nu$ requirement for a 1 dB power penalty at a BER of 10^{-9} for some of the principal modulation/demodulation schemes.

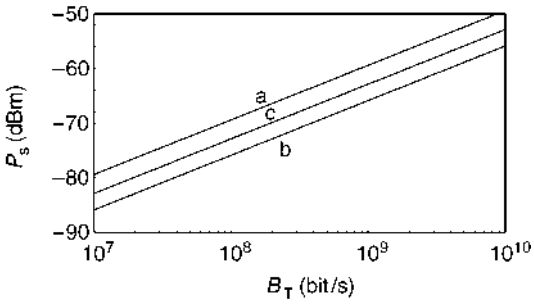

Figure 12 Required signal power versus bit rate for a BER = 10^{-9} for various shot-noise limited demodulation schemes: (a) PSK homodyne; (b) ASK envelope; and (c) CPFSK. The signal wavelength is 1550 nm and the detector is assumed to have unity quantum efficiency.

Table 2 Required normalized beat linewidth at 1 dB power penalty for a BER = 10^{-9} for various modulation/demodulation schemes

	Modulation scheme	*$\Delta\nu$ (%)*
Homodyne	PSK	6×10^{-4} (balanced loop) 0.01 (decision-driven loop)
Synchronous heterodyne	PSK	0.2–0.5
Asynchronous heterodyne	PSK	0.3–0.7
	ASK (envelope)	3
	FSK (dual-filter)	10–20
Differential detection	CPFSK	0.33 ($m = 0.5$) 0.66 ($m = 1.0$)
	DPSK	0.33

PSK Homodyne Detection – Phase Locked Loop Schemes

Practical homodyne receivers require an OPLL to lock the local-oscillator frequency and phase to that of the signal lightwave. There are a number of OPLL schemes possible for PSK homodyne detection, including the balanced PLL, Costas-type PLL and the decision-driven PLL.

The balanced PLL scheme, shown in Figure 13, requires that the PSK signal uses a phase shift of less than π between a 0 and 1, $E_s(t)$ and $E_{LO}(t)$ can be written as

$$E_s(t) = \sqrt{2P_s}\cos\left[2\pi f_0 t + a(t)\theta + \phi_{N,s}(t) + \frac{\pi}{2}\right] \qquad [20]$$

$$E_{LO}(t) = \sqrt{2P_{LO}}\cos[2\pi f_0 t + \phi_{N,LO}(t)] \qquad [21]$$

where f_0 is the common signal and local oscillator optical frequency and $\phi_{N,s}(t)$ and $\phi_{N,LO}(t)$ are the signal and local oscillator phase noise, respectively. When $\theta \neq \pi/2$, $E_s(t)$ contains an unmodulated carrier component (residual carrier) in quadrature with the information-bearing signal. $E_s(t)$ and $E_{LO}(t)$ are combined by an optical 180° hybrid. The output optical signals from the π-hybrid are given by

$$E_1(t) = \frac{1}{\sqrt{2}}[E_s(t) + E_{LO}(t)] \qquad [22]$$

$$E_2(t) = \frac{1}{\sqrt{2}}[E_s(t) - E_{LO}(t)] \qquad [23]$$

The voltage at the output of the summing circuit is given by

$$v_L(t) = 2RR_L\sqrt{P_s P_{LO}}\{a(t)\sin\theta\cos[\phi_e(t)] + \cos\theta\sin[\phi_e(t)]\} \qquad [24]$$

where the photodetectors have the same responsivity and load resistance R_L. The phase error is given by

$$\phi_e(t) = \phi_{N,s}(t) - \phi_{N,LO}(t) - \phi_c(t) \qquad [25]$$

$\phi_c(t)$ is the controlled phase determined by the control signal $v_c(t)$ at the output of the loop filter. v_L contains two terms; an information-bearing signal proportional to $\sin\theta$, which is processed by the data detection circuit, and a phase error signal proportional to $\cos\theta$ used by the PLL for locking. The power penalty, due to the residual carrier transmission, is $10\log_{10}(1/\sin^2\theta)$ dB.

This scheme is not practical because the $\Delta\nu$ needed for a power penalty of <1 dB is typically less than 10^{-5}. This would require that $\Delta\nu_{IF} < 10$ kHz for

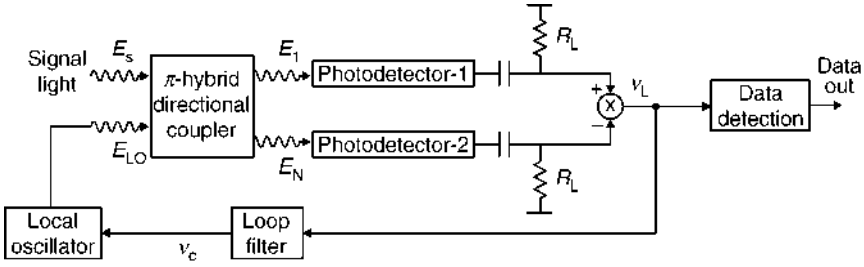

Figure 13 PSK homodyne balanced PLL receiver.

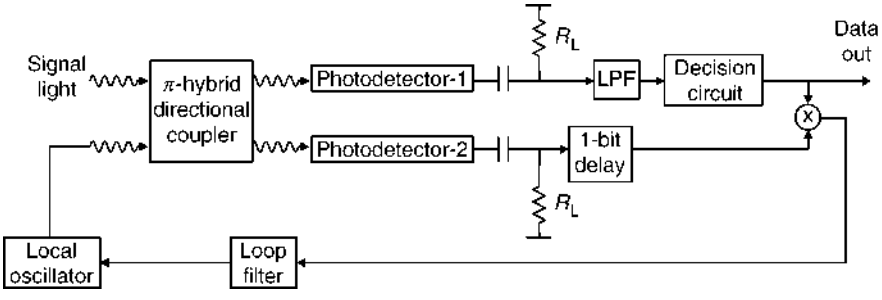

Figure 14 PSK decision-driven PLL.

a data rate of 1 Gb/s. The narrow signal and local-oscillator linewidths required are only achievable with external cavity semiconductor lasers.

The Costas-type and decision-driven PLLs also utilize π-hybrids and two photodetectors but as they have full suppression of the carrier ($\theta = \pi/2$), nonlinear processing of the detected signals is required to produce a phase error signal that can be used by the PLL for locking. In the decision-driven PLL, shown in Figure 14, the signal from one of the photodetector outputs is sent to a decision circuit and the output of the circuit is multiplied by the signal from the second photodetector. The mixer output is sent back to the local oscillator laser for phase locking. The $\Delta\nu$ required for a power penalty of <1 dB is typically less than 2×10^{-4}. This is superior to both the balanced PLL and Costas-type PLL receivers.

Heterodyne Phase Locking

In practical synchronous heterodyne receivers phase locking is performed in the electronic domain utilizing techniques from radio engineering. In the PSK case, electronic analogs of the OPLL schemes used in homodyne detection can be used. Synchronous heterodyne schemes have better immunity to phase noise than the homodyne schemes.

Asynchronous Systems

Asynchronous receivers do not require optical or electronic phase locking. This means that they are less sensitive to phase noise than synchronous receivers. They do have some sensitivity to phase noise because IF filtering leads to phase-to-amplitude noise conversion.

Differential Detection

Differential detection receivers have sensitivities to phase noise between synchronous and asynchronous receivers because, while they do not require phase locking, they use phase information in the received signal.

Phase-Diversity Receivers

Asynchronous heterodyne receivers are relatively insensitive to phase noise but require a much higher receiver bandwidth for a given bit rate. Homodyne receivers only require a receiver bandwidth equal to the detected signal bandwidth but require an OPLL, which is difficult to implement. The phase diversity receiver, shown in Figure 15, is an asynchronous homodyne scheme, which does not require the use of an OPLL and has a bandwidth approximately equal to synchronous homodyne detection. This is at the expense of increased receiver complexity and reduced sensitivity compared to synchronous homodyne detection. The phase-diversity scheme can be used with ASK, DPSK, and CPFSK modulation. ASK modulation uses an squarer circuit for the demodulator component, while DPSK and CPFSK modulation use a delay line and mixer.

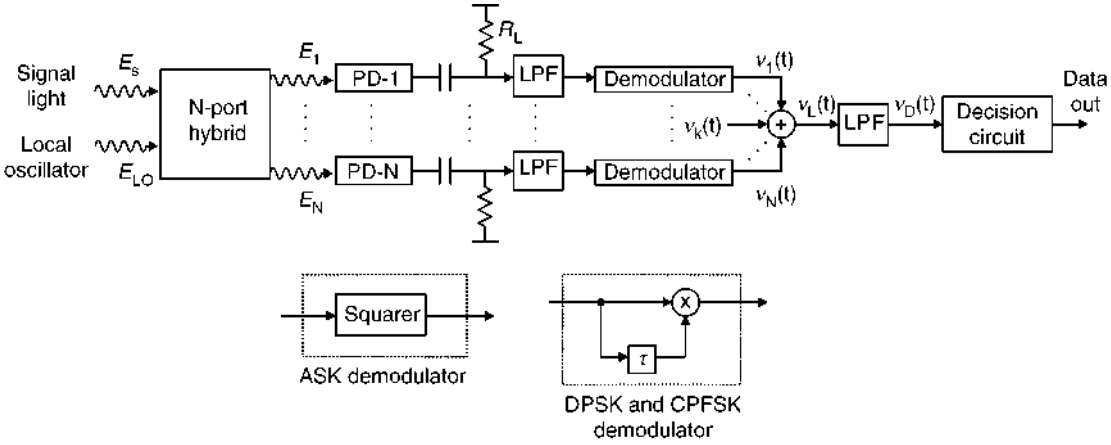

Figure 15 Multiport phase-diversity receiver. PD: photodetector.

In the case of ASK modulation, the outputs of the N-port hybrid can be written as

$$E_k(t) = \frac{1}{\sqrt{N}}\left\{a(t)\sqrt{2P_s}\cos\left[2\pi f_s t + \phi_s(t) + \frac{2\pi k}{N}\right] + \sqrt{2P_{LO}}\cos[2\pi f_{LO}t + \phi_{LO}(t)]\right\}, \quad k = 1,2,\ldots,N,\ N \geq 3 \qquad [26]$$

For $N = 2$, the phase term is $\pi/2$ for $k = 2$. The voltage input to the k-th demodulator (squarer), after the dc terms have been removed by the blocking capacitor, is given by

$$v_k(t) = a(t)\frac{RR_L}{N}2\sqrt{P_s P_{LO}}\cos\left[2\pi(f_s - f_{LO}) + \phi_s(t) - \phi_{LO}(t) + \frac{2\pi k}{N}\right] \qquad [27]$$

These voltages are then squared and added to give an output voltage

$$v_L(t) = \alpha\left(\frac{RR_L}{N}\right)^2 4P_s P_{LO}a(t)\sum_{k=1}^{N}\left\{1 + \cos\left[4\pi \times \left(f_s - f_{LO}\right)t + 2\phi_s(t) - 2\phi_{LO}(t) + \frac{4\pi k}{N}\right]\right\} \qquad [28]$$

where α is the squarer parameter. The lowpass filter at the summing circuit output effectively integrates $v_L(t)$ over a bit period. If $2(f_s - f_{LO}) \ll B_T$ and the difference between the signal and local oscillator phase noise is small within a bit period, then the input voltage to the decision circuit is

$$v_D(t) = \alpha\frac{(RR_L)^2}{N}4P_s P_{LO}a(t) \qquad [29]$$

which is proportional to the original data signal. In shot-noise limited operation, the receiver sensitivity is 3 dB worse than the ASK homodyne case. The scheme is very tolerant of laser phase noise. The $\Delta\nu$ for a power penalty of < 1 dB is of the same order as for ASK heterodyne envelope detection.

Polarization

The mixing efficiency between the signal and local-oscillator lightwaves is a maximum when their polarization states are identical. In practice the polarization state of the signal arriving at the receiver is unknown and changes slowly with time. This means that the IF photocurrent can change with time and in the worst case, when the polarization states of the signal and local-oscillator are orthogonal, the IF photocurrent will be zero. There are a number of possible solutions to this problem.

1. Polarization maintaining (PM) fiber can be used in the optical link to keep the signal polarization from changing. It is then a simple matter to adjust the local-oscillator polarization to achieve optimum mixing. However, the losses associated with PM fiber are greater than for conventional single-mode (SM) fiber. In addition most installed fiber is SM fiber.
2. An active polarization controller, such as a fiber coil using bending-induced birefringence, can be used to ensure that the local-oscillator polarization tracks the signal polarization.
3. A polarization scrambler can be used to scramble the polarization of the transmitted signal at a speed greater than the bit rate. The sensitivity degradation due to the polarization scrambling is 3 dB compared to a receiver with perfect polarization matching. However, this technique is only feasible at low bit rates.
4. The most general technique is to use a polarization-diversity receiver, as shown in Figure 16. In this scheme, the signal light and local-oscillator

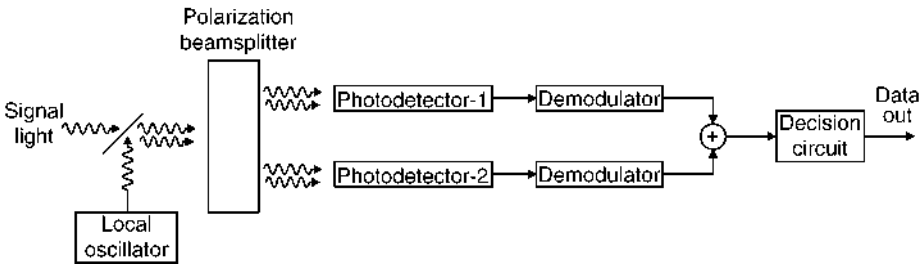

Figure 16 Polarization-diversity receiver.

Table 3 General comparisons between various modulation/demodulation schemes. The sensitivity penalty is relative to ideal homodyne PSK detection

Modulation/demodulation scheme	*Sensitivity penalty (dB)*	*Immunity to phase noise*	*IF bandwidth/bit rate*	*Complexity*
Homodyne PSK	0	Very poor	1	Requires an OPLL
Synchronous heterodyne PSK	3	Poor	3–6	Requires electronic phase locking
Asynchronous heterodyne FSK (dual-filter)	6.5	Excellent	3–6	High bandwidth IF circuits and two envelope detectors required
Differential detection CPFSK	3.5	Moderate	≈2	Relatively simple
DPSK	3.5	Moderate	≈2	Relatively simple

are split into orthogonally polarized lightwaves by a polarization beamsplitter. The orthogonal components of the signal and local-oscillator are separately demodulated and combined. The resulting decision signal is polarization independent. The demodulator used depends on the modulation format.

Comparisons Between the Principal Demodulation Schemes

General comparisons between the demodulation schemes discussed above are given in Table 3. FSK modulation is relatively simple to achieve. FSK demodulators have good receiver sensitivity and low immunity to phase noise. PSK modulation is also easy to achieve and gives good receiver sensitivity. ASK modulation has no real merit compared to FSK or PSK.

In general, homodyne schemes have the best theoretical sensitivity but require optical phase locking and are very sensitive to phase noise. Phase diversity receivers are a good alternative to homodyne receivers, in that they do not require an OPLL. Asynchronous heterodyne receivers have similar sensitivities to synchronous receivers but are much less complex. However, they require an IF bandwidth much greater than the data rate.

Differential detection schemes have good receiver sensitivity, do not require wide bandwidth IF circuits and have moderate immunity to phase noise. The advantage of CPFSK over DPSK is that the immunity to phase noise can be increased by increasing the FSK modulation index.

Polarization diversity receivers can be used with any modulation/demodulation scheme and offer the most general approach to overcoming polarization effects.

Table 4 Coherent systems experiments. The penalty is with respect to the theoretical sensitivity

Receiver	*Bit rate (Mb/s)*	*Sensitivity (dBm)*		*Penalty (dB)*
		dBm	*Photons/bit*	
ASK	400	−47.2	365	9.7
FSK (dual-filter)	680	−39.1	1600	16
CPFSK ($m = 0.5$)	1000	−37	1500	18.6
differential detection	4000	−31.3	1445	18.6
DPSK	400	−53.3	45.1	3.5
	2000	−39.0	480	13.8

System Experiments

Most coherent communication system experiments date from the 1980s and early 1990s and were based on the technology available at that time. This restricted the bit rates possible to typically <4 Gb/s. Some coherent systems experimental results obtained around this time are listed in Table 4. All of the

listed experimental trials used lasers (usually external cavity lasers) with negligible phase noise in the 1550 nm region. It can be seen that the measured sensitivities deteriorate at higher bit rates. This is attributable to the nonuniform response and bandwidth limitations of the modulator, photodetector, and demodulator. There is no doubt that the performance could be improved by the use of current optoelectronic technologies.

Conclusion

Coherent optical communication systems offer the potential for increased sensitivity and selectivity compared to conventional IM/DD systems. However, coherent optical receivers are more complex and have more stringent design parameters compared to IM/DD receivers. Moreover, advances in optoelectronic component design and fabrication along with the advent of reliable optical amplifiers has meant that coherent systems are not a viable alternative to IM/DD systems in the vast majority of optical communication systems.

List of Units and Nomenclature

e	Electronic charge
h	Planck constant

See also

Modulators: Modulation and Demodulation of Optical Signals. **Optical Amplifiers:** Basic Concepts; Erbrium Doped Fiber Amplifiers for Lightwave Systems; Optical Amplifiers in Long-Haul Transmission Systems. **Optical Communication Systems:** Architectures of Optical Fiber Communication Systems; Lightwave Transmitters; Wavelength Division Multiplexing. **Polarization:** Introduction.

Further Reading

Barry JR and Lee EA (1990) Performance of coherent optical receivers. *IEEE Proceedings* 78: 1369–1394.

Cvijetic M (1996) *Coherent and Nonlinear Lightwave Communications*. Boston, MA: Artech House.

Kazovsky LG (1989) Phase- and polarization-diversity coherent optical techniques. *Journal of Lightwave Technology* 7: 279–292.

Kazovsky L, Benedetto S and Willner A (1996) *Optical Fiber Communication Systems*. Boston, MA: Artech House.

Linke RA and Gnauck AH (1988) High-capacity coherent lightwave systems. *Journal of Lightwave Technology* 6: 1750–1769.

Ryu S (1995) *Coherent Lightwave Systems*. Boston, MA: Artech House.

COHERENT TRANSIENTS

Contents

Coherent Transient Spectroscopy in Atomic and Molecular Vapors
Foundations of Coherent Transients in Semiconductors
Ultrafast Studies of Semiconductors

Coherent Transient Spectroscopy in Atomic and Molecular Vapors

P R Berman, University of Michigan, Ann Arbor, MI, USA

R G Brewer, Stanford University, Stanford, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical spectroscopy is a method that allows one to determine atomic and molecular transition frequencies, as well as relaxation rates in atomic and molecular vapors. The transition frequencies serve as 'fingerprints' that can be used to identify and classify atoms and molecules. Optical spectroscopic methods fall into two broad categories, continuous wave (cw) or stationary spectroscopy and time-dependent or transient spectroscopy. In cw spectroscopy, the absorption or emission line shapes are measured as a function of the frequency of a probe field. On the basis of theoretical line shape formulas, it is then possible to extract the relevant transition frequencies and relaxation rates from the line shapes. In transient optical spectroscopy, pulsed optical fields are used to create atomic state populations or coherences between atomic states. Following the

excitation, the time-evolution of the atoms is monitored, from which transition frequencies and relaxation rates can be obtained.

The earliest coherent transient effects were observed with atomic nuclear spins in about 1950 by H C Torrey, who discovered transient spin nutation, and by E L Hahn who detected spin echoes and free induction decay (FID). The optical analogs of these and other nuclear magnetic resonance (NMR) effects have now been detected using either optical pulses or techniques involving Stark or laser frequency switching.

In this brief introduction to coherent optical transients, several illustrative examples of coherent transient phenomena are reviewed. The optical Bloch equations are derived and coupled to Maxwell's Equations. The Maxwell–Bloch formalism is used to analyze free precession decay, photon echoes, stimulated photon echoes, and optical Ramsey fringes. Experimental results are presented along with the relevant theoretical calculations. In addition, coherent transient phenomena involving quantized matter waves are discussed. Although the examples considered in this chapter are fairly simple, it should be appreciated that sophisticated coherent transient techniques are routinely applied as a probe of complex structures such as liquids and solids or as a probe of single atoms and molecules.

Optical Bloch Equations

Many of the important features of coherent optical transients can be illustrated by considering the interaction of a radiation field with a two-level atom. The lower state $|1\rangle$ has energy $-\hbar\omega_0/2$ and upper state $|2\rangle$ has energy $\hbar\omega_0/2$. For the moment, the atom is assumed to be fixed at $\mathbf{R} = 0$ and all relaxation processes are neglected. The incident electric field is

$$\mathbf{E}(\mathbf{R} = 0, t) = \frac{1}{2}\boldsymbol{\varepsilon}[E(t)e^{-i\omega t} + E(t)e^{i\omega t}] \qquad [1]$$

where $E(t)$ is the field amplitude, $\boldsymbol{\varepsilon}$ is the field polarization and ω is the carrier frequency. The time dependence of $E(t)$ allows one to consider pulses having arbitrary shape. A time-dependent phase for the field could have been included, allowing one to consider the effect of arbitrary frequency 'chirps,' but such effects are not included in the present discussion. It is convenient to expand the atomic state wave function in a field interaction representation as

$$|\psi(t)\rangle = c_1(t)e^{i\omega t/2}|1\rangle + c_2(t)e^{-i\omega t/2}|2\rangle \qquad [2]$$

The atom-field interaction potential is

$$V(\mathbf{R}, t) = -\mu \cdot \mathbf{E}(\mathbf{R}, t) \qquad [3]$$

where μ is a dipole moment operator. When eqn [2] is substituted into Schrödinger's equation and rapidly varying terms are neglected (rotating-wave approximation), one finds that the state amplitudes evolve according to

$$i\hbar\frac{\mathrm{d}\mathbf{c}}{\mathrm{d}t} = \tilde{\mathbf{H}}\mathbf{c}; \qquad \tilde{\mathbf{H}} = (\hbar/2)\begin{pmatrix} -\delta_0 & \Omega_0(t) \\ \Omega_0(t) & \delta_0 \end{pmatrix} \qquad [4]$$

where $\mathbf{c}$ is a vector having components (c_1, c_2)

$$\delta_0 = \omega_0 - \omega \qquad [5]$$

is an atom-field detuning

$$\Omega_0(t) = -\frac{\mu E(t)}{\hbar} = -\frac{\mu}{\hbar}\sqrt{\frac{2S(t)}{\varepsilon_0 c}} \qquad [6]$$

is a Rabi frequency, $\mu = \langle 1|\mu\cdot\varepsilon|2\rangle = \langle 2|\mu\cdot\varepsilon|1\rangle$ is a dipole moment matrix element, ε_0 is the permittivity of free space, and $\mathbf{S}(t)$ is the time-averaged Poynting vector of the field. Equation [4] can be solved numerically for arbitrary pulse envelopes.

Expectation values of physical observables are conveniently expressed in terms of density matrix elements defined by

$$\rho_{ij} = c_i c_j^* \qquad [7]$$

which obey equations of motion

$$\begin{aligned} \dot{\rho}_{11} &= -i\Omega_0(t)(\rho_{21} - \rho_{12})/2 \\ \dot{\rho}_{22} &= i\Omega_0(t)(\rho_{21} - \rho_{12})/2 \\ \dot{\rho}_{12} &= -i\Omega_0(t)(\rho_{22} - \rho_{11})/2 + i\delta_0\rho_{12} \\ \dot{\rho}_{21} &= i\Omega_0(t)(\rho_{22} - \rho_{11})/2 - i\delta_0\rho_{21} \end{aligned} \qquad [8]$$

An alternative set of equations in terms of real variables can be obtained if one defines new parameters

$$\begin{aligned} u &= \rho_{12} + \rho_{21}; & \rho_{12} &= (u + iv)/2 \\ v &= i(\rho_{21} - \rho_{12}); & \rho_{21} &= (u - iv)/2 \\ w &= \rho_{22} - \rho_{11}; & \rho_{22} &= (m + w)/2 \\ m &= \rho_{11} + \rho_{22}; & \rho_{11} &= (m - w)/2 \end{aligned} \qquad [9]$$

which obey

$$\begin{aligned} \dot{u} &= -\delta_0 v \\ \dot{v} &= \delta_0 u - \Omega_0(t)w \\ \dot{w} &= \Omega_0(t)v \\ \dot{m} &= 0 \end{aligned} \qquad [10]$$

The last of these equations reflects the fact that $\rho_{11} + \rho_{22} = 1$, while the first three can be rewritten as

$$\dot{\mathbf{U}} = \mathbf{\Omega}(t) \times \mathbf{U} \qquad [11]$$

where the *Bloch vector* $\mathbf{U}$ has components (u, v, w) and the *pseudofield vector* $\mathbf{\Omega}(t)$ has components $[\Omega_0(t), 0, \delta_0]$.

Equations [8] or [10] constitute the optical Bloch equations without decay. The vector $\mathbf{U}$ has unit magnitude and traces out a path on the Bloch sphere as it precesses about the pseudofield vector. The component w is the population difference of the two atomic states, while u and v are related to the quadrature components of the atomic polarization (see below).

Equations [8] and [10] can be generalized to include relaxation. A simplified relaxation scheme that has a wide range of applicability is one in which state 2 decays spontaneously to state 1 with rate γ_2, while the relaxation of the coherence ρ_{12} is characterized by a complex decay parameter Γ_{12}, resulting from phase-changing collisions with a background gas. When such processes are included in eqn [8], they are modified as

$$\dot{\rho}_{11} = -i\Omega_0(t)(\rho_{21} - \rho_{12})/2 + \gamma_2\rho_{22} \qquad [12a]$$

$$\dot{\rho}_{22} = i\Omega_0(t)(\rho_{21} - \rho_{12})/2 - \gamma_2\rho_{22} \qquad [12b]$$

$$\dot{\rho}_{12} = -i\Omega_0(t)(\rho_{22} - \rho_{11})/2 - \gamma\rho_{12} + i\delta\rho_{12} \qquad [12c]$$

$$\dot{\rho}_{21} = i\Omega_0(t)(\rho_{22} - \rho_{11})/2 - \gamma\rho_{21} - i\delta\rho_{21} \qquad [12d]$$

where $\gamma = \gamma_2/2 + \mathrm{Re}(\Gamma_{12})$ and the detuning $\delta = \delta_0 - \mathrm{Im}(\Gamma_{12})$ is modified to include the collisional shift. With the addition of decay, the length of the Bloch vector is no longer conserved. The quantity γ_2 is referred to as the longitudinal relaxation rate. Moreover, one usually refers to $T_1 = \gamma_2^{-1}$ as the longitudinal relaxation time and $T_2 = \gamma^{-1}$ as the transverse relaxation time. In the case of purely radiative broadening, $\gamma_2 = 2\gamma$ and $T_1 = T_2/2$.

Maxwell–Bloch Equations

As a result of atom-field interactions, it is assumed that a polarization is created in the medium of the form

$$\mathbf{P}(\mathbf{R}, t) = \frac{1}{2}\boldsymbol{\varepsilon}\left[P(Z,t)e^{i[kZ-\omega t]} + P^*(Z,t)e^{-i[kZ-\omega t]}\right] \qquad [13]$$

which gives rise to a signal electric field of the form

$$\mathbf{E}_s(\mathbf{R}, t) = \frac{1}{2}\boldsymbol{\varepsilon}\left[E_s(Z,t)e^{i[kZ-\omega t]} + E_s^*(Z,t)e^{-i[kZ-\omega t]}\right] \qquad [14]$$

The z-axis has been chosen in the direction of $\mathbf{k}$. It follows from Maxwell's equations that the quasi-steady-state field amplitude is related to the polarization by

$$\frac{\partial E_s(Z,t)}{\partial Z} = \frac{ik}{2\varepsilon_0}P(Z,t) \qquad [15]$$

To arrive at eqn [15], it is necessary to assume that transverse field variations can be neglected, that the *phase matching* condition $k = \omega/c$ is met, and that the complex field amplitudes $P(Z,t)$ and $E_s(Z,t)$ vary slowly in space compared with e^{ikZ} and slowly in time compared with $e^{i\omega t}$.

The polarization $\mathbf{P}(\mathbf{R}, t)$, defined as the average dipole moment per unit volume, is given by the expression

$$\mathbf{P}(\mathbf{R}, t) = \mathcal{N}\Big[\boldsymbol{\mu}_{21}\langle\rho_{12}(Z,t)\rangle e^{-i[kZ-\omega t]} + \boldsymbol{\mu}_{12}\langle\rho_{21}(Z,t)\rangle e^{i[kZ-\omega t]}\Big] \qquad [16]$$

where $\mathcal{N}$ is the atomic density and $\rho_{12}(Z,t)$ and $\rho_{21}(Z,t)$ are single-particle density matrix elements that depend on the atomic properties of the medium. As such, eqn [16] provides the link between Maxwell's equations and the optical Bloch equations. The $\langle\,\rangle$ brackets in eqn [16] indicate that there may be additional averages that must be carried out. For example, in a vapor, there is a distribution of atomic velocities that must be summed over, while, in a solid, there may be an inhomogeneous distribution of atomic frequencies owing to local strains in the media. By combining eqns [9], [13], [15] and [16], and using the fact that $\mu = \mu_{12}\cdot\varepsilon = \mu_{21}\cdot\varepsilon$, one arrives at

$$\begin{aligned}\frac{\partial E_s(Z,t)}{\partial Z} &= \frac{ik\mathcal{N}\mu}{\varepsilon_0}\langle\rho_{21}(Z,t)\rangle \\ &= \frac{ik\mathcal{N}\mu}{2\varepsilon_0}\langle u(Z,t) - iv(Z,t)\rangle\end{aligned} \qquad [17]$$

which, together with eqn [12], constitute the Maxwell–Bloch equations. From eqn [17], it is clearly seen that the coherence $\langle\rho_{21}(Z,t)\rangle$ drives the signal field. If $\langle\rho_{21}(Z,t)\rangle$ is independent of Z, the intensity of radiation exiting a sample of length L is proportional to

$$I(L,t) = \frac{k\mathcal{N}\mu L}{\varepsilon_0}{}^2|\langle\rho_{21}(t)\rangle|^2 \qquad [18]$$

Alternatively, the outgoing field can be heterodyned with a reference field to extract the field amplitude rather than the field intensity.

Coherent Transient Signals

The Maxwell–Bloch equations can be used to describe the evolution of coherent transient signals. In a typical experiment, one applies one or more 'short' radiation pulses to an atomic sample and monitors the radiation field emitted by the atoms following the pulses. A 'short' pulse, is one satisfying

$$|\delta|\tau, ku\tau, \gamma\tau, \gamma_2\tau \ll 1 \qquad [19]$$

where τ is the pulse duration. Conditions [19] allow one to neglect any effects of detuning (including Doppler shifts or other types of inhomogeneous broadening) or relaxation during the pulse's action. The evolution of the atomic density matrix during each pulse is determined solely by the pulse area defined by

$$\theta = \int_{-\infty}^{\infty} \Omega_0(t)\mathrm{d}t \qquad [20]$$

In terms of density matrix elements in a standard interaction representation defined by

$$\rho_{12}^I = \rho_{12}e^{-i[kZ+\delta t]}, \quad \rho_{21}^I = \rho_{21}e^{i[kZ+\delta t]},$$
$$\rho_{22}^I = \rho_{22}, \qquad \rho_{11}^I = \rho_{11}$$

the change in density matrix elements for an interaction occurring at time T_j is given by

$$\begin{pmatrix} \rho_{11}^I \\ \rho_{22}^I \\ \rho_{12}^I \\ \rho_{21}^I \end{pmatrix}^+ = \frac{1}{2}\begin{pmatrix} 1+\cos\theta_j & 1-\cos\theta_j & i\sin\theta_j e^{i\Phi(T_j)} & -i\sin\theta_j e^{-i\Phi(T_j)} \\ 1-\cos\theta_j & 1+\cos\theta_j & -i\sin\theta_j e^{i\Phi(T_j)} & i\sin\theta_j e^{-i\Phi(T_j)} \\ i\sin\theta_j e^{-i\Phi(T_j)} & -i\sin\theta_j e^{-i\Phi(T_j)} & 1+\cos\theta_j & (1-\cos\theta_j)e^{-2i\Phi(T_j)} \\ -i\sin\theta_j e^{i\Phi(T_j)} & i\sin\theta_j e^{i\Phi(T_j)} & (1-\cos\theta_j)e^{2i\Phi(T_j)} & 1+\cos\theta_j \end{pmatrix}\begin{pmatrix} \rho_{11}^I \\ \rho_{22}^I \\ \rho_{12}^I \\ \rho_{21}^I \end{pmatrix}^- \qquad [21]$$

where $\mp$ superscripts refer to times just before and after the pulse

$$\Phi(T_j) = \mathbf{k}\cdot\mathbf{R}(T_j) + \delta T_j$$

$$\mathbf{R}(T_j) = \mathbf{R} - \mathbf{v}(t - T_j)$$

θ_j is the pulse area, and $\mathbf{v}$ is the atomic velocity. Between pulses, or following the last pulse, the atoms evolve freely in the absence of any applied fields. The density matrix elements evolve as

$$\dot{\rho}_{11}^I = \gamma_2\rho_{22}^I; \quad \dot{\rho}_{22}^I = -\gamma_2\rho_{22}^I;$$
$$\dot{\rho}_{12}^I = -\gamma\rho_{12}^I; \quad \dot{\rho}_{21}^I = -\gamma\rho_{21}^I \qquad [22]$$

A coherent transient signal is constructed by piecing together field interaction zones and free evolution periods. There are many different types of coherent transient signals. A few illustrative examples are given below.

Free Polarization Decay

Free polarization decay (FPD), also referred to as optical FID, is the optical analog of free induction decay (FID) in NMR. This transient follows preparation of the atomic sample by any of three methods: (1) an optical pulse resonant with the atomic transition; (2) a Stark pulse that switches the transition frequency of a molecule (atom) into resonance with an incident *cw* laser beam; or (3) an electronic pulse that switches the frequency of a *cw* laser into coincidence with the molecular (atomic) transition frequency. In methods (2) and (3), the FPD emission propagates coherently in the forward direction along with the *cw* laser beam, automatically generating a heterodyne beat signal at a detector that can be orders of magnitude larger than the FPD signal. The FPD field intensity itself exiting the sample is equal to

$$I(L,t) = \left(\frac{k\mathcal{N}\mu L}{\varepsilon_0}\right)^2 |\langle\rho_{21}(t)\rangle|^2$$
$$= \left(\frac{k\mathcal{N}\mu L}{2\varepsilon_0}\sin\theta\right)^2 \left|\int \mathrm{d}\mathbf{v}\, W_0(\mathbf{v}) e^{-[\gamma - i(\delta + \mathbf{k}\cdot\mathbf{v})]t}\right|^2 \qquad [23]$$

For a Maxwellian velocity distribution

$$W_0(\mathbf{v}) = \frac{1}{(\pi u^2)^{3/2}} e^{-(v/u)^2} \qquad [24]$$

where u is the most probable atomic speed, one finds

$$I(L,t) = \left(\frac{k\mathcal{N}\mu L}{2\varepsilon_0}\sin\theta\right)^2 e^{-2\gamma t} e^{-(kut)^2/2} \qquad [25]$$

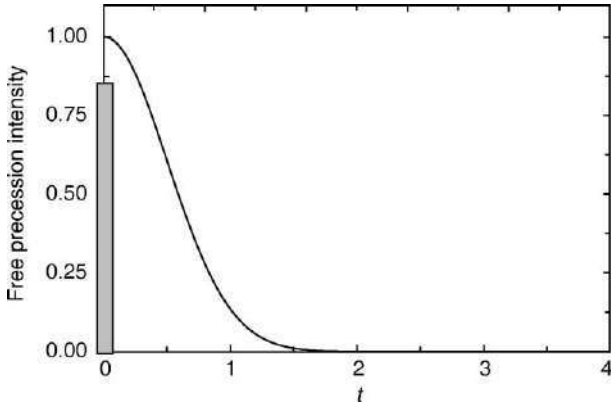

Figure 1 Pulse sequence and signal intensity for free precession decay. In this and subsequent figures, time is measured in units of $2/ku$.

If $ku \gg \gamma$, the signal decays mainly owing to inhomogeneous broadening, as shown in Figure 1. Owing to different Doppler shifts for atoms having different velocities, the optical dipoles created by the pulse lose their relative phase in a time of order $T_2^* = 2/ku$. The FPD signal can be used to measure T_2^*, which can be viewed as an inhomogeneous, transverse relaxation time. At room temperature ku/γ is typically of order of 100.

When FPD signals are observed using laser frequency switching, method 3 above, the *cw* field can modify the molecular velocity distribution, exciting only those atoms having $\mathbf{k}\cdot\mathbf{v} = -\delta \pm \gamma'$, where γ' is a power broadened homogeneous width. In the linear field regime, the FPD signal again decays on a time scale of order $(ku)^{-1}$, since the velocity distribution is unchanged to first order in the field amplitude. When one considers nonlinear interactions with the field, however, the *cw* field excites a narrow velocity subclass that is no longer subject to inhomogeneous broadening. These atoms give rise to a contribution to the FPD signal that decays with rate $T_2^{-1} = (\gamma + \gamma')$. Thus by using nonlinear atom-field interactions, one can extract homogeneous decay rates in situations where there is large inhomogeneous broadening. The price one pays is that only a small percentage of the atoms (those having velocities for which the applied field is resonant) contribute to the signal.

A FPD signal obtained on the D_2 transition in Cs is shown in Figure 2, where $T_2^* = 1.4$ ns. Oscillations in the signal originate from ground state hyperfine splitting. Figure 3 represents a FPD signal obtained in NH_2D at 10.6 μm using the method of Stark switching with *cw* state preparation. The oscillations are the heterodyne beat signal while the slowly varying increase in the signal is the result of optical nutation of molecules switched into resonance by the Stark pulse. The FPD signal manifests itself as a reduction of the amplitude of the oscillation with time. This amplitude decays with the (power-broadened) homogeneous decay rate γ'.

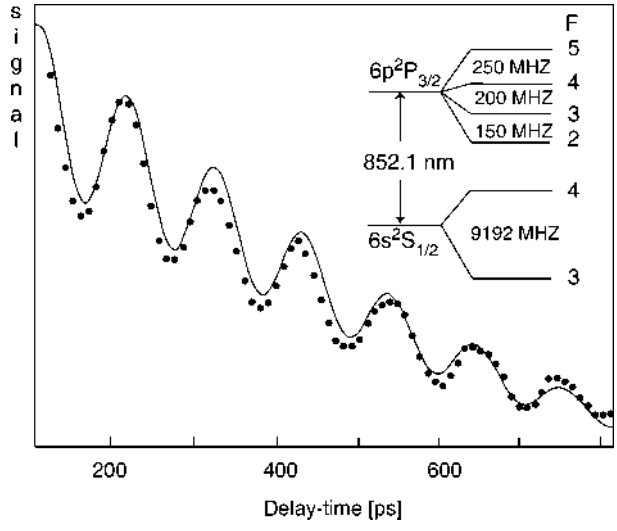

Figure 2 Optical free precession decay in Cs using an excitation pulse having 20 ns duration. Reproduced with permission from Lehmitz H and Harde H (1986) In: Prior Y, Ben-Reuven A and Rosenbluh M (eds) *Methods of Laser Spectroscopy*, pp. 109–112. New York: Plenum.

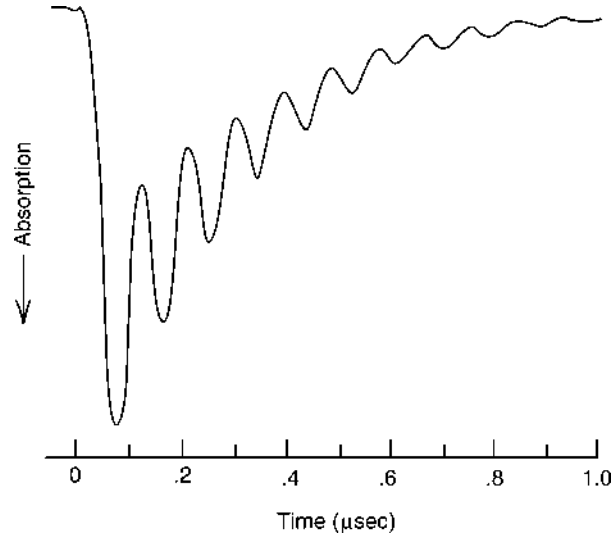

Figure 3 Optical free precession in NH_2D using the Stark switching technique. Reproduced with permission from Brewer RG and Shoemaker RL (1972) Optical free induction decay. *Physical Review A* 6: 2001. Copyright (1972) by the American Physical Society.

Photon Echo

Although the FPD signal produced by short excitation pulses decays in a time of order $(ku)^{-1}$, the coherence of individual atoms decays in a much longer time, T_2.

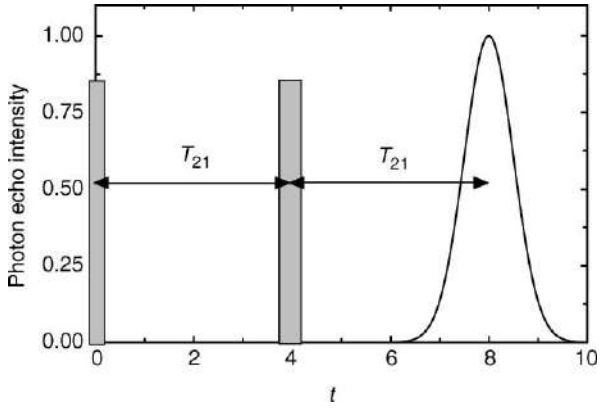

Figure 4 Pulse sequence and signal intensity for the photon echo.

The question arises as to whether it is possible to bring all the atomic dipoles back into phase so they can radiate a coherent signal. The photon echo accomplishes this goal, although it has very little to do with either photons or echoes. The pulse sequence and echo signal are shown in Figure 4. The first pulse creates a phased array of dipoles, which begin to dephase owing to inhomogeneous broadening. At time T_{21} the second pulse provides a nonlinear interaction that couples ρ_{12} and ρ_{21}. As a result of the second pulse, the dipoles begin a rephasing process that is completed at time $t = 2T_{21}$. The 'echo' appears at this time. The echo intensity exiting the sample is

$$\begin{aligned} I(L,t) &= \left(\frac{k\mathcal{N}\mu L}{\varepsilon_0}\right)^2 |\langle\rho_{21}(t)\rangle|^2 \\ &= \left(\frac{k\mathcal{N}\mu L}{2\varepsilon_0}\sin\theta_1 \sin^2(\theta_2/2)\right)^2 \\ &\quad \times e^{-4\gamma T_{21}} e^{-k^2u^2(t-2T_{21})^2/2} \end{aligned} \qquad [26]$$

If the echo intensity is measured as a function of T_{21}, one can extract the homogeneous decay rate $\gamma = T_{21}^{-1}$. It is interesting to note that the echo intensity near $t = 2T_{21}$ mirrors the FPD intensity immediately following the first pulse. For experimental reasons, it is often convenient to use a different propagation vector for the second pulse. If $\mathbf{k}_1$ and $\mathbf{k}_2$ are propagation vectors of the first and second pulses, the echo signal is in the direction $\mathbf{k} = 2\mathbf{k}_2 - \mathbf{k}_1$, provided $|\mathbf{k}_2 - \mathbf{k}_1|/k_1 \ll 1$.

The echo signal is sensitive only to $\rho_{21}(t)$ and $\rho_{12}(t)$ in the entire time interval of interest. Any interaction that results in a degradation of the dephasing–rephasing process for these density matrix elements leads to a decrease in the echo intensity. As such, echo signals can serve as a probe of transverse relaxation. Transverse relaxation generally falls into two broad categories. First, there are dephasing processes which produce an exponential damping of the coherences and contribute to γ. Second there is *spectral diffusion*; in a vapor, spectral diffusion is produced by velocity-changing collisions that change the effective field frequency seen by the atoms owing to the Doppler effect. Such terms enter the optical Bloch equations as *integral* terms, transforming the equations into differentio-integral equations. In general, the phase-changing and velocity-changing aspects of collisions are entangled; however, if the collisional interaction is *state-independent*, as it is for some molecular transitions, then collisions are purely velocity-changing in nature, leading to an echo that decays exponentially as T_{21}^3 for early times and T_{21} for later times. For electronic transitions, collisions are mainly phase-changing in nature, but there is a velocity-changing contribution that persists in the forward diffractive scattering cone. Photon echo techniques are sufficiently sensitive to measure the relaxation caused by such diffractive scattering.

The photon echo was first observed by NA Kurnit and co-workers, using a pulsed ruby laser (Figure 5). In Stark switching experiments, the photon echo follows two Stark pulses as shown in Figure 6. The effects of velocity-changing collisions on the photon echo signal in a CH_3F molecular vapor is represented in Figure 7. The signal exhibits exponential decay varying as T_{21}^3 for early times and T_{21} for later times. A theoretical interpretation of the observed echo decay permits a determination of the binary collision parameters for CH_3F, namely, the characteristic velocity jump $\Delta u_{\mathrm{rms}} = 200\ \mathrm{cm\,s^{-1}}$ and the cross section of $\sigma = 580\ \text{Å}^2$.

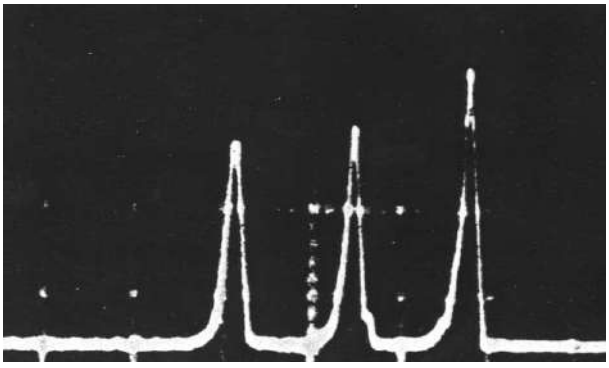

Figure 5 A photon echo signal from ruby. Time increases to the right with a scale of 100 ns/division. The pulse on the right is the echo signal, while the first two pulses are the (attenuated) input pulses. Reprinted with permission from Kurnit NA, Abella ID and Hartmann SR (1964) Observation of a photon echo. *Physical Review Letters* 13: 567. Copyright (1964) by the American Physical Society.

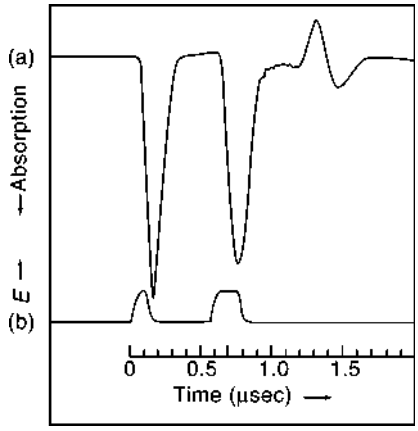

Figure 6 A photon echo signal in $^{13}CH_3F$ using the Stark switching method. The third pulse in trace (a) is the echo heterodyne beat signal. This signal is preceded by 'nutation' transient signals following each of the two Stark pulses shown in trace (b). Reprinted with permission from Brewer RG and Shoemaker RL (1971) Photo echo and optical nutation in molecules. *Physical Review Letters* 27: 631. Copyright (1971) by the American Physical Society.

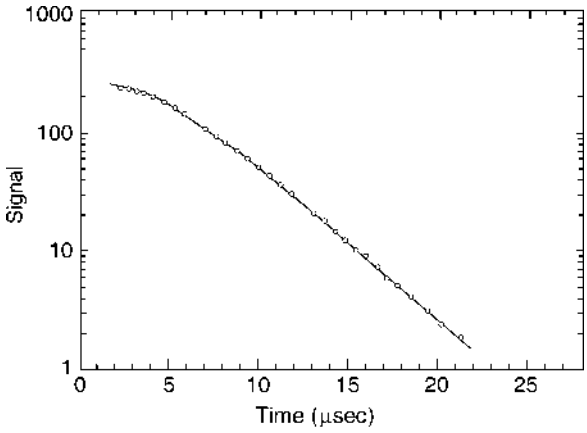

Figure 7 Decay of the photon echo signal in $^{13}CH_3F$ as a function of pulse separation. Reprinted with permission from Schmidt J, Berman PR and Brewer RG (1973) Coherent transient study of velocity-changing collisions. *Physical Review Letters* 31: 1103. Copyright (1973) by the American Physical Society.

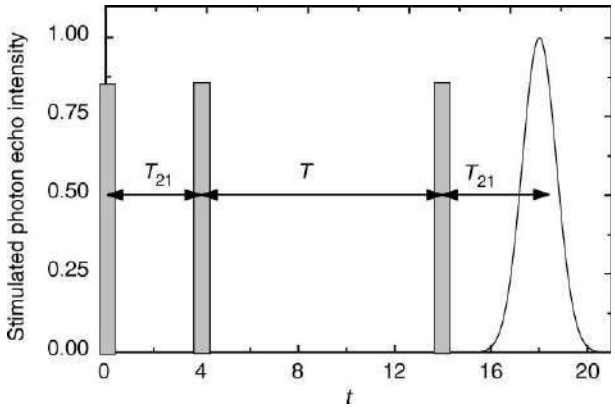

Figure 8 Pulse sequence and signal intensity for the stimulated photon echo.

Stimulated Photon Echo

In contrast to conventional photon echoes, *stimulated photon echoes* can be used to *simultaneously* measure both transverse and longitudinal relaxation times. Stimulated photon echoes have become an important diagnostic probe of relaxation in condensed matter systems as well as atomic vapors. The pulse sequence is shown in Figure 8 and consists of three pulses, having areas $\theta_1, \theta_2, \theta_3$, and equal propagation vectors $\mathbf{k}$. The time interval between the first two pulses is $T_{21} \gg T_2^*$ and pulse 3 occurs at time $t = T_{21} + T$. The maximum echo amplitude is given by

$$I(L, t_e) = \frac{1}{2}\left(\frac{k\mathcal{N}\mu L}{2\varepsilon_0}\sin\theta_1 \sin\theta_2 \sin\theta_3\right)^2 \times e^{-2\gamma_2 T} e^{-4\gamma T_{21}} \qquad [27]$$

and occurs when $t_e - (T_{21} + T) = T_{21}$. The optimal pulse sequence consists of three $\pi/2$ pulses.

If slightly different $\mathbf{k}$ vectors are chosen such that $\mathbf{k}_1 = \mathbf{k}_3 \approx \mathbf{k}_2$ the echo intensity varies as

$$e^{-2\gamma_2 T} e^{-4\gamma T_{21}} e^{-|\mathbf{k}_1 - \mathbf{k}_2|^2 u^2 (T + 2T_{21})^2/2}$$

By varying the angle between $\mathbf{k}_1$and $\mathbf{k}_2$, one can determine the Doppler width ku. By monitoring the echo signal as a function of $T_{21}(T)$, one obtains information on the transverse (longitudinal) relaxation.

The situation changes dramatically if there is an *additional* ground state level (level 0) to which the excited state can decay. In that case, the system is 'open' since population can leave the 'two-level' subsystem. The total population of the 1–2 state subsystem is no longer conserved, requiring an *additional* decay rate to account for relaxation. Let us suppose that all states decay with rate Γ_t as a result of their finite time in the laser beams. Moreover, let $\Gamma_{2,1}$ and $\Gamma_{2,0}$ be the decay rates of level 2 to levels 1 and 0, respectively. Assuming that $\gamma_2 T \gg 1$, one finds that the factor $e^{-2\gamma_2 T}$ in eqn [27] must be replaced by $(\Gamma_{2,0}/(\Gamma_{2,1} + \Gamma_{2,0}))^2 e^{-2\Gamma_t T}$. If $\Gamma_{2,0} \neq 0$, there is a long-lived component in the ground state population that contributes to the echo signal. One can exploit this feature of open systems to study spectral diffusion or velocity-changing collisions with very high sensitivity. Echoes can occur for time separations T much greater than the excited state lifetime. The creation of such long-lived stimulated photon echoes can be attributed to the fact that, in open systems, a portion of the velocity modulation created by the first two pulses in the ground state remains following spontaneous decay from the excited state.

The stimulated photon echo observed on the D_1 transition in sodium is shown in Figure 9. This is

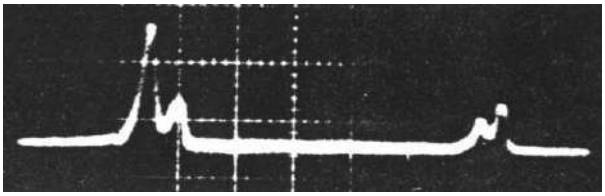

Figure 9 Stimulated photon echo observed on the D_1 transition in sodium. Reprinted with permission from Mossberg T, Flusberg A, Kachru R and Hartmann SR (1979) Total scattering cross section for Na on He measured by stimulated photon echoes. *Physical Review Letters* 42: 1665. Copyright (1979) by the American Physical Society.

an 'open' system, owing to ground state hyperfine structure. For the data shown, T is 17 times the 16 ns lifetime of the excited state. The first three pulses represent light scattered from the three input pulses and the fourth pulse is the echo. The echo appears at $t = T + 2T_{21}$.

Optical Ramsey Fringes

We have seen that coherent transients can be used to measure relaxation rates. With a slight modification of the stimulated photon echo geometry, it is also possible to use coherent transient signals to measure transition frequencies. If one chooses $\mathbf{k}_3$ in the $-\mathbf{k}_1$ direction instead of the $\mathbf{k}_1$ direction, it is possible to generate a phase matched signal in the

$$\mathbf{k} = \mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_3 = -\mathbf{k}_2$$

direction. To simplify matters, it is assumed that $\mathbf{k}_1 \approx \mathbf{k}_2 = -\mathbf{k}$ and terms of order $|\mathbf{k}_1 - \mathbf{k}_2| u \times (T + 2T_{21})$ are neglected.

The averaged density matrix element in the vicinity of the echo is

$$\langle \rho_{21}(t) \rangle = (i/8) \sin\theta_1 \sin\theta_2 \sin\theta_3 e^{-\gamma_2 T} e^{-2\gamma T_{21}} \times e^{i\delta[(t-T_{21}-T)+T_{21}]} e^{-k^2u^2|(t-T_{21}-T)-T_{21}|^2/4} \quad [28]$$

An echo is formed at time $t_e = T + 2T_{21}$, just as for the stimulated photon echo, but there is an additional phase factor, given by $e^{2i\delta T_{21}}$ when $t = t_e$, that is absent for the nearly collinear geometry. This phase factor is the optical analog of the phase factor that is responsible for the generation of Ramsey fringes in the microwave domain. The phase factor can be measured directly by heterodyning the signal field with a reference field or, indirectly, by converting the off-diagonal density matrix element into a population by the addition of a *fourth* pulse in the $\mathbf{k}_3$ direction at time $t = T + 2T_{21}$. In either case, the signal varies as $\cos(2\delta T_{21})$. The central fringe, corresponding to $\delta = 0$ can be isolated using one of two methods. If the experiment is carried out using an atomic beam rather than atoms in a cell, $T_{21} = L/u_0$, will be different for atoms having different u_0 (L = spatial separation of the first two pulses and u_0 is the longitudinal velocity of the atoms). On averaging over a distribution of u_0, the fringe having $\delta = 0$ will have the maximum amplitude. Experiments of this type allow one to measure optical frequencies with accuracy of order T_{21}^{-1}. For experiments using temporally-separated pulses acting on atoms in a cell, it is necessary to take data as a function of δ for several values of T_{21}, and then average the data over T_{21}. It might seem remarkable that narrow pulses having large bandwidths can be used to measure optical frequencies with arbitrary accuracy as T_{21} is increased. Of course, it is the *total* duration of the pulse sequence rather than the duration of a single pulse that is the relevant time parameter. For the optical coherence, T_{21} is the appropriate time parameter (during the interval T, it is population rather than coherence that contributes to the signal). The price one pays by using large T_{21} is a signal that decays as $e^{-4\gamma T_{21}}$.

Figure 10 illustrates an optical Ramsey fringe signal on the 657 nm intercombination line in Ca. Four-field zones were used. The most probable value of T_{21} was about 10^{-5} s for an effusive beam having a most probable longitudinal speed equal to 800 ms^{-1}, giving a central fringe width of order 60 kHz.

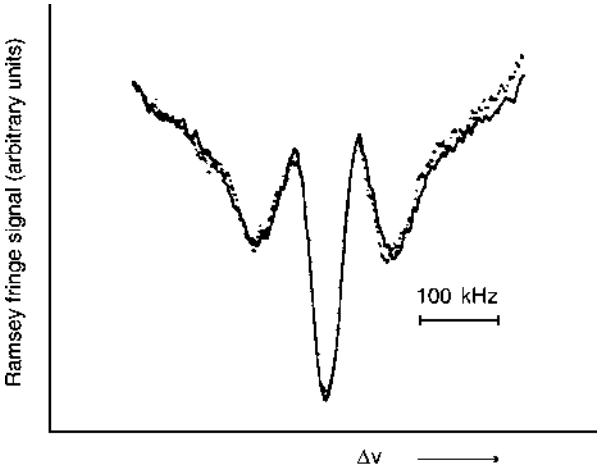

Figure 10 An optical Ramsey fringe signal on the 657 nm intercombination line in Ca. Four field zones were used. The solid and dotted lines represent runs with the directions of the laser field reversed, to investigate phase errors in the signals. Reprinted from Ito N, Ishikawa J and Moringa A (1994) Evaluation of the optical phase shift in a Ca Ramsey fringe stabilized optical frequency standard by means of laser-beam reversal. *Optics Communications* 109: 414, with permission from Elsevier.

Atom Interferometry

Matter wave atom interferometers rely on the wave nature of the center-of-mass motion for their operation. Such interferometers illustrate some interesting features of coherent optical transients not found in NMR. As an example, one can consider the response of a cold atomic ensemble to two, off-resonant standing-wave optical pulses separated in time by T. The electric field amplitude of pulse $j(j = 1, 2)$ is given by $\mathbf{E}_j(\mathbf{Z}, t) = \boldsymbol{\varepsilon} E_j(t)\cos(kZ)\cos(\omega t)$. The net effect of the field is to produce a spatially-modulated, ac Stark or light shift of the ground state energy. As a result, pulse j modifies the ground state amplitude by the phase factor $\exp[i\theta_j \cos(2kZ)]$. In other words, the standing wave field acts as a phase grating for the atoms. The phase factor can be interpreted in terms of momentum exchange between the two traveling wave components of the standing wave field. Owing to the fields' interaction with the atoms, all even integral multiples of $2\hbar k$ can be exchanged by the fields, imparting impulsive momenta of $2n\hbar k$ (n is a positive or negative integer) to the atoms. For an atom having momentum $\mathbf{P}$ and mass M, the frequency change associated with this momentum change is

$$E_{P,\mathbf{P}\pm 2n\hbar k\hat{Z}}/\hbar = \left(\frac{P^2}{2M} - \frac{|\mathbf{P} \pm 2n\hbar k\hat{Z}|^2}{2M}\right)\Big/\hbar = \mp\frac{2nP_Z k}{M} - \omega_{2nk} \qquad [29]$$

where

$$\omega_\kappa = \frac{\hbar\kappa^2}{2M}$$

There are two contributions to the frequency eqn [29]. The first part is independent of $\hbar$ and represents a classical Doppler shift, while the second part is proportional to $\hbar$ and represents a quantum, matter-wave effect. The quantum contribution will become important for times of order ω_{2nk}^{-1}. As with the echoes described previously, the Doppler effect results in a dephasing and rephasing of the optical response. However, in contrast to the previous echoes, the ground state density is no longer constant for times $t > \omega_{2nk}^{-1}$, when matter wave effects begin to play a role. At such times the phase grating created by the pulses can be transformed into amplitude gratings of the atomic density.

Long interaction times are possible, limited only by the time the atoms spend in the field. As such, atom interferometers offer exciting new possibilities for precision measurements of rotation rates, fundamental constants such as $\hbar$, gravitational acceleration, and gravitational gradients. Moreover, one can take advantage of the nonlinear atom-field interaction to create atomic density patterns having period $\lambda/2n$ from a standing wave optical field having wavelength λ. In this manner one has the possibility to construct atom interferometers and atom structures that operate on the nano-scale (spacings on the order of tens of nanometers).

Conclusion

Optical coherent transients are now used routinely as probes of atomic vapors, liquids, and condensed matter. Sophisticated techniques have been developed using fast and ultrafast pulses (having duration of 10 to 100 femtoseconds) to probe such systems and obtain the relevant relaxation rates. The manner in which optical coherent transients can be used to extract information on transition frequencies and relaxation rates in atomic and molecular vapors has been reviewed. In atomic vapors, optical coherent transients provide a means for obtaining both longitudinal and transverse relaxation rates. Moreover, echo experiments serve as an extremely sensitive probe of velocity changing collisions. Similar techniques are being rediscovered as an important tool in the analysis of cold atomic vapors and Bose–Einstein condensates. Only the most basic optical coherent transient phenomena have been described in this chapter. More elaborate, multi-pulse excitation schemes are joining established methods in attempts to understand the complex decay dynamics that one encounters in many branches of physics, chemistry and biology.

Acknowledgments

The research of PRB is supported by the National Science Foundation under Grants No. PHY-0098016 and the FOCUS Center Grant, and by the U. S. Army Research Office under Grant No. DAAD19-00-1-0412.

List of Units and Nomenclature

$\mathbf{E}(\mathbf{R}, t)$	Electric field vector
$E(t)$	Electric field amplitude
$E_s(Z, t)$	Complex signal electric field amplitude
$I(L, t)$	Signal intensity exiting the sample
$\mathbf{k}$	Field propagation vector
L	Sample length
$\mathcal{N}$	Atomic density
$\mathbf{P}$	Center-of-mass momentum
$\mathbf{P}(\mathbf{R}, t)$	Polarization vector
$P(Z, t)$	Complex polarization field amplitude
T_{21}, T	Time interval between pulses
u	Most probable atomic speed

(u, v, w)	Elements of Bloch vector
$\mathbf{U}(t)$	Bloch vector
$\mathbf{v}$	Atomic velocity
$W_0(\mathbf{v})$	Atomic velocity distribution
γ	Transverse relaxation rate
γ_2	Excited state decay rate or longitudinal relaxation rate
δ	Atom-field detuning
θ	Pulse area
μ	Dipole moment matrix element
ρ_{ij}^{I}	Density matrix element in 'normal' interaction representation
$\rho_{ij}(Z, t)$	Density matrix element in a field interaction representation
τ	Pulse durations
ω	Field frequency
ω_k	Recoil frequency
ω_0	Atomic transition frequency
$\Omega(t)$	Generalized Rabi frequency
$\mathbf{\Omega}(t)$	Pseudofield vector
$\Omega_0(t)$	Rabi frequency

See also

Interferometry: Gravity Wave Detection; Overview.

Further Reading

Allen L and Eberly J (1975) *Optical Resonance and Two-Level Atoms*, Chapters 2,4,9, New York: Wiley.

Berman PR and Steel DG (2001) *Coherent optical transients*. In: Bass M, Enoch JM, Van Stryland E and Wolf WL (eds) Handbook of Optics, vol. IV. chap. 24 New York: McGraw-Hill. (Material presented in this chapter is a condensed version of this reference.).

Brewer RG (1977) Coherent optical spectroscopy. In: Balian R, Haroche S and Liberman S (eds) *Frontiers in Laser Spectroscopy, Les Houches, Session XXVII*, vol. 1, pp. 341–396. Amsterdam: North Holland.

Brewer RG (1977) Coherent optical spectroscopy. In: Bloembergen N (ed.) *Nonlinear Spectroscopy*, Proc. Int. School Phys., Enrico Fermi, Course 64, Amsterdam: North-Holland.

Brewer RG (1977) *Coherent optical transients*. In *Physics Today*, vol. 30(5), 50–59.

Brewer RG and Hahn EL (1984) Atomic memory. In *Scientific American*, vol. 251(6), 50–57.

Journal of Optical Society America (1986) Special issue on coherent optical spectroscopy, Vol. A3(4).

Levenson MD and Kano SS (1988) *Introduction to Nonlinear Laser Spectroscopy*, Chapter 6. Boston: Academic Press.

Meystre P and Sargent M III (1999) *Elements of Quantum Optics*, 3rd edn. Chapter 12. Berlin: Springer-Verlag.

Shen YR (1984) *The Principles of Nonlinear Optics*, Chapter 17. New York: Wiley.

Shoemaker RL (1978) Coherent transient infrared spectroscopy. In: Steinfeld JI (ed.) *Laser and Coherence Spectroscopy*, pp. 197–371. New York: Plenum.

Zinth W and Kaiser W (1988) *Ultrafast Coherent Spectroscopy*. In: Kaiser W (ed.) *Ultrashort Laser Pulses and Applications*, vol. 60, pp. 235–277. Berlin: Springer-Verlag.

Foundations of Coherent Transients in Semiconductors

T Meier and S W Koch, Philipps-University, Marburg, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Electromagnetic irradiation of matter generates a coherent superposition of the excited quantum states. The attribute 'coherent' relates to the fact that the material excitations originally have a well defined phase dependence which is imposed by the phase of the excitation source, often a laser in the optical regime. Macroscopically, the generated superposition state can be described as an optical polarization which is determined by the transition amplitudes between the participating quantum states.

The optical polarization is a typical non-equilibrium quantity that decays to zero when a system relaxes to its equilibrium state. Coherent effects are therefore only observable in a certain time window after pulsed photo-excitation, or in the presence of a continuous-wave (cw) beam. As coherent transients one usually refers to phenomena that can be observed during or shortly after pulsed laser excitation and critically depend on the presence of the induced optical polarization.

Many materials such as atoms, molecules, metals, insulators, semiconductors including bulk crystals, heterostructures, and surfaces, as well as organic and biological structures are studied using coherent optical spectroscopy. Depending on the particular system, the states participating in the optical transitions, the interactions among them, and the resulting time-scale for the decay of the induced polarization may be very different. As a result, the time window during which coherent effects are

observable can be as long as seconds for certain atomic transitions, or it may be as short as a few femtoseconds (10^{-15}s), e.g., for metals, surfaces, or highly excited semiconductors.

Coherent spectroscopy and the analysis of coherent transients has provided valuable information on the nature and dynamics of the optical excitations. Often it is possible to learn about the interaction processes among the photoexcitations and to follow the temporal evolution of higher-order transitions which are only accessible if the system is in a non-equilibrium state.

The conceptually simplest experiment which one may use to observe coherent transients is to time resolve the transmission or reflection induced by a single laser pulse. However, much richer information can be obtained if one excites the system with several pulses. A pulse sequence with well-controlled delay times makes it possible to study the dynamical evolution of the photo-excitations, not only by time resolving the signal, but also by varying the delay. Prominent examples of such experiments are pump-probe measurements, which usually are performed with two incident pulses, or four-wave mixing, for which one may use two or three incident pulses.

Besides the microscopic interaction processes, the outcome of an experiment is determined by the quantum mechanical selection rules for the transitions and by the symmetries of the system under investigation. For example, if one wants to investigate coherent optical properties of surface states one often relies on phenomena, such as second-harmonic or sum-frequency generation, which give no signal in perfect systems with inversion symmetry. Due to the broken translational invariance, such experiments are therefore sensitive only to the dynamics of surface and/or interface excitations.

In this article the basic principles of coherent transients are presented and several examples are presented. The basic theoretical description and its generalization for the case of semiconductors are introduced.

Basic Principles

In the absence of free charges and currents, Maxwell's equations show that the electromagnetic field interacts with matter via the optical polarization. This polarization **P**, or more precisely its second derivative with respect to time $(\partial^2/\partial t^2)\mathbf{P}$, appears as a source term in the wave equation for the electric field **E**. Consequently, if the system is optically thin such that propagation effects within the sample can be ignored and if measurements are performed in the far field region, i.e., at distances exceeding the characteristic optical wavelength λ, the emitted electric field resulting from the polarization is proportional to its second time derivative, $\mathbf{E} \propto (\partial^2/\partial t^2)\mathbf{P}$. Thus the measurement of the emitted field dynamics yields information about the temporal evolution of the optical material polarization.

Microscopically the polarization is determined by the transition amplitudes between the different states of the system. These may be the discrete states of atoms or molecules, or the microscopic valence and conduction band states in a dielectric medium, such as a semiconductor. In any case, the macroscopic polarization **P** is computed by summing over all microscopic transitions p_{cv} via $\mathbf{P} = \sum_{c,v} (\boldsymbol{\mu}_{cv} p_{cv} + c.c.)$, where $\boldsymbol{\mu}_{cv}$ is the dipole matrix element which determines the strength of the transitions between the states v and c, and $c.c.$ denotes the complex conjugate. If ε_c and ε_v are the energies of these states, their dynamic quantum mechanical evolution is described by the phase factors $e^{-i\varepsilon_c t/\hbar}$ and $e^{-i\varepsilon_v t/\hbar}$, respectively. Therefore, each p_{cv} is evolving in time according to $e^{-i(\varepsilon_c - \varepsilon_v)t/\hbar}$. Assuming that we start at $t = 0$ with $p_{cv}(t = 0) = p_{cv,0}$, which may be induced by a short optical pulse, we have for the optical polarization $\mathbf{P}(t) = \sum_{c,v} (\boldsymbol{\mu}_{cv} p_{cv,0} e^{-i(\varepsilon_c - \varepsilon_v)t/\hbar} + c.c.)$. Thus $\mathbf{P}(t)$ is given by a summation over microscopic transitions which all oscillate with frequencies proportional to the energy differences between the involved states. Hence, the optical polarization is clearly a coherent quantity which is characterized by amplitude and phase. Furthermore, the microscopic contributions to $\mathbf{P}(t)$ add up coherently. Depending on the phase relationships, one may obtain either constructive superposition, interference phenomena like quantum beats, or destructive interference leading to a decay (dephasing) of the macroscopic polarization.

Optical Bloch Equations

The dynamics of photo-excited systems can be conveniently described by a set of equations, the optical Bloch equations, named after Felix Bloch (1905–1983) who first formulated such equations to analyze the spin dynamics in nuclear magnetic resonance. For the simple case of a two-level model, the Bloch equations can be written as

$$i\hbar \frac{\partial}{\partial t} p = \Delta\varepsilon p + \mathbf{E}\cdot\boldsymbol{\mu} I \qquad [1]$$

$$i\hbar \frac{\partial}{\partial t} I = 2\mathbf{E}\cdot\boldsymbol{\mu}(p - p^*) \qquad [2]$$

Here, $\Delta\varepsilon$ is the energy difference and I the inversion, i.e., the occupation difference between upper and lower state. The field **E** couples the polarization to the product of the Rabi energy $\mathbf{E}\cdot\boldsymbol{\mu}$

and the inversion I. In the absence of the driving field, i.e., $\mathbf{E} \equiv 0$, eqn [1] describes the free oscillation of p discussed above.

The inversion is determined by the combined action of the Rabi energy and the transition p. The total occupation N, i.e., the sum of the occupations of the lower and the upper states, remains unchanged during the optical excitation since light does not create or destroy electrons; it only transfers them between different states. If N is initially normalized to 1, then I can vary between -1 and 1. $I = -1$ corresponds to the ground state of the system, where only the lower state is occupied. In the opposite extreme, i.e., for $I = 1$, the occupation of the upper state is 1 and the lower state is completely depleted.

One can show that eqns [1]–[2] contain another conservation law. Introducing $\mathcal{P} = p + p^* = 2Re[p]$ and $J = i(p - p^*) = -2Im[p]$, which are real quantities and often named polarization and polarization current, respectively, one finds that the relation

$$\mathcal{P}^2 + J^2 + I^2 = 1 \qquad [3]$$

is fulfilled. Thus the coupled coherent dynamics of the transition and the inversion can be described on a unit sphere, the so-called Bloch sphere. The three terms $\mathcal{P}, J$, and I can be used to define a three-dimensional Bloch vector $\mathbf{S} = (\mathcal{P}, J, I)$. With these definitions, one can reformulate eqns [1]–[2] as

$$\frac{\partial}{\partial t}\mathbf{S} = \mathbf{\Omega} \times \mathbf{S} \qquad [4]$$

with $\mathbf{\Omega} = (-2\boldsymbol{\mu}\cdot\mathbf{E}, 0, \Delta\varepsilon)/\hbar$ and $\times$ denoting the vector product.

The structure of eqn [4] is mathematically identical to the equations describing either the angular momentum dynamics in the presence of a torque or the spin dynamics in a magnetic field. Therefore, the vector $\mathbf{S}$ is often called pseudospin. Moreover, many effects which can be observed in magnetic resonance experiments, e.g., free decay, quantum beats, echoes, etc. have their counterparts in photo-excited optical systems.

The vector product on the right-hand side of eqn [4] shows that $\mathbf{S}$ is changed by $\mathbf{\Omega}$ in a direction perpendicular to $\mathbf{S}$ and $\mathbf{\Omega}$. For vanishing field the torque $\mathbf{\Omega}$ has only a z-component. In this case the z-component of $\mathbf{S}$, i.e., the inversion, remains constant, whereas the x- and y-components, i.e., the polarization and the polarization current, oscillate with the frequency $\Delta\varepsilon/\hbar$ on the Bloch sphere around $\mathbf{\Omega}$.

Semiconductor Bloch Equations

If one wants to analyze optical excitations in crystallized solids, i.e., in systems which are characterized by a periodic arrangement of atoms, one has to include the continuous dispersion (bandstructure) into the description. In translationally invariant, i.e., ordered systems, this can be done by including the crystal momentum $\hbar\mathbf{k}$ as an index to the quantities which describe the optical excitation of the material, e.g., p and I in eqns [1]–[2]. Hence, one has to consider a separate two-level system for each crystal momentum $\hbar\mathbf{k}$. As long as all other interactions are disregarded, the bandstructure simply introduces a summation over the uncoupled contributions from the different $\mathbf{k}$ states, leading to inhomogeneous broadening due to the presence of a range of transition frequencies $\Delta\varepsilon(\mathbf{k})/\hbar$.

For any realistic description of optical processes in solids, it is essential to go beyond the simple picture of non-interacting states and to treat the interactions among the elementary material excitations, e.g., the Coulomb interaction between the electrons and the coupling to additional degrees of freedom, such as lattice vibrations (electron–phonon interaction) or other bath-like subsystems. If crystals are not ideally periodic, imperfections, which can often be described as a disorder potential, need to be considered as well.

All these effects can be treated by proper extensions of the optical Bloch equations introduced above. For semiconductors the resulting generalized equations are known as semiconductor Bloch equations, where the microscopic interactions are included at a certain level of approximation. For a two-band model of a semiconductor, these equations can be written schematically as

$$i\hbar\frac{\partial}{\partial t}p_{\mathbf{k}} = \Delta\varepsilon_{\mathbf{k}}p_{\mathbf{k}} + \Omega_{\mathbf{k}}(n^c_{\mathbf{k}} - n^v_{\mathbf{k}}) + i\hbar\frac{\partial}{\partial t}p_{\mathbf{k}}|_{\text{corr}} \qquad [5]$$

$$i\hbar\frac{\partial}{\partial t}n^c_{\mathbf{k}} = (\Omega^*_{\mathbf{k}}p_{\mathbf{k}} - \Omega_{\mathbf{k}}p^*_{\mathbf{k}}) + i\hbar\frac{\partial}{\partial t}n^c_{\mathbf{k}}|_{\text{corr}} \qquad [6]$$

$$i\hbar\frac{\partial}{\partial t}n^v_{\mathbf{k}} = -(\Omega^*_{\mathbf{k}}p_{\mathbf{k}} - \Omega_{\mathbf{k}}p^*_{\mathbf{k}}) + i\hbar\frac{\partial}{\partial t}n^v_{\mathbf{k}}|_{\text{corr}} \qquad [7]$$

Here $p_{\mathbf{k}}$ is the microscopic polarization and $n^c_{\mathbf{k}}$ and $n^v_{\mathbf{k}}$ are the electron populations in the conduction and valence bands (c and v), respectively. Due to the Coulomb interaction and possibly further processes, the transition energy $\Delta\varepsilon_{\mathbf{k}}$ and the Rabi energy $\Omega_{\mathbf{k}}$ both depend on the excitation state of the system, i.e., they are functions of the time-dependent polarizations and populations. This leads, in particular, to a coupling among the excitations for all different values of the crystals momentum $\hbar\mathbf{k}$. Consequently, in the presence of interactions, the optical excitations can no longer be described as independent two-level systems but have to be treated

as a coupled many-body system. A prominent and important example in this context is the appearance of strong exciton resonances which, as a consequence of the Coulomb interaction, show up in the absorption spectra of semiconductors energetically below the fundamental bandgap.

The interaction effects lead to significant mathematical complications since they induce couplings between all the different quantum states in a system and introduce an infinite hierarchy of equations for the microscopic correlation functions. The terms given explicitly in eqns [5]–[7] arise in a treatment of the Coulomb interaction on the Hartree-Fock level. Whereas this level is sufficient to describe excitonic resonances, there are many additional effects, e.g., excitation-induced dephasing due to Coulomb scattering and significant contributions from higher-order correlations, like excitonic populations and biexcitonic resonances, which make it necessary to treat many-body correlation effects that are beyond the Hartree-Fock level. These contributions are formally included by the terms denoted as $|_{\mathrm{corr}}$ in eqns [5]–[7]. The systematic truncation of the many-body hierarchy and the analysis of controlled approximations is the basic problem in the microscopic theory of optical processes in condensed matter systems.

Examples

Radiative Decay

The emitted field originating from the non-equilibrium coherent polarization of a photo-excited system can be monitored by measuring the transmission and the reflection as function of time. If only a single isolated transition is excited, the dynamic evolution of the polarization and therefore of the transmission and reflection is governed by radiative decay. This decay is a consequence of the coupling of the optical transition to the light field described by the combination of Maxwell- and Bloch-equations. Radiative decay simply means that the optical polarization is converted into a light field on a characteristic time scale $2T_{\mathrm{rad}}$. Here, T_{rad} is the population decay time, often also denoted as T_1-time. This radiative decay is a fundamental process which limits the time on which coherent effects are observable for any photo-excited system. Due to other mechanisms, however, the non-equilibrium polarization often vanishes considerably faster.

The value of T_{rad} is determined by the dipole matrix element and the frequency of the transition, i.e., $T_{\mathrm{rad}}^{-1} \propto |\boldsymbol{\mu}|^2 \Delta\omega$, with $\Delta\omega = \Delta\varepsilon/\hbar$. The temporal evolution of the polarization and the emitted field are proportional to $e^{-i\Delta\omega t - t/(2T_{\mathrm{rad}})}$. Usually one measures the intensity of a field, i.e., its squared modulus, which evolves as $e^{-t/T_{\mathrm{rad}}}$ and thus simply shows an exponential decay, the so-called *free-induction decay* (see Figure 1). T_{rad} can be very long for transitions with small matrix elements. For semiconductor quantum wells, however, it is in the order of only 10 ps, as the result of the strong light-matter interaction.

The time constant on which the optical polarization decays is often called T_2. In the case where this decay is dominated by radiative processes, we thus have $T_2 = 2T_{\mathrm{rad}}$.

Superradiance

The phenomenon of superradiance can be discussed considering an ensemble of N two-level systems which are localized at certain positions $\mathbf{R}_i$. In this case

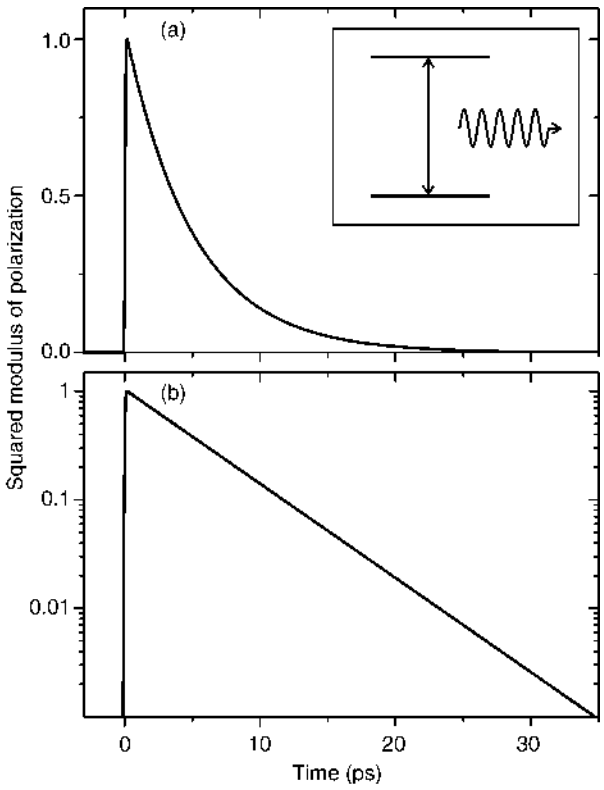

Figure 1 A two-level system is excited by a short optical pulse at $t = 0$. Due to radiative decay (inset) and possibly other dephasing mechanisms, the polarization decays exponentially as function of time with the time constant T_2, the dephasing time. The intensity of the optical field which is emitted as a result of this decay is proportional to the squared modulus of the polarization, which falls off with the time constant $T_2/2$. The decay of the squared modulus of the polarization is shown in (a) on a linear scale and in (b) on a logarithmic scale. The dephasing time was chosen to be $T_2 = 10$ ps which models the radiative decay of the exciton transition of a semiconductor quantum well.

Maxwell's equations introduce a coupling among all these resonances since the field emitted from any specific resonance interacts with all other resonances and interferes with their emitted fields. As a result, this system is characterized by N eigenmodes originating from the radiatively coupled optical resonances.

A very spectacular situation arises if one considers N identical two-level systems regularly arranged with a spacing that equals an integer multiple of $\lambda/2$, where λ is the wavelength of the system's resonance, i.e., $\lambda = c/\Delta\omega$, where c is the speed of light in the considered material (see Figure 2a). In this case all emitted fields interfere constructively and the system behaves effectively like a single two-level system with a matrix element increased by $\sqrt{N}$. Consequently, the radiative decay rate is increased by N and the polarization of the coupled system decays N-times faster than that of an isolated system (see Figure 2b). This effect is called superradiance.

It is possible to observe superradiant coupling effects, e.g., in suitably designed semiconductor heterostructures. Figure 2c compares the measured time-resolved reflection from a single quantum well (dashed line) with that of a Bragg structure, i.e., a multiple quantum-well structure which consists of 10 individual wells that are separated by $\lambda/2$ (solid line), where λ is the wavelength of the exciton resonance. For times greater than about 2 ps the direct reflection of the exciting pulse has decayed sufficiently and one is left with the exponential decay of the remaining signal. Figure 2c shows that this decay is much more rapid for the Bragg structure than for the single quantum well, due to superradiance introduced by the radiative coupling among the quantum wells.

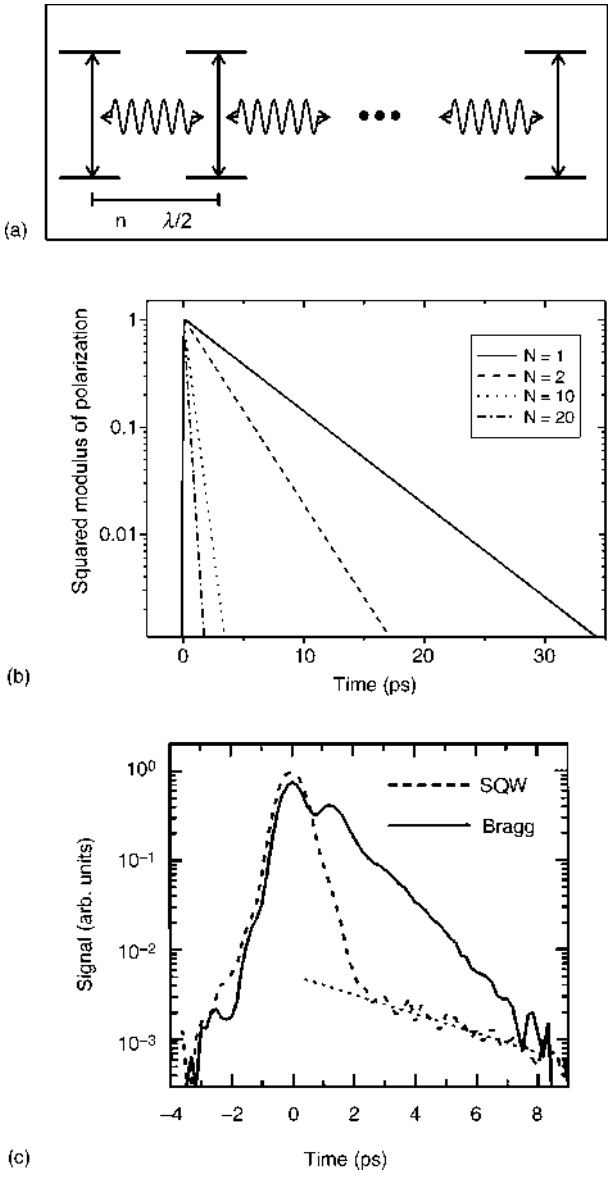

Figure 2 (a) N identical two-level systems are depicted which are regularly arranged with a spacing that equals an integer multiple of $\lambda/2$, where λ is the wavelength of the system's resonance. Due to their coupling via Maxwell's equations all fields emitted from the two-level systems interfere constructively and the coupled system behaves effectively like a single two-level system with an optical matrix element increased by $\sqrt{N}$. (b) The temporal decay of the squared modulus of the polarization is shown on a logarithmic scale. The radiative decay rate is proportional to N and thus the polarization of the coupled system decays N-times faster than that of an isolated system. This effect is called superradiance. (c) Measured time-resolved reflection of a semiconductor single quantum well (dashed line) and a $N = 10$ Bragg structure, i.e., a multi quantum well where the individual wells are separated by $\lambda/2$ (solid line), on a logarithmic scale. [Part (c) is reproduced with permission from Haas S, Stroucken T, Hübner M, *et al.* (1998) Intensity dependence of superradiant emission from radiatively coupled excitons in multiple-quantum-well Bragg structures. *Physical Review B* 57: 14860. Copyright (1998) by the American Physical Society].

Destructive Interference

As the next example we now consider a distribution of two-level systems which have slightly different transition frequencies characterized by the distribution function $g(\Delta\omega - \bar{\omega})$ of the transition frequencies which is peaked at the mean value $\bar{\omega}$ and has a spectral width of $\delta\omega$ (see Figure 3a). Ignoring the radiative coupling among the resonances, the optical polarization of the total system evolves after excitation with a short laser pulse at $t = 0$ proportional to $\int d\omega\, g(\Delta\omega - \bar{\omega}) e^{-i\Delta\omega t} \propto \tilde{g}(t) e^{-i\bar{\omega}t}$, where $\tilde{g}(t)$ denotes the Fourier transform of the frequency distribution function. $\tilde{g}(t)$ and thus the optical polarization decays on a time-scale which is inversely proportional to the spectral width of the distribution function $\delta\omega$. Thus the destructive interference of many different transition frequencies results in a rapid decay of the polarization (see Figure 3b).

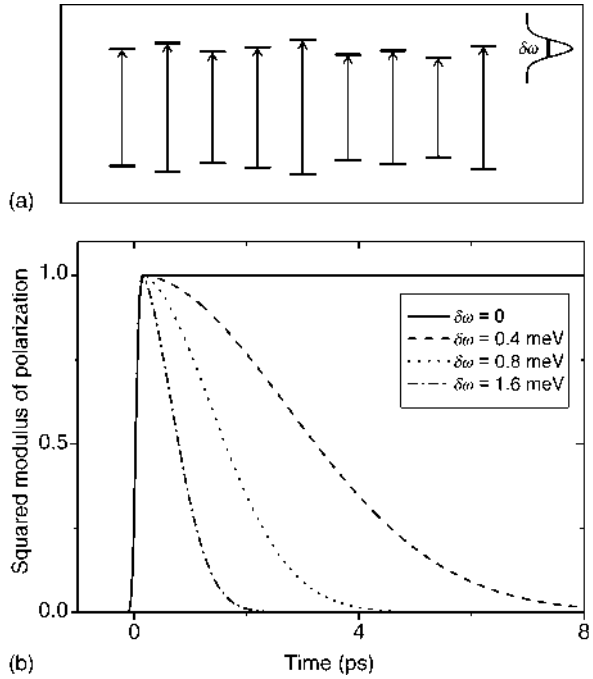

Figure 3 (a) An ensemble of two-level systems is shown where the resonance frequencies are randomly distributed according to a Gaussian distribution function of width $\delta\omega$ around an average value. (b) The temporal dynamics of the squared modulus of the polarization of ensembles of two-level systems after excitation with a short optical pulse at $t = 0$ is shown on a linear scale. Since the dephasing time is set to infinity, i.e., $T_2 \to \infty$, the polarization of an ensemble of identical two-level system ($\delta\omega = 0$) does not decay. However, for a finite width of the distribution, $\delta\omega > 0$, the individual polarizations of the ensemble oscillate with different frequencies and, therefore, due to destructive interference the polarization of the ensemble decays as function of time. Since the Fourier transform of a frequency-domain Gaussian is a Gaussian in the time domain, the dashed, dotted, and dashed dotted line have a Gaussian shape with a temporal width which is inversely proportional to $\delta\omega$.

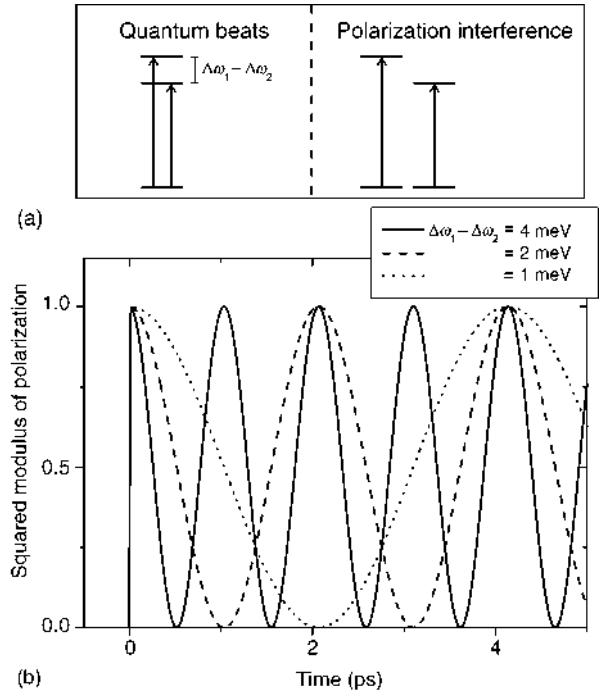

Figure 4 (a) Two optical resonances with frequency difference $\Delta\omega_1 - \Delta\omega_2$ may be realized by either a three-level system or by two uncoupled two-level systems. After impulsive excitation with an optical pulse the linear polarization of both types of systems show a modulation of the squared modulus of the polarization with the time period $2\pi/(\Delta\omega_1 - \Delta\omega_2)$. For the intrinsically coupled three-level system these modulations are called quantum beats, whereas for an interference of uncoupled systems they are named polarization interferences. Using nonlinear optical techniques, e.g., four-wave mixing and pump probe, it is possible to distinguish quantum beats and polarization interferences since coupled and uncoupled systems have different optical nonlinearities. (b) The temporal dynamics of the squared modulus of the polarization is shown for systems with two resonances and frequency difference $\Delta\omega_1 - \Delta\omega_2$, neglecting dephasing, i.e., setting $T_2 \to \infty$. After excitation with a short optical pulse at $t = 0$ the squared modulus of the polarization is periodically modulated with a time period $2\pi/(\Delta\omega_1 - \Delta\omega_2)$.

In the spectral domain this rapid decay shows up as inhomogeneous broadening. Depending on the system under investigation, there are many different sources for such an inhomogeneous broadening. One example is the Doppler broadening in atomic gases or disorder effects such as well-width fluctuations in semiconductor quantum wells or lattice imperfections in crystals.

In the nonlinear optical regime it is, under certain circumstances, possible to reverse the destructive interference of inhomogeneously broadened coherent polarizations. For example, in four-wave mixing, a second pulse may lead to a rephasing of the contributions with different frequencies, which results in the photon echo (see discussion below).

Quantum Beats

The occurrence of quantum beats can be understood most easily in a system where the total optical polarization can be attributed to a finite number of optical transitions. Let us assume for simplicity that all these transitions have the same matrix element. In this case, after excitation with a short laser pulse at $t = 0$ the optical polarization of the total system evolves proportional to $\sum_i e^{-i\Delta\omega_i t}$. The finite number of frequencies results in a temporal modulation with time periods $2\pi/(\Delta\omega_i - \Delta\omega_j)$ of the squared modulus of the polarization which is proportional to the emitted field intensity. For the case of two frequencies the squared modulus of the polarization is proportional to $[1 + \cos((\Delta\omega_1 - \Delta\omega_2)t)]$, i.e., due to the interference of two contributions the polarization varies between a maximum and zero (see Figure 4b).

In the linear optical regime it is impossible to distinguish whether the optical transitions are

uncoupled or coupled. As shown in Figure 4, the two uncoupled two-level systems give the same linear polarization as a three-level systems where the two transitions share a common state. It is, however, possible to decide about the nature of the underlying transitions if one performs nonlinear optical spectroscopy. This is due to the fact that the so-called quantum beats, i.e., the temporal modulations of the polarization of an intrinsically coupled system, show a different temporal evolution as the so-called polarization interferences, i.e., the temporal modulations of the polarization of uncoupled systems. The former ones are also much more stable in the presence of inhomogeneous broadening than the latter ones.

In semiconductor heterostructures, quantum-beat spectroscopy has been widely used to investigate the temporal dynamics of excitonic resonances. Also the coupling among different optical resonances has been explored in pump-probe and four-wave-mixing measurements.

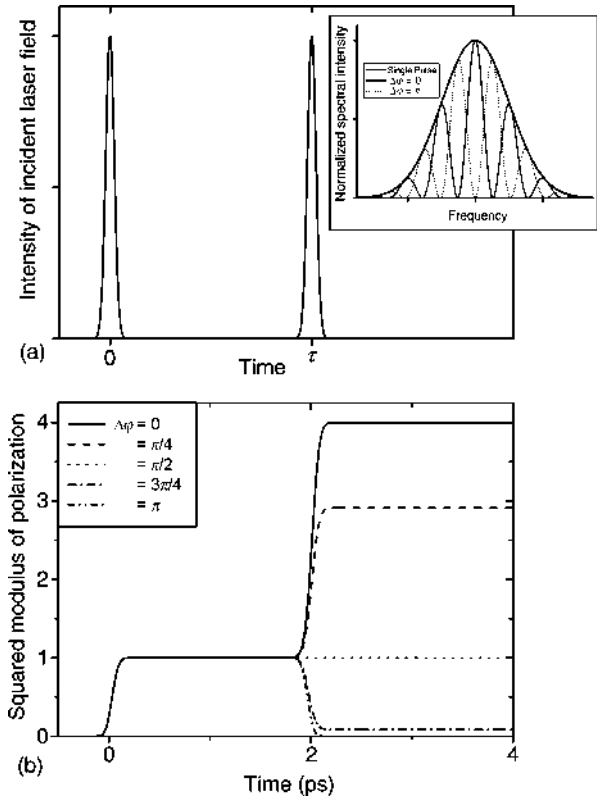

Figure 5 (a) Two Gaussian laser pulses separated by the time delay τ can lead to interferences. As shown by the thick solid line in the inset, the spectral intensity of a single pulse is a Gaussian with a maximum at the central frequency of the pulse. The width of this Gaussian is inversely proportional to the duration of the pulse. The spectral intensity of the field consisting of both pulses shows interference fringes, i.e., is modulated with a spectral period that is inversely proportional to τ. As shown by the thin solid and the dotted lines in the inset, the phase of the interference fringes depends on the phase difference $\Delta\varphi$ between the two pulses. Whereas for $\Delta\varphi = 0$ the spectral intensity has a maximum at the central frequency, it vanishes at this position for $\Delta\varphi = \pi$. (b) The temporal dynamics of the squared modulus of the polarization is shown for a two-level system excited by a pair of short laser pulses neglecting dephasing, i.e., setting $T_2 \to \infty$. The first pulse excites at $t = 0$ an optical polarization with a squared modulus normalized to 1. The second pulse, which has the same intensity as the first one, also excites an optical polarization at $t = 2$ ps. For $t > 2$ ps the total polarization is given by the sum of the polarizations induced by the two pulses. Due to interference, the squared modulus of the total polarization depends sensitively on the phase difference $\Delta\varphi$ between the two pulses. For constructive interference the squared modulus of the total polarization after the second pulse is four times bigger than that after the first pulse, whereas it vanishes for destructive interference. One may achieve all values in between these extremes by using phase differences $\Delta\varphi$ which are no multiples of π. It is thus shown that the second pulse can be used to coherently control the polarization induced by the first pulse.

Coherent Control

In the coherent regime the polarization induced by a first laser pulse can be modified by a second, delayed pulse. For example, the first short laser pulse incident at $t = 0$ induces a polarization proportional to $e^{-i\Delta\omega t}$ and a second pulse arriving at $t = \tau > 0$ induces a polarization proportional to $e^{-i\varphi}e^{-i\Delta\omega(t-\tau)}$. Thus for $t \geq \tau$ the total polarization is given by $e^{-i\Delta\omega t}(1 + e^{-i\Delta\varphi})$ with $\Delta\varphi = \varphi - \Delta\omega\tau$. If $\Delta\varphi$ is an integer multiple of 2π, the polarizations of both pulses interfere constructively and the amplitude of the resulting polarization is doubled as compared to a single pulse. If, on the other hand, $\Delta\varphi$ is an odd multiple of π, the polarizations of both pulses interfere destructively and the resulting polarization vanishes after the action of both pulses. Thus by

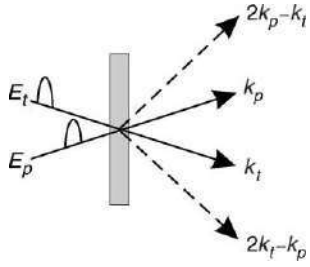

Figure 6 Transient optical nonlinearities can be investigated by exciting the system with two time delayed optical pulses. The pump pulse (E_p) puts the system in an excited state and the test (probe) pulse (E_t) is used to measure its dynamics. The dynamic nonlinear optical response may be measured in the transmitted or reflected directions of the pump and test pulses, i.e., $\pm k_p$ and $\pm k_t$, respectively. In a pump-probe experiment one often investigates the change of the test absorption induced by the pump pulse, by measuring the absorption in the direction k_t with and without E_p. One may also measure the dynamic nonlinear optical response in scattering directions like $2k_p - k_t$ and $2k_t - k_p$ as indicated by the dashed lines. This is what is done in a four-wave-mixing or photon-echo experiment.

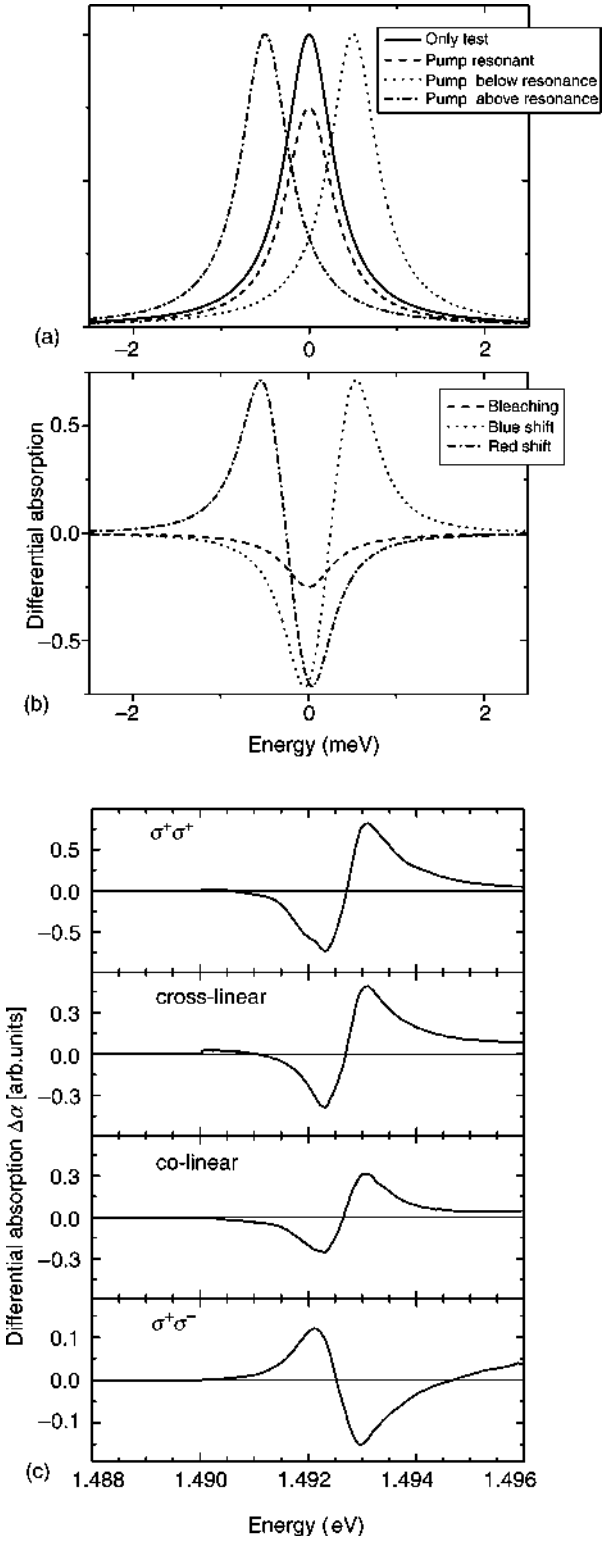

Figure 7 (a) Absorption spectra of a two-level system. The solid line shows the linear optical absorption spectrum as measured by a weak test pulse. It corresponds to a Lorentzian line that is centered at the transition frequency $\Delta\omega$, which has been set to zero. The width of the line is inversely proportional to the dephasing time T_2. In the presence of a pump pulse which resonantly excited the two-level system, the absorption monitored by the test pulse is reduced in amplitude, i.e., bleached, since the pump puts the system in an excited state (dashed line). In the optical Stark effect the frequency of the pump is nonresonant with the transition frequency. If the pump is tuned below (above) the transition frequency, the absorption is shifted to higher (lower) frequencies, i.e., shifted to the blue (red) part of the spectrum, see dotted (dashed-dotted) line. (b) The differential absorption obtained by taking the difference between the absorption in the presence of the pump and the absorption without pump. The dashed line shows the purely negative bleaching obtained using a resonant pump. The dotted and the dashed-dotted lines correspond to the dispersive shape of the blue and red shift obtained when pumping below and above the resonance, respectively. (c) Measured differential absorption spectra of a semiconductor quantum well which is pumped off resonantly 4.5 *meV* below the 1s heavy-hole exciton resonance. The four lines correspond to different polarization directions of the pump and probe pulses as indicated. [Part (c) is reproduced with permission from Sieh C, Meier T, Jahnke F, *et al.* (1999) Coulomb memory signatures in the excitonic optical Stark effect. *Physical Review Letters* 82: 3112. Copyright (1999) by the American Physical Society.]

varying the phase difference of the two pulses it is possible to coherently enhance or destroy the optical polarization (see Figure 5b).

One can easily understand the coherent control in the frequency domain by considering the overlap between the absorption of the system and the pulse spectrum. For excitation with two pulses that are temporally delayed by τ, the spectrum of the excitation shows interference fringes with a spectral oscillation period that is inversely proportional to τ. The positions of the maxima and the minima of the fringes depend on the phase difference between the two pulses. For constructive interference, the excitation spectrum is at a maximum at the resonance of the system, whereas for destructive interference the excitation spectrum vanishes at the resonance of the system (see Figure 5a).

Coherent control techniques have been applied to molecules and solids to control the dynamical evolution of electronic wavepackets and also the coupling to nuclear degrees of freedom. In this context, it is sometimes possible to steer certain chemical reactions into a preferred direction by using sequences of laser pulses which can be chirped, i.e., have a time-dependent frequency. Furthermore, in semiconductors and heterostructures, coherent control has been used to optically inject electronic currents on an ultrafast time-scale.

Transient Absorption Changes

In a typical pump-probe experiment one excites the system with a pump pulse (E_p) and probes its dynamics with a weak test pulse (E_t) (see Figure 6). With such experiments one often measures the differential absorption $\Delta\alpha(\omega)$ which is defined as

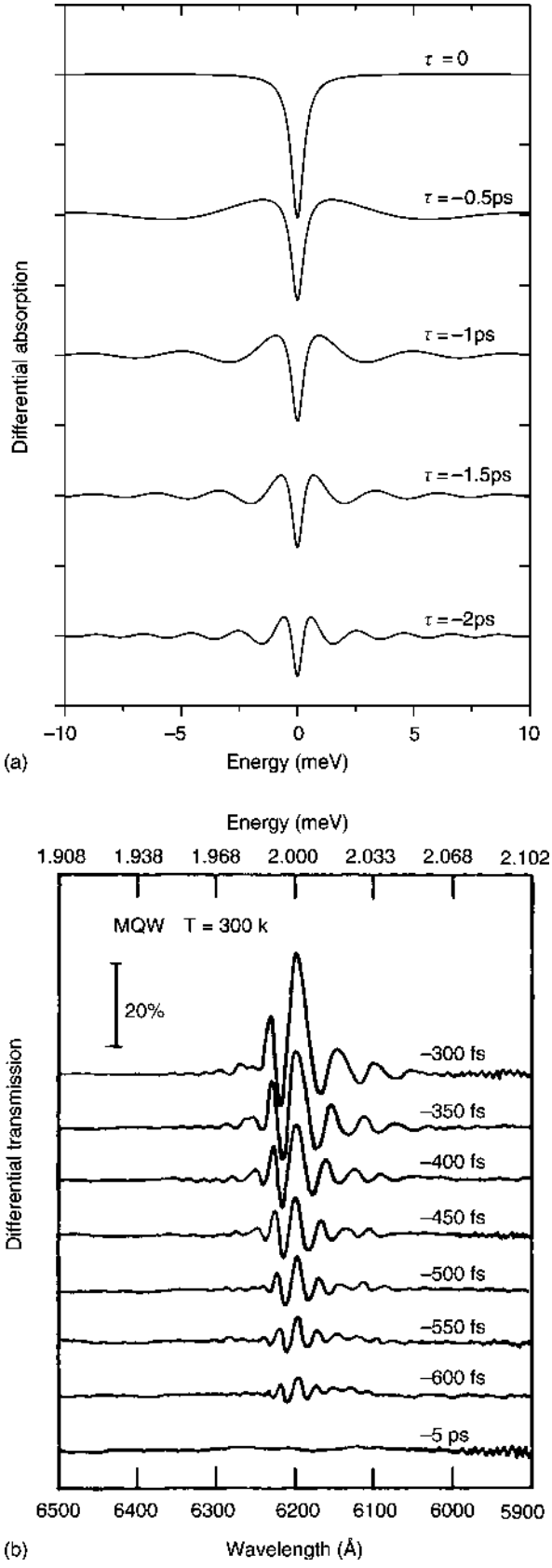

the difference between the probe absorption in the presence of the pump $\alpha_{\text{pump on}}(\omega)$ and the probe absorption without the pump $\alpha_{\text{pump off}}(\omega)$.

For resonant pumping and for a situation where the pump precedes the test (positive time delays $\tau > 0$), the absorption change is usually negative in the vicinity of the resonance frequency $\Delta\omega$, indicating the effect of absorption bleaching (see Figure 7a,b). There may be positive contributions spectrally around the original absorption line due to resonance broadening and, at other spectral positions, due to excited state absorption, i.e., optical transitions to energetically higher states which are only possible if the system is in an excited state. The bleaching and the positive contributions are generally present in coherent, and also in incoherent situations, where the polarization vanishes but occupations in excited states are present.

For detuned pumping, the resonance may be shifted by the light field as for example, in the optical Stark effect. Depending on the excitation configuration and the system, this transient shift may be to higher (blue shift) or lower energies (red shift) (see Figure 7a,b). With increasing pump-probe time delay the system gradually returns to its unexcited state and the absorption changes disappear.

As an illustration we show in Figure 7c experimentally measured differential absorption spectra of a semiconductor quantum well which is pumped spectrally below the exciton resonance. For a two-level system one would expect a blue shift of the absorption, i.e., a dispersive shape of the differential absorption with positive contributions above and negative contributions below the resonance frequency. This is indeed observed for most polarization directions of the pump and probe pulses. However, if both pulses are oppositely circularly polarized, the experiment shows a resonance shift in the reverse direction, i.e., a red shift. For the explanation of this behavior one has to consider the optical selection rules of the quantum well system. One finds that the signal should actually vanish for oppositely circularly

Figure 8 (a) Differential absorption spectra are shown for resonant pumping of a two-level system for various time delays τ between the pump and the test pulse. When the pump precedes the test, i.e., $\tau < 0$, the differential absorption exhibits spectral oscillations with a period which is inversely proportional to the time delay. When τ approaches zero, these spectral oscillations vanish and the differential absorption develops into purely negative bleaching. (b) Measured differential transmission spectra of a multi quantum well for different negative time delays as indicated. [Part (b) is reproduced with permission from Sokoloff JK, Joffre M, Fluegel B, *et al.* (1988) Transient oscillations in the vicinity of excitons and in the band of semiconductors. *Physics Review B* 38: 7615. Copyright (1988) by the American Physical Society.]

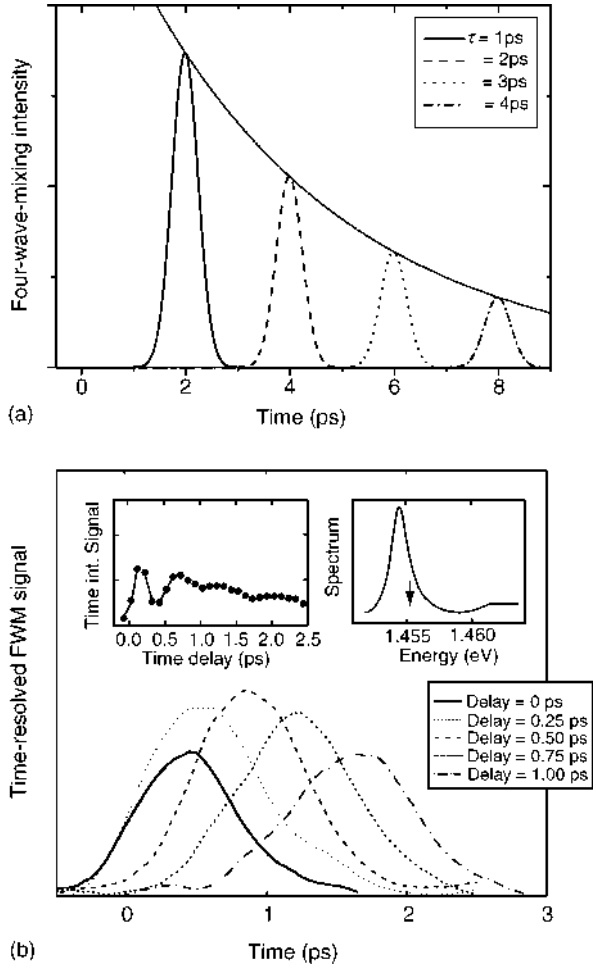

Figure 9 (a) The intensity dynamics as measured in a four-wave-mixing experiment on an ensemble of inhomogeneously broadened two-level systems for various time delays τ of the two incident pulses. The first pulse excites the system at $t = 0$. In the time period $0 < t < \tau$ the individual polarizations of the inhomogeneously broadened system oscillate with their respective frequencies and due to destructive interference the total polarization decays, cp. **Figure 3**. The second pulse arriving at $t = \tau$ leads to a partial rephasing of the individual polarizations. Due to phase conjugation all individual polarizations are in phase at $t = 2\tau$ which results in a macroscopic measurable signal, the photon echo. Due to dephasing processes the magnitude of the photon echo decays with increasing τ, see solid, dashed, dotted and dashed-dotted lines. Describing the dephasing via a T_2 time corresponds to an exponential decay as indicated by the thin dotted line. By measuring the decay of the echo with increasing τ, one thus gets experimental information on the dephasing of the optical polarization. (b) Time-resolved four-wave-mixing signal measured on an inhomogeneously broadened exciton resonance of a semiconductor quantum well structure for different time delays as indicated. The origin of the time axis starts at the arrival of the second pulse, i.e., for a non-interacting system the maximum of the echo is expected at $t = \tau$. The insets show the corresponding time-integrated four-wave-mixing signal and the linear absorption. [Part (b) is reproduced with permission from Jahnke F, Koch M, Meier T, *et al.* (1994) Simultaneous influence of disorder and Coulomb interaction on photon echoes in semiconductors. *Physics Review B* 50: 8114. Copyright (1994) by the American Physical Society.]

polarized pulses, as long as many-body correlations are neglected. Thus in this case the entire signal is due to many-body correlations and it is their dynamics which gives rise to the appearance of the red shift.

Spectral Oscillations

For negative time delays $\tau < 0$, i.e., if the test pulse precedes the pump, the absorption change $\Delta\alpha(\omega)$ is characterized by spectral oscillations around $\Delta\omega$ which vanish as τ approaches zero (see **Figure 8a**). The spectral period of the oscillations decreases with increasing $|\tau|$. **Figure 8b** shows measured differential transmission spectra of a multiple quantum-well structure for different negative time delays. As with the differential absorption, also the differential transmission spectra are dominated by spectral oscillations around the exciton resonance whose period and amplitude decrease with increasing $|\tau|$.

The physical origin of the coherent oscillations is the pump-induced perturbation of the free-induction decay of the polarization generated by the probe pulse. This perturbation is delayed by the time τ at which the pump arrives. The temporal shift of the induced polarization changes in the time domain leads to oscillations in the spectral domain, since the Fourier transformation translates a delay in one domain into a phase factor in the conjugate domain.

Photon Echo

In the nonlinear optical regime one may (partially) reverse the destructive interference of a coherent, inhomogeneously broadened polarization. For example, in four-wave mixing, which is often performed with two incident pulses, one measures the emitted field in a background-free scattering direction (see **Figure 5**). The first short laser pulse excites all transitions at $t = 0$. As a result of the inhomogeneous broadening the polarization decays due to destructive interference (see **Figure 3**). The second pulse arriving at $t = \tau > 0$ is able to conjugate the phases ($e^{i\varphi} \rightarrow e^{-i\varphi}$) of the individual polarizations of the inhomogeneously broadened system. The subsequent unperturbed dynamical evolution of the polarizations leads to a measurable macroscopic signal at $t = 2\tau$. This photon echo occurs since at this point in time all individual polarizations are in phase and add up constructively (see **Figure 9a**). Since this rephasing process leading to the echo is only possible as long as the individual polarizations remain coherent, one can analyze the loss of coherence (dephasing) by measuring the decay of the photon echo with increasing time delay.

Echo experiments are frequently performed using more complicated geometries, e.g., more than

two laser pulses. Further echo phenomena can also be observed as the result of the coupling of the photoexcitation to nuclear degrees of freedom and also due to spatial motion of the electrons, i.e., electronic currents.

Time-resolved four-wave-mixing signal, measured on an inhomogeneously broadened exciton resonance of a semiconductor quantum well structure for different time delays, is presented in Figure 9b. Compared to Figure 9a, the origin of the time axis starts at the arrival of the second pulse, i.e., for a non-interacting system the maximum of the echo is expected at $t = \tau$. Due to the inhomogeneous broadening the maximum of the signal is shifting to longer times with increasing delay. Further details of the experimental results, in particular, the exact position of the maximum and the width of the signal, cannot be explained on the basis of non-interacting two-level systems, but require the analysis of many-body effects in the presence of inhomogeneous broadening.

See also

Coherent Control: Applications in Semiconductors; Experimental; Theory. **Spectroscopy:** Nonlinear Laser Spectroscopy.

Further Reading

Allen L and Eberly JH (1975) *Optical Resonance and Two-Level Atoms*. New York: Wiley.

Baumert T, Helbing J, Gerber G, *et al.* (1997) Coherent control with femtosecond laser pulses. *Advanced Chemical Physics* 101: 47–82.

Bergmann K, Theuer H and Shore BW (1998) Coherent population transfer among quantum states of atoms and molecules. *Review of Modern Physics* 70: 1003–1025.

Bloembergen N (1965) *Nonlinear Optics*. New York: Benjamin.

Brewer RG (1977) *Coherent Optical Spectroscopy*. In: Balian, *et al.* (ed.). Course 4 in Les Houches Session XXVII, *Frontiers in Laser Spectroscopy*, vol. 1. Amsterdam: North-Holland Publ.

Cohen-Tannoudji C, Dupont-Roc J and Grynberg G (1989) *Photons and Atoms*. New York: Wiley.

Haug H and Koch SW (2004) *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 4th edn. Singapore: World Scientific.

Koch SW, Peyghambarian N and Lindberg M (1988) Transient and steady-state optical nonlinearities in semiconductors. *Journal of Physics* C 21: 5229–5250.

Macomber JD (1976) *The Dynamics of Spectroscopic Transitions*. New York: Wiley.

Mukamel S (1995) *Principles of Nonlinear Optical Spectroscopy*. New York: Oxford.

Peyghambarian N, Koch SW and Mysyrowicz A (1993) *Introduction to Semiconductor Optics*. Englewood Cliffs, NJ: Prentice-Hall.

Pötz W and Schroeder WA (eds) (1999) *Coherent Control in Atoms, Molecules, and Semiconductors*. Dordrecht: Kluwer.

Schäfer W and Wegener M (2002) *Semiconductor Optics and Transport Phenomena*. Berlin: Springer.

Shah J (1999) *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*, 2nd edn. Berlin: Springer.

Yeazell JA and Uzer T (eds) (2000) *The Physics and Chemistry of Wave Packets*. New York: Wiley.

Ultrafast Studies of Semiconductors

J Shah, Arlington, VA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Linear optical spectroscopy, using the techniques of absorption, transmission, reflection and light scattering, has provided invaluable information about the electronic and vibrational properties of atoms, molecules, and solids. Optical techniques also possess some additional unique strengths: the ability to

1. generate nonequilibrium distributions functions of electrons, holes, excitons, phonons, etc. in solids;
2. determine the distribution functions by optical spectroscopy;
3. determine the nonequilibrium distribution functions on femtosecond timescales;
4. determine femtosecond dynamics of carrier and exciton transport and tunneling; and
5. investigate interactions between various elementary excitations as well as many-body processes in semiconductors.

Researchers have exploited these unique strengths to gain fundamental new insights into nonequilibrium, nonlinear, and transport physics of semiconductors and their nanostructures over the past four decades.

A major focus of these efforts has been devoted to understanding how a semiconductor in

thermodynamic equilibrium, excited by an ultrashort optical pulse, returns to the state of thermodynamic equilibrium. Four distinct regimes can be identified:

(a) the coherent regime in which a well-defined phase relationship exists between the excitation created by the optical pulse and the electromagnetic (optical) field creating it;
(b) the nonthermal regime in which the coherence (well-defined phase relationships) has been destroyed by various collision and interference processes but the distribution function of excitations is nonthermal (i.e., cannot be described by a Maxwell–Boltzmann distribution or its quantum (degenerate) counterparts);
(c) the hot carrier regime in which the distributions for various excitations are thermal, but with different characteristic temperatures for different excitations and thermal bath; and
(d) the isothermal regime in which the excitations and the thermal bath are at the same temperature but there is an excess of excitation (e.g., electron–hole pairs) compared to thermodynamic equilibrium.

Various physical processes take the semiconductor from one regime to the next, and provide information not only about the fundamental physics of semiconductors but also about the physics and ultimate performance limits of electronic, optoelectronic, and photonic devices.

In addition to coherent and relaxation dynamics, ultrafast studies of semiconductors provide new insights into tunneling and transport of carriers, and demonstrate novel quantum mechanical phenomena and coherent control in semiconductors.

The techniques used for such studies have also advanced considerably over the last four decades. Ultrafast lasers with pulse widths corresponding to only a few optical cycles (1 optical cycle ≈ 2.7 fs at 800 nm) have been developed. The response of the semiconductor to an ultrafast pulse, or a multiple of phase-locked pulses, can be investigated by measuring the dynamics of light emitted by the semiconductor using a streak camera (sub-ps time resolution) or a nonlinear technique such as luminescence up-conversion (time resolution determined by the laser pulse width for an appropriate nonlinear crystal). The response of the semiconductor can also be investigated using a number of two or three beam pump-probe techniques such as transmission, reflection, light scattering or four-wave mixing (FWM). Both the amplitude and the phase of the emitted radiation or the probe can be measured by phase-sensitive techniques such as spectral interferometry. The lateral transport of excitation can be investigated by using time-resolved spatial imaging and the vertical transport and tunneling of carriers can be investigated using the technique of 'optical markers'. Electro-optic sampling can be used for transmission/reflection studies or for measuring THz response of semiconductors. More details on these techniques, and indeed many topics discussed in this brief article, can be found in the Further Reading.

Coherent Dynamics

The coherent response of atoms and molecules is generally analyzed for an ensemble of independent (noninteracting) two-level systems. The statistical properties of the ensemble are described in terms of the density matrix operator whose diagonal components relate to population of the eigenstates and off-diagonal components relate to coherence of the superposition state. The time evolution of the density matrix is governed by the Liouville variant of the Schrödinger equation $i\hbar\dot{\rho} = [H, \rho]$ where the system Hamiltonian $H = H_0 + H_{\text{int}} + H_R$ is the sum of the unperturbed, interaction (between the radiation field and the two-level system), and relaxation Hamiltonians, respectively. Using a number of different assumptions and approximations, this equation of motion is transformed into the optical Bloch equations (OBE), which are then used to predict the coherent response of the system, often using iterative procedures, and analyze experimental results.

Semiconductors are considerably more complex. Coulomb interaction profoundly modifies the response of semiconductors, not only in the linear regime (e.g., strong exciton resonance in the absorption spectrum) but also in the coherent and nonlinear regime. The influence of Coulomb interaction may be introduced by renormalizing the electron and hole energies (i.e., introducing excitons), and by renormalizing the field–matter interaction strength by introducing a renormalized Rabi frequency. These changes and the relaxation time approximation for the relaxation Hamiltonian lead to an analog of the optical Bloch equations, the semiconductor Bloch equations which have been very successful in analyzing the coherent response of semiconductors.

In spite of this success, it must be stressed that the relaxation time approximation, and the Boltzmann kinetic approach on which it is based, are not valid under all conditions. The Boltzmann kinetic approach is based on the assumption that the duration of the collision is much shorter than the interval between the collisions; i.e., the collisions are instantaneous. In the non-Markovian regime where these assumptions are not valid, each collision does not strictly

conserve the energy and momentum and the quantum kinetic approach becomes more appropriate. This is true for photo-excited semiconductors as well as for the quantum transport regime in semiconductors. Also, semiconductor Bloch equations have been further extended by including four-particle correlations. Finally, the most general approach to the description of the nonequilibrium and coherent response of semiconductors following excitation by an ultrashort laser pulse is based on nonequilibrium Green's functions.

Experimental studies on the coherent response of semiconductors can be broadly divided into three categories:

1. investigation of novel aspects of coherent response not found in other simpler systems such as atoms;
2. investigation of how the initial coherence is lost, to gain insights into dephasing and decoherence processes; and
3. coherent control of processes in semiconductors using phase-locked pulses.

Numerous elegant studies have been reported. This section presents some observations on the trends in this field.

Historically, the initial experiments focused on the decay of the coherent response (time-integrated FWM as a function of delays between the pulses), analyzed them in terms of the independent two-level model, and obtained very useful information about various collision processes and rates (exciton–exciton, exciton–carrier, carrier–carrier, exciton–phonon, etc.). It was soon realized, however, that the noninteracting two-level model is not appropriate for semiconductors because of the strong influence of Coulomb interactions. Elegant techniques were then developed to explore the nature of coherent response of semiconductors. These included investigation of time- and spectrally resolved coherent response, and of both the amplitude and phase of the response to complement intensity measurements. These studies provided fundamental new insights into the nature of semiconductors and many-body processes and interactions in semiconductors. Many of these observations were well explained by the semiconductor Bloch equations. When lasers with <10 fs pulse widths became laboratory tools, the dynamics was explored on a time-scale much shorter than the characteristic phonon oscillation period (≈ 115 fs for GaAs longitudinal optical phonons), the plasma frequency (≈ 150 fs for a carrier density of $5 \times 10^{17}\,\mathrm{cm}^{-3}$), and the typical electron–phonon collision interval (≈ 200 fs in GaAs). Some remarkable features of quantum kinetics, such as reversal of the electron–phonon collision, memory effects, and energy nonconservation, were demonstrated. More recent experiments have demonstrated the influence of four-particle correlations on coherent nonlinear response of semiconductors such as GaAs in the presence of a strong magnetic field.

The ability to generate phase-locked pulses with controllably variable separation between them provides an exciting opportunity to manipulate a variety of excitations and processes within a semiconductor. If the separation between the two phase-locked pulses is adjusted to be less than the dephasing time of the system under investigation, then the amplitudes of the excitations produced by the two phase-locked pulses can interfere either constructively or destructively depending on the relative phase of the carrier waves in the two pulses. The nature of interference changes as the second pulse is delayed over an optical period. A number of elegant experiments have demonstrated coherent control of exciton population, spin orientation, resonant emission, and electrical current. Phase-sensitive detection of the linear or nonlinear emission provides additional insights into different aspects of semiconductor physics.

These experimental and theoretical studies of ultrafast coherent response of semiconductors have led to fundamental new insights into the physics of semiconductors and their coherence properties. Discussion of novel coherent phenomena is given in a subsequent section. Coherent spectroscopy of semiconductors continues to be a vibrant research field.

Incoherent Relaxation Dynamics

The qualitative picture that emerges from investigation of coherent dynamics can be summarized as follows. Consider a direct gap semiconductor excited by an ultrashort laser pulse of duration τ_L (with a spectral width $h\Delta\nu_L$) centered at photon energy $h\nu_L$ larger than the semiconductor bandgap E_g. At the beginning of the pulse the semiconductor does not know the pulse duration so that coherent polarization is created over a spectral region much larger than $h\Delta\nu_L$. If there are no phase-destroying events during the pulse (dephasing rate$1/\tau_D \ll 1/\tau_L$), then a destructive interference destroys the coherent polarization away from $h\nu_L$ with increasing time during the excitation pulse. Thus, the coherent polarization exists only over the spectral region $h\Delta\nu_L$ at the end of the pulse. This coherence will be eventually destroyed by collisions and the semiconductor will be left in an incoherent (off-diagonal elements of the density matrix are 0) nonequilibrium state with peaked electron and hole population distributions whose central energies and energy widths are

determined by $h\nu_L$, $h\Delta\nu_L$ and E_g, if the energy and momentum relaxation rates $(1/\tau_E, 1/\tau_M)$ are much smaller than the dephasing rates.

The simplifying assumptions that neatly separate the excitation, dephasing and relaxation regimes are obviously not realistic. A combination of all these (and some other) physical parameters will determine the state of a realistic system at the end of the exciting pulse. However, after times $\sim \tau_D$ following the laser pulse, the semiconductor can be described in terms of electron and hole populations whose distribution functions are not Maxwell–Boltzmann or Fermi–Direct type, but nonthermal in a large majority of cases in which the energy and momentum relaxation rates are much smaller than the dephasing rates. This section discusses the dynamics of this incoherent state.

Extensive theoretical work has been devoted to quantitatively understand these initial athermal distributions and their further temporal evolution. These include Monte Carlo simulations based on the Boltzmann kinetic equation approach (typically appropriate for time-scales longer than $\sim$100 fs, the carrier–phonon interaction time) as well as the quantum kinetic approach (for times typically less than $\sim$100 fs) discussed above. We present below some observations on this vast field of research.

Nonthermal Regime

A large number of physical processes determines the evolution of the nonthermal electron and hole populations generated by the optical pulse. These include electron–electron, hole–hole, and electron–hole collisions, including plasma effects if the density is high enough, intervalley scattering in the conduction band and intervalence band scattering in valence bands, intersub-band scattering in quantum wells, electron–phonon, and hole–phonon scattering processes. The complexity of the problem is evident when one considers that many of these processes occur on the same time-scale. Of these myriad processes, only carrier–phonon interactions and electron–hole recombinations (generally much slower) can alter the total energy in electronic systems. Under typical experimental conditions, the redistribution of energy takes place before substantial transfer of energy to the phonon system. Thus, the nonthermal distributions first become thermal distributions with the same total energy. Investigation of the dynamics of how the nonthermal distribution evolves into a thermal distribution provides valuable information about the nature of various scattering processes (other than carrier–phonon scattering) described above.

This discussion of separating the processes that conserve energy in the electronic system and those that transfer energy to other systems is obviously too simplistic and depends strongly on the nature of the problem one is considering. The challenge for the experimentalist is to devise experiments that can isolate these phenomena so each can be studied separately. If the laser energy is such that $h\nu_L - E_g < \hbar\omega_{LO}$, the optical phonon energy, then majority of the photo-excited electrons and holes do not have sufficient energy to emit an optical phonon, the fastest of the carrier–phonon interaction processes. The initial nonthermal distribution is then modified primarily by processes other than phonon scattering and can be studied experimentally without the strong influence of the carrier–phonon interactions. Such experiments have indeed been performed both in bulk and quantum well semiconductors. These experiments have exhibited spectral hole burning in the pump-probe transmission spectra and thus demonstrated that the initial carrier distributions are indeed nonthermal and evolve to thermal distributions. Such experiments have provided quantitative information about various carrier–carrier scattering rates as a function of carrier density. In addition, experiments with $h\nu_L - E_g > \hbar\omega_{LO}$ and $h\nu_L - E_g >$ intervalley separation have provided valuable information about intravalley as well as intervalley electron–phonon scattering rates. Experiments have been performed in a variety of different semiconductors, including bulk semiconductors and undoped and modulation-doped quantum wells. The latter provide insights into intersub-band scattering processes and quantitative information about the rates.

These measurements have shown that under typical experimental conditions, the initial nonthermal distribution evolves into a thermal distribution in times $<$ or $\sim$1 ps. However, the characteristic temperatures of the thermal distributions can be different for electrons and holes. Furthermore, different phonons may also have different characteristic temperatures, and may even be nonthermal even when the electronic distributions are thermal. This is the hot carrier regime that is discussed in the next subsection.

Hot Carrier Regime

The hot carriers, with characteristic temperatures T_c (T_e and T_h for electrons and holes, respectively), have high energy tails in the distribution functions that extend several kT_c higher than the respective Fermi energies. Some of these carriers have sufficient energy to emit an optical phonon. Since the carrier–optical phonon interaction rates and the phonon energies are

rather large, this may be the most effective energy loss mechanism in many cases even though the number of such carriers is rather small. The emitted optical phonons have relatively small wavevectors and nonequilibrium populations larger than expected for the lattice temperature. These optical phonons are often referred to as hot phonons although nonequilibrium phonons may be a better term since they probably have nonthermal distributions. The dynamics of hot phonons will be discussed in the next subsection, but we mention here that in case of a large population of hot phonons, one must consider not only phonon emission but also phonon absorption. Acoustic phonons come into the picture at lower temperatures when the fraction of carriers that can emit optical phonons is extremely small or negligible, and also in the case of quantum wells with sub-band energy separation smaller than optical phonons.

The dynamics of hot carriers have been studied extensively to obtain a deeper understanding of carrier–phonon interactions and to obtain quantitative measures of the carrier–phonon interaction rates. Time resolved luminescence and pump-probe transmission spectroscopy are the primary tools used for such studies. For semiconductors like GaAs, the results indicate that polar optical phonon scattering (longitudinal optical phonons) dominates for electrons whereas polar and nonpolar optical phonon scattering contribute for holes. Electron–hole scattering is sufficiently strong in most cases to maintain a common electron and hole temperature for times $>\sim 1$ ps. Since many of these experiments are performed at relatively high densities, they provide important information about many-body effects such as screening. Comparison of bulk and quasi-two-dimensional semiconductors (quantum wells) has been a subject of considerable interest. The consensus appears to be that, in spite of significant differences in the nature of electronic states and phonons, similar processes with similar scattering rates dominate both types of semiconductors. These studies reveal that the energy loss rates are influenced by many factors such as the Pauli exclusion principle, degenerate (Fermi–Dirac) statistics, hot phonons, and screening and many-body aspects.

Hot carrier distribution can be generated not only by photo-excitation, but also by applying an electric field to a semiconductor. Although the process of creating the hot distributions is different, the processes by which such hot carriers cool to the lattice temperature are the same in two cases. Therefore, understanding obtained through one technique can be applied to the other case. In particular, the physical insights obtained from the optical excitation case can be extremely valuable for many electronic devices that operate at high electric fields, thus in the regime of hot carriers. Another important aspect is that electrons and holes can be investigated separately by using different doping. Furthermore, the energy loss rates can be determined quantitatively because the energy transferred from the electric field to the electrons or holes can be accurately determined by electrical measurements.

Isothermal Regime

Following the processes discussed above, the excitations in the semiconductor reach equilibrium with each other and the thermal bath. Recombination processes then return the semiconductor to its thermodynamic equilibrium. Radiative recombination typically occurs over a longer time-scale but there are some notable exceptions such as radiative recombination of excitons in quantum wells occurring on picosecond time-scales.

Hot Phonons

As discussed above, a large population of nonequilibrium optical phonons is created if a large density of hot carriers has sufficient energy to emit optical phonons. Understanding the dynamics of these nonequilibrium optical phonons is of intrinsic interest.

At low lattice temperatures, the equilibrium population of optical phonons is insignificant. The optical phonons created as a result of photo-excitation occupy a relatively narrow region of wavevector (k) space near $k = 0$ ($\sim 1\%$ of the Brillouin zone). This nonequilibrium phonon distribution can spread over a larger k-space by various processes, or anharmonically decay into multiple large-wavevector acoustic phonons. These acoustic phonons then scatter or decay into small-wavevector, low-energy acoustic phonons that are eventually thermalized.

Pump-probe Raman scattering provides the best means of investigating the dynamics of nonequilibrium optical phonons. Such studies, for bulk and quantum well semiconductors, have provided invaluable information about femtosecond dynamics of phonons. In particular, they have provided the rate at which the nonequilibrium phonons decay. However, this information is for one very small value of the phonon wavevector so the phonon distribution function within the optical phonon branch is not investigated. The large-wavevector acoustic phonons are even less accessible to experimental techniques. Measurements of thermal phonon propagation provide some information on this subject. Finally, the nature of phonons is considerably more complex in

quantum wells, and some interesting results have been obtained on the dynamics of these phonons.

Tunneling and Transport Dynamics

Ultrafast studies have also made important contributions to tunneling and transport dynamics in semiconductors. Time-dependent imaging of the luminescing region of a sample excited by an ultrashort pulse provides information about spatial transport of carriers. Such a technique was used, for example, to demonstrate negative absolute mobility of electrons in p-modulation-doped quantum wells. The availability of near-field microscopes enhances the resolution of such techniques to measure lateral transport to the subwavelength case.

A different technique of 'optical markers' has been developed to investigate the dynamics of transport and tunneling in a direction perpendicular to the surface (vertical transport). This technique relies heavily on fabrication of semiconductor nanostructures with desired properties. The basic idea is to fabricate a sample in which different spatial regions have different spectral signatures. Thus if the carriers are created in one spatial region and are transported to another spatial region under the influence of an applied electric field, diffusion, or other processes, the transmission, reflection, and/or luminescence spectra of the sample change dynamically as the carriers are transported.

This technique has provided new insights into the physics of transport and tunneling. Investigation of perpendicular transport in graded-gap superlattices showed remarkable time-dependent changes in the spectra, and analysis of the results provided new insight into transport in a semiconductor of intermediate disorder. Dynamics of carrier capture in quantum wells from the barriers provided information not only about the fundamentals of capture dynamics, but also about how such dynamics affects the performance of semiconductor lasers with quantum well active regions.

The technique of optical markers has been applied successfully to investigate tunneling between the two quantum wells in an a-DQW (asymmetric double quantum well) structure in which two quantum wells of different widths (and hence different energy levels) are separated by a barrier. The separation between the energy levels of the system can be varied by applying an electric field perpendicular to the wells. In the absence of resonance between any energy levels, the wavefunction for a given energy level is localized in one well or the other. In this nonresonant case, it is possible to generate carriers in a selected quantum well by proper optical excitation. Dynamics of transfer of carriers to the other quantum well by tunneling can then be determined by measuring dynamic changes in the spectra. Such measurements have provided much insight into the nature and rates of tunneling, and demonstrated phonon resonances. Resonant tunneling, i.e., tunneling when two electronic levels are in resonance and are split due to strong coupling, has also been investigated extensively for both electrons and holes. One interesting insight obtained from such studies is that tunneling and relaxation must be considered in a unified framework, and not as sequential events, to explain the observations.

Novel Coherent Phenomena

Ultrafast spectroscopy of semiconductors has provided valuable insights into many other areas. We conclude this article by discussing some examples of how such techniques have demonstrated novel physical phenomena.

The first example once again considers an a-DQW biased in such a way that the lowest electron energy level in the wide quantum well is brought into resonance with the first excited electron level in the narrow quantum well. The corresponding hole energy levels are not in resonance. By choosing the optical excitation photon energy appropriately, the hole level is excited only in the wide well and the resonant electronic levels are excited in a linear superposition state such that the electrons at $t = 0$ also occupy only the wide quantum well. Since the two electron eigenstates have slightly different energies, their temporal evolutions are different with the result that the electron wavepacket oscillates back and forth between the two quantum wells in the absence of damping. The period of oscillation is determined by the splitting between the two resonant levels and can be controlled by an external electric field. Coherent oscillations of electronic wavepackets have indeed been experimentally observed by using the coherent nonlinear technique of four-wave mixing. Semiconductor quantum wells provide an excellent flexible system for such investigations.

Another example is the observation of Bloch oscillations in semiconductor superlattices. In 1928 Bloch demonstrated theoretically that an electron wavepacket composed of a superposition of states from a single energy band in a solid undergoes a periodic oscillation in energy and momentum space under certain conditions. An experimental demonstration of Bloch oscillations became possible only recently by applying ultrafast optical techniques to semiconductor superlattices. The large period of a superlattice (compared to the atomic period in a

solid) makes it possible to satisfy the assumptions underlying Bloch's prediction. The experimental demonstration of Bloch oscillations was also performed using four-wave mixing techniques. This provides another prime example of the power of ultrafast optical studies.

Summary

Ultrafast spectroscopy of semiconductors and their nanostructures is an exciting field of research that has provided fundamental insights into important physical processes in semiconductors. This article has attempted to convey the breadth of this field and the diversity of physical processes addressed by this field. Many exciting developments in the field have been omitted out of necessity. The author hopes that this brief article inspires the readers to explore some of the Further Reading.

See also

Coherence: Coherence and Imaging. **Coherence Control:** Experimental; Theory. **Coherent Transients:** Ultrafast Studies of Semiconductors. **Nonlinear Optics, Basics:** Four-Wave Mixing. **Scattering:** Scattering Theory.

Further Reading

Allen L and Eberly JH (1975) *Optical Resonance and Two-Level Atoms*. New York: Wiley Interscience.

Chemla DS (1999) Chapter 3. In: Garmire E and Kost AR (eds) *Nonlinear Optics in Semiconductors*, pp. 175–256. New York: Academic Press.

Chemla DS and Shah J (2001) Many-body and correlation effects in semiconductors. *Nature* 411: 549–557.

Haug H (2001) Quantum kinetics for femtosecond spectroscopy in semiconductors. In: Willardson RK and Weber ER (eds) *Semiconductors and Semimetals*, vol. 67, pp. 205–229. London: Academic Press.

Haug H and Jauho AP (1996) *Quantum Kinetics in Transport and Optics of Semiconductors*. Berlin: Springer-Verlag.

Haug H and Koch SW (1993) *Quantum Theory of the Optical and Electronic Properties of Semiconductors*. Singapore: World Scientific.

Kash JA and Tsang JC (1992) Chapter 3. In: Shank CV and Zakharchenya BP (eds) *Spectroscopy of Nonequilibrium Electrons and Phonons*, pp. 113–168. Amsterdam: Elsevier.

Mukamel S (1995) *Principles of Nonlinear Optical Spectroscopy*. New York: Oxford University Press.

Shah J (1992) *Hot Carriers in Semiconductor Nanostructures: Physics and Applications*. Boston, MA: Academic Press.

Shah J (1999) *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*. 2nd edn. Heidelberg, Germany: Springer-Verlag.

Steel DG, Wang H, Jiang M, Ferrio KB and Cundiff ST (1994) In: Phillips RT (ed.) *Coherent Optical Interactions in Solids*, pp. 157–180. New York: Plenum Press.

Wegener M and Schaefer W (2002) *Semiconductor Optics and Transport*. Heidelberg, Germany: Springer-Verlag.

Zimmermann R (1988) *Many-Particle Theory of Highly Excited Semiconductors*. Leipzig, Germany: Teubner.

COLOR AND THE WORLD

G N Rao, Adelphi University, Garden City, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

The human visual system is capable of detecting a small part of the electromagnetic spectrum which is called light. The visible region has a typical wavelength range of about 380–780 nm. Under optimal conditions the range is 360–830 nm. The most important point is that color is a perception. It is how human beings perceive the visual part of the electromagnetic spectrum. The color of an object perceived depends on the physical characteristics of the radiation illuminating the object, spectral scattering and reflectivity of the object, physiology of the eye, photochemistry of the sensors in the retina, the complex data processing that takes place in the visual system, and the psychology of color perception. In this review, we present an overview of color in nature and the world. Initially, color and wavelength is presented, then the human visual system and color perception. Then we go on to discuss measuring color and colorimetry and then color in nature, followed by a brief account of color in art and holography and the structural colors in nature.

Color and Wavelength (Spectral and Nonspectral Colors)

The ingenious yet simple experiments performed by Isaac Newton in the year 1666 and his ideas presented

in his famous treatise 'Opticks', are the foundations of our understanding of color and color perception. Based on his experiments, Newton showed that different regions of the visible spectrum are perceived as different colors and the solar radiation (white light) consists of spectral colors violet, blue, cyan, green, yellow, orange, and red. Thus, natural day light is quite colorful though we perceive it as colorless. Table 1 gives approximate wavelength ranges of the spectral colors. Superposition of all the spectral colors results in a perception of white.

Colors can be produced in a number of ways:

1. *Gas discharges* emit characteristic radiation in the visible region. For example, neon discharge is red in color, argon discharge blue, sodium discharge yellow, helium-neon laser gives red radiation and argon ion laser blue-green radiation. A variety of fluorescent lamps are available and they provide continuous spectral power distribution with characteristic lines of mercury in the blue-green region. Gas-filled sodium and fluorescent lamps offer long life and high efficiency and are popular for indoor and outdoor lighting.
2. *Transition metal* (Cr, Mn, Fe, Co, Ni, Cu) compounds or transition metal impurities are responsible for the colors of the minerals, paints, gems, and pigments. They involve inorganic compounds of transition metal or rare earth ions with unpaired d or f orbitals. About 1% chromium sesquioxide (Cr_2O_3) in colorless aluminum oxide results in beautiful red ruby, a gemstone. The Cr^{3+} impurity in pure beryllium aluminum silicate gives a beautiful emerald green resulting in sapphire, another gemstone. Transition metal impurities are added to glass in the molten state to prepare colored glass. For example, adding Cr^{3+} gives green, Mn^{3+} purple, Fe^{3+} pale yellow, Co^{3+} reddish blue, Ni^{2+} brown, and Cu^{2+} greenish blue color.
3. *Color in Organic Compounds* – Most natural and synthetic dyes and biological (vegetable and animal origin) colorants are complicated organic molecules. They absorb light, exciting molecular orbitals. The dyes are used for food coloration, clothing, photographic, and sensitizer purposes.
4. *Colors of metals, semiconductor materials and color centers* – The colors of these materials are caused by the electronic transitions involving the energy bands.

   Diode lasers emit radiation corresponding to the bandgap of the semiconductor materials used for their fabrication. They are inexpensive, compact, and can be operated using dry cell batteries. They are extensively used in compact disks, barcode readers, and optical communications.

   Light emitting diodes (LED) are also prepared using semiconductor materials and they emit radiation corresponding to the bandgap of the semiconductor materials used for their fabrication. LEDs are available in a variety of colors, red (GaAsP), orange (AlInGaP), yellow (InGaAs), green (CdSSe), blue (CdZnS) and are widely used in clocks, toys, tuners, displays, electronic billboards, and appliances.

   Color centers may be produced by irradiating some alkali halides and glass materials with electromagnetic radiation (gamma rays, X-rays, ultraviolet, and visible radiation) or with charged or uncharged particles (electrons, protons, neutrons). The irradiation produces an electron hole pair and the electron is trapped forming an electron center. The electron or hole, or both, form a color center absorbing part of the visible radiation. The color centers can be reversed by heating to high temperatures.
5. *Optical Phenomenon* lead to spectacular colors in nature and in biological materials. Scattering of light by the atmosphere is responsible for colors of the sky, beautiful sunsets and sunrise, twilight, blue moon, and urban glows. Dispersion and polarization cause rainbows and halos. Interference gives rise to the beautiful colors of thin films on water, foam, bubbles, and some biological colors (butterflies). Diffraction is responsible for the colors of liquid crystals, coronae, color of gems (opal), color of animals (sea mouse), the colors of butterfly (*Morpho rhetenor*), and diffraction gratings.

Table 1 Approximate wavelength ranges of different spectral colors

Color	*Wavelength range (nm)*
Violet	380–435
Blue	435–500
Cyan	500–520
Green	520–565
Yellow	565–597
Orange	597–625
Red	625–780

The Human Visual System and Color Perception

Photoreceptors in the Eye: Rods and Cones

As light enters the eye, the cornea and the eye lens focus the image on the retina. The retina has two kinds of photoreceptors: rods and cones so named

because of their shapes. Typically, there are about 7 million cones and 110–125 million rods. Cones are activated during photopic (daylight) conditions and the rods during scotopic (night) conditions. Cones are responsible for color perception whereas rods can perceive only white and gray shades. Cones are predominantly located near the fovea, the central region of the retina. When light falls on the rods and the cones, the photosensitive material in them, namely rhodopsin, is activated, generating an electrical signal which is further processed in the visual system and transmitted to the visual cortex. Considerable data processing takes place within the retina.

Color Perception

Subjective nature of color

Color is a perception and is subjective. There is no way one can define color in an absolute or quantitative means, even though it can be precisely characterized by the spectral power distribution (SPD) of the radiation. For monochromatic light of different wavelengths we see the spectral colors detailed in Table 1. In general, the SPD of light from an object is complex and to predict the exact color of such a spectrum is quite difficult. However, in the following we discuss some simple color schemes.

Additive color mixing

Newton's experiments demonstrated that using the three primary color light sources, blue, green, and red and by varying their relative intensities, one can generate most of the colors by additive color mixing. Additive color mixing is employed in stage lighting, and large screen color television projection systems. If we are concerned only with the hue and ignore saturation and brightness, the superposition of colored light beams of equal brightness on a white screen gives the following simple additive color scheme:

$$\text{Green} + \text{Red} \Rightarrow \text{Yellow}$$

$$\text{Blue} + \text{Red} \Rightarrow \text{Magenta}$$

$$\text{Blue} + \text{Green} \Rightarrow \text{Cyan}$$

$$\text{Blue} + \text{Green} + \text{Red} \Rightarrow \text{White}$$

Subtractive color mixing

When white light passes through, say, a red filter, the radiation corresponding to the red radiation is transmitted and all other parts of the visible spectrum are absorbed. If we insert more red filters the brightness of the light transmitted becomes less though the hue is likely to be more saturated. If you insert red and blue filters together in a white beam, you will see no light transmitted because the red filter blocks the entire spectrum except the radiation corresponding to red color which in turn is blocked by the blue filter.

Subtractive color scheme is also applicable when paints of various colors are mixed together. The subtractive primary colors are magenta, yellow, and cyan. The artists often call them red, yellow, and blue. By appropriately mixing the primary colored paints one can, in principle, generate most of the colors.

How many colors can we differentiate?

If we do a simple experiment beginning from the lowest wavelength of the visible spectrum, say 380 nm, and gradually move to the long wavelength end of the visible spectrum, one can differentiate about 150 spectral colors. For each spectral color there will be a number of colors that vary in brightness and saturation. Additionally, we can have additive mixtures of these spectral colors, which in turn have variations in brightness and saturation. It is estimated that under optimal conditions we can differentiate as many as 7 to10 million different colors and shades.

Temporal Response of the Human Visual System

Positive afterimages

The visual system responds to changes in the stimulation in time. However, the response of the visual system is both delayed and persists for some time after the stimulation is off. Because of the persistence of vision, one would see positive afterimages. The afterimages last about 1/20 s under normal conditions but are shorter at high light conditions. A rapid succession of images show results in a perception of continuous motion which is the basis for movies and TV.

Close your eyes for about five minutes so that any after images of the objects you have seen before are cleared. Open your eyes to see a bright colored object and then close your eyes again; you will see positive afterimages of the object in its original colors. Even though the positive afterimage at the initial stages will be of the same colors as the original, they gradually change color showing that different cones recover at different rates.

Negative afterimages

If you stare at a color picture intensely for about a minute and look at a white background, such as a sheet of white paper, you will see the complementary colors of the picture. The image of the object falls on a particular part of the retina and if we look at a green colored object, the cones that are more sensitive to the mid-wavelength region are activated

and are likely to be saturated. If you look at a bright sheet of white paper immediately, the cones that are sensitive to the middle wavelength region are inactivated since they need some time to replenish the active chemical whereas the cones that have maximum sensitivity for the low and high wavelengths are active, therefore you will see the object in complementary colors. Yellow will be seen after looking at blue image, cyan after looking at red image and vice versa.

Spatial Resolution of the Human Visual System

Lateral inhibition (color constancy)

By and large the perceived color of an object is independent of the nature of the illuminating white light source. The color of the object does not change markedly under different white light sources of illumination. For example, we see basically the same color of the object in the daytime in the light of the blue sky or under an incandescent light source even though the spectral power distributions of these sources are different and the intensity of light reaching the eye in both cases is also drastically different. The color constancy is a result of chromatic lateral inhibition. The human visual system has a remarkable way of comparing the signals from the object and manipulating them with the signals generated from the surroundings and picking up the correct color independent of the illuminating source. The color constancy and simultaneous lightness contrast are due to lateral inhibition. Lateral inhibition was exploited by a number of artists to provide certain visual effects that would not have been possible otherwise. For example, Georges Seurat in his famous painting 'La Poseuse en Profil' improved the light intensity contrast in different regions of his painting by employing edge enhancement effects resulting from lateral inhibition. Victor Vasarely's art work 'Arcturus' demonstrates the visual effects of lateral inhibition in different colors.

Spatial assimilation

The color of a region assimilates the color of the neighboring region, thus changing the perception of its color. Closely spaced colors, when viewed from a distance, tend to mix partitively and are not seen as distinct colors but as a uniform color due to additive color mixing. Georges Seurat in his painting 'Sunday Afternoon on the Island of La Grande Jatte', uses thousands of tiny dots of different colors to produce a variety of color sensations, a technique which is known as pointillism. The visual effects of spatial assimilation were clearly demonstrated by Richard Anuszkiewicz in his painting 'All Things Do Live in the Three'.

Color Perception Models

Thomas Young, in 1801, postulated that there are three types of sensors thus proposing trichromacy to explain the three attributes of color perception: hue, saturation, and brightness. Data on microspectrophotometry of excised retinas, reflection densitometry of normal eyes, and psychophysical studies of different observers confirm the existence of three types of cones. As shown in Figure 1, the three types of cones have different responsivities to light. The cones that have maximum responsivity in the short wavelength region are often refereed as S-cones, the cones that have maximum in the intermediate wavelength range as I-cones, and the cones that have maximum in the long wavelength region as L-cones.

The three response curves have considerable overlap. The overlap of the response curves and the complex data processing that takes place in the visual system enable us to differentiate as many as 150 spectral colors and about seven to ten million different colors and shades. An important consequence of the existence of only three types of color sensors is that different spectral power distributions can produce identical stimuli resulting in the perception of the same color. Such spectral power distributions are called metamers.

The neural impulses generated by the cones are processed through a complex cross-linking of bipolar cells, horizontal cells, amacrine cells, and communicated to ganglion cells leading to the optic chiasmas through the optic nerve. Information gathered by about 100 million photoreceptors, after considerable processing, is transmitted by about a million ganglion cells.

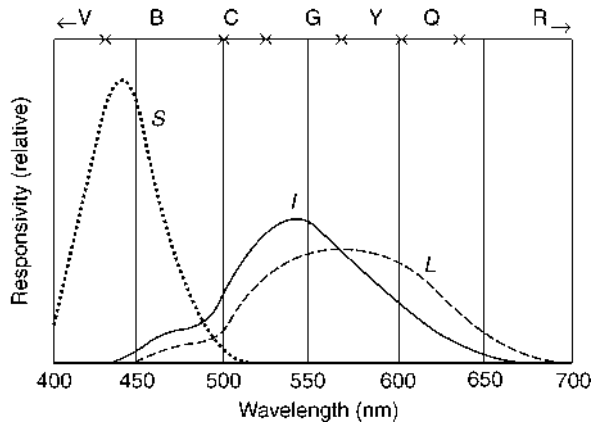

Figure 1 Relative spectral absorption of the three types of cones. *S*, *I*, and *L* respectively stand for the cones that have maximum responsivity for the short, intermediate, and long wavelength regions.

The zone theory of color vision assumes two separate zones where the signals are sequentially processed. In the first zone the signals are generated by the cones and in the second zone, the signals are processed to generate one achromatic and two opponent chromatic signals. Possibly other zones exist where the signals generated in the first and the second zones are further processed and communicated to the visual cortex where they are processed and interpreted based on the temporal and spatial response and comparisons with the information already available in the memory.

Measuring Color and Colorimetry

Since color is a human perception, how do we measure color? Colorimetry is based on the human visual system and how different observers perceive color. Though, the spectral power distribution of light is helpful for characterizing spectral hues, it is very difficult to infer the color from the spectral power distribution.

The three attributes of color perception are hue, saturation, and brightness. Hue is the color such as red, green, and blue. Saturation refers to the purity of the color. A color is said to be more saturated if the whiteness in the color is less and vice versa. Brightness is related to the intensity of light. For the sun at sunrise the hue is red, saturated, and low brightness. About half an hour after sunrise, the hue is yellow, less saturated, and higher brightness. Though a variety of color schemes are available in the literature, in the following we discuss the Munsell system and the C.I.E. diagram only.

Munsell Color System

Albert Munsell, a painter and art teacher, devised a color atlas in 1905 by arranging different colors in an ordered three-dimensional space. He used the three color attributes hue, chroma, and value corresponding to hue, saturation, and brightness. In this system there are 10 basic hues, each of which is divided into 10 equal gradations, resulting in 100 equally spaced hues. Each hue has a chart with a number of small chips arranged in rows and columns. Under daylight illumination, the chips in any column are supposed to have colors of equal chroma and the chips in any row are supposed to be of equal brightness. The brightness increases from the bottom of the chart to the top in steps that are perceptually equal. The saturation of the color increases from the inside edge to the outside edge of the chart in steps that are also perceptually equal. Each chip is identified by hue, value/chroma. In the Munsell scale, black is given a value of 0 and white 10, and the nine grays are uniformly placed in between. Though the Munsell system has gone through many revisions, typically it has a total of about 1450 chips. The Munsell color scheme, though subjective in nature, is widely used because of its simplicity and ease.

The Commission Internationale de l'Echlairage (C.I.E.) Diagram

Even though the visible light reaching the eye has a very complex SPD, color perception is mostly dependent on the signals generated by the three different types of cones. However, we do not know precisely the absorption and sensitivity of the cones and therefore are not in a position to calculate the signal strengths generated by the three different cones. Additionally, the response of the cones is specific to the individual. Therefore, the C.I.E. system is based on the average response of a large sample of normal observers. The C.I.E. system employs three color matching functions $\bar{x}(\lambda), \bar{y}(\lambda)$, and $\bar{z}(\lambda)$ (Figure 2), based on the psychological observations of a large number of standard observers to calculate the 'tristimulus values' X, Y, and Z defined as

$$X = c\int S(\lambda)\bar{x}(\lambda)\mathrm{d}\lambda \qquad [1]$$

$$Y = c\int S(\lambda)\bar{y}(\lambda)\mathrm{d}\lambda \qquad [2]$$

$$Z = c\int S(\lambda)\bar{z}(\lambda)\mathrm{d}\lambda \qquad [3]$$

Here, λ is the wavelength, $S(\lambda)$ the spectral power distribution, and c is the normalization constant. The integration is carried out over the visible region, normally 380 to 780 nm. The system is based on the assumption of additivity and linearity. The sum of the tristimulus values $(X + Y + Z)$ is normalized to 1. We choose $y(\lambda)$ so that the Y tristimulus value is proportional to the luminance which is a quantitative measure of the intensity of light leaving the surface. We further calculate x and y which are called chromaticity coordinates:

$$x = \frac{X}{(X+Y+Z)} \qquad [4]$$

$$y = \frac{Y}{(X+Y+Z)} \qquad [5]$$

The psychological color is then specified by the coordinates (Y, x, y). Relationship between colors is usually displayed by plotting the x and y values on

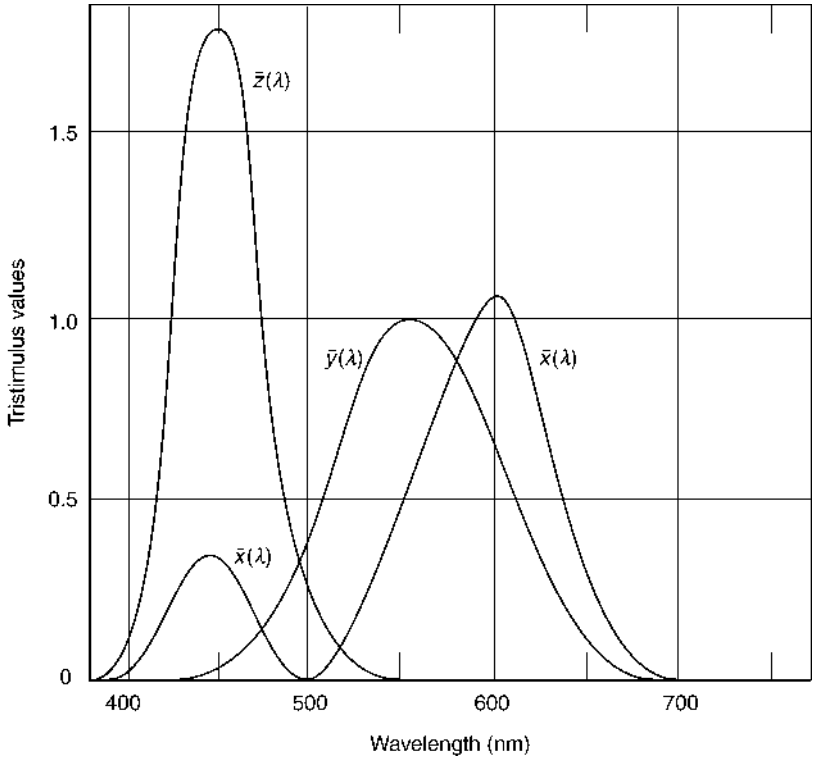

Figure 2 Color-matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ of the C.I.E. 1931 standard observer.

a two-dimensional Cartesian Coordinate system with z as an implicit parameter ($z = 1 - x - y$) which is known as the C.I.E. chromaticity diagram (Figure 3). In this plot, the monochromatic hues are on the perimeter of a horseshoe shaped curve and is called spectrum locus. A straight line joining the ends of the spectrum is called the line of purples. Every chromaticity is represented by means of two coordinates (x, y) in the chromaticity diagram. The point corresponding to $x = 1/3$ and $y = 1/3$ (also $z = 1/3$) is the point of 'equal energy' and represents achromatic point. The complementary colors are on the opposite sides of the achromatic point. Complementary wavelength of a color is obtained by drawing a line from the color through the achromatic point to the perimeter on the other side.

Continuum (white) sources

A perfect white source is the one that has constant SPD over the entire visible region. Such a source will not distort the color of the objects they illuminate. A number of sources in general are considered as white sources; they include incandescent lamps, day skylight, a variety of fluorescent lamps, etc. It should be noted that the SPD of these sources is not constant and varies with the source. Typical SPDs of commonly used light sources are given in Figure 4.

Color temperature

The thermal radiation emitted by a black body has a characteristic SPD which depends solely on its temperature. The wavelength λ of the peak of this distribution is inversely proportional to its absolute temperature. As the absolute temperature of the blackbody increases, the peak shifts towards shorter wavelengths (blue region) and also the width of the distribution decreases. Since the peak of the black body radiation is dependent on the absolute temperature, the color of the blackbody can be approximately defined by its temperature.

The color temperature of a body is that temperature at which the SPD of the body best matches with that of the blackbody radiation at that temperature. The concept of color temperature is often used to characterize the sources. Low color temperature sources have a reddish tinge and high color temperature sources tend to be slightly bluish. For example, the color temperature of red looking star Antares is about 5000 K and that of the bluish white looking Sirius is about 11 000 K. One should note that the color temperature of a body is different from its actual temperature. The fluorescent lamps have a

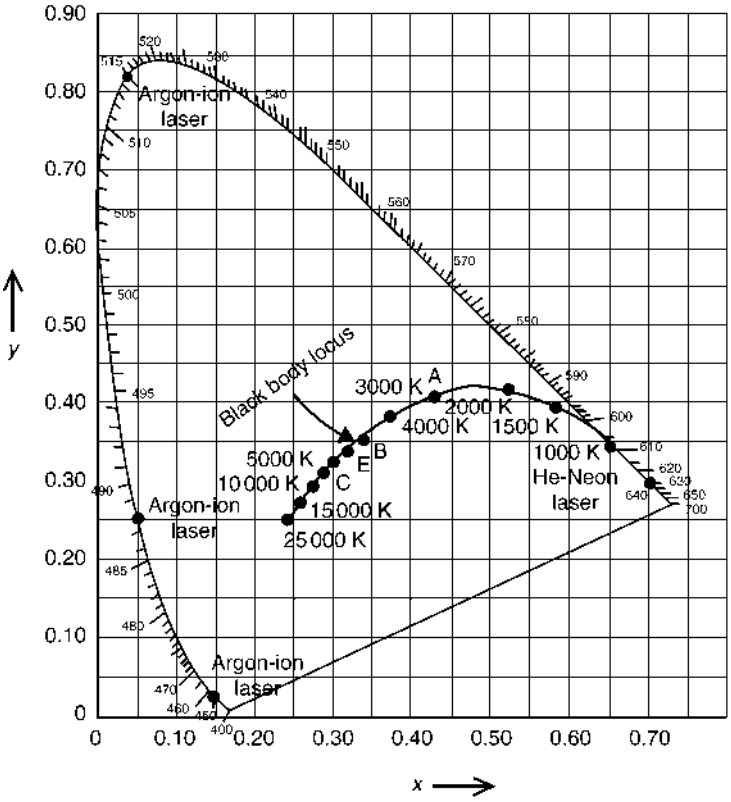

Figure 3 Chromaticity diagram of the C.I.E. 1931 standard observer. The chromaticities of the incandescent sources and their color temperatures are given inside the curve. A, B, and C are the C.I.E. standard illuminants. The chromaticities of the helium-neon laser (red, 633 nm) and the argon-ion laser (green 515 nm, blue 488 nm, and indigo, 458 nm) are also shown.

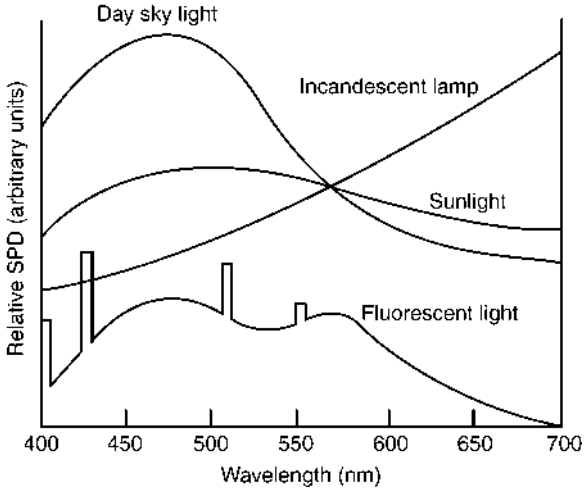

Figure 4 Relative SPDs of commonly used white light sources.

color temperature range of 5000–7500 K, incandescent lamps about 3000 K, direct sunlight ~4870 K and overcast skylight has a color temperature of ~7000 K.

The C.I.E. chromaticity diagram along with the color temperatures of the commonly employed sources, is shown in Figure 3.

Color in the Nature

As the solar radiation passes through the atmosphere, it is partly absorbed and partly scattered by the atmospheric constituents: air, water vapor, and dust particles. Absorption depends on the molecular properties of the air molecules, water vapor, and the dust particles, and is strongly wavelength dependent. Absorbed energy increases the internal energy of the molecule finally manifesting itself as thermal energy. When the sun is at the zenith, the solar radiation passes through one air mass loosing approximately 32% of its power. The SPD of the solar radiation above the atmosphere and at the Earth's surface at air mass 1 is shown in Figure 5. Scattering and absorption of the solar radiation by the atmosphere are responsible for a variety of interesting colorful views in the nature and they are briefly discussed in this section.

Atmospheric Colors

1. *Colors of the sky.* Scattering occurs over most of the visible part of the electromagnetic spectrum and is a function of the frequency of

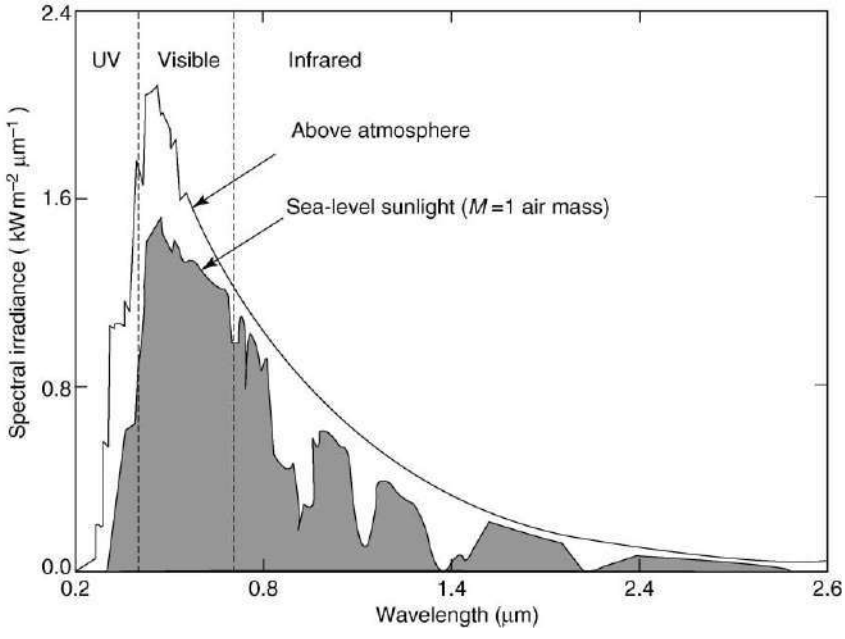

Figure 5 Spectral irradiance (kW m^{-2} μm^{-1}) of the solar radiation measured above the atmosphere and at the Earth's surface at sea level on a clear day and the sun at the zenith.

the radiation. According to Lord Raleigh's theory of scattering, the scattering probability varies inversely as a function of the fourth power of the wavelength of the radiation. If we consider the radiation corresponding to blue (400 nm) and the radiation corresponding to say red (700 nm), the photons corresponding to blue radiation are scattered by about 9.4 times that of photons corresponding to the red radiation. During noontime, when we look at the sky, not directly at the sun, the sky looks unsaturated blue because of the predominant scattering experienced by the high frequency radiation corresponding to blue color. However, it should be noted that the daytime skylight contains all the spectral colors of the solar radiation though the spectral power distribution has changed resulting in an increase in the intensity of the high frequency (blue) radiation. During noontime at the horizon, the sky looks close to white in color with a bluish tinge, because the solar radiation has passed through many air masses. As the thickness of the atmosphere increases, the intensity of the blue radiation decreases and ultimately as the radiation reaches the horizon, all the spectral colors are scattered about the same level resulting in white or bluish white sky.

2. *Polarization of the solar radiation by the atmosphere*. Light scattered by a molecule is polarized. The blue skylight is polarized with a maximum polarization of about 75% to 85% at 90° from the sun.
3. *Distant mountains seem bluish in color.* When an artist wishes to paint a distant mountain in perspective, he/she paints it blue; why? This is called airlight. The low wavelength part of the solar radiation is scattered more prominently, giving it a blue perspective when observed from a distance. However, it should be mentioned that the distant mountains may look any color depending on the scattering and absorption of the solar radiation sometimes resulting in 'Purple Mountain's majesty.'
4. *Sun looks red at the sunrise and sunset*. At sunset as the sun approaches the horizon, the sun changes its color from dazzling white to yellow, orange, bright red, and dull red and the order is reversed at sunrise. At the sunsets and sunrise, the solar radiation travels through many airmasses compared to approximately one airmass when the sun is at the zenith. Because of the large mass of air the radiation has to pass through, first the blue part of the spectrum is removed and the transmitted green to red spectral colors give the appearance of yellow. As this spectrum passes through more airmass, green is scattered most, leaving spectral yellow, orange, and red. As the radiation further passes through air, yellow and orange are also scattered, leaving only red.

5. *Twilight*. After sunset (dusk), or before the sunrise (dawn), the sky is illuminated by light which in turn illuminates the landscape. At sunsets and sunrise the solar radiation passes through a large mass of air. At a solar depression of about 0°, most of the blue is lost by scattering, resulting in a yellow arch and bright sky, and as the solar depression increases to about 10°, the arch will be reddish and the brightness of the sky reduced.
6. *Alpenglow*. Snow-covered mountains glow orange yellow from the low sun and after the sunset the same mountains look purple because of the reflection of the twilight.
7. *Twinkling of stars*. The stars twinkle. The local variations in the density of air are responsible for the twinkling of stars. As the density of air changes, so does the index of refraction and the starlight is continually refracted and dispersed. Twinkling is more prominent when the stars are on the horizon or on cold clear and windy nights.
8. *Mirages*. Mirages are refracted images produced because of the existence of refractive index gradients in the atmosphere. This results in the sight of shimmering water on a highway, inverted view of a car, etc.
9. *Aurora Borealis*. Aurorae are diffused slowly moving lights seen at high latitudes. They are often greenish yellow with different shapes. The lights in the northern hemisphere are called 'Aurora Borealis' and the lights in the southern hemisphere 'Aurora Australis.' They are caused primarily by the high-energy electrons deflected by the Earth's magnetic field towards the Earth's magnetic poles interacting with the atmospheric molecules resulting in molecular excitation and ionization. When the molecules de-excite, the characteristic electromagnetic radiation is commonly emitted. The commonly seen yellow-green light is due to the molecular emission of oxygen at 557 nm.
10. *Urban glows*. The scattered light from cities can be observed from space. The metropolitan Los Angeles can be observed from about 250 km in space. Extensively used sodium lamps with strong spectral yellow and blue-green mercury and fluorescent lights have unique color spectra of their own.

Colors due to Water and Water Droplets

1. *What is the color of pure water?* The transmission coefficient of pure water as a function of wavelength for different thicknesses of water column is shown in Figure 6. Our perception of the color of water changes with the thickness of water. For 1 mm thick water column, all the wavelengths are transmitted, resulting in the perception of white; 1 m thick water column looks unsaturated bluish green. As the thickness of water increases, the bluish green color will get more and more saturated. The color of water perceived by us depends on: (i) light reflected from the water surface; (ii) the light refracted at the water surface and diffused and scattered by the water molecules; and (iii) the refracted light reaches the bottom and gets reflected, which is further diffused and scattered and refracted at the water surface again. What we see is an additive sum of all these components.

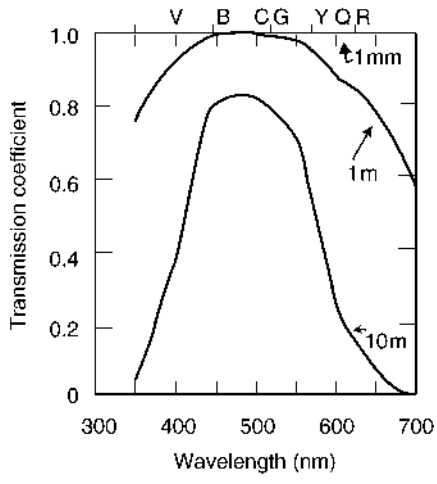

Figure 6 Transmission coefficient of water as a function of wavelength for different thicknesses of water (1 mm, 1 m, and 10 m).

2. *Refraction through water–air interface*. When light travels from an optically denser medium to an optically lighter medium, the light is deviated away from the normal. This is the reason why the apparent depth of a fish located in a pond is less than the true depth and a pencil struck in water seems to bend. If you are a diver when you look at the horizon from under water, the hemisphere seems to be compressed into a circle of diameter 97.2° instead of 180° which is known as 'Optical Manhole.'
3. *Glitter*. When you look at the reflection of the sun or the moon in wavy water, you notice a bright elongated reflected light known as glitter. It has a brilliant reddish orange color when the sun is above the horizon. Since the water is wavy, light reflected by a certain part of the wave at a certain time has the correct angle of reflection to reach your eye, which is known as glint. An ensemble of a large number of glints

coming from different waves and different parts of the waves results in glitter. Since you are looking at the reflected light, it will be polarized.

4. *Sky pools and landpools*. The light coming from a wavy water surface of a lake consists of oval shaped bright colored patterns that continually change in shape. This results from the reflection of skylight by the wavy water. If the waves are not turbulent the reflected light reaching your eye continually changes resulting in bright colored patterns. Landpools are caused by a nearby sloping landscape and the reflected light pattern from the landscape.
5. *Thin films*. Oil floats on water and spreads into thin films because of its low surface tension. When viewed in reflected natural white light, the light reflected from the top surface of the oil and the light reflected from the interfacial surface of oil and water interfere to produce colorful patterns. Brilliant colored patterns are commonly seen on the roads after rain because of gasoline on water. Foam ordinarily consists of bubbles ranging typically 10^{-4} mm to several mm. Even though each bubble produces colorful light due to the interference effects, the light coming out of the foam adds up to white light.
6. *Wet surfaces look darker*. A water spot on a concrete road looks darker. Since concrete is not porous a thin layer of water sticks to the surface. At the water surface part of the light incident is reflected, and a large part is transmitted. The transmitted light is diffusively scattered, part of which goes through total internal reflection and only a small part reaches the eye, making the surface look darker.

   When a fabric is wet it looks darker because of a slightly different mechanism. The surface of the fabric gets coated with a thin layer of water (refractive index = 1.33) which is smaller than the refractive index of the fabric. When the cloth is wet more light is transmitted, and less gets reflected than before. The transmitted light penetrates the fabric and gets scattered, resulting in a darker appearance.
7. *Rainbows*. Water with a refractive index of 1.33 is a dispersive medium. White light consisting of different wavelengths travel at different speeds in the medium and emerge at different directions. If the dispersion is adequate, different spectral colors can be seen by the naked eye. The dew covered lawn on a sunny day displays the spectral colors. If you stand close to the dew drop without blocking the sun's radiation falling on the drop, at some angle one should see the spectral colors. The primary rainbow is produced by the water droplets in the atmosphere after rain. The radius of the primary rainbow is about 42° with a width of about 2°. The color of the primary rainbow would be blue to red from the center of the rainbow. A rainbow is observed at the antisolar point. The brightness of the rainbow depends on the ambient sunlight, the size of the droplets, and whether the sun is low or high in the sky. The secondary rainbow, which is fainter, is often seen outside the primary rainbow at about 51°. The secondary rainbow will have red to blue from the center of the secondary rainbow. The secondary rainbow is approximately half as bright as the primary rainbow. The secondary rainbow is caused by two total internal reflections within the droplet whereas the primary rainbow is the result of one total internal reflection only. Because of the total internal reflection process, light is about 95% tangentially polarized in the primary rainbow.
8. *Heiligenschein*. Heiligenschein or the 'holy light' can be seen around the shadow of your head on a lawn on a sunny morning when dew drops are present. Usually, the dew drops are held by the tiny hairs of the grass at some distance from the blade of the grass. As light falls on the droplet, the spherical drop converges light at its focus and if the grass blade is located at that point, the light gets retroreflected approximately towards the sun resulting in holy light around your shadow.
9. *Coronae*. Colored concentric patchy rings seen around the moon and the sun are called coronae. They are primarily produced by the water droplets in the thin clouds. Even though the water droplets are randomly located, the diffracted light from the droplets has a defined angle and the superposition of the diffraction patterns from different water droplets results in colored ring structures. Coronae can be easily seen around the full moon viewed through thin clouds.
10. *The Glory*. On a foggy day with the sun at your back, one may see bright colored rings in the form of arches of about 5° to 10° at the solar point. The light in the glory is polarized and the ring structure is produced by reflection of the solar radiation by the fog.
11. *Clouds*. Clouds contain suspended water droplets or ice crystals. The size of the droplets ranges from less than a micrometer to about 100 micrometers (10^{-6} m). When the air is supersaturated with water vapor, the water condenses into water droplets, resulting in a definite shape to the cloud. When the sun's radiation falls on

the cloud, the light is scattered by the water droplets. Even though one would expect that the scattered light may have some color distribution, since we look at the scattered light from a very large number of droplets, the summation of all the colors results in white. Some clouds look dark or gray because, either they are in the shadow of other clouds or they are so thick that most of the light is absorbed. During sunset and sunrise, the clouds look bright orange because of the colored light illuminating them. Even though all the clouds are basically white, they may appear colored because of the selective absorption of certain colors in the atmosphere or due to the colored background radiation. Often, you see a silver lining at the edges of a cloud, because at the edges the density of water is low and when the sun illuminates it from the back, the radiation is scattered in the forward direction quite efficiently, resulting in a bright and shiny 'silver lining' to the cloud.

12. *Lightning*. Lightning is caused by an intense electrical discharge due to the high potential differences developed. On a cloudy and rainy day, as the water droplets grow in size, the falling drops get distorted to ellipsoidal shape and get polarized. The top smaller portion becoming negative and the lower larger part becoming positive. As the deformation increases, the drop ultimately breaks and the smaller drops being negatively charged and the larger ones positively charged. The wind currents easily lift the buoyant small drops to higher altitudes and the heavier positively charged droplets fall towards the Earth. The final result is that the upper region of the cloud is negatively charged whereas the lower region of the cloud is positively charged. The portion of the Earth under the cloud becomes negatively charged due to induction. The localized ions in the positively charged cloud produce a strong electric field which is further augmented by the induced negative charges on the Earth below. The strong electric field ionizes the air molecules setting up a current discharge to the earth. This discharge is called stepped leader. Following the stepped leader, the charges in the upper region of the cloud also discharge resulting in what is called return stroke. The sequence of leader and return strokes occur typically about ten times in a second. Since the clouds are charged and they are closer to each other, lightening is more frequent between the oppositely charged clouds than cloud and Earth. It is estimated that the lightning between the clouds is about five times more frequent compared to the lightening events between a cloud and the Earth. In this process, the nitrogen and oxygen molecules in air are excited and ionized. The flash is produced by the de-excitation of these molecular species. Since the temperatures involved are quite high (30 000 K), the flash has a strong unsaturated bluish hue. Some hydrogen lines are also observed in the flash. Hydrogen is produced as a result of the dissociation of water molecules during the discharge process.
13. *Ice and Halos*. Snow reflects most of the light falling with a reflectivity of about 95%. Snowflakes are crystalline and have hexagonal structure. In nature they occur in assorted sizes and shapes. As white light falls on them, light may be simply reflected from their surfaces making them shining white, which are often called glints. The light may also pass through the crystalline material get dispersed and emerge with different colors, which are known as sparkles.

    The halos appear as thin rings around the sun or the moon. They are produced because of the minimum deviation of light in snow crystals. The 22° halo is most common and is often seen a number of times in a year. The halos are circular and have a faint reddish outer edge and a bluish white diffused spot at the center. They often look white because of the superposition of different colors resulting in a more or less white color.

Color in Art and Holography

Sources of Color

There are two major color sources, dyes and paint pigments. Dyes are usually dissolved in a solvent. The dye molecules have selective absorption resulting in selective transmission. Paint pigments are powdered and suspended in oil or acrylic. The paint particles reflect part of the radiation falling on them, selectively absorb part of the radiation and transmit the balance. Lakes are made of translucent materials such as tiny grains of alumina soaked in a dye of appropriate color.

In the case of water colors and printer's inks, part of the light incident is reflected from the surface and the remaining transmitted which is selectively absorbed by the dye. The transmitted light reflected by the paper, passes through the dye again and emerges. Each transmission produces color by subtractive color scheme and the emerging light is mixed partitively by the observer.

In the case of paints, the color depends on the size and concentration of the pigment particles and the

medium in which they are suspended. Part of the light incident on the artwork gets reflected at the surface and the remaining selectively absorbed and transmitted which in turn is further reflected and transmitted by other pigment particles each time following a subtractive color scheme. The light reaching the canvas gets reflected and passes through the medium and emerges. Again, all the emerging beams are partitively mixed by the observer.

Holograms may be broadly categorized as transmission and reflection holograms. Both transmission and reflection (volume) holograms are prepared by producing the interference pattern of the reference and object laser beams. Transmission holograms can be viewed in the presence of the reference laser beam that was employed for its preparation. Spectacular three-dimensional colored images are reconstructed by the reflection holograms depending on the angle of incidence of the white light and the direction of observation. Embossed holograms are popular for advertising, security applications (credit cards), display, and banknotes. Some studios prepare three-dimensional holographic portraits that can be viewed in natural light. In the future, three-dimensional videos and even holographic three-dimensional movies may be a reality.

Structural Colors in Nature

Nanoscale Photonic Lattices

Newton observed that iridescence (change in color with the direction of observation) is caused by optical interference. The colors are produced by the interference or the light diffracting characteristics of the microstructures and not by the selective reflection or absorption of light, as in the case of pigments. Photonic band structures consist of periodic arrangement of dielectric materials even though they are transparent, have a bandgap resulting in inhibition of certain wavelengths. Light with frequencies in the gaps between the bands get reflected resulting in the specific color.

The structural colors of light can be seen in the animal world and minerals. For example, opal, a sedimentary gem does not have crystalline structure, but consists of tiny spheres of silica packed together. The diffraction of light through these structures results in spectacular colors determined by the size of the spheres and their periodicity. The spines of the sea mouse Aphrodite are covered with natural photonic structures that produce iridescence resulting in brilliant colors depending on the direction of the incident light and the direction of observation.

In the case of the butterfly, *Morpho rhetenor*, the metallic blue color is produced by the periodic structure of its wings. Its wings have layers of scales, each of about 200 mm long and 80 mm wide and about 1300 thin parallel structures per mm which form a diffraction grating.

The bird feathers viewed through an optical microscope show the structural components which are responsible for their coloration. The structural details of the peacock's eye pattern employing the scanning electron microscopy show photonic structures on peacock barbules. Barbules are the strap like branches on the peacock feathers. The four principal colors on the peacock feathers blue, green, yellow, and brown were identified wth the specific structures. The interference between the light reflected from the front surface to that reflected from the back was shown to be responsible for the brown color in the peacock's eye. The photonic structure of the green barbule has a periodic structure of melanin cylinders with a spacing of about 150 nm whereas the blue barbule has a periodic structure of 140 nm, and yellow barbules have a periodicity of 165 nm.

See also

Dispersion Management. **Holography, Techniques:** Color Holography. **Photon Picture of Light**. **Polarization:** Introduction. **Scattering:** Scattering Theory.

Further Reading

Blau SK (2004) Light as a feather: structural elements give peacock plumes their color. *Physics Today* 57: 18–21.

Color Science, http://www.physics.sfasu.edu/astro/color.html.

Colourware, Colour FAQ, http://www.colourware.co.uk/cpfaq.htm.

CVRL Color & Vision database, http://www.cvrl.org/.

Hecht E (2002) *Optics*, 4th edn. San Francisco, CA: Addison Wesley.

Lynch DK and Livingston W (2001) *Color and Light in Nature*, 2nd edn. Cambrige, UK: Cambridge University Press.

Minnaert MGJ (1993) *Light and Color in the Outdoors*. Heidelbergh: Springer-Verlag.

Nassau K (2000) *The Physics and Chemistry of Color*. New York: John Wiley and Sons.

Poynton CA, Color technology, http://www.Poynton.com/ColorFAQ.html.

Tayeb G, Gralak B and Enoch S (2003) Structural colors in nature and butterfly-wing modeling. *Optics and Photonics News* 14(2): 38–43.

Wyszecki G and Stiles WS (2000) *Color Science*. New York: John Wiley.

D

DETECTION

Contents

Fiber Sensors
Heterodyning
Image Post-Processing and Electronic Distribution
Smart Pixel Arrays

Fiber Sensors

M J Connelly, University of Limerick, Limerick, Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The growth in the fiber optic communications and optoelectronics industries has led to the development and commercialization of many types of optical components. Fiber sensor devices and systems are a major user of such technologies. This has resulted in the development of commercial fiber sensors that can compete with conventional sensors. Fiber sensors are capable of measuring a wide variety of stimuli including: mechanical (displacement, velocity, acceleration, strain, pressure); temperature; electromagnetic (electric field, magnetic field, current); radiation (X-ray, nuclear); chemical composition; flow and turbulence in liquids; and biomedical. The main advantages of fiber sensors are that they are low cost, compact and lightweight, robust, passive, immune to electromagnetic interference, and highly sensitive.

Fiber sensors can be grouped into two basic classes: intrinsic and extrinsic. In an intrinsic fiber sensor the sensing is carried out by the optical fiber itself. In an extrinsic sensor the fiber is simply used to carry light to and from an external optical device where the sensing takes place. Point fiber sensors are localized to discrete regions; quasi-distributed sensors utilize point sensors at various locations along a fiber and distributed sensors are capable of sensing over the entire length of fiber.

A basic fiber sensor system, as shown in Figure 1, consists of a light source, fiber sensor, optical detector, and signal processing electronics. A measurand causes some change (intensity, phase, polarization, spectrum, etc.) in the light propagating through the sensor. This change is detected and processed to give an output signal proportional to the measurand. A large part of fiber sensor research concerns the development of fiber sensors sensitive to particular measurands and appropriate signal processing techniques. We describe some of the most important fiber sensors and also consider quasi-distributed and distributed sensing, including the use of multiplexing techniques.

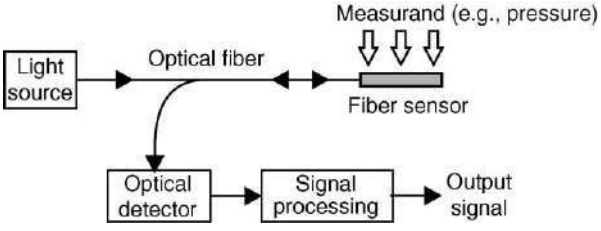

Figure 1 Basic fiber sensor system.

Intensity-Based Fiber Sensors

The intensity modulation (IM) of light is a simple method for optical sensing. There are several mechanisms that can produce a measurand-induced change in the optical intensity propagated by an optical fiber. Perhaps the simplest type of IM fiber sensor is the microbend sensor shown in Figure 2. The sensor consists of two grooved plates between which passes an optical fiber. The upper plate can move in response to pressure. When the fiber is bent sufficiently, light escapes into the fiber cladding and is lost. The greater the pressure on the plates the more loss occurs.

Coupling-based fiber sensors are useful for measurement of displacement or dynamic pressure. Transmission and reflective configurations are possible as shown in Figure 3. The transmission coupling-based sensor consists of two fibers with a small gap between them. The amount of light coupled to the second fiber depends on the fiber acceptance angles and the distance between the fibers. One of the fibers can move in response to vibration or pressure thereby changing the distance between the fibers and hence the coupling loss. The reflection-based sensor operates in a similar fashion, where light is reflected from a flexible diaphragm back into a collecting fiber. The reflected light intensity changes as the diaphragm is flexed. Once the coupling relationship between the input fiber, diaphragm and collecting fiber is known, intensity changes can be related to the applied displacement or pressure.

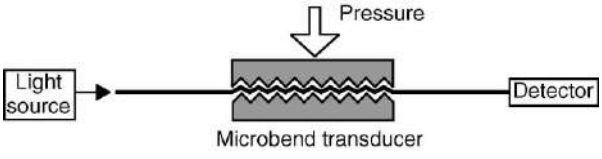

Figure 2 Microbend fiber sensor. The transducer moves in response to pressure and in doing so changes the bending radius of the fiber and thereby the fiber loss.

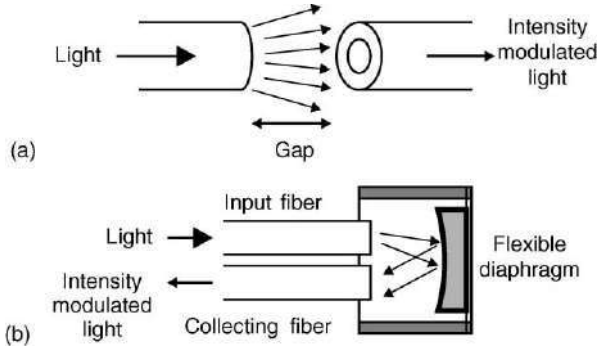

Figure 3 (a) Transmission and (b) reflective coupling-based fiber sensors.

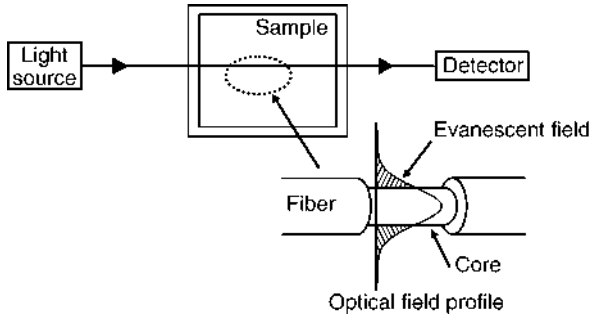

Figure 4 Evanescent wave fiber chemical sensor.

Evanescent wave fiber sensors exploit the fact that some of the energy in the guided mode of an optical fiber penetrates a short distance from the core into the cladding. The penetration of light energy into the cladding is called the evanescent wave. It is possible to design a sensor where energy is absorbed from the evanescent wave in the presence of certain chemicals as shown in Figure 4. This is achieved by stripping the cladding from a section of the fiber and using a light source having a wavelength that can be absorbed by the chemical that is to be detected. The resulting change in light intensity is a measure of the chemical concentration. Measurements can also be performed in a similar fashion by replacing the cladding with a material such as an organic dye whose optical properties can be changed by the chemical under investigation. Evanescent wave fiber sensors have found many applications in the biomedical field, such as blood component meters for detection of cholesterol and uric acid concentrations in blood.

A linear position sensor based on time division multiplexing is shown in Figure 5. It uses a square-wave modulated light source, optical delay loops and an encoded card. The delay loops separate the return signal from the encoded card by a time that is greater than the pulse duration. The encoded return signal can be decoded to determine the card position and velocity.

Interferometric Sensors

Interferometric fiber sensors operate on the principle of interference between two or more light beams to convert phase differences to intensity changes. Common configurations used include Michelson, Mach–Zehnder, and Sagnac interferometers, polarimetric systems, grating and etalon-based interferometers and ring resonators. Interferometric fiber sensors have extremely high sensitivity and are able to resolve path differences of the order of 10^{-6} of the light

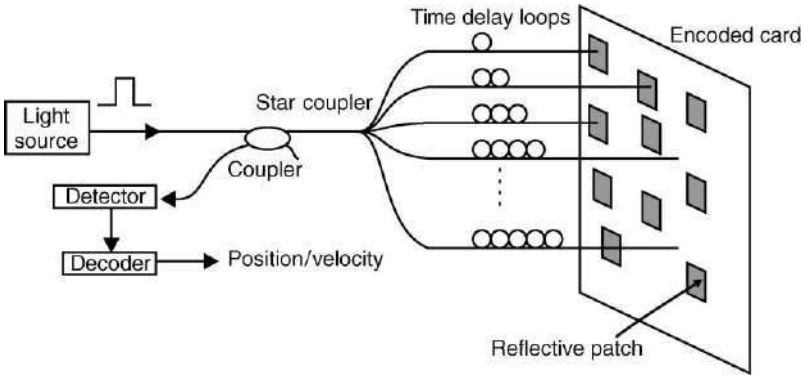

Figure 5 Linear position sensor using time division multiplexing.

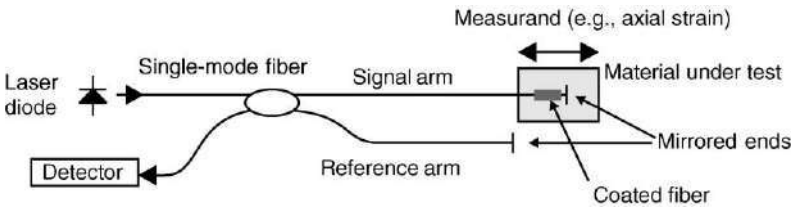

Figure 6 Michelson interferometric strain sensor. Part of the signal arm is embedded in a material, such as concrete (in civil engineering structures) or carbon composite (used in the aerospace industry).

source wavelength. To realize high sensitivity, spectrally pure light sources such as semiconductor lasers must be used. In addition, single mode fibers and components are used to maintain the spatial coherence of the light beams. Factors that affect the performance of interferometric sensors include the optical source phase noise and the polarization states of the interfering light beams.

In two-beam interferometric sensors, such as the Michelson interferometric strain sensor shown in Figure 6, one of the light beams travels through a fiber where its phase can be modulated by the measurand. The two reflected beams recombine on a detector, the output of which is of the form $1 + V \cos \phi$ as shown in Figure 7. ϕ is the relative phase shift between the return beams and V the interferometer visibility. Compensating techniques must be used to ensure that the interferometer sensitivity is maximized. Active techniques involve the insertion of an active device, such as a fiber stretcher, in one of the interferometer arms to control the mean phase difference between the beams. Passive schemes have the advantage in that no active components are required, but usually involve complex signal processing schemes such as active homodyne or synthetic heterodyne demodulation. The sensitivity of the fiber to the measurand can be further improved through the use of specialized coatings.

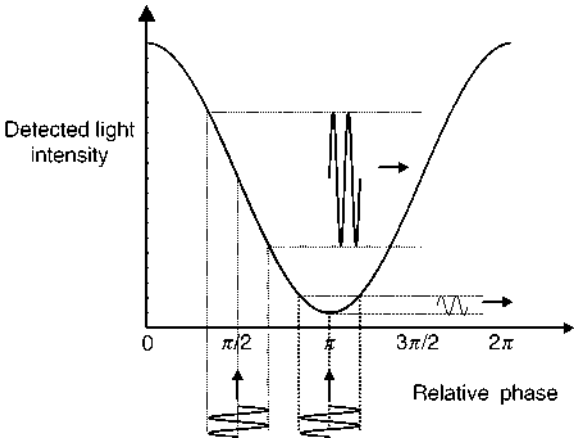

Figure 7 Two-beam interferometer transfer function. The sensitivity is maximized when the mean phase difference between the detected light beams is an odd integer multiple of $\pi/2$. When this is the case the interferometer is said to be operating in quadrature.

Compact interferometric sensors can be constructed using intrinsic or extrinsic Fabry–Perot type configurations, the most common of which are shown in Figure 8. If the Fabry–Perot cavity reflectivities are $\ll 1$, then it can be considered to be a two-beam interferometer.

Polarimetric fiber sensors relate changes induced in the polarization state of light to the measurand.

These sensors usually use high birefringence (Hi-Bi) fiber. The phase difference between the two orthogonally polarized modes of Hi-Bi fiber is given by

$$\phi = \Delta\beta l \qquad [1]$$

where $\Delta\beta$ is the fiber (linear) birefringence and l the fiber length. Both $\Delta\beta$ and l can be changed by the measurand. A polarimetric strain sensor using Hi-Bi fiber embedded in carbon fiber composite, is shown in Figure 9. Light from a linearly polarized He–Ne laser is launched into the Hi-Bi fiber, through a half-waveplate and lens. The waveplate is used to rotate the light plane of polarization so that it is at 45° to the principal axes of the fiber. This ensures that half of the input light power is coupled to each of the fast and slow modes of the Hi-Bi fiber. The output light from the Hi-Bi fiber is passed through a polarization beamsplitter. The beamsplitter separates the light into two orthogonally polarized beams, which are then detected. The output from each of the detectors is given by $V_1 = I_0 \sin^2\phi$ and $V_2 = I_0 \cos^2\phi$, respectively, where I_0 is the input light intensity. The state of polarization (SOP) of the output light from the Hi-Bi fiber is given by

$$\mathrm{SOP} = \frac{V_1 - V_2}{V_1 + V_2} = \cos^2\phi + \sin^2\phi \qquad [2]$$

It is a simple matter to calculate the SOP, from which ϕ can be determined. The form of eqn [2] is very similar to the two-beam interferometer transfer function, shown in Figure 7.

A polarimetric electric current sensor based on the Faraday effect is shown in Figure 10. The Faraday effect provides a rotation of the light's polarization state when a magnetic field is parallel to the optical path in glass. If an optical fiber is closed around a current-carrying conductor, the Faraday rotation is directly proportional to the current. By detecting the polarization rotation of light in the fiber, the current can be measured.

Linearly polarized light is equivalent to a combination of equal intensity right and left circularly polarized components. The linear state of polarization of the polarized laser light rotates in the presence of a magnetic field because the field produces

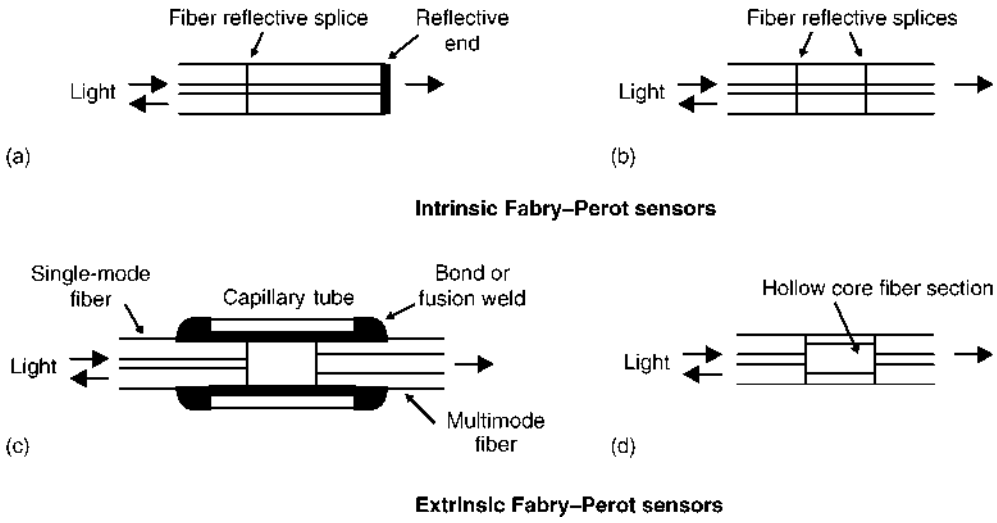

Figure 8 Interferometric fiber Fabry–Perot configurations.

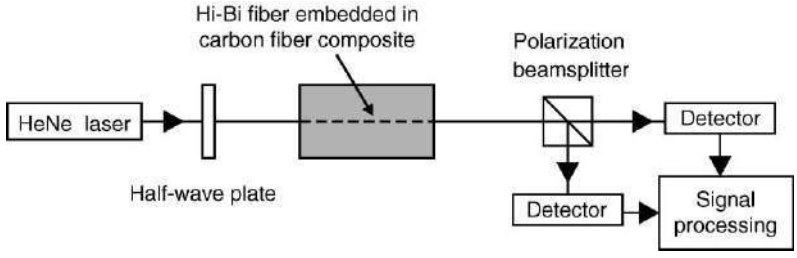

Figure 9 Polarimetric strain sensor system.

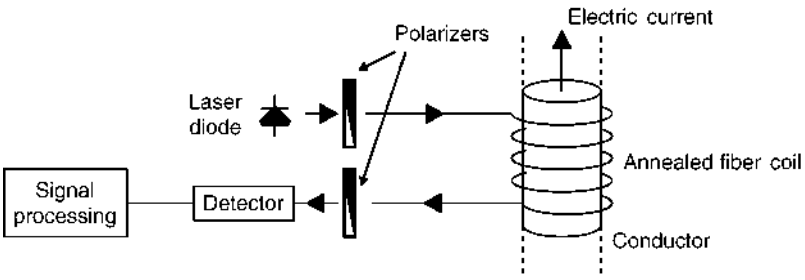

Figure 10 Polarimetric electric current sensor system.

circular birefringence in the fiber. This means that right and left circularly polarized light will travel at different speeds and accumulate a relative phase difference given by

$$\phi = VBl \qquad [3]$$

where V is the Verdet constant (rad Tesla^{-1} m^{-1}) of the glass, B the magnetic field flux density (Tesla), and l the length of the fiber exposed to the magnetic field. In the polarimetric current sensor, the polarization rotation is converted to an intensity change by the output polarizer. The output to the detector is proportional to $1 + \sin(2\phi)$, from which ϕ and the current can be determined.

The linear birefringence of conventional silica fiber is much greater than its circular birefringence. To make a practical current sensor, the linear birefringence must be removed from the fiber and this can be achieved by annealing the fiber. Annealing involves raising the temperature of the fiber, during manufacture, to a temperature above the strain point for a short period of time and slowly cooling back to room temperature. This reduces stresses in the glass, which are the principal cause of linear birefringence, and also cause physical and chemical changes to the glass. Waveguide-induced birefringence cannot be removed by annealing, but can be significantly reduced by twisting the fiber.

White light – or more accurately low-coherence – interferometry utilizes broadband sources such as LEDs and multimode lasers in interferometric measurements. Optical path differences (OPDs) are observed through changes in the interferometric fringe pattern. A processing interferometer is required in addition to the sensing interferometer to extract the fringe information.

The processing of white light interferometry signals relies on two principal techniques. The first technique involves scanning the OPD of the processing interferometer to determine regions of optical path balance. The second technique involves determination of the optical spectrum using an optical spectrum analyzer. The resulting fringe pattern fringe spacing is then related to the OPD of the sensing interferometer.

An example of the optical path scanning technique is shown in Figure 11a. The sensing interferometer is designed such that its OPD is much greater than the coherence length of the light source, so at its output no interference fringes are observed. The output of the sensing interferometer is fed to the processing interferometer. The OPD of the processing interferometer is scanned using a piezoelectric fiber stretcher driven by a suitable scanning voltage. As the processing interferometer is scanned, interference fringes are observed at two distinct points as shown in Figure 11b. The first set of fringes occurs when the OPD of the processing interferometer is within the coherence length of the optical source (close to zero). The second set of fringes occurs when the OPDs in the two interferometers are equal. The OPD in the sensing interferometer can be determined by measuring the OPD in the processing interferometer between the two positions of maximum visibility in the output signal from the detector. This, in turn, can be related to OPD changes due to the action of the measurand, in this case strain.

Sagnac interferometric sensors can be used to create highly sensitive gyroscopes that can be used to sense angular velocity (e.g., in aircraft navigation systems). It is based on the principle that the application of force (e.g., centrifugal force) will alter the wavelength of light as it travels around a coil of optical fiber. A basic open-loop fiber gyroscope is shown in Figure 12. The broadband source (e.g., superluminescent diode) is split into two counter propagating light beams traveling in the clockwise and anticlockwise directions. The polarizer is used to ensure the reciprocity of the counter propagating waves. The inherent nonlinear response of the gyroscope can be overcome by using a phase modulator and signal processing techniques. In the ideal case the detector output is proportional to $1 + \cos\phi_s$. The Sagnac phase shift ϕ_s between the

two returning beams is given by

$$\phi_s = \frac{8\pi AN\Omega}{c\lambda_0} \qquad [4]$$

A is the area enclosed by a single loop of the fiber, N the number of turns, Ω the component of the angular velocity perpendicular to the plane of the loop, λ_0 the free-space wavelength of the optical source, and c the speed of light in a vacuum. Sensitivities greater than 10^{-8} rad/s and dynamic ranges in excess of 40 dB have been achieved with open-loop fiber gyroscopes. More advanced gyroscopes can greatly improve on this performance.

Fiber Grating Sensors

Many types of fiber gratings can be used in sensing applications including Bragg, long-period, and

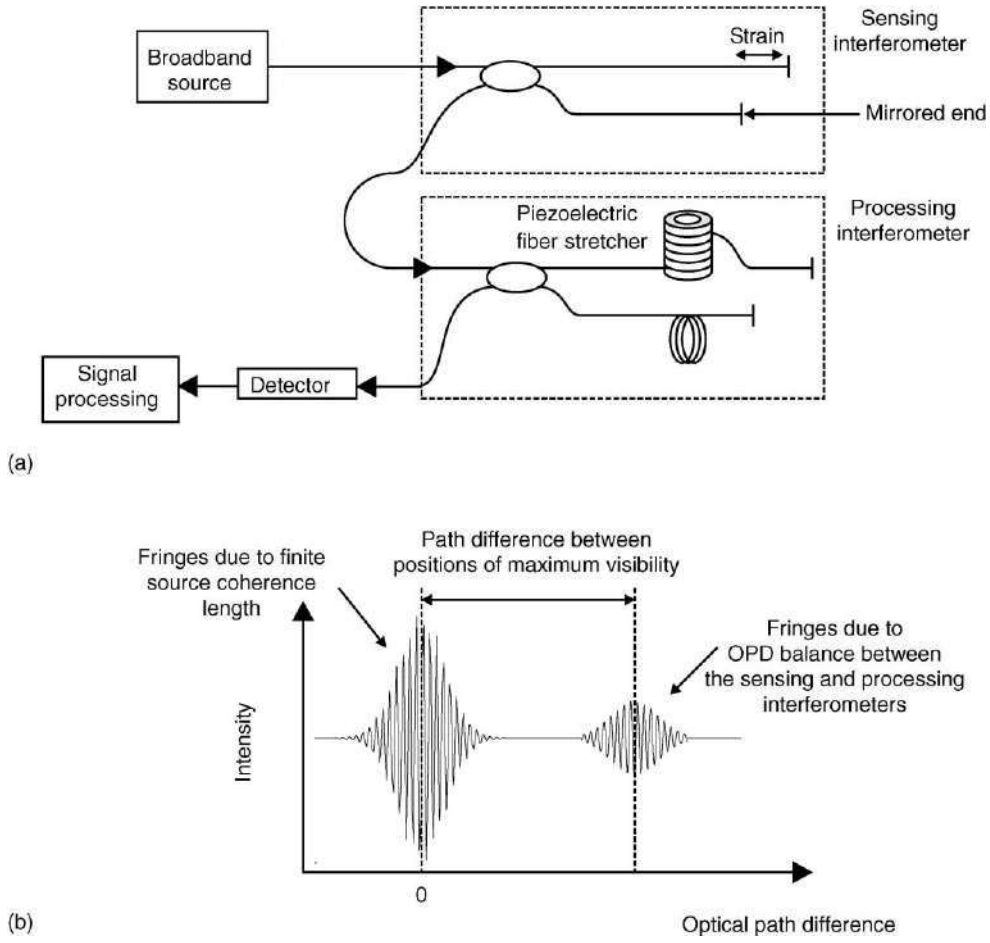

Figure 11 Low-coherence interferometric sensor. (a) Schematic diagram. (b) Output of the scanned processing interferometer.

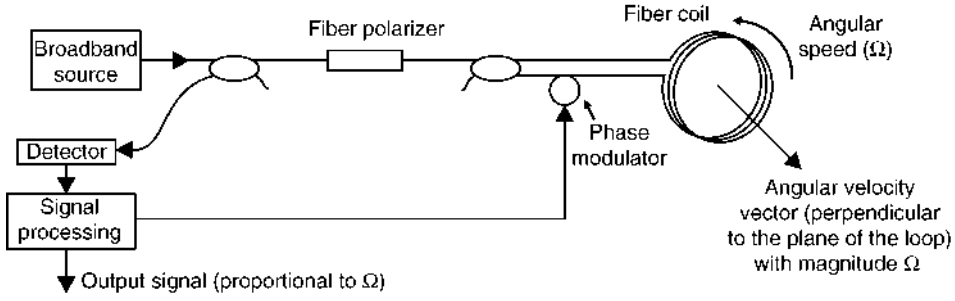

Figure 12 Basic open-loop fiber gyroscope.

chirped gratings. Because fiber gratings are small and have a narrow wavelength response, they can be used for both point and quasi-distributed sensing. They can be embedded in composite materials for smart structure monitoring and also in civil engineering structures such as bridges.

Fiber Bragg gratings (FBGs) are the most popular type of grating sensor. A typical FBG consists of a short section (typically a few mm) of single-mode fiber with a periodic modulation (typically 100s of nm) of the core refractive index. The index modulation causes light in the forward propagating core mode to be coupled into the backward core mode. This causes the FBG to act as a highly wavelength-selective rejection filter. The wavelength of peak reflectivity is the Bragg wavelength $\lambda_B = 2n_{eff}\Lambda$, where n_{eff} is the effective refractive index of the guided mode in the fiber and Λ the index modulation period. Both n_{eff} and Λ can be changed by an external measurand, resulting in a shift in λ_B.

The basic principle of FBG sensors is the measurement of an induced shift in the wavelength of an optical source due to a measurand, such as strain or temperature. A basic reflective FBG sensor system is shown in Figure 13. A broadband light source is used to interrogate the grating, from which a narrowband slice is reflected. The peak wavelength of the reflected spectrum can be compared to λ_B, from which strain or temperature can be inferred. The shift in the Bragg wavelength $\Delta\lambda_B$ with applied microstrain $\Delta(\mu\varepsilon)$ and change in temperature ΔT, for silica fiber is given approximately by

$$\frac{\Delta\lambda_B}{\lambda_B} \approx 0.78 \times 10^{-6}\Delta(\mu\varepsilon) + 6.67 \times 10^{-6}\Delta T \qquad [5]$$

A wavelength resolution of ~1 pm is required (at 1.3 μm) to resolve a temperature change of 0.1 °C or a strain change of 1 $\mu\varepsilon$. The response of the grating to strain can be improved through the use of specialized coatings. Polyimide coatings are commonly used as an efficient strain transducer for gratings embedded in composite materials. Because thermal effects in materials are usually very slow, it is relatively easy to measure dynamic strain (>1 Hz). However, in structural monitoring it can be necessary to distinguish between wavelength shifts due to static strain and those due to temperature. One technique is to use two collocated gratings whose response to strain and temperature is significantly different. In addition to strain and temperature measurement, grating sensors have also been used to measure flow, vibration, electromagnetic fields and chemical effects.

An example of quasi-distributed strain sensing, using a wavelength division multiplexed array of FBGs, is shown in Figure 14. Each FBG in the array has a unique Bragg wavelength. The return light from the FBG array is passed through a tunable narrowband Fabry–Perot filter. As the filter is tuned, the wavelengths returned by the individual FBGs can be analyzed and the strain present at each grating determined.

FBGs can be used as narrowband reflectors for creating fiber laser sensors, capable of measuring temperature, static strain, and very high-resolution dynamic strain. The basic form of an FBG laser sensor system shown in Figure 15 consists of a doped fiber section between two identical FBGs. The doped fiber is optically pumped to provide gain and thereby enable lasing to occur. Single-mode or multi-mode lasing is possible depending on the cavity length. In single-mode operation the FBG laser linewidth can be much smaller than the linewidth of diode lasers. This means that FBG laser sensors have greater sensitivities compared to passive FBG sensors.

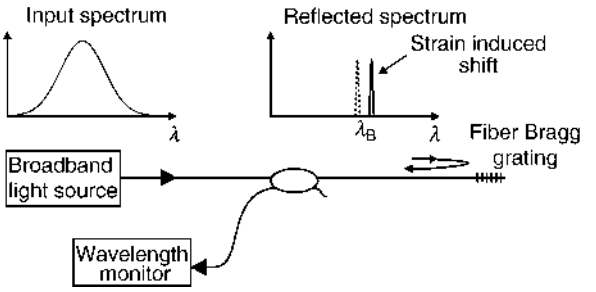

Figure 13 Basic reflective FBG sensor system.

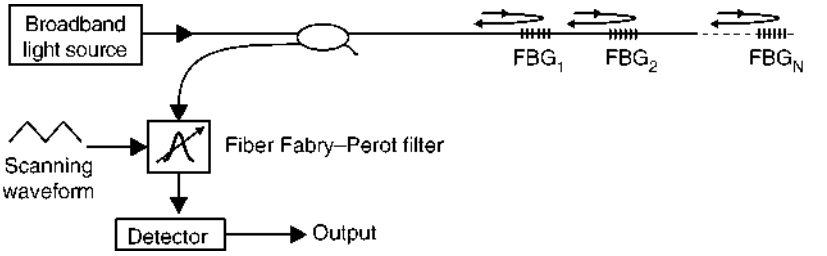

Figure 14 Quasi-distributed strain sensing using a wavelength division multiplexed array of FBGs.

When the cavity is subject to weak dynamic strain, the laser output is frequency modulated. This frequency modulation can be detected by using interferometric techniques. The main advantage of the FBG laser sensor over direct interferometry is that it is possible to obtain comparable strain sensitivities using a much shorter length of fiber.

Long-period fiber gratings (LPFGs) are attracting much interest for use in sensing applications. They are more sensitive to measurands than FBGs and easier to manufacture. A typical LPFG has a length of tens of mm with a grating period of 100s of μm. Its operation is different to an FBG in that coupling occurs between the forward propagating core mode and co-propagating cladding modes. The high attenuation of the cladding modes results in a series of minima occurring in the transmission spectrum of the fiber. This means that the spectral response is strongly influenced by the optical properties of the cladding and surrounding medium. This can be exploited for chemical sensing as shown in Figure 16, where a broadband source is used to interrogate an LPFG. The wavelength shifts of the output spectrum minima can be used to determine the concentration of particular chemicals in the substance surrounding the grating. The longest wavelength attenuation bands are the most sensitive to the refractive index of the substance surrounding the grating. This is because higher order cladding modes extend a greater distance into the external medium. LPFGs can also be used as strain, temperature, refractive index, bend, and load sensors.

Fiber Laser Doppler Velocimeter

Laser Doppler velocimetry (LDV) is a technique used for measuring velocity, especially of fluid flows. A fiber LDV system and associated scattering geometry is shown in Figure 17. In LDV two coherent laser beams intersect in a small measurement volume where they can interfere. The light reflected by a seed particle passing through the measurement volume is modulated at a frequency proportional to the spatial frequency (Doppler difference frequency Δf) of the interference fringes and the component of its velocity

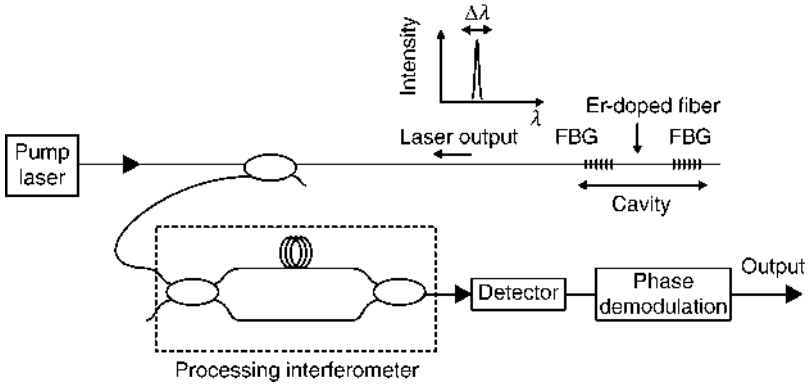

Figure 15 FBG laser sensor with interferometric detection.

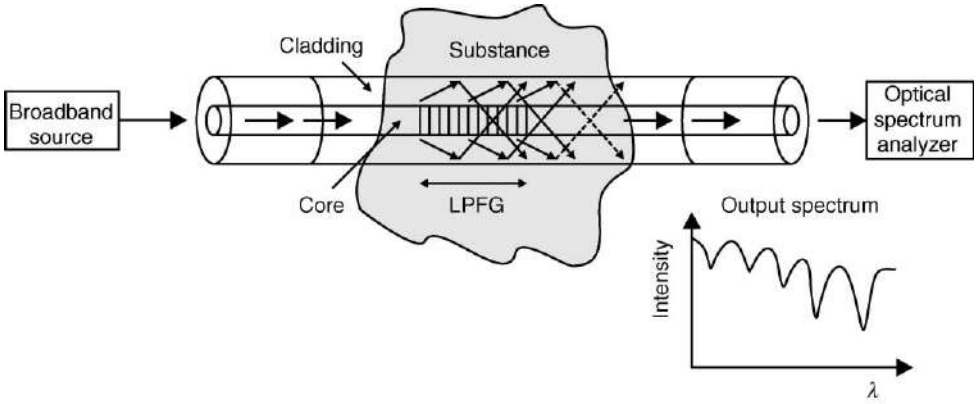

Figure 16 Detail of LPFG sensor.

normal to the interference fringes. Δf is given by

$$\Delta f = \frac{2nV\cos\beta}{\lambda_0}\sin(\theta/2) \qquad [6]$$

where n is the fluid refractive index, V the particle velocity and λ_0 the laser free-space wavelength. The output of the detector is processed to extract Δf and therefore $V\cos\beta$. Δf is independent of the scattered light direction so collection of the scattered light using a lens increases the system sensitivity. The form of eqn [6] indicates that the direction of flow cannot be ascertained. The simplest technique to resolve this ambiguity is to apply a frequency shift Δf_s to one of the input beams. This can be achieved by the use of a piezoelectric frequency shifter. The frequency shift causes a phase shift to appear between the two beams. The phase shift increases linearly with time. This results in a fringe pattern of spacing s, which moves with constant velocity $V_f = s\Delta f_s$. In this case the measured Δf will be less than or greater than Δf_s, depending on whether the particle is moving with or against the fringe motion. There will then be an unambiguous detectable range of velocities from zero to V_f.

Luminescence-Based Fiber Sensors

Luminescence-based fiber sensors are usually based on fluorescence or amplified spontaneous emission occurring in rare earth materials. They can be used in many applications such as chemical, humidity, and temperature sensing. It is possible to connect a fiber to luminescent material or to introduce luminescent dopants into the fiber. An example of the latter, used to detect chemical concentration is shown in Figure 18. A laser pulse causes the doped section of the fiber to luminesce at a longer wavelength than the

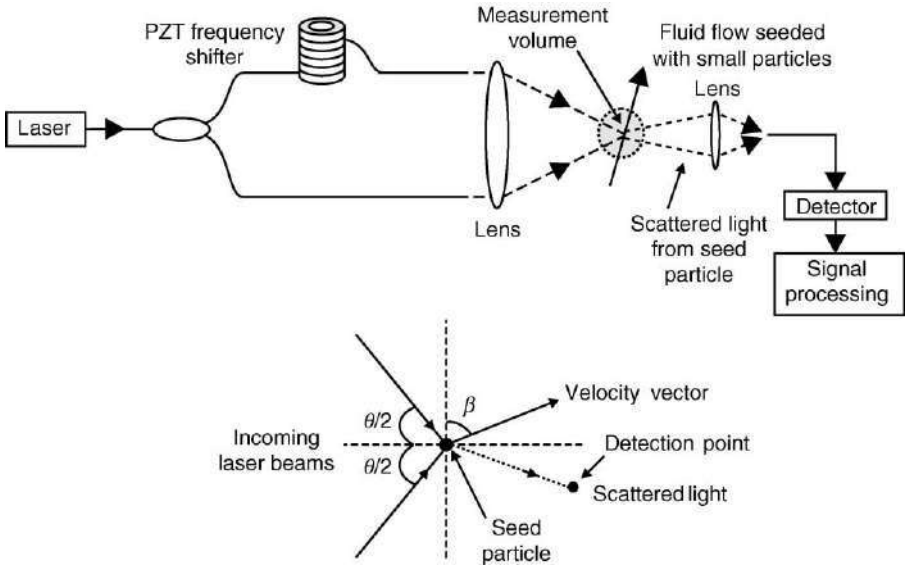

Figure 17 Fiber Doppler velocimeter and scattering geometry.

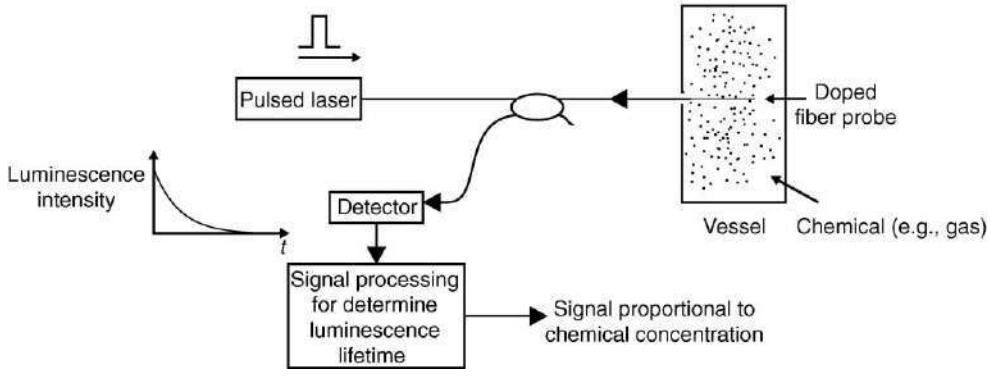

Figure 18 Chemical sensor based on fluorescence in doped optical fiber.

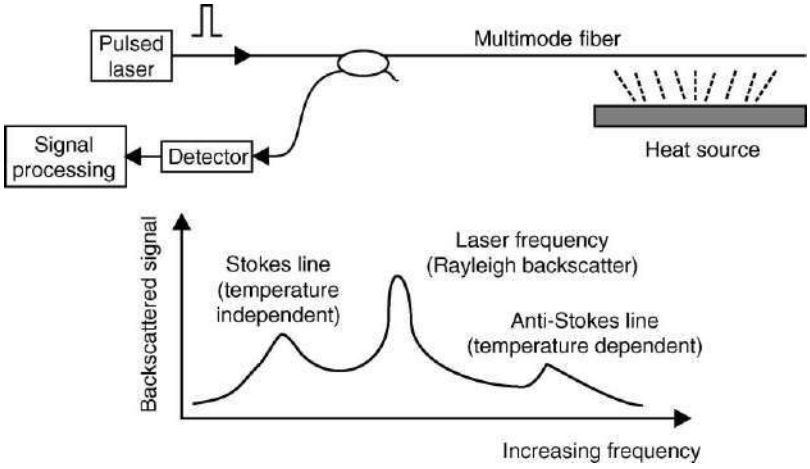

Figure 19 Anti-Stokes Raman thermometry system.

laser. The luminescence intensity *I(t)* at the detector has an exponential decay profile given by

$$I(t) = I_0 \exp\big(-(k_1 + k_2 C)t\big) \qquad [7]$$

where I_0 is the initial luminescence intensity, t time, k_1, k_2 constants, and C the chemical concentration. The luminescence time constant $1/(k_1 + k_2 C)$ can be determined by comparing the luminescence intensity at various times after excitation by the laser pulse, from which C can be determined. The use of time division multiplexing allows quasi-distributed measurement of chemical concentration. The use of plastic optical fiber for luminescence-based sensors is attracting much interest.

Distributed Fiber Sensing

We have seen that both point and quasi-distributed sensing are possible using fiber sensors. Distributed sensing can be achieved through the use of linear or nonlinear backscattering or forward scattering techniques. In linear backscattering systems, light backscattered from a pulse propagating in an optical fiber is time resolved and analyzed to obtain the spatial distribution of the measurand field, e.g., polarization optical time domain reflectometry analyzes the polarization state of backscattered light to determine the spatial distribution of electromagnetic fields. Nonlinear backscattering schemes use effects such as Raman or Brillouin scattering. An important example of the former is the anti-Stokes Raman thermometry system shown in Figure 19. A light pulse is transmitted down the sensing fiber. Spontaneous Raman scattering causes Stokes and anti-Stokes photons to be generated along the fiber. Some of these photons travel back along the fiber to a fast detector. The intensity of the Stokes line is temperature independent. The anti-Stokes line intensity is a function of temperature. The ratio of the two intensities provides a very accurate measurement of temperature. The measurement location is determined by timing of the laser pulse.

See also

Environmental Measurements: Laser Detection of Atmospheric Gases. **Fiber Gratings**. **Interferometry:** Overview; Phase Measurement Interferometry; White Light Interferometry. **Optical Materials:** Smart Optical Materials.

Further Reading

Culshaw B and Dakin JP (1988, 1989, 1997) *Optical Fiber Sensors*, vols. I, II, III. Norwood, MA: Artech House.

Grattan KTV and Meggitt BT (1994, 1997, 1999, 1999) *Optical Fiber Sensor Technology*, vols. I, II, III and IV. Boston, CA: Kluwer Academic.

Grattan KTV and Meggitt BT (2000) *Optical Fiber Sensor Technology*. Boston, MA: Kluwer Academic.

Kersey AD, Davis MA, Patrick HJ, *et al.* (1997) Fiber grating sensors. *Journal of Lightwave Technology* 15: 1442–1463.

Krohn DA (1999) *Fiber Optic Sensors: Fundamentals and Applications*, 3rd edn. Research Triangle Park, NC: Instrumentation Society of America.

Lopez-Higuera JM (2001) *Handbook of Optical Fiber Sensing Technology*. New York: Wiley.

Othonos A and Kalli K (1999) *Fiber Bragg Gratings: Fundamentals and Applications in Telecommunications and Sensing*. Norwood, MA: Artech House.

Udd E (1991) *Fiber Optic Sensors: An Introduction for Engineers and Scientists*. New York: Wiley.

Udd E (1995) *Fiber Optic Smart Structures*. New York: Wiley.

Yu FTS and Yin S (2002) *Fiber Optic Sensors*. New York: Marcel Dekker.

Heterodyning

T-C Poon, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Heterodyning, also known as frequency mixing, is a frequency translation process. Heterodyning has its root in radio engineering. The principle of heterodyning was discovered in the late 1910s by radio engineers experimenting with radio vacuum tubes. Russian cellist and electronic engineer Leon Theremin invented the so-called Thereminvox, which was one of the earliest electronic musical instruments that generates an audio signal by combining two different radio signals. American electrical engineer Edwin Armstrong invented the so-called super-heterodyne receiver. The receiver shifts the spectrum of the modulated signal, that is, the frequency contents of the modulated signal, so as to demodulate the audio information from it. Commercial amplitude modulation (AM) broadcast receivers nowadays are super-heterodyne type. After illustrating the basic principle of heterodyning by some simple mathematics, we will discuss some examples on how the principle of heterodyning is used in radio and in optics.

For a simple case, when signals of two different frequencies are heterodyned or mixed, the resulting signal produces two new frequencies, the sum and difference of the two original frequencies. Figure 1 illustrates how signals of two frequencies are mixed or heterodyned to produce two new frequencies, $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$, by simply multiplying the two signals $\cos(\omega_1 t + \theta_1)$ and $\cos(\omega_2 t)$, where ω_1 and ω_2 are radian frequencies of the two signals and θ_1 is the phase angle between the two signals. Note that when the frequencies of the two signals to be heterodyned are the same, i.e., $\omega_1 = \omega_2$, the phase information of $\cos(\omega_1 t + \theta_1)$ can be extracted to get $\cos\theta_1$ if we use an electronic lowpass filter (LPF) to filter out the term $\cos(2\omega_1 t + \theta_1)$. This is shown in Figure 2. Heterodyning is often referred to as homodyning for the mixing of two signals of the same frequency.

For a general case of heterodyning, we can have a signal represented by $s(t)$ with its spectrum $S(\omega)$, which is given by the Fourier transform of $s(t)$, and when it is multiplied by $\cos(\omega_2 t)$, the resulting spectrum is frequency-shifted to new locations in the frequency domain as:

$$\mathcal{F}\{s(t)\cos(\omega_2 t)\} = \frac{1}{2}S(\omega - \omega_2) + \frac{1}{2}S(\omega + \omega_2) \quad [1]$$

where $\mathcal{F}\{s(t)\} = S(\omega)$ and $\mathcal{F}\{\cdot\}$ denotes the Fourier transform of the quantity being bracketed. The spectrum of $s(t)\cos(\omega_2 t)$, along with the spectrum of $s(t)$, is illustrated in Figure 3. It is clear that multiplying $s(t)$ with $\cos(\omega_2 t)$ is a process of heterodyning, as we have translated the spectrum of the signal $s(t)$. This process is known as modulation in communication systems.

In order to appreciate the process of heterodyning let us now consider, for example, heterodyning in radio. In particular, we consider AM. While the frequency content of an audio signal (from 0 to around 3.5 kHz) is suitable to be transmitted over a pair of wires or coaxial cables, it is, however, difficult to be transmitted in air. By impressing the audio information onto a higher frequency, say 550 kHz, as one of the broadcast bands in AM radio, i.e., by

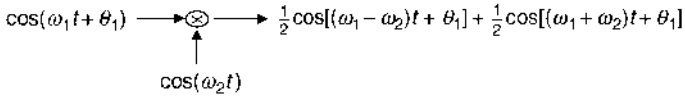

Figure 1 Heterodyning of two sinusoidal signals.

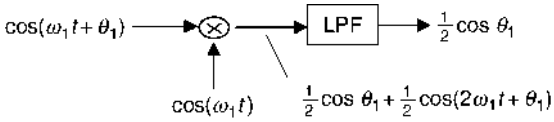

Figure 2 Heterodyning becomes homodyning when the two frequencies to be mixed are the same: homodyning allows the extraction of the phase information of the signal.

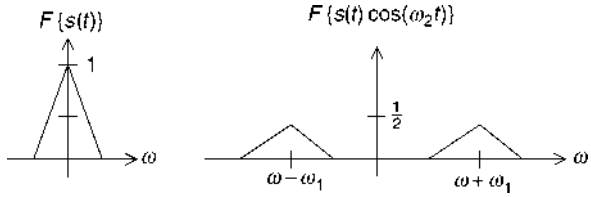

Figure 3 Spectrum of $s(t)$ and $s(t)\cos(\omega_2 t)$. It is clear that spectrum of $s(t)$ has been translated to new locations upon multiplying a sinusoidal signal.

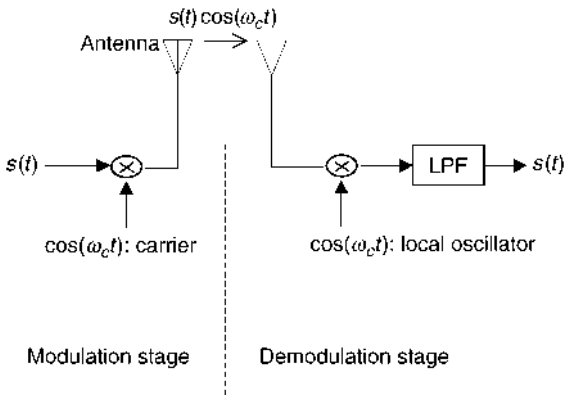

Figure 4 Radio link: heterodyning in the demodulation stage is often known as heterodyne detection.

modulating the audio signal, the resulting modulated signal can now be transmitted using antennas of reasonable dimensions. In order to recover (demodulate) the audio signal from the modulated signal, the modulated signal is first received by an antenna in a radio receiver, multiplied by the so-called local oscillator with the same frequency as the signal used to modulate the audio signal, then followed by a lowpass filter to eventually obtain the audio information back. The situation is illustrated in Figure 4. We shall now describe the process mathematically with reference to Figure 4.

We denote $s(t)$ as the audio signal. After multiplying by $\cos(\omega_c t)$, which is usually called a carrier in radio in the modulation stage, where ω_c is the radian frequency of the carrier. The modulated signal $s(t)\cos(\omega_c t)$. is transmitted via an antenna. When the modulated signal is received in the demodulation stage, the modulated signal is then multiplied by the local oscillator of signal waveform $\cos(\omega_c t)$. The output of the multiplier is given by

$$s(t)\cos(\omega_c t)\cos(\omega_c t) = s(t)\tfrac{1}{2}[1+\cos(2\omega_c t)] \qquad [2]$$

Note that the two signals in the demodulation stage, $s(t)\cos(\omega_c t)$ and $\cos(\omega_c t)$, are heterodyned to produce two new frequencies, the sum $\omega_c + \omega_c = 2\omega_c$ to give the term $\cos(2\omega_c t)$ and the difference $\omega_c - \omega_c = 0$ to give the term $\cos(0) = 1$. Since the two frequencies to be multiplied in the demodulation stage are the same, this is homodyning as explained above. Now by performing lowpass filtering (LPF), we can recover our original audio information $s(t)$. Note that if the frequency of the local oscillator in the demodulation stage within the receiver is higher than the frequency of the carrier used in the modulation stage, heterodyning in the receiver is referred to as superheterodyning, which most receivers nowadays use for amplitude modulation. In general, we have two heterodyning processes in the radio system just described, one in the modulation stage and the other in the demodulation stage. However, it is unusual to speak of heterodyning in the modulation stage, and so we just refer to the process in the demodulation stage as 'heterodyne detection.'

In summary, heterodyne detection can extract the information from a modulated signal and can also extract the phase information of a signal if homodyning is used. We shall see, in the next section, how optical heterodyne detection is employed.

Optical information is usually carried by coherent light such as a laser. Let $\psi_p(x, y)$ be a complex amplitude of the light field, which may physically represent a component of the electric field. We further assume that the light field is oscillating at temporal frequency ω_0. Therefore, we can write the light field as $\psi_p \exp(i\omega_0 t)$. By taking the real part of $\psi_p \exp(i\omega_0 t)$, i.e., $\mathrm{Re}[\psi_p \exp(i\omega_0 t)]$, we recover the usual physical real quantity. For a simple example, if we let $\psi_p = A\exp(-ik_0 z)$, where A is some constant and k_0 is the wavenumber of the light, $\mathrm{Re}[\psi_p \exp(i\omega_0 t)] = \mathrm{Re}[A\exp(-ik_0 z)\exp(i\omega_0 t)] = A\cos(\omega_0 t - k_0 z)$, which is a plane wave propagating along the positive z direction in free space. To detect light energy, we use a photodetector (PD), as shown in Figure 5. Assuming a plane wave for simplicity, as $\psi_p = A$, we have also taken $z = 0$ at the surface of the photodetector. As the photodetector responds to intensity, i.e., $|\psi_p|^2$, which gives the current, i, as output by spatially integrating the intensity:

$$i \propto \int_S |\psi_p \exp(i\omega_0 t)|^2 \mathrm{d}x\mathrm{d}y = A^2 S \qquad [3]$$

where S is the surface area of the photodetector. We can see that the photodetector current is proportional to the intensity, A^2, of the incident light. Hence the output current varies according to the intensity of the

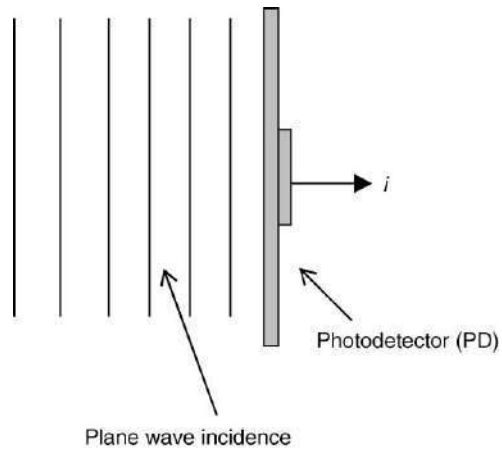

Figure 5 Optical direction detection or optical incoherent detection.

optical signal intensity. This mode of photodetection is called direct detection or incoherent detection in optics.

Let us now consider the heterodyning of two plane waves on the surface of the photodetector. We assume an information-carrying plane wave, also called the signal plane wave, $A_s \exp\{i[(\omega_0 + \omega_s)t + s(t)] \times \exp(-ik_0 x \sin\phi)\}$, and a reference plane wave, $A_r \exp(i\omega_0 t)$, or called a local oscillator in radio. The situation is shown in Figure 6.

Note that the frequency of the signal plane wave is ω_s higher than that of the reference signal, and it is inclined at an angle ϕ with respect to the reference plane wave, which is normal incident to the photodetector. Also, the information content $s(t)$ is in the phase of the signal wave. In the situation shown in Figure 6, we see that the two plane waves are interfered on the surface of the photodetector, giving the total light field ψ_t, given by

$$\psi_t = A_r \exp(i\omega_0 t) + A_s \exp\{i[(\omega_0 + \omega_s)t + s(t)]\} \times \exp(-ik_0 x \sin\phi) \quad [4]$$

Again, the photodetector responds to intensity, giving the current

$$i \propto \int_S |\psi_t|^2 \mathrm{d}x\mathrm{d}y = \int_{-a}^{a}\int_{-a}^{a} [A_r^2 + A_s^2 + 2A_rA_s \cos(\omega_s t + s(t) - k_0 x \sin\phi)]\mathrm{d}x\mathrm{d}y \quad [5]$$

where we have assumed that the photodetector has a $2a \times 2a$ square area. The current can be evaluated to be:

$$i(t) \propto 2a(A_r^2 + A_s^2) + 4A_rA_s \frac{\sin(k_0 a \sin\phi)}{k_0 \sin\phi} \cos(\omega_s t + s(t)) \quad [6]$$

The current output has two parts: the DC current and the AC current. The AC current at frequency ω_s is commonly known as the heterodyne current.

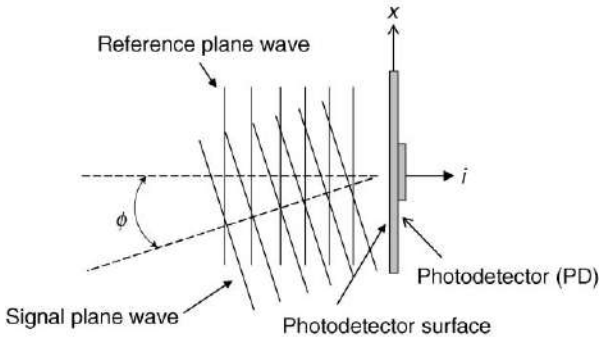

Figure 6 Optical heterodyne detection.

Note that the information content $s(t)$ originally embedded in the phase of the signal plane wave has now been preserved and transferred to the phase of the heterodyne current. The above process is called optical heterodyning. In optical communications, it is often referred to as optical coherent detection. In contrast, if the reference plane wave has not been used for the detection, we have the incoherent detection. The information content carried by the signal plane wave would be lost, as it is evident that for $A_r = 0$, the above equation gives only a DC current at a value proportional to the intensity of the plane wave, A_s^2.

Let us now consider some of the practical issues encountered in coherent detection. Again, the AC part of the current given by the above equation is the heterodyne current $i_{het}(t)$, given by

$$i_{het}(t) \propto A_rA_s \frac{\sin(k_0 a \sin\phi)}{k_0 \sin\phi} \cos(\omega_s t + s(t)) \quad [7]$$

We see that since the two plane waves propagate in slightly different directions, the heterodyne current output is degraded by a factor of

$$\frac{\sin(k_0 a \sin\phi)}{k_0 \sin\phi} = a\mathrm{sinc}(k_0 a \sin\phi)$$

where $\mathrm{sinc}(x) = \sin(x)/x$. For small angles, i.e., $\sin\phi \approx \phi$, the current amplitude falls off as $\mathrm{sinc}(k_0 a\phi)$. Hence, the heterodyne current is at a maximum when the angular separation between the signal plane wave and the reference plane wave is zero, i.e., the two plane waves are propagating exactly in the same direction. The current will go to zero when $k_0 a\phi = \pi$, or $\phi = \lambda_0/2a$, where λ_0 is the wavelength of the laser light. To see how critical it is for the angle ϕ to be aligned in order to have any heterodyne output, we assume the size of the photodetector $2a = 1$ cm and the laser used is red, i.e., $\lambda_0 \approx 0.6$ μm; ϕ is calculated to be about 2.3×10^{-3} degrees. Hence to be able to work with coherent detection, we need to have precise optomechanical mounts for angular rotation.

Finally we describe a famous application of heterodyning in optics – holography. As we have seen, the mixing of two light fields with different temporal frequencies will produce heterodyne current at the output of the photodetector. But we can record the two spatial light fields of the same temporal frequency with photographic films instead of using electrical devices. Photographic films respond to light intensity as well. We discuss holographic recording of a point source object as a simple example. Figure 7 shows a collimated laser split into two plane waves and recombined by the use of two mirrors (M) and two beamsplitters (BS). One plane wave is used to illuminate the pinhole aperture (our point source object), and the other

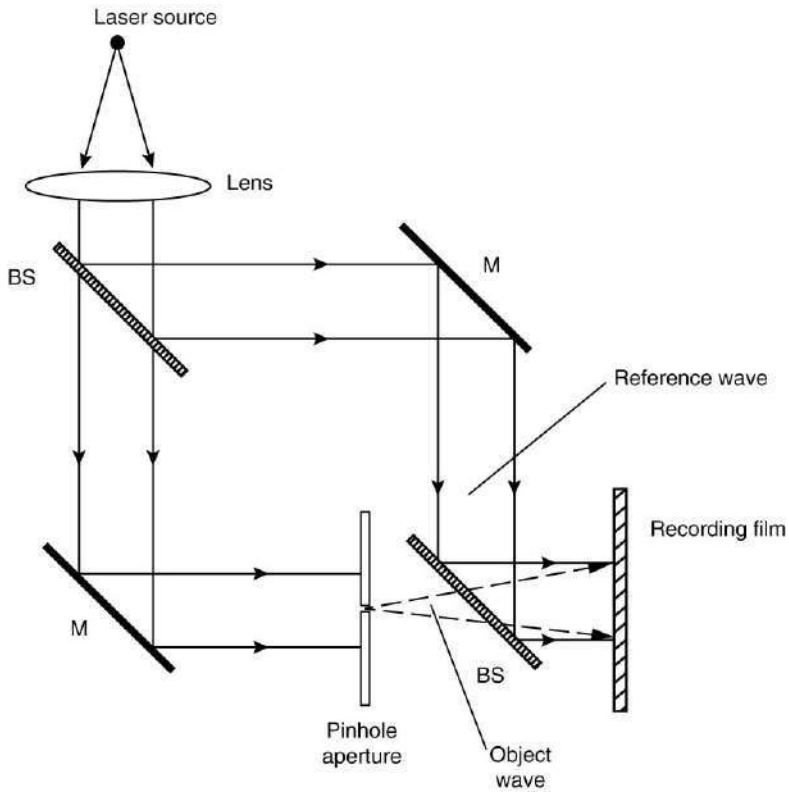

Figure 7 Holographic recording of a point source object.

illuminates the recording film directly. The plane wave that is scattered by the point source object, which is located z_0 away from the film, generates a diverging spherical wave. This diverging wave is known as an object wave in holography. The object wave arising from the point source object on the film is given by, according to Fresnel diffraction:

$$\psi_0 = A_0 \exp(-ik_0 z_0) \frac{ik_0}{2\pi z_0} \times \exp[-ik_0(x^2+y)^2/2z_0] \exp(i\omega_0 t) \quad [8]$$

This object wave is a spherical wave, where A_0 is the amplitude of the spherical wave.

The plane wave that directly illuminates the photographic plate is known as a reference wave, ψ_r. For the reference plane wave, we assume that the plane wave has the same phase with the point source object at a distance z_0 away from the film. Its field distribution on the film is, therefore, $\psi_r = A_r \exp(-ik_0 z_0)\exp(i\omega_0 t)$, where A_r is the amplitude of the plane wave. The film now records the interference of the reference wave and the object wave, i.e., what is recorded on the film as a 2D pattern is given by $t(x, y) \propto |\psi_r + \psi_0|^2$. The resulting recorded 2D pattern, $t(x, y)$, is called the hologram of the object. This kind of recording is known as holographic recording, distinct from a photographic recording in which the reference wave does not exist and hence only the object wave is recorded. Now:

$$\begin{aligned} t(x,y) &\propto |\psi_r + \psi_0|^2 \\ &= \left| A_r \exp(-ik_0 z_0)\exp(i\omega_0 t) + A_0 \exp(-ik_0 z_0) \right. \\ &\quad \left. \times \frac{ik_0}{2\pi z_0} \exp[-ik_0(x^2+y)^2/2z_0] \exp(i\omega_0 t) \right|^2 \\ &= A + B \sin\left\{ \frac{k_0}{2z_0}[(x^2+y)^2/2z_0] \right\} \end{aligned} \quad [9]$$

where A and B are some inessential constants. Note that the result of recording two spatial light fields of the same temporal frequency preserves the phase information (noticeably the depth parameter z_0) of the object wave. This is considered optical homodyning as it is clear from the result shown in Figure 2.

The intensity distribution being recorded on the film, upon being developed, will have transmittance given by the above equation. This expression is called the Fresnel zone plate, which is the hologram of a point source object and we shall call it the point-object hologram. Figure 8 shows the hologram for a particular value of z_0 and k_0. The importance of this phase-preserving recording is that when we

illuminate the hologram with a plane wave, called the reconstruction wave, a point object is reconstructed at the same location of the original point source object if we look towards the hologram as shown in Figure 9, that is the point source is reconstructed at a distance z_0 from the hologram as if it were at the same distance away from the recording film. For an arbitrary 3D object, we can imagine the object as a collection of points, and therefore, we can see that we have a collection of Fresnel zone plates on the hologram. Upon reconstruction, such as the illumination of the hologram by a plane wave, the observer would see a 3D virtual object behind a hologram.

As a final example, we discuss a state-of-the-art holographic recording technique called optical scanning holography, which employs the use of optical heterodyning and electronic homodyning to achieve real-time holographic recording without the use of films. We will take the holographic recording of a point object as an example. Suppose we superimpose a plane wave and a spherical wave of different temporal frequencies and use the resulting intensity pattern as an optical scanning beam to raster scan a point object located z_0 away from the point source that generates the spherical wave. The situation is shown in Figure 10. Point C is the point source that generates the spherical wave (shown with dashed lines) on the pinhole object, our point object. This point source, for example, can be generated by a focusing lens. The solid parallel rays represent the plane wave. The interference of the two waves generates a Fresnel zone plate type pattern on the pinhole object:

$$
\begin{aligned}
&I_s(x, y;\ z_0, t) \\
&= \left| A_r \exp(-ik_0 z_0) \exp(i\omega_0 t) + A_0 \exp(-ik_0 z_0) \frac{ik_0}{2\pi z_0} \right. \\
&\quad \left. \times \exp[-ik_0(x^2 + y)^2/2z_0] \exp[i(\omega_0 + \Omega)t] \right|^2 \\
&= 1 + \left(\frac{1}{\lambda_0 z_0}\right)^2 + \frac{1}{\lambda_0 z_0} \sin\left[\frac{\pi}{\lambda_0 z_0}(x^2 + y^2) - \Omega t\right]
\end{aligned}
\qquad [10]
$$

assuming $A_r = A_0 = 1$ for simplicity. Note that the plane wave is at temporal frequency ω_0, and the spherical wave is at temporal frequency $\omega_0 + \Omega$. The expression $I_s(x, y;\ z_0, t)$ is a temporally modulated Fresnel zone plate and is known as the time-dependent Fresnel zone plate (TDFZP). If we freeze time, say at $t = t_0$, we have the Fresnel zone plate pattern on the pinhole object as shown in Figure 10. However, if we let the time run in the above expression, physically we will have running zones that would be moving away from the center of the pattern. These running zones are the result of optical heterodyning of the plane wave and the spherical wave of different temporal frequencies. It is this TDFZP that is used to raster scan a 3D object to obtain holographic information of the scanned object and such technique is called optical scanning holography.

Upon scanning by the TDFZP, the photodetector, captures all the transmitted light and delivers a current, which consists of a heterodyne current at

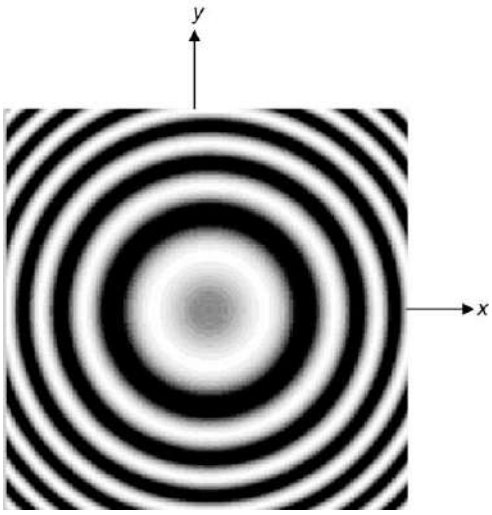

Figure 8 Point-object hologram.

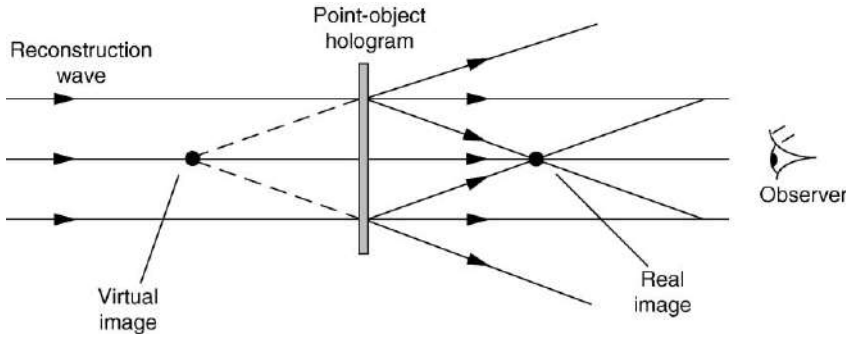

Figure 9 Reconstruction of a point-object hologram.

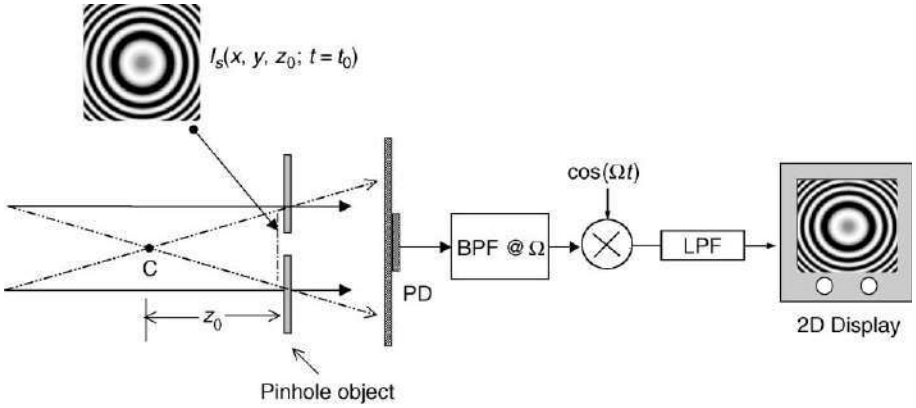

Figure 10 Optical scanning holography (use of optical heterodyning and electronic homodyning to record holographic information upon scanning the object in two dimensions).

frequency Ω. After electronic bandpass filtering (BPF) at Ω, the heterodyne current is homodyned electronically by cos(Ω*t*) and lowpass filtered (LPF) to extract the phase information of the current. When this final scanned and processed current is display in a 2D display, as shown in Figure 10, we have the Fresnel zone plate, which is the hologram of the pinhole object, on the display. We can photograph this 2D display to obtain a transparency and have it illuminated by a plane wave to have the reconstruction as shown in Figure 9, or, since the hologram is now in electronic form, we can store it in a PC and reconstruct it digitally. When holographic reconstruction is performed digitally, we have so-called digital holography.

See also

Diffraction: Fresnel Diffraction. **Holography, Techniques:** Digital Holography.

Further Reading

Banerjee PP and Poon TC (1991) *Principles of Applied Optics*. MA: Irwin.

Palais JC (1988) *Fiber Optic Communications*. New Jersey: Prentice-Hall.

Poon TC and Banerjee PP (2001) *Contemporary Optical Image Processing with MATLAB*. Oxford: Elsevier.

Poon TC and Qi Y (2003) Novel real-time joint-transform correlation using acousto-optic heterodyning. *Applied Optics* 42: 4663–4669.

Poon TC, Wu MH, Shinoda K and Suzuki Y (1996) Optical scanning holography. *Proc. IEEE* 84: 753–764.

Pratt WK (1969) *Laser Communication Systems*. New York: Wiley.

Stremler FG (1990) *Introduction to Communication Systems*. MA: Addison-Wesley.

VanderLugt A (1992) *Optical Signal Processing*. New York: Wiley.

Image Post-Processing and Electronic Distribution

K M Iftekharuddin, The University of Memphis, Memphis, TN, USA

F Ahmed, The Catholic University of America, Washington, DC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The spectacular development of imaging sensors, communication, and internet infrastructure has immensely facilitated the creation, processing, and distribution of electronic digital images. One nontrivial aspect of the digital image is its complete application dependence. The image may be created using a number of different acquisition methods and sensors. It is then processed in different ways, depending upon myriad of digital imaging applications. Consequently, the effectiveness of a useful electronic image processing engine may involve development of robust pre-processing techniques, different types of goal-directed processing such as detection, recognition, identification, classification, tracking, reconstruction and/or registration, as well as post-processing algorithms for different application areas such as multimedia, biomedical, astronomical, defense, consumer, industrial, etc. A notional schematic of this electronic image processing cycle is shown in Figure 1. Note that

while the goal-directed techniques may be different, based on problems at hand, the pre- and post-processing techniques and tools are very similar. Thus, most of the common image pre-processing techniques are also applicable for image post-processing purposes.

Electronic image post-processing techniques involve three broad domains such as: (i) spatial; (ii) frequency; and (iii) time-frequency. Each of these domain techniques may be grouped into different types of processing such as: (i) spatial domain technique to include filtering, spatial processing such as histogram modification, morphological processing, texture processing, etc.; (ii) frequency domain technique to include filter; and (iii) time-frequency domain technique to include short-time Fourier transform, wavelet, and Wigner-ville transforms.

On the other hand, it is useful to identify the typical lifecycle of a digital image from the distribution perspective as shown in Figure 2. The image may be created using a number of different acquisition methods and sensors as mentioned above. The processing block next combines all different types of processing, as mentioned in Figure 1. The image may then be transmitted over the internet or it can leave the digital world to go to the print world. It may then go through wear/tear, subsequently being scanned back to digital domain again. After some processing, it may again be distributed over the internet.

Consequently, the image may be copied legally or fraudulently, processed, and used by a user. Further, the threats and vulnerabilities of the internet have also increased proportionately. The rapid development and availability of high-end computers, printers, and scanners has made the tools of counterfeiting easily accessible and more affordable. With these products, one may counterfeit digital images, which are easy to distribute and modify. As a result, copyright ownership poses a formidable challenge to image post-processing and distribution.

In the subsequent sections, we briefly review relevant background in image post-processing and distribution techniques. We also present a few representative application examples of image post-processing and copyright protection in image distribution.

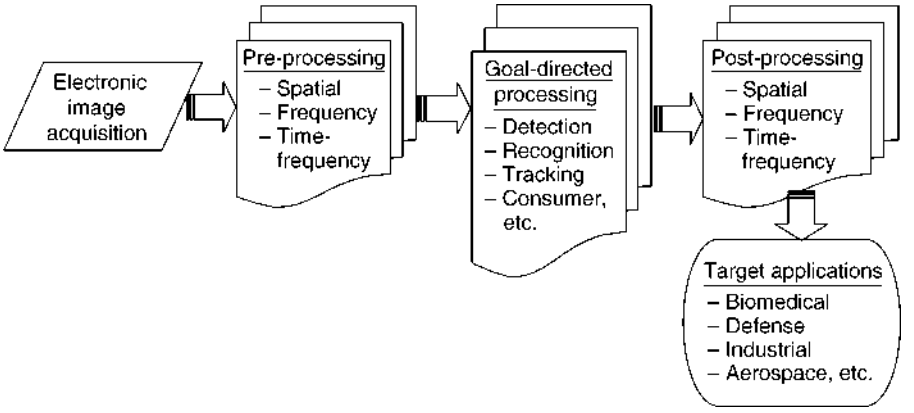

Figure 1 Notional schematic of electronic image processing.

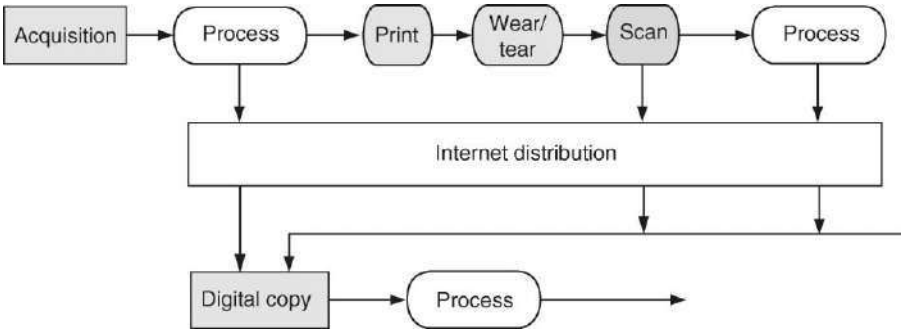

Figure 2 Typical lifecycle of a digital image distribution cycle.

Techniques in Image Post-Processing

Image Formation

We begin by describing the image formation technique briefly. Consider that a three-dimensional (3D) object or scene is imaged onto a 2D imaging plane by means of a recording system such as a camera. If the image formation process is linear, the recorded image may be modeled as the output of the system shown in Figure 3, which is given mathematically by

$$g(x,y) = s\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y;s,t)f(s,t)\mathrm{d}s\,\mathrm{d}t\right\} + n(x,y) \quad [1]$$

where $g(x,y)$ denotes the recorded output image, and $f(x,y)$ is the ideal electronic image. The function $h(x,y;s,t)$ is the 2D impulse response or the point-spread function (PSF) of the image formation system. The PSF determines the radiant energy distribution in the image plane due to a point source located in the object plane. Usually the PSF is normalized such that

$$\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y;s,t)\mathrm{d}s\,\mathrm{d}t\right\} = 1 \quad [2]$$

The noise contribution $n(x,y)$ is shown as an additive random process that is statistically uncorrelated with the image. This is a simplification because noises such as film-grain noise and the noise caused by photon statistics, which often corrupt images, are not uncorrelated with the input. This simplification nonetheless leads to reasonable and useful results. One of the primary goals of image pre-processing (and post-processing as well) is to devise techniques that reduce this additive random noise.

The PSF is a function of only the argument differences $x - s$ and $y - t$ for stationary image and object fields. Thus, it is possible to envision an image formation system that is space invariant; hence the PSF is independent of position. In this case, we may rewrite the superposition integral in eqn [1] into a more familiar convolution integral as follows:

$$g(x,y) = \left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-s,y-t)f(s,t)\mathrm{d}s\,\mathrm{d}t\right\} + n(x,y) \quad [3]$$

Equivalently, eqn [3] is written as

$$g(x,y) = \{h(x,y)^{*}f(x,y)\} + n(x,y) \quad [4]$$

where $(*)$ is used to denote 2D convolution. Converting from a continuous image $f(x,y)$ to its discrete representation $f(i,j)$ requires the process of sampling. In the ideal sampling system, $f(x,y)$ is multiplied by an ideal 2D impulse train given as

$$f(i,j) = \sum_{m=-\infty}^{+\infty}\sum_{m=-\infty}^{+\infty} f(mX_0,nY_0)\,\delta(x-mX_0,y-Y_0) \quad [5]$$

where X_0 and Y_0 are the sampling distances or intervals, $\delta(^{*},^{*})$ is the ideal impulse function. Note that the square sampling implies that $X_0 = Y_0$. Sampling with an impulse function corresponds to sampling with an infinitesimally small point. If one is interested in image processing, one should choose a sampling density based upon the classical Nyquist sampling theory. Thus, in discrete domain implementation, eqns [3] and [4] are expressed as

$$g(i,j) = \left\{\sum_{\forall(k,l)} h(i,j;k,l)f(k,l)\right\} + n(i,j),\ 0 \le i,j \le N-1 \quad [6]$$

and,

$$g(i,j) = s\left\{\sum_{\forall(k,l)} h(i-k,j-l)f(k,l)\right\} + n(i,j) \quad [7]$$

Thus, we may rewrite eqn [6] in discrete domain as follows:

$$g(i,j) = h(i,j)^{*}f(i,j) + n(i,j) \quad [8]$$

where $(*)$ represents convolution operation. Equation [8] also describes the spatial domain filtering wherein $h(x,y;s,t)$ may be considered as the spatial domain filter mask. The discrete Fourier transform may be used to yield the frequency domain model as follows:

$$G(u,v) = H(u,v)F(u,v) + N(u,v) \quad [9]$$

where G, H, F, and N are corresponding Fourier transforms and u and v are the dummy variables. In eqn [9], $H(u,v)$ is again the frequency domain filter.

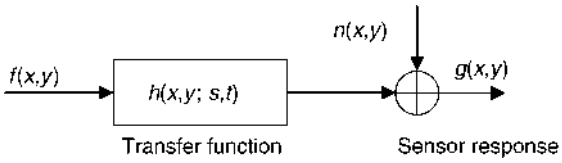

Figure 3 Model for image formation and recording.

We discuss the frequency domain filtering concept further, later in this article.

Spatial Domain Processing

The image pixel domain operations are frequently described as the spatial domain processing. These techniques are important for image pre- as well as post-processing operations. We describe two representative spatial domain processing techniques such as histogram processing and fractal processing for texture analysis below.

Histogram processing

One of the important issues in post-processing is image enhancement. Distortions in image may be introduced due to different noise factors, relative motion between a camera and the object, a camera that is out of focus, or atmospheric turbulence. Noise may be due to thermal measurement errors, recording medium, transmission medium, or digitization process. The histogram processing operation is one of the simplest techniques for enhancing the distorted image for certain type of noises. Histogram equalization is a contrast enhancement technique with the objective to obtain a new enhanced image with a uniform histogram. This can be achieved by using the normalized cumulative histogram as the gray scale mapping function. Histogram modeling is usually introduced using continuous domain functions. Consider that the images of interest contain continuous intensity levels in the interval [0,1] and that the transformation function τ which maps an input image $f(x, y)$ onto an output image $w(x, y)$ is continuous within this interval. We assume that the transfer function, $K_f = \tau(K_w)$, is single-valued and monotonically increasing such that it is possible to define the inverse law $K_w = \tau^{-1}(K_f)$. An example of such a transfer function is illustrated in Figure 4. All pixels in the input image, with densities in the interval D_f to $D_f + dD_f$, are mapped using the transfer function $K_f = \tau(K_w)$, such that they assume an output pixel density value in the interval from D_w to $D_w + dD_w$. Consider that if the histogram h is regarded as a continuous probability density function p, describing the distribution of the intensity levels, then we obtain:

$$p_w(w) = p_f(D_f)/d(D_f) \qquad [10]$$

where

$$D_f = du(f)/df \qquad [11]$$

In the case of histogram equalization, the output probability densities are all an equal fraction of the maximum number of intensity levels in the input image, thus generating a uniform intensity distribution. Sometimes it is desirable to control the shape of the output histogram in order to enhance certain intensity levels in an image. This can be accomplished by the histogram specialization operator which maps a given intensity distribution into a desired distribution using a histogram equalized image as an intermediate stage.

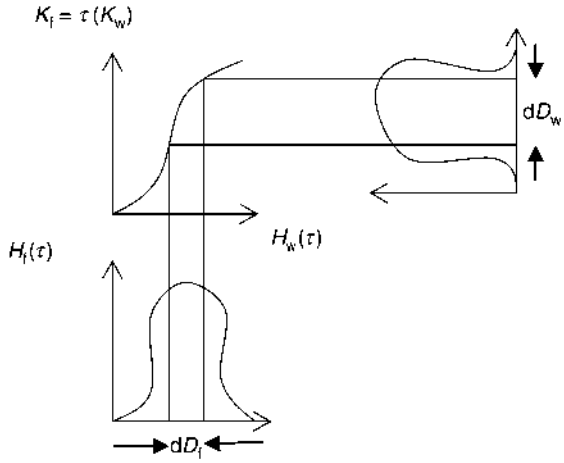

Figure 4 A histogram transformation function.

Fractals for Texture Analysis

The fractal aspect of an image has successfully been exploited for image texture analysis. The fractal concept developed by Mandelbrot, who coined the term fractal from the Latin 'fractus', provides a useful tool to explain a variety of naturally occurring phenomena. A fractal is an irregular geometric object with an infinite nesting of structure at all scales. Fractal objects can be found everywhere in nature such as coastlines, fern trees, snowflakes, clouds, mountains, and bacteria. Some of the most important properties of fractals are self-similarity, chaos, and noninteger fractal dimension (FD). The FD of an object can be correlated with the object properties. Object textures in images are candidates for characterization using fractal analysis because of their highly complex structure. The fractal theory developed by Mandelbrot is based, in part, on the work of mathematicians Hausdorff and Besicovitch. The Hausdorff–Besicovitch dimension, D_H, is defined as:

$$D_H = \lim_{\varepsilon \to 0^+} \frac{\ln N}{\ln 1/r} \qquad [12]$$

where N is the number of elements of box size r required to form a cover of the object. The FD can be defined as the exponent of the number of self-similar pieces (N) with magnification factor ($1/r$) into which a figure may be broken. The FD is a noninteger

value in contrast to objects that lie strictly in Euclidean space. The equation for FD is as follows:

$$\mathrm{FD} = \frac{\ln\,(\text{number of self-similar pieces})}{\ln\,(\text{magnification factor})} = \frac{\ln N}{\ln(1/r)} \qquad [13]$$

Frequency Domain Image Filtering

Ignoring noise contribution, we may obtain the basic filtering expression using eqns [7] and [8] as follows:

$$f(i, j) = \Im^{-1}[G(u, v)/H(u, v)] \qquad [14]$$

where $\Im^{-1}$ is the inverse Fourier transform and $[1/H(u, v)]$ is the inverse frequency domain filter. The key importance of this result is that, instead of performing a convolution as shown in eqn [7] to obtain the output of the system, we obtain the inverse of the Fourier transform of the impulse response $h(x, y; s, t)$ and the output image $g(x, y)$, multiply them and then take the inverse Fourier transform of the product to obtain $f(x, y)$. This is the basis for all the filtering as well as other types of image post-processing such as image reconstruction works. The filter function in eqn [9] may have different forms and shapes based on different types of functions such as uniform, Gaussian, lowpass, highpass, etc. While these are all linear filters, there are examples of nonlinear filters such as median, mean, etc. that are equally applicable for image post-processing.

Time-Frequency Domain Processing

One of the fundamental aspects of early image representation is the notion of scale. An inherent property of objects in the world and details in image is that they only exist as meaningful entities over certain ranges of scale. Consequently, a multiscale representation is of crucial importance if one aims at describing the structure of the world, or more specially the structure of projections of the three-dimensional world onto two-dimensional electronic images. The multiscale image post-processing may be facilitated by the wavelet (WT) computation.

A wavelet is, by definition, an amplitude-varying short waveform with a finite bandwidth. Naturally, wavelets are more effective than the sinusoids of Fourier analysis for matching and reconstructing signal features. In wavelet transformation and inversion, all transient or periodic data features can be detected and reconstructed by stretching or contracting a single wavelet to generate the matching building blocks. Consequently, wavelet analysis provides many flexible and effective ways to post-process images that surpass classical techniques – making it very attractive for data analysis, and modeling. The continuous WT is defined as

$$\Phi(a, b) = |a|^{-1/2} \int_{-W}^{W} x(t)\Psi\left[\frac{t-b}{a}\right]\mathrm{d}t \qquad [15]$$

where a and b are scaling and translation factors and $\Psi(t)$ is the mother wavelet. The WT may be viewed as a multiresolution analysis with shifted and scaled versions of the mother wavelet. Corresponding WT coefficients $c_{m,n}$ are

$$c_{m,n} = |a_0|^{-m/2} \int x(t)\Psi([a_0^{-m}t - nb_0])\mathrm{d}t \qquad [16]$$

The reconstruction of the signal is achieved using the admissibility condition, $\int \Psi(t)\mathrm{d}t = 0$, as follows:

$$x(t) = c \int\int_{a>0} \left[\Phi(a, b)\frac{\Psi a, b(t)}{a}\right]\mathrm{d}a\,\mathrm{d}b \qquad [17]$$

where c is a constant. For compact dyadic wavelets, binary scaling and dyadic translation of the mother wavelet form the basis functions. A dyadic translation is a shift by the amount $n/2^m$ that is an integer multiple of the binary scale factor and of the width of the wavelet as well. The discrete WT (DWT) is the natural choice for digital implementation of WT. The DWT basis and 'sub-band coding' process for discretization of a and b are identified. The iterative bandpass filter that finally leads to the wavelet is given as

$$\Psi m, n(x) = 2^{m/2}\Psi(2^{-m}x - n) \qquad [18]$$

The above eqn [18] forms the basis for 'sub-band' implementation for multiresolution analysis (MRA) of wavelet processing.

Image Post-Processing Examples

Some of the important application areas for electronic image post-processing involve the following:

(i) Biomedical image analysis and object detection;
(ii) Image compression for transmission and distribution;
(iii) Satellite image processing and noise removal;
(iv) IR/SAR/detection and clutter removal; and
(v) Image classification and machine vision.

In this article, we provide examples of some of these image-post-processing applications.

Example I: Spatial Domain Processing: Image Post-Processing in Biomedical Image

The mega voltage X-ray images are the X-ray images taken using high voltage X-radiations. These images are often characterized by low resolution, low contrast, and blur. The importance of mega voltage X-ray images arises in the context of radiation therapy for cancer treatment wherein the mega voltage X-radiation is used as a therapeutic agent. Figure 5 shows such mega voltage X-ray images, also known as an electronic portable digital (EPID) image with a very low level of detail. Figure 5a shows poor dynamic range. In order to improve the contrast of this image, without affecting the structure (i.e., geometry) of the information contained therein, we can apply the histogram equalization operation discussed in eqn [10]. Figure 5b presents the result of the histogram image enhancement post-processing process. The low level of detail in the bony anatomy is easily visible in Figure 5b.

Example II: Frequency Domain Processing: Image Post-Processing in Biomedical Tissue Detection

In medical imaging, texture is a fundamental characteristic that can be exploited to analyze tissues and pathologies in a biomedical image. The image texture is often characterized by the random nature of the objects such as tissues and body structures, noise, imaging modality, imaging limitations, and a variety of measurement parameters affecting the image acquisition process. Extensive research suggests that the fractal analysis may be combined with MRA techniques to post-process the texture content of an image in general and tumor detection in particular. In addition, the stochastic fractal Brownian motion (fBm) is well suited to describe texture and tumor characteristics since fBm offers an elegant integration of fractal and multiresolution analysis.

The fBm is a part of the set of $1/f$ processes, corresponding to a generalization of the ordinary Brownian motion $B_{\mathrm{H}}(s)$. It is a nonstationary zero-mean Gaussian random function, defined as

$$B_{\mathrm{H}}(t) - B_{\mathrm{H}}(s) = \frac{1}{\Gamma(H+0.5)} \times \left\{ \int_{-\infty}^{0} [(t-s)^{H-0.5} - s^{H-0.5}] \mathrm{d}B(s) + \int_{0}^{\infty} (t-s)^{H-0.5} \mathrm{d}B(s) \right\}, \quad B_{\mathrm{H}}(0) = 0 \qquad [19]$$

where the Hurst coefficient H, restricted to $0 < H < 1$, is the parameter that characterizes fBm; t and s correspond to different observation times of the process B_{H}, and Γ to the Euler's Gamma function. From the frequency domain perspective, it is known that some of the most frequently seen structures in fractal geometry, generally known as $1/f$ processes, show a power spectrum satisfying the power law relationship:

$$S(\omega) \propto \frac{k}{|\omega|^{\gamma}} \qquad [20]$$

where ω corresponds to the spatial frequency, and $\gamma = 2H + 1$. The observations of stationarity and self-similarity of fBm suggest that the use of time-frequency signal decomposition techniques, such as MRA, are well suited for the analysis of fractal signals. Thus, the fBm analysis framework, along with eqn [13], may be successfully exploited to estimate H, and hence, FD as follows:

$$\mathrm{FD} = D_{\mathrm{E}} + 1 - H \qquad [21]$$

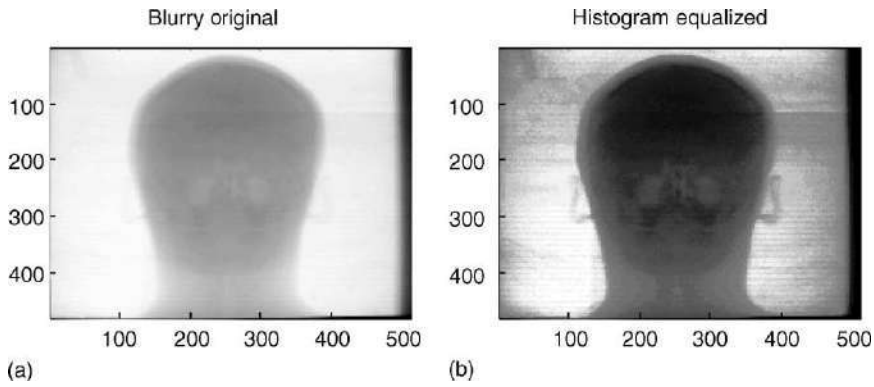

Figure 5 (a) Original EPID image; and (b) Histogram enhanced image.

where D_E is the Euclidean dimension that contains the fBm (i.e., the position of each point of the process is described with the vector $x = (x_1, \ldots, x_E)$).

We exploit the wavelet MRA combined with complementary signal processing techniques such as time-frequency descriptions and parameter estimation of stochastic signals in fBm framework for tumor detection in CT images. We demonstrate an example of image post-processing for tumor segmentation. We test our models to estimate FD in brain MR images as shown in Figures 6a and b, respectively. The images are scanned pixel-by-pixel with a sliding window that defines the subimage for analysis. Pre-processing, such as median filtering, is used to remove noise from the original images.

The results for the spectral analyses of CT image are shown in Figure 7. The results suggest that the FD values do not depend on the mother wavelet used. The window size that offers the best results is the 16×16 pixels, while better discrimination between low and high H values is obtained using biorthogonal wavelet. To improve tumor detection performance of the spectral technique, we apply post-processing filters on the results in Figures 8a and b, respectively. In Figure 8, for an example, we show the filtered fractal analysis results corresponding to a 16×16 pixels window with Daubechies 4. It is possible to obtain an even better description of the location and shape of the tumor after applying a two-dimensional post-processing filter to the matrices derived from the variance estimation algorithms with biorthogonal 1.5 wavelet and an 8×8 pixels window, respectively. The results are shown in Figure 9 and are similar for both averaging and Gaussian filters as well.

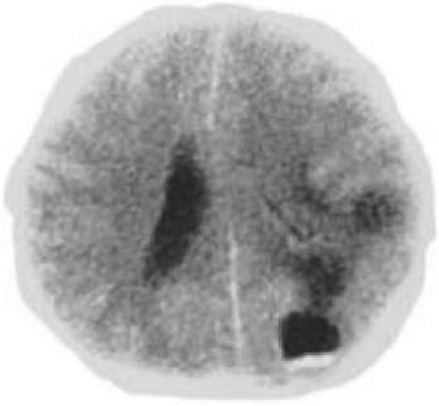

Figure 6 Biomedical images used to test the proposed algorithms.

Example III: Time-Frequency Domain Filtering: Image Post-Processing in Noise Reduction (De-Noising)

This section discusses an example of image post-processing in image noise removal (or de-noising) using a scale-frequency technique such as wavelets. The wavelet transform has been proved to be very successful in reducing noise (de-noising) in an image. We recently propose a new family of wavelets namely Joint Harmonic Wavelet Transform (JHWT) that offers better speckle noise reduction for image post-processing. Unlike wavelets generated by discrete dilation functions, as shown in eqn [18], whose shape cannot be expressed in functional form, JHWT has a simple structure, which can be written, in the

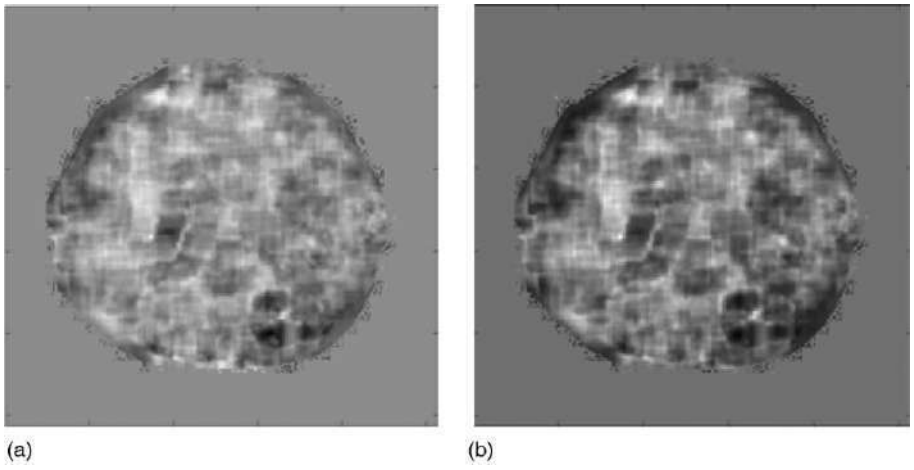

Figure 7 MR-Fractal analysis using spectral method (a) and (b) correspond to a 16×16 analyzing window for Daubechies 4 and biorthogonal 1.5 wavelets, respectively.

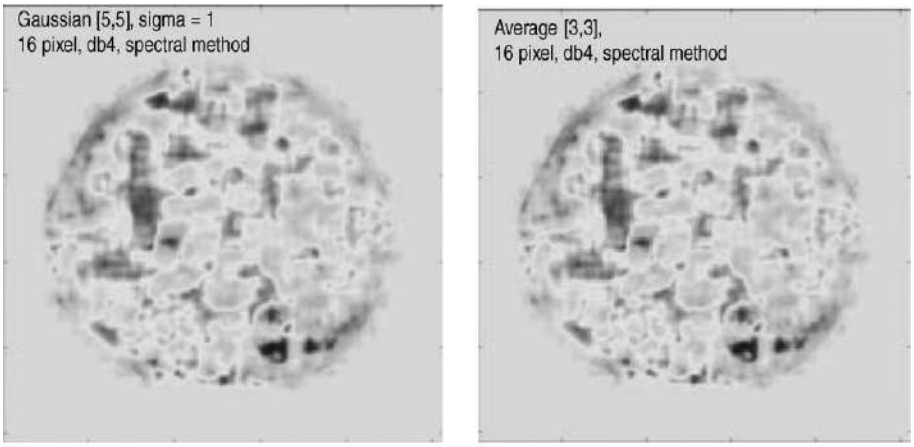

(a) (b)

Figure 8 Post-processing results using: (a) 5×5 Gaussian filtering with $\sigma = 1$ and (b) 3×3 averaging filter.

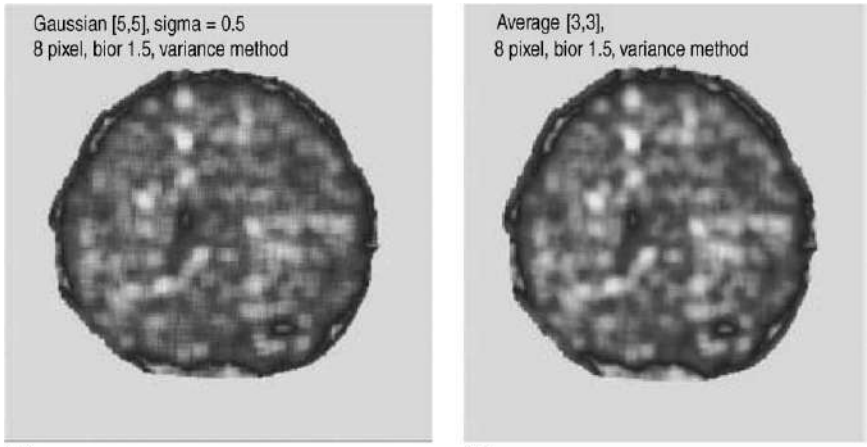

(a) (b)

Figure 9 Post-processing on the fractal analysis results of **Figure 8** for (a) 5×5 Gaussian filtering with $\sigma = 1$, and (b) 3×3 averaging filtering.

frequency domain:

$$W(\omega) = \int_{-\infty}^{\infty} w(x)\exp(-\mathrm{i}\omega x)\mathrm{d}x, \\ = \begin{cases} 1 & \text{for } 2\pi \le \omega < 4\pi \\ 0 & \text{elsewhere} \end{cases} \qquad [22]$$

Now, at level j, we let $g(x) = w(2^j x)$ and shift it by $k/2^j$ to obtain the harmonic wavelet as:

$$g(x - k/2^j) = \{\exp\{\mathrm{i}4\pi(2^j x - k)\} \\ - \exp\{\mathrm{i}2\pi(2^j x - k)\}\}/\mathrm{i}2\pi(2^j x - k) \qquad [23]$$

where integers j and k define the scale and translation of the wavelet $w(x)$, respectively on the x-axis. Now consider a real function $f(t)$ is represented by the sequence f_r of length $(n-1)$, where $r = 0$ to $2^n - 1$. To obtain JHWT coefficients, we take FFT of f_r such that,

$$|F_{2^j+s}|^2 = (1/n^2)\sum_{k=0}^{n-1} |a_{2^j+k}|^2 \exp(-\mathrm{i}4\pi sk/n), \qquad [24]$$

where $r = 0$ to $n-1$ and $s = 0$ to $n-1$

where j denotes the level of the wavelet, k is the translation, m and n are adjacent levels of wavelets with $n = 2^j$, and $s = m - n$. The JHWT in eqn [24] offers all the properties and advantages of standard discrete wavelet transform. Putting $n = 2^j$ and $m = s + n$, eqn [24] yields all the JHWT coefficients from

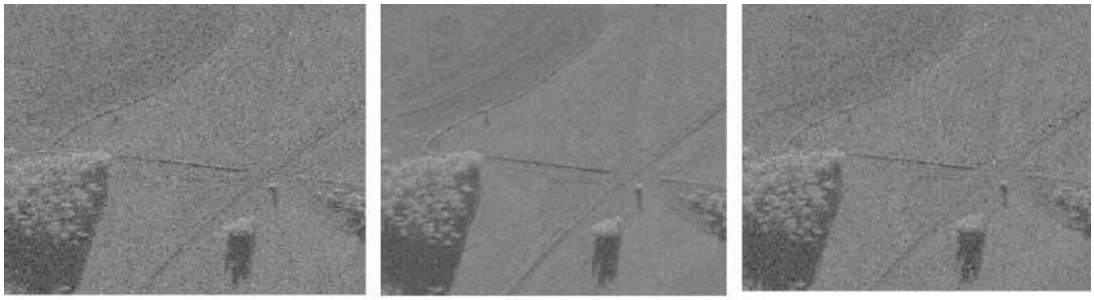

(a) (b) (c)

Figure 10 (a) Noisy image; (b) de-noised image with JHWT-filtering; and (c) de-noised image with WP-filtering.

the Fourier coefficients as

$$\sum_{j=0}^{N-1} |F_{2^j+4}|^2 = |a_0|^2 + \sum_{j=0}^{n-2} (1/2^j) \times \sum_{k=0}^{2^j-1} \left(|a_{2^j+k}|^2 + |a_{N-2^j-k}|^2 \right) + (a_{N/2})^2 = (1/N) \sum_{r=0}^{N-1} f_r^2 \qquad [25]$$

Unlike conventional correlation techniques, the perfect octave band limited characteristics of JHWT eliminate the necessity of any multiplication. We show an example of image de-noising in Figure 10 using JHWT.

Image Post-processing, Transmission, and Distribution

One of the areas wherein image post-processing plays a vital role is in image transmission. A significant problem for any digital library is the network bandwidth required to transmit large digital images. The bandwidth consumption depends on the number of images used by a single user, and with the number of users requesting the services of the library. The popularity of database browsers such as Gopher and Mosaic has serious implications for image transmission and distribution. There has been active research in innovative new compression schemes and transmission protocols that reduce bandwidth requirements for many users along with software servers and visualization. Further, compression of digital image and video data plays an important role in reducing the transmission and distribution cost for visual communication.

One of the requirements for effective image transmission and distribution is reduced data size for faster speed. Image compression schemes are often used for image data reduction. Most of the successful image compression schemes exploit subdividing the image into blocks for increased compression. However, this subdivision of image renders a blocky effect on the compressed image. There has been intense research in removing the blocky effect on the compressed image. To achieve high bit rate reduction while maintaining the best possible perceptual quality, post-processing techniques provide one attractive solution. There has been a significant amount of work done in post-processing of compressed image/video. Some of them are developed from the image enhancement point of view while others are from image restoration. In the case of image restoration, the post-processing is considered as a reconstruction problem. Image enhancement methods are performed based on the local contents of the decoded image as discussed above.

The image post-processing algorithms are, generally, very effective in removing blocking artifacts. A straightforward approach to remove blocking artifacts is to apply lowpass filtering to the region, where artifacts occur. Consequently, the filter must be adapted to the block boundaries, i.e., the image partition. Further, if high frequency details such as edges and texture are to be preserved, then lowpass filter coefficients should be selected appropriately in regions where such details occur. Therefore, adaptive image post-processing such as spatial filtering methods may be applied to deal with this problem. In low-bit-rate image and video compression and distribution systems, high frequency transformed coefficients are often lost due to aggressive quantization or uneven error protection schemes. However, with the block-based encoding and transmission methods, the amount of high frequency texture loss is uneven between adjacent blocks. Thus, it is possible to exploit this texture-level correlation between adjacent image blocks to reconstruct the lost texture.

A relevant problem of robust transmission and distribution of compressed electronic images involves unreliable networks. The pre- and post-processing of the image data may play a vital role in addressing this problem. With the rapid growth of mobile wireless

communication systems, wireless multimedia has recently undergone enormous development. However, to design an efficient multimedia communication system over wireless channels, there still exist many challenges due to severe wireless channel conditions, special characteristics of the compressed multimedia data, and resource limitation, such as power supply. Since multimedia data, such as an image, contain redundancy, to efficiently utilize limited resources, source compression may be necessary. Further, as mentioned above, the ease of image creation, processing, transmission, and distribution has also magnified the associated counterfeiting problems. We discuss the relevant issues of image distribution in the following sections.

Techniques in Image Distribution

The spectacular development of imaging sensors, communication, and internet infrastructure has also made the counterfeiting tools easily accessible and more affordable. In addition, the threats and vulnerabilities of the internet have also increased manyfold. All of these together pose a formidable challenge to the secure distribution and post-processing of images.

Requirements of Secure Digital Image Distribution

In light of the above-mentioned vulnerabilities of image distribution channels and processes, below are a few requirements of secure image distribution:

- *Data Integrity and Authentication*: Digital image is easy to tamper with. Powerful publicly available image processing software packages such as Adobe Photoshop or PaintShop Pro make digital forgeries a reality. Integrity of the image content and authentication of the source and destination of the image distribution are, therefore, of utmost importance.
- *Digital Forensic Analysis*: Consider a litigation involving a medical image that is used for diagnosis and treatment of a disease, as shown in the post-processing example in Figures 7 or 8 above. Integrity of the image again is vital for the proper forensic analysis.
- *Copyright Protection*: Digital copyright protection has become an essential issue to keep the healthy growth of the internet and to protect the rights of on-line content providers. This is something that a textual notice alone cannot do, because it can so easily be forged. An adversary can steal an image, use an image processing program to replace the existing copyright notice with a different notice and then claim to own the copyright himself. This requirement has resulted in a generalized digital rights management initiative that involves copyright protection technologies, policies, and legal issues.
- *Transaction Tracking/Fingerprinting*: In the continuous (re)distribution of digital images through internet and other digital media (CD, camera, etc.), it is important to track the distribution points to track possible illegal distribution of images.

Next, we will discuss how the digital watermarking technology can help enforce these security requirements.

Example of secure image distribution: digital watermarking

Digital watermarking is the process of embedding a digital code (watermark) into a digital content like image, audio, or video. The embedded information, sometimes called a watermark, is dependent on the security requirements mentioned above. For example, if it is copyright application, the embedded information could be the copyright notice. Figure 11 shows a block diagram of the digital watermarking process.

Consider that we want to embed a message M in an image, f_o, also known as the original image. For the sake of brevity, we drop the indices representing the pixel coordinates. The message is first encoded using source coding and optionally with the help of error correction and detection coding, represented in Figure 11 as $e(M)$ such that

$$W_m = e(M) \qquad [26]$$

The encoded message W_m is then combined with the key-based reference pattern W_k, in addition to the scaling factor α, to result in W_a which is the signal that is actually added to the original image, given as

$$W_a = \alpha(f_o)[W_M \otimes W_K] \qquad [27]$$

Note that the scaling coefficient α may be a constant or image dependent or derived from a model of image perceptual quality. When the scaling factor α is dependent on the original image, we can have informed embedding which may result in more perceptually adaptive watermarking. The operation involving W_m and W_k is the modulation part, which is dependent on the actual watermarking algorithm used. As an example, in a typical spread-spectrum based watermarking, this operation is usually an exclusive NOR (XNOR) operation. The final operation is an additive or multiplicative embedding of W_a with the original image. We show a simple additive watermarking process as

$$f_w = f_o + W_a \qquad [28]$$

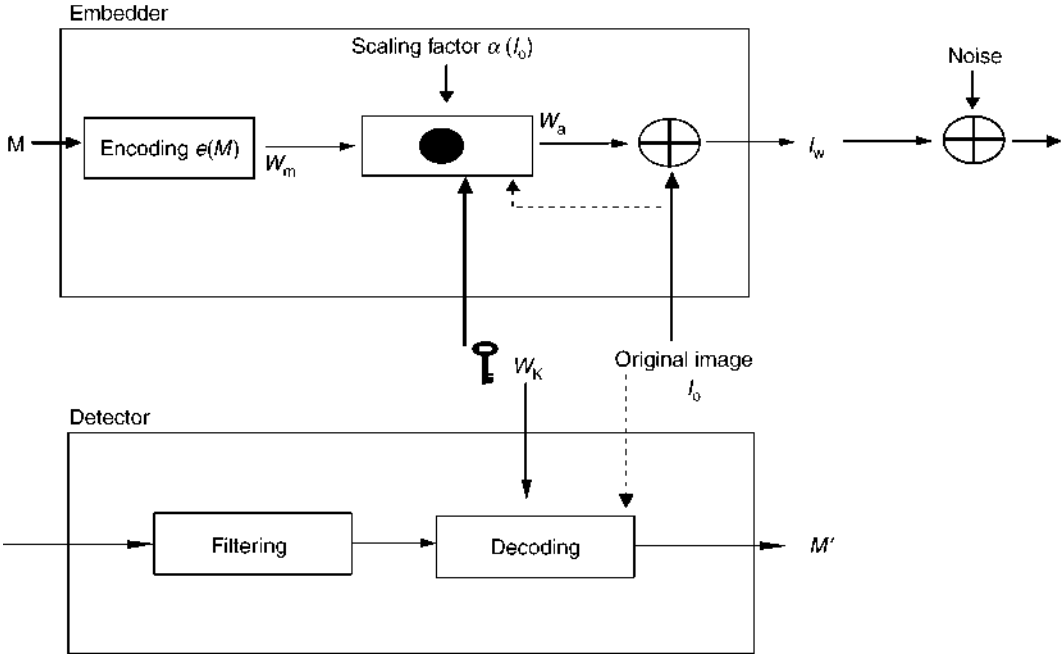

Figure 11 A simple watermarking process: embedder and detector.

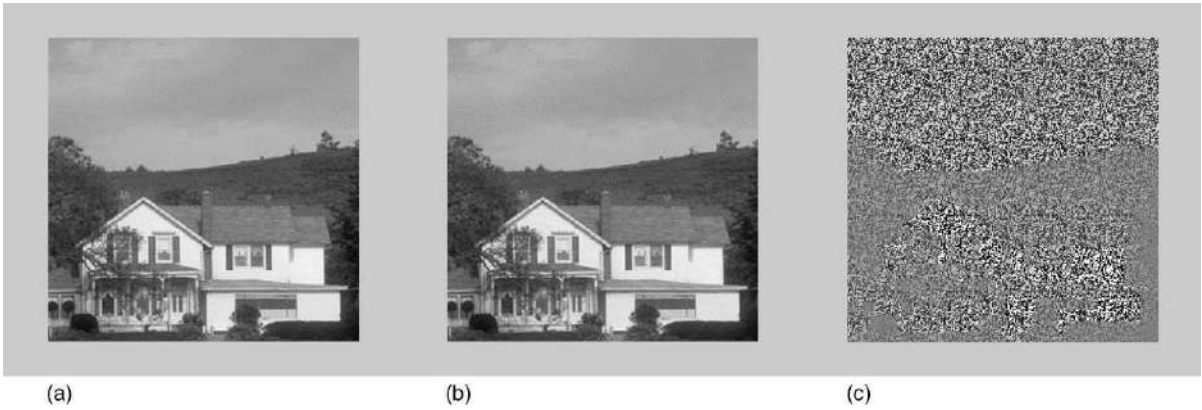

Figure 12 (a) Original image, (b) watermarked image, and (c) the difference image.

The result of this process is the watermarked image f_w which may actually need some clipping and clamping operations to have the image pixel values in the desired range.

The detector, as shown in Figure 11, does the inverse operations of the embedder. If the original image is known, as in the case of nonblind watermarking, it is first subtracted from the watermarked image. Blind watermarking detector, on the other hand, uses some filtering with a goal to estimate and subtract the original image component. The result of filtering is then passed to the decoder, which collects the message bits with the help of the shared key, and finally gets back the decoded message M'.

Figure 12 shows the original image, watermarked image, and the difference image using a spread-spectrum based watermarking algorithm, respectively. The enhanced difference image represents a noise-like image, wherein the scaling coefficient α is clearly dependent on the original image. As evident from Figure 12b, the hidden message is imperceptible as well as the change in the original image. The imperceptibility is one of the primary criteria of digital watermarking.

There are other criteria such as unobtrusiveness, capacity, robustness, security, and detectability. Thus, digital watermarking is a multidimensional problem, which attempts to make a trade-off among a number of different criteria. The inserted

watermark should not be obtrusive to the intended use of the original image. Because the watermarking process does not increase the size of the original image, it may be desirable to add as much information as possible. However, generally speaking, the more information one adds, the more severe will be the impact on the perceptual quality of the image. From the detection point of view, watermark should be robust enough to tolerate a range of image post-processing and degradation. In addition, the embedded watermark should also be secure enough so that it may not be removed from the watermarked image. Finally, detectability of the embedded watermark is an important criterion that places some constraints on the embedding algorithm.

In order to fully appreciate the use of watermarking for image distribution, it is instructive to look at different classes of watermark, which is delineated in Figure 13.

For an example, depending on the robustness criterion of watermarking, they can be classified into three categories such as robust, fragile, and semi-fragile. Note that while the robustness of watermark is important, it is not always equally desirable in all applications. Let us take the authentication as an example. If you, as the owner of a digital image, embed some authentication information into the image using watermarking, you might want to see the authentication fail, when some one steals your image. This is an example of nonrobust watermark, where the fragility of the embedded watermark is required to detect any forging done on the digital image.

Now, consider how watermark provides the security requirements for secure image distribution. In particular, we look into the authentication of watermark. Figure 14 shows the use of watermark for image authentication, which has some similarities with the cryptographic authentication using digital signature. However, there are potential benefits to using watermarks in content authentication and verification. Unlike cryptography, watermarking offers both in-storage and in-transit authentication. Watermarking is also faster compared to encryption, which is very important in internet-based image distribution. By comparing a watermark against a known reference, it may be possible to infer not just that an alteration occurred but what, when and where changes have occurred. As shown in Figure 14, first we identify some features from the image and then we compute the digital signature, which is then embedded into the image. The detector takes a test image, computes the signature and extracts the embedded

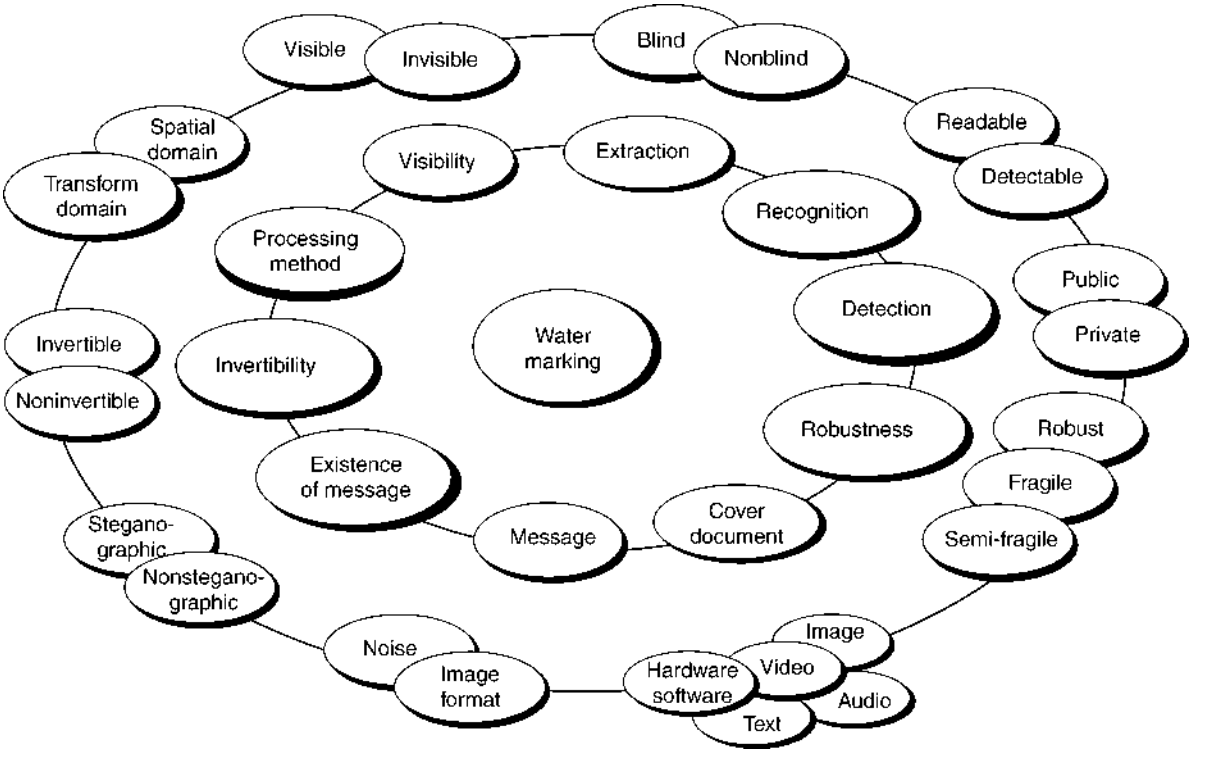

Figure 13 Watermark classifications.

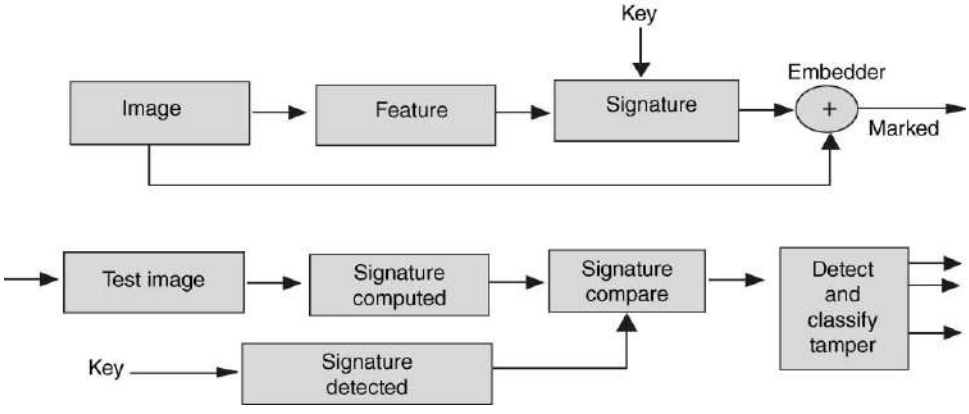

Figure 14 Image authentication using watermark.

signature, and then compares the two. If they match, the image is authenticated. If they do not match, then the image may have gone through some forging/tampering effects and the algorithm may optionally detect the tamper area as well.

One of the limitations of the watermarking approach in secure image distribution is that the technology is not yet widely deployed, and nor is the protocol satisfactorily standardized. Other major limitation is that the watermarking process cannot be made sufficiently robust to arbitrary types of different attacks. In reality, therefore, the hybrid combination of cryptography and watermarking is expected to improve the secure distribution of images across the unreliable internet.

Concluding Remarks

In this chapter, we discussed the principles of image post-processing and distribution. We showed a few representative examples for post-processing and distribution. We also explained the relevant issues in image distribution and present digital watermarking as one of the solutions for secure image distribution.

Acknowledgment

One of the authors (KMI) wishes to acknowledge the partial support of this work through a grant (RG-01-0125) from the Whitaker Foundation.

See also

Imaging: Information Theory in Imaging.

Further Reading

Ahmed F and Moskowitz IS (2004) A correlation-based watermarking method for image authentication applications, accepted for publication in *Optical Engineering*.

Barlaud M (1994) *Wavelets in Image Communication, Advances in Image Communication 5*. Amsterdam: Elsevier.

Chui A (1992) *An Introduction to Wavelets*. San Diego: Academic Press.

Cox NI, Bloom J and Miller M (2001) *Digital Watermarking: Principles & Practice*. Francis, CO: Morgan Kauffman Publishers.

Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory* 36(5): 961–1005.

Gonzalez RC and Woods RE (1993) *Digital Image Processing*. New Jersey: Addison-Wesley Publishing Company.

Iftekharuddin KM and Parra C (2004) Multiresolution-fractal analysis of medical images. Under Review. *Neuroimaging*, vol. 1. New York: Wiley/IEEE Press, Chapter 4.

Jain AK (1989) *Fundamental of Digital Image Processing*. New Jersey: Prentice-Hall.

Lu N (1997) *Fractal Imaging*. San Diego: Academic Press.

Mallat S (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7): 674–693.

Mandelbrot BB (1983) *The Fractal Geometry of Nature*. San Francisco, CA: Freeman.

Schmucker AM and Wolthusen SD (2003) *Techniques and Applications of Digital Watermarking and Content Protection*. Norwood, MA: Artech House.

Serra J (1982) *Image Analysis and Mathematical Morphology*. San Diego: Academic Press.

Stallings W (2003) *Cryptography and Network Security – Principles and Practices*, 3rd edn. New Jersey: Prentice-Hall.

Wu M and Liu B (2002) *Multimedia Data Hiding*. New York: Springer.

Smart Pixel Arrays

P Seitz, Centre Suisse d'Electronique et de Microtechnique SA, Zurich, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

For more than 50 years silicon has been the semiconductor on which the enormous progress of microelectronics is based. Since silicon is also an excellent detector for electromagnetic radiation, in particular in the visible spectrum between 400 and 700 nm, the advances in semiconductor technology have directly influenced the progress of solid-state image sensing. Once the relentlessly shrinking dimensions of transistors have reached a critical minimum limit, it has become economically feasible to co-integrate analog and digital circuitry with each light-sensitive pixel. This capability is at the heart of the technological revolution called smart pixel arrays.

The first section relates how the shrinkage of minimum feature size on silicon chips paralleled the reduction of solid-state pixel geometry, as a function of time. This development ceased once the pixel size reached the optical diffraction limit, which marked the birth of the field of smart pixel arrays. In the second section, the basic functionality and the most important physical limitations of conventional pixels and photosensors are given. The third section deals with opto-electronic functionality that is preferred today, of smart pixel arrays; in particular, the capability of spatio-temporal demodulation is very much in demand. In the fourth section, some of the most useful modular building blocks for smart pixel arrays are presented, together with a few key applications. An outlook on photonic systems-on-chip (SoC) is given in the fifth section, with a discussion of single-chip machine vision systems (so-called seeing chips), organic semiconductors for even higher levels of integration of photonic microsystems, and smart pixel arrays that are sensitive to physical parameters other than electromagnetic radiation.

Silicon Technology for Image Sensing

Long before it was realized that silicon would be the material on which modern semiconductor industry and the information technology revolution is based, in 1940 Russell Ohl discovered that silicon is an excellent detector of electromagnetic radiation. In particular, silicon is an excellent detector for light in the visible and near infrared part of the spectrum, for wavelengths between 400 and 1,000 nm. This is very fortunate because it allows designers of image sensors to profit directly from the amazing progress of silicon technology. Since the 1970s, the minimum feature size on silicon chips showed a reliable exponential reduction of about 10% per year, as illustrated in Figure 1. This is a direct consequence of the observation by Intel's Gordon Moore, sometimes called 'Moore's law', that the number of transistors per integrated circuit doubles about every two years. The evolution of minimum feature size is paralleled by the evolution of minimum pixel size of solid-state image sensor, as illustrated in Figure 2. Once the pixel size reached a record low value of 2.4 μm in the year 1996, no further reduction in pixel size seemed desirable, because the optical diffraction limit was essentially reached.

The combination of technological availability of smaller and smaller electronic components and pixel size that demanded no further reduction, inspired researchers to provide each pixel with an increasing amount of electronic functionality, all of which is available simultaneously for massively parallel information processing. This represents the heart of the smart pixel array revolution; pixels that not only sense the incident electromagnetic radiation but that are capable of pre-processing the acquired information in a relevant fashion.

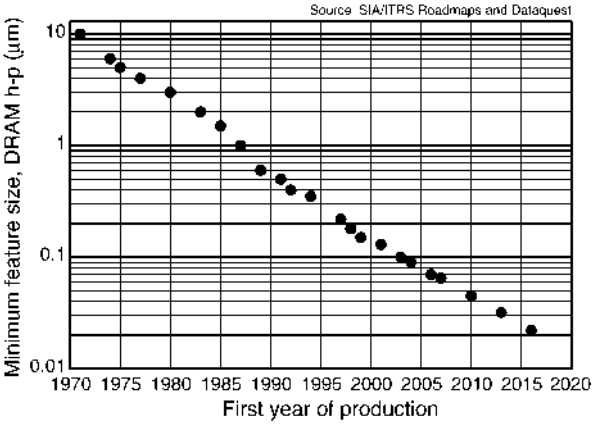

Figure 1 Evolution of minimum feature size (DRAM half-pitch) in state-of-the-art silicon technology, showing an exponential reduction for the past 30 years, which is expected to continue for at least another decade.

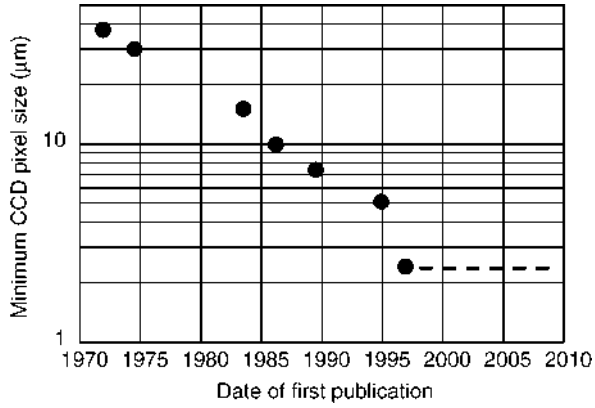

Figure 2 Evolution of minimum pixel pitch in CCD image sensors, tracking the exponential reduction of minimum feature size shown in Figure 1, until a lower limit was reached in 1997, imposed by optical diffraction. Since then, no smaller pixel pitches have been described in the literature; this is indicated graphically with the hatched 'diffraction limit' line.

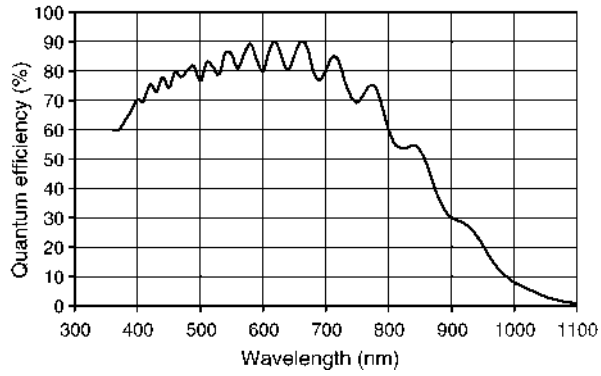

Figure 3 External quantum efficiency of a photosensor realized with an industry standard CMOS process, showing quantum efficiencies above 50% in the whole visible spectral region.

Basic Functionality and Physical Limitations of Conventional Solid-State Photosensors

The prime task of a pixel is the detection of incident electromagnetic radiation. The detection process consists of six basic steps:

1. The incident photon impinges on the surface of the image sensor, and it has to reach the bulk of the semiconductor.
2. In the bulk of the semiconductor the photon has to interact with the semiconductor, converting its energy into a number of electron–hole pairs.
3. The photogenerated charge carriers must be transported to the sensor surface.
4. At the surface, the photogenerated charge needs to be collected and accumulated in the storage part of the pixel.
5. The accumulated photocharge must be transferred to the input of an electronic circuit capable of detecting this charge.
6. The charge detection circuit converts photocharge into a corresponding voltage, exhibiting a monotonous (but not necessarily linear) transfer function.

Since the interaction length of electromagnetic radiation in silicon is in the tens of micrometers for incident photon energies between 1 eV and 10 keV (corresponding to a wavelength range of about 0.1 to 1,000 nm), silicon is an excellent detector for this large spectral range. This is illustrated for the visible and near infrared spectral range in Figure 3, showing the absolute external quantum efficiency of a pixel realized with an industry standard silicon process. In the visible spectral range, absolute external quantum efficiencies close to 100% can be reached, even with standard silicon processes for the microelectronics industry that were not optimized for photosensing applications.

Photocharge transport to the surface of the semiconductor is mainly provided by diffusion, i.e. random thermal motion of the charge carriers that exhibit a velocity of about 10^5 m/s in silicon at room temperature. The mean distance L traveled by a diffusing charge carrier during time t is given by

$$L = \sqrt{Dt} \qquad [1]$$

with the diffusivity D, which has a value of around 40 cm^2/s for electrons in low-doped silicon at room temperature.

Once the photocharge carriers are close to the surface of the semiconductor, they are transported by drift in an electric field E, exhibiting a linear dependence of the transport distance L with time t:

$$L = \mu E t \qquad [2]$$

with the mobility μ which has a value of around 1300 cm^2/Vs for electrons in low-doped silicon at room temperature.

There are two basic principles for creating the electric field with which the photocharge carriers are transported to their storage site within the pixel. A photodiode consists of a semiconductor junction between a p-doped and an n-doped region, as illustrated in Figure 4. Such a photodiode also represents a capacitance on which photocharge can be accumulated and stored. Photodiodes are used as photosensors in one

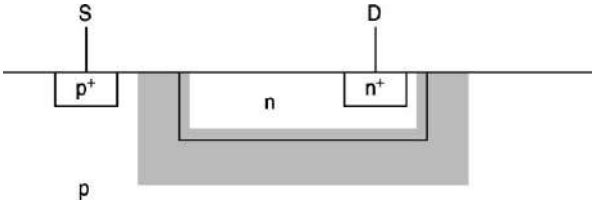

Figure 4 Cross-section of a photodiode, consisting of an n-doped region in a p-doped substrate. The substrate is contacted using a heavily p-doped diffusion S, and the n-region is contacted using a heavily n-doped diffusion D. The space–charge region is illustrated as hatched area at the pn-junction.

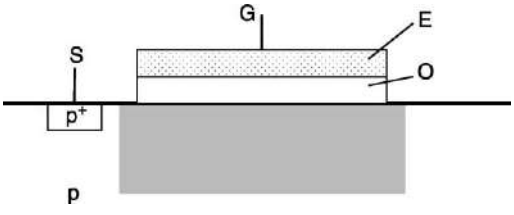

Figure 5 Cross-section of an MOS capacitance, the basic element of a CCD image sensor. On top of a p-substrate a thin transparent insulator O and a transparent electrode E are fabricated. The substrate is contacted with a heavily p-doped diffusion S, and the electrode E is contacted with the gate contact G. The hatched area illustrates the space–charge region under the electrode.

popular type of image sensor, so-called APS (active pixel sensor) or CMOS (complementary metal oxide semiconductor) imagers.

An alternative to a photodiode is an MOS (metal oxide semiconductor) capacitor, consisting of a transparent electrode that covers an insulating layer on top of semiconducting material, as illustrated in Figure 5. An MOS capacitor is the basic element of a charge coupled device (CCD), the other widespread type of image sensor.

The charge transfer between photodiodes and photodetector circuits is based on electrical conduction, occurring essentially with the speed of light. In CCDs, the charge transfer is effectuated with clocks, and excellent charge transfer efficiencies can be observed at clock frequencies of several tens of MHz.

The major limitation in semiconductor photosensors is the last element in the line, i.e. the electronic charge detection circuit. In Figure 6, a so-called source follower is illustrated, which is by far the most frequently used photocharge detection circuit. Its performance is limited by the thermal (Johnson) noise in the channel of the first field effect transistor (FET). This corresponds to a statistical uncertainty with which an amount of photocharge can be measured at the gate of the FET. The input referred charge noise

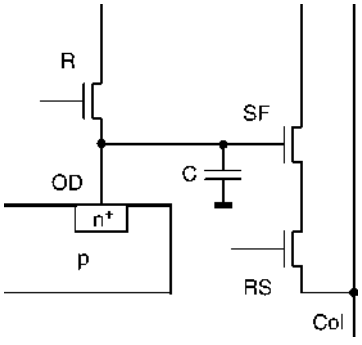

Figure 6 Schematic diagram of a photosensor's output diffusion OD connected to a reset transistor R and a source follower transistor SF, the most widely used charge detection circuit for image sensors and smart pixel arrays. The photocharge is stored on the effective capacitance C at the gate of transistor SF, which is read out using a row select transistor RS, connecting the source follower to the readout column Col.

variance s_Q is given approximately by

$$s_Q \cong C^2 \frac{4kTB}{g} \qquad [3]$$

where C is the effective capacitance seen by the gate of the FET, k is Boltzmann's constant, T is the absolute temperature, B is the effective bandwidth at the FET's channel, and g is the transconductance of the FET.

Desired Optoelectronic Functionality in Smart Pixel Arrays

From the previous section it becomes clear that even conventional pixel and photosensors can be improved in several desirable respects:

- The spectral response can be improved in the ultraviolet (50–400 nm), or it can be expanded significantly into the X-ray region (below 0.1 nm) and into the infrared region (above 1,000 nm).
- The speed of the photosensors can be increased by replacing diffusion based photocharge transfer mechanisms with the directed, more efficient drift in electric fields. This is true for the vertical as well as for the lateral charge transport direction.
- The sensitivity of the electronic charge detection process can be improved in two ways. It can be achieved by decreasing the effective detection capacitance C at the input of the charge detection circuit, as can easily be seen from eqn [3]. An alternative is to make use of a physical charge amplification process (i.e. avalanche multiplication in a high electric field), that has low intrinsic noise.

- A derivative improvement that is desired is the increase of the dynamic range D/R of the complete detection process, as defined, for example, by

$$D/R = 10^{10} \log \frac{Q_{\max}^2}{s_Q} \qquad [4]$$

where $Q_{\max}$ denotes the maximum (or saturation) charge and s_Q is the total charge noise variance of the electronic detection circuit.

In addition to the above described improvements of conventional pixel functionality, a huge demand exists for the capability to measure the parameters of modulated electromagnetic signals, i.e. to demodulate such signals. This demand is related to the large number of very powerful electronic and opto-electronic measurement principles, capable of improving the sensitivity of the measurement process significantly by employing modulated signals, either in the temporal or in the spatial domain. For this reason, smart pixel arrays are sought that offer this combination of photodetection and demodulation.

Modular Building Blocks for Smart Pixel Arrays

Towards the end of the last century, a large body of literature has been created, addressing various approaches for the realization of one of the above described desired opto-electronic functionalities. The result is a powerful 'toolbox' of modular building blocks for custom smart pixel arrays, with which optimized solutions for optical measurement problems and practical applications can be constructed.

High-Sensitivity Charge Detection

From the noise formula (eqn [3]) two possibilities for the improvement of the sensitivity of image sensors can be deduced: A first approach is the reduction of the capacitance value. An excellent example for this is the invention of the 'double gate' FET, in which the charge to be measured is not placed on a gate electrode at the surface but resides in the semiconductor bulk. In this way the effective capacitance of the pixel can be reduced by at least one order of magnitude compared to conventional pixels. The resulting capacitance of less than 1 fF leads to a readout noise of about one electron at room temperature and for a readout bandwidth of several MHz. Unfortunately the small capacitance allows the storage of only around ten thousand electrons. For such small charge packets, quantum noise is still relevant, leading to significantly degraded image quality, even when the images are taken under favorable lighting conditions.

A second approach to high-sensitivity photosensing exploits the dependence of the charge detection variance on the measurement bandwidth, as indicated in eqn [3]. One method for exploiting this consists of reading out the image sensor at a significantly reduced bandwidth of a few 10 kHz. At the same time, the image sensor is cooled to temperatures of around −100 °C. This becomes necessary because readout at such slow speeds increases the time, to read one complete image, to many minutes, requiring a substantial reduction of the dark current. Instead of working at low readout bandwidths, it is also possible to read out the same information repeatedly and non-destructively at high speed, and to average many measurements of the same amount of photocharge. Both methods have been shown to lead to readout noise levels of about one electron; on the other hand, both approaches require very long times for the acquisition of complete images.

This problem can be resolved by exploiting the massive parallelism that smart pixel arrays offer: low-bandwidth techniques can be employed without losing effective readout speed. Pixels and columns are read out with low-bandwidth amplifiers that are all active simultaneously, and only the final off-chip readout amplifier needs to operate at a large bandwidth. In practice, this approach has been used successfully in single-chip digital cameras with Megapixel resolution, exhibiting a readout noise of only a very few electrons.

A completely different approach is based on the use of a physical amplification mechanism with low intrinsic noise, before detecting the amplified charge packets with a conventional electronic circuit; the avalanche multiplication effect in high electric fields is very well suited to this approach. Each pixel is supplied with its own combination of stabilized avalanche multiplier and conventional charge detection circuit. This approach has been used for the realization of a fully integrated 2D array of single photon detectors that can be fabricated with industry standard CMOS processes. Similar performance is obtained with the Impactron™ series of high-sensitivity CCD image sensors offered by Texas Instruments Inc., making use of a single multistage avalanche multiplier at the output of the image sensor. For both high-sensitivity smart pixel arrays, single-photoelectron performance is reported at room temperature and at readout frequencies of several MHz, as required for video applications.

Photosensing with Improved Dynamic Range

Conventional solid-state image sensors and cameras usually have a very limited dynamic range, as defined in eqn [4]. Typical values are between 40 and 80 dB, rarely exceeding four orders of magnitude. In natural scenes, however, much higher *D*/*R* values can be encountered. To cope with this, the human eye is capable of spanning a dynamic range of more than 150 dB. This realization has spurred the invention of several methods with which the *D*/*R* of smart pixel arrays can be increased.

One approach consists of acquiring several images of the same scene with different exposure times, with the subsequent combination into one high-dynamic-range composite picture. *D*/*R* values exceeding 120 dB have been obtained with such multi-exposure methods. Fast-moving objects, however, will have relocated from one exposure to the next, resulting in false, blurred representations of these objects in the composite picture.

This can be overcome using nonlinear response characteristics similar to the ones found in natural vision. Impinging light produces a photocurrent that is highly linear with the incident light intensity. Because the voltage across the photodiode is logarithmically dependent on the photocurrent, a logarithmic voltage-illumination response is obtained, with a *D*/*R* exceeding 120 dB. Such logarithmic pixels are well-suited for applications providing high illumination levels but they suffer from excessive noise and slow response speed at low light levels.

This disadvantage led to the development of linear-logarithmic pixels, exhibiting a high-sensitivity linear response at low light levels, and a gradual changeover to logarithmic response once a certain, programmable light level is reached. The design principle is illustrated in Figure 7. An otherwise conventional CMOS/APS pixel is supplied with a device that exhibits logarithmic voltage–current response, such as a diode. As long as the cathode is at a higher voltage than the anode, no current flows across the diode, and the pixel response is linear. Once the photocurrent has lowered the cathode voltage below the anode voltage, the pixel response becomes logarithmic, as described above. The *D*/*R* of such a linear-logarithmic pixel reaches unprecedented high values; almost 200 dB have already been obtained in practice. Because of the simplicity and robustness of this principle, such linear-logarithmic smart pixel arrays have found their way into commercial sensor and camera products that are available commercially, as illustrated in Figure 8. A circuit board and a light bulb are imaged in linear mode (Figure 8a), showing highly saturated areas, and in linear-logarithmic mode (Figure 8b), in which even the light bulb's filament can clearly be seen.

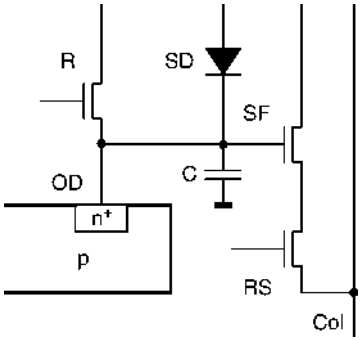

Figure 7 Schematic diagram of a high-dynamic-range pixel, consisting of a conventional APS/CMOS pixel, as illustrated in Figure 6, in combination with a shunt diode SD. The photocurrent lowers the voltage at the gate of transistor SF continuously. Below a certain value of this voltage, the diode starts to become forward biased, leading to a linear-logarithmic illumination response, with a dynamic range that can be as high as 200 dB.

Extension of the Spectral Sensitivity Range

The spectral sensitivity of silicon is essentially restricted to the spectral wavelength range between 0.1 and 1,100 nm. With a suitable choice of materials that cover the silicon sensor's surface, this sensitive wavelength range can be extended. For high-energy radiation, scintillator materials are employed, capable of converting incident high-energy photons into a proportional number of visible photons. Scintillators render smart image sensors particularly sensitive in the UV (300–400 nm), deep UV (below 300 nm), X-ray (0.002–1 nm), and gamma ray (below 0.002 nm) spectral region. Typically, the scintillator materials are fabricated as platelets that are glued on the surface of the produced CMOS/CCD image sensor.

The extension of the sensitive spectral range to the infrared can be achieved with semiconductor materials with smaller energy bandgaps than silicon. Examples include Ge, SiGe, SiC, PtSi, IrSi, InAs, InGaAs, or InGaAsP. Typically, these materials are deposited on the produced CMOS/CCD image sensor in an ultra-high vacuum environment. An alternative is the use of bump-bonding, with which the CMOS/CCD image sensor can be connected to a small-bandgap material chip at each pixel site. With these technologies, silicon-based smart pixel arrays become sensitive to wavelengths of several microns.

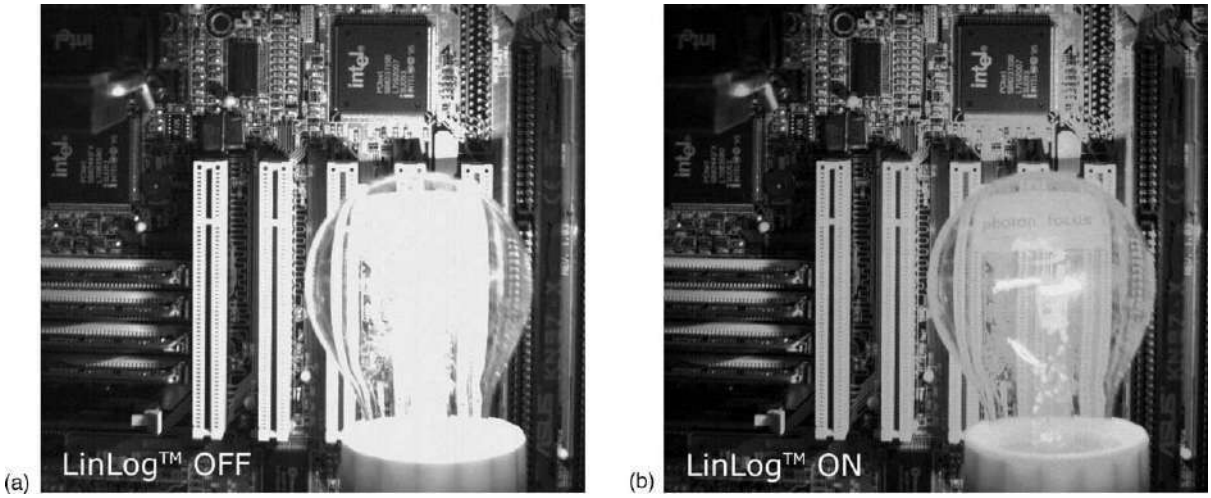

Figure 8 Images of a scene with high dynamic range, taken with an image sensor exhibiting linear illumination response (a) and linear-logarithmic illumination response (b). A similar pixel architecture as illustrated in Figure 7 is employed. (Courtesy of Photonfocus AG, Lachen, Switzerland.)

High-Speed Image Sensing

Increasing the speed of the photosensing process necessitates improvements in two key areas. Photocharges need to be collected and transported faster, and the acquired information must be read out faster.

A comparison of eqns [1] and [2] immediately reveals that the replacement of diffusion by drift transport provides the desired speed improvement. In the vertical direction, the electric field of photodiodes or MOS capacitances reaches more deeply into the semiconductor bulk if lower doping levels of the substrate are used or the bias voltage is increased. Improving the lateral transport speed is more difficult. A highly successful approach in so-called drift-field pixels is illustrated in Figure 9. On top of a thin (a few tens of nanometers) transparent insulator, a highly resistive layer – often of polysilicon – is placed. When a voltage difference is applied at the two ends, a small current flows across the resistive layer, inducing a linear potential variation at the semiconductor surface. A simple calculation shows that very small transit times for photocharges can be obtained in this way. Assuming a (quite large) pixel size of 25 μm and a voltage difference of 5 V, the maximum transit time of photoelectrons that have to move from one end of the pixel to the other, is less than 1 ns. This is more than two orders of magnitude faster than can be achieved with diffusion transport.

Smart pixel arrays provide a simple solution to the problem of speeding up the readout of the acquired picture information; a large degree of parallelism. This is illustrated with the architecture of a split-field high-speed image sensor, shown schematically in Figure 10. The photosensitive area is split into two parts, so that the column amplifiers of one half have time to settle to their proper voltages, while the columns of the other half are being read out. No time is lost, therefore, when another row of the image is selected for readout. A large number of output amplifiers are provided for highly parallel readout of the image data, each with a typical readout rate of 40–100 MHz. In this way, sustainable pixel rates of several GHz can be achieved in practice.

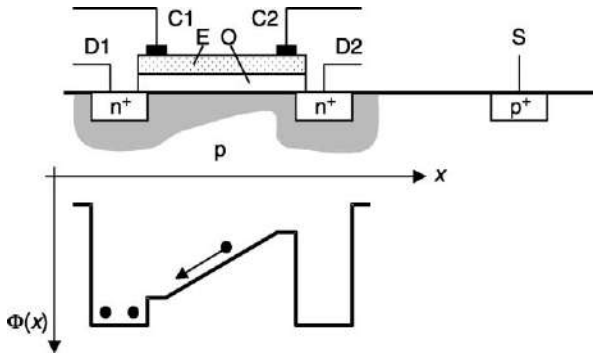

Figure 9 Cross-section of a drift field pixel, exploiting the lateral drift field at the semiconductor–oxide interface. The corresponding sloped potential distribution $\Phi(x)$ is created with a highly resistive transparent electrode E on top of a thin transparent insulator O; this conductor has two contacts C1 and C2 at its ends, to which a voltage difference is applied. Depending on the sign of the slope, the photoelectrons are collected either in diffusion D1 or D2. The hatched area indicates the space-charge region.

Static Spatial Demodulation

The spatially varying geometry of the photosensitive part of a pixel can be used for the static spatial

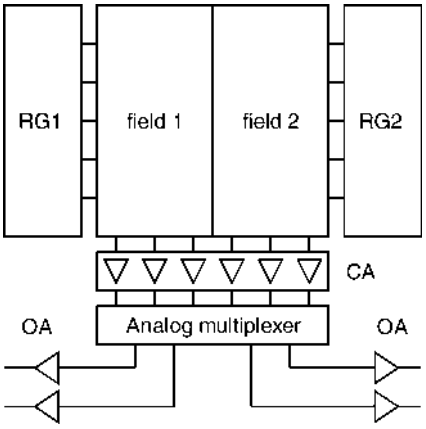

Figure 10 Block diagram of a high-speed image sensor, making use of the large amount of parallelism possible in smart pixel arrays. The light-sensitive area is split into two fields whose rows can be addressed independently with two separate row address generators RG. In this way, one row of one field can be selected and the corresponding column amplifiers CA can reach equilibrium, while the columns of the other field are being read out using several parallel output amplifiers OA.

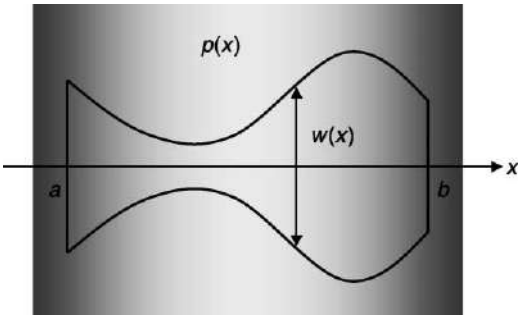

Figure 11 Illustration of a one-dimensional intensity distribution $p(x)$ of light which is incident on a shaped photodiode with position-dependent width $w(x)$, between the two boundaries a and b.

demodulation of an impinging light field. When a one-dimensional intensity distribution $p(x)$ is incident on a photodiode with spatially varying width $w(x)$ between the limits a and b, as illustrated in Figure 11, an output signal r is produced:

$$r = \int_a^b p(x)w(x)\mathrm{d}x \qquad [5]$$

By shaping the photodiode appropriately, any type of parametrized linear transform can be realized, as illustrated by the photomicrograph in Figure 12. In this practical example, four photodiodes are shaped sinusoidally and cosinusoidally, weighted with a slowly varying windowing function. Their signals are used for determining the Fourier coefficients

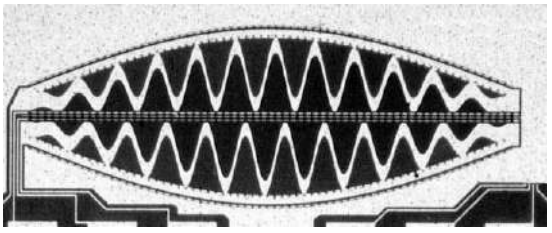

Figure 12 Photomicrograph of two photodiode pairs that are shaped as spatial sine and cosine functions, weighted with a slowly varying windowing function. The four photosignals can be employed to determine the phase and amplitude of a one-dimensional intensity signal that has the same spatial frequency as the sine and cosine. In practice, the accuracy of the phase determination is as high as 0.1% of the spatial wavelength. The width of the photomicrograph corresponds to about 1 mm in reality.

(the sine and cosine transform) of an incident light distribution for a fixed spatial frequency. This is employed, for example, in cost-effective optical encoders, capable of measuring linear displacement with an accuracy below 50 nm.

Dynamic Spatial Convolution

Many image-processing tasks require the convolution of the pixel data $P_{i,j}$ at the position (i, j) in an image with a two-dimensional convolution kernel $w_{k,l}$:

$$R_{k,l} = \sum_{i=-N}^{N} \sum_{j=-N}^{N} P_{i,j} w_{k-i,l-j} \qquad [6]$$

Instead of employing a purely geometrical solution, as described in the previous section, an alternate approach is found when interpreting the convolution eqn [6] in the following way. The resulting image R is obtained by displacing the original image P laterally by an amount (k,l), by multiplying the shifted pixel values with the corresponding weight $w_{k,l}$, and by adding up all these products.

The lateral displacement is realized with a CCD structure, implementing the 2D lateral movement of the acquired image charge packets, the multiplications with the convolution kernel values are realized by using various exposure times for each lateral shift position, proportional to the corresponding weight value, and the summation is just the natural charge accumulation of CCD structures. The problem of negative weights is solved by separately determining the positive and negative parts of the weight function, each stored separately in the pixel, and by calculating the difference of the two values during readout, either on-chip or off-chip. In this way, such convolution image sensors

can be used for the determination of complete, programmable, real-time image convolutions, with arbitrary kernels that are fully under the dynamic control of digital logic or microprocessors. Practical experience with such convolution imagers shows that the convolution results exhibit inaccuracies of less than 1%.

Temporal Demodulation

In many relevant optical measurement techniques, temporally modulated light fields are employed. For this reason, smart pixel arrays are desired, providing synchronous demodulation of these signals. An electromagnetic signal $S(t)$ that is harmonic as a function of time with given frequency ω, having unknown modulation amplitude A, offset B and phase delay φ, is described by

$$S(t) = A\sin(\omega t + \varphi) + B \qquad [7]$$

The three unknowns A, B and φ can be measured if three or more sampling points are sampled synchronously per period in so-called 'lock-in' pixels. An example of a lock-in pixel employing the drift-field transport principle with four accumulation taps, is illustrated in Figure 13. Demodulation frequencies of several tens of MHz have already been demonstrated in practice.

The most important application of such synchronous demodulation pixels is in optical time-of-flight range cameras. Such a camera incorporates a light source, usually an array of infrared or visible LEDs, that is modulated at a few tens of MHz. The modulated light is reflected diffusely by the various objects, and the camera's lens images it onto a lock-in image sensor. Each of its pixels carries out synchronous sampling/accumulation of photocharges. At the end of the exposure time, the accumulated photocharge samples are read out, and the three modulation parameters of the reflected light are determined for each pixel. The phase shift is a direct measure of the local distance. It has been shown that for a wide range of operating conditions, the distance resolution of such optical range cameras is already limited just by the photo shot noise. A typical result of such a depth image is shown in Figure 14.

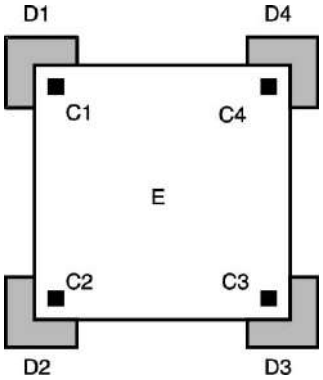

Figure 13 Schematic top view of a lock-in pixel based on the lateral drift-field approach illustrated in Figure 9. A square electrode is fabricated with a highly resistive transparent material on top of a thin insulator. The electrode contacts C1 to C4 are employed to create a 'reeling' potential distribution that rotates with the modulation frequency of the incident light that requires synchronous demodulation, so that the photocharges are collected at diffusions D1 to D4, respectively, at the appropriate time during each period.

Photonic Systems-on-chip

The microelectronics revolution is based on the key notion of integration. More functionality integrated monolithically on the same chip leads to faster, smaller, better performing, and cheaper products. The same is true for smart pixel arrays, for which progress through integration is attained in several respects.

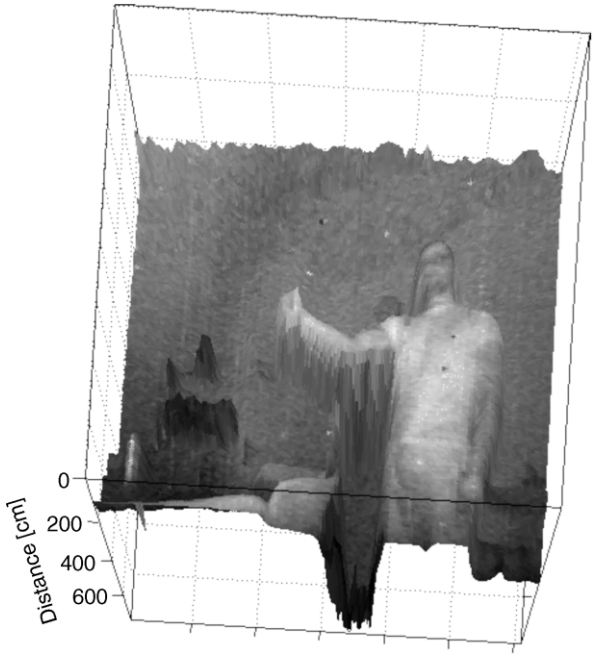

Figure 14 Example of a distance picture (a person) out of a 3D range image sequence, taken with an optical time-of-flight distance camera making use of lock-in pixels. Depending on the available light level, the background illumination and the modulation frequency, such solid-state range cameras can have a resolution of a few mm.

Seeing Chips

Already, in the mid-1980s, researchers were excited about the possibility of integrating local signal processing capabilities in each pixel, implementing some functions of biological vision systems. It was envisioned that an increasing amount of bio-inspired functionality would lead to smart image sensors, exhibiting a primitive sense of vision. The idea of a 'seeing chip', with the capability to interpret a scene, was born.

Unfortunately, even today, our understanding of the enormously successful biological vision systems has not yet reached a mature enough state allowing us to implement signal processing functions that approximate any 'image understanding'. Published algorithms are still incapable of interpreting a scene and 'perceiving' its contents. Obviously, no set of vision algorithms is yet known with which seeing chips could be realized. Although progress is slower than envisioned, single-chip machine vision systems of growing complexity and increasing functionality are being developed.

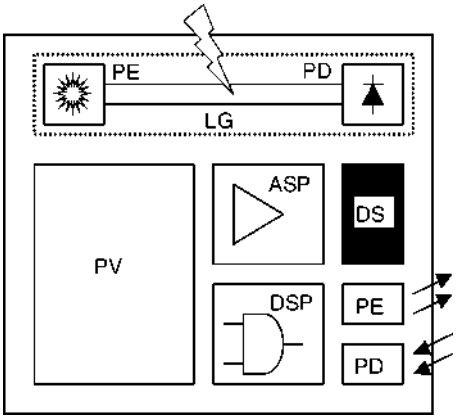

Figure 15 Schematic block diagram of a monolithically integrated photonic microsystem, that is expected to be producable in the near future with organic semiconductors. All optoelectronic devices are fabricated on the same planar surface: photovoltaic power supplies (PV); analog and digital signal processing (ASP and DSP); photodetectors (PD); photoemitters (PE) – used for bidirectional communication or as parts of a photonic detector subsystem – a light guiding structure (LG) that can be sensitized for specific sensing applications, and a display subsystem (DS).

Organic Semiconductors for Monolithic Photonic Microsystems

The realization that organic materials with conjugated double bonds, in particular noncrystalline polymers, are direct bandgap semiconductors, has recently led to a flurry of R&D activities in the domain of organic photonics. Organic light-emitting diodes (OLEDs), color displays, lighting panels, photodiodes, image sensors, photovoltaic cells, RF transponders, and tags, and several other organic photonic devices have not only been demonstrated but many are already commercially available. The special appeal of organic semiconductors stems from their potential for monolithic, low-cost integration of complete, large-area photonic microsystems, even on thin, flexible substrates. In Figure 15, such a monolithic photonic microsystem is illustrated schematically, combining a photovoltaic power supply, a photonic detector subsystem, analog and digital signal processing, as well as bidirectional optical communications with the outside world, possibly including also an on-chip display. The photonic detector subsystem typically consists of a light source, a light guiding structure that is sensitized – for example chemically – and a photodetector/image sensor subsection. In such a monolithic photonic microsystem, the term 'smart pixel array' obtains a much broader meaning, opening up a huge number of practical applications in essentially all domains of our daily and professional lives. This domain of organic semiconductors for integrated photonic microsystems is still largely unexplored, and it is offering exciting opportunities for research and product development.

Smart Pixel Arrays Beyond Electromagnetic Field Imaging

The sensing capabilities of smart pixel arrays are not restricted to electromagnetic radiation; this technology can also be used for the acquisition of other two- or three-dimensional physical fields. As an example, electrostatic imaging can be used for the temporally and spatially resolved acquisition of surface potential distributions in living cells. Electrostatic potential values of several hundred mV can be acquired with a precision of 1 mV and a temporal resolution below 1 millisecond. Other acquisition modalities include magnetostatic field imaging, for example by employing the Hall effect, or imaging of local electrostatic charge or current density distributions.

Summary

The relentless advances of semiconductor technology make it possible to integrate an increasing amount of photosensitive, analog, and digital functionality in the same smart pixel and on the same smart pixel array. Unusual photosensor devices can be fabricated with commercially available

semiconductor processes, implementing innovative designs and layouts, as well as novel electronic circuits, with which the physical limits of photosensing and photocharge processing may be reached. Novel opto-electronic microsystems can be developed, based on application-specific smart pixel arrays, either by realizing known optical measurement techniques more efficiently or by implementing new techniques, whose realization requires the availability of nonconventional image sensing functionality.

Already, smart pixel arrays are opening up a large number of new opportunities in photonics research and product development. The continuously increasing levels of integration lead to new forms of system-on-chip solutions for a wide range of practical applications. Smart pixel arrays can perform at the basic limits imposed by physics, and they have become indispensable tools in almost any type of research laboratory and any field of science and engineering.

List of Units and Nomenclature

Bandwidth	B	s^{-1}
Boltzmann's constant	k	J/K
Capacitance	C	F
Charge	Q	C
Charge carrier mobility	μ	$m^2/(V\ s)$
Charge noise variance	s_Q	C^2
Diffusivity	D	m^2/s
Distance	L	m
Dynamic range	D/R	dB
Field effect transistor transconductance	g	A/V
Integrated power measured with shaped photodiode	r	W
Intensity of electromagnetic wave	S, A, B	W/m^2
One-dimensional light intensity distribution	$p(x)$	W/m^2
One-dimensional width distribution	$w(x)$	m
Phase delay	φ	–
Pixel value at position (i,j) in an image	$P_{i,j}; R_{i,j}$	–
Temperature	T	K
Temporal frequency	ω	s^{-1}
Time	t	s
Two-dimensional convolution kernel	$w_{k,l}$	–

See also

Electromagnetically Induced Transparency. **Photonic Crystals:** Photonic Crystal Lasers, Cavities and Waveguides. **Semiconductor Materials:** Group IV Semiconductors, Si/SiGe.

Further Reading

Forrest SR (2000) Active optoelectronics using thin-film organic semiconductors. *IEEE Journal on Selected Topics in Quantum Electronics* 6: 1072–1083.

Fossum ER (1997) CMOS image sensors: electronic camera-on-a-chip. *IEEE Transactions on Electron Devices* 44: 1689–1698.

Hosticka BJ, Seitz P and Simoni A (2004) Optical time-of-flight range imaging for solid-state 3D vision. In: Grimes G and Dickey EC (eds) *Encyclopedia of Sensors*. Stevenson Ranch, CA: American Scientific Publishers (in print).

Koch C (1989) Seeing chips: analog VLSI circuits for computer vision. *Neural Computation* 1: 184–200.

Lutz G (1999) *Semiconductor Radiation Detectors*. Berlin: Springer.

Mead C (1989) *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley.

Saleh BEA and Teich MC (1991) *Fundamentals of Photonics*. New York: John Wiley and Sons.

Seitz P (1999) Solid-state image sensing. In: Jaehne B (ed.) *Handbook of Computer Vision and Applications*. London: Academic Press.

Seitz P and Blanc N (2003) CCD and APS/CMOS technology for smart pixels and image sensors. *Proceedings of SPIE* 5251 (in print).

Sze SM (1994) *Semiconductor Sensors*. New York: John Wiley and Sons.

Theuwissen AJP (1995) *Solid State Imaging with Charge Coupled Devices*. Dordrecht, Germany: Kluwer.

DIFFRACTION

Contents

Diffraction Gratings
Fraunhofer Diffraction
Fresnel Diffraction

Diffraction Gratings

J Turunen and T Vallius, University of Joensuu, Joensuu, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The classic textbook example of a diffraction grating is a periodic arrangement of parallel opaque wires: if a plane wave is incident upon such a structure, it is divided into a multitude of plane waves propagating in well-defined directions. In particular, the dramatic wavelength-dependence of these directions has attracted great interest ever since the potential of diffraction gratings in spectroscopy was realized in the late 1700s. Today spectroscopy is only one, though important, application of diffraction gratings. Gratings of many different forms belong to the basic building blocks in modern wave-optics-based design of optical elements and systems.

Grating Equations

Some examples of the wide variety of possible grating structures are illustrated in Figure 1, which shows the profiles of some commonly encountered diffraction gratings that are invariant in the y direction. These are known as linear gratings. Each grating may be of reflection type (the refractive index n_{III} is complex) or of transmission type (n_{III} is real). In all cases n_1 and n_2 can both be real or one of them can be complex. Here we assume that the incident field is a plane wave with wave vector $\boldsymbol{k}$ in the x–z plane; more general illumination waves can be represented as superpositions of plane waves.

Figure 1a illustrates a binary grating: in the modulated region $0 < z < H$ the refractive index alternates between n_1 and n_2, changing from one value to another at certain transition points within the period d. The structure in Figure 1b represents a sinusoidal surface-relief grating as an example of smooth grating profiles. Finally, Figure 1c shows a triangular grating profile.

The periodicity of a grating, illuminated by a plane wave, usually leads to the appearance of a set of reflected and transmitted plane waves (diffraction orders of the grating) that propagate in directions determined by the grating period d, the illumination wavelength λ, and the refractive indices n_{I} and n_{III} (see Figure 2a). These directions are obtained from the transmission grating equation:

$$n_{\mathrm{III}} \sin \theta_m = n_{\mathrm{I}} \sin \theta + m\lambda/d \qquad [1]$$

where m is the index of the diffraction order and θ_m is its propagation angle. For reflected orders (superscript r) one simply replaces n_{III} by n_{I} in eqn [1].

Obviously only a limited number of orders satisfy the condition $|\theta_m| < \pi/2$ and thereby represent conventional propagating plane waves. Higher, so-called evanescent orders are inhomogeneous waves that propagate along the grating surface and decay exponentially in the z direction. However, they cannot be ignored in grating analysis, as we shall see.

In the case of a biperiodic grating with a period $d_x \times d_y$ the propagation direction of the incident beam is defined by two angles θ and ϕ illustrated in Figure 2b. The grating equations now read as:

$$n_{\mathrm{III}} \sin \theta_{mn} \cos \phi_{mn} = n_{\mathrm{I}} \sin \theta \cos \phi + m\,\lambda/d_x \qquad [2]$$

$$n_{\mathrm{III}} \sin \theta_{mn} \sin \phi_{mn} = n_{\mathrm{I}} \sin \theta \sin \phi + n\lambda/d_y, \qquad [3]$$

where θ_{mn} and ϕ_{mn} define the propagation directions of the diffraction order with indices m and n. For reflected orders n_{III} is again replaced by n_{I}. These expressions apply also to the special case of linear gratings (limit $d_y \to \infty$) illuminated by a plane

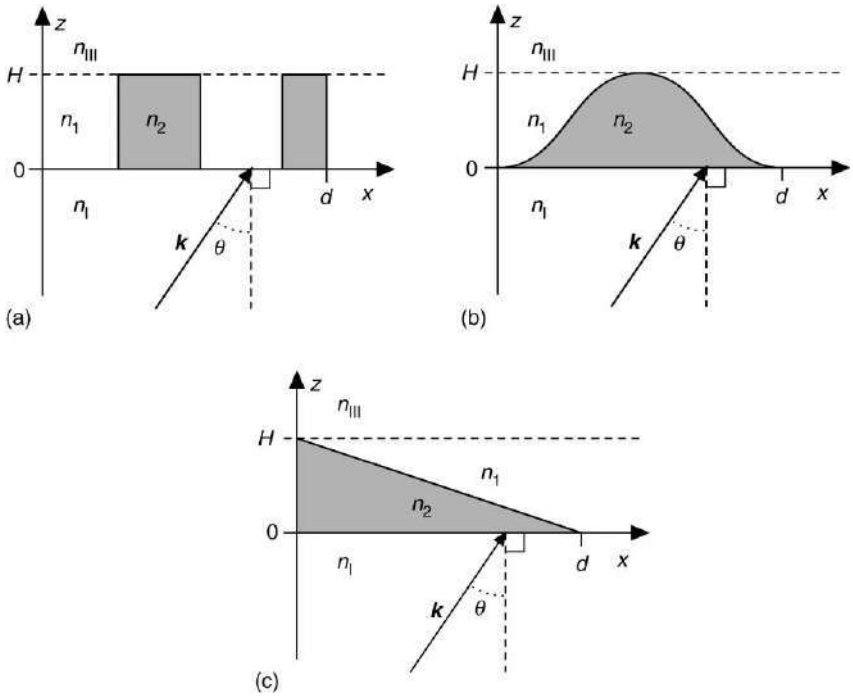

Figure 1 Examples of linear gratings. (a) Binary grating; (b) sinusoidal grating; (c) triangular grating. Here d is the grating period, H is the thickness of the modulated region, θ is the angle of incidence, and n denotes refractive index.

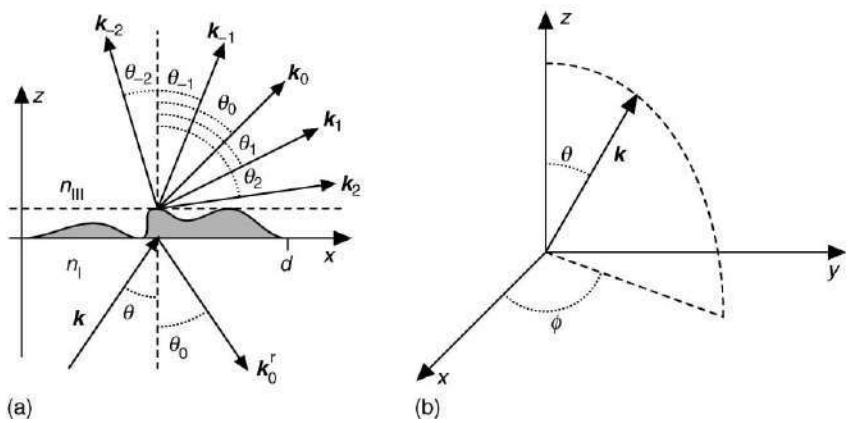

Figure 2 Diffraction by gratings. (a) Linear grating; (b) definitions of direction angles in the case of a biperiodic grating.

wave not propagating in the $x-z$ plane (conical incidence).

Some two-dimensionally periodic gratings are illustrated in Figure 3. The structure illustrated in Figure 3a is an extension of a binary grating to the biperiodic geometry: here a single two-dimensional feature is placed within the grating period but, of course, more features with different outlines could be included as well. Moreover, each period could contain a continuous 'mountain-range' or consist of a 'New-York'-style pixel structure. Figure 3b illustrates a structure consisting of an array of rectangular-shaped holes pierced in a flat screen, while Figure 3c defines an array of isolated particles on a flat substrate. If the grid or the particles are metallic, surrounded by dielectrics from all sides, one speaks of inductive and capacitive structures, respectively: in the former case relatively free-flowing currents are induced in the metallic grid structure, but in the latter case the isolated metal particles become polarized (and therefore capacitive) when illuminated by an electromagnetic field.

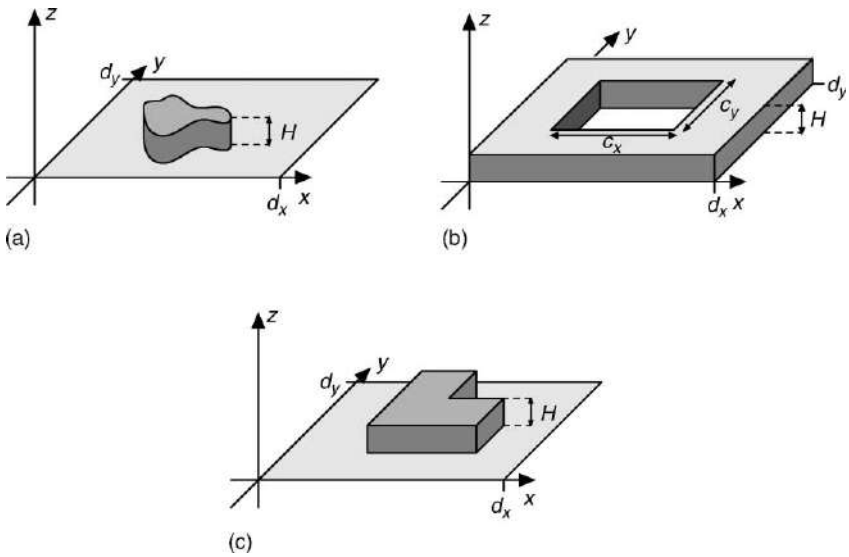

Figure 3 Examples of biperiodic gratings. (a) General binary grating; (b) inductive structure; (c) capacitive structure.

Grating Theory

Gratings equations provide exact knowledge of the propagation directions of the diffraction orders, but they give no information about the distribution of light among different orders. To obtain such information, one must in general solve Maxwell's equations exactly (rigorous diffraction theory), but within certain limits simpler methods are available. In all cases, however, the transmitted and reflected fields are represented as superpositions of plane waves that propagate in the directions allowed by the grating equations (these are known as Rayleigh expansions).

The most popular method for rigorous grating analysis is the Fourier Modal Method (FMM), though many alternative analysis methods exist. To understand the principle of FMM, consider first binary gratings. The field inside the modulated region $0 < z < H$ is represented as a superposition of waveguide modes, determined by imagining that the modulated region would extend from $x = -\infty$ to $x = \infty$. With given transition points the electric permittivity, its inverse, and the electric field inside the grating, can be represented in the form of Fourier series and one ends up with a matrix eigenvalue problem with finite dimensions when the matrices are truncated. A numerical solution of this problem provides a set of orthogonal modes that can propagate in the structure with certain propagation constants. The relative weights of the modes are then determined by applying the electromagnetic boundary conditions at the planes $z = 0$ and $z = H$ to match the modal expansions inside the grating with the plane-wave (Rayleigh) expansions of the fields in the homogeneous materials on both sides of the grating. This matching procedure, which is numerically implemented by solving a set of simultaneous linear equations, yields the complex amplitudes of the electric and magnetic field components associated with the reflected and transmitted diffraction orders of the grating.

Of main interest are the diffraction efficiencies, η_m (or η_{mn} for biperiodic gratings) of the orders, which tell the distribution of energy between different propagating orders. In electromagnetic theory the diffraction efficiency is defined as the z-component of the time-averaged Poynting vector associated with the particular order, divided by that of the incident plane wave.

Gratings with z-dependent surface profiles can be handled similarly if they are 'sliced' into a sufficient number of layers, in each of which the grating is modeled as a binary structure. The electromagnetic boundary conditions are applied at each slice boundary and the approach models the real continuous structure with arbitrary degree of accuracy when the number of the slices is increased; in practice it is usually sufficient to use approximately 20–30 slices for sufficiently accurate modeling of any continuous grating profile by FMM.

The main problem with rigorous techniques such as FMM is bad scaling. If, in the case of 1D modulated (or y-invariant) gratings, the period is increased by a factor of two, the computation time required to solve the eigenvalue problem is increased by a factor of around eight. Thus we have a d^3 type

scaling law. The situation is much worse for bi-periodic gratings, for which the exponent is around six. Thus no foreseeable developments in computer technology will solve the problem, and approximate analysis methods have to be sought. A multitude of such methods are under active development, but in this context we discuss the simplest example only.

If $d \gg \lambda$ and the grating profile does not contain wavelength-scale features, one can describe the effects of a grating on the incident field merely by considering it as a complex-amplitude-modulating object, which imposes a position-dependent phase shift and amplitude transmission (or reflection) coefficient on the incident field. This model is known as the thin-element approximation (TEA).

At normal incidence the plane-wave expansion of any scalar field component U immediately behind the grating (at $z = H$) takes the form:

$$U(x, y, H) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} T_{mn} \times \exp[\mathrm{i}2\pi(mx/d_x + ny/d_y)] \qquad [4]$$

where

$$T_{mn} = \frac{1}{d_x d_y} \int_0^{d_x} \int_0^{d_y} U(x, y, H) \times \exp[-\mathrm{i}2\pi(mx/d_x + ny/d_y)]\mathrm{d}x\,\mathrm{d}y \qquad [5]$$

are the complex amplitudes of the transmitted diffraction orders. In TEA the field at $z = H$ may be connected to the field at $z = 0$ (a unit-amplitude plane wave is assumed here) by optical path integration along a straight line though the modulated region of the grating:

$$U(x, y, H) = \int_0^H \exp\left[\mathrm{i}\frac{2\pi}{\lambda}\hat{n}(x, y, z)\right]\mathrm{d}z \qquad [6]$$

where $\hat{n}(x, y, z) = n(x, y, z) + \mathrm{i}\kappa(x, y, z)$ is the complex refractive-index distribution in the region $0 < z < H$. With TEA the diffraction efficiencies can be calculated from a simple formula:

$$\eta_{mn} = |T_{mn}|^2 \qquad [7]$$

which, however, ignores Fresnel reflections.

Equations [5]–[7] allow one to obtain closed-form expressions for the diffraction efficiencies η_{mn} in many grating geometries. For example, a y-invariant triangular profile with $H = \lambda/|n_{\mathrm{I}} - n_{\mathrm{III}}|$ gives $\eta_{-1} = 100\%$ if $n_1 = n_{\mathrm{III}}$, $n_2 = n_{\mathrm{I}}$, and $\theta = 0$. For a Q-level stair-step-quantized version of this grating we obtain:

$$\eta_m = \sin c^2(1/Q) \qquad [8]$$

where $\sin cx = \sin(\pi x)/(\pi x)$. It should be noted, however, that these results are valid only for gratings with $d \gg \lambda$, even if they are corrected by Fresnel transmission losses at the interface between the two media.

Fabrication of Gratings

Gratings can be manufactured in a number of different ways, perhaps the simplest being to plot an equidistant array of straight black and white lines and to reduce the scale of the pattern photographically. This method can be readily adapted to the generation of more complex grating structures and diffractive elements, but it is not suitable for the production of high-quality gratings required in, for example, spectroscopy.

Ruling engines represent the traditional way of making gratings with periods of the order of the wavelength of light, especially for spectroscopic applications. These are machines equipped with a triangularly shaped diamond ruling edge and an interferometrically controlled two-dimensional mechanical translation stage that moves the substrate with respect to the ruling edge. Triangular-profile gratings of high quality can be produced with such machines, but the fabrication process is slow and the cost of the manufactured gratings is therefore high. In addition, even with real-time interferometric control of the ruling-edge position, it is impossible to avoid small local errors in the grating geometry. Such errors are particularly harmful if they are periodic since, according to the grating equation, these errors will generate spurious spectral lines known as ghosts. Such lines generated by the grating 'super-period' are weak but they may be confused with real spectral lines in the source spectrum. If the ruling errors are random, they result in 'pedestal-type' background noise. While such errors do not introduce ghosts, they nevertheless reduce the visibility of weak original spectral lines.

The introduction of lasers in the early 1960s made it possible to form stable and high-contrast interference patterns over a sufficient time-scale to allow the exposure of photographic films and other photosensitive materials such as polymeric photoresists. This interferometric technique offers a direct method for the fabrication of large diffraction gratings without appreciable periodic errors. With this technique it is possible to make highly regular

gratings for spectroscopy, but it does not permit easy generation of arbitrarily shaped grating structures. The two intersecting plane waves can also be generated with a linear grating using diffraction orders $m = +1$ and $m = -1$, suppressing the zero order by appropriate design of the grating profile and choosing the period in such a way that all higher transmitted orders are evanescent. This so-called phase mask technique is suitable, for example, for the exposure of Bragg gratings in fibers.

To generate arbitrary gratings one usually employs lithographic techniques. Some of the possible processes are illustrated in Figure 4. The exposure can be performed with scanning photon, electron, or ion beams, etc. The process starts with substrate preparation, i.e., coating it with a beam-sensitive layer known as resist, and in the case of electron beam exposure with a thin conductive layer to prevent charging. Following the exposure the resist is developed chemically, and in this process either the exposed or nonexposed parts of the resist are dissolved (depending on whether the resist is of positive or negative type), leading to a surface-relief profile. Several alternative fabrication steps may follow the development, depending on the desired grating type. If a high-contrast resist is used, a slightly 'undercut' profile as illustrated in Figure 4c is obtained. This can be coated with a metal layer and the resist can be 'lifted off' chemically. Thus a binary metallic grating is formed. To produce a binary dielectric profile one bombards the substrate with ions either purely mechanically (ion beam milling) or with chemical assistance (reactive ion etching), and finally the residual metal is dissolved.

Alternatively, the particle beam dose can be varied with position to obtain an analog profile in the resist after development. This resist profile can be transformed into the substrate by proportional etching, which leads to an analog dielectric surface-relief structure (see Figure 4f). This profile can be made

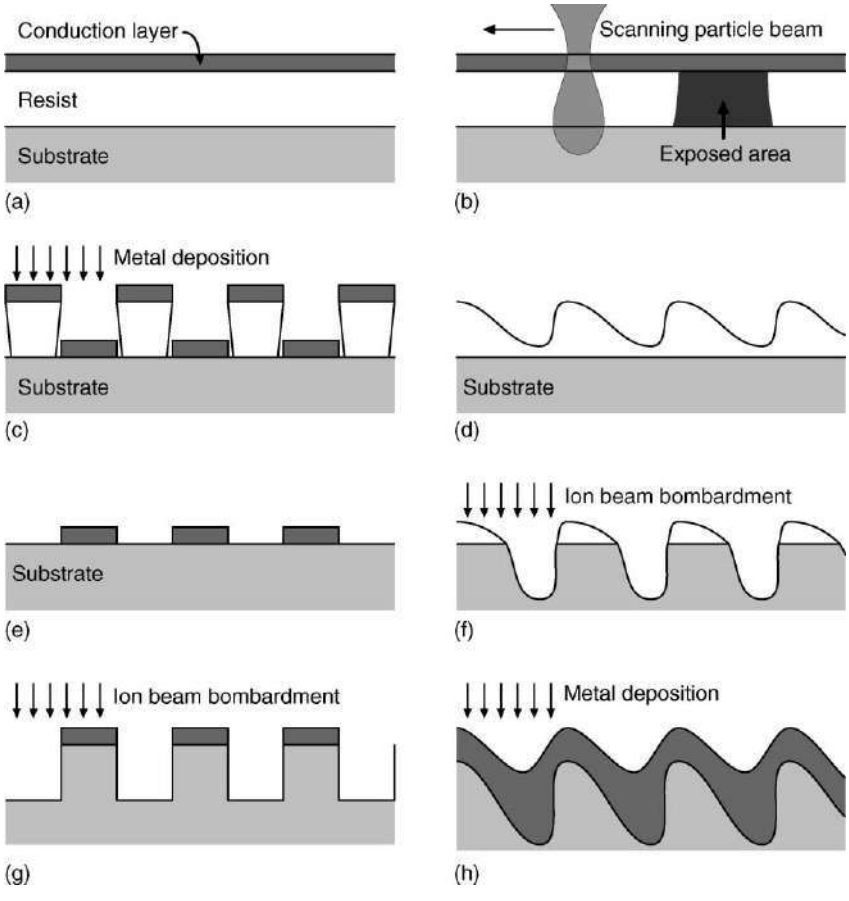

Figure 4 Lithographic processes for grating fabrication. (a) Substrate after preparation for electron beam exposure; (b) exposure by a scanning electron beam; (c) slightly undercut profile in hight-contrast resist after development, under metal film evaporation; (d) analog resist profile after variable-dose exposure and resist development; (e) finished binary metallic grating after resist lift-off; (f) transfer of the analog surface-relief profile into the substrate by proportional etching; (g) etching of a binary profile into the substrate with a metal mask; (h) growth of a metal film on an analog surface profile.

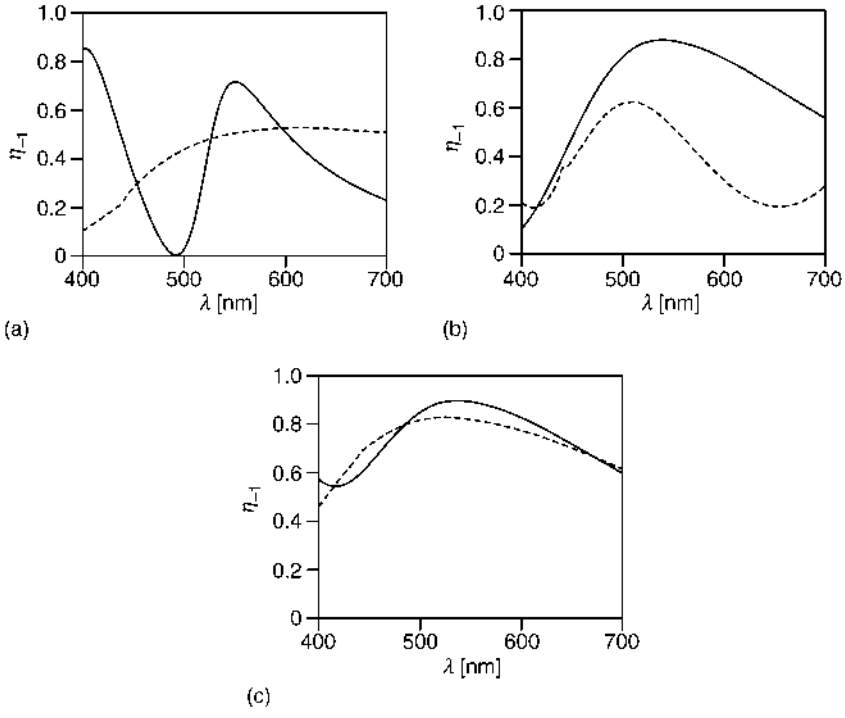

Figure 5 Spectral diffraction efficiency curves of spectroscopic gratings in TE polarization. (a) Binary; (b) sinusoidal; and (c) triangular reflection gratings illuminated at Bragg angle. Solid lines: $d = \lambda_0$. Dashed lines: $d = 2\lambda_0$.

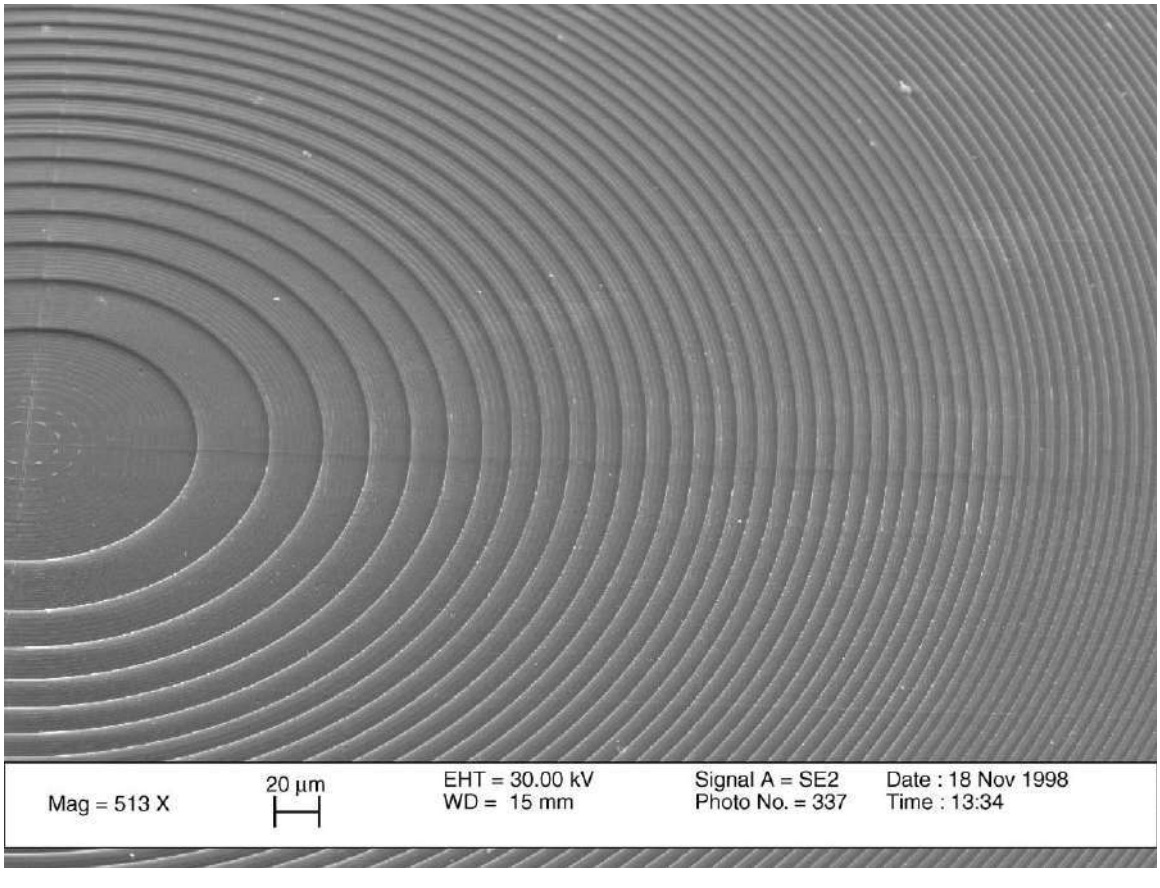

Figure 6 A scanning electron micrograph of a part of a diffractive lens.

reflective by depositing a thin metal film on top of it. If large-scale fabrication of identical gratings is the goal, a thick ($\sim 100\ \mu$m) metal layer (for example nickel) can be grown on the surface by electroplating. After separation from the substrate, the electroplated metal film, which is a negative of the original profile, can be used as a stamp in replicating the original profile in plastics using methods such as hot embossing, ultraviolet radiation curing, and injection molding. In particular, the latter method provides the key to high-volume and low-cost production of various types of diffraction gratings and other grating-based diffractive structures in the industrial scale.

Some Grating Applications

We now proceed to describe a collection of representative grating applications. More exhaustive coverage of grating applications can be found in the Further Reading section at the end of this article.

Spectroscopic Gratings

The purpose of a spectroscopic grating is to divide the incident light into a spectrum, i.e., to direct different frequency components of the incident beam into different directions given by the grating (eqn [1]). When the grating period is reduced, the angular dispersion $\partial\theta_m/\partial\lambda$, and therefore the resolution of the spectrograph, increases.

In the numerical examples presented in Figure 5 we assume a reflection grating in a Littrow mount, i.e., $\theta_{\mathrm{m}}^{\mathrm{r}} = -\theta$. This mount is common in spectroscopy; it obviously requires that the grating is rotated when the spectrum is scanned. We see from eqn [1] that in the Littrow mount:

$$\sin\theta = \frac{m\lambda}{2d} \qquad [9]$$

By inserting this result back into eqn [1] and differentiating we obtain a quantitative measure for dispersion in the Littrow mount:

$$\frac{\partial\theta_m}{\partial\lambda} = -\frac{m}{2d\cos\theta} \qquad [10]$$

We consider TE polarization (electric field is linearly polarized in the y-direction) and order $m = -1$, optimize H to give maximum efficiency at $\lambda_0 = 550$ nm, and plot the diffraction efficiency η_{-1} over the visible wavelength range for two values of the grating period, $d = \lambda_0$ and $d = 2\lambda_0$. Rigorous diffraction theory is used with $n_{\mathrm{I}} = n_2 = 1$, and the metal is assumed to be aluminum (n_1 and n_{III} are complex-valued).

In Figure 5a binary gratings with one groove (width $d/2$) per period is considered. Parametric optimization gives $H = 422$ nm if $d = 1\lambda$, and $H = 168$ nm if $d = 2\lambda$. Figure 5b illustrates the results for sinusoidal gratings with $H = 345$ nm if $d = 1\lambda$, and

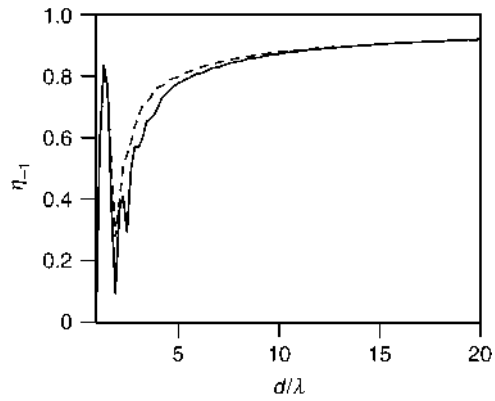

Figure 7 Diffraction efficiency versus wavelength for triangular transmission gratings illuminated at normal incidence. Solid lines: TE polarization. Dashed lines: TM polarization.

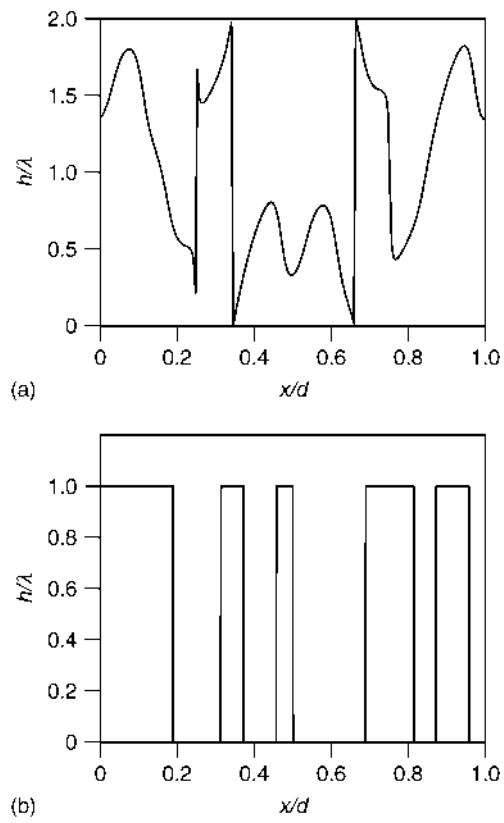

Figure 8 Surface-relief profiles of (a) continuous and (b) binary $1 \rightarrow 8$ beamsplitter gratings.

$H = 520$ nm if $d = 2\lambda$. Finally, triangular gratings with $H = 431$ nm if $d = 1\lambda$, and $H = 316$ nm if $d = 2\lambda$, are considered. The triangular grating profile is seen to be the best choice in general.

Diffractive Lenses

Diffractive lenses are gratings with locally varying period and groove orientation. Figure 6 illustrates the central part of a diffractive microlens fabricated by lithographic methods. It is seen to consist of concentric zones with radially decreasing width.

The phase transmission function of an ideal diffractive focusing lens can be obtained by a simple optical path calculation, and is given by

$$\phi(r) = \frac{2\pi}{\lambda_0}(f - \sqrt{f^2 + r^2}) \approx -\frac{\pi}{\lambda_0 f} r^2 \qquad [11]$$

where f is the focal length, r is the radial distance from the optical axis, and the approximate expression is valid in the paraxial region $r \ll f$. Here λ_0 is the design wavelength, which determines the modulation depth of the locally blazed grating profile:

$$H = \frac{\lambda_0}{n(\lambda_0) - 1} \qquad [12]$$

and $n(\lambda_0)$ is the refractive index of the dielectric material at $\lambda = \lambda_0$.

The zone boundaries r_n are defined by the condition $\phi(r_n) = -2\pi n$, and therefore we obtain from eqn [11]:

$$r_n = \sqrt{2nf\lambda_0 + (n\lambda_0)^2} \approx \sqrt{2nf\lambda_0} \qquad [13]$$

Obviously, when the numerical aperture of the lens is large, the zone width in the outer regions of the lens is reduced to values of the order of λ. Figure 7 illustrates the local diffraction efficiency in different part of a diffractive lens: we use the order $m = -1$ of a locally triangular diffraction grating of the

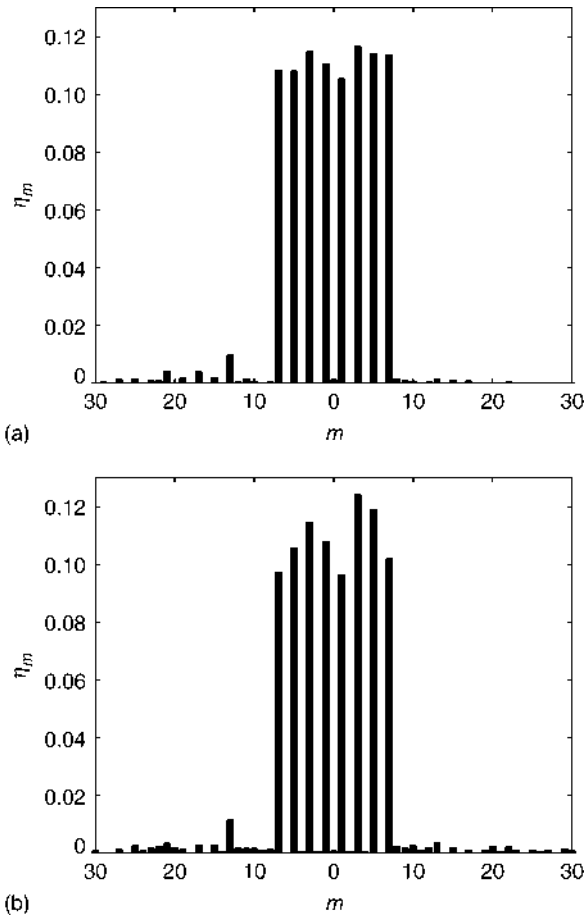

Figure 9 Diffraction pattern produced by a continuous-profile beamsplitter grating when (a) $d = 200\lambda$ and (b) $d = 50\lambda$.

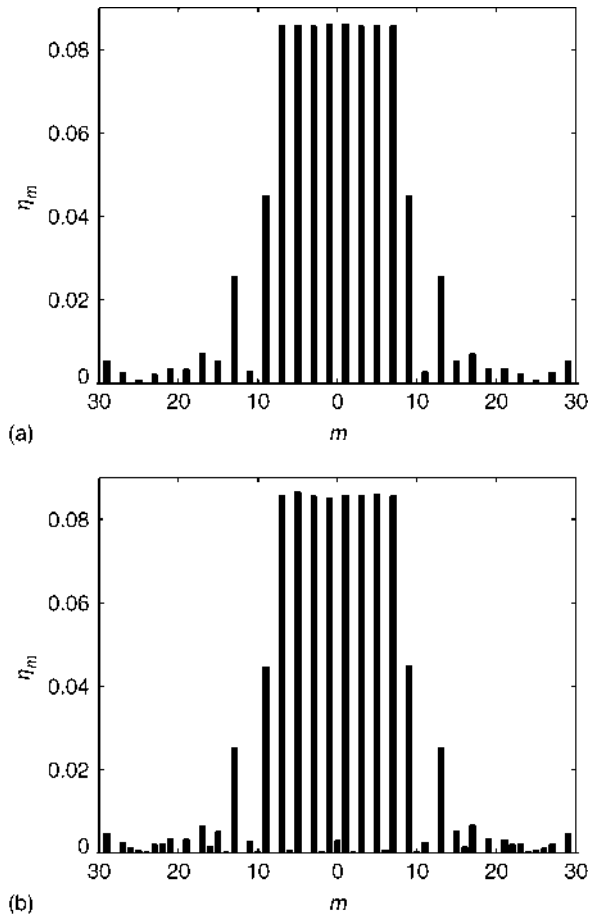

Figure 10 Diffraction pattern produced by a binary beamsplitter grating grating when (a) $d = 200\lambda$ and (b) $d = 50\lambda$.

type shown in Figure 1c with $\theta = 0$, assuming that $n_{\mathrm{I}} = n_2 = 1.5$ (fused silica) and $n_1 = n_{\mathrm{III}} = 1$. The efficiency η_{-1} is calculated as a function of λ using rigorous diffraction theory for both TE and TM polarization (the electric vector points in the groove direction in the TE case and is perpendicular to it in the TM case). It is seen that the local efficiency of the lens becomes poor in regions where the local grating period is small, and thus diffractive lenses operate best in the paraxial region (it is, however, possible to obtain improved non-paraxial performance by optimizing the local grating profile).

A factor that limits the use of diffractive lenses with polychromatic light is their large chromatic aberration: in the paraxial limit the focal length depends on the wavelength according to the formula:

$$f(\lambda) = \frac{\lambda_0}{\lambda} f \qquad [14]$$

However, since the dispersion of a grating has an opposite sign compared to the diffraction of a prism, a weak positive diffractive lens can be combined with a relatively strong positive refractive lens to form a single-element achromatic lens. Such a hybrid lens is substantially thinner and lighter than a conventional all-refractive achromatic lens made of crown and flint glasses.

Beam Splitter Gratings

Multiple beamsplitters, also known as array illuminators, are gratings with sophisticated periodic structure that are capable of transforming an incident plane wave into a set of diffraction orders with a specified distribution of diffraction efficiencies. Most often one wishes to obtain a one- or two-dimensional array of diffracted plane waves with equal efficiency. This goal can be achieved by appropriate design of the grating profile, which may be of binary, multilevel, or continuous form.

Figure 8 illustrates binary and continuous grating profiles capable of dividing one incident plane wave into eight diffracted plane waves (central odd-numbered orders of the grating) with equal efficiency in view of TEA. Figure 9 shows the distribution of energy among the different diffraction orders in and around the array produced by the continuous profile, evaluated by rigorous theory for two different grating periods, while Figure 10 gives corresponding results for the binary grating.

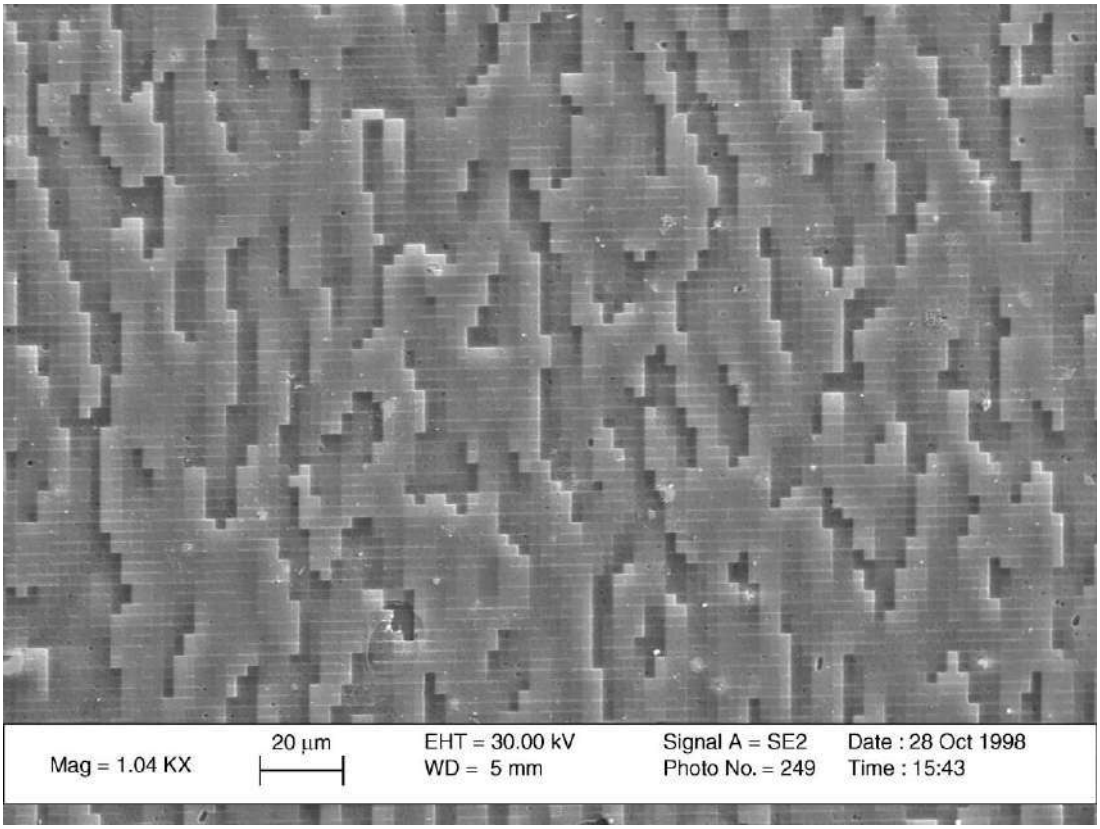

Figure 11 Scanning electron micrograph of a sophisticated beamsplitter grating.

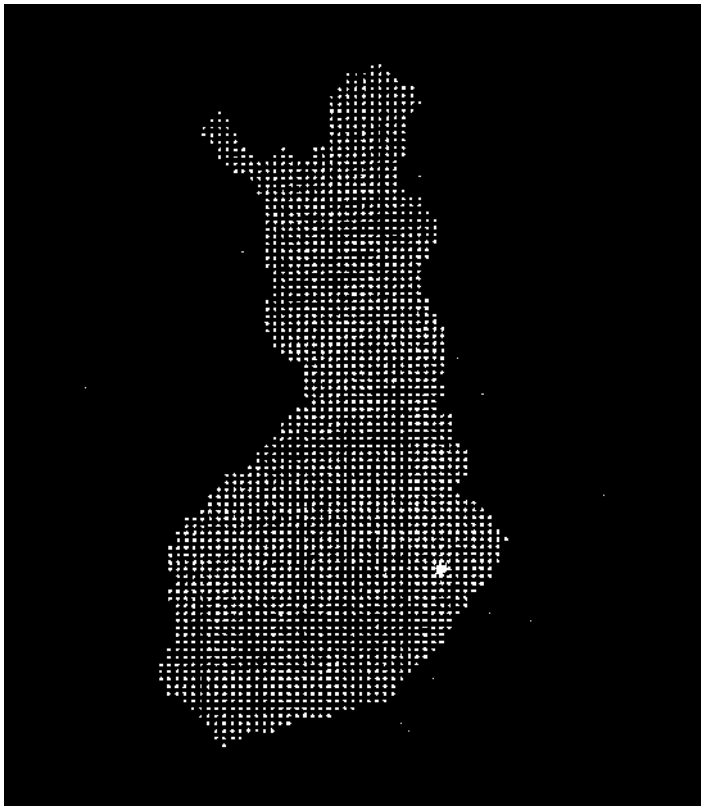

Figure 12 The far-field diffraction pattern produced by the grating in Figure 11.

The fraction of incident energy that ends up in the desired eight orders is 72% for the binary grating and 96% for the continuous grating according to TEA, making the latter more attractive from a theoretical viewpoint. However, the continuous profile is more difficult to fabricate accurately than the binary one, especially for small values of the ratio d/λ. Moreover, the array uniformity produced by the continuous profile degrades faster than that of the binary profile when the period is reduced.

An example of the structure of a biperiodic grating that is capable of forming a large, shaped array of diffraction orders with equal efficiency is presented in Figure 11. Figure 12 shows the far-field diffraction pattern produced by this element: the shape of Finland with the location of the town of Joensuu highlighted by placing the zero diffraction order on the appropriate position on the map (the apparent nonuniformity in the array is mostly due to the finite resolution of the CCD detector).

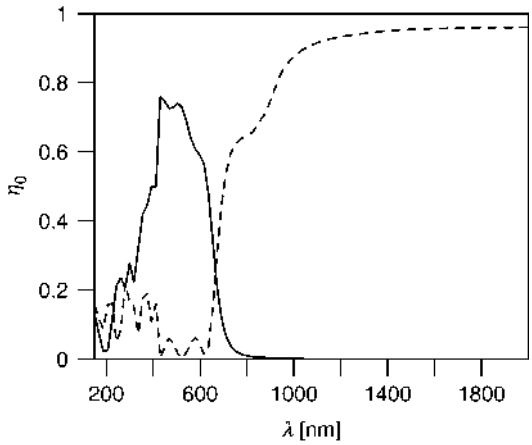

Figure 13 Zero order spectral transmission (solid line) and reflection (dashed line) of an aluminium inductive grid filter with 400 nm period.

Inductive Grid Filters

As a final example we consider a scaled-down version of a device well known from the door of a microwave oven: one can see in through the holes pierced in the metallic screen in the oven door, but the microwaves are trapped inside because their wavelengths are large in comparison with the dimensions of the holes. Another example of these inductive grid filters is found in large radio telescope mirrors for centimeter-scale

wavelengths, which are wide-grid structures rather than smoothly curved metal surfaces.

Figure 13 illustrates the spectral transmission and reflection curves of an inductive grid filter of the type illustrated in Figure 3b. We assume that $d_x = d_y = d = 400$ nm, $c_x = c_y = c = 310$ nm, $H = 500$ nm, and that the grid is a self-supporting structure ($n_{\mathrm{I}} = n_{\mathrm{III}} = 1$) made of aluminum. Only the zero reflected and transmitted orders propagate when $\lambda > d$, but other orders (with rather low efficiency) appear for smaller wavelengths. We see that the filter reflects infrared radiation effectively, but transmits a substantial part of radiation at visible and shorter wavelength regions.

List of Units and Nomenclature

Complex field amplitude	U
Diffraction angles [rad]	θ, ϕ
Diffraction efficiency	η
Focal length [m]	f, f_0
Grating period [m]	d
Grating thickness [m]	H
Number of quantization levels	Q
Phase function [rad]	$\phi(r)$
Refractive index	n
Wavelength [m]	λ, λ_0
Wave vector [m^{-1}]	$\mathbf{k}$

See also

Fiber Gratings.

Further Reading

Chandezon J, Maystre D and Raoult G (1980) A new theoretical method for diffraction gratings and its numerical application. *Journal of Optics* 11: 235–241.

Gaylord TK and Moharam MG (1985) Analysis and applications of optical diffraction by gratings. *Proceedings of IEEE* 73: 894–937.

Herzig HP (ed.) (1997) *Micro-optics: Elements, Systems and Applications*. London: Taylor & Francis.

Hutley MC (1982) *Diffraction Gratings*. London: Academic Press.

Li L (2001) Mathematical reflections on the fourier modal method in grating theory. In: Bao G, Cowsar L and Masters W (eds) *Mathematical Modelling in Optical Science*, pp. 111–139. Philadelphia: SIAM.

Li L, Chandezon J, Granet G and Plumey J-P (1999) Rigorous and efficient grating-analysis method made easy for optical engineers. *Applied Optics* 38: 304–313.

Petit R (ed.) (1980) *Electromagnetic Theory of Gratings*. Berlin: Springer-Verlag.

Rayleigh JWS (1907) On the dynamical theory of gratings. *Proceeding of the Royal Society A* 79: 399.

Solymar L and Cooke DJ (1981) *Volume Holography and Volume Gratings*. London: Academic Press.

Turunen J, Kuittinen M and Wyrowski F (2000) Diffractive optics: electromagnetic approach. In: Wolf E (ed.) *Progress in Optics XL*. Amsterdam: Elsevier.

Turunen J and Wyrowski F (eds) (1997) *Industrial and Commercial Applications of Diffractive Optics*. Berlin: Akademie Verlag.

van Renesse RL (1993) *Optical Document Security*. Boston, MA: Artech House.

Fraunhofer Diffraction

B D Guenther, Duke University, Durham, NC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Geometrical optics does a reasonable job of describing the propagation of light in free space unless that light encounters an obstacle. Geometrical optics predicts that the obstacle would cast a sharp shadow. On one side of the shadow's edge would be a bright uniform light distribution, due to the incident light, and on the other side there would be darkness. Close examination of an actual shadow edge reveals, however, dark fringes in the bright region and bright fringes in the dark region. This departure from the predictions of geometrical optics was given the name diffraction by Grimaldi.

There is no substantive difference between diffraction and interference. The separation between the two subjects is historical in origin and is retained for pedagogical reasons. Since diffraction and interference are actually the same physical process, we expect the observation of diffraction to be a strong function of the coherence of the illumination. With incoherent sources, geometrical optics is usually all that is needed to predict the performance of an optical system and diffraction can be ignored except at dimensions on the scale of a wavelength. With light sources having a large degree of spatial coherence, it is impossible to neglect diffraction in any analysis.

Even though diffraction produced by most common light sources and objects is a small effect, it is possible to observe diffraction without special equipment. By viewing a light source through a slit formed by two fingers that are nearly touching, fringes can be observed. Diffraction is used by photographers and television cameramen to provide artistic highlights in an image containing small light sources. They accomplish this by placing a screen in front of the camera lens. Light from point sources in the field of view is diffracted by the screen and produces 'stars' in the image. A third place where diffraction effects are easily observed is when a mercury or sodium streetlight, a few hundred meters away, is viewed through a sheer curtain.

Fresnel originally developed an approximate scalar theory based on Huygens' principle which states that:

> Each point on a wavefront can be treated as a source of a spherical wavelet called a secondary wavelet or a Huygens' wavelet. The envelope of these wavelets, at some later time, is constructed by finding the tangent to the wavelets. The envelope is assumed to be the new position of the wavefront.

Rather than simply using the wavelets to construct an envelope, Fresnel's scalar theory assumes that the Huygens' wavelets interfere to produce a new wavefront (*see* **Diffraction:** Fresnel Diffraction).

A rigorous diffraction theory is based on Maxwell's equations and uses the boundary conditions associated with the obstacle to calculate a field scattered by the obstacle. The origins of this scattered field are currents induced in the obstacle by the incident field. The scattered field is allowed to interfere with the incident field to produce a resultant diffracted field. The application of the rigorous theory is very difficult and for most problems an approximate scalar theory developed by Kirchhoff and Sommerfeld is used.

Kirchhoff Theory of Diffraction

Kirchhoff's theory was based on the elastic theory of light but can be reformulated into a vector theory. Here we will limit our discussion to the scalar formulation. Given an incident wave, φ, we wish to calculate the optical wave at point P_0, in Figure 1, located at $\mathbf{r}_0$, in terms of the wave's value on a surface, S, that we construct about the observation point.

To obtain a solution to the wave equation at the point, P_0, we select as a Green's function the one proposed by Kirchhoff (Box 1), a unit amplitude spherical wave, expanding about the point at $\mathbf{r}_0$, denoted by Ψ, and then apply Green's theorem (see Box 2). We will run into boundary condition problems as we proceed with the derivation of the solution. Sommerfeld removed the problems by assuming a slightly more complex Green's function (see Box 1). We will limit our discussion to the Kirchhoff Green's function.

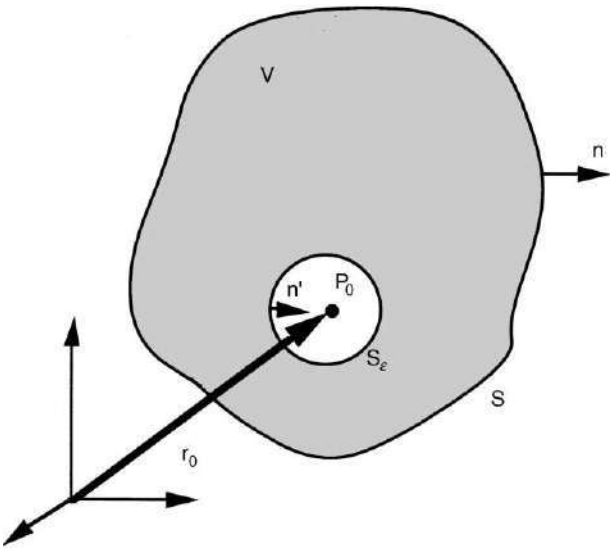

Figure 1 Region of integration for solution of Kirchhoff's diffraction integral. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

Green's theorem requires that there be no sources (singularities) inside the surface S but our Green's function [1] is based on a source at r_0. We eliminate the source by constructing a small spherical surface S_ε, of radius ε, about $\mathbf{r}_0$, excluding the singularity at $\mathbf{r}_0$ from the volume of interest, shown in gray in Figure 1. The surface integration that must be performed in Green's theorem is over the surface $S' = S + S_\varepsilon$.

Within the volume enclosed by S', the Green's function, Ψ, and the incident wave, φ, satisfy the scalar Helmholtz equation so that the volume integral can be written as

$$\iiint_V (\Psi\nabla^2\varphi - \varphi\nabla^2\Psi)\mathrm{d}\nu = -\iiint_V (\Psi\varphi k^2 - \varphi\Psi k^2)\mathrm{d}\nu \qquad [7]$$

The right side of [7] is identically equal to zero. This fact allows us to use [6] to produce a simplified statement of Green's theorem:

$$\iint_{S'} \left(\Psi\frac{\partial\varphi}{\partial n} - \varphi\frac{\partial\Psi}{\partial n}\right)\mathrm{d}s = 0 \qquad [8]$$

Box 1

Green's Function

1. **Kirchhoff Green's function**: Assume that the wave, Ψ, is due to a single point source at $\mathbf{r}_0$. At an observation point, $\mathbf{r}_1$, the Green's function is given by

$$\Psi(\mathbf{r}_1) = \frac{e^{-i\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}, \qquad [1]$$

where $r_{01} = |\mathbf{r}_{01}| = |\mathbf{r}_1 - \mathbf{r}_0|$ is the distance from the source of the Green's function, at $\mathbf{r}_0$, to the observation position, at $\mathbf{r}_1$.

2. **Sommerfeld Green's function**: Assume that there are two point sources, one at P_0 and one at P_0' (see Figure 2 where the source, P_0', is positioned to be the mirror image of the source at P_0 on the opposite side of the screen, thus

$$r_{01} = r_{01}' \qquad \cos(\hat{\mathbf{n}}, \mathbf{r}_{01}) = -\cos(\hat{\mathbf{n}}, \mathbf{r}_{01}') \qquad [2]$$

A Green's function that could be used is

$$\Psi' = \frac{e^{-i\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}} - \frac{e^{-i\mathbf{k}\cdot\mathbf{r}_{01}'}}{r_{01}'} \qquad [3a]$$

or

$$\Psi' = \frac{e^{-i\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}} + \frac{e^{-i\mathbf{k}\cdot\mathbf{r}_{01}'}}{r_{01}'} \qquad [3b]$$

Box 2

Green's Theorem

To calculate the complex amplitude, $\tilde{E}(\mathbf{r})$, of a wave at an observation point defined by the vector $\mathbf{r}$, we need to use a mathematical relation known as Green's Theorem. Green's theorem states that, if φ and Ψ are two scalar functions that are well behaved, then

$$\iint_S (\varphi\nabla\Psi\cdot\hat{\mathbf{n}} - \Psi\nabla\varphi\cdot\hat{\mathbf{n}})\mathrm{d}s = \iiint_V (\varphi\nabla^2\Psi - \Psi\nabla^2\varphi)\mathrm{d}\nu \qquad [4]$$

The vector identities

$$\nabla\Psi\cdot\hat{\mathbf{n}} = \frac{\partial\Psi}{\partial n} \qquad \nabla\varphi\cdot\hat{\mathbf{n}} = \frac{\partial\varphi}{\partial n} \qquad [5]$$

allow Green's theorem to be written as

$$\iint_S \left[\varphi\frac{\partial\Psi}{\partial n} - \Psi\frac{\partial\varphi}{\partial n}\right]\mathrm{d}s = \iiint_V [\varphi\nabla^2\Psi - \Psi\nabla^2\varphi]\mathrm{d}\nu \qquad [6]$$

This equation is the prime foundation of scalar diffraction theory but only the proper choice of Ψ, φ and the surface, S, allows direct application to the diffraction problem.

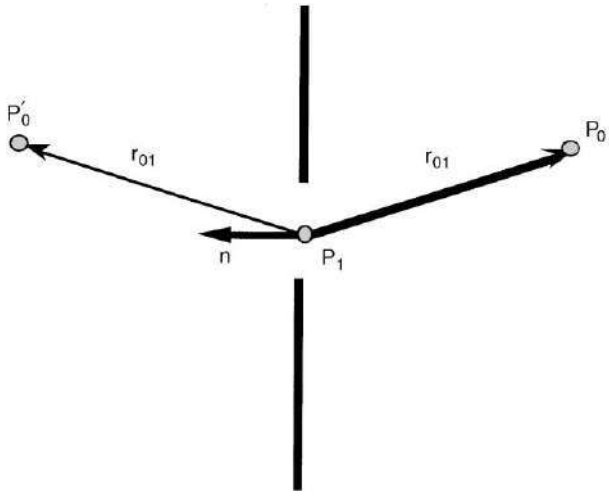

Figure 2 Geometry for the Sommerfeld Green's function. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

Because the integral over the combined surfaces, S and S_ε, is equal to zero, the integral over the surface S must equal the negative of the integral over the surface, S_ε:

$$-\iint_{S_\varepsilon}\left(\Psi\frac{\partial\varphi}{\partial n}-\varphi\frac{\partial\Psi}{\partial n}\right)\mathrm{d}s=\iint_{S}\left(\Psi\frac{\partial\varphi}{\partial n}-\varphi\frac{\partial\Psi}{\partial n}\right)\mathrm{d}s \quad [9]$$

To perform the surface integrals, we must evaluate the Green's function on the surface, S′. We therefore select $\mathbf{r}_1$, the vector defining the observation point, to be on either of the two surfaces that make up S′. The derivative, $\partial\Psi/\partial n$, to be evaluated is

$$\frac{\partial\Psi}{\partial n}=\frac{\partial\Psi}{\partial r}\frac{\partial r}{\partial n}$$
$$=\left(-\mathrm{i}k\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}-\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}^2}\right)\cos(\mathrm{n},\mathbf{r}_{01}) \quad [10]$$

where $\cos(\hat{\mathbf{n}},\mathbf{r}_{01})$ is the cosine of the angle between the outward normal $\hat{\mathbf{n}}$ and the vector $\mathbf{r}_{01}$, the vector between points P_0 and P_1. (Note from Figure 1 that the outward normal on S_ε is inward, toward P_0, while the normal on S is outward, away from P_0.)

For example if $\mathbf{r}_1$ were on S_ε, then

$$\cos(\mathbf{n},\mathbf{r}_{01})=-1, \quad [11]$$

$$\frac{\partial\Psi(\mathbf{r}_1)}{\partial n}=\left[\frac{1}{\varepsilon}+\mathrm{i}k\right]\frac{\mathrm{e}^{-\mathrm{i}k\varepsilon}}{\varepsilon}, \quad [12]$$

where ε is the radius of the small sphere we constructed around the point at P_0.

We now use [10] to rewrite the two integrals in [9]. The integral over the surface S_ε is

$$\iint_{S_\varepsilon}\left(\Psi\frac{\partial\varphi}{\partial n}-\varphi\frac{\partial\Psi}{\partial n}\right)\mathrm{d}s$$
$$=\iint_{S_\varepsilon}\left[\frac{\partial\varphi}{\partial n}\frac{\mathrm{e}^{-\mathrm{i}k\varepsilon}}{\varepsilon}-\varphi\frac{\mathrm{e}^{-\mathrm{i}k\varepsilon}}{\varepsilon}\left(\frac{1}{\varepsilon}+\mathrm{i}k\right)\right]$$
$$\times\varepsilon^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \quad [13]$$

while the integral over the surface S is

$$\iint_{S}\left[\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}\frac{\partial\varphi}{\partial n}-\varphi\frac{\partial}{\partial n}\left(\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}\right)\right]r^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \quad [14]$$

The omitted volume contained within the surface S_ε is allowed to shrink to zero by taking the limit as $\varepsilon\to 0$. Equation [14] will not be affected by taking the limit. The first and last terms of the right side of [13] go to zero as $\varepsilon\to 0$ because they contain $\varepsilon\mathrm{e}^{-\mathrm{i}k\varepsilon}$. The second term in [13] contains $\mathrm{e}^{-\mathrm{i}k\varepsilon}$ which goes to 1 as $\varepsilon\to 0$. Therefore, in the limit as $\varepsilon\to 0$, [9] becomes

$$\varphi(\mathbf{r}_0)=\frac{1}{4\pi}\iint_{S}\left[\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}\frac{\partial\varphi}{\partial n}-\varphi\frac{\partial}{\partial n}\left(\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}\right)\right]\mathrm{d}s \quad [15]$$

which is sometimes called the integral theorem of Kirchhoff. By carefully selecting the surface of integration in [15], we can calculate the diffraction field observed at P_0 produced by an aperture in an infinite opaque screen.

Assume that a source at P_2 in Figure 3 produces a spherical wave that is incident on an infinite, opaque screen from the left. To find the field at P_0 in Figure 3, we apply [15] to the surface $S_1+S_2+\Sigma$, where S_1 is a plane surface adjacent to the screen, Σ is that portion of S_1 in the aperture and S_2 is a large spherical surface, of radius R, centered on P_0, see Figure 3. The first question to address is how to calculate the contribution from S_2 or better yet how do we show we can ignore contributions from S_2. As R increases, Ψ and φ will decrease as $1/R$. However, the area of integration increases as R^2 so the $1/R$ fall-off is not a sufficient reason for neglecting the contribution of the integration over S_2.

On the surface S_2, the Green's function and its derivative are

$$\Psi=\frac{\mathrm{e}^{-\mathrm{i}kR}}{R} \quad [16]$$

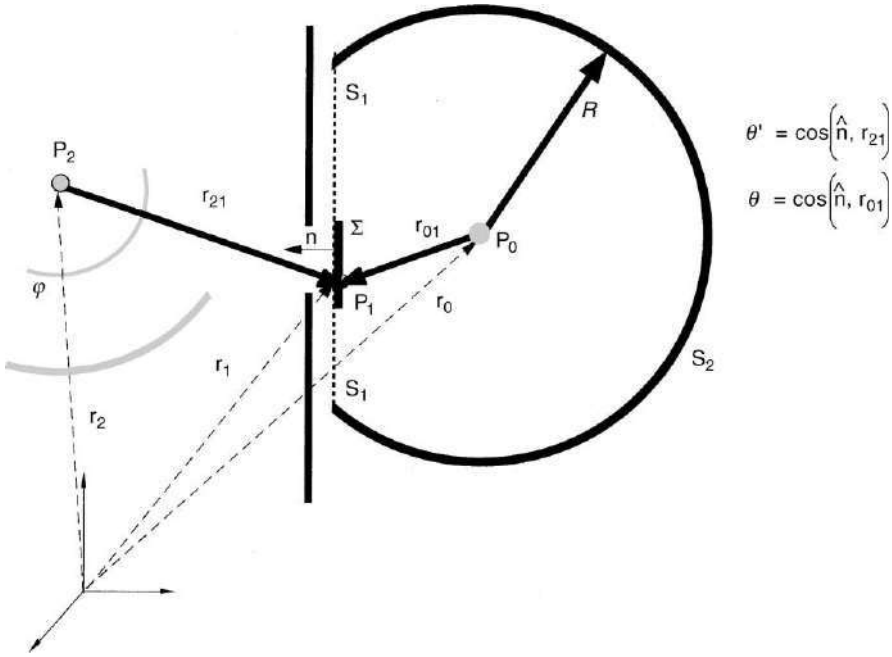

Figure 3 Integration surface for diffraction calculation. P_2 is the source and P_0 is the point where the field is to be calculated. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

$$\frac{\partial \Psi}{\partial n} = \left(-\mathrm{i}k - \frac{1}{R}\right)\frac{\mathrm{e}^{-\mathrm{i}kR}}{R} \approx -\mathrm{i}k\Psi \qquad [17]$$

where the approximate equality is for large R. Thus, the integral over S_2 for large R is

$$\iint_{S_2}\left[\Psi\frac{\partial\varphi}{\partial n} + \varphi(\mathrm{i}k\Psi)\right]\mathrm{d}s$$

$$= \iint_{S_2}\Psi\left(\frac{\partial\varphi}{\partial n} + \mathrm{i}k\varphi\right)R^2 \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \qquad [18]$$

The quantity $R\Psi = \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}}$ is finite as $R \to \infty$ so for the integral to vanish, we must have

$$\lim_{R\to\infty} R\left(\frac{\partial\varphi}{\partial n} + \mathrm{i}k\varphi\right) = 0 \qquad [19]$$

This requirement is called the Sommerfeld radiation condition and is satisfied if $\varphi \to 0$ as fast as $1/R$ (for example, a spherical wave). Since the illumination of the screen will be a spherical wave, or at least a linear combination of spherical waves, we should expect the integral over S_2 to make no contribution (it is exactly zero).

Another way to insure that surface S_2 makes no contribution is to assume that the light turns on at time t_0. At a later time, t, when we desire to know the field at P_0, the radius of S_2 is $R > c\,(t - t_0)$ which physically means that the light has not had time to reach S_2. In this situation, there can be no contribution from S_2. This is not a perfect solution because when the wave is of finite duration it can no longer be considered monochromatic.

We should expect the major contribution in the integral over S_1 to come from the portion of the surface in the aperture, Σ. We make the following assumptions about the incident wave, φ (these assumptions are known as St. Venaut's hypothesis or Kirchhoff's boundary conditions)

1. We assume that, in the aperture, φ and $\partial\varphi/\partial n$ have the values they would have if the screen were not in place.
2. On the portion of S_1 not in the aperture, φ and $\partial\varphi/\partial n$ are identically zero.

The Kirchhoff boundary conditions allow the screen to be neglected, reducing the problem to an integration only over the aperture. It is surprising that the assumptions about the contribution of the surface S_2 and the screen yield very accurate results (as long as polarization is not important, the aperture is large with respect to the wavelength, and we don't get too close to the aperture).

Mathematically the Kirchhoff boundary conditions are incorrect. The two Kirchhoff boundary conditions imply that the field is zero everywhere behind the screen, except in the aperture, which makes Ψ and $\partial\Psi/\partial n$ discontinuous on the boundary of the aperture. Another problem with the boundary conditions is

that Ψ and $\partial\Psi/\partial n$ are known only if the problem has already been solved.

The Sommerfeld Green's functions will remove the inconsistencies of the Kirchhoff theory. If we use the Green's function [3a] then Ψ vanishes in the aperture while $\partial\Psi/\partial n$ can assume the value required by Kirchhoff's boundary conditions. If we use the Green's functions [3b] then the converse holds, i.e. $\partial\Psi/\partial n$ is zero in the aperture. This improved Green's function makes the problem harder with little gain in accuracy so we will retain the Kirchhoff formalism.

As a result of the above discussion, the surface integral [15] is only performed over the surface, Σ, in the aperture. The evaluation of the Green's function on this surface can be simplified by noting that normally $r_{01} \gg \lambda$, i.e., $k \gg 1/r_{01}$, thus on Σ,

$$\begin{aligned}\frac{\partial\Psi}{\partial n} &= \cos(\mathbf{n}, \mathbf{r}_{01})\left(-\mathrm{i}k - \frac{1}{r_{01}}\right)\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}} \\ &\approx -\mathrm{i}k\cos(\hat{\mathbf{n}}, \mathbf{r}_{01})\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}\end{aligned} \qquad [20]$$

Substituting this approximate evaluation of the derivative into [15] yields

$$\varphi(\mathbf{r}_0) = \frac{1}{4\pi}\iint_{\Sigma}\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}\left[\frac{\partial\varphi}{\partial n} + \mathrm{i}k\varphi\cos(\hat{\mathbf{n}}, \mathbf{r}_{01})\right]\mathrm{d}s \qquad [21]$$

The source of the incident wave is a point source located at P_2, with a position vector, $\mathbf{r}_2$, measured with respect to the coordinate system and a distance $|\mathbf{r}_{21}|$ away from P_1, a point in the aperture (see **Figure 3**). The incident wave is therefore a spherical wave of the form

$$\varphi(\mathbf{r}_{21}) = A\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{21}}}{r_{21}} \qquad [22]$$

which fills the aperture. Here also we will assume that $r_{21} \gg \lambda$ so that the derivative of the incident wave assumes the same form as [20]. Then [21] can be written

$$\begin{aligned}\varphi(\mathbf{r}_0) = \frac{\mathrm{i}A}{\lambda}\iint_{\Sigma}&\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot(\mathbf{r}_{21}+\mathbf{r}_{01})}}{r_{21}r_{01}} \\ &\times\left[\frac{\cos(\hat{\mathbf{n}}, \mathbf{r}_{01}) - \cos(\hat{\mathbf{n}}, \mathbf{r}_{21})}{2}\right]\mathrm{d}s\end{aligned} \qquad [23]$$

This relationship is called the Fresnel–Kirchhoff diffraction formula. It is symmetric with respect to $\mathbf{r}_{01}$ and $\mathbf{r}_{21}$, making the problem identical when the source and measurement point are interchanged.

A physical understanding of [23] can be obtained if it is rewritten as

$$\varphi(\mathbf{r}_0) = \iint_{\Sigma}\Phi(\mathbf{r}_{21})\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}}\mathrm{d}s \qquad [24]$$

The field at P_0 is due to the sum of an infinite number of secondary Huygens sources in the aperture Σ. The secondary sources are point sources radiating spherical waves of the form

$$\Phi(\mathbf{r}_{21})\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{01}}}{r_{01}} \qquad [25]$$

with amplitude $\Phi(\mathrm{r}_{21})$, defined by

$$\Phi(\mathbf{r}_{21}) = \frac{\mathrm{i}}{\lambda}\left[A\frac{\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_{21}}}{r_{21}}\right]\left[\frac{\cos(\hat{\mathbf{n}}, \mathbf{r}_{01}) - \cos(\hat{\mathbf{n}}, \mathbf{r}_{21})}{2}\right] \qquad [26]$$

The imaginary constant, i, causes the wavelets from each of these secondary sources to be phase shifted with respect to the incident wave. The obliquity factor, in the amplitude [26]

$$\tfrac{1}{2}[\cos(\hat{\mathbf{n}}, \mathbf{r}_{01}) - \cos(\hat{\mathbf{n}}, \mathbf{r}_{21})] \qquad [27]$$

causes the secondary sources to have a forward directed radiation pattern.

If we had used the Sommerfeld Green's function, the only modification to our result would be a change in the obliquity factor to $\cos(\mathbf{n},\mathbf{r}_{01})$. In our discussions below we will assume that the angles are all small, allowing the obliquity factor to be replaced by 1 so that in the remaining discussion the choice of Green's function has no impact.

We will assume the source of light is at infinity, $z_1 = \infty$ in **Figure 3**, so that the aperture is illuminated by a plane wave, traveling parallel to the z-axis. With this assumption

$$\theta' = \cos(\mathrm{n}, \mathbf{r}_{21}) \approx -1 \qquad [28]$$

We will also make the paraxial approximation which assumes that the viewing position is close to the z-axis, leading to

$$\theta = \cos(\mathbf{n}, \mathbf{r}_{01}) \approx 1 \qquad [29]$$

With these assumptions eqn [24] becomes identical to the Huygens–Fresnel integral discussed in the section on Fresnel diffraction:

$$\tilde{E}(\mathbf{r}_0) = \frac{\mathrm{i}}{\lambda}\iint_{\Sigma}\frac{\tilde{E}_i(\mathbf{r})}{R}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\,\mathrm{d}s \qquad [30]$$

The modern interpretation of the Fresnel–Kirchhoff (or now in the small-angle approximation, the Fresnel–Huygens) integral is to view it as a convolution integral. By considering free space to be a linear system, the result of propagation can be calculated by convolving the incident (input) wave with the impulse response of free space (our Green's function),

$$\frac{i}{\lambda}\frac{e^{-i\mathbf{k}\cdot\mathbf{R}}}{R} \qquad [31]$$

The job of calculating diffraction has only begun with the derivation of the Fresnel–Kirchhoff integral. In general, an analytic expression for the integral cannot be found because of the difficulty of performing the integration over R. There are two approximations that will allow us to obtain analytic expressions of the Fresnel–Kirchhoff integral. In both approximations, all dimensions are assumed to be large with respect to the wavelength. In one approximation, the viewing position is assumed to be far from the obstruction; the resulting diffraction is called Fraunhofer diffraction and will be discussed here. The second approximation, which leads to Fresnel diffraction, assumes that the observation point is nearer the obstruction, to be quantified below. This approximation, which is the more difficult mathematically, is discussed in the article on **Diffraction:** Fresnel Diffraction.

Fraunhofer Approximation

In Fraunhofer diffraction, we require that the source of light and the observation point, P_0, be far from the aperture so that the incident and diffracted wave can be approximated by plane waves. A consequence of this requirement is that the entire waveform passing through the aperture contributes to the observed diffraction.

The geometry to be used in this derivation is shown in Figure 4. The wave incident normal to the aperture, Σ, is a plane wave and the objective of the calculation is to find the departure of the transmitted wave from its geometrical optical path. The calculation will provide the light distribution, transmitted by the aperture, as a function of the angle the light is deflected from the incident direction. We assume that diffraction makes only a small perturbation on the predictions of geometrical optics. The deflection angles encountered in this derivation are, therefore, small, as assumed above, and we will be able to use the paraxial approximation.

The distance from a point P, in the aperture, to the observation point P_0, of Figure 4, is

$$R^2 = (x-\xi)^2 + (y-\eta)^2 + Z^2 \qquad [32]$$

From Figure 4 we see R_0 is the distance from the center of the screen to the observation point, P_0,

$$R_0^2 = \xi^2 + \eta^2 + Z^2 \qquad [33]$$

The difference between these two vectors is

$$\begin{aligned} R_0^2 - R^2 &= (R_0 - R)(R_0 + R) \\ &= \xi^2 + \eta^2 + Z^2 - Z^2 - (x^2 - 2x\xi + \xi^2) \\ &\quad - (y^2 - 2y\eta + \eta^2) \\ &= 2(x\xi + y\eta) - (x^2 + y^2) \end{aligned} \qquad [34]$$

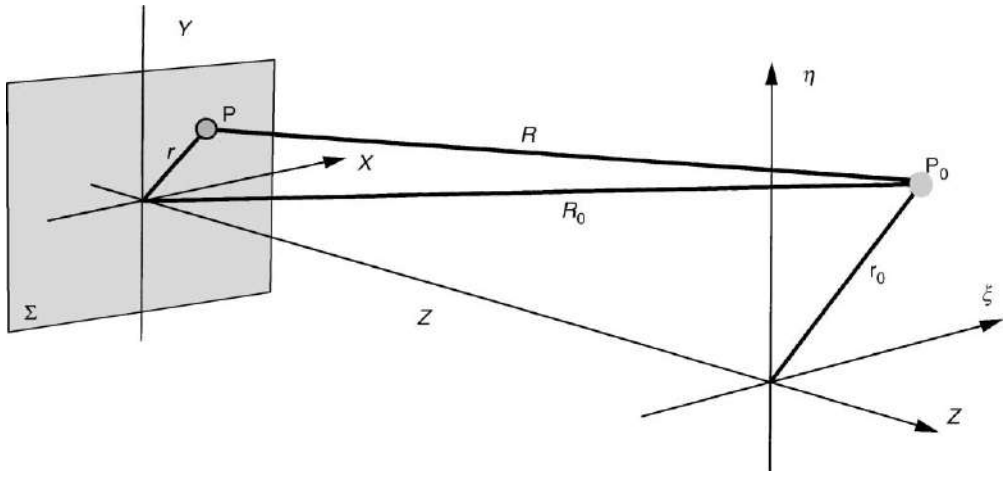

Figure 4 Geometry for Fraunhofer diffraction. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

The equation for the position of point P in the aperture can be written in terms of [34]:

$$r = R_0 - R = \frac{R_0^2 - R^2}{R_0 + R},$$
$$= [2(x\xi + y\eta) - (x^2 + y^2)]\frac{1}{R_0 + R} \qquad [35]$$

The reciprocal of $(R_0 + R)$ can be written as

$$\frac{1}{R_0 + R} = \frac{1}{2R_0 + R - R_0} = \frac{1}{2R_0}\left(1 + \frac{R - R_0}{2R_0}\right)^{-1} \qquad [36]$$

Now using [36]

$$\begin{aligned} R_0 - R &= \left[\frac{x\xi + y\eta}{R_0} - \frac{x^2 + y^2}{2R_0}\right]\left[1 - \frac{R_0 - R}{2R_0}\right]^{-1} \\ &= \left[\frac{x\xi + y\eta}{R_0} - \frac{x^2 + y^2}{2R_0}\right]\left[1 - \frac{r}{2R_0}\right]^{-1} \end{aligned} \qquad [37]$$

If the diffraction integral [30] is to have a finite (nonzero) value, then

$$k|R_0 - R| \ll kR_0 \qquad [38]$$

This requirement insures that all the Huygens's wavelets, produced over the aperture from the center out to the position r, will have similar phases and will interfere to produce a nonzero amplitude at P_0. This requirement results in

$$\frac{1}{1 - \dfrac{r}{2R_0}} \approx 1 \qquad [39]$$

From this equation we are tempted to state that the aperture dimension must be small relative to the observation distance. However, for the exponent to provide a finite contribution in the integral, it is kr that is small, not the aperture dimension, r.

Using the approximation for $(R_0 - R)$, we obtain the diffraction integral

$$\tilde{E}_P = \frac{\mathrm{i}\alpha\, \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_0}}{\lambda R_0} \iint_{\Sigma} f(x, y)\mathrm{e}^{+\mathrm{i}k\left(\frac{x\xi + y\eta}{R_0} - \frac{x^2 + y^2}{2R_0}\right)}\, \mathrm{d}x\, \mathrm{d}y \qquad [40]$$

The change in the amplitude of the wave due to the change in R, as we move across the aperture, is neglected, allowing R in the denominator of the Huygens–Fresnel integral to be replaced by R_0 and moved outside of the integral. We have introduced the complex transmission function, $f(x, y)$, of the aperture to allow a very general aperture to be treated. If the aperture function described the variation in absorption of the aperture as a function of position, as would be produced by a photographic negative, then $f(x, y)$ would be a real function. If the aperture function described the variation in transmission of a biological sample, it might be entirely imaginary.

The argument of the exponent in [40] is

$$\mathrm{i}k\left(\frac{x\xi + y\eta}{R_0} - \frac{x^2 + y^2}{2R_0}\right) = \mathrm{i}2\pi\left(\frac{x\xi + y\eta}{\lambda R_0} - \frac{x^2 + y^2}{2\lambda R_0}\right) \qquad [41]$$

If the observation point, P_0, is far from the screen, we can neglect the second term and treat the phase variation across the aperture as a linear function of position. This is equivalent to assuming that the diffracted wave is a collection of plane waves. Mathematically, the second term in [41] can be neglected if

$$\frac{x^2 + y^2}{2\lambda R_0} \ll 1 \qquad [42]$$

This is called the far-field approximation and the theory yields Fraunhofer diffraction. If the quadratic term of [41] is on the order of 2π then the fraction is

$$\frac{x^2 + y^2}{2\lambda R_0} \approx \Theta(1) \qquad [43]$$

and we must retain the quadratic term and the theory yields Fresnel diffraction.

We define spatial frequencies in the x and y direction by

$$\begin{aligned} \omega_x &= \frac{2\pi}{\lambda}\sin\theta_x = -\frac{2\pi\xi}{\lambda R_0} \\ \omega_y &= \frac{2\pi}{\lambda}\sin\theta_y = -\frac{2\pi\eta}{\lambda R_0} \end{aligned} \qquad [44]$$

We define the spatial frequencies with a negative sign to allow equations involving spatial frequencies to have the same form as those involving temporal frequencies. The negative sign is required because ωt and $\mathbf{k}\cdot\mathbf{r}$ appear in the phase of the wave with opposite signs and we want the Fourier transform in space and in time to have the same form.

With the variables defined in [44], the integral becomes

$$\tilde{E}_P(\omega_x, \omega_y) = \frac{\mathrm{i}\alpha\, \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_0}}{\lambda R_0} \iint_{\Sigma} f(x, y)\mathrm{e}^{-\mathrm{i}(\omega_x x + \omega_y y)}\, \mathrm{d}x\, \mathrm{d}y \qquad [45]$$

The result of our derivation is that the Fraunhofer diffraction field, $\tilde{E}_P$, can be calculated by performing the two-dimensional Fourier transform of the aperture's transmission function.

Definition of Fourier Transform

In this discussion the function $F(\omega)$ is defined as the *Fourier transform* of $f(t)$:

$$\mathcal{F}\{f(t)\} \equiv F(\omega) \equiv \int_{-\infty}^{\infty} f(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau \qquad [46]$$

The transformation from a temporal to a frequency representation given by [46] does not destroy information; thus, the inverse transform can also be defined

$$\mathcal{F}^{-1}\{F(\omega)\} = f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}\omega \qquad [47]$$

By assuming that the illumination of an aperture is a plane wave and that the observation position is in the far field, the diffraction of an aperture is found to be given by the Fourier transform of the function describing the amplitude transmission of the aperture. The amplitude transmission of the aperture, $f(x, y)$, may thus be interpreted as the superposition of mutually coherent plane waves.

Diffraction by a Rectangular Aperture

We will now use Fourier transform theory to calculate the Fraunhofer diffraction pattern from a rectangular slit and point out the reciprocal relationship between the size of the diffraction pattern and the size of the aperture.

Consider a rectangular aperture with a transmission function given by

$$f(x, y) = \begin{cases} 1 & |x| \le x_0, |y| \le y_0 \\ 0 & \text{all other } x \text{ and } y \end{cases} \qquad [48]$$

Because the aperture is two-dimensional, we need to apply a two-dimensional Fourier transform but it is not difficult because the amplitude transmission function is separable, in x and y. The diffraction amplitude distribution from the rectangular slit is simply one-dimensional transforms carried out individually for the x and y dimensions:

$$\tilde{E}_P = \frac{\mathrm{i}\alpha}{\lambda R_0}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_0}\int_{-x_0}^{x_0} f(x)\mathrm{e}^{-\mathrm{i}\omega_x x}\,\mathrm{d}x\int_{-y_0}^{y_0} f(y)\mathrm{e}^{-\mathrm{i}\omega_y y}\,\mathrm{d}y \qquad [49]$$

Since both $f(x)$ and $f(y)$ are defined as symmetric functions, we need only calculate the cosine transforms to obtain the diffracted field:

To calculate the Fourier transform defined by [46] we can rewrite the transform as

$$F(\omega) = \int_{-\infty}^{\infty} f(\tau)\cos\,\omega\tau\,\mathrm{d}\tau - \mathrm{i}\int_{-\infty}^{\infty} f(\tau)\sin\,\omega\tau\,\mathrm{d}\tau \qquad [50]$$

If $f(\tau)$ is a real, even function then the Fourier transform can be obtained by simply calculating the cosine transform:

$$\int_{-\infty}^{\infty} f(\tau)\cos\,\omega\tau\,\mathrm{d}\tau \qquad [51]$$

$$\tilde{E}_P = \mathrm{i}\frac{4x_0y_0\alpha}{\lambda R_0}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_0}\frac{\sin\,\omega_x x_0}{\omega_x x_0}\frac{\sin\,\omega_y y_0}{\omega_y y_0} \qquad [52]$$

The intensity distribution of the Fraunhofer diffraction produced by the rectangular aperture is

$$I_P = I_0\frac{\sin^2\omega_x x_0}{(\omega_x x_0)^2}\frac{\sin^2\omega_y y_0}{(\omega_y y_0)^2} \qquad [53]$$

The maximum intensities in the x and y directions occur at

$$\omega_x x_0 = \omega_y y_0 = 0 \qquad [54]$$

The area of this rectangular aperture is defined as $A = 4x_0y_0$, resulting in an expression for the maximum intensity of

$$I_0 = \left[\frac{\mathrm{i}4x_0y_0\alpha\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_0}}{\lambda R_0}\right]^2 = \frac{A^2\alpha^2}{\lambda^2 R_0^2} \qquad [55]$$

The minima of [53] occur when $\omega_x x_0 = n\pi$ or when $\omega_y y_0 = m\pi$. The location of the zeros can be specified as a dimension in the observation plane or, using the paraxial approximation, in terms of the angle defined by [44]

$$\begin{aligned} \sin\,\theta_x \approx \theta_x \approx \frac{\xi}{R_0} = \frac{n\lambda}{2x_0} \\ \sin\,\theta_y \approx \theta_y \approx \frac{\eta}{R_0} = \frac{m\lambda}{2y_0} \end{aligned} \qquad [56]$$

The dimensions of the diffraction pattern are characterized by the location of the first zero, i.e., when $n = m = 1$, and are given by the observation plane coordinates, ξ and η. The dimensions of the diffraction pattern are inversely proportional to the dimensions of the aperture. As the aperture dimension expands, the width of the diffraction pattern decreases until, in the limit of an infinitely wide aperture, the diffraction pattern becomes a delta function.

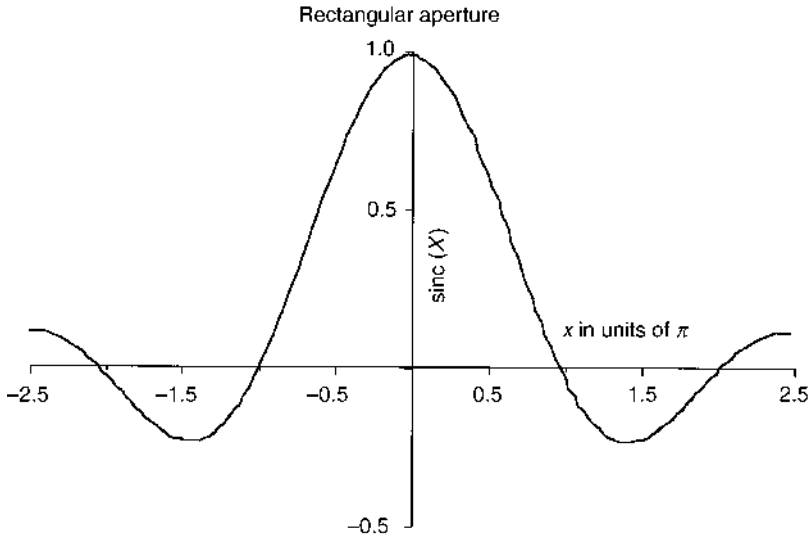

Figure 5 The amplitude of the wave diffracted by a rectangular slit. This sinc function describes the light wave's amplitude that would exist in both the *x* and *y* directions. The coordinate would have to be scaled by the dimension of the slit. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

Figure 5 is a theoretical plot of the amplitude of the diffraction wave from a rectangular slit.

Diffraction from a Circular Aperture

To obtain the diffraction pattern from a circular aperture we first convert from the rectangular coordinate system we have used to a cylindrical coordinate system. The new cylindrical geometry is shown in Figure 6. At the aperture plane

$$\begin{aligned} x &= s\cdot\cos\varphi & y &= s\cdot\sin\varphi \\ f(x,y) &= f(s,\varphi) & \mathrm{d}x\,\mathrm{d}y &= s\,\mathrm{d}s\,\mathrm{d}\varphi \end{aligned} \qquad [57]$$

At the observation plane

$$\xi = \rho\cdot\cos\theta \qquad \eta = \rho\cdot\sin\theta \qquad [58]$$

In the new, cylindrical, coordinate system at the observation plane, the spatial frequencies are written as

$$\begin{aligned} \omega_x &= -\frac{k\xi}{R_0} = -\frac{k\rho}{R_0}\cos\theta \\ \omega_y &= -\frac{k\eta}{R_0} = -\frac{k\rho}{R_0}\sin\theta \end{aligned} \qquad [59]$$

From Figure 6 we see that the observation point, P, can be defined in terms of the angle ψ, where

$$\sin\psi = \frac{\rho}{R_0} \qquad [60]$$

This allows an angular representation of the size of the diffraction pattern if it is desired.

Using [57] and [59], we may write

$$\begin{aligned} \omega_x x + \omega_y y &= -\frac{ks\rho}{R_0}(\cos\theta\cos\varphi + \sin\theta\sin\varphi), \\ &= -\frac{ks\rho}{R_0}\cos(\theta-\varphi) \end{aligned} \qquad [61]$$

The Huygens–Fresnel integral for Fraunhofer diffraction can now be written in terms of cylindrical coordinates:

$$\tilde{E}_P = \frac{\mathrm{i}\alpha}{\lambda R_0}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_0}\int_0^{\frac{a}{2}}\int_0^{2\pi} f(s,\varphi)\mathrm{e}^{-\mathrm{i}k\frac{s\rho}{R_0}\cos(\theta-\varphi)}\,s\,\mathrm{d}s\,\mathrm{d}\varphi \qquad [62]$$

We will use this equation to calculate the diffraction amplitude from a clear aperture of diameter a, defined by the equation

$$f(s,\varphi) = \begin{cases} 1 & s \le \frac{a}{2},\ \text{all}\ \varphi \\ 0 & s > \frac{a}{2} \end{cases} \qquad [63]$$

The symmetry of this problem is such that

$$f(s,\varphi) = f(s)g(\varphi) = f(s) \qquad [64]$$

$$\mathcal{F}\{f(s,\phi)\} = \mathcal{F}\{f(s)\} = F(\rho) \qquad [65]$$

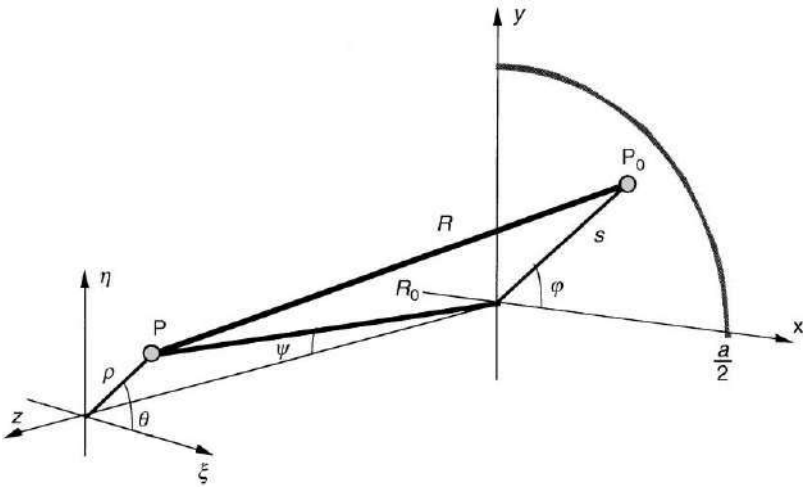

Figure 6 Geometry for calculation of Fraunhofer diffraction from a circular aperture. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

$$F(\rho, \theta) = \int_0^\infty f(s)s\,\mathrm{d}s \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}\rho \cdot s\,\cos(\varphi-\theta)}\,\mathrm{d}\varphi \qquad [66]$$

$$F(\rho) = \int_0^\infty f(s)\,s\,\mathrm{d}s \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}\rho \cdot s\,\cos\varphi}\,\mathrm{d}\varphi \qquad [67]$$

The second integral in [67] belongs to a class of functions called the Bessel function defined by the integral

$$\mathrm{J}_n(s\rho) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{\mathrm{i}[s\rho\,\sin\,\varphi - n\varphi]}\mathrm{d}\varphi \qquad [68]$$

In [67] the integral corresponds to the $n = 0$, zero-order, Bessel function. Using this definition, we can rewrite [67] as

$$F(\rho) = \int_0^\infty f(s)\mathrm{J}_0(s\rho)s\,\mathrm{d}s \qquad [69]$$

This transform is called the Fourier–Bessel transform or the Hankel zero-order transform.

Using these definitions, the transform of the aperture function, $f(s)$, is

$$F(\rho) = \int_0^{\frac{a}{2}} \mathrm{J}_0(s\rho)s\,\mathrm{d}s \qquad [70]$$

We use the identity

$$x\mathrm{J}_1(x) = \int_0^x \chi\mathrm{J}_0(\chi)\,\mathrm{d}\chi \qquad [71]$$

to obtain

$$F(\rho) = \frac{a}{2\rho}\mathrm{J}_1\left(\frac{\rho a}{2}\right) \qquad [72]$$

We can now write the amplitude of the diffracted wave as

$$\tilde{E}_P = \frac{\mathrm{i}\alpha}{\lambda}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_0}\left[\frac{\pi a}{k\rho}\mathrm{J}_1\left(\frac{ka\rho}{2R_0}\right)\right] \qquad [73]$$

A plot of the function contained within the bracket of [73] is given in Figure 7. If we define

$$u = \frac{ka\rho}{2R_0} \qquad [74]$$

then the spatial distribution of intensity in the diffraction pattern can be written in a form known as the Airy formula,

$$I = I_0\left[\frac{2\mathrm{J}_1(u)}{u}\right]^2 \qquad [75]$$

where we have defined

$$I_0 = \left(\frac{\alpha A}{\lambda R_0}\right)^2 \qquad [76]$$

with A representing the area of the aperture, $A = \pi(\frac{a}{2})^2$.

The intensity pattern described by [75] is called the Airy pattern. The intensity at $u = 0$ is the same as was obtained for a rectangular aperture of the same area, because, in the limit,

$$\lim_{u\to 0}\left[\frac{2\mathrm{J}_1(u)}{u}\right] = 1 \qquad [77]$$

For the Airy pattern, 84% of the total area is contained in the disk between the first zeros of [75]. Those zeros occur at $u = \pm 1.22\pi$, which corresponds

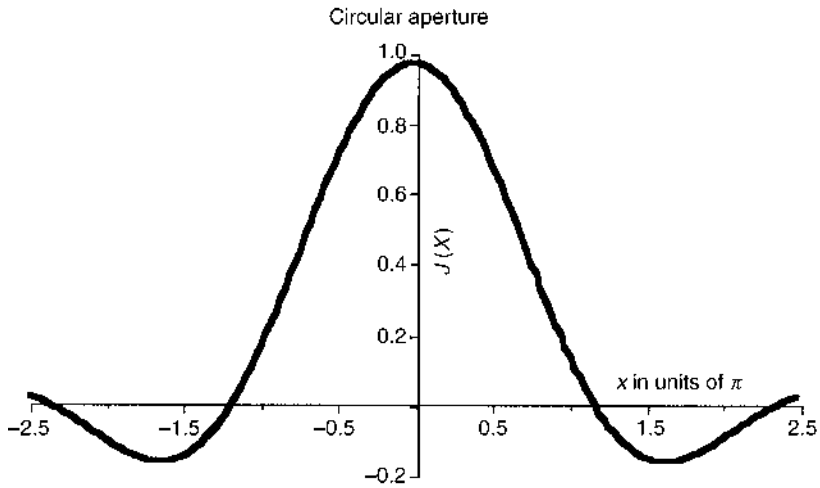

Figure 7 Amplitude of a wave diffracted by a circular aperture. The observed light distribution is constructed by rotating this Bessel function around the optical axis. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

to a radius in the observation plane of

$$\rho = \frac{1.22(\lambda R_0)}{a} \qquad [78]$$

91% of the light intensity is contained within the circle bounded by the second minimum at $u = 2.233\pi$. The intensities in the secondary maxima of the diffraction pattern of a rectangular aperture [53] are much larger than the intensities in the secondary maxima of the Airy pattern of a circular aperture of equal area. The peak intensities, relative to the central maximum, of the first three secondary maxima of a rectangular aperture are 4.7%, 1.6%, and 0.8%, respectively. For a circular aperture, the same quantities are 1.7%, 0.04%, and 0.02%, respectively.

Array Theorem

There is an elegant mathematical technique for handling diffraction from multiple apertures, called the array theorem. The theorem is based on the fact that Fraunhofer diffraction is given by the Fourier transform of the aperture function and utilizes the convolution integral.

The convolution of the functions $a(t)$ and $b(t)$ is defined as

$$g(t) = a(t) \otimes b(t) = \int_{-\infty}^{\infty} a(t)b(\tau - t)\,\mathrm{d}t \qquad [79]$$

The Fourier transform of a convolution of two functions is the product of the Fourier transforms of

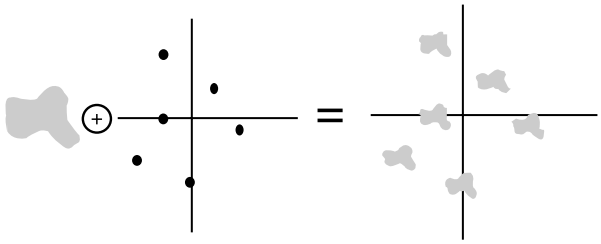

Figure 8 The convolution of an aperture with an array of delta functions will produce an array of identical apertures, each located at the position of one of the delta functions. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

the individual functions:

$$\mathcal{F}\{a(t) \otimes b(t)\} = \mathcal{F}\left\{\int_{-\infty}^{\infty} a(t)b(\tau - t)\,\mathrm{d}t\right\} = A(\omega)B(\omega) \qquad [80]$$

We will demonstrate the array theorem for one dimension, where the functions represent slit apertures. The results can be extended to two dimensions in a straightforward way.

Assume that we have a collection of identical apertures, shown on the right of Figure 8. If one of the apertures is located at the origin of the aperture plane, its transmission function is $\psi(x)$. The transmission function of an aperture located at a point, x_n, can be written in terms of a generalized aperture function, $\psi(x - \alpha)$, by the use of the sifting property of the delta function

$$\psi(x - x_n) = \int \psi(x - \alpha)\delta(\alpha - x_n)\,\mathrm{d}\alpha \qquad [81]$$

The aperture transmission function representing an array of apertures will be the sum of the

distributions of the individual apertures, represented graphically in Figure 8 and mathematically by the summation

$$\Psi(x) = \sum_{n=1}^{N} \psi(x - x_n) \qquad [82]$$

The Fraunhofer diffraction from this array is $\Phi(\omega_x)$, the Fourier transform of $\Psi(x)$,

$$\Phi(\omega_x) = \int_{-\infty}^{\infty} \Psi(x) e^{-i\omega_x x}\, dx \qquad [83]$$

which can be rewritten as

$$\Phi(\omega_x) = \sum_{n=1}^{N} \int_{-\infty}^{\infty} \psi(x - x_n) e^{-i\omega_x x}\, dx \qquad [84]$$

We now make use of the fact that $\psi(x - x_n)$ can be expressed in terms of a convolution integral. The Fourier transform of $\psi(x - x_n)$ is, from the convolution theorem [80], the product of the Fourier transforms of the individual functions that make up the convolution:

$$\begin{aligned}\Phi(\omega_x) &= \sum_{n=1}^{N} \mathcal{F}\{\psi(x - \alpha)\}\mathcal{F}\{\delta(\alpha - x_n)\} \\ &= \mathcal{F}\{\psi(x - \alpha)\} \sum_{n=1}^{N} \mathcal{F}\{\delta(\alpha - x_n)\} \\ &= \mathcal{F}\{\psi(x - \alpha)\}\mathcal{F}\left\{\sum_{n=1}^{N} \delta(\alpha - x_n)\right\} \qquad [85]\end{aligned}$$

The first transform in [85] is the diffraction pattern of the generalized aperture function and the second transform is the diffraction pattern produced by a set of point sources with the same spatial distribution as the array of identical apertures. We will call this second transform the array function.

To summarize, the array theorem states that the diffraction pattern of an array of similar apertures is given by the product of the diffraction pattern from a single aperture and the diffraction (or interference) pattern of an identically distributed array of point sources.

N Rectangular Slits

An array of N identical apertures is called a diffraction grating in optics. The Fraunhofer diffraction patterns produced by such an array have two important properties; a number of very narrow beams are produced by the array and the beam positions are a function of the wavelength of illumination of the apertures and the relative phase of the waves radiated by each of the apertures.

Because of these properties, arrays of diffracting apertures have been used in a number of applications.

- At radio frequencies, arrays of dipole antennas are used to both radiate and receive signals in both radar and radio astronomy systems. One advantage offered by diffracting arrays at radio frequencies is that the beam produced by the array can be electrically steered by adjusting the relative phase of the individual dipoles.
- An optical realization of a two-element array of radiators is Young's two-slit experiment and a realization of a two-element array of receivers is Michelson's stellar interferometer. Two optical implementations of arrays, containing more diffracting elements, are diffraction gratings and holograms. In nature, periodic arrays of diffracting elements are the origin of the colors observed on some invertebrates.
- Many solids are naturally arranged in three-dimensional arrays of atoms or molecules that act as diffraction gratings when illuminated by X-ray wavelengths. The resulting Fraunhofer diffraction patterns are used to analyze the ordered structure of the solids.

The array theorem can be used to calculate the diffraction pattern from N rectangular slits, each of width a and separation d. The aperture function of a single slit is equal to

$$\psi(x, y) = \begin{cases} 1 & |x| \le \frac{a}{2}, |y| \le y_0 \\ 0 & \text{all other } x \text{ and } y \end{cases} \qquad [86]$$

The Fraunhofer diffraction pattern of this aperture has already been calculated and is given by [52]:

$$\mathcal{F}\{\psi(x, y)\} = K \frac{\sin \alpha}{\alpha} \qquad [87]$$

where the constant K and the variable α are

$$\begin{aligned} K &= \frac{i2ay_0\alpha}{\lambda R_0} e^{-i\mathbf{k}\cdot\mathbf{R}_0} \frac{\sin \omega_y y_0}{\omega_y y_0} \\ \alpha &= \frac{ka}{2} \sin \theta_x \end{aligned} \qquad [88]$$

The array function is

$$A(x) = \sum_{n=1}^{N} \delta(x - x_n) \quad \text{where } x_n = (n - 1)d \qquad [89]$$

and its Fourier transform is given by

$$\mathcal{F}\{A(x)\} = \frac{\sin N\beta}{\sin \beta} \qquad [90]$$

$$\beta = \frac{kd}{2}\sin \theta_x$$

The Fraunhofer diffraction pattern's intensity distribution in the x-direction is thus given by

$$I_\theta = I_0 \underbrace{\frac{\sin^2\alpha}{\alpha^2}}_{\text{shape factor}} \underbrace{\frac{\sin^2 N\beta}{\sin^2\beta}}_{\text{grating factor}} \qquad [91]$$

We have combined the variation in intensity in the y-direction into the constant I_0 because we assume that the intensity variation in the x-direction will be measured at a constant value of y.

A physical interpretation of [91] views the first factor as arising from the diffraction associated with a single slit; it is called the shape factor. The second factor arises from the interference between light from different slits; it is called the grating factor. The fine detail in the spatial light distribution of the diffraction pattern is described by the grating factor and arises from the coarse detail in the diffracting object. The coarse, overall light distribution in the diffraction pattern is described by the shape factor and arises from the fine detail in the diffracting object.

Young's Double Slit

The array theorem makes the analysis of Young's two-slit experiment a trivial exercise. This application of the array theorem will demonstrate that the interference between two slits arises naturally from an application of diffraction theory. The result of this analysis will support a previous assertion that interference describes the same physical process as diffraction and the division of the two subjects is an arbitrary one.

The intensity of the diffraction pattern from two slits is obtained from [91] by setting $N = 2$:

$$I_\theta = I_0 \frac{\sin^2\alpha}{\alpha^2}\cos^2\beta \qquad [92]$$

The sinc function describes the energy distribution of the overall diffraction pattern, while the cosine function describes the energy distribution created by interference between the light waves from the two slits. Physically, α is a measure of the phase difference between points in one slit and β is a measure of the phase difference between similar points in the two slits. Zeros in the diffraction intensity occur whenever $\alpha = n\pi$ or whenever $\beta = \frac{1}{2}(2n+1)\pi$. Figure 9 shows the interference maxima, from the grating factor, under the central diffraction peak, described by the shape factor. The number of interference maxima contained under the central maximum is given by

$$\frac{2d}{a} - 1 \qquad [93]$$

In Figure 9 three different slit spacings are shown with the ratio d/a equal to 3, 6, and 9, respectively.

The Diffraction Grating

In this section we will use the array theorem to derive the diffraction intensity distribution of a large number of identical apertures. We will discover that the positions of the principal maxima are a function of the illuminating wavelength. This functional relationship has led to the application of a diffraction grating to wavelength measurements.

The diffraction grating normally used for wavelength measurements is not a large number of diffracting apertures but rather a large number of reflecting grooves cut in a surface such as gold or aluminum. The theory to be derived also applies to these reflection gratings but a modification must be made to the theory. The shape of the grooves in the reflection grating can be used to control the fraction of light diffracted into a principal maximum and we will examine how to take this into account. A grating whose groove shape has been controlled to enhance the energy contained in a particular principal maximum is called a blazed grating. The use of special groove shapes is equivalent to the modification of the phase of individual elements of an antenna array at radio frequencies.

The construction of a diffraction grating for use in an optical device for measuring wavelength was first suggested by David Rittenhouse, an American astronomer, in 1785, but the idea was ignored until Fraunhofer reinvented the concept in 1819. Fraunhofer's first gratings were fine wires spaced by wrapping the wires in the threads of two parallel screws. He later made gratings by cutting (*ruling*) grooves in gold films deposited on the surface of glass. H.A. Rowland made a number of well designed ruling machines, which made possible the production of large-area gratings. Following a suggestion by Lord Rayleigh, Robert Williams Wood (1868–1955) developed the capability to control the shape of the individual grooves.

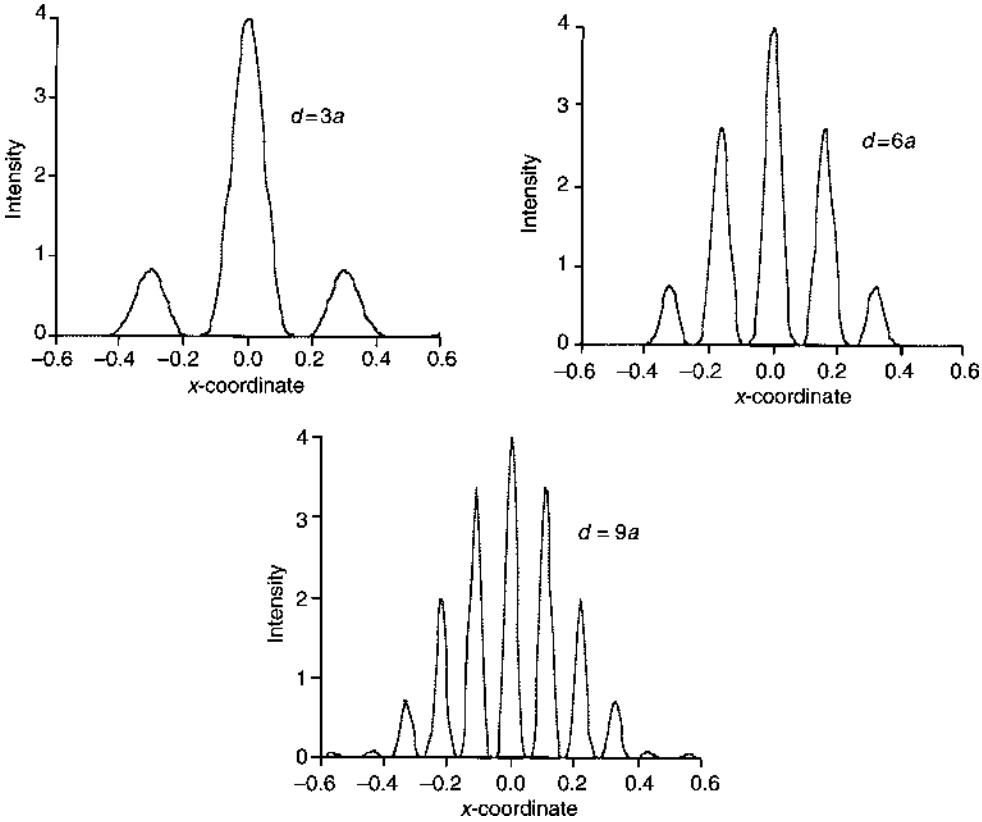

Figure 9 The number of interference fringes beneath the main diffraction peak of a Young's two-slit experiment with rectangular apertures. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

If N is allowed to assume values much larger than 2, the appearance of the interference fringes, predicted by the grating factor, changes from a simple sinusoidal variation to a set of narrow maxima, called principal maxima, surrounded by much smaller, secondary maxima. To calculate the diffraction pattern we must evaluate [91] when N is a large number.

Whenever $N\beta = m\pi$, $m = 0, 1, 2, \ldots$, the numerator of the second factor in [91] will be zero, leading to an intensity that is zero, $I_\theta = 0$. The denominator of the second factor in [91] is zero when $\beta = l\pi$, $l = 0, 1, 2, \ldots$ If both conditions occur at the same time, the ratio of m and N is equal to an integer, $m/N = l$, and instead of zero intensity, we have an indeterminate value for the intensity, $I_\theta = 0/0$. To find the actual value for the intensity, we must apply L'Hospital's rule

$$\lim_{\beta \to l\pi} \frac{\sin N\beta}{\sin \beta} = \lim_{\beta \to l\pi} \frac{N \cos N\beta}{\cos \beta} = N \qquad [94]$$

L'Hospital's rule predicts that whenever

$$\beta = l\pi = \frac{m}{N}\pi \qquad [95]$$

where l is an integer, a principal maximum in the intensity will occur with a value given by

$$I_{\theta P} = N^2 I_0 \frac{\sin^2 \alpha}{\alpha^2} \qquad [96]$$

Secondary maxima, much weaker than the principal maxima, occur when

$$\beta = \left(\frac{2m+1}{2N}\right)\pi \qquad m = 1, 2, \ldots \qquad [97]$$

(When $m = 0$, m/N is an integer, thus the first value that m can have in [97] is $m = 1$.) The intensity of each secondary maximum is given by

$$\begin{aligned} I_{\theta S} &= I_0 \frac{\sin^2 \alpha}{\alpha^2} \frac{\sin^2 N\beta}{\sin^2 \beta} \\ &= I_0 \frac{\sin^2 \alpha}{\alpha^2} \left[\frac{\sin^2\left(\frac{2m+1}{2}\right)\pi}{\sin^2\left(\frac{2m+1}{2N}\right)\pi} \right] \\ &= I_0 \frac{\sin^2 \alpha}{\alpha^2} \left[\frac{1}{\sin\left(\frac{2m+1}{2N}\right)\pi} \right]^2 \end{aligned} \qquad [98]$$

The quantity $(2m+1)/2N$ is a small number for large values of N, allowing the small angle approximation to be made:

$$I_{\theta S} \approx I_0 \frac{\sin^2\alpha}{\alpha^2}\left[\frac{2N}{\pi(2m+1)}\right]^2 \qquad [99]$$

The ratio of the intensity of a secondary maximum and a principal maximum is given by

$$\frac{I_{\theta S}}{I_{\theta P}} = \left[\frac{2}{\pi(2m+1)}\right]^2 \qquad [100]$$

The strongest secondary maximum occurs for $m=1$ and, for large N, has an intensity that is about 4.5% of the intensity of the neighboring principal maximum.

The positions of principal maxima occur at angles specified by the grating formula.

$$\beta = \frac{kd\sin\theta}{2} = l\pi = \frac{m}{N}\pi \qquad [101]$$

The angular position of the principal maxima (the diffracted angle θ_d) is given by the Bragg equation

$$\sin\theta_d = \frac{l\lambda}{d} \qquad [102]$$

where l is called the grating order. This simple relationship between the angle and the wavelength can be used to construct a device to measure wavelength, the grating spectrometer.

The model we have used to obtain this result is based on a periodic array of identical apertures. The transmission function of this array is a periodic square wave. If, for the moment, we treat the grating as infinite in size, we discover that the principal maxima in the diffraction pattern correspond to the terms of the Fourier series describing a square wave.

The zero order, $m=0$, corresponds to the temporal average of the periodic square wave and has an intensity proportional to the spatially averaged transmission of the grating. Because of its equivalence to the temporal average of a time-varying signal, the zero-order principal maximum is often called the dc term.

The first-order principal maximum corresponds to the fundamental spatial frequency of the grating and the higher orders correspond to the harmonics of this frequency.

The dc term provides no information about the wavelength of the illumination. Information about the wavelength of the illuminating light can only be obtained by measuring the angular position of the first or higher order principal maxima.

Grating Spectrometer

The curves shown in Figure 10 display the separation of the principal maxima as a function of $\sin\theta_d$. The angular separation of principal maxima can be converted to a linear dimension by assuming a distance, R, from the graing to the observation plane. In the lower right-hand curve of Figure 10, a distance of 2 meters was assumed. Grating spectrometers are classified by the size in meters of R used in their design. The larger R, the easier it is to resolve wavelength differences. For example, a 1 meter spectrometer is a higher-resolution instrument than a 1/4 meter spectrometer.

The fact that the grating is finite in size causes each of the orders to have an angular width that limits the resolution with which the illuminating wavelength can be measured. To calculate the resolving power of the grating, we first determine the angular width of a principal maximum. This is accomplished by measuring the angular change, of the principal maximum's position, when β changes from $\beta = l\pi = m\pi/N$ to $\beta = (m+1)\pi/N$, i.e., $\Delta\beta = \pi/N$. Using the definition of β

$$\beta = \frac{kd\sin\theta_d}{2} \qquad [103]$$

$$\Delta\beta = \frac{\pi d\cos\theta_d\Delta\theta}{\lambda}$$

and the angular width is

$$\Delta\theta = \frac{\lambda}{Nd\cos\theta_d} \qquad [104]$$

The derivative of the grating formula gives

$$\Delta\lambda = \frac{d}{l}\cos\theta_d\Delta\theta \qquad [105]$$

$$\Delta\theta = \frac{l\Delta\lambda}{d\cos\theta_d}$$

Equating this spread in angle to eqn [104] yields

$$\frac{l\Delta\lambda}{d\cos\theta} = \frac{\lambda}{Nd\cos\theta} \qquad [106]$$

The resolving power of a grating is therefore

$$\frac{\lambda}{\Delta\lambda} = Nl \qquad [107]$$

The improvement of resolving power with N can be seen in Figure 10. A grating 2 inches wide and containing 15 000 grooves per inch would have a resolving power in second order ($l=2$) of 6×10^4. At a wavelength of 600 nm this grating could resolve two waves, differing in wavelength by 0.01 nm.

The diffraction grating is limited by overlapping orders. If two wavelengths, λ and $\lambda+\Delta\lambda$, have successive orders that are coincident, then

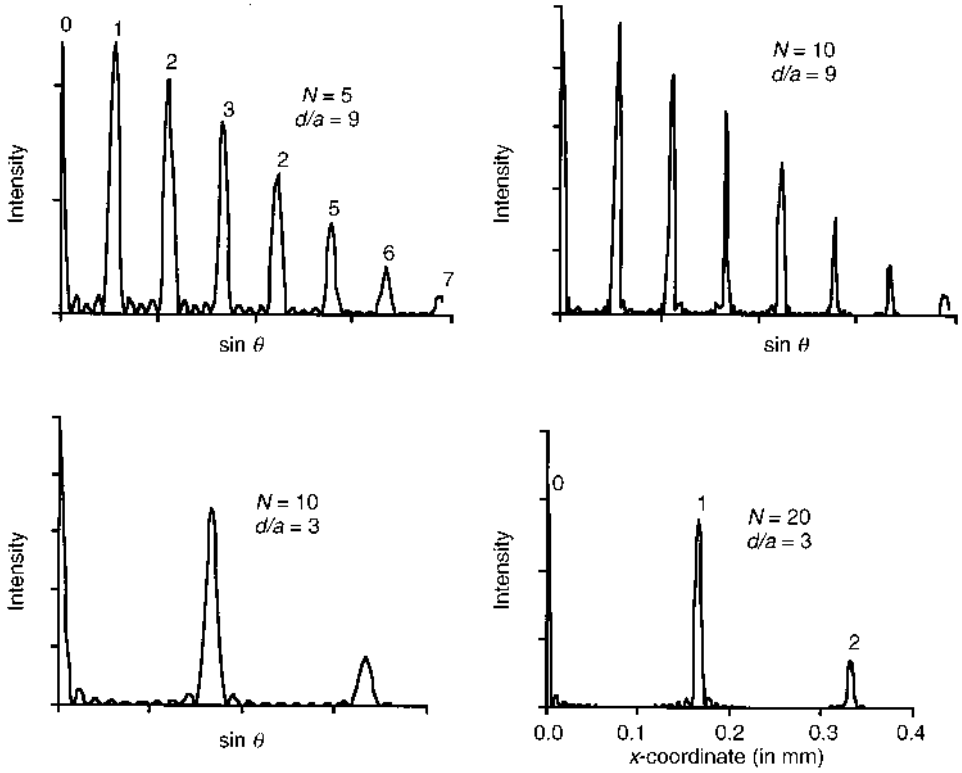

Figure 10 Decrease in the width of the principal maxima of a transmission grating with an increasing number of slits. The various principal maxima are called orders, numbering from zero, at the origin, out to as large as seven in this example. Also shown is the effect of different ratios, d/a, on the number of visible orders. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

from eqn [102]

$$(l+1)\lambda = l(\lambda + \Delta\lambda) \qquad [108]$$

The minimum wavelength difference for which this occurs is defined as the free spectral range of the diffraction grating

$$(\Delta\lambda)_{\mathrm{SR}} = \frac{\lambda}{l} \qquad [109]$$

Blazed Gratings

We have been discussing amplitude transmission gratings. Amplitude transmission gratings have little practical use because they waste light. The light loss is from a number of sources.

1. Light is diffracted simultaneously into both positive and negative orders (the positive and negative frequencies of the Fourier transform). The negative diffraction orders contain redundant information and waste light.
2. In an amplitude transmission grating, light is thrown away because of the opaque portions of the slit array.
3. The width of an individual aperture leads to a shape factor for a rectangular aperture of $\mathrm{sinc}^2\alpha = (\sin^2\alpha)/\alpha^2$, which modulates the grating factor and causes the amplitude of the orders to rapidly decrease. This can be observed in **Figure 10**, where the second order is very weak. Because of the loss in intensity at higher orders, only the first few orders ($l = 1, 2$ or 3) are useful. The shape factor also causes a decrease in diffracted light intensity with increasing wavelength for higher-order principal maxima.
4. The location of the maximum in the diffracted light, i.e., the angular position for which the shape factor is a maximum, coincides with the location of the principal maximum due to the zero-order interference. This zero-order maximum is independent of wavelength and is not of much use.

One solution to the problems created by transmission gratings would be the use of a grating that modified only the phase of the transmitted wave. Such gratings would operate using the same physical processes as a microwave phased array antenna, where the location of the shape factor's maximum is controlled by adding a constant phase shift to each antenna element. The construction of an optical

transmission phase grating with a uniform phase variation across the aperture of the grating is very difficult. For this reason, a second approach, based on the use of reflection gratings, is the practical solution to the problems listed above.

By tilting the reflecting surface of each groove of a reflection grating, Figure 11, the position of the shape factor's maximum can be controlled. Problems 1, 3, and 4 are eliminated because the shape factor maximum is moved from the optical axis out to some angle with respect to the axis. The use of reflection gratings also removes Problem 2 because all of the incident light is reflected by the grating.

Robert Wood, in 1910, developed the technique of producing grooves of a desired shape in a reflective grating by shaping the diamond tool used to cut the grooves. Gratings, with grooves shaped to enhance their performance at a particular wavelength, are said to be blazed for that wavelength. The physical properties on which blazed gratings are based can be understood by using Figure 11. The groove faces can be treated as an array of mirror surfaces. The normal to each of the groove faces makes an angle θ_B with the normal to the grating surface. We can measure the angle of incidence and the angle of diffraction with respect to the grating normal or with respect to the groove normal, as shown in Figure 11. From Figure 11, we can write a relationship between the angles

$$\theta_i = \varphi_i - \theta_B \qquad -\theta_d = -\varphi_d + \theta_B \qquad [110]$$

(The sign convention used defines positive angles as those measured in a counterclockwise rotation from the normal to the surface. Therefore, θ_B is a negative angle.) The blaze angle provides an extra degree of freedom that will allow independent

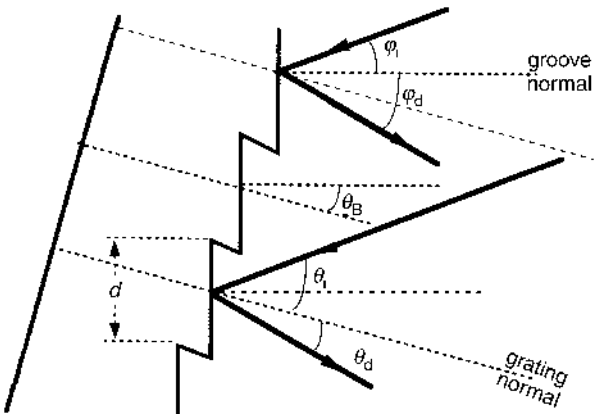

Figure 11 Geometry for a blazed reflection grating. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

adjustment of the angular location of the principal maxima of the grating factor and the zero-order, single-aperture, diffraction maximum. To see how this is accomplished, we must determine, first, the effect of off-axis illumination of a diffraction grating.

Off-axis illumination is easy to incorporate into the equation for the diffraction intensity from an array. To include the effect of an off-axis source, the phase of the illuminating wave is modified by changing the incident illumination from a plane wave of amplitude, E, traveling parallel to the optical axis, to a plane wave with the same amplitude, traveling at an angle θ_i to the optical axis: $\tilde{E}\,e^{-ikx\sin\theta_i}$. (Because we are interested only in the effects in a plane normal to the direction of propagation, we ignore the phase associated with propagation along the z-direction, $kz\cos\theta_i$.) The off-axis illumination results in a modification of the parameter for single-aperture diffraction from

$$\alpha = \frac{ka}{2}\sin\theta_d \qquad [111]$$

to

$$\alpha = \frac{ka}{2}(\sin\theta_i + \sin\theta_d) \qquad [112]$$

and for multiple aperture interference from

$$\beta = \frac{kd}{2}\sin\theta_d \qquad [113]$$

to

$$\beta = \frac{kd}{2}(\sin\theta_i + \sin\theta_d) \qquad [114]$$

The zero-order, single-aperture, diffraction peak occurs when $\alpha = 0$. If we measure the angles with respect to the groove face, this occurs when

$$\alpha = \frac{ka}{2}(\sin\varphi_i + \sin\varphi_d) = 0 \qquad [115]$$

The angles are therefore related by

$$\sin\varphi_i = -\sin\varphi_d \qquad [116]$$

$$\varphi_i = -\varphi_d$$

We see that the single-aperture, diffraction maximum (the shape factor's maximum) occurs at the same angle that reflection from the groove faces occurs. We can write this result in terms of the angles measured with respect to the grating normal

$$\theta_i = -(\theta_d + 2\theta_B) \qquad [117]$$

The blaze condition requires the single aperture diffraction maximum to occur at the lth principal

maximum, for wavelength λ_B. At that position

$$l\pi = \frac{2\pi d}{\lambda_B}(\sin\theta_i + \sin\theta_d) \qquad [118]$$

$$l\lambda_B = 2d\sin\tfrac{1}{2}(\theta_i + \theta_d)\cos\tfrac{1}{2}(\theta_i - \theta_d)$$

For the special geometrical configuration called the Littrow condition, where $\theta_i = \theta_d$, we find that [117] leads to the equation

$$l\lambda_B = 2d\sin\theta_B \qquad [119]$$

By adjusting the blaze angle, the single-aperture diffraction peak can be positioned on any order of the interference pattern. Typical blaze angles are between 15° and 30° but gratings are made with larger blaze angles.

See also

Diffraction: Fresnel Diffraction.

Further Reading

Born M and Wolf E (1970) *Principles of Optics*. New York: Pergamon Press.

Guenther RD (1990) *Modern Optics*. New York: Wiley.

Haus HA (1984) *Waves and Fields in Optoelectronics*. Englewood Cliffs, NJ: Prentice-Hall.

Hecht E (1998) *Optics*. 3rd edn. Reading, MA: Addison-Wesley.

Jackson JD (1962) *Classical Electrodynamics*. New York: Wiley.

Klein MV (1970) *Optics*. New York: Wiley.

O'Neill EL (1963) *Introduction to Statistical Optics*. Reading, MA: Addison-Wesley.

Rossi B (1957) *Optics*. Reading, MA: Addison-Wesley.

Rubinowicz A (1984) Tahe Miyamoto–Wolf diffraction wave. In: Wolf E (ed.) *Progress in Optics*, vol. IV, p. 201. Amsterdam: North-Holland.

Sommerfeld A (1964) *Optics*. New York: Academic Press.

Stone JM (1963) *Radiation and Optics*. New York: McGraw-Hill.

Fresnel Diffraction

B D Guenther, Duke University, Durham, NC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Fresnel was a civil engineer who pursued optics as a hobby, after a long day of road construction. Based on his own very high-quality observations of diffraction, Fresnel used the wave propagation concept developed by Christiaan Huygens (1629–1695) to develop a theoretical explanation of diffraction. We will use a descriptive approach to obtain the Huygens–Fresnel integral by assuming an aperture can be described by N pinholes which act as sources for Huygens' wavelets. The interference between these sources will lead to the Huygens–Fresnel integral for diffraction.

Huygens' principle views wavefronts as the product of wavelets from various luminous points acting together. To apply Huygens' principle to the propagation of light through an aperture of arbitrary shape, we need to develop a mathematical description of the field from an array of Huygens' sources filling the aperture. We will begin by obtaining the field from a pinhole that is illuminated by a plane wave,

$$\tilde{\mathbf{E}}_i(\mathbf{r}, t) = \tilde{\mathbf{E}}_i(\mathbf{r})e^{i\omega t}$$

Following the lead of Fresnel, we will use theory of interference to combine the fields from two pinholes and then generalize to N pinholes. Finally by letting the areas of each pinhole approach an infinitesimal value, we will construct an arbitrary aperture of infinitesimal pinholes. The result will be the Huygens–Fresnel integral.

We know that the wave obtained after propagation through an aperture must be a solution of the wave equation,

$$\nabla^2\tilde{\mathbf{E}} = \mu\varepsilon\frac{\partial^2\tilde{\mathbf{E}}}{\partial t^2}$$

We will be interested only in the spatial variation of the wave so we need only look for solutions of the Helmholtz equation

$$(\nabla^2 + k^2)\tilde{\mathbf{E}} = 0$$

The problem is further simplified by replacing this vector equation with a scalar equation,

$$(\nabla^2 + k^2)\tilde{E}(x, y, z) = 0$$

This replacement is proper for those cases where $\mathbf{n}E(x,y,z)$ [where $\mathbf{n}$ is a unit vector] is a solution of the vector Helmholtz equation. In general, we cannot substitute $\mathbf{n}E$ for the electric field $\mathbf{E}$ because of Maxwell's equation

$$\nabla\cdot\mathbf{E} = 0$$

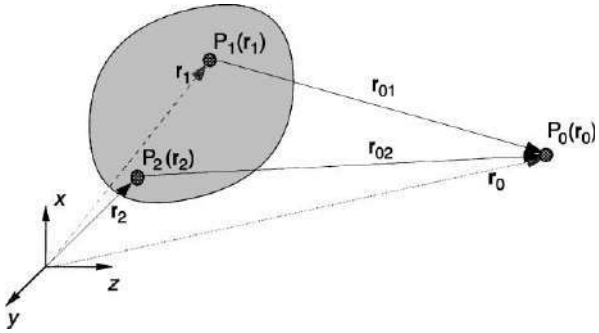

Figure 1 Geometry for application of Huygens' principle to two pinholes. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

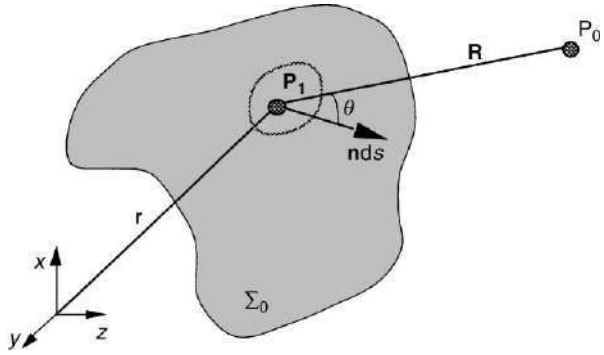

Figure 2 The geometry for calculating the field at P_0 using [9]. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

Rather than working with the magnitude of the electric field, we should use the scalar amplitude of the vector potential. We will neglect this complication and assume the scalar, E, is a single component of the vector field, **E**. A complete solution would involve a scalar solution for each component of **E**.

The pinhole is illuminated by a plane wave, and the wave that leaves the pinhole will be a spherical wave which is written in complex notation as

$$\tilde{E}(\mathbf{r})e^{i\omega t} = A\frac{e^{-i\delta}e^{-ik\cdot r}}{r}e^{i\omega t}$$

The complex amplitude

$$\tilde{E}(\mathbf{r}) = A\frac{e^{-i\delta}e^{-ik\cdot r}}{r} \qquad [1]$$

is a solution of the Helmholtz equation. The field at P_0, in Figure 1, from two pinholes: one at P_1, located a distance $\mathbf{r}_{01} = |\mathbf{r}_0 - \mathbf{r}_1|$ from P_0, and one at P_2, located a distance $\mathbf{r}_{02} = |\mathbf{r}_0 - \mathbf{r}_2|$ from P_0, is a generalization of Young's interference. The complex amplitude is

$$\tilde{E}(\mathbf{r}_0) = \frac{A_1}{r_{01}}e^{-ik\cdot r_{01}} + \frac{A_2}{r_{02}}e^{-ik\cdot r_{02}} \qquad [2]$$

We have incorporated the phase δ_1 and δ_2 into the constants A_1 and A_2 to simplify the equations.

The light emitted from the pinholes is due to a wave, E_i, incident onto the screen from the left. The units of E_i are per unit area so to obtain the amount of light passing through the pinholes, we must multiply E_i by the areas of the pinholes. If $\Delta\sigma_1$ and $\Delta\sigma_2$ are the areas of the two pinholes, respectively, then

$$A_1 \propto \tilde{E}_i(\mathbf{r}_1)\Delta\sigma_1 \qquad A_2 \propto \tilde{E}_i(\mathbf{r}_2)\Delta\sigma_2$$

$$\tilde{E}(\mathbf{r}_0) = C_1\frac{\tilde{E}_i(\mathbf{r}_1)}{r_{01}}e^{-ik\cdot r_{01}}\Delta\sigma_1 + C_2\frac{\tilde{E}_i(\mathbf{r}_2)}{r_{02}}e^{-ik\cdot r_{02}}\Delta\sigma_2 \qquad [3]$$

where C_i is the constant of proportionality. The constant will depend on the angle that $\mathbf{r}_{0i}$ makes with the normal to $\Delta\sigma_i$. This geometrical dependence arises from the fact that the apparent area of the pinhole decreases as the observation angle approaches 90°.

We can generalize [3] to N pinholes

$$\tilde{E}(\mathbf{r}_0) = \sum_{j=1}^{N} C_j\frac{\tilde{E}_i(\mathbf{r}_j)}{r_{0j}}e^{-ik\cdot r_{0j}}\Delta\sigma_j \qquad [4]$$

The pinhole's diameter is assumed to be small compared to the distances to the viewing position but large compared to a wavelength. In the limit as $\Delta\sigma_j$ goes to zero, the pinhole becomes a Huygens source. By letting N become large, we can fill the aperture with these infinitesimal Huygens' sources and convert the summation to an integral.

It is in this way that we obtain the complex amplitude, at the point P_0, see Figure 2, from a wave exiting an aperture, Σ_0, by integrating over the area of the aperture. We replace $\mathbf{r}_{0j}$, in [4], by **R**, the position of P_1, the infinitesimal Huygens' source of area, ds, measured with respect to the observation point, P_0. We also replace $\mathbf{r}_j$ by **r**, the position of the infinitesimal area, ds, with respect to the origin of the coordinate system. The discrete summation [4] becomes the integral

$$\tilde{E}(\mathbf{r}_0) = \int\int_A C(\mathbf{r})\frac{\tilde{E}_i(\mathbf{r})}{R}e^{-ik\cdot\mathbf{R}}ds \qquad [5]$$

This is the Fresnel integral. The variable $C(\mathbf{r})$ depends upon θ, the angle between **n**, the unit vector normal to the aperture, and **R**, shown in Figure 2. We now need to determine how to treat $C(\mathbf{r})$.

The Obliquity Factor

When using Huygens' principle, a problem arises with the spherical wavelet produced by each

Huygens' source. Part of the spherical wavelet propagates backward and results in an envelope propagating toward the light source, but such a wave is not observed in nature. Huygens neglected this problem. Fresnel required the existence of an obliquity factor to cancel the backward wavelet but was unable to derive its functional form. When Kirchhoff placed the theory of diffraction on a firm mathematical foundation, the obliquity factor was generated quite naturally in the derivation. In this intuitive derivation of the Huygens–Fresnel integral, we will present an argument for treating the obliquity factor as a constant at optical wavelengths. We will then derive the constant value it must be assigned.

The obliquity factor, $C(\mathbf{r})$, in [5], causes the amplitude per unit area of the transmitted light to decrease as the viewing angle increases. This is a result of a decrease in the effective aperture area with viewing angle. When Kirchhoff applied Green's theorem to the scalar diffraction problem, he found the obliquity factor to have an angular dependence given by

$$\frac{\cos(\hat{\mathbf{n}}, \mathbf{R}) - \cos(\hat{\mathbf{n}}, \mathbf{r}_s)}{2} \qquad [6]$$

which includes the geometrical effect of the incident wave arriving at the aperture, at an angle with respect to the normal to the aperture from a source located at a position $\mathbf{r}_s$, with respect to the aperture. If the source is at infinity, then the incident wave can be treated as a plane wave, incident normal to the aperture, and we may simplify the angular dependence of the obliquity factor to

$$\frac{1 + \cos\theta}{2}$$

where $\theta \rightarrow (\hat{\mathbf{n}}, \mathbf{R})$ is the angle between the normal to the aperture and the vector $\mathbf{R}$. This is the configuration shown in Figure 2. The obliquity factor provides an explanation of why it is possible to ignore the backward-propagating wave that occurs in application of Huygens' principle. For the backward wave, $\theta = \pi$, and the obliquity factor is zero.

The obliquity factor increases the difficulty of working with the Fresnel integral and it is to our benefit to be able to treat it as a constant. We can neglect the angular contribution of the obliquity factor by making an assumption about the resolving power of an optical system operating at visible wavelengths.

Assume we are attempting to resolve two stars that produce plane waves at the aperture of a telescope with an aperture diameter, a. The wavefronts from the two stars make an angle ψ with respect to each other, Figure 3:

$$\tan\psi \approx \psi = \frac{\Delta x}{a}$$

The smallest angle, ψ, that can be measured is determined by the smallest length, Δx, that can be measured. We know we can measure a fraction of wavelength with an interferometer but, without an interferometer, we can measure a length no smaller than λ, leading to the assumption that $\Delta x \geq \lambda$. This reasoning leads to the assumption that the smallest angle we can measure is

$$\psi \geq \frac{\lambda}{a} \qquad [7]$$

The resolution limit established by the above reasoning places a limit on the minimum separation that can be produced at the back focal plane of the telescope. The minimum distance on the focal plane between the images of star 1 and 2 is given by

$$d = f\psi \qquad [8]$$

From [7]

$$d = \frac{\lambda f}{a}$$

The resolution limit of the telescope can also be expressed in terms of the cone angle produced when the incident plane wave is focused on the back focal plane of the lens. From the geometry of Figure 3, the cone angle is given by

$$\tan\theta = \frac{a}{2f}$$

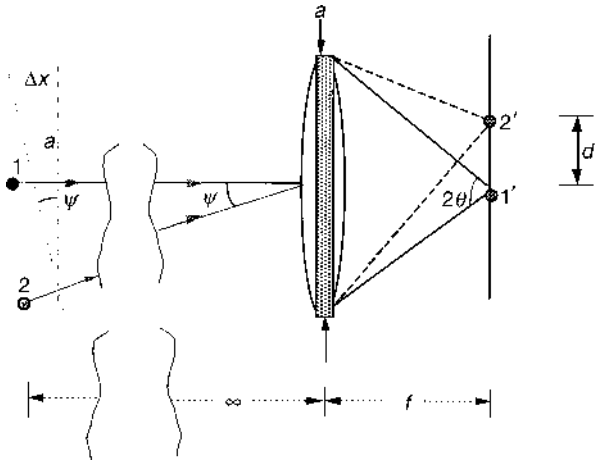

Figure 3 Telescope resolution. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

The separation between the stars 1 and 2 at the back focal plane of the lens in Figure 3 is thus

$$d = \frac{\lambda}{2\tan\theta} \qquad [9]$$

If we assume that the minimum d is 3λ (this is four times the resolution of a typical photographic film at visible wavelengths), then the largest obliquity angle that should be encountered in a visible imaging system is

$$\tan\theta = \frac{\lambda}{2d} = \frac{\lambda}{6\lambda} = \tfrac{1}{6} = 0.167$$

$$\theta = 9.5^\circ \approx 10^\circ$$

yielding a value for $\cos\theta = 0.985$ and $(1 + \cos\theta)/2 = 0.992$. The obliquity factor has only changed by 0.8% over the angular variation of 0° to 10°. The obliquity factor as a function of angle is shown in Figure 4 for an incident plane wave. The obliquity factor undergoes a 10% change when θ varies from 0° to about 40°; therefore, the obliquity factor can safely be treated as a constant in any optical system that involves angles less than 40°.

While we have shown that it is safe to ignore the variability of C, we still have not assigned a value to the obliquity factor. To find the proper value for C, we will compare the result obtained using [5] with the result predicted by using geometric optics.

We illuminate the aperture Σ_0 in Figure 5 with a plane wave of amplitude α, traveling parallel to the z-axis. Geometrical optics (equivalently the propagation of an infinite plane wave) predicts a field at P_0, on the z-axis, a distance z_0 from the aperture, given by

$$\tilde{E}_{\text{geom}} = \alpha e^{-ikz_0} \qquad [10]$$

The area of the infinitesimal at P_1 (the Huygens' source) is

$$ds = r\,dr\,d\phi$$

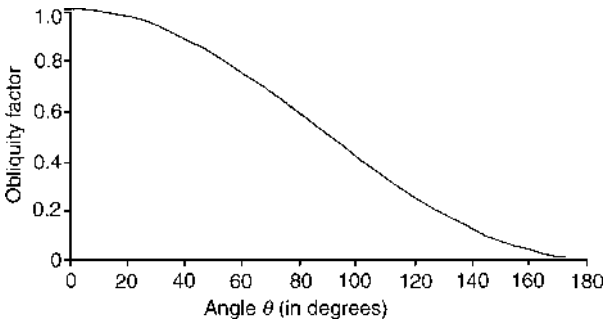

Figure 4 The obliquity factor as a function of the angle defined in eqn [6] for an incident plane wave. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

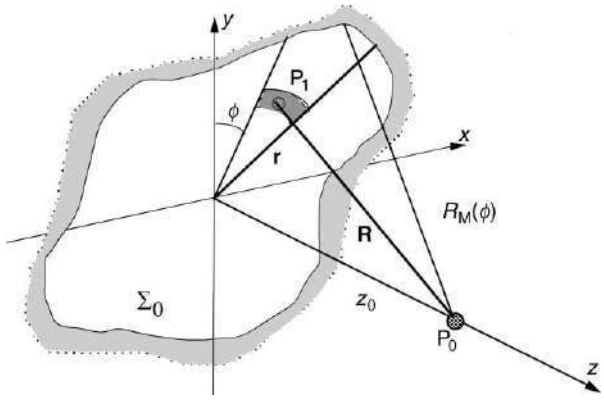

Figure 5 Geometry for evaluating the constant in the Fresnel integral. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

Based on our previous argument, the obliquity factor can be treated as a constant, C, that can be removed from under the integral. The incident wave is a plane wave whose value at $z = 0$ is $E(r) = \alpha$. Using these parameters, the Fresnel integral [5] can be written as

$$\tilde{E}(z_0) = C\alpha \int\int \frac{e^{-i\mathbf{k}\cdot\mathbf{R}}}{R} r\,dr\,d\phi \qquad [11]$$

From the geometry in Figure 5, the distance from the Huygens source to the observation point is

$$z_0^2 + r^2 = R^2$$

where z_0 is a constant equal to the distance from the observation plane to the aperture plane. The variable of integration can be written in terms of R:

$$r\,dr = R\,dR$$

The limits of integration over the aperture extend from $R = z_0$ to the maximum value of R, $R = R_M(\phi)$.

$$\tilde{E}(z_0) = C\alpha \int_0^{2\pi} \int_{z_0}^{R_M(\phi)} e^{-ikR} dR\,d\phi \qquad [12]$$

The integration over R can now be carried out to yield

$$\tilde{E}(z_0) = \underbrace{\frac{C\alpha}{ik} e^{-ikz_0} \int_0^{2\pi} d\phi}_{\text{Geometrical Optics}} - \underbrace{\frac{C\alpha}{ik} \int_0^{2\pi} e^{-ikR_M(\phi)} d\phi}_{\text{Diffraction}} \qquad [13]$$

The first integration in [13] is easy to perform and contains the amplitude at the observation point due to geometric optics. The second term may be interpreted as interference of the waves diffracted by the boundary of the aperture. An equivalent statement is that the second term is the interference of the waves scattered from the aperture's boundary, an interpretation of diffraction that was first suggested by Young.

The aperture is irregular in shape, at least on the scale of a wavelength, thus in general, kR_M will vary over many multiples of 2π as we integrate around the aperture. For this reason, we should be able to ignore the second integral in [13] if we confine our attention to the light distribution on the z-axis. After neglecting the second term, we are left with only the geometrical optics component of [13]:

$$\tilde{E}(z_0) = \frac{2\pi C}{ik}\alpha e^{-ikz_0} \qquad [14]$$

For [14] to agree with the prediction of geometric optics [10] the constant C must be equal to

$$C = \frac{ik}{2\pi} = \frac{i}{\lambda} \qquad [15]$$

The Huygens–Fresnel integral can be written, using the value for C just derived,

$$\tilde{E}(\mathbf{r}_0) = \frac{i}{\lambda}\int\int_{\Sigma}\frac{\tilde{E}_i(\mathbf{r})}{R}e^{-i\mathbf{k}\cdot\mathbf{R}}\,ds \qquad [16]$$

The job of calculating diffraction has only begun with the derivation of the Huygens–Fresnel integral [16]. Rigorous solutions exist for only a few idealized obstructions. To allow discussion of the general properties of diffraction, it is necessary to use approximate solutions. To determine what type approximations we can make let's look at the light propagation path shown in Figure 6.

For the second term in [13] to contribute to the field at point P_0, in Figure 6, the phase of the exponent that we neglected in [13] must not vary over 2π when the integration is performed over the aperture, i.e.,

$$\Delta = \mathbf{k}\cdot(\mathbf{R}_1 - \mathbf{R}'_1 + \mathbf{R}_2 - \mathbf{R}'_2) < 2\pi$$

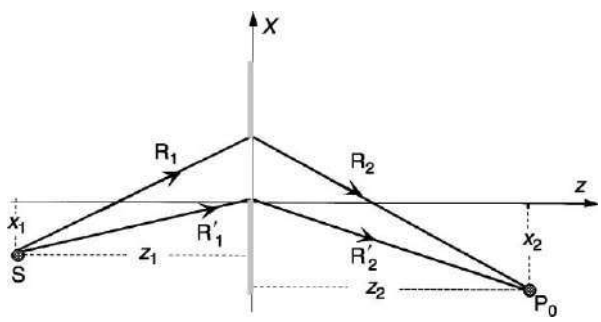

Figure 6 One-dimensional slit used to establish Fresnel approximations. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

From the geometry of Figure 6, the two paths from the source, S, to the observation point, P_0, are

$$R_1 + R_2 = \sqrt{z_1^2 + (x_1 + b)^2} + \sqrt{z_2^2 + (x_2 + b)^2}$$

$$R'_1 + R'_2 = \sqrt{z_1^2 + x_1^2} + \sqrt{z_2^2 + x_2^2}$$

By assuming that the aperture, of width b, is small compared to the distances z_1 and z_2, the difference between these distances, Δ, can be rewritten as a binomial expansion of a square root

$$\sqrt{1+b} = 1 + \frac{b}{2} - \frac{b^2}{8} + \ldots \qquad [17]$$

thus

$$\Delta \approx k\left(\frac{x_1}{z_1} + \frac{x_2}{z_2}\right)b + \frac{k}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)b^2 + \ldots \qquad [18]$$

If we assume that the second term of this expansion is small, i.e.,

$$\frac{k}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)b^2 \ll 2\pi$$

$$\frac{b^2}{2}\frac{1}{z_1} + \frac{1}{z_2} < \lambda$$

we see that the phase of the wavefront in the aperture is assumed to have a quadratic dependence upon aperture coordinates. Diffraction predicted by this assumption is called Fresnel diffraction.

To discover the details of the approximation consider Figure 7. In the geometry of Figure 7, the Fresnel integral [16] becomes

$$\tilde{E}_{P_0} = \frac{i\alpha}{\lambda}\int\int_{\Sigma} f(x,y)\frac{e^{-i\mathbf{k}\cdot(\mathbf{R}+\mathbf{R}')}}{RR'}\,dx\,dy \qquad [19]$$

As we mentioned in our discussion of [13] the integral that adds diffractive effects to the geometrical optics field is nonzero only when the phase of the integrand is stationary. For Fresnel diffraction, we can insure that the phase is nearly constant if R does not differ appreciably from Z, or R' from Z'. This is equivalent to stating that only the wave in the aperture, around a point in Figure 7 labeled **S**, called the stationary point, will contribute to E_P. The stationary point, **S**, is the point in the aperture plane where the line, connecting the source and observation positions, intersects the plane. Physically, only light propagating over paths nearly equal to the path predicted by geometrical optics (obtained from Fermat's principle) will contribute to E_{P_0}.

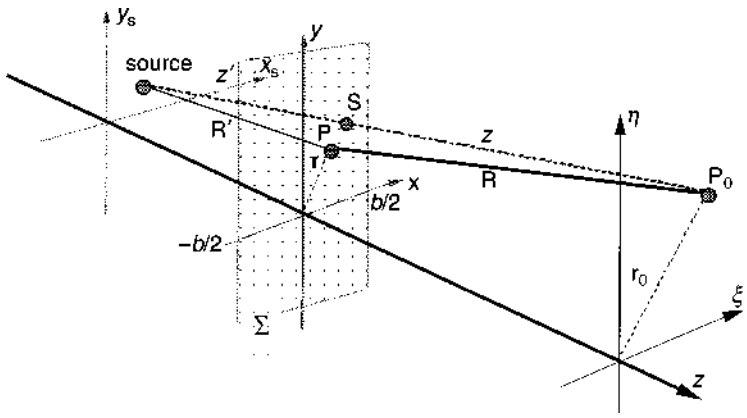

Figure 7 Geometry for Fresnel diffraction. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

The geometry of Figure 7 allows the distances R and R' to be written as

$$R^2 = (x-\xi)^2 + (y-\eta)^2 + Z^2 \qquad [20]$$
$$R'2 = (x_s - x)^2 + (y_s - y)^2 + Z'2$$

To solve these expressions for R and R', we apply the binomial expansion [17] and retain the first two terms:

$$R + R' \approx Z + Z' + \left[\frac{(x-\xi)^2}{2Z} + \frac{(x_s - x)^2}{2Z'}\right] + \left[\frac{(y-\eta)^2}{2Z} + \frac{(y_s - y)^2}{2Z'}\right] \qquad [21]$$

In the denominator of [19], R is replaced by Z and R' by Z'. (By making this replacement, we are implicitly assuming that the amplitudes of the spherical waves are not modified by the differences in propagation distances over the area of integration. This is a reasonable approximation because the two propagation distances are within a few hundred wavelengths of each other). In the exponent, we must use [21], because the phase changes by a large fraction of a wavelength, as we move from point to point in the aperture.

If the wave, incident on the aperture, were a plane wave we can assume $R' = \infty$ and [19] becomes

$$\tilde{E}_{P_0} = \frac{i\alpha}{\lambda}\frac{e^{-ikZ}}{Z}\int\int_{\Sigma} f(x,y)e^{-\frac{ik}{2Z}[(x-\xi)^2+(y-\eta)^2]}dx\,dy \qquad [22]$$

The physical interpretation of [22] states that when a plane wave illuminates the obstruction, the field at point P_0 is a spherical wave, originating at the aperture, a distance Z away from P_0:

$$\frac{e^{-ikZ}}{Z}$$

The amplitude and phase of this spherical wave are modified by the integral containing a quadratic phase dependent on the obstruction's spatial coordinates.

By defining three new parameters,

$$\rho = \frac{ZZ'}{Z+Z'} \quad \text{or} \quad \frac{1}{\rho} = \frac{1}{Z} + \frac{1}{Z'} \qquad [23a]$$

$$x_0 = \frac{Z'\xi + Zx_s}{Z+Z'} \qquad [23b]$$

$$y_0 = \frac{Z'\eta + Zy_s}{Z+Z'} \qquad [23c]$$

the more general expression for Fresnel diffraction of a spherical wave can be placed in the same format as [22]. The newly defined parameters x_0 and y_0 are the coordinates, in the aperture plane, of the stationary point, **S**. The parameters in [23] can be used to express, after some manipulation, the spatial dependence of the phase, in the integrand of [19]

$$R + R' = Z + Z' + \frac{(\xi - x_s)^2 + (\eta - y_s)^2}{2(Z+Z')} + \left[\frac{(x-x_0)^2 + (y-y_0)^2}{2\rho}\right]$$

A further simplification to the Fresnel diffraction integral can be made by obtaining an expression for the distance, D, between the source and the observation point:

$$D = Z + Z' + \frac{(\xi - x_s)^2 + (\eta - y_s)^2}{2(Z+Z')} \qquad [24]$$

We use this definition of D to also write

$$\frac{1}{ZZ'} = \frac{Z+Z'}{ZZ'}\frac{1}{Z+Z'} \approx \frac{1}{\rho D}$$

Using the parameters we have just defined, we may rewrite [19] as

$$\tilde{E}_{P_0} = \frac{i\alpha}{\lambda\rho D} e^{-ikD} \iint f(x,y) e^{-\frac{ik}{2\rho}\left[(x-x_0)^2+(y-y_0)^2\right]} dx\,dy \qquad [25]$$

With the use of the variables defined in [23], the physical significance of the general expression of the Huygens–Fresnel integral can be understood in an equivalent fashion as was [22]. At point P_0, a spherical wave

$$\frac{e^{-ikD}}{D}$$

originating at the source, a distance D away, is observed, if no obstruction are present. Because of the obstruction, the amplitude and phase of the spherical wave are modified by the integral in [25]. This correction to the predictions of geometrical optics is called Fresnel diffraction.

Mathematically, [25] is an application of the method of stationary phase, a technique developed in 1887 by Lord Kelvin to calculate the form of a boat's wake. The integration is nonzero only in the region of the critical point we have labeled **S**. Physically, the light distribution, at the observation point, is due to wavelets from the region around **S**. The phase variations of light, coming from other regions in the aperture, are so rapid that the value of the integral over those spatial coordinates is zero. The calculation of the integral for Fresnel diffraction is complicated because, if the observation point is moved, a new integration around a different stationary point in the aperture must be performed.

Rectangular Apertures

If the aperture function, $f(x,y)$, is separable in the spatial coordinates of the aperture, then we can rewrite [25] as

$$\tilde{E}_{P_0} = \frac{i\alpha}{\lambda\rho D} e^{-ikD} \int_{-\infty}^{\infty} f(x) e^{-ig(x)} dx \int_{-\infty}^{\infty} f(y) e^{-ig(y)} dy$$

$$\tilde{E}_{P_0} = A[C(x) - iS(x)][C(y) - iS(y)]$$

A represents the spherical wave from the source, a distance D away,

$$A = \frac{i\alpha}{2D} e^{-ikD}$$

If we treat the aperture function as a simple constant, C and S are integrals of the form

$$C(x) = \int_{x_1}^{x_2} \cos[g(x)]dx \qquad [26a]$$

$$S(x) = \int_{x_1}^{x_2} \sin[g(x)]dx \qquad [26b]$$

The integrals, $C(x)$ and $S(x)$, have been evaluated numerically and are found in tables of Fresnel integrals. To use the tabulated values for the integrals, [26] must be written in a general form

$$C(w) = \int_0^w \cos\left(\frac{\pi u^2}{2}\right) du \qquad [27]$$

$$S(w) = \int_0^w \sin\left(\frac{\pi u^2}{2}\right) du \qquad [28]$$

The variable u is an aperture coordinate, measured relative to the stationary point, $\mathbf{S}(x_0,y_0)$, in units of

$$\sqrt{\frac{\lambda\rho}{2}}$$

$$u = \sqrt{\frac{2}{\lambda\rho}}(x - x_0) \text{ or } u = \sqrt{\frac{2}{\lambda\rho}}(y - y_0) \qquad [29]$$

The parameter w in [27] and [28] specifies the location of the aperture edge relative to the stationary point **S**. The parameter w is calculated through the use of [29] with x and y replaced by the coordinates of the aperture edge.

The Cornu Spiral

The plot of $S(w)$ versus $C(w)$, shown in Figure 8, is called the Cornu spiral in honor of M. Alfred Cornu (1841–1902) who was the first to use this plot for graphical evaluation of the Fresnel integrals.

To use the Cornu spiral, the limits w_1 and w_2 of the aperture are located along the arc of the spiral. The length of the straight line segment drawn from w_1 to w_2 gives the magnitude of the integral. For example, if there were no aperture present, then, for the x-dimension, $w_1 = -\infty$ and $w_2 = \infty$. The length of the line segment from the point $(-1/2,1/2)$ to $(1/2,1/2)$ would be the value of E_{P_0}, i.e., $\sqrt{2}$. An identical value is obtained for the y-dimension, so that

$$\tilde{E}_{P_0} = 2A = \frac{\alpha e^{-ikD}}{D}$$

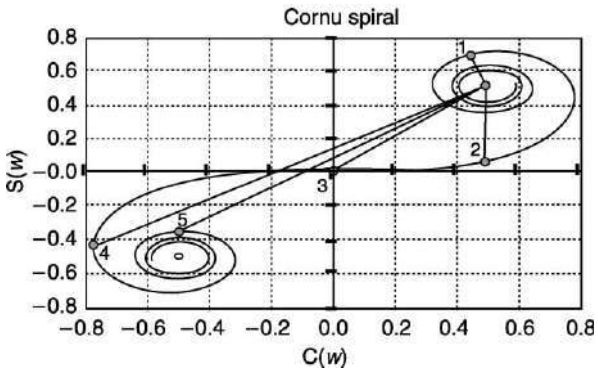

Figure 8 Cornu spiral obtained by plotting the values for $C(w)$ versus $S(w)$ from a table of Fresnel integrals. The numbers indicated along the arc of the spiral are values of w. The points 1–5 correspond to observation points shown in **Figure 9** and are used to calculate the light intensity diffracted by a straight edge. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

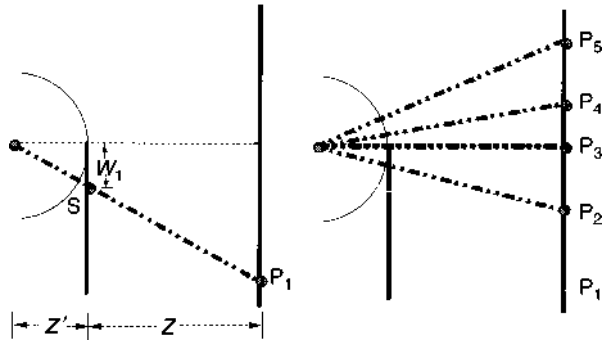

Figure 9 A geometrical construct to determine w_1 for the stationary point S associated with five different observation points. The values of w are then used in **Figure 8** to calculate the intensity of light diffracted around a straight edge. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

This is the spherical wave that is expected when no aperture is present.

As an example of the application of the Fresnel integrals, assume the diffracting aperture is an infinitely long straight edge reducing the problem to one dimension:

$$f(x, y) = \begin{Bmatrix} 1 & x \geq x_1 \\ 0 & x < x_1 \end{Bmatrix}$$

The value of E_{P_0} in the y-dimension is from above $\sqrt{2}$ since $w_1 = -\infty$ and $w_2 = \infty$. In the x-dimension, the value of w at the edge is

$$w_{x_1} = \sqrt{\frac{2}{\lambda\rho}}(x_1 - x_0)$$

We will assume that the straight edge blocks the negative half-plane so that $x_1 = 0$. The straight edge is treated as an infinitely wide slit with one edge located at infinity so that $w_2 = \infty$ in the upper half-plane. When the observation point is moved, the coordinate x_0 of **S** changes and the origin of the aperture's coordinate system moves. New values for the w's must therefore be calculated for each observation point.

Figure 9 shows the assumed geometry. A set of observation points, P_1–P_5, on the observation screen are selected. The origin of the coordinate system in the aperture plane is relocated to a new position (given by the position of **S**) when a new observation point is selected. The value of w_1, the position of the edge with respect to **S**, must be recalculated for each observation point. The distance from the origin to the straight edge, w_1, is positive for P_1 and P_2, zero for P_3, and negative for P_4 and P_5.

Table 1 Fresnel integrals for straight edge

w_1	$C(w_1)$	$S(w_1)$	I_P/I_0	P_j
∞	0.5	0.5	0	
2.0	0.4882	0.3434	0.01	
1.5	0.4453	0.6975	0.021	P_1
1.0	0.7799	0.4383	0.04	
0.5	0.4923	0.0647	0.09	P_2
0	0	0	0.25	P_3
−0.5	−0.4923	−0.0647	0.65	
−1.0	−0.7799	−0.4383	1.26	P_4
−1.2	−0.7154	−0.6234	1.37	
−1.5	−0.4453	−0.6975	1.16	
−2.0	−0.4882	−0.3434	0.84	P_5
−2.5	−0.4574	−0.6192	1.08	
−∞	−0.5	−0.5	1.0	

Figure 8 shows the geometrical method used to calculate the intensity values at each of the observation points in **Figure 9** using the Cornu spiral. The numbers labeling the straight line segments in **Figure 8** are associated with the labels of the observation points in **Figure 9**.

To obtain an accurate calculation of Fresnel diffraction from a straight edge, a table of Fresnel integrals provides the input to the following equation

$$I_P = I_0\left\{\left[\tfrac{1}{2} - C(w_1)\right]^2 + \left[\tfrac{1}{2} - S(w_1)\right]^2\right\}$$

where $I_0 = 2A^2$. **Table 1**, shows the values extracted from the table of Fresnel integrals to find the relative intensity at various observation points. The result obtained by using either method for calculating the light distribution in the observation plane, due to the straight edge in **Figure 9**, is plotted in **Figure 10**. The relative intensities at the observation points

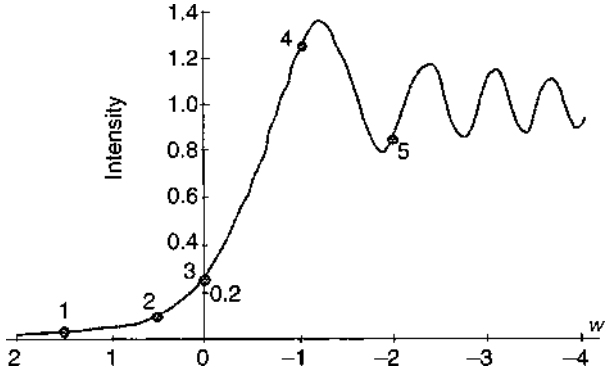

Figure 10 Light diffracted by a straight edge with its edge located at $w = 0$. The points labeled 1–5 were found using the construction in **Figure 9** to obtain w. This was then used in **Figure 8** to find a position on the Cornu spiral. The length of the lines shown in **Figure 8** from the (1/2,1/2) point to the numbered positions led to the intensities shown. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

depicted in Figure 9 are labeled on the diffraction curve of Figure 10.

Fresnel Zones

Fresnel used a geometrical construction of zones to evaluate the Huygens–Fresnel integral. The Fresnel zone is a mathematical construct, serving the role of a Huygens source in the description of wave propagation. Assume that at time t, a spherical wavefront from a source at P_1 has a radius of R'. To determine the field at the observation point, P_0, due to this wavefront, a set of concentric spheres of radii, $Z, Z + (\lambda/2), Z + 2(\lambda/2), \ldots, Z + j(\lambda/2), \ldots$ are constructed, where Z is the distance from the wavefront to the observation point on the line connecting P_1 and P_0 (see Figure 11). These spheres divide the wavefront into a number of zones, $\zeta_1, \zeta_2, \ldots, \zeta_j \ldots$, called Fresnel zones, or half-period zones.

We treat each zone as a circular aperture illuminated from the left by a spherical wave of the form

$$\tilde{E}_j(\mathbf{R}') = \frac{A e^{-i\mathbf{k}\cdot\mathbf{R}'}}{R'} = \frac{A e^{-ikR'}}{R'}$$

R' is the radius of the spherical wave. The field at P_0 due to the jth zone is obtained by using [5]

$$\tilde{E}_j(P_0) = \frac{A}{R'} e^{-i\mathbf{k}\cdot\mathbf{R}'} \int\int_{\zeta_j} C(\varphi) \frac{e^{-i\mathbf{k}\cdot\mathbf{R}}}{R} ds \qquad [30]$$

For the integration over the jth zone, the surface element is

$$ds = R' 2 \sin\theta d\theta \, d\phi \qquad [31]$$

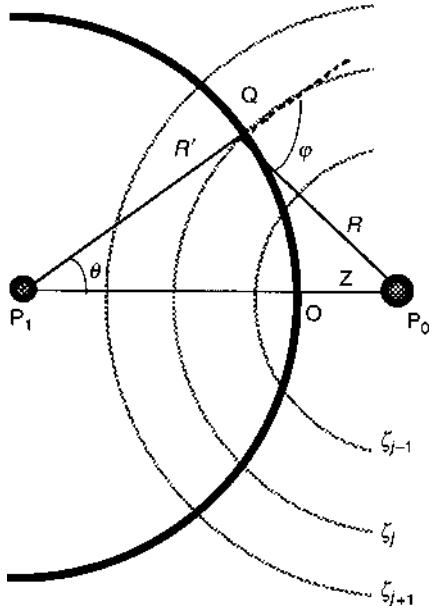

Figure 11 Construction of Fresnel zones observed from a position P_0 on a spherical wave originating from a source at point P_1. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

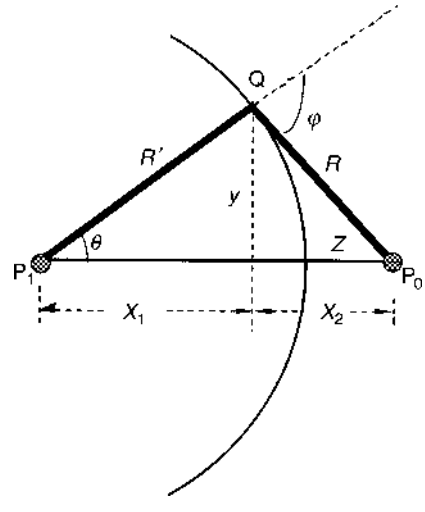

Figure 12 Geometry for finding the relationship between R and R' yielding eqn [32]. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

The limits of integration extend over the range

$$Z + (j-1)\tfrac{\lambda}{2} \le R \le Z + j\tfrac{\lambda}{2}$$

The variable of integration is R; thus, a relationship between R' and R must be found. This is accomplished by using the geometrical construction shown in Figure 12.

A perpendicular is drawn from the point Q on the spherical wave to the line connecting P_1 and P_0, in Figure 12. The distance from the source to the

observation point is $x_1 + x_2$ and the distance from the source to the plane of the zone is the radius of the incident spherical wave, R'. The distance from P_1 to P_0 can be written

$$x_1 + x_2 = R' + Z$$

The distance from the observation point to the zone is

$$R^2 = x_2^2 + y^2 = [(R' + Z) - x_1]^2 + (R'\sin\theta)^2$$
$$= R'2 + (R' + Z)^2 - 2R'(R' + Z)\cos\theta$$

The derivative of this expression yields

$$R\,\mathrm{d}R = R'(R' + Z)\sin\theta\,\mathrm{d}\theta \qquad [32]$$

Substituting [32] into [31] gives

$$\mathrm{d}s = \frac{R'}{R' + Z} R\mathrm{d}R\mathrm{d}\phi \qquad [33]$$

The integration over ϕ is accomplished by rotating the surface element about the P_1P_0 axis. After integrating over ϕ between the limits of 0 and 2π, we obtain

$$\tilde{E}_j(P_0) = \frac{2\pi A}{R' + Z} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}'} \int_{Z+(j-1)\frac{\lambda}{2}}^{Z+\frac{j\lambda}{2}} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} C(\varphi)\mathrm{d}R \qquad [34]$$

We assume that $R',\ Z \gg \lambda$ so that the obliquity factor is a constant over a single zone, i.e., $C(\varphi) = C_j$. The obliquity factor is not really a constant as we earlier assumed but rather a very slowly varying parameter of j, changing by less than two parts in 10^4 when 'j' changes by 500 (for $Z = 1$ meter and $\lambda = 500$ nm). In fact it only changes by 2 parts in 10^{-7} across one zone. We are safe in applying the assumption that the obliquity factor is a constant over a zone and this allows the integral in [34] to be calculated:

$$\tilde{E}_j(P_0) = \frac{2\pi \mathrm{i} C_j A}{k(R' + Z)} \mathrm{e}^{-\mathrm{i}k(R'+Z)} \mathrm{e}^{-\mathrm{i}kj\frac{\lambda}{2}} \left[1 - \mathrm{e}^{-\mathrm{i}k\frac{\lambda}{2}}\right] \qquad [35]$$

Using the identity $k\lambda = 2\pi$ and the definition for the distance between the source and observation point [24] modified for this geometry, $D = R' + Z$ [35] can be simplified to

$$\tilde{E}_j(P_0) = 2\mathrm{i}\lambda(-1)^j \frac{C_j A}{D} \mathrm{e}^{-\mathrm{i}kD} \qquad [36]$$

The physical reasons for the behavior predicted by [36] are quite easy to understand. The distance from P_0 to a zone changes by only $\lambda/2$ as we move from zone to zone and the area of a zone is almost a constant, independent of the zone number; thus, the amplitudes of the Huygens' wavelets from each zone should be approximately equal. The alternation in sign, from zone to zone, is due to the phase change of the light wave from adjacent zones because the propagation paths for adjacent zones differ by $\lambda/2$.

To find the total field strength at P_0 due to N zones, the collection of Huygens' wavelets is added:

$$E(P_0) = \sum_{j=1}^{N} \tilde{E}_j(P_0) = \frac{2\mathrm{i}\lambda A}{D} \mathrm{e}^{-\mathrm{i}kD} \sum_{j=1}^{N} (-1)^j C_j \qquad [37]$$

To evaluate the sum, the elements of the sum are regrouped and rewritten as

$$-\sum_{j=1}^{N} (-1)^j C_j = \frac{C_1}{2} + \left(\frac{C_1}{2} - C_2 + \frac{C_3}{2}\right) + \left(\frac{C_3}{2} - C_4 + \frac{C_5}{2}\right) + \ldots$$

Because the C's are very slowly varying functions of j, even out to 500 zones, we are justified in setting the quantities in parentheses equal to zero. With this approximation, the summation can be set equal to one of two values, depending upon whether there is an even or odd number of terms in the summation

$$-\sum_{j=1}^{N} (-1)^j C_j = \left\{ \begin{array}{ll} \frac{1}{2}(C_1 + C_N) & N \text{ odd} \\ \frac{1}{2}(C_1 - C_N) & N \text{ even} \end{array} \right\}$$

For very large N, the obliquity factor approaches zero, $C_N \to 0$, as was demonstrated in Figure 4. Thus, the theory has led us to the conclusion that the total field produced by an unobstructed wave is equal to one half the contribution from the first Fresnel zone, i.e., $E = E_1/2$. Stating this result in a slightly different way, we obtain a surprising result – the contribution from the first Fresnel zone is twice the amplitude of the unobstructed wave!

The zone construction can be used to analyze the effect of the obstruction of all or part of a zone. For example, by constructing a circular aperture with a diameter equal to the diameter of the first Fresnel zone, we have just demonstrated that it is possible to produce an intensity at the point P_0 equal to four times the intensity that would be observed if no aperture were present. To analyze the effects of a smaller aperture, we subdivide a half-period zone into a number of subzones such that there is a constant phase difference between each subzone. The individual vectors form a curve know as the vibrational curve. Figure 13 shows a vibrational curve produced by the addition of waves from nine subzones of the first Fresnel zone. The vibrational curve in Figure 13 is an arc with the appearance of a half-circle.

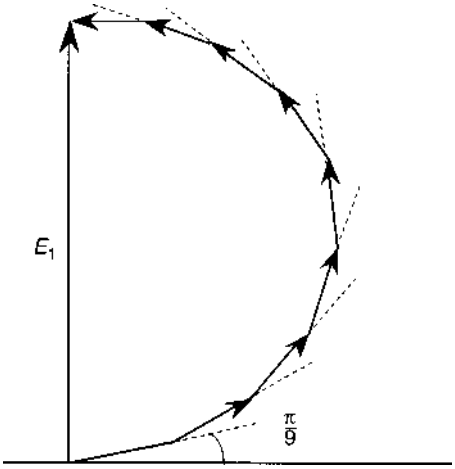

Figure 13 Vector addition of waves from nine subzones constructed in the first Fresnel zone. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

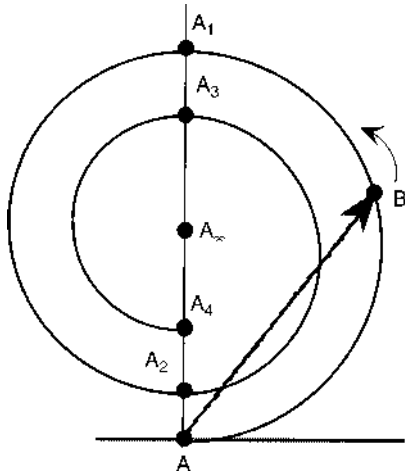

Figure 14 Vibration curve for determining the Fresnel diffraction from a circular aperture. The change of the diameter of the half-circles making up the spiral has been exaggerated for easy visualization. The actual changes are one part in 10^{-5}. The position of point B is determined by the aperture size. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

If the radius of curvature of the arc were calculated, we would discover that it is a constant, except for the contribution of the obliquity factor,

$$\rho = \frac{\lambda R'}{R' + Z} C(\varphi)$$

Because the obliquity factor for a single zone is a constant to 2 parts in 10^{-7}, the radius of curvature of the vibration curve can be considered a constant over a single zone. If we let the number of subzones approach infinity, the vibrational curve becomes a semicircle whose chord is equal to the wavelet produced by zone one, i.e., E_1. If we subdivide additional zones, and add the subzone contributions, we create other half-circles whose radii decrease at the same rate as the obliquity factor. The vibrational curve for the total wave is a spiral, constructed of semicircles which converges to a point halfway between the first half-circle, see Figure 14. The length of the vector from the start to the end of the spiral is $E = E_1/2$, as we derived above. When this same construction technique is applied to a rectangular aperture, the vibrational curve generated is the Cornu spiral.

Circular Aperture

The vector addition technique, described in Figure 13, can be used to evaluate Fresnel diffraction, at a point P_0, from a circular aperture and yields the intensity distribution along the axis of symmetry of the circular aperture. The zone concept will also allow a qualitative description of the light distribution normal to this axis.

To develop a quantitative estimate of the intensity at an observation point on the axis of symmetry of the circular aperture, we construct a spiral, Figure 14, to represent the Fresnel zones of a spherical wave incident on the aperture. The point B on the spiral shown in Figure 14 corresponds to the portion of the spherical wave unobstructed by the screen. The length of the chord, AB, represents the amplitude of the light wave at the observation point P_0. As the diameter of the aperture increases, B moves along the spiral, in a counterclockwise fashion, away from A. The first maximum occurs when B reaches the point labeled A_1 in Figure 14; the aperture has then uncovered the first Fresnel zone. At this point, the amplitude is twice what it would be with no obstruction. Four times the intensity!

If the aperture's diameter continues to increase, B reaches the point labeled A_2 in Figure 14 and the amplitude is very nearly zero; two zones are now exposed in the aperture. Further maxima occur when an odd number of zones are in the aperture and further minima when an even number of zones are exposed. Figure 15 shows an aperture containing four exposed Fresnel zones. The amplitude at the observation point would correspond to the chord drawn from A to A_4 in Figure 14.

The aperture diameter can be fixed and the observation point P_0 can move along, or perpendicular to, the axis of symmetry of the circular aperture. As P_0 is moved away from the aperture, along the symmetry axis, i.e., as Z increases, the radius of the Fresnel zones increase without limit. For small values of n, the radius of the nth zone can be approximated by

$$r_n \approx \sqrt{nZ\lambda} \qquad [38]$$

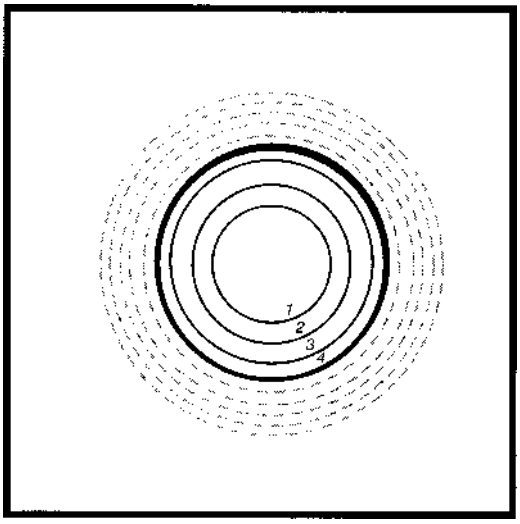

Figure 15 Aperture with four Fresnel zones exposed. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

At Z_{max}, the light intensity is a maximum, given by the chord length from A to A_1 in Figure 14. If $\lambda = 500$ nm and $a = 0.5$ mm then this maximum occurs when $Z = 0.5$ m

$$Z_{max} = \frac{a^2}{\lambda} \qquad [39]$$

If we start at Z_{max} and move toward the aperture, along the axis, as Z decreases in value, a point will be reached when the intensity on the axis becomes a minimum. The value of Z where the first minimum in intensity is observed is equal to

$$Z_{min} = \frac{a^2}{2\lambda}$$

In Figure 14 the chord would extend from A to A_2.

As the observation point, P_0, is moved along the axis toward the aperture, and Z assumes values less than Z_{min}, the point B in Figure 14 spirals inward, toward the center of the spiral and the intensity cycles through maximum and minimum values. The cycling of the intensity, as the observation point moves toward the aperture, will not continue indefinitely. At some point, the field on the axis will approach the field observed without the aperture because the distance between Z_{max} and Z_{min} shrinks to the wavelength of light making the intensity variation unobservable.

For values of Z that exceed Z_{max} of [39], the aperture radius, a, will be smaller than the radius of the first zone, and Fraunhofer diffraction will be observed because the aperture contains only one zone and the phase in the aperture is a constant.

Opaque Screen

If the screen containing a circular aperture, of radius a, is replaced by an opaque disk of radius a, the intensity distribution on the symmetry axis, behind the disk, is found to be equal to the value that would be observed with no disk present. This prediction was first derived by Poisson, to demonstrate that wave theory was incorrect; however, experimental observation supported the prediction and verified the theory.

We construct a spiral, shown in Figure 16, similar to the one for the circular aperture in Figure 14. The point B on the spiral represents the edge of the disk. The shaded portion of the spiral from A to B does not contribute because that portion of the wave is covered by the disk and the zones, associated with that portion of the wave, cannot be seen from the observation point. The amplitude at P_0 is the length of the chord from B to A_∞ shown in Figure 16. If the observation point moves toward the disk, then B moves along the spiral toward A_∞. There is always intensity on the axis for this configuration, though it slowly decreases until it reaches zero when the observation point reaches the disk; this corresponds to point B reaching point A_∞ on the spiral. Physically, zero intensity occurs when the disk blocks the entire light wave. There are no maxima or minima observed as the disk diameter, a, increases or as the observation distance changes. If the observation point is moved perpendicular to the symmetry axis, a set of concentric bright rings are observed. The origin of these bright rings can be explained using Fresnel zones, in a manner similar to the one used to explain the bright rings observed in a Fresnel diffraction pattern from a circular aperture.

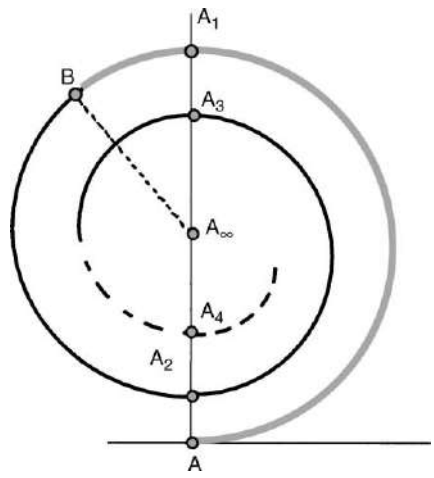

Figure 16 Vibration curve for opaque disk. The shaded region makes no contribution because it is associated with the portion of the wave obstructed by the opaque object. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

Zone Plate

In the construction of Fresnel zones, each zone was assumed to produce a Huygens wavelet, out of phase with the wavelets produced by its nearest neighbors. If every other zone were blocked, then there would be no negative contributions to [37]. The intensity on-axis would be equal to the square of the sum of the amplitudes produced by the in-phase zones – exceeding the intensity of the incident wave. An optical component made by the obstruction of either the odd or the even zones could therefore be used to concentrate the energy in a light wave.

The boundaries of the opaque zones, used to block out-of-phase wavelets, are seen from [38] to increase as the square root of the integers. An array of opaque rings, constructed according to this prescription is called a zone plate, see Figure 17.

A zone plate will perform like a lens with focal length

$$f = \pm \frac{r_1^2}{\lambda} \qquad [40]$$

The zone plate, shown in Figure 17, will act as both a positive and negative lens. What we originally called the source can now be labeled the object point, O, and what we called the observation point can now be labeled the image point, I. The light passing through the zone plate is diffracted into two paths, labeled C and D in Figure 17. The light waves, labeled C, converge to a real image point I. For these waves the zone plate performs the role of a positive lens with a focal length given by the positive value of [40]. The light waves, labeled D in Figure 17, appear to originate from the virtual image point labeled $\mathbb{I}$. For these waves the zone plate performs the role of a negative lens with a focal length given by the negative value of [40].

The zone plate will not have a single focus, as is the case for a refractive optical element, but rather will have multiple foci. As we move toward the zone plate, from the first focus, given by [40], the effective Fresnel zones will decrease in diameter. The zone plate will no longer obstruct out-of-phase Fresnel zones and the light intensity on the axis will decrease. However, additional maxima, of the on-axis intensity, will be observed at values of Z for which the first zone plate opening contains an odd number of zones. These positions can also be labeled as foci of the zone plate; however, the intensity at each of these foci will be less than the intensity at the primary focus.

Lord Rayleigh suggested that an improvement of the zone plate design would result if, instead of blocking every other zone, we shifted the phase of alternate zones by 180°. The resulting zone plate, called a phase-reversal zone plate, would, more efficiently, utilize the incident light. R.W. Wood was the first to make such a zone plate. Holography provides an optical method of constructing the phase-reversal zone plates in the visible region of the spectrum. Semiconductor lithography has been used to produce zone plates in the x-ray region of the spectrum.

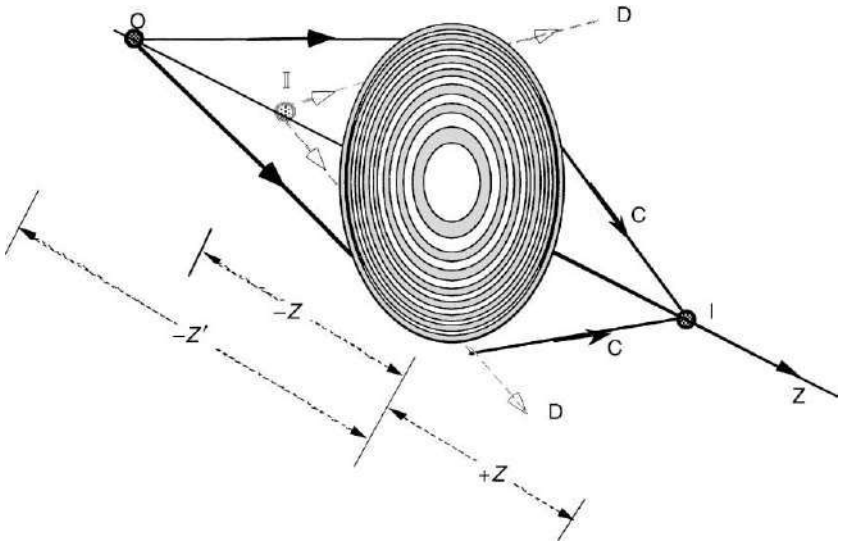

Figure 17 A zone plate acts as if it were both a positive and a negative lens. Light from the object, O, is diffracted into waves traveling in both the D and the C directions. The light traveling in the C direction produces a real image of O at I. The light traveling in the D direction appears to originate from a virtual image at I. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

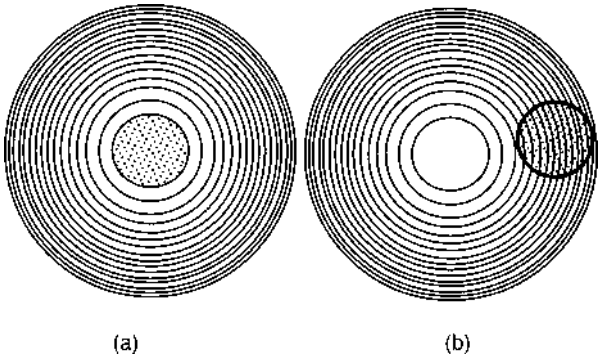

Figure 18 (a) A set of Fresnel zones has been constructed about the optical path taken by some light ray. If the optical path of the ray is varied over the cross-hatched area shown in the figure, then the optical path length does not change. This cross-hatched area is equal to the first Fresnel zone and is described as the neighborhood of the light ray. (b) The neighborhood defined in (a) is moved so that it surrounds an incorrect optical path for a light ray. We see that this region of space would contribute no wave amplitude at the observation point because of the destructive interference between the large number of partial zones contained in the neighborhood. Reprinted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley & Sons.

The resolving power of a zone plate is a function of the number of zones contained in the plate. When the number of zones exceeds 200, the zone plate's resolution approaches that of a refractive lens.

Fermat's Principle

The Fresnel zone construction provides physical insight into the interpretation of Fermat's principle which states that, if the optical path length of a light ray is varied in a neighborhood about the true path, there is no change in path length. By constructing a set of Fresnel zones about the optical path of a light ray, we can discover the physical meaning of a neighborhood.

The rules for constructing a Fresnel zone require that all rays passing through a given Fresnel zone have the same optical path length. The true path will pass through the center of the first Fresnel zone constructed on the actual wavefront. A neighborhood must be the area of the first Fresnel zone, for it is over this area that the optical path length does not change. Figure 18a shows the neighborhood, about the true optical path, as a cross-hatched region equal to the first Fresnel zone.

Light waves do travel over wrong paths but we do not observe them because the phase differences for those waves that travel over the 'wrong' paths are such that they destructively interfere. By moving the neighborhood, defined in the previous paragraph as the first Fresnel zone, to a region displaced from the true optical path, we can use the zone construction to see that this statement is correct. In Figure 18b the neighborhood is constructed about a ray that is improper according to Fermat's principle. We see that this region of space would contribute no energy at the observation point because of destructive interference between the large number of partial zones contained in the neighborhood.

This leads us to another interpretation of diffraction. If an obstruction is placed in the wavefront so that we block some of the normally interfering wavelets, we see light not predicted by Fermat's principle, i.e., we see diffraction.

Further Reading

Backer BB and Copson EJ (1969) *The Mathematical Theory of Huygens' Principle*. London: Oxford University Press.
Born M and Wolf E (1970) *Principles of Optics*. New York: Pergamon Press.
Guenther RD (1990) *Modern Optics*. New York: Wiley.
Haus HA (1984) *Waves and Fields in Optoelectronics*. Englewood Cliffs, NJ: Prentice-Hall.
Hecht E (1998) *Optics*, 3rd edn. Reading, MA: Addison-Wesley.
Klein MV (1970) *Optics*. New York: Wiley.
Rubinowicz A (1984) *Tahe Miyamoto–Wolf diffraction wave*. In: Wolf E (ed.) Progress in Optics, vol. IV.
Sayanagi I (1967) Pinhole Imagery. *Journal of the Optical Society of America*. 57: 1091–1099.
Sears FW (1949) *Optics*. Reading, MA: Addison-Wesley.
Sommerfeld A (1964) *Optics*. New York: Academic Press.
Stegliani DJ, Mittra R and Semonin RG (1967) Resolving power of a zone plate. *Journal of the Optical Society of America* 57: 610–613.
Stone JM (1963) *Radiation and Optics*. New York: McGraw-Hill.

DIFFRACTIVE SYSTEMS

Contents

Aberration Correction with Diffractive Elements
Applications of Diffractive and Micro-Optics in Lithography
Design and Fabrication of Diffractive Optical Elements
Diffractive Laser Resonators
Diffractives in Animals
Microstructure Fibers
Omnidirectional Surfaces and Fibers
Wave Optical Modeling and Design

Aberration Correction with Diffractive Elements

N Lindlein, University of Erlangen-Nürnberg, Erlangen, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Diffractive optical elements (DOEs) can be used for the correction of aberrations in optical systems because of two fundamental properties:

1. The phase function Φ of a DOE can be chosen quite arbitrarily so that for a given wavelength λ the DOE can act like an aspheric optical element.
2. The dispersion of a DOE due to the local grating equation is very high and its sign is opposite to that of the dispersion of a refractive optical element with normal dispersion of the refractive index.

The first property allows the correction of monochromatic aberrations in an optical system similar to the case of using an aspheric surface in a refractive optical system. The second property allows the construction of hybrid achromatic optical elements that consist of a refractive lens and a diffractive lens whose optical powers have the same sign (e.g., two positive lenses for a positive achromatic element or two negative lenses for a negative achromatic element).

This article will first describe two methods for the simulation of DOEs by ray tracing because ray tracing is the basic simulation tool for aberration correction. Afterwards, the properties of DOEs for the correction of monochromatic and chromatic aberrations and some applications will be discussed. The name DOE will always mean either binary, or multiple-step or blazed surface relief diffractive optical elements which change only the phase of the incident wave and not the amplitude. All examples will be given for transmission elements but in principle reflective elements can also be described with quite similar formulas.

Ray Tracing Simulation of DOEs

First of all it seems straightforward to use a diffractive analysis to simulate a diffractive optical element. However, the global diffractive analysis of a general DOE with perhaps 1000 or more grating structures and locally changing grating period and especially the diffractive analysis of a complete optical system containing DOEs and other elements is so complex that it can be done only for very simple DOEs and systems. Therefore, the diffractive analysis of a DOE is mostly restricted to a local analysis using some periods (e.g., 10–50 periods) to calculate, e.g., the different local diffraction orders, their diffraction efficiencies and possibly local phase shifts. For the global analysis another simulation method has to be used.

The most important simulation method for the correction of aberrations in traditional optical systems is ray tracing. Therefore, ray tracing methods are also used for DOEs. These methods can calculate the path of a ray for a given diffraction order, but they cannot calculate the diffraction efficiency for this ray. For the calculation of the diffraction efficiency local diffractive analysis methods (scalar or rigorous methods depending on the local grating frequency and the angle of the incident ray) have to be used

in addition. In particular, for applications with a broad wavelength spectrum (e.g., an achromatic lens) or a broad spectrum of angles of the incident rays, the diffraction efficiency has to be analyzed carefully to obtain the energy distribution in the different diffraction orders. But this is beyond the scope of this article about aberration correction and we assume that we know the local diffraction efficiency for each ray.

Sweatt Model for the Ray Tracing Simulation of DOEs

An aspheric refractive surface with height h as a function of the lateral coordinates x and y and a refractive index n below the surface and air (refractive index 1) above the surface can be approximately replaced by a DOE with the phase function (see Figure 1)

$$\Phi(x, y) = 2\pi \frac{(n-1)h(x, y)}{\lambda} \qquad [1]$$

Here, λ is the wavelength of the light for which the DOE is designed and the phase function is a continuous representation of the structure of the DOE. To encode the phase function of the DOE the wrapped phase function $\Phi \bmod 2\pi$ has to be considered. The encoding itself can be done by building a blazed, a multiple-step or a binary structure (see Figure 1), whereby the first-order diffraction efficiency decreases from the blazed to the binary structure (at maximum 100% diffraction efficiency for the blazed structure and about 40% for the binary structure). The replacement of an aspheric refracting surface by a DOE using eqn [1] is only valid in the paraxial approximation since the phase function of the DOE is encoded in a plane whereas the refracting surface of the aspheric is typically curved. But, the agreement between the aspheric surface and the DOE with a phase function calculated with the help of eqn [1] increases with increasing refractive index n since then the curvature of the aspheric is quite small and the refraction also nearly takes place in a plane. Of course, the refractive index of real materials cannot exceed a value of more than about 3.5 (for silicon used in the near infrared). But, in a numerical simulation the refractive index can be arbitrarily high, e.g., several thousand. This is the reason why the so called Sweatt model for tracing rays through DOEs uses a weakly curved aspheric surface with height h_{sweatt} and a very high refractive index n_{sweatt} to simulate a diffractive phase function Φ:

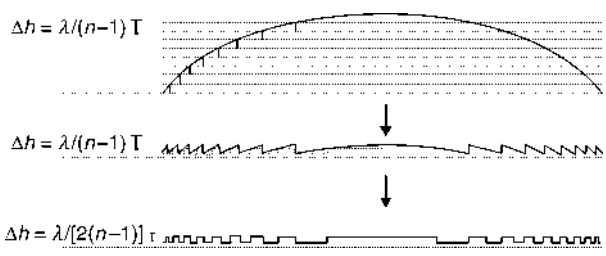

Figure 1 Transition from an aspheric refracting surface to a blazed diffractive optical element and finally to a binary diffractive optical element.

$$h_{\text{sweatt}}(x, y) = \frac{\lambda}{n_{\text{sweatt}} - 1} \frac{\Phi(x, y)}{2\pi} \qquad [2]$$

To take into account the wavelength dispersion of the DOE the refractive index n_{sweatt} in the Sweatt model has to be proportional to the wavelength λ

$$n_{\text{sweatt}}(\lambda) = n_0 \frac{\lambda}{\lambda_0} \qquad [3]$$

where n_0 is the refractive index at the wavelength λ_0.

The validity of the Sweatt model increases with increasing refractive index and becomes nearly exact for $n_{\text{sweatt}} \to \infty$ (in practice values of several thousand for the refractive index are normally sufficient and a rule of thumb is to take the numerical value of the utilized wavelength specified either in nanometers for infrared light or angstroms for visible light). The Sweatt model was introduced as a means to simulate DOEs in ray tracing programs unable to implement DOEs directly but capable of simulating aspheric refractive surfaces. However, a disadvantage of this model is the numerical error generated by the finite value of the refractive index.

Modern ray tracing programs therefore use a more appropriate model based on the local grating equation.

Local Grating Model for the Ray Tracing Simulation of DOEs

The basic equation for the simulation of DOEs by the local grating model is that the phase Φ_{out} of a diffracted ray in the plane of the DOE can be calculated from the phase Φ_{in} of the incident ray by adding the phase function Φ of the DOE multiplied by the diffraction order m:

$$\Phi_{\text{out}}(x, y, z_0) = \Phi_{\text{in}}(x, y, z_0) + m\Phi(x, y) \qquad [4]$$

Here, z_0 is the z-coordinate of the DOE plane which is assumed to be perpendicular to the z-axis (see Figure 2). The diffraction order m will be only one for an ideal blazed DOE but it can have an arbitrary integer value for binary DOEs. In practice, the optical designer will calculate the system for one diffraction order and analyze the system for other diffraction orders afterwards if necessary.

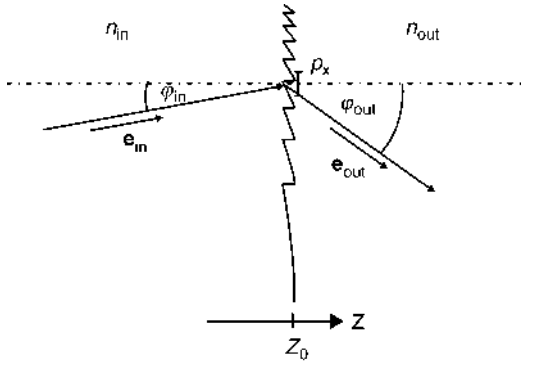

Figure 2 Schematic view in two-dimensions of the path of a ray which is diffracted at a DOE.

For the phases Φ_{in} of the incident and Φ_{out} of the diffracted ray the eikonal equation of geometrical optics is valid,

$$|\nabla\Phi_j|^2 = k_j^2 = \left(\frac{2\pi n_j}{\lambda}\right)^2 \qquad [5]$$

where the index j can have the two values $j \in \{\text{in}, \text{out}\}$. n_{in} or n_{out} are the refractive indices of the materials in which the incident or diffracted rays propagate. λ is the wavelength in vacuum. This form of the eikonal equation is obtained from the more common form $|\nabla L_j|^2 = n_j^2$, which is derived in any textbook of optics, by multiplying it with $(2\pi/\lambda)^2$ and using the definition of the optical phase $\Phi_j = 2\pi L_j/\lambda$. Here, L_j is the optical path length which is often called the eikonal. The curves of constant phase $\Phi_j(x, y, z) = \text{constant}$ are the wavefronts and the gradient $\nabla\Phi_j(x, y, z)$ points in the direction of the local ray which is perpendicular to the wavefront at the point (x, y, z) and describes its direction of propagation. Furthermore, the unit vectors $\mathbf{e}_{in}$ and $\mathbf{e}_{out}$ along the ray direction of the incident and diffracted ray, respectively, can be calculated by normalizing the gradient:

$$\mathbf{e}_j = \frac{\nabla\Phi_j}{|\nabla\Phi_j|} = \frac{\lambda}{2\pi n_j}\nabla\Phi_j\,; \qquad j \in (\text{in}, \text{out}) \qquad [6]$$

Since $\mathbf{e}_j$ is a unit vector, i.e., $e_{x,j}^2 + e_{y,j}^2 + e_{z,j}^2 = 1$, it can be seen from eqn [6] (and also from the original eqn [5]) that the z-component $e_{z,j}$ of the ray direction vector can be calculated (apart from its sign) if the two partial derivatives of the phase with respect to x and y are known. This means that it is sufficient to know the phase in a plane as stated in eqn [4].

By calculating the partial derivatives of eqn [4] with respect to x and y the equations for the ray tracing at a DOE are derived:

$$\begin{aligned}
\frac{\partial\Phi_{out}(x,y)}{\partial x} &= \frac{\partial\Phi_{in}(x,y)}{\partial x} + m\frac{\partial\Phi(x,y)}{\partial x} \\
&\Rightarrow n_{out}e_{x,out}(x,y) \\
&= n_{in}e_{x,in}(x,y) + m\frac{\lambda}{2\pi}\frac{\partial\Phi(x,y)}{\partial x} \\
\frac{\partial\Phi_{out}(x,y)}{\partial y} &= \frac{\partial\Phi_{in}(x,y)}{\partial y} + m\frac{\partial\Phi(x,y)}{\partial y} \\
&\Rightarrow n_{out}e_{y,out}(x,y) \\
&= n_{in}e_{y,in}(x,y) + m\frac{\lambda}{2\pi}\frac{\partial\Phi(x,y)}{\partial y}
\end{aligned} \qquad [7]$$

The local grating frequencies ν_x and ν_y in x and y direction are calculated from the phase function Φ by:

$$\begin{aligned}
\nu_x(x,y) &= \frac{1}{2\pi}\frac{\partial\Phi(x,y)}{\partial x} \\
\nu_y(x,y) &= \frac{1}{2\pi}\frac{\partial\Phi(x,y)}{\partial y}
\end{aligned} \qquad [8]$$

Therefore, given $\mathbf{e}_{in}$ and eqn [6], the ray direction vector $\mathbf{e}_{out}$ of the diffracted ray is

$$n_{out}\mathbf{e}_{out} = n_{out}\begin{pmatrix} e_{x,out} \\ e_{y,out} \\ e_{z,out} \end{pmatrix} = \begin{pmatrix} n_{in}e_{x,in} + m\lambda\nu_x \\ n_{in}e_{y,in} + m\lambda\nu_y \\ \sqrt{n_{out}^2 - (n_{in}e_{x,in} + m\lambda\nu_x)^2 - (n_{in}e_{y,in} + m\lambda\nu_y)^2} \end{pmatrix} \qquad [9]$$

The dependence of the lateral coordinates x and y is left out in eqn [9] to simplify notation. It can be easily checked that this equation fulfills the required condition for the unit vector: $e_{x,out}^2 + e_{y,out}^2 + e_{z,out}^2 = 1$.

As stated above, for a blazed DOE fabricated to the exact depth and illuminated by the design wavelength only the first diffraction order $m = 1$ is generated. But in practice other diffraction orders m are generated due to fabrication errors or illuminating with other wavelengths. In particular, by using binary DOEs there are inherently different diffraction orders. Therefore, it may be necessary to analyze the optical system for several values m one after the other. Each value m defines one wavefront generated by the DOE and the wavefront can be calculated by evaluating the optical path lengths along the rays at different points. In practice, the optical system has to be analyzed for those values m with sufficiently high

diffraction efficiency. Then, it can be checked whether undesired diffraction orders of the DOE are blocked in the optical system by aperture stops or whether they will disturb the output of the optical system. The local diffraction efficiency of each ray for a value m can be calculated to a good approximation by using a diffractive analysis of the corresponding local grating, i.e., that of an infinitely extended grating which would have the same grating period and the same structure shape (e.g., depth) as the DOE at the specific point.

If the y component of the incident ray and the local grating frequency ν_y are both zero the ray components in the x direction can be expressed via the sine of the angle of incidence φ_{in} or the angle of the diffracted ray φ_{out} (see Figure 2). In this case the first row of eqn [9] can be written as

$$n_{\text{out}} \sin \varphi_{\text{out}} = n_{\text{in}} \sin \varphi_{\text{in}} + m\frac{\lambda}{p_x} \qquad [10]$$

where the local grating period $p_x = 1/\nu_x$ has been introduced. This is the well-known grating equation.

In this article eqns [4]–[10] assume that the DOE is on a planar substrate and that x and y are the lateral coordinates on the substrate. However, the equations can also be written in a vector formulation which is valid for DOEs on curved substrates (see Further Reading).

So, we have now a powerful simulation method to design and analyze in the following several optical systems containing DOEs to correct monochromatic or chromatic aberrations.

Correction of Monochromatic Aberrations

For practical applications it is important to have an efficient algorithm to calculate the phase function of a DOE to correct the aberrations in a given optical system. Several different cases have to be distinguished. An easy case is the design of a DOE to produce a desired output wavefront Φ_{out} from a known input Φ_{in}. In this case, eqn [4] can be used to calculate the phase function Φ of the DOE directly. Two examples are given in the following. Afterwards, some aspects of a more general numerical optimization procedure are described.

Aberration Correction for Interferometrically Recorded Holograms

Although the following example of the aberration correction of interferometrically recorded holograms by using DOEs seems to be quite special it shows exemplarily several principal steps which are necessary to design a DOE for correcting monochromatic aberrations.

In this article the name DOE is used for diffractive elements with a microstructured surface (binary, multistep or blazed). Such elements have the property that their diffraction efficiency depends only weakly on the wavelength or the angle of incidence of the illuminating light. Another interesting class of diffractive elements are holograms in volume media, which are recorded interferometrically using two coherent recording waves. The refractive index in a volume hologram is modulated periodically and the thickness has to amount to several periods of the interference pattern along the z-direction in order to have a so-called thick hologram with a high diffraction efficiency. If there are only a few periods along the z-direction the quite small modulation of the refractive index (typically about 0.02) will result in a low diffraction efficiency. Therefore, if the angle between the two recording waves is small the resulting interference pattern will have large periods and only a few periods are recorded in the photosensitive layer along the z-direction. But this means that the holographic optical element (HOE) can only have a small diffraction efficiency. From this fact, it is clear that also by using the recorded HOE the deflection angle between the incident and the diffracted wave must not be smaller than a certain value because otherwise the thickness has to be very big or the diffraction efficiency decreases. Therefore, such elements with high diffraction efficiency, which will be named in the following holographic optical elements (HOE), are always off-axis elements because for on-axis elements all rays in the neighborhood of the axis are only deflected by a small angle and then the diffraction efficiency would be low. For the correct wavelength and angle of incidence, so that the Bragg condition of holography is fulfilled, HOEs, like blazed DOEs, exhibit only one dominant diffraction order $m = m_{\text{dom}}$ (normally $m_{\text{dom}} = 1$). But, if the angle of incidence or the wavelength change, the diffraction efficiency in the order m_{dom} decreases rapidly and light passes through the HOE without being deflected, i.e., most of the light is then in the order $m = 0$. For this reason, a HOE can be used for example for a head-up display to fade in some information under the Bragg angle without hindering the normal view onto the environment.

Traditional materials for HOEs like dichromated gelatin (DCG) are sensitive only in the blue–green spectral region (e.g., for light of an argon ion laser with $\lambda = 488$ nm). However, once recorded it will be illuminated in many applications (e.g., optical interconnects) by a wavelength in the red–infrared spectral region. Unfortunately, the wavelength mismatch between recording and reconstruction illumination introduces aberrations.

If the two desired reconstruction waves at the wavelength λ_2 have the phases $\Phi_{r,\lambda 2}$ and $\Phi_{o,\lambda 2}$ in the surface plane of the HOE (the subscripts r and o refer to interferometry reference and object, respectively), the phase function Φ of the HOE which has to be recorded is

$$\Phi(x, y) = \Phi_{o,\lambda 2}(x, y) - \Phi_{r,\lambda 2}(x, y) \qquad [11]$$

If our discussion here is restricted to holographic lenses, the two reconstruction waves will be spherical (or plane) waves (see Figure 3b) with the center of curvature at the points $\mathbf{r}_{r,\lambda 2}$ and $\mathbf{r}_{o,\lambda 2}$ and eqn [11] is:

$$\Phi(x,\ y) = \frac{2\pi\sigma_{o,\lambda 2}}{\lambda_2}\sqrt{(x - x_{o,\lambda 2})^2 + (y - y_{o,\lambda 2})^2 + z_{o,\lambda 2}^2} - \frac{2\pi\sigma_{r,\lambda 2}}{\lambda_2}\sqrt{(x - x_{r,\lambda 2})^2 + (y - y_{r,\lambda 2})^2 + z_{r,\lambda 2}^2} \qquad [12]$$

The parameters $\sigma_{o,\lambda 2}$ and $\sigma_{r,\lambda 2}$ are either $+1$ for a divergent spherical wave or -1 for a convergent spherical wave. The surface of the HOE itself is in the plane at $z = 0$.

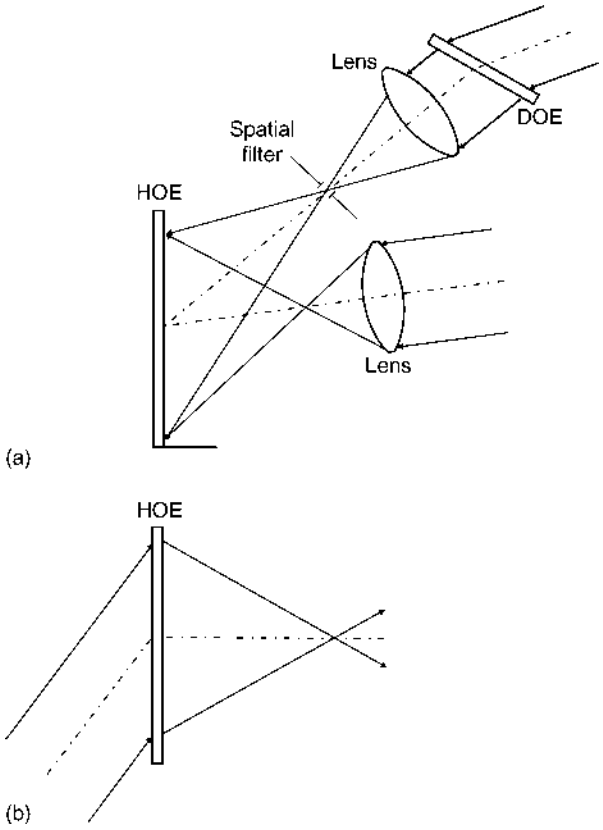

Figure 3 Schematic view of the (a) recording and (b) reconstruction setup for a holographic lens. The recording is made at wavelength λ_1 and the reconstruction at wavelength λ_2.

The task is now to find two recording waves which construct nearly the same three-dimensional interference pattern in the hologram at the recording wavelength λ_1. This means that eqn [12] has to be fulfilled exactly for the difference of the phases of the two recording waves in the plane $z = 0$ and, additionally, the Bragg condition has to be fulfilled at least approximately. So, the next step is to calculate one of the recording waves with phase $\Phi_{r,\lambda 1}$ so that the Bragg condition is fulfilled. This can be done approximately with a third-order theory resulting in a spherical wave or nearly exactly with an iterative numerical procedure resulting generally in an aspheric recording wave. If $\Phi_{r,\lambda 1}$ is known, the other recording wave with phase $\Phi_{o,\lambda 1}$ is calculated by

$$\Phi_{o,\lambda 1}(x, y) = \Phi(x, y) + \Phi_{r,\lambda 1}(x, y) \qquad [13]$$

At least this recording wave will be aspheric. To generate such an aspheric wave it is traced by ray tracing to a plane where the two recording waves are spatially separated. There, either a DOE or a combination of a DOE and a lens are used to generate the aspheric wave. The calculations of the phase function of the DOE are done again by using an equation like [11] whereby one wave is the propagated aspheric wave (possibly propagated also through a lens) and the other is normally a plane off-axis wave with which the DOE will be illuminated in the real recording setup (see Figure 3a). Here, the off-axis angle is necessary to allow spatial filtering of disturbing diffraction orders of the DOE which will be normally fabricated as a binary DOE.

Experiments of our own showed (see the literature) that nearly diffraction-limited holographic lenses with a numerical aperture of NA = 0.4 can be fabricated by correcting the aberrations of about 28 wavelengths peak-to-valley (for the reconstruction at 633 nm wavelength) due to the wavelength mismatch between recording at 488 nm wavelength and reconstruction at 633 nm wavelength by using a DOE in the recording setup.

Testing of Aspheric Surfaces by Using a DOE as Aberration Compensating Element

An important application for DOEs is testing aspheric surfaces and nowadays more and more companies use DOEs to test aspheric surfaces which are used for example in the objectives of lithographic wafer steppers to produce highly integrated microelectronic circuits.

In this testing setup the DOE is illuminated by a wave with phase Φ_{in} to generate a wave with phase Φ_{out} that is incident perpendicular to the aspheric surface (see Figure 4). After reflection at the asphere

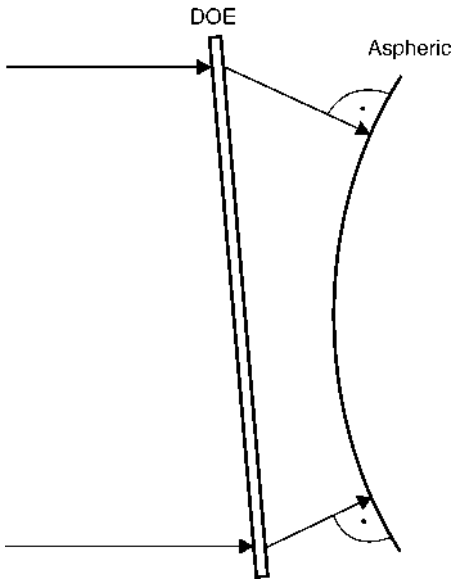

Figure 4 Scheme showing the basic elements for testing an aspheric surface.

the wave is retro-reflected and takes the same path if the asphere has no surface deviations and no misalignments. This means that the setup is completely symmetric for a reflection at the aspheric surface. Therefore, the optical path lengths have to be identical for all rays on the asphere. To calculate the phase function of the DOE the following steps are made:

- A grid of N rays starting perpendicularly on the aspheric surface and having all phase zero (due to the fact that the phase has to be identical it can be set to zero) is traced to the plane of the DOE. N depends on the shape and numerical aperture of the asphere and the grid will be in most cases two dimensional and regular or one dimensional (along one radial section of the asphere) and regular if the asphere and the system are completely rotationally symmetric. As a rule of thumb the number N of rays has to be more than twice the number of polynomial coefficients if the asphere is described by a polynomial.
- The phase of these rays is the function $-\Phi_{\text{out}}$ (the negative sign is necessary because the rays take the opposite direction) and is now known at some points (x_i, y_i) with $i \in \{1, 2, \ldots, N\}$.
- The phase Φ_{in} of the incident wave which is in practice either a plane wave or a spherical wave is calculated in the plane of the DOE. If the incident wave is generated by a lens having aberrations it is again calculated by ray tracing, otherwise it can be calculated analytically.
- Now, the phase function Φ of the DOE can in principle be calculated with the help of eqn [4].

However, the phase Φ_{out} (and possibly also Φ_{in}) is known only at some points (x_i, y_i), but it has to be known at all points of the DOE. Therefore in practice an analytical representation of the phase functions of the incident and output wave has to be calculated. This can be done by fitting a polynomial to the phase values using a least squares fit. In this case the coordinates, at which the phase values are known, can be quite arbitrary as long as the sampling is dense enough. For a rotationally symmetric system the polynomial should also be rotationally symmetric, i.e., a polynomial P_{rot} depending only on the radius coordinate r

$$r = \sqrt{x^2 + y^2} \qquad [14]$$

$$P_{\text{rot}}(r) = \sum_{j=0}^{G} a_j r^j \qquad \text{or} \qquad P_{\text{rot}}(r) = \sum_{j=0}^{G/2} a_{2j} r^{2j} \qquad [15]$$

where G is the degree of the polynomial, i.e., the highest power in r. In these two cases the number N of sampling points lying along one radial section of the rotationally symmetric optical system should be larger than $2G$ or G, respectively.

For general non-symmetric systems a Cartesian polynomial P_{cart} can be used, e.g.,

$$P_{\text{cart}}(x, y) = \sum_{j=0}^{G} \sum_{k=0}^{G-j} a_{jk} x^j y^k \qquad [16]$$

Here, the number of coefficients of the polynomial is $(G + 1)(G + 2)/2$ and the number N of sampling points should again be larger than twice the number of coefficients, i.e., $N > (G + 1)(G + 2)$.

For both types of polynomials the coefficients a_j or a_{jk} have to be calculated so that the merit function M

$$M(\mathbf{a}) = \sum_{i=1}^{N} (P(x_i, y_i) - \Phi_j(x_i, y_i))^2; \quad j \in \{\text{in, out}\} \qquad [17]$$

is minimized. The symbol P means either P_{rot} or P_{cart}. The vector $\mathbf{a}$ symbolizes the coefficients $\mathbf{a}_j$ or $\mathbf{a}_{jk}$. By differentiating M with respect to all coefficients and setting the partial derivatives to zero a system of linear equations is obtained which can be solved by standard procedures. As mentioned before, the number N of independent sampling points has to be at least double the size of the number of coefficients of the polynomial. Otherwise, the polynomial can 'swing' between the sampling points and the polynomial fits well to the desired phase function only at the sampling points but not at other points in between.

For very irregular aspherics with a fast change of the local curvature it might by necessary not to take a

global polynomial fit but a spline fit, i.e., a set of adjacent local polynomials with several continuity conditions at the border area of each local polynomial. In this case the total number of coefficients used to describe the phase functions and Φ_{in} and Φ_{out} of the DOE will be increased compared to the description by a global polynomial. Then, the number N of rays for the calculation has also to be increased accordingly.

So, in this example the DOE acts as a null-compensator for the asphere and at first sight it is not a typical example for aberration correction. However, by regarding the complete system the DOE compensates the aberrations, i.e., deviations from a spherical wave, generated by the asphere which are primary and higher-order spherical aberrations if the asphere is rotationally symmetric and more general aberrations otherwise. This example shows how the phase function of a DOE can be designed straightforwardly for correcting aberrations if there is only one well-defined wavefront propagating through the optical system. This design method can also be used for other optical systems where the input and the output wavefront of the DOE are well defined.

Numerical Optimization of DOEs in Optical Systems

In a general optical system, where a DOE is not only used to compensate the aberrations for one specific incident wave but for different incident waves (optical system with a field), it is not possible to calculate the optimum phase function in such a straightforward way as given above. However, a DOE is in principle similar to an asphere in an optical system, so that the design rules for including aspherics in optical systems can be used in some cases as long as monochromatic aberrations are considered. Nevertheless, one should keep in mind that a DOE is, according to the Sweatt model, an asphere with high refractive index and small height, i.e., the deflection of the light rays occurs in a plane (or on a curved surface if the DOE is situated on a curved substrate). Therefore, the design rules for aspheric surfaces can be used only if the high refractive index of the Sweatt model is used in the design.

In the general case only a numerical optimization can be used. For this, it is very important to select an appropriate mathematical representation of the phase function of the DOE.

For rotationally symmetric optical systems also the phase function of the DOE should be rotationally symmetric. Therefore, in such a case it is useful to write the phase function Φ as a polynomial in the radius coordinate r as given in eqns [14] and [15]. Then, the $(G + 1)$ or $(G/2 + 1)$ coefficients a_j have to be optimized, whereby a_2 is proportional to the focal power of the DOE, whereas the other coefficients represent aspheric terms. Depending on the numerical aperture and on the application the degree G of the polynomial can be increased in order to achieve a better solution. Nevertheless, the degree in the optimization should be as small as necessary in order to avoid 'swinging effects' and numerical instabilities of the polynomial. To determine the smallest adequate degree an iterative method has to be used by starting with a quite small degree (e.g., 4). If the result of the optimization is not good enough the degree has to be increased until the result of the optimization is satisfying.

For non-rotationally symmetric optical systems a Cartesian polynomial like in eqn [16] can be taken and all the $(G + 2)(G + 1)/2$ coefficients a_{jk} have to be optimized. Since the number of free parameters is in this case larger than the number of free parameters in the rotationally symmetric case the convergence of the optimization may take a long time and the risk to break down in a local minimum of the multidimensional parameter space increases.

In principle, the calculation of a DOE used, e.g., for testing an aspheric surface can also be done by optimizing the coefficients of the polynomial representing the phase function of the DOE. Nevertheless, as a general rule it can be said that a direct calculation of the phase function is always more efficient than a numerical optimization, if the direct calculation is possible. In practice, it can be seen that the maximum degree of a polynomial, which is directly calculated by the method described in the last section, is always smaller than the degree which is necessary in a numerical optimization to achieve convergence of the optimization.

Of course, in such cases where several different types of aberrations (on-axis aberrations like spherical aberration, off-axis aberrations like astigmatism and coma or field aberrations like field curvature or distortion) have to be corrected in an optical system a direct calculation of the phase function is not possible so that only the numerical optimization can provide a solution.

Up to now only monochromatic aberrations have been discussed. But as stated at the beginning of this article the strong dispersion of a DOE having opposite sign to the dispersion of a refractive element offers additional applications of DOEs by correcting chromatic aberrations. The principle of doing this is described in the following.

Correction of Chromatic Aberrations

The focal power $1/f$ (f is the focal length of the lens) of a thin refractive lens consisting of two spherical

surfaces with radii of curvature R_1 and R_2 and made of a material (glass) with refractive index n is

$$\frac{1}{f(\lambda)} = (n(\lambda) - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad [18]$$

where λ is the wavelength of the illuminating light.

This equation is valid for all wavelengths and so also for the commonly used wavelength $\lambda_d = 587.6$ nm (yellow d-line of helium) which is near the center of the visible spectrum. Then $1/R_1 - 1/R_2$ can be expressed as $1/((n_\mathrm{d} - 1)f_\mathrm{d})$ and the focal power depends on the wavelength as

$$\frac{1}{f(\lambda)} = \frac{n(\lambda) - 1}{n_\mathrm{d} - 1}\frac{1}{f_\mathrm{d}} \qquad [19]$$

where n_d and f_d are the refractive index and the focal length at $\lambda_\mathrm{d} = 587.6$ nm, respectively. So, by combining two thin lenses made of materials with refractive indices n_1 and n_2 and focal lengths f_1 and f_2 the focal power of the doublet assuming a zero distance between the two lenses is:

$$\begin{aligned}\frac{1}{f_\mathrm{ges}(\lambda)} &= \frac{1}{f_1(\lambda)} + \frac{1}{f_2(\lambda)} \\ &= \frac{n_1(\lambda) - 1}{n_{1,\mathrm{d}} - 1}\frac{1}{f_{1,\mathrm{d}}} + \frac{n_2(\lambda) - 1}{n_{2,\mathrm{d}} - 1}\frac{1}{f_{2,\mathrm{d}}} \qquad [20]\end{aligned}$$

For an achromatic doublet the focal power should be equal for two different wavelengths which are commonly the wavelengths $\lambda_\mathrm{F} = 486.1$ nm (blue F-line of hydrogen) and $\lambda_\mathrm{C} = 656.3$ nm (red C-line of hydrogen) which mark the limits of the well visible spectrum (of course the visible spectrum itself is broader but the sensitivity of the human eye to wavelengths below λ_F or beyond λ_C is quite bad). Therefore, the following equations result:

$$\begin{aligned}\frac{1}{f_\mathrm{ges}(\lambda_\mathrm{F})} = \frac{1}{f_\mathrm{ges}(\lambda_\mathrm{C})} &\Rightarrow \frac{n_1(\lambda_\mathrm{F}) - n_1(\lambda_\mathrm{C})}{n_{1,\mathrm{d}} - 1}\frac{1}{f_{1,\mathrm{d}}} \\ &= -\frac{n_2(\lambda_\mathrm{F}) - n_2(\lambda_\mathrm{C})}{n_{2,\mathrm{d}} - 1}\frac{1}{f_{2,\mathrm{d}}} \qquad [21]\end{aligned}$$

By introducing the Abbe number V_d

$$\begin{aligned}V_\mathrm{d} &= \frac{n_\mathrm{d} - 1}{n_\mathrm{F} - n_\mathrm{C}} \\ &= \frac{n(\lambda_\mathrm{d} = 587.6\ \mathrm{nm}) - 1}{n(\lambda_\mathrm{F} = 486.1\ \mathrm{nm}) - n(\lambda_\mathrm{C} = 656.3\ \mathrm{nm})} \qquad [22]\end{aligned}$$

equation [21] can be simplified to

$$V_{1,\mathrm{d}} f_{1,\mathrm{d}} = -V_{2,\mathrm{d}} f_{2,\mathrm{d}} \qquad [23]$$

Now, the wavelength dispersion of all conventional materials (glasses, polymers, etc.) is such that the refractive index decreases with increasing wavelength, i.e. the material has a so-called normal dispersion. This means that the Abbe number V_d is always positive and ranges from about 20 for high-dispersive glasses (e.g., SF59) to more than about 80 for low-dispersive glasses (e.g. Ultran 20). So, to fabricate a refractive achromatic doublet with a positive focal power it is necessary to take a positive lens with focal length f_1 made of a low-dispersive glass with a large Abbe number $V_{1,\mathrm{d}}$ and a negative lens with focal length f_2 made of a high-dispersive glass with a small Abbe number $V_{2,\mathrm{d}}$ (see Figure 5a). Therefore, the resulting focal power of the achromatic doublet

$$\frac{1}{f_{\mathrm{ges,d}}} = \frac{1}{f_{1,\mathrm{d}}} + \frac{1}{f_{2,\mathrm{d}}} = \frac{1}{f_{1,\mathrm{d}}}\left[1 - \frac{V_{2,\mathrm{d}}}{V_{1,\mathrm{d}}}\right] \qquad [24]$$

is always smaller than the focal power of the positive lens with focal power $1/f_{1,\mathrm{d}}$ alone.

The situation is completely different if a refractive and a diffractive lens are used. From the Sweatt model

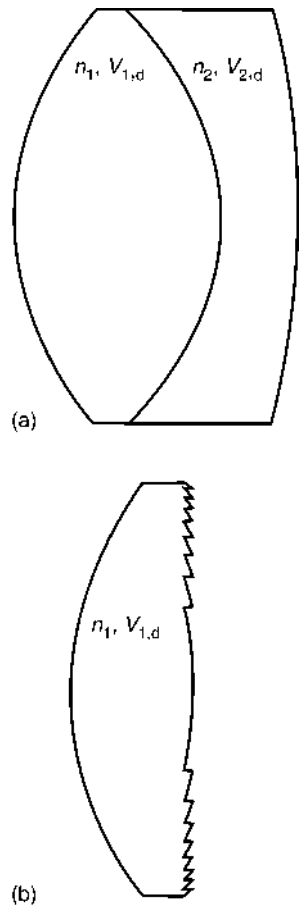

Figure 5 Structure of achromats: (a) refractive achromatic doublet consisting of a positive lens with low dispersion and a negative lens with high dispersion; (b) hybrid achromatic lens consisting of a positive refractive lens and a positive diffractive lens.

of a DOE (eqn [3]) it is known that the refractive index n_{sweatt} is proportional to the wavelength λ

$$n_{\text{sweatt}}(\lambda) = \frac{n_0}{\lambda_0}\lambda = \frac{n_d}{\lambda_d}\lambda \qquad [25]$$

Therefore, the Abbe number of a DOE is

$$\begin{aligned} V_d &= \frac{n_d - 1}{n_F - n_C} \\ &= \frac{\lambda_d}{\lambda_F - \lambda_C}\left(1 - \frac{1}{n_d}\right) \overset{n_d \to \infty}{=} \frac{\lambda_d}{\lambda_F - \lambda_C} \\ &= -3.452 \end{aligned} \qquad [26]$$

Here, we used the fact that the accuracy of the Sweatt model increases with increasing refractive index so that the Abbe number of a DOE is achieved in the limit of an infinite refractive index. So, it can be seen that the Abbe number of a DOE is a small negative constant. This means that the dispersion of a DOE is high and opposite in sign to that of a refractive element. Therefore, according to eqns [23] and [26] a positive refractive lens with a large focal power (i.e., small focal length) and a positive diffractive lens with a small focal power (i.e., long focal length) can be combined to form an achromatic doublet with a focal power larger than the focal power of one of the two lenses alone (see Figure 5b). By using eqns [19] and [25] the focal power of a diffractive lens can be calculated with the Sweatt model (n very big $\rightarrow n - 1 \approx n$):

$$\frac{1}{f_{\text{DOE}}(\lambda)} = \frac{\lambda}{\lambda_d f_d} \qquad [27]$$

In practice, the lenses are neither thin nor is the distance between them zero. In this case the equations for the design of a hybrid achromatic doublet are more complex (see Further Reading). Nevertheless, the given equations can be used for a first design and to understand the principal behavior.

Another advantage of a hybrid achromatic doublet is that the DOE can in addition behave like an asphere. Therefore, e.g., the spherical aberration can be corrected with the aspheric terms of the DOE provided that the spherical aberration is considerably smaller than the phase term encoded in the DOE to correct the chromatic aberrations. Otherwise, the correction of the monochromatic (spherical) aberration will disturb the correction of the chromatic aberrations. A comparison of a simple hybrid achromat, consisting of a planoconvex lens with the diffractive structure made directly onto the plane surface of the refractive lens and a conventional refractive achromatic doublet shows that the optical performance of the hybrid achromat is a little bit worse than that of the refractive achromatic doublet regarding, for example, off-axis aberrations. A simple explanation of this is that a refractive achromatic doublet has three spherical surfaces and two glasses, i.e., in total five parameters, whereas a hybrid achromat has one spherical surface, one glass and the DOE with a paraxial term and an aspheric term, i.e. in total about four parameters. Of course, the merit of these different types of parameters is not identical so that this assessment is only qualitative.

In the case of hybrid achromatic elements for a broad wavelength range (e.g., for the whole visible wavelength range) the dependence of the diffraction efficiency on the wavelength may be a limiting factor because the other diffraction orders behave like stray light and reduce the maximum contrast of the image. Therefore, special laminated blazed diffractive optical elements made of laminae of two or more different materials are described in the literature which show a nearly achromatic behavior of the diffraction efficiency but the 'normal' dispersion of DOEs. DOEs are in the meantime used in zoom lenses (see the cited patent in Further Reading) to correct chromatic aberrations.

Summary

In this article the possibilities of correcting monochromatic and chromatic aberrations in optical systems by using diffractive optical elements have been discussed. There are several advantages and disadvantages of using DOEs.

Advantages:

- DOEs offer many degrees of freedom in the design. By using modern lithographic fabrication techniques the fabrication of a DOE with a strongly aspheric phase function is just as expensive as fabricating a simple Fresnel zone lens with a quadratic phase function. So, DOEs can be used in some cases instead of aspheric surfaces which are more expensive. In addition, the lithographic fabrication technique guarantees a quite good accuracy concerning the phase function of the DOE compared to the more difficult fabrication of well-defined aspheric surfaces.
- The strong chromatic dispersion with the opposite sign to that of a refractive element allows the correction of chromatic aberrations. DOEs also give in this case an additional degree of freedom in the design because aspheric terms of the phase function can correct in addition monochromatic aberrations of the system as long as these aberrations are considerably smaller than the quadratic term of the phase function which corrects the chromatic aberrations.

Disadvantages:

- DOEs often exhibit not only one diffraction order but several diffraction orders due to fabrication errors, a broad spectrum of application wavelengths or because some DOEs naturally have several diffraction orders like, for example, binary DOEs. This can produce disturbing light in the optical system and limit the application.
- The strong dispersion can restrict the application in monochromatic systems because it requires in some cases (e.g., for testing aspheric surfaces) not only a relative constancy, i.e., constant but the absolute value does not have to be hit exactly, of the illuminating wavelength but an absolute constancy because otherwise systematic aberrations are introduced by the DOE.

So, there are many advantages of DOEs compared to other optical elements but on the other hand a DOE is not a magic bullet in correcting aberrations in optical systems. Depending on the system a careful assessment of the advantages and disadvantages has to be performed in order to find the best solution.

List of Units and Nomenclature

Abbe number	real number	V_d
Angle	[rad]	φ
Cartesian coordinates	[m]	x, y, z
Coefficients of a polynomial	different negative powers of m	a
Components of the ray vector	real numbers	e_x, e_y, e_z
Degree of a polynomial	integer number	G
Diffraction order	integer number	m
Focal length	[m]	f
Grating frequencies in x and y direction	[m^{-1}]	ν_x, ν_y
Height of a lens surface	[m]	h
Local grating periods in x and y direction	[m]	p_x, p_y
Merit function	real valued function	M
Modulus of the wave vector	[m^{-1}]	k
Nabla differential operator	[m^{-1}]	∇ $\begin{pmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{pmatrix}$
Number of sampling points/rays	integer number	N
Optical phase, phase function	[rad]	Φ
Polynomial	real valued function	P
Radial coordinate	[m]	r
Refractive index	real number	n
Sign variable	either +1 or −1	σ
Symbol for partial differentiation		∂
Unit vector in ray direction	vector with three real components	$\mathbf{e}$
Wavelength	[m]	λ

See also

Geometrical Optics: Aberrations.

Further Reading

Born M and Wolf E (1997) *Principles of Optics*, 6th edn. Cambridge, UK: Cambridge University Press.

Fairchild RC and Fienup JR (1982) Computer-originated aspheric holographic optical elements. *Optical Engineering* 21: 133–140.

Falkenstörfer O, Lindlein N, Keinonen T and Schwider J (1996) Diffraction-limited holographic lenses in dichromated gelatine. *Optical Engineering* 35: 2026–2033.

Herzig HP (1997) *Micro-Optics*. London: Taylor & Francis.

Kogelnik H (1969) Coupled wave theory for thick hologram gratings. *The Bell System Technology Journal* 48: 2909–2947.

Kress B and Meyrueis P (2000) *Digital Diffractive Optics*. Chicester, UK: Wiley.

Latta JN (1971) Computer-based analysis of holography using ray tracing. *Applied Optics* 10: 2698–2710.

Lindlein N, Schwider J, Schrader M, *et al.* (1993) Wavelength mismatch correcting elements for volume holograms written on a laser pattern generator. *Journal of Modern Optics* 40: 647–661.

MacGovern AJ and Wyant JC (1971) Computer generated holograms for testing optical elements. *Applied Optics* 10: 619–624.

Meier RW (1965) Magnification and third-order aberrations in holography. *Journal of the Optical Society of America* 55: 987–992.

Nishio A (1999) Zoom lens with diffractive optical elements. United States Patent 5978153.

Offner A (1966) Ray tracing through a holographic system. *Journal of the Optical Society of America* 56: 1509–1512.

Sinzinger S and Jahns J (1999) *Microoptics*. Weinheim: Wiley-VCH.

Stone Th and George N (1988) Hybrid diffractive-refractive lenses and achromats. *Applied Optics* 27: 2960–2971.
Sweatt WC (1977) Describing holographic optical elements as lenses. *Journal of the Optical Society of America* 67: 803–808.
Upatnieks J, Vander Lugt A and Leith E (1966) Correction of lens aberrations by means of holograms. *Applied Optics* 5: 589–593.
Welford WT (1975) A vector raytracing equation for hologram lenses of arbitrary shape. *Optical Communications* 14: 322–323.
Winick K (1982) Designing efficient aberration-free holographic lenses in the presence of a construction-reconstruction wavelength shift. *Journal of the Optical Society of America* 72: 143–148.

Applications of Diffractive and Micro-Optics in Lithography

V Kettunen, Heptagon, Zurich, Switzerland

H P Herzig, University of Neuchâtel, Neuchâtel, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Most aspects of modern life are thoroughly influenced by the micro-electronics industry through microchips, such as microprocessors and memory chips. Optical lithography has been a key factor in the rapid pace of integration that drives the development of the micro-electronics industry. The photolithography process not only facilitates the fabrication of the chips, but also directly influences the limits of the obtainable chip performance in terms of speed or storage capacity, through the minimum feature size that can be created in the lithography steps. The increasing resolution demands of the IC industry, outlined in the International Technology Roadmap for Semiconductors as technology nodes, mean continuing demand for improvements in the photolithography systems.

The challenges to the improvements are encapsulated in the equation for the resolution of an optical lithography system:

$$R = k_1 \lambda / NA \qquad [1]$$

which expresses the resolution R in terms of the smallest resolvable half-pitch as a function of the wavelength λ, the numerical aperture NA of the exposure system and the unit less constant k_1, which is used as a measure of lithographic difficulty, with smaller values indicating a more challenging process. Traditionally, the increase in the resolution has been obtained by decreasing the used wavelength, with the current mark for volume production systems currently standing at 193 nm. Within each wavelength, further reduction has been traditionally obtained by increasing the numerical aperture towards the mathematical limit of 1. Finally, several so-called Resolution Enhancement Techniques (RET) have been introduced to reduce k_1, in order to reach subwavelength technology nodes within each lithography wavelength generation. It is in connection with these techniques that micro-optics has found a role in modern lithography systems.

In this chapter we discuss the ways micro-optics are being utilized in lithography systems. We will start with a brief overview of conventional lithography systems, of the essential parts of such systems as well as the special resolution enhancement techniques utilized in the systems. We will then discuss the utilization of micro-optics within the context of lithography systems, along with the unique challenges in terms of design and fabrication that are associated with such applications. Finally, we will briefly discuss nonconventional lithography applications, especially those based on utilization of micro-optics. It should be noted that the motivation of this article is to describe the issues related to the use of micro-optical elements in a modern lithography system and is not intended as a review of the modern lithography systems, nor is it a review of micro-optics.

Overview of Conventional Lithography Systems

Photolithography is a process whereby a specified pattern or mask is transferred onto a wafer using photons. In the process a photosensitive material, commonly known as a resist, is initially used to record the pattern that is generated when the mask is illuminated. After exposure, the resist is developed. The resist is referred to as positive or negative, depending on whether it is removed or remains, respectively, after the development of the irradiated regions. After its development, the pattern in the resist is then transferred onto the wafer using suitable chemical processes, such as wet chemical etching or dry plasma etching.

Several different lithographical techniques exist. The oldest and simplest is contact printing, where a

photomask is brought into contact with a layer of photoresist on the wafer. Although this technique can reproduce high-resolution features, the contact between the wafer and the mask can lead to problems with cross contamination, and can damage either the wafer or the mask. Contact printing is mainly used nowadays in laboratory environments for small series of photolithography steps. A slightly more complicated technique is proximity printing, in which the photomask is held in close proximity to the surface of the wafer without actual contact. This avoids the problems of contamination and damage, but due to diffraction effects, reduces the maximum resolution of the process. Nevertheless, proximity printing remains a valuable technique for lithography applications, that do not require the highest possible resolution.

Modern volume production photolithography systems are based on the use of a projection system to image a mask onto a wafer through a complex system of lenses. There exist several variations of the basic idea. In some systems the entire mask is imaged at once onto the wafer. This approach works only when the feature size on the mask is reasonably large, in the range of a few microns. Reproducing smaller features with this approach is generally not possible, as the demands on the optical quality of the projection system, in the form of allowed aberrations, become extremely difficult to meet, especially when the physical size of the optical components increases with the reducing feature size. Therefore, so-called wafer steppers are used in modern microchip fabrication. The basic idea of a wafer stepper is to image only a small region of the mask onto the wafer, and pattern the entire surface area in consecutive steps.

The centerpiece of a wafer stepper is the exposure system. Its task is to image desired patterns from the photomask to their proper positions on the surface of the wafer. In many systems this process includes a size reduction between the patterns on the photomask and on the wafer. This reduces the dimensional accuracy requirements, such as feature size and positioning for the mask fabrication by whatever is the de-magnification of the lithography system. However, it also increases the complexity of the exposure system, as significance of aberrations is increased. The requirements of microchip production mean that the image of the photomask with submicron features must be reproduced on the wafer with a dimensional accuracy of only a few tens of nanometers and aligned to a specific position within a fraction of the linewidth. Furthermore, these tolerances must be guaranteed over the entire exposure field that is typically several square centimeters. It is not surprising that the resulting complex piece of optics costs several millions dollars.

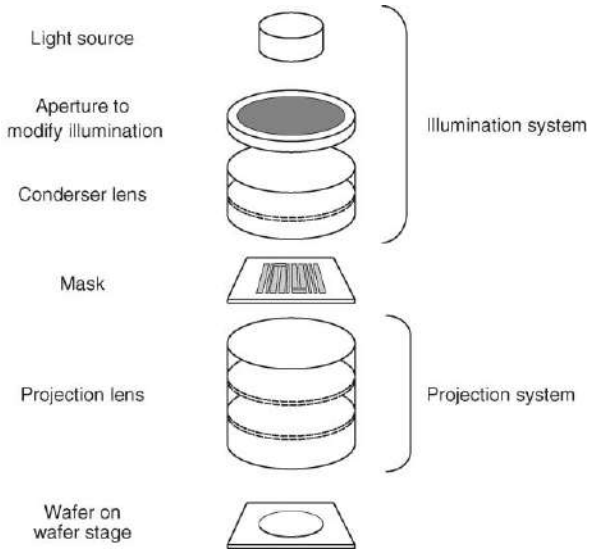

Figure 1 Schematic illustration of the exposure system.

The exposure system consists of a projection lens, illumination system, and wafer management system, as illustrated schematically in Figure 1.

A typical projection lens is a complicated lens system. As the resolution of a lens system is inversely proportional to the diameter of the lenses, the lenses in photolithographic systems can have diameters up to 250 mm, and the whole system is arranged inside a massive barrel-frame. The aim of the projection lens in the entire exposure system is to image the pattern from the photomask to the wafer. The projection lens is primarily responsible for the resolution and must therefore fulfill the highest optical requirements. To produce the smallest possible image, diffraction effects alone must essentially limit the resolution of the system. This means that the optical aberrations over the exposure field of several square centimeters must be virtually eliminated, by the wavefront deviations being less than a fraction of the wavelength.

The illumination system transfers the illumination source through the photomask to the entrance of the projection system. The main requirements for the illumination system are to collect most of the light into the system and to ensure that the irradiance of the photomask is uniform to within a few percent. In addition, resolution enhancement techniques, based on off-axis illumination, require that the illumination pattern on the photomask has a specific shape. We will discuss these requirements later in more detail. In general terms, the illumination system is less complicated than the projection system, especially with regard to the elimination of aberrations, due to the fact that precise imaging is not required in the illumination system.

Although the wafer management system is outside the scope of this article, we note in passing that it maintains the position, orientation, and movement of the wafers through the exposure system. Again, high precision is required to ensure the successful exposure of the desired pattern to the wafer.

Resolution Enhancement Techniques

In most modern lithography systems the achievable maximum resolution is increased via the use of one or more special resolution enhancement techniques (RET). These are essentially creative refinements of the lithography technology, based on theoretical optics papers published often decades before their first application in the lithography process in the late 1980s and early 1990s. There are essentially three RET approaches: Off-Axis Illumination (OAI), Optical Proximity Correction (OPC), and Phase Shifted Masks (PSM).

In optical proximity correction the basic idea is to predistort the mask pattern in order to overcome the proximity effects caused by the optics and the different processes. This is done either using a set of pre-existing rules describing the connection between a mask pattern and its realization on the wafer, using advanced models to predict the proximity effects during the lithography process, or with some combination of both. The key difficulties with OPC are related to the fabrication of the predistorted mask and to the accuracy of the used distortion rules or proximity effect models. Finally, it should be noted that OPC is strictly speaking not a true resolution enhancement technique, as it only enhances the effective resolution and not the theoretical resolution limit of the process defined by the optics.

In phase shifted masks the improvement in the resolution is obtained by altering the phase of light passing through different portions of the mask, creating regions of destructive interference in the image. This leads to higher contrast, which helps to improve the effective patterning resolution. However, as with OPC, there is no true resolution enhancement in terms of the theoretical resolution limit. There are two main difficulties with application of PSM. First, it is not straightforward to introduce required phase shifts into an arbitrary design layout, because some geometrical shapes, such as junctions are harder to realize than others. Another problem lies with the mask fabrication, which requires careful calibration of the material layers and can be very complicated and costly compared to fabrication of conventional binary masks.

For the purposes of this article, only off-axis illumination is discussed, as both OPC and PSM are essentially mask-based techniques and thus provide no opportunity for micro-optics. Off-axis illumination refers to any illumination scheme that significantly reduces or eliminates the light hitting the mask at near normal incidence, leading to the diffraction pattern of the mask being shifted on the projection lens. The principle behind the resolution enhancement, obtained by introducing such a shift, is schematically illustrated in Figure 2 for periodic pattern.

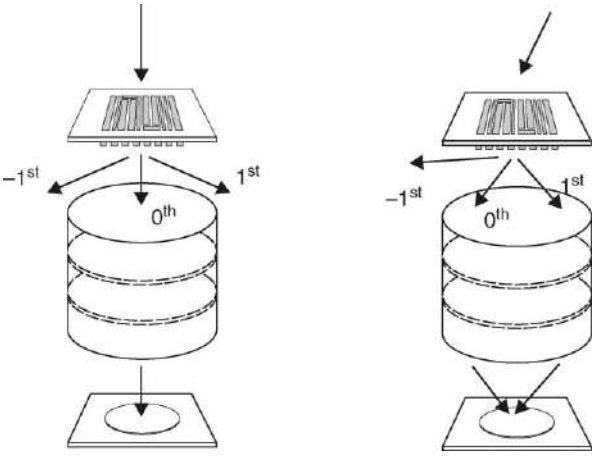

Figure 2 Schematic illustration of effect of off-axis illumination on image formation.

The mask pattern can be imaged if at least two diffraction orders generated by the mask pass through the pupil of the projection lens. Using on-axis illumination, the minimum feature size of the system is reached when the period of the mask pattern is so short that the ±1st diffraction orders can no longer pass through the pupil. If the illumination is moved off-axis, the diffraction pattern is shifted and the second diffraction order (in this case +1st) again passes the pupil. Thus the resolution has been increased. A secondary improvement in contrast may be obtained by shaping the diffraction orders falling onto the pupil, such that the filling of the pupil is optimized, i.e., such that less unwanted light contributing to the background illumination passes the pupil. Finally, the depth of focus of the system can be optimized by ensuring that the contributing orders (in this case 0th and +1st) are symmetric with regards to the optical axis.

The conventional way to realize off-axis illumination schemes is to introduce a suitable aperture to the illumination optics between the source and the mask. In order to effectively use off-axis illumination, the shape and size of the aperture, i.e., the illumination pattern, must be optimized for the specific mask shape and pitch. This optimization leads to four basic shapes, as illustrated in Figure 3. In most modern lithography systems these shapes

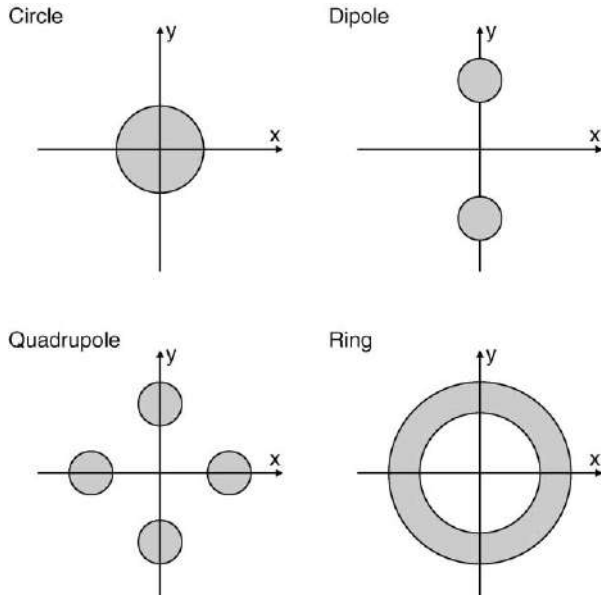

Figure 3 The four typical illumination patterns: disk, ring or annular, quadrupole, and dipole.

are available with different values of partial coherence factor, governed by the ratio of the numerical aperture of the illumination lens and projection lens, allowing control of proximity effects and edge sharpness.

Overview of Micro-Optics

There exists no generally agreed definition of micro-optics, but it typically means optical systems built to significantly smaller scales than conventional table-top systems, using components such as micro-lenses and microprisms with physical dimensions in the range of a couple of millimeters at most. Micro-optical systems also include so-called diffractive optical elements, which are usually optical components whose operation cannot be explained by Snell's law or the law of reflection. Such elements are usually macroscopically planar microstructures, consisting of features with dimensions from a couple of wavelengths to a few tens of microns, and are designed by advanced numerical algorithms based on diffraction theory.

In the past decade, micro-optics has emerged as a powerful new addition to the field of optics and optical systems, following improvements in micro-optics fabrication techniques, both via adaptation from established microelectronics fabrication processes and development of novel approaches, such as variable dose direct writing using either UV light or an electron beam. Additionally, emergence of replication techniques, such as injection molding or UV casting, allows the low-cost mass production of micro-optical elements, making micro-optics economically feasible for a wide range of consumer applications. Consequently, micro-optical elements, utilizing either refractive or diffractive surfaces, are now found in industrial applications such as material processing, in the optical telecom field, and in consumer electronics such as the digital video camera.

One of the major strengths of micro-optics, as compared to conventional optics, is their ability to realize complex optical functions into a compact form. This is especially true for micro-optics using diffractive optical elements, as diffractive elements may be designed to realize several optical functions such as focusing, filtering, or beam splitting at once allowing integration of several classical optical components into a single diffractive element.

Micro-Optics in Conventional Lithography Systems

The simplicity of contact or proximity printing means that there is little need or opportunity to incorporate additional functionality through inclusion of micro-optical elements. The only exception to this will be discussed in the final section of this article, where we will consider a nonconventional technique, which utilizes micro-optics and can be incorporated into proximity printing systems. However, in this section we will limit the discussion to projection-based lithography systems such as wafer steppers.

Let us consider the possible uses of micro-optics in the three main parts of the exposure system identified in the earlier section. It is obvious that the wafer management system, which is essentially mechanical, cannot be enhanced by inclusion of micro-optics. In principle, micro-optics could be incorporated into the projection system to reduce its weight and, to a lesser degree, complexity. However, it is important to note that using micro-optics to reduce the physical size, i.e., the diameter of the lenses in the projection system is not possible, as such reduction would increase the theoretical minimum feature size the system can resolve which is inversely proportional to the lens diameter. In practical terms, even the most advanced fabrication approaches used to realize micro-optical elements cannot yet produce elements with the precision and quality sufficient to meet the optical requirements associated with a lithographic projection lens. In general, the state of the art in micro-optics is such that it does not make sense to replace refractive lenses with diffractive ones unless the weight of the optics is a critical parameter in

determining the merit of the system. This is not the case for wafer steppers.

Even though the optical requirements for the lenses in the illumination system are not as high as in the projections system, they are in practice still far too demanding for micro-optical elements. Therefore, the only area where micro-optics could be used in the illumination system is the generation of various illumination patterns needed for the off-axis illumination schemes used for resolution enhancement, as well as overall homogenization of light. An element meeting these goals is essentially a custom-designed diffuser, where both the generated angular range and the relative intensity of illumination in any given angular direction are controlled to a high degree.

Custom Diffusers for Illumination

As mentioned earlier, one of the strengths of micro-optics lies in its ability to combine different optical functions into a single element. When used in lithographic illumination systems, micro-optical elements can serve a dual purpose. The element can serve as an artificial aperture, effectively distributing the light over some predetermined angular range and at the same time minimizing light outside the range, allowing generation of various illumination patterns needed for the resolution enhancement via off-axis illumination. Unlike conventional solutions based on the use of actual apertures to generate the desired illumination patterns, distribution of light using micro-optical elements can be done with efficiencies up to 95%, with typical target efficiency being in the range of 90%, depending on the pattern complexity. At the same time, the element can be used to smooth the light distribution, improving the uniformity of the illumination. Typically the required uniformity error, i.e., allowed variations from the average illumination, is in the range of 5%, but can be as low as 1%.

In addition to the above-mentioned high-efficiency and -uniformity targets, optical quality requirements of modern lithography systems introduce unique challenges to the micro-optics used in such systems. First, micro-optical elements used in lithography systems must also exhibit low amounts of stray light, especially in formation of hot-spots. Typically, hot-spots are allowed to be only a few percent of the average illumination level. Because of the fact that the size, shape, and profile of the illumination beam can vary, partly due to variations in the source and partly because such variations improve the functionality of the lithography system, by, for example, controlling the partial coherence factor, the micro-optical elements must be virtually space-invariant. The elements must perform the same function regardless of where and how the illumination coming from the source hits the element. Furthermore, the elements themselves must have very low absorption, because the high power used could otherwise lead to destruction of the elements. This means that the materials must be carefully selected. Finally, the optical wavelengths used in advanced lithography systems are 248 nm or less. Because the feature size of a diffractive element scales down with the wavelength, the features in elements used are much smaller than in the case of elements used for applications in the visible region. Consequently the accurate fabrication of such elements becomes a formidable challenge, because most fabrication errors, such as mask misalignment, etch depth error or the way the sharp corners in diffractive elements get rounded during fabrication, increase in significance when the feature size is decreased.

There exist two distinct approaches for obtaining such custom-designed diffusers. We will now discuss the benefits and weaknesses of these two techniques in terms of efficiency, pattern uniformity, stray light, and ease of fabrication, in order to illustrate the trade-off a designer creating micro-optical elements for a lithography system has to consider.

Aperture Modulated Diffusers

One approach to acquire custom diffusers for illumination systems is to use an array of refractive or diffractive micro-lenses. More generally, the lens phase can be replaced by any phase obtained by geometrical considerations, i.e., by a geometrical map transform between the input and output planes of the optical system. The basic principle of an aperture modulated diffuser is to use the phase function to accurately control the angular divergence of the diffuser, while the shape of the desired intensity distribution is created by introducing a properly shaped aperture in which the phase function is contained. This basic principle is shown in Figure 4.

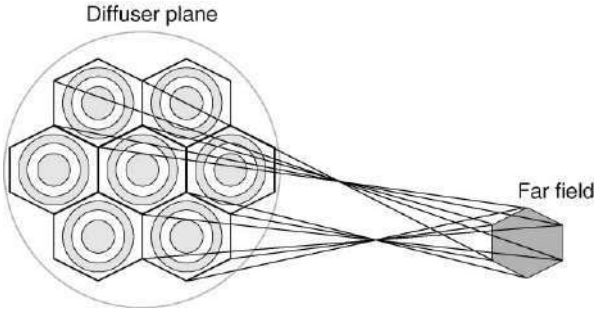

Figure 4 Basic principle of aperture modulated diffuser. The lens f-number determines the angular extend of the far field pattern, while the lens aperture shape determines the shape of the pattern.

The main benefit of using a lens phase lies in the ease of the design process, which essentially consists of selecting the numerical aperture of the lens to meet the angular size of the desired pattern, while more complicated phase profiles can be used to improve some aspects of the performance of the diffuser.

In terms of optical performance, aperture modulated diffusers typically achieve efficiency of up to 95% and low levels of stray light, easily fulfilling the target specifications outlined earlier, especially when refractive surfaces are used. However, since the lens aperture in the element plane controls the shape of the intensity distribution, the range of intensity distribution shapes is limited. This is necessary because it must be possible to tile the required aperture shape in order to realize an array of elements with high fill factor, i.e., to realize a space-invariant diffuser. Furthermore, introducing an aperture into the element plane means that intensity oscillations, due to boundary diffraction at the edge of the apertures, cannot be avoided. Consequently, the uniformity of the generated intensity distributions suffers. This is especially true in cases where the desired intensity pattern has sharp corners, where the intensity oscillations due to boundary diffraction, can easily be up to 50% of the average intensity of the desired intensity pattern. Figure 5 illustrates a typical far field intensity distribution obtainable with aperture modulated diffusers.

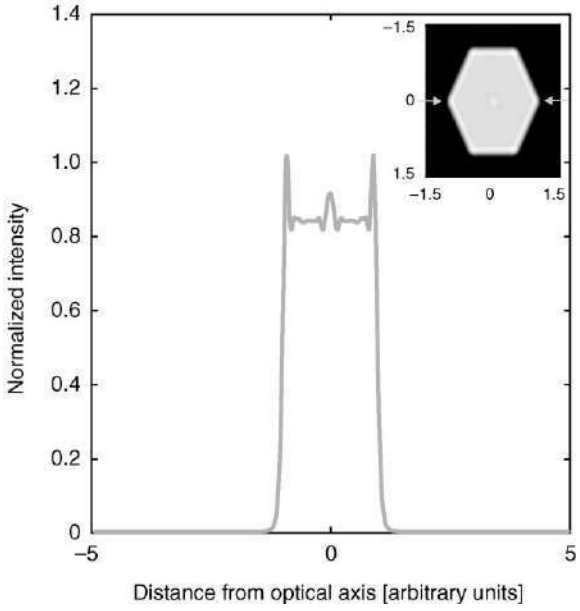

Figure 5 Typical intensity distribution obtainable with an aperture modulated diffuser. The overall distribution is smooth apart from the slow intensity oscillations at the edges resulting from diffraction at the aperture boundary.

One needs to consider several issues with regard to the fabrication of aperture modulated diffusers. For refractive aperture modulated diffusers, the main issue is the accurate realization of the required surface profile. The fabrication of large arrays of refractive elements having arbitrary surface profiles is still challenging. Resist reflow techniques, where the micro-lenses are realized by melting small islands of resist, lead to a spherical lens profile. Complicated and hard-to-control profile reshaping techniques are required if dramatic alterations from the spherical lens profiles are needed. Even though fabrication through direct writing techniques, such as laser writing offers more freedom, realization of arbitrary refractive surface remains a challenge. Therefore, use of refractive elements is mainly limited to cases where the lens profile can be close to the spherical shape. The most important of such cases are patterns with uniform circular or rectangular distributions. The former can be realized by proper selection of the numerical aperture of the lens, while the latter can be realized by combining two properly selected cylindrical lenses on opposite sides of the substrate.

For diffractive aperture modulated diffusers there are considerably less limitations in terms of what phase profiles can be attained. As long as the opening angles, i.e., the angular size of the pattern is below 10 degrees, the phase can be realized as a diffractive element. With diffractive elements, the main issue to consider is therefore the effect of typical fabrication errors such as profile depth and shape errors. First, the typical consequence of fabrication errors in the element is loss of efficiency via increase of stray light. Fortunately the errors do not generate sharp stray light peaks, with the obvious exception of the axis hot-spot. This results from the fact that both lens functions and more complicated phase functions designed by geometrical considerations, exhibit locally periodic structures, and the introduction of common fabrication errors then leads to redistribution of light away from the desired signal orders of these local structures to higher or lower orders. On the level of the whole intensity distribution, such redistribution appears as a constant rise in the background noise level, but will not lead to sharp intensity fluctuations.

Another issue related to the fabrication of aperture modulated diffusers appears when diffuser elements are fabricated using a series of photomasks, as it is typically the case with a relatively large size of the element. Although mask misalignment is essentially a random error, i.e., the direction and amount of the error are difficult to predict before hand, it easily leads to systematic and partly symmetric uniformity problems with aperture modulated diffusers.

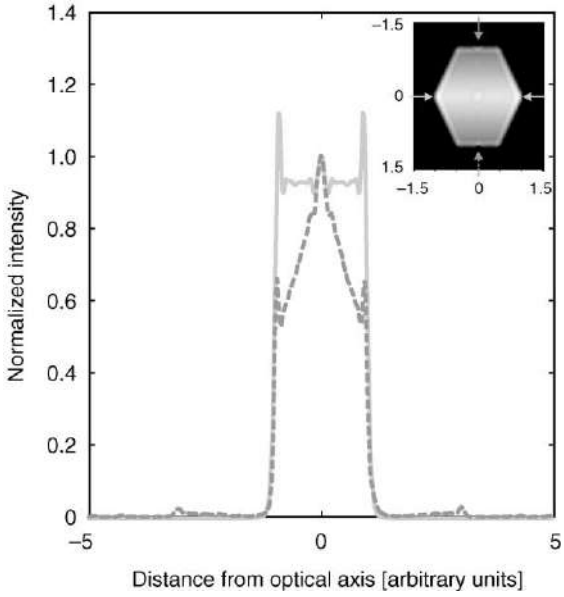

Figure 6 Effect of mask misalignment on the intensity distribution with aperture modulated diffusers. Essentially random error leads to symmetric and systematic uniformity error due to locally periodic nature of the design.

Again the reason for this lies in the locally periodic nature of the elements. A single mask misalignment along some direction results in an erroneous profile shape in the structures, which are periodic in the direction of the alignment error. Consequently, the relative intensity distribution along that direction is changed. On the other hand, structures, where the period is perpendicular to the misalignment, are realized correctly. Furthermore, the local period of the phase function is often changing from the center of the element to the edge, meaning that the relative magnitude of the mask misalignment is different in different parts of the element. These facts lead to asymmetry in the overall intensity distribution, which is especially harmful with off-axis illumination patterns, as the asymmetry results in loss of the resolution gain obtained with off-axis illumination. Furthermore, as the source of the asymmetry is essentially a random error, compensation in the design is very difficult. Figure 6 illustrates the above-mentioned phenomena.

Fan-Out Based Diffusers

An alternative technique uses specially designed fan-out elements to produce a large number of closely spaced diffraction orders arranged in a desired pattern, as illustrated in Figure 7. Such an approach offers all the flexibility of diffractive elements, allowing generation of nearly arbitrary intensity distributions, especially if the elements can be realized

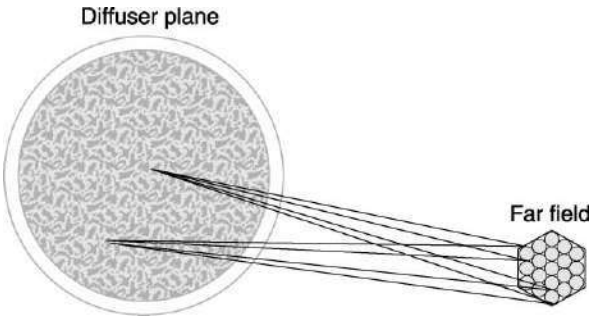

Figure 7 Basic principle of fan-out based diffusers. A grating is used to generate a high number of closely spaced diffraction orders arranged in the shape of the desired pattern.

as so-called multilevel elements, i.e., if the surface profile can have more than two distinct levels.

The flexibility in the design comes at a cost, where specially developed design tools, such as the iterative Fourier transform algorithm, are required. Fortunately such tools are starting to become commercially available as emphasis and interest in wave-optical engineering, an extension of the classical optical engineering approaches, increases.

In terms of optical performance, fan-out based diffusers offer high uniformity as there are no boundary diffraction effects as with aperture modulated diffusers. Furthermore, by choosing the beam separation of the fan-out to be clearly smaller than the divergence of the incidence beam, it is possible to average out any local intensity variations from order to order, leading to a smooth overall intensity distribution. The remaining variations in the intensity distribution generated by a fan-out based diffuser can be as small as 1% of the average intensity.

As a trade-off for the improved performance in terms of uniformity compared to aperture modulated diffusers, fan-out based diffusers typically exhibit lower efficiency and, consequently, higher levels of stray light, even when no fabrication errors are present. This is a result of additional design freedoms being used to remove unwanted intensity oscillations from the far field regions, belonging to the desired intensity distribution, by introducing some noise outside these regions. Typically the efficiency of a fan-out based diffuser is in the range of 85–90%, or 5–10 percentage points lower than the efficiency of a comparable aperture modulated diffuser (should one exist). Furthermore, unlike aperture modulated diffusers, fan-out based diffusers can have stray light that forms hot-spots. If care is not taken during the design, these hot-spots may be up to 10% of the average intensity of the illumination pattern, even in the absence of fabrication errors. Figure 8 illustrates

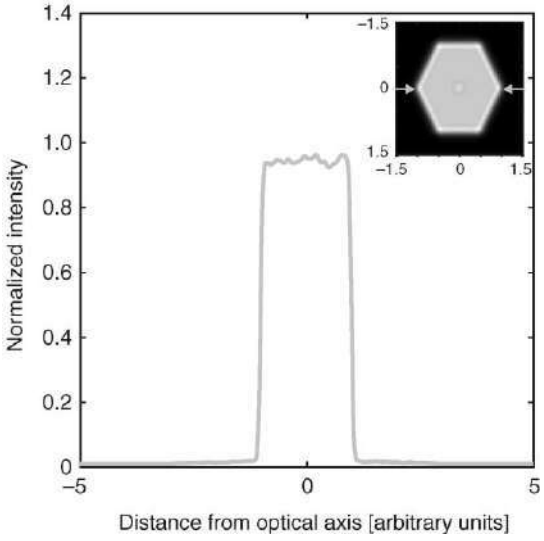

Figure 8 Typical intensity distribution obtainable with a fan-out based diffuser. Overall distribution is smooth, but stray light appears outside the main pattern.

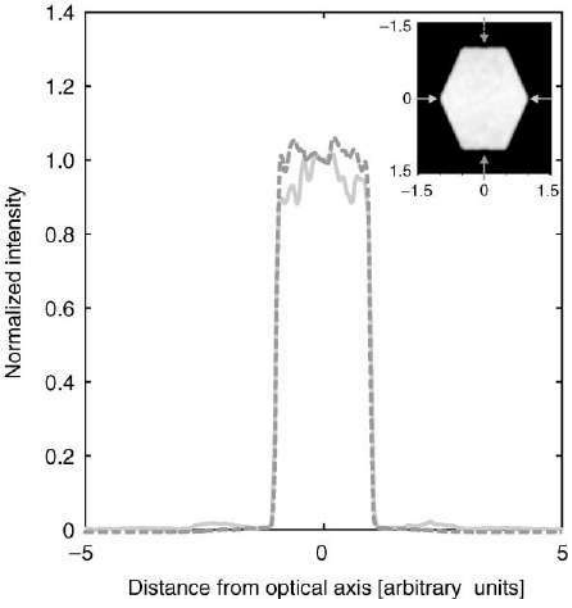

Figure 9 Effect of mask misalignment on the intensity distribution with fan-out based diffusers. The uniformity error appears randomly, allowing statistical compensation schemes.

a typical far field intensity distribution obtainable with fan-out based diffusers.

Fan-out based diffusers offer several advantages in terms of fabrication compared to mapping based diffusers. First, random errors, such as mask misalignment, do not lead to symmetric or systematic uniformity problems, but instead result in random local oscillations in the intensity pattern as illustrated in Figure 9. This allows special compensation schemes where the element is divided into segments, each of which has different fabrication tolerances but generates the same far field intensity distribution. If the fabrication tolerances of the individual designs are balanced correctly against the typical fabrication errors of the used fabrication methods, the loss of uniformity due to fabrication errors can be eliminated almost entirely. Furthermore, the flexibility of diffractive design allows for fabrication error-tolerant designs. This follows from the ability to combine several optical functions in a single optical element. The typical trade-off with designs with relaxed fabrication tolerances is that the optical performance of the element in terms of efficiency is worse than that of a conventional design if fabrication errors are not present.

Micro-Optics and Non-Conventional Lithography

In the previous section we have considered micro-optics in connection with conventional lithography approaches and shown that micro-optical elements can find a significant role in such systems. We conclude this article by considering briefly nonconventional lithography systems, focusing especially on a technique called micro-lens lithography, which is fundamentally based on the use of micro-optical elements.

The most straightforward way to improve existing lithography approaches is to further reduce the wavelength used in the exposure process. In X-ray lithography, the light source used in conventional lithography approaches is replaced by a cyclotron emitting light with wavelengths deep in the X-ray region. This leads to minimum resolution far lower than achievable by conventional techniques, but with the additional difficulty and cost of maintaining and running such a source.

The techniques that are currently considered as serious alternatives for conventional photolithography approaches are based on the use of particles other than photons in the fundamental exposure process. Techniques such as electron- or ion-beam lithography are routinely used to realize structures with feature size and positioning resolution of some tens of nanometers for use in both micro-electronics and micro-optics. The main drawback of these approaches is the cost involved in patterning large surface areas. Unlike optical lithography techniques, particle beam lithography cannot expose large surface areas at once, instead requiring painstaking scanning of the surface area with a relatively small writing beam. This leads to processing times that scale linearly with the required surface area, and consequently the throughput of such techniques is small. Therefore, particle beam lithography is not yet widely used in commercial fabrication of micro-electronics. However, it is worth noting that especially

electron beam lithography is widely used in the origination of micro-optical components, as the relatively high cost of creating the master component can be offset in the final product by the inexpensive replication process. Such an approach allows the replicated end-product to meet the high optical requirements achievable only by high resolution surface control, yet remain affordable for use in, for example, consumer electronics. This has lead some in the micro-optics community to conclude that electrons are better suited for fabrication of components controlling photons, while photons are better suited for fabrication components controlling electrons.

In addition to particle beam lithography, other noteworthy techniques exist. For our purposes, the most interesting technique is the so-called micro-lens array lithography, where one or more micro-lens arrays are used to image the pattern of the photomask to the resist layer on the substrate. Although the resolution of this approach is limited to 3–5 microns, for applications such as with flat panel displays, micromechanics, or multichip module fabrication, this technique is of considerable interest.

It is well known that the geometrical aberrations scale down proportionally with the physical lens size. Thus, large areas can be imaged favorably by use of arrays of micro-lenses, where each lens images only a small section of the whole field. Such a setup is shown in Figure 10.

With this setup it is possible to expose a comparatively large area with a single exposure, thus eliminating the need for costly stepping operations and therefore increasing the throughput of the system. Analysis of the basic system shows that an array based on spherical micro-lenses of 1 mm focal length and 300 micron aperture can produce a nearly diffraction limited resolution on the order of 3–5 micrometers over a 300 micron field.

In terms of conventional lithography systems, the micro-lens array lithography has most in common with proximity printing. Both techniques achieve comparable resolutions, yet avoid damaging the mask or substrate by keeping the two separated by some finite working distance. However, micro-lens array lithography offers two distinct advantages over conventional proximity printing. First, the working distance between the imaging system and the substrate surface can be much larger, which further minimizes the risk of mask or substrate damage. Furthermore, added working distance allows lithography steps on nonflat surfaces, e.g., inside holes or grooves where conventional proximity printing would not be possible. Another advantage of the micro-lens array lithography is the larger depth of focus. With proximity printing, the resolution rapidly decreases with increasing distance. In micro-lens array lithography, the optimum resolution is obtained at the image plane, while a region of reasonable resolution extends forward and backward from this plane (Figure 11). This extension means that in the depth of field, resolution is less sensitive to changes in the working distance, which makes it better suited for practical applications.

In addition to the optical benefits of micro-lens array lithography, there also exist other practical benefits. Many micro-devices have a repetitive and

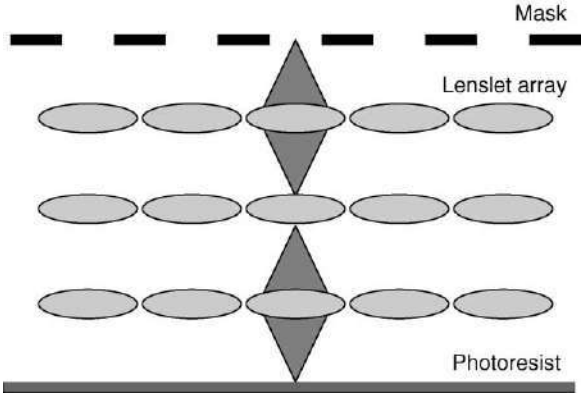

Figure 10 Basic geometry of micro-lens array lithography.

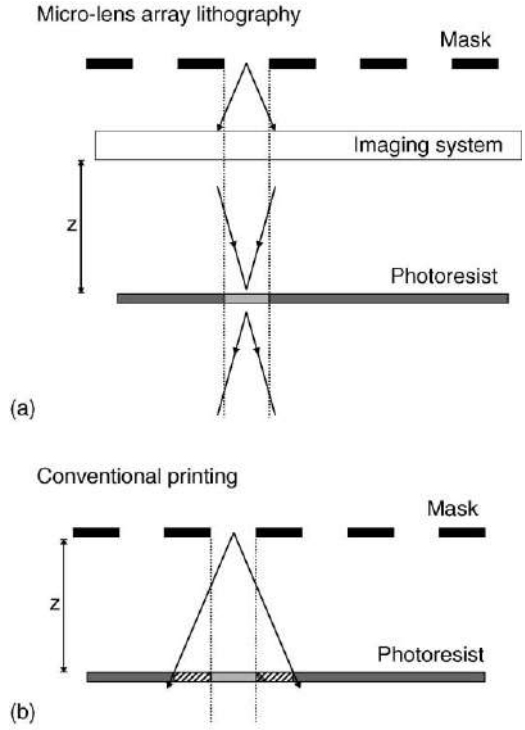

Figure 11 Depth of focus with micro-lens array lithography and conventional printing. (a) With micro-lens array lithography, the image of the mask is extended both forward and backward from the working plane at distance *z*. (b) In conventional printing, the resolution is rapidly reduced with increasing working distance *z*.

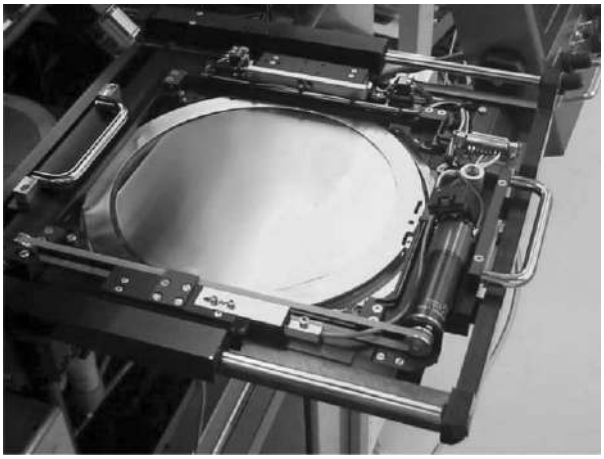

Figure 12 Ultra-flat micro-lens projection system (courtesy of R. Völkel, SUSS MicroOptics, Neuchâtel, Switzerland).

discrete nature with respect to some unit element. Such a unit element lends itself well to the imaging properties particular to lens arrays, as each lens of the array can be made to correspond to the unit cell of the particular micro-device making the imaging highly efficient. A photograph of a micro-lens projection system is shown in Figure 12.

One could also imagine that instead of an array of imaging systems, an array of pattern generators is used that generates a specific pattern, such as lines or points. In the ideal case such a system would replace the mask and the imaging system altogether. However, this approach is only possible for simple patterns and large production series.

Outlook

As long as optical lithography stays attractive, micro-optics will be used as beam-homogenizers and sophisticated diffusers to generate optimized illumination patterns. How far micro-optics will be used in the imaging system is difficult to predict. However, it can be assumed that the difference between micro-optics and macro-optics fabrication technology will disappear.

Recent research is also investigating near-field optics to further reduce the resolution limit. Micro-optics technology is well suited to realize such systems. A serious problem of near-field optics is the short distance (10 nm–100 nm) between the optical element and the interface to be patterned. In addition, near-field optics is based on scanning systems, which increases the exposure time considerable. However, there is still room for improvements or as Richard Feynman said 'there is plenty of room at the bottom'.

Further Reading

Dändliker R, Gray S, Clube F, Herzig HP and Völkel R (1995) Non-conventional techniques for optical lithography. *Microelectronic Engineering* 27: 205–211.

Herzig HP (1998) *Micro-Optics: Elements, Systems and Applications*. London: Taylor & Francis.

Herzig HP and Kipfer P (1999) Aperture modulated diffusers (AMDs). In: Asakura T (ed.) *International Trends in Optics and Photonics*. Berlin: Springer.

Naumann H and Schröder G (1992) *Bauelemente der Optik*. München, Germany: Carl Hanser Verlag.

Sheats JR and Smith BW (eds) (1998) *Microlithography: Science and Technology*. New York: Marcel Decker.

Turunen J and Wyrowski F (eds) (1997) *Diffractive Optics for Industrial And Commercial Applications*. Berlin: Wiley-VCH.

Völkel R, Herzig HP, Nussbaum P, Dändliker R and Hugle WB (1996) Microlens array imaging system for photolithography. *Optical Engineering* 35: 3323–3330.

Design and Fabrication of Diffractive Optical Elements

D W Prather, T Dillon, A Sure and X Gao, University of Delaware, Newark, DE, USA

J N Mait, US Army Research Laboratory, Adelphi, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Diffractive optical elements (DOEs) are passive devices that redirect and focus light through the division and mutual interference of a propagating electromagnetic wave. This is in contrast to refractive elements, which redirect and focus light through variations in indices of refraction. From a historical perspective, diffraction was traditionally considered a detriment in optics and caused images to blur. However, we now know that it is possible to use diffraction to enhance the performance of optical systems. For example, DOEs can be used to correct chromatic aberrations and to implement optical functions that conventional refractive and reflective optics cannot, such as multiple beamsplitting.

Further, recent advances in computing and fabrication technologies have increased the number of applications in which diffractive elements are used, which has established diffractive optics as yet another tool in an optical designer's kit.

The first modern diffractive element was the hologram. A hologram is a photographic record of the constructive and destructive interference between two coherent beams known as the object and write beams. The object beam contains information about the object recorded. The write beam is necessary to record the hologram but contains no information itself. When the hologram is played back with a third beam, the read beam, the recorded interference pattern diffracts it in such a way that it reconstructs the object beam.

Unfortunately, the functions that a natural hologram can produce are limited to those whose wavefields can be produced in a laboratory. Thus, although it is possible to record a hologram that can focus or deflect a single beam, it is difficult to record a hologram that splits a single beam into multiple foci. However, the advent of computing technology made it possible to design such holograms analytically. The first diffractive element designed with the aid of a computer was demonstrated in 1966. Since then the ability to create elements that respond arbitrarily has had widespread application in imaging systems, telecommunications, and sensing.

The two most basic diffractive elements are the grating and the Fresnel zone plate. The diffractive behavior of a grating is predominately transverse to the optic axis. It causes light to split into discrete beams, referred to as diffraction orders, each with a different angle of propagation relative to the optic axis. The diffractive behavior of a Fresnel zone plate also divides incoming light into discrete orders, however, the orders are produced along the optic axis. This is useful in producing the diffractive equivalent of a refractive lens. Arbitrary transverse and longitudinal behavior is realized by combining characteristics of these basic two diffractive elements. In the remainder of this article, we discuss the design and fabrication of general diffractive elements.

Design

Once an optical system designer determines the functional behavior required for a particular application, for example, focusing or multiple beamsplitting, he or she needs to design a diffractive element that realizes this function. However, unlike refractive and reflective optical elements, whose material properties and surface curvatures determine their functionality, the functionality of a diffractive element is determined by a pattern of wavelength scale features. Determining this pattern is the goal of diffractive design but to do so one must first understand the physical relationship between the DOE and the wavefield it produces.

Scalar diffraction theory is one of the simpler models that relates the effect of a surface relief, or diffractive profile, $d(x, y)$, on an incident wavefield $U_{\text{in}}(x, y)$:

$$U_{\text{out}}(x, y) = \exp[\mathrm{i}\theta(x, y)]U_{\text{in}}(x, y) \qquad [1]$$

where the phase transformation $\theta(x, y)$ is related to the diffractive profile $d(x, y)$ by

$$\theta(x, y) = 2\pi(n - 1)d(x, y)/\lambda \qquad [2]$$

The constant n is the refractive index of the optical substrate and λ is the wavelength of illumination in a vacuum.

The effects of the phase change are not apparent until the wavefield $U_{\text{out}}(x, y)$ propagates over some distance z. In scalar diffraction theory the relationship between $U_{\text{out}}(x, y)$ and the wavefield $U_z(x, y)$ at a plane a distance z away, is described by the Rayleigh–Sommerfeld diffraction equation:

$$U_z(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_{\text{out}}(\zeta,\eta) \times \frac{\exp\left\{\mathrm{i}(2\pi z/\lambda)\sqrt{1+(x-\zeta/z)^2+(y-\eta/z)^2}\right\}}{\mathrm{i}z\lambda[1+(x-\zeta/z)^2+(y-\eta/z)^2]}\mathrm{d}\zeta\,\mathrm{d}\eta \qquad [3]$$

Huygens' principle is reflected in this formulation by the convolution between the transmitted field $U_{\text{out}}(x,y)$ and a spherical wave $h(x,y)$, which represents free space propagation over the distance z:

$$U_z(x,y) = U_{\text{out}}(x,y)^{**}h(x,y) \qquad [4]$$

where

$$h(x,y) = \frac{\exp\left\{\mathrm{i}(2\pi z/\lambda)\sqrt{1+(x/z)^2+(y/z)^2}\right\}}{\mathrm{i}z\lambda[1+(x/z)^2+(y/z)^2]} \qquad [5]$$

The Rayleigh–Sommerfeld formulation is valid so long as the angles at which energy diffracts from the aperture are small. If the distance z is sufficiently large that the angles stay clustered about the optic axis, one can replace the spherical wave response of freespace

with a parabolic wave:

$$h(x, y) \approx \frac{\exp(\mathrm{i}2\pi z/\lambda)}{\mathrm{i}z\lambda}\exp[\mathrm{i}(\pi/\lambda z)(x^2 + y^2)] \qquad [6]$$

This simplification is referred to as the paraxial, or Fresnel, approximation. The diffraction integral is now given by:

$$\begin{aligned} U_z(x, y) \approx\ & \frac{\exp(\mathrm{i}2\pi z/\lambda)}{\mathrm{i}z\lambda}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_{\mathrm{out}}(\zeta, \eta) \\ & \times \exp\{\mathrm{i}(\pi/\lambda z)[(x-\zeta)^2 + (y-\eta)^2]\}\mathrm{d}\zeta\,\mathrm{d}\eta \\ =\ & \frac{\exp(\mathrm{i}2\pi z/\lambda)}{\mathrm{i}z\lambda}\exp[\mathrm{i}(\pi/\lambda z)(x^2 + y^2)] \\ & \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_{\mathrm{out}}(\zeta, \eta)\exp[\mathrm{i}(\pi/\lambda z)(\zeta^2 + \eta^2)] \\ & \times \exp\{-\mathrm{i}2\pi[\zeta(x/\lambda z) + \eta(y/\lambda z)]\}\mathrm{d}\zeta\,\mathrm{d}\eta \end{aligned} \qquad [7]$$

The second representation results from an expansion of the exponent and allows the diffraction integral to be written as a Fourier transform. The quadratic phase terms can be removed if a positive lens with focal length z is placed next to the diffractive element and a second positive lens with focal length z is placed in the image plane (see **Figure 1**). Under these conditions, $U_{\mathrm{out}}(x, y)$ and $U_z(x, y)$ form a Fourier transform pair:

$$\begin{aligned} U_z(x, y) &= \mathrm{FT}[U_{\mathrm{out}}(x, y)]|_{u=x/\lambda z, v=y/\lambda z} \\ &= \mathrm{FT}\{\exp[\mathrm{i}\theta(x, y)]\}^{**}\mathrm{FT}[U_{\mathrm{in}}(x, y)]|_{u=x/\lambda z, v=y/\lambda z} \end{aligned} \qquad [8]$$

The objective of diffractive design is to find a surface relief $d(x, y)$ that generates a desired response $q(x, y)$ somewhere in the response $U_z(x, y)$. However, eqns [3] and [7] neglect the fabrication process, which affects the fields that can actually be realized. A diffractive optic designer can incorporate fabrication constraints into their design either directly or indirectly. In a direct approach, the designer represents the diffractive structure parametrically, for example, a multilevel phase element can be represented by a discrete set of allowed phase values and a binary phase element can be represented by the contours that separate one phase value from the other. A designer defines the diffractive element by specifying and modifying these parameters.

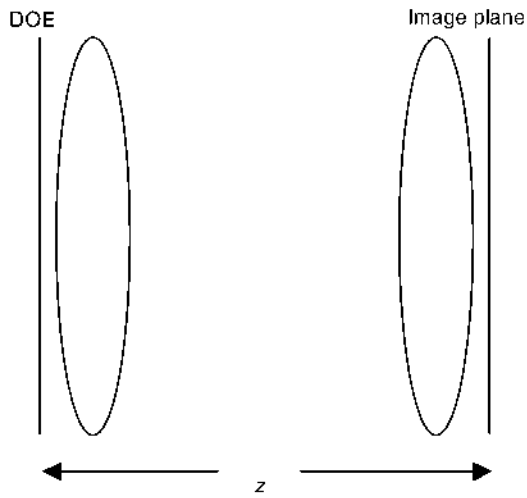

Figure 1 Optical Fourier transform module.

In an indirect approach, the designer first determines an intermediate function without consideration for fabrication and afterward applies the fabrication constraints to this function. For example, if $f(u, v)$ is the desired response one wishes to generate in the z-plane using $U_{\mathrm{in}}(x, y)$ as the input, from eqn [8], the optical function required to do is

$$Q(x, y) = F(x, y)/U_{\mathrm{in}}(x, y) \qquad [9]$$

where $F(x, y)$ is the Fourier transform of $f(u, v)$. The designer must then find a phase $\theta(x, y)$ that satisfies the fabrication constraints and insures that $\exp[\mathrm{i}\theta(x, y)] \approx Q(x, y)$.

For example, fabrication using multiple binary masks yields a phase structure quantized to 2^N levels. An effective design should, therefore, yield a quantized phase structure that minimizes reconstruction error. A naïve approach is to set the magnitude of $Q(x, y)$ in eqn [9] to unity and quantize the phase. However, this introduces considerable error. A direct approach to design searches only over the set of quantized phase structures for the one that, for example, generates the desired response with maximum energy and minimum error. An indirect design seeks first a phase structure whose response produces a minimum amount of error between it and $Q(x, y)$ and then quantizes this structure.

Performance of the element can be improved if one exploits certain freedoms available in design. For example, all systems have a finite-sized detector. It is, therefore, unnecessary to specify the performance of the diffractive element everywhere in the image plane. Its behavior within the detector window is all that is important. If only the intensity of the response is specified, then the phase of the response within the window is also a design freedom. Finally, one can also control the amount of energy contained in the signal window by scaling the amplitude of the response. This allows a designer to balance diffraction efficiency with error in the signal window.

Two general nonlinear optimization algorithms for the design of DOEs are represented in **Figure 2**. Indicative of the nature of data flow, **Figure 2a** is

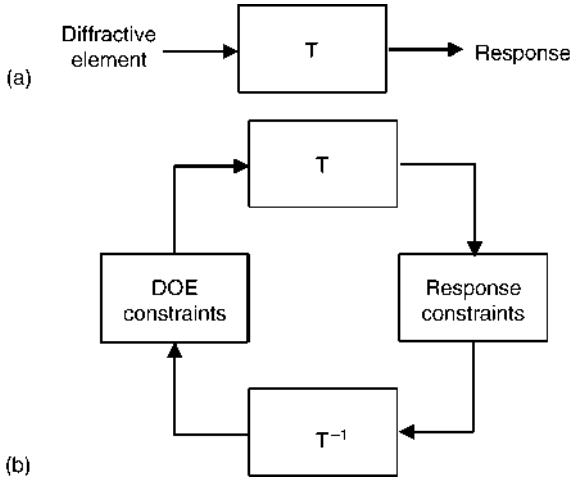

Figure 2 Algorithms used for diffractive design. (a) Unidirectional and (b) bidirectional algorithms.

referred to as a unidirectional algorithm and Figure 2b as bidirectional. In a unidirectional algorithm, changes are made only to the diffractive element. The changes are driven by how well the element performed in a previous iteration. If models exist for both forward and inverse propagation ($\mathbf{T}$ and $\mathbf{T}^{-1}$, respectively), a designer can use a bidirectional algorithm, which allows modification of both the diffractive element and its response. This allows bidirectional algorithms to converge to a solution, in general, more rapidly than unidirectional algorithms. In scalar theory, $\mathbf{T}$ is a Fourier transform but more rigorous models exist. If the model of the optical system cannot be inverted, a unidirectional algorithm must be used.

Fabrication

Regardless of which design algorithm is used to determine $d(x, y)$, ultimately it needs to be realized physically. Although numerous techniques exist for doing so, by far the most common are based on microlithographic processes. The process requires exposing a layer of photoresist deposited on top of an optical substrate and etching the photoresist to transfer the pattern into the substrate. These methods are especially appealing because they leverage the existing manufacturing infrastructure of the microelectronics industry and, as a consequence, have enabled the efficient and economical mass manufacture of DOEs.

Two basic methods exist for patterning photoresist, direct writing using a modulated beam and indirect patterning using masks. Further, the mask process can be either a single or multistep process. The multistep approach uses N masks and N exposures to fabricate a multilevel diffractive element that has 2^N phase levels (see Figure 2). However, profiles that contain very small features are difficult to fabricate in this manner. As a consequence, alternative processes using single-step grayscale lithography have also been developed.

Although grayscale photolithography is more time-efficient and less prone to mask alignment errors than multistep lithography, the photoresist must respond linearly to illumination. Unfortunately, most photoresists used to fabricate microelectronic circuits are highly nonlinear. However, one can obtain a linear response by adjusting the exposure and development process. In the following, we review fabrication using both multiple binary masks and a single grayscale mask.

Multilevel Fabrication Using Binary Masks

In this fabrication method, one must first generate a sequence of N masks that are related, in a recursive manner, to the quantized phase $\theta(x, y)$:

$$m_k(x, y) = \mathrm{mod}(M_k(x, y), 2), \quad k = [1, N] \qquad [10]$$

where

$$M_{k+1}(x, y) = \frac{M_k(x, y) - m_k(x, y)}{2} \qquad [11]$$

and

$$M_1(x, y) = \theta(x, y)\frac{N}{2\pi} \qquad [12]$$

The masks can be produced using either optical or electron beam (e-beam) pattern generators from integrated circuit fabrication. The pattern generators expose a thin layer of photoresist on the surface of a chrome-covered quartz substrate. Development of the photoresist leaves the desired diffractive pattern in the unexposed areas. The diffractive pattern is transferred to the chrome layer by etching any chrome not covered by photoresist. After the chrome etch process is completed, any remaining photoresist is washed away to yield the finished lithographic mask. This process is repeated for each mask.

The completed mask set is used to fabricate the DOE using techniques similar to those used to generate the mask set. An optical substrate is coated with a thin layer of photoresist, placed in intimate contact with the first lithographic mask $m_1(x, y)$, and illuminated with an ultraviolet (UV) exposure lamp. During photoresist development, exposed resist is washed away, which leaves a binary pattern in the remaining developed photoresist.

When the pattern is transferred into the substrate material via etching, the photoresist acts as an etch stop. For the first mask, the etch depth into the substrate corresponds to a π-phase shift. The process is repeated for each subsequent mask, that is, the substrate is coated with a layer of photoresist, UV illumination exposes the photoresist through the mask, the photoresist is developed, and the substrate etched by the remaining pattern. The etch depth of each mask is half that of the previous mask. Figure 3 illustrates this process.

The substrate can be etched using either wet or dry techniques. In wet etching, one uses reactive chemicals, such as hydrofluoric acid, to etch the substrate. However, wet etching is isotropic, that is, substrate etching occurs equally in all directions. As a result, it is difficult to preserve vertical sidewalls and to achieve high fidelity in etching high-aspect ratio features. As a consequence, dry etching is more commonly used.

A dry etch uses a beam of ions to remove substrate material. The process for removal can be either chemical or kinetic. The process used is determined by the value of the plasma power that creates the source of ions relative to the bias power used to accelerate the ions toward the substrate. If the bias power is large and the plasma power small, a small stream of ions are accelerated rapidly toward the substrate. In this case, the operational mode is similar to a micromilling machine, wherein inert particles chip away at the surface by virtue of their kinetic energy. Alternatively, a small bias power and a large plasma power produce a large population of ions with no preferential directionality. This yields an isotropic process similar to wet etching.

Two methods for performing dry etching are reactive ion etching (RIE) and inductively coupled plasma (ICP) etching. However, in RIE systems the bias and plasma power are not independent due to capacitive coupling. Thus, adjusting one affects the other. ICP, on the other hand provides independent control over the bias and plasma powers. Independent control allows one to control precisely the anisotropic nature of an etch process. As mentioned above, this is achieved by varying the bias power relative to the plasma power, which controls the directionality of the acceleration.

Independent control is important because the ideal process for a given DOE depends on the substrate material and the characteristics of the diffractive profile. Independent control allows one to adjust the respective powers, as well as gas flow rates, and their partial pressures, to determine an optimal etch process. Dry etching allows one to incorporate endpoint detection, which provides precise control

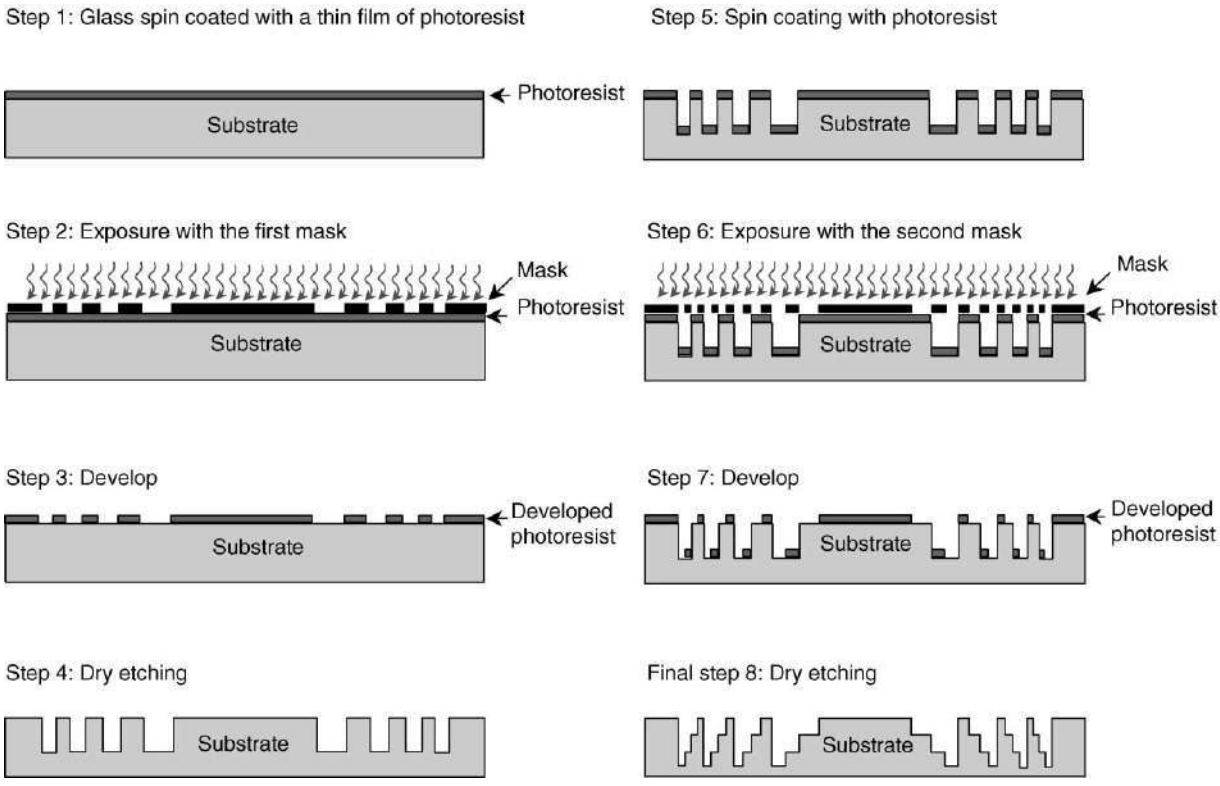

Figure 3 Fabrication of a multilevel quantized phase DOE using binary masks.

over the etch depth. As presented below, achieving the proper etch depth is important to generating responses with high diffraction efficiency.

Care must be used when determining the number of phase levels appropriate for a particular application. The number of phase levels has a direct impact on the portion of input energy one can hope to diffract into the desired response. This is referred to as the diffraction efficiency. According to scalar diffraction theory, the relationship between the maximum diffraction efficiency η and the number of masks N is:

$$\eta = \mathrm{sinc}^2\left(\frac{1}{N}\right) \qquad [13]$$

Thus, using two masks to create a 4-level phase element, as opposed to a single mask to create a binary-phase element, doubles the upper bound on the diffraction efficiency from 41% to 81%. The maximum diffraction efficiency for 8-level and 16-level phase elements is 95 and 99%, respectively.

However, this increase in diffraction efficiency incurs fabrication costs. After the first etch step, each subsequent exposure requires careful mask alignment. Alignment is accomplished using a standard tool from integrated circuit fabrication known as a mask aligner. The quality of a mask aligner is determined by its precision and the majority of commercially available aligners can register a mask and substrate to within less than a micrometer. This degree of alignment accuracy is sufficient to fabricate many useful multilevel diffractive elements.

To summarize, a multilevel surface relief profile uses a set of N masks to generate a surface profile that exhibits 2^N phase levels. The etch depth for the kth mask is $d_k = \lambda/2^k(n-1)$. In the next section, we describe a process for fabricating continuous DOE profiles in a single expose and etch cycle.

Continuous DOE Profiles Using Grayscale Lithography

Grayscale photolithography has become increasingly popular because of its flexibility and convenience in fabricating continuous diffractive profiles. Methods including directly writing partially exposed e-beam resist, resist reflow lithography, and using grayscale masks in a single step exposure. Here we discuss only the latter method because the processes are similar to those used in multiple mask fabrication. Thus, the same facilities can be used to produce both. In particular, we consider the application of high-energy-beam-sensitive (HEBS) glass to grayscale lithography and the fabrication of continuous profile DOEs.

The process for creating a diffractive element using HEBS glass is similar to fabricating a binary-phase element from a single mask. However, mask creation is markedly different. HEBS glass is a white crown glass within which a thin buried layer of silver ions is incorporated. The layer of silver ions is created via ion exchange in an aqueous acidic solution and essentially serves the same function as silver grains in photographic film. When the glass is exposed to a beam of electrons, the silver-containing crystal complexes are chemically reduced and the glass darkens. By controlling the electron dose, one controls the opacity of the glass to produce a grayscale pattern.

Although the attraction of single-step grayscale lithography is reduced costs in fabrication, the process incurs high calibration costs to insure quality pattern transfer through several steps. The parameter that has the greatest impact on the overall process is the desired size of the smallest diffractive feature. The minimum feature size affects the selection of the accelerating potential, beam current, addressing grid, and electron dose of the e-beam used to expose the HEBS glass.

The e-beam accelerating potential sets the resolution limit of the process and affects the sensitivity of the glass to the electron dose, the total charge per area incident upon the glass. For example, for a constant electron dose, increasing the accelerating potential increases the darkening response of the material. The large accelerating potential produces a large flux, which achieves the desired electron dose in a short period using energetic electrons. This, in turn, enhances the sensitivity of the glass. However, increasing the accelerating potential also increases scattering, which limits the minimum feature size achievable. Thus, low beam energies are generally used to write small features. But this, unfortunately, reduces the range of available optical densities because the material is less sensitive at low beam energies. To compensate for the reduced sensitivity, it is possible to increase the electron dose by increasing the exposure time. However, this undermines to some extent using a low energy beam to achieve small features due to broadening that results from long exposure times.

Once the accelerating potential for writing the mask is chosen, the beam current and addressing grid need to be specified. The beam current is related to the spot size and, therefore, affects directly the length of time to write the mask. High current levels produce large spots. Since write times can easily reach into

hours for large structures, high current levels are desirable to reduce them.

The addressing grid represents the physical grid that the electron beam is translated over in order to complete the exposure. This grid should be chosen as finely as possible, however, too fine a grid can significantly increase the required dwell time on each grid point. In this respect, a compromise is required. High beam currents produce short dwell times, which may necessitate using a coarse address grid to insure the beam is capable of performing the exposure.

Once the process parameters are set, calibration between the HEBS transmission and the electron dose is necessary. Analyzing the optical density of written test structures by some experimental means and determining its functional dependence on the electron dose completes the calibration. This allows one to specify the electron doses required to achieve the features dictated by the device design.

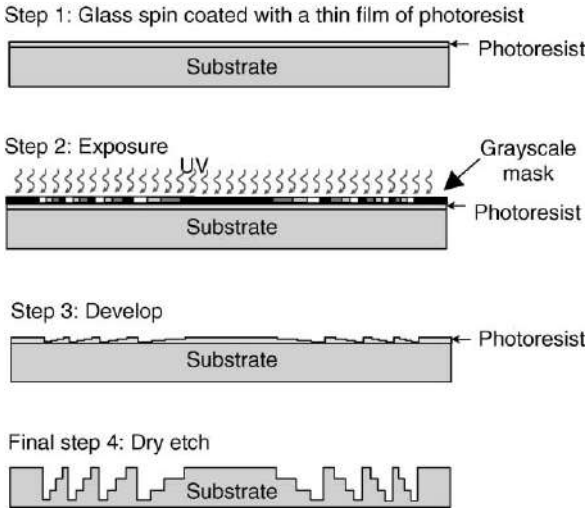

Figure 4 Fabrication of a continuous phase DOE using a grayscale mask.

If resolution requirements cannot be met through judicious selection of the accelerating potential, then some of the contrast in the mask can be sacrificed for an increase in available resolution. As we have mentioned, reducing the maximum optical density, or contrast, can alleviate scattering-induced broadening. It is possible, however, to compensate to some degree for the reduced contrast during the photolithography process, where UV exposure and develop times provide additional latitude in the pattern transfer process.

Figure 4 represents the fabrication of a continuous profile DOE using a grayscale mask. The substrate is spin-coated with a thin film of photoresist, placed in intimate contact with the grayscale mask, and UV illuminated to expose the resist film. The resist is developed and the pattern transferred into the substrate using either wet or dry etching techniques. Because the solubility of the resist is proportional to exposure, developing the photoresist produces a variable height profile.

Grayscale lithography is challenging because modern photoresists are also designed to be highly nonlinear yet linear response is required. Traditional photoresists are also designed to amplify mask contrast and produce the sharpest resist profile, thus they either dissolve away quickly and completely, or remain impervious to developer. The exposure threshold of traditional resists is, therefore, steep and considerable effort is required to

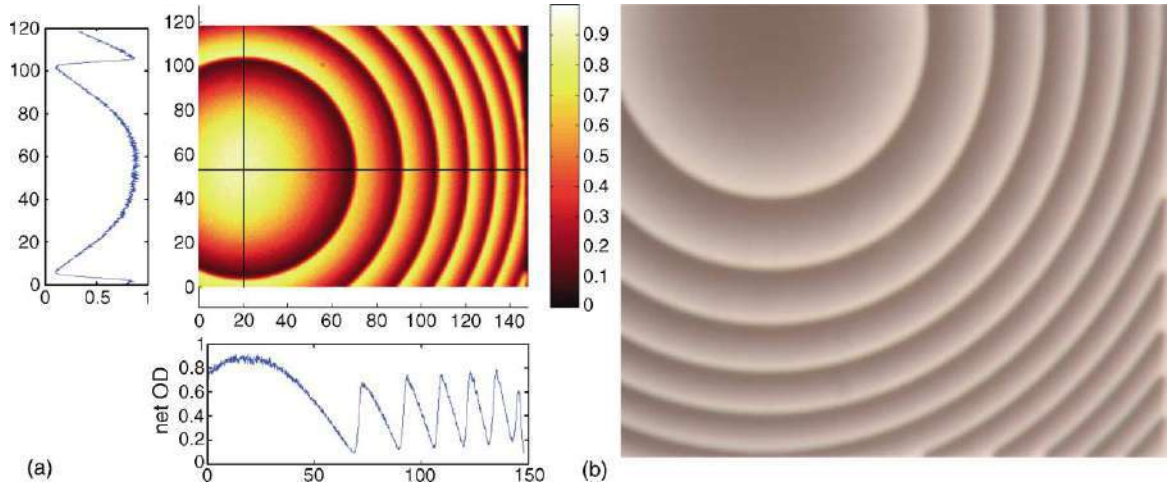

Figure 5 HEBS glass fabrication. (a) Data map of the optical density cross-section of an exposed HEBS grayscale mask and (b) its corresponding visible image.

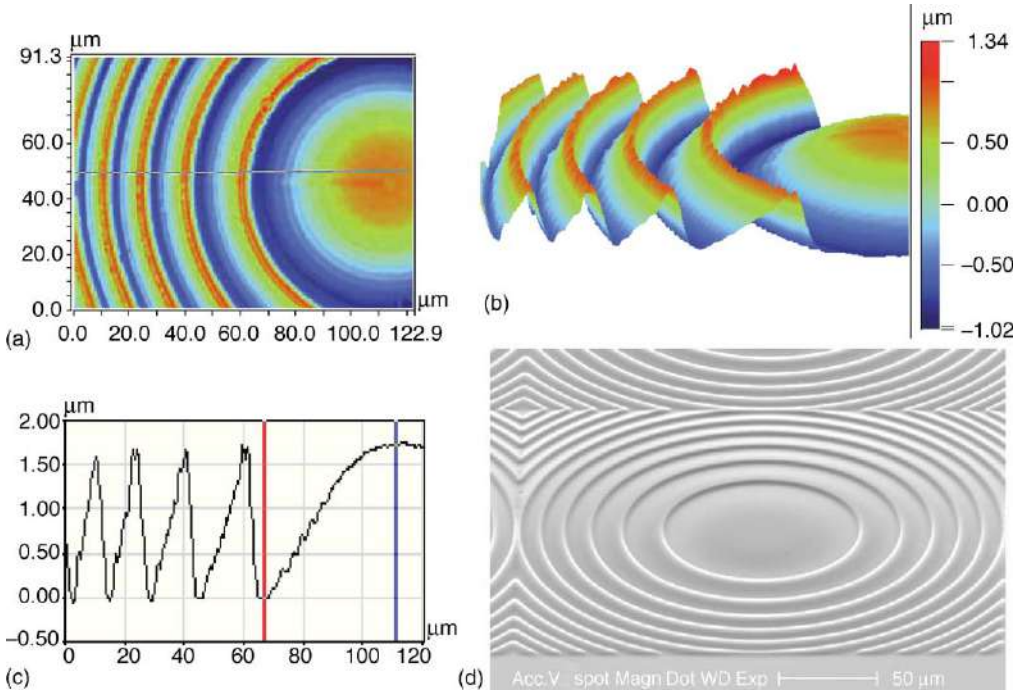

Figure 6 Characterization of a continuous phase diffractive lens. (a) Two-dimensional profile. (b) Three-dimensional profile. (c) Line scan of the two-dimensional plot. (d) Image of element from a scanning electron microscope.

insure that resist exposure is on the linear portion of the curve.

Although the resist exposure process is sensitive to all parameters, the behavior of a resist can be characterized using simulation models. The details of the photoresist models can be found elsewhere, but the simulations predict that the greatest utilization of the available resist occurs for short exposure times (seconds), and that these short exposures require long developing times (minutes). Further, the models predict that low mask contrast does not adversely affect the usable resist height. It is, therefore, not unreasonable to use a low contrast mask to achieve small features.

Through simulation it is possible to determine process parameters that maximize the linear response range of the photoresist. However, these parameters must be validated in the laboratory. The results presented in Figures 5 and 6 provide convincing evidence of the efficacy of these techniques.

Final Remarks

Advances in passive photonic devices have been made possible by leveraging fabrication processes from the electronics industry. Modifying standard photolithographic techniques and photoresist processes has produced an industrial capacity for fabricating diffractive elements. This capability will no doubt lead to future advances in telecommunications, signal processing, and imaging.

See also

Diffraction: Fresnel Diffraction. **Diffractive Systems:** Aberration Correction with Diffractive Elements. **Holography, Techniques:** Overview.

Further Reading

Dascher W, Long P, Stein R, Wu C and Lee SH (1997) Cost-effective mass fabrication of multilevel diffractive optical elements by use of a single optical exposure with a gray-scale mask on high-energy beam-sensitive glass. *Applied Optics* 36: 4675–4680.

Dillon T, Balcha A, Murakowski JA and Prather DW (2004) Continuous-tone grayscale mask fabrication using high-energy-beam-sensitive (HEBS) glass. *Journal of Microlithography, Microfabrication, and Microsystems*.

Fienup JR (1980) Iterative method applied to image reconstruction and to computer-generated holograms. *Optical Engineering* 19: 297–306.

Goodman JW (1968) *Introduction to Fourier Optics*. New York: McGraw-Hill.

Kohler M (1998) Etching in microsystem technology. In: Turunen J and Wyrowski F (eds) *Diffractive Optics for Industrial and Commercial Applications*. Berlin: Wiley-VCH.

LeCompte M, Gao X and Prather DW (2001) Photoresist characterization and linearization procedure for the grayscale fabrication of diffractive optical elements. *Applied Optics* 40: 5921–5927.

Leith EN and Upatnieks J (1962) Reconstructed wavefronts and communication theory. *Journal of the Optical Society of America A* 52: 1123–1130.

Lohmann AW and Paris DP (1967) Binary Fraunhofer holograms generated by computer. *Applied Optical* 6: 1739–1748.

Mait JN (1995) Understanding diffractive optical design in the scalar domain. *Journal of the Optical Society of America A* 12: 2145–2158.

Stern MB and Rubico-Jay T (1994) Dry etching for coherent refractive microlens arrays. *Optical Engineering* 33: 3547–3551.

Swanson GJ (1989) Binary optics technology: the theory and design of multi-level diffractive optical elements, Technical Report 854, MIT.

Swanson GJ (1991) Binary optics technology: theoretical limits on the diffraction efficiency of multi-level diffractive optical elements, Technical Report 914, MIT.

Swanson GJ and Veldkamp WB, (1990) High-efficiency, multilevel diffractive optical elements. United States Patent 4,895,790.

Wang MR and Su H (1998) Laser direct-write gray-level mask and one-step etching for diffractive microlens fabrication. *Applied Optics* 37: 7568–7576.

Wu CK (1994) High Energy Beam Sensitive Glasses. U.S. Patent 5,285,517.

Wyrowski F (1991) Digital holography as part of diffractive optics. *Report on Progress in Physics* : 1481–1571.

Diffractive Laser Resonators

U D Zeitner, Fraunhofer Institute of Applied Optics and Precision Engineering, Jena, Germany

F Wyrowski, Friedrich-Schiller University, Jena, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

All lasers consist of three basic functional components: an active medium, an optical arrangement to pump the active medium, and a resonator, which we consider in detail here. The basic function of the resonator is to back-couple a portion of the emitted laser light into the active medium for further amplification. Simultaneously, in steady-state laser operation, the geometry of the resonator defines the spatial field distribution of the generated light field.

In the most simple, and most common, case the laser resonator consists of two planar or spherical mirrors. This configuration was proposed in 1958 and has been treated extensively in the literature. However, these configurations are only special cases of a more general resonator concept in which arbitrary mirror surface profiles are used. Historically, resonators consisting of surface profiled mirrors played only a small role in laser physics. This was due, not only to difficulties in their numerical design and analysis, but also to the lack of suitable fabrication methods for the mirrors.

However, since the 1990s, these difficulties have been overcome and in 1992 it was demonstrated for the first time experimentally a stable CO_2 laser resonator that has a user-defined mode shape generated by surface structured mirrors.

In the following we derive the basic design rules for stable and unstable diffractive laser resonators and discuss the advantages and disadvantages of the technology.

Modes in Laser Resonators

Before presenting the details of diffractive laser resonators we look first at some general properties of the electromagnetic field inside a laser resonator. As is typical in laser resonator theory, we use scalar and paraxial descriptions of the fields.

The complex amplitude $U(x, y, z)$ represents one scalar, monochromatic component of the electromagnetic field. To handle wave propagation inside the resonator, we use a general operator formalism in which the influence of each optical component on the wave $U(x, y, z)$ is represented by an operator. One of the most important operators is the so-called round trip operator $\mathbf{Z}$, which describes the effect of a round trip inside the resonator on the wave $U(x, y, z)$.

If $U(x, y, z)$ denotes the wave at the beginning of a resonator round trip, the resulting wave $\tilde{U}(x, y, z)$ at its end is

$$\tilde{U}(x, y, z) = \mathbf{Z}U(x, y, z) \qquad [1]$$

With reference to Figure 1, the round trip operator can be decomposed into a series of operators that represent the different elements in the resonator. For the common two-mirror resonator a round trip from mirror 1 to mirror 2, reflection at mirror 2, back propagation to mirror 1, and final reflection at

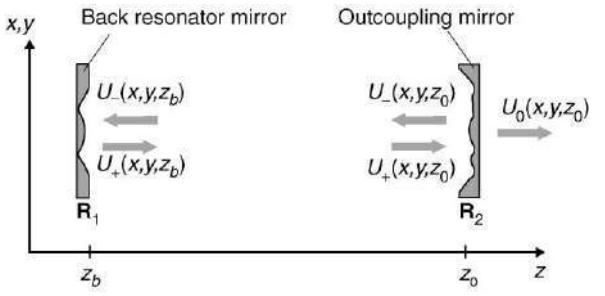

Figure 1 General diffractive laser resonator arrangement.

mirror 1, is represented by

$$\mathbf{Z} = \mathbf{R}_1\mathbf{P}_{\Delta z}\mathbf{R}_2\mathbf{P}_{\Delta z} \qquad [2]$$

The operators $\mathbf{R}_1$ and $\mathbf{R}_2$ represent the reflection at mirrors 1 and 2, respectively, and the propagation along the resonator length Δz is represented by the propagation operator $\mathbf{P}_{\Delta z}$.

In steady-state laser operation, the wave circulating inside the resonator reproduces itself, up to a constant factor γ, after a complete round trip:

$$\tilde{U}(x, y, z) = \gamma U(x, y, z) \qquad [3]$$

Together eqns [1] and [3] give the eigenvalue equation:

$$\gamma U(x, y, z) = \mathbf{Z}U(x, y, z) \qquad [4]$$

which, in general, has a set of solutions $U_p(x, y, z)$.

Thus, in steady-state, the waves $U_p(x, y, z)$ circulating inside the resonator are eigenfunctions of the round trip operator $\mathbf{Z}$ with the connected eigenvalue γ_p. The eigenfunctions of the resonator are called modes and the real-valued eigenvalues γ_p are connected with the round trip losses V, according to the equation:

$$V = 1 - |\gamma_p|^2 \qquad [5]$$

As we can see, the spatial properties of the modes are determined by the optical elements in the resonator. In the simple two-mirror resonator, the mirror profiles, which are related to the operators $\mathbf{R}_1$ and $\mathbf{R}_2$, can be used to specify the shape of the modes.

To derive the rules for resonator design we must look closer at the orthogonality of the resonator modes. The normalized eigenfunctions of the round trip operator $\mathbf{Z}$ do not obey the usual power-orthogonality relationship:

$$\int\int U_m(x, y, z)U_n^*(x, y, z)\,\mathrm{d}x\,\mathrm{d}y = \delta_{mn} \qquad [6]$$

but instead satisfy the biorthogonality relationship:

$$\int\int U_{m;+}(x, y, z)U_{n;-}(x, y, z)\,\mathrm{d}x\,\mathrm{d}y = \delta_{mn} \qquad [7]$$

The δ_{mn} is the so-called Kronecker symbol and $*$ denotes the complex conjugate. In eqn [7], $U_{m;+}(x, y, z)$ is the mode in plane z traveling in the $+z$ direction and $U_{m;-}(x, y, z)$ is the mode in the same plane z propagating in the opposite direction. The physical reason for the difference between eqns [6] and [7] are diffractive mode losses at limiting apertures. However, if a mode $U_{m;+}(x, y, z)$ travels without losses through the whole resonator, a power normalization:

$$\int\int U_m(x, y, z)U_m^*(x, y, z)\,\mathrm{d}x\,\mathrm{d}y = 1 \qquad [8]$$

is valid in all planes z. Equations [7] and [8] will be used in the next section to derive the design strategy for resonators with predefined fundamental modes.

If the mirrors, either planar or spherical, are assumed to be infinite in extent, it is possible to derive analytic solutions for the modes. This yields either the Gauss–Hermite or Gauss–Laguerre modes, which are orthogonal in the sense of eqn [6]. For resonators whose mirror shape differs from either a plane or sphere, the analytical calculation of its eigenfunctions or resonator modes is difficult and in most cases even impossible.

However, in such cases the analysis can be performed numerically. One of the most common methods is the Fox-Li-algorithm which uses the physical propagation of the modes through the resonator. The algorithm begins with a random field distribution in some plane of the resonator. This field is then numerically propagated through the resonator until a round trip is completed. The effect of the resonator optical elements must be accounted for in this procedure. At the end of the resonator round trip, the starting field will have changed due to diffraction and serve as a new starting wave for a succeeding round trip calculation. This procedure is repeated until a static field distribution is obtained, which is then an eigenfunction of the resonator.

Design Principles

The condition in eqn [7] on the biorthogonality of resonator modes, together with the eqn [8] valid for lossless modes, can be used to develop principles to design resonators with an arbitrary fundamental mode shape. Consider a stable laser resonator characterized by negligible diffraction loss of the fundamental mode. Following the discussion above, its lossless modes must simultaneously satisfy

eqns [7] and [8]. A comparison of both equations yields:

$$U_+(x, y, z) = U_-^*(x, y, z) \qquad [9]$$

For reasons of simplicity we neglect here, and in the following, the subscript m for the mode order if only one mode is considered. Because this equation must hold in all planes z, it is satisfied in the mirror resonator planes z_0 and z_b. Further, by definition of the operators $\mathbf{R}_1$ and $\mathbf{R}_2$:

$$U_+(x, y, z_b) = \mathbf{R}_1 U_-(x, y, z_b) \qquad [10]$$

$$U_-(x, y, z_0) = \mathbf{R}_2 U_+(x, y, z_0) \qquad [11]$$

The combination of eqns [10] and [11] with eqn [9] yields the following expressions for the mirror operators:

$$\mathbf{R}_1 = \frac{U_+(x, y, z_b)}{U_+^*(x, y, z_b)} \qquad [12]$$

$$\mathbf{R}_2 = \frac{U_-(x, y, z_0)}{U_-^*(x, y, z_0)} \qquad [13]$$

Equations [12] and [13] imply that lossless modes are phase-conjugated on reflection at the resonator mirrors. Because we have not made any assumptions on the shape of the mode, these equations represent a general property of lossless resonator modes. Furthermore, they provide a general design rule for the construction of stable laser resonators with predefined fundamental mode shape.

If one considers limiting apertures inside the resonator, an equivalent statement for lossy modes can be made by making assumptions about phase conjugation in a lossy system and showing that they lead to a contradiction. Without loss of generality, we assume an aperture on the outcoupling resonator mirror that clips the mode. We now consider the wave $U_-(x, y, z_0)$ after reflection at this mirror and assume this wave is still phase-conjugated after propagation through the resonator during its reflection at the back resonator mirror according to:

$$U_+(x, y, z_b) = U_-^*(x, y, z_b) \qquad [14]$$

If now the wave $U_+(x, y, z_b)$ is back propagated to the outcoupling mirror, we end up with the wave $U_+(x, y, z_0)$ which must, because of the phase conjugation in z_b, have the same amplitude as the wave $U_-(x, y, z_0)$ we started with. As a result, because both waves are connected to each other via the reflection operator of the outcoupling mirror by eqn [11], $\mathbf{R}_2$ cannot influence the amplitude distribution which, however, is a contradiction to the assumption of a limiting aperture. Thus, we finally can state that lossy modes do not have the property to be phase conjugated during their reflection at both mirrors. As a consequence, the design of diffractive mirrors for unstable laser resonators, in general characterized by energy outcoupling via diffractive losses, is more complex than simply calculating a phase-conjugating mirror structure for stable laser resonators.

Up to now, design methods described in the literature for unstable laser resonators with user-defined output beam shapes have all been based on an iterative calculation of the mirror reflection function. Because a detailed description of these methods is rather complex, we restrict ourselves to a sketch of the principal ideas behind them.

Without loss of generality, we consider an unstable laser resonator from which the light is coupled out by a mirror with spatially varying transmission and reflection. The second mirror is fully reflective within its aperture. The particular application defines only few design constraints which are in most cases the light wavelength, the dimensions of the resonator, and the shape of the outcoupled beam $U_0(x, y, z_0)$. As a consequence of the outcoupling principle, the shape of the mode inside the resonator is not finally defined by the outcoupled beam. In regions where the outcoupling mirror is reflective and not transmissive, the mode amplitude can be chosen freely. Also the amplitude of the mode in the plane of the back resonator mirror is not fixed but influenced by the amplitude and phase of the mode in the plane of the outcoupling mirror.

The general starting situation for an unstable resonator design is sketched in Figure 2. The design procedure must yield the complex amplitudes of four waves: the mode directly in front of the outcoupling mirror $U_-(x, y, z_0)$, the mode after reflection at the outcoupling mirror $U_+(x, y, z_0)$, and the complex amplitudes of the forward and reverse modes in the plane of the back resonator mirror $U_+(x, y, z_b)$ and $U_-(x, y, z_b)$, respectively. Since we assume no losses at

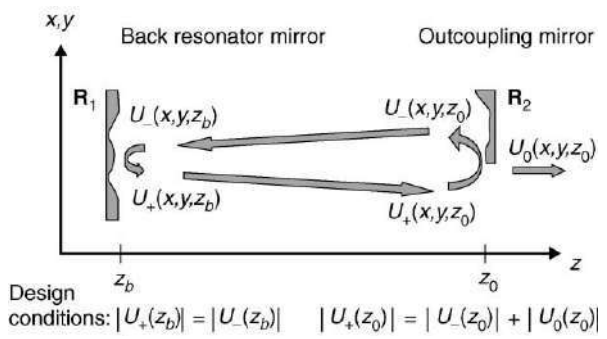

Figure 2 Starting situation for design of an unstable diffractive laser resonator.

this mirror, the amplitudes must satisfy the condition:

$$|U_+(x, y, z_b)| = |U_-(x, y, z_b)| \qquad [15]$$

If we further assume no absorption at the outcoupling mirror, the amplitudes of the three waves in z_0 are related according to:

$$|U_-(x, y, z_0)| + |U_0(x, y, z_0)| = |U_+(x, y, z_0)| \qquad [16]$$

To design the resonator one starts with the choice of initial intensity distributions for the wavefields $U_-(x, y, z_0)$ and $U_+(x, y, z_b)$. In order to obtain a good filling of the available mode volume, a homogeneous intensity across the mirror apertures is recommendable. To avoid strong diffraction effects at the mirror edges, the corresponding intensity distributions should be smoothed out. Together with the desired outcoupled beam amplitude $|U_0(x, y, z_0)|$ one can calculate the two remaining amplitudes for the wave fields $U_-(x, y, z_b)$ and $U_+(x, y, z_0)$ according to eqns [15] and [16].

An iterative algorithm, such as the well known Gerchberg–Saxton algorithm, can be used to find the phases of the beams $U_-(x, y, z_0)$ and $U_+(x, y, z_0)$ that insure the wavefields have almost equal amplitudes once they are propagated or back-propagated, respectively, to the plane z_b. In order to find wavefields that satisfy eqns [15] and [16], two successive iteration runs with different constraints on the complex amplitudes have to be performed. During the first iteration run no change in the initial amplitudes in plane z_0 are allowed. The iteration only look for phases for $U_-(x, y, z_0)$ and $U_+(x, y, z_0)$ which result in the amplitude $|U_+(x, y, z_b)|$ once they are propagated and back-propagated, respectively, to plane z_b. Usually the results of a Gerchberg–Saxton algorithm, with such constraints, are wavefields $U_+(x, y, z_b)$ and $U_-(x, y, z_b)$ with similar but not equal amplitude. Thus, in order to determine wavefields that satisfy eqn [15], a second iteration run with the Gerchberg–Saxton algorithm has to be added in which slight variations in wavefield amplitudes also in plane z_0 are allowed. Because the application of the constraints to ensure convergence of the iteration is rather complex, we do not go into details here but refer the reader to the Further Reading at the end of this article.

Upon completion of the iterations, the mirror reflectance operators $\mathbf{R}_1$ and $\mathbf{R}_2$ are calculated according to:

$$\mathbf{R}_1 = \frac{U_+(x, y, z_b)}{U_-(x, y, z_b)} \qquad [17]$$

$$\mathbf{R}_2 = \frac{U_-(x, y, z_0)}{U_+(x, y, z_0)} \qquad [18]$$

The design method described so far yields a mathematical expression for the resonator mirror reflection operator. Practical realization of the resonator requires this expression be converted into a description of a physical property of the mirror. In most practical cases this is the surface profile. Since the design method considers only phase-only functions, the physical description of the reflection operators are:

$$\begin{aligned} \mathbf{R}_1 &= \exp[-i\phi_1(x, y)] \text{ and} \\ \mathbf{R}_2 &= \exp[-i\phi_2(x, y)] \end{aligned} \qquad [19]$$

Using the so-called thin-element-approximation, the surface profiles $h_1(x, y)$ and $h_2(x, y)$ of the two mirrors follow directly from the phases $\phi_1(x, y)$ and $\phi_2(x, y)$:

$$\begin{aligned} h_1(x, y) &= -\frac{\lambda}{4\pi}\phi_1(x, y) \text{ and} \\ h_2(x, y) &= -\frac{\lambda}{4\pi}\phi_2(x, y) \end{aligned} \qquad [20]$$

with λ being the laser wavelength. The expressions in eqn [20] contain a factor of 1/2 due to reflection. These surface profiles are commonly fabricated by microlithographic techniques. For stable laser resonators, where the mirrors perform a static phase conjugation, the mirror profile is a surface of equal phase of the fundamental mode.

Use of General Resonator Modes

Applications of diffractive laser resonators can be classified into beam shaping for a particular task external to the laser and resonator optimization via modifying the eigen-space of the round trip operator.

Internal beam shaping for external tasks is relevant to applications where it is advantageous to use a fixed beam profile other than a Gaussian, as is the case for instance in materials processing. Shaping the beam internally, avoids the energy losses usually associated with external beam shaping techniques. However, lasers that have a nonGaussian outcoupled beam profile are often less flexible in their general applicability.

However, there is considerable potential for diffractive laser resonators to enhance resonator characteristics and thereby improve the lasing operation itself. The conventional Gaussian-shaped fundamental mode in customary laser resonators varies considerably in intensity from its central maximum to its outer wings. The transfer of energy from the pumped active medium into the propagating mode is dependent upon the local transverse intensity and is

optimal for homogeneous illumination. As a result, the efficiency with which energy is transferred into the fundamental mode can be increased by using diffractive resonator mirrors to match the shape of the lateral fundamental mode to that of the cross-section of the active medium.

An important resonator property, especially in high power laser systems, is the difference of round trip losses V between the fundamental mode and the higher, usually undesired, laser modes. This difference is referred to as modal discrimination. With no limiting apertures, the wavefronts of all Gauss–Hermite or Gauss–Laguerre modes of common laser resonators are phase conjugated upon reflection at the spherical mirrors. As we have shown above, there is a tight connection between the phase conjugation property of resonator modes and their property of having no round trip losses. Thus, the phase conjugation property of Gauss–Hermite and Gauss–Laguerre modes directly implies small round trip losses. In fact, in resonators without apertures they all have an eigenvalue $\gamma_p = 1$ and according to eqn [5] the round trip losses are zero. High-order modes, whose diameters are large, can be suppressed by decreasing the size of the mirror aperture, however, because the difference in mode diameter between modes of neighboring order m is small, the achievable discrimination is limited to a maximum ratio $V_{m+1}/V_m \approx 10$. The structured surface profile of the mirrors in a diffractive laser resonator whose fundamental mode is nonGaussian is tightly adapted to the phase of the design mode. The wavefront of the higher-order modes in these resonators often do not fit the mirror profile and it can be observed that their round trip losses are significantly higher. Thus, diffractive laser resonators usually show an improved modal discrimination compared to resonators with spherical mirrors. This effect can be further improved by inserting into the resonator additional phase plates with diffuser-like properties.

Shaping the surface profile of resonator mirrors or insertion of additional phase plates significantly improves the flexibility for the resonator design. As is demonstrated by the examples below, there are resonator configurations in which high modal discrimination or adaptation of the mode volume to the cross-section of the active medium are combined with a Gaussian shape of the outcoupled beam.

In the design procedure described so far, the resonators were treated as passive optical systems. The propagation of the mode through the resonator has not been influenced by the active medium. However, in real laser systems, this simplification is often not applicable. Due to the pumping process and the connected heat insertion, the refractive index distribution inside the active medium is not homogeneous, which affects the propagation of the modes. Furthermore, changes or fluctuations in the pump energy and cooling mechanisms introduce temporal changes in the optical function associated with the refractive index of the active medium. Further, although static or averaged, optical power of the active medium are included in the round trip operator and calculation of the mirror profiles, experiments with lasers that have diffractive resonator mirrors indicate that the temporal fluctuations can strongly influence the shape and round trip losses of the fundamental resonator mode. In general, diffractive laser resonators are more sensitive to deviations from design than conventional resonators.

Examples

We consider as an example the adaptation of fundamental mode shape to the cross-section of the active medium. In solid state lasers, such as Nd:YAG lasers, the active medium is often a crystal rod with a circular cross-section. Instead of using this sharp-edged cross-section shape as intensity distribution for the desired fundamental mode, a smoothed version, which is approximated by a super-Gaussian intensity distribution in the center plane of the rod, is chosen to be the fundamental mode. The super-Gaussian intensity distribution has a more desirable propagation characteristic than a sharp circular one. However, even these modes do not exhibit the advantageous property of propagation-invariant beam shape characteristic of Gaussian beams.

In our first example, we consider the design of a stable diffractive laser resonator with a super-Gaussian fundamental mode shape inside the active medium and a Gaussian beam coupled out from the resonator. The resonator configuration is based upon a beam shaping arrangement for a Gauss to super-Gauss beam transformation. Its optical setup is sketched in Figure 3. In the plane of the outcoupling

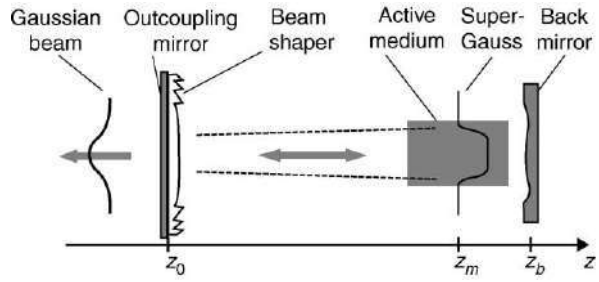

Figure 3 Setup for the Gauss-to-super-Gauss resonator.

mirror z_0, the fundamental mode has a Gaussian amplitude $U_g(x, y, z_0)$. A phase element, which is located directly in front of the outcoupling mirror, shapes the mode amplitude into a super-Gaussian one $U_{sg}(x, y, z_m)$ in the center plane z_m of the active medium. For this special beam shaping problem it is possible to calculate a phase plate surface profile that transforms the beams without energy losses. The corresponding calculation procedure is described in the Further Reading section at the end of this article.

The mirror profiles have to be realized to satisfy the previously derived phase conjugation property for the described beam. Since the Gaussian beam has its minimum waist at the outcoupling mirror, its phase is planar in that plane. Thus, to insure phase conjugation a plane mirror is all that is required in front of the beam shaping phase plate. In the plane of the back resonator mirror the beam has a modulated wavefront due to its transformation from a Gaussian to a super-Gaussian amplitude distribution. Thus, a surface modulated mirror profile is required to phase conjugate this beam during reflection. Figures 4a and b represent the phase profile of the beam shaping element and the back resonator mirror, respectively, for the following set of parameters: laser wavelength $\lambda = 1064$ nm, radius of the Gaussian and super-Gaussian beams in the plane z_0 and z_m, respectively, $w_g = 1$ mm and $w_{sg} = 1$ mm, order of the super-Gauss $n = 10$, distance $z_m - z_0 = 200$ mm, and resonator length $z_b - z_0 = 250$ mm. The amplitude distribution for the fundamental mode in this resonator simulated with the Fox–Li-algorithm is shown in Figure 4c for the plane of the outcoupling mirror and in Figure 4d for the center of the active medium.

In our second example, the flexibility of diffractive laser resonators is demonstrated by a resonator with a user-defined far-field intensity distribution. In our case the far-field intensity $|U_f(x, y, z_f)|^2$ is shaped like the Greek letter lambda, λ. Because we are concerned only with the intensity of the wavefield, the phase $\phi_f(x, y, z_f)$ of the far-field $U_f(x, y, z_f)$ can be chosen freely. As shown in the following, this phase freedom can be used to improve the interaction of the mode with the active medium, as was the case in our first design example. Again, the mode shape in the center of the active medium has a super-Gaussian intensity distribution, which insures high extraction efficiency. We used an iterative Fourier-transform algorithm to calculate a far-field phase that yields a nearly super-Gaussian intensity distribution in the outcoupling plane of the resonator. The resulting wave:

$$U_0(x, y, z_0) = a_0(x, y, z_0) \exp[i\varphi_0(x, y, z_0)] \qquad [21]$$

is generated directly as the laser output. The amplitude $a_0(x, y, z_0)$ is generated as a mode inside the cavity using a surface-modulated back resonator mirror. This performs a static phase conjugation of the desired mode in the mirror plane. The phase $\varphi_0(x, y, z_0)$ is provided by an additional element on the outside of the outcoupling mirror. A sketch of the resonator configuration is shown in Figure 5.

For the experimental realization of this example the thermal index gradient of the active medium was included into the resonator design and the two diffractive elements were fabricated on a quartz substrate using electron beam lithography and ion-beam etching. The measured mode shape in the plane of the outcoupling mirror and the far-field

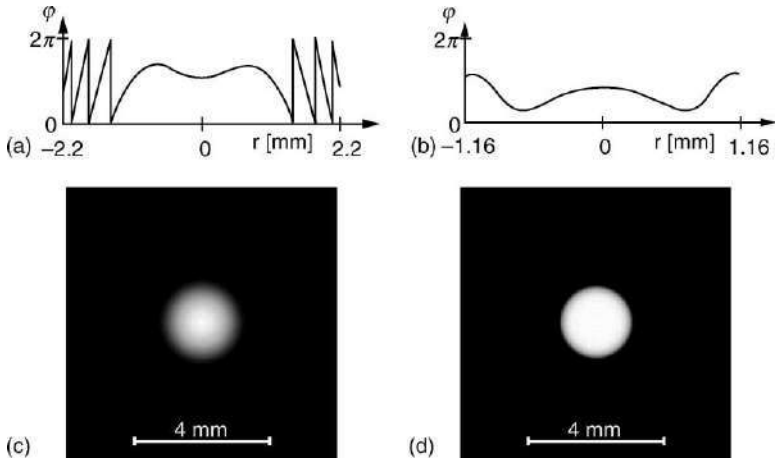

Figure 4 Gauss-to-super-Gauss resonator: (a) phase of beam shaping element; (b) phase of back resonator mirror, simulated amplitude distributions for the fundamental resonator mode; (c) in the plane of the outcoupling mirror; and (d) in the center of the active medium.

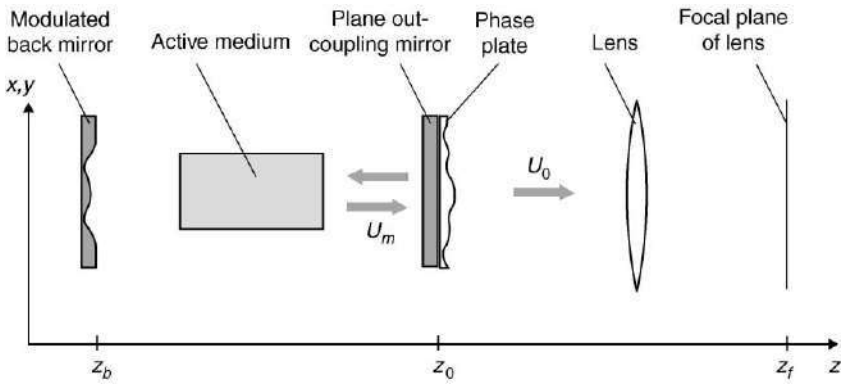

Figure 5 Configuration for resonator internal generation of a user defined far-field intensity distribution. (With permission from Zeitner UD, Wyrowski F and Zellmer H (2000) External design freedom for optimization of resonator originated beam shaping. *IEEE Journal of Quantum Electronics* 36: 1105–1109. © 2001 IEEE.)

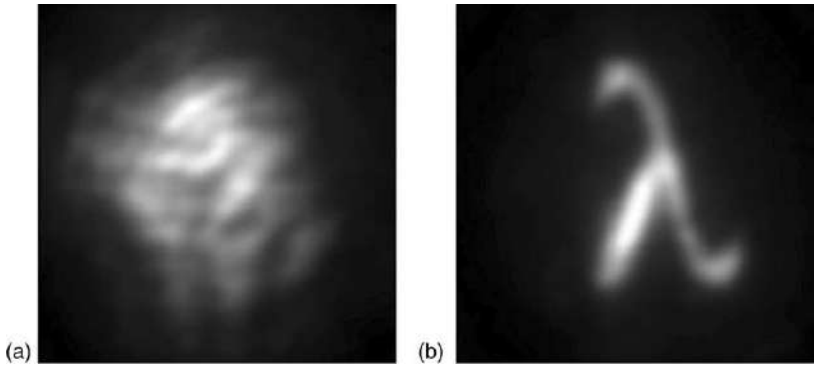

Figure 6 Measured mode intensities for the resonator of **Figure 5**. (a) In the plane of the outcoupling mirror; and (b) in the far-field of the resonator. (With permission from Zeitner UD, Wyrowski F and Zellmer H (2000) External design freedom for optimization of resonator originated beam shaping. *IEEE Journal of Quantum Electronics* 36: 1105–1109. © 2000 IEEE.)

intensity distribution are shown in **Figures 6a** and **b**, respectively. The mode in the outcoupling plane of the resonator fills an area of about 2 mm diameter with nearly homogeneous intensity. The far-field shows the desired λ-shaped intensity distribution. The variations in intensity are due to the nonlinear response of the active medium. As mentioned above, the response can be improved if these effects are included in the design of the resonator mirrors.

Conclusion

Diffractive laser resonators are a generalization of common spherical mirror resonators. The use of arbitrary resonator mirror profiles and additional internal phase plates offer a huge flexibility for the generation of user defined mode shapes and optimization of resonator properties like extraction efficiency and modal discrimination. The drawback of the use of structured mirrors is an increased adjustment effort, because the diffractive mirrors have more degrees of freedom to be aligned. Furthermore, diffractive resonators tend to be more sensitive against changes of the optical properties of the active medium than spherical mirror resonators. Such changes may occur due to temporal fluctuations in pump power or cooling power.

Most promising applications for diffractive resonators are:

- lasers for materials processing which require a fixed intensity pattern for the sample treatment and a fixed laser power;
- high efficient homogeneous laser illumination, for example, for data projection or sensing applications; and
- suppression of higher order modes and maintaining single-mode operation for high brightness lasers.

Some example applications and solutions have been presented here which cannot be achieved by conventional resonator optics. For stable laser resonators, the phase-conjugation property of the

modes is used for a straightforward calculation of the required diffractive mirror profiles. Unstable resonator geometries with diffractive outcoupling require more complex iterative design procedures. However, especially for high-power and high-gain lasers, the latter sometimes have more advantageous properties, such as better discrimination of undesired higher-order modes and more simple mirror cooling options.

See also

Lasers: Carbon Dioxide Lasers. **Phase Control:** Phase Conjugation and Image Correction.

Further Reading

Aagedal H, Wyrowski F and Schmid M (1997) Paraxial beam splitting and shaping. In: Turunen J and Wyrowski F (eds) *Diffractive Optics for Industrial and Commercial Applications*, pp. 165–188. Berlin: Akademie Verlag.

Belanger PA, Lachance RL and Pare C (1992) Super-Gaussian output from a CO_2 laser by using a graded-phase mirror resonator. *Optics Letters* 17: 739–741.

Fox A and Li T (1961) Resonant modes in a maser interferometer. *Bell Systems Technical Journal* 40: 453–488.

Gerchberg R and Saxton W (1972) A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* 35: 237–246.

Herzig H (1997) *Micro Optics*. London: Taylor & Francis.

Leger JR (1997) Diffractive laser resonators. In: Turunen J and Wyrowski F (eds) *Diffractive Optics for Industrial and Commercial Applications*, pp. 165–188. Berlin: Akademie Verlag.

Leger JR, Chen D and Dai K (1994) High modal discrimination in a Nd:YAG laser resonator with internal phase gratings. *Optics Letters* 19: 1976–1978.

Makki S and Leger JR (2001) Mode shaping of a graded-reflectivity-mirror unstable resonator with an intracavity phase element. *IEEE Journal of Quantum Electronics* 37: 80–86.

Siegman AE (1986) *Lasers*. Mill Valley, CA: Univ. Science Books.

Zeitner UD and Wyrowski F (2001) Design of unstable laser resonators with user-defined mode shape. *IEEE Journal of Quantum Electronics* 37: 1594–1599.

Zeitner UD, Wyrowski F and Zellmer H (2000) External design freedom for optimization of resonator originated beam shaping. *IEEE Journal of Quantum Electronics* 36: 1105–1109.

Diffractives in Animals

A R Parker, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Structural coloration involves the selective reflectance of incident light by the physical nature of a structure. Although the color effects often appear considerably brighter than those of pigments, structural colors often result from completely transparent materials. In addition, animal structures can be designed to provide the opposite effect to a light display: they can be antireflectors, causing 'all' of the incident light to penetrate a surface (like glass windows, smooth surfaces cause some degree of reflection).

The study of natural structural colors began in the seventeenth century, when Hooke and Newton correctly explained the structural colors of silverfish (Insecta) and peacock feathers (Figure 1), respectively. Nevertheless, accurate, detailed studies of the mechanisms of structural colors did not begin until the introduction of the electron microscope in 1942.

Structural colors generally may be formed by one of three mechanisms: thin-film reflectors, diffraction gratings, or structures causing scattering of light waves. This is a simplistic characterization resulting from the application of geometric optics to the reflective elements. Some structures, however, rather fall between the above categories, such as photonic crystals (Figure 2) – these become apparent when quantum optics is considered. In some cases, this has led to confusion over the reflector type. For example, the reflectors in some scarab beetles have been categorized by different authors as multilayer reflectors, three-dimensional diffraction gratings, photonic crystals, and liquid crystal displays. Perhaps all are correct! Since the above categories are academic, I place individual cases of structural colors in their most appropriate, not unequivocal, category. Also, in this article I will employ geometric optical treatments only.

The array of structural colors found in animals today results from millions of years of evolution. Structures that produce metallic colors have also been identified in extinct animals. Confirmation of this

Figure 1 Newton's subject.

fact, from ultrastructural examination of exceptionally well-preserved fossils, such as those from the Burgess Shale (Middle Cambrian, British Columbia), 515 million years old, permits the study of the role of light in ecosystems throughout geological time, and consequently its role in evolution. In some cases, such as for ostracod crustaceans, diffraction gratings have been found to increase in their optical sophistication throughout 350 million years of evolution.

Multilayer Reflectors

Light may be strongly reflected by constructive interference between reflections from the different interfaces of a stack of thin films (of actual thickness d) of alternately high and low refractive index (n). For this to occur, reflections from successive interfaces must emerge with the same phase and this is achieved when the so-called 'Bragg Condition' is fulfilled. The optical path difference between the light reflected from successive interfaces is an integral number of wavelengths and is expressed in the equation:

$$2nd \cos \Theta = (m + 1/2)\lambda \qquad [1]$$

from which it can be seen that the effect varies with angle of incidence (Θ, measured to the surface normal), wavelength (λ), and the optical thickness of the layers (nd). There is a phase change of half a wavelength in waves reflected from every low to high refractive index interface only. The optimal narrowband reflection condition is therefore achieved where the optical thickness (nd) of every layer in the stack is a quarter of a wavelength. In a multilayer consisting of a large number of layers with a small variation of index the process is more selective than one with a smaller number of layers with a large difference of index. The former therefore gives rise to more saturated colors corresponding to a narrow spectral bandwidth and these colors therefore vary more with a change of angle of incidence. Both conditions can be found in animals – different colored effects are appropriate for different functions under different conditions. For an oblique angle of incidence, the wavelength of light that interferes constructively will be shorter than that for light at normal incidence. Therefore, as the angle of the incident light changes, the observed color also changes (**Figure 3**).

Single layer reflectors are found in nature, where light is reflected, and interferes, from the upper and lower boundaries (**Figure 4**). A difference in the thickness of the layer provides a change in the color observed from unidirectional polychromatic light. The wings of some houseflies act as a single thin film and appear different colors as a result of this phenomenon. A single quarter-wavelength film of guanine in cytoplasm, for example, reflects about 8% of the incident light. However, in a multilayer reflector with ten or more high index layers, reflection efficiencies can reach 100%. Thus, animals possessing such reflectors may appear highly metallic.

The reflectance of the multilayer system increases very rapidly with increasing number of layers. If the dimensions of the system deviate from the quarter-wave condition (i.e., nd is not equal for all layers), then the reflector is known as 'nonideal' in a theoretical sense (may be 'ideal' for some natural situations). 'Nonideal' reflectors have a reduced proportional reflectance (not always a significant reduction) for a given number of layers and this

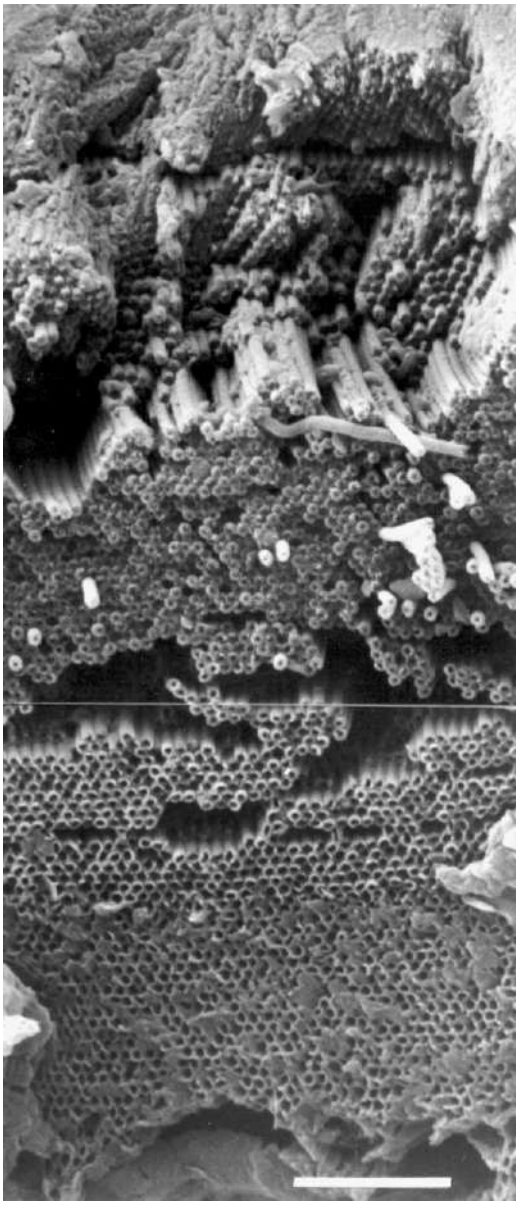

Figure 2 Scanning electron micrograph of a cross-section of the wall of a cylindrical spine of the sea mouse *Aphrodita* sp. (Polychaeta). The wall is composed of small cylinders with varying internal diameters (increasing with depth in the stack), arranged in a hexagonal array, that form a photonic crystal fiber. The opal-type photonic crystal – hexagonally packed nano-spheres (250 nm diameter) – are also found in animals (e.g. weevils). Scale bar represents 8 μm.

reflectance has a narrower bandwidth. A narrow bandwidth, less conspicuous reflection is sometimes selected for in animals, this will be discussed later in this article. Multilayer reflectors polarize light incident at Brewster's angles. This is about 54° for a quarter-wave stack of guanine and cytoplasm. At very oblique angles, all wavelengths are strongly reflected.

Multilayer reflectors are common in animals. They are usually extra-cellular, produced by periodic secretion and deposition, but sometimes occur within cells. Guanine ($n = 1.83$) is a common component in invertebrate reflectors because it is one of very few biological materials with a high refractive index and is readily available to most invertebrates as a nitrogenous metabolite. However, arthropods, including insects, crustaceans, and spiders, have largely ignored guanine in favor of pteridines. Also surprising is that the reflector material of closely related species, e.g., the molluscs *Pecten* (scallop) and *Cardium* (cockle), may differ.

Multilayers produce effects in beetle cuticle from highly metallic colors ('ideal' system) to rather dull greens ('nonideal' system in combination with scattering; Figure 5), and colors from the wings of many butterflies. Often in butterflies, layers of chitin ($n =$ about 1.56) are supported by vertical vanes of the scales. Air ($n = 1.0$) fills in the spaces and provides the alternate layers of the system. Under white light at normal incidence, the blue color of the butterfly *Arhopala micale* turns to green when the air is replaced by acetone. This is due to an increase in the refractive index of the low n layer (the actual thickness of this layer remains the same but the optical thickness increases) that effectively makes the system less 'ideal'. If the refractive indices of both layers were made equal, there would be no internal interference. A layer of melanin often underlies the reflector and intensifies the metallic colored effect by absorbing the transmitted portion of incident light. For example, in beetles, the elytra of *Anopognathus parvulus* appears metallic gold, green, or yellow in reflected light, and diffuse brown in transmitted light. Individual butterfly scales have been examined in detail to reveal a number of variations of quarter-wave stacks, sometimes in combination with other optical structures, to provide a range of colored effects.

The crustaceans *Limnadia* (Conchostraca), *Tanais tennicornis* (Tanaidacea), *Ovalipes molleri* (Decapoda), and the males of *Sapphirina* (Copepoda) all bear multilayer reflectors in their cuticles, in different forms (Figure 6). In contrast to the usual continuous thin layers, male sapphirinids have ten to fourteen layers of interconnecting hexagonal platelets within the epidermal cells of the dorsal integument. The reflector of *O. molleri* comprises layers that are corrugated and also slightly out of phase (Figure 7). At close to normal incidence this structure reflects red light, but at an angle of about 45° blue light is reflected. The corrugation, however, functions to broaden the reflectance band, at the expense of reducing the intensity of reflection.

Figure 3 Butterfly (*Parthenos sylvia* (Nymphalidae) from Malaysia), showing the change in color with angle of wings. This animal may have evolved a temporal signal rather than a static one.

A broadband wavelength-independent reflectance, appearing silver or mirror-like to the human eye, can be achieved in a multilayer stack in at least three ways in invertebrates (Figure 8). These are (a) a composite of regular multilayer stacks each tuned to a specific wavelength, (b) a stack with systematically changing optical thicknesses with depth in the structure, termed a 'chirped' stack, and (c) a disordered arrangement of layer thicknesses about a mean value, termed a 'chaotic' stack (Figure 8). The nauplius eye of the copepod *Macrocyclops* (Crustacea) has regularly arranged platelets ~100 nm thick in stacks of 20–60, achieving the first condition. Silver beetles and the silver and gold chrysalis of butterflies in the genera *Euoplea* and *Amauris*, owe their reflection to the second condition. The mirror-like reflectors in the scallop *Pecten* eye comprise alternating layers of cytoplasm ($n = 1.34$) and guanine crystals ($n = 1.83$) and approximate an 'ideal' quarter-wave system in the same manner as within fish skin using the third mechanism. The ommatidia of the superposition compound eyes of *Astacus* (Crustacea) are lined with a multilayer of isoxanthopterin (a pteridine) crystals, which again fall into the third category. Multilayer reflectors can also be found in the eyes of certain spiders, butterflies, and possibly flies, where they assist vision, as discussed further below.

Squid and cuttlefish, for example, possess mirror-like reflectors in photophores (light organs) and

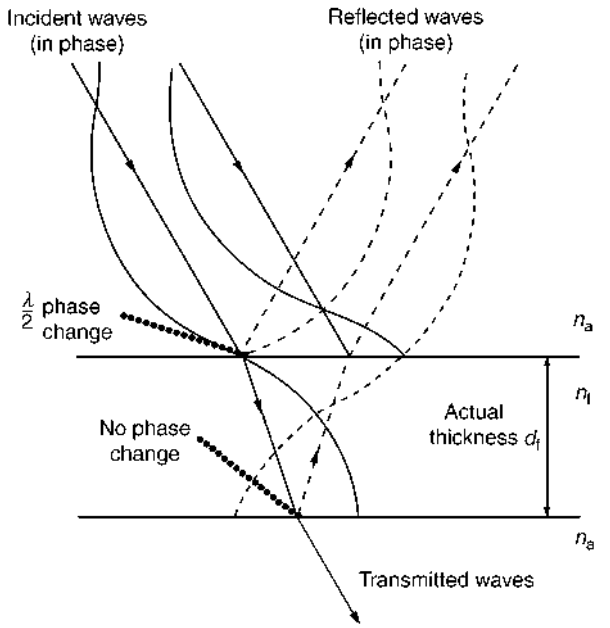

Figure 4 Schematic diagram of thin-film reflection. The direction of wave (straight line) and profile of electric (or magnetic) component are illustrated. Incident waves are indicated by a solid line, reflected waves by broken lines. Refraction occurs at each media interface. The refractive index of the film (n_f) is greater than the refractive index of the surrounding medium (n_a). Constructive interference of the reflected waves is occurring. As the angle of incidence changes, different wavelengths constructively interfere. At normal incidence constructive interference occurs where $n_f \times d_f = \lambda/4$.

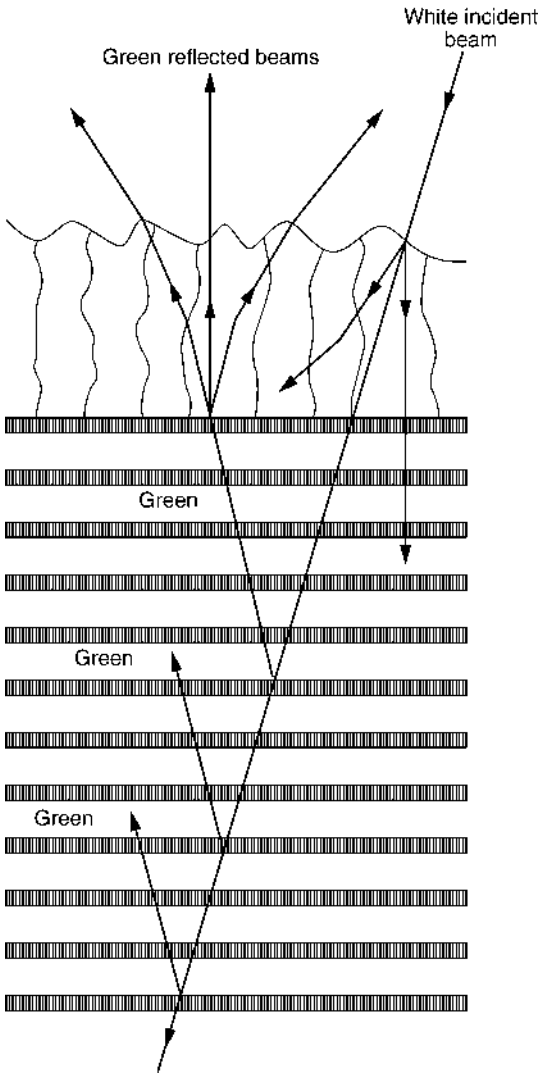

Figure 5 Generalized diagram of a multilayer reflector in the cuticle of the green beetle *Calloodes grayanus* (high refractive index material is shown shaded). Outer layer causes scattering.

iridophores. Iridophores are cells that, in this case, contain groups (iridosomes) of flexible layers of thin lamellae with cytoplasm between, forming a quarter-wave stack. The platelets of both squids and octopods develop from the rough endoplasmic reticulum and are separated by extracellular space. Euphausiid crustaceans possess photophores with very elaborate mirror-like reflectors. Up to 60 dense layers, about 70 nm thick and 75–125 nm apart, are formed from the aggregation of granules (probably a type of chitin) and surround the main photogenic mass. An intricate ring consisting of very flattened cells, forming the dense layers of a multilayer reflector (about 175 nm thick, separated by 90 nm) surrounds the lens of the photophore and reflects blue light at acute angles of incidence.

Dead invertebrates may not display their original colors. Following death, one (or both) of the layers in a multilayer reflector may become gradually reduced. For example, water may be lost from the system. This occurs in beetles of the genus *Coptocycla*; their brassy yellow color quickly changes through green, blue, and violet until the brown of melanin is finally observed. The color progression may subsequently be reversed by water uptake. This is an important consideration when examining fossils for multilayer reflectors. Multilayer reflectors have been identified in the shells of some ammonites, preserved as fossils.

Diffraction Gratings

When light interacts with a periodic surface consisting for example of a series of parallel grooves, it may be deviated from the direction of simple transmission or reflection. For this to happen, light which is scattered or diffracted from successive grooves should be out of phase by integral values of 2π. This occurs when, for a given direction of propagation, the optical path difference via successive grooves is $M\lambda$, where M is an integer known as the circle number.

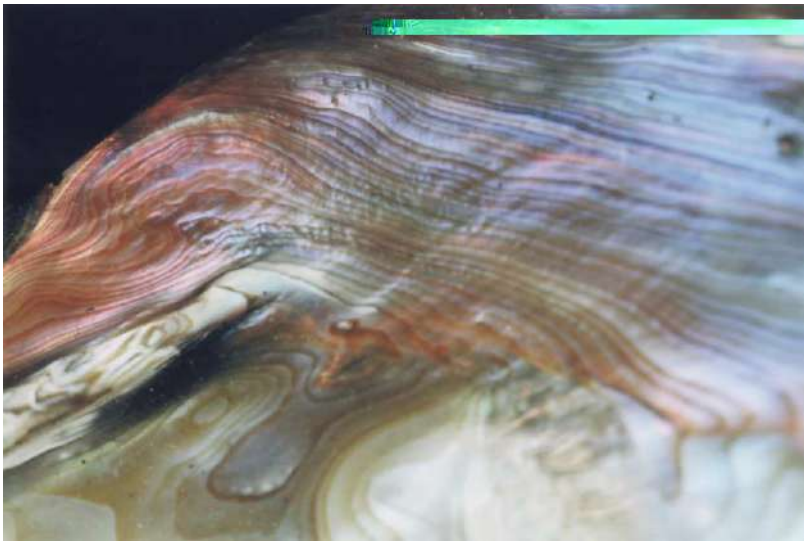

Figure 6 Bivalve shell with outer pigment removed to reveal an incidental multilayer reflector. Layers evolved to provide strength and prevent crack propagation.

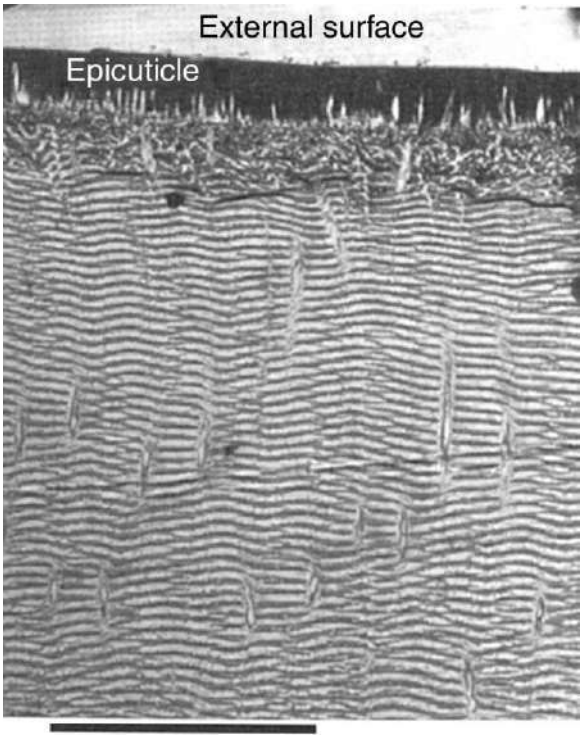

Figure 7 Transmission electron micrograph of a multilayer reflector in the cuticle of a swimming paddle of the crab *Ovalipes molleri* (Crustacea: Decapoda). Layers of reflector are slightly sinuous and out of phase; note the unusual side branches of each high refractive index (dark) layer, which provide support for these solid layers within a liquid matrix. Scale bar represents 5 μm.

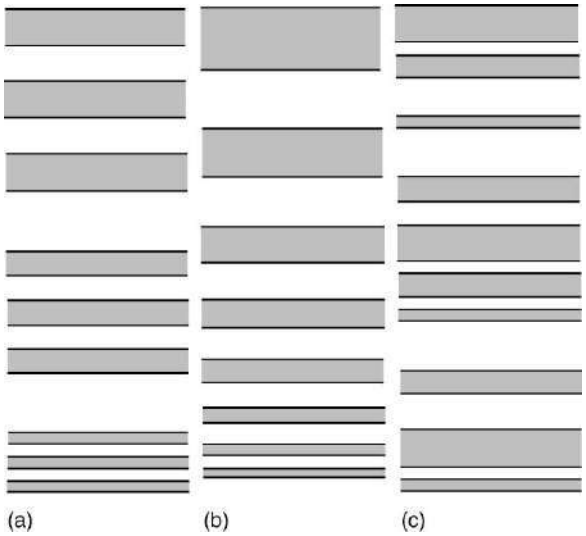

Figure 8 Three ways of achieving a broadband, wavelength-independent reflector in a multilayer stack (high refractive index material is shown shaded). (a) Three quarter-wave stacks, each tuned to a different wavelength. (b) A 'chirped' stack. (c) A 'chaotic' stack.

This may be expressed by the grating equation:

$$2d(\sin\alpha - \sin\beta) = M\lambda \qquad [2]$$

where α and β are angles of incidence and diffraction, and d is the period.

As with multilayers, the effect gives rise to coloration because different wavelengths are diffracted into different directions. Although the effect changes with angle of incidence, it is less critical than it is with thin films and the visual appearance is different. For a parallel beam of white light incident upon a multilayer, one wavelength will be reflected as determined by the Bragg Condition. The same beam

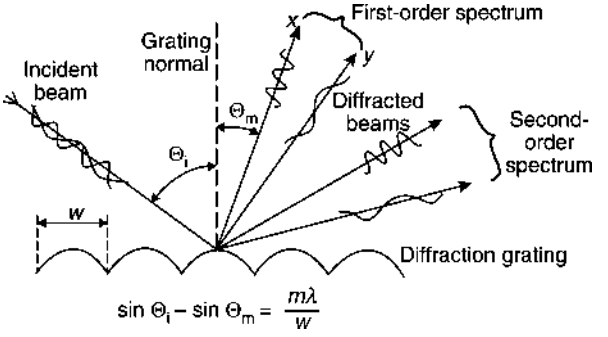

Figure 9 Reflection-type diffraction grating (w = periodicity, m = order of reflected beam). For white light of fixed angle of incidence (Θ_i), the color observed is dependent on the point of observation (e.g., violet light can be seen at point x, red at point y, within the first order spectrum).

incident upon a grating will be dispersed into spectra (e.g., Figure 9). The complete spectrum reflected nearest to the perpendicular (grating normal) is the first order. The first-order spectrum is reflected over a smaller angle than the second-order spectrum, and the colors are more saturated and appear brighter within the former. Diffraction gratings have polarizing properties, but this is strongly dependent on the grating profile.

Diffraction gratings were believed to be extremely rare in nature, but have recently been revealed to be common among invertebrates. They are particularly common on the seate or setules (hairs) of Crustacea, such as on certain first antennal setules of male Myodocopina ostracods or 'seed shrimps' (Crustacea) (Figure 10). Here, the grating is formed by the external surface of juxtaposed rings with walls circular in cross-section (Figure 11). The width of the rings, and consequently the periodicity of the grating, is about 700 nm in *Azygocypridina lowryi*. Different colors are viewed with varying angles of observation under a fixed light source. The ostracod *Euphilomedes carcharodonta*, for example, additionally houses a diffraction grating on the rostrum, a continuous flattened area of the carapace that is corrugated to form periodic ridges. The dark brown beetle *Serica sericea*, bears gratings on its elytra with 800 nm periodicity, which causes a brilliant iridescence in sunlight.

Many polychaetes possess gratings on their setae (hairs). For example, the opheliid *Lobochesis longiseta* bear gratings with periodicities in the order of 500 nm (Figure 12), appearing iridescent. The wings of the neurochaetid fly *Neurotexis primula* bear diffraction gratings only on their dorsal surfaces, and the iridescent effect remains after the insect is gold coated for electron microscopy. These gratings cause iridescence with a higher reflectance than the

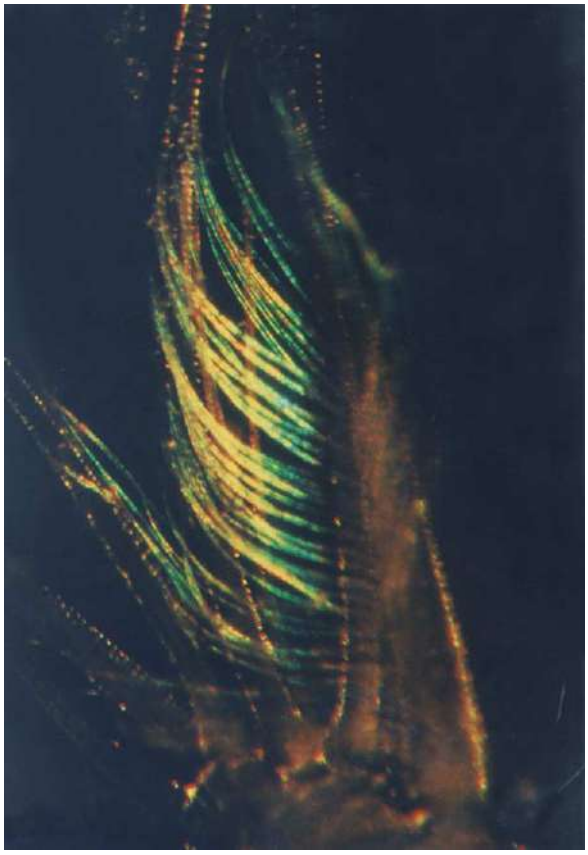

Figure 10 Diffraction gratings on the halophores (hairs) on the first antenna of the seed shrimp *Azygocypridina lowyri* (Ostracoda).

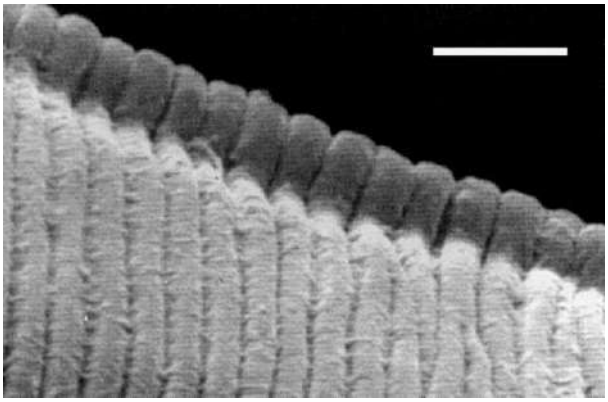

Figure 11 Scanning electron micrograph of diffraction grating on a halophore (setule) of *Azygocypridina lowryi* (Ostracoda). Scale bar represents 2 μm.

iridescence of the membranous wings of other insects, which reflect light by interference. Iridescence caused by interference disappears after gold coating because transmission of light through the outer surface is prevented.

Very closely spaced, fine setules may also form the ridges of a diffraction grating. Cylindroleberidid

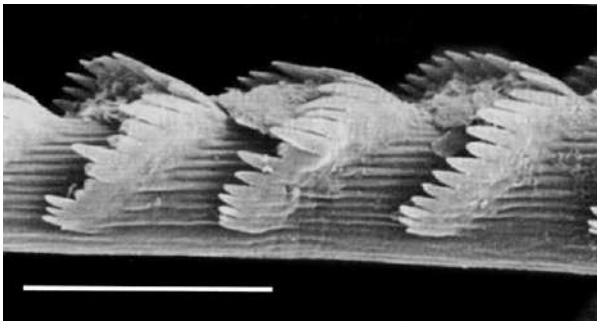

Figure 12 Scanning electron micrograph of diffraction gratings on a hair of *Lobochesis longiseta* (Polychaeta). Scale bar represents 10 μm.

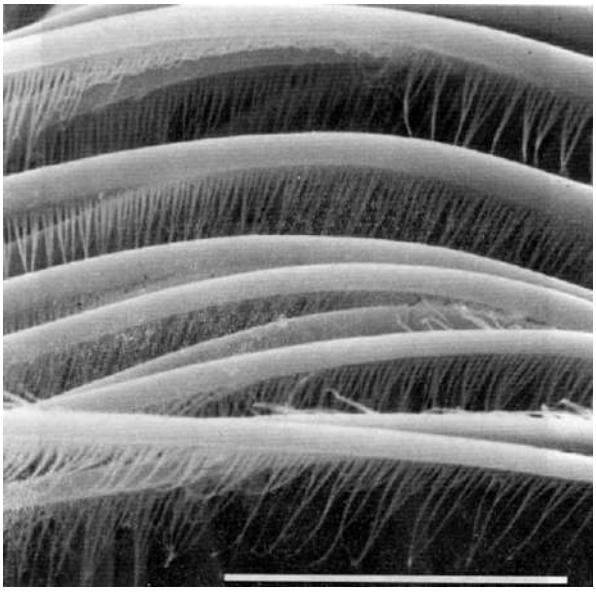

Figure 13 Scanning electron micrograph of *Tetraleberis brevis* (Ostracoda), hairs of comb of maxilla; setules (orientated vertically) form the ridges of a grating. Scale bar represents 2 μm.

ostracods (seed-shrimps) possess a comb on their maxilla bearing numerous setules on each seta (hair), collectively forming a grating with a periodicity of about 500 nm (Figure 13).

The 'helicoidal' arrangement of the microfibrils comprising the outer 5–20 μm of the cuticle (exocuticle) of certain scarabeid beetles, such as *Plusiotis resplendens*, also gives rise to metallic colors. Here, the fibrils are arranged in layers, with the fibril axis in each layer arranged at a small angle to the one above, so that after a number of layers the fibrillar axis comes to lie parallel to the first layer. Thus, going vertically down through the cuticle, two corresponding grating layers will be encountered with every 360° rotation of the fibrils – the 'pitch' of the system. Polarized light encounters an optically reinforcing plane every half turn of the helix. The system can be treated as a three-dimensional diffraction grating, with a peak reflectance at $\lambda = 2nd$, where d is the separation of analogous planes, or half the pitch of the helix. The diffracted light resembles that from a linear grating except for the polarization; the three-dimensional grating reflects light that is circularly or elliptically polarized. It should be noted that the diffracted color does not depend on the total film thickness as it does in interference, but on the layer repeat distance within the film as in the diffraction grating (analogous to 'liquid crystals').

When each groove of a grating is so formed that it independently, by means of geometrical optics, redirects the light in the direction of a chosen diffracted order, it is known as a blazed grating. In a blazed reflection grating, each groove consists of a small mirror (or prism) inclined at an appropriate angle (i.e., the grating has a 'saw-tooth' profile). Blazed gratings have been identified on the wing scales of the moth *Plusia argentifera* (Plusinae).

When the periodicity of a grating reduces much below the wavelength of light, it becomes a zero-order grating and its effect on light waves changes. This difference in optical effect occurs because when the periodicity of the grating is below the wavelength of light the freely propagating diffracted orders are suppressed and only the zero-order is reflected when the illumination is normal to the plane of the grating. To describe accurately the optical properties of a zero-order grating, rigorous electromagnetic theory is required. In contrast to gratings with freely propagating orders, zero-order structures can generate saturated colors even in diffuse illumination. Such structures occur on the setae (hairs) of the first antenna of some isopod crustaceans, such as the giant species of *Bathynomus*. Here, there are diffracted orders and the spectral content of the light within the grating is controlled by the groove profile. In an optical system that only accepts the zero-order, what is seen is white light minus that diffracted into the ± 1 orders.

A zero-order grating can cause total transmission (i.e., there is no reflection). Such antireflective structures are found on the corneal surfaces of each ommatidium (visual unit) in the eye of *Zalea minor* (Diptera) (Figure 14). The periodicity of the corneal gratings of this fly is 242 nm. Another form of antireflection grating is formed on the transparent wings of the hawkmoth *Cephonodes hylas* on the corneal surface of each visual unit (ommatidium) of the eyes of moths and butterflies (Figure 15). Here, optical-impedance matching is achieved by means of a hexagonal array of tapered cylindrical protuberances, each of about 250 nm diameter, thus forming a 'tri-grating' with grooves transecting at 120°.

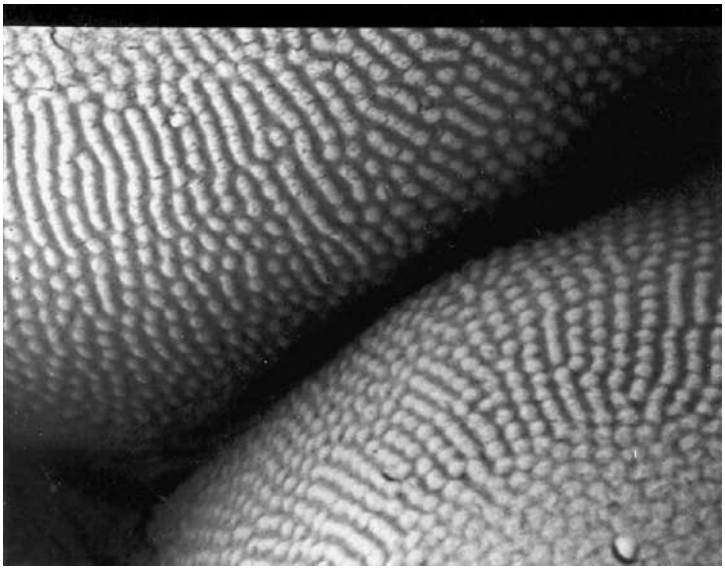

Figure 14 Scanning electron micrograph of the compound eye of *Zalea minor* (Diptera). The corneal surface at the junction of two ommatidia is shown with antireflection gratings (positioned on sloping regions, i.e., at high angles to the normal of the eye) phasing into moth-eye-type protuberances. Scale bar represents 2 μm.

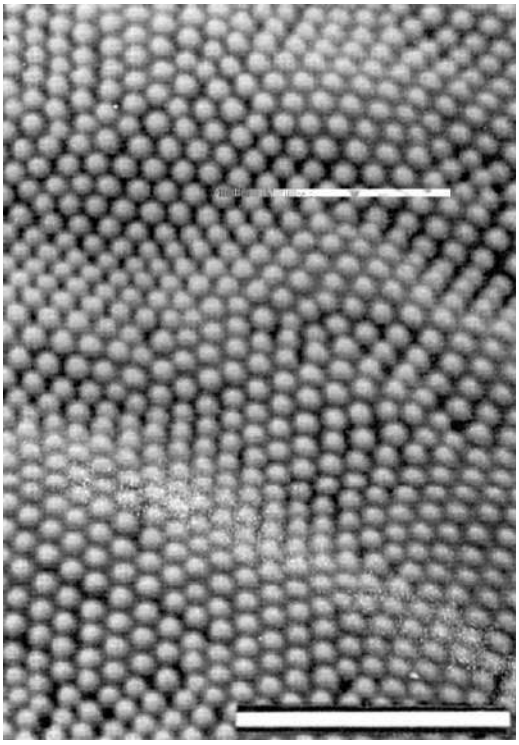

Figure 15 Tri-grating (antireflector) on the corneal surface of a butterfly eye. Scale bar represents 2 μm.

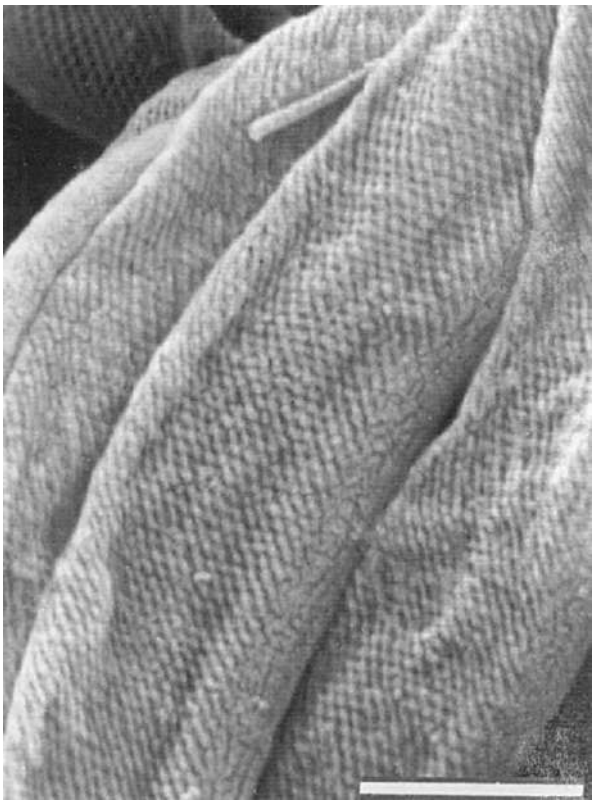

Figure 16 Bi-grating on the callynophores (hairs) of the amphipod crustacean *Waldeckia australiensis*. Scale bar represents 2 μm.

The protuberances provide a graded transition of refractive index between the air and the cornea/wing. Hence the refractive index at any depth is the average of that of air and the corneal/wing material.

The grooves of a grating may also create parallel rows in two directions, forming a bi-grating. Bi-gratings can be found in some crustaceans and flies. In the amphipod crustacean *Waldeckia australiensis*, two effectively superimposed gratings subtend angles of about 60/120° (Figure 16).

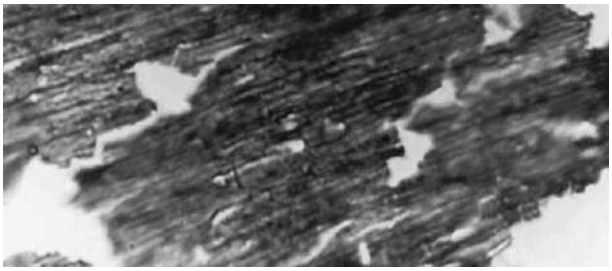

Figure 17 Light micrograph of the surface of a hair of *Canadia spinosa* (Middle Cambrian, 515 million years old, British Columbia), isolated by acid maceration of the rock matrix, showing gratings running longitudinally. Periodicity 0.9 μm.

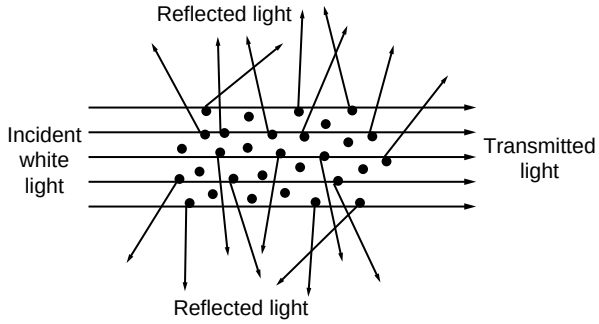

Figure 18 Scattering of white light by small particles (represented by black circles). The reflected light is randomly directed at different angles.

In terms of fossil colors, diffraction gratings are responsible for the nacreous luster of brachiopod shells, such as those from the Devonian, around 360 million years old. Here, tabular aragonite platelets averaging 600 nm in thickness, each comprising a linear diffraction grating, form layers and consequently a three-dimensional diffraction grating. Antireflective, zero-order gratings have been identified on the eye of an Eocene fly, 45 million years old, preserved in Baltic amber. Linear diffraction gratings causing color have been discovered on the sclerites of *Wiwaxia corrugata* from the Burgess Shale (Middle Cambrian, 515 million years old) (Figure 17). This finding has led to the 'light switch theory', in that the evolution of the first eye caused the big bang in animal evolution.

Scattering (Incoherent Reflectors)

Simple, equal scattering of all spectral wavelengths results in the observation of a diffuse white effect. This commonly arises from the effects of a nonperiodic arrangement of colloidally dispersed matter where the different materials involved have different refractive indices (Figure 18), or from solid colorless materials in relatively concentrated, thick layers. In the colloidal system, the particles are larger than the wavelength of light and can be thought of as mirrors oriented in all directions. The reflection is polarized unless the incident light is at normal incidence on the system and, in the colloidal system, spherical or randomly arranged particles are involved.

The colloidal system involves either a gas-in-solid, gas-in-liquid, liquid-in-liquid (emulsions), or solid-in-liquid. For example, the gas-in-liquid system is partly responsible for the white body and/or tentacles of certain anemones. Light is reflected and refracted at the surfaces of the particles of matter or spaces (with dimensions >1 μm), regardless of the color of the materials involved (except opaque brown and black compounds, such as melanin). In insects, the materials involved typically have very low transparencies.

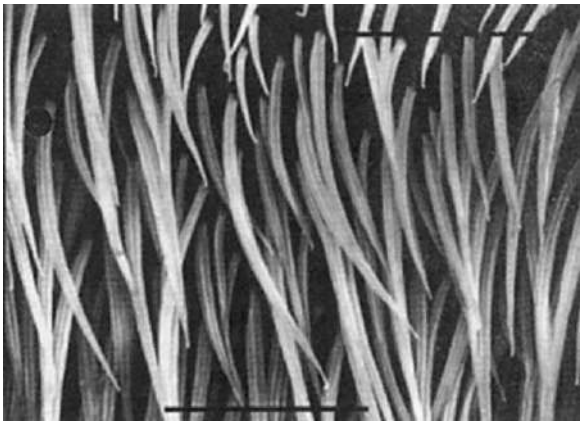

Figure 19 Scanning electron micrograph of a section of a white reflecting patch on the abdomen of the fly *Amenia* sp., showing closely packed, randomly arranged hairs that scatter incident light in all directions (Figure 11). Scale bar represents 20 μm.

An unordered (as opposed to periodic) group of closely spaced setae (hairs), such as those in patches on the fly *Amenia* sp., may form a white reflection via random scattering or reflection (Figure 19). However, if the arrangement becomes periodic to some degree, a diffraction grating may be formed, such as the grating of *Tetraleberis brevis* (Ostracoda) (Figure 13).

From some scales of butterfly wings, light is scattered uniformly and completely in all directions, due to the chaotic disposition of the surfaces. Matt or pearly whites may be observed depending on the complexity of the arrangement of the structures, which affects the relative degree of scattering. The structures may be so small that the molecular topography of the surface has an effect. The chromatic effects of the butterfly scales are greatly intensified if a dark, absorbing pigment screen lies beneath. This screen prevents reflection of white or

colored light from the background that would dilute or alter the color. Additionally, if a dark pigment such as melanin is interspersed with the scattering elements, the reflection will appear a shade of gray or brown. The cells of the reflector in the photophore of a beetle ('fire-fly') are packed with sphaerocrystals of urate that cause a diffuse reflection (Figure 20).

Reflection and refraction that occur at the interfaces of strata with different refractive indices may result in the display of white light. The degree of whiteness depends upon the difference in refractive indices. This mechanism is evident in the shells of many lamellibranch molluscs. Between the outer, often pigmented layer and the mantle, is a thick middle layer of crystalline calcium carbonate. The inner surface of this (nacreous) layer is lined with multiple laminations of the same salt. In most species these laminations are sufficiently thick (>1 μm) to render the inner lining white, although in some species they become thin so as to form a multilayer reflector. Calcium carbonate similarly produces whiteness in Foraminifera and in calcareous sponges, corals, echinoderms, and crustacean cuticles. Also, in the class of white solids is silica in diatom tests and skeletons of hexactinellid sponges.

Other forms of scattering also exist and result in a blue colored effect (red when the system is viewed in transmission). Tyndall or Mie scattering occurs in a colloidal system where the particle size approximates the wavelength of light. Here, diffraction is important. Rayleigh scattering occurs in molecules in a two-photon process by which a photon is absorbed and raises the molecule to an excited electronic state from which it re-radiates a photon when it returns to the ground state. Diffraction is not involved here.

Tyndall scattered light is polarized under obliquely incident light. The intensity of the resultant blue is increased when it is viewed against a dark background, such as melanin. The relative sizes of particles determine the shade of blue. If the particles responsible for the scattering coalesce to form particles with a diameter greater than about 1 μm, then white light is observed (see above). A gradation from blue to white scattering ('small' to 'large' particles) occurs on the wings of the dragonfly *Libellula pulchella*.

Scattered blues can also be found in other dragonflies. In the aeschnids and agrionids, the epidermal cells contain minute colorless granules and a dark base. The males of libellulids and agrionids produce a waxy secretion that scatters light similarly over their dark cuticle. The green of the female *Aeschna cyanea* is the combined result of Tyndall scattering and a yellow pigment, both within the epidermal cells (degradation of the yellow pigment turns the dead dragonfly blue).

Scattered blues are also observed from the skin of the cephalopod *Octopus bimaculatus*, where a pair of ocelli are surrounded by a broad blue ring. Blue light is scattered from this region as a result of fine granules of purine material within cells positioned above melanophore cells. The color and conspicuousness of the ring are controlled by the regulation of the melanophores, by varying the distribution of melanin

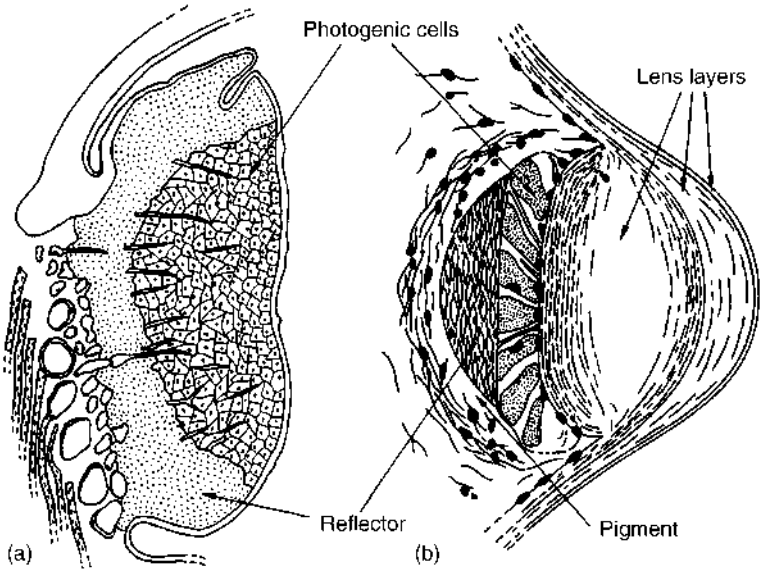

Figure 20 Sections of photophores (bioluminescent organs), external surface of animals on the right. (a) The 'fire-fly' beetle *Pyrophorus* sp., abdominal photophore of male – reflector is based on a scattering system. (b) The shrimp *Sergestes prehensilis*, showing lens layers and absorptive pigment – reflector is based on a multilayer type.

and consequently the density of the absorbing screen. The squid *Onychia caribbaea* can produce rapidly changing blue colors similarly. The bright blue patterns produced by some nudibranch molluscs result from reflecting cells containing small vesicular bodies, each composed of particles about 10 nm in diameter and therefore appropriate for Rayleigh scattering.

See also

Diffraction: Diffraction Gratings. **Scattering:** Scattering Theory.

Further Reading

Fox DL (1976) *Animal Biochromes and Structural Colours*. Berkeley, CA: University of California Press.

Fox HM and Vevers G (1960) *The Nature of Animal Colours*. London: Sidgwick and Jackson Ltd.

Gale M (1989) Diffraction, beauty and commerce. *Physics World* 2: 24–28.

Harvey EN (1952) *Bioluminescence*. New York: Academic Press Inc.

Herring PJ, Cambell AK, Whitfield M and Maddock L (1990) *Light and Life in the Sea*. Cambridge, UK: Cambridge University Press.

Hutley MC (1982) *Diffraction Gratings*. London: Academic Press.

Land MF and Nilsson D-E (2002) *Animal Eyes*. Oxford, UK: Oxford University Press.

Pain S (1998) Seeing the light. *New Scientist* 2161: 42–46.

Parker AR (1999) Light-reflection strategies. *American Scientist* 87: 248–255.

Parker AR (2003) *In the Blink of an Eye*. London: Simon and Schuster/Cambridge (USA): Perseus Publishing.

Parker AR, Welch VL, Driver D and Martini N (2003) An opal analogue discovered in a weevil. *Nature* 426: 786–787.

Parker AR, McPhedran RC, McKenzie DR, Botten LC and Nicorovici N-AC (2001) Aphrodite's iridescence. *Nature* 409: 36–37.

Microstructure Fibers

R S Windeler, OFS Laboratories, Murray Hill, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Microstructure fibers have unique properties that cannot be obtained from traditional fibers (i.e. all glass, doped silica fibers) and can deliver functionalities superior to many of today's best transmission and specialty fibers. Their unique properties are obtained from an intricate cross-section of high and low index regions that traverse the length of the fiber. The vast majority of fibers consist of silica for the high-index region and air for the low-index region. These fibers are known by several different names including microstructure fiber, holey fiber, and photonic crystal fiber. The term microstructure fiber is used in this chapter because it is the most generic, encompassing all the fiber types.

Microstructure fiber properties vary greatly and are determined by the size and arrangement of air holes in the fiber. For example, fibers have been fabricated such that the numerical aperture of the core or cladding approaches one. They have been fabricated with unique dispersion profiles, such as a zero-group velocity dispersion in the near visible regime or an ultraflat dispersion over hundreds of nanometers wide. They have been fabricated to generate a super continuum over two octaves wide. Some guide light with an air core via bandgap guidance. In addition, the air holes in these microstructure fibers can be filled with highly tunable materials, giving one the capability of controlling the fiber properties for use in energy efficient devices.

Microstructure fibers can have doped regions like traditional fibers. These hybrid fibers combine the benefits of traditional and microstructure fibers. The doped core typically guides most of the light, but its guidance properties can be strongly influenced by the air regions. Applications for hybrid fibers include dispersion compensation, polarization maintaining, and bend insensitive fibers.

Loss is always an important factor when determining whether a microstructure fiber will compete with or replace a traditional optical fiber. The index profiles that make these fibers so special can also lead to fibers that have an inherently high loss at connections or along the length of the fiber. Loss at connections typically occurs due to mode mismatch between the two fibers, undesired hole collapse, or poor alignment. Loss along the fiber length can occur due to impurities, hole surface roughness, or poor confinement.

The dispersion profile (zero dispersion wavelength, dispersion slope, normal or anomalous) of microstructure fibers can be optimized more than a traditional fiber, because the hole size and placement

can be arranged to make the cladding index strongly wavelength dependent. Dispersion causes a pulse to spread because the phase velocity is wavelength dependent. Normal dispersion is when longer wavelengths travel faster than shorter wavelengths. Anomalous dispersion is just the opposite – shorter wavelengths travel faster than longer wavelengths. Dispersion determines the amount of interaction between different wavelengths. For applications such as transmission fiber, one wants little interaction and for high nonlinear applications, one wants significant interaction.

This chapter examines the methods for fabricating microstructure fibers and reviews the different types of microstructure fibers and their properties. Fibers that guide by total internal reflection are reviewed separately from bandgap guided fibers because their properties are significantly different. Lastly, a description and example of an active device is presented in which the air regions of a microstructure fiber are filled with controllable material.

Microstructure Fiber Fabrication

At first glance, fabrication of microstructure fibers looks similar to traditional fibers in that the fibers are fabricated in a two-step process. First a preform is fabricated and then it is drawn (stretched) into a fiber. However, when the process is examined in more detail, both the preform fabrication and draw depart significantly from traditional methods. First, a novel preform fabrication method is used to incorporate air holes that run the length of the preform. Second, novel drawing procedures are used to keep the air holes open as the preform diameter is reduced by several orders of magnitude during draw.

Preform Fabrication

Microstructure preforms are cylinders of amorphous material usually less than a few centimeters in diameter that have an index profile running the length of the preform similar to the desired fiber. The most common preform consists of air and a single material such as silica, polymer, or high nonlinear glass. These air-containing preforms are relatively easy to fabricate but are difficult to draw because the structures can deform.

The basic parameters of the fiber are usually determined by the geometry of the holes in the preform. These include hole diameter (d), hole position, pitch (center-to center hole spacing, Λ), core diameter (a), and number of layers. However, changes in the size and shape of the holes in the fiber can be made purposely or accidentally during draw, which will cause deviations between the fiber and preform profiles.

Several methods are used to fabricate microstructure preforms including stacking, casting, extrusion, and drilling. Each process has its advantages and disadvantages. By far, the most common method is stacking in which tubes, rods, and core rods (rods containing doped cores fabricated by traditional optical fiber techniques) are stacked in a close packed pattern such as a triangular or hexagonal lattice (Figure 1). The bundle of tubes and rods is then bound together with a large tube, called an overclad tube. Fibers can be made with complicated or asymmetric index profiles by strategically placing tubes with the same outer diameter (for good stacking) but a different inner diameter to change the index in that location. For example, a fiber with smaller holes on opposite sides of the core produces high birefringence.

The main advantages of stacking are that no special equipment is needed for fabricating microstructure preforms and that doped cores are easily added. However, there are several disadvantages. First, the stacking method is limited to simple geometries of air holes because the tubes are stacked in a close pack arrangement. Second, unless hexagonal tubes are used, interstitial areas are created between the tubes that may not be desired in the final fiber. Third, the method is labor intensive and requires significant glass handling.

A variation of the stacking method is used for fabricating air-clad fibers. Air-clad preforms are created by placing a layer of small tubes around the perimeter of a large core rod. The assembly is then overcladed for strength (Figure 2).

Extrusion and casting of glass powder, polymer, or sol–gel slurry are also used for making single material

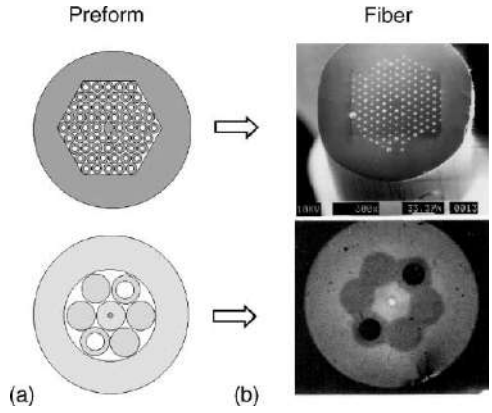

Figure 1 (a) Tubes, rods, and core rods are stacked together in a close packed arrangement and held together with an overclad tube. (b) During draw the desired air region stay open forming a microstructure fiber.

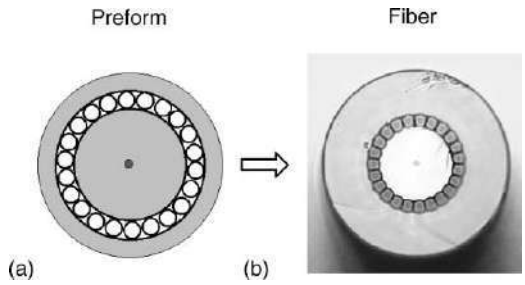

Figure 2 (a) An air-clad preform is fabricated by stacking a ring of small, thin walled tubes around a large core rod. (b) Photograph of the resulting fiber.

microstructure preforms. The main advantage of extrusion and casting is that complicated structures can be fabricated in which the position, size, and shape of the air regions are independent of one another. The disadvantage of this method is that preform fabrication is more difficult compared to the other fabricate methods. This method may become more common as complex air structures are needed to make advanced microstructure fibers.

The last method consists of drilling holes in a preform or rod. Drilling is well understood and used for other specialty fibers. The advantage of this method is that it is easy to drill holes of various sizes in any position in a preform, including doped regions. The disadvantages of drilling are that the holes cannot be drilled very deep compared to a traditional preform length; the distance between holes may be limited due to cracking; and the fiber may experience high loss due to surface roughness of the holes and impurities incorporated during drilling.

Fiber Draw

It is significantly more difficult to draw a microstructure preform with air holes than a traditional preform. This is because the air holes tend to close due to surface tension. This force can be overcome or minimized using two different techniques. The first is to draw the preform under very high tension (low temperature). Minimal hole collapse occurs when the draw tension is significantly larger than the surface tension. The problem with this method is that the break rate increases substantially at higher tension.

The second method is to apply an external pressure to the holes to counteract the surface tension. If a single pressure source is used, the holes can be made larger or smaller during draw by changing the pressure as needed. However, the process can become unstable because a large hole needs a lower pressure to maintain its size than a small hole. If holes of different sizes exist, a larger hole will grow at the expense of a smaller one regardless of the pressure used. To minimize the instability, this method is typically performed in conjunction with high-tension drawing.

To alleviate the draw problem, a second solid material can be used in the preform to obtain the desired index profile. With the air regions removed from the preform, the drawing process becomes significantly easier. In addition, when a preform consists of two solid materials, one of the materials does not have to be continuous, as in single material preforms. This allows for simpler designs such as concentric rings, which may be easier to fabricate and model (Figure 13c). The disadvantage of this approach is finding two materials that have compatible physical and chemical properties, have low loss, and have a large index of refraction difference.

Microstructure Fiber Types

Microstructure fibers are categorized by their method of guidance, properties, and function. Index guided microstructure fibers most closely resemble traditional fibers and are described first. These fibers guide light by total internal reflection and have a core index of refraction greater than the cladding. Bandgap fibers are examined next. The unique guidance of bandgap fibers allows the core index to be lower than the cladding index. Tunable microstructure fibers are described last. These microstructure fibers are filled with tunable materials so that the fiber's properties can be actively manipulated.

Index Guided Fibers

Most microstructure fibers guide light by total internal reflection. The size and spacing of air holes determine the guiding properties. For example, a solid silica core surrounded by a cladding consisting of small air holes that are spread apart creates a small core-to-cladding index difference similar to what can be achieved in traditional fibers. At the other extreme, a solid silica core surrounded by a cladding of large holes, which are closely spaced together, forms a large core-to-cladding index difference. This structure creates a large numerical aperture in the core and optically isolated regions within the fiber. The ability to create regions of significantly different indices in various parts of the fiber distinguishes the index guided microstructure fibers from traditional fiber. Different types of microstructure fiber are formed by varying the size, spacing and pattern of the air holes.

Different techniques are used to model these fibers, depending on spacing and size of the air holes relative to the wavelength of the guided light. If the hole diameter and spacing is much smaller than the

wavelength of the guided light, the cladding index can be modeled as an average index weighted by the volume fraction of the two materials (typically air and silica). As the hole diameter or spacing approaches the wavelength of the guided light, complicated models are used such as finite difference method, finite element method, beam propagation method, or multipole method.

Endlessly single-mode fiber

The endlessly single-mode fiber consists of a solid core surrounded by a two dimensional array of small air holes in the cladding (Figure 3). The size and position of the air holes are arranged such that only the fundamental mode exists in the core regardless of the wavelength of the guided light.

The prevention of higher order modes can be understood by examining the V number. The V number is a dimensionless parameter that describes the guiding nature of the waveguide. In traditional fiber, when the V number is less than 2.405, the fiber will guide only the fundamental mode. The V number is given by

$$V = \frac{2\pi a}{\lambda}(n_{\text{core}}^2 - n_{\text{clad}}^2)^{1/2} \qquad [1]$$

where a is the core diameter, λ is the wavelength of the guided light, and n_{core} and n_{clad} are the index of refractions of the core and cladding, respectively. In traditional fiber, the core and cladding index (n_{core} and n_{clad}) are for the most part constant, so the V number increases as the wavelength decreases.

However, for the endlessly single mode microstructure fiber, the V number does not increase indefinitely with decreasing wavelength, but approaches a constant value. Obtaining a V number independent of wavelength is understood by examining the distance the light propagates into the cladding as a function of wavelength. As the wavelength becomes shorter, light penetrates a shorter distance into the cladding, interacting with less of the air regions. This causes the effective cladding index (n_{clad}) to increase with decreasing wavelength, such that V approaches a constant value.

Microstructure fibers are made endlessly single-moded over all wavelengths by properly designing the hole diameter-to-pitch ratio (d/Λ). When the d/Λ is low enough, the microstructure fiber guides light only in the single mode regardless of wavelength and core size (Figure 4). However, these large core diameter fibers are limited by long wavelength bend loss in the same way as large core single mode traditional fibers.

High delta core fiber

High delta core microstructure fibers (HDCMF) consist of a small (typically less than 2 microns) solid silica core surrounded by one or more layers of large air holes closely spaced together (Figure 5). This creates a very large index difference between the core and cladding. A core-to-cladding index difference of 0.4 is easily achieved in a HDCMF, which is an order of magnitude larger than can be achieved with traditional fiber.

Even though the core is very small, these fibers are almost always multimoded due to the large index difference between the core and cladding. However, because of its small core, it is difficult to launch anything other than the fundamental mode into the fiber. Once a mode is guided in the fiber it remains in that mode due to the large difference in effective indices between modes (Figure 6). When higher-order modes are purposely generated in the fiber, they also

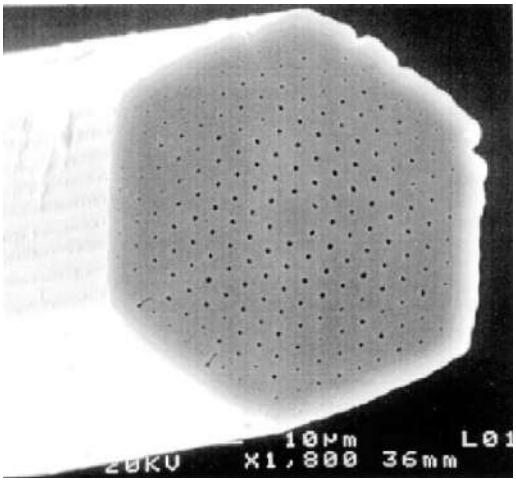

Figure 3 Photograph of an endlessly single mode fiber (Reprinted with permission from Birks TA, Knight JC and Russell P (1997) Endlessly single mode photonic crystal fiber. *Optics Letters* 22: 961–963).

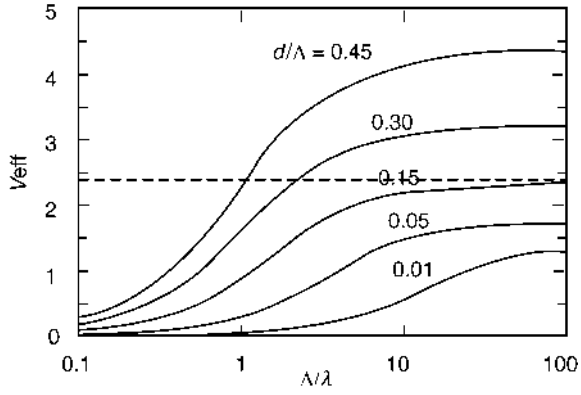

Figure 4 The effective V number as a function of Λ/λ for various d/Λ (Reprinted with permission from Birks TA, Knight JC and Russell P (1997) Endlessly single mode photonic crystal fiber. *Optics Letters* 22: 961–963).

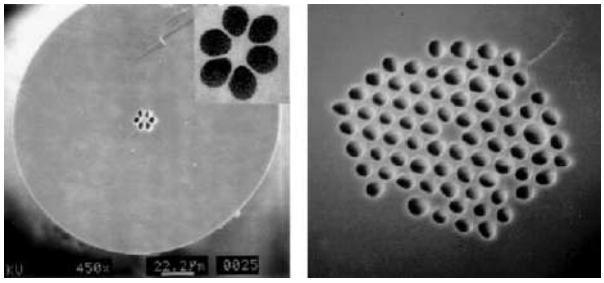

Figure 5 High delta core fibers consist of a small solid core surrounded by at least one layer of large air holes closely spaced together. Right-hand panel reprinted with permission from Ranka JK, Windeler RS and Stentz AJ (2000) Optical properties of high-delta air-silica microstructure optical fibers. *Optics Letters* 25: 796–798.

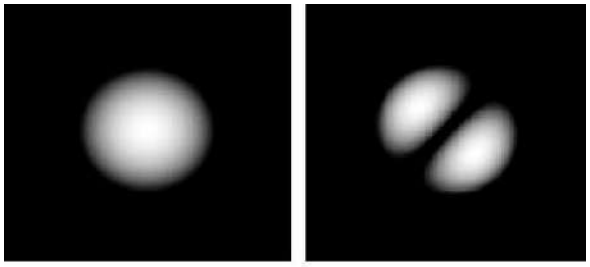

$n_{eff} = 1.425$ $n_{eff} = 1.390$

Figure 6 Individual modes do not couple well due to large differences in effective indices between modes. The figure shows the mode profile and effective index of the fundamental and the next higher mode for a fiber with a two-micron diameter core with a core-to-cladding index difference of 0.45 (Reprinted with permission from Ranka JK, Windeler RS and Stentz AJ (2000) Optical properties of high-delta air-silica microstructure optical fibers. *Optics Letters* 25: 796–798).

guide over long lengths without coupling to other modes (Figure 7). This strong guidance makes the fiber extremely bend insensitive.

HDCMF can have a unique dispersion profile because it can guide the fundamental mode exclusively in the multimode regime. This allows greater flexibility in controlling the shape of the dispersion profile, especially the ability to move the zero dispersion to shorter wavelengths. The HDCMF is the first silica fiber with anomalous dispersion of single-mode light below 1270 nm (the wavelength at which the material dispersion of silica is zero). These fibers are often made to have a zero-group velocity dispersions around 770 nm for use with Ti-sapphire lasers.

HDCMF is ideal for performing nonlinear experiments in the near-visible regime because the fiber's small, high-index core creates high intensity light, and the fiber's low dispersion creates long interaction times. For example, a broadband continuum can be generated over two octaves using less than a meter of HDCMF fiber (Figures 8 and 9).

Tapered microstructure fiber (TMF)

Coupling light in and out of HDCMF is difficult and results in a large loss due to its small core. When free space optics are used to couple the light, frequent realignment of the beam is required to minimize loss due to drifts in the equipment. Splicing HDCMF fiber to traditional fiber also results in high loss due to a large modal mismatch. These coupling problems are solved with tapered microstructure fibers.

Tapered microstructure fibers (TMF) consist of a traditional germanium single-mode core surrounded by a small silica cladding (Figure 10). The inner

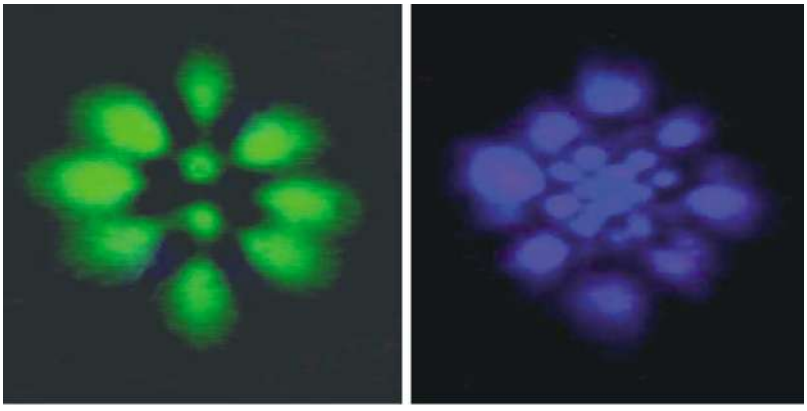

Far-field pattern, 532 nm light
5 meter fiber length

Far-field pattern, 355 nm light
2 meter fiber length

Figure 7 Far-field patterns show that higher order modes propagate long distances in the HDCMF without coupling to other modes (Reprinted with permission from Ranka JK, Windeler RS and Stentz AJ (2000) Optical properties of high-delta air-silica microstructure optical fibers. *Optics Letters* 25: 796–798).

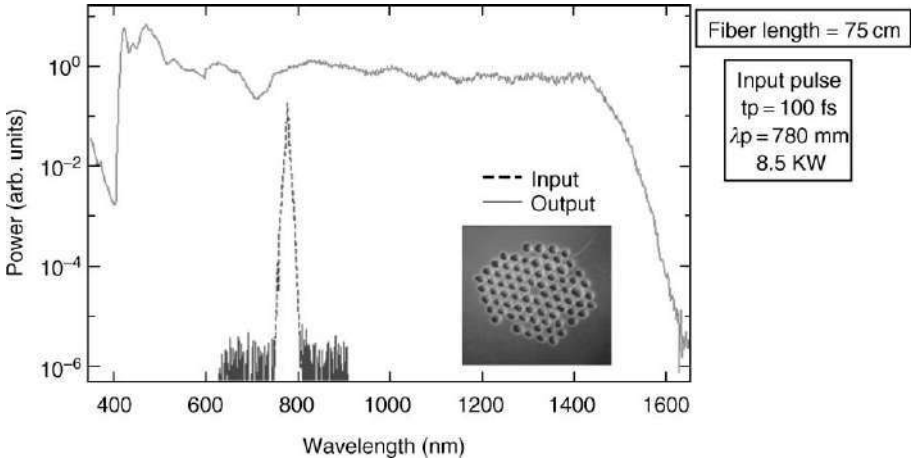

Figure 8 Plot of a supercontinuum spectrum over two octaves wide generated from a 75 cm piece of HDCMF (Reprinted with permission from Ranka JK, Windeler RS and Stentz AJ (2000) Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm. *Optics Letters* 25: 25–27).

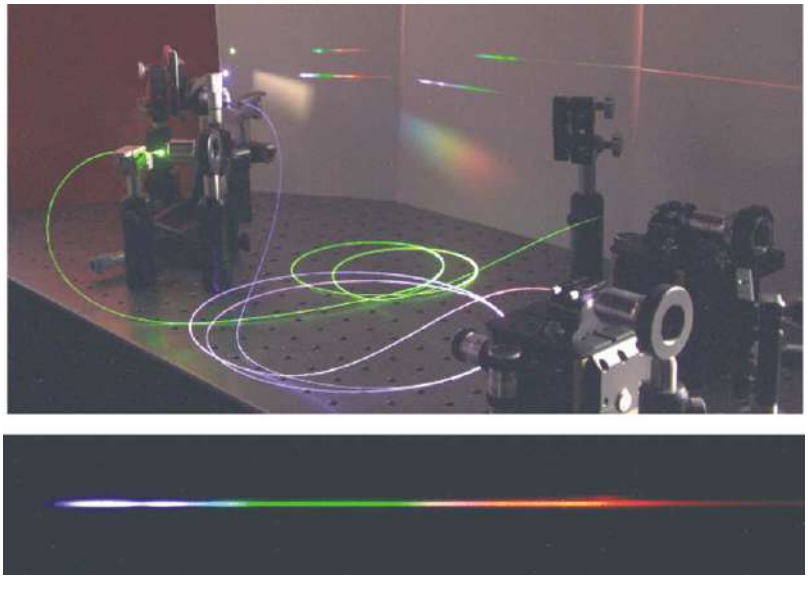

Figure 9 Photograph of fiber generating the supercontinuum. The bottom photo shows the continuum after passing through a prism.

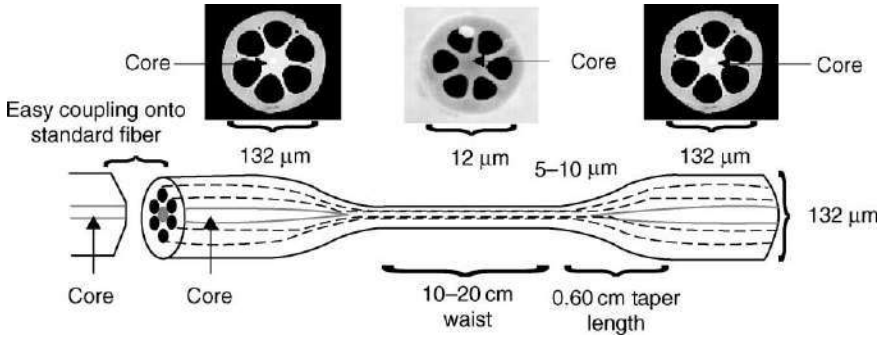

Figure 10 Schematic drawing of the tapered microstructure fiber (Reprinted with permission from Chandalia JK, Eggleton BJ, Windeler RS, Kosinski SG, Liu X and Xu C (2001) Adiabatic coupling in tapered air-silica microstructures optical fiber. *IEEE Photonics Tech. Letters* 13(1) (©2001 IEEE)).

cladding is surrounded by a layer of air holes and then a protective outer silica cladding for strength. Since the core of the TMF is very similar to a traditional single-mode fiber core, the splice loss is low. Splice losses below 0.075 dB are typical.

Properties similar to the HDCMF are obtained by adiabatically tapering the TMF while maintaining the cross-sectional profile of the fiber. In the taper region, the fundamental mode is no longer guided by the germanium core and evolves into the fundamental mode of the silica region, where it is confined by the ring of air holes. The taper waist is typically ten to twenty centimeters long and has properties identical to the HDCMF. When the fiber is adiabatically expanded back to its original size, the light is guided back into the standard diameter germanium core, making it easy to splice to traditional fiber with low loss.

The same effect can be obtained by stretching a traditional fiber from 125 microns down to 2 microns. However, there are several disadvantages compared to the tapered microstructure fiber. The taper region is much longer in the traditional fiber; the optimal diameter of the silica core is harder to maintain; the taper region must remain clean and uncoated; and the taper region is much weaker.

Air-clad fiber

Air-clad fibers consist of a doped core surrounded by a silica inner cladding. The inner cladding is surrounded by a ring of air, which in turn is surrounded by an outer silica cladding for strength (Figure 2). Thin silica webs connect the inner and outer claddings to hold the inner fiber region in place. The region consisting of air and thin silica webs creates an optical barrier, preventing most of the light from escaping the inner cladding. Such fibers are referred to as double clad fiber.

In traditional double-clad fibers, a material of lower index (typically another layer of glass or a layer of polymer) immediately surrounds the inner silica cladding. A key optical parameter describing the light guiding properties of the inner cladding is the numerical aperture (NA). The NA is defined as the sine of the largest (acceptance) angle of light that will be guided in the inner cladding. The upper NA values of traditional double-clad fibers (~0.22 and ~0.45 for glass and polymer outer claddings, respectively) are limited by the refractive indices of the available cladding materials.

For the air-clad fiber, the NA is determined by the ability of light to escape from the inner cladding through the silica webs. The NA of the air-clad fiber can be calculated by modeling the web as an infinitely long slab waveguide. As seen in Figure 11, the NA increases as the webs become thinner or the wavelength of light becomes longer. To obtain NA similar to what is achieved by coating a traditional fiber with a low index polymer (NA ~0.45), the web thickness must be about equal to the wavelength of the light. The NA of the air-clad fiber increases dramatically as the web thickness becomes less than half the wavelength of the guided light. Numerical apertures above 0.90 have been experimentally demonstrated.

The large NA values achievable with air-clad fibers are an important advantage for advanced double-clad

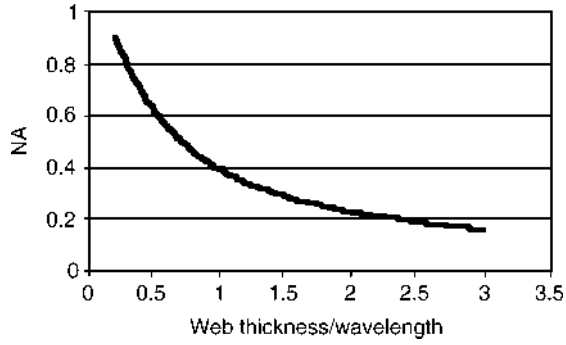

Figure 11 Predicted NA as a function of the web thickness divided by the wavelength using an infinite slab model.

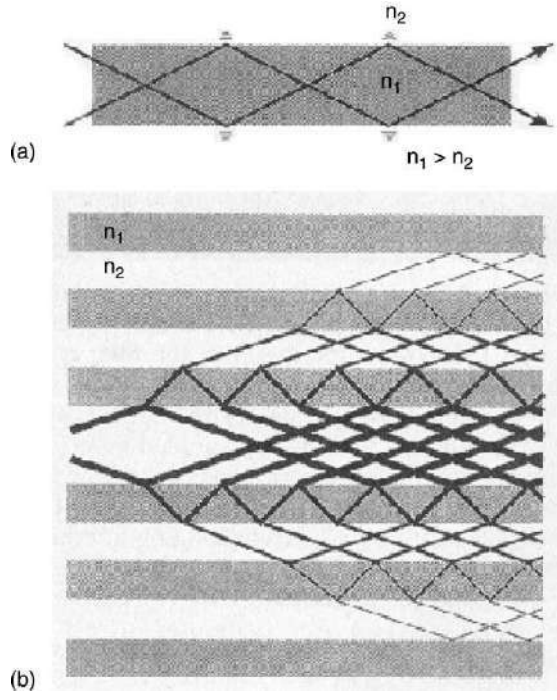

Figure 12 Schematic of the mechanism of (a) index guided light and (b) band gap guided light (Reprinted with permission from Cregan RF, Mangan BJ, Knight JC, Birks TA, Russell PJ, Roberts PJ and Allan DC (1999) Single-mode photonic band gap guidance of light in air. *Science* 285. Copyright 1999 American Association for the Advancement of Science).

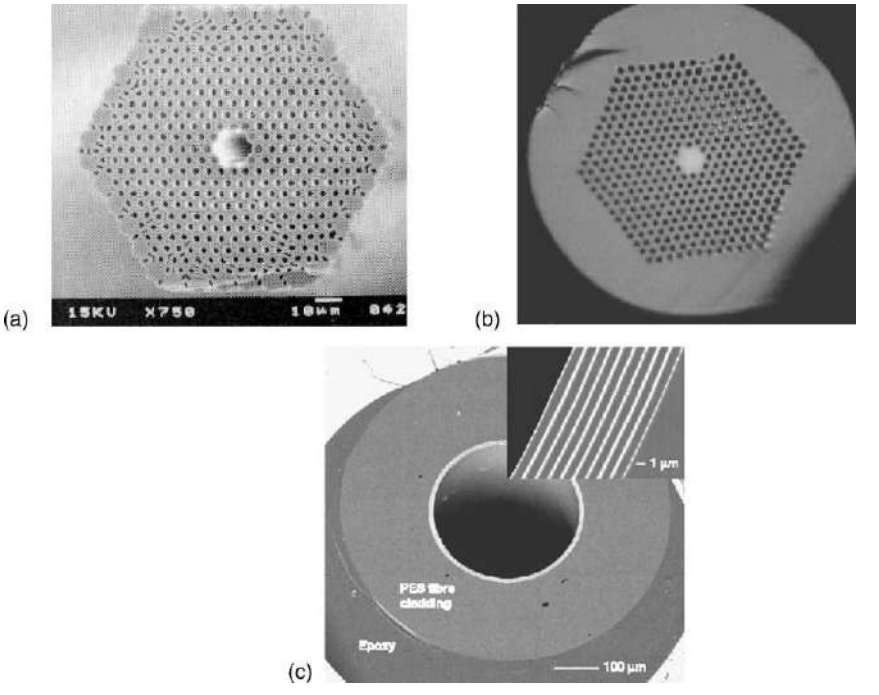

Figure 13 Photographs and drawing of three types of bandgap fibers: (a) triangular array cladding with an air core (Cregan RF, Mangan BJ, Knight JC, Birks TA, Russell PJ, Roberts PJ and Allan DC (1999) Single-mode photonic band gap guidance of light in air. *Science* vol. 285), (b) tunable with a silica core, (c) concentric rings with an air core (Reprinted with permission from Temelkuran *et al. Nature* 420. Copyright 1999 American Association for the Advancement of Science).

amplifiers and lasers. A double-clad amplifier or laser (also known as cladding pumped amplifier or laser) contains a double-clad fiber with a rare earth doped core. High-power pump light, launched into and guided within the inner cladding, excites the rare earth ions as the pump light crosses through the core. The excited ions, which emit light with the same wavelength and phase as the signal, amplify the signal. The fiber's efficiency and the rare earth inversion level increase with the amount of pump light absorbed by the core. The light absorbed by the core, in turn, increases with the core-to-inner cladding area ratio and the pump light intensity in the inner cladding.

Traditional fibers have a practical limit to their core-to-cladding ratio. The maximum core diameter is limited by bend loss. And, while ultimately the minimum fiber diameter is limited by draw capabilities, in practice it is limited by the ability to couple pump light into a fiber having a limited inner-cladding NA. For a pump source of given brightness, the product of beam diameter and NA cannot be increased. Such light can be successfully coupled into a larger inner cladding with smaller NA or a smaller inner cladding with larger NA.

Air-clad fibers have other advantages over traditional double-clad fibers. The outer glass cladding in air-clad fibers is optically inactive. This allows the inner cladding diameter to be decoupled from the physical fiber diameter, eliminating draw-related constraints on minimum inner cladding size. Since only the inner cladding is optically active, its size can be optimized (made smaller) while keeping the total fiber diameter at the standard 125 microns. In addition, the pump light does not interact with the polymer coating, which can be important if the polymer properties are affected by the surroundings.

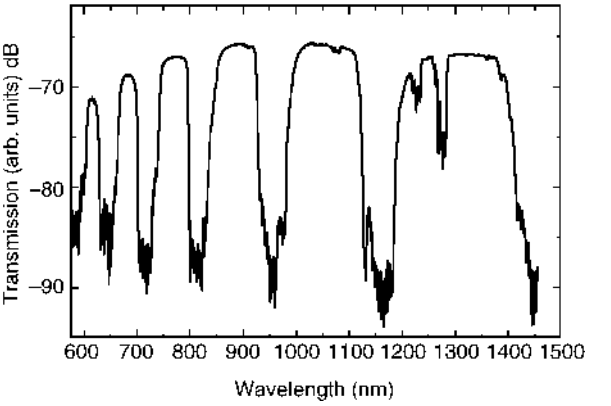

Figure 14 Spectrum of a typical bandgap fiber.

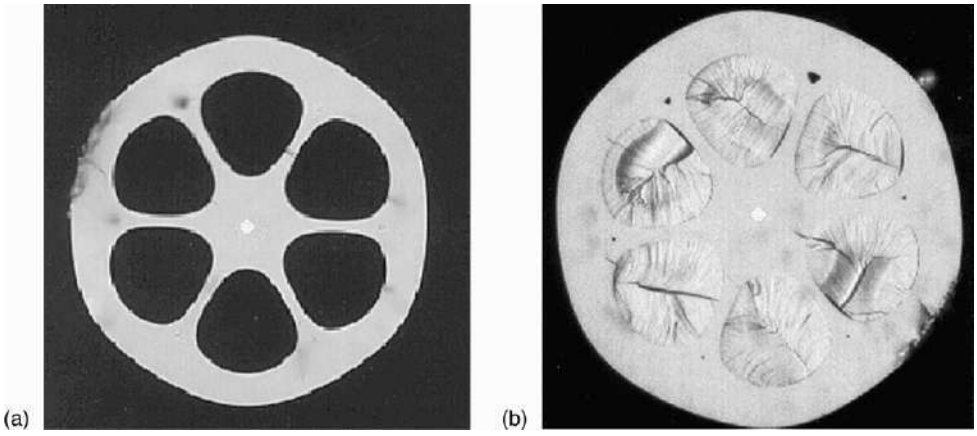

Figure 15 Holes in the microstructure fibers are filled with controllable materials to affect the fiber properties actively. Photos show (a) original fiber and (b) fiber after polymer is inserted and cured in the microstructure fiber.

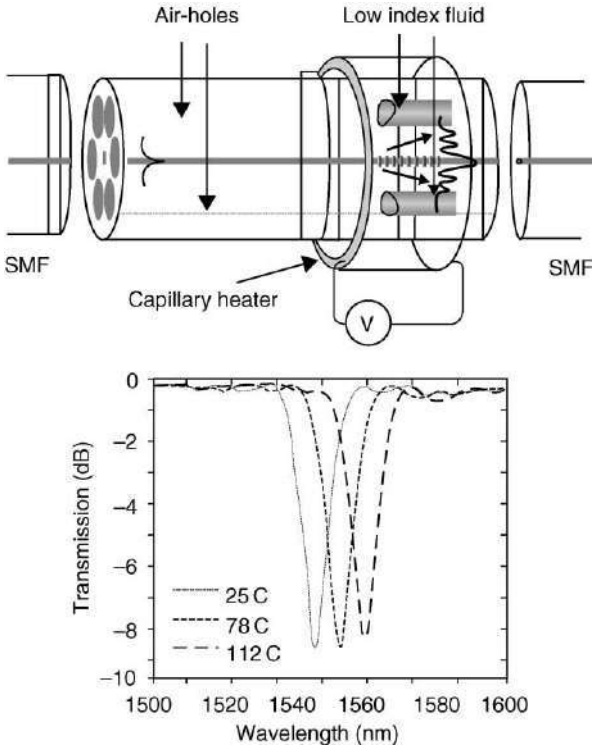

Figure 16 Schematic of a tunable device in which low index liquid is positioned over the grating. The plot shows the shift in the spectrum as a function of temperature (Reproduced from Kerbage C, Windeler RS, Eggleton BJ, Dolinskoi M, Mach P and Rogers JA (2002) Tunable devices based on dynamic positioning of micro-fluids in microstructure optical fiber. *Optics Communications* 204: 179–184, with permission from Elsevier).

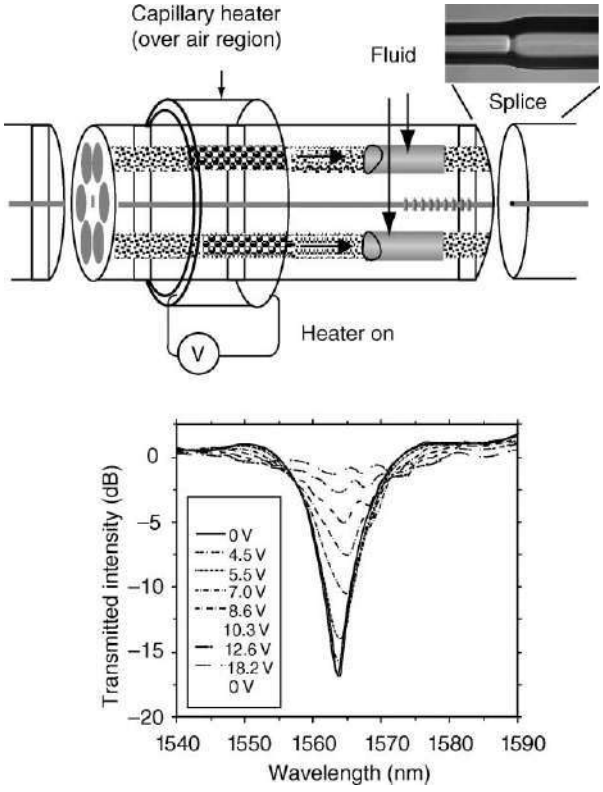

Figure 17 Schematic of a tunable device in which a high index liquid plug is moved over the grating. The plot shows the intensity of the spectrum as a function of temperature (Reproduced from Kerbage C, Windeler RS, Eggleton BJ, Dolinskoi M, Mach P and Rogers JA (2002) Tunable devices based on dynamic positioning of micro-fluids in microstructure optical fiber *Optics Communications* 204: 179–184, with permission from Elsevier).

Bandgap Guided Fibers

Bandgap fibers guide light in the core (also referred to as a defect) by confining the light through constructive interference due to Bragg scattering (Figure 12). Unlike traditional fibers, this mechanism allows light to propagate in a core that has an index lower than the cladding. Bandgap fibers with an air core could theoretically be very low loss, propagate high powers, have a large effective area core, and

exhibit very low nonlinearities. In addition, the fibers can be used in novel devices by filling the core with special gasses or liquids for nonlinear processes.

Bandgap fibers can be generalized into three types (Figure 13). The first type of fiber has an air core surrounded by a cladding that consists of a periodic array of air holes in silica. Fibers of this type have been designed to guide a single mode and hold the greatest promise of guiding light over long distances with very low loss. The second type of bandgap fiber consists of a silica core and a silica cladding with a periodic array of holes which are filled with a high-index, tunable liquid. This design allows the bandgap properties to be tuned and is of interest for use in devices. The third fiber type consists of rings of alternating high- and low-index dielectric material with a center air core. They typically have a large, multimode core which is ideal for sending high powers. In addition, very little light is in the dielectric materials so the fiber can be designed to guide any wavelength with relatively little loss.

The spectrum of a bandgap fiber is quite different than an index guided fiber (Figure 14). Frequencies that lie within the bandgap cannot propagate in the cladding and are confined to the defect in the lattice (the core). Bandgap fibers with ideal geometry are predicted to have losses over an order of magnitude lower than can be obtained in an ideal traditional fiber. Although the loss of bandgap fibers has not been demonstrated to be less than traditional fiber at telecommunication wavelengths, bandgap fibers have been shown to have much lower loss than traditional fibers at other wavelengths.

Tunable Microstructure Fiber Devices

Unique tunable devices are made by filling the air holes of microstructure fibers with materials whose properties can be controlled actively (Figure 15). The materials can be positioned in the core or cladding to affect the corresponding modes. Most commonly, the index of refraction is changed using temperature sensitive liquids or polymers, although

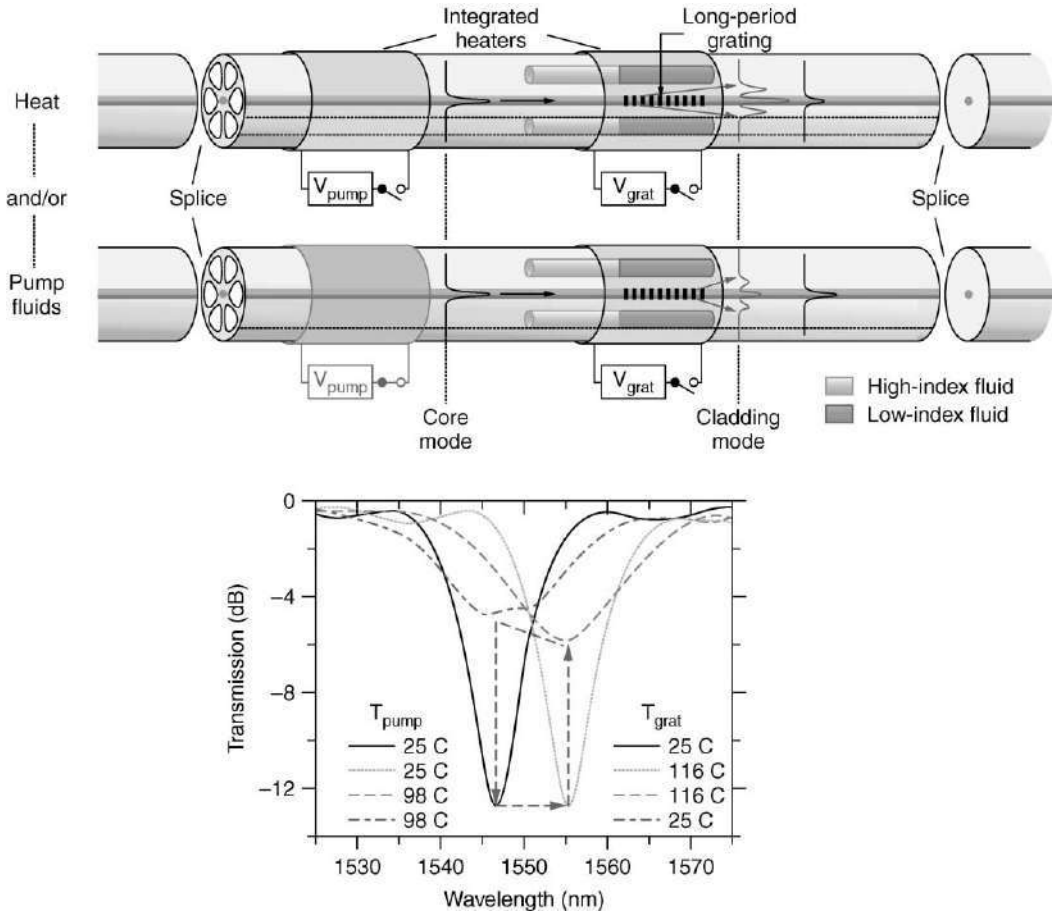

Figure 18 Schematic of a tunable device using both high and low index liquids with two heaters. The plot shows that the position and strength of the spectrum are adjusted independently (Reproduced with permission from Mach P, Dolinski M, Baldwin KW, Rogers JA, Kerbage C, Windeler RS and Eggleton BJ (2002) Tunable microfluidic optical fiber. *Applied Physics Letters* 80(23). Copyright (2002) by the American Physical Society).

materials that respond to electric or magnetic fields enable a quicker response time. These material-filled active fibers can be used to fabricate devices like tunable filters, switches, broadband attenuators, and fibers with tunable birefringence. Below, a few examples of tunable filters are presented in which the device properties depend on the type and position of the liquid inserted in the fiber holes.

A microstructure tunable filter works by controlling the spectral position, shape, or strength of a longitudinal long-period grating written in the fiber. A grating is written in the fiber before filling the holes with tunable liquid. The position of the peak absorption wavelength is controlled by placing a liquid, with an index close to silica and strongly temperature dependent, in air holes in the cladding region over the grating (**Figure 16**). The cladding index is tuned by varying the temperature of the liquid, using a small thin film heater located over the liquid. The position of the peak absorption wavelength moves as the cladding index is changed.

The strength or depth of the filter is controlled by inserting a moveable plug of high-index fluid in a sealed air section of a microstructure fiber (**Figure 17**). The strength of the grating is determined by the fraction of the grating surrounded by high index fluid. The position of the liquid plug relative to the grating can be finely adjusted with a thin film heater located over the air region adjacent to the material plug. As heat is applied to the air, the air expands and pushes the liquid plug over the grating, decreasing the coupling of the fundamental mode to higher order cladding modes. **Figure 18** shows an example in which the position and strength of a filter are varied independently using two heaters and two adjacent materials.

See also

Photonic Crystals: Photonic Crystal Lasers, Cavities and Waveguide.

Further Reading

Birks TA, Knight JC and Russell P (1997) Endlessly single mode photonic crystal fiber. *Optics Letters* 22(13).

Bjarklev A, Broeng J and Bjarklev AS, *Photonic Crystal Fibers*, Boston: Kluwer Academic Publishers.

Broeng J, Mogilevstev D, Bjarklev A and Barkou SE (1999) Photonic crystal fibers: A new class of optical waveguides. *Optical Fiber Technology*, vol. 5.

Chandalia JK, Eggleton BJ, Windeler RS, Kosinski SG, Liu X and Xu C (2001) Adiabatic coupling in tapered air-silica microstructures optical fiber. *IEEE Photonics Tech. Letters* 13(1).

Cregan RF, Mangan BJ, Knight JC, Birks TA, Russell PJ, Roberts PJ and Allan DC (1999) Single-mode photonic band gap guidance of light in air. *Science* 285.

Eggleton BJ, Westbrook PS, White CA, Kerbage C, Windeler RS and Burdge GL (2000) Cladding-mode resonance in air-silica microstructure optical fiber. *Journal Lightwave Technology* 18(8).

Joannopoulos JD, Meade RD and Winn JN (1995) *Photonic Crystals*. New Jersey: Princeton University Press.

Johnson SG, Ibanesc M, Skorobogatiy M, Weisberg M, Engeness O, Soljacic M, Jacobs SA, Joannopoubs JD and Fink Y (2001) Low-loss asymptotically single-mode propagation in large-core OmniGuide fibers. *Optics Express* 9(13).

Kerbage C, Hale A, Yablon A, Windeler RS and Eggleton BJ (2001) Integrated all fiber variable attenuator based on hybrid microstructure fiber. *Applied Physics Letters* 79(19).

Kerbage C, Windeler RS, Eggleton BJ, Dolinskoi M, Mach P and Rogers JA (2002) Tunable devices based on dynamic positioning of micro-fluids in microstructure optical fiber. *Optics Communications* 204: 179–184.

Knight JC, Birks TA, Russell PJ and Atkin DM (1996) All-silica single-mode optical fiber with photonic crystal cladding. *Optics Letters* 21(19).

Knight JC, Birks TA and Russell PJ (2001) 'Holey' silica fibers. In: Markel VA and George TF (eds). *Optics of nanostructured materials*. New York: John Wiley & Sons.

Mach P, Dolinski M, Baldwin KW, Rogers JA, Kerbage C, Windeler RS and Eggleton BJ (2002) Tunable microfluidic optical fiber. *Applied Physics Letters* vol. 80.

Ranka JK, Windeler RS and Stentz AJ (2000) Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm. *Optics Letters* 25(1).

Ranka JK, Windeler RS and Stentz AJ (2000) Optical properties of high-delta air-silica microstructure optical fibers. *Optics Letters* 25: 796–798.

Omnidirectional Surfaces and Fibers

S Hart, G Benoit and Y Fink, Massachusetts Institute of Technology, Cambridge, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Omnidirectional Reflecting Mirrors

Mirrors are probably the most prevalent of optical devices. Known to the ancients and used by them as objects of worship and beauty, mirrors are currently employed for imaging, solar energy collection, and also in laser cavities. Their intriguing optical properties have captured the imagination of scientists as well as artists and writers.

One can distinguish between two types of mirrors, the age-old metallic, and the more recent multilayer dielectric. Metallic mirrors reflect light over a broad range of frequencies incident from arbitrary angles (i.e., omnidirectional reflectance). However, at infrared and optical frequencies a few percent of the incident power is typically lost due to absorption.

Multilayer dielectric mirrors are used primarily to reflect a narrow range of frequencies incident from a particular angle or particular angular range. Unlike their metallic counterparts, dielectric reflectors can be extremely low loss. The ability to reflect light of arbitrary angle of incidence for all-dielectric structures has been associated with the existence of a complete photonic bandgap, which can exist only in a system with a dielectric function that is periodic along three orthogonal directions. In fact, a sufficient condition for the achievement of omnidirectional reflection in a periodic system with an interface, is the existence of an overlapping bandgap regime in phase space above the light cone of the ambient media.

Consider a system that is constructed of alternating dielectric layers coupled to a homogeneous medium – characterized by n_0 (such as air with $n_0 = 1$), at the interfaces. Electromagnetic waves are incident upon the multilayer film from the homogeneous medium. While such a system was analyzed extensively in the literature the possibility of omnidirectional reflectivity was not recognized until recently.

The generic system is described by the index of refraction profile in Figure 1, where h_1 and h_2 are the layer thickness, and n_1 and n_2 are the indices of refraction of the respective layers. The incident wave has a wavevector $\vec{k} = k_x\hat{e}_x + k_y\hat{e}_y$, and frequency of $\omega = c|k|$. The wavevector together with the normal to the periodic structure defines a mirror plane of symmetry which allows us to distinguish between two independent electromagnetic modes: transverse electric (TE) modes and transverse magnetic (TM) modes. For the TE mode the electric field is perpendicular to the plane, as is the magnetic field for the TM mode.

General features of the transport properties of the finite structure can be understood when the properties of the infinite structure are elucidated. In a structure with an infinite number of layers, translational symmetry along the direction perpendicular to the layers leads to Bloch wave solutions of the form:

$$u_K(x, y) = E_K(y)\, e^{iKy}\, e^{ik_x x} \qquad [1]$$

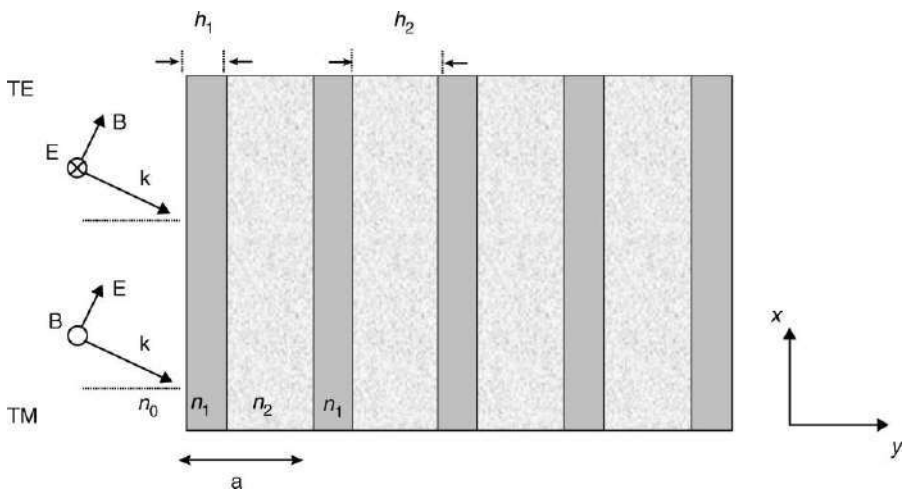

Figure 1 Schematic of the multilayer system showing the layer parameters (n_α, h_α – index of refraction and thickness of layer α), the incident wavevector $\vec{k}$ and the electromagnetic mode convention.

where $E_K(y)$ is periodic, with a period of length a, and K is the Bloch wave number. These waves represent solutions to an eigenvalue problem and are completely and uniquely defined by the specification of K, k_x and ω.

Solutions can be propagating or evanescent, corresponding to real or imaginary Bloch wave numbers, respectively. It is convenient to display the solutions of the infinite structure by projecting the $\omega(K, k_x)$ function onto the $\omega - k_x$ plane; Figure 2a,b are examples of such projected structures.

The gray background areas highlight phase space where K is strictly real, that is, regions of propagating states, while the white areas represent regions containing evanescent states. The shape of the projected bandstructures for the multilayer film can be understood intuitively. At $k_x = 0$, the bandgap for waves traveling normal to the layers is recovered. For $k_x > 0$, the bands curve upward in frequency. As $k_x \to \infty$, the modes become largely confined to the slabs with the high index of refraction and do not couple between layers (and are therefore independent of k_x).

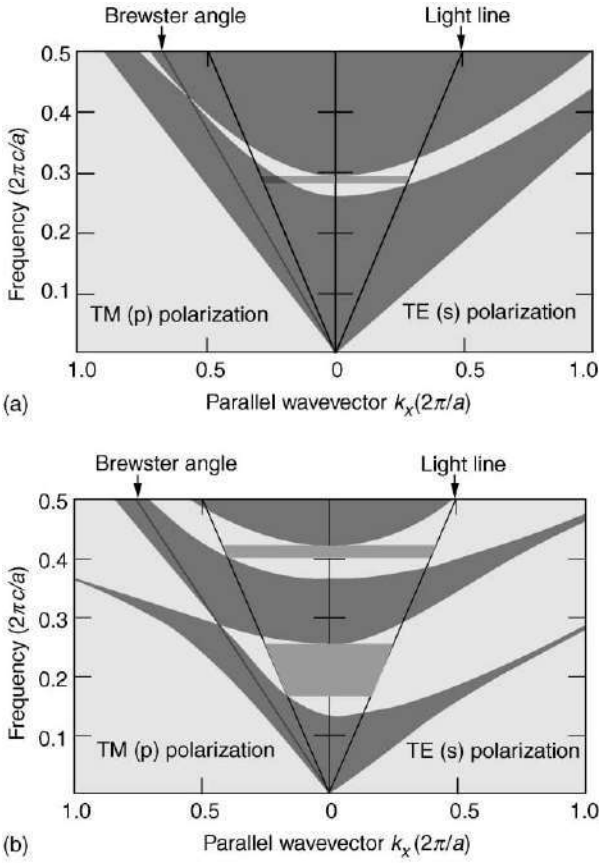

Figure 2 (a) Projected bandstructure of a multilayer film with the light line and Brewster line, exhibiting a reflectivity range of limited angular acceptance with ($n_1 = 2.2$ and $n_2 = 1.7$ and a thickness ratio of $h_2/h_1 = 2.2/1.7$). (b) Projected bandstructure of a multilayer film together with the light line and Brewster line, showing an omnidirectional reflectance range at the first and second harmonic (propagating states – dark gray, evanescent states – white, omnidirectional reflectance range – light grey). The film parameters are $n_1 = 4.6$ and $n_2 = 1.6$ with a thickness ratio of $h_2/h_1 = 1.6/0.8$. These parameters are similar to the actual polymer–tellurium film parameters measured in the experiment. Reproduced with permission from Fink Y, Winn JN, Fan S, *et al.* (1998) A dielectric omnidirectional reflector. *Science* 282: 1679–1682. Copyright 1998 American Association for the Advancement of Science.

In a finite structure, the translational symmetry in the directions parallel to the layers is preserved, hence k_x remains a conserved quantity and can be used to label solutions even though these solutions will no longer be of the Bloch form. The relevance of the band diagram to finite structures is that it allows for the prediction of regions of phase space where waves are evanescently decaying in the multilayer structure.

Since we are primarily interested in waves originating from the homogeneous medium external to the periodic structure, we will focus only on the portion of phase space lying above the light line. Waves originating from the homogeneous medium satisfy the condition $\omega \geq ck_x/n_0$, where n_0 is the refractive index of the homogeneous medium, and therefore they must reside above the light line. States of the homogeneous medium with $k_x = 0$ are normal incident, and those lying on the $\omega = ck_x/n_0$ line with $k_y = 0$ are incident at an angle of 90°.

The states in Figure 2a, that are lying in the restricted phase space defined by the light line and that have a (ω, k_x) corresponding to the propagating solutions (gray areas) of the crystal can propagate in both the homogeneous medium and in the structure. These waves will partially or entirely transmit through the film. Those with (ω, k_x) in the evanescent regions (white areas) can propagate in the homogeneous medium but will decay in the crystal – waves corresponding to this portion of phase space will be reflected off the structure.

The multilayer system leading to Figure 2a represents a structure with a limited reflectivity cone, since for any frequency one can always find a k_x vector for which a wave at that frequency can propagate in the crystal – and hence transmit through the film. The necessary and sufficient criterion for omnidirectional reflectivity at a given frequency is that there exists no transmitting state of the structure inside the light cone – this criterion is satisfied by frequency ranges marked in light gray in Figure 2b. In fact, the system leading to Figure 2b exhibits two omnidirectional reflectivity ranges.

The omnidirectional range is defined from above by the normal incidence bandedge $\omega_h(k_y = \pi/a, k_x = 0)$ (Figure 2b) and below by the intersection of the top of the TM allowed bandedge with the light line $\omega_l(k_y = \pi/a, k_x = \omega_l/c)$ (Figure 2b).

A dimensionless parameter used to quantify the extent of the omnidirectional range is the range to mid-range ratio defined as $(\omega_h - \omega_l)/\frac{1}{2}(\omega_h + \omega_l)$.

Figure 3 is a plot of this ratio as a function of n_2/n_1 and n_1/n_0, where ω_h and ω_l are determined by solutions of eqn [2] with quarter wave layer thickness. The contours in this figure represent various equi-omnidirectional ranges for different material index parameters and could be useful for design purposes.

At normal incidence, there is no distinction between TM and TE modes. At increasingly oblique angles, the gap of the TE mode increases, whereas the gap of the TM mode decreases. In addition, the center of the gap shifts to higher frequencies. Therefore, the criterion for the existence of omnidirectional reflectivity can be restated as the occurrence of a frequency overlap between the gap at normal incidence and the gap of the TM mode at 90°. Analytical expressions for the range to mid-range ratio can be obtained by setting:

$$\begin{aligned}\omega_h &= \frac{2c}{h_2 n_2 + h_1 n_1}\cos^{-1}\left(-\left|\frac{n_1 - n_2}{n_1 + n_2}\right|\right)\\ \omega_l &= \frac{2c}{h_2\sqrt{n_2^2 - 1} + h_1\sqrt{n_1^2 - 1}}\\ &\quad\times\cos^{-1}\left(\left|\frac{n_1^2\sqrt{n_2^2 - 1} - n_2^2\sqrt{n_1^2 - 1}}{n_1^2\sqrt{n_2^2 - 1} + n_2^2\sqrt{n_1^2 - 1}}\right|\right)\end{aligned} \qquad [2]$$

Moreover, the maximum range width is attained for thickness values that are not equal to the quarter wave stack though the increase in bandwidth gained by deviating from the quarter wave stack is typically only a few percent.

In general, the TM mode defines the lower frequency edge of the omnidirectional range; an example can be seen in Figure 2b for a particular choice of the indices of refraction. This can be proven by showing that:

$$\left.\frac{\partial\omega}{\partial k_y}\right|_{TM} \geq \left.\frac{\partial\omega}{\partial k_y}\right|_{TE} \qquad [3]$$

in the region that resides inside the light line. The physical reason for eqn [3] lies in the vectorial nature of the electric field. In the upper portion of the first band the electric field concentrates its energy in the high dielectric regions. Away from normal incidence, the electric field in the TM mode has a component in the direction of periodicity, this component forces a larger portion of the electric field into the low dielectric regions. The group velocity of the TM mode is therefore enhanced. In contrast, the electric field of the TE mode is always perpendicular to the direction of periodicity and can concentrate its energy primarily in the high dielectric region. While the omnidirectional reflection criteria can be used to confine light in various geometries, it is the cylindrical fiber configuration that appears to present significant application opportunities.

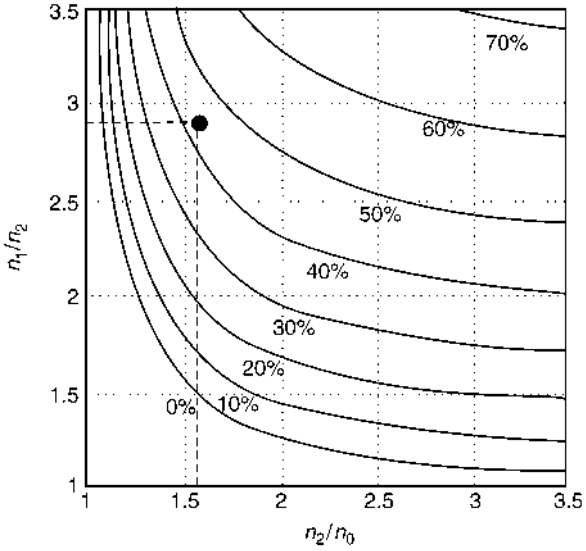

Figure 3 The range to mid-range ratio $(\omega_h - \omega_l)/\frac{1}{2}(\omega_h + \omega_l)$, for the fundamental frequency range of omnidirectional reflection, plotted as contours. Here, the layers were set to quarter wave thickness and $n_1 > n_2$. The ratio for our materials is approximately 45%. ($n_1/n_2 = 2.875$, $n_2/n_0 = 1.6$) is located at the intersection of the dashed lines (black dot). Reproduced with permission from Fink Y, Winn JN, Fan S, *et al.* (1998) A dielectric omnidirectional reflector. *Science* 282: 1679–1682. Copyright 1998 American Association for the Advancement of Science.

Mesostructured Fibers for External Reflection Applications

Polymer fibers are ubiquitous in applications such as textile fabrics, due to their excellent mechanical properties and the availability of low-cost, high-volume processing techniques; however, the control over their optical properties has so far remained relatively limited. Conversely, dielectric mirrors are used to precisely control and manipulate light in high performance optical applications, but the fabrication of these typically fragile mirrors has been mostly restricted to planar geometries and typically involves multiplicity of deposition sequences in a high vacuum thin film deposition system.

Fabrication Approach

The fabrication of extended lengths of omnidirectionally reflecting fibers with large bandgaps and

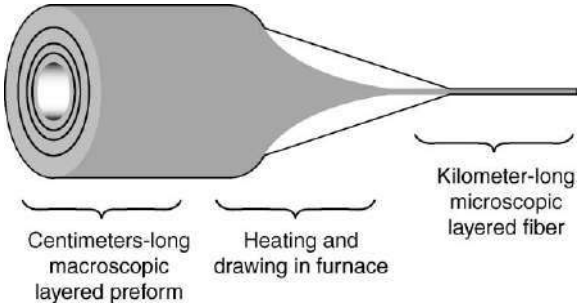

Figure 4 Conceptual preform based fabrication process for meso-structured fibers.

high layer counts poses considerable challenges. To illustrate the nature of this formidable task, one needs to consider the necessity of maintaining the uniformity of sub-100 nm layer thicknesses over kilometer-length scales; creating continuous layers with an aspect ratio of $\sim 10^{10}$ in a single process! To meet this and other challenges associated with the fabrication of mesostructured fibers, we have developed a preform-based fabrication approach (Figure 4). A scaled-up version of the final fiber, called a preform, is fabricated which shares the geometry and materials of the final fiber but exhibits macroscopic lateral features. The preform is heated up and drawn under tension into the fiber using a simple cylindrical furnace. The macroscopic layers are reduced in the process to microscopic dimensions while maintaining the overall geometry and symmetry of the original preform. Conservation of mass determines the final length of the fiber – typically this length is equal to the lateral reduction factor squared multiplied by the length of the preform. The nature of the process and requirements on the fiber's optical properties lead to the definition of materials selection rules. First, in order to achieve omnidirectional reflectivity, one needs to identify two solid materials exhibiting an index contrast given by the plot in Figure 3. Second, to enable the codrawing of the two dissimilar materials both will need to have viscosities that are lower than $\sim 10^8$ poise at the drawing temperature. In order to maintain high draw speeds, the majority component needs to be amorphous and the adhesion between these two materials needs to be sufficient to prevent delamination. Finally, their thermal expansion coefficients need to be close or alternatively at least one of the materials needs to be capable of relieving the stress due to CTE mismatch.

Materials Selection Criteria

Pairs of materials that are compatible with the process and fiber property requirements have been identified, including: high glass transition temperature (T_g), thermoplastic polymers such as poly(ether-sulfone) (PES), poly (ether-imide) (PEI) and members of the chalcogenide glass family, such as arsenic triselenide (As_2Se_3) or arsenic trisulfide. Pairs of these materials will have substantially different refractive indices, as shown in Figure 5, which is a measurement of the real and imaginary indices of refraction of PES and As_2Se_3, obtained using a broadband spectroscopic ellipsometer (SOPRA GES5). They nevertheless exhibit similar thermo-mechanical properties within a certain thermal processing window.

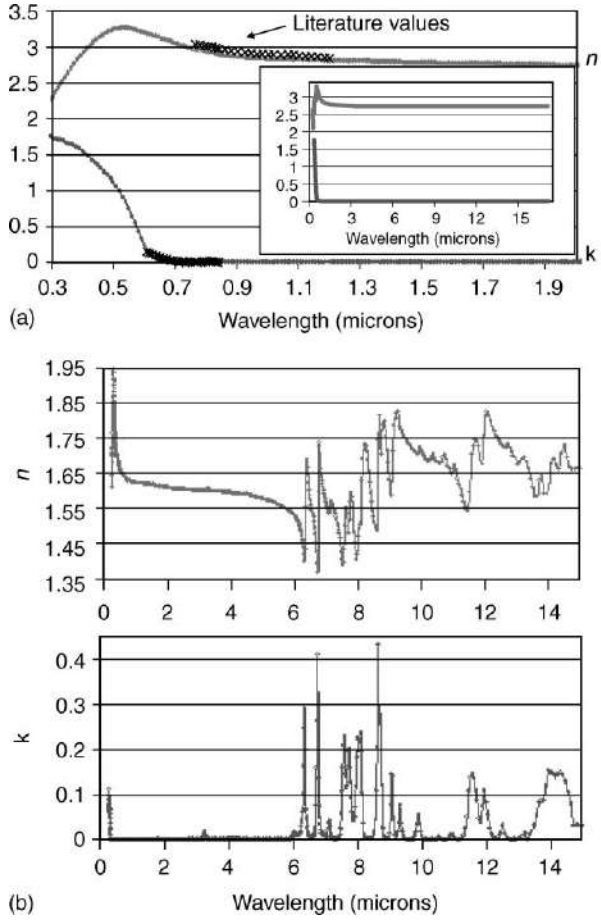

Figure 5 (a) Real part (n – light gray) and imaginary part (k – dark gray) of the refractive index of annealed As_2Se_3. The black crosses correspond to literature values. (b) Real part (n – upper) and imaginary part (k – lower) of the refractive index of PES.

Adhesion and extensional viscosity in the fluid state are difficult to measure in general, and the measurement of high-temperature surface tension is quite involved. Thus, limited data on these properties are available and it was necessary to empirically identify materials that could be used to draw out mirror-fibers. Various high-index chalcogenide (S, Se, and Te containing) glasses and low-index polymers were

identified as potential candidates based on their optical properties and overlapping thermal softening regimes. Adhesion and viscosity matching were tested by thermal evaporation of a chalcogenide glass layer on top of a polymer film or rod and elongation of the coated substrate at elevated temperatures. The choice of a high-temperature polymer, PES, and a simple chalcogenide glass, As_2Se_3, resulted in excellent thermal co-deformation without film cracking or delamination. Approximate matching of extensional viscosity in this manner was also demonstrated using As_2Se_3 and PEI. The properties, processing, and applications of chalcogenide glasses have been explored extensively elsewhere. One advantage in choosing As_2Se_3 for this application is that not only is it a stable glass, but it is a stoichiometric compound that can be readily deposited in thin films through thermal evaporation or sputtering without dissociation. Additionally, As_2Se_3 is transparent to IR radiation from approximately 0.8 to 17 μm as shown in the ellipsometric data (Figure 5a), and has a refractive index of ~2.8 in the mid-IR. PES is a high-performance, dimensionally stable thermoplastic with a refractive index of ~1.6 and good transparency to EM waves in a range extending from the visible regime into the mid-IR (~5 μm), as shown in Figure 5b.

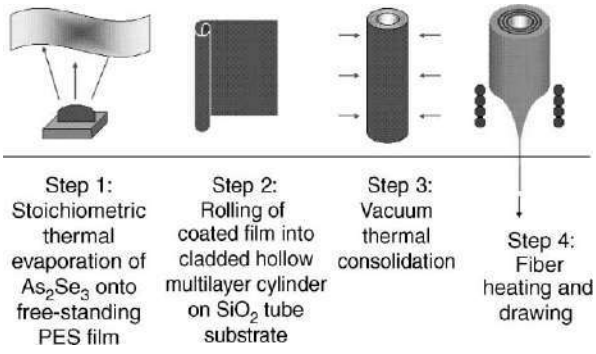

Figure 6 Multilayer preform fabrication sequence.

Preform Construction Process and Fiber Draw

The selected materials were used to construct a multilayer preform rod, which essentially is a macroscale version of the final fiber. In order to fabricate the dielectric mirror fiber preform, an As_2Se_3 film was deposited through thermal evaporation on either side of a free-standing PES film which was then rolled on top of a PES tube substrate, forming a structure having 21 alternating layers of PES and As_2Se_3, as shown in Figure 6.

The resulting multilayer fiber preform was subsequently drawn down using an optical fiber draw tower into hundreds of meters of multilayer fiber with a precisely controlled submicron layer thickness, creating a photonic bandgap in the mid-IR. Fibers of outer diameters varying from 175–500 μm with a typical standard deviation of 10 μm from target, were drawn from the same preform to demonstrate adjustment of the reflectivity spectra through thermal deformation. The spectral position of the photonic bandgap was controlled by the optical monitoring of the outer diameter (OD) of the fiber during draw, which was later verified by reflectivity measurements on single and multiple fibers of different diameters. Scanning electron micrographs (SEMs) of the cross-section of these fibers are depicted in Figure 7.

Bandstructure for Multilayer Fibers for External Reflection Applications

In theoretically predicting the spectral response of these fibers, it is helpful to calculate the photonic

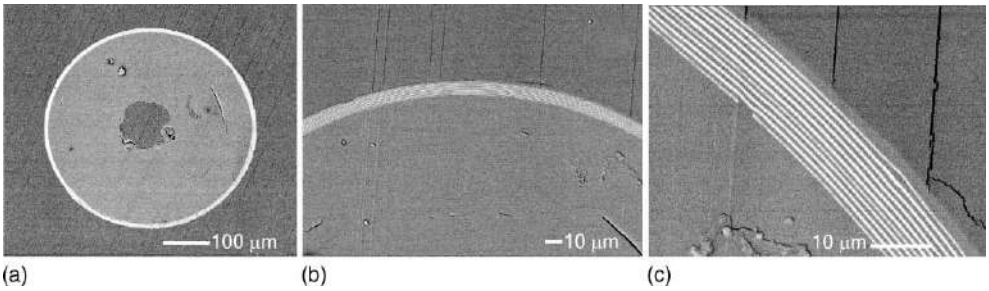

Figure 7 SEM micrographs of 400 μm OD fiber cross-section. The entire fiber is embedded in epoxy. (a) shows the entire fiber cross-section, with mirror structure surrounding the PES core; (b) demonstrates that the majority of the fiber exterior is free of significant defects and that the mirror structure adheres well to the fiber substrate; and (c) reveals the ordering and adhesion within the alternating layers of As_2Se_3 (bright layers) and PES. Stresses developed during sectioning caused some cracks in the mounting epoxy that are deflected at the fiber interface. Fibers from this batch were used in the reflectivity measurements recorded below in Figure 9a. Reproduced with permission from Hart SD, Maskaly GR, Temelkuran B, Prideaux PH, Joannopoulos JD and Fink Y (2002) External reflection from omnidirectional dielectric mirror fibers. *Science* 296: 510–513. Copyright 2002 American Association for the Advancement of Science.

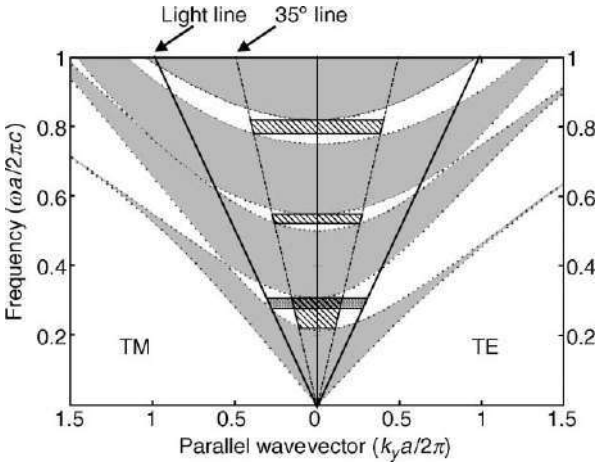

Figure 8 Photonic band diagram for a one-dimensional photonic crystal having a periodic refractive index alternating between 2.8 and 1.55. Gray regions represent propagating modes within the structure, while white regions represent evanescent modes. Hatched regions represent photonic bandgaps where high reflectivity can be expected for external EM waves over an angular range extending from normal to 35° incidence. The lower dark shaded trapezoid represents a region of external omnidirectional reflection.

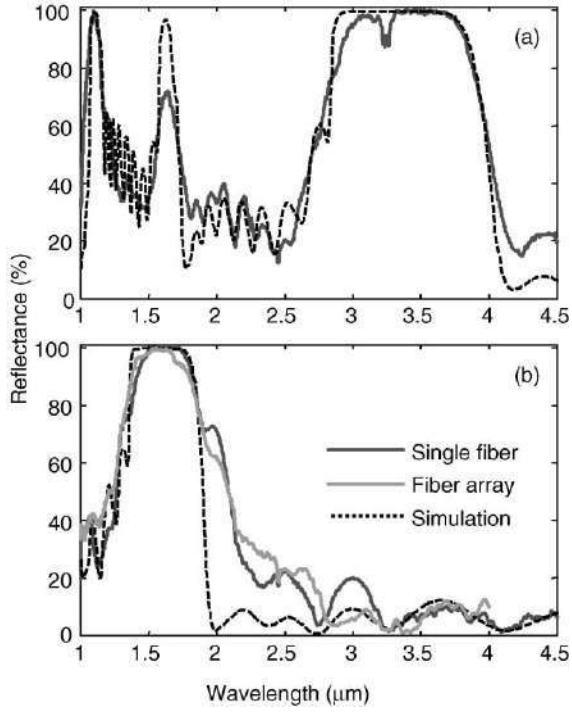

Figure 9 Measured reflectance spectra for 400 μm OD (a) and 200 μm OD (b) dielectric mirror fibers relative to gold-coated fibers of the same diameter. (a) shows a single-fiber reflectivity measurement, while (b) compares single-fiber reflectivity to that measured from a multifiber array. Simulations were performed using the transfer matrix method.

bandstructure that corresponds to an infinite one-dimensional photonic crystal. This allows for the analysis of propagating and evanescent modes in the structure, corresponding to real or imaginary Bloch wave number solutions.

The electric or magnetic field vector is parallel to the mirror layer interfaces for the TE and TM polarized modes, respectively. The parallel wavevector (k_y) is the component of the incident electromagnetic (EM) wavevector that is parallel to the layer interfaces. The phase space accessible from an external ambient medium is contained between the light lines (defined by the glancing-angle condition $\omega = ck_y/n_0$), and the modes between the 35° lines correspond to those sampled experimentally. Axes are normalized to the thickness a of one mirror bilayer (a period consisting of one high- and one low-index layer). Figure 8 depicts the photonic band diagram for an infinite structure having similar periodicity and refractive indices to the mirror structures fabricated here. Three photonic bandgaps are present where high reflectivity is expected within the 0–35° angular range, and the fundamental gap contains a region of external omnidirectional reflectivity.

Optical Characterization of 'Mirror Fibers'

Mirror fiber reflectivity was measured from both single fibers and parallel fiber arrays using a Nicolet/SpectraTech NicPlan Infrared Microscope and Fourier Transform Infrared Spectrometer (Magna 860). The microscope objective (SpectraTech 15 ×, Reflachromat) used to focus on the fibers had a numerical aperture (NA) of 0.58. This results in a detected cone where the angle of reflection with respect to the surface normal of the structure could vary from normal incidence to ~35°, which is determined by the NA of the microscope objective. As a background reference for the reflection measurements, we used gold-coated PES fibers of matching diameters. Dielectric mirror fibers, drawn to 400 μm OD, exhibited a very strong reflection band centered at 3.4 μm wavelength (Figure 9a). Measured reflectivity spectra agree well with planar-mirror transfer matrix method (TMM) simulations, where the reflectivity was averaged across the aforementioned angular range for both polarization modes. Fibers drawn down to 200 μm OD show a similar strong fundamental reflection band centered near 1.7 μm (Figure 9b). This shifting of the primary photonic bandgap clearly illustrates the precise tuning of the reflectivity spectra over wide frequency ranges through thermal deformation processing. Strong optical signatures are measurable from single fibers as small as 200 μm OD. Fiber array measurements,

simultaneously sampling reflected light from multiple fibers, agree quite well with single-fiber data (Figure 9b).

These reflectivity results are strongly indicative of uniform layer thickness control, good interlayer adhesion, and low interdiffusion through multiple thermal treatments. This was confirmed by SEM inspection of fiber cross-sections (Figure 7). The layer thicknesses observed ($a = 0.90$ μm for the 400 μm fibers; $a = 0.45$ μm for the 200 μm fibers) correspond well to the measured reflectivity spectra. The fibers have a hole in the center, due to the choice of a hollow rod as the preform substrate, which experienced some nonuniform deformation during draw. The rolled-up mirror structure included a double outer layer of PES for mechanical protection, creating a noticeable absorption peak in the reflectivity spectrum at ~3.2 μm (Figure 9a).

A combination of spectral and direct imaging data demonstrates excellent agreement with the photonic band diagram. The measured gap width (range to mid-range ratio) of the fundamental gap for the 400 μm OD fiber is 27%, compared to 29% in the photonic band diagram.

Tunable 'Fabry–Perot' Fibers

The fabrication of fibers surrounded by or lined with alternating layers of materials with a large disparity in their refractive indices, presents interesting opportunities for passive and active optical devices. While a periodic multilayer structure, such as the one reported above, leads to the formation of photonic bandgaps and an associated range of high reflectivity, it is the incorporation of intentional deviations from periodicity, also called 'defects', which allows for the creation of localized electromagnetic modes in the vicinity of the defect. These structures, sometimes called optical cavities, in turn can provide the basis for a large number of interesting passive and active optical devices such as vertical cavity surface emitting lasers (VCSELs), bi-stable switches, tuneable dispersion compensators, tuneable drop filters, etc. Here we report on the fabrication of a fiber surrounded by a Fabry–Perot cavity structure and demonstrate that by application of axial mechanical stress, the spectral position of the resonant Fabry–Perot mode can be reversibly tuned.

Structure and Optical Properties of the Fabry–Perot Fibers

The fabrication technique described above allows for the accurate placement of optical cavities that can encompass the entire or partial fiber circumference. The fibers discussed above are made of As_2Se_3 and PES and have a low index Fabry–Perot cavity (Figure 10).

This structure was achieved by introducing an extra polymer layer in the middle of the periodic multilayer structure of the preform, thus generating a defect mode in the photonic bandgaps of the drawn fibers. The position of the bandgap center is linearly related to the optical thickness of the layers by the Bragg condition. Through accurate outer diameter control afforded by the laser micrometer mounted on the draw tower we have been able to place gaps (Figure 11) at wavelengths ranging from 11 microns to below 1 micron. Such cost-effective tuneable optical filters could lead to applications such as optical

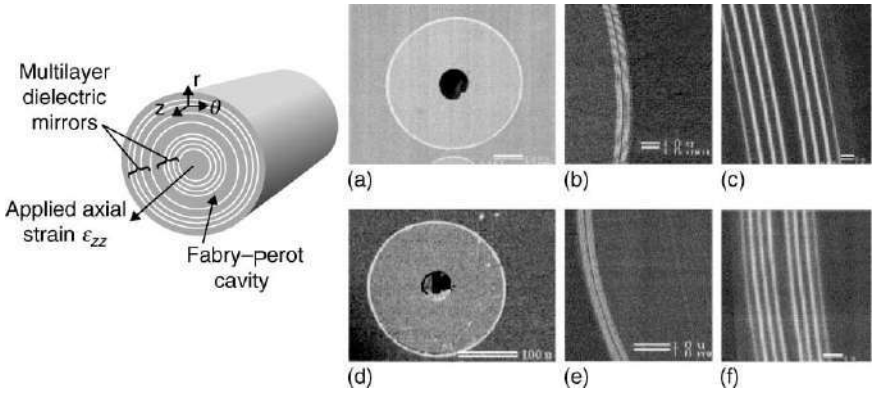

Figure 10 Schematic of the structure of dielectric mirror fibers made of As_2Se_3 (light gray) and PES (dark gray) with low index Fabry–Perot cavity. The local cylindrical coordinate system is represented as well as the applied axial strain ε_{zz}. Typical radii are ~150 μm ± 100 μm. Backscattered SEM micrographs of the cross-sections of a 460 micron diameter fiber (a, b, c) and of a 240 micron diameter fiber (d, e, f) embedded in epoxy and microtomed. (a) and (d) show the entire cross-section of the fibers, (b) and (e) demonstrate long-range layer uniformity, and (c) and (f) reveal the ordering and adhesion of the Fabry–Perot cavity structure. Bar scales have been redrawn for clarity. Reproduced with permission from Benoit G, *et al.* (2003) Static and dynamic properties of optical cavities in photonic bandgap gains. *Advanced Materials* 15: 2053–2056.

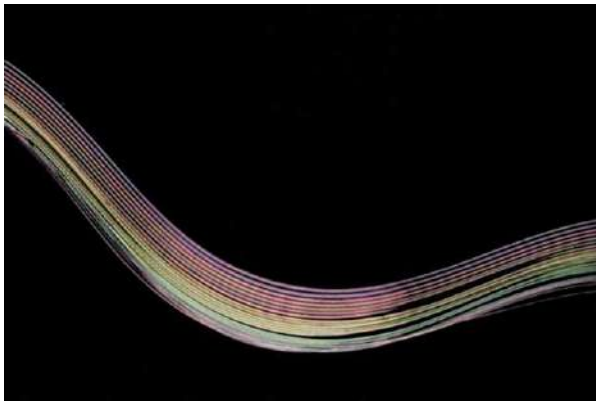

Figure 11 Array of parallel fibers with outer diameter ranging from ~420 μm (bottom) to ~100 μm (top). The colors are due to the narrow 4th order photonic bandgap in the visible.

switches for wavelength-division-multiplexing (WDM) systems and sensors.

The cross-sectional structures of a 240 micron and a 460 micron diameter fiber were observed by an SEM using a backscattered electron detector (Figure 10). The structure is composed of a hollow core polymer rod surrounded by six bilayers of As_2Se_3 and PES, separated in the middle by an extra polymer layer, which forms the Fabry–Perot cavity. An extra polymer layer protects the fiber surface. For the 240 micron (460 micron) diameter fiber, the glass layers are ~135 nm (250 nm) thick, except for the first and the last ones which are half as thick due to the fabrication technique; the polymer layers are ~270 nm (540 nm) thick and the defect layer is ~610 nm (1170 nm) thick.

Reflectivity spectra measurements were performed under a microscope using a (Nicolet SpectraTech NicPlan) Infrared Microscope and Fourier Transform Infrared Spectrometer (Magna 860) with a lens numerical aperture (NA) corresponding to 30 degrees of angular spread. They exhibited a single-mode Fabry–Perot resonant mode at 1.74 and 3.2 μm for the 240 and 460 micron diameter fiber, respectively (Figure 12).

Because of the range of incident angles, the measured quality factor (Q-factor, defined as its spectral position divided by the full width half maximum (FWHM)) was equal to 31. Using the different thicknesses and the real and imaginary part of the refractive index (n & k) of As_2Se_3 and PES carefully measured with a broadband (300 nm to 16 microns) spectroscopic ellipsometer (Sopra GES-5), the reflectivity spectra of these fibers were computed with the TMM by approximating them as a one-dimensional planar stacking. By averaging the calculated spectra over the accessible incident angles

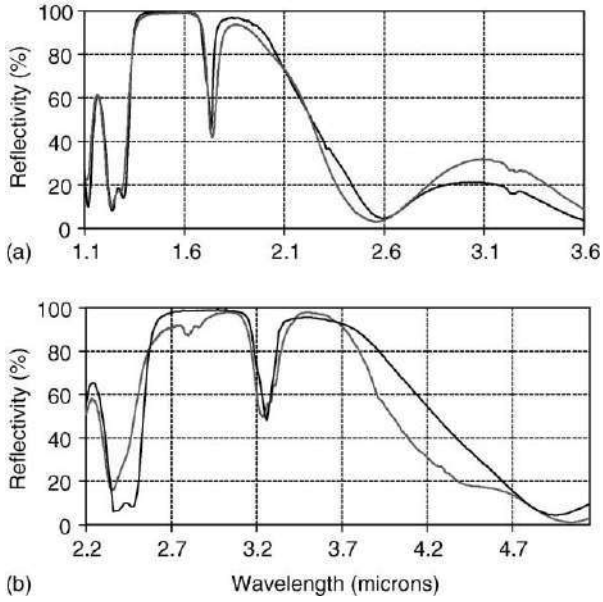

Figure 12 Computed (black lines) and measured (grey lines) reflectivity spectra for the 240 micron (a) and of the 460 micron diameter fibers (b) with Fabry–Perot resonant modes at 1.74 and 3.2 μm, respectively. Reproduced with permission from Benoit G, *et al.* (2003) Static and dynamic properties of optical cavities in photonic bandgap gains. *Advanced Materials* 15: 2053–2056.

and polarizations, we obtained a very good agreement between the simulation and the measurements.

Simulation of the Opto-Mechanical Behavior of the Fabry–Perot Fibers

We focused our analysis on the elastic regime, which ultimately limits the operational range of these fibers. The fiber's Young's modulus (E) can be approximated by modeling this multilayer structure as independent parallel springs under axial (along the z-axis of the fiber) strain assumed to be equal for all the layers, leading to a Young's modulus of 2.64 GPa (using $E_{\mathrm{PES}} = 2.4$ GPa for bulk PES and $E_{\mathrm{As_2Se_3}} = 15$ GPa reported for 1.5 μm thick films).

Neglecting the possible strain-induced refractive index variation of the materials, the normalized shift of the Fabry–Perot resonant mode can be related to the applied axial strain by calculating the radial stresses σ_{rr} (resulting from the difference between the Poisson ratios of the materials – $\nu_{\mathrm{PES}} = 0.45$ and $\nu_{\mathrm{As_2Se_3}} = 0.289$ – and the adhesion condition between the layers) and displacements u_r in each layer under axial strain. Starting from the equilibrium equations:

$$\frac{\partial \sigma_{zz}}{\partial z} = 0 \qquad [4]$$

$$\frac{\partial \sigma_{rr}}{\partial r} + \frac{\sigma_{rr} - \sigma_{\theta\theta}}{r} = 0 \Leftrightarrow \frac{\mathrm{d}^2 u_r}{\mathrm{d}^2 r} + \frac{1}{r}\frac{\mathrm{d}u_r}{\mathrm{d}r} - \frac{u_r}{r^2} = 0 \qquad [5]$$

and using Hooke's law and Lame's equations, we can derive a general expression for u_r and σ_{rr}, with two unknowns A and B per layer:

$$u_r = \frac{A}{r} + Br \qquad [6]$$

$$\sigma_{rr} = -\frac{2GA}{r^2} + 2(\eta + G)B + \lambda\varepsilon_{zz} \qquad [7]$$

where $G = E/[2(1+\nu)]$ is the shear modulus and $\eta = E\nu/[(1+\nu)(1-2\nu)]$ the Lame modulus. The continuity of the displacement and the radial stress at each interface, plus the boundary conditions (σ_{rr} vanishes at free surfaces) can be expressed as a linear system whose unique solution allows us to relate the radial strain in each layer to the applied axial strain as a linear relation $\varepsilon_{rr} = C\varepsilon_{zz}$, where C can be interpreted as an effective Poisson ratio. Finally, taking into account that the Fabry–Perot resonant mode itself is linearly shifted within the bandgap because of the different effective Poisson ratios of the glass and polymer layers, we obtain a linear relation between the normalized shift of the Fabry–Perot resonant mode and the applied axial strain:

$$\frac{\Delta\lambda}{\lambda} = -0.373\varepsilon_{zz} \qquad [8]$$

All the layers (except the outer protective polymer layer) are under tensile radial stress whose maximum (0.22 MPa under 1% axial strain) is located at the interface between the layers and the polymer core where delamination is most likely to occur.

Mechanical Tuning Experiment and Discussion

Measurements were performed on fibers ~30 cm long, which were fixed at one end with epoxy to a load cell (Transducer Techniques MDB-2.5) while the other end was attached with strong tape to a pole mounted on a stepper rotational stage (Newport PR50) (Figure 13). This end of the fiber was also screwed to the pole to further secure it in place.

The diameter of the pole was equal to 2.1 cm, leading to a normalized shift precision below 0.005%. The uncertainty on the Young's modulus, due to the precision of the load cell, was lower than 30 MPa. All the reflectivity spectra were normalized to a background taken with a flat gold mirror. The measurements were realized as far as possible from the fixed ends of the fiber where edge effects are likely to occur. These edge effects result in a reduction of the length of the fiber that deforms uniformly and consequently increase the real strain far from the edges by a factor of 1.14 and 1.15 for the 240 and the 460 micron diameter fiber, respectively (determined experimentally by measuring the position of two reference points on the fiber) compared to the strain calculated from the rotation of the pole. Moreover, to avoid measuring slight variations in the spectral position of the bandgap resulting from outer diameter variations, typically of the order of 4–5 microns over meters of fibers, the measurements were realized at a fixed reference position on the fiber.

By focusing on the fundamental bandgap of the 240 and the 460 micron diameter fiber, we demonstrated the tuning of the Fabry–Perot resonant mode under increasing axial strain (Figure 14).

A drop in the reflectivity of 13% was observed at 1.71 μm (dash line) for the 240 micron diameter fiber when increasing the applied axial strain from 0.23% (light gray) to 1.07% (dark gray). The normalized-shift versus strain curves (Figure 15) appeared to be linear for both fibers up to approximately 0.9% (dash line) with a slope equal to −0.3859 and −0.3843 for the 240 and the 460 micron diameter fiber, respectively, close to the predicted value (−0.373, eqn [8]).

The normalized shift was equal to −0.347% for 0.9% applied axial strain, which seems to be the limit of the elastic regime and corresponds to small applied

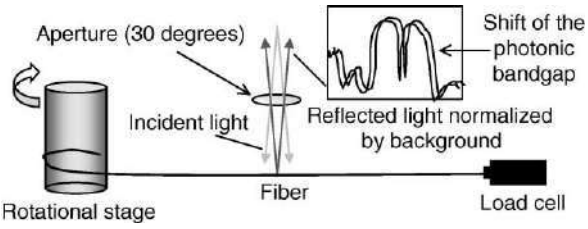

Figure 13 Experimental setup used for mechanical tuning demonstration. Reproduced with permission from Benoit G, *et al.* (2003) Static and dynamic properties of optical cavities in photonic bandgap gains. *Advanced Materials* 15: 2053–2056.

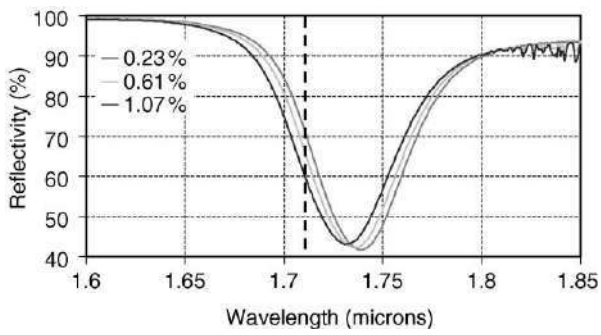

Figure 14 Reflectivity versus wavelength plot showing the shift of the Fabry–Perot resonant mode of the 240 micron diameter fiber for three increasing values of the applied axial strain. Reproduced with permission from Benoit G, *et al.* (2003) Static and dynamic properties of optical cavities in photonic bandgap gains. *Advanced Materials* 15: 2053–2056.

loads: 76 and 300 g for the 240 and 460 micron diameter fiber, respectively. Under higher strains, the normalized-shift versus strain curves were no longer linear and the measured loads would start decreasing slightly with time, which could be a sign of delamination between the layers and the polymer core, of plastic deformation and/or of relaxation in the layers. The stress–strain curves also exhibit an elastic regime up to approximately 1% strain with a corresponding Young's modulus equal to 2.59 GPa for the 240 μm OD fiber and to 2.39 GPa for the 460 μm OD fiber. These values are slightly lower than the predicted value, possibly due to relaxation effects in the polymer and the glass layers (for As_2Se_3 and PES, $T_g = 175$ and 220 °C, respectively).

Wavelength-Scalable Hollow Optical Fibers with Large Photonic Bandgaps for CO_2 Laser Transmission

Hollow optical transmission fibers offer the potential to circumvent fundamental limitations associated with conventional index guided fibers and thus have been the subject of active research in recent years. Here we report on the materials selection, design, fabrication, and characterization of extended lengths of hollow optical fiber lined with an interior omnidirectional dielectric mirror. These fibers consist of a hollow air core surrounded by multiple alternating submicron-thick layers of a high-refractive-index glass and a low-index polymer, resulting in large infrared photonic bandgaps. These gaps provide strong confinement of optical energy in the hollow fiber core and lead to light guidance in the fundamental and up to fourth-order gaps. We show that the fiber transmission windows can be scaled over a large wavelength range covering at least 0.75 to 10.6 microns. The utility of our approach is further demonstrated by the design and fabrication of tens of meters of hollow photonic bandgap fibers for 10.6 micron radiation transmission. We demonstrate transmission of carbon dioxide (CO_2) laser light with high power-density through more than 4 meters of hollow fiber and measure the losses to be less than 1.0 dB/m at 10.6 microns. This establishes suppression of fiber waveguide losses by orders of magnitude compared to the intrinsic fiber material losses.

Silica optical fibers have been extremely successful in telecommunications applications, and other types of solid-core fibers have been explored at wavelengths where silica is not transparent. However, all fibers that rely on light propagation principally through a solid material have certain fundamental limitations stemming from nonlinear effects, light absorption by electrons or phonons, material dispersion, and Rayleigh scattering that limit maximum optical transmission power and increase attenuation losses. These limitations have, in turn, motivated the study of a fundamentally different light-guiding methodology: the use of hollow waveguides having highly reflecting walls. Light propagation through air in a hollow fiber eliminates or greatly reduces the problems of nonlinearities, thermal lensing, and end-reflections, facilitating high-power laser guidance and other applications which may be impossible using conventional fibers. Hollow metallic or metallo-dielectric waveguides have been studied fairly extensively and found useful practical application, but their performance has been bounded by the notable losses occurring in metallic reflections at visible and infrared (IR) wavelengths, as well as by the limited length and mechanical flexibility of the fabricated waveguides. Hollow all-dielectric fibers, relying on specular or attenuated total reflection, have also been explored, but high transmission losses have prevented their broad application. More recently, all-dielectric fibers,

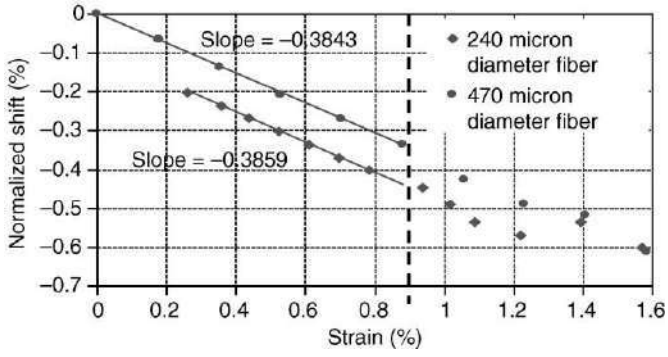

Figure 15 Normalized shift of the Fabry–Perot resonant mode (defined as $\Delta\lambda/\lambda$) versus applied axial strain for the 240 micron (diamonds) and the 460 micron diameter fiber (dots). The curve obtained for the 240 micron diameter fiber is lower because the experimental 0% strain is likely to correspond to a nonzero positive strain, necessary to keep this thinner fiber straight for the measurement (equivalent to a load less than 10 g). Reproduced with permission from Benoit G, *et al.* (2003) Static and dynamic properties of optical cavities in photonic bandgap gains. *Advanced Materials* 15: 2053–2056.

consisting of a periodic array of air holes in silica, have been used to guide light through air using narrow photonic bandgaps. Solid-core, index-guiding versions of these silica photonic crystal fibers have also been explored for interesting and important applications, such as very large core single-mode fibers, nonlinear enhancement and broadband supercontinuum generation, polarization maintenance, and dispersion management. However, the air-guiding capabilities of such waveguides thus far remain inferior to transmission through solid silica, due to various factors such as the difficulties in fabricating long, uniform fibers which must have a high volume fraction of air and many air–hole periods, as well as by the large electromagnetic (EM) penetration depths associated with the small photonic bandgaps achievable in these air–silica structures. In our fiber, the hollow core is surrounded by a solid high-refractive-index-contrast multilayer structure leading to large photonic bandgaps and omnidirectional reflectivity. The pertinent theoretical background and recent analyses indicate that such fibers may be able to achieve ultralow losses and other unique transmission properties. The large photonic bandgaps result in very short EM penetration depths within the layer structure, significantly reducing radiation and absorption losses while increasing robustness. Omnidirectional reflectivity is expected to reduce intermode coupling losses.

To achieve high index contrast in the layered portion of the fiber, we combined a chalcogenide glass with a refractive index of ~2.8 As_2Se_3, and a high-performance polymer with a refractive index of ~1.55 PES. We recently demonstrated that these materials could be thermally co-drawn into precisely layered structures without cracking or delamination, even under large temperature excursions. The same polymer was used as a cladding material, resulting in fibers composed of ~98% polymer by volume (not including the hollow core) and thus combine high optical performance with polymeric processability and mechanical flexibility. We fabricated a variety of fibers by depositing a 5–10 micron thick As_2Se_3 layer through thermal evaporation onto a 25–50 micron thick PES film and the subsequent 'rolling' of that coated film into a hollow multilayer tube called a fiber preform. This hollow macroscopic preform was consolidated by heating under vacuum and cladded with a thick outer layer of PES; the layered preform was then placed in an optical fiber draw tower and drawn down into tens or hundreds of meters of fiber having well-controlled submicron layer thicknesses. The nominal positions of the photonic bandgaps were determined by laser monitoring of the fiber OD during the draw process. Typical standard deviations in the fiber OD were ~1% of the OD. The resulting fibers were designed to have large hollow cores, useful in high-energy transmission.

SEM analysis (Figure 16) reveals that the drawn fibers maintain proportionate layer thickness ratios and that the PES and As_2Se_3 films adhere well during rigorous thermal cycling and elongation. Within the multilayer structure shown in Figure 1, the PES layers (gray) have a thickness of 900 nm, and the As_2Se_3 layers (bright) are 270 nm thick (except for the first and last As_2Se_3 layers, which are 135 nm). Broadband fiber transmission spectra were measured with a Fourier transform infrared (FTIR) spectrometer (Nicolet Magna 860), using a parabolic mirror to couple light into the fiber and an external detector. The results of these measurements are shown in the lower panel of Figure 2 for fibers having two different layer structures. For each spectrum, light is guided at the fundamental and high-order photonic bandgaps. Also shown in the upper panel of Figure 2, is the corresponding photonic band diagram for an infinite periodic multilayer structure calculated using the experimental parameters of our fiber (layer thicknesses and indices). Good agreement is found between the positions of the measured transmission peaks and the calculated bandgaps, corroborated with the SEM-measured layer thicknesses, verifying that transmission is dominated by the photonic bandgap mechanism. In order to demonstrate 'wavelength scalability' (i.e., the control of transmission through the fiber's structural parameters) another fiber was produced, having the same cross-section but with thinner layers. We compared the transmission

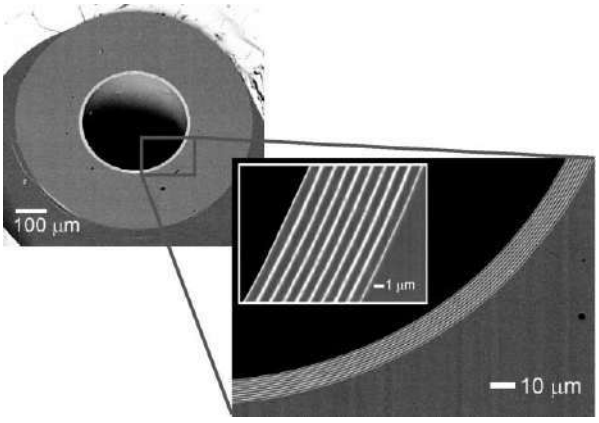

Figure 16 Cross-sectional SEM micrographs at various magnifications of hollow cylindrical multilayer fiber mounted in epoxy. The hollow core appears black, the PES layers and cladding gray, and the As_2Se_3 layers bright white. This fiber has a fundamental photonic bandgap at a wavelength of ~3.55 microns. Reproduced with permission from Temelkuran B, Hart SD, Benoit G, Joannopoulos JD and Fink Y (2002) Wavelength-scalable hollow optical fibres with large photonic bandgaps for CO_2 laser transmission. *Nature* 420: 650–653.

spectra for the original 3.55 micron bandgap fibers to the fiber with the scaled-down layer thicknesses.

Figure 17 shows the shifting of the transmission bands, corresponding to fundamental and high-order photonic bandgaps, from one fiber to the next. The two fibers analyzed in Figure 17 were fabricated from the same fiber preform using different draw-down ratios (fibers with a fundamental bandgap centered near 3.55 microns have an OD of 670 microns; those with a gap at 3.1 microns have an OD of 600 microns). The high-order bandgaps are periodically spaced in frequency, as expected for such a photonic crystal structure.

The wavelength scalability of our fibers was further demonstrated in the fabrication of hollow fibers designed for the transmission of 10.6 micron EM radiation. This not only shows that these structures can be made to guide light at extremely disparate wavelengths, but that specific useful bandgap wavelengths can be accurately targeted during fabrication and fiber drawing. Powerful and efficient CO_2 lasers are available that emit at 10.6 microns and are used in such applications as laser surgery and materials processing, but waveguides operating at this wavelength have remained limited in length or loss levels. Using the fabrication techniques outlined above, we produced fibers having hollow core diameters of 700–750 microns and ODs of 1300–1400 microns with a fundamental photonic bandgap spanning the 10–11 micron wavelength regime, centered near 10.6 microns. Figure 5 depicts a typical FTIR transmission spectrum for these fibers, measured using ~30 cm long straight fibers.

In order to quantify the transmission losses in these 10.6 micron bandgap hollow fibers, fiber cutback measurements were performed. This involved the comparison of transmitted intensity through ~4 meters of straight fiber with the intensity of transmission through the same section of fiber cut to shorter lengths (Figure 18 inset). This test was performed on multiple sections of fiber, and the results found to be nearly identical for the different sections tested. The measurements were performed using a 25 watt CO_2 laser (GEM-25, Coherent-DEOS) and high power detectors (Newport 818T-10). The fiber was held straight, fixed at both ends as well as at multiple points in the middle to prevent variations in the input coupling and propagation conditions during fiber cutting. The laser beam was sent through focusing lenses as well as 500 micron diameter pinhole apertures and the input end face of the fiber was coated with a metal film to prevent accidental

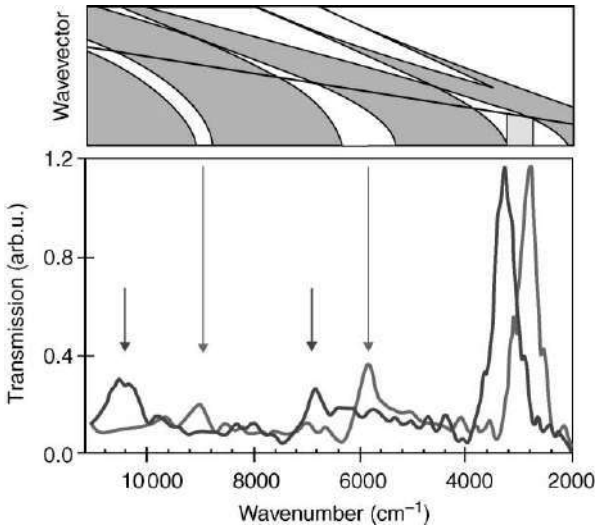

Figure 17 Upper panel: Calculated photonic bandstructure associated with the dielectric mirror lining of the hollow fiber. Modes propagating through air and reflected by the fiber walls lie in the bandgaps (white) and within the light cone defined by the glancing-angle condition (black line). The gray regions represent modes radiating through the mirror. The fundamental bandgap has the widest range-to-mid-range ratio, with a region of omnidirectional reflectivity highlighted in black. Lower panel: Comparison of transmission spectra for two different hollow fibers of ~30 cm length having similar structure but scaled photonic crystal (multilayer) period dimensions. The spectrum in gray is from a fiber with a fundamental photonic bandgap at 3.55 microns; the black spectrum is from a fiber where the corresponding bandgap is at 3.1 microns. High-order bandgaps (indicated by arrows) are periodically spaced in frequency. Reproduced with permission from Temelkuran B, Hart SD, Benoit G, Joannopoulos JD and Fink Y (2002) Wavelength-scalable hollow optical fibres with large photonic bandgaps for CO_2 laser transmission. *Nature* 420: 650–653.

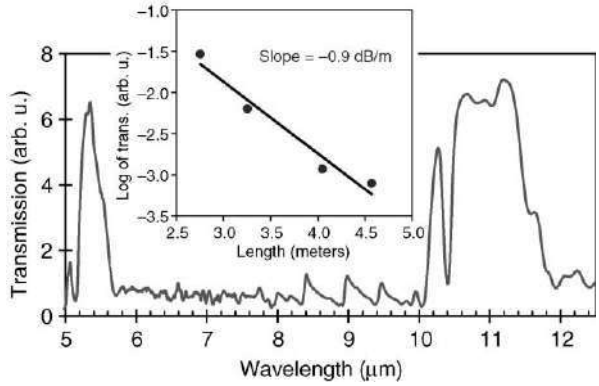

Figure 18 Typical transmission spectrum of hollow fibers designed to transmit CO_2 laser light. The fundamental photonic bandgap is centered near a wavelength of 10.6 microns and the second-order gap is at ~5 microns. Inset: Log of transmitted power (arbitrary units) versus length of fiber (meters). The slope of this graph is the loss in dB/m. The measured fiber has a hollow core diameter of 700 microns. Reproduced with permission from Temelkuran B, Hart SD, Benoit G, Joannopoulos JD and Fink Y (2002) Wavelength-scalable hollow optical fibres with large photonic bandgaps for CO_2 laser transmission. *Nature* 420: 650–653.

laser damage from misalignment. The transmission losses in the fundamental bandgap at 10.6 microns were measured to be 0.95 dB/m, as shown in the inset of Figure 18, with an estimated measurement uncertainty of 0.15 dB/m. These loss measurements are comparable to some of the best reported loss values for other types of waveguides operating at 10.6 microns. A bending analysis for fibers with a bandgap centered at 10.6 microns revealed bending losses below 1.5 dB for 90 degree bends with bending radii from 4–10 cm. We expect that these loss levels could be lowered even further by increasing the number of layers, through optimization of the layer thickness ratios and by creating a cylindrically symmetric multilayer fiber with no inner seam (present here because of the 'rolling' fabrication method). In addition, using a polymer with lower intrinsic losses should greatly improve the transmission characteristics.

One reasonable figure of merit for optical transmission losses through hollow all-dielectric photonic bandgap fibers is to compare the hollow fiber losses to the intrinsic losses of the materials used to make the fiber. As_2Se_3 has been explored as an IR-transmitting material, yet the losses at 10.6 microns reported in the literature are ~10 dB/m for highly purified material, and more typically are greater than 10 dB/m for commercially available materials such as those used in our fabrication. Based on FTIR transmission and spectroscopic ellipsometer measurements that we have performed on PES, the optical losses associated with propagation through solid PES should be greater than 40 000 dB/m at 10.6 microns. This demonstrates that guiding light through air in our hollow bandgap fibers leads to waveguide losses that are orders of magnitude lower than the intrinsic fiber material losses, which has been one of the primary goals of hollow photonic bandgap fiber research. These comparatively low losses are made possible by the very short penetration depths of EM waves in the high refractive index contrast photonic crystal structure, allowing these materials to be used at wavelengths that may have been thought improbable. Another long-standing motivation of infrared fiber research has been the transmission of high-power laser light. As a qualitative demonstration of the potential of these fibers for such applications, both straight and smoothly bent fibers of lengths varying from 0.3–2.5 meters were used to transmit enough CO_2 laser energy to burn holes through paper. The maximum laser power density coupled into our fibers in these trials was approximately 300 W/cm^2, more than sufficient to burn a homogeneous polymer material, including PES. No damage to the fibers was observed when the laser beam was properly coupled into the hollow fiber core. These results indicate the feasibility of using hollow multilayer photonic bandgap fibers as a low-loss wavelength-scalable transmission medium for high-power laser light.

See also

Diffractive Systems: Design and Fabrication of Diffractive Optical Elements. **Fiber and Guided Wave Optics:** Fabrication of Optical Fiber. **Lasers:** Carbon Dioxide Lasers. **Spectroscopy:** Fourier Transform Spectroscopy.

Further Reading

Abeles F (1950) Investigations on the propagation of sinusoidal electromagnetic waves in stratified media. Application to thin films. *Annales de Physique* 5: 706.

Benoit G and Fink Y, Spectroscopic Ellipsometry Database, http://mit-pbg.mit.edu/Pages/Ellipsometry.html

Born M and Wolf E (1980) *Principles of Optics*, 6th edn. New York: Pergamon Press, p. 67.

Engeness TD, Ibanescu M, Johnson SG, *et al.* (2003) Dispersion tailoring and compensation by modal interactions in OmniGuide fibers. *Optics Express* 11: 1175–1198. http://www.opticsexpress.org/abstract.cfm?URI = OPEX-11-10-1175

Fink Y, Winn JN, Fan S, *et al.* (1998) A dielectric omnidirectional reflector. *Science* 282: 1679–1682.

Fink Y, Ripin DJ, Fan S, Chen C, Joannopoulos JD and Thomas EL (1999) Guiding optical light in air using an all-dielectric structure. *Journal of Lightwave Technology* 17: 2039–2041.

Hart SD, Maskaly GR, Temelkuran B, Prideaux PH, Joannopoulos JD and Fink Y (2002) External reflection from omnidirectional dielectric mirror fibers. *Science* 296: 510–513.

Ibanescu M, Johnson SG, Soljacic M, *et al.* (2003) Analysis of mode structure in hollow dielectric waveguide fibers. *Physics Review E* 67: 046608-1–8.

Johnson SG, Ibanescu M, Skorobogatiy M, *et al.* (2001) Low-loss asymptotically single-mode propagation in large-core OmniGuide fibers. *Optics Express* 9: 748–779. http://www.opticsexpress.org/abstract.cfm?URI = OPEX-9-13-748

Kuriki K, Shapira O, Hart SD, *et al.* (2004) Hollow multilayer photonic bandgap fibers for NIR applications. *Optics Express* 12: 8.

Soljacic M, Ibanescu M, Johnson SG, Joannopoulos JD and Fink Y (2003) Optical bistability in axially modulated OmniGuide fibers. *Optics Letters* 28: 516–518.

Temelkuran B, Hart SD, Benoit G, Joannopoulos JD and Fink Y (2002) Wavelength-scalable hollow optical fibres with large photonic bandgaps for CO_2 laser transmission. *Nature* 420: 650–653.

Yeh P, Yariv A and Hong C-H (1977) Electromagnetic propagation in periodic stratified media. 1. General theory. *Journal of the Optical Society of America* 67(4): 423.

Yeh P, Yariv A and Marom E (1978) Theory of Bragg fiber. *Journal of the Optical Society* 68: 1196–1201.

Wave Optical Modeling and Design

F Wyrowski, Friedrich-Schiller University, Jena, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Why do ray optics dominate the literature on optical engineering? The major practical reason may be that historically optical engineering has dealt with systems for image forming. These systems transform divergent spherical waves from points in an object, into spherical waves converging to focal points in the image. Thus, the basic operation of an imaging system performed by a lens is changing the radius of curvature of spherical fields. The propagation of spherical waves is modeled well by geometrical optics, so long as truncation by apertures and other high frequency aberrations are not significant. Moreover, the basic merit function in imaging evaluates the appearance of the image point at its correct location by so-called dot diagrams. Evaluation of this merit function requires only a ray bundle representation of the output field. The combination of a ray representation with geometrical optics propagation is called raytracing, which explains the dominance of raytracing in conventional optical engineering. Because the design of illumination systems is also often based on principles of imaging, raytracing remains a suitable technique in that case too. However, systems that perform general light transformations must apply wave-optical engineering techniques, which constitutes a generalization of conventional optical engineering. In this sense it includes the modeling of image-forming systems as a special case.

Modeling Principles

Modeling in wave-optical engineering addresses the propagation of electromagnetic fields through systems as illustrated in Figure 1. The introduction of operators and operands as mathematical tools simplifies a systematic discussion of the modeling concepts. Formally, an electromagnetic field may be expressed by the operand $\mathbf{f} \in \mathsf{F}$ with the set of fields F. The actual form of $\mathbf{f} \in \mathsf{F}$ depends on the physical approach chosen to model electromagnetic fields. For instance, $\mathbf{f} \in \mathsf{F}$ may stand for a ray bundle, a complex amplitude, or a coherence function. The photon source generates a field which acts as the input field $\mathbf{f}$ into the optical system. The quality of the output field $\mathbf{f}_{\text{out}}$ is evaluated with the help of a merit function $\mathbf{\Omega}$. This function may consist of M components which yields:

$$\mathbf{\Omega}(\mathbf{f}_{\text{out}}) = \{\Omega_1(\mathbf{f}_{\text{out}}), \ldots, \Omega_\mu(\mathbf{f}_{\text{out}}), \ldots, \Omega_M(\mathbf{f}_{\text{out}})\} \quad [1]$$

with $\Omega_\mu : \mathsf{F} \rightarrow \mathsf{R}^+$ and $\mu = 1, \ldots, M$. Each component Ω_μ defines a mathematical way to obtain one positive valued quantity for a given $\mathbf{f}_{\text{out}}$. Examples are the rms-value to evaluate the quality of a wavefront, M^2-value to judge the profile of a laser beam, center of gravity of a light distribution, light transformation efficiency, and the value of an overlap integral to determine the coupling efficiency into a fiber. Any property of the output field can be evaluated by an appropriate definition of a function Ω_μ, as long as the demanded field parameters are accessible for the field model in use. However, it should be mentioned, that it is not clear *a priori*, that all quality criteria Ω_μ can be satisfied simultaneously. How the different Ω_μ are related is dependent on the application.

The propagation of the input field through the system is mathematically summarized by the operator equation:

$$\mathbf{f}_{\text{out}} = S(p_\alpha)\mathbf{f}_{\text{in}} \quad [2]$$

with the operator $S : \mathsf{F} \rightarrow \mathsf{F}$. p_α stands for all parameters p_1, p_α, p_{A} which specify the system, for instance position of elements, refractive index, surface profiles, and layer thickness. Obviously a full electromagnetic model of S does not exist for arbitrary systems. The propagation problem eqn [2] is much too complex for a wave-optical solution. However, for a sequential approach the situation is different. If the propagation through a system is performed sequentially through the elements and the homogenous regions, at least approximate wave-optical methods are available to model the propagation S. In a sequential approach, either the transmitted or reflected field is considered per region

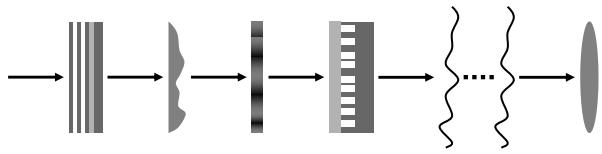

Figure 1 An optical system may include various types of elements to perform the desired optical function. Layered media, free-form surfaces, index-modulated regions, microstructured interfaces, lenses, and other structured media offer the flexibility needed to realize general light transformations. Typically structured media and homogenous regions follow each other.

of the system. Although this is an approximation, it turns out to be very useful in wave-optical engineering. With regard to Figure 2 the sequential approach is mathematically expressed by

$$\mathbf{f}_{\text{out}} = S(p_\alpha)\mathbf{f}_{\text{in}} = \mathcal{P}_J(n_J, \Delta z_J) S_{J-1}(p_{\alpha_{J-1}}) \cdots S_2(p_{\alpha_2}) \times \mathcal{P}_1(n_1, \Delta z_1)\mathbf{f}_{\text{in}} \quad [3]$$

$\mathcal{P}_j(n_j, \Delta z_j)$ expresses the propagation through a homogenous dielectric with refractive index n_j and extent Δz_j. This type of propagation is also called free propagation. Sometimes the term free propagation is restricted to propagation in vacuum; Here it is used more generally. $S_j(p_{\alpha_j})$ models the propagation through an optical component which is specified by the parameters p_{α_j}. Optical engineering needs physical models for the field operand $\mathbf{f}$, the free propagation $\mathcal{P}$, and the propagation S_j through elements. The most fundamental demand in wave-optical engineering is a wave-optical model to express the field $\mathbf{f}$ in homogenous regions. It is common to restrict to harmonic fields. The generalization to more general radiation is often possible by suitable superposition of harmonic fields. This approach assumes linear optical systems as typical in optical engineering. From Maxwell's equations it follows that harmonic fields in homogenous regions are completely specified if two independent field components are determined in one plane z_0 within the region. Therefore, the electromagnetic field operand takes the form:

$$\mathbf{f}(x, y, z_0) = [E_x(x, y, z_0), E_y(x, y, z_0)] \quad [4]$$

with the complex amplitudes E_x and E_y of the x and y components of the electric field vector. Because the knowledge of $\mathbf{f}(x, y, z_0)$ allows the direct calculation of the missing complex amplitudes E_z, H_x, H_y, H_z via Maxwell's equations, the entire electromagnetic harmonic field in a homogenous region is specified by eqn [4].

With respect to the two components of the field operand $\mathbf{f}$ propagation operators S are 2×2-matrices of the form:

$$S = \begin{Bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{Bmatrix} \quad [5]$$

In the case of a diagonal matrix, that is $S_{xy} = S_{yx} = 0$, E_x and E_y propagate independently and are considered decoupled channels of the system. If in addition $S_{xx} = S_{yy}$, both channels behave identically, and it is sufficient to model only one channel. This is often referred to as the scalar model. Of course, also in the scalar model the electromagnetic field still consists of two independent components and maintains its vectorial nature.

Numerical implementation of eqn [3] for the field operand of eqn [4] constitutes the backbone of modeling in wave-optical engineering.

Free Propagation

The propagation $\mathcal{P}$ of harmonic fields $\mathbf{f}$ through homogenous regions of a system has a rigorous electromagnetic solution in form of the spectrum of plane waves integral which yields:

$$\begin{aligned} \mathbf{f}(x,y,z \geq z_0) &= \mathcal{P}_{\text{SPW}}\mathbf{f}(x,y,z_0) \\ &= \begin{pmatrix} \mathcal{P}_{\text{SPW}} & 0 \\ 0 & \mathcal{P}_{\text{SPW}} \end{pmatrix} \begin{pmatrix} E_x(x,y,z_0) \\ E_y(x,y,z_0) \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{F}^{-1}([\mathcal{F}E_x(x,y,z_0)]e^{i(k_x x + k_y y + k_z \Delta z)}) \\ \mathcal{F}^{-1}([\mathcal{F}E_y(x,y,z_0)]e^{i(k_x x + k_y y + k_z \Delta z)}) \end{pmatrix} \end{aligned} \quad [6]$$

with the Fourier transform $\mathcal{F}$ and $\Delta z = z - z_0$. The term $\mathcal{F}E(x,y,z_0)$ is called spectrum of planewaves (SPW). The operator matrix is diagonal and $\mathcal{P}_{\text{SPW}} = \mathcal{P}_{xx} = \mathcal{P}_{yy}$, that means the x- and y-components propagate independent of each other according to the same formula in homogenous dielectrics.

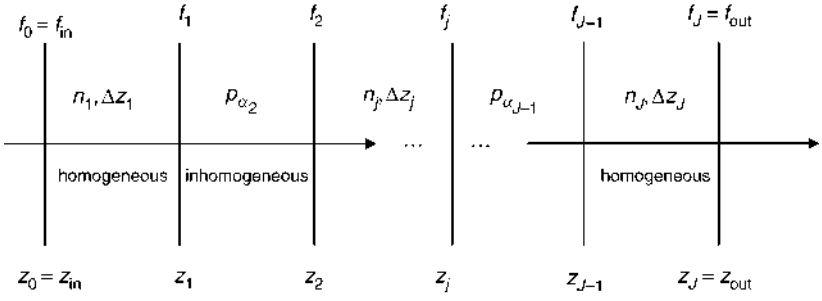

Figure 2 Formally a system may be subdivided into a sequence of homogenous dielectrics like air and inhomogeneous regions which include the elements. We indicate a region by the letter *j*.

In this sense, free propagation is a scalar problem which allows the convenient use of $U(x,y,z)$ instead of two field components for the sake of simpler mathematical expressions.

Equation [6] allows the derivation of other rigorous versions of the free propagation operator $\mathcal{P}_{\mathrm{SPW}}$, for instance Rayleigh's integral formula, which is well-known from diffraction theory. Moreover, eqn [6] leads to important approximate operators for far-field propagation and the propagation of paraxial fields. Propagation of a field into its far field is given by the operator:

$$\begin{aligned} U(r\mathbf{s}) &= \mathcal{P}_{\mathrm{far}} U(x, y, z_0) \\ &= -\mathrm{i}2\pi k s_z [\mathcal{F} U(x, y, z_0)]|_{(ks_x, ks_y)} \frac{\exp(\mathrm{i}kr)}{r} \end{aligned} \quad [7]$$

with the wave number $k = 2\pi n/\lambda$, $r = |\mathbf{r}| = (x^2 + y^2 + (\Delta z)^2)^{1/2}$, and the direction vector $\mathbf{s} = (s_x, s_y, s_z) = \mathbf{r}/r$. Here, and in what follows, λ denotes the vacuum wavelength. $|_{(ks_x, ks_y)}$ indicates the variables to be inserted into the Fourier transformed field. Though the far-field formula is derived for $r \to \infty$, it is also of great practical value for the propagation of fields generated by nonparaxially emitting sources like laser diodes. Roughly speaking eqn [7] predicts a spherical wave in the far field which is modulated by the spectrum of planewaves of the field in z_0. $\mathcal{P}_{\mathrm{far}}$ propagates from the plane z_0 to the spherical surface with radius r and not to a plane z.

The paraxial approximation of eqn [6] follows if the spectrum of planewaves $\mathcal{F}U(x, y, z_0)$ possesses significant values only for small k_x and k_y, that is if both electric field components propagate approximately along the z-axis. The resulting propagation operator is known as Fresnel's integral formula. In practice, its formulation with the Fourier transform $\mathcal{F}$ is typically more suitable:

$$\begin{aligned} U(x, y, z \geq z_0) &= \mathcal{P}_{\mathrm{paraxial}} U(x, y, z_0) \\ &= \frac{k \exp[\mathrm{i}k\Delta z]}{\mathrm{i}2\pi\Delta z} \exp\left[\frac{\mathrm{i}k(x^2 + y^2)}{2\Delta z}\right] \\ &\quad \times \left| \left\{ \mathcal{F}\left(\exp\left[\frac{\mathrm{i}k(x^2 + y^2)}{\Delta z}\right] U(x, y, z_0) \right) \right\} \right|_{(x,y)n/\lambda\Delta z} \end{aligned} \quad [8]$$

$|_{(x,y)n/\lambda\Delta z}$ indicates the variables to be inserted after Fourier transform. The far-field approximation of eqn [8] leads to the well-known Fraunhofer diffraction formula.

Equations [6]–[8] provide fundamental operators for the propagation of fields in wave-optical engineering. All integrals are based on Fourier transforms which allows the application of the fast Fourier transform (FFT) in practice. Dependent on the propagation technique, one or two FFTs must be performed. In practice, the sampling of fields is of great concern. Naturally the amount of data should be as small as possible. According to the rigorous propagation operator eqn [6] the (x, y)-coordinate system and therefore the sampling period in z_0 and z are identical, that is, scaling does not occur. As illustrated in Figure 3, that can lead to an enormous amount of data for large propagation distances. For demonstration purpose the field amplitude depicted at the top is chosen. The field window has the size 1 mm × 4 mm. The wavelength λ is 632.8 nm, $n = 1$, and the sampling distance 10 μm. That results in 100 × 400 sampling points. Propagation of this field with the spectrum of planewave operator maintains the size of the field window. The resulting field for a propagation distance $\Delta z = 10$ mm is shown in the middle. Because of diffraction, the field widens. As long as the initial window is large enough to encompass the propagated field, the propagation method works well. At the bottom, the resulting field for $\Delta z = 200$ mm is shown. The field exceeds the size of the window and that leads to numerical errors (aliasing). This can be avoided by embedding the initial field into a window which is large enough with respect to the expected size of the propagated field. As a result, one obtains the

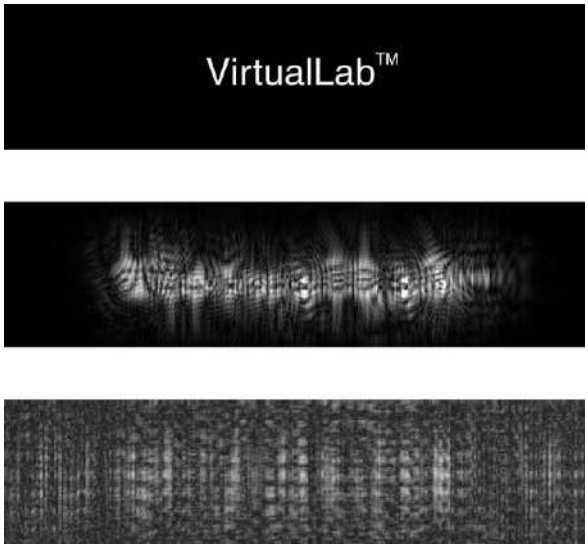

Figure 3 Illustration of numerical properties of the spectrum of plane wave propagation technique. The field amplitude depicted at the top is numerically propagated in free space and the distributions in the middle and the bottom are obtained. The simulations in this and all following figures were performed. The simulations in this and all following figures were performed with the wave-optical engineering software VirtualLab™ from LightTrans GmbH.

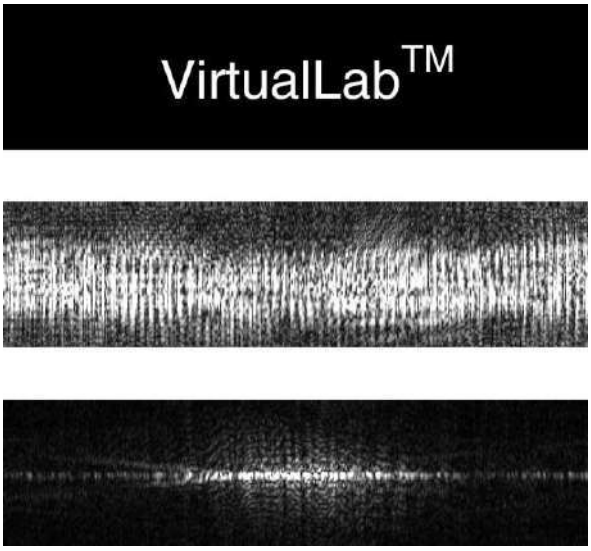

Figure 4 Starting from the same input field and the same distances as in Figure 3 the propagated fields depicted in the middle and the bottom are obtained using the paraxial propagation operator. Because of the inherent scaling the field window sizes are 632.8 μm × 632.8 μm and 12.656 mm × 12.656 mm, respectively. The number of sampling points remains unchanged. Obviously, the field in the middle is too small to encompass the propagated field and numerical errors occur. For the long distance the field size is perfectly matched to the field size obtained by propagation.

demanded accuracy at cost of increasing data memory requirements and computation time. Thus in practice the SPW propagation operator is particularly suitable for short distances. Fortunately the paraxial propagation operator eqn [6] possesses a complementary behavior. Because of $(x, y)n/\lambda\Delta z$, the (x, y)-coordinate system changes its scaling with the distance. As illustrated in Figure 4, that makes it well suited for large propagation distances but problems occur for small ones. Therefore, all paraxial propagation problems have an efficient numerical solution by a suitable combination of $\mathcal{P}_{\text{paraxial}}$ and $\mathcal{P}_{\text{SPW}}$. In the nonparaxial case the far-field operator eqn [7] often helps to realize a numerically efficient propagation for large distances. However, there are sometimes situations, in which free propagation of nonparaxial fields causes significant numerical problems.

Propagation Through Elements

The propagation of harmonic fields $\mathbf{f}$ through elements is a challenging task. For example, Figure 5 illustrates propagation through smooth or microstructured interfaces between two homogenous media. The field $\mathbf{f}_1(x, y, z_1)$ should be determined if $\mathbf{f}_0(x, y, z_0)$, the shape of the interface, and the

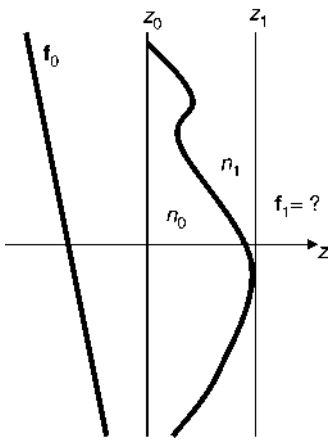

Figure 5 A planewave $\mathbf{f}_0$ propagates through an interface between two homogeneous dielectrics with refractive index n_0 and n_1. The interface is of general shape and therefore a rigorous solution of the propagation problem does not exist. The virtual planes in z_0 and z_1 encompass the interface in a way which minimizes the distance $z_1 - z_0$.

refractive indices n_0 and n_1 are known. The reflection mode of operation can be formulated analogously. It is remarkable that this basic problem possesses a rigorous solution only for a few special interface geometries. Of particular importance are the rigorous solutions for plane and periodic interfaces. In the first case, the solution comprizes Snell's law of refraction, the law of reflection, and Fresnel's formulae. Grating theory provides the solution for periodic interfaces. It is possible to use these methods also for nonperiodic interfaces but at the cost of immense computational effort. Thus, they are only practicable for structures with detail no larger than a few wavelengths. For general interfaces, accurate approximate methods are required. One approach is to decompose a general propagation problem into a set of simpler problems that can be solved rigorously. The results are then combined to obtain an approximate solution of the original problem. Two basic concepts for subdividing the general problem are: (i) Decomposition of the incident field into laterally truncated fragments which illuminate elementary fractions of the interface. (ii) Decomposition of the response of the interaction of the incident field with the interface into approximately independent ones related to local interface features. The first technique is referred to as local elementary interface approximation (LEIA) and the second one as local independent response approximation (LIRA). Elementary interfaces of particular importance are plane, periodic, and spherical ones.

The local plane interface approximation (LPIA) is of particular importance in wave-optical engineering. Figure 6 illustrates the basic concept for the example

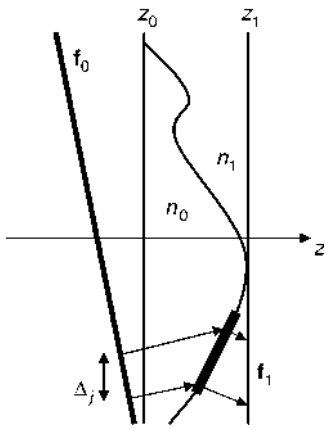

Figure 6 If a lateral fraction of the input field is propagated towards the interface, it is often possible to laterally restrict the major part of the propagated field to a local plane interface. Because of diffraction it is never possible to do that rigorously. Neglecting diffraction in the region between z_0 and z_1 provides a simple model to propagate a section of the incident plane field through the interface. Fresnel's formulae and Snell's law of refraction can then be applied directly to obtain the transmitted plane field, which is propagated to the plane z_1.

of a plane incident field. It has been proven that the approximation of geometrical optics propagation within the regions of the component, here that means between z_0 and z_1, provides results which are surprisingly accurate in numerous situations of practical interest. A full wave-optical version of LPIA is currently under research and development. In general, however, LPIA is applicable if the characteristic lateral extent of the local modulation of the interface is not close to the wavelength. It should be emphasized that LPIA is, as most other approximate propagation techniques, a 2×2 matrix, like eqn [5], with nonvanishing components. That means, for instance, that in general, an E_y-component occurs in z_1 even if $E_y = 0$ in z_0. That is, the two field channels experience crosstalk. A further approximation in geometrical optics LPIA, popular in Fourier optics holography, and diffractive optics is the thin element approximation (TEA). Geometrical optics LPIA leads to TEA if refraction and the local dependence of Fresnel losses at the interface are neglected. Obviously, that is allowed only if all angles involved, including the slope of the interface and the direction of the input field, are small. Therefore, TEA is a paraxial and geometrical optics approximation. Moreover, small local slopes of the interface lead to a restriction to thin elements, which explains the name TEA. It is also referred to as scalar approximation, because the operator matrix becomes diagonal with $S_{\mathrm{TEA}} = S_{xx} = S_{yy}$, that means the two field channels propagate independently according the same formula through the interface. In the thin element only approximation the dependence of the optical path through the element on position (x, y) is considered and mathematically takes the form:

$$\begin{aligned} U(x, y, z_1) &= S_{\mathrm{TEA}}\, U(x, y, z_0) \\ &= c_{\mathrm{Fresnel}}\, U(x, y, z_0) \exp\left[\frac{\mathrm{i}2\pi n_1(z_1 - z_0)}{\lambda}\right] \\ &\quad \times \exp\left[\frac{\mathrm{i}2\pi h(x, y)(n_0 - n_1)}{\lambda}\right] \end{aligned} \qquad [9]$$

with the vacuum wavelength λ. The profile height $h(x, y)$ is expressed as the distance from z_0 to the interface. Propagation through an interface between dielectrics modeled with TEA affects the phase of the incident field proportional to the profile height. The magnitude remains unchanged besides a constant factor c_{Fresnel}, due to Fresnel's equations. If complex valued refractive indices are taken into consideration, absorption results in an amplitude modulation completely analogous to the phase modulation described in eqn [9]. This way it is also possible to address the propagation through apertures and the combination of S_{TEA} with $\mathcal{P}$ leads to the well-known diffraction integrals. In fact, Kirchhoff's boundary conditions of field propagation through a hole in a completely opaque screen can be interpreted as TEA for complex valued refractive indices.

The simple form of the TEA operator allows a straightforward numerical implementation. The field operand and the profile are equidistantly sampled and the resulting field is sampled analogously. That makes the combination of S_{TEA} with numerical implementations of the free propagation operators $\mathcal{P}$ easy. As soon as refraction is considered in geometrical optics LPIA, the equidistant sampling of the output field is destroyed which enforces additional interpolation effort before a free propagation by fast Fourier transforms can be implemented, because fast numerical Fourier transform algorithms require an equidistant sampling.

In wave-optical engineering the combination of geometrical optics LPIA with a wave-optical field model is mandatory. In this context it is instructive to mention, that its combination with a ray-bundle representation of the incident electromagnetic field leads to raytracing through the interface, that is it reduces to the basic technique in conventional optical engineering.

Similar to LPIA it is possible to consider linear gratings as elementary local interface and the local linear grating approximation (LLGA) results. It is of special concern for all types of diffractive elements, for which the interface can locally considered

periodically. An important example is given for diffractive lenses in the outer zones of high numerical aperture lenses. In the center and for low numerical aperture TEA is appropriate for propagating through diffractive lenses.

Design Principles

The design problem in wave-optical engineering can be stated as follows: specify system parameters p_α in a way that ensures for a given input field $\mathbf{f}_{\text{in}}$ an output field $\mathbf{f}_{\text{out}}$ which satisfies the quality criteria defined by the merit functions Ω_μ. In addition, one may demand a system which is most simple, particularly cheap, or various other system properties of concern in practice.

Design constitutes an optimization problem. Restrictions to the system parameters and the merit function sets design constraints as well as design freedoms.

System parameters can be restricted, for instance, by the materials available, the maximum size of a system, and types of element profiles. Within these limits the system parameters are chosen freely during optimization to achieve the optical function.

The merit functions determines which parameters of the output field are important and should be emphasized, for instance the intensity distribution. All output fields, whose parameters satisfy the constraints satisfy the design problem. In most applications, not all field parameters are constrained. For instance, in laser materials processing or illumination applications, the merit function is related to energy quantities and then the phase of the output field can be used as a free parameter distribution. Moreover, the output field is typically not constrained in the entire output plane, but only in so-called signal regions, which introduces freedom of amplitude outside the signal regions. Hence, dependent on the application, there exist field parameters that can be chosen partly or completely free during design.

Successful design strategies use the freedoms to improve the matrix function. It is often not clear a priori, if a solution does exist. Then the field constraints are satisfied as good as possible in terms of the merit function.

In imaging systems, typical system parameters include, for instance, distances, refractive indices, curvature of lenses, and parameters to specify aspherical surfaces. For high-end lens systems, the resulting number of parameters can be large, but they are still small enough to utilize parameter optimization algorithms if the designer starts with a reasonable initial guess. In wave-optical engineering, the number of system parameters can become huge due to the generalized nature of the structures, which typically prevents the use of system parameter optimization techniques. For example, the profile of diffractive elements is typically specified by 10^4–10^6 parameters. Thus a direct optimization of the system parameters, which requires the evaluation of $S(p_\alpha)$ (see eqn [3]) after any change of system parameters p_α, is not realistic.

Instead of optimizing the system parameters directly, which is referred to as design in the structural embodiment of the system, a technique wellknown from Fourier optics and diffractive optics can be generalized for wave-optical engineering. It does not start with structure related element parameters but replaces all or at least some of the operators $S_j(p_{\alpha_j})$ of eqn [3] by mathematical operators $\mathcal{T}_j(t_{\tau_j})$. t_τ denotes the parameters of $\mathcal{T}$. The propagation operator through a system expressed in this functional embodiment has the form:

$$\begin{aligned}\mathbf{f}_{\text{out}} &= S_{\text{functional}}\mathbf{f}_{\text{in}} \\ &= \mathcal{P}_J(n_J, \Delta z_J)\mathcal{T}_{J-1}(t_{\tau_{J-1}})\cdots\mathcal{T}_2(t_{\tau_2}) \\ &\quad \times \mathcal{P}_1(n_1, \Delta z_1)\mathbf{f}_{\text{in}}\end{aligned} \qquad [10]$$

if all propagations through elements are expressed in functional form. A very important example of a functional operator $\mathcal{T}$ is given by a simple multiplication of both field components with identical or different complex functions $t(x, y)$, that is:

$$\mathcal{T}_t(t_\tau)\mathbf{f}_{\text{in}} = \begin{pmatrix} t_{xx}(x, y) & 0 \\ 0 & t_{yy}(x, y) \end{pmatrix}\mathbf{f}_{\text{in}}(x, y) \qquad [11]$$

The values of both $t(x, y)$ form the parameters t_τ of $\mathcal{T}$. The functions $t(x, y)$ are also called ideal transmission functions of elements. Of special interest are phase-only transmission functions $t(x, y) = \exp[i\tau(x, y)]$ (see discussion below). If $t(x, y) = t_{xx}(x, y) = t_{yy}(x, y)$, the design is simplified. Fortunately, in most actual applications this assumption is valid.

Designing in the functional embodiment requires the evaluation of $S_{\text{functional}}$ instead of S of eqn [3], which reduces the numerical complexity enormously. That improves the likelihood for developing efficient design algorithms to optimize the distances Δz_j, the refractive indices n_j, and parameters t_{τ_j}.

The design in the functional embodiment does not deliver the structure parameters of elements but only parameters t_τ of the operator $\mathcal{T}$ which gives information about how a field should be changed at the position z where $\mathcal{T}$ takes effect. Thus, in a next design step, elements must be obtained which cause the required effect. This so-called structure design

step is also generally challenging because it requires the solution of an inverse propagation problem through interfaces. For a further discussion the most important $\mathcal{T}$, that is the transmission type operator of eqn [11], is considered. For the sake of simplicity $t(x, y) = t_{xx}(x, y) = t_{yy}(x, y)$ is assumed. Then, a solution which uses geometrical optics LPIA in combination with phase-only transmissions is known in the form of a recursive structure design strategy. In its paraxial approximation, that is TEA, the solution becomes very simple. For a transmission $t(x, y) = \exp[\mathrm{i}\tau(x, y)]$, obtained from a functional design, an element with a surface profile:

$$h(x, y) = \frac{\lambda\tau(x, y)}{2\pi(n_{\text{element}} - n_{\text{free}})} \qquad [12]$$

results from eqn [9]. The refractive index of the element is denoted by n_{element} and the homogenous region behind the element has the index n_{free}, eqn [12] constitutes a direct relationship between the phase values of the transmission function and the profile $h(x, y)$. This structure design technique is accurate for all paraxial situations. However, it is also a reasonable method for nonparaxial structure design since it provides an initial profile for further optimization. Obviously it is applicable for phase-only transmission functions only. That underlines the importance of this type of functional operator for the design in the functional embodiment. Often the fabrication of profiles is done with lithographic techniques which result in surface reliefs with a few discrete height levels, that means the profile $h(x, y)$ is quantized. According to eqn [12] a quantization of $h(x, y)$ directly leads to quantized phase values $\tau(x, y)$. This quantization must be taken into account when designing $\tau(x, y)$. An example should serve to discuss further aspects of structure design according to eqn [9]. A fundamental transmission in various applications is the phase of a spherical wave and its paraxial version, the quadratic phase. An example of a spherical phase $\tau(x, y)$ is shown in Figure 7 where the 2π-modularity of the exponential function is taken into account – as normal for phase data which has been obtained by numerical calculations. Inserting this phase into eqn [9] leads to the profile of a diffractive lens. Because of the use of inverse TEA for structure design it works as demanded for moderate numerical apertures and starts to suffer from aberrations for large numerical apertures because of electromagnetic effects in the very outer zones. It is also possible to unwrap the phase before inserting it into eqn [9]. Then a smooth surface, that is a refractive spherical lens, results. It is well-known that a spherical lens works accurately in the paraxial domain and suffers from aberrations for moderate and large numerical apertures, which are more severe than for the diffractive lens. The use of inverse geometrical optics LPIA as a structure design technique is not restricted to the paraxial domain and therefore it leads to a suitable aspherical surface in order to ensure a spherical phase transmission for larger numerical apertures too. This is illustrated in Figure 8. The lens example shows another advantage of the two-step design method, which consists of the design in the functional embodiment prior to the one in the structural form of the element. In the functional

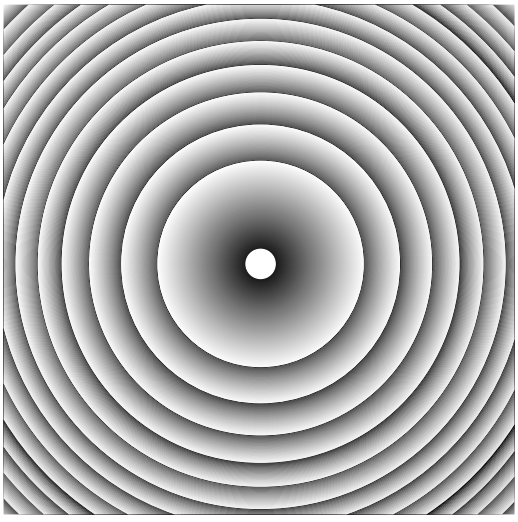

Figure 7 The phase of a spherical phase in the 2π-modulo representation. This phase can be unwrapped which leads to a smooth spherical phase which varies from 0 to various 2π. The noisy edges are due to the use of a cartesian coordinate system.

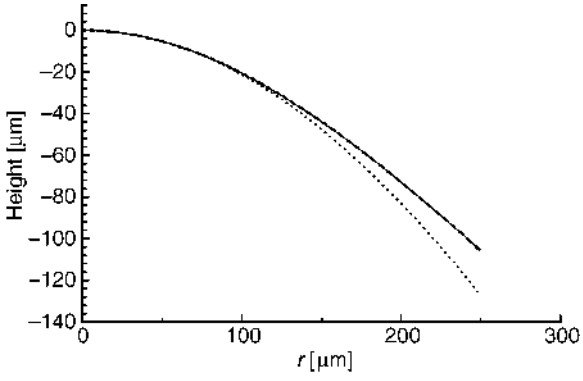

Figure 8 Application of the thin lens approximation for structure design leads to a spherical surface for a spherical phase-only transmission. The corresponding surface is depicted by the dotted line. If geometrical optics LPIA is used for structure design, an aspherical surface is obtained as required. The aspherical surface is depicted by the solid line. The height of the surfaces are shown as function of the radius *r*. Naturally the difference between TEA and geometrical optics LPIA occurs for larger *r* only.

form the transmission is obtained. It is not necessary to decide about the use of a diffractive or a refractive element in an early stage of the design. Moreover, it provides the designer with insight into selecting optical elements to achieve the function. If the phase-only transmission can be unwrapped, a refractive as well as a diffractive element can be used. An optical engineer has the freedom to decide about the suitable choice taking fabrication, size, costs, and other additional constraints into account. However, as soon as the phase-only transmission includes so-called vortices, unwrapping is not possible and diffractive optics is mandatory. An example of a vortex is shown in Figure 9. It turns out that vortices are more likely to appear in wave-optical engineering the less image-like the optical function is (see also last section). It is important to note, that functional design assumes a diagonal operator matrix, that is, it excludes crosstalk of the two field channels. Transferring the system into a structural form may violate this assumption because of vectorial effects in the nonparaxial domain. For instance, geometrical optics LPIA in general does not correspond to a diagonal operator. Therefore, it is necessary to analyze the final system for $\mathbf{f}_{\text{in}}$ – even in the case of linear polarized light – using more accurate propagation techniques than S_{TEA} if it is not ensured that the paraxial approximation is accurate enough.

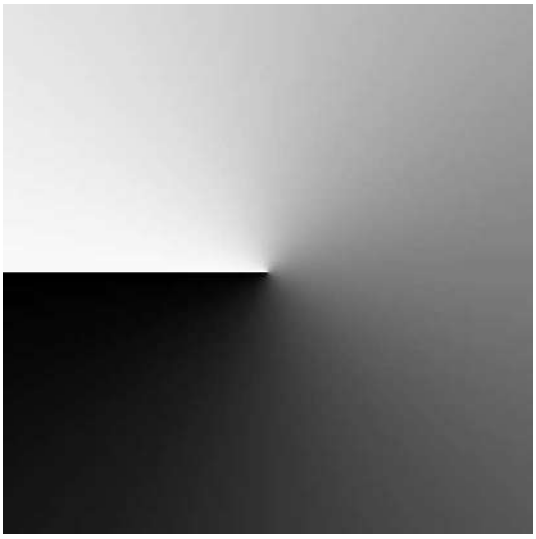

Figure 9 Illustration of a single vortex of charge one. The phase varies from 0 to 2π. Because of the vortex, this phase is not unwrappable, that means a smooth phase function cannot be obtained. The transition in the phase function requires a microstructure in the profile of the element, which generates this phase-only transmission.

The so-called thin lens, or ideal lens, in conventional optical engineering constitutes a special case of the two-step design strategy. In the first design step, distances and lens focal lengths are determined, without taking into account the actual shape of a lens. In the second step lenses, or even lens systems, are inserted according to the position and the focal length specified in the first step. Thus, wave-optical engineering extends our understanding by broadening established methods in optical engineering.

Fundamental Systems

A smart approach to design minimizes the number of elements required to achieve a desired light transformation. Thus, it is logical to try a functional design with a system of the form:

$$\begin{aligned} U_{\text{out}}(x,y) &= S_{\text{functional}} U_{\text{in}}(x,y) \\ &= \mathcal{P}(n_{\text{free}}, \Delta z_{\text{out}}) \exp[\mathrm{i}\tau(x,y)] \\ &\quad \times \mathcal{P}(n_{\text{free}}, \Delta z_{\text{in}}) U_{\text{in}}(x,y) \end{aligned} \qquad [13]$$

which is illustrated in Figure 10. One phase-only transmission, properly placed between the input and the output plane, should realize the desired transformation of the function. Scalar terminology is used, because both components are decoupled in the functional embodiment.

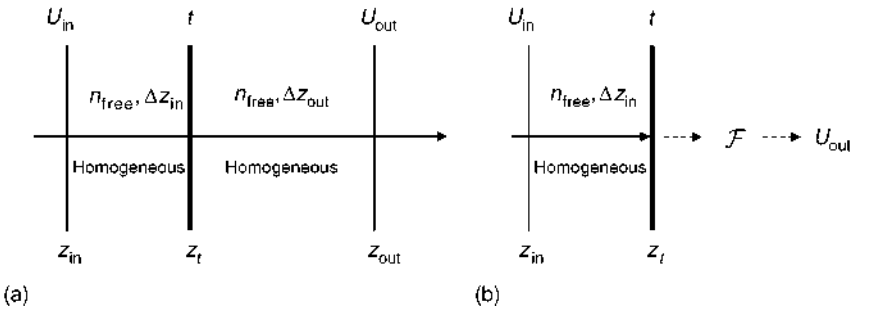

Figure 10 A basic system for light transformation can be described functionally by two propagations through homogeneous dielectrics with refractive index n_{free} and one phase-only transmission function. (b) If the second propagation is replaced by an operation in functional form, here the Fourier transform, the setup shown on the right side results.

If the output field is completely specified by the merit function, for instance: if the complex amplitude of a laser beam should be transformed into another one, the so-called amplitude matching strategy is a very powerful technique to solve the design problem stated by eqn [13]. In amplitude matching one determines the Distances Δz_{in} and Δz_{out} such that the inverse propagated U_{out} and the forward propagated U_{in} possess identical magnitudes. If such distances exist, the transformation can be realized using a phase-only transmission in the plane z_t. The phase simply follows from the ratio of the propagated fields. If no distances exist where amplitude matching occurs, the design problem has no solution and at least one more transmission function must be introduced.

As mentioned above, in numerous applications the output field is not specified completely. In particular, phase is often a free parameter. Then, an even simpler system approach than the one suggested by eqn [13], often leads to reasonable solution. Instead of $\mathcal{P}(n_{\text{free}}, \Delta z_{\text{out}})$, a Fourier transform is assumed, then if,

$$U_{\text{out}}(k_x, k_y) = S_{\text{functional}} U_{\text{in}}(x, y)$$
$$= \mathcal{F}\{\exp[\mathrm{i}\tau(x, y)]\mathcal{P}(n_{\text{free}}, \Delta z_{\text{in}}) U_{\text{in}}(x, y)\} \qquad [14]$$

The output field is expressed in k-coordinates as long as the system is discussed in functional form. For a functional design the distance Δz_{in} is of no concern, however it must be chosen together with the structure design as discussed below. Thus, in functional design, one concentrates only on the design of the transmission function. The iterative Fourier transform algorithm (IFTA), an extremely powerful algorithm, has been developed to accomplish this design. Its principle is illustrated in Figure 11. The algorithm sequentially applies constraints on the transmission and in the Fourier plane. Properly controlled, the algorithm converges to a Fourier pair for which the transmission plane constraint is satisfied completely and the Fourier domain constraint is satisfied as accurately as required by the merit function. The initial distribution is of enormous importance for a successful transmission design. Typically the initial distribution is defined in the Fourier plane as shown in Figure 11. A field $U_{\text{sig}}(k_x, k_y)$ is chosen, which satisfies the merit function. It is called the signal field. Because of design freedoms such as phase, typically there exists a set of signal fields F_{sig}. If phase freedom is available, any field with the desired amplitude and an arbitrary phase is an element of F_{sig} and can be used as initial distribution of the IFTA. Dependent on the application, the choice of the phase is of fundamental concern. In the following we assume phase freedom is available, but IFTA is also useful if this is not the case.

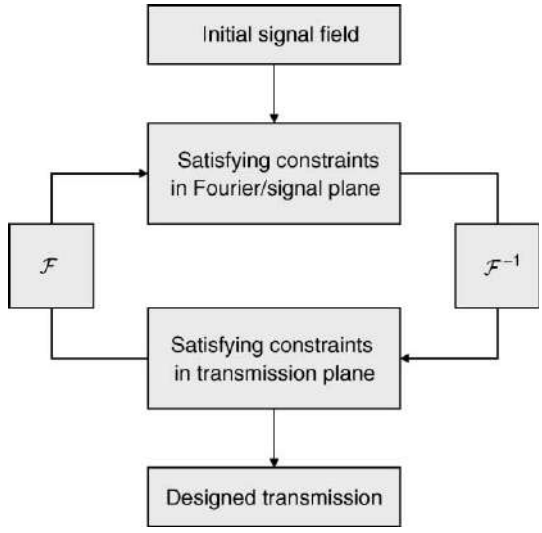

Figure 11 Illustration of the basic principle of the iterative Fourier transform algorithm (IFTA). By a sequential satisfaction of constraints in the signal and the transmission plane, the design in the functional embodiment can be accomplished. If both constraints are satisfied accurate enough, the IFTA is terminated with a final satisfaction of the transmission plane constraints. The depicted IFTA version can be generalized by replacing $\mathcal{F}$ by $\mathcal{P}$ or any S.

It is helpful to distinguish between two types of signal fields. If amplitudes for only a finite number of discrete k-values, (i.e., angles or directions), need to be determined, the transmission acts as a diffractive beamsplitter. In all other cases more general light transformations are concerned.

The constraints in the transmission plane follow from the structure design. In most cases the structure design is done by TEA according to eqn [12]. Then, phase-only transmissions are required. Typically the fabrication of the corresponding element profile demands a phase quantization in addition. Both constraints have been intensively investigated for the IFTA.

As mentioned above, practically signal fields are always restricted to one or more signal regions of finite extent. For example, a detector has finite extent. Outside of these regions the signal may be unrestricted or only partially restricted. This freedom in amplitude outside the signal regions is fundamental for a successful design of phase-only transmissions. It has been shown, that the only phase-only function whose Fourier transform exists over a finite area is a linear phase. The transform of all other phase-only functions contribute energy outside of the signal regions. In other words, the use of amplitude freedom is mandatory. Energy losses due to stray light outside

the signal region cannot be avoided. As a result, it is not possible to generate the desired response with 100% efficiency. For any given signal field U_{sig} an upper bound $\eta_{bound}(U_{sig})$ of the achievable efficiency $\eta(U_{sig})$ of light transformation with a phase-only transmission has been derived, that is $\eta(U_{sig}) \leq \eta_{bound}(U_{sig})$. The equal sign is only valid for the special case of a linear phase transmission, for which the signal field consists of one k-direction only. Then $\eta_{bound}(U_{sig}) = \eta(U_{sig}) = 1$ is obtained. After structure design for a linear phase-only transmission with TEA, a so-called blazed grating results. If TEA is a sufficient approximation one concludes that a blazed grating has a theoretical efficiency of 100%. For all other signals the theoretical efficiency is smaller than 100%. Moreover, the upper bound is a function of U_{sig} and thus also of its phase. Therefore, in a first step of a transmission design it is recommendable to perform a phase synthesis which maximizes the upper bound. This first design step delivers information about achieveable efficiency and a suitable initial signal field for the IFTA.

If fabrication requires phase quantization, amplitude freedom can be used to separate quantization noise from the signal regions. Unfortunately this reduces the achievable efficiency even more. If the phase to be quantized is homogenously distributed, the upper bound for a phase-only transmission quantized to Q equidistant phase levels is given by:

$$\eta_{bound}(U_{sig}, Q) = \mathrm{sinc}^2\left[\frac{1}{Q}\right]\eta_{bound}(U_{sig}) \qquad [15]$$

For the special case of linear phase, the well-known formula $\eta(Q) = \mathrm{sinc}^2[1/Q]$ follows. The values of $\mathrm{sinc}^2[1/Q]$ and therefore of quantized blazed gratings in the paraxial approximation are 0.405, 0.812, 0.950, 0.987, and 0.997 for $Q = 2$, 4, 8, 16, and 32, respectively. This result does not mean that, for example, a diffractive beamsplitter with a splitting ratio $1 : M_{beams}$ and 16 height levels will provide ~99% efficiency. According to eqn [15], the theoretical values are smaller and depend on M_{beams} because $\eta_{bound}(U_{sig})$ depends via U_{sig} on M_{beams}. If fabrication processes and the inherent imperfections are taken into consideration, the values decrease even more. As a rule of thumb, it is possible to claim that efficiencies larger than 90% are always challenging to obtain. With sophisticated design and fabrication techniques, efficiencies larger than 80% are reasonable.

Up to now, we have discussed only the importance of the signal field phase to efficiency. But the phase of the initial signal field has another extremely important effect on the result of the transmission design. From mapping-type considerations, a smooth initial signal field phase can be constructed in order to realize a light transformation. An example is illustrated in Figure 12. It shows the central part of the phase of the transmission function which transforms a Gaussian input beam into a flat-top line beam. Because of the smooth signal phase also the transmission phase is smooth and therefore unwrappable. This approach is typical for a so-called beam shaping design. It requires a well-defined input beam and high accuracy in adjustment of the beam shaping optics to obtain the output field with good quality. If, instead of a smooth initial phase, a random one is applied, the situation illustrated in Figure 13 (left) is achieved. Here the phase of the transmission looks random, which is typical for beamsplitter and diffuser. Because such phase functions possess numerous vortices, they do not allow an unwrapping. Thus, diffractive elements are the logical result of structure design. An example is shown in Figure 13 (right).

Figure 14 shows how light from a laser diode passes through the diffuser element of Figure 13 (right). The radiation is transformed into an intensity pattern that depicts the 'skyline' of Jena. Although globally the intensity possesses the requested shape, a closer look shows that the intensity distribution is speckled. However, in numerous applications this is not a problem so long as the size of the speckles is well adapted to the resolution of the detector.

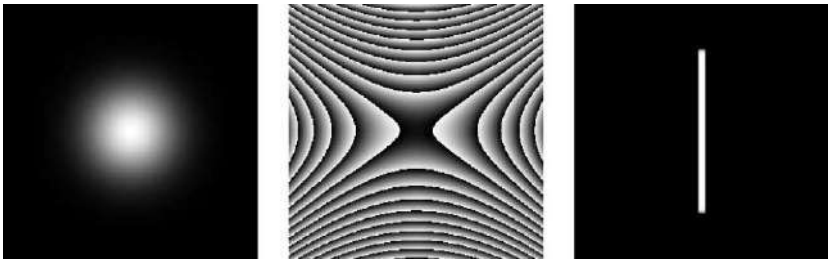

Figure 12 The input beam on the left propagates through the transmission function the central part of the phase of which is shown in the middle and in the signal plane the amplitude shown on the right results. This Gaussian-to-line transformation is a typical beam shaping design task.

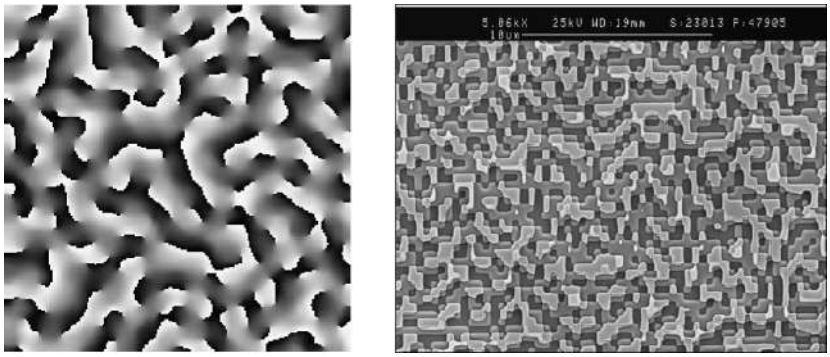

Figure 13 A typical phase distribution of a transmission (left) which results from design with IFTA for a random initial signal field phase. Numerous vortexes prevent its unwrapping. Therefore, structure design leads to diffractive elements. Then, a quantization of the phase is demanded in most cases. An example is shown on the right-handed side. It depicts a portion of a four level diffuser which results in the output field shown in **Figure 14**. The pixel size is 400 nm and the corresponding element was designed and fabricated in a cooperation of LightTrans GmbH and the Institute of Applied Physics at the University of Jena.

This technique is called diffuser-type light transformation. The appearance of the global intensity distribution is robust against variations in the input beam and thus is well-suited also for partial coherent illumination, for example radiation from an excimer laser and LED. In that case, even the appearance of speckles is reduced or vanishes completely. The mapping-type and the diffuser-type approach to choose the initial signal phase constitute two extreme cases of a great variety to control basic aspects of light transformation using different initial phases.

After choosing a suitable initial signal field $U_{\mathrm{sig}}(k_x, k_y)$ the IFTA typically performs well and a transmission function is obtained, which leads to $|U_{\mathrm{out}}(k_x, k_y)|^2 \propto |U_{\mathrm{sig}}(k_x, k_y)|^2$ in the signal region with minimum error. In the next step the structural design must be performed. The structure of the element is calculated via eqn [12].

Besides designing structure to produce the desired transmission function one needs to realize optically the Fourier transformation. Placing the observation plane in the far field behind the element is the most straightforward way to achieve this. However, field Fourier transformation can also be realized by lenses as shown in **Figure 15**. From simple geometric considerations using paraxial approximation one obtains:

$$\alpha = \frac{2\pi n_{\mathrm{free}}}{\lambda f_{\mathrm{e}}} \qquad [16]$$

Each system produces a Fourier transform that is scaled specially by α, where the output field is considered in a dielectric with refractive index n_{free}. The effective focal length f_{e} is equal to the focal length of the lens if the element is placed in front of the lens (see top of **Figure 15**), the distance of the element to the lens is of no concern for the resulting intensity,

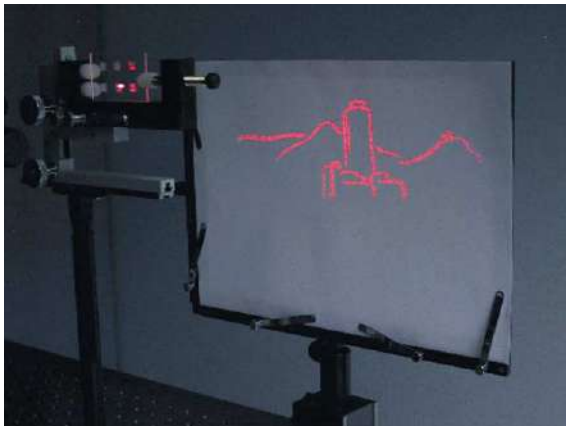

Figure 14 Optical demonstration of the use of a diffuser (see in Figure (right)) transforming a laser diode beam into an intensity displaying the 'skyline' of Jena. No additional optical element is needed! Stray light appears but is laterally separated from the desired signal and thus it is easy to be filtered.

however it is recommended to minimize the distance between element and lens for elements with deflection angles close to the paraxial limit. That avoids light losses due to truncation by the lens aperture. If the focal length itself is chosen as the distance of the element to the lens, a so-called $2f$-setup results, which generate, in addition to the correct magnitude, the correct phase of the Fourier transformation in the focal plane. If an alternative distance to the lens is used, an additional quadratic phase occurs in the focal plane. If the element is placed behind the lens (see middle row of **Figure 15**), f_{e} is the distance between element and focal plane. This setup provides some freedom in adjusting the output scaling. On the other hand, it requires an accurate adjustment of position to obtain a specific scaling. This is not required for the first setup. In order to obtain a specific

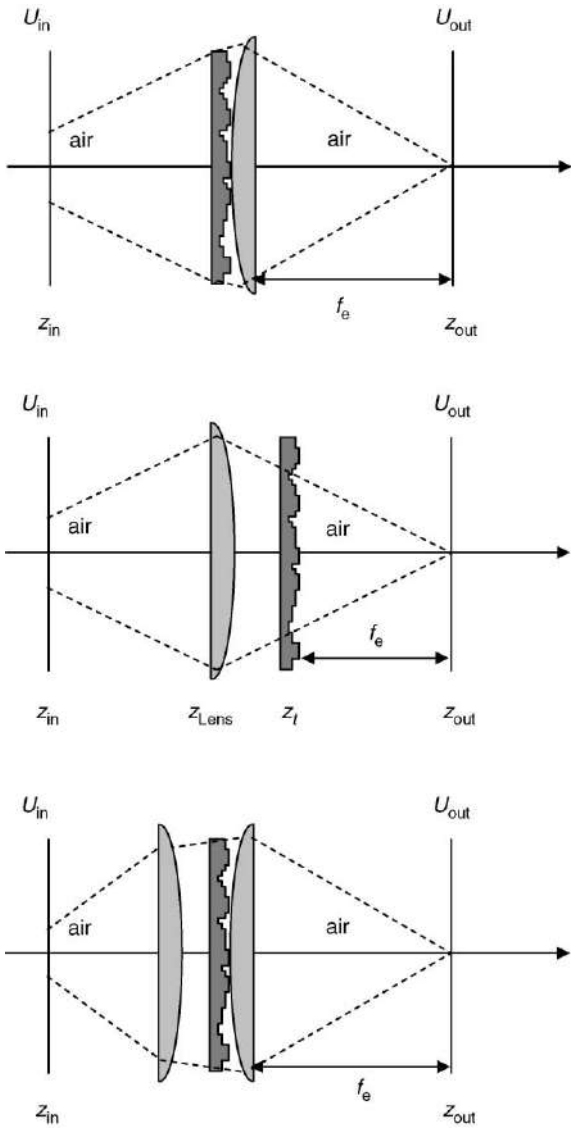

Figure 15 Three setups to realize a Fourier transformation for a beam shaping or diffuser/splitter element, which deflects the light only in a paraxial way. In all cases the output is obtained in the focal plane of the lens.

scaling of the output field $U_{out}(\alpha x, \alpha y)$, the effective focal length and the size of $U_{sig}(k_x, k_y)$ in the k-space must be chosen properly according to eqn [16]. In addition to the size of the output field, its resolution is also of fundamental importance. According to eqn [14] and the convolution theorem the output field is given by

$$U_{out}(k_x, k_y) = \mathcal{F}\ \exp[i\tau(x, y)] * \mathcal{F}\mathcal{P}(\Delta z_{in})U_{in}(x, y) \qquad [17]$$

where $*$ denotes the convolution. For a constant transmission function one would obtain the focus of the input beam. In functional embodiment that is $\mathcal{F}\mathcal{P}(\Delta z_{in})U_{in}(x, y)$ only. In the structural form, the focus of the beam must be calculated by propagating the input field through the system without the transmission function. The Fourier transformed transmission function, that is $\mathcal{F}\ \exp[i\tau(x, y)]$, is convolved with this beam focus in order to obtain the resulting output field. In beamsplitting that means a replication of the beam. In general the output field can be understood as being composed of numerous replicated beam foci which are overlapping. That makes plausible that the size of the focus of the input beam limits the resolution of the whole output field. The size of the focus is determined by the basic beam focusing lens system and not by the diffractive or refractive element. In conclusion, a highly resolved output field requires a focusing optic with high numerical aperture. For examples, it is not possible to obtain a flat-top beam profile from a Gaussian beam where focus has the same size as the flat-top should have. The steeper the sidewalls of the flat-top profile, the smaller must be the focal spot without the beam shaping element. The choice of a suitable numerical aperture of the lens system must be done before the transmission design is performed because it also determines the extent of the transmission function. The distance Δz_{in} must be chosen in such a way that ensures that the propagated input field illuminates the whole aperture of the system.

If a high numerical aperture on the output side is required to achieve a high resolution, the setup in the middle of Figure 15 should be avoided and the first setup is preferable. Otherwise the element would be illuminated by a highly nonparaxial field which violates the structure design with TEA. With the same argument, but now on the input side, an input field with high divergence should be at least partly collimated to achieve a paraxial illumination of the element. Then, the setup at the bottom of Figure 15 results.

For nonparaxial light transformation, that means the output field is large and hence the transformation requires nonparaxial deflection of the input field; the described transmission and structure design is still useful to obtain an initial design for a further optimization by wave-optical modeling techniques.

Though a Fourier approach to obtain general light transformations seems rather restricted, it covers a broad field of applications. The combination of suitable beam optics with the refractive free-form or diffractive element can be very simple (see optical setup of Figure 14) or very complex. That mainly depends on the resolution and size of the demanded output field as it is well known from imaging systems.

Each design should be finalized with a complete wave-optical analysis, typically including adjustment and fabrication error tolerancing, to achieve a full

understanding of the optical properties of the system. After this final investigation it is ready for fabrication and application providing a further innovation through optics and photonics on the base of wave-optical engineering.

See also

Fourier Optics.

Further Reading

Aagedal H, Schmid M, Beth T, Teiwes S and Wyrowski F (1996) Theory of speckles in diffractive optics and its application to beam shaping. *Journal of Modern Optics* 43(7): 1409–1421.

Aagedal H, Schmid M, Egner S, Müller-Quade J, Beth T and Wyrowski F (1997) Analytical beam shaping with application to laser-diode arrays. *Journal of the Optical Society of America A* 14(7): 1549–1553.

Aagedal H, Wyrowski F and Schmid M (1997) Paraxial beamsplitting and shaping. In: Turunen J and Wyrowski F (eds) *Diffractive Optics for Industrial and Commercial Applications*, chapter 6, pp. 165–188. Berlin: Akademie Verlag.

Born M and Wolf E (1980) *Principles of Optics*, 6th edn. New York: Pergamon Press.

Bryngdahl O (1974) Optical map transformations. *Optics Communications* 10(2): 164–168.

Bühling S, Schimmel H and Wyrowski F (2003) Wave-optical engineering with virtuallab. *Proceedings SPIE*, 5182: 24–33.

Bühling S, Schimmel H and Wyrowski F (2003) Challenges in industrial applications of diffractive beam splitting. *Proceedings SPIE* 5182: 123–131.

Collier RJ, Burckhardt CB and Lin LH (1971) *Optical Holography*. London: Academic Press.

Fienup JR (1980) Iterative method applied to image reconstruction and to computer-generated holograms. *Optical Engineering* 19(3): 297–305.

Gerchberg RW and Saxton WO (1972) A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* 35(2): 237–246.

Goodmann JW (1968) *Introduction to Fourier Optics*. New York: McGraw-Hill.

Goodman JW and Silvestri AM (1970) Some effects of Fourier-domain phase quantization. *IBM Journal of Reserved Development* 14: 478–484.

Krackhardt U, Mait JN and Streibl N (1992) Upper bound on the diffraction efficiency of phase-only fanout elements. *Applied Optics* 31(1): 27–37.

Lajunen H, Tervo J, Turunen J, Vallius T and Wyrowski F (2003) Simulation of light propagation by local spherical interface approximation. *Applied Optics* 34: 6804–6810.

Mandel L and Wolf E (1995) *Optical Coherence and Quantum Optics*. Cambridge: Cambridge University Press.

Pfeil AV and Wyrowski F (2000) Wave-optical structure design with the local plane-interface approximation. *Journal of Modern Optics* 47(13): 2335–2350.

Pfeil AV, Wyrowski F, Drauschke A and Aagedal H (2000) Analysis of optical elements with the local plane-interface approximation. *Applied Optics* 39(19): 3304–3313.

Schimmel H, Bühling S and Wyrowski F (2003) Use of ray tracing in wave-optical engineering. *Proceedings SPIE* 5182: 6–14.

Schimmel H and Wyrowski F (2000) Amplitude matching strategy for wave-optical design of monofunctional systems. *Journal of Modern Optics* 47(13): 2295–2321.

Sergienko N, Stamnes JJ, Kettunen V, Kuittinen M, Turunen J, Vahimaa P and Friberg AT (1999) Comparison of electromagnetic and scalar methods for evaluation of diffractive lenses. *Journal of Modern Optics* 46(1): 65–82.

Smith WJ (1997) *Practical Optical System Layout*. New York: McGraw-Hill.

Smith WJ (2000) *Modern Optical Engineering*, 3rd edn. Bellingham: SPIE Press.

Turunen J (1997) Diffraction theory of microrelief gratings. In: Herzig HP (ed.) *Micro-Optics Elements, Systems and Applications*, chapter 2, pp. 31–52. London: Taylor & Francis.

Turunen J and Wyrowski F (1997) *Diffractive Optics for Industrial and Commercial Applications*. Berlin: Akademie Verlag.

Vallius T, Jefimovs K, Kettunen V, Kuittinen M, Laakkonen P and Turunen J (2001) Design of non-paraxial array illuminators by step-transition perturbation approach. *Journal of Modern Optics* 48(12): 1869–1879.

Wyrowski F (1990) Diffractive optical elements: Iterative calculation of quantized, blazed phase structures. *Journal of the Optical Society of America A* 7: 961–969.

Wyrowski F (1991) Digital phase encoded inverse filter for optical pattern recognition. *Applied Optics* 30: 4650–4657.

Wyrowski F (1991) Considerations on convolution and phase factors. *Optics Communications* 81(6): 353–358.

Wyrowski F (1991) Upper bound of the diffraction efficiency of diffractive phase elements. *Optics Letters* 16(24): 1915–1917.

Wyrowski F (1993) Design theory of diffraction elements in the paraxial domain. *Journal of the Optical Society of America A* 10: 1553–1561.

Wyrowski F (2003) Comments on wave-optical engineering. *Proceedings SPIE* 5182: 1–5.

Wyrowski F and Turunen J (2002) Wave-optical engineering. In: Guenther AH (ed.) *International Trends in Applied Optics*, chapter 21, pp. 471–496. Bellingham: SPIE Press.

DISPERSION MANAGEMENT

A E Willner, Y-W Song, J Mcgeehan and Z Pan, University of Southern California, Los Angeles, CA, USA

B Hoanca, University of Alaska Anchorage, Anchorage, AK, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical communications systems have grown explosively in terms of the capacity that can be transmitted over a single optical fiber. This trend has been fueled by two complementary techniques, those being the increase in data-rate-per-channel coupled with the increase in the total number of parallel wavelength channels. However, there are many considerations as to the total number of wavelength-division-multiplexed (WDM) channels that can be accommodated in a system, including cost, information spectral efficiency, nonlinear effects, and component wavelength selectivity.

Dispersion is one of the critical roadblocks to increasing the transmission capacity of optical fiber. The dispersive effect in an optical fiber has several ingredients, including intermodal dispersion in a multimode fiber, waveguide dispersion, material dispersion, and chromatic dispersion. In particular, chromatic dispersion is one of the critical effects in a single mode fiber (SMF), resulting in a temporal spreading of an optical bit as it propagates along the fiber. At data rates ≤ 2.5 Gbit/s, the effects of chromatic dispersion are not particularly troublesome. For data rates ≥ 10 Gbit/s, however, transmission can be tricky and the chromatic dispersion-induced degrading effects must be dealt with in some way, perhaps by compensation. Furthermore, the effects of chromatic dispersion rise quite rapidly as the bit rate increases – when the bit rate increases by a factor of four, the effects of chromatic dispersion increase by a factor of 16! This article will only deal with the management of chromatic dispersion in single mode fiber.

One of the critical limitations of optical fiber communications comes from chromatic dispersion, which results in a pulse broadening as it propagates along the fiber. This occurs as photons of different frequencies (created by the spreading effect of data modulation) travel at different speeds, due to the frequency-dependent refractive index of the fiber core. Compounding the problems cause by chromatic dispersion is the fact that as bit rates rise, chromatic dispersion effects rise quadratically with respect to the increase in the bit rate. One can eliminate these effects using fiber with zero chromatic dispersion, known as dispersion shifted fiber (DSF). However, with zero dispersion, all channels in a wavelength-division-multiplexed (WDM) system travel at the same speed, in-phase, and a number of deleterious nonlinear effects such as cross-phase modulation (XPM) and four-wave mixing (FWM) result. Thus, in WDM systems, some amount of chromatic dispersion is necessary to keep channels out-of-phase, and as such chromatic dispersion compensation is required. Any real fiber link may also suffer from 'dispersion slope' effects, in which a slightly different dispersion value is produced in each WDM channel. This means that while one may be able to compensate one channel exactly, other channels may progressively accumulate increasing amounts of dispersion, which can severely limit the ultimate length of the optical link and the wavelength range that can be used in a WDM system. This article will address the concepts of chromatic dispersion and dispersion slope management followed by some examples highlighting the need for tunability to enable robust optical WDM systems in dynamic environments. Some dispersion monitoring techniques are then discussed and examples given.

Chromatic Dispersion in Optical Fiber Communication Systems

In any medium (other than vacuum) and in any waveguide structure (other than ideal infinite free space), different electromagnetic frequencies travel at different speeds. This is the essence of chromatic dispersion. As the real fiber-optic world is rather distant from the ideal concepts of both vacuum and infinite free space, dispersion will always be a concern when one is dealing with the propagation of electromagnetic radiation through fiber. The velocity in fiber of a single monochromatic wavelength is constant. However, data modulation causes a broadening of the spectrum of even the most monochromatic laser pulse. Thus, all modulated data have a nonzero spectral width which spans several wavelengths, and the different spectral components of modulated data travel at different speeds. In particular, for digital data intensity modulated on an optical carrier, chromatic

dispersion leads to pulse broadening – which in turn leads to chromatic dispersion limiting the maximum data rate that can be transmitted through optical fiber (see Figure 1).

Considering that the chromatic dispersion in optical fibers is due to the frequency-dependent nature of the propagation characteristics, for both the material (the refractive index of glass) and the waveguide structure, the speed of light of a particular wavelength λ will be expressed as follows, using a Taylor series expansion of the value of the refractive index as a function of the wavelength:

$$v(\lambda) = \frac{c_0}{n(\lambda)} = \frac{c_0}{n_0(\lambda_0) + \frac{\partial n}{\partial \lambda}\delta\lambda + \frac{\partial^2 n}{\partial \lambda^2}(\delta\lambda)^2} \qquad [1]$$

Here, c_0 is the speed of light in vacuum, λ_0 is a reference wavelength, and the terms in $\partial n/\partial\lambda$ and $\partial^2 n/\partial\lambda^2$ are associated with the chromatic dispersion and the dispersion slope (i.e., the variation of the chromatic dispersion with wavelength), respectively. Transmission fiber has positive dispersion, i.e. longer wavelengths result in longer propagation delays. The units of chromatic dispersion are picoseconds per nanometer per kilometer, meaning that shorter time pulses, wider frequency spread due to data modulation, and longer fiber lengths will each contribute linearly to temporal dispersion. Higher data rates inherently have both shorter pulses and wider frequency spreads. Therefore, as network speed increases, the impact of chromatic dispersion rises precipitously as the square of the increase in data rate. The quadratic increase with the data rate is a result of two effects, each with a linear contribution. On one hand, a doubling of the data rate makes the spectrum twice as wide, doubling the effect of dispersion. On the other hand, the same doubling of the data rate makes the data pulses only half as long (hence twice as sensitive to dispersion). The combination of a wider signal spectrum and a shorter pulse width is what leads to the overall quadratic impact. Moreover, the data modulation format used can significantly affect the sensitivity of a system to chromatic dispersion. For example, the common nonreturn-to-zero (NRZ) data format, in which the optical power stays high throughout the entire time slot of a '1' bit, is more robust to chromatic dispersion than is the return-to-zero (RZ) format, in which the optical power stays high in only part of the time slot of a '1' bit. This difference is due to the fact that RZ data have a much wider channel frequency spectrum compared to NRZ data, thus incurring more chromatic dispersion. However, in a real WDM system, the RZ format increases the maximum allowable transmission distance by virtue of its reduced duty cycle (compared to the NRZ format), making it less susceptible to fiber nonlinearities as can be seen in Figure 2.

A rule for the maximum distance over which data can be transmitted is to consider a broadening of the pulse equal to the bit period. For a bit period B, a dispersion value D and a spectral width $\Delta\lambda$, the dispersion-limited distance is given by

$$L_D = \frac{1}{D \cdot B \cdot \Delta\lambda} = \frac{1}{D \cdot B \cdot (cB)} \propto \frac{1}{B^2} \qquad [2]$$

(see Figure 3). For example, for single mode fiber, $D = 17$ ps/nm/km, so for 10 Gbit/s data the distance is $L_D = 52$ km. In fact, a more exact calculation

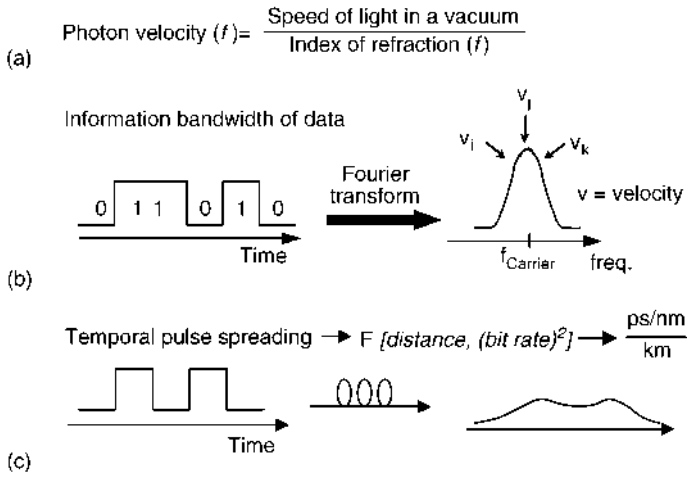

Figure 1 The origin of chromatic dispersion in data transmission. (a) Chromatic dispersion is caused by the frequency-dependent refractive index in fiber. (b) The nonzero spectral width due to data modulation. (c) Dispersion leads to pulse broadening, proportional to the transmission distance and the data rate.

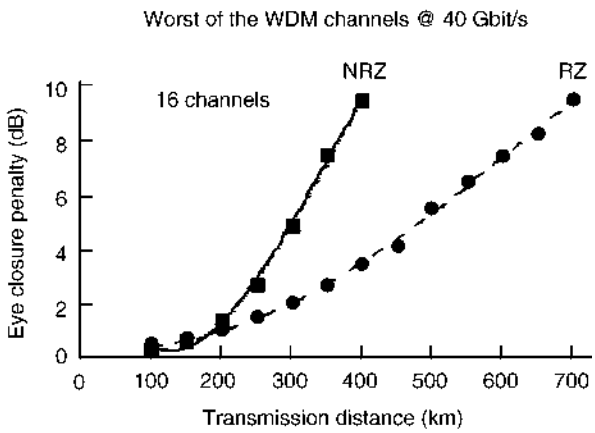

Figure 2 Performances of RZ and NRZ formats in a real fiber transmission link. (Reproduced with permission from Hayee I and Willner AE (1999) NRZ versus RZ in 10–40-Gb/s dispersion-managed WDM transmission systems. *IEEE Photon. Tech. Lett.* 11(8): 991–993.)

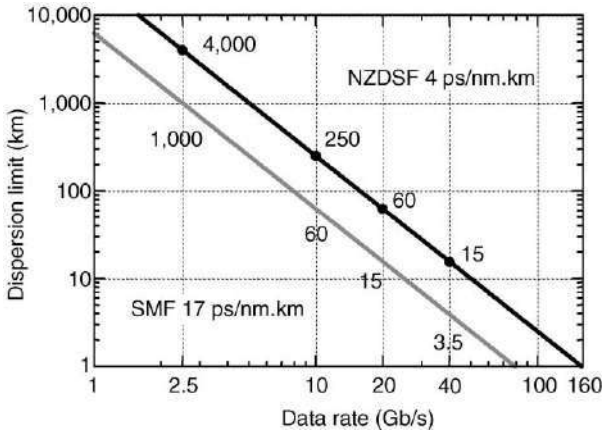

Figure 3 Transmission distance limitations due to uncompensated dispersion in SMF as a function of data rate for intensity modulated optical signals. (Reproduced with permission from Garrett LD (2001) Invited Short Course, *Optical Fiber Communication Conference*.)

shows that for 60 km, the dispersion induced power penalty is less than 1 dB (see Figure 4). The power penalty for uncompensated dispersion rises exponentially with transmission distance, and thus to maintain good signal quality, dispersion compensation is required.

Chromatic Dispersion Management

Optical Nonlinearities as Factors to be Considered in Dispersion Compensation

Even though it is possible to manufacture fiber with zero dispersion, it is not practical to use such fiber for WDM transmission, due to large penalties induced by fiber nonlinearities. Most nonlinear effects originate from the nonlinear refractive index of fiber, which is

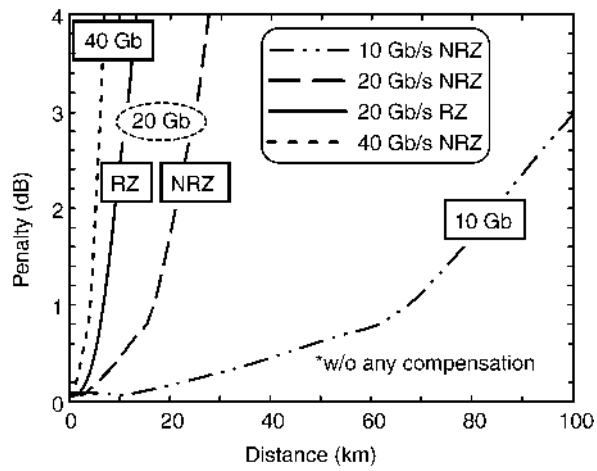

Figure 4 Power penalties due to uncompensated dispersion in single mode fiber (SMF) as a function of distance and data rate. (Reproduced with permission from Garrett LD (2001) Invited Short Course, *Optical Fiber Communication Conference*.)

not only dependent on the frequency of light but also on the intensity (optical power), and is related to the optical power as:

$$\bar{n}(f, P) = n(f) + n_2 \frac{P}{A_{\text{eff}}} \qquad [3]$$

where $n(f)$ is the linear part of the refractive index, P is the optical power inside the fiber, and n_2 is the nonlinear-index coefficient for silica fibers. The typical value of n_2 is 2.6×10^{-20} m^2/W. This number takes into account the averaging of the polarization states of the light as it travels in the fiber. The intensity dependence of the refractive index gives rise to three major nonlinear effects.

Self-phase modulation (SPM)

A million photons 'see' a different glass than does a single photon, and a photon traveling along with many other photons will slow down. SPM occurs because of the varying intensity profile of an optical pulse on a single WDM channel. This intensity profile causes a refractive index profile and, thus, a photon speed differential. The resulting phase change for light propagating in an optical fiber is expressed as:

$$\Phi_{\text{NL}} = \gamma P L_{\text{eff}} \qquad [4]$$

where the quantities γ and L_{eff} are defined as:

$$\gamma = \frac{2\pi n_2}{\lambda A_{\text{eff}}} \quad \text{and} \quad L_{\text{eff}} = \frac{1 - e^{-\alpha L}}{\alpha} \qquad [5]$$

where A_{eff} is the effective mode area of the fiber and α is the fiber attenuation loss. L_{eff} is the effective nonlinear length of the fiber that accounts for fiber loss, and γ is the nonlinear coefficient measured

in rad/km/W. A typical range of values for γ is between 10–30 rad/km/W. Although the nonlinear coefficient is small, the long transmission lengths and high optical powers, that have been made possible by the use of optical amplifiers, can cause a large enough nonlinear phase change to play a significant role in state-of-the-art lightwave systems.

Cross-phase modulation (XPM)

When considering many WDM channels co-propagating in a fiber, photons from channels 2 through N can distort the index profile that is experienced by channel 1. The photons from the other channels 'chirp' the signal frequencies on channel 1, which will interact with fiber chromatic dispersion and cause temporal distortion. This effect is called cross-phase modulation. In a two-channel system, the frequency chirp in channel 1, due to power fluctuation within both channels, is given by

$$\Delta B = \frac{\mathrm{d}\Phi_{\mathrm{NL}}}{\mathrm{d}t} = \gamma L_{\mathrm{eff}} \frac{\mathrm{d}P_1}{\mathrm{d}t} + 2\gamma L_{\mathrm{eff}} \frac{\mathrm{d}P_2}{\mathrm{d}t} \qquad [6]$$

where, $\mathrm{d}P_1/\mathrm{d}t$ and $\mathrm{d}P_2/\mathrm{d}t$ are the time derivatives of the pulse powers of channels 1 and 2, respectively. The first term on the right-hand side of the above equation is due to SPM, and the second term is due to XPM. Note that the XPM-induced chirp term is double that of the SPM-induced chirp term. As such, XPM can impose a much greater limitation on WDM systems than can SPM, especially in systems with many WDM channels.

Four-wave-mixing (FWM)

The optical intensity propagating through the fiber is related to the electric field intensity squared. In a WDM system, the total electric field is the sum of the electric fields of each individual channel. When squaring the sum of different fields, products emerge that are beat terms at various sum and difference frequencies to the original signals. Figure 5 depicts that if a WDM channel exists at one of the four-wave-mixing beat-term frequencies, then the beat term will interfere coherently with this other WDM channel and potentially destroy the data.

Dispersion Maps

While zero-dispersion fiber is not a good idea, a large value of the accumulated dispersion at the end of a fiber link is also undesirable. An ideal solution is to have a 'dispersion map,' alternating sections of positive and negative dispersion as can be seen in Figure 6. This is a very powerful concept: at each point along the fiber the dispersion has some nonzero value, eliminating FWM and XPM, but the total dispersion at the end of the fiber link is zero, so that no pulse broadening is induced (Table 1). The most advanced systems require periodic dispersion compensation, as well as pre- and post-compensation (before and after the transmission fiber).

The addition of negative dispersion to a standard fiber link has been traditionally known as 'dispersion compensation,' however, the term 'dispersion management' is more appropriate. SMF has positive dispersion, but some new varieties of nonzero dispersion-shifted fiber (NZDSF) come in both positive and negative dispersion varieties. Some examples are shown in Figure 7. Reverse dispersion fiber is also now available, with a large dispersion

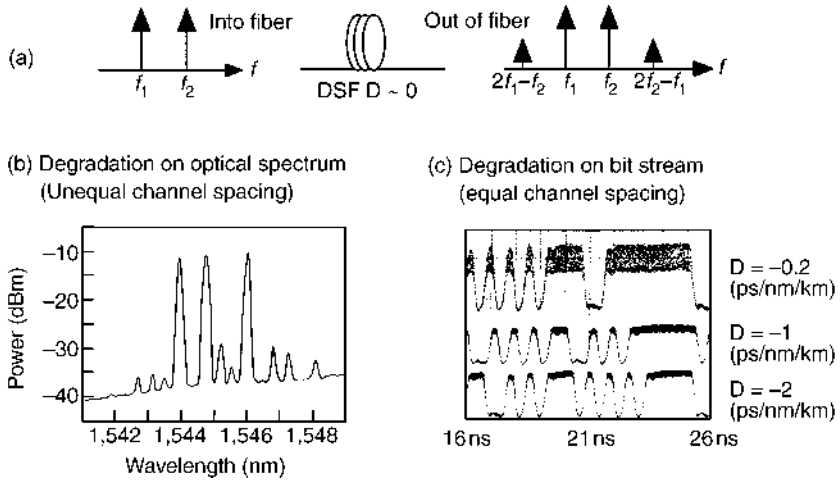

Figure 5 (a) and (b) FWM induces new spectral components via nonlinear mixing of two wavelength signals. (c) The signal degradation due to FWM products falling on a third data channel can be reduced by even small amounts of dispersion. (Reproduced with permission from Tkach RW, Chraplyvy AR, Forghieri F, Gnauck AH and Derosier RM (1995) Four-photon mixing and high-speed WDM systems. *Journal of Photon Technology* 13(5): 841–849.)

comparable to that of SMF, but with the opposite sign. When such flexibility is available in choosing both the magnitude and sign of the dispersion of the fiber in a link, dispersion-managed systems can be fully optimized to the desired dispersion map using a combination of fiber and dispersion compensation devices (see Figure 8).

Dispersion is a linear process, so first-order dispersion maps can be understood as linear systems. However, the effects of nonlinearities cannot be ignored, especially in WDM systems, with many tens of channels, where the launch power may be very high. In particular, in systems deploying dispersion compensating fiber (DCF), the large nonlinear coefficient of the DCF can dramatically affect the dispersion map.

Corrections to Linear Dispersion Maps

Chromatic dispersion is a necessity in WDM systems, to minimize the effects of fiber nonlinearities. A chromatic dispersion value as small as a few ps/nm/km is usually sufficient to make XPM and FWM negligible. To mitigate the effects of nonlinearities but maintain small amounts of chromatic dispersion, NZDSF is commercially available. Due to these nonlinear effects, chromatic dispersion must be managed, rather than eliminated.

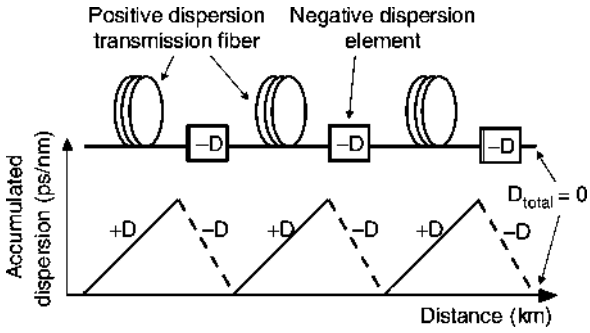

Figure 6 Dispersion map of a basic dispersion managed system. Positive dispersion transmission fiber alternates with negative dispersion compensation elements such that the total dispersion is zero end-to-end.

If a dispersion-management system was perfectly linear, it would be irrelevant whether the dispersion along a path is small or large, as long as the overall dispersion is compensated to zero (end to end). Thus, in a linear system the performance should be similar, regardless of whether the transmission fiber is SMF, and dispersion compensation modules are deployed every 60 km, or the transmission fiber is NZDSF (with approximately a quarter of the dispersion value of SMF) and dispersion compensation modules are deployed every 240 km. In real life, optical nonlinearities are very important, and recent results seem to favor the use of large, SMF-like, dispersion values in the transmission path and correspondingly high dispersion compensation devices. A recent study of performance versus channel spacing showed that the capacity of SMF could be more than four times that of NZDSF. This is because the nonlinear coefficients are much higher in NZDSF than in SMF, and for dense WDM the channel interactions become a limiting factor. A critical conclusion is that not all dispersion compensation maps are created equal: a simple calculation of the dispersion compensation, to cancel the overall dispersion value, does not lead to optimal dispersion map designs.

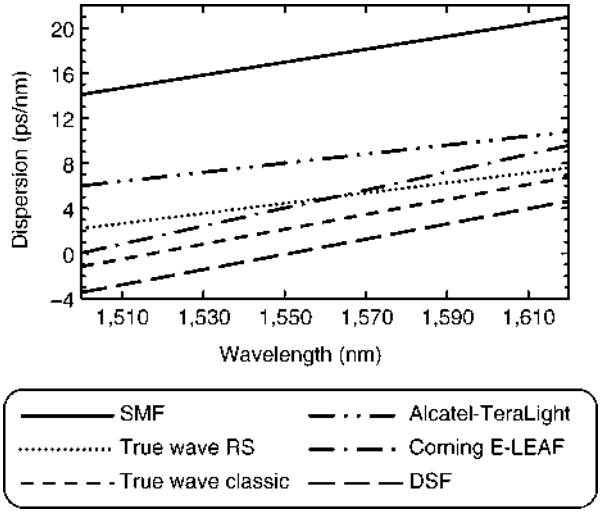

Figure 7 Chromatic dispersion characteristics of various commercially available types of transmission fiber.

Table 1 Commercially available fibers and their characteristics

	Dispersion @ 1,550 nm [ps/nm/km]	*Dispersion slope [ps/nm^2/km]*	*Attenuation [dB]*	*Mode field diameter [μm]*	*PMD [ps/km$^{0.5}$]*
SMF	18	0.08	≤0.25	11	≤0.5
Tera Light	8.0	0.058	≤0.2	9.2	≤0.04
TW-RS	4.4	0.043	≤0.25	8.4	≤0.03
LEAF	4.0	0.085	≤0.25	9.6	≤0.08
Standard DCF	−90	−0.22	≤0.5	5.2	≤0.08

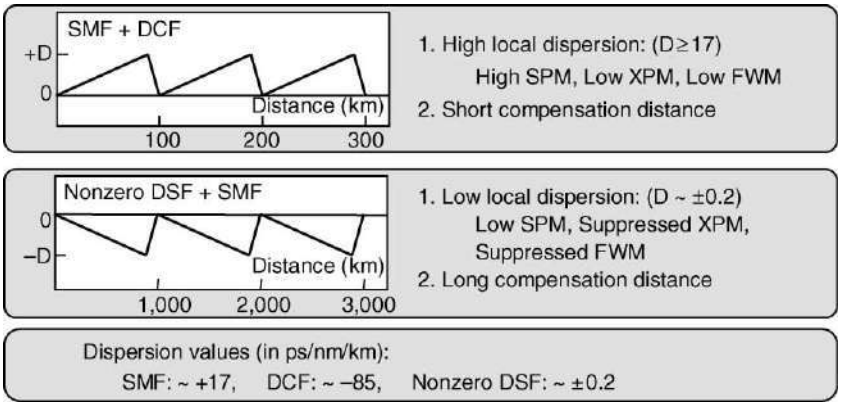

Figure 8 Various dispersion maps for SMF-DCF and NZDSF-SMF.

Additionally, several solutions have been shown to be either resistant to dispersion, or have been shown to rely on dispersion itself for transmission. Such solutions include chirped pulses (where prechirping emphasizes the spectrum of the pulses so that dispersion does not broaden them too much), dispersion assisted transmission (where an initial phase modulation tailored to the transmission distance leads to full-scale amplitude modulation at the receiver end due to the dispersion), and various modulation formats robust to chromatic dispersion and nonlinearities.

Dispersion Management Solutions

Fixed Dispersion Compensation

From a systems point of view, there are several requirements for a dispersion compensating module: low loss, low optical nonlinearity, broadband (or multichannel) operation, small footprint, low weight, low power consumption, and clearly low cost. It is unfortunate that the first dispersion compensation modules, based on DCF only, met two of these requirements: broadband operation and low power consumption. On the other hand, several solutions have emerged that can complement or even replace these first-generation compensators.

Dispersion compensating fiber (DCF)

One of the first dispersion compensation techniques was to deploy specially designed sections of fiber with negative chromatic dispersion. The technology for DCF emerged in the 1980s and has developed dramatically since the advent of optical amplifiers in 1990. DCF is the most widely deployed dispersion compensator, providing broadband operation and stable dispersion characteristics, and the lack of a dynamic, tunable DCF solution has not reduced its popularity.

As can be seen in Figure 9, the core of the average dispersion compensating fiber is much smaller than that of standard SMF, and beams with longer wavelengths experience relatively large changes in mode size (due to the waveguide structure) leading to greater propagation through the cladding of the fiber, where the speed of light is greater than that of the core. This leads to a large negative dispersion value. Additional cladding layers can lead to improved DCF designs that can include negative dispersion slope to counteract the positive dispersion slope of standard SMF.

In spite of its many advantages, DCF has a number of drawbacks. First, it is limited to a fixed compensation value. In addition, DCF has a weakly guiding structure and has a much smaller core cross-section, 19 μm^2, compared to the 85 μm^2 of SMF. This leads to higher nonlinearity, higher splice losses, as well as higher bending losses. Last, the length of DCF required to compensate for SMF dispersion is rather long, about one-fifth of the length of the transmission fiber for which it is compensating. Thus DCF modules induce loss, and are relatively bulky and heavy. The bulk is partly due to the mass of fiber, but also due to the resin used to hold the fiber securely in place. One other contribution to the size of the module is the higher bend loss associated with the refractive index profile of DCF; this limits the radius of the DCF loop to 6–8 inches, compared to the minimum bend radius of 2 inches for SMF.

Traditionally, DCF-based dispersion compensation modules are usually located at amplifier sites. This serves several purposes. First, amplifier sites offer relatively easy access to the fiber, without requiring any digging or unbraiding of the cable. Second, DCF has high loss (usually at least double that of standard SMF), so a gain stage is required

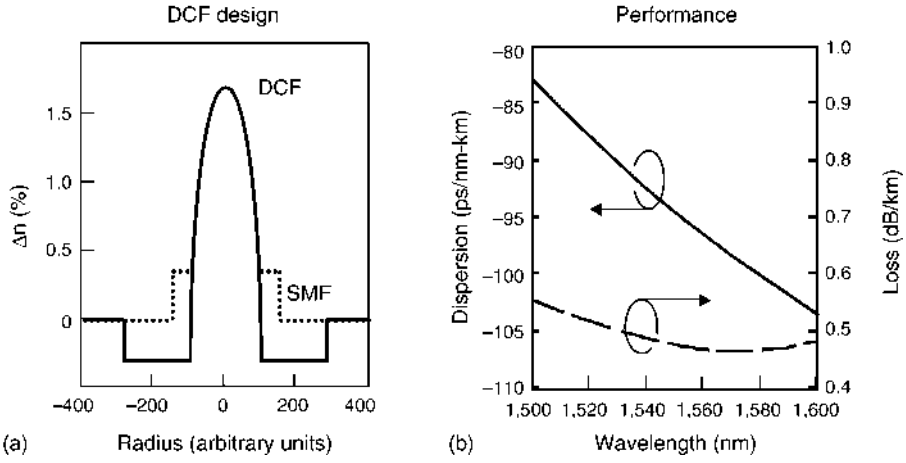

Figure 9 Typical DCF (a) refractive index profile and (b) dispersion and loss as a function of wavelength. Δn is defined as refractive index variation relative to the cladding.

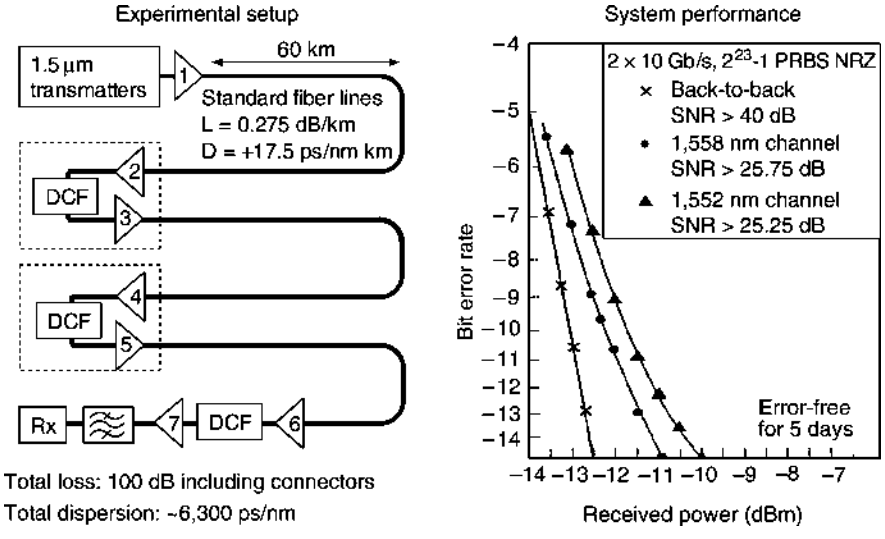

Figure 10 System demonstration of dispersion compensation using DCF. (Reproduced with permission from Park YK, Yeates PD, Delavaux J-MP, *et al.* (1995) A field demonstration of 20-Gb/s capacity transmission over 360 km of installed standard (non-DSF) fiber. *Photon. Technol. Lett.* 7(7): 816–818.)

before the DCF module to avoid excessively low signal levels. DCF has a cross-section four times smaller then SMF, hence a higher nonlinearity, which limits the maximum launch power into a DCF module. The compromise is to place the DCF in the mid-section of a two-section EDFA. This way, the first stage provides pre-DCF gain, but not to a power level that would generate excessive nonlinear effects in the DCF. The second stage amplifies the dispersion compensated signal to a power level suitable for transmission though the fiber link. This launch power level is typically much higher than could be transmitted through DCF without generating large nonlinear effects. Many newer dispersion compensation devices have better performance than DCF, in particular lower loss and lower nonlinearities. For this reason, they may not have to be deployed at the mid-section of an amplifier. Figure 10 shows the real demonstration results using the DCF.

Chirped fiber Bragg gratings

Fiber Bragg gratings have emerged as major components for dispersion compensation because of their low loss, small footprint, and low optical nonlinearity. Bragg gratings are sections of single-mode fiber in which the refractive index of the core is modulated in a periodic fashion, as a function of the spatial coordinate along the length of the fiber. When the spatial periodicity of the modulation matches what is known as a Bragg condition with respect to the wavelength of light propagating through the grating,

the periodic structure acts like a mirror, reflecting the optical radiation that is traveling through the core of the fiber. An optical circulator is traditionally used to separate the reflected output beam from the input beam.

When the periodicity of the grating is varied along its length, the result is a chirped grating which can be used to compensate for chromatic dispersion. The chirp is understood as the rate of change of the spatial frequency as a function of position along the grating. In chirped gratings the Bragg matching condition for different wavelengths occurs at different positions along the grating length. Thus, the roundtrip delay of each wavelength can be tailored by designing the chirp profile appropriately. Figure 11 compares the chirped FBG with uniform FBG. In a data pulse that has been distorted by dispersion, different frequency components arrive with different amounts of relative delay. By tailoring the chirp profile such that the frequency components see a relative delay which is the inverse of the delay of the transmission fiber, the pulse can be compressed back. The dispersion of the grating is the slope of the time delay as a function of wavelength, which is related to the chirp.

The main drawback of Bragg gratings is that the amplitude profile and the phase profile as a function of wavelength have some amount of ripple. Ideally, the amplitude profile of the grating should have a flat (or rounded) top in the passband, and the phase profile should be linear (for linearly chirped gratings) or polynomial (for nonlinearly chirped gratings). The grating ripple is the deviation from the ideal profile shape. Considerable effort has been expended on reducing the ripple. While early gratings were plagued by more than 100 ps of ripple, published results have shown vast improvement to values close to ±3 ps.

Higher order mode dispersion compensation fiber

One of the challenges of designing standard DCF is that high negative dispersion is hard to achieve unless the cross-section of the fiber is small (which leads to high nonlinearity and high loss). One way to reduce both the loss and the nonlinearity is to use a higher-order mode (HOM) fiber (LP_{11} or LP_{02} near cutoff instead of the LP_{01} mode in the transmission fiber).

Such a device requires a good-quality mode converter between LP_{01} and LP_{02} to interface between the SMF and HOM fiber. HOM fiber has a dispersion per unit length greater than six times that of DCF. Thus, to compensate for a given transmission length in SMF, the length of HOM fiber required is only one sixth the length of DCF. Thus, even though losses and nonlinearity per unit length are larger for HOM fiber than for DCF, they are smaller overall, because of the shorter HOM fiber length. As an added bonus, the dispersion can be tuned slightly by changing the cutoff wavelength of LP_{02} (via temperature tuning). A soon to be released HOM fiber-based commercial dispersion compensation module is not tunable, but can fully compensate for dispersion slope.

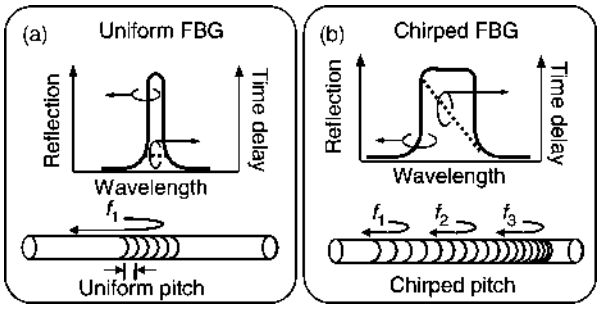

Figure 11 Uniform and chirped FBGs. (a) A grating with uniform pitch has a narrow reflection spectrum and a flat time delay as a function of wavelength. (b) A chirped FBG has a wider bandwidth, a varying time delay and a longer grating length.

Tunable Dispersion Compensation

The need for tunability

In a perfect world, all fiber links would have a known, discrete, and unchanging value of chromatic dispersion. Network operators would then deploy fixed dispersion compensators periodically along every fiber link to exactly match the fiber dispersion. Unfortunately, several vexing issues may necessitate that dispersion compensators are tunability, that they have the ability to adjust the amount of dispersion to match system requirements.

First, there is the most basic business issue of inventory management. Network operators typically do not know the exact length of a deployed fiber link nor its chromatic dispersion value. Moreover, fiber plants periodically undergo upgrades and maintenance, leaving new and nonexact lengths of fiber behind. Therefore, operators would need to keep in stock a large number of different compensator models, and even then the compensation would only be approximate. Second, we must consider the sheer difficulty of 40 Gbit/s signals. The tolerable threshold for accumulated dispersion for a 40 Gbit/s data channel is 16 times smaller than at 10 Gbit/s. If the compensation value does not exactly match the fiber to within a few percent of the required dispersion value, then the communication link will not work. Tunability is considered a key enabler for this bit rate (see Figures 12 and 13). Third, the accumulated dispersion changes slightly with temperature, which begins to be an issue for 40 Gbit/s systems and 10 Gbit/s ultra long-haul systems. In fiber, the zero-dispersion wavelength changes with temperature at a

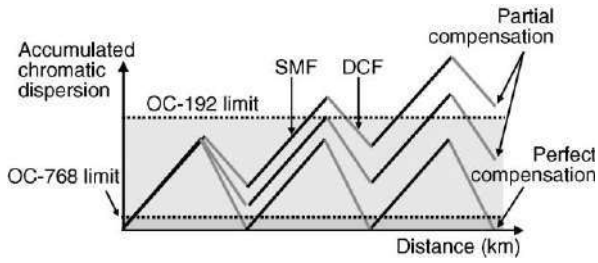

Figure 12 The need for tunability. The tolerance of OC-768 systems to chromatic dispersion is 16 times lower than that of OC-192 systems. Approximate compensation by fixed in-line dispersion compensators for a single channel may lead to rapid accumulation of unacceptable levels of residual chromatic dispersion.

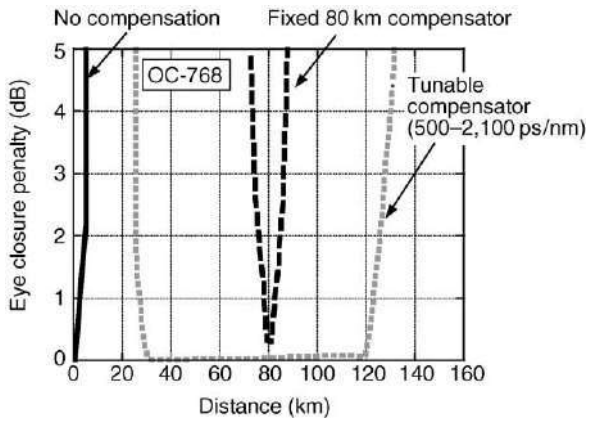

Figure 13 Tunable dispersion compensation at OC-768 (40 Gb/s) is essential for achieving a comfortable range of acceptable transmission distances (80 km for tunable, only ~4 km for fixed compensation).

typical rate of 0.03 nm/°C. It can been shown that a not-uncommon 50 °C variation along a 1,000 km 40 Gbit/s link can produce significant degradation (see Figure 14). Fourth, we are experiencing the dawn of reconfigurable optical networking. In such systems, the network path, and therefore the accumulated fiber dispersion, can change. It is important to note that even if the fiber spans are compensated span-by-span, the pervasive use of compensation at the transmitter and receiver suggests that optimization and tunability based on path will still be needed.

Other issues that increase the need for tunability include: (i) laser and (de)mux wavelength drifts for which a data channel no longer resides on the flat-top portion of a filter, thereby producing a chirp on the signal that interacts with the fiber's chromatic dispersion; (ii) changes in signal power that change both the link's nonlinearity and the optimal system dispersion map; and (iii) small differences that exist in transmitter-induced signal chirp.

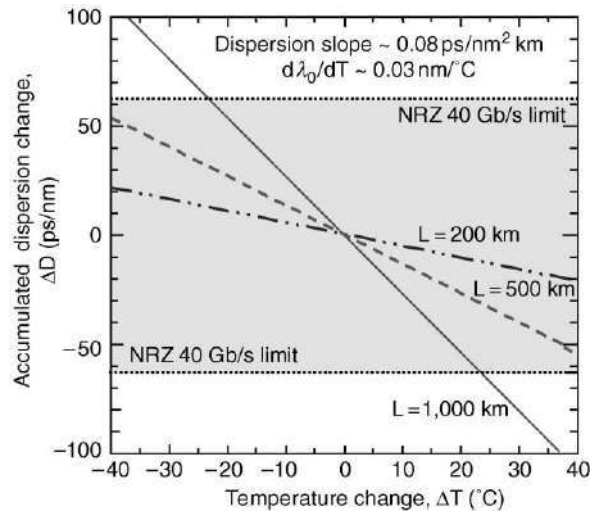

Figure 14 Accumulated dispersion changes as a function of the link length and temperature fluctuation along the fiber link.

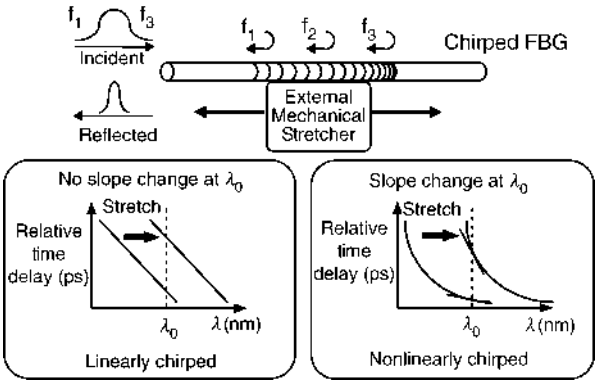

Figure 15 Tuning results for both linearly and nonlinearly chirped FBGs using uniform stretching elements. The slope of the dispersion curve at a given wavelength λ_0 is constant when the linearly chirped grating is stretched, but changes as the nonlinearly chirped grating is stretched.

Approaches to tunable dispersion compensation

A host of techniques for tunable dispersion compensation have been proposed in recent years. Some of these ideas are just interesting research ideas, but several have strong potential to become viable technologies.

Fiber gratings offer the inherent advantages of fiber compatibility, low loss, and low cost. If a FBG has a refractive-index periodicity that varies nonlinearly along the length of the fiber, it will produce a time delay that also varies nonlinearly with wavelength (see Figure 15). Herein lies the key to tunability. When a linearly chirped grating is stretched uniformly by a single mechanical element, the time delay curve is shifted towards longer wavelengths, but the slope of the ps-vs.-nm curve remains constant at all wavelengths within the passband. When a

nonlinearly-chirped grating is stretched, the time delay curve is shifted toward longer wavelengths, but the slope of the ps-vs.-nm curve at a specific channel wavelength changes continuously. Ultimately, tunable dispersion compensators should accommodate multichannel operation. Several WDM channels can be accommodated by a single chirped FBG in one of two ways: fabricating a much longer (i.e., meters-length) grating, or using a sampling function when writing the grating, thereby creating many replicas of transfer function of the FBG in the wavelength domain (see Figure 16).

One free-space-based tunable dispersion compensation device is the virtually imaged phased array (VIPA), based on the dispersion of a Fabry–Perot interferometer. The design requires several lenses, a movable mirror (for tunability), and a glass plate with a thin film layer of tapered reflectivity for good mode matching. Light incident on the glass plate undergoes several reflections inside the plate. As a result, the beam is imaged at several virtual locations, with a spatial distribution that is wavelength-dependent.

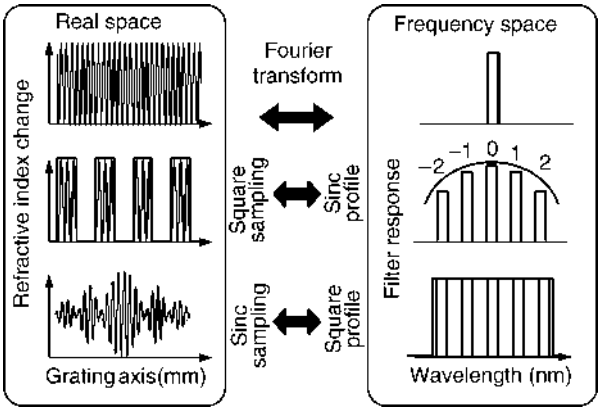

Figure 16 The concept of 'sampled' FBGs, where a superstructure is written on top of the grating that produces a Fourier transform in the frequency domain, leading to multiple grating passbands.

Several devices used for dispersion compensation can be integrated on a chip, using either an optical chip media (semiconductor-based laser or amplifier medium) or an electronic chip. One such technology is the micro-ring resonator, a device that, when used in a structure similar to that of an all-pass filter (see Figure 17), can be used for dispersion compensation on a chip-scale. Although these technologies are not yet ready for deployment as dispersion compensators, they have been used in other applications and have the potential to offer very high performance at low cost.

As the ultimate optical dispersion compensation devices, photonic bandgap fibers (holey fibers) are an interesting class in themselves (see Figure 18). These are fibers with a hollow structure, with holes engineered to achieve a particular functionality. Instead of being drawn from a solid preform, holey fibers are drawn from a group of capillary tubes fused together. This way, the dispersion, dispersion slope, the nonlinear coefficients could in principle all be precisely designed and controlled, up to very small or very large values, well outside the range of those of the solid fiber.

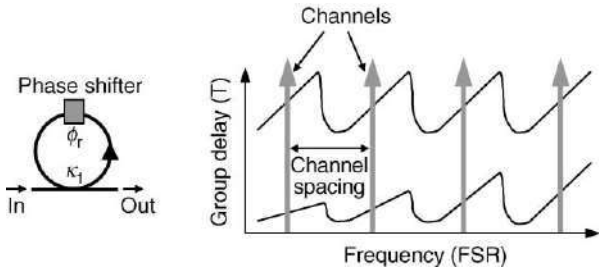

Figure 17 Architecture of an all-pass filter structure for chromatic dispersion and slope compensation. (Reproduced with permission from Madsen CK, Lenz G, Bruce AJ, *et al.* (1999) Integrated all-pass filters for tunable dispersion and dispersion slope compensation. *Photon. Technol. Lett.* 11(12): 1623–1625.)

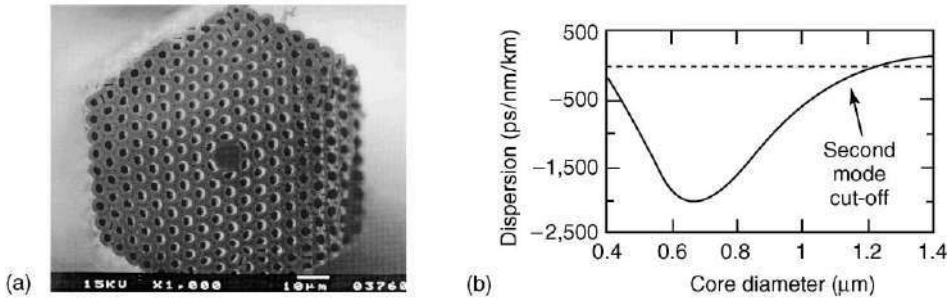

Figure 18 (a) SEM image of a photonic crystal fiber (holey fiber), and (b) net dispersion of the fiber at 1550 nm as a function of the core diameter. (Reproduced with permission from Birk TA, Mogihetsev D, Knight JC and Russell PSt.J (1999) Integrated all-pass filters for tunable dispersion and dispersion slope compensation. *Photon. Technol. Lett.* 11(6): 674–676.)

Dispersion Slope Mismatch

Transmission fiber, especially fiber with dispersion compensation built in, may suffer from a dispersion slope in which a slightly different dispersion value is produced for each WDM channel (see Figure 19). Even though the compensator would be able to cancel the dispersion of the fiber at the design wavelength, there will be residual dispersion left at the other wavelength channels unless the compensator can match the slope of the dispersion curve of the fiber as well. Some solutions for dispersion slope compensation are described in this section.

First, DCF, with negative dispersion slope, is a prime candidate for deployment as a dispersion slope compensator even though it cannot easily be made tunable. By designing the DCF with the same ratio of dispersion to dispersion slope as that of a real fiber link, new types of DCF can be used to compensate for both dispersion and dispersion slope, much like DCF is used for dispersion compensation today. DCF's popularity, wideband functionality, and stable dispersion characteristics make this a particularly attractive solution. Designing the DCF to match the dispersion characteristics of the transmission fiber is the critical engineering challenge in this slope compensation scheme. Second, third-order nonlinearly chirped FBG can act as a tunable dispersion slope compensator. A simple modification of the nonlinearly chirped FBG allows tuning of the compensated dispersion slope value via stretching the grating. The grating is prepared such that the time delay as a function of wavelength has a cubic profile that covers several WDM channels over a continuous bandwidth of many nanometers. Since the resulting dispersion curve is quadratic over the grating bandwidth, the dispersion slope experienced by the WDM channels can be tuned by stretching the grating using a single mechanical element (see Figure 20). Third, combining the VIPA, with either a 3D mirror or diffraction grating, can also provide tunable free-space dispersion slope compensation. Slope tuning is achieved by dynamically controlling the MEMS-based 3D mirror or the diffraction grating. Fourthly, using an FBG with many spaced thin-film heater sections can also enable tunable dispersion slope. Each heater can be individually electrically controlled, allowing the time delay profile of the grating to be dynamically tuned via changing the temperature along the length of the grating. Advanced applications of this technique can allow alterations to the entire time delay profile of the grating, providing a truly flexible dispersion and dispersion slope compensation mechanism.

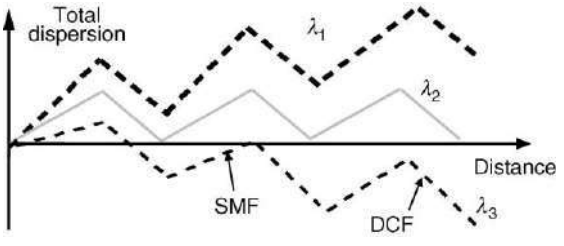

Figure 19 The chromatic dispersion slope mismatch caused by the different slopes of transmission fiber (SMF or NZDSF) and DCF.

Chromatic Dispersion Monitoring

Another important issue related to dispersion management is dispersion monitoring techniques. In a reconfigurable system, it is necessary to reconfigure

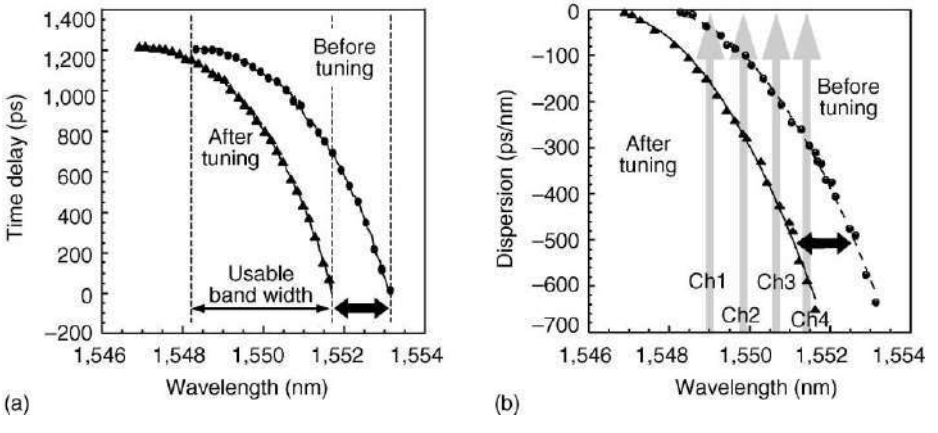

Figure 20 Tunable dispersion slope compensation using a third-order nonlinearly chirped FBG. (a) Cubic time delay curves of the grating, and (b) quadratic dispersion curves showing the change in dispersion in each channel before and after tuning. (Reproduced with permission from Song YW, Motoghian SMR, Starodubov D, *et al.* (2002) Tunable dispersion slope compensation for WDM systems using a non-channelized third-order-chirped FBG. *Optical Fiber Communication Conference* 2002. Paper ThAA4.)

any tunable chromatic dispersion compensation modules on the fly as the network changes. An in-line chromatic dispersion monitor can quickly measure the required dispersion compensation value while data are still being transmitted through the optical link. This is very different from the more traditional chromatic dispersion measurement techniques where dark fiber is used and the measurement is done off-line over many hours (or days).

Chromatic dispersion monitoring is most often done at the receiving end, where the Q-factor of the received data or some other means is employed to assess the accumulated dispersion. Existing techniques that can monitor dispersion in-line, fast, and with relatively low cost include: (i) general performance monitoring using the bit-error rate (BER) or eye opening, but this approach cannot differentiate among different degrading effects; (ii) detecting the intensity modulation induced by phase modulation at the transmitter; (iii) extracting the bit-rate frequency component (clock) from photodetected data and monitoring its RF power (see Figure 21); (iv) inserting a subcarrier at the transmitter and subsequently monitoring the subcarrier power degradation; (v) extracting several frequency components from the optical data using very narrow-band optical filters and detecting the optical phase; (vi) dithering the optical-carrier frequency at the transmitter and measuring the resultant phase modulation of the clock extracted at receiver with an additional phase-locked loop; and (vii) using an optical filter to select the upper and lower vestigial sideband (VSB) signals in transmitted optical data and determine the relative group delay caused by dispersion (see Figure 22).

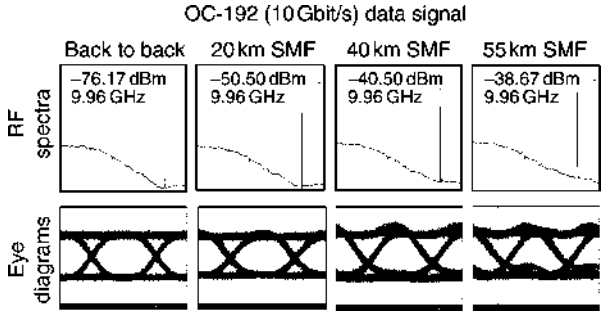

Figure 21 Clock regenerating effect due to chromatic dispersion for NRZ data. As the amount of residual dispersion increases, so does the amount of power at the clock frequency. This power can be used to monitor the amount of uncompensated chromatic dispersion. (Reproduced with permission from Pan Z, Yu Q, Xie Y, *et al.* (2001) Chromatic dispersion monitoring and automated compensation for NRZ and RZ data using clock regeneration and fading without adding signalling. *Optical Fiber Communication Conference 2001.* Paper WH5.)

Conclusion

Chromatic dispersion is a phenomenon with profound implications for optical fiber communications systems. It has negative effects, broadening data pulses, but it also helps reduce the effects of fiber nonlinearities. For this reason, managing dispersion, rather than trying to eliminate it altogether, is the key. Fixed dispersion components are suitable for point-to-point OC-192 systems. Tunable dispersion components are essential for dispersion management in reconfigurable and OC-768 systems. These dispersion compensation elements must have low loss, low nonlinearity, and must be cost effective.

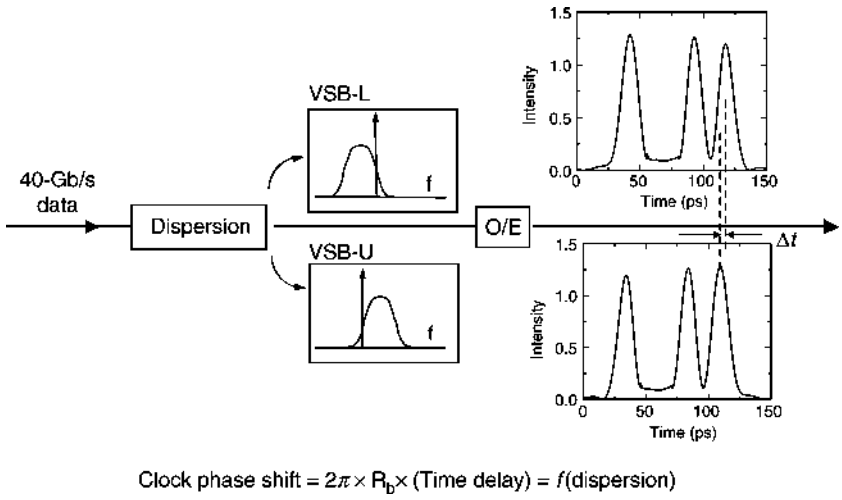

Figure 22 Chromatic dispersion monitoring using the time delay (Δt) between two VSB signals, which is a function of chromatic dispersion. (Reproduced with permission from Yu Q, Yan L-S, Pan Z and Willner AE (2002) Chromatic dispersion monitor for WDM systems using vestigial-sideband optical filtering. *Optical Fiber Communication Conference 2002.* Paper WE3.)

Although several technologies have emerged that meet some or all of the above requirements, no technology is a clear winner. The trend is towards tunable devices, or even actively self-tunable compensators, and such devices will allow system designers to cope with the shrinking system margins and with the emerging rapidly reconfigurable optical networks.

List of Units and Nomenclature

Chromatic dispersion	The time-domain pulse broadening in an optical fiber caused by the frequency dependence of the refractive index. This results in photons at different frequencies traveling at different speeds [ps/nm/km].
Chromatic dispersion slope	Different wavelengths have different amounts of dispersion values in an optical fiber [ps/nm^2/km].
Conventional single mode fiber (SMF)	Fiber that transmits only a single optical mode by virtue of a very small core diameter relative to the cladding. Provides lowest loss in the 1,550 nm region and has zero dispersion at 1,300 nm.
Cross-phase modulation (XPM)	A nonlinear Kerr effect in which a signal undergoes a nonlinear phase shift induced by a copropagating signal at a different wavelength in a WDM system.
Dispersion compensating fiber (DCF)	Optical fiber that has both a large negative dispersion and dispersion slope around 1,550 nm.
Fiber Bragg gratings (FBGs)	A small section of optical fiber in which there is a periodic change of refractive index along the core of the fiber. An FBG acts as a wavelength-selective mirror, reflecting only specific wavelengths (Bragg wavelengths) and passing others.
Four-wave mixing (FWM)	A nonlinear Kerr effect in which two or more signal wavelengths interact to generate a new wavelength.
Kerr effect	Nonlinear effect in which the refractive index of optical fiber varies as a function of the intensity of light within the fiber core.
Modulation	The process of encoding digital data onto an optical signal so it can be transmitted through an optical network.
Nonlinear effects	Describes the nonlinear responses of a dielectric to intense electromagnetic fields. One such effect is the intensity dependence of the refractive index of an optical fiber (Kerr effect).
Nonzero dispersion-shifted fiber (NZDSF)	Optical fiber that has a small amount of dispersion in the 1,550 nm region in order to reduce deleterious nonlinear effects.
Polarization mode dispersion (PMD)	Dispersion resulting from the fact that different light polarizations within the fiber core will travel at different speeds through the optical fiber [ps/km$^{0.5}$].
Self-phase modulation (SPM)	A nonlinear Kerr effect in which a signal undergoes a self-induced phase shift during propagation in optical fiber.
Wavelength division multiplexing (WDM)	Transmitting many different wavelengths down the same optical fiber at the same time in order to increase the amount of information that can be carried.

See also

Nonlinear Optics, Basics: Four-Wave Mixing. **Photonic Crystals:** Nonlinear Optics in Photonic Crystal Fibers.

Further Reading

Agrawal GP (1997) *Fiber-Optic Communication Systems*. Rochester, NY: John Wiley & Sons.

Agrawal GP (1997) *Nonlinear Fiber Optics*, 2nd edn. Academic Press.

Gowar J (1993) *Optical Communication Systems*, 2nd edn. Prentice Hall.

Kaminow IP and Koch TL (1997) *Optical Fiber Telecommunications IIIA & IIIB*. Academic Press.

Kashyap R (1999) *Fiber Bragg Gratings*. Academic Press.

Kazovsky L, Benedetto S and Willner A (1996) *Optical Fiber Communication Systems*. Artech House.

Willner AE and Hoanca B (2002) Fixed and tunable management of fiber chromatic dispersion. In: Kaminow IP and Li T (eds) *Optical Fiber Telecommunications IV*. Academic Press.

Yariv A and Yeh P (1984) *Optical Waves in Crystals*. John Wiley & Sons.

DISPLAYS

R L Donofrio, Display Device Consultants LLC, Ann Arbor, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

This article is a brief review of display technology being used today. It emphasizes some of the optical aspects of displays. The details of the basic construction of the display devices are also discussed to aid in understanding. The displays covered are the CRT, VFD, FED, PDP, LED, OLED, LCD-AMLCD, transmissive and transflective LCDs.

Definition of a Display

A display can be defined as an object which transfers information to the viewer and/or an output unit that gives a visual representation of data.

Optical Characteristics of a Display

A display has (in general) the optical characteristics of luminance, reflectivity, contrast ratio, color, surface gloss, resolution, and image sharpness (sometimes called small area contrast). Measurement methods can be obtained from the EIA (Electronics Industries Alliance), JT-31 (Optical Characteristics of Display Devices), SAE (Society of Automotive Engineers and ISO Standards) and VESA Standards, to name a few.

Luminance

The luminance of a display is its light output and is given in units of candela/square meter or cd/m^2 (in the English system fL). Luminance is measured using a photometer, photometer/colorimeter, or spectroradiometer.

Illuminance

The illuminance is the light incident on the surface being measured. Its units are Lumen per m^2 or Lux (and in the English system Ft.cd). This is measured using illuminance meters.

Reflectivity

The reflectivity for CRTs (cathode ray tubes) is measured with incident light at 45 degrees (in some cases for LCDs, a smaller 30 degree angle is used). The reflected light from the surface is measured with a detector normal to (perpendicular to) the surface. Devices such as 'tube face reflectivity' (TFR) can be used to measure this parameter and it is published in the EIA-JT-31 measurement methods. Other similar methods are used for flat panel displays.

Color

Although color perception is different for different individuals, standard responses to color stimuli have been developed and we can measure the color of different objects and display images. Color can be measured with such equipment as a colorimeter or spectroradiometer. The output data are given in various color systems. Color systems, such as the 1931 CIE (Commission International de I'Eclairage) Color System (x, y), the 1960 CIE (u, v) or 1976 CIE (u', v') color coordinates are all used. All these color systems can be converted to each other by simple mathematical relations and these conversions are usually available to the user in the data output. The choice of the color space/system is dependent on application but the system in most general use for displays is the 1976 CIE (u', v'). This color system has become more widely used because equally perceived color differences in this color system are at almost equal distances in this color space. In reality, the quest for a system with equal distances equating to equal color differences can be achieved through a complicated matrix approach in which the constants used in the matrix vary from point to point in the color space.

Gloss

Gloss is a measure of surface reflectivity under Snell's Law criteria (equal incidence and reflection angles). The measurement angles of 20, 45, 60, and 85 degrees are used to measure the gloss of common surfaces. Display glass surface measurements tend to use the 60 degree gloss angle. The surface of a display can be made rougher to scatter incident light away from the viewer or it can be optically coated to reduce the reflected light to achieve improved viewing. Some times both methods are used. Reduced gloss surfaces also are involved with the reduction in surface glare. The level of increased roughness of the surface affects the user's ability to see detail.

Contrast Ratio

When people talk about contrast they generally mean 'contrast ratio'. The contrast ratio (CR) of an area on a display is a unitless parameter and is defined as the excited luminance (L_e) divided by

the unexcited luminance (L_u). The excited luminance (L_e) equals the glass transmission (T) times some excited internal luminance (L_{ei}) and for simplicity, L_u equals T times an internal zero ambient unexcited luminance light level of (L_{ui}). Factors, such as internal scattered light in LCDs and scattered electrons and scattered light in CRTs, give a nonzero un-excited luminance L_u.

When dealing with the CR, the effect of ambient illumination A contributes to the numerator and the denominator. Thus, when the ambient illumination is increased, it lowers the contrast ratio. In a simple case where the ambient illumination is reflected from the surface of the display whose reflectivity is ρ, the contrast ratio is given below:

$$\mathrm{CR} = (L_e + A\rho)/(L_u + A\rho) \qquad [1]$$

For the case of the reflectivity of a phosphor screen (reflectance R) coated panel with a glass transmission (T), the reflectivity to first approximation is

$$\rho = cRT^2 \qquad [2]$$

where c is a constant.

So we see that changes in contrast can come about through changing the display's glass panel transmission. For example, if the L_e of the white field of a CRT was 100 fL with a 90% panel glass (along with $R = 0.5$) and had a L_u of 2 fL, then the zero ambient contrast would be 50. However, with an ambient illumination of 50 Ftcd the contrast would be 4.76. Now, when we use a 50% transmission panel glass at the ambient illumination of 50 Ft.cd the CR is increased to approximately 8.39. Equations [1] and [2] show how to calculate the contrast. It should be noted that decreased panel glass transmission or the use of 'dark glass' increases contrast because the reflectivity is a 'T squared' effect and the luminance L is a function 'T'. There are second-order effects involved with the reduction of reflected light and unexcited luminance which make the equations more complicated. These second-order effects will not be covered as they are beyond the basic discussion of this article.

In comparing display CR, one must compare the displays at the same ambient illumination. Additionally, you will find that a number of flat panel display advertising reports will report the CR at zero ambient illumination. As we have shown, these CR values are the highest ones for the display. The CR for emissive displays tends to be constant over a wide range of viewing angles, whereas the contrast ratio for liquid crystal displays (LCDs) is more variable over large-angle viewing. Recent new electrode position developments, such as the 'in-plane switching' technology for active matrix liquid crystal displays (AMLCDs) has gone a long way to reduce this reduced angle-of-view effect in that type of LCD display.

Figure 1 shows how a viewer in a sport utility vehicle (SUV) would see the changes in contrast with viewing angle for two types of twisted nematic LCDs (TN-LCDs). One of the TN LCDs does not

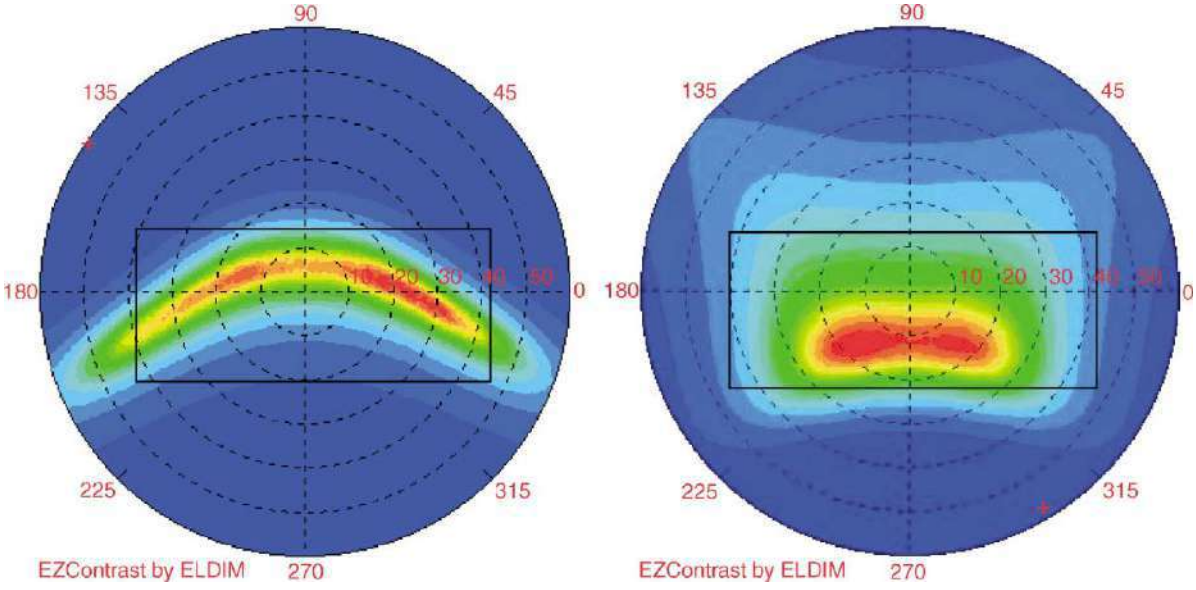

Figure 1 A typical rectangular viewing envelope for a mid-size SUV vehicle superimposed on LCD iso-contrast plots. Left is standard TN LCD. Right, is film-compensated TN LCD. Reproduced from Smith-Gillespie R, *et al.* (2001) Design requirements for automotive entertainment displays. *Proceedings of the 8th Annual Symposium on Vehicle Displays (2001).* Permission for reprint courtesy of the Society for Information Displays.

have the film compensation and the other does. For relative contrast appraisal, the red areas show the highest contrast regions and the blue the lowest. It is obvious that different technologies can give significant changes in viewed contrast.

Phosphor Aging

The fall off of luminance as the phosphor screen is bombarded with the electron beam is called phosphor aging. This effect is due to two main factors: (i) the thermal quenching of the phosphor luminescence (which is reversible); and (ii) a nonreversible aging due to damage to the screen components (such as phosphor, binder, and aluminum coating). It has been found that the aging (when we are using an electron beam) is a function of the number of electrons bombarding the screen. Aging models help us to understand and predict the amount of aging a tube will show after a fixed amount of time. The first aging model in 1951 was on ion bombardment and showed that:

$$I = I_0/(1 + AN) \qquad [3]$$

where I_0 is the initial luminance before aging, and I is the final luminance.

In 1961 a similar (one parameter) Model $I = I_0/(1 + CN)$ was used by Pfhanl, where C is the aging constant to half luminance and N is the number of coulombs. In 1981, a two-component model for low current aging was presented, where:

$$I = I_0/(1 + N(B/Q^D)) \qquad [4]$$

Here we have two constants B and D, and Q is the cumulated charge in number of coulombs. When $D = 0$ then the equation given above goes to the Pfhanl case.

Later, in 1996 the three-parameter model was described and in 1997 the Gaussian beam aging model was described.

Ambient Illumination and Reflection

To see the information shown on a display, the information must be spatially detected by the human eye over the reflected ambient illumination (such as room and sunlight). The ambient illumination is generally from localized or diffuse sources. As we have seen previously, the effect of diffuse illumination can be reduced through the use of neutral density filters. But, when we are dealing with very high ambient light such as that entering the canopy of a fighter aircraft then, the monochromatic information from the display can be viewed using selective spectral filters. However, for the usual computer monitor, the 'specular reflection' (Snell's Law type of reflection) off the front of the display is due to a localized external light source. This reflection is reduced by methods such as anti-glare (AG) and anti-reflecting (AR) surfaces applied to the front surface of the display. A flat versus curved front surface of the display can also be a glare reduction technique. Surface roughness is used in AG surfaces to scatter a portion of the specular light away from the viewer. Surface roughness can be achieved through surface abrading, surface plastic coating, surface silicate coating, and surface acid etching to name a few methods. AR surfaces use optical interference techniques to reduce the reflected light to the viewer. In some cases, both methods are applied to a display surface. Surface gloss of the display is usually measured with a gloss meter in 60 degree gloss values but 20, 45, and 80 degree angles are also used. Resolution is measured in line pairs/mm or pixels/mm and image sharpness is associated with modulation transfer function (MTF) measurements. We should note here that resolution and image sharpness are not the same thing. High resolution is equated with the high frequency limit of the MTF measurement. Since the reduction in specular reflection through the use of surface roughness can affect the image sharpness, we see that a system of trade-offs can be developed to optimize a display for a given viewing environment. AR coatings improve glare reduction without scattering and thus give higher-resolution capability to the displays but this is usually obtained at a higher price then the AG coatings.

Types of Displays

Displays can be divided into two types; emissive and nonemissive. By emissive we mean that the display surface gives off light. In the nonemissive, the display or a small portion of the display acts as a shutter or spatial filter which also can be reflective in nature. In this nonemissive mode, the display can pass light from a source in back of or in front of the shutter (or both, as we shall see later).

Table of Emissive and Nonemissive Displays

Thus from Table 1 we see that under the title of emissive displays we have the cathode ray tube (CRT), projection CRT, vacuum fluorescent display (VFD), field emissive display (FED), electroluminescent (EL) display, light emitting diode (LED), organic light emitting diode (OLED), and plasma. These devices, which are in the realm of emissive displays, use certain types of emissive materials called

Table 1 Table of emissive and nonemissive displays

Emissive	*Nonemissive*
CRT	LCD and AMLCD
VFD	Bi-Stable LCD
FED	Digital Mirror D
EL – inorganic	LCOS (Liquid Crystal on Si)
LED	Electronic Inks
OLED	
PLASMA	

lumophors (or phosphors). The lumophors are called cathodoluminescent, photoluminescent, and electroluminescent, depending whether the excitation action was due to the interaction of the phosphor with cathode rays, photons, or low-voltage electrons. Cathodoluminescent lumophors are used in cathode ray tubes (CRTs – television tubes, monitor tubes, projection CRTs), (VFDs), and (FEDs).

EL displays can use inorganic lumophors (phosphors) or organic lumophors such as those used in OLEDs. Under photoluminescence displays, we have 'plasma displays' (plasma TV), IR-up conversion, and fluorescent lamps. In plasma displays, the excitation of the plasma in a localized region gives rise to UV photons which strike the phosphor screen in that region. In IR-up conversion, IR photons are absorbed in such a way that two IR photons can excite an electron with twice the energy of one IR photon. This then allows the use of IR lasers to create displays through the excitation of these up-conversion phosphors. In fluorescent lamps, the UV created by the excitation of Hg or Xe, strikes the special lamp phosphors (coated on the inside of the glass envelope), which become luminescent.

Under the heading of shutters we can consider LCDs which consist of AMLCDs, super twisted nematic liquid crystal displays (STN LCDs) and bistable LCDs. Another shutter type is the liquid crystal on silicon (LCOS) display and the movable digital mirror display (DMD), and the digital light processing (DLP) projecting engines. Another spatial display type is both normal and dynamic inks. A dynamic ink can be a 'polarizable' ink display in which a two shade (light/dark) ink sphere element can rotate under the presence of an electric field to show the light or dark side. Although they are very interesting, we will not cover ink displays in this article.

The shutter action is also used in some projection TV sets (or projectors) in which the LCD either transmits or reflects (such as LCOS) the light from an arc lamp. The DMD movable mirrors reflect the digital information to the lens system and the unwanted rays are reflected away from the lens to a black light capture area.

The Cathode Ray Tube (CRT)

The CRT is a display which has been in use for over 100 years (the Braun CRT tube was developed in 1898). The CRT uses a phosphor screen which is bombarded by energetic electrons and emits visible light radiation. The single electron gun CRT was used for years in early monochrome television sets and monochrome computer displays. Monochrome green, amber, and white phosphors have been used in monitor tubes. Some color selectivity can be obtained with the single electron gun using the following three methods; (i) voltage penetration phosphors, (ii) current sensitive phosphors and (iii) beam index methods. In a CRT, the penetration of the electron beam follows Hans Bethe's energy loss equation for electron penetration in solids and thus is a function of the electron energy or CRT voltage.

Hans Bethe's energy loss equation is:

$$\mathrm{d}E/\mathrm{d}x = -2\Pi e^4(N_e/E)\ln(\alpha E/I) \qquad [5]$$

where: E = energy of the incident electron; x = depth of material; N_e = density of atomic electrons; I = mean excitation energy; and α = a constant.

In the voltage penetration tube, a cascaded screen is used to give a color change when the electron voltage is increased. For example, if a cascade screen is made from a green phosphor layer placed on top of the red phosphor layer then the electrons strike the green phosphor first. Then with increased voltage, the electrons pass through the top phosphor and excite the bottom red phosphor. In the current sensitive color tube, the phosphor screen is composed of (for example) red and green phosphor whose luminance versus current curves are different. In these cases, we are referring to the linearity of the change in luminescence from the phosphor as a result of a change in electron beam current. The red phosphor (generally a rare earth oxide) is a linear phosphor, whereas the green is nonlinear. Changing the current thus can change the color of the display. In the beam index tube, a UV (or secondary electron emitters have also been used) emitting index layer is put along side a red, green, or blue triad (for example, between the blue and the red phosphor stripes). The scanning electron beam, through the use of a UV sensor (or secondary electron detector) mounted in the rear portion of the tube and the scanning electronics, can determine the location of the red, green, and blue phosphors. The electron beam can then excite only the individual phosphor areas of interest. With this tube, generally, the

electron beam is at a low current to give a small enough cross-section to only excite one color and thus has limited application.

Another application of the monochrome is the projection CRT system. Here, three tubes for red, green, and blue are used to project an image on a viewing screen. The terminology is such that, when the CRTs are in the back of the screen and the viewer in front of the screen, it is called a 'rear' projection set and if the CRTs and the viewer are in front of the screen it is a 'front' projection set. The principle here is that, either using one CRT tube at a time or by a superposition two or three of the CRTs, the color of the displayed image will be within the color gamut defined by the primary colors.

The Color CRT

The more ubiquitous method of obtaining a color display is the device called the shadow mask CRT (and the similar Trinitron CRT). This color CRT is designed in such a manner that through the use of an aperture mask, (seen in Figure 1) the electrons from the red electron gun strike the red phosphor and likewise electrons from the green and blue electron guns strike the green and blue phosphors. The combination allows the viewer to see a color image. This color separation action is seen in Figure 2. Although Figure 2 shows a delta tri-color design, there are also in-line systems in which the mask consists of an array of slots which are somewhat rectangular in nature. In this array, many slots can be on a line, or a mask made can be made of a series of wires as used in the Trinitron design.

The disadvantage of the shadow mask technique is that the mask can capture as much as 80% of the electrons from the electron guns. This yields lower screen luminance and lower contrast due to scattered electrons. As the screen size gets larger from 19 inches to 36 inches or more, additional temperature compensation methods must be used to reduce the effects of mask heating. A number of methods to control the shift of the electron beam on the phosphor as the mask heats up have been developed. Initially, axially positioned bimetal springs were used on masks to compensate for the heating effect. Then lighter masks were used with corner lock suspension systems. Another method is using mask materials with lower coefficients of expansion such as INVAR.

Contrast in color CRTs has been improved by first applying a carbon coating between the panel glass and the phosphor screen. This requires additional process steps which form 'windows' in the carbon coating (matrix) through which the excited phosphor is seen. The use of pigmented phosphors also helped to reduce the reflectivity of the phosphor/matrix screen.

The phosphors used in the shadow mask color CRT are the red Y_2O_2S:Eu or, in some cases, the Y_2O_3:Eu phosphor. The green phosphor is ZnS:Cu and the blue phosphor is ZnS:Ag.

The CRT seen in Figure 3 shows the new high-definition television-type (HDTV) tube with the 16×9 aspect ratio as contrasted with the older 4×3 aspect ratio (screen width $\times$ height).

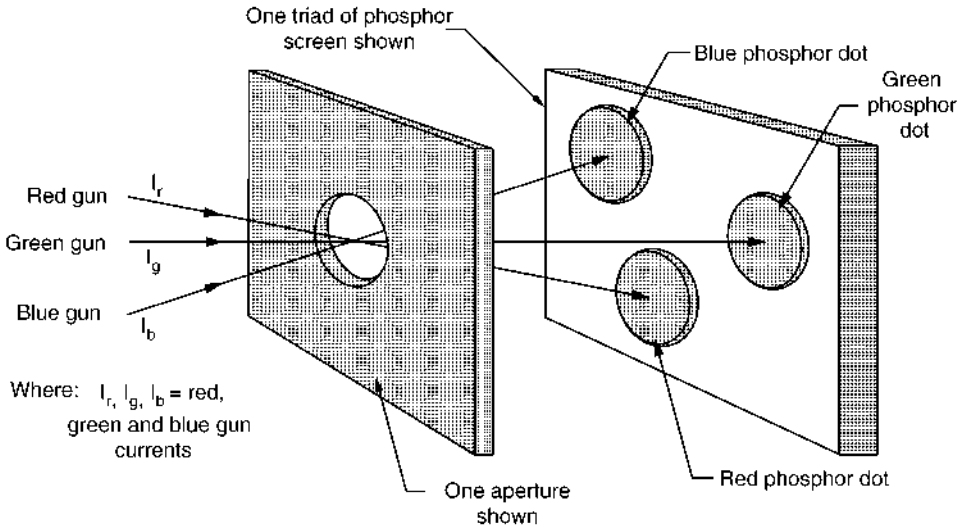

Figure 2 The action of a shadow mask.

Figure 3 A typical HDTV color CRT. Reproduced from Donofrio RL (2002) The future of direct-view CRTs. *SID Information Display Magazine* 6: 10. Permission for reprint courtesy of the Society for Information Display.

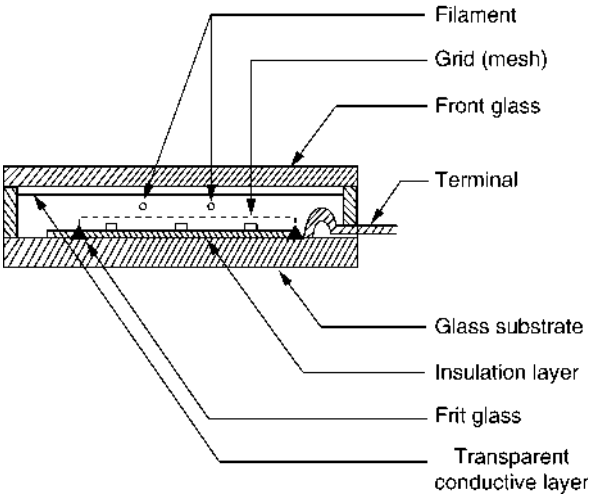

Figure 4 The cross-section of a vacuum fluorescent display assembly.

The Vacuum Fluorescent Display (VFD)

The VFD can be thought of as a low-voltage CRT. Just as in the CRT, electrons from a cathode are controlled by a grid which allows them to be accelerated and strike the phosphor screen anode. However, the electrons in a CRT can be accelerated with a voltage of 30 000 V, whereas the VFD accelerates the electrons with voltages of 12 to 40 V.

In the operation of the VFD, electrons are emitted by the filament. The filament is a very thin tungsten wire coated with a tri-carbonate oxide (barium, strontium, and calcium carbonates) which are the same constituents of the CRT oxide cathode. The filament is held between a filament support and an anchor which applies tension to the filament wire. A current is passed through the filament to achieve a temperature to ~600 °C. At that temperature, thermonic emission occurs. Between the filament and the phosphor elements is a stainless steel mesh grid. By the application of a positive or negative voltage on this grid, electrons can either be cut off or accelerated to the phosphor on the anode. The anode can be a conductor such as graphite, coated with a phosphor layer such as ZnO:Zn (zinc oxide). This is a low-voltage cathodoluminescent phosphor which emits a green color with an emission maxima of about 505 nm. Other phosphors are also used to obtain red, orange, and blue (Figure 4).

In addition to the filament, grid and anode, the VFD is a vacuum device (like the CRT) and contains a getter to assist achieving and maintaining a good vacuum and an internal transparent conductive coating to protect the display from external electrostatic fields. The VFD display is used frequently in the auto industry and aging of the phosphor is very small, achieving 100 000 hrs to half of the initial luminance. VFD units have been made with phosphor elements on either side of the filament. This gives rise to a 3D effect. Graphic and fixed segment VFDs have been made for multimedia products and 16 × 16 graphics for Kanji and other characters.

Experimentally, 2D VFD arrays have been made similar to the FEDs covered in the next section.

Field Emissive Displays (FEDs)

The FED is similar to the CRT and the VFD in that electrons from a source bombard a phosphor screen. In both the CRT and the VFD thermionic electrons are used. However, in the FED the electrons arise from field emission. Field emission devices create electrons though quantum mechanical tunneling from the emitter surface into the vacuum by the action of a large electrical field. This process is governed by the Fowler–Nordheim equation. This equation takes the form:

$$J = A(\beta V)^2/\Phi \exp(-B\Phi^{3/2}/\beta V) \qquad [6]$$

where, J is the electron current density; V is the applied voltage; Φ is the work function of the emitting tip; A & B are constants; and β is the field enhancement factor.

The emitters can be of the Molybdenum Spindt type and also the polydiamond or diamond-like carbon (DLC) coated emitters. We have the field enhancement with a sharp tip and reduced work function with the DLC. There has been renewed interest in FEDs, with the development of the 'carbon nanotube' as an electron field emitting source. The carbon nanotube is a fullerene material, which is a distinct form of carbon. This 'bucky ball'-like

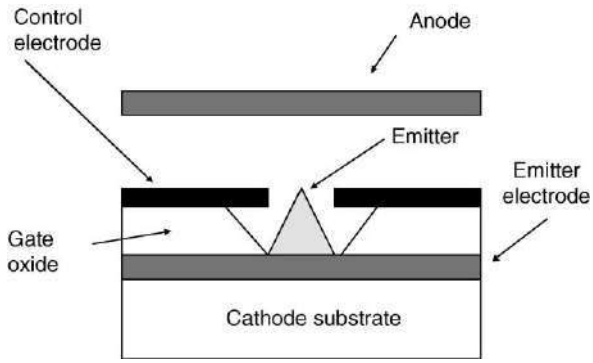

Figure 5 Field emission display cross-section. Reproduced from Green P and Haven D (1998) Fundamentals of emissive displays. *SID Short Course S-3*. Permission for reprint courtesy of the Society for Information Display.

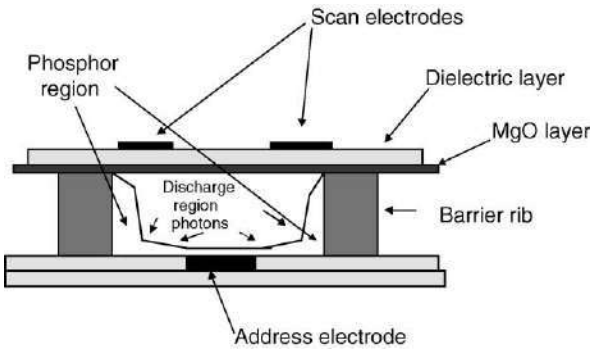

Figure 6 ACPDP partial cross-section view. Reproduced from Green P and Haven D (1998) Fundamentals of emissive displays. *SID Short Course S-3*. Permission for reprint courtesy of the Society for Information Display.

material is cylindrical instead of spherical in nature, has great strength, and high conductivity.

There are two types of FEDs, a low-voltage (LVFED) and a high-voltage (HVFED). The HVFED uses an aluminum film over the phosphor to protect the phosphor screen from aging. The HVFED can use the high voltage (2–10 KV) electrons to bombard the aluminized screen and it can use the more efficient, somewhat standard CRT phosphors ZnS:Cu green, ZnS:Ag Blue, and Y_2O_3:Eu red. The LVFED (12–200 V) cannot use the standard CRT phosphors but uses the low-voltage phosphor such as the ZnO:Zn green phosphor. This does not use an aluminum layer, since about 2 KV is needed to penetrate the Al layer. The commonly used CRT red phosphor Y_2O_2S:Eu was found unsuitable for HVFED use due to release of the sulfur during electron bombardment.

Figure 5 shows that for this basic FED structure, the electrons from the Spindt emitter are controlled by the voltages of the anode, control electrode and emitter electrode.

Plasma Displays

Earlier plasma displays, made using excited Ne gas with the characteristic amber glow required no phosphor and were used in the printing industry. Present plasma displays used for television and information signs are photoluminescent displays. As shown in Figure 6, the excitation of the plasma gas made of elements such as He, Ne, Ar, and Xe give rise to a glow discharge and the creation of vacuum ultraviolet photons that strike the screen and cause it to luminesce. In construction, the display is similar to the FED in that it is a matrix addressed device. It has been described in literature that the similarity goes even further with the need for anode and cathode surfaces, a means of evacuating and a perimeter seal. There are two types of plasma displays, DC and AC. The AC design appears to be the more successful. Typical PDP phosphors for the triad are (Y,Gd)BO_3:Eu for the red, Zn_2SiO_4:Mn for the green, and $BaMgAl_{10}O_{17}$:Eu (called BAM for short) for the blue. Figure 6 below shows the basic structure of an AC plasma display. One phosphor element is shown with barrier ribs and address and scan electrodes. The discharge of the plasma takes place in the region of the phosphor and excites the photoluminescent phosphor, which is shaped (in this example) along the side of the barrier ribs and over the address electrode.

Electroluminescence

Electroluminescence was first seen in the 1930s by Destriau. EL is the generation of (nonthermal) light by use of an electric field. High electric field EL units emit light through the impact of high energy electrons in luminescent centers in a semi-conducting material. The electric field can be alternating current (AC) or direct current (DC). Powder phosphor AC EL displays appear to be one of the successful inorganic EL display methods. Because of their rugged nature, alternating current electroluminescence ACEL displays have been used in trucks, tractors and the M1 Abrams Tank. Thick-film EL panels are also used in some vehicle displays as back lights for LCDs. These light panels are free from mercury and are more environmentally friendly.

Inorganic LED Displays

Light-emitting diodes are being used as backlights for LCD displays, replacing some of the fluorescent light backlighting methods. They are also used in indicator lights. Large displays have been made of LEDs. Using a matrix array row and column connectors,

we can digitally select a given color LED and pattern in an LED array and generate a color image. Similar methods were carried out many years ago using incandescent lights. However, now they are bright and fast reacting. LEDs operate by semiconductor P–N junction rules and the flow of electrons and holes under forward bias operation give rise to junction light emission. Materials such as GaP:Zn,O are used to create a red LED with an emission maxima wavelength of 699 nm.

Red, green, and blue-UV/blue LEDs are now available, with an increasing demand for white LEDs. Two methods are used to produce a white LED. One method is to use a blue LED with the front of the LED coated with a green photoluminescent phosphor. If the phosphor is thin enough, the blue light penetrates the screen and excites the green to give a blue/green white. Another method is to use a blue LED and coat it with a white photoluminescent phosphor.

Organic LEDs

The basic process of light generation in OLEDs is the formation of a P–N junction. Electrons injected from the N-type material and holes from the P-type material recombine at the junction and light is generated. There are presently two competing methods to fabricate an OLED. The first developed by Kodak in 1987 is the small molecule molecular weight material approach. In OLED operation, the electron hole pair recombine to form excitons (on the average one singlet and three triplet excitons are formed for every four electron hole pairs). The light for the Kodak OLED is from the recombination of singlet excitons and the energy transferred to the fluorescent dye. The terminology is such that we have small molecule-OLEDs (sm-OLEDs) and polymer-OLED (PLED). In later developments, the sm-OLED device structure was reported to have three layers. The first is the hole transport layer (hole transport) enhancing layer copper phthalocyanine (CuPc), a layer NPB (60 nm), the light emitting aluminum tris-8-hydroxyquinolate Alq layer (75 nm), and electron transport layer. The electron transport layer is in contact with the three layer structure. The top electrode is a low work function material, such as MgAg (75–200 nm thick) called an electron injection layer and topped with a transparent indium tin oxide (ITO) layer for transparent OLEDS. All of these coatings are applied in a vacuum chamber.

In the phosphorescent material method, the light emission is from the triplet excitons through the use of phosphorescent dyes and the polymer-based devices. The common base material is a polyphenylene-vinylene PPV and other polymers such as polyfluorene have been used. The construction of the phosphorescent OLED is similar in that the ITO is also used as a starting substrate but the hole transport layer is applied by spin casting from solution. Spin casting is also used for the electron transfer material. After drying (removal of solvents) a calcium cathode material is applied. A contrast of 3000/1 has been achieved (for a military type application) using a green OLED with an AR coated circular polarizer and at an ambient light level of 500 lux, and a 6/1 contrast ratio has been achieved with an ambient illumination of 50 000 lux and an output of 370 cd/m^2. Additionally, OLEDs can be made on flexible (F) substrates and are called FOLEDs. The use of a transparent (T) cathode gives rise to a top emission OLED or a TOLED.

The first OLED display used in auto products was developed by Pioneer and used in a radio display and the Ford Motor Company has shown a White OLED display in one of their 'Concept Cars'. This display allows them to have a reconfigurable instrument

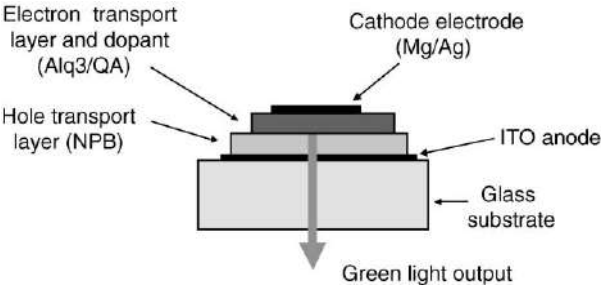

Figure 7 OLED partial cross-section – Kodak type. Reproduced from Green P and Haven D (1998) Fundamentals of emissive displays. *SID Short Course S-3*. Permission for reprint courtesy of the Society for Information Display.

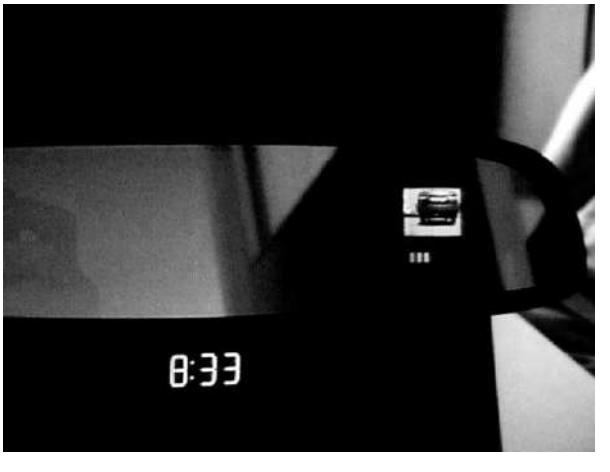

Figure 8 Philips poled display in an auto mirror. Reproduced with permission of Philips Mobile Display Systems.

panel (IP) and is made up of a number of 5 inch × 8 inch displays across the instrument panel (Figure 7).

The POLED as seen in Figure 8 is now being used by Philips Mobile Systems in a rear view car mirror, as seen below. It is the display from a video camera looking at the rear of a car to see any obstacles to backing up safely.

Thin Film EL (TFEL) Displays

A simple type of TFEL display is the color by white structure shown below. In this case, the color is generated using a white phosphor and red, green, and blue optical filters. To make this ACEL display function properly, insulators are used on either side of the phosphor. In Figure 9, the ITO conducting stripes (depicted) horizontally on the top and vertically on the bottom, give the means of addressing each color pixel. The white phosphor, for example, is a ZnS:Mn, SrS:Ce blend which gives a white electroluminescence.

Other phosphors such as ZnS:Mn can be used but its yellow luminescence requires the use of a green and red filter and only two metal electrodes on the bottom to give a two-color display. Better red green color separation can be achieved through the use of a $Zn_{1-x}Mg_xS$:Mn and ZnS:Mn phosphor mix. New EL blue phosphor developments have allowed larger color gamuts to be achieved.

One of the important characteristics of the alternating current thin-film electroluminescence (ACTFEL) is that it has an operating temperature range from −40 °C to 85 °C which is key for automotive use. It has a viewing angle of 160 degrees with a life of more than 50 000 hours. Active matrix (active addressed) AMEL structures allow fast

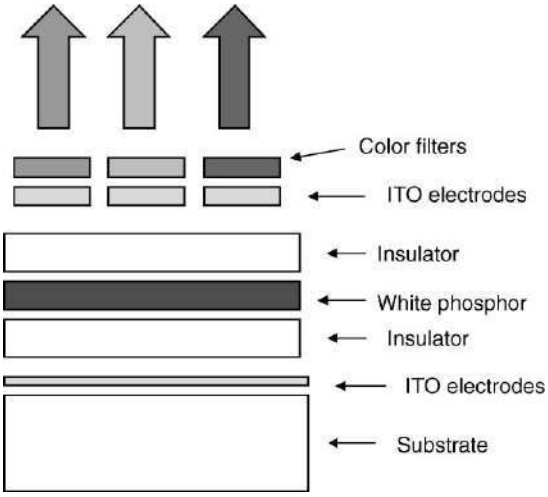

Figure 9 Color thin film electroluminescent display (TFEL) cross-section. Reproduced from Dakkala A (1998) New EL displays for vehicular applications. In: Donofrio RL and Musa S (eds) *Flat Panel and Vehicle Display '98.* Permission for reprint courtesy of the Society for Information Display.

Figure 10 AMLCDs used in automobiles. Reproduced from Donofrio RL (2003) The road show. *SID Information Display Magazine* 7: 26. Permission for reprint courtesy of the Society for Information Display.

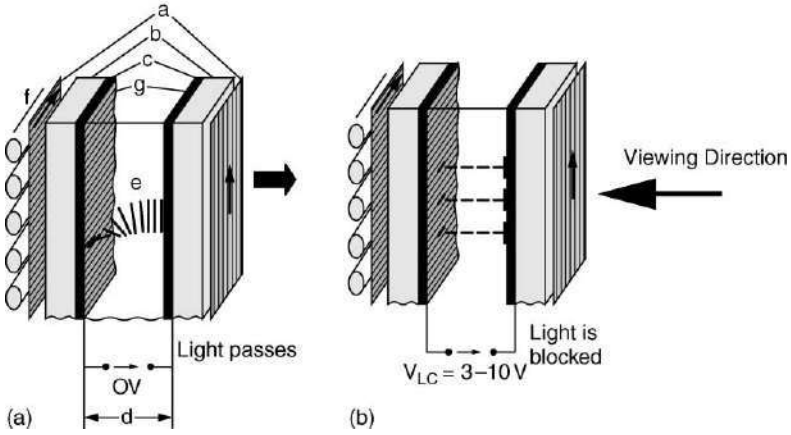

Figure 11 The operation of a simple liquid crystal display. Reproduced from Lueder E (1997) Fundamentals of passive and active-addressed liquid crystal displays. *SID Short Course S-1* (1997). Permission for reprint courtesy of the Society for Information Display.

response and moving graphic displays. These have found application in 'heads up displays' (HUD) for the military and medical industries and in consumer goods such as portable computers, pagers, cell phones, and 3D gaming.

The Active Matrix LCD

The active matrix liquid crystal display (AMLCD) device consists of a series of liquid crystal cells or windows whose characteristics can be controlled by an applied voltage to pass or restrict the passage of light through the windows via the external polarizers and the polarizing action of the LCD cells. These displays are used in laptop and tabletop computers, TVs, and for automotive displays (for driving information and entertainment). Figure 10 shows some present applications for AMLCDs in the automobile.

When the simple LCD is used without a transistor matrix, it is termed a passive display rather than an active one. Figure 11 below shows how a simple LCD cell operates.

The action shown in this figure is the twisted alignment nature of the liquid crystal material. When the polarizer on the left has its high transmission axis horizontal and the polarizer on the right of this cell has perpendicular alignment then the light from the backlight (f) is passed through the crossed polarizers. However, when voltage is applied to the cell, the liquid crystal is no longer twisted and the crossed polarizers prevent the light from going though the cell to the viewer. If the polarizers were not crossed but the polarization transmission axis was in the same direction in both polarizers, then the opposite effect would occur.

In order to show a color image, the LCD is designed to have three or more cells clustered per color pixel. The following (Figure 12) is the transmission profiles of each of these red, green, and blue filters.

The backlight used is a fluorescent lamp with a white phosphor (made up of red, green, and blue photoluminescent phosphors). Figure 13 shows what would happen if only the blue phosphor was used and also shows the superposition of the mercury lines that penetrate the phosphor and contribute to the final spectral energy distribution.

Reflective LCD

For applications where there is suitable external light to view the display, no internal light source is used.

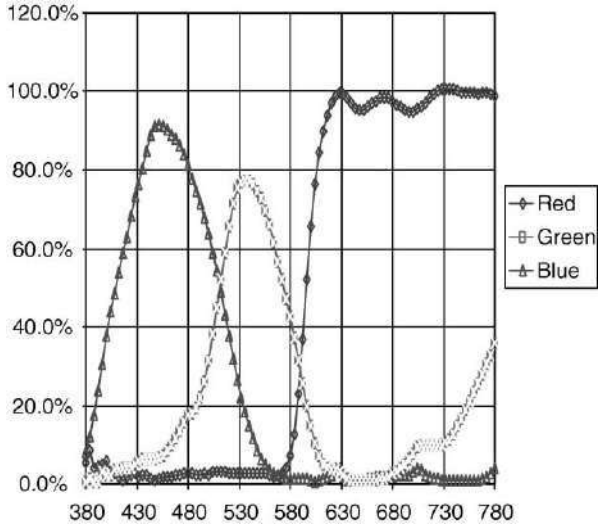

Figure 12 Typical color LCD filter transmission characteristics. Reproduced from Donofrio RL and Abileah A (1998) LCD backlighting system – phosphors and colors. In: Donofrio RL and Musa S (eds) *Flat Panel and Vehicle Display '98, p. 123*. Permission for reprint courtesy of the Society for Information Display.

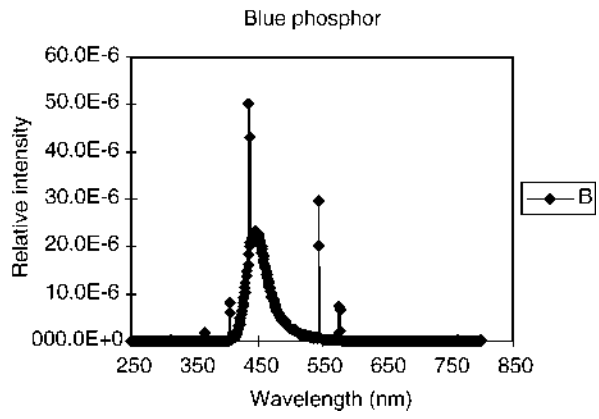

Figure 13 The relative luminance spectrum of the LCD. Backlight blue phosphor showing mercury lines. Reproduced from Donofrio RL and Abileah A (1998) LCD backlighting system – phosphors and colors. In: Donofrio RL and Musa S (eds) *Flat Panel and Vehicle Display '98, p. 123*. Permission for reprint courtesy of the Society for Information Display.

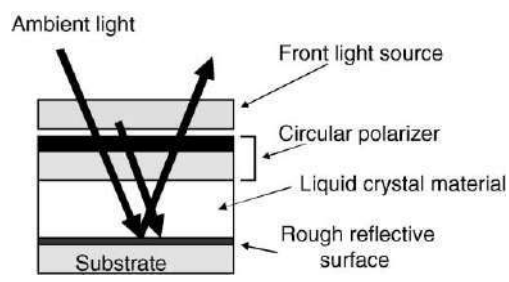

Figure 14 The action and construction of a reflective LCD. Reproduced with permission from Watanabe R and Tomita O (2003) Active-Matrix LCDs for Mobile Telephones in Japan. *SID Information Display Magazine* 7(3): 30–35. Permission for reprint courtesy of the Society for Information Display.

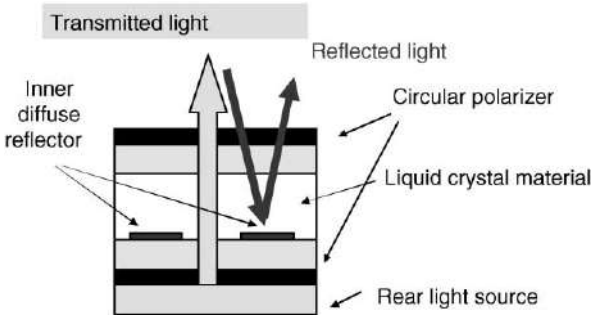

Figure 15 The action and cross-section of a transflective LCD. Reproduced with permission from Watanabe R and Tomita O (2003) Active-Matrix LCDs for Mobile Telephones in Japan. *SID Information Display Magazine* 7(3): 30–35. Permission for reprint courtesy of the Society for Information Display.

This reflective LCD consists of a polarizer and retarder plates stacked on the front side of the cell. The cell with spacers contains the liquid crystal material and on the bottom of the cell is located a reflecting surface. It has been found that a dimpled reflecting surface makes a diffuse reflecting surface (Figure 14).

Transflective LCD

The importance of the transflective device is that it has better contrast under high ambient illumination than the standard transmission back lighted LCD and fair contrast for low ambient illumination.

The construction of this device consists of front and rear circular polarizers, back light, reflective elements, and novel filters. The back light is in the rear of the display. However, like the reflective LCD, internal reflectors are used on each pixel. The filters are designed spatially to be more transmissive for the reflective light than for the back light. Additionally the cell gap for the reflective light is smaller then the transmitted light region because the reflected light passes the cell twice whereas the transmitted light passes the cell once. This type of device seen in Figure 15, is finding uses in the auto industry and where the LCD is needed for both low and high ambient light levels.

Conclusions

In this brief review of displays we have tried to cover many different types of displays. The reader will in time find that there are many other variations on the displays discussed. The object is to give the reader an up-to-date review of the topics involved with displays and the various types of display in use today. By its nature it is a broad approach.

See also

Color and the World. **Instrumentation:** Photometry. **Light Emitting Diodes**. **Materials for Nonlinear Optics:** Liquid Crystals for NLO.

Further Reading

Donofrio RL (1981) Low Current Density Aging. *Proceedings of the Electro-Chemical Society*, Abstract #152, May 11, 1981.

Donofrio RL (2003) The road show. *Information Display* 7(3): 26–29.

Donofrio RL and Abileah A (1998) LCD backlighting. *Flat Panel and Vehicle Displays '98* : 119–124. Metropolitan Detroit Chapter of the SID.

Donofrio RL, Musa S and Wise K (eds) (1999) Vehicle Displays and Microsensors '99. *Proceedings of the 6th Strategic Forum and Technical Symposium*, Sept 22–23, 1999.

Flat Panel Display Measurements Standard, Version 1.0 VESA (Video Electronics Standards Association) May 15, 1998.

Forrest S (2002) *Fundamentals of Organic Light-Emitting Devices*. Short Course S-2 SID.

Green P and Haven D (1998) *Fundamentals of Emissive Displays*, Short Course S-3, SID.

Keller P (1991) *The Cathode-Ray Tube, Technology, History, and Applications*. New York: Palisades Press.

Leverence HW (1968) *An Introduction to Luminescence of Solids*. Dover Publications.

MacDonald W and Lowe A (1997) *Display Systems – Design and Applications*. SID Series in Display Technology, Wiley and Sons Ltd.

Sherr S (1993) *Electronic Displays*, 2nd edn. John Wiley.

Shionoya S and Yen WM (1998) *Phosphor Handbook*. CRC Press.

Stupp EH and Brennesholtz MS (1999) *Projection Displays*. SID Series in Display Technology, Wiley and Sons Ltd.

Watanabe R and Tomita O (2003) Active-Matrix LCDs for Mobile Telephones in Japan. *SID Information Display Magazine* 7(3): 30–35.

Web site atis.org/tg2k/. American National Standard TI.523-2001 'Def of Display Device'.

Wu ST and Yang DK (2001) *Reflective Liquid Crystal Displays*. SID Series in Displays Technology, Wiley and Sons Ltd.

E

ELECTROMAGNETICALLY INDUCED TRANSPARENCY

J P Marangos, Imperial College of Science, Technology and Medicine, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In this article we examine how the process of electromagnetically induced transparency (EIT) can be used to increase the efficiency of non-linear optical frequency mixing schemes. We illustrate the basic ideas by concentrating upon the enhancement of resonant four-wave mixing schemes in atomic gases. To start we introduce the quantum interference phenomenon that gives rise to EIT. The calculation of EIT effects in a three-level atomic medium using a density matrix approach is presented. Next we examine the modified susceptibilities that result and that can be used to describe nonlinear mixing, and show how large enhancements in nonlinear conversion efficiency can result. Specific examples are briefly discussed along with a further discussion of the novel aspects of refractive index and pulse propagation modifications that result from EIT. The potential benefits to nonlinear optics of electromagnetically induced transparency are discussed. In an article of this nature it is not possible to credit all of the important contributions to this field, but we include a bibliography for further reading indicating some of the seminal contributions.

Electromagnetically induced transparency is the term applied to the process by which an otherwise opaque medium can be rendered transparent through laser-induced quantum mechanical interference processes. The linear susceptibility is substantially modified by EIT. Absorption is cancelled due to destructive interference between excitation pathways. The dispersion of the medium is likewise modified such that where there is no absorption the refractive index is unity and there can be a large value of normal dispersion leading to low values of group velocity. In contrast to the linear susceptibility the nonlinear susceptibility involving these same resonant laser fields will undergo constructive rather than destructive interference. This leads to a strong enhancement in nonlinear frequency mixing. For a normally transparent medium the nonlinear optical couplings will be small unless large-amplitude fields are applied. The most important consequence of EIT, in contrast to the usual case when a medium is transparent, is that the dispersion is not vanishing and the nonlinear couplings can be very large.

To understand how this comes about we must consider the system illustrated in Figure 1a where because of the resonant condition satisfied for the applied fields the atom can be simplified to just three-levels. The important parameters in this model system are that state |2> is metastable and has a dipole-allowed coupling to state |3>. The coupling between states |1> and |3> is also dipole-allowed

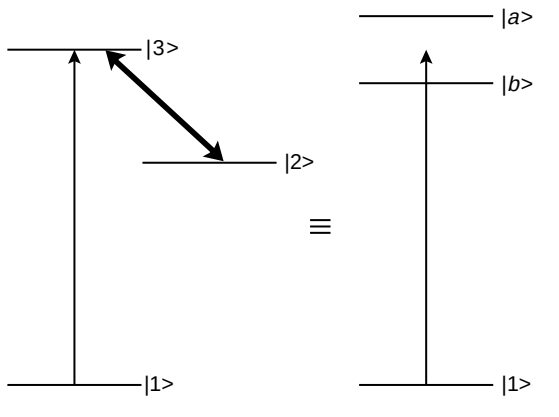

Figure 1 A three-level atomic system coupled to laser fields in the lambda configuration is shown on the left-hand side. In the limit of a strong coupling field this is equivalent to the dressed states |a> and |b> coupled to the ground state by the probe field alone as shown on the right-hand side.

and as a consequence state |3> can decay radiatively to either |1> or |2>. Let a coupling field be applied resonantly to the |2> − |3> transition, which may be cw or pulsed but should in the latter case be transform-limited. We define the Rabi coupling as $\Omega_{23} = \frac{\mu_{23}E}{\hbar}$ where E is the laser electric field amplitude and μ_{23} is the dipole moment of the transition. The Rabi frequency needs to be larger than the inhomogeneous broadening of the sample. A second field, the probe, typically of much lower intensity, is then applied in resonance with the |1> − |3> transition. If the condition $\Omega_{23} \gg \Omega_{13}$ is satisfied then it is convenient to replace the bare atomic states |2> and |3> with the dressed states (see Figure 1b):

$$|a> = \frac{1}{\sqrt{2}}[|2> + |3>] \qquad [1a]$$

$$|b> = \frac{1}{\sqrt{2}}[|2> - |3>] \qquad [1b]$$

Transitions from state |1> induced by the probe field to this pair of near-resonant dressed states are subject to exact cancellation at resonance if |2> is perfectly metastable. This is because the only contribution to the excitation amplitude comes from the matrix elements $<1|\mathbf{r}\cdot\mathbf{E}|3>$ as the |1> − |2> transition is dipole forbidden. The contributions from the equally detuned states |a> and |b> thus enter into the overall amplitude with opposite signs and equal magnitude as can be seen by inspection of eqns [1a] and [1b]. This leads to cancellation of the absorption amplitude. This type of absorption cancellation is well known and closely related to the so-called *dark states*.

Theoretical Treatment of EIT in a Three-Level Medium

It was realized by several workers in the 1970s that laser-induced interference effects could lead to a cancellation of absorption at certain frequencies. To gain a more quantitative understanding of the effects of the coupling field upon the optical properties of a dense ensemble of three-level atoms we require a treatment that computes the optical susceptibility of the medium. A treatment originally carried out by Harris *et al.* for a Λ scheme similar to that illustrated in Figure 1 was the first to derive the modified susceptibilities that will be discussed below. In that treatment the state amplitudes in the three-level atom were solved in the steady-state limit and from these the linear and nonlinear susceptibilities (see below) are then derived. In what follows we adopt a density matrix treatment as employed by various workers. This readily allows the inclusion of dephasing as well as population decay terms. The critical parameters in this system are the strengths of the fields (described in terms of the Rabi couplings), the detuning of the fields from resonance $\Delta\omega_{13} = \omega_{13} - \omega_P$ and $\Delta\omega_{23} = \omega_{23} - \omega_C$ (see Figure 2), the radiative decays from |3> to |1> and |2>, γ_c and γ_d and from |2> − |1> γ_a (although the latter is anticipated to be smaller). Extension to other configurations of the three states, such as a V or ladder scheme is implicit within this general treatment.

Figure 2 illustrates the system considered. This treatment will also address the nonlinear response in a four-wave mixing scheme completed by the two-photon resonant coupling applied between state |1> and |2>.

We write the interaction Hamiltonian as:

$$H = H_0 + V \qquad [2]$$

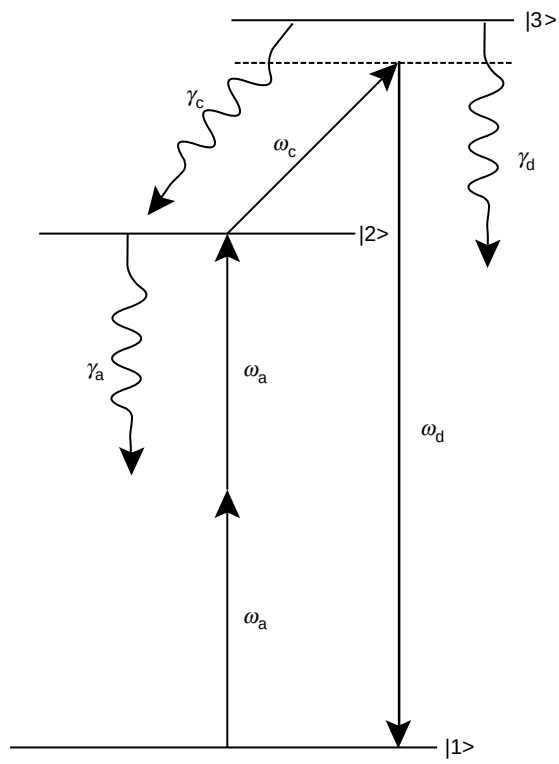

Figure 2 A four-wave mixing scheme incorporating the three-level lambda system in which electromagnetically induced transparency has been created for the generated field ω_d by the coupling field ω_c. The decay rates γ_i from the three atomic levels are also shown. For a full explanation of this system see text.

where H_0 is the unperturbed Hamiltonian of the system and is written as

$$H_0 = \hbar\omega_1|1\rangle\langle 1| + \hbar\omega_2|2\rangle\langle 2| + \hbar\omega_3|3\rangle\langle 3| \quad [3]$$

and V is the interaction Hamiltonian and can be expressed

$$\begin{aligned} V = {} & \hbar\Omega_a e^{-i\omega_a t}|2\rangle\langle 1| + \hbar\Omega_c e^{-i\omega_c t}|2\rangle\langle 3| \\ & + \hbar\Omega_d e^{-i\omega_d t}|3\rangle\langle 1| + c.c \end{aligned} \quad [4]$$

Note that the Rabi frequencies in the equation can be described as $\hbar\Omega_{ij} = \mu_{ij}|E(\omega_{ij})|$, where μ_{ij} is the dipole moment of the transition between states |i> and |j>, the Rabi frequency Ω_a is a two-photon Rabi frequency that characterizes the coupling between the laser field *a* and the atom for this two-photon transition. We have assumed for simplicity that $\omega_a = \omega_b = \omega_{12}/2$.

Assuming all the fields lie close to the resonance, the rotating wave approximation can be applied to the interaction picture and the interaction Hamiltonian V^I is given as

$$\begin{aligned} V^I = {} & \hbar\Omega_a e^{i\Delta_a t}|2\rangle\langle 1| + \hbar\Omega_c e^{i\Delta_c t}|2\rangle\langle 3| \\ & + \hbar\Omega_d e^{i\Delta_d t}|3\rangle\langle 1| + c.c \end{aligned} \quad [5]$$

where Δ_a, Δ_c and Δ_d refer the detunings of the fields and can be written as:

$$\begin{aligned} \Delta_a &= \omega_{12} = 2\omega_a \\ \Delta_c &= \omega_{32} - \omega_c \\ \Delta_d &= \omega_{13} - \omega_d \end{aligned} \quad [6]$$

For the evaluation of the density matrix with this interaction V^I, the Schrödinger equation can be restated in terms of the density matrix components. This form is called the Liouville equation and can be written as:

$$\hbar\frac{\partial}{\partial t}\varrho_{ij}(t) = -i\sum_k H_{ik}(t)\varrho_{ij}(t) + i\sum_k \varrho_{ij}(t)H_{ik}(t) + \Gamma_{ij} \quad [7]$$

where Γ_{ij} represents phenomenologically added decay terms (i.e. spontaneous decays, collisional broadening, etc.). This formalism leads to a set of nine differential equations for nine different density matrix elements that describe the three-level system.

To remove the optical frequency oscillations, a coordinate transform is needed and to incorporate the relevant oscillatory detuning terms into the off-diagonal elements we make the substitution:

$$\begin{aligned} \tilde{\varrho}_{12} &= \varrho_{12}e^{-i\Delta_a t} \\ \tilde{\varrho}_{23} &= \varrho_{23}e^{-i\Delta_c t} \\ \tilde{\varrho}_{31} &= \varrho_{31}e^{-i\Delta_d t} \end{aligned} \quad [8]$$

This operation conveniently eliminates the time dependencies in the rate equations and the equations of motion for the density matrix are given by:

$$\begin{aligned} \frac{\partial}{\partial t}\varrho_{11} &= i\frac{1}{2}\Omega_a\tilde{\varrho}_{12} + i\frac{1}{2}\Omega_d\tilde{\varrho}_{13} - i\frac{1}{2}\Omega_a^*\tilde{\varrho}_{21} \\ &\quad - i\frac{1}{2}\Omega_d^*\tilde{\varrho}_{31} + \Gamma_{11} \\ \frac{\partial}{\partial t}\varrho_{22} &= i\frac{1}{2}\Omega_c^*\tilde{\varrho}_{23} + i\frac{1}{2}\Omega_a^*\tilde{\varrho}_{21} - i\frac{1}{2}\Omega_c\tilde{\varrho}_{32} \\ &\quad - i\frac{1}{2}\Omega_a\tilde{\varrho}_{12} + \Gamma_{22} \\ \frac{\partial}{\partial t}\varrho_{33} &= i\frac{1}{2}\Omega_c^*\tilde{\varrho}_{23} - i\frac{1}{2}\Omega_d^*\tilde{\varrho}_{31} + i\frac{1}{2}\Omega_d\tilde{\varrho}_{13} \\ &\quad - i\frac{1}{2}\Omega_c\tilde{\varrho}_{32} + \Gamma_{33} \\ \frac{\partial}{\partial t}\varrho_{23} &= i\frac{1}{2}\Omega_c\tilde{\varrho}_{22} - i\frac{1}{2}\Omega_c\tilde{\varrho}_{33} - i\Delta_a\tilde{\varrho}_{23} \\ &\quad + i\frac{1}{2}\Omega_d\tilde{\varrho}_{21} - i\frac{1}{2}\Omega_a\tilde{\varrho}_{13} + \Gamma_{23} \\ \frac{\partial}{\partial t}\varrho_{21} &= i\Omega_a\tilde{\varrho}_{22} + i\Omega_a\tilde{\varrho}_{33} + i\Omega_d\tilde{\varrho}_{23} - i\Delta_c\tilde{\varrho}_{21} \\ &\quad - i\frac{1}{2}\Omega_c\tilde{\varrho}_{31} - i\frac{1}{2}\Omega_a\tilde{\varrho}_{31} + \Gamma_{21} \\ \frac{\partial}{\partial t}\varrho_{31} &= i\frac{1}{2}\Omega_d\tilde{\varrho}_{22} + i\Omega_d\tilde{\varrho}_{33} + i\Omega_d\tilde{\varrho}_{23} - i\Delta_c\tilde{\varrho}_{21} \\ &\quad - i\frac{1}{2}\Omega_c\tilde{\varrho}_{31} - i\frac{1}{2}\Omega_a\tilde{\varrho}_{31} + \Gamma_{31} \end{aligned} \quad [9]$$

Using the set of equations above and the relation $\tilde{\varrho}_{ij} = \tilde{\varrho}_{ij}^*$ we obtain equations for $\tilde{\varrho}_{12}$, $\tilde{\varrho}_{32}$, and $\tilde{\varrho}_{13}$. For the incoherent population relaxation the decays can be written:

$$\begin{aligned} \Gamma_{11} &= \gamma_a\varrho_{22} + \gamma_d\varrho_{33} \\ \Gamma_{22} &= -\gamma_a\varrho_{22} + \gamma_c\varrho_{33} \\ \Gamma_{33} &= -(\gamma_c + \gamma_d)\varrho_{33} \end{aligned} \quad [10]$$

and for the coherence damping:

$$\begin{aligned} \Gamma_{21} &= -\left\{\frac{1}{2}[\gamma_a + \gamma_c] + \gamma_{21}^{col}\right\}\varrho_{21} \\ \Gamma_{23} &= -\left\{\frac{1}{2}[\gamma_a + \gamma_c + \gamma_d] + \gamma_{23}^{col}\right\}\varrho_{23} \\ \Gamma_{31} &= -\left\{\frac{1}{2}[\gamma_d + \gamma_c] + \gamma_{31}^{col}\right\}\varrho_{31} \end{aligned} \quad [11]$$

where γ_{ij}^{col} represents collisional dephasing terms which may be present.

This system of equations can be solved by various analytical or numerical methods to give the individual density matrix elements. Analytical solutions are possible if we can assume for instance that $\Omega_c \gg \Omega_a$, Ω_d and that there are continuous-wave fields so a steady-state treatment is valid. For pulsed fields or in the case where the generated field may be significant numerical solutions are in general required.

We are specifically interested in the optical response to a probe field at ω_d (close to resonance with the $|1\rangle - |3\rangle$ transition) that is governed by the magnitude of the coherence ϱ_{13}. We find ϱ_{13} making the assumption that only the coupling field is strong, i.e., that $\Omega_c \gg \Omega_a, \Omega_d$ holds. From this quantity the macroscopic polarization is obtained from which the susceptibility can be computed. The macroscopic polarization P at the transition frequency ω_{13} can be related to the microscopic coherence ϱ_{13} via the expression:

$$P_{13} = N\mu_{13}\varrho_{13} \qquad [12]$$

where N is the number of equivalent atoms in the ground state within the medium, and μ_{13} is the dipole matrix element associated with the transition. In this way real and imaginary parts of the linear susceptibility χ at frequency ω can be directly related to ϱ_{13} via the macroscopic polarization since the latter can be defined as:

$$P_{13}(\omega) = \varepsilon_0\chi(\omega)E \qquad [13]$$

where E is the electric field amplitude at frequency ω_d. The linear susceptibility (real and imaginary parts) is given by the following expressions:

$$\mathrm{Re}\chi_D^{(1)}(-\omega_d,\omega_d) = \frac{|\mu_{13}|^2 N}{\varepsilon_0\hbar} \times \left[\frac{-4\Delta_{21}(|\Omega_2|^2 - 4\Delta_{21}\Delta_{32}) + 4\Delta_{31}\gamma_a^2}{4\Delta_{31}\Delta_{21} - \gamma_d\gamma_a - |\Omega_2|^2)^2 + 4(\gamma_d\Delta_{21} + \gamma_a\Delta_{31})^2}\right] \qquad [14]$$

$$\mathrm{Im}\chi_D^{(1)}(-\omega_d,\omega_d) = \frac{|\mu_{13}|^2 N}{\varepsilon_0\hbar} \times \left[\frac{8\Delta_{21}^2\gamma_d + 2\gamma_a(|\Omega_2|^2 + \gamma_d\gamma_a)}{(4\Delta_{31}\Delta_{21} - \gamma_d\gamma_a - |\Omega_2|^2)^2 + 4(\gamma_d\Delta_{21} + \gamma_a\Delta_{31})^2}\right] \qquad [15]$$

We now consider the additional two-photon resonant field ω_a. This leads to a four-wave mixing process that generates a field at ω_d (i.e., the probe frequency). The nonlinear susceptibility that describes the coupling of the fields in a four-wave mixing process is given by the expression:

$$\chi_D^{(3)}(-\omega_d,\omega_a,\omega_a,\omega_c) = \frac{\mu_{23}\mu_{13}N}{6\varepsilon_0\hbar^3((\Delta_{21} - j\gamma_a/2)(\Delta_{31} - j\gamma_d/2) - (|\Omega_c|^2/4))^*} \times \sum_i \mu_{i1}\mu_{i3}\left(\frac{1}{\omega_i - \omega_a} + \frac{1}{\omega_i - \omega_b}\right)$$

Nonlinear Optical Processes

The dependence of the susceptibilities upon the detuning is plotted in Figure 3. In this plot the effects of inhomogeneous (Doppler) broadening are

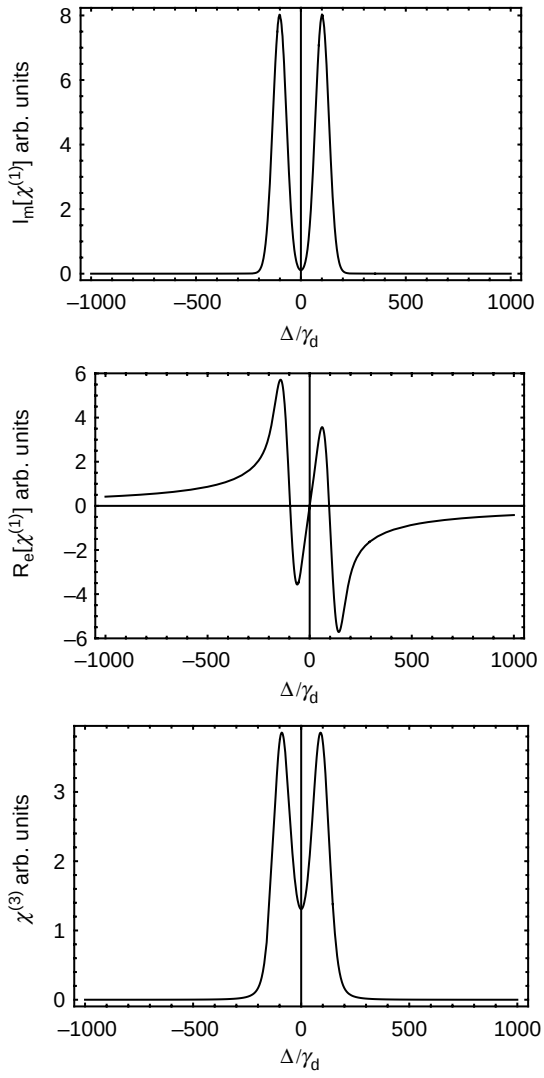

Figure 3 Electromagnetically induced transparency is shown in the case of significant Doppler broadening. The Doppler averaged values of the imaginary $\mathrm{Im}[\chi^{(1)}]$ and real $\mathrm{Re}[\chi^{(1)}]$ parts of the linear susceptibility and the nonlinear susceptibility $\chi^{(3)}$ are plotted in the three frames. The Doppler width is taken to be $20\gamma_d$, $\Omega_c = 100\gamma_d$ and $\gamma_d = 50\gamma_a$, $\gamma_c = 0$.

also incorporated by computing the susceptibilities over the inhomogeneous profile (see below for more details).

By inspection of eqn [14] we see that the absorptive loss at the minimum of $\mathrm{Im}[\chi^{(1)}]$ varies as γ_a/Ω_c^2. This dependence is a consequence of the interference discussed above. In the absence of interference (i.e., just simply Autler–Townes splitting) the minimum loss would vary as γ_d/Ω_c^2. Since $|3\rangle - |1\rangle$ is an allowed decay channel (in contrast to $|2\rangle - |1\rangle$) it follows that $\gamma_b \gg \gamma_a$ and so the absorption is much less as a consequence of EIT. $\mathrm{Re}[\chi^{(1)}]$ in eqn [15] and Figure 3 is also modified significantly. The resonant value of refractive index is equal to the vacuum value where the absorption is a minimum, the dispersion is normal in this region with a gradient determined by the strength of the coupling laser, a point we will return to shortly. For an unmodified system the refractive index is also unity at resonance but in that case there is high absorption and steep anomalous absorption. Reduced group velocities result from the steep normal dispersion that accompanies EIT. Inspection of the expression [16] and Figure 3 shows that $\chi^{(3)}$ is also modified by the coupling field. The nonlinear susceptibility depends upon $1/\Omega_c^2$ as is expected for a laser dressed system, however in this case there is not destructive but rather constructive interference, between the field-split components. This result is of great importance: it ensures that the absorption can be minimized at frequencies where the nonlinear absorption remains large.

As a consequence of constructive interference the nonlinear susceptibility remains resonantly enhanced whilst the linear susceptibility vanishes or becomes very small at resonance due to destructive interference. Large nonlinearity accompanying vanishing absorption (transparency) of course match conditions for efficient frequency mixing as a large atomic density can then be used. Moreover the dispersion (controlling the refractive index) also vanishes at resonance; this leads to perfect phase matching (i.e., zero wavevector mismatch between the fields) in the limit of a simple three-level system. As a result of these features a large enhancement of the conversion efficiency in this type of scheme can be achieved.

To compute the generated field strength Maxwell's equations must be solved using these expression for the susceptibility to describe the absorption, refraction, and nonlinear coupling in the medium. We will treat this within the slowly varying envelope approximation since the fields are cw or nanosecond pulses. To be specific we assume that the susceptibilities are time independent, i.e., that we are in the steady-state (cw) limit. We make the assumptions also that there is no pump field absorption and that we have plane waves. Under these assumptions the generated field amplitude A_d is given in terms of the other field amplitudes A_i by:

$$\frac{\partial}{\partial z}A_\mathrm{d} = \mathrm{i}\frac{\omega_\mathrm{d}}{4c}\chi^{(3)}A_\mathrm{a}^2 A_\mathrm{c}\mathrm{e}^{-\mathrm{i}\Delta k_\mathrm{d} z} - \frac{\omega_\mathrm{d}}{2c}\mathrm{Im}[\chi^{(1)}]A_\mathrm{d} + \mathrm{i}\frac{\omega_\mathrm{d}}{2c}\mathrm{Re}[\chi^{(1)}]A_\mathrm{d} \qquad [16]$$

where the wavevector mismatch is given by:

$$\Delta k_\mathrm{d} = k_\mathrm{d} + k_\mathrm{c} - 2k_\mathrm{a} \qquad [17]$$

The wavevector mismatch will be zero on resonance for the three-level atom considered in this treatment. In fact the contribution to the refraction from all the other atomic levels must be taken into account whilst computing Δk_d and it is implicit that these make a finite contribution to the wavevector mismatch.

We can solve this first-order differential equation with the boundary condition $A_\mathrm{d}(z=0)=0$ to find the generated intensity $I(\omega_\mathrm{d})$ after a length z:

$$I(\omega_\mathrm{d}) = \frac{3n\omega_\mathrm{d}^2}{8Z_0c^2}|\chi^{(3)}|^2|A_\mathrm{a}|^4|A_\mathrm{c}|^2 \times \frac{\left\{1+\mathrm{e}^{-\frac{\omega_\mathrm{d}}{c}\mathrm{Im}[\chi^{(1)}]z} - 2\mathrm{e}^{-\frac{\omega_\mathrm{d}}{c}\mathrm{Im}[\chi^{(1)}]z}\cos\left(\Delta k + \frac{\omega_\mathrm{d}}{2c}\mathrm{Re}[\chi^{(1)}]\right)z\right\}}{\frac{\omega_\mathrm{d}^2}{4c^2}\mathrm{Im}[\chi^{(1)}]^2 + \left\{\Delta k + \frac{\omega_\mathrm{d}}{2c}\mathrm{Re}[\chi^{(1)}]\right\}^2}$$

where Z_0 is the impedance of free space. This expression is quantitatively correct for the case of the assumptions made. More generally the qualitative predictions and general scaling remain valid in the limit where the pulse duration is significantly longer than the time required to traverse the medium. Note that both real and imaginary parts of $\chi^{(1)}$ and $\chi^{(3)}$ play an important role, as we would expect for resonant frequency mixing. The influence of that part of the medium refraction which is due to other levels is contained in the terms with Δk. In the case of a completely transparent medium this becomes a major limit to the conversion efficiency.

Propagation and Wave-Mixing in a Doppler Broadened Medium

Doppler shifts arising from the Maxwellian velocity distribution of the atoms in the medium lead to a corresponding distribution in the detunings for

the various members of the atomic ensemble. The response of the medium, as characterized by the susceptibilities, must therefore include the Doppler effect by performing a weighted sum over possible detunings. The weighting is determined by the Gaussian form of the Maxwellian velocity distribution. From this operation the effective values of the susceptibilities at a given frequency are obtained, and these quantities can be used to calculate the generated field. This step is of considerable practical importance as in most up-conversion schemes it is not possible to achieve Doppler-free geometries and the use of laser cooled atoms, although in principle possible, limits the atomic density that can be employed. The interference effects persist in the dressed profiles providing the coupling laser Rabi frequency is comparable to or larger than the inhomogeneous width. This is because the Doppler profile follows a Gaussian distribution which falls off much faster in the wings of the profile than the Lorentzian profile due to the natural broadening.

In the case considered with weak probe field, excited state populations and coherences remain small. The two-photon transition need not be strongly driven (i.e., a small two-photon Rabi frequency can be used) but a strong coupling laser is required. The coupling laser must be intense enough that its Rabi frequency is comparable to or exceeds the inhomogenous widths in the system (i.e., Doppler width), and a laser intensity of above 1 MW cm^{-2} is required for a typical transition. This is trivially achieved even for unfocused pulsed lasers, but does present a serious limit to cw lasers unless a specific Doppler-free configuration is employed. The latter is not normally suitable for a frequency up-conversion scheme if a significant up-conversion factor is required, e.g., to the vacuum ultraviolet (VUV); however recent experiments report significant progress in cw frequency up-conversion using EIT and likewise a number of other possibilities, e.g., laser-cooled atoms and standing-wave fields, have been proposed.

A transform-limited single-mode laser pulse is essential for the coupling laser field since a multimode field will cause an additional dephasing effect on the coherence, resulting in a deterioration of the quality of the interference. In contrast, whilst it is advantageous for the field driving the two-photon transition to be single mode (in order to achieve optimal temporal and spectral overlap with the EIT hole induced by the dressing laser), this is not essential since this field does not need to drive the coherence responsible for interference.

When a pulsed laser field is used additional issues must be considered. The group velocity is modified for pulses propagating in the EIT large reductions, e.g., by factors down to $<c/100$, in the group velocity have been observed. Another consideration beyond that found in the simple steady-state case is that the medium can only become transparent if the pulse contains enough energy to dress all the atoms in the interaction volume. The minimum pulse energy to prepare a transparency is:

$$E_{\text{preparation}} = \frac{f_{13}}{f_{23}} NL\hbar\omega \qquad [18]$$

where f_{ij} are the oscillator strengths of the transitions and NL the product of the density and the length. Essentially the number of photons in the pulse must exceed the number of atoms in the interaction volume to ensure all atoms are in the appropriate dressed state. This puts additional constraints on the laser pulse parameters.

Up-conversion to the UV and vacuum UV has been enhanced by EIT in a number of experiments. Only pulsed fields have so far been up-converted to the VUV with EIT enhancement. The requirements on a minimum value of $\Omega_c > \Delta_{\text{Doppler}}$ constrains the conversion efficiency that can be achieved because the $1/\Omega_c^2$ factor in eqn [17] ultimately leads to diminished values of $\chi^{(3)}$. The use of gases of higher atomic weight at low temperatures is therefore highly desirable in any experiment utilizing EIT for enhancement of four-wave mixing to the VUV. Conversion efficiencies, defined in terms of the pulse energies by E_d/E_a or E_d/E_c of a few percent have been achieved using the EIT enhancement technique. It is typically most beneficial to maximize the conversion efficiency defined by the first ratio since ω_a is normally in the UV and is the lower energy of the two applied pulses.

Nonlinear Optics with a Pair of Strong Coupling Fields in Raman Resonance

An important extension of the EIT concept occurs when two strong fields are applied in Raman resonance between a pair of states in a three-level system. Considering the system illustrated in Figure 1 we can imagine that both applied fields are now strong. Under appropriate adiabatic conditions the system evolves to produce the maximum possible value for the coherence $\varrho_{12} = 0.5$. Adiabatic evolution into the maximally coherent state is achieved by adjusting either the Raman detuning or the pulse sequence (counter to-intuitive order). The pair of fields may also be in single-photon resonance with a

third level, in which case the EIT-like elimination of absorption will be important. This situation is equivalent to the formation of a darkstate, since neither of the two strong fields is absorbed by the medium. For sufficiently strong fields the single-photon condition need not be satisfied and a maximum coherence will still be achieved.

An additional field applied to the medium can participate in sum- or difference-frequency mixing with the two Raman resonant fields. The importance of the large value of coherence is that it is the source polarization that drives the new fields generated in the frequency mixing process. Complete conversion can occur over a short distance that greatly alleviates the constraints usually set by phase-matching in nonlinear optics. Recently near unity conversion efficiencies to the far-UV were reported in an atomic lead system where maximum coherence had been created. In a molecular medium large coherence between vibrational or rotational levels has also been achieved using adiabatic pulse pairs. Efficient multi-order Raman sideband generation has been observed to occur. This latter observation may lead the way to synthesizing very short duration light pulses since the broadband Raman sideband spectrum has been proved to be phase-locked.

Pulse Propagation and Nonlinear Optics for Weak CW Fields

In a Doppler-free medium a new regime can be accessed. This is shown in Figure 4 where the possibility now arises of extremely narrow transparency dips since very small values of Ω_c are now sufficient to induce EIT. The widths of these features are typically subnatural and are therefore accompanied by very steep normal dispersion, which corresponds to a much reduced group velocity. The ultraslow propagation of pulses is one of the most dramatic manifestations of EIT in this regime. Nonlinear susceptibilities are now very large as there is constructive interference controlling the value and the splitting of the two states is negligible compared to their natural width. Nonlinear optics at very low light levels, i.e., at the few-photon limit, is possible in this regime.

Propagation of pulses is significantly modified in the presence of EIT. Figure 4 shows the changes to $\mathrm{Re}[\chi^{(1)}]$ that arise. Within the transparency dip there exists a region of steep normal dispersion. In the vicinity of resonance this is almost linear and it becomes reasonable to consider the leading term only that describes the group velocity. An analysis of the

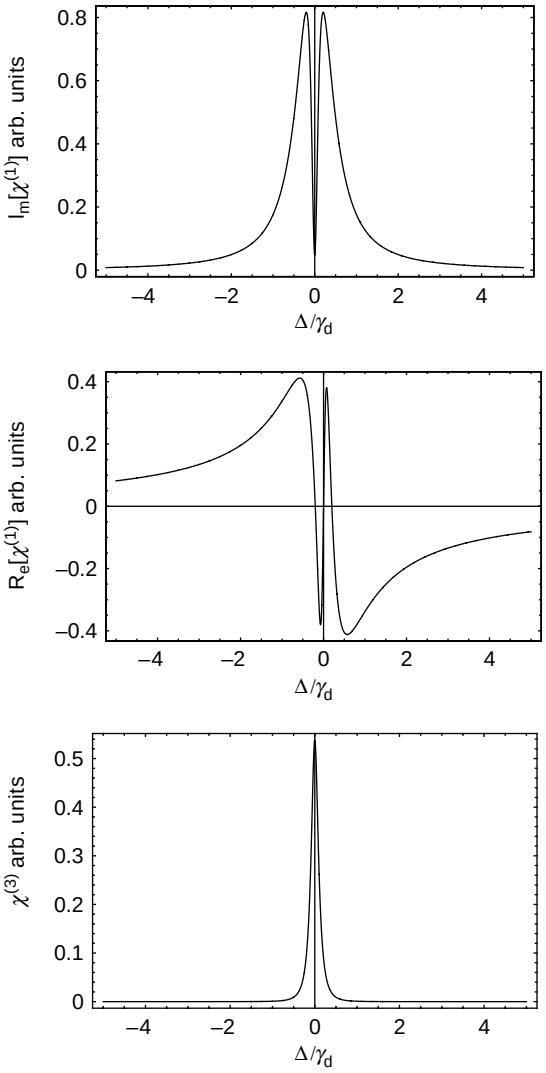

Figure 4 Electromagnetically induced transparency is shown in the case where there is no significant Doppler broadening. The values of the imaginary $\mathrm{Im}[\chi^{(1)}]$ and real $\mathrm{Re}[\chi^{(1)}]$ parts of the linear susceptibility and the nonlinear susceptibility $\chi^{(3)}$ are plotted in the three frames. We take $\Omega_c = \gamma_d/5$ and $\gamma_d = 50\gamma_a$, $\gamma_c = 0$.

refractive changes has been provided by Harris who expanded the susceptibilities (both real and imaginary parts) of the dressed atom around the resonance frequency to determine various terms in $\mathrm{Re}[\chi^{(1)}]$. The first term of the series (zero order) $\mathrm{Re}[\chi^{(1)}](\omega_{13}) = 0$ corresponds to the vanishing dispersion at resonance. The next term $\partial[\mathrm{Re}\chi^{(1)}](\omega)/\partial\omega$ gives the slope of the dispersion curve; at ω_{13} this takes the value:

$$\frac{\partial \mathrm{Re}\chi^{(1)}(\omega_{13})}{\partial\omega} = \frac{|\mu_{13}|^2 N 4(\Omega_c^2 - \gamma_a^2)}{2\pi\varepsilon_0(\Omega_c^2 + \gamma_a\gamma_d)} \qquad [19]$$

The slope of the dispersion profile leads to a reduced group velocity v_g:

$$\frac{1}{v_g} = \frac{1}{c} + \frac{\pi}{\lambda}\frac{\partial\chi^{(1)}}{\partial\omega} \qquad [20]$$

From the expression for $\partial\chi/\partial\omega$ we see that this slope is steepest (and so v_g mimimum) in the case where $\Omega_c \gg \Gamma_2$ and $\Omega_c^2 \gg \Gamma_2\,\Gamma_3$ but is still small compared to Γ_3 (i.e., $\Omega_C < \Gamma_3$) and hence $\partial\chi/\partial\omega \propto 1/\Omega_c^2$. In the limit of small Ω_c the following expression for v_g therefore holds:

$$v_g = \frac{\hbar c \varepsilon_0 [\Omega_c]^2}{2\omega_d |\mu_{13}|^2 N} \qquad [21]$$

Extremely low group velocities, down to a few meters per second, are achieved in this limit using excitation of the hyperfine split ground states in either laser cooled atomic samples or Doppler-free configurations in finite temperature samples. Recently similar light slowing has been observed in solids. Storage of the optical pulse within the medium has also been achieved by adiabatically switching off the coupling field and thus trapping the optical excitation as an excitation within the hyperfine ground states for which the storage time can be very long (>1 ms) due to very low dephasing rates. Since the storage scenario should be valid even for single photons this process has attracted considerable attention recently as a means to enable quantum information storage.

Extremely low values of Ω_c are sufficient to induce complete transparency (albeit in a very narrow dip) and at this location the nonlinear response is resonantly enhanced. Very high efficiency nonlinear frequency mixing and phase conjugation at low light levels have been reported under these conditions. It is expected that high-efficiency nonlinear optical processes will persist to ultralow intensities (the few photon level) in an EIT medium of this type.

One example of highly enhanced nonlinear interactions is the predicted large values of the Kerr type nonlinearity (nonlinear refractive index). The origin of this can be seen by considering the steep dispersion profile in the region of the transparency dip in Figure 4. Imagine that we apply an additional field ω_f, perhaps a very weak one, at a frequency close to resonance between state $|2\rangle$ and a fourth level $|4\rangle$. The ac Stark shift caused by this new field to the other three level, although small, will have a dramatic effect upon the value of the refractive index because of the extreme steepness of the dispersion profile. Essentially even a very small shift of the resonant wavelength causes a large change in the refractive index for a field applied close to this frequency. It is predicted that strong cross-phase modulations will be created in this process between the fields ω_f and ω_d, even in the quantum limit for these fields. This is predicted to lead to improved schemes for quantum nondemolition measurements of photons through the measurement of the phase-shifts they induce (through cross-phase modulation) on another quantum field. This type of measurement has direct application in quantum information processing.

See also

Nonlinear Optics, Applications: Phase Matching; Raman Lasers. **Scattering:** Raman Scattering.

Further Reading

Arimondo E (1996) Coherent population trapping in laser spectroscopy. *Progress in Optics* 35: 257–354.

Harris SE (1997) Electromagnetically induced transparency. *Physics Today* 50: 36–42.

Harris SE and Hau LV (1999) Non-linear optics at low light levels. *Physical Review Letters* 82: 4611–4614.

Harris SE, Field JE and Imamoglu A (1990) Nonlinear optical processes using electromagnetically induced transparency. *Physical Review Letters* 64: 1107–1110.

Hau LV, Harris SE, Dutton Z and Behroozi CH (1999) Light speed reduction to 17 metres per second in an ultra-cold atomic gas. *Nature* 397: 594–598.

Marangos JP (2001) Electromagnetically induced transparency. In: Bass M, *et al.* (ed.) *Handbook of Optics*, vol. IV, ch. 23. New York: McGraw-Hill.

Merriam AJ, Sharpe SJ, Xia H, *et al.* (1999) Efficient gas-phase generation of vacuum ultra-violet radiation. *Optics Letters* 24: 625–627.

Schmidt H and Imamoglu A (1996) Giant Kerr nonlinearities obtained by electromagnetically induced transparency. *Optics Letters* 21: 1936–1938.

Scully MO (1991) Enhancement of the index of refraction via quantum coherence. *Physical Review Letters* 67: 1855–1858.

Scully MO and Suhail Zubairy M (1997) Quantum Optics. Cambridge, UK: Cambridge University Press.

Sokolov AV, Walker DR, Yavuz DD, *et al.* (2001) Femtosecond light source for phase-controlled multi-photon ionisation. *Physical Review Letters* 87: 033402-1.

Zhang GZ, Hakuta K and Stoicheff BP (1993) Nonlinear optical generation using electromagnetically induced transparency in hydrogen. *Physical Review Letters* 71: 3099–3102.

ENVIRONMENTAL MEASUREMENTS

Contents

Doppler Lidar
Hyperspectral Remote Sensing of Land and the Atmosphere
Laser Detection of Atmospheric Gases
Optical Transmission and Scatter of the Atmosphere

Doppler Lidar

R M Hardesty, National Oceanic and Atmospheric Administration, Boulder, CO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Lidars developed for the measurement of atmospheric winds make use of the Doppler effect, in which radiation scattered or emitted from a moving object is shifted in frequency as a result of the movement of the particles. Atmospheric Doppler lidars irradiate a volume of atmosphere with a pulse of very narrowband, laser-produced radiation, then detect the change in frequency of the radiation backscattered from atmospheric aerosol particles or molecules present in the volume. Although Doppler lidars are conceptually similar to Doppler weather radars, lidar wavelengths are 3–4 orders of magnitude shorter than radar wavelengths, producing some important differences in propagation and scattering characteristics. Radiation at wavelengths in the optical region of the spectrum is efficiently scattered by aerosols and molecules, therefore Doppler lidars, unlike commonly used Doppler weather radars, do not require the presence of hydrometeors or insects to obtain useful results. Also, optical radiation can be tightly collimated, virtually eliminating effects of ground clutter and enabling lidar probing of small volumes to within a few meters of terrain or structures.

The primary disadvantage of lidar techniques is the severe attenuation of the optical radiation by cloud water droplets and fog. Doppler lidars do not typically probe beyond the edge of most atmospheric clouds – the one exception being tenuous ice clouds such as cirrus, which often are characterized by low optical extinction and high backscatter, making them excellent lidar targets. Doppler lidars also do not have the extended range, which can exceed 100 km, available from contemporary meteorological radars when scatterers are present.

In clear air, however, the characteristics and capabilities of Doppler lidar are well suited for obtaining detailed wind and turbulence observations for a wide variety of applications. Lidar beams can be easily scanned to characterize motions within very confined three-dimensional spaces such as shallow atmospheric boundary layers, narrow canyons, and turbulent structures. The relative compactness of Doppler lidars makes deployment on aircraft or other moving platforms extremely viable, and a Doppler lidar deployed on a satellite has been proposed as a way to obtain global measurements of atmospheric wind fields. By scanning a lidar beam from an orbiting satellite and analyzing the returns backscattered from clouds, aerosols, and molecules, a satellite-based instrument could provide important wind information for numerical forecast models.

Doppler Lidar Fundamentals

Backscattered Light

As laser light propagates through the atmosphere, some of the incident energy is scattered by the atmospheric molecules and constituents through which the light passes. In lidar applications, the light backscattered (scattered directly back toward the source) is collected and analyzed. The backscattering properties of an atmospheric particle depend on its refractive index, shape, and size. In the special case when the particle is much smaller than the wavelength of the incident radiation, as is the case for laser radiation scattered by atmospheric molecules, the scattering process is characterized as Rayleigh scattering. Because, in the Rayleigh scattering regime the energy backscattered by a particle increases proportionally as the inverse of the fourth power of the wavelength, Doppler lidar systems designed for molecular scatter operate at short wavelengths. Molecular Doppler lidar systems typically operate in the visible or ultraviolet spectral regions.

Aerosol particles, the other component of the atmosphere that scatters laser light, produce Mie scattering, which applies when the diameter of the scatterers is not orders of magnitude smaller than the incident wavelength. Aerosol particles present in the atmosphere include dust, soot, smoke, and pollen, as well as liquid water and ice. Although for Mie scattering the relationship between the power backscattered by an ensemble of aerosol particles and the incident wavelength is not simply characterized, most studies have shown that the energy backscattered increases roughly proportional to the first or second power of the inverse of the incident wavelength, depending on the characteristics of the scattering aerosol particles. In a polluted environment, such as in the vicinity of urban areas, an abundance of large particles often results in a roughly linear relationship between the inverse of the incident wavelength and the backscattered energy, while in more pristine environments, such as the free troposphere, the inverse wavelength/backscatter relationship can approach or exceed a square-law relationship.

The primary objective in Doppler lidar is to measure the Doppler frequency shift of the scattered radiation introduced by the movements of the scattering particles. Figure 1 shows a typical spectrum of the radiation collected at the lidar receiver for a volume of atmosphere irradiated by a monochromatic laser pulse. Molecular scattering produces the broadband distribution in Figure 1, where the broadening results from the Doppler shifts of the radiation backscattered from molecules moving at their thermal velocities. The width of the molecular velocity distribution in the atmosphere ranges from ~ 320 to 350 m s^{-1}, scaling as the square root of the temperature. In the center of the spectrum is a much narrower peak resulting from scattering of the light by aerosol particles. Because the thermal velocity of the much larger aerosol particles is very low, the width of the distribution of the aerosol return is determined by the range of velocities of particles moved about by small-scale turbulence within the scattering volume. This is typically on the order of a few m s^{-1}. Also shown in Figure 1 is an additional broadband distribution, due to scattered solar radiation collected at the receiver. If the laser source is not monochromatic, the backscattered signal spectrum is additionally broadened, with the resulting spectrum being the convolution of the spectrum shown in Figure 1 with the spectrum of the laser pulse. As seen in the figure, the entire spectrum is Doppler-shifted in frequency, relative to the frequency of the laser pulse. The object for a Doppler lidar system is to measure this shift, given by $\delta f = 2v_{\mathrm{rad}}/\lambda$, where v_{rad} is the component of the mean velocity of the particles in the direction of propagation of the lidar pulse and λ is the laser wavelength.

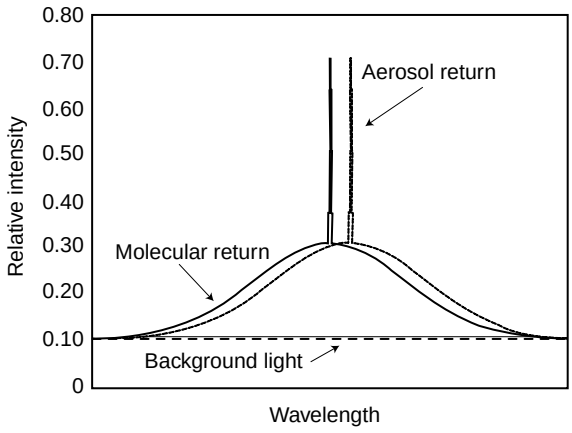

Figure 1 Distribution of wind velocities in a lidar return, showing Doppler shift resulting from wind-induced translation of the scatterers. The ratio of aerosol to molecular signals increases with increasing wavelength.

Elements of a Doppler Lidar

Doppler lidar systems can be designed primarily to measure winds from aerosol-scattered radiation, molecule-scattered radiation, or both. The type of system places specific requirements on the primary components that comprise a Doppler lidar system. A Doppler lidar consists of a laser transmitter to produce pulses of energy that irradiate the atmospheric volume of interest, a receiver which collects the backscattered photons and estimates the energy and Doppler shift of the return, and a beam pointing mechanism for directing the laser beam and receiver field of view together in various directions to probe different atmospheric volumes. Independent of whether the primary scatterers are molecules or aerosol particles, the system design criteria for a Doppler system are driven by a fundamental relationship between the error in the estimate of mean frequency shift δf_1, the bandwidth of the return f_2 (proportional to the distribution of velocities), and the number of incident backscattered photons detected N as

$$\delta f_1 \propto f_2/N^{1/2} \qquad [1]$$

Thus, the precision of the velocity estimate is improved by increasing the number of detected signal photons and/or decreasing the bandwidth of the backscattered signal. It is obvious from the equation that a significantly greater number of photons are required to achieve the same precision in a Doppler measurement from a molecular backscattered signal,

characterized by higher bandwidth f_2, compared to the number required to compute velocities from aerosol returns. The improved measurement precision gained from a narrow bandwidth return also indicates that the laser transmitter in a Doppler lidar system should be characterized by narrow spectral width (considerably narrower than the spectral broadening caused by the distribution of velocities), and to increase the number of photons detected, maximum transmit energy.

Narrowband, single-frequency pulses are obtained by employing a low-power, precisely frequency-stabilized master oscillator laser, whose radiation is either used to seed a stabilized, higher-energy laser cavity (injection seeding) or amplified in one or more optical amplifiers (master-oscillator, power amplifier, (MOPA)) to produce frequency-stabilized pulses.

The lidar receiver gathers the backscattered photons and extracts the wind velocity as a function of the range to the scattering volume. This requires a telescope, to gather and focus the scattered radiation, and a receiver to detect the scattered radiation and determine the Doppler shift. Frequency analysis in a Doppler lidar receiver is carried out using one of two techniques: coherent detection (also known as heterodyne detection) or direct detection (alternately labeled incoherent detection). The techniques differ fundamentally. In a heterodyne receiver, frequency analysis is carried out on a digitized time series created by mixing laser radiation with the backscattered radiation, while in direct detection lidar an interferometer optically analyzes the backscattered radiation to produce a light pattern which contains the information on the frequency content.

Coherent (Heterodyne) Doppler Lidar

Description

Coherent or heterodyne lidar is implemented by optically mixing the backscattered laser light with radiation from a stable, continuous-wave, local oscillator (LO) laser whose frequency is precisely controlled to be a known frequency offset, typically on the order of tens of MHz, from that of the laser transmitter (Figure 2). The mixing process at the face of an optical detector generates, after high pass filtering, an electrical signal with amplitude proportional to amplitude of the backscattered electromagnetic field and frequency equal to the difference between the backscattered field frequency and the LO laser field frequency. This signal is sampled, and then digitally processed to estimate the range-dependent mean frequency shift of the backscattered signal, from which the radial wind component can be

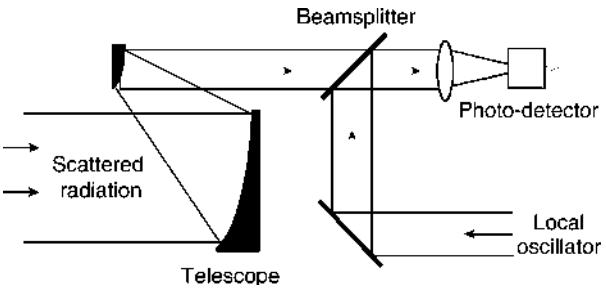

Figure 2 Schematic of coherent detection of backscattered radiation.

derived. A property of coherent lidar is that, because of constructive and destructive interference of the returns from individual scatterers, single pulse returns are characterized by random fluctuations in the amplitude and phase of the detected time series and Fourier spectrum. Consequently, averaging (sometimes referred to as accumulation) over multiple pulses usually is necessary to produce a sufficiently precise estimate of the signal mean frequency estimate. Because the phase of the detected time series is random within each pulse, averaging must be carried out in the power spectrum or autocorrelation function domain, rather than on the detected time series.

The optical mixing process in coherent Doppler lidar provides both benefits and design challenges. Because receiver efficiencies (equivalent fraction of incident photons entering the receiver that are converted to electrons) are quite high, wind velocities can be estimated from very weak backscattered signals. Also, signal processing is performed on a time series derived from the mixed signal, enabling the use of electrical filters to produce a narrow receiver bandwidth and effectively eliminating broadband solar background light as a major source of noise. As a result, unlike many lidars coherent Doppler lidar performance is not degraded under daylight conditions.

The primary limitation of coherent detection results from the added photon noise from the local oscillator radiation (which is usually much stronger than the backscattered radiation). The effect of local oscillator photon noise is to define a system noise level below which very weak signals cannot be practically extracted, even with substantial multiple-pulse averaging. In addition, because efficient mixing requires phase coherence of the backscattered signal field across the detector, coherent lidar performance at longer ranges is degraded by strong optical turbulence along the path of the laser pulse. These adverse effects of turbulence increase for shorter wavelength systems.

Although measurements at wavelengths near 1.06 μm have been demonstrated, coherent Doppler lidar systems used for regular atmospheric probing operate in the eye-safe, infrared portion of the spectrum at wavelengths longer than 1.5 μm. Currently, the two most common system wavelengths for coherent lidar wind systems are in atmospheric window spectral regions around 2 and 10.6 μm. Early coherent Doppler lidar measurements, beginning in the 1970s, employed CO_2 laser transmitters and local oscillators operating at wavelengths near 10.6 μm. Pulsed CO_2 laser transmitters with as much as 10 J of energy have since been demonstrated, and systems with 1 J lasers have probed the atmosphere to ranges of 30 km or more. In the late 1980s, solid state laser transmitters operating near 2 μm wavelengths were introduced into coherent lidar wind-measuring systems. The compact size and potential reliability advantages of solid-state transmitters, in which the transmitter laser is optically pumped by an array of laser diodes, provide advantages over larger CO_2 laser technology. Also, because for a given measurement accuracy the range-resolution obtainable is proportional to wavelength, 2 μm instruments have an enhanced capability to probe small-scale features. However, although development of higher energy, single-frequency coherent lidar sources operating at 2 μm region, as well as at 1.5 μm, is currently an active research area, solid state lasers with pulse energies greater than several tens of mJ have yet to be incorporated into lidar systems for atmospheric probing.

Applications of Coherent Doppler Lidar

Coherent lidars have been used to measure winds for a variety of applications, and from an assortment of platforms, such as ships and aircraft. Since these lidars operate in the infrared where aerosol scattering dominates molecular scattering, they require aerosol particles to be present at some level to obtain usable returns. Although clouds also provide excellent lidar targets, most of the more useful applications of coherent lidars involve probing the atmospheric boundary layer or lower troposphere where aerosol content is highest. Because of the capability to scan the narrow lidar beam directly adjacent to terrain, a unique application of lidar probing is the measurement of wind structure and evolution in complex terrain such as mountains and valleys. Over the past two decades, Doppler lidar studies have been used, for example, to study structure of damaging windstorms on the downwind side of mountain ranges, advection of pollution by drainage flows from valleys, and formation of mountain leeside turbulence as a potential hazard to landing aircraft. The interactions of wind flows with complex terrain can produce dangerous conditions for transportation, especially aircraft operations. Figure 3 shows a strong mountain wave measured by a Doppler lidar near Colorado Springs, during an investigation of the effects of downslope winds and turbulence on aircraft operations at the Colorado Springs airport. The mountains are on the right of the figure, where the strong wave with wavelength of about 6 km can be seen. Note the reversal in winds near the location where the wave descends to the surface. The Colorado Springs study was prompted by the crash of a passenger jet while landing at the airport during a period when the winds were down the slope of the Rocky Mountains just to the west of the airport.

A Doppler lidar was recently deployed in an operational mode for wind shear detection at the new Hong Kong International Airport. Because winds flowing over mountainous terrain just south of the

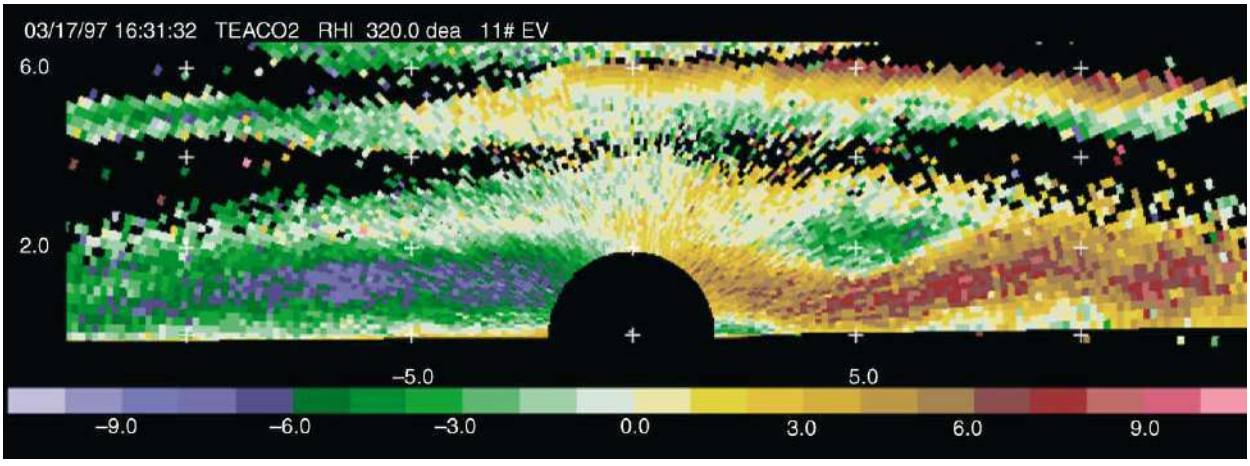

Figure 3 Vertical scan of the winds in a mountain wave measured by a 10.6 μm coherent Doppler lidar near Colorado Springs, Colorado. Courtesy L. Darby, NOAA.

airport can produce mountain waves and channeled flows, wind shear is frequently encountered during landing and takeoff operations. Figure 4 shows a gust front situated just to the west of the airport, characterized by a sharp wind change (which would produce a corresponding change in airspeed) approaching 13 m s^{-1} over a distance of less than one kilometer. Providing a capability to detect such events near the airport and warn pilots during approach or preparation for takeoff was the primary reason for deployment of the lidar. Within the first year of operation the Hong Kong lidar was credited with improving wind shear detection rates and providing more timely warnings for pilots of the potentially hazardous conditions. As of August 2003 this lidar had logged more than 10 000 hours of operation.

Because coherent Doppler lidars are well matched to applications associated with probing small-scale, turbulent phenomena, they have also been deployed at airports to detect and track wing tip vortices generated by arriving or departing aircraft on nearby parallel runways. In the future, a network of ground-based lidars could provide information on vortex location and advection speed as well as wind shear, leading to a potential decrease in congestion at major airports. Also, compact lidar systems looking directly ahead of research aircraft have shown the capability to detect wind changes associated with potentially hazardous clear-air turbulence. In the future, a Doppler lidar installed on the commercial aircraft fleet could potentially look ahead and provide a warning to passengers to fasten seat belts before severe turbulence is encountered.

The high resolution obtainable in a scanning lidar enables visualization of finely structured wind and turbulence layers. Figure 5 shows an image of turbulence associated with a nocturnal low-level jet just 50 m above the surface obtained at night over flat

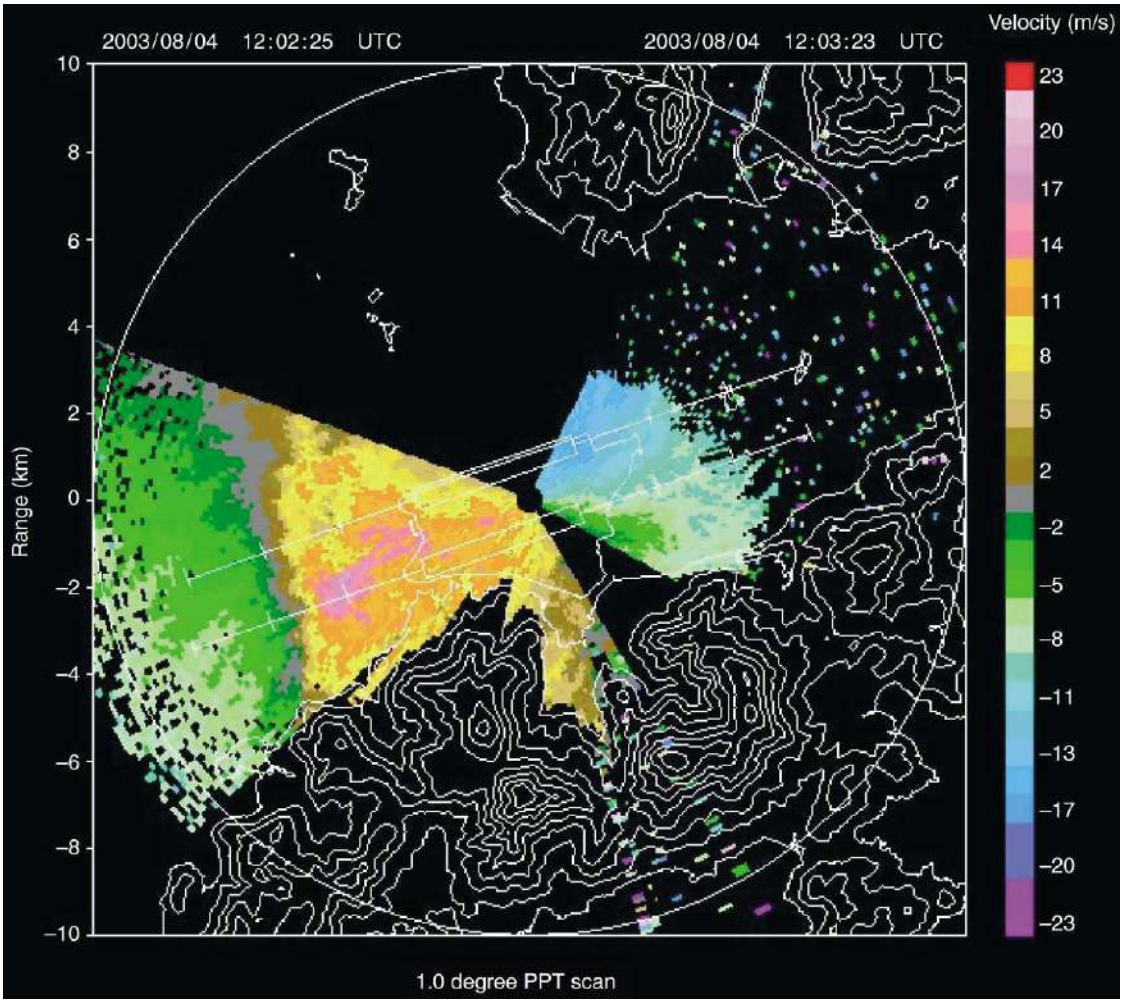

Figure 4 Horizontal depiction of the radial component of the wind field measured at Hong Kong Airport by a 2.02 μm Doppler lidar during a gust front passage, showing strong horizontal shear just west of the airport. Courtesy C. M. Shun, Hong Kong Observatory.

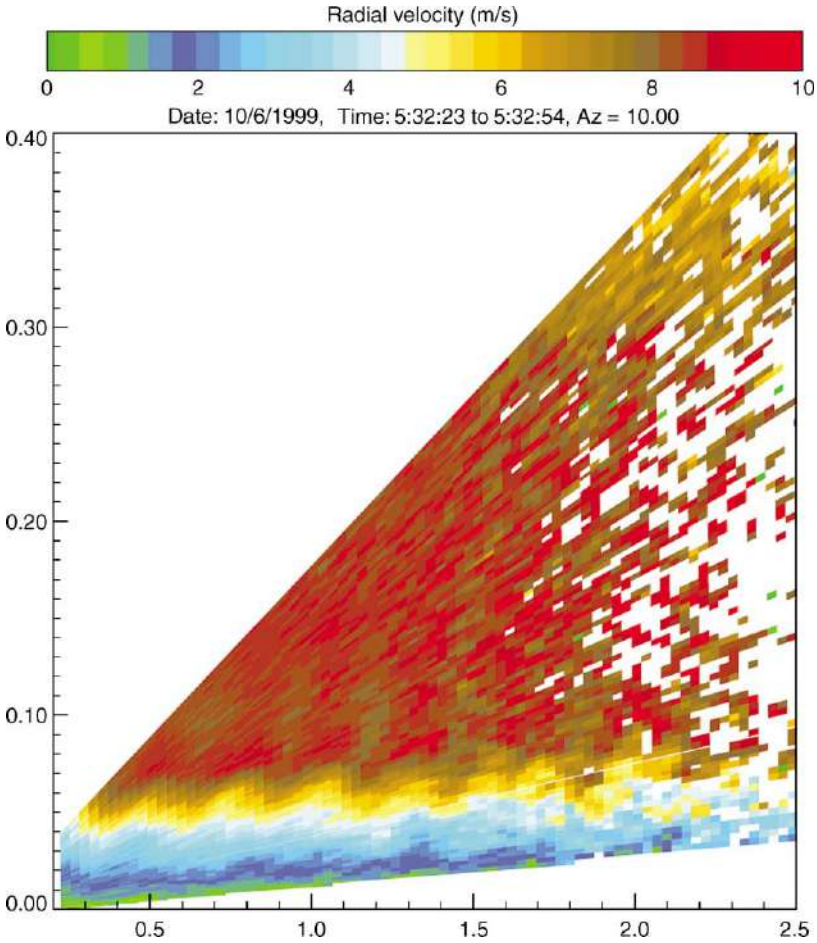

Figure 5 Turbulence structure within a nocturnal stable layer at a height of 50 m. Note the stretching of the vertical scale. Courtesy R. Banta, NOAA.

terrain in Kansas. Although a low-level jet was present on just about every evening during the experiment, similar images obtained at different times by the Doppler lidar illustrated markedly different characteristics, such as a wide variation in the observed mechanical turbulence along the interface. Such observations enable researchers to improve parameterizations of turbulence in models and better understand the conditions associated with vertical turbulent transport and mixing of atmospheric constituents such as pollutants.

By deploying Doppler lidars on moving platforms such as ships and aircraft, the spatial coverage of the measurements can be greatly increased. Although removal of platform motion and changes in orientation are not trivial, aircraft-mounted lidar can map out such features as the low-level jet and the structure of boundary layer convective plumes. Figure 6 shows horizontal and vertical motions associated with convective plumes measured along an approximately 60 km path over the southern Great Plains of the United States. From such measurements estimates of turbulence intensity, integral scale, and other boundary layer characteristics can be computed.

Direct Detection Doppler Lidar

Description

Direct detection or incoherent Doppler lidar has received significant attention in recent years as an alternative to coherent lidar for atmospheric wind measurements. In contrast to coherent lidar, in which an electrical signal is processed to estimate Doppler shift, an optical interferometer, usually a Fabry–Perot etalon, serves as the principal element in a direct detection lidar receiver for determining the frequency shift of the backscattered radiation.

One implementation of a direct-detection Doppler lidar receiver is the 'fringe imaging' technique. In this design, the interferometer acts as a spectrum analyzer. The backscatter radiation is directed through a

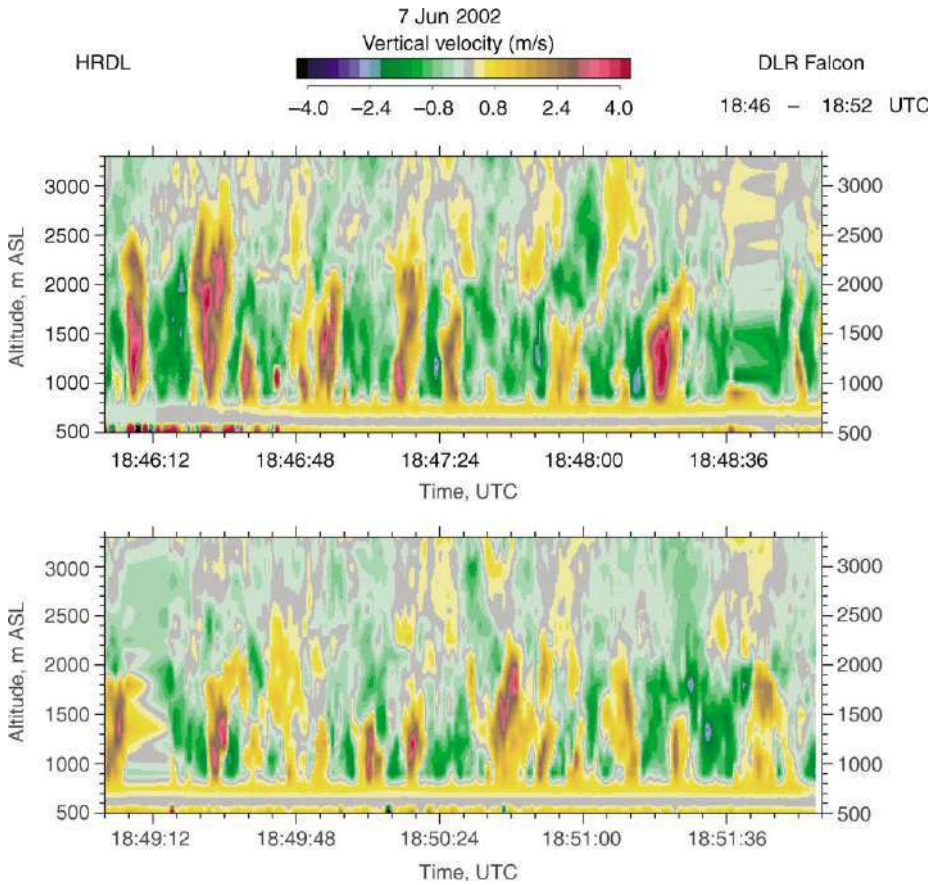

Figure 6 Vertical motion in convective plumes measured by a vertically pointing airborne coherent Doppler lidar. Total horizontal extent of the measurements is approximately 60 km, reddish colors correspond to upward motions.

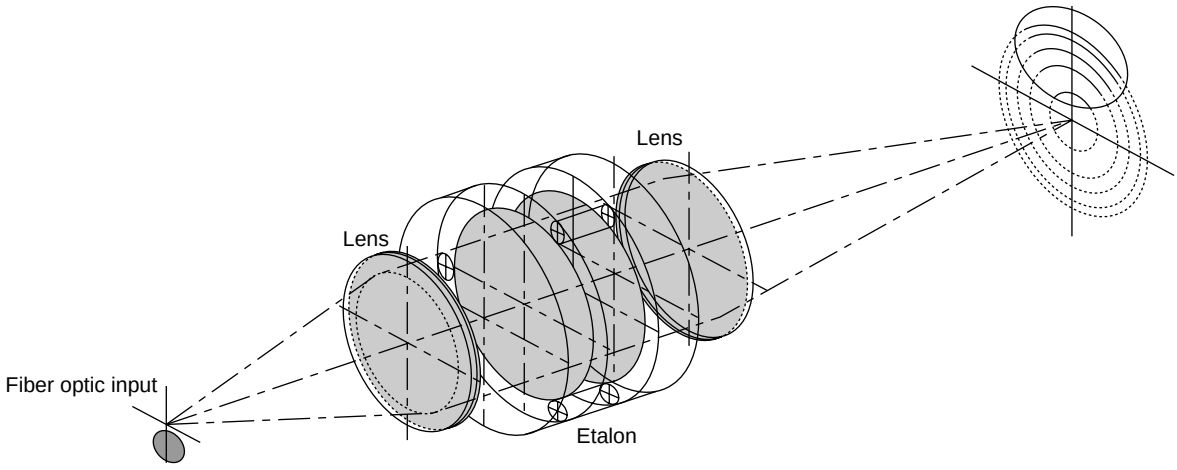

Figure 7 Schematic of Fabry–Perot etalon in a direct detection, fringe-imaging lidar receiver. Courtesy P. Hays, Michigan Aerospace Corporation.

Fabry–Perot interferometer, which produces a ring pattern in the focal plane (Figure 7). The spectral content information of the incident radiation is contained in the radial distribution of the light. Each ring corresponds to an order of the interferometer and is equivalent to a representation of the backscattered signal frequency spectrum. As the mean frequency of the backscattered radiation changes, the rings move inward or outward from the center. To extract a spectrum of the backscattered light, one or

more of the rings are imaged onto a two-dimensional detector, and the resulting pattern analyzed. The rings can either be detected using a multi-element ring detector, or can be converted to a linear pattern with a circle-to-line converter optic and then imaged on to a charge-coupled device array device.

An alternate direct detection receiver configuration, generally called the double edge technique, is to employ two interferometers as bandpass filters, with the center wavelength of each filter set above and below the laser transmitter wavelength, as shown in Figure 8. The incoming radiation is split between the two interferometers, and the wavelength shift is computed by examining the ratio of the radiation transmitted by each interferometer. Both the double edge and fringe-imaging techniques have demonstrated wind measurements to heights well into the stratosphere. The major challenge associated with the double edge receiver is optimizing the instrument when both aerosol and molecular scattered radiation are present, since in general the change in transmission as a function of velocity is different for the aerosol and molecular signals. Double edge receivers optimized for both aerosol and molecular returns place the bandpass of the etalon filters at the precise wavelength where the change in transmission with change in Doppler shift is the same for both aerosol and molecular returns.

For both types of direct detection receivers described above, much of the radiation incident on the interferometer is reflected out of the system, reducing the overall efficiency of the receiver. Recently, designs that incorporate fiber optics to collect a portion of the reflected radiation and 'recycle' it back into the etalon have been demonstrated as a method to improve the transmission efficiency of the etalon in a fringe imaging receiver.

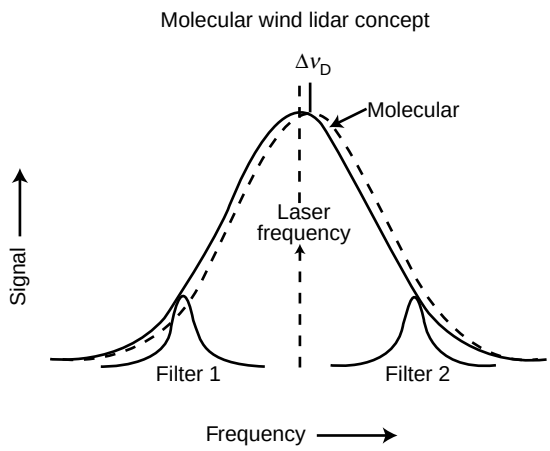

Figure 8 Spectrum of molecular lidar return showing placement of bandpass filters for a dual channel (double edge) direct detection receiver. Courtesy B. Gentry, NASA.

Doppler Wind Measurements Based on Molecular Scatter

One of the primary advantages of direct detection Doppler lidar is its capability for measurements based on scatter from atmospheric molecules. Measurement of Doppler shifts from molecular scattered radiation is challenging because of the large Doppler-broadened bandwidth of the return. Because one wants to measure a mean wind velocity with a precision of a few m s^{-1} or better, in the order of 10^5 photons are required. Some combination of multiple pulse averaging, powerful lasers, and large receiver optics is required to obtain these high photon counts from backscattered returns.

Molecular-scatter wind measurements have been demonstrated in the visible spectral region at 532 nm wavelength as well as at 355 nm in the near ultraviolet. The ultraviolet region has the dual advantages of enhanced molecular scatter and less restrictive laser eye-safety restrictions. Figure 9 shows the time series of a wind profile measured in the troposphere from both aerosols and clouds using a molecular-scatter, ground-based 355-nm wavelength Doppler fringe-imaging lidar. The figure shows measurements from receiver channels optimized for the wideband molecular signal and the narrowband aerosol return. For this measurement, backscattered photons were collected by a 0.5 m receiver aperture, averaged for 1 minute, and processed. In the absence of clouds, direct detection Doppler lidars have measured wind profiles continuously from the surface to beyond 15 km height. Figure 10 shows that the estimated wind error for the same 355 nm lidar is less than 1 m s^{-1} to about 10 km, and less than 4 m s^{-1} at 15 km height.

Heterodyne and Direct-Detection Doppler Trade-Offs

Lively debates within the lidar community have occurred over the past decade regarding the relative merits of heterodyne versus direct detection Doppler lidars. To a large extent, the instruments are complementary. Generally, heterodyne instruments are much more sensitive when significant aerosols are present. Coherent lidar processing techniques have been developed that can produce accurate wind measurement rates using only a few lidar pulses with as few as 100 detected photons, such that several wind observations per second can be obtained for nominal pulse rates. This inherent sensitivity has led to numerous applications in which a lidar beam has been scanned rapidly over a large volume to obtain time-varying, three-dimensional wind measurements.

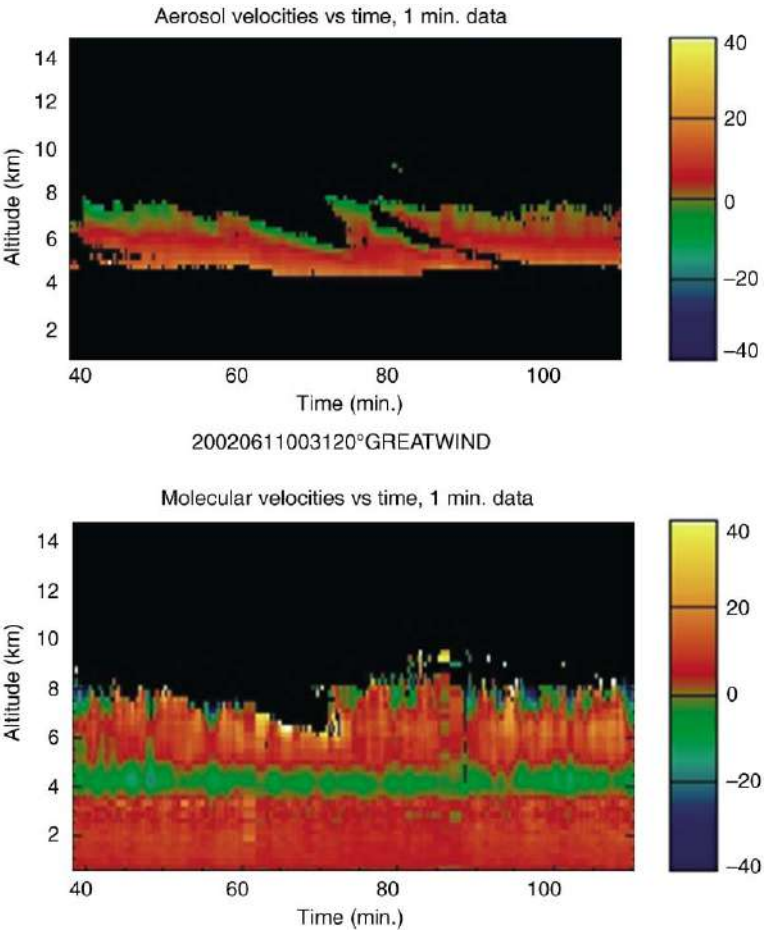

Figure 9 Time series of radial wind speed profiles measured by a 355 nm fringe-imaging direct detection lidar aerosol channel (top) and molecular channel (bottom) at Maura Loa Observatory, HI. Change in wind speed and blocking of the return by clouds is clearly seen. Courtesy C. Nardell, Michigan Aerospace Corp.

The primary advantage of direct detection instruments is their demonstrated capability to measure winds from molecular-backscattered returns in the middle and upper troposphere. In very pristine air, direct detection offers the only method for long-range wind measurements, even though significant averaging may be required. Direct-detection lidars have the additional advantage of being unaffected by atmospheric refractive turbulence. Unlike heterodyne lidars, which require a very pure laser pulse and a diffraction-limited receiver field of view matched to and precisely aligned with the transmitted beam, direct detection systems can have a wider bandwidth transmitter and a receiver field of view several times diffraction-limited. In direct detection lidar design, the field of view is usually constrained by the need to limit background light during daytime operation. A major design challenge for direct detection instruments is holding the Fabry–Perot etalon plate spacing stable over a range of temperatures and in high vibration environments.

Global Wind Measurements

A satellite-based Doppler lidar has frequently been proposed as a way to measure wind fields over most of the Earth. At present, winds are the one major meteorological variable not well-measured from orbiting platforms. Measurement of winds is especially important over regions of the Earth that are not currently well sampled, such as over Northern Hemisphere oceans, as well as over most of the tropics and Southern Hemisphere. Wind profile information is currently obtained from radiosondes and by tracking cloud and water vapor inhomogeneities using satellite imagers. Doppler lidar wind measurements would greatly augment the current data set by providing wind estimates throughout the troposphere under clear conditions, and highly height-resolved observations down to cloud tops when cloud decks are present. Observing system simulation experiments conducted in recent years indicate that satellite-based lidar global wind

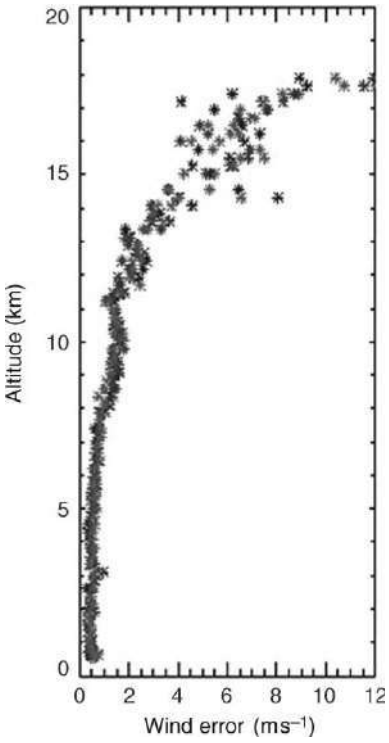

Figure 10 Estimated error in the radial velocity estimate versus altitude for a 355 nm direct detection, fringe imaging Doppler lidar system at Mauna Loa observatory on Hawaii. The lidar provides data for assimilation into local mesoscale forecast models. Courtesy C. Nardell, Michigan Aerospace Corp.

measurements could lead to a significant improvement in long-term forecast skill, provided the wind fields can be observed with sufficient accuracy and spatial resolution.

In a Doppler lidar wind mission, a satellite carrying a lidar system would orbit the Earth in a nearly polar orbit. The pulsed laser beam would be scanned conically about the nadir to obtain different components of the wind velocity. The scanning could be either continuous or 'stop and stare'. After sufficient returns are averaged at a given pointing angle to produce an acceptable estimate, the radial component of the velocity would be computed and assimilated directly into numerical analysis and forecast models.

Doppler lidar measurement of winds from space is theoretically feasible but technologically difficult. Depending on the orbital height, the scattering volume is anywhere from 450 to ~850 km from the satellite, which challenges the sensitivity of current system types. Because weight and power consumption are critical parameters for space systems, telescope diameter and laser power cannot be easily increased to obtain the necessary sensitivity. Similarly, the ability to average returns from multiple pulses is also limited by time limitations. Because a satellite moves at about 7 km s^{-1}, in order to obtain measurements over a horizontal distance of 300 km (the resolution of the radiosonde network) only about 45 seconds are available to make enough observations from multiple look angles to obtain a useful measurement. It should also be noted that, as a result of the high orbital velocity of the satellite, precise knowledge of beam pointing is extremely critical for measurements from satellites. For a lidar operating at a nadir angle of 45 degrees, an error in the knowledge of pointing angle of just 1 mrad results in an error of about 5 ms^{-1} in the measured radial component of the wind.

Despite the challenge of employing a satellite-based Doppler lidar, efforts are continuing to develop the appropriate technology and to assess the impact of the observations. The European Space Agency is planning a Doppler wind lidar demonstration mission for the late 2000s that would incorporate a nonscanning, direct-detection instrument with both aerosol and molecular channels. Doppler lidar technology research and numerical simulations aimed at satellite-based wind sensing is ongoing at several research centers within the United States, Europe, and Japan. One option currently being studied is a 'hybrid' lidar system that would combine a direct detection lidar for measurement of winds in the clear free troposphere with a low-energy coherent system that would obtain its observations from clouds and the aerosol-rich boundary. Although a hybrid instrument reduces some requirements on system power and aperture size, combining the two techniques on a single platform will likely require innovative engineering and new approaches to beam combining and scanning.

See also

Environmental Measurements: Optical Transmission and Scatter of the Atmosphere. **Scattering:** Scattering Theory.

Further Reading

Baker WE, Emmitt GD, Robertson F, *et al.* (1998) Lidar-measured wind from space: a key component for weather and climate prediction. *Bulletin of the American Meteorological Society* 79: 581–599.

Banta RM, Olivier LD and Levinson DH (1993) Evolution of the Monterey Bay sea breeze layer as observed by pulsed Doppler lidar. *Journal of Atmospheric Science* 50 (24): 3959–3982.

Banta RM, Olivier LD, Gudiksen PH and Lange R (1996) Implications of small-scale flow features to modeling dispersion over complex terrain. *Journal of Applied Meteorology* 35: 330–342.

Banta RM, Darby LS, Kaufmann P, Levinson DH and Zhu C-J (1999) Wind flow patterns in the Grand Canyon as revealed by Doppler lidar. *Journal of Applied Meteorology* 38: 1069–1083.

Frehlich R (1996) Coherent Doppler lidar measurements of winds in the weak signal regime, 1997. *Applied Optics* 36: 3491–3499.

Gentry B, Chen H and Li SX (2000) Wind measurements with a 355 nm molecular Doppler lidar. *Optics Letters* 25: 1231–1233.

Grund CJ, Banta RM, George JL, *et al.* (2001) High-resolution Doppler lidar for boundary-layer and cloud research. *Journal of Atmosphere and Oceanic Technology* 18: 376–393.

Huffaker RM and Hardesty RM (1996) Remote sensing of atmospheric wind velocities using solid state and CO_2 coherent laser systems. *Proceedings of the IEEE* 84: 181–204.

Menzies RT and Hardesty RM (1989) Coherent Doppler lidar for measurements of wind fields. *Proceedings of the IEEE* 77: 449–462.

Post MJ and Cupp RE (1990) Optimizing a pulsed Doppler lidar. *Applied Optics* 29: 4145–4158.

Rees D and McDermid IS (1990) Doppler lidar atmospheric wind sensor: reevaluation of a 355 nm incoherent Doppler lidar. *Applied Optics* 29: 4133–4144.

Rothermel J, Cutten DR, Hardesty RM, *et al.* (1998) The multi-center airborne coherent lidar atmospheric wind sensor. *Bulletin of the American Meteorological Society* 79: 581–599.

Skinner WR and Hays PB (1994) Incoherent Doppler lidar for measurement of atmospheric winds. *Proceedings of SPIE* 2216: 383–394.

Souprayen C, Garnier A and Hertzog A (1999) Rayleigh–Mie Doppler wind lidar for atmospheric measurements. II: Mie scattering effect, theory, and calibration. *Applied Optics* 38: 2422–2431.

Souprayen C, Garnier A, Hertzog A, Hauchecorne A and Porteneuve J (1999) Rayleigh–Mie Doppler wind lidar for atmospheric measurements. I: Instrumental setup, validation, first climatological results. *Applied Optics* 38: 2410–2421.

Hyperspectral Remote Sensing of Land and the Atmosphere

W H Farrand, Space Science Institute, Boulder, CO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Hyperspectral remote sensing is a true marriage of imaging technology with spectroscopy. Hyperspectral remote sensing systems (also known as imaging spectrometer systems) fully sample the optical wavelength range of interest, whether it be the reflected solar range (0.35 to 2.5 μm) or the range of thermal emission from the Earth's surface (3.0 to 14 μm). A hyperspectral sensor views the Earth's surface in a series of contiguous and spectrally narrow image bands. Figure 1 presents the concept of hyperspectral remote sensing, wherein each spatial element of an image has an associated full resolution spectrum. The calibrated reflectance or emittance spectra collected by a hyperspectral system are meant to be directly comparable with that of materials measured in the laboratory. With such high spectral resolution, it thus becomes possible to do reflectance or emittance spectroscopy of the Earth's surface from an overhead perspective.

Underlying Principles

Optical remote sensing is done over wavelength intervals, or windows, in which the atmosphere is largely transparent. For the reflective solar portion of the spectrum, the atmosphere is mostly transparent from approximately 0.35 to 2.5 μm. Water, CO_2, and other gases have absorptions of varying strength in this range. This window is subdivided into the range of human vision (the visible) and the region of slightly longer wavelengths known as the near infrared; together they are the visible/near infrared or VNIR and extend from approximately 0.35 to 1.0 μm. The region from approximately 1.0 to 2.5 μm is known as the short wavelength infrared or SWIR. In the range of emitted terrestrial radiation there are two window regions. The first extends from approximately 3 to 5 μm and is known as the medium wavelength infrared or MWIR, and the second, the long wavelength infrared or LWIR, extends from approximately 8 to 14 μm.

The light that is intercepted by the entrance aperture of a sensor is known by the quantity of radiance which is measured in units of microwatts per square centimeter per nanometer per unit of solid angle. Physical quantities related to the material properties of Earth surface materials are reflectance and emittance. Both properties are the result of ratios and so are unitless. One definition of reflectance is that it is the ratio of radiance reflected from a material, divided by the radiance reflected from an identically illuminated perfectly diffuse reflector. Likewise, a definition for emittance is the ratio of radiance emitted from a material, divided by the ratio

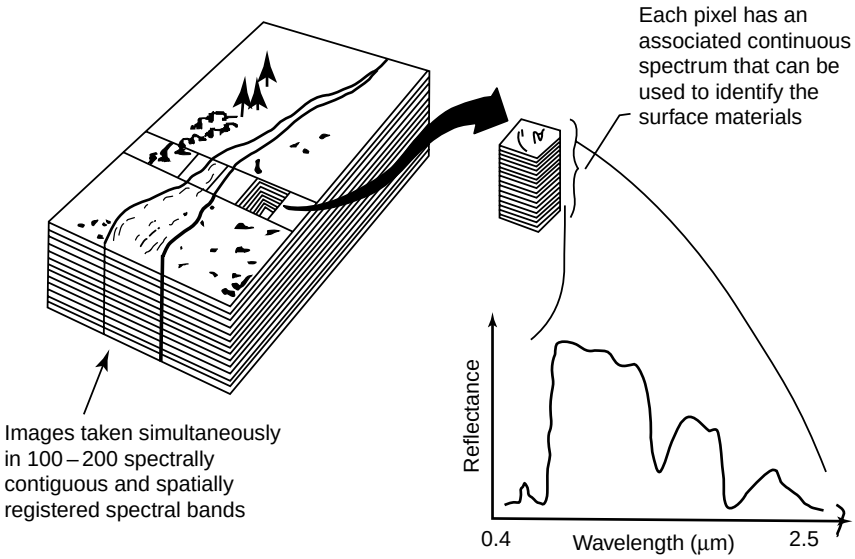

Figure 1 Hyperspectral remote sensing concept.

of radiance emitted from a perfectly emitting material of the same temperature.

Materials covering the Earth's surface, or gases in the atmosphere, can be identified in hyperspectral data on the basis of absorption features, or bands, in the spectra recorded by the sensor. Absorption bands result from the preferential absorption of energy in some wavelength interval. The principal types of processes by which light can be absorbed include, in order of decreasing energy of the process, electronic charge transfers, electronic crystal field effects, and molecular vibrations. The former two processes are observed in minerals and manmade materials that contain transition group metal cations, most usually, iron. Charge transfer absorptions are the result of cation-to-anion or cation-to-cation electron transfers. Crystal field absorptions are explainable by the quantum mechanical theory of atomic structure, wherein an atom's electrons are contained in orbital shells. The transition group metals have incompletely filled *d* orbital shells. When a transition metal cation, such as iron, is surrounded by anions, and potentially selects other cations, an electric field, known as the crystal field, exists. Crystal field absorptions occur when radiant energy causes that cation's orbital shell energy levels to be split by interaction with the crystal field. The reflectance spectra of different iron-bearing minerals are unique, because the iron cation has a unique positioning with respect to the anions and other cations that compose the mineral.

A material's component molecules have bonds to other molecules and as a result of interaction with radiant energy, these bonds can stretch or bend. These molecular vibrations, and overtones of those vibrations, are less energetic processes than the electronic absorption processes described above and so the resulting absorption features occur at longer wavelengths. Fundamental vibrational features occur in the MWIR and LWIR. Overtones and combination overtones of these vibrations are manifested as weaker absorption features in the SWIR. These overtone features in the SWIR include absorptions diagnostic of carbonate and certain clay minerals. Absorption features caused by vibrational overtones of C–H, O–H, and N–H stretches are also manifested in the SWIR and these absorption features are characteristic of important vegetative biochemical components such as cellulose, lignin, starch, and glucose. An example of the effect of the changes in reflected energy recorded at a sensor, due to the absorption of incident energy by green vegetation and other surface materials, is provided in Figure 2.

Laboratory spectrometers have been used for many years to analyze the diagnostic absorptions of materials caused by the above effects. More recently, spectrometers were mounted on telescopes to determine the composition of the Moon and the other solid bodies in our solar system. Technology progressed to where profiling spectrometers (instruments that measure a successive line of points on the surface) could be mounted on aircraft. The next logical step was the construction of imaging spectrometers (hyperspectral sensors) that measured spectra in two spatial dimensions. In fact, the data produced by a hyperspectral sensor are often thought of as an image cube (Figure 3) because they consist of three dimensions: two spatial and one spectral.

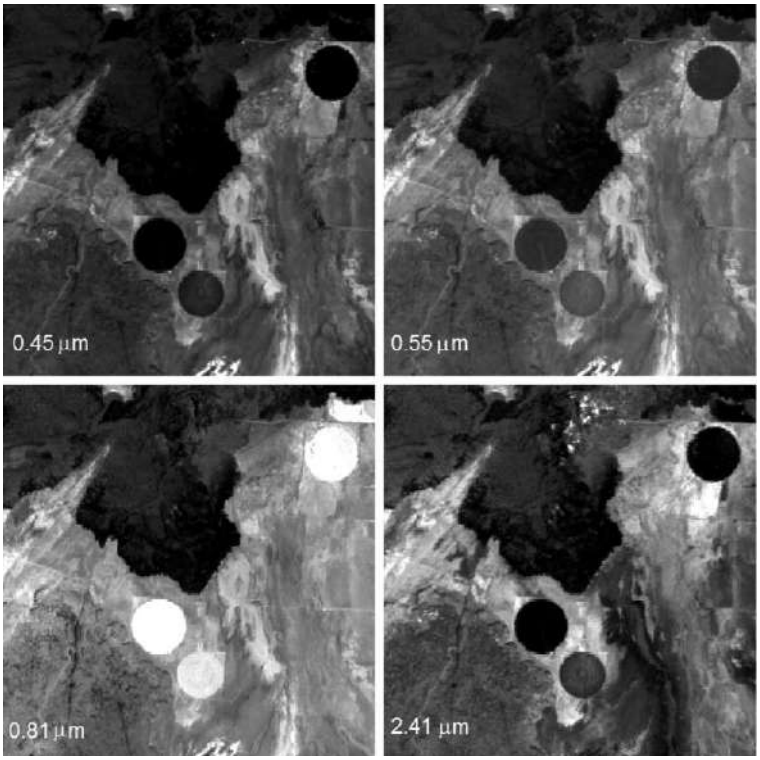

Figure 2 Portion of a scene from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) from Utah that includes a basalt flow, circularly irrigated grasses sparsely vegetated plains, and damp alluvial sediments. AVIRIS channels centered at 0.45, 0.55, 0.81, and 2.41 μm are shown. Green vegetation is brightest in the 0.81 μm band which samples the peak of the near infrared plateau (see **Figure 7**) just beyond the 'red edge'. The green vegetation and damp sediments are darkest in the 2.41 μm band where water has a very low reflectance (note also **Figure 8**). The bright patches in the basalt flow in the 2.41 μm image are occurrences of oxidized red cinders which have a high reflectance in the SWIR.

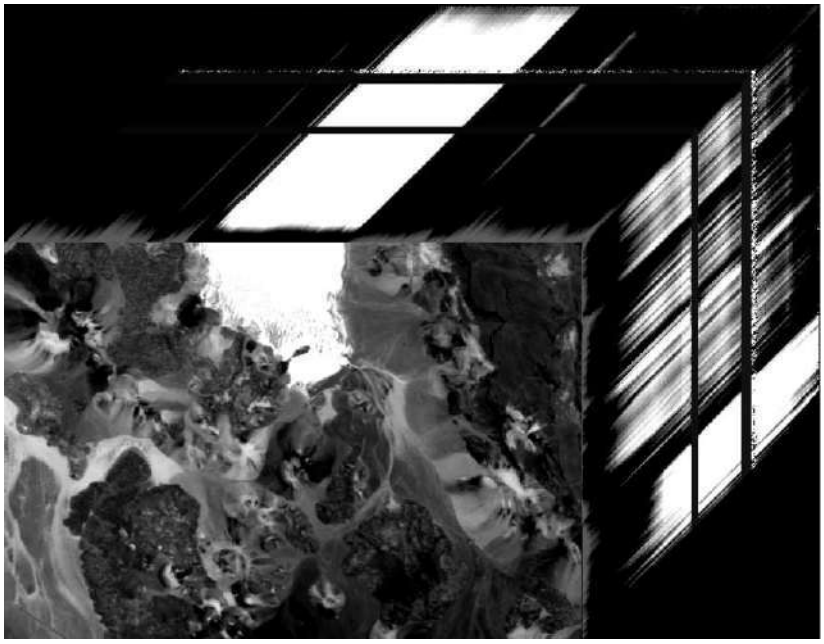

Figure 3 Visualization of a hyperspectral image cube. A 1.7 μm image band from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the Lunar Crater Volcanic Field, Nevada is shown with each of the 224 rightmost columns and 224 topmost lines arrayed behind it. In the lower right-hand corner, the low reflectance of ferric oxide-rich cinders in the blue and green show up as dark in the stacking of columns with the high reflectance in the infrared showing up as bright.

Advances in hyperspectral sensor technology have been accompanied by advances in associated technologies, such as computing power and the evolution of data processing algorithms.

Data Processing Approaches

An early impediment to the widespread use of hyperspectral data was the sheer volume of the data sets. For instance, a single scene from NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) is 614 samples by 512 lines by 224 bands. With its short integer storage format, such a scene takes up 140.8 Mbytes. However, since the dawn of the hyperspectral era, in the 1980s, computing power has expanded immensely and computing tasks, which once seemed prohibitive, are now relatively effortless. Also, numerous processing algorithms have been developed which are especially well suited for use with hyperspectral data.

Spectral Matching

In one class of algorithms, the spectrum associated with each spatial sample is compared against one or a group of model spectra and some similarity metric is applied. The model spectrum can be measured by a laboratory or field portable spectrometer of pure materials or can be a single pixel spectrum, or average of pixel spectra, covering a known occurrence of the material(s) to be mapped. A widely used similarity metric is the Spectral Angle Mapper (SAM). SAM is the angle, θ, obtained from the dot product of the image pixel spectrum, t, and the library spectrum, r, as expressed by

$$\theta = \cos^{-1}\left(\frac{t * r}{\|t\| * \|r\|}\right) \qquad [1]$$

SAM is insensitive to brightness variations and determines similarity based solely on spectral shape. Lower θ values mean more similar spectra.

In a related approach, matching is done, not on the entire spectrum, but rather on characteristic absorption features of the materials of interest. In Figure 4, the geometry of an absorption band is plotted. To map a given mineral, with a specific absorption feature, a high resolution laboratory spectrum is resampled to the spectral resolution of the hyperspectral sensor. The spectrum is subsampled to match the specific absorption feature (i.e., only bands from the short wavelength shoulder of the absorption to the long wavelength shoulder are included). A straight line continuum (calculated, based on a line between the two-band shoulders) is divided out from both the laboratory and the sensor spectra. Contrast between the library and the sensor spectra is mitigated through the use of an additive constant. This constant is incorporated into a set of equations which are solved through the use of standard least squares. The result of such a band-fitting algorithm is two data numbers per spatial sample of the hyperspectral scene. First the band depth (or, alternatively, band area) of the feature is calculated, and second, a goodness-of-fit parameter is calculated. These two parameters are most often combined in a band-depth times goodness-of-fit image. More sophisticated versions of the band-mapping algorithm target multiple absorption features and are, essentially, expert systems.

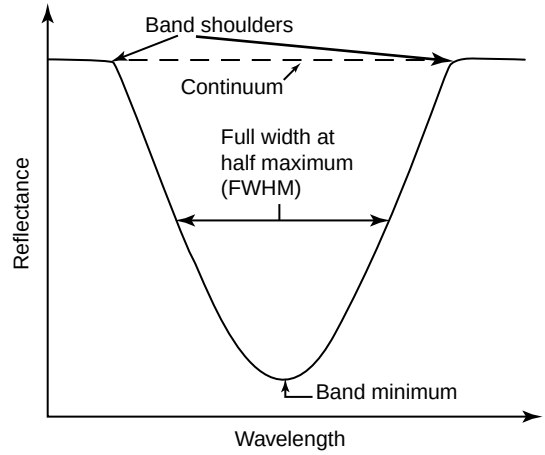

Figure 4 Parameters that describe an absorption band.

Spectral Mixture Analysis

A fundamentally different methodology for analyzing hyperspectral data sets is to model the measured spectrum of each spatial sample as a linear combination of endmember spectra. This methodology is based on the fact that the ground area corresponding to any given spatial sample likely will be covered by more than one material, with the consequence that the measured reflectance or emittance spectrum is a mixed pixel spectrum. The objective of linear SMA is to try to determine the fractional abundance of the component materials, or endmembers. The basis of linear spectral unmixing is to model the response of each spatial element as a linear combination of endmember spectra. The basis equation for linear SMA is

$$r(x, y) = \alpha \mathbf{M} + \varepsilon \qquad [2]$$

where $r(x, y)$ = the relative reflectance spectrum for the pixel at position (x, y), α = the vector of endmember abundances, $\mathbf{M}$ = the matrix of endmember spectra, and ε = the vector of residuals between the modeled and the measured reflectances. Application

of SMA results in a series of fraction images for each endmember wherein the data numbers range, ideally, between 0 and 1. Fraction image pixels with a digital number (DN) of 0 are taken to be devoid of the endmember material. Fraction image pixels with a DN of 1.0 are taken to be completely covered with the endmember material.

Techniques related to SMA have been developed which also map the abundance of a target material on a per pixel basis. These include implementations highly similar to SMA such as Orthogonal Subspace Projection (OSP) wherein, as with SMA, all endmembers must be determined a priori. Other approaches such as Constrained Energy Minimization (sometimes referred to as 'Matched Filter') do not require a priori knowledge of the endmembers. Instead, only the reflectance or emittance spectrum of the target need be known and the undesired spectral background is estimated, and ultimately compensated for, using the principal eigenvectors of the sample correlation or covariance matrix of the hyperspectral scene.

Expert Systems and Artificial Neural Networks

Another means of extracting information from hyperspectral data sets is to use computer approaches that in some way mimic the human thought process. This can take the form of an expert system which applies a set of rules or tests as each pixel spectrum is successively analyzed. For example, if a spectrum has an absorption at 2.2 μm it could tentatively be classified as being caused by a clay mineral. Additionally, if the absorption is a doublet, the specific clay mineral is likely a kaolinite. In practice, expert system codes are lengthy and complex although excellent results can be demonstrated from their use, albeit at the expense of processing time. Another approach in this vein is the use of an artificial neural network (ANN) for data processing. The use of ANNs is motivated by their power in pattern recognition. ANN architectures are well suited for a parallel processing approach and thus have the potential for rapid data processing.

Hyperspectral Remote Sensing of the Land

Geology

Early efforts with airborne hyperspectral sensors focused on geologic remote sensing. This was, at least in part, a consequence of the mineralogic diversity of the Earth's surface and the fact that many minerals have absorption features which are unique and diagnostic of the mineral's identity. The expectation, which was soon borne out by results from these early sensors, was that, given exposure of the surface, mineralogic maps as detailed as the spatial resolution of the sensor could be derived from hyperspectral data. As some of the processing techniques discussed above have become more readily available, it has become possible to produce mineralogic maps from hyperspectral data far more rapidly and cost-effectively than geologic maps produced by standard means (e.g., by walking over the ground and manually noting which lithologies are underfoot).

The propensity for transition metals, most especially iron, to absorb energy through charge transfer and crystal field effects, is noted above. Fe–O and Fe^{2+}–Fe^{3+} charge transfers cause a profound absorption in the reflectance spectrum of Fe-bearing minerals shortwards of 0.4 μm. The wing of this absorption causes the low blue and green reflectance of Fe-bearing minerals. Crystal field bands cause an absorption feature in the 0.9 to 1.0 μm region. The ability to discriminate subtle differences among Fe-bearing minerals in hyperspectral data has proven extremely valuable in applications such as mineral exploration, volcanology, and in the characterization of abandoned mine lands.

In the SWIR, absorptions caused by vibrational overtones of molecular bonds within minerals. These include absorptions in the 2.2 and 2.3 μm region which are characteristic of many Al- and Mg-bearing clay minerals. Absorptions in the SWIR are narrower in width than the Fe-generated crystal field bands discussed previously. In fact, the requirement to efficiently resolve these vibrational overtone absorption bands (which have full width at half maximum (FWHM) bandwidths of 20 to 40 nm) helped to drive the selection of the nominal 10 nm bandwidth of early and most current hyperspectral sensors. Certain clay minerals can be indicators of hydrothermal activity associated with economic mineral deposits; thus, the SWIR is an important spectral region for mineral exploration. Clay minerals, by definition, have planar molecular structures that are prone to failure if subjected to shearing stresses. The ability to uniquely identify and map these minerals, using hyperspectral data, has thus been used to good effect to identify areas on volcanoes and other hydrothermally altered terrains that could be subject to landslides.

Minerals with adsorbed or molecularly bound OH^- and/or water have vibrational overtone absorptions near 1.45 and 1.9 μm, although these features are masked in remotely sensed data by prominent atmospheric water vapor bands at 1.38 and 1.88 μm. The reflectance of water and OH-bearing minerals decrease to near zero at wavelengths longwards of

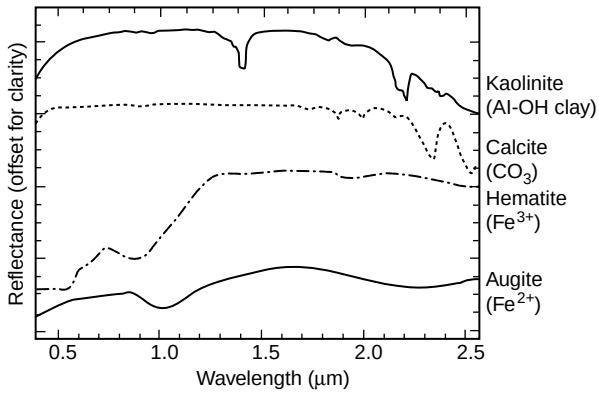

Figure 5 Reflectance spectra of representative minerals. Kaolinite represents clay minerals with its diagnostic 2.2 μm band caused by an overtone vibration of the Al–OH bond. The 2.335 μm band of calcite is caused by an overtone vibration of the carbonate molecule. The features shortwards of 1.0 μm in the hematite spectrum are caused by crystal field and charge transfer absorptions due to Fe^{3+}. The bands near 1.0 and longwards of 2.0 μm in the augite spectrum are the result of Fe^{2+} crystal field bands.

2.5 μm, due to the presence of water and OH fundamental vibrations near 3.0 μm. Minerals bearing the carbonate (CO_3^{-2}) molecule are abundant on the Earth's surface due to the ubiquitous occurrences of limestone and dolomite. Calcite and dolomite have overtone absorptions centered at 2.335 and 2.315 μm, respectively. A good test for the spectral resolution and spectral calibration of a hyperspectral sensor is its ability to successfully discriminate calcite from dolomite on the basis of the aforementioned absorption features. Figure 5 shows representative spectra of carbonate, clay, and ferric and ferrous iron-bearing minerals.

Absorptions resulting from fundamental molecular absorptions are manifested in the LWIR. These absorptions are of great geologic interest because they include absorptions resulting from vibrations of the Si–O bond and silicate minerals form the bulk of the Earth's crust. The wavelength at which the Si–O stretching feature occurs is dictated by the level of polymerization (or molecule-to-molecule bonding) of the silicate mineral. Highly polymerized framework silicate minerals, such as quartz and feldspar, have a shorter wavelength absorption than do silicate minerals, such as olivine which are composed of disconnected SiO_4 molecules. In Figure 6, laboratory emission spectra of highly and poorly polymerized silicate minerals are shown.

Vegetation and the Environment

While different minerals are, by definition, generally composed of diverse component molecules, different species of vegetation represent variations of the same general biochemical constituents (e.g., chlorophyll, proteins, lignin, cellulose, sugar, starch, etc.). In the thermal IR, vegetation has a generally flat emittance, thus it is in the reflective solar spectrum that most vegetative remote sensing is performed. The major features in the reflectance spectrum of green vegetation are shown in Figure 7. In the visible, the major absorption features are caused by the presence of chlorophyll. Chlorophyll has strong absorptions in the blue and the red, leaving a reflectance maximum (the green peak) at 0.55 μm. In the NIR, scattering in the leaf structure causes high reflectance leaving a profound absorption edge (the red edge) between

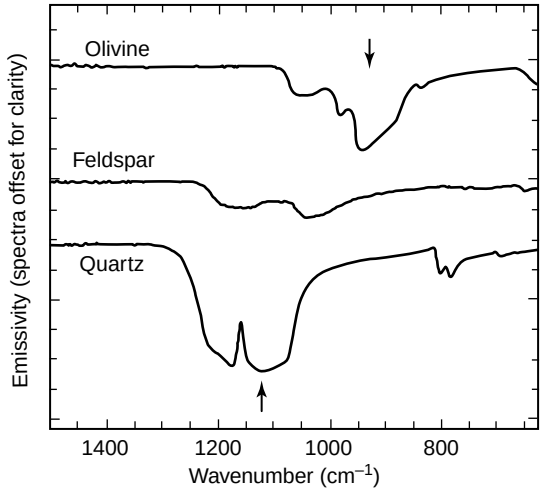

Figure 6 Emissivity spectra of SiO_4-bearing minerals showing the shift in band minimum to higher wave number with increasing polymerization.

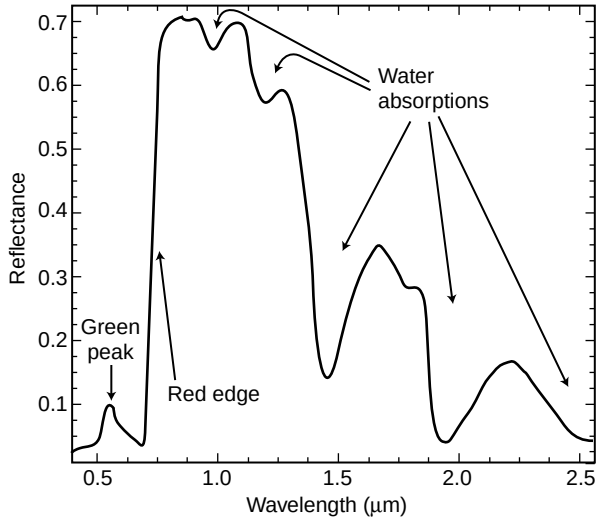

Figure 7 Major spectral features in the reflectance spectrum of green vegetation.

0.7 and 0.78 μm. Past the red edge and into the SWIR, the spectrum of green vegetation is dominated by water in the leaf, with leaf water absorptions occurring at 0.97, 1.19, 1.45, 1.93, and 2.50 μm. Studies of vegetation, using multispectral systems, have made use of a number of broadband indices. Hyperspectral systems allow for the discrimination of individual absorption features and subtle spectral shape differences in vegetation spectra which are unresolvable using broadband multispectral systems. For example, the absorptions caused by chlorophyll can be used for studies of chlorophyll content in vegetative canopies. The unique interaction of chlorophyll with leaf structures can provide a morphology to the green peak that is unique to a given species and thus is mappable using a spectral feature fitting approach.

A desired goal in the study of vegetation and ecosystem processes is to be able to monitor the chemistry of forest canopies. The foliar biochemical constituents making up forest canopies have associated absorption features in the SWIR, that result from vibrational overtones and combinations of C–O, O–H, C–H, and N–H molecular bonds. However the absorptions from these biochemical components overlap so that the chemical abundance of any one plant component cannot be directly related to any one absorption feature. An even more serious challenge is introduced by the strong influence of water in the SWIR in green vegetation spectra. Water constitutes 40 to 80% of the weight of leaves. However, the unique characteristics of high-quality hyperspectral data (e.g., excellent radiometric and spectral calibration, high signal-to-noise ratio, high spectral resolution) have been used to detect even these subtle features imprinted on the stronger water features.

The amount of water present in plant leaves is, in itself, a valuable piece of information. Hyperspectral data can be used to determine the equivalent water thickness present in vegetative canopies. Conversely, the amount of dry plant litter and/or loose wood can be estimated using spectral mixture analysis and related techniques. Taken together, all these data: vegetation species maps, determinations of leaf water content and the relative fractions of live vegetation versus litter, can be used to characterize woodlands for forest fire potential. The ability to map different species of vegetation has also proved useful in the field of agriculture, for assessing the impact of invasive weed species on farm and ranch land. Being able to assess the health of crops, in terms of leaf water content, is also important for agriculture.

Another subtle vegetative spectral feature, that provides information on plant vitality, is the aforementioned red edge. Shifts in the position of the red edge have been linked to vegetation stress. Studies have shown that vegetation stress and consequent shifts in the position of the red edge can be caused by a number of factors, including insufficient water intake or the intake of pernicious trace metals. Vegetation growing over mineral deposits can display red edge shifts and this can be used to assist in mineral exploration efforts.

Another application area, in which hyperspectral remote sensing has found great utility, is in the analysis of snow-covered areas. Over the reflective solar spectrum, the reflectance of snow varies from values near zero in the SWIR to values near one in the blue. Differences in the grain size of snow result in differences in reflectance as is illustrated in Figure 8. The spectral variability of snow makes it easily distinguishable from other Earth surface materials.

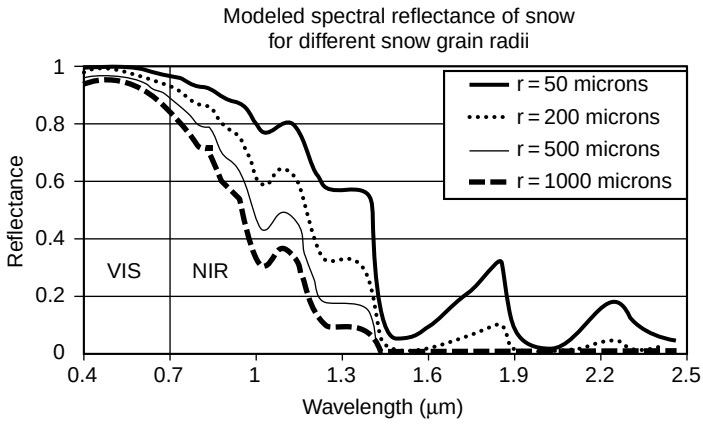

Figure 8 Differences in snow reflectance spectra resulting from differences in snow grain size. Courtesy of Dr Anne Nolin of Oregon State University.

Linear spectral unmixing, using hyperspectral data, has been shown to successfully discriminate between snow, vegetation, rock, and clouds and to accurately map snow cover fraction in mixed pixels. Accurate maps of snow cover, types of snow, and fractions of liquid water admixed with snow grains, are required for forecasting snowmelt runoff and stream discharge in watersheds dominated by snow cover.

Hyperspectral Remote Sensing of the Atmosphere

The physical quantity most often sought in land remote sensing studies is surface reflectance. However, the quantity recorded by a hyperspectral sensor is radiance at the entrance aperture of the sensor. In order to obtain reflectance, the background energy level of the Sun and/or Earth must be removed and the scattering and absorbing effects of the atmosphere must be compensated for. In the VNIR through SWIR, there are seven atmospheric constituents with significant absorption features: water vapor, carbon dioxide, ozone, nitrous oxide, carbon monoxide, methane, and oxygen. Molecular scattering (commonly called Rayleigh scattering) is strong in the blue but decreases rapidly with increasing wavelength. Above 1 μm, its effect is negligible. Scattering caused by atmospheric aerosols, or Mie scattering, also is more prominent at shorter wavelengths and decreases with increasing wavelength, but the dropoff of effects from Mie scattering is not as profound as that for Rayleigh scattering. Consequently, effects from aerosol scattering can persist into the SWIR.

There are three primary categories of methods for removing the effects of atmosphere and solar insolation from hyperspectral imagery, in order to derive surface reflectance. Methods of atmospheric corrections can be considered as being either an image-based, empirical, or model-based approach. An image-based, or 'in-scene' approach uses only data measured by the instrument. Empirical methods make use of the remotely sensed data in combination with field measurements of reflectance, to solve a simplified equation of at-sensor radiance such as eqn [3].

$$L_s = A\rho + B \qquad [3]$$

where L_s is the at-sensor radiance, ρ is the reflectance of the surface, and A and B are quantities that incorporate, respectively, all multiplicative and additive contributions to the at-sensor radiance. All the quantities expressed in eqn [3] can be considered as varying as a function of wavelength, λ. Approximations for A and B from eqn [3] constitute a set of gains and offsets derived from the empirical approach. In some empirical approaches, atmospheric path radiance is ignored and only a multiplicative correction is applied.

Model-based approaches seek to model what the at-sensor radiance should be on a pixel-by-pixel basis, including the contribution of the atmosphere. The at-sensor radiance, L_s, at any wavelength, λ, can be expressed as

$$L_s = \frac{1}{\pi}(E\rho + M_T)\tau_\theta + L_u \qquad [4]$$

where E is the irradiance at the surface of the Earth, ρ is the reflectance of the Earth's surface, M_T is the spectral radiant exitance of the surface at temperature, T, τ_θ is the transmissivity of the atmosphere at zenith angle θ, and L_u is the spectral upwelling path radiance of the atmosphere. The ability to solve eqn [4] and do atmospheric correction on a pixel-by-pixel basis is appealing, in that it negates the shortcomings of an empirical approach where a correction based on calibration targets in one part of a scene might not be appropriate for another part of the scene, due to differences in atmospheric path-length or simple atmospheric heterogeneity. Model-based approaches are also able to take advantage of the greater spectral dimensionality of hyperspectral data sets for the derivation of atmospheric properties (e.g., the amount of column water vapor, CO_2 band depths, etc.) directly from the data.

The main atmospheric component, affecting imagery in the reflective solar portion of the spectrum, is water vapor. Model-based atmospheric correction techniques determine the column water vapor for each pixel in the scene, based on the depth of the 940 and/or 1130 nm atmospheric water bands. Thus for each pixel, an appropriate amount of water can be removed. The maps of atmospheric water vapor distribution are themselves of interest to atmospheric scientists. Absorption features caused by well-mixed gases, such as the 760 nm O_2 band, can be used by model-based programs, to produce images of approximate scene topography. More advanced model-based approaches also solve for the scattering effects of aerosols on the hyperspectral data.

As noted above, atmospheric constituents, including industrial effluents, can be detected and spatially mapped by hyperspectral sensors. The fundamental vibrational absorption of gases of interest occurs in the MWIR and the LWIR. Overtones of these gaseous molecular vibrations occur in the VNIR to SWIR, but are generally too weak to be detected. The ability to detect anthropomorphically produced gases depends on a number of factors, including the

temperature difference between the gas cloud and the background atmosphere, the concentration of gas within the cloud, and the size of the cloud (e.g., the path length of light through the cloud).

See also

Imaging: Infrared Imaging. **Instrumentation:** Spectrometers.

Further Reading

Clark RN (1999) Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz AN (ed.) *Remote Sensing for the Earth Sciences*, pp. 3–57. New York: John Wiley and Sons.

Farrand WH and Harsanyi JC (1997) Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho through the use of a Constrained Energy Minimization technique. *Remote Sensing Environment* 59: 64–76.

Gao B-C and Goetz AFH (1995) Retrieval of equivalent water thickness and information related to biochemical components of vegetation canopies from AVIRIS data. *Remote Sensing Environment* 52: 155–162.

Gao B-C, Heidebrecht KH and Goetz AFH (1993) Derivation of scaled surface reflectances from AVIRIS data. *Remote Sensing Environment* 44: 165–178.

Goetz AFH (1989) Spectral remote sensing in geology. In: Asrar G (ed.) *Theory and Applications of Optical Remote Sensing*, pp. 491–526. New York: John Wiley & Sons.

Goetz AFH, Vane G, Solomon JE and Rock BN (1985) Imaging spectrometry for Earth remote sensing. *Science* 228: 1147–1153.

Hapke B (1993) *Theory of Reflectance and Emittance Spectroscopy*. Cambridge, UK: Cambridge University Press.

Kirkland LE, Herr KC and Salisbury JW (2001) Thermal infrared spectral band detection limits for unidentified surface materials. *Applied Optics* 40: 4852–4862.

Martin ME, Newman SD, Aber JD and Congalton RG (1998) Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing Environment* 65: 249–254.

Painter TH, Roberts DA, Green RO and Dozier J (1998) The effect of grain size on spectral mixture analysis of snow-covered area with AVIRIS data. *Remote Sensing Environment* 65: 320–332.

Pieters CM and Englert PAJ (eds) *Remote Geochemical Analysis: Elemental and Mineralogical Composition*. Cambridge, UK: Cambridge University Press.

Pinzon JE, Ustin SL, Castaneda CM and Smith MO (1997) Investigation of leaf biochemistry by hierarchical foreground/background analysis. *IEEE Transactions on Geoscience and Remote Sensing* 36: 1–15.

Roberts DA, Green RO and Adams JB (1997) Temporal and spatial patterns in vegetation and atmospheric properties using AVIRIS. *Remote Sensing Environment* 62: 223–240.

Wessman CA, Aber JD and Peterson DL (1989) An evaluation of imaging spectrometry for estimating forest canopy chemistry. *International Journal of Remote Sensing* 10: 1293–1316.

Laser Detection of Atmospheric Gases

E V Browell, W B Grant and S Ismail, NASA Langley Research Center, Hampton, VA, USA

Published by Elsevier Ltd.

Introduction

Lidar (light detection and ranging) systems are able to measure profiles of atmospheric aerosols, clouds, and gases by transmitting a pulsed laser beam into the atmosphere and collecting the backscattered radiation from aerosols and molecules in the atmosphere with a receiver located near the transmitter. The differential absorption lidar (DIAL) approach is the most widely used technique for measuring a variety of gases. In this approach, two closely spaced laser wavelengths are used: one which is absorbed by the gas of interest, and the other which is only weakly, or not at all, absorbed by the gas. A differential with respect to range and wavelength is calculated, to determine the average gas concentration along any segment of the lidar path, using the familiar Beer–Lambert law for an absorbing medium.

The DIAL equation can be expressed in its simple form as

$$N = \left(1/\left(2(R_2 - R_1)(\sigma_{\text{on}} - \sigma_{\text{off}})\right)\right) \times \ln\left(P_{\text{off}}(R_2)P_{\text{on}}(R_1)\right)/\left(P_{\text{off}}(R_1)P_{\text{on}}(R_2)\right) \quad [1]$$

where N is the average gas concentration, R is the range, σ_{on} and σ_{off} are the absorption cross sections at the on- and off-line wavelengths, and $P_{\text{on}}(R)$ and $P_{\text{off}}(R)$ are the powers received at the on- and off-line wavelengths. However, since it is performed in the atmosphere and not a laboratory, great care must be taken to ensure that the data are analyzed properly to minimize random and systematic errors. Random errors arise from noise in the backscattered signal, the solar background signal, and the inherent

detector noise, and this type of error can be reduced by signal averaging. Systematic errors arise from uncompensated instrument and atmospheric effects and must be carefully considered in system design and operation and data processing.

Another approach, the Raman lidar approach, uses inelastic scattering (scattered/emitted light has different wavelength than the illuminating light) from gases where the wavelength shift corresponds to vibrational or rotational energy levels of the molecules. Any illuminating laser wavelength can be used, but since the Raman scattering cross-section varies as λ^{-4}, where λ is the wavelength, the shorter laser wavelengths, such as in the near UV spectral region, are preferred. While even shorter wavelengths in the solar blind region below ~300 nm would permit operation in daytime, the atmospheric attenuation, due to Rayleigh scattering and UV molecular absorption, limits the measurement range. High-power lasers are used due to the low Raman scattering cross-sections. The Raman lidar measurements have to be carefully calibrated against a profile measured using another approach, such as is done for water vapor (H_2O) measurements with the launch of hygrometers on radiosondes, since the lidar system constants and atmospheric extinctions are not usually well known or modeled. In order to obtain mixing ratios of H_2O with respect to atmospheric density, the H_2O signals are ratioed to the nitrogen Raman signals. However, care must be exercised in processing these data due to the spectral dependences of atmospheric extinction.

Surface-Based Lidar Systems

Surface-based lidar systems measure the temporal evolution of atmospheric profiles of aerosols and gases. The first lidar systems used to remotely measure atmospheric gases were Raman lidar systems. The Raman lidar systems were thought to be very promising since one high-power laser could be used to measure a variety of gases. They are easier to develop and operate than DIAL systems, because the laser does not have to be tuned to a particular gas absorption feature. However, the Raman scattering cross-section is low, and the weak, inelastically scattered signal is easily contaminated by daylight background radiation, resulting in greatly reduced performance during daytime. Raman lidar systems have been developed primarily for measuring H_2O and retrieving atmospheric temperature.

Surface-based UV DIAL systems, that are part of the Network for the Detection of Stratospheric Change (NDSC) have made important contributions to the understanding of stratospheric O_3. There are DIAL systems at many locations around the Earth, with sites in Ny-Ålesund, Spitzbergen (78.9°N, 11.9°E), Observatoire Haute Provence, France (43.9°N, 5.7°E), Table Mountain, California (34.4°N, 118.2°W), Mauna Loa, Hawaii (19.5°N, 155.6°W), and Lauder, New Zealand (45.0°S, 169.7°E). First established in the 1980s, these DIAL systems are strategically located so that O_3 in different latitude bands can be monitored for signs of change. In addition, they can provide some profiles for comparison with space-based O_3 measuring instruments. The parameters of a typical ground-based UV DIAL system used in the NDSC are given in Table 1.

Airborne DIAL Systems

Airborne lidar systems expand the range of atmospheric studies beyond those possible by surface-based lidar systems, by virtue of being located in aircraft that can be flown to high altitudes and to remote locations. Thus, they permit measurements at locations inaccessible to surface-based lidar systems. In addition, they can make measurements of large atmospheric regions in times that are short compared with atmospheric motions, so that the large-scale patterns are discernible. Another advantage of airborne lidar operation is that lidar systems perform well in the nadir (down) direction since the atmospheric density and aerosol loading generally increases with decreasing altitude towards the surface, which helps to compensate for the R^{-2} decrease in the lidar

Table 1 Parameters for the Jet Propulsion Laboratory's Mauna Loa DIAL systems for stratospheric O_3 measurements

L	XeCl	Nd:YAG (3rd Harmon.)
Wavelength (nm)	308 (on)	355 (off)
Pulse energy (mJ)	300	150
Pulse repetition frequency (Hz)	200	100
R		
Area (m^2)	0.79	
Optical efficiency (%)	~40	
Wavelengths–Rayleigh (nm)	308	355
N_2 Raman (nm) (for aerosol correction)	332	387
Measurement range* (km)	15–55	
Vertical resolution (km)	3 at bottom, 1 at O_3 peak, 8–10 at top	
Measurement averaging time (hours)	1.5	
Measurement accuracy	<5% at peak, 10–15% at 15 km, >40% at 45 km	

*Chopper added to block beam until it reaches 15 km in order to avoid near field signal effects.

signal with range. For the zenith direction, the advantage is that the airborne lidar system is closer to the region being measured.

Field Measurement Programs

Global O_3 Measurements

The first airborne DIAL system, which was developed by the NASA Langley Research Center (LaRC), was flown for O_3 and aerosol investigations in conjunction with the Environmental Protection Agency's Persistent Elevated Pollution Episodes (PEPE) field experiment, conducted over the east coast of the US in the summer of 1980. This initial system has evolved into the advanced UV DIAL system that will be described in the next section. Airborne O_3 DIAL systems have also been developed and used in field measurement programs by several other groups in the United States, Germany, and France.

The parameters of the current NASA LaRC airborne UV DIAL system are given in Table 2. The on-line and off-line UV wavelengths of 288.2 and 299.6 nm are used for DIAL O_3 measurements during tropospheric missions, and 301 and 310 nm are used for stratospheric missions. This system also transmits 1,064 and 600 nm beams for aerosol and cloud measurements. The time delay between the on- and off-wavelength pulses is 400 μs, which is sufficient time for the return at the first set of wavelengths to end, but short enough that the same region of the atmosphere is sampled. This system has a demonstrated absolute accuracy for O_3 measurements of better than 10% or 2 ppbv (parts per billion by volume), whichever is larger, and a measurement precision of 5% or 1 ppbv with a vertical resolution of 300 m and an averaging time of 5 minutes (about 70 km horizontal resolution at typical DC-8 ground speeds).

The NASA LaRC airborne UV DIAL systems have made significant contributions to the understanding of both tropospheric and stratospheric O_3, aerosols, and clouds. These systems have been used in 18 international and 3 national field experiments over the past 24 years, and during these field experiments, measurements were made over, or near, all of the oceans and continents of the world. A few examples of the scientific contributions made by these airborne UV DIAL systems are given in Table 3.

The NASA LaRC airborne UV DIAL system has been used extensively in NASA's Global Tropospheric Experiment (GTE) program which was started in the early 1980s, had as its primary mission the study of tropospheric chemistry in remote regions of the Earth, in part to study and gain a better understanding of atmospheric chemistry in the unperturbed atmosphere. A related goal was to document the Earth's atmosphere in a number of places during seasons when anthropogenic influences are largely absent, then return to these areas later to document the changes. The field missions have included campaigns in Africa, Alaska, the Amazon Basin,

Table 2 Parameters of the NASA LaRC airborne UV DIAL system

L [illegible]: Nd:YAG-pumped dye lasers, frequency doubled into the UV			
Pulse repetition frequency (Hz)	30		
Pulse length (ns)	8–12		
Pulse energy (mJ) at 1.06 μm	250–300		
Pulse energy (mJ) at 600 nm	50–70		
UV pulse energy (mJ)			
For troposphere at 288/300 nm	20		
For stratosphere at 301/310 nm	20		
Dimensions (l × w × h) (cm)	594 × 102 × 109		
Mass (kg)	1735		
Power requirement (kW)	30		
		Wavelength region (nm)	
R [illegible]	289–311	572–622	1064
Area (m^2)	0.086	0.086	0.864
Receiver optical efficiency (%)	30	40	30
Detector quantum efficiency (%)	26 (PMT)	8 (PMT)	40 (APD)
Field-of-view (mrad)	≤1.5	≤1.5	≤1.5
[illegible]			
Measurement range (km)	up to 10–15 (nadir and zenith)		
Vertical resolution (m)	300–1500, depending on range		
Horizontal resolution (km)	≤70		
Measurement accuracy	≤10% or 2 ppbv, whichever is greater		

Table 3 Examples of significant contributions of airborne O_3 DIAL systems to the understanding of tropospheric and stratospheric O_3 (see also **Figures 1–5**)

[illegible]:
Air mass characterizations in a number of remote regions
Continental pollution plume characterizations (see **Figures 1** and **2**)
Study of biomass burn plumes and the effect of biomass burning on tropospheric O_3 production (see **Figure 3**)
Case study of warm conveyor belt transport from the tropics to the Arctic
Study of stratospheric intrusions and tropopause fold events (see **Figure 4**)
Observation of the decay of a cutoff low
Power plant plume studies
[illegible]:
Contributions to the chemical explanation of behavior of Antarctic O_3
Quantification of O_3 depletion in the Arctic (see **Figure 5**)
Polar stratospheric clouds – particle characterizations
Intercomparison with surface-based, airborne and space-based instruments
Quantification of O_3 reduction in tropical stratosphere after the June 1991 eruption of Mount Pinatubo and characterization of the tropical stratospheric reservoir edge
Cross-vortex boundary transport

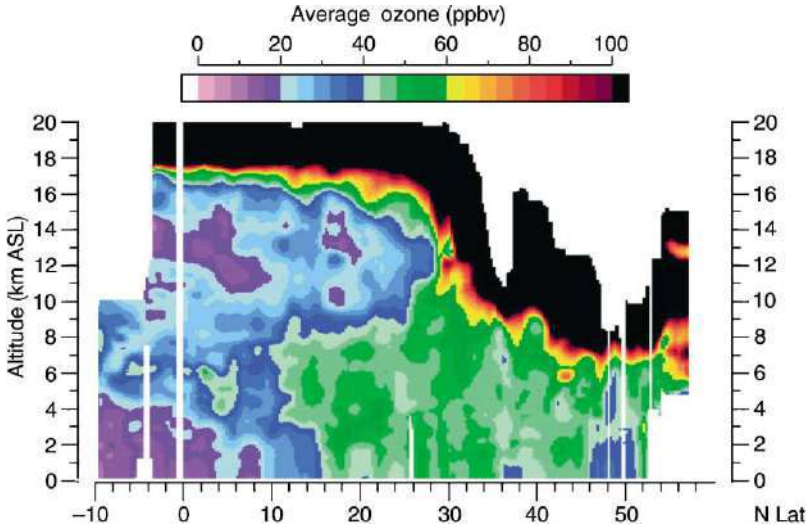

Figure 1 Latitudinal distribution of ozone over the western Pacific Ocean obtained during the second Pacific Exploratory Mission (PEM West B) in 1994.

Canada, and North America, and over the Pacific Ocean from Antarctica to Alaska, with concentrated measurements off the east coast of Asia and in the tropics. One of the bigger surprises of the two decades of field missions, was the discovery in the mid-1990s, that the tropical and South Pacific Ocean had very high tropospheric O_3 concentrations in plumes from biomass burning in Africa and South America during the austral spring. There were few indications from surface-based or space-based measurements that there were extensive biomass burn plumes in the area, primarily since the plumes were largely devoid of aerosols due to being stripped out during cloud convective lofting. Once over the ocean, horizontal transport appears to proceed relatively unimpeded unless a storm system is encountered.

The NASA LaRC airborne UV DIAL system has also been flown in all the major stratospheric O_3 campaigns starting in 1987 with the Airborne Antarctic Ozone Experiment (AAOE) to determine the cause of the Antarctic ozone hole. The UV DIAL system documented the O_3 loss across the ozone hole region. Later, when attention was turned to the Arctic and the possibility of an ozone hole there, the system was used to produce an estimate of O_3 loss during the winter season as well as to better characterize the polar stratospheric cloud (PSC) particles. The UV DIAL system was also used to study O_3 loss in the tropical stratospheric reservoir following the eruption of Mount Pinatubo in June 1991. The loss was spotted by the Microwave Limb Sounder (MLS) on the Upper Atmospheric Research Satellite (UARS), but the MLS

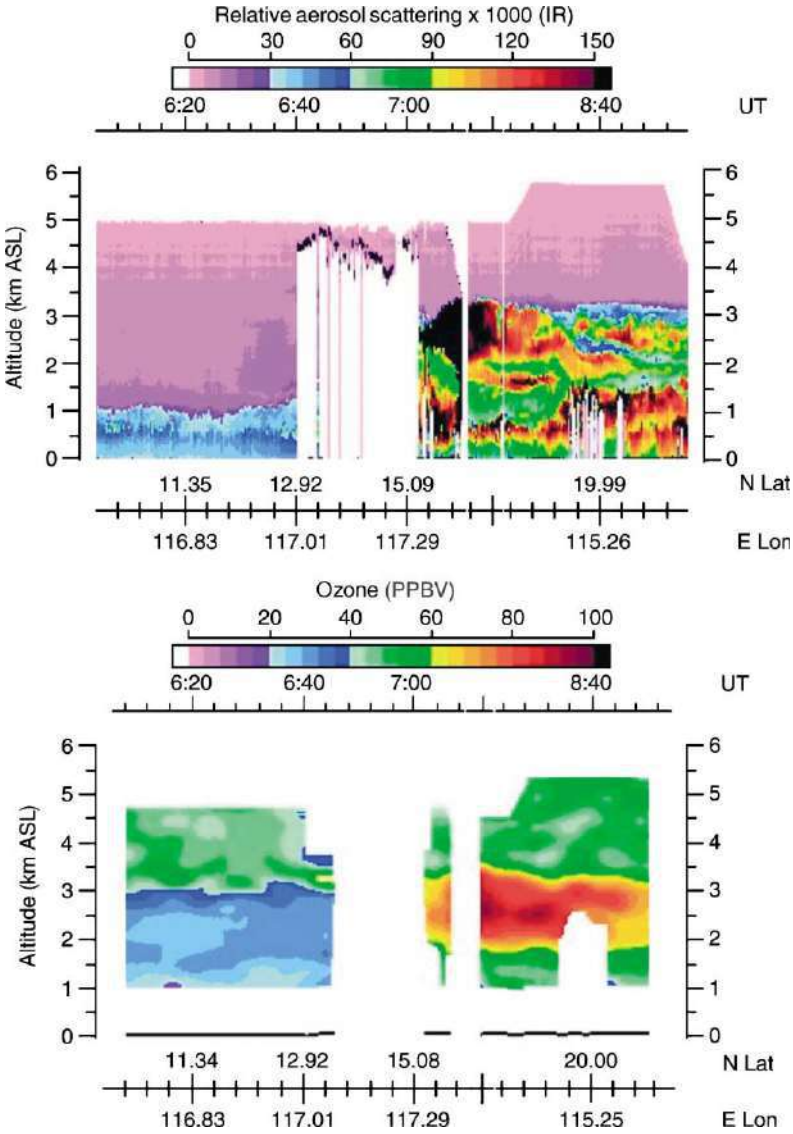

Figure 2 Pollution outflow from China over the South China Sea (right side of figure) with clean tropical air on south side of a front (left side of figure), observed during PEM West B in 1994.

was unable to study the loss in detail due to its low vertical resolution (5 km) compared to the small-scale (2–3 km) features of the O_3 loss. Other traditional space-based O_3 measuring instruments also had difficulty during this period, due to the high aerosol loading in the stratosphere following the eruption.

Space-Based O_3 DIAL System

In order to obtain nearly continuous, global distributions of O_3 in the troposphere, a space-based O_3 DIAL system is needed. A number of key issues could be addressed by a space-based O_3 DIAL system including: the global distribution of photochemical O_3 production/destruction and transport in the troposphere; location of the tropopause; and stratospheric O_3 depletion and dynamics. High-resolution airborne O_3 DIAL and other aircraft measurements show that to study tropospheric processes associated with biomass burning, transport of anthropogenic pollutants, tropospheric O_3 chemistry and dynamics, and stratosphere–troposphere exchange, a vertical profiling capability from space with a resolution of 2–3 km is needed, and this capability cannot currently be achieved using passive remote sensing satellite instruments. An example of the type of latitudinal O_3 cross-section that could be provided by a space-based O_3 DIAL system is shown in Figure 1. This figure shows many different aspects of O_3 loss and production; vertical and horizontal transport; and stratosphere–troposphere exchange that occurs from the tropics to high latitudes.

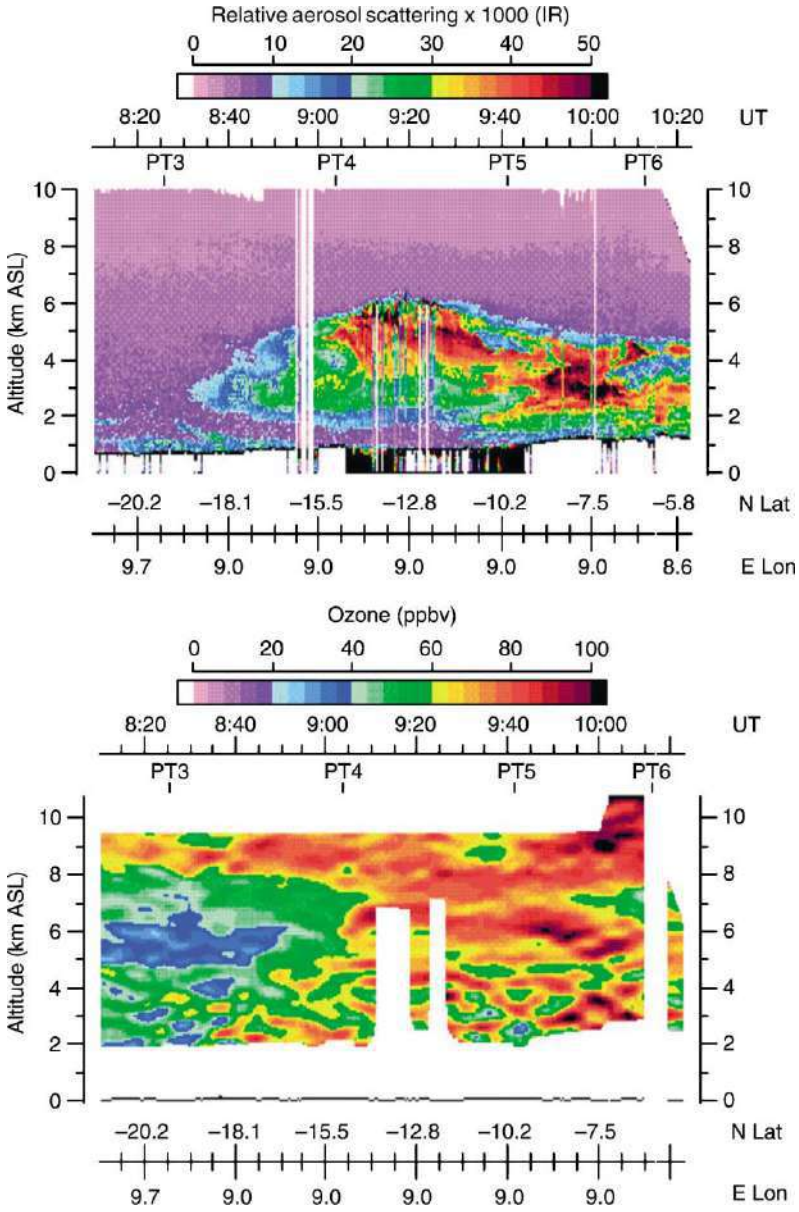

Figure 3 Biomass burning plume over the Atlantic Ocean arising from biomass burning in the central part of western Africa, observed during the TRACE-A mission in 1992.

This type of data would be available from just one pass from a space-based O_3 DIAL system. A space-based O_3 DIAL system optimized for tropospheric O_3 measurements (see system description in Table 4a,b) would also permit high-resolution O_3 measurements in the stratosphere (1 km vertical, 100 km horizontal), along with high-resolution aerosol measurements (100 m vertical, 10 km horizontal). In addition, these DIAL measurements will be useful in assisting in the interpretation of passive remote sensing measurements and in helping to improve their data processing algorithms.

Global H_2O Measurements

H_2O and O_3 are important to the formation of OH in the troposphere, and OH is at the center of most of the chemical reactions in the lower atmosphere. In addition H_2O is an excellent tracer of vertical and horizontal transport of air masses in the troposphere, and it can be used as a tracer of stratosphere–troposphere exchange. Increased aerosol sizes, due to high relative humidities, can also affect heterogeneous chemical processes and radiation budgets in the boundary layer and in cloud layers. Knowledge of H_2O is important to weather forecasting and climate

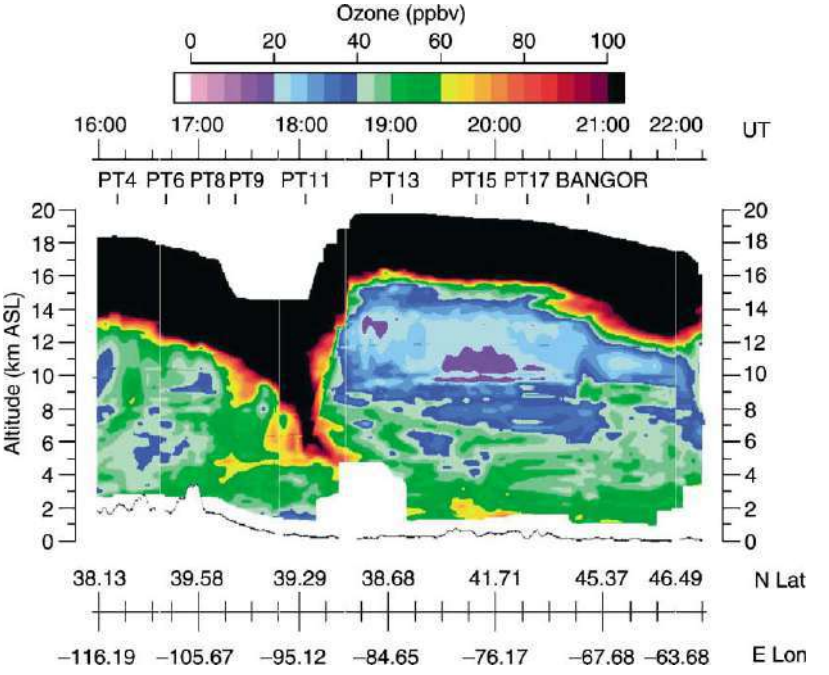

Figure 4 Ozone distribution observed on flight across US during the SASS (Subsonic Assessment) Ozone and Nitrogen Experiment (SONEX) in 1997. A stratospheric intrusion is clearly evident on left side of figure, and low ozone air from tropics transported to mid-latitudes can be seen in the upper troposphere on the right.

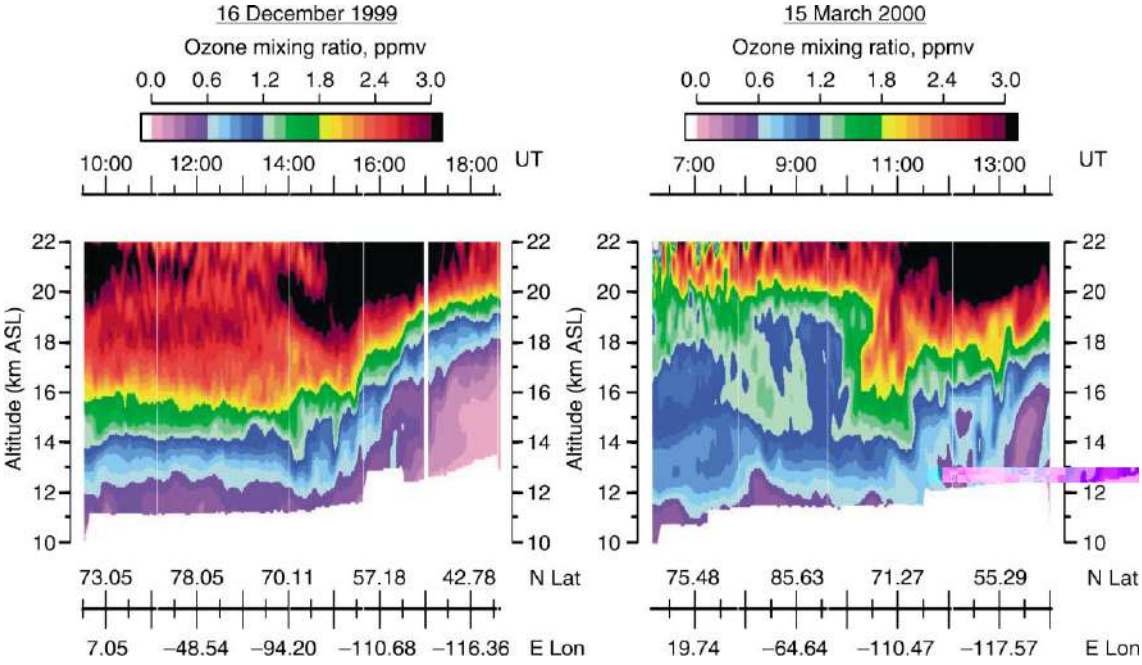

Figure 5 Ozone cross-sections in the stratosphere measured in the winter of 1999/2000 during the SOLVE mission. The change in ozone number density in the Arctic polar vortex due to chemical loss during the winter is clearly evident at latitudes north of 72°N.

predictions. Thus, H_2O distributions can be used in several different ways to better understand chemical, transport, radiation and meteorological processes in the global troposphere.

H_2O Raman Lidar Systems

The first Raman lidar measurements of H_2O were made in the late 1960s, but not much progress in using the Raman approach for H_2O was made until a

system using an abandoned searchlight mirror was employed by the NASA Goddard Space Flight Center in the mid-1980s, to show that if the signal collector (telescope) was large enough, useful Raman measurements could be made to distances of several kilometers. Until recently, Raman lidar measurements of H_2O were largely limited to night-time, due to the high sunlight background interference. Measurements are now made during daytime, using very narrow telescope fields of view and very narrowband filters in the receiver. A good example of a Raman lidar system used to measure H_2O is that at the Department of Energy's Atmospheric Radiation Measurement – Cloud and Radiation Test Bed (ARM-CART) site in Oklahoma. The site is equipped with a variety of instruments aimed at studying the radiation properties of the atmosphere. The system parameters are given in Table 4a,b.

Raman lidar systems have been used for important measurements of H_2O from a number of ground-based locations. Some of the important early work dealt with the passage of cold and warm fronts. The arrival of the wedge-shaped cold front, pushing the warm air up, was one of these studies. Raman lidar have also been located at the ARM-CART site in Kansas and Oklahoma, where they could both provide a climatology of H_2O as well as provide correlative measurements of H_2O for validation of space-based instruments.

Table 4a Parameters for the US Department of Energy's surface-based Raman lidar system for measurements of H_2O

L	Nd:YAG Laser
Wavelength (nm)	355
Pulse energy (mJ)	400
Pulse repetition frequency (Hz)	30
R	
Area (m^2)	1.1
Wavelengths (nm)	
Water vapor	407
Nitrogen	387
Measurement range (km)	
Night-time	near surface to 12
Daytime	near surface to 3
Range resolution (m)	39 at low altitudes, 300 >9 km
Measurement accuracy (%)	
Night-time	<10 <7 km (1 min avg) 10–30 at high altitudes (10–30 min avg)
Daytime	10 <1 km; 5–15 for 1–3 km (10 min avg)

Table 4b Parameters for the NASA Goddard Space Flight Center's Scanning Raman lidar system for measurements of H_2O

L	Nd:YAG Laser	XeF
Wavelength (nm)	355	351
Pulse energy (mJ)	300	30–60
Pulse repetition frequency (Hz)	30	400
R		
Area (m^2)	1.8	1.8
Wavelengths (nm)		
Water vapor	407	403
Nitrogen	387	382
Measurement range (km)		
Night-time	near surface to 12	
Daytime	near surface to 4	
Range resolution (m)	7.5 at low altitudes, 300 >9 km	
Measurement accuracy (%)		
Night-time	<10 <5 km (10 sec and 7.5 m range resolution) <10 <7 km (1 min avg) 10–30 at high altitudes (10–30 min avg)	
Daytime	<10 <4 km (5 min avg)	

H_2O DIAL Systems

H_2O was first measured with the DIAL approach using a temperature-tuned ruby laser lidar system in the mid-1960s. The first aircraft-based H_2O DIAL system was developed at NASA LaRC, and was flown in 1982, as an initial step towards the development of a space-based H_2O DIAL system. This system was based on Nd:YAG-pumped dye laser technology, and it was used in the first airborne H_2O DIAL atmospheric investigation, which was a study of the marine boundary layer over the Gulf Stream. This laser was later replaced with a flashlamp-pumped solid-state alexandrite laser, which had high spectral purity, i.e., little out-of-band radiation, a requirement since water vapor lines are narrow, and this system was used to make accurate H_2O profile measurements across the lower troposphere.

A third H_2O DIAL system, called LASE (Lidar Atmospheric Sensing Experiment) was developed as a prototype for a space-based H_2O DIAL system, and it was completed in 1995. This was the first fully autonomously operating DIAL system. LASE uses a Ti:sapphire laser that is pumped by a double-pulsed, frequency-doubled Nd:YAG to produce laser pulses in the 815 nm absorption band of H_2O (see Table 5). The wavelength of the Ti:sapphire laser is controlled by injection seeding with a diode laser that is frequency locked to a H_2O line using an

Table 5 Parameters of LASE H_2O DIAL system

L	Ti:sapphire
Wavelength (nm)	813–818
Pulse energy (mJ)	100
Pulse-pair repetition frequency	5 (on- and off-line pulses separated by 300 μs)
Linewidth (pm)	<0.25
Stability (pm)	<0.35
Spectral purity (%)	>99
Beam divergence (mrad)	<0.6
Pulse width (ns)	35
R	
Area (m^2)	0.11
Receiver optical efficiency (%)	50 (night), 35 (day)
Avalanch Photodiode (APD) detector quantum efficiency (%)	80
Field-of-view (mrad)	1.0
Noise equivalence power ($W\ Hz^{-0.5}$)	2×10^{-14}
Excess noise factor	3
Measurement range (altitude) (km)	15
Range resolution (m)	300–500
Measurement accuracy (%)	5

Table 6 Examples of significant contributions of airborne H_2O DIAL systems to the understanding of H_2O distributions (see also **Figures 6–8**)

NA A L H₂O DIAL LA E
Study of marine boundary layer over Gulf Stream
Observation of H_2O transport at a land/sea edge
Study of large-scale H_2O distributions across troposphere (see **Figure 6**)
Correlative and remote measurements
Observations of boundary layer development (see **Figure 7**)
Cirrus cloud measurements
Hurricane studies (see **Figure 8**)
Relative humidity effects on aerosol sizes
Ice supersaturation in the upper troposphere
Studies of stratospheric intrusions
H_2O distributions over remote Pacific Ocean
O H₂O DIAL
Boundary layer humidity fluxes
Lower-stratospheric H_2O studies

absorption cell. Each pulse pair consists of an on-line and off-line wavelength for the H_2O DIAL measurements. To cover the large dynamic range of H_2O concentrations in the troposphere (over 3 orders of magnitude), up to three line pair combinations are needed. LASE uses a novel approach of operating from more than one position on a strongly absorbing H_2O line. In this approach, the laser is electronically tuned at the line center, side of the line, and near the wing of the line to achieve the required absorption cross-section pairs (on and off). LASE has demonstrated measurements of H_2O concentrations across the entire troposphere using this 'side-line' approach. The accuracy of LASE H_2O profile measurements was determined to be better than 6% or 0.01 g/kg, whichever is larger, over the full dynamic range of H_2O concentrations in the troposphere. LASE has participated in over eight major field experiments since 1995. See **Table 6** for a listing of topics studied using airborne H_2O DIAL systems (**Figures 6–8**).

Space-Based H_2O DIAL System

The technology for a space-based H_2O DIAL system is rapidly maturing in the areas of: high-efficiency, high-energy, high-spectral-purity, long-life lasers with tunability in the 815- and 940-nm regions; low-weight, large-area, high-throughput, high-background-rejection receivers; and high-quantum-efficiency, low-noise, photon-counting detectors. With the expected advancements in lidar technologies leading to a 1-J/pulse capability, a space-based H_2O DIAL system could be flown on a long-duration space mission this decade.

Space-based DIAL measurements can provide a global H_2O profiling capability, which when combined with passive remote sensing with limited vertical resolution, can lead to 3-dimensional measurements of global H_2O distributions. High vertical resolution H_2O (≤1 km), aerosol (≤100 m), and cloud top (≤50 m) measurements from the lidar along the satellite ground-track, can be combined with the horizontally contiguous data from nadir passive sounders to generate more complete high-resolution H_2O, aerosol, and cloud fields for use in the various studies indicated above. In addition, the combination of active and passive measurements can provide significant synergistic benefits leading to improved temperature and relative humidity measurements. There is also strong synergism with aerosol and cloud imaging instruments and with future passive instruments that are being planned or proposed for missions addressing atmospheric chemistry, radiation, hydrology, natural hazards, and meteorology.

Tunable Laser Systems for Point Monitoring

Tunable diode laser (TDL) systems are also used to measure atmospheric gases on a global scale from aircraft and balloons. TDLs are small lasers that emit extremely narrowband radiation and can be tuned in the near IR spectral region using a combination of temperature and current. TDLs can be built into very sensitive and compact systems and located on the surface or flown on aircraft or balloons and used to measure such species as CO, HCl, CH_4, and oxides of nitrogen such as N_2O and NO_2. They derive their

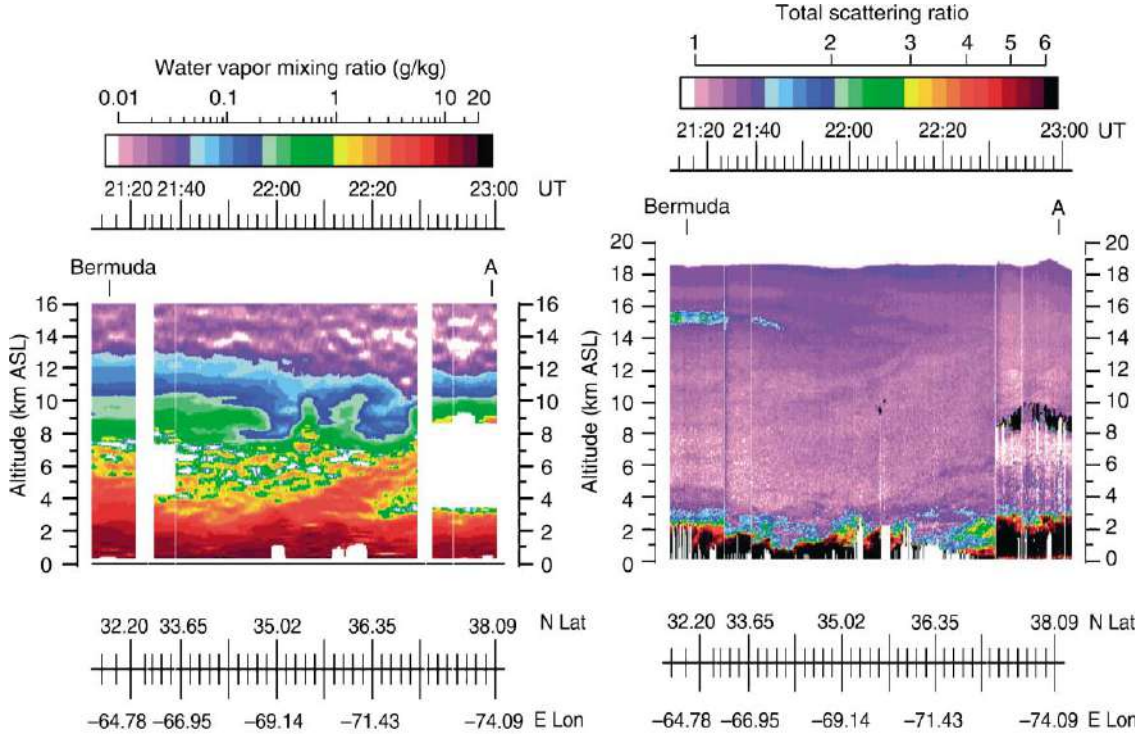

Figure 6 LASE measurements of water vapor (left) and aerosols and clouds (right) across the troposphere on an ER-2 flight from Bermuda to Wallops during the Tropospheric Aerosol Radiative Forcing Experiment (TARFOX) conducted in 1996.

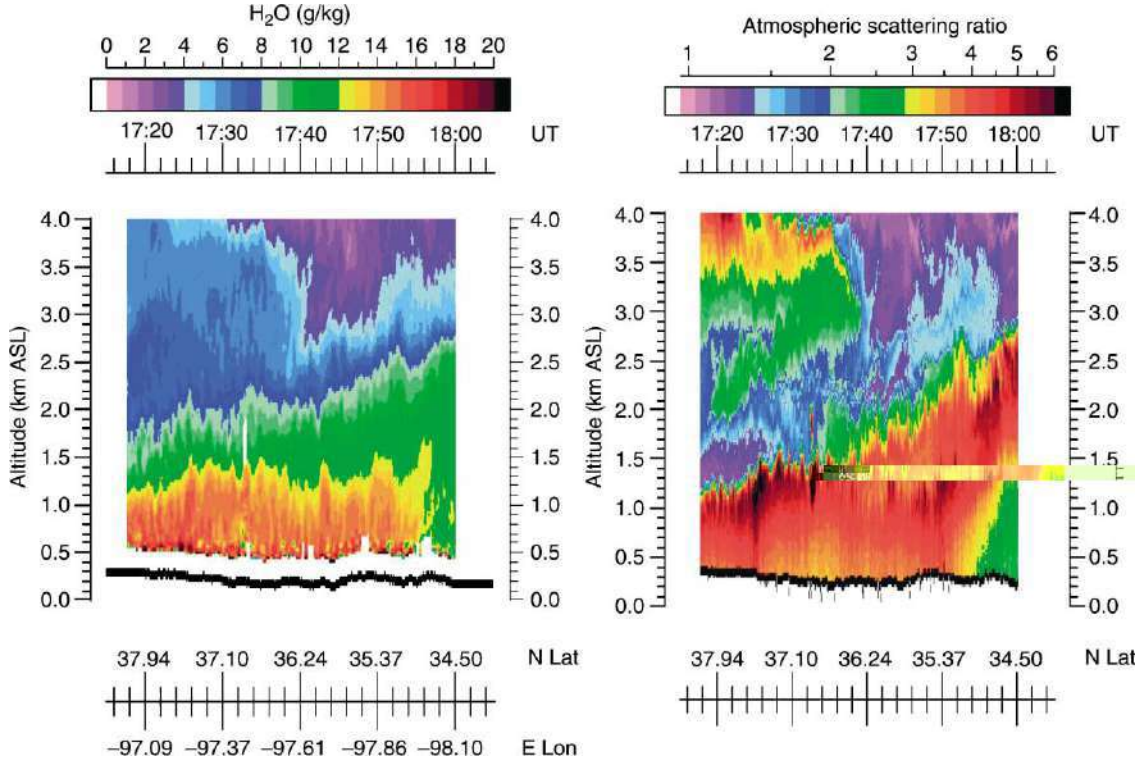

Figure 7 Water vapor and aerosol cross-section obtained on a flight across a cold front during the Southern Great Plains (SGP) field experiment conducted over Oklahoma in 1997.

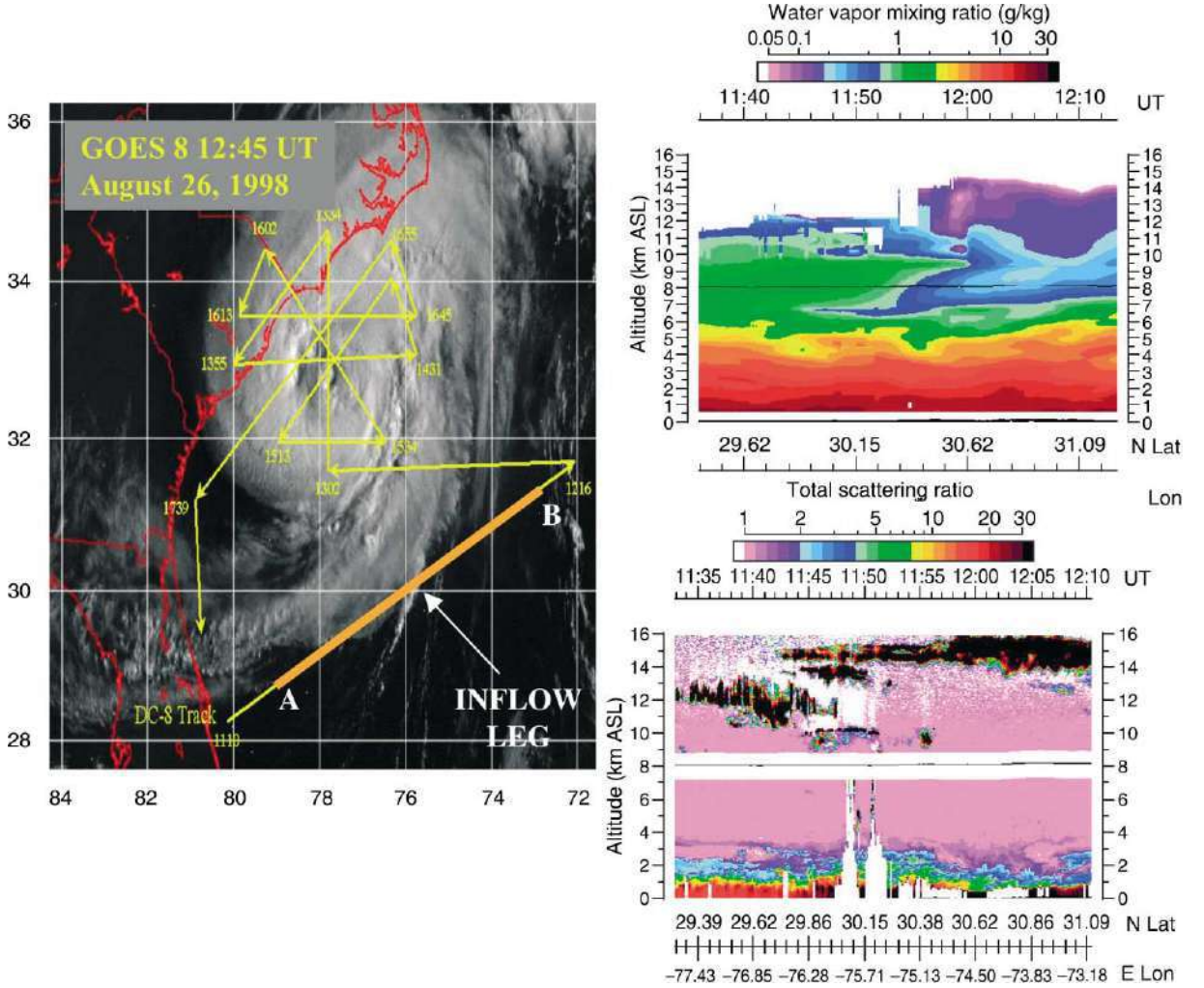

Figure 8 Measurements of water vapor, aerosols, and clouds in the inflow region of Hurricane Bonnie during 1998 Convection and Moisture Experiment (CAMEX-3). A rain band can be clearly seen at the middle of Leg-AB on the satellite and LASE cross-sections.

high sensitivity to the fact that the laser frequency is modulated at a high frequency, permitting a small spectral region to be scanned rapidly. The second- or fourth-harmonic of the scan frequency is used in the data acquisition, effectively eliminating much of the low-frequency noise due to mechanical vibrations and laser power fluctuations. In addition, multipass cells are employed, thereby generating long paths for the absorption measurements.

TDL systems have been used in a number of surface-based measurement programs. One system was mounted on a ship doing a latitudinal survey in the Atlantic Ocean and monitored NO_2, formaldehyde (HCHO), and H_2O_2. Another TDL system was located at the Mauna Loa Observatory and was used to monitor HCHO and H_2O_2 during a photochemistry experiment, finding much lower concentrations of both gases than models had predicted. The TDL system used to measure HCHO has been used in a number of additional ground-based measurement programs and has also been flown on an aircraft in a couple of tropospheric missions.

One TDL system, called DACOM (Differential Absorption CO Measurement), has been used in a large number of NASA GTE missions. It makes measurements of CO in the 4.7 μm spectral region and CH_4 in the 3.3 or 7.6 μm spectral region. This instrument is able to make measurements at a 1 Hz rate and with a precision of 0.5–2.0%, depending on the CO value, and an accuracy of ±2%. DACOM has been very useful in determining global distributions of CO due to the wide-ranging nature of the GTE missions. It has also contributed to

the characterization and understanding of CO in air masses encountered during the GTE missions (Table 7).

A pair of TDL instruments, the Airborne Tunable Laser Absorption Spectrometer (ATLAS) and the Aircraft (ER-2) Laser Infrared Absorption Spectrometer (ALIAS), have been flown on several NASA missions to explore polar O_3 chemistry and atmospheric transport. The ATLAS instrument measures N_2O, and ALIAS is a 4-channel spectrometer that measures a large variety of gases, including HCl, N_2O, CH_4, NO_2, and CO and is currently configured to measure water isotopes across the tropopause.

A new class of lasers, tunable quantum-cascade (QC) lasers, are being added to the list of those available for *in situ* gas measurement systems. Using a cryogenically cooled QC laser during a series of 20 aircraft flights beginning in September 1999 and extending through March 2000, measurements were made of CH_4 and N_2O up to ~20 km in the stratosphere over North America, Scandinavia, and Russia, on the NASA ER-2. Compared with its companion lead salt diode lasers, that were also flown on these flights, the single-mode QC laser, cooled to 82 K, produced higher output power (10 mW), narrower laser linewidth (17 MHz), increased measurement precision (a factor of 3), and better spectral stability (~0.1 cm^{-1} K). The sensitivity of the QC laser channel was estimated to correspond to a minimum-detectable mixing ratio of approximately 2 ppbv of CH_4.

Laser Long-Path Measurements

Laser systems can also be used in long-path measurements of gases. The most important of such programs entailed the measurement of hydroxyl radical (OH) in the Rocky Mountains west of

Table 7 Examples of significant contributions of airborne tunable diode laser systems to the understanding of atmospheric chemistry and transport

B [illegible]
The first [illegible] measurements of the suite NO_2, NO, O_3, and the NO_2 photolysis rate to test NO_x (NO_2 + NO) photochemistry
The first [illegible] stratospheric measurements of NO_x over a full diurnal cycle to test N_2O_5 chemistry
J [illegible] measurements of NO_2 and HNO_3 over the 20–35 km region to assess the effect of Mt. Pinatubo aerosol on heterogeneous atmospheric chemistry
Measurements of HNO_3 and HCl near 30 km
Measurements of CH_4, HNO_3, and N_2O for validation of several satellite instruments
Intrusions from midlatitude stratosphere to tropical stratospheric reservoir
A [illegible]
C [illegible]
Measurement of CO from biomass burning from Asia, Africa, Canada, Central America, and South America
Detection of thin layers of CO that were transported thousands of miles in the upper troposphere
Observed very high levels of urban pollution in plumes off the Asian continent
Emission indices for many gases have been calculated with respect to CO
M [illegible]
Determined CH_4 flux over the Arctic tundra, which led to rethinking of the significance of tundra regions as a global source of CH_4
Found that biomass burning is a significant source of global CH_4
Found strong enhancements of CH_4 associated with urban plumes
A [illegible]
R [illegible]
Extreme denitrification observed in Antarctic winter vortex from NO_y:N_2O correlation study
Observed very low N_2O in Antarctic winter vortex, which helped refute the theory that the O_3 hole is caused by dynamics
Contributed to the study of transport out of the lower stratospheric Arctic vortex by Rossby wave breaking
Measurement of concentrations of gases involved in polar stratospheric O_3 destruction and production
Activation of chlorine in the presence of sulfate aerosols
M [illegible]
Measurements in aircraft exhaust plumes in the lower stratosphere, especially of reactive nitrogen species and CO
Vertical profiles of CO in the troposphere and lower stratosphere
Determination of the hydrochloric acid and the chlorine budget of the lower stratosphere
Measurement of NO_2 for testing atmospheric photochemical models
Trends in HCl/Cl_y in the stratosphere <21 km, 1992–1998
Gas concentration measurements for comparison with a balloon-borne Fourier transform spectrometer observing the Sun
Near-IR TDL laser hygrometers (ER-2, WB57, DC-8) for measuring H_2O and total water in the lower stratosphere and upper troposphere
Remnants of Arctic winter vortex detected many months after breakup
[illegible]
Tropical entrainment time scales inferred from stratospheric N_2O and CH_4 observations

Boulder, Colorado. OH was measured using a XeCl excimer laser operating near 308 nm and transmitting a beam to a retroreflector 10.3 km away. The measurements were quite difficult to conduct, primarily since OH abundance is very low (10^5–$10^7\,cm^{-3}$), yielding very low absorption (~0.02% for an abundance of $10^5\,cm^{-3}$) over the 20.6 km path. In addition, the excimer laser could generate OH from ambient H_2O and O_3, unless the laser energy density was kept low. To help ensure good measurements, a white light source was also employed during the measurements to monitor H_2O, O_3, and other gases. The measurements were eventually very successful and led to new values for OH abundances in the atmosphere.

Laser-Induced Fluorescence (LIF)

The laser-induced fluorescence approach has been used to measure several molecular and ionic species *in situ*. The LIF approach has been used to measure OH and HO_2 (HO_x) on the ground and on aircraft platforms. One airborne LIF system uses a diode-pumped Nd:YAG-pumped, frequency-doubled dye laser to generate the required energy near 308 nm. The laser beam is sent into a White cell where it can make 32–36 passes through the gas in the cell to increase the LIF signal strength. NO is used to convert HO_2 to OH. The detection limit in 1 minute of about 2–3 ppqv (10^{-15}) above 5 km altitude, which translates into a concentration of about 4×10^4 molec/cm^3 at 5 km and 2×10^4 molec/cm^3 at 10 km altitude. One of the interesting findings from such measurements is that HO_x concentrations are up to 5 times larger than model predictions based on NO_x concentrations, suggesting that NO_x emissions from aircraft could have a greater impact on O_3 production than originally thought.

NO is detected using the LIF technique in a two-photon approach: electrons are pumped from the ground state using 226 nm radiation and from that state to an excited state using 1.06 μm radiation. The 226 nm radiation is generated by frequency doubling a dye laser to 287 nm and then mixing that with 1.1 μm radiation derived from H_2 Raman shifting of frequency-mixed radiation from a dye laser and a Nd:YAG laser. From the excited level, 187–201 nm radiation is emitted. In order to measure NO_2, it is first converted to the photofragment NO via pumping at 353 nm from a XeF excimer laser. One of the interesting findings from airborne measurements in the South Pacific is that there appeared to be a large missing source for NO_x in the upper troposphere.

Summary

DIAL and Raman lidar systems have played important roles in studying the distribution of gases such as O_3 and H_2O on local, regional, and global scales, while TDL systems have played corresponding roles for such gases as CO, HCl, and oxides of nitrogen. It is anticipated that these approaches will continue to yield valuable information on these and other gases, as new and more capable systems are developed. Within the next decade, it is expected that a DIAL system will be placed in orbit to make truly global measurements of O_3, H_2O, and/or carbon dioxide.

See also

Imaging: Lidar. **Scattering:** Raman Scattering.

Further Reading

Browell EV (1994) Remote sensing of trace gases from satellites and aircraft. In: Calvert J (ed.) *Chemistry of the Atmosphere: The Impact on Global Change*, pp. 121–134. Cambridge, MA: Blackwell Scientific Publications.

Browell EV, Ismail S and Grant WB (1998) Differential absorption lidar (DIAL) measurements from air and space. *Applied Physics. B* 67: 399–410.

Dabas A, Loth C and Pelon J (eds) (2001) *Advances in Laser Remote Sensing*. Selected Papers presented at the 20th International Laser Radar Conference (ILRC), Vichy, France, 10–14 July 2000, pp. 357–360. Palaiseau Cedex, France: A. Edition d'Ecole Polytechnique.

Goldsmith JEM, Bisson SE, Ferrare RA, *et al.* (1994) Raman lidar profiling of atmospheric water vapor: simultaneous measurements with two collocated systems. *Bulletin American Meteorological Society* 75: 975–982.

Goldsmith JEM, Blair FH, Bisson SE and Turner DD (1998) Turn-key Raman lidar for profiling atmospheric water vapor, clouds, and aerosols. *Applied Optics* 37: 4979–4990.

Grant WB (1995) Lidar for atmospheric and hydrospheric studies. In: Duarte F (ed.) *Tunable Laser Applications*, pp. 213–305. New York: Marcel Dekker Inc.

Grant WB, Browell EV, Menzies RT, Sassen K and She C-Y (eds) (1997) *Laser Applications in Remote Sensing*. SPIE Milestones series, 690 pp, 86 papers.

Grisar R, Booettner H, Tacke M and Restelli G (1992) *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*. Proceedings of the International Symposium held in Freiburg, Germany, 17–18 October 1991, 372 pp. Dordrecht, The Netherlands: Kluwer Academic Publishers.

McDermid IS, Walsh TD, Deslis A and White M (1995) Optical systems design for a stratospheric lidar system. *Applied Optics* 34: 6201–6210.

Mégie G (1988) Laser measurements of atmospheric trace constituents. In: Measures RM (ed.) *Laser Remote Chemical Analysis*, pp. 333–408. New York: J Wiley & Sons.

Podolske JR and Loewenstein M (1993) Airborne tunable diode laser spectrometer for trace gas measurements in the lower stratosphere. *Applied Optics* 32: 5324–5330.

Pougatchev NS, Sachse GW, Fuelberg HE, *et al.* (1999) Pacific Exploratory Mission-Tropics carbon monoxide measurements in historical context. *Journal of Geophysics Research* 104: 26,195–26,207.

Sachse GW, Collins Jr JE, Hill GF, *et al.* (1991) Airborne tunable diode laser sensor for high precision concentration and flux measurements of carbon dioxide and methane. *Proceedings of the SPIE International Society for Optical Engineering* 1443: 145–156.

Svanberg S (1994) Differential absorption lidar (DIAL). In: Segrist MW (ed.) *Air Monitoring Spectroscopic Techniques*, pp. 85–161. New York: J Wiley & Sons.

Webster CR, *et al.* (1994) Aircraft (ER-2) laser infrared absorption spectrometer (ALIAS) for in situ measurements of HCl, N_2O, CH_4, NO_2, and HNO_3. *Applied Optics* 33: 454–472.

Whiteman DN, Melfi SH and Ferrare RA (1992) Raman lidar system for the measurement of water vapor and aerosols in the Earth's atmosphere. *Applied Optics* 31: 3068–3082.

Zanzottera E (1990) Differential absorption lidar techniques in the determination of trace pollutants and physical parameters of the atmosphere. *Critical Reviews in Analytical Chemistry* 21: 279–319.

Optical Transmission and Scatter of the Atmosphere

S M Adler-Golden and A Berk,
Spectral Sciences, Inc., Burlington, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The ability to understand and model radiative transfer (RT) processes in the atmosphere is critical for remote sensing, environmental characterization, and many other areas of scientific and practical interest. At the Earth's surface, the bulk of this radiation, which originates from the sun, is found in the ultraviolet to infrared range between around 0.3 and 4 μm. This article describes the most significant RT processes for these wavelengths, which are absorption (light attenuation along the line of sight LOS) and elastic scattering (redirection of the light). Transmittance, T, is defined as one minus the fractional extinction (absorption plus scattering). At longer wavelengths (in the mid- and long-wave infrared) the major light source is thermal emission.

A few other light sources are mentioned here. The moon is the major source of visible light at night. Forest fires can be a significant source of mid-wave infrared radiation. Manmade light sources include continuum sources such as incandescent lamps and spectrally narrow sources such as fluorescent lamps and lasers. In general, spectrally narrow sources need to have a different RT treatment than continuum sources due to the abundance of narrow spectral absorption lines in the atmosphere, as is discussed below.

The challenge of atmospheric RT modeling is essentially to solve the following equation that describes monochromatic light propagation along the LOS direction:

$$\left(\frac{1}{k}\right)\frac{\mathrm{d}I}{\mathrm{d}u} = -I + J \qquad [1]$$

where I is the LOS radiance (watts per unit area per unit wavelength per steradian), k is the extinction coefficient for the absorbing and scattering species (per unit concentration per unit length), u is the material column density (in units of concentration times length), and J is the radiance source function. I is a sum of direct (i.e., from the sun) and diffusely scattered components. The source function represents the diffuse light scattered into the LOS, and is the angular integral over all directions Ω_i of the product of the incoming radiance, $I(\Omega_i)$, and the scattering phase function, $p(\Omega_o, \Omega_i)$:

$$J(\Omega_o) = \int_{\Omega_i} p(\Omega_o, \Omega_i) I(\Omega_i) \mathrm{d}\Omega_i \qquad [2]$$

The scattering phase function describes the probability density for incoming light from direction Ω_i scattering out at angle Ω_o, and is a function of the difference (scattering) angle θ.

The direct radiance component, I', is described by eqn [1] with the source function omitted. Integrating along the LOS leads to the well-known Beer's Law

equation for transmittance:

$$T = I'/I'_0 = \exp(-ku) \qquad [3]$$

where I'_0 is the direct radiance at the boundary of the LOS. The quantity $ku = \ln(1/T)$ is known as the optical depth.

Atmospheric Constituents

The atmosphere has a large number of constituents, including numerous gaseous species and suspended liquid and solid particulates. Their contributions to extinction are depicted in Figure 1. The largest category is gases, of which the most important are water vapor, carbon dioxide, and ozone. Of these, water vapor is the most variable and carbon dioxide the least, although the CO_2 concentration is gradually increasing. In atmospheric RT models, carbon dioxide is frequently taken to have a fixed and altitude-independent concentration, along with other 'uniformly mixed gases' (UMGs). The concentration profiles of the three major gas species are very different. Water vapor is located mainly in the lowest 2 km of the atmosphere. Ozone has a fairly flat profile from the ground through the stratosphere (~30 km). The UMGs decline exponentially with altitude, with a scale height of around 8 km.

Gases

Gas molecules both scatter and absorb light. Rayleigh scattering by gases scales inversely with the fourth power of the wavelength, and is responsible for the sky's blue color. For typical atmospheric conditions, the optical depth for Rayleigh extinction is approximately $0.009/\lambda^4$ per air mass (λ is in μm and air mass is defined by the vertical column from ground to space). The Rayleigh phase function has a $(1 + \cos^2\theta)$ dependence.

Absorption by gases may consist of a smooth spectral continuum (such as the ultraviolet and visible electronic transitions of ozone) or of discrete spectral lines, which are primarily rotation lines of molecular vibrational bands. In the lower atmosphere (below around 25 km altitude), the spectral shape of these lines is determined by collisional broadening and described by the normalized Lorentz line shape formula:

$$f_{\text{Lorentz}}(\nu) = \frac{\alpha_c/\pi}{\alpha_c^2 + (\nu - \nu_0)^2} \qquad [4]$$

Here ν is the wavenumber (in cm^{-1}) [$\nu = (10\,000\ \mu\text{m}/\text{cm})/\lambda$], ν_0 is the molecular line transition frequency and α_c is the collision-broadened half-width (in cm^{-1}), which is proportional to pressure. The constant of proportionality, known as the pressure-broadening parameter, has a typical value on the order of $0.06\ \text{cm}^{-1}\ \text{atm}^{-1}$ at ambient temperature. The extinction coefficient $k(\nu)$ is the product of $f_{\text{Lorentz}}(\nu)$ and the integrated line strength S, which is commonly in units of $\text{atm}^{-1}\ \text{cm}^{-2}$.

At higher altitudes in the atmosphere the pressure is reduced sufficiently that Doppler broadening

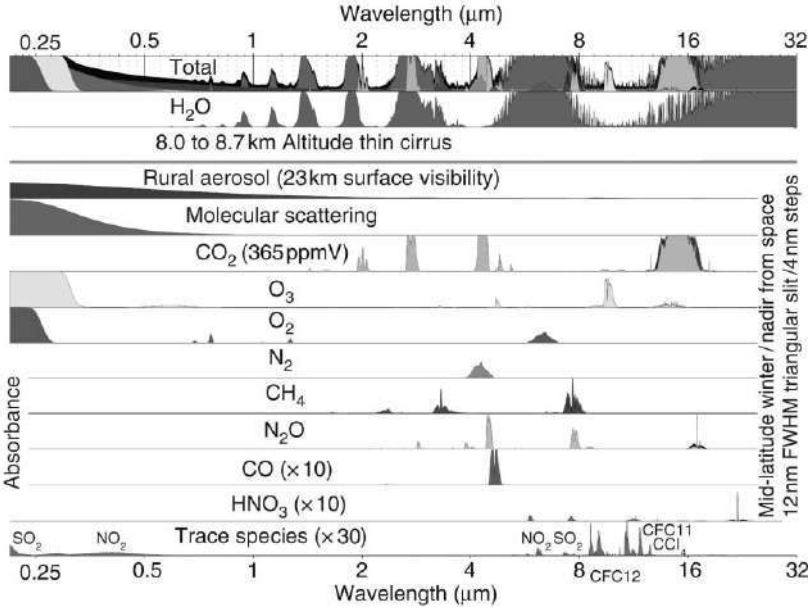

Figure 1 Spectral absorbance (1 – transmittance) for the primary sources of atmospheric extinction. The 12 nm resolution data were generated by MODTRAN for a vertical path from space with a mid-latitude winter model atmosphere.

becomes competitive with collisional broadening, and the Lorentz formula becomes inaccurate. The general lineshape for combined collisional and Doppler broadening is the Voigt lineshape, which is proportional to the real part of the complex error (probability) function, w:

$$f_{\mathrm{Voigt}}(\nu) = \frac{1}{\alpha_{\mathrm{d}}\sqrt{\pi}}\mathrm{Re}\left[w\left(\frac{\nu-\nu_0}{\alpha_{\mathrm{d}}} + \frac{\alpha_{\mathrm{c}}}{\alpha_{\mathrm{d}}}\mathrm{i}\right)\right] \qquad [5]$$

Here, α_{d} is the Doppler $1/e$ half-width.

Comprehensive spectral databases have been compiled of the transition frequencies, strengths, and pressure-broadened half-widths for atmospherically important molecules throughout the electromagnetic spectrum. Perhaps the most notable of these databases is HITRAN, which was developed by the US Air Force Research Laboratory and is currently maintained at the Harvard-Smithsonian Center for Astrophysics in Cambridge, MA.

Liquids and Solids

The larger particulates in the atmosphere (greater than a few μm in radius) typically belong to clouds. Low-altitude clouds consist of nearly spherical water droplets, while high-altitude cirrus clouds are mainly a collection of ice crystals. Other large particulates include sand dust. Their light scattering is close to the geometric limit at visible and ultraviolet wavelengths. This means that the extinction is nearly wavelength-independent, and the scattering phase function and single-scattering albedo may be reasonably modeled with ray-tracing techniques that account for the detailed size and shape distributions of the particles. However, Mie scattering theory is typically used to calculate cloud optical properties because it is exact for spherical particles of any size.

The smaller particulates in the atmosphere belong to aerosols, which are very fine liquid particles, and dusts, which are solids such as minerals and soot. These particulates are concentrated mainly in the lower 2 km or so of the atmosphere; however, they are also present at higher altitudes in smaller concentrations. Their optical properties are typically modeled using Mie theory. The wavelength dependence of the scattering is approximately inversely proportional to a low power of the wavelength, typically between 1 and 2, as befits particulates intermediate in size between molecular and geometric-limit. The scattering phase functions have a strong forward-scattering peak; values of the asymmetry parameter g (the average value of $\cos\theta$) typically range from 0.6 to 0.8 at solar wavelengths.

Solution Methods

Geometry

Atmospheric properties are primarily a function of altitude, which determines pressure, temperature and species concentration profiles. Accordingly, most RT methods define a stratified atmosphere. The most accurate treatments of transmission and scattering account for the spherical shape of the layers and refraction; however, most RT models use a plane-parallel approximation for at least some computations, such as multiple scattering.

Spectral Resolution

Optical instruments have finite, and frequently broad, wavelength responses. Nevertheless, modeling their signals requires accounting for the variation of absorption on an extremely fine wavelength scale, smaller than the widths of the molecular lines.

'Exact' monochromatic methods

The most accurate RT solution method involves explicitly solving the RT problem for a very large number of monochromatic wavelengths. This line-by-line method is used in a number of RT models, such as FASCODE. It allows Beer's law to be applied to combine transmittances from multiple LOS segments, and provides an unambiguous definition of the optical parameters. It is suitable for use with spectrally structured light sources, such as lasers. The one major drawback of this method is that it is computationally intensive, and therefore may not be practical for problems where large wavelength ranges, multiple LOS views and multiple atmospheric conditions need to be treated.

To alleviate the computational burden of monochromatic calculations, some approximate methods have been developed that model RT in finite spectral intervals, as described below.

Statistical band models

The band model method represents spectral lines in a narrow interval, $\Delta\nu$, statistically using such parameters as the total line strength, the mean pressure-broadening parameter, and the effective number of lines in the interval. An example of a popular band model-based RT algorithm is MODTRAN, which is described below.

A key to the success of band models is the availability of approximate analytical formulas for the integrated absorption for an individual molecular transition of strength S, known as the single-line total

equivalent width, W_{sl}:

$$W_{sl} = \int_{-\infty}^{\infty} [1 - \exp(-Suf_{\nu})]\mathrm{d}\nu \qquad [6]$$

In the optically thin (small absorption) limit, W_{sl} is proportional to the molecular species column density, while in the optically thick (large absorption) limit it scales as the square root of the column density. A further assumption made by MODTRAN is that the line centers are randomly located within the interval, i.e., spectrally uncorrelated. With this assumption, the net transmittance can be expressed as the product of the transmittances for each individual line, whether the line belongs to the same molecular species or a different species.

A particular challenge in band models is treatment of the inhomogeneous path problem – that is, the variations in path properties along a LOS and their effect on the statistical line parameters, which arise primarily from differences in pressure and hence line width. The Curtis–Godson path averaging method provides a reasonable way to define 'equivalent' homogeneous path parameters for the band model. Another challenge is to define an effective extinction optical depth for each layer in order to solve the RT problem with scattering. One option in MODTRAN is to define it by computing the cumulative transmittance through successive layers in a vertical path.

Correlated-*k* model

Another well-known approximate RT algorithm for spectral intervals is the correlated-*k* method. This method starts with an 'exact' line-by-line calculation of extinction coefficients (k's) within the interval on a fine spectral grid, from which a *k*-distribution (vs. cumulative probability) is computed. A database of *k* values and probabilities summing to 1 is built from these *k*-distributions for a grid of atmospheric pressures and temperatures, and for all species contributing to the spectral interval. Inhomogeneous paths are handled by recognizing that the size order of the k's is virtually independent of pressure and typically only weakly dependent on temperature. LOS transmittances, LOS radiances, and fluxes are calculated by interpolating the database over temperature and pressure to define the k's for each LOS segment, solving the monochromatic RT equation at each fixed distribution location, and finally integrating over the distribution. The correlated-*k* method has been found to be quite accurate for atmospheric paths containing a single molecular species; however, corrections must be applied for spectral intervals containing multiple species.

Scattering Methods

When the diffuse light field is of interest, scattering methods are used to calculate the source function J of eqns [1] and [2].

Single scattering

If scattering is weak, the approximation may be made that the solar radiation scatters only once. Thus the integral over the scattering phase function, eqn [2], is straightforwardly calculated using the direct radiance component, which is given by eqn [3]. The neglect of multiple scattering (i.e., the diffuse contribution to the source function) means that the diffuse radiance is underestimated; however, single scattering is sufficiently accurate for some atmospheric problems in clear weather and at infrared wavelengths.

Multiple scattering

A number of different methods have been developed to solve the multiple scattering problem. Two-stream methods, which are the simplest and fastest, resolve the radiance into upward and downward directions. These methods generally produce reasonably accurate values of hemispherically averaged radiance, which are also referred to as horizontal fluxes or irradiances.

A much more accurate approach to the multiple scattering problem is the method of discrete ordinates. It involves expansion of the radiation field, the phase function and the surface reflectance as a series of spherical harmonics, leading to a system of linear integral equations. Evaluation of the integrals by Gaussian quadrature leads to a solvable system of linear differential equations. An important approximation called delta-M speeds up convergence of the discrete ordinates method, especially when scattering phase functions are strongly forward peaked, by representing the phase function as the sum of a forward direction δ-function and a remainder term. For most scattering problems, the solution is converged upon with a modest number (~8 to 16) of quadrature points (streams).

A very different type of multiple scattering technique, called the Monte Carlo method, is based on randomly sampling a large number of computer-simulated 'photons' as they travel through the atmosphere and are absorbed and scattered. The basic idea here is that sensor radiance can be expressed as a multiple path integral over the local source terms, and Monte Carlo methods solve integrals by sampling the integrand. The major advantage of this method is that it is flexible enough to allow for all of the complexity of a realistic atmosphere, often neglected by other methods. The major drawback is its

computational burden, as a large number of photons is required for reasonable convergence; the Gaussian error in the calculation declines with the square root of the number of photons. The convergence problem is moderated to a large extent by using the physics of the problem being solved to bias the selection of photon paths toward those trajectories which contribute most; mathematically, this is equivalent to requiring that the integrand be sampled most often where its contributions are most significant.

An Example Atmospheric RT Model: MODTRAN

MODTRAN, developed collaboratively by the Air Force Research Laboratory and Spectral Sciences, Inc., is the most widely used atmospheric radiation transport model. It defines the atmosphere using stratified layering and computes transmittances, radiances, and fluxes using a moderate spectral resolution band model with IR through UV coverage. The width of the standard spectral interval, or bin, in MODTRAN is 1 cm^{-1}. At this resolution, spectral correlation among extinction sources is well characterized as random. Thus, the total transmittance from absorption and scattering of atmospheric particulates and molecular gases is computed as the product of the individual components.

Rayleigh, aerosol, and cloud extinction are all spectrally slowly varying and well represented by Beer's Law absorption and scattering coefficients on a 1 cm^{-1} grid. Calculation of molecular absorption is more complex because of the inherent spectral structure and the large number of molecular transitions contributing to individual spectral bins. As illustrated in Figure 2, MODTRAN partitions the spectral bin molecular attenuation into 3 components:

- *Line center absorption* from molecular transitions centered within the spectral bin;
- *Line tail absorption* from the tails of molecular lines centered outside of the spectral bin but within 25 cm^{-1}; and
- H_2O and CO_2 *continuum absorption* from distant (>25 cm^{-1}) lines.

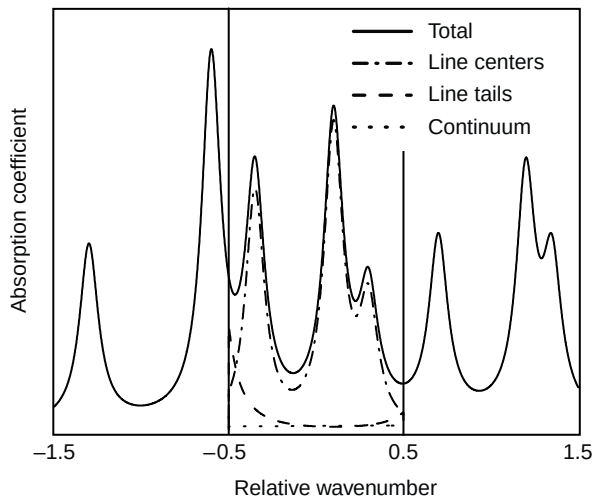

Figure 2 Components of molecular absorption. The line center, line tail and continuum contributions to the total absorption are illustrated for the central 1 cm^{-1} spectral bin.

Within the terrestrial atmosphere, only H_2O and CO_2 have sufficient concentrations and line densities to warrant inclusion of continuum contributions. These absorption features are relatively flat and accurately modeled using 5 cm^{-1} spectral resolution Beer's Law absorption coefficients.

Spectral bin contributions from neighboring line tails drop off, often rapidly, from their spectral bin edge values, but the spectral curves are typically simple, containing at most a single local minimum. For MODTRAN, these spectral contributions are pre-computed for a grid of temperature and pressure values, and fit with Padé approximants, specifically the ratio of quadratic polynomials in wavenumber. These fits are extremely accurate and enable line tail contributions to be computed on an arbitrarily fine grid. MODTRAN generally computes this absorption at a resolution equal to one-quarter the spectral bin width, i.e., 0.25 cm^{-1} for the 1.0 cm^{-1} band model.

The most basic *ansatz* of the MODTRAN band model is the stipulation that molecular line center absorption can be approximated by the absorption of n identical Voigt lines randomly distributed within the band model spectral interval, $\Delta\nu$. Early in the development of RT theory, Plass derived the expression for the transmittance from these n randomly distributed lines:

$$T = \left(1 - \frac{W'_{sl}}{\Delta\nu}\right)^n \qquad [7]$$

MODTRAN's evolution has resulted in a fine tuning of the methodology used to define both the effective line number n and the in-band single-line equivalent width W'_{sl}. The effective line number is initially estimated from a relationship developed by Goody, in which lines are weighted according the to square root of their strength, but MODTRAN combines nearly degenerate transitions into single lines because these multiplets violate the random distribution assumption. The initial effective line number values are refined to insure a match with higher resolution transmittance predictions degraded to the band model resolution. The in-band equivalent

width is computed for an off-centered Voigt line of strength S. Lorentz and Doppler half-widths are determined as strength-weighted averages. The off-center distance is fixed to insure that the weak-line Lorentz equivalent width exactly equals the random line center value.

MODTRAN scattering calculations are optimally performed using the DISORT discrete ordinates algorithm developed by Stamnes and co-workers. Methods for computing multiple scattering such as DISORT require additive optical depths, i.e., Beer's Law transmittances. Since the in-band molecular transmittances of MODTRAN do not satisfy Beer's Law, MODTRAN includes a correlated-k algorithm option. The basic band model *ansatz* is re-invoked to efficiently determine k-distributions; tables of k-data are pre-computed as a function of Lorentz and Doppler half-widths and effective line number, assuming spectral intervals contain n randomly distributed identical molecular lines. Thus, MODTRAN k-distributions are statistical, only dependent on band model parameter values, not on the exact distribution of absorption coefficients in each spectral interval.

Applications

Among the many applications of atmospheric transmission and scattering calculations, we briefly describe two complementary ones in the area of remote sensing, which illustrate many of the RT features discussed earlier as well as current optical technologies and problems of interest.

Earth Surface Viewing

The first example is Earth surface viewing from aircraft or spacecraft with spectral imaging sensors. These include hyperspectral sensors, such as AVIRIS, which have typically a hundred or more contiguous spectral channels, and multispectral sensors, such as Landsat, which typically have between three and a few tens of channels. These instruments are frequently used to characterize the surface terrain, materials and properties for such applications as mineral prospecting, environmental monitoring, precision agriculture, and military uses. In addition, they are sensitive to properties of the atmosphere such as aerosol optical depth and column water vapor. Indeed, in order to characterize the surface spectral reflectance, it is necessary to characterize and remove the extinction and scattering effects of the atmosphere.

Figure 3 also shows an example of data collected by the AVIRIS sensor at ~3 km altitude over thick vegetation. The apparent reflectance spectrum is the observed radiance divided by the Top-of-Atmosphere horizontal solar flux. Atmospheric absorption by water vapor, oxygen, and carbon dioxide is evident, as well as Rayleigh and aerosol scattering. Figure 3 shows the surface reflectance spectrum inferred by modeling and then removing these atmospheric effects (this process is known as atmospheric removal, compensation, or correction). It has the characteristic smooth shape expected of vegetation, with strong chlorophyll absorption in the visible and water bands at longer wavelengths. A detailed analysis of such a spectrum may yield information on the vegetation type and its area coverage and health.

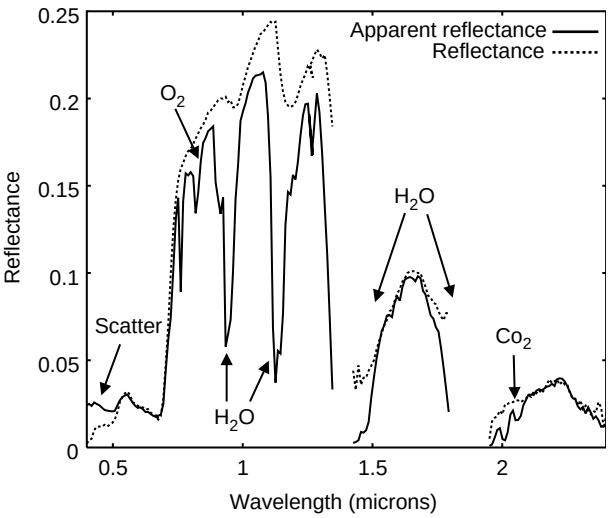

Figure 3 Vegetation spectrum viewed from 3 km altitude before (apparent reflectance) and after (reflectance) removal of atmospheric scatter and absorption.

Sun and Sky Viewing

The second remote sensing example is sun and sky viewing from the Earth's surface with a spectral radiometer, which can yield information on the aerosol content and optical properties as well as estimates of column concentrations of water vapor, ozone, and other gases. Figure 4 shows data from a Yankee Environmental Systems, Inc. multi-filter rotating shadow-band radiometer, which measures both 'direct flux' (the direct solar flux divided by the cosine of the zenith angle) and diffuse (sky) flux in narrow wavelength bands. The plot of ln(direct signal) versus the air mass ratio is called a Langley plot, and is linear for most of the bands, illustrating Beer's Law. The extinction coefficients (slopes) vary with wavelength consistent with a combination of Mie and Rayleigh scattering. The water-absorbing 940 nm band has the lowest values and a curved

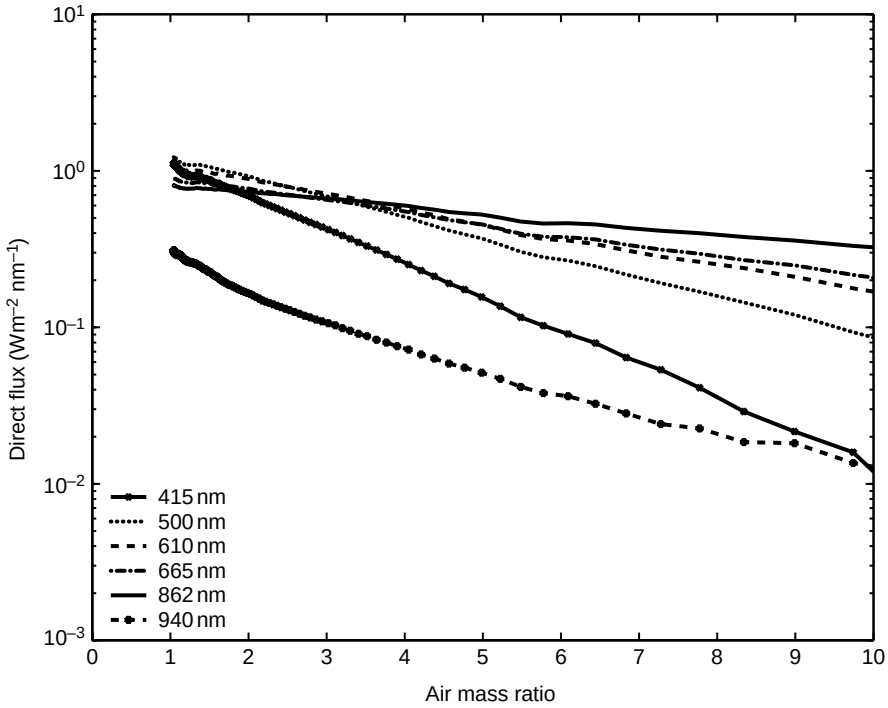

Figure 4 Direct solar flux versus air mass at the surface measured by a Yankee Environmental Systems, Inc. multi-filter rotating shadow-band radiometer at different wavelengths.

plot, in accordance with the square-root dependence of the equivalent width for optically thick lines. The diffuse fluxes, which arise from aerosol and Rayleigh scattering, have a very different dependence on air mass than the direct flux. In particular, the diffuse contributions often increase with air mass for high sun conditions.

The ratio of the diffuse and direct fluxes is related to the total single-scattering albedo, which is defined as the ratio of the total scattering coefficient to the extinction coefficient. Results from a number of measurements showing lower than expected diffuse-to-direct ratios have suggested the presence of black carbon or some other continuum absorber in the atmosphere, which would have a significant impact on radiative energy balance at the Earth's surface and in the atmospheric boundary layer.

List of Units and Nomenclature

Line shape function [cm]	f_ν
Lorentz line shape function [cm]	$f_{\mathrm{Lorentz}}(\nu)$
Voigt line shape function [cm]	$f_{\mathrm{Voigt}}(\nu)$
Scattering asymmetry parameter	g
Direct plus diffuse radiance (W cm^{-1} sr^{-1} or W cm^{-2} μm^{-1} sr^{-1})	I
Direct radiance (W cm^{-1} sr^{-1} or W cm^{-2} μm^{-1} sr^{-1})	I'
Direct radiance at the LOS boundary (W cm^{-1} sr^{-1} or W cm^{-2} μm^{-1} sr^{-1})	I'_0
Source function (W cm^{-1} sr^{-1} or W cm^{-2} μm^{-1} sr^{-1})	J
Extinction coefficient (cm^{-1} atm^{-1})	k
Effective number of lines in bin	n
Scattering phase function (sr^{-1})	p
Line strength (cm^{-2} atm^{-1})	S
Transmittance	T
Column density (atm cm)	u
Complex error (probability) function	w
Single-line total equivalent width (cm^{-1})	W_{sl}
Single-line in-band equivalent width (cm^{-1})	W'_{sl}
Collision-broadened half-width at half maximum (cm^{-1})	α_{c}
Doppler half-width at $1/e$ (cm^{-1})	α_{d}
Spectral interval (bin) width (cm^{-1})	$\Delta\nu$
Scattering angle (radian)	θ
Wavelength (μm)	λ
Wavenumber (cm^{-1})	ν
Molecular line transition frequency	ν_0
Direction of incoming light (sr)	Ω_{i}
Direction of outgoing light (sr)	Ω_{o}

See also

Environmental Measurements: Laser Detection of Atmospheric Gases. **Instrumentation:** Spectrometers. **Scattering:** Scattering from Surfaces and Thin Films.

Further Reading

Berk A, Anderson GP, Acharya PK, Bernstein LS, Shettle EP, Adler-Golden SM, Lee J and Muratov L (2004) MODTRANS User's Manual. Airforce Research Laboratory, Hanscom AFB, MA, in press.

Chandrasekhar S (1960) *Radiative Transfer*. New York: Dover Publications, Inc., Originally published by Oxford University Press, London (1950).

Fenn RW, Clough SA, Gallery WO, Good RE, Kneizys FX, Mill JD, Rothman LS, Shettle EP and Volz FE (1985) Optical and infrared properties of the atmosphere. In: Jursa AS (ed.) *Handbook of Geophysics and the Space Environment*, chap. 18, US Air Force Geophysics Laboratory, Hanscom AFB, MA, AFGL-TR-85-0315, available from NTIS ADA 167000.

Goody RM and Yung YL (1989) *Atmospheric Radiation, Theoretical Basis*, 2nd edn. New York: Oxford University Press.

HITRAN Special Issue, J. Quant. Spectrosc. Rad. Transfer, vol. 82, Nos. 1–4, 15 November–15 December (2003).

Houghton J (2001) *The Physics of Atmospheres*, 3rd edn, Cambridge, UK: Cambridge University Press.

Killinger DK, Churnside JH and Rothman LS (1994) Atmospheric optics. In *Handbook of Optics*, 2nd edn, 2-Volume Set, chap. 44. New York: McGraw-Hill.

Liou KN (1997) *An Introduction to Atmospheric Radiation*, 2nd edn. London: Academic Press.

Smith FG (ed.) (1993) *The Infrared and Electro-Optical Systems Handbook*, vol. 2, *Atmospheric Propagation of Radiation*. Ann Arbor: Environmental Research Institute of Michigan.

Thomas GE and Stamnes K (1999) *Radiative Transfer in the Atmosphere and Ocean*. Cambridge, UK: Cambridge University Press.

Van de Hulst HC (1981) *Light Scattering by Small Particles*. New York: Dover Publications, Inc., Originally published in 1957, New York: John Wiley & Sons, Inc.

F

FIBER AND GUIDED WAVE OPTICS

Contents

Overview
Dispersion
Fabrication of Optical Fiber
Light Propagation
Measuring Fiber Characteristics
Nonlinear Effects (Basics)
Nonlinear Optics
Optical Fiber Cables
Passive Optical Components

Overview

A Mickelson, University of Colorado, Boulder, CO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

It is said that some in ancient Greece knew that light could be guided inside slabs of transparent material. From this purported little-known curiosity in the ancient world, guided wave optics has grown to be the technology of the physical level of the systems which transfer most of the information about the world; the waveguide of the worldwide telecommunications network is the optical fiber. The optical fiber itself is but a passive waveguide, and guided wave optics is the technology which includes all of the passive and active components which are necessary to prepare optical signals for transmission, regenerate optical signals during transmission, route optical signals through systems and code onto optical carriers, and decode the information from optical carriers, to more conventional forms. In this article, some introduction to this rather encompassing topic will be given.

This article will be separated into four parts. In the first section, discussion will be given to fiber optics, that is, the properties of the light guided in optical waveguides which allow the light to be guided in distinctly nonrectilinear paths over terrestrial distances. The second section will then turn to the components which can be used along with the fiber optical waveguides in order to form useful systems. These components include sources and detectors as well as optical amplifiers. In the third section, we will discuss the telecommunications network which has arisen due to the availability of fiber optics and fiber optic compatible components. The closing section will discuss integrated optics, the field of endeavor which has such great promise to form the future of optical technology.

Fiber Optics

Fiber optics is a term which generally refers to a technology in which light (actually infrared, visible, or ultraviolet radiation) is transmitted through the transparent cores of small threads of composite material. These threads, or fibers as they are called, when surrounded by a cladding material and coated with polymer for environmental protection (see Figure 1) can be coiled like conventional wire and

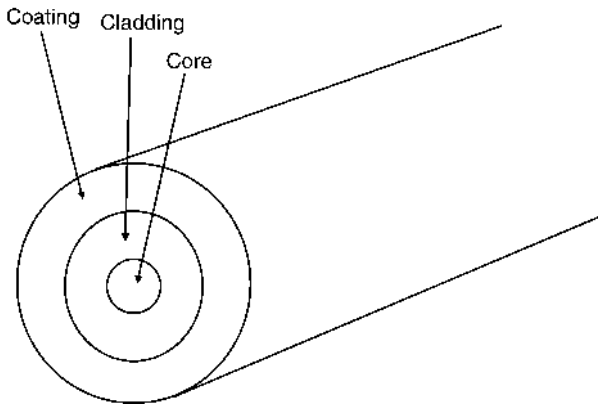

Figure 1 A depiction of the structure of an optical fiber. The innermost cylinder is the core region in which the majority of the light is confined. The next concentric region is the cladding region which is still made of a pure material but one of a lower index of refraction than the innermost core region, such that the light of the lightwave decays exponentially with extension into this region. The outermost region is a coating which protects the fused silica of the core and cladding from environmental contaminants such as water. For telecommunications fibers, the core is comprised of fused silica doped with germanium and typically has an index of refraction of 1.45, whereas the cladding differs from this index by only about 1%. The coating is often a polyimide plastic which has a higher index (perhaps 1.6) but no light guided light should see the cladding-coating interface.

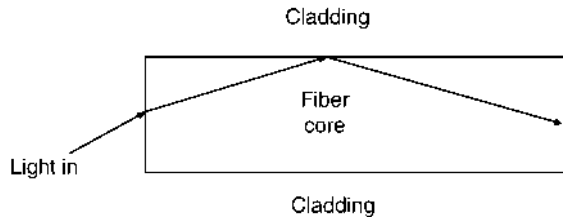

Figure 2 Schematic depiction of the geometrical optics interpretation of the coupling of light into an optical fiber. Here a ray that is incident on the center of the fiber core at an upward angle is first refracted at the fused silica–air interface into the fused silica core. The ray is then totally internally reflected at the core cladding interface such that the ray power remains in the core. The ray picture of light coupling is applicable to single rays coupling to a multimode fiber. When there are multiple rays, interference between these rays must be taken into account which is difficult to do in the ray picture. When there is only one (diffraction limited) mode in the fiber, the interference between a congruence of (spatially coherent) incident rays is necessary to describe the coupling. This description is more easily effected by a quasi-monochromatic mode picture of the propagation.

when cabled can resemble (a very lightweight flexible version of) conventional transmission wire. Although most of the optical fiber in use today is fabricated by a process of gas phase chemical deposition of fused silica doped with various other trace chemicals, fiber can be made from a number of different material systems and in a number of different configurations for use with various types of sources. In what follows in this opening section, we will limit discussion to the basic properties of the light guided by the fiber and leave more technological discussion to following sections.

There are two complementary mathematical descriptions of the propagation of light in an optical waveguide. In the ray description, light incident on a fiber endface is considered to be made up of a bundle of rays. In a uniform homogeneous medium, each ray is like an arrow that exhibits rectilinear propagation from its source to its next interface with a dissimilar material. These rays satisfy Snell's laws of reflection and refraction at interfaces between materials with dissimilar optical properties that they encounter along their propagation path. That is, at an interface, a fraction of the light is reflected backwards at an angle equal to the incident angle and a portion of the light is transmitted in a direction which is more directed toward the normal to the interface when the index increases across the boundary and is directed more away from the normal when the index decreases. At the input endface of a waveguide, a portion of the energy guided by each ray is refracted due to the change in refractive index at the guide surface and then exhibits a more interesting path within the fiber. In a step index optical fiber, where the index of refraction is uniformly higher in the fiber core than in a surrounding cladding, the rays will propagate along straight paths until encountering the core cladding interface. Guided rays (see Figure 2) are totally internally reflected back into the fiber core to again be totally internally reflected at the next core cladding interface and so on. Radiating rays (see Figure 3) will be only partially reflected at the core cladding interface and will rapidly die out in propagating down the fiber when it is taken into account that typical distances between successive encounters with the boundary may be sub-millimeter and total propagation distances may be many kilometers. In graded index fibers, the refractive index within the fiber core varies continuously from a maximum somewhere within the core to a minimum at which the core ends and attaches continuously to a cladding. The ray paths within such fibers are curved and the guided rays are characterized by the fact that once in the fiber they never encounter the cladding. Radiating rays encounter the cladding and are refracted out of the fiber. This description of fiber propagation is quite simple and pleasing but does not take into account that each ray actually is carrying a clock that remembers how long it has been following its given path. When two rays come together, they can

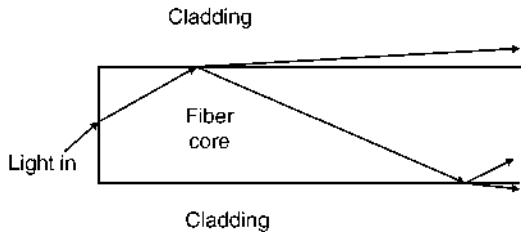

Figure 3 Schematic depiction of the existence of a cut-off angle for coupling into an optical fiber. A ray incident from air on the center of the fused silica fiber core is refracted at the air fused silica interface and then split when it comes to the core cladding interface within the fiber. Were the refracted angle of this ray zero (along the core cladding interface) we would say the ray is totally internally reflected and no power escapes. As it is, there is refracted power into the cladding and the ray will rapidly attenuate as it propagates along the fiber. Although most telecommunications fiber in this day and age is graded index (the rays will curve rather than travel along straight paths), a typical repeat period in such a fiber is a millimeter, whereas propagation distances may be 100 km. With a loss of even only 0.1% at each reflection form the cladding, the light light at cut-off will be severely attenuated in one meter, no loss, one kilometer.

either add or subtract, depending on the reading on their respective clocks. When the light is nearly monochromatic, the clocks' readings are simply a measure of the phase of the ray when it was radiated from its source and the interference between rays can be quite strong. This interference leads to conditions which only allow certain ray paths to be propagated. When the light has a randomly varying phase in both space and time (and is therefore polychromatic), the interference all but disappears. But the condition that allows coupling to a specific ray path (mode) also is obliterated by the phase randomness, and coupling to a single mode guide, for example, becomes inefficient.

When we need to preserve phase relations in order to preserve information while it propagates along a fiber path, we need to use single modes excited by nearly transform limited sources, that is sources whose time variation exhibits well-defined (nonrandom) phase variation. As a monochromatic wave carries no information, we will require a source which can be modulated. When modulation can be impressed on a carrier without loss of phase information of either information or carrier, information can be propagated at longer distances than when the frequency spectrum of the carrier is broadened by noise generated during the modulation. We refer to such a source that can be modulated without broadening as a coherent or nearly coherent source. We cannot easily adapt the ray picture to the description of propagation of such coherent radiation in an optical fiber. Instead we must resort to using the time harmonic form of Maxwell's equations, the equations which describe electromagnetic phenomena, whereas rays can be described by a simplification of the time-dependent form of these equations. In the time harmonic approach, we assume that the source is monochromatic and then solve for a set of modes of the fiber at the assumed frequency of the source. These modes have a well-defined phase progression as a function of carrier frequency and possess a given shape in the plane transverse to the direction of propagation. Information can be included in the propagation by assuming the quasi-monochromatic variation of the source, that is, one assumes that the source retains the phase relations between the various modulated frequency components even while its amplitude and phase is being varied externally. When one assumes that time harmonically (monochromatic) varying fields are propagating along the fiber axis, one obtains solutions of Maxwell's equations in the form of a summation of modes. These are guided modes, ones that propagate down the fiber axis without attenuation. There are also radiation modes that never couple into a propagating mode in the fiber. These modes are analogous to the types of rays we see in the ray description of fiber propagation. There are also another set of modes which are called evanescent modes which show up at discontinuities in the fiber or at junctions between the fiber and other components. In the modal picture of fiber propagation, each of these modes is given an independent complex coefficient (that is, a coefficient with both amplitude and phase). These coefficients are determined first by any sources in the problem and then must be recalculated at each discontinuity plane along the propagation path in the fiber. When the source(s) in the problem is not transform limited, the phases of the coefficients become smeared out and the coupling problems at discontinuities take on radically different solutions.

In some limit, these randomized solutions must appear as the ray solutions. Optical fibers are generally characterized by their numerical aperture (NA), as well as the number which characterizes its transverse dimension. The numerical aperture is essentially the sine of the maximum angle into which the guide will radiate into free space. When the guide is multimoded, then the transverse dimension is generally given as a diameter. When a guide is single-moded, the transverse dimension is generally given as a spotsize, that is, by a measure of the size of a unity magnification image of the fiber endface. Multimode guides can be excited by even poorly coherent sources so long as they radiate into the NA or capture angle at the input of the guide. Single-mode fibers require an excitation which matches the shape and size of their fundamental mode.

Active Fiber Compatible Components

Our discussion of propagation in fiber optic waveguides would be a rather sterile one if there were not a number of fiber compatible components that can generate, amplify, and detect light streams available to us in order to construct fiber optic systems. In this section, we will try to discuss some of the theory of operation of active optical components in order to help elucidate some of the general characteristics of such components and the conditions that inclusion of such components in a fiber system impose on the form that the system must take.

In a passive component such as an optical fiber, we need only consider the characteristics of the light that is guided. Operation of active components requires that the optical fields interact with a medium in such a manner that energy can be transferred from the field through an excitation of the medium to the controlling electrical stream and/or vice versa. Generally one wants the exchange to go only one way, that is from field to electrical stream or from electrical stream to field. In a detector this is not difficult to achieve, whereas in lasers and amplifiers it is quite hard to eliminate the back reaction of the field on the device. Whereas a guiding medium can be considered as a passive homogeneous continuum, the active medium has internal degrees of freedom of its own as well as a granularity associated with the distribution of the microscopic active elements. At least, we must consider the active medium as having an active index of refraction with a mind (set of differential equations anyway) of its own. If one also wants to consider noise characteristics, one needs to consider the lattice of active elements which convert the energy. The wavelength of operation of that active medium is determined by the energy spacing between the upper and lower levels of a transition which is determined by the microscopic structure of these grains, or quanta, that make up the medium. In a semiconductor, we consider these active elements to be so uniformly distributed and ideally spaced that we can consider the quanta (electron hole pairs) to be delocalized but numerous and labeled by their momentum vectors. In atomic media such as the rare earth doped optical fibers which serve as optical amplifiers, we must consider the individual atoms as the players. In the case of the semiconductor, a current, flowing in an external circuit, controls the population of the upper (electronic) and lower (hole) levels of each given momentum state within the semiconductor. When the momentum states are highly populated with electrons and holes, there is a flow of energy to the field (at the transition wavelength) and when the states are depopulated, the field will tend to be absorbed and populate the momentum states by giving up its energy to the medium. The situation is similar with the atomic medium except that the pump is generally optical (rather than electrical) and at a different wavelength than the wavelength of the field to be amplified. That is to say, one needs to use a minimum of three levels of the localized atoms in the medium in order to carry out an amplification scheme.

The composition of a semiconductor determines its bandgap, that is, the minimum energy difference between the electron and hole states of a given momentum value. A source, be it a light emitting diode (LED) or laser diode (see Figure 4) will emit light at a wavelength which corresponds to an energy slightly above the minimum gap energy (wavelength equals the velocity of light times Planck's constant divided by energy) whereas a detector can detect almost any energy above the bandedge. Only semiconductors with bandgaps which exhibit a minimum energy at a zero momentum transfer can be made to emit light strongly. Silicon does not exhibit a direct gap and although it can be used as a detector, it cannot be used as a source material. The silicon laser is and has been the 'Holy Grail' of electronics because of the ubiquity of silicon electronics. To even believe that a 'Holy Grail' exists requires a leap of faith. Weak luminescence has been

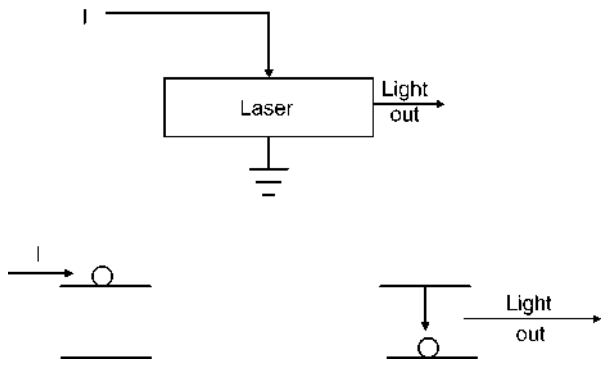

Figure 4 A schematic depiction of the workings of a semiconductor laser light source. The source is fabricated as a diode, and normal operation of this diode is in forward bias, that is, the p-side of the junction is biased positively with respect to ground which is attached to the n-side of the junction. With the p-side positively biased, current should freely flow through the junction. This is not quite true as there is a heterojunction region between the p- and n-regions in which electrons from the n-side may recombine with holes from the p-side. This recombination gives off a photon which is radiated. In light emitting diodes (LEDs), the photon simply leaves the light source as a spontaneously emitted photon. In a laser diode, the heterojunction layer serves as a waveguide and the endfaces of the laser as mirrors to provide feedback and allow laser operation in which the majority of the photons generated are generated in stimulated processes.

observed in certain silicon structures. This luminescence has been used for mid-infrared sources (where thermal effects mask strong luminescence) and in concert with rare earth dopants to make visible light emitting diodes. The silicon laser is still only a vision.

Source materials are, therefore, all made of compound semiconductors. Detectors are in essence the inverse of sources (see Figure 5) but also must detect at wavelengths above their gap energy. For primarily historical reasons, as will be further discussed below, telecommunications employs essentially three windows of wavelengths for different applications, wavelengths centered about 0.85 microns in the first window, wavelengths about 1.3 microns in the second window, and wavelengths about 1.55 microns in the third window. Materials made of layers of different compositions of AlGaAs and mounted on GaAs substrates are used in the first band, while the other bands require the increased degree of freedom allowed by the quaternary alloys of InGaAsP mounted on InP substrates. Other material mixes are possible, these above-mentioned materials are the most common. Rare earth ions exhibit many different transitions. The most useful ones have proven to be the transitions of Er in the third telecommunications window, the one that covers the spectral region to either side of 1.55 microns. Although both Nd and Pr can be made to amplify near the 1.3 micron window, amplifiers made of these materials have not proven to be especially practical.

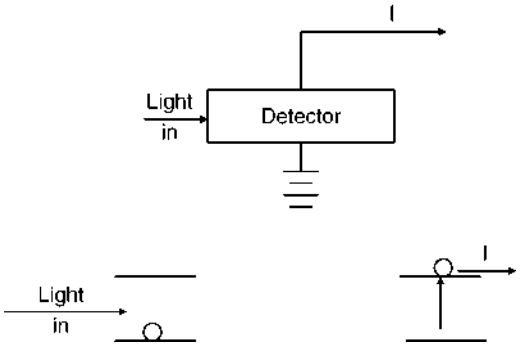

Figure 5 A schematic depiction of the workings of a semiconductor detector. As with a semiconductor source, the detector is fabricated to operate as a diode, that is a p-n junction, but in the case of the detector this diode is to be operated only in reverse bias, that is, with the n-material being biased positively with respect to the ground which is attached to the p-material's contact. In this mode of operation, no current should flow through the junction as there are no free carriers in the material, they have all been drawn out through the contacts. A photon of sufficiently high energy can change this situation by forming an electron hole pair in the material. The electron and hole are thereafter drawn to the n- and p-contacts, respectively, causing a detectable current to flow in the external circuit.

Telecommunications Technology

The physical transmission layer of the worldwide telecommunications network has come to be dominated by fiber optic technology, in particular, by wavelength division multiplexed or WDM single-mode fiber optics operating in the 1.55 micron telecommunications window. In the following, we will first discuss how this transformation to optical communications took place and then go on to discuss some of the specifics of the technology.

Already by the middle of the 1960s, it was clear that changes were going to have to take place in order that the exponential growth of the telephone system in the United States, as well as in Europe, could continue. The telephone system then, pretty much as now, consisted of a hierarchy of tree structures connected by progressively longer lines. A local office is used to connect a number of lines emanating in a tree structure to local users. The local offices are connected by trunk lines which emanate from a toll office. The lines from there on up the hierarchy are long distance ones which are termed long lines and can be regional and longer. The most pressing problem in the late 1960s was congestion in the so-called trunk lines which connect the local switching offices. The congestion was naturally most severe in urban areas. These trunk lines were generally one kilometer in length at that time. The problem was that there was no more space in the ducts that housed these lines. A solution was to use time division multiplexing or TDM to increase the traffic that could be carried by each of the lines already buried in the conduit. The problem was that the twisted pair lines employed would smear out the edges of the time varying bit streams carrying the information at the higher bitrates (aggregated rates due to the multiplexing) due to the inherent dependence of signal propagation velocity on frequency known as dispersion. After discussion and even development of a number of possible technologies, in 1975 a demonstration of a fiber optic system which employed multimode semiconductor lasers feeding multimode optical fibers all operating at 0.85 micron wavelength, proved to be the most viable model for trunk line replacement. The technology was successful and already in 1980 advances in single-mode laser and single-mode fiber technology operating at the 1.3 micron wavelength had made the inclusion of fiber into the long lines viable as well. For roughly the decade from 1985 onward, single-mode fiber systems dominated long line replacement for terrestrial as well as transoceanic links. The erbium doped fiber amplifier which operated in the 1.55 micron wavelength third telecommunication window had proven

itself to be viable for extending repeater periods by around 1990. Development of the InGaAsP quaternary semiconductor system allowed for reliable lasers to be manufactured for this window as development of strained lattice laser technology in the InGaAs system allowed for efficient pump lasers for the fiber amplifiers. As the optical amplifiers could amplify across the wavelength band that could be occupied by many aggregated channels of TDM signals, the move of the long line systems to the third telecommunications window was accompanied by the adoption of wavelength division multiplexing or WDM. That is, electronics becomes more expensive as the analog bandwidth it can handle increases. Cost-effective digital rates are now limited to a few Gb/s, and analog rates to perhaps 20 GHz. Optical center frequencies are in the order of 2×10^{14} Hz and purely optical filters can be made to cleanly separate information signals spaced as closely as 100 GHz apart on optical carriers. An optical carrier can then be made to carry hundreds of channels of 10 Gb/s separated by 100 GHz and this signal can be propagated thousands of kilometers through the terrestrial fiber optics network. Such is today's Worldwide Web.

Integrated Optics as a Future of Guided Wave Optics

As electronics has been pushed to ever greater levels of integration, its power has increased and its price has dropped. A goal implicit in much of the development of planar waveguide technology has been that optics could become an integrated technology in the same sense as electronics. This has not occurred and probably will not occur. This is not to say that integrated optical processing circuits are not being developed and that they will be not employed in the future. They most definitely will.

Integrated optics was coined as a term when it was used in 1969 to name a new program at Bell Laboratories. This program was aimed at investigation of all the technologies that were possible with which to fabricate optical integrated circuits. The original effort was in no way driven by fiber optics as the Bell System only reluctantly moved to a fiber optic network solution in 1975. The original integrated optics efforts were based on such technologies as the inorganic crystal technologies that allowed for large second-order optical nonlinearities, a necessity in order that a crystal also exhibits a large electro-optic coefficient. An electro-optic coefficient allows one to change the index of refraction of a crystal by applying a low frequency or DC electromagnetic field across the crystal.

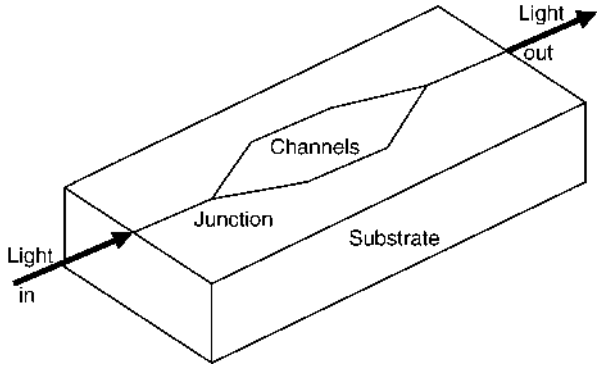

Figure 6 A schematic depiction of an integrated optical device which is often used as a high speed modulator. The device is an integrated version of a Mach–Zehnder interferometer. The basic idea behind its operation is that spatially and temporally coherent light is input into a single mode optical channel. That channel is then split into two channels by a Y-junction. Although not depicted in the figure, the two arms should not have completely equivalent propagation paths. That is to say, that if the light in one of those paths propagates a slightly longer distance (as measured in terms of phase fronts of the wave) then one cannot recombine the power from the two arms. That is, in the second Y-junction, a certain amount of the light which we are trying to combine from the two arms will be radiated out from the junction. In fact, if the light were temporally incoherent (phase fronts not well defined), exactly half of the light would be radiated form the junction. This is a statement of the brightness theorem which was thought to be a law of propagation before there were coherent sources of light. In a high speed modulator, the substrate is an electro-optic crystal and electrodes are placed over the two arms which apply the electrical signal to be impressed on the optical carrier to the channels.

Lithium niobate technology is a technology that has lasted to the present as an optical modulator technology (see **Figure 6**) and an optical switch technology (see **Figure 7**), as well as a technology for parametric wavelength conversion. Unfortunately, although low loss passive waveguide can be fabricated in lithium niobate, the crystal is not amenable to any degree of monolithic integration. Glass was also investigated as an integration technology from the early days of integrated optics, but the lack of second-order nonlinearity in glass strictly limited its applicability. Attention for a period turned to monolithic integration in the semiconductor materials which were well progressing as for use as lasers, detectors, and integrated circuits. Semiconductor crystals, however, are too pure to allow for low loss light propagation which requires the material defects to be so numerous that optical wavelengths cannot sample them. Passive waveguides in semiconductors incur huge losses that can only be mitigated by almost constant optical amplification which, unfortunately, cannot track the rapid optical field variations necessary for information transfer. Early on, attention turned to silicon

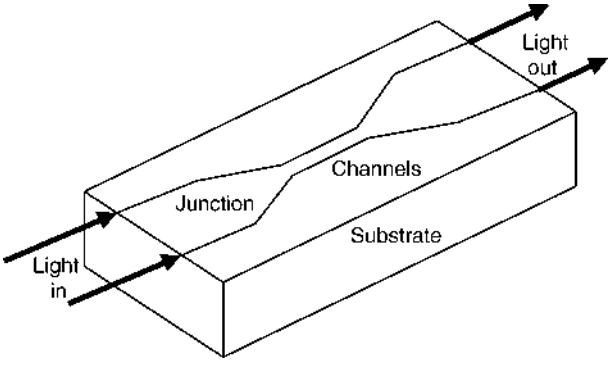

Figure 7 A schematic depiction of an integrated optical device which is often used as an optical switch. This device is interferometric but really has no free space optics counterpart as the Mach–Zhender interferometer does. In this interferometer, spatially and temporally coherent light is input into a single mode input channel. The two single optical mode input channels, which are initially so far apart from each other in terms of their mode sizes that they are effectively uncoupled, are then brought ever closer together until they are strongly coupled and their modes are no longer confined to their separate waveguides but are shared between the two channels. As the channels are symmetric, but the initial excitation, if from a single waveguide alone, is not, the pattern must evolve with propagation. Were both input channels excited, symmetry would still require a symmetry of their phases. If the phases were either equal to each other or completely out of phase, then there would be no evolution because we would have excited coupler modes. If the coupled region is the proper length, light input to one channel will emerge from the other after the channels are pulled apart. If the substrate is an electro-optic crystal and electrodes are placed over the channels, application of a voltage can affect the evolution of the interference in such a manner so as to 'decouple' the channels and switch the light back to the original channel.

optical bench technology, a technology which has presently spawned micro-electro-mechanical-system (MEMs) technology and still persists as an expensive vehicle to hybrid electro-optical integration. In silicon bench, individual components are micro-mounted with solder or silicon pop-up pieces on a silicon wafer which serves as a micro-optical bench. The technology is expensive and hybrid but pervasive in the highest end of cost classes of optical systems.

The tremendous success of the fiber optical network spawned a movement in which groups tried to achieve every possible system functionality in fibers themselves. A notable success has been the optical fiber amplifier. Significantly less successful in the fiber optic network has been the fiber Bragg grating. Attempts to carry out WDM functionalities in all fiber configuration leads to unwieldy configurations requiring large amounts of fiber and significant propagation delays, although, Bragg grating sensors have proven to be somewhat useful in sensing applications. WDM functions have been implemented in a large number of different hybrid configurations, involving various types of glasses, fused silica on silicon and simple hybrids of birefringent crystals.

The telecommunications network is not the driver that it once was. That this is so is perhaps the strongest driving force yet for integrated optics. The driver now is new applications that must be implemented in the most efficient manner. The most efficient manner is quite generally the one in which there is a maximum degree of integration. In the long run, polymer has always been the winning technology for optics of any kind, due to the flexibility of the technology and the drop in cost that accompanies mass production. Indeed, polymer integrated optics progresses. There are also a number of researchers involved in investigating strongly guiding optics, so-called photonic crystal optics. That is, in order to achieve low loss as well as low cost, the fiber and integrated optic waveguides up to the present have used small variations in optical properties of material to achieve guidance at the cost of having structures that are many wavelengths in size. In radio frequency (RF) and microwave technology, for example, guides are tiny fractions of a wavelength. This can lead to an impedance matching problem. RF and microwave impedance matching costs (versus circuit impedance matching costs) are generally quite high, precluding mass application. That radio frequency systems, such as cell phone transceivers link to the rest of the world by antenna, allows that the overall circuit dimension can be kept less than a single wavelength and impedance matching can be foregone in this so-called circuit limit. It is usually hard to keep an overall microwave system to less than a wavelength in overall extent except in the cell phone case or in a microwave oven. Optical miniaturization will require high index contrast guides, and will require optical impedance matching. There are few complete optical systems which will comprise less than an optical wavelength in overall extent. But then this so-called photonic crystal technology will likely be polymer compatible and there may be ways to find significant cost reduction. The future of integrated optics is unclear but bright.

See also

Fiber and Guided Wave Optics: Dispersion; Fabrication of Optical Fiber; Light Propagation; Measuring Fiber Characteristics; Nonlinear Effects (Basics); Nonlinear Optics; Optical Fiber Cables; Passive Optical Components. **Optical Communication Systems:** Wavelength Division Multiplexing.

Further Reading

Agrawal GP (1995) *Fiber-Optic Communication Systems*. New York: Wiley.
Bette, DeMarchis G and Iannone (1995) *Coherent Communications*. New York: Wiley.
Haken H (1985) *Light*. New York: Elsevier North-Holland.
Hecht E (1998) *Optics*. Reading, MA: Addison-Wesley.
Klein MV and Furtak TE (1986) *Optics*. New York: Wiley.
Marcuse D (1982) *Light Transmission Optics*. New York: Van Nostrand Reinhold.
Marcuse D (1991) *Theory of the Dielectic Optical Waveguide*. Boston, MA: Academic Press.
Mickelson AR, Lee Y-C and Basavanhally NR (1997) *Optoelectronic Packaging*. New York: Wiley.
Saleh BAA and Teich MC (1991) *Fundamentals of Photonics*. New York: Wiley.
Yariv A (1997) *Optical Electronics in Modern Communications*. New York: Van Nostrand Reinhold.

Dispersion

L Thévenaz, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Dispersion designates the property of a medium to propagate the different spectral components of a wave with a different velocity. It may originate from the natural dependence of the refractive index of a dense material to the wavelength of light, and one of the most evident and magnificent manifestations of dispersion is the appearance of a rainbow in the sky. In this case dispersion gives rise to a different refraction angle for the different spectral components of the white light, resulting in an angular dispersion of the sunlight spectrum and explicating the etymology of the term dispersion. Despite this spectacular effect, dispersion is mostly seen as an impairment in many applications, in particular for optical signal transmission. In this case, dispersion causes a rearrangement of the signal spectral components in the time domain, resulting in temporal spreading of the information and eventually in severe distortion.

There is a trend to designate all phenomena resulting in a temporal spreading of information as a type of dispersion, namely the polarization dispersion in optical fibers that should be properly named as polarization mode delay (PMD). In this article chromatic dispersion will be solely addressed, that designates the dependence of the propagation velocity on wavelength and that corresponds to the strict etymology of the term.

Effect of Dispersion on an Optical Signal

An optical signal may always be considered as a sum of monochromatic waves through a normal Fourier expansion. Each of these Fourier components propagates with a different phase velocity if the optical medium is dispersive, since the refractive index n depends on the optical frequency ν. This phenomenon is linear and causal and gives rise to a signal distortion that may be properly described by a transfer function in the frequency domain.

Let $n(\nu)$ be the frequency-dependent refractive index of the propagation medium. When the optical wave propagates in a waveguiding structure the effective wavenumber is properly described by a propagation constant β, that reads:

$$\beta(\nu) = \frac{2\pi\nu}{c_0} n(\nu) \qquad [1]$$

where c_0 is the vacuum light velocity. The propagation constant corresponds to an eigenvalue of the wave equation in the guiding structure and normally takes a different value for each solution or propagation mode. For free-space propagation this constant is simply equal to the wavenumber of the corresponding plane wave.

The complex electrical field $E(z,t)$ of a signal propagating in the z direction may be properly described by the following expression:

$$E(z,t) = A(z,t)\, \mathrm{e}^{\mathrm{i}(2\pi\nu_0 t - \beta_0 z)}$$
$$\text{with} \quad \nu_0\text{: the central optical frequency}$$
$$\text{and} \quad \beta_0 = \beta(\nu_0) \qquad [2]$$

where $A(z,t)$ represents the complex envelope of the signal, supposed to be slowly varying compared to the carrier term oscillating with the frequency ν_0. Consequently, the signal will spread over a narrow band around the central frequency ν_0 and the propagation constant β can be conveniently approximated by a limited expansion to the second order:

$$\beta(\nu) = \beta_0 + \left.\frac{\mathrm{d}\beta}{\mathrm{d}\nu}\right|_{\nu=\nu_0} (\nu - \nu_0) + \left.\frac{\mathrm{d}^2\beta}{\mathrm{d}\nu^2}\right|_{\nu=\nu_0} (\nu - \nu_0)^2 \qquad [3]$$

For a known signal $A(0,t)$ at the input of the propagation medium, the problem consists in

determining the signal envelope $A(z,t)$ after propagation over a distance z. The linearity and the causality of the system make possible a description using a transfer function $H_z(\nu)$ such as:

$$\tilde{A}(z,\nu) = H_z(\nu)\,\tilde{A}(0,\nu) \qquad [4]$$

where $\tilde{A}(z,\nu)$ is the Fourier transform of $A(z,t)$.

To make the transfer function $H_z(\nu)$ explicit, let us assume that the signal corresponds to an arbitrary harmonic function:

$$A(0,t) = A_0\,\mathrm{e}^{\mathrm{i}2\pi ft} \qquad [5]$$

Since this function is arbitrary and the signal may always be expanded as a sum of harmonic functions through a Fourier expansion, there is no loss of generality. The envelope identified as a harmonic function actually corresponds to a monochromatic wave of optical frequency $\nu = \nu_0 + f$ as defined by eqn [2]. Such a monochromatic wave will experience the following phase shift through propagation:

$$\begin{aligned} E(z,t) &= A_0\,\mathrm{e}^{\mathrm{i}[2\pi(\nu_0+f)t-\beta(\nu_0+f)z]} \\ &= A_0\,\mathrm{e}^{\mathrm{i}2\pi ft}\mathrm{e}^{\mathrm{i}[2\pi\nu_0 t-\beta(\nu_0+f)z]} \end{aligned} \qquad [6]$$

On the other hand, an equivalent expression may be found using the linear system described by eqns [2] and [4]:

$$\begin{aligned} E(z,t) &= A(z,t)\,\mathrm{e}^{\mathrm{i}(2\pi\nu_0 t-\beta_0 z)} \\ &= A_z\,\mathrm{e}^{\mathrm{i}2\pi ft}\,\mathrm{e}^{\mathrm{i}(2\pi\nu_0 t-\beta_0 z)} \end{aligned} \qquad [7]$$

Since eqns [6] and [7] must represent the same quantities and using the definition in eqn [4], a simple comparison shows that the transfer function must take the following form:

$$H_z(\nu) = \mathrm{e}^{-\mathrm{i}[\beta(\nu)z-\beta_0 z]} \qquad [8]$$

The transfer function takes a more analytical form using the approximation in eqn [3]:

$$H_z(\nu) = \mathrm{e}^{-\mathrm{i}2\pi(\nu-\nu_0)\tau_\mathrm{D}}\,\mathrm{e}^{-\mathrm{i}\pi D_\nu(\nu-\nu_0)^2 z} \qquad [9]$$

In the transfer function interpretation the first term represents a delay term. It means that the signal is delayed after propagation by the quantity:

$$\tau_\mathrm{D} = \frac{1}{2\pi}\frac{\mathrm{d}\beta}{\mathrm{d}\nu_0}z = \frac{z}{V_\mathrm{g}} \qquad [10]$$

where V_g represents the signal group velocity. This term therefore brings no distortion for the signal and thus states that the signal is replicated at the distance z with a delay τ_D.

The second term is the distortion term which is similar in form to a diffusion process and normally results in time spreading of the signal. In the case of a light pulse it will gradually broaden while propagating along the fiber, like a hot spot on a plate gradually spreading as a result of heat diffusion. The effect of this distortion is proportional to the distance z and to the coefficient D_ν, named group velocity dispersion (GVD):

$$D_\nu = \frac{1}{2\pi}\frac{\mathrm{d}^2\beta}{\mathrm{d}\nu^2} = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{1}{V_\mathrm{g}}\right) \qquad [11]$$

It is important to point out that the GVD may be either positive (normal) or negative (anomalous) and the distortion term in the transfer function in eqn [9] may be exactly cancelled by propagating in a medium with D_ν of opposite sign. It means that the distortion resulting from chromatic dispersion is reversible and this is widely used in optical links through the insertion of dispersion compensators. These are elements made of specially designed fibers or fiber Bragg gratings showing an enhanced GVD coefficient, with a sign opposite to the GVD in the fiber.

From the transfer function in eqn [9] it is possible to calculate the impulse response of the dispersive medium:

$$h_z(t) = \frac{1}{\sqrt{i|D_\nu|z}}\mathrm{e}^{\mathrm{i}\pi\frac{(t-\tau_\mathrm{D})^2}{D_\nu z}} \qquad [12]$$

so that the distortion of the signal may be calculated by a simple convolution of the impulse response with the signal envelope in the time domain.

The effect of dispersion on the signal can be more easily interpreted by evaluating the dispersive propagation of a Gaussian pulse. In this particular case the calculation of the resulting envelope can be carried out analytically. If the signal envelope takes the following Gaussian distribution at the origin:

$$A(0,t) = A_0\,\mathrm{e}^{-\frac{t^2}{\tau_0^2}} \qquad [13]$$

with τ_0 the $1/e$ half-width of the pulse, the envelope at distance z is obtained by convoluting the initial envelope with the impulse response $h_z(t)$:

$$\begin{aligned} A(z,t) &= h_z(t)\otimes A(0,t) \\ &= A_0\sqrt{\frac{\mathrm{i}z_0}{z+\mathrm{i}z_0}}\,\mathrm{e}^{\mathrm{i}\pi\frac{(t-\tau_\mathrm{D})^2}{D_\nu(z+\mathrm{i}z_0)}} \end{aligned} \qquad [14]$$

where

$$z_0 = -\frac{\pi\tau_0^2}{D_\nu} \qquad [15]$$

represents the typical dispersion length, that is the distance necessary to make the dispersion effect noticeable.

The actual pulse spreading resulting from dispersion can be evaluated by calculating the intensity of the envelope at distance z:

$$|A(z,t)|^2 = A_0 \frac{\tau_0}{\tau(z)} e^{-\frac{2(t-\tau_D)^2}{\tau^2(z)}} \qquad [16]$$

that is still a Gaussian distribution centered about the propagation delay time τ_D, with $1/e^2$ half-width:

$$\tau(z) = \tau_0\sqrt{1+(z/z_0)^2} \qquad [17]$$

The variation of the pulse width $\tau(z)$ is presented in Figure 1 and clearly shows that the pulse spreading starts to be nonnegligible from the distance $z = z_0$. This gives a physical interpretation for the dispersion length z_0. It must be pointed out that there is a direct formal similarity between the pulse broadening of a Gaussian pulse in a dispersive medium and the spreading of a free-space Gaussian beam as a result of diffraction. Asymptotically for distances $z \gg z_0$, the pulsewidth increases linearly:

$$\tau(z) \simeq |D_\nu| \frac{z}{\pi\tau_0} \qquad [18]$$

It must be pointed out that the width increases proportionally to the dispersion D_ν, but also inversely proportionally to the initial width τ_0. This results from the larger spectral width corresponding to a narrower pulsewidth, giving rise to a stronger dispersive effect.

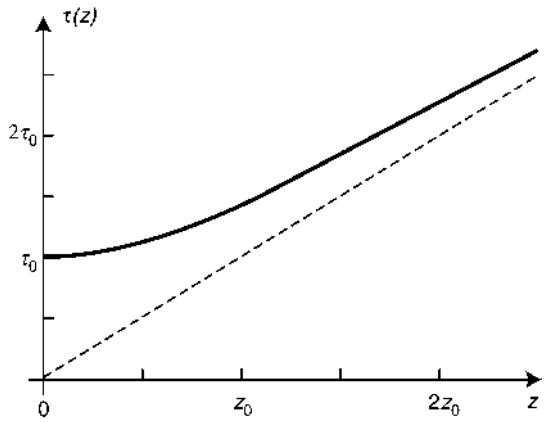

Figure 1 Variation of the $1/e^2$ half-width of a Gaussian pulse, showing the pulse spreading effect of dispersion. The dashed line shows the asymptotic linear spreading for large propagation distance.

The chromatic dispersion does not modify the spectrum of the transmitted light, as any linear effect. This can be straightforwardly demonstrated by evaluating the intensity spectrum of the signal envelope at any distance z, using the eqns [4] and [8]:

$$|\tilde{A}(z,\nu)|^2 = |H_z(\nu)\tilde{A}(0,\nu)|^2 = |e^{-i[\beta(\nu)z-\beta_0 z]}|^2|\tilde{A}(0,\nu)|^2 = |\tilde{A}(0,\nu)|^2 \qquad [19]$$

It means that the pulse characteristics in the time and frequency domains are no longer Fourier-transform limited, since after the broadening due to dispersion, the spectrum should normally spread over a narrower spectral width. This feature results from a re-arrangement of the spectral components within the pulse. This can be highlighted by evaluating the distribution of instantaneous frequency through the pulse. The instantaneous frequency ω_i is defined as the time-derivative of the wave phase factor $\phi(t)$ and is uniformly equal to the optical carrier pulsation $\omega_i = 2\pi\nu_0$ for the initial pulse, as can be deduced from the phase factor at $z = 0$ by combining eqns [2] and [13]. This constant instantaneous frequency means that all spectral components are uniformly present within the pulse at the origin.

After propagation through the dispersive medium the phase factor $\phi(t)$ can be evaluated by combining eqns [2] and [14] and evaluating the argument of the resulting expression. The instantaneous frequency ω_i is obtained after a simple time derivative of $\phi(t)$ and reads:

$$\omega_i(t) = \omega_0 + \frac{2\pi z}{D_\nu(z^2+z_0^2)}(t-\tau_D) \qquad [20]$$

For $z > 0$ the instantaneous frequency varies linearly over the pulse, giving rise to a frequency chirp. The frequency components are re-arranged in the pulse, so that the lower frequency components are in the leading edge of the pulse for a normal dispersion $D_\nu > 0$ and the higher frequencies in the trailing edge. For an anomalous dispersion $D_\nu < 0$, the arrangement is opposite, as can be seen in Figure 2. The effect of this frequency chirp can be visualized in Figure 3, showing that the effect of dispersion is equivalent to a frequency modulation over the pulse. The chirp is maximal at position z_0 and the pulsewidth takes its minimal value τ_0 when the chirp is zero. It is evident with this description that the pulse broadening may be entirely compensated through propagation in a medium of opposite

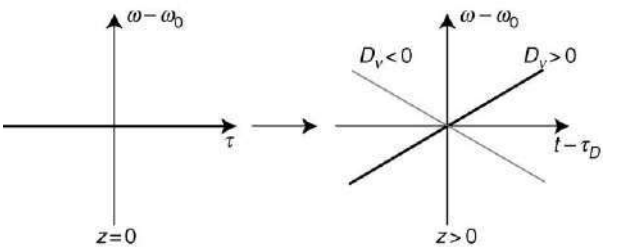

Figure 2 Distribution of the instantaneous frequency through a Gaussian pulse. At the origin the distribution is uniform (left) and the dispersion induces a frequency chirp that depends on the sign of the GVD coefficient D_ν.

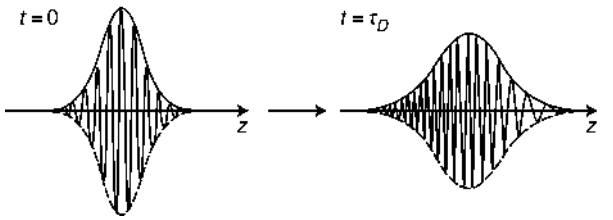

Figure 3 The dispersion results in a pulse broadening together with a frequency chirp, here for a normal GVD, that can be seen like a frequency modulation.

group velocity dispersion; this is equivalent to reversing the time direction in Figure 3. Moreover a pre-chirped pulse can be compressed to its Fourier-transform limited value after propagation in a medium with the proper dispersion sign. This feature is widely used for pulse compression after pre-chirping through propagation in a medium subject to optical Kerr effect.

The above description of the propagation of a Gaussian pulse implicitly states that the light source is perfectly coherent. In the early stages of optical communications it was not at all the case, since most sources were either light emitting diodes or multi-mode lasers. In this case the spectral extent of the signal was much larger than actually required for the strict need of modulation. Each spectral component may thus be considered as independently propagating the signal and the total optical wave can be identified to light merged from discrete sources emitting simultaneously the same signal at a different optical frequency. The group velocity dispersion will cause a delay $\delta\tau$ between different spectral components separated by a frequency interval $\delta\nu$ that can be simply evaluated by a first-order approximation:

$$\delta\tau = \frac{\mathrm{d}\tau_\mathrm{D}}{\mathrm{d}\nu}\delta\nu = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{z}{V_\mathrm{g}}\right)\delta\nu = D_\nu z\delta\nu \qquad [21]$$

where eqns [10] and [11] have been used. The delay $\delta\tau$ is thus proportional to the GVD coefficient D_ν, to the propagation distance z and the frequency separation $\delta\nu$. This description can be extended to a continuous frequency distribution with a spectral width σ_ν, resulting to the following temporal broadening σ_τ:

$$\sigma_\tau = |D_\nu|\sigma_\nu z \qquad [22]$$

Traditionally the spectral characteristics of a source are given in units of wavelength and the GVD coefficient is expressed in optical fibers accordingly. Following the same description as above, the temporal broadening σ_τ, for a spectral width σ_λ in wavelength units, reads:

$$\sigma_\tau = |D_\lambda|\sigma_\lambda z \qquad [23]$$

Since equal spectral widths must give equal broadening the value of the GVD in units of wavelength can be deduced from the natural definition in frequency units:

$$D_\lambda = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{1}{V_\mathrm{g}}\right) = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{1}{V_\mathrm{g}}\right)\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = -\frac{c_0}{\lambda^2}D_\nu \qquad [24]$$

It must be pointed out that the coefficient D_λ takes a sign opposite to D_ν; in other words, a normal dispersion corresponds to a negative GVD coefficient D_λ. It is usually expressed in units of picoseconds of temporal broadening, per nanometer of spectral width and per kilometer of propagating distance, or ps/nm km. For example, a pulse showing a spectral width of 1 nm propagating through a 100 km fiber having a dispersion D_λ of $+10$ ps/nm km, will experience, according to eqn [23], a pulse broadening σ_τ of $10 \times 1 \times 100 = 1000$ ps or 1 ns.

Material Group Velocity Dispersion

Any dense material shows a variation of its index of refraction n as a function of the optical frequency ν. This natural property is called material dispersion and is the dominant contribution in weakly guiding structures such as standard optical fibers. This natural dependence results from the noninstantaneous response of the medium to the presence of the electric field of the optical wave. In other words, the polarization field $P(t)$ corresponding to the material response will vary with some delay or inertia to the change of the incident electric field $E(t)$. This delay between cause and effect generates a memory-type response of the medium that may be described using a time-dependent medium susceptibility $\chi(t)$. The relation between medium polarization at time t and incident field results from the weighted superposition

of the effects of $E(t')$ at all previous times $t' < t$. This takes the form of the following convolution:

$$P(t) = \varepsilon_0 \int_{-\infty}^{+\infty} \chi(t - t')E(t')\mathrm{d}t' \qquad [25]$$

Through application of a simple Fourier transform this relation reads in the frequency domain as:

$$P(\nu) = \varepsilon_0 \chi(\nu)E(\nu) \qquad [26]$$

showing clearly that the noninstantaneous response of the medium results in a frequency-dependent refractive index using the standard relationship with the susceptibility χ:

$$n(\nu) = \sqrt{1 + \chi(\nu)} \quad \text{where} \quad \chi(\nu) = \mathrm{FT}\{\chi(t)\} \qquad [27]$$

This means that a beautiful natural phenomenon such as a rainbow, originates on a microscopic scale from the sluggishness of the medium molecules to react to the presence of light. For signal propagation, it results in a distortion of the signal and in most cases in a pulse spreading, but the microscopic causes are in essence identical. This noninstantaneous response is tightly related to the molecules' vibrations that also give rise to light absorption. For this reason it is convenient to express the propagation in an absorptive medium by adding an imaginary part to the susceptibility $\chi(\nu)$:

$$\chi(\nu) = \chi'(\nu) + \mathrm{i}\chi''(\nu) \qquad [28]$$

so that the refractive index $n(\nu)$ and the absorption coefficient $\alpha(\nu)$ reads in a weakly absorbing medium:

$$n(\nu) = \sqrt{1 + \chi'(\nu)}$$
$$\alpha(\nu) = -\frac{2\pi\chi''(\nu)}{\lambda n(\nu)} \qquad [29]$$

Since the response of the medium, given by the time-dependent susceptibility $\chi(t)$ in eqn [25], is real and causal, the real and imaginary part of the susceptibility in eqn [28] are not entirely independent and are related by the famous Kramers–Kronig relations:

$$\chi'(\nu) = \frac{2}{\pi}\int_0^{\infty} \frac{s\chi''(\nu)}{s^2 - \nu^2}\mathrm{d}s$$
$$\chi''(\nu) = \frac{2}{\pi}\int_0^{\infty} \frac{\nu\chi'(\nu)}{\nu^2 - s^2}\mathrm{d}s \qquad [30]$$

Absorption and dispersion act in an interdependent way on the propagating optical wave and knowing either the absorption or the dispersion spectrum is theoretically sufficient to determine the entire optical response of the medium. The interdependence between absorption and dispersion is typically illustrated in Figure 4. The natural tendency is to observe a growing refractive index for increasing frequencies in low absorption regions. In this case, the dispersion is called normal and this is the most observed situation in transparency regions of a material, which obviously offer the largest interest for optical propagation. In an absorption line the tendency is opposite, a diminishing index for increasing wavelength, and such a response is called anomalous dispersion. The dispersion considered here is the phase velocity dispersion, represented by the slope of $n(\nu)$, that must not be mistaken with the group velocity dispersion (GVD) that only matters for signal distortion. The difference between these two quantities is clarified below.

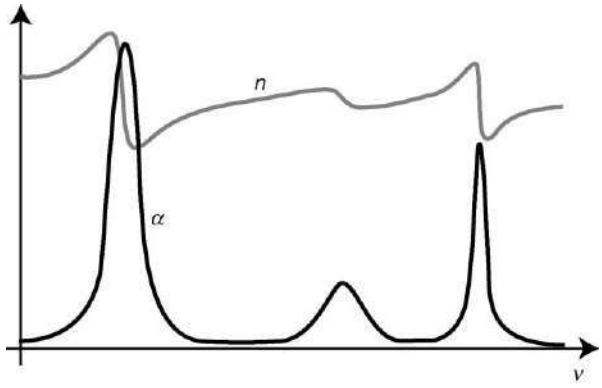

Figure 4 Typical optical response of a transparent medium, showing the spectral interdependence between the absorption coefficient α and the index of refraction n.

To demonstrate that a frequency-dependent refractive index $n(\nu)$ gives rise to a group velocity dispersion and thus a signal distortion we use eqns [1] and [10], so the group velocity V_g can be expressed:

$$V_g = \frac{c_0}{N} \quad \text{with} \quad N = n + \nu\frac{\mathrm{d}n}{\mathrm{d}\nu} = n - \lambda\frac{\mathrm{d}n}{\mathrm{d}\lambda} \qquad [31]$$

N is called group velocity index and differs from the phase refractive index n only if n shows a spectral dependence. In a region of normal dispersion $(\mathrm{d}n/\mathrm{d}\nu) > 0$, the group index is larger than the phase index and this is the situation observed in the great majority of transparent materials. From eqn [31] and using eqn [24] the GVD coefficient D_λ can be expressed as a function of the refractive index $n(\lambda)$:

$$D_\lambda = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{1}{V_g}\right) = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{N}{c_0}\right) = -\frac{\lambda}{c_0}\frac{\mathrm{d}^2 n}{\mathrm{d}\lambda^2} \qquad [32]$$

The GVD coefficient is proportional to the second derivative of the refractive index n with respect to

wavelength and is therefore minimal close to a point of inflexion of $n(\lambda)$. As can be seen in Figure 4 such a point of inflexion is always present where absorption is minimal, at the largest spectral distance from two absorption lines. It means that the conditions of low-group velocity dispersion and high transparency are normally fulfilled in the same spectral region of an optical dielectric material. In this region the phase velocity dispersion is normal at any wavelength, but the group velocity is first normal for shorter wavelengths, then is zero at a definite wavelength corresponding to the point of inflexion of $n(\lambda)$, and finally becomes anomalous for longer wavelengths. The situation in which normal phase dispersion and anomalous group dispersion are observed simultaneously is in no way exceptional.

In pure silica the zero GVD wavelength is at 1273 nm, but is subject to be moderately shifted to larger wavelengths in optical fibers, as a result of the presence of doping species to raise the index in the fiber guiding core. This shift normally never exceeds 10 nm using standard dopings; larger shifts are observed resulting from waveguide dispersion and this aspect will be addressed in the next section. The zero GVD wavelength does not strictly correspond to the minimum attenuation in silica fibers, because the dominant source of loss is Rayleigh scattering in this spectral region and not molecular absorption. This scattering results from fluctuations of the medium density as observed in any amorphous materials such as vitreous silica and is therefore a collective effect of many molecules that does not impinge on the microscopic susceptibility $\chi(t)$. It has therefore no influence on the material dispersion characteristics and this explains the reason for the minimal attenuation wavelength at 1550 nm, mismatching and quite distant from the zero material GVD at 1273 nm.

Material GVD in amorphous SiO_2 can be accurately described using a three-term Sellmeier expansion of the refractive index:

$$n(\lambda) = \sqrt{1 + \sum_{j=1}^{3} \frac{C_j \lambda^2}{\lambda^2 - \lambda_j^2}} \qquad [33]$$

and performing twice the wavelength derivative. The coefficients C_j and λ_j are found in most reference handbooks and result in the GVD spectrum shown in Figure 5. Such a spectrum explains the absence of interest for propagation in the visible region through optical fibers, the dispersion being very important in this spectral region. It also explains the large development of optical fibers in the 1300 nm region as a consequence of the minimal material dispersion there. It must be pointed out that it is quite easy to set up propagation in an anomalous GVD regime in optical fibers, since this regime is observed in the lowest attenuation spectral region. Anomalous dispersion makes possible interesting propagation features when combined with a third-order nonlinearity as the optical Kerr effect, namely soliton propagation and efficient spectral broadening through modulation instability.

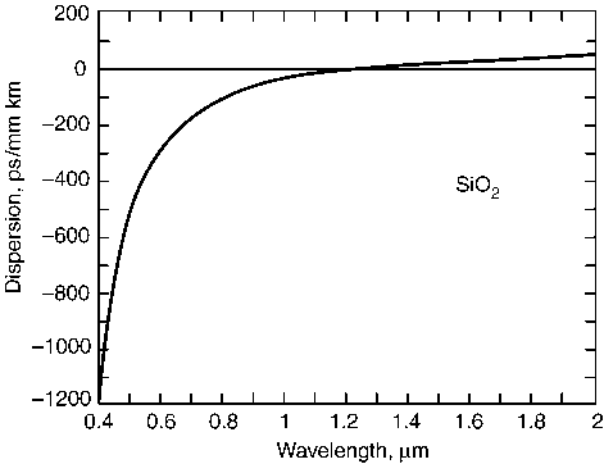

Figure 5 Material dispersion of pure silica. The visible region (0.4–0.7 μm) shows a strong normal GVD that decreases when moving into the infrared and eventually vanishes at 1273 nm. In the minimum attenuation window (1550 nm) the material GVD is anomalous.

Waveguide Group Velocity Dispersion

Solutions of the wave equation in an optical dielectric waveguide such as an optical fiber are discrete and limited. These solutions, called modes, are characterized by an unchanged field distribution along the waveguides and by a uniform propagation constant β over the wavefront. This last feature is particularly important if one recall that the field extends over regions presenting different refractive indices in a dielectric waveguide. For a given mode, the propagation constant β defines an effective refractive index n_{eff} for the propagation by similarity to eqn [1]:

$$\beta = \frac{2\pi\nu}{c_0} n_{\text{eff}} \qquad [34]$$

The value of this effective refractive index n_{eff} is always bound by the value of the core index n_1 and of the cladding index n_2, so that $n_2 < n_{\text{eff}} < n_1$. For a given mode, the propagation constant β, and so the effective refractive index n_{eff}, only depend on a quantity called normalized frequency V that essentially scales the light frequency to the waveguide

optical parameters:

$$V = \frac{2\pi}{\lambda} a \sqrt{n_1^2 - n_2^2} \qquad [35]$$

where a is the core radius. Figure 6 shows the dependence of the propagation constant β of the fundamental mode LP_{01} on the normalized frequency V. The variation is nonnegligible in the single-mode region and gives rise to a chromatic dispersion, since V depends on the wavelength λ, even in the fictitious case of dispersion-free refractive indices in the core and the cladding materials. This type of chromatic dispersion is called waveguide dispersion.

To find an expression for the waveguide dispersion in function of the guiding properties, let us define another normalized parameter, the normalized phase constant b, such as:

$$\beta = \frac{2\pi}{\lambda} \sqrt{n_2^2 + b(n_1^2 - n_2^2)} \qquad [36]$$

The parameter b takes values in the interval $0 < b < 1$, is equal to 0 when $n_{\text{eff}} = n_2$ at the mode cutoff, and is equal to 1 when $n_{\text{eff}} = n_1$. This latter situation is never observed and is only asymptotic for very large normalized frequencies V. Solving the wave equation provides the dispersion relation $b(V)$ and it is important to point out that this relation between normalized quantities depends only on the shape of the refractive index profile. Step-index, triangular or multiple-clad index profiles will result in different $b(V)$ relations, independently of the actual values of the refractive indices n_1 and n_2 and of the core radius a.

From the definitions in eqns [10] and [35] and in the fictitious case of an absence of material dispersion, the propagation delay per unit length reads:

$$\frac{1}{V_g} = \frac{1}{2\pi}\frac{d\beta}{d\nu} = \frac{1}{2\pi}\frac{d\beta}{dV}\frac{dV}{d\nu} = \frac{\lambda}{2\pi} V \frac{d\beta}{dV} \qquad [37]$$

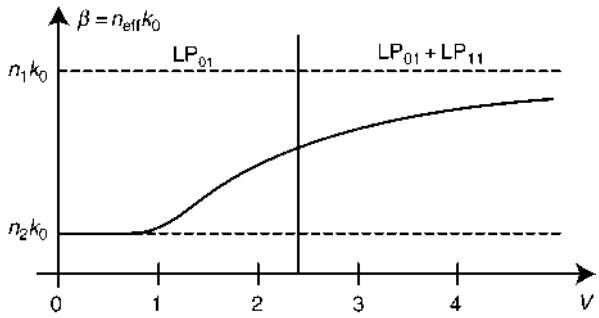

Figure 6 Propagation constant β as a function of the normalized frequency V in a step-index optical fiber. The single mode region is in the range $0 < V < 2.405$.

so that the waveguide group velocity dispersion can be expressed using the relation in eqn [24]:

$$D_\lambda^{\text{w}} = \frac{d}{d\lambda}\left(\frac{1}{V_g}\right) = \frac{d}{d\nu}\left(\frac{1}{V_g}\right)\frac{d\nu}{d\lambda}$$
$$= -\frac{\nu}{\lambda}\frac{d}{d\nu}\left(\frac{1}{V_g}\right) = -\frac{1}{2\pi c_0} V^2 \frac{d^2\beta}{dV^2} \qquad [38]$$

The calculation of the combined effect of material and waveguide dispersions results in very long expressions in which it is difficult to highlight the relative effect of each contribution. Nevertheless, by making the assumption of weak guidance:

$$\frac{n_1 - n_2}{n_2} \simeq \frac{N_1 - N_2}{N_2} \ll 1 \qquad [39]$$

where N_1 and N_2 are the group indices in the core and the cladding, respectively, defined in eqn [31], the complete expression can be drastically simplified to obtain for the delay per unit length:

$$\frac{1}{V_g} = \frac{1}{c_0}\left[N_2 + (N_1 - N_2)\frac{d(bV)}{dV}\right] \qquad [40]$$

and for the total dispersion:

$$D_\lambda = D_2 + (D_1 - D_2)\frac{d(bV)}{dV}$$
$$- (N_1 - N_2)\frac{N_2}{n_1}\frac{1}{\lambda c_0} V \frac{d^2(bV)}{dV^2} \qquad [41]$$

where

$$D_1 = -\frac{\lambda}{c_0}\frac{d^2 n_1}{d\lambda^2} \quad \text{and} \quad D_2 = -\frac{\lambda}{c_0}\frac{d^2 n_2}{d\lambda^2} \qquad [42]$$

are the material GVD in the core and the cladding, respectively.

The first two terms in eqn [41] represent the contribution of material dispersion weighted by the relative importance of core and cladding materials for the propagating mode. In optical fibers, the difference between D_1 and D_2 is small, so that this contribution can be often well approximated by D_2, independently of any guiding effects.

The last term represents the waveguide dispersion and is scaled by 2 factors:

- *The core-cladding index difference* $n_1 - n_2 \simeq N_1 - N_2$. The waveguide dispersion will be significantly enhanced by increasing the index difference between core and cladding.
- *The shape factor* $V(d^2(bV)/dV^2)$. This factor uniquely depends on the shape of the refractive index profile and may substantially modify the

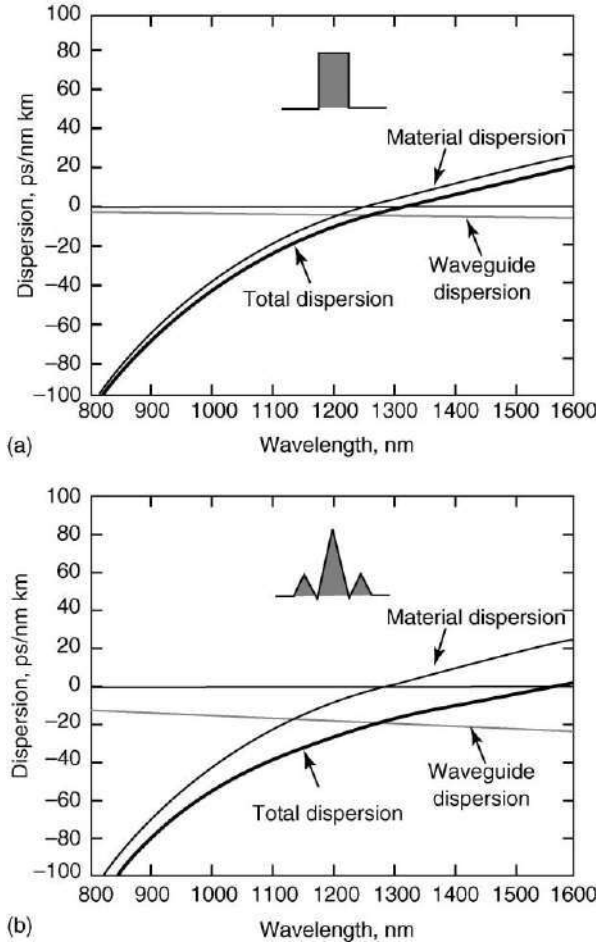

Figure 7 Material, waveguide and total group velocity dispersions for: (a) step-index fiber; (b) triangular profile fiber. The waveguide dispersion can be significantly enhanced by changing the shape of the index profile, making possible a shift of the zero GVD to the minimum attenuation window.

spectral dependence of the waveguide dispersion, making possible a great variety of dispersion characteristics.

Due to this degree of freedom brought by waveguide dispersion it is possible to shift the zero GVD wavelength to the region of minimal attenuation at 1550 nm in silica optical fibers. Figure 7a shows the total group velocity dispersion of a step-index core fiber, together with the separate contributions of material and waveguide GVD. In this case, the material GVD is clearly the dominating contribution, while the small waveguide GVD results in a shift of the zero GVD wavelength from 1273 nm to 1310 nm. A larger shift could be obtained by increasing the core-cladding index difference, but this also gives rise to an increased attenuation from the doping and no real benefit can be expected from such a modification. In Figure 7b, the core shows a triangular index profile to enhance the shape factor in eqn [41], so that the contribution of waveguide GVD is significantly increased with no impairing attenuation due to excessive doping. This makes it possible to realize the ideal situation of an optical fiber showing a zero GVD at the wavelength of minimum attenuation. These dispersion-shifted fibers (DSF) have now become successful in modern telecommunication networks.

Nevertheless, the absence of dispersion favors the efficiency of nonlinear effects and several classes of fibers are now proposed, showing a small but nonzero GVD at 1550 nm, with positive or negative sign, nonzero DSF (NZDSF). By interleaving fibers with positive and negative GVDs it is possible to propagate along the optical link in a dispersive medium and thus minimize the impact of nonlinearities, while maintaining the overall GVD of the link close to zero and canceling any pulse spreading accordingly.

List of Units and Nomenclature

Chromatic dispersion [ps nm^{-1} km^{-1}]	D_λ
Electric field [V m^{-1}]	E
Group index	N
Group velocity [m s^{-1}]	V_g
Group Velocity Dispersion GVD [s^2 m^{-1}]	D_ν
Linear attenuation coefficient [m^{-1}]	α
Medium susceptibility	χ
Normalized frequency	V
Optical frequency [s^{-1}]	ν
Propagation constant [m^{-1}]	β
Propagation delay [s]	τ_D
Polarization density field [A s m^{-2}]	P
Refractive index	n
Vacuum light velocity [m s^{-1}]	c_0
Wavelength [m]	λ

See also

Dispersion Management. **Fiber and Guided Wave Optics:** Overview. **Fourier Optics**. **Nonlinear Optics, Basics:** Kramers–Kronig Relations in Nonlinear Optics.

Further Reading

Ainslie BJ and Day CR (1986) A review of single-mode fibers with modified dispersion characteristics. *IEEE Journal of Lightwave Technology* LT-4(8): 967–979.

Bass M (ed.) (1994) *Handbook of Optics*, sponsored by the Optical Society of America, 2nd edn, vol. II, chap. 10 & 33. New York: McGraw-Hill.

Buck JA (1995) In: Goodman JW (ed.) *Fundamentals of Optical Fibers*, chap. 1, 5, 6. New York: Wiley Series in Pure and Applied Optics.

Gloge D (1971) Dispersion in weakly guiding fibers. *Applied Optics* 10: 2442–2445.
Jeunhomme LB (1989) *Single Mode Fiber Optics: Principles and Applications*. Optical Engineering Series, No. 23. New York: Marcel Dekker.
Marcuse D (1974) *Theory of Dielectric Optical Waveguides*. Series on Quantum Electronics – Principles and Applications, chap. 2. New York: Academic Press.
Marcuse D (1980) Pulse distortion in single-mode fibers. *Applied Optics* 19: 1653–1660.
Marcuse D (1981) Pulse distortion in single-mode fibers. *Applied Optics* 20: 2962–2974.
Marcuse D (1981) Pulse distortion in single-mode fibers. *Applied Optics* 20: 3573–3579.
Mazurin OV, Streltsina MV and Shvaiko-Shvaikovskaya TP (1983) *Silica Glass and Binary Silicate Glasses (Handbook of Glass Data; Part A)* Series on Physical Sciences Data. Amsterdam: Elseviervol. 15, pp. 63–75.
Murata H (1988) *Handbook of Optical Fibers and Cables*. Optical Engineering Series, No. 15, chap. 2. New York: Marcel Dekker.
Saleh BEA and Teich MC (1991) In: Goodman JW (ed.) *Fundamentals of Photonics*, chap. 5, 8 & 22. New York: Wiley Series in Pure and Applied Optics.

Fabrication of Optical Fiber

D Hewak, University of Southampton, Southampton, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The drawing of optical fibers from silica preforms has, over a short period of time, progressed from the laboratory to become a manufacturing process capable of producing millions of kilometers of telecommunications fiber a year. Modern optical fiber fabrication processes produce low-cost fiber of excellent quality, with transmission losses close to their intrinsic loss limit. Today, fiber with transmission losses of 0.2 dB per kilometer of fiber are routinely drawn through a two-stage process that has been refined since the 1970s.

Although fibers of glass have been fabricated and used for hundreds of years, it was not until 1966 that serious interest in the use of optical fibers for communication emerged. At this time, it was estimated that the optical transmission loss in bulk glass could be as low as 20 dB km^{-1} if impurities were sufficiently reduced, a level at which practical applications were possible. At this time, no adequate fabrication techniques were available to synthesize glass of high purity, and fiber-drawing methods were crude.

Over the next five years, efforts worldwide addressed the fabrication of low-loss fiber. In 1970, a fiber with a loss of 20 dB km^{-1} was achieved. The fiber consisted of a titania doped core and pure silica cladding. This result generated much excitement and a number of laboratories worldwide actively began researching optical fiber. New fabrication techniques were introduced, and by 1986, fiber loss had been reduced close to the theoretical limit.

All telecommunications fiber that is fabricated today is made of silica glass, the most suitable material for low-loss fibers. Early fiber research studied multicomponent glasses, which are perhaps more familiar optical materials; however, low-loss fiber, could not be realized, partly due to the lack of a suitable fabrication method. Today other glasses, in particular the fluorides and sulfides, continue to be developed for speciality fiber applications, but silica fiber dominates in most applications.

Silica is a glass of simple chemical structure containing only two elements, silicon and oxygen. It has a softening temperature of about 2,000°C at which it can be stretched, i.e. drawn into fiber. An optical fiber consists of a high purity silica glass core, doped with suitable oxide materials to raise its refractive index (Figure 1). This core, typically on the order of 2–10 microns in diameter, is surrounded by silica glass of lower refractive index. This cladding layer extends the diameter to typically 125 microns. Finally a protective coating covers the entire structure. It is the phenomenon of total internal reflection at the core cladding interface that confines light to the core and allows it to be guided. The basic requirements of an optical fiber are as follows:

1. The material used to form the core of the fiber must have a higher refractive index than the cladding material, to ensure the fiber is a guiding structure.

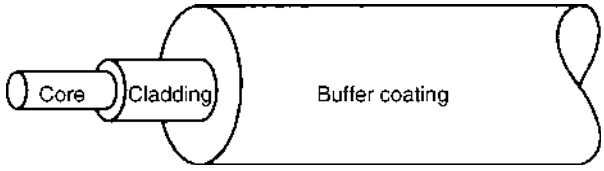

Figure 1 Structure of an optical fiber.

2. The materials used must be low loss, providing transmission with no absorption or scattering of light.
3. The materials used must have suitable thermal and mechanical properties to allow them to be drawn down in diameter into a fiber.

Silica (SiO_2) can be made into a glass relatively easily. It does not easily crystallize, which means that scattering from unwanted crystalline centers within the glass is negligible. This has been a key factor in the achievement of a low-loss fiber. Silica glass has high transparency in the visible and near-infrared wavelength regions and its refractive index can be easily modified. It is stable and inert, providing excellent chemical and mechanical durability. Moreover, the purification of the raw materials used to synthesize silica glass is quite straightforward.

In the first stage of achieving an optical fiber, silica glass is synthesized by one of three main chemical vapor processes. All use silicon tetrachloride ($SiCl_4$) as the main precursor, with various dopants to modify the properties of the glass. The precursors are reacted with oxygen to form the desired oxides. The end result is a high purity solid glass rod with the internal core and cladding structure of the desired fiber.

In the second stage, the rod, or preform as it is known, is heated to its softening temperature and stretched to diameters of the order of 125 microns. Tens to hundreds of kilometers of fiber are produced from a single preform, which is drawn continuously, with minimal diameter fluctuations. During the drawing process one or more protective coatings are applied, yielding long lengths of strong, low-loss fiber, ready for immediate application.

Preform Fabrication

The key step in preparing a low-loss optical fiber is to develop a technique for completely eliminating transition metal and OH ion contamination during the synthesis of the silica glass. The three methods most commonly used to fabricate a glass optical fiber preform are: the modified chemical vapor deposition process (MCVD); the outside vapor deposition process (OVD); and the vapor-axial deposition process (VAD).

In a typical vapor phase reaction, halide precursors undergo a high temperature oxidation or hydrolysis to form the desired oxides. The completed chemical reaction for the formation of silica glass is, for oxidation:

$$SiCl_4 + O_2 \rightarrow SiO_2 + 2Cl_2 \qquad [1]$$

For hydrolysis, which occurs when the deposition occurs in a hydrogen-containing flame, the reaction is:

$$SiCl_4 + 2H_2O \rightarrow SiO_2 + 4HCl \qquad [2]$$

These processes produce fine glass particles, spherical in shape with a size of the order of 1 nm. These glass particles, known as soot, are then deposited and subsequently sintered into a bulk transparent glass. The key to low-loss fiber is the difference in vapor pressures of the desired halides and the transition metal halides that cause significant absorption loss at the wavelengths of interest. The carrier gas picks up a pure vapor of, for example, $SiCl_4$, and any impurities are left behind.

Part of the process requires the formation of the desired core/cladding structure in the glass. In all cases, silica-based glass is produced with variations in refractive index produced by the incorporation of dopants. Typical dopants used are germania (GeO_2), titania (TiO_2), alumina (Al_2O_3), and phosphorous pentoxide (P_2O_5) for increasing the refractive index, and boron oxide (B_2O_2) and fluorine (F) for decreasing it. These dopants also allow other properties to be controlled, such as the thermal expansion of the glass and its softening temperatures. In addition, other materials, such as the rare earth elements, have also been used to fabricate active fibers that are used to produce optical fiber amplifiers and lasers.

MCVD Process

Optical fibers were first produced by the MCVD method in 1974, a breakthrough that completely solved the technical problems of low-loss fiber fabrication (Figure 2).

As shown schematically in Figure 3, the halide precursors are carried in the vapor phase by oxygen carrier gas into a pure silica substrate tube. An oxyhydrogen burner traverses the length of the tube, which it heats externally. The tube is heated to temperatures of about 1,400°C which then oxidizes the halide vapor materials. The deposition temperature is sufficiently high to form a soot made of glassy particles which are deposited on the inside wall of the substrate tube but low enough to prevent the softened silica substrate tube from collapsing. The process usually takes place on a horizontal glass-working lathe.

During the deposition process, the repeated traversing of the burner forms multiple layers of soot. Changes to the precursors entering the tube and thus the resulting glass composition are introduced for layers which will form the cladding and then the core. The MCVD method allows germania to be doped into

Figure 2 Fabrication of an optical fiber preform by the MCVD method.

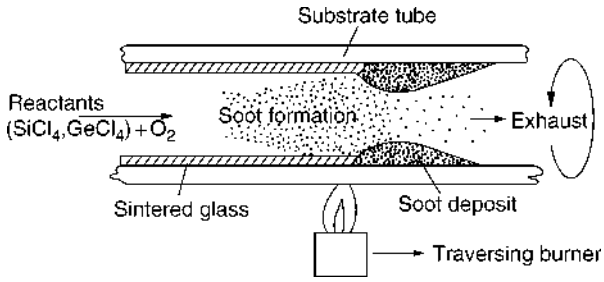

Figure 3 Schematic of preform fabrication by the MCVD method.

the silica glass and the precise control of the refractive index profile of the preform. When deposition is complete, the burner temperature is increased and the hollow, multilayered structure is collapsed to a solid rod. A characteristic of fiber formed by this process is a refractive index dip in the center of the core of the fiber. In addition, the deposition takes place in a closed system, which dramatically reduces contamination by OH^- ions and maintains low levels of other impurities. The high temperature allows high deposition rates compared to traditional chemical vapor deposition (CVD) and large performs, built up from hundreds of layers can be produced.

The MCVD method is still widely used today though it has some limitations, particularly on the preform size that can be achieved and thus the manufacturing cost. Diameter of the final preform is determined by the size of the initial silica substrate tube, which, due to the high purity required, accounts for a significant portion of the cost of the preform. A typical preform fabricated by MCVD yields about 5 km of fiber. Its success has spurred improvements to the process, in particular to address the fiber yield from a single preform.

OVD Process

The outside vapor deposition (OVD) methods, also known as outside vapor phase oxidation (OVPO), synthesize the fine glass particles within a burner flame. The precursors, oxygen, and fuel for the burner, are introduced directly into the flame. The soot is deposited onto a target rod that rotates and traverses in front of the burner. As in MCVD, the preform is built up layer by layer, though now initially by depositing the core glass and then building up the cladding layers over this. After the deposition process is complete, the preform is removed from the target rod and is collapsed and sintered into a transparent glass preform. The center hole remains, but disappears during the fiber drawing process. This technique has advantages in both size of preform which can be obtained and the fact that a high-quality silica substrate tube is no longer required. These two advantages combine to make a more economical process. From a single preform, several hundred kilometers of fiber can be produced.

VAD Process

The most recent refinement to the fabrication process was developed again to aid the mass production of high-quality fibers. In the VAD process, both core and cladding glasses are deposited simultaneously. Like OVD, the soot is synthesized and deposited by flame hydrolysis, as shown in Figure 4, the precursors are blown from a burner, oxidized, and deposited onto a silica target rod.

Burners for the VAD process consist of a series of concentrate nozzles. The first delivers an inert carrier and the main precursors $SiCl_4$, the second delivers an inert carrier glass and the dopants, the third delivers hydrogen fuel, and the fourth delivers oxygen. Gas flows are up to a liter per minute and deposition rates can be very high. With the VAD process, both core and cladding glasses can be deposited simultaneously. The main advantage is that this is a continuous process, as the soot which forms the core and cladding glasses are deposited axially onto the end of the rotating silica rod, it is slowly drawn upwards into a furnace which sinters and consolidates the soot into a transparent glass. The upward motion is such that the end at which deposition is occurring remains in a fixed position and essentially the preform is grown from this base.

The advantages of the VAD process are a preform without a central dip and, most importantly, the mass production associated with a continuous process. The glass quality produced is uniform and the resulting fibers have excellent reproducibility and low loss.

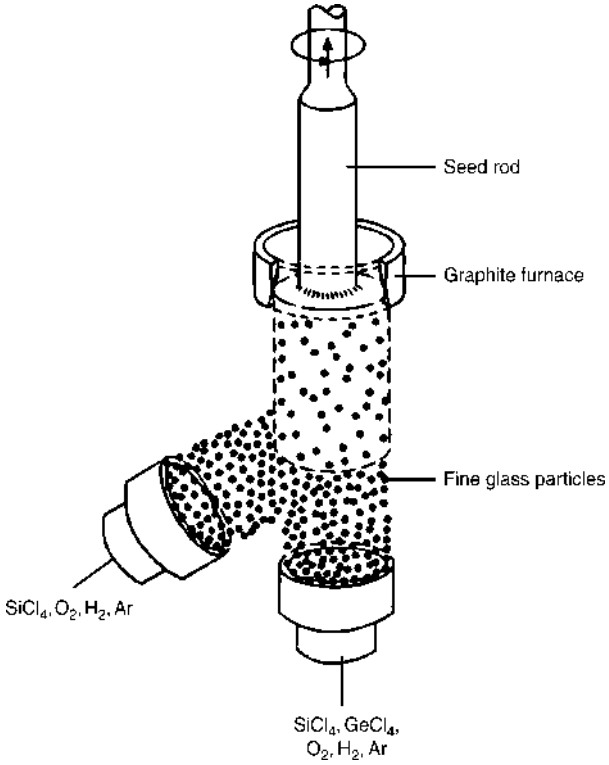

Figure 4 Schematic of preform fabrication by the VAD method.

Figure 5 A commercial scale optical fiber drawing tower.

Other Methods of Preform Fabrication

There are other methods of preform fabrication, though these are not generally used for today's silica glass based fiber. Some methods, such as fabrication of silica through sol–gel chemistry could not provide the large low-loss preforms obtained by chemical vapor deposition techniques. Other methods, such as direct casting of molten glass into rods, or formation of rods and tubes by extrusion, are better suited to and find application in other glass chemistries and speciality fibers.

Fiber Drawing

Once a preform has been made, the second step is to draw the preform down in diameter on a fiber drawing tower. The basic components and configuration of the drawing tower have remained unchanged for many years, although furnace design has become more sophisticated and processes like coating have become automated. In addition, fiber drawing towers have increased in height to allow faster pulling speeds (Figure 5).

Essentially, the fiber drawing process takes place as follows. The preform is held in a chuck which is mounted on a precision feed assembly that lowers the preform into the drawing furnace at a speed which matches the volume of preform entering the furnace to the volume of fiber leaving the bottom of the furnace. Within the furnace, the fiber is drawn from the molten zone of glass down the tower to a capstan, which controls the fiber diameter, and then onto a drum. During the drawing process, immediately below the furnace, is a noncontact device that measures the diameter of the fiber. This information is fed back to the capstan which speeds up to reduce the diameter or slows down to increase the fiber diameter. In this way, a constant diameter fiber is produced. One or more coatings are also applied in-line to maintain the pristine surface quality as the fiber leaves the furnace and thus to maximize the fiber strength.

The schematic drawing in Figure 6 shows a fiber drawing tower and its key components. Draw towers are commercially available, ranging in height from 3 m to greater than 20 m, with the tower height increasing as the draw speed increases.

Typical drawing speeds are on the order of $1\ \mathrm{m\,s^{-1}}$. The increased height is needed to allow the fiber to cool sufficiently before entering the coating applicator, although forced air-cooling of the fiber is commonplace in a manufacturing environment.

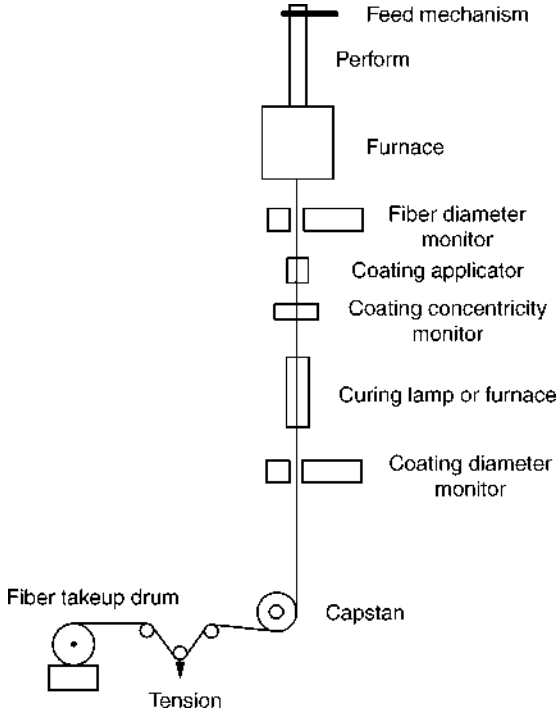

Figure 6 Schematic diagram of an optical fiber drawing tower and its components.

Furnaces

A resistance furnace is perhaps the most common and economical heating source used in a fiber drawing tower. A cylindrical furnace, with graphite or tungsten elements, provides heating to the preform by blackbody radiation. These elements must be surrounded by an inert gas, such as nitrogen or argon, to prevent oxidation. Gas flow and control is critical to prevent variations in the diameter of the fiber. In addition, the high temperature of the elements may lead to contamination of the preform surface that can then weaken the fiber.

An induction furnace provides precise clean heating through radio frequency (RF) energy that is inductively coupled to a zirconia susceptor ring. This type of furnace is cleaner than the graphite resistance type and is capable of continuous running for several months. It also has the advantage that it does not require a protective inert atmosphere and consequently the preform is drawn in a turbulence-free environment. These advantages result in high-strength fibers with good diameter control.

Diameter Measurement

A number of noncontact methods of diameter measurement can be applied to measure fiber diameter. These include laser scanning, interferometry, and light scattering techniques. To obtain precise control of the fiber diameter, the deviation between the desired diameter and the measured diameter is fed into the diameter control system. In order to cope with high drawing speeds, sampling rates as high as 1,000 times per second are used. The diameter control is strongly affected by the gas flow in the drawing furnace and is less affected by the furnace temperature variation. The furnace gas flow can be used to achieve suppression of fast diameter fluctuations. This is used in combination with drawing speed control to achieve suppression of both fast and slow diameter fluctuations. Current manufacturing processes are capable of producing several hundreds of kilometers of fiber with diameter variations of ± 1 micron.

There are two major sources of fiber diameter fluctuations: short-term fluctuations caused by temperature fluctuations in the furnace; and long-term fluctuations caused by variations in the outer diameter of the preform. Careful control of the furnace temperature, the length of the hot zone, and the flow of gas minimize the short-term fluctuations. Through optimization of these parameters, diameter errors of less than ± 0.5 microns can be realized. The long-term fluctuations in diameter are controlled by the feedback mechanism between the diameter measurement and the capstan.

Fiber Coating

A fiber coating is primarily used to preserve the strength of a newly drawn fiber and therefore must be applied immediately after the fiber leaves the furnace. The fiber coating apparatus is typically located below the diameter measurement gauge at a distance determined by the speed of fiber drawing, the tower height, and whether there is external cooling of the fiber. Coating is usually one of the limiting factors in the speed of fiber drawing. To minimize any damage or contamination of the pristine fiber leaving the furnace, this portion of the tower, from furnace to the application of the first coating, is enclosed in a clean, filtered-air chamber.

A wide variety of coating materials has been applied; however, conventional, commercial fiber generally relies on a UV curable polymer as the primary coating. Coating thickness is typically 50–100 microns. Subsequent coatings can then be applied for specific purposes. A dual coating is often used with an inner primary coating that is soft and an outer, secondary coating which is hard. This ratio of low to high elastic modulus can minimize stress on the fiber and reduce bending loss.

Alternative coatings include hermetic coatings of a low melting temperature metal, ceramics, or amorphous carbon. These can be applied in-line before the polymeric coating. Metallic coatings are applied by passing the fiber through a molten metal while ceramic or amorphous coatings utilize an in-line chemical vapor deposition reactor.

Types of Fiber

The majority of silica fiber drawn today is single mode. This structure consists of a core whose diameter is chosen such that, with a given refractive index difference between the core and cladding, only a single guide mode propagates at the wavelength of interest. With a discrete index difference between core and cladding, it is often referred to as a step index fiber. In typical telecommunications fiber, single mode operation is obtained with core diameters of 2–10 microns with a standard outer diameter of 125 microns.

Multimode fiber has core diameters considerably larger, typically 50, 62.5, 85, and 110 microns, again with a cladded diameter of 125 microns. Multimode fibers are often graded index, that is the refractive index is a maximum in the center of the fiber and smoothly decreases radially until the lower cladding index is reached. Multimode fibers find use in non-telecommunication applications, for example optical fiber sensing and medicine.

Single mode fibers, which are capable of maintaining a linear polarization input to the fiber, are known as polarization preserving fibers. The structure of these fibers provides a birefringence that removes the degeneracy of the two possible polarization modes. This birefringence is a small difference in the effective refractive index of the two polarization modes that can be guided and it is achieved in one of two ways. Common methods for the realization of this birefringence are an elliptical core in the fiber, or through stress rods, which modify the refractive index in one orientation. These fiber structures are shown in Figure 7.

While the loss minimum of silica-based fiber is near 1.55 microns, step index single-mode fiber offers zero dispersion close to 1.3 micron wavelengths and dispersion at the loss minimum is considerable. A modification of the structure of the fiber, and in particular a segmented refractive index profile in the core, can be used to shift this dispersion minimum to 1.55 microns. This fiber, illustrated in Figure 7 is known as dispersion shifted fiber. Similarly fibers, with a relatively low dispersion over a wide wavelength range, known as dispersion flattened fibers, can be obtained by the use of multiple cladding layers.

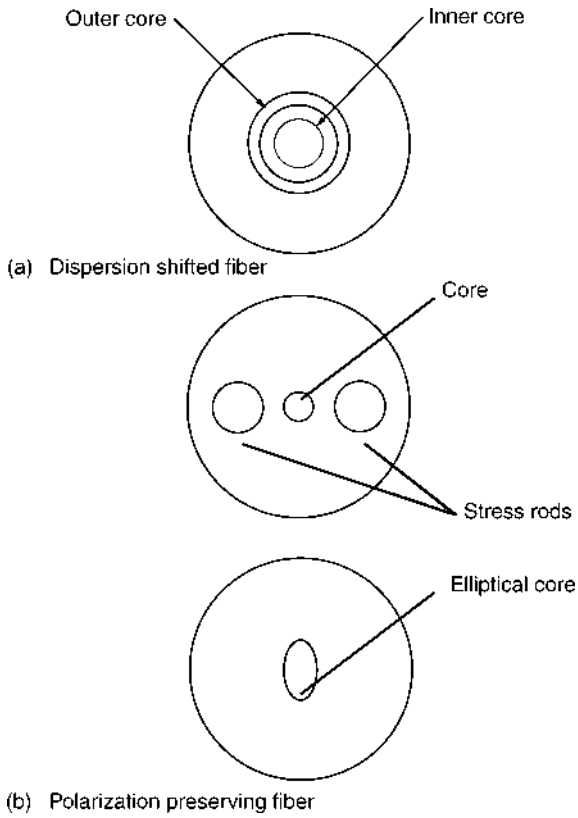

Figure 7 Structure of (a) dispersion shifted fiber and (b) two methods of achieving polarization preserving fiber.

List of Units and Nomenclature

Dopants	Elements or compounds added, usually in small amounts, to a glass composition to modify its properties.
Fiber drawing	The process of heating and thus softening an optical fiber preform and then drawing out a thin thread of glass.
Fiber loss	The transmission loss of light as it propagates through a fiber, usually measured in dB of loss per unit length of fiber. Loss can occur through the absorption of light in the core or scattering of light out of the core.
Glass	An amorphous solid formed by cooling from the liquid state to a rigid solid with no long range structure.
Modified chemical vapor deposition (MCVD)	A process for the fabrication of an optical preform where gases flow into the inside of a rotating tube, are heated and react

	to form particles of glass which are deposited onto the wall of the glass tube. After deposition, the glass particles are consolidated into a solid preform.
Outside vapor deposition (OVD)	A process for the fabrication of an optical preform where glass soot particles are formed in an oxy-hydrogen flame and deposited on a rotating rod. After deposition, the glass particles are consolidated into a solid preform.
Preform	A fiber blank, a bulk glass rod consisting of a core and cladding glass composite which is drawn into fiber.
Refractive index	A characteristic property of glass, which is defined by the speed of light in the material relative to the speed of light in a vacuum.
Silica	A transparent glass formed from silicon dioxide.
Vapor-axial deposition	A process similar to OVD, where the core and cladding layers are deposited simultaneously.

See also

Detection: Fiber Sensors. **Fiber and Guided Wave Optics:** Dispersion; Light Propagation; Measuring Fiber Characteristics; Nonlinear Effects (Basics); Optical Fiber Cables. **Fiber Gratings**. **Optical Amplifiers:** Erbrium Doped Fiber Amplifiers for Lightwave Systems. **Optical Communication Systems:** Architectures of Optical Fiber Communication Systems; Historical Development; Wavelength Division Multiplexing. **Optical Materials:** Optical Glasses.

Further Reading

Agrawal GP (1992) *Fiber-Optic Communication Systems*. New York: John Wiley & Sons.

Bass M and Van Stryland EV (eds) (2001) *Fiber Optics Handbook: Fiber, Devices, and Systems for Optical Communications*. New York: McGraw-Hill Professional Publishing.

Fujiura K, Kanamori T and Sudo S (1997) Fiber materials and fabrications. In: Sudo S (ed.) *Optical Fiber Amplifiers*, ch. 4, pp. 193–404. Boston, MA: Artech House.

Goff DR and Hansen KS (2002) *Fiber Optic Reference Guide: A Practical Guide to Communications Technology*, 3rd edn. Oxford, UK: Butterworth-Heinemann UK.

Hewak D (ed.) (1998) *Glass and Rare Earth Doped Glasses for Optical Fibres*. EMIS Datareview Series, INSPEC. London: The Institution of the Electrical Engineers.

Keck DB (1981) Optical fibre waveguides. In: Barnoski MK (ed.) *Fundamentals of Optical Fiber Communications*. New York: Academic Press.

Li T (ed.) (1985) *Optical Fiber Communications: Fiber Fabrication*. New York: Academic Press.

Personick SD (1985) *Fiber Optics: Technology and Applications*. New York: Plenum Press.

Schultz PC (1979) In: Bendow B and Mitra SS (eds) *Fiber Optics*. New York: Plenum Press.

Light Propagation

F G Omenetto, Los Alamos National Laboratory, Los Alamos, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Optical fibers have become a mainstay of our modern way of life. From phone calls to the internet, their presence is ubiquitous. The unmatched ability that optical fibers have to transmit light is an amazing asset that, besides being strongly part of our present, is poised to shape the future.

The foundation of the unique ability of optical fibers to transmit light hinges on the well-known physical concept of total internal reflection. This phenomenon is described by considering the behavior of light traveling from one material (A) to a different material (B). When light traverses the interface of the two materials with a specific angle, its direction of propagation is altered with respect to the normal to the surface, according to the well-known Snell law which relates the incident angle (onto the interface between A and B) to the refracted angle (away from the interface between A and B). This is written as $n_A \sin i = n_B \sin r$, where i is the angle of incidence, r the angle of refraction, and n_A and n_B are the refractive indices of the materials A and B. Under appropriate conditions, a critical value of the incidence angle (i_{CR}) exists for which the light does not propagate into B and is totally reflected back into A. This value is given by $i_{CR} = \arcsin(n_A/n_B)$. It follows that total internal reflection occurs when light is traveling from a medium with a higher index of

refraction into a medium with a lower index of refraction. As an approximation, this is the basic idea behind light propagation in fibers.

Optical fibers are glass strands that include a concentric core with a cladding wrapped around it. The index of refraction of the core is slightly higher than the index of refraction of the cladding, thereby creating the necessary conditions for the confinement of light within them. Such index mismatches between the core and the cladding are determined by using different materials for each, by doping the glass matrix before pulling the fiber to create an index gradient, by introducing macroscopic defects such as air-holes in the structure of the fiber cladding (such as in the recently developed photonic crystal fibers), or by using materials other than glass, such as polymers or plastic.

The real physical picture of light guiding in fibers is more complex than the conceptual description stated above. The light traveling in the fiber is, more precisely, an electromagnetic wave with frequencies that lie in the optical range, and the optical fiber itself constitutes an electromagnetic waveguide. As such, the waveguide supports guided electromagnetic modes which can take on a multitude of shapes (i.e., of energy distribution profiles), depending on the core/cladding index of refraction profiles. These profiles are often complex and determine the physical boundary conditions that influence the field distribution within the fiber. The nature of the fiber modes has been treated extensively in the literature where an in-depth description of the physical solutions can be found.

There are many distinguishing features that identify different optical fiber types. Perhaps the most important one, in relation to the electromagnetic modes that the fiber geometry supports, is the distinction that is made between single-mode fibers and multimode fibers. As the name implies, a single-mode fiber is an optical fiber whose index of refraction profile between core and cladding is designed in such a way that there is only one electromagnetic field distribution that the fiber can support and transmit. That is to say, light is propagated very uniformly across the fiber. A very common commercially used fiber of this kind is the Corning SMF-28 fiber, which is employed in a variety of telecommunication applications. Multimode fibers have larger cores and are easier to couple light into, yet the fact that many modes are supported implies that there is less control over the behavior of light during its propagation, since it is more difficult to control each individual mode.

The condition for single-mode propagation in optical fibers is determined during the design and manufacture of the fiber by its geometrical parameters and is specific to a certain operation wavelength. A quantity V is defined as $V = (2\pi/\lambda) \times a(n_{\text{core}}^2 - n_{\text{cladding}}^2)^{1/2}$, where a is the core radius and λ is the wavelength of light. For example, a step-index fiber (i.e., a fiber where the transition from the index of refraction of the core to the index of refraction in the cladding is abrupt), supports a single electromagnetic mode if $V < 2.405$.

Light propagation in optical fibers, however, is affected by numerous other factors besides the electromagnetic modes that the fiber geometry allows.

First, the ability to efficiently transmit light in optical fiber is dependent on the optical transparency of the medium that the fiber is made of. The optical transparency of the medium is a function of the wavelength (λ) of light that needs to be transmitted through the fiber. Optical attenuation is particularly high for shorter wavelengths and the transmission losses for ultraviolet wavelengths become considerably higher than in the near infrared. The latter region, due to this favorable feature, is the preferred region of operation for optical fiber based telecommunication applications. This ultimate physical limit is determined by the process of light being scattered off the atoms that constitute the glass, a process called Rayleigh scattering. The losses due to Rayleigh scattering are inversely proportional to the fourth power of the wavelength (losses $\propto \lambda^{-4}$), and therefore become quite considerable as λ is decreased.

The ability of optical fibers to transmit light over a certain set of wavelengths, is quantified by their optical attenuation. This parameter is defined as the ratio of the output power versus the input power and is usually measured in units of decibels (dB), i.e., attenuation (dB) $= 10 \log_{10}(P_{\text{out}}/P_{\text{in}})$.

The values that are found in the literature relate optical attenuation to distance and express the attenuation per unit length (such as dB/km). Single-mode fibers used in telecommunications are usually manufactured with doped silica glasses designed to work in the near infrared, in a wavelength range between 1.3 and 1.6 μm (1 μm $= 10^{-6}$ m) which provides the lowest levels of Rayleigh scattering and is located before the physical infrared absorption of silica (~1.65 μm) kicks in. Such fibers, which have been developed and refined for decades, have attenuation losses of less that 0.2 dB/km at a wavelength of 1.55 μm. Many other materials have been used to meet specific optical transmission needs. While silica remains by far the most popular material used for optical fiber manufacturing, other glasses are better suited for different portions of the spectrum: quartz glass can be particularly transmissive in the ultraviolet whereas plastic is sometimes

conveniently used for short distance transmission in the visible.

The ability to collect light at the input of the fiber represents another crucial and defining parameter. Returning to total internal reflection, such an effect can only be achieved by sending light onto an interface at a specific angle. In fibers, the light needs to be coupled into a circular core and reach the core cladding interface at that specific geometrical angle. The measure of the ability to couple light in the fiber is referred to as 'coupling efficiency'. The core dimension defines the geometry of the coupling process which, in turn, determines the efficiency of the propagation of light inside the fiber. In order to be coupled efficiently and therefore guided within the fiber, the light has to be launched within a range of acceptable angles. Such a range is usually defined or derivable from the numerical aperture (NA) of the fiber which is approximately $\mathrm{NA} = (n_{\mathrm{core}}^2 - n_{\mathrm{cladding}}^2)^{1/2}$, or can be alternatively described as a function of the critical angle for total internal reflection within the fiber by $\mathrm{NA} = n_{\mathrm{core}} \sin i_{\mathrm{CR}}$. Also, light has to be focused into the core of the fiber and therefore the interplay between core size, focusability, and NA (i.e., acceptance angle) of the fiber influence the coupling efficiency into the fiber (which in turn affects transmission). Typical telecommunication fibers have a 10 μm core (a NA of ~0.2) and have very high coupling efficiencies (of the order of 70–80%). Specialized single mode fibers (such as photonic crystal fibers) can have core diameters down to a few microns, significantly lower coupling efficiencies, and high numerical apertures. Multimode fibers, in contrast, are easy to couple into given their large core size.

Other factors that contribute to losses during propagation are due to the physical path that the actual optical fiber follows. A typical optical fiber strand is rarely laid along a straight line. These types of losses are called bending losses and can be intuitively understood by thinking that when the fiber is straight, the light meets its total reflection condition at the core-cladding interface, but when the fiber bends, the interface is altered and varies the incidence conditions of light at the interface. If the bend is severe, light is no longer totally reflected and escapes from the fiber.

Another important issue that influences propagation arises when the light that travels through the fiber is sent in bursts or in pulses of light. Pulsed light is very important in fiber transmission because the light pulse is the carrier of optical information, most often representing a binary digit of data. The ability to transmit pulses reliably is at the heart of modern day telecommunication systems. The features that affect the transmission of pulses through optical fibers depend, again, largely on the structure and constituents of the fiber. One of the most important effects on pulsed light is dispersion. This phenomenon causes the broadening in time of the light pulses during their travel through the fiber. This can be easily explained by a Fourier analogy by thinking that a short pulse in time (such as the 10–20 picosecond pulse duration that is used for optical communications, 1 picosecond $= 10^{-12}$ seconds) has a large frequency content (i.e., is composed of a variety of 'colors'). Since the speed of light through bulk media, and therefore through the fiber constituent, depends on wavelength, different 'colors' will travel at different speeds, arriving at the end of the fiber slightly delayed with respect to one another. This causes a net pulse broadening and is dependent on the properties of the fiber that are depicted in the dispersion curve of the fiber. This type of dispersion is called chromatic dispersion and is dependent on the distance the pulse travels in the fiber. A measure of the pulse broadening can be calculated by multiplying the dispersion value for the fiber (measured in ps/km) at the wavelength of operation times and the distance in kilometers that the pulse travels.

There are other types of dispersion that affect pulses traveling through fibers: light with different polarization travels at a different speed, thereby causing an analogous problem to chromatic distortion. In multimode fibers, different modes travel with different properties. Dispersion issues are extremely important in the design of the next-generation telecommunication systems. As a push is made toward higher transmission capacity, pulses become shorter and these issues have to be dealt with carefully. Specifically, transmission of light in optical fibers becomes more complex because the interaction between the light and the fiber material constituents becomes nonlinear. As pulses become shorter, their energy is confined in a smaller temporal interval, making their peak power (which is equal to energy/time) increase. If the peak power becomes sufficiently high, the atoms that form the glass of which the fiber is made are excited in a nonlinear fashion adding effects that need to be addressed with caution. Energetic pulses at a certain wavelength change their shape during propagation and can become severely distorted or change wavelength (i.e., undergo a frequency conversion process), thereby destroying the information that they were meant to carry.

Several physical manifestations of optical nonlinearity affect pulsed light propagation, giving rise to new losses but also providing new avenues for efficient transmission.

Perhaps the most successful example of the positive outcome of nonlinear effects in fibers is the optical soliton. The latter is an extremely stable propagating pulse that is transmitted through the fiber undistorted, since it is a result of a peculiar balance between the linear distortion and nonlinear distortion that cancel each other by operating in a specific region of dispersion of the optical fiber (the anomalous dispersion region).

Nonlinear effects in fibers are an extremely rich area of study which, besides several drawbacks, carries immense opportunities if the nonlinearities can be controlled and exploited.

See also

Dispersion Management. **Nonlinear Optics, Applications:** Pulse Compression via Nonlinear Optics. **Solitons:** Optical Fiber Solitons, Physical Origin and Properties.

Further Reading

Hecht J (2001) *Understanding Fiber Optics*. New York: Prentice Hall.

Snyder AW and Love JD (1983) *Optical Waveguide Theory*. Boston: Kluwer Academic Press.

Agrawal GP (2001) *Nonlinear Fiber Optics*. San Diego: Academic Press.

Measuring Fiber Characteristics

A Girard, EXFO, Quebec, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The optical fiber is divided in two types: multimode and singlemode. Each type is used in different applications and wavelength ranges and is consequently characterized differently. Furthermore, the corresponding test methods also vary.

The optical fiber characteristics may be divided into four categories:

- The optical characteristics (transmission related);
- The dimensional characteristics;
- The mechanical characteristics; and
- The environmental characteristics.

These categories will be reviewed in the following sections, together with their corresponding test methods.

Fiber Optical Characteristics and Corresponding Tests Methods

The following sections will describe the following optical characteristics:

- attenuation;
- macrobending sensitivity;
- microbending sensitivity;
- cut-off wavelength;
- multimode fiber bandwidth;
- differential mode delay for multimode fibers;
- chromatic dispersion;
- polarization mode dispersion;
- polarization crosstalk; and
- nonlinear effects.

Attenuation

The spectral attenuation of an optical fiber follows exponential power decay from the power level at a cross-section 1 to the power level at cross-section 2, over a fiber length L, as follows:

$$P_2(\lambda) = P_1(\lambda) \cdot e^{-\gamma(\lambda)L} \qquad [1]$$

$P_1(\lambda)$ is the optical power transmitted through the fiber core cross-section 1, expressed in mW; $P_2(\lambda)$ is the optical power transmitted through the fiber core cross-section 2 away from cross-section 1, expressed in mW; $\gamma(\lambda)$ is the spectral attenuation coefficient in linear units, expressed in km^{-1}; and L is the fiber length expressed in km.

Attenuation may be characterized at one or more specific wavelengths or as a function of wavelength. In the later case, attenuation is referred to spectral attenuation. Figure 1 illustrates such power decay.

Equation [1] may be expressed in relative units as follows:

$$\log_{10} P_2 = (\log_{10} P_1) \cdot - \gamma L \cdot \log_{10} e \qquad [2]$$

P is expressed in dBm units using the following definition.

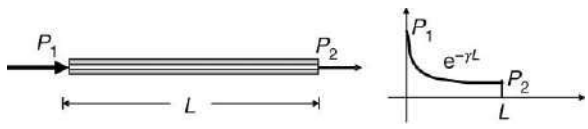

Figure 1 Power decay in an optical fiber due to attenuation.

The power in dBm is equal to 10 times the logarithm in base 10 of the power in mW; or

$$0\ \text{dBm} = 10\log_{10}(1\ \text{mW}) \qquad [3]$$

Then

$$\gamma(\lambda)[\text{km}^{-1}] = [(10\log_{10} P_1) - (10\log_{10} P_2)] \div L\log_{10} e \qquad [4]$$

$$\alpha(\lambda)[\text{dB/km}] = [P_1(\text{dBm}) - P_2(\text{dBm})]/L \qquad [5]$$

A new relative attenuation coefficient $\alpha(\lambda)$ with units of dB/km has been introduced, which is related to the linear coefficient $\gamma(\lambda)$ with units of km^{-1} as follows:

$$\alpha(\lambda) = (\log_{10} e)\cdot\gamma(\lambda) \approx 4{,}343\gamma(\lambda) \qquad [6]$$

$\alpha(\lambda)$ is the spectral attenuation coefficient of a fiber and is illustrated in Figure 2.

Test methods for attenuation

The attenuation may be measured by the following methods:

- cut-back method;
- backscattering method; and
- insertion loss method.

Cut-back method. The cut-back method is a direct application of the definition in which the power levels P_1 and P_2 are measured at two points of the fiber change of input conditions. P_2 is the power emerging from the far end of the fiber and P_1 is the power emerging from a point near the input after cutting the fiber.

The output power P_2 is recorded from the fiber under test (FUT) placed in the measurement setup. Keeping the launching conditions fixed, the FUT is cut to the cut-back length, for example 2 m from the launching point. The FUT attenuation, between the points where P_1 and P_2 have been measured, is calculated using P_1 and P_2, from the definition equations provided above.

Backscattering method. The attenuation coefficient of a singlemode fiber is characterized using bidirectional backscattering measurements. This method is also used for:

- attenuation uniformity;
- optical continuity;
- physical discontinuities;
- splice losses; and
- fiber length.

An optical time domain reflectometer (OTDR) is used for performing such characterization. Adjustment of laser pulsewidth and power is used to obtain a compromise between resolution (a shorter pulsewidth provides a better resolution but at lower power) and dynamic range/fiber length (higher power provides better dynamic range but with longer pulsewidth). An example of such an instrument is shown in Figure 3.

The measurement is applicable to various FUT configurations (e.g., cabled fiber in production or deployed in the field, fiber on a spool, etc.). Two unidirectional backscattering curves are obtained, one from each end of the fiber (Figure 4).

Each backscattering curve is recorded on a logarithmic scale, avoiding the parts at the two ends of the curves, due to parasitic reflections.

The FUT length is found from the time interval between the two ends of the backscattering loss

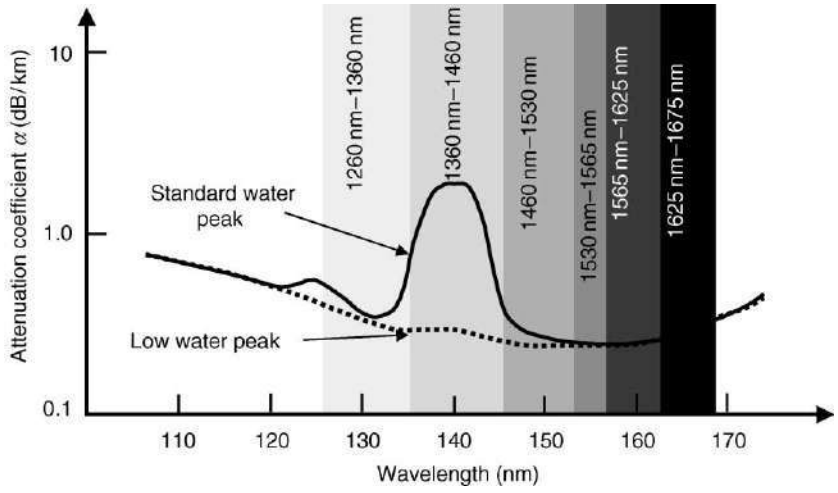

Figure 2 Typical spectral attenuation of a singlemode fiber.

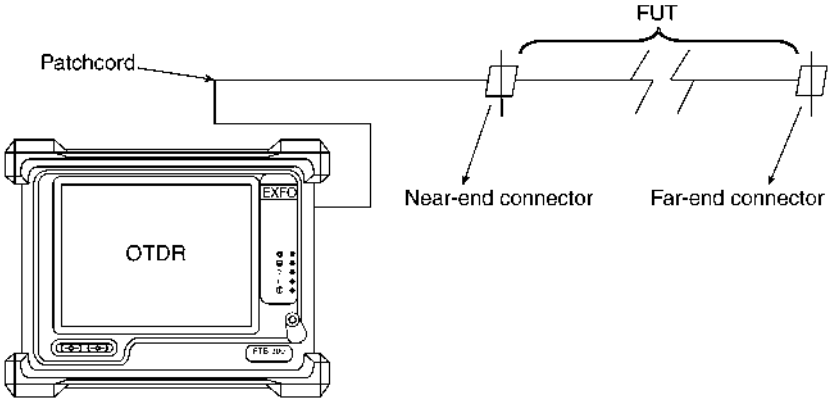

Figure 3 The backscattering method (OTDR).

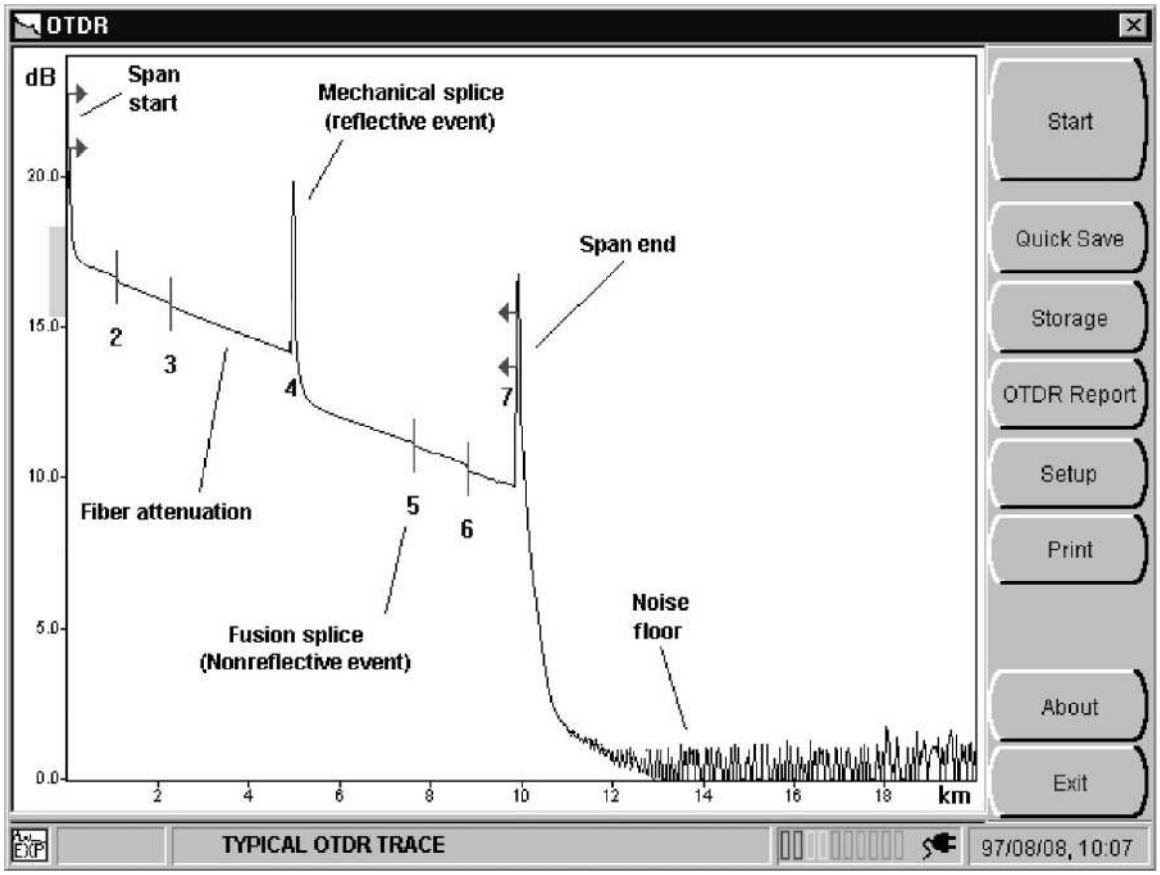

Figure 4 Unidirectional OTDR backscattering loss measurement.

curve together with the FUT group index of refraction, n_g, as:

$$L_f = c_{fs} \cdot \frac{\Delta T_f}{n_g} \qquad [7]$$

c_{fs} is the velocity of light in free space.

The bidirectional backscattering curve is obtained using two unidirectional backscattering curves and calculating the average loss between them. The end-to-end FUT attenuation coefficient is obtained from the difference between two losses divided by the difference of their corresponding distances.

Insertion loss method. The insertion loss method consists of the measurement of the power loss due to

the FUT insertion between a launching and a receiving system, previously interconnected (reference condition). The powers P_1 and P_2 are evaluated in a less straightforward way than in the cut-back method. Therefore, this method is not intended for use in manufacturing.

The insertion loss technique is less accurate than the cut-back, but has the advantage of being non-destructive for the FUT. Therefore, it is particularly suitable in the field.

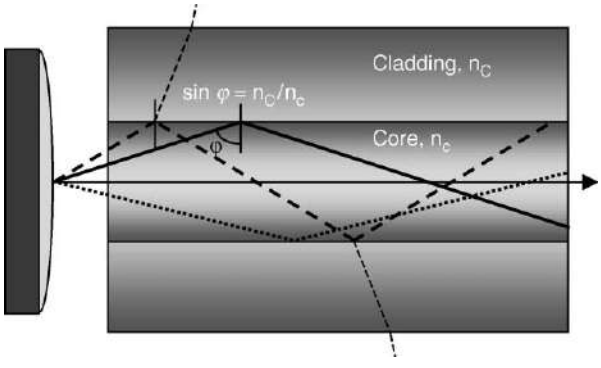

Figure 5 Total internal reflection and macrobending effect on the light rays.

Macrobending Sensitivity

Macrobending sensitivity is the property by which there is a certain amount of light leaking (loss) into the cladding when the fiber is bent and the bending angle is such that the condition of total internal reflection is no longer met at the core-cladding interface. Figure 5 illustrates such a case.

Macrobending sensitivity is a direct function of the wavelength: the longer the wavelength and/or the smaller the bending diameter, the more loss the fiber experiences. It is recognized that 1625 nm is a wavelength that is very sensitive to macrobending (see Figure 6).

The macrobending loss is measured by the power monitoring method (OTDR, especially for field assessment, see Figure 6) or the cut-back method.

Microbending Sensitivity

Microbending is a fiber property by which the core-cladding concentricity randomly changed along

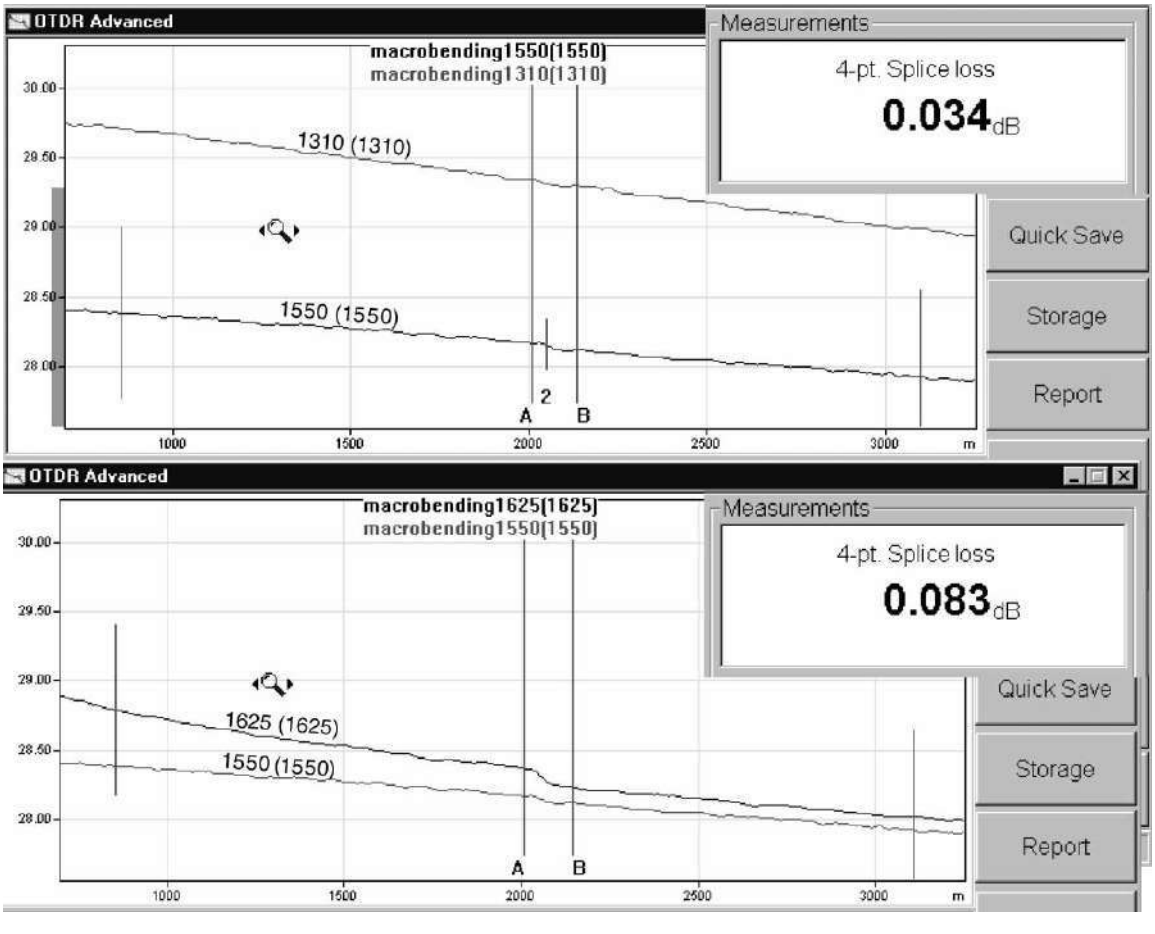

Figure 6 Macrobending sensitivity as a function of wavelength.

the fiber length. It causes the core to wobble inside the cladding and along the fiber length.

Four methods are available for characterizing microbending sensitivity in optical fibers:

- expandable drum for singlemode fibers and optical fiber ribbons over a wide range of applied linear pressure or loads;
- fixed diameter drum for step-index multimode, singlemode, and ribbon fibers for a fixed linear pressure;
- wire mesh and applied loads for step-index multimode and singlemode fibers over a wide range of applied linear pressure or loads; and
- 'basketweave' wrap on a fixed diameter drum for singlemode fibers.

The results from the four methods can only be compared qualitatively. The test is nonroutine for general evaluation of optical fiber.

Cut-off Wavelength

The cut-off wavelength is the shortest wavelength at which a single mode can propagate in a singlemode fiber. This parameter can be computed from the fiber refractive index profile (RIP). At wavelengths below the cut-off wavelength, several modes propagate and the fiber is no longer singlemode, but multimode.

In optical fibers, the change from multimode to singlemode behavior does not occur at a specific wavelength, but rather over a smooth transition as a function of wavelengths. Consequently, from a fiber-optic network standpoint, the actual threshold wavelength for singlemode performance is more critical. Thus an effective cut-off wavelength is described below.

The cut-off wavelength is defined as the wavelength greater than the ratio between the total power, in the higher-order modes and the fundamental mode, which has decreased to less than 0.1 dB. According to this definition, the second-order mode LP_{11} undergoes 19.3 dB more attenuation than the fundamental LP_{01} mode when the modes are equally excited.

Because the cut-off wavelength depends on the fiber length, bends, and strain, it is defined on the basis of the following three cases:

- fiber cut-off wavelength;
- cable cut-off wavelength; and
- jumper cable cut-off wavelength.

Fiber cut-off wavelength: Fiber cut-off wavelength λ_{fc} is defined for uncabled primary-coated fiber and is measured over 2 m with one loop of 140 mm radius loosely constrained, with the rest of the fiber kept essentially straight. The presence of a primary coating on the fiber usually will not affect λ_{fc}. However, the presence of a secondary coating may result in λ_{fc} being significantly shorter than that of the primary coated fiber.

Cable cut-off wavelength: Cable cut-off wavelength is measured prior to installation on a substantially straight 22 m cable length prepared by exposing 1 m of primary-coated fiber at both ends, the exposed ends each incorporating a 40 mm radius loop. Alternatively, this parameter may be measured on 22 m primary-coated uncabled fiber in the same configuration as for the λ_{fc} measurement.

Jumper cable cut-off wavelength: Jumper cable cut-off wavelength is measured over 2 m with one loop of 76 mm radius, or equivalent (e.g., split mandrel), with the rest of the jumper cable kept essentially straight.

Multimode Fiber Bandwidth for Multimode Fibers

The −3 dB bandwidth of a multimode optical fiber (or modal bandwidth) is defined as the lowest frequency where the magnitude of the baseband frequency response in optical power has decreased by 3 dB relative to the power at zero frequency. Modal bandwidth is also called intermodal dispersion as it takes into account the dispersion between the modes of propagation of the transmitted signal into the multimode fiber.

Various methods of reporting the results are available, but the results are typically expressed in terms of the −3 dB (optical power) frequency. Figure 7 illustrates modal bandwidth.

The bandwidth or pulse broadening may be normalized to a unit length, such as GHz · km, or ns/km.

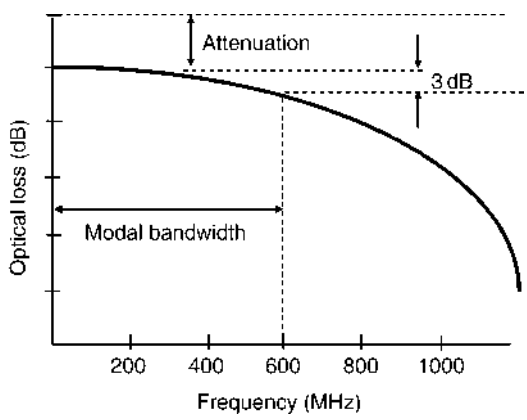

Figure 7 Determination of modal bandwidth.

Two methods are available for determining transmission capacity of multimode fibers:

- the frequency domain measurement method in which the baseband frequency response is directly measured in the frequency domain by determining the fiber response to a sinusoidally modulated light source; or
- the optical time domain measurement method (pulse distortion) in which the baseband response is measured by observing the broadening of a narrow pulse of light.

Differential Mode Delay for Multimode Fibers

Differential mode delay (DMD) characterizes the modal structure of a graded-index glass-core multimode fiber. DMD is useful for assessing the bandwidth performance of a fiber when used with short-pulse, narrow spectral-width laser sources.

The output from a singlemode probe fiber excites the multimode FUT at the test wavelength. The probe spot is scanned across the FUT endface, and the optical pulse delay is determined at specified radial offset positions between an inner and an outer limit. DMD is the difference in optical pulse delay time between the FUT fastest and slowest modes excited for all radial offset positions between and including the inner and the outer limits.

The related critical issues influencing DMD are the temporal width of the optical pulse, jitter in the timing, the finite bandwidth of the optical detector, and the mode broadening due to the source spectral width and the FUT chromatic dispersion.

The test method is commonly used in production and research facilities, but is not easily accomplished in the field. DMD may be a good predictor of the source launching conditions. DMD may be normalized to a unit length, such as ps/m.

Chromatic Dispersion

Chromatic dispersion in a singlemode fiber is a combination of material dispersion and waveguide dispersion (see Figure 8), and it contributes to pulse broadening and distortion in a digital signal.

Material dispersion is produced by the dopants used in glass and is important in all fiber types. Waveguide dispersion is produced by the wavelength dependence of the index of refraction and is critical in singlemode fibers only.

From the point of view of the transmitter, this is due to two causes:

- The presence of wavelengths in the source optical spectrum. Each wavelength has a different phase delay and group delay (different group velocities) along the fiber, because they travel under different index of refraction (or phase) as the index varies as a function of wavelengths, as shown in Figure 9.
- The other cause is the modulation of the source, which itself has two effects:
  - As bit-rates increase, the spectral width of the modulated signal increases and can become comparable to or exceed the spectral width of the source.
  - Chirp occurs when the source wavelength spectrum varies during the pulse. By convention, positive chirp at the transmitter occurs when the spectrum during the rise/fall of the pulse shifts towards shorter/longer wavelengths respectively. For a positive fiber dispersion coefficient, longer wavelengths are delayed relative to shorter wavelengths. Hence if the sign of the product of chirp and dispersion is positive, the two processes combine to produce pulse broadening. If the product is negative, pulse compression can occur over an initial fiber length until the pulse reaches a minimum width and then broadens again with increasing dispersion.

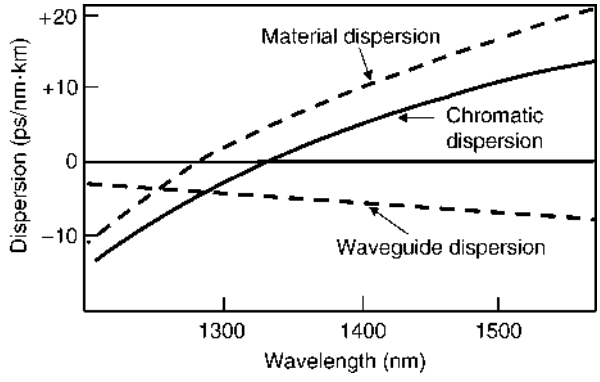

Figure 8 Contribution of the material and waveguide dispersions to the chromatic dispersion.

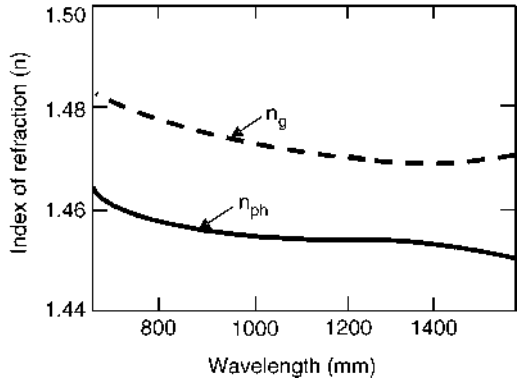

Figure 9 Difference between the phase and the group index of refraction.

The electric field propagating into the FUT may be simply described as follows:

$$E(t, z) = E_0 \sin(\omega t - \beta z) \qquad [8]$$

$\omega = 2\pi c/\lambda$ [rad/s] is the angular frequency; $\beta = kn = (\omega/c)n$ is the effective index; as β has units of m^{-1} it is often referred to as wavenumber or even sometimes propagation constant; k is the propagation constant.

The group delay τ_g is then given by:

$$d\beta/d\omega = \beta_1 = \tau_g \qquad [9]$$

An example of the group delay spectral distribution is shown in Figure 10. Assuming n is not complex; β_1 is the first-order derivative of β.

The group velocity v_g is given by

$$v_g = (d\beta/d\omega)^{-1} = -(\lambda^2/2\pi c)(d\beta/d\lambda)^{-1} \qquad [10]$$

The FUT input–output delay is given by

$$\tau_d = L/v_g \qquad [11]$$

L is the FUT length.

The group index of refraction n_g is given by

$$n_g = c/v_g = n - \lambda(dn/d\lambda) \qquad [12]$$

The dispersion parameter or dispersion coefficient D (ps/nm · km) is given by

$$D = -(\omega/\lambda)(d\tau_g/d\omega) = -(2\pi c/\lambda^2)(d^2\beta/d\omega^2)$$
$$= -(\lambda/c)(d^2n/d\lambda^2) \qquad [13]$$

$$d^2\beta/d\omega^2 = \beta_2 \qquad [14]$$

An example of the spectral distribution of D obtained from the group delay is shown in Figure 10.

β_2 (ps^2/km) is the group velocity dispersion parameter, so D may be related to β_2, as follows:

$$D = -(\omega/\lambda)\beta_2 \qquad [15]$$

When β_2 is positive then D is negative and vice-versa. The region where β_2 is positive is called normal dispersion while the negative-β_2 region is called anomalous dispersion.

At λ_0, β_1 is minimum, and $\beta_2 = 0$, then $D = 0$.

The dispersion slope S (ps/nm^2 · km), also called the differential dispersion parameter or second-order dispersion, is given by

$$S = dD/d\lambda = (\omega/\lambda)\beta_3 + (2\omega/\lambda^2)\beta_2 \qquad [16]$$

$$\beta_3 = d\beta_2/d\omega = d^3\beta/d\omega^3$$

At λ_0, β_1 is minimum, $\beta_2 = 0$, then $D = 0$; but S is not zero and depends on β_3. An example of the spectral distribution of S and S_0 is illustrated in Figure 10.

Overall, the general expression of β is given by

$$\beta(\omega) = \beta_0 + (\omega - \omega_0)\beta_1 + (1/2)(\omega - \omega_0)^2\beta_2$$
$$+ (1/12)(\omega - \omega_0)^3\beta_3 + \cdots \qquad [17]$$

Figure 11 illustrates the difference between dispersion unshifted fiber (ITU-T Rec. G.652), dispersion shifted fiber (ITU-T Rec. G.653) and nonzero dispersion shifted fiber (ITU-T Rec. G.655).

Test methods for the determination of chromatic dispersion

All methods measure the group delay at a specific wavelength over a range and use agreed fitting functions to evaluate λ_0 and S_0.

In the phase shift method, the group delay is measured in the frequency domain, by detecting, recording, and processing the phase shift of a sinusoidal modulating signal between a reference, a secondary fiber path, and the channel signal.

Setup variances exist and some do not require the secondary reference path. For instance, by using a reference optical filter at the FUT output it is possible to completely decouple the source from the phase-meter. With such an approach, chromatic dispersion may now be measured in the field over very long links using optical amplifiers (see Figure 12).

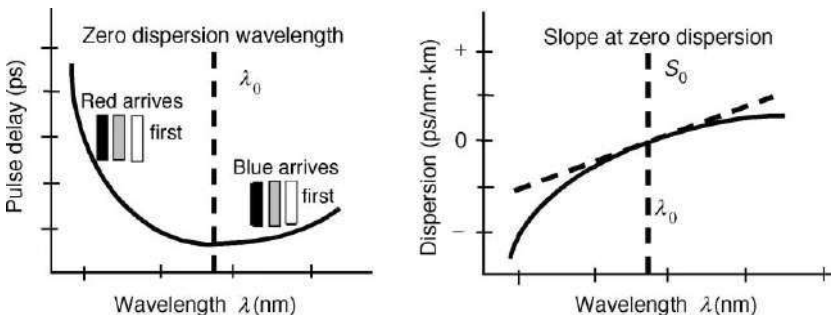

Figure 10 Relation between the pulse (group) delay and the (chromatic) dispersion.

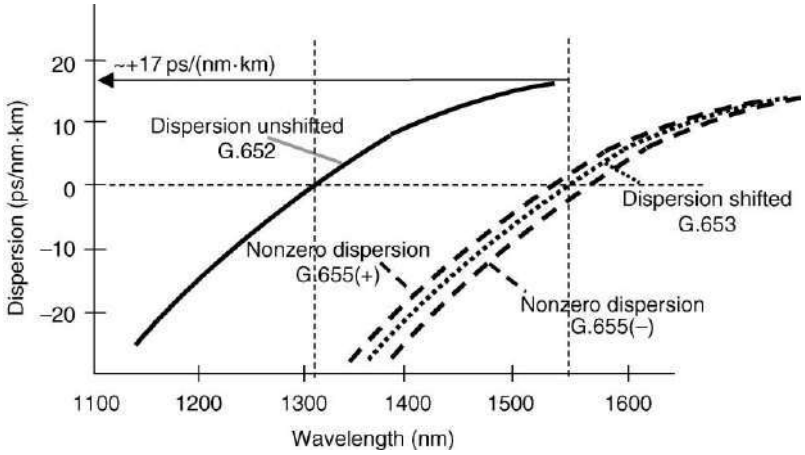

Figure 11 Chromatic dispersion for various types of fiber.

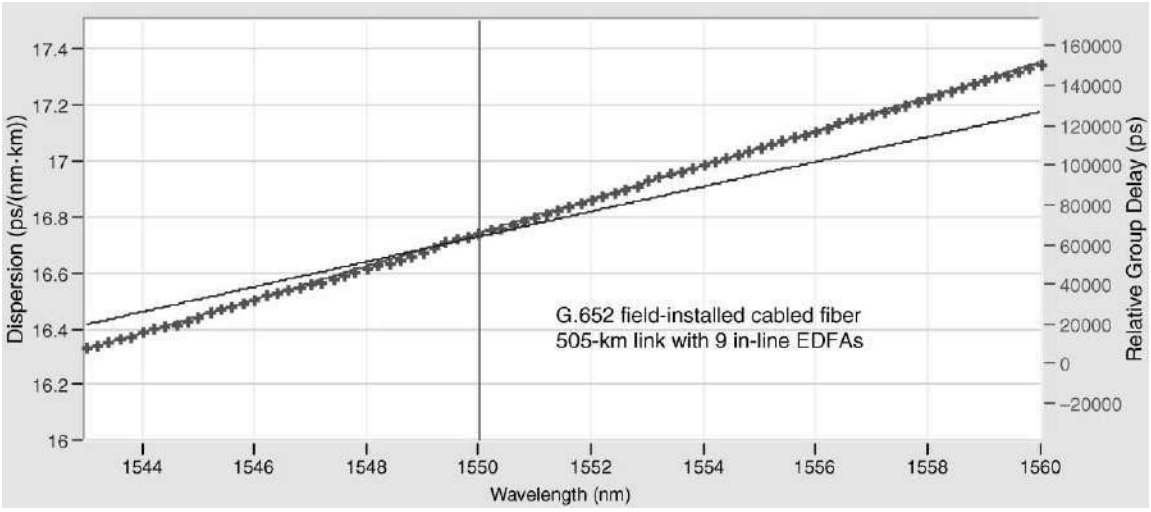

Figure 12 Test results for field-related chromatic dispersion using the phase-shift technique.

In the differential phase shift method, two detection systems are used together with two wavelength sources at the same time. In this case the chromatic dispersion may be determined directly from the two group delays. This technique usually offers faster and more reliable results but costs much more than the phase-shift technique which is usually preferred.

In the interferometric method, the group delay between the FUT and a reference path is measured by a Mach–Zehnder interferometer. The reference delay line may be an air path or a singlemode fiber standard reference material (SRM). The method can be used to determine the following characteristics:

- longitudinal chromatic dispersion homogeneity; and
- effect of overall or local influences, such as temperature changes and macrobending losses.

In the pulse delay method, the group delay is measured in the time domain, by detecting, recording, and processing the delay experienced by pulses at various wavelengths.

Polarization Mode Dispersion

Polarization mode dispersion (PMD) causes an optical pulse to spread in the time domain and may impair the performance of a telecommunications system. The effect can be related to differential phase and group velocities and corresponding arrival time of different polarization components of the pulse signal. For a sufficiently narrowband source, the effect can be related to a differential group delay (DGD), $\Delta\tau$, between a pair of orthogonally polarized principal states of polarization (PSP) at a given wavelength or optical frequency (see Figure 13a).

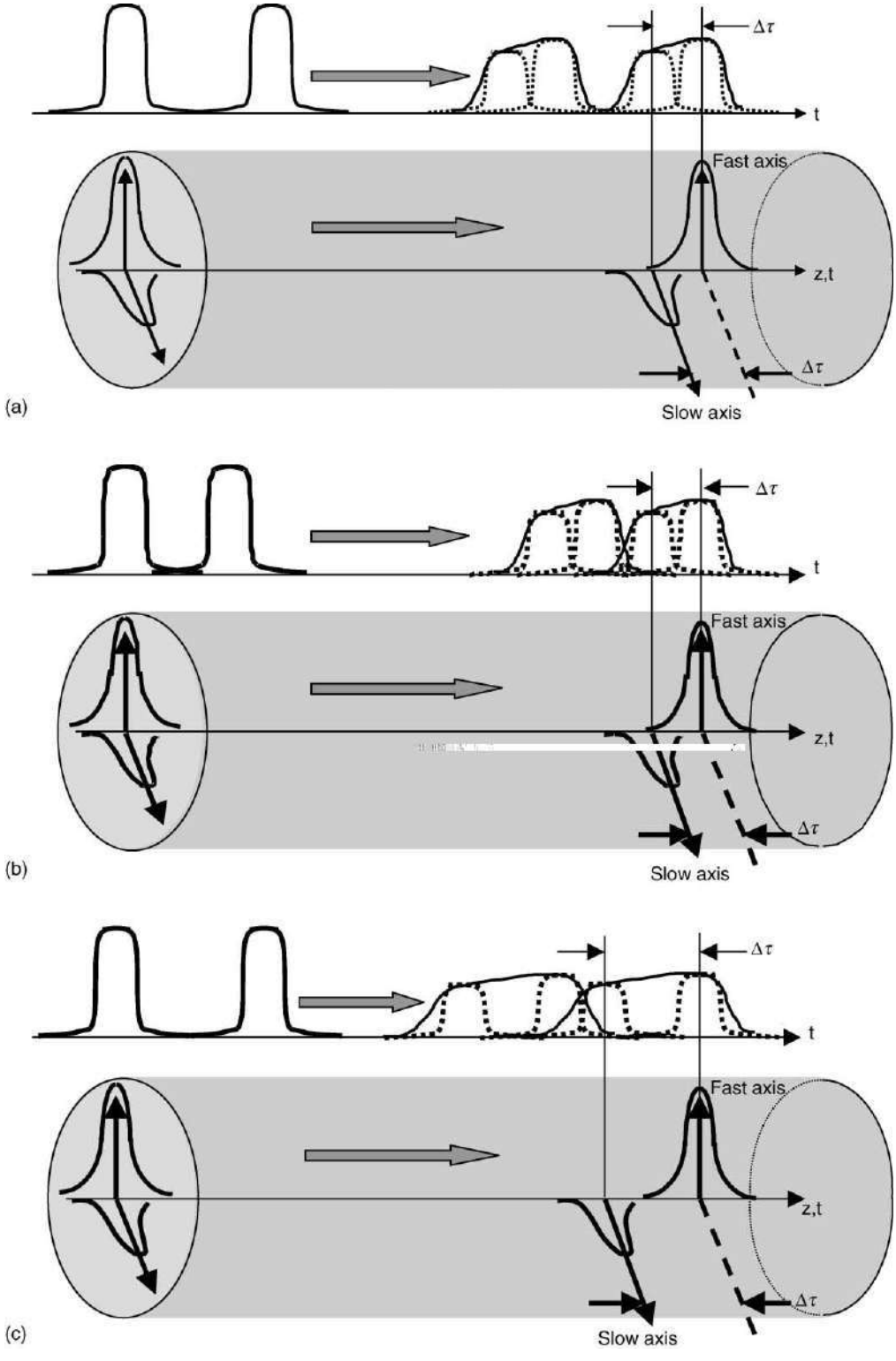

Figure 13 PMD effect on the pulse broadening and its possible pulse impairment. (a) The pulse is spread by the DGD $\Delta\tau$ but the bit rate is too low to create an impairment. (b) The pulse is spread by the DGD $\Delta\tau$ but the bit rate is high enough to create an impairment. (c) The DGD $\Delta\tau$ is large enough even at low bit rate to make the pulse spreading and creating impairment.

In an ideal circular symmetric fiber, the two PSPs propagate with the same velocity. However:

- a real fiber is not perfectly circular;
- the core is not perfectly concentric with the cladding;
- the core may be subjected to microbending;
- the core may present localized clusters of dopants; and
- the environmental conditions may stress the deployed cable and affect the fiber.

Each time the fiber undergoes local stresses and consequently birefringence. These asymmetry characteristics vary randomly along the fiber and in time, lead to a statistical behavior of PMD.

For a deployed cabled fiber at a given time and optical frequency, there always exist two PSPs such that the pulse spreading due to PMD vanishes, if only one PSP is excited. On the contrary, the maximum pulse spread due to PMD occurs when both PSPs are equally excited, and is related to the difference in their group delays, the DGD associated with the two PSPs. For broadband transmission, the DGD statistically varies as a function of wavelengths or frequencies and result in an output pulse that is spread out in the time domain (see **Figures 13a–c**). In this case, the spreading can be related to the RMS (root mean square) of DGD values $\langle \Delta\tau^2 \rangle^{1/2}$. However, if a known distribution such as the Maxwell distribution may be fit to the DGD distribution probability, then a mean (or average) value of the DGD $\langle \Delta\tau \rangle$ may be correlated to the RMS value and used as a system performance predictor in particular with a maximum value of the DGD distribution associated to a low probability of occurrence. This maximum DGD may then be used to define the quality of service that would tolerate values lower than this maximum DGD.

Test methods for polarization mode dispersion

Three methods are generically used for measuring PMD. Other methods or analyses may exist but they are generally not standardized or are limited in their applications.

- Stokes parameter evaluation (SPE)
  - Jones matrix eigenanalysis (JME)
  - Poincaré sphere analysis (PSA)
- Fixed analyzer (FA)
  - Extrema counting (EC)
  - Fourier transform (FT)
- Interferometry (INTY)
  - Traditional analysis (TINTY)
  - General analysis (GINTY).

All methods use a linearly polarized source at the FUT input and are suitable for laboratory measurements of factory lengths of fiber and cable. However, the interferometric method is the only method appropriate for measurements of cabled fiber that may be moving or vibrating such as is found in the field.

Stokes parameter evaluation. SPE determines PMD by measuring a response to a change of narrowband light (from a tuneable light source with broadband detector – JME, or a broadband source with a filtered detector such as an interferometer – PSA) across a wavelength range. The Stokes vector of the output light is measured for each wavelength. The change of these Stokes vectors with angular optical frequency (wavelength), ω and with the change in input SOP (state of polarization), yields the DGD as a function of wavelength.

For both JME and PSA analyses, three distinct and known linear SOPs (orthogonal on the Poincaré sphere) must be launched for each wavelength. **Figure 14** illustrates the test setup and examples of test results.

The JME and PSA method are mathematically equivalent.

Fixed analyzer. FA determines PMD by measuring a response to a change of narrowband light across a wavelength range. For each SOP, the change in output power that is filtered through a fixed polarization analyzer, relative to the power detected without the analyzer, is measured as a function of wavelength. **Figure 15** illustrates a test setup and examples of test results.

The resulting measured function can be analyzed in one of two ways:

- by counting the number of peaks and valleys (EC) of the curve and application of a formula. This analysis is considered as a frequency domain approach; and
- by taking the Fourier transform (FT) of the measured function. This FT is equivalent to the pulse spreading obtained by TINTY.

Interferometry. INTY uses a broadband light source and an interferometer. The fringe pattern containing the source auto-correlation, together with the PMD related cross-correlation of the emerging electromagnetic field, is determined by the interference pattern of the output light, i.e., the interferogram. The PMD determination for the wavelength range associated with the source spectrum is based on the envelope of the fringe

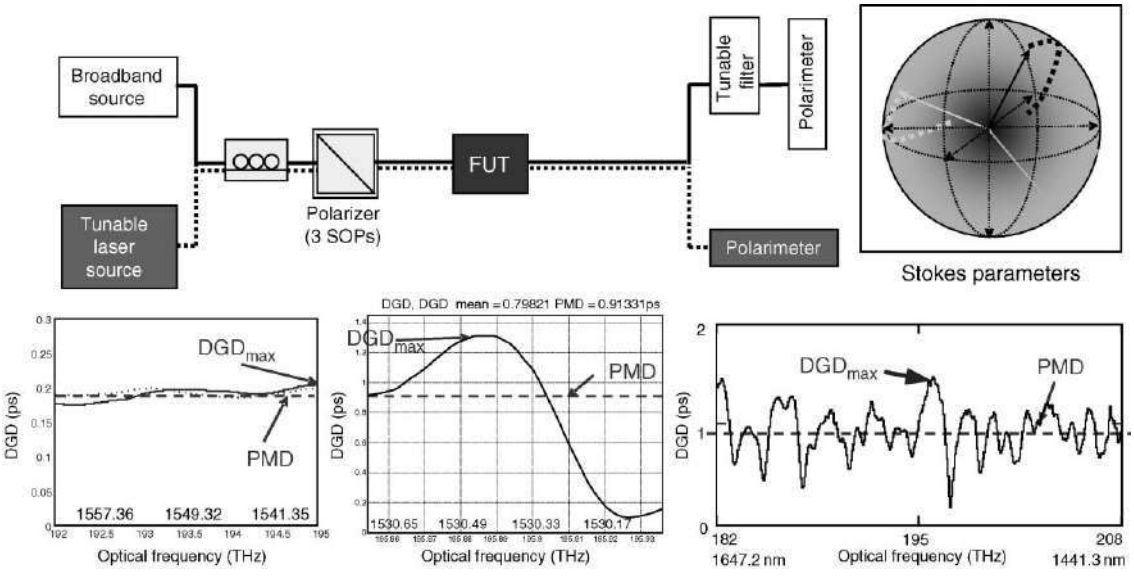

Figure 14 PMD by Stokes parameter evaluation method.

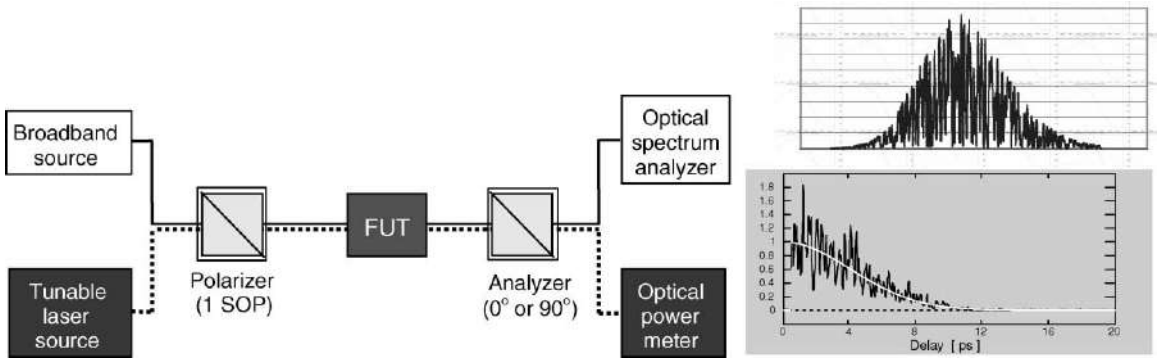

Figure 15 PMD by fixed analyzer method.

pattern interferogram. Two analyses are available to obtain the PMD:

- TINTY uses a set of specific operating conditions for its successful applications and a basic setup; and
- GINTY uses no limiting operating conditions, but in addition to the same basic setup, also using a modified setup compared to TINTY.

Figure 16 illustrates the test setup for both approaches and examples of test results.

Polarization Crosstalk

Polarization crosstalk is a characteristic of energy mixing/transfer/coupling between the two PSPs in a PMF (polarization maintaining fiber) when their isolation is imperfect. It is the measure of the strength of mode coupling or output power ratio between the PSPs within a PMF.

A PMF is an optical fiber capable of transmitting, under external perturbations, such as bending or lateral pressures, both HE_{11}^{x} and HE_{11}^{y} polarization modes whose electric field vector directions are orthogonally to each other and which have different propagation constants β_x and β_y.

Two methods are available for measuring the polarization crosstalk of PMF:

- *power ratio method*, which uses the maximum and minimum values of output power at a specified wavelength, and is applicable to fibers and connectors jointed to a PMF, and to two or more PMFs joined in series; and
- *in-line method*, which uses an analysis of the Poincaré sphere, and is applicable to single or cascaded sections of PMF, and to PMF interconnected with optical devices.

Nonlinear Effects

When the power of the transmission signal is increased to achieve longer span lengths at high bit

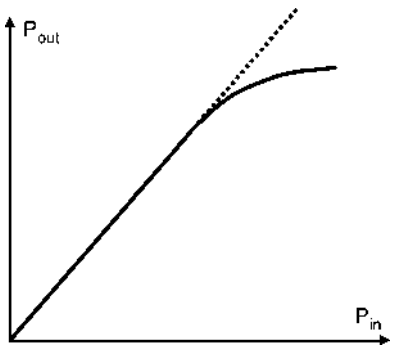

Figure 17 Power output-to-power input relationship for production of nonlinear effects.

expressed as follows:

$$n = n_0 + n_2 I \qquad [18]$$

n is the nonlinearity dependent index; n_0 is the linear part of the index; n_2 is the nonlinear index, also called the Kerr nonlinear index (2.2 to 3.4×10^{-16} cm^2/W); and I is the optical intensity inside the fiber.

The field propagation at a distance L into the fiber is described by the following equation:

$$E_{\text{out}}(L) = E_{\text{in}}(0)\exp[-a/2 + \mathrm{i}\beta + \gamma P(L,t)/2]L \qquad [19]$$

$a/2$ is the attenuation; $\mathrm{i}\beta$ is the phase of the wave; and $\gamma P(L,t)/2$ is the nonlinearity term;

$$\gamma = 2\pi n_2/\lambda A_{\text{eff}} \qquad [20]$$

γ is the nonlinearity coefficient and may be a complex number; A_{eff} is the fiber core effective area; $P(L,t)$ is the total power; and λ is the signal wavelength and t the time variable.

The nonlinear coefficient is defined as n_2/A_{eff}. This coefficient plays a critical role in the fiber and is closely related to system performance degradation due to nonlinearities when very high power is used.

Methods for measuring the nonlinear coefficient. Two methods are available for measuring the nonlinear coefficient:

- Continuous-wave dual-frequency (CWDF); and
- Pulsed single-frequency (PSF).

In the CWDF method, light from two wavelengths is injected into the fiber. At higher power, the light from the two wavelengths beat due to the nonlinearity and produce an output spectrum that is spread. The relationship of the power level to a particular spreading is used to calculate the nonlinear coefficient.

In the PSF method, the pulsed light from a single wavelength is injected into the fiber. Very short pulses (<1 ns) and their input peak power must be measured and related to the nonlinear spreading of the output spectrum.

Stimulated Brillouin scattering

In an intensity modulated system using a source with a narrow linewidth, significant optical power is transferred from the forward-propagating signal to a backward-propagating signal when the SBS power threshold is exceeded. At that point periodic regions of index of refraction produce a grating traveling at speed of sound away from the source. The grating reflects backward part of the incident light. The reflected sound waves (acoustic phonons) scatter light back to the source. Phase matching (or momentum conservation) dictates that the scattered light preferentially travels in the backward direction. The scattered light will be Doppler-shifted (downshifted or Brillouin-shifted) by approximately 11 GHz (at 1550 nm, for G.652 fiber). The scattered light has a very narrow spectrum (very coherent) and very close to the carrier signal and may be very detrimental.

Stimulated Raman scattering

SRS is an interaction between light and the fiber molecular vibrations as adjacent atoms vibrate in opposite directions (an 'optical phonon'). Some of the energy of the main carrier (optical pump wave) is transferred to the molecules, thereby further increasing the amplitude of their vibrations. If the vibrations become large enough, a threshold is reached at which the local index of refraction changes. These local changes then scatter light in all directions similar to Rayleigh scattering. However, unlike Rayleigh scattering, the wavelength of the Raman scattered light is shifted to longer wavelengths by an amount that corresponds to the molecular vibration frequencies and the Raman signal spreads over a large spectrum.

Self-phase modulation

SPM is the effect that a powerful pulse has on its own phase, considering that in eqn [18], $I(t)$ varies in time:

- $I(t) \rightarrow n(t) = n_0 + n_2 I(t) \rightarrow$ modulates the phase $\beta(t)$ of the pulse; and
- $\mathrm{d}I/\mathrm{d}t \rightarrow \mathrm{d}n/\mathrm{d}t \rightarrow \mathrm{d}\beta/\mathrm{d}t$(chirp) $\rightarrow$ broadening in the frequency domain $\rightarrow$ broadening in the time domain.

$I(t)$ peaks at the center of the pulse (peak power) and consequently increases the index of refraction. A higher index causes the wavelengths in the center of the pulse to accumulate phase more quickly than at the wings of the pulse:

- this causes wavelength stretching (shift to longer wavelengths) at pulse leading edge (risetime); and
- this causes wavelength compression (shift to shorter wavelengths) at pulse trailing edge (falltime).

The pulse will then broaden with negative (normal) dispersion and shorten with positive (anomalous) dispersion. SPM may then be used for dispersion compensation considering that self-phase modulation imposes $C > 0$ (positive chirp). It can cancel the dispersion if properly manage as a function of the sign of the dispersion. SPM is one of the most critical nonlinear effects for the propagation of soliton or very short pulses over very long distance.

Cross-phase modulation

XPM is the effect that a powerful pulse has on the phase of an adjacent pulse from another WDM system channel traveling in phase or at slightly the same group velocity. It concerns spectral interference between two WDM channels:

- the increasing $I(t)$ at the leading edge of the interfering pulse shifts the other pulse to longer wavelength; and
- decreasing $I(t)$ at the trailing edge of the interfering pulse shifts the other pulse to shorter wavelengths.

This produces spectral broadening, which dispersion converts to temporal broadening depending on the sign of the dispersion. XPM effect is similar to SPM except it depends on the channel count.

Four-wave (four-photon) mixing

FWM is the by-product production effect from two or more WDM channels. For two channels $I(t)$ modulates the phase of each signal (ω_1 and ω_2). An intensity modulation appears at the beat frequency $\omega_1 - \omega_2$.

Two sideband frequencies are created in a similar way as harmonics generation. New wavelengths are created in a number equal to $N^2(N-1)/2$, where $N =$ number of original wavelengths.

Fiber Dimension Characteristics and Corresponding Tests Methods

Table 1 provides a list of the various fiber dimensional characteristics and their corresponding test methods.

Table 1 Fiber dimensional characteristics

Attribute	*Measured parameter*
Fiber geometry	Core/cladding diameter Core/cladding noncircularity Core-cladding concentricity error
Numerical aperture	
Mode field diameter	
Coating geometry	
Length	

Fiber Geometry Characteristics

The fiber geometry is related to the core and cladding characteristics.

Core

The core center is the center of a circle which best fits the points at a constant level in the near-field intensity profile emitted from the central region of the fiber, using wavelengths above and/or below the cut-off wavelength.

The RIP can be measured by refracted near field (RNF) or transverse interferometry techniques and transmitted near field (TNF).

The core concentricity error is the distance between the core center and the cladding center. This definition applies very well for multimode fibers. The distance between the center of the near field profile and the center of the cladding is also used for singlemode fibers.

The mode field diameter (MFD) represents a measure of the transverse electromagnetic field intensity of the mode in a fiber cross-section and it is defined from the far-field intensity distribution.

The MF is the singlemode field distribution of the LP_{01} mode, giving rise to a spatial intensity distribution in the fiber.

The MF concentricity error is the distance between the MF center and the cladding center.

The core noncircularity is a measure of the core ellipticity. This parameter is one of the causes for creating birefringence in the fiber and consequently PMD.

Cladding

The cladding is the outermost region of constant refractive index in the fiber cross-section.

The cladding center is the center of a circle best fitting the outer limit (boundary) of the cladding.

The cladding diameter is the diameter of the circle defining the cladding center.

The cladding noncircularity is a measure of the difference between the diameters of the two circles defined by the cladding tolerance field divided by the nominal cladding diameter.

Coating

The primary coating is one or more layers of protective material applied to the cladding during or after the drawing process to protect the cladding surface (e.g., a 250 μm protective coating). The secondary coating is one or more layers of protective material applied over the primary coating in order to give additional protection or to provide a particular structure.

Measurement of the fiber geometrical attributes
The fiber geometry is measured by the following methods:

- TNF;
- RNF;
- Side-view technique/transverse interference;
- TNF image technique; and
- Mechanical diameter.

Test instrumentation may incorporate two or more methods such as the one shown in Figure 18.

Transmitted near-field technique. The cladding diameter, core concentricity error, and cladding noncircularity are determined from the near-field intensity distribution. Figure 19 provides a series of examples of test results from TNF measurements.

Refracted near-field technique. The RIP across the entire fiber (core and cladding) can be directly obtained from the RNF measurement, as shown in Figure 20.

The geometrical characteristics of the fiber can be obtained from the refractive index distribution using suitable algorithms:

- core/cladding diameter;
- core/cladding concentricity error;
- core/cladding noncircularity;
- maximum numerical aperture (NA); and
- index and relative index of refraction difference.

Figure 21 illustrates the core geometry.

Side-view technique/transverse interference. The side-view method is applied to singlemode fibers to determine the core concentricity error, cladding diameter and cladding noncircularity by measuring the intensity distribution of light that is refracted inside the fiber. The method is based on an interference microscope focused on the side view of an FUT illuminated perpendicular to the FUT axis. The fringe pattern is used to determine the RIP.

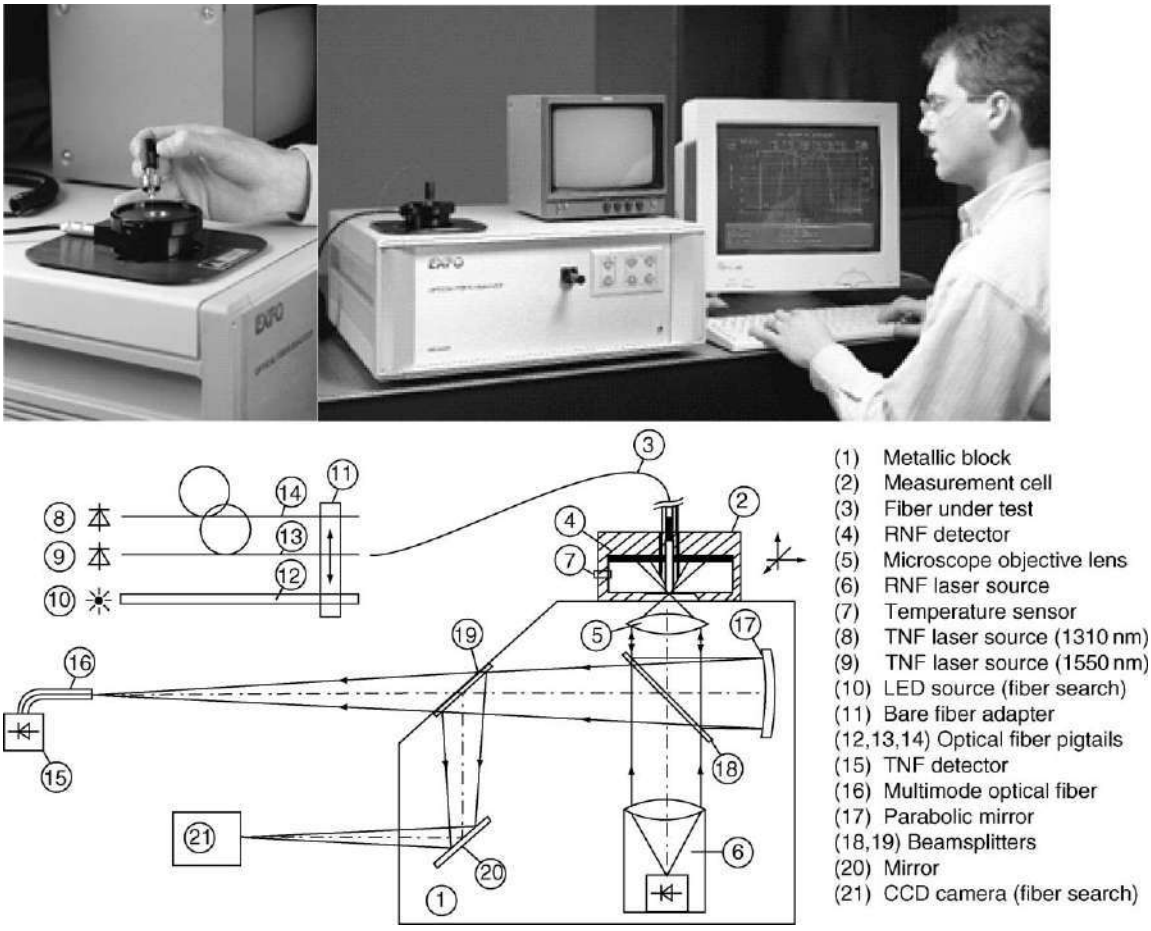

Figure 18 RNF/TNF combined instrumentation.

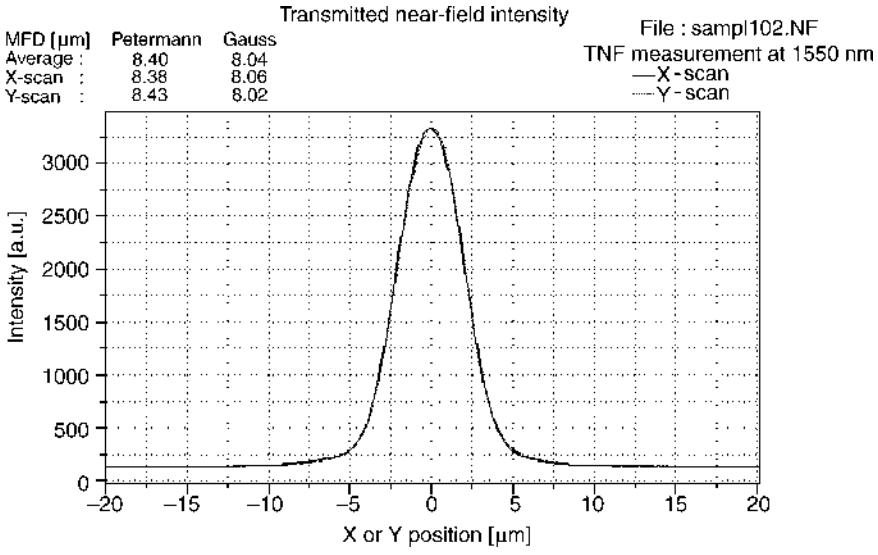

Figure 19 MFD measurement by TNF.

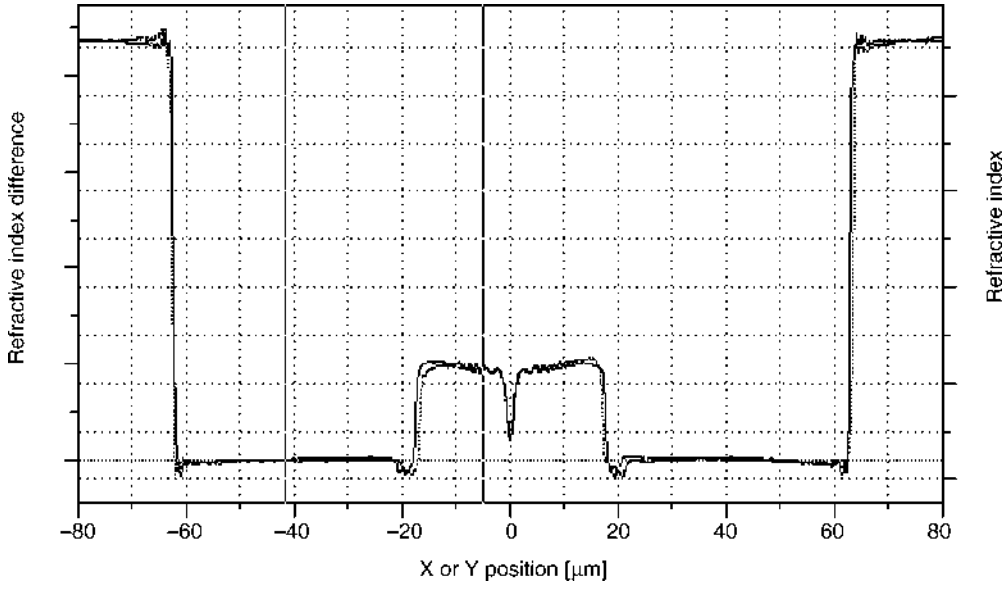

Figure 20 RIP from RNF measurement.

TNF image technique. The TNF image technique, also called near-field light distribution, is used for the measurement of the geometrical characteristics of singlemode fibers. The measurement is based on analysis of magnified images at the FUT output. Two subsets of the method are available:

- grey-scale technique which performs an x–y near-field scan using a video system; and
- Single near-field scanning technique performing a one-dimensional scan.

Mechanical diameter. This is a precision mechanical diameter measurement technique used to accurately determine the cladding diameter of silica fibers. The technique uses an electronic micrometer such as based on a double-pass Michelson interferometer. The technique is used for providing calibrated fibers to the industry as SRM.

Numerical Aperture

The NA is an important attribute for multimode fibers in order to predict their launching efficiency, joint loss at splices and micro/macrobending characteristics.

A method is available for the measurement of the angular radiant intensity distribution (far-field)

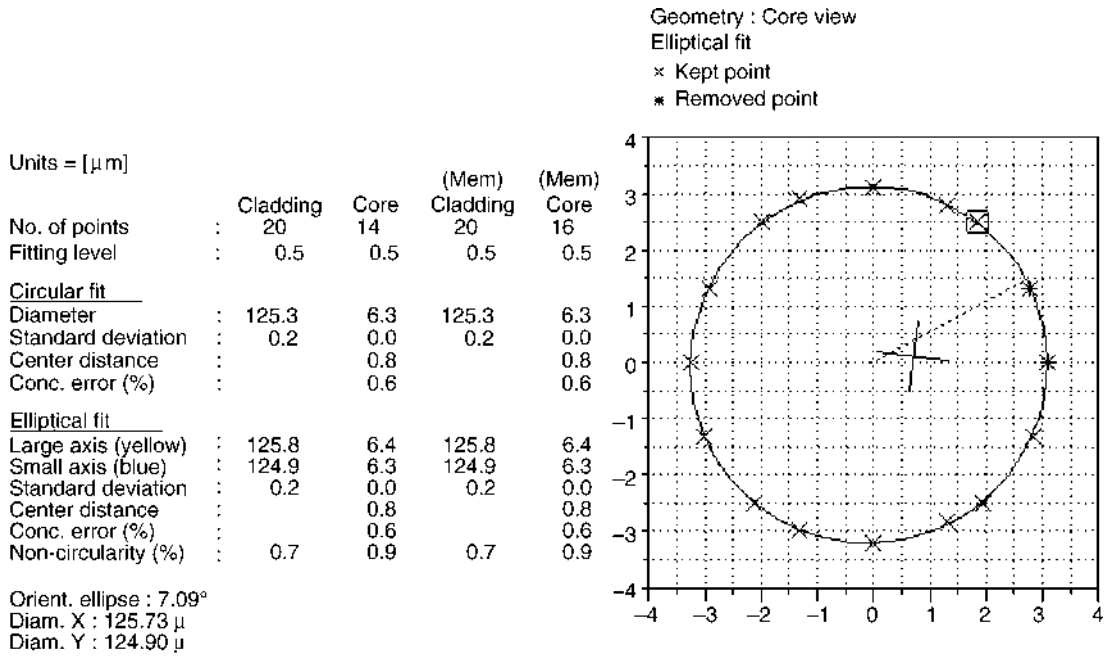

Figure 21 Core geometry by RNF measurement.

distribution or the RIP at the output of an FUT. NA can be determined from the analysis of the test results.

Mode Field Diameter

The definition of the MFD is given in the section describing the core center above. Four measurement methods are available:

- direct far-field scan determines MFD from the far-field intensity distribution;
- variable aperture in far field determines MFD from the complementary aperture transmission function $a(x)$, $x = d \tan\theta$ being the aperture radius and d the distance between the aperture and the FUT:

$$\mathrm{MFD} = \frac{\lambda}{\pi d}\left[\int_0^\infty a(x)\frac{x}{(x^2+d^2)^2}\,\mathrm{d}x\right]^{-1/2} \qquad [21]$$

Equation [21] is valid for small-θ approximation.

- near-field scan determines MFD from the near-field intensity distribution I_{NF}, r being the radial coordinate:

$$\mathrm{MFD} = 2\left[2\frac{\int_0^\infty rI_{\mathrm{NF}}(r)\,\mathrm{d}r}{\int_0^\infty r\left[\frac{\mathrm{d}I^{1/2}(r)}{\mathrm{d}r}\right]^2\mathrm{d}r}\right]^{1/2} \qquad [22]$$

Equation [22] is valid for small-θ approximation; and

- bidirectional backscattering uses an OTDR and bidirectional measurements to determine MFD by comparing the FUT results with a reference fiber.

Effective Area

A_{eff} is a critical nonlinearity parameter and is defined as follows:

$$A_{\mathrm{eff}} = \frac{2\pi[\int_0^\infty I(r)r\,\mathrm{d}r\,]^2}{\int_0^\infty I(r)^2 r\,\mathrm{d}r} \qquad [23]$$

$I(r)$ is the field intensity distribution of the fiber fundamental mode at radius r. The integration in the equation is carried out over the entire fiber cross-section. For a Gaussian approximation:

$$I(r) = \exp - 2\left(\frac{2r}{\mathrm{MFD}}\right)^2 \qquad [24]$$

which yields:

$$A_{\mathrm{eff}} = \frac{\pi}{4}\mathrm{MFD}^2 \qquad [25]$$

Three methods are available for the measurement of A_{eff}:

- direct far-field;
- variable aperture in far-field; and
- near-field.

Mechanical Measurement and Test Methods

Table 2 describes the various mechanical characteristics and their corresponding test methods.

Proof Stressing

The proof stress level is the value of tensile stress or strain applied to a full fiber length over a short period of time. The method for fiber proof stressing is the longitudinal tension which describes procedures for applying tensile loads to a length of fiber. The fiber stress is calculated from the applied tension. The tensile load is applied over short periods of time but not too short in order for the fiber to experience proof stress.

Residual Stress

Residual stress is the stress built up by the thermal expansion difference between core and cladding during the fiber drawing process or splicing. Methods are available for measuring residual stress based on polarization effects. A light beam produced by a rotating polarizer propagates to the x-axis direction while the beam polarization is in the $y-z$ plane and the fiber longitudinal axis in the z-axis. The light experiences different phase shift in the y- and z-axis due to the FUT birefringence. The photoelastic effect gives the relationship between this phase change and residual stresses.

Stress Corrosion Susceptibility

The stress corrosion susceptibility is related to the dependence of crack growth on applied stress. It depends on the environmental conditions and static and dynamic values may be observed.

Environmental Characteristics

Table 3 lists the fiber characteristics related to the effect of the environment.

Table 2 Fiber mechanical characteristics

Attribute
Proof stress
Residual stress
Stress corrosion susceptibility
Tensile strength
Stripability
Fiber curl

Table 3 Fiber environmental characteristics

Attribute
Hydrogen aging
Nuclear gamma irradiation
Damp heat
Dry heat
Temperature cycling
Water immersion

Hydrogen Aging for Low-Water-Peak Single-Mode Fiber

Hydrogen aging on low-water peak fibers, such as G.652.C, is based on a test performed at 1.0 atmosphere of hydrogen pressure at room temperature over a period of one month. Other proportional combinations are possible.

Nuclear Gamma Irradiation

Nuclear radiation is considered on the basis of a steady state response of optical fibers and cables exposed to gamma radiation and to determine the level of related radiation-induced attenuation produced in singlemode or multimode cabled or uncabled fibers.

The fiber attenuation generally increases when exposed to gamma radiation. This is primarily due to the trapping of radiolytic electrons and holes at defect sites in the glass (i.e., the formation of 'color centers'). Two regimes are considered:

- the low dose rate suitable for estimating the effect of environmental background radiation; and
- the high dose rate suitable for estimating the effect of adverse nuclear environments.

The effects of environmental background radiation are measured by the attenuation (cut-back method). The effects of adverse nuclear environments are tested by power monitoring before, during and after FUT exposure.

See also

Fiber and Guided Wave Optics: Nonlinear Effects (Basics). **Imaging:** Interferometric Imaging. **Interferometry:** Gravity Wave Detection; Overview. **Polarization:** Introduction. **Scattering:** Scattering Phenomena in Optical Fibers.

Further Reading

Agrawal GP (1997) *Fiber-Optic Communication Systems*, 2nd edn. New York: John Wiley & Sons.

Agrawal GP (2001) *Nonlinear Fiber Optics*, 3rd edn. San Diego, CA: Academic Press.
Girard A (2000) *Guide to WDM Technology and Testing*. Quebec: EXFO, pp. 194.
Hecht J (1999) *Understanding Fiber Optics*, 3rd edn. Upper Saddle River, NJ: Prentice Hall.
Masson B and Girard A (2004) *FTTx PON Guide, Testing Passive Optical Networks*. Quebec: EXFO, pp. 56.
Miller JL and Friedman E (2003) *Optical Communications Rules of Thumb*. New York: McGraw-Hill, pp. 428.
Neumann E-G (1988) *Single-Mode Fibers, Fundamentals*. Berlin: Springer-Verlag.

Nonlinear Effects (Basics)

G Millot and P Tchofo-Dinda, Université de Bourgogne, Dijon, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Many physical systems in various areas such as condensed matter or plasma physics, biological sciences, or optics, give rise to localized large-amplitude excitations having a relatively long lifetime. Such excitations lead to a host of phenomena referred to as nonlinear phenomena. Of the many disciplines of physics, the optics field is probably the one in which practical applications of nonlinear phenomena have been the most fruitful, in particular, since the discovery of the laser in 1960. This discovery has thus led to the advent of a new branch in optics, referred to as *nonlinear optics*. The applications of nonlinear phenomena in optics include the design of various kinds of laser sources, optical amplifiers, light converters, light-wave communication systems for data transmission purposes, to name a few.

In this article, we present an overview of some basic principles of nonlinear phenomena that result from the interaction of light waves with dielectric waveguides such as optical fibers. These nonlinear phenomena can be broadly divided into two main categories, namely, parametric effects and scattering phenomena. Parametric interactions arise whenever the state of the dielectric matter is left unchanged by the interaction, whereas scattering phenomena imply transitions between energy levels in the medium. More fundamentally, parametric interactions originate from the electron motion under the electric field of a light wave, whereas scattering phenomena originate from the motion of heavy ions (or molecules).

Linear and Nonlinear Signatures

The macroscopic properties of a physical system can be obtained by analyzing the response of the system under an external excitation. For example, consider at time t the response of a system, such as an amplifier, to an input signal $E_1 = A\sin(\omega t)$. In the low-amplitude limit of the output signal, the response R_1 of the system is proportional to the excitation:

$$R_1 = \alpha_1 E_1 \qquad [1]$$

where α_1 is a constant. This type of behavior corresponds to the so-called linear response. In general, a physical system executes a linear response when the superposition of two (or more) input signals E_1 and E_2 yields a response which is a superposition of the output signals, as schematically represented in Figure 1:

$$R = \alpha_1 R_1 + \alpha_2 R_2 \qquad [2]$$

Now, in almost all real physical systems, if the amplitude of an excitation E_1 becomes sufficiently large, distortions will occur in the output signals. In other words, the response of the system will no longer be proportional to the excitation, and consequently, the law of superposition of states will no longer be observed. In this case, the response of the system may take the following form:

$$R = \alpha_1 E_1 + \alpha_2 E_1^2 + \alpha_3 E_1^3 + \ldots \qquad [3]$$

which involves not only a signal at the input frequency ω, but also signals at frequencies 2ω, 3ω, and so on. Thus, harmonics of the input signal are generated. This behavior, called *nonlinear response*, is at the origin of a host of important phenomena in many

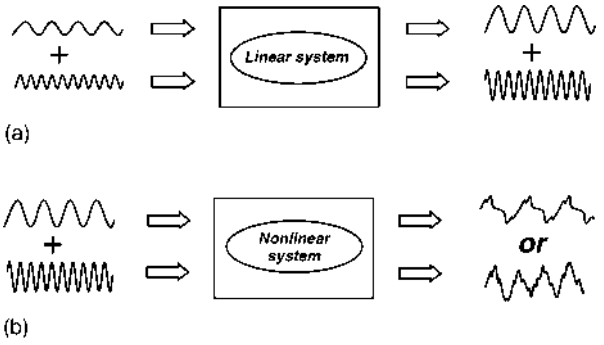

Figure 1 Schematic representation of linear and nonlinear responses of a system, to two input signals.

branches of sciences, such as in condensed matter physics, biological sciences, or in optics.

Physical Origin of Optical Nonlinearity

Optical nonlinearity originates fundamentally from the action of the electric field of a light wave on the charged particles of a dielectric waveguide. In contrast to conductors where charges can move throughout the material, dielectric media consist of *bound* charges (ions, electrons) that can execute only relatively limited displacements around their equilibrium positions. The electric field of an incoming light wave will cause the positive charges to move in the polarization direction of the electric field whereas negative charges will move in the opposite direction. In other words, the electric field will cause the charged particles to become dipoles, as schematically represented in Figure 2.

Then each dipole will vibrate under the influence of the incoming light, thus becoming a source of radiation. The global light radiated by all the dipoles represents the scattered light. When the charge displacements are proportional to the excitation (incoming light), i.e., in the low-amplitude limit of scattered radiation, the output light vibrates at the same frequency as that of the excitation. This process corresponds to Rayleigh scattering. On the other hand, if the intensity of the excitation is sufficiently large to induce displacements that are not negligible with respect to atomic distances, then the charge displacements will no longer be proportional to the excitation. In other words, the response of the medium, which is no longer proportional to the excitation, becomes nonlinear. In this case, the scattered waves will be generated not only at the excitation frequency (the Kerr effect), say ω, but also at frequencies that differ from ω (e.g., 2ω, 3ω). On the other hand, it is worth noting that when all induced dipoles vibrate coherently (that is, their relative phase does not vary randomly), their individual radiation may, under certain conditions, interfere constructively and lead to a global field of high intensity.

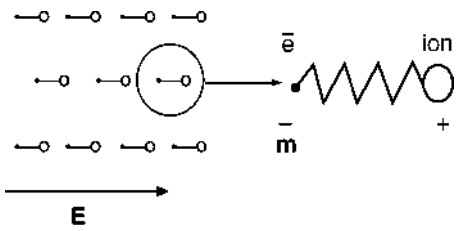

Figure 2 Electric dipoles in a dielectric medium under an external electric field.

The condition of constructive interference is the phase-matching condition.

In practice, the macroscopic response of a dielectric is given by the polarization, which corresponds to the total amount of dipole moment per unit volume of the dielectric. As the mass of an ion is much larger than that of an electron, the amplitude of the ion motion is generally negligible with respect to that of the electrons. As a consequence, the electron motion generally leads to the dominant contribution in the macroscopic properties of the medium. The behavior of an electron under an optical electric field is similar to that of a particle embedded in an anharmonic potential. A very simple model (called the Lorentz model) that provides a deep insight into the dielectric response consists of an electron of mass m and charge $-e$ connected to an ion by an elastic spring (see Figure 2). Under the electric field $E(t)$, the electron executes a displacement $x(t)$ with respect to its equilibrium position, which is governed by the following equation:

$$\frac{d^2x}{dt^2} + 2\lambda\frac{dx}{dt} + \omega_0^2 x + (a^{(2)}x^2 + a^{(3)}x^3 + \cdots) = -\frac{e}{m}E(t) \qquad [4]$$

where $a^{(2)}$, $a^{(3)}$, and so on, are constant parameters, ω_0 is the resonance angular frequency of the electron, and λ is the damping coefficient resulting from the dipolar radiation. When the amplitude of the electric field is sufficiently large, then the restoring force on the electrons becomes a nonlinear function of x; hence the presence of terms such as $a^{(2)}x^2$, $a^{(3)}x^3$, and so on, in eqn [4]. In this situation, the macroscopic response of the dielectric is the polarization

$$P = -\sum ex(\omega, 2\omega, 3\omega, \ldots) \qquad [5]$$

where $x(\omega, 2\omega, 3\omega, \ldots)$ is the solution of eqn [4] in the frequency domain, and the summation extends over all the dipole moments per unit volume. In terms of the electric field E the polarization may be written as

$$P = \varepsilon_0(\chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \cdots) \qquad [6]$$

where $\chi^{(1)}$, $\chi^{(2)}$, $\chi^{(3)}$, and so on, represent the susceptibility coefficients. Figure 3 (top left) illustrates schematically the polarization as a function of the electric field.

In particular, one can clearly observe that when the amplitude of the incoming electric field is sufficiently small (bottom left in Figure 3), the polarization (top right) is proportional to the electric field (bottom left), thus implying that the electric dipole radiates a wave having the same frequency as that of the incoming light (bottom right). On the other hand, Figure 4

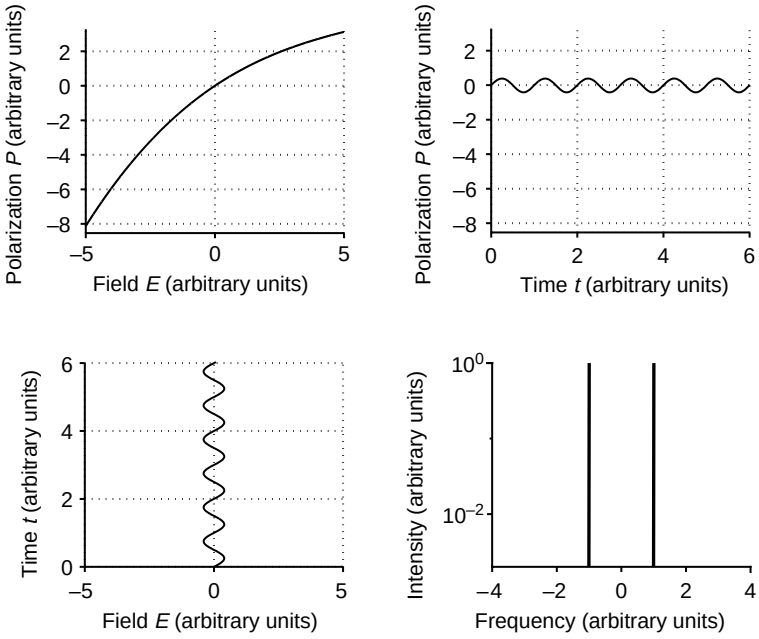

Figure 3 Polarization induced by an electric field of small amplitude. Nonlinear dependence of the polarization as a function of field amplitude (top left) and time dependence of the input electric field (bottom left). Time dependence of the induced polarization (top right) and corresponding intensity spectrum (bottom right).

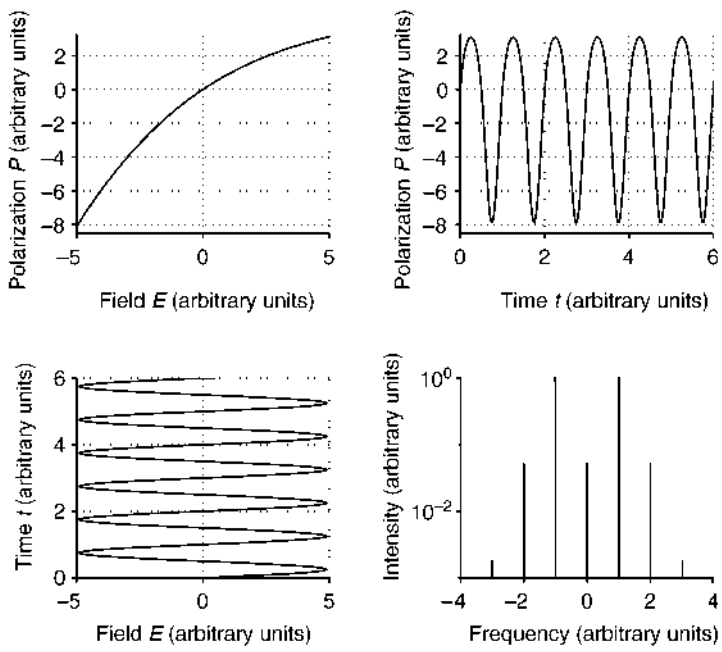

Figure 4 Polarization induced by an incoming electric field of large amplitude.

shows that for an electric field of large amplitude, the polarization is no longer proportional to the electric field, leading to radiation at harmonic frequencies (see Figure 4, bottom right).

This nonlinear behavior leads to a host of important phenomena in optical fibers, which are useful for many optical systems but detrimental for other systems.

Parametric Phenomena in Optical Fibers

In anisotropic materials, the leading nonlinear term in the polarization, i.e., the $\chi^{(2)}$ term, leads to phenomena such as the harmonic generation or optical rectification. This $\chi^{(2)}$ term vanishes in homogeneous isotropic materials such as cylindrical

optical fibers, and there the leading nonlinear term becomes the $\chi^{(3)}$ term. Thus, most of outstanding nonlinear phenomena in optical fibers originate from the third-order nonlinear susceptibility $\chi^{(3)}$. Some of those phenomena are described below.

Optical Kerr Effect

The optical Kerr effect is probably the most important nonlinear effect in optical fibers. This effect induces an intensity dependence of the refractive index, which leads to a vast wealth of fascinating phenomena such as self-phase modulation (SPM), cross-phase modulation (CPM), four-wave mixing (FWM), modulational instability (MI) or optical solitons. The Kerr effect can be conveniently described in the frequency domain through a direct analysis of the polarization, which takes the following form:

$$P_{\mathrm{NL}}(\omega) = \frac{3}{4}\varepsilon_0\chi^{(3)}(\omega)|E(\omega)|^2E(\omega) \qquad [7]$$

The constant 3/4 comes from the symmetry properties of the tensor $\chi^{(3)}$. Setting $P_{\mathrm{NL}}(\omega) = \varepsilon_0\varepsilon_{\mathrm{NL}}E(\omega)$, where $\varepsilon_{\mathrm{NL}} = \frac{3}{4}\chi^{(3)}|E|^2$ is the nonlinear contribution to the dielectric constant, the total polarization takes the form

$$P(\omega) = P_{\mathrm{L}} + P_{\mathrm{NL}} = \varepsilon_0[\chi^{(1)}(\omega) + \varepsilon_{\mathrm{NL}}]E(\omega) \qquad [8]$$

As eqn [8] shows, the refractive index n, at a given frequency ω, is given by

$$n^2 = 1 + \chi^{(1)} + \varepsilon_{\mathrm{NL}} = (n_0 + \Delta n_{\mathrm{NL}})^2 \qquad [9]$$

with $n_0^2 = 1 + \chi^{(1)}$. In practice $\Delta n_{\mathrm{NL}} \ll n_0$, and then, the refractive index is given by

$$n(\omega, |E|^2) = n_0(\omega) + n_2^e|E|^2 \qquad [10]$$

where n_2^e is the nonlinear refractive index defined by $n_2^e = 3\chi^{(3)}/(8n_0)$. The linear polarization P_{L} is responsible for the frequency dependence of the refractive index, whereas the nonlinear polarization P_{NL} causes an intensity dependence of the refractive index, which is referred to as the optical Kerr effect. Knowing that the wave intensity I is given by $I = \alpha|E|^2$, with $\alpha = \frac{1}{2}\varepsilon_0 c n_0$, the refractive index can be then rewritten as

$$n(\omega, I) = n_0(\omega) + n_2 I \qquad [11]$$

with $n_2 = n_2^e/\alpha = 2n_2^e/(\varepsilon_0 c n_0)$. For fused silica fibers one has typically: $n_2 = 2.66 \times 10^{-20}\ \mathrm{m^2\ W^{-1}}$. For example, an intensity of $I = 1\ \mathrm{GW\ cm^{-2}}$ leads to $\Delta n_{\mathrm{NL}} = 2.66 \times 10^{-7}$, which is much smaller than $n_0 \approx 1.45$.

Four-Wave Mixing

The four-wave mixing (FWM) process is a third-order nonlinear effect in which four waves interact through an energy exchange process. Let us consider two intense waves, $E_1(\omega_1)$ and $E_2(\omega_2)$, with $\omega_2 > \omega_1$, called pump waves, propagating in an optical fiber. Hereafter we consider the simplest case when waves propagate with the same polarization. In this situation the total electric field is given by

$$\begin{aligned} E_{\mathrm{tot}}(\mathbf{r}, t) &= E_1 + E_2 \\ &= A_1(\omega_1)\exp[\mathrm{i}(\mathbf{k}_1\cdot\mathbf{r} - \omega_1 t)] \\ &\quad + A_2(\omega_2)\exp[\mathrm{i}(\mathbf{k}_2\cdot\mathbf{r} - \omega_2 t)] \end{aligned} \qquad [12]$$

where $\mathbf{k}_1$ and $\mathbf{k}_2$ are the wavevectors of the fields E_1 and E_2, respectively. Equation [7], which gives the nonlinear polarization induced by a single monochromatic wave, remains valid provided that the frequency spacing between the two waves is relatively small, i.e., $|\Delta\omega| = |\omega_2 - \omega_1| \ll \omega_0 = (\omega_1 + \omega_2)/2$. In this context, eqn [7] leads to

$$P_{\mathrm{NL}} \approx \frac{3}{4}\varepsilon_0\chi^{(3)}(\omega_0)|E_{\mathrm{tot}}|^2E_{\mathrm{tot}} \qquad [13]$$

Substituting eqn [12] in eqn [13] yields

$$\begin{aligned} P_{\mathrm{NL}} = 2n_0n_2\varepsilon_0\Big[&\left(|E_1|^2 + 2|E_2|^2\right)E_1 \\ &+ \left(|E_2|^2 + 2|E_1|^2\right)E_2 + E_1^2E_2^* + E_1^*E_2^2\Big] \end{aligned} \qquad [14]$$

The term $|E_1|^2E_1$ in the right-hand side of eqn [14] represents the self-induced Kerr effect on the wave ω_1. The term $|E_2|^2E_1$, which corresponds to a modification of the refractive index seen by the wave ω_1, due to the presence of the wave ω_2, represents the cross-Kerr effect. Thus, the refractive index at frequency ω_1 depends simultaneously on the intensities of the two pumps, i.e., $n = n(\omega_1, |E_1|^2, |E_2|^2)$. Similarly the third and fourth terms in the right-hand side of eqn [14] illustrate the self-induced and cross-Kerr effects on the wave ω_2, respectively. Note that the self-induced Kerr effect is responsible for a self-phase modulation, which induces a spectral broadening of a pulse propagating through the optical fiber, whereas the cross-Kerr effect leads to a cross-phase modulation that induces a spectral broadening of one wave in the presence of a second wave. The two last terms in the right-hand side of eqn [14] correspond to the generation of new waves at frequencies $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$, respectively. The wave with the smallest frequency $2\omega_1 - \omega_2 = \omega_{\mathrm{s}}$ is called the Stokes wave, whereas the wave with the highest frequency $2\omega_2 - \omega_1 = \omega_{\mathrm{as}}$ is called the anti-Stokes wave. These two newly generated waves interact

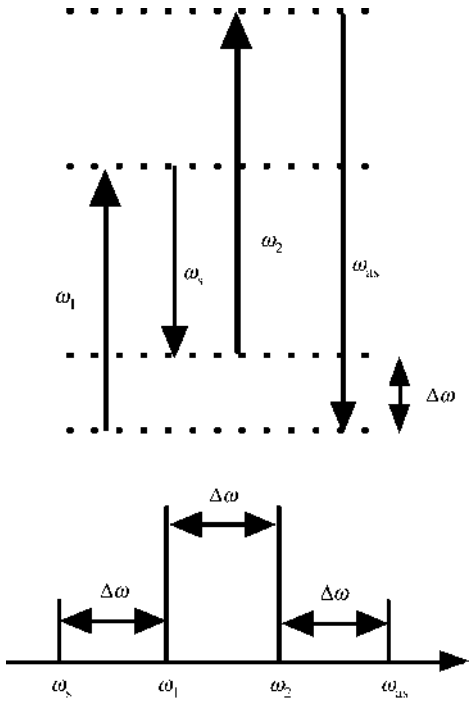

Figure 5 Schematic diagram representing a non-degenerate four-wave mixing process (top) and the corresponding frequency spectrum (bottom).

with the two pumps through an energy exchange process. This interaction is commonly referred to as a four-wave mixing process.

From a quantum mechanical point of view, four-wave mixing corresponds to a process in which two photons with frequencies ω_1 and ω_2 are annihilated with simultaneous creation of two photons at frequencies ω_s and ω_{as}, respectively. The new waves are generated at frequencies ω_s and ω_{as} such that

$$\omega_1 + \omega_2 = \omega_s + \omega_{as} \qquad [15]$$

which states that the total energy is conserved during the interaction. But the condition for this FWM process to occur is that the total momentum is conserved, that is,

$$\Delta\mathbf{k} = \mathbf{k}_s + \mathbf{k}_{as} - \mathbf{k}_1 - \mathbf{k}_2 = 0 \qquad [16]$$

or equivalently,

$$n(\omega_s)\omega_s + n(\omega_{as})\omega_{as} - n(\omega_1)\omega_1 - n(\omega_2)\omega_2 = 0 \qquad [17]$$

Equation [16] is known as the phase-matching condition. Figure 5 displays a schematic diagram of a four-wave mixing process with the corresponding frequency spectrum.

FWM with different pump frequencies $\omega_1 \neq \omega_2$ is called 'nondegenerate FWM', whereas the case $\omega_1 = \omega_2$ is referred to as 'degenerate FWM', or more simply, as three-wave mixing.

Conclusion

Optical fibers constitute a key device for many areas of optical sciences, and in particular for ultrafast optical communications. The nonlinear phenomena that arise through the nonlinear refractive index change (induced mainly by the Kerr effect) have been largely investigated these last two decades for applications such as all-optical wavelength conversion, parametric amplification, generation of new optical frequencies or ultrahigh repetition rate pulse trains. However, in many other optical systems such as optical communication systems, these nonlinear phenomena become harmful effects, and in this case control processes are being developed in order to suppress or at least reduce their impact in the system.

List of Units and Nomenclature

Nonlinear refractive index n_2 [$m^2\ W^{-1}$]
Optical intensity I [$GW\ cm^{-2} = 10^9\ W\ cm^{-2}$]

CPM: cross phase modulation
FWM: four wave mixing
MI: modulational instability
SPM: self-phase modulation

See also

Fiber and Guided Wave Optics: Light Propagation; Measuring Fiber Characteristics. **Lasers:** Optical Fiber lasers. **Optical Amplifiers:** Semiconductor Optical Amplifiers. **Optical Communication Systems:** Wavelength Division Multiplexing. **Scattering:** Scattering Theory. **Solitons:** Soliton Communication Systems.

Further Reading

Agrawal GP (2001) *Nonlinear Fiber Optics*. San Diego, CA: Academic Press.
Bloembergen N (1977) *Nonlinear Optics*. Reading, MA: Benjamin.
Boyd RW (1992) *Nonlinear Optics*. San Diego, CA: Academic Press.
Butcher PN and Cotter DN (1990) *The Elements of Nonlinear Optics*. Cambridge, UK: Cambridge University Press.
Dianov EM, Mamyshev PV, Prokhorov AM and Serkin VN (1989) *Nonlinear Effects in Optical Fibres*. Switzerland: Harwood Academic Publishers.
Hellwarth R (1977) Third-order optical susceptibilities of liquids and s\olids. *Progress in Quantum Electronics* 5: 1–68.
Pocholle JP, Papuchon M, Raffy J and Desurvire E (1990) Non linearities and optical amplification in single mode fibres. *Revue Technique Thomson-CSF* 22: 187–268.
Shen YR (1984) *Principles of Nonlinear Optics*. New York: Wiley.

Nonlinear Optics

K Thyagarajan and A K Ghatak, Indian Institute of Technology, New Delhi, India

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The invention of the laser provided us with a light source capable of generating extremely large optical power densities (several MW/m^2). At such large power densities, matter behaves in a nonlinear fashion and we come across new optical phenomena such as second-harmonic generation (SHG), sum and difference frequency generation, intensity dependent refractive index, mixing of various frequencies, etc. In SHG, an incident light beam at frequency ω interacts with the medium and generates a new light wave at frequency 2ω. In sum and difference frequency generation, two incident beams at frequencies ω_1 and ω_2 mix with each other producing sum $(\omega_1 + \omega_2)$ and difference $(\omega_1 - \omega_2)$ frequencies at the output. Higher-order nonlinear effects, such as self-phase modulation, four-wave mixing, etc. can also be routinely observed today. The field of nonlinear optics dealing with such nonlinear interactions is gaining importance due to numerous demonstrated applications in many diverse areas such as optical fiber communications, all-optical signal processing, realization of novel sources of optical radiation, etc.

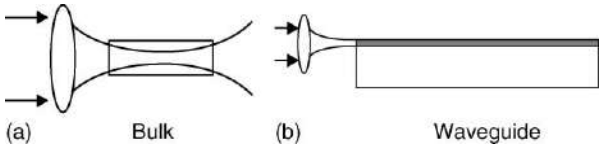

Figure 1 (a) In bulk media, tighter focusing produces larger intensities, but over shorter interaction lengths. (b) In optical waveguides diffraction effects are balanced by waveguiding and the interaction lengths are much larger.

Nonlinear optical interactions become prominent when the optical power densities are high and interaction takes place over long lengths. The usual method to increase optical intensity is to focus the light beam using a lens system. For a given optical power, the tighter the focusing, the larger will be the intensity for a given optical power; however, greater will be the divergence of the beam. Thus tighter focusing produces larger intensities, but over shorter interaction lengths (see Figure 1a). Optical waveguides, in which the light beam is confined to a small cross-sectional area, are currently being explored for realizing efficient nonlinear devices. In contrast to bulk media, in waveguides, diffraction effects are balanced by waveguiding and the beam can have small cross-sectional areas over much longer interaction lengths (see Figure 1b). A simple optical waveguide consists of a high index dielectric medium surrounded by a lower index dielectric medium so that light waves can be trapped in the high index region by the phenomenon of total internal reflection. Figure 2 shows a planar waveguide, a channel waveguide and an optical fiber. In the planar waveguide a film of refractive index n_f is deposited/diffused on a substrate of refractive index n_s and has a cover of refractive index n_c (with n_s, $n_c < n_f$). The waveguide has typical cross-sectional dimensions of a few micrometers. Unlike planar waveguides, in which guidance takes place only in one dimension, in channel waveguides the diffused region in a substrate is surrounded on all sides by lower index media. Optical fibers are structures with cylindrical symmetry and have a central cylindrical core of doped silica surrounded by a concentric cylindrical cladding of pure silica which has a slightly lower refractive index. Light guidance in all these waveguides takes place through the phenomenon of total internal reflection.

In contrast to bulk interactions requiring beam focusing, in the case of optical waveguides, the beam can be made to have extremely small cross-sectional

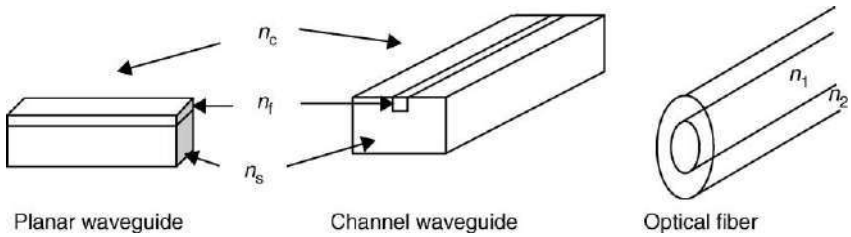

Figure 2 A planar waveguide, a channel waveguide and an optical fiber.

areas (~25 μm^2) over very long interaction lengths (~20 mm in integrated optical waveguides and tens of thousands of kilometers in the case of optical fibers). This leads to very much increased nonlinear interaction efficiencies even at moderate powers (approximately a few tens of mW).

In the following sections, we will discuss some of the important nonlinear interactions that are being studied with potential applications to various branches of science and engineering. Obviously it is impossible to cover all aspects of nonlinear optics – several books have been written in this area (see Further Reading section at the end of this article) – what we will do is to discuss the physics of some of the important nonlinear effects.

Apart from intensity and length of interaction, one of the most important requirements for many efficient nonlinear interactions is the requirement of phase matching. The concept of phase matching can be easily understood from the point of view of SHG. In SHG, the incident wave at frequency ω generates a nonlinear polarization at frequency 2ω and this nonlinear polarization is responsible for the generation of the wave at 2ω. Now, the phase velocity of the nonlinear polarization wave at 2ω is the same as the phase velocity of the electromagnetic wave at frequency ω, which is usually different from the phase velocity of the electromagnetic wave at frequency 2ω; this happens due to wavelength dispersion in the medium. When the two phase velocities are unequal, the polarization wave at 2ω (which is the source) and the electromagnetic wave at 2ω pass in and out of phase with each other as they propagate through the medium. Due to this, the energy flowing in from ω to 2ω cannot add constructively and the efficiency of second-harmonic generation is limited. If the phase velocities of the waves at ω and 2ω are matched then the polarization wave and the wave at 2ω remain in phase leading to drastically increased efficiencies. This condition is referred to as phase matching and plays a very important role in most nonlinear interactions.

Nonlinear Polarization

In a linear medium, the electric polarization P is assumed to be a linear function of the electric field E:

$$P = \varepsilon_0 \chi E \qquad [1]$$

where, for simplicity, a scalar relation has been written. The quantity χ is termed as linear dielectric susceptibility. At high optical intensities (which corresponds to high electric fields), all media behave in a nonlinear fashion. Thus eqn [1] is modified to

$$P = \varepsilon_0(\chi E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \ldots) \qquad [2]$$

where $\chi^{(2)}, \chi^{(3)}, \ldots$ are higher order susceptibilities giving rise to the nonlinear terms. The second term on the right-hand side is responsible for SHG, sum and difference frequency generation, parametric interactions, etc. while the third term is responsible for third-harmonic generation, intensity dependent refractive index, self-phase modulation, four-wave mixing, etc. For media possessing an inversion symmetry, $\chi^{(2)}$ is zero and there is no second-order nonlinear effect. Thus silica optical fibers, which form the heart of today's communication networks, do not possess the second-order nonlinearity.

We will first discuss second-harmonic generation and parametric amplification which arises due to the second-order nonlinear term and then go on to effects due to third-order nonlinear interaction.

Second-Harmonic Generation in Crystals

The first demonstration of SHG was made in 1961 by focusing a 3 kW ruby laser pulse ($\lambda_0 = 6943$ Å) on a quartz crystal. An incident beam from a ruby laser (red color) after passing through a crystal of KDP, gets partly converted into blue light which is the second harmonic. Ever since, SHG has been one of the most widely studied nonlinear interactions.

We consider a plane wave of frequency ω propagating along the z-direction through a medium and consider the generation of the second-harmonic frequency 2ω as the beam propagates through the medium. Now, the field at ω generates a polarization at 2ω, which acts as a source for the generation of an electromagnetic wave at 2ω. Corresponding to the frequencies ω and 2ω, the electric fields are assumed to be given by

$$E^{(\omega)} = \frac{1}{2}(E_1(z)e^{i(\omega t - k_1 z)} + \text{c.c.}) \qquad [3]$$

and

$$E^{(2\omega)} = \frac{1}{2}(E_2(z)e^{i(2\omega t - k_2 z)} + \text{c.c.}) \qquad [4]$$

respectively; c.c. stands for the complex conjugate of the preceding quantities. The quantities:

$$k_1 = \omega\sqrt{(\varepsilon_1 \mu_0)} = (\omega/c)n^{\omega} \qquad [5]$$

and

$$k_2 = 2\omega\sqrt{(\varepsilon_2 \mu_0)} = (2\omega/c)n^{2\omega} \qquad [6]$$

represent the propagation vectors at ω and 2ω, respectively; ε_1 and ε_2 represent the dielectric permittivities at ω and 2ω, and n^{ω} and $n^{2\omega}$ represent

the corresponding wave refractive indices. It should be noted that the amplitudes E_1 and E_2 are assumed to be z dependent – this is because at $z = 0$ (where the beam is incident on the medium) the amplitude E_2 is zero and this would develop as the beam propagates through the medium. We will now develop an approximate theory for the generation of the second harmonic.

We start with the wave equation:

$$\nabla^2 E - \varepsilon\mu_0 \frac{\partial^2 E}{\partial t^2} = \mu_0 \frac{\partial^2 P_{\mathrm{NL}}}{\partial t^2} \qquad [7]$$

where P_{NL} is the nonlinear polarization. An incident wave at frequency ω generates a polarization at 2ω, which acts as a source for the generation of an electromagnetic wave at 2ω.

In order to consider SHG, we write the wave equation corresponding to 2ω with the nonlinear polarization at 2ω given by

$$P_{\mathrm{NL}}^{(2\omega)} = \frac{1}{2}(\tilde{P}_{\mathrm{NL}}^{(2\omega)}\, \mathrm{e}^{2\,\mathrm{i}(\omega t - k_1 z)} + \mathrm{c.c.}) \qquad [8]$$

where

$$\tilde{P}_{\mathrm{NL}}^{(2\omega)} = dE_1E_1 \qquad [9]$$

and

$$d = \frac{\varepsilon_0 \chi^{(2)}}{2} \qquad [10]$$

represents the effective nonlinear coefficient and depends on the nonlinear material, the polarization states of the fundamental and the second harmonic and also on the propagation direction. Simple manipulations give us, for the rate of change of amplitude of the second harmonic wave:

$$\frac{\mathrm{d}E_2}{\mathrm{d}z} = -\frac{i\mu_0\, dc\omega}{n^{2\omega}} E_1^2(z)\mathrm{e}^{\mathrm{i}(\Delta k)z} \qquad [11]$$

where

$$\Delta k = k_2 - 2k_1 = (2\omega/c)(n^{2\omega} - n^{\omega}) \qquad [12]$$

and we have assumed:

$$\mathrm{d}^2E_2/\mathrm{d}z^2 \ll k_2(\mathrm{d}E_2/\mathrm{d}z) \qquad [13]$$

In order to solve eqn [11] we neglect depletion of the fundamental field, i.e., $E_1(z)$ is almost a constant and the quantity E_1^2 on the right-hand side can be assumed to be independent of z. If we now integrate eqn [11], we obtain:

$$E_2(z) = -\frac{i\mu_0\, \mathrm{d}c\omega}{n^{2\omega}} z E_1^2 \mathrm{e}^{\mathrm{i}\sigma} \frac{\sin\sigma}{\sigma} \qquad [14]$$

where

$$\sigma = \frac{1}{2}(\Delta k)z = (\omega/c)(n^{2\omega} - n^{\omega})z \qquad [15]$$

Now, the powers associated with the beams corresponding to ω and 2ω are given by:

$$P_1 = \frac{n^{\omega}}{2c\mu_0} S|E_1|^2, \qquad P_2 = \frac{n^{2\omega}}{2c\mu_0} S|E_2|^2 \qquad [16]$$

where S represents the cross-sectional area of the beams. Substituting for $|E_2|^2$ from eqn [16], we get after some elementary simplifications:

$$\eta = \frac{P_2}{P_1} = \frac{2c^3\mu_0^3 d^2\omega^2}{(n^{\omega})^2 n^{2\omega}} z^2 \frac{P_1}{S}\left(\frac{\sin\sigma}{\sigma}\right)^2 \qquad [17]$$

when η represents the SHG efficiency. Note that the efficiency of SHG increases if P_1/S, the intensity of the fundamental wave increases. Also η increases if the nonlinear coefficient d increases (obviously) and if the frequency increases. However, for a given power P_1, the conversion efficiency increases if the area of the beam decreases – thus a focused beam will have a greater SHG efficiency. The most important factor is $(\sin\sigma/\sigma)^2$ which is a sharply peaked function around $\sigma = 0$, attaining a maximum value of unity at $\sigma = 0$. Thus for maximum SHG efficiency:

$$\sigma = 0 \Rightarrow n^{2\omega} = n^{\omega} \qquad [18]$$

i.e., the refractive index at 2ω must be equal to the refractive index at ω – this is known as the phase matching condition.

The phase matching condition can be pictorially represented by a vector diagram (see Figure 3). In Figure 3, $\mathbf{k}_1\omega$ and $\mathbf{k}_2\omega$ represent the wave vectors of the fundamental and the second harmonic respectively. To achieve reasonably effective SHG, it is very important to satisfy the phase-matching condition. Efficiencies under nonphase matched operation can be orders of magnitude lower than under phase-matched operation.

We see from eqn [17] that the smallest z for which η is maximum is:

$$z = L_{\mathrm{c}} = \frac{\pi}{\Delta k} = \frac{\pi c}{2\omega(n^{2\omega} - n^{\omega})} \qquad [19]$$

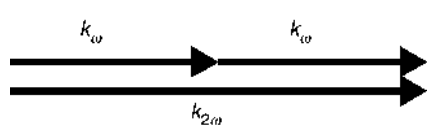

Figure 3 Vector diagram representing the phase matching condition for SHG.

where we have used eqn [15] for Δk. The length L_c is called the phase coherence length and represents the maximum crystal length up to which the second-harmonic power increases. Thus, if the length of the crystal is less than L_c, the second-harmonic power increases almost quadratically with z. For $z > L_c$, the second-harmonic power begins to reduce again.

In general, because of dispersion, it is very difficult to satisfy the phase-matching condition. However, in a birefringent medium, for example with $n_o > n_e$, it may be possible to find a direction along which the refractive index of the *o*-wave for ω equals the refractive index of the *e*-wave for 2ω (see Figure 4). For media with $n_e > n_o$, the direction would correspond to that along which the refractive index of the *e*-wave for ω equals the refractive index for the *o*-wave for 2ω. This can readily be understood by considering a specific example. We consider the SHG in KDP corresponding to the incident ruby laser wavelength ($\lambda_0 = 0.6943\ \mu\text{m}$, $\omega = 2.7150 \times 10^{15}$ Hz). For this frequency:

$$\left\{ \begin{array}{ll} n_o^{\omega} = 1.50502, & n_e^{\omega} = 1.46532 \\ n_o^{2\omega} = 1.53269, & n_e^{2\omega} = 1.48711 \end{array} \right\} \qquad [20]$$

The refractive index variation for the *e*-wave is given by:

$$n_e(\theta) = \left(\frac{\sin^2\theta}{n_e^2} + \frac{\cos^2\theta}{n_o^2} \right)^{-\frac{1}{2}} \qquad [21]$$

where θ is the angle that the wave propagation direction makes with the optic axis. Now, as can be seen from the above equations:

$$n_e < n_e(\theta) < n_o \qquad [22]$$

Since n_o at 0.6943 μm lies between n_e and n_o at 0.34715 μm, there will always exist an angle θ_m along which $n_e^{2\omega}(\theta_m) = n_o^{\omega}$. We can solve for θ_m and obtain:

$$\cos\theta_m = \left[\frac{(n_o^{\omega})^2 - (n_e^{2\omega})^2}{(n_o^{2\omega})^2 - (n_e^{2\omega})^2} \right]^{1/2} \frac{n_o^{2\omega}}{n_o^{\omega}} \qquad [23]$$

For the values given by eqn [20], we find $\theta_m = 50.5°$.

Quasi Phase Matching (QPM)

As mentioned earlier, phase matching is extremely important for any efficient nonlinear interaction. Recently the technique of quasi phase matching (QPM) has become a convenient way to achieve phase matching at any desired wavelength in any material. QPM relies on the fact that the phase mismatch in the two interacting beams can be compensated by periodically readjusting the phase of interaction through periodically modulating the nonlinear characteristics of the medium at a spatial frequency equal to the wavevector mismatch of the two interacting waves. Thus in SHG, when the nonlinear polarization at 2ω and the electromagnetic wave at 2ω have an accumulated phase difference of π, then the sign of the nonlinear coefficient is reversed so that the energy flowing from the polarization to the wave can add constructively with the existing energy (see Figure 5). Thus, by properly choosing the period of the spatial modulation of the nonlinear coefficient, one could achieve phase matching. This scheme, QPM, is being very widely studied for application to nonlinear interactions in bulk and in waveguides.

In a ferroelectric material such as lithium niobate, the signs of the nonlinear coefficients are linked to the direction of the spontaneous polarization. Thus a

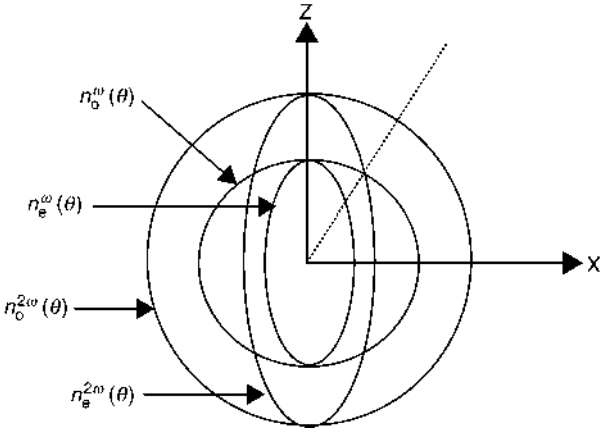

Figure 4 In a birefringent medium, for example with $n_o > n_e$, it may be possible to find a direction along which the refractive index of the *o*-wave for ω equals the refractive index of the *e*-wave for 2ω.

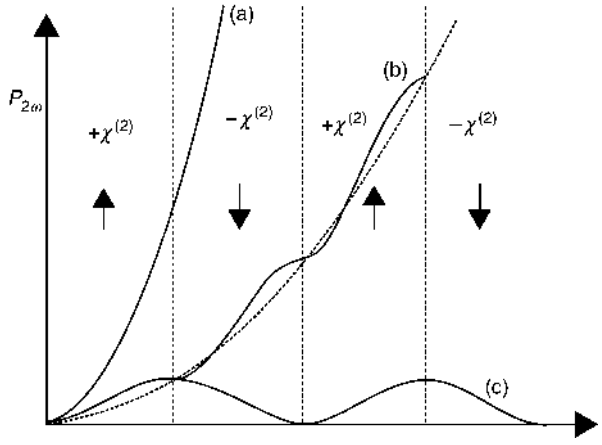

Figure 5 By reversing the sign of the nonlinear coefficient after every coherence length, the energy in the second harmonic can be made to grow. (a) Perfect phase matching; (b) Quasi phase matching; (c) Nonphase matched.

periodic reversal of the domains of the crystal can be used for achieving QPM (see Figure 6). This is the currently used technique to obtain high-efficiency SHG and other nonlinear interactions in $LiNbO_3$, $LiTaO_3$, and KTP. The most popular technique today, to achieve periodic domain reversal in $LiNbO_3$, is the technique of electric field poling. In this method, a high electric field pulse is applied to properly oriented lithium niobate crystal using lithographically defined electrode patterns to produce a permanent periodic domain reversed pattern. Such a periodically domain reversed $LiNbO_3$ crystal with the periodically reversed domains going through the entire depth of the crystal is also referred to as PPLN (periodically poled lithium niobate, pronounced 'piplin') and is now commercially available.

In order to analyze SHG in a periodically poled material, let us assume that the nonlinear coefficient d varies sinusoidally with a period Λ. In such a case we have:

$$d = d_0 \sin(Kz) \qquad [24]$$

where d_0 is the amplitude of modulation of the nonlinear coefficient and $K(= 2\pi/\Lambda)$ represents the spatial frequency of the periodic modulation. For easier understanding we are assuming the modulation to be sinusoidal; in general, the modulation will be periodic but not sinusoidal. Any periodic modulation can be written as a superposition of sinusoidal and cosinusoidal variations. Thus our discussion is valid for one of the Fourier components of the variation.

By using eqn [24], eqn [11] for the variation of the amplitude of the second harmonic becomes:

$$\frac{dE_2}{dz} = -\frac{i\mu_0 d_0 c\omega}{n^{2\omega}} E_1^2(z) e^{i(\Delta k)z} \sin(Kz) \qquad [25]$$

which can be written as:

$$\frac{dE_2}{dz} = -\frac{\mu_0 d_0 c\omega}{2n^{2\omega}} E_1^2(z) e^{i(\Delta k)z} [e^{iKz} - e^{-iKz}] \qquad [26]$$

Using similar arguments as earlier, it can be shown that if $\Delta k - K = 0$, then only the second term within the brackets in eqn [26] contributes to the growth of the second harmonic and similarly if $\Delta k + K = 0$, then only the first term within the brackets contributes to the growth of the second harmonic. The first condition implies that:

$$2\frac{2\pi}{\lambda_0}(n^{2\omega} - n^{\omega}) - \frac{2\pi}{\Lambda} = 0 \qquad [27]$$

where $\Lambda = 2\pi/K$ represents the spatial period of the modulation of the nonlinear coefficient, and λ_0 is the wavelength of the fundamental. Thus the modulation period Λ required for QPM SHG is:

$$\Lambda = \frac{\lambda_0}{2(n^{2\omega} - n^{\omega})} = 2\,L_c \qquad [28]$$

Figure 7 shows the vector diagram corresponding to QPM SHG. The phase mismatch between the fundamental and second harmonic is compensated by the spatial frequency vector of the periodic variation of d.

In the case of waveguides, n^{ω} and $n^{2\omega}$ would represent the effective indices of the modes at the fundamental and the second-harmonic frequency. It may be noted that the index difference between the fundamental and the second harmonic is typically 0.1 and thus the required spatial periods for a fundamental wavelength of 800 nm is $\sim 4\ \mu$m.

The main advantage of QPM is that it can be used at any wavelength within the transparency window of the material; only the period needs to be correctly chosen for a specific fundamental wavelength. One can also choose appropriate polarization to make use of the largest nonlinear optical coefficients. Another advantage is the possibility of varying the domain reversal period (chirp) to achieve specific interaction characteristics such as increased bandwidth, etc.

In general, the spatial variation of the nonlinear grating is not sinusoidal. In this case the efficiency of interaction would be determined by the Fourier component of the spatial variation at the spatial frequency corresponding to the period given by eqn [28]. Also, since the required periods are very small, it is possible to use a higher spatial period of modulation and use one of the Fourier components for the nonlinear interaction process. Thus, in the case of periodic reversal of the nonlinear coefficient

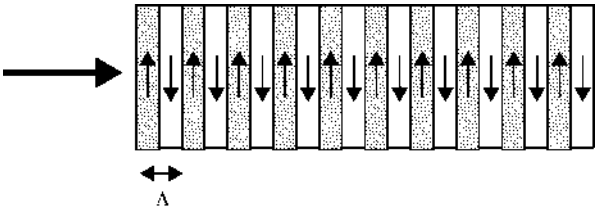

Figure 6 QPM can be achieved by a periodic reversal of the domains of a ferroelectric material. Arrows represent the direction of the spontaneous polarization of the crystal.

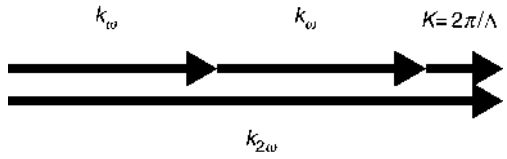

Figure 7 Vector diagram corresponding to QPM-SHG.

with a spatial period given by:

$$\Lambda_g = m\frac{\lambda_0}{2(n^{2\omega} - n^{\omega})}; \qquad m = 1, 3, 5, \ldots \qquad [29]$$

which happens to be the mth harmonic of the fundamental spatial frequency required for QPM, the corresponding nonlinear coefficient that would be responsible for SHG would be the Fourier amplitude at that spatial frequency. This can be taken into account by defining an effective nonlinear coefficient:

$$d_{QPM} = \frac{2d_0}{m\pi} \qquad [30]$$

Of course the largest effective nonlinear coefficient is achieved by using the fundamental frequency with $m = 1$. Higher spatial periods are easier to fabricate but would lead to reduced nonlinear efficiencies.

As compared to bulk devices, in the case of waveguides, the interacting waves are propagating in the form of modes having specific intensity distributions in the transverse plane. Because of this, the nonlinear interaction also depends on the overlap between the periodically inverted nonlinear medium, the fields of the fundamental and second-harmonic waves. Thus if $E_\omega(x, y)$ and $E_{2\omega}(x, y)$ represent the electric field distributions of the waveguide modes at the fundamental and second-harmonic frequency, then the efficiency of SHG depends on the following overlap integral:

$$I_{ovl} = \int\int d_0(x, y)E_\omega^2(x, y)E_{2\omega}(x, y)\,dx\,dy \qquad [31]$$

where $d_0(x, y)$ represents the transverse variation of the nonlinear coefficient of the waveguide. Thus optimization of SHG efficiency in the case of waveguide interactions has to take account of the overlap integral.

Since QPM relies on periodic phase matching, it is highly wavelength dependent. Thus the required period Λ is different for different fundamental frequencies. Any deviation in the frequency of the fundamental would lead to a reduction in the efficiency due to deviation from QPM condition. Thus the pump laser needs to be highly stabilized at a frequency corresponding to the fabricated period.

Apart from SHG, QPM has been used for other three-wave interaction processes such as difference frequency generation, parametric amplification, etc. Among the many materials that have been studied for SHG using QPM, the important ones are lithium niobate, lithium tantalite, and potassium titanyl phosphate (KTP). Many techniques have been developed for periodic domain reversal to achieve a periodic variation in the nonlinear coefficient. This includes electric field poling, electron bombardment, thermal diffusion treatment, etc.

Third-Order Nonlinear Effects

In the earlier sections, we discussed effects arising out of second-order nonlinearity, i.e., the term proportional to E^2 in the nonlinear polarization. This nonlinear term is found only in media not possessing an inversion symmetry. Thus amorphous materials or crystals possessing inversion symmetry do not exhibit second-order effects. The lowest-order nonlinear effects in such a medium is of an order three wherein the nonlinear polarization is proportional to E^3.

Self-phase modulation (SPM), cross-phase modulation (XPM) and four-wave mixing (FWM) represent some of the very important consequences of third-order nonlinearity. These effects have become all the more important as they play a significant and important role in wavelength division multiplexed optical fiber communication systems.

Self-Phase Modulation (SPM)

Consider the propagation of a plane light wave at frequency ω through a medium having $\chi^{(3)}$ nonlinearity. The polarization generated in the medium is given by

$$P = \varepsilon_0\chi E + \varepsilon_0\chi^{(3)}E^3 \qquad [32]$$

If we consider a single frequency wave with an electric field given by

$$E = E_0\cos(\omega t - kz) \qquad [33]$$

then

$$P = \varepsilon_0\chi E_0\cos(\omega t - kz) + \varepsilon_0\chi^{(3)}E_0^3\cos^3(\omega t - kz) \qquad [34]$$

Expanding $\cos^3\theta$ in terms of $\cos\theta$ and $\cos 3\theta$, we obtain the following expression for the polarization at frequency ω:

$$P = \varepsilon_0\left(\chi + \frac{3}{4}\chi^{(3)}E_0^2\right)E_0\cos(\omega t - kz) \qquad [35]$$

For a plane wave given by eqn [33], the intensity is

$$I = \frac{1}{2}c\varepsilon_0 n_0 E_0^2 \qquad [36]$$

where n_0 is the refractive index of the medium at low intensity. Then

$$P = \varepsilon_0\left(\chi + \frac{3}{2}\frac{\chi^{(3)}}{c\varepsilon_0 n_0}I\right)E \qquad [37]$$

The polarization P and electric field are related through the following equation:

$$P = \varepsilon_0(n^2 - 1)E \qquad [38]$$

where n is the refractive index of the medium. Comparing eqns [37] and [38], we get

$$n^2 = n_0^2 + \frac{3}{2}\frac{\chi^{(3)}}{c\varepsilon_0 n_0} I \qquad [39]$$

where

$$n_0^2 = 1 + \chi \qquad [40]$$

Since the last term in the eqn [39] is usually very small, we get

$$n = n_0 + n_2 I \qquad [41]$$

where

$$n_2 = \frac{3}{4}\frac{\chi^{(3)}}{c\varepsilon_0 n_0^2} \qquad [42]$$

is the nonlinear coefficient.

For fused silica $n_0 \approx 1.47$, $n_2 \approx 3.2 \times 10^{-20}$ m^2/W and if we consider power of 100 mW having a cross-sectional area of 100 μm^2, the resultant intensity is 10^9 W/m^2 and the corresponding change in refractive index is

$$\Delta n \approx n_2 I \approx 3.2 \times 10^{-11} \qquad [43]$$

Although this is very small, when the beam propagates over an optical fiber over long distances (a few hundred to a few thousand kilometers), the accumulated nonlinear effects can be significant.

In the case of an optical fiber, the light beam propagates as a mode having a specific transverse electric field distribution and thus the intensity is not constant across the cross-section. In such a case, it is convenient to express the nonlinear effect in terms of the power carried by the mode (rather than in terms of intensity). If the linear propagation constant of the mode is represented by β, then in the presence of nonlinearity, the effective propagation constant is given by

$$\beta_{\mathrm{NL}} = \beta + \frac{k_0 n_2}{A_{\mathrm{eff}}} P \qquad [44]$$

where $k_0 = 2\pi/\lambda_0$, P is the power carried by the mode. The quantity A_{eff} represents the effective transverse cross-sectional area of the mode and is defined by

$$A_{\mathrm{eff}} = \frac{\left[\int_0^\infty \int_0^{2\pi} \psi^2(r) r \,\mathrm{d}r\,\mathrm{d}\phi\right]^2}{\int_0^\infty \int_0^{2\pi} \psi^4(r) r \,\mathrm{d}r\,\mathrm{d}\phi} \qquad [45]$$

where $\psi(r)$ represents the transverse mode field distribution of the mode. For example, under the Gaussian approximation:

$$\psi(r) = \psi_0\, \mathrm{e}^{-r^2/w_0^2} \qquad [46]$$

where ψ_0 is a constant and $2w_0$ represents the mode field diameter (MFD), we get:

$$A_{\mathrm{eff}} = \pi w_0^2 \qquad [47]$$

It is usual to express the nonlinear characteristic of an optical fiber by the coefficient given by

$$\gamma = \frac{k_0 n_2}{A_{\mathrm{eff}}} \qquad [48]$$

Thus, for the same input power and same wavelength, smaller values of A_{eff} lead to greater nonlinear effects in the fiber. Typically:

$$A_{\mathrm{eff}} \sim 50\text{–}80\ \mu\mathrm{m}^2 \quad \text{and} \quad \gamma \approx 2.4\ \mathrm{W}^{-1}\,\mathrm{km}^{-1} \text{ to } 1.5\ \mathrm{W}^{-1}\,\mathrm{km}^{-1} \qquad [49]$$

When a light beam propagates through an optical fiber, the power decreases because of attenuation. Thus, the corresponding nonlinear effects also reduce. Indeed, the phase change suffered by a beam in propagating from 0 to L is given by

$$\phi = \int_0^L \beta_{\mathrm{NL}}\,\mathrm{d}z = \beta L + \gamma \int_0^L P\,\mathrm{d}z \qquad [50]$$

If α represents the attenuation coefficient, then

$$P(z) = P_0\,\mathrm{e}^{-\alpha z} \qquad [51]$$

and we get

$$\phi = \beta L + \gamma P_0 L_{\mathrm{eff}} \qquad [52]$$

where

$$L_{\mathrm{eff}} = \frac{1 - \mathrm{e}^{-\alpha L}}{\alpha} \qquad [53]$$

is referred to as the effective length. For $\alpha L \gg 1$, $L_{\mathrm{eff}} \approx 1/\alpha$ and for $\alpha L \ll 1$, $L_{\mathrm{eff}} \approx L$.

The effective length represents the length of the fiber over which most of the nonlinear effects has accumulated. For a loss coefficient of 0.20 dB/km, $L_{\mathrm{eff}} \approx 21$ km.

If we consider a fiber length much longer than L_{eff}, then to have reduced impact of SPM, we must have

$$\gamma P_0 L_{\mathrm{eff}} \ll 1 \qquad [54]$$

or

$$P_0 \ll \frac{1}{\gamma L_{\mathrm{eff}}} \approx \frac{\alpha}{\gamma} \qquad [55]$$

For $\alpha = 4.6 \times 10^{-2}$ km^{-1} (which corresponds to an attenuation of 0.2 dB/km) and $\gamma = 2.4$ W^{-1} km^{-1}, we get

$$P_0 \ll 19\ \mathrm{mW} \qquad [56]$$

Propagation of a Pulse

When an optical pulse propagates through a medium, it suffers from the following effects:

(i) attenuation;
(ii) dispersion; and
(iii) nonlinearity.

Attenuation refers to the reduction in the pulse energy due to various mechanisms, such as scattering, absorption, etc. Dispersion is caused by the fact that a light pulse consists of various frequency components and each frequency component travels at a different group velocity. Dispersion causes the temporal width of the pulse to change; in most cases it results in an increase in pulse width, however, in some cases the temporal width could also decrease. Dispersion is accompanied by chirping, the variation of the instantaneous frequency of the pulse within the pulse duration. Since both attenuation and dispersion cause a change in the temporal variation of the optical power, they closely interact with nonlinearity in deciding the pulse evolution as it propagates through the medium.

Let $E(x, y, z, t)$ represent the electric field variation of an optical pulse. It is usual to express E in the following way:

$$E(x, y, z, t) = \frac{1}{2}[A(z, t)\psi(x, y)\mathrm{e}^{\mathrm{i}(\omega_0 t - \beta_0 z)} + \text{c.c.}] \quad [57]$$

where $A(z, t)$ represents the slowly varying complex envelope of the pulse, $\psi(x, y)$ represents the transverse electric field distribution, ω_0 represents center frequency, and β_0 represents the propagation constant at ω_0.

In the presence of attenuation, second-order dispersion and third-order nonlinearity, the complex envelope $A(z, t)$ can be shown to satisfy the following equation:

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A - \beta_1\frac{\partial A}{\partial t} + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - i\gamma|A|^2 A \quad [58]$$

Here

$$\beta_1 = \left|\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\right|_{\omega=\omega_0} = \frac{1}{v_g} \quad [59]$$

represents the inverse of the group velocity of the pulse, and

$$\beta_2 = \left|\frac{\mathrm{d}^2\beta}{\mathrm{d}\omega^2}\right|_{\omega=\omega_0} = -\frac{\lambda_0^2}{2\pi c}D \quad [60]$$

where D represents the group velocity dispersion (measured in ps/km nm).

The various terms on the RHS of the eqn [58] represent the following:

I term:	attenuation
II term:	group velocity term
III term:	second-order dispersion
IV term:	nonlinear term

If we change to a moving frame defined by coordinates $T = t - \beta_1 z$, eqn [58] becomes

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial T^2} - i\gamma|A|^2 A \quad [61]$$

If we neglect the attenuation term, we obtain the following equation which is also referred to as the nonlinear Schrödinger equation:

$$\frac{\partial A}{\partial z} = i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial T^2} - i\gamma|A|^2 A \quad [62]$$

The above equation has a solution given by

$$A(z, t) = A_0 \operatorname{sech} \sigma T \, \mathrm{e}^{-\mathrm{i}gz} \quad [63]$$

with

$$A_0^2 = -\frac{\beta_2}{\gamma}\sigma^2, \qquad g = -\frac{\sigma^2}{2}\beta_2 \quad [64]$$

Equation [63] represents an envelope soliton and has the property that it propagates undispersed through the medium. The full width at half maximum (FWHM) of the pulse envelope will be given by $\tau_f = 2\tau_0$, where

$$\operatorname{sech}^2 \sigma\tau_0 = \frac{1}{2} \quad [65]$$

which gives the FWHM τ_f:

$$\tau_f = 2\tau_0 = \frac{2}{\sigma}\ln(1 + \sqrt{2}) \approx \frac{1.7627}{\sigma} \quad [66]$$

The peak power of the pulse is:

$$P_0 = |A_0|^2 = \frac{|\beta_2|}{\gamma}\sigma^2 \quad [67]$$

Replacing σ by τ_f, we obtain

$$P_0\tau_f^2 \approx \frac{\lambda_0^2}{2cy}D \quad [68]$$

where we have used eqn [60]. The above equation gives the required peak for a given τ_f for the formation of a soliton pulse.

As an example, we have $\tau_f = 10$ ps, $\gamma = 2.4\ \mathrm{W}^{-1}\ \mathrm{km}^{-1}$, $\lambda_0 = 1.55\ \mu\mathrm{m}$, $D = 2$ ps/km nm and the required peak power will be $P_0 = 33$ mW.

Soliton pulses are being extensively studied for application to long distance optical fiber

communication. In actual systems, the pulses have to be optically amplified at regular intervals to compensate for the loss suffered by the pulses. The amplification could be carried out using erbium doped fiber amplifiers (EDFAs) or fiber Raman amplifiers.

Spectral Broadening due to SPM

In the presence of only nonlinearity, eqn [61] becomes

$$\frac{\mathrm{d}A}{\mathrm{d}z} = -i\gamma|A|^2 A \qquad [69]$$

whose solution is given by

$$A(z,t) = A(z=0,t)\mathrm{e}^{-\mathrm{i}\gamma P z} \qquad [70]$$

where $P = |A|^2$ is the power in the pulse. If P is a function of time, then time dependent phase term at $z = L$ becomes

$$\mathrm{e}^{\mathrm{i}\phi(t)} = \mathrm{e}^{\mathrm{i}[\omega_0 t - \gamma P(t) L]} \qquad [71]$$

We can define an instantaneous frequency as

$$\omega(t) = \frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega_0 - \gamma L \frac{\mathrm{d}P}{\mathrm{d}t} \qquad [72]$$

for a Gaussian pulse:

$$P = P_0\, \mathrm{e}^{-2T^2/\tau_0^2} \qquad [73]$$

giving

$$\omega(t) = \omega_0 + \frac{4\gamma L T P_0\, \mathrm{e}^{-2T^2/\tau_0^2}}{\tau_0^2} \qquad [74]$$

Thus the instantaneous frequency within the pulse changes with time, leading to chirping of the pulse (see Figure 8). Note that since the pulsewidth has not changed, but the pulse is chirped, the frequency spectrum of the pulse has increased. Thus SPM leads to the generation of new frequencies. By Fourier transform theory, an increased spectral width implies that the pulse can now be compressed in the temporal domain by passing it through a medium with the proper sign of dispersion. This is indeed one of the standard techniques to realize ultrashort femtosecond optical pulses.

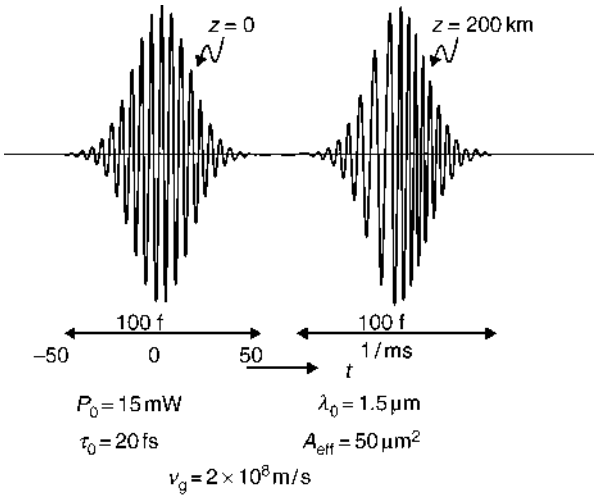

Figure 8 Due to self phase modulation, the instantaneous frequency within the pulse changes with time leading to chirping of the pulse.

Cross Phase Modulation (XPM)

Like SPM, cross-phase modulation also arises due to the intensity dependence of refractive index, leading to spectral broadening. Unlike SPM, in the case of XPM, intensity variations of a light beam at a particular frequency modulate the phase of light beam at another frequency. If the signals at both frequencies are pulses, then due to difference in group velocities of the pulses, there is a walk off between the two pulses, i.e., if they start together, they will separate as they propagate through the medium. Nonlinear interaction takes place as long as they physically overlap in the medium. Smaller the dispersion, smaller will be the difference in group velocities (assuming closely spaced wavelengths) and the longer they will overlap. This would lead to stronger XPM effects. At the same time, if two pulses pass through each other, then since one pulse will interact with both the leading and the trailing edge of the other pulse, XPM effects will be nil provided there is no attenuation. In the presence of attenuation in the medium, the pulse will still get modified due to XPM. A similar situation can occur when the two interacting pulses are passing through an optical amplifier.

To study XPM, we assume simultaneous propagation of two waves at two different frequencies through the medium. If ω_1 and ω_2 represent the two frequencies, then one obtains for the variation of the amplitude A_1 of the frequency ω_1:

$$\frac{\mathrm{d}A_1}{\mathrm{d}z} = -i\gamma(\tilde{P}_1 + 2\tilde{P}_2)A_1 \qquad [75]$$

where $\tilde{P}_1$ and $\tilde{P}_2$ represent the powers at frequencies ω_1 and ω_2, respectively. The first term in eqn [75] represents SPM while the second term corresponds to XPM. If the powers are assumed to attenuate at the same rate, i.e.:

$$\tilde{P}_1 = P_1\, \mathrm{e}^{-\alpha z}, \qquad \tilde{P}_2 = P_2\, \mathrm{e}^{-\alpha z} \qquad [76]$$

then the solution of eqn [75] is

$$A_1(L) = A_1(0)\, \mathrm{e}^{-\mathrm{i}\gamma(P_1+2P_2)L_{\mathrm{eff}}} \qquad [77]$$

where, as before, L_{eff} represents the effective length of the medium. When we are studying the effect of

power at ω_2 on the light beam at frequency ω_1, we will refer to the wave at frequency ω_2 as pump, and the wave at frequency ω_1 as probe or signal. From eqn [77] it is apparent that the phase of signal at frequency ω_1 is modified by the power at another frequency. This is referred to as XPM. Note also that XPM is twice as effective as SPM.

Similar to the case of SPM, we can now write for the instantaneous frequency in the presence of XPM as eqn [72]:

$$\omega(t) = \omega_0 - 2\gamma L_{\text{eff}} \frac{\mathrm{d}P_2}{\mathrm{d}t} \qquad [78]$$

Hence the part of the signal that is influenced by the leading edge of the pump will be down-shifted in frequency (since in the leading edge $\mathrm{d}P_2/\mathrm{d}t > 0$) and the part overlapping with the trailing edge will be up-shifted in frequency (since $\mathrm{d}P_2/\mathrm{d}t < 0$). This leads to a frequency chirping of the signal pulse just as in the case of SPM.

If the probe and pump beams are pulses, then XPM can lead to induced frequency shifts depending on whether the probe pulse interacts only with the leading edge or trailing edge or both, as they both propagate through the medium. Let us consider a case when the group velocity of pump pulse is greater than that of the probe pulse. Thus, if both pulses enter the medium together, then since the pump pulse travels faster, the probe pulse will interact only with the trailing edge of the pump. Since in this case $\mathrm{d}P_2/\mathrm{d}t$ is negative, the probe pulse suffers a blue-induced frequency shift. Similarly if the pulses enter at different instants but completely overlap at the end of the medium, then $\mathrm{d}P_2/\mathrm{d}t > 0$ and the probe pulse would suffer a red-induced frequency shift. Indeed, if the two pulses start separately and walk through each other, then there is no induced shift due to cancellation of shifts induced by leading and trailing edges of the pump. Figure 9 shows measured induced frequency shift of 532 nm probe pulse as a function of the delay between this pulse and a pump pulse at 1064 nm, when they propagate through a 1 m long optical fiber.

When pulses of light at two different wavelengths propagate through an optical fiber, due to different group velocities of the pulses, they pass through each other, resulting in what could be termed as a collision. In the linear case, the pulses will pass through without affecting each other, but when intensity levels are high, XPM induces phase shifts in both pulses. We can define a parameter termed walk off length L_{wo} which is the length of the fiber required for the interacting pulses to walk off relative to each other. The walk off length is given by

$$L_{\text{wo}} = \frac{\Delta\tau}{D\Delta\lambda} \qquad [79]$$

where D represents the dispersion coefficient, and $\Delta\lambda$ represents the wavelength separation between the interacting pulses. For return to zero (RZ) pulses, $\Delta\tau$ represents the pulse duration while for nonreturn to zero (NRZ) pulses, $\Delta\tau$ represents the rise term or fall time of the pulse. Closely spaced channels will thus interact over longer fiber lengths, thus leading to greater XPM effects. Larger dispersion coefficients will reduce L_{wo} and thus the effects of XPM. Since the medium is attenuating, the power carried by the pulses decreases as they propagate and thus leading to reduced XPM effect. The characteristic length for attenuation is the effective length L_{eff}, defined by eqn [53]. If $L_{\text{wo}} \ll L_{\text{eff}}$, then over the length of interaction of the pulses, the intensity levels do not change appreciably and the magnitude of the XPM-induced effects will be proportional to the wavelength spacing $\Delta\lambda$. For small $\Delta\lambda$'s, $L_{\text{wo}} \gg L_{\text{eff}}$ and the interaction length is now determined by the fiber losses (rather than by walk off) and the XPM-induced effects become almost independent of $\Delta\lambda$. Indeed, if we consider XPM effects between a continuous wave (cw) probe beam and a sinusoidally industry modulated pump beam, then the amplitude of the XPM-induced

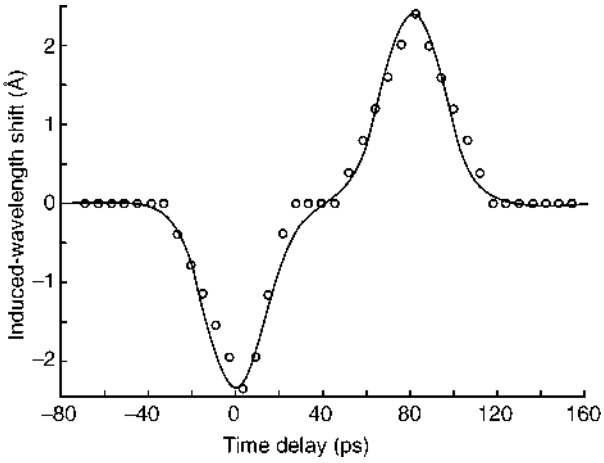

Figure 9 Measured induced frequency shift of 532 nm probe pulse as a function of the delay between this pulse and a pump pulse at 1064 nm, when they propagate through a 1 m long optical fiber. (Reproduced from Baldeck PL, Alfano RR and Agrawal GP (1998) Induced frequency shift of copropagating ultrafast optical pulses. *Applied Physics Letters* 52: 1939–1941 with permission from the American Institute of Physics.)

phase shift ($\Delta\Phi_{pr}$) in the probe beam is given by:

$$\Delta\Phi_{pr} \approx 2\gamma P_{2m} L_{eff} \quad \text{for} \quad L_{wo} \gg L_{eff}$$
$$\Delta\Phi_{pr} \approx 2\gamma P_{2m} L_{wo} \quad \text{for} \quad L_{wo} \ll L_{eff} \qquad [80]$$

Here P_{2m} is the amplitude of the sinusoidal power modulation of the pump beam.

XPM-induced intensity interference can be studied by simultaneously propagating an intensity modulated pump signal and a cw probe signal at a different wavelength. The intensity modulated signal will induce phase modulation on the cw probe signal and the dispersion of the medium will convert the phase modulation to intensity modulation of the probe. Thus, the magnitude of the intensity fluctuation of the probe signal serves as an estimate of the XPM induced interference. Figure 10 shows the intensity fluctuations on a probe signal at 1550 nm, induced by a modulated pump for a channel separation of 0.6 nm. Figure 11 shows the variation of the RMS value of probe intensity modulation with the wavelength separation between the intensity modulated signal and the probe. The experiment has been performed over four amplified spans of 80 km of standard single mode fiber (SMF) and nonzero dispersion shifted fiber (NZDSF). The large dispersion in SMF has been compensated using dispersion compensating chirped gratings. The probe modulation, in the case of SMF, decreases approximately linearly with $1/\Delta\lambda$ for all $\Delta\lambda$s; the modulation is independent of $\Delta\lambda$. In contrast the NZDSF shows a constant modulation for $\Delta\lambda \leq 1$ nm.

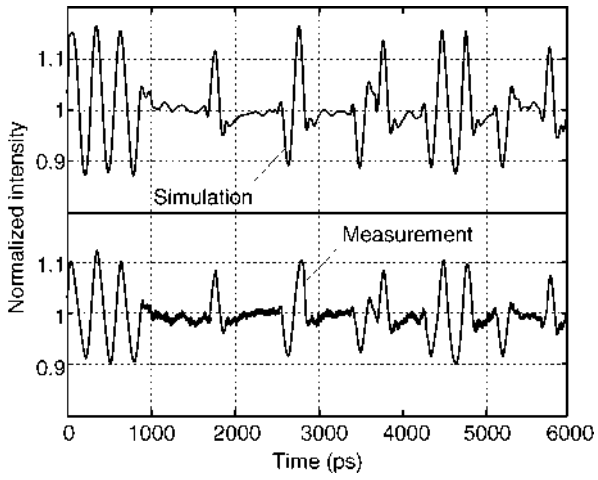

Figure 10 Intensity fluctuations induced by cross phase modulation on a probe signal at 1550 nm by a modulated pump for a channel separation of 0.6 nm. (Reproduced with permission from Rapp L (1997) Experimental investigation of signal distortions induced by cross phase modulation combined with dispersion. *IEEE Photon Technical Letters* 9: 1592–1595, © 2004 IEEE.)

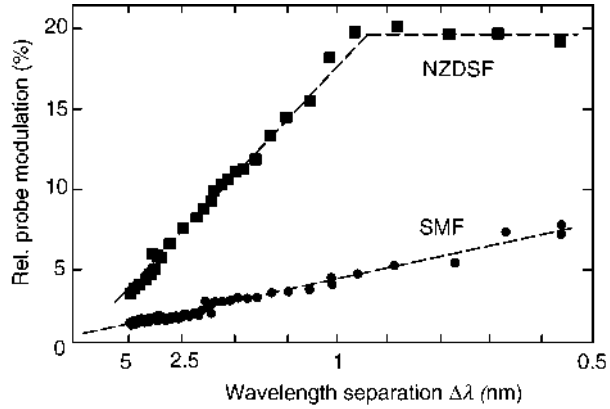

Figure 11 Variation of the RMS value of probe intensity modulation with the wavelength separation between the intensity modulated signal and the probe. (Reproduced with permission from Shtaif M, Eiselt M and Garret LD (2000) Cross-phase modulation distortion measurements in multispan WDM systems. *IEEE Photon Technical Letters* 12: 88–90, © 2004 IEEE.)

This is consistent with the earlier discussion in terms of L_{wo} and L_{eff}.

Four-Wave Mixing (FWM)

Four-wave mixing (FWM) is a nonlinear interaction that occurs in the presence of multiple wavelengths in a medium, leading to the generation of new frequencies. Thus, if light waves at three different frequencies ω_2, ω_3, ω_4 are launched simultaneously into a medium, the same nonlinear polarization that led to intensity dependent refractive index, leads to nonlinear polarization component at a frequency:

$$\omega_1 = \omega_3 + \omega_4 - \omega_2 \qquad [81]$$

This nonlinear polarization, under certain conditions, leads to the generation of electromagnetic waves at ω_1. This process is referred to as four-wave mixing due to the interaction between four different frequencies. Degenerate four-wave mixing corresponds to the case when two of the input waves have the same frequency. Thus, if $\omega_3 = \omega_4$, then inputting waves at ω_2 and ω_3 leads to the generation of waves at the frequency

$$\omega_1 = 2\omega_3 - \omega_2 \qquad [82]$$

During FWM process, there are four different frequencies present at any point in the medium. If we write the electric field of the waves as

$$E_i = \frac{1}{2}[A_i(z)\psi_i(x, y)e^{i(\omega_i t - \beta_i z)} + \text{c.c.}], \qquad [83]$$
$$i = 1, 2, 3, 4$$

where, as before, $A_i(z)$ represents the amplitude of the wave, $\psi_i(x, y)$ the transverse field distribution, and β_i the propagation constant of the wave. The total electric field is given by

$$E = E_1 + E_2 + E_3 + E_4 \qquad [84]$$

Substituting for the total electric field in the equation for nonlinear polarization, the term with frequency ω_1 comes out as

$$P_{\mathrm{NL}}^{(\omega_1)} = \frac{1}{2}[P_{\mathrm{NL}}^{(\omega_1)}\,\mathrm{e}^{\mathrm{i}(\omega_1 t - \beta_1 z)} + \mathrm{c.c.}] \qquad [85]$$

where

$$P_{\mathrm{NL}}^{(\omega_1)} = \frac{3\varepsilon_0}{2}\chi^{(3)}A_2^* A_3 A_4 \psi_2\psi_3\psi_4\,\mathrm{e}^{-\mathrm{i}\Delta\beta z} \qquad [86]$$

and

$$\Delta\beta = \beta_3 + \beta_4 - \beta_2 - \beta_1 \qquad [87]$$

In writing eqn [86], we have only considered the FWM term, neglecting the SPM and XPM terms.

Substituting the expression for $P_{\mathrm{NL}}^{(\omega_1)}$ in the wave equation for ω_1 and making the slowly varying approximation (in a manner similar to that employed in the case of SPM and XPM), we obtain the following equation for $A_1(z)$:

$$\frac{\mathrm{d}A_1}{\mathrm{d}z} = -2i\gamma A_2^* A_3 A_4\,\mathrm{e}^{-\mathrm{i}\Delta\beta z} \qquad [88]$$

where γ is defined by eqn [48] with $k_0 = \omega/c$, ω representing the average frequency of the four interacting waves, and A_{eff} is the average effective area of the modes.

Assuming all waves to have the same attenuation coefficient α and neglecting depletion of waves at frequencies ω_2, ω_3 and ω_4, due to nonlinear conversion we obtain for the power in the frequency ω_1 as

$$P_1(L) = 4\gamma^2 P_2 P_3 P_4 L_{\mathrm{eff}}^2 \eta\,\mathrm{e}^{-\alpha L} \qquad [89]$$

where

$$\eta = \frac{\alpha^2}{\alpha^2 + \Delta\beta^2}\left[1 + \frac{4\mathrm{e}^{-\alpha L}\sin^2\dfrac{\Delta\beta L}{2}}{(1 - \mathrm{e}^{-\alpha L})^2}\right] \qquad [90]$$

and L_{eff} is the effective length (see eqn [53]). Maximum four-wave mixing takes place when $\Delta\beta = 0$, since in such a case $\eta = 1$. Now:

$$\Delta\beta = \beta(\omega_3) + \beta(\omega_4) - \beta(\omega_2) - \beta(\omega_1) \qquad [91]$$

Since the frequencies are usually close to each other, we can make a Taylor series expansion about any frequency, say ω_2. In such a case, we obtain

$$\Delta\beta = (\omega_3 - \omega_2)(\omega_3 - \omega_1)\left|\frac{\mathrm{d}^2\beta}{\mathrm{d}\omega^2}\right|_{\omega=\omega_2} \qquad [92]$$

In optical fiber communication systems, the channels are usually equally spaced. Thus, we assume the frequencies to be given by

$$\omega_4 = \omega_2 + \Delta\omega, \qquad \omega_3 = \omega_2 - 2\Delta\omega \text{ and } \omega_1 = \omega_2 - \Delta\omega \qquad [93]$$

Using these frequencies and eqn [60], eqn [92] gives us:

$$\Delta\beta = -\frac{4\pi D\lambda^2}{c}(\Delta\nu)^2 \qquad [94]$$

where $\Delta\omega = 2\pi\Delta\nu$. Thus maximum FWM takes place when $D = 0$. This is the main problem in using wavelength division multiplexing (WDM) in dispersion shifted fibers which are characterized by zero dispersion at the operating wavelength of 1550 nm, as FWM will then lead to crosstalk among various channels. FWM efficiency can be reduced by using fiber with nonzero dispersion. This has led to the development of nonzero dispersion shifted fiber (NZDSF) which have a finite nonzero dispersion of about ± 2 ps/km nm at the operating wavelength.

From eqn [94], we notice that for a given dispersion coefficient D, FWM efficiency will reduce as $\Delta\nu$ increases.

In order to get a numerical appreciation, we consider a case with $D = 0$, i.e., $\Delta\beta = 0$. For such a case $\eta = 1$. If all channels were launched with equal power P_{in} then:

$$P_1(L) = 4\gamma^2 P_{\mathrm{in}}^3 L_{\mathrm{eff}}^2\,\mathrm{e}^{-\alpha L} \qquad [95]$$

Thus the ratio of power generated at ω_1, due to FWM and that existing at the same frequency, is

$$\frac{P_{\mathrm{g}}}{P_{\mathrm{out}}} = \frac{P_1(L)}{P_{\mathrm{in}}\,\mathrm{e}^{-\alpha L}} = 4\gamma^2 P_{\mathrm{in}}^2 L_{\mathrm{eff}}^2 \qquad [96]$$

Typical values are $L_{\mathrm{eff}} = 20$ km, $\gamma = 2.4\ \mathrm{W}^{-1}\ \mathrm{km}^{-1}$. Thus:

$$\frac{P_{\mathrm{g}}}{P_{\mathrm{out}}} \approx 0.01 P_{\mathrm{in}}^2(\mathrm{mW}^2) \qquad [97]$$

Figure 12 shows the output spectrum measured at the output of a 25 km-long dispersion shifted fiber ($D = -0.2$ ps/km nm) when three 3 mW wavelengths are launched simultaneously. Notice the generation of many new frequencies by FWM. Figure 13 shows the

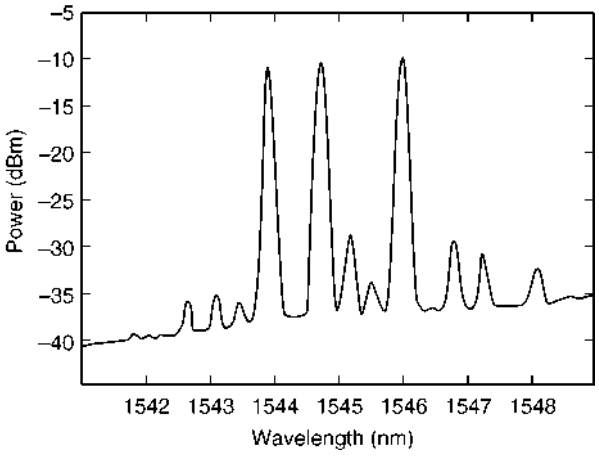

Figure 12 Generation of new frequencies because of FWM when waves at three frequencies are incident in the fiber. (Reproduced with permission from Tkach RW, Chraplyvy AR, Forghiari F, Gnanck AH and Derosier RM (1995) Four-photon mixing and high speed WDM systems. *Journal of lightwave Technology* 13: 841–849, © 2004 IEEE.)

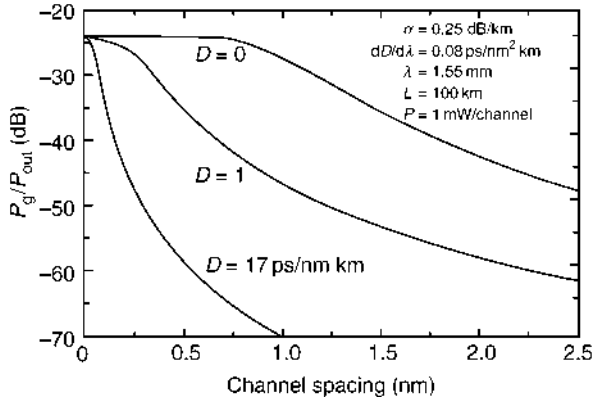

Figure 13 Ratio of generated power to the output as a function of channel spacing $\Delta\lambda$ for different dispersion coefficients. (Reproduced with permission from Tkach RW, Chraplyvy AR, Forghiari F, Gnanck AH and Derosier RM (1995) Four-photon mixing and high speed WDM systems. *Journal of lightwave Technology* 13: 841–849, © 2004 IEEE.)

ratio of generated power to the output as a function of channel spacing $\Delta\lambda$ for different dispersion coefficients. It can be seen that by choosing a nonzero value of dispersion, the four-wave mixing efficiency can be reduced. Larger the dispersion coefficient, smaller can be the channel spacing for the same crosstalk.

Since dispersion leads to increased bit error rates in fiber optic communication systems, it is important to have low dispersion. On the other hand, lower dispersion leads to crosstalk due to FWM. This problem can be resolved by noting that FWM depends on the local dispersion value in the fiber, while the pulse spreading at the end of a link depends on the overall dispersion in the fiber link. If one chooses a link made up of positive and negative dispersion coefficients, then by an appropriate choice of the lengths of the positive and negative dispersion fibers, it would be possible to achieve a zero total link dispersion while at the same time maintaining a large local dispersion. This is referred to as dispersion management in fiber optic systems.

Although FWM leads to crosstalk among different wavelength channels in an optical fiber communication system, it can be used for various optical processing functions such as wavelength conversion, high-speed time division multiplexing, pulse compression, etc. For such applications, there is a concerted worldwide effort to develop highly nonlinear fibers with much smaller mode areas and higher nonlinear coefficients. Some of the very novel fibers that have been developed recently include holey fibers, photonic bandgap fibers, or photonic crystal fibers which are very interesting since they possess extremely small mode effective areas ($\sim 2.5\ \mu m^2$ at 1550 nm) and can be designed to have zero dispersion even in the visible region of the spectrum. This is expected to revolutionize nonlinear fiber optics by providing new geometries to achieve highly efficient nonlinear optical processing at lower powers.

Supercontinuum Generation

Supercontinuum (SC) generation is the phenomenon in which a nearly continuous spectrally broadened output is produced through nonlinear effects on high peak power picosecond and subpicosecond pulses. Such broadened spectra find applications in spectroscopy, wavelength characterization of optical components such as a broadband source from which many wavelength channels can be obtained by slicing the spectrum.

Supercontinuum generation in an optical fiber is a very convenient technique since it provides a very broad bandwidth (>200 nm), the intensity levels can be maintained high over long interaction lengths by choosing small mode areas, and the dispersion profile of the fiber can be appropriately designed by varying the transverse refractive index profile of the fiber.

The spectral broadening that takes place in the fiber is attributed to a combination of various third-order effects such as SPM, XPM, FWM, and Raman scattering. Since dispersion plays a significant role in the temporal evolution of the pulse, different dispersion profiles have been used in the literature to achieve broadband SC.

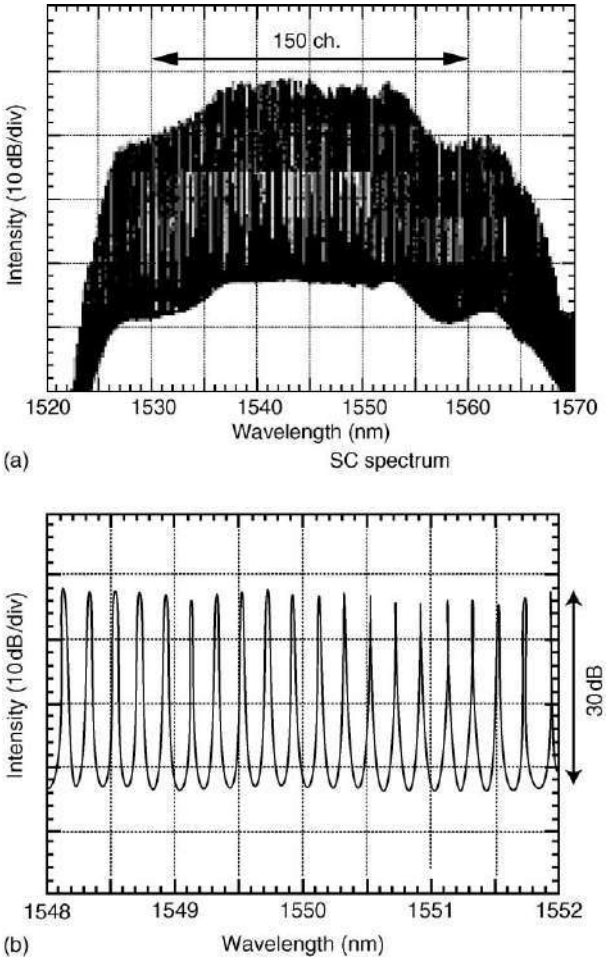

Figure 14 Super Continuum spectrum generated by passing 25 GHz optical pulse train at 1544 nm generated by a mode locked laser diode and amplified by an erbium doped fiber. (Reproduced from Yamada E, Takara H, Ohara T, Satc K, Morioka T, Jinguji K, Itoh M and Ishi M (2001) A high SNR, 150 Ch. Supercontinuum CW optical source with high SNR and precise 25 GHz spacing for 10 Gbit/s DWDM systems. *Electronics Letters* 37: 304–306 with permission from IEEE.)

Some studies have used dispersion decreasing fibers, dispersion flattened fibers, while others have used a constant anamolous dispersion fiber followed by a normal dispersion fiber.

Figure 14 shows the SC spectrum generated by passing 25 GHz optical pulse train at 1544 nm generated by a mode locked laser diode and amplified by an erbium doped fiber. The fiber used for SC generation is a polarization maintaining dispersion flattened fiber. The output is a broad spectrum containing more than 150 spectral components at a spacing of 25 GHz with a flat top spectrum over 18 nm. The output optical powers range from −9 dBm to +3 dBm. Such sources are being investigated as attractive solutions for dense WDM (DWDM) optical fiber communication systems.

Conclusions

The advent of lasers provided us with an intense source of light and led to the birth of nonlinear optics. Nonlinear optical phenomena exhibit a rich variety of effects and coupled with optical waveguides and fibers, such effects can be observed even at moderate optical powers. With the realization of compact femtosecond lasers, highly nonlinear holey optical fibers, quasi phase matching techniques, etc, the field of nonlinear optics is expected to grow further and lead to commercial exploitation such as in compact blue lasers, soliton laser sources and multiwavelength sources and for fiber optic communication.

List of Units and Nomenclature

Attenuation	The decrease in optical power of a propagating mode usually expressed in dB/km
Birefringent medium	A medium characterized by two different modes of propagation characterized by two different polarization states of a plane light wave along any direction
Cross-phase modulation (XPM)	The phase modulation of a propagating light wave due to the nonlinear change in refractive index of the medium brought about by another propagating light wave
Effective length (L_{eff})	The length over which most of the nonlinear effect is accumulated in a propagating light beam
Four-wave mixing (FWM)	The mixing of four propagating light waves due to intensity dependent refractive index of the medium. This mixing leads to the generation of new frequencies from a set of two or three input light beams
Mode effective area (A_{eff})	The cross sectional area of the mode which determines the nonlinear effects on the propagating mode. This is different from the core area
Nonlinear polarization (P_{NL})	When the response of the medium to an applied electric field is not proportional to the applied electric field, then this results in a nonlinear polarization
Optical waveguides	Devices which are capable of guiding optical beams by overcoming diffraction effects using

	the phenomenon of total internal reflection
Phase coherence length	The minimum crystal length upto which the second harmonic power generated by the nonlinear interaction increases
Phase matching	Condition under with the nonlinear polarization generating an electromagnetic wave propagates at the same phase velocity as the generated electromagnetic wave
Pulse dispersion	The broadening of a light pulse as it propagates in an optical fiber
Quasi-phase matching (QPM)	Periodic modulation of the nonlinear coefficient to achieve efficient nonlinear interaction
Second-harmonic generation (SHG)	The generation of a wave of frequency 2ω from an incident wave of frequency ω through nonlinear interaction in a medium
Self-phase modulation (SPM)	The modulation of the phase of a propagating light wave due to the nonlinear change in refractive index of the medium brought about by itself
Supercontinuum generation (SC)	The phenomenon in which a nearly contiunuous spectrally broadband output is produced through nonlinear effects on high peak power picosecond and sub-picosecond pulses
Walk off length (L_{wo})	Length of the fiber required for two interacting pulses at different wavelengths to walk off relative to each other

See also

Nonlinear Optics, Applications: Phase Matching. **Nonlinear Optics, Basics:** $\chi^{(2)}$–Harmonic Generation; $\chi^{(3)}$–Third-Harmonic Generation; Four-Wave Mixing. **Spectroscopy:** Second Harmonic Spectroscopy.

Further Reading

Agrawal GP (1989) *Nonlinear Fiber Optics*. Boston, MA: Academic Press.

Armstrong JA, Bloembergen N, Ducuing J and Pershan PS (1962) Interactions between light waves in a nonlinear dielectric. *Physics Review* 127: 1918.

Baldeck PL, Alfano RR and Agrawal GP (1998) Induced frequency shift of copropagating ultrafast optical pulses. *Applied Physics Letters* 52: 1939–1941.

Baldi P (1992) *Generation de fluorescence parametrique guidee sur niobate et tantalite de lithium polarizes periodiquement. Etudes preliminaire d'un oscillateur parmetrique optique integree*. Doctoral Thesis of the University of Nice, France.

Chiang TK, Kagi N, Marhic ME and Kazovsky LG (1996) Cross phase modulation in fiber links with multiple optical amplifiers and dispersion compensators. *Journal of Lightwave Technology* 14: 249–259.

Franken PA, Hill AE, Peter CW and Weinreich G (1961) Generation of optical harmonics. *Physics Review Letters* 7: 118.

Ghatak AK and Thyagarajan K (1987) *Optical Electronics*. Cambridge, UK: Cambridge University Press.

Ghatak AK and Thyagarajan K (1998) *Introduction to Fiber Optics*. Cambridge, UK: Cambridge University Press.

Lim EJ, Fejer MM and Byer RL (1989) Second-harmonic generation green light in periodically poled planar lithium niobate waveguide. *Electronic Letters* 25: 174.

Lim EJ, Fejer MM, Byer RL and Kozovsky WJ (1989) *Electronic Letters* 25: 731–732.

Myers LE and Bosenberg WR (1997) Periodically poled lithium niobate and quasi phase matcher parametric oscillators. *IEEE Journal of Quantum Electronics* 33: 1663–1672.

Newell AC and Moloney JV (1992) *Nonlinear Optics*. Redwood City, CA: Addison Wesley Publishing Co.

Rapp L (1997) Experimental investigation of signal distortions induced by cross phase modulation combined with dispersion. *IEEE Photon Technical Letters* 9: 1592–1595.

Shen YR (1989) *Principles of Nonlinear Optics*. New York: Wiley.

Shtaif M, Eiselt M and Garret LD (2000) Cross-phase modulation distortion measurements in multispan WDM systems. *IEEE Photon Technical Letters* 12: 88–90.

Tkach RW, Chraplyvy AR, Forghiari F, Gnanck AH and Derosier RM (1995) Four-photon mixing and high speed WDM systems. *Journal of Lightwave Technology* 13: 841–849.

Wang QZ, Lim QD, Lim DH and Alfano R (1994) High resolution spectra of self phase modulation in optical fibers. *Journal of Optical Society of America* B-11: 1084.

Webjorn J, Siala S, Nan DW, Waarts RG and Lang RJ (1997) Visible laser sources based frequency doubling in nonlinear waveguides. *IEEE Journal of Quantum Electronics* 33: 1673–1686.

Yamada M, Nada N, Saitoh M and Watanabe K (1993) *Applied Physics Letters* 62: 435–436.

Yamada E, Takara H, Ohara T, Sato K, Morioka T, Jinguji K, Itoh M and Ishi M (2001) A high SNR, 150 Ch. Supercontinuum CW optical source with high SNR and precise 25 GHz spacing for 10 Gbit/s DWDM systems. *Electronics Letters* 37: 304–306.

Yariv A (1997) *Optical Electronics in Modern Communications*, 5th ed. New York: Oxford University Press.

Optical Fiber Cables

G Galliano, Telecom Italia Lab, Torino, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In order to be put to practical use in the telecommunication network, optical fibers must be protected from environmental and mechanical stresses which can change their transmission characteristics and reliability.

The fibers, during manufacturing, are individually protected with a thin plastic layer (extruded during the drawing) to preserve directly their characteristics from damage due to the environment and handling. This coating surrounding the fiber cladding (with an external diameter of 125 μm) is known as the primary coating, which is obtained with a double layer of UV resin for a maximum external diameter of about 250 μm.

Primary coated fiber, at the end of the manufacturing process, is subjected to a dynamic traction test (proof-test) to monitor its mechanical strength. This test, carried out on all fibers manufactured, consists of the application of a well-defined strain (from 0.5% to 1%) for a fixed time (normally 1 s) and thus permits the exclusion of, from successive cabling, those fibers with poor mechanical characteristics.

At the end of the manufacturing process the optical fibers are not directly usable. It is necessary to surround the single fiber (or groups of fibers) with a structure which provides protection from mechanical and chemical hazards and allows for stresses that may be applied during cable manufacture, installation, and service.

Suitable protection is provided using fiber conditioning within the cable (secondary coating or loose cabling), while entire protection is obtained with some cable elements, such as strength members, core assembling, filling, and outer protections. Moreover, the cable structure and its design is heavily influenced by the application (e.g., number of fibers), installation (laying and splicing), and environmental conditions of service (underground, aerial, or submarine).

In this article, the main characteristics of optical cables (with both multimode or single mode fibers) are described and analyzed, together with suitable solutions.

Secondary Fiber Coating

The fiber, before cabling, can be protected with a secondary coating (a tight jacketing) to preserve it during cable stranding. However, fiber with only a primary coating may be cabled with a suitable cable protection.

In the first case the fiber can be directly stranded to form a cable core; in the second case, a single group of fibers must be inserted into a loose structure (tubes or grooves) with a proper extra-length. These two solutions are not always distinguishable; in fact, tight fibers inserted into loose tubes or grooves are often used.

The implementation of the secondary coating and the choice of proper materials requires particular attention, in order to avoid microbending effects on the fiber, which can cause degradation of the transmission characteristics. Microbending is caused by the pressure of the fiber on a microrough surface, or by the fiber buckling, due to the contraction of the structure (e.g., secondary coating) containing it. Microbending causes an increase in fiber attenuation at long wavelengths (1300 and 1550 nm) for both single and multimode fibers, but is particularly emphasized in multimode fibers.

Tight Jacket

In a tight protection, the primary coating fiber is surrounded by a second plastic jacket in contact with the fiber. This jacket mainly protects the fiber from lateral stresses that can cause the microbending.

The fiber can be individually protected with a single or double layer of plastic material or in a ribbon structure. In Figure 1 four different types of tight jacket are shown. In Figure 1a and b the fibers are singularly protected with a single or double layer up to an external diameter of 0.9 mm. The second type is normally preferred because the external hard plastic gives a good fiber protection and the inner soft plastic avoids microbending effects on the fiber.

In a ribbon structure, from 4 to 16 primary coated fibers are laid close and parallel and covered with plastic. Two types of ribbon are normally classified 'encapsulated' (Figure 1c) or 'edge bonded' (Figure 1d). In the first case, the fibers are assembled with a common secondary coating (one or two layers), in the second case, an adhesive is used to stick together the fibers with only the primary coating.

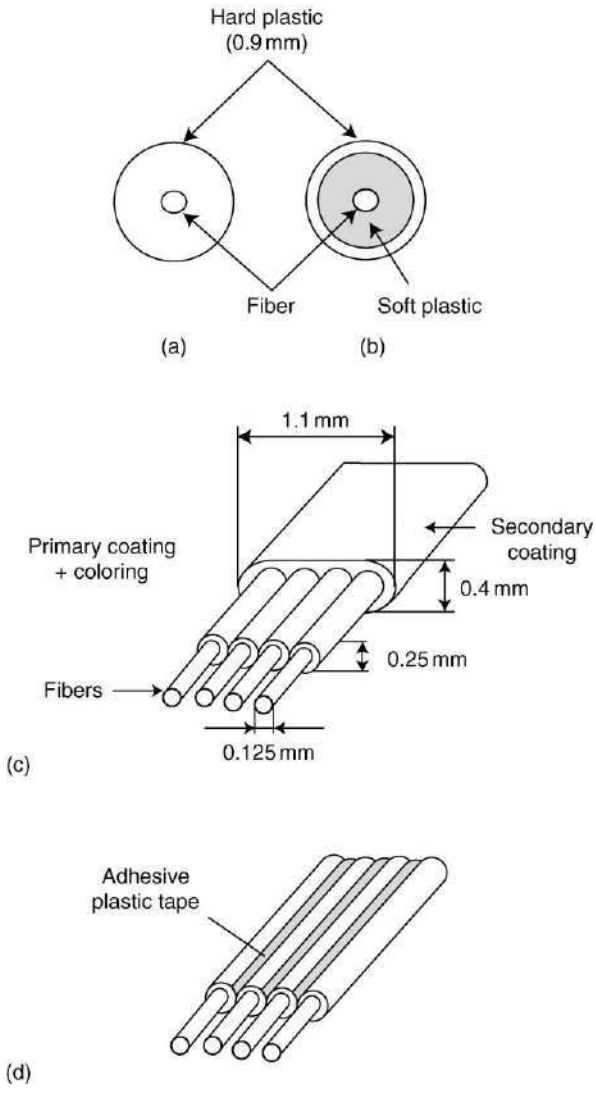

Figure 1 Different types of tight coatings.

Care should be taken in applying tight protection, in particular for a ribbon structure, to avoid an attenuation increase due to the action of the jacket on the fiber (in particular for the thermal contraction of the coating materials). The ribbon structure is designed to permit the simultaneous junction of multiple fibers; for this reason the ribbon geometry must be precision controlled.

Loose Jacket

In a loose protection the fiber can be inserted into a cable with only the primary coating. A thin layer of colored plastic may be added onto the primary coating to identify the fiber in the cable. The fibers are arranged in a tube or in a groove on a plastic core (slotted core). The diameter of the tube or the dimension of the core must be far larger than the diameter of the fiber so that the fiber is completely free inside the structure. The fibers can be inserted into a tube or into a slotted core, singularly or in groups. The space inside the tube around the fibers can be empty or filled with a suitable jelly. Some examples of loose structures are shown in Figure 2. In the tube structure, the external diameter of the jacket depends on the number of the fibers: for a single fiber up to 2 mm (Figure 2a), and for a multifiber tube (Figure 2b) up to 3 mm (for 10 fibers).

Examples of slotted core structures are shown in Figure 2c and d, respectively, for one fiber and for a group of fibers, the dimensions of grooves depending on the number of fibers.

Plastic materials with a high Young modulus are normally used for the tube that is extruded on the single or on the group of fibers. Instead, for the slotted core, plastic materials with good extrusion are used to obtain a precise profile of the grooves.

The loose structure guarantees freedom to the fibers in a defined interval of elongation and compression. Out of this interval, microbending effects can arise which increase fiber attenuation. Axial compression generally occurs when the cable is subjected to low temperatures and is generated from the thermal contraction of the cable materials that is different from that of the fibers. Axial elongation normally occurs when the cable is pulled or by the effects of high temperature.

For a correct design of a cable with a loose structure, the extra length of the fiber with respect to the tube or groove dimension and the stranding pitch must be evaluated, starting from the range of axial compression and elongation. In fact, care must be taken to prevent the fiber from being strained or forced against the tube or groove walls in the designed range.

For the tube solution, the extra length of the fibers is controlled during the tube extrusion; for the slotted core solution, the length is controlled during cabling, giving a controlled tension to the slotted core. The loose structure is often used for the cabling of fibers with tight jacketing, joining the advantages of the tight protection and loose cable. This solution is typical for ribbon cables.

Fiber Identification

To identify the fibers in a cable, a proper coloration with defined color codes must be given to the single fibers. In the tight protection, the tight buffer is directly colored during extrusion. In the loose protection, a thin colored layer is added on the primary coating of the fiber; tubes or slotted core may

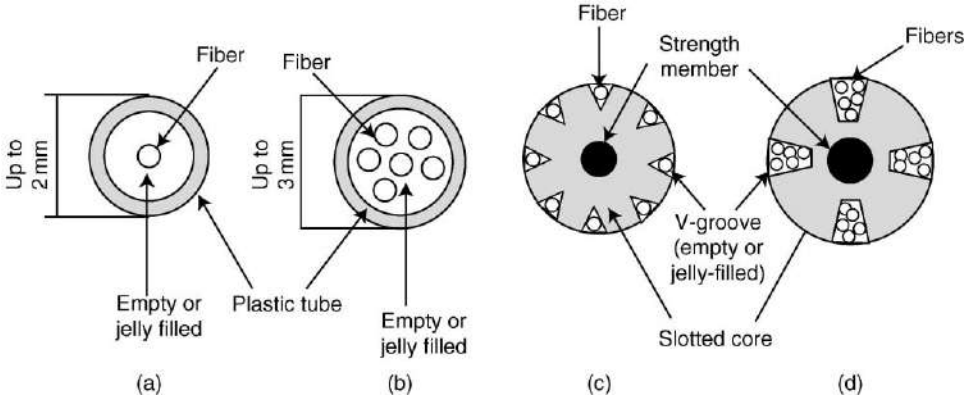

Figure 2 Different types of loose coatings.

be also colored to identify groups of fibers in high potentiality cables.

Cable Components

The fundamental components of an optical cable are: the cable core with the optical fibers, the strength member (one or more) which provide the mechanical strength of the cable, sheaths which provide the external protection, and filling materials.

Cable Core

The optical fibers with the secondary coating (tight or loose) are rejoined together in a cable core. For tight fibers or loose tubes, the cable core is obtained by stranding the fibers or the tubes around central elements that normally act as strength members too (see below).

The stranding pitch must be long enough to avoid excessive bending of the fibers and short enough to guarantee stress apportionment among the fibers in cable bends and maximum elongation of the cable without fiber strain. Normally, a helical stranding is used, but often SO stranding, that consists in an inversion of stranding direction every three of more turns, is employed. The second stranding type makes the cabling easier but may reduce its performance.

The unit core or the slotted core are made of plastic and normally incorporate a central strength member.

Strength Member

The strength member may be defined as an element that provides the mechanical strength of the cable, to withstand elongation and contraction during the installation and to ensure thermal stability. Fiber elongation, when the cable is pulled or the fiber compressed due to low temperatures, depends greatly

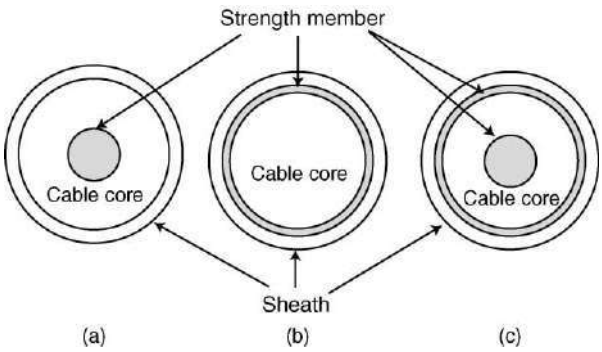

Figure 3 Different locations of the strength member in an optical cable.

on the characteristics of the strength member inserted in the cable.

As the maximum elongation values suggested for the fibers are in the range from 20 to 30% of the proof-test value, the maximum elongation allowed for a tight cable is ~0.1% instead of ~0.3% for a loose cable. To warrant these performances, a material with a high Young modulus is normally used as a strength member. Also, the strength member must be light (to avoid excessive cable weight) and flexible: these properties make it easier when laying the cable in ducts.

The strength member may be metallic or non-metallic and may be inserted in the core (Figure 3a) or in the outer part of the cable (Figure 3b), normally under the external sheath. Often two strength members are placed in an optical cable: the first one in the core and the second one under the sheath (Figure 3c).

Metallic strength members, generally one or more wires of steel, are inserted in the center or in the outer part of the cable. The steel, for its low thermal expansion coefficient and high Young modulus, behaves as a good strength member, particularly

when it is arranged in the central core. Steel is a nonexpensive material, but needs protection against corrosion and electrical-induced voltages, such as by lightning or current flow in the ground.

If cables with high flexibility are requested, steel is not suitable and aramid fiber yarn or glass-fiber reinforced plastic (G-FRP) are used, particularly when the strength member is placed in the outer part of the cable. Glass fiber elements are generally used as a compressive strength member in the central cable core to avoid contraction at low temperatures, but they may be used as strength elements when the traction strength is low.

Aramid yarns are not active as resisting compression elements but are normally arranged in the outer part of the optical cable. Aramid yarns have a high Young modulus (as with steel) and low specific weight: in commercial form they consist of a large number of filaments gathered into a layer. Normally two layers stranded in opposite directions are used to avoid cable torsion during the installation.

Metal-free cables are protected by G-FRP and aramid yarns. This solution results in optical cables which are insensitive to electrical interferences.

Sheaths

The functions of the sheaths are to protect an optical cable during its life, in particular from mechanical and environmental stresses. During the installation or when the cable is in service, it may be subjected to mechanical stresses, for example, presence of humidity or water and chemical agents. The sheaths act as a barrier to avoid degradations of the fibers in the internal cable core.

The plastic sheaths (one or more, depending on the cable type and structure) are extruded over the core. This extrusion is a delicate operation in cable manufacturing because the cable core may be subjected to tension, while during the cooling the thermal compression of the plastic produces a stress on all the cable elements. Examples of cable sheaths are shown in Figure 4.

Different plastic materials, such as polyvinylchloride (PVC), polyethylene, and polyurethane are normally used. PVC has excellent mechanical properties, flexibility, and is flame retardant but is permeable to humidity. On the other hand, high-density polyethylene has a permeability lower than PVC, with good mechanical properties and low coefficient of friction, but it is more flammable than PVC and less flexible. Finally, polyurethane, for its softness and high flexibility, is often used for inner sheaths. When low toxicity is requested (e.g., in indoor cables), halogen-free materials are employed.

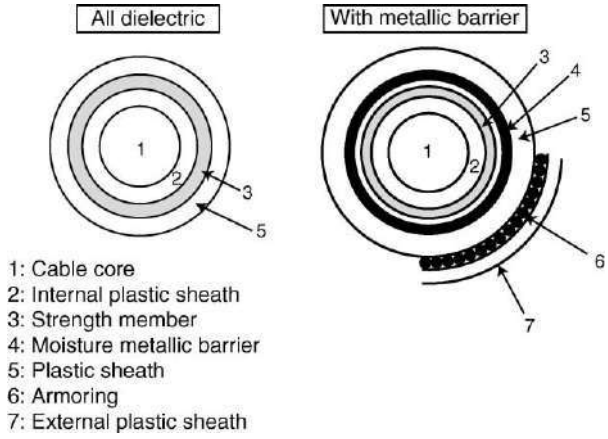

Figure 4 Examples of optical cable sheaths.

Often a metallic barrier is used under the external sheath to prevent moisture entering the cable core. The type most widely used is an aluminum ribbon (of a few tenths of a millimeter) bonded into the external polyethylene sheath, during the sheath extrusion. This structure is normally indicated as a LAP (laminated aluminum polyethylene) sheath.

When the cable is directly buried, a metallic armor (corrugated steel tape or armoring wires) is generally inserted. The metallic armor protects the cable core against radial stresses but it is also used as protection against rodents and insects.

Filling Compounds and Other Components

Fillings are used in optical cables to avoid the presence of moisture and water propagation in the cable core should a failure occur in the cable sheath. Suitable jelly compounds, with constant viscosity at a wide range of temperatures and chemical stabilities, are generally inserted in the cable core and, sometimes, in the outer protection layer. In loose structures, the jelly is in direct contact with the fibers: the compound must be compatible with them and guarantee the fiber freedom. Special compounds may be used to adsorb and permanently fix any hydrogen present in the cable core. The presence of hydrogen gas, that may be generated from the internal materials of the cable (mainly metals), is harmful to the fibers because it causes permanent attenuation increases and thus damage to them.

With the generic terminology 'other components', some other components used in the optical cables are included. They are often similar to those used in conventional cables. The most important are: insulated conductors, plastic fillers, and cushion layers and tapes around the fiber core.

Insulated copper conductors may be incorporated in the cable core, for example, to carry power supply

or as service channels, or to detect the presence of moisture in the cable (in this case, the insulation is missing or partial). The conductors may cause problems due to the voltages induced by power lines or lighting.

Plastic fillers are often used to complete the cable core geometry. Cushion layers are used to protect the cable core against radial compression. Normally, they are based on plastic materials wound around the core. Tapes are wound around the cable core to hold the assemblies together and secondly to provide a heat barrier during the sheath extrusion. Mylar tapes are usually employed.

Optical Cable Structures, Types, and Applications

Cable Structures

The structure of an optical cable is strictly dependent on the construction method and the application type that determines the design and the grade of external protection. The main core (or inner) structures of an optical cable can be classified as: stranded structures (tight and loose); slotted core cable; or ribbon cable. In this section, a few examples of cable structures are presented. Below, separate sections are devoted to submarine cables and to special application cables, for their particular structures and features.

In the stranded tight structure, a low number of tight or loose fibers is stranded around a strength member to form a cable unit. Some of this cable unit may be joined to constitute a cable with a medium/high number of fibers (Figure 5a). Alternatively the fibers may be arranged in a multilayer structure (Figure 5b).

The main advantages of a tight stranded structure are the small dimensions of the cable core, even when a large number of fibers are involved, and also the facility of handling. Care must be taken in the design of the strength member: in fact, the fibers are immediately subjected to tension when the cable is pulled into the ducts or compressed when subjected to low temperatures.

In the stranded loose structures, two categories of cables are defined, one fiber per tube and a group of fibers (up to 20) per tube (Figure 5c). As already mentioned, in the loose structures, the fibers are free inside the tube and for this reason the cable is less critical than the tight cable. The external dimensions of loose cables are larger than tight cables, but the solution with multiform tubes allows for cables with a high number of fibers in relatively small dimensions.

Slotted core cables (Figure 2c,d) contain fibers with only the primary coating and may be divided, as in the stranded loose structure, into one fiber per groove or into a group of fibers per groove. The cable design and the choice of a suitable plastic material for the slotted core joined to a proper strength member, make it possible to obtain cables with fiber stability at a wide range of temperature. As in the loose stranded solutions, the slotted core solution is adopted to obtain cables with a high number of fibers in small dimensions.

In the ribbon solution, the fibers (from 4 to 16) are placed in a linear ribbon array. Many ribbons are stacked together to constitute the optical unit. The ribbon cable normally has a loose structure (loose tube or slotted core). Three structures are shown in Figure 6. In the first example (classical AT & T cable), 12 ribbons of 12 fibers are directly stacked and stranded in a tube (Figure 6a). In the second and third examples, ribbons are inserted into tubes or grooves (Figure 6b,c). Several tubes or slotted cores can be assembled together to give a cable with very high fiber density. The ribbon approach may be efficient when a large number of fibers must be

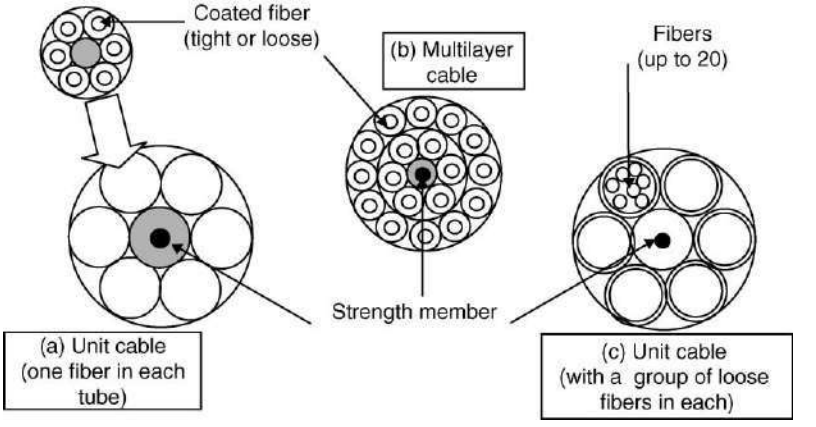

Figure 5 Structures of stranded core cables.

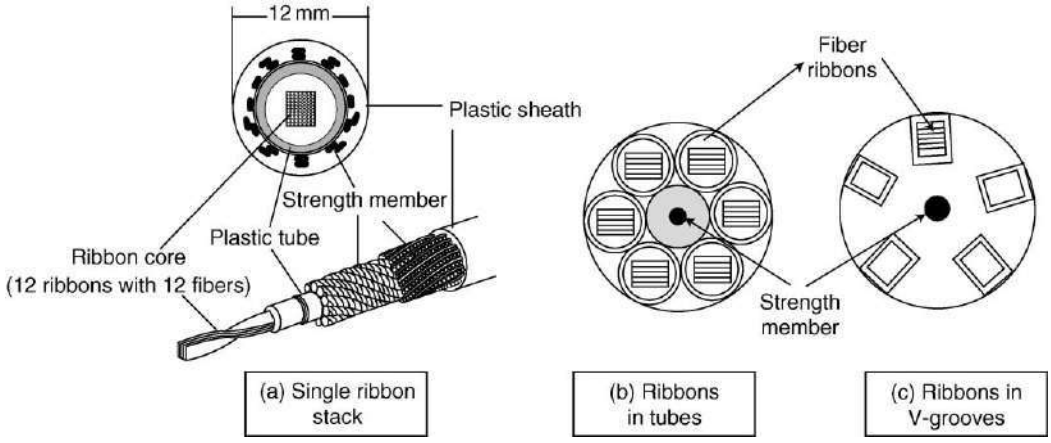

Figure 6 Structures of ribbon cables.

simultaneously handled, spliced, and terminated at connectors.

Types and Applications

The choice of an optical cable is heavily established by many factors, but mainly on the type of installation. On this basis, some different typology of cables may be identified: aerial cables, duct cables, direct buried cables, indoor cables, underwater cables, etc.

Aerial cables can be clasped onto a metal strand or can be self supporting. In the first case, the cable is not subjected to a particular tension but requires good thermal and mechanical performances. This solution is chosen when the cable must withstand strong ice and wind and when long distances between the poles are required. All the structures mentioned above may be adopted. In the second case, the cable is normally exposed to high mechanical stresses during its life and high tensile strength must be guaranted to maintain the fiber elongation at a safety level. Generally, loose structures are preferred for their greater fiber freedom. Often, in self-supporting cables, the strength member is external to the cable core. An example is shown in Figure 7.

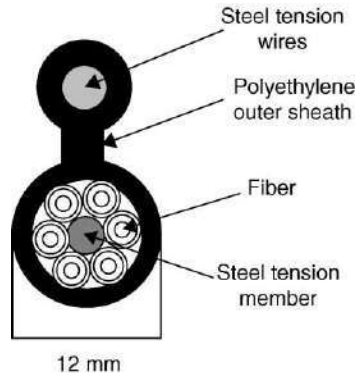

Figure 7 Self-supporting aerial cable.

Cables pulled in ducts must be resistant to pulling and torsion forces, and light and flexible to permit the installation of long sections. The cable must also protect the fibers against water and moisture that may be present in ducts or manholes. Usually the cable is filled with a jelly compound, a metallic barrier is normally used and, often, an armoring is added when protection against rodents is necessary. All the structures described above may be employed in duct cables.

The characteristics of direct buried cables are very similar to those of duct cables, but additional protections, such as metallic armoring, are required to avoid the risk of damages from digging and other earthmoving work.

Cables for indoor applications normally require a smaller number of fibers. Their main characteristics are: small dimensions, flexibility, small curvature radius, and ease in handling and jointing. For indoor applications, the sheath must not permit flame propagation and must have a low emission of toxic gases and dark smoke. Tight cables are preferably used for indoor applications due to their compactness and small dimensions.

Underwater cables are employed in the crossing of rivers, in lakes and lagoons, and for shallow water applications in general. Such cables must have stringent requirements such as resistance to moisture and water penetration and propagation, pulling resistance (during the installation, and recovery in event of failure), and high resistance to static pressure and core structure compatible with land cables (when it is part of a ground link). In addition, due to the presence of metallic structures, attention should be paid to hydrogen effects.

Submarine Cables

Submarine cable is manufactured for laying in deep-sea conditions. In designing a submarine cable, it is necessary to provide high reliability that the mechanical and transmission characteristics of the optical fibers will be stable over a long period of time. Different cables are used for submarine applications, but generally follow the basic scheme of Figure 8.

The central core (with a few fibers – up to 12 – in a slotted core or in a tight structure) is surrounded by a double layer of steel wires that act as strength members against tensile action and water pressure. Metal pipes at the inner or outer sides of the strength member, act as a water barrier and as power supply conductors (for the supply of the undersea regenerators).

The outer polyethylene insulating sheath is the ultimate protection for the ordinary deep-sea cable. For cables requiring special protection, steel wire armoring (anti-attack from fishes), and further polyethylene are added.

The cable must have stringent mechanical requirements, such as resistance to traction, torsion, crushing, impact, and to shark attack. The cable must be suitable for installation using standard cable-laying ships. Furthermore the cable materials must have low content of hydrogen and emissions. If necessary, a hydrogen absorber may be included in the cable core.

Special Cables

This category comprises cables not specifically used in ordinary telecommunication networks, but cables used for specific applications and according to specified requirements. Cables for military use are normally employed for temporary plant restoration. They must be light but crush resistant and with good mechanical characteristics. Cables for mobile applications (robots, elevators, aircrafts, submarines, etc.) generally require high flexibility and tensile strength. Special cables may be installed in particular environments, where they must be insensitive, for instance, to nuclear radiation, chemical agents, and high temperatures.

Optical cables can be installed in power lines, together with the power conductors, or into the ground wires of high-voltage overhead power lines. In this last type of cable, the central stranded steel wire, making up the ground wire, is replaced with a metal tube containing the optical unit (normally a small groups of optical fibers inserted in a tube or in a slotted core).

Blown Fiber

This technology consists of the installation of a single fiber or of small fiber bundles (or cables) into pre-installed tubes or ducts using a flux of compressed air. Initially, cables with empty tubes are installed and successively, at the time of need, the tubes can be filled by blowing fibers (singly, or in bundles) or cables. The installation time is normally fast; some compressors can blow tens of meters in a minute. The typical maximum blowing distance is about 1 km for horizontal planes, but it is reduced

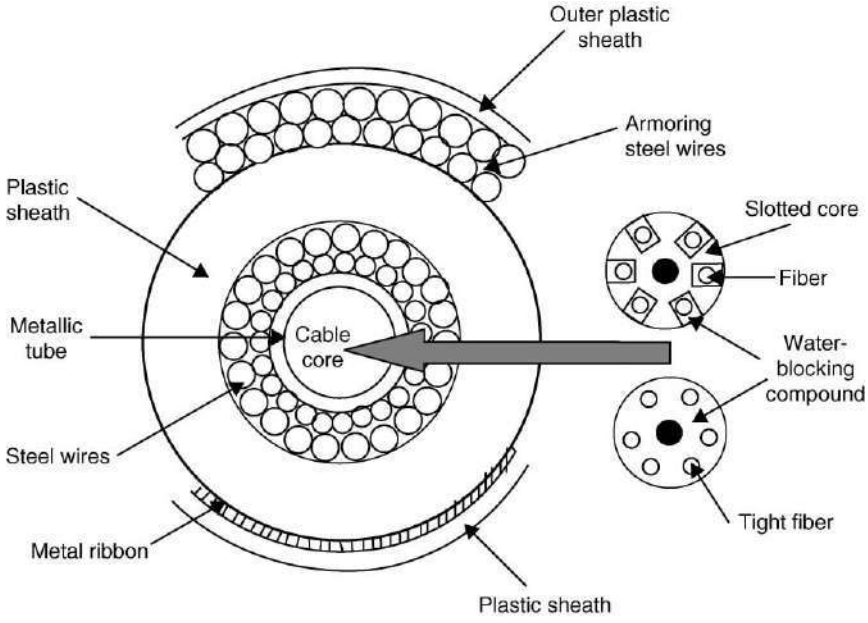

Figure 8 Basic structure of a submarine cable.

along winding roads. The advantage of this technique is to produce variable and flexible structures, increasing the number of fibers to fit the network need. Besides, substitutions or change of fibers can be quickly carried out without substitution of the cable.

See also

Fiber and Guided Wave Optics: Dispersion. **Light Emitting Diodes**.

Further Reading

Griffioen W (1989) The installation of conventional fiber-optic cables in conduits using the viscous flow of air. *Journal of Lightwave Technology* 7(2): 297–302.

Hughes H (1997) *Telecommunications Cables. Design, Manufacture and Installation*. : John Wiley and Sons.

Keiser G (1986) *Optical Fiber Communications*. McGraw-Hill.

Murata H (1988) *Handbook of Optical Fibers and Cables*. Marcel Dekker.

Technical Staff of CSELT (1990) *Fiber Optic Communications Handbook Second Edition*. : TAB Books.

Passive Optical Components

D Suino, Telecom Italia Lab, Torino, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Modern telecommunications networks are based on fiber optics as signal transporters. In these networks the signals travel in the form of light confined within the fibers. As light power travels through the fibers from the source apparatus to the receiver, a number of components are employed to manage the optical signal. All other components that are not used to manage the signal are called 'passive optical components'. They serve several functions, such as to join two different fibers to distribute the signal onto more optical branches to limit the optical power to select particular wavelengths, and so on.

These components can be located anywhere in the network, depending on their function and on the network architecture. However, their main location is near the apparatus in which it is necessary to manage the signal before it is sent through the network or at the junction in the network where it is sent on to the receiver.

The main types of passive components that are found in a network can be grouped as follows:

- Joint devices:
  - Optical connectors;
  - Mechanical splices.
- Branching devices:
  - Nonwavelength-selective branching devices;
  - WDM.
- Attenuators;
- Filters;
- Isolators;
- Circulators.

Optical fibers are also divided into several groups (multimode 50/125, multimode 62.5/125, singlemode, dispersion shifted, NZD, and so on). There are also components of different typology that match up with each particular fiber. However, the general concept in terms of functionality, technology, and characteristic parameters, are the same among passive optical components of the same type (connectors, branching devices, or attenuators) but of different groups (single-mode, multimode, etc.).

Optical Joint Devices

The function of an optical joint is to perform a junction between two fibers, giving optical continuity to the line. Two different classes of joint can be considered: the connectors and the splices. The joints made by using connectors, are flexible points in the network, which means they can be opened and reconnected several times in order, for example, to reconfigure the network plan. However, a splice is a fixed joint and usually cannot be opened and remated. There are two type of splice: mechanical splices, holding joints to fibers by glue or mechanical crimp, and fusion splices fusing the two fiber heads with a suitable fusion splicer.

General Considerations

Considering that fibers can be described as pipes, in the core of which flows light, continuity between two fibers means that the two cores must be aligned in a stable and accurate way. Obviously this alignment must be carried out with minimum power loss in the junction. The main factors that

affect the power loss in a joint point between two fibers are:

- x = Lateral offset;
- z = Longitudinal offset;
- θ = Angular misalignment; and
- w_T/w_R = Mode field diameter ratio (in the single mode fiber) or numerical aperture ratio (in the multimode fiber).

These geometrical mismatching between the two fibers is represented in Figure 1.

The function that describes the joint attenuation is different for single-mode and multimode connectors, but the parameters that contribute to it are the same. The general formulation for these functions are complicated and, in particular, for the multimode connector it involves an integral on the shape distribution of the light in the fiber core. These functions become simpler in particular conditions: for example, if we consider a single-mode joint without a gap between the two fibers (that is the more usual condition in the modern connectors) the z parameter is nil, and in this case the function that describes the attenuation of the connector is

$$A(\mathrm{dB}) = -10\log\left\{\left(\frac{2w_T w_R}{w_T^2 + w_R^2}\right)^2 \times \exp\left[-2\frac{(x\lambda)^2 + (\pi n w_T w_R \sin(\vartheta))^2}{\lambda^2(w_T^2 + w_R^2)}\right]\right\} \quad [1]$$

where n is the fiber core refractive index and λ is the wavelength of the light.

Figure 2 shows the trend of eqn [1], where the parameter ranges in the graphs are typical for the most common products on the market. It is evident that the main contribution to attenuation arises from the lateral offset.

For connectors of good quality, both for single-mode and multimode fibers, the attenuation values are in the range of a few tenths of dB (typically less than 0.3 dB). Another important factor for a joint, in particular for connectors used in networks for high speed or analog signals, is the return loss (RL). This is a parameter that gives a measure of the quantity of light back-reflected onto the optical discontinuity of the connector, due to the Fresnel effect. The RL is defined as the ratio in dB between the incident light power (P_{inc}) and the reflected light power (P_{ref}):

$$\mathrm{RL} = -10\log\left(\frac{P_{ref}}{P_{inc}}\right) \quad [2]$$

Fresnel reflection arises when the light passes between two media with different refractive indices. In the case of connectors in which the two fibers are not in contact, the light passes from the silica of the fiber

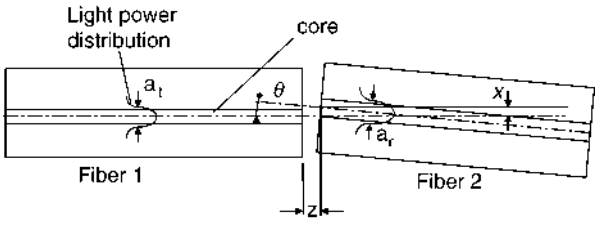

Figure 1 Main parameter that affects the joint power loss.

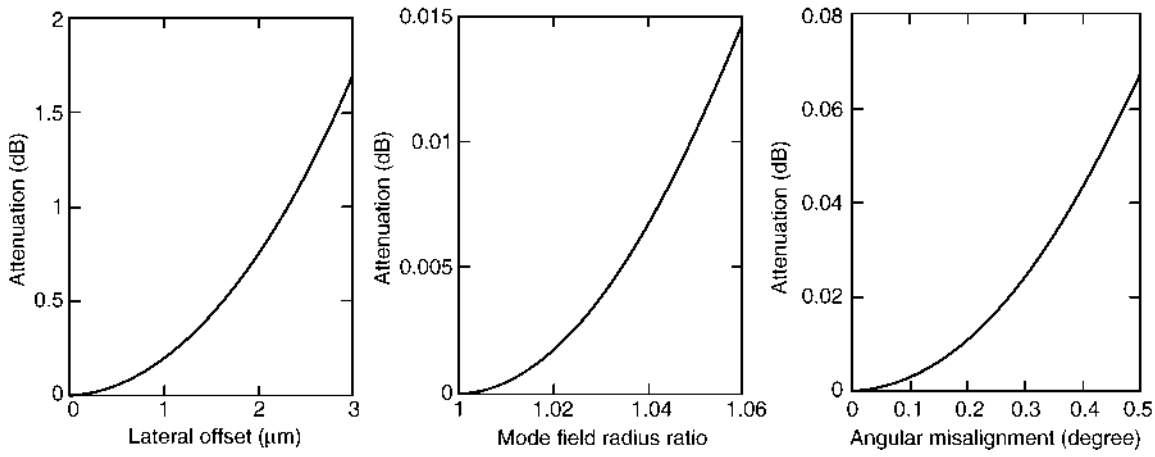

Figure 2 Single mode connector attenuation as a function of the geometrical parameters. Each curve is plotted keeping the other parameter constant.

core to the air in the gap. This causes the reflection of 4% of the incident light that, as RL, is about 14 dB.

To improve the RL performances of an optical joint, it is necessary to reduce the refractive index difference. This can be achieved either by index matching material between the two fibers that reduces the differences between the refractive index of the fiber and the gap, or by performing a physical contact between the fiber heads. The first technique is typically used in mechanical splices but, due to the problem related with the pollution contamination of the index matching, it is not a good solution for connectors that are to be opened and reconnected several times.

The physical contact solution is the most frequently adopted for these connectors. Physical contact is obtained by polishing the ferrule and the fiber end-faces in a convex shape, with the fiber core positioned in the apex (Figure 3). With this technique, the typical values of return loss for a single mode connector are in the range from 35 dB to 55 dB, depending on the polishing grade. The residual back-reflected light is generated in the thin layer of the fiber end surface, because of slight changes in the refractive index due to the polishing process.

Higher return loss values (small quantity of reflected light) are achieved with the angled physical contact (APC) technique. This is obtained by polishing the convex ferrule end face angled with respect to the fiber axis. In this way, the light is not back reflected into the fiber core, but into the cladding, and is thus eliminated (Figure 4). In APC connectors, typical return loss values are greater than 60 dB.

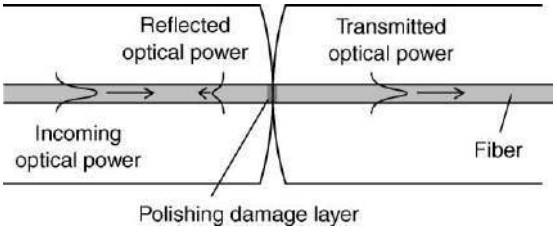

Figure 3 Back reflection effects in PC connectors.

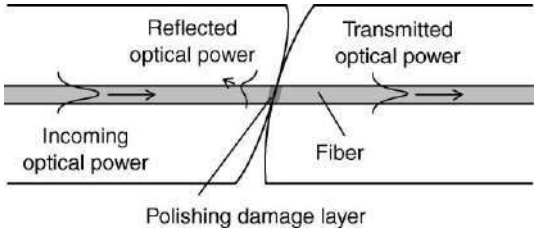

Figure 4 Back reflection effects in APC connectors.

Optical Connectors

On the market are several types of optical connectors. It is possible to divide them into three main groups:

- Connectors with direct alignment of bare fiber;
- Connectors with alignment by means of a ferrule; and
- Multifiber connectors.

Connectors with direct alignment of bare fiber

In this type of connector the bare fibers are aligned on an external structure that is usually a V-groove (Figure 5). Two configurations can be performed: plug and socket or plug–adapter–plug. In the first case, the fiber in the socket is fixed and the one in the plug aligns to it. In the plug–adapter–plug configuration, two identical plugs containing the bare fibers are locked onto an adapter in which the fibers are placed on the aligned structure. These connectors are cheaper than the connectors with a ferrule, (described below), but this is unreliable. In fact, the bare fibers are delicate and brittle and they can be affected by the external mechanical or environmental stresses.

Connectors with alignment by means ferrule

In this type of connector, the fibers are fixed in a secondary alignment structure, usually of cylindrical shape, called a ferrule. The fiber alignment is obtained by inserting the two ferrules containing the fibers into a sleeve. Usually this type of connector is in the plug–adapter–plug configuration (Figure 6). With this technique, the precision of the alignment is moved from the V-groove precision to the dimensional tolerance of the ferrule. Typical tolerance values on the parameters for a cylindrical ferrule with a diameter of 2.5 mm, are in the range of a tenth of a micrometer, and must be verified on the diameter, eccentricity between the hole (through which the fiber is fixed) and cylindricity of the external surface.

Multifiber connectors

In some types of optical cable, the fibers are organized in multiple structures, in which many fibers are glued together to form a ribbon. The ribbon can have

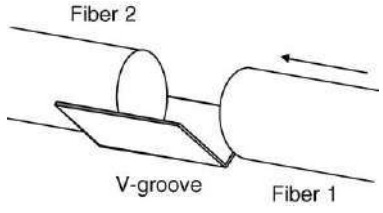

Figure 5 Scheme of the bare fibers alignment on a V-groove.

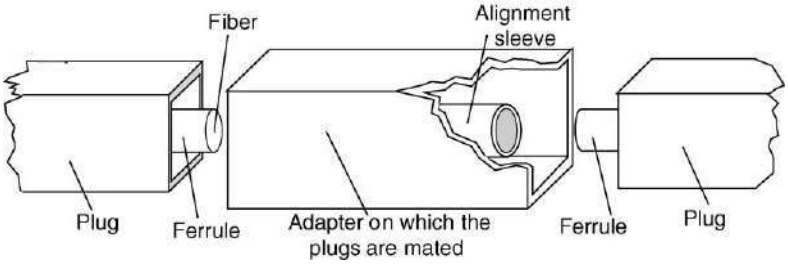

Figure 6 Scheme of a plug–adapter–plug connector with alignment by ferrules and sleeve.

2, 4, 8, 12, or 16 fibers, for example. When these cables are used, it is convenient to maintain the ribbon fiber structure in the connectors (until the point in which it is necessary to use the fibers singularly). Multifiber connectors are typically based on a secondary alignment on a plastic ferrule of rectangular shape in which the fibers maintain the same geometry that they have in the ribbon. The rectangular ferrules of the two plugs are normally aligned by means of two metallic pins placed in two hole in line with the fiber. In Figure 7, a scheme of a multifiber connector is shown.

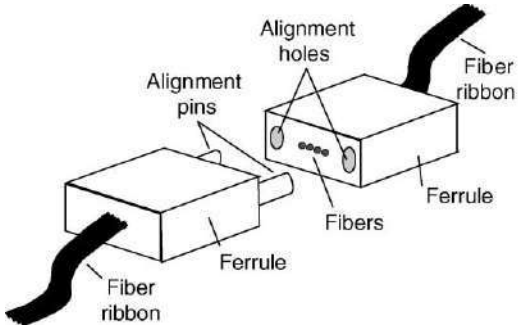

Figure 7 Multifiber connector.

Mechanical Splices

Mechanical splices are yet another way to join together two fibers. All the main considerations for the connectors are true for the mechanical splices. The difference is that the mechanical splice is a fixed joint and, normally, it cannot be opened and re-mated. Mechanical splices, both for single fiber and for multifiber structures, are the most usually manufactured. In a mechanical splice, the fibers are aligned directly in a V-groove and fixed by glue or a mechanical holder.

These components are cheaper than fusion splices but their optical, mechanical, and environmental functions are lower than the fusion splices. Figure 8, shows a mechanical splice for multifibers in which the fibers are crimped on a V-groove array. Usually, to avoid high optical reflection in the junction point, this type of joint is filled with index matching material.

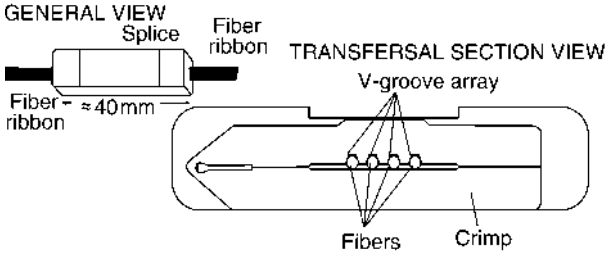

Figure 8 Example of a multifiber mechanical splice.

Branching Devices

Branching devices are passive components devoted to distributing the optical power from an input fiber to two or more fibers. These components can be classified in two broad classes, on the basis of the wavelength dependence. The nonwavelength-selective branching devices, called also couplers or splitters, share the input power to two or more fibers independently from the light wavelength. The components that perform the light distribution on more fibers on the basis of the different wavelengths of the input light are called a WDM (wavelength division (de)multiplexer).

Branching devices have three or more ports, bare fiber or connectors, for the input and/or output of optical power, and share optical power among these ports. A generic branching device can be represented as a multiport device having N input ports and M output ports, as shown in the Figure 9.

The optical characteristics of a branching device can be represented by a square transfer matrix of $n \times n$ coefficients, where n is the total number of the component ports. Each coefficient t_{ij} in the matrix is the fractional optical power transferred among designated ports, that is the ratio of the optical power P_{ij} transferred out of port j with respect to input

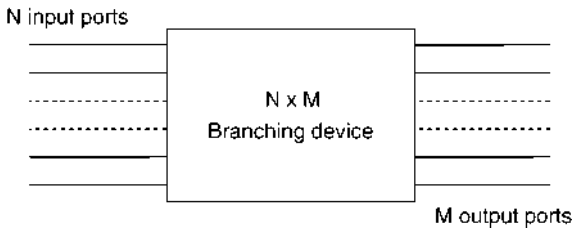

Figure 9 Scheme of a generic N × M branching device.

power P_i into port i:

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & & & \\ \vdots & & t_{ij} & \vdots \\ & & & \ddots \\ t_{n1} & & & t_{nn} \end{bmatrix} \quad [3]$$

where:

$$t_{ij} = \frac{P_{ij}}{P_i} \quad [4]$$

So, according to the relationship between two ports, the coefficients represent several parameters. For example, if the port i is an input port and j is an output port, t_{ij} is the transferred power signal through the component, while if they exist as two isolated ports the coefficients represent the crosstalk between the two ports.

The three main techniques to obtain this type of component are:

(i) All fiber components;
(ii) Micro-optics; and
(iii) Integrated optics.

Nonwavelength-Selective Branching Devices

A nonwavelength-selective branching device is a component having three or more ports, which shares an input signal among the output ports in a predetermined fashion. Usually these components are completely reversible and so they also work as a combiner of signals from more input ports onto a single output port.

There are several types of couplers for different applications:

- Y-coupler with one input port and two output ports;
- X-coupler with two input ports and two output ports; and

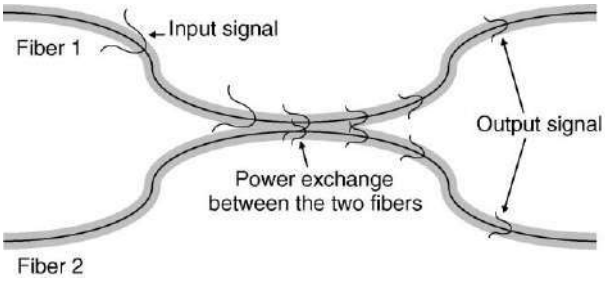

Figure 10 Scheme of a fusion 2 × 2 coupler.

- Star coupler with more than two ports in input and/or in output.

Moreover the couplers can be symmetrical, and in this case they divide the input power light onto n equal output flows, or asymmetrical devices in which the quantity of power light in the output ports is shared in a predefined nonuniform way.

The techniques used to carry out this type of device are mainly the fusion technique or planar technique. The first one is based on the phenomenon of evanescent wave coupling: when the core of two optical guides (as fibers) are located in close proximity, a power exchange from a guide to another is possible. This configuration can be obtained by stretching the fibers under controlled fusion (see Figure 10).

The planar technique, that is the simpler integrated optic application, is based on the optical guides with appropriate shapes being placed directly onto a silicon wafer by means of lithographic processes (Figure 11).

The fusion technique has the advantage of being 'all in fiber' so the component is ready to be inserted into the optical networks by means of usual joint techniques. But for a high number of ports, the coupling ratio among the branches cannot be controlled in a precise way and the cost of the fusion device is proportional to the number of the branches. On the contrary, a device obtained with the planar technology must be coupled with fibers normally used and this operation may be difficult due to the different geometry of the planar guide (rectangular) and the fiber (circular). On the other hand, this technology allows a high precision in the optical characteristics of the component and the cost of the device is independent of the branch number.

WDM

A wavelength-selective branching device, usually called WDM (Wavelength Depending Multiplexer),

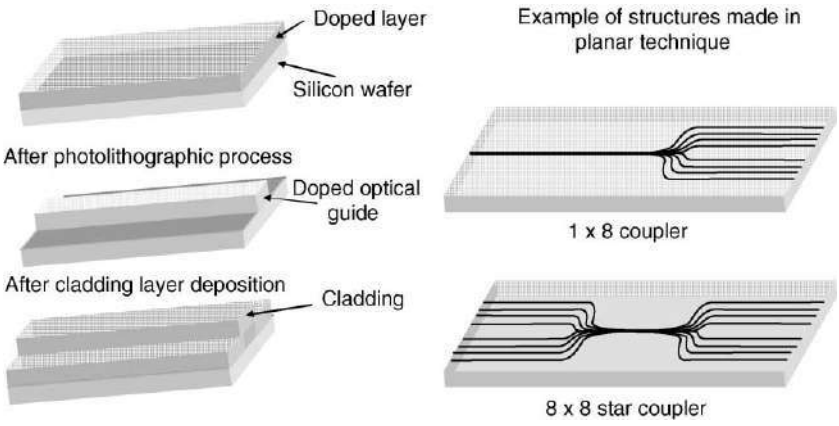

Figure 11 Scheme of the process to obtain a planar coupler and an example of a device.

is a component having three or more ports which shares an input signal among the output ports in a predetermined fashion, depending on the wavelength, so that at least two different wavelength ranges are nominally transferred between two different couples of ports.

WDM devices may be divided in three categories on the basis of the bandwidth of the channels (the spacing between two adjacent wavelength discriminated by the device):

(i) DWDM (Dense WDM) device operates for channel spacing equal or less than 1000 GHz (about 6–8 nm in the range of typical optical wavelength used in telecommunications systems, 1310 or 1550 nm);
(ii) CWDM (Coarse WDM) device operates for channel spacing less than 50 nm and greater than 1000 GHz; and
(iii) WWDM (Wide WDM) device operates for channel spacing equal or greater than 50 nm.

These components are used to combine, at the input side of an optical link in only one fiber, more optical signals with different wavelengths and separate them at the receiving end. This technology allows to send along a fiber, more communication channels, thus increasing the total bit rate transmitted. These types of components can be obtained using filters to select single a wavelength (as a diffractive grating that resolves the light into its monochromatic components) or using the phenomena of evanescent field coupling occurring between two adjacent fiber cores (as described for coupler devices), controlling in a precise way the coupling zone; this is, in fact, dependent on the wavelength (Figure 12).

Using wavelength filters, as they select a singular wavelength, it is necessary to perform a cascade filter structure for separating more than two wavelengths.

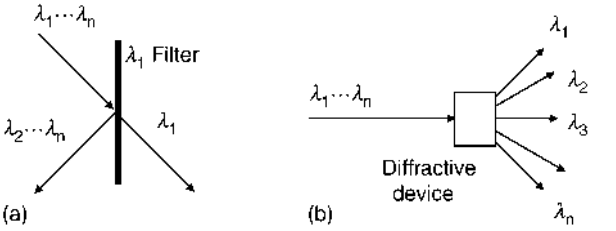

Figure 12 (a) Scheme of the filter technique and (b) diffractive device technique.

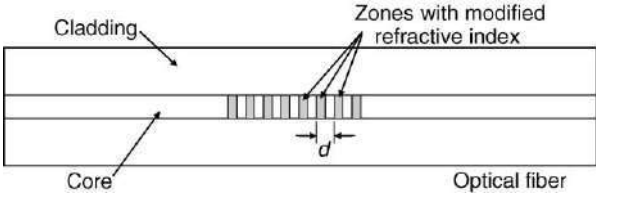

Figure 13 Scheme of FBG.

Micro-optic technologies are generally used for multimode fiber due to the critical collimation problem. For the single-mode fiber the evanescent field coupling technique is applied directly to fiber or to planar optical guides.

Another efficient technique to create WDM devices is with a fiber Bragg grating (FBG) wavelength filter directly applied to the fiber. FBGs consist in periodic modulation (see Figure 13) of the refractive index along the core of the fiber, and they act as reflection filters.

During the fabrication process (see Figure 14) the core of the optical fiber is exposed to UV radiation, and this exposure increases the refractive index of the core in defined zones, with a periodic shape obtained by the interference fringe created by crossing two coherent beams.

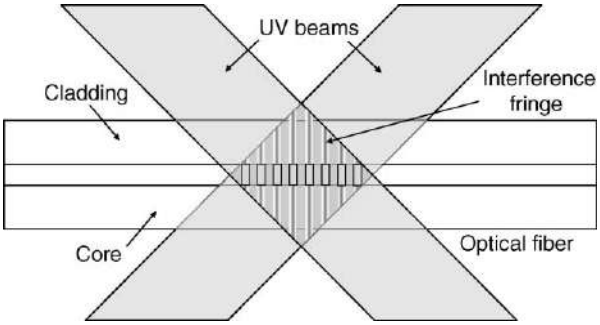

Figure 14 Scheme of FBG production.

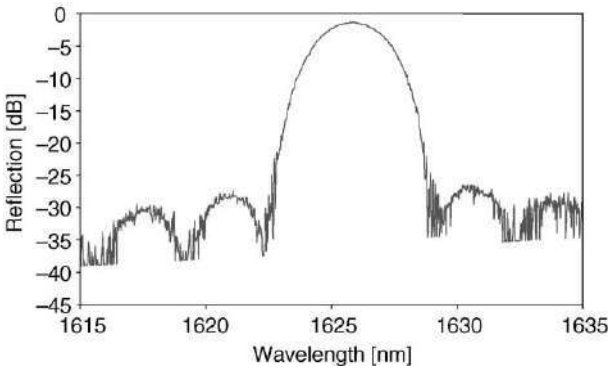

Figure 15 Transmission characteristic of FBG.

Figure 15 shows a representative spectral curve of a FBG filter centered at 1625 nm. The central wavelength, the width, and the value of reflection depend on the grating period (d) and on the deposited UV energy in the fiber core.

Other Passive Optical Components

In an optical network, besides the devices described above, there are a number of other passive optical components that perform particular functions on the optical signal; the main ones are:

Attenuators These are devices used to attenuate the optical power, for example, in front of a receiver when the optical line is short and the power of the received signal is still too high for the detector, or when it is necessary to equalize the signals arriving from different optical lines. It exists both in tunable and fixed versions. The first ones allow adjustment of the attenuation value, however, the second ones have a fixed and predefined value. The fixed optical attenuators can be obtained by inserting a filter or stressing a fiber in a controlled and permanent way. The tunable attenuators are based on variable optical parameters as, for example, the distance or the lateral offset in an optical connection. They exist both as connect lengths of fiber (attenuated optical jumpers) or as compact components similar to a double connector (plug stile attenuator).

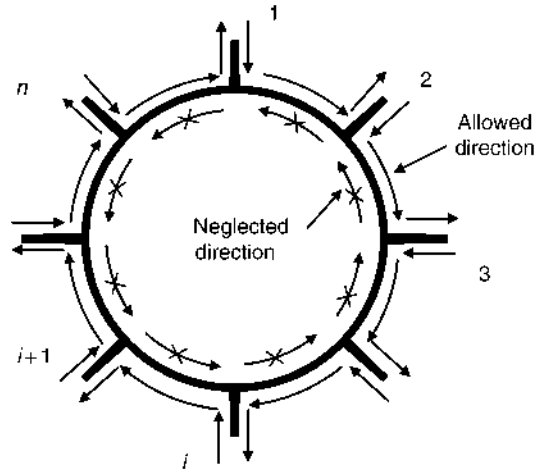

Figure 16 Principle scheme of an optical circulator.

Filters These are components with two ports employed to select or to filter some fixed wavelengths. They can be divided into the following categories:

- short-wave pass (only wavelengths lower than or equal to a specified value are passed);
- long-wave pass (only wavelengths greater than or equal to a specified value are passed);
- band-pass (only an optical window is allowed); and
- notch (only an optical window is inhibited).

It is also possible to have a combination of the above categories. The optical filtering can be obtained by inserting one or more filtering devices (as interferential films) into a fiber or creating a filter directly on the fiber by the FBG described above.

Isolators These are nonreciprocal two-port optical device in which the light flows in the forward direction, and not in the reverse one. The isolators are used to suppress

backward reflections along an optical fiber transmission line, while having minimum insertion loss in the forward direction. Fiber optic isolators are commonly used to avoid back reflections into laser diodes and optical amplifiers, which can make the laser and amplifiers oscillate unstably, and cause noise in the fiber optic transmission system.

Circulators These are passive components having three or more ports numbered sequentially $(1, 2, \ldots, n)$ through which optical power is transmitted nonreciprocally from port 1 to port 2,..., from port (i) to port $(i+1), \ldots$ and from port n to port 1 (Figure 16).

See also

Diffraction: Fresnel Diffraction. **Optical Communication Systems:** Wavelength Division Multiplexing.

Further Reading

Adam MJ (1981) *An Introduction to Optical Waveguides*. New York: John Wiley & Sons.

Blonder GE, *et al.* (1989) Glass waveguides on silicon for hybrid optical packaging. *IEEE Journal of Lightwave Technology* 7: 1530.

Digonnet M and Shaw HJ (1983) Wavelength multiplexing in single-mode fiber couplers. *Applied Optics* 22: 484.

Jeunhomme LB (1983) *Single-Mode Fiber Optics – Principles and Applications*. New York: Marcel Dekker.

Kashima N (1995) *Passive Optical Components for Optical Fiber Transmission*. Norwood, MA: Artech House.

Kashima N (1993) *Optical Transmission for the Subscriber Loop*. Norwood, MA: Artech House.

Miller CM (1986) *Optical Fiber Splices and Connectors – Theory and Methods*. New York: Marcel Dekker.

SC86B IEC (1999) *IEC 60874-1: Connectors for Optical Fibres and Cables – Part 1: Generic Specification*, 4th edn. Geneva: International Electrotechnical Commission.

SC86B IEC (1999) *IEC 61073-1: Optical Fibres and Cables – Mechanical Splices and Fusion Splice Protectors – Part 1: Generic Specification*, 3rd edn. Geneva: International Electrotechnical Commission.

SC86B IEC (1999) *IEC 60869-1: Fibre Optic Attenuators – Part 1: Generic Specification*, 3rd edn. Geneva: International Electrotechnical Commission.

SC86B IEC (2000) *IEC 60875-1: Non-wavelength Selective Fibre Optic Branching Devices – Part 1:Generic Specification*, 4th edn. Geneva: International Electrotechnical Commission.

SC86B IEC (2000) *IEC 62077-1: Fibre Optic Circulators – Part 1: Generic Specification*, 1st edn. Geneva: International Electrotechnical Commission.

SC86B IEC (2000) *IEC 61977-1 Ed. 1.0: Fibre Optic Filters – Generic Specification*, 1st edn. Geneva: International Electrotechnical Commission.

SC86B IEC (2000) *IEC 61202-1: Fibre Optic Isolators – Part 1: Generic Specification*, 1st edn. Geneva: International Electrotechnical Commission.

Technical Staff of CSELT (1990) *Fiber Optic Communications Handbook*, 2nd edn, TAB Books. New York: McGraw-Hill.

Yokohama I, *et al.* (1988) Novel mass-fabrication of fiber couplers using arrayed fiber ribbons. *Electronic Letters* 24: 1147.

FIBER GRATINGS

P S Westbrook, OFS Laboratories, Murray Hill, NJ, USA

B J Eggleton, University of Sydney, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

Fiber gratings are periodic index variations inscribed into the core of an optical fiber, and are important devices for manipulating fiber guided light. Since their first fabrication in the late 1970s, fiber grating use and manufacture has increased significantly and they are now employed commercially as optical filters, reflectors and taps in telecommunications systems, and as strain and temperature sensors in various industries. We review optical fibers and the fundamentals of fiber grating characteristics as well as fabrication. Finally we consider several of the major industrial and scientific applications of fiber gratings.

Fiber Modes

A fiber grating couples light between the modes of an optical fiber, and therefore a knowledge of these modes is essential to understanding the characteristics of fiber gratings. In what follows, we emphasize single-mode fibers used near 1550 nm because these are very important in most practical applications that

use fiber gratings. A single-mode fiber can be viewed as a thin cylinder of silica with a central raised index region of small enough diameter so that only a single mode is guided through total internal reflection at the core boundary. The electric field (E-field) of the fundamental mode propagating in the core is described by a characteristic transverse E-field profile and a propagation constant k:

$$E = E(r, \varphi)\mathrm{e}^{-\mathrm{i}(\omega t - kz)} \qquad k = 2\pi n_{\mathrm{eff}}/\lambda \qquad [1]$$

Here r and φ are the radial and azimuthal directions, z is measured longitudinally along the fiber, ω is the radial frequency, and λ is the vacuum wavelength, and the E-field points in the same direction everywhere. As the second equation indicates, the fiber mode propagates in the fiber as though it has an effective refractive index n_{eff}.

The silica cladding acts to isolate the core mode from the outside, hence the core mode penetrates very little into the cladding. However, if the outer cladding surface allows total internal reflection, the cladding may also guide light. Such 'cladding modes' are also important in understanding fiber gratings, since the grating couples the core mode with the cladding modes. As shown in Figure 1, the cladding modes extend outside the core into the cladding region, and their effective index is therefore always less than that of the core mode. We also note that in the case when the outer coating of the fiber does not support total internal reflection, 'leaky' cladding modes result. In the limit that there is no reflection at the cladding boundary a continuum of 'radiation modes' results. Although these modes are not guided, they are exploited in a variety of applications and can be carefully tailored by a suitably designed fiber grating. Note that the radiation and cladding modes exist regardless of the presence of a grating; a grating simply couples power from the core mode to the cladding/radiation modes; the grating is typically only a small perturbation on the index profile of the waveguide and the corresponding mode fields.

Fiber Grating Theory

Fiber gratings are longitudinal periodic variations in the refractive index (or, more generally, the dielectric function) of the core and/or cladding of an optical fiber. The scattering from any grating may be understood simply by considering the scattering off each successive period and adding these contributions while taking the phase of the E-field into account. Figure 2 illustrates this analysis in the case of a fiber Bragg reflector. When the reflections from each period of the grating add constructively, a strong back-reflection results. This is known as Bragg reflection. The condition for Bragg reflection is given by:

$$\lambda_{\mathrm{Bragg}} 2 n_{\mathrm{eff}} = \Lambda_{\mathrm{grating}} \qquad [2]$$

Here λ_{Bragg} is the vacuum wavelength of the light (equivalent to its frequency, $\omega = 2\pi c/\lambda$), n_{eff} is the effective index of the mode, and $\Lambda_{\mathrm{grating}}$ is the period of the grating refractive index modulations. This Bragg reflection is evident in a narrow dip in the

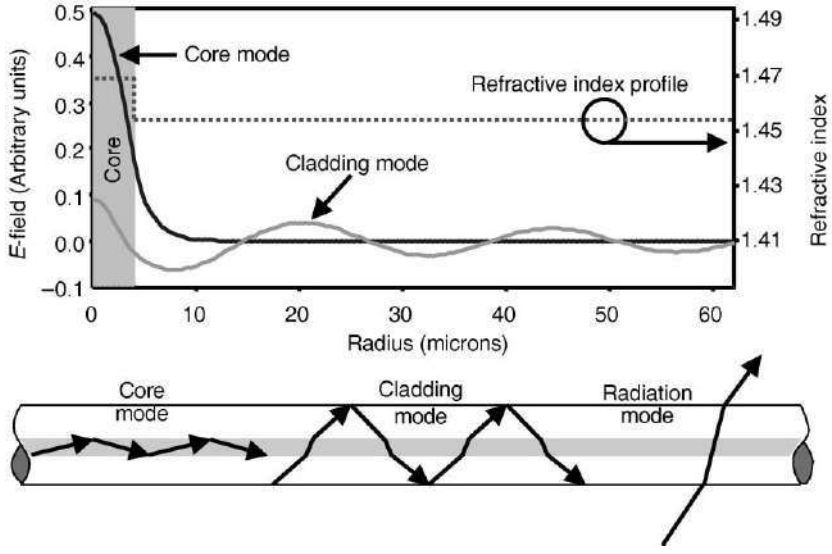

Figure 1 Fiber modes. Refractive index profile of a step index, single-mode fiber (dashed line). Radial profiles of mode E-fields of the fundamental (LP01) core mode, guided within the raised index region defined by the core, and one of the many higher-order cladding modes guided by the outer, air–silica interface. Also shown is a fiber indicating the ray paths for core, cladding, and radiation modes.

transmission spectrum of the grating (see Figure 3a). The ability of a fiber grating to selectively reflect the core light at one wavelength is one of its most important practical properties since it can be used as a narrowband filter or reflector. As a fiber component, the fiber grating has the advantage of compactness and provides in-line wavelength filtering.

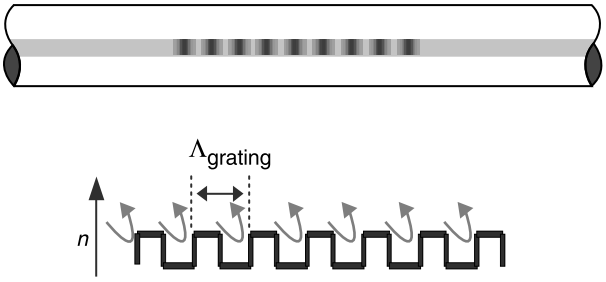

Figure 2 Bragg reflection. Each period of the grating reflects a small fraction of the incident light. When the period of the grating is adjusted so that these reflections add coherently, a strong Bragg reflection results and a narrow dip is observed in the transmission spectrum of the fiber (see Figure 3a). As shown in Figures 3 and 4, Bragg reflection can also occur into cladding modes.

The condition for Bragg reflection may be expressed more generally as a phase-matching condition. Phase matching refers to the condition when the phases of all the grating-scattered waves are the same and there is constructive interference. More specifically, this condition expresses the relationship between the spatial frequencies of the incident light, the scattered light, and the grating modulation, necessary for constructive interference in the scattered light. For a grating, the phase matching condition may be expressed in terms of the grating wavevector, $K_{grating} = 2\pi/\Lambda_{grating}$ and of the two mode propagation constants, $k_{incident}$ and $k_{scattered}$ (defined in eqn [1])

$$k_{incident} - K_{grating} = k_{scattered} \qquad [3]$$

Using phase matching, we can understand the various transmission spectra that result from fiber

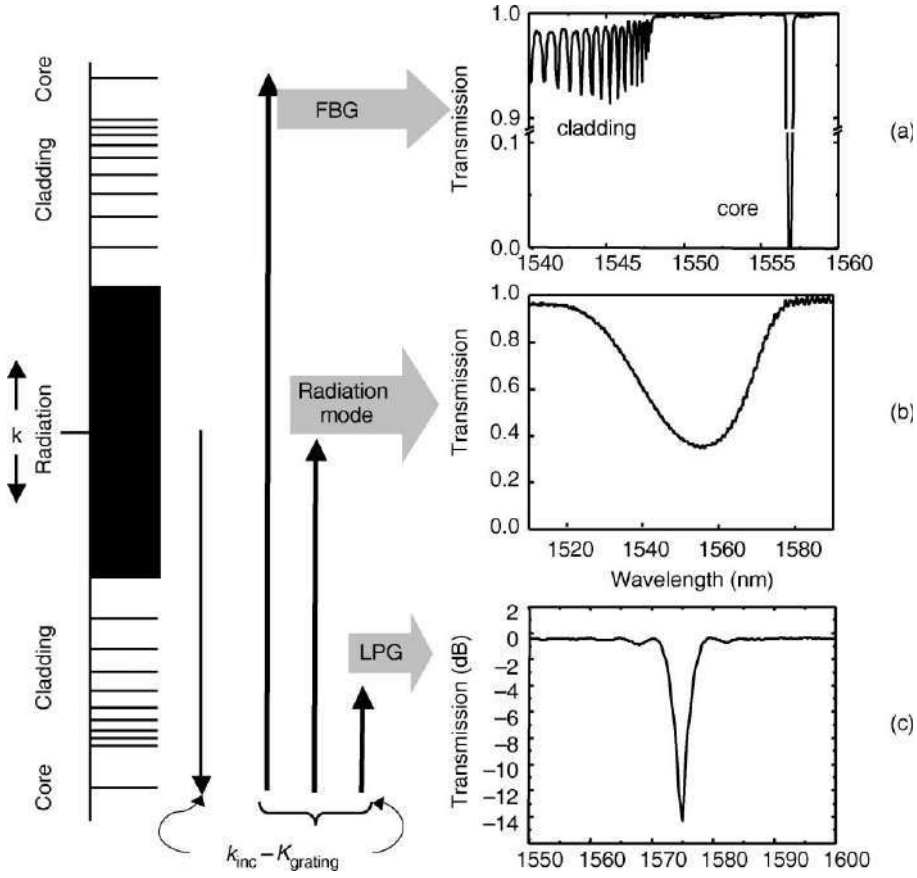

Figure 3 Phase matching. Fiber grating resonances occur when the phase matching condition is satisfied: $k_{inc} - K_{grating} = k_{scat}$. Propagation constants ($k = 2\pi n_{eff}/\lambda$) of the fiber modes are indicated as hash marks on the vertical line. The region indicates the regime of unguided radiation modes. The fiber grating is represented by its longitudinal spatial frequency (wavevector) $K_{grating} = 2\pi/\Lambda_{grating}$, where $\Lambda_{grating}$ is the period of the grating. Assuming an incident forward-going core mode, the grating wavevector then phase matches to (a) counter-propagating core and cladding modes (fiber Bragg grating, FBG), (b) continuum of radiation modes, or (c) a copropagating cladding mode (long-period grating, LPG).

gratings. Figure 3 shows the transmission characteristic observed for core guided light in a single-mode fiber containing three different types of gratings: Bragg reflector, radiation mode coupler, and copropagating mode coupler. Each grating has a different type of phase matching condition. Figure 3 also depicts the phase matching condition in eqn [3] graphically. The vertical line contains a mark representing the propagation constants, k, of modes of a single mode fiber (both positive and negative for the two propagation directions) and a solid black region representing the radiation mode continuum. Bragg reflection into backward-propagating core and cladding modes occurs for short grating periods ($K_{\text{grating}} \sim 2k_{\text{core}}$). Such gratings are known as fiber Bragg gratings (FBGs) and a typical transmission spectrum with a strong Bragg reflection and several cladding mode resonances is shown in Figure 3a (a typical Bragg reflection spectrum is shown in Figure 8 below; the FBG grating period was ~500 nm). Core-cladding mode reflection results in loss resonances for wavelengths shorter than the core–core Bragg reflection. When the grating period is very long (in general, $K_{\text{grating}} < k_{\text{core}}(n_{\text{cladding}} - 1)/n_{\text{cladding}}$; typically Λ_{grating} is ~1000 times that of an FBG), phase matching to copropagating cladding modes occurs. Such gratings are known as long-period gratings (LPGs). A typical LPG spectrum is shown in Figure 3c. The LPG resonance corresponds to coupling between the core mode and a single copropagating cladding mode. For wavevectors in between these two regimes, coupling to the continuum of radiation modes is possible. An example of such a grating spectrum is shown in Figure 3b. We note that fiber cladding modes can be either well guided (resulting in the sharp grating resonances of Figure 3a) or poorly guided (resulting in a broad continuum of Figure 3b) depending on the fiber surface. The radiation mode approximation is valid in the limit that the cladding mode is very poorly guided (i.e., has high loss) within the grating region.

While phase matching yields the resonance wavelengths of a fiber grating, the strengths of these resonances must be derived from a full solution of Maxwell's equations. The most common approximation used in these calculations is known as coupled mode theory (CMT). In CMT, the longitudinal refractive index variation is assumed to be a small perturbation to the waveguide modes, which results only in a change in the propagation constant of the modes but does not change the transverse dependence of the mode E-fields (see eqn [1]). CMT then assumes that the E-field amplitude of each mode varies slowly (compared to wavelength) along the grating. A pair of coupled equations relating the two mode amplitudes as a function of position along the grating can then be derived, and these allow for computation of the grating reflection and transmission. The coupling between two modes is determined by an overlap integral κ_{12} between the product of the E-fields of the two modes (E_1 and E_2) and the periodic refractive index variation:

$$\kappa_{12} \propto \int\int_{\text{Area}} E_1(r,\phi)E_2(r,\phi)\delta n(r,\phi)\,\mathrm{d}A \qquad [4]$$

where $\delta n(r,\phi)$ is derived from the refractive index modulation:

$$\delta n(r,\phi,z) = \delta n(r,\phi)\cos(K_{\text{grating}}z) \qquad [5]$$

This overlap integral is useful because it allows for a rough comparison of the strength of resonances for different modes without the need to obtain a full CMT solution. In the simple case of Bragg reflection of the core mode in a standard single-mode fiber (see Figure 1):

$$\kappa = \frac{\pi\delta n\eta}{\lambda}, \qquad [6]$$

where $\eta \approx 0.8$ is the overlap integral of the core mode intensity with the core.

To understand fiber grating spectra we first discuss a 'uniform' Bragg grating, of length L. Such a grating has uniform index modulation along its length and is shown in Figure 4a. Depending on the length and refractive index modulation of the grating, it may be classified as either weak or strong. The division between these two regimes may be understood by considering the light scattered by each successive period of the grating. The fraction of light scattered from each period can be approximated by $\delta n/2n$ using the Snell's law reflectance formula. The boundary between a strong and a weak grating occurs when the grating is either long enough or strong enough that the sum of the scattering from all periods is of order 1. This division may also be expressed in terms of κ and L: $\kappa L \sim 1$.

In the weak limit, the incident light propagates through the grating with very little scattering. In this case, both the LPG and FBG resonances have transmission spectra that are approximately related to the Fourier transform of the grating index profile. (For LPGs, this may not hold if modal dispersion is large; see the discussion below.) In the case of a uniform grating profile, this results in a $1 - \text{sinc}^2$ wavelength dependence of the transmission spectrum and a spectral width $\delta\lambda/\lambda \sim \Lambda_{\text{grating}}/L = 1/N_{\text{periods}}$ as shown in Figure 4a. An important feature of the uniform grating is the side lobes. The side lobes result from the sudden 'turn on' of the grating. In the FBG,

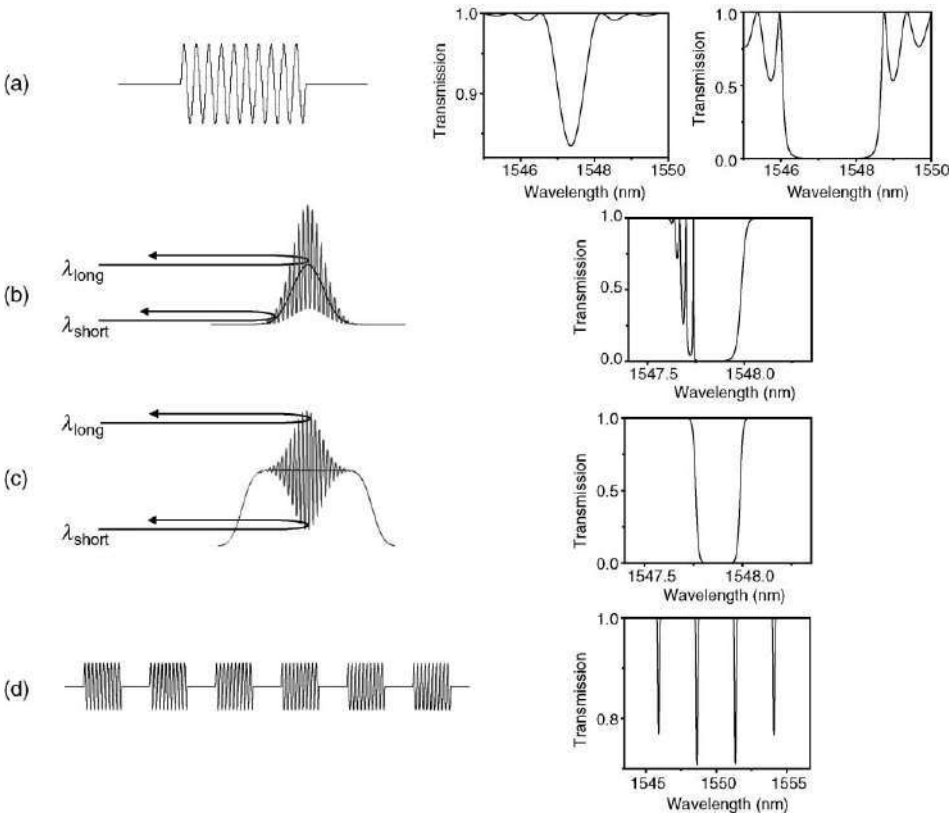

Figure 4 Fiber Bragg grating refractive index profiles and spectra. (a) A uniform profile with transmission spectra for both strong and weak index modulations. (b) An 'unapodized' Gaussian grating profile showing Fabry–Perot-like structure within the grating resonance. (c) A '100% dc apodized' Gaussian grating profile showing a symmetric transmission spectrum. (d) A superstructure grating exhibiting four resonance peaks.

they can be understood as resonances of the effective Fabry–Perot cavity defined by the sharp grating boundaries.

In the strong limit, a substantial fraction of light is scattered out of the incident mode, and the FBG and LPG transmission spectra become different. As shown in Figure 4a, the FBG spectrum widens to a width determined only by δn: $\delta\lambda/\lambda = \delta n/n$. The incident light penetrates only a short distance ($\sim \lambda/\delta n$) into the grating before being completely reflected. The strong LPG, on the other hand, becomes 'overcoupled' as the second mode is reconverted into the incident mode. The result is a higher transmission for the incident core mode (provided that the scattered mode propagates without loss).

While uniform gratings are easy to model and manufacture, most practical fiber gratings have a nonuniform profile. In a nonuniform grating profile, both the refractive index modulation amplitude, $\delta n_{AC}(z)$, and the average effective index value, $\delta n_{DC}(z)$, may vary along the fiber grating length:

$$\delta n(z) = \delta n_{DC}(z) + \delta n_{AC}(z)\cos(K_{grating}z) \qquad [7]$$

By controlling the ac and dc index change, a greatly improved filter may be fabricated. Figure 4b shows the index profile and the grating transmission spectrum of a fiber Bragg grating with an 'unapodized' Gaussian profile. Note that due to the smooth increase in modulation amplitude, the sidebands of the uniform grating have been eliminated. However, because the dc index (and hence Bragg condition) is not constant within the grating, sharp resonant structure is observed within the grating resonance. This structure can be understood as Fabry–Perot resonances in the cavity formed by the nonuniform Bragg condition within the grating. Figure 4c shows a grating with a '100% dc apodized' Gaussian profile. This grating has a constant dc refractive index within the grating and therefore the resonance is smooth and symmetric. Most useful bandpass filters require a profile close to 100% apodization. Although not shown in Figure 4, a smoothly varying index profile also eliminates the sidelobes in LPG resonances.

Other important types of nonuniform fiber gratings include chirped fiber gratings and superstructure fiber gratings. In a chirped fiber grating the grating period

slowly changes along the length of the grating. As discussed below, chirped fiber gratings may be used as chromatic dispersion compensators. They may also be used as broadband reflectors. LPGs may also be chirped to produce a broadband response. Superstructure gratings have oscillations in the period and/or amplitude of the refractive index modulations. The resulting grating can have several resonance peaks (see Figure 4d). The properties of nonuniform gratings may be computed using CMT and also understood intuitively using 'band-diagrams' and an effective index model of the fiber grating response.

Grating spectra showing radiation mode coupling are unlike FBGs or LPGs because there is no conversion of the radiation modes back to the incident core mode. The sharp resonances smooth out into a broad continuum. The grating transmission spectrum depends less on the refractive index profile of the grating than on the overlap of the core mode and grating with the radiation modes. Radiation mode coupling is greatly affected by tilting the grating planes with respect to the axis of the fiber. As discussed below, such tilted gratings may be designed as loss filters and optical fiber taps. Grating tilt has also been applied to decrease core mode back-reflection.

LPG resonances (and to some degree FBGs) may also exhibit strong dependence on the modal dispersion (or variation of the propagation constant, k, with wavelength). This wavelength dependence can dominate the wavelength dependence arising from CMT, making the resonance more or less narrow than predicted by CMT without modal dispersion. In an extreme case, very broadband LPG resonances (40 nm) may be produced between the core mode and a specially designed higher-order mode whose propagation constant varies such that it is resonant with the core mode over a large bandwidth.

Fiber Grating Fabrication

Photosensitivity

The key discovery leading to the study and practical application of fiber gratings was the phenomenon of UV-induced photosensitivity in the conventional germanosilicate optical fibers used in telecommunications systems. UV irradiation of the germanosilicate core region can produce refractive index changes in the range 0.01 to 0.0001. Such refractive index changes are large enough to produce useful filters and reflectors. The refractive index change is most efficiently produced by irradiation close to 242 nm or around 193 nm, where germanosilicate glass has strong absorptions. Numerous other dopants have been shown to be photosensitive at wavelengths ranging from 484 nm to 155 nm. These include P, Ce, Pb, Sn, N, Sn-Ge, Sb, and Ge-B. Other dopants, such as Al and F, show very little photosensitivity.

A second major discovery enabling practical manufacture of fiber gratings was the effect of 'hydrogen loading'. The fiber is placed in a pressurized hydrogen atmosphere until molecular hydrogen has diffused into the core region to a level of a few mole percent (a level similar to that of the photosensitive dopants). UV irradiation of the resulting fiber produces index changes an order of magnitude larger than in the unloaded fiber. Hydrogen loading allows gratings to be easily imprinted in conventional single-mode fibers, whose Ge content is too low to allow for strong gratings. The increase in photosensitivity is true for most photosensitive dopants. Index changes in excess of 10^{-2} have been produced with hydrogen loading. In order to avoid increased OH absorption in the telecommunications band (1520–1630 nm), deuterium (D_2) is normally used in place of hydrogen.

The origins of UV photosensitivity are not fully understood; however, at least two mechanisms are agreed to contribute to the index change. Firstly, as is evident from Figure 5, the UV absorption spectrum is modified during UV exposure. This bleaching corresponds to a change in refractive index in the infrared which may be computed using the Kramers–Kronig relations. In particular, the absorption band at

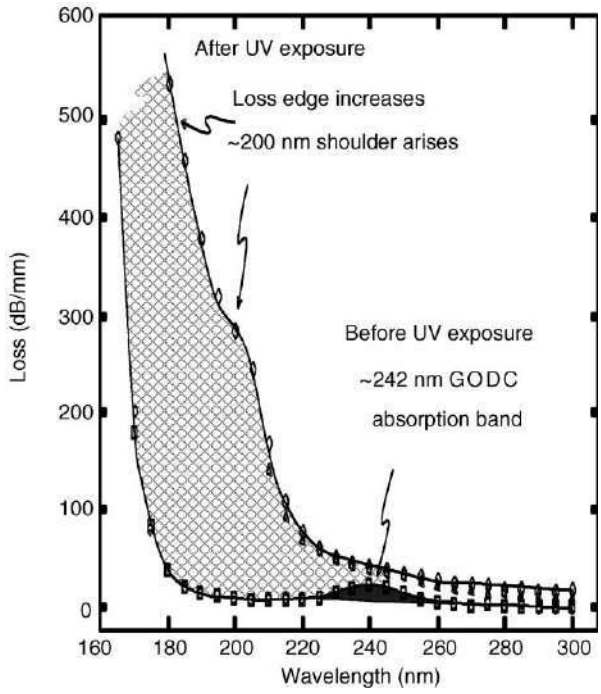

Figure 5 UV absorption spectrum of Ge-doped silica before and after irradiation with 242 nm light, showing bleaching of the Ge-oxygen deficient defect feature at 242 nm.

242 nm resulting from germanium oxygen deficient centers is bleached, indicating that these defects are modified during exposure. The infrared index change results both from this bleaching and from the change in the UV absorption edge at ~190 nm. Secondly, absorption of UV causes the silica matrix to contract. This compaction and the resulting stress distribution also change the refractive index.

Both mechanisms are enhanced by the presence of hydrogen in the silica lattice. The hydrogen reacts with oxygen, thus increasing the number of UV-absorbing germanium oxygen-deficient centers. As a result, in hydrogen loaded fibers, more of the Ge atoms participate in the index change, explaining the large increase in photosensitivity. The number of oxygen-deficient defects may also be increased during fiber manufacture by collapsing the fiber preform in a low-oxygen atmosphere. Such fibers show increased photosensitivity without the need for hydrogen loading. Co-doping a Ge-doped fiber with boron has also been shown to increase the level of photosensitivity of a fiber for a given level of index change in the core.

An important aspect of UV-induced index changes is that they require thermal annealing to ensure long-term stability. Sufficient exposure to elevated temperatures can completely erase the refractive index change in some Ge-doped gratings. Gratings can be completely erased by several hours of heating to greater than 500 °C. Thermal stability has been described by an activation energy model which assumes a distribution of energies for the defects giving rise to the refractive index change. Therefore, at a given anneal temperature a fixed fraction of the index change will quickly decay, leaving only stable defects that easily survive at lower temperatures. By annealing at several temperatures for a fixed time, a relationship between temperature and time required for a given refractive index decay may then be derived and used in accelerated aging tests of fiber gratings. For 20-year grating reliability of ~1% decrease in UV-induced refractive index change, a typical Ge-doped grating requires annealing at 140 °C for 12 hours.

Grating Inscription

The first method used in UV grating inscription is known as 'internal writing'. In this method light propagating in the fiber causes a standing wave to form within the core, and UV photosensitivity from a two-photon process results in permanent grating formation. Both LPG and FBG resonances have been fabricated using this method. The technique is inflexible, however, because the grating parameters are not easily controlled.

Another important advance that made possible the widespread use of fiber gratings is the 'transverse side-writing' technique. In this method, a UV interference pattern is projected through the side of the fiber onto the core region. The first demonstration of this technique employed a free space interferometer to produce the UV fringes. A significant improvement on this technique employs a zero-order nulled phase mask to generate the interference pattern. The phase mask acts to split the incident beam, and these two then form the interference pattern. The method is shown in Figure 6. The advantage of the phase mask

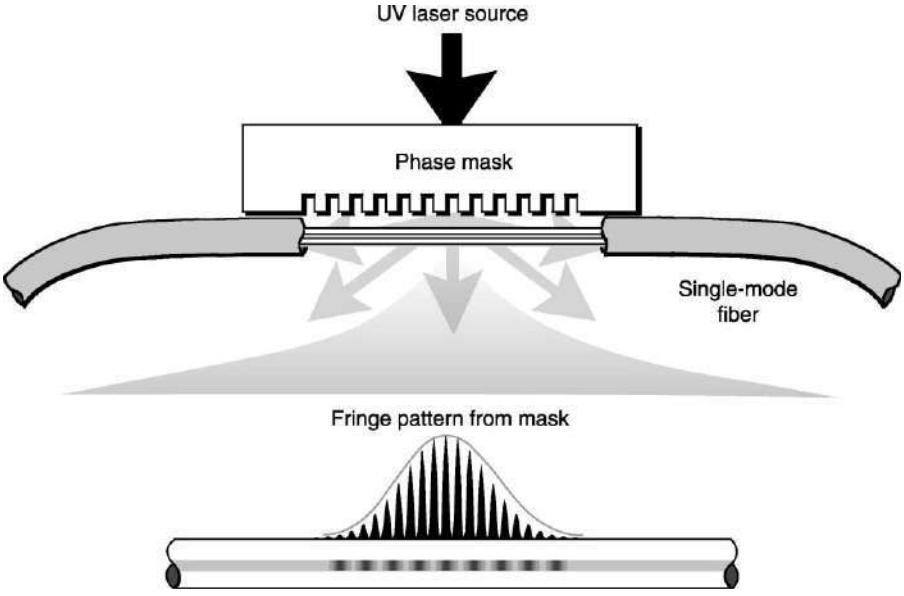

Figure 6 Phase mask fabrication of a fiber grating.

technique is that the fringes are very stable, since the UV beams propagate only a short distance. Another important technique for writing fiber gratings is the 'point-by point' writing method. In a point-by-point technique the UV interference pattern is modulated and translated along a fiber in such a way that a long grating may be formed. Although complex and sensitive, such methods can yield gratings as long as a few meters and also allow for the fabrication of complex grating profiles. Point-by-point methods have been used to fabricate broadband chirped dispersion compensating gratings and Bragg reflectors with very low dispersion.

LPGs may also be fabricated with UV irradiation. An amplitude mask is typically used. Requirements for beam coherence are greatly reduced because of the long period (typically a few hundred microns). Other techniques may also be used to fabricate LPGs. The fiber may be periodically deformed by heating or with periodic microbends. Dynamic LPGs may also be produced by propagating acoustic waves along the fiber.

Applications

Fiber gratings are key enabling technologies in a number of important applications. In this section we review these, highlighting the unique functionalities that can be provided by fiber gratings.

Fiber Grating Strain and Temperature Sensitivity

Fiber gratings are sensitive to both strain and temperature and may therefore be used as sensors. Both LPG and FBG resonances change with strain and temperature. LPGs show greater but more complicated sensitivity and typically FBGs are used in such applications. Strain sensitivity arises from the fact that gratings are extended in length. Because fibers are very thin (diameter $\sim 125\ \mu m$), significant extension is possible. A smaller strain sensitivity arises from the stress optic effect. Typical sensitivity at 1550 nm is ~ 1 pm/1 $\mu\varepsilon$ in FBGs. Temperature sensitivity arises from temperature dependence of the refractive index and gives rise to a change order 0.011 nm/°C of temperature change of the FBG. FBG strain sensors have been applied to map stress distributions and can be employed in a distributed manner with many Bragg reflectors in one fiber being examined with one detector.

These sensitivities may also be used to make fiber grating filters tunable. Temperature changes are more stable but typically give 1–2 nm of change in an FBG resonance. Strain tuning is more difficult to package commercially but can give up to 4 nm in extension and more than 10 nm in compression. Passive thermal stabilization of grating resonances may also be achieved by designing packages in which strain and temperature variations cancel each other.

Fiber Grating Stabilized Diode Lasers and Fiber Lasers

One of the first important commercial applications of fiber gratings was in stabilization of fiber coupled diode lasers. Normally these lasers operate on several cavity modes making the output unstable. If feedback is provided in the form of a narrowband reflection then the laser output will be only in this narrow bandwidth. The fiber grating is implemented in the fiber pigtail of the diode laser and typically provides a few percent reflectivity over a bandwidth of ~ 1 nm. The application is depicted schematically in Figure 7.

More generally, fiber gratings have been used to realize fiber laser cavities and fiber distributed feedback lasers. One important example is the cascaded Raman resonator fiber laser, in which several FBGs recycle Raman pump light to produce high-power, fiber coupled light at any design wavelength.

Optical Add/Drop Multiplexers

Another important commercial FBG application is the optical add/drop multiplexer (see Figure 8). These are key components in wavelength division multiplexed (WDM) networks. WDM networks transmit large amounts of data through a single fiber by modulating several optical carrier wavelengths and sending all the wavelengths through a single fiber. The optical add/drop multiplexer allows individual wavelengths to be added or dropped from a given fiber.

Fiber Bragg gratings are useful in this application because of the very narrow resonances that are achievable (typically a few tenths of a nm at 1550 nm). For wavelengths close to the Bragg wavelength an FBG strongly reflects. The reflectivity and bandwidth are determined by the depth of the index modulation and grating length, as discussed earlier. These degrees of freedom make fiber gratings ideal filters to be used in wavelength division multiplexed lightwave networks, in which multiple channels are transmitted along long lengths of optical fiber. Figure 8 illustrates schematically the principle of

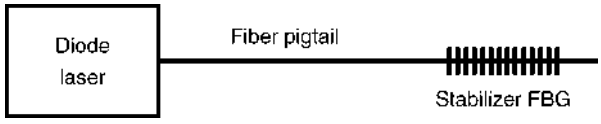

Figure 7 Fiber Bragg grating stabilized diode laser. The fiber grating reflector in the laser pigtail provides a small level of feedback to stabilize the laser output power and wavelength.

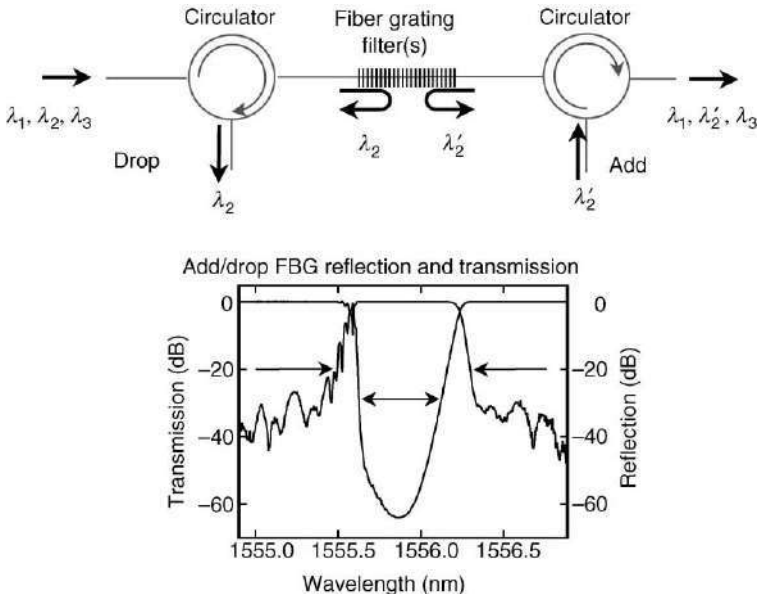

Figure 8 Top: Add/drop node employing fiber grating reflector. Bottom: characteristic reflection and transmission spectrum of an add/drop fiber grating filter.

a fiber grating based add/drop multiplexer. The grating is located between two three-port optical circulators. (A circulator is a bulk optic element that allows transmission only from port n to port $n+1$.) Different channels are incident upon the grating from the left and only one particular channel undergoes reflection at the grating. Note that by carefully designing the grating properties the reflectivity in the adjacent channel can be made to be negligible. This one channel is reflected and re-routed by the three-port circulator. The remainder of the channels are transmitted through the grating. The grating also enables the 'add' functionality. In this case light originating from a different part of the optical network enters the link again through an optical circulator and is reflected off the fiber grating.

Gain-Flattening Loss Filters

A primary strength of fiber gratings is the diversity of filter shapes that is achievable through modification of the grating profile. Gain-flatteners are an important application requiring such filter design. Gain-flattening filters are important in WDM telecommunications systems in order to smooth the wavelength dependence of fiber amplifiers (such as Er or Raman amplifiers). Both LPGs and FBGs may be employed as gain flattening filters. The desired loss spectrum is translated into a corresponding grating profile. LPGs have the advantage of low return loss, but the LPG resonance wavelengths and strengths are very sensitive to the fiber geometry, making the computation of the grating profile less accurate. FBGs may also be used in both tilted and untilted form; however they typically have back-reflections that require careful fiber grating growth or an isolator to suppress back-reflections. Figure 9 shows an example of a gain equalizing loss filter using LPGs.

Dispersion Compensator

Chromatic dispersion can limit the transmission distances and capacities of long-haul fiber links. Chromatic dispersion arises from the variation in propagation velocity with wavelength. An initially sharp pulse is thus dispersed and begins to overlap with adjacent pulses resulting in degraded signal quality. A key technology for long-haul fiber links is therefore 'dispersion compensation'. The most common solution has been 'dispersion compensating fiber' (DCF), which is simply fiber in which the chromatic dispersion has been engineered to be the exact opposite of the dispersion in the fiber link. Limitations of DCF include nonlinearities, higher-order dispersion, and lack of tunability. Fiber gratings offer an alternative approach that overcomes the problems.

Fiber Bragg gratings can provide strong chromatic dispersion when operated in reflection with a three-port circulator (see Figure 10). In this case the grating period is made to vary along the length of the grating. This means that different wavelengths in a pulse reflect at different positions in the grating. This gives rise to wavelength-dependent group delay and therefore chromatic dispersion. The advantage of the fiber grating compared to DCF is that the device can be made to be very compact, with potentially lower

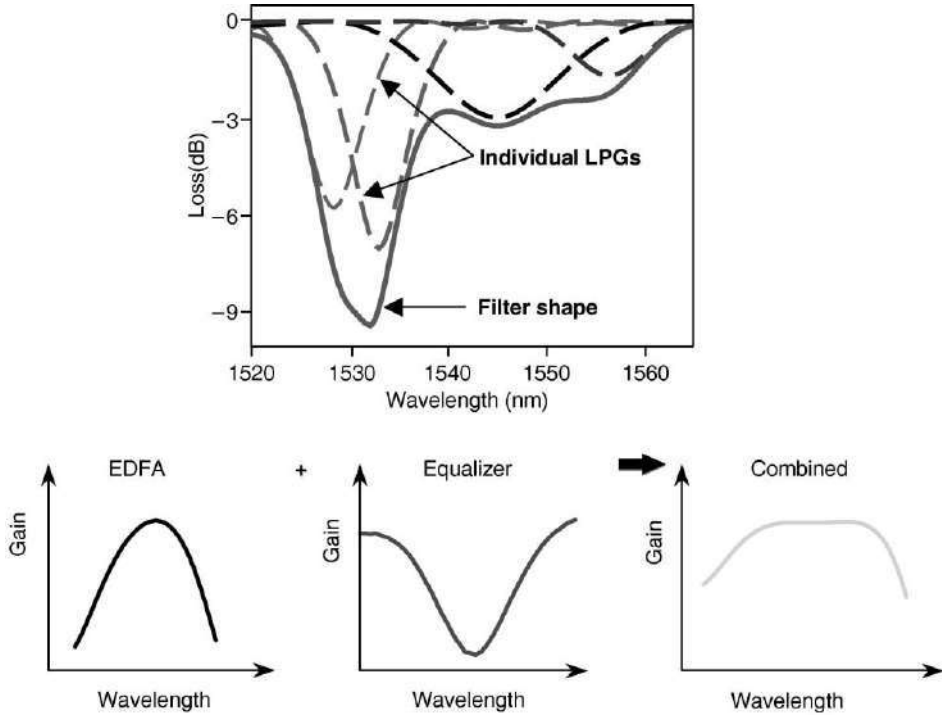

Figure 9 Long-period grating gain equalizer. Several individual gratings (dashed line) are combined to produce the overall shape (solid line).

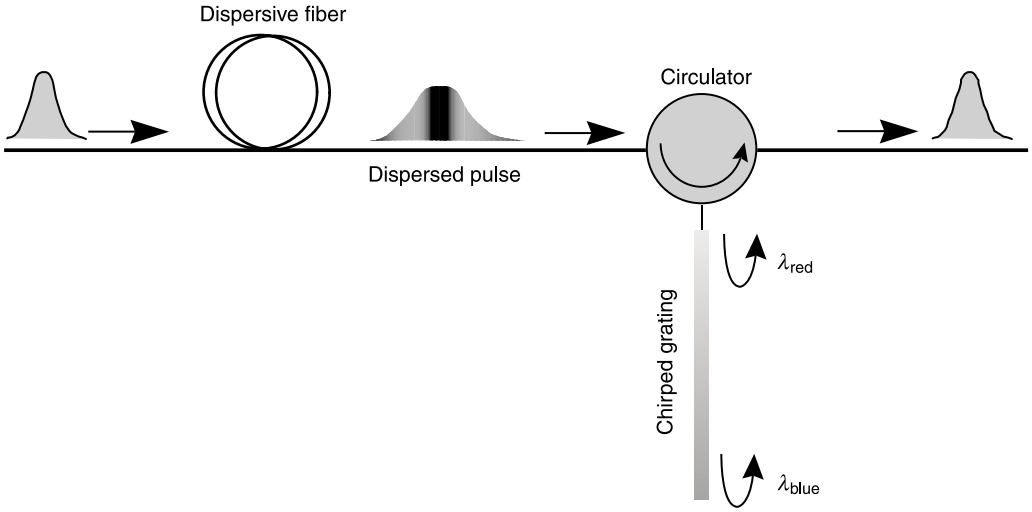

Figure 10 Chirped fiber grating dispersion compensator.

insertion loss, reduced optical nonlinearity, and controllable optical characteristics. In particular fiber gratings offer the possibility of engineering the chromatic dispersion to allow for compensation of dispersion slope and other degradations such as polarization mode dispersion.

Optical Monitors

Fiber gratings may also be employed as compact optical taps coupling core guided light to the radiation modes propagating outside the fiber (see Figure 3b). When a conventional Bragg grating is tilted with respect to the fiber axis it can scatter light out of the fiber in a highly directional manner. The angle of scatter is also wavelength dependent because of the phase matching condition discussed above. This wavelength dependence may form the basis of a compact wavelength monitor. The fiber grating is used to simultaneously couple light out of the fiber and separate the wavelengths. A simple

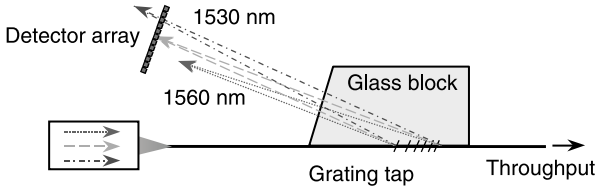

Figure 11 Fiber grating wavelength monitor. The fiber grating acts to scatter light out of fiber and simultaneously disperse the light.

detector array can then measure a spectrum of all the channels in a typical WDM system (see Figure 11).

Fiber grating taps may also scatter light in a polarization sensitive manner. This may be achieved by tilting the fiber grating by 45° with respect to the fiber axis. The resulting broadband taps may be employed in compact polarization monitoring schemes in which detectors are placed on the outside of the fiber to detect light at specific polarizations.

Advanced Topic in Fiber Gratings

Fiber Gratings in Air–Silica Microstructured Optical Fibers

Air–silica microstructure fibers (ASMFs), also known as 'holey' fibers, are fibers with air regions running along their length and allow for more complicated modal properties than conventional silica-doped fiber. It has been shown that fiber gratings can be UV written into photosensitive regions of various types of microstructure fibers. ASMF gratings open new possibilities for fiber grating devices. In addition, they are useful in studying the guidance properties of ASMFs, because they can selectively excite higher-order modes of the fiber. The phase matching condition may be used to derive the effective indices of the higher-order modes and mode profiles may be recorded by imaging a cleaved surface. Such measurements agree well with standard beam propagation computations. Figure 12 shows an example of a mode field image generated with a FBG.

One important application of ASMF gratings is a widely tunable LPG filter. Such filters employ a 'hybrid waveguide' formed by infusing a polymer into the air regions. Because the polymer index is much more sensitive to temperature than silica, the guidance properties of the cladding modes may be tuned. Thus, widely tunable LPG filters may be formed. Moreover, because the cladding modes are confined within an 'inner cladding' such filters are insensitive to external index, allowing for instance, the use of on-fiber, thin film heaters for power-efficient temperature tuning.

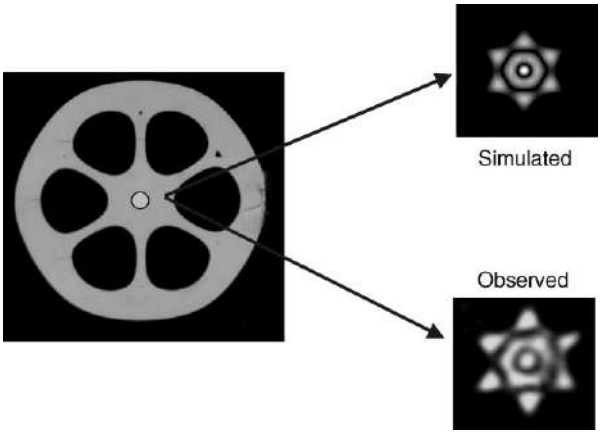

Figure 12 Cross-section of an air–silica microstructure fiber along with measured and simulated modes guided by the air regions.

Nonlinear Optics in Fiber Gratings

The linear properties of fiber gratings are well understood and as explained earlier lead to a number of interesting phenomena and applications. When the peak intensity of the optical field is very high a new class of nonlinear effects can occur. These optical nonlinearities arise from the intrinsic nonlinearity of the material system in which the grating is inscribed. In the case of optical fibers, the dominant nonlinearity is third-order nonlinearity which manifests itself in the refractive index of the material depending on the peak intensity of the optical field: $n = n_0 + n_2 I$. This nonlinearity gives rise to a number of dramatic nonlinear effects in fiber gratings, including all-optical switching, bistability, multistability, soliton generation, pulse compression, and wavelength conversion.

To understand these effects we consider an intense optical field incident upon a fiber grating with a central wavelength tuned inside the stopband of the grating where the reflectivity is high. At low intensities the pulse is reflected and only a small amount of light leaks through the grating. As the intensity of the incident optical field is increased the optical nonlinearity becomes significant and results in shifting of the local Bragg wavelength of the grating. This in turn can lead to an increase in transmission of light through the grating and bistable operation.

Theoretical studies of the continuous wave nonlinear properties of periodic media show that such media can exhibit switching as well as bistability and multistability. This prediction has been confirmed

experimentally in a variety of geometries, such as in semiconductors, and in colloidal crystals. Experiments studying nonlinear effects in gratings can also be performed in optical fiber gratings. In these experiments short optical pulses are typically used in order to obtain the very high intensities. This has several key advantages. The first of these is that long, high-quality Bragg gratings can be permanently imprinted into the core of optical fibers by side illumination with interfering ultraviolet beams (see above). This allows for very long interaction lengths (compared to picosecond pulse lengths) and enables nonlinear propagation phenomena to be studied in detail. These include a number of propagation phenomena, such as the formation and propagation of Bragg solitons, modulational instability, and pulse compression. The second advantage is that although the nonlinearity is weak, it has a response time of the order of a few femtoseconds, and thus acts essentially instantaneously on the scale of all but the shortest optical pulses. Finally, the ability to engineer the fiber grating offers a number of unique geometries and potential applications of nonlinear fiber gratings.

A key idea in the interpretation of nonlinear optics in fiber Bragg gratings is the Bragg soliton. Briefly, in these solitons the enormous dispersion provided by the fiber Bragg grating is balanced by the fiber nonlinearity. This is analogous to the optical soliton that has been demonstrated in standard single-mode fiber where the dispersion is predominantly associated with the intrinsic material properties of silica. Since the dispersion of a Bragg grating can be five orders of magnitude larger than that of unprocessed fiber, Bragg solitons can be observed in gratings of only a few centimeters in length, thus rivaling the compactness of other soliton geometries, such as spatial solitons. Of course this implies that the required optical intensities to generate Bragg solitons are correspondingly higher than in uniform media.

Figure 13 shows an example of experimental results when the peak intensity of a pulse incident upon a fiber grating was approximately 10 GW/cm^2 and the incident pulse is tuned to the anomalous dispersion regime of the grating. In particular, Figure 13 shows the transmitted intensity versus time for different values of detuning. By varying the strain on the grating, we can change the detuning of the input pulse, and thus tune to different points on the dispersion curve. Note also from Figure 13 that far from the photonic bandgap edge, where the grating-induced dispersion is negligible, the pulse is unaffected and propagates through the grating at the speed of light. Closer to the edge of the photonic bandgap the pulse is substantially compressed due to the formation of a Bragg soliton. In addition to this compression, the transmitted pulse is substantially delayed in time as a result of the reduced group velocity of the pulse propagating through the grating.

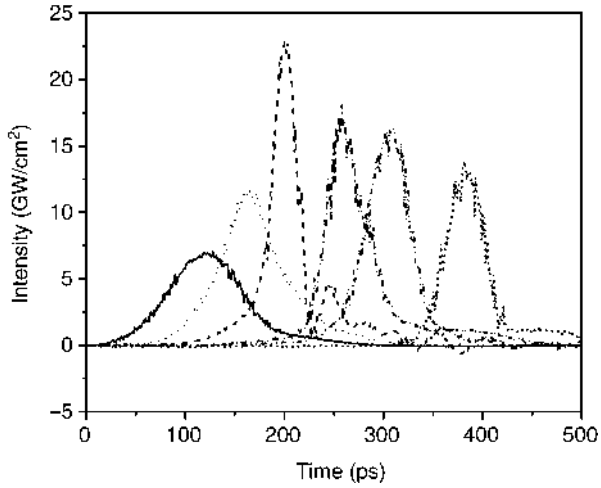

Figure 13 Pulse compression associated with nonlinear pulse propagation through a fiber Bragg grating. Compression is associated with the formation of a Bragg soliton when the incident pulse is tuned just outside of the photonic bandgap formed by the grating.

List of Units and Nomenclature

cw	Continuous wave
FBG	Fiber Bragg grating
LPG	Long-period grating

See also

Detection: Fiber Sensors. **Fiber and Guided Wave Optics:** Fabrication of Optical Fiber; Light Propagation; Measuring Fiber Characteristics; Nonlinear Optics; Optical Fiber Cables. **Lasers:** Optical Fiber Lasers. **Optical Amplifiers:** Optical Amplifiers in Long-Haul Transmission Systems. **Optical Communication Systems:** Architectures of Optical Fiber Communication Systems; Basic Concepts. **Optical Materials:** Optical Glasses. **Photon Picture of Light**.

Further Reading

Adams MJ (1981) *An Introduction to Optical Waveguides*. Chichester, UK: Wiley.

Agrawal GP (2001) *Applications of Nonlinear Fiber Optics*. San Diego, CA: Academic Press.

Eggleton BJ, Westbrook PS, Windeler RS, Spalter S and Strasser TA (1999) Grating resonances in air-silica microstructured optical fibers. *Optical Letters* 24: 1460–1462.

Eggleton BJ, Lenz G and Litchinitser NM (2000) Optical pulse compression schemes that use nonlinear Bragg gratings. *Fiber Integrated Optics* 19: 383–421.

Espindola RP, Windeler RS, Abramov AA, *et al.* (1999) External refractive index insensitive air-clad long period fiber grating. *Electronic Letters* 35: 327–328.
Hill KO, Fujii Y, Johnson DC and Kawasaki BS (1978) Photosensitivity in optical fiber waveguides: applications to reflection filter fabrication. *Applied Physics Letters* 32: 647–649.
IEEE Journal of Lightwave Technology (1997) Special issue on fiber gratings, photosensitivity and poling 15(8).
Kashyap R (1999) *Fiber Bragg Gratings*. New York: Academic Press.
Kogelnik H (1990) Theory of optical waveguides. In: Tamir T (ed.) *Guided-Wave Optoelectronics*. Berlin: Springer-Verlag.
Lemaire PJ, Atkins RM, Mizrahi V and Reed WA (1993) High-pressure H_2 loading as a technique for achieving ultrahigh UV photosensitivity and thermal sensitivity in GeO_2 doped optical fibers. *Electronic Letters* 29: 1191–1193.
Marcuse D (1974) *Theory of Dielectric Optical Waveguides*. New York: Academic Press.
Meltz G, Morey WW and Glenn WH (1989) Formation of Bragg gratings in optical fibres by transverse holographic method. *Optical Letters* 14: 823–825.
Othonos A and Kalli K (1999) *Fiber Bragg Gratings: Fundamentals and Applications in Telecommunications and Sensing*. Norwood, MA: Artech House.
Polladian L (1993) Graphical and WKB analysis of nonuniform Bragg gratings. *Physical Review E* 48: 4758–4767.
Westbrook PS, Eggleton BJ, Windeler RS, *et al.* (2001) Cladding mode loss in hybrid polymer-silica microstructured optical fiber gratings. *IEEE Photonic Technology Letters* 12: 495–497.

FOURIER OPTICS

S Jutamulia, University of Northern California, Petaluma, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Coherent optical information processing is almost entirely based on the Fourier-transform property of a lens. A Fourier-transform lens is actually an ordinary lens. If the input transparency is placed in the front focal plane of the lens and illuminated with coherent collimated light (planewave), the amplitude function in the back focal plane of the lens will be the Fourier transform of the input transparency as shown in Figure 1. A coherent optical processor usually consists of two lenses, as shown in Figure 2, thus it performs two Fourier transforms successively. The first lens transforms the input function from space domain into frequency domain, and the second lens transforms back the frequency function from frequency domain to the output function in space domain.

Fourier Transform Property of Lens

Recently, the Fourier-transform property of lenses has been explained, simply using the concept of linear system and geometrical optics. In contrast, the Fourier-transform property of lenses is traditionally analyzed, based on the Fresnel–Kirchhoff diffraction theory involving quadratic phase terms (eqns [2] and [3]). In this article, we adopt the simple geometrical theory based on geometrical optics.

We start our discussion by looking at Figure 1. Consider that a coherent point source is at the center of the front focal plane of the lens, which is denoted by plane (x, y). The point source generates a collimated beam parallel to the optical axis. The amplitude function at the back focal plane of the lens, which is denoted by plane (u, v), can be described as unity. Consequently, we can express mathematically:

$$f(x, y) = \delta(x, y) \qquad [1]$$

$$F(u, v) = 1 \qquad [2]$$

where $f(x, y)$ is the amplitude function in the front focal plane, and $F(u, v)$ is the amplitude function in the back focal plane.

If the point source shifts a distance a along the y-axis, the amplitude function in the front focal plane is

$$f(x, y) = \delta(x, y + a) \qquad [3]$$

The off-axis point source generates a tilted collimated beam. If we place a screen in the plane (u, v), we will see uniform intensity across the screen similar to that which would be seen when the source is centered on the axis. Mathematically, it means:

$$|F(u, v)|^2 = 1 \qquad [4]$$

However, the phase of the amplitude function is not uniform, since the collimated beam is not parallel to

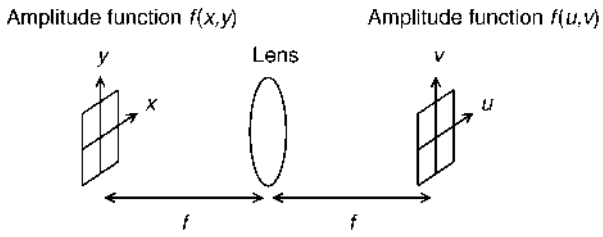

Figure 1 A lens Fourier-transforms amplitude function $f(x, y)$ in the front focal plane to amplitude function $F(u, v)$ in the back focal plane.

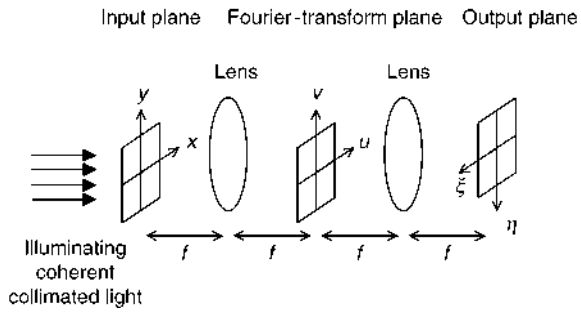

Figure 2 A pair of lenses form 4-f coherent optical processor.

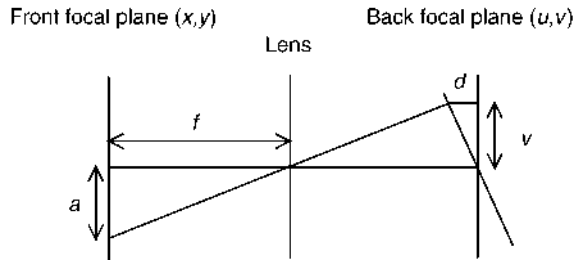

Figure 3 Calculation of phase function in the back focal plane (u, v) based on geometrical optics.

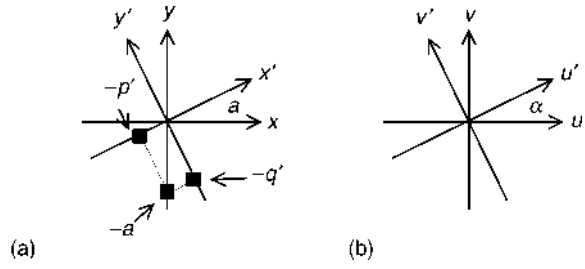

Figure 4 (a) Front focal plane coordinate system (x, y) is rotated by α. (b) Similarly, back focal plane coordinate system (u, v) is rotated by α.

the optical axis. In other words, plane (u, v) is not parallel to the wavefront.

If the phase at the center of plane (u, v) is considered zero, the phase at any point at plane (u, v) can be calculated referring to Figure 3. Using simple geometry, we know that

$$d(u, v) = \frac{a}{f} v \qquad [5]$$

where $d(u, v)$ is the optical path difference. Correspondingly, the phase is

$$\phi(u, v) = \frac{2\pi}{\lambda f} av \qquad [6]$$

where λ is the wavelength of light.

In complex representation, the amplitude function is expressed as

$$F(u, v) = A(u, v) \exp[i\phi(u, v)] \qquad [7]$$

where $A(u, v)$ and $\phi(u, v)$ are magnitude and phase of the amplitude function, respectively. From eqn [4], we know that

$$A(u, v) = |F(u, v)| = 1 \qquad [8]$$

Substituting eqns [8] and [6] into [7] yields

$$F(u, v) = \exp\left(i \frac{2\pi}{\lambda f} av\right) \qquad [9]$$

Equation [9] is the amplitude function in the back focal plane of the lens, when the amplitude function in the front focal plane of the lens is given by eqn [3].

Consider now that coordinate (x,y) is rotated by angle α to new coordinates (x',y') as shown in Figure 4a. The amplitude function in the front focal plane now becomes

$$f(x', y') = \delta(x' + p', y' + q') \qquad [10]$$

where

$$p' = a \sin \alpha \qquad [11a]$$

$$q' = a \cos \alpha \qquad [11b]$$

Correspondingly, coordinate (u, v) is also rotated by the same angle α to new coordinates (u', v') as shown in Figure 4b. The old coordinates u and v can be described in terms of new coordinates u' and v' as follow:

$$u = u' \cos \alpha - v' \sin \alpha \qquad [12a]$$

$$v = u' \sin \alpha + v' \cos \alpha \qquad [12b]$$

Substituting eqn [12b] into eqn [6] yields

$$\phi(u', v') = \frac{2\pi}{\lambda f} a(u' \sin\alpha + v' \cos\alpha) \qquad [13]$$

Further, substituting eqns [11a] and [11b] into [13] yields

$$\phi(u', v') = \frac{2\pi}{\lambda f}(p'u' + q'v') \qquad [14]$$

Thus, the amplitude function in the back focal plane of lens is

$$F(u', v') = \exp\left[i\frac{2\pi}{\lambda f}(p'u' + q'v') \right] \qquad [15]$$

For convenience, we now change parameters x' and y' to x and y, u' and v' to u and v, p' and q' to $-p$ and $-q$. Equations [10] and [15] become

$$f(x, y) = \delta(x - p, y - q) \qquad [16]$$

$$F(u, v) = \exp\left[-i\frac{2\pi}{\lambda f}(pu + qv) \right] \qquad [17]$$

We may simply consider that $f(x, y)$ is an input function and $F(u, v)$ is an output function of a system, since they are both physically the same amplitude functions. If there are two mutually coherent point sources at $(0, 0)$ and $(0, -a)$ with different amplitudes k and l in the front focal plane, the input function is

$$f(x, y) = k\delta(x, y) + l\delta(x, y + a) \qquad [18]$$

These two point sources will generate two collimated beams. Since they are mutually coherent, two collimated beams will interfere, and the resultant amplitude function in the back focal plane is simply the superposition of two beams, which is

$$F(u, v) = k + l \exp\left(i\frac{2\pi}{\lambda f} av \right) \qquad [19]$$

Thus, the lens is a linear system when the light is coherent. Furthermore, if input function $f(x, y)$ is a group of mutually coherent point sources with different amplitudes (i.e., an optical transparency illuminated with a planewave), it can be expressed as follows:

$$f(x, y) = \int\int_{-\infty}^{\infty} f(p, q)\delta(x - p, y - q)\mathrm{d}p\,\mathrm{d}q \qquad [20]$$

Since the lens is a linear system, output function will be the same superposition as follows:

$$F(u,v) = \int\int_{-\infty}^{\infty} f(p,q)\exp\left[-i\frac{2\pi}{\lambda f}(pu+qv) \right]\mathrm{d}p\,\mathrm{d}q \qquad [21]$$

By again changing parameters p and q to x and y, respectively, eqn [21] becomes

$$F(u,v) = \int\int_{-\infty}^{\infty} f(x,y)\exp\left[-i\frac{2\pi}{\lambda f}(xu+yv) \right]\mathrm{d}x\,\mathrm{d}y \qquad [22]$$

which is the optical Fourier transform of the input function $f(x,y)$.

4-f Coherent Optical Processor

If the amplitude function in the front focal plane of a lens is denoted by $f(x, y)$, then the amplitude function in the back focal plane is simply $F(u, v)$ given in eqn [22], which is the Fourier transform of $f(x, y)$. Mathematically, $f(x, y)$ can also be derived from $F(u, v)$ (eqn [4]). After neglecting the normalization factor, it can be expressed as follows:

$$f(x, y) = \int\int_{-\infty}^{\infty} F(u,v)\exp\left[i\frac{2\pi}{\lambda f}(xu+yv) \right]\mathrm{d}u\,\mathrm{d}v \qquad [23]$$

The normalization factor $(1/\lambda f)^2$ is caused by the coordinate scale factor λf in the Fourier-transform plane (u,v).

Equations [22] and [23] are Fourier transform and inverse Fourier transform, respectively. The operation of Fourier transform is represented by a notation $F\{\cdot\}$, and inverse Fourier transform is $F^{-1}\{\cdot\}$.

Two Fourier transform lenses can be combined into a two-lens processor, which is better known as a 4-f optical processor, as shown in Figure 2. Figure 2 illustrates the principle of a 4-f optical processor, in which $f(x, y)$, $F(u, v)$, and $f(\xi, \eta)$ are amplitude functions in input, Fourier transform, and output planes, respectively. Since $F(u, v)$ is the function of spatial frequency, Fourier-transform plane is also called a frequency plane. Since a lens performs only Fourier transform, to perform inverse Fourier transform, the output plane must be rotated by 180°. Thus, orientations of ξ and η correspond to $-x$ and $-y$.

In practice, $f(x, y)$ is represented by an optical mask, which can be a film transparency or a spatial light modulator. The optical mask with $f(x, y)$ amplitude transmittance is illuminated with collimated coherent light to produce the amplitude function $f(x, y)$. The Fourier transform $F(u, v)$ is physically formed in the frequency plane. The frequency content of the input function $f(x, y)$ can be directly altered by placing an optical mask in the frequency plane. Finally, an image with amplitude function $f(\xi, \eta)$ is formed in the output plane. Notice that all detectors, including our eyes, detect intensity that is the square of amplitude.

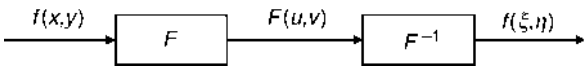

Figure 5 Block diagram of 4-f coherent optical processor.

According to eqn [22], $F(u, v)$ can be expressed in terms of $f(x, y)$ as

$$F(u, v) = F\{f(x, y)\} \qquad [24]$$

Similarly, the amplitude function in the output plane must be

$$f(x', y') = F\{F(u, v)\} \qquad [25]$$

where (x', y') is the coordinate of the output plane having the same orientation as (x, y) in the input plane. This function can be expressed in terms of inverse Fourier transform as follows:

$$f(-x', -y') = F^{-1}\{F(u, v)\} \qquad [26]$$

A rotated coordinate (ξ, η), shown in Figure 2, substitutes $(-x', -y')$, such that

$$f(\xi, \eta) = F^{-1}\{F(u, v)\} \qquad [27]$$

Thus, the successive operations can be simply illustrated by the diagram shown in Figure 5.

Spatial Filtering

The origin of optical information processing can be traced as far back as the work of Abbe of 125 years ago. His interest at that time concerned the theory of microscopy in order to establish a scientific approach to microscope design. He found that a larger aperture resulted in higher resolution. An object under the microscope, whose size was of the order of the wavelength of light, diffracted the illuminating light, i.e., the straight light was bent. If the aperture of the objective lens was not large enough to collect all of the diffracted light, the image did not correspond exactly to that object. Rather, it related to a fictitious object whose complete diffracted light pattern matched the one collected by the objective lens. These lost portions are associated with the higher spatial frequency. Their removal will result in a loss in image sharpness and resolution. The basic ideas of Abbe's theory were later clarified in a series of experiments by Porter in the early 1900s.

Since the late 1940s, the possibility of applying coherent optical techniques to image processing had been explored. In 1960, Cutrona *et al.* presented an optical system using lenses to perform Fourier transforms. Thus, every Fourier transformation in the processor can be realized by a lens. The power of this architecture is its capability of forming Fourier spectra and also being cascaded.

Consider an image formed by an optical system such as a camera; because of the error or limitation of the system, the perfect point in the object may smear. If we consider that the original object is the collection of a large number of very fine points, the formed image will be the collection of the same number of blurred spots, instead of fine points. Thus, the image formation can be expressed by convolution. The blurred spot is represented by the impulse response $h(x, y)$, which is the characteristic of the optical system.

When the input is a point (an impulse), its amplitude function is a delta function $\delta(x, y)$, and the output amplitude function is $h(x, y)$. The observed image is the intensity function $|h(x, y)|^2$. Since it indicates the spread of a point, $|h(x, y)|^2$ is called the point spread function.

The output image function $o(x, y)$ is the convolution of input function $f(x, y)$ and the impulse response $h(x, y)$ as follows:

$$o(x, y) = \int\int_{-\infty}^{\infty} f(p, q)h(x - p, y - q)\mathrm{d}p\,\mathrm{d}q \qquad [28]$$

Notice that $h(x - p, y - q)$ is the blurred spot centered at (p, q). $H(u, v)$, $F(u, v)$, and $O(u, v)$ are Fourier transforms of $h(x, y)$, $f(x, y)$, and $o(x, y)$, respectively. Substitution of

$$h(x-p, y-q) = \int\int_{-\infty}^{\infty} H(u,v) \times \exp\left[i\frac{2\pi}{\lambda f}[(x-p)u+(y-q)v]\right]\mathrm{d}u\,\mathrm{d}v \qquad [29]$$

then

$$F(u,v) = \int\int_{-\infty}^{\infty} f(p,q)\exp\left[-i\frac{2\pi}{\lambda f}(pu+qv)\right]\mathrm{d}p\,\mathrm{d}q \qquad [30]$$

yields

$$o(x,y) = \int\int_{-\infty}^{\infty} F(u,v)H(u,v)\exp\left[i\frac{2\pi}{\lambda f}(xu+yv)\right]\mathrm{d}u\,\mathrm{d}v \qquad [31]$$

Since

$$o(x,y) = \int\int_{-\infty}^{\infty} O(u,v)\exp\left[i\frac{2\pi}{\lambda f}(xu+yv)\right]\mathrm{d}u\,\mathrm{d}v \qquad [32]$$

we obtain

$$O(u,v) = H(u,v)F(u,v) \qquad [33]$$

The function $H(u,v)$, which is the Fourier transform of the impulse response $h(x,y)$, is called transfer function. Actually, $H(u,v)$ changes the spectrum of input function $F(u,v)$ to become $O(u,v)$ and the spectrum of the output function becomes $H(u,v)F(u,v)$.

The first application of spatial filtering may be the restoration of a photograph. The image formed by a camera is a collection of spots. If the image is in focus, the size of the spots is very fine, and thus the image looks sharp. If the image is out of focus, the spot size is fairly large and the image is blurred. Interestingly, an out-of-focus image can be restored using spatial filtering technique.

Consider that the original image is $f(x, y)$. Because it is out of focus, a point in the image becomes a spot expressed by $h(x, y)$. If the image is in focus, $h(x, y)$ is a delta function. If the in-focus image is $f(x, y)$, its frequency function is $F(u, v)$. The defocused image $o(x, y)$ is expressed by eqn [28], and its frequency function is $F(u, v)H(u, v)$. To produce the in-focus image, the frequency function must be $F(u, v)$. Thus $F(u, v)H(u, v)$ must be multiplied by the filter function of $1/H(u, v)$ to produce $F(u, v)$ only. The filter of $1/H(u, v)$ can be made from a photographic plate to provide its magnitude, and a glass plate having two values of thickness to provide $+1$ and -1 signs. The glass plate is essentially a phase filter. The photographic plate with varying density function is an amplitude filter. By passing frequency function $F(u, v)H(u, v)$ through the combination of amplitude and phase filters, it will result in $F(u, v)$ only. When $F(u, v)$ is again Fourier transformed by a lens, an in-focus image $f(x, y)$ will be produced.

Complex Matched Spatial Filtering

The restoration of a photograph mentioned above, utilizes the filter function $1/H(u, v)$, which has positive and negative real values. We will see an application in which the filter function is a complex function. A complex function filter can be produced by a single hologram, instead of using a combination of a phase filter and an amplitude filter. The complex matched spatial filter can perform a correlation, as first demonstrated by Vander Lugt. The complex filter is actually a Fourier transform hologram proposed by Leith and Upatnieks.

The synthesis of a complex filter is shown in Figure 6. First, a function $g(x, y)$ is displayed in the input plane by a transparency or spatial light modulator and illuminated with a collimated coherent light beam. Its Fourier-transform $G(u, v)$ is generated in the back focal plane of the lens. The hologram is made in the Fourier-transform plane by recording the interference pattern of $G(u, v)$ and the reference beam $R(u, v)$. $R(u, v)$ is an oblique collimated beam, having magnitude of unity, which is expressed mathematically as

$$R(u, v) = \exp(iau) \qquad [34]$$

and

$$a = \frac{2\pi \sin \alpha}{\lambda} \qquad [35]$$

where α is the angle of the reference beam and λ is the wavelength of light.

The transmittance of the recorded hologram will be

$$\begin{aligned} T(u, v) &= |G(u, v) + R(u, v)|^2 \\ &= |G(u, v)|^2 + 1 + G(u, v)\exp(-iau) \\ &\quad + G^*(u, v)\exp(iau) \end{aligned} \qquad [36]$$

where * denotes the complex conjugate. The hologram, which is called a complex matched spatial filter, is then placed back in the Fourier-transform plane, as shown in Figure 7. A new input function $f(x, y)$ is displayed in the input plane. The Fourier-transform of input function, $F(u, v)$ will be generated in the Fourier-transform plane. The amplitude of light,

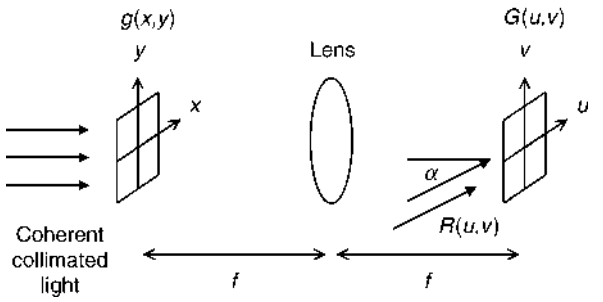

Figure 6 Synthesis of complex matched spatial filter is the recording of Fourier-transform hologram.

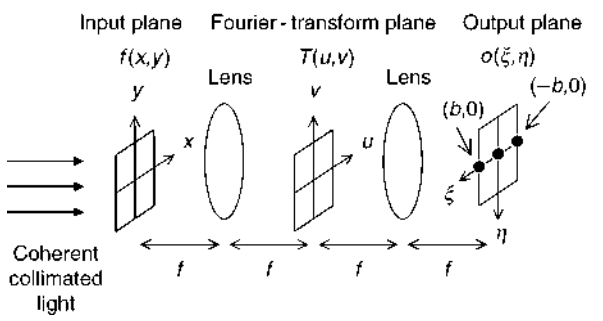

Figure 7 Optical processing using complex matched spatial filter in a 4-f processor.

immediately after the complex matched spatial filter, is

$$F(u,v)T(u,v) = (|G(u,v)|^2 + 1)F(u,v) + F(u,v)G(u,v)\exp(-iau) + F(u,v)G^*(u,v)\exp(iau) \quad [37]$$

The second lens then produces the inverse Fourier transform of $F(u,v)T(u,v)$ in the output plane:

$$F^{-1}\{F(u,v)T(u,v)\} = F^{-1}\{(|G(u,v)|^2 + 1)F(u,v)\} + F^{-1}\{F(u,v)G(u,v)\exp(-iau)\} + F^{-1}\{F(u,v)G^*(u,v)\exp(iau)\} \quad [38]$$

The first term will be at the center of the output plane. The second and the third terms will appear at $(b,0)$ and $(-b,0)$ in (ξ,η) plane, respectively, where

$$b = f\sin\alpha \quad [39]$$

and f is the focal length of lens.

We are now in a position to analyze the inverse Fourier transform of $F(u,v)G(u,v)$ and $F(u,v)G^*(u,v)$. Referring to eqns [28]–[33], the inverse Fourier transform of $F(u,v)G(u,v)$ is the convolution of $f(\xi,\eta)$ and $g(\xi,\eta)$. Thus, the complex matched spatial filtering can also be applied to the restoration of photography mentioned previously without using the combination of amplitude and phase filters.

The inverse Fourier transform of $F(u,v)G^*(u,v)$ is derived as follows:

$$o_3(\xi,\eta) = \int\int_{-\infty}^{\infty} F(u,v)G^*(u,v) \times \exp\left[i\frac{2\pi}{\lambda f}(\xi u + \eta v)\right]\mathrm{d}u\,\mathrm{d}v \quad [40]$$

Substitute

$$F(u,v) = \int\int_{-\infty}^{\infty} f(\alpha,\beta) \times \exp\left[-i\frac{2\pi}{\lambda f}(\alpha u + \beta v)\right]\mathrm{d}\alpha\,\mathrm{d}\beta \quad [41]$$

and

$$G^*(u,v) = \int\int_{-\infty}^{\infty} g^*(-p,-q) \times \exp\left[-i\frac{2\pi}{\lambda f}(pu + qv)\right]\mathrm{d}p\,\mathrm{d}q \quad [42]$$

into eqn [40], one obtains

$$o_3(\xi,\eta) = \int\int_{-\infty}^{\infty} f(\alpha,\beta)\int\int_{-\infty}^{\infty} g^*(-p,-q) \times \int\int_{-\infty}^{\infty} \exp\left[i\frac{2\pi}{\lambda f}(\xi-\alpha-p)u + (\eta-\beta-q)v\right]\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}p\,\mathrm{d}q\,\mathrm{d}\alpha\,\mathrm{d}\beta$$

$$o_3(\xi,\eta) = \int\int_{-\infty}^{\infty} f(\alpha,\beta)\int\int_{-\infty}^{\infty} g^*(-p,-q) \times \delta(\xi-\alpha-p,\eta-\beta-q)\mathrm{d}p\,\mathrm{d}q\,\mathrm{d}\alpha\,\mathrm{d}\beta$$

$$o_3(\xi,\eta) = \int\int_{-\infty}^{\infty} f(\alpha,\beta)g^*(\alpha-\xi,\beta-\eta)\mathrm{d}\alpha\,\mathrm{d}\beta \quad [43]$$

Equation [43] is the correlation of $f(x,y)$ and $g(x,y)$. It is apparent that the correlation output $o_3(\xi,\eta)$ will achieve the maximum when $f(x,y)$ is identical to $g(x,y)$. If $f(x,y)$ and $g(x,y)$ are different, the correlation function will be a flat function with minor random fluctuation. Optical correlators using complex matched spatial filters have been applied to automatic pattern recognition.

Joint Transform Correlator

Alternatively, the joint transform correlator can perform a correlation without recording a complex matched spatial filter in advance. In addition to the correlation, the joint transform correlator can also be used as a general optical processor.

A transparency with amplitude transmittance $f(x,y)$ is centered at $(-a,0)$, and a second transparency with amplitude transmittance $g(x,y)$, is centered at $(a,0)$. The two transparencies are illuminated by a coherent collimated beam and jointly Fourier transformed by a single lens as shown in Figure 8. In the back focal plane of the lens, the amplitude function $S(u,v)$ is

$$S(u,v) = \int\int_{-\infty}^{\infty} [f(x+a,y) + g(x-a,y)] \times \exp\left[-i\frac{2\pi}{\lambda f}(xu + yv)\right]\mathrm{d}x\,\mathrm{d}y \quad [44]$$

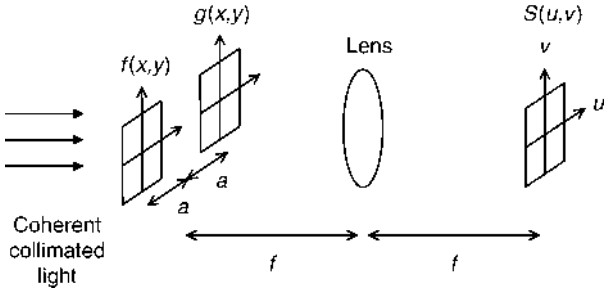

Figure 8 Joint Fourier transform of two input functions.

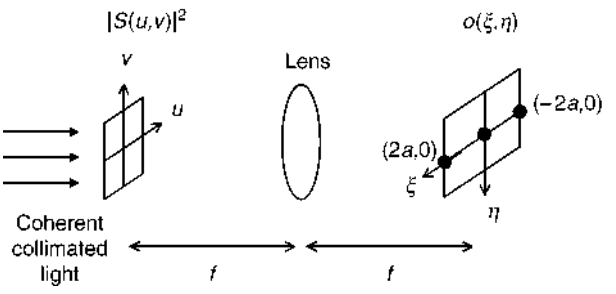

Figure 9 Fourier transform of the joint power spectrum.

$S(u,v)$ is the joint Fourier spectrum of $f(x,y)$ and $g(x,y)$. It can be written as

$$S(u,v) = F(u,v)\exp\left[i\frac{2\pi}{\lambda f}au\right] + G(u,v)\exp\left[-i\frac{2\pi}{\lambda f}au\right] \quad [45]$$

where $F(u,v)$ and $G(u,v)$ are Fourier transforms of $f(x,y)$ and $g(x,y)$, respectively.

If a square-law detector is placed in the back focal plane of the lens, the recorded pattern of joint power spectrum $|S(u,v)|^2$ is

$$|S(u,v)|^2 = |F(u,v)|^2 + F(u,v)G^*(u,v)\exp\left[i\frac{4\pi}{\lambda f}au\right] + F^*(u,v)G(u,v)\exp\left[-i\frac{4\pi}{\lambda f}au\right] + |G(u,v)|^2 \quad [46]$$

where * denotes the complex conjugate. A photographic plate is a square-law detector and can be used to record $|S(u,v)|^2$. When the developed plate is placed in the input plane, as shown in Figure 9, the input transmittance is $|S(u,v)|^2$ as given in eqn [46].

The inverse Fourier transform of the joint power spectrum $|S(u,v)|^2$, in the back focal plane of the lens, is

$$o(\xi,\eta)=\int\int_{-\infty}^{\infty}|S(u,v)|^2\exp\left[i\frac{2\pi}{\lambda f}(\xi u+\eta v)\right]\mathrm{d}u\,\mathrm{d}v \quad [47]$$

which consists of four terms, according to eqn [46]. We now analyze this equation term by term. The first term is

$$o_1(\xi,\eta)=\int\int_{-\infty}^{\infty}|F(u,v)|^2\exp\left[i\frac{2\pi}{\lambda f}(\xi u+\eta v)\right]\mathrm{d}u\,\mathrm{d}v \quad [48]$$

which is the auto-correlation of $f(x,y)$

$$o_1(\xi,\eta)=\int\int_{-\infty}^{\infty}f(\alpha,\beta)f^*(\alpha-\xi,\beta-\eta)\mathrm{d}\alpha\,\mathrm{d}\beta \quad [49]$$

The second term is

$$o_2(\xi,\eta)=\int\int_{-\infty}^{\infty}F(u,v)G^*(u,v)\exp\left[i\frac{4\pi}{\lambda f}au\right]\times\exp\left[i\frac{2\pi}{\lambda f}(\xi u+\eta v)\right]\mathrm{d}u\,\mathrm{d}v \quad [50]$$

which is

$$o_2(\xi,\eta)=\int\int_{-\infty}^{\infty}f(\alpha,\beta)g^*[\alpha-(\xi+2a),\beta-\eta]\mathrm{d}\alpha\,\mathrm{d}\beta \quad [51]$$

Equation [51] is the correlation between $f(x,y)$ and $g(x,y)$, shifted by $-2a$ in the ξ-axis direction.

Similarly, the third term will be the correlation between $g(x,y)$ and $f(x,y)$, shifted by $2a$ in the ξ-axis direction:

$$o_3(\xi,\eta)=\int\int_{-\infty}^{\infty}g(\alpha,\beta)f^*[\alpha-(\xi-2a),\beta-\eta]\mathrm{d}\alpha\,\mathrm{d}\beta \quad [52]$$

And the fourth term is

$$o_4(\xi,\eta)=\int\int_{-\infty}^{\infty}g(\alpha,\beta)g^*(\alpha-\xi,\beta-\eta)\mathrm{d}\alpha\,\mathrm{d}\beta \quad [53]$$

which is the auto-correlation of $g(x,y)$ and overlaps with the first term at the center of the output plane.

From the analysis given above, it is understood that the correlation of $f(x,y)$ and $g(x,y)$, and the correlation of $g(x,y)$ and $f(x,y)$, are shifted by $-2a$ and $2a$, respectively, along the ξ-axis. The correlation outputs are centered at $(-2a,0)$ and $(2a,0)$ in the back focal plane of the lens. If $f(x,y)$ and $g(x,y)$ are the same, the two correlation outputs are identical. The first and fourth terms are the autocorrelations of $f(x,y)$ and $g(x,y)$, centered at $(0,0)$. Thus the second and third terms, which are correlation signals, can be isolated. A versatile real-time joint transform

correlator, using inexpensive liquid crystal televisions, is widely used in laboratory demonstration.

See also

Information Processing: Coherent Analogue Optical Processors. **Spectroscopy:** Fourier Transform Spectroscopy.

Further Reading

Abbe E (1873) Beitrage zur Theorie des Mikroskops und der mikroskopichen Wahrnehmung. *Archiv Mikroskopie Anatomie* 9: 413–468.

Bracewell RN (1965) *The Fourier Transform and Its Applications*. New York: McGraw-Hill.

Cutrona LJ, Leith EN, Palermo CJ and Porcello LJ (1960) Optical data processing and filtering systems. *IRE Transactions on Information Theory* IT-6: 386–400.

Goodman JW (1968) *Introduction to Fourier Optics*. New York: McGraw-Hill.

Jutamulia S (1992) Joint transform correlators and their applications. *Proceedings of the SPIE* 1812: 233–243.

Jutamulia S and Asakura T (2002) Optical Fourier-transform theory based on geometrical optics. *Optical Engineering* 41: 13–16.

Leith EN and Upatnieks J (1962) Reconstructed wavefronts and communication theory. *Journal of the Optical Society of America* 52: 1123–1130.

Porter AB (1906) On the diffraction theory of microscopic vision. *Philosophical Magazine* 11: 154–166.

Rau JE (1966) Detection of differences in real distributions. *Journal of the Optical Society of America* 56: 1490–1494.

Tsujiuchi J (1963) Correction of optical images by compensation of aberrations and by spatial frequency filtering. *Progress in Optics* 2: 133–180.

Vander Lugt A (1964) Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory* IT-10: 139–145.

Weaver CS and Goodman JW (1966) A technique for optically convolving two functions. *Applied Optics* 5: 1248–1249.

Yu FTS (1983) *Optical Information Processing*. New York: Wiley.

Yu FTS and Ludman JE (1988) Joint Fourier transform processor. *Microwave Optical Technology Letters* 1: 374–377.

Yu FTS, Jutamulia S, Lin TW and Gregory DA (1987) Adaptive real-time pattern recognition using a liquid crystal TV based joint transform correlator. *Applied Optics* 26: 1370–1372.

G

GEOMETRICAL OPTICS

Contents

Lenses and Mirrors
Aberrations
Prisms

Lenses and Mirrors

A Nussbaum, University of Minnesota, Minneapolis, MN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Lenses are carefully shaped pieces of transparent material used to form images. The term 'transparent' does not necessarily mean that the lens is transmitting visible light. The element germanium, well-known in the transistor industry, is a light-gray opaque crystal which forms images with infrared radiation, and provides the optics for night-vision instruments. The lenses used in cameras, telescopes, microscopes, and other well-known applications, are made from special glasses or – more recently – from plastics. Hundreds of different optical glass formulas are listed in the manufacturers' catalogues. Plastics have come into use in recent years because it is now possible to mold them with high precision. The way in which lenses form images has been studied for several hundred years, resulting in an enormous and very complicated literature. In this article, we show how this theory can be simplified by introducing the restriction that lenses are perfect. This limitation provides an easy introduction to the image forming process. A companion article **Geometrical Optics:** Aberrations, will deal with imperfect components, showing how their defects degrade images and how corrections are made.

The Cardinal Points and Planes of an Optical System

The location, size and orientation of an image are items of information which a designer needs to know. The nature of the image can be determined by a simple model of an optical system, based on two experimental observations. The first is well-known and is pictured in Figure 1. Parallel rays of light passing through a double convex lens (a simple magnifying glass, for example) should meet at the focal point or focus, designated as F'. For rays from the right side of the lens, another such point F exists at an equal distance from the lens. The other experiment uses the lens to magnify (Figure 2a). The amount of magnification decreases as the lens is brought closer to 'ISAAC NEWTON' (Figure 2b), and when the lens touches the paper, object and image are approximately the same size. The locations of object and image corresponding to equality in size are called the unit or principal points H and H', respectively, and the set of four points F, F', H, and H' are the cardinal points. Therefore, the planes through these points normal to the axis of the lens are the cardinal planes. Figure 3 shows an object whose image we can now determine. The coordinate orientation in this figure may appear strange at first. It is customary in optics to designate the lens axis as OZ, and the other two axes form a right-handed system. The x-axis then lies in the

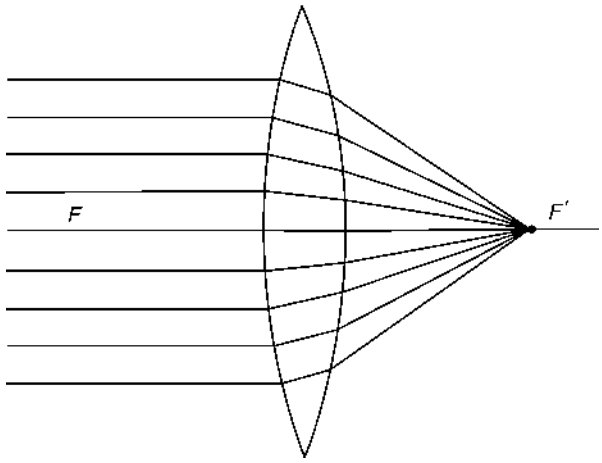

Figure 1 Focal points of a double convex lens.

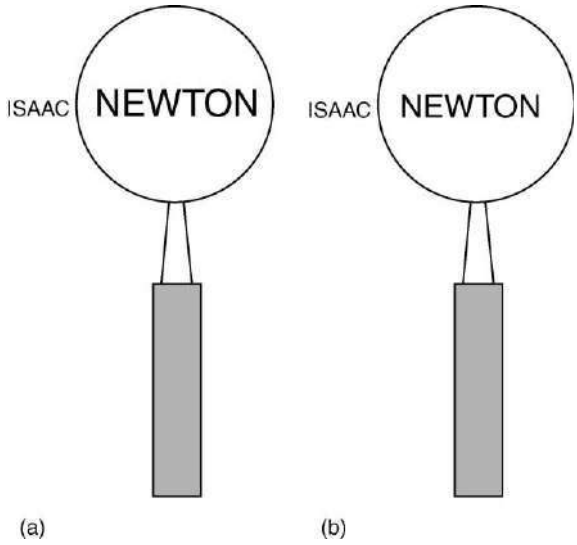

Figure 2 Magnification by a double convex lens. The magnification in (a) is reduced when the lens is brought closer to the image (b).

plane of the figure; the y-axis, coming out of the paper and perpendicular to it, is not shown. The object, of height x, is placed to the left of the focal point F. The cardinal points occur in the order F, H, H', F', since we expect – based on the experiment corresponding to Figure 2 – that the unit points are very close to the lens. Another fact to be demonstrated later is that H and H' are actually inside the lens. Two rays are shown leaving the point P at the top of the object. The ray which is parallel to the axis strikes the unit plane through H and continues to the right, completely missing the lens. The ray then meets the unit plane through H' at a point which must be a distance x from the axis. To understand why, we recognize that if the object were relocated so as to lie

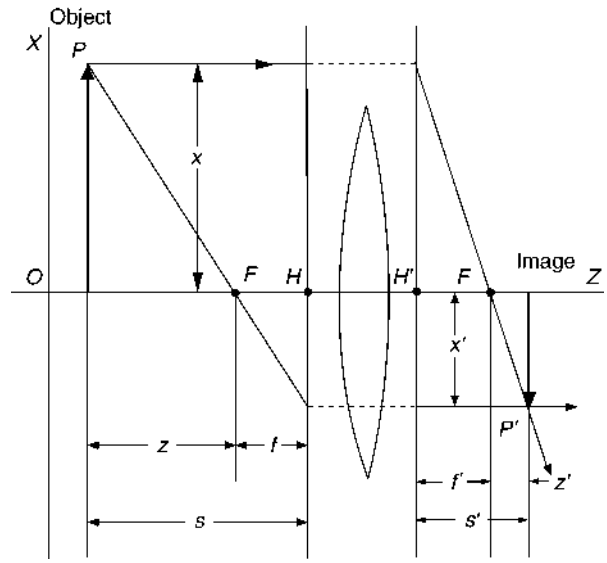

Figure 3 Ray tracing using cardinal points and planes.

on the object-space unit plane, the ray emerging at the other unit plane will be at a distance x from the axis; as required by the definition of these planes. We then assume that the ray is bent or refracted at the unit plane at H' and proceeds to, and beyond, the focal point F'. The second ray, from P through the focal point F, is easily traced since the lens is indifferent to the direction of the light. Assume temporarily that we know the location of the image point P'. Let a parallel ray leave this point and travel to the left. The procedure just given indicates that this ray will strike the unit plane at H', emerge at the other unit plane, be refracted so as to pass through F and reach the object point P. Then reversing the direction of this ray, it will start at P, emerge at the unit plane H' and travel parallel to the axis, intersecting the other ray at P'. The distance between F and H is called the object space focal length f, with f' being the corresponding image space quantity. Simple geometry, making use of the similar triangles and the distances indicated in this diagram, can be used to derive the Newton lens equation or the equivalent Gauss lens equation. Detailed derivations of these two equations will be found in either of the sources listed in the Further Reading at the end of this article. It is not usually made clear in physics texts that these equations apply only to the special case of a single, thin lens, and we shall not apply them here. The image in this figure is real, inverted, and reduced. That is, it is smaller than the object, oriented in the opposite direction, and can be projected onto a screen located at the image position. This result can be verified by using an ordinary magnifier to form the image of a distant light source on a sheet of white paper.

Paraxial Matrices

To consider how light rays behave in optical systems, we introduce the index of refraction n of a material medium, defined as the ratio of the velocity of light c in a vacuum to its velocity v in that medium, or:

$$n = \frac{c}{v} \qquad [1]$$

The velocity of light in air is approximately the same as it is in a vacuum. A light ray crossing the interface between air and a denser medium such as glass will be bent towards the normal to the surface. This is the phenomenon of refraction and the amount of bending is governed by Snell's law, which has the form:

$$n \sin\theta = n' \sin\theta' \qquad [2]$$

where n and n' are the indices of refraction of the two media and θ and θ' are the angles as measured from the normal. Figure 4 shows a ray of light leaving an object point P and striking the first surface of a lens at point P_1. It is refracted there, and proceeds to the second surface. All rays shown lie in the z, x-plane or meridional plane, which passes through the symmetry axis OZ. The amount of refraction is specified by Snell's law, eqn [2], which we shall now simplify. The sine of an angle can be approximated by the value of the angle itself, if this value is small when expressed in radians (less than 0.1 rad), and

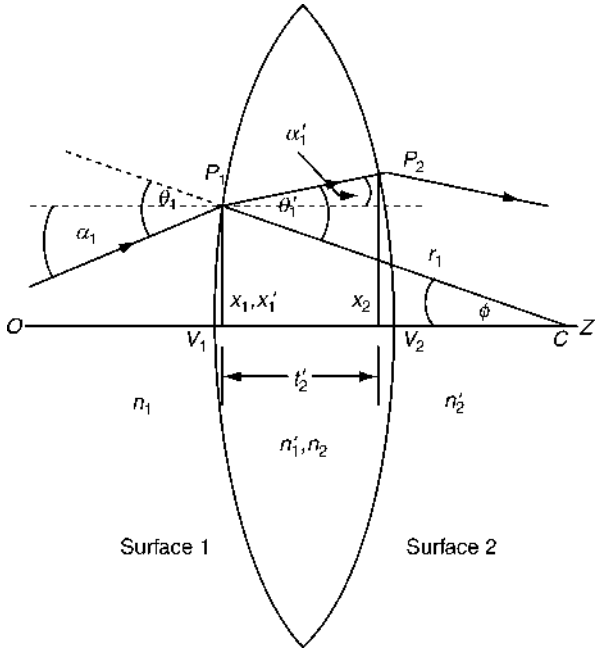

Figure 4 Ray passing through a double convex lens.

Snell's law simplifies to

$$n_1\theta_1 = n_1'\theta_1' \qquad [3]$$

This is called the paraxial form of Snell's law; the word paraxial means 'close to the axis'. It turns out, however, that the Snell's law angles are not convenient to work with, and we eliminate them with the terms:

$$\theta_1 = \alpha_1 + \phi, \qquad \theta_1' = \alpha_1' + \phi \qquad [4]$$

where α_1 is the angle which the incident ray makes with the axis OZ, α_1' is the corresponding angle for the refracted ray, and ϕ is the angle which the radius r_1 (the line from C to P_1) of the lens surface makes at the center of curvature C. This particular angle can be specified as

$$\sin\phi = \frac{x_1}{r_1} \qquad [5]$$

but using the paraxial approximation simplifies this to

$$\phi = \frac{x_1}{r_1} \qquad [6]$$

Substituting for the angles gives

$$n_1'\alpha_1' = \frac{n_1 - n_1'}{r_1} + n_1\alpha_1 \qquad [7]$$

Notice that the distance from the point P_1 to the axis is labeled as either x_1 or x_1'. This strange notation leads to the trivial relation

$$x_1 = x_1' \qquad [8]$$

and this can be combined with eqn [7] to obtain the matrix equation:

$$\begin{pmatrix} n_1'\alpha_1' \\ x_1' \end{pmatrix} = \begin{pmatrix} 1 & -k_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} n_1\alpha_1 \\ x_1 \end{pmatrix} \qquad [9]$$

where the constant k_1 is called the refracting power of surface 1 and is defined as

$$k_1 = \frac{n_1' - n_1}{r_1} \qquad [10]$$

To review the procedure for matrix multiplication, the first element $n_1'\alpha_1'$ in the one-column product matrix is calculated by multiplying the upper left-hand corner of the 2×2 matrix with the first element of the one-column matrix to its right, obtaining $1(n_1\alpha_1)$, and to this is added the product $-k_1x_1$ of the upper right-hand element in the 2×2 matrix and the second member of the one-column matrix. This square matrix is called

the refraction matrix R_1 for surface 1, defined as

$$R_1 = \begin{pmatrix} 1 & -k_1 \\ 0 & 1 \end{pmatrix} \qquad [11]$$

Next, we look at what happens to the ray as it travels from surface 1 to surface 2. As it goes from P_1 to P_2, its distance from the axis becomes

$$x_2 = x'_1 + t'_1 \tan \alpha'_1 \qquad [12]$$

or using the paraxial approximation for small angles:

$$x_2 = x'_1 + t'_1 \alpha'_1 \qquad [13]$$

Another approximation we shall use is to regard t'_1 as being equal to the distance between the lens vertices V_1 and V_2. Using the identity:

$$\alpha_2 = \alpha'_1 \qquad [14]$$

leads to a second matrix equation:

$$\begin{pmatrix} n_2\alpha_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ t'_1/n'_1 & 1 \end{pmatrix} \begin{pmatrix} n'_1\alpha'_1 \\ x'_1 \end{pmatrix} \qquad [15]$$

where the 2×2 translation matrix T_{21} is defined as

$$T_{21} = \begin{pmatrix} 1 & 0 \\ t'_1/n'_1 & 1 \end{pmatrix} \qquad [16]$$

To get an equation which combines a refraction followed by a translation, we take the one-column matrix on the left side of eqn [9] and substitute it into eqn [15] to obtain

$$\begin{pmatrix} n_2\alpha_2 \\ x_2 \end{pmatrix} = T_{21} R_1 \begin{pmatrix} n_1\alpha_1 \\ x_1 \end{pmatrix} \qquad [17]$$

Note that the 2×2 matrices appear in right to left order and that the multiplication of two square matrices is an extension of the rule given above. This equation gives the location and slope of the ray which strikes surface 2 after a refraction and a translation. To continue the ray trace at surface 2, we introduce a second refraction matrix R_2 by expressing k_2 in terms of the two indices at this surface and the radius. Then eqn [17] is extended to give the relation:

$$\begin{pmatrix} n'_2\alpha'_2 \\ x'_2 \end{pmatrix} = R_2 T_{21} R_1 \begin{pmatrix} n_1\alpha_1 \\ x_1 \end{pmatrix} \qquad [18]$$

This process can obviously be applied to any number of components. The product of the three 2×2 matrices in the above equation is known as the system matrix S_{21} and it completely specifies the effect of the lens on the incident ray passing through it. It is also written as

$$S_{21} = R_2 T_{21} R_1 = \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \qquad [19]$$

where the four quantities a, b, c, and d are known as the Gaussian constants. They are used extensively in specifying the behavior of a lens or an optical system. We now state the sign and notation conventions which are necessary for the application of paraxial matrix methods:

1. Light normally travels from left to right.
2. The optical systems we deal with are symmetrical about the z-axis. The intersections of the refracting or reflecting surfaces with this axis are the vertices and are designated in the order encountered as V_1, V_2, etc.
3. Positive directions along the axes are measured from the origin in the usual Cartesian fashion, so that horizontal distances (that is, along the z-axis) are positive if measured from left to right. Angles are positive when measured up from the z-axis.
4. Quantities associated with the incident ray are unprimed; those for the refracted ray are primed.
5. A subscript denotes the associated surface.
6. If the center of curvature of a surface is to its right, the radius is positive, and vice versa.
7. There is a special rule for mirrors to be explained below.

Using the Gaussian Constants

Figure 5 shows the double convex lens of Figure 4 with the assumed locations of the cardinal points indicated. These positions, which we shall now determine accurately, are at distances designated as l_F, l'_F, l_H, and l'_H and are measured from the associated vertex. The object position can be called t, a positive quantity which is measured to the right from object to the first vertex, or it can be called t_1 if measured in the opposite direction. For the first choice, the matrix specifying the translation from object to lens will have the quantity t/n_1 in its lower left-hand corner. However, it is both logical and convenient to use the first vertex as the reference point. Hence, we replace t/n_1 in the translation matrix with the quantity $-t_1/n_1$, and remember to specify t_1 as a negative number when calculating the image position. The equation connecting object and image is obtained by starting with this matrix, multiplying it by the system matrix of eqn [19], and finally by a translation matrix

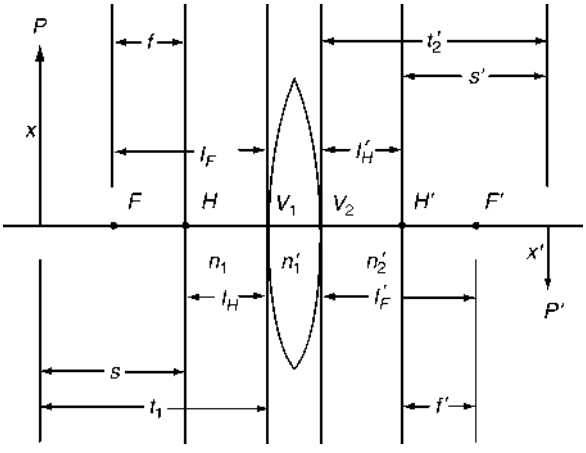

Figure 5 Definitions of cardinal plane locations.

corresponding to the translation t_2' from lens to image to obtain:

$$\begin{pmatrix} n_2'\alpha_2' \\ x' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ t_2'/n_2' & 1 \end{pmatrix} \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ -t_1/n_1 & 1 \end{pmatrix} \begin{pmatrix} n_1\alpha_1 \\ x \end{pmatrix} \qquad [20]$$

where α or α_1 is the angle that the ray from the top of the object makes with the z-axis. This equation takes the initial value of the inclination α and of the position x of the ray leaving the object, and determines the final values, α' and x' at the image. The product of the three 2×2 matrices on the right-hand side can be consolidated into a single matrix called the object-image matrix $S_{P'P}$. Then eqn [20] can be written as

$$\begin{pmatrix} n_2'\alpha_2' \\ x' \end{pmatrix} = S_{P'P} \begin{pmatrix} n_1\alpha_1 \\ x \end{pmatrix} \qquad [21]$$

and this matrix has the complicated form:

$$S_{P'P} = \begin{pmatrix} b + \dfrac{at_1}{n_1} & -a \\ \dfrac{bt_2'}{n_2'} + \dfrac{at_1t_2'}{n_1n_2'} - d - \dfrac{ct_1}{n_1} & c - \dfrac{at_2'}{n_2'} \end{pmatrix} \qquad [22]$$

If we put this matrix into the previous equation, we obtain an unsatisfactory result: the value of x' will depend on the angle α_1 made by the incident ray, as can be seen by multiplying the two matrices on the right side. The ratio of x' to x is called the magnification m; that is

$$m = \frac{x'}{x} \qquad [23]$$

A perfect image can be formed only if the magnification is determined solely by the object distance and the constants of the lens. To eliminate this difficulty, the lower left-hand element in the matrix, eqn [22] is required to be equal to zero, and the magnification is then:

$$m = \frac{x'}{x} = c - \frac{at_2'}{n_2'} \qquad [24]$$

The determinant of this matrix is unity, since it is the product of three matrices whose individual determinants are unity. It follows that

$$\frac{1}{m} = b + \frac{at_1}{n_1} \qquad [25]$$

so that

$$\begin{pmatrix} n_2'\alpha_2' \\ x' \end{pmatrix} = \begin{pmatrix} 1/m & -a \\ 0 & m \end{pmatrix} \begin{pmatrix} n_1\alpha_1 \\ x \end{pmatrix} \qquad [26]$$

Using the fact that the lower left-hand element of the matrix eqn [22] is zero shows that

$$\frac{t_2'}{n_2'} = \frac{d + ct_1/n_1}{b + at_1/n_1}, \qquad \frac{t_1}{n_1} = \frac{d - bt_2'/n_2'}{-c + at_2'/n_2'} \qquad [27]$$

This equation connects the object distance t_1 with the image distance t_2', both quantities being measured from the appropriate vertex. This relation is known as the generalized Gauss lens equation. It applies to all lens systems, no matter how complicated.

We can determine the location of the unit planes by letting $m = 1$. This t_2' in eqn [27] is the location l_H' of the image space unit plane, and it becomes

$$l_H' = \frac{n_2'(c-1)}{a} \qquad [28]$$

This relation expresses the location of the unit plane on the image side as the distance from V_2 to H'. Similarly, the location of H with respect of V_1 is

$$l_H = \frac{n_1(1-b)}{a} \qquad [29]$$

To locate the focal planes, consider a set of parallel rays coming from infinity and producing a point image at F'. The terms containing t_1 in eqn [27] are

much larger than d or b, and this relation becomes

$$l'_F = \frac{n'_2 c}{a} \qquad [30]$$

Letting t'_2 become infinite in the right-hand half of eqn [27], the location of F is given by the equation

$$l_F = \frac{-n_1 b}{a} \qquad [31]$$

The focal lengths f and f' shown in Figure 5 were defined earlier as the distances between their respective focal and unit planes. Hence

$$f' = l'_F - l'_H = \frac{n'_2}{a} \qquad [32]$$

and

$$f = l_F - l_H = \frac{-n_1}{a} \qquad [33]$$

If we let $n_1 = n'_2 = 1$ for air, these become

$$-f = f' = \frac{1}{a} \qquad [34]$$

The Gaussian constant a is thus the reciprocal of the focal length in air. Even when the lens is not in air, it is still true that $f' = -f$ if the lens is in a single medium. It is also true regardless of whether or not the lens is symmetric.

It is customary in optics to work with a consistent set of units, such as centimeters or millimeters, and the sign conventions used for x and y are the standard Cartesian rules. As mentioned above, curved surfaces which open to the right have a positive radius and vice versa. As an example, let a double convex lens have radii of 2.0 and 1.0, respectively, an index of refraction of 1.5, and a thickness of 0.5. These quantities are then denoted as

$$\begin{aligned} &r_1 = +2.0, \qquad r_2 = -1.0, \qquad t'_1 = t_2 = +0.5 \\ &n_1 = 1.0, \qquad n'_1 = 1.5 = n_2, \qquad n'_2 = 1.0 \end{aligned} \qquad [35]$$

By eqn [10]:

$$k_1 = \frac{n'_1 - n_1}{r_1} = \frac{1.5 - 1.0}{2.0} = 0.25 \qquad [36]$$

and

$$k_2 = \frac{1.0 - 1.5}{-1.0} = 0.50 \qquad [37]$$

The system matrix S_{21} is

$$\begin{aligned} S_{21} &= R_2 T_{21} R_1 \\ &= \begin{pmatrix} 1 & -0.50 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0.5/1.5 & 1 \end{pmatrix} \begin{pmatrix} 1 & -0.25 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0.83 & -0.71 \\ 0.33 & 0.92 \end{pmatrix} \end{aligned} \qquad [38]$$

and we note that the determinant is

$$(0.83)(0.92) - (0.71)(-0.33) = 1.00 \qquad [39]$$

This property of the system matrix provides a very useful check on the accuracy of matrix multiplications. The locations of the four cardinal points can be determined from the formulas given above and their locations are shown in the scale drawing of Figure 6. Since we know how to locate the cardinal points of a lens, we are in a position to ray trace in a precise manner, using the method of Figure 3 and adding a third ray. After locating F, H, H', and F', we no longer need the lens; the cardinal planes with the proper separation are fully equivalent to the lens they replace. This is very much like what electrical engineers do when they replace a complicated circuit with a black box; this equivalent circuit behaves just like the original. Ray tracing becomes more complicated when we deal with diverging lenses (Figure 7), but the procedure gives above still applies. Parallel rays spread apart and do not come to a focus on the right-hand side of the lens. However, if the refracted

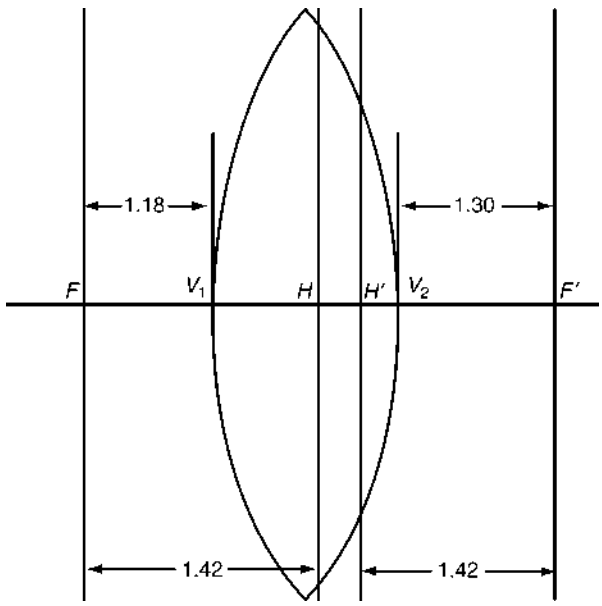

Figure 6 Positions of cardinal planes for an asymmetric lens.

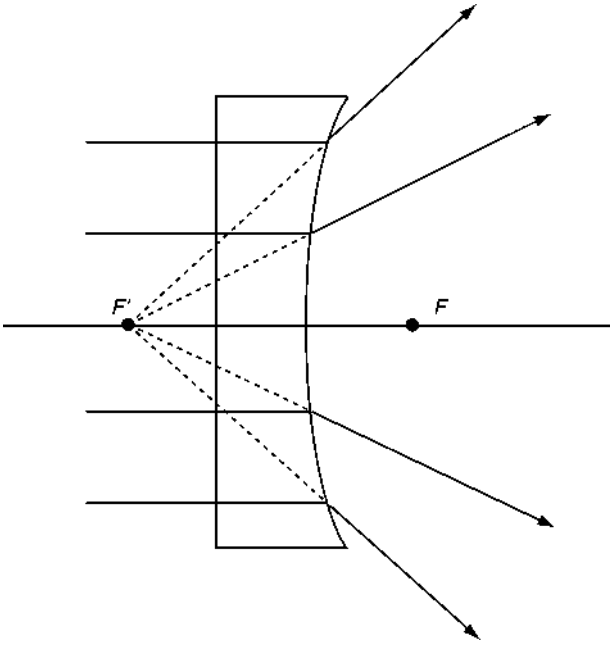

Figure 7 Ray tracing for a planar-concave lens.

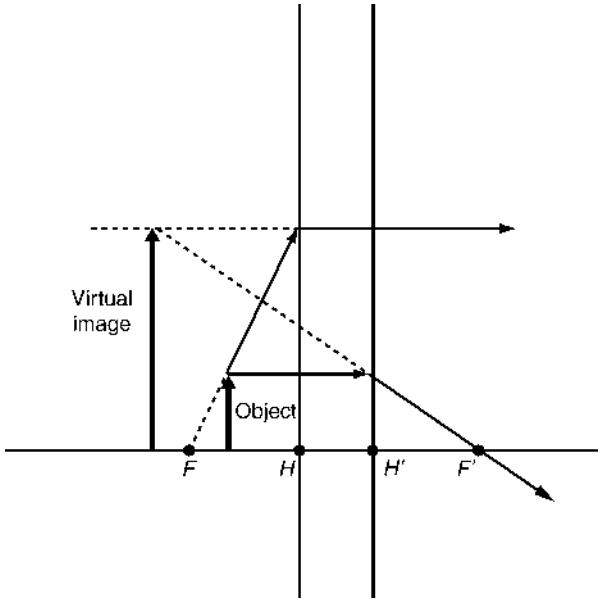

Figure 8 Formation of a virtual image by a convex lens.

rays are extended backwards, these extensions will meet as indicated. This behavior can be verified quantitatively by using the formulas developed above. It will be found that F and F' have exchanged places, but H and H' retain their original positions inside the lens. If an object is placed between the focal point F and the first vertex of a convex lens (Figure 8), then our usual procedure produces two rays which do not intersect in image space. The ray from the object which is parallel to the axis is

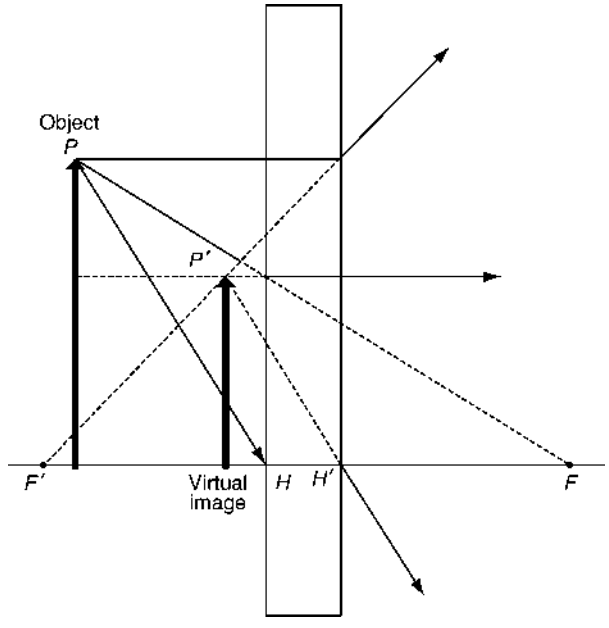

Figure 9 Formation of a virtual image by a concave lens.

refracted downward so that it does not intersect the other ray in image space. However, the extensions to the left of these two rays meet to form an image which is erect, magnified, and virtual. This is the normal action of a magnifying lens; the image can be seen but cannot be projected onto a screen, unlike the real, inverted image of Figure 3. This ray tracing diagram confirms what would be seen when a double convex lens is used to magnify an object lying close to the lens. For the double concave lens of Figure 9, a parallel ray leaves an object point P and is refracted upward at the unit plane H' in such a way that its extension, rather than the ray itself, passes through F'. The ray headed for F is refracted at H before it can reach the object-space focal plane and becomes parallel to the axis. The third ray, going from P to H, emerges at H' parallel to its original direction and its extension to the left passes through the image point P' already determined by the intersection of the other two rays. The resulting virtual image is upright and reduced. All three rays in this diagram behave exactly like the corresponding rays in Figure 3; the only change is the use of the extensions to locate the image. The eye receives the diverging rays from P and believes that they are coming from P'.

Nodal Points and Planes

Let a ray leave a point on the z-axis at an angle α_1, and have an angle α'_2 when it reaches image space.

Since $x = 0$ at the starting point, eqn [26] shows that

$$n_2'\alpha_2' = n_1\alpha_1/m \qquad [40]$$

The ratio of the final angle to the initial angle in this equation is the angular magnification μ, which is then

$$\mu = n_1/mn_2' \qquad [41]$$

The locations of object and image for unit angular magnification are called the nodal points, labeled as N and N'. The above definition shows that the linear and angular magnification are reciprocals of one another when object and image are in air or the same medium. The procedure that located the unit points will show that the nodal points have positions given by

$$\frac{l_N}{n_1} = \frac{(n_2'/n_1) - b}{a} \qquad [42]$$

and

$$\frac{l_N'}{n_2'} = \frac{c - (n_1/n_2')}{a} \qquad [43]$$

When the two indices are identical, these relations are identical to those for the points H and H'. That is why the rays from P to H and H' to P' in Figure 3 are parallel.

Compound Lenses

A great advantage of the paraxial matrix method is the ease of dealing with systems having a large number of lenses. To start simply, consider the cemented doublet of Figure 10: a pair of lenses for which surface 2 of the first element and surface 1 of the second element match perfectly. The parameters of this doublet are given in the manner shown below.

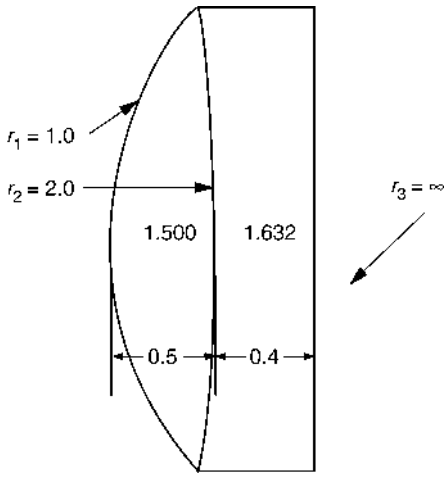

Figure 10 Specification for a cemented doublet.

It is understood that the first entry under r is r_1, the second entry is r_2, and so on. Note that this doublet has only three surfaces; if the two parts were separated by an air space, producing an air-spaced doublet, then there would be four surfaces to specify. The values of the indices n' and the spacings t' are placed between the values of r. The constants of the doublet are then

r	n'	t'
1.0		
	1.500	0.5
−2.0		
	1.632	0.4
∞		

[44]

Note that surface 3, which is flat, has a radius of infinity. Numbering the vertices in the usual manner, the system matrix is

$$S_{31} = R_3T_{32}R_2T_{21}R_1 \qquad [45]$$

where

$$k_2 = (n_2' - n_2)/r_2 = (1.632 - 1.500)/-2.0 \qquad [46]$$

Approximations for Thin Lenses

An advantage of paraxial matrix optics is the ease of obtaining useful relations. For example, combining the several expressions for the Gaussian constant a, we obtain:

$$a = -\frac{n_1}{f} = \frac{n_2'}{f'} = k_1 + k_2 - \frac{k_1k_2t_1'}{n_1'} \qquad [47]$$

and using the definitions of k_1 and k_2 with the lens in air leads to

$$\frac{1}{f'} = (n_1' - 1)\left[\frac{1}{r_1} - \frac{1}{r_2} + \frac{(n_1' - 1)t_1'}{n_1'r_1r_2}\right] \qquad [48]$$

which is the lensmakers' equation. It tells how to find the focal length of a lens from a knowledge of its material and geometry. This derivation is much more direct than what you will usually find. This equation becomes simpler when we are dealing with thin lenses – those for which the third term in brackets can be neglected by assuming that the lens thickness

is approximately zero. Then

$$\frac{1}{f'} = (n_1' - 1)\left[\frac{1}{r_1} - \frac{1}{r_2}\right] \qquad [49]$$

which is the form often seen in texts. We then realize that the original form is valid for lenses of any thickness, but the thin lens form can be used if the spacing is much less than the two radii. It is also seen that when the thickness is approximately zero, then

$$b = c = 1 \qquad [50]$$

and the unit planes have locations

$$l_H = l_H' = 0 \qquad [51]$$

Thus, they are now coincident at the center of the lens. A ray from any point on the object would pass undeviated through the geometrical center of the lens, as is often shown in the elementary books. These new values of the Gaussian constants give a system matrix of the form:

$$S_{21} = \begin{pmatrix} 1 & -1/f' \\ 0 & 1 \end{pmatrix} \qquad [52]$$

This matrix contains a single constant, the focal length of the lens, so that the preliminary design of an optical system is merely a matter of specifying the focal length of the lenses involved and using the resulting matrices to determine the object-image relationship. We also note that this matrix looks like a refraction matrix; the non-zero element is in the upper right-hand corner. This implies that the refraction–translation–refraction procedure that actually occurs can be replaced, for a thin lens, by a single refraction occurring at the coinciding unit planes. Ophthalmologists take advantage of the thin lens approximation by specifying focal lengths in units called diopters. The reciprocal of the focal length in meters determines its refraction power in diopters. For example, a lens with $f' = 10$ cm will be a 10 diopter lens. If two lenses are placed in contact, the combined power is the sum of the individual powers, for multiplying matrices of the above form gives the upper right-hand element as $(-1/f_1' - 1/f_2')$.

The Paraxial System for Design or Analysis

The material given in this article can serve as the basis for an organized way of taking the specifications of an optical system and using them to gain a full understanding of this system. We shall demonstrate it with a practical example which has some interesting features. Chemical engineers have long known that the index of refraction of a liquid can be determined by observing the empty bore of a thick glass tube with a telescope having a calibrated eyepiece and then measuring the magnification of the bore when the liquid flows through it. Consider a liquid with an index of 1.333 and a tube with bore radius of 3 cm, outer radius of 4 cm, and glass of index equal to 1.500. The first step is to specify the parameters of the optical system. These are the radii $r_1 = -3$, $r_2 = -4$ and the thickness $t_1' = 1$, regarding the tube as a concentric lens with an object to its left; the indices $n_1 = 4/3$ (fractions are convenient), $n_1' = 3/2$, $n_2' = 1$. Calculating the elements of the two refraction matrices and the translation matrix, the system matrix is

$$S_{21} = \begin{pmatrix} 1 & -1/8 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2/3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1/18 \\ 0 & 1 \end{pmatrix} \qquad [53]$$

Multiplying, the Gaussian constants are

$$a = 2/27, \quad b = 11/12, \quad c = 28/27, \quad d = -2/3 \qquad [54]$$

An important check is to verify that $bc - ad = 1$, as is the case here. The expressions given previously for the positions of the cardinal points then lead to the values

$$\begin{aligned} &l_H = 3/2, \qquad l_H' = 1/2, \qquad l_F = -16\ 1/2, \\ &l_F' = 14, \qquad l_N = -3, \qquad l_N' = -4 \end{aligned} \qquad [55]$$

These points are shown in Figure 11. Although the same scale has been used for both horizontal and vertical distances, this is usually not necessary; the vertical scale, which merely serves to verify the magnification, can be arbitrary. The entire liquid column is the object in question, but it is simpler to use a vertical radius as the object. Then the image, generated as described in all the previous examples,

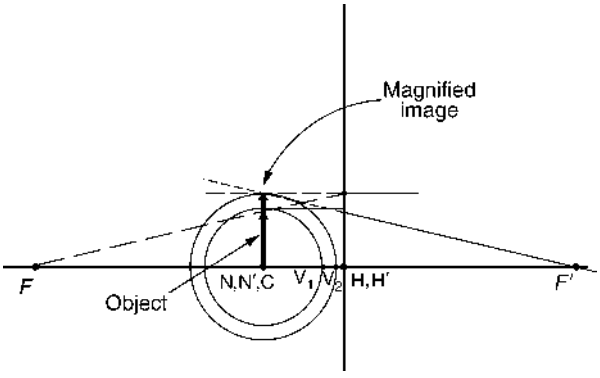

Figure 11 Ray tracing diagram for a thick-walled glass tube.

coincides with the object and it is enlarged, erect, and virtual. We confirm the image location shown by using the generalized Gaussian lens equation, giving

$$t' = [(-2/3) + (28/27)(-3)/(4/3)]/[(11/12) + (2/27)(-3)/(4/3)] = -4 \qquad [56]$$

and this agrees with the sketch. The denominator of this expression was shown to be the reciprocal of the magnification $1/m$ with a value of 3/4, and m itself can be calculated from t' as $c - at' = 4/3$; this is another important check on the accuracy of the calculation; the sketch obeys this conclusion, and the purpose of this analysis is confirmed.

Reflectors

To show how reflecting surfaces are handled with matrices, consider a plane mirror located at the origin and normal to the z-axis (Figure 12). The ray which strikes it at an angle α_1 leaves at an equal angle, resulting in specular reflection. Since $k = 0$ for a flat surface, the refraction matrix reduces to the unit matrix and by eqn [9]:

$$n_1'\alpha_1' = n_1\alpha_1 \qquad [57]$$

This contradicts Figure 12, since α_1 and α_1' are opposite in sign.

If, however, we specify that

$$n_1' = -n_1 \qquad [58]$$

then the difficulty is removed. Hence, reflections involve a reversal of the sign on the index of refraction, and two successive reflections restore the original sign. For a mirror, the refraction power is

$$k_1 = \frac{n_1' - n_1}{r_1} = \frac{-1-(1)}{r_1} = -\frac{2}{r_1} \qquad [59]$$

The system matrix for a single refracting surface thus becomes

$$S_{11} = R_1 = \begin{pmatrix} 1 & -k_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2/r_1 \\ 0 & 1 \end{pmatrix} \qquad [60]$$

with

$$a = -2/r_1, \qquad b = c = 1, \qquad d = 0 \qquad [61]$$

The connection between object and image can be expressed as

$$\begin{pmatrix} n_1'\alpha_1' \\ x_1' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ t'/n_1' & 1 \end{pmatrix} \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -t/n_1 & 1 \end{pmatrix} \begin{pmatrix} n_1\alpha_1 \\ x_1 \end{pmatrix} \qquad [62]$$

and using the same procedure that was applied to eqns [28]–[31], the locations of the six cardinal points are

$$l_F = -n_1 b/a, \quad l_F' = n_1' c/a$$
$$l_H = n_1(1-b)/a, \quad l_H' = n_1'(c-1)/a \qquad [63]$$
$$\frac{l_N}{n_1} = \frac{(n_1'/n_1) - b}{a}, \quad \frac{l_N'}{n_1'} = \frac{c - (n_1/n_1')}{a}$$

so that for the spherical mirror:

$$l_F = \frac{(-1)(1)}{-2/r_1} = \frac{r_1}{2} = l_F' \qquad [64]$$

$$l_H = l_H' = 0 \qquad [65]$$

and

$$l_N = \frac{-1-1}{-2/r_1} = r_1 \qquad [66]$$

but

$$\frac{l_N'}{-1} = \frac{1-(-1)}{-2/r_1}, \quad l_N' = r_1 \qquad [67]$$

Since r_1 is negative, these equations show that the unit points lie on the vertex, the foci coincide halfway between the center of curvature and the

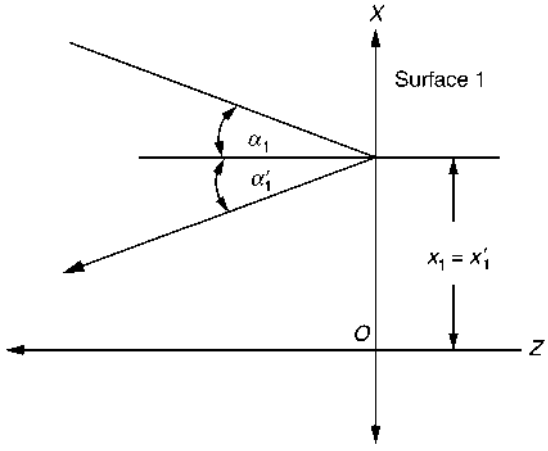

Figure 12 Specular reflection.

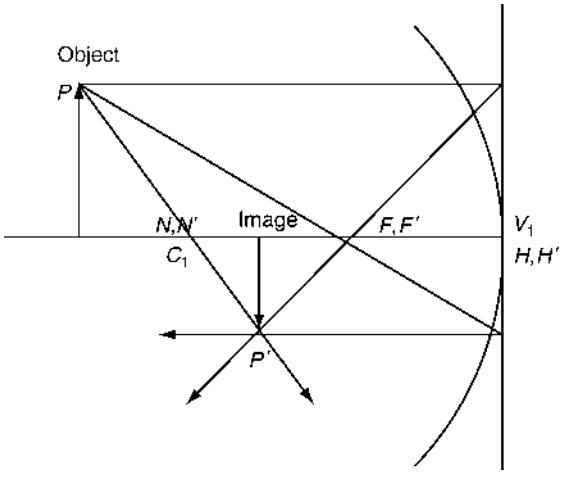

Figure 13 Ray tracing for a concave mirror.

vertex, and the nodal points are at this center. The perfect focusing is a consequence of the paraxial approximation. The true situation is presented in **Geometrical Optics:** Aberrations. Ray tracing for this mirror uses the procedure developed for lenses, which applies to optical systems of any complexity. Figure 13 shows an object to the left of the center of curvature. The ray from P parallel to the axis goes to H' and then through F', while the ray through F goes to H and then becomes parallel to the axis. Their intersection at P' produces a real, inverted image. The third set of rays represents something different: it requires a knowledge of the nodal point locations. The ray from P to N should be parallel to the ray from N' to P', by definition; in this example, they meet this requirement by being colinear.

As an example of the power of matrix optics to simplify the design or analysis of an optical system involving lenses and mirrors, consider an object that is 15 units to the left of a converging lens, with focal length $f' = 10$. A concave mirror, of radius $r_1 = 16$, is 20 units to the right of the lens. The refraction power of the mirror is $k_1 = (-1 - 1)/-16 = 1/8$ and using the thin lens form of the system matrix, the combined matrix is

$$\begin{pmatrix} 1 & -1/8 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 20 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1/10 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -3/2 & 1/40 \\ 20 & -1 \end{pmatrix} \quad [68]$$

This gives the constants a, b, c, and d, from which we find that the image is $-4\ 4/9$ units to the left of the mirror, its magnification is $-8/9$, and it is inverted. Compare the simplicity of this procedure with the usual way of doing this calculation, which involves first finding the image in the lens and then using that image as a virtual object.

Conclusion

It has been shown that the use of paraxial matrices provides a simple approach to an understanding of the kind of optical systems that provide perfect images. Further details and derivations of expressions given here will be found in the Further Reading below.

See also

Geometrical Optics: Aberrations.

Further Reading

Brouwer W (1964) *Matrix Methods in Optical Instrument Design*. New York: WA Benjamin.

Nussbaum A (1998) *Optical System Design*. Upper Saddle River, NJ: Prentice-Hall.

Aberrations

A Nussbaum, University of Minnesota, Minneapolis, MN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The article **Geometrical Optics:** Lenses and Mirrors, describes image formation by components which are optically perfect. The images may be magnified or reduced and they may be erect or inverted, but they are otherwise faithful replicas of the object. This behavior is a consequence of ray tracing based on an approximate or paraxial form of Snell's law which we write as

$$n\theta = n'\theta' \quad [1]$$

where n and n' are the indices of refraction on either side of an interface and θ and θ' are the angles of incidence and refraction. If, however, the following exact form of Snell's law:

$$n \sin\theta = n' \sin\theta' \quad [2]$$

governs the ray behavior, as is the case when the rays have large inclinations with respect to the symmetry axis, then it is necessary to devise a more involved ray

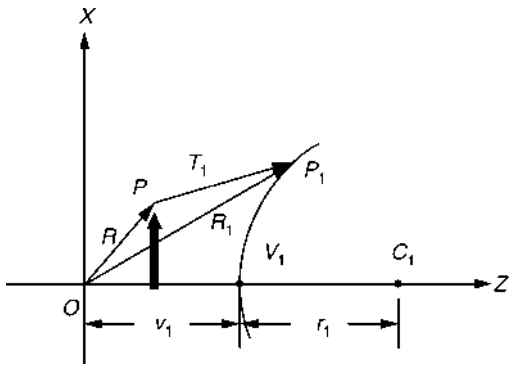

Figure 1 Ray tracing in the meridional plane.

tracing procedure. Figure 1, which shows the first surface of a lens, will be used to explain how this procedure works. A ray leaves the object point P and strikes the lens surface at the point P_1. Rays which lie completely in the plane ZOX are said to be meridional; this is the plane which passes through the symmetry axis, just as meridians on the Earth's surface are determined by planes through the geographic axis. By regarding the various distances in the figure as vectors, it is possible to start with the initial coordinates of the ray and with the direction-cosines which specify its slope and find the coordinates of the point P' where the ray strikes the lens surface. A derivation of the equation governing this process will be found in the text of A Nussbaum (see Further Reading). This derivation involves only elementary algebra and vector analysis, and will be found simpler than the usual textbooks treatment of ray tracing procedures. The equation is

$$T_1 = \frac{F}{-E + \sqrt{E^2 - c_1 F}} \qquad [3]$$

where

$$E = Bc_1/2 = c_1[(z - v_1)N + xL] - N \qquad [4]$$

and

$$F = Cc_1 = c_1\left[(z - v_1)^2 + x^2\right] - 2(z - v_1) \qquad [5]$$

Knowing the coordinates z and x of the starting point P of the ray, as well as its starting cosines N and L, plus the curvature c_1 or radius $r_1 = 1/c_1$ of the lens surface and the location v_1 of the surface's vertex, we can readily calculate the length T_1 of the ray. Then its components on the two axes give the coordinates of P'. Next, we need to find the slope of the ray after it is refracted at the lens surface, and this can be calculated from Snell's law combined with the behavior of the incident and refracted rays regarded as vectors. Again we invoke the reference cited above for a derivation of the equations giving the value of the direction-cosines after refraction; these expressions are

$$n_1' N_1' = K_1/c_1 + n_1 N_1 - K_1 z_1 \qquad [6]$$

and

$$n_1 L_1' = n_1 L_1 - K_1 x_1 \qquad [7]$$

where K_1 is known as the refracting power and has the definition

$$K_1 = c_1(n_1' \cos\theta' - n_1 \cos\theta) \qquad [8]$$

The ray can now be traced to the next surface knowing the lens thickness, and then to the image plane. These two sets of equations, when incorporated into a computer program of about two dozen lines, will trace meridional rays to any desired accuracy through an optical system containing an arbitrary number of lenses and spherical reflectors. As an example, consider a symmetric double convex lens with radii of 50 units at surface 1, −50 units at surface 2, a thickness of 15 units, and an index of 1.50. Rays parallel to the axis – one at 0.2 units above the axis, a second one at 2 units above the axis, and a third one at 20 units above the axis – are started well to the left and traced to the paraxial focal plane, which is at 47.368 units from the second surface. The ray that is 0.2 units above the axis will cross the focal plane at the axis; this is a paraxial ray, passing through the paraxial focal point F'. The ray which is 2 units above the axis crosses the focal plane slightly below the focal point, and is therefore not quite paraxial, while the ray which starts at 20 units above the axis falls well short of the focal point, intersecting the axis before it reaches the focal plane. This behavior indicates that the lens possesses a defect known as spherical aberration and these calculations imply that spherical aberration should be defined as the failure of nonparaxial rays to conform to the behavior of paraxial rays. This definition does not appear in optics texts, and the few definitions that are given appear to be incorrect. As clearly indicated in **Geometrical Optics:** Lenses and Mirrors, the focal point of a lens is strictly a paraxial concept and it is a constant, obtainable from the parameters of the lens or lens system. It is meaningless to speak of a focal point as defined by a pair of nonparaxial rays; this infinite set of ray-pairs determines a corresponding number of intersection points, all of which fall short of the true focal point F'. This situation is nicely illustrated by the next figure, which shows a large number of parallel rays, the central ones being paraxial and the others behaving differently. The rays which are very close to the axis will meet as

expected at the focal point F'. The rays a little farther out – the intermediate rays – will cross the axis just a little to the left of F', and those near the edge of the lens – the *marginal* rays – will fall very short, as indicated. Rotating this diagram about the axis, we realize that spherical aberration produces a very blurry circular image at the paraxial plane. The three-dimensional envelope of the group of rays produced by this rotation is known as the caustic surface and the narrowest cross-section of this surface, slightly to the left of the paraxial focal plane, is called the circle of least confusion. If this location is chosen as the image plane, then the amount of spherical aberration can be reduced somewhat. Other methods of improving the image will be considered below. Figure 2 was obtained by adding a graphics step to the computer program described above. Figure 3 shows how to define spherical aberration quantitatively. Two kinds of aberration are specified in this figure. The place where the ray crosses the axis, lying to the left of F', is at a distance, measured from the paraxial focus, called the longitudinal spherical aberration (LSA), while the distance below the axis at the focal plane is the transverse spherical aberration (TSA). To reduce this aberration, note in Figure 2 that the amount of refraction at the two surfaces of the lens for marginal rays is not the same. Equalizing the refraction at the two surfaces will improve the situation. Suppose we alter the lens while keeping the index, thickness, and focal length constant; only the radii will change. This can be done if the new radii satisfy the paraxial equation giving the focal length in terms of the lens constants. The process of varying the shape of lens, while holding the other parameters constant, is called *bending the lens*. This can be easily accomplished by a computer program, obtaining what are known as optimal shape lenses, which are commercially available. To study their effect on spherical aberration, define the shape factor σ of a lens as

$$\sigma = \frac{r_2 + r_1}{r_2 - r_1} \qquad [9]$$

from which $\sigma = 0$ for a symmetric lens ($r_1 = -r_2$). Figure 4a shows the monitor screen for a symmetric lens, and Figure 4b illustrates the result of lens bending. This lens has had the curvature of the first surface raised and the second surface has been flattened to keep the focal length constant. Let us now consider what lens designers can do about spherical aberration. Brouwer in his book (see Further Reading) gives the specifications for a cemented doublet consisting of a double convex lens followed by a lens with two concave surfaces. The longitudinal spherical aberration of the front lens alone varies as shown in Figure 5. The way that the designer arranged for the marginal rays to meet at the paraxial focus, rather than falling short, was to recognize that a diverging lens following the front element would refract the rays away from – rather than closer to – the axis. When this second component is added, we would expect the behavior shown in Figure 6. Note that the value of LSA for the doublet has been enormously reduced, as indicated by the horizontal scale change. Spherical aberration can be reduced in more elaborate systems as part of the design process. Telescopes, in particular, make use of corrector plates which are placed in front of the large mirrors.

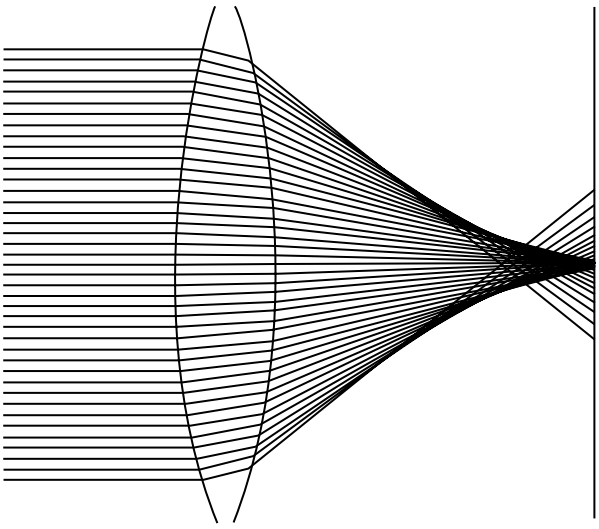

Figure 2 Spherical aberration for parallel meridional rays.

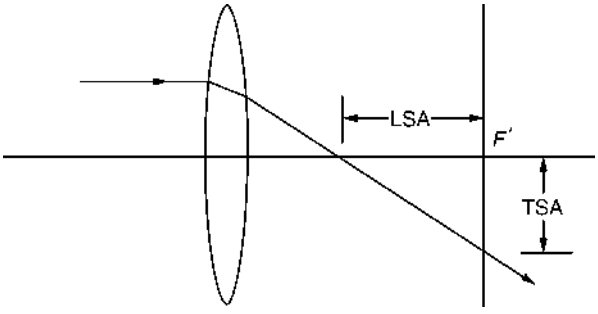

Figure 3 Definitions of transverse and longitudinal spherical aberration.

Coma

We have seen how to trace nonparaxial, meridional rays and found that spherical aberration appears because the intermediate and marginal rays do not behave like the central rays. We now wish to deal

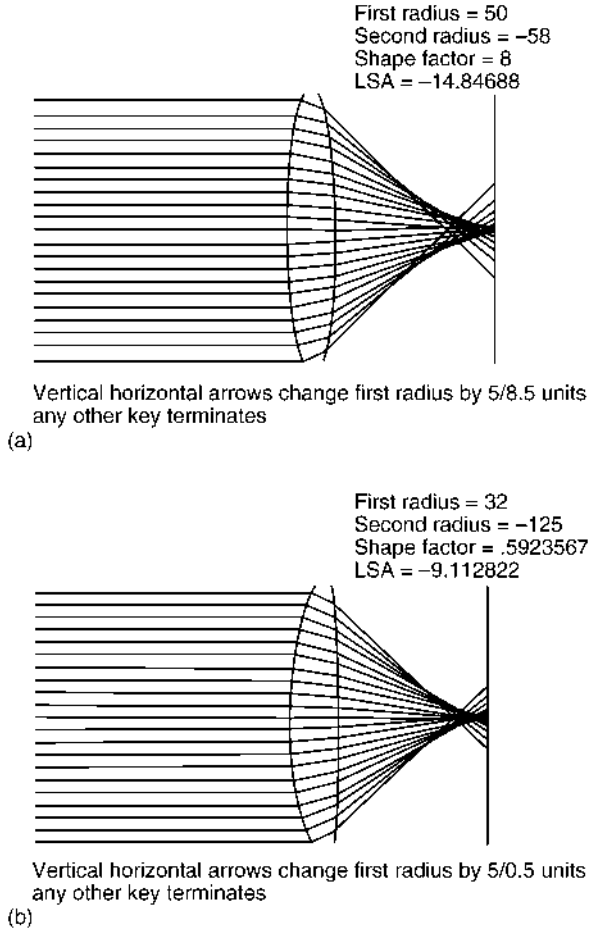

Figure 4 Lens bending as performed numerically.

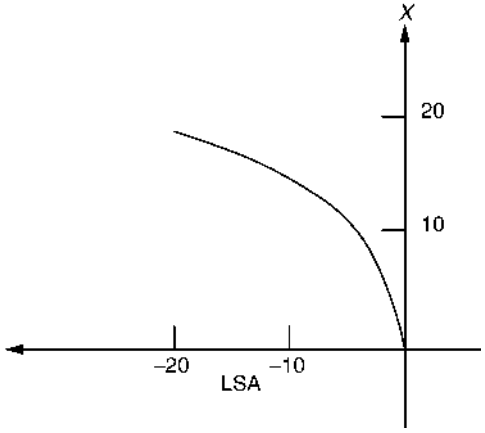

Figure 5 Spherical aberration for a single lens.

with rays that do not lie in a meridional plane; these are nonmeridional or skew rays. Calculating the behavior of such a ray would appear to be a complicated problem, but it turns out to be a simple extension of what has already been accomplished. Let a skew ray start at an arbitrary point in space with coordinates x, y, and z and let it have direction

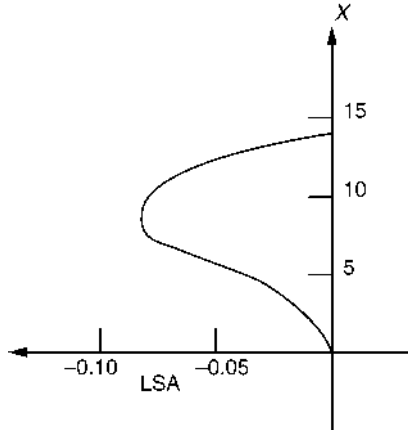

Figure 6 Reduction of spherical aberration by a doublet.

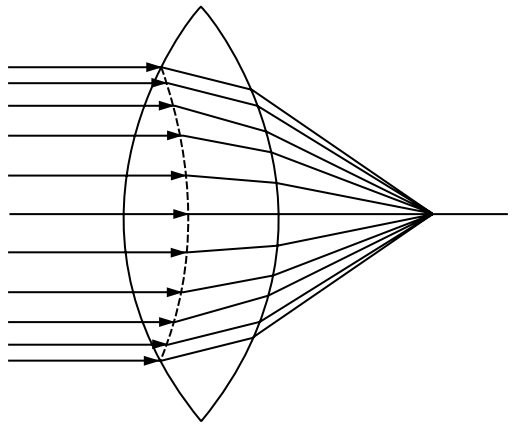

Figure 7 Configuration used to illustrate coma.

cosines L, M, and N with respect to OX, OY, and OZ. The points P and P_1 in Figure 1 are now out of the plane ZOX and when the vectors in this figure are expressed as the sum of three, rather than two, components, the equations given above for meridional rays acquire terms in y and M which have the same form as those for x and L. Now trace a set of skew rays chosen to lie on a cylinder (Figure 7) which is centered about the z-axis of a lens, so that all the rays are meridional. These rays are fairly far from the axis, so that they are also nonparaxial, but they all meet at a common image point, forming a cone whose apex is this image point. Then give this cylinder a downward displacement while holding fixed the intersection point of each ray with the dotted circle on the front of the lens. The tilting of the cylinder changes all its rays – except the top and the bottom rays – from meridional to skew. In addition, the cone on the image side will then tilt upwards; it should not change in any other way if the skew rays continue to meet at a well-defined apex. But this is not what happens in a simple lens, as will now be explained.

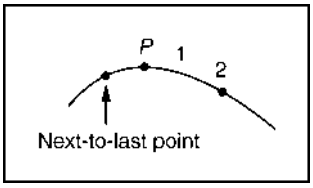

Figure 8 Starting a coma plot.

Let the top and bottom meridional rays meet at a point P in image space (Figure 8) and use this point to determine the location of an image plane. The skew ray just below the uppermost ray in Figure 7 will pass through this plane fairly close to the intersection point. Let us guess (and confirm later) that it will pass through a point slightly to the right of P and a little below it, as we observe this plane from the lens position. This is point 1 in Figure 8. The next skew ray on the front of the lens will strike the image plane still farther below and to the right, producing point 2 in the figure. As we continue to trace these rays down the front half of the circle, we realize that a ray almost at the bottom of Figure 7 will have to be very close to the lower meridional ray and that its image point is just below but to the left of P; this is the point labeled 'next-to-last' in Figure 8. All of these points join to form a closed curve. The back halves of the dotted circle and of the ray cylinder are not shown in the figure, but since we are again starting at P in Figure 8, the pattern will repeat itself. That is, we get a second closed curve which – by symmetry – should be identical to the first curve. This is a most unusual effect; double-valued phenomena are quite rare. To confirm this prediction, we use the skew ray equations, and trace the pattern formed by a set of concentric cylinders of different sizes. The appearance of the coma pattern generated by the computer is shown in Figure 9. Each half-circle on the front side of the lens produces a closed figure (an approximation to a circle) as an image. If coma were the only aberration (when it is called ideal coma), the plots would be the perfect, coinciding circles of Figure 10. The comet-like shape explains where the name comes from. We can picture the generation of these circles in the way shown in this figure. The ray through the geometrical center of the lens produces a point in the paraxial focal plane. The top and bottom rays (the meridional rays) from the circles of increasing size produce points which define image planes that are successively closer to the lens; this is a consequence of spherical aberration. Moving the top ray parallel to itself and along the half-circle generates the image plane pattern. The lines tangential to the coma circles make an angle of 60° with each other, as can be seen by measurement on the computer output. Most optics

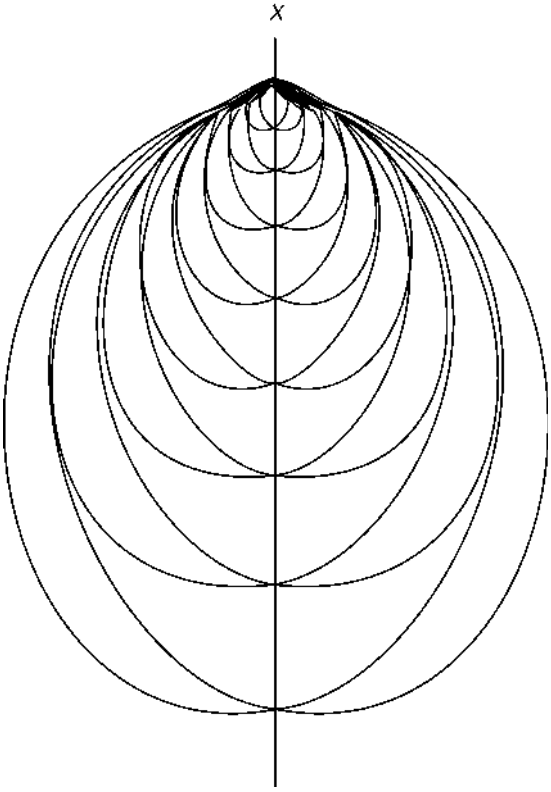

Figure 9 Coma as calculated numerically.

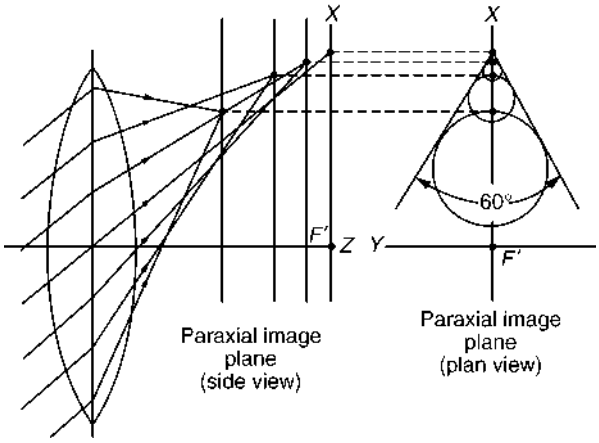

Figure 10 Customary way of illustrating coma.

books look at coma in a way different from that given here. Coma has been defined by WJ Smith (see Further Reading) as the variation of magnification with aperture, analogous to the definition of spherical aberration as the variation of focal length with aperture. And as was previously mentioned, focal length and magnification are purely paraxial concepts, not applicable to the description of aberrations. The definition that comes from the arguments

given above is that coma is the failure of skew rays to match the behavior of meridional rays, just as spherical aberration is the failure of meridional rays to match the behavior of paraxial rays. Most texts show coma in its ideal form.

Astigmatism

We have shown that spherical aberration is the failure of meridional rays to obey the paraxial approximation and coma is the failure of skew rays to match the behavior of meridional rays. To see the connection between these two aberrations, consider the object point P of Figure 11. This very complicated diagram can actually be very helpful in understanding the third of the five geometrical aberrations. Let P be the source of a meridional fan: this is the group of rays with the top ray labeled PA and the bottom ray is PB. If the lens is completely corrected for spherical aberration, this fan will have a sharp image point P_T' lying directly below the z-axis. Now let P also be the source of a fan bounded by the rays PC and PD. This fan is at right-angles to the other one; we can think of these rays as having the maximum skewness possible. If the lens has no coma, this fan also will produce a sharp image point P_s' which, for a converging lens, will be farther from the lens than the tangential focal point. In other words, even though the lens has been fully corrected for both aberrations, the two corrections will not necessarily produce a common image. In the figure, the meridional fan is called a tangential fan and the skew fan is called a sagittal fan; these are the terms commonly used by lens designers. The failure of the sagittal and tangential rays to produce a single image in a lens corrected for both spherical aberration and coma is known as astigmatism. Astigmatism, coma, and spherical aberration are the point aberrations that an optical system can have. We have already noted that spherical aberration is the only one that can be associated with a point on the z-axis; the others require an off-axis object. Astigmatism is very common in the human eye; to see how it shows up, look at the sagittal image point P_s' of Figure 11. This point is a distance z_s, from the x,y-plane. A plane through this point will have a line image on it, the sagittal line, due to the tangential fan, which has already come to a focus and is now diverging. The converse effect, producing a tangential line, occurs at the other image plane. If we locate an image plane halfway between the two, then both fans contribute to the image and in the ideal case it will be a circle. As the image plane is moved forwards or backwards, these images become ellipses and eventually reduce to a sagittal or a tangential line, as shown in Figure 12. Because there are two different

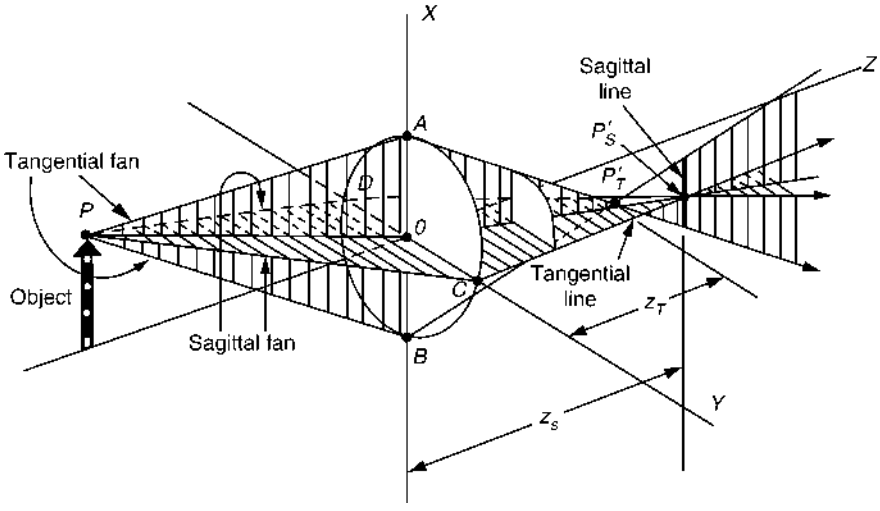

Figure 11 Tangential and sagittal fans displaying astigmatism.

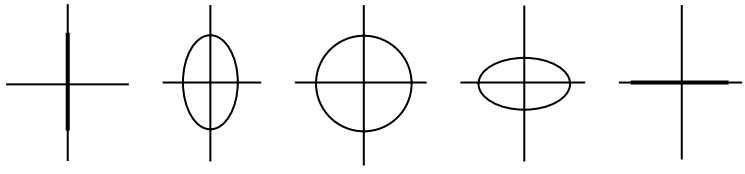

Figure 12 Astigmatic image patterns.

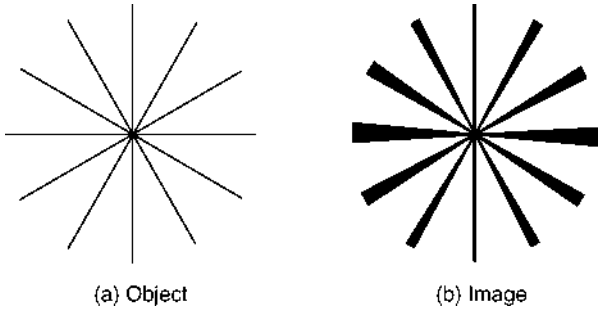

Figure 13 Image degraded by astigmatism.

image planes, an object with spokes (Figure 13) will have an image for which the vertical lines are sharp and the others get gradually poorer as they become more horizontal, or vice versa, depending on which image plane is chosen. Eye examinations detect astigmatism if the spokes appear to go from black to gray. This example of the effect of astigmatism was often based on a radial pattern of arrows, which explains the origin of the word sagittal, derived from the Latin 'sagitta' for arrow.

Curvature of Field and Distortion

Having covered the three point aberrations, there are two lens defects associated with extended objects. If we move the object point P in Figure 11 closer to or farther from the z-axis, we would expect the positions of the tangential and sagittal focal planes to shift, for it is only when the paraxial approximation holds that these image points are independent of x. Hence, we obtain the two curves of Figure 14, which shows what astigmatism does to the image of a two-dimensional object. If the astigmatism could be eliminated, the effect would be to make these curved image planes coincide, but we have no guarantee that the common image will be flat, or paraxial. The resulting defect is called Petzval curvature or curvature of field. For a single lens, the Petzval surface can be flattened by a stop in the proper place, and this is usually done in inexpensive cameras. Petzval curvature is associated with the z-axis. If we take the object in Figure 11 and move it along the y-axis, then all rays leaving it are skew and this introduces distortion, the aberration associated with the coordinates normal to the symmetry axis. Distortion is what causes vertical lines to bulge outward (barrel distortion) or inward (pincushion distortion) at the image plane, as shown in Figure 15. A pinhole camera will have no distortion, since there are no skew rays, and a single thin lens with a small aperture will have very little. Placing a stop near the lens to reduce astigmatism and curvature of field introduces distortion because, as

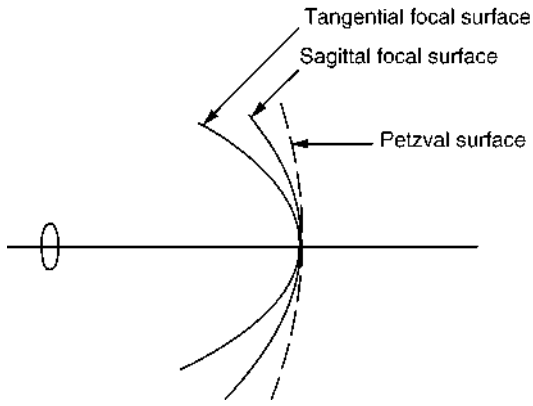

Figure 14 Petzval of field curvature.

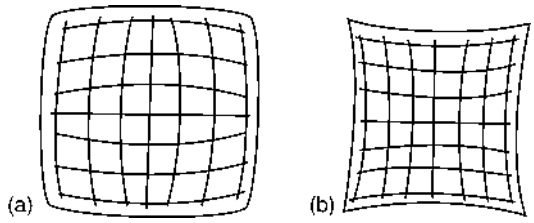

Figure 15 Barrel and pincushion distortion.

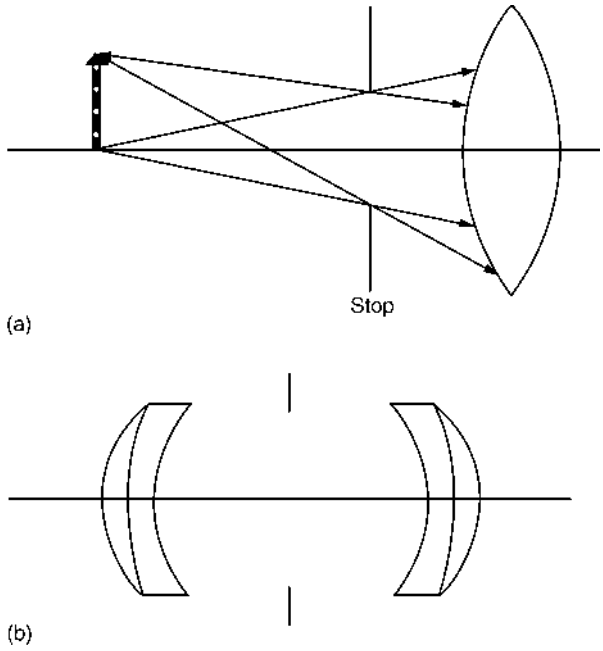

Figure 16 (a) Distortion-creating stop. (b) Symmetric-component lens that reduces distortion.

shown in Figure 16a, the rays for object points far from the axis are limited to off-center sections of the lens. The situation in this figure corresponds to barrel distortion; placing the stop on the other side of the lens produces pincushioning. Distortion can be

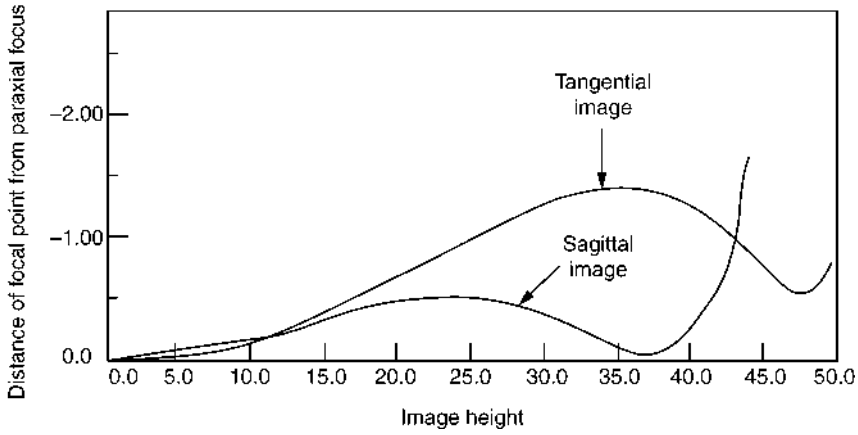

Figure 17 Reduction of astigmatism through design.

therefore reduced by a design which consists of two identical groupings (Figure 16b), with the iris diaphragm placed between them; the distortion of the front half cancels that of the back half. Figure 17 shows the output for a calculation using approximate formulas rather than exact ray tracing. Note that the designer has arranged for astigmatism to vanish at the lens margin. This example indicates the nature of astigmatism in a system for which the coma and the spherical aberration are low but not necessarily zero; it is the failure of the tangential and sagittal image planes to coincide.

Nonspherical Surfaces

So far, we have considered optical systems that use only spherical or plane reflecting and refracting surfaces. However, many astronomical telescopes use parabolic and hyperbolic surfaces. In recent years, nonspherical or aspheric glass lenses have become more common because they can be mass produced by computer-controlled machinery and plastic lenses can be molded with very strict tolerances. Ray tracing procedures given above can easily be extended to the simplest kind of aspheric surfaces: the conic sections. Iterative procedures may be applied to surfaces more complicated than the conic sections. A well-known example is the curved mirror; a spherical reflecting surface will have all the aberrations mentioned here, but a parabolic mirror will bring parallel rays to a perfect focus at the paraxial focal point, thus eliminating the spherical aberration. Headlamp reflectors in automobiles are a well-known example. A plano-convex lens with this defect corrected will be one whose first surface is flat and whose second surface is a hyperbola with its eccentricity equal to the index of refraction of the glass composing the lens. One form of astronomical telescope, which has a large primary concave mirror to collect the light and reflect it back to a smaller secondary convex mirror, is the Ritchey–Chrétien design of 1924. Both mirrors are hyperbolic in shape, completely eliminating spherical aberration. This is the design of the Hubble space telescope. The two-meter diameter primary mirror, formed by starting with a spherical mirror and converted to a hyperbola by computer-controlled polishing, was originally fabricated incorrectly and caused spherical aberration. From a knowledge of the computer calculations, it was possible to design, build, and place in orbit a large correcting optical element which removed all the aberrations from the emerging image.

Conclusion

The three-point aberrations and the two extended object aberrations have been defined and illustrated. Simple ways of reducing them have been mentioned, but the best way of dealing with aberrations is in the process of lens design. In addition, there is a completely separate defect known as chromatic aberration. All the above calculations are based on ray tracing for a single wavelength (or color) of visible light. However, the index of refraction of a transparent material varies with the wavelength of the light passing through it, and a design good for one value of the wavelength will give an out-of-focus image for any other wavelength. It has been known for many years that an optical system composed of two different kinds of glass will provide a complete correction for two colors – usually red and blue – and other colors will be partially corrected. Such lenses are said to be achromatic. For highly demanding applications, such as faithful color reproductions

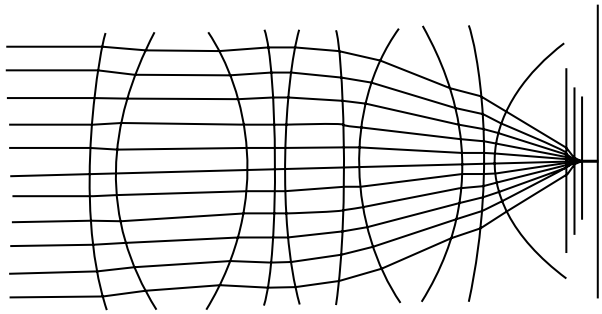

Figure 18 High-power high-resolution microscope objective.

of works of art, it is possible to design elaborate and expensive lenses which correct for yellow light as well. This kind of optical system is said to be apochromatic. With perfect imaging at the two ends and the middle of the visible spectrum, the color error elsewhere is virtually nonexistent. As an example of a design which is free of all aberrations – geometric and chromatic – Figure 18 shows the ray trace for a high-power aberration-free and apochromatic microscope objective. Note that the trace shows parallel rays brought to a focus inside the cover glass; in use, the light source would be a point on the slide and the light would converge at the eyepiece.

See also

Geometrical Optics: Lenses and Mirrors; Prisms.

Further Reading

Brouwer W (1964) *Matrix Methods in Optical Instrument Design*. New York: WA Benjamin.

Laikin M (1990) *Lens Design*. New York: Marcel Dekker.

Nussbaum A (1998) *Optical System Design*. Upper Saddle River, NJ: Prentice-Hall.

Nussbaum A and Phillips RA (1976) *Contemporary Optics for Scientists and Engineers*. Englewoods Cliffs, NJ: Prentice Hall.

Smith WJ (1990) *Modern Optical Engineering*, 2nd edn. New York: McGraw-Hill Book Co.

Prisms

A Nussbaum, University of Minnesota, Minneapolis, MN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Prisms are solid structures made from a transparent material (usually glass). Figure 1 shows the simplest type of prism – one that has many uses. The vertical sides are two identical rectangles at right angles to each other and a larger third rectangle is at 45° to the other two. The top and bottom are 45°–45°–90° triangles. The most common applications of prisms are:

1. To bend light in a specified direction
2. To fold an optical system into a smaller space
3. To provide proper image orientation
4. To combine or split optical beams, using partially reflecting surfaces
5. To disperse light in optical instruments such as spectrographs.

Single Prisms as Reflectors

A prism like that of Figure 1 can act as a very effective reflector; it is more efficient than a mirror since there is no metal coating to absorb light. Figure 2 shows a light ray entering the prism at right angles to one of the smaller sides, reflected at the larger side, and emerging at right angles to the other small side. The path shown is determined by Snell's law. This law has the form:

$$n \sin \theta = n' \sin \theta' \qquad [1]$$

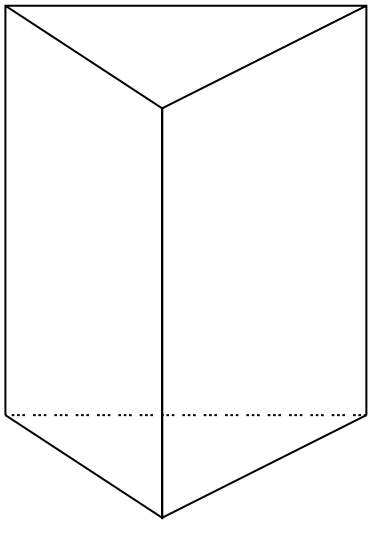

Figure 1 A 45–45–90 degree prism.

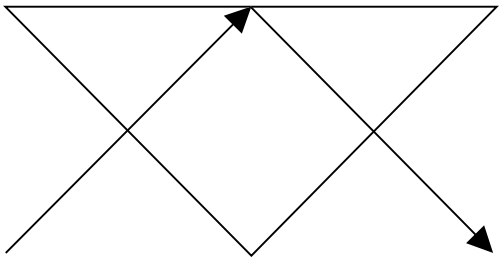

Figure 2 Total internal reflection by a prism.

This indicates that a ray of light making an angle θ to the normal at the interface between media of indices of refraction n and n', will be refracted and the angle to the normal becomes θ'. Notice that the light ray strikes the surface where the reflection takes place at an angle of 45° to the normal; this angle is too large to permit it to emerge from the prism after refraction. To understand why, we recognize that when the ray is able to leave the glass, Snell's law requires that the emerging angle be larger than the incident angle. The largest possible incident angle – known as the critical angle – will correspond to an emerging angle of 90°; that is, this ray will be parallel to the interface. For an index of 1.50, as an example, the critical angle θ_c satisfies Snell's law in the form:

$$1.5 \sin \theta_c = 1.0 \sin 90° \qquad [2]$$

from which $\theta_c = 41.8°$. The ray shown in the figure will therefore remain in the glass. Furthermore, since the indices appearing on both sides of Snell's law are now those of the glass, then the incident and refracted angles must be identical, and the prism reflects the ray in the same way that a mirror would. This phenomenon is known as total internal reflection.

This prism has an important application as an optical component of a fingerprint machine, whose action depends on modifying the total internal reflection which normally occurs. Figure 3 shows the prism and the reflecting behavior of Figure 2, with a person's finger pressed against the long side. When the ray strikes a place on the surface corresponding to a valley in the fingerprint pattern, the index at this region is that of air, and the ray is reflected in the usual manner. But a ray striking a portion of the prism which lies immediately below a raised section of the skin will make an angle with the normal which is now smaller than the critical angle, since the index is much greater than that of air, so that the ray will cross the interface and be absorbed by the finger. The reflected beam, with the fingerprint pattern sharply reproduced, is sent to a detector and the resulting signal goes to a computer, where it is processed and used to drive a printer. The fingerprint cards generated by this kind of equipment are sharper than those done by the traditional smeared-ink method. It is possible to observe the phenomenon just described by looking at your fingers while holding a glass of water – the fingerprints will show up in an enhanced manner. In addition to the better quality, the manufacturer of one version of this equipment has incorporated a feature in their computer program which can detect attempts by the person being fingerprinted to degrade the image.

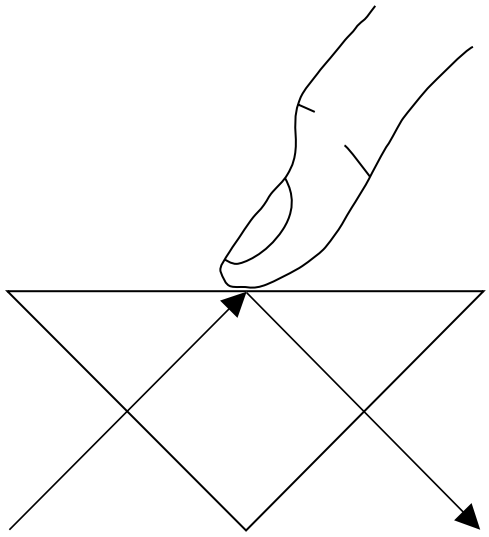

Figure 3 Use of a prism to produce a fingerprint.

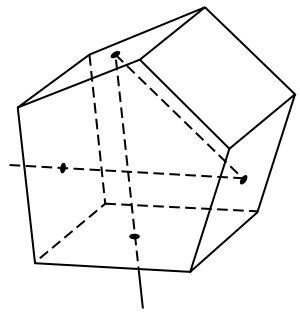

Figure 4 A pentaprism.

A more elaborate single prism – one with five sides – is the pentaprism of Figure 4, used extensively in single lens reflex cameras. This prism does not take advantage of the total internal reflection process, as described above. Below the bottom face is a mirror at a 45° angle. Light from the camera's lens is reflected upward and emerges from the left-hand face where the scene being photographed is observed by the eye.

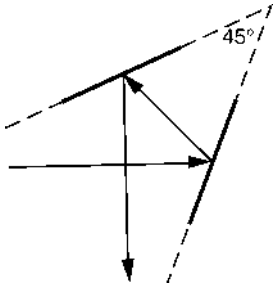

Figure 5 Action of coated faces of a pentaprism.

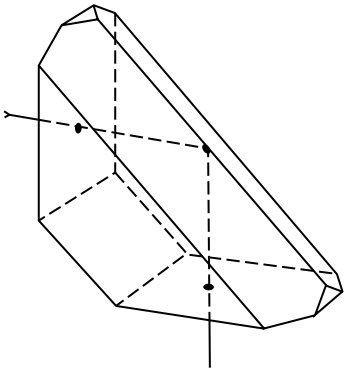

Figure 6 Amici prism.

Since there are two reflections, the reversal caused by a single mirror is eliminated. The camera's mirror is hinged and when the exposure button is pressed, it flies up and the incoming light can reach the film. The two reflecting faces, which lie at an angle of 45° to one another (Figure 5), are coated. To see why this is necessary, note that the two angles in the isosceles triangle formed by the light ray and the face extensions must equal (180° − 45°)/2 or 67.5°. This ray must then lie at an angle of 90° − 67.5° to the normal, or 22.5°, which is much less than the critical angle needed for reflection.

There are at least a dozen other forms of prism which have been designed and used. For example, the Amici prism (Figure 6) is a truncated right-angle prism with a roof section added to the hypotenuse face. It splits the image down the middle and thereby interchanges the right and left portions. These prisms must be very carefully made to avoid producing a double image. One use is in telescopes to correct the reversal of an image.

Double Prism Reflectors

Another extensive application of prisms is their use in pairs as components of telescopes and binoculars. The most widely used is a pair of Porro prisms: two right angle prisms arranged as shown in Figure 7. The double change in direction has two effects; it eliminates image reversal and it shortens the spacing between objective and eyepiece lenses, making the binoculars lighter and more compact. It is possible to calculate how light rays behave in Porro prisms with a ray tracing procedure. This method involves the use of 4×4 matrices and its development is rather complicated. A simple understanding of the way light passes through this optical component can be based on Figure 8, which shows the two prisms separated for clarity. An arbitrary ray (the heavy line) coming from the object will be reflected at a 45° angle at each of the four faces that it strikes. Using two line segments at right angles as an object, the ray from the top of the vertical arrow and the ray from the end of the horizontal arrow will behave as indicated. The image is then identical to the object but rotated by 180°. Since the optics of the binoculars – the combined effect of the objective and the eyepiece – result in an inverted image, the Porro prism corrects this problem.

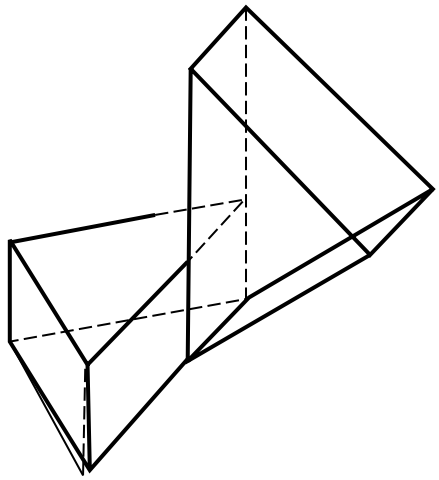

Figure 7 A pair of Porro prisms.

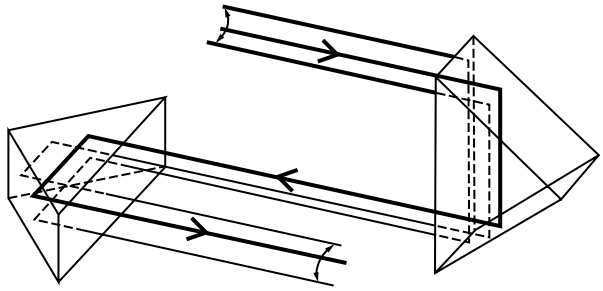

Figure 8 Action of Porro prisms.

Prisms as Instruments

Prisms are the crucial optical element in many laboratory instruments. To show how they can be used to measure the index of refraction of transparent materials – solids or liquids – consider the prism of Figure 9, with apex angle A. Light comes to the prism at an incident angle i, enters the prism at the angle of refraction r, crosses the prism and is refracted a second time when it emerges, the associated angles being designated as r' and i'. When the ray leaves the prism, it is traveling in a direction which represents an angular deviation δ from its original orientation. This deviation is the sum of the individual changes in direction at each face, or

$$\delta = (i - r) + (r' - i') \qquad [3]$$

For the triangle formed by the projected normals, which meet at the angle A:

$$r = r' + A \qquad [4]$$

Hence

$$\delta = i + i' - A \qquad [5]$$

The minimum value of the deviation is an important property of the prism; we find it by differentiation to obtain:

$$\mathrm{d}\delta = \mathrm{d}i + \mathrm{d}i' = 0 \qquad [6]$$

Differentiation of the form of Snell's law valid at each interface, namely:

$$\sin i = n \sin r, \quad \sin i' = n \sin r' \qquad [7]$$

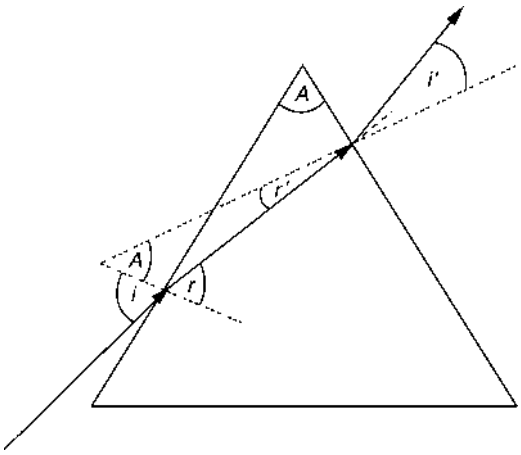

Figure 9 Calculation of prism deviation.

gives

$$\mathrm{d}i = n \cos r \, \mathrm{d}r / \cos i \qquad [8]$$

$$\mathrm{d}i' = n \cos r' \, \mathrm{d}r' / \cos i'$$
$$= n \cos r' \, \mathrm{d}r / \cos i' \qquad [9]$$

where $\mathrm{d}r'$ can be replaced by $\mathrm{d}r$, as obtained from the equation connecting r, r', and A. It now follows that:

$$\mathrm{d}\delta = n\Big[\big\{\cos r / \sqrt{(1 - n^2 \sin^2 r)}\big\} - \big\{\cos(A - r) / \sqrt{(1 - n^2 \sin^2(A - r)}\big\}\Big]\mathrm{d}r = 0 \qquad [10]$$

from which

$$\cos^2 r\big\{1 - n^2 \sin^2(A - r)\big\} = \cos^2(A - r)\{1 - n^2 \sin^2 r\} \qquad [11]$$

or:

$$\cos^2 r = \cos^2(A - r) \qquad [12]$$

and finally:

$$r = A/2, \quad r' = A/2, \quad i' = i \qquad [13]$$

The differentiation thus leads to the conclusion that the ray enters and leaves the prism in a symmetrical manner. The amount of deviation is now:

$$\delta = 2i - A \qquad [14]$$

or:

$$i = (\delta + A)/2 \qquad [15]$$

and:

$$n = \sin i / \sin r = \sin\{(\delta + A)/2\} / \sin(A/2) \qquad [16]$$

Using this expression to calculate n as a function of δ, it is found that δ is a minimum, so that a measurement of the minimum deviation produced by a prism gives the value of its index of refraction in a direct and accurate way. This method can also be used for liquids; a triangular glass cell with walls of reasonable thickness is made very accurately and used to measure δ. The refraction due to the walls does not affect the measurement, since the refraction occurring at the entrance wall is equal but opposite to that at the exit wall.

The description of prism behavior given so far assumes that we are dealing with monochromatic light; that is light of a single wavelength or single

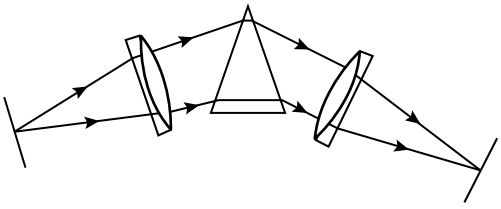

Figure 10 A prism spectrometer.

color in the visible spectrum. However, the index n of a prism varies with wavelength and white light passing through a prism will undergo dispersion – each component of a beam will emerge at a different angle and can be observed. The prisms that are part of expensive glass chandeliers display such an effect. This property of a prism is used in spectrometers – instruments which can measure the wavelength of light. Figure 10 shows a schematic diagram of this kind of instrument of high quality. Note that the light is brought into and taken out of the spectrometer with doublets. As shown in **Geometrical Optics:** Aberrations, doublets can be designed to be achromatic – that is, they focus light throughout the spectrum in approximately the same way. Various kinds of spectrometers have been put into use, with the ability to resolve closely spaced wavelengths being a crucial factor in the design.

See also

Geometrical Optics: Aberrations; Lenses and Mirrors.

Further Reading

Brouwer W (1964) *Matrix Methods in Optical Instrument Design*. New York: WA Benjamin.

Longhurst RS (1967) *Geometrical and Physical Optics*, 2nd edn. New York: John Wiley & Sons.

MIL-HDBK-141 (1962) *Military Standardization Handbook*. Washington, DC: US Government Printing Office.

H

HOLOGRAPHY, APPLICATIONS

Contents

Art Holography
High-Resolution Holographic Imaging and Subsea Holography
Holographic Recording Materials and Their Processing

Art Holography

A Pepper, Shearwater Foundation, Nottingham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Holography captures the phase and amplitude of light reflected from, or transmitted through, an object. In visual terms it captures 'true' three-dimensional space and any objects which are placed within that space. This is very different from the various optical illusions, including perspective drawing and painting, which 'trick' the eye into thinking it is seeing a dimensional image on a flat surface. Not surprisingly, artists have been intrigued by display holography since reports about it became more public in the early 1960s.

Since much of the information about the holographic process was originally published in professional scientific and optics journals, which the artistic community rarely read, or had access to, the transfer of practical information from optical scientist to visual artist was slow. As less technical reports about holography began to appear in publications, such as *Scientific American*, and art journals, such as *Leonardo*, more artists became aware of the possibilities holography offered. It is now over 50 years since its invention by Dennis Gabor, and art holography can be found in public and private collections around the world, ranging from major museums, through academic art galleries to large individual collections.

For the purpose of this article, holography can be defined as a section of a broad optical field, encompassing display holograms produced by artists, creative individuals, and groups. These works are often produced through independent or self-driven research, but can also be the result of a commissioned project. Holographic art works tend not to be produced specifically for commercial gain (although some do sell for large amounts and can be editioned in the same way prints or photographs, to increase their salability or make them more accessible to a wider audience). It is incredibly difficult to assess the number of artists working in this field today, as no specific research has been carried out, but it can be estimated as between 50–100 individuals worldwide, which includes those who have shown work in curated group and solo exhibitions. This relatively small number has made an enormous impact and includes artists who have made holography their main medium of expression, as well as those, more established in other media, who have extended their creative vocabulary by embracing the holographic process. Salvador Dali for example, had holograms made of his sculptural compositions in the early 1970s, and Bruce Nauman used the process to capture a series of self-portraits in the 1960s. Apart from these well known and established contemporary artists, much of the pioneering work has been driven by individuals

who have concentrated on holography as their medium of choice.

Some artists build their own optical studio in which to experiment in a hands-on environment, some collaborate with scientific and research establishments, where the work, facilities and expertise is shared, and others design and supervise the production of their work which is produced for them by a commercial company. All of them produce a unique contribution to the visual arts.

Early Exploration

Holography offers a unique method of recording and reproducing three dimensions on a flat surface without the need for special glasses (for stereoscopic viewing), lenses (such as lenticular viewing) or linear perspective which, since the Renaissance, has been our main method of tricking the eye into believing that an image recedes into the distance and displays three dimensions.

Many of the very early works produced by artists concentrated on the obvious still life nature of the medium and explored the unusual phenomenon of objects 'frozen' in holographic space. Almost all of these early works were recorded using continuous wave helium neon lasers, either in scientific/research facilities, which artists had arranged to use, or in collaboration with established scientists and optical imaging researchers.

As early as 1969, British artist Margaret Benyon was experimenting with this new medium. Her laser transmission hologram 'Picasso', is a recording of images taken from Pablo Picasso's painting 'Demoiselles D'Avignon'. Here Margaret Benyon took a copy of the painting, cut out the characters, placed them in a three-dimensional spatial arrangement, with real objects, and then made a hologram of this 'scene'. Picasso attempted to show, through many of his paintings, multiple views of the same subject on a two-dimensional surface, one of the main thrusts of Cubism in the late twentieth century. Benyon achieved this in 'real space' by recording a hologram of the flat reproductions. By doing so, she demonstrated that holography automatically achieved what Picasso, and many other key contemporary visual artists, had been trying to do for years.

Holographic Black Holes

It has been known, from the very early days of holographic research that, when using a continuous wave laser to record a hologram, if any object in the scene moves more than half the wavelength of light, no interference pattern will be produced and therefore no recording of the object will take place. Unlike photography, where if an object moves during the exposure, it will become blurred, in holography it will simply not be recorded. Clearly, if no recording can be made, optical researchers working with holography, while noting the phenomenon, would not necessarily feel the need to make holograms which would fail. Several artists, on the other hand, not confined by research deadlines, have acknowledged this strange phenomenon and exploited this limitation of the holographic process. One of the very early examples is 'Hot Air' (Figure 1), a laser transmission hologram by Margaret Benyon and now in the Australian National Gallery Collection. Produced in 1970, it shows a three-dimensional composition of a jug, cup, and the artist's hand. The jug and cup were filled with hot water and during the recording of the holographic plate, with a continuous wave laser, the heat from the hot water disturbed the air above these objects. This tiny amount of movement resulted in the area above the jug and cup not being recorded. Similarly, the artist's hand moved slightly during the exposure and therefore did not record. Strangely, when viewing the resulting laser transmission hologram, the hand is visible. What is displayed is a holographic hand-shaped black hole, the three-dimensional recording of the space once occupied by the hand; the dimensional recording of the absence of an object.

Figure 1 'Hot Air', Margaret Benyon 1970. Laser transmission hologram. Copyright © 1970 M. Benyon. Reproduced with permission.

> Artists in display holography place continual demands on the technology. Their natural tendency is to break rules. I believe that most successful holographic artists belong to this 'suck it and see' category of experimentalists. On hearing about the 'rules' of holography, they break them, just to see what happens. (Margaret Benyon, PhD Thesis, 1994).

Benyon started by breaking the rules and has continued to develop from these early pioneering laser transmission works, which helped establish a visual 'vocabulary' including many important social and political issues.

Other artists have exploited the 'limitations' of the holographic process to produce visually stunning works which question how we look at the three-dimensional world and question what is 'real'. Rick Silberman (USA) developed a shadowgram technique where the hologram is made of a laser illuminated ground glass screen. Objects are placed in front of this luminous surface and their three-dimensional silhouette is recorded. The viewer of the hologram is looking at the space where the object used to be. A pivotal work in this field is 'The Meeting' 1979 (Figure 2), a white light reflection hologram which had been recorded in such a way as to cause the three-dimensional shadow of a wine glass to protrude out of the holographic plate and be viewable several centimeters in front of it, between the observer and the holographic emulsion. The actual wine glass used in the recording has been broken and what remains (the base, stem, and small part of the bowl of the glass) is placed on a platform attached to the holographic plate. When viewing the work head on, the missing parts of the broken glass are recreated, in three dimensions, by the holographic shadow. When viewed obliquely, from the side for example, only the broken (real), glass is visible. Not only does this piece have an element of performance in the recording, breaking, mounting, and reconstruction of the wine glass, it also explores questions of visual perception and concepts of reality.

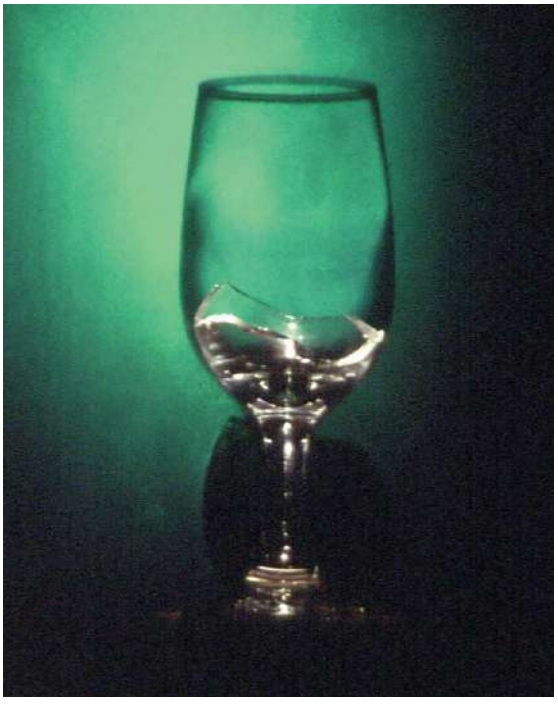

Figure 2 'The Meeting', Rick Silberman 1979. White light reflection hologram (shadowgram). Copyright © 1980 ATP. Reproduced with permission.

The recording of the shadow of an object, or objects, has now become commonplace in creative holography, with many artists employing techniques to mix different objects together, change their color and rearrange them in space. One advantage of this technique is that dependence on absolute stability of the objects used is reduced. The objects are 'encouraged' to move during the recording process and so create 'negative' images. Steve Weinstock (USA), Marie-Andrée Cossette (Canada), Michael Wenyon and Susan Gamble (USA/UK) have all contributed to this area in some of their early works.

Public Display

As the technology and optical processes of holography have developed, so greater flexibility, and public accessibility, have emerged. The single most important aspect has been the ability to display holograms in white light.

Emmett Leith and Uris Upatnieks invented laser transmission holography in 1964. They demonstrated the practicalities of Gabor's original invention, by utilizing laser light, to produce three-dimensional images. To display these recorded images, laser light needed to be shone through the holographic plate, at the original angle used by the recording reference beam (off-axis). The results were breathtaking, with full visual parallax in all directions. Although it was the laser light which made the recording and reconstruction of these unusual images possible, the laser was also a limiting factor. Each time a hologram was displayed, a laser was required. This relegated the technique to displays in scientific conferences, laboratories, or other locations with optical equipment. What was needed was a technique allowing holograms to be displayed without the need for this expensive and dangerous laser light. White light holography was the answer.

Although a laser is still required to record the original hologram, by using a copying technique, a

restricted aperture or one step process, the resulting holograms can be reconstructed using a small source of white light such as a spotlight or even direct sunlight. As expensive lasers were no longer needed to display holograms, large public exhibitions could be mounted. Gallery and museum spaces could include holography in their exhibitions and individuals could buy and display holography in domestic locations. This technique also benefited the commercial display market, where holograms could now be used for advertising and promotional events.

Stephen Benton (USA) (1942–2003), inventor of white light transmission (rainbow) holography in 1968, made a major contribution to the use of the medium by artists. His technique not only allowed holograms to be viewed in white light, but they were extremely bright and offered the possibility of creating multicolor images. His 'Crystal Beginnings' 1977, (Figure 3), shows just how much space can be created with very simple points of light. Benton even taught master classes for artists so that he could explain how his technique functioned and explain the mathematics involved. He also worked with several artists to help them realize their individual projects.

Harriet Casdin-Silver (USA) was one of the first pioneering artists to work with Benton and produced several key works now seen as integral to the art of holography. Their early collaboration 'Cobweb Space', 1972, sculpted pure light and was one of the very first art holograms to be displayed using Benton's rainbow hologram technique. A later work, 'Equivocal Forks', 1977, exists in different formats; as a laser transmission hologram showing the pseudoscopic display of domestic folks which protrude, inside out, from the holographic plate, and as a number of white light transmission holograms, of the same image, which were shown in an outdoor installation mounted by The Center for Advanced Visual Studies, part of the Massachusetts Institute of Technology. This demonstrated to a wide, nonscientific, audience, that holograms could be accessible, easily illuminated and weatherproof. Casdin-Silver has continued to produce art holography and has, most recently, worked with large-scale portraiture.

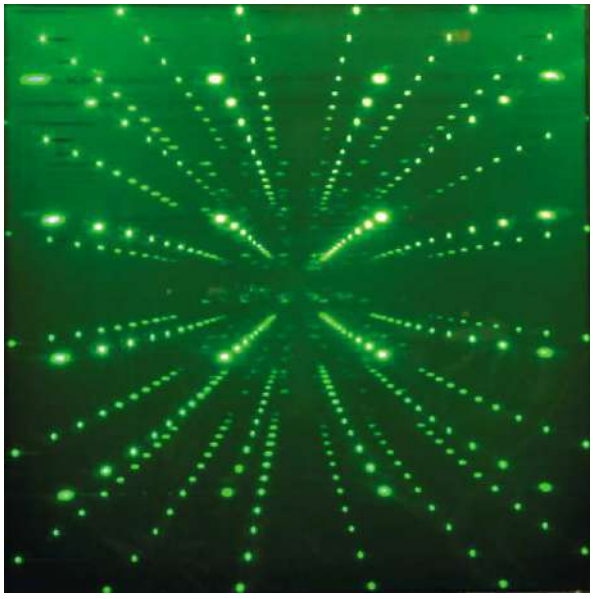

Figure 3 'Crystal Beginnings' Stephen Benton 1977, White light transmission (rainbow) hologram. Jonathan Ross Collection, London, UK. One of the best known works by the inventor of the rainbow hologram. Copyright © 2004 J. Ross. Reproduced with permission.

Although Benton's process increased the accessibility of holography, an earlier technique, invented by Yuri Denisyuk (Russia) in 1962, allowed holograms to be recorded in a single step and then displayed using white light. The political situation between the USSR and the rest of the world during the early 1960s meant that information and results from much of this early work was not openly accessible, and communication between Denisyuk and other key researchers in holography was limited. However, as soon as details of the Denisyuk technique became more widely available in the 1970s, artists began to embrace it as a practical alternative.

The work of Nicholas Phillips (UK), whose research improved chemical processing of display holograms during the late 1970s, helped artists and the rest of the holographic field to achieve much better results when making reflection holograms, a technique favored in Europe. He was also heavily involved with the rock band, 'The Who', and their Holoco company which mounted two very successful exhibitions of display holography at the Royal Academy of Art, London, at this time. Artists who might not have come across the holographic phenomenon became aware of its potential due to its presence in one of the UK's more traditional art venues. Other large exhibitions in the late 1970s and early 1980s in USA, Sweden, and Japan also highlighted the immense possibilities holography offered to a nonscientific community of gallery visitors.

As interest in all aspects of holography increased, dedicated spaces opened to sell, display, and archive the output from artists, scientists, and commercial producers. One of the most influential of these was the Museum of Holography, located in the SoHo art district of Lower Manhattan, New York (Figure 4). From 1976 through to 1992, it regularly mounted group and solo exhibitions by artists, giving many of

Figure 4 Dr Dennis Gabor, inventor of holography with, Posy Jackson, Director of the Museum of Holography, during his visit to the Museum in March, 1977. Copyright © 1977 Paul D. Barefoot. Reproduced with permission.

them their first major exposure to a museum-going public. There were displays relating to the scientific, commercial, and industrial usage of the medium, as well as exceptionally popular traveling exhibitions. 'Through the Looking Glass', which traveled worldwide, allowed early display holography to be experienced by thousands of people who would not normally have come into contact with the medium. In the Summer of 1981, over 300,000 visitors saw 'Through the Looking Glass' during the 11 weeks it was on view at the Israel Museum, Jerusalem.

After the Museum of Holography closed in 1992, its entire collection was purchased by the MIT Museum, Boston, USA, and now forms part of the largest museum collection of creative and scientific holography in the world. The New York museum was not an isolated venue. During the late 1970s and early 1980s, museums of holography opened in Chicago, Paris, and Pulheim (near Cologne). Galleries were also founded in many major cities to sell art and commercial holograms.

Stability Solutions

Although being able to display holograms using inexpensive sources of white light allowed the medium to become much more accepted as a display process, there has always been the cost and complexity of the optical laboratories needed to record the holograms in the first place. Few artists were able to equip their studios with 'off the shelf' optical components and precision-built isolation tables most often found in scientific and research labs. The solution was to replace the commercially built isolation table with much less expensive sand tables.

Sand box holography revolutionized the field for independent artists. By building a wood or brick box filled with sand, resting on car inner tubes, artists were able to hand-make their own isolation tables. Inexpensive optics, mounted onto plastic pipes were simply thrust into the sand, allowing complex optical recording setups to be constructed. Surprisingly, many key artworks have been produced using this technique, which opened the field to anyone with basic construction skills and a laser.

The laser has remained an expensive element in the holography studio but, most recently, the prospect of replacing this with laser diodes has offered a much more economic alternative. Not only can light from these diodes provide more compact optical setups to record holograms, they can also be used to display them. Significant progress has been made in both diode recording and display, which is opening up new opportunities for the visual artist. Almost all of the early practitioners in the field were attracted to holography because of the laser transmission holograms they saw. They have a depth and dimensional quality which is extremely different from the white light viewable holograms we tend to see today. Now, with the possibility of displaying holograms illuminated by inexpensive laser diodes, the laser transmission hologram is set, once again, to be a viable medium for artists.

Object Holography

Many artists were attracted to holography because of its unusual three-dimensional properties. However, they did not have the resources to make holograms of large objects or scenes, so scaled these down and constructed small sculptural settings for recording. One of the pioneering artists in this area was Dan Schweitzer (1946–2001). Originally an actor and familiar with theatrical staging, many of his holograms are populated by tiny figures placed in specially built 'theatrical' settings. Holography not only recorded these specially constructed miniature 'rooms', but the resulting images were realistic enough to suggest to the viewer that they were looking at a scaled-down life-size event. We do this regularly when looking at photographs and accept that they are a small-scale version of a larger scene. Traditional holography does not offer this 'scaling down' in the same way as photography. One celebrated example is Schweitzer's 'Thendara', 1978, where a sculpted, miniature figure sits and looks towards a window frame. He also produced a companion piece in which the physical model, used to make the earlier hologram, is on display as a piece of sculpture. Here the miniature window is the frame for a hologram displaying a previously recorded model of a landscape. As Schweitzer's technical skills

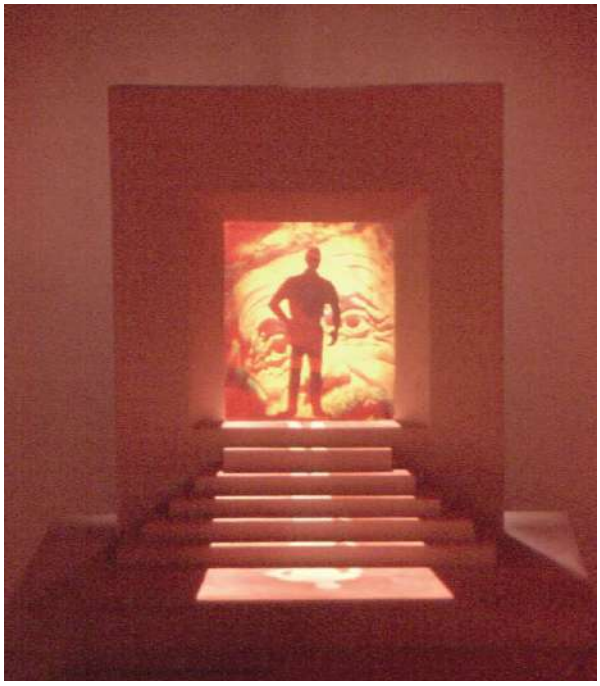

Figure 5 'Od Vent', Dan Schweitzer 1980, White light transmission hologram in sculptural mount. Copyright © 1980 ATP. Reproduced with permission.

developed, he began to produce complex multi-layered works incorporating still-life models, hologram fragments, and optical animation to construct visual spaces which could not exist in the real world.

In his work 'Od Vent' (1980) (Figure 5), Schweitzer framed an exceptionally complex white light transmission hologram into a theatrical model of a doorway. The miniature stairs lead up to an opening, a portal into another world, in which the hologram is mounted. Illuminated from behind, the hologram displays multiple exposures, optical animation, and several superimpositions of objects as an observer moves from left to right in front of the piece. In this complex holographic world, the silhouette of the figure, standing in the door, casts a bright (negative) shadow down the miniature stairs. Not only are we able to look into an illusionary world through the holographic doorway, where at a certain point Einstein looks back at us, but the peculiarities of this holographic world extend into our own world as seen by the 'real' (negative) shadow cast down onto the stairs.

Another artist, who combined the use of multiple exposed objects displayed in a single space, is John Kaufman (USA). He established a reputation by recording objects (often rocks) collected from around his Californian holography studio. These found objects were captured in three dimensions and were then, through a multiple exposure and emulsion swelling technique, combined into a single still-life composition. He became expert at masking out areas of his objects so that they could appear in different, intense colors, and in many cases used the holographic process and multiple exposures to mix colors directly in the holographic space he was working with. In 'Stone Head' (1982) (Figure 6), for example, this relatively 'normal' stone is transformed by the use of intense color and the three-dimensional volume it occupies.

When considering how the hologram should be framed and displayed, several artists have utilized the mount in which the hologram sits as an integral part of a much more complex work. Frithioff Johansen (Denmark) recorded a hologram of smoke, freezing it in time and space using a pulsed laser (see below). This was then converted into a white light transmission hologram and mounted in a frame of granite. 'Hardware-Software 1' 1987–1988, allows the gallery visitor to not only view the massiveness of the granite structure, but also the delicate sculptural solidity of smoke 'frozen' in space.

Rebecca Deem (USA) recorded what appears to be an extremely simple hologram of a luminous tube, mounted onto a swivel at the end of an elegant metal support. In 'W/Hole Message' viewers can look into the small holographic plate and see the image of the tube extending away from them into the volume behind the plate. By flipping the hologram, on its swivel, to turn it over, the image in the hologram is also flipped and the tube appears in the viewer's space, between them and the plate. This three-dimensional pseudoscopic image, one of the attractive and unusual aspects of holographic display, is inside out, back to front, and upside down. The gallery visitor really does have a chance to view the whole message.

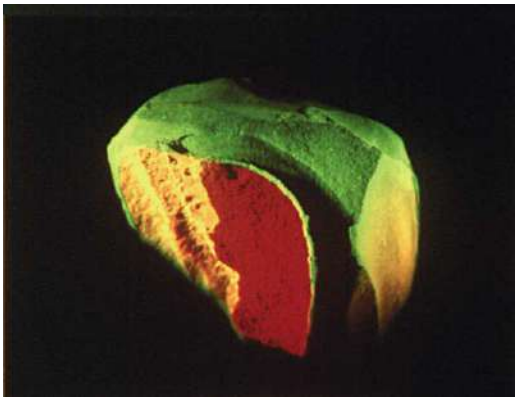

Figure 6 'Stone Head', John Kaufman 1982, White light reflection hologram. Copyright © 1982 John Kaufman. Reproduced with permission.

Figure 7 'Breakfast', Pearl John 2002. Laser etched toast displayed in a white light reflection hologram. Copyright © 2002 Pearl John. Reproduced with permission.

Text has become an important element in holography, as it is often incorporated in, onto, or around holographic displays. Eduardo Kac (Brazil) has produced several holograms, which display poems, words, and letters floating within the holographic space he has created.

Pearl John (UK) most recently produced a series of still-life holograms using a pulsed laser but with the added subtlety of writing onto the surface of her objects. In 'Breakfast' 2002 (Figure 7), for example, the slices of toast, making up the composition, have been laser etched with text which appears 'burnt' into the surface of the bread.

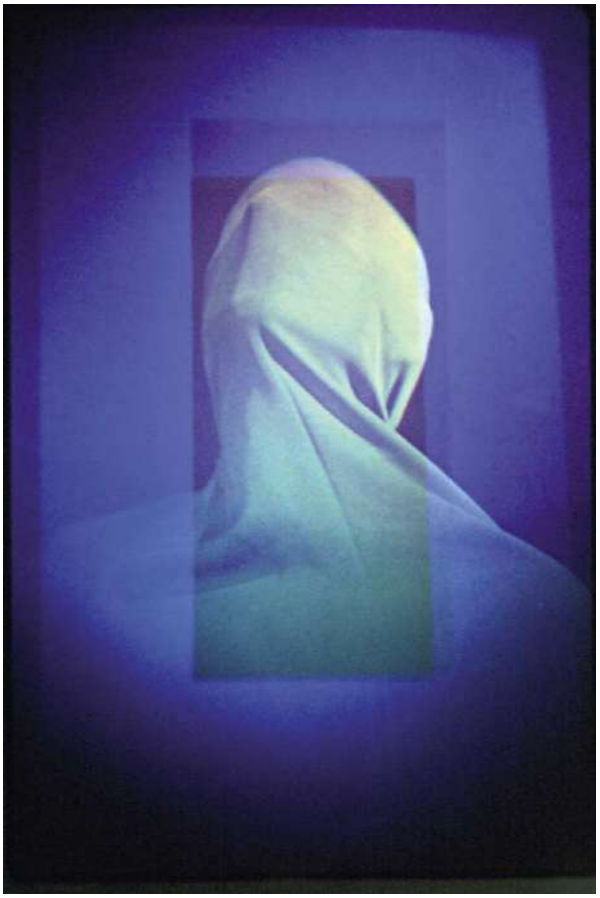

Figure 8 'Awakening', Ana Maria Nicholson, 1993. White light (pulsed) reflection hologram. Copyright © 1993 Ana Maria Nicholson. Reproduced with permission.

Pulsed Lasers

Continuous wave lasers placed limits on the holographic recording process and dictated that no movement in the recorded object could take place. We have seen that this 'problem' was exploited by artists at an early stage, but there was also a desire to capture living objects, in the form of the human figure. Pulsed laser holography was the solution. Because this type of laser generates a bright flash of coherent laser light, the intensity and duration of the pulse is enough to 'freeze' any physical movement during the recording process, allowing the recording of people, animals, or anything which moves and could be brought into an holography studio. Again, the cost and accessibility of this specialist equipment was prohibitive to artists, but they found ways of persuading scientists with pulsed lasers to help them. Early adopters of this technique were artists such as Ana Maria Nicholson (USA), Melissa Crenshaw and Sydney Dinsmore (Canada), Margaret Benyon (UK), Anait (USA), and Alexander (USA/UK). Recordings could now be made of live figurative compositions and individual portraits. Many of the artists, not content with producing the holographic version of a traditional photographic portrait, began to manipulate these recordings, distorting, multiple exposing and inverting these unusual three-dimensional images of people.

Holographic portraits do have an unusual quality, which painters or photographers have found difficult to deal with but art holographers have embraced. Some, like Ana Maria Nicholson (USA), have combined subtle processing techniques, to introduce color 'layers' into the portraits, combined with nonstandard views or compositions (see 'Awakening' 1993 (Figure 8)). Other artists, who emerged in the 1980s, produced complex, highly staged compositions, such as those from Patrick Boyd (UK), or they offered us a different view of the 'self portrait' like those by Martin Richardson (UK). Others, such as Shu-Min Lin (Taiwan), used the combination of multiple views of human figures recorded onto the same holographic plate. By walking past these displays, an animated effect could be experienced.

So much interest was generated in pulsed holography that several studios emerged for use by artists, including independent facilities and those within educational establishments. No longer was this the domain of the well funded scientific research facility. Several of these 'studios' still exist and are offering artists access to equipment through specialized artist-in-residence programs.

Abstraction

Although holography is an exceptional process for recording every detail of an object and the volume it occupies, many artists have wanted to work with the abstract and kinetic possibilities the process offers. Light can be recorded and displayed in three dimensions, it can be animated, inverted, compressed, sculpted, and projected. Douglas Tyler (USA), in his In the 'Dream Point Series' 1983 (Figure 9), recorded small laser transmission holograms but displayed them using normal white light, producing a smearing effect of the three-dimensional object. He then mounted these onto transparent surfaces (Plexiglas) and drew on the surface of the plastic to incorporate etched lines and shapes. Rudie Berkhout (Netherlands/USA) began his holographic exploration with stunning abstract landscapes made of luminous, multicolor, animated objects which took advantage of his mastery of the white light transmission technique. Most recently he has been recording reflection holograms which display undulating colored 'waves' of light as an observer moves past the hologram. These have been included in specially produced paintings for large-scale architectural commissions.

Sam Moree (USA) combines multiple images, exposures, diffraction gratings, and smaller holograms into a single work. The finished holograms display a complex series of image fragments and colors within the depth of the holographic space he has constructed. Ruben Nuñez (Venezuela), originally approached holography from a background in glass production and used the ability of the holographic process to 'capture' light reflecting and passing through his glass objects. The resulting white light transmission holograms display an abstract 'photonic' display of form and color. Fred Unterseher (USA), uses holography to record optical distortions and diffraction gratings which would display abstract patterns of light. Many of these early works were mandala images which displayed an intense kinetic effect as an observer walked past the surface of the holographic plate.

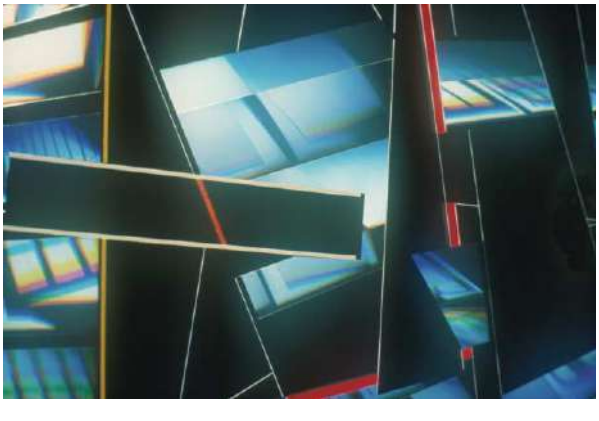

Figure 9 'Dream Passages': 32″ × 138″ (3 panels 32″ × 46″). Douglas E. Tyler 1983, Laser transmission holograms, tape and Plexiglas, illuminated with white light. Detail of center panel. Copyright © 1983 D.E. Tyler. Reproduced with permission.

Pseudo Holograms

In 1973, Lloyd Cross invented the multiplex (integral) hologram technique which used sequential two-dimensional images recorded on movie film combined with white light (rainbow) hologram techniques. The result was a holographic stereogram, which gave viewers the illusion of seeing animated images in three dimensions. The final holographic sheet was displayed either curved or as a complete cylinder, which was often motorized so that viewers could stand in front of the unit and watch the animation take place as the hologram rotated. Such a display does give the impression that an observer can walk all the way round the image, enhancing its perceived three-dimensional effect.

As the original images are recorded photographically, there are several advantages – anything which can be recorded on inexpensive black and white movie film can be converted into a holographic display. Short animated sequences can be presented while accepted film techniques such as zoom, pan, tilt, and transitions can be incorporated into the finished sequence of images. Although very popular for advertizing and promotional campaigns during the 1980s, particularly in the USA, several artists used this technique. They could make the movie film sequences themselves and then send this film to a commercial producer for a hologram to be made for them. This offered an interesting balance in the creative process. The artist could concentrate on the filmic image, using facilities and photographic techniques they were familiar with and not have to learn the holographic process or build a holographic laboratory themselves. Artists such as Anait (USA) produced several experiments in the area, one of which showed her inside the holographic display,

Figure 10 'Woman' Harriet Casdin-Silver 1979. Multiplex hologram (white light transmission). Copyright © 1980 ATP. Reproduced with permission.

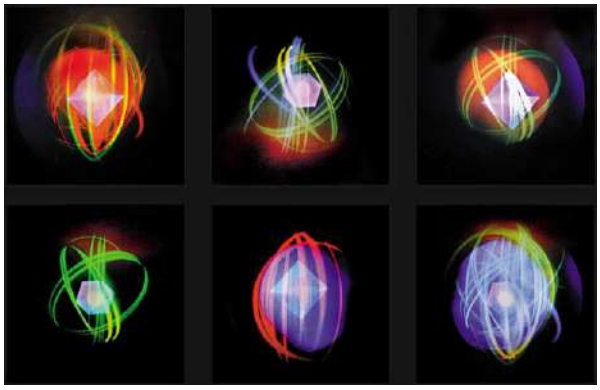

Figure 11 'Neuro Hologram', Ikuo Nakamura, 1993. White light transmission hologram with multiple, computer controlled, lighting. Copyright © 1993 Ikuo Nakamura. Reproduced with permission.

writing, with spray cream, seemingly onto the holographic surface. Dieter Jung (Germany), used this technique to display a specially written poem which appeared to spiral around inside of the display.

Harriet Casdin-Silver used the technique to produce an unusual self-portrait. 'Woman', 1979 (Figure 10), is a multiplex hologram of the artist's head. Black and white film images were recorded of Casdin-Silver's head, through 360 degrees, and then each frame of film was converted into a thin white light transmission hologram. The artist chose not to display the resulting image in a normal cylindrical display, but to 'flatten' it out so that the front, sides, and back of her head could be seen, distorted, in a single head-on view.

As video and digital imaging processes have developed over the years they too have been combined with this pseudo holographic technique, opening up a wide avenue of possibilities. Video footage, computer generated modeling, or any process which can produce sequential 2-D images, can now be converted into an animated 3-D hologram. As digital video cameras and computer 3-D modeling software drop in price, and increase in sophistication, a whole new area of display has become available to the digital and video artist.

Multi Media

As in other areas of the visual arts, more than one medium or process are often combined. Doris Vila (USA) has produced complex environments combining large-scale holographic images which, rather than being illuminated with a single source of light, are reconstructed with multiple spotlights or via video projectors. The lighting is activated by sensors in the exhibition space, which are computer controlled and made to trigger certain sequences dependent on the input they receive from the movement of visitors.

Traditional illumination of a hologram involves a single point source of light shining onto, or through, the holographic plate. This will reconstruct a single image. If multiple lights are shone onto the hologram, multiple reconstructions of the single image will result. By using many independently controlled lights on one hologram, Vila has been able to create animation and kinetic effects as the resulting multiple images overlap. Using light from a video projector also allows the artist to combine a reconstructed holographic image with two-dimensional animated footage from the video projector.

Other artists to combine holograms with structures, interactive lighting and environments, sound, and installational displays include: Georges Dyens (Canada), Marie-Christiane Mathieu (Canada), Matthew Schreiber (USA), Philippe Boissonnet (Canada) Ana MacArthur (USA) and Shunsuke Mitamura (Japan).

Ikuo Nakamura (Japan/USA) has used multiple lighting of a single computer generated hologram to produce animated effects controlled by the viewer. In his 'Neuro Hologram' 1993 (Figure 11), brain waves from the viewer, wearing a special set of sensors on their head, are measured and converted into control signals for a number of lights behind the hologram. By relaxing, or becoming agitated, the viewer can change their brain waves and actively control the appearance of the hologram, causing multiple views of the holographic image to fade in and out, giving an impression of abstract animation.

Architectural Scale

One of the great limitations of holography has been the size of the holographic plates on which the light

sensitive emulsion is coated. These tend to be in the region of 20 × 23 cm, 30 × 40 cm, or meter-square formats. Generally, the larger the finished hologram, the larger the laboratory needed to produce it. This impacts a great deal on artists who do not have the facilities, or finances, to produce 'large' holograms. A solution has been to combine smaller holographic plates together, as a mosaic, to give the size and scale needed. In recent years, several artists have completed public commissions for large-scale interior and exterior holographic installations.

Setsuko Ishii, (Japan) has produced several installations in public spaces where individual holographic plates are mounted together to form a larger composite structure. Many of these pieces use holograms recorded on dichromate gelatin plates which produce an exceptionally bright image, ideal for public spaces, where visibility, from multiple angles, is important.

Michael Bleyenberg (Germany) produced an entire exterior holographic wall (13 × 5 meters) during 2002 (Figure 12), for the extension of a research building in Bonn. The installed graphic wall, 'EyeFire' (AugenFeuer), is made up of 26 smaller panels, each constructed from tiny holographic 'pixels' (holographic optical elements, HOE) to give an intense color effect. Working in collaboration with the Institute for Light and Building Technology, Cologne, Bleyenberg was able to design his graphic panels on a computer screen and have this information translated into the thousands of tiny holographic pixels which, when seen together, make up the complex and high intensity colour design.

Figure 12 'EyeFire/AugenFeuer', Michael Bleyenberg 2000. White light transmission holograms mounted onto building exterior. Copyright © 2000 M. Bleyenberg. Reproduced with permission.

Dieter Jung (Germany) has produced commissioned art works for large interior architectural spaces. In 'Perpetuum Mobile' 2001, located at the European Patent Office, The Hague, holograms are placed on the floor, which can be walked over, and an extensive holographic mobile structure hangs overhead. The light, and color, from the abstract holograms in the mobile sections, can be reflected in the abstract holographic floor and combine to produce an interior lightscape impossible to produce with more traditional media.

Holograms can be mass produced using a mechanical embossing process which transfers the optical contents of the original master hologram onto plastic sheeting for duplication in large runs. One novel use of this material has been the interior of the Canadian pavilion at Expo '92 (Figure 13). Melissa Crenshaw (Canada/USA) used huge quantities of embossed holographic material to cover the 3000 square foot interior walls of the specially built pavilion. Lighting was installed to illuminate this optical diffracting wall and the entire structure was covered in glass, over which water flowed. The resulting interior appeared to be a wall of liquid light which changed color and hue depending on where it was viewed from.

Sally Weber (USA) used holographic embossed material for several outdoor commissions. Structures were clad in embossed holographic sheets and made water resistant. Sunlight illuminated the installations to produce surfaces of constantly changing color. In direct (bright) sunlight, intense rainbow colors would be diffracted by the holographic surfaces. On an overcast day, pastel, more muted,

Figure 13 Interior of the Canadian Pavilion Expo '92, Spain. Holographic wall water feature by Melissa Crenshaw. White light transmission (embossed) hologram foil mounted into pavilion walls. Copyright © 1992 M. Crenshaw. Reproduced with permission.

Figure 14 Signature of Source, Sally Weber 1997. Site specific window installation at the Karl Ernst Osthaus-Museum, Hagen, Germany. Copyright © 1997 Sally Weber. Reproduced with permission.

colors would appear. Although she is often known for these large-scale outdoor embossed works, one of her permanent commissions is housed in the Karl Ernst Osthaus-Museum, Hagen, Germany, where specially made holographic diffraction gratings are incorporated into an existing building. 'Signature of Source' 1997, (Figure 14), is a permanent, site-specific installation producing a line of light which enters the museum space from the skylight it is built into. This abstract line of pure color becomes a sculptural element within the space of the museum and is generated by the action of light from outside the building being manipulated and 'formed' as it passes through the holographic window.

One of the most celebrated and very early attempts to incorporate large-scale holograms into art works

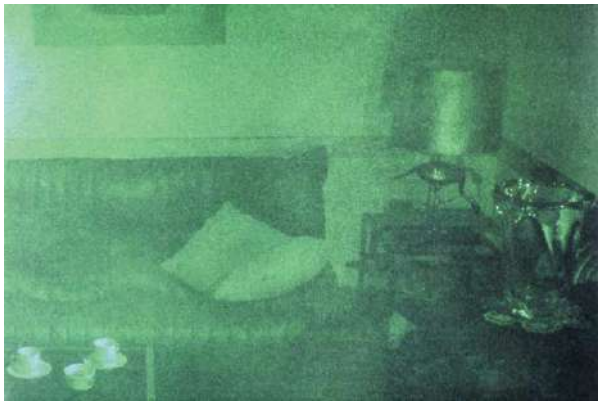

Figure 15 'There's no Place Like Home', Paula Dawson 1979–1980. Laser transmission hologram of an entire room. Copyright © 1980 Paula Dawson. Reproduced with permission.

was the installation 'There's No Place Like Home' by Paula Dawson (Australia) in 1979–1980 (Figure 15). A laser transmission hologram of an entire room was recorded. The room was specially constructed, life size, and recorded onto a single holographic plate. The resulting hologram was then built into the window of a small 'house' constructed in the gallery. Visitors were able to approach the house, look into the window and 'see' the interior with its furniture, fixtures, and fittings. They would then enter the house and walk past the space where the room appeared to exist. When viewed from this vantage point there was nothing there except an empty space. Artists have often questioned and researched the concepts of 'inside and outside', voyeurism, reality, normality, and 'existence', but few have had the tenacity to produce an installation with such visual and conceptual impact.

Twenty-First Century Art

Art holography has developed and matured over the past 50 years to become a viable and significant element in the contemporary visual arts. Critics are still asking 'is it art?' in the same way they used to worry about whether photography was art. While they consider this problem, practitioners in the medium are continuing to find ways to push past its technical and visual boundaries (Figure 16). They continue to collaborate with the scientific and research world and use holography as a medium to comment on life and society in the early twenty-first century.

> …I hope, that one day I will be able to give you, artists, something really powerful; panoramic holograms,

MANIFESTO

.A code.
.Spatial.
.Didactic.
.Photonic.
.Futuristic.
.A discipline.
.Illumination.
.A photograph.
.Science↔Art.
.A precision tool.
.Illusion concrete.
.Pure chiaroscuro.
.An automatic Eden.
.An indivisible whole.
.Painting↔Sculpture.
.3D on 2D intimating nD.
.Material and immaterial.
.Writing the whole picture.
.The presence of absence.
.Making the invisible visible.
.An artificial 3D imaging system.
.Re-evaluating our visual traditions.
.Part of an innate human grammar.
.A bridge into the future with a past.
.The smile without the Cheshire Cat.
.Neither geometric nor organic in form.
.The reconstruction of light wave-fronts.
.Shattering the illusionist paradigm in art.
.A perceptual screen from the 'real world'.
.A potential mass-communications medium.
.A visible medium and a subvisible principle.
.A model for the complexities of brain function.
.Not art, but a condition in which art can happen.
.Mirrored in the life process, our own physiology.
.A mirror to our evolutionary progress from objects to systems.
.Realizing space time/time reversed/back as front/inside as out.
.A model for non-fragmentary thinking, language and behaviour.
.A diminutive shadow-theatre in which metaphors are strongly lit.
.A possible visual model for quantum theory and relativity theory.
.The most complete form of reproductive imaging that exists to date.
.A means of making statements about the world, or presenting pieces of it.
.The fulfillment of a longstanding cultural vision of what imaging ought to be like.
.Joined with high technology to raise the ceiling on humanity, or record Doomsday.
.The automatic execution of the ultimate cubist notion of a simultaneous view at every given point.
.An expression for mystical aphorisms such as 'the universe in a grain of sand' and 'we are all one'.

HOLOGRAPHY AS ART

Figure 16 'Holography as Art', Margaret Benyon 1980. Manifesto. Copyright © 1980 M. Benyon. Reproduced with permission.

extending to infinity, in natural colors. Dennis Gabor, Inventor of holography in a letter to artist Margaret Benyon 28th February, 1973.

See also

Coherence: Coherence and Imaging. **Diffraction:** Diffraction Gratings. **Holography, Techniques:** Overview. **Imaging:** Multiplex Imaging. **Interferometry:** White Light Interferometry. **Introductory Article:** Early Development of the He–Ne Laser (Title TBC).

Further Reading

Benyon M (1994) *How is Holography Art?* PhD thesis, Royal College of Art, London.

Brill LM (ed.) (1989) Holography as an art medium, Leonardo. *Journal of the International Society of Arts, Sciences and Technology*, 22 (3 and 4).

Coyle R and Hayward P (1995) *Apparition, Holographic Art in Australia*. University of Sydney, NSW, Australia: Power Publications.

Casdin-Silver H (1998) *The Art of Holography Exhibition Catalogue*. Lincoln, MA, USA: DeCordova Museum and Sculpture Park.

Holographische Visionen, Bilder durch Licht zum Leben erweckt (1991) Exhibition Catalogue, Museum für Holographie und neue Visuelle Medien, Pulheim, Germany: Edition Braus.

Holosphere, Formerly published by the Museum of Holography. New York: Boston, USA: Archieved at the MIT Museum.

Jackson R and Burns J (1976) *Holografi: Det 3-Dimensionella Mediet, Exhibition Catalogue*. Museum of Holography, New York: Stockholm Cultural Center.

Jung D (ed.) (2003) *Holographic Network*. Germany: Rasch Verlag, Bramsach.

Orazem V and Schepers J (1992) *Olografia, Holografie, Holography, Exhibition Catalogue*. Rocca, Paolina, Italy: Centro Espositivo.

Lipp A and Zec P (1985) *Mehr Licht/More Light, Exhibition Catalogue*. Kabel-Verlag: Hamburger Kunsthalle.

Pepper AT (ed.) (1992) *The Creative Holography Index: The International Catalogue for Holography*. Germany: Monand Press.

Pepper AT (ed.) (1996) *Art in Holography 2*. International Symposium on Art in Holography. Nottingham, UK: University of Nottingham.

Margaret Benyon M (1989) *Phases: Exhibition Catalogue*. New York, USA: Museum of Holography.

Tyler DE (ed.) (1990) *Report, International Congress on Art in Holography*. Notre Dame, Indiana, USA: Saint Mary's College.

Tyler DE (ed.) (2003) *Leading Lights, Women in Holography, Exhibition Catalogue*. Notre Dame, Indiana, USA: Saint Mary's College.

High-Resolution Holographic Imaging and Subsea Holography

J Watson, University of Aberdeen, Aberdeen, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Holography is well-known to most of us: there can be few who are not familiar with the white-light display hologram seen in exhibitions and shops and on credit cards. The impact of holography, though, spreads far and wide into areas as diverse as vibration analysis, image processing, holographic-optical elements, and optical computers. One aspect often overlooked is its suitability for high-precision imaging and accurate dimensional measurement. It is the unique ability of holography to recreate, from a flat photographic film, a three-dimensional image that is optically indistinguishable from the original which sets it apart from other imaging techniques. In particular, holography comes into its own when recording objects in a hazardous environment or where access is difficult, such as underwater or in nuclear fuel element inspection.

We describe here the use of holography for imaging, identification, and measurement of aquatic organisms. A detailed knowledge of the distribution and dynamics of marine and freshwater organisms and particles, such as plankton, is crucial to our understanding of such systems and how they affect our environment. Traditional methods of gathering *in situ* data on aquatic particles, such as net-collection and electronic counting or photography, are not usually suited to observing precise spatial relationships and can destroy the more delicate organisms. The overriding benefit of holography is that it permits nonintrusive and nondestructive observation and analysis of organisms in their natural environment, while preserving their relative spatial distribution. Holography also offers a permanent archive. The use of a short-duration pulsed laser is essential, in order to reduce the effects of vibration or movement in any of the optical components or objects being recorded. Consequently, the object scene is effectively 'frozen' at the recording instant, thereby allowing fast moving particles to be imaged. Compared to a photograph, a hologram captures vastly more data. Sequential holograms can record changes within the volume over a defined period of time and the wide recording dynamic range afforded by a hologram allows images to be captured that would otherwise be lost in noise.

Methodology of Holographic Recording and Replay

Of the many possible recording methods of holography two in particular, 'in-line' and 'off-axis', find use for high-resolution *in situ* imaging underwater. The concepts and fundamental principles outlined here are generally applicable wherever high-precision measurement is needed from a hologram.

In an in-line reference beam hologram (ILH), a single laser beam (Figure 1) is directed through the

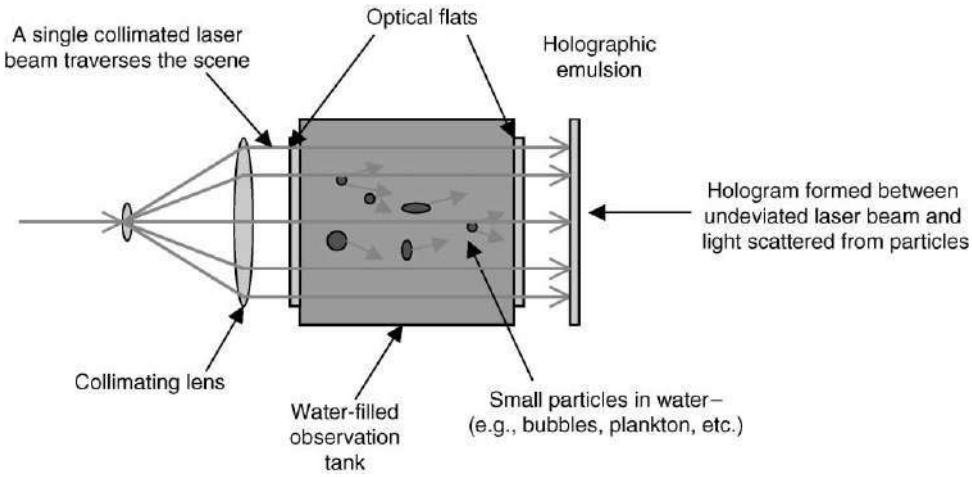

Figure 1 Recording an in-line hologram.

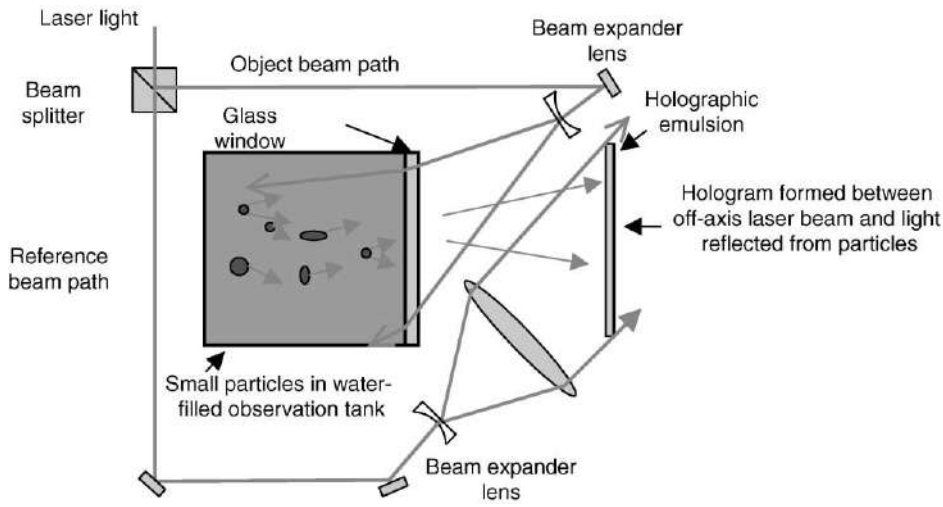

Figure 2 Recording an off-axis hologram.

sample volume towards the holographic plate and records the optical interference between light diffracted by the object and the undeviated portion of the illuminating beam. The subject volume is illuminated in transmission and the scene needs an overall transparency of about 80% so that speckle noise does not seriously degrade image quality. Particles down to around 5 μm dimension can be readily recorded and identified. The particular benefits of ILH are its geometric simplicity, minimization of laser coherence and energy requirements, and high resolution over a large depth of field.

For recording an off-axis reference beam hologram (OAH), a two beam geometry is utilized: one beam illuminates the scene and the other directly illuminates the holographic film at an oblique incidence angle (Figure 2). Interference occurs between diffuse light reflected from the scene and the angularly separated reference beam. On replay, the real and virtual images are similarly angularly separated which makes their interrogation easier. OAH is primarily applied to opaque subjects of large volume. Although there is no real upper limit to the size of particles that can be recorded (this is determined by available energy and coherence of the laser) the practical lower limit to the resolution that can be achieved is around 50 to 100 μm. The scene can be front, side or back illuminated (or all three combined) allowing high concentrations of large particles to be recorded, provided that they reflect sufficient light to the film, and allowing subject visibility to be optimized. OAH provides a more complete record than ILH, in the sense that it gives information on the surface

structure of the object, while ILH records only a silhouette.

Holograms are replayed or reconstructed in a variety of ways, depending on the recording parameters and subsequent processing of the recording medium. In scientific and engineering applications the requirements of high resolution and image fidelity demand that they are replayed with a laser, and viewed with some auxiliary optical system and (usually) a video camera. Depending on the orientation of the hologram, either the virtual image, which is apparently located behind the hologram and seen through it, or the real image, which is formed in space in front of the hologram, will be visible.

Considering first the off-axis hologram in the virtual image mode, the processed hologram is replaced in the position in which it was recorded and illuminated with an exact duplicate of the original reference wave, in terms of its wavelength, curvature, and beam angle. In this mode, which might be termed 'normal' viewing of an off-axis hologram, a virtual image (Figure 3) is observed as if viewing the scene through a window. Other than the colour, this image is almost optically indistinguishable from the original scene; it retains features such as parallax and three-dimensionality and does not suffer from perspective distortion or defocusing of the image.

For data extraction and measurement from a hologram, creation of the virtual image is not the best method of observation. Here the projected real image (Figure 4) is most useful. If we illuminate an off-axis hologram from behind, with a wave which is the exact conjugate of the original (i.e., one which possesses the same wavelength as the original but with the opposite direction and curvature), we generate an image that appears to float in the real space in front of the observer. Often a collimated reference beam is used in recording, since reconstruction is then a simple case of turning the hologram around. No lens is needed to form this image, and it is optically identical to the original wave that it is reversed left-to-right and back-to-front (pseudoscopic). Planar sections of the real image can be optically interrogated by directly projecting onto a video camera (often with the lens removed) which can be moved throughout the 3D volume. It is this 'optical sectioning' feature of a holographic recording that is so useful when nondestructive (noninvasive) evaluation must be made, at high resolution, of collections of objects which extend over a considerable depth.

A series of images taken from the real image of one hologram and captured on a video camera (Figure 5) illustrates how the viewing plane can be moved sequentially through the reconstruction to show firstly one marine organism and then another. The axial separation of the organisms is about 35 mm.

To replay an in-line hologram, it is also placed in a conjugate of the recording beam (Figure 6). The replayed hologram simultaneously forms two images, one virtual, the other real, which are located on the optic axis on opposite sides of the holographic plate. These images are co-axial, and located (for the typical case of collimated recording and replay beams) at equal distances in front of and behind the hologram plane. Although a true 3D real image is produced it has almost no parallax, and cannot easily be seen by the unaided eye. This image has inverted depth perspective and is seen in darkfield against the out-of-focus virtual image.

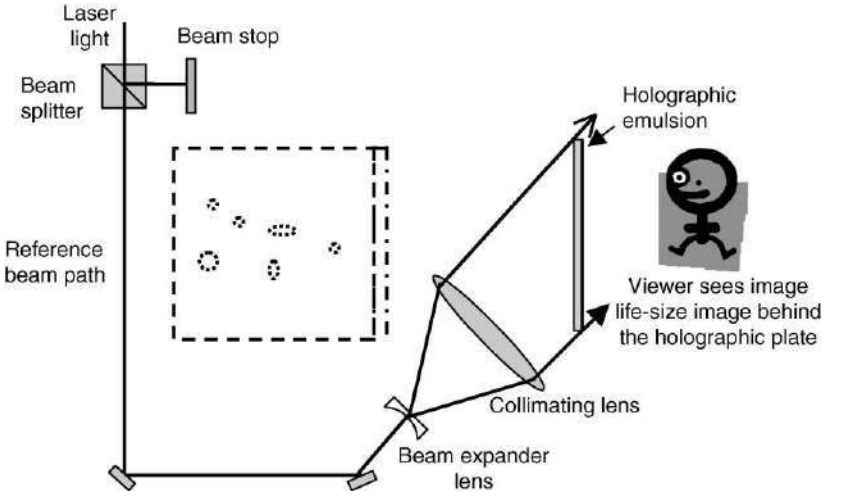

Figure 3 Replay of an off-axis hologram in virtual image mode.

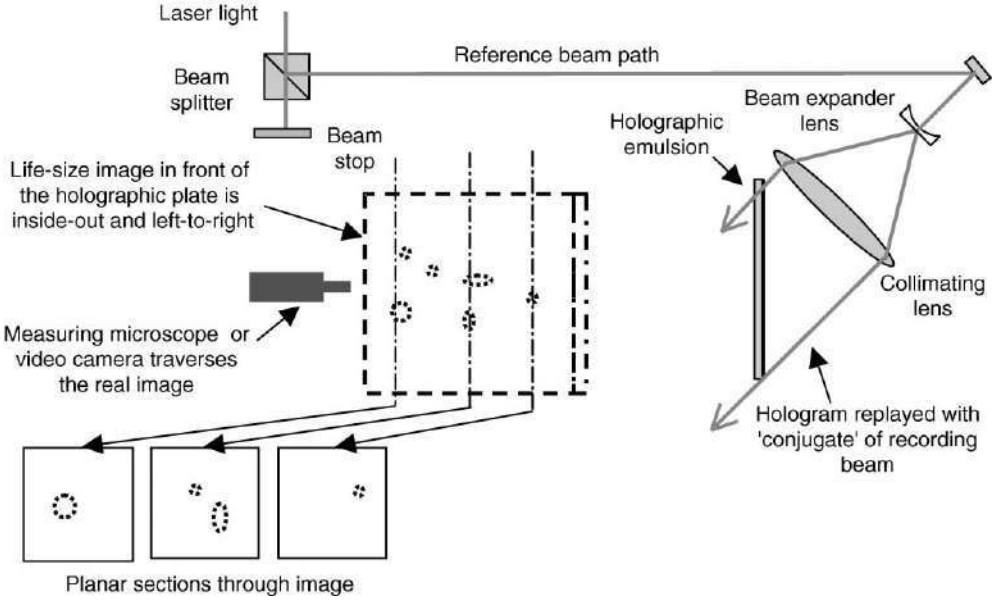

Figure 4 Replay of an off-axis hologram in real image mode.

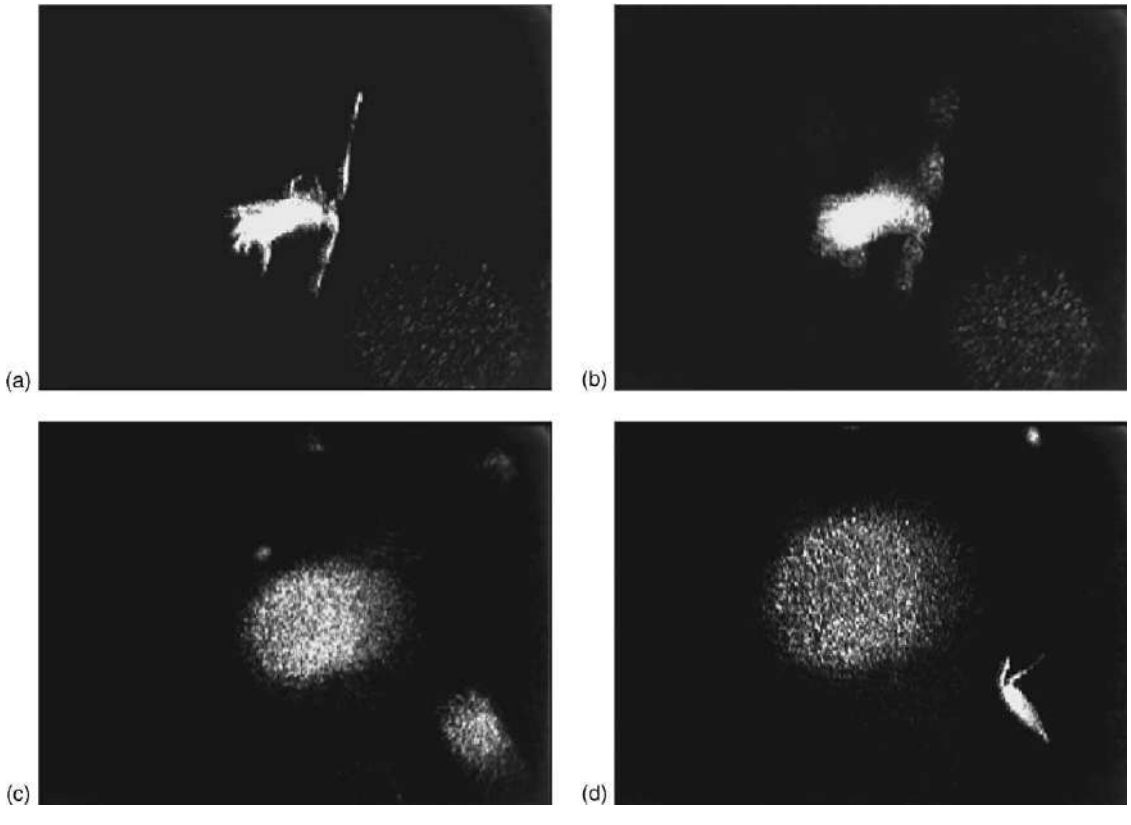

Figure 5 A series of images taken from a single hologram showing the concept of moving the viewing plane through the real image.

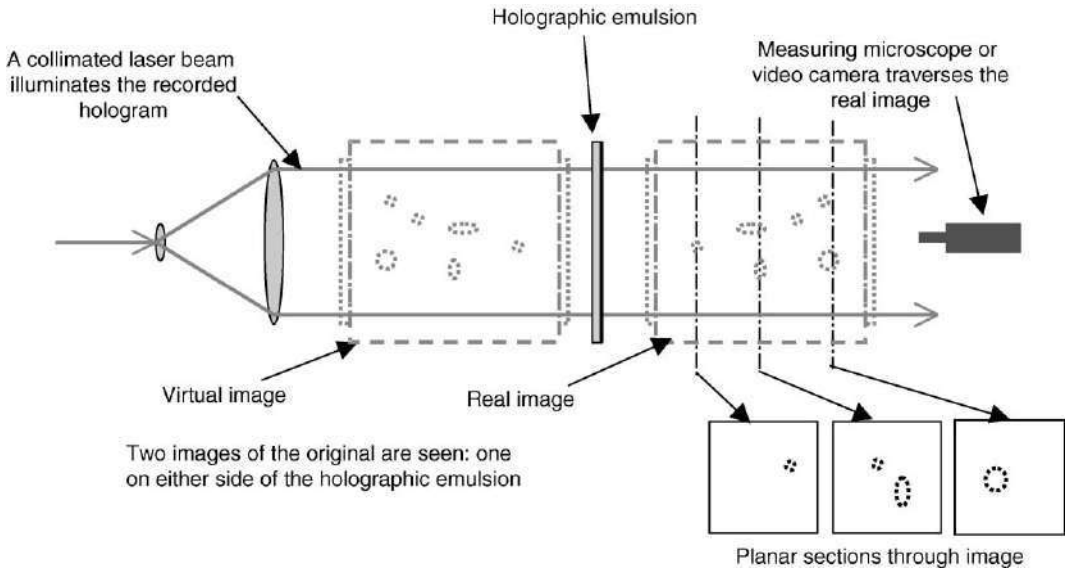

Figure 6 Replay of an in-line hologram.

Holographic Image Quality

In general, the usefulness of holography for accurate inspection and measurement is dependent on its ability to reproduce an image of the object which is low in optical aberrations and high in image resolving power. Regardless of whether or not the hologram is recorded underwater, high resolution, high contrast, and low noise are the dominant requirements, rather than image brightness. In practice, loss of resolution and image degradation can occur at any stage in the recording and replay of a hologram. All the primary monochromatic aberrations to be found in any optical system (spherical aberration, astigmatism, coma, distortion, and field curvature), may be present in the holographic image. The most significant factors affecting this are departure from conjugacy of the reconstruction beam; positional mismatch and misalignment between reference beam and film; degradation of the reconstruction beam due to poor optical quality of the film substrate or optical components; and, distortion of the fringe pattern recorded in the holographic film as a result of chemical processing. For precise and accurate reconstruction of the real image, the reconstruction reference beam should be the exact phase conjugate of that used in recording. Under these conditions, the lateral, longitudinal, and angular magnifications of the real image with all equal unity and aberrations will be reduced to a minimum.

Diffraction-limited resolution of a holographically produced real image, in the absence of all aberrations, is usually defined from the ability to distinguish two points in the image separated by a distance r. The resolution as defined in this way is dependent upon reconstruction wavelength, λ_c, the separation between hologram and reconstructed image, s, and the effective aperture, D, of the hologram:

$$r = 1.22 s\lambda_c/D \qquad [1]$$

This relationship is well-known in imaging optics: it defines the ideal resolving power of a lens and is equally applicable to holography. In holography, though, the presence of speckle effects introduced by the coherence of the light and the finite aperture of the viewing system influence the resolving power. In practice, the minimum resolution is increased by a factor of two to three to take account of speckle. For off-axis holograms the choice of film size (effective aperture) has a considerable bearing on image resolution. Typically, for 100 mm diameter film the theoretical image resolution (neglecting speckle effects) is around 6 μm (150 lp/mm) with a 532 nm laser at a target distance of 1 m. For in-line holograms (with collimated illumination) the film size also defines the transverse dimensions of the volume recorded.

Wavelength matching between recording and reconstruction beams require that, if possible, the same type of laser be used in both replay and recording. Of course, this is not possible in field situations when the use of a pulsed laser is essential to ensure that the fringe pattern is not blurred by any movement in the system. In such applications, laboratory replay of the hologram may require the use of a tuneable laser to match wavelengths

precisely, although in practice an argon laser provides a useful selection of wavelengths. A mismatch between record and replay wavelengths will lead to distortion in the image, because longitudinal and lateral magnifications will no longer have the same value.

In contrast to off-axis holograms, the interference patterns on an in-line hologram are localized about the geometric shadow of the objects. The lack of a spatially separate reference beam leads to a requirement that there be a large fraction (>80%) of the beam unobscured. This limits the size of particles that can be recorded and sets an upper limit on the recordable concentration of particles (around 40 particles cm^{-3} for 20 μm diameter particles and a recorded depth of 1 m). There is also a need to balance the size of the particles with the object-to-hologram distance. The range of object size recordable on an in-line hologram has an upper limit set by the requirement to record in the far-field of the object, so that the conditions of Fraunhofer recording are satisfied:

$$z > d^2/\lambda \qquad [2]$$

where z is the object-to-film distance and d is the maximum dimension of the object to be recorded. In-line holograms are relatively easy to replay as the only requirement is to place the hologram normal to the replay beam. Following this criterion, the upper limit for good size measurement is about 2 to 3 mm particle size over a recording distance of a meter. Unlike off-axis holograms which are extremely sensitive to misalignment, the in-line geometry minimizes the effects of film rotation about its axes. A collimated recording beam is by far the easiest to duplicate upon replay. Small deviations from collimation can significantly affect conjugacy and hence resolution, introducing aberrations (mainly spherical) and a depth-dependent magnification (e.g., a cuboid replays as a trapezoidal prism).

In general, the interference field is captured on photographic emulsions; these may be laid on acetate or polyester film bases or on glass plates. If film is used, it is essential that it be held as flat as possible between thin flat glass plates or under vacuum, and should not be stretched or put under any strain during exposure. An in-line hologram does not require an extremely high-resolution recording medium. The fringe spacing is not as fine as the typical carrier fringe frequency used in off-axis holograms, and holograms can be made on conventional slow photographic emulsions. In practice, however, ultra-fine grained holographic film is almost always used since the intensity of the forward scattered light (a function of grain size) is then reduced: this is important in increasing the signal-to-noise ratio, as the replayed image is viewed against a background of the attenuated replay beam. Since the proportion of the illuminating beam that goes into the images is very low, this lack of angular separation requires that the average optical density of the hologram is high. The high optical density attenuates the undiffracted light, thus the image appears bright against a dark background, resulting in contrast reversal.

The exposed film is chemically processed to render the interference field permanent. This process is a crucial step in the holographic procedure, and especially critical for good off-axis holograms. Some of the factors, which have to be considered, include image brightness, image resolution, reconstruction wavelength, emulsion shrinkage, humidity, and noise level. It has been well established that for bright off-axis holograms on silver halide film, bleaching of the emulsion with its consequent conversion into a phase hologram is essential to maximize diffraction efficiency. Some forms of bleaching produce a nonuniform shrinkage of the emulsion which will give rise to astigmatism in the reconstructed image. Bleaching significantly increases the background noise levels of the hologram and introduces a high degree of variability into the result. In practice, the best repeatability will be obtained with amplitude processing if the low brightness can be tolerated.

Underwater Holography

When holography is extended to the inspection of underwater scenes, the hologram is recorded in water but the image is replayed, in the laboratory, in air. Because of the refractive index difference between recording and replay spaces, optical aberrations will be introduced into the reconstructed image, and will seriously impair the potential for precision measurement, unless some method of aberration reduction is incorporated. In ILH, only spherical aberration is significant since both reference and object beam angles are normal to the recording plane. However, in OAH the dominant aberrations are astigmatism and coma, which increase with the field angle of the subject in the reconstructed image. These limit resolution and introduce uncertainty in co-ordinate location, since the focal 'point' is now distributed over a finite region of image space. A key point here is an appreciation that the additional aberrations introduced are entirely due to the refractive index mismatch between object and image spaces and are unconnected with the holographic process itself: they are in essence the same distortions seen when looking into a fish-tank. Furthermore, the water

itself may be expected to influence the quality of the images produced. An increase in the overall turbidity of the water will adversely affect both in-line and off-axis techniques and would be expected to create a background noise that will reduce image fidelity.

Techniques, such as immersing the holographic plate in water during recording or direct replay back into water, have been used to minimize underwater aberrations, but such methods are generally impractical in the field. One practical solution, unique to holography, compensates the change of effective wavelength of light as it passes through water, by allowing a deliberate mismatch of recording and replay reference beams to balance the refractive index mismatch.

For holograms recorded and replayed in air, a usual prerequisite is that the reconstruction wavelength should remain unchanged from that used in recording. By noting the dependence of wavelength on the refractive index, we can apply the more general condition that it is the ratio λ/n that must remain constant. This relationship suggests that a hologram, immersed in water, and recorded at a wavelength λ_a will produce an aberration-free image in air when replayed at a reconstruction wavelength, λ_c, which is itself equivalent to the wavelength of light when passing through water, λ_w, i.e.:

$$\lambda_c = \lambda_w = \lambda_a n_a / n_w \qquad [3]$$

where n_a and n_w are the respective refractive indices of air and water. Thus, in principle, the aberrations may be completely eliminated by the introduction of an appropriate wavelength change between recording and replay. If a green (532 nm) laser is used in recording, the ideal replay wavelength is about 400 nm (i.e., 532 nm/1.33). However, no convenient laser lines exist at exactly this wavelength. Furthermore, complete correction assumes that the entire recording system be located in water. Since this is both impractical and undesirable, holograms are usually recorded with the holographic film in air behind a planar glass window. The additional glass and air paths affect the compensation of aberrations. However, third-order aberration theory shows that if the window-to-air path length ratio is appropriately chosen for a specific replay wavelength, then aberration balancing occurs and residual aberrations are reduced to a minimum over a wide range of field angles and object locations. For an air gap of 120 mm and a BK7 window of 30 mm thickness, a blue replay wavelength of 442 nm (HeCd laser) achieves a good performance over a full field angle of about 40°.

In most underwater applications the objects (or the camera) are in motion during the exposure. The effect of this motion is to blur out the finer fringes, and thus reduce resolution and contrast. In-plane motion is the most severe, and adopting the experimentally verified criterion that the maximum allowable motion is less than one-tenth of the minimum required fringe spacing of the smallest object, then for in-line holograms the maximum object motion must be less than one-tenth of the object's dimension. For particles of 10 μm dimension, and a typical Q-switched YAG laser pulse duration of 20 ns, a high transverse velocity of up to 50 m s^{-1} can be tolerated. Off-axis holograms are more demanding in their requirements and the maximum allowed velocity is reduced to about 3 m s^{-1}. This is, however, more than adequate for most field applications of the technique.

Underwater Holographic Cameras

The first-known use of holography for recording living marine plankton was by Knox in 1966, who recorded in-line holograms (ILH), using a short coherence length ruby laser, of a variety of living marine plankton species in a tank. However, to avoid problems with refractive index and wavelength changes, the holographic plate was immersed directly in water, which is completely impractical for field use. It was not until the late 1970s that quantitative data from replayed holograms of marine particles recorded *in situ* were obtained by Carder and his team. Small volumes (10 or 20 cm^3) were recorded using ILH and a low-power HeNe laser, which limited the technique to slowly moving objects. Later, Heflinger used the off-axis geometry with both pulsed argon-ion (514 nm) and pulsed xenon (535 nm) lasers to record plankton on 35 mm film. In order to compensate for the aberrations introduced by replaying the real image in air, the real image was projected back into the water tank using a conjugate beam from a continuous-wave argon laser. The sides of their tank were flexible so that the real image could be placed close to the exit window and viewed via an external microscope. Knox later developed a ship-borne in-line holographic camera which was deployed from the *RV Ellen B Scripps*. A pulsed xenon laser was used to record water volumes of up to 10^5 cm^3 to a water depth of 100 m.

It was not until the late 1990s that significant use of underwater holography for plankton recording came to the fore again. Joseph Katz at Johns Hopkins University, Baltimore, USA developed a camera based

around a ruby laser to record in-line holograms on rolls of film, in a 63 mm diameter sample volume that can be varied from 100 to 680 mm in length. The camera housing utilizes a two-tube design with one tube holding the laser and another containing film holder and power units. The recording volume is situated 890 mm above the camera to minimize disturbance to flow. The camera is designed as a 'Lagrangian drifter' to move with the water current and was successfully deployed in Chesapeake Bay, Baltimore, and Solomon's Island, Maryland. Holograms were recorded of various species of phytoplankton and zooplankton and represented a significant advance in the deployment of sub-sea holography.

From the early 1980s, a team led by Aberdeen University, Scotland, investigated the use of underwater holography for offshore applications. Primary interest was in inspection of subsea pipelines for corrosion and damage, and much of the work concentrated on an analysis of the optical aberrations introduced by recording in water and replaying in air and methods of compensating for such aberrations. However, this work was later adapted for holography of plankton and led to the consequent development of the HoloMar system. The HoloMar camera was the first to incorporate simultaneous ILH and OAH recording.

More recent developments have utilized digital or electronic recording of holograms in the in-line mode, and it is to be expected that much more use will be made of 'e-holography' in the future.

For field applications of holography, it is essential to record the holograms using a pulsed laser in order to create stable interference fringes. For subsea purposes, a laser with a wavelength in the blue-green region of the spectrum will match the oft-quoted peak transmission window of seawater. In practice, a frequency-doubled Nd-YAG laser with a wavelength of 532 nm is a good choice, although ruby lasers operating at 694 nm are also acceptable if transmission distance is not an issue. Usually OAH requires more energy and places stricter constraints on coherence properties than ILH. Beam quality is governed by quality of the optical components in the beam path and also by the output beam of the laser itself. Because off-axis holography uses diffuse front/side illumination, energies of up to several hundred millijoules are often required for large (up to a meter in diameter) poorly reflective objects, and coherence lengths of greater than a meter are needed to record large scene depths. To obtain a long coherence length, and thus enable large-volume objects to be recorded, it is necessary that the laser operate in a single longitudinal cavity mode so that the bandwidth of the laser line is kept small. To enable smooth even illumination of both subject and holographic emulsion, good spatial beam quality is needed. This usually dictates laser output beams with either a Gaussian (TEM_{00}) mode profile or a flat profile, and spatial filtering during beam expansion. Compared to off-axis holography, the demands placed upon lasers for ILH are minimal. The only essential requirement is for a good spatial mode structure. Energy requirements are moderate: as little as 10 mJ is enough since little light is lost during the passage of the beam through the recording volume. The laser coherence length can be quite short, and only depends on the

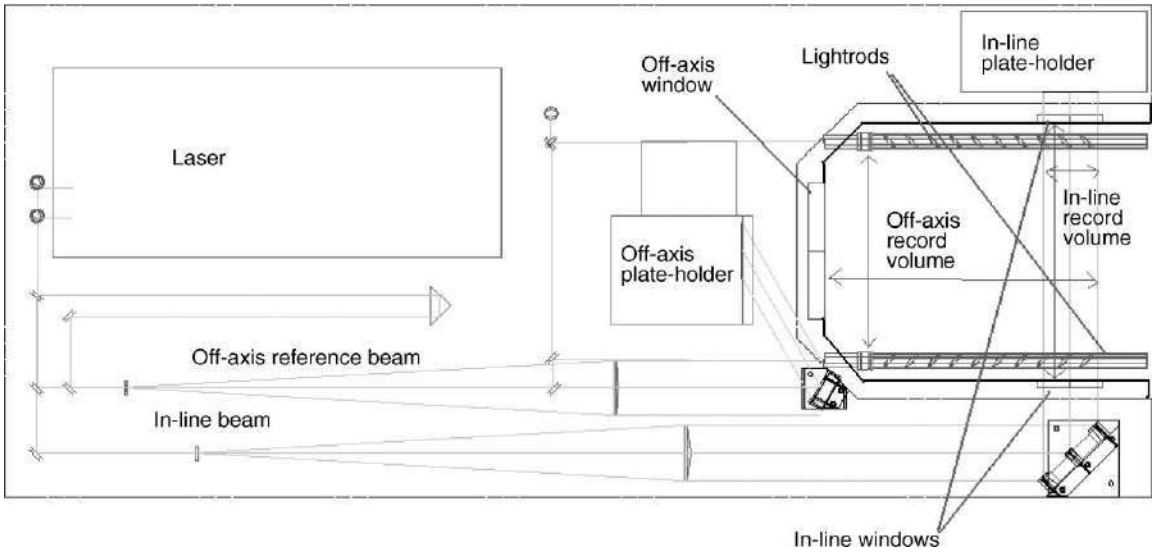

Figure 7 The HoloMar camera configuration showing the separate in-line and off-axis beam paths.

effective hologram area, thus lasers operating on more than one longitudinal mode are usually quite suitable.

The HoloMar System

The HoloMar system (shown diagrammatically in Figure 7) was developed to record simultaneous in-line and off-axis holograms of partially overlapping, but orthogonal, views of the water column. Although the recording views are orthogonal to each other, the arrangement provides an element of cross-correlation between holograms. It is designed around a Q-switched, frequency-doubled Nd-YAG laser with an output energy of 700 mJ in a single pulse of less than 10 ns duration and a coherence length in excess of 2 m. The laser output is split into two beams: one of 100 mJ and the other of 600 mJ energy. The 100 mJ beam is further split into two equal energy beams and expanded and collimated to a diameter of about 90 mm. One path forms the illuminating beam for the in-line mode and the other path forms the reference beam for the off-axis mode. The in-line path records a water column of 470 mm by 90 mm diameter (9,000 cm^3) at 400 mm from the front face of the housing, whereas the off-axis hologram records almost the entire volume (50,000 cm^3) delineated by the front window to the end of the arms that house the in-line system (see Figure 7). Three 'lightrods', encompassing hollow perspex tubes with 10 glass beam-splitter plates distributed equally along the length, positioned at 45° to the main axis, provide roughly even side illumination to the off-axis volume. The front face of the camera is shown in Figure 8; the lightrods are positioned on either side of the recording volume (two on one side, one on the other). The exit window for the in-line beam path and the off-axis window can also be seen.

Two motor-driven plate holders and transport mechanisms hold up to 20 plates for in-line and 24 for off-axis holograms. One pair of plates can be exposed every 10 seconds, or they can be exposed separately. The camera is remotely operated from the surface via a control console (PC). A network of micro-controllers serves as the backbone of the in-camera control system and communication between topside and camera modules is implemented using a Controller Area Network (CAN) bus via an umbilical cable. The entire camera is enclosed in a water-tight stainless steel housing of 1,000 mm diameter by 2,200 mm length. The prototype system is capable of either ship deployment or attachment to a fixed buoy and allows recording down to a water depth of 100 m.

The HoloMar camera is shown, in Figure 9, being lowered into Loch Etive, a sea loch on the west coast of Scotland, on its maiden dive. Figures 10 and 11 show holographic images of plankton recorded on this cruise. The first shows a view from a virtual

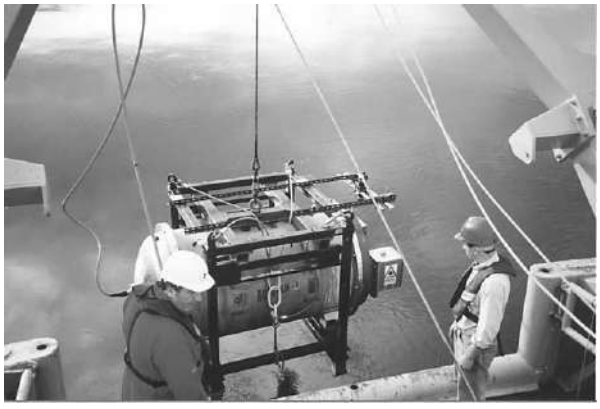

Figure 9 The HoloMar camera being lowered into Loch Etive, Scotland.

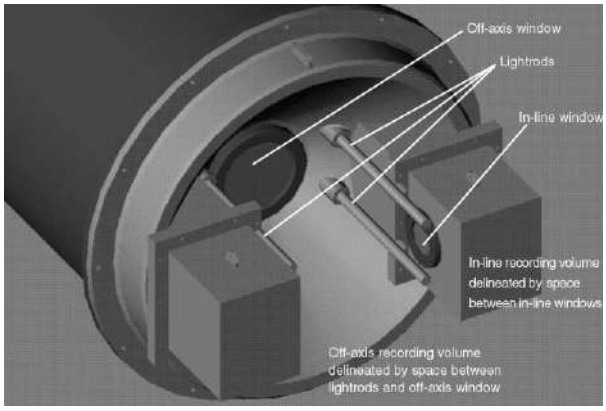

Figure 8 Front-end of the HoloMar camera showing the 'lightrods' and in-line and off-axis windows.

Figure 10 A photograph taken from a virtual image replay of an off-axis hologram recorded in Loch Etive.

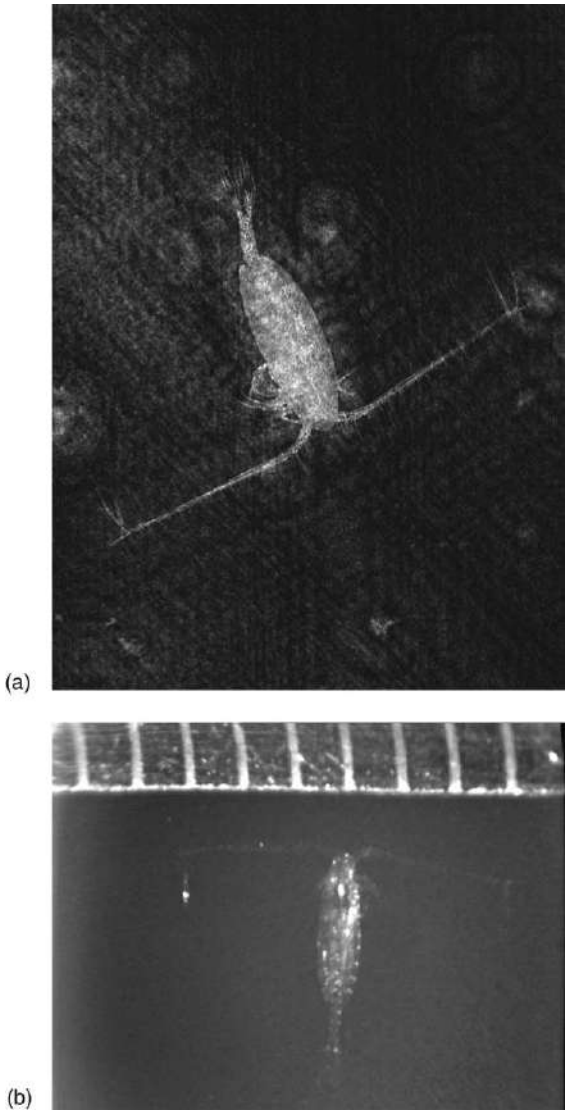

Figure 11 Images of calanoid copepods from (a) an in-line hologram recorded at 70 m depth and (b) an off-axis hologram recorded at 60 m depth, the millimeter scale is added at the replay stage. Each organism is approximately 4 mm long.

replay of an off-axis hologram at a wavelength of 514 nm: this is what a viewer would see when looking from the inside of the holocamera into the laser-illuminated water. Each bright dot signifies an object of interest; two fiducial wires can also be seen fixed diagonally across the arms at the end of the camera, some 300 mm from the off-axis window. In the lower left of the image a long 'string-like' object about 40 mm long is probably 'floc', the dead organic matter that forms the food-stuff of many zooplankton. Objects like this would be destroyed in net-collection. Figure 11 shows images taken from the real image projection of in-line and off-axis holograms. The holograms are replayed using a collimated beam at wavelengths of 514, 488, 457, or 442 nm, depending on location in the image volume. Some of the contrasting features between in-line and off-axis holograms are apparent: the in-line image shows clearly the outline of the organism with detail resolvable down to about 10 μm, whereas the off-axis hologram, although not achieving this resolution, reveals the body structure of the organism.

List of Units and Nomenclature

d	dimension of object to be recorded
D	effective aperture of hologram
n	refractive index
n_a	refractive index of air
n_w	refractive index of water
r	minimum resolvable separation between two points on the hologram
s	distance between hologram and image location
z	distance between hologram and object location
λ	wavelength
λ_a	in-air wavelength of recording beam
λ_c	reconstruction wavelength
λ_w	effective wavelength of recording beam in water

See also

Holography, Applications: Holographic Recording Materials and Their Processing. **Holography, Techniques:** Computer-Generated Holograms. **Phase Control:** Phase Conjugation and Image Correction.

Further Reading

Beers JR, Knox C and Strickland JDH (1970) A permanent record of plankton samples using holography. *Limnology and Oceanography* 15: 967–970.

Carder KL (1979) Holographic microvelocimeter for use in studying ocean particle dynamics. *Optical Engineering* 18: 524–525.

Hariharan P (1984) *Optical Holography*, 2nd edn. Cambridge, UK: Cambridge University Press.

Heflinger LO, Stewart GL and Booth CR (1978) Holographic motion pictures of microscopic plankton. *Applied Optics* 17: 951–954.

Hobson PR and Watson J (2002) The principles and practice of holographic recording of plankton. *Journal of Optics A: Pure and Applied Optics* 4: S34–S49.

Katz J, Donathay PL, Zhang J, King S and Russell K (1999) Submersible holocamera for detection of particle

characteristics and motions in the ocean. *Deep-Sea Research* 46: 1455–1481.
Kilpatrick JM and Watson J (1993) Underwater hologrammetry: reduction of aberrations by index compensation. *Journal of Physics D: Applied Physics* 26: 177–182.
Knox C (1966) Holographic microscopy as a technique for recording dynamic microscopic subjects. *Science* 153: 989–990.
Malkiel E, Alquaddoomi O and Katz J (1999) Measurements of plankton distribution in the ocean submersible holography. *Measurement Science and Technology* 10: 1141–1152.
Owen RB and Zozulya AA (2000) In-line digital holographic sensor for monitoring and characterizing marine particulates. *Optical Engineering* 39: 2187–2197.
Stewart GL, Beers JR and Knox C (1970) Application of holographic techniques to the study of marine plankton in the field and the laboratory. *Proceedings of SPIE* 41: 183–188.
Sun H, Dong H, Player MA, *et al.* (2002) In-line digital holography for the study of erosion processes in sediments. *Measurement Science and Technology* 13: L1–L6.
Watson J, Alexander S, Craig G, *et al.* (2001) Simultaneous in-line and off-axis subsea holographic recording of plankton and other marine particles. *Measurement Science and Technology* 12: L9–L15.

Holographic Recording Materials and Their Processing

H I Bjelkhagen, De Montfort University, Leicester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The first off-axis transmission holograms in the USA were recorded on Kodak 649F spectroscopic silver halide material (now discontinued). For the first reflection holograms, which were recorded in Russia, a very fine-grain emulsion (Lippmann emulsion) had to be prepared to resolve the fine interference pattern associated with such holograms. Since then various types of recording materials for holography have been developed and manufactured, including improved silver halide emulsions.

The following provides a brief description of the holographic recording process. For recording holograms and holographic optical elements (HOEs), dichromated gelatin, photopolymer materials, and photoresist materials are employed, in addition to silver halide materials. Existing holographic recording materials are described, as well as suitable processing methods to obtain high-quality holograms and HOEs.

Holographic Recording Materials

To produce a hologram, a micropattern, caused by interference between the object beam and the reference beam, is recorded in a light-sensitive high-resolution material. The quality of a holographic image depends on a number of factors, such as: the geometry and stability of the recording setup; the coherence of the laser light; the reference and object beam ratio; the type of hologram produced; the size of the object and its distance from the recording material; the recording material and the emulsion substrate used; the processing technique applied as well as the reconstruction conditions. The material must comply with certain requirements to be suitable for this purpose. The most important of these concerns the resolving power of the material. The recording material must be able to resolve the highest spatial frequencies of the interference pattern created by the maximal angle θ between the reference and the object beams in the recording setup (Figure 1).

If λ is the wavelength of the laser light used for the recording of a hologram and n the refractive index of the emulsion, then the closest separation d_e between the fringes in the interference pattern created by the angle θ between the reference and the object beams in the recording setup is

$$d_e = \frac{\lambda}{2n\sin(\theta/2)} \qquad [1]$$

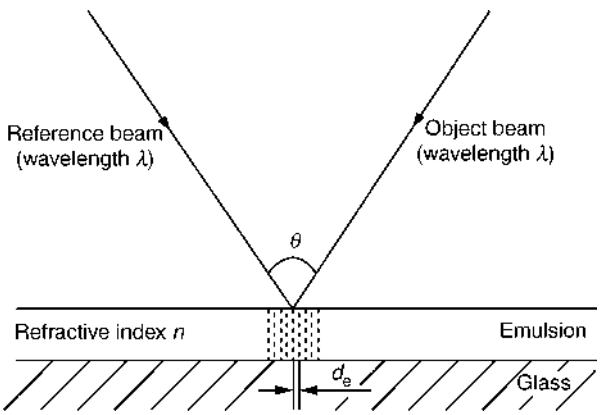

Figure 1 Demand on resolution for recording a hologram.

For high-quality holograms the resolution limit of the material must be higher than the minimum value obtained according to the above formula. For a reflection hologram recorded in blue light ($\lambda = 400\,\text{nm}$) with an angle of 180° between the beams, a minimum resolving power of 7600 lines $(\text{mm})^{-1}$ is required.

In holography, the resolution of the holographic image and the resolving power of the recording material are not directly related to the way in which they are in photography.

Equation [1] gives the minimum resolving power needed to record a hologram. The resolution of the image depends on the area (diameter D) of the recording material, the recording laser wavelength (λ), and the distance (L) between the recording material and the object (Figure 2).

Theoretically, the resolution of the holographic image is the diffraction-limited resolution that can be obtained when the information is collected over an aperture equal to the size of the recording holographic plate. In principle, the larger the holographic plate the better the resolution will be. To obtain high image resolution, it is important that the recorded interference fringes will not change or be distorted during processing. To prevent that from happening, a stable support for the emulsion (like a glass plate) is needed and the processing method applied must not affect the recorded fringe pattern in the emulsion. In coherent imaging systems laser speckles are present which may also have an effect upon image resolution.

There are two main types of holograms: amplitude and phase holograms. In a pure amplitude hologram only the absorption varies with the exposure (after processing), whereas in a pure phase hologram, either the refractive index or the emulsion thickness varies with the exposure. If the hologram is thin, $d \approx 0$, phase variations are then caused by surface relief variations only. If the hologram is thick and has a negligible surface relief ($\Delta d = 0$), phase variations are caused by index variations only. In many cases, the phase modulation in a hologram is a combination of the two different extreme types (a complex hologram).

For amplitude holograms, amplitude transmission T_a against exposure or log exposure is used (Figure 3). For a phase hologram, the corresponding curve is the phase shift against the log exposure relation (Figure 4).

One of the most important hologram characteristics is its diffraction efficiency η. It is defined as the diffracted intensity E_i of the wanted diffraction order of the hologram in relation to the incident

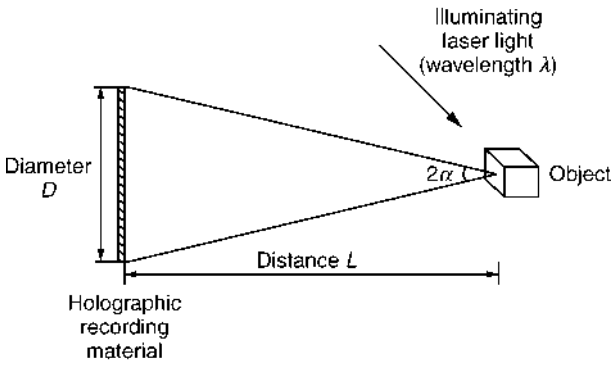

Figure 2 Resolution of a holographic image depends on the area (diameter D) of the recording material, the recording laser wavelength λ, and the distance L between the recording material and the object.

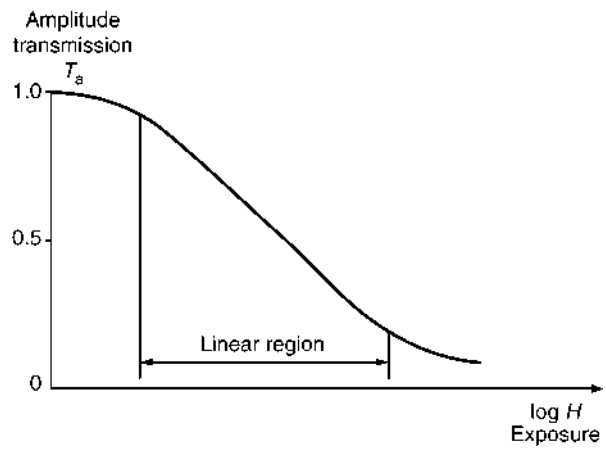

Figure 3 Amplitude transmission T_a versus the logarithm of exposure H used for the characterization of amplitude transmission holograms.

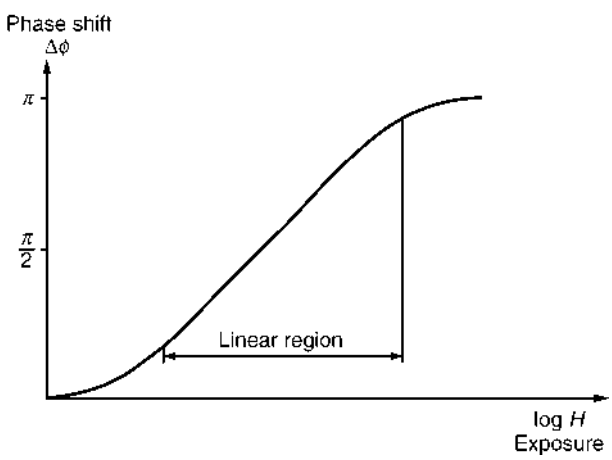

Figure 4 Phase shift $\Delta\phi$ versus the logarithm of exposure H used for the characterization of phase transmission holograms.

intensity of the reconstruction beam E_r : $\eta = E_i/E_r$. In the definition of η, the incidental light losses caused by surface reflection and base absorption should be subtracted, but are often ignored. Other important hologram characteristics are the signal-to-noise ratio (SNR) and the dynamic range of the recording material.

The resolution capability of an image reproduction process is normally described by the Modulation Transfer Function (MTF). Briefly, a test pattern containing a sinusoidal variation in illuminance combined with a continuous variation in spatial frequency along one direction is recorded. The modulation M of the pattern in the test target is

$$M = (H_{\mathrm{MAX}} - H_{\mathrm{MIN}})(\mathrm{H}_{MAX} + \mathrm{H}_{MIN})^{-1} \qquad [2]$$

where **H** is the exposure incident on the photographic material. When this pattern is recorded in the material, light scattering will take place in the emulsion, which will reduce the original contrast of the pattern. Therefore, the modulation of the pattern will be decreased, in particular at high spatial frequencies. The effective exposure modulation M' will then be

$$\mathrm{M}' = (H'_{\mathrm{MAX}} - H'_{\mathrm{MIN}})(H'_{\mathrm{MAX}} + H'_{\mathrm{MIN}})^{-1} \qquad [3]$$

where $\mathbf{H}'$ is exposure in the emulsion.

The original modulation M is constant and accurately known; it is also independent of the spatial frequency. After the tested emulsion has been processed, the corresponding 'exposed' modulation is obtained from the density variations. The ratio between modulation M' in the emulsion and modulation M of the incident exposure is called the modulation transfer factor, also called response

$$R = M(M')^{-1} \qquad [4]$$

If the response is plotted as the function of spatial frequency, this curve will then be the modulation transfer function of the material. The MTF for both a holographic emulsion and a conventional photographic emulsion are illustrated in Figure 5.

There are many ways of categorizing holograms but here we are mainly interested in definitions related to the recording material and its processing. One of the valid criteria is the thickness of the recording layer (as compared to the interference fringe spacing within the layer), i.e., the layer coated on the material substrate. Holograms can thus be classified into 'thin' or 'thick' (sometimes also called 'plane' or 'volume' holograms, respectively). To distinguish between the two types, the Q-parameter is normally

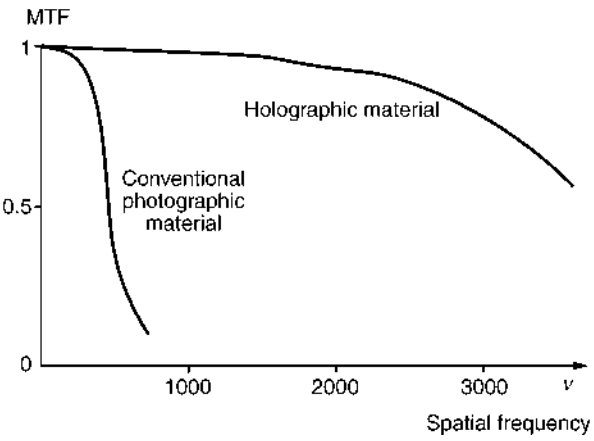

Figure 5 Graphical representation of the modulation transfer function (MTF) curves for both a conventional emulsion and a holographic emulsion are shown.

used; it is defined in the following way:

$$Q = 2\pi\lambda d(n\Lambda^2)^{-1} \qquad [5]$$

where λ is the wavelength of the illuminating light, d is the thickness of the layer, n is the refractive index of the emulsion, and Λ the spacing between the recorded fringes. A hologram is considered thick if $Q \geq 10$, and thin if $Q \leq 1$. Holograms with Q-values between 1 and 10 can be treated as either thin or thick.

Recording Materials

The most common holographic materials used for recording holograms are the following: silver halide emulsions; dichromated gelatin emulsions; photopolymer materials; photoresist materials; thermoplastic materials; bacteriorhodopsin; and photorefractive crystals.

The following materials can also be considered for holographic recordings, but not many practical applications have been reported so far: chalcogenide glasses; ferroelectric-photoconductors; liquid crystals; magneto-optic films; metal and organic dye ablative films; photochromic and photodichroic materials; photothermorefractive glass; and transparent electrophotographic films.

Sensitivity of Photographic and Holographic Materials

The time of exposure for a given material will depend on the sensitivity of the material used, as

well as on the intensity of the interference pattern. Some holographic materials must be processed after exposure in a specific way to obtain a hologram. The recorded intensity variations are converted during this processing step to local variations in optical density or refractive index and/or thickness of the recording layer.

Exposure H is defined as the incident intensity E times the time t of exposure of the recording material. If the intensity is constant during the whole exposure time, which is usually the case, then

$$H = Et \qquad [6]$$

Holographic materials are usually characterized using radiometric units. The radiometric equivalent of illuminance is irradiance. The unit of irradiance is $W(m)^{-2}$ and the exposure will then be expressed in $J(m)^{-2}$. The sensitivity of a holographic emulsion is most often expressed in $\mu J(cm)^{-2}$ or $mJ(cm)^{-2}$. Knowing the sensitivity of the material used and having measured the irradiance at the position of the holographic plate, the exposure time can be calculated using the above formula, i.e.,

$$\text{Exposure time} = \text{sensitivity/irradiance}$$

Holographic materials are sensitized in such a way that they are optimized for laser wavelengths commonly used in holography.

Silver Halide Materials

A silver halide recording photographic material is based on one type, or a combination of silver halide crystals embedded in a gelatin layer. The emulsion is coated on a flexible or stable substrate material. Silver-halide grain sizes vary from about 10 nanometers for the ultra-fine-grain holographic emulsions to a few micrometers for highly sensitive photographic emulsions (Table 1).

Silver salts are only sensitive to UV light, violet, and deep blue light. Therefore special sensitizers (dyes) must be added to the emulsion to make it sensitive to other parts of the spectrum. Orthochromatic emulsions are sensitive to green light. If the material has been sensitized to both green and red light, it is said to be panchromatic.

In the following, silver halide materials manufactured by photographic and holographic companies are described. The main difference between the new materials and the previous emulsions from Agfa, Ilford, etc., is that for the current silver halide emulsions, the grain sizes are smaller. The present holographic silver halide emulsion manufacturing process is more or less based on the Russian ultrafine-grain emulsion technology.

Table 1 Silver halide emulsion grain sizes

Type of emulsion	*Average grain diameter [nm]*
Ultra-fine-grain holographic emulsion	10–30
Fine-grain holographic emulsion	30–50
Fast holographic emulsion	50–100
Chlorobromide paper emulsion	200
Lithographic emulsion	200–350
Fine-grained photographic emulsion	350–700
Fast photographic emulsion	1000–2000
Fast medical X-ray emulsion	2500

Substrates for Holographic Emulsions

The material on which the emulsion is coated has a strong bearing on the final quality of the hologram. The best choice is often a glass plate as it is mechanically stable and optically inactive. High resolution imaging, hologram interferometry, holographic optical elements (HOEs) and spatial filters are a few examples where a very stable emulsion support is important. Mass produced holograms are mainly recorded on film substrates. However, the use of film has many advantages as compared to that of glass (breakage, weight, cost, size, etc.). Film substrates are used exclusively in the production of large-format holograms. Film substrates are of mainly two types: a polyester (polyethylene terephthalate); or a cellulose ester, commonly triacetate (cellulose triacetate) or acetate–butyrate. Polyester is mechanically more stable than triacetate film and it is also less sensitive to humidity. Because of the higher tensile strength, the polyester film can be made thinner than the triacetate film. On the other hand, polyester is birefringent, which can cause problems when recording holograms. In Table 2 most current commercial silver halide materials for holography are listed.

Processing of Silver Halide Emulsions

In silver halide emulsions a latent image is formed during exposure, which later is developed to a silver image. Processing (development, fixing, and bleaching) silver halide emulsions requires some expertise to obtain high-quality holograms. With the exception of DCG emulsions, other holographic materials are easier to process and more straightforward. Here it is not possible to describe all techniques employed and recipes used for silver halide processing. The reader is referred to the Further Reading section for more details.

Table 2 Commercial holographic recording materials

Material	*Thickness [μm]*	*Spectral sensitivity [nm]*	*Sensitivity [μJ (cm)$^{-2}$] at*				*Resolv. power [lp (mm)$^{-1}$]*	*Grain size [nm]*
			442	*514*	*663*	*694*		
SILVER HALIDE EMULSIONS:								
Slavich								
Red PFG-01	7	<700	–	–	80	–	>3000	35–40
Red PFG-03M	7	<700	–	–	1500	–	>5000	10–20
Green VRP-M	7	<550	–	80	–	–	>3000	35–40
Pan PFG-03C	9	400–700	1000	2000	1000	–	>5000	10–20
Colourholographic								
Red BB-700	7	<700	–	–	50	150	>2500	50–60
Red BB-640	7	<650	–	–	150	–	>4000	20–25
Green BB-520	7	<540	150	150	–	–	>4000	20–25
Blue BB-450	7	<470	150	–	–	–	>4000	20–25
Kodak								
Red 131PX	9	<650	2	–	0.5	–	>1250	70
Red 131CX	9	<650	2	–	0.5	–	>1250	70
Red 120PX	6	<750	60	–	40	40	>2500	58
Red 120CX	6	<750	60	–	40	40	>2500	50
FilmoTec-ORWO								
Red HF65	10	<650	–	–	1500	–	5000	30
Green HF53	10	<550	–	700	–	–	5000	30
Ultimate								
Ultimate 15	7	<700	–	150	150	150	>5000	15
Ultimate 08	7	<650	120	200	200	–	>7000	8
DICHROMATED GELATIN EMULSIONS								
Slavich								
Blue PFG-04	16	<515	1×10^5	2.5×10^5	–	–	10000	–
FilmoTec-ORWO								
Blue GF 40	6/20	*	*	–	–	–	–	–
* delivered: non-sensitized								
Holotec								
Custom orders only								
THERMOPLASTIC MATERIALS								
Tavex America								
Pan TCC-2	–	<800	1	1	1	1	1500	–
MD Diffusion								
Custom orders only								
PHOTORESIST MATERIALS								
Towne Technologies								
UV-Blue Shipley 1800	1.5–2.4	<450	1.5×10^5	–	–	–	1000	–
Hoya Corporation								
Custom orders only								
BACTERIORHODOPSIN MATERIALS								
MIB GmbH								
BR-WT B-type	30–100	<650	–	80×10^3	–	–	5000	–
BR-D96N M-type	30–100	<650	30×10^3	–	–	–	5000	–

The developing agent is the most important constituent in a holographic developer. In addition to the developing agent, a developer consists of the following components: a preservative (or antioxidant); a weak silver-solvent agent; an accelerator (or activator); a restrainer; and a solvent which is usually water. Employing a developer which contains a silver-solvent agent, solution-physical development

or semi-physical development can be performed. Such a developer is particularly suitable for processing the ultrafine-grain emulsions.

Holographic Developers

Silver halide materials for holography are all of the fine-grain type. Such materials require the application of special processing techniques to achieve the best possible results. For the processing of amplitude holograms a high-contrast developer is required, such as Kodak D-19. For holograms which are going to be bleached (phase holograms), developers based on ascorbic acid, hydroquinone, metol, pyrocatechol, or pyrogallol are frequently used. The following developers are recommended for the new silver halide materials: D-19; CW-C2; AAC; MAA-3; PAAP; and for pulsed laser recordings, the SM-6 developer. Colloidal development, employing a highly diluted developer which contains silver solvents, is recommended for the ultrafine-grain emulsions. Colloidal silver grains formed in the emulsion during such development are very small which means that very little absorption occurs. The method produces holograms with high diffraction efficiency combined with very low light scattering noise. The Russian GP-2 developer is recommended for colloidal development. Recipes for common holographic developers are found in Table 3.

Bleach Baths

Bleaching is used to convert a developed hologram (containing a silver image) into a phase hologram, to obtain a higher diffraction efficiency than is possible for amplitude holograms. During bleaching the silver image is converted into a water-soluble or an insoluble silver compound. Holographic bleaching techniques can be divided into three categories:

1. Conventional or direct (rehalogenating) bleaching;
2. Fixation-free rehalogenating bleaching;
3. Reversal (complementary) or solvent bleaching.

In conventional bleaching, the developed silver image is, after fixing, converted into a transparent silver halide compound, i.e., after the unexposed silver halide crystals have been removed. This method is not recommended for ultrafine-grain emulsions. Instead the fixation-free rehalogenating bleaching is preferred. The hologram is bleached directly after development, without fixing, leaving the unexposed silver halide crystals in the emulsion. A rehalogenating bleach bath consists of: an oxidizing agent; an alkali halide (often KBr); and a buffer. Among the rehalogenating bleaches, the following bleach baths work well: PBQ-bleach; Ferric EDTA; and the PBU-bleaches (PBU-amidol or the PBU-metol).

In reversal bleaching, the developed silver image is not fixed. Instead it is converted into a soluble silver complex which is removed from the emulsion during bleaching, leaving only the original unexposed silver halide grains in the emulsion. A reversal bleach bath consists of an oxidizing agent and a buffer. For ultrafine-grain silver halide emulsions, the reversal method is not recommended. The very small unexposed silver halide grains are also affected (partly dissolved) during reversal bleaching, resulting in low diffraction efficiency. For materials with larger grains, a more diluted version of the dichromate Kodak R-9 bleach can be

Table 3 Developers for silver halide emulsions

Constituents	*CWC2*	*GP-2*	*MAA-3*	*PAAP*	*AAC*	*SM6*
Methylphenidone		0.2 g				
Catechol	10 g					
Ascorbic acid	5 g		10 g	18 g	16 g	18 g
Hydroquinone		5 g				
Metol			2.5 g			
Phenidone				0.5 g		6 g
Sodium sulfite	5 g	100 g				
Sodium phosphate (diabasic)				28.4 g		28.4 g
Urea	50 g					
Potassium hydroxide		5 g				
Sodium hydroxide				12 g		12 g
Sodium carbonate	30 g		55.6 g		60 g	
Ammonium thiocyanate		12 g				
Distilled water	1 l	1 l	1 l	1 l	1 l	1 l
Dilution:	no	15 ml dev. + 400 ml water	no	no	no	no

Table 4 Bleach baths for silver halide emulsions

Constituents	*Ferric-EDTA (rehal.)*	*PBQ (rehal.)*	*PBU-(amidol/metol) (rehal.)*	*Copper (rehal.)*	*Chrome (reversal)*	*PSSB (reversal)*
Amidol *or* metol*			1 g			
Cupric bromide			1 g			
Potassium persulfate			10 g			20 g
Potassium dichromate					4 g	
p-Benzoquinone		2 g				
Ferric III sulfate	30 g					30 g
Copper sulfate				35 g		
Disodium EDTA	30 g					
Boric acid		1.5 g				
Citric acid			50 g			
Potassium bromide	30 g	30 g	20 g	30 g		
Sodium hydrogen sulfate	30 g			30 g		30 g
Sulfuric acid					4 ml	
Distilled water	1 l	1 l	1 l	1 l	1 l	1 l

*To be added after all other constituents have been dissolved.

used. Recipes for common holographic bleach baths are found in Table 4.

Commercial Silver Halide Emulsions

In Table 2 most of the current silver halide holographic emulsions are listed. The main manufacturer of holographic silver halide emulsions is the Micron branch of the SLAVICH Joint Stock Company photographic company located outside Moscow (2 pl. Mendeleeva, 152140 Pereslavl-Zalessky, Russia, www.slavich.com). Both film (triacetate) and glass plates are manufactured. Many different sizes are produced, including 115 cm by 10 m film rolls. Recommended holographic developers: CW-C2, SM-6. Many rehalogenating holographic bleach baths work well and, in particular, the PBU-type bleaches are recommended.

Colourholographic Ltd is a new company based in England (Braxted Park, Great Braxted, Witham CM8 3BX, England; www.colourholographic.com). The Colourholographic materials are based on the HRT emulsion, previously manufactured in Germany by Dr Birenheide. At the present time only glass plates are manufactured in England. Recommended processing solutions for the new HRT materials are a sodium-sulfite-free ascorbic acid–metol developer followed by a rehalogenating ferric EDTA or PBU bleaches. For reflection holograms, a pyrogallol developer and the ferric-EDTA bleach can be used.

Eastman Kodak's holographic materials are produced in America (343 State Street, Rochester, New York 14650, USA; www.kodak.com/go/PCB products). These plates are usually made to order only. Materials can be ordered with (marked -01) or without (marked -02) anti-halation layer. The 120 emulsion can be used for recording holograms using a pulsed ruby laser. Kodak recommends the developer Kodak D-19 for the processing.

FilmoTec is located in Germany (Chemie Park Bitterfeld-Wolfen, Röntgenstrasse, D-06766 Wolfen, Germany; www.filmotec.de). The company manufactures ORWO holographic emulsions which are coated only on triacetate film, 104 cm wide in 10 and 30 m lengths.

In France, M. Yves Gentet has started a small-scale emulsion manufacturing company: Atelier de Création d'Art en Holographie Yves Gentet (50, rue Dubourdieu, F-33800 Bordeaux, France; www.perso.wanadoo.fr/holographie) Monochromatic and panchromatic ultrafine-grain emulsions are produced called 'Ultimate'. Plates and film up to 60 cm by 80 cm are produced. An ascorbic acid–metol developer is recommended followed by a Ferric EDTA bleach. In addition, the Russian GP-2 developer can be used. Also a special developer can be ordered from the company.

Konica in Japan is manufacturing a green-sensitive emulsion for the use in the American holographic VOXEL 3D medical imaging system and, in addition, for the Japanese market only.

Dichromated Gelatin Materials

Dichromated gelatin (DCG) is an excellent recording material for volume phase holograms and HOEs. The grainless material has its highest sensitivity in the UV region, but extends into the blue and green parts of the spectrum. There are possibilities to sensitize the

emulsion in the red part of the spectrum as well. However, DCG is most often exposed with blue laser wavelengths. Depending on processing parameters, diffraction efficiency and bandwidth can be controlled. It is easy to obtain high diffraction efficiency combined with a large SNR.

During the exposure of a DCG emulsion to UV or blue light, the hexavalent chromium ion (Cr^{6+}) is photo-induced to trivalent chromium ion (Cr^{3+}) which causes cross-linking between neighboring gelatin molecules. The areas exposed to light are hardened and become less soluble than the unexposed areas. Developing consists of a water wash which removes the residual or unreacted chemical compounds. Dehydration of the swollen gelatin follows after the material has been immersed in isopropanol, which causes rapid shrinkage resulting in voids and cracks in the emulsion, thus creating a large refractive index modulation. The underlying mechanism is not completely clear because high modulation can also be caused by the binding of isopropanol molecules to chromium atoms at the cross-linked sites. The DCG material has a rather low sensitivity of about 100 mJ$(cm)^{-2}$. In many cases DCG plates are prepared by fixing an unexposed silver halide emulsion and then sensitizing it in a dichromate solution. It is also, for example, possible to make a 10 μm thick emulsion on a 8″ by 10″ glass plate using 1 g ammonium dichromate mixed with 3 g photographic grade gelatin in 25 ml deionized water. Then the emulsion is spin-coated on the glass substrate or the emulsion is applied by the doctor blade coating technique. Many HOE manufacturers prefer to produce their own emulsions but there are also commercial DCG emulsions available on the market.

Processing of DCG holograms are performed in graded alcohol (propanol) baths starting with water–propanol solutions of high water content and ending with pure propanol (cold or hot) baths. Depending on the wanted HOE characteristics, one has to carefully control the propanol/water mixtures as well as the temperature of the different baths. Cold baths produce better uniformity and lower noise. Warm baths can yield high index modulation but often with increased noise. The bandwidth can be controlled by the processing temperature and the ratio of propanol/water mixtures.

Commercial DCG Materials

Slavich in Russia is one manufacturer of presensitized dichromated plates for holography. The DCG emulsion is marked PFG-04. Plates up to a size of 30 by 40 cm can be ordered.

The FilmoTec DCG emulsions are only available on 190 μm triacetate film, 104 cm wide in 10 and 30 m lengths; emulsion thickness 6 or 20 μm. Note that the film needs to be sensitized in dichromate solution before recording. This means that the company only supplies large-format gelatin-coated film.

Holotec is a German-based company which offers presensitized DCG coated on both plates and film (Jülicher Strasse 191, D-52070 Aachen, Germany; www.holotec.de). The Holotec emulsion is the high-quality DCG emulsions developed by Professor Stojanoff in Aachen. The company can supply large-format DCG glass plates (m^2 size) or film (PET-Polyethylenterephtalat), pre-sensitized and ready to use.

Photopolymer Materials

Photopolymer materials have become popular for recording phase holograms and HOEs, in particular for mass production of holograms since some photopolymer materials only require dry processing techniques.

A photopolymer recording material, such as the DuPont material, consists of three parts: a photopolymerizable monomer; an initiator system (initiates polymerization upon exposure to light); and a polymer (the binder). First, an exposure is made to the information-carrying interference pattern. This exposure polymerizes a part of the monomer. Monomer concentration gradients, formed by variation in the amount of polymerization due to the variation in exposures, give rise to diffusion of monomer molecules from the regions of high concentration to the regions of lower concentration. The material is then exposed to regular light of uniform intensity until the remaining monomer is polymerized. A difference in the refractive index within the material is obtained. The DuPont material requires only a dry processing technique (exposure to UV light and a heat treatment) to obtain a hologram. The DuPont photopolymer materials have a coated film layer thickness of about 20 μm. The photopolymer film is generally coated in a 12.5″ width on a 14″ wide Mylar® polyester base which is .002″ thick. The film is protected with a .00092″ thick Mylar® polyester cover sheet.

The recording of a hologram on DuPont polymer is rather simple. The film has to be laminated to a piece of clean glass or attached to a glass plate using an index-matching liquid. Holograms can be recorded manually, but in order to produce large quantities of holograms, a special machine is required. For hologram replication a laser line scanning technique can provide the highest production rate.

The photopolymer material needs an exposure of about 10 mJ(cm)$^{-2}$.

After the exposure is finished, the film has to be exposed to strong white or UV light. DuPont recommends about 100 mJ(cm)$^{-2}$ exposure at 350–380 nm. After that, the hologram is put in an oven at a temperature of 120 °C for two hours in order to increase the brightness of the image. The process is simple and very suitable for machine processing using, for example, a baking scroll oven.

Commercial Pholymer Materials

Currently the only manufacturer of holographic photopolymer materials is E.I. du Pont de Nemours & Co. (DuPont Holographics, POB 80352, Wilmington, DE 19880, USA; www.dupont.com/holographics). In the past it was possible to obtain photopolymer materials from DuPont, but in 2002 the company changed its policy and today it is producing materials mainly for DuPont Holographics, but approved customers may still be able to buy material from DuPont. The reason for the restrictions is that many applications of mass-produced holograms are in the field of document security, where DuPont's photopolymers are used to produce optical variable devices (OVDs).

Photoresist Materials

Photoresist materials in holography are used for making masters for display and security holograms. The recorded relief image in the photoresist plate is then used to make the nickel shim needed as the tool for embossing holograms into plastic materials.

The exposure to actinic radiation produces changes in the photoresist layer that result in a solvency differentiation as a function of exposure. Processing is done with a suitable solvent dissolving either the unexposed or exposed regions, depending on whether the resist is of the negative or the positive type. For optical recording, positive photoresist (exposed resist removed during development) is preferred to the negative type because of the higher resolving power and low scatter. On the resulting surface, relief pattern particles of nickel are deposited by electrolysis to make a mold, which can then be used as an embossing tool. The photoresist process can be used for making transmission holograms only. If an embossed hologram is mirror-backed by using, for example, an aluminum coating process, it can be used in the reflection reconstruction mode as well. Photoresists are UV and deep-blue sensitive. The material is spincoated on glass substrates to obtain a thickness of between 0.5 and 2 μm, then it is baked at about 75 °C for 15 minutes. A typical photoresist for holography (e.g., Shipley Microposit 1350) has a sensitivity of about 10 mJ(cm)$^{-2}$. The most common laser wavelengths for recording photoresist holograms are the following wavelengths: 413 nm (Krypton-ion), 442 nm (Helium–Cadmium), and 458 nm (Argon-ion). At higher wavelengths, e.g., 488 nm, the sensitivity is rather low. Development of Shipley resist is performed in Shipley 303A developer, diluted 1 part developer + 5 parts de-ionized water. Normally only a 10 second development time is needed (under agitation). After that the plate has to be rinsed in de-ionized water for 2 minutes. The plate needs to be dried quickly with a high pressure steam of compressed nitrogen or dry, filtered compressed air.

Commercial Resist Materials

The main American manufacturer of photoresist is Shipley Co. (1457 Macarthur Rd., #101, Whitehall, PA 18052, USA). There are a few companies which make plates based on Shipley resist.

Towne Technologies, Inc. (6–10 Bell Ave., Sommerville, NJ 08876, USA; www.townetech.com) produces spin-coated plates using Shipley resist (Shipley S-1800 series). Towne plates have a sublayer of iron oxide which enhances adhesion of the resist during electroplating operations. In addition, the sublayer eliminates unwanted backscatter from the glass substrate. The plates are developed in Shipley 303A developer.

Hoya Corporation (7-5 Naka-Ochiai 2-chome, Shinjuku-ku, Tokyo, Japan; www.hoya.co.jp) is a producer of mask blanks coated with 0.5 μm thick Shipley S-1800 resist (SLW plates). Different types of glass as well as quartz plates are produced. Hoya supply plates for holography, however, custom orders only.

Thermoplastic Materials

The holographic thermoplastic is a multilayer structure coated on glass or film. The substrate is first coated with a conducting layer, e.g., evaporated gold, then a photoconductor, e.g., poly-n-vinyl carbazole, PVK, that has been sensitized with, e.g., 2,4,7-trinitro-9-fluorenone (TNF), and on top of this layer a thermoplastic coating is deposited (usually a styrene-methacrylate material). The recording of a hologram starts with a uniform charging of the surface of the thermoplastic material using a corona charger. The charge is divided between the photoconductor and the thermoplastic layer. Exposure and the consequent photogeneration in the photoconductor cause charges of opposite signs to migrate to the interface with the thermoplastic layer and

the substrate. This will not change the charge, but only decrease the surface potential. Before the image can be developed, the material has to be recharged with the uniform corona charger. This step adds additional charges to the exposed areas, in proportion to the reduced potential which is also proportional to the exposure. The material is then heated to a softening temperature of the thermoplastic layer which will be deformed due to the electrostatic forces acting on it. The material is then cooled and the image is fixed as a relief surface pattern in the thermoplastic layer. The processing time after the exposure necessary to obtain the image is between ten seconds and half a minute. If the material is heated to a somewhat higher temperature, the pattern disappears and the image is erased. The thermoplastic material can now be used for the recording of another hologram, repeating the recording procedure. Such a cycling technique can be repeated hundreds of times, without any serious effect on the quality of the image. That, and the fast dry processing made such materials popular for the use in holographic cameras for nondestructive testing. The sensitivity is between 10 and 100 $\mu J(cm)^{-2}$ over the whole visible electromagnetic spectrum.

Commercial Thermoplastic Materials

In the past, several companies produced thermoplastic materials and recording equipment for holography. Since there is very little demand for such materials today, there is only one company which can deliver thermoplastic equipment for holography: Tavex America Inc. (14 Garden Road, Natick, MA 01760, USA; www.tavexamerica.com). Tavex America manufactures a thermoplastic camera, TCC-2 in which 40 mm by 40 mm plates are used. For obsolete systems, thermoplastic film can still be obtained from MD Diffusion in France (93 rue d'Adelshoffen, F67300 Schiltigheim, France).

Bacteriorhodopsin

Photochromic materials, a class to which bacteriorhodopsin (BR) belongs, have not been used much in holography. Photochromics are real-time recyclable materials which need no processing for development. They can easily be erased and reused over and over again. In holography the only successful material has been the BR film which was introduced only a few years ago. BR is a biomolecule with several applications in photonics, comprising a light-driven molecular proton pump with exceptional photoelectric and photochromic properties. BR is a photochromic protein, found in the photosynthetic system of a salt-marsh bacterium, termed purple membrane (PM), of the cell membrane of *Halobacterium salinarum*. BR thin films are made from polymer suspensions of the isolated purple membrane. Some of the desirable properties of BR films are real-time writing and erasing in microsecond time-scales, durability allowing millions of write/erase cycles without degradation, very high spatial resolution of 5000 $lp(mm)^{-1}$, reasonably good light sensitivity of approximately 10 $mJ(cm)^{-2}$, the encoding of both amplitude and phase patterns, and the use of visible laser light for writing and reading. The composition of the films may also be altered by chemical means to control the optical characteristics, and the BR molecules themselves may be mutagenically changed for performance optimization. Particular properties that can be enhanced in these ways are image lifetimes and the diffraction efficiency of holograms recorded in the films. One advantage of this material is that it is reversible. Therefore it has been considered to be used in holographic storage systems and for real-time holographic interferometry applications. BR film can be used for recording polarization holograms because of BR films inherent photo-inducable optical anisotropy. The material is also possible to expose with pulsed lasers, which makes it suitable for fast holographic recording.

The BR film works in the following way. When illuminated with green light, the BR molecule undergoes an absorption resulting in a maximum shift of about 160 nm towards the red and its color changes from purple to yellow. The yellow image can be erased by illuminating the BR material with blue light. The recording–erasing cycle can be repeated about a million times. Suitable BR-films for holography are obtained by embedding purple membranes into inert matrices like polyvinylalcohol or polyacrylamide. For hologram recording and erasure two photochemical conversions $B \rightarrow M$ (B-type holograms) and $M \rightarrow B$ (M-type holograms) can be employed. After excitation by light, BR cycles through a sequence of spectrally distinguishable intermediates and returns to the initial state. The time associated with this thermal relaxation process depends on several aspects, for example, type of BR (wildtype or variant), pH value of film-matrix, film-temperature, etc. For example, the M intermediate is populated from the B state (absorption spectrum peaked at 570 nm) by a green pumping beam and information is recorded with blue light, which initiates the photochemical $M \rightarrow B$ transition. The M state has its absorption peak at 410 nm. In BR variants in which the thermal relaxation of the M state has been prolonged, photocontrolled switching between these two states can be achieved. Only when

the pumping beam is incident upon the BR film can information be recorded with blue light. The pumping beam can be used as a readout beam for the recorded hologram.

Commercial BR Materials

A holographic camera based on a BR film is manufactured in Germany. The holographic system, FringeMaker-*plus*, developed at the University of Marburg is marketed by Munich Innovation Biomaterials (MIB) GmbH. (Hans-Meerwein-Str., D-35032 Marburg, Germany; www.mib-biotech.de). MIB is also a manufacturer of both wildtype and genetically modified BR film for holography and other scientific/technical applications. BR-films offered by MIB can be used in two different recording configurations: the transition from the initial B to M-state (induced with yellow light) serves as basis for optical data recording and processing (B-type recording); or the photochemically induced transition from M to B state (M-type recording). In the latter case, yellow light is used to control the population of the M-state and blue light is used to write the information. MIB offers BR-films with normal and slow thermal relaxation characteristics. Typical relaxation time ranges (RTR), after which the photochemically excited BR molecules have returned to the original state, are between 0.3 and 80 s. The BR film is sealed between two windows of high-quality optical glass. The BR film layer has a thickness of 30–100 μm, dependent on the optical density of the film. The diameter of the film device is 25 mm and the clear aperture is 19 mm.

Photorefractive Crystal Materials

Photorefractive crystals are electro-optic materials which are suitable for recording volume phase holograms in real-time. These crystals are important components in holographic storage systems and are seldom used in other holographic image applications. Holograms recorded in crystals, such as: lithium niobate ($LiNbO_3$); barium titanate ($BaTiO_3$); potassium tantalate niobate (KTN); barium sodium neobate (SBN); bismuth silicon oxide (BSO); and bismuth germanium oxide (BGO), consist of bulk space charge patterns. An interference pattern acting upon a crystal will generate a pattern of electronic charge carriers that are free to move. Carriers are moving to areas of low optical illumination where they get trapped. This effect will form patterns of net space charge, creating an electrical field pattern. As these crystals have strong electro-optic properties, the electrical field pattern will create a corresponding refractive index variation pattern, which means a phase hologram. These holograms can be immediately reconstructed with a laser beam differing from the beams creating the hologram. A hologram can also be erased and a new hologram can be recorded in the crystal. Holograms can be fixed by converting the charge patterns to the patterns of ions which are not sensitive to light. In some materials the hologram can be fixed by heating the crystal to above 100 °C based on thermally activated iconic conductivity. The resulting iconic is frozen upon cooling the crystal back to room temperature. More details about these materials and storage applications are found in the Further Reading.

See also

Holography, Techniques: Overview.

Further Reading

Bjelkhagen HI (1993) *Silver Halide Recording Materials for Holography and Their Processing*. Springer Series in Optical Sciences, vol. 66. Heidelberg: Springer-Verlag.

Bjelkhagen HI (ed.) (1996) *Selected Papers on Holographic Recording Materials*. SPIE Milestone Series, vol. MS130. Bellingham, WA, USA.

Lessard RA (ed.) (1995) *Selected Papers on Photopolymers, Physics, Chemistry, and Applications*. SPIE Milestone Series, vol. MS114. Bellingham, WA, USA.

Sincerbox GT (ed.) (1994) *Selected Papers on Holographic Storage*. SPIE Milestone Series, vol. MS95. Bellingham, WA, USA.

Smith HM (ed.) (1977) *Holographic Recording Materials*, Springer Topics in Applied Physics, vol. 20. Berlin: Springer-Verlag.

HOLOGRAPHY, TECHNIQUES

Contents

Overview
Color Holography
Computer-Generated Holograms
Digital Holography
Holographic Interferometry
Sandwich Holography and Light in Flight

Overview

C Shakher and A K Ghatak, Indian Institute of Technology, New Delhi, India

© 2005, Elsevier Ltd. All Rights Reserved.

Major Milestones

1948:	Essential concept for holographic recording by Dennis Gabor.
1960:	The first successful operation of a laser device by Theodore Maiman.
1962:	Off-axis technique of holography by Leith and Upatnieks.
1962:	Yu N Denisyuk suggested the idea of three-dimensional holograms based on thick photoemulsion layers. His holograms can be reconstructed in ordinary sunlight. These holograms are called Lippmann–Bragg holograms.
1964:	Leith and Upatnieks pointed out that a multicolor image can be produced by a hologram recorded with three suitably chosen wavelengths.
1969:	S A Benton invented 'Rainbow Holography' for display of holograms in white light. This was a vital step to make holography suitable for display applications.

Holography is the science of recording an entire optical wavefront, both amplitude and phase information, on appropriate recording material. The record is called a hologram. Unlike conventional photography, which records a three-dimensional scene in a two-dimensional format, holography records true three-dimensional information about the scene.

Holography was invented by Gabor in 1948 and his first paper introduced the essential concept for holographic recording – the reference beam. Holography is based on the interference between waves and, it provides us with a way of storing all the light information arriving at the film in such a way that it can be regenerated later.

The use of the reference beam is utilized because the physical detectors and recorders are sensitive only to light intensity. The phase is not recorded but is manifest only when two coherent waves of the same frequency are simultaneously present at the same location. In that case, the waves combine to form a single wave whose intensity depends not only on intensities of the two individual waves, but also on the phase difference between them. This is key to holography. The film record, or hologram, can be considered as a complicated diffraction grating. Holograms bear no resemblance to conventional photographs in that an image is not actually recorded. In fact, the interferometric fringes which are recorded on the recording material are not visible to an unaided eye because of extremely fine interfringe spacing (~ 0.5 micrometer). The fringes which are visible on the recording material are the result of dust particles in the optical system used to produce the hologram.

Gabor's original technique is now known as in-line holography. In this arrangement, the coherent light source as well as the object, which is a transparency containing small opaque details on a clear background, is located along the axis normal to the photographic plate. With such a system, an observer focusing on one image observes it superposed on the out-of-focus twin image as well as a strong coherent background. This constitutes the most serious problem of Gabor's original technique.

The first successful technique for separating the twin images was developed by Leith and Upatnieks in 1962. This is called the off-axis or side band technique of holography. We shall consider mainly the off-axis technique here.

Basic Holography Principle

Light from a laser or any light beam is characterized by spatial coherence (l_c) and temporal coherence (τ_c), which is discussed in all textbooks on optics. For typical laboratory hologram recording, the required degree of coherence of laser radiation is determined by the type of object and geometrical arrangement being used for the recording. The following condition must hold:

$$L \ll c\tau_c \qquad [1]$$

where τ_c is the coherence time, and L the maximum path length difference between the two waves chosen to record the hologram. To begin recording, two wavefronts are derived from the same laser source. One wavefront is used to illuminate the object and another is used as a reference wavefront. The wavefront derived by illuminating the object is superimposed with the reference wave and the interference pattern is recorded. These object and reference waves must:

- be coherent (derived from the same laser); and
- have a fixed relative phase at each point on the recording medium.

If the above conditions are not met, the fringes will move during the exposure and the holographic record will be smeared. Therefore, to record the hologram one should avoid air currents and vibrations. The recording medium must have sufficient resolution to record the fine details in the fringe pattern. This will be typically of the order of the wavelength. To create a three-dimensional image from the holographic process one then has to record and reconstruct the hologram. The object and reference waves are derived from the same laser.

Holograms record an interference pattern formed by interference of object and reference wavefronts, as explained above. We may mention here that for the two plane waves propagating at an angle θ between them, the spacing of the interference fringes is given by

$$\alpha = \frac{\lambda_0}{2\sin(\theta/2)} \qquad [2]$$

where λ_0 is the free space wavelength.

Processing of the hologram (developing, fixing, and washing of the recording material) yields a plate with alternating transparent and opaque parts, variation of refractive index, or variation of height corresponding to intensity variation in the fringe pattern. Such a plate can be regarded as a complicated diffraction grating. In this process, the hologram is illuminated with monochromatic, coherent light. The hologram diffracts this light into wavefronts which are essentially indistinguishable from the original waves which were diffracted from the object. These diffracted waves produce all the optical phenomena that can be produced by the original waves. They can be collected by a lens and brought into focus, thereby forming an image of the original object, even though the object has since been removed. If the reconstructed waves are intercepted by the eye of an observer, the effect is exactly as if the original waves were being observed; the observer sees what appears to be the original object in true three-dimensional form. As the observer changes his viewing position, the perspective of the image scene changes; parallactic effects are evident and the observer must refocus when the observation point is changed from a near to a distant object in the scene. Assuming that both the construction and reconstruction of the hologram are made with the same monochromatic light source, there is no visual test which can be made to distinguish between the real object and the reconstructed image of the object. It is as if the hologram were a window through which the apparent object is viewed.

Hologram of a Point Object

Consider a hologram recorded with a collimated reference wave normal to the recording plate and a point object inclined at a certain angle. If the hologram is illuminated once again with the same collimated reference wave, it reconstructs two images, one virtual true image and the other real image. However, the two images differ in one very important respect.

While the virtual image is located in the same position as the object and exhibits the same parallax properties, the real image is formed at the same distance from the hologram but in front of it. Corresponding points on the real and virtual images are located at equal distances from the plane of the hologram; the real image has the curious property that its depth is inverted. Such an image is not formed with a normal optical system; it is therefore called a pseudo image as opposed to a normal or orthoscopic image.

This depth inversion results in conflicting visual clues, which make viewing of the real image

psychologically unsatisfactory. Thus, if O_1 and O_2 are two elements in the object field, and if O_1 blocks the light scattered by O_2 at a certain angle, the hologram records information only on the element O_1 at this angle and records no information about this part of O_2. An observer viewing the real image from the corresponding direction then cannot see this part of O_2, which, contrary to normal experience, is obscured by O_1, even though O_2 is in front of O_1.

Production of an Orthoscopic Real Image

An orthoscopic real image of an object can be produced by recording two holograms in succession. In the first step, a hologram is recorded of the object with a collimated reference beam. When this hologram is illuminated once again with the collimated reference beam that was used to record it, it reconstructs two images of the object at unit magnification, one of them being an orthoscopic virtual image, while the other is a pseudoscopic real image. A second hologram is then recorded of this real image with a second collimated reference beam.

When the second hologram is illuminated with a collimated beam it reconstructs a pseudoscopic virtual image located in the same position as the real image formed by the second hologram is an orthoscopic image. Since a collimated reference beam is used throughout, the final real image is the same size as the original object and free from aberrations.

In addition to these characteristic linked intensities with the three-dimensional nature of the reconstruction, the holographic recording has several other properties. Each portion of the hologram can reproduce the entire image scene. If a smaller and smaller portion of the hologram is used for reconstruction, there is loss of image intensity and reconstruction. When a hologram is reversed, such as in contact printing processes, it will still reconstruct a positive image indistinguishable from the image produced by the original hologram.

Simple Mathematical Description of Holography

Let the object and reference waves be given by

$$U_o = O(x, y)e^{i\phi(x,y)} \qquad [3]$$

and

$$U_r = R\, e^{iky\sin\theta} \qquad [4]$$

U_r describes the reference (plane wave) propagating at angle θ to the z-axis.

The intensity at the hologram plane is

$$\begin{aligned} I &= |U_r + U_o|^2 \\ &= |U_r|^2 + |U_o|^2 + U_r^* U_o + U_r U_o^* \end{aligned} \qquad [5]$$

As can be seen from the above equation, the amplitude and phase of the wave are encoded as the amplitude and phase modulation of a set of interference fringes. The material used to record the patterns of interference fringes described above is assumed to provide linear mapping of the intensity incident during the reconstruction process into amplitude transmitted by or reflected from the recorded material. Usually both light detection and wavefront modulation are performed by photographic plate/film. We assume the amplitude transmission properties of the plate/film after processing to be described by

$$T = T_o - b(E - E_o) \qquad [6]$$

where the exposure E, at the film is $E = I\tau$; here τ is the exposure time. T_0 is the transmittance of the unexposed plate.

If the hologram is reilluminated by the reference wave, the transmitted wave amplitude will be

$$U_t = U_r(T_o - b\tau(|U_o|^2 + U_r^* U_o + U_r U_o^*)) \qquad [7]$$

The first term is reference wave times a constant. The second term is the reference wave modulated by $|U_o|^2 = O(x, y)^2$; it gives small-angle scattering about the reference wave direction. The third term is proportional to U_o. It is the same as the original object wave (note this is only so if the reconstruction wave is identical to the reference wave). The fourth term is

$$-b\tau U_r^2 U_0^* = -b\tau R^2\, e^{iky2\sin\theta} O(x, y)e^{-i\phi(x,y)} \qquad [8]$$

This is essentially a wave traveling in a direction $\sin^{-1}(2\sin\theta)$ to the z-axis, with the correct object amplitude modulation but its phase reversed in sign, producing a conjugate wave.

Types of Holograms

Primarily, the holograms are classified as thin and thick, based on the thickness of the recording medium. When the thickness of the recording medium is small compared with the average spacing of the interference fringes, the hologram can be treated as a thin hologram. Such holograms are characterized by

spatially varying complex amplitude transmittance:

$$\mathbf{t}(x, y) = [\mathbf{t}(x, y)]\exp[-i\phi(x, y)] \qquad [9]$$

Thin holograms can be further categorized as thin amplitude holograms or thin phase holograms. If amplitude transmittance of the hologram is such that $\phi(x, y)$ is constant while $\mathbf{t}(x, y)$ varies over the hologram, the hologram is termed an amplitude hologram. For a lossless phase hologram, $|\mathbf{t}(x, y)| = 1$, so that the complex amplitude transmittance is caused by variation in phase.

When the thickness of the hologram recording material is large compared to the fringe spacing of the interference fringes, the holograms may be considered as volume holograms. These may be treated as a three-dimensional system of layers corresponding to a periodic variation of absorption coefficient or refractive index, and the diffracted amplitude is at a maximum when Bragg's condition is satisfied. In general, the behavior of thin and thick holograms corresponds to Raman Nath and Bragg diffraction regimes. The distinctions between two regimes is made on the basis of a parameter Q, which is defined by the relation:

$$Q = \frac{2\pi\lambda_o d}{n_o \Lambda^2} \qquad [10]$$

where

Λ = spacing of fringe on hologram, measured normal to the surface;
d = thickness of recording medium;
n_o = mean refractive index of the medium; and
λ_o = wavelength of light.

Small values of Q ($Q < 1$) correspond to thin gratings, while large values of Q ($Q > 10$) correspond to volume gratings. For values of Q between 1 and 10, intermediate properties are exhibited. However, this criterion is not always adequate. The boundaries between the thin and thick holograms are given by the value of

$$P = \frac{\lambda_o^2}{\Lambda^2 n_o n_1} \qquad [11]$$

where P^{-2} is the relative power diffracted into higher orders, and $P < 1$ for thin holograms and $P > 10$ for thick holograms. For P having values between 1 and 10, the hologram may be thin or thick, depending on the other parameters involved.

Holograms can also be classified based on whether they are reconstructed in transmitted light or reflected light. For transmission holograms, during the recording stage, two interfering wavefronts make equal but opposite angles to the surfaces of recording medium and are incident on it from the same side. Reflection holograms reconstruct images in reflected light. They are recorded such that the interfering wavefronts are symmetrical with respect to the surface of the recording medium but are incident on it from opposite sides. When the angle between the interfering wavefronts is maximum (180°), the spacing between the fringe planes of the recording medium is minimum. Under such conditions, reflection holograms may have wavelength sensitivity high enough to be reconstructed, even with white light.

An important development in the field of holography was the invention of rainbow holograms by Benton in 1969. This provides a method for utilizing white light for illumination when viewing the holograms. The technique does so by minimizing the blur introduced by color dispersion in transmission holograms, at the price of giving up parallax information in one dimension.

Recording Materials

An ideal recording medium for holography should have a well matching spectral sensitivity corresponding to available laser wavelengths, linear transfer characteristics, high resolution, and low noise. It should also be either inexpensive or indefinitely recyclable. Toward achieving the above-mentioned properties, several materials have been studied, but none has been found so far that meets all the requirements. Materials investigated include: silver halide photographic emulsion; dichromated gelatin plates/films; silver halide sensitized gelatin plates/films; photo-resists; photo polymer systems; photochromics; photo thermoplastics; and ferro-electric crystals. Recently, the use of storage and processing capabilities of computers, together with CCD cameras, has been used for recording holograms.

Silver halide photographic emulsions are a commonly used recording material for holography, mainly because of relatively high sensitivity and easy availability. Manufacture, use, and processing of these emulsions have been well standardized. The need for wet processing and drying may constitute a major drawback. Dichromated gelatin can be considered an almost ideal recording material for volume phase holograms, as it has a large refractive index modulation capability, high resolution, low absorption and scattering.

Because of these features, dichromated gelatin has been extensively investigated. The most significant disadvantage is its comparatively low sensitivity. Pennington *et al.* developed an alternative technique, which combines the advantage of silver halide emulsions (high sensitivity) with that of dichromated

gelatin (low absorption scattering and high stability). It involves exposing silver halide photographic emulsion and then processing it, so as to obtain a volume phase hologram consisting solely of hardened gelatin.

Thin phase holograms can be recorded on photoresists, which are light-sensitive organic films yielding relief image after exposure and development. They offer advantages of easy replication using thermoplastics but are slow in response and undergo nonlinear effects at diffraction efficiency greater than ~0.05. Shipley AZ-1350 is a widely used photoresist, with maximum sensitivity in the ultraviolet, dropping rapidly for longer wavelengths towards the blue.

Photopolymers are being keenly investigated because they offer advantages such as ease of handling, low cost, and real-time recording for the application of holography and nonlinear optics. However, they have low sensitivity and short shelf life.

Thin phase surface relief holograms can be recorded in a thin layer of thermoplastics. These materials have a reasonably high sensitivity over the whole visible spectrum, fairly high diffraction efficiency, and do not require wet processing. Their application in holographic interferometry, optical information processing, and in making compact holographic devices has been widely reported.

Photorefractive crystals such as $Bi_{12}SiO_{20}$, $LiNbO_3$, $Bi_{12}GeO_{20}$, $BaTiO_3$, $LiTaO_3$, etc., as recording materials, offer high-angular sensitivity and provide capability to read and write volume holographic data in real time. Besides the materials discussed here, several other materials have been investigated for recording holograms; these include photocromics, elastomere devices, magneto-optic materials, etc.

Application of Holography

Holography can be constructed not only with the light waves of lasers, but also with sound waves, microwaves, and other waves in the electromagnetic spectrum of radiation. Holograms made with ultraviolet light or X-rays can record images of objects/particles smaller than the wavelength of visible light, e.g., atoms or molecules. Acoustical holography uses sound waves to see through solid objects. Holography has a vast scope of practical applications, which have been classified into two major categories:

1. applications requiring three-dimensional images for visual perception; and
2. applications in which holography is used as a measuring tool.

The unique ability of holography – to record and reconstruct both electromagnetic and sound waves – makes it a valuable tool for education, science, industry, and business. Below are some of the important applications:

1. Holographic interferometry (HI) is one of the most powerful measuring tools. In HI, two states of an object, i.e., initial and deformed state are recorded on the same photographic plate. After reconstruction of the light wave corresponding to two states of an object, interferences and deformations are displayed in terms of the interference pattern. The change of distance of one tenth of a micron, or lower, can be resolved. HI provides scientists/engineers with crucial data for design of critical machine parts of power-generating equipment, in the aircraft industry, automobile industry, and nuclear installations (say, for example, in the design of containers used to transport nuclear materials, improve the design of aircraft wings and turbine blades, etc.). Presently, HI is being widely used in mechanical engineering, acoustics, and aerodynamics, for nondestructive testing, to investigate oscillation in diaphragms and flow around various objects, respectively.
2. Microwave holography can detect objects deep within spaces, by recording the radio waves they emit.
3. Another important application of holography is the design of optical elements, which possess special properties. A holographic recording of a concave mirror behaves in much the same way as the mirror itself, i.e., it can focus the light. In some cases chromatism can be introduced in the design of elements so that a location of the point, where the beams are focused, depends on the wavelength. This can be achieved by accurately choosing the recording arrangement of the focusing elements, these elements are, in fact, diffraction gratings. These can have low noise levels, freedom from astigmatism, and have other useful properties. Holographic optical elements have found applications in supermarket scanners to read barcodes, headup displays in fighter aircraft to observe critical cockpit instruments, etc.
4. A telephone credit card used in Europe and other developed countries has embossed surface holograms which carry a monetary value. When the card is inserted into the telephone, a card reader discerns the amount due and deducts the appropriate amount to cover the cost of the call.

5. Holography is having applications in analog and digital computers, offering remarkable opportunities to realize various logical operators, devices for identifying images based on matched filtering, and in computer memory units. The basic advantage of holographic memory is that a relatively large volume of information can be stored and that there are a limited number of ways to change the record. The arrival of the first prototype of optical computers, which use holograms as storage material for data, could have a dramatic impact on the holographic market. The yet-to-be-unveiled optical computers will be able to deliver trillions of bits of information faster than the current generation of computers.
6. Optical tweezers are going to become an important tool for the study of micro-organisms/bacteria, etc.
7. Holograms can be used to locate and retrieve information without knowing its location in the storage medium, only needing to know some of its content.
8. The links between computer science and holography are now well-established and are developing, with at least two aspects making computer-generated holograms extremely interesting. First, such holograms enable us to obtain visual 3-dimensional reconstructions of imagined objects. For example, one can reconstruct three dimensions of a model of an object still in the design stage. Second, computer-generated holograms can be used to reconstruct lightwaves with specified wave fronts. This means specially computed and manufactured holograms may function as optical elements that transform the incident light wave into desired wavefronts.
9. Another important applications of holography is its utilization to compensate for the distortion that occurs when viewing objects through optically heterogeneous mediums. It can be achieved based on the principle of beam reversal by the hologram.
10. Finally, let us consider the application of holography to art. The development of holography gives very effective ways of creating qualitative three-dimensional images. Thus, a new independent area of holographic creative work representational/artistic holography has appeared. The art of holographic depiction has developed along two major routes. The first is creation of the view hologram, used as holograms of natural objects that are to be displayed in exhibitions and museums; these are also known as artistic holograms. Portrait holography is also classified under this category. The progress in portrait holography is hampered partly because of imperfection of pulsed lasers and partly because of the deterioration of photographic material when exposed to pulsed electromagnetic radiation.

    The principle behind the creation of elusion by using composite holograms is also very convincing for the display of objects. To synthesize the composite holograms, the photographs of various aspects of a scene are printed onto the photographic plate. The synthesis techniques for preparing the composite hologram are very complicated, while the images created by holograms are still far from perfect. However, there is no doubt that composite holography opens holography up as an artistic technique. Recently rainbow holography has been very popular to display such objects.

See also

Holography, Applications: Art Holography. **Holography, Techniques:** Color Holography; Computer-Generated Holograms; Digital Holography; Holographic Interferometry. **Phase Control:** Phase Conjugation and Image Correction.

Further Reading

Benton SA (1969) On a method for reducing the information content of holograms. *Journal of the Optical Society of America* 59: 1545.

Denisyuk YuN (1962) Photographic reconstruction of the optical properties of an object in its own scattered radiation field. *Sov. Phys.-Dokl.* 7: 543.

Denisyuk YuN (1963) On the reproduction of the optical properties of an object by the wave field of its scattered radiation, Pt. I. *Optical Spectroscopy* 15: 279.

Denisyuk YuN (1965) On the reproduction of the optical properties of an object by the wave field of its scattered radiation, Pt II. *Optical Spectroscopy* 18: 152.

Gabor SD (1948) A new microscopic principle. *Nature* 161: 777.

Gabor D (1949) Microscopy by reconstructed wavefronts. *Proceedings of the Royal Society* A197: 454.

Gabor D (1951) Microscopy by reconstructed wavefronts: II. *Proceedings of the Physical Society* B64: 449.

Hariharan P (1983) *Optical Holography: Principle, Techniques and Applications*. Cambridge, UK: Cambridge University Press.

Leith EN and Upatnieks J (1962) Reconstructed wavefronts and communication theory. *Journal Optical of the Society of America* 52: 1123.

Leith EN and Upatnieks J (1963) Wavefront reconstruction with continuous-tone objects. *Journal of the Optical Society of America* 53: 1377.
Leith EN and Upatnieks J (1964) Wavefront reconstruction with diffused illumination and three-dimensional objects. *Journal of the Optical Society of America* 54: 1295.
Pennington KS and Lin LH (1965) Multicolor wavefront reconstruction. *Applied Physics Letters* 7: 56.
Stroke GW and Labeyrie AE (1966) White-light reconstruction of holographic images using the Lippmann-Bragg diffraction effect. *Physics Letters* 20: 368.

Color Holography

H I Bjelkhagen, De Montfort University, Leicester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Since the appearance of the first laser-recorded monochromatic holograms in the 1960s, the possibilities of recording full-color high-quality holograms have now become a reality. Over many years holographic techniques have been created to produce holograms with different colors, referred to as pseudo- or multicolor holograms. For example, in the field of display holography, it is common to make multiple-exposed reflection holograms using a single-wavelength laser and emulsion thickness manipulation to obtain an image with different colors. In this way many beautiful pseudocolor holograms have been made by artists. Applying Benton's rainbow hologram technique (for example, the one used for embossing holograms on credit cards), it has been possible to record transmission holograms as well as mass-produce embossed holograms with 'natural' colors. However, a rainbow hologram has a very limited viewing position from which a correct color image can be seen.

This article will acquaint the reader with the topic of color holography (full-color or true-color holography). This describes a holographic technique to obtain a color 3D image of an object where the color rendition is as close as possible to the color of the real object. A brief review of the development of color holography, the current status and the prospects of this new 3D imaging technique, are presented. Also included are recording techniques using three lasers providing red, green and blue (RGB) laser light as well as computer-generating techniques for producing full parallax 3D color images.

The Development of Color Holography

In theory, the first methods for recording color holograms were established in the early 1960s. Already Leith and Upatnieks (1964) proposed multicolor wavefront reconstruction in one of their first papers on holography. The early methods concerned mainly transmission holograms recorded with three different wavelengths from a laser or lasers, combined with different reference directions to avoid cross-talk. Such a color hologram was then reconstructed (displayed) by using the original laser wavelengths from the corresponding reference directions. However, the complicated and expensive reconstruction setup prevented this technique from becoming popular. More suitable for holographic color recording is reflection holography, which can be reconstructed and viewed in ordinary white light.

Most likely, the three-beam transmission technique might eventually become equally applicable, provided inexpensive multicolor semiconductor lasers become widely available. Such RGB lasers can be used to display a very realistic, deep, and sharp holographic image of, for example, a room.

Relatively few improvements in color holography were made during the 1960s and 1970s. Not until the end of the 1970s did a few examples of color holograms appear. During the 1980s, new and improved techniques were introduced. A review of various transmission and reflection techniques for color holography was published by Hariharan (1983). From 1990 until today, many high-quality color holograms have been recorded, mainly due to the introduction of new and improved panchromatic recording materials. On the market are ultra-fine-grain silver halide emulsions (manufactured by Slavich in Russia) as well as photopolymer materials (manufactured by E. I. du Pont de Nemours & Co. in the USA).

Color reflection holography presents no problems with regard to the geometry of the recording setup, but the final result is highly dependent on the recording material used and the processing techniques applied. There are some problems associated with recording color holograms in general and specifically in silver halide emulsions:

- Scattering occurring in the blue part of the spectrum typical of many holographic coarse-grain silver halide emulsions makes them unsuitable for the recording of color holograms.

- Multiple storage of interference patterns in a single emulsion reduces the diffraction efficiency of each individual recording. The diffraction efficiency of a three-color recording in a single-layer emulsion is lower than a single wavelength recording in the same emulsion.
- During wet processing, emulsion shrinkage frequently occurs, causing a wavelength shift. White-light-illuminated reflection holograms normally show an increased bandwidth upon reconstruction, thus affecting the color rendition.
- The fourth problem, related to some extent to the recording material itself, is the selection of appropriate laser wavelengths and their combination in order to obtain the best possible color rendition of the object.

In the beginning, when no suitable panchromatic emulsion existed, the sandwich technique was used to make color reflection holograms. Two plates were sandwiched together, in which, for example, two different types of recording materials were used. The most successful demonstration of the sandwich recording technique was made by Kubota (1986) in Japan. He used a dichromated gelatin (DCG) plate for the green (515 nm) and the blue (488 nm) components, and an Agfa 8E75 silver halide plate for the red (633 nm) component of the image. Kubota's sandwich color hologram of a Japanese doll recorded in 1986, demonstrated the potential of high-quality color holography for the first time. Color holograms have also been recorded in red-sensitized DCG materials.

Extensive work in the field of reflection color holography was performed by Hubel and Solymar (1991). Primarily they employed Ilford silver halide materials for the recording of color holograms applying the sandwich technique.

Very important is the research on recording materials by Usanov and Shevtsov (1990) in Russia. Their work is based on the formation of a microcavity structure in gelatin, by applying a special processing technique to silver halide emulsions. Such holograms recorded in a silver halide emulsion have a high diffraction efficiency and exhibit a typical DCG hologram quality.

Not until panchromatic ultra-fine-grain silver halide emulsions were introduced in Russia in the early 1990s, was it possible to record high-quality color holograms in a single emulsion layer as demonstrated by the present author. Although it is now possible to produce color holograms, they need to be displayed correctly. An important improvement would be to find an edge-illuminating technique to make color holograms easier to display. The display problem still remains the main obstacle preventing color holograms, as well as monochrome holograms, being more widely used.

Recording of Color Holograms

Currently, most transmission color holograms are of the rainbow type. Large-format holographic stereograms made from color-separated movie or video recordings have been produced. There are also some examples of embossed holograms in which a correct color image is reproduced. In order to generate a true color rainbow hologram, a special setup is required in which the direction of the reference beam can be changed in between the recordings of the color-separated RGB primary images. However, as already mentioned, holographic color images of the rainbow type can reconstruct a correct color image only along a horizontal line in front of the film or plate and, thus, of less interest for serious color holography applications. Reflection holography can offer full parallax, large field of view 3D color images where the colors do not change when observed from different directions.

Silver Halide Materials

To be able to record high-quality color reflection holograms it is necessary to use extremely low light-scattering recording materials. This means, for example, the use of ultra-fine-grain silver halide emulsions (grain size about 10 nm). Currently, the only producer of a commercial holographic panchromatic ultra-fine-grain silver halide material is the Micron branch of the Slavich photographic company located outside Moscow. The PFG-03c emulsion comes in standard sizes from 63 mm × 63 mm format up to 300 mm × 406 mm glass plates and is also available on film. Some characteristics of the Slavich material are presented in Table 1.

By using suitable processing chemistry for the PFG-03c emulsion it has been possible to obtain high-quality color holograms. These holograms are recorded in a single-layer emulsion, which greatly simplifies the recording process as compared with many earlier techniques. The aforesaid single-layer technique is described below.

Table 1 Characteristics of the Slavich panchromatic emulsion

Silver halide material	PFG-03C
Emulsion thickness	7 μm
Grain size	12–20 nm
Resolution	~10,000 lp(mm)$^{-1}$
Blue sensitivity	~1.0–1.5 × 10^{-3} J(cm)$^{-2}$
Green sensitivity	~1.2–1.6 × 10^{-3} J(cm)$^{-2}$
Red sensitivity	~0.8–1.2 × 10^{-3} J(cm)$^{-2}$
Color sensitivity peaked at	633 nm and 530 nm

Photopolymer Materials

The color holography photopolymer material from DuPont is another alternative recording material for color holograms. In particular, this type of material is suitable for mass production of color holograms. Although, being less sensitive than the ultra-fine-grain silver halide emulsion, it has special advantages of easy handling and dry processing (only UV-curing and baking). The DuPont color photopolymer material has a coated film layer thickness of about 20 μm. The photopolymer film is generally coated in a 12.5-inch width on a 14-inch wide Mylar® polyester base which is 0.002-inch thick. The film is protected with a 0.00092-inch thick Mylar® polyester cover sheet. However, the panchromatic polymer material is only supplied to specially selected and approved hologram producers. The reason for this is that some applications of color holograms are in the field of document security where optically variable devices (OVDs) are produced.

The recording of a color hologram on DuPont polymer is simple. The film has to be laminated to a piece of clean glass or attached to a glass plate using an index-matching liquid. To obtain the right color balance, the RGB sensitivity depends on the particular material, but typically red sensitivity is lower than green and blue sensitivities. It is difficult to obtain high red-sensitivity of photopolymer materials. Simultaneous exposure is the best recording technique for photopolymer materials. Holograms can be recorded manually, but in order to produce large quantities of holograms, a special machine is required. For hologram replication the scanning technique can provide the highest production rate. In this case, three scanning laser lines are needed, which can be adjusted in such a way that all three simultaneously can scan the film. The color photopolymer material needs an overall exposure of about 10 mJ(cm)$^{-2}$.

After the exposure is finished, the film has to be exposed to strong white or UV light. DuPont recommends about 100 mJ(cm)$^{-2}$ exposure at 350–380 nm. After that, the hologram is put in an oven at a temperature of 120 °C for two hours in order to increase the brightness of the image.

Laser Wavelengths for Color Holograms

The problem of choosing optimal primary laser wavelengths for color holography is illustrated in the 1976 CIE chromaticity diagram (Figure 1) which indicates suitable laser wavelengths for color holograms.

It may seem that the main aim of choosing the recording wavelengths for color holograms would be to cover as large an area of the chromaticity diagram as possible. However, there are many other

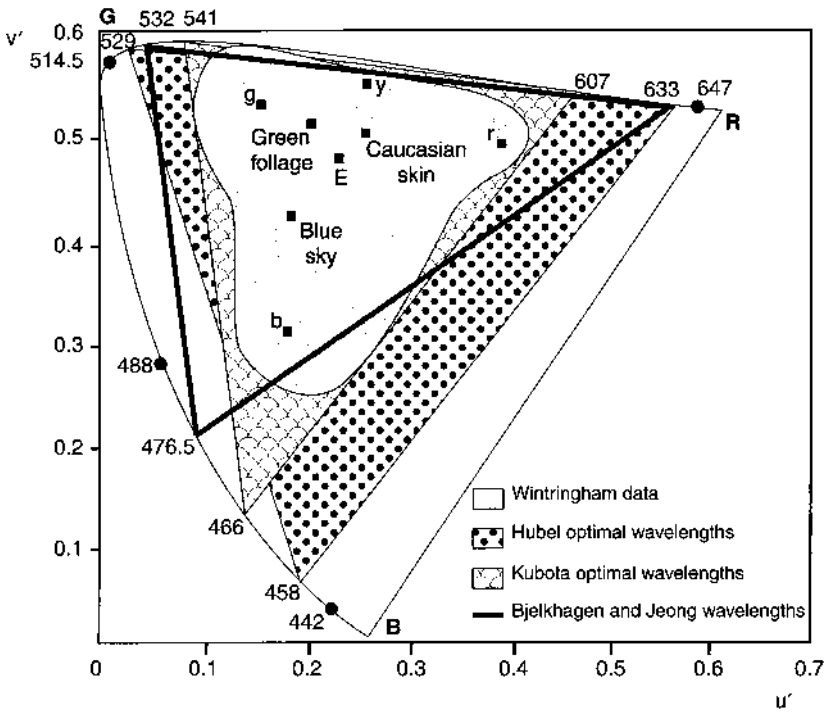

Figure 1 The 1976 CIE uniform scales chromaticity diagram shows the gamut of surface colors and positions of common laser wavelengths. Optimal color-recording laser wavelengths are also indicated.

considerations that must be taken into account when choosing the wavelengths for color holograms. One of these important considerations is the question of whether three wavelengths are really sufficient for color holography. The wavelength selection problem for color holography has been treated in several papers, for example by Peercy and Hesselink (1994).

Hubel and Solymar (1991) provided a definition of color recording in holography:

> A holographic technique is said to reproduce 'true' colors if the average vector length of a standard set of colored surfaces is less than 0.015 chromaticity coordinate units, and the gamut area obtained by these surfaces is within 40% of the reference gamut. Average vector length and gamut area should both be computed using a suitable white light standard reference illuminant.

An important factor to bear in mind when working with silver halide materials is that a slightly longer blue wavelength than the optimal one might give higher-quality holograms (better signal-to-noise ratio) because of reduced Rayleigh scattering during the recording. Another important factor to consider is the reflectivity of the object at primary spectral wavelengths. It has been shown that the reflectivity of an object at three wavelength bands, peaked at 450, 540, and 610 nm, has an important bearing on color reconstruction. These wavelengths can also be considered optimal for the recording of color holograms even though the laser lines are very narrow-band. A triangle larger than necessary can be considered in order to compensate for color desaturation (color shifting towards white) that takes place when reconstructing color reflection holograms in white light. According to Hubel's color rendering analysis, the optimal wavelengths are 464, 527, and 606 nm for the sandwich silver halide recording technique. If the calculations are performed to maximize the gamut area instead, the following set of wavelengths is obtained: 456, 532, and 624 nm. However, when approached from a different viewpoint, the optimal trio of wavelengths based on the reconstructing light source of 3400 K, a 6 μm thick emulsion with a refractive index of 1.63 and an angle of 30° between the object and the reference beam the following wavelengths were obtained: 466.0, 540.9, and 606.6 nm. Peercy and Hesselink (1994) discussed wavelength selection by investigating the sampling nature of the holographic process. During the recording of a color hologram the chosen wavelengths point-sample the surface-reflectance functions of the object. This sampling on color perception can be investigated by the tristimulus value of points in the reconstructed hologram which is mathematically equivalent to integral approximations for the tristimulus integrals. Peercy and Hesselink used both Gaussian quadrature and Riemann summation for the approximation of the tristimulus integrals. In the first case they found the wavelengths to be 437, 547, and 665 nm. In the second case the wavelengths were 475, 550, and 625 nm. According to Peercy and Hesselink, the sampling approach indicates that three monochromatic sources are almost always insufficient to preserve all of the object's spectral information accurately. They claim that four or five laser wavelengths are required.

Only further experiments will show how many wavelengths are necessary and which combination is the best for practical purposes. Another factor that may influence the choice of the recording wavelengths is the availability of cw lasers currently in use in holographic recording, for example, argon ion, krypton ion, diode-pumped solid state (DPSS) frequency-doubled Nd:YAG, helium neon, and helium cadmium lasers (Table 2).

The recent progress in DPSS laser technology has made available both red and blue DPSS lasers. These lasers are air-cooled, small, and each laser requires less than a hundred watts of electric power to operate. Usually, a set of three DPSS lasers will be the best choice of cw lasers for color holography in the future.

Setup for Recording Color Holograms

A typical reflection hologram recording setup is illustrated in Figure 2.

Table 2 Wavelengths from cw lasers

Wavelength [nm]	*Laser type*	*Single line power [mW]*
442	Helium cadmium	<100
457	DPSS blue	<500
458	Argon ion	<500
468	Krypton ion	<250
476	Krypton ion	<500
477	Argon ion	<500
488	Argon ion	<2000
497	Argon ion	<500
502	Argon ion	<400
514	Argon ion	<5000
521	Krypton ion	<100
529	Argon ion	<600
531	Krypton ion	<250
532	DPSS green	<3000
543	Green neon	<10
568	Krypton ion	<100
633	Helium neon	<80
647	Krypton ion	<2000
656	DPSS red	<1000

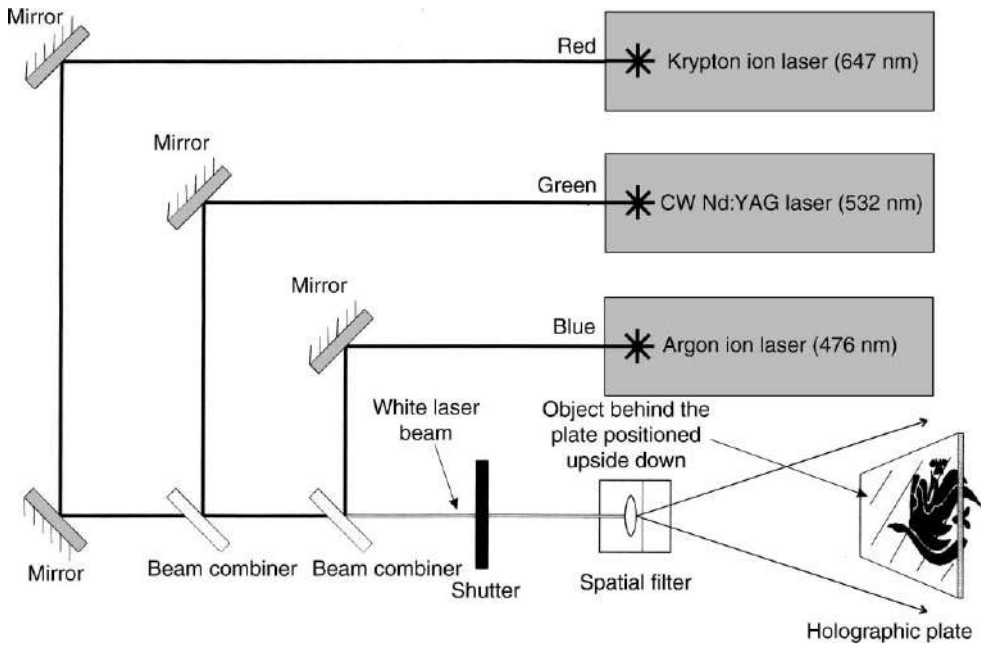

Figure 2 The setup for recording color reflection holograms.

The different laser beams necessary for the exposure of the object pass through the same beam expander and spatial filter. A single-beam Denisyuk arrangement is used, i.e., the object is illuminated through a recording holographic plate. The light reflected from the object constitutes the object beam of the hologram. The reference beam is formed by the three expanded laser beams. This 'white' laser beam illuminates both the holographic plate and the object itself through the plate. Each of the three primary laser wavelengths forms its individual interference pattern in the emulsion, all of which are recorded simultaneously during the exposure. In this way, three holographic images (a red, a green, and a blue image) are superimposed upon one another in the emulsion.

Three laser wavelengths are employed for the recording: 476 nm, provided by an argon ion laser, 532 nm, provided by a cw frequency-doubled Nd:YAG laser, and 647 nm, provided by a krypton laser. Two dichroic filters are used for the combining of the three laser beams. The 'white' laser beam goes through a spatial filter, illuminating the object through the holographic plate.

By using the dichroic filter beam combination technique, it is possible to perform simultaneous exposure recording, which makes it possible to control independently the RGB ratio and the overall exposure energy in the emulsion. The RGB ratio can be varied by individually changing the output power of the lasers, while the overall exposure energy is controlled solely by the exposure time. The overall energy density for exposure is about 3 mJ(cm)$^{-2}$.

In the initial experiments performed by the author in 1993, a specially designed test object consisting of the 1931 CIE chromaticity diagram, a rainbow ribbon cable, pure yellow dots, and a cloisonné elephant, was used for the color balance adjustments and exposure tests. Later, another test target was employed – the Macbeth ColorChecker® chart, which was used for color rendering tests.

Color reflection holograms can also be produced using copying techniques so that projected real images can be obtained, however, normally associated with a restricted field of view. For many display purposes, the very large field of view obtainable in a Denisyuk color hologram is often more attractive.

Processing of Color Holograms

The dry processing of color holograms recorded on photopolymer materials has already been described. The process is simple and very suitable for machine processing using, for example, a baking scroll oven. The processing of silver halide emulsions is more difficult and critical. The Slavich emulsion is rather soft, and it is important to harden the emulsion before the development and bleaching takes place. Emulsion shrinkage and other emulsion distortions caused by the active solutions used for the processing must be avoided. In particular, when recording master color holograms intended for photopolymer replication, shrinkage control is extremely important. The processing steps are summarized in Table 3.

Table 3 Color holography processing steps

1.	Tanning in a formaldehyde solution	6 min
2.	Short rinse	5 s
3.	Development in the CWC2 developer	3 min
4.	Wash	5 min
5.	Bleaching in the PBU-amidol bleach	~5 min
6.	Wash	10 min
7.	Soaking in acetic acid bath (printout prevention)	1 min
8.	Short rinse	1 min
9.	Washing in distilled water with wetting agent added	1 min
10.	Air drying	

It is very important to employ a suitable bleach bath to convert the developed silver hologram into a phase hologram. The bleach must create an almost stain-free clear emulsion so as not to affect the color image. In addition, no emulsion shrinkage can be permitted, as it would change the colors of the image. Washing and drying must also be done so that no shrinkage occurs. Finally, to prevent any potential emulsion thickness variation by humidity variations, the emulsion needs to be protected by a glass plate sealed onto the hologram plate.

Evaluation of Color Holograms

Recorded color holograms of the two test targets are presented here. The illuminating spotlight to reconstruct the recorded color holograms was a 12-Volt 50-Watt halogen lamp. This type of spotlight is suitable for displaying color holograms. The selection of a suitable lamp for the reconstruction of color holograms is much more important than the selection of lamps for monochrome hologram display. The color balance for the recording of a color hologram must be adjusted to the type of spotlight that is going to be used for the display of the finished hologram. Figure 3 shows a typical normalized spectrum obtained from a white area of the color test target hologram.

This means that the diffraction efficiency of each color component is obtained assuming a flat spectrum of the illuminating source. One should note the high diffraction efficiency in blue, needed to compensate for the low blue light emission of the halogen spotlight. The noise level, mainly in the blue part of the spectrum, is visible and low. The three peaks are exactly at the recording wavelengths; i.e., 647, 532, and 476 nm.

In Table 4 some results of the Macbeth ColorChecker® investigation are presented. The 1931 CIE x and y coordinates are measured at both the actual target and the holographic image of the target. The measured fields are indicated in the table by color and the corresponding field number.

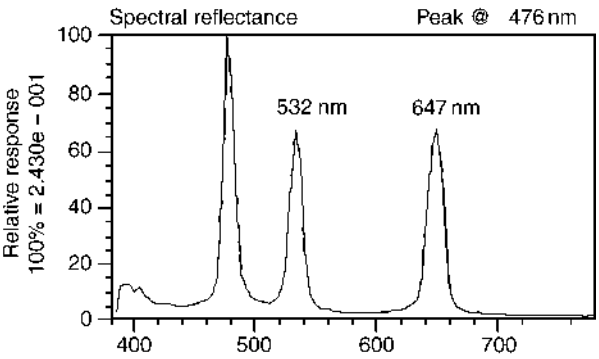

Figure 3 Normalized spectrum from a white area of a color test target hologram.

Color reproductions of two color holograms by the author are presented. A photograph of the 100 mm by 120 mm hologram of the CIE test object is shown in Figure 4. Featured in Figure 5 is a 200 mm by 250 mm hologram of a large Russian egg.

Computer-Generated Color Holograms

Today it is not possible to obtain a computer-generated hologram (CGH) with the same high information content as the laser-recorded ones of real objects. What may become possible in the future is a technique to compute and record the interference pattern that is stored in a Denisyuk color hologram which, upon illumination, can produce an image like the laser-recorded ones of real objects. Therefore, the best compromise is to use holographic stereograms which can provide high-quality computer-generated images. A holographic stereogram is created by using a series of 2D photographic images or 2D images displayed on a LCD screen from which the hologram is recorded. In order to obtain a high-quality holographic image with a large field of view, many 2D images are needed. As mentioned earlier, some color CGHs of the rainbow transmission type have been produced. However, image color changes as a function of viewer position in that type of hologram, making them less attractive than computer-generated reflection holograms. Over the last few years there has been rapid progress in color CGHs of the reflection type. Remarkable results have been achieved by Klug *et al.* (1997) at Zebra Imaging Inc. A new technique, to record very large full-parallax color reflection CGHs, has been developed. Color holograms can be produced, having both vertical and horizontal parallax, with a 100° field of view. The generation of a holographic hardcopy of either digitized images or computer graphics models is based on the following technique. The 'object'

Table 4 Chromaticity coordinates from color hologram recording tests using the Macbeth ColorChecker®

Object	*White #19*	*Blue #13*	*Green #14*	*Red #15*	*Yellow #16*	*Magenta #17*	*Cyan #18*
CIE *x y*	*x*/*y*	*x*/*y*	*x*/*y*	*x*/*y*	*x*/*y*	*x*/*y*	*x*/*y*
Target	0.435/0.405	0.295/0.260	0.389/0.514	0.615/0.335	0.517/0.450	0.524/0.322	0.285/0.380
Image	0.354/0.419	0.335/0.362	0.337/0.449	0.476/0.357	0.416/0.437	0.448/0.338	0.295/0.366

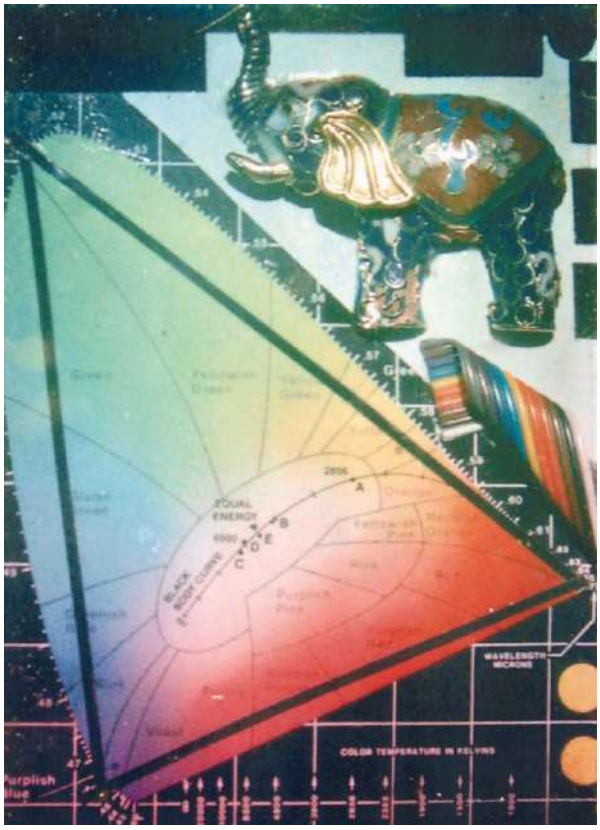

Figure 4 Hologram of the CIE test object recorded with 476, 532, and 647 nm laser wavelengths. Notice the rainbow ribbon cable, the pure yellow dots, the full range of colors, and a balanced white in the center of the CIE color test target.

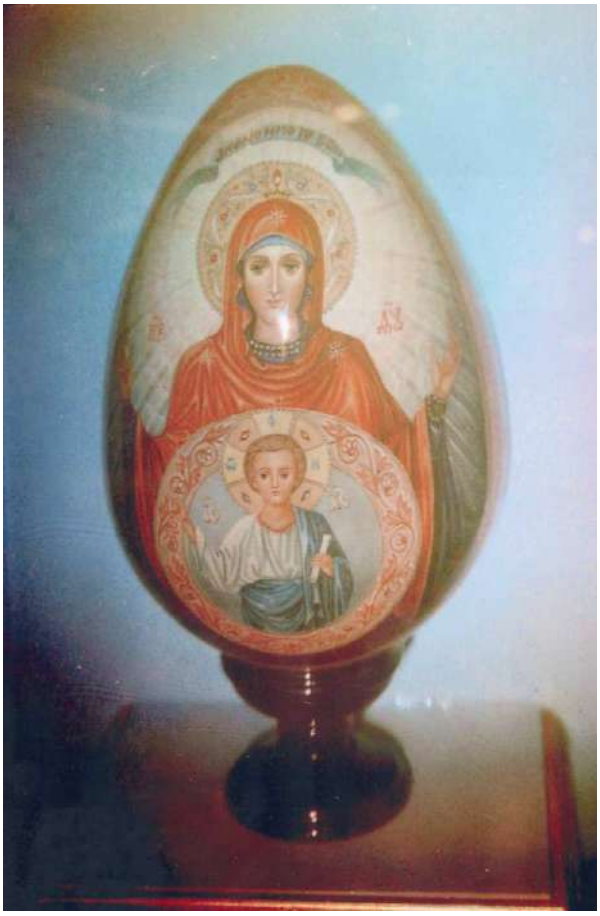

Figure 5 Color hologram of a large Russian egg, size 100 mm by 120 mm.

sub-beam is directed through a sequence of digital images on a liquid-crystal screen. Each resulting exposure, about two millimeters square, is called a 'hogel'. The full-color hogels are the holographic building blocks of a finished CGH image. In an automated step-and-repeat process, 9×10^4 hogels are formed on a flat square tile of DuPont panchromatic photopolymer film by simultaneous RGB laser exposures. The 60 cm by 60 cm tile itself is the finished hologram, or to obtain larger holograms, the 3D image is made in 60 cm by 60 cm tiled increments. So far, the largest image created was of Ford's P2000 Prodigy concept car in which ten such holograms tiled together made up the large color reflection hologram (1.2 m by 3 m) which is reproduced in Figure 6.

This technique has opened the door to real 3D computer graphics. However, generating such a large hologram is a very time-consuming process. The Ford P2000 hologram took almost two weeks (300 hours) to produce, since each individual panel requires a 24-hour recording time.

The Future of Color Holography

The manufacturing of large-format color reflection holograms is now possible. Mass production of color holograms on photopolymer materials has started in Japan. Although good color rendition can be obtained, some problems remain to be solved, such

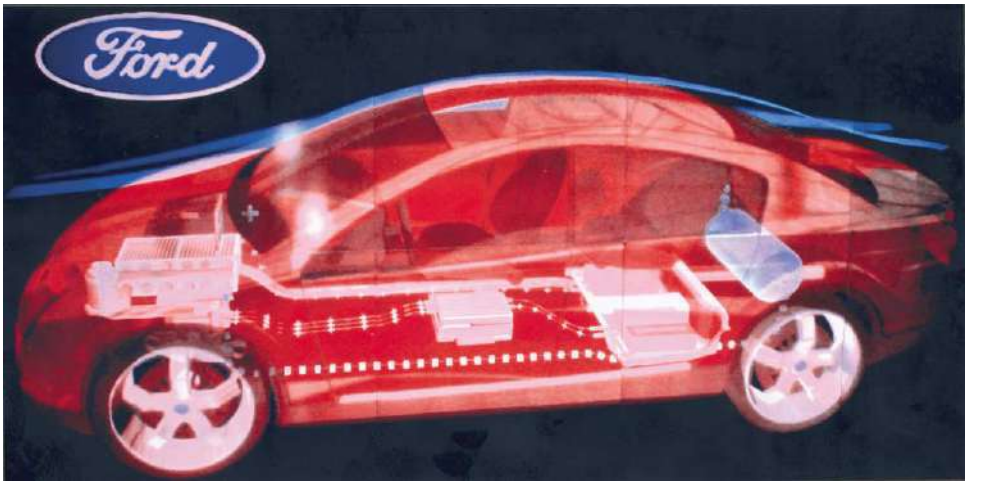

Figure 6 Computer-generated color hologram of Ford's P2000 Prodigy concept car. A ten tile full parallax Mosaic hologram produced by Zebra Imaging, Inc., size 1.2 m by 3 m.

as desaturation, image resolution, signal-to-noise ratio and dynamic range. Other limitations concerning holographic color recording include the fact that some colors we see are the result of fluorescence, which cannot be recorded in a hologram. There are some differences in the recorded colors in one of the test charts shown here. However, color rendition is a very subjective matter. Different ways of rendition may be preferable for different applications, and different people may have different color preferences.

At the present time, research on processing color reflection holograms recorded on ultra-fine-grain materials is still in progress. Work is carried out on microcavity structure holograms in order to produce low-noise reflection holograms with high diffraction efficiency. For producing large quantities of color holograms, the holographic photopolymer materials are most suitable and will be further improved in the future.

Another important field of research is the development of a three-wavelength pulsed laser. Employing such a laser, dynamic events, as well as portraits, can be recorded in a holographic form. Currently, research work is in progress at the French German Research Institute ISL, Saint Louis, France, as well as at the Geola company in Lithuania.

The virtual color image behind a color holographic plate represents the most realistic image of an object that can be obtained today. The extensive field of view adds to the illusion of beholding a real object rather than an image of it. The wavefront reconstruction process recreates accurately the three-wavelength light scattered off the object during the recording of the color hologram. This 3D imaging technique has many obvious applications, in particular, in displaying unique and expensive artifacts. There are also many potential commercial applications of this new feature of holographic imaging, provided that the display problem can be solved using some sort of integrated lighting.

Today, it is technologically possible to record and replay acoustical waves with very high fidelity. Hopefully, holographic techniques will be able to offer the same possibility in the field of optical waves, wavefront storage, and reconstruction. Further development of computer-generated holographic images will make it possible to display extremely realistic full-parallax 3D color images of nonexisting objects, which could be applied in various spheres, for example, in product prototyping, as well as in other applications in 3D visualization and three-dimensional art. Eventually it will become possible to generate true-holographic 3D color images in real-time, as computers become faster and possess greater storage capacity. However, the most important issue is the development of extremely high-resolution electronic display devices which are necessary for electronic holographic real-time imaging.

See also

Holography, Techniques: Computer-Generated Holograms.

Further Reading

Bjelkhagen HI (1993) *Silver Halide Recording Materials for Holography and Their Processing*. Springer Series in Optical Sciences, vol. 66. New York: Springer-Verlag.

Bjelkhagen HI, Jeong TH and Vukičević D (1996) Color reflection holograms recorded in a panchromatic ultrahigh-resolution single-layer silver halide emulsion. *Journal of Imaging Science & Technology* 40: 134–146.

Bjelkhagen HI, Huang Q and Jeong TH (1998) Progress in color reflection holography. In: Jeong TH (ed.) *Sixth International Symposium on Display Holography*, Proc. SPIE, vol. 3358, pp. 104–113. Bellingham, WA: The International Society for Optical Engineering.

Hariharan P (1983) Colour holography. In: Wolf E (ed.) *Progress in Optics*, vol. 20, pp. 263–324. Amsterdam: North-Holland.

Hubel PM and Solymar L (1991) Color reflection holography: theory and experiment. *Applied Optics* 30: 4190–4203.

Klug MA, Klein A, Plesniak W, Kropp A and Chen B (1997) Optics for full parallax holographic stereograms. In: Benton SA and Trout TJ (eds) *Practical Holography XI and Holographic Materials III*, Proc. SPIE, vol. 3011, pp. 78–88. Bellingham, WA: The International Society for Optical Engineering.

Kubota T (1986) Recording of high quality color holograms. *Applied Optics* 25: 4141–4145.

Leith EN and Upatnieks J (1964) Wavefront reconstruction with diffused illumination and three-dimensional objects. *Journal of the Optical Society of America* 54: 1295–1301.

Peercy MS and Hesselink L (1994) Wavelength selection for true-color holography. *Applied Optics* 33: 6811–6817.

Usanov YuE and Shevtsov MK (1990) Principles of fabricating micropore silver-halide-gelatin holograms. *Optics and Spectroscopy (USSR)* 69: 112–114.

Watanabe M, Matsuyama T, Kodama D and Hotta T (1999) Mass-produced color graphic arts holograms. In: Benton SA (ed.) *Practical Holography XIII*, Proc. SPIE, vol. 3637, pp. 204–212. Bellingham, WA: The International Society for Optical Engineering.

Computer-Generated Holograms

W J Dallas, University of Arizona, Tucson, AZ, USA

A W Lohmann, University of Erlangen-Nürnberg, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

A hologram is a tool for shaping the amplitude and the phase of a light wave. Shaping the amplitude distribution is easy, one way is by a photographic plate, another way is by binary pixels as used in 'half-tone printing'. The phase distribution can be shaped by RE-fraction on a thin film of variable thickness, or by DIF-fraction on a diffraction grating that has been deliberately distorted.

Both kinds of holograms can be manufactured based on a computer design, the refractive version being called a 'kinoform', and the diffractive version being known as a computer-generated hologram (CGH). We will explain the fundamentals of computer holography with emphasis on the 'Fourier–CGH', and will also highlight the large variety of applications.

From the Classical Hologram to the Computer-Generated Hologram: CGH

Holography is a method to form images, consisting of two steps: recording and reconstruction. In the recording step, some interference fringes are recorded on a photographic plate (Figure 1a). The two interfering waves originate from the object and from a reference light source. After being developed by the usual photo-chemical treatment the photographic plate is called a hologram. In the reconstruction step, the hologram acts as a diffracting object that is illuminated by a replica of the reference light. Due to diffraction, the light behind the hologram is split into three parts (Figure 1b). One part proceeds to the observer who perceives a virtual object where the genuine object had been originally during the recording step.

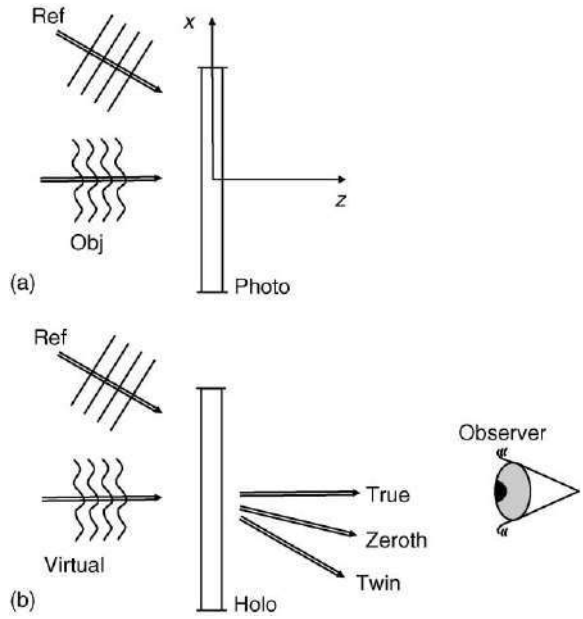

Figure 1 Holographic image formation in two steps: (a) recording the hologram, (b) reconstructing the object wave.

This was a brief description of 'classical-holography'. In 'computer holography' the first step, recording, is synthetic. In other words, the propagation of the light from the object to the photographic plate is simulated digitally. The simulation includes the addition of the object wave to the reference wave and the subsequent modulus square process, which describes the transition from the two complex amplitudes to the hologram irradiance:

$$|u_0(x) + u_R(x)|^2 = I_H(x) \qquad [1]$$

The amplitude transmittance of the hologram is

$$T_H(x) = c_0 + c_1 I_H(x) = c_1 u_0 u_R^* + \cdots \qquad [2]$$

The coefficients c_0 and c_1 represent the photochemical development. In the reconstruction process (Figure 1b) the transmittance T_H is illuminated by the reference beam. Hence, we get

$$u_R(x) T_H(x) = c_1 u_0(x) |u_R(x)|^2 + \cdots \qquad [3]$$

If the reference intensity $|u_R(x)|^2$ is constant, this term represents a reconstruction of the object wave $u_0(x)$. The omitted parts in eqns [2] and [3] describe those beams that are ignored by the observer (Figure 1b). The hologram transmittance $T_H(x)$ is an analog signal:

$$0 \le T_H \le 1 \qquad [4]$$

Before the 1960s, when computer holograms were invented, the available plotters could produce only binary transmittance patterns. Hence, it was necessary to replace the analog transmission $T_H(x)$ by a binary transmission $B_H(x)$, which is possible without loss of image quality. In addition, binary computer holograms have better light efficiency and a better signal-to-noise ratio. They are also more robust when being copied or printed.

Another advantage of a CGH is that the object does not have to exist in reality, which is important for some applications. The genuine object may be difficult to manufacture or, if it is three-dimensional, difficult to illuminate. The CGH may be used to implement an algorithm for optical image processing. In that case, the term 'object' becomes somewhat fictitious. We will discuss more about applications, towards the end of this article.

From a Diffraction Grating to a Fourier CGH

Now we will explain, in some detail, the 'Fourier type' CGH. Fourier holograms are more popular than two other types: 'image-plane' CGHs and 'Fresnel-type' CGHs. The Fresnel-type will be treated in the section on 'software' because of some interest in 3-D holograms. Certain hardware issues will also be described.

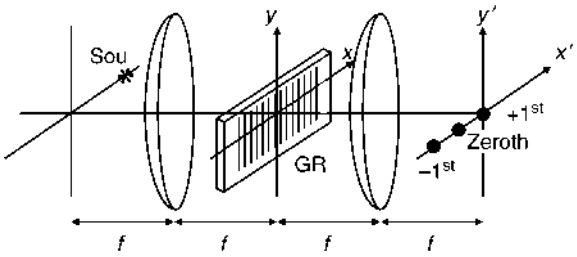

Figure 2 Grating diffraction, but with an off-axis source.

The explanation of Fourier CGHs begins with grating diffraction (Figure 2). The only uncommon feature in Figure 2 is the off-axis location of the source. It is arranged such that the plus-first diffraction order will hit the center of the output plane.

We will now convert a simple Ronchi grating (Figure 3a) into a Fourier CGH. The Ronchi grating transmittance can be expressed as the Fourier series:

$$G(x) = \sum_{(m)} C_m \exp\left(\frac{2\pi i m x}{D}\right); \quad C_0 = \frac{1}{2};$$

$$C_m = \frac{\sin(\pi m/2)}{\pi m} \qquad [5]$$

Now we shift the grating bars by an amount PD. Then we modify the width from $D/2$ to WD, as shown in Figure 3b. The Fourier coefficients are now

$$C_m = \frac{\exp(2\pi i m P)\sin(\pi m W)}{\pi m} \qquad [6]$$

The C_{+1} coefficient is responsible for the complex amplitude at the center of the readout plane:

$$C_{+1} = \frac{\exp(2\pi i P)\sin(\pi W)}{\pi} \qquad [7]$$

We are able now to control the beam magnitude, $A = (\frac{1}{\pi})\sin(\pi W)$, by the relative bar width W and the phase $\phi = 2\pi P$ by the relative bar shift P. This particular kind of phase shift is known as 'detour phase'.

So far, the light that ends up at the center of the exit plane, leaves the grating as a plane wave with a wave vector parallel to the optical axis. Now we distort the grating on purpose (Figure 3c):

$$W \to W(x, y); \qquad P \to P(x, y) \qquad [8]$$

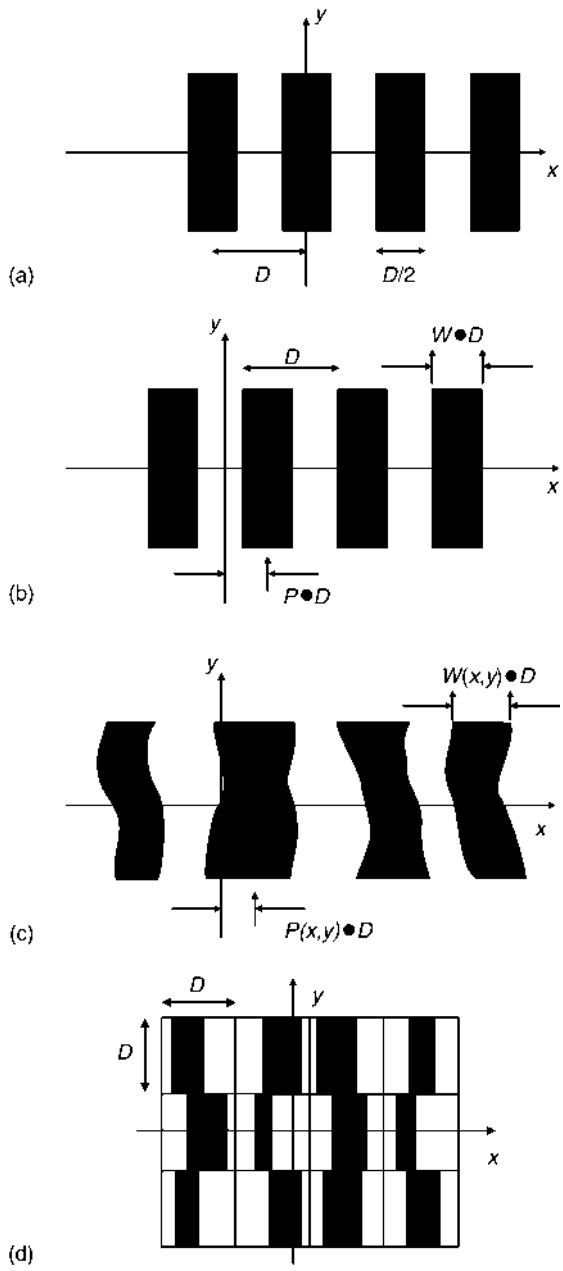

Figure 3 Modifying a grating into a hologram in 3 steps: (a) Ronchi grating, (b) Modified width and location of the grating bars, (c) Distorted grating bars, (d) as in **Figure 3c**, but now discrete distortions.

These distortions should be mild enough that the plus-first order light behind the distorted grating can be described by the complex amplitude:

$$\left(\frac{1}{\pi}\right)\sin[\pi W(x,y)]\exp[2\pi i P(x,y)] = u_{\mathrm{H}}(x,y) \qquad [9]$$

This is a generalization of eqn [7]. The phase $2\pi P(x,y)$ covers the range from $-\pi$ to $+\pi/2$ if

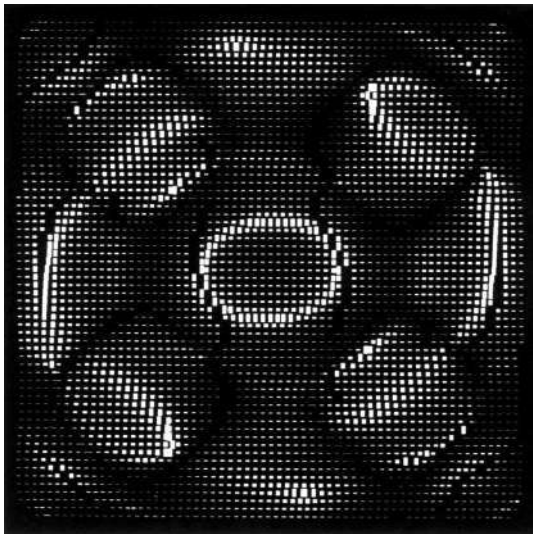

Figure 4 CGH with 64 × 64 cells.

the shift is bounded by $|P| \leq \frac{1}{2}$. That is enough to cover the range of complex amplitudes within the circle of $|u_{\mathrm{H}}| \leq \frac{1}{\pi}$. We will return shortly to the condition 'mild-enough distortions' in the context of **Figure 4**.

Most plotters move their pens only in the x-direction and y-direction and printers place dots on a grid. Therefore it is convenient to replace the continuous (x, y) variations of W and P by piecewise constant variations, as in **Figure 3d**. That restriction turns out to be acceptable, as demonstrated by the first computer holograms. The theoretical proof involves the sampling theorem.

Recall that the light propagation through a 2-f system (**Figure 2**, between grating and output plane) can be described by a Fourier transformation:

$$\int\int u_{\mathrm{H}}(x,y)\exp\left[\frac{-2\pi i(xx'+yy')}{\lambda f}\right]\mathrm{d}x\,\mathrm{d}y = \tilde{u}_{\mathrm{H}}\left(\frac{x'}{\lambda f},\frac{y'}{\lambda f}\right) \qquad [10]$$

Suppose now that we wish to see a particular image: $v(x', y')$. Hence, we request that

$$v(x',y') = \tilde{u}_{\mathrm{H}}\left(\frac{x'}{\lambda f},\frac{y'}{\lambda f}\right)\mathrm{RECT}\left(\frac{x'}{\lambda f}\right) \qquad [11]$$

The RECT function indicates that only the PLUS first diffraction order is of interest. As a consequence we have to select the grating distortions $W(x, y)$ and $P(x, y)$ such that the hologram $u_{\mathrm{H}}(x, y)$ is the

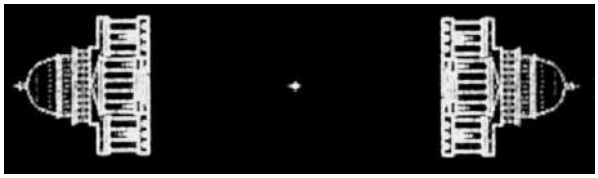

Figure 5 A reconstruction from a 1024 × 1024 CGH. The real image and the symmetrical twin image appear. The zeroth-order is blocked.

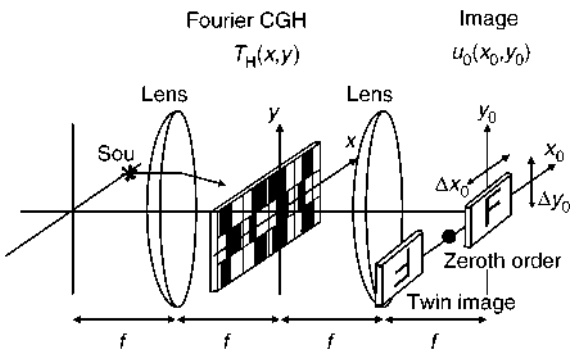

Figure 6 Reconstruction setup for a Fourier CGH.

Fourier transform of $v(x', y')$:

$$u_H(x,y) = \int\int v(x',y') \exp\left[\frac{2\pi i(xx'+yy')}{\lambda f}\right] dx'\,dy' \qquad [12]$$

The condition 'mild enough distortions' has been treated in the literature. There is not enough space here to present that part of the CGH theory. But we show an actual CGH with 64 × 64 cells (Figure 4). This CGH is no longer a grating, but, it shows enough resemblance to our heuristic derivation. Hence, the CGH theory appears to be trustworthy. A CGH looks less regular if the image $v(x', y')$ contains a random phase that simulates a diffuser. The diffuser spreads light across the CGH and so levels out unacceptably large fluctuations in the amplitude.

With 512 × 512 or 1024 × 1024 cells in a CGH one can obtain an image of the quality seen in Figure 5. The reconstruction setup is shown in Figure 6, here with the letter F as the image.

About Some CGH Algorithms

The computational effort for getting the desired hologram amplitude of a Fourier hologram u_H is reasonable due to the powerful FFT algorithm. However, to use it the data must be discrete. In other words, the CGH consists of cells, centered at $x_m = mD$, $y_n = nD$ as shown in Figure 3d. The questions now are: 'What is the proper cell size D?' and 'How many cells are needed?' The answers to those questions depend on the parameters of the image. If the zero order (see Figure 6) should be at the distance Δx_0, then the quasi-grating period D should obey:

$$\frac{\lambda f}{D} \geq \Delta x_0 \qquad [13]$$

And if the size of an image pixel should be as fine as δx_0, the size Δx_H of the hologram should obey:

$$\frac{\lambda f}{\Delta x_H} \leq \delta x_0 \qquad [14]$$

This condition is plausible because the finite size Δx_H of the hologram acts like a resolution-limiting aperture. The combination of eqns [13] and [14] yields

$$\frac{\Delta x_0}{\delta x_0} \leq \frac{\Delta x_H}{D} \qquad [15]$$

This means that the number of image pixels $\Delta x_0/\delta x_0$ is bounded by the number of CGH cells $\Delta x_H/D$. The generalization to two dimensions is straightforward. Again, Fourier CGHs benefit from the fact that the light propagation from the virtual object to the hologram plane can be described by a Fourier transformation. Hence, the FFT can be used.

And now we move to the synthesis of a Fresnel hologram. A Fresnel hologram is recorded at a finite distance z from the object. Hence, we have to simulate the wave propagation in free space from the object (at $z = 0$) to the hologram at a distance z. This free space propagation can be described (in the paraxial approximation) as a convolution of object and quasi-spherical wave:

$$u_0(x) \rightarrow \int_{-\infty}^{\infty} u_0(x') \exp\left[\frac{-i\pi(x-x')^2}{\lambda z}\right] dx' = u(x,z) \qquad [16]$$

In the Fourier domain the corresponding operation is:

$$\tilde{u}_0(\mu) \rightarrow \tilde{u}_0(\mu) \exp[-i\pi\lambda z\mu^2] = \tilde{u}(\mu,z) \qquad [17]$$

The indirect execution of eqn [16], based on eqn [17], is easy: a Fourier transformation of the object converts $u_0(x)$ into $\tilde{u}_0(\mu)$, which is then multiplied by the quadratic phase factor, yielding $\tilde{u}(\mu,z)$. An inverse Fourier transformation produces $u(x,z)$. The two Fourier transformations consist typically of $4N \log N$ multiply/add operations. N is the number of pixels: image size Δx_0, divided by the pixel size δx_0.

In the case of a rectangular object $u_0(x, y)$, the number of pixels is $\Delta x_0 \Delta y_0 / \delta x_0 \delta y_0$.

Now to describe the algorithmic effort for computing the convolution, the $u_0(x)$ has N pixels, and, the lateral size of the exponential is essentially $M = z/\delta z$, where δz represents 'depth of focus'. M describes the lateral size of the diffraction cone, measured in pixel size units δx_0, and the convolution involves MN multiply/adds. A comparison of these two algorithms is dictated by the ratio

$$\frac{M}{4 \log N} \qquad [18]$$

The direct convolution is preferable if the distance z is fairly small. The integration limits of eqn [16] depend on the oscillating phase $\pi x^2/\lambda z$. Effectively, nothing is contributed to this integral, when this phase varies by more than 2π over the length, δx_0, of an object pixel. Note that the pixel size of $u(x, z)$ does not change with the distance z, because the bandwidth $\Delta\mu(z) = \Delta\mu_0$ is z-independent. This is so because the power spectrum is z-invariant:

$$|\tilde{u}(\mu, z)|^2 = |\tilde{u}_0(\mu)|^2 \qquad [19]$$

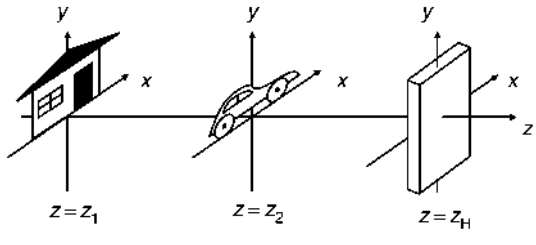

Figure 7 Synthesis of a hologram for a three-dimensional object.

It is easy to proceed from here to the CGH synthesis of a 3-D object. One starts with an object detail $u_1(x)$ at z_1, for example, the house as in Figure 7. Then, one propagates to the next object detail at z_2. This second detail may act as a multiplication (transmission) and as addition (source elements). This process is repeated as the wave moves towards the plane of the hologram. There, a reference wave has to be added, as in the Fourier case.

A very powerful algorithm for the CGH synthesis is 'IFTA' (iterative Fourier transform algorithm). Suppose, we want to design a Fourier CGH, which looks like a person and whose reconstructed image shows the signature of that person (Figure 8a,b). In other words, the two intensities

$$|u_{\mathrm{H}}(x, y)|^2 \text{ and } |\tilde{u}_{\mathrm{H}}(\mu, \nu)|^2 \qquad [20]$$

are determined. But, the phase distributions of u_{H} and $\tilde{u}_{\mathrm{H}}$ are still at our disposal. The IFTA algorithm will yield those phases in many cases at the expense of ten, hundred or more Fourier transformations.

The IFTA algorithm is sometimes called 'Fourier Ping-Pong' because it bounces back and forth between the (x, y) and the (μ, ν) domain. The amplitudes $|u_{\mathrm{H}}|$ and $|u_0|$ are enforced at every opportunity. But the associated phases are free to vary and to converge, eventually. If so, u_{H} and u_0 are completely known. The IFTA will not always converge, for example not if

$$|u_{\mathrm{H}}(x, y)| = \delta(x, y) \text{ and } |\tilde{u}_{\mathrm{H}}(\mu, \nu)| = \delta(\mu, \nu) \qquad [21]$$

But the chances are fairly good if both amplitudes are often close to one and seldom near zero.

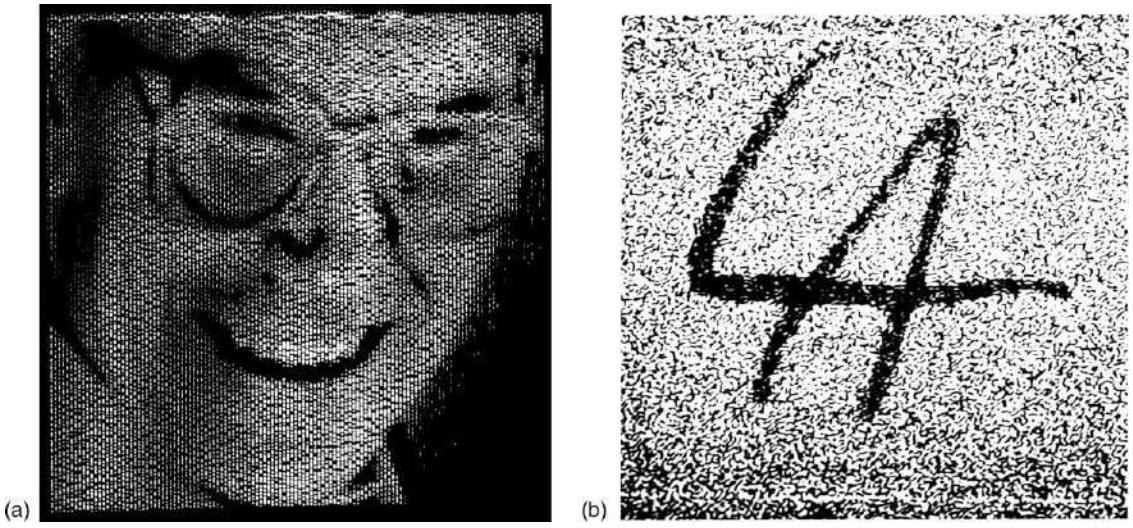

Figure 8 IFTA-CGH: (a) The Fourier CGH. (b) The optical reconstruction thereof. (Courtesy of H.O. Bartelt.)

Finally, an image plane CGH is trivial from an algorithmic point of view. The simulated propagation from object plane to image plane is usually an identity, apart from the finite bandwidth of the image-forming system. Nevertheless, image plane holograms can be quite useful, for example for interferometric testing of a nonspherical mirror of an astronomical telescope. The CGH is laid out to provide a wavefront that is suitable for an interferometric 'null test'. The CGH serves as a 'synthetic prototype'.

A fourth kind of CGH, to be named near-field CGH may emerge. It operates in the near-field and it employs evanescent waves as they occur, if object details are comparable in size with, or even smaller than, the wavelength. Conceivable topics are 'super resolution' and 'sub-lambda lithography'. Fundamentals and history of classical evanescent holography are reported in the Further Reading.

Some Hardware Aspects

We now describe some hardware aspects of the CGHs. Remembering the CGH cell structure in Figure 3d, we saw the amplitude encoded as relative width W (eqn [6]). Instead of the width, one may use the relative height $H \leq 1$ as the amplitude parameter. More efficient, but more difficult to manufacture are saw tooth-shaped hologram cells. The saw tooth prism may be approximated by a phase stair, which is possible with today's lithographic technology. The corresponding theory is called phase quantization.

Some CGH Applications

Some applications had been mentioned already before, e.g., 3-D display (see Figure 7). The data volume can be very high, hence, shortcuts such as eliminating the vertical perspective, are necessary. The largest CGHs for visual light are used for testing astronomical telescopes by means of interferometry. The IFTA algorithm has been used, among other things, for designing a CGH which acts as an optimal diffuser. Speckles and light efficiency are the critical issues in diffuser design. Another project, that also benefits from IFTA, is beam shaping. That is a broad topic with aims at, for example, homogenizing the output of laser, beam structuring for welding, cutting, and scanning.

A CGH can also be used as a spatial filter for image processing. For example, an input image may be Hilbert-transformed in the setup shown in Figure 9a. The filter (Figure 9b) is a grating, of which one sideband is shifted by one-half of the grating period. Such a geometric shift generates the π-phase shift that is needed for the Hilbert transformation. The same setup can also be used to implement Zernike's phase contrast. Now the grating is distorted such that the zero-frequency region at the optical axis is reduced in amplitude by (typically) a factor of 0.2. In addition, the phase is shifted by $\pi/2$ due to a fringe shift of a quarter period (Figure 10a). In the image plane one can see an ordinary image on-axis, and two phase-contrast images in the plus-minus first diffraction orders. One of them has positive phase contrast, the

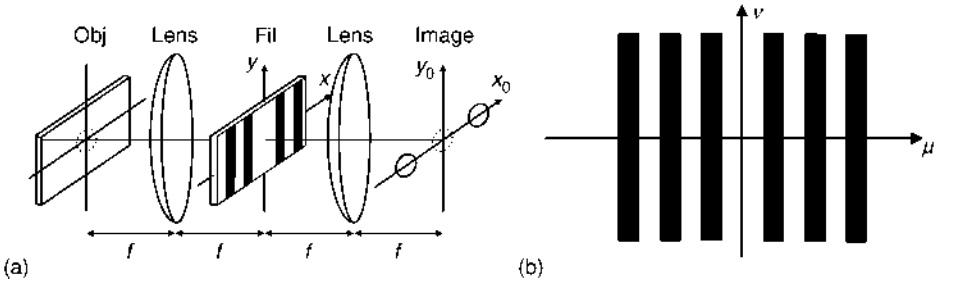

Figure 9 (a) Hilbert setup. (b) Hilbert filter.

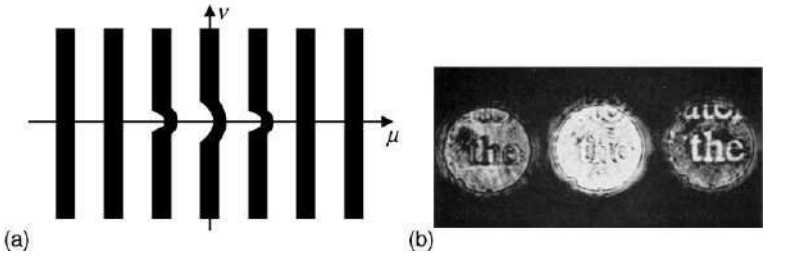

Figure 10 (a) Phase contrast filter. (b) Phase contrast output.

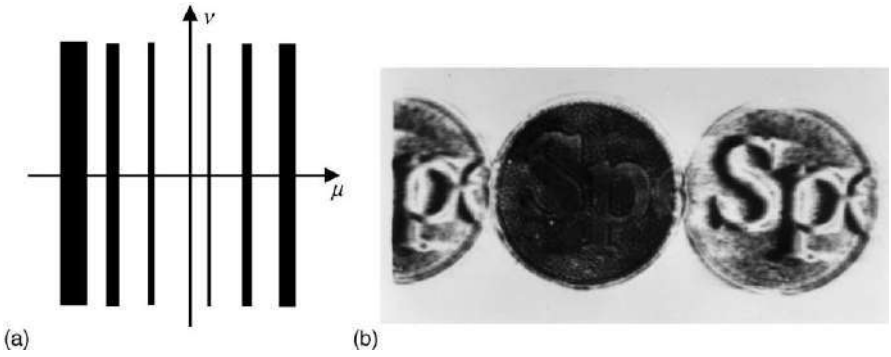

Figure 11 (a) Differentiation filter. (b) Differentiation output.

other one has negative phase contrast (Figure 10b). Another simple spatial filter (Figure 11a) produces a derivative of the input $u(x, y) : \partial u(x, y)/\partial x$ (Figure 11b). Equations [22]–[24] explain this differential filtering experiment.

$$u(x, y) \rightarrow \frac{\partial u(x, y)}{\partial x} = v(x, y) \qquad [22]$$

$$\tilde{u}(\mu, \nu) \rightarrow 2\pi i\mu \cdot \tilde{u}(\mu, \nu) = \tilde{v}(\mu, \nu) \qquad [23]$$

$$\tilde{p}(\mu, \nu) = 2\pi|\mu| \cdot \left\{ \begin{matrix} +i(\mu > 0) \\ -i(\mu > 0) \end{matrix} \right\} \qquad [24]$$

The phase portion of the filter (Figure 11a) is the same as for the Hilbert filter, but, the filter amplitude now enhances the higher frequencies. The object used for getting the results of Figure 10b and 11b was a phase-only object: a bleached photograph.

A very famous spatial filtering experiment is 'matched filtering', which can be used for pattern recognition. The original matched filters were classical Fourier holograms of the target pattern. Computer hologram's can do the same job, even with spatially incoherent objects, this work being initiated by Katyl.

Fourier holograms are also attractive for data storage, due to their robustness against local defects. However, a local error in the Fourier domain spreads out all over the image domain. The computed hologram has an advantage over classical holograms because of the freedom to incorporate any error detection and correction codes for reducing the bit error rate even further.

One of the most recent advances is in matched filtering with totally incoherent light. This step allows moving spatial filtering out of the well-guarded laboratory into hostile outdoor environments.

Finally, a word about terminology; several terms are used in place of 'CGH': synthetic hologram, digital hologram, holographic optical element (HOE), diffractive optical element (DOE) or digital optical element.

CGHs have also been made for other waves, such as acoustical waves, ocean waves, and microwaves. Digital antenna steering represents a case of computer holography.

See also

Diffraction: Fraunhofer Diffraction; Fresnel Diffraction. **Fourier Optics**. **Holography, Techniques:** Computer-Generated Holograms; Holographic Interferometry; Overview. **Imaging:** Information Theory in Imaging; Inverse Problems and Computational Imaging; Three-Dimensional Field Transformations. **Information Processing:** Coherent Analogue Optical Processors; Free-Space Optical Computing; Incoherent Analog Optical Processors. **Optical Communication Systems:** Free-Space Optical Communications. **Optical Processing Systems**. **Phase Control:** Wavefront Coding.

Further Reading

Andrés P (1999) *Opt. Lett.* 245: 1331.
Brown BR and Lohmann AW (1966) *Appl. Opt.* 5: 967.
Brown BR and Lohmann AW (1969) *IBM J. R&D* 13: 160.
Bryngdahl O (1973) *Progr. Opt.* 11: 168.
Bryngdahl O and Wyrowski F (1990) *Progr. Opt.* 28: 1–86.
Dallas WJ (1971) *Appl. Opt.* 10: 673.
Dallas WJ (1973) *Appl. Opt.* 12: 1179.
Dallas WJ (1980) *Topics in Appl. Phys.* 40: 291–366.
Dallas WJ (1999–2003) http://www.radiology.arizona.edu/dallas/CGH.htm. University of Arizona, Tucson, AZ.
Dresel T, Beyerlein M and Schwider J (1996) *Appl. Opt.* 35: 4615.
Fienup JR (1981) *Proc. SPIE* 373: 147.
Fienup JR (1982) *Appl. Opt.* 21: 2758.
Gerchberg RW and Saxton WO (1972) *Optik* 35: 237.
Goodman JW and Silvestri AM (1970) *IBM J. R&D* 14: 478.

Hansler RL (1968) *Appl. Opt.* 7: 1863.
Hauk D and Lohmann AW (1958) *Optik* 15: 275.
Ichioka Y and Lohmann AW (1972) *Appl. Opt.* 11: 2597.
Katyl RH (1972) *Appl. Opt.* 11: 1255.
Kozma A and Kelly DL (1965) *Appl. Opt.* 4: 387.
Lee WH (1978) *Progr. Opt.* 16: 219–232.
Lesem LB, Hirsch PM and Jordan JA (1969) *IBM J. R&D* 13: 150.
Lohmann AW and Paris DP (1967) *Appl. Opt.* 6: 1746.
Lohmann AW and Paris DP (1968) *Appl. Opt.* 7: 651.
Lohmann AW and Sinzinger S (1995) *Appl. Opt.* 34: 3172.
Lohmann AW and Werlich HW (1971) *Appl. Opt.* 10: 670.
McGovern AJ and Wyant JC (1971) *Appl. Opt.* 10: 619.
Pe'er A, Wang D, Lohmann AW and Friesem AA (1999) *Opt. Lett.* 24: 1469; (2000) 25: 776.
Sinzinger S and Jahns J (1999) *Micro Optics*. New York: Wiley.
Vander Lugt AB (1964) *IEEE, IT* 10: 139.
Wyrowski F and Bryngdahl O (1991) *Rep. Prog. Phys.* 1481–1571.
Yaroslavskii LP and Merzlyakov NS (1980) *Methods in Digital Holography*. New York: Plenum Press.

Digital Holography

W Osten, University of Stuttgart, Stuggart, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The sinusoidal nature of light, its ability to interfere, and the precise knowledge of its wavelength make it possible to measure several metrological quantities, such as distances and displacements, with high accuracy. However, the high frequency of light results in the difficulty that the primary quantity to be measured, i.e., the phase of the light wave, cannot be observed directly. All quadratic sensors (CCD, CMOS, photo plates) are only able to record intensities. Therefore, one makes use of a trick: the transformation of phase changes into recordable intensity changes as the basic principle of interferometry and holography. Holography provides the possibility to record and to reconstruct the complete information of optical wave fields. Numerous potential applications such as 3D-imaging of natural objects, fabrication of diffractive optical elements, and interferometric investigation of complex technical components become possible. Further progress has been made by recording holograms directly with electronic targets and not in photographic emulsions. In this way the independent numerical reconstruction of phase and intensity can be accomplished by the computer. One important consequence is that interferometric techniques, such as 3D-displacement analysis and 3D-contouring, can be implemented easily and with high flexibility. Furthermore, digitally recorded holograms can be transmitted via telecommunication networks and optically reconstructed at any desired place using an appropriate spatial light modulator. Consequently, the complete optical information of complex objects can be transported between different places with high velocity. This opens the possibility to compare nominal identical but physically different objects (master sample comparison) available at different locations with interferometric sensitivity. We call this remote metrology.

Direct Phase Reconstruction by Digital Holography

In digital speckle pattern interferometry (DSPI), the object is focused onto the target of an electronic sensor. Thus, an image plane hologram is formed as a result of the interference with an inline reference wave. In contrast with DSPI, a digital hologram is recorded without imaging. The target records the superposition of the reference and the object wave in the near-field region – a so-called Fresnel hologram. The basic optical setup in digital holography for recording holograms is the same as in conventional holography (**Figure 1a**). A laser beam is divided into two coherent beams, one illuminating the object and forming the object wave, the other entering the target directly and forming the reference wave. On this basis, very compact solutions are possible. **Figure 1b** shows a holographic camera where the interferometer, containing a beamsplitter and some wavefront shaping components, is mounted in front of the camera target.

For a description of the principle of digital Fresnel holography, we use a simplified version of the optical setup (**Figure 2a**). The object is modeled by a plane rough surface that is located in the (x, y)-plane and illuminated by laser light. The scattered wavefield forms the object wave $u(x, y)$. The target of an electronic sensor (e.g., a CCD or a CMOS) used for recording the hologram is located in the (ξ, η)-plane at the distance d from the object. Following the basic principles of holography, the hologram $h(\xi, \eta)$ originates from the interference of the object wave $u(\xi, \eta)$ and the reference wave $r(\xi, \eta)$

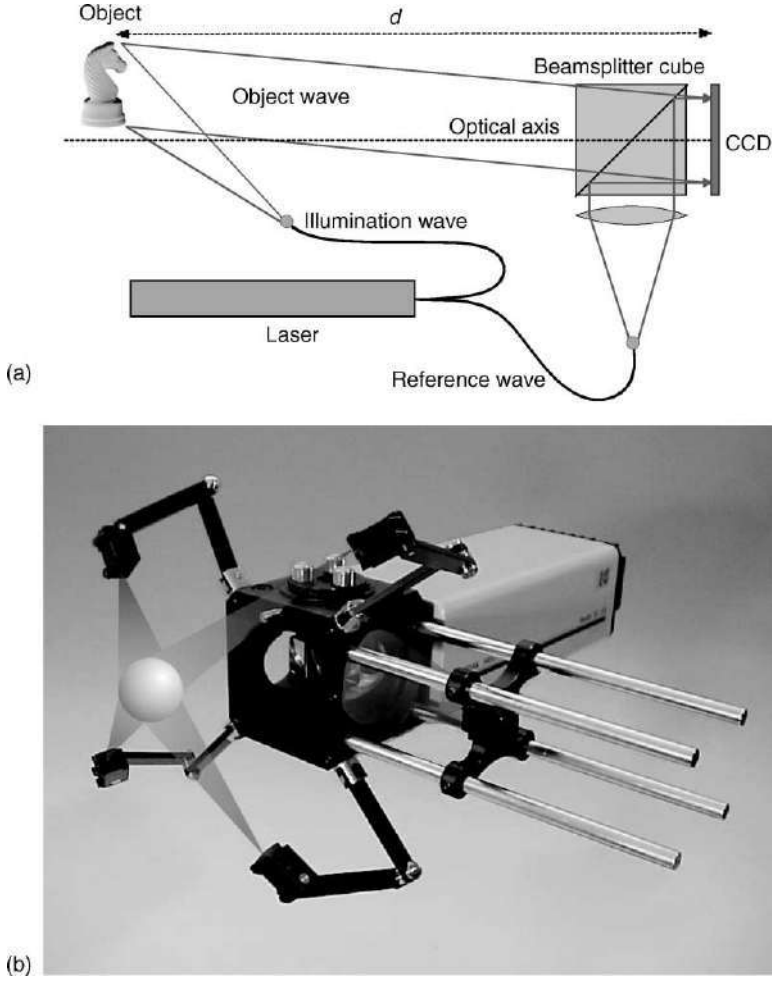

Figure 1 Setup for recording digital holograms. (a) Schematic setup for recording a digital hologram onto a CCD target. (b) Implementation of a holographic camera by mounting a compact holographic interferometer in front of a CCD-target (4 illumination directions are used). Reproduced from Seebacher S, Osten W, Baumbach Th and Jüptner W (2001) The determination of material parameters of microcomponents using digital holography. *Optics and Lasers in Engineering* 36(2): 103–126, copyright © 2001, with permission from Elsevier.

in the (ξ, η)-plane:

$$h(\xi, \eta) = |u(\xi, \eta) + r(\xi, \eta)|^2$$
$$= rr^* + ru^* + ur^* + uu^* \qquad [1]$$

The transformation of the intensity distribution into a gray value distribution, that is stored in the image memory of the computer, is considered by the characteristic function t of the sensor. This function is generally only approximately linear:

$$T = t[h(\xi, \eta)] \qquad [2]$$

Because the sensor has a limited spatial resolution, the spatial frequencies of the interference fringes in the hologram plane – the so-called micro-interferences – have to be considered. The fringe spacing g and the spatial frequency f_x, respectively, are determined by the angle β between the object and the reference wave (Figure 2b):

$$g = \frac{1}{f_x} = \frac{\lambda}{2\sin(\beta/2)} \qquad [3]$$

with λ as the wavelength. If we assume that the discrete sensor has a pixel pitch (distance between adjacent pixels) $\Delta\xi$, the sampling theorem requires at least two pixels per fringe for a correct reconstruction of the periodic function:

$$2\Delta\xi < \frac{1}{f_x} \qquad [4]$$

Consequently, we obtain for small angles β:

$$\beta < \frac{\lambda}{2\Delta\xi} \qquad [5]$$

Modern high-resolution CCD or CMOS chips have a pitch, $\Delta\xi$, of about 4 μm. In this case, a maximum angle between the reference and the

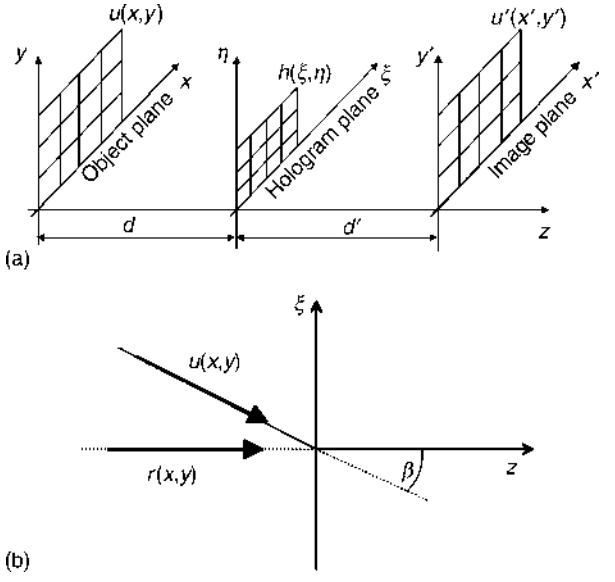

Figure 2 Geometry for recording and reconstruction of digital holograms. (a) Schematic setup for Fresnel holography. (b) Interference between the object and reference wave in the hologram plane.

object wave of only 4° is acceptable. The practical consequence of the restricted angle resolution in digital holography is a limitation of the effective object size that can be stored holographically by an electronic sensor. However, this is only a technical handicap. Larger objects can be placed at a sufficient distance from the hologram, or reduced optically by imaging with a negative lens.

The reconstruction is done by illuminating the hologram with a so-called reconstruction wave $c(\xi, \eta)$:

$$u'(x', y') = t[h(\xi, \eta)]c(\xi, \eta) \qquad [6]$$

The hologram, $t[h(\xi, \eta)]$, diffracts the wave, $c(\xi, \eta)$, in such a way that images of the object wave are reconstructed. Usually four terms are reconstructed if the wave, $u'(x', y')$, propagates in space. An assumed linear characteristic function, $t(h) = \alpha h + t_0$, delivers:

$$u' = Tc = t(h)c \qquad [7]$$

and

$$u' = \alpha\lfloor cu^2 + cr^2 + cur^* + cru^* \rfloor + ct_0 \qquad [8]$$

with two relevant image terms, $[cur^*]$ and $[cru^*]$. The appearance of the image terms depends on the concrete shape of the reconstruction wave. Usually the reference wave, $c = r$, or its conjugated version, $c = r^*$, is applied. In the case of the conjugated reference wave, a direct or real image will be reconstructed due to a converging image wave that can be imaged on a screen at the location of the original object.

However, in digital holography the reconstruction of the object wave in the image plane, $u'(x', y')$, is done by numerical simulation of the physical process, as shown in Figure 3a. The reconstruction wave with a well-known shape equal to the reference wave, $r(\xi, \eta)$, propagates through the hologram, $h(\xi, \eta)$. Following the Huygens principle, each point, $P(\xi, \eta)$, on the hologram, acts as the origin of a spherical elementary wave. The intensity of these elementary waves is modulated by the transparency, $h(\xi, \eta)$. In a given distance, $d' = d$, from the hologram, a sharp real image of the object can be reconstructed as the superposition of all elementary waves. For the reconstruction of a virtual image, $d' = -d$ is used.

Consequently, the calculation of the wavefield, $u'(x', y')$, in the image plane starts with the pointwise multiplication of the stored and transformed intensity values, $t[h(\xi, \eta)]$, with a numerical model of the reference wave, $r(\xi, \eta)$. If the hologram is assumed to be uniformly illuminated by a normally incident, monochromatic plane wave of amplitude 1 the reference wave can be modeled by $r(\xi, \eta) = 1$. After the multiplication the resulting field in the hologram plane is propagated in free space. In the distance, d', the diffracted field, $u'(x', y')$, can be found by solving the well-known Rayleigh–Sommerfeld diffraction formula, that is also known as the

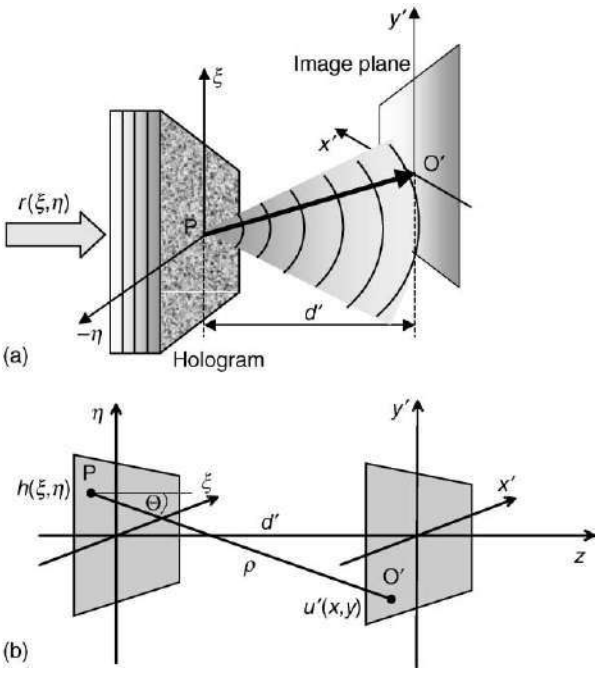

Figure 3 Reconstruction of digital holograms. (a) Principle of wave front reconstruction. (b) Light propagation by diffraction (Huygens–Fresnel principle).

Huygens–Fresnel principle:

$$u'(x',y') = \frac{1}{\mathrm{i}\lambda}\int\int_{-\infty\cdots\infty} t[h(\xi,\eta)]r(\xi,\eta) \times \frac{\exp(\mathrm{i}k\rho)}{\rho}\cos\theta\,\mathrm{d}\xi\,\mathrm{d}\eta \qquad [9]$$

with

$$\rho(\xi-x',\eta-y') = \sqrt{d'^2+(\xi-x')^2+(\eta-y')^2} \qquad [10]$$

as the distance between a point $\mathrm{O}'(x',y',z=d')$ in the image plane and a point $\mathrm{P}(\xi,\eta,z=0)$ in the hologram plane, and

$$k = \frac{2\pi}{\lambda} \qquad [11]$$

as the wave number. The obliquity factor, $\cos\theta$, represents the cosine of the angle between the outward normal and the vector joining P to O′. This term is given as:

$$\cos\theta = \frac{d'}{\rho} \qquad [12]$$

and therefore the Huygens–Fresnel principle can be rewritten:

$$u'(x',y') = \frac{d'}{\mathrm{i}\lambda}\int\int_{-\infty\cdots\infty} t[h(\xi,\eta)]r(\xi,\eta)\frac{\exp(\mathrm{i}k\rho)}{\rho^2}\mathrm{d}\xi\mathrm{d}\eta \qquad [13]$$

This numerical reconstruction delivers the complex amplitude of the wavefront. Consequently, the phase distribution, $\phi(x',y')$, and the intensity, $I(x',y')$, can be calculated directly from the reconstructed complex function, $u'(x',y')$:

$$\phi(x',y') = \arctan\frac{\mathrm{Im}|u'(x',y')|}{\mathrm{Re}|u'(x',y')|} \qquad [-\pi,\pi] \qquad [14]$$

$$I(x',y') = u'(x',y')u'^*(x',y') \qquad [15]$$

The direct approach to the phase yields several advantages for imaging and metrology applications that are discussed later.

Reconstruction Principles in Digital Holography

Different numerical reconstruction principles have been investigated: the Fresnel approximation, the convolution approach, the lens-less Fourier approach, the phase shifting approach, and the phase-retrieval approach. In this section the main techniques are briefly described.

The Fresnel Approximation

If the distance, d, between the object and hologram plane, and equivalently, $d' = d$, between the hologram and image plane, is large compared to $(\xi - x')$ and $(\eta - y')$, then the denominator of eqn [13] can be replaced by d'^2 and the parameter ρ in the numerator can be approximated by a binomial expansion for the square root (10) where only the first two terms are considered:

$$\rho \approx d'\left[1 + \frac{(\xi-x')^2}{2d'^2} + \frac{(\eta-x')^2}{2d'^2}\right] \qquad [16]$$

The resulting expression for the field at (x',y') becomes

$$u'(x',y') = \frac{\exp(\mathrm{i}kd')}{\mathrm{i}\lambda d'}\int\int_{-\infty\cdots\infty} t[h(\xi,\eta)]r(\xi,\eta) \times \exp\left[\frac{\mathrm{i}k}{2d'}\left\{(\xi-x')^2+(\eta-y')^2\right\}\mathrm{d}\xi\,\mathrm{d}\eta\right] \qquad [17]$$

Equation [17] is a convolution integral that can be expressed as

$$u'(x',y) = \int\int_{-\infty\cdots\infty} t[h(\xi,\eta)]r(\xi,\eta)H \times (\xi-x',\eta-y')\mathrm{d}\xi\,\mathrm{d}\eta \qquad [18]$$

with the convolution kernel:

$$H(x',y') = \frac{\exp(\mathrm{i}kd')}{\mathrm{i}\lambda d'}\exp\left[\frac{\mathrm{i}k}{2d'}(x'^2+y'^2)\right] \qquad [19]$$

Another notation of eqn [17] is found if the term $\exp[(\mathrm{i}k/2d')(x'^2+y'^2)]$ is taken out of the integral:

$$u'(x',y') = \frac{\exp(\mathrm{i}kd')}{\mathrm{i}\lambda d'}\exp\left[\mathrm{i}\frac{k}{2d'}(x'^2+y'^2)\right]\int\int_{-\infty\cdots\infty} \times\left\{t[h(\xi,\eta)]r(\xi,\eta)\exp\left[\mathrm{i}\frac{k}{2d'}(\xi^2+\eta^2)\right]\right\} \times\exp\left[-\mathrm{i}\frac{2\pi}{\lambda d'}(\xi x'+\eta y')\right]\mathrm{d}\xi\,\mathrm{d}\eta \qquad [20]$$

or

$$u'(x',y') = \frac{\exp(\mathrm{i}kd')}{\mathrm{i}\lambda d'}\exp\left[\mathrm{i}\frac{\mathrm{k}}{2d'}(x'^2+y'^2)\right] \times\mathrm{FT}_{\lambda d'}\left\{t[h(\xi,\eta)]r(\xi,\eta)\exp\left[\mathrm{i}\frac{k}{2d'}(\xi^2+\eta^2)\right]\right\} \qquad [21]$$

where $\mathrm{FT}_{\lambda d'}$ indicates the 2-dimensional Fourier transform that has been modified by a factor $1/(\lambda d')$. Equation [21] makes clear that the diffracted

wavefront consists of the Fourier transform (FT) of the digitally stored hologram, $t[h(\xi,\eta)]$, multiplied by the reference wave, $r(\xi,\eta)$, and the chirp function, $\exp\{(i\pi/\lambda d')(\xi^2+\eta^2)\}$. This Fourier transform is scaled by the constant factor, $1/(id'\lambda)$, and a phase factor that is independent of the processed hologram.

The discrete sampling in digital processing requires the transformation of the infinite continuous integral in eqn [20] into a finite discrete sum. This results in the finite Fresnel transform:

$$u'(m,n) = \sum_{j=0}^{M-1}\sum_{l=0}^{N-1} t[h(j\Delta\xi, l\Delta\eta)]r(j\Delta\xi, l\Delta\eta) \times \exp\left[\frac{i\pi}{d'\lambda}(j^2\Delta\xi^2 + l^2\Delta\eta^2)\right] \times \exp\left\{-i2\pi\left(\frac{jm}{M}+\frac{ln}{N}\right)\right\} \quad [22]$$

where constants and pure phase factors preceding the sums have been omitted. The main parameters are the pixel number $M\times N$ and the pixel pitches $\Delta\xi$ and $\Delta\eta$ in the two directions, which are defined by the used sensor chip.

The discrete phase distribution, $\phi(m,n)$, of the wavefront and the discrete intensity distribution, $I(m,n)$, on the rectangular grid of $M\times N$ sample points can be calculated from the reconstructed complex function, $u'(m,n)$, by using eqns [14] and [15], respectively. The wrapped phase distribution is computed directly without the need of additional temporal or spatial phase modulation such as phase shifting or spatial heterodyning.

For interferometric applications with respect to displacement and shape measurement the phase difference, $\delta(m,n)$, of two reconstructed wave fields, $u_1'(m,n)$ and $u_2'(m,n)$, needs to be calculated. The algorithm can be reduced to the following equation:

$$\delta(m,n) = \phi_1(m,n) - \phi_2(m,n) \quad [23]$$

Figure 4 shows the principle of displacement measurement by digital holography. Besides the investigation of surface displacements in the region of the wavelength, digital holography can be applied advantageously for the measurement of the shape of complex objects. In Figure 5, a turbine blade was investigated with the 2-wavelength contouring method. Both, the digitally reconstructed mod 2π-phase and its demodulated version after-phase unwrapping, are presented. Phase unwrapping is necessary in digital holography, since the fringe-counting problem still remains. However, there are some efficient approaches to overcome the difficulties of this process.

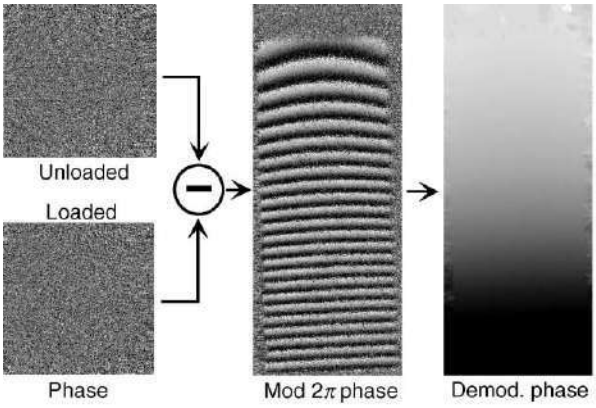

Figure 4 The principle of digital holographic interferometry demonstrated on an example of a loaded small beam. The interference phase is the result of the subtraction of the two phases which correspond to the digital holograms of both object states.

Following eqn [22] the pixel spacing in the reconstructed image is

$$\Delta x' = \frac{d'\lambda}{M\Delta\xi} \quad \text{and} \quad \Delta y' = \frac{d'\lambda}{N\Delta\eta} \quad [24]$$

Additional to the real image, a blurred virtual image and a bright dc-term, the zero-order diffracted field, is reconstructed. This term can effectively be eliminated by preprocessing the stored hologram or the different terms can be separated by using the off-axis instead of the in-line technique. However, the spatial separation between the object and the reference field requires a sufficient space-bandwidth product of the used CCD-chip, as discussed in the section on direct phase reconstruction above.

Numerical Reconstruction by the Convolution Approach

In the section above, we have already mentioned that the connection between the image term, $u'(x',y')$, and the product, $t[h(\xi,\eta)]r(\xi,\eta)$, can be described by a linear space-invariant system. The diffraction formula [17] is a superposition integral with the impulse response:

$$H(x'-\xi, y'-\eta) = \frac{\exp(ikd')}{i\lambda d'} \times \exp\left[\frac{ik}{2d'}\{(\xi-x')^2+(\eta-y')^2\}\right] \quad [25]$$

A linear space-invariant system is characterized by a transfer function G, that can be calculated as the

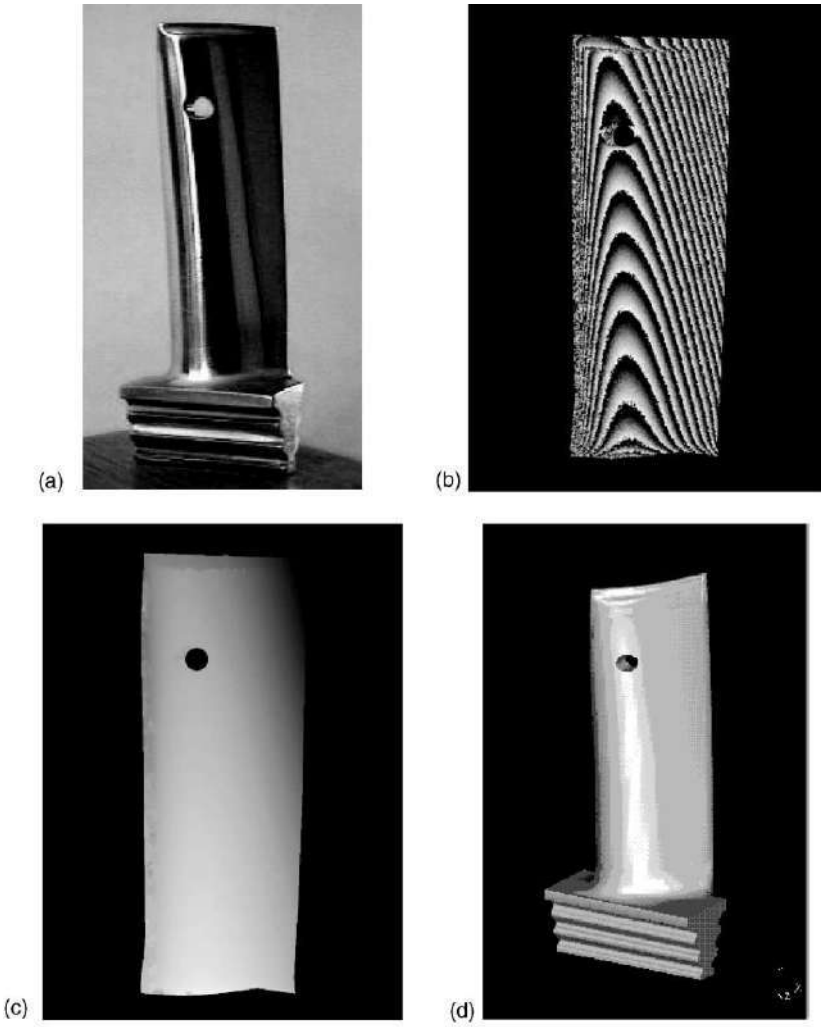

Figure 5 Phase reconstruction by digital holographic interferometry on an example of 2-wavelength contouring of a turbine blade. (a) Image of the turbine blade. (b) Mod-2π phase distribution. (c) Unwrapped phase distribution. (d) CAD reconstruction of the turbine blade.

Fourier-transform of the impulse response:

$$G(f_x, f_y) = \mathrm{FT}\{H(\xi, \eta, x', y')\} \qquad [26]$$

with the spatial frequencies f_x and f_y. Consequently, the convolution theorem can be applied, which states that the Fourier transform of the convolution, $t[h(\xi, \eta)]r(\xi, \eta)$, with H is the product of the individual Fourier transforms, $\mathrm{FT}\{t[h(\xi, \eta)]r(\xi, \eta)\}$ and $\mathrm{FT}\{H\}$. Thus $u'(x', y')$ can be calculated by the inverse Fourier transform of the product of the Fourier-transformed convolution partners:

$$u'(x', y') = \mathrm{FT}^{-1}\{\mathrm{FT}[t(h)r]\mathrm{FT}[H]\} \qquad [27]$$

$$u'(x', y') = [t(h)r] \otimes H \qquad [28]$$

with $\otimes$ as the convolution symbol. The computing effort is comparatively high: two complex multiplications and three Fourier transforms. The main difference to the Fresnel transform is the different pixel size in the reconstructed field:

$$\Delta x' = \Delta\xi \quad \text{and} \quad \Delta y' = \Delta\eta \qquad [29]$$

Numerical Reconstruction by the Lensless Fourier Approach

We have already mentioned that the limited spatial resolution of the sensor restricts the angle resolution of the digital hologram. The sampling theorem requires that the angle between the object beam and the reference beam at any point of the electronic sensor be limited in such a way that the

microinterference spacing is larger than double the pixel size. In general, the angle between the reference beam and the object beam varies over the sensor's surface, and so does the maximum spatial frequency. Thus, for most holographic setups, the full spatial bandwidth of the sensor cannot be used. However, it is important to use the entire spatial bandwidth of the sensor because the lateral resolution of the reconstructed image depends on a complete evaluation of the entire information obtained from the sensor.

Even if the speed of digital signal processing is increasing rapidly, algorithms should be as simple and as fast to compute as possible. For the Fresnel approach and the convolution approach, several fast Fourier transforms and complex multiplications are necessary. Therefore, a more effective approach, such as the subsequently described algorithm, seems promising.

The lensless Fourier approach is the fastest and most suitable algorithm for small objects. The corresponding setup is shown in Figure 6. It allows us to choose the lateral resolution in a range from a few microns to hundreds of microns without any additional optics. Each point, (ξ, η), on the hologram is again considered as a source point of a spherical elementary wavefront (Huygen's principle). The intensity of these elementary waves is modulated by $t[h(\xi, \eta)]$ – the amplitude transmission of the hologram. The reconstruction algorithm for lensless Fourier holography is based on the Fresnel reconstruction. Here again, $u(x, y)$ is the object wave in the object plane, $h(\xi, \eta)$ the hologram, $r(\xi, \eta)$ the reference wave in the hologram plane, and $u'(x', y')$ the reconstructed wavefield.

For the specific setup of lensless Fourier holography, a spherical reference wave is used with an origin at the same distance from the sensor as the object itself. In the case that $d' = -d$, $x = x'$, and $y = y'$, both virtual images are reconstructed and the reconstruction algorithm then reads:

$$u'(x', y') = \frac{\exp(\mathrm{i}kd)}{\mathrm{i}\lambda d}\, \mathrm{e}^{-\mathrm{i}\frac{k}{2d}(x^2+y^2)} \times \mathrm{FT}_{\lambda d'}\left\{t[h(\xi,\eta)]\cdot r(\xi,\eta)\cdot \mathrm{e}^{-\mathrm{i}\frac{k}{2d}(\xi^2+\eta^2)}\right\} \quad [30]$$

with $\mathrm{FT}_{\lambda d}$ as the 2-dimensional Fourier transformation which has been scaled by a factor $1/\lambda d$. In this recording configuration, the effect of the spherical phase factor associated with the Fresnel diffraction pattern of the object is eliminated by the use of a spherical reference wave, $r(\xi, \eta)$, with the same average curvature:

$$r(x, y) = \text{const.} \times \exp\left(\mathrm{i}\frac{\pi}{\lambda d}(\xi^2 + \eta^2)\right) \quad [31]$$

This results in a more simple reconstruction algorithm which can be described by

$$u'(x', y') = \text{const.} \times \exp\left(-\mathrm{i}\frac{\pi}{\lambda d}(x^2 + y^2)\right) \times \mathrm{FT}_{\lambda d}\{h(\xi, \eta)\} \quad [32]$$

Besides the simpler and faster reconstruction algorithm (only one Fourier transformation has to be computed), the Fourier algorithm uses the full space-bandwidth product (SBP) of the sensor chip because it adapts the curvature of the reference wave to the curvature of the object wave.

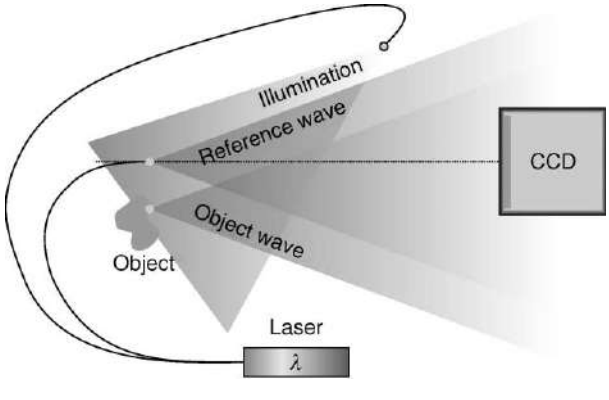

Figure 6 Scheme of a setup for digital lensless Fourier holography of diffusely reflecting objects.

Influences of Discretization

For the estimation of the lateral resolution in digital holography, three different effects related to discretization have to be considered: averaging, sampling, and the limited sensor size. We assume a quadratic sensor with $N \times M$ pixels of the size $\Delta\xi \times \Delta\xi$. Each pixel has a light-sensitive region with a side length $\gamma\Delta\xi = \Delta\xi_{\mathrm{eff}}$. γ^2 is the so-called fill factor $(0 \le \gamma \le 1)$, that indicates the active area of the pixel. The sensor averages the incident light which has to be considered because of possible intensity fluctuations over this area. The continuous expression for the intensity $I(\xi, \eta) = t[h(\xi, \eta)]$, registered by the sensor, has to be integrated over the light-sensitive area. This can be expressed mathematically by the convolution of the incident intensity $I(\xi, \eta)$, with a rectangle function:

$$I^1(\xi, \eta) \propto I \otimes \mathrm{rect}_{\Delta\xi_{\mathrm{eff}}\cdot\Delta\xi_{\mathrm{eff}}} \quad [33]$$

The discrete sampling of the light field is modeled by the multiplication of the continuous assumed hologram with the 2-dimensional comb-function:

$$I^2(\xi, \eta) \propto I^1(\xi, \eta) \times \mathrm{comb}_{\Delta\xi.\Delta\xi} \qquad [34]$$

Finally, the limited sensor size requires that the comb-function has to be truncated at the borders of the sensor. This is considered by the multiplication with a 2-dimensional rect-function of the size $N \times \Delta\xi$:

$$I^3(\xi, \eta) \propto I^2 \times \mathrm{rect}_{N\Delta\xi N.\Delta\xi} \qquad [35]$$

The consequences are amplitude distortions, aliasing and speckling that have to be considered in the reconstruction procedure.

Advantages of Digital Holography

Besides the electronic processing and the direct access to the phase, some further advantages recommend digital holography for several applications such as metrology and microscopy:

- The availability of the independently reconstructed phase distributions of each individual state of the object and interferometer, respectively, offers the possibility to record a series of digital holograms with increased load amplitude. In the evaluation process, the convenient states can be compared interferometrically. Furthermore, a series of digital holograms with increasing load can be applied to unwrap the mod 2π-phase temporally. In this method, the total object deformation is subdivided into many measurement steps, in which the phase differences are smaller than 2π. By adding up those intermediate results, the total phase change can be obtained without any further unwrapping. This is an important feature, since it is essential to have an unwrapped phase to be able to calculate the real deformation data from the phase map. Figure 7 shows an example of such a measurement. The left image shows the wrapped deformation phase for a clamped coin which was loaded with heat. The right image shows the temporal unwrapped phase which has been obtained by dividing the total deformation into 85 sub-measurements. Thus, the displacement of the object can be observed almost in real time during the loading process.
- The independent recording and reconstruction of all states also gives a new degree of freedom for optical shape measurement. In the case of multiwavelength contouring, each hologram can be stored and reconstructed independently with its corresponding wavelength. This results in a drastic decrease of aberrations and makes it possible to use larger wavelength differences for the generation of shorter synthetic wavelengths.
- Because all states of the inspected object can be stored and evaluated independently, only seven digital holograms are necessary to measure the 3D-displacement field and shape of an object under test: one hologram for each illumination direction before and after the loading, respectively, and one hologram with a different wavelength (or a different source-point of illumination) which can interfere with one of the other holograms to perform two-wavelength-contouring (or two-source-point-contouring) for shape measurement. If four illumination directions are used, nine holograms are necessary.
- The direct access to all components of the wavefield makes it possible to correct wavefront aberrations effectively by computer.
- The digital hologram containing all information of the wavefront can be transmitted via the internet to any location and can be reconstructed digitally by computer or as an analogue version by a spatial light modulator. Consequently, remote access to wavefronts, that are generated at distant places, is possible.
- Finally, the possibility to miniaturize complex holographic sensor setups and to use this method for remote comparative interferometry, makes digital holography a versatile tool for the solution of numerous inspection and measurement problems.

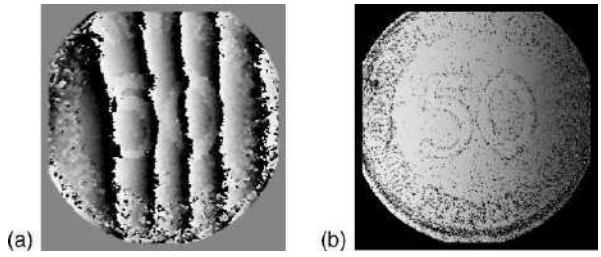

Figure 7 Temporal unwrapping by digital holography: (a) Wrapped phase of a thermal loaded coin. (b) Temporally unwrapped phase.

See also

Diffraction: Fresnel Diffraction. **Fourier Optics**. **Holography, Techniques:** Holographic Interferometry.

Further Reading

Butters JN and Leendertz JA (1971) Holographic and video techniques applied to engineering measurement. *Journal of Measurement Control* 4: 349–354.

Creath K (1993) Temporal phase measurement methods. In: Robinson DW and Reid GT (eds) *Interferogram Analysis*. Bristol and Philadelphia: IOP Publishing Ltd.

Demetrakopoulos T and Mittra T (1974) Digital and optical reconstruction of images from suboptical diffraction patterns. *Applied Optics* 13: 665–670.

Denisyuk YuN (1962) Photographic reconstruction of the optical properties of an object in its own scattered radiation field. *Doklady Akademii Nauk SSSR* 144: 1275–1279.

Gabor D (1949) Microscopy by reconstructed wavefronts. *Proceedings of the Royal Society, A* 197: 454–487.

Goodman JW (1996) *Introduction to Fourier Optics*. New York: McGraw Hill Comp. Inc.

Goodman JW and Lawrence RW (1967) Digital image formation from electronically detected holograms. *Applied Physics Letters* 11: 77–79.

Kolenovic E, Osten W, Klattenhoff R, Lai S, Kopylow C and Jüptner W (2003) Miniaturized digital holography sensor for distal three-dimensional endoscopy. *Applied Optics* 42(25): 5167–5172.

Kreis T (1996) *Holographic Interferometry*. Berlin: Akademie Verlag.

Kreis Th and Jüptner W (1997) Principles of digital holography. In: Jüptner W and Osten W (eds) *Proc. Fringe'97, 3rd International Workshop on Automatic Processing of Fringe Patterns*, pp. 353–363. Berlin: Akademie Verlag.

Kreis Th and Jüptner W (1997) The suppression of the dc-term in digital holography. *Optical Engineering* 36: 2357–2360.

Kujawinska M (1993) Spatial phase measurement methods. In: Robinson DW and Reid GT (eds) *Interferogram Analysis*. Bristol and Philadelphia: IOP Publishing Ltd.

Lee W-H (1978) Computer generated holograms: techniques and applications. In: Wolf E (ed.) *Progress in Optics*, vol. XVI, pp. 121–230. Amsterdam: North Holland.

Leith EN and Upatnieks J (1961) New techniques in wavefront reconstruction. *JOSA* 51: 1469–1473.

Osten W (2000) Holography and virtual 3D-testing. Proc. International Berlin Workshop HoloMet 2000, New Prospects of Holography and 3D-Metrology, *Strahltechnik* 14: 14–17.

Osten W (2003) Active metrology by digital holography. *Proceedings of SPIE* 4933: 96–110.

Osten W and Kujawinska M (2000) Active phase measuring metrology. In: Rastogi PK and Inaudi D (eds) *Trends in Optical Nondestructive Testing and Inspection*, pp. 45–69. Amsterdam: Elsevier Science B.V.

Osten W, Jüptner W and Seebacher S (1998) *The Qualification of Large Scale Approved Measurement Techniques for the Testing of Microcomponents*. Proc. 18th Symposium on Seebacher, Experimental Mechanics of Solids, Jachranka, pp. 43–55.

Osten W, Baumbach T, Seebacher S and Jüptner W (2001) Remote shape control by comparative digital holography. *Proc. Fringe 2001*, Bellingham: Elsevier Science, pp. 373–382.

Osten W, Seebacher S and Jüptner W (2001) The application of digital holography for the inspection of microcomponents. *Proceedings of SPIE* 4400: 1–15.

Osten W, Baumbach T and Jüptner W (2002) Comparative digital holography. *Optics Letters* 27(20): 1764–1766.

Pedrini G, Schedin S and Tiziani HJ (2001) Aberration compensation in digital holographic reconstruction of microscopic objects. *Journal of Modern Optics* 48: 1035–1041.

Powell RL and Stetson K (1965) Interferometric vibration analysis of three-dimensional objects by wavefront reconstruction. *JOSA* 55: 1593–1598.

Schnars U (1994) Direct phase determination in hologram interferometry with use of digitally recorded holograms. *Journal of the Optical Society of America A* 11: 2011–2015.

Schnars U and Jüptner W (1994) Direct recording of holograms by a CCD target and numerical reconstruction. *Applied Optics* 33(2): 179–181.

Seebacher S (2001) Application of digital holography for 3d-shape and deformation measurement of micro components. PhD thesis, University of Bremen.

Seebacher S, Osten W and Jüptner W (1997) 3-D deformation analysis of micro-components using digital holography. *Proceedings of SPIE* 3098: 382–391.

Seebacher S, Osten W, Baumbach Th and Jüptner W (2001) The determination of material parameters of microcomponents using digital holography. *Optics and Lasers in Engineering* 36(2): 103–126.

Takeda M, Taniguchi K, Hirayama T and Kogho H (1996) Single transform Fourier–Hartley fringe analysis for holographic interferometry. In: Füzessy Z, Jüptner W and Osten W (eds) *Simulation and Experiment in Laser Metrology*, pp. 67–73. Berlin: Akademie Verlag.

Wagner C, Seebacher S, Osten W and Jüptner W (1999) Digital recording and numerical reconstruction of lensless Fourier holograms in optical metrology. *Applied Optics* 38(22): 4812–4820.

Wagner C, Osten W and Seebacher S (2000) Direct shape measurement by digital wavefront reconstruction and multi-wavelength contouring. *Optical Engineering* 39(1): 79–85.

Yamaguchi I and Zhang T (1979) Phase-shifting digital holography. *Optical Letters* 22: 1268–1270.

Yaroslavsky LP and Merzlyakov NS (1980) *Methods of Digital Holography*. New York: Consultants Bureau.

Zhang Y, Pedrini G, Osten W and Tiziani H (2003) Image reconstruction for in-line holography with the Yang-Gu algorithm. *Applied Optics* 42(32): 6452–6457.

Holographic Interferometry

P Rastogi, Swiss Federal Institute of Technology, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The development of holographic interferometry has redefined our perception of optical interferometry. The unique property of holographic interferometry to bring a wavefront, generated at some earlier time, stored in a hologram and released at a later time, to interfere with a comparison wavefront, has made possible the interferometric comparison of a rough surface, which is subject to stress, with its normal state. Assuming the surface deformations to be very small, the interference of the two speckled wavefronts forms a set of interference fringes that are indicative of the amount of displacement and deformation undergone by the diffuse object.

Holographic interferometry is an important and exciting area of research and development. It has established itself as a highly promising noninvasive technique in the field of optical metrology. Techniques within its folds for the measurement of static and dynamic displacements, topographic contours, and flow fields have been developed and demonstrated with success in a wide range of problems. Significant and extensive contributions to the development of holographic interferometry have been realized both from the theoretical and experimental perspectives. A wide range of procedures have been developed which make it not only possible to measure surface displacements and deformations of engineering structures to an accuracy of a fraction of a micrometer but also to detect material flaws and inhomogeneities having escaped the manufacturing process. The methodology of quantitative analysis of holographic interferograms has undergone extensive development during the last decade. This article provides a brief review of the holographic techniques with emphasis placed on bringing out its relevance in experimental mechanics and nondestructive testing. The series of books given in the Further Reading section at the end of this article treat the field in detail.

Basic Methods in Holographic Interferometry

The ability of holography to store a wavefront and release it for reconstruction at a later time offers access to the possibility to compare wavefronts, which have albeit existed at different times. There are several schemes that have been established to obtain the interferometric comparison of the wavefronts.

Double-Exposure Holographic Interferometry

In frozen form of wavefront comparison, two holograms of the object are recorded on a photographic plate. The first hologram is made with the object in its initial, unstressed state and the second hologram is recorded with the object in its final, stressed state. Upon reconstruction, the hologram releases the two stored wavefronts, one corresponding to the object in its unstressed state and the other corresponding to the object in its stressed state (Figure 1). The virtual image is overlaid by a set of

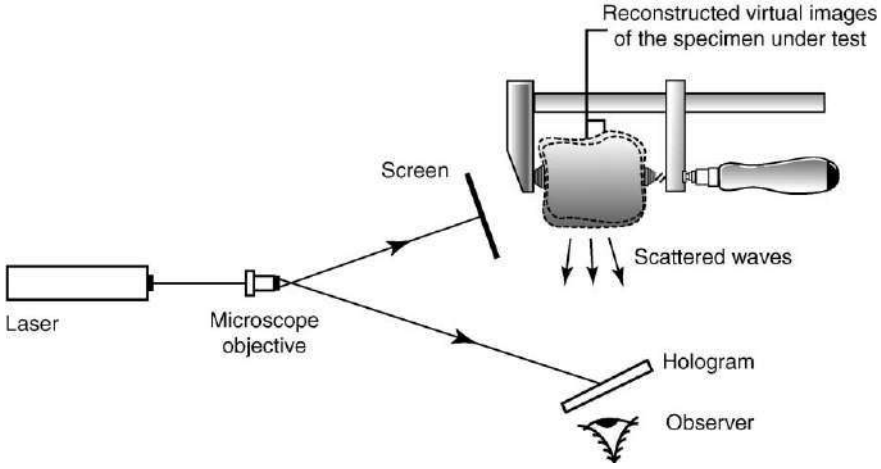

Figure 1 Layout for observing a double exposure holographic interferogram.

bright and dark interference fringes, which are due to the displacement of the object between the two exposures.

The fringe pattern is permanently stored on the holographic plate. The interference fringes denote the loci of points that have undergone an equal change of optical path between the light source and the observer. The image intensity of the holographic interferogram is

$$I(x, y) = I_0(x, y)\{1 + V(x, y)\cos(\varphi(x, y) - \varphi'(x, y))\} \quad [1]$$

where $I_0(x, y)$ is the background intensity, $V(x, y)$ is the fringe contrast, and φ and φ' are the phases of the waves from the object in its initial and deformed states, respectively.

A bright fringe is produced whenever

$$\varphi(x, y) - \varphi'(x, y) = 2n\pi \qquad n = 0, 1, 2, \ldots \quad [2]$$

where n is the fringe number. The wave scattered by a rough surface shows rapid variations of phase, which have little correlation across the wavefront. The phase difference varies randomly over the ray directions contained within the aperture of the viewing system, which would normally prevent the observation of fringes with large aperture optics near such points. However, a surface may exist at a distance from the object at which the value of phase difference is stationary over the cone of ray pairs defined by the viewing aperture. Interference fringes are localized in the region where the variation in phase difference is minimum over the range of viewing directions. This approach has been widely used to compute fringe localization for any object displacement and for any arbitrary illumination and observation geometry.

Real-Time Holographic Interferometry

A hologram is made of an arbitrarily shaped rough surface. After development, the hologram is placed back in exactly the same position. Upon reconstruction, the hologram produces the original wavefront. A person looking through the hologram sees a superposition of the original object and its reconstructed image (Figure 2). The object wave interferes with the reconstructed wave to produce a dark field due to destructive interference. If the object is now slightly deformed, interference fringes are produced which are related to the change in shape of the object. A dark fringe is produced, whenever:

$$\varphi(x, y) - \varphi'(x, y) = 2n\pi \qquad n = 0, 1, 2, \ldots \quad [3]$$

The method is very useful for determining the direction of the object displacement, and for compensating on the interferogram the influence of the rigid body motion of the object when subjected to a stress field.

The optical setup for real-time holography uses an immersion tank, which is mounted on a universal stage and contains the photographic plate. The holographic plate is followed by a vidicon camera, which visualizes the interferograms directly onto the TV monitor screen connected to a video-recording tape to memorize the information. The use of thermoplastic plates as recording material has gained in importance in recent years for implementing real-time and double exposure holographic interferometry setups, as it can be processed rapidly *in situ* using electrical and thermal processing.

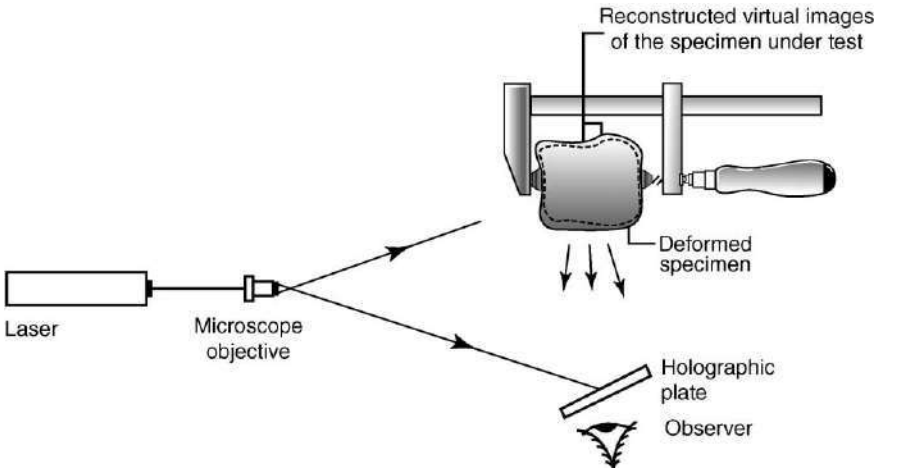

Figure 2 Layout for observing a holographic interferogram in real-time.

The plate is also cost-effective as it is erasable and reusable at least several hundred times. In thermoplastic recording the information is recorded as a thickness variation corresponding to a charge intensity pattern deposited on the thermoplastic layer. Self-developing or photorefractive crystals are other interesting alternatives to the use of film as the recording medium.

Digital Holographic Interferometry

The method consists of recording two holograms corresponding to undeformed and deformed states of an object using a CCD camera. These images are stored in a digital image processing system. The two holograms are then reconstructed separately by numerical methods from the digitaly stored holograms. The advantage of digital holography is that it allows for computing both the intensity and the phase of a holographically stored wavefront during the numerical reconstruction. The contours of constant optical phase change, due to deformation, are obtained by subtracting the reconstructed phases of the undeformed from the deformed object wavefield.

A hologram can be recorded on a CCD as long as the sampling theorem is fulfilled; that is to say, each period of the spatial variation of the hologram intensity is sampled by at least two pixels of the CCD array. The condition imposes a limit on the angle that the object beam can make with the reference beam, which must necessarily be small given that the resolution of CCDs is relatively low. The applications of the method are thus limited mainly by the pixel size in CCD array. The maximum spatial frequency $f_{\max}$ that must be resolved by the recording medium is determined by the maximum angle $\Delta\theta_{\max}$ between the reference and the object wave:

$$f_{\max} = \frac{2}{\lambda}\sin\frac{\Delta\theta_{\max}}{2} \qquad [4]$$

where λ is the wavelength of the laser source. Since the CCD cameras have resolutions of around 100 lines/mm, the maximum angle between the object and reference waves is limited to a few degrees.

Calculation of Phase Change on Object Loading

A diffusely reflecting object is deformed such that a point P on the surface moves to P′ as shown in Figure 3. In double-exposure or real-time holographic interferometry the two wavefronts related to the states of a surface before and after deformation are reconstructed simultaneously and compared. The change of optical path length due to displacement d is

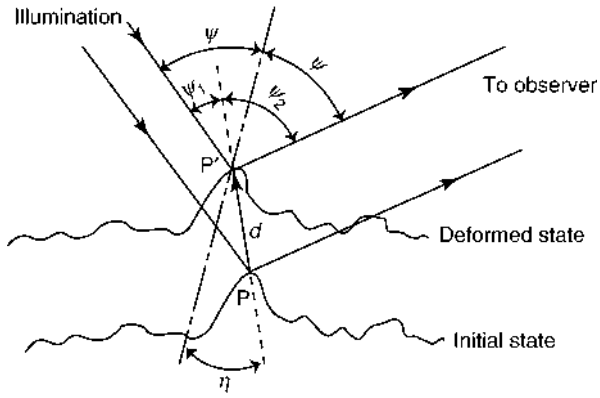

Figure 3 Calculation of the change in optical phase due to surface displacement.

$$\Delta\varphi = \frac{2\pi d}{\lambda}(\cos\psi_1 + \cos\psi_2) \qquad [5]$$

where ψ_1 and ψ_2 are the angles which the illumination and observation waves make with respect to the direction of displacement; eqn [5] can be written in the form:

$$\Delta\varphi = \frac{2\pi d}{\lambda}\cos\eta\cos\psi \qquad [6]$$

where ψ is the angle of bisection between the incident and viewing directions, and η is the angle which the bisector makes with the direction of displacement. The term $d\cos\eta$ being the resolved part of the displacement PP′ in the direction of the bisector, implies that the fringe pattern provides the measure of the displacement component along the bisector of the angle between the incident and viewing directions.

Digital Phase Measurement

The last decade has seen a rapid development of techniques for the automatic and precise reconstruction of phases from fringe patterns. These are based on the concepts of fringe tracking, Fourier transform, carrier frequency, and phase shift interferometry. Quantitative measurement of wavefront phase, for example, can be obtained in a few seconds using phase shift interferometry. With this technique one records a series of holograms by introducing artificially known steps of phase differences in the interference image in eqn [1]:

$$I_i(x,y) = I_0(x,y)\{1 + V(x,y)\cos(\Delta\varphi(x,y) + \Delta\psi_i)\};\quad i = 0, 1, 2, \ldots \qquad [7]$$

where $\Delta\psi_i$ is the phase shift which is produced, for example, by shifting the phase of the reference wave. A minimum of three intensity patterns is required to calculate the phase at each point on the object surface. Known phase shifts are commonly introduced using a piezoelectric transducer, which serves to shorten or elongate the reference beam path by a fraction of wavelength; the phase-shifter is placed in one arm of the interferometer. Numerous phase shifting techniques have been developed and which can be incorporated to holographic interferometry to generate a phase map, $\Delta\varphi(x, y)$, corresponding to the object displacement. The calculated phase $\Delta\varphi(x, y)$ is independent of the terms $I_0(x, y)$ and $V(x, y)$, and considerably reduces the dependence of the accuracy of measurements on the fringe quality.

The introduction of phase shift interferometry not only provides accurate phase measurements but also eliminates the problem of phase sign ambiguity of the interference fringes. Dynamic phase shifting techniques allow for producing time sequences of deformation maps for studying objects subjected to time-varying loads. The concept is based on considering the normal phase variations caused by a dynamic phenomenon as equivalent to the phase shifting that is introduced in a holographic interferometer in order to perform the phase evaluation task. This method extends the possibility of applying the method to continuous deformation measurements.

Basic Interferometers for Static Displacement Measurement

Equation [6] shows that observation along a single direction yields information only about the resolved part of the surface displacement in one particular direction. The approach to measuring three-dimensional vector displacement would be to record holograms from three different directions in order to obtain three independent resolved components of displacement. The computation of the resulting system of equations would then allow for determining the complete displacement vector.

Out-of-Plane Displacement Component Measurement

The out-of-plane component of displacement is usually measured by illuminating and viewing the object surface in a near normal direction. The fringe equation corresponding to the line of sight displacement component is given by

$$w = \frac{n\lambda}{2} \qquad [8]$$

where w is the out-of-plane component of displacement. An example of a fringe pattern depicting the out-of-plane displacements in a rectangular aluminum plate, clamped along the edges and drawn out at the center by means of a bolt, subjected to three-point loading, is shown in Figure 4.

In-Plane Displacement Component Measurement

The in-plane component of displacement can be measured by illuminating the object along two directions symmetric to the surface normal and observing along a common direction. This is schematized in Figure 5. The in-plane displacements, in a direction containing the two beams, are mapped as a moiré between the interferograms due to each illumination beam. The moiré appears as a family of fringes:

$$u = \frac{n\lambda}{2\sin\theta_e} \qquad [9]$$

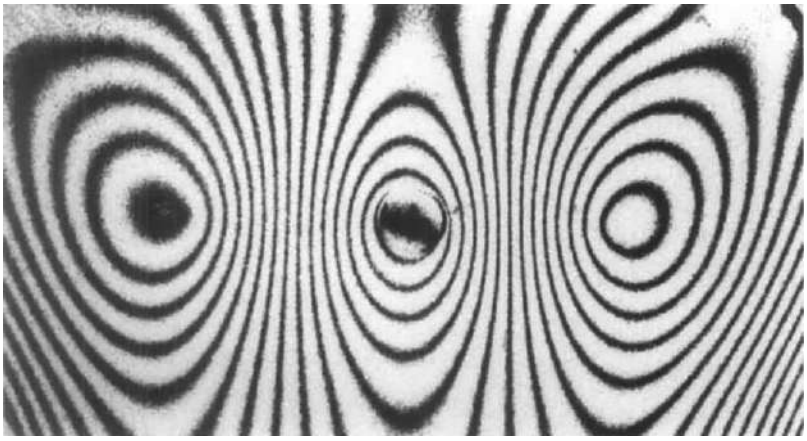

Figure 4 Example of interferogram corresponding to the whole field distribution of out-of-plane component of displacement, w, on an aluminum plate subjected to three-point bending.

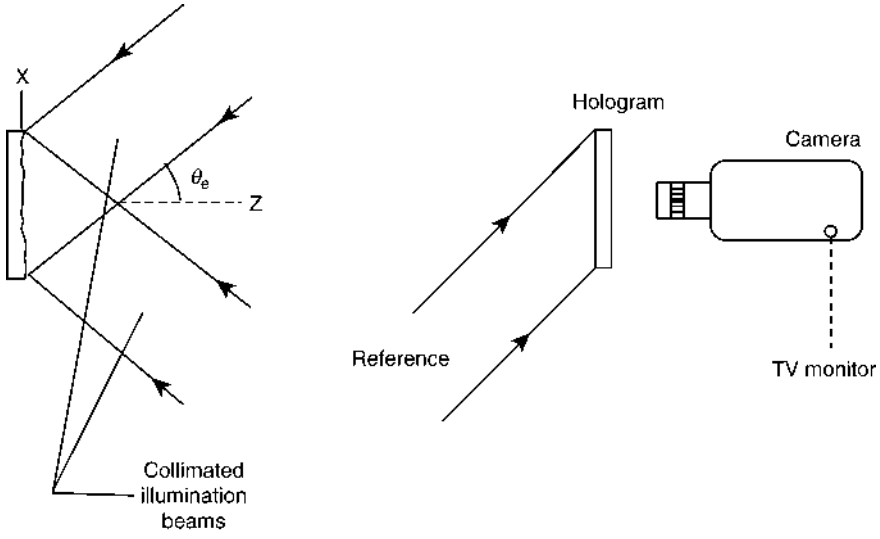

Figure 5 Schematic of the geometrical configuration for measuring the in-plane component of displacement, *u*.

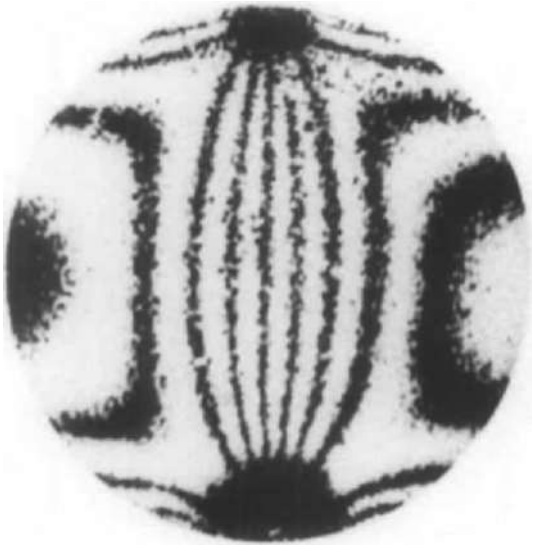

Figure 6 Fringe contours showing the in-plane displacement response of a disk under diametral compression.

where u is the displacement component in the x-direction and θ_e is the angle which the illumination beams make with the surface normal. Figure 6 shows the u pattern corresponding to a disk under diametral compression. The moiré pattern is obtained after optically filtering the interference image.

Comparative Holographic Interferometry

Another technique of significant interest in non-destructive testing is comparative holography that provides the contours of path variations related to the difference in displacements or shapes of two objects. The method provides information about the resolved part of the difference in displacement vector along the direction bisecting the illumination and observation rays. In the case of illumination and observation normal to the object surface, the fringe equation becomes

$$\Delta w = \frac{n\lambda}{2} \qquad [10]$$

where Δw is the difference in displacement component along z direction. The technique provides a tool to compare the mechanical responses of two nominally identical specimens subjected to the same loading and also the shapes of two nominally identical specimens. These features are useful in detecting anomalies in a test specimen with respect to the flaw-free master specimen. One such example is shown in Figure 7a, which displays the presence of a defect in a test specimen. In this case two plates, one flaw-free and the other with a flaw, are subjected to identical loadings. The fringe map in Figure 7b is related to the difference in displacements of test and master specimens subjected to unequal loadings. The presence of flaws is detected in both interferograms.

Basic Interferometers for Dynamic Displacement Measurement

Another important application of holographic interferometry is in the study of the resonant modes of vibration of an object.

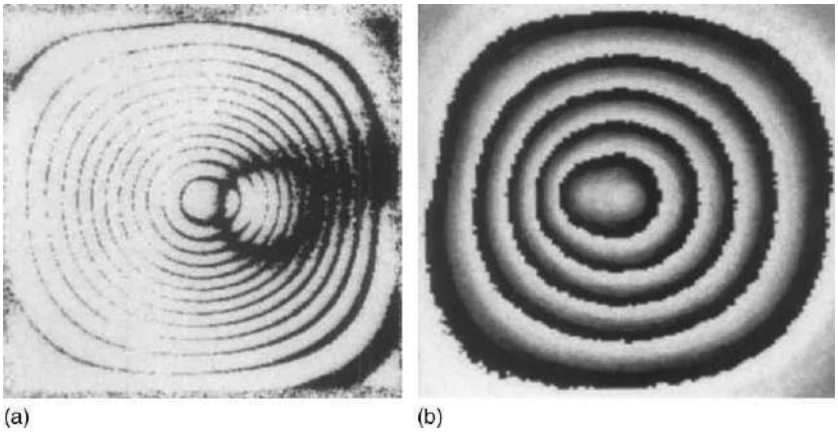

Figure 7 Comparative holography reveals the presence of defects in aluminum plates using (a) identical and (b) unequal loadings.

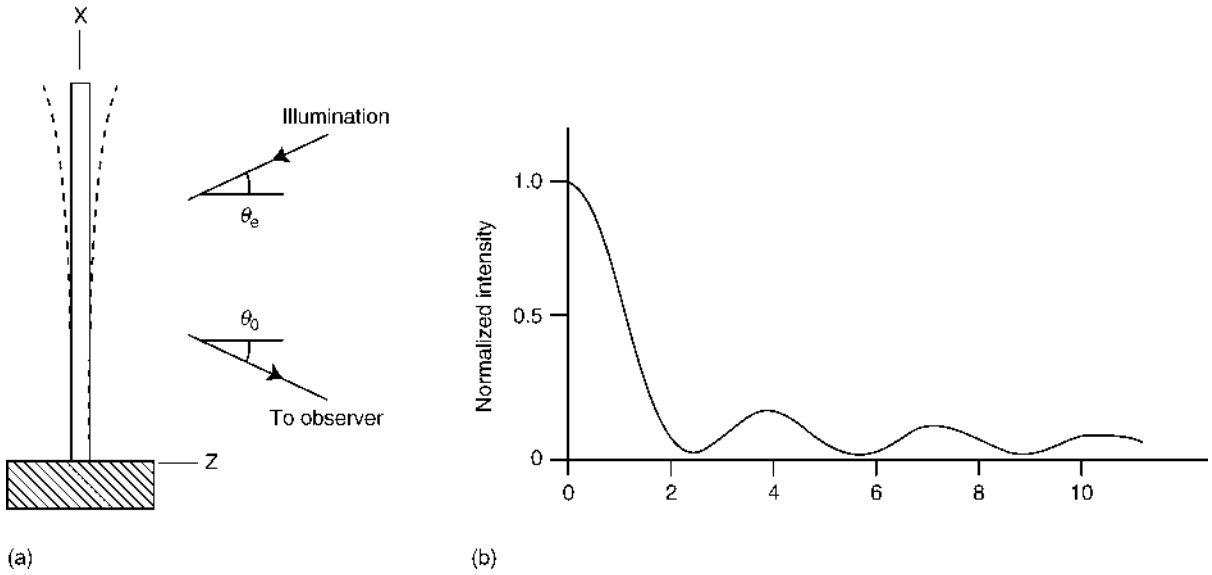

Figure 8 (a) Schematic of the geometrical configuration of a vibrating cantilever beam; (b) plot of fringe function for time-average holography.

Frozen Time-Average Holographic Interferometry

Consider that a hologram is recorded of an object, which is vibrating sinusoidally in a direction normal to its surface (Figure 8a). The exposure time is supposed to be much longer than the period of vibration. The intensity distribution of the image reconstructed by this hologram is

$$I(x, y) = I_o(x, y) J_o^2\left[\frac{2\pi}{\lambda} d(x, y)(\cos\theta_e + \cos\theta_o)\right] \quad [11]$$

where J_o is the zero-order Bessel function of the first kind. The virtual image is modulated by the $J_o^2(\xi)$ function. The dark fringes correspond to the zeros of the function $J_o^2(\xi)$. A plot of the function, shown in Figure 8b, is characterized by a comparatively brighter zero-order fringe, which corresponds to the nodes, a decreasing intensity, and an unequal spacing between the successive zeros. The interferogram in Figure 9 shows the vibrational modes of a helicopter component. The nodes that represent zero motion are clearly seen as the brightest areas in the time average reconstructed pattern.

Real-Time Time-Average Holographic Interferometry

The possibility of studying the response of vibrating objects in real-time extends the usefulness of the technique to identify the resonance states (resonant vibration mode shapes and resonant frequencies) of the object. A single exposure is made of the object in its state of rest. The plate is processed, returned to its original condition and reconstructed. The observer looking through the hologram at the sinusoidally

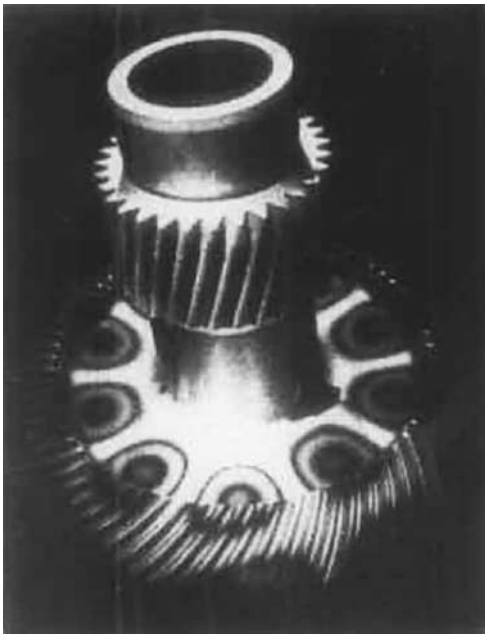

Figure 9 Fringe contours showing the mode shape of a vibrating helicopter component.

vibrating object sees the time-averaged intensity:

$$I(x,y) = I_o(x,y)\left[1 - J_o\left\{\frac{2\pi}{\lambda}d(x,y)(\cos\theta_e + \cos\theta_o)\right\}\right] \quad [12]$$

Stroboscopic Holographic Interferometry

Stroboscopic holography is another interesting variant in the study of vibrations. The hologram is recorded by exposing the photographic plate twice for short time intervals during a vibration cycle. The pulsed exposures are synchronized with the vibrating surface, which is equivalent to making the surface virtually stationary during the recording. Reconstruction of the hologram yields cosinusoidal fringes, which are characteristic of double exposure holography. In real-time stroboscopic holography, a hologram of a nonvibrating object is first recorded. If the vibrating object is illuminated stroboscopically and viewed through the hologram, the reconstructed image from the hologram interferes directly with the light scattered from the object to generate live fringes.

Flow Measurement

Application of holographic interferometry to flow visualization and the measurement of spatial refractive index, density or temperature distributions have led to an advancement of understanding in areas such as aerodynamics, plasma diagnostics, and heat transfer. In the holographic interferometer shown in Figure 10 two consecutive exposures are made, usually the first exposure without flow and the second in the presence of a flow field. Double pulsed holography is used if the flow field is changing rapidly. The optical phase change due to flow between the exposures is

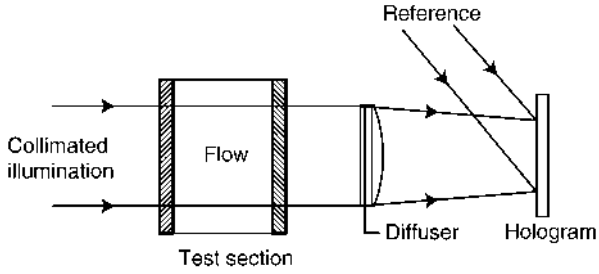

Figure 10 Schematic of a configuration for studying flow.

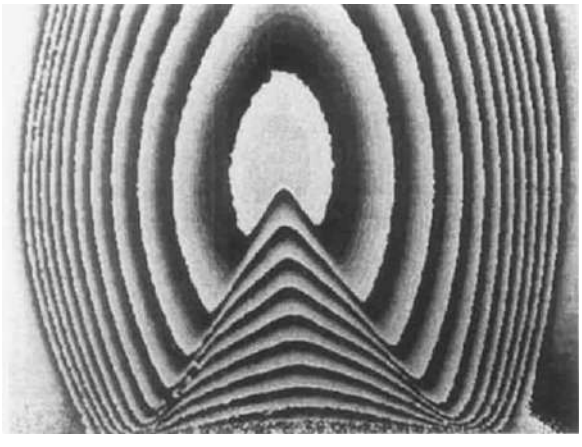

Figure 11 Fringe contours showing the phase distribution of a test flame.

$$\Delta\varphi(x,y) = \frac{2\pi}{\lambda}\int_0^t \{n'(x,y,z) - n'_o\}\mathrm{d}z \quad [13]$$

where $n'(x,y,z)$ is the refractive index distribution during the second exposure, n'_o is the uniform refractive index during the first exposure, and t is the length of the test section. Assuming that the properties of the flow are constant in the z direction, the expression for phase change can be expressed as

$$\Delta\tau(x,y) = KL\{\rho(x,y) - \rho_o\} \quad [14]$$

where K is the Gladstone–Dale constant, ρ is the density of the gas, and ρ_o is the density in a no-flow situation. The interference pattern contours the change in the density field of the flow. The change in density per fringe is given by λ/Kt. An example of application related to temperature measurements in an axisymmetric premixed flame is illustrated in Figure 11.

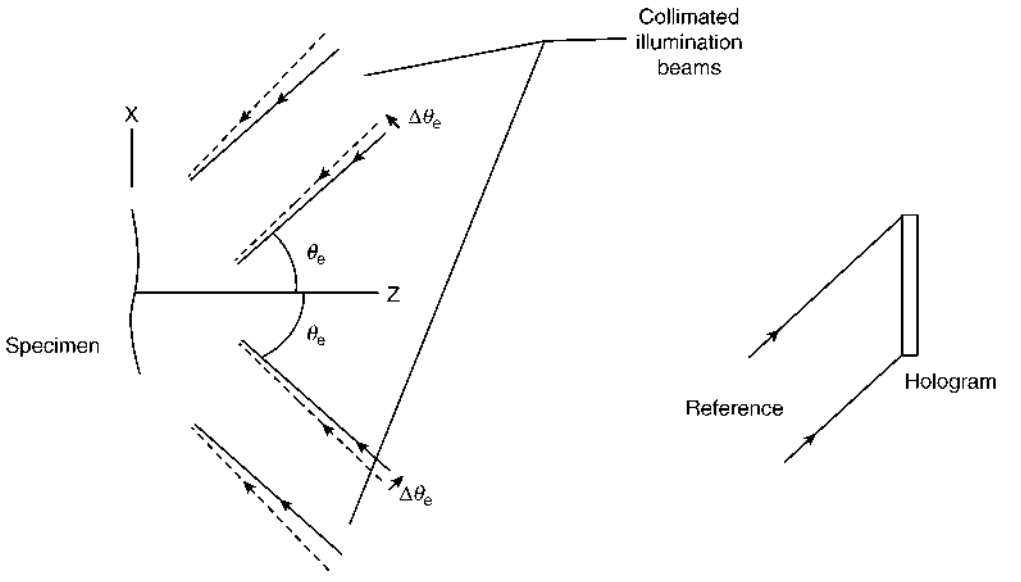

Figure 12 Schematic diagram of a two-beam method for shape measurement.

Surface Contouring

A practical way to display the shape of an object is to obtain contour maps showing the intersection of the object with a set of equidistant planes orthogonal to the line of sight.

The Dual Refractive Index Method

The method requires placing the model in a glass tank filled with a liquid (or gas) of refractive index n'_1. A hologram of the object is recorded, and the liquid contained in the tank is replaced by one of refractive index n'_2. An observer looking through the hologram sees the object surface modulated by a set of interference fringes arising from the change of optical phase in the light rays traversing the two liquids. The contour interval is given by

$$\Delta z = \frac{\lambda}{2(n'_1 - n'_2)} \qquad [15]$$

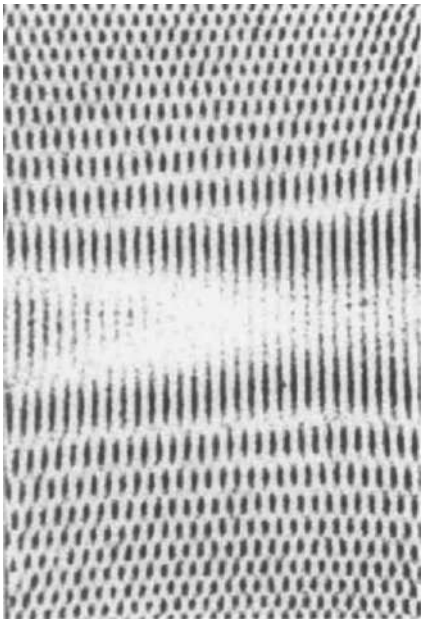

Figure 13 Example of a moiré topographic contour pattern obtained using the two-beam method.

The Dual Illumination Method

In this method, the object is illuminated obliquely by collimated beams along two directions symmetric to the surface normal. The scattered waves are recorded on a holographic plate and the object and its reconstructed image are viewed through the hologram. Two sets of parallel equidistant fringes are projected onto the object surface by introducing small tilts, $\Delta\theta_e$, to the illumination beams (Figure 12). The contour sensitivity per fringe is given by

$$\Delta z = \frac{\lambda}{2\sin\theta_e \sin\Delta\theta_e} \qquad [16]$$

The sensitivity of the method can be tuned in a wide range. An example of topographic contour pattern is shown in Figure 13. The distance between two adjacent moiré contour planes is 172 μm.

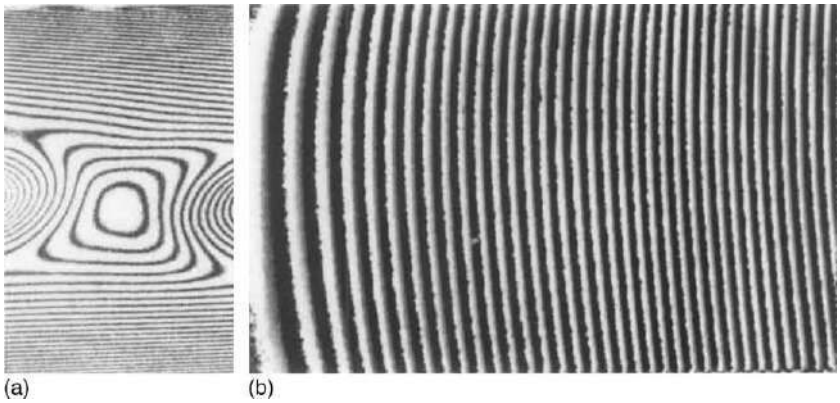

(a) (b)

Figure 14 Examples of (a) shape measurement: the distance between two adjacent contour planes is 208 μm; and, (b) very large out-of-plane deformation measurement on a cantilever beam. The distance between two adjacent contour planes is 19 μm.

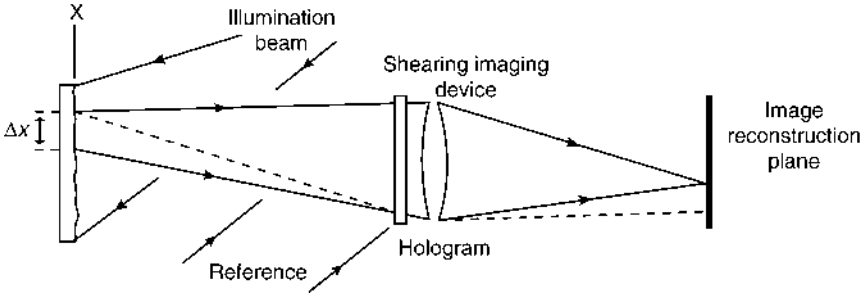

Figure 15 Schematic diagram of the principle used in holographic shearing interferometry.

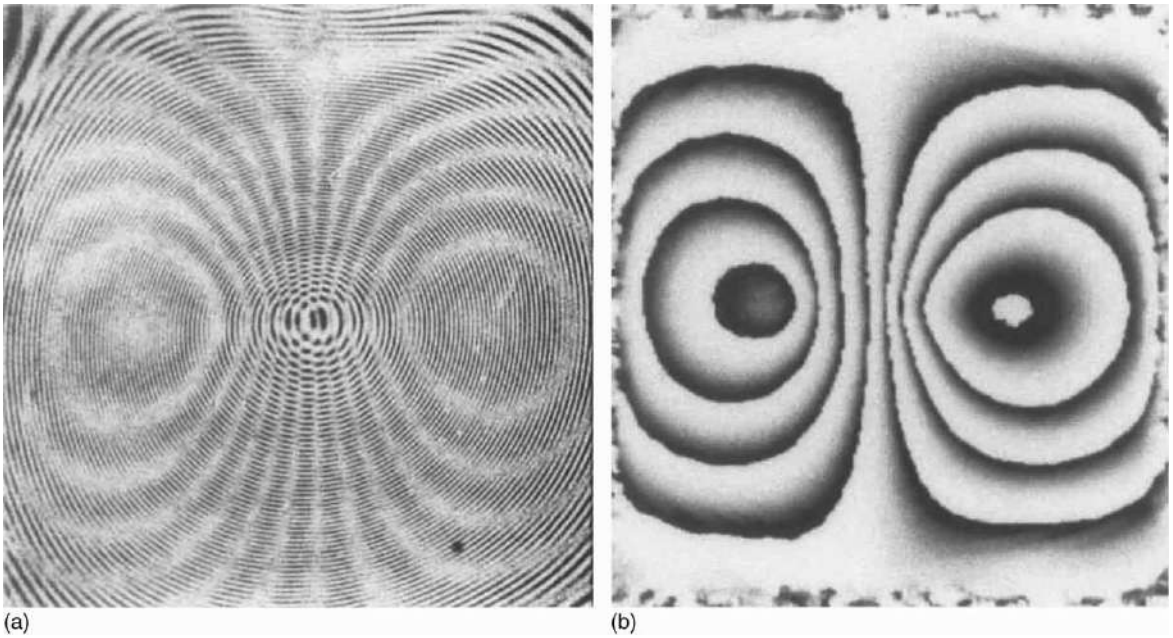

(a) (b)

Figure 16 Examples of (a) moiré fringe contour, and (b) phase distribution corresponding to slope change produced on a centrally loaded aluminum plate clamped along its boundary.

In another method a collimated beam obliquely illuminates the object and a set of parallel equidistant fringes is projected onto the object surface by introducing a small tilt, $\Delta\theta_e$, to the illumination beam. Providing an appropriate reference beam rotation and a holographic plate translation generates contouring surfaces normal to the line of sight. A typical result using one beam illumination is shown in Figure 14a. The distance between two adjacent moiré contour planes is 208 μm.

This approach has been used to obtain measurement of large out-of-plane deformations undergone by a deformed object. The fringe pattern in Figure 14b corresponds to the out-of-plane displacement of a cantilever beam subjected to load at its free end. The distance between two adjacent contour planes is 19 μm.

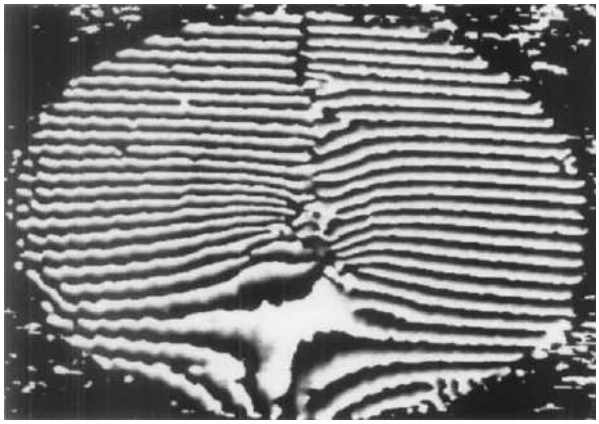

Figure 17 Example taken from the study of the fracture process zone in concrete: phase distribution corresponds to in-plane displacements on a notched concrete specimen subjected to a wedge splitting test.

Holographic Shearing Interferometry

Holographic shearing interferometry directly provides the patterns of slope change contours by laterally shearing wavefronts diffracted from the object surface. There are numerous ways to achieve shearing, such as, by introducing a Michelson type configuration, an inclined glass plate device, or a split-lens assembly in a holographic interferometer. The role of the shearing device is to enable the observation of a point on the object along two distinct neighboring directions. As a result of lateral wavefront shearing in the interferometer, a point in the image plane receives contributions from two different points on the object (Figure 15). Assuming that the illumination beam lies in the x–z plane and makes an angle θ_e with the z-axis, and the observation is carried along the direction normal to the object surface, the change of optical phase due to deformation is

$$\Delta\varphi = \frac{2\pi}{\lambda}\left[\frac{\partial u}{\partial x}\sin\theta_e + \frac{\partial w}{\partial x}(1+\cos\theta_e)\right]\Delta x \qquad [17]$$

where Δx is the object shear along the x-direction. For $\theta_e = 0$, eqn [11] reduces to

$$\frac{\partial w}{\partial x} = \frac{n\lambda}{2\Delta x} \qquad [18]$$

Equation [18] displays a family of holographic moiré fringes corresponding to the contours of constant slope change. The fringe pattern shown in Figure 16a illustrates a moiré corresponding to contours of constant slope change for a centrally loaded square aluminum plate clamped along its boundary. An example of phase distribution

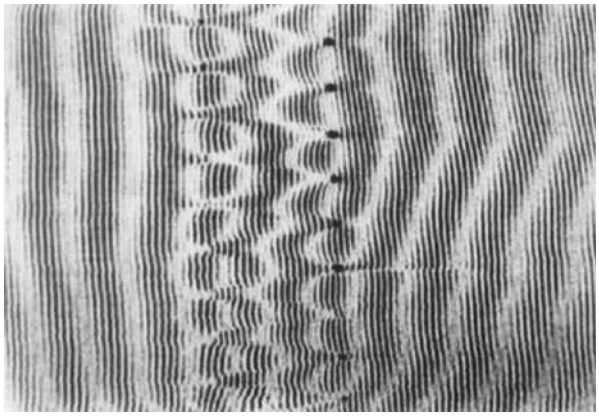

Figure 18 Example of application to nondestructive testing on a fingertip-joint wood specimen under tension using an in-plane sensitive holographic moiré setup.

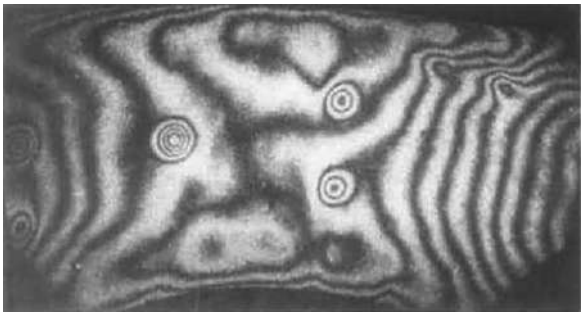

Figure 19 Application to the detection of defects in a tyre. Courtesy of H. Steinbichler, Steinbichler GmbH.

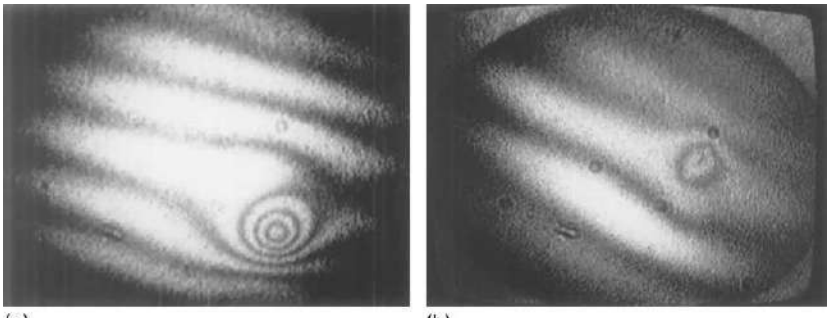

(a) (b)

Figure 20 Fringe contours showing (a) a specific limb movement of a chick embryo during incubation, and (b) the hatching behavior. The point on the shell receiving blows is clearly visible.

corresponding to slope change along x-direction, is shown in Figure 16b.

Examples of Applications

Holographic interferometry has firmly established itself as an attractive and useful class of technique in nondestructive testing. Defects in a specimen are made apparent by the presence of anomalies in the interference fringe distribution of the deformed state of the specimen. The presence of flaws such as voids, disbonds, delaminations, and cracks weakens the strength of the specimen. The resulting strain concentration areas produced by the presence of flaws in the specimen are detected through their effect on the surface deformation map. The flaws in the specimen manifest themselves as distortion of the fringes, pockets of fringe concentrations, or localized discontinuities of the fringes in the full-field display of the deformation field of a test specimen in response to the applied load.

Figure 17 shows phase distribution corresponding to in-plane displacements of a notched fiber-reinforced concrete specimen subjected to a wedge-splitting test. The application is taken from the study of a fracture process zone in fiber-reinforced concrete. Another example of the application of holography in nondestructive testing is shown in Figure 18. The fringe pattern corresponds to in-plane displacements on a fingertip joint wood specimen under tension. Figure 19 shows an example of an application of holography in the tyre industry. The interferogram reveals, through the changes in the fringe pattern, defects in the tyre. The picture in Figure 20a is taken from the study of the embryonic behavior of embryos during incubation. The photograph shows a specific limb movement near the air space. The method also allows for a lively display of the hatching behavior of the chick. The point on the shell receiving blows is clearly visible in Figure 20b.

Conclusions

The aim of this article is to introduce readers to holographic techniques available for the noninvasive and whole-field investigation of the time-evolution and spatial distribution of the displacement and refractive index fields. Diverse techniques to obtain contourgrams of three-dimensional objects are also presented. The method offers strong potential in nondestructive testing applications, such as in the detection of flaws, in the conception of prototypes, in the evaluation of the performances of structural components, and in the verification of the reliability of finite-element computer codes.

See also

Holography, Techniques: Digital Holography; Overview.
Interferometry: Overview; Phase Measurement Interferometry.

Further Reading

Erf RK (1974) *Holographic Non-Destructive Testing*. New York: Academic.

Ostrovsky YI, Butusov MM and Ostrovskaya GV (1980) *Interferometry by Holography*. Berlin: Springer-Verlag.

Ostrovsky YI, Schepinov VP and Yakovlev VV (1991) *Holographic Interferometry in Experimental Analysis*. Berlin: Springer-Verlag.

Rastogi PK (1994) *Holographic Interferometry – Principles and Methods*. Berlin: Springer-Verlag.

Schumann W and Dubas M (1979) *Holographic Interferometry*. Berlin: Springer-Verlag.

Schumann W, Zürcher JP and Cuche D (1985) *Holography and Deformation Analysis*. Berlin: Springer-Verlag.

Vest CM (1979) *Holographic Interferometry*. New York: Interscience.

Sandwich Holography and Light in Flight

N Abramson, Royal Institute of Technology, Stockholm, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction to Holographic Interferometry

Holography is a photographic principle that produces 3D images that are so real that it is possible to take measurements inside this image with a precision of a fraction of a thousandth of a millimeter. Contours are formed within the holographic image that, like the relief lines of a map, indicate the displacements of every point on the object surface. The method is named holographic interferometry (*see* **Holography, Techniques:** Holographic Interferometry) and it can be used to measure deformation, vibration and also, with slightly lower resolution, dimension.

In real time holography the deformed object is directly compared to the holographic image of its undeformed state. In double exposure holography the comparison is made between two holographic recordings on the same plate, while in sandwich holography the two states are recorded on two different plates that are sandwiched together for comparison.

Sandwich holography is useful in industrial environments, because it can measure much larger deformations and also compensate for unwanted movement of the object or the setup between the two exposures. Further on in the process it will produce any sign of deformation and comparison can be made at any time without referring back to original setup. However, processing of the two plates is needed before the comparison can be made and the plates have to be positioned with high accuracy.

Interferometric Sensitivity

To calculate the interferometric sensitivity let us name (A) the point from which the laser light is emitted, (B) the point behind the hologram plate from which the observation is made, and (C) the displaced point on the object surface to be measured. Draw an ellipsoid through (C) with (A) and (B) as focal points. If (C) is displaced along the surface of the ellipsoid, the path length ACB will remain unchanged and no interference fringes are formed, while a displacement normal to this surface produces maximal interference. A diagram, the holodiagram, is based on these ellipsoids and used for the making and evaluation of holograms.

The bisector to ACB will be the normal to the ellipsoids. If the path length ACB is repeatedly increased by δL a diagram like Figure 1 is formed where the separation (s) of the ellipses, and thus the interferometric sensitivity, varies over the diagram. If δL is the wavelength of the laser light, one interference fringe is formed each time an object point has moved the distance (s) because of a deformation of the object between the two exposures. The displacement (d) can be calculated:

$$d = ns \qquad [1]$$

$$s = k0.5\delta L \qquad [2]$$

$$k = 1/\cos\alpha \qquad [3]$$

where n is number of fringes between fixed and studied point, k is constant along arcs of circles through A and B, and α is half the angle ACB.

When two slightly dissimilar patterns are placed one on top of the other, a moire pattern is formed that shows up the difference. The moire analogy to holographic interferometry is based on the fact that one fringe is formed for every ellipse that is crossed by an object point displaced between the two exposures.

If δL is the wavelength (λ) of the laser light the displacement (d) is

$$d = nk0.5\lambda \qquad [4]$$

If two-wavelength holography is used to measure the 3D shape of an undisturbed object, it will be seen intersected by ellipses where the separation (s) is

$$s = k0.5\lambda_1\lambda_2/(\lambda_1 - \lambda_2) \qquad [5]$$

If δL is the coherence length or the pulse length of the laser source, the object will be seen intersected by one single bright fringe, the thickness (s) of which is

$$s = k\delta L \qquad [6]$$

Everything outside that ellipse will be dark during reconstruction. This phenomenon can be used either to optimize the use of a limited coherence length or to explain the results of 'light-in-flight recordings', as described later.

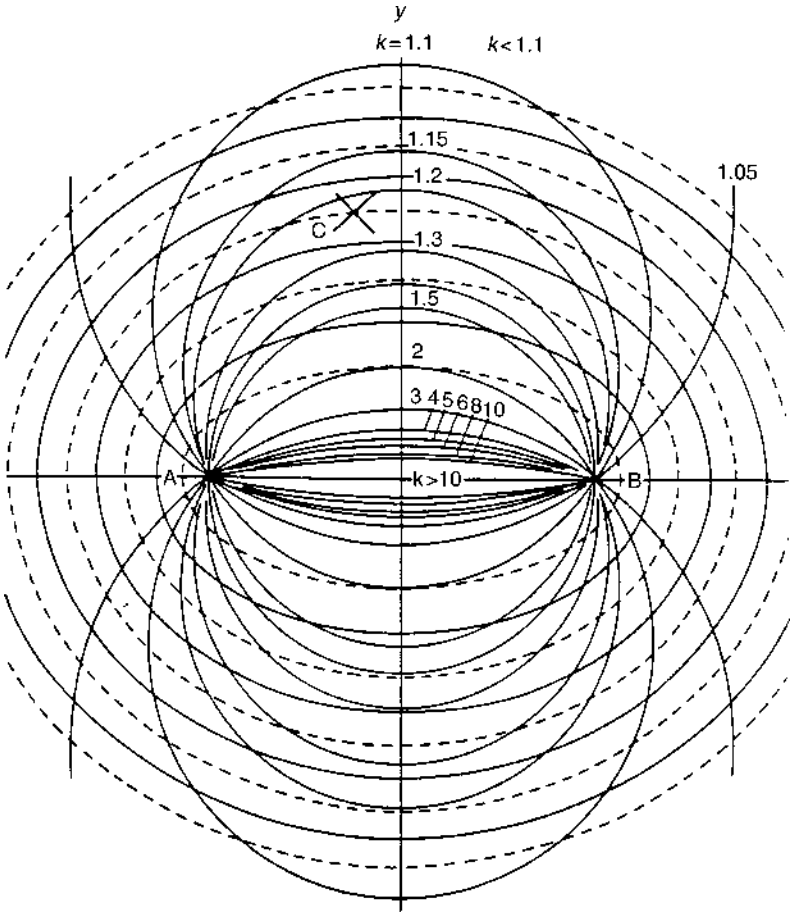

Figure 1 The holodiagram used for the evaluation of holograms.

The Principle of Sandwich Holography

In the following discussion on sandwich holography, we will only study the simplified example where the directions of illumination and observation are normal to the studied displacement and therefore α is equal to zero. The reflected light from an object point combined with the reference light produces a set of hyperbolas (Young's fringes) on the hologram plate.

When the object point is displaced between exposures on two different hologram plates, the hyperbolas will be positioned differently on the plates. There is a one-to-one relationship between the ellipsoids and the hyperbolas. Thus one point on the hologram plate will be crossed by one hyperbola if the corresponding object point is moved so that it crosses one of the ellipses of the holodiagram. Thus, the moire effect of the displaced hyperbolas at the hologram plate represents the moire effect of the ellipses in the object space. The interference pattern in the reconstructed object image can therefore be manipulated by moving one plate in relation to the other during reconstruction.

We have found that the simplest way to do this is to glue the two plates together separated by the glass thickness of one plate. To make the situation identical during exposure and reconstruction, the plates also have to be recorded in pairs. Thus, the first exposure can be made with one front plate and one back plate in contact in the hologram holder (Figure 2). The second exposure is made in exactly the same position with two new plates. After processing the front plate of the second exposure is glued in front of the back plate of the first exposure.

Then a reconstruction beam similar to the reference beam, is used to reconstruct the sandwich hologram. The interference fringes seen on the image of the object are manipulated by tilting the sandwich hologram, which results in an addition or subtraction of new fringes as in the following examples:

1. If the object is rotated by the small angle ϕ_1, the number of fringes can, during reconstruction, be

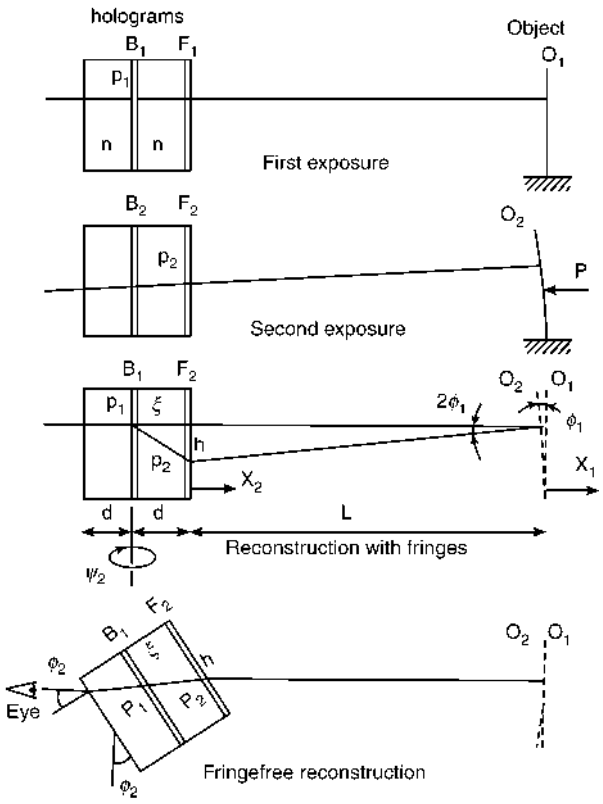

Figure 2 The exposures and reconstructions of a sandwich hologram.

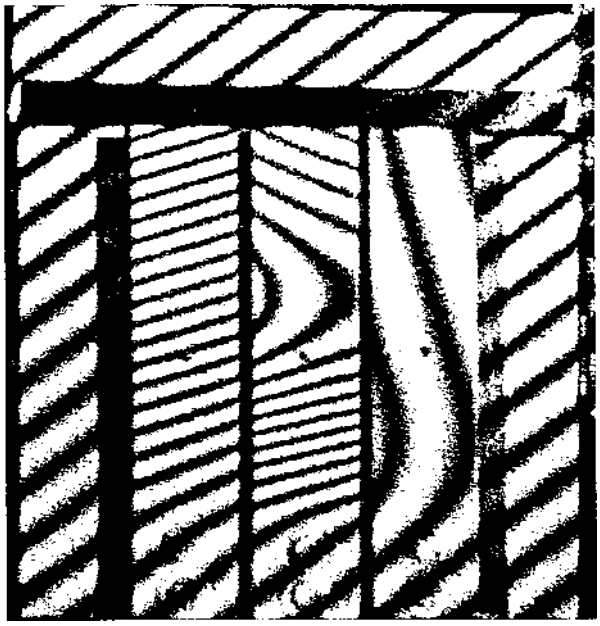

Figure 3 The movements of three steel bars in relation to a frame are to be measured by the fringes of sandwich holography. However, the unwanted motion of the total set-up makes the evaluation impossible.

eliminated by rotating the sandwich in the same direction by the much larger angle ϕ_2.

$$\phi_1 = \left[\frac{d}{2Ln}\right]\phi_2 \qquad [7]$$

2. If the sandwich during reconstruction is rotated around an axis perpendicular to that of the object rotation, new fringes are introduced at the angle γ, which is analogous to, but much larger than, the object rotation ϕ_1.

$$\phi_1 = \left[\frac{d}{2Ln}\right]\varphi_2 tg\gamma \qquad [8]$$

3. When the sandwich is rotated as described in (2) fringes are formed, the radius of curvature (r_2) of which are much larger than those of the bent object (r_1).

$$r_1 = \left[\frac{2Ln}{d}\right]\left[\frac{1}{\varphi_2}\right]r_2 \qquad [9]$$

Examples

The object consisted of three vertical steel bars (Figure 3) that were fixed by screws at their lower end to a rigid frame that surrounded all the bars. The middle bar was also supported at the upper end. Forces were applied at the middle of the bars. The right bar is deformed away from the observer while the other two are deformed towards him. After deformation of the three objects the second pair of plates were exposed. Then all four plates were processed.

The back plate of the first exposure (B_1 in Figure 2) was repositioned in the plate holder in front of the plate of the second exposure (F_2). The two plates were bonded together to form a sandwich hologram. O_1 and O_2 represent the two positions of one object, which between the two exposures was bent at angle ϕ_1 toward the hologram. (L) is the distance between the front hologram and the object (~1 m); (L) is also the distance between the point of illumination and the object. (D) is the distance between the point source of the reference beam and the hologram plate ($D = 2L$); (d) is the thickness of one hologram plate (~1.3 mm); (n) is the refractive index of the plates (~1.5 mm).

When the two plates were properly repositioned in the plate holder, fringes were seen everywhere, both on the steel bars and on the supposedly fixed frame which apparently had made an accidental movement (Figure 3). By tilting the plate holder in different directions, the fringes on the frame could be eliminated and the hologram could be evaluated as if the frame had remained fixed (Figure 4).

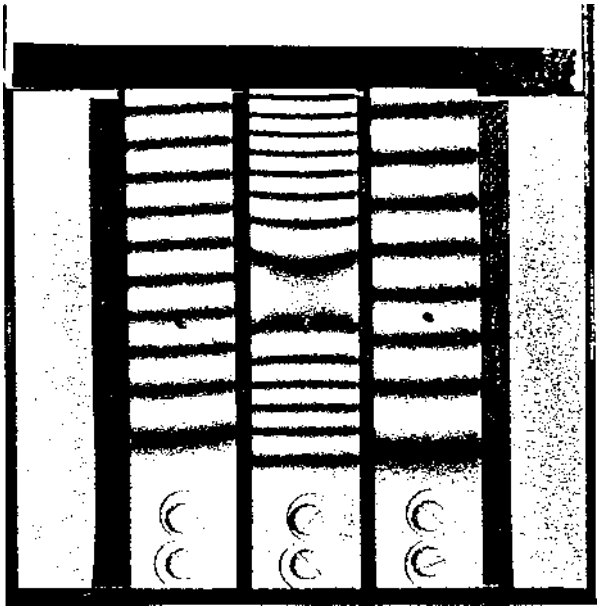

Figure 4 The unwanted fringes caused by the motion of the total set-up have been eliminated by a tilt of the sandwich hologram.

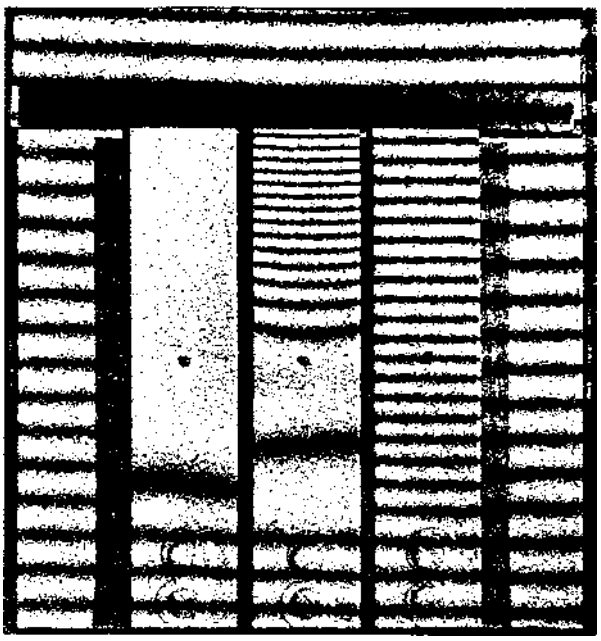

Figure 5 Another tilt of the sandwich hologram reveals that the left and the right bar have been deformed in opposite directions.

The plate holder was then tilted backwards at angle ϕ_2 and horizontal fringes were found on the frame while the top of the left bar became fringe-free (Figure 5). The number of fringes on the right bar increased, proving that it had been tilted in the opposite direction. If the sandwich were instead rotated around a vertical axis at angle φ_2, a set of vertical fringes were formed on the frame, while the bars became covered by fringes, the inclination and curvature of which represent a magnified view of the deformed bars as seen in profile (Figure 6).

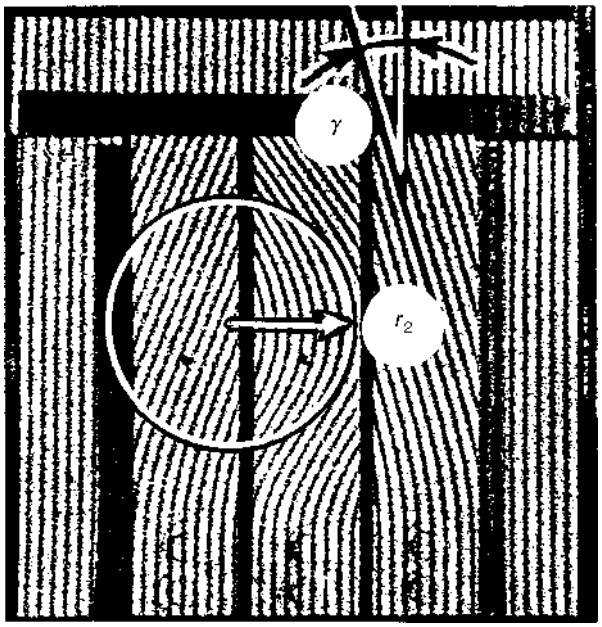

Figure 6 A rotation of the sandwich hologram around a vertical axis produces fringes that reveal direction of tilt and bending radius of all steel bars.

Introduction to Light-in-flight Recording by Holography

Recording of light pulses in flight have been possible since the beginning of the 1990s, using two-photon fluorescence or ultrafast Kerr cells driven by laser pulses. Very fast optical phenomena, such as refractive-index changes in laser-produced plasmas, have been recorded by holography using short illumination pulses.

In 1968, Staselko *et al.* published a method of studying the time coherence pattern of wave trains from pulsed lasers by the use of autocorrelation in which the unknown pulse is compared to itself. The method we name 'light-in-flight recording by holography' (LIF) is based on studying both the temporal and the spatial shape of pulses by correlation, which means that we study the change in shape of a pulse before and after it has passed through the transformation we want to study. For this process, we want to start with as short a pulse (or short coherence length) as possible and then compare the transformed pulse to the original pulse.

In a hologram only those parts of an object will be recorded for which the path length from the laser to the holographic plate via the object does not differ from the path length of the reference beam by more than the pulse length (or coherence length) of the laser light used for the recording. If that length is short, during reconstruction a large object will be seen intersected by an imaginary surface in the form of an ellipsoid, one of its two foci being the point from which the spherical wavefronts of illumination are emitted (spatial filter (A) in Figure 1) and the other being the point of observation (B). A spherical wavefront appears distorted into an ellipsoid if it is divergent, and into a hyperboloid if it is convergent, while a flat wavefront appears distorted into a paraboloid. The general rule is that the apparent wavefront has the shape of a mirror that focuses light from (A) to (B).

By careful planning, it is possible to optimize the geometry of the holographic setup in such a way that the utilized portions of the ellipsoids and hyperboloids can be approximated into flat surfaces. Thus, the object is seen intersected by one flat imaginary interference surface, the depth of which can be altered during reconstruction by changing the point of observation at the hologram plate.

Light of short duration, or short coherence length, can, in this way, be used in a holographic method for finding depth contours. If, however, the 3D shape of the object is known a priori, e.g., if it is a flat surface, this method can just as well be used to study the 3D shape of the illuminating wavefront. As mentioned before, the intersections of the ellipsoids can be approximated by intersections of spherical wave fronts of illumination, if the distance between object and hologram plate is long and if the flat surface is perpendicular to the line of sight.

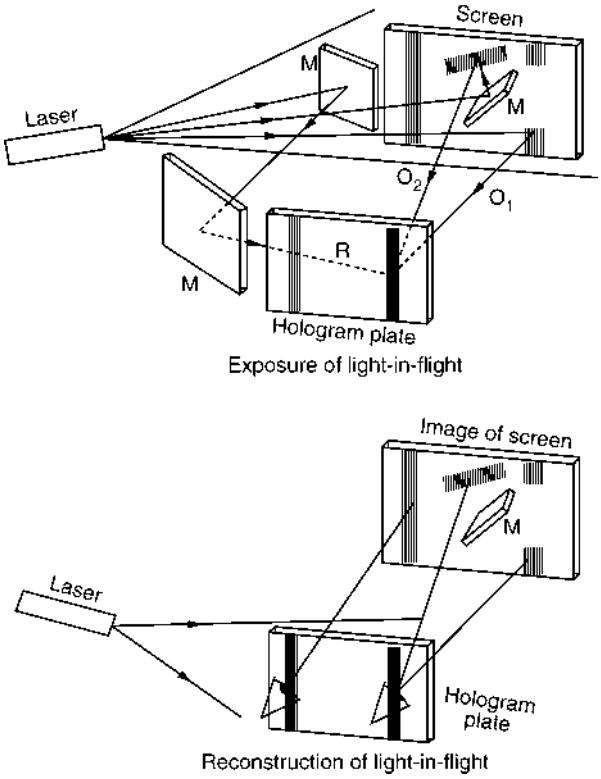

Figure 7 During exposure of the light-in-flight hologram, the short pulse from the laser is used to obliquely illuminate both the object and the hologram plate. Only those parts will be recorded where path length for object light and reference light are equal.

Examples of Wavefront Studies

As an example we will describe an experiment that consists of an opaque white diffusely reflecting flat screen, that is illuminated from the left at an oblique angle, as seen in Figure 7. A mirror is fixed to the screen so that some of the light is reflected upwards. During recording of the hologram, the reference pulse is reflected by the two mirrors (M) so that it arrives at the middle of the hologram plate at the same time as the object pulse, in the form of diffuse light, arrives from the middle of the screen. As the reference pulse passes over the hologram plate from left to right, it works like a curtain shutter (focal-plane shutter) recording early object light to the left and later light to the right. This 'light shutter' moves across the plate with a velocity faster than light which is possible as it is just an intersection between pulsefront and plate. The difference in time of the recordings made at the two ends of the hologram plate is the time it takes for the reference pulse to move across the plate, or about 800 picoseconds. During reconstruction with a continuous laser the early arriving object light is seen through the left part of the plate, while the later arriving light is seen through its right part. If the eye, or a TV camera scans from left to right, the pulse is seen moving continuously over the object screen, with some of it reflected upwards by the mirror (Figure 8a–d). The corresponding pulse length is 100 ps (~30 mm), and to minimize pulsefront distortions the screen is perpendicular to the line of sight.

The photo in Figure 9 shows the result of another experiment. A spherical wave expands from a point source at left and moves to the right. The spherical wavefront intersects a flat observation screen consisting of a white-coated aluminum plate of 20 × 30 cm. A cylindrical lens is fixed, with its axis normal to the screen, so that it focuses the arriving light almost parallel to the screen surface. Figure 8 is a composite of several photographs taken through different parts

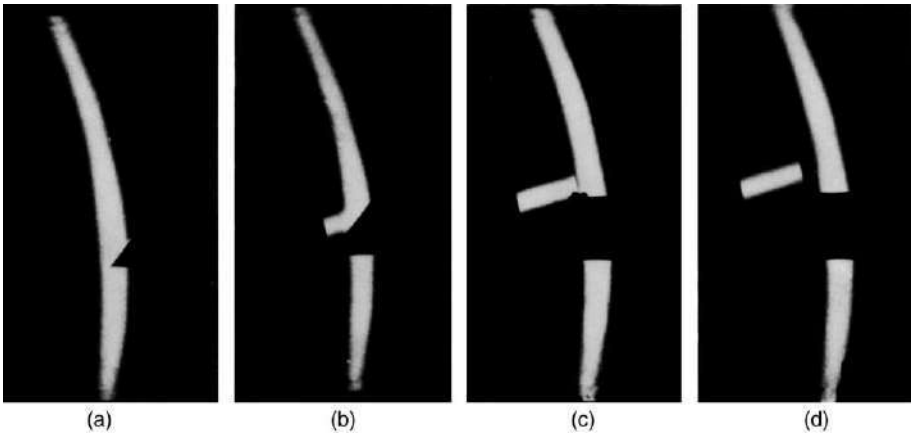

Figure 8 Resulting reconstructed images of the light reflected by a mirror at about 45°. These four photos were taken during reconstruction while the camera behind the hologram plate was moved from left to right in between each of the exposures.

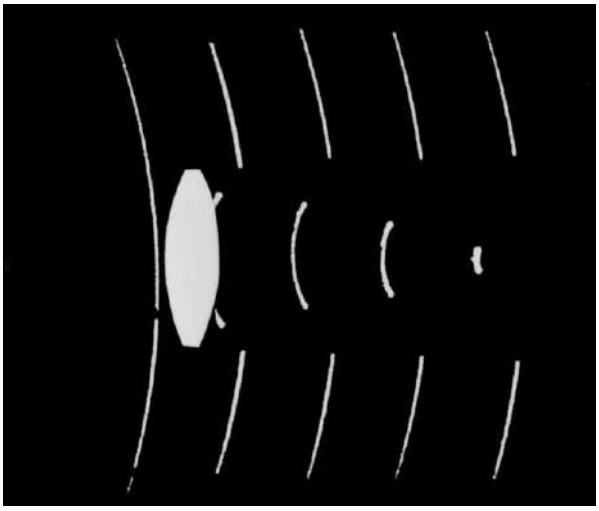

Figure 9 Light focused by a lens. A composite of several photos taken from different parts of the hologram plate. An image of the lens is included to make the picture clear.

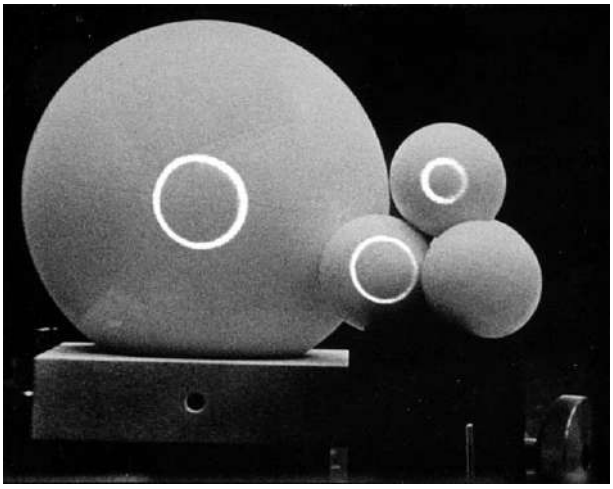

Figure 10 Spheres intersected by a thin sheet of light.

of the hologram plate. The pulse length was in this case 10 ps (~3 mm).

Measuring the Shape of 3D Objects

The wavefront (pulse front) of a short pulse forms a thin sheet of light. The intersection of this sheet by a 3D body produces on its surface a thin line representing a cross-section. It is advantageous if this cross-section represents a flat surface, which will be approximately the case if the illumination and observation beams are collimated or arrive from points at large distances.

By moving the point of observation along the hologram plate from left to right, the intersecting sheet of light can be moved in depth forward or backward. Determination of fringe order number (by finding the zero fringe) no longer represents any problem because only the zero order fringe exists. Therefore, this way of contouring has great advantages, especially when the observations and calculations are made by computer. Figure 10 shows a set of spheres intersected by a thin light sheet.

The intersecting light sheet can be manipulated in yet another way, i.e., it can be rotated. If the point of observation is moved away from the plate, the right part of an object will be seen through the right part of the plate were the recording made later than at the left part. During the time interval between left and right records, the intersecting light sheet has moved so that it intersects the object further away, resulting in a tilted cross-section of the object. This way it has been possible to rotate the cross-section up to 90 degrees, so that if, for example, a face is illuminated from the

front by a flat sheet of light, the reconstructed intersection will show the face in profile.

In a limited way it is also possible to produce results similar to those of LIF by using a CCD camera instead of the holographic plate. The great advantage is its compatibility with a PC, while the main disadvantage is a lower resolution.

Other Uses

Like all methods of 'gated viewing', LIF can be used to look through scattering media, e.g., to look through human tissue. By observing only the first light the influence of the later arriving scattered light is eliminated. Thus, LIF can be used as a diagnostic method based on the idea that the scattering, absorption, or refractive index of cancer is different from that of healthy tissue. Another use of LIF is to image just a thin section of a gas or liquid filled with particles. In this way it is possible to observe the distribution and velocity in a certain localization without influence of the scattering from particles in front of or behind the studied section.

Up to now we have discussed gated viewing by LIF in the laboratory using picosecond light pulses and distances of up to a few meters. Instead we could look at astronomical objects with light pulses of days (e.g., a supernova) and distances of many light years. Even in these cases we will see particles in space illuminated by ellipsoids with (A) being the supernova and (B) a telescope on Earth. While each ellipsoid in the laboratory may represent a pathlength of some nanoseconds, each ellipsoid in space might represent light years.

However, the methods we have shown here are not limited to the study of the apparent shape of spheres of light. They can also be used to study other objects moving at ultrahigh velocities. If an exploding supernova throws out particles in all directions at a rather slow velocity, it can be seen surrounded by an expanding sphere. If, instead, this sphere expanded with the speed of light (which is not possible) it would appear to us just like an ellipsoid of the holodiagram. If, however, it expanded at a very high velocity, but lower than that of light, the sphere would appear to us in the shape of an egg with its pointed end towards the observer. The velocity of expansion could, in this case, appear superluminal (higher than the speed of light). If the velocity is (v), its approaching velocity will to a first approximation appear to be the true velocity divided by ($1 - v/c$).

Finally, we will show a relationship between LIF and special relativity, by looking at a modification of the 'Minkowski light cones'. Figure 11 shows a diagram

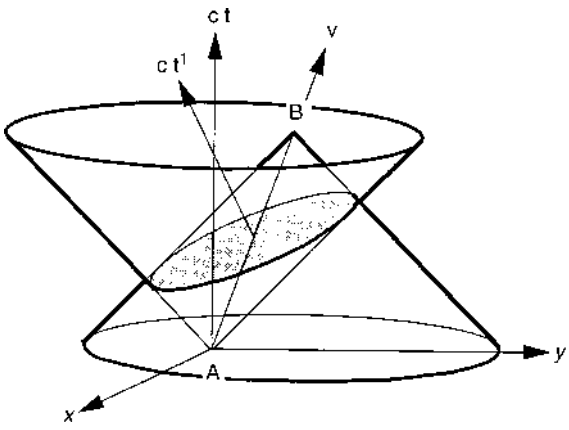

Figure 11 A modified Minkowski diagram consisting of one cone of illumination (A) and one cone of observation (B). The separation in time and space between (A) and (B) could either be static as in LIF or caused by a relativistic velocity of the observer from (A) to (B).

where (x) and (y) represent space coordinates, while (z) represents time multiplied by (c). A sphere from a short pulse of light that expands from the point A is represented by a cone with its apex at (A). A sphere from a short observation that converges to (B) is represented by a cone with its apex at (B). Only those object points, that are positioned at the elliptic intersection of the two cones, can be seen illuminated. This ellipse is identical to one of the ellipses of the holodiagram. The distance in time and space separating (A) and (B) could be either caused by the fixed distance between source of laser light and hologram plate in holographic LIF, or by a high velocity of the observer. Thus, the Lorentz contraction in special relativity could be explained as a measuring error because the high velocity of the observer traveling from (A) to (B) causes his sphere of observation to be transformed into an ellipsoid of observation. The tilt of the elliptic intersection indicates which observations that appear to be simultaneous depending on the separation of (A) and (B) along the y-axis, both in LIF and in relativity.

See also

Holography, Techniques: Holographic Interferometry.

Further Reading

Abramson N (1971) Moiré patterns and hologram interferometry. *Nature* 231: 65–67.

Abramson N (1981, 1986) *The Making and Evaluation of Holograms*. London: Academic Press.

Abramson N (1996) *Light in Flight, or, The Holodiagram – The Columbi Egg of Optics*. USA: SPIE Press, vol. PM 27.

An W and Carlsson T (1999) Digital measurement of three-dimensional shapes using light-in-flight speckle interferometry. *Optical Engineering* 38: 1366–1370.

Dugay MA and Mattick AT (1971) Ultrahigh speed photography of picosecond light pulses and echoes. *Applied Optics* 10: 2162–2170.

Garrington S (1992) Invisible jet sheds new light. *Nature* 355: 771–772.

Giordmaine J and Rentzepis P (1967) Two-photon excitation of fluorescence by picosecond light pulses. *Applied Physics Letters* 11: 216.

Hinrichs H and Hinsch KD (1997) Light-in-flight holography for flow visualization and velocimetry in three-dimensional flows. *Optics Letters* 22: 828–830.

Malin D and Allen D (1990) Echoes of the supernova. *Sky and Telescopes* 22–25, January.

Powell R and Stetson K (1965) Interferometric analysis by wavefront reconstruction. *Journal of the Optical Society of America* 55: 1593–1598.

Spears K and Serafin J (1989) Chrono-coherent imaging for medicine. *IEEE Transactions on Medical Engineering* 36: 1210–1220.

Staselko DI, Denisyuk YN and Smirnow AG (1969) Holographic recording of the time-coherence pattern of a wave train from a pulsed laser source. *Optics and Spectroscopy* 26: 413.

IMAGING

Contents

Information Theory in Imaging
Inverse Problems and Computational Imaging
Adaptive Optics
Hyperspectral Imaging
Imaging Through Scattering Media
Infrared Imaging
Interferometric Imaging
Lidar
Multiplex Imaging
Photon Density Wave Imaging
Three-Dimensional Field Transformations
Volume Holographic Imaging
Wavefront Sensors and Control (Imaging Through Turbulence)

Information Theory in Imaging

F O Huck and C L Fales, NASA Langley Research Center, Hampton, VA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Information theory is an intuitively attractive mathematical approach for combining observation with communication, or sensing with processing. So far, however, this approach has remained largely in the realm of signal processing and transmission, as first formalized by Shannon and Wiener over 50 years ago. Recent extensions of this approach to imaging have mostly addressed the interpretation of images, without explicitly accounting for the critical limiting factors that constrain image gathering and restoration. Yet, to be useful in the characterization of imaging, information theory must account for these constraints and lead to a close correlation between predicted and actual performance. The characterization that is outlined here traces the rate of information from the scene to the observer; and it pairs this rate, first, with the theoretical minimum data rate to assess the efficiency with which information can be transmitted, and, second, with the maximum-realizable fidelity to assess the quality with which images can be restored. Further details, especially for the mathematical development of figures of merit and for the design of imaging systems, can be found in the Further Reading.

Model of Imaging

Description

Modern imaging systems usually have the following three stages, as depicted in Figure 1:

1. *Image gathering*, to capture the radiance field that is either reflected or emitted by the scene and transform this field into a digital signal;

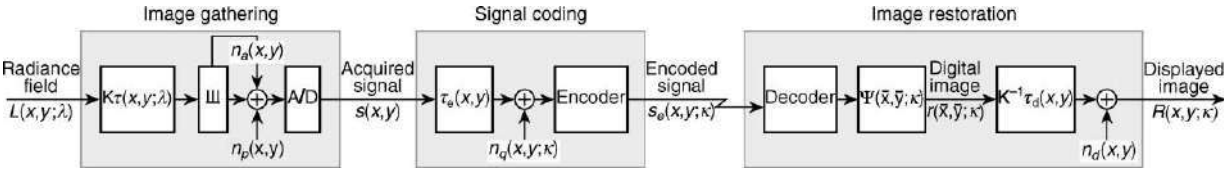

Figure 1 Model of digital imaging systems.

2. *Signal coding*, to encode the acquired signal for the efficient transmission and/or storage of data and decode the received signal for the restoration of images. The encoding may be either lossless (without loss of information) or lossy (with some loss of information), if a higher compression is required;
3. *Image restoration*, to produce a digital representation of the scene and transform this representation into a continuous image for the observer.

Human vision may be included in this model as a fourth stage, to characterize imaging systems in terms of the information rate that the eye of the observer conveys from the displayed image to the higher levels of the brain.

These stages are similar to each other in one respect: each contains one or more transfer functions, either continuous or discrete, followed by one or more sources of noise, also either continuous or discrete. The continuous transfer function of the image-gathering device (and of the human eye) is followed by sampling that transforms the captured radiance field into a discrete signal with analog magnitudes. In most devices the photodetection mechanism conveys this signal serially into an analog-to-digital (A/D) converter to produce a digital signal. However, analog signal processing in a parallel structure, akin to that in human vision, has many advantages that are increasingly emulated by neural networks.

This model of imaging differs from the classical model of communication that Shannon and Wiener addressed in two fundamental ways:

1. *The continuous-to-discrete transformation in image gathering*. Whereas communication is constrained critically only by bandwidth and noise, image gathering is constrained also by the compromise between blurring and aliasing, due to limitations in the response of optical apertures and in the sampling density of photodetection mechanisms. This additional constraint requires the rigorous treatment of insufficiently sampled signals throughout the imaging system.
2. *The sequence of image gathering followed by signal coding*. Whereas the characterization of communication addresses the perturbations that occur at the encoder and decoder or during transmission, the characterization of imaging systems must also address the perturbations in the image-gathering process prior to coding. These additional perturbations require a clear distinction between the information rate of the encoded signal and the associated theoretical minimum data rate or, more generally, between information and entropy.

The mathematical model of imaging that is outlined here accounts for the appropriate deterministic and stochastic (random) properties of natural scenes and for the critical limiting factors that constrain the performance of imaging systems. The corresponding formulations of the figures of merit are sufficiently general to account for a wide range of radiance field properties, photodetection mechanisms, and signal processing structures. However, the illustrations emphasize: (a) the salient properties of natural scenes that dominate the captured radiance field in the visible and near-infrared region under normal daylight conditions; and (b) the photodetector arrays that are used most frequently as the photodetection mechanism of image-gathering devices.

Radiance Field

The radiance field $L(x, y; \lambda)$, reflected by natural scenes, may be modeled approximately as

$$L(x, y; \lambda) = \frac{1}{\pi} I(\lambda)\rho(\lambda)\rho(x, y) \qquad [1]$$

where $I(\lambda)$ and $\rho(\lambda)$, respectively, represent the spectral irradiance and reflectance properties as a function of the wavelength λ, and $\rho(x, y)$ represents the random reflectance properties as a function of the spatial coordinates (x, y). Figure 2 illustrates a model of $\rho(x, y)$ for which the power spectral density (PSD) $\hat{\Phi}_\rho(\upsilon, \omega)$ corresponds closely to that typical of natural scenes, as given by

$$\hat{\Phi}_\rho(\upsilon, \omega) = \frac{2\pi\mu_\rho^2\sigma_\rho^2}{[1 + (2\pi\mu_\rho\zeta)^2]^{3/2}} \qquad [2]$$

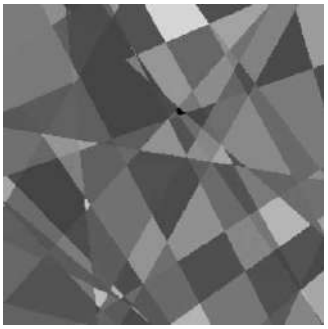

Figure 2 Model of reflectance $\rho(x, y)$.

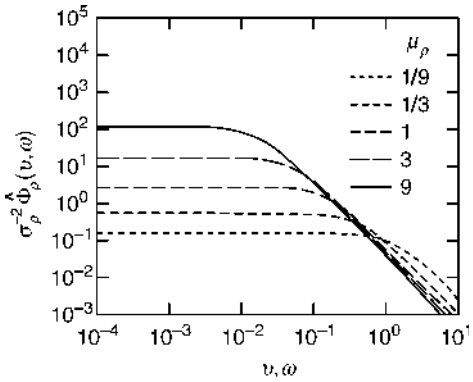

Figure 3 Normalized PSD of the reflectance $\rho(x, y)$.

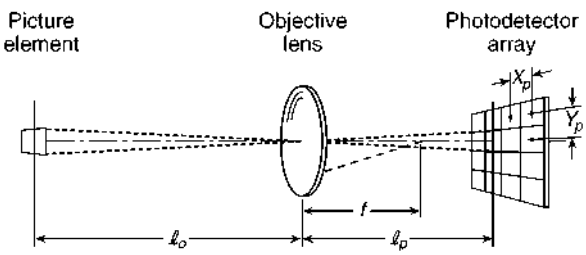

Figure 4 Optical geometry of image-gathering device.

where $\zeta^2 = \upsilon^2 + \omega^2$ and (υ, ω) are the spatial frequency coordinates of the Fourier-transform domain (see Figure 3). This model consists of random polygons, whose boundaries are distributed according to Poisson probability with a mean separation of μ_ρ, and whose magnitudes are distributed according to independent zero-mean Gaussian statistics of variance σ_ρ^2. The constraints associated with eqn [1] are satisfied by letting the mean reflectance $\bar{\rho} = 0.5$ and the standard deviation $\sigma_\rho \leq 0.2$.

Image Gathering

Figure 4 depicts the basic optical geometry of an image-gathering device, in which the objective lens forms an image of the captured radiance field across the photodetector array, and the array transforms this

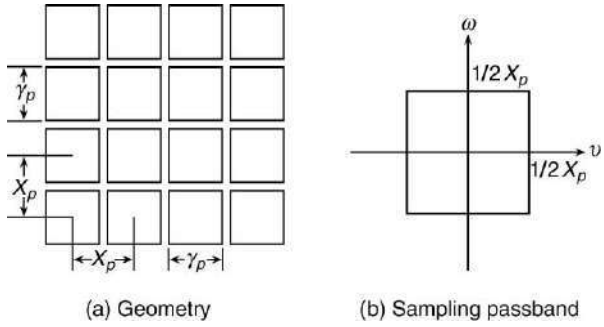

Figure 5 Photodetector-array geometry and sampling passband.

image into a discrete signal. The sampling lattice (X, Y), which determines the resolution of this device, is formed in angular units by the separation (X_p, Y_p) of the photodetector-array apertures and the distance ℓ_p as given for one of the dimensions by

$$X = 2\tan^{-1}\frac{X_p}{2\ell_p} \approx \frac{X_p}{\ell_p}\,\text{rad} = \frac{57.3X_p}{\ell_p}\,\text{deg} \qquad [3]$$

Normally, the sampling lattice is square, as shown in Figure 5, so that $X_p = Y_p$. Then the corresponding area of the photodetector apertures is $A_p = \gamma_p^2$, and the solid angle Ω of the instantaneous field of view (IFOV) is $\Omega \approx A_p/\ell_p^2$ ster.

The sampling lattice projected onto the scene at the distance ℓ_o has the dimensions $(X_o, Y_o) \approx (X\ell_o, Y\ell_o)$. Because these dimensions represent the finest spatial detail that normally can be restored, it is convenient to measure the mean spatial detail of the scene relative to the sampling lattice, as given by $\mu = \mu_\rho(X_o Y_o)^{-1/2}$, which reduces to $\mu = \mu_\rho X_o^{-1}$ for $X = Y$. Hence, $\mu = 1$ indicates that the sampling interval is equal to the mean spatial detail.

The image-gathering process that transforms the continuous radiance field $L(x, y; \lambda)$ into the discrete signal $s(\mathbf{x}, \mathbf{y})$ may be modeled by the expression

$$s(\mathbf{x}, \mathbf{y}) = K\rho(\mathbf{x}, \mathbf{y}) * \tau(\mathbf{x}, \mathbf{y}) + n_p(\mathbf{x}, \mathbf{y}) \qquad [4]$$

where the discrete coordinates $(\mathbf{x}, \mathbf{y}) = (mX, nY)$, and the symbol $*$ denotes continuous spatial convolution. This model accounts, in the simplest form, for the reflectance-to-signal conversion gain K, the reflectance $\rho(x, y)$ of the scene, the spatial response (SR) $\tau(x, y)$ of the image-gathering device, the continuous-to-discrete transformation by the photodetector-array lattice, and the discrete noise $n_p(\mathbf{x}, \mathbf{y})$ of the photodetectors. The discrete coordinates are

defined by the sampling function

$$\underline{\text{Ш}} \equiv \underline{\text{Ш}}(\mathbf{x}, \mathbf{y}) = XY \sum_{\mathbf{x},\mathbf{y}} \delta(x - \mathbf{x}, y - \mathbf{y}) \qquad [5]$$

where $\delta(x, y)$ is the Dirac delta function; and the gain is given by

$$K = \frac{1}{\pi} A_\ell \Omega \int I(\lambda)\rho(\lambda)r(\lambda)\mathrm{d}\lambda \qquad [6]$$

where $A_\ell = \pi D^2/4$ is the area of the objective lens with diameter D, and $r(\lambda)$ is the responsivity of the photodetectors. This gain relates the magnitude of the signal $s(\mathbf{x}, \mathbf{y})$ directly to that of the reflectance $\rho(x, y)$; and, together with the variances σ_ρ^2 and σ_p^2 of the reflectance and the photodetector noise, respectively, it also determines the rms-signal-to-rms-noise ratio (SNR) $K\sigma_\rho/\sigma_p$.

The DFT of $s(\mathbf{x}, \mathbf{y})$ is the periodic spectrum $\tilde{s}(\upsilon, \omega)$ given by

$$\tilde{s}(\upsilon, \omega) = K\hat{\rho}(\upsilon, \omega)\hat{\tau}(\upsilon, \omega) * \hat{\underline{\text{Ш}}} + \tilde{n}_p(\upsilon, \omega) \qquad [7]$$

where $\hat{\rho}(\upsilon, \omega)$ is the spectrum of the reflectance, $\hat{\tau}(\upsilon, \omega)$ is the spatial frequency response (SFR) of the image-gathering device, and $\tilde{n}_p(\upsilon, \omega)$ is the spectrum of the photodetector noise. The symbol $\hat{\underline{\text{Ш}}}$ is the Fourier transform of the sampling function $\underline{\text{Ш}}$. It accounts for the sidebands generated by the sampling process and is given by

$$\hat{\underline{\text{Ш}}} \equiv \hat{\underline{\text{Ш}}}(\upsilon, \omega) = \sum_{\mathbf{v},\mathbf{w}} \delta(\upsilon - \mathbf{v}, \omega - \mathbf{w}) \qquad [8]$$

where $(\mathbf{v}, \mathbf{w}) = (m/X, n/Y)$. The corresponding sampling passband is

$$\hat{B} = \left\{(\upsilon, \omega);\ |\upsilon| < \frac{1}{2X},\ |\omega| < \frac{1}{2Y}\right\} \qquad [9]$$

The above expression for $\tilde{s}(\upsilon, \omega)$ can be rewritten more conveniently as

$$\tilde{s}(\upsilon, \omega) = K\hat{\rho}(\upsilon, \omega)\hat{\tau}(\upsilon, \omega) + \hat{n}(\upsilon, \omega) \qquad [10]$$

where

$$\hat{n}(\upsilon, \omega) = \hat{n}_\mathrm{a}(\upsilon, \omega) + \tilde{n}_p(\upsilon, \omega) \qquad [11]$$

is the spectrum of the total noise in the acquired signal. The first term $\hat{n}_\mathrm{a}(\upsilon, \omega)$ represents the aliased signal components caused by insufficient sampling, as given by

$$\hat{n}_\mathrm{a}(\upsilon, \omega) = K\hat{\rho}(\upsilon, \omega)\hat{\tau}(\upsilon, \omega) * \hat{\underline{\text{Ш}}}_\mathrm{s} \qquad [12]$$

where

$$\hat{\underline{\text{Ш}}}_\mathrm{s} \equiv \hat{\underline{\text{Ш}}}_\mathrm{s}(\upsilon, \omega) = \sum_{\mathbf{v},\mathbf{w}\neq(0,0)} \delta(\upsilon - \mathbf{v}, \omega - \mathbf{w}) \qquad [13]$$

If, for spatial coordinates at either the scene or the photodetector array, the unit of (x, y) is meters, then the unit of (υ, ω) is cycles m^{-1}; and if, for angular coordinates centered at the objective lens, the unit of (x, y) is radians, then the unit of (υ, ω) is cycles rad^{-1}. Either way, it is convenient to normalize the sampling intervals to unity (i.e., $X = Y = 1$). Then the area of the sampling passband is $|\hat{B}| = 1$, the unit of (x, y) is samples, and the unit of (υ, ω) is cycles sample^{-1}.

The SFR $\hat{\tau}(\upsilon, \omega)$ of the image-gathering device is

$$\hat{\tau}(\upsilon, \omega) = \hat{\tau}_\ell(\upsilon, \omega)\hat{\tau}_p(\upsilon, \omega) \qquad [14]$$

where $\hat{\tau}_\ell(\upsilon, \omega)$ and $\hat{\tau}_p(\upsilon, \omega)$ are the SFRs of the objective lens and photodetector apertures, respectively. This SFR may be modeled approximately by the circularly symmetric Gaussian shape

$$\hat{\tau}(\upsilon, \omega) = \exp[-(\zeta/\zeta_\mathrm{c})^2] \qquad [15]$$

where, as shown in Figure 6, the optical-response index ζ_c controls the trade-off between the blurring that is caused by the decrease of the SFR inside the sampling passband and the aliasing that is caused by the extension of the SFR beyond this passband.

Signal Coding

Signal coding may be divided into two stages, as depicted in Figure 1. The first stage, which is represented by the digital operator $\tau_\mathrm{e}(\mathbf{x}, \mathbf{y})$ and the quantization noise $n_q(\mathbf{x}, \mathbf{y}; \kappa)$, accounts for manipulations of the acquired signal that seek to extract those

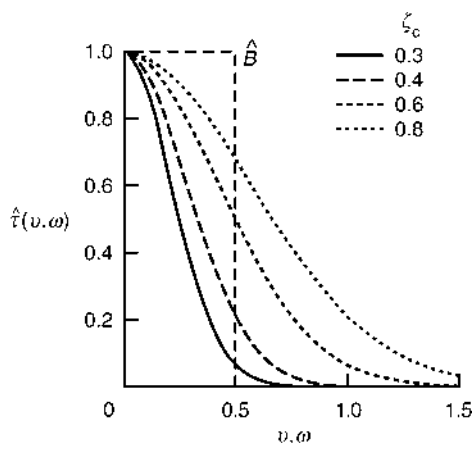

Figure 6 SFRs $\hat{\tau}(\upsilon, \omega)$ of the image-gathering device relative to the sampling passband $\hat{B}$ for unit sampling intervals (i.e., $X = Y = 1$).

features of the captured radiance field considered to be perceptually most significant, which usually entails some loss of information. The second stage, which is represented by the encoder, accounts for manipulations that seek to reduce redundancies in the acquired signal, without loss of information. The resultant encoded signal $s_e(\mathbf{x}, \mathbf{y}; \kappa)$ may be expressed in the same form as eqn [4] for the acquired signal $s(\mathbf{x}, \mathbf{y})$ as

$$s_e(\mathbf{x}, \mathbf{y}; \kappa) = s(\mathbf{x}, \mathbf{y}) \otimes \tau_e(\mathbf{x}, \mathbf{y}) + n_q(\mathbf{x}, \mathbf{y}; \kappa) \qquad [16]$$

where the symbol $\otimes$ denotes discrete spatial convolution, and $n_q(\mathbf{x}, \mathbf{y}; \kappa)$ is the quantization noise. If the data transmission and/or storage are error-free, as is assumed here, then the signal that leaves the decoder remains identical to the signal that enters the encoder.

The DFT of $s_e(\mathbf{x}, \mathbf{y}; \kappa)$ is the periodic spectrum $\tilde{s}_e(\upsilon, \omega; \kappa)$ given in the same form as eqn [10] by

$$\tilde{s}_e(\upsilon, \omega; \kappa) = K\hat{\rho}(\upsilon, \omega)\hat{\Gamma}_e(\upsilon, \omega) + \hat{n}_e(\upsilon, \omega; \kappa) \qquad [17]$$

where

$$\hat{\Gamma}_e(\upsilon, \omega) = \hat{\tau}(\upsilon, \omega)\tilde{\tau}_e(\upsilon, \omega) \qquad [18]$$

is the throughput SFR from image gathering to transmission, and

$$\hat{n}_e(\upsilon, \omega; \kappa) = \hat{n}(\upsilon, \omega)\tilde{\tau}_e(\upsilon, \omega) + \tilde{n}_q(\upsilon, \omega; \kappa) \qquad [19]$$

is the accumulated noise in the encoded signal. For η-bit quantization, where $\eta = \log \kappa$ and κ is the number of uniformly spaced quantization levels, $\tilde{n}_q(\upsilon, \omega; \kappa)$ may be treated (for $\eta \geq 4$) as independent Gaussian noise with the PSD

$$\tilde{\Phi}_q(\upsilon, \omega; \kappa) = \sigma_q^2 = (\sigma_{s_e}/\kappa)^2 \qquad [20]$$

where $\sigma_{s_e}^2$ is the variance given approximately by

$$\sigma_{s_e}^2 \approx K^2 \int_{\hat{B}} \int \hat{\Phi}_\rho(\upsilon, \omega)|\hat{\Gamma}_e(\upsilon, \omega)|^2 \mathrm{d}\upsilon \mathrm{d}\omega \qquad [21]$$

Image Restoration

Image restoration, as depicted in Figure 1, transforms the decoded signal $s_e(\mathbf{x}, \mathbf{y}; \kappa)$ into the digital image $r(\bar{\mathbf{x}}, \bar{\mathbf{y}}; \kappa)$, as modeled by the expression

$$r(\bar{\mathbf{x}}, \bar{\mathbf{y}}; \kappa) = \overline{\overline{\mathbb{I}}} s_e(\mathbf{x}, \mathbf{y}; \kappa) \otimes \Psi(\bar{\mathbf{x}}, \bar{\mathbf{y}}; \kappa) \qquad [22]$$

The operation indicated by $\overline{\overline{\mathbb{I}}} s_e(\mathbf{x}, \mathbf{y}; \kappa)$ (called zero padding) prepares the decoded signal for image restoration with interpolation, and the operator $\Psi(\bar{\mathbf{x}}, \bar{\mathbf{y}}; \kappa)$ produces the digital image $r(\bar{\mathbf{x}}, \bar{\mathbf{y}}; \kappa)$ on the interpolation lattice $\overline{\overline{\mathbb{I}}}$ that is Z times denser than the sampling lattice $\underline{\underline{\mathbb{I}}}$. If the interpolation in the image restoration is sufficiently dense (normally, for $Z \geq 4$), then the blurring and raster effects of the image-display process are suppressed entirely. This condition leads not only to the best realizable image quality but also to the simplest expression for the continuous displayed image, as given in the same form as eqns [10] and [17] by

$$\hat{R}(\upsilon, \omega; \kappa) = \hat{\rho}(\upsilon, \omega)\hat{\Gamma}_r(\upsilon, \omega; \kappa) + \hat{n}_r(\upsilon, \omega; \kappa) \qquad [23]$$

where

$$\hat{\Gamma}_r(\upsilon, \omega; \kappa) = \hat{\tau}(\upsilon, \omega)\tilde{\tau}_e(\upsilon, \omega)\hat{\Psi}(\upsilon, \omega; \kappa) \qquad [24]$$

is the throughput SFR from image gathering to display, and

$$\begin{aligned}\hat{n}_r(\upsilon, \omega; \kappa) &= K^{-1}\hat{n}_e(\upsilon, \omega; \kappa)\hat{\Psi}(\upsilon, \omega; \kappa) \\ &= \hat{n}_{ar}(\upsilon, \omega; \kappa) + \hat{n}_{pr}(\upsilon, \omega; \kappa) \\ &\quad + \hat{n}_{qr}(\upsilon, \omega; \kappa)\end{aligned} \qquad [25]$$

is the accumulated noise in the restored image. This noise may be expressed explicitly by substituting the noise components given above. The noise due to the granularity in image-display mediums is not accounted for here.

The Wiener filter, that minimizes the mean-square error due to the perturbations in image gathering and coding, is given by

$$\hat{\Psi}(\upsilon, \omega; \kappa) = \frac{\hat{\Phi}'_\rho(\upsilon, \omega)\hat{\Gamma}_e^*(\upsilon, \omega)}{\hat{\Phi}'_\rho(\upsilon, \omega)|\hat{\Gamma}_e(\upsilon, \omega)|^2 + \hat{\Phi}'_{n_e}(\upsilon, \omega; \kappa)} \qquad [26]$$

where $\hat{\Phi}'(\upsilon, \omega) = \sigma_\rho^{-2}\hat{\Phi}_\rho(\upsilon, \omega)$ is the normalized PSD of the radiance field and

$$\begin{aligned}\hat{\Phi}'_{n_e}(\upsilon, \omega; \kappa) = (K\sigma_\rho)^{-2}\Big[&\hat{\Phi}_n(\upsilon, \omega)|\tilde{\tau}_e(\upsilon, \omega)|^2 \\ &+ \tilde{\Phi}_q(\upsilon, \omega; \kappa)\Big]\end{aligned}$$

is the normalized PSD of the accumulated noise in the encoded signal. If the photodetector noise is white so that the PSD $\tilde{\Phi}_p(\upsilon, \omega)$ is equal to its variance σ_p^2 and the PSD $\tilde{\Phi}_q(\upsilon, \omega; \kappa)$ is equal to its variance σ_q^2, then the PSD of the noise simplifies to

$$\hat{\Phi}'_{n_e}(\upsilon, \omega; \kappa) = \hat{\Phi}'_{n_e}(\upsilon, \omega) + \tilde{\Phi}'_q(\upsilon, \omega; \kappa) \qquad [27]$$

where

$$\begin{aligned}\hat{\Phi}'_{n_e}(\upsilon, \omega) = \Big[&\hat{\Phi}'_\rho(\upsilon, \omega)|\hat{\tau}(\upsilon, \omega)|^2 * \underline{\underline{\hat{\mathbb{I}}}}_s \\ &+ (K\sigma_\rho/\sigma_p)^{-2}\Big]|\tilde{\tau}_e(\upsilon, \omega)|^2\end{aligned} \qquad [28]$$

and

$$\hat{\Phi}'_q(\upsilon, \omega; \kappa) = (K\sigma_\rho/\sigma_{s_e})^{-2}\kappa^{-2} \qquad [29]$$

These simplifying conditions permit the Wiener filter given above and the figures of merit given below to be expressed as a function of the SNR $K\sigma_\rho/\sigma_p$, which can be specified without accounting explicitly for either the gain K or the variances σ_ρ^2 and σ_p^2.

Figures of Merit

Information Rate $\mathcal{H}$

The information rate $\mathcal{H}$ (in bits sample^{-1}) of the encoded signal $s_e(\mathbf{x}, \mathbf{y}; \kappa)$ is defined by

$$\mathcal{H} = \mathcal{E}[\tilde{s}_e(\upsilon, \omega; \kappa)] - \mathcal{E}[\tilde{s}_e(\upsilon, \omega; \kappa)|\hat{\rho}(\upsilon, \omega) : (\upsilon, \omega) \in \hat{B}] \quad [30]$$

where the first term, $\mathcal{E}[\cdot]$, represents the entropy of $\tilde{s}_e(\upsilon, \omega; \kappa)$ and the second term, $\mathcal{E}[\cdot|\cdot]$, represents the conditional entropy of this signal when the reflectance spectrum $\hat{\rho}(\upsilon, \omega)$ within the sampling passband $\hat{B}$ is known. For the assumptions that have been made about image gathering and coding, the conditional entropy becomes the entropy of the accumulated noise $\hat{n}_e(\upsilon, \omega; \kappa)$, so that $\mathcal{H}$ simplifies to

$$\begin{aligned}\mathcal{H} &= \mathcal{E}[\tilde{s}_e(\upsilon, \omega; \kappa)] - \mathcal{E}[\hat{n}(\upsilon, \omega; \kappa) : (\upsilon, \omega) \in \hat{B}] \\ &= \frac{1}{2}\int_{\hat{B}}\int \log\left[1 + \frac{\hat{\Phi}'_\rho(\upsilon, \omega)|\hat{\Gamma}_e(\upsilon, \omega)|^2}{\hat{\Phi}'_{n_e}(\upsilon, \omega; \kappa)}\right] d\upsilon d\omega\end{aligned} \quad [31]$$

The theoretical upper bound of $\mathcal{H}$ is Shannon's channel capacity $\mathcal{C}$ for a bandwidth-limited system, with white photodetector noise and an average power limitation (here the variance σ_ρ^2), as given by

$$\mathcal{C} = \frac{1}{2}\log\left[1 + (K\sigma_\rho/\sigma_p)^2\right] \quad [32]$$

$\mathcal{H}$ can reach $\mathcal{C}$ only when

$$\hat{\Phi}_\rho(\upsilon, \omega) = \begin{cases} \sigma_\rho^2, & (\upsilon, \omega) \in \hat{B} \\ 0, & \text{elsewhere} \end{cases}$$

$$\hat{\tau}(\upsilon, \omega) = \begin{cases} 1, & (\upsilon, \omega) \in \hat{B} \\ 0, & \text{elsewhere} \end{cases} \quad [33]$$

However, neither of these two conditions can occur in practice because both the PSD $\hat{\Phi}_\rho(\upsilon, \omega)$ of natural scenes and the SFR $\hat{\tau}(\upsilon, \omega)$ of image-gathering devices decrease gradually with increasing spatial frequency, as depicted in Figures 3 and 6, respectively. Hence, $\mathcal{H}$ can never reach $\mathcal{C}$.

Theoretical Minimum Data Rate $\mathcal{E}$

The theoretical minimum data rate $\mathcal{E}$ (also in bits sample^{-1}) that is required to convey the information rate $\mathcal{H}$ is defined by

$$\mathcal{E} = \mathcal{E}[\tilde{s}_e(\upsilon, \omega; \kappa)] - \mathcal{E}[\tilde{s}_e(\upsilon, \omega; \kappa)|\tilde{s}(\upsilon, \omega)] \quad [34]$$

where the second term, $\mathcal{E}[\cdot|\cdot]$ is the uncertainty of $\tilde{s}_e(\upsilon, \omega; \kappa)$, when $\tilde{s}(\upsilon, \omega)$ is known. For the assumptions that have been made about the coding process, the conditional entropy becomes the entropy $\mathcal{E}[\tilde{n}_q(\upsilon, \omega; \kappa)]$ of the quantization noise, so that $\mathcal{E}$ simplifies to

$$\begin{aligned}\mathcal{E} &= \mathcal{E}[\tilde{s}_e(\upsilon, \omega; \kappa)] - \mathcal{E}[\tilde{n}_q(\upsilon, \omega; \kappa)] \\ &= \frac{1}{2}\int_{\hat{B}}\int \log\left[1 + \frac{\hat{\Phi}'_\rho(\upsilon, \omega)|\hat{\Gamma}_e(\upsilon, \omega)|^2 + \hat{\Phi}'_{n_e}(\upsilon, \omega)}{\tilde{\Phi}'_q(\upsilon, \omega; \kappa)}\right] d\upsilon d\omega\end{aligned} \quad [35]$$

This expression for $\mathcal{E}$ represents the entropy of completely decorrelated data.

Information Efficiency $\mathcal{H}/\mathcal{E}$

The information efficiency of completely decorrelated data is defined by the ratio $\mathcal{H}/\mathcal{E}$. This ratio reaches its upper bound $\mathcal{H}/\mathcal{E} = 1$ when $\hat{\Phi}'_{n_e}(\upsilon, \omega) \ll \hat{\Phi}'_\rho(\upsilon, \omega)|\hat{\Gamma}_e(\upsilon, \omega)|^2$ and $\hat{\Phi}'_{n_e}(\upsilon, \omega) \ll \tilde{\Phi}'_q(\upsilon, \omega; \kappa)$, because then both eqn [31] for $\mathcal{H}$ and eqn [35] for $\mathcal{E}$ reduce to the same expression

$$\mathcal{H} = \mathcal{E} = \frac{1}{2}\int_{\hat{B}}\int \log\left[1 + \frac{\hat{\Phi}'_\rho(\upsilon, \omega)|\hat{\Gamma}_e(\upsilon, \omega)|^2}{\tilde{\Phi}'_q(\upsilon, \omega; \kappa)}\right] d\upsilon d\omega \quad [36]$$

This upper bound can be approached with a small loss in $\mathcal{H}$ only when the aliasing and photodetector noise are small. Hence, the electro-optical design that optimizes $\mathcal{H}$ also optimizes $\mathcal{H}/\mathcal{E}$. However, a compromise always remains between $\mathcal{H}$ and $\mathcal{H}/\mathcal{E}$ in the selection of the number of quantization levels: $\mathcal{H}$ favors fine quantization, whereas $\mathcal{H}/\mathcal{E}$ favors coarse quantization.

Maximum-Realizable Fidelity $\mathcal{F}$

The maximum-realizable fidelity $\mathcal{F}$ of the images restored with the Wiener filter $\hat{\Psi}(\upsilon, \omega; \kappa)$ can be expressed as a function of the throughput SFR $\hat{\Gamma}_r(\upsilon, \omega; \kappa)$ given by eqn [24] as

$$\mathcal{F} = \int\int \hat{\Phi}'_\rho(\upsilon, \omega)\hat{\Gamma}_r(\upsilon, \omega; \kappa)\, d\upsilon d\omega \quad [37]$$

or, equivalently, as a function of the integrand $\hat{\mathcal{H}}(\upsilon, \omega; \kappa)$ of the information rate $\mathcal{H}$ given by eqn [31] as

$$\mathcal{F} = \int\int \hat{\Phi}'_{\rho}(\upsilon, \omega)\left[1 - 2^{-\hat{\mathcal{H}}(\upsilon,\omega;\kappa)}\right] \mathrm{d}\upsilon \mathrm{d}\omega \qquad [38]$$

This relationship between fidelity and information rate can be extended to other goals. For example, the enhancement of spatial features (e.g., edges and boundaries), for robotic vision, can be accommodated by letting $\tilde{\tau}_{\mathrm{e}}(\upsilon, \omega)$ be the desired feature-enhancement filter.

Performance and Design Trade-offs

Electro-optical Design

Figure 7 presents the information rate $\mathcal{H}$ (for lossless encoding) as a function of the optical-response index ζ_{c} and the mean spatial detail μ. The information rate $\mathcal{H}$ versus ζ_{c} is given for several SNRs $K\sigma_{\rho}/\sigma_{p}$ and $\mu = 1$, and $\mathcal{H}$ versus μ is given for the three informationally optimized designs specified in Table 1.

The curves show that $\mathcal{H}$ can reach its highest possible value only when both of the following conditions are met:

1. The relationship between the SFR $\hat{\tau}(\upsilon, \omega)$ and the sampling passband $\hat{B}$ is chosen to optimize $\mathcal{H}$ for the available SNR $K\sigma_{\rho}/\sigma_{p}$. This design is said to be informationally optimized.
2. The relationship between the sampling passband $\hat{B}$ and the PSD $\hat{\Phi}_{\rho}(\upsilon, \omega)$ is chosen to optimize $\mathcal{H}$. For the PSDs $\hat{\Phi}_{\rho}(\upsilon, \omega)$ typical of natural scenes, the best match occurs when the sampling interval is near the mean spatial detail of the scene (i.e., when $\mu \approx 1$).

These conditions are consistent with those for which the information rate $\mathcal{H}$ reaches Shannon's channel capacity $\mathcal{C}$. However, blurring and aliasing constrain $\mathcal{H}$ to values that are smaller than $\mathcal{C}$ by a factor of nearly two (i.e., $\mathcal{H} \lesssim \mathcal{C}/2$). Hence, these conditions replace the role of white and band-limited signals in classical communication theory, to establish an upper bound on the information rate in imaging systems.

The above two conditions appeal intuitively when the problem of image restoration is considered:

1. The result that the SFR $\hat{\tau}(\upsilon, \omega)$ that optimizes $\mathcal{H}$ depends on the available SNR is consistent with the observation that, in one extreme, when the SNR is low, substantial blurring should be avoided because the photodetector noise would constrain the enhancement of fine spatial detail even if aliasing were negligible. In the other extreme, when the SNR is high, substantial aliasing should be avoided so that this enhancement is relieved from any constraints except those that the sampling passband inevitably imposes.

Table 1 Informationally optimized designs

Design	$K\sigma_{\rho}/\sigma_{p}$	ζ_{c}	$\mathcal{H}^{*}$
1	256	0.3	4.4
2	64	0.4	3.3
3	16	0.5	2.2

$^{*}\mu = 1$.

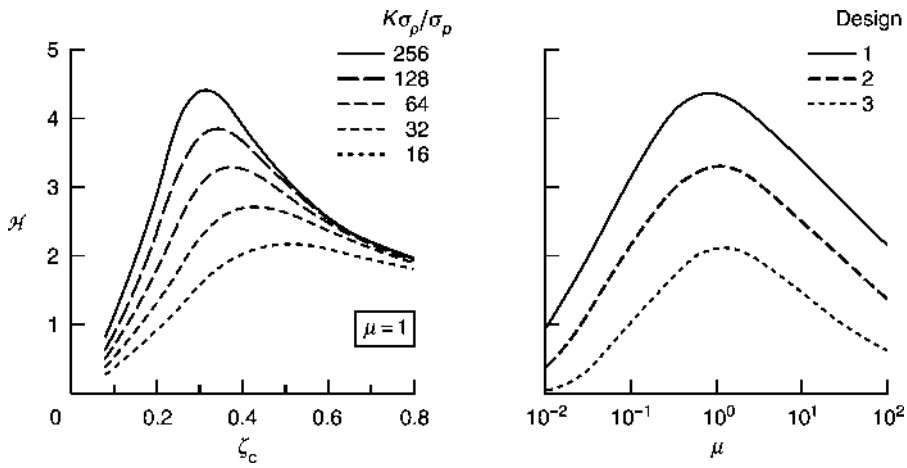

Figure 7 Information rate $\mathcal{H}$ versus the optical-response index ζ_{c} and the mean spatial detail μ (relative to the sampling interval). The designs are specified in Table 1, and the SFRs $\hat{\tau}(\upsilon, \omega)$ for ζ_{c} are shown in Figure 6.

2. The result that $\mathcal{H}$ reaches its highest value when the sampling interval is near the mean spatial detail of the scene is consistent with the observation that, ordinarily, it would not be possible to restore spatial detail that is finer than the sampling interval, whereas it would be possible to restore much coarser detail from fewer samples.

Information Rate and Data Rate

Figure 8 presents an information-entropy $\mathcal{H}(\mathcal{E})$ plot that characterizes the relationship between the information rate $\mathcal{H}$ of the encoded signal and the associated theoretical minimum data rate, or entropy, $\mathcal{E}$ for η-bit quantization. The encoder SFR $\tilde{\tau}_e(\upsilon, \omega)$ is assumed to be unity. The plot illustrates the trade-off between $\mathcal{H}$ and $\mathcal{E}$ in the selection of the number of quantization levels for encoding the acquired signal. It shows, in particular, that the design that realizes the highest information rate $\mathcal{H}$ also enables the encoder to realize the highest information efficiency $\mathcal{H}/\mathcal{E}$. This result appeals intuitively, because the perturbations due to aliasing and photodetector noise that interfere with the image restoration can also be expected to do so with the signal decorrelation.

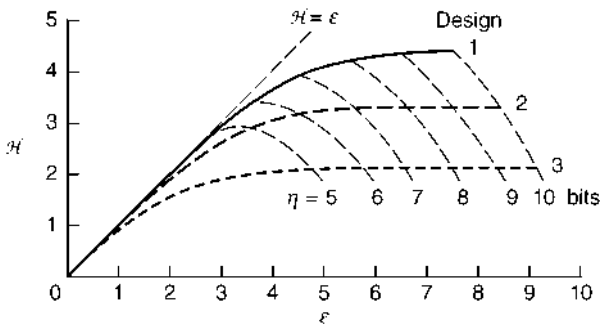

Figure 8 The information-entropy $\mathcal{H}(\mathcal{E})$ plot of the information rate $\mathcal{H}$ versus the associated theoretical minimum data rate $\mathcal{E}$ for η-bit quantization. The designs are specified in Table 1.

Throughput SFR

Figure 9 illustrates the SFRs $\hat{\tau}(\upsilon, \omega)$ and $\hat{\Psi}(\upsilon, \omega; \kappa)$ of the image-gathering device and Wiener filter, respectively, and of their product, the throughput SFR $\hat{\Gamma}_r(\upsilon, \omega; \kappa)$, for the three designs specified in Table 1. As the accumulated noise becomes negligible, $\hat{\Gamma}_r(\upsilon, \omega; \kappa)$ approaches the ideal SFR of the classical communication channel, which is unity inside the sampling passband and zero outside.

Information Rate, Fidelity and Robustness

The PSD $\hat{\Phi}_\rho(\upsilon, \omega)$ of a scene is seldom known when images are restored. Moreover, even if the PSD were known to be the one given by eqn [2], then the mean

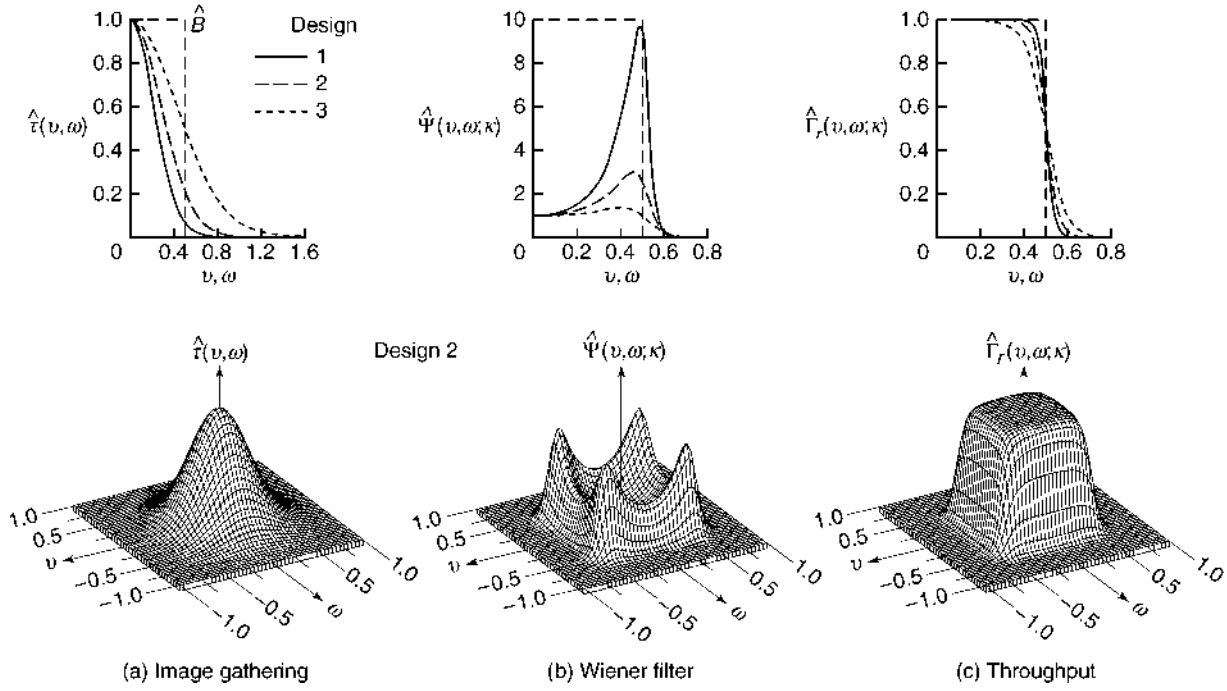

Figure 9 SFRs of image gathering, restoration, and throughput. The designs are specified in Table 1.

spatial detail μ would usually remain uncertain because it depends, not only on the properties of the scene, but also on the angular resolution and viewing distance of the image-gathering device.

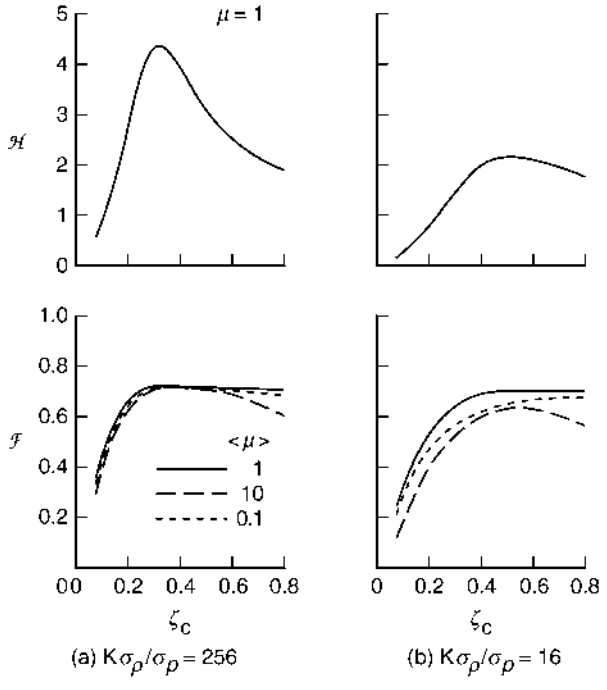

Figure 10 Information rate $\mathcal{H}$ and maximum-realizable fidelity $\mathcal{F}$ versus the optical-response index ζ_c for two SNRs $K\sigma_\rho/\sigma_p$. $\mathcal{F}$ is given for the matched ($\langle\mu\rangle = 1$) and two mismatched Wiener restorations.

In practice, therefore, image restorations have to rely on estimates of the statistical properties of scenes. The tolerance of the quality of the restored image to errors in these estimates is referred to as the robustness of the image restoration.

Figure 10 presents the information rate $\mathcal{H}$ and the corresponding maximum-realizable fidelity $\mathcal{F}$ for matched and mismatched Wiener restorations. These restorations represent the digital image $R(\bar{x}, \bar{y}; \kappa)$ or, equivalently, the continuous image $R(x, y; \kappa)$ displayed on a noise-free medium. The matched restorations use the correct value of μ, whereas the mismatched restorations use wrong estimates of μ. These curves reveal that the informationally optimized designs produce the highest fidelity and robustness, and that both improve with increasing $\mathcal{H}$. Moreover, these curves reveal that, ordinarily, one cannot go far wrong in Wiener restorations by assuming that the mean spatial detail is equal to the sampling interval (i.e., $\mu = 1$). This observation appeals intuitively because spatial detail that is much smaller cannot be resolved, and detail that is larger is not degraded substantially by blurring and aliasing.

Information Rate and Visual Quality

Figures 11 and 12 present Wiener restorations and their three components, as given by eqn [23], where $\rho(x, y) * \Gamma_r(x, y; \kappa)$ accounts for blurring, $n_{ar}(x, y; \kappa)$ accounts for aliasing, and $n_{pr}(x, y; \kappa)$ accounts for

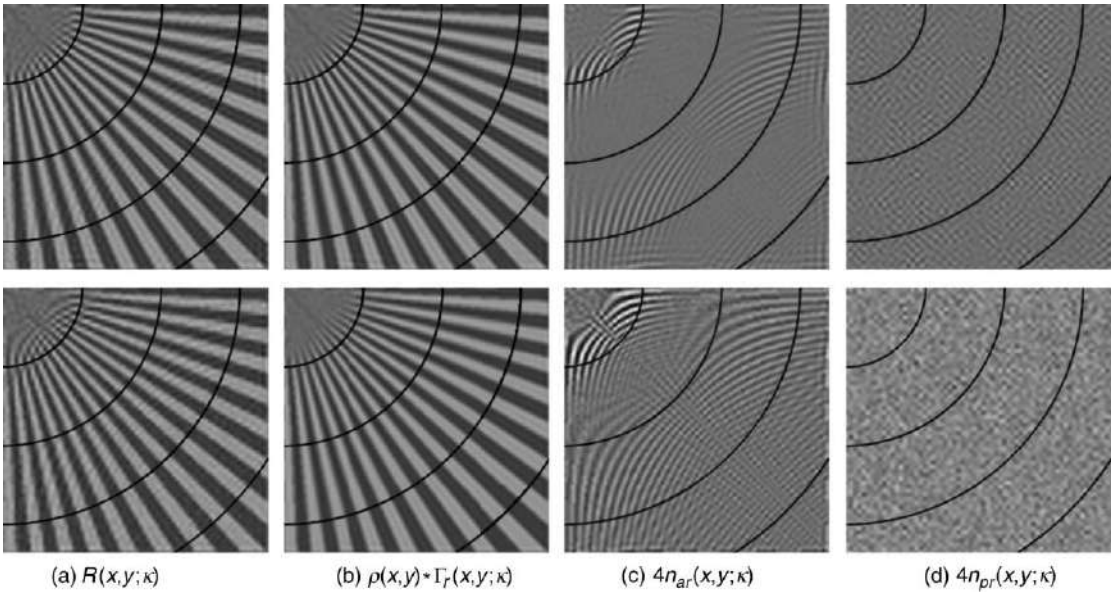

Figure 11 Images restored with the Wiener filter and their three components for designs 1 (top) and 3 (bottom).

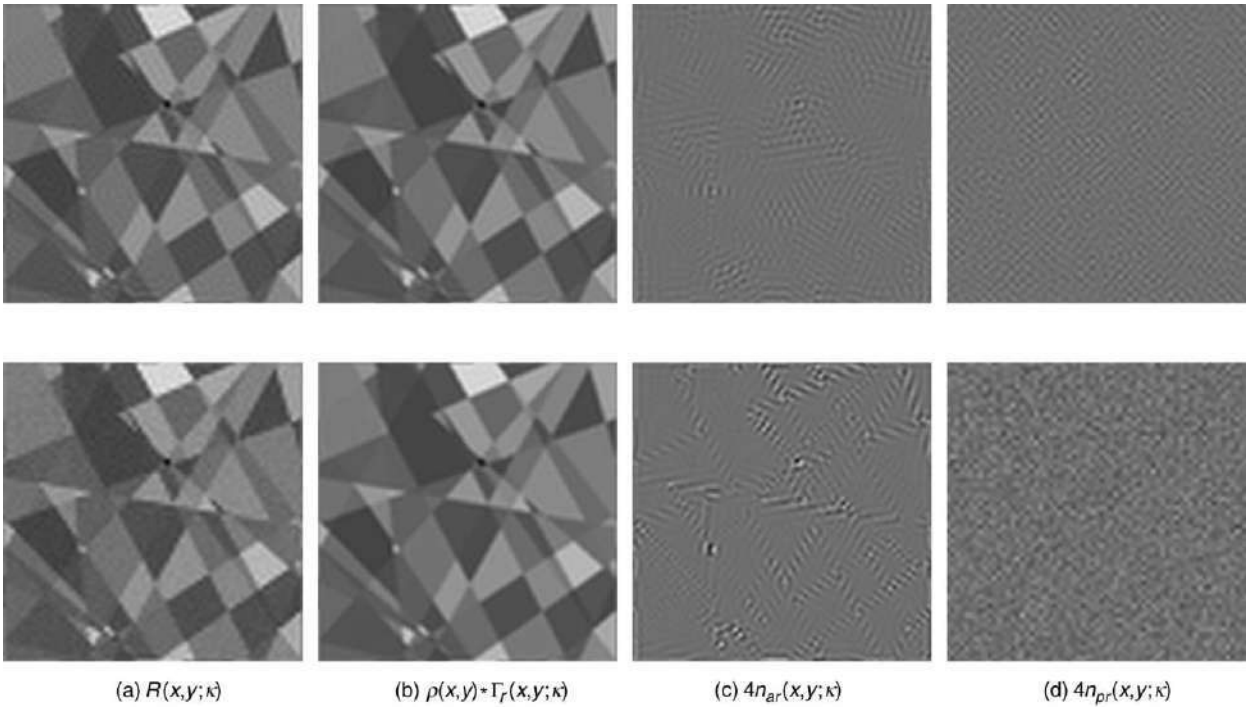

Figure 12 Images restored with the Wiener filter and their three components for designs 1 (top) and 3 (bottom).

photodetector noise. The contrast of the illustrations of the two noise components has been increased by a factor of four to improve their visibility. The extraneous structure that aliasing produces in the images of the resolution wedges is commonly referred to as Moiré pattern. The corresponding distortion in the images of the random polygons emerges more subtly as jagged (or staircase) edges.

The images are formed from a set of 64×64 samples with a $Z = 4$ interpolation and have a 4×4 cm format. If these images were displayed in a smaller format, then the jagged edges could not be resolved by the observer and would appear to be the result of blurring or photodetector noise instead of aliasing. Aliasing has, therefore, often been overlooked as a significant source of image degradation.

The Wiener restoration produces images in which fine spatial detail near the limit of resolution normally has a high contrast approaching that in the scene. However, these images also tend to contain visually annoying defects, due to the accumulated noise and the ringing near sharp edges (Gibbs phenomenon). Hence, it is often desirable to combine this restoration with an enhancement filter that gives the user some control over the trade-off among fidelity, sharpness and clarity.

Figure 13 presents images for which the Wiener restorations given in Figures 11 and 12 are enhanced for visual quality by suppressing some of the defects due to aliasing and ringing. This enhancement improves clarity at the cost of a small loss in sharpness. The images are produced again from 64×64 pixels, but they are displayed in three different formats. The smallest format, with a density of 64 pixels/cm, corresponds to a display of 256×256 pixels in a 4×4 cm area, which is typical of the images found in the prevalent digital image-processing literature. As can be observed, this small format leads to images that are indistinguishable from each other because they hide many of the distortions that larger formats clearly reveal.

In general, if the images are displayed in a format that is large enough to reveal the finest spatial detail near the limit of resolution of the imaging system, then distortions due to the perturbations in this system become also visible. When no effort is spared to reduce these distortions without a perceptual loss of resolution and sharpness, then it becomes strongly evident that the best visual quality with which images can be restored improves with increasing information rate, even after the fidelity has essentially reached its upper bound. This improvement continues until it is gradually ended by the unavoidable compromise among sharpness, aliasing, and ringing as well as by the granularity of the image display.

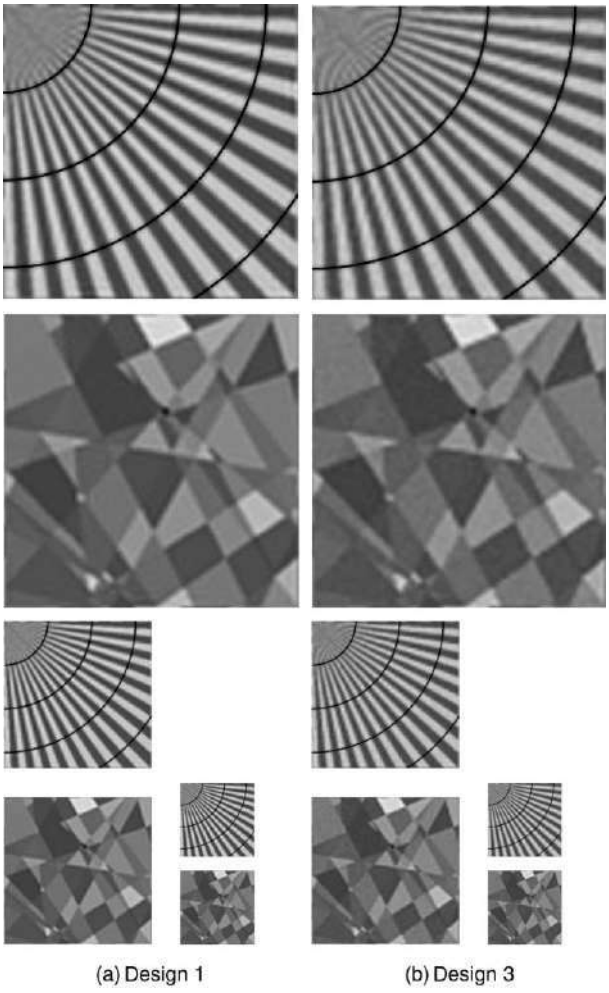

Figure 13 Images restored with the Wiener filter and enhanced for visual quality. The large, medium, and small formats contain 16, 32 and 64 pixels/cm, respectively.

Concluding Remarks

The performance of image-gathering devices and the human eye is constrained by the same critical limiting factors. When these factors are accounted for properly, then it emerges that the design of the image-gathering device that is optimized for the highest realizable information rate corresponds closely, under appropriate conditions, to the design of the human eye that evolution has optimized for viewing the real world. This convergence of information theory and evolution toward the same design clearly supports the extension of information theory to the characterization of imaging systems.

Numerous other information-theoretic characterizations of imaging are now emerging. However, none of these characterizations accounts for the critical limiting factors that constrain image gathering and restoration, as outlined here. Hence, it is not possible for them to address the efficiency and accuracy with which images can be conveyed. Moreover, none of these characterizations combines observation with communication, or sensing with processing, in a general formalization. Instead, each of them still deals with a particular measurement or imaging problem. Hence, the role of information theory in imaging is far from mature. Many challenges remain: to develop general formalizations, to account for physical phenomena, and to correlate predictions with measurements.

See also

Information Processing: Optical Digital Image Processing. **Modulators:** Electro-optics.

Further Reading

Fales CL, Huck FO, Alter-Gartenberg R and Rahman Z (1996) Image gathering and digital restoration. *Philosophical Transactions of the Royal Society of London* A354: 2249–2287.

Huck FO and Fales CL (2001) Characterization of image systems. *Encyclopedia of Imaging Science and Technology*. New York: John Wiley & Sons.

Huck FO, Fales CL and Rahman Z (1996) An information theory of visual communication. *Philosophical Transactions of the Royal Society of London* A354: 2193–2248.

Huck FO, Fales CL and Rahman Z (1997) *Visual Communication: An Information Theory Approach*. Boston, CA: Kluwer Academic Publishers.

Lesurf JCG (1995) *Information and Measurement*. Bristol, UK: Institute of Physics Pub.

O'Sullivan JA, Blahut RE and Snyder DL (1998) Information-theoretic image formation. *IEEE Transactions Information Theory* 44: 2094–2123.

Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423 and 28: 623–656.

Shannon CE and Weaver W (1964) *The Mathematical Theory of Communication*. Urbana, IL: U. Illinois Press.

Wiener N (1949) *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: John Wiley & Sons.

Inverse Problems and Computational Imaging

M Bertero and P Boccacci, University of Genova, Genova, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

A digital image is an array (2D image) or a cube (3D image) of quantized numbers, providing a representation of a physical object or a map of some distributed property of the object (density of matter, energy, etc.). It can be obtained by digitizing the original image (for instance, a photograph) or can be directly formed by the outputs of the imaging system (digital camera, microscope, telescope, etc.). When visualized on a suitable image display device, a digital image must provide a representation of the physical object which can be easily interpreted and analyzed. In many cases, however, this condition is not satisfied. We give two typical examples.

The first is that of an image degraded by blurring which can be introduced by defects of the optical systems (such as aberrations) or by image motion, image defocusing, etc. The representation of the object does not have the quality which should be desirable for its interpretation.

The second case is even more important and arises when the output of the imaging system is not directly related to the physical quantity to be imaged. For instance, in X-ray tomography, the output of a detector is the attenuation of an X-ray pencil crossing the body while the quantity to be imaged is the density of the body.

In the situations described above, the desired image can be obtained by solving a number of mathematical problems. If we denote f as the object, namely the target of the imaging system, and g as the image of f provided by the system, then the problems to be solved can be summarized as follows:

- develop a physical model of the system to obtain a mathematical relationship between the object f and the image g; the computation of g from a given f is usually called the solution of the direct problem;
- solve the problem of obtaining the physical quantity f from given outputs g of the system; this second step is usually called the solution of the inverse problem.

The digital image of f, obtained by solving an inverse problem, is a computed one, generated by the computer where the algorithm for the solution of the inverse problem has been implemented.

It may be convenient to formulate both the direct and the inverse problem in terms of functions rather than arrays, cubes, etc., because such a formulation allows the use of powerful tools of functional analysis for their investigation. Next, the results obtained in the continuous case can be used for understanding the features of the corresponding discrete problem. In most important applications, for instance, the relationship between f and g is linear so that, in the discrete version of the direct problem, one obtains g by applying a suitable matrix A to f while, in the discrete version of the inverse problem, one obtains f from g by solving a linear algebraic system. However, the solution of the inverse problem is not so simple and the difficulties can be understood by looking at the corresponding continuous problem.

The basic reason relies on the fact that inverse problems are ill-posed in the sense of Hadamard: the solution may not exist; even if it exists it may not be unique; and, even if it exists and is unique, it may not depend continuously on the data. The last statement means that a small perturbation of the data, such as that caused by noise or any kind of experimental error, can completely modify the solution.

Inverse problems are ill-posed because imaging systems do not transmit complete information about the physical object. Therefore, the problem is to extract the useful information contained in the data or, in mathematical terms, to look for approximate solutions which are stable against noise. This is the purpose of a mathematical theory introduced in its general form by the Russian mathematician, Tikhonov and known as regularization theory.

We illustrate the general framework outlined above by means of a specific example – image deconvolution, also known as image deblurring, image restoration, etc. The object f is the image which should be recorded in the absence of degradation, while g is the image corrupted by aberrations or other causes of blurring. Both f and g are functions of 2D (or 3D) space variables $\mathbf{x}$ which are the coordinates of a point in the image domain. If the imaging system is isoplanatic, then it is described by a space-invariant point spread function (PSF) $K(\mathbf{x})$, which is the image of a point source located in the center of the image domain. The PSF provides the response of the system to any point source wherever it is located. On the other hand, its Fourier transform, the transfer function $\hat{K}(\omega)$ (the hat denotes Fourier transform

and ω, the spatial frequencies associated to the space variables $\mathbf{x}$), tells us how a signal of a fixed frequency is propagated through the linear system, so that the blurring can also be viewed as a sort of frequency filtering.

Under these assumptions the image g is the convolution product of object f and PSF K:

$$g(\mathbf{x}) = \int K(\mathbf{x} - \mathbf{x}')f(\mathbf{x}')\mathrm{d}\mathbf{x}' \qquad [1]$$

and the solution of the direct problem is just the computation of a convolution product.

The image g, as given by eqn [1], is the so-called noise-free image. The actually recorded image g_r is affected by a noise term and, in general, it can be written in the following form:

$$g_r(\mathbf{x}) = g(\mathbf{x}) + n(\mathbf{x}) \qquad [2]$$

where $n(\mathbf{x})$ is a function describing noise contamination. It is a random function and, therefore, it is unknown even if one understands its statistical properties. We point out that eqn [2] does not imply that the noise is additive or signal independent. For instance, in microscopy and astronomy, the contamination of the image is due both to photon noise, which satisfies Poisson statistics, and read-out noise, which is basically Gaussian and white. Therefore the noise term in eqn [2] must only be intended as the difference between the noisy and the noise-free image.

In a first attempt at approaching the inverse problem, it seems quite natural to ignore the noise term $n(\mathbf{x})$ and to solve eqn [1] for $f(\mathbf{x})$ with $g(\mathbf{x})$ replaced by $g_r(\mathbf{x})$. Then, by taking the Fourier transform of both sides of this equation and using the well-known convolution theorem, one finds the following relationship:

$$\hat{g}_r(\omega) = \hat{K}(\omega)\hat{f}(\omega) \qquad [3]$$

which relates the Fourier transform of f and g to the transfer function of the imaging system.

The solution of this equation looks elementary. However, a first problem is due to the fact that an optical system is generally band-limited; its band Ω is the bounded domain of spatial frequencies where the transfer function is different from zero. Since the noise is not band-limited or, at least, has a band much broader than that of the optical system, outside Ω the right-hand side of eqn [3] is zero while the left-hand side is not; in other words, the solution of the deconvolution problem, as formulated above, does not exist because no object f satisfies eqn [3] everywhere. In this particular case, one can circumvent the problem by applying a suitable filter to $g_r(\mathbf{x})$, in order to suppress the out-of-band noise.

The second problem is the nonuniqueness of the solution of the problem with the filtered data. This is due to the so-called invisible objects, namely objects whose Fourier transform is zero on Ω, so that the corresponding images are exactly zero. If we find a solution of the deconvolution problem, by adding an arbitrary invisible object to this solution, we find another solution of the same problem. A well-defined solution can be obtained by requiring its Fourier transform to be zero outside Ω. In other words, we set to zero what is not transmitted by the imaging system.

The standard way for approaching the previous questions is to look for solutions in the least-squares sense, namely for objects which minimize the functional:

$$\varepsilon^2(f) = \|K * f - g_r\|^2 \qquad [4]$$

where $*$ denotes the convolution product, as defined in eqn [1], and the right-hand side is the square of the L^2-norm (which can also be interpreted as energy) of the discrepancy between the recorded image g_r and the computed image $K * f$. If one looks for a least-squares solution with minimal energy then one automatically re-obtains the solution discussed above, namely an object whose Fourier transform in Ω is given by

$$\hat{f}_{\mathrm{inv}}(\omega) = \frac{\hat{g}_r(\omega)}{\hat{K}(\omega)} \qquad [5]$$

and is zero outside Ω. Such a solution is that provided by the so-called inverse filter.

However, there is an additional difficulty in that the transfer function usually tends to zero at the boundary of the band while the Fourier transform of the noise, hence that of g_r, does not, or tends to zero in a different way. It follows that $\hat{f}_{\mathrm{inv}}(\omega)$ is divergent or very large at the boundary of the band. We conclude that its inverse Fourier transform does not exist or, if it exists, is affected by large artifacts due to noise propagation. The latter is the typical situation in the case of digital images. In **Figure 1** we give the result of a numerical simulation showing the typical effect produced by the inverse filter. The object is a beautiful picture of a galaxy recorded by the Hubble Space Telescope (HST), while the PSF is a computed one describing the blurring of HST before installation of the optics correction. The blurred image is obtained by convolving the object with the PSF and by adding white noise. The restored image produced by the inverse filter is completely corrupted by noise.

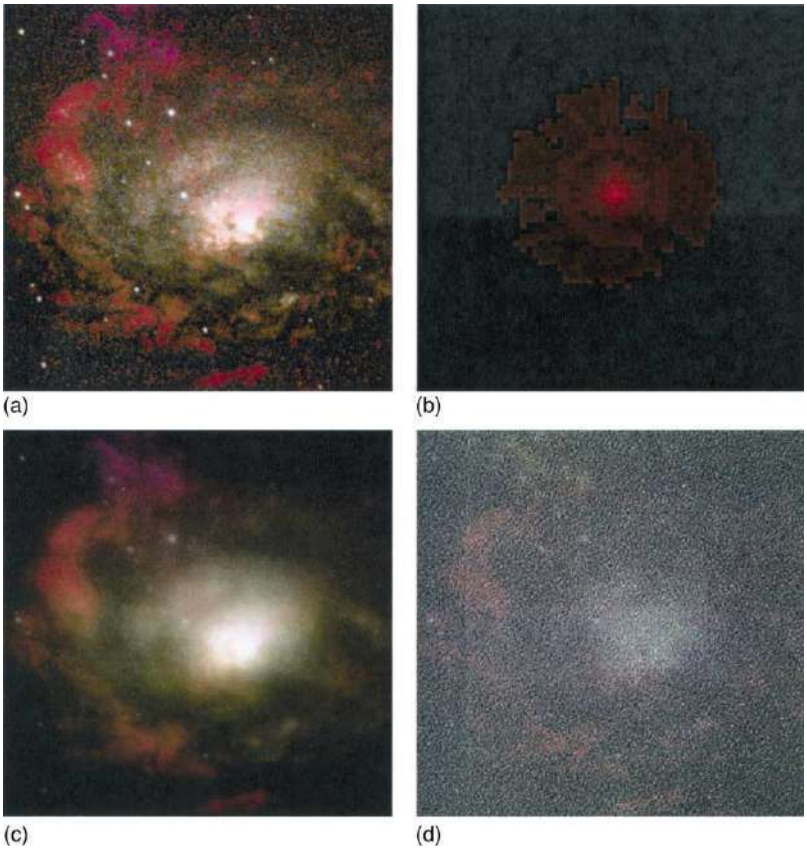

Figure 1 Illustrating the effect of the inverse filter: (a) Image of the Circinus galaxy taken on 10 April 1999, with the Wide Field Planetary Camera 2 of the Hubble Space Telescope; (b) the PSF used for blurring the RGB components of the image (here shown in red); (c) the blurred image obtained by convolving the components of the image in (a) with the corresponding PSFs and adding noise; (d) the restoration obtained by applying the inverse filter to the three components of (c). Part (a) courtesy of NASA/Space Telescope Science Institute.

The inverse filter provides a solution which fits the data in the best possible way and therefore also fits the noise in the frequency domains where the signal is weakly transmitted by the imaging system. This can be achieved at the cost of an energy (L^2-norm) of the solution which is too large. Therefore, such a conclusion suggests that one must search for a reasonable compromise between data fitting and energy of the solution. This is the basic idea of regularization theory, at least in its most simple form. The mathematical formulation is obtained by looking for objects which minimize the functional:

$$\Phi_\mu(f) = \|K * f - g_r\|^2 + \mu\|f\|^2 \qquad [6]$$

where μ, the so-called regularization parameter, is a parameter controlling the trade-off between data fitting and energy of the solution: when μ is small, the data fitting is good and the energy is large while, when μ is large, the data fitting is poor and the energy is small.

For a given μ, the function f_μ, which minimizes the functional of eqn [6] is called the regularized solution of the problem. It is easy to show that its Fourier transform is given by

$$\hat{f}_\mu(\omega) = \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2 + \mu}\hat{g}_r(\omega) \qquad [7]$$

The regularized solutions form a one-parameter family of functions. When μ is small these functions are strongly contaminated by noise whose effect is gradually reduced when μ increases. For too large values of μ, one re-obtains blurred versions of the original object.

Such behavior is illustrated by the sequence of regularized solutions given in Figure 2 and corresponding to the example of Figure 1. The regularization parameter increases starting from left to right and from top to bottom. Its values are $\mu = 0$ (inverse filter), 0.001, 0.01, 0.05, 0.1, and 0.5 (the noise is of the order of a few percent and the PSF is normalized in such a way that the sum of all its value is 1). The sequence makes evident the well-established theoretical result of the existence of an optimal value of the regularization parameter. Such an optimal value can

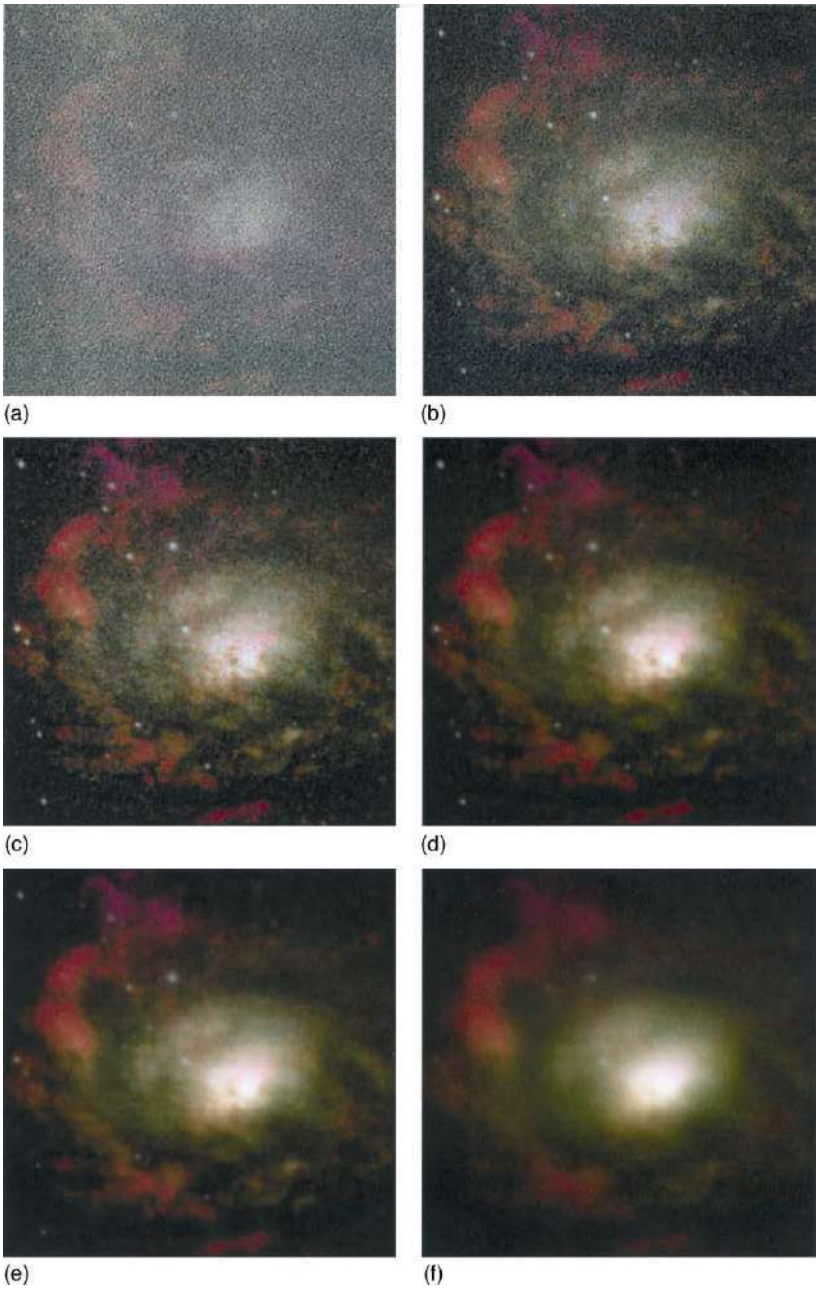

Figure 2 Illustrating the effect of the regularization parameter. The blurred image is that shown in **Figure 1c**. The sequence is obtained by increasing the value of the regularization parameter: (a) $\mu = 0$ (inverse filter); (b) $\mu = 0.001$; (c) $\mu = 0.01$; (d) $\mu = 0.05$; (e) $\mu = 0.1$; (f) $\mu = 0.5$. It is evident that the best restoration is that shown in (c).

be determined in the case of numerical simulations as that represented in Figure 2. However, its estimation in the case of real images is more difficult. Several methods for the selection of μ have been proposed. We only mention the so-called discrepancy principle. It consists of determining the value of μ, such that the value of the discrepancy functional at f_μ, $\varepsilon^2(f_\mu)$ coincides with an estimate of the energy of the noise. In other words, the data are fitted with an accuracy comparable with their uncertainty.

As shown by eqn [7], the regularized solution has the same band Ω of the imaging system, because $\hat{K}^*(\omega)$ is zero outside Ω. As is known, a band-limited function can be represented in terms of its samples taken at a rate which is roughly proportional to the size of the band. This is the content of the famous Shannon sampling theorem (more precisely its 2D extension). On the other hand, the sampling distance can be taken as a measure of the resolution provided by the restored image so that one concludes that the

restoration method discussed above does not produce an improvement of the resolution of the imaging system. It certainly produces an improvement of the quality of the image which makes possible a visual verification of that resolution.

In many circumstances, the term super-resolution is used for describing methods which allow an improvement of resolution beyond the limit provided by the sampling theorem. Therefore, a super-resolving method must produce a restored image with a band broader than that of the imaging system. Therefore, super-resolution is related to the problem of out-of-band extrapolation.

Such an approach was already proposed in the 1960s, when it was observed that the Fourier transform of an object with a finite spatial extent (all objects have this property) is analytic; then the theorem of unique analytic continuation implies that the Fourier transform of the object can be determined everywhere from its values on the band of the imaging system.

Unfortunately this problem is ill-posed, so that out-of-band extrapolation is not feasible in practice. Only recently it has been recognized that a considerable amount of super-resolution is possible when the spatial extent of the object is not much greater than the resolution distance of the instrument. For example, super-resolving methods could be used for detecting unresolved binary star in astronomical images.

The important point is that a super-resolving method must implement explicitly the finite extent of the object: the domain of the object must be estimated and the method must search for a solution which is zero outside this domain. This introduces an important question in object restoration and, more generally, in the theory of inverse problems, namely the design of regularized solutions satisfying additional conditions (constraints). This question is also important from the theoretical point of view: the problem is ill-posed because of insufficient information on the object as transmitted by the imaging system; the use of additional constraints reduces the class of the solutions which are compatible with the data and therefore can improve the restoration. This is the use of a priori information in the solution of inverse problems.

A constraint which has been widely investigated and used is the positivity of the solution, i.e., all the values of the restored image must be positive or zero. The physical meaning of the constraint is obvious. Its beneficial effect is to reduce ringing artifacts which affect the regularized solutions previously discussed, in cases where the object contains bright spots over a black background or sharp intensity variations from zero to some positive value. The ringing is essentially related to the well-known Gibbs effect in the truncation of the Fourier series.

The requirement of positivity is the requirement of a particular lower bound on the values of the solution. Therefore, one can consider more general constraints in requiring a given lower and/or upper bound on the values of the solution, eventually combined with a constraint on the domain to produce, for instance, a super-resolved positive solution. The most general form of these constraints requires that the solution belongs to a given closed and convex set $\mathcal{C}$ in some functional space. A general method producing regularized solutions in a convex set is an iterative method which is essentially a gradient method for the minimization of the least-squares functional, with a projection on the set $\mathcal{C}$ at each iteration. In the case of object deconvolution, the least-squares functional is that given in eqn [4] and the method has the following form: if f_k is the result of the k-th iteration, then f_{k+1} is given by

$$f_{k+1} = P_{\mathcal{C}}[f_k + \tau K^{\mathrm{T}} * (g_r - K * f_k)] \qquad [8]$$

where $P_{\mathcal{C}}$ is the convex projection onto the set $\mathcal{C}$, τ is a relaxation parameter and $K^{\mathrm{T}}(\mathbf{x}) = K(-x)$. In general, the algorithm, known as the projected Landweber method, is initialized with $f_0 = 0$. An important property is that it has a regularization effect, in the sense that iterations must not be pushed to convergence but stopped after a suitable number of iterations to avoid strong noise contamination. In Figure 3 we compare, in a specific example, the result obtained by means of the linear regularization method and that obtained by means of the projected iterative method with a lower and upper bound on the solution. The reduction of ringing is evident.

Among the methods producing positive solutions we must mention the most popular one in astronomy, the so-called Richardson–Lucy method, which is an iterative method for Maximum-Likelihood estimation. If the PSF has been normalized in such a way that the sum of all its values is one, it takes the following form:

$$f_{k+1} = f_k\left(K^{\mathrm{T}} * \frac{g_r}{K * f_k}\right) \qquad [9]$$

The method is normally initialized with a uniform image and it must not be pushed to convergence to avoid noise amplification. Another algorithm suggested by statistical arguments and producing positive solution is the so-called Maximum Entropy Method (MEM).

The image restoration methods described above have wide applications both in microscopy and

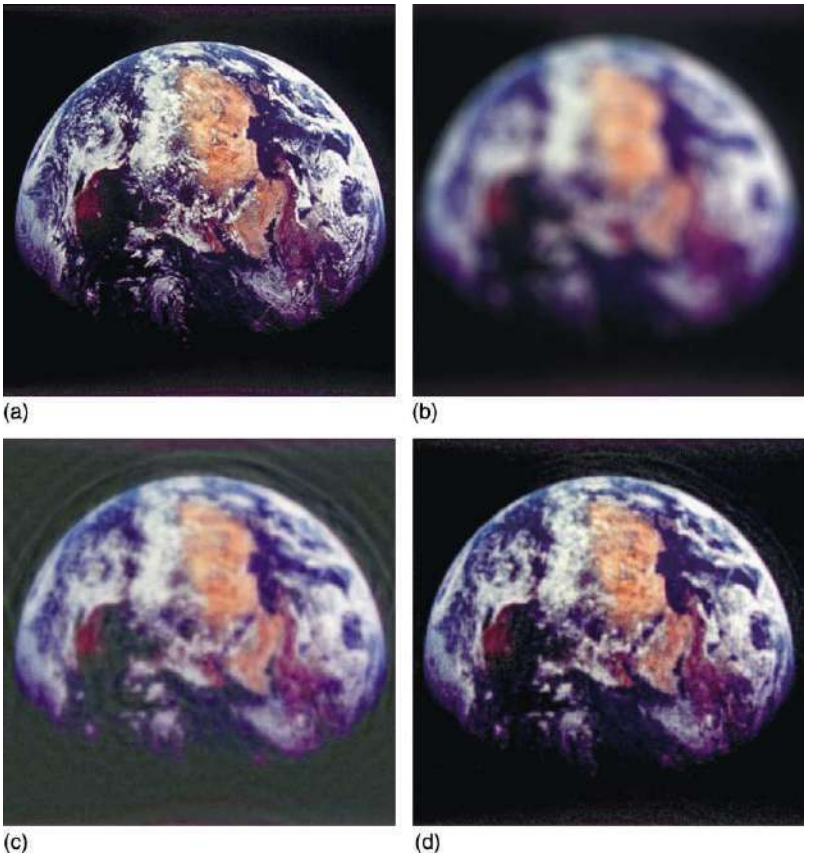

Figure 3 Illustrating the effect of constraints on image restoration: (a) Image of the Earth taken during the Apollo 11 mission; (b) blurred image of (a) obtained by assuming out-of-focus blur and adding noise; (c) optimal restoration provided by the regularization method; the ringing around the boundary of the Earth is evident; (d) optimal restoration obtained by means of the projected Landweber method with a lower and an upper bound on the solution at each iteration. Part (a) courtesy of NASA.

astronomy. For instance, the minimization of the functional eqn [6], with the additional constraint of positivity, has been used for deconvolving images in microscopy to reduce the effect of the missing cone. On the other hand, the Richardson–Lucy method was widely used for deconvolving the images of HST before installation of corrective optics.

Nowadays, image deconvolution is becoming more important for the ground-based telescopes equipped with adaptive optics. Indeed, even if adaptive optics is able to provide a considerable compensation of the atmospheric blur, a further improvement can be obtained by deconvolving the detected images. The PSF is provided by the image of a suitable guide star.

Finally, it is worth mentioning the Large Binocular Telescope, in construction on the Mount Graham in Arizona, because this instrument requires image restoration methods for a full exploitation of its imaging properties. The telescope (Figure 4) consists of two 8 m mirrors on a common mount: both mirrors are equipped with adaptive optics and are combined interferometrically to reach the resolution

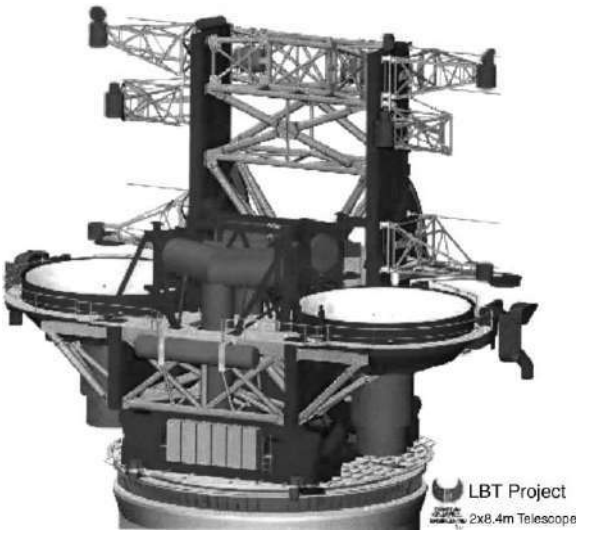

Figure 4 Picture of the Large Binocular Telescope (LBT) in construction on the Mount Graham in Arizona. It consists of two mirrors on the same mount. The images of the two mirrors are combined interferometrically to produce a high-resolution image in the direction of the line joining the centers of the mirrors (baseline).

of a 22.8 m mirror in the direction of the baseline. The telescope can be rotated to record images of the same astronomical target with different orientations of the baseline, i.e., the line joining the centers of the two mirrors. Finally the images must be processed by means of multiple images deconvolution methods to obtain a unique high-resolution image, equivalent to that produced by a 22.8 m mirror.

The applications of inverse problems with major social impact are in the area of medical imaging. The spectacular success of X-ray computerized tomography (CT), invented by GH Hounsfield at the beginning of the 1970s, has stimulated the development of other imaging modalities, based essentially on the same principle, such as positron emission tomography (PET) and single photon emission computerized tomography (SPECT). Computed images are also provided by magnetic resonance imaging (MRI). In such a case, however, the computational problem is rather simple since it consists essentially in Fourier transform inversion.

The basic principle of X-ray CT is described as a finely collimated source S (see Figure 5a) emitting a pencil of X-rays which propagates through the body along a straight line L up to a well collimated receiver R. If we know both the intensity I_0 of the source and the intensity I detected by the receiver, the logarithm of the ratio I_0/I is the integral of the linear attenuation function (roughly proportional to the density function of the body) along the line L. If the source and the receiver are moved along two parallel lines, having direction θ and defining the plane Π (in practice a planar slice of the body under consideration), one obtains the integrals of the linear attenuation function $f(\mathbf{x})$ along all parallel lines orthogonal to θ. The scanning variables in the plane Π are defined in Figure 5b: φ is the angle formed by the unit vector θ with the x-axis and s is the signed distance between the origin and the integration line L. The integrals along all parallel lines orthogonal to θ define a function of s which is called the projection of $f(\mathbf{x})$ in the direction θ:

$$P_\theta(s) = \int_{-\infty}^{+\infty} f(s\cos\varphi - u\sin\varphi, s\sin\varphi + u\cos\varphi)\mathrm{d}u \qquad [10]$$

The set of the projections of $f(\mathbf{x})$ for all possible θ is called the Radon transform of $f(\mathbf{x})$, in honor of the mathematician Johann Radon, who first investigated the problem of recovering a function of two variables from its line integrals. Today, such a problem is known as Radon transform inversion or object restoration from projections. Sampled values of the Radon transform, i.e., sampled values of the projections for a number of possible directions between 0 and π, are the outputs of the acquisition system of the medical equipment known as CT-scanners (Figure 6). Their data-processing system contains the implementation of an algorithm for Radon transform inversion and the final result is a set of computed images providing maps of slices of the human body.

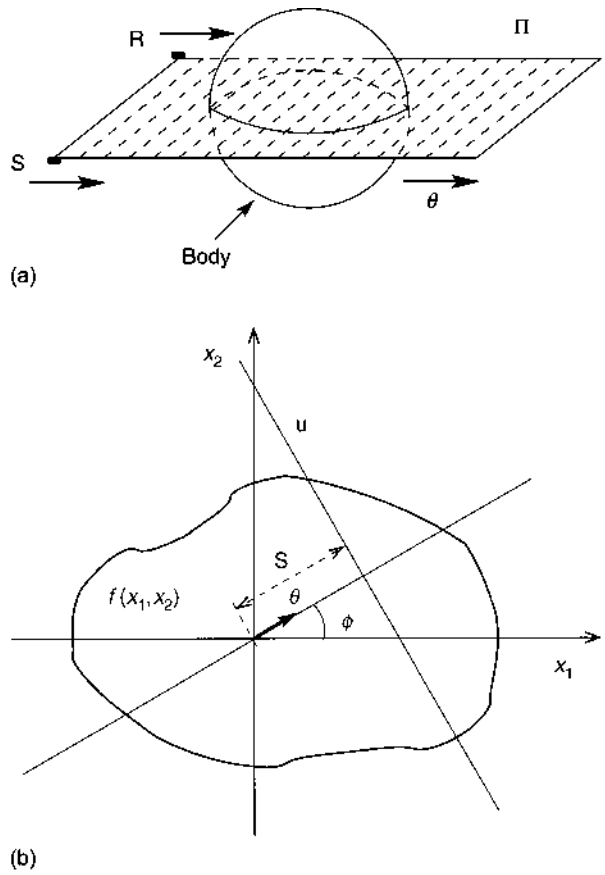

Figure 5 (a) Scheme of the scanning procedure in X-ray tomography: the source(S)–receiver(R) pair is moved along a direction θ orthogonal to the line S–R, defining the plane Π to be imaged; when the scanning has been completed the system is rotated by a certain angle, the scanning procedure is repeated, and so on. (Reproduced with permission from Bertero M and Boccacci P (1998) *Introduction to Inverse Problems in Imaging.* Bristol, UK: IOP Publishing.) (b) Definition of the variables used in the representation of the Radon transform; φ is the angle formed by the S–R line with the x_1-axis; s is the signed distance between the origin and the integration line (orthogonal to θ); and u is the coordinate on the integration line.

The plot of the Radon transform of f in the (s, φ)-plane, obtained by representing its values as gray levels, is called the sinogram of f. In Figure 7, we give the image of the sinogram of a brain slice. The horizontal rows provide representations of the projections of the phantom. In the image, the angle φ defining the direction θ takes values between 0 and 2π

while the variable s takes values between $-a$ and a, if a is the radius of a disc containing the phantom. As is evident, the sinogram is symmetric with respect to the point $(0, \pi)$. Each point in the rectangle corresponds to a straight line crossing the phantom and the straight lines through a fixed point of the phantom describe a sinusoidal curve in the (s, φ)-plane. This property has given rise to the name sinogram.

The sinogram is, in a sense, the image of the body slice under investigation, as provided directly by a CT-scanner. However, an inspection of this image does not easily provide information about the body. Therefore, it must be processed to obtain a more significant image. Radon transform inversion is an example of inverse and ill-posed problem and its solution must be treated with extra care.

The crucial point in Radon transform inversion is the Fourier slice theorem, whose content is the following: the Fourier transform of the projection in the direction θ of the function f, is the Fourier transform of f on the straight line passing through the origin and having direction θ:

$$\hat{P}_\theta(\omega) = \hat{f}(\omega\theta) \qquad [11]$$

This theorem clarifies the information content of CT data: each projection provides the Fourier transform of the function f along a well-defined straight line in the plane of the spatial frequencies. It is also the starting point for deriving the basic algorithm of data inversion in tomography, the filtered back-projection (FBP). It is obtained from Fourier transform inversion in polar coordinates and is a two-step algorithm: the first step is a filtering of the projections by means of a ramp-filter; the second is the back-projection of the filtered sinogram. More precisely the two steps are as follows:

- *Filtering*: for each θ the projection $P_\theta(s)$ is replaced by a filtered projection given by

$$Q_\theta(s) = \int_{-\infty}^{+\infty} |\omega| \hat{P}_\theta(\omega) e^{i\omega s}\, \mathrm{d}\omega \qquad [12]$$

  where the multiplication by the ramp filter $|\omega|$ derives from the Jacobian of polar coordinates;
- *Back-projection*: this operation is the dual of the projection operation: indeed the projection assigns to a straight line L with coordinates (s, φ), hence to a point of the domain of definition of the Radon transform, the integral of f along L; on the other hand, the back-projection assigns the value of the Radon transform at (s, φ) to all points of the straight line L with the same coordinates (a pictorial representation of projection and back-projection is given in Figure 8); as a consequence the back-projection operator $R^\#$ assigns to each

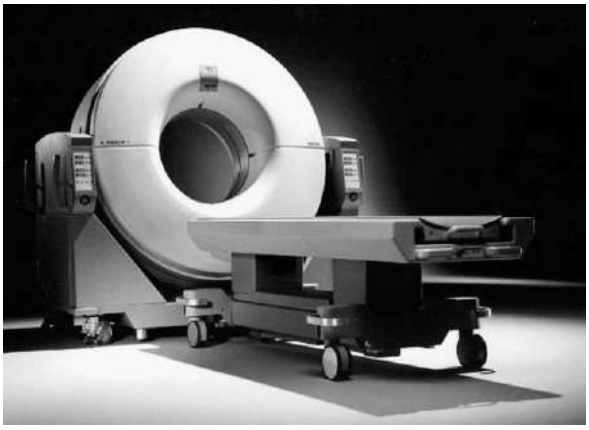

Figure 6 Picture of a scanner for X-ray tomography. The annular part contains the acquisition and scanning systems.

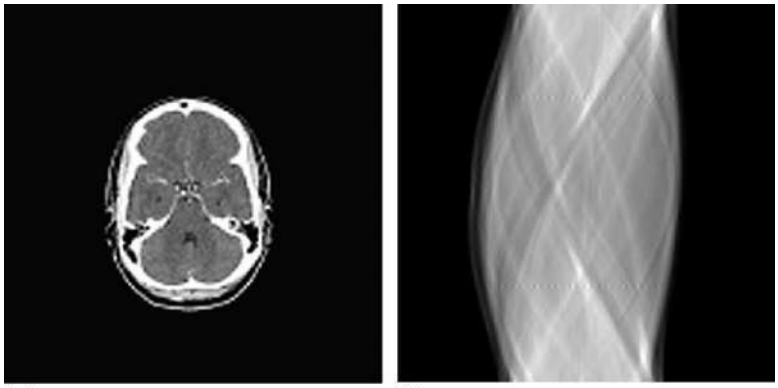

(a) (b)

Figure 7 (a) Picture of a brain slice; (b) the corresponding sinogram. As explained in the text, the sinogram is a gray-level representation of the Radon transform in the s, φ plane. Each horizontal line in (b) corresponds to a projection of (a).

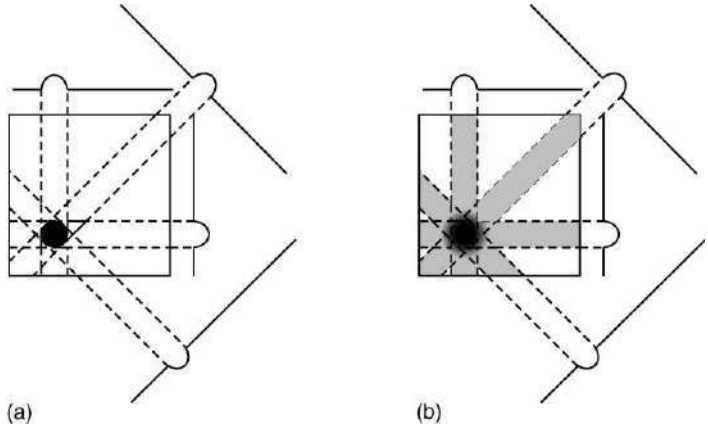

Figure 8 (a) The effect of the projection operation is shown by drawing the projections of a circular phantom corresponding to three directions. (b) shows the result obtained by applying the back-projection operation to the three projections given in (a). It is evident that the picture provided by this operation has a maximum on the domain of the circular phantom.

point $\mathbf{x}$ the sum (integral) of all values of the Radon transform corresponding to straight lines passing through $\mathbf{x}$:

$$R^{\#}P_{\theta}(\mathbf{x}) = \int_{0}^{2\pi} P_{\theta}(x\cos\varphi + y\sin\varphi)\mathrm{d}\varphi \qquad [13]$$

The back-projection operator, when applied directly to the projections as in eqn [13], provides a blurred image of f; in order to get a satisfactory restoration it must be applied to the filtered projections $Q_{\theta}(s)$. This point is illustrated in Figure 9.

We point out that the ramp filter introduced in eqn [12] amplifies the high frequency components of the projections hence the noise corrupting these components. This remark clarifies that Radon transform inversion is an ill-posed problem and requires some kind of regularization. This is obtained by introducing an additional low-pass filter which reduces the effect of the high frequencies. The most frequently used combination of ramp and low-pass filter is the so-called Shepp–Logan filter.

X-ray tomography is also called transmission computerized tomography (TCT) because the image is obtained by detecting the X-rays transmitted by the body. It provides information about anatomical details of human organs because the map of the linear attenuation function is essentially the map of the density of the tissues.

A different type of information is obtained by the so-called emission computerized tomography (ECT) which is based on the administration of radio-nuclide-labeled agents known as radio-pharmaceuticals. Their distribution in the body of the patient depends on factors such as blood flow, metabolic processes, etc. Then a map of this distribution is obtained by detecting the γ-rays produced by the decay of the radio-nuclides. Therefore, ECT yields functional information, in the sense that the images produced by ECT show the function of the tissues of the organs.

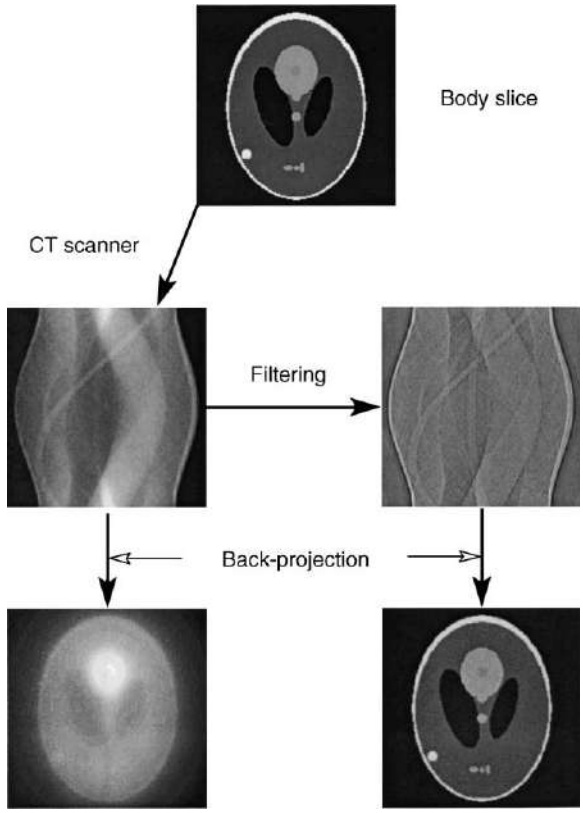

Figure 9 Illustrating the effect of the filtered back-projection. Shown here is: a slice of the so-called Shepp–Logan phantom; the corresponding sinogram (indicated by the arrow CT scanner); the filtered sinogram (indicated by the arrow filtering; the restorations obtained by applying the back-projection operation both to the original sinogram and to the filtered sinogram.

Two different modalities of ECT are usually considered:

- single photon emission computerized tomography (SPECT), which makes use of radio-isotopes where a single γ-ray is emitted per nuclear disintegration;
- positron emission tomography (PET), which makes use of β^+-emitters, where the final result of a nuclear disintegration is a pair of γ-rays, propagating in opposite directions, produced by the annihilation of the emitted positron in the tissues.

In both cases it is necessary to detect the γ-rays coming from well-defined regions of the body and, to this purpose, different methods are used in each case.

In SPECT, the discrimination of the γ-rays is obtained by means of a collimator, consisting of holes in a large lead slab covering the crystal detector face. In PET the collimation is obtained by means of pairs of detectors in coincidence. This technique, which is also called electronic collimation, is more accurate than the physical collimation used in SPECT and allows the design of systems with a great efficiency.

The basic reconstruction method in TCT, namely FBP, is often used in the reconstruction of SPECT and PET data. However, several corrections are necessary in practice. The principal ones are due to the collimator blur and to the scattering of photons in the body. If one develops a more accurate model of data acquisition by taking into account these effects, then the restoration problem implies the inversion of a very large matrix which is sparse and ill-conditioned. To this purpose iterative regularization methods are used such that the basic operation required at each step is matrix-vector multiplication. Other imaging modalities, which have been proposed for medical applications, are electrical impedance and microwave tomography. The underlying inverse problems are basically nonlinear so that the computational cost of the methods for their solution is, in general, too high for clinical applications. However the research in these areas is very active and, therefore, the situation could be improved in the near future.

See also

Tomography: Tomography and Optical Imaging.

Further Reading

Bertero M (1989) Linear inverse and ill-posed problems. In: Hawkes PW (ed.) *Advances in Electronics and Electron Physics*. vol. 75, pp. 1–120. New York: Academic Press.

Bertero M and Boccacci P (1998) *Introduction to Inverse Problems in Imaging*. Bristol, UK: IOP Publishing.

Bertero M and De Mol C (1996) Super-resolution by data inversion. In: Wolf E (ed.) *Progress in Optics*. vol. XXXVI, pp. 129–178. Amsterdam, Elsevier.

Engl HW, Hanke M and Neubauer A (1996) *Regularization of Inverse Problems*. Dordrecht, Germany: Kluwer.

Groetsch CW (1993) *Inverse Problems in Mathematical Sciences*. Braunsschweig, Germany: Vieweg.

Herman GT (ed.) (1979) *Image Reconstruction from Projections*. Berlin: Springer.

Herman GT (1980) *Image Reconstruction from Projections. The Fundamentals of Computerized Tomography.* New York: Academic Press.

Huang TS (ed.) (1975) *Picture Processing and Digital Filtering*. Berlin: Springer.

Kak AC and Slaney M (1988) *Principles of Computerized Tomographic Imaging*. New York: IEEE Press.

Krestel E (ed.) (1990) *Imaging Systems for Medical Diagnostics*. Berlin: Siemens Aktiengesellschaft.

Natterer F (1986) *The Mathematics of Computerized Tomography.* New York: Wiley.

Tikhonov AN and Arsenin VY (1977) *Solutions of Ill-Posed Problems*. Washington, DC: Winston/Wiley.

Adaptive Optics

C Pernechele, Astronomical Observatory of Padova, Padova, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The image quality of a perfect optical system is ideally limited by its diffraction figure. In the real world there are many effects which degrade that physical limit, resulting in poorer image quality at the focal plane. These effects may be static (e.g., time-independent, as, for example, error in the alignment of the optical train) or dynamic (e.g., variable with time, such as optical misalignment introduced by thermo-mechanical distortion). Active and adaptive optics, the topics of this article, are technologies able to correct for the optical aberrations induced by dynamical effects.

There are a variety of causes liable for dynamical image degradation, among these time-variable optical misalignments, pointing and tracking errors (for motorized instruments), and atmospheric perturbations. In practice, both atmospheric turbulence and misaligned instrument introduce random spatial and temporal perturbations on the optical beam, distorting the spherical (ideal) wavefront incoming on to the focal plane. This leads to loss of both the diffraction limited condition and image resolution. A typical situation for astronomical application is shown in Figure 1. A target (a star in this case) emits light whose wavefront, being far from the observing system, can be considered plane. When this wavefront passes through the atmosphere it is distorted with a temporal frequency of the order of hundreds of Hertz. This dynamically distorted wavefront fills the aperture of the optical system and will once more be distorted by the dynamic misalignment of the optical system itself. The final wavefront will be deformed with respect to the ideal (spherical) wavefront.

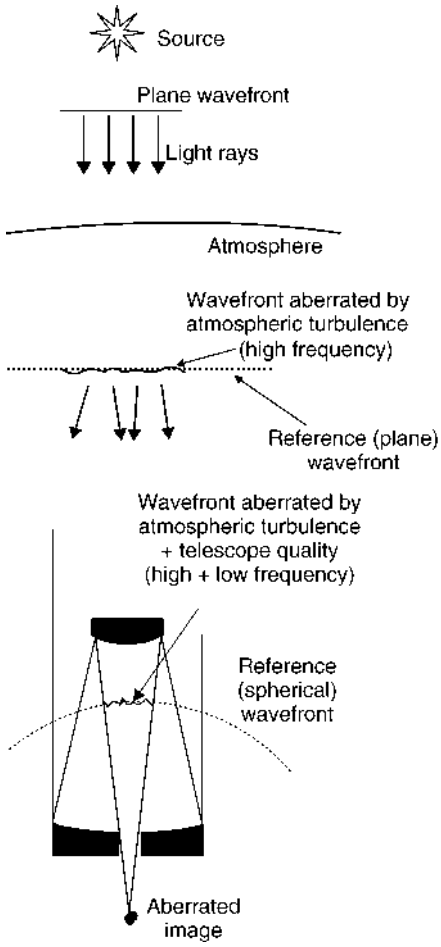

Figure 1 Dynamical wavefront aberration.

Dynamic optical misalignment may depend on gravitational and thermal deformation, while pointing error may appear as very slow (drift), slow (wander), or fast (jitter) tilt of the image on the focal plane. Atmospheric perturbations are produced by air turbulence and may be natural (as in the astronomical or satellite tracking cases) or induced by the experiment itself (thermal blooming in high power laser applications). All these effects appear on very different time-scales, as shown in Figure 2. As noticeable, atmospheric turbulence introduces aberrations of all orders, at a frequency above about 1 Hz. Pointing and tracking errors introduce an image tilt with different time-scale. Gravitational effects induce mechanical deformation of the optics, mainly in astronomical applications because of the massive components (the telescope primary mirror in particular). Thermal deformation produces optical misalignment among optical components mainly resulting in defocus. Figure 2 is purely illustrative, because some effects may produce wavefront aberrations at different frequencies than indicated there. For example, drift may be faster than indicated in the figure, and thermal blooming (a thermal effect) introduces aberrations at a higher frequency, and so on.

The correction of the wavefront errors is usually made by means of devices able to change the phase wave somewhere in the light path. Though the concept for correcting aberrated wavefront at slow or high frequency is the same, for example, the use of one or more deformable elements in the optical path, the technologies developed for each area are quite

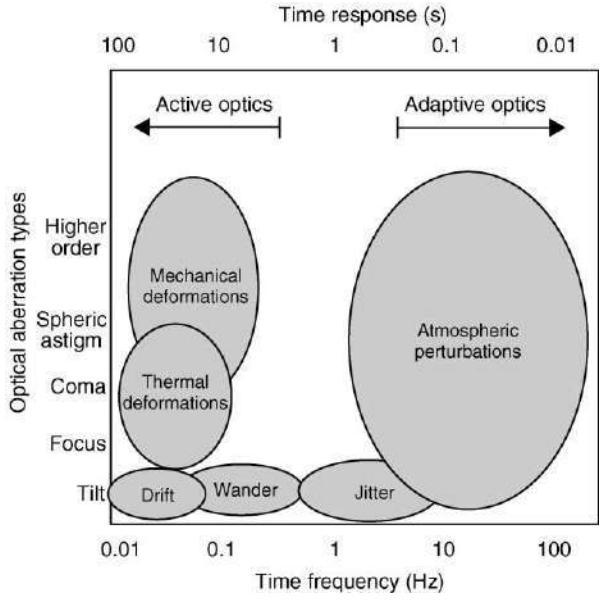

Figure 2 Wavefront aberration time-scale and associated aberrations.

different: variations at slow frequency (approximately below 0.1 Hz in Figure 2) are indicated as quasi-static errors and the techniques developed to overcome them are properly called active optics, while adaptive optics indicates the techniques to overcome the unwanted effects at high frequency (above about 1 Hz in frequency). The frequency value dividing the active from adaptive optics is not well defined and depends on the field of application and its scientific engineering communities. The important point is that a correct use of a combination of the two techniques may ultimately lead to bringing the imaging system close to its diffraction limit.

Slow Corrections: Active Optics

The analysis of the wavefront is required before its correction, and one or more deformable/movable mirrors and their control system, to operate the corrections. A typical system for low frequency correction is shown in Figure 3. A telescope contains three mirrors (M1, M2, and M3) and an instrument to detect its scientific target, usually placed on its optical axis. The dynamic deformations present in this kind of instrument are mainly third-order spherical (thermal deformation of M1), third-order astigmatism (a saddle-horse deformation of M1 due to gravity when the telescope is pointing away from the zenith), and third-order coma (due to gravitational decentering between M1 and M2). Since a telescope has a tracking system some pointing error may be present, causing motion of the image in the focal plane. The observation of an off-axis target (typically a star for astronomical applications) is made to measure the deformations of the incoming wavefront by means of a wavefront sensor (WS) system. This step is completed using some of the wavefront sensing technique described elsewhere in this encyclopedia. The WS must measure the information on the wavefront distortion with adequate spatial resolution to compensate for aberrations across the full aperture of the system. Once the wavefront distortion has been measured by the WS, the needed corrections are sent to the control system which apply the appropriate mirror deformation/movements. In the example of Figure 3 only the primary mirror (entrance pupil) is deformable and it is used for correcting spherical and astigmatism aberrations. M2 is used to compensate the decentering coma by means of lateral displacement. M3 may be tilted to correct for low-frequency pointing errors. In astronomical applications active optics is essential for telescopes with primary-mirror diameters larger than approximately four meters which cannot be manufactured rigidly at reasonable cost. Active optics requires a thin deformable or a segmented primary mirror maintained on the correct shape at every elevation. Using this method, low-frequency correction has been achieved to date in telescopes with primary mirrors of the order of ten meters.

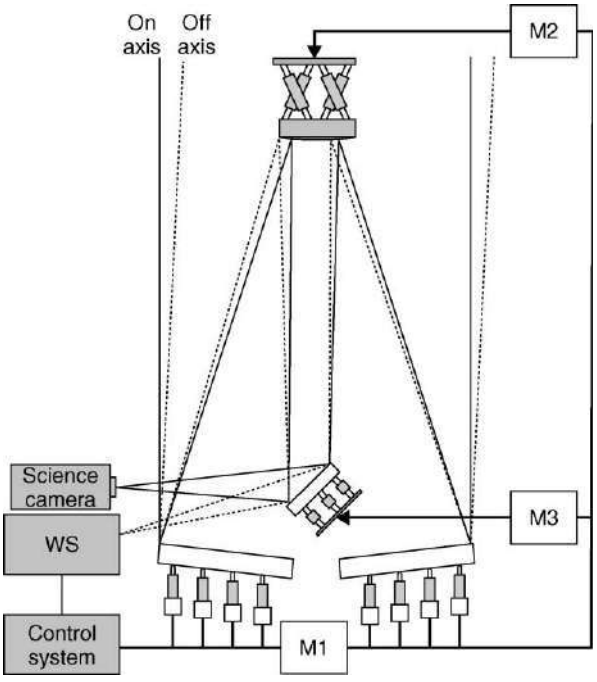

Figure 3 A typical system for correcting low-frequency dynamical aberrations (active optics).

Fast Corrections: Adaptive Optics

It is clear that corrections at high frequency are impossible using the approach discussed above, because of the large dimensions and weight of the mirror and the low response time required. For an adaptive correction it is necessary to image a smaller system pupil in the optical train and make the active correction of the wavefront. A typical configuration is shown in Figure 4. The light from the telescope focal plane (corrected for low frequency errors by means of active optics) is collimated and a pupil is located in this beam. Behind the active mirror a beamsplitter is used to send part of the light into the WS and measure wavefront distortions. The analysis results are sent to the control system which performs the active mirror deformation for the wavefront compensation. If the WS and the control system are sufficiently fast (typically faster than the time-variations of the wavefront distortion), this system is able to produce a plane wavefront which can be focused at the diffraction limit of the system. For this reason

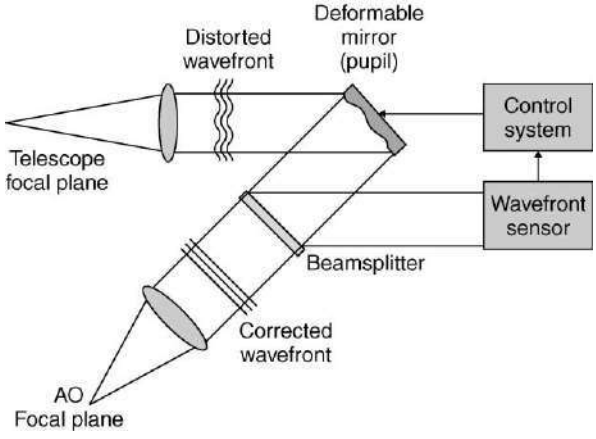

Figure 4 A typical setup for correcting high-frequency dynamical aberrations (adaptive optics).

adaptive optics are often referred to as real-time system.

The seminal paper on the adaptive optics concept was written by H.W. Babcock of the Mount Wilson and Palomar observatories in 1953. Since then a great improvement has been made, especially in military and astronomical fields. In recent years, however, many applications of adaptive optics have developed in fields different from traditional ones, such as precision manufacturing by means of high-power lasers, telecommunications, aero-optics, medical physics, and ophthalmology.

We describe adaptive optics with particular reference to astronomical applications. We provide a brief description of other applications at the end of this article.

Imaging Through the Atmosphere

The optical effects of the atmospheric turbulence are due to local temperature variations that induce fluctuations in the refractive index of air. The refractive index changes depend on the air density as well as the range of the temperature variations. Air density is highest at sea level and decreases exponentially with altitude. Optical effects of turbulence decrease with the altitude of the site, the reason for locating astronomical telescopes on mountains. The effects of turbulence on a point source image are depicted in Figure 5. Hereafter we consider an ideal telescope, for example, one without need for active correction at low frequencies. We consider the source 1 (continous line). The radiation emitted by the source is a spherical wavefront that can be treated as a plane, because of the very long distance to the Earth's atmosphere.

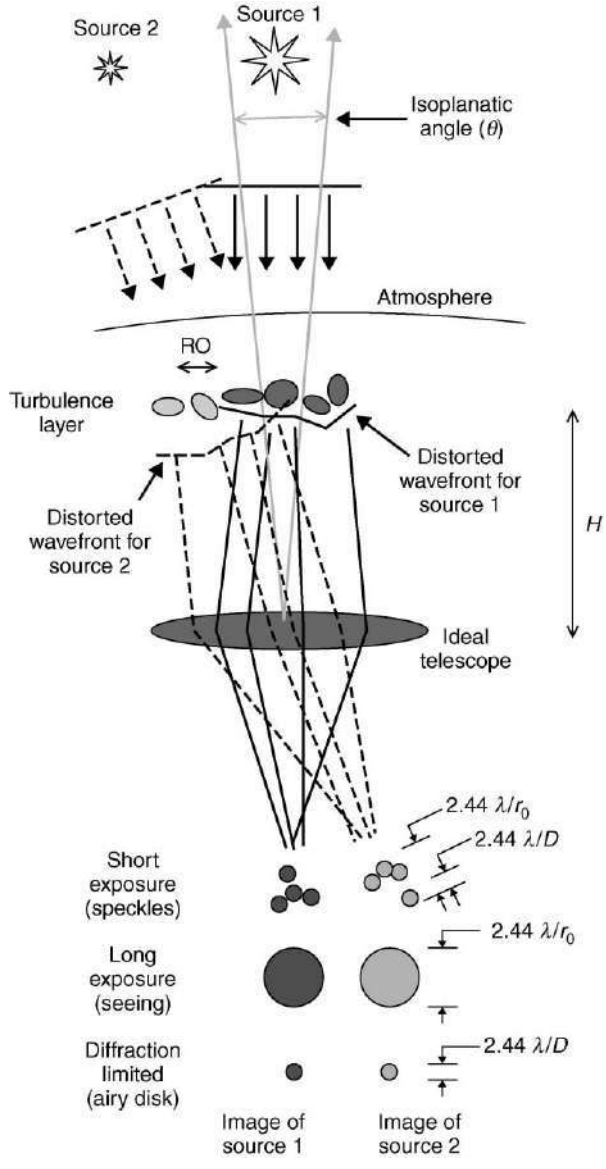

Figure 5 Imaging through the atmosphere.

In absence of wavefront distortion, the angular diameter of the image (Airy disk), is approximately 2.44 λ/D, with λ the wavelength of the observation and D the telescope aperture (the mirror diameter). In this case the image is said to be diffraction limited, and it is the best possible performance for the observation. This is the performance of a space-based ideal telescope.

In the presence of the atmosphere the incoming wavefront is distorted by the turbulence and when it reaches the telescope aperture it is no longer a plane (see in Figure 5). The so-called Fried's parameter r_0 (also called atmospheric coherence length) is the most common parameter used to quantify the distortion effects of atmospheric turbulence. It is defined as the

distance over which the wavefront phase is highly correlated (in this way it provides a measure of the turbulence scale size). For an undistorted plane wave, r_0 is infinity, while for atmospheric distortion it typically varies between 5 cm (bad sites) and 20 cm (excellent sites). We depict Fried's parameter schematically in Figure 5, as bubbles of dimension r_0 in the turbulent layer (referred to as Fried's cells). r_0 is a statistical parameter, which depends on time and wavelength ($r_0 = r_0(t, \lambda^{6/5})$). Its variability with time may be considerable, sometimes changing by a factor of 2 within a few seconds. A small telescope, with a diameter less than r_0, will always operate at this diffraction limit. When the r_0 is smaller than the telescope aperture, the angular size of the image is different if we consider long- or short-time exposures. If the exposure time is very short (less than hundredths of a second) it is possible to image a structure in the focal plane (see short exposure image in Figure 5). This structure, which varies with time, is due to high-order wavefront distortions that occur within the aperture. The short exposure image is broken into a large number of speckles of diameter 2.44 λ/D, which are randomly distributed over a circular region of approximately 2.44 λ/r_0 in angular diameter. A lower-order wavefront distortion is correlated with the size of the largest disturbances produced by atmospheric turbulence (the so-called outer scale) which varies considerably (from 1 m to over 1 km). Such disturbances produce a time-dependent tilt of the overall wavefront that causes fluctuations in the angle of arrival of the light from a star and randomly displace the image. For D of the order of the Fried parameter r_0, image motion dominates, and a simple tip-tilt mirror (not just deformable) provides useful correction. For D greater than r_0, speckles dominate and a deformable mirror is necessary to account for the wavefront distortion. For long exposures the dimension of the image is determined by the ratio λ/r_0 rather than λ/D. The angular resolution is diminished due to the atmospheric turbulence (see Figure 5). For a 4 m telescope, atmospheric distortion degrades the spatial resolution by more than one order of magnitude compared to the diffraction limit, and the intensity at the center of the star image is lowered by a factor of 100 or more. Even at the best astronomical sites, ground-based telescopes observing at visible wavelengths cannot, because of atmospheric turbulence alone, achieve an angular resolution better than telescopes of about 20 cm diameter.

Most of the distortions imposed by the atmosphere are in the wave-phase and an initially plane wavefront traveling 20 km through the turbulent atmosphere accumulates, across the diameter of a large telescope, phase errors of a few micrometers.

Turbulence is distributed along the propagation path through the atmosphere, and then the wavefront errors become uncorrelated as the field angle increases. This effect is known as angular isoplanatism and is one of the major limitations to the astronomical adaptive optics. The structure becomes completely uncorrelated out of the isoplanatic angle, approximately equal to r_0/H, where H is the altitude of the turbulence layer. This decorrelation is sketched in Figure 5, where the short exposure images for objects 1 and 2 have different speckles structure.

In Earth's atmosphere significant turbulence occurs at altitudes up to 20 km, often with large peaks in the vicinity of the tropopause, at around 10 km there exist, in practice, more than one parameter H in Figure 5, one for each turbulence layer. This three-dimensional distribution of the turbulence considerably restricts the angle over which adaptive compensation is effective and leads to the design of multiconjugate adaptive optics, as we next explain.

Adaptive System Design

As stated above, an active system is composed of at least one wavefront sensor and one deformable mirror. Typically, the wavefront is corrected by making independent tip, tilt, and piston movements of the deformable mirror surface, as shown in Figure 6. As a rule of thumb, a deformable mirror used for atmospheric correction should have one actuator for each Fried's cell, such that the total number of actuators should be equal to the number of Fried's cells filling the telescope aperture, i.e., D^2/r_0^2. For instance, a near-perfect correction for an observation in visible light (0.6 μm) with an 8 m telescope placed in a good astronomical site ($r_0 = 20$ cm), would require a deformable mirror with approximately 6400 actuators in its adaptive optics system. Recalling that r_0 is strongly dependent on the wavelength (i.e., proportional to $\lambda^{6/5}$), similar performance in the infrared at 2 μm requires only 250 actuators. This is one reason why adaptive correction is easier in the infrared. A large number

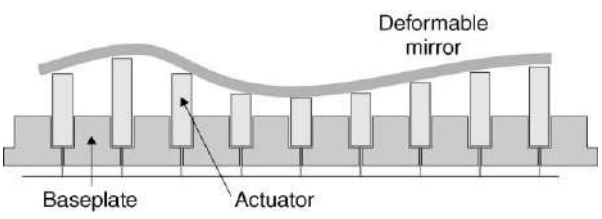

Figure 6 A sketch of a deformable mirror.

of Fried's cells requires a similarly large number of subapertures in the wavefront sensor. Commonly used WSs are based on modifications of the classic Hartmann screen test or of curvature type. In the former case the adaptive device is made with individual piezoelectric actuators (see Figure 6), while in the latter the correction is achieved with a bimorph adaptive mirror, made of two bonded piezoelectric plates. With both methods, wavefront sensing is done on a reference off-axis object (Figure 3), or even on the observed object itself (Figure 4) if it is bright enough. In the former case the angular distance of the reference object must fall inside the isoplanatic angle and must be sufficiently luminous. In typical astronomical sites the isoplanatic angle at a wavelength of 0.6 μm is ~5 arcsec, while it can be ~20 arcsec at 2 μm. This is an additional reason why the adaptive optics system is most efficient in the infrared range. Correction in the visible requires a reference star approximately 25 times brighter than in the infrared. Most current astronomical systems are designed to provide diffraction-limited images in the near-infrared (1 to 2 μm) with the capability for partial correction in the visible. However, some military systems for satellite observations do provide full correction in the visible on at least 1 m class telescopes.

The control system is generally a specialized computer that calculates, from the WS measurements, the commands to send the actuators of the deformable mirror. The calculation must be performed quickly (within 0.5 to 1 ms), otherwise the state of the atmosphere may have changed making the wavefront correction inaccurate. The required computing power can exceed several hundred million operations for each command set sent to a 250 actuator deformable mirror. As in active optics systems, zonal or modal control methods are used. In zonal control, each zone or segment of the mirror is controlled independently by wavefront signals that are measured for the subaperture of that zone. In modal control, the wavefront is expressed as the best-fit linear combination of polynomial modes of the wavefront distorted by the atmospheric perturbations.

Future Development

As explained above, a convenient method for measuring the wavefront distortion is to look for an off-axis reference star (also known as natural guide star, NGS) within the isoplanatic angle. However, it is generally impossible to find a bright reference star close to an arbitrary astronomical target. Conditions are far better in the infrared than in the visible because atmospheric turbulence (and in particular its high spatial frequencies) is, for a given AO correction, less pronounced at longer wavelengths. The spatial and temporal sampling of the disturbed wavefront can be reduced, which in turn can permit the use of fainter reference stars. Coupled with the larger isoplanatic angle in the infrared, this gives a much better outlook for AO correction than in the visible.

Nevertheless, even for observations at 2.2 μm, the sky coverage achievable by this technique (equal to the probability of finding a suitable reference star in the isoplanatic patch around the chosen target) is of the order of 0.5 to 1%. It is therefore unsurprising that most scientific applications of AO have to date been limited to the study of objects which provide their own reference object. The most promising way to overcome the isoplanatic angle limitation is by the use of artificial reference stars, also referred to as laser guide stars (LGS). These are patches of light created by the back-scattering of pulsed laser light of sodium atoms in the high mesosphere or by molecules and particles located in the low stratosphere. The laser beam is focused at an altitude of about 90 km in the first case (sodium resonance) and 10 to 20 km in the second case (Rayleigh diffusion). Such an artificial reference star can be created as close to the astronomical target as desired, and a wavefront sensor measuring the scattered laser light is used to correct the wavefront aberrations on the target object.

Military laboratories have reported the successful operation of adaptive optics devices at visible wavelengths using laser guide star on both a 60 cm telescope and a 1.5 meter telescope, achieving an image resolution of 0.15 arcsec. Nevertheless, there are still physical limitations of a LGS. The first problem, focus anisoplanatism, or cone-effect, was realized long ago. Since the artificial star is created at a relatively low altitude, back-scattered light collected by the telescope forms a conical beam, which does not cross the same turbulence-layer areas as light from the astronomical target. This leads to a phase estimation error, which in principle can be solved by the simultaneous use of several laser guide stars surrounding the targeted object. A more significant problem is image motion or tilt determination. Since the paths of the outgoing and returning light rays are the same, the centroid of the artificial light appears to be stationary in the sky, while the apparent position of an astronomical source suffers lateral motions (also known as tip/tilt). The simplest solution is to supplement the AO system and LGS with a tip/tilt corrector set on a (generally) faint close NGS. Performance is then limited by poor photon statistics in the tip/tilt correction.

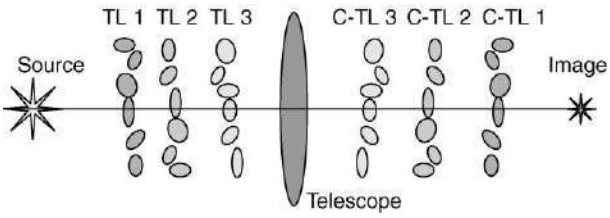

Figure 7 Conjugated planes of turbulent layers.

Adaptive optics with a multicolor laser probe is another concept that has been studied to solve the tilt determination problem of laser beacon based AO. Only applicable to sodium resonant scattering at 90 km, it excites different states of the sodium atoms and makes use of the slight variation in the refraction index of air with wavelength. Its main drawback is the limited returned flux, due to the saturation of mesospheric sodium layer.

The quality of a site for adaptive optics is determined not only by the median seeing, but also by the height and thickness of the turbulent layers (often only a few thin layers dominate). Some adaptive-optics systems can be adapted to correct for turbulence originating at a particular height by using several reference stars. This technique is known as multiconjugate adaptive optics (MCAO) and is still in his infancy. This concept deals with a 3-D view of the atmosphere, and for this reason it is also called turbulence tomography. Several wavefront sensors are needed for tomography. Turbulence tomography is also useful to correct for the cone effect when using LGS. When several deformable mirrors are placed on the appropriate conjugate planes, as shown in Figure 7, anisoplanatism reduces and the size of the corrected field of view increases. The situation is roughly depicted in Figure 7, where each turbulent layer (indicated as TL) has a proper conjugated plane (C-TL).

Another challenging optical solution is one of utilizing large secondary adaptive mirrors. This application seems especially interesting at thermal wavelengths, where any additional mirror raises the significant thermal background detected by instruments.

Other Adaptive Optics Applications

In recent years the adaptive optics technique has been developed in many applications beside astronomy and the military, for example, as in high-power laser welding or cutting, in ophthalmology and in aero-optics.

In high-power laser applications, thermal blooming takes place when laser light is absorbed by the gases and particles that make up the atmosphere. Since energy is conserved, the energy lost by the laser (its intensity diminishes) is gained by the atmosphere that heats up. This inhomogeneity at the boundary of the laser beam generates wavefront distortions by turbulence, as shown in Figure 8. This means that laser light that passes through the air at a later time interacts with a different atmosphere than the light that passed by earlier. The net effect is that, in the case of a Gaussian laser beam, light following a wavefront which passed at an earlier time, 'sees' a concave lens that deflects the successive rays out of the beam (away from the axis) causing the beam to appear to grow in radius or 'bloom'. Adaptive optics techniques for real-time compensation of blooming in laser beams were developed in the late 1960s. Given the high coherence and monochromaticity of a laser, the technique is more straightforward than for more common white-light applications of the astronomical applications.

Density fluctuations in a compressible shear layer are sufficiently high to cause a time varying index of refraction of the layer. This situation is common in a turbulent flowfield, and the study of the wavefront aberrations induced by this flowfield, when it is of relatively short propagation length in the near-field of the optical system, is referred to as aero-optics. Turbulent flowfield induces various types of degradations in the performances of an optical system. For example the jitter and defocusing of laser-based weapons cause a reduction in the amount of energy delivered to the target. Tracking systems experience bore-site errors and imaging systems experience blurring. Recent work suggests that adaptive optics compensation may successfully overcome these problems.

In ophthalmology the adaptive technique finds application in studying the retina of the eye. In this case the eye itself produces wavefront aberration, and the retina is seen as blurred. Techniques in ophthalmic laser surgery are being developed to use adaptive optics to correct for detected distortions.

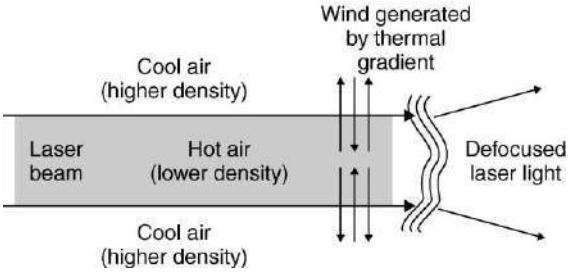

Figure 8 Thermal blooming effect in high-power laser applications.

In the field of telecommunications, studies are in progress for using adaptive techniques in data transmission between ground-based stations and satellites.

See also

Environmental Measurements: Doppler Lidar; Optical Transmission and Scatter of the Atmosphere. **Instrumentation:** Astronomical Instrumentation; Telescopes. **Phase Control:** Wavefront Coding; Wavefront Sensors and Control (Imaging Through Turbulence).

Further Reading

Beckers JM (1993) Adaptive optics for astronomy: principles, performance and application. *ARAA* 31: 13–62.

Fried DL (1965) Statistics of a geometric representation of wavefront distortion. *JOSA* 55: 1427–1435.

Hardy JW (1998) *Adaptive Optics for Astronomical Telescopes*. New York: Oxford University Press.

Jumper EJ and Fitzgerald EJ (2001) Recent advances in aero-optics. *Progress in Aerospace Sciences* 37: 299–339.

Liang J, Williams DR and Miller DT (1997) Supernormal vision and high-resolution retinal imaging through adaptive optics. *JOSA A* 14(11): 2884–2892.

Lukin VP (1995) *Atmospheric Adaptive Optics*. Bellingham: SPIE Optical Engineering Press.

Ragazzoni R, Marchetti E and Valente G (2000) Adaptive-optics corrections available for the whole sky. *Nature* 403(6765): 54–56.

Roddier F (1999) *Adaptive Optics in Astronomy*. Cambridge: Cambridge University Press.

Roggeman M and Welsh B (1996) *Imaging Through Turbulence*. Boca Raton, FL: CRC Press.

Tyson RK (2000) *Introduction to Adaptive Optics*. Bellingham: SPIE Optical Engineering Press.

Welford W (1974) *Aberrations of the Symmetrical Optical System*. New York: Academic Press.

Wilson RW and Jenkins CR (1996) Adaptive optics for astronomy: theoretical performance and limitations. *MNRAS* 268: 39.

Hyperspectral Imaging

M L Huebschman, R A Schultz and H R Garner, University of Texas, Dallas, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Hyperspectral imaging techniques on a microscopic level have paralleled the hyperspectral remote astronomical sensing which has developed over the years. It is especially apparent in remote sensing of our own planet's land masses and atmosphere, such as in the LANDSAT satellite program. In biology, biochemistry, and medicine, hyperspectral imaging and multispectral imaging are often interchanged, although there is a distinct difference in the two techniques.

A multispectral imager produces images of each of a number of discrete but relatively wide spectral bands. Multispectral imagers employ several wide bandpass filters to collect the data for the images. A hyperspectral imager produces images over a contiguous range of spectral bands and, thus, offers the potential for spectroscopic analysis of data. Hyperspectral imagers may use gratings or interferometers to produce the contiguous range of spectral bands to record.

Fluorescent dyes (also referred to as fluorophores or fluorochromes) and recently quantum dots, as markers, are a very powerful tool for analyzing organic tissues for researchers in cytology (study of cell function), cytogenetics (study of the human chromosomes), histology (study of structure and function of tissues), microarray scanning (analysis of concentrations of multiple varied probes), and biological material science. Without the correlation of the spectral and spatial data in biological images, there would be little useful information for the researcher. Therefore, hyperspectral imaging in the biological sciences must combine imaging, spectroscopy, and fluorescent probe techniques. Such imaging requires high spatial resolution to present a true physical picture, and the spectroscopy must also have high spectral resolution in order to determine the quantities of biochemical probes present. The correlation must provide the color-position relation for identification of meaningful information.

This description will begin with an initial review of multispectral imaging to lead into a discussion of hyperspectral imaging. The discussion will examine aspects of hyperspectral imaging based on an interferometric method and on a grating dispersion method similar to the methods in multispectral imaging.

Image Parameters

The analysis of fluorophores involves the excitation of the dye with a shorter wavelength and thus higher energy than the emission spectrum. The use of the appropriate excitation wavelength will yield a maximum emission intensity at a characteristic wavelength. The quality of a fluorescent image generally combines contrast and brightness. Contrast depends on the ratio of emission intensity to that of background light. Consequently, the use of filters to block the excitation wavelength and prevent it from overwhelming the field is critical in providing high contrast for the detection of fluorophore emission. However, the use of narrow filters to restrict the background can severely limit the brightness of the fluorescent image. This is due to the barrier filters allowing only a portion of the actual emission spectrum to pass. Brightness is an intrinsic property of each individual fluorophore, but is also dependent on the ability to illuminate with the peak excitation wavelength for that fluorophore.

The balance between contrast and brightness may be further complicated by applications designed to facilitate the evaluation of multiple fluorophores. The selection of appropriate excitation and emission filters within a single set to allow for the simultaneous viewing of multiple fluorophores can be very inhibiting. Although quad-band beamsplitters are currently available, commercial advances to date have yielded only triple-pass combinations for simultaneous detection, such as one of which permits the simultaneous detection of the widely used fluorophore combination fluorescein-5-isothiocyanate (FITC), Texas Red and 4′,6-diamidino-2-phenylindole (DAPI) (e.g., Chroma Technology Corp.).

Although the capture of digital images has propelled this field, efforts to collect multicolored images as digital files also suffer from technical limitations. Color CCD cameras often utilize filters or liquid crystals to segregate the red, green, and blue wavelengths onto the imaging chip. The employment of these cameras is restricted to resolving colors as discriminated by the filters used. For scientific applications, then, the greatest sensitivity and selectivity has been achieved through black and white CCD cameras with optimized filters for a specific application. While this approach is widely used, it suffers from the inherent limits of the spectra of data that can be analyzed since it is defined by the characteristics of the optical filters employed. Thus, the system is restricted to the use of fluorophores that can be effectively resolved from each other by filters, and the filters selected limit the data collected to a portion of the total emission spectrum for any one fluorophore.

Multispectral Imaging

Multispectral imaging systems in biomedical research are composed of a microscope, filters and/or interferometer for spectral selection, a charged coupled device (CCD) camera to record the data, and computer control for acquisition, analysis, and display (Figure 1a). The introduction of CCD cameras has offered a means to document results and capture images as digital computer files. This has facilitated the analysis of results with more and more complex probe sets and has allowed the use of more sophisticated computer algorithms for data analysis. The other major component of the multispectral imaging system is a light source to excite the sample's fluorescent dyes whose emissions are selected by filters or an interferometer for recording on the CCD camera.

An example of the use of multispectral imaging is in the identification of the 24 human chromosomes. A scheme was presented for the analysis of human chromosomes by differentially labeling each chromosome with a unique combination of 5 fluorophore-labeled nucleotides (building blocks of DNA and RNA, consisting of a nitrogenous base, a five-carbon sugar, and a phosphate group). These complex probes are simultaneously hybridized to metaphase chromosomes and analyzed to determine which of the fluorophores are present. Combinations of one, two, or three of the five fluorophores are enough to distinguish a particular chromosome from the others.

An interferometric multispectral system used to perform this type of analysis is the Spectral Karyotyping (chromosome typing) or SKY System (Figure 1a). This technology does not represent hyperspectral imaging, but is presented for comparison. Illumination light passes through a triple band-pass dichroic filter which allows for the simultaneous excitation of all the fluorochromes in, for instance, Speicher's scheme. Modern multi-band pass filters add little to image distortion while effectively blocking unwanted autofluorescence or excitation illumination. The fluorescence emissions from the sample are split into two separate beams by the interferometer. The beams are then recombined with an optical path difference (OPD) between them such that the two beams are no longer in register and interfere with each other. By adjusting the OPD, a particular wavelength band can be selected for, which results in constructive interference and thus can be collected by the CCD camera pixels. After adjusting the OPD to record

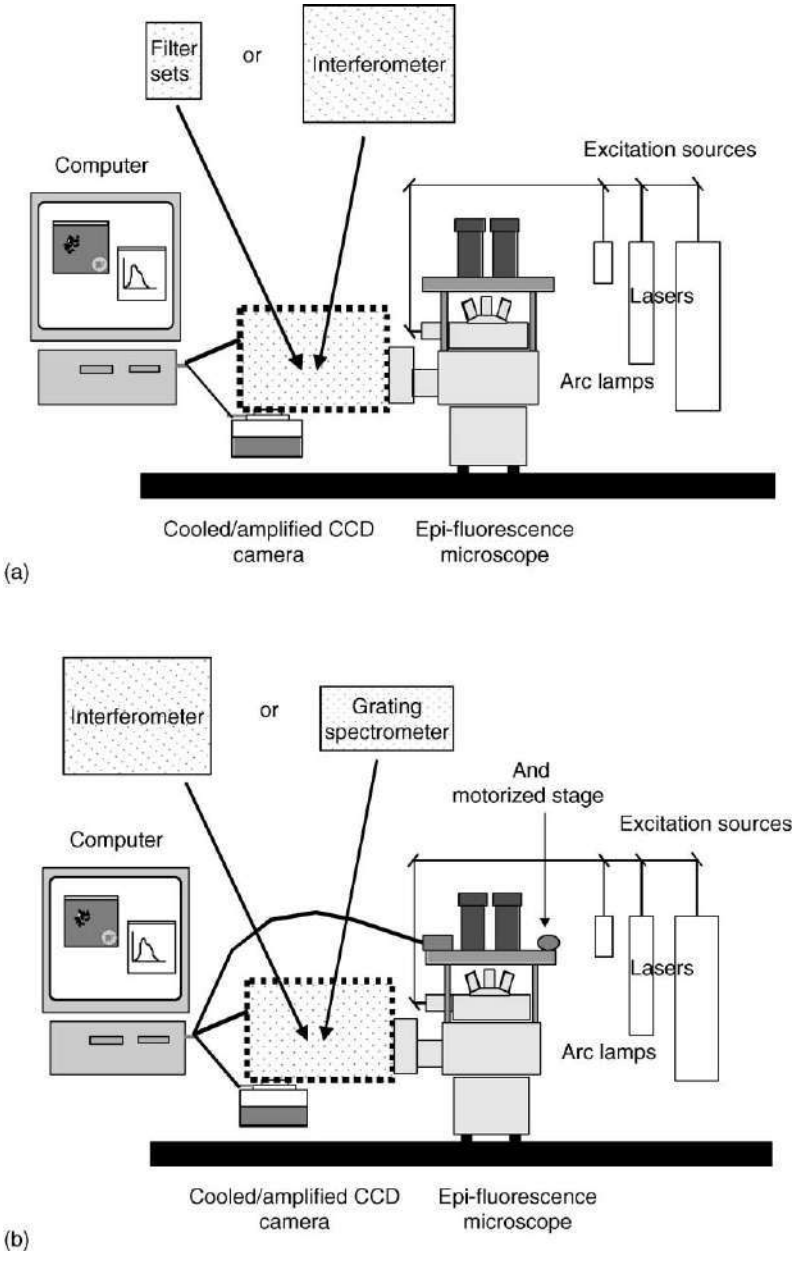

Figure 1 The instruments for multispectral and hyperspectral imaging are similar. (a) Multispectral imagers use interferometers or bandpass filters to collect noncontiguous emission wavelengths. (b) Hyperspectral imagers employ interferometers or gratings to collect a contiguous wavelength band. Computers, microscopes, CCD cameras, and excitation light sources are common to all the techniques. The systems which use filter sets or interferometers use motorized filter placement systems or automated interferometers. The gratings systems employ a motorized stage for the sample movement. Adapted with permission from Schultz RA, Nielsen T, Zavaleta JR, Ruch R, Wyatt R and Garner HR (2001) Hyperspectral imaging: a novel approach for microscopic analysis. *Cytometry* 43: 239–247.

the 5 dye wavelengths, each pixel's 5 records are then compared to the known probe-labeling scheme to identify the corresponding chromosome material. The image of the metaphase chromosome spread can be karyotyped using specific pseudo-colors for maximal visual discrimination of the chromosomes.

Another nonhyperspectral method is the Multi-Fluorescence *In Situ* Hybridization or M-FISH System. It uses five single-band-pass excitation/emission filter sets where each filter set is specific for one of the five fluorochromes used in the combinatorial labeling process (Figure 1a). Again, five separate exposures are taken (one exposure for each filter set) and combined to create a multicolor image. Most M-FISH systems now have motorized filter set selectors which minimize any image registration

shifts that may occur between exposures. For each pixel, the presence/absence of signal from each of the five fluorochromes is assessed and then matched to the probe-coding scheme to determine which chromosome material is present. As with the SKY System, an image of the metaphase spread is converted into a karyotype and specific pseudo-colors are used to maximize visual discrimination of the chromosomes.

Hyperspectral Imaging

As has been stated, the main difference between multi- and hyper-spectral imaging is the collection of a broad contiguous band of wavelengths. The apparatus and methods for hyperspectral imaging are similar to multispectral imaging (Figure 1). Instead of collecting sections of the emission spectrum using filters or interferometers, hyperspectral imaging collects a continuous emission spectrum by dispersing the wavelength of light using gratings and measuring the wavelength intensity, or by dispersing Fourier components of light using interferometers and measuring interference intensity (Figure 1b). Multispectral images are limited to measuring total intensity within the selected wavelength band. Determining the abundance of more than one dye per pixel is less accurate when only a small segment of each dyes' spectrum is passed by the emission filter. This has not been considered a limitation in the past because most research depended on just identifying the presence and not the abundance of several spectrally broad and overlapping dyes which also occupied relatively large spatial areas. With the advent of smaller and smaller featured microarrays, the desire to use more dyes in identification schemes, and the use of quantum-dot markers which can be used to indicate both genetic makeup and quantity, the limitations of multispectral imaging has led to a move to hyperspectral imaging.

In contrast to the fluorescence multispectral imaging approaches, a hyperspectral image is fully three-dimensional. Every XY spatial pixel of the image has an arbitrarily large number of wavelengths allied with it. An intensity cube of data is collected. Two of the axes of the cube are spatial and one axis is wavelength (see Figure 2). At each of the 3D points, there is an associated intensity of emitted energy. Therefore, it is a four-dimensional information space. The individual intensity signature of differing emitting bodies can then be separated to give a complete view. Spectral decomposition is a mathematical/algorithmical process by which individual emission spectra of the dyes can be uniquely separated from a composite (full spectral) image by knowing the signature of each dye emission spectra component measured separately. Spectral decomposition may be performed using standard algorithms with codes which are readily available.

Interferometric Hyperspectral Imaging

The system architecture of the interferometric hyperspectral imager resembles that of the SKY System (Figure 1). But, instead of just sequencing the interferometer to select a set of specific wavelengths to record as in the SKY System, the interferometer in the hyperspectral imager is sequenced through

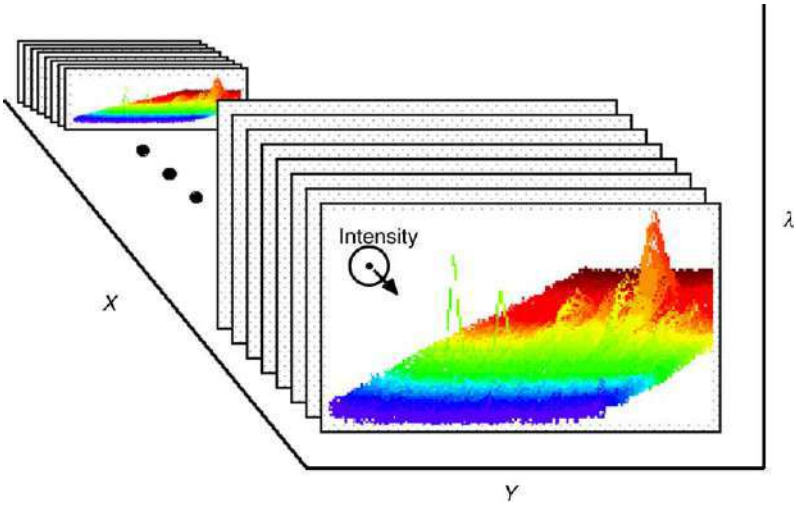

Figure 2 Intensity data cube. In this data cube, each CCD exposure captures a measurement in intensity (number of counts) by wavelength for a small spatial area in Y and ΔX. Many displacements in X with subsequent CCD acquisitions yield a 'cube' of data in X, Y, and λ. The intensity counts may be thought of as another dimension, $I(x,y,\lambda)$. Another form of the data cube is with each CCD acquisition comprising intensity counts for a spatial X and Y area for a small $\Delta\lambda$ and multiple displacements in λ for each acquisition.

discrete optical path differences to record the intensity at each pixel for each OPD. Fourier spectroscopy techniques are used to transform the OPD intensities at each pixel into the wavelength intensities at each pixel. Each pixel may be thought of as a Fourier spectrometer which is not dependent on any other pixel. Then, spectral decomposition algorithms can be run to determine which dyes are present (see the next section).

Like the multifilter methods, the interferometer approach provides a real two-dimensional image that can be observed, focused, and optimized prior to measurement. But like the filtering methods, the excitation source must be filtered for sufficient contrast to view the fluorescent emissions. Interferometers have high optical throughput, relative to other methods, and can have high and variable spectral resolution. The spectral range can cover from the ultraviolet to the infrared. The high and variable spectral resolution allows the user to trade off sensitivity, spectral resolution, and acquisition time in a way that is not possible with any other technique. And interferometers do not cause polarization of light as a filter may.

However, interferometric instruments are susceptible to image registration and lateral coherence difficulties. They are limited by the extreme precision required of the optical components which results in the continuous need for monitoring and adjustment, since aberrations are not tolerated in interferometric instruments. The throughput is reduced by loss due to the beamsplitter design and from the mirrors. Additionally, the Fourier transform instrument deconvolutes based on a two-step process; the correlation of the light passed by an interferometer followed by an inverse Fourier transform in software.

To mitigate the extreme precision and sensitivity of the most interferometric systems, Sagnac interferometers have been employed because of their common path intrinsic stability. The common path compensates any misalignment, vibration, or temperature affect. The OPD is provided by the Sagnac affect. When the Sagnac interferometer is rotated, the path traveled in one direction will be different from that of the light on the common path in the other direction. Although, this interferometer method is trading one difficulty for another, it is somewhat easier to manage the rotation of a physical system than to maintain the exacting precision of mirror alignments in a standard interferometer.

Spatial Resolution

The spatial resolution of a Sagnac Hyperspectral Imager is set by the size of the pixels of the CCD camera and the microscope system magnification, since the interferometric hyperspectral viewer sees the entire field of view at once. Let the x- and y- dimensions of the CCD pixel be δx and δy, respectively, and the magnification be M, then the spatial resolutions in the x and y directions are $2\delta x/M$ and $2\delta y/M$, respectively. For CCD pixel sizes of 10 microns square, binning together two pixels in the x- and two pixels in the y- direction, and an overall magnification of 60, the x and y resolutions are both 0.66 microns.

Field of View

For N pixel rows and C pixel columns, the field of view will be $N\delta y/M$ and $C\delta x/M$ for the length of the area along the y- and x- axis, respectively. For a CCD array of 1536 columns by 1024 rows, and the parameters above, the field of view is 171 microns by 256 microns, x and y, respectively.

The sensitivity, dynamic range, and spectral response are set by the CCD camera optical components.

Spectral Resolution

Spectral resolution will depend on the ability to vary the OPD and the Fourier transform. While the interferometer is used basically as a filter in the multispectral imager to select a wavelength, in the hyperspectral imager it is used to vary the OPD for all the wavelengths whose energy is integrated at the CCD camera pixels.

The electric field, E, for emissions at one wavelength, λ, or wave number, $k = 1/\lambda$, radial frequency, ω, and amplitude, A, can be written as

$$E(k) = A(k)\,\mathrm{e}^{-\mathrm{i}(2\pi kx - \omega t)} \qquad [1]$$

and the intensity which is measured is the magnitude squared of the electric field:

$$I = E^{*}E \qquad [2]$$

Emissions from fluorescent sources are usually noncoherent and composed of a range of wavelengths, even with only one dye present. The intensity measured at a pixel is the superposition of many wave numbers. After traveling a certain OPD, L, the intensity integrated over all wave numbers may be written as

$$I_{\mathrm{pxl}}(L) = \frac{1}{2}\left[\int I(k)\mathrm{d}k + \int I(k)\cos(2\pi kL)\mathrm{d}k\right] \qquad [3]$$

or

$$I_{\mathrm{pxl}}(L) = C_0 + C_1 \cdot \mathrm{Re}\{\mathrm{FT}[I(k)]\} \qquad [4]$$

where FT stands for Fourier transform.

The inverse transform is

$$\int I_{\mathrm{pxl}}(L)\,\mathrm{e}^{-\mathrm{i}2\pi k' L}\,\mathrm{d}L = C_0 \int \mathrm{e}^{-\mathrm{i}2\pi k' L}\,\mathrm{d}L + C_1 \int\int I(k)\mathrm{e}^{-\mathrm{i}2\pi k L'}\,\mathrm{d}k \cdot \mathrm{e}^{\mathrm{i}2\pi k' L}\,\mathrm{d}L \qquad [5]$$

$$\int I_{\mathrm{pxl}}(L)\mathrm{e}^{-\mathrm{i}2\pi k' L}\,\mathrm{d}L = C_3 + C_4 I(k)$$

The integration constant, C_3, can be ignored for large wave numbers, as in the case of the visible range, since C_3 will be proportional to the inverse of the wave number. The intensity as a function of wave number is then

$$I(k) = C_5 \int I_{\mathrm{pxl}}(L)\mathrm{e}^{-\mathrm{i}2\pi k L}\,\mathrm{d}L \qquad [6]$$

For absolute quantitative measurements, the constant C_5 can be determined by calibration with a known emission. For discrete measurements made by changing the optical path length, then the working formula is

$$I_{\mathrm{pxl}}(k) = \sum_{L_i} I_{\mathrm{pxl}}(L_i)\mathrm{e}^{-i2\pi k L_i}\,\Delta L \qquad [7]$$

This is the discrete Fourier transform which must be preformed for each pixel to determine the intensity as a function of wavelength. Per the Nyquist sampling tenets:

$$\Delta L = \frac{1}{(N/2)\Delta k} \qquad \text{or} \qquad \Delta k = \frac{2}{N\Delta L} \qquad [8]$$

where Δk is the discrete wave number interval and N the number of different OPDs. Sagnac interferometers can make rotations which give at least a change of $\Delta L = \lambda/4$, so substituting in eqn [8]:

$$\Delta k = k_1 - k_0 = \frac{1}{\lambda_1} - \frac{1}{\lambda_0} = \frac{\Delta\lambda}{\lambda^2 + \lambda\Delta\lambda} = \frac{8}{N\lambda} \qquad [9]$$

$$\frac{\Delta\lambda}{\lambda^2} \approx \frac{8}{N\lambda} \qquad \text{or} \qquad \frac{\Delta\lambda}{\lambda} \approx \frac{8}{N}$$

A 5% resolution in wavelength (or 20 nm at 400 nm) requires $N = 160$ and, of course, a 2.5% resolution requires $N = 360$. Therefore, about 400 OPD changes and acquisitions of the intensity fringes will give calculated resolutions of less than 10 nm in wavelength.

Analysis of the spectrum based on this resolution can be conducted to determine which dyes are present. Software techniques for displaying data will be described in the next section.

Grating-Dispersion Hyperspectral Imaging

A hyperspectral imaging microscope collects the complete visible emission spectrum from a microscope slide by using a grating spectrometer to provide the wavelength dispersion. This microscope correlates spatial and spectral information with minimal use of optical filters. A hyperspectral microscope system is composed of a standard epi-fluorescence microscope (Olympus IX70) equipped with objectives lens (including 1.25 × and 10 × Plan Fluorite, 40 × Plan Apochromat Dry, and 60 ×, and 100 × Plan Apochromat Oil), as well as a set of standard filter cubes (Chroma Technology) to allow for traditional visualization of dyes through the eyepiece. (Prototype typical components are listed in parentheses.) Multiple excitation sources can be utilized including standard 100 W mercury and xenon arc lamps, tungsten brightfield illumination, a 532 nm solid state laser (Brimrose), a helium neon laser (Uniphase), an argon ion laser (Uniphase), and a pulsed-doubled nitrogen dye laser (Laser Science, Inc.) (Figure 1b).

One side port of microscope is optically coupled to a spectrograph (Acton Research SpectraPro® 556i). A narrow entrance slit into the spectrograph allows for only one line (ΔX) of the image to be captured at a time. A 50-micron slit permits 80% of maximum throughput for the spectrometer. The spectrograph separates the polychromatic light illuminated from this single line into its light components. Normal operation uses a grating with 50 grooves/mm and a 600 nm-blaze wavelength, which allows wavelengths of approximately 400–780 nm to be measured. Different spectrograph gratings can be used for special applications from near UV to near IR. The center wavelength and grating can be shifted under computer control. The output of the spectrograph is then captured with an air/liquid-cooled CCD camera (Photometrics-Quantix) and an optional image intensifier (Video Scope International VS4-1845). The camera has a resolution of 1536 × 1024 pixels with a depth of 12 bits per pixel. The 1536 pixels are aligned in the spatial direction to allow the greatest spatial resolution. During normal operation the wavelength resolution is 'binned' to 128 pixels, sufficient to resolve ~8 different spectra in a single excitation scan. Higher resolution can be employed for special applications.

A line of excitation light can be provided by a line generator when laser excitation is used or by a combination of slits and cylindrical lenses when using traditional lamps or burners (mercury or xenon). This line of light illuminates the sample in the same position that maps the resulting emission light

through the imaging spectrograph and onto the camera. Some advantages of illumination with a line of light include decreased sample bleaching, reduced background due to scattering, and efficient use of available excitation power. A traditional bandpass, a low-pass dichroic filter, or a linear variable filter (Raynard Corp.) may be employed to select or restrict the excitation wavelength during specific applications.

Linear Variable Filter (LVF)

The linear variable filter (LVF) (Raynard Corporation) can be used to replace the standard excitation filter cube to provide a capability for continuous excitation wavelength band selection. The LVF is $1'' \times 2.25''$ and is coated such that it continuously filters light from ~400 nm to ~800 nm along its length. Custom filters that span other wavelengths are available. The coating and therefore wavelength is uniform along a short distance, making it an ideal method for selecting an excitation wavelength band to paint across the slide, excite the slide's contents, and map the contents' emission into the slit of the imaging spectrograph. The bandwidth passed is about 21 nm wide. The LVF introduces ~50% attenuation of the intensity.

Spatial Resolution

The spatial resolution of a grating spectrograph Hyperspectral Imager is set by the size of the pixels of the CCD camera in the y- direction and the microscope system magnification. However, differing from the interferometric system, the spatial resolution in the x- direction depends on the spectrometer slit width and the microscope system magnification. As before, let the x- and y- dimensions of the CCD pixel be δx and δy, respectively, and the magnification be M. Let w_s be the slit width of the spectrometer, and note than the slit width is always going to be larger than a few δx. Then the spatial resolution in the y-direction is again $2\delta y/M$, but the spatial resolution in the x- direction is now $2w_s/M$. Again, for CCD pixel sizes of 10 microns square, binning together two pixels in the y-direction, and an overall magnification of 60, the y- resolution is 0.66 microns, as it was for the interferometer method. For a slit width of 50 microns, the x- resolution is 1.67 microns, independent of the number of pixels binned in the x-direction. The control software moves the stage the right number of steps to minimize the overlap or skipped areas. The slit width can be reduced, but at a loss of throughput or increased exposure time. The control software can compensate for different spatial resolution to display the image as 1:1.

Field of View

For N pixel rows (y- dimension) and C pixel columns (x- dimension), the field of view will be $N\delta y/M$ for the length of the area along the y-axis. If the spectrometer slit length is less than the height of the CCD, then the length is just h_s/M, where h_s is the slit length. We orient our CCD array so that the 1536-pixel dimension is along the y-axis. The y-view is thus 256 microns as in the interferometric example. To achieve the same field of view as the Sagnac method imager, we would need to take 103 stage movements and acquisitions. This highlights the difference in the two hyperspectral methods. The grating method collects the complete spectrum on each acquisition with no spectrometer adjustment but needs to step down the slide to produce the full spatial image. The interferometer method collects the complete spatial image on each acquisition with no position adjustment but needs to step the interferometer through a range of optical path differences to acquire multispectral content.

The sensitivity, dynamic range, and spectral response are set by the CCD camera optical components.

Spectral Resolution

Spectral resolution depends on spectrometer instrument function which includes the slit width, the dispersion of the grating, the CCD pixel size, and number of pixels binned together. The dispersion for normal operation was determined to be ~40 nm/mm or 0.4 nm/pixel in the 10 micron CCD pixel example. For example, binning of 8 pixels yields a spectral resolution of 3.2 nm at all wavelengths. To achieve this spectral resolution, the interferometer method would require 1000 OPD increments and acquisitions.

Analysis of the spectrum based on this resolution can be conducted to determine which dyes are present. Although the spatial resolutions differ considerably in the two methods, as long as the dye peaks are separated by the spectral resolution spacing, the spectral decomposition algorithms will find all dyes within the spatial resolution area. We have demonstrated that the analysis algorithms can resolve 5 dyes with peaks separated by more than 6 nm in various combinations and known quantity ratios against the dye standards. The program successfully identifies the dyes present with a 91% accuracy in the quantities.

Software

Software (*Hyper*Scope, a 'C' language program) controls all the hardware components for acquisition.

It includes features for analyzing and displaying the resultant $X-Y$ images and spectral information for the grating spectrometer Hyperspectral Imager. During a scan, individual pictures are taken in the $Y-\lambda$ plane, while the stage is incremented in the X-direction. A collection of such $Y-\lambda$ scan files, which can total from 50–1000, is merged to build an image cube (Figure 2). Another advantage of the single-line acquisition mode is that time-dependent features can be monitored continuously. The software was designed to permit repeated collection of a single line at variable time points. The utility of this feature has been demonstrated by capturing images of fluorescent objects moving through UV-transparent capillaries.

The software was designed to display the resulting $X-Y$ plane extracted in various ways from the image cube. A graph of the emission spectrum for any pixel within that image may be viewed by clicking on any pixel of interest in the $X-Y$ image (Figure 3c). A standard linear curve-fitting algorithm is used to determine the contribution of individual dyes to the measured emission spectrum. The software also features a windowing technique (integration of the emission spectra over a fixed window) which emulates standard filtering. An overlay feature within the software allows the curve fitting results to be displayed in false color and compared by viewing the composite for multiple fluorophores in a single image. Additionally, it permits the contribution of a fluorophore to be emphasized for visualization purposes by either displaying the contribution of a single fluorescent signature in the absence of the others that were present in the image or by scaling individual display intensity values up or down. Overlapping pseudo-colors assigned to represent each fluorophore (either by curve fitting or by windowing) may be displayed as added, averaged, maximum or minimum values. For each new fluorophore–excitation source combination, the characteristic emission spectra is acquired and added to a library of available spectra that can be used as components in the spectral decomposition. This permits the user to introduce new fluorophores at any time. Moreover, background spectra (defined as a region of interest by the user) can be acquired and used as one element of the decomposition. This background component can be turned off in the false color view (as any of the spectral components may be) resulting in enhanced noise reduction and visual contrast improvement.

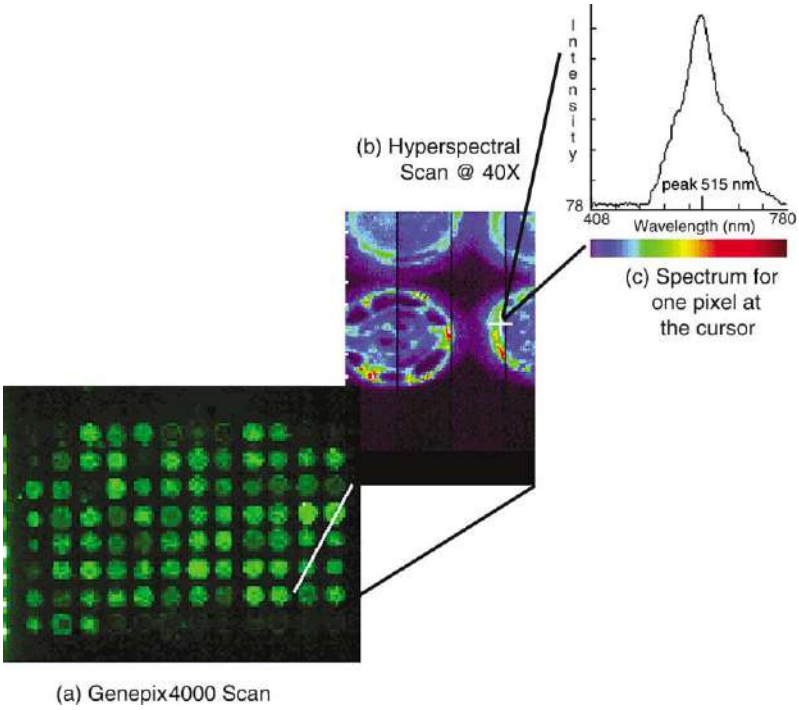

Figure 3 The benefit of spectral imaging is demonstrated in a spatial high-resolution scan of an expression microarray. (a) A portion of the scan of an expression microarray on a microscope slide using the Genepix4000 scanner. It records the intensity emitted by the Cy3 dye when excited with a 532 nm laser. (b) A portion of the scan of the same microarray with the Hyperspectral Imaging Microscope using a 40X objective and Hg excitation lamp. False color is displayed to indicate the intensity variations. (c) A *Hyper*Scope computer program display of the spectrum recorded when the cursor is placed over one pixel in a scan.

Description of Acquisition

In order to acquire a full image, a slide is placed on the microstage mover (Ludl Electronics) that allows for precise movement of the sample across the image acquisition plane. The user can position the stage via a joystick or via software to move to predetermined coordinates. The image is then scanned, with the number of scans being selected at the time of the image acquisition. Furthermore, the sample can be scanned at one magnification or resolution and then rescanned at another. Each $Y-\lambda$ image is saved as a single $Y-\lambda$ file for each increment along the sample. Thus, each image file corresponds to the excitation line of light and the emission collection region that maps through the imaging spectrograph to the CCD camera. Saving individual files permits unlimited $Y-\lambda$ image acquisition at the same Y coordinate, valuable for time-dependent studies. This feature also provides some robustness by enabling a partial $X-Y$ image to be displayed should acquisition of one or more Y scans fail for any reason. After acquiring each image, the stage is moved automatically, one or more steps in the X direction, to acquire emissions at the next line and generate the next $Y-\lambda$ file. Therefore, only the fluorescent probes in a line at a single X-location are examined at each acquisition time. Step sizes can be adjusted by the software to give: (i) the best resolution; (ii) the best 1:1 image; (iii) the minimum overlap; or (iv) the fast acquisition time.

After a variable number of rows of associated emission have been captured (typically 50–1000 scans), an image cube is loaded into memory comprising all the $Y-\lambda$ data files. A graph of the spectrum from any pixel (or bin) in the $X-Y$ plane may be displayed in a separate window by clicking on the cube (Figure 3c). A typical acquisition scan of 250 $Y-\lambda$ files, which are used to construct a $250 \times 768 \times 128$ ($\times 12$ bit) image cube, consists of approximately 100 Mb of information.

MicroArray Scanning and Customized Excitation

Another use of this technology is to analyze spotted microarrays. The Hyperspectral Imager has been compared to commercial two-color scanners with comparable sensitivity at approximately 0.5 fluors per square micron (Figure 3). Increased sensitivity (500–10 000 gain) can be provided by using a microchannel plate amplifier (Model VS4-1845, Videoscope International, Ltd.) installed between the CCD camera and the imaging spectrograph.

Since the hyperspectral imaging microscope collects images that reflect only the $Y-\lambda$ plane, it contains the potential to use a single line of light as a source for image illumination. Therefore, lights systems being developed, which are capable of generating a line of light composed of variable or complex wavelengths, could be employed to precisely control excitation. The incorporation of such an excitation source in hyperspectral imaging microscopy could be useful in many clinical and basic research applications.

See also

Environmental Measurements: Hyperspectral Remote Sensing of Land and the Atmosphere. **Imaging:** Interferometric Imaging.

Further Reading

Anguiano A (2000) Fluorescence in situ hybridization (FISH): Overview and medical applications. *Journal of Clinical Ligand Assay* 23: 33–42.

Berland K, Jacobson K and French T (1998) Electronic cameras for low-light microscopy. *Methods in Cell Biology* 56: 19–44.

Bernheim A, Vagner-Capodano AM and Couturier J (2000) From cytogenetics to cytogenomics of tumors. *S-Medical Science* 16: 528–539.

Bezrookove V, Hansson K, van der Burg M, *et al.* (2000) Individuals with abnormal phenotype and normal G-banding karyotype: improvement and limitations in the diagnosis by the use of 24-colour FISH. *Human Genetics* 106: 392–398.

Boland MV and Murphy RF (1999) Automated analysis of patterns in fluorescence-microscope images. *Trends in Cell Biology* 9: 201–202.

Cardullo RA and Alm EJ (1998) Introduction to image processing. *Methods in Cell Biology* 56: 91–115.

Garini Y, Katzir N, Cabib D, Buckwald RA, Soenksen D and Malik Z (1996) Spectral bio-imaging. In: Wang XF and Herman B (eds) *Fluorescence Imaging Spectroscopy and Microscopy*, pp. 87–122. New York: John Wiley and Sons.

Garini Y, Macville MV, du Manoir S, *et al.* (1996) Spectral Karyotyping. *Bioimaging* 4: 65–72.

Gothot A, Grosdent J and Palulus J (1996) A strategy for multiple immunophenotyping by image cytometry: Model studies using latex microbeads labeled with seven Streptaviden-bound fluorochromes. *Cytometry* 24: 214–225.

Gribble SM, Roberts I, Grace C, Andrews KM, Green AR and Nacheva EP (2000) Cytogenetics of the chronic myeloid leukemia-derived cell line K562: Karyotype clarification by multicolor fluorescence *in situ* hybridization, comparative genomic hybridization, and locus-specific fluorescence in situ hybridization. *Cancer Genet Cytogenetics* 118: 1–8.

Han H, Gao X, Su J and Nie S (2001) Quantum-dot-tagged microbeads for multiplexed optical coding of bimolecules. *Nat Biotech.* 19: 631–635.

Huebschman ML, Schultz RA and Garner HR (2002) Characteristics and capabilities of the hyperspectral imaging microscope. *Eng in Med and Bio*. 21: 104–117.
Inoue T and Gliksman N (1998) Techniques for optimizing microscopy and analysis through digital image processing. *Methods in Cell Biology* 56: 63–90.
Jain AK (1997) *Fundamental of Digital Image Processing*, pp. 80–128. New Jersey: Prentice-Hall.
Jalal SM and Law ME (1999) Utility of multicolor fluorescent in situ hybridization in clinical cytogenetics. *Genet Med*. 1: 181–186.
Johannes C, Chudoba I and Obe G (1999) Analysis of X-ray-induced aberrations in human chromosome 5 using high-resolution multicolour banding FISH (mBAND). *Chromosome Research* 7: 625–633.
Kruse F (1993) The Spectral Image Processing System (SIPS) – Interactive visualization and analysis of imaging spectrometer data. *Remote Sensing Environment* 44: 145–163.
Lu YJ, Morris JS, Edwards PAW and Shipley J (2000) Evaluation of 24-color multifluor-fluorescence *in-situ* hybridization (M-FISH) karyotyping by comparison with reverse chromosome painting of the human breast cancer cell line T-47. *Chromosome Research* 8: 127–132.
Malik Z, Cabib D, Buckwald RA, Talmi A, Garini Y and Lipson SG (1996) Fourier transform multipixel spectroscopy for quantitative cytology. *Journal of Microscopy* 182: 133–140.
Press WH, Flannery BP, Teukolsky SA and Vetterling WT (1986) *Numerical Recipes, the Art of Scientific Computing*, pp. 546. New York: Cambridge University Press.
Raap AK (1998) Advances in fluorescence in situ hybridization. *Mutat. Res. Fundam. Mol. Mech. Mutagen* 400: 287–298.
Sakamoto M, Pinkel D, Mascio L, *et al.* (1995) Semiautomated DNA probe mapping using digital imaging microscope: II. system performance. *Cytrometry* 19: 60–69.
Schrock E, duManoir S, Veldman T, *et al.* (1996) Multicolor spectral karyotyping of human chromosomes. *Science* 273: 494–497.
Schultz RA, Nielsen T, Zavaleta JR, Ruch R, Wyatt R and Garner HR (2001) Hyperspecteral imaging: a novel approach for microscopic analysis. *Cytometry* 43: 239–247.
Schwartz S (1999) Molecular cytogenetics: show me the colors. *Genet Med*. 1: 178–180.
Speicher MR, Ballard SG and Ward DC (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genetics* 12: 368–375.
Wang YL (1998) Digital deconvolution of fluorescence images for biologists. *Methods in Cell Biology* 56: 305–315.
Zhao L, Hayes K and Glassman A (2000) Enhanced detection of chromosomal abnormalities with the use of RxFISH multicolor banding technique. *Cancer Genet Cytogenetics* 118: 108–111.

Imaging Through Scattering Media

A C Boccara, Ecole Supérieure de Physique et Chimie Industrielles and Centre National de la Recherche Scientifique, Paris, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

We are often faced with the problem of seeing through scattering media such as smoke, fog, clothes, blood, tissues, ceramics, walls, etc. Tremendous effort has been made to introduce techniques which will allow us to overcome the intrinsic difficulty of selecting 'good' photons which are able to provide valuable information on structures which appear fuzzy or, for most of the time, completely hidden.

Although a number of concepts and techniques that we will discuss in this article could be applied to a large number of scattering samples, we will restrict ourselves to biological tissues.

Studying light propagation through biological tissues is becoming more and more popular among physicists, biologists, engineers, and doctors. Indeed this research domain appears to be growing rapidly as seen by the number of published articles, conferences, and international meetings. Optical techniques are attractive and are starting to compete with other well-established techniques (magnetic resonance imaging, X-ray imaging, ultrasonic imaging, positron emission tomography, etc.), because they are non-ionizing, nondestructive, they can reach high resolution and are usually much cheaper. Optical techniques reveal an optical contrast which is valuable in providing morphologic as well as functional information.

Most biological tissues are weakly absorbing in the deep red and near infrared region of the spectrum; the main difficulty in performing imaging arises from the high level of scattering. Figure 1 shows the picture of a small flash light bulb placed in the mouth: it appears fuzzy and red.

In Figure 2 we consider various photon trajectories between a short pulsed source S and a fast detector D.

Unscattered photons are called 'ballistic' photons, their number decreases exponentially, and they are completely damped after one or two millimeters.

Then there are scattered photons which propagate close to the ballistic trajectory reaching the detector; these are named 'snake-like' photons and can be used for imaging purposes.

Most of the energy which reaches D is related to the scattered photons which may exhibit very long trajectories (typically ~10 times the distance S–D).

The description of various experimental methods used to overcome the difficulty introduced by scattering will be associated with the various photon families that we have mentioned: ballistic, snake-like, and scattered photons. We will first give a few orders of magnitude of the parameters which will dictate the way light interacts with the tissues. Let us emphasize that these parameters also carry the information related to the contrast mechanism. The ultimate goal is to realize an 'optical biopsy', noninvasive but as close as possible to the regular biopsy followed by histopathology.

Figure 1 Biological tissues are highly scattering but the absorption is low in the red (as seen in this digital visible photography) and the near infrared.

Optical Parameters

Scattering Cross-Section

As shown in Figure 3, biological tissues are constituted of structures which exhibit various characteristic scales: from a few nanometers in membranes to about ten micrometers for an entire cell.

As we know from scattering theories these various scales correspond to various scattering properties. For small particles ('small' means much smaller than the optical wavelength λ) we are in the Rayleigh regime which is based on the evaluation of a time-varying dipole induced by a uniform time-varying (at the light frequency scale) electric field. This dipole radiates in the far field and its scattering cross-section, which is the ratio of the scattered power to the incoming irradiance for a sphere of radius r, is given by

$$\sigma = \frac{\langle P \rangle}{I} = \left(\frac{n^2 - 1}{n^2 + 2}\right)^2 \frac{8\pi}{3} r^6 k^4 \qquad [1]$$

n being the refractive index of the sphere divided by the index of the surrounding medium and $k = 2\pi/\lambda$. The scattering diagram for such particles is isotropic if we use unpolarized light. This cross-section is much smaller than the 'geometrical' section S of the particle; for instance, if $n = 1.5$, $a = 10$ nm, $\lambda = 0.6$ μm, $\sigma = 1510^{-8}$ μm^2, which is much smaller than the actual section (3×10^{-4} μm^2).

For larger particles the sphere (or the ellipse) interacting with a plane polarized wave is rather complicated because we are not restricted to the dipolar approximation (uniform field), but induced multipolar moments and their radiating fields have to be taken into account. This theory was proposed by Mie and can be found in optics textbooks. As an example, Figure 4 shows the scattering cross-section of a micron-size particle as a function of the

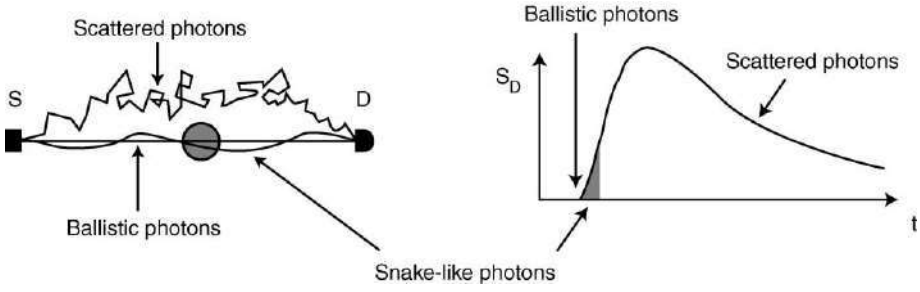

Figure 2 Between an ultrafast light source S and a detector D can be distinguished various classes of photons which can be used for imaging.

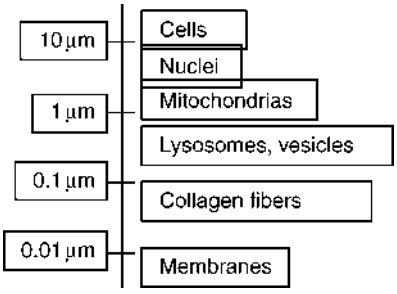

Figure 3 Various sizes of scatterers may be found in cell structures.

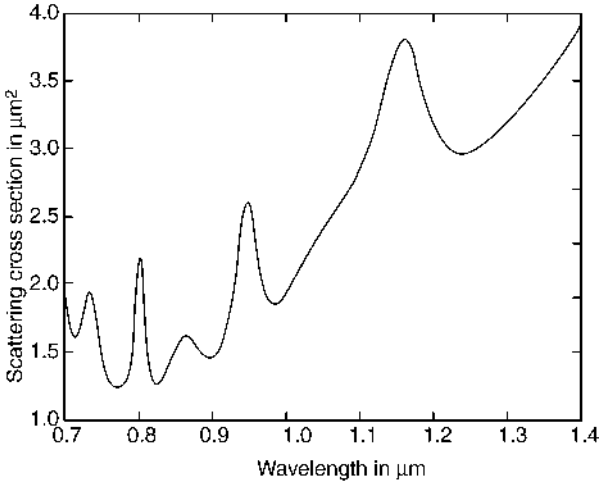

Figure 4 In the Mie scattering regime resonances are observed when varying the wavelength (here a 1 μm sphere, refractive index: 2).

wavelength: a number of 'resonances' can be observed.

The scattering diagram is no longer isotopic, it exhibits lobes, and forward scattering is much higher than back-scattering. Scattering is characterized by a scattering coefficient, μ_s, which is the inverse mean free path l_s of photons between scattering events:

$$l_s = 1/\mu_s = \sigma\rho \qquad [2]$$

where σ is the scattering cross-section and ρ the number of scatterers by unit volume.

The scattering coefficients μ_s of tissues are found within the range 100–1000 cm^{-1}. After a path length L, the number of ballistic photons will be damped by a factor $\exp(-L/l_s)$, so tissue imaging cannot rely only on ballistic photons as they survive only to a path length larger than about one mm. One needs to consider in more detail the scattering properties of the scatterers.

The dominant sizes of the scatterers in tissue (cells) demand consideration of the scattering in the Mie regime. Such a scattering event does not result in isotropic scattering angles. Instead, the scattering in tissue is biased in the forward direction.

Anisotropic scattering is quantified in a coefficient, g, which is defined as the mean cosine of the scattering angle θ, $p(\theta)$ being the probability of a particular scattering angle:

$$g \equiv \langle\cos(\theta)\rangle \equiv \frac{\int_0^\pi p(\theta)\cos(\theta)\sin(\theta)\mathrm{d}\theta}{\int_0^\pi p(\theta)\sin(\theta)\mathrm{d}\theta} \qquad [3]$$

For isotropic scattering, $g = 0$. For complete forward scattering, $g = 1$. In tissues, g varies from 0.7 to 0.95.

Photon migration, especially in the so-called diffusion regime, is often based on isotropic scattering such as for phonons (heat) or charge carriers scattering. For diffusion-like models, it has been shown that one may use an isotropic scattering model with a corrected scattering coefficient, μ_s', and obtain equivalent results using:

$$\mu_s' = \mu_s(1 - g) \qquad [4]$$

One can consider $l_s' = 1/\mu_s'$ as the length over which the incoming photon loses the memory of its initial direction, μ_s' is usually called the transport-corrected scattering coefficient.

Absorption Cross-Section

The absorption coefficient, μ_a (cm^{-1}) represents the inverse mean path length of photons before absorption. $1/\mu_a$ is also the distance in a medium at which intensity is attenuated by a factor of $1/e$ (Beer's Lambert law). Absorption in tissue is strongly wavelength dependent and is due to chromophores and water. Among the chromophores in tissue, the dominant component is the hemoglobin in blood. In the visible range, the blood absorption is relatively high, whereas it is much weaker in the near infrared. Water absorption is low in the visible and NIR region and increases above 1 μm, with tissues turning completely opaque above 2 μm. Thus, for greatest penetration of light in tissue, wavelengths in the 700–1300 nm spectrum are used. This region of the spectrum is called 'the therapeutic window'. Different spectra of chromophores allow one to separate the contribution of varying functional species in tissue such as quantification of oxy- and deoxy-hemoglobin to study tissue oxygenation.

Finally, when light is propagating in scattering media of thickness L, the damping of the incoming energy involves a combination of its absorption and

scattering properties, and is given by

$$\exp(-L/a) \qquad \text{with} \qquad a^2 = 1/3\mu_a(\mu_s' + \mu_a) \qquad [5]$$

This formula reflects the fact that the photon's pathlength is increased by scattering, leading to an increase of the damping, which is related to absorption.

These few examples provide a 'static' view of light behavior in scattering media. Studying in a more detailed way, the spatial and temporal distribution of light, is more difficult.

Photon Migration in Scattering Media

Photon migration models have been inspired by other fields, such as astrophysics, soft matter physics, and neutronics, where physical media are random in both space and time. To avoid difficulties linked to the microscopic complexity such investigations rely on a statistical basis.

The most widely used theory is the time-dependent diffusion approximation to the transport equation:

$$\vec{\nabla}\cdot(D\vec{\nabla}\Phi(\vec{r},t)) - \mu_a\Phi(\vec{r},t) = \frac{1}{c}\frac{\partial\Phi(\vec{r},t)}{\partial t} - S(\vec{r},t) \qquad [6]$$

where $\vec{r}$ and t are spatial and temporal coordinates, c is the speed of light in tissue, and D is the diffusion coefficient related to the absorption and scattering coefficients by the formula:

$$D = \frac{1}{3[\mu_a + \mu_s']} \qquad [7]$$

The quantity $\Phi(\vec{r},t)$ is called the fluence, defined as the power divided by the unit area. This equation does not reflect the scattering angular dependence; the use of μ_s' instead of μ_s ensures that isotropic scattering has been reached. Indeed, for the use of the diffusion theory which implies anisotropic scattering, the diffusion coefficient is expressed in terms of the transport-corrected scattering coefficient; $S(\vec{r},t)$ is the source term.

The diffusion equation has been solved analytically for different types of measurements such as reflection and transmission modes, assuming that the optical properties remain invariant throughout the tissue. To incorporate the finite boundaries, the method of images has been used.

The diffusion approximation equation in the frequency-domain is the Fourier transform of the time-domain equation. This leads to the equation:

$$\vec{\nabla}\cdot(D\vec{\nabla}\Phi(\vec{r},\omega)) - \left[\mu_a + \frac{i\omega}{c}\right]\Phi(\vec{r},\omega) + S(\vec{r},\omega) = 0 \qquad [8]$$

Here the time variable is replaced by the frequency ω which is the modulation angular frequency of the source. In this model, the fluence can be seen as a complex number describing the amplitude and phase of the so-called 'photon density wave' in addition to a DC component:

$$\begin{aligned}\Phi(\vec{r},\omega) &= \Phi_{AC}(\vec{r},\omega) + \Phi_{DC}(\vec{r},0)\\ &= I_{AC}\exp(i\theta) + \Phi_{DC}(\vec{r},0)\end{aligned} \qquad [9]$$

where θ is the phase shift of the diffusing wave whose wavelength is:

$$2\pi\sqrt{\frac{2c}{3\mu_s'\omega}} \qquad [10]$$

Although analytical techniques provide a rapid understanding of most phenomena involved in static or dynamic situations, real biological samples are much more complex than homogeneous isotropic media. In these cases, direct and inverse algorithms map the spatially varying optical properties.

Monte Carlo simulations are very popular nowadays; a number of useful programs are available on the websites of various groups. Results may include time or polarization dependence of the photons along their paths.

The other method used in tissue optics is the random walk theory (RWT) on a lattice. RWT models the diffusion-like motion of photons in turbid media in a probabilistic manner. Using RWT, an expression may be derived for the probability of a photon arriving at any point and time given a specific starting point and time.

Imaging in the Ballistic Regime: Taking Advantage of Various Coherence Properties

Shallow Tissue Imaging Through Selection of Ballistic Photons

When imaging thin (less than 2 mm) tissue samples, it is possible to use photons which have not been scattered, called the ballistic photons which can be used to form high-resolution images in the same manner as if no scattering had taken place.

Their number, however, decreases exponentially with propagation distance and this ballistic signal is usually hidden by multiply scattered photons that obscure the image. For relatively shallow tissue depths it is possible to use various spatial filtering techniques to block the multiply scattered photons and there is much current research that aims to

extend optical imaging and biopsy to depths of a few mm.

Confocal microscopy rejects much of the scattered light through spatial filtering and has been shown to form 3-D images at tissue depths of up to 200–300 μm, both *in vitro* and *in vivo*, which corresponds to about 50 dB of damping for the ballistic photons.

Two-photon fluorescence microscopy is particularly interesting in this regime since it ensures that all the detected fluorescence photons originate from the desired image plane. This technique also ensures a better penetration (using IR excitation) and a better resolution (nonlinear response) than single photon confocal microscopes.

Imaging to much greater tissue depths does not appear to be practical using optical microscopy and spatial filtering alone and more sophisticated techniques must also be employed to discriminate in favor of the ballistic photons. These techniques exploit either the fact that the ballistic signal photons retain spatial coherence of incident light or the fact that the ballistic photons keep the optical phase memory that multiply diffused photons have lost.

One of the most successful ballistic light-imaging techniques based on the latter filtering approach is optical coherence tomography (OCT) which uses coherence gating using low temporal coherence light and an interferometric detection. Coherent detection is used to discriminate against the multiply scattered photons that scatter back into the trajectories of ballistic photons. The ballistic light signal is measured by interference with a reference beam and the resulting pattern is detected with high sensitivity using homodyne or heterodyne detection. By using low coherence length sources and matching the time-of-flight of the reference beam signal to that of the desired ballistic light, OCT can acquire depth-resolved images in scattering media such as biological tissues. For powers in the mW range for tissue irradiation in the infrared, the unscattered ballistic (or more precisely single back scattered) component of the light will be limited by the shot noise detection limit after propagating through approximately 25 scattering mean free paths (MFP) of a scattering medium (or more than 100 dB). This corresponds to a 'typical' tissue thickness of about 1 mm tissue depth for the reflection geometry which is potentially useful for clinical applications such as screening for skin cancer, or for retinal examination.

OCT has proved clinically useful as a means of acquiring depth-resolved images in the eye and is being developed for application in strongly scattering biological tissue. Although the image acquisition requires pixel-by-pixel scanning, the use of new high-power superluminescent diodes in the 100 mW range or of ultrafast lasers to provide high average power, low coherence length radiation (sub-two cycle lasers) has resulted in an *in vivo* OCT system providing real-time imaging.

There are a number of other approaches to coherent imaging, including whole-field 2-D acquisition techniques which remove the need to scan transversely pixel by pixel.

Whole-field imaging using a temporally and spatially incoherent source and a well-balanced interferometric microscope intrinsically provides higher image acquisition rates and can take advantage of inexpensive spatially incoherent, broadband sources such as high-power LEDs, and white light thermal sources.

As an example, Figure 5 shows Linnik-type white light interferometer with optical path modulation and Figure 6 an image of an *ex-vivo* rat retina where one can recognize the main feature expected from histology.

Holography, which works in the Fourier space rather than in the object or image space, is also an interferometric technique able to discriminate between ballistic and scattered photons. Since the development of electronic holography, which takes advantage of available arrays of detectors with multimillion pixels, this approach is rapidly growing.

Real-time 3-D imaging systems based on photorefractive holography in multi-quantum-wells devices, which can potentially acquire depth-resolved images at thousands of frames per second, have been demonstrated.

These ballistic light techniques can extend the depth of tissue imaging to the mm range when

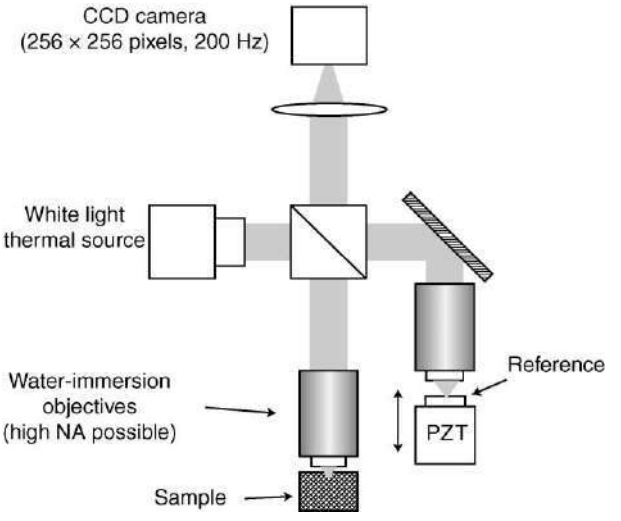

Figure 5 The full-field OCT setup based on a Linnik interferometer. The PZT actuator is used to modulate the path difference between the two arms.

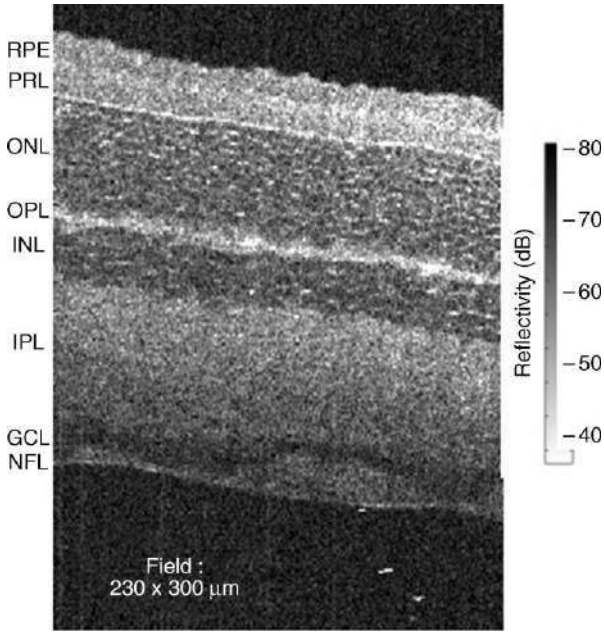

Figure 6 OCT rat retinal image: retinal pigmented epithelium. PRL: photoreceptor layer. ONL: outer nuclear layer. OPL: outer plexiform layer. INL: inner nuclear layer. IPL: inner plexiform layer. GCL: ganglion cell layer. NFL: nerve fiber layer.

imaging in reflection. When it is necessary to image through cms, rather than mms of tissue, however, there will be no detectable ballistic light signal and one must extract useful information from the scattered photons.

Imaging in the Snake-Like Regime: Taking Advantage of Time (or Frequency) Gating

The first approach is to use time gating to select the earliest arriving scattered light, which will provide the most useful image information.

For moderate tissue depths one can exploit the fact that biological tissue is highly forward scattering, with most photons being only slightly deviated from their original direction upon each scattering event. After a few cm of tissue, there can be a significant number of photons that have followed reasonably well-defined 'snake-like' paths about the original direction through the tissue, arriving at the detector after the ballistic light but before the fully diffuse photons whose trajectories are effectively randomized. This early arriving light can be selectively gated using ultrafast lasers and a variety of detectors such as streak cameras, time-gated optical image intensifiers, and time-gated fast photon counting systems, which provide picoseconds time gates.

An alternative, sometimes cheaper, approach to temporal gating is the use of (high) frequency signal discrimination (which is the Fourier transform of the pulse experiment). The periodic solution of the diffusion equation with a periodic source term gives rise to the so-called 'photon density waves'. Such diffusive waves, like thermal waves, are used in the near-field regime (their amplitudes are damped by $\exp(-2\pi)$ after propagating over a single wavelength path) and they can reveal subsurface structures through their refractive and diffusive behavior. Coherence gating and polarization gating can be used when the early arriving light still retains some degree of coherence of the incident light. Indeed polarization discrimination has been used to select those photons which undergo few scattering events and therefore preserve a fraction of their initial polarization state, from the photons which experience multiple scattering resulting in a complete loss of memory of their initial polarization state.

As an example, Figure 7 shows the image of sub-mm light sources (LEDs) taken through 1 cm of polystyrene solution ($\mu_s' \sim \mu_s \sim 10\ \mathrm{cm}^{-1}$) by selecting photons which kept the memory of their initial circular polarization (here a few less than 1%). One can see the sources quite clearly.

For many important biomedical applications, such as mammography or imaging brain function, it is necessary to penetrate through 5–10 cm of tissue, after which all the detected signal is diffuse. Detecting earlier arriving photons can provide more information but, if the time window is too narrow, the detected signal becomes too weak to use with acceptable data acquisition times.

This problem is, however, being addressed with some success using statistical models of photon transport, with different degrees of approximation ranging from full Monte Carlo simulation of photon propagation to the diffusion equation. The approach is to address the inverse problem, i.e. to measure the scattered light signal as comprehensively as is practical and to calculate what distribution of absorption and scattering properties would have produced the measured signal. This provides a means to quantify the optical properties of biological tissue, averaged over a volume corresponding to a particular distribution of photon paths, and to form relatively low-resolution tomographic images.

Multiplying the number of sources and detectors and correlating the various intensity distributions over the whole structure under examination gives, as expected, better precision in localizing the size and position of the local structure to be identified.

Despite the multiple input and multiple output approaches, real-life situations with the previously

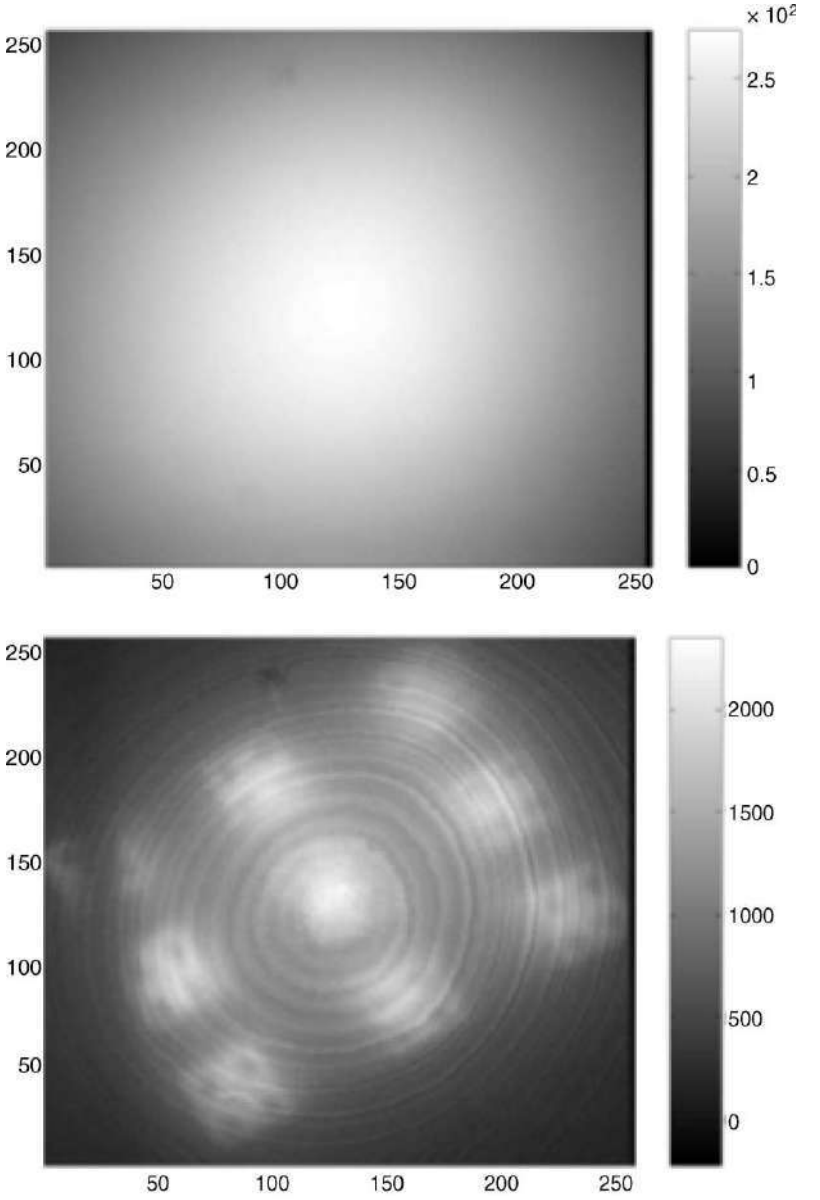

Figure 7 Polarization imaging: here the degree of circular polarization is measured in the upper image. The lower image shows the field without polarization filtering.

mentioned problem of the inhomogeneity of the optical properties of human tissues means that the results are still poor in term of resolution and sub-centimeter structures can hardly be observed in breast imaging, for instance.

One important exception to this limited success in finding precisely the position, the size, and the optical properties of hidden structures is the case of 'activation' studies. This corresponds to a 'differential' situation in which an unknown structure is locally changing with time. Here the goal is not to reach a full description of the unknown structure but only to characterize the local changes induced by the activation.

The main field of application of this approach is certainly in brain activation studies. A matrix of light emitters (usually at two wavelengths which differentiate oxy- and deoxy-hemoglobin, for instance 780 and 840 nm) is coupled to a matrix of light detectors. The signals are usually balanced and any local change which breaks the symmetry of the scattered photon paths induces a differential signal, easy to detect, because there is no background signal before activation takes place.

Moreover it has been demonstrated that for weak perturbations (weak variations of the optical parameters), the inverse problem can be rigorously solved. This geometrical selection provides a cheap and simple setup with rather spectacular results. Sometimes a modulation is added and the phase shift of the photon density wave is enhanced by the differential approach.

Imaging in the Multiple Scattering Regime: Taking Advantage of all the Photons

Light Only: Transillumination and Fluorescence Imaging

The optical systems used for CW measurements are rather simple and require only the alignment of a number of light sources and detectors. Compared to time-gated techniques due to multiple scattering a much stronger path length distribution is present in the signal and the result is a loss of localization and resolution. In the so-called transillumination geometry, collimated detection is used to isolate multiply scattered photons close to the normal emergence angle.

Direct imaging with the required resolution using CW techniques in very thick tissues, such as breast tissue, has not been established even with sophisticated multi-sources/multi-detectors combinations (typically a few tens of sources and detectors at three different wavelengths).

Despite rapid progress in the modeling of light behavior in scattering media it has not yet been demonstrated that CW imaging can provide images with suitable resolution in thick tissue and time-dependent measurement techniques discussed earlier are dominantly used.

For fluorescence imaging, one has to account for the strong attenuation of light as it passes through tissue. Fluorescent dyes (porphyrin, indo-cyanine, and more recently quantum dots) have been developed that are excited and re-emit in the 'biological window' at near-infrared (NIR) wavelengths, where scattering and absorption coefficients are lower, which is favorable for deep fluorescence imaging. Fluorescent intensity in deep tissues depends on the location, size, concentration, and intrinsic characteristics (e.g., lifetime, quantum efficiency) of the fluorophores, and on the scattering and absorption coefficients of the tissue at both the excitation and emission wavelengths. In order to extract intrinsic characteristics of fluorophores within tissue, it is necessary to describe the statistics of photon path lengths which depend on all these different parameters. The amount of light emitted at the site of the fluorophore is directly proportional to the total amount of light impinging on the fluorophore.

Using fluorescence contrast has the unique feature of carrying two spectroscopic contrasts related to both excitation and fluorescent photons.

The next two techniques take advantage of another way to handle the optical information by coupling optics and acoustics.

Coupling Light and Sound: Acousto-Optic and Photo-Acoustic

Although the two approaches that we will now describe are different in their basic principles as well as in the experimental setups, their names are similar – opto-acoustics (sometimes called photo-acoustics) and acousto-optics.

Opto- (or Photo-)Acoustics

In this technique a pulsed (or modulated) electromagnetic beam irradiates the structure under examination. Despite their tortuous paths (about 10 to 100 cm) photons propagate through the volume in a few nanoseconds. Local absorbing centers then absorb the electromagnetic energy, they experience a fast thermal expansion, and become sources of acoustic waves.

The optical problem of tomographic reconstruction is then transformed into a simpler problem of acoustic reconstruction of source distribution because the speed of the sound is much slower than the speed of the light. More precisely we can compute the signal generated by a short light pulse impinging on a scattering medium which contains an absorber (thickness e) as can be seen on Figure 8, when $e\mu_a \ll 1$.

The amount of power per unit surface which is deposited in the absorbing slice is $E\mu_a$ (E is the fluence in W/m^2). When the light pulse duration τ is short enough, despite the propagation in the scattering medium ($\tau \ll e/v$ where v is the speed of the sound),

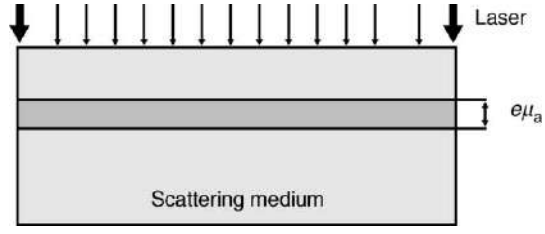

Figure 8 Photo acoustic signal generation: 1-D model.

we find a variation of the mass by volume unit:

$$\Delta\rho \sim \rho\beta\Delta T \sim \rho\frac{\beta E\mu_a e}{C_v e} \qquad [11]$$

The corresponding pressure is:

$$\Delta\rho = \frac{v^2}{\gamma}\Delta\rho \qquad \text{or} \qquad \Delta\rho \sim \rho\frac{v^2\beta E}{C_\rho}\mu_a \qquad [12]$$

Although this approach is new, compared to the purely optical transillumination approach, it has led to a large number of studies in which time gating brings new information to the accurate determination of the US emitter depth. It has been applied to breast tumor examination to complement standard imaging techniques such as echography and X-ray mammography. Recently this technique has demonstrated submillimeter resolution for *in vivo* blood distribution in rat brain imaging.

Acousto-Optic

Here a DC single frequency laser is used to illuminate the sample. For a typical biological sample a few cm thick the trajectories are distributed between the sample thickness value and more than ten times this value. Due to the good temporal coherence of the source all the scattered photons are likely to interfere when they emerge from the sample volume. These interferences are seen as a speckle field which can be observed at a suitable distance from the sample surface.

As seen in Figure 9, adding an ultrasonic field will mainly induce a small (smaller than the optical wavelength) displacement of the scatterers (it also induces local variation of the refractive index) which will cause a speckle modulation at the ultrasonic frequency.

If the zone irradiated by the ultrasonic field is optically absorbing, less photons will emerge from this zone than from a nonabsorbing one and thus less modulation will be seen when scanning across the sample volume. This modulation can be detected by selecting a single speckle grain and averaging over a number of uncorrelated speckle distributions to obtain a suitable average value (for instance with latex particles). This is not always possible, in particular because the overall geometry is rather stable in a semisolid tissue. Using a single detector which receives many speckle grains gives less contrast. A multiple detector and a parallel processing of many speckle grains improve greatly the speed and the signal-to-noise ratio: a typical image of absorbing spheres, embedded in a scattering phantom, is shown in Figure 10.

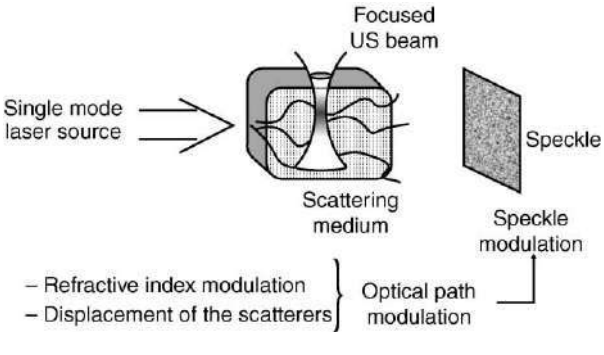

Figure 9 Acousto-optic imaging principle: photons are tagged by the acoustic field.

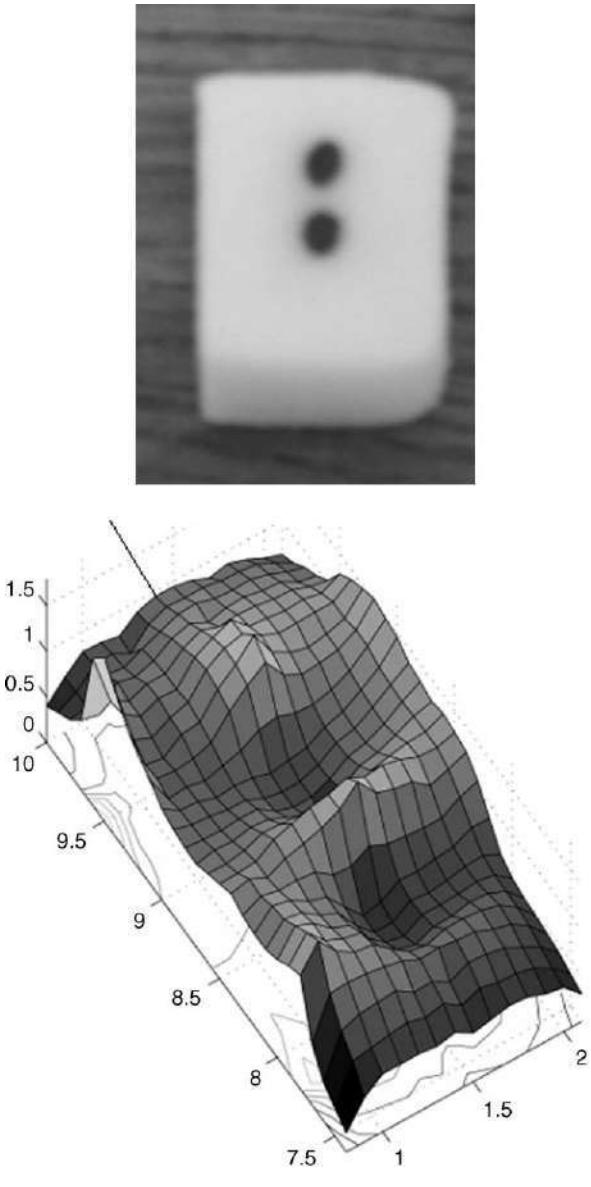

Figure 10 Acousto-optic imaging (lower image) of two absorbing spheres embedded in a scattering medium (upper image shows a section of the gel sample).

Conclusion

The complexity associated with multiple scattering in turbid media, such as tissues of the human body, prevents an easy use of light as a routine tool for in-depth examination, whereas superficial examinations have always been sources of diagnostics.

Optical techniques will obviously progress to better characterization of tissues' optical properties, improvement of laser properties and detection techniques, as well as new mathematical approaches of the inverse problem.

We believe that this field of research is still open to new ideas, and new experimental schemes leading to new breakthroughs in the delicate problem of using light in highly scattering media.

Acknowledgements

I would like to warmly thank all the research staff (Arnaud Dubois, Benoit Forget, François Ramaz, Vincent Loriette) and PhD students (Michael Atlan, Kate Grieve, Gael Moneron, Juliette Selb) at ESPCI whose results have been used to illustrate this article. I also acknowledge discussions with my colleagues Pr Paul French (Imperial College London) and Dr Amir Gandbjbakhche (NIH Bethesda).

See also

Coherence: Coherence and Imaging. **Holography, Techniques:** Digital Holography. **Imaging:** Interferometric Imaging. **Interferometry:** White Light Interferometry. **Microscopy:** Confocal Microscopy; Imaging Multiple Photon Fluorescence Microscopy. **Tomography:** Optical Coherence Tomography; Tomography and Optical Imaging.

Further Reading

Boccara AC and Fink M (eds) (2001) *Optical and Acoustical Imaging of Biological Media*. Compte Rendu de l'Académie des Sciences, Physics and Astrophysics série IV tome 2, no. 8. Paris: Elsevier.

Born M and Wolf E (1997) *Principles of Optics*. New York: Oxford University Press.

Goodman JW (2000) *Statistical Optics*. New York: Wiley-Interscience.

Hecht E and Zajac A (1979) *Optics*. Reading, MA: Addison-Wesley.

Sebbah P (ed.) (2001) *Waves and Imaging through Complex Media*. Dordrecht: Kluwer Academic Publishers.

Van de Hulst HC (1981) *Light Scattering by Small Particles*. New York: Dover.

Van Tiggelen B and Skipetrov S (eds) (2003) *Wave Scattering in Complex Media: From Theory to Applications*. Dordrecht: Kluwer Academic Publishers.

Infrared Imaging

K Krapels and R G Driggers, US Army Night Vision and Electronic Sensors Directorate, Fort Belvoir, VA, USA

Published by Elsevier Ltd.

Introduction

All objects above 0 Kelvin emit electromagnetic radiation associated with the thermal activity on the surface of the object. For terrestrial temperatures (around 300 Kelvin), objects emit a good portion of the electromagnetic flux in the infrared part of the electromagnetic spectrum. The visible band spans wavelengths from 0.4 to 0.7 micrometers (infrared engineers typically use micrometer/μm for wavelength rather than the nanometer or wavenumber more commonly used in other fields). Infrared imaging devices convert energy in the infrared portion of the electromagnetic spectrum into images that can be created by the human eye. The unaided human eye cannot image infrared energy because the lens of the eye is opaque to infrared radiation.

The infrared spectrum begins at the red end of the visible spectrum where the eye can no longer sense energy. It spans from 0.7 μm to 100 μm. The infrared spectrum is, by common convention, broken into five different bands (this may vary according to the application). The bands are typically defined as: near infrared (NIR) from 0.7 to 1.0 μm; short wavelength infrared (SWIR) from 1.0 to 3.0 μm; mid-wavelength infrared (MWIR) from 3.0 to 5.0 μm; long wavelength infrared (LWIR) from 8.0 to 14.0 μm; and far infrared (FIR) from 14.0 to 100 μm. These bands are depicted graphically in Figure 1, which shows the atmospheric transmission for a 1-kilometer horizontal ground path for a 'standard' day in the United States. These types of transmission graphs can be tailored for any conditions, using sophisticated atmospheric models such as MODTRAN (from www.ontar.com). Note that

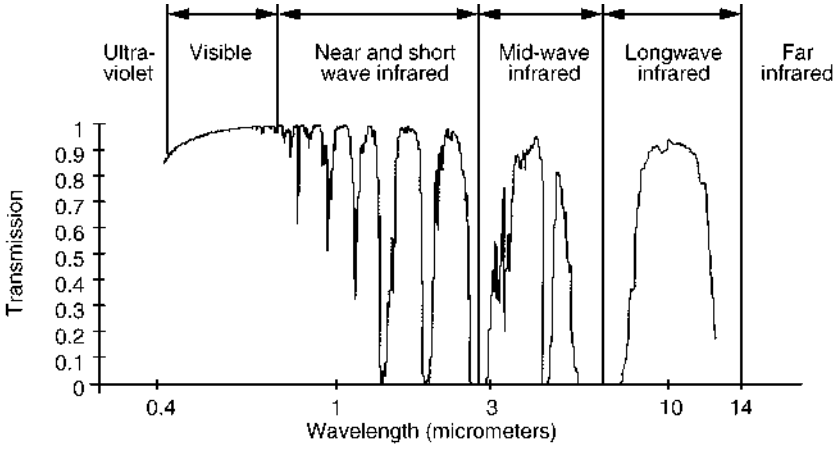

Figure 1 Visible and infrared spectra.

Figure 2 Visible image–reflected energy.

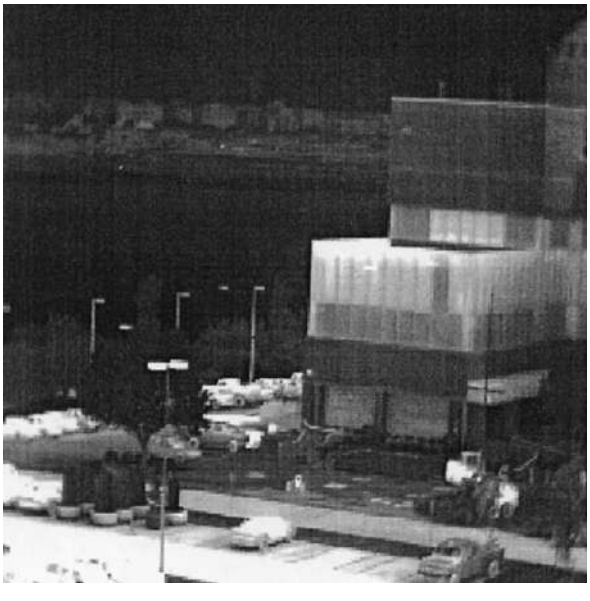

Figure 3 Infrared image–self-emitted energy.

there are many atmospheric 'windows', such that an imager designed with such a band selection can see through the atmosphere.

The primary difference between a visible spectrum camera and an infrared imager is the physical phenomonology of the radiation from the scene being imaged. The energy used by a visible camera is predominantly reflected solar, or some other illuminating energy, in the visible spectrum. The energy imaged by infrared imagers (Forward Looking Infrared, commonly known as FLIRs) in the MWIR and LWIR bands, is primarily self-emitted radiation. From Figure 1, the MWIR band has an atmospheric window in the 3 to 5 μm region and the LWIR band has an atmospheric window in the 8 to 12 μm region. The atmosphere is opaque in the 5 to 8 μm region, so it would be pointless to construct a camera that responds to this waveband.

Figures 2 and 3 are images that show the difference in the source of the radiation sensed by the two types of cameras. The visible image (Figure 2) is light that was provided by the Sun, propagated through Earth's atmosphere, reflected off the objects in the scene, traversed through a second atmospheric path to the sensor, and then imaged with a lens and a visible band detector. A key here is that the objects in the scene are represented by their reflectivity characteristics. The image characteristics can also change by any change in atmospheric path or source characteristic change. The atmospheric path characteristics from the Sun to the objects change frequently, because the Sun angle changes throughout the day and also weather and

cloud conditions change. The visible imager characterization model is therefore a multipath problem that is extremely difficult.

The LWIR image given in Figure 3 is obtained primarily by the emission of radiation by objects in the scene. The amount of electromagnetic flux depends on the temperature and emissivity of the objects. A higher temperature and a higher emissivity correspond to a higher flux. The image shown is 'white-hot' or the whiter a point in the image the higher the flux leaving the object. It is interesting to note that trees have a natural self-cooling process since a high temperature can damage foliage. Objects that have absorbed a large amount of solar energy are hot and are emitting large amounts of infrared radiation. This is sometimes called 'solar loading'.

The characteristics of the infrared radiation emitted by an object are described by Planck's black-body law in terms of spectral radiant emittance

$$M_\lambda = \varepsilon(\lambda)\frac{c_1}{\lambda^5(e^{c_2/\lambda T}-1)}\ \mathrm{W/cm^2\ \mu m} \qquad [1]$$

where c_1 and c_2 are constants of 3.7418×10^4 W μm^4/cm^2 and 1.4388×10^4 μm K. The wavelength, λ, is provided in micrometers and $\varepsilon(\lambda)$ is the emissivity of the surface. A black-body source is defined as an object with an emissivity of 1.0, so it is a perfect emitter. Source emissions of black-bodies at typical terrestrial temperatures are shown in Figure 4. Often, in modeling, the terrestrial background temperature is assumed to be 300 K. The source emittance curves are shown for other temperatures for comparison purposes. One curve corresponds to an object colder than the background and two curves correspond to temperatures hotter than the background.

Planck's equation describes the spectral shape of the source as a function of wavelength. It is readily apparent that the peak shifts to the left (shorter wavelengths) as the body temperature increases. If the temperature of a black-body is increased to that of the Sun (5900 Kelvin), the peak of the spectral shape would decrease to 0.55 μm or green light. This peak wavelength is described by

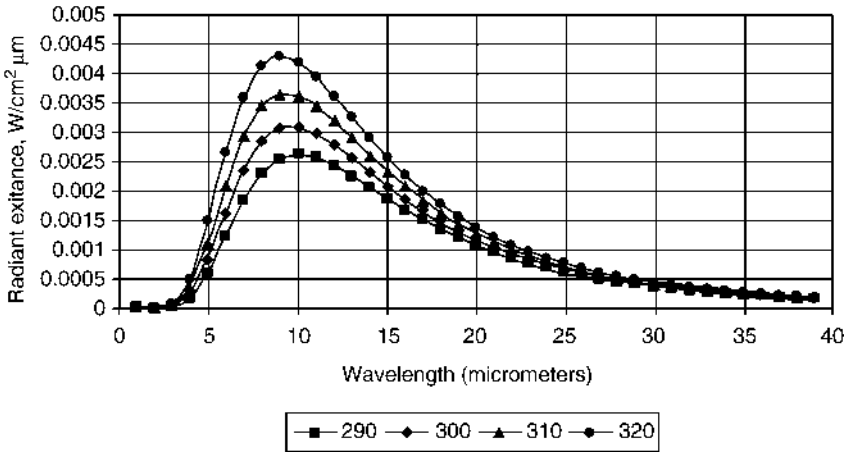

Figure 4 Planck's blackbody radiation curves for four temperatures from 290 K to 320 K.

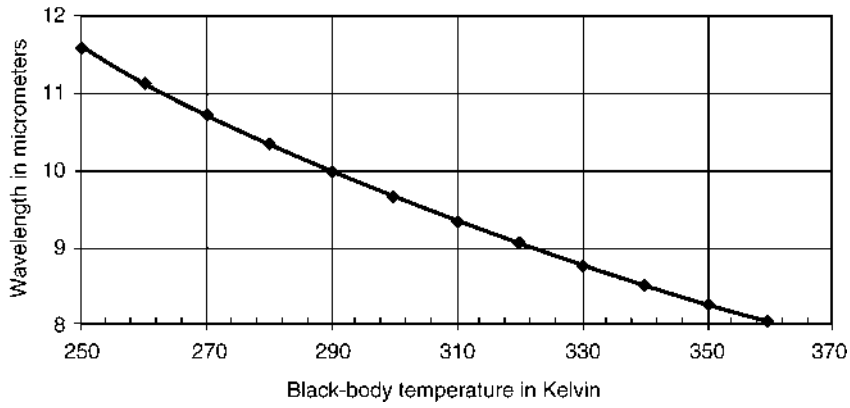

Figure 5 Location of peak of black-body radiation, Wien's Law.

Wien's displacement law

$$\lambda_{max} = 2898/T, \ \mu m \qquad [2]$$

Figure 5 shows the peak wavelength as a function of temperature through the longwave region. It is important to note that the difference between the black-body curves is the 'signal' in the infrared bands. For an infrared sensor, if the background is at 300 K and the target is at 302 K, the signal is the difference in flux between these curves. Signals in the infrared ride on very large amounts of background flux. However, in the visible band this is not the case. For example, consider the case of a white target on a black background. The black background is generating no signal, whereas the white target is generating a maximum signal (given that the sensor gain has been adjusted). Dynamic range may be fully utilized in a visible sensor. In the case of an IR sensor, a portion of the dynamic range is used by the large background flux radiated by everything in the scene. This flux never has a small value, hence sensitivity and dynamic range requirements are much more difficult to satisfy in IR sensors than in visible sensors. Table 1 summarizes this characteristic numerically for the 300 K background and targets shown in Figure 4.

A typical infrared imaging system scenario is depicted in Figure 6. The scene consists of two major components, the target and the background. In an IR scene, the majority of the energy is remitted from the constituents of the scene. This emitted energy is transmitted through the atmosphere to the sensor. As it propagates through the atmosphere it is degraded by absorption and scattering. Obscuration by intervening objects and additional energy emitted by the path also affect the target energy. All of these contributors, which are not the target of interest, essentially reduce one's ability to discriminate the target. So, at the entrance aperture for the sensor, the target signature is reduced from the values observed at the target. Then, the signal is degraded by the optics and scanner (as applicable) of the sensor. The energy is then sampled by the detector array and converted to electrical signals. Various electronics amplify and condition this signal before it is presented to either a display for human interpretation or an algorithm like an automatic target recognizer for machine interpretation. A linear systems approach to modeling allows the components' transfer functions to be treated separately as contributors to the overall system performance. This approach allows for straightforward modifications to a performance model for changes in the sensor or environment when performing trade-off analyses.

Table 1 Signal/dynamic range limitation in IR bands

Blackbody temperature	*Exitance in 8–12 μm band (w/cm^{-2})*	*Contrast*
290 (−10Δ*T* K target)	1.25E − 02	−8.48%
300 (background)	1.48E − 02	
310 (+10Δ*T* target)	1.74E − 02	7.97%
320 (+20Δ*T* target)	2.02E − 02	15.39%

History of Infrared Imaging

Night vision systems began to be developed extensively during World War II. Satisfying the requirement to image at night was approached with two different methodologies. The first method was image intensification (I^2). This method involves amplification of any small light that was available and displaying it directly to the eye. Typically I^2 devices are sensitive to visible light and a small portion of the short-wave IR (SWIR, 0.7–3.0 μm) band. They are often classified (along with visible band imagers) as electro-optical (EO) imagers. With an I^2 imager, there must be some source of illumination for them to function well (as little as starlight is sufficient

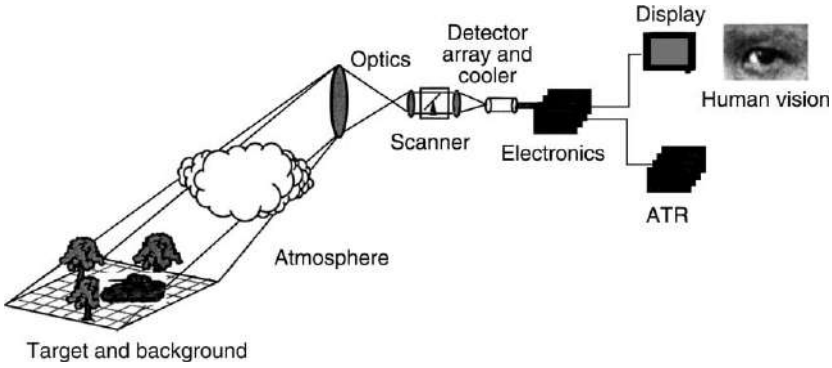

Figure 6 Typical infrared imaging scenario.

Table 2 Applications of infrared imaging

Category	*User*	*Application*
Military	Intelligence analyst	Intelligence surveillance and reconnaissance
	Pilot, driver	Navigation/pilotage
	Gunner, weapons officer	Targeting/fire control
	Air defense	Search and track
Commercial	*Civil government*	
	Police	Surveillance, fugitive tracking
	Fire	Rescue, hot-spot location
	Environmental	
	EPA	Emission tracking
	Interior department	Resource management
	NIMA\USGS	Mapping
	NOAA\MWS	Weather forecasting
	Industrial	
	Manufacturers	Machine vision
	Maintenance	Non-destructive testing
	Medical	
	Doctors	Diagnostic imaging

for approx 20/40 acuity with current systems). The second type of night vision device is the infrared or thermal imager (also known as FLIRs). FLIRs are sensitive to the radiation in either the mid-wave IR (MWIR, 3–5 μm) or long-wave IR (LWIR, 8–14 μm) bands.

The first infrared cameras used photographic film that was sensitive to infrared wavelengths. Intelligence and reconnaissance imaging from aircraft drove the infrared system development as a night augmentation to existing visible cameras. These imagers used either continuous moving film or stationary film, like an ordinary camera. A slit aperture was used in the continuous moving film cameras and aircraft motion provided the scan of the scene along the film to build a continuous image. The introduction of television, as electronic imaging in the visible spectrum, led to the development of electronic imagers in the infrared band. Electronic infrared imagers were not restricted to looking down and using aircraft motion for scan, hence these devices came to be known as FLIRs (Forward Looking InfraRed). Today, an electronic infrared imager may be called a FLIR or a thermal imager, with the trend being to use FLIR to describe military applications.

Early FLIRs were often accompanied by illumination devices. They were not very sensitive compared to modern FLIRs so active illumination was necessary. Infrared searchlights or illuminators were used to get higher signal-to-noise ratios. These types of FLlRs are called 'active' IR imagers, while FLIRs that do not use illuminators are 'passive' imagers. Generally, active imagers are used in civil applications and are declining in military use. Broadly, applications for thermal imagers fall into two categories; military and commercial. Table 2 lists some of the purposes and users for modern thermal imagers. The design and performance criteria vary widely for some of the applications.

Types of Infrared Imagers

Infrared imagers are classified by different characteristics: scan type, detector material/cooling requirements, and detector physics. The scan type refers to the method used to construct the electronic image. The camera may use a single detector which is raster scanned over the input scene to build an image. Alternatively, a parallel scan uses a linear array of detectors scanned across a scene to build an image. The latest advances in materials have led to staring arrays of detectors. In a staring system, a detector is present in two dimensions to represent each image pixel and no mechanical motion of the focal plane is necessary to construct the image. There are some hybrid FLIR types that combine the different imaging techniques. Usually this combination leads to an improvement in signal-to-noise ratio or to reduce undersampling.

The second classification, by detector material/cooling requirements, usually describes FLIRs built using differing materials. For example, a typical LWIR FLIR material is mercury cadmium telluride (HgCdTe or MCT). In order to achieve high sensitivity, these devices are cooled to decrease dark current. Usually the cooling is based on liquid nitrogen or a cryogenic cooler, and the detectors operate at 77 K. Another common detector material is indium antimonide (InSb), which is used for MWIR FLIRs and is also cooled. A newer class of infrared cameras is not cooled, being referred to as 'uncooled' FLIRs. These uncooled FLIRs are microbolometers (resistive elements) or pyrometers (capacitive elements) and have less sensitivity than cooled imagers. Typically, the cooled imagers are comprised of photon detectors while the uncooled imagers are based on thermal detectors, the uncooled FLIRs being used in lower-performance applications.

Infrared Imager Performance

There are three general categories of infrared sensor performance characterizations. The first is sensitivity and the second is resolution. When end-to-end, or human-in-the-loop (HITL), performance is required, the third type of performance characterization

describes the visual acuity of an observer through a sensor. The former two are both related to the hardware and software that comprises the system, while the latter includes both the sensor and the observer. Sensitivity is determined through radiometric analysis of the scene/environment and the quantum electronic properties of the detectors. Resolution is determined by analysis of the physical optical properties, the detector array geometry, and other degrading components of the system in much the same manner as complex electronic circuit/signals analysis.

Sensitivity describes how the sensor performs with respect to input signal level. It relates noise characteristics, responsivity of the detector, light gathering of the optics, and the dynamic range/quantization of the sensor. Radiometry describes how much light leaves the object and background and is collected by the detector. Optical design and detector characteristics are of considerable importance in sensor sensitivity analysis. In infrared systems, noise equivalent temperature difference (NETD) is often a first-order description of the system sensitivity. The 3D noise model describes more detailed representations of sensitivity parameters (see Further Reading).

The second type of measure is resolution. Resolution is the ability of the sensor to image small targets and to resolve fine detail in large targets. Modulation transfer function (MTF) is the most widely used resolution descriptor in infrared systems. Alternatively, it may be specified by a number of descriptive metrics such as the optical Rayleigh criterion or the instantaneous field-of-view of the detector. Where these metrics are component-level descriptions, the system MTF is an all-encompassing function that describes the system resolution.

Sensitivity and resolution can be competing system characteristics and they are the most important issues in initial studies for a design. For example, given a fixed sensor aperture diameter, an increase in focal length can provide an increase in resolution, but may decrease sensitivity. Typically, visible band systems have plenty of sensitivity and are resolution-limited, whereas infrared imagers have been more sensitivity-limited. With staring infrared sensors, the sensitivity has seen significant improvements.

Often metrics, such as NETD and MTF, are considered to be separable. However, in an actual sensor, sensitivity and resolution performance are not independent. As a result, minimum resolvable temperature difference (MRT or MRTD) has become the primary performance metric for infrared systems. MRT is a quantitative performance measure in terms of both sensitivity and resolution and a simple MRT curve is shown in Figure 7. The performance is

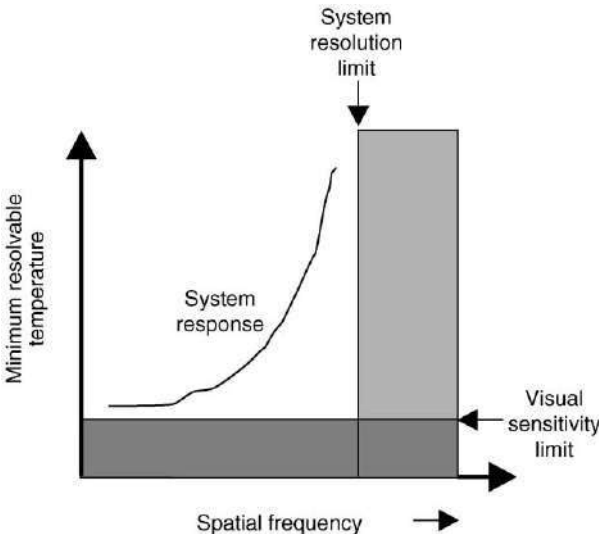

Figure 7 IR imager performance.

bounded by the sensor's limits and the observer's limits. The temperature difference (or thermal contrast) required to image smaller details in a scene increases with detail size. The inclusion of observer performance yields a single sensor performance characterization, MRT. It describes sensitivity as a function of resolution, and includes the human visual system.

Many different sensor characteristics may be considered to increase the fidelity of a sensor model. A list of typical infrared imaging system parameters is given in Figure 8. When considering performance in the system sense, the groupings are often blocks in the system diagram, each with separate transfer functions. This approach works well unless the linear shift invariance assumption is not valid.

General Characteristics of Infrared Imagers

There are a number of imaging characteristics that make infrared systems a little different than conventional visible imaging systems. These include source flux levels, detector charge storage and sensitivity, detector size, diffraction blur, sampling, and uniformity characteristics.

First, we consider the source flux levels. The daytime and night-time flux levels (in photons per second per square centimeter) on Earth in the visible (0.3 to 0.7 μm) is 1.5×10^{17} and around 1×10^{10}, respectively. In the mid-wave (3 to 5 μm), the daytime and night-time flux levels are 4×10^{15} and 2×10^{15} μm where the flux is a combination of

Illumination
- Spectral irradiance

Target
- Reflectivity
- Emissivity
- Size and spatial characteristics
- Temperature

Atmosphere
- Weather conditions
- Obscurants
- Transmission
- Optical transfer function
- Scattering

Sensor optics
- Lens transmissions
- Mirror reflectivity
- Filter transmission
- Aberrations
- Diffraction
- Aperture size and shape

Sensor detector
- Responsivity
- Detectivity
- Noise characteristics
- Detector angular substense

Sensor electronics
- Temporal characteristics
- Digital and/or analog
- Filter characteristics

Display
- Resolution
- Brightness and contrast

Human psychophysics
- Temporal response
- Image transfer function
- Brightness dependence

ATR or AiTR response
- Discrimination probability
- False alarm rate

Figure 8 Factors in infrared sensor modeling.

emitted and solar reflected flux. In the longwave, the flux is primarily emitted where both day and night yield a 2×10^{17} μm frequency, the AMOP drops rapidly to zero. The effect is to extend the prediction of the MRT beyond the half sample rate and introduce a parameter called minimum temperature difference perceived (MTDP) to replace MRT in the range performance estimate and also in the lab to characterize an undersampled system.

The concept of MTDP follows from observation of standard 4-bar images as the pattern frequency passes the imager half sample rate. The image changes from four bars to three, two, and finally one (can be perceived). The phase of the target image on the detector array may need to be adjusted to observe this progression. The standard MRT measurement requires that all four bars be resolvable by the observer during the measurement process. For use in the lab, the MTDP is defined as the minimum temperature difference at which two, three, or four bars can be resolved in the image of the standard 4-bar test pattern by an observer, with the test pattern positioned at optimum phase. The optimum phase is the phase at which two, three, or four bars are 'best perceived'. TRM3 uses the standard definition of MRT as for adequately sampled imagers. The TRM3 Approach Model is implemented if the imager is considered undersampled as defined when the prefilter MTF at the half-sample frequency is larger than 10%. The MTDP equation is given by

$$\mathrm{MTDP}(f) = \frac{\frac{\pi}{2}\mathrm{SNR}_{\mathrm{th}}\Psi(f)}{\mathrm{AMOP}(f)} \qquad [3]$$

where $\mathrm{SNR}_{\mathrm{th}}$ is a threshold signal-to-noise ratio and Ψ is the total system filtered noise. MTDP is used in the same manner as MRTD in range calculations.

Another approach is the triangle orientation discrimination (TOD) threshold. In the TOD, the test pattern is a (nonperiodic) equilateral triangle in four possible orientations (apex up, down, left, or right), and the measurement procedure is a robust 4AFC (four alternative forced-choice) psychophysical procedure. In this procedure, the observer has to indicate which triangle orientation he sees, even if he is not sure. Variation of triangle contrast and/or size leads to a variation in the percentage correct between 25% (complete guess) and 100%, and by interpolation the exact 75% correct threshold can be obtained. A complete TOD curve (comparable to an MRTD curve) is obtained by plotting the contrast thresholds as a function of the reciprocal of the triangle angular size (Figure 9).

The TOD method has a large number of theoretical and practical advantages: it is suitable for undersampled and well-sampled electro-optical and optical

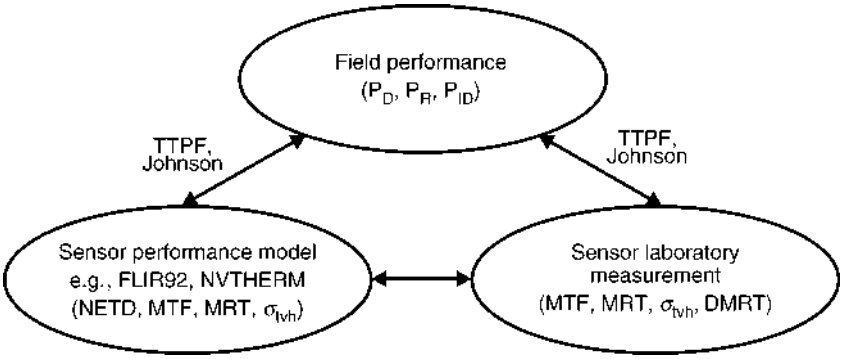

Figure 9 Modeling, testing and performance triangle.

imaging systems in both the thermal and visual domains, it has a close relationship to real target acquisition, and the observer task is easy. The results are free from observer bias and allow statistical significance tests. The method may be implemented in current MRTD test equipment with little effort, and the TOD curve can be used easily in a target acquisition (TA) model such as ACQUIRE. Two validation studies with real targets show that the TOD curve predicts TA performance for under-sampled and well-sampled imaging systems very well. Currently, a theoretical sensor model to predict the TOD (comparable to NVTherm or TRM3) is under development. The lab measurement and field performance appear sound, but the model is not yet available.

Testing Infrared Imagers: NETD, MTF, and MRTD (Sensitivity, Resolution and Acuity)

In general, imager sensitivity is a measure of the smallest signal that is detectable by a sensor. Sensitivity is determined using the principles of radiometry and the characteristics of the detector. For infrared imaging systems, noise equivalent temperature difference (NETD) is a measure of sensitivity. The system intensity transfer function (SITF) can be used to estimate the NETD, which is the system noise rms voltage over the noise differential output. NETD is the smallest measurable signal produced by a large target (extended source).

Equation [4] describes NETD as a function of noise voltage and the system intensity transfer function. The measured values are determined from a line of video stripped from the image of a test target, as depicted in Figure 10. A square test target is placed before a black-body source. The delta T is the difference between the black-body temperature and the mask. This target is then placed at the focal point of an off-axis parabolic mirror. The mirror serves the purpose of a long optical path length to the target, yet relieves the tester from concerns over atmospheric losses to the temperature difference. The image of the target is shown in Figure 10b. The SITF slope for the scan line in Figure 10c is the $\Delta S/\Delta T$ where ΔS is the signal measured for a given ΔT. The N_{rms} is the background signal on the same line.

$$\mathrm{NETD} = \frac{N_{rms}[\mathrm{volts}]}{\mathrm{SITF_Slope}[\mathrm{volts/K}]} \qquad [4]$$

Resolution is a general term that describes the size of resolvable features in an imager's output. With infrared imaging systems, the resolution is described by the system modulation transfer function (MTF). Consider Figure 11 for the determination of MTF, from either point spread function or edge spread function (psf or esf). In Figure 11a, a test target is placed at the focal point of a collimator. The target may be a point, edge, or line as shown in Figure 11b. The IR system images this target. The output of a detector scan, or a line of detectors in the staring array case, is taken for analysis, in Figure 11c. For the edge spread function, the spatial derivative is taken to get the psf. The Fourier transform is then calculated to give the system MTF as shown in Figure 11d. The point spread function is really the impulse response of the system, so a smaller psf is desirable, where a wide MTF is desirable. Such a psf/MTF combination gives a higher resolution.

The MRTD of an infrared system is defined as the differential temperature of a four bar target that makes the target just resolvable by a particular observer. It is a measure of observer visual acuity when a typical observer is using the infrared imager. It results in a descriptor, which is a function not just of a single value. It is a plot of sensitivity as a function of spatial frequency.

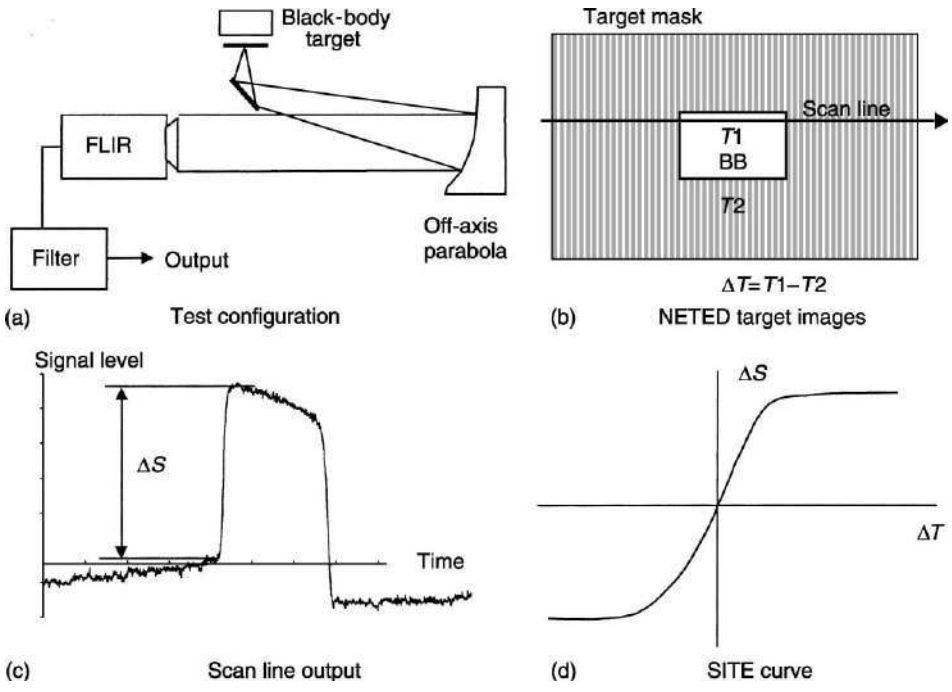

Figure 10 NETD measurement.

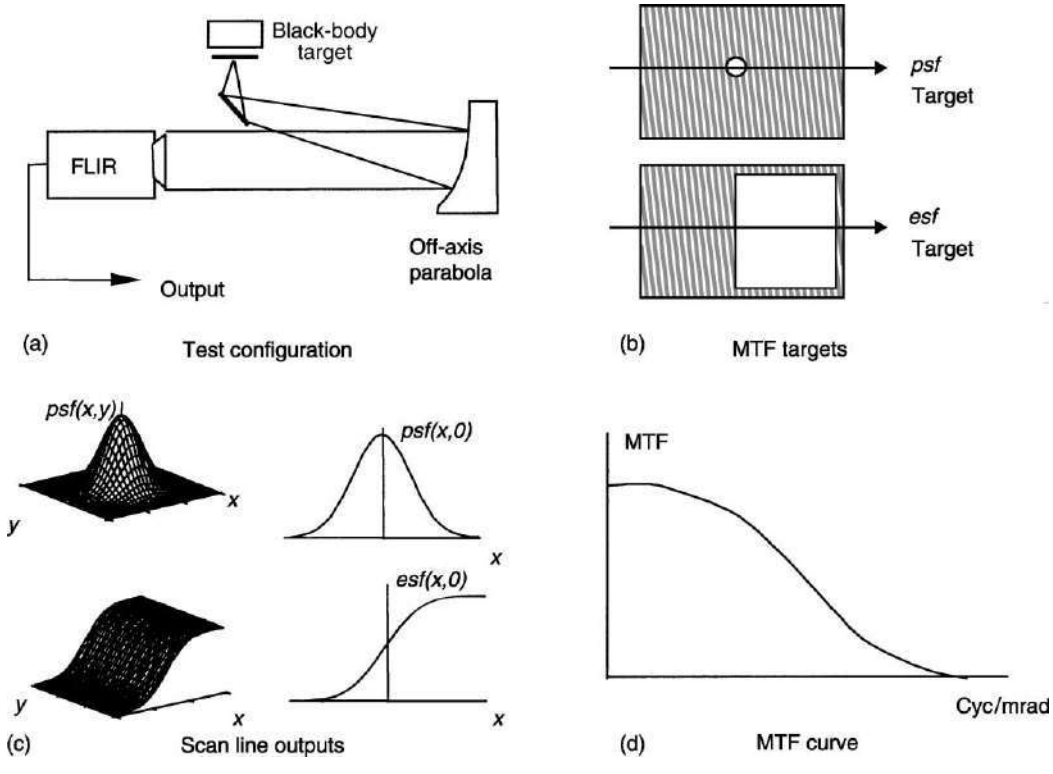

Figure 11 MTF measurement.

This parameter can be extended to field performance using Johnson's criteria.

As a laboratory measurement, it uses a 7:1 aspect ratio bar target. The procedure is shown in Figure 12. A bar target mask, with a black-body illuminator, is placed at the focal plane of a collimator (Figure 12a and 12b). The temperature is increased from a small value until the bars are just resolvable to a trained

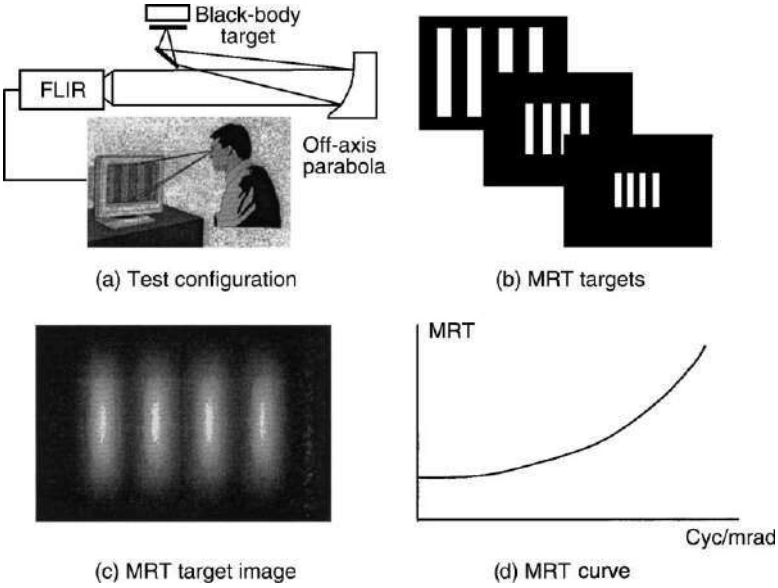

Figure 12 MRTD measurement.

observer (Figure 12c). Then the temperature difference is decreased until the bars are no longer visible. These are averaged for a particular target. The same procedure is performed with a negative contrast bar target. The average of the positive and negative values is the MRT for a particular spatial frequency. These data points are plotted and a curve may be fitted to interpolate/extrapolate performance at other than the discrete spatial frequencies of the targets, as in Figure 12d.

Summary

We have provided a general description of infrared imaging systems in terms of characteristics, modeling, field performance, and performance measurement. The characteristics of infrared imagers continue to change. Significant changes have occurred in the past five years, to include the development of higher performance uncooled imagers and ultra-narrow field of view photon systems. Large format detector arrays are commercially available in the mid-wave and are becoming more available in the longwave. Still in the research phase are dual-band focal plane level. At first sight, it appears that the longwave flux characteristics are as good as a daytime visible system; however, there are two other factors limiting performance. First, the energy bandgaps of infrared photons are much smaller than those photons in the visible, so the detectors suffer from higher dark current. The detectors are usually cooled to reduce this effect. Second, the reflected light in the visible is modulated with target and background reflectivities that typically range from 7 to 20%. In the infrared, all terrestrial objects are emitting with an emission temperature of around 300 Kelvin. Typically, a two-degree equivalent black-body difference in photon flux is considered a high contrast target to background flux. The flux difference between two black-bodies of 302 K compared to 300 K can be calculated in a manner similar to that shown in Figure 4. This is the difference in signal that provides an image, so note the difference in signal compared to the ambient background flux. In the longwave, the signal is three percent of the mean flux and in the mid-wave, it is six percent of the mean flux. This means that there is a large flux pedestal associated imaging in the infrared. Unfortunately, there are two problems accompanying this large pedestal. First, photon noise is the dominant noise and is determined by the mean of the pedestal and is compared to the small signal differences. Second, the charge well storage in infrared detectors is limited to around 10^7 charge carriers. A longwave system in a hot desert background would generate 10^{10} charge carriers in a 33 millisecond integration time. Smaller F-numbers, spectral bandwidths, and integration times are used so that the charge well does not saturate. This results in SNR of 10 to 30 times below the ideal. It has been suggested that some on-chip compression may be a solution to the well pedestal problem. In many scanning FLIR systems, the pedestal is eliminated by AC-coupling the detector signals. Infrared focal plane array (IRFPA) readout circuits have been previously designed and fabricated to perform charge skimming and charge partitioning.

Another major difference in infrared systems, from that of visible systems, is the size of the detector and diffraction blur. For infrared photon detectors, typical sizes range from 25 to 50 μm, where visible systems can be fabricated with 3 μm detectors. Also, the diffraction blur for the longwave is more than ten times larger than the visible blur and mid-wave blur is eight times larger than visible blur. Therefore, the image blur, due to diffraction and detector size, is much larger in an infrared system than a visible system. Also, it is common for infrared staring arrays to be sampling limited where the sample spacing is larger than the diffraction blur and the detector size. Dither and microscanning are frequently used to enhance performance.

Finally, both infrared photon and thermal detectors in staring arrays have responsivities that vary dramatically and it is current practice to correct for the resulting nonuniformity (nonuniformity correction or NUC). The nonuniformity can cause fixed pattern noise in the image that can limit the performance of the system even more than temporal noise (especially in static imaging or stabilized target acquisition imaging).

Modeling Infrared Imagers

Otto Schade Sr. developed the earliest imaging system models and performance measures. His work describing television in the 1950s and 1960s, pioneered the way for the characterizations of imaging sensors used by engineers today. His performance measure for still pictures, moving pictures and television systems was based on an observer resolving a three bar periodic target.

John Johnson developed the technique of relating the acquisition of military targets in the field to laboratory measurements of resolvable bar targets. He divided the discrimination tasks into four categories: detection, orientation, recognition, and identification. This technique was termed an 'equivalent bar pattern approach'. It related performance on a simple test target to performance with complex object targets. Johnson viewed scale models and bar targets in the laboratory against a bland background. The smallest discernible barchart target yielded the maximum resolvable bar pattern frequency. These results, tabulated in Table 3 as cycles across the minimum dimension, were the basis for the discrimination methodology in widespread use today. Note that detection took only one cycle, but as the tasks got more specific or difficult, the requirement went as high as 8 cycles across the target minimum dimension.

Recognizing that typical vehicle aspect ratios were somewhat limited, Johnson and Lawson conducted further experiments. Paul Moser conducted some of his own, reanalyzed Johnson and Lawson's, and developed the concept that this task was related to the average or critical dimension instead of the minimum dimension. This led to the conversion of line pairs to resolution elements, introducing the second dimension. Additionally, he performed a similar experiment on marine targets, which is summarized in Table 4. Lloyd and Sendall developed the MRT metric, which combines both resolution and sensitivity characteristics. This measurement of MRT is described in detail in a following article, but there is a theoretical model that is used to evaluate new sensor designs and concepts. It relates the minimum temperature difference between bar pairs at which they are just resolvable as a function of spatial frequency. MRT describes the infrared imager sensitivity as a function of resolution. More specifically, MRT is a measure of thermal contrast sensitivity as a function of spatial frequency.

Ratches *et al.* developed the NVL (US Army Night Vision and Electronic Sensors Directorate formerly known as the Night Vision Laboratory) Static Performance Model, which predicted the end-to-end performance. It started with the target signature and carried through to the observer. The MRT, a laboratory measurement or a modeled value, was related to the probabilities of discrimination through the target transfer probability function, cumulative percentages related to Johnson's criteria.

Table 3 Johnson's criteria

Target	*Resolution per minimum dimension*			
Broadside view	*Detection*	*Orientation*	*Recognition*	*Identification*
Truck	0.9	1.25	4.5	8
M-48 Tank	0.75	1.2	3.5	7
Stalin Tank	0.75	1.2	3.5	6
Centurion Tank	0.75	1.2	3.5	6
Half-track	1	1.5	4	5
Jeep	1.2	1.5	4.5	5.5
Command car	1.2	1.5	4.3	5.5
Soldier (standing)	1.5	4.8	3.8	8
105 howitzer	1	1.5	4.8	6
Average	1 ± 0.25	1.4 ± 0.35	4.0 ± 0.8	6.4 ± 1.5

Table 4 Moser's data

Discrimination task	Line pairs/critical dimension
Detect ship	Aperiodic treatment
Classify as combatant	4
Recognize type	10

D'Agostino and Webb created the 3D noise model, both spatial and temporal considerations, for analyzing system noise. This technique divided the noise components into manageable and understandable components. It also simplified the integration of these effects into models.

The introduction of focal plane array (FPA) imagers has created a requirement to improve the model performance to account for the difference between scanning arrays and staring arrays. This change has made sampling an important area for improvement. A new model called NVTHERM is a result of the NVESD model improvement effort, which addresses undersampled imaging system performance.

Sensor characterization is seen in three ways: theoretical models, field performance, and laboratory measurements (Figure 9). Theoretical models are developed that describe sensor sensitivity, resolution, and human performance for the purpose of evaluating new conceptual sensors. Acquisition models, and other models, are developed to relate the theoretical models to field performance. This link allows theoretical models to be converted to field performance quantities (e.g., probabilities of detection, recognition, and identification). Field performance is measured in the field so that the theoretical models can be refined and become more accurate with advanced sensor developments. Since field performance activities are so expensive, methods for the direct measurement of sensor performance are developed for the laboratory. Field performance testing of every infrared sensor built, including buy-off, acceptance, and life-cycle testing, is ridiculous and out of the question. Laboratory measurements of sensor performance are developed such that, given these measurements, the field performance of a system can be predicted. Sensor characterization programs require accurate sensor models, field performance estimates with acquisition models, and repeatable laboratory measurements.

There are a few alternatives to the US NVTherm model. One candidate approach to undersampled imager modeling is Germany's TRM3, or the MDTP approach. The impact of undersampling in the image of a 4-bar target can readily be seen by simply observing the change in the image as a function of spatial frequency and phase. The spatial frequency is defined as line pairs per angular extent of the target, and the phase specifies the relative location of the target image to the detector array raster. These effects are obviously not independent of each other, but for each target there can be found an optimal phase where the observer can see the maximum amplitude modulation in the image of the target. MRT measurements in the past have utilized this variation with phase by allowing the observer to optimize the displayed image through target or system line-of-sight changes during the measurement process. For undersampled imagers, TRM3 addresses the problem of the MRT calculation's inability to predict beyond the half sample rate of the sensor, by replacing the MTF in the denominator of the MRT equation with an appropriately scaled term called the average modulation at optimum phase (AMOP). AMOP is the average signal difference in the image of the 4-bar standard pattern, with the test pattern positioned at optimum phase. AMOP oscillates between the presample MTF and the bar modulation. Beyond a frequency of 0.8 times the sample rate or 1.6 times the half sample arrays and quantum well detector systems. These systems will find their place in applications to include both military and commercial sectors.

Further Reading

Bijl P and Valeton MJ (1998) Triangle orientation discrimination: the alternative to minimum resolvable temperature difference and minimum resolvable contrast. *Optical Engineering* 37(7): 1976–1983.

D'Agostino J and Webb C (1991) 3-D analysis framework and measurement methodology for imaging system noise. *Proceedings of SPIE* 1488: 110–121.

Driggers RG, Cox P and Edwards T (1998) *Introduction to Infrared and Electro-Optical Systems*. Boston, MA: Artech House, p. 8.

Holst GC (1995) *Electro-Optical Imaging System Performance*. Winter Park, FL: JCD Publishing, p. 347.

Holst GC (1995) *Electro-Optical Imaging System Performance*. Winter Park, FL: JCD Publishing, p. 432.

Johnson J (1958) Analysis of image forming systems. *Proceedings of IIS* 249–273.

Johnson J and Lawson W (1974) Performance modeling methods and problems. *Proceedings of IRIS*.

Ratches JA (1973) *NVL Static Performance Model for Thermal Viewing Systems*. USA Electronics Command Report ECOM 7043, AD-A011212.

Schade OH (1948) Electro-optical characteristics of television systems. *RCA Review* IX(1–4).

Sendall R and Lloyd JM (1970) Improved specifications for infrared imaging systems. *Proceedings of IRIS* 14(2): 109–129.

Vollmerhausen RH and Driggers RG (1999) NVTHERM: next generation night vision thermal model. *Proceedings of IRIS Passive Sensors* 1.

Wittenstein W (1998) Thermal range model TRM3. *Proceedings of SPIE* 3436.

Interferometric Imaging

D L Marks, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Interferometric imaging is image formation by measuring the interference between electromagnetic signals from incoherently emitting or illuminated sources. To learn about interferometric image formation using coherent light, *see* **Holography, Techniques:** Overview. To understand what optical coherence is, *see* **Coherence:** Overview. To learn what kinds of sources produce incoherent light, *see* **Incoherent Sources:** Lamps. To learn about interferometric instruments, *see* **Interferometry:** Overview. To learn about interferometric instruments utilizing incoherent light, *see* **Interferometry:** White Light Interferometry. To better understand the effect of the coherence of light sources on image formation, *see* **Coherence:** Coherence and Imaging.

Interferometric imaging attempts to improve the resolution of images by interferometrically combining the signals from several apertures. By doing so, a resolution can be achieved that is superior to that achievable through standard image formation by any of the individual apertures. This is especially useful in astronomy, where resolving more distant and smaller stellar objects is always desirable. It is also finding increasing use in microscopy. It has several advantages over conventional noninterferometric imaging methods. The method can combine the light gathered from several imaging instruments to form an image superior to that which can be formed by any one of the instruments. It can also obviate the need to produce a single very large aperture to achieve the equivalent resolution, which in many cases is of impractical size. In addition, because it can produce a phase-resolved measurement, the data can be processed more flexibly to form a computed image estimate. There are also considerable disadvantages to interferometric imaging compared to conventional imaging methods. The amount of signal gathered is far less than could be gathered with an equivalent single aperture. It is difficult to mechanically or feedback stabilize the time delay between two widely separated apertures to obtain an accurate phase measurement. Often only a small bandwidth of the source light can be utilized, further reducing the available signal. The interference component of the combined signal can be very small. Typically an interferometric image must be computed from the data, rather than being directly measured on photographic film or an electronic focal plane array. Even with these disadvantages, interferometric imaging is frequently more feasible than building a single large aperture to achieve the highest available resolution images.

In contrast to interferometric imaging, a standard image forming instrument such as a telescope has a resolution limited by the telescope mirror aperture in the absence of aberrations and atmospheric turbulence. As telescope mirrors become very large, they become bulky, extremely heavy, and difficult to manipulate, as well as deform under their own weight and temperature gradients. However, adaptive optics is being more frequently utilized to dynamically correct for instrument aberrations as well as atmospheric turbulence. The images captured by telescopes are typically intensity-only images, which are amenable to image processing and deconvolution, but nevertheless lack phase information. Interferometric imaging is an alternative, where the resolution of a large aperture can be attained by measuring the interference between individual points of the electromagnetic field within an area equivalent to the large aperture extent. The interferometric combining of light from sub-apertures to achieve the resolution of a larger aperture is called aperture synthesis.

Unlike image formation by a lens, interferometric images are never physically formed. Conventional images are formed when diverging spherical waves emanating from sources are focused by a lens to converging spherical waves on a photographic film or an electronic sensor. In this case, the image information is directly contained in the exposure at each position on the sensor. Interferometric measurements contain the image information as statistical correlations between the electromagnetic fields at pairs of spatial points. The statistical properties of electromagnetic waves are modeled by optical coherence theory. Interferometric imaging works by measuring the statistical correlations between various pairs of points in the electromagnetic field, and then uses optical coherence theory to infer what the image of the source of the radiation is.

The light that emanates from most natural and many artificial sources is incoherent. Incoherent radiation is optical random noise. It is analogous to the sound of static heard on a radio receiving no signal. Because it is random, the fluctuations of the electromagnetic field produced by an incoherent source are completely unrelated to the fluctuations of any other source. The origin of the randomness is

usually from microscopic processes such as the thermal motions of electrons or the spontaneous emission of an excited atom. Because these microscopic processes tend to act independently of one another, the light produced by different sources tends to be mutually incoherent. Because incoherent sources do not radiate a deterministic field, the average amount of power they radiate is usually specified rather than the field itself, which is random.

However, as the fields propagate away from their sources, the fields can become partially coherent. To see this, consider the light waves propagating away from a single point-like incoherent source. The field produced at the source point consists of random fluctuations in time, but these fluctuations travel outward in a spherical wave at the speed of light. The fluctuations of the field of any two points an equal distance from the source will be the same, or perfectly correlated, because they arrive from the source at the same time. This is illustrated in Figure 1a, where a random wave is emanated from a point incoherent source. Because of the complete correlation, we can say that these two points are fully spatially coherent.

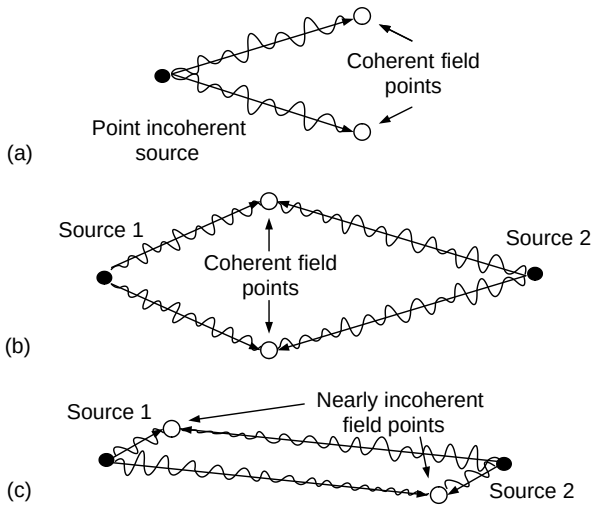

Figure 1 The distance between incoherent sources and two field points determines the coherence between the field points. (a) Because there is only one source, the field everywhere is determined by the radiation of that source. Therefore, the same fluctuations of the wave arrive at two points an equal distance from the source at the same time, and produce identical fields at both points, and so are completely coherent. (b) Two field points are now the same distance from two incoherent sources. Because both sources contribute identical fields to both points (because the fluctuations of the wave from both sources arrive at both points at the same time), the superposition of the fields from both sources at both field points are identical, and so the fields are coherent. (c) Two field points are now near two sources. Because each field point receives radiation mostly from its nearby source and very little from the other source, the fields at each field point are almost completely mutually incoherent.

So while the source itself produces random noise, the fluctuations of the field at points distant from the source become correlated because they originate from the same source.

Partially coherent waves occur when the measured field contains a superposition of the fields from more than one incoherent source. If there are now two point sources instead of one, each field point in space will receive radiation from both sources. The amount of the contribution from each source to each point in space depends on the distance between the field point and the source. In general, because both sources can be different distances from the two field points, the fluctuations at the field points will not be the same. For example, two field points that are the same distances from both sources will be perfectly correlated, or coherent, because they receive the same combination of fields from both sources. This situation is illustrated in Figure 1b. However, if one considers two points that are near both respective sources, the points will usually receive light from the nearby source. Because the sources are incoherent, and each point is mostly receiving radiation from its closer source, these points will be nearly uncorrelated, as shown in Figure 1c.

A spatial distribution of incoherent sources radiating various amounts of power will create a pattern of correlations in the field remote from the source. By using interferometry, the correlation between pairs of field points remote from the sources can be measured. The interference between two points in the electromagnetic field is achieved by relaying the two fields (e.g., with mirrors) to a common location that has an equal travel time for the fields from both points, where the fields are superimposed. The power of the superimposed beams is then detected, by a photocathode, for example. If the two signals being combined are incoherent, then no interference will take place, and the total power measured on the photodetector is simply the sum of the power of the two signals. However, if the two signals are partially coherent, there will be a deviation in the measured power from the sum of the power of the two signals. The magnitude of this deviation relative to the power of the constituent signals indicates the degree of partial coherence of the signal.

An interferometer, such as depicted in Figure 2, gathers the light from two areas of the electromagnetic field remote from an object and combines them together at the same point to form an interferogram. In this example, two telescopes of a known separation gather light from a star. The primary mirror of each telescope is of insufficient size to resolve the star as anything but a point. However, a telescope is used in practice to collect enough light to produce a

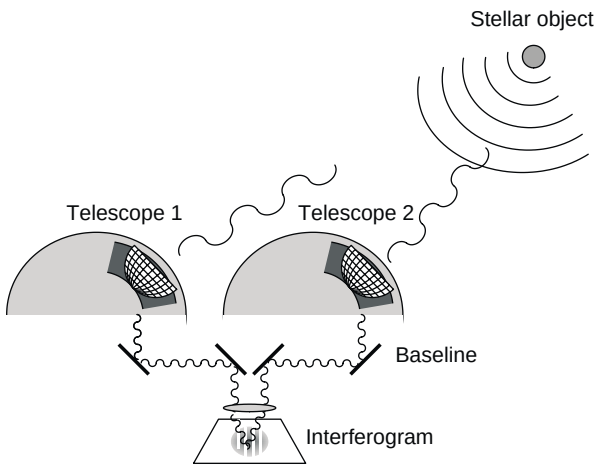

Figure 2 Two telescopes receive radiation from a stellar object. The light is collimated and directed down the baseline to a point halfway between the telescopes, where the two images are interfered together onto a focal plane.

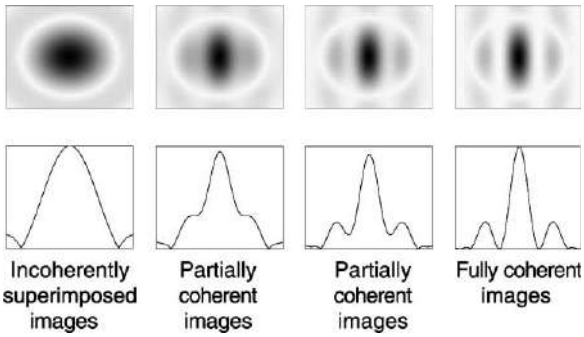

Figure 3 The image of two interfering wavefronts from a point object with various states of coherence.

measurable interference signal. Each telescope primary mirror forms a point image of the star. At the image plane of each telescope a mirror directs the point image to a measurement plane equidistant between the two telescopes. This line that joins the two telescopes is called the baseline, and it is the baseline length rather than the individual telescope mirror size that determines the minimally resolvable size. The image from each telescope is superimposed by mirrors (or perhaps beamsplitters) and imaged onto a focal plane, to get an image similar to that of Figure 3. If the light arriving from one telescope is blocked, then the image formed looks like an Airy ring pattern formed by an image, a point-like object, which is the incoherent image of Figure 3. However, when both signals are allowed to interfere, the image will have vertical stripes superimposed on it. If these stripes are very prominent, so that the image is fully darkened inside the stripes, then an image similar to the rightmost image of Figure 3 is obtained, complete intereference is occurring and the two field points at which the telescopes are situated are coherent or completely correlated. If the stripes disappear, then no interference is occurring and the two field points at which the telescopes reside are incoherent. These examples are shown in Figure 3, with images of two superimposed points with varying degrees of coherence on the top row, and a line plot of the measured field through the center of the images in the bottom row. The degree to which the contrast of the stripes is enhanced by coherence is referred to as the 'modulation' and indicates the degree or magnitude of partial coherence.

In addition to having a degree of coherence, the correlation between two electromagnetic field points also has a relative phase shift. This phase shift is given by the relative position of the interference maxima, as shown in Figure 4. As the relative phase between the

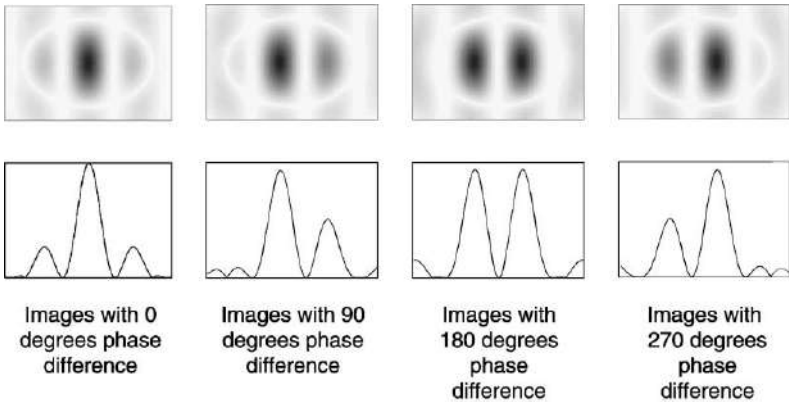

Figure 4 The image of two interfering wavefronts from a point object with various phase differences between the two wavefronts.

fields collected by the telescopes changes (which is a function of the path length difference between the light arriving at the two telescopes), the peaks of the interferogram between the two telescope images shift position. By observing the depth of the formed interference fringes and the position of the fringes, both the amplitude and the phase of the correlation between the two field points can be simultaneously measured. The amplitude of the modulation and the phase of the interference are used to compute the complex degree of coherence that is used in the computation of interferometric images.

In practice, the phase of the complex degree of coherence is much more difficult to measure than its magnitude. This is because while the amplitude tends to change very slowly as the positions of the two correlated points change, the phase is sensitive to changes in relative path length on the order of an optical wavelength. Even though a long baseline produces better resolution, the baseline length must be known to a small fraction of a wavelength to obtain a meaningful phase measurement. Minute vibrations in the positions of the mirrors on the baseline can easily cause fluctuations in the path length greater than a wavelength. This is observed as wild oscillations in the position of the fringes on the image plane. In addition, atmospheric turbulence can also produce large random phase shifts. Achieving a stable phase measurement is one of the most challenging engineering aspects of interferometric imaging.

To model realistic sources, we must determine the coherence produced by sources of a finite size. Finite size incoherent radiators can be regarded as an arrangement of an infinite number of point-like incoherent sources radiating various amounts of power. When an interferogram is made of a field which has two superimposed incoherent sources, the power of the interferograms that each would make alone is added together, rather than the fields of the interferograms. This is because when the fields of the two sources interfere, the phases of both sources are varying randomly over time. The two fields will randomly vary between constructively and destructively interfering with each other. Over a long time, the constructive and destructive interference averages out to no interference. We can then calculate what the interferogram of many incoherent sources combined is by computing the interferogram of a point source and adding the power of many such interferograms together.

To figure out what the power of an interferogram of an arbitrary object is, first consider what the interferogram of a single point source is. Consider a situation depicted in Figure 5, but with a single point

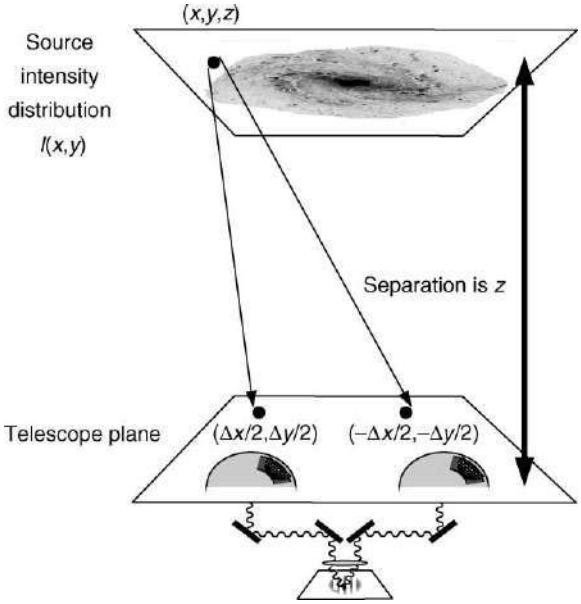

Figure 5 Diagram of source and telescope plane used for derivation of Van Cittert–Zernike theorem. A source at a distance z from the telescope plane has a power density $I(x,y)$. The telescope plane has two movable apertures symmetric about the origin of the telescope plane separated by $(\Delta x,\Delta y)$. The light from both apertures is directed to the center of the baseline at the origin of the telescope plane, where it is interfered on the sensor.

illuminator at the source plane positioned at (x,y,z) with radiant power I. We will have a receiving plane at a distance z from the source plane with two telescopes located symmetrically about the origin at $(\Delta x/2,\Delta y/2,0)$ and $(-\Delta x/2,-\Delta y/2,0)$. The fields received at the telescopes from the point will be interfered together at the center of the baseline, which is at the origin. We will assume that the point source produces a narrow band of radiation centered at wavelength λ. The point source produces a spherical wave that is received at both telescopes. The interference power is proportional to the two intensities measured at each telescope with the relative phase at the two telescopes causing constructive or destructive interference:

$$P \propto I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\phi \qquad [1]$$

where I_1 and I_2 are the intensities of the field at each telescope, and ϕ is the phase difference between the two fields at each telescope. We will assume that the point (x,y,z) is very far away from the telescopes, and so the intensities at the two telescopes are equal and proportional to the point intensity I, so that $P \propto I(1+\cos\phi)$. The phase ϕ is given by the difference in distance from the the source point at (x,y,z) to the receivers at $(\Delta x/2,\Delta y/2,0)$ and $(-\Delta x/2,-\Delta y/2,0)$. The phase is given by the difference in the Euclidean

distances:

$$\phi = \frac{2\pi}{\lambda}\left(\sqrt{z^2 + (x + \Delta x/2)^2 + (y + \Delta y/2)^2} - \sqrt{z^2 + (x - \Delta x/2)^2 + (y - \Delta y/2)^2}\right) \approx \frac{2\pi(x\Delta x + y\Delta y)}{z\lambda} \quad [2]$$

where the second term gives the difference in the distance to a first-order approximation in x and y. The interference intensity relation is then $P \propto I(1 + \cos(2\pi(x\Delta x + y\Delta y)/z\lambda))$.

If the source is now of a finite site, we can regard it as radiating an intensity (power per unit area) $I(x, y)$. To compute the total intensity of the interferogram, we integrate over all the sources in the source plane:

$$P(\Delta x,\Delta y) \propto \int I(x,y)(1+\cos(2\pi(x\Delta x+y\Delta y)/z\lambda))\mathrm{d}x\mathrm{d}y = \int I(x,y)\mathrm{d}x\mathrm{d}y + \int I(x,y)\cos(2\pi(x\Delta x+y\Delta y)/z\lambda)\mathrm{d}x\mathrm{d}y \quad [3]$$

This result can be interpreted as follows. The intensity of the interference when the two telescopes are separated by a distance $(\Delta x,\Delta y)$ is proportional to the real part of the two-dimensional Fourier transform of the intensity distribution $I(x,y)$ evaluated at spatial frequencies $(\Delta x/z\lambda,\Delta y/z\lambda)$, plus a constant intensity. This result, related to the Van Cittert–Zernike theorem, is a commonly used result in interferometric imaging. It is covered in more detail in **Coherence:** Overview.

By moving the telescopes to various horizontal and vertical separations and measuring the interference intensity, the real part of the Fourier transform of the intensity is inferred. In practice the imaginary part can also be found by varying the baseline optical path length difference by a quarter wavelength. With these samples of the Fourier transform, the inverse Fourier transform computes the image intensity distribution $I(x,y)$. This is how simple computational image formation occurs with interferometric data.

One application of the Van Cittert–Zernike theorem and interferometric imaging is the measurement of the diameter of stars. When observing coherence, the correlations between field points tend to decrease as the separation between the points grows. By observing this decrease of the coherence, the angular size of a star can be measured. This method was pioneered by Albert Michelson in 1890 with the invention of the Michelson Stellar Interferometer. It is needed because the aperture size required to measure the angular diameter of most nearby stars in the Milky Way is larger than the size of most telescope apertures. Stars usually have a circular profile that produces a circularly symmetric pattern of correlations. The coherence between field points as the distance between them is increased does not decrease monotonically, but disappears at certain separations, and has many revivals. The first separation distance at which interference disappears (and the degree of coherence is zero) for a circular profile object is $d_0 = 0.61\lambda/\alpha$, where λ is the wavelength of the light measured from the star, α is the angular diameter (in radians), and d_0 is the first separation from zero at which the interference disappears. In practice, each wavelength produces a null in interference at a different separation, so all but a small bandwidth of the star's light is filtered out before measurement to achieve a suitably strong interference null. For example, for the star Sirius, which has an angular diameter of 0.0068 arc seconds viewed from the Earth, measuring at a wavelength of $\lambda = 500$ nm (blue-green in the visible spectrum), the first separation distance at which no interference will be observed is at 18.5 m. This is larger than the largest available telescope aperture, and therefore an image cannot be directly formed that can resolve the size of Sirius.

Modern examples of stellar interferometry include the European Southern Observatory in Paranal, Chile, the CHARA Array at Mt Wilson, California, USA, and the NASA Jet Propulsion Laboratory Keck Interferometer in Hawaii, USA. The ESO main telescope system is the Very Large Telescope Interferometer (VLTI) consisting of four 8.2 m diameter telescopes as well as several smaller ones that can be moved independently, and will achieve milliarcsecond resolution through the interferometric combination of the telescopes. These telescopes are connected by underground tunnels through which the collected light from each telescope is combined. To measure the phase of the interference, extreme positioning precision is required in the relative delay between the light collected by the telescopes along the baseline, of 50 nm over 120 m, or 1 part in 2400 million, equivalent to 1.6 cm over the circumference of the Earth. Achieving this requires combinations of electronic and mechanical feedback systems with laser stabilized interferometric position measurement. The CHARA array has a very large 330-meter baseline, with 1-meter telescope apertures. The Keck Interferometer consists of two 10 m diameter telescopes with a baseline separation of 85 m. One of the main goals of achieving such high

resolution is to be able to observe planets and planet formation around distant stars, to identify possible signs of life in other solar systems.

A similar principle is used for interferometric image formation in radio astronomy. The chief differences are the methods used to collect, detect, and correlate the electromagnetic radiation. Radio astronomical observatories, such as the Very Large Array in New Mexico, USA, and MERLIN at the Jodrell Bank Observatory in Cheshire, UK, use collection dishes and radio antennas to directly measure the electric field, which is not feasible at optical frequencies. These fields are converted to electrical voltages, which can then be directly correlated using electronic mixers, rather than detecting the correlation indirectly through interference. Because the wavelength is much larger, the stability requirements are greatly relaxed. Radio and optical frequency interferometric images are formed in essentially the same way, using the Van Cittert–Zernike theorem.

Interferometric imaging methods have also become of interest for microscopy applications. Aperture synthesis is not needed in microscopy, because the objects of interest are very small and therefore single apertures can be used to achieve diffraction-limited resolution. However, interferometric detection of partially coherent light affords greater flexibility because an image may be computed from the data rather than being directly detected. Typically this is achieved by creating an incoherent hologram of the object, typically through shearing interferometry. For example, by using a Rotational Shearing Interferometer, one can measure an incoherent hologram that can be used to produce a representation of an object with a greatly increased depth-of-field. In addition, direct detection of the partial coherence of sources can allow the methods of computed tomography to be used to produce three-dimensional images of volumes of incoherently emitting objects. Because of the versatility that interferometric detection can provide, interferometric imaging will surely be increasingly used outside of astronomy.

See also

Coherence: Coherence and Imaging; Overview. **Holography, Techniques:** Overview. **Incoherent Sources:** Lamps. **Interferometry:** Overview; White Light Interferometry.

Further Reading

Baldwin JE and Haniff CA (2002) The application of interferometry to optical astronomical imaging. *Philosophical Transactions of the Royal Society of London A* 360: 969–986.

Beckers JM (1991) Interferometric imaging with the very large telescope. *Journal of Optics* 22: 73–83.

Born M and Wolf E (1997) *Principles of Optics*. New York: Cambridge University Press.

Goodman JW (1996) *Introduction to Fourier Optics*. New York: McGraw-Hill Inc.

Goodman JW (2000) *Statistical Optics*. New York: Wiley-Interscience.

Mandel L and Wolf E (1995) *Optical Coherence and Quantum Optics*. New York: Cambridge University Press.

Marks DL, Stack RA, Brady DJ, Munson DC Jr and Brady RB (1999) Visible cone-beam tomography with a lensless interferometric camera. *Science* 284: 2164–2166.

Monnier JD (2003) Optical interferometry in astronomy. *Reports on Progress in Physics* 66: 789–857.

Potuluri P, Fetterman MR and Brady DJ (2001) High depth-of-field microscopic imaging using an interferometric camera. *Optics Express* 8: 624–630.

Thompson AR, Moran JM and Swenson GW (2001) *Interferometry and Synthesis in Radio Astronomy*. New York: Wiley-Interscience.

Lidar

M L Simpson and D P Hutchinson, Oak Ridge National Laboratory, Oak Ridge, TN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Light detection and ranging (LIDAR) refers to an optical system that measures any or all of a variety of target parameters including range, velocity, and chemical constituents. Usually a LIDAR system consists of a transmitting laser source, transmitter optics, receiver optics, and a detector. Several configurations of LIDAR systems are possible, based on the type of laser source and detection scheme selected. Basically, sources can be continuous wave (CW), pulsed, or chirped. Sources that are CW are most often used to measure velocity or chemical constituents. Pulsed and chirped sources

provide the additional capability of measuring range. LIDAR receivers are either incoherent (direct detection – amplitude only) or coherent (heterodyne detection – amplitude and frequency). Generally, for velocity detection, a coherent receiver with a local oscillator (LO) laser is required. A coherent receiver is also needed if a chirped laser source is used, for instance, in frequency modulation continuous wave (FM-CW) LIDAR.

For the special case of imaging LIDAR, the detector has multiple elements. In an imaging LIDAR system, the detector, or focal plane array (FPA), and the receiver optics are designed to image independent volumes in object space onto individual elements in the FPA. Each detector element, therefore, contains information for a specific volume in space and can be thought of as an independent LIDAR system. The advantage of an imaging LIDAR system is that multiple measurements can be made on extended target(s) at the same time. Thus, characterization of rapidly changing events (like wake vortices), speckle reduction by spatial averaging, synthetic aperture radar, multiple target tracking, and target identification are all possible without the sophisticated algorithms and hardware platforms required in single channel scanning systems. In addition because the multiple pixels allow parallel image capture, a true imaging LIDAR can acquire data much more rapidly than a scanning system.

In this chapter, for clarity, the emphasis will be on imaging LIDAR systems applied to hard body targets for the measurement of range and velocity. The basic configurations presented here, though, are extensible to distributed targets and measurements of wind profiles and atmospheric constituents. Also, since general textbooks are available that cover LIDAR systems and incoherent imaging systems, the concentration will be on architectures, optics, detectors, and design considerations that are specifically applicable to coherent imaging LIDAR.

Frequency Modulated Continuous Wave (FM-CW) LIDAR

One LIDAR technique, which is particularly suited for imaging due to the simplicity of the processing, is FM-CW LIDAR. FM-CW LIDAR simultaneously measures range and velocity (Doppler shift) by using a frequency chirped waveform. Referring to Figure 1, assume that a continuous wave (CW) laser frequency is linearly modulated (chirped) by a triangular wave about an offset frequency, such that the modulated signal is given as:

$$f(t) = f_o + \alpha\tau \qquad [1]$$

where f_o is the subcarrier offset frequency and α is the linear modulation rate in frequency/second, and τ is a time variable. The transmitted beam is focused by the optical system onto the target surface, reflected (or absorbed or scattered) and arrives at the receiver after τ seconds, where

$$\tau = 2\frac{R}{c} \qquad [2]$$

and R is the distance to the target surface, and c is the speed of light. The returning signal is optically mixed with the LO on the detector and produces a sum/difference beat frequency (Δf). Figure 1 shows the incident wave with its linear or chirped frequency characteristics and the delayed version returned from the target. Only the difference frequency can be optically detected since only it falls within the detector's bandwidth. Assuming the target is not moving, the result is

$$\Delta f = f(t+\tau) - f(t) = f_o + \alpha(t+\tau) - (f_o + \alpha t) = \alpha\tau \qquad [3]$$

where $\alpha = B/T$, the chirp rate of the modulator in units of Hz/s, B is the modulation bandwidth, and T is the period of the chirp. Substituting into eqn [3] for

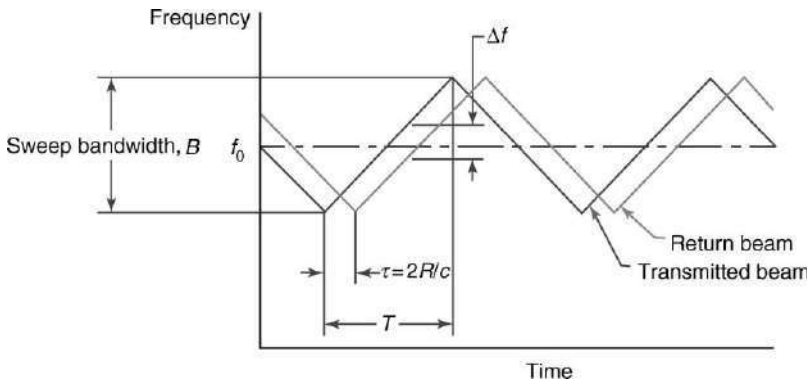

Figure 1 Triangular-wave modulation Envelope on the transmitted beam (black) and return beam (gray).

τ and α we have

$$\Delta f = \frac{B}{T}\left(2\frac{R}{c}\right) \quad [4]$$

Finally, one can solve for the distance to the target, R, as a function of the beat frequency and chirp rate:

$$R = \frac{cT}{2B}\Delta f \quad [5]$$

The difference frequency, Δf, is the same for the up and down ramp of the modulation signal. Figure 2 demonstrates the return signal for the case of a target moving toward the LIDAR system. The difference frequency now contains a component proportional to the target induced Doppler shift given by

$$f_{\mathrm{d}} = 2\frac{v}{\lambda} \quad [6]$$

where v is the line-of-sight target velocity and λ is the laser wavelength. From Figure 2, the Doppler shift causes the difference frequency on the up ramp to be less than the range induced frequency shift given by eqn [5] and the net difference frequency is greater on the down slope. In this case, the difference frequency is given by

$$\Delta f = \frac{B}{T}\left(2\frac{R}{c}\right) + f_{\mathrm{d}} \quad [7]$$

The range and Doppler contributions to the difference frequency may be separated by measuring the frequency difference over both the up and down chirp of the waveform. The Doppler frequency $f_{\mathrm{d}} = \Delta f(\mathrm{up}) - \Delta f(\mathrm{down})$ and the range is given by the average of $\Delta f(\mathrm{up})$ and $\Delta f(\mathrm{down})$ over a complete cycle of the modulating triangular wave. Therefore, substituting in eqns [5] and [6], the range may be determined from the equation

$$R = \frac{cT}{2B}\left\{\Delta f(\mathrm{up}) - \Delta f(\mathrm{down})\right\} \quad [8]$$

and the line-of-sight (LOS) velocity is

$$v = \frac{\lambda}{2}\left\{\frac{\Delta f(\mathrm{up}) + \Delta f(\mathrm{down})}{2}\right\} \quad [9]$$

The range resolution of a FM-CW LIDAR system is the range for which the difference frequency equals 1 cycle in the sweep time interval of T or $\Delta f = 1/T$. Then, from eqn [5], the range resolution, ΔR, is given by

$$\Delta R = \frac{cT}{2B}\frac{1}{T} = \frac{c}{2B} \quad [10]$$

Therefore, the range resolution is only a function of the sweep bandwidth and not the laser wavelength. However, the Doppler shift is dependent on the laser frequency. The accuracy with which the difference frequency may be determined is a function of the number of cycles over which the measurement is made. The number of cycles is maximized for the case of $\tau = T/2$ or T should be set for twice the time delay for the maximum range of interest.

Numerical Example

A range resolution of 30 cm requires a chirp bandwidth, B, of 500 MHz. For a total range of interest of 10 km, the chirp time, T, should be 133.33 μs or the chirp rate should be 3750 Hz (1 up and 1 down ramp per cycle). The difference frequency for a 10 km range target will be 250 MHz, which demonstrates the need for wideband detectors to capture high-resolution range images (Table 1).

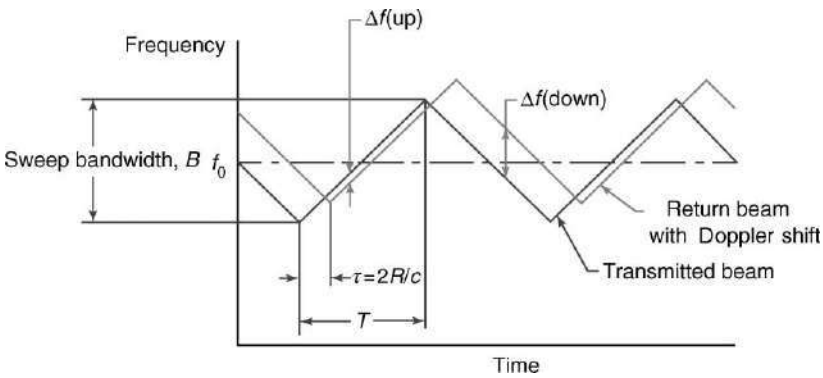

Figure 2 Lidar return with Doppler shift proportional to target velocity.

Table 1 Summary of FM-CW LIDAR characteristics

Parameter	*Value*
Range resolution (m)	0.3
Chirp rate (Hz)	3750
Chirp bandwidth (Hz)	500×10^6
Maximum range (km)	10

Optical Configuration for Imaging LIDAR

Imaging LIDAR systems can be constructed for virtually any wavelength range with appropriate sources, optical components, and detectors. The design of an optical receiver system capable of recording coherent images requires additional considerations over those necessary for a video camera or single-channel receiver. First, the collection optics must have resolution capabilities consistent with both the detector grid dimensions and the coherence properties of the source. Next, the local oscillator (LO) power must be applied to each array element. For a typical infrared (HgCdTe) detector, one to ten milliwatts of LO power might be required to reach the shot noise limit. For large arrays, this may result in unacceptable heat loads, even for momentary application. Thus, detector arrays with substantial spacing between detector elements benefit from the application of individual LO 'beamlets' to shield the insensitive portions of the array substrate. These beamlets can be generated by illuminating a mask containing an array of holes with the LO Gaussian laser beam. This scheme is illustrated in Figure 3.

In the configuration shown in the figure, the signal beam (from the target) and the LO beamlets are combined at the beamsplitter prior to being focused onto the detector plane by the objective lens. The detector plane coincides with the far-field image plane of the objective lens, and any additional optical elements utilized to enhance the camera's resolution are not shown. An expanded LO beam illuminates an aperture array located at the far-field object plane of the ancillary lens. For instance, if the objective lens and the ancillary lens are identical, the LO mask is imaged onto the detector plane with unity magnification. The detector surface is assumed to be flat with uniform video and heterodyne quantum efficiency.

Several aspects of this design are discussed below. We first discuss how the optical parameters may be adjusted to give nearly-Gaussian LO profiles at each detector position. The receiver antenna array is then shown to consist of nonoverlapping sub-beams spaced with the detector array geometry. Signal considerations for both incoherent and coherent sources are then discussed, and it is found that optimal detector size and array-spacing is dependent on the coherence properties of the intended object field.

Aperture Array

As an alternative to holographic LO beam array generation, the LO mask design presented here consists of an array of apertures with size and spacing equal to the detector dimensions. The LO is expanded to coherently illuminate the entire mask with approximately uniform intensity. The mask is located at the front focal plane of the ancillary lens and its image is located at the back focal plane (detector plane) of the objective lens. The coherent image of the mask formed at the detector plane by the ancillary-objective lens combination can be described by a two-stage Fraunhofer diffraction process.

Finite dimensions of the ancillary pupil results in spatial filtering which can affect the image profile of the LO mask. If desired, the ancillary pupil may be chosen to provide nearly-Gaussian LO beamlet profiles with beam waists at each detector location.

It is useful to first consider the coherent image of a single aperture uniformly illuminated with plane LO wavefronts. In one dimension, this wave train diffracts from the aperture to form a far-field diffraction pattern at the Fraunhofer plane of the ancillary lens given by

$$E_{\mathrm{F}}(\beta) = E_{\mathrm{F}} \frac{\sin(\beta)}{\beta} \qquad [11]$$

where $\beta = \frac{1}{2} kb \sin(\theta)$, $k = 2\pi/\lambda$, λ is the wavelength, b is the aperture width, and θ is measured from the optical axis. For small θ, $\beta = \pi b \nu/(\lambda f)$ where ν measures a point on the Fraunhofer plane and f is the focal length of the ancillary lens. The finite diameter D (determined by an aperture stop in the Fraunhofer plane) of the ancillary lens spatially filters the diffraction pattern and this limits the range of β values in eqn [11]. We define an aperture-limited pupil as one which limits β to the range $-\pi < \beta < \pi$; this will be the case when $f/D = b/2\lambda$.

Consider now a linear aperture array consisting of N individual apertures of width b and separation $a > b$. Assume that each aperture is illuminated by plane, uniform, mutually coherent LO wavefronts of equal intensity. In this case, the LO profile at the Fraunhofer plane is characterized by the diffraction pattern for a coherently illuminated linear diffraction grating:

$$E_{\mathrm{F}}(\theta) = E_0 \mathrm{sinc}\,(\beta) \frac{\sin(N\alpha)}{N \sin(\alpha)} \qquad [12]$$

where $\alpha = \frac{1}{2} ka \sin\theta$ and where β, θ, and k are defined above.

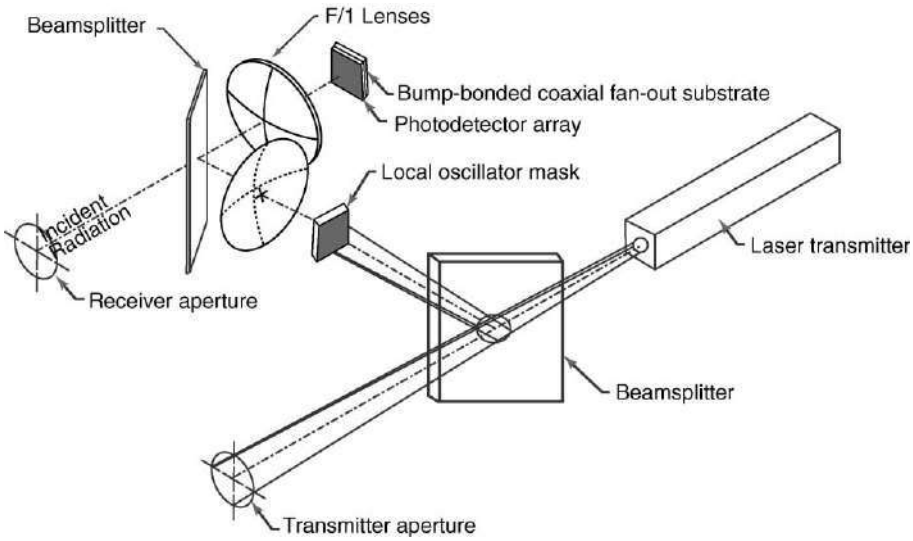

Figure 3 The individual beamlets are generated by the local oscillator or aperture mask. Reproduced with permission from Simpson ML, Bennett CA, Emery MS, *et al.* (1997) Coherent imaging with two-dimensional focal-plane arrays: design and applications. *Applied Optics* 36: 6913–6920.

Propagation of the LO from the ancillary lens Fraunhofer plane to the detector plane can be described by applying the Fraunhofer diffraction integral to eqn [12] with integration limits appropriate to the ancillary pupil dimensions. The larger pupil gives increased illumination at the detector edges while the aperture-limited pupil gives a nearly Gaussian LO profile with $1/e$ points close to the detector edges.

Receiver Antenna Array

The requirement of wavefront matching between the LO field and the signal field leads to the notion of an 'antenna beam' or equivalently, 'backward propagating LO' (BPLO) to define the field of view of an individual heterodyne receiver element. This beam is formed by allowing a reverse projection of the local oscillator field to diffract from the detector surface, through all collection optics in the signal beam path to a point on the object field of the imaging system. A receiver array will produce an array of such BPLO beams which must maintain the detector array geometry in the object field if the LO and signal beams are to mix efficiently.

As above, spatial filtering by the receiver pupil (assumed to be located at the Fraunhofer plane of the objective lens) will affect the BPLO profile. In the Fraunhofer plane of the objective lens, the uniformly illuminated single-detector BPLO profile is given by eqn [11], and we define an aperture-limited detector as one which limits the range of β to $\pm\pi$. Similarly, eqn [12] characterizes the BPLO profile resulting from a detector array uniformly illuminated with an extended uniform LO wavefront. The uniform illumination BPLO profiles in the object field differ from the detector plane profiles only in horizontal scale which now is consistent with the magnified detector dimensions.

For our purposes, we may define the far field of the objective lens according to the condition for maximum Gaussian beam collimation:

$$\omega_d \omega_0 = \frac{f\lambda}{\pi} \qquad [13]$$

where ω_0 is an effective BPLO Gaussian beam waist at the Fraunhofer plane and ω_d is an effective BPLO beam waist at the detector. Under this circumstance, the BPLO Rayleigh range is $Z_D = \pi\omega_0^2/\lambda = f(\omega_0/\omega_d)$. Identifying ω_0/ω_d as the magnification (notice that for the BPLO, the magnification is greater than unity) gives $d_o = (d_i/f)Z_D \sim Z_D$ where d_o is the object distance and d_i is the image distance. Thus the far-field of the objective lens is at least one BPLO Rayleigh range away.

The object field BPLO profiles for an aperture-limited detector array consists of nearly Gaussian sub-beams with $1/e$ points located approximately at the magnified detector edge locations. Since the far-field object distance exceeds Z_D, the divergence of each BPLO sub-beam is constant and remains distinct and nonoverlapping for all $d_o > Z_D$.

Imaging Applications Involving Incoherent Sources

For an object plane in the far field of aperture-limited collection optics, the nearly Gaussian-profile BPLO propagates from the receiver aperture according to

the condition for maximum Gaussian beam collimation. The waist is at the Fraunhofer plane where the BPLO wavefront is nearly plane, while at the object field the BPLO wavefronts are spherical with radius equal to

$$R(d_o) = d_o\left[1 + \left(\frac{Z_D}{d_o}\right)^2\right] \qquad [14]$$

where we have set $d_o - f \approx d_o$.

It will generally be the case that an imaging system will be used to analyze extended objects whose size exceeds the resolution limit. Following convention, we define a spatially coherent source as one which is illuminated with spatially coherent light with appreciable structure and roughness of depth within the coherence length of the illuminating light, and where the target motion does not appreciably perturb the speckle field during the measurement process. The speckle effects, which so seriously degrade the information content in images formed of spatially coherent objects, are not present in images formed of incoherent objects such as thermal emitters. Speckle may be reduced by time-averaging speckle fluctuations resulting from target motion, atmospheric effects, and independent transmitter pulses, and in the limit such objects become incoherent pseudothermal sources. Similar remarks apply to spatial averages over detector subarrays.

Independent radiators on the object field produce disturbances which can diffract to illuminate the entire receiver aperture as well as a finite coherence area in the image field A_{ci} given by

$$A_{ci} \gg \frac{\lambda^2}{\Omega_{ci}} \qquad [15]$$

where Ω_{ci} is the solid angle subtended by the receiver aperture at the image field. Alternatively, we may interpret A_{ci} as the area of an aperture in the object plane which, if illuminated with wavefronts characterized by eqn [14] (reciprocal illumination), would just fill the receiver aperture with a central diffraction order. Thus the combination of independent disturbances within the detector's field of view on the object plane, combine to form a spatially coherent disturbance over the receiver aperture and within A_{ci}.

According to the Van Cittert–Zernike theorem, the complex degree of spatial coherence μ_s is a unity amplitude wavetrain that propagates exactly as a signal beam resulting from the reciprocal coherent illumination of an object-field aperture of area A_{ci}. Thus, μ_s and the BPLO are Helmholtz-reciprocal pairs whose overlap determines the mixing efficiency with which heterodyne signal is generated from radiation collected from an extended incoherent source. Following convention, we refer to the overlap of μ_s and the BPLO as the effective receiver aperture.

Detectors larger than the aperture-limit, whose area exceeds A_{ci}, exhibit reduced receiver apertures and a corresponding decrease in heterodyne mixing efficiency. For example, varying the receiver pupil, while maintaining constant LO illumination, results in a heterodyne signal proportional to the received power for detector diameters much less than the aperture-limit, while detector diameters much greater than the aperture-limit produce heterodyne signal only in proportion to the square-root of the received power. Thus, for constant LO illumination and incoherent sources, increasing the entendue past the aperture limit will always increase the heterodyne signal, but less than for direct detection.

Optimal Detector Size

As noted above, an aperture-limited system is one where detectors of diameter d are matched with an objective lens with $f/D = d/2\lambda$. Commercially available detectors for use at 10 μm are commonly 100 μm wide suggesting the use of a $f/5$ lens, however aspheric lenses as fast as $f/1$ are also commercially available. Apart from aberrations, it would be advantageous to utilize a 5×5 subarray of 20 μm detectors with an $f/1$ lens in place of a single 100 μm detector for applications involving coherent sources. Applications involving incoherent thermal sources would generally favor a wide detector/fast lens combination since thermal sources provide weak heterodyne signals.

Noise Analysis and Performance

Radar Range Equation

The laser power requirements for the LIDAR system depend on the signal-to-noise ratio of the FPA, the i.f. amplifier performance, atmospheric transmission, and the optics system design. The following calculations are used to determine the performance of the LIDAR system for a single pixel of the FPA. The radar range equation is

$$P_r = P_t\left(\frac{G_t}{4\pi R^2}\right)\left(\frac{\sigma_{pt}}{4\pi R^2}\right)\left(\frac{\pi D^2}{4}\right)\eta_{atm}\eta_{sys} \qquad [16]$$

where P_r = received power; P_t = transmitted power; G_t = antenna gain; σ_{pt} = radar cross-section; R = range; D = transmitting and receiving aperture diameter; η_{atm} = atmospheric transmission; and η_{sys} = optical system efficiency.

The beam divergence of the transceiver is

$$\theta_t = \left[\left(\frac{D}{R}\right)^2 + \left(\frac{\lambda}{D}\right)^2\right]^{0.5} \quad [17]$$

where λ = wavelength and the antenna gain is

$$G_t = \frac{4\pi}{\theta_t^2} \quad [18]$$

The radar cross-section for a lambertian diffuse point target is $\sigma_{pt} = 4\rho_t dA$. The target reflectivity is ρ_t and dA is the target area, which is effectively the size of a pixel of the FPA projected onto the target. For purposes of the following calculations, a surface reflectivity of 0.1 is assumed. The pixel size at the target (assuming Gaussian beam optics) is given by

$$dA = \pi\left(\frac{D}{3}\right)^2\left[1 + \left[\frac{\lambda R}{\pi}\left(\frac{3}{D}\right)^2\right]^2\right] \quad [19]$$

where D = transmitter and receiver mirror diameter and it is assumed that the target is larger than dA.

Finally, the signal-to-noise (SNR) ratio for a coherent receiver is given by

$$\mathrm{SNR} = \frac{\eta P_r}{h\nu B_{if}} \quad [20]$$

where η = detector quantum efficiency; P_r = received power from eqn [16]; h = Planck's constant; ν = frequency of laser; and B_{if} = post-detection bandwidth of signal processor.

Numerical example

Assuming an atmospheric transmission efficiency, η_{atm}, of 98%, an optical system efficiency, η_{sys}, of 30%, and a detector efficiency, η, of 10%, the SNR was calculated as a function of the transmitting/receiving telescope (assumed to be a mirror) diameter, D, with target range as a parameter and plotted in Figure 4. This calculation assumed that the pixel field-of-view (FOV) at the target was illuminated with 1 watt of power. Even at a range of 10 km, the 32 × 32 element imaging LIDAR receiver with a SNR of 100 for each pixel will require only 75 watts (approximately 74-milliwatts per pixel) of total transmitted power for a mirror diameter of 30 cm. From Figure 4, it appears that the telescope mirror size could be reduced to a diameter of 20 cm and still maintain an excellent SNR. However, the SNR is not the only consideration on the transmitting telescope diameter. The cross-range target resolution is determined by the projected pixel size at the target. This target pixel size is plotted versus the telescope mirror diameter in Figure 5. Diffraction from the transmitting and receiving optic sets the limiting resolution that can be achieved with the LIDAR system and is only a function of the diameter and wavelength. A transmit/receive telescope aperture of 30 cm seems to represent a good trade-off between cross-range resolution and SNR.

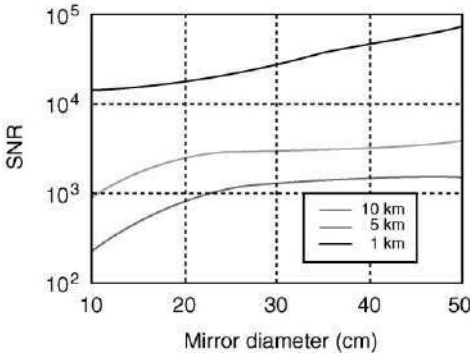

Figure 4 The SNR saturates for a mirror diameter greater than 30 cm at long range.

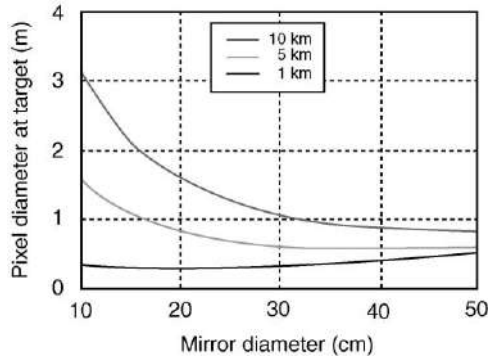

Figure 5 The pixel size determines the target cross-range image resolution.

Applications and Future Directions

The remote sensing community has recognized the advantages of coherent receivers over conventional direct IR detection with respect to increased measurement sensitivity and multidimensional information capability. These advantages, however, come at a price of added system complexity, power, size, weight, and cost. In addition, choices of components, particularly in the long wavelength infrared (LWIR) region of the spectrum, have been limited to liquid nitrogen-cooled nonimaging detectors and table-top-sized gas lasers. Thus, implementations of imaging LIDAR systems have been primarily single detector, single source scanning systems which require high pulse repetition frequency lasers, complex tracking systems, and platforms, and augmentation of automated target recognition input to compensate for multiple, simultaneous measurements on the

same target. New technology developments in quantum well infrared photodetector (QWIP) detectors and arrays, and in HgCdZnTe technology are now paving the way toward thermo-electrically cooled LWIR imaging array detectors. Also waveguide, folded cavity, laser resonator designs are providing compact, lightweight packages for CO_2 lasers. Coherent receivers, that use the new focal plane arrays and compact sources, provide the capability of making parallel LIDAR measurements without the added hardware and software overhead of scanning systems. In the following section, several examples are provided that demonstrate the versatility of imaging LIDAR systems that use detector arrays for measurement applications such as chemical plumes, velocity gradients, and speckle reduction.

Wideband heterodyne receivers are useful in measuring chemical species whose absorption peaks are not well aligned with the laser source line. If the species absorption peak corresponds directly with the laser line, an absorption measurement can be made using direct detection (amplitude only). For off-line (off-resonance) measurements, heterodyning is needed, where the bandwidth of the heterodyne receiver dictates how far away from the laser line an absorption peak can be measured. For example, the CO_2 laser line spacing in the 9 to 12 micron region of the spectrum averages around 50 GHz. With the new QWIP detectors theoretically having heterodyne bandwidths of 30 GHz or more, full high-resolution spectral coverage for the long-wave IR is within grasp. For chemical detection with heterodyne receivers, there are two modes of operation. The first is termed passive heterodyne radiometry where the source is a black-body radiator and the absorption is measured for chemical species located between the blackbody and the receiver. The second mode of operation is thermoluminescence where a laser transmitter is used to excite a chemical species in the atmosphere and the resulting emission (luminescence) is measured with the heterodyne receiver. The advantage of imaging in both cases is the ability to capture chemical plumes in the presence of wind and other anomalies.

To illustrate the ability of heterodyne imaging receivers to spectroscopically resolve chemical plumes, Figure 6 shows a series of 10×10 pixel images of a small bottle of concentrated ammonium hydroxide against a 212 °C blackbody recorded in both passive heterodyne radiometry and direct detection modes. Figure 6a shows a heterodyne image of the bottle with the top in place where, in Figure 6b the top is removed and the absorbing plume is clearly imaged. The heterodyne images in Figures 6a and 6b were recorded with the LO adjusted to emit the 9R(30) line which is known to

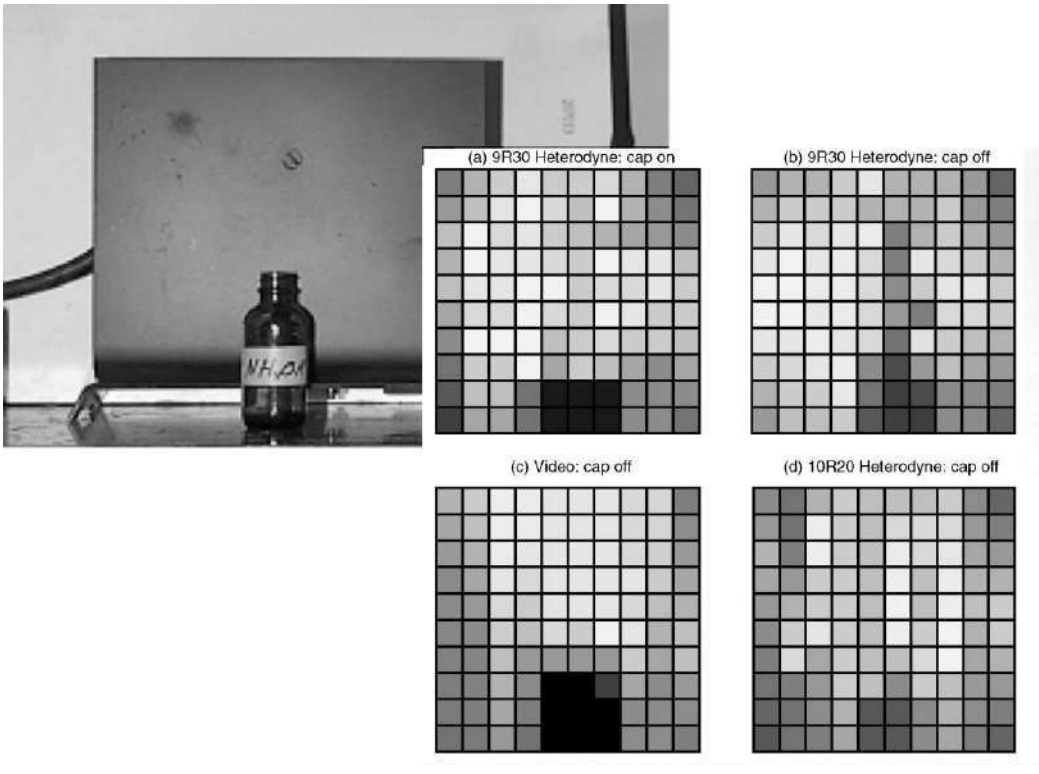

Figure 6 Ammonia plume characterization.

be absorbed strongly by NH_3. In Figure 6c, the same image was recorded by direct detection and fails to detect the ammonia plume. Similarly, in Figure 6d, the LO was adjusted to emit a nearby line known not to be absorbed by NH_3 and the plume is again not detected. Both the objective and ancillary lens were $f/1$ 38 mm, and the object field was about three meters away. The top of the bottle was approximately 2 cm in diameter and is located at the bottom of the field of view. Measurements utilizing hot ammonium hydroxide imaged against a cold background resulted in similar images (albeit at substantially reduced signal to noise ratios) where the plume was bright and the background was dark.

Another use of imaging LIDAR is the measurement of velocity profiles. Applications include wake vortex measurements for aircraft, wind shear detection, and vibration signature measurements for identification. In another experiment, active mode images were recorded by illuminating the target with a portion of the LO laser beam which had been shifted in frequency by 40 MHz using an acousto-optic modulator. The resulting Doppler image is recorded using an 8×8 HgCdTe array with discrete electronics for each pixel element in the array. Figure 7 shows the active image of a vertical squirrel cage fan where each pixel is rendered to represent the peak Doppler shift measured with the detector at the corresponding location in the image plane. The moving target scatters incoherently and thus speckle effects do not degrade the image. The measured Doppler shifts are consistent with the known rotational velocity of the target.

Speckle effects associated with coherent mixing of laser light reflected from an extended target (constructive and destructive interference) are a major source of noise in LIDAR measurements. These speckle effects can be reduced by signal averaging. For single detector LIDAR, averaging over many transmitter pulses can reduce speckle but this technique also reduces temporal resolution.

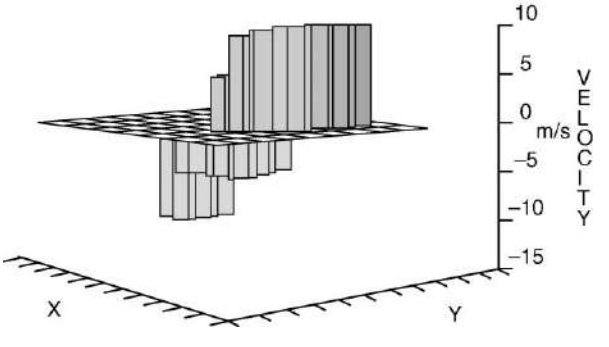

Figure 7 Doppler image of a rotating squirrel cage fan.

The alternative considered here is to sacrifice spatial resolution by averaging over an array of detectors, all of which record the signal returned from a single transmitter pulse. Detectors separated by a coherence area diameter will record signals that approach statistical independence, and in this case speckle effects can be reduced by an amount approaching the square-root of the number of detectors in the subarray.

The following laboratory experiment demonstrates speckle reduction by spatial averaging with a fixed focal plane array. The receiver used in this experiment consists of a 3×3 array of HgCdTe detectors. The 50-micron diameter detectors are arranged in a square pattern, with 100 micron center-to-center spacing. Custom electronics multiplex the output of each detector for subsequent processing. As with the velocity profile experiment, LIDAR illumination is produced by frequency shifting a portion of the LO beam by 40 MHz with an acousto-optic modulator prior to illuminating the object field. Polarization of the illumination beam in the experiment is parallel to the LO prior to scattering from the target. The measured receiver bandwidth, limited mostly by the dewar design, is about 100 MHz, which permits the acquisition of the 40 MHz heterodyne signal. The collection optic consists of a 38 mm $f/2$ asphere which produces a 2 mm diameter pixel image at a 3 m object distance.

The target for this experiment is a 30 cm diameter disk coated with an aluminum powder. The particle size of this powder ranges from 1 to 100 microns as determined by electron microscopy. The disk is attached to a stepper motor that is incremented between frames to provide statistically independent speckle fields. Figure 8 shows the distribution of

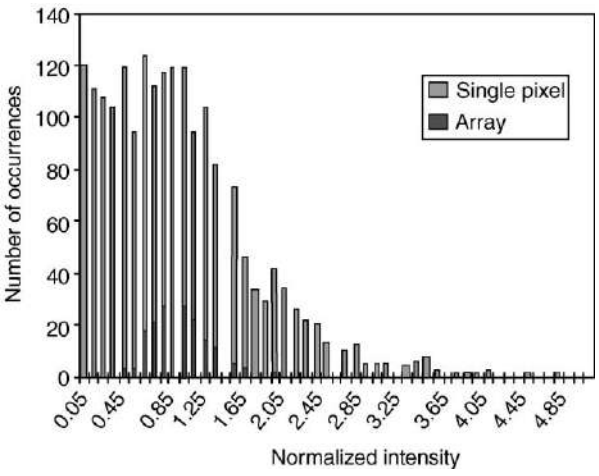

Figure 8 Speckle reduction using spatial averaging over a coherent array.

measurements taken with a single pixel superimposed on the distribution taken by averaging over the focal plane array. A reduction in speckle noise by a factor of the square root of the number of averaged pixels is obtained as predicted.

In this chapter, an overview of imaging LIDAR systems has been presented. Imaging LIDAR system design, performance, and application were discussed. With new technology developments in wide-bandwidth QWIP detector arrays, thermo-electrically cooled detectors, and compact sources, substantial potential exists for imaging LIDAR systems in the next generation in remote sensing. Due to the parallel, multifunction measurement capability of these new receivers a single sensor system can provide three-dimensional information, including range, shape, velocity, and chemical composition. In addition to the sensor fusion advantages, snapshot imaging allows the capture of high-speed, rapidly evolving events that elude present scanning systems.

See also

Environmental Measurements: Doppler lidar. **Imaging:** Infrared Imaging.

Further Reading

Born M and Wolf E (1964) *Principles of Optics*. London: Pergamon.

Forrestor AT (1956) On coherence properties of light waves. *American Journal of Physics* 24: 192–196.

Goodman JW (1975) Statistical properties of laser speckle patterns. In: Dainty JC (ed.) *Laser Speckle and Related Phenomena*, Chap. 2, pp. 9–75. New York: Springer-Verlag.

Haner AB and Isenor NR (1970) Intensity correlations from pseudothermal light sources. *American Journal of Physics* 38: 748–750.

Holzhauser E and Massig JH (1978) An analysis of optical mixing in plasma scattering experiments. *Plasma Physics* 20: 867–877.

Jelalian AV (1992) *Laser Radar Systems*, pp. 3–5. Norwood, WA: Artech House.

Kogelnik H and Li T (1966) Laser beams and resonators. *Applied Optics* 5: 1550–1566.

Kusters JF, Rye BJ and Walker AC (1989) Spatial weighting in laboratory light scattering experiments. *Applied Optics* 28: 657–664.

Martienssen W and Spiller E (1964) Coherence and fluctuations in light beams. *American Journal of Physics* 32: 919–926.

Measures RM (1984) *Laser Remote Sensing Fundamentals and Applications*. New York: John Wiley.

Rye BJ (1979) Antenna parameters for incoherent backscatter heterodyne lidar. *Applied Optics* 18: 1390–1398.

Shannon RR (1997) *The Art and Science of Optical Design*. New York: Cambridge University Press.

Siegman AE (1966) The antenna properties of optical heterodyne receivers. *Applied Optics* 5: 1588–1594.

Simpson ML, Bennett CA, Emery MS, *et al.* (1997) Coherent imaging with two-dimensional focal-plane arrays: design and applications. *Applied Optics* 36: 6913–6920.

Veldkamp WB (1983) Holographic local oscillator beam mutiplexing for array heterodyne detection. *Applied Optics* 22: 891–900.

Veldkamp WB and Van Allen EJ (1983) Binary holographic LO beam multiplexer for IR imaging detector arrays. *Applied Optics* 22: 1497–1506.

Zernike F (1938) The concept of degree of coherence and its application to optical problems. *Physica* V8: 785–795.

Multiplex Imaging

A Lacourt, Université de Franche-Comté, Besancon, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The ability to project an image onto a screen in a darkened room has been known since antiquity and Leonardo da Vinci (1515) precisely described principles of the Camera oscura. When one exposes a paper sheet directly to sunlight, it is uniformly lighted. However, if one pierces a little hole through the paper sheet then an image of the sun can be displayed on a screen beyond the hole (Figure 1). Similarly, the shape of the light spot we can see in a dark room, from the rays passing through a small aperture of a closed shutter, is not that of the aperture. Generally, the spot has a circular shape, as an image of the sun. Actually, it is neither a shadow of the hole nor an image of the sun but a combination of both. Light incident on an aperture contains the entire information about the light distribution of the object, but this information has to be 'deblurred', or decoded. The aperture acts as an elementary spatial 2D-filter that selects and preserves the direction and intensity of light rays coming from each point of the scene. The projection on the screen of each pencil of light passing through the hole is a small light spot, the geometry of which is the same as the aperture, and

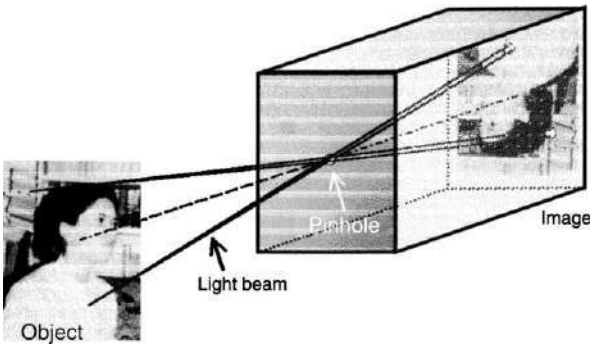

Figure 1 Principle of darkroom.

the intensity is proportional to that of the emitting point. The smaller the hole, the more detail the image exhibits, but with less luminosity. Thus, we must adopt a compromise between the resolution and the illumination of the image.

Now, if the mask has several holes, the amount of light reaching the screen, being proportional to the whole aperture area, is extensively increased. On the other hand, as each hole displays an image on the screen, the overlapping of these multiple images gives a new but blurred image. However, the information exists and we should be able to retrieve it if we know the characteristics of the complex aperture. This is the basic concept of multiplex imaging. The crucial question is: what is a good aperture design and a good way to reconstruct images having high resolution and with high illumination levels? The suitable masks with this aim are named synthetic apertures.

Porta (1589) and Kepler (1611) found the perfect solution, by setting up a lens in front of a large hole. Indeed, one can regard a lens as a wide aperture made of continuously juxtaposed microprisms (Figure 2). The lens obtains such refractive properties that the image displayed by each of these microprisms is moved toward a center. Moreover, as the angle of the microprisms varies continuously, it focuses on a single point, all the rays coming from a given object point. Thus, the blurring is suppressed and the illumination is increased. The way was opened toward the invention of optical instruments and photography and the 'pinhole' has fallen into oblivion for a long time.

Two new problems awake interest in synthetic apertures, around 1960. First, the domain of imaging has extended out of visible light: infrared and ultraviolet light, γ-rays, X-rays, ultrasound waves, and others. Materials suitable for making lenses or mirrors, operating with such waves, are expensive or often do not exist. For example, most of the transparent materials have a refractive index close to 1 for γ-rays and X-rays. However, some are opaque enough to enable the use of masks.

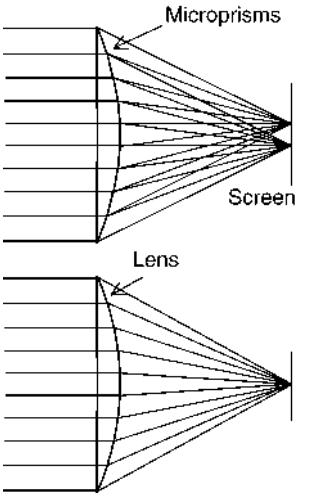

Figure 2 A lens can be modeled as a juxtaposition of microprisms with continously variable angles.

The second challenge appeared with progress in astronomy. In order to explore the universe more, further from Earth and closes to the Big Bang, astrophysicists needed telescopes providing very high luminosity and high resolution, since distant stars, nebulae, and other galaxies are extremely weak and small luminous objects. This exploration can be achieved by increasing the diameter of the primary mirror of telescopes. Indeed, due to diffraction of light by the pupil of the primary mirror, the theoretical limit of resolution of a telescope (i.e., the smaller distance between two points that one can discriminate), is inversely proportional to the diameter of the mirror. In addition, the illumination is proportional to its area. However, the working of large mirrors is a technical prowess that is very expensive and has high risks. Defect of a fraction of micrometer when polishing the surface can have dramatic consequences, e.g., Hubble. Also, such mirrors, being heavy, can bend out of shape under their own weight. Nevertheless, outstandingly large telescopes have been implemented, based on Adaptive optics: the deformations of the mirror are measured in real time by an interference method and compensated by activating a mosaic of jacks (New Technology Telescope – NTT – European program, 1989). Another way consists in replacing the solid-state mirror by a mosaic of 36 mirrors accurately directed (Keck telescope, USA). However, in either case, the mirror diameter does not exceed 8 to 10 meters. Synthetic apertures could be a solution to that difficulty, as earlier demonstrated by Fizeau (1868) and Michelson (1890).

In the first part of this article, any fundamentals are briefly recalled about properties of light and imaging.

The second part describes recent experimental methods and results obtained in synthetic aperture imaging. The last part reports the important programs in progress today relating to stellar interferometry.

Background

Electromagnetic Waves Propagation

Light power, which is involved in image formation, is carried by the electric field of an electromagnetic wave (Figure 3). The electric field $\vec{E}$ has properties of a vector orthogonal to the direction of propagation $\vec{N}$ of the light ray. It is characterized by its amplitude $\vec{E}_0$, the angular time frequency of vibration ω, and the wavevector $\vec{k}$, which specifies the propagating direction of the ray. Let $P(\vec{d})$ be a current point on a plane wavefront, defined by the position vector $\vec{d}$ from the origin of the reference frame. The electric field propagation of a monochromatic wave is expressed as:

$$\vec{E}(\vec{d}, t) = \vec{E}_0 \exp[\mathrm{i}(\vec{k}\cdot\vec{d} - \omega t)] \qquad [1]$$

The light power carried by the electromagnetic field is proportional to $I = \|\vec{E}_0\|^2 = \vec{E}\cdot\vec{E}^*$ where $\vec{E}^*$ is the complex conjugate of $\vec{E}$. The time frequency $\nu = \omega/2\pi$, or the wavelength $\lambda = c/\nu$, where c is the electromagnetic wave celerity, characterize the hue of the light. The wavevector $\vec{k} = \vec{N}\omega/c$, where $\|\vec{N}\| = 1$ is an angular space frequency vector, as $\vec{n} = \vec{N}/\lambda = \vec{N}\nu/c$ is a space frequency vector. We have $\vec{k} = 2\pi\vec{n}$. As λ is a constant for a given monochromatic light, $\vec{n}$ defines the direction of propagation of a ray.

Spatially Coherent Light

Consider a very small source stated as the point source S_1, emitting uniform lightwaves in any direction (Figure 4). The wave reaching a farfield point, P, from S_1, propagates along a direction defined by the wavevector $\vec{k}_1$. The vectors $\vec{k}_1$ and $\vec{d}_1 = \overrightarrow{S_1P}$ are collinear. Another point source, S_2 is characterized from P by the wavevector $\vec{k}_2$, such that $\vec{k}_2$ and $\vec{d}_2 = \overrightarrow{S_2P}$ are collinear. When the distances d_1 and d_2 are close to each other, the amplitudes $\vec{E}_{01}$ and $\vec{E}_{02}$ of the waves at P are proportional to those of the corresponding emitting sources, S_1 and S_2. According to the principle of superposition, the resulting field at the point P, $\vec{E} = \vec{E}_1 + \vec{E}_2$, and the instant wavevector $\vec{k}$, are unique vectors.

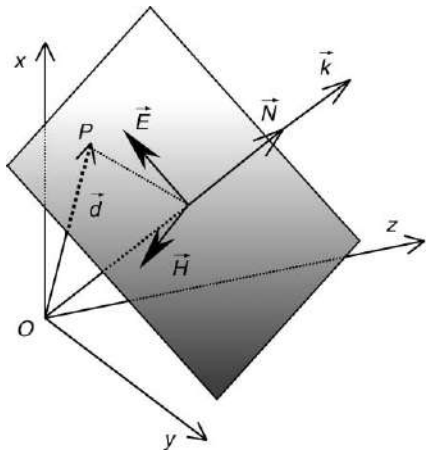

Figure 3 Electromagnetic wave propagation.

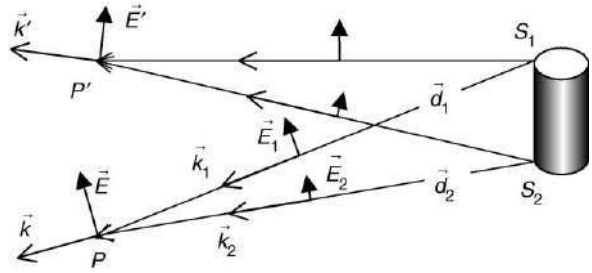

Figure 4 Spatially coherent light.

If S_1 and S_2 are two synchronous sources, the amplitude $\vec{E}_0$ of $\vec{E}$ and the wavevector $\vec{k}$ are constant versus time at any point P. This is still the case when an extended object is made of an arrangement of synchronous points. That situation defines a spatially coherent object. Thus, one cannot recover multiple data about the object's distribution, from the field framework at a single point P, since the relationship between the two is multi-unequivocal. Nevertheless, due to diffraction, one can retrieve them by knowing the complex amplitude distribution of light on a set of wavevectors available over an extensive area of a screen. Let $\vec{E}_0(\vec{r})$ be the amplitude of the electric field at the object point $M(\vec{r})$ and let $\vec{E}_0(\vec{n})$ be the one diffracted by the object to farfield toward the direction corresponding to the space frequency $\vec{n}$. Fourier transforms give the connection between $\vec{E}_0(\vec{r})$, and $\vec{E}_0(\vec{n})$:

$$\left\{\begin{aligned} \vec{E}_0(\vec{n}) &= \int \vec{E}_0(\vec{r})\exp[\mathrm{i}2\pi\vec{n}\cdot\vec{r}]\mathrm{d}\vec{r}, \\ \vec{E}_0(\vec{r}) &= \int \vec{E}_0(\vec{n})\exp[-\mathrm{i}2\pi\vec{n}\cdot\vec{r}]\mathrm{d}\vec{n}. \end{aligned}\right\} \qquad [2]$$

This is the situation encountered when an object is illuminated by a laser or by a point source such as a single star.

Spatially Incoherent Light

When the point sources S_1 and S_2 are independent – spatially incoherent – both wavevectors $\vec{k}_1$ and $\vec{k}_2$

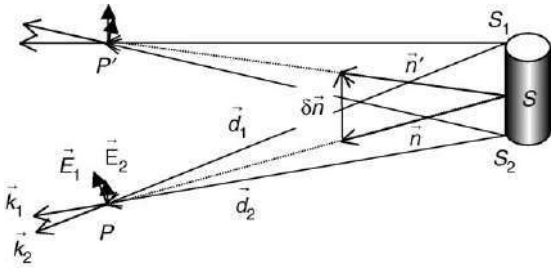

Figure 5 Spatially incoherent light.

coexist at the point P, each characterizing the angular position of S_1 and S_2 (Figure 5). If S_1 and S_2 are part of an extended intensity distribution $I_0(\vec{r})$, fans of wavevectors and fields are mixed at P, each coming from an independent point source S. Together, they carry the whole data on the object distribution. Those data are partially uncorrelated at any two points P and P' on a screen, related to the different optical paths from any point S to P and P' and to the finite time of coherence of light. The correlation between data available to farfield in two directions $\vec{n}$ and $\vec{n}'$ from the object, is given by the mutual intensity function, $\Gamma(\delta\vec{n}) = \int \vec{E}(\vec{n})\cdot\vec{E}^*(\vec{n}+\delta\vec{n})\mathrm{d}\vec{n}$, where $\delta\vec{n} = \vec{n}' - \vec{n}$. Fourier transforms give the relation between $I_0(\vec{r})$ and $\Gamma(\delta\vec{n})$:

$$\left\{\begin{array}{l} \Gamma(\delta\vec{n}) = \displaystyle\int I(\vec{r})\exp[\mathrm{i}2\pi\delta\vec{n}\cdot\vec{r}]\mathrm{d}\vec{r}, \\ I(\vec{r}) = \displaystyle\int \Gamma(\delta\vec{n})\exp[-\mathrm{i}2\pi\delta\vec{n}\cdot\vec{r}]\mathrm{d}\delta\vec{n} \end{array}\right\} \quad [3]$$

Therefore, by drawing a map of $\Gamma(\delta\vec{n})$, one can recover the object distribution $I_0(\vec{r})$ by a Fourier transformation. Most of time, one uses the spatial degree of coherence function, $\gamma(\delta\vec{n}) = \Gamma(\delta\vec{n})/\Gamma(\vec{0})$. When the object illuminates a screen, at a large distance d, the degree of coherence of light between two points P and P' on the screen is $\gamma(\delta\vec{\rho}/\lambda d)$, with $\delta\vec{\rho} = \overrightarrow{PP'}$. The 2D extension of $\gamma(\delta\vec{\rho}/\lambda d)$ defines the area of coherence of light on the screen. The width of the area of coherence, $\Delta\rho$, and that of the object distribution, Δr, are inversely proportional:

$$\Delta\rho\cdot\Delta r \sim \lambda d \quad [4]$$

If a mask perforated by two pinholes, separated by a distance $\delta\vec{\rho}$, replaces the screen, one can observe Young fringes of interference beyond the mask, the visibility of them being precisely $\gamma(\delta\vec{\rho}/\lambda d)$. Therefore, one can built a map of $\gamma(\delta\vec{\rho}/\lambda d)$ versus $\delta\vec{\rho}$ from a set of 2D measurements of visibilities of fringes, then reconstruct the object distribution by a Fourier transformation.

Notice that the entire data of the object distribution is available from a single coherence area. That is, a pinhole covering exactly one area of coherence would select the entire available data from the object and would display to farfield the most accurate image possible. Generally, such a pinhole is so small that practically no light passes through it. However, we can regard each area of coherence as an independent channel of information. Then an imaging device based on a juxtaposition of such reduntant channels can be considered as a multiplex imaging device.

Self-luminous objects such as extended or double stars, and artificial thermic sources are spatially incoherent sources.

Multiplex Imaging

Methods of multiplex – or lensless – imaging are performed either with coherent of incoherent light. The very different properties of the two types of light leads to completely different methods, even if all are related to holography. Holography with coherent light is an interference method for local recording of both the direction and modulus of the wavevectors and the amplitude of the field contained in many juxtaposed grains of speckle, scattered by the object. Each grain is an independent channel of imaging.

Synthetic Apertures with Spatially Incoherent Objects

As interferences are impossible by this method with incoherent light, we must improve other methods of imaging. Synthetic aperture is a concept consisting of using a redundant multichannel system to increase the luminosity of images without loss of resolution. The basic principle is that of a multiple holes darkroom. The processing is in two steps: recording and reading the image. The recording step consists in illuminating a photosensitive receptor with the object through a mask (Figure 6). Generally, the receptor is a

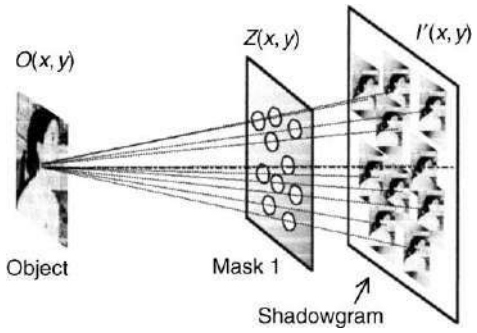

Figure 6 Principle of synthetic aperture: recording.

photographic plate. Obviously, the light distribution on the plate is blurred. Indeed, any elementary transparencies on the mask project on the receptor inverted images of the object located the shadow of the mask. Such a distribution is expressed by a convolution. Let $O(x, y)$ be the conical projection of the object on the receptor and $Z(x, y)$ that of the mask. The shadows distribution on the receptor is (the variable being normalized):

$$I'(x', y') = Z(x, y) * O(x, y) = \int Z(x, y)O(x' - x, y' - y)\mathrm{d}x\,\mathrm{d}y \quad [5]$$

The reading step is a decorrelation processing from the recorded intensity (Figure 7). It consists in achieving a new correlation of I' with an arbitrary function T:

$$I''(x, y) = T(x, y) * [Z(x, y) * O(x, y)] = [T(x, y) * Z(x, y)] * O(x, y) \quad [6]$$

$R[x, y] = [T(x, y) * Z(x, y)]$ is the system point spread function (SPSF) of the processing. It expresses the image distribution of a single-point. The perfect image is obtained when $R(x, y)$ is a 2D *Dirac* – or *delta*– distribution, $\delta(x, y)$:

$$R(x, y) = \delta(x, y), \quad \text{as} \quad \int \delta(x, y)O(x' - x, y' - y) \times \mathrm{d}x\,\mathrm{d}y = O(x', y') \quad [7]$$

Fresnel zone plates

First Mertz and Young used Fresnel zone plate (FZP) as a synthetic aperture for X-ray imaging purposes. FZPs are made of alternate transparent and opaque concentric crowns of equal area. The radius of the nth circle is $r_1\sqrt{n}$, r_1 being the smallest. When illuminated by a cylindrical light beam, the FZP is a multiple foci diffractive lens (Figure 8).

The reading step is either analogical or numerical. Through the analogical method, we consider the recorded plate as a pseudohologram (Figure 9). When the 'shadowgram' is lighted with a parallel beam of laser light, the FZPs recorded around each point concentrates the light on its foci. Then an image of the object is reconstructed on the focal planes; a method that is simple and elegant. Nevertheless, the processing is not linear since we record an intensity distribution at the recording step, while the reading step by diffraction involves amplitude. Moreover, the S/N ratio decreases as the amount of data increases, recorded on the photographic plate because of its finite dynamic.

Mertz and Young had the idea of using a second FZP to perform the deconvolution processing by optical methods. In this case, the SPSF, $R(x, y)$, is the autocorrelation function of the FZP transparency (Figure 10). It presents a narrow central peak, emerging from a pyramidal ground (Figure 11). The optical setup is the same as the recording step.

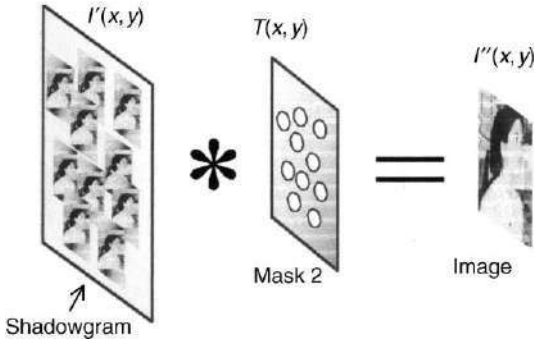

Figure 7 Synthetic aperture: reading.

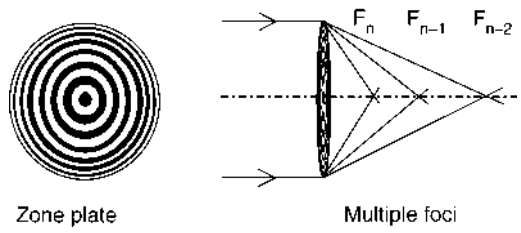

Figure 8 Properties of FZP.

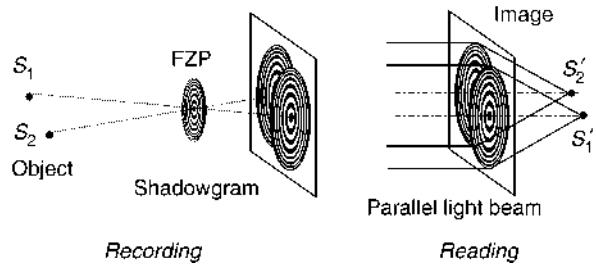

Figure 9 Recording and analogical reading with a FZP.

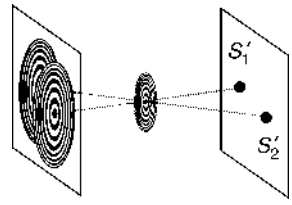

Figure 10 Optical decorrelation.

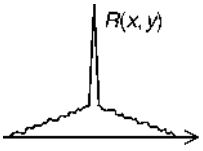

Figure 11 System point spread function with a Fresnel zone plate.

The FZP scale factor only must be adjusted to compensate for the magnification ratio generated by the conical projection at the recording step.

In either digital or analogical deconvolution, the image reconstructed is accurate, with good sharpness, if the number of zones is sufficient. Ceglio *et al.* obtained relevant images of 8 μm targets with FZP having 100 zones. Nevertheless, the multiple foci of FZP induces decreasing of the S/N ratio which is an important limitation. Besides, we pointed out above that the larger the object, the more the decrease in the S/N ratio.

Random and uniformly redundant arrays

Dicke and Ables substituted a random array to the FZP. The mask is made of randomly distributed holes and the expected advantage is doubled. First, one can reduce the transparent area of the mask by less than 50% in order to increase the S/N ratio when the object is wide. Second, the autocorrelation function of such an array is theoretically a Dirac distribution when its width is infinite. Thus, one can hope to narrow the system point response. However, when the width is finite, the pyramidal ground of the point response mentioned above is present and the advantage is not clearly demonstrated. In order to correct this effect, Chris Brown used a 'mismatched' function, for the postprocessing deconvolution. This function is calculated, first by achieving the complement of the recording mask function, then subtracting from it. Thus, the 1 transparencies result in $+1$ while the 0 transparencies are changed by -1. Brown shows images digitally reconstructed of X-ray sources, 30 sec. of arc wide. Fenimore and Cannon conceived a uniformly redundant array. The center random motif of the mask is half replicated on each side with only the center area being used for the recording step. The deconvolution is calculated by taking into account the whole area in a mismatched way. The result is an exact delta function. The authors show images of 200 μm-wide microspheres with a similar resolution but a signal 7100 times stronger than with a comparable pinhole camera.

Theta Rotating Interferometer

The theta shearing interferometer enables us to record complete data from the Fourier transform of a spatially incoherent object (Figure 12). The source can be any natural one, but monochromatic. This is easily practicable using a good filter or monochromator.

The device is based on a Michelson interferometer, the plane mirrors being replaced by two roof prisms. One of the roof prisms is fixed so the other can rotate around the axis of the interferometer. As a result, we have two images of the object. Let θ be the angle between the two edges of the prisms. The angle between the images is 2θ. Since the original object is noncoherent, and two homologous points being mutually coherent, the light they emit interferes, resulting in fringes.

The intensity, space frequency, and direction of the fringes characterize the intensity and the position of the point source. One shows that the farfield visibility distribution of fringes $V(\delta\vec{n})$ and the object distribution $O(\vec{r})$ are related by Fourier transforms:

$$\left\{\begin{aligned} V(\delta\vec{n}) &= \int O(\vec{r})\exp\left[\mathrm{i}2\pi\frac{\delta\vec{n}\cdot\vec{r}}{2\sin\theta}\right]\frac{\mathrm{d}\vec{r}}{2\sin\theta}, \\ O(\vec{r}) &= \int V(\delta\vec{n})\exp\left[-\mathrm{i}2\pi\frac{\delta\vec{n}\cdot\vec{r}}{2\sin\theta}\right]\mathrm{d}\delta\vec{n} \end{aligned}\right\} \quad [8]$$

Then, one can see that the visibility of the interference pattern is an image of the degree of coherence $\gamma(\delta\vec{n})$, with a scale factor $1/2\sin\theta$.

Such a pattern recorded on a photographic plate is a Fourier hologram of the noncoherent object. An image can be reconstructed by diffraction at infinity or near a focus (see Figure 13, examples performed in 1972 by the author). The technique was successfully applied to the determination of the modulation transfer function of optical systems. C Roddier and F Roddier applied the method to astronomy.

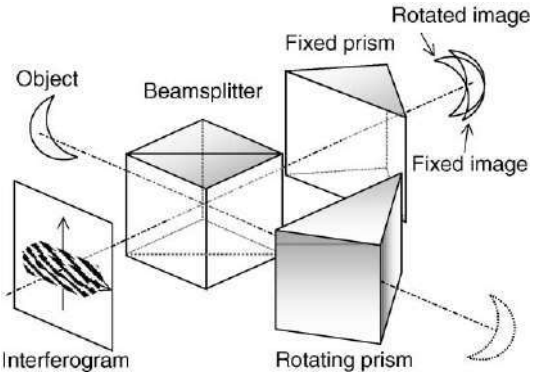

Figure 12 Theta shearing interferometer.

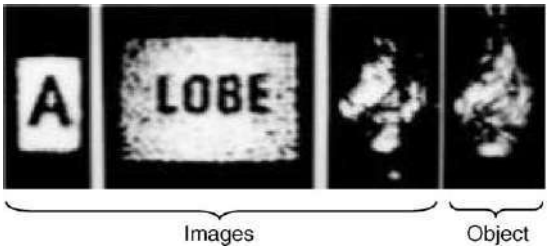

Figure 13 Incoherent light holograms.

For instance, they examined Alpha Orionis and discovered the evolution of its dust envelope and the possible existence of a companion.

High Resolution Imaging in Astronomy

In astronomy, the limit of resolution depends chiefly on two factors. The first is a technical one. Adaptive optics and very large telescopes (VLT) brought spectacular progress to clarity and resolution of telescopes. The 8.2 m VLT of Kueyen set up in 1989, shows images easily resolved of HIC 69495 Centauri, separated by 0.12 sec. of arc. Nevertheless, working of large mirrors for telescopes reaches its limits. Therefore, their clarity remains too weak. On the other hand, atmospheric turbulences make the refractive index of air unstable and the position of the images in the focal plane of telescopes unsettled. Thus, the point response is a granular spot named 'speckle'. Short exposure times relieve that inconvenience, but they prohibit the observation of weak stars. In 1970, Labeyrie proposed a method of speckle interferometry consisting of a statistical analysis of a set of speckles. The result is a display of a diffraction-limited image. He resolved many tens of single and double stars. Only Hubble, above the atmosphere, is able to do it by direct imaging. However, the primary mirror of Hubble being only 2.40 m wide, has clarity much lower. Stellar interferometry is a very promising alternative solution.

Principles of Stellar Interferometry

Stellar interferometry is based on eqn [3]. The light emitted by double or wide stars is noncoherent, so that one can access an image of them trough 2D measurements of the degree of coherence of light. We showed above that one could achieve this by using interference devices such as Young holes. Eqn [3] shows that if we measure the visibility of Young fringes at increasing distances $\delta\rho$ between the pinholes, the distance $\Delta\rho$ from which the visibility is zero gives the angular width of the star, $\Delta\alpha$, under the formula:

$$\Delta\rho \cdot \Delta\alpha \sim \lambda \qquad [9]$$

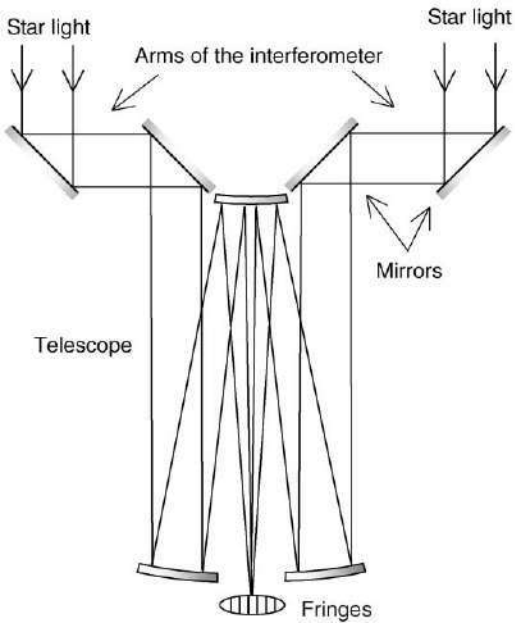

Figure 14 Stellar interferometer of Michelson.

where $\Delta\alpha = \Delta r/d$. In the case of a double star, the visibility varies periodically versus $\delta\rho$ when the pinholes are aligned along the axis of two stars. Let $\Delta\alpha$ be the angular distance between the stars: the smallest distance $\Delta\rho$ for which the visibility is zero is given by:

$$\Delta\rho \cdot \Delta\alpha = \lambda/2 \qquad [10]$$

The larger the uttermost distance between the holes, the more the limit of resolution of the processing is potentially sharp.

In order to exploit this idea, Stephan placed and rotated a mask with two holes in front of the objective of a refractive telescope. He established that the diameter of the stars he observed was much less than 0.16 sec. of arc. Then, Michelson built up, on top of a 30 cm refractive telescope, a stellar interferometer consisting of two arms catching the light at 7 m of distance from each other (Figure 14). He measured correctly the diameter of the satellites of Jupiter. Hale carried the distance of the base up to 15 m in 1920.

The Two Telescopes Interferometer of Labeyrie

In 1975, Labeyrie improved the idea by setting up two separate 25 cm diameter Cassegrain-coudé telescopes on a 12 m baseline, oriented north–south (Figure 15). The setting and mechanical stability of the device are extremely sensitive. The tolerances do not exceed any μm. A mobile table near the common focal plane of the two telescopes compensates for the optical path differences due to stars tracking. In spite

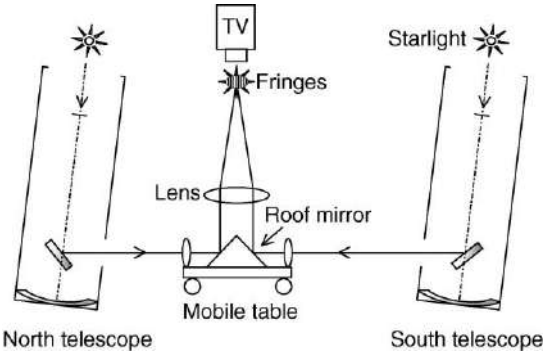

Figure 15 The two telescopes interferometer of Labeyrie.

of the acute difficulties, he succeeded in observing fringes on Vega. He obtained a resolution of 0.01 sec. of arc versus 0.5 sec. with a single 25 cm telescope. Later, one telescope was rendered mobile bringing the base width variable between 4 and 67 m. The resolution reaches up to 0.00015 sec. of arc. Note that the rotation of Earth also enables east–west analysis with resolution of 0.005 sec., and today, the 25 cm telescopes are replaced by 1.5 m ones.

The Very Large Telescope Interferometer (VLTI)

The two telescopes interferometer of Labeyrie is a first step toward a very ambitious project: the telescope at synthetic aperture and very large base. Proofs of reliability of such a principle was already given by the very large array (VLA) experimented in radio astronomy. Labeyrie with Lena, and the team of the ESO (Ecole Supérieure d'Optique, Paris), imagined a very large telescope interferometer (VLTI) made of 6 telescopes of 2 m, in a 2D structure on a base of 100 m. The European VLTI was built at Cerro Paranal. It is made of four 8.2 m telescopes corrected in real time by adaptive optics (NTT technology) and three 1.8 m telescopes. The fourth 8.2 m telescope was put in service in 2001. Some days, each of the 8.2 m telescopes used separately gives a resolution close to that of Hubble. However, the clarity is much better because of the larger primary mirror diameter (8.2 m against 2.4 m). In May 2003, the team of J-G Cuby observed a galaxy the farthest away ever seen, only 900 millions years old since the Big Bang. Obviously, a drastic gain of resolution (by a factor 10) is expected from the interference configuration.

See also

Coherent Lightwave Systems. **Diffraction:** Fresnel Diffraction. **Imaging:** Adaptive Optics; Interferometric imaging. **Instrumentation:** Telescopes. **Interferometry:** Overview.

Further Reading

Labeyrie A (1976) High resolution techniques in optical astronomy. In: Wolf E (ed.) Progress in Optics, vol. XIV, pp. 49–85. New York: North-Holland.

Mertz L (1965) *Transformations in Optics*. Amsterdam, Oxford: John Wiley & Sons, Inc.

Photon Density Wave Imaging

V Toronov, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Photon density wave imaging (PDWI) employs diffused light, usually in the near-infrared band, for imaging inhomogeneous structures in a highly scattering medium. The other names for PDWI are diffuse optical tomography and photon migration imaging. The term photon density wave (PDW) refers to the wave of light density spreading in the highly scattering medium from the source whose power changes in time. Therefore, rigorously speaking, photon density wave imaging uses intensity-modulated light sources, such as pulsed lasers, or sources whose power is harmonically modulated. However, many applications employ steady-state light sources. Most applications of PDWI are in the field of imaging of biological tissues, including medicine and cognitive neuroscience.

As in the other imaging methods, such as x-ray, computed tomography (CT) or ultrasound imaging, the problem of PDWI consists in the reconstruction of the internal structure of the medium from the measurements made outside the medium (on its surface). In the case of PDWI the internal structure of the medium is represented by the spatial distribution of its optical properties. The reconstruction procedure is essentially an iterative mathematical algorithm, which includes the modeling of the light propagation inside the medium assuming a certain distribution of the optical properties, and the

correction of this distribution to match the model prediction with the actual measurement. The modeling part of the imaging problem is usually referred to as the forward problem, and the correction procedure is known as an inverse problem.

Forward Problem: Photon Density Waves

Unlike optically transparent media, in a highly scattering medium the coherence of light from the coherent source very rapidly decays with distance. This happens due to the interference of many randomly and multiply scattered light waves at every point of the medium that is distanced from the source much farther than the photon mean free path l, which is the mean distance at which a light wave undergoes a single scattering event. For biological tissues, where light scattering occurs on cell membranes and other heterogeneities of the index of refraction, l is typically about 1 mm. Although high scattering eliminates the coherence of the electromagnetic waves at the short time intervals corresponding to oscillation at optical frequencies ($\sim 10^{-15}$ s), in the case of an intensity-modulated light source, the macroscopic wave of the density of light energy, called 'photon density wave', forms in the highly scattering medium. The properties of PDW depend on the type of the source modulation and on the local optical properties of the medium, which are the scattering coefficient μ_s [cm^{-1}] and the absorption coefficient μ_a [cm^{-1}].

Because of the complexity and incoherence of the macroscopic electromagnetic field in highly scattering media, i.e. media with $\mu_a \ll \mu_s$, the fundamental Maxwell equations of electrodynamics do not provide a convenient approach to the solution of the problem of light propagation in such media at distances much larger than l. Therefore, statistical approaches have been developed, such as the photon transport model and its diffusion approximation. In the photon transport model the light is characterized by the radiance $L(\mathbf{r}, \hat{\Omega}, t)$ [W cm^{-2} steradian^{-1}] (where the steradian is the unit solid angle), which is the power of radiation near the point $\mathbf{r}$ propagating within a small solid angle around the direction $\hat{\Omega}$ at time t. The photon transport model accounts for the possible nonisotropy of the radiation flux. For the given spatio-temporal distribution of the optical properties, μ_a and μ_s, and for the given spatial and angular distribution of light source power $S(\mathbf{r}, \hat{\Omega}, t)$ [W cm^{-3} steradian^{-1}] the radiance $L(\mathbf{r}, \hat{\Omega}, t)$ can be found from the linear radiation transfer equation, also called the photon transport equation, or simply the transport equation (TE):

$$\frac{1}{v}\frac{\partial L(\mathbf{r}, \hat{\Omega}, t)}{\partial t} + \nabla \cdot L(\mathbf{r}, \hat{\Omega}, t)\hat{\Omega} + (\mu_s + \mu_a)L(\mathbf{r}, \hat{\Omega}, t) = \mu_s \int L(\mathbf{r}, \hat{\Omega}', t) f(\hat{\Omega}, \hat{\Omega}') \mathrm{d}\hat{\Omega}' + S(\mathbf{r}, \hat{\Omega}, t) \qquad [1]$$

where $f(\hat{\Omega}, \hat{\Omega}')$ is the phase function, which represents the portion of light energy scattered from the direction $\hat{\Omega}$ to the direction $\hat{\Omega}'$, and v is the speed of light in the medium. TE is equivalent to the Boltzmann transport equation describing the transport of particles undergoing scattering.

Equation [1] is difficult to solve even numerically. Therefore, a hierarchy of approximations to the PTM was developed. This hierarchy, known as the P_N approximation, is based on the expansion of the radiance in a series of spherical harmonics $Y_{nm}(\hat{\Omega})$ truncated at $n = N$. In the P_1 approximation the coefficients of the series are proportional to the radiation fluence

$$\Phi(\mathbf{r}, t) = \int L(\mathbf{r}, \hat{\Omega}, t) \mathrm{d}\hat{\Omega}, \qquad [2]$$

and the flux

$$J(\mathbf{r}, t) = \int L(\mathbf{r}, \hat{\Omega}, t)\hat{\Omega} \mathrm{d}\hat{\Omega}, \qquad [3]$$

both having units of [W cm^{-2}]. By substituting the P_1 series for $L(\mathbf{r}, \hat{\Omega}, t)$ and $S(\mathbf{r}, \hat{\Omega}, t)$ into eqn [1] and assuming isotropic light sources and sufficiently slow change of source intensity, one can obtain the equation that includes only $\Phi(\mathbf{r}, t)$ as the unknown. Usually, this equation

$$v\mu_a U(\mathbf{r}, t) + \frac{\partial U(\mathbf{r}, t)}{\partial t} - D\Delta U(\mathbf{r}, t) = Q(\mathbf{r}, t) \qquad [4]$$

is written in terms of the value $U(\mathbf{r}, t) = \frac{1}{v}\Phi(\mathbf{r}, t)$ [J cm^{-3}], which has units of energy density, and is called the photon density. In eqn [4] $D = v/(3\mu_a + 3\mu_s')$ [cm^2/s] is the light diffusion coefficient, and $\mu_s' = \mu_s(1 - g_1)$ [cm^{-1}] is the reduced scattering coefficient; g_1 is the average cosine of the scattering angle. The source term $Q(\mathbf{r}, t)$ [J cm^{-3} s^{-1}] in the right-hand side of the equation [4] describes the spatial distribution of light sources (in the isotropic approximation, i.e., neglecting the dipole and higher momenta of $S(\mathbf{r}, \hat{\Omega}, t)$) and the temporal modulation of source intensity.

Equation [4] is mathematically equivalent to the equation describing the density of the medium consisting of the particles undergoing linear diffusion. Therefore, it is called the diffusion equation (DE), and the P_1 approximation is also known as the diffusion

approximation. The diffusion approximation is valid under the following three conditions: (a) the light sources are isotropic or the distance between the source and the detector is much larger than l; (b) characteristic source modulation frequencies ν are low enough that $\nu\mu_s'/(2\pi\nu) \gg 1$; and (c) $\mu_a \ll \mu_s'$. In the biomedical applications of PDWI the condition (a) is fulfilled when the source–detector distance is larger than 1 cm. The condition (b) is valid when the characteristic modulation frequencies of the light source are less than 1 GHz. The condition (c) is fulfilled in most biological tissues, for which the scattering is highly isotropic, i.e. $g_1 \ll 1$. Thus, the forward problem of PDWI, i.e., the modeling of light propagation, can be solved using the diffusion approximation in a large variety of situations.

PDWs described by equation [4] are the scalar damped traveling waves, characterized by the coherent changes of light energy (photon density) in different locations on time intervals from microseconds to nanoseconds. Usually pulsed or harmonically modulated light sources are used to generate PDWs. The properties of the PDWs from a harmonic source are fundamental, because the PDW from the source of any type of modulation can be considered as a superposition of harmonic waves from sources having different frequencies using the Fourier transform.

In order to get an idea of the basic properties of PDW, one can consider the analytical solution of [4] for the case of a harmonically modulated point-like source placed in an infinite homogeneous scattering medium. This solution is

$$U(r,t) = U_{dc}(r) + U_{ac}(r)\exp[-i\omega t + i\phi(r)] \qquad [5]$$

where U_{ac}, U_{dc}, and ϕ are the ac, dc and phase, respectively, of the wave measured at the distance r from the source, and $\omega = 2\pi\nu$. U_{ac}, U_{dc}, and ϕ depend on the medium optical properties, source intensity Q_0, modulation amplitude A_0, and initial phase ψ_0 as

$$U_{ac}(r) = \frac{Q_0 A_0}{4\pi D} \times \frac{\exp\left[-r\left(\frac{\nu\mu_a}{2D}\right)^{1/2}((1+x^2)^{1/2}+1)^{1/2}\right]}{r} \qquad [6a]$$

$$U_{dc}(r) = \frac{Q_0}{4\pi D}\,\frac{\exp\left[-r\left(\frac{\nu\mu_a}{D}\right)^{1/2}\right]}{r} \qquad [6b]$$

$$\phi(r) = r\left(\frac{\nu\mu_a}{2D}\right)^{1/2}\sqrt{(1+x^2)^{1/2}-1} \qquad [6c]$$

where $x = \omega/(\nu\mu_a)$. Equations [5]–[6] describe a spherical wave whose wavelength at $\mu_a = 0.1\ \mathrm{cm}^{-1}$ and $\mu_s' = 8\ \mathrm{cm}^{-1}$ (typical for biological tissues) and at modulation frequency $\nu = 100$ MHz, is close to 35 cm, so that the phase of the wave changes approximately by 10° per cm. In the same time, the wave decays exponentially at a distance of about 3 cm, so that PDWs are strongly damped. A similar analytical solution in the form of a spherical wave can be obtained for the case of a semi-infinite homogeneous medium and point source placed on the surface of the medium (see Figure 1). These analytical solutions are employed to measure μ_a and μ_s' of the homogeneous medium. The method is based on fitting measured ac, dc, and phase of the PDW as functions of source–detector distance by the analytical solution.

Inhomogeneities in the optical properties of a turbid medium cause distortion of the PDW from sphericity (see Figure 1). PDWs in homogeneous media of limited size or with strong surface curvature also usually have complex structure. For a limited number of configurations one can obtain solutions to the DE using analytical methods of mathematical physics, such as the method of Green's functions. However, in practice the complexity of the inhomogeneities and surfaces usually requires numerical solution of the DE. This can be achieved using numerical methods for the solution of partial differential equations, such as the finite element method and the finite difference method. The finite difference method can also be used to solve the TE. This may be necessary in situations when the DE is not valid, for example to accurately account for the influence of the cerebrospinal fluid (which has relatively low scattering) on light transport in the human skull. However, the solution of the TE requires formidable computational resources, which increase proportionally to the volume of the medium. Therefore, hybrid DE-TE methods were developed, in

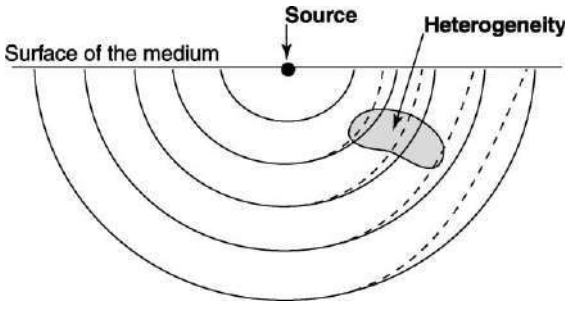

Figure 1 Equiphase surfaces of the PDW induced by a point source in a semi-infinite homogeneous medium (solid lines) and in a medium including heterogeneity (dashed lines).

which DE is used to describe light propagation in most of the medium except small low-scattering areas described by the TE.

One should note that although the term 'photon' suggests quantum properties of light, no consistent application of quantum theory to problems related to PDWI has been developed so far. Also, it should be noted that the notion of a photon as a particle of light having certain energy, momentum, and localization at the same time is not completely consistent with quantum electrodynamics. However, since the theoretical models of light transport in a scattering medium (such as TE and DE) are mathematically equivalent to the models describing systems of randomly moving particles, the notion of a photon as a 'particle of light' may be accepted as adequate and is widely used in the limits of PDWI.

Inverse Problem and Imaging Algorithm

Mathematically the forward problem can be expressed in the form of a general nonlinear operator F:

$$G = F[\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}] \qquad [7]$$

where G is the set of measured values (ac, dc, and phase), $\mathbf{r}$ is the coordinate vector of a medium element (voxel), and $\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}$ denotes the set of tissue optical properties corresponding to whole collection of voxels. The problem of finding the spatial distribution of medium optical properties can be formalized as

$$\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} = F^{-1}[G] \qquad [8]$$

Neither the transport model nor the diffusion model allow inversion in a closed analytical form for a general boundary configuration and spatial distribution of optical properties. The basic reason for the complexity of the inverse problem is in the complexity of the light paths in the highly scattering medium. This situation is opposite to that in CT, where the x-ray radiation propagates through the medium along straight lines almost without scattering, which significantly simplifies the forward problem and allows its analytical inversion. From the mathematical point of view, the complexity of the inverse problem of PDWI is due to its nonlinearity and ill-posedness. Only approximate methods of inversion have been developed so far.

The simplest method is called backprojection by analogy with the method of the same name used in CT. The method attempts to reconstruct the actual spatial distribution of the optical properties μ_a and/or μ_s' from the 'measured' bulk values μ_a^m and $\mu_s'^m$ obtained at a variety of locations assuming homogeneity of the medium. The measured values are considered as a result of the volumetric 'projection'

$$\{\mu_a^m(\mathbf{r}), \mu_s'^m(\mathbf{r})\} = \int \{\mu_a(\mathbf{r}'), \mu_s'(\mathbf{r}')\} h(\mathbf{r} - \mathbf{r}') d^3\mathbf{r}' \qquad [9]$$

of the actual distribution, where $\mathbf{r}$ and $\mathbf{r}'$ are the locations of the measurement and reconstruction, respectively. A weight function $h(\mathbf{r})$ accounts for the contribution of the particular region to a measured value and should be derived from the forward model. Different versions of the backprojection method use different weight functions $h(\mathbf{r})$. In the backprojection method the values of μ_a and/or μ_s' are obtained by inverting the convolution [9]. In practice the convolution integral is usually estimated by a sum over a limited number of voxels for a limited number of measurements. Therefore, the inversion of the convolution reduces to the solution of the set of algebraic equations corresponding to [9]. The advantage of the method is high speed, which allows real-time imaging of the dynamic processes, for example, functional hemodynamic changes in the brain. However, the backprojection method provides a phenomenological rather than a self-consistent inversion of the forward problem, and its quantitative accuracy is low.

A more consistent method is based on the perturbation approach to the solution of the inverse problem [8]. If we have an estimate $\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}_e$ that is close to the ideal solution, then $G_e = F[\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}_e]$ (see eqn [7]) is close to S. Then the difference $G - G_e$ can be written as a Taylor series:

$$G - G_e = J[\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}_e]\{\Delta\mu_a(\mathbf{r}), \Delta\mu_s'(\mathbf{r})\} + \ldots \qquad [10]$$

where $J[\ldots]$ is the Jacobian, and $\{\Delta\mu_a(\mathbf{r}), \Delta\mu_s'(\mathbf{r})\} = \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} - \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}_e$. Neglecting terms after the first one in [10], the problem of finding the correction $\{\Delta\mu_a(\mathbf{r}), \Delta\mu_s'(\mathbf{r})\}$ to the estimate $\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}_e$ reduces to the calculation of the Jacobian and to the solution of a linear algebraic equation.

This approach provides a basis for the PDWI algorithm, which is the following. First one makes an initial guess at the optical properties of the medium. Then one calculates the ac, dc, and phase of the PDW using a model of light propagation through the tissue and the geometry of the source–detector configuration. The calculated values are compared with the measured ones. If the error is above a set tolerance level, one calculates corrections to the current values of μ_a and μ_s' by solving eqn [10]. The algorithm is completed when the required error level is achieved. Thus, the PDWI algorithm iteratively applies both

the forward and inverse problems. Since at each of many iterations one should solve the forward problem, which in the simplest case is a partial differential equation, the algorithm typically requires considerable computational resources.

One should note that the success of the method is largely dependent on the closeness of the initial guess to the actual values of the optical properties. In many studies the 'first guess' of the tissue optical properties was obtained by measuring the bulk absorption and reduced scattering coefficients. However, because of the ill-posedness of the inverse problem, homogeneous initial conditions did not always lead to the correct reconstruction. A more productive approach to the problem of the initial conditions is to assign approximate initial values of the optical properties known for example from *in vitro* measurements to different structures revealed by nonoptical imaging techniques, such as magnetic resonance imaging (MRI) or CT.

Applications

PDWI was proposed and developed as a noninvasive biomedical imaging technique. Although the spatial resolution of PDWI cannot compete with that of CT or MRI, PDWI has the advantages of low cost, high temporal resolution, and high biochemical specificity. The latter is based on the optical spectroscopic approach. Biological tissues are relatively transparent to light in the near-infrared band between 700 and 1000 nm (a near-infrared window). The major biological chromophores in this window are oxy- and deoxyhemoglobin, and water. One can separate contributions of these chromophores into light absorption by employing light sources at different wavelengths. Multi-wavelength PDWI can also be used to measure the redox state of the cytochrome oxidase, which is the terminal electron acceptor of the mitochondrial electron transport chain and responsible for 90% of cellular oxygen consumption.

One of the hot topics in cognitive neuroscience is the hemodynamic mechanism of the oxygen supply to neurons and removal of products of cerebral metabolism. High temporal resolution and biochemical specificity make PDWI a unique noninvasive tool for studies of functional cerebral hemodynamics, although light can penetrate no deeper than the very surface of the adult brain. Figure 2 displays a typical arrangement of light sources and detectors used in PDWI of functional cerebral hemodynamics, and the images of the activated area in the brain.

Because of its low cost, noninvasiveness, and compact size of the optical sensors, PDWI is also considered to be a promising tool for mammography and structural imaging of anomalies in the brain. PDWI is an especially promising tool for neonatology, because the infant tissues have low absorption.

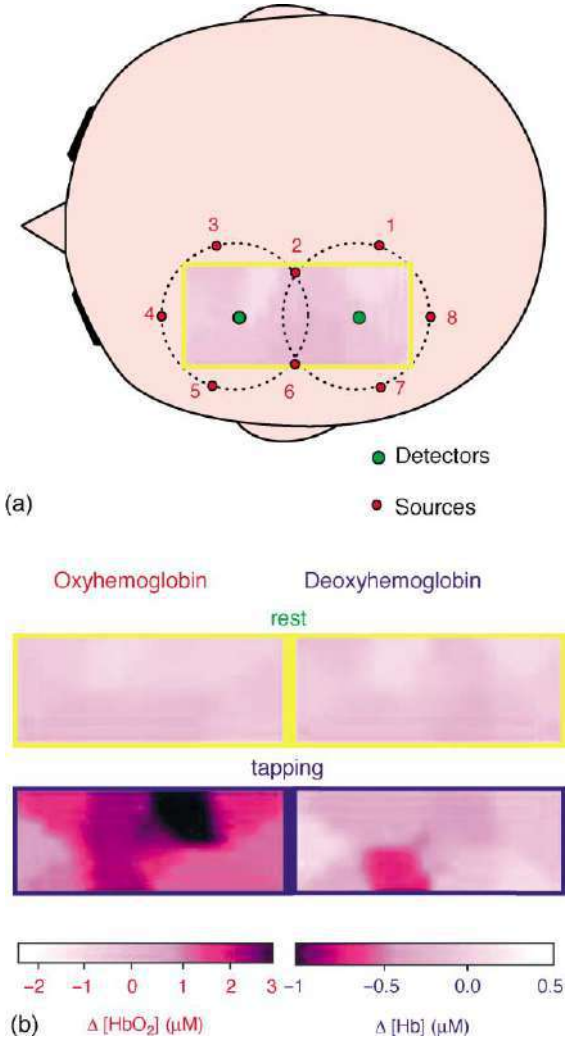

Figure 2 (a) Geometrical arrangement of the 16 sources and two detectors on the head of a human subject. Each one of the eight small red circles (numbered 1–8) represents one pair of source fibers at 758 and 830 nm. Two larger green circles represent detector fiber bundles. The rectangle inside the head outline is the $4 \times 9\ cm^2$ imaged area. (b) Maps of oxyhemoglobin ($[HbO_2]$, left panels) and deoxyhemoglobin ([Hb] right panels) concentration changes at rest (top) and at the time of maximum response during the third tapping period (bottom). The measurement protocol involved hand-tapping with the right hand (contralateral to the imaged brain area). Adapted with permission from Franceschini MA, Toronov V, Filiaci M, Gratton E, Fantini S (2000) On-line optical imaging of the human brain with 160-ms temporal resolution. *Optics Express* **6**(3), 49–57 (Figures 1 and 4).

PDWI instruments employ lasers, light-emitting diodes and gas-discharge tubes as sources of light. Among the detectors there are photomultiplier

tubes, avalanche photodiodes, and CCD cameras. Both sources and detectors can be coupled with the tissue by means of optical fibers. A number of commercially produced near-infrared instruments for PDWI are available on the market. According to the type of source modulation, there are three groups of instruments: frequency domain (harmonically modulated sources), time domain (pulsed sources), and steady state. Steady-state instruments are simpler and cheaper than frequency-domain and time-domain ones, but their capabilities are much lower. Time-domain instruments can theoretically provide a maximum of information about the tissue structure, but technically it is difficult to build a time-domain instrument with high temporal resolution and high signal-to-noise ratio. Frequency-domain instruments provide temporal resolution up to several milliseconds and have high quantitative accuracy in the determination of optical properties.

Nomenclature

Absorption coefficient, μ_a [cm^{-1}]
Alternating part of PDW, alternating current (ac)
Average cosine of the scattering angle, g_1
Circular modulation frequency, ω [s^{-1}]
Computed tomography (CT)
Difference between the estimate and actual optical properties, $\{\Delta\mu_a(\mathbf{r}), \Delta\mu_s'(\mathbf{r})\} = \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\} - \{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}_e$ [cm^{-1}]
Diffusion coefficient, $D = v/(3\mu_a + 3\mu_s')$ [cm^2/s]
Diffusion equation (DE)
Distribution of light source power $S(\mathbf{r}, \hat{\Omega}, t)$ [W cm^{-3} steradian^{-1}]
Divergence operator, ∇ [cm^{-1}]
Estimate of optical properties, $\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}_e$ [cm^{-1}]
General nonlinear operator, F
Jacobian, J
Laplace operator, Δ [cm^{-2}]
Magnetic resonance imaging (MRI)
Measured bulk optical properties (absorption and reduced scattering coefficients) $\{\mu_a^m \mu_s'^m\}$ [cm^{-1}]
Modulation frequency, ν [s^{-1}]
Non-alternating part of PDW, direct current (dc)
Optical properties of the medium (absorption and reduced scattering coefficients) as functions of spatial coordinate, $\{\mu_a(\mathbf{r}), \mu_s'(\mathbf{r})\}$ [cm^{-1}]
Phase as a function of source–detector distance $\phi(r)$
Phase function, $f(\hat{\Omega}, \hat{\Omega}')$
Photon density, $U(\mathbf{r}, t)$ [J cm^{-3}]
Photon density wave imaging (PDWI)
Photon density wave (PDW)
Photon mean free path, l [cm]
Radiance, $L(\mathbf{r}, \hat{\Omega}, t)$ [W cm^{-2} steradian^{-1}]
Radiation fluence, $\Phi(\mathbf{r}, t)$ [W cm^{-2}]
Radiation flux, $J(\mathbf{r}, t)$ [W cm^{-2}]
Reduced scattering coefficient, μ_s' [cm^{-1}]
Scattering coefficient, μ_s [cm^{-1}]
Set of measured values (ac, dc, and phase), G
Set of predicted values (ac, dc, and phase), G_e
Source term $Q(r, t)$ [J cm^{-3} s^{-1}]
Source–detector distance, r [cm]
Spatial coordinate, $\mathbf{r}$ [cm]
Speed of light in medium, v [cm/s]
Spherical harmonics $Y_{nm}(\hat{\Omega})$
Transport equation (TE)
Weight function, $h(\mathbf{r})$
$x = \omega/(v\mu_a)$

See also

Coherence: Coherence and Imaging. **Fiber and Guided Wave Optics:** Light Propagation. **Imaging:** Imaging Through Scattering Media. **Optical Materials:** Heterogeneous Materials. **Scattering:** Scattering Theory.

Further Reading

Arridge SR (1999) Optical tomography in medical imaging. *Inverse Problems* 15: R41–R93.

Boas DA (1996) *Diffuse Photon Probes of Structural and Dynamical Properties of Turbid Media: Theory and Biomedical Applications*. PhD Thesis, University of Pennsylvania.

National Research Council (US), Institute of Medicine (US) (1996) *Mathematics and Physics of Emerging Biomedical Imaging*. Washington, DC: National Academy Press.

O'Leary MA (1996) *Imaging with Diffuse Photon Density Waves*. PhD Thesis, University of Pennsylvania.

Special Issue (1997) Near-infrared spectroscopy and imaging of living systems. *Philosophical Transactions of the Royal Society B* 352(1354).

Tuchin V (2002) *Handbook of Optical Biomedical Diagnostics*. Bellingham, WA: SPIE Press.

Tuchin VV, Luo Q and Ulyanov SS (2001) *Imaging of Tissue Structure and Function*. Proceedings of SPIE 4427. Bellingham, WA: SPIE Press.

Villringer A and Chance B (1997) Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neuroscience* 20: 435–442.

Villringer A and Dirnagl U (1993) *Optical Imaging of Brain Function and Metabolism*. New York: Plenum Press.

Villringer A and Dirnagl U (1997) *Optical Imaging of Brain Function and Metabolism 2: Physiological Basis and Comparison to Other Functional Neuroimaging Methods*. New York: Plenum Press.

Three-Dimensional Field Transformations

R Piestun, University of Colorado, Boulder, CO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Conventional optical imaging is understood as an isomorphic mapping from an object space onto an image space. The object space is composed of scattering or emitting objects. The scattering objects can be opaque, transparent, or partially absorbing, while emitting objects can be as diverse as stars in astronomical imaging or fluorescent molecules in microscopic imaging. The image space is an optical field representation of the object.

In most classical applications of optical imaging the goal is to optimize the mapping between one plane in the object space and one plane in the image space. Modern optical systems, though, are aimed at acquiring three-dimensional information regarding the object. The image is thus a three-dimensional field representation of the object.

While many imaging systems are aimed at obtaining visually appealing images, many modern imaging systems register the field information in some type of sensitive device such as film or CCD cameras for later processing. This processing is typically performed electronically by a computer and the final image is ultimately a mathematical digital representation of the object. This type of imaging is called computational imaging. An example of computational imaging is optical tomography, which involves the registration of a set of images that do not necessarily resemble the object, and off-line digital processing to obtain a three-dimensional representation of the object.

In this article we will focus on imaging in the conventional sense but will also focus on the three-dimensional aspects of the image.

Geometrical Optics Transformations

In many optical problems it is appropriate to consider the wavelength of the electromagnetic wave as infinitely small. This approximation is the origin of the field of geometrical optics in which the propagation of light is formulated in terms of geometrical laws. Accordingly, the energy is transported along light rays and their trajectories are determined by certain differential or integral equations that can be directly derived from Maxwell's equations. One of these formulations uses Fermat's principle that states the principle of the shortest optical path. The optical path between two points P_1 and P_2 is defined as:

$$OP = \int_{P_1}^{P_2} n(P)\mathrm{d}s \qquad [1]$$

where n is the refractive index of the medium. Fermat's principle states that the optical path of a real ray is always a local minimum.

In ideal optical imaging systems, every point P_o of the so-called object space will generate a sharp image point P_i, that is a point where infinite rays emerging from P_o pass through P_i. Perfect imaging occurs when every curve defined by object points is geometrically similar to its image. Perfect imaging can occur between three-dimensional spaces or between surfaces.

An interesting result of geometrical optics is Maxwell's theorem. This theorem states that if an optical instrument produces a sharp image of a three-dimensional domain, the optical length of any curve in the object is equal to the optical length of its image. More explicitly, if the object space has index n_o and the image space index n_i, the theorem states that:

$$\int_{C_\mathrm{o}} n_\mathrm{o}\,\mathrm{d}s_\mathrm{o} = \int_{C_\mathrm{i}} n_\mathrm{i}\,\mathrm{d}s_\mathrm{i} \qquad [2]$$

where C_i is the image of the object curve C_o.

An important consequence of Maxwell's theorem is a theorem by Carathéodory which states that any imaging transformation has to be a projective transformation, an inversion, or a combination of them. Another consequence of this theorem is that if the two media are homogeneous, perfect imaging can only occur if the magnification is equal to the ratio of the refractive indices in the object and image space. If the media have the same index, the imaging is trivial with unit magnification. The simplest example of a perfect imaging instrument is a planar mirror. Therefore, if we want to achieve nontrivial imaging between homogeneous spaces with equal index, we must give up either the perfect imaging (similarity) or the sharpness of the image.

The geometrical optics approximation is very useful in the design of optical imaging systems. However, it fails to provide a complete description

of wave optical effects, notably diffraction. Therefore, the following sections provide a description of field transformations based on wave optics.

Systems Approach to 3D Imaging

In this article we deal with linear imaging systems. Hence the relation between three-dimensional (3D) input (object) signals and 3D output (image) signals can be completely specified by the 3D impulse response or point spread function (PSF) of the system $h(x, y, z; x'y', z')$. This function specifies the response of the system at image space coordinates (x, y, z) to a δ-function input (point source) at object coordinates (x', y', z'). The optical system object and image signals are related by the superposition integral:

$$\mathrm{i}(x, y, z) = \iiint_{\Omega} \mathrm{o}(x', y', z') h(x, y, z; x'y', z') \mathrm{d}x' \, \mathrm{d}y' \, \mathrm{d}z' \qquad [3]$$

where Ω is the object domain. It is clear that for a generic linear imaging system the image signal is completely specified by the response to input impulses located at all possible points of the image domain.

If the system is shift invariant, the PSF due to a point source located anywhere in the object domain will be a linearly shifted version of the response due to a reference point source. This approximation is in general not valid and the system's PSF is a function of the position of the point source in the object domain. Shift invariance is traditionally divided into radial and axial shift invariance. Radial shift variance is the change in the image of a point source as a function of its radial distance from the optical axis. Radial shift variance is caused by lens aberrations. Axial shift variance is the change in the image of a point source as a function of its axial position, z', in the object space. Axial shift variance is inherent to any rotationally symmetric lens system because they can only have one pair of conjugate surfaces, i.e., object–image corresponding surfaces.

In most practical cases the image is detected at a single plane, $z = z_1$ transverse to the optical axis Oz. In those situations, we need to specify the impulse response on this plane:

$$h(x, y, z_1; x'y', z') = h_{z'}(x, y; x'y') \qquad [4]$$

The condition for lateral shift invariance is then expressed as $h_{z'}(x, y; x'y') = \tilde{h}_{z'}(x - x', y - y')$. In this case, the superposition integral is a two-dimensional (2D) convolution and it is possible to define the transfer function of the system that is the 2D Fourier transform of the impulse response $\tilde{h}_{z'}(x - x', y - y')$.

Point Spread Function for Coherent, Incoherent, and Partially Coherent Illumination

The input–output relations of an optical imaging system relate different magnitudes according to the type of illumination. If the illumination is coherent, the system is linear in the complex amplitude, while if the illumination is fully incoherent the system is linear in intensity. The 3D impulse response of the incoherent (h_{Inc}) and coherent (h_{Coh}) systems are related as follows:

$$h_{\mathrm{Inc}}(x, y, z; x'y', z') = \left| h_{\mathrm{Coh}}(x, y, z; x'y', z') \right|^2 \qquad [5]$$

Notwithstanding, many optical systems cannot be considered fully coherent or fully incoherent. In these situations the system is said partially coherent. If the optical path differences within the optical system are much shorter than the coherence length of the illumination, the light effectively has complete temporal coherence and the magnitude of interest is the mutual intensity. The mutual intensity is defined as the statistical correlation of the complex wavefunction $U(\mathbf{r}, t)$ between different points in space:

$$J(\mathbf{r}_1, \mathbf{r}_2) = \langle U^*(\mathbf{r}_1, t) U(\mathbf{r}_2, t) \rangle \qquad [6]$$

where $\mathbf{r}_i = (x_i, y_i, z_i), i = 1, 2$. The illumination is then called quasi-monochromatic. For the case of partially coherent quasi-monochromatic illumination, the system is linear in the mutual intensity. The partially coherent impulse response (h_{PC}) is also related to the coherent impulse response as follows:

$$h_{\mathrm{PC}}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}'_1, \mathbf{r}'_2) = h^*_{\mathrm{Coh}}(\mathbf{r}_1; \mathbf{r}'_1) h_{\mathrm{Coh}}(\mathbf{r}_2; \mathbf{r}'_2) \qquad [7]$$

The 3D superposition integral that relate object and image mutual intensities is now:

$$J_{\mathrm{image}}(\mathbf{r}_1, \mathbf{r}_2) = \iiint_{\Omega} h_{\mathrm{PC}}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}'_1, \mathbf{r}'_2) \times J_{\mathrm{object}}(\mathbf{r}'_1, \mathbf{r}'_2) \mathrm{d}\mathbf{r}'_1 \, \mathrm{d}\mathbf{r}'_2 \qquad [8]$$

Therefore, we conclude that knowledge of the coherent PSF is enough to characterize the optical system for illumination of various degrees of coherence.

Spatial Frequency Content of the 3D Point Spread Function

As previously stated, the coherent impulse response of the optical system completely defines its imaging characteristics. Therefore, it is justified to wonder

what the limitations are in the specification of this response. In this section we analyze the frequency domain of coherent and incoherent impulse responses.

The impulse response of the optical system will satisfy the corresponding wave equations regardless of the particulars of the optical system. In particular, assuming a homogeneous isotropic medium in the imaging domain, the coherent response $h(\mathbf{r})$ to an impulse at $\mathbf{r}'$ will satisfy the homogeneous Helmholtz equation in the image domain:

$$\nabla^2 h(\mathbf{r}) + k^2 h(\mathbf{r}) = 0 \qquad [9]$$

where $h(\mathbf{r})$ is the complex amplitude and $k = 2\pi/\lambda$, with λ the wavelength in the medium. If the field at a given origin of coordinates $z = 0$, $f(x, y)$, is known, $h(\mathbf{r})$ can be calculated using the angular spectrum representation as follows:

$$h(x, y, z) = \int\int_{-\infty}^{\infty} F(f_x, f_y) \times \exp[i2\pi(f_x x + f_y y + f_z z)]\mathrm{d}f_x\, \mathrm{d}f_y \qquad [10]$$

where $f_z = \sqrt{\lambda^{-2} - (f_x^2 + f_y^2)}$ and $F(f_x, f_y) = \mathrm{FT}\{f(x, y)\}$; FT represents a 2D Fourier transformation and f_x, f_y are the spatial frequencies along the x and y axes respectively.

Equation [10] can be written as a 3D integral as follows:

$$h(x, y, z) = \int\int_{-\infty}^{\infty} F(f_x, f_y)\cdot\delta\left[f_z - \sqrt{\lambda^{-2} - f_x^2 - f_y^2}\right] \times \exp[i2\pi(f_x x + f_y y + f_z z)]\mathrm{d}f_x\, \mathrm{d}f_y\, \mathrm{d}f_z \qquad [11]$$

where δ represents the dirac impulse. Next we represent the 3D impulse response $h(x, y, z)$ as a 3D Fourier decomposition:

$$h(x, y, z) = \int\int_{-\infty}^{\infty} H(f_x, f_y, f_z) \times \exp[i2\pi(f_x x + f_y y + f_z z)]\mathrm{d}f_x\, \mathrm{d}f_y\, \mathrm{d}f_z \qquad [12]$$

where $H(f_x, f_y, f_z) = \mathrm{FT}^{3\mathrm{D}}\{h(x, y, z)\}$. Comparing eqns [11] and [12] we can infer that:

$$H(f_x, f_y, f_z) = F(f_x, f_y)\cdot\delta\left[f_z - \sqrt{\lambda^{-2} - f_x^2 - f_y^2}\right] \qquad [13]$$

This relation implies that the domain of the Fourier transform of the impulse response is confined to a spherical shell of radius $1/\lambda$. Alternatively, $H(f_x, f_y, f_z)$ is zero everywhere, except from the surface of the sphere.

The previous derivation did not consider any limitation in the range of spatial frequencies allowed through the system. We now assume a finite aperture that will introduce an effective radial cutoff frequency $f_c < 1/\lambda$ and will further limit the spatial spectral domain of $h(x, y, z)$. This is depicted in Figure 1. Note that the transverse and longitudinal spatial frequencies are related, hence the longitudinal spatial frequencies are located within a bandpass of width $\Delta f_z = 1 - \sqrt{\lambda^{-2} - f_c^2}$.

Let us now consider the frequency content of possible incoherent PSFs. From eqn [5] we derive the 3D Fourier transform of the incoherent impulse response:

$$\begin{aligned} H_{\mathrm{Inc}}(f_x, f_y, f_z) &= \mathrm{FT}\left\{h_{\mathrm{Inc}}(x, y, z; x'y', z')\right\} \\ &= \mathrm{FT}\left\{|h_{\mathrm{Coh}}(x, y, z; x'y', z')|^2\right\} \\ &= H \otimes H(f_x, f_y, f_z) \end{aligned} \qquad [14]$$

where $\otimes$ represents a 3D correlation. The domain of H_{Inc} now becomes a volume of revolution as shown in Figure 2. Moreover, because h_{Inc} is real, H_{Inc} is conjugate symmetric, i.e., $H_{\mathrm{Inc}} = H_{\mathrm{Inc}}^*$. Now, the transverse cutoff frequency is $f_{\mathrm{c}}^{\mathrm{Inc}} = 2f_{\mathrm{c}}$ while the longitudinal cutoff frequency is $f_{\mathrm{cl}}^{\mathrm{Inc}} = 2\Delta f_z$.

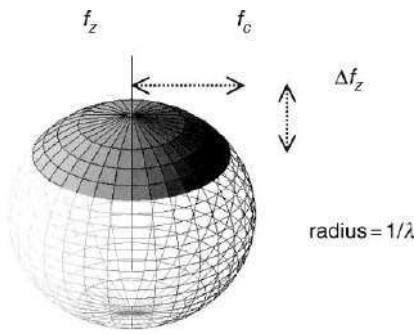

Figure 1 Three-dimensional spatial-frequency domain of the coherent point spread function.

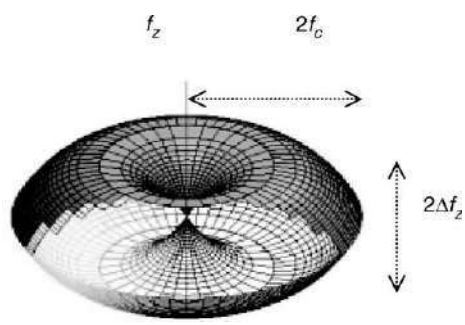

Figure 2 Three-dimensional spatial-frequency domain of the incoherent point spread function.

Diffraction Calculation of the 3D Point Spread Function

The 3D-image field created by a point source in the object space can be predicted by proper modeling of the optical system. In the absence of aberrations, the distortion of the PSF is originated in the diffraction created by the limiting aperture of the optical system. Diffraction is thus the ultimate limitation to the resolution of the optical system. Under the design conditions of the system, i.e., assuming all aberrations are corrected, the image of a point source is the diffraction pattern produced when a spherical wave converging to the (geometrical) image point is diffracted by the limiting aperture of the system. This is called a diffraction limited image.

There are many formulations to describe the 3D diffraction pattern obtained with a converging wave through an aperture. Here we review the Kirchhoff formulation. Let us consider the aperture **A** and the field at a point P_d originated from a converging spherical wave focused at point P_F (see Figure 3). We define the vectors **r** joining a generic point on the aperture P_0 and P_F, and **s** joining P_0 and P_d. The Kirchhoff formulation states that the complex amplitude $U(P_d/P_F)$ is the integral over the aperture of the complex amplitude at each point propagated to P_d by a spherical wave $\exp(iks)/s$ modified by the directional factor $\cos(\hat{\mathbf{n}}, \mathbf{r}) - \cos(\hat{\mathbf{n}}, \mathbf{s})$. $\hat{\mathbf{n}}$ is the unit vector normal to the aperture:

$$U(P_d/P_F) = \frac{-iC}{2\lambda} \int\int_A \frac{\exp[-ik(r-s)]}{rs} \times [\cos(\hat{\mathbf{n}}, \mathbf{r}) - \cos(\hat{\mathbf{n}}, \mathbf{s})] dS \qquad [15]$$

where $r = \|\mathbf{r}\|, s = \|\mathbf{s}\|$, and C is the amplitude of the spherical wave.

The directional factor varies slowly and in many situations can be approximated by a constant equal to 2 for small image-field angles. The amplitude over the aperture is also slowly varying and can be approximated by a constant, i.e., $C/r \approx C_0$. Different formulations of the Kirchhoff integral differ in the estimates of the factor $1/s$ and the phase $k(r-s)$. For example, one such formulation leads to the following expression for the 3D diffraction pattern for a focal point on the axis of an optical system with circular aperture:

$$U(P_d/P_F) = \frac{-ika^2 C_0}{z_d} \exp\left[ik\left(z_d - z_F + \frac{r_d^2}{2z_d}\right)\right] \times \int_0^1 J_0\left(\frac{ka\rho r_d}{z_d}\right) \exp\left[\frac{-ika^2\rho^2(z_d - z_F)}{2z_d z_F}\right] \rho d\rho \qquad [16]$$

where a is the radius of the aperture, r_d is the radial coordinate of P_d, and ρ is the radial coordinate of P_0.

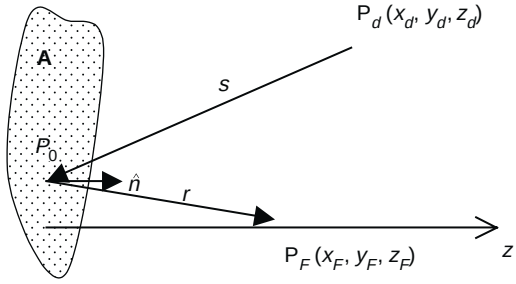

Figure 3 Diffraction of a spherical wave by a planar aperture.

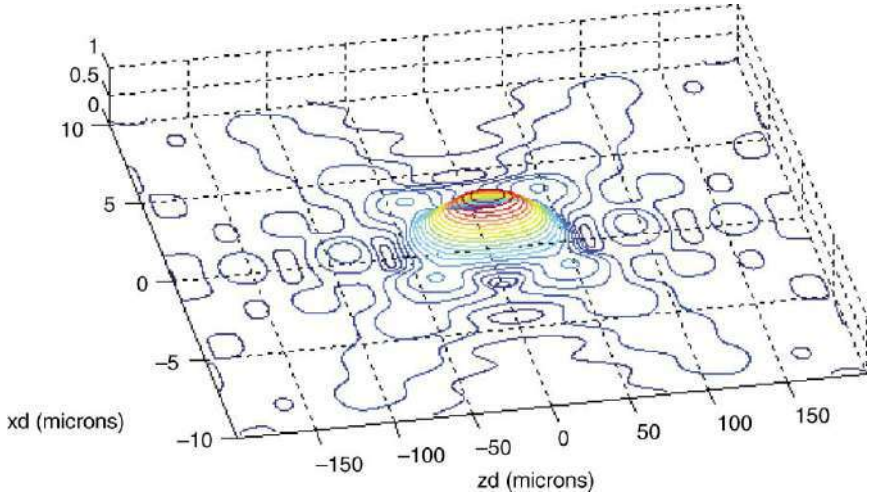

Figure 4 Lines of equal intensity in a meridional plane near the focus of a converging spherical wave diffracted at a circular aperture ($\lambda = 0.5\ \mu$, $a = 0.5$ cm, $z_F = 4$ cm).

To illustrate the use of eqn [16], we calculated the intensity around the focus of a spherical wave diffracted through a circular aperture of radius $a = 0.5$ cm, focusing at distance $z_F = 4$ cm from the screen. We assumed the wavelength to be $\lambda = 0.5\ \mu$. Figure 4 depicts the magnitude of the complex amplitude in a meridional plane near the focus of the converging spherical wave.

Models that take into account off-axis focal points have also been developed. For the most accurate calculations, it is necessary to consider the full electromagnetic nature of the optical field.

Conclusions

Classical imaging theory can be applied to the 3D imaging transformation from object to image. The systems approach to the imaging problem implies that knowledge of the 3D coherent PSF suffices to predict the 3D response of the system under illumination of varying degrees of coherence. The PSF response can be predicted using well-known formulations of diffraction.

See also

Coherent Lightwave Systems. **Fourier Optics**.

Further Reading

Born M and Wolf E (1999) *Principles of Optics*, 7th edn. New York: Cambridge University Press.

Gibson SF and Lanni F (1989) Diffraction by a circular aperture as a model for three-dimensional optical microscopy. *Journal of the Optical Society of America A* 6: 1357–1367.

Goodman JW (1996) *Introduction to Fourier Optics*. New York: McGraw-Hill.

Li Y and Wolf E (1984) Three-dimensional intensity distributions near the focus in systems of different Fresnel numbers. *Journal of the Optical Society of America A* 1: 801–808.

Piestun R and Shamir J (2002) Synthesis of 3D light fields and applications. *Proceedings of the IEEE* 90: 222–244.

Saleh BEA and Teich MC (1991) *Fundamentals of Photonics*. New York: Wiley.

Stamnes JJ (1986) *Waves in Focal Regions*. Bristol: Hilger.

Volume Holographic Imaging

G Barbastathis, Massachusetts Institute of Technology, Cambridge, MA, USA

© 2005 Published by Elsevier Ltd.

Volume holographic imaging (VHI) refers to a class of imaging techniques which use a volume hologram in at least one location within the optical path. The volume hologram acts as a 'smart lens,' which processes the optical field to extract spatial information in three dimensions (lateral as well as longitudinal) and spectral information.

Figure 1 is a generic VHI system. The object is either illuminated by a light source (e.g., sunlight or a pump laser) as shown in the figure, or it may be self-luminous. Light scattered or emitted by the object is first transformed by an objective lens and then illuminates the volume hologram. The role of the objective is to form an intermediate image which serves as input to the volume hologram. The volume hologram itself is modeled as a three-dimensional (3D) modulation $\Delta\varepsilon(\mathbf{r})$ of the dielectric index within a finite region of space. The light entering the hologram is diffracted by $\Delta\varepsilon(\mathbf{r})$ with efficiency η, defined as

$$\eta = \frac{\text{Power diffracted by the volume hologram}}{\text{Power incident to the volume hologram}} \quad [1]$$

We assume that diffraction occurs in the Bragg regime. The diffracted field is Fourier-transformed by the collector lens, and the result is sampled and measured by an intensity detector array (such as a CCD or CMOS camera).

Intuitively, we expect that a fraction of the illumination incident upon the volume hologram is Bragg-matched and is diffracted towards the Fourier-transforming lens. The remainder of the incident illumination is Bragg-mismatched, and as result is transmitted through the volume hologram un-diffracted. Therefore, the volume hologram acts as a filter which admits the Bragg-matched portion of the object and rejects the rest. When appropriately designed, this 'Bragg imaging filter' can exhibit very rich behavior, spanning the three spatial dimensions and the spectral dimension of the object.

To keep the discussion simple, we consider the specific case of a transmission geometry volume

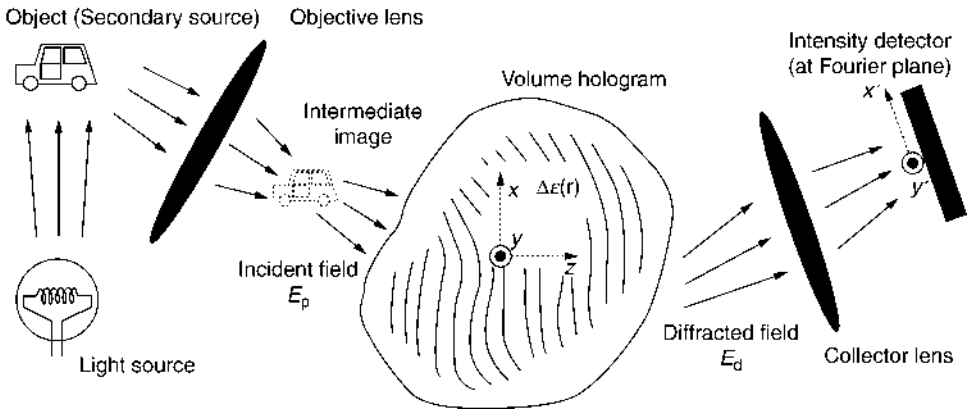

Figure 1 Volume holographic imaging (VHI) system.

hologram, described in Figure 2. The volume hologram is created by interfering a spherical wave and a planewave, as shown in Figure 2a. The spherical wave originates at the coordinate origin. The planewave is off-axis and its wavevector lies on the xz plane. As in most common holographic systems, the two beams are assumed to be at the same wavelength λ, and mutually coherent. The volume hologram results from exposure of a photosensitive material to the interference of these two beams.

First, assume that the object is a simple point source. The intermediate image is also approximately a point source, that we refer to as 'probe,' located somewhere in the vicinity of the reference point source. Assuming the wavelength of the probe source is the same as that of the reference and signal beams, volume diffraction theory shows that:

(i) If the probe point source is displaced in the y direction relative to the reference point source, the image formed by the volume hologram is also displaced by a proportional amount. Most common imaging systems would be expected to operate this way.
(ii) If the probe point source is displaced in the x direction relative to the reference point source, the image disappears (i.e., the detector plane remains dark).
(iii) If the probe point source is displaced in the z direction relative to the reference point source, a defocused and faint image is formed on the detector plane. 'Faint' here means that the fraction of energy of the defocused probe transmitted to the detector plane is much smaller than the fraction that would have been transmitted if the probe had been at the origin.

Now consider an extended, monochromatic, spatially incoherent object and intermediate image, as in Figure 2f. According to the above description, the volume hologram acts as a 'Bragg slit' in this case: because of Bragg selectivity, the volume hologram transmits light originating from the vicinity of the y axis, and rejects light originating anywhere else. For the same reason, the volume hologram affords depth selectivity (like the pinhole of a confocal microscope). The width of the slit is determined by the recording geometry, and the thickness of the volume hologram. For example, in the transmission recording geometry of Figure 2a, where the reference beam originates a distance z_0 away from the hologram, the planewave propagates at angle θ with respect to the optical axis z (assuming $\theta \ll 1$ radian), and the hologram thickness is L (assuming $L \ll z_0$), the width of the slit is found to be

$$\Delta x \approx \frac{\lambda z_0}{L\theta} \qquad [2]$$

The imaging function becomes richer if the object is polychromatic. In addition to its Bragg slit function, the volume hologram exhibits then dispersive behavior, like all diffractive elements. In this particular case, dispersion causes the hologram to image simultaneously multiple Bragg slits, each at a different color and parallel to the original slit at wavelength λ, but displaced along the z axis. Light from all these slits finds itself in focus at the detector plane, thus forming a 'rainbow image' of an entire slice through the object, as shown in Figure 2g.

To further exploit the capabilities of volume holograms, we recall that in general it is possible to 'multiplex' (superimpose) several volume gratings

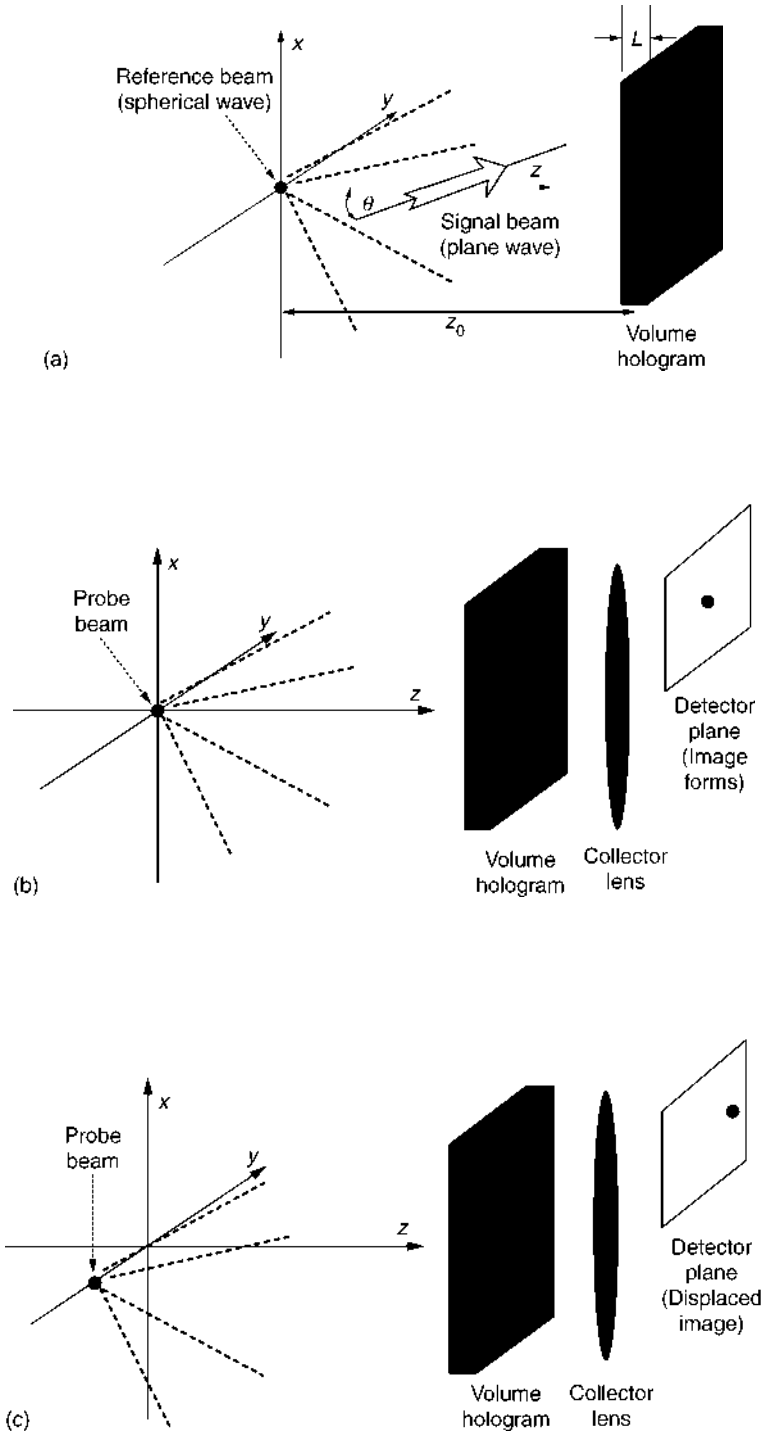

Figure 2 (a) Recoding of a transmission-geometry volume hologram with a spherical wave and an off-axis plane wave with its wave-vector on the *xz* plane. (b) Imaging of a probe point source that replicates the location and wavelength of the reference point source using the volume hologram recorded in part (a). (c) Imaging of a probe point source at the same wavelength but displaced in the *y* direction relative to the reference point source. (d) Imaging of a probe point source at the same wavelength but displaced in the *x* direction relative to the reference point source. (e) Imaging of a probe point source at the same wavelength but displaced in the *z* direction relative to the reference point source. (f) 'Bragg slitting:' Imaging of an extended monochromatic, spatially incoherent object using a volume hologram recorded as in (a). (g) Joining Bragg slits from different colors to form rainbow slices: Imaging of an extended polychromatic object using a volume hologram recorded as in (a). (h) Multiplex imaging of several slices using a volume hologram formed by multiple exposures.

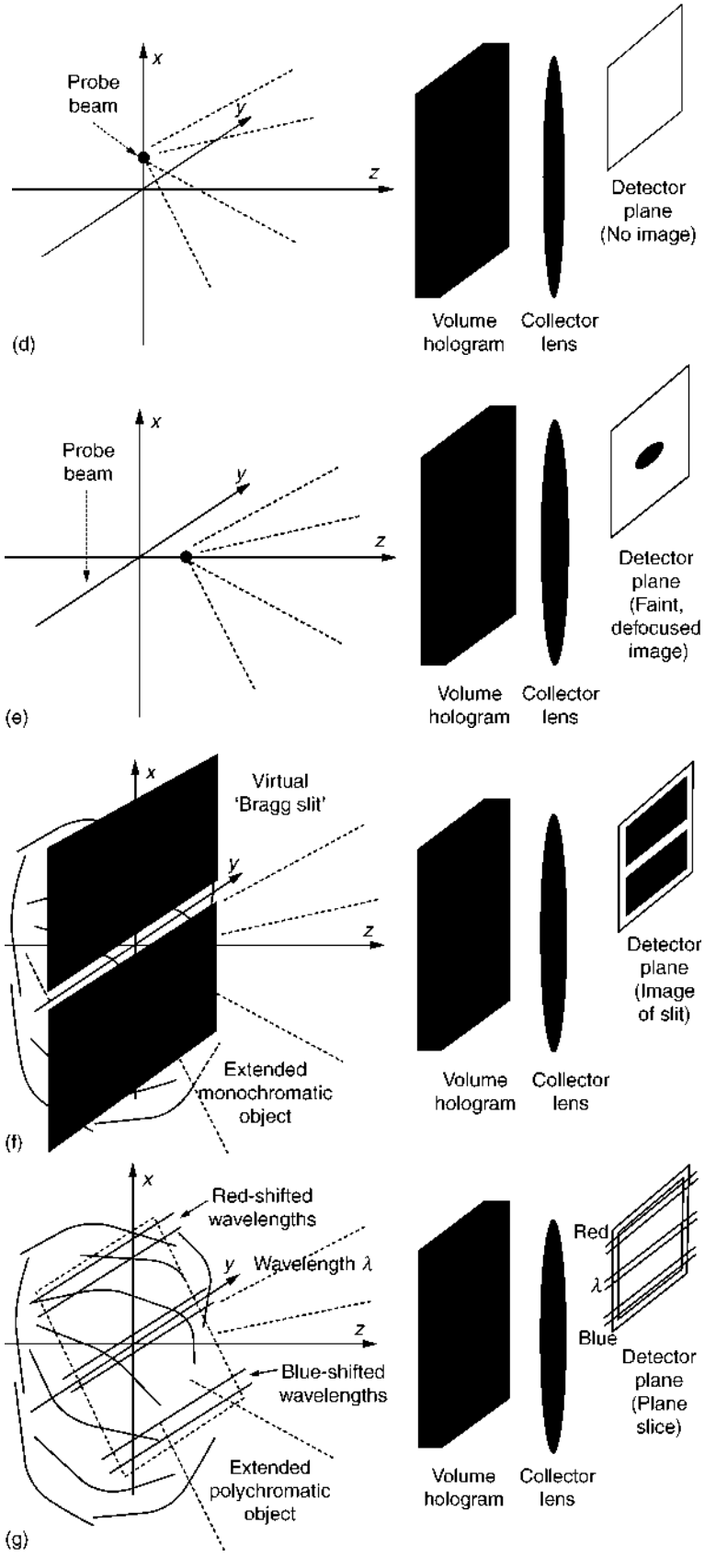

Figure 2 *Continued.*

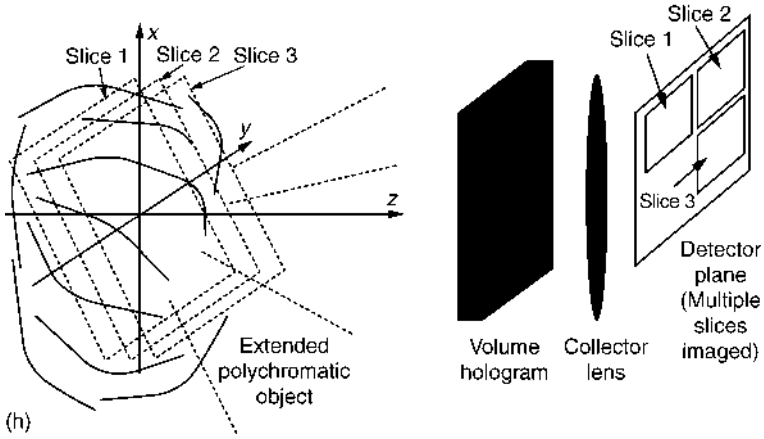

Figure 2 *Continued.*

within the same volume by successive exposures. In the imaging context, suppose that we multiplex several gratings similar to the grating described in Figure 2a but with spherical reference waves originating at different locations and plane signal waves at different orientations. When the multiplexed volume hologram is illuminated by an extended polychromatic source, each grating forms a separate image of a rainbow slice, as described earlier. By spacing appropriately the angles of propagation of the plane signal waves, we can ensure that the rainbow images are formed on nonoverlapping areas on the detector plane, as shown in Figure 2h. This device is now performing true four-dimensional (4D) imaging: it is separating the spatial and spectral components of the object illumination so that they can be measured independently by the detector array. Assuming the photon count is sufficiently high and the number of detector pixels is sufficient, this 'spatio-spectral slicing' operation can be performed in real time, without need for mechanical scanning.

The complete theory of VHI in the spectral and spatial domains is given in the Further Reading section below. Experimental demonstrations of VHI have been performed in the context of a confocal microscope where the volume hologram performs the function of a 'Bragg pinhole' as well as a real-time 4D imager. The limited diffraction efficiency η of volume holograms poses a major concern for VHI systems. Holograms, however, are known to act as matched filters. It has been shown that the matched filtering nature of volume holograms as imaging elements is superior to other filtering elements (e.g., pinholes or slits) in an information-theoretic sense. Efforts are currently underway to strengthen this property by design of volume holographic imaging elements which are more elaborate than described here.

See also

Holography, Techniques: Overview. **Imaging:** Three-Dimensional Field Transformations.

Further Reading

Barbastathis G and Brady DJ (1999) Multidimensional tomographic imaging using volume holography. *Proceedings of the IEEE* 87(12): 2098–2120.

Barbastathis G and Sinha A (2001) Information content of volume holographic images. *Trends in Biotechnology* 19(10): 383–392.

Barbastathis G, Balberg M and Brady DJ (1999) Confocal microscopy with a volume holographic filter. *Optics Letters* 24(12): 811–813.

Barbastathis G, Sinha A and Neifeld MA (2001) Information-theoretic treatment of axial imaging. In: *OSA Topical Meeting on Integrated Computational Imaging Systems*, pp. 18–20. Albuquerque, NM.

Liu W, Psaltis D and Barbastathis G (2004) Real time spectral imaging in three spatial dimensions. *Optics Letters* in press.

van-der Lugt A (1964) Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory* IT-10(2): 139–145.

Wavefront Sensors and Control (Imaging Through Turbulence)

C L Matson, Air Force Research Laboratory, Kirtland, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In many situations, distant objects are imaged using optical or near-infrared imaging systems. Examples include terrestrial surveillance from space, tactical surveillance from airborne imaging systems, and ground-based astronomical imaging. Ideally, the resolution in these images is limited only by the wavelength of light λ used to create the images divided by the imaging system's effective diameter D. The ratio λ/D is called the diffraction limit of the imaging system. Unfortunately, when the imaging system is located in the Earth's atmosphere, it is often atmospheric turbulence fluctuations that limit the level of detail in the long-exposure images, not the diffraction limit of the telescope. For example, the Keck telescopes, located on top of Mauna Kea in Hawaii, are 10 m diameter telescopes that have a diffraction-limited angular resolution of 0.2 μrad for an imaging wavelength of 2 μm. However, atmospheric turbulence limits the achievable resolution in long-exposure images to the order of 1–5 μrad at this wavelength.

The important discovery by Anton Labeyrie in 1970 that short-exposure images of objects retain meaningful diffraction-limited information about the object, where 'short' means that the exposure time is less than the atmospheric turbulence coherence time, has revolutionized imaging through turbulence. Atmospheric coherence times depend upon a number of factors, including wind speeds, wind altitudes, and observation wavelength. At astronomical observatories for visible observation wavelengths, a rule of thumb is that atmospheric turbulence stays essentially constant for exposure times of one to ten milliseconds under average clear-sky weather conditions. A variety of techniques that exploit this discovery have been developed. These techniques can be grouped into three different categories: techniques that remove blurring from the detected intensity images (post-detection compensation), techniques that modify the optical field prior to detecting the light (pre-detection compensation), and hybrid techniques that use both post-detection and pre-detection compensation. In this article, a variety of these techniques are presented that are representative of the myriad of methods that have been developed during the past decades. To simplify the discussion, it is assumed that the blurring due to atmospheric turbulence is the same at every point in the image (isoplanaticity), that the images are formed using incoherent light as emitted by the sun and stars, and that a single filled aperture is used to create the images. Interferometric techniques that use multiple apertures to estimate the mutual coherence function of the optical field and calculate an image from it, are discussed in a separate article (*see* **Imaging:** Interferometric Imaging).

The emphasis in this article is on identifying and removing atmospheric blurring in images. As part of this process, regularization techniques must be included in the image reconstruction algorithms in order to obtain the best possible resolution for an acceptable level of noise. Although regularization techniques are not discussed in this article, information can be found in the Further Reading section at the end of this article (*see* **Imaging:** Inverse Problems and Computational Imaging).

Imaging and Atmospheric Turbulence

In Figure 1, a conceptual diagram illustrates the impact that atmospheric turbulence has upon resolution in classical imaging systems. A distant star is being imaged by a telescope/detector system. In the absence of atmospheric turbulence, the field emitted by the star has a planar wavefront over the extent of the lens since the star is a long distance from the imaging system and thus unresolved by the telescope. When imaged by the telescope, the diffraction pattern of the telescope is obtained, resulting in a diffraction-limited image (Figure 2). However, in the presence of atmospheric turbulence near the telescope, the planar wavefront is randomly distorted by the turbulence, resulting in a corrugated wavefront with a spatial correlation scale given by the Fried parameter r_o. For visible wavelengths, at good astronomical observing sites, r_o ranges from 5 cm to 20 cm. When this corrugated wavefront is converted to an image by the telescope, a distorted image is obtained (Figure 3). Because atmospheric turbulence is time-varying, the distortions in the image also change over time. If the imaging shutter on the telescope remains open for many occurrences of atmospheric turbulence, the resulting image will be an average of all the individually distorted images, resulting in what is known as the seeing disk (Figure 4). It is the seeing disk that typically limits spatial resolution in long-exposure images, not the diffraction pattern of the

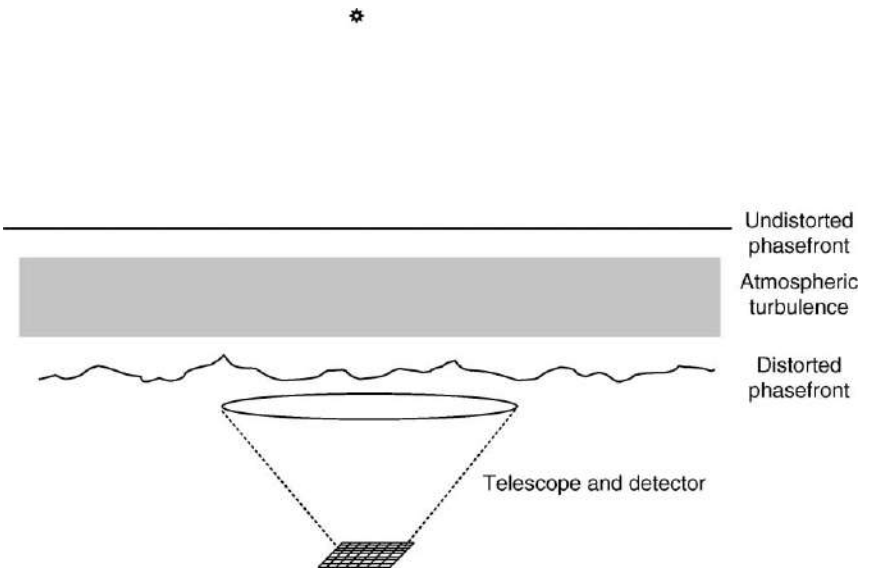

Figure 1 Conceptual diagram of imaging through atmospheric turbulence.

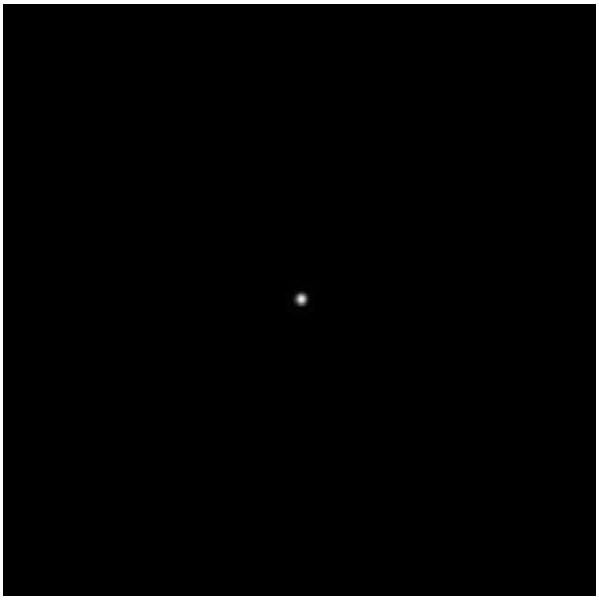

Figure 2 Computer-simulated diffraction-limited image of a point source.

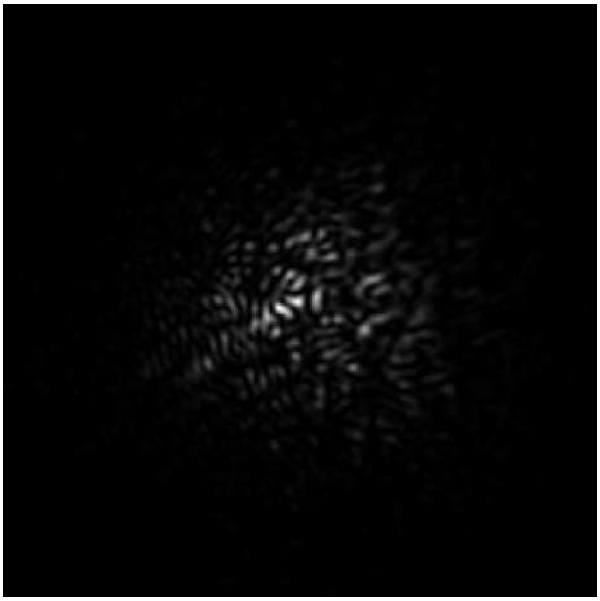

Figure 3 Computer-simulated short-exposure image of a point source through atmospheric turbulence.

telescope. As a result, the effective spatial resolution in long-exposure images in the presence of atmospheric turbulence is given by λ/r_o, not λ/D.

Atmospheric turbulence distorts a field's wavefront by delaying it in a way that is spatially and temporally random. The origin of these delays is the random variation in the index of refraction of the atmosphere as a result of turbulent air motion brought about by temperature variations. A Fourier optics imaging model is useful in characterizing the effects of atmospheric turbulence on spatial resolution in images. Let $i(\vec{x})$ be an image of an object $o(\vec{x})$, where $\vec{x}$ is a two-dimensional spatial vector, and let $I(\vec{f})$ and $O(\vec{f})$ be their spatial Fourier transforms, where $\vec{f}$ is a two-dimensional spatial frequency vector. In the absence of atmospheric turbulence, the diffraction-limited Fourier transform $I(\vec{f})$ is given by

$$I_d(\vec{f}) = O(\vec{f})H_d(\vec{f}) \qquad [1]$$

where the subscript 'd' denotes the diffraction-limited case and where $H_d(\vec{f})$ is the diffraction-limited

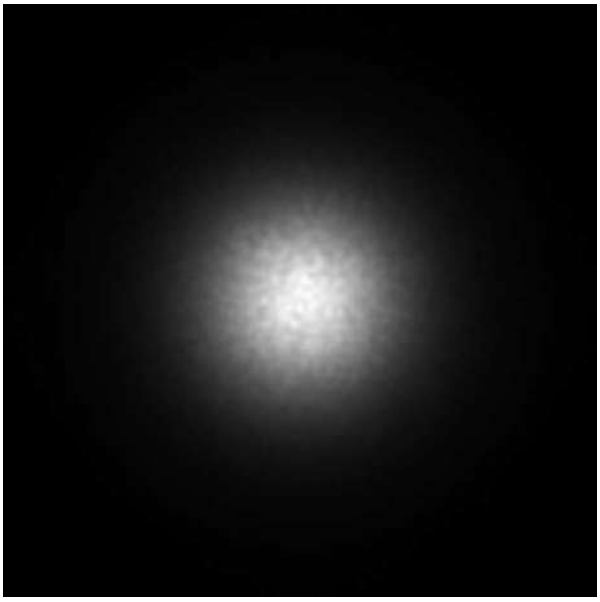

Figure 4 Computer-simulated long-exposure image of a point source through atmospheric turbulence.

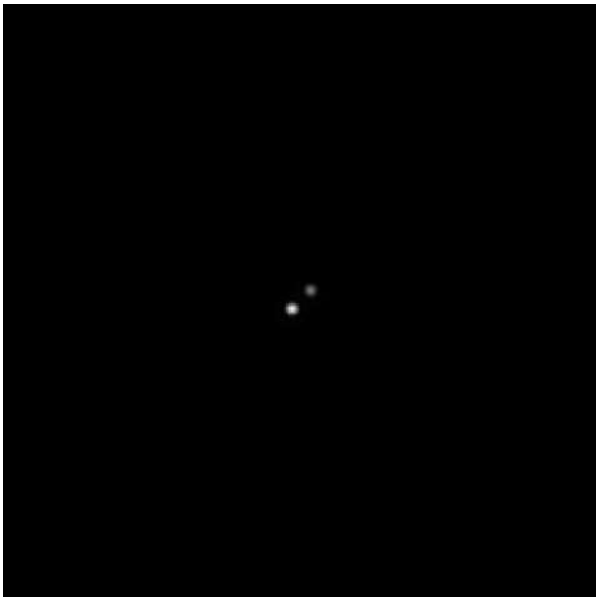

Figure 6 Computer-simulated diffraction-limited image of a binary star.

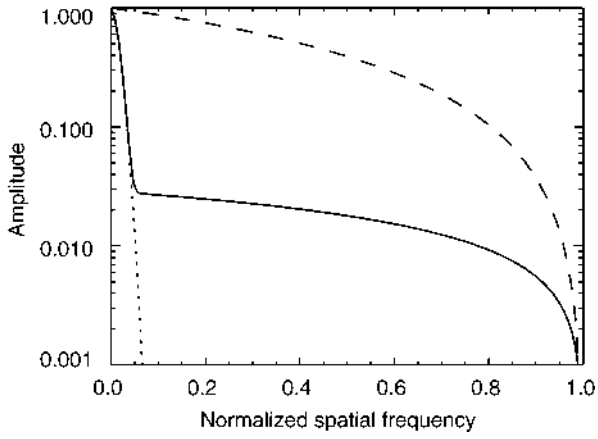

Figure 5 Plots of the Fourier amplitudes of the diffraction-limited transfer function (dashed line), Fourier amplitudes of the long-exposure transfer function (dotted line), and the square root of the average short exposure energy spectrum. The spatial frequency axis is normalized to one at the diffraction limit of the telescope.

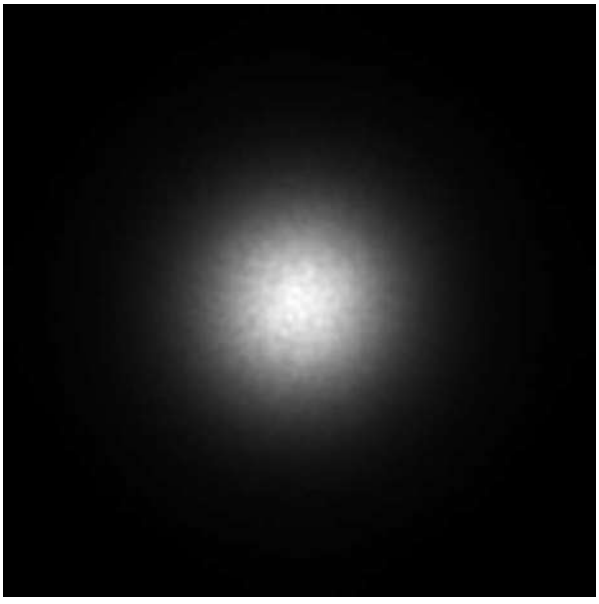

Figure 7 Computer-simulated long-exposure image of a binary star.

telescope transfer function. In the presence of atmospheric turbulence, the Fourier transform of a long exposure image $i_{le}(\vec{x})$ is given by

$$I_{le}(\vec{f}) = O(\vec{f})H_d(\vec{f})H_{le}(\vec{f}) = I_d(\vec{f})H_{le}(\vec{f}) \qquad [2]$$

where the subscript 'le' denotes long exposure. If the turbulence satisfies Kolmogorov statistics, as is typically assumed, then:

$$H_{le}(\vec{f}) = \exp\left[-3.44\left(\frac{\lambda d|\vec{f}|}{r_o}\right)^{5/3}\right] \qquad [3]$$

where d is the distance between the imaging system's exit pupil and detector plane. As can be seen from eqn [2], the effect of atmospheric turbulence on a long-exposure image is a multiplication of the diffraction-limited image's Fourier transform by a transfer function due to atmospheric turbulence. In Figure 5, an amplitude plot of a diffraction-limited transfer function is displayed along with an amplitude plot of a long-exposure transfer function for a 2.3 m

diameter telescope, an imaging wavelength of 0.5 μm, and an r_o value of 10 cm. Notice that the long exposure transfer function significantly reduces the amount of high spatial frequency information in an image. Diffraction-limited and long-exposure turbulence-limited images of a binary star are shown in Figures 6 and 7, illustrating the significant loss of spatial resolution brought on by atmospheric turbulence.

Post-Detection Compensation

Speckle Imaging

Labeyrie's 1970 discovery was motivated by his analysis of short-exposure star images. In these images (see Figure 3), a number of speckles are present, where the individual speckle sizes are roughly the size of the diffraction-limited spot of the telescope. From this, he realized that high spatial frequency information is retained in short-exposure images, albeit distorted. He then showed that the energy spectra $E_I(\vec{f})$ of the individual short-exposure images can be averaged to obtain an estimate of the average energy spectrum that retains high signal-to-noise ratio (SNR) information at high spatial frequencies. The square root of the average energy spectrum (the amplitude spectrum) of a star, calculated using Labeyrie's method (also known as speckle interferometry), is shown in Figure 5, along with the diffraction-limited and long-exposure amplitude spectra. Notice that meaningful high spatial frequency information is retained in the average amplitude spectrum out to the diffraction limit of the telescope. Because the atmospherically corrupted amplitude spectrum estimate is significantly lower than the diffraction-limited amplitude, a number of short exposure images must be used to build up the SNRs at the higher spatial frequencies.

To reconstruct an image, both the Fourier amplitudes and Fourier phases of the image must be known. In fact, it is well known that the Fourier phase of an image carries most of the information in the image. Because only the Fourier amplitudes are available using Labeyrie's method, other techniques were subsequently developed to calculate the Fourier phase spectrum from a sequence of short-exposure images. Techniques that combine amplitude spectrum estimation using Labeyrie's method along with phase spectrum estimation are called speckle imaging techniques. The two most widely used phase spectrum estimation methods are the Knox–Thompson (or cross spectrum) and bispectrum algorithms. Both of these algorithms calculate Fourier-domain quantities that retain high spatial frequency information about the Fourier phase spectrum when averaged over multiple short-exposure images. The cross spectrum and bispectrum techniques calculate the average cross spectrum $C(\vec{f}, \Delta\vec{f})$ and bispectrum $B(\vec{f}_1, \vec{f}_2)$, respectively, that are given by

$$C(\vec{f}, \Delta\vec{f}) = I(\vec{f})I^*(\vec{f} + \Delta\vec{f}) \qquad [4]$$

$$B(\vec{f}_1, \vec{f}_2) = I(\vec{f}_1)I(\vec{f}_2)I^*(\vec{f}_1 + \vec{f}_2) \qquad [5]$$

where the superscript $*$ denotes complex conjugate. Although the cross spectrum is defined for all values of $\vec{f}$ and $\Delta\vec{f}$, only for values of $\Delta\vec{f}$ or $\vec{f}$ less than r_o/λ are the cross spectrum elements obtained with sufficiently high SNRs to be useful in the reconstruction process. The equivalent constraint on the bispectrum elements is that either $\vec{f}_1, \vec{f}_2$, or $\vec{f}_1 + \vec{f}_2$ must be less than r_o/λ. Once the average cross spectrum or bispectrum is calculated from a series of short-exposure images, the phase spectrum $\phi_I(\vec{f})$ of the underlying image can be calculated from their phases using one of a number of algorithms based upon the following equations:

$$\phi_I(\vec{f} + \Delta\vec{f}) = \phi_I(\vec{f}) - \phi_C(\vec{f} + \Delta\vec{f}) \qquad [6]$$

for the cross spectrum and:

$$\phi_I(\vec{f}_1 + \vec{f}_2) = \phi_I(\vec{f}_1) + \phi_I(\vec{f}_2) - \phi_B(\vec{f}_1 + \vec{f}_2) \qquad [7]$$

for the bispectrum. Because the phases are only known for units of 2π, phasor versions of eqns [6] and [7] must be used. For the cross spectrum technique, each short-exposure image must be centroided prior to calculating its cross spectrum and adding it to the total cross spectrum. The bispectrum technique is insensitive to tilt, so no centroiding is necessary; however, the bispectrum technique requires setting two image phase spectrum values to arbitrary values to initialize the phase reconstruction process. Typically, the image phase spectrum values at $\vec{f} = (0, 1)$ and $\vec{f} = (1, 0)$ are set equal to zero, which centroids the reconstructed image but does not affect its morphology.

To obtain an estimated Fourier transform $I_e(\vec{f})$ of the image from the energy and phase spectrum estimates, the square root of the average energy spectrum is combined with the average phase spectrum calculated from either the bispectrum or cross spectrum. The result is given by

$$I_e(\vec{f}) = [E_I(\vec{f})]^{1/2} \exp[j\phi_I(\vec{f})] \qquad [8]$$

where $j = \sqrt{-1}$. The result in eqn [8] includes atmospheric attenuation of the Fourier amplitudes, so it must be divided by the amplitude spectrum of the atmosphere to obtain the desired object spectrum estimate. The atmospheric amplitude spectrum estimate is obtained by using Labeyrie's technique on a sequence of short-exposure images of a nearby unresolved star. No phase spectrum correction is necessary because it has been shown theoretically and verified experimentally that atmospheric turbulence has a negligible effect upon $\phi_{\mathrm{I}}(\vec{f})$.

Blind Deconvolution

Speckle imaging techniques are effective as well as easy and fast to implement. However, it is necessary to obtain a separate estimate of the atmospheric energy spectrum using an unresolved star in order to reconstruct a high-resolution image. Usually, the star is not co-located with the object and the star images are not collected simultaneously with the images of the object. As a result, the atmospheric energy spectrum estimate may be inaccurate. At the very least, collecting the additional star data expends valuable observation time. In addition, there is often meaningful prior knowledge about the object and the data, such as the fact that measured intensities are positive (positivity), that many astronomical objects are imaged against a black background (support), and that the measured data are band-limited, among others. Speckle imaging algorithms do not exploit this additional information. As a result, a number of techniques have been developed to jointly estimate the atmospheric blurring effects as well as an estimate of the object being imaged using only the short-exposure images of the object and available prior knowledge. These techniques are called blind deconvolution techniques. Two of the most common techniques are known as multi-frame blind deconvolution (MFBD) and projection-based blind deconvolution (PBBD).

The MFBD algorithm is based upon maximizing a figure of merit to determine the best object and atmospheric short-exposure point spread function (PSF) estimates given the data and available prior knowledge. This figure of merit is created using a maximum likelihood problem formulation. Conceptually, this technique produces estimates of the object and the atmospheric turbulence short-exposure PSFs that most likely generated the measured short-exposure images. Mathematically, the MFBD algorithm obtains these estimates by maximizing the conditional probability density function (pdf) $\mathrm{p}[i_{\mathrm{m}}(\vec{x})|o(\vec{x}); h_{\mathrm{m}}(\vec{x}),\ m = 1, \ldots, M]$, where $i_{\mathrm{m}}(\vec{x})$ is the mth of M short exposure images and is given by

$$i_{\mathrm{m}}(\vec{x}) = h_{\mathrm{m}}(\vec{x}) * o(\vec{x}) + n_{\mathrm{m}}(\vec{x}) \qquad [9]$$

where $h_{\mathrm{m}}(\vec{x})$ and $n_{\mathrm{m}}(\vec{x})$ are the atmospheric/system blur and detector noise functions for the mth image, respectively, and the asterisk denotes convolution. The noise is assumed to be zero mean with variance $\sigma^2(\vec{x})$. The conditional pdf $\mathrm{p}[i_{\mathrm{m}}(\vec{x})|o(\vec{x}); h_{\mathrm{m}}(\vec{x}),\ m = 1, \ldots, M]$ is the pdf of the noise with a mean equal to $h_{\mathrm{m}}(\vec{x}) * o(\vec{x})$ and, as a function of $o(\vec{x})$ and $h_{\mathrm{m}}(\vec{x})$, it is known as a likelihood function. For spatially independent Gaussian noise, as is the case for camera read noise, the likelihood function is given by

$$\mathrm{L_G}[o(\vec{x}); h_{\mathrm{m}}(\vec{x}),\ m = 1, \ldots, M] = \prod_m \prod_{\vec{x}} \frac{1}{\sqrt{2\pi\sigma^2(\vec{x})}} \times \exp\left\{ -\left[i(\vec{x}) - (o * h_{\mathrm{m}})(\vec{x})\right]^2 / 2\sigma^2(\vec{x}) \right\} \qquad [10]$$

For Poisson noise, the likelihood function is given by

$$\mathrm{L_P}[o(\vec{x}); h_{\mathrm{m}}(\vec{x}), \mathrm{m} = 1, \ldots, M] = \prod_m \prod_{\vec{x}} \frac{\left[(o * h_{\mathrm{m}})(\vec{x})\right]^{i(\vec{x})} \exp\left\{ -\left[(o * h_{\mathrm{m}})(\vec{x})\right] \right\}}{i(\vec{x})!} \qquad [11]$$

In practice, logarithms of eqns [10] and [11] are taken prior to searching for the estimates of $o(\vec{x})$ and $h_{\mathrm{m}}(\vec{x})$ that maximize the likelihood function.

In general, there are an infinite number of estimates of $o(\vec{x})$ and $h_{\mathrm{m}}(\vec{x})$ that reproduce the original dataset when convolved together. However, when additional prior knowledge is included to constrain the solution space, such as positivity, support, and the knowledge that $h_{\mathrm{m}}(\vec{x})$ is band-limited, along with regularization, the MFBD algorithm almost always converges to the correct estimates of $o(\vec{x})$ and $h_{\mathrm{m}}(\vec{x})$.

The PBBD technique also uses prior knowledge constraints to jointly estimate $o(\vec{x})$ and $h_{\mathrm{m}}(\vec{x})$. However, instead of creating and maximizing a likelihood function, the PBBD technique uses the method of convex projections to find estimates of $o(\vec{x})$ and $h_{\mathrm{m}}(\vec{x})$ that are consistent with the measured short-exposure images and all the prior knowledge constraints. The data consistency constraint and the prior knowledge constraints are mathematically formulated as convex sets and the PBBD technique repeatedly takes the current estimates of $o(\vec{x})$ and $h_{\mathrm{m}}(\vec{x})$ and projects them onto the convex constraint sets. As long as all the sets are truly convex, the

algorithm is guaranteed to converge to a solution that is in the intersection of all the sets as long as the algorithm doesn't stagnate at erroneous solutions called traps. These traps are the projection equivalent of local minima in cost function minimization algorithms.

A key component of this algorithm is the ability to formulate the measured short-exposure images and prior knowledge constraints into convex sets. As an example, consider the convex set $C_{\text{pos}}[u(\vec{x})]$ that corresponds to the positivity constraint. This set is given by

$$C_{\text{pos}}[u(\vec{x})] \triangleq \{u(\vec{x}) \text{ such that } u(\vec{x}) \geq 0\} \qquad [12]$$

It is straightforward to show that $C_{\text{pos}}[u(\vec{x})]$ is a convex set. Other constraints can be converted into convex sets in a similar manner. Although all constraints can be formulated into sets, most but not all of these sets are convex.

The projection process is straightforward to implement. For example, the projections of the current estimates of $o(\vec{x})$ and $h_{\text{m}}(\vec{x})$ onto $C_{\text{pos}}[u(\vec{x})]$ are carried out by zeroing out the negative values in these estimates. Similarly, the projections of $o(\vec{x})$ and $h_{\text{m}}(\vec{x})$ onto the convex sets that correspond to their support constraints are carried out by zeroing out all their values outside of their supports. In practice, some versions of PBBD algorithms have been formulated to use minimization routines as part of the projection process in order to minimize the possibility of stagnation.

Deconvolution from Wavefront Sensing

A key distinguishing feature of post-detection compensation algorithms is the type of additional data required to estimate and remove atmospheric blurring in the measured short-exposure images. Blind deconvolution techniques need no additional data. Speckle imaging techniques estimate average atmospheric blurring quantities and need an estimate of the average atmospheric energy spectrum. Because average quantities are calculated when using speckle imaging techniques, only an estimate of the average atmospheric energy spectrum is needed, not the specific energy spectra that corrupted the measured short-exposure images. As a result, the additional data need not be collected simultaneously with the short-exposure images. The technique described in this section, deconvolution from wavefront sensing (DWFS), seeks to deconvolve the atmospheric blurring in each short-exposure image by estimating the specific atmospheric field phase perturbations that blur each short-exposure image. Thus, the additional data must be collected simultaneously with the short-exposure images. The atmospheric field phase perturbations are calculated from data obtained with a wavefront sensor using a different region of the optical spectrum than used for imaging so that no light is taken from the imaging sensor. Most commonly, the wavefront sensor data consists of the first derivatives of the phasefront (phase differences) as obtained with Shack–Hartmann or shearing interferometer wavefront sensors, but curvature sensing wavefront sensors are also used that produce data that are proportional to the second derivatives of the phasefront.

The original algorithm to implement the DWFS technique produced an estimate $\hat{O}(\vec{f})$ of the Fourier transform $O(\vec{f})$ of the object being imaged using the following estimator:

$$\hat{O}(\vec{f}) = \frac{\sum_m I_{\text{m}}(\vec{f})\tilde{H}^*_{\text{m}}(\vec{f})}{\sum_m |\tilde{H}_{\text{m}}(\vec{f})|^2} \qquad [13]$$

where $\tilde{H}_{\text{m}}(\vec{f})$ is the overall system transfer function (including the telescope and atmospheric effects) for the mth short-exposure image estimated from the wavefront sensor data. To create $\tilde{H}_{\text{m}}(\vec{f})$, the atmospheric field phase perturbations corrupting the mth image must be calculated from the wavefront sensor measurements. This topic is discussed in more detail in **Imaging**: Adaptive Optics. Then $\tilde{H}_{\text{m}}(\vec{f})$ is obtained using standard Fourier optics techniques. Although eqn [13] is easily implemented given an existing system that includes a wavefront sensor, it suffers from the fact that it is a biased estimator. An unbiased version of eqn [13] is given by

$$\hat{O}(\vec{f}) = \frac{\sum_m I_{\text{m}}(\vec{f})\tilde{H}^*_{\text{m}}(\vec{f})}{\sum_n H_{\text{n,ref}}(\vec{f})\tilde{H}^*_{\text{n,ref}}(\vec{f})} \qquad [14]$$

where $\tilde{H}_{\text{n,ref}}(\vec{f})$ and $H_{\text{n,ref}}(\vec{f})$ are estimates of the overall system transfer function obtained from data collected on a reference star using wavefront sensor data and image-plane data, respectively. Thus, to obtain an unbiased estimator, two more datasets must be collected. The number of frames in the reference star dataset need not be the same as for the imaging dataset. As for speckle imaging, the atmospheric statistics must be the same for the reference star and imaging datasets.

Another technique that uses at least one additional data channel along with the measured image is called

phase diversity. In the additional data channels, images are also recorded, but with known field phase aberrations intentionally introduced to add known diversity to the unknown field phase corrupting the measurements. The most commonly introduced field phases aberration in an additional channel is defocus due to its ease of implementation. In the case of defocus, the additional data channel along with the imaging channel can be viewed as samples of the focal volume of the imaging system. Both the noise-free object and the unknown field phase aberrations can be recovered from the measured data using optimization techniques to enforce data consistency along with other prior knowledge constraints. These types of systems can use higher-order aberrations as well. The phase diversity concept has been generalized to an approach known as computational imaging, where an imaging system is designed by taking into account its intended use, the expected image degradation mechanisms, and the fact that the image will be the product of both an optical system and a postprocessing algorithm.

Pre-Detection Compensation

The primary pre-detection-only compensation technique in use today employs dynamical electro-optical systems to sense and remove the phases in the impinging wavefront due to atmospheric turbulence. This technique, known as adaptive optics, provides much higher-quality data than post-detection compensation techniques, when there is enough light from the object being imaged or a guidestar (artificial or natural) to drive the adaptive optics system at a bandwidth high enough to keep up with the temporal changes in atmospheric turbulence. It consists of a wavefront sensor that measures the portion of the phasefront in the pupil of the telescope that is due to atmospheric turbulence, a wavefront reconstructor that generates the deformable mirror drive signals from the wavefront sensor data, and a deformable mirror that applies the conjugate of the atmospheric phase distortions to a reimaged pupil. Some of the light reflected off the deformable mirror is then sent to the imaging sensor. Adaptive optics systems are discussed in much greater detail in **Imaging**: Adaptive Optics. Here, it suffices to say that no adaptive optics system produces imagery that is diffraction limited in quality, although, under the right conditions, the images can be near diffraction limited. However, atmospheric seeing conditions fluctuate significantly over a night as well as seasonally. Therefore, virtually all adaptive optics images benefit from applying post-compensation techniques, as is discussed next.

Hybrid Techniques

Adaptive Optics and Post-Processing

Often, an adaptive optics system does not fully correct for atmospheric turbulence. This can occur when atmospheric conditions become worse than the system is designed to handle. Also, an adaptive optics system may be intentionally designed to only partially correct for atmospheric turbulence in order to lower the cost of the system. As a result, there still may be significant atmospheric blurring in systems that use adaptive optics. For this reason, additional post-processing of adaptive optics imagery can result in significant improvement in image quality. All of the techniques discussed in the section on post-detection compensation can be applied to adaptive optics imagery. Applying these techniques should never degrade image quality, and will improve the quality if the adaptive optics system doesn't fully correct for atmospheric turbulence. Generally speaking, the amount of improvement decreases as the performance of the adaptive optics system increases. Because the adaptive optics system's performance parameters can vary as atmospheric conditions vary, the blind deconvolution algorithms are especially applicable to processing adaptive optics imagery since they estimate the overall system transfer function from the imaging data. Because speckle imaging techniques require a separate estimate of the overall system transfer function, it can be difficult and time-consuming to obtain all of the additional data necessary to apply these techniques to adaptive optics data. Since these post-detection compensation techniques require short-exposure images, the adaptive optics system needs to have a camera capable of collecting short-exposure images. Finally, even if an adaptive optics system fully corrects for the atmosphere, deconvolution techniques that remove the diffraction-limited transfer function amplitude attenuation can improve image quality.

Partially Redundant Pupil Masking

Although adaptive optics systems can function very well, currently they are expensive and require significant maintenance. The primary benefit of adaptive optics systems is that they remove atmospheric turbulence phase distortions in the pupil of the telescope, thus permitting light from all areas of the pupil to interfere in phase when creating an image. An interferometric view of imaging reveals that multiple regions of the light in the pupil contribute to a single spatial frequency in the image. When the phasefront in the pupil is distorted

by the atmosphere, these multiple regions tend to not interfere in phase, reducing the amplitude of the image's Fourier component at that spatial frequency. This amplitude reduction effect occurs for all spatial frequencies greater than the limit imposed by the atmosphere, r_o/λ. Therefore, techniques have been developed to modify the wavefront by masking part of the pupil prior to measuring the field with an imaging sensor. This class of techniques has various names including optical aperture synthesis, pupil masking, and aperture masking. Here this class of techniques will be referred to as partially redundant pupil masking (PRPM).

PRPM techniques function by masking the pupil in such a way as to minimize the amount of out-of-phase interference of light on the imaging detector. Early efforts used pupil masks that were opaque plates with r_o-sized holes in them. Because the atmospheric coherence length is r_o, all the light from an r_o-sized hole is essentially in phase. By choosing the holes in the mask so that only two holes contribute to a single spatial frequency in the image, out-of-phase interference problems are removed. This type of mask is called a nonredundant pupil mask. Of course, a significant amount of light is lost due to this masking, so only bright objects benefit from this particular implementation. Often, the pupil mask needs to be rotated in order to obtain enough spatial frequencies in the image to be able to reconstruct a high-quality estimate of the object. To increase the light levels in the image, masks have been made to allow some out-of-phase interference for a given spatial frequency, thus trading off the amplitude of the spatial frequencies allowed in the image with the amount of light in the image. Masks with this property are called partially redundant pupil masks.

One version of PRPM is called the variable geometry pupil (VGP) technique. It is envisioned to work in conjunction with an adaptive optics system that does not fully correct the wavefront for atmospheric phase distortions. It employs a spatial light modulator or a micro-electromechanical mirror between the deformable mirror and the imaging sensor. Either of these two devices is driven by signals from the wavefront sensor in the adaptive optics system in such a way that the portions of the pupil where the phasefront has a sufficiently large second derivative is blocked, while the remainder of the phasefront is transmitted.

All PRPM techniques benefit from post-compensation processing. For the case of nonredundant pupil masks, post-processing techniques are required to convert the discrete-spatial-frequency information into an image with reasonable quality. These techniques are essentially the same techniques as used in interferometric imaging, not those described in this article. When partially redundant pupil masks are used (including VGP masks), the image may include enough spatial frequency content that the methods described in the post-detection compensation section can be applied.

See also

Imaging: Adaptive Optics; Interferometric Imaging; Inverse Problems and Computational Imaging.

Further Reading

Ayers GR, Northcott MJ and Dainty JC (1988) Knox–Thompson and triple-correlation imaging through atmospheric turbulence. *Journal of the Optical Society of America A* 5: 963–985.

Bertero M and Boccacci P (1998) *Introduction to Inverse Problems in Imaging*. Bristol, UK: Institute of Physics Publishing.

Craig IJD and Brown JC (1986) *Inverse Problems in Astronomy*. Bristol, UK: Adam Hilger Ltd.

Dainty JC (1984) Stellar speckle interferometry. In: Dainty JC (ed.) *Topics in Applied Physics, vol. 9: Laser Speckle and Related Phenomena*, 2nd edn, pp. 255–320. Berlin: Springer-Verlag.

Goodman JW (1985) *Statistical Optics*. New York: John Wiley and Sons.

Haniff CA and Buscher D (1992) Diffraction-limited imaging with partially redundant masks I: Infrared imaging of bright objects. *Journal of the Optical Society of America A* 9: 203–218.

Katsaggelos AK and Lay KT (1991) Maximum likelihood identification and restoration of images using the expectation-maximization algorithm. In: Katsaggelos AK (ed.) *Digital Image Restoration*, pp. 143–176. Berlin: Springer-Verlag.

Nakajima T and Haniff CA (1993) Partial adaptive compensation and passive interferometry with large ground-based telescopes. *Publications of the Astronomical Society of the Pacific* 105: 509–520.

Roggemann MC and Welsh B (1996) *Imaging Through Turbulence*. Boca Raton, FL: CRC Press.

Schulz TJ (1993) Multiframe blind deconvolution of astronomical images. *Journal of the Optical Society of America A* 10: 1064–1073.

Tyson RK (1991) *Principles of Adaptive Optics*. San Diego, CA: Academic Press.

Youla DC and Webb H (1982) Image restoration by the method of convex projections: Part 1 – Theory. *IEEE Transactions on Medical Imaging* M1-1: 81–94.

INCOHERENT SOURCES

Contents

Lamps
Synchrotrons

Lamps

V Roberts, Roberts Research and Consulting, Inc., Burnt Hills, NY, USA

© 2005 Roberts Research and Consulting, Inc. Published by Elsevier Ltd. All rights reserved.

Introduction

Electric lamps are by far the most common source of light used on Earth, other than our sun. These sources are composed of a very large number of individual radiating elements, all acting independently. This independence causes the light generated by electric lamps to have a very low degree of spatial and temporal coherence. When compared to lasers, lamps are usually considered to be incoherent sources.

Electric lamps are widely used for general lighting, such as residential, commercial and industrial space, and task lighting. A much less frequent use of incoherent sources, but perhaps of more interest to the reader, is as wavelength, intensity, and total light output standard used to calibrate various coherent light sources. We will discuss the general characteristics of incandescent, gas discharge, and solid state lamps, with sources in the Further Reading section at the end of this article, which are used as calibration standards for coherent sources.

Definitions

Before discussing individual lamp technologies, we need to define some of the key terms that are used for describing the performance of electric lamps. Most of these definitions are based on or abstracted from the IESNA Lighting Handbook:

- *Efficacy*: The total light output of a lamp, measured in lumens, divided by the power input in watts. Efficacy has the units of lumens per watt, or $\mathrm{l\,m\,W^{-1}}$, and can be greater than 100. The efficacy of common light sources varies from $10\,\mathrm{l\,m\,W^{-1}}$ to over $200\,\mathrm{l\,m\,W^{-1}}$.
- *Correlated Color Temperature* (*of a light source*) (*CCT*): the absolute temperature of a blackbody whose chromaticity most nearly resembles that of the light source. CCT is measured in Kelvin (K). The CCT of typical light sources varies from 2000 K to 6000 K, with the most popular sources being in the range from 2700 K to perhaps 4500 K.
- *Color Rendition Index* (*of a light source*) (*CRI*): a measure of the degree of color shift objects undergo when illuminated by the light source as compared with those same objects when illuminated by a reference source of comparable color temperature. The highest possible CRI is 100. The lowest theoretical CRI is less than zero, but CRI less than 50 is generally considered to have no value in determining the color quality of a light source.
- *Photopic Luminous Efficiency Function* (V_λ): a measure of the sensitivity of the human eye in lumens per watt of radiation received, as a function of wavelength, for luminance levels higher than $\sim 3\,\mathrm{cd\,m^{-2}}$. V_λ is based on relatively few observations, and individuals may differ significantly from the standard V_λ function. V_λ provides the basis for measuring the efficacy of all light sources.
- *Visible Light*: generally accepted to mean radiation between 380 nm and 780 nm.

Incandescent Lamps

Incandescent lamps are quasi-blackbody thermal radiators. A true blackbody radiator would generate light according to Planck's Law, as shown in eqn [1] and Figure 1, for various temperatures. Most incandescent lamps use tungsten filaments, and the emissivity of tungsten, the ratio of its radiative output to that of a blackbody at the same temperature, is about 0.44 in the visible region of the spectrum and about 0.33 in the infrared. The radiative output of a tungsten incandescent lamp is, therefore, less than half that of a blackbody of the same temperature and the intensity versus wavelength curve does not exactly follow Plank's law, being slightly favorable to visible

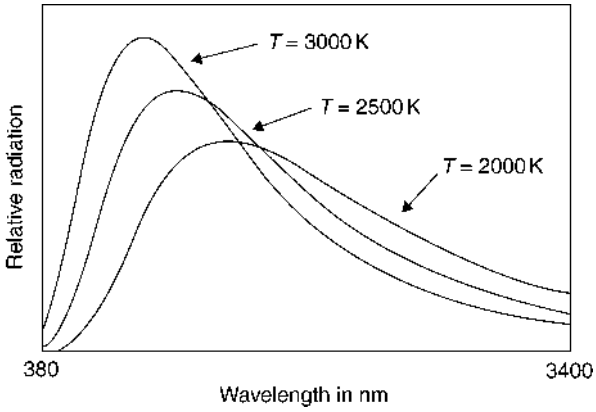

Figure 1 Radiation from blackbody source at various temperatures. © 2005 Roberts Research and Consulting, Inc. Published by Elsevier Ltd. All rights reserved.

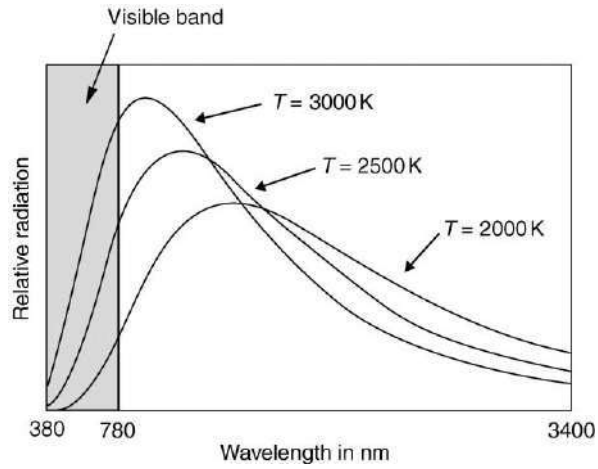

Figure 2 Radiation from blackbody source at various temperatures with visible band. © 2005 Roberts Research and Consulting, Inc. Published by Elsevier Ltd. All rights reserved.

radiation:

$$E(\lambda) = \frac{2\pi hc^2 10^{-9}}{\lambda^5(\exp^{hc/k\lambda T} - 1)} \qquad [1]$$

where:

E = power per square meter of surface area per nm of wavelength
λ = wavelength in meters
h = Plank's constant, 6.63×10^{-34} Joule seconds
c = speed of light, 3×10^8 meters $\sec^{-1}$
κ = Boltzman's constant, 1.38×10^{-23} Joule K^{-1}
T = temperature of the source in K

Tungsten lamps have excellent color quality, reflected in a perfect CRI of 100, are inexpensive, available in a great variety of sizes, shapes, and operating voltages, have a small source size that permits excellent light control with relatively small reflectors and lenses, run equally well on AC or DC power, can be easily dimmed, can have very long life, are insensitive to operating position, are relatively insensitive to ambient temperature, contain no hazardous materials, and light virtually instantly, though not as fast as LEDs. However, incandescent lamps have one major disadvantage; they are very inefficient. Most of the energy consumed by incandescent lamps is radiated in the infrared, while only 5% to 10% is radiated in the visible portion of the spectrum. The reason for this is shown in Figure 2, which shows the visible region of the spectrum superimposed on the blackbody radiation curves.

There are a number of techniques that have been used or are being investigated to increase the efficacy of incandescent lamps. First, as can be seen in Figure 2, increasing the filament temperature shifts the wavelength for peak emission toward lower wavelengths, and will result in a larger fraction of the radiated energy being in the visible region. Unfortunately, increasing filament temperature also increases the rate of tungsten evaporation and therefore shortens the life of tungsten lamps, since evaporation of tungsten is the dominant incandescent lamp failure mode. In a similar manner, the life of incandescent lamps can be increased to arbitrarily long times by decreasing filament temperature. This technique, which is the means used in far too many 'long life' incandescent lamps, unfortunately leads to a decrease in the already low efficacy of incandescent lamps. The approximate dependences of life, efficacy, and color temperature on operating voltage, for common incandescent lamps operating between 90% to 110% of rated voltage, are shown in eqns [2], [3], and [4]:

$$\frac{\text{LIFE}_1}{\text{LIFE}_2} = \left(\frac{\text{VOLTS}_2}{\text{VOLTS}_1}\right)^{13} \qquad [2]$$

$$\frac{\text{EFFICACY}_1}{\text{EFFICACY}_2} = \left(\frac{\text{VOLTS}_1}{\text{VOLTS}_2}\right)^{1.9} \qquad [3]$$

$$\frac{\text{COLOR_TEMPERATURE}_1}{\text{COLOR_TEMPERATURE}_2} = \left(\frac{\text{VOLTS}_1}{\text{VOLTS}_2}\right)^{0.42} \qquad [4]$$

Combining eqns [2] and [3] yields eqn [5]:

$$\frac{\text{LIFE}_1}{\text{LIFE}_2} = \left(\frac{\text{EFFICACY}_2}{\text{EFFACACY}_1}\right)^{6.84} \qquad [5]$$

Under the assumption that the color temperature of an incandescent lamp is approximately the same as the operating temperature of the filament, we can combine eqns [2] and [4] to obtain eqn [6], and combine eqns [3] and [4] to obtain eqn [7]:

$$\frac{\text{LIFE}_1}{\text{LIFE}_2} = \left(\frac{\text{TEMPERATURE}_2}{\text{TEMPERATURE}_1}\right)^{30.95} \qquad [6]$$

$$\frac{\text{EFFACACY}_1}{\text{EFFICACY}_2} = \left(\frac{\text{TEMPERATURE}_1}{\text{TEMPERATURE}_2}\right)^{4.52} \qquad [7]$$

One way to allow operation with higher filament temperatures, while not decreasing filament life, is to fill the lamp with a high molecular weight and/or high pressure gas, as either will retard tungsten evaporation. Lamps using krypton or xenon in place of the normal argon fill allow for a moderate increase in filament temperature. Better yet is a fill of halogen gas, such as HBr, CH_3Br, or CH_2Br^2 plus Kr or Xe at total pressures that range from 2 to 15 atmospheres. The high total pressure inhibits tungsten evaporation from the filament, and the halogen fill gas reacts with tungsten that deposits on the wall of the lamp to remove this tungsten and maintain light output over the life of the lamp. The tungsten removed from the wall of the lamp is redeposited on the tungsten filament, increasing lamp life under some circumstances. Due to the high gas pressures involved, tungsten halogen incandescent lamps are constructed in small 'filament tubes' to minimize both the amount of gas needed and the total explosive energy. The filament tubes are manufactured from quartz or special high temperature glass. For general purpose applications, the quartz or high temperature glass filament tube is enclosed in a larger glass envelope to protect the user should the filament tube explode and to protect the filament tube from environmental contaminants, such as salts from human fingers.

Since 90% to 95% of the energy consumed by an incandescent lamp is radiated as infrared energy, it is easy to see that the potential exists to dramatically increase the efficacy of incandescent lamps by reflecting some of this infrared energy back onto the filament, thus allowing the filament to be maintained at its normal operating temperature with far less electrical power. The reflector must have very low absorption in the visible region of the spectrum, high reflectivity in the IR and a reasonably sharp transition between the two, because the greatest IR emission is just below the visible band. The reflector and filament must also be precisely positioned so that the reflected IR radiation hits the filament, and the reflector must also be low enough in cost to be commercially viable. Successful IR reflectors have been made from multilayer dielectric films. These are similar to the coatings used to make high reflectivity mirrors for He–Ne lasers, but they have much broader bandwidth and lower reflectivity. The reflector-filament alignment problem is most easily solved using the geometry of a tungsten halogen filament tube as the carrier for the IR reflector. The typical efficacy gain is about 50%: a 60-watt IR-halogen lamp produces about as much light as a 90-watt halogen lamp.

Other methods being investigated to increase the efficacy of incandescent lamps involve modification of the filament to decrease its emissivity in the IR and/or increase its emissivity in the visible, and hence turn it into more of a 'selective radiator' than natural tungsten. Three general methods are under investigation: one involves creating a dense pattern of submicron size cavities on the surface of the tungsten to inhibit IR radiation; a second involves coating the tungsten with a material that is itself a selective radiator that favors the visible portion of the spectrum, while a third involves replacing the tungsten filament with a material that is a natural selective radiator. No products exist that employ these techniques, but active R&D projects are underway that may yield a more efficient incandescent lamp in the future.

Incandescent lamps can be used as standards for total flux and color temperature if carefully constructed and operated. The National Institutes for Standards and Technology (NIST) in the US sells 1000-watt tungsten halogen lamps calibrated to an accuracy of 0.6% in luminous intensity and 8 K in color temperature at a color temperature of 2856 K. Similar service are provided by standards organizations in other countries.

Discharge Lamps

The most efficient lamps produced today are based on electric discharges in gasses. These include 'low pressure' lamps such as fluorescent lamps and low pressure sodium lamps; and 'high pressure' or high intensity discharge (HID) lamps such as metal halide, high pressure sodium, high pressure mercury, and sulfur lamps. These lamps are all more efficient than incandescent lamps because they are 'selective radiators' instead of blackbody sources and have been designed to generate as much radiant energy as possible in the visible band while minimizing energy produced outside the visible band. The best white light discharge lamps have conversion efficiencies from electric power to visible light of 25% to 30%, making them almost five times as efficient as the typical incandescent lamp.

Fluorescent Lamps

Fluorescent lamps, as shown in Figure 3, are generally constructed in a cylindrical soda lime glass envelope with a length to diameter ratio significant greater than one. Most contain thermionic electrodes at each end to generate an electric field that establishes and sustains the discharge. The lamps are filled with about 2 Torr of Argon and/or another rare gas, and a few mg of liquid mercury, while the inside surface of the tube is coated with a phosphor that converts UV radiation generated by excited mercury atoms into visible light.

As with any other low pressure gas discharge, various discharge regions are aligned along the axis of the tube:

- Aston Dark Space
- Cathode Glow
- Cathode Dark Space
- Negative Glow
- Faraday Dark Space
- Positive Column
- Anode Glow
- Anode Dark Space

The axial lengths of all these regions, except for the positive column, are fixed by the properties of the cathode and anode, the fill gas type and pressure, and the discharge current. The length of the positive column is variable and fills the space between the Faraday Dark Space and the Anode Glow. The cathode and anode regions of the discharge have a total axial length of about twice the diameter of the discharge tube. Therefore, if the length-to-diameter ratio of the discharge tube is substantially greater than one, the majority of the tube is filled by the Positive Column, as shown in Figure 3. The positive column is the part of the discharge that creates the UV photons that excite the phosphor and cause it to generate visible light. The remainder of this discussion will therefore focus on the positive column.

The basic operation of a fluorescent lamp positive is shown in Figure 4. Free electrons are accelerated by the electric field established between the two electrodes. These electrons collide with both the mercury and rare gas atoms, but selectively excite and ionize mercury atoms, since the excitation and ionization energies of mercury are significantly lower than that of any rare gas used in the lamps. Ionization of mercury creates more free electrons to sustain the discharge. The gas density in the lamp is low enough to require that most of the recombination of mercury ions and free electrons takes place on the wall of the tube, since this recombination requires a three-body collision. The excited mercury atoms generate both UV and visible photons, and the UV photons generate visible light once they reach the phosphor coated on the inner surface of the lamp.

Figure 3 Linear fluorescent lamp. © 2005 Roberts Research and Consulting, Inc. Published by Elsevier Ltd. All rights reserved.

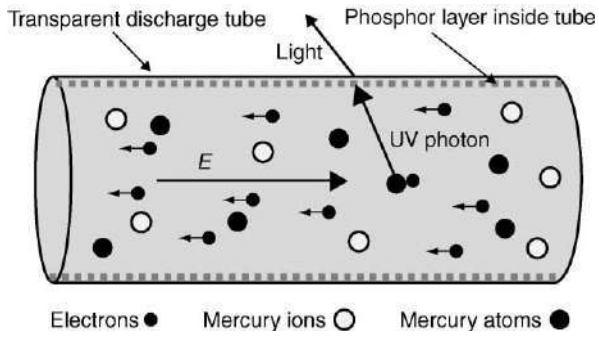

Figure 4 Operation of fluorescent lamp positive column. © 2005 Roberts Research and Consulting, Inc. Published by Elsevier Ltd. All rights reserved.

The high efficiency of fluorescent lamps is directly related to the fact that 60% to 65% of the electric power consumed by the positive column of the fluorescent lamp discharge is converted into 245 nm UV photons as a result of the resonance transition from the 6^3P_1 state of mercury at 4.86 eV to the 6^1S_0 ground state. The resonance transition from the 6^1P_1 state of mercury generates 185 nm UV photons that carry about 10% of the energy. About 3% of the positive column energy ends up in mercury visible lines at 405 nm, 436 nm, 546 nm, 577 nm, and 579 nm.

The 245 nm and 185 nm photons generated by the excited mercury atoms travel only a fraction of a mm before they are absorbed by a ground-state mercury atom. However, due to the relative absence of other de-excitation mechanisms, a large fraction of these absorbed photons are re-emitted by the absorbing atoms, and most of the UV photons generated in the core of the discharge eventually reach the phosphor coated interior wall of the discharge tube where they are converted into visible light (and waste heat).

The phosphors used in modern fluorescent lamps are a blend of three rare-earth phosphors:

- Blue: $BaMg_2Al_{16}O_{27}:Eu^{2+}$
- Green: $CeMgAl_{11}O_{19}(Ce^{3+}):Tb^{3+}$
- Red: $Y_2O_3:Eu^{3+}$

Rare earth phosphors provide a combination of high efficiency, high color rendition index, and high resistance to degradation that far exceeds the performance of the halophosphate phosphors used in older generations of fluorescent lamps.

However, due to the Stokes Shift, there is a substantial energy loss converting the UV photons produced by excited mercury atoms into visible light, even using a 'perfect' phosphor that has quantum efficiency of 1.0. The average energy of a visible photon is 2.43 eV, using the 380 nm to 780 nm limits for visible radiation discussed at the start of this article. When one 245 nm photon that has 4.885 eV of energy is converted to one visible photon with an average energy of 2.43 eV, a bit over 50% of the photon's energy is lost. The situation is worse for the 185 nm photons, which each have 6.707 eV of energy. They lose 64% of their energy when they are converted to visible photons, so it is good that the fluorescent lamp discharge generates six to seven 245 nm photons for each 185 nm photon. The Stokes shift is responsible for the largest single energy loss mechanism in fluorescent lamps. Work is underway to develop phosphors with a quantum efficiency greater than 1.0, so less energy will be lost in the UV-to-visible light conversion process, but no practical phosphors have been developed to date.

The mercury pressure for optimum UV production is about 6×10^{-3} Torr, which is the vapor pressure of mercury over a liquid pool at 40 °C. At higher mercury pressure there is too much absorption by ground state mercury atoms, and at lower pressures there is an insufficient number of mercury atoms for excitation by the electrons. The diameter of most linear lamps is chosen so that the coolest spot on the interior of the lamp wall will be at 40 °C when the lamp is operated at its design power in an ambient temperature of 25 °C. Usually this 'cold spot' will be near the middle of the lamp, but in some higher power linear lamps the wall temperature is well over 40 °C. In these lamps a cold spot is often created at the end of the lamp by use of a longer electrode mount and a heat shield mounted between the electrode and the end of the lamp. Obviously, the performance of fluorescent lamps designed in this manner is highly sensitive to ambient temperature. In some smaller, high-power density lamps, most commonly compact fluorescent lamps (CFLs), it is not possible to create a region with a temperature as low as 40 °C. In these lamps the mercury pressure is controlled by amalgamating the mercury with Pb, Bi, or Sn, creating a compound that both reduces the mercury vapor pressure at any given temperature and also reduces the variability of mercury vapor pressure with temperature. These mercury amalgams have the secondary advantage of making the fluorescent lamp less sensitive to ambient temperature.

Mercury, which is the key ingredient in fluorescent lamps, is considered to be a hazardous material by regulatory agencies in many countries. At least as early as the mid-1980s researchers began to study ways to reduce or eliminate mercury from fluorescent lamps.

Development of a mercury-free fluorescent lamp is very difficult and the only mercury-free fluorescent lamp that exists, as of mid-2003 has an efficacy of about 10 l mW^{-1}, one-tenth the efficacy of the best mercury-based fluorescent lamp (Osram Sylvania LINEX® Linear Excimer Lamp System, Model LINEX A3-10W40). There are three key problems that have to be solved to develop an efficient mercury-free fluorescent lamp system:

(i) Identify a nontoxic material that generates UV photons and has its optimum gas density at a convenient and safe temperature. The most promising candidate is xenon, used as either an atomic radiator or an excimer.

(ii) Develop a phosphor that produces high-quality white light with high efficiency when excited by UV photons from the chosen mercury-replacement material. This has proven a major challenge. The resonance line of atomic Xe has a wavelength of 147 nm, while the output from a xenon excimer discharge is at 172 nm. The Stokes shift loss converting each of these to visible light is 71% for atomic xenon and 66% for xenon excimers. To reduce the Stokes shift loss, a quantum-splitting phosphor, one that generates more than one visible photon for each UV photon, is needed. Most of the effort to develop mercury-free fluorescent lamps has been concentrated in this area. No viable phosphor has been developed, but work continues.

(iii) Develop a 'reservoir' for xenon or the other working gas. The mass of 'active' gas needed for a fluorescent lamp is small; a 4-foot, 1.5-inch diameter lamp needs 0.1 mg of mercury in the gas phase to operate properly. However, during life, some of this gas is consumed through physical attachment or chemical reaction and must be replaced. The drop of liquid mercury or small mass of mercury amalgam used in conventional fluorescent lamps provides a reservoir that maintains the mercury gas pressure in the presence of these consumption mechanisms.

In contrast to the failure to develop a mercury-free fluorescent lamp, there has been great success in

reducing the amount of mercury used in each lamp. A typical 4-foot, 1.5-inch diameter fluorescent lamp produced in the 1970s would have 50 mg to 100 mg of mercury. This high amount of mercury was mostly the result of the type of manufacturing equipment used, but mercury consumption required that each lamp had 15 mg to 20 mg of mercury in order to survive for its rated 20 000-hour life. The development of coatings and other materials that reduce mercury consumption, when combined with new manufacturing technology, has allowed the amount of mercury used in modern 4-foot fluorescent lamps to be reduced to 3 mg to 5 mg, and further reductions are possible with improved technology.

As stated above, the length-to-diameter ratio of most fluorescent lamps is significantly greater than one. The reason for this is that there is a 10 to 15 volt drop at the electrodes, and the power lost in these regions does not produce a significant amount of light. A large length-to-diameter ratio increases the voltage and power of the positive column, where the bulk of the UV photons are generated, relative to the power lost in the electrode regions. When the energy crunch hit in the mid-1970s, there was strong interest in developing a fluorescent lamp that could replace the inefficient incandescent lamp. However, the need for high length-to-diameter ratios meant that it would not be practical to make a fluorescent lamp shaped like an incandescent lamp. This need created the technology to manufacture CFLs which are constructed of long tubes bent or coiled into shapes that allow them to fit into spaces occupied by incandescent lamps.

Fluorescent lamps and other discharge lamps have a negative incremental voltage vs. current (V/I) characteristic, which means that their operating voltage decreases as the operating current increases. Due to this, such discharges need to be operated from a current limited power source, commonly known as a 'ballast' in the lighting industry. Ballasts were originally simple series inductors and lamps operated at the power line or mains frequency, though many line frequency ballasts are now considerable more complex than an inductor. In the 1940s, it was discovered that the efficacy of fluorescent lamps could be increased by operation at frequencies of a few kHz. It was not until the 1970s, however, until the cost of power electronics came down to the point where it was possible to make affordable high frequency electronic ballasts. Electronic ballasts operating in the 20 kHz to 100 kHz range are now used with almost all CFLs and in most commercial applications using linear fluorescent lamps.

The desire for fluorescent lamps shaped like incandescent lamps also sparked the development of electrodeless fluorescent lamps. Since the need for high length-to-diameter ratios is driven by the losses at the lamp electrodes, researchers realized they could make fluorescent lamps in short, wide shapes if they eliminated the electrodes. As shown in Figure 5, an electrodeless lamp is just like a transformer, except it has a single-turn discharge secondary instead of a multiturn copper or aluminum secondary. An open core version of an electrodeless fluorescent lamp with integral ballast is shown in Figure 6. The electric field is created by a time-varying magnetic field. Because of the relative high voltage-per-turn imposed by the discharge secondary, these lamps operate only at high frequency. There are three electrodeless fluorescent lamps on the market. The closed core design

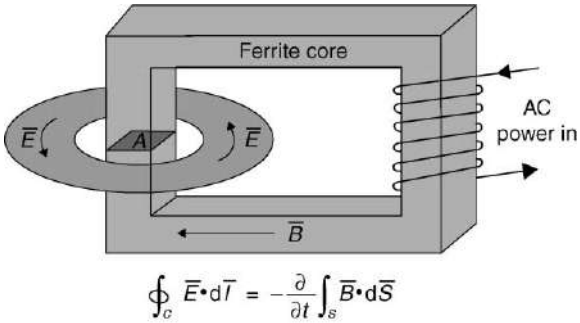

Figure 5 Basic electrodeless fluorescent lamp with closed ferrite core. © 2005 Roberts Research and Consulting, Inc. Published by Elsevier Ltd. All rights reserved.

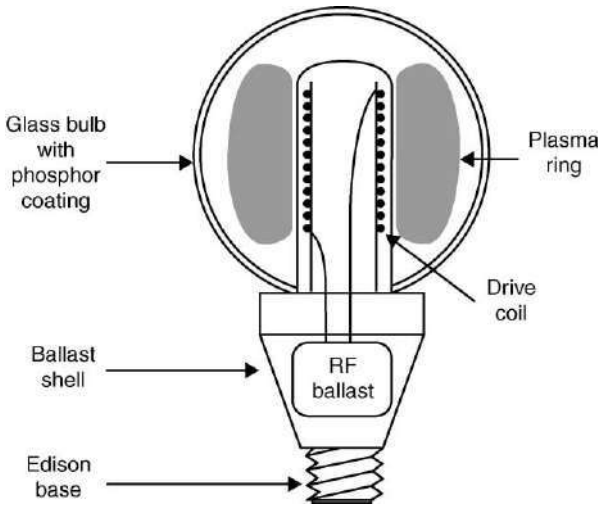

Figure 6 Open core electrodeless lamp with integral ballast. © 2005 Roberts Research and Consulting, Inc. Published by Elsevier Ltd. All rights reserved.

shown in Figure 7 operates near 250 kHz with an external ballast, while the open core design shown in Figure 8 operates near 2.5 MHz. Figure 9 shows an open core design with an integral ballast that also operates near 2.5 MHz. Some people have suggested that electrodeless fluorescent lamps are fundamentally different from normal fluorescent lamps and the high frequency field drives the phosphor to emit light. This is incorrect. Other than the manner in which the electric field is created, and the new shapes this allows, electrodeless fluorescent lamps operate in the same manner as normal fluorescent lamps.

Due to the strong dependence of fluorescent performance on mercury temperature, these lamps are not useful as flux standards. However, the five mercury visible lines listed above, plus the 365 nm line of mercury, make the ordinary fluorescent lamp an ideal source for wavelength calibration of spectrometers. Germicidal lamps, which do not use phosphor but are constructed in quartz envelopes and use the same type of mercury-rare gas discharge as fluorescent lamps, are a convenient source of 254 nm radiation that can be used to calibrate UV spectrometers. Care should be taken when using these lamps, as the 254 nm radiation can damage human eyes and skin.

High Intensity Discharge Lamps

If the mercury pressure in a discharge lamp is raised from 6×10^{-3} Torr to an atmosphere or

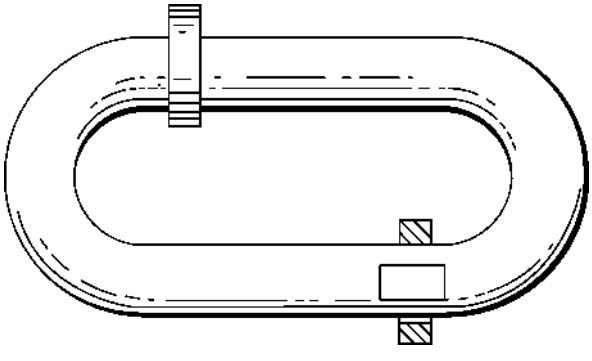

Figure 7 Closed core electrodeless fluorescent lamp, Osram Sylvania Icetron/Endura Lamp, from US Patent 5,834,905.

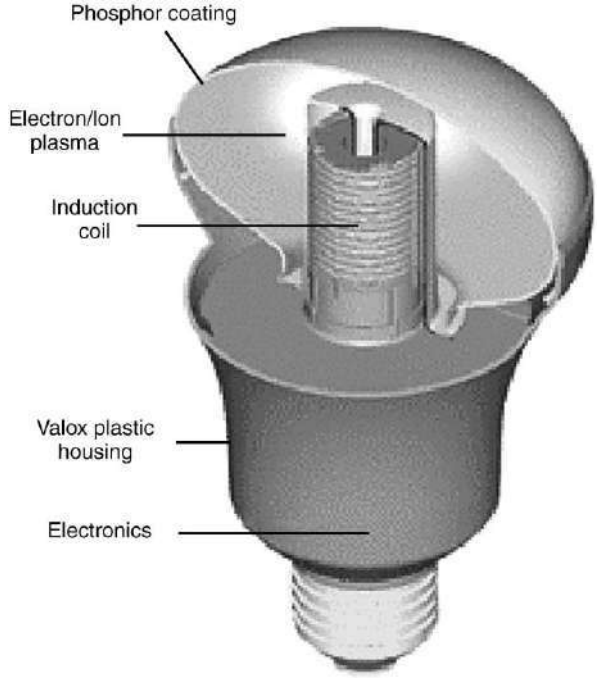

Figure 9 GE Genura® open core electrodeless fluorescent lamp with integral ballast. Courtesy of General Electric Company.

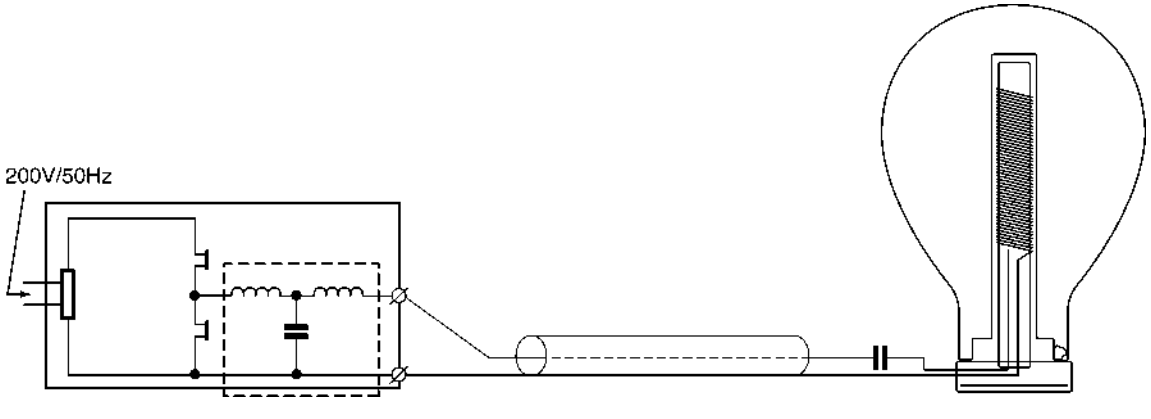

Figure 8 Open core electrodeless fluorescent lamp with separate ballast, Philips QL lamp, from US Patent 6,373,198 B1.

greater, the 245 nm and 185 nm resonance lines are trapped via absorption by ground-state mercury atoms and the radiation output from the discharge shifts to the five optically thin visible lines listed in the discussion of fluorescent lamps, plus the 365 nm near-UV line and a number of lines near 300 nm. The result is a high pressure mercury lamp that operates with an efficacy of about 55 $\mathrm{l\,m\,W^{-1}}$ and has a CRI of about 50. The high-mercury vapor pressure requires a wall temperature of at least 360 °C, which in turn requires that the arc tube be constructed from quartz and operated at high power density. A 400-watt high-pressure mercury arc tube has a diameter of 22 mm and length of about 100 mm. In contrast, a 32-watt fluorescent lamp has a diameter of 25.4 mm and a length of about 1200 mm. This accounts for the classification of these high pressure lamps as high intensity discharge (HID) lamps. Due to the high temperature of the arc tube, most HID lamps are constructed with the arc tube mounted inside a hard glass outer jacket that protects both the user and the arc tube. The high-pressure mercury lamp is a direct visible radiator, but the arc tube does generate a small amount of radiation in the 254 nm to 365 nm region, that radiation is trapped by the outer glass jacket. Some high-pressure mercury lamps have a phosphor coating on the inside of the outer jacket to convert this UV radiation into visible radiation, mostly in regions of the spectrum where the high-pressure mercury arc is deficient in radiation, thus improving the color of the lamp.

Most high-pressure mercury lamps operate at a mercury pressure of about four atmospheres. Because their spectrum is dominated by mercury line emission, the output of these lamps is deficient in the red portion of the spectrum and the lamps have poor color rendition. In 1992, Philips received a US patent for a new type of ultra-high pressure mercury lamp that they call UHP for 'Ultra High Performance'. This lamp operates at a mercury pressure of 200 atmospheres or more. The output spectrum is characterized by a significant amount of continuum radiation from quasi-molecular states of mercury and this continuum includes energy in the red portion of the spectrum that is missing from conventional high pressure mercury lamps. UHP lamps also include a small amount of halogen, such as bromine, which creates a tungsten-halogen cycle similar to that discussed earlier in relation to tungsten-halogen incandescent lamps. During lamp operation, tungsten is sputtered and evaporated from the electrodes. In the absence of the halogen fill, this tungsten rapidly deposits on the walls, leading to very short lamp life. The halogen combines with most of the tungsten before it has a chance to reach the wall of the arc tube. The halogen will also combine with any tungsten that has deposited on the wall of the arc tube. The tungsten-halogen compound formed is gaseous at the operating temperature of the lamp and will diffuse back toward the core of the discharge where the high temperature of the arc dissociates the tungsten-halogen compound. The tungsten is deposited back on the electrodes, thereby greatly extending the life of the lamp. Due to their very high pressures, UHP lamps are constructed in small discharge tubes and have a very short discharge length, typically 1 to 2 mm, making these lamps a virtual point source. Due to their excellent spectral characteristics and small source size, UHP lamps have found application in various projection applications, including LCD-based electronic projectors.

The performance of high-pressure mercury lamps can be substantially improved by adding metals such as sodium and scandium to the high-pressure mercury arc. The sodium emits near the peak of the eye sensitivity curve, providing high efficacy, while the scandium has many transitions in the visible, which fill out the spectrum and provide high CRI. If metallic elements are added directly to the high-pressure mercury arc they would deposit on or react with the wall of the arc tube. To prevent this they are added as iodides, such as NaI or ScI_3. This creates a regenerative cycle in which the NaI and ScI_3 are heated and dissociated in the 5000 K core of the arc, allowing the free metals to radiate. As the metals diffuse toward the cooler sections of the arc near the walls, they recombine with the iodine to reform the iodides. These metal halide lamps have produce light with an efficacy of about 90 $\mathrm{l\,m\,W^{-1}}$ to 100 $\mathrm{l\,m\,W^{-1}}$ and a CRI of about 65 to 75.

The high-pressure sodium (HPS) lamp is an HID lamp that uses hot sodium vapor to generate a golden white light. The lamp uses a mixture of sodium, mercury, and rare gas such as xenon. The core temperature of the sodium-mercury-xenon discharge is typically about 4000 K, while the temperature at the edge of the discharge and the walls of the ceramic arc tube is typically about 1500 K. At low pressure, emission from sodium is dominated by the resonance lines at 589.0 and 589.6 nm, so the light is virtually monochromatic. However, in the HPS lamp, the sodium pressure is typically 75 Torr, which is high enough to almost completely trap emission from the resonance lines, forcing radiant energy to escape in the wings of

these lines, thereby increasing the color quality of the light. Since sodium will attack quartz, this mixture is contained in a translucent alumina (aluminum oxide) ceramic arc tube. As with other HID lamps, the alumina arc tube is contained in a hard glass outer bulb. HPS lamps are the most efficient generators of 'white light' available. A typical 400 watt HPS lamp produces light with an efficacy of 125 $\mathrm{l\,m\,W^{-1}}$, this high efficacy being due to the fact that the emission matches the peak of the eye sensitivity curve. The light has a CRI of only 22; however, the HPS lamp is widely used for roadway and other outdoor lighting applications.

Solid State Sources

The newest incoherent sources to be used for general purpose lighting are solid state devices: light emitting diodes (LEDs), and organic light emitting diodes (OLEDs). LEDs are very similar to laser diodes but without the two mirrors used to form the laser cavity in laser diods. Photons are generated as a result of electron-hole recombination as electrons move across the energy gap of a biased p-n junction. By appropriate selection of materials, photons can be created at visible wavelengths. Low brightness infrared and red LEDs have been available for quite some time. It was not until the mid-1980s that high brightness red LEDs were developed. These were followed by increasingly more challenging yellow, green, and then blue high brightness LEDs, developed in 1994 by Shuji Nakamura, of Nichia Chemical Industries. Nakamura added phosphor to his high brightness blue LEDs in 1996 and developed the first 'white' LED, thus triggering the current level of excitement concerning the use of LEDs to replace incandescent and discharge lamps. Monochromatic high brightness LEDs have already replaced color-filtered incandescent lamps in many applications, such as traffic lights, vehicle signal lights, and signage. In these applications, LEDs are the clear winner in life and efficacy. However, when compared to unfiltered sources, the best white LEDs are about as efficient as tungsten halogen incandescent lamps and far less efficient than fluorescent and metal halide lamps. They also cost far more per lumen than even the most expensive discharge lamp. LEDs have high internal quantum efficiency, so the main technical challenges are to fabricate devices to minimize internal light trapping in order to improve device efficiency and reduce the cost per lumen produced.

Organic light emitting diodes (OLEDs) generate light by the passage of electric current through special organic molecules, using a process known as electrophosphorescence. The electrophosphorescent material is usually coated on a transparent and perhaps flexible substrate, and overcoated with a transparent conductor. OLEDs currently have lower efficacy and shorter life than LEDs, but they also have lower manufacturing costs. If LEDs are the 'new' point sources that replace incandescent lamps and perhaps CFLs, then OLEDs are the 'new' diffuse sources that some say will replace liner fluorescent lamps.

See also

Light Emitting Diodes.

Further Reading

Cobine JD (1958) *Gaseous Conductors: Theory and Engineering Applications*. New York: Dover Publications.

de Groot JJ and van Vliet JAJM (1986) *The High Pressure Sodium Lamp*. Philips Technical Library. Deventer–Antwerprn: MacMillian Education.

Elenbass W (1951) *The High Pressure Mercury Vapor Discharge*. Amsterdam: North-Holland Publishing Company.

Fisher HE (1996) *High Pressure Mercury Discharge Lamp*, U.S. Patent, 5,497,049.

Fisher HE and Hörster H (1998) *High-Pressure Mercury-Vapor Discharge Lamp*, U.S. Patent 5,109,181.

Godyak VA (2002) Bright Idea: Radio-frequency light sources. *IEEE Industry Applications Magazine* 8: 42–49.

Hirsh MN and Oskam HJ (1978) *Gaseous Electronics, Volume 1, Electrical Discharges*. New York: Academic Press.

Mitchel AGC and Zemansky MW (1971) *Resonance Radiation and Excited Atoms*, 2nd edn. London: Cambridge University Press.

Nichia's Shuji Nakamura: Dream of the Blue Laser Diode, ScienceWatch, January/February 2000. Available at: http://www.sciencewatch.com/jan-feb2000/sw_jan-feb2000_page3.htm

Pekarski P, Hechtfischer U, Heusler G, Koerber A, Moonch H, Retsch JP, Weichmann U and Kitz A (2003) *UHP Lamps for Projection Systems*, XXVI International Conference on Phenomena in Ionized Gasses, Greifswald, Germany. Available at: http://www.icpig.uni-greifswald.de/proceedings/data/Pekarski_1

Raizer YP (1991) *Gas Discharge Physics*. Berlin: Springer-Verlag.

Rea MS (2000) *IESNA Lighting Handbook*, 9th edn. New York: Illuminating Engineering Society of North America.

Roberts VD (2003) *Radio-Frequency Lighting*. McGraw-Hill Yearbook of Science & Technology, pp. 349–352. New York: The McGraw-Hill Companies, Inc.
Schewe PF and Stein B (1999) *Physics News Update*, Number 437 (Story #2), The American Institute of Physics Bulletin of Physics News. Available at: http://www.aip.org/enews/physnews/1999/split/pnu437-2.htm
Waymouth JF (1971) *Electric Discharge Lamps*. Cambridge, Massachusetts: MIT Press.
Wharmby DO (1993) Electrodeless Lamps for Lighting: A Review. *IEE Proceedings-A* 140(6): 465–473.

Synchrotrons

R Clarke, University of Michigan, Ann Arbor, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Since their discovery by Röntgen in 1895, X-rays have been used in a wide variety of applications ranging from medical and industrial radiography to crystallographic studies of the structure of biomolecules. The short wavelength ($\lambda \sim 0.1$ nm) and high penetrating power of hard X-rays makes them an ideal probe of the internal structure and composition of condensed matter. Indeed, Watson and Crick's elucidation of the double helix structure of DNA in the early 1950s, based on the X-ray diffraction patterns of Rosalind Franklin, has revolutionized biology and provided an atomistic view of the stuff of life.

The quest for ever higher spatial resolution, and the desire to obtain high-quality X-ray data on shorter time-scales, have both been prime motivations for the development of high intensity hard X-ray sources since the mid-1970s. This evolutionary trend has led to the design and construction of large synchrotron radiation facilities, serving many thousands of users and enabling forefront experimental studies across a broad range of disciplines. One such example of a national facility, dedicated to high-brightness X-ray synchrotron radiation, is the Advanced Photon Source (APS), located at Argonne National Laboratory in the United States (Figure 1). Facilities of a similar scope and purpose have also been constructed in Europe and Japan, and a number of new high brightness sources are currently under construction in several countries around the world (see Table 1).

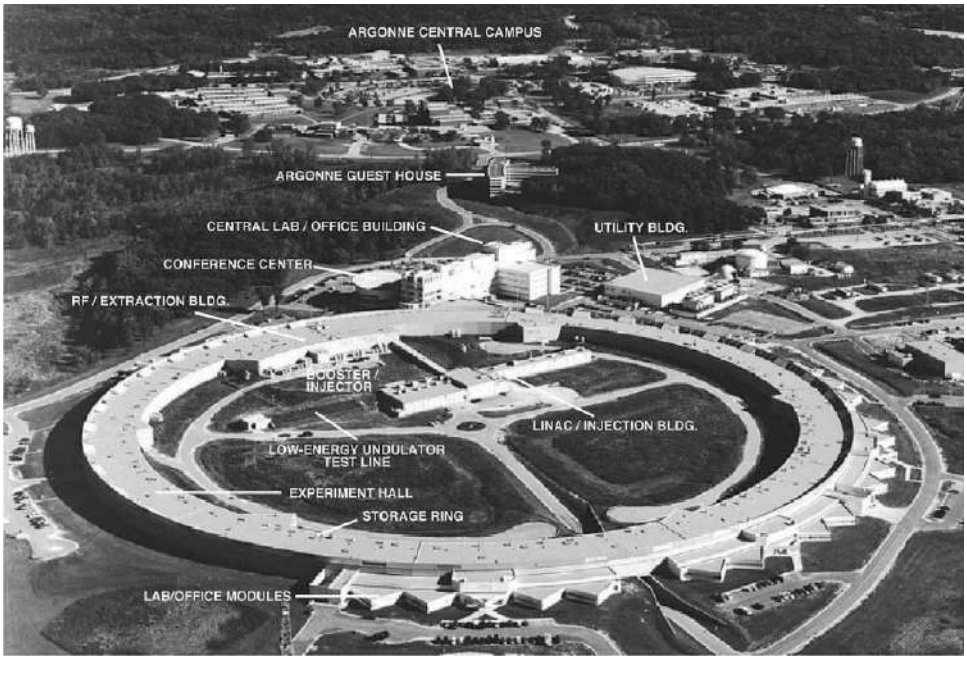

Figure 1 Aerial view of the Advanced Photon Source, Argonne National Laboratory. Courtesy of Advanced Photon Source.

Table 1 Third-generation synchrotron radiation sources

Location	*Facility*	*Electron energy (GeV)*
Grenoble, France	European Synchrotron Radiation Facility (ESRF)	6
Argonne, Illinois, USA	Advanced Photon Source (APS)	7
Nishi Harima, Japan	8-GeV Super Photon Ring (SPring-8)	8
Berkeley, CA, USA	Advanced Light Source (ALS)	1.5–1.9
Canberra, Australia	Boomerang[a]	3
Didcot, UK	DIAMOND[a]	3
Orsay, France	SOLEIL (LURE)	2.75
Karlsruhe, Germany	Angströmquelle-Karlsruhe (ANKA)	2.5
Hamburg, Germany	PETRA III[b]	6
Indore, India	INDUS II[a]	2.5
Villigen, Switzerland	Swiss Light Source (SLS)	2.4
Saskatoon, Canada	Canadian Light Source[a]	2.9
Stanford, CA, USA	SPEARS	3
Upton, New York, USA	NSLS II[b]	2.5–2.8
Trieste, Italy	Elettra	2.0–2.4

[a]Under construction as of March 2004.
[b]Planning stage.

In this article we will draw on examples from APS to illustrate the characteristics and performance of the current generation of synchrotron sources in the hard X-ray region.

From the earliest days of particle physics studies in the 1930s and 40s, it was realized that copious amounts of X-rays could be produced by accelerating charged particles in a curved trajectory, as in a cyclotron or a synchrotron. For particle physics such radiation is highly detrimental because it limits the ultimate energy of particle accelerators and forces their circumference to become inordinately large in order to reduce radiation losses.

Originally considered to be a major source of concern in particle accelerator design, synchrotron radiation soon began to emerge as a boon to X-ray science. This came about in the early 1970s after the capability was developed to produce and store significant circulating currents of high-energy charged particles (electrons or positrons) for colliding beam experiments. At that time X-ray scientists were tolerated as 'parasitic' experimenters at such storage rings, taking X-ray radiation from the bending magnets which maintained the stored current in an overall circular trajectory. The PETRA (Positron–Electron Tandem Ring Accelerator) facility at HASYLAB/DESY in Hamburg, Germany, and the Cornell Electron Storage Ring (CESR)/Cornell High-Energy Synchrotron Source (CHESS) at Cornell University, Ithaca, New York, are examples of accelerator facilities that were primarily intended for particle physics experiments and then became partly dedicated to X-ray production. Such shared synchrotron facilities are referred to as 'first-generation sources'.

The first generation sources demonstrated several attractive features of synchrotron radiation, including a high degree of collimation in the vertical direction with a divergence half-angle given approximately by the relativistic factor $\gamma^{-1} = m_0c^2/E$, where m_0c^2 is the rest-mass energy (0.51 MeV) of the electron (or positron) and E is the energy of the stored particle beam. For a bending magnet source with $E = 2–3$ GeV the beam would have a vertical collimation of ~ 0.3 mrad (17 mdeg). This would produce an X-ray beam with a vertical extent of ~ 3 mm at a distance of 10 m from the bending magnet source. The horizontal fan of bend magnet radiation is broad and is usually limited by a mask or a slit to a few mrad. The most useful characteristic of bend magnet radiation is its large bandwidth whereby the X-ray spectrum is continuous and increasing in intensity (see Figure 2) until some critical energy above which the intensity begins to fall off. Bending magnet radiation is therefore useful for X-ray measurements requiring a 'white' beam, for example, Laue X-ray diffraction, energy dispersive diffraction, and Extended X-ray Absorption Fine Structure (EXAFS) analysis.

Beginning in the 1980s, several second-generation storage rings were constructed in various locations around the world. The design of these facilities for the first time was optimized for the production of bright beams of X-ray and XUV (soft X-ray to ultraviolet) and their operation was dedicated exclusively to synchrotron radiation science. The increase in brightness provided by these new storage ring sources, compared to standard laboratory X-ray sources (e.g., rotating anode generators), was dramatic (5 or 6 orders of magnitude) and this opened up many new experiments that were not previously feasible. However, it was soon realized that the capabilities of such sources could be extended to

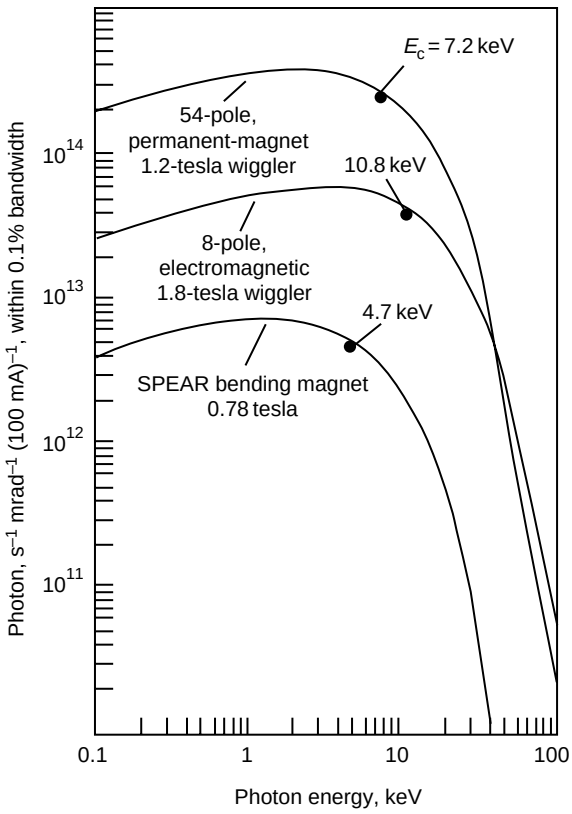

Figure 2 Comparison of bending magnet and wiggler sources. With permission from Stanford Synchrotron Radiation Laboratory, Stanford Linear Accelerator Center, Menlo Park, CA.

even higher levels of intensity and to significantly higher X-ray energies.

One of the limitations of bending magnet sources relates to the X-ray critical energy alluded to above:

$$E_c = 3hc\gamma^3/4\pi\rho \qquad [1]$$

where ρ is the radius of curvature imposed on the particle beam trajectory by the bending magnet. For a given particle acceleration energy, E, eqn [1] leads to a critical X-ray energy given by:

$$E_c\,(\text{keV}) = 2.2\,E^3\,(\text{GeV})/\rho\,(\text{m}) \qquad [2]$$

For a typical second-generation machine with $E = 2.5$ GeV and $\rho = 7$ m (limited by the magnetic field generated by the dipole electromagnet), the critical energy is only about 5 keV. There is useable flux well above this energy, but X-ray photons become scarce in the energy range above 10 keV for a bending magnet source with these characteristics. Since E is effectively fixed by the design parameters of the storage ring, the only possibility to achieve higher photon energies is to install additional magnetic devices which will introduce perturbations of the trajectory with effectively smaller radius of curvature in the denominator of eqn [2]. Such a device, known as a 'wiggler', shifts the effective critical X-ray energy to higher values in proportion to the strength of the magnetic field B (see Figure 2). Using wigglers based on superconducting magnets ($B \sim 5$ Tesla), E_c can be raised to >20 keV, a value which would require accelerator energies of more than 6 GeV using conventional bending magnets alone. In addition to shifting the output spectrum to higher energies, it is possible to enhance the intensity at a given energy by producing multiple wiggles with no net displacement of the particle trajectory. Multipole wigglers are able to produce significant increases in flux in the hard X-ray range. The 27-pole hybrid wiggler installed on beamline X25 at the National Synchrotron Light Source, a second generation facility at Brookhaven National Laboratory, Upton, New York, is one of the most successful designs of its kind, providing $\sim 6 \times 10^{14}$ photons/sec per 1% bandwidth at 8 keV. When focused with a toroidal X-ray mirror, this beam is about an order of magnitude brighter than the standard bending magnet sources at NSLS. For reasons that will become clear later in this article, the radiation from multiple wiggles in such a device adds incoherently.

Insertion Devices

The idea of using wigglers to enhance the characteristics of synchrotron radiation sources has led to the design and construction of storage rings with many straight sections to accommodate wigglers and other types of 'insertion devices'. Such facilities are referred to as third-generation sources. Coming on-line in the mid-1990s, several such sources are now in mature operation and are supporting a large international community of X-ray scientists.

One of the most important types of insertion device to emerge from the development of third-generation facilities is the undulator. Conceptually, the idea behind an undulator is similar to that of the wiggler, i.e., it provides additional tightly curved transverse displacements in the particle beam. There are two important differences: in an undulator the displacements are smaller and they are more numerous, being produced by a linear periodic array of many short alternating dipoles (Figure 3). Under these conditions the synchrotron radiation emitted from each 'undulation' of the particle beam trajectory adds coherently and therefore the peak intensity from such a device can be enhanced by a factor of N^2 where N is the number of dipole magnet pairs, or periods.

A simple theoretical description of the basic operation of an undulator can be discussed in terms of radiation from a charged particle moving at

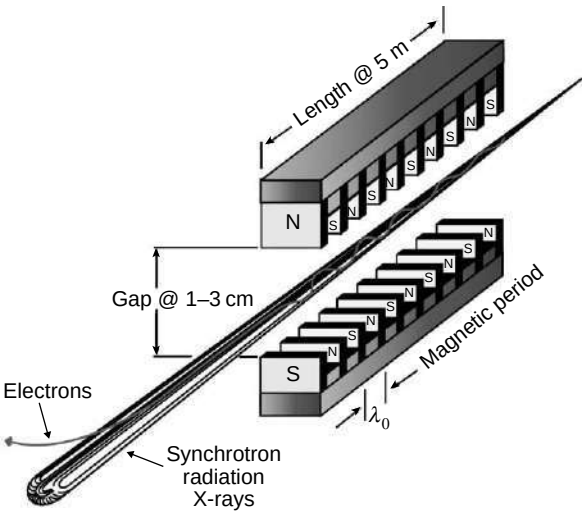

Figure 3 Schematic of undulator. Courtesy of Advanced Photon Source.

relativistic speeds through an alternating magnetic field. The magnetic field in the vertical direction (y) is taken to be periodic:

$$B_y = B_0 \sin[(2\pi/\lambda_u)z] \qquad [3]$$

where z is in the direction of average particle motion.

The transverse particle trajectory is described by:

$$x = (\lambda_u \phi_{max}/2\pi) \sin[(2\pi/\lambda_u)z] \qquad [4]$$

where λ_u is the magnet period, and ϕ_{max} is the maximum particle deflection angle:

$$\phi_{max} = (eB_0\lambda_u c)/2\pi m_0 c^2 \gamma = K/\gamma \qquad [5]$$

where

$$K = 93.4B_0 \text{ (Tesla)}\lambda_u \text{ (meters)} \qquad [6]$$

The field strength parameter K serves as a useful quantity for discussing the characteristics of various insertion device designs. For $K \leq 1$, which defines the undulator regime, the transverse angular deflections are less than or on the order of the angular spread of the emitted radiation and in this case the radiation from each of the deflections adds coherently. This is distinguished from the case of a multipole wiggler where $K \gg 1$ leading to incoherent superposition.

Most of the interesting features of undulator radiation are revealed by considering the trajectory of the particle in a Lorentz frame moving at speed v^* with respect to the laboratory frame. In this frame the particle will execute oscillatory motion with the period of oscillations T' given by the Lorentz transformation:

$$T' = \gamma^*(T - \beta^*\lambda_u/c) = \lambda_u/(v^*\gamma^*) \qquad [7]$$

with a corresponding oscillation frequency $\omega' = 2\pi/T' = (2\pi c/\lambda_u)\gamma^*$.

For $K \ll 1$, the motion in the co-moving frame is purely transverse and nonrelativistic. However, for $K \sim 1$, the oscillatory motion in this frame can have a longitudinal component and in this case the transverse motion becomes nonharmonic with frequency components at odd multiples of the fundamental frequency ω'. This nonharmonic motion is a consequence of the fact that in the laboratory frame the particle is moving with constant speed $\approx c$ but its path length is slightly longer than λ_u, owing to the transverse motion. This constrains the motion in the co-moving frame to be nonharmonic.

Undulator Spectrum

It now remains to transform ω' into the laboratory frame so that we can obtain the spectral distribution of the output of the undulator. Invoking the Doppler transformation:

$$\omega = \gamma^*\omega'(1 + \beta^* \cos\theta') \qquad [8a]$$

$$\tan\theta = (\sin\theta')/[\gamma^*(\cos\theta' + \beta^*)] \qquad [8b]$$

where θ is the angle at which the radiation is observed:

$$\omega_n(\theta) = (4\pi cn\gamma^2)/(\lambda_u[1 + \tfrac{1}{2}K^2 + (\gamma\theta)^2]) \qquad [9a]$$

or in terms of wavelength:

$$\lambda_n(\theta) = (\lambda_u/2n\gamma^2)[1 + \tfrac{1}{2}K^2 + (\gamma\theta)^2] \qquad [9b]$$

Equation [9] is the central expression describing the spectral output of an undulator. Notice how the macroscopic period of the undulator λ_u is transformed into a microscopic wavelength $\lambda_n(\theta)$ by the relativistic factor γ^2. For third-generation hard X-ray sources with particle energies in the 6–8 GeV range, $\gamma \sim 14\,000$, taking cm scales down to the angstrom level.

Equation [9] describes a spectrum of monochromatic peaks, for on-axis radiation ($\theta = 0$), with wavelengths that increase with increasing K. Since $K \sim B_0\lambda_u$ (eqn [6]) any particular harmonic can be tuned either by changing B_0 or λ_u. In most undulator designs, λ_u is fixed by the construction of the permanent magnet array, so the tuning is conveniently accomplished by varying B_0. In practice this is done by changing the vertical distance between the poles

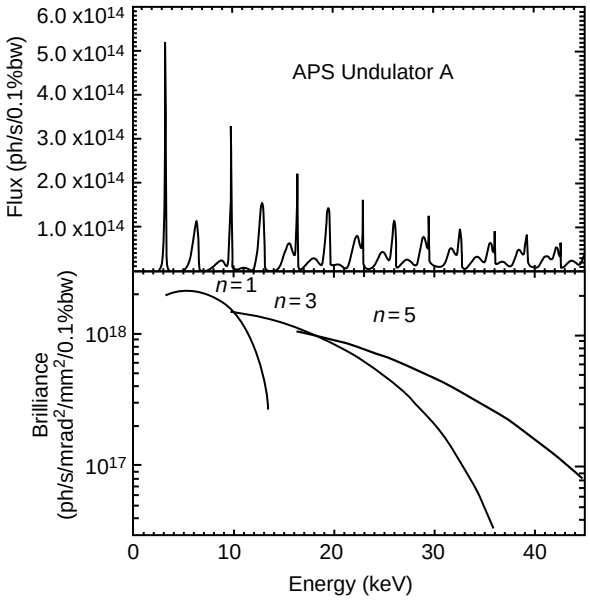

Figure 4 Undulator spectrum (upper panel) and tuning range for first, third and fifth harmonics (lower panel) of Advanced Photon Source Undulator A.

of the magnets (the 'undulator gap') similar to opening the jaws of a vise. Typical tuning ranges for the $n = 1$, 3, and 5 harmonics are illustrated in Figure 4.

Since the number of undulator oscillations of the particle beam is finite, the nth harmonic is not strictly monochromatic and has a spectral broadening of order $1/nN$, where N is the number of oscillator periods. In practice with N typically ~ 100, a bandwidth of $\sim 1\%$ is expected for the first harmonic. Additional broadening of a similar magnitude might also be expected from the θ-dependence in eqn [9], depending on the acceptance angle of the optics that is used in a particular experiment. This quasi-monochromatic radiation is sometimes referred to as 'pink beam'.

The storage ring energy of the APS was carefully chosen so that the tuning ranges of the different harmonics would overlap in energy, thus providing a continuous tuning of the X-ray energy over a wide range (Figure 4). Note also that the discrete spectral characteristics of the undulator can be deliberately smeared out to mimic the broadband (white beam) characteristics of a bend magnet (with consequent reduction in brightness). This is done simply by tilting the jaws of the undulator with respect to each other so that B_0 is made to vary along the length of the undulator. The ability to taper the undulator in this way is a standard feature of most designs.

The radiation from an undulator is sharply peaked in the forward direction in both the vertical and horizontal directions. The half-angle of the emission cone is given by:

$$\theta_{1/2} = (1/\gamma)[(1 + \tfrac{1}{2}K^2)/2Nn]^{1/2} = [\lambda_n/L]^{1/2} \qquad [10]$$

where L is length of the undulator. For a typical third-generation source undulator, such as the Type A undulator at the Advanced Photon Source, $L = 2.4$ m, giving $\theta_{1/2} \approx 6$ μrad for the first harmonic tuned to 12 keV. This corresponds to an X-ray beam size of ~ 180 μm at a distance 30 m from the source.

The very tight collimation of undulator radiation (eqn [10]) highlights the importance of distinguishing between the brightness and the 'brilliance' of the beam. The latter term takes into account the natural collimation of the beam, as reflected in its units of photons/sec/mm^2/mrad^2/0.1% bandwidth. Undulator radiation is well matched to the acceptance angle of ideal crystal monochromators and is $\sim 10\,000$ times more brilliant than the radiation from a bend magnet source. The high brilliance of undulator X-ray beams is extremely advantageous in X-ray imaging applications, X-ray microprobe measurements, and in experiments where the sample size is very small (e.g., high-pressure diamond anvil cell measurements).

In practice the emission cone of the source might be broadened a little when the undulator characteristics are convolved with those of the storage ring. The largest effect is in the horizontal direction where the photon beam divergence is dominated by the electron beam divergence. This is a consequence of the particular conditions under which the storage ring is operated in order to stabilize the particle beam orbit.

Equation [10] demonstrates that preserving the exquisite performance capabilities of an undulator insertion device places severe demands on the stability and precision of the storage ring orbit. An important measure of this precision is characterized by the 'emittance' of the particle beam defined in terms of the area of phase space (position and angular divergence) occupied by the stored beam. In general, storage rings are designed to achieve very low emittances and in current third-generation sources they are close to being diffraction limited. Table 2 summarizes the characteristics of the APS storage ring operated in 'high-emittance' and 'low-emittance' modes.

The large power densities resulting from the extreme brightness of undulator sources (up to 160 kW/mrad^2 for APS Undulator A) necessitates careful design of the X-ray optical components that are used to condition the beam. For example, if a double-bounce Silicon crystal pair is used as the first optical component to monochromatize the undulator

Table 2 Advanced photon source storage ring parameters

Parameter	*Low emittance*	*High emittance*
Storage ring energy, E	7.0 GeV	7.0 GeV
Storage ring current, I	100 mA	100 mA
Beam energy spread, $\delta E/E$	0.096%	0.096%
Horizontal emittance, ε_x	3.5×10^{-9} m-rad	7.7×10^{-9} m-rad
Vertical emittance, ε_y	3.5×10^{-11} m-rad	7.7×10^{-11} m-rad
Coupling constant	1%	1%
Horizontal beta function, β_x	14.4 m	16.1 m
Vertical beta function, β_y	4.0 m	4.3 m
Dispersion function, η_x	0.124 m	0.0 m
Horizontal beam size, σ_x	254 μm	351 μm
Vertical beam size, σ_y	12 μm	18 μm
Horizontal beam divergence, $\sigma_{x'}$	15.6 μrad	21.8 μrad
Vertical beam divergence, $\sigma_{y'}$	3.0 μrad	4.2 μrad

beam, the first crystal is normally cooled with liquid nitrogen to dissipate the heat. Alternately, high thermal conductivity diamond crystals are sometimes used as monochromators.

Helical Undulators

The use of undulators to produce extremely bright, tunable X-ray beams has enabled many new applications of X-ray science, as well as offering significant improvements to existing techniques in terms of quality of data and the time required to accumulate a high degree of statistical accuracy. One of the areas that has benefited significantly from novel undulator design is research on magnetic materials and other spin-dependent phenomena. It is often advantageous to probe such systems with circularly polarized X-rays, for example to study magnetic X-ray circular dichroism (MXCD). Early MXCD measurements utilized the left and right elliptical polarization of bending magnet radiation above and below the plane of the ring. Such measurements are difficult because the intensity falls off dramatically out of the plane. The problem to produce high-brightness beams of circularly polarized X-rays is effectively solved by designing an undulator which produces a transverse periodic helical magnetic field of constant magnitude. Several helical undulators are now in operation at third-generation sources. The soft X-ray region (500–1500 eV) is particularly interesting because it covers the L-edges of the 3d transition elements.

Coherence

While this article describes what would be nominally considered as 'incoherent' sources in the parlance of conventional optics, it should be pointed out that undulators are in fact excellent sources of coherent X-rays. This is by virtue of the high degree of collimation inherent in the emission of radiation from the undulations of a highly relativistic charged particle (eqn [10]). The spatial (transverse) coherence angle is given by:

$$\lambda_n/2d \qquad [11]$$

where d is the source size. This is the angle within which the beam is spatially coherent.

For APS undulator A, the source size is ~15 μm (vertical) × 240 μm (horizontal), giving a coherence angle in the vertical direction of 6.5 μrad and 0.4 μrad in the horizontal direction. The X-ray beam divergence is approximately 4 μrad (vertical) by 15 μrad (horizontal). This means that the beam has complete spatial coherence in the vertical direction and is coherent over about 3% of the horizontal fan. At a distance of 30 m from the source, an on-axis pinhole of ~12 micron diameter will be completely illuminated by coherent X-rays. Transversely coherent X-ray beams are finding novel applications in X-ray photon correlation spectroscopy, coherent X-ray scattering, and phase reconstruction (lensless) imaging.

Timing Structure

X-ray storage rings produce pulsed X-rays. In the Advanced Photon Source, one or more long trains of electron bunches (15 ns long) are accelerated by the linac to 325 MeV. The electron bunches are injected into a booster that ramps the single bunch from 325 MeV to 7 GeV in about 225 ms. At the end of the booster cycle, the bunch is extracted and injected into the storage ring. The bunch can be injected into any of 1296 'buckets' that are spaced by 2.842 ns, forming any arbitrary pattern after several injection cycles.

The APS storage ring rf frequency of 351.927 MHz determines the bucket spacing of 2.842 ns. The circumference of 1104 m gives a revolution frequency of 271.5 kHz which allows for 1296 buckets in 3683 ns. The storage ring is filled with a selected bunch pattern depending on the requirements for a particular operations cycle. The standard 'singlets' bunch pattern at APS is 102 mA of average current in a train of 23 bunches (4.43 mA per bunch) each spaced by 153 ns (54 buckets). There is a gap of 306 ns (108 buckets) in this pattern, useful as

a timing guide. The singlets pattern has a lower lifetime than patterns with less charge per bunch, due to natural internal scattering processes. A 'top-up' mode has been devised to overcome the problem of overall intensity fall-off due to the exponential decay of the stored current. In top-up mode, new bunches are injected into the storage ring every few minutes to maintain the stored current at 102 mA.

A second fill pattern, known as the 'hybrid' pattern, is again 102 mA but distributed in one bunch of 5 mA, plus 8 groups of 7 consecutive bunches (1.73 mA/bunch), spaced by 68 ns (24 buckets). The 8 groups are diametrically opposite the 5 mA bunch, allowing for a 1.5 μs gap on both sides of the 5 mA bunch. This hybrid mode permits a single bunch to be isolated using a mechanical chopper or a fast framing camera. Since the RMS width of the single bunch is ~36 ps, relatively fast time-resolved experiments can be performed stroboscopically using this pulsed X-ray beam. In another implementation of time-resolved experiments, the singlet pattern is used, isolating each bunch with a fast avalanche photodiode detector, or an X-ray streak camera. With the streak camera it is possible to achieve time resolutions of a few picoseconds. This approach has been used to study impulsive phonon excitations and ultrafast strain wave propagation in solids.

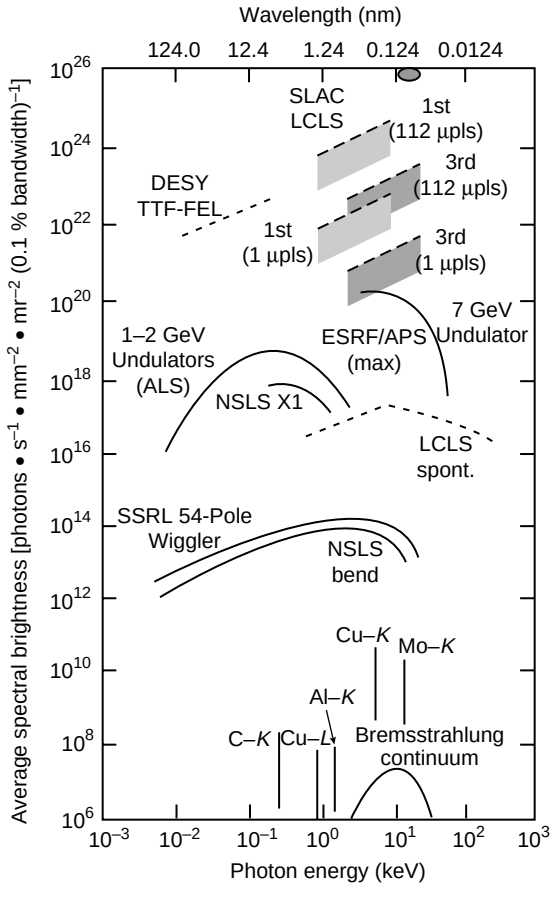

Figure 5 Comparison of spectral brilliance of X-ray sources. Courtesy of Linear Coherent Light Source, Stanford Linear Acceleration Center.

The Future

This article highlights current (third-generation) storage ring sources based on undulator-type insertion devices. These sources will provide high-brightness tunable X-rays in support of a wide range of experiments for many years to come. The brightness of these sources will increase steadily into the future as continual improvements are made, such as further reducing the emittance of the particle beam, improved undulator design, increasing the average storage ring current, and making changes to the storage ring lattice. These incremental improvements will be important to maintain the vitality and usefulness of national user facilities based on third-generation sources. New storage ring sources with designs and specifications optimized for low emittance and new types of undulators (e.g., in-vacuum undulators) will come online and will undoubtedly lead to new technical advances, supporting new and exciting science. It seems likely that linear accelerator (linac) based undulator sources will provide a promising route towards higher brilliance in the future. The superconducting Energy Recovery Linac (ERL) proposed by CHESS is an interesting concept which uses regeneration of the electron beam to overcome the high energy cost of single pass particle acceleration.

A new generation of X-ray sources (the 'fourth generation') is envisioned to take on the task of providing vastly brighter beams (>10 orders more peak brilliance than at present, see Figure 5) with much shorter pulses (~100 fs). These facilities will also be based on linear accelerators. Two such *Free Electron Laser* sources are currently being designed and should be operational towards 2010: the Linac Coherent Light Source (LCLS), to be built at the Stanford Linear Accelerator Center (SLAC), and the X-ray Free Electron Laser (XFEL) Source, to be constructed by the TESLA collaboration at HASYLAB, Hamburg.

See also

Coherence: Overview. **Physical Applications of Lasers:** Free Electron Lasers in Physics.

Further Reading

Brown G and Lavender W (1991) Synchrotron radiation spectra. In: Brown GS and Moncton DE (eds) *Handbook on Synchrotron Radiation*, vol. 3. Amsterdam: North Holland.

Duke PJ (1993) Synchrotron radiation sources. In: Michette AG and Buckley CJ (eds) *X-Ray Science and Technology*. Bristol: Institute of Physics Publishing Ltd.

Duke PJ (2000) *Synchrotron Radiation: Production and Properties*. Oxford, UK: OUP.

Mills DM (ed.) (2002) *Third Generation Hard X-ray Synchrotron Radiation Sources*. Berlin: Wiley, VCH.

INFORMATION PROCESSING

Contents

All-Optical Multiplexing/Demultiplexing
Coherent Analog Optical Processors
Free-Space Optical Computing
Incoherent Analog Optical Processors
Optical Bit-Serial Computing
Optical Digital Image Processing
Optical Neural Networks

All-Optical Multiplexing/Demultiplexing

Z Ghassemlooy, Northumbria University, Newcastle upon Tyne, UK

G Swift, Sheffield Hallam University, Sheffield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The ever-increasing aggregate demand of electrically based time division multiplexing systems should have coped with the steady growth rate of voice traffic. However, since 1990, the explosive growth of the Internet and other bandwidth-demanding multimedia applications, has meant that long-haul telecommunication traffic has been increasingly dominated by data, not voice traffic. Such systems suffer from a bandwidth bottleneck due to speed limitations of the electronics. This limits the maximum data rate to considerably less than the THz bandwidth offered by an optical fiber. Optical technology is proposed as the only viable option and is expected to play an ever increasing role in future ultrahigh-speed links/networks. There are a number of multiplexing techniques, such as space division multiplexing (SDM), wavelength division multiplexing (WDM), and optical time division multiplexing (OTDM), that are currently being applied to effectively utilize the bandwidth of optical fiber as a means to overcome the bandwidth bottleneck imposed by electrical time division multiplexing (TDM). In SDM a separate optical fiber is allocated to each channel, but this is the least preferred option for increasing channel numbers. In WDM, a number of different data channels are allocated to discrete optical wavelengths for transmission over a single fiber. Dense WDM technology has been improving at a steady rate in recent years, with the latest systems capable of operating at a data rate of >1 T bps, using a large number of wavelengths over a single fiber link. However, there are a number of problems associated with the WDM systems such as:

- Performance of WDM is highly dependent on the nonlinearities associated with fiber, i.e.:
  - *Stimulated Raman scattering*: degrades the signal-to-noise (SNR) as the number of channels increases;
  - *Four-wave mixing*: limits the channel spacing;
  - *Cross-phase modulation*: limits the number of channels.
- Relatively static optical paths, thus offering no fast switching with high performance within the network;

- Switching is normally carried out by separating each wavelength of each fiber onto different physical outputs. Space switches are then used to spatially switch the separated wavelengths, an extremely inefficient way of utilizing network resources;
- The need for amplifiers with high gain and flat spectra.

In order to overcome these problems, OTDM was introduced that offers the following:

- Flexible high bandwidth on demand (>1 Tbit/s compared to the bit rates of 2.5–40 Gbit/s per wavelength channel in WDM systems);
- The total bandwidth offered by a single channel network is equal to DWDM;
- In a network environment, OTDM provides potential improvements in:
  - Network user access time, delay and throughput, depending on the user rates and statistics.
  - Less complex end node equipment (single-channel versus multichannels).
- Self-routing and self-clocking characteristics;
- Can operate at second- and third-transmission windows:
  - 1500 nm (like WDM) due to Erbium doped fiber amplifier (EDFA);
  - 1300 nm wavelengths.
- Offers both broadcast and switched based networks.

Principle of OTDM

Figure 1 show the generic block diagram of a point-to-point OTDM transmission link, where N optical data channels, each of capacity M Gbps, are multiplexed to give an aggregate rate of $N \times M$ Gbps. The fundamental components are a pulsed light source, an optical modulator, a multiplexer, channel, add/drop unit, and a demultiplexer. The light source needs to have good stabilities and be capable of generating ultrashort pulses (<1 ps). Direct modulation of the laser source is possible but the preferred method is based on external modulation where the optical signal is gated by the electronic data. The combination of these techniques allows the time division multiplexed data to be encoded inside a subnanosecond time slot, which is subsequently interleaved into a frame format. Add/drop units provide added versatility (see Figure 2) allowing the 'adding' and 'dropping' of selected OTDM channels to interchange data at chosen points on the link. At the receiving end, the OTDM pulse stream is demultiplexed down to the individual channels at the initial M Gbps data rate. Data retrieval is then within the realm of electronic devices and the distinction between electronic and optical methods is no longer relevant. Demultiplexing requires high-speed all optical switching and can be achieved using a number of methods, which will be discussed in more detail below.

The optical multiplexing (or interleaving) can be carried out at the bit level (known as bit interleaving) or at the packet level (known as packet interleaving), where blocks of bits are interleaved sequentially. This is in accord with the popular conception of packet switched networks. The high data rates required and consequently narrow time slots necessitate the need for strict tolerances at processing nodes, e.g., switches. As such, it is important that the duration

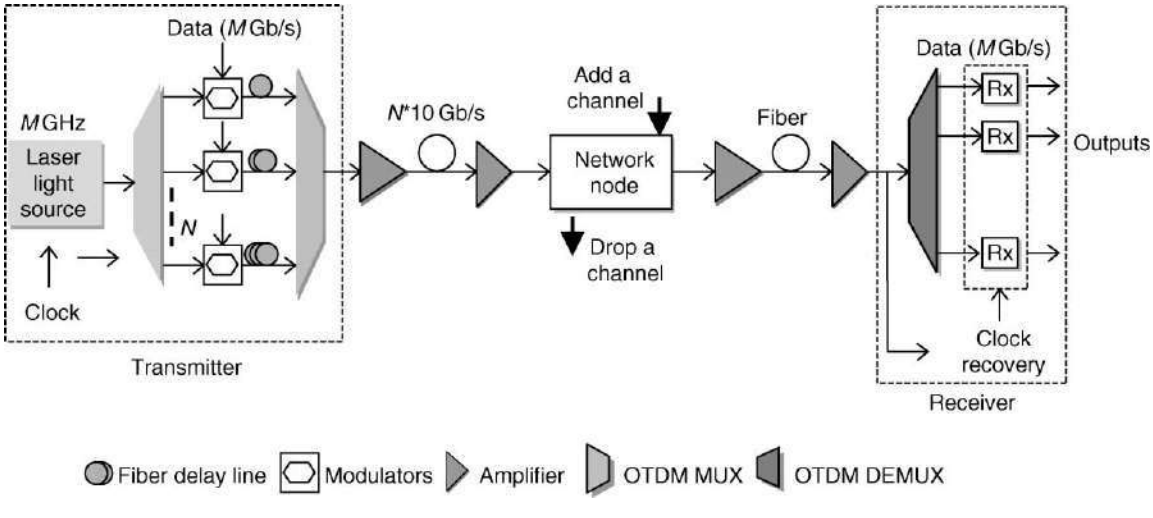

Figure 1 Block diagram of a typical OTDM transmission system.

of the optical pulses is chosen to be significantly shorter than the bit period of the highest multiplexed line rate, in order to reduce the crosstalk between channels.

Bit Interleaved OTDM

A simple conceptual description of a bit interleaved multiplexer is shown in Figure 3. It uses a number of different length optical fiber delay lines (FDL) to interleave the channels. The propagation delay of each FDL is chosen to position the optical channel in its corresponding time slot in relation to the aggregate OTDM signal. Prior to this, each optical pulse train is modulated by the data stream. The output of the modulators and an undelayed pulse train, labeled the framing signal, are combined, using a star coupler or combiner, to produce the high bit rate OTDM signal (see Figure 3b). As shown in Figure 3b, the framing pulse has a higher intensity for clock recovery purpose. At the demultiplexer, the incoming OTDM pulse train is split into two paths (see Figure 4a). The lower path is used to recover the framing pulse by means of thresholding, this is then delayed by an amount corresponding to the position of the *i*th (wanted) channel (see Figure 4b). The delayed framing pulse and the OTDM pulse stream are then passed through an AND gate to recover the *i*th channel. The AND operation can be carried out all optically using, for example, a nonlinear loop mirror, a terahertz optical asymmetric demultiplexer, or a soliton-trapping gate.

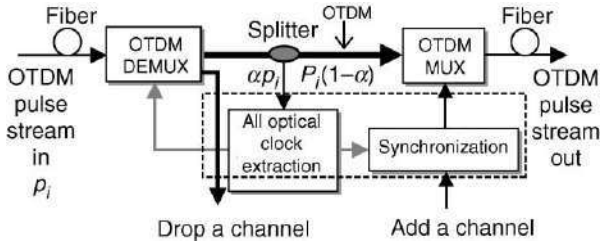

Figure 2 Add/drop unit in an OTDM network node.

Packet Interleaved OTDM

Figure 5a shows the block diagram of a system used to demonstrate packet interleaving. Note that the time interval between successive pulses now needs to be much less than the bit interval T. This is achieved by passing the low bit rate output packet of the modulator into a compressor, which is based on a feed forward delay line structure. The feed forward delay lines are based on a cascade of passive M–Z interferometers with unequal arm lengths. This configuration generates the delay times required, i.e., $T-\tau, 2(T-\tau)\ldots(2^{n-1})(T-\tau)$, etc., where $n = \log_2 k$ is the number of stages, T and τ are the

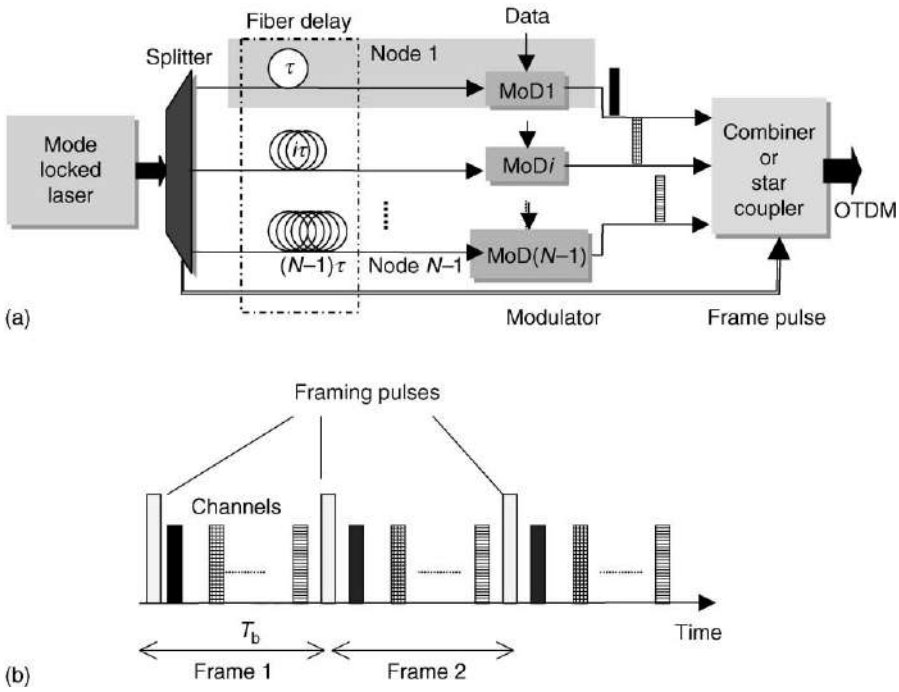

Figure 3 Bit interleaved OTDM; (a) block diagram and (b) timing waveforms.

incoming bit duration and the outgoing bit duration, respectively, and k is the packet length in bits. Pulse i's location at the output is given by $(2^n - 1)(T - t) + (i - 1)\tau$. The signal at the input of the compressor is:

$$I_{\text{in}}(t) = \sum_{i=0}^{k-1} \delta(t - iT)A_i \qquad [1]$$

Where $\delta(\cdot)$ is the optical carrier pulse shape, and A_i is the ith data bit in the packet.

As shown in Figure 5b, each bit is split, delayed, and combined to produce the four bit packets O_i:

$$O_i(t) = \frac{1}{2^{n+1}} \sum_{i=0}^{k-1} I_i[t - i(T - \tau)] \qquad [2]$$

Note the factor of $1/(2^{n-1})$ is due to the signal splitting at the 2×2 3-dB coupler at the input of each stage.

The combined signal is shown in Figure 5b, and is defined as:

$$I_{\text{out}} = \sum_{i=0}^{k-1} O_i = \frac{1}{2^{n+1}} \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} I_i[t - (i+j)T + j\tau]A_i \qquad [3]$$

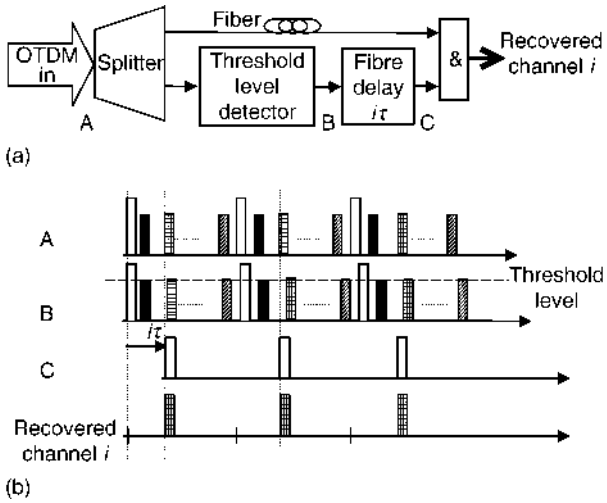

Figure 4 Bit interleaving demultiplexer: (a) block diagram and (b) typical waveforms.

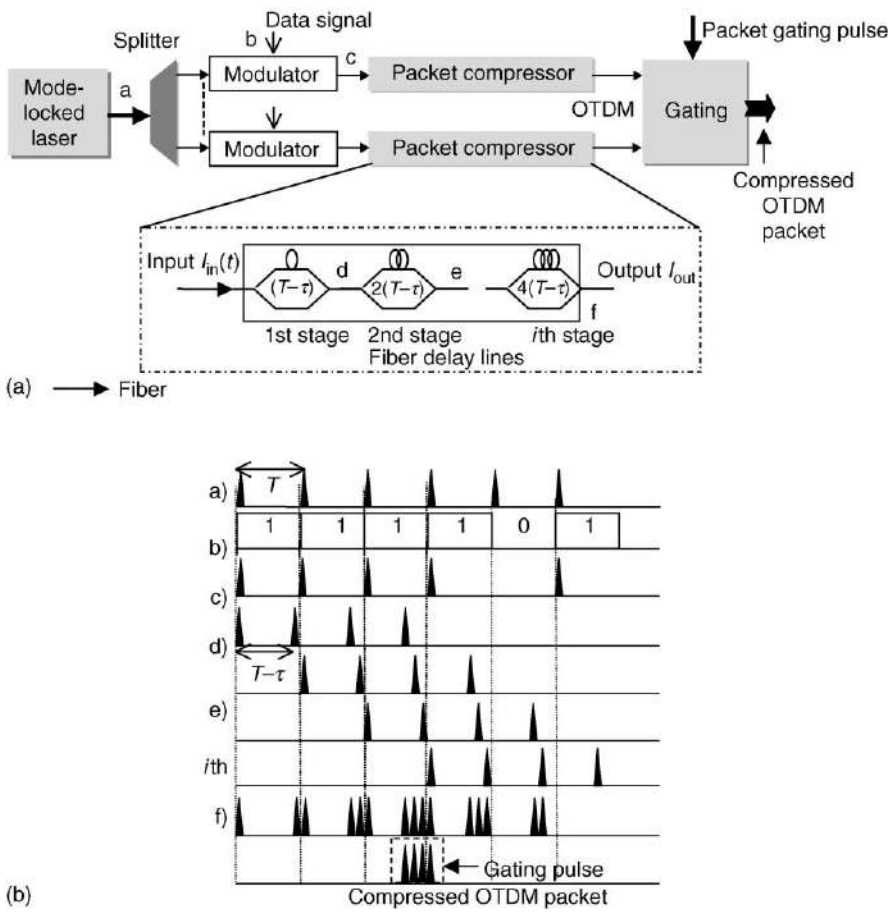

Figure 5 Packet interleaving multiplexer: (a) block diagram and (b) typical waveforms.

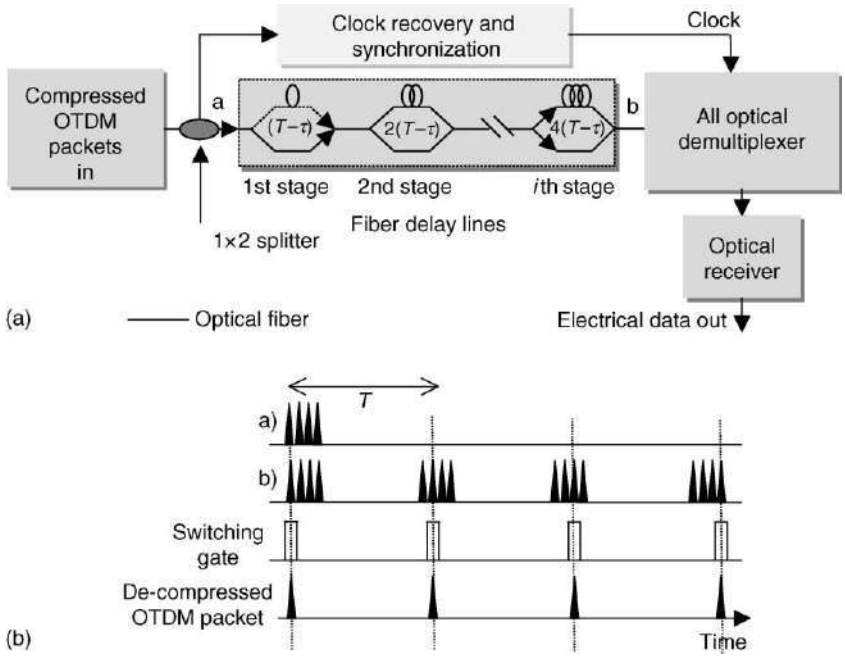

Figure 6 Packet interleaving demultiplexer: (a) block diagram and (b) typical waveforms.

For a packet delay of $(i+j)T$ packets are being built up. For the case $(i+j)=3$, four bits are compressed into a packet (see Figure 5b). An optical gating device is used to select the only complete usable compressed copy of the incoming packets from the unwanted copies.

At the receiving end, demultiplexing is carried out by decompressing the OTDM packet stream using the same delay line structure (see Figure 6). Multiple copies of the compressed packet are generated, and each is delayed by $(T-\tau)$. An optical demultiplexer placed at the output of the delay lines generates a sequence of switching window at the rate of $1/T$. By positioning the switching pulses at the appropriate position, a series of adjacent packet (channel) bits from each of the copied compressed packets is extracted (see Figure 6b). The output of the demultiplexer is the decompressed packet signal, which can now be processed at a much slower rate using electronic circuitry.

Components of an OTDM System

Optical Sources

In an ultrahigh speed OTDM system, it is essential that the optical sources are capable of generating transform-limited subnanosecond pulses having low duty cycles, tuneability, and a controllable repetition rate for synchronization. A number of suitable light sources are: gain-switch distribute feedback laser (DFB), active mode locked lasers (MLL) (capable of generating repetitive optical pulses), and harmonic mode-locking Erbium-doped fiber (EDF) lasers. Alternative techniques include super-continuum pulse generation in a dispersion shifted fiber with EDF pumping, adiabatic soliton compression using dispersion-flattened dispersion decreasing fiber, and pedestal reduction of compressed pulses using a dispersion-imbalanced nonlinear optical loop mirror. As shown in Figure 1, the laser light source is power split to form the pulse source for each channel. These are subsequently modulated with an electrical data signal (external modulation). External modulation is preferred in an OTDM system as it can achieve narrow carrier linewidth, thus reducing the timing jitter of the transmitted pulse. For ultrahigh bit rate OTDM systems, the optical pulses emerging from the external modulators may also need compressing. One option is to frequency-chirp the pulses and pass them through an anomalous dispersive medium. Using this approach, it is essential that the frequency chirp be linear throughout the duration of the pulse, in addition the midpoint of the linear-frequency chirp should coincide with the center of the pulse. When a frequency-chirped pulse passes through a dispersive medium, different parts of the pulse travel at different speeds, due to a temporal variation of frequency. If the trailing edge travels faster than the leading edge, the result would be pulse compression. An optical fiber cable or a semiconductor laser amplifier (SLA) can be used to create the frequency chirp.

Multiplexers

Multiplexing of the pulses generated by the optical sources can be implemented either passively or actively. The former method is commonly implemented using a mono-mode optical fiber. This method has the advantage of being simple and cost-effective. The latter method uses devices more complex in nature, for example, electro-optic sampling switches, semiconductor optical amplifiers, and integrated optics.

An integrated active multiplexer can be made by integrating semiconductor laser amplifiers (SOA) into a hybrid planar light-wave circuit (PLC). If the optical path length is kept short then temperature control is made easier. An alternative to this approach is to use lithium niobate (LiNb) technology. Periodically poled LiNb (PPLN) based OTDM multiplexers offer compact size and low noise (due to the absence of amplifier spontaneous emission noise and pattern effect, which is a feature of SOA based devices). Figure 7 is a schematic of a PLC-based OTDM multiplexer composed of two PLCs (R and L), composed of 1×8 and 8×1 couplers, eight 2×1 couplers, and eight different path length PPLN waveguides. The input clock pulse is split into eight by the 1×8 coupler in PLC-L, and are combined with the modulated optical pulse trains using the 2×1 couplers. The outputs of the 2×1 couplers are then passed through a PPLN waveguide to generate a return-to-zero (RZ) optically modulated signal. These are then combined, using the 8×1 coupler in PLC-R. When the path-length difference between the waveguides is set to one time slot (equal to 1/8 of the base data rate), then the output of the 8×1 coupler is the required high-bit rate OTDM pulse.

Demultiplexers

In contrast to multiplexing, demultiplexing must be performed as an active function. It can be implemented electro-optically or optically. The former method needs to complete demultiplexing of all channels in order to extract a single channel (see Figure 8). The demultiplexer in Figure 8 uses two LiNb Mach–Zehender (M–Z) modulators in tandem. The first and second modulators are driven with a sinusoidal signal of amplitudes $2V_\pi$ and V_π, respectively, to down-covert the N-bit rate to $N/2$ and $N/4$, respectively. Channels can be selected by changing either the DC-bias V + to the M–Zs, or the electrical phase delay. At ultrahigh speed implementation of an electro-optics demultiplexer becomes increasingly difficult due to the higher drive voltage requirement by the M–Z. An alternative is to use all optical demultiplexing based on the nonlinear effect in a fiber and optical active devices offering switching resolution in the order of picoseconds. There are a number of methods available to implement all optical demultiplexing. The most popular methods that use fast phase modulation of an optical signal are based on M–Z and Sagnac interferometers. Four-wave mixing is another popular method.

Mach–Zehnder (M–Z) Interferometers

The key to interferometric switching is the selection of an appropriate material for the phase modulation. Semiconductor materials, often in the form of an SLA, are suitable devices for this purpose.

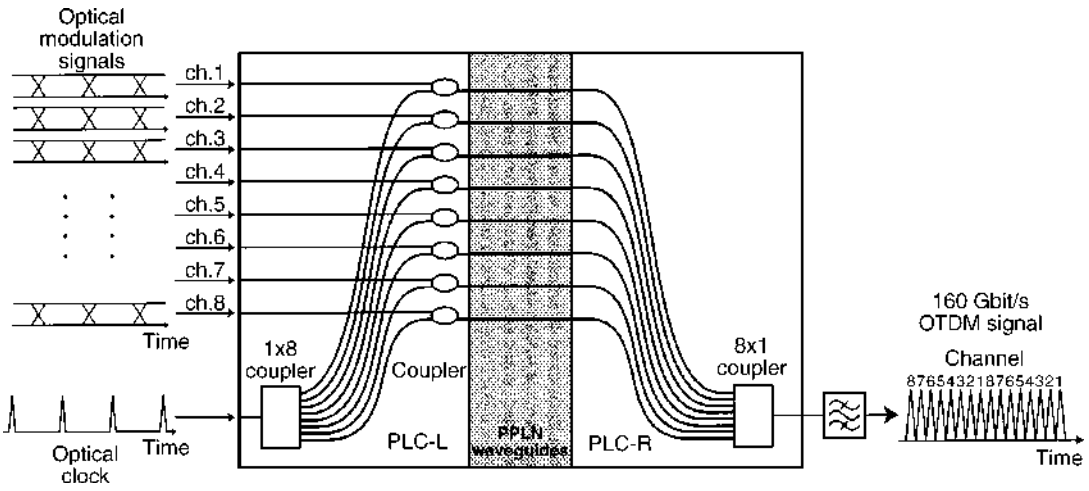

Figure 7 Configuration of a PLC-OTDM-MUX. Reproduced with permission from Ohara T, Takara H and Shake I, *et al.* (2003). 160-Gb/s optica-time-division multiplexing with PPLN hybrid integrated planar lightwave circuit. *IEEE Photonics Letters* 15(2): 302–304. © 2003 IEEE.

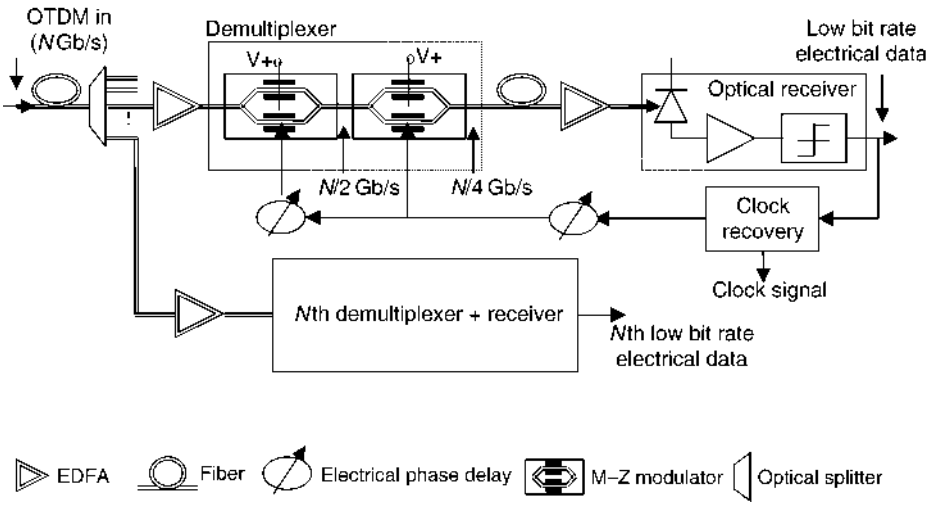

Figure 8 Electro-optics demultiplexer and receiver system block diagram.

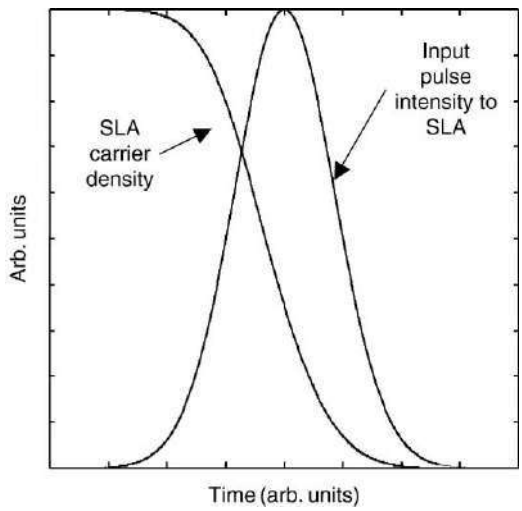

Figure 9 SLA carrier density response to input pulse.

The refractive index of an SLA is a function of the semiconductor carrier density, which can be modulated optically resulting in fast modulation. If a high-intensity pulse is input to an SLA the carrier density changes nonlinearly, as shown in Figure 9. Phase modulation is affected via the material chirp index (dN/dn), which represents the gradient of the refractive index carrier density curve for the material. The phase modulation is quite strong and, in contrast to the intensity-based nonlinearity in an optical fiber, (see Kerr effect below), is a consequence of a resonant interaction between the optical signal and the material. The nonlinearity is relatively long-lived and would be expected to limit the switching speed when using semiconductors. Semiconductors, when in excitation from an optical field, tend to experience the effect quite quickly (ps) with a slow release (hundreds of picoseconds) time. Advantage can be taken of this property by allowing the slow recovery to occur during the time between channels of an OTDM signal, as this time may be of the order of hundreds of pico-seconds or more. A Mach–Zehnder configuration using two SLAs, one in each arm of the interferometer placed asymmetrically, is shown in Figure 10. An optical control pulse entering the device via port 3 initiates the nonlinearity. The transmission equation relating the input signal (port 1) to the output port is given by

$$\frac{I_{\text{out}}(t)}{I_{\text{in}}(t)} = 0.25(G_{\text{SLA1}}(t) + G_{\text{SLA2}}(t) \pm 2\sqrt{G_{\text{SLA1}}(t)G_{\text{SLA2}}(t)}\cos\Delta\phi(t)) \qquad [4]$$

where $G_{\text{SLA1}}(t)$ and $G_{\text{SLA2}}(t)$ refer to the gain profiles of the respective SLAs, and $\Delta\phi(t)$ is the time-dependent phase difference between them. Assuming that the gain and phase profiles of an excited amplifier are given by $G(t)$ and $\phi(t)$, respectively, then the signals passing through the upper and lower arms experience optical properties of the material given by:

$$G(t),\ G(t - T_{\text{d}}),\ \phi(t) \text{ and } \phi(t - T_{\text{d}}) \qquad [5]$$

where T_{d} is given by $2L_{\text{d}}N_{\text{SLA}}/c$, L_{d} is the distance between SLA1 and SLA2, N_{SLA} is the SLA index, and c is the speed of light in a vacuum. As the width of the $\Delta\phi(t)$ profile is dependent on the distance between the SLAs, placing them in close proximity allows high-resolution switching. Less emphasis has been placed on the SLA gain as this is considered to be less effective when phase differences of π are reached. It only remains for the gain and phase

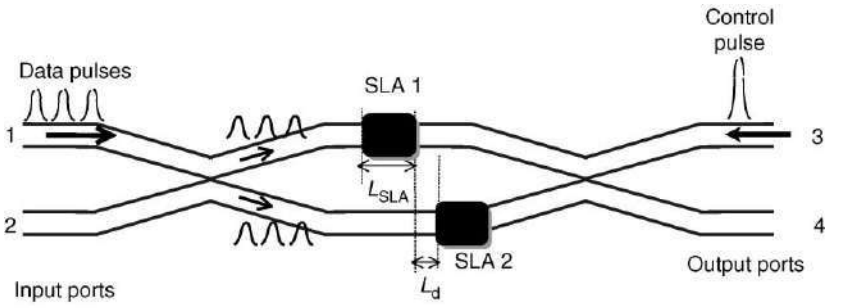

Figure 10 Asymmetric TWSLA Mach–Zehnder devices.

modulation to recover, and as indicated previously, this is allowable over a time-scale commensurate with tributary rates.

Sagnac Interferometers

There are two main types; the nonlinear optical loop mirror (NOLM), and the terahertz optical asymmetric demultiplexer (TOAD).

NOLM

In this method of switching the inherent nonlinearity of an optical fiber known as the Kerr effect is used. The phase velocity of any light beam passing through the fiber will be affected by its own intensity and the intensity of any other beams present. When the intrinsic third-order nonlinearity of silica fibers is considered via the intensity-dependent nonlinear refraction component, then the signal phase shift is

$$\Delta\phi_{\text{signal}} = \frac{2\pi}{\lambda_{\text{ds}}} n_2 L I_{\text{ds}} + 2\frac{2\pi}{\lambda_{\text{ds}}} n_2 L I_{\text{c}} \qquad [6]$$

where n_2 is the Kerr coefficient, I_{ds} is the intensity of the data signal to be switched, I_{c} is the intensity of the control signal used to switch the data signal, λ_{ds} is the data signal wavelength, and L is the fiber length. The optical loop mirror consists of a long length of mono-mode fiber formed into a fiber coupler at its free ends (see Figure 11). The input to the loop comprises the high-frequency data stream plus a control pulse at the frame rate. The data split at the coupler and propagate around the loop in contra directions (clockwise E_{CW} and counter-clockwise E_{CCW}) recombining back at the coupler. In the absence of a control pulse, the pulse exits via port 1. If a particular pulse in the loop (in this example E_{CW}) is straddled by the control pulse (see Figure 12), then that pulse experiences cross-phase modulation, according to the second term on the right-hand side of eqn [6], and undergoes a phase

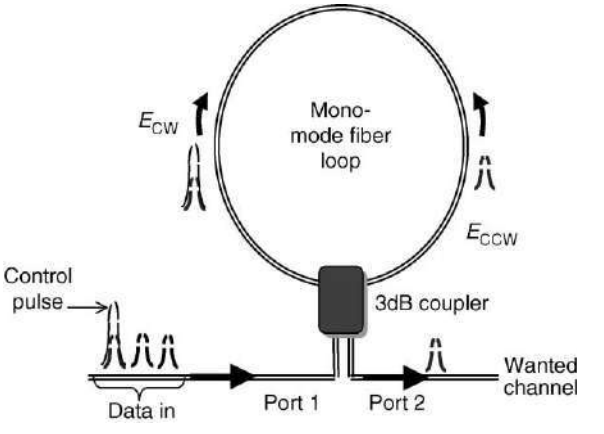

Figure 11 Nonlinear optical loop mirror.

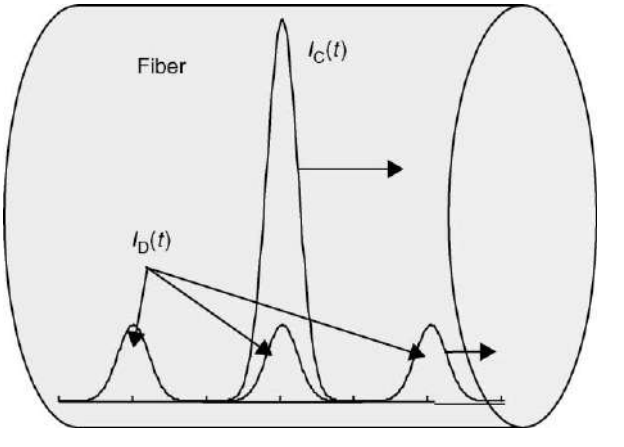

Figure 12 Control and data pulses propagation in the fiber.

change relative to E_{CW}. The difference in the phase between E_{CW} and E_{CCW} causes the pulse to exit via port 2. The phase shift profile experienced by the co-propagating pulse is

$$\Delta\phi(t) = \frac{4\pi}{\lambda_{\text{ds}}} n_2 \int_0^L I_{\text{c}}(t)\,\mathrm{d}x \qquad [7]$$

Assuming unity gain around the loop, the transmittance of the NOLM is

$$T_x = 1 - \cos^2\left(\frac{\Delta\phi}{2}\right) = \Delta\phi(t)$$

$$= 1 - \cos^2\left(\frac{2\pi}{\lambda_{ds}} n_2 \int_0^L I_C(t)\,dx\right) \qquad [8]$$

As it stands, the switching resolution is determined by the width of the control pulse; however, it is possible to allow the signal and control to 'walk off' each other, allowing the window width to be increased by an amount determined by the 'walk off'. The phase shift is now

$$\Delta\phi(t) = 2\frac{2\pi}{\lambda_{ds}} n_2 \int_0^L I_C(t - T_w x)dx \qquad [9]$$

where the parameter T_w, is the walk off time per unit length between control and signal pulses. The increased window width is accompanied by a lower peak transmission (see Figure 13).

TOAD

The TOAD uses a loop mirror architecture that incorporates an SLA and only needs a short length of fiber loop (see Figure 14a). The SLA carrier density is modulated by a high-intensity control pulse, as in the M–Z-based demultiplexer. The operation of the TOAD is similar to the NOLM, where the loop transmittance is determined by the phase difference between CW and CCW traveling pulses. Strictly speaking, the gain of the SLA must also be taken into account; however, without any loss of generality the effect is adequately described by considering only the phase property. Figure 14a shows the timing diagram associated with a TOAD demultiplexer. The control pulse, shown in Figure 14b(1), is incident at the SLA at a time t_1; the phase profiles for the CCW and CCW data pulses are shown in Figures 14b(2) and (3), respectively. The resulting transmission window is shown in Figure 14b(4). The window width is given by $t_{asy} = 2\Delta x/c_f$, where Δx is the SLA offset from the loop center and c_f is the speed of light in the fiber. As in NOLM, if the phase difference is of sufficient magnitude, then data can be switched to port 2. The switch definition is, in principle, determined by how close the SLA is placed to the loop center when the asymmetry is relatively large. However, for small asymmetries the switching window is asymmetric, which is due to the CW and CCW gain profiles being different (see Figure 15). Assuming the phase modulation dominates the transmission then the normalized transmission for a small asymmetry loop is as depicted in Figure 16. The gain response (not shown) would have a similar shape and temporal position as the phase response.

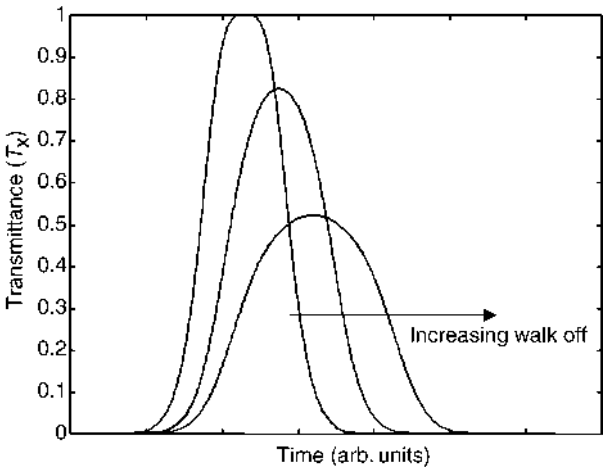

Figure 13 Transmittance profiles with the walk-off time as a parameter.

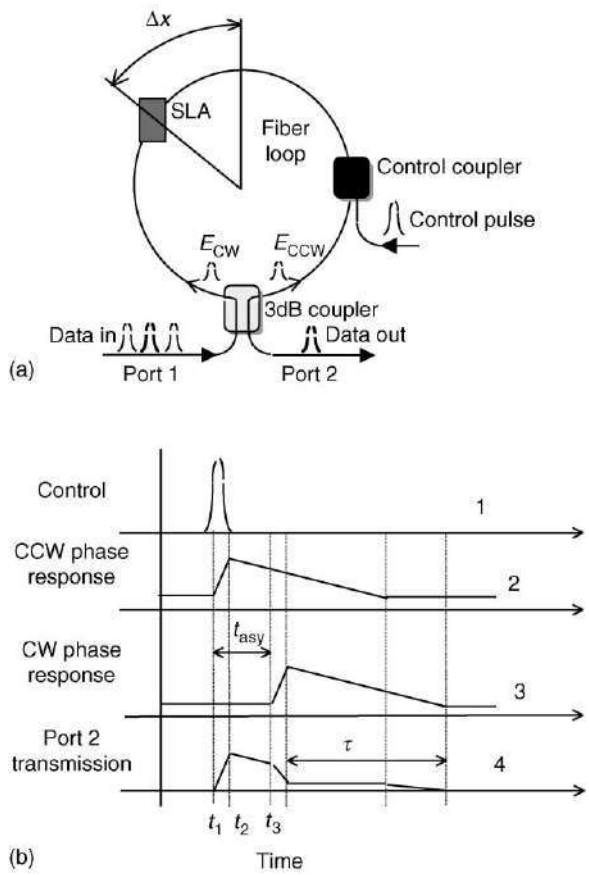

Figure 14 TOAD: (a) architecture and (b) timing diagrams.

Four-Wave Mixing (FWM)

FWM demultiplexing uses a concept whereby two optical signals of different wavelengths are mixed together in a nonlinear medium to produce harmonic components. The nonlinearity in this case arises from the third-order nonlinear susceptibility $\chi^{(3)}$ of an optical fiber, such that the polarization P induced on a pair of electric fields propagating through the fiber is

$$P = \varepsilon_0 \chi^{(3)} (E_{ds} + E_c)^3 \qquad [10]$$

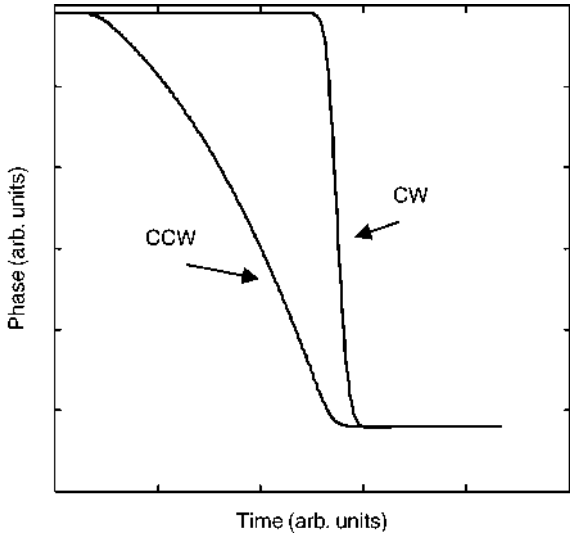

Figure 15 Phase responses for CW and CCW components of TOAD.

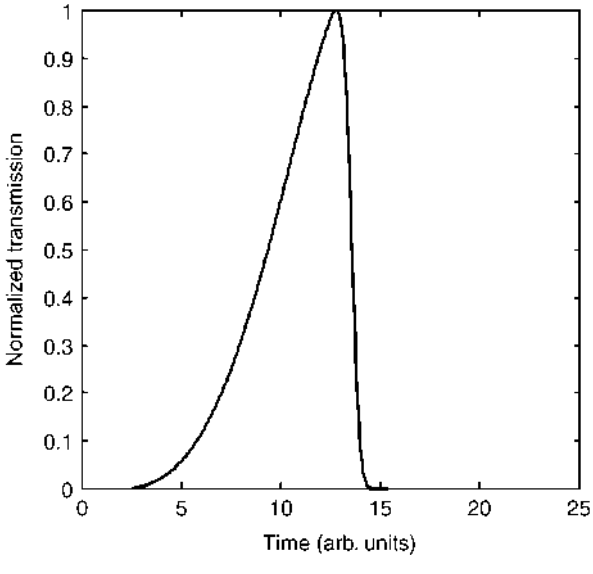

Figure 16 TOAD transmission window profile for small asymmetry loop.

where ε_0 is the vacuum permittivity of the medium, E_{ds} and E_c are the electric fields of data and control signals, respectively. The mixing of the two signals takes place in a long length of fiber in which the control signal propagates alongside the channel to be demultiplexed (see Figure 17). The control signal is of sufficient intensity to cause the fiber to operate in the nonlinear regime. The nonlinear relationship causes a number of frequencies to be generated, with the ones of interest having a frequency given by

$$f_{FWM} = 2f_c - f_{ds} \qquad [11]$$

An optical filter is then used to demultiplex the required channel from the composite signal. The nonlinearity is detuned from the resonant frequency of the fiber glass and as such tends to be weak, requiring long lengths of fiber to give a measurable effect. The power in the demultiplexed signal, for given data and control signal wavelengths and fiber material, depends on the input power and the fiber length according to

$$P_{FWM} = kP_{ds}P_c L_E^2 \qquad [12]$$

where k is a constant, P_{ds} is the signal launch power, P_c is the control signal launch power, and L_E is the fiber effective length. FWM is essentially an inefficient method as power is wasted in the unused frequency components of the four-wave signal. More in line with integrated structures, the nonlinear properties of a semiconductor laser amplifier can be used. Here a relatively high-power optical signal is input to the SLA (see Figure 18). This enables saturation and operation in the nonlinear regime

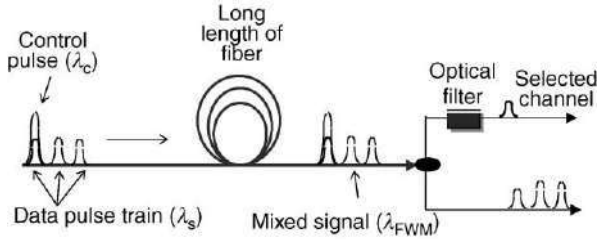

Figure 17 Block diagram of FWM demultiplexer.

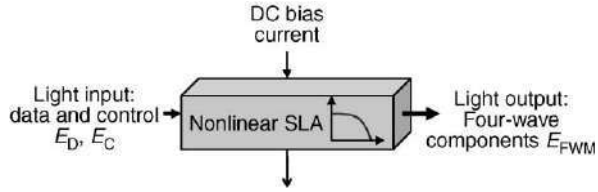

Figure 18 Four-wave mixing in SOA.

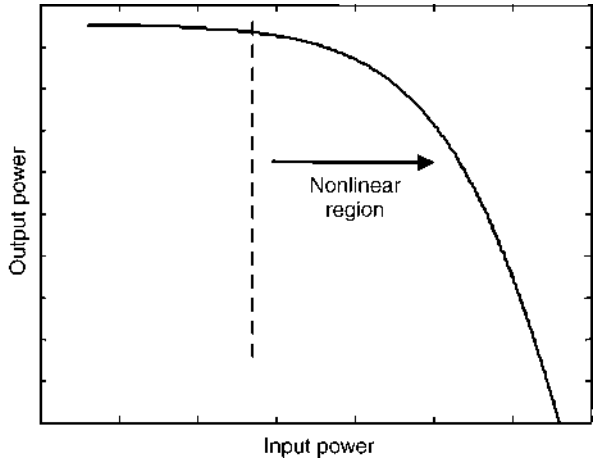

Figure 19 SLA output–input power curve.

(see Figure 19). Operating the SLA in saturation allows the nonlinear effects to produce the FWM components as in the fiber method.

Clock and Data Synchronization in OTDM

In common with electronically based TDM systems, clock recovery is fundamental to the recovery of data in ultrahigh-speed OTDM systems. Two main methods are proposed: (i) clock signal transmitted with the OTDM signal (i.e., multiplexed); and (ii) extraction of clock signal from the incoming OTDM pulse stream. In a packet-based system, synchronization between the clock and the data packet is normally achieved by sending a synch pulse with each packet. However, techniques based on optical-phased locked loops are proving popular and remove the need for a separate clocking signal.

Clock Multiplexing

(i) *Space division multiplexing*: This is conceptually the simplest to implement, where the clock signal is carried on a separate fiber from the data. However, it is susceptible to any differential delay between the different paths taken by clock and data due to temperature variation. It is difficult to justify in systems where the installed fiber base is at a premium.

(ii) *Wavelength division multiplexing*: Here different wavelengths are allocated to the clock, and payload. It is only really practical for predetermined path lengths between nodes in single-hop networks such as point–point links or broadcast-and-select star networks. It also suffers from random delays between the clock and the payload, which is problematic in an asynchronous packet switched based network, where the optical path length a packet may take is nondeterministic.

(iii) *Orthogonal polarization*: This is suitable for small links, where separate polarizations are used for the clock and data. However, in large networks it is quite difficult to maintain the polarization throughout the transmission link due to polarization mode dispersion and other nonlinear effects.

(iv) *Intensity division multiplexing*: This uses higher-intensity optical clock pulses to differentiate it from the data pulses as discussed above. However, in long-distance transmission links, it is difficult to maintain both the clock intensity and its position, due to the fiber nonlinearity.

(v) *Time division multiplexing*: In this scheme a single clock pulse, which has the same wavelength, polarization, and amplitude as the payload pulses, is separated in time, usually ahead of the data pulses.

Synchronization – Optical Phased Locked Loops (PLL)

The PLL is a common technique used for clock recovery in electronic TDM systems. However, the speed of conventional electronic PLLs tends to be limited by the response of the phase comparators used. There are a number of approaches based on opto-electronic PLL. Opto-electronic PLLs based on four-wave mixing in a traveling wave laser amplifier are complex and can suffer from frequency modulation in the recovered clock. However, others based on balanced photo-detectors, result in low timing jitter and good phase stability. In contrast, a number of all optical methods clock recovery scheme exist, one technique based on the TOAD (see above) is depicted in Figure 20. The high-speed data stream enters the TOAD at a rate $n \times R$ where n is the number of channels per OTDM frame, and R is the frame rate. A pulse generator, such as a mode locked fiber laser (MLFL) clocked by a local oscillator (LO), is used as the TOAD control input (MLFL-C). The OTDM data is switched by the TOAD at a frequency of say, $R + \Delta f$ Hz. Thus, the switching window samples data pulses at a rate higher or lower than $n \times R$ Hz and uses this signal for cross correlation in the PLL unit. The output of the phase comparator is used to regulate a voltage controlled oscillator (VCO)

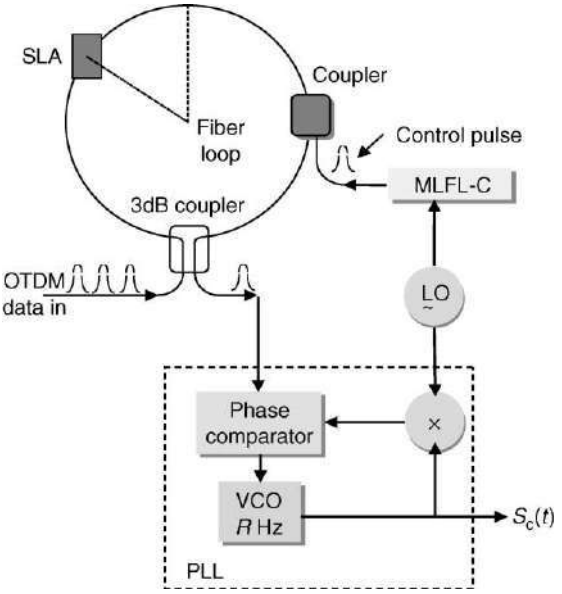

Figure 20 Ultrafast clock recovery using TOAD.

running at R Hz, which in turn feeds the control signal. The PLL circuit locks into the clock frequency, generating the clock signal $S_c(t)$.

OTDM Bit-Error Rate (BER) Performance

Figure 21 shows a typical block diagram of a high-speed optical receiver for an OTDM system, composed of a NOLM or a TOAD demultiplexer, an optical pre-amplifier, an optical bandpass filter, and a conventional optical receiver using a PIN photodiode. Due to the crosstalk introduced in the demultiplexing process, the demultiplexed optical signal contains not only the target channel but may also contain nontarget channels with reduced amplitude (see the inset in Figure 21). The intensity of the demultiplexed optical signal is boosted by the optical pre-amplifier (EDFA). Amplified spontaneous emission (ASE) from the optical preamplifier adds a wide spectrum of optical fields onto the demultiplexed optical signal. Although an optical bandpass filter (BPF) can reduce the ASE, it still remains one of the major noise sources. The function of the optical filter is to reduce the excess ASE to within the range of the signal spectrum. The PIN photodiode converts the received optical power into an equivalent electrical current, which is then converted to a voltage signal by an electrical amplifier. Finally, the output voltage is sampled periodically by a decision circuit for estimating the correct state (mark/space) of each bit.

The BER performance of an OTDM system deteriorates because of the noise and crosstalk

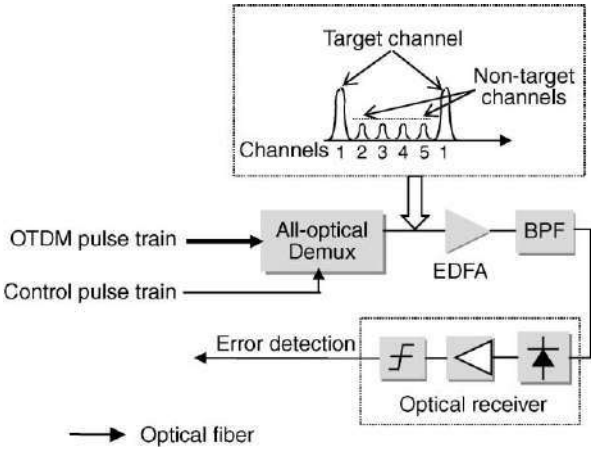

Figure 21 Block diagram of OTDM receiver.

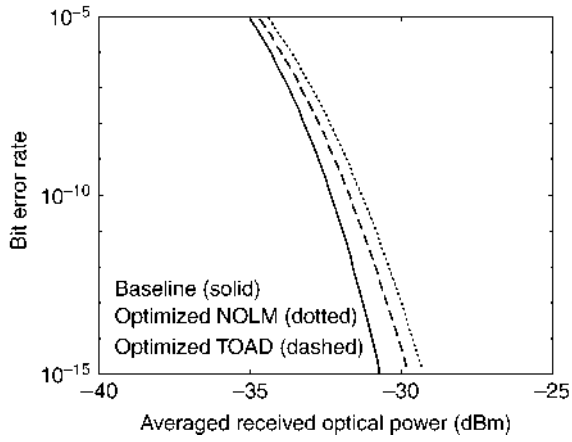

Figure 22 BER versus the average received optical power for 100 Gb/s (10 channels) OTDM system.

introduced by the demultiplexer and is:

$$\mathrm{BER} = \frac{1}{\sqrt{2\pi}} \frac{\exp\left(-Q^2/2\right)}{Q} \qquad [13]$$

where Q is defined as:

$$Q = \frac{\bar{I}_m - \bar{I}_s}{\sqrt{\sigma^2_{\mathrm{RIN,m}} + \sigma^2_{\mathrm{RIN,s}} + \sigma^2_{\mathrm{amp,m}} + \sigma^2_{\mathrm{amp,s}} + \sigma^2_{\mathrm{rec,m}} + \sigma^2_{\mathrm{rec,s}}}} \qquad [14]$$

and $\bar{I}_m$ and $\bar{I}_s$ are the average photocurrents for a mark and a space, respectively. $\sigma_{x,i}$ are the variances of the relative intensity noise (RIN), further optical pre-amplifier and receiver for a mark and a space.

For 100 Gb/s (10 channels) OTDM system, the BER against average received optical power for optimized NOLM and TOAD demultiplexers, is shown in Figure 22.

List of Units and Nomenclature

A_i	ith data bit in a packet
c	Speed of light in a vacuum
c_f	Speed of light within the fiber
E_c	Electric fields of control signals
E_{ds}	Electric fields of data signals
f_{FWM}	Frequency due to four-wave mixing
$G_{SLA}(t)$	Gain profiles of SLA
I_c	Intensity of the control signal
I_D	Intensity of the data signal
$\overline{I}_m$	Average photo current for a mark
$\overline{I}_s$	Average photo current for a space
k	Packet length in bits
L	Fiber length
L_d	Distance between two SLAs
L_E	Fiber effective length
n	Number of stages
n_2	Kerr coefficient
N_{SLA}	SLA index
P	Polarization
P_c	Control signal launch power
P_{ds}	Data signal launch power
t_{asy}	The window width is given by $t_{asy} = 2\Delta x/c_f$
T	Incoming bit interval
T_w	Walk off time per unit length between control and data pulses
T_x	Transmittance of the NOLM
$\delta(\cdot)$	Optical carrier pulse shape
$\Delta\phi(t)$	Time-dependent phase difference
Δx	SLA offset from the fiber loop center
ε_0	Vacuum permittivity of the medium
λ_s	Data signal wavelength
$\sigma_{amp,m}$	Variances for optical pre-amplifier for a mark
$\sigma_{amp,s}$	Variances for optical pre-amplifier for a space
$\sigma_{rec,m}$	Variances for receiver for a mark
$\sigma_{rec,s}$	Variances for receiver for a space
$\sigma_{RIN,m}$	Variances of RIN for a mark
$\sigma_{RIN,s}$	Variances of RIN for a space
τ	Outgoing bit interval
$\phi(t)$	Phase profile of SLA
$\chi^{(3)}$	Third-order non-linear susceptibility of an optical fiber

See also

Interferometry: Overview. **Nonlinear Optics, Basics**: Four-Wave Mixing. **Optical Communication Systems**: Optical Time Division Multiplexing; Wavelength Division Multiplexing.

Further Reading

Agrawal GP (2002) *Fiber-Optic Communication Systems*, 3rd edn. New York: J Wiley and Sons Inc.

Chang K (ed.) (2003) *Handbook of Optical Components and Engineering*, 2nd edn. New York: Wiley and Sons Inc.

De Marchis G and Sabella R (1999) *Optical Networks Design and Modelling*. Dordrecht: Kluwer Academic.

Hamilton SA, Robinson BS, Murphy TE, Savage SJ and Ippen EP (2002) 100 Gbit/s optical time-division multiplexed networks. *Journal of Lightwave Technology* 20(12): 2086–2100.

Jhon YM, Ki HJ and Kim SH (2003) Clock recovery from 40 gbps optical signal with optical phase-locked loop based on a terahertz optical asymmetric demultiplexer. *Optics Communications* 220: 315–319.

Kamatani O and Kawanishi S (1996) Ultrahigh-speed clock recovery with phase locked loop based on four wave mixing in a traveling-wave laser diode maplifier. *Journal of Lightwave Technology* 14: 1757–1767.

Nakamura S, Ueno Y and Tajima K (2001) Femtosecond switching with semiconductor-optical-amplifier-based symmetric-Mach–Zehnder-type all-optical switch. *Applied Physics Letters* 78(25): 3929–3931.

Ohara T, Takara H, Shake I, *et al.* (2003) 160-Gb/s optica-time-division multiplexing with PPLN hybrid integrated planar lightwave circuit. *IEEE Photonics Letters* 15(2): 302–304.

Ramaswami R and Sivarajan KN (2002) *Optical Network; a Practical Perspective*. Morgan Kaufman.

Reman E (2001) Trends and evolution of optical networks and technology. *Alcatel Telecommun. Rev. 3rd Quarter* 173–176.

Sabella R and Lugli P (1999) *High Speed Optical Communications*. Dordrecht: Kluwer Academic.

Seo SW, Bergman K and Prucnal PR (1996) Transparent optical networks with time division multiplexing. *Journal of Lightwave Technology* 14(5): 1039–1051.

Sokolloff JO, Prucnal PR, Glesk I and Kane M (1993) A terahertz optical asymmetric demultiplexer (TOAD). *IEEE Photonics Technology Letters* 5(7): 787–790.

Stavdas A (ed.) (2001) *New Trends in Optical Network Design and Modelling*. Dordrecht: Kluwer Academic.

Ueno Y, Takahashi M, Nakamura S, Suzuki K, Shimizu T, Furukawa A, Tamanuki T, Mori K, Ae S, Sasaki T and Tajima K (2003) Control scheme for optimizing the interferometer phase bias in the symmetric-Mach–Zehnder all-optical switch (invited paper). Joint special issue on recent progress in optoelectronics and communications. *IEICE Trans. Electron.* E86-C(5): 731–740.

Coherent Analog Optical Processors

H H Arsenault, S Roy and D Lefebvre, University of Laval, Quebec, PQ, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Linear filtering is a useful tool commonly used in image processing. The correlation operation can be defined as:

$$f(x,y)*h(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)h(x-x',y-y')\mathrm{d}x'\mathrm{d}y' \quad [1]$$

where $*$ denotes the correlation. The operation of correlation and the Fourier transform

$$G(\mu,\nu)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)\mathrm{e}^{-j2\pi(\mu x+\nu y)}\mathrm{d}x\mathrm{d}y \quad [2]$$

are intimately related through the correlation theorem of Fourier theory, which states that the correlation between two functions $f(x,y)$ and $h(x,y)$ is equal to the inverse Fourier transform of the product of the Fourier transforms:

$$f(x,y)*h(x,y)=FT^{-1}[F(\mu,\nu)H^*(\mu,\nu)] \quad [3]$$

where $F(\mu,\nu)$ and $H(\mu,\nu)$ are the Fourier transforms of $f(x,y)$ and of $h(x,y)$ respectively, and where H^* designates the complex conjugate of H. The ability to carry out and display Fourier transforms in two dimensions, by means of optical systems, is at the root of optical information processing.

One of the main advantages of correlation is that the operations do not require segmentation of the regions of interest. This means that the locations of objects do not have to be segmented from the scene and that the detection is shift-invariant, that is the correlation peaks appear at locations in the scene corresponding to locations of the targets. These advantages are clearer for detection in uncontrolled conditions, for example for detecting vehicles in natural scenes. But they are also useful in security applications, such as identity verification or for detecting defects in assembly lines.

The Optical Fourier Transform

The Fourier transform of eqn [2] is widely used in the field of information processing. The inverse Fourier transform is simply defined as:

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}G(\mu,\nu)\mathrm{e}^{i2\pi(\mu x+\nu y)}\mathrm{d}\mu\mathrm{d}\nu \quad [4]$$

We will now discuss how to perform Fourier transforms by means of four different optical setups.

FT Without a Lens

If we illuminate a transparency $U(x_0,y_0)$ with a coherent plane wave, the diffraction pattern we observe at a point d from the input plane is given by the following relation, known as the Fresnel approximation for diffraction (Figure 1):

$$U(x_1,y_1)=\frac{\exp\left(i\frac{2\pi d}{\lambda}\right)}{j\lambda d}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}U(x_0,y_0)\times\exp\left\{i\frac{\pi}{\lambda d}[(x_0-x_1)^2+(y_0-y_1)^2]\right\}\mathrm{d}x_0\mathrm{d}y_0 \quad [5]$$

where λ is the wavelength of the light and d is the distance between the diffracting screen and the observed pattern.

In order to use the Fresnel approximation we need only to satisfy the geometrical optics paraxial conditions. If we add an additional approximation, we can further simplify this equation to yield the Fourier transform of $U(x_0,y_0)$. The extra condition is known as the Fraunhofer or far-field approximation, which is only satisfied when the diffracted image is observed far away from the diffracting screen or when the extent of $U(x_0,y_0)$ is very small. Mathematically this condition is expressed by:

$$\frac{\pi}{\lambda d}(x_0^2+y_0^2)\ll 1 \quad [6]$$

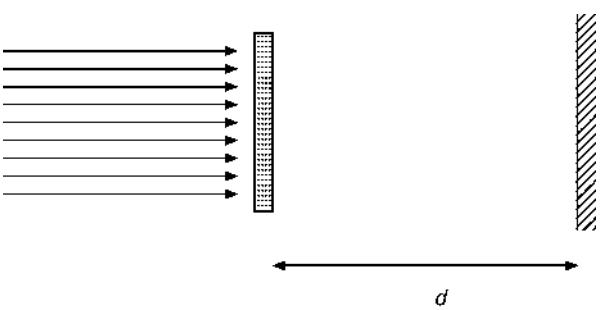

Figure 1 Fresnel diffraction without a lens.

After expanding the squares in the exponentials, there will be three terms: one can go outside the integral and one will be approximately equal to unity so the diffracted pattern becomes

$$U(x_1, y_1) = \frac{\exp\left(i\frac{2\pi d}{\lambda}\right)}{i\lambda d} \times \exp\left[i\frac{\pi}{\lambda d}(x_1^2 + y_1^2)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_0, y_0) \times \exp\left[-i\frac{2\pi}{\lambda d}(x_0x_1 + y_0y_1)\right]\mathrm{d}x_0\mathrm{d}y_0 \quad [7]$$

This is the Fourier transform of $U(x_0, y_0)$, scaled by a factor λd, with an extra quadratic phase factor

$$\exp\left[\frac{i\pi}{\lambda d}(x_1^2 + y_1^2)\right]$$

The factor $\exp(i2\pi d/\lambda)/i$ is usually of no importance since it is a phase factor distributed across the output plane. In some applications we only want to know the power spectrum (squared modulus) of the Fourier transform. In such cases the quadratic phase factor disappears because the squared modulus of any complex exponential is equal to 1. Performing a Fourier transform without a lens is possible within the Fraunhofer approximation. But if it is required to obtain the Fourier transform of an image which extends about 1 cm from the optical axis using a coherent He–Ne laser ($\lambda = 632.8$ nm), the condition to be satisfied is $d \gg 496$ m: the required observation distance d is at least a few kilometers. This kind of optical setup is usually impractical, but can fortunately be alleviated by means of lenses.

Object on the Lens

Instead of having only the transparency in the input plane, we place a lens immediately after it as shown in Figure 2. For mathematical simplicity the lens and input plane are superimposed (more realistic cases are considered below). The lens acts as a phase factor

$$\exp\left[-i\frac{\pi}{\lambda f}(x_0^2 + y_0^2)\right]$$

where f is the focal length of the lens. Now in the input plane the object $U(x_0, y_0)$ is multiplied by the phase factor of the lens:

$$U'(x_0, y_0) = U(x_0, y_0)\exp\left[-i\frac{\pi}{\lambda f}(x_0^2 + y_0^2)\right] \quad [8]$$

Inserting this expression into the integral formula for Fresnel diffraction yields the diffraction pattern observed in the focal plane of the lens:

$$U(x_1, y_1) = \frac{\exp\left(i\frac{2\pi f}{\lambda}\right)}{i\lambda f} \times \exp\left[i\frac{\pi}{\lambda f}(x_1^2 + y_1^2)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_0, y_0) \times \exp\left[-i\frac{2\pi}{\lambda f}(x_0x_1 + y_0y_1)\right]\mathrm{d}x_0\mathrm{d}y_0 \quad [9]$$

The lens exactly compensates for the other phase factor that was in the integral. This expression for the Fraunhofer approximation is almost the same as before, but the focal length f of the lens replaces the distance d. So by placing a lens immediately after the input plane we can have a much more compact setup since lenses are available with a wide range of focal lengths, usually from a few centimeters to a few meters. Another way to look at this is to consider that the effect of a lens is to bring the pattern at infinity to the focal plane of the lens – except for the phase factor.

Object Before the Lens

Another setup used to perform the Fourier transform is shown in Figure 3, where the lens is placed somewhere between the input and the output planes. Using the Fresnel formula for diffraction it can be

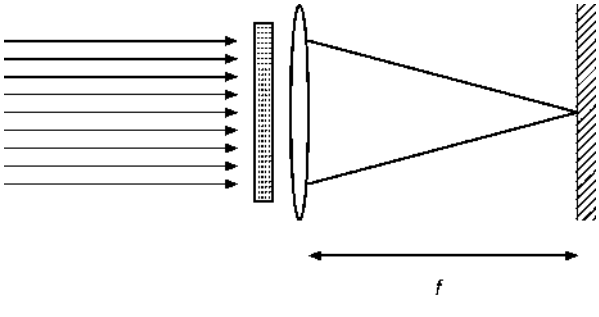

Figure 2 Diffraction with the object on the lens.

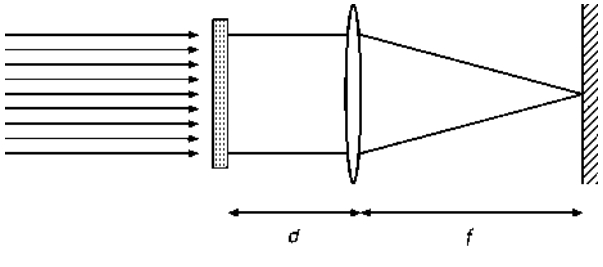

Figure 3 Diffraction with the object before the lens.

shown that the diffraction pattern in the focal plane is:

$$U(x_1, y_1) = \frac{1}{i\lambda f} \times \exp\left[i\frac{\pi}{\lambda f}\left(1 - \frac{d}{f}\right)(x_1^2 + y_1^2)\right] \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_0, y_0) \times \exp\left[-i\frac{2\pi}{\lambda f}(x_0x_1 + y_0y_1)\right] \mathrm{d}x_0\mathrm{d}y_0 \quad [10]$$

where d is the distance between the input plane and the lens. If we set d equal to the focal length f of the lens, the quadratic phase factor in front in the integral vanishes and we have an exact Fourier transform. This setup with the object and filter in the front and back focal planes of the lens respectively is the one most frequently seen in scientific articles and is known as the *f-f* Fourier transform system.

Object After the Lens

Still another configuration is the one shown in Figure 4, where the lens is placed at a distance d before the input plane. For this configuration the Fresnel diffraction yields the following pattern in the focal plane of the lens:

$$U(x_1, y_1) = \frac{\exp\left[i\frac{\pi}{\lambda d}(x_1^2 + y_1^2)\right]}{i\lambda d}\frac{f}{d}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_0, y_0) \times \exp\left[-i\frac{2\pi}{\lambda d}(x_0x_1 + y_0y_1)\right] \mathrm{d}x_0\mathrm{d}y_0 \quad [11]$$

Once again the Fourier transform of $U(x_0, y_0)$ is multiplied by a quadratic phase factor. It is interesting to note that, in this case, the Fourier transform is scaled by a factor λd. This can be used to control the scale of the Fourier transform by changing the distance between the lens and the object.

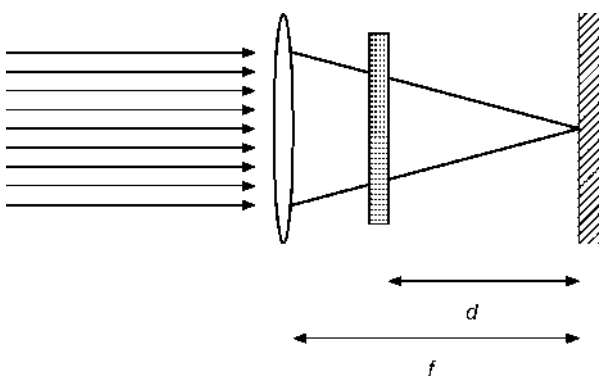

Figure 4 Diffraction with the object after the lens.

Note that although two of the three methods yield Fourier transforms that are accompanied by a quadratic phase factor, the latter is not necessarily an impediment to correlation, contrary to popular belief. Indeed all three configurations have been used for correlation pattern recognition, and each configuration has its advantages.

Optical Correlation

The last section showed how to obtain Fourier transforms by means of an optical setup. The proposed optical setups cannot perform the inverse Fourier transform. Because the Fourier transform and its inverse are very similar, the Fourier transform is performed instead, yielding the mirror image of the correlation:

$$(f * h)(-x, -y) = \mathrm{FT}[F(\mu, \nu)H^*(\mu, \nu)] \quad [12]$$

The correlation can be considered as the result of three operations: a Fourier transform; a product; and an additional Fourier transform. The Fourier transform of a 2D function can be obtained by means of coherent optics. The equivalent of multiplication in optics is transmissivity, so multiplication is accomplished by sending light successively through two superimposed transparencies.

There are two main categories of optical correlators, serial correlators and joint transform correlators.

Serial Correlators

A serial correlator functions in two stages. The first stage performs the Fourier transform of the input image $f(x, y)$, which is multiplied in the Fourier plane with a transparency $H(\mu, \nu)$. The filter $H(\mu, \nu)$ can be a hologram, a spatial light modulator (SLM) or a simple transparency. A second Fourier transform displays in the output plane the correlation of the input image with the impulse response $h(x, y)$, which is the Fourier transform of the filter function $H(\mu, \nu)$.

Serial correlators directly implement the two Fourier transforms and the multiplication operation in a pipelined manner. The input plane and the correlation plane are conjugate planes of the system, that is in the absence of a filter, the output plane displays an inverted image of the input $f(x, y)$. Between those conjugate planes, at the conjugate plane of the source with respect to the first lens, is the Fourier plane where the spatial frequency filter is placed. When the input is illuminated with parallel

light, the frequency plane is at the back focal plane of the first lens. A SLM can be placed in the input plane to feed the images to be correlated; a second SLM can be placed in the Fourier plane to display the filter. The filter is often calculated by a digital computer, and usually takes the form of a hologram if phase information must be used in the filter.

The 4-*f* correlator

The 4-*f* correlator is the most intuitive optical correlator system. We showed above that an *f*-*f* system can be used to optically obtain a Fourier transform (Figure 5). If $f(x, y)$ is in the front focal plane, its Fourier transform $F(\mu, \nu)$ appears at the back focal plane of the first lens. If a filter function $H^*(\mu, \nu)$ is inserted at this point, by means of a SLM or a hologram, then a second *f*-*f* system can perform a second Fourier transform to yield the correlation plane at the back focal plane of the second lens.

Convergent beam correlator

This kind of correlator consists of two conjugate systems. This time, the Fourier transform is obtained by placing the input object in a convergent beam after the first lens. The second lens is placed so that the output plane is conjugate to the input plane, and as before, the filter is placed at the point of convergence of the light after the first lens. It is important to note that this is not the focal plane of the lens, but the conjugate plane of the source of light. In order to have an imaging system between the input and output planes, the distances must satisfy the conjugate relation:

$$\frac{1}{s_0} + \frac{1}{s_i} = \frac{1}{f_2} \qquad [13]$$

where s_i and s_0 are defined in Figure 6. As previously discussed for convergent light, the Fourier transform will have an additional parabolic phase factor. This must be taken into account when constructing the filter H.

This configuration has some advantages over the 4-*f* correlator. By moving the object between the first lens and the filter, it is possible to control the scale of the Fourier transform. And because the first lens is imaging only a single point on the optical axis, only spherical aberration needs to be compensated for. In 4-*f* systems, the first lens must be corrected for all aberrations for the whole input plane, and for a given extent of the input object, the first lens must be larger than that for the convergent system in order to avoid space variance effects. Unfortunately, specially designed expensive lenses are required to correct for those aberrations, so well-corrected Fourier transform lenses for 4-*f* systems are expensive, whereas a simple doublet is sufficient to correct for on-axis spherical aberration, which is all that is required for the convergent light system. In addition, the convergent light system does not require an expensive collimator that is required in the 4-*f* system to obtain the parallel illuminating beam. Aberration correction requirements for the second lens are less stringent

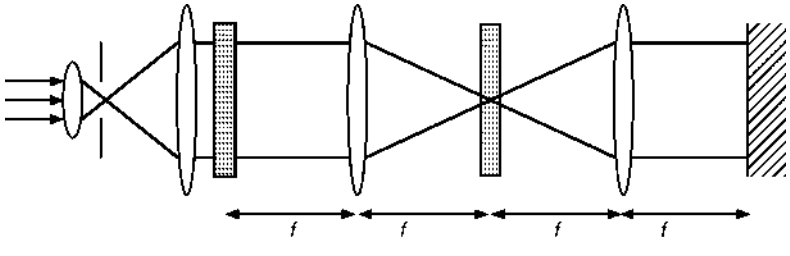

Figure 5 4-*f* correlator setup.

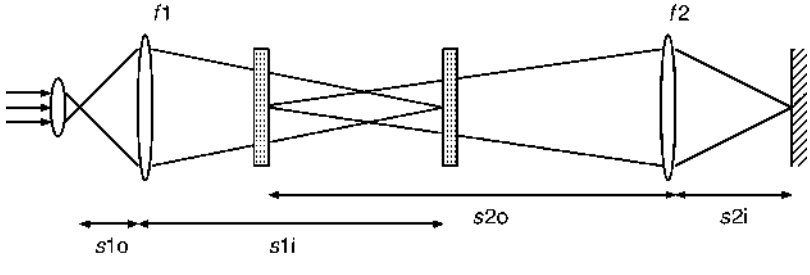

Figure 6 Convergent beam correlator setup.

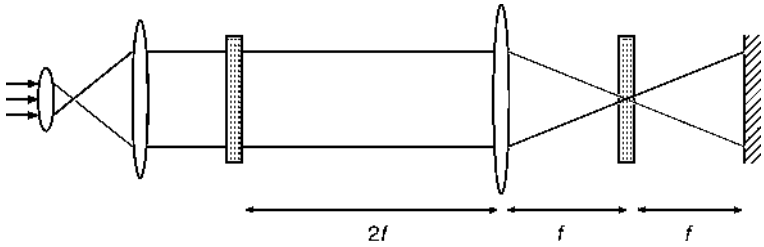

Figure 7 Single lens correlator setup.

than for the first, because in most cases, a slight deformation of the output plane can be acceptable, whereas comparable distortions in the Fourier plane can cause misregistrations between the Fourier transform and the filter that can seriously degrade the correlation.

Divergent beam correlators

Instead of placing the input object in the convergent beam, as in the previous case, another configuration that has been used places the input object in the divergent beam before the first lens. This configuration is rarely seen today, because it exacerbates the space variance and aberration problems of the 4-f system, and the control it allows over the scale of the Fourier transform can be achieved with fewer penalties by means of the convergent light system.

Single lens correlator

Consider an imaging system using only one lens illuminated with a parallel beam of coherent light (Figure 7). The input and output planes are located at distances $2f$ on either side of the lens – the shortest conjugate distances possible. If a filter $H(\mu, \nu)$ is placed at the focal plane of the lens on the object's Fourier transform, the correlation will appear on the output plane. This system requires satisfying impractical conditions: aberration corrections for both the Fourier transform and for the imaging operation, so it is rarely seen.

Filters

Many types of filters can be used for different applications. Table 1 lists some commonly used filters for pattern recognition. The classical matched filter yields the optimum signal to noise ratio in the presence of additive noise. The phase-only filter (POF) is frequently used because of its high discrimination capability. The inverse filter is sometimes used because it theoretically yields very sharp correlation peaks, but is very sensitive to noise, and when the spectrum $F(\mu, \nu)$ has some zeros, the realization of $1/F(\mu, \nu)$ can be impossible. A close relative of the inverse filter is the Wiener filter, which avoids the problem of the inverse filter by introducing a quantity $S(\mu, \nu)$ in the denominator of the filter function. The function $S(\mu, \nu)$ can sometimes be a constant. $H(\mu, \nu)$ is usually chosen to minimize the effect of noise or to optimize some parameter such as the peak to sidelobe ratio, or the discrimination ability of the system.

Table 1 Commonly used filters for image processing

Name	*Filter*
Matched filter	$H_{\mathrm{CMF}}(\mu, \nu) = F(\mu, \nu)$
Phase-only filter	$H_{\mathrm{POF}}(\mu, \nu) = \dfrac{F(\mu, \nu)}{\lvert F(\mu, \nu)\rvert}$
Inverse filter	$H_{\mathrm{INV}}(\mu, \nu) = \dfrac{1}{F(\mu, \nu)}$
Wiener filter	$H_{\mathrm{W}}(\mu, \nu) = \dfrac{F^*(\mu, \nu)}{\lvert F(\mu, \nu)\rvert^2 + S(\mu, \nu)}$

Joint Transform Correlators

In serial transform correlators, the filter is generally a complex function. Spatial light modulators are usually conceived with only one degree of freedom for each pixel, which can be optimized for amplitude modulation or for phase modulation. In order to achieve a fully complex filter, two modulators are required, or some spatial bandwidth of a modulator is sacrificed to encode a computer-generated hologram.

The joint-transform correlator (JTC) does not require a fully complex filter. The filter is stored in the spatial domain so there is usually no need for the computation of a filter in the spectral domain. In the JTC, the correlation of a scene with a reference image is contained in the power spectrum of an input plane consisting of the scene placed side by side with the reference. The correlation terms are revealed when the Fourier transform of the power spectrum is

computed. The joint transform correlation can be summarized in three steps:

1. The scene and the reference are located side by side in the input plane.
2. The Fourier transform of the input plane is computed or carried out optically by one of the methods previously described. The transform intensity is measured and constitutes the joint power spectrum. Some optional enhancements, based on spatial filtering and/or nonlinear processing on the power spectrum, can also be performed.
3. The Fourier transform of this joint power spectrum is carried out, yielding the correlation plane containing the correlation of the reference with the scene.

A mathematical analysis is required to understand the location of the correlation peaks. The side-by-side input scene $s(x, y)$ and reference image $r(x, y)$ are

$$i(x, y) = r(x - x_r, y - y_r) + s(x - x_s, y - y_s) \qquad [14]$$

The joint power spectrum $E(\mu, \nu)$ is equal to the modulus squared of the Fourier transform of the input plane:

$$\begin{aligned} I(\mu, \nu) &= R(\mu, \nu)\mathrm{e}^{-i2\pi(\mu x_r + \nu y_r)} + S(\mu, \nu)\mathrm{e}^{-i2\pi(\mu x_s + \nu y_s)} \\ E(\mu, \nu) &= \|I(\mu, \nu)\|^2 = \\ &\quad \|R(\mu, \nu)\|^2 + \|S(\mu, \nu)\|^2 \\ &\quad + R(\mu, \nu)S^*(\mu, \nu)\mathrm{e}^{-i2\pi[\mu(x_r - x_s) + \nu(y_r - y_s)]} \\ &\quad + R^*(\mu, \nu)S(\mu, \nu)\mathrm{e}^{-i2\pi[\mu(x_s - x_r) + \nu(y_s - y_r)]} \end{aligned} \qquad [15]$$

The joint power spectrum contains the Fourier transforms of the autocorrelation of the input scene, the autocorrelation of the reference, the correlation of the scene with the reference and its complex conjugate. The correlation peaks are obtained from the Fourier transform of the joint power spectrum:

$$\begin{aligned} C(x', y') = {} & R_{ss}(x', y') + R_{rr}(x', y') \\ & + R_{sr}(x' - (x_r - x_s),\ y' - (y_r - y_s)) \\ & + R_{sr}(x' + (x_r - x_s),\ y' + (y_r - y_s)) \end{aligned} \qquad [16]$$

where

$$R_{ab}(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(\xi, \zeta) b^*(\xi - x', \zeta - y') \mathrm{d}\xi \mathrm{d}\zeta \qquad [17]$$

Multiple targets and multiple references

The input scene of the JTC may contain multiple targets. Each target in the scene will correlate with the other targets. The additional correlation peaks can clutter the correlation plane, so some care must be taken when positioning the scene with respect to the reference. Let $s(x, y)$ contain multiple targets $s_i(x, y)$:

$$s(x, y) = \sum_{i=1}^{N} s_i(x - x_i, y - y_i) \qquad [18]$$

The joint power spectrum becomes:

$$\begin{aligned} E(\mu, \nu) = {} & \|R(\mu, \nu)\|^2 + \sum_{i=1}^{N} \|S_i(\mu, \nu)\|^2 + \sum_{i=1}^{N} \sum_{\substack{k=1 \\ i \neq k}}^{N} \\ & \times S_i(\mu, \nu) S_k^*(\mu, \nu) \mathrm{e}^{-i2\pi[\mu(x_i - x_k) + \nu(y_i - y_k)]} + \text{C.C.} \\ & + \sum_{i=1}^{N} R(\mu, \nu) S_i^*(\mu, \nu) \mathrm{e}^{-i2\pi[\mu(x_r - x_i) + \nu(y_r - y_i)]} + \text{C.C.} \end{aligned} \qquad [19]$$

where C.C. denotes the complex conjugate of the preceding term. The correlation plane will now be:

$$\begin{aligned} C(x', y') = {} & R_{rr}(x', y') + \sum_{i=1}^{N} R_{s_i s_i}(x', y') \\ & + \sum_{i=1}^{N} \sum_{\substack{k=1 \\ i \neq k}}^{N} R_{s_i s_k}(x' - (x_i - x_k),\ y' - (y_i - y_k)) \\ & + \sum_{i=1}^{N} R_{rs_i}(x' - (x_r - x_i),\ y' - (y_{ri} - y_i)) \\ & + \sum_{i=1}^{N} \sum_{\substack{k=1 \\ i \neq k}}^{N} R_{s_k s_i}(x' + (x_i - x_k),\ y' + (y_i - y_k)) \\ & + \sum_{i=1}^{N} R_{s_i r}(x' + (x_r - x_i),\ y' + (y_{ri} - y_i)) \end{aligned} \qquad [20]$$

The second lines of eqns [9] and [20] are the cross-correlation terms between the targets. In order to avoid confusion between reference–target correlations and target–target correlations, special care must be taken in positioning the reference with respect to the scene. All the target–target correlations will be inside a centered box that has twice the length and twice the width of the input scene, so the reference should be placed at least at twice the length or width of the scene.

Multiple targets in the input scene can be viewed as multiple references, and the correlation of multiple

references with multiple targets are all computed simultaneously. This property has led to applications in multiclass pattern recognition problems and in optical neural networks, where each neuron of a layer is coded as a reference object. This is one of the advantages of the JTC compared to serial correlators, since serial correlators are limited to one filter at a time.

Example

Here is an example of joint transform correlation. The input plane is shown in Figure 8a. The plane is composed of the input scene in the top third of the input plane and the reference is placed in the center. This input scene contains two targets. The joint power spectrum is shown in Figure 8b with its pixels mapped to a logarithmic intensity scale. The correlation plane is shown in Figure 8c and d with the central peak truncated. The central peak contains the autocorrelation terms R_{ss} and R_{rr}. On each side of the autocorrelation peak are the cross-correlations between the different targets of the input scene. The correlation between the reference image and the input scene is located in the top part of the correlation plane. The correlation plane is symmetrical with respect to the origin, so the correlation peaks are reproduced on the bottom half of the output plane.

Joint transform correlator architectures

Different architectures have been used for JTCs. Three setups are presented here. This is not an exhaustive list. The simplest optical setup shown in Figure 9 is an optoelectronic implementation of a power spectrum machine. A convergent beam

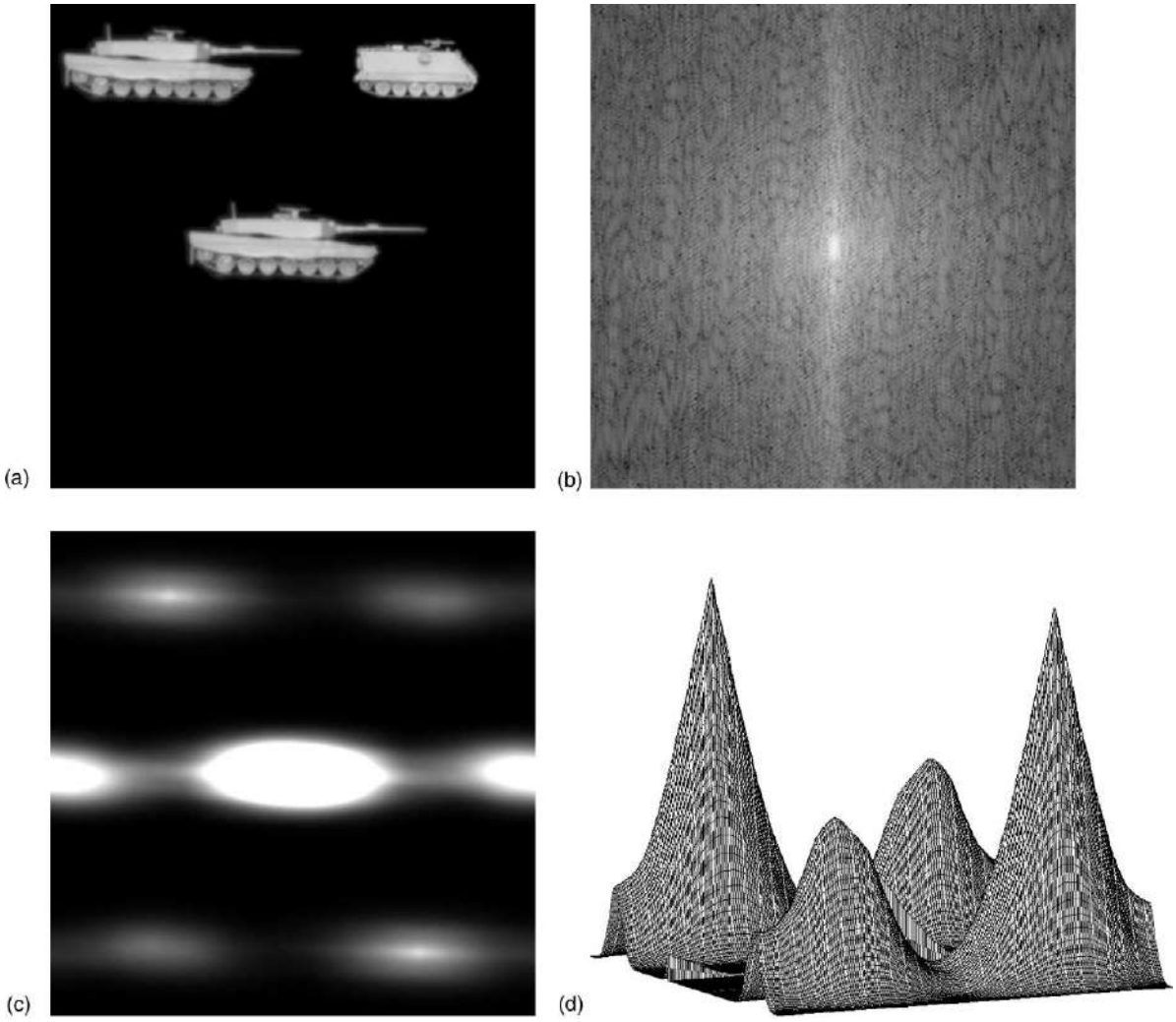

Figure 8 Joint transform correlation example. (a) Input scene; (b) joint power spectrum; (c) correlation plane (top view); (d) correlation plane (mesh view).

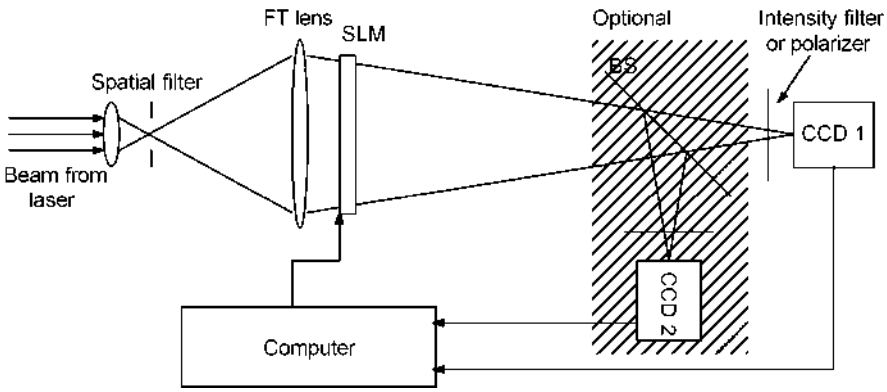

Figure 9 Joint transform correlator implementation as an optical power spectrum machine.

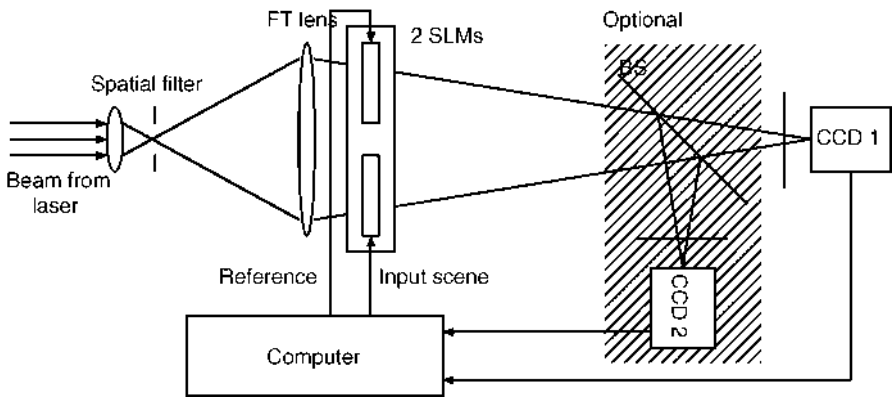

Figure 10 Optical JTC using two SLMs in the input plane.

optical Fourier transform system is used to produce the Fourier transform of an image displayed on a single spatial light modulator (SLM). A CCD camera is placed in the Fourier plane to measure the joint power spectrum. The joint transform correlation is a two-step process: the input plane containing the scene and the reference is first displayed on the SLM, and the camera simultaneously captures the joint power spectrum. In the second step, the joint power spectrum is displayed on the SLM, so the correlation plane is now on the camera. An optional second camera with intensity filters can be added to this setup. In this case, one camera serves to digitize the joint power spectrum during step 1 and the second camera captures the correlation plane at step 2 and transmits it to a display device. This optional configuration is useful to keep both the correlation plane and the joint power spectrum within the dynamic range of their specific cameras.

The setup shown in Figure 9 has two main disadvantages: the correlation is performed in two steps and requires some processing to construct the input plane, and only half of the spatial light modulator can be used to display the input scene. There are alternative setups to alleviate those drawbacks, but they require the use of two or three SLMs. Two modulators can be used in the input plane to simulate a full modulator for the scene, as shown in Figure 10. The original setup can also be doubled with one optical Fourier transform system for the joint power spectrum and one for the correlation plane (Figure 11). With three modulators, it is possible to perform the joint transform correlation in a single step with a full modulator for the input scene.

Comparison to serial transform correlators

The joint transform correlator has some advantages over the serial transform correlator. Since the phase of the Fourier transform is destroyed in the JTC, there are much fewer constraints on the design. So it is advantageous to use an optical convergent beam setup to obtain the Fourier transform. As mentioned before, there is no need for a fully complex spatial light modulator in the joint transform correlator.

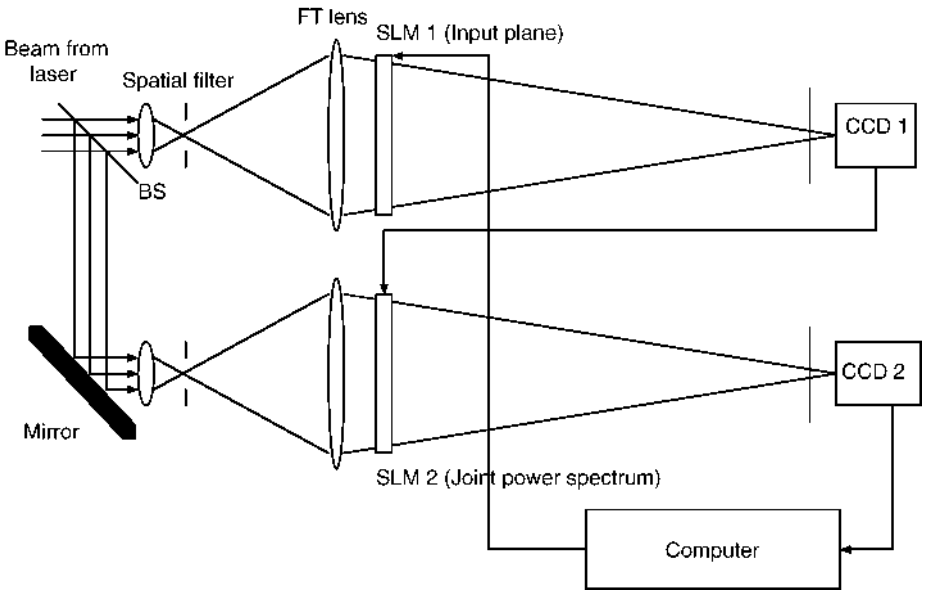

Figure 11 JTC implementation using two power spectrum machines.

This allows the use of commercially available spatial light modulators primarily designed for projectors or for displays, where phase modulation is not required. Since there is a wider market for such devices, they are cheaper.

For pattern recognition, the joint transform correlator has some properties that a serial transform correlator cannot reproduce. Some pattern recognition techniques require a linear combination of multiple amplitude correlations of the scene with different references. This is impossible with serial correlators because only the intensity of the correlation plane is detected. In a JTC, the joint power spectrum is linked to the correlation plane by the Fourier transform. So it is possible to obtain linear combinations of correlation planes by performing the corresponding operations on the joint power spectra. Some recent pattern recognition techniques such as morphological correlation and the sliced orthogonal nonlinear generalized (SONG) correlation, have been optically implemented using a JTC, but the required operations could not be carried out by means of serial optical correlators.

Whereas optical serial transform correlators always carry out linear filtering operations, optical joint transform correlators can be made to carry out nonlinear filtering. This is achieved by imposing a nonlinear characteristic curve on the joint power spectrum, instead of the usual squaring operation inherent to the detection process. The nonlinearity can be intrinsic to the components (saturation, dynamic range) or obtained by introducing electronics between the CCD camera and the second SLM, or by using an optically addressed spatial light modulator. Nonlinear processing can enhance the weights of the high frequencies in the correlation, resulting in sharper correlation peaks and improved discrimination. To illustrate this effect, the joint power spectrum of the example of Figure 8 has been modified with a logarithmic transformation:

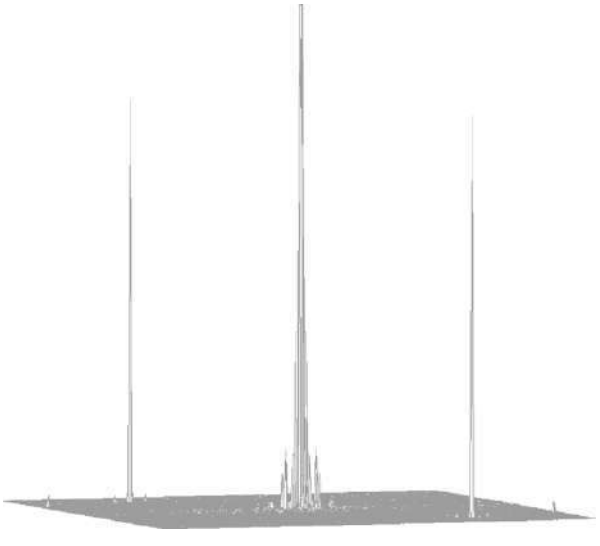

Figure 12 Correlation plane with a logarithmic transformation of the joint power spectrum.

$$E(\mu, \nu) = \ln(1 + \beta \|I(\mu, \nu)\|^2) \qquad [21]$$

with $\beta = 0.25$. The correlation plane for the transformed joint power spectrum is shown in Figure 12.

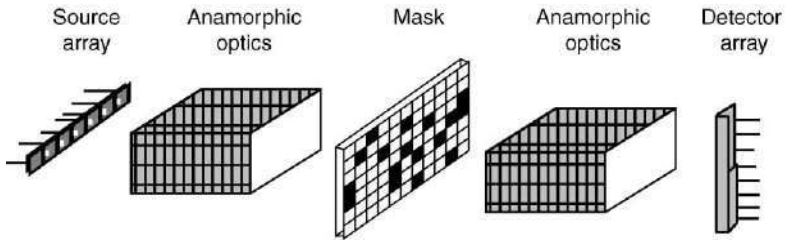

Figure 13 Optical vector–matrix multiplier.

It is clear that the discrimination has improved over that of the previous case.

Optical Vector–Matrix Multiplier

An operation that is frequently required in information processing is the product of a matrix and a vector. This can be accomplished by means on anamorphic optics as shown in Figure 13.

Here the input vector is a set of vertical fans of light from the source array, that are spread out, each over one row of the matrix mask by means of anamorphic optics. The inverse operation is carried out in the second half, so that the vertical array of spots in the output contains the vector–matrix product between the intensities of the source array and the transmissivities of the mask array.

A vector–matrix product can also be carried out by means of a conventional optical correlator as shown in Figure 14.

The input vector **a** is placed in a parallel beam at the focal plane of a conventional 2D correlator. A hologram in the Fourier plane contains a vector array **H**. The vector–matrix product **aH** appears along the axis in the correlation plane.

The above figure shows an input vector $f(x,b)\delta(y-b)$ containing the vector components $f(x1, x2, \ldots, xN)$ arrayed along the horizontal axis x, and shifted a distance a from the horizontal axis. If the Fourier transform of the mask in the filter plane is $h(x, y)$, then the output plane has the distribution:

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\mu,\upsilon)\delta(\mu-b)h(x+\mu,y+\upsilon)\mathrm{d}\mu\mathrm{d}\upsilon \qquad [22]$$

$$g(x,y)=\int_{-\infty}^{\infty} f(\mu,a)h(x+\mu,b-y)\mathrm{d}\mu \qquad [23]$$

$$g(x,b)=\int_{-\infty}^{\infty} f(\mu,b)h(x+\mu,b)\mathrm{d}\mu \qquad [24]$$

$$g(0,b)=\int_{-\infty}^{\infty} f(\mu,b)h(\mu,b)\mathrm{d}\mu \qquad [25]$$

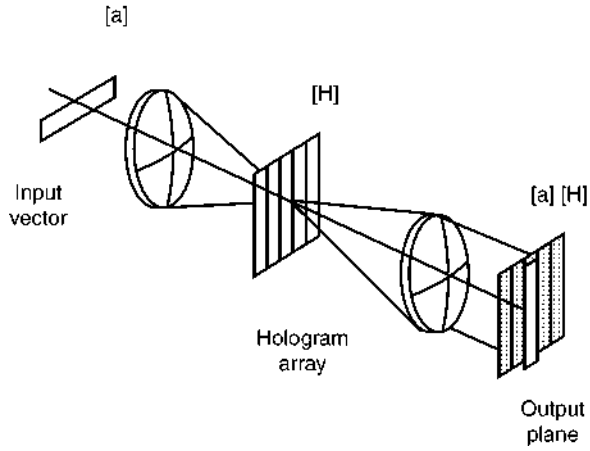

Figure 14 A vector–matrix product using a conventional 2D optical correlator.

So the output value along the y-axis at a distance a from the horizontal axis is equal to the scalar product of the vector $f(x,b)$ with the vector $h(x,b)$. Now if input vectors are stacked up one over the other in the input plane, each one will yield along the output y-axis a value that is the scalar product of the corresponding vector with the vector **h**. The set of values along the y-axis is therefore the vector–matrix product of the matrix constituted by the set of vectors in the input plane, with the vector **h** that is the impulse response of the filter. The cost for this operation is some considerable loss of light, since only the light along the y-axis is used in the output.

Applications of the optical vector–matrix product are optical crossbar, numerical matrix operations such as the solving of linear equations, optical neural networks, spectrum analysis and discrete Fourier transforms.

Conclusion

It has been estimated that serial coherent optical correlators should be able to attain data rates of up to 10^{12} operations per second, using the fastest spatial

light modulators available. Joint transform correlators are somewhat slower, due to the input plane being divided between the object and the reference, to geometrical constraints in placement of targets, and due to the need to use electronics to carry out some of the operations. Until now, speeds have been limited by the data rates at the input, but as the speeds of optically addressed SLMs increase, the problem could be shifted to the output, where the lack of optical high spatial bandwidth threshold devices could cause a bottleneck.

List of Units and Nomenclature

A	Vector or matrix
CCD	Charge coupled device
$F(\mu, \nu)$	Fourier transform of $f(x, y)l$
FT	Fourier transform
FT^{-1}	Inverse Fourier transform
H^*	Complex conjugate of H
JTC	Joint-transform correlator
POF	Phase-only filter
R_{ab}	Correlation between functions a and b
(x, y)	Spatial coordinates
λ	Wavelength of light
(μ, ν)	Spatial frequencies
$*$	Correlation
SLM	Spatial light modulator
SONG	Sliced orthogonal nonlinear generalized correlation

See also

Coherence: Coherence and Imaging. **Diffraction:** Fraunhofer Diffraction; Fresnel Diffraction. **Fourier Optics**. **Geometrical Optics:** Aberrations. **Holography, Techniques:** Computer-Generated Holograms. **Imaging:** Information Theory in Imaging. **Information Processing:** Incoherent Analog Optical Processors; Optical Neural Networks.

Further Reading

Arsenault HH, Szoplik T and Macukow B (eds) (1989) *Optical Processing and Computing*. New York: Academic Press.

Bartelt H (1987) Unconventional correlators. In: Horner JL (ed.) *Optical Signal Processing*, pp. 97–128. New York: Academic Press.

Garcia-Martinez P, Arsenault HH and Ferreira C (2001) Time sequential nonlinear correlations for optical pattern recognition. Optoelectronic Information processing: optics for information systems. *Proceedings of SPIE* CR81-12: 186–207.

Goodman JW (1988) *Introduction to Fourier Optics*. New York: McGraw-Hill.

Javidi B (1994) Nonlinear joint transform correlators. In: Javidi B and Réfrégier P (eds) *Workshop on Optical Pattern Recognition*. Bellingham, WA: SPIE – The International Society for Optical Engineering, PM-12.

Javidi B and Horner JL (eds) (1994) *Real-Time Optical Information Processing*. New York: Academic Press.

Javidi B and Johnson KM (eds) (1996) *Optoelectronic Devices and Systems for Processing*. Bellingham, WA: SPIE – The International Society for Optical Engineering, CR65.

Lu XL, Yu FTS and Gregory D (1990) Comparison of Vander Lugt and joint transform correlators. *Applied Physics B* 51: 153–164.

McAulay AD (1991) *Optical Computer Architectures*. New York: Wiley & Sons.

Preston K (1972) *Coherent Optical Computers*. New York: McGraw-Hill.

Vander Lugt AB (1992) *Optical Signal Processing*. New York: John Wiley & Sons.

Free-Space Optical Computing

A A S Awwal, Lawrence Livermore National Laboratory, Livermore, CA, USA

M Arif, Infineon Technologies, Longmont, CO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Free space optical computing encompasses a vast field, from digital and discrete to analog and hybrid optical computing. This article gives an overview of various algorithms, such as shadow-casting, symbolic substitution to nonlinear optics-based processors. Some of the core devices, such as modulators and nonlinear logic elements, are discussed. The optical computing concepts, originated with nonlinear optics and analog optical computing, are more commonly known as optical signal processing. Now they have culminated into digital optical computing, opening the door to highly specialized parallel computing architecture.

Parallel computation is a direct approach to enhance the speed of digital computations which are otherwise inhibited by the sequential processing bottleneck. Physical limitations – such as limited interconnections, interactions amongst electrons, with the consequent inability to share a data bus simultaneously, and constant propagation delay of electronic circuits – place an upper bound on the speed of the fastest operation possible with an electronic computer, without a modification to the computing architecture. Optics provides an architectural advantage over electronics, due to the fact that the most natural way to use optics lends itself to a parallel implementation. Optical signals propagate in parallel, cross each other without interference (whereas charged electrons interact heavily with each other), have the lowest possible propagation delay of all signals, and can provide million channel free-space interconnections with simple lenses (whereas an unrealistic number of connections are necessary for electrons). Thus, optics is an excellent candidate for implementing parallel computing machines using the 3D space connectivity. While electronic architecture is naturally 2D, optical free space architecture is 3D, exploiting the third dimension of space. The communication bottleneck forces electronic computers to update the computation state space sequentially, whereas the use of optics will empower one to change the entire computation space in parallel.

Historical Perspective

Historically, digital optical computing was proposed as a natural extension of nonlinear optics. Just as transistors, which are nonlinear electronic devices, led to the realization of the computing logic using transistors, it was envisioned that nonlinear optical effects would give rise to optical transistors which would lead to development of optical logic gates. One of the limitations of nonlinear optical device-based computation is the high laser power requirement to demonstrate any nonlinear optical effect.

One of the effects used to demonstrate nonlinearity is to change the refractive index with the incident intensity. In other words, most materials will behave linearly with low optical energy. The refractive index expressed as $n = n_0 + n_2 I$ is nominally linear because the nonlinear co-efficient n_2 is usually very small. In the presence of high power, the $n_2 I$ is significant enough to cause an optical path change by altering the refractive index. When such a medium is placed inside a Fabry–Perot resonator, it will allow one to control the transmission of the filter by changing the incident intensity. Thus, at high power, it shows a nonlinear behavior or bistability. The term optical bistability was coined, which meant the presence of two stable states of output for the same intensity value.

Another development took place at the University of Michigan by the development of a precision optical processor; this was an optical processor which could process radar signals using optics, this is the example of an analog optical computing processor.

A third area which is in between the digital and the analog are the discrete level devices which most often take the form of matrix-based processors. Free space optical matrix vector multipliers are examples of such processors.

The latest field of optical computing research is in the algorithms and architecture area. The motivation here is to develop new algorithms which take advantage of optics. The thing that is lacking is the presence of appropriate device technology that can implement the algorithm in a given architecture. In terms of practical realization, analog optical computing is the most successful one. Adaptive optic processors have found success both in astronomy as well as vision sciences, and more recently in optical imaging through the atmosphere.

Development of free space optical computing has two main driving forces: the technology (device) base and the algorithm base. In this classification, since the field is in its infancy, both device-based classes and architectural (algorithm) classes are listed side-by-side as separate entities, ignoring the natural overlap that may be present in some instances. For example, nonlinear optical devices could be implemented into a variety of architecture that is listed below. Holographic processor is another technology-based class while some other architectures, such as, symbolic substitution may involve use of holograms, but it does not exclude it being a class by itself. The following list gives a classification of optical free space computing:

1. Digital Optical Computing:
   (a) Pattern coded computing or cellular logic:
      (i) Optical shadow-casting processor;
      (ii) Symbolic substitution;
      (iii) Theta logic based computation.
   (b) Parallel arithmetic circuits:
   (c) Hybrid optical computing:
   (d) Nonlinear optical device based:
   (e) Threshold logic.
2. Discrete optical computing:
   (a) Optical fuzzy logic:
   (b) Optical matrix processors:
      (i) Vector matrix multiplier;
      (ii) Differential equation solver;

(iii) Linear algebra processor;
(iv) DFT processor;
(v) Programmable logic array;
(vi) Associative memory processor.

(c) Optical neural network.

3. Analog optical computing:
   (a) Adaptive optics processor;
   (b) Partially coherent processing (white light processing):
   (c) Fourier transform based processor:
   (d) Nonlinear contact screen based:
   (e) Holographic processors:
   (f) Linear optical processors (image differentiation).

We start by discussing some of the device based classes and their underlying concepts: (a) nonlinear optics; (b) SLMs; and (c) hologram based.

Core Device Technologies Used for Free-Space Optical Computing

One of the most difficult challenges of optical computing (OC) is inventing devices that can realize aforementioned key advantages of optics. There exist, however, several key technology areas, where a breakthrough in any one of these will pave the way to real-world optical computing. We discuss here some of the major device technologies that relate to OC.

Spatial light modulator (SLM) converts information from different forms to the optical domain. As a result, SLM is used as an input device to an optical computing system. High contrast ratio, very high speed, and parallel addressing scheme are desirable properties of an SLM as an input device. For example, the SLM could be used to input a 2-dimensional signal or an image into a Fourier transform-based processor. Several applications of SLM are possible:

1. *Optical encoding*: Pattern coded logic can perform logic operations if inputs can be encoded using certain pattern;
2. *Encoding complex filter*: It could encode phase, amplitude or both; and
3. *As phase modulator*: for adaptive optics.

Two types of SLMs are liquid-crystal based and micro-electromechanical (MEM) based. The former uses liquid crystals to change the phase of transmitted or reflected light leading to modulation of phase. Although the most successful commercial application of MEMs is its use in consumer electronics, such as projection TVs or video projectors, in free-space optical computing it can be used as:

(a) an interconnection device;
(b) an SLM for matrix; or
(c) a phase modulating device in AO close loop control of optical phase aberrations or phase only filter.

The heart of the MEMs is tiny electrically controllable mirrors which can steer the incoming beam in a different direction or bend to create a relative optical path difference for the reflected beam.

Algorithms, where an SLM plays a major role, are pattern coded computing, optical matrix processors, adaptive optic processors, partially coherent processing, Fourier transform based processors, and linear optical processors.

Nonlinear Optical Element

A nonlinear optical element is an architecture or device structure which allows for utilization of the nonlinearity of the optical material that produces detectable changes of optical intensity. Figure 1 shows the basic block diagram for an optical computing basic element. To achieve an optical logic gate, one needs a scheme where light can be controlled by light, as current is controlled by current in a transistor. An optical device showing optical bistability can be used for this purpose. Optical bistability is defined as the situation where two different output intensities are possible for a given input intensity.

For example, as shown in Figure 2 below, at the intensity level of C, there are two stable states, one HIGH and one LOW. Just as the transistor has two stable states ON and OFF, the optical bistable device exhibits two states ON and OFF. The fact that such nonlinear behavior could implement logic behavior

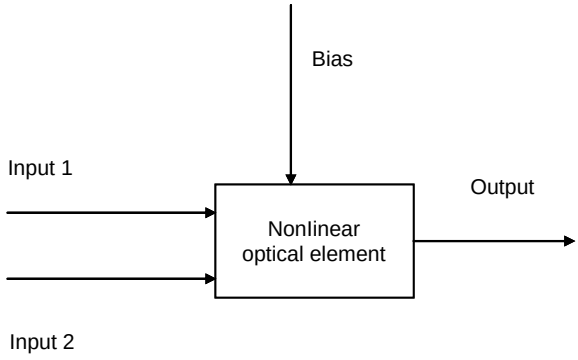

Figure 1 Block diagram of an optical logic gate using nonlinear optical element.

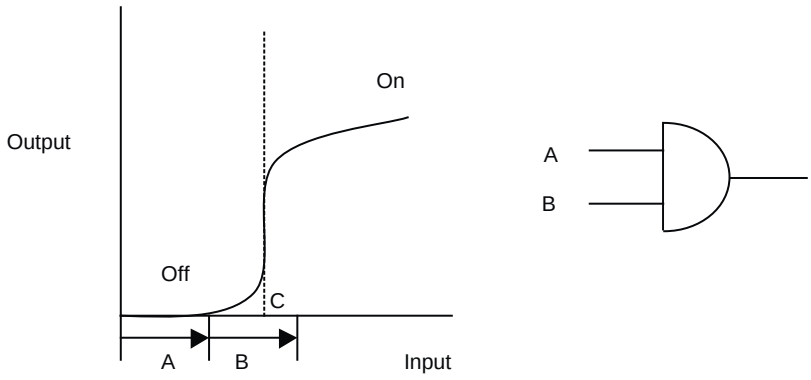

Figure 2 Bistable behavior, a typical AND behavior with inputs A and B.

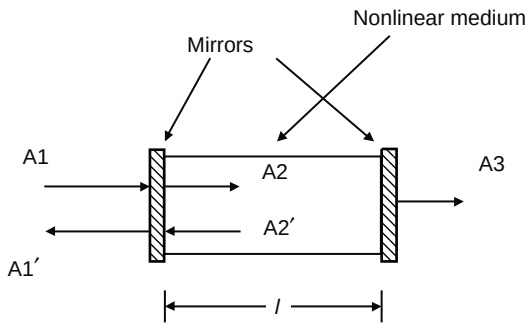

Figure 3 Bistable optical device, nonlinear medium in a Fabry–Perot cavity.

can be explained as follows. Assume that a two input device possesses a nonlinear characteristic as shown in Figure 2. Assume that the input intensity value of 1 is indicated by an intensity level equal to A. When both A and B are equal to 1, then the combined intensity switches the device to HIGH output state. However, if only one input is present, the output will be LOW. Thus, this bistable device acts as a logical AND gate.

To achieve optical bistability, two features are required, namely, nonlinearity and feedback. One way to achieve this is by placing a nonlinear medium in a Fabry–Perot cavity. Figure 3 shows a schematic of these devices. In Figure 3, A_1 is the incident field, A_1' is the reflected field, A_2 and A_2' are the forward and backward fields inside the cavity, respectively, and l is the cavity length.

The coupled wave equation for this system is given by

$$A_2' = \rho A_2 e^{2ikl-\alpha l} \qquad [1a]$$

$$A_2 = \tau A_1 + \rho A_2' \qquad [1b]$$

where, ρ is the amplitude reflectivity and τ is the amplitude transmissivity. The wave vector $k = n\omega/c$ and α is the absorption coefficient. In deriving the above equations, it was assumed that k and α are spatially invariant and real. Solving for A_2, from eqns [1a] and [1b] yields

$$A_2 = \frac{\tau A_1}{1 - \rho^2 e^{2ikl-\alpha l}} \qquad [2]$$

Equation [2] is the so-called Airy's equation which describes the properties of the Fabry–Perot cavity. From eqn [2] it can be seen that if k or α is a sufficiently nonlinear function, this device will show bistability. In general, both k and α can show nonlinearity. Hence optical bistability can be either absorptive or dispersive.

Absorptive Bistability

If the absorption coefficient α depends nonlinearly on the intensity, assuming k is constant, after some manipulation, eqn [2] can be written as

$$I_2 = \frac{TI_1}{[1 - R(1 - \alpha l)]^2} \qquad [3]$$

where, $I_i \propto |A_i|^2$, $T = |\tau|^2$ and $R = |\rho|^2$. In deriving eqn [3], it was assumed that the mirror separation is such that the system is in resonance and $\alpha l \ll 1$. If α depends on the intensity, it can be assumed that it follows the relationship of two level saturable absorbers, for simplicity, which is

$$\alpha = \frac{\alpha_0}{1 + I/I_s} \qquad [4]$$

where, α_0 denotes the unsaturated absorption coefficient, I is the local value of intensity, and I_s is the saturation intensity. Using eqn [4] in eqn [3] yields

$$I_1 = TI_2\left(1 + \frac{C_0}{1 + 2I_2/I_s}\right)^2 \qquad [5]$$

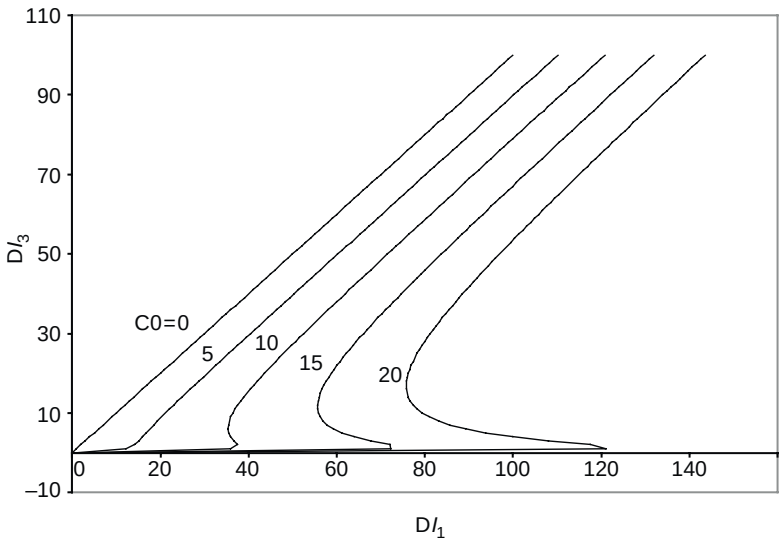

Figure 4 The input–output relationship of absorptive bistability predicted by eqns [5] and [6].

where, $C_0 = R\alpha_0 l/(1-R)$. Also the output intensity I_3 is related to I_2 by

$$I_3 = TI_2 \qquad [6]$$

Using eqns [5] and [6], it can be shown that there can be multiple solutions to the above input–output relationship which predicts the optical bistability. Figure 4 shows a plot of DI_1 versus DI_3, where $D = 2/TI_s$, as a function of various C_0. As can be seen from Figure 4, for a high enough value of C_0, the system possesses multiple solutions.

Dispersive Bistability

The intensity dependence of the refractive index can be written as

$$n = n_0 + n_2 I \qquad [7]$$

If the refractive index n varies nonlinearly with intensity, assuming $\alpha = 0$ and using eqn [7], eqn [2] can be written as

$$A_2 = \frac{\tau A_1}{1-\rho^2 e^{2ikl}} = \frac{\tau A_1}{1-Re^{i\delta}} \qquad [8]$$

where, $\delta = \delta_0 + \delta_2$. δ_0 and δ_2 are given by

$$\delta_0 = \varphi + 2n_0\frac{\omega c}{l}$$
$$\delta_2 = 2n_2 I\frac{\omega c}{l} \qquad [9]$$

where, φ is the phase associated with ρ and the fact that $k = n\omega/c$ is used.

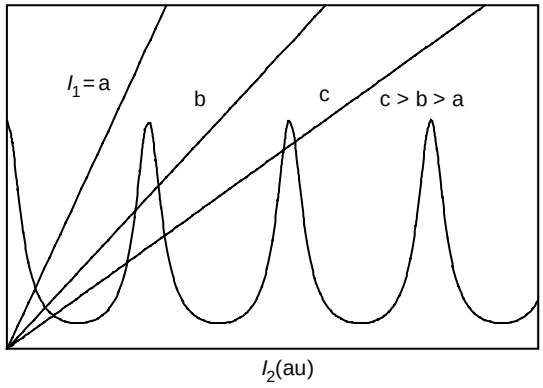

Figure 5 Plots of both sides of eqn [10] as a function of I_2 and for increasing values of input intensity I_1. The oscillatory curve represents the right-hand side of eqn [10] and the straight lines represent the left-hand side with increasing values.

After some manipulation, eqn [8] can be rewritten in terms of intensity as

$$\frac{I_2}{I_1} = \frac{1/T}{1+(4R/T^2)\sin^2(\delta/2)} \qquad [10]$$

and

$$\delta = \delta_0 + \left(4n_2\frac{\omega}{c}l\right)I_2 \qquad [11]$$

Equation [10] can be solved graphically to show the introduction of bistable behavior as a function of input intensity. Figure 5 shows the plots of both sides of eqn [10] as a function of I_2 and for increasing values of input intensity I_1. The oscillatory curve represents the right-hand side of eqn [10] and the

straight lines represent the left-hand side with increasing values of I_1. It can be readily seen from Figure 5 that at sufficiently high values of I_1, multiple solutions are possible for eqn [10]. This again will give rise to the bistable behavior.

Nonlinear Interference Filters

The optical bistability from the nonlinear interference filters (or etalons) can be used to achieve optical logic gates. The transmission property of these etalons can be modified by another optical beam, known as a bias beam. A typical logic gate and the transmission property are shown in the following figure. If the bias beam is such that, $P_{\text{switch}} - (P_a + P_{\text{bias}}) < P_b$, then the transmitted beam shown in Figure 6b readily follows an AND gate. The truth table for such configurations is shown in Figure 6d.

Four-Wave Mixing

Four Wave Mixing (FWM) can be used to implement various computing functions. Using the photorefractive effect and FWM in a third-order nonlinear optical material, optical phase conjugate beams can be obtained. A schematic of FWM and phase conjugation is shown in Figure 7. The forward pump beam interferes with the probe beam to produce a grating in the photorefractive material. The backward pump beam is then diffracted to produce a beam which retraces the path of the probe beam but has an opposite phase. It can be seen from Figure 7, that the optical phase conjugation can be readily used as an AND logical operator. By using the four-wave mixing, an optical counterpart of transistors has been introduced. FWM have also been used in AND-based optical symbolic substitution (OSS) operations. FWM is also used to produce matched filter-based optical correlators.

Another device showing the optical bistability is the self electro-optic effect device (SEED). SEEDs were used to implement optical logic circuits. The SEEDs are produced by putting multiple quantum wells (MQW) in the intrinsic region of a p-i-n structure. The MQWs are created by placing alternating thin layers of high (barriers) and low (wells) bandgap materials. An electric field is applied to the SEEDs externally. The absorption coefficient is a function of applied bias voltage. When a light beam is incident on these devices, it creates a photocurrent, consequently lowering the bias voltage across the MQW. This, in turn, increases the absorption and when the input optical power is high enough, the system sees a peak absorption and suddenly switches to a lower state and thus shows the bistable behavior. A schematic of SEED is shown in Figure 8.

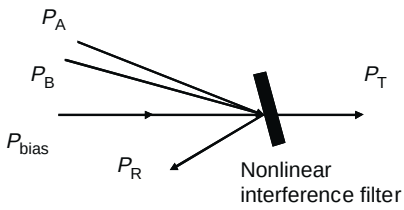

(a)

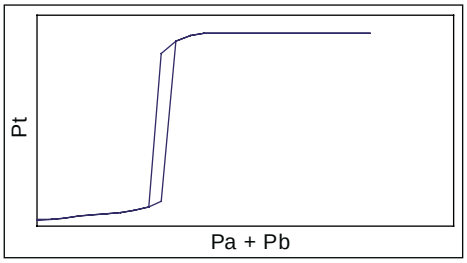

(b)

Truth table for an AND function from the above setup from the transmitted beam

P_A= input 1	P_B = input 2	P_T = output
0	0	0
0	1	0
1	0	0
1	1	1

(d)

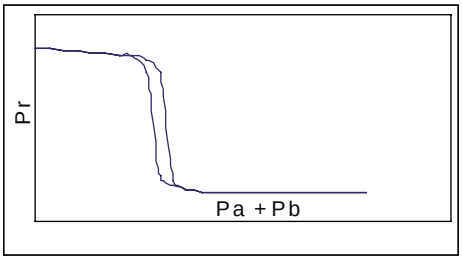

(c)

Figure 6 A typical optical logic gate configuration using nonlinear interference filter: (a) schematic; (b) the transmitted power output; (c) reflected power output (both as functions of inputs); and (d) the truth table for AND operation with beams A and B as inputs.

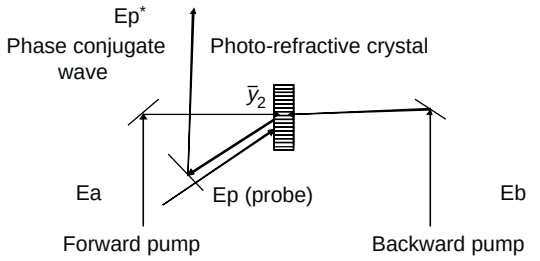

Figure 7 Four-wave mixing setup.

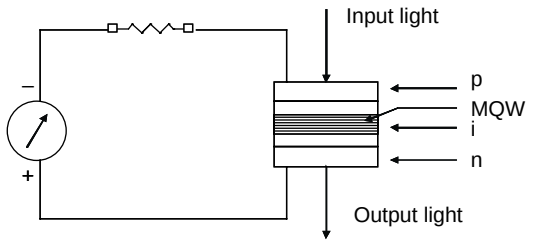

Figure 8 Schematic of self electro-optic effect devices.

Replacing the resistive load by a photodiode, a diode-based SEED (D-SEED) can be constructed, which is analogous to the way diode transistor logic was developed from resistor transistor logic in electronics. Electrically connecting two MQW p-i-n structure symmetric SEEDs (S-SEED) are constructed, which are less sensitive to optical power fluctuations and effectively provide isolation between input and output.

Pattern Coded Optical Computing: Symbolic Substitution or Cellular Logic

This is a keyword which is equivalent to simple digital logic operation. In digital logic, a truth table can describe the relationship between input and output variable. In symbolic substitution logic, this relationship is expressed as a 2D pattern rule. In any specific rule (equivalent to a row in the truth table) of SS logic, a specific 2D pattern is identified and replaced by another pattern. For example, in the case of binary logic, we can express the addition of two binary bits with the following truth table:

Rule number	*Input*		*Output*	
	A	*B*	*C*	*S*
1	0	0	0	0
2	0	1	0	1
3	1	0	0	1
4	1	1	1	0

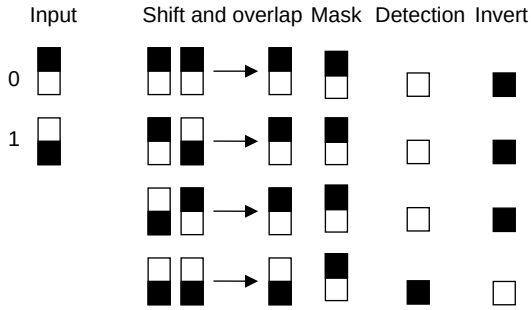

Figure 9 Symbolic substitution example, the two inputs are coded using dual rails. The second column shows the four combinations of inputs (00, 01, 10, and 11). They are shifted and overlapped creating the fourth column. A mask detects the output as a dark pixel appearing through the transparent opening. The dark is inverted to create a detection output.

Here A and B are the two digits to be added, C and S are the sum and the carry bit, respectively. To convert this operation to an SS operation, we can say that a two bit binary pattern is identified and replaced by a sum bit and a carry bit. If we now express the above in terms of symbolic substitution, we will say let 0 and 1 be represented by the following 2D symbol. Then the rule for the carry bit generation is shown in Figure 9.

Note that each of these rules has to be implemented individually. However, one can implement these by only generating the 1 output or the zero output, whichever leads to the minimum rule. Here implementing the last rule is convenient, one can make a copy of the whole pattern and then overlap it with its left shift version. This will result in a dark pixel at the origin. This is inverted and then produces a detection output. Natural extension of SS is in signed digit computing, where carry free addition can be achieved, allowing parallel implementation of numerical algorithm using free space optics.

Optical Shadow Casting (OSC)

All digital logic is nonlinear. The nonlinearity of the digital logic is converted into nonlinearity of coding. The OSC system originally proposed by Tanida and Ichioka uses two-dimensional spatially encoded patterns as data input and light emitting diodes (LEDs) as light sources. In the original shadow-casting system, the inputs were represented by vertical and horizontal stripes of opaque and transparent bars. With this system, programmability was attained by changing the LED pattern and it was possible to realize 16 logical operations between two

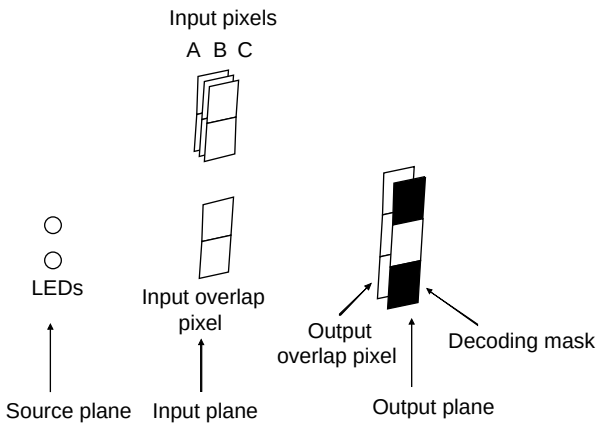

Figure 10 An optical shadow-casting system.

variables as well as a half addition as an example of arithmetic operation. Later, polarized codes were introduced into shadow-casting systems. This improved technique, called the polarization-encoded optical shadow casting (POSC), enabled complex combinational logic units like binary and trinary full adders, as well as image processing applications, to be realized.

The lensless OSC system, as shown in Figure 10, consists of spatially encoded 2D binary pixel patterns as the inputs. The input patterns are kept in close contact at the input plane and the resulting input overlapped pixel pattern is illuminated by a set of LEDs from the source plane. The light originating from each of the LEDs located at the source plane produces a shadow of the input overlap pixel pattern at the output plane. The overlap of these relatively displaced shadows results in an output overlap pixel pattern at the output plane. A decoding mask placed at the output plane is then used to spatially filter and detect the logical output. To prevent crosstalk, specific spacing among the different processing elements must be maintained.

Discrete Processors: Optical Matrix Processor

Most of the discrete optical processing schemes revolve around optical matrix–vector multiplication. Many numerically useful operations such as: (a) linear algebraic operations; (b) signal processing algorithms; (c) Boolean logic; (d) digital multiplication by analog convolution; and (e) neural networks, can be formulated as matrix–vector multiplication. An example of the digital multiplication by analog convolution (DMAC) can be understood by the following numerical example:

```
1   3
1   4
    5   2
1   3
1   8   2
_____________
                                  Convolution
                                        3        Original number
                Flipped number 4    1              1
Ordinary
multiplication                      4   1          4+3=7
                                        4   1      12
                                  Overlap with     Partial
                                  the original     products
                                                   added

1  7   12  = 1×100 + 7×10 + 12×1 = 182
```

The two numbers are convolved as shown above. This is done by reversing one of the numbers (4 1), while keeping the other number in the same order (1 3), then sliding the reversed number to the right, and collecting the element by element products. The second partial products are 4×1 and 1×3 which are then added to give the convolution results. Surprisingly, the weighted number in mixed radix (since numbers that have digits higher than 9, are no longer decimal) has the same value as the product shown in the upper left. It is possible to have fast parallel convolution optically, resulting in fast parallel multiplier. However, an efficient electronic post processor is necessary to convert the mixed radix numbers to the radix of the to-be-multiplied numbers.

A basic optical matrix–vector multiplier is shown in Figure 11. Here the vector, represented by an expanding column of light, is projected onto the 2D matrix, possibly represented by a glass plate or electronically addressable spatial light modulator. As the light representing the vector passes through the transmitting medium, each element of the vector (**I**) gets multiplied by a column of matrix represented by the transmittance of the matrix (T). The product (**I**T)

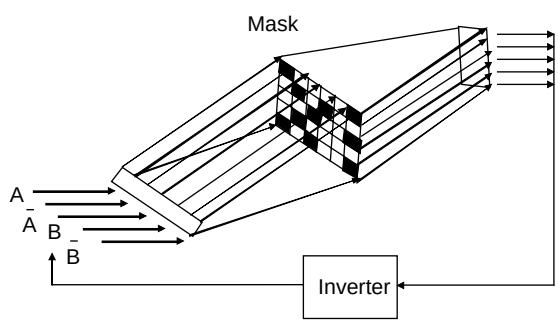

Figure 11 A matrix–vector multiplier.

is integrated horizontally using a cylindrical lens. This leads to an addition operation of the partial products, resulting in the generation of the output product vector. Mathematically:

$$y = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_1 + a_{12}b_2 + a_{13}b_3 + a_{14}b_4 \\ a_{21}b_1 + a_{22}b_2 + a_{23}b_3 + a_{24}b_4 \\ a_{31}b_1 + a_{32}b_2 + a_{33}b_3 + a_{34}b_4 \\ a_{41}b_1 + a_{42}b_2 + a_{43}b_3 + a_{44}b_4 \end{bmatrix} \qquad [12]$$

The above can be achieved if the vector b is expanded as

$$\begin{bmatrix} b_1 & b_2 & b_3 & b_4 \\ b_1 & b_2 & b_3 & b_4 \\ b_1 & b_2 & b_3 & b_4 \\ b_1 & b_2 & b_3 & b_4 \end{bmatrix}$$

and a point-by-point multiplication is performed between the expanded vector and the original matrix. Then the addition is performed by the cylindrical lenses.

Another simple application of vector matrix multiplier is in digital logic, known as programmable logic array (PLA). PLA is an electronic device that consists of a set of AND gates followed by a set of OR gates, which can be organized in AND–OR format suitable for realizing any arbitrary general purpose logic operation. AND logic is performed when light representing a logic level is passed through a transparency representing the other variable; OR is achieved by converging light to a common detector.

$$\text{Assume a function } F = AC + \bar{A}BD + A\bar{B}C \qquad [13]$$

where $\bar{A}$ represents the logical inverse of A. Using DeMorgans' theorem in digital logic, this can be expressed as

$$F = \overline{(\bar{A} + \bar{C})} + \overline{(A + \bar{B} + \bar{C})} + \overline{(\bar{A} + B + \bar{C})}$$

$$= \bar{y}_1 + \bar{y}_2 + \bar{y}_2 \qquad [14]$$

Now the y_is can be generated by the vector matrix product, as shown below:

$$\text{where} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} A \\ \bar{A} \\ B \\ \bar{B} \\ C \\ \bar{C} \\ D \\ \bar{D} \end{pmatrix} \qquad [15]$$

Note here that the AND has been converted to OR, which needs to be inverted and then ORed to generate the function F. A setup shown in Figure 11 can be used to accomplish this. In step 1, the y_is are generated, and then in step 2 they are inverted and summed by a second pass through the same system with a different mask. Other applications of vector–matrix multipliers are in neural networks, where the interconnection weights are represented by the matrix and the inputs are the vector.

Analog Optical Computing

Correlation is a way to detect the presence of signal in additive noise. Optics offer a fast way of performing correlations. The practical motivation of this area comes from the fact that a lens performs a Fourier transform of a two-dimensional signal at its focal plane. The most famous analog optical computing is known as optical matched filter. The theoretical basis of an optical matched filter is in the Fourier theorem of matched filtering which states that the cross correlation between two signals can be formed by multiplying the Fourier transform of the signal and the complex conjugate of the second signal and followed by a subsequent Fourier transform operation. Mathematically:

$$\text{Correlation}(f, g) = \mathrm{F}^{-1}[\mathrm{F}\{f\}^{*}\text{conjugate}(\mathrm{F}\{g\})] \qquad [16]$$

where F represents a Fourier transform operation and F^{-1} represents an inverse transform. Since the lens can easily perform the Fourier transform, and multiplication can be performed by light passing through a medium, the only problem that has to be

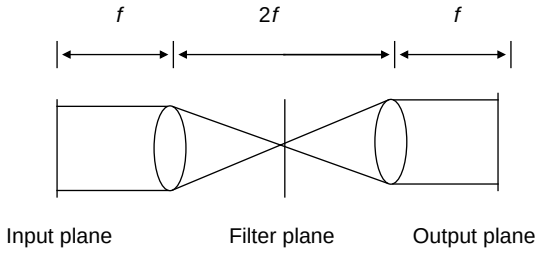

Figure 12 An optical 4-*f* set up. Lenses perform Fourier transform at their focal plane.

solved is to represent the signal and the filter in the optical domain. Since an SLM can represent both the real signal and the complex filter, thus the whole operation can be performed in a 4-*f* setup as shown in Figure 12. The input is displayed in the input plane by an SLM, the first lens performs a Fourier transform. This filter (conjugate F{*g*}) is encoded using another SLM and placed at the filter plane. When light from the F{*f*} passes through the filter, it performs the required multiplication. The second lens performs a second Fourier transform and produces the correlation output at the output plane. Variations of the basic setup exist as a joint transform correlator (JTC) or binary phase-only filter (BPOF). While the JTC achieves the product of the complex domain by adding the two signals and squaring, the BPOF is achieved by simplifying the complex domain filter by numerically approximating it with the binarized phase. Other applications of linear optical processing can be performed in the basic 4-*f* optical setup where filters such as low-pass or high-pass filters, can be set up in the correlation plane.

The hologram base computing treats holograms as a legitimate way of signal transformation from one form to another, just as a lens transforms from one domain to another. Holograms are complex representations that can act on a signal. A lens can be approximated by a hologram, in reality functions of many optical elements, such as lens, grating, prism, and other transformation such as beamsplitting, it can be combined into single computing element known as computer-generated hologram (CGH). The smartness of this type of system depends on how many operations have been replaced by a single CGH.

It is envisioned that optical computing is better suited for special purpose computing rather than general purpose computing. However, the future of optical computing lies in the marriage of appropriate device technology with the right architecture at the right applications. The development of quantum computing may hold new promises for optical computing.

List of Units and Nomenclature

DMAC	Digital multiplication by analog convolution
JTC	Joint transform correlator
OC	Optical computing
PLA	Programmable logic array
QC	Quantum computing
SLM	Spatial light modulator

See also

Nonlinear Optics, Basics: Four-Wave Mixing; Nomenclature and Units. **Quantum Optics:** Quantum Computing with Atoms.

Further Reading

Arsenault H, Szoplik T and Macukow B (eds) (1989) *Optical Processing and Computing*. San Diego, CA: Academic Press.

Boyd RW (1992) *Nonlinear Optics*. San Diego, CA: Academic Press.

Gaskill J (1978) *Linear Systems Fourier Transform and Optics*. New York: Wiley.

Gibbs HM (ed.) (1982) *Optical Bistabiligy II*. New York: Plenum Press.

Goodman JW (1996) *Introduction to Fourier Optics*, 2nd edn. New York: McGraw Hill.

Ichioka Y and Tanida J (1984) Optical parallel logic gates using a shadow-casting system for optical digital computing. *Proceedings of the IEEE* 72: 787.

Karim MA and Awwal AAS (1992) *Optical Computing: An Introduction*. New York: Wiley.

Koren I (1993) *Computer Arithmetic Algorithms*. New Jersey: Prentice Hall.

McAulay A (1992) *Optical Computing Architecture*. New York: Wiley.

Reynolds GO, DeVelis JB, Parrent GB and Thompson BJ (eds) (1989) *Physical Optics Notebook: Tutorials in Fourier Optics*. Bellingham, CA: SPIE Optical Engineering Press.

Saleh BEA and Teich MC (1991) *Fundamentals of Photonics*. New York: John Wiley & Sons.

VanderLugt A (1992) *Optical Signal Processing*. New York: Wiley.

Yu FTS and Jutamulia S (1992) *Optical Signal Processing, Computing and Neural Networks*. New York: Wiley.

Incoherent Analog Optical Processors

S Jutamulia, University of Northern California, Petaluma, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

After low-cost lasers became widely available, it seemed that incoherent analog optical processing was no longer important. In general, incoherent optical processing has no apparent significant advantages, as compared with its coherent optical counterpart or digital computer processing. However, although nobody now attempts to build an incoherent optical processor, such as an incoherent correlator, most instruments are still operating using incoherent or natural light. These incoherent instruments relate to imaging, including cameras, microscopes, telescopes, projection displays, and lithographic equipment. From this point of view, it is still very important to study the principles of incoherent analog optical processing, in order to further progress or improve the incoherent image.

Incoherent Image Formation

The image formation in a coherent optical system can be explained using the linear system. A perfect point $\delta(x, y)$ in the input plane is imaged in the output plane as the impulse response of the system $h(x, y)$. When the input is only a single point, the output amplitude is simply $h(x, y)$ and the output intensity is $|h(x, y)|^2$. Accordingly, the function of $|h(x, y)|^2$ is called point spread function (PSF).

If we consider that the input object is the collection of a large number of very fine points, under coherent light illumination, the output image will be the collection of the same number of blurred spots. The shape of each blurred spot is $h(x, y)$. If the amplitude function of input object is $f(x, y)$, the amplitude function of output image is

$$o(x, y) = \int\int_{-\infty}^{\infty} f(p, q)h(x - p, y - q)\mathrm{d}p\ \mathrm{d}q \qquad [1]$$

which is a convolution of $f(x, y)$ and $h(x, y)$. The intensity function of output image is

$$|o(x, y)|^2 = \left|\int\int_{-\infty}^{\infty} f(p, q)h(x - p, y - q)\mathrm{d}p\ \mathrm{d}q\right|^2 \qquad [2]$$

According to the convolution theorem, in the Fourier transform domain or frequency domain, we have

$$O(u, v) = F(u, v)H(u, v) \qquad [3]$$

where $O(u, v)$, $F(u, v)$, and $H(u, v)$ are the Fourier transforms of $o(x, y)$, $f(x, y)$, and $h(x, y)$, respectively. $H(u, v)$ is the transfer function of the coherent imaging system. To distinguish it from the incoherent imaging system, it is also called coherent transfer function (CTF).

The same optical imaging system is now illuminated with incoherent light instead of coherent light. The amplitude impulse response is still the same $h(x, y)$, and the point spread function is also the same $|h(x, y)|^2$. Similarly, the input object is the collection of perfect points, and the output image is the collection of blurred images of each point. However, since the illuminating light is incoherent, light from any point in the input object is not coherent with light from any other points. Thus, there is no interference with light from different points. The output image is simply the addition of intensity patterns generated by each point in the input object. The amplitude function of the output image is not computable. However, the intensity function of the output image is

$$|o(x,y)|^2 = \int\int_{-\infty}^{\infty} |f(p,q)|^2 |h(x-p, y-q)|^2 \mathrm{d}p\,\mathrm{d}q \qquad [4]$$

In the frequency domain, we now have

$$O_I(u,v) = F_I(u,v)H_I(u,v) \qquad [5]$$

where $O_I(u,v)$, $F_I(u,v)$, $H_I(u,v)$ are Fourier transforms of $|o(x,y)|^2$, $|f(x,y)|^2$, and $|h(x,y)|^2$, respectively. The intensity impulse response $|h(x,y)|^2$ is the PSF. Note that the index I of O_I, F_I, and H_I denotes intensity. The function $H_I(u,v)$ is now the transfer function of the incoherent imaging system, which is called optical transfer function (OTF).

Referring to Fourier analysis, if $H(u, v)$ is the Fourier transform of $h(x, y)$, the Fourier transform of $|h(x, y)|^2$ is the autocorrelation of $H(u, v)$ as follows:

$$H_I(u, v) = \int\int_{-\infty}^{\infty} H(p, q)H^*(p - u, q - v)\mathrm{d}p\ \mathrm{d}q \qquad [6]$$

The OTF (the function $H_I(u, v)$) is the autocorrelation of CTF (the function $H(u, v)$), which is a complex

function in general. It can be written as

$$\mathrm{OTF} = |\mathrm{OTF}| \exp(i\phi) \qquad [7]$$

We can further write

$$\mathrm{MTF} = |\mathrm{OTF}| \qquad [8]$$

where MTF stands for modulation transfer function, and

$$\mathrm{PTF} = \phi \qquad [9]$$

where PTF stands for phase transfer function.

Measurement of MTF

The MTF of an incoherent imaging system can be measured empirically using an input that is a cosine grating having intensity transmittance function:

$$|f(x)|^2 = 1 + \cos(2\pi ax) \qquad [10]$$

where a is the frequency of grating. The output or the image intensity function will be

$$|o(x)|^2 = 1 + m\cos(2\pi ax + \varphi) \qquad [11]$$

where m, the contrast, can be measured by

$$m = \frac{|o(x)|^2_{\max} - |o(x)|^2_{\min}}{|o(x)|^2_{\max} + |o(x)|^2_{\min}} \qquad [12]$$

In the frequency domain, the Fourier transform of the input function is

$$F_I(u) = \delta(u) + \tfrac{1}{2}\delta(u-a) + \tfrac{1}{2}\delta(u+a) \qquad [13]$$

The Fourier transform of the output function will be

$$O_I(u) = F_I(u)\mathrm{OTF}(u) \qquad [14]$$

Substitution of eqn [13] yields

$$O_I(u) = \delta(u)\mathrm{OTF}(0) + \tfrac{1}{2}\delta(u-a)\mathrm{OTF}(a) + \tfrac{1}{2}\delta(u+a)\mathrm{OTF}(-a) \qquad [15]$$

Notice that

$$\mathrm{OTF}(0) = 1 \qquad [16]$$

due to normalization, and

$$\mathrm{OTF}(a) = \mathrm{OTF}(-a) \qquad [17]$$

Because OTF is an autocorrelation function, it must be a symmetric function. Substitution of eqns [16] and [17] into eqn [15] yields

$$O_I(u) = \delta(u) + \tfrac{1}{2}\delta(u-a)\mathrm{OTF}(a) + \tfrac{1}{2}\delta(u+a)\mathrm{OTF}(a) \qquad [18]$$

Remembering that OTF is a complex function given in eqn [7], we can write

$$\mathrm{OTF}(a) = \mathrm{MTF}(a)\exp(i\phi(a)) \qquad [19]$$

Substitution of eqn [19] into eqn [18] yields

$$O_I(u) = \delta(u) + \tfrac{1}{2}\delta(u-a)\mathrm{MTF}(a)\exp(i\phi(a)) + \tfrac{1}{2}\delta(u+a)\mathrm{MTF}(a)\exp(i\phi(a)) \qquad [20]$$

The inverse Fourier transform of eqn [20] is

$$|o(x)|^2 = 1 + \mathrm{MTF}(a)\cos(2\pi ax + \phi(a)) \qquad [21]$$

By comparing eqn [21] and eqn [11], we find

$$\mathrm{MTF}(a) = m \qquad [22]$$

and the PTF

$$\phi(a) = \varphi \qquad [23]$$

Therefore, the MTF at frequency $u = a$ of the incoherent imaging system can be measured using eqn [12]. To get the complete MTF(u), the measurement is repeated using cosine gratings with different frequencies. The PTF can also be measured at the same time.

Incoherent Spatial Filtering

In the 4f coherent optical processor, the first lens performs the Fourier transform, and the second lens performs the inverse Fourier transform, as shown in Figure 1. The Fourier transform domain or frequency domain is materialized in the frequency plane that is the back focal plane of the first lens.

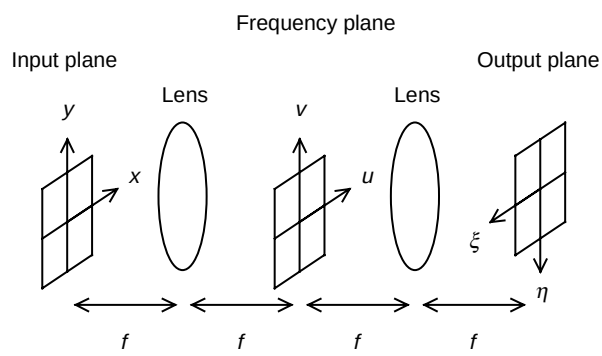

Figure 1 4f coherent optical processor consists of two Fourier transform lenses.

Thus, the Fourier transform of the object can be visualized in the frequency plane. For example, if the object is a grating, we will see its frequency components in the frequency plane. However, if the coherent illumination (e.g., from a laser) is replaced with incoherent illumination (e.g., from a lightbulb), the pattern of Fourier transform in the frequency plane disappears. Does it mean that we can no longer perform spatial filtering? No, we still can perform spatial filtering, although the pattern of Fourier transform is no longer visualized.

Consider that we have a spatial filter in the frequency plane whose physical form can be expressed by the function $H(u, v)$. For illustration, this spatial filter is simply a small one-dimensional window $H(u)$, as shown in Figure 2a. If the 4f optical processor is illuminated with coherent light, the CTF of the system is $H(u)$ itself as shown in Figure 2b. However, if the 4f optical processor is illuminated with incoherent light, we will use OTF instead of CTF in the analysis of linear system. The OTF is the autocorrelation of CTF, which is graphically depicted in Figure 2c. In this example, since the OTF is real and positive, the MTF is identical to the OTF.

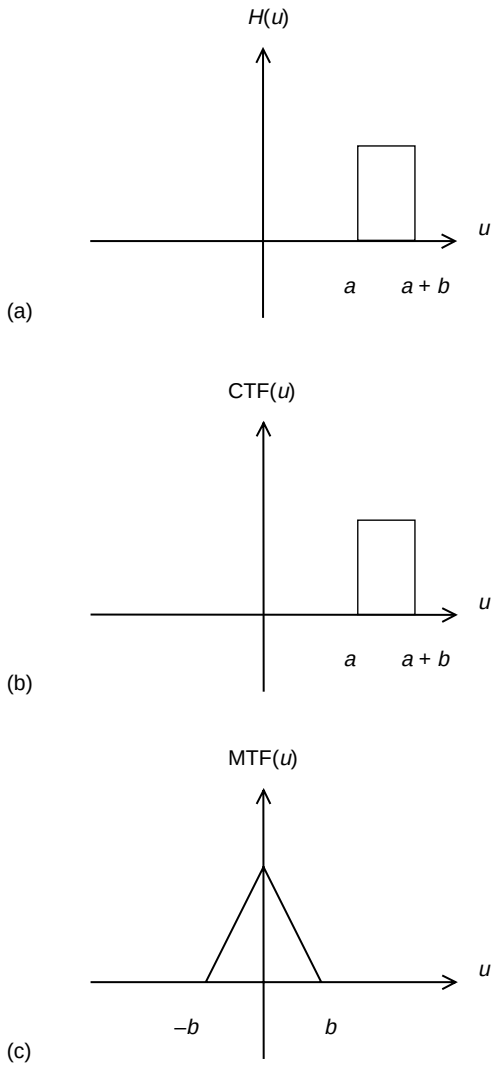

Figure 2 (a) One-dimensional window in frequency plane functions as a spatial filter. (b) CTF and (c) OTF or MTF produced by spatial filter (a).

For simplicity, we analyze and show a one-dimensional filter only in a two-dimensional graph. It certainly can be extended to the two-dimensional filter. And we need a three-dimensional graph to show the two-dimensional autocorrelation.

It is important to note that, although the CTF is a high pass filter that passes frequencies in the band of $u = a$ to $u = a + b$, the MTF is a low pass filter that passes frequencies from $u = -b$ to $u = b$, independent of a. This indicates that an incoherent processor cannot enhance high frequency. It is always a low pass processor. It is interesting to note that the OTF is independent of the location of the filter $u = a$, but it is determined by the width of the filter $\Delta u = b$ only.

Accordingly, an aperture made up of randomly distributed pinholes behaves as a low pass filter in an incoherent imaging system. The cut-off frequency of such a filter can be derived from the diameter of the pinhole, which is equivalent to Figure 2b. The procedure for using this filter is simple. If a camera is used to take a picture, the random pinhole filter can be simply attached to the camera lens. The picture taken will not consist of high-frequency components, because the filter acts like a low pass filter. The low pass filtering can also be done by reducing the aperture of camera. However, by doing so, the light entering the camera is also reduced. The use of a random pinhole filter removes high frequencies without reducing the light intensity. When the random pinhole filter is placed in the frequency plane of a 4f coherent optical processor, it can be used to reduce the speckle noise. Note that the speckle noise is the main drawback in coherent optical processing.

Incoherent Complex Matched Spatial Filtering

Lohmann and Werlich pointed out that the Vander Lugt correlator can also be operated with incoherent light. In the coherent Vander Lugt correlator the correlation term is:

$$o_3(\xi, \eta) = \iint_{-\infty}^{\infty} f(\alpha, \beta) g*(\alpha - \xi, \beta - \eta) \mathrm{d}\alpha \, \mathrm{d}\beta \qquad [24]$$

Its intensity distribution is

$$|o_3(\xi,\eta)|^2 = \left| \iint_{-\infty}^{\infty} f(\alpha,\beta) g*(\alpha-\xi,\beta-\eta) \mathrm{d}\alpha \mathrm{d}\beta \right|^2 \quad [25]$$

However, when the input object is illuminated with incoherent light instead of coherent light, the intensity distribution of the correlation term becomes

$$|o_3(\xi,\eta)|^2 = \iint_{-\infty}^{\infty} |f(\alpha,\beta)|^2 |g(\alpha-\xi,\beta-\eta)|^2 \mathrm{d}\alpha \mathrm{d}\beta \quad [26]$$

Equation [26] shows the correlation of $|f(x,y)|^2$ and $|g(x,y)|^2$. Therefore, when $f(x,y)$ is identical to $g(x,y)$, the correlation peak always appears regardless of whether the illumination is coherent or incoherent. However, the detailed structures of the coherent and incoherent correlation outputs are different. Note that the joint transform correlator cannot be operated using incoherent light, since no joint Fourier spectra will be formed.

Depixelization of Projection Display

Recently, a technique to remove the pixel structure from an image projected by liquid crystal projection display has been reported. A liquid crystal projection display must have a pixel structure to form an image. Can we remove this pixel structure from the projected image to obtain a movie-like image? It is well known that a single hole in the frequency plane can perform a coherent spatial filtering, such that the pixel structure will disappear from the output image.

Because of the pixel structure, we get multiple Fourier spectra in the frequency plane. By passing only one spectrum order through the hole in the frequency plane, the pixel structure can be removed as shown in Figure 3.

It is interesting to note that selecting any one of the spectra will produce the image at the same location.

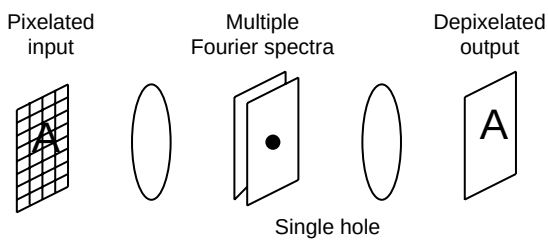

Figure 3 Removal of pixel structure using a single hole in frequency plane. Most of the energy is wasted and the output intensity is dim.

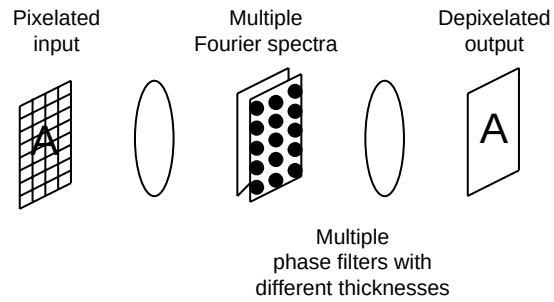

Figure 4 Removal of pixel structure using multiple phase filters with different thickness to cover each Fourier spectrum order. No energy is wasted and the output intensity is very bright.

However, if two spectra are passed, on the top of the produced image, Young interference fringes are also produced. If all spectra are passed, the produced interference fringes are in fact the pixel structure itself. By passing only one spectrum order and blocking most spectrum orders, we lose most of the energy. The projected image will be very dim, although the pixel structure is removed.

To overcome this problem, we may cover each spectrum order with a transparent material having different thickness as shown in Figure 4. Thus, every spectrum order is passed and delayed by a phase filter with a different thickness. If the delay produced by the phase filter is larger than the coherent length of light, every spectrum order is no longer coherent to each other. In other words, the resultant image is the sum of the intensities of all images. As a result, the pixel structure does not appear, and no intensity is lost. For a white light source, since the frequency bandwidth is large, the coherent length is typically on the order of several tens of μm.

This technique can significantly improve the quality of the liquid crystal projection display. Figure 5 shows the experimental result of a depixelated projection image. Figure 5a shows the projected image of an input with pixel structure when no phase filter is applied in the frequency plane. Figure 5b shows the projected image of the same input with pixel structure when phase filters are applied in the frequency plane. The pixel structure is successfully removed as shown in Figure 5b.

Computed Tomography

The computed tomography or CT using X-ray is usually considered beyond optical information processing. However, it is interesting to review the principle of the X-ray CT, since the X-ray source is incoherent, and the CT image reconstruction utilizes Fourier transformation that is commonly used in

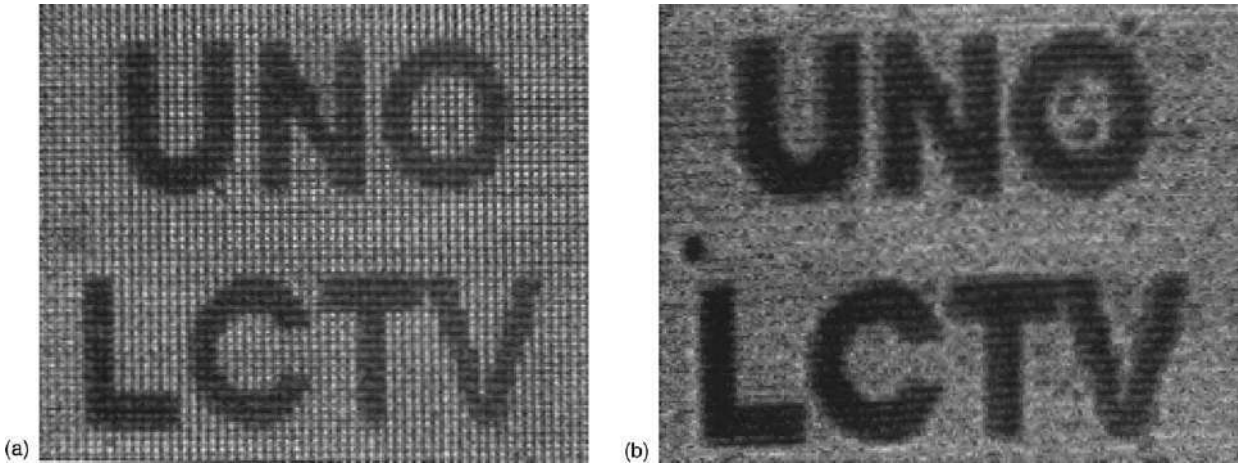

Figure 5 (a) Projected image consists of pixel structure when no phase filter is applied in the frequency plane. (b) Depixelated image is produced when phase filters are applied in the frequency plane.

optical information processing. The X-ray CT is used for taking the cross-section picture of the human body. Note that a conventional X-ray picture only provides a projection instead of a cross-section picture.

Consider the case when an X-ray penetrates through an object characterized by its attenuation coefficient or density function $f(x, y)$ as shown in Figure 6. The detected X-ray intensity is

$$I = I_0 \exp - \left[\int f(x, y) \mathrm{d}s \right] \qquad [27]$$

where I_0 is the original intensity. Equation [27] can be written as

$$\int f(x, y) \mathrm{d}s = -\ln \frac{I}{I_0} \qquad [28]$$

Apparently, we cannot directly measure $f(x, y)$, although we can obtain $\int f(x, y)\mathrm{d}s$ from the measurement of I and I_0. However, we want to get the cross-section picture which is the density function $f(x, y)$. The basic idea is to compute $f(x, y)$ from its Fourier transform $F(u, v)$, and $F(u, v)$ is derived from eqn [28]. $F(u, v)$ can be expressed as

$$F(u, v) = \int \int_{-\infty}^{\infty} f(x, y) \exp[-\mathrm{i}2\pi(ux + vy)] \mathrm{d}x \, \mathrm{d}y \qquad [29]$$

For $v = 0$, eqn (29) becomes

$$F(u, 0) = \int \int_{-\infty}^{\infty} f(x, y) \exp[-\mathrm{i}2\pi ux] \mathrm{d}x \, \mathrm{d}y \qquad [30]$$

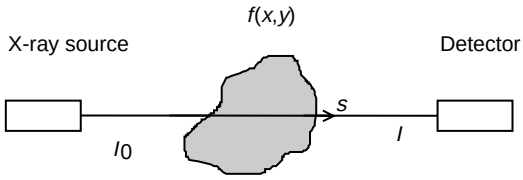

Figure 6 X-ray passes through object having density function $f(x, y)$.

which can be written

$$F(u, 0) = \int_{-\infty}^{\infty} p(x) \exp[-\mathrm{i}2\pi ux] \mathrm{d}x \qquad [31]$$

where

$$p(x) = \int_{-\infty}^{\infty} f(x, y) \mathrm{d}y \qquad [32]$$

The projection function $p(x)$ given by eqn [32] can be obtained using an X-ray parallel beam as shown in Figure 7. Then $F(u, 0)$ can be computed using computer based on eqn [31]. $F(u, 0)$ is the Fourier transform along the u-axis. To get other data, we rotate the coordinate (u, v) to (u', v') by a small angle α. Correspondingly, the coordinate (x, y) is also rotated to (x', y') by the same angle α while the object is not rotated as shown in Figure 8.

From the measurement shown in Figure 8, we can get $p(x')$, which, in turn, provides $F(u', v' = 0)$. By increasing the rotation angle α, we will get other data of $F(u'', v'' = 0)$. The completion of rotation of 180 degrees will give us the data in the frequency domain (u, v) as shown in Figure 9.

After the data $F(u, v)$ is collected, as shown in Figure 9, $F(u, v)$ can be inverse Fourier transformed to

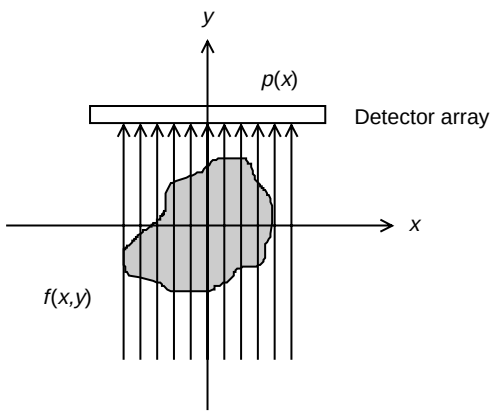

Figure 7 X-ray parallel beam passes through the object to produce projection function $p(x)$.

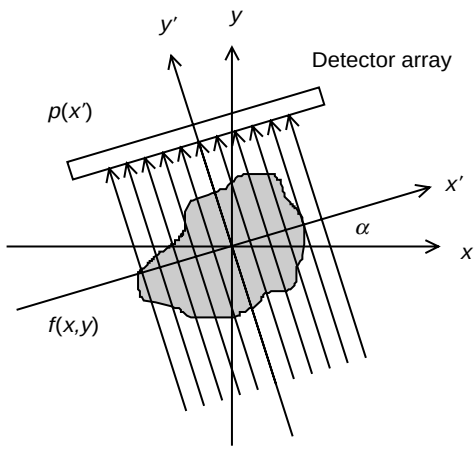

Figure 8 Rotated X-ray parallel beam produces the rotated projection function $p(x')$.

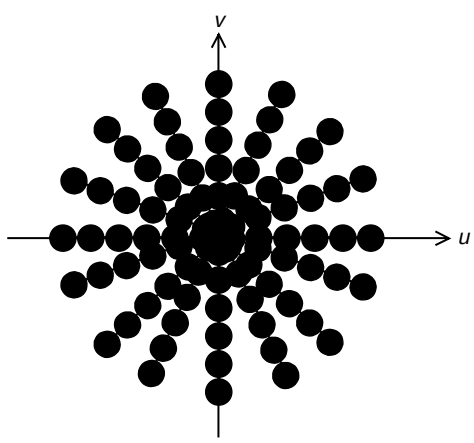

Figure 9 $F(u, v)$ computed from a set of rotated projection functions.

produce $f(x, y)$. There are two approaches in the inverse Fourier transform. The first approach is to interpolate the computed value $F(u, v)$ such that the value of $F(u, v)$, on a regular grid structure, can be defined before an inverse Fourier transform can be taken.

The second approach is to apply polar coordinates, instead of Cartesian coordinates, to the inverse Fourier transform, such that no interpolation in frequency domain (u, v) is required. The inverse Fourier transform in Cartesian coordinates is as follows:

$$f(x, y) = \int\int_{-\infty}^{\infty} F(u, v) \exp[\mathrm{i}2\pi(ux + vy)]\mathrm{d}u\,\mathrm{d}v \quad [33]$$

It can be written in polar coordinates as:

$$f(x,y) = \int_0^{2\pi}\int_0^{\infty} F(\rho\cos\phi, \rho\sin\phi) \times \exp[\mathrm{i}2\pi\rho(x\cos\phi + y\sin\phi)]|\rho|\mathrm{d}\rho\,\mathrm{d}\phi \quad [34]$$

where

$$u = \rho\cos\phi \quad [35]$$

$$v = \rho\sin\phi \quad [36]$$

Therefore $f(x,y)$ can be obtained from the collected $F(u,v)$, as shown in Figure 9, without further interpolation using eqn [34]. Note that eqns [31], [33], and [34] are computed using a computer program.

See also

Spectroscopy: Fourier Transform Spectroscopy.

Further Reading

Bracewell RN (1965) *The Fourier Transform and its Applications*. New York: McGraw-Hill.

Hounsfield GN (1973) Computerized transverse axial scanning (tomography): Part I. *British Journal of Radiology* 46: 1016–1022.

Iizuka K (1985) *Engineering Optics*. New York: Springer-Verlag.

Jutamulia S, Toyoda S and Ichihashi Y (1995) Removal of pixel structure in liquid crystal projection display. *Proceedings of SPIE* 2407: 168–176.

Klein MV (1970) *Optics*. New York: Wiley.

Lohmann AW and Werlich HW (1971) Incoherent matched filtering with Fourier holograms. *Applied Optics* 10: 670–672.

Reynolds GO, DeVelis JB, Parrent GB Jr and Thompson BJ (1989) *The New Physical Optics Notebook: Tutorials in Fourier Optics*. Bellingham, WA: SPIE.

Vander Lugt A (1964) Signal detection by complex spatial filtering. *IEEE Trans. Inf. Theo.* IT-10: 139–145.

Yang X, Jutamulia S and Li N (1996) Liquid-crystal projection image depixelization by spatial phase scrambling. *Applied Optics* 35: 4577–4580.

Yu FTS and Wang EY (1973) Speckle reduction in holography by means of random spatial sampling. *Applied Optics* 12: 1656–1659.

Optical Bit-Serial Computing

A D McAulay, Lehigh University, Bethlehem, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Why Optical Bit-Serial Computing?

Optical bit-serial computing will lower cost and improve dataflow rates in optical telecommunications and associated computing. Moreover, the electromagnetic immunity of optics will enhance security for computing as well as for communications. Also, absence of sparking with optics enables safe operation, even in hazardous environments.

Currently, optical bit-serial communication involves passing terabits of data per second (10^{12} bps) – thousands of encyclopedias per second – along an optical fiber by switching light of multiple colors on and off, representing '1' and '0' bits respectively. Replacing electronic computing functions, constrained to tens of gigabits per second (10^9 bps), with integrated optics, ones of much higher speed, enables faster flow rates – allowing continuation of the annual doubling of capacity-distance that has occurred since the 1970s. Cost is reduced by eliminating the demultiplexing and multiplexing for optic-electronic-optic (OEO) conversions. Two illustrative areas of optical bit-serial computing for replacing electronic computing functions are described: computing at up to 40 Gbps with semiconductor optical amplifiers (SOAs) and computing at higher bit rates with integrated optic components. Several books address the relevant technical background: free-space optical information processing, optical computer architectures, optical fiber communications, and integrated optics.

Computing with Semiconductor Optical Amplifiers

Semiconductor optical amplifiers, SOAs, allow nonlinear operations (switching, logic) at low power (mW) because of amplification – typically up to 1000. An SOA is a laser diode with antireflection coatings in place of reflecting facets. Light pulses at bit rates in excess of 40 Gbps are amplified by absorbing power from an electronic pump. SOAs are used in the following, at bit rates up to 40 Gbps, for nonlinear operations in cross-gain, cross-phase, or cross-polarization.

SOA-XGM for Logic

Figure 1 shows an SOA in cross-gain modulation (XGM) performing a NOR function. Either A or B inputs at a '1' level, driving the SOA through couplers into saturation, lowering its gain. A continuous wave at C, having a different frequency λ_3, from that of A, λ_1 and B, λ_2, experiences the decrease in gain, resulting in an output '0' level at frequency λ_3. The filter in Figure 1 can be avoided by counter-propagating A and B through the SOA from right to left. XOR gates use interference between two incoming streams of bits (locked to identical frequency, phase, polarization) to correlate headers on internet protocol (IP) address headers – similar to matching zip codes in postal mail routing machines. Bit-serial-adders are constructed from combinations of XOR and NOR gates; subsequently the adders are used to construct ripple-carry-adders for word addition. In addition, NOR gates are used to construct flip-flops; hence registers which form short-term memory.

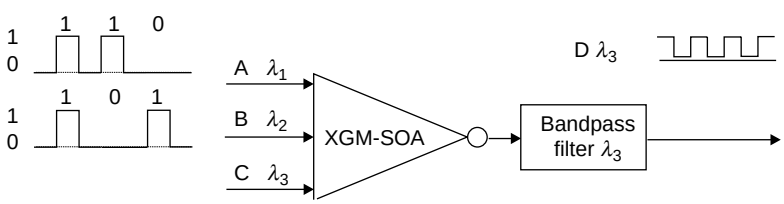

Figure 1 Optical XGM-SOA NOR gate.

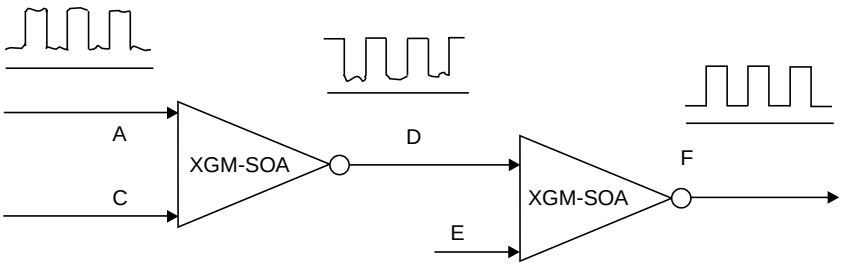

Figure 2 Optical XGM-SOA 2R signal regeneration.

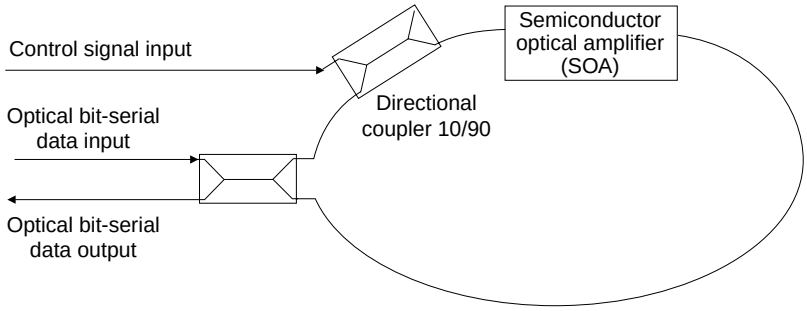

Figure 3 XPM-SOA in Sagnac interferometer for optical manipulation of bits at high data rates.

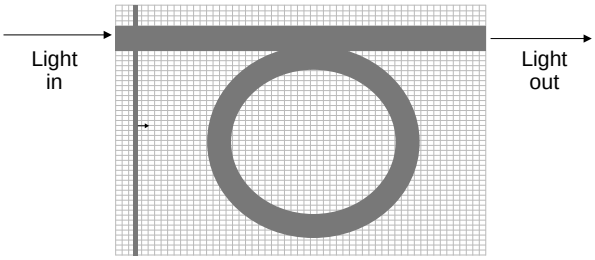

Figure 4 Integrated optic Optiwave layout for microring resonator.

SOA-XGM for Frequency Conversion

Ignoring input B, the bit information in Figure 1 has been transferred from carrier frequency A onto the carrier frequency C, with inversion. This enables switching of signals in wavelength division multiplexing (WDM), for example, color reuse is simplified and a color red is routed to the location where blue was previously routed; this is equivalent to time slot interchangers for switching in time division multiplexing (TDM).

SOA-XGM for 2-R Signal Regeneration

Figure 2 show how two SOAs may be used in sequence to restore signal levels distorted by propagation or computation; reamplification and reshaping (2-R). The '1' level at the output of the first SOA is clipped by gain saturation, removing noise, this becoming the '0' level at the output of the second SOA. The second SOA clips the '1' level at the output.

SOA-XPM with Interferometer

In cross-phase modulation (XPM), the signal at A, Figure 1, changes the phase of carrier for signal C passing through the SOA (signal levels may be less than for XGM). Inclusion of an SOA in an interferometer, Mach–Zehnder or Sagnac, converts phase to intensity for frequency conversion, switching or high-speed bit manipulation. Figure 3 shows a control pulse entering a coupler at the left and providing a control input to the SOA at one frequency. The bits circulating round the loop at a different frequency can be switched on and off by the control pulse, to synchronize or manipulate high bit rate signals or for frequency conversion.

Computing with Integrated Optics

Combining several discrete optical components into a single optical chip is more beneficial than in electronic very large-scale integration (VLSI) because optical connections require greater precision; moreover bit rates are higher because photons, unlike electrons, have no mass or charge (only spin).

Integrated Optic Microring Resonator

Figure 4 shows a microring resonator in an integrated optic chip for optical filtering, light entering at the left couples into the loop. For loop lengths that are a multiple of wavelength, light cycling the loop is in phase and resonance occurs. At resonant frequencies,

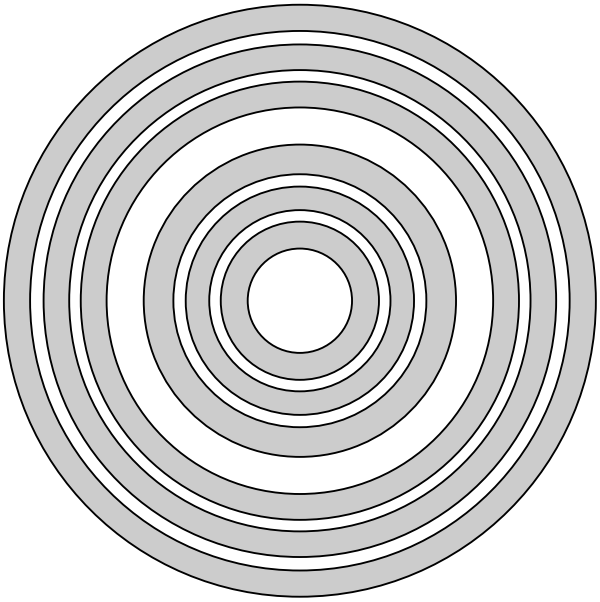

Figure 5 Integrated optic Optiwave layout for Bragg microring resonator.

light exiting the loop is out of phase with that in the linear guide, causing cancellation of the input. As frequency varies away from resonance, light passes and the device acts as a notch filter to block light at the resonant frequencies. With nonlinear Kerr material, low-level light intensity may be used to change a resonant frequency to make a switch.

Bragg Integrated Optic Microring Resonator

Performance is improved by using a Bragg microring resonator, Figure 5, which confines light in the central light colored ring by dielectric mirror reflection from Bragg gratings on either side. The refractive index is less in the ring than in the surrounding regions, in contrast to a conventional waveguide (used in Figure 4) that guides by total internal reflection. The higher reflectivity, smaller ring radius and lower refractive index – these help to reduce power loss, volume, and travel time around the ring. This increases filter selectivity, drives up peak power for enhanced nonlinearity for switching, decreases stored energy for fast response, and increases frequency range (free spectral range, *FSR*). In this case, an equivalent photonic crystal may be more expensive because it requires higher resolution lithography and greater refractive index contrast.

Photonic Crystal Mach–Zehnder Switch

Recently developed photonic crystal technology allows a reduction in integrated optic chip size from centimeters to submillimeters. The 2D or 3D periodic structure of dielectric variations emulates crystals and provides dielectric mirror reflections in multiple dimensions for lossless waveguide bends, high Q resonators, and slowing of wave propagation to enhance effective nonlinearity of materials. Figure 6 shows a proposed 2D photonic crystal Mach–Zehnder interferometer switch in which light is split into two paths by a coupler and propagation velocity is slowed for enhancing nonlinearity by using coupled-resonator optical waveguides (CROWs), in which alternate posts of different dielectric constant remain in the waveguides to form coupled cavities.

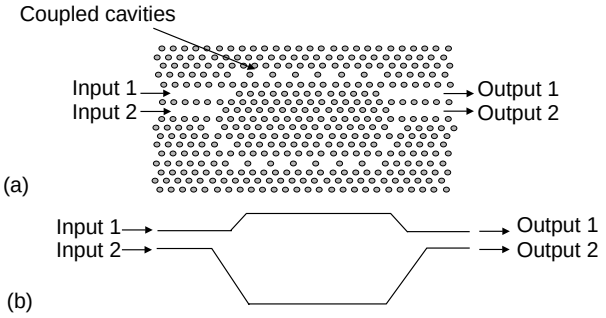

Figure 6 Proposed photonic crystal Mach–Zehnder interferometer.

Conclusion

Optical bit-serial computing can reduce cost and increase flow-rate for expanding telecommunication network capacity, while providing enhanced security from electromagnetic sabotage and tampering. Advances in integrated optics (microring resonators), progress in optoelectronic components (SOAs), improved materials and fabrication techniques – these technologies enable cost-effective optical bit-serial computing solutions for future telecommunication evolution.

See also

All-Optical Signal Regeneration. **Information Processing:** All-Optical Multiplexing/Demultiplexing.

Further Reading

Agrawal G (2002) *Fiber-Optic Communication Systems*. New York: Wiley.

Connelly MJ (2002) *Semiconductor Optical Amplifiers*. Boston, MA: Kluwer Academic Publishers.

Goodman JW (1996) *Introduction to Fourier Optics*, 2nd edn. New York: McGraw-Hill.

Heuring VP, Jordan HF and Pratt JP (1992) A bit serial architecture for optical computing. *Applied Optics* 31(17): 3213–3224.
Hunsperger RG (2002) *Integrated Optics*, 5th edn. New York: Springer.
Joannopoulios JD, Meade RD and Winn JN (1995) *Photonic Crystals*. Princeton, NJ: Princeton University Press.
McAulay AD (1991) *Optical Computer Architectures*. New York: Wiley.
McAulay AD (2004) Novel all-optical flip-flop using semiconductor optical amplifiers in innovating frequency-shifting inverse-threshold pairs. *Optical Engineering* 43(5): 1115–1120.
Scheuer J and Yariv A (2003) Two-dimensional optical ring resonators based on radial Bragg resonance. *Optics Letters* 28(17): 1528–1530.
Xu L, Glesk I, Baby V and Prucnal PR (2004) All-optical wavelength conversion using SOA at nearly symmetric position in a fiber-based Sagnac interferometric loop. *IEEE Photonics Technical Letters* 16(2): 539–541.
Yariv A (1997) *Optical Electronics in Modern Communications*, 5th edn. New York: Oxford University Press.

Optical Digital Image Processing

B L Shoop, United States Military Academy, West Point, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Optical approaches to image processing provide a natural method by which to obtain the inherent advantages associated with optics including massive parallelism, high-speed processing, and the inherent compatibility with image formats. Early applications of optics employed analog processing techniques with the most common being the optical Fourier transform, matrix-vector processors, and correlators. Later, advances in digital signal processing (DSP), digital storage, and digital communication systems demonstrated the potential for higher resolution, improved flexibility and functionality, and increased noise immunity over analog techniques and quickly became the preferred method for accurate signal processing. Consequently, optical processing architectures, that incorporated digital processing techniques, were explored. Advances in very large-scale integration (VLSI) electronic circuitry and optoelectronic devices later enabled smart pixel technology, which integrates the programmability and processing of electronic circuitry with the two-dimensional (2D) nature of an optical architecture and made possible a particular realization of optical digital image processing. This chapter describes an application of optical digital image processing, based on smart pixel technology, called digital image halftoning.

Natural images are, by definition, continuous in intensity and color. Halftoning is the process by which a continuous-tone, gray-scale image is converted to one containing only binary output pixels. Halftoning can be thought of as an image compression technique whereby a high-resolution image is transformed to a low-resolution image containing only black-and-white pixels. The transformation from a continuous-tone, gray-scale image to one containing only binary-valued pixels is also similar, in principle, to low-resolution analog-to-digital (A/D) conversion.

An early manual approach to printing continuous-tone images, using only the presence or absence of ink, is the Mezzotint. Ludwig von Siegen (1609–1680), an amateur Dutch printmaker, first invented the mezzotint, or 'half tone process', in 1640. The process later came into prominence in England during the early eighteenth century. Mezzotints are produced on copper plates where the surface of the plate is roughed with a tool called the Mezzotint Rocker, shaped like a wide chisel with a curved and serrated edge. By rocking the toothed edge backwards and forwards over the plate, a rough burr is produced which holds the ink. The dark regions of an image were roughed in a random fashion, while the areas to be lightened were scraped and burnished. This process was found to be especially useful for the reproduction of paintings, due to its ability to capture the subtlest gradations of tone from rich, velvety blacks to glowing highlights. The mezzotint is, therefore, an early artistic predecessor to the modern day halftone.

Optical halftoning has been in use by commercial printers for over 100 years. Commercial halftone screens are based on a discovery made by William Henry Fox Talbot (1800–1877), in 1852. He demonstrated the feasibility of optical halftoning by photographing an image through a loosely woven fabric or 'screen'. In the 1890s, this process came into practical use when the halftone screen, consisting of two ruled glass plates cemented together, became commercially available. Commercial halftone screens produce an effect of variably sized dots on the photographic plate that gives the illusion of a continuous tone image.

Digital image halftoning, sometimes referred to as spatial dithering, is the process of converting an

electronic continuous-tone, gray-scale image to one containing only binary-valued pixels, for the purpose of displaying, printing, or storing the image. The underlying concept is to provide the viewer of the image with the illusion of viewing a continuous-tone image when, in fact, only black and white pixel values are used in the rendering. This process is particularly important in applications such as laser printing, bilevel displays, xerography, and more recently, facsimile.

There are a number of different methods by which this digital image halftoning can be accomplished, which are generally classified as either point or neighborhood processes. A point process is one which computes the output pixel value based strictly on some characteristic of the corresponding input pixel. In contrast, a neighborhood process computes a single output pixel based on a number of pixels in a neighborhood or region of the input image. Ordered dither, which is considered a point process, produces an output by comparing a single continuous-tone input value against a deterministic periodic array of threshold values. If the value of the input pixel under consideration is greater than the corresponding threshold value, the corresponding output pixel is rendered white. If the intensity is less that the threshold, it is rendered black. Dispersed-dot ordered dither is a subset of ordered dither in which the halftone dots are of a fixed size while clustered-dot ordered dither uses variable-sized dots and simulates the variable-sized dots of a printer's halftone screen in the rendering. Among the advantages of point process halftoning, in general and ordered dither halftoning specifically, are simplicity and speed of implementation. The primary disadvantage is that ordered dither produces patterns in the halftoned image, which are visually undesirable.

In contrast, halftoning using the error diffusion algorithm, first introduced by Floyd and Steinberg in 1975, employs neighborhood operations and is currently the most popular neighborhood process. In this approach to halftoning, the output pixel value is determined, not solely by the value of the input pixel and some deterministic pattern, but instead by a weighted average of values from pixels in a neighborhood surrounding the specific pixel being computed. Here, the error of the quantization process is computed and spatially distributed or diffused within a local neighborhood in order to influence future pixel quantization decisions within that neighborhood and thereby improve the overall quality of the halftoned image. In classical unidirectional error diffusion, the image is processed sequentially, proceeding from the upper-left to the lower-right. Starting in the corner of the image, the first pixel is thresholded and the quantizer error is calculated. The error is then diffused to neighboring pixels that have not yet been processed according to the scaling and interconnect defined by the error diffusion filter. The remainder of the pixels in the image are subsequently processed with each quantizer input, now being the algebraic sum of the corresponding original input pixel intensity and the weighted error from previously processed pixels. While error diffusion produces halftone images of superior quality, the unidirectional processing of the algorithm continues to introduce visual artifacts that are directly attributable to the algorithm itself.

In an effort to overcome the implementation constraints of serial processing and improve overall halftone image quality, a number of different parallel architectures have been investigated, including several based on neural network algorithms. While neural network approaches can provide distinct advantages in improved halftone image quality, they also present challenges to hardware implementations, including speed of convergence and the physical interconnect requirements. These challenges, in addition to the natural compatibility with imaging media, provide the motivation for developing optical image processing architectures for these types of applications.

The Error Diffusion Algorithm

Figure 1 shows a block diagram of the error diffusion architecture used in digital image halftoning. The input image is represented as $x_{m,n}$, $w_{m,n}$ is the impulse response of a 2D causal, unity gain error diffusion filter, and the output quantized image is described by $y_{m,n} \in \{-1, 1\}$. The quantizer $q[u_{m,n}]$ provides the one-bit thresholding functionality necessary to convert each analog input pixel to a low-resolution digital output pixel. The quantizer error $\varepsilon_{m,n}$ is computed as the difference between the output and input to the quantizer and distributed to adjacent pixels according to the weighting and interconnection specified by the error diffusion filter. The unity gain constraint ensures

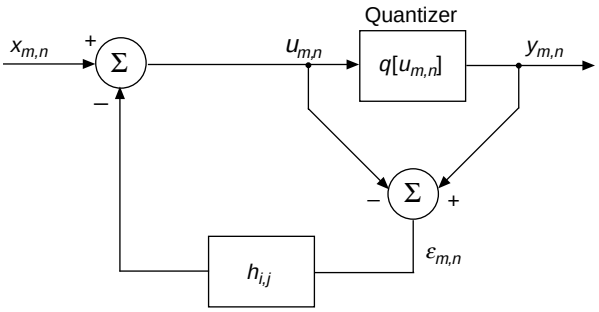

Figure 1 Block diagram of recursive error diffusion architecture.

that no amplification or attenuation of the error signal occurs during the error diffusion process and preserves the average intensity of the image.

In this architecture, the error associated with the quantizer decision at spatial coordinates (m, n) is diffused within a local 2D region to influence adjacent quantization decisions. All of the state variables in this architecture are scalar values since this architecture represents a recursive algorithm in which each pixel in the image is processed sequentially. In the general case, where $w_{m,n}$ is assumed to be a 2D finite impulse response (FIR) filter with coefficients that span a region of support defined by $R_{m,n}$, the quantizer input state $u_{m,n}$ can be written as

$$u_{m,n} = x_{m,n} - \sum_{i,j \in R_{m,n}} w_{i,j} \varepsilon_{m-i,n-j} \qquad [1]$$

Here, $w_{i,j} = 0$, for all $i = j$, since the error signal is diffused spatially within the region of support $R_{m,n}$ and does not influence the pixel under consideration. In this unidirectional error diffusion, the 2D error diffusion filter is therefore noncausal. Figure 2 shows the impulse response of several popular error diffusion filters that have been used in rectangular grid architectures. Figure 2a shows the original error diffusion filter weights developed by Floyd and Steinberg. They argued that the four weights were the smallest number of weights that would produce good halftone images. Figure 2b is a filter that contains a larger region of support developed by Jarvis, Judice, and Ninke in 1976. Finally, the filter in Figure 2c, developed by Stucki, contains the same region of support with coefficients that are a multiple of 2 for digital compatibility and computational efficiency. Here, '•' represents the pixel being processed and the weights describe the local region over which the error is diffused. The normalization factors ensure that the filter coefficients sum to one and therefore meet the unity gain criterion.

To understand the effect of error diffusion and its impact on the quantizer error, it is instructive to describe the effect of error diffusion mathematically. Consider the relationship between the quantizer error $\varepsilon_{m,n} \equiv y_{m,n} - u_{m,n}$ and the overall quantization error $\hat{\varepsilon}_{m,n} \equiv y_{m,n} - x_{m,n}$ in the frequency domain. If we assume that the error is uncorrelated with the input and has statistical properties consistent with a white process, z-transform techniques can be applied to show that the feedback architecture provides the following noise-shaping characteristics:

$$H_{\mathrm{ns}}(z_1, z_2) \equiv \frac{\hat{E}(z_1, z_2)}{E(z_1, z_2)} = 1 - W(z_1, z_2) \qquad [2]$$

Here, $\hat{E}(z_1, z_2)$, $E(z_1, z_2)$, and $W(z_1, z_2)$ represent the z-transforms of $\hat{\varepsilon}_{m,n}$, $\varepsilon_{m,n}$, and $w_{m,n}$, respectively. Equation [2] shows that the noise-shaping characteristic of the error diffusion architecture is directly related to the spectral characteristics of $W(z_1, z_2)$. Appropriate selection of $W(z_1, z_2)$ can spectrally shape the quantizer noise in such a way that minimizes the effect of the low-resolution quantization process on the overall halftoning process. An error diffusion filter with low-pass spectral characteristics produces an overall noise-shaping characteristic $1 - W(z_1, z_2)$, which has the desired high-pass frequency characteristics. This noise-shaping function suppresses the quantization noise within those frequencies occupied by the image and spectrally shapes the noise to high frequencies, which are less objectionable to the human visual system. Circular symmetry is another important characteristic in the error diffusion filter since the human visual system is particularly sensitive to directional artifacts in the image.

In the following qualitative description of digital image halftoning, a 348×348 natural image of the Cadet Chapel at West Point was selected as the test image. This image provides a variety of important image characteristics, with regions of uniform gray-scale, edges, and areas with fine detail. This image was scanned from a high-resolution black and white photograph at 150 dots per inch (dpi) and then printed using a 300 dpi laser printer.

Figure 3 shows the test image of the Cadet Chapel rendered using dispersed-dot ordered dither and results in 256 gray-levels at 300 dpi. Figure 4 shows the halftone image of the Cadet Chapel using the error diffusion algorithm and the Floyd–Steinberg filter coefficients shown in Figure 2a. The unidirectionality of the processing and the causality of the diffusion filter result in undesirable visual artifacts in the halftone image. These include directional hysteresis, which is manifested as 'snakes' running from upper-left to lower-right, and transient behavior near boundaries, which appears as 'shadows' below and to the right of sharp-intensity changes. The directional

$$\left(\frac{1}{16}\right)\times \begin{array}{ccc} & \bullet & 7 \\ 3 & 5 & 1 \end{array}$$

(a)

$$\left(\frac{1}{48}\right)\times \begin{array}{ccccc} & & & \bullet & 7 & 5 \\ 3 & 5 & 7 & 5 & 3 \\ 1 & 3 & 5 & 3 & 1 \end{array}$$

(b)

$$\left(\frac{1}{42}\right)\times \begin{array}{ccccc} & & & \bullet & 8 & 4 \\ 2 & 4 & 8 & 4 & 2 \\ 1 & 2 & 4 & 2 & 1 \end{array}$$

(c)

Figure 2 Three common error diffusion filters used in rectangular grid architectures. (The '•' represents the origin.)

hysteresis is particularly objectionable in uniform gray intensity areas such as the cloud structure in the upper-left of Figure 4. Similar artifacts are also present in images halftoned using the error diffusion algorithm and other causal diffusion kernels. The logical conclusion to draw from this limited, qualitative analysis is that if we could diffuse the error symmetrically and simultaneously process the entire image in parallel, we could reduce some of these visual artifacts and thereby improve overall halftone image quality.

Figure 3 Original gray-scale image of the Cadet Chapel rendered using dispersed-dot ordered dither.

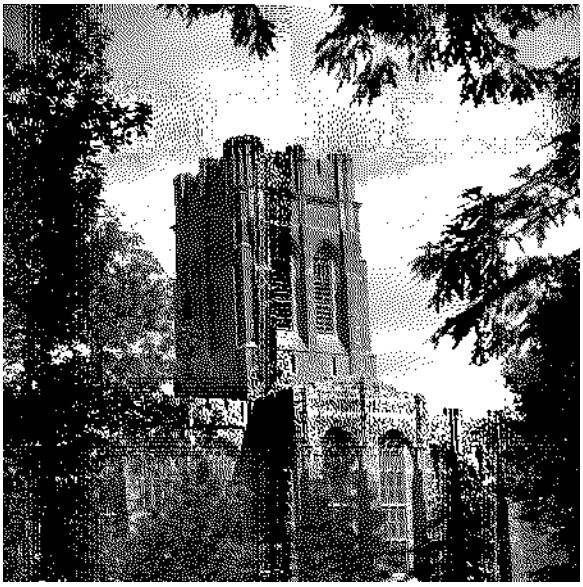

Figure 4 Halftone image of the Cadet Chapel produced using the error diffusion algorithm and the Floyd–Steinberg weights.

The Error Diffusion Algorithm and Neural Networks

The popularity of the neural network-based approach to signal and image processing lies in the ability to minimize a particular performance metric associated with a highly nonlinear system of equations. Specifically, the problem of creating a halftone image can be cast in terms of a nonlinear quadratic optimization problem where the performance measure to be minimized is the difference between the original and the halftone images.

The Hopfield-Type Neural Network

Tank and Hopfield first proposed a mathematical description of the functionality of the neural processing network for signal processing applications. Figure 5 shows an electronic implementation of a four-neuron architecture. Here, a single neuron is comprised of both a standard and inverting amplifier and the synapses or neural interconnections are represented by the physical connections between the input and output of the amplifiers. If the input to amplifier i is connected to the output of amplifier j by a resistor with resistance $R_{i,j}$, the amplitude of the connection $T_{i,j}$ is the conductance $1/R_{i,j}$.

The dynamic behavior of an N-neuron Hopfield-type neural network can be described by the following system of N nonlinear differential equations:

$$c\frac{\mathrm{d}u_i(t)}{\mathrm{d}t} = -u_i(t) + \sum_j T_{i,j}\Im[u_j(t)] + x_i \qquad [3]$$

where $i = 1, 2, \ldots N$, $\Im[\cdot]$ is a monotonically increasing sigmoid function, x_i is an input vector containing N

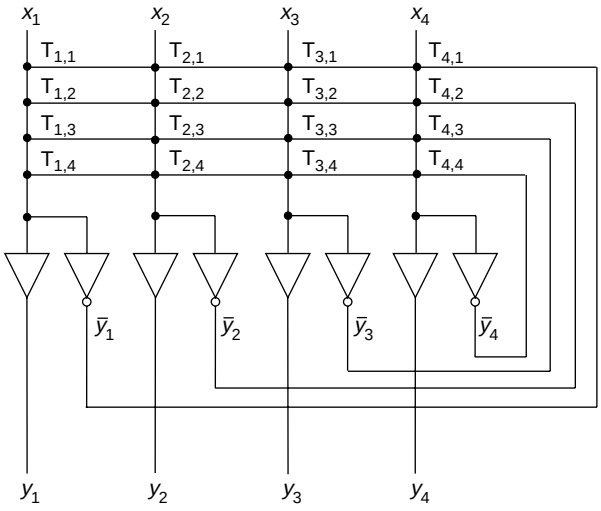

Figure 5 Electronic realization of a four-neuron Hopfield-type neural network.

elements, and c is a scaling factor. In equilibrium, eqn [3] implies that

$$u_i = x_i + \sum_j T_{i,j} \Im[u_j] \qquad [4]$$

Hopfield showed that when the matrix of interconnection weights $\mathbf{T}$ is symmetric with zero diagonal elements and the high-gain limit of the sigmoid $\Im[\cdot]$ is used, the stable states of the N functions $y_i(t) = \Im[u_j]$ are the local minima of the energy function:

$$E = -\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{T}\mathbf{y} - \mathbf{x}^{\mathrm{T}}\mathbf{y} \qquad [5]$$

where $\mathbf{y} \in \{-1, 1\}$ is the N-vector of quantized states. As a result, if the neural network can be shown to be stable, as the network converges the energy function is minimized. In most neural network applications, this energy function is designed to be proportional to a performance metric of interest and therefore as the network converges, the energy function and consequently the performance metric is also minimized.

The Error Diffusion Neural Network

An understanding of the operation and characteristics of the Hopfield-type neural network can be applied to the development and understanding of a 2D extension of the error diffusion algorithm. Figure 6 shows an electronic implementation of a four-neuron error diffusion-type neural network, where the individual neurons are represented as amplifiers and the synapses by the physical connections between the input and output of the amplifiers.

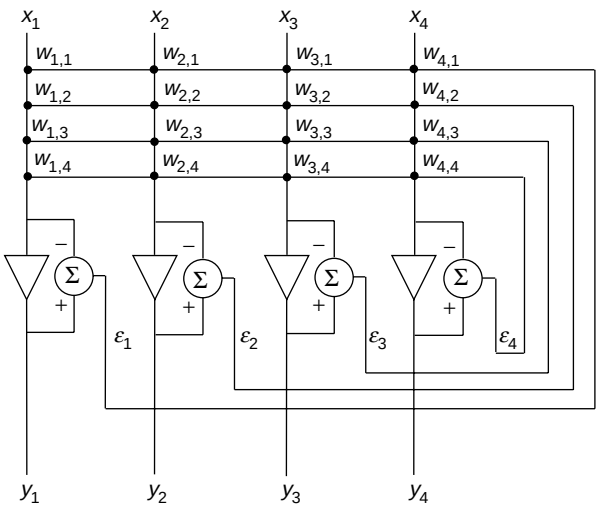

Figure 6 Four-neuron electronic implementation of the error diffusion neural network.

In equilibrium, the error diffusion neural network can be shown to satisfy

$$\mathbf{u} = \mathbf{W}(\mathbf{y} - \mathbf{u}) + \mathbf{x} \qquad [6]$$

Here, $\mathbf{u}$ is the state vector of neuron inputs, $\mathbf{W}$ is a matrix containing the interconnect weights, $\mathbf{x}$ is the input state vector, and $\mathbf{y}$ is the output state vector. For an $N \times N$ image, $\mathbf{W}$ is an $N^2 \times N^2$ sparse, circulant matrix derived from the original error diffusion weights $w_{m,n}$. If we define the coordinate system such that the central element of the error diffusion kernel is $(i, j) = (0, 0)$, then the matrix $\mathbf{W}$ is defined as $W(i, j) = -w[(j - i) \operatorname{div} N, (j - i) \bmod N]$ where:

$$x \operatorname{div} y = \left\lfloor \frac{x}{y} \right\rfloor \text{ if } xy \geq 0 \text{ or } \left\lceil \frac{x}{y} \right\rceil \text{ if } xy < 0 \qquad [7]$$

An equivalence to the Hopfield network can therefore be described by

$$\mathbf{u} = \mathbf{A}(\mathbf{W}\mathbf{y} + \mathbf{x}) \qquad [8]$$

where $\mathbf{A} = (\mathbf{I} + \mathbf{W})^{-1}$. Effectively, the error diffusion network includes a pre-filtering of the input image $\mathbf{x}$ by the matrix $\mathbf{A}$ while still filtering the output image $\mathbf{y}$ but now with a new matrix, $\mathbf{AW}$. Recognizing that $\mathbf{AW} = \mathbf{I} - \mathbf{A}$ and adding the arbitrary constant $k = \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, we can write the energy function of the error diffusion neural network as

$$\hat{E}(x, y) = \mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \qquad [9]$$

This energy function is a quadratic function, which can be factored into

$$\hat{E}(x, y) = [\mathbf{B}\underbrace{(\mathbf{y} - \mathbf{x})}_{\text{error}}]^{\mathrm{T}}[\mathbf{B}\underbrace{(\mathbf{y} - \mathbf{x})}_{\text{error}}] \qquad [10]$$

where $\mathbf{A} = \mathbf{B}^{\mathrm{T}}\mathbf{B}$. From eqn [10] we find that as the error diffusion neural network converges and the energy function is minimized, so too is the error between the output and input images.

If the neurons update independently, the convergence of the error diffusion network is guaranteed if

$$\forall k : [\mathbf{AW}]_{k,k} \geq 0 \qquad [11]$$

We find in practice, that even in a synchronous implementation, the halftoned images converge to a solution, which results in significantly improved halftone image quality over other similar halftoning algorithms.

The Error Diffusion Filter

The purpose of the error diffusion filter is to spectrally shape the quantization noise in such a way that the

quantizer error is distributed to higher spatial frequencies, which are less objectionable to the human visual system. In this application, a feedback filter for the error diffusion neural network was designed using conventional 2D filter design techniques resulting in the 7×7 impulse response shown in Figure 7. Figure 8 shows the same image of the Cadet Chapel halftoned using the error diffusion neural network algorithm and the 2D error diffusion filter shown in Figure 7. There is clear improvement in halftone image quality over the image produced using sequential error diffusion in Figure 4. Notice particularly the uniform distribution of pixels in the cloud formation in the upper-left of Figure 8. Also noteworthy is the improvement around the fine detail portions of the tree branches and next to the vertical edges of the chapel.

0.0003	0.0019	0.0051	0.0068	0.0051	0.0019	0.0003
0.0019	0.0103	0.0248	0.0328	0.0248	0.0103	0.0019
0.0051	0.0248	0.0583	0.0766	0.0583	0.0248	0.0051
0.0068	0.0328	0.0766	•	0.0766	0.0328	0.0068
0.0051	0.0248	0.0583	0.0766	0.0583	0.0248	0.0051
0.0019	0.0103	0.0248	0.0328	0.0248	0.0103	0.0019
0.0003	0.0019	0.0051	0.0068	0.0051	0.0019	0.0003

Figure 7 Impulse response of a 7×7 symmetric error diffusion filter. ('•' represents the origin.)

Figure 8 Halftone image of the Cadet Chapel using the error diffusion neural network and the 7×7 symmetric error diffusion filter shown in Figure 7.

Smart Pixel Technology and Optical Image Processing

The concept underlying smart pixel technology is to integrate electronic processing and individual optical devices on a common chip, to take advantage of the complexity of electronic processing circuits and the speed of optical devices. Arrays of these smart pixels then allow the advantage of parallelism that optics provides. There are a number of different approaches to smart pixel technology that differ primarily in the kind of opto-electronic devices, the type of electronic circuitry, and the method of integration of the two. Common opto-electronic devices found in smart pixels include: light-emitting diodes (LEDs); laser diodes (LDs); vertical cavity surface emitting lasers (VCSELs); multiple quantum well (MQW) modulators; liquid crystal devices (LCDs); and photodetectors (PDs). The most common types of electronic circuitry are silicon-based semiconductors such as complementary metal oxide semiconductors (CMOS) and compound semiconductors such as gallium arsinide (GaAs)-based circuitry. There are a number of different approaches to smart pixels, which generally differ in the way in which the electronic and optical devices are integrated. Monolithic integration, direct epitaxy, and hybrid integration are the three most common approaches in use today.

Smart pixel technology provides a natural methodology by which to implement optical image processing architectures. The opto-electronic devices provide a natural optical interface while the electronic circuitry provides the ability to perform either analog or digital computation. It is important to understand that any functionality that can be implemented in electronic circuitry can be integrated into a smart pixel architecture. The only limitations arise from physical space constraints imposed by the integration of the opto-electronic devices with the electronic circuitry. While smart pixels can be fabricated with digital or analog circuitry, the smart pixel architecture described subsequently uses mixed-signal circuitry.

A Smart Pixel Implementation of the Error Diffusion Neural Network

A smart pixel hardware implementation of the error diffusion neural network provides the potential to simultaneously achieve the computational complexity of electronic circuitry and the parallelism and high-speed switching of optics. The specific smart pixel architecture described here integrates MQW modulators called self electro-optic effect devices (SEEDs) with silicon CMOS

VLSI circuitry, using a hybrid integration approach called flip-chip bonding. This type of smart pixel is commonly referred to as CMOS-SEED smart pixel technology.

To provide an example of the application of smart pixel technology to digital image halftoning, a 5×5 CMOS-SEED smart pixel array was designed and fabricated. The CMOS circuitry was produced using a 0.5 μm silicon process and the SEED modulators were subsequently flip-chip bonded to silicon circuitry using a hybrid integration technique. The central neuron of this new smart pixel array consists of approximately 160 transistors while the complete 5×5 array accounts for over 3,600 transistors. A total of 50 optical input/output channels are provided in this implementation.

Figure 9 shows the circuitry associated with a single neuron of the smart pixel error diffusion neural network. All state variables in each circuit of this architecture are represented as currents. Beginning in the upper-left of the circuit and proceeding clockwise, the input optical signal incident on the SEED is continuous in intensity and represents the individual input analog pixel intensity. The input SEED at each neuron converts the optical signal to a photocurrent and subsequently, current mirrors are used to buffer and amplify the signal. The width-to-length ratio of the metal oxide semiconductor field effect transistors (MOSFETs) used in the CMOS-SEED circuitry provide current gain to amplify the photocurrent. The first circuit produces two output signals: $+I_u$ which represents the state variable $u_{m,n}$ as the input to the quantizer and $-I_e$ which represents the state variable $-\varepsilon_{m,n}$ as the input to the feedback differencing node. The function of the quantizer is to provide a smooth, continuous threshold functionality for the neuron that produces the output signal I_{out}, corresponding to the output state variable $y_{m,n}$. This second electronic circuit is called a modified wide-range transconductance amplifier and produces a hyperbolic tangent sigmoidal function when operated in the sub-threshold regime. The third circuit takes as its input I_{out}, produces a replica of the original signal, and drives

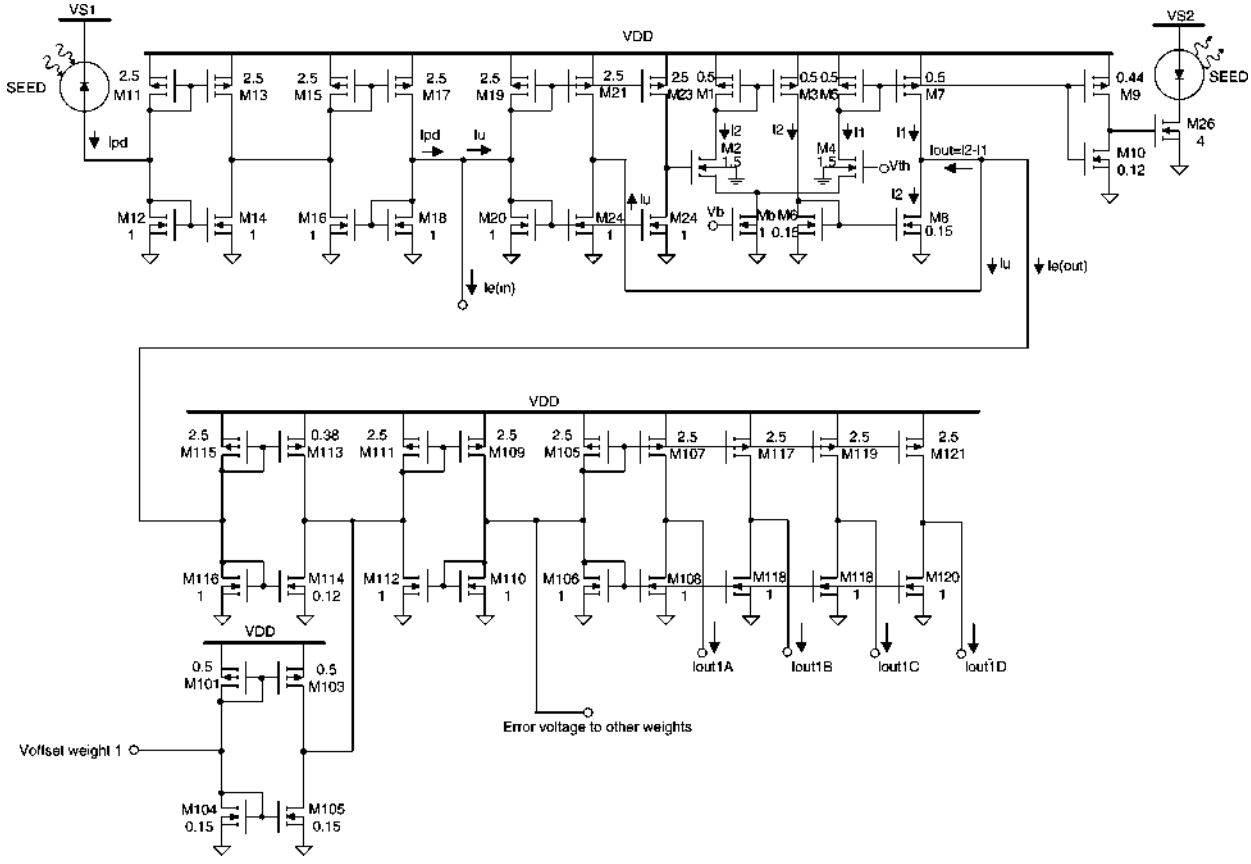

Figure 9 Circuit diagram of a single neuron and a single error weight of the 5×5 error diffusion neural network based on a CMOS-SEED-type smart pixel architecture. (Reprinted with permission from Shoop BL, *Photonic Analog-to-Digital Conversion*, Springer Series in Optical Sciences, Vol. 81 [1], Fig. 8.6, p. 222. Copyright 2001, Springer-Verlag GmbH & Co. KG.)

the output optical SEED. In this case, the output optical signal is a binary quantity represented as the presence or absence of light. When the SEED is forward-biased, light is generated through electroluminescence. The last circuit at the bottom of the schematic implements a portion of the error weighting and distribution function of the error diffusion filter. The individual weights are implemented by scaling the width-to-length ratio of the MOSFETs to achieve the desired weighting coefficients. The neuron-to-neuron interconnections are accomplished using the four metalization layers of the 0.5 μm silicon CMOS process. In this design, the error diffusion filter was limited to a 5×5 region of support because of the physical constraints of the circuitry necessary to implement the complete error diffusion neural network in silicon circuitry. The impulse response of the 5×5 filter used in this particular smart pixel architecture is shown in Figure 10. The error weighting circuitry at the bottom of Figure 9 represents only the largest weight (0.1124) with interconnects to its four local neighbors ($I_{\mathrm{outA}} - I_{\mathrm{outD}}$).

0.0023	0.0185	0.0758	0.0185	0.0023
0.0185	0.0254	0.1124	0.0254	0.0185
0.0758	0.1124	•	0.1124	0.0758
0.0185	0.0254	0.1124	0.0254	0.0185
0.0023	0.0185	0.0758	0.0185	0.0023

Figure 10 Error diffusion filter coefficients used in the 5×5 CMOS-SEED error diffusion architecture. ('•' represents the origin.)

Figure 11 shows a photomicrograph of a 5×5 error diffusion neural network, which was fabricated using CMOS-SEED smart pixels. Figure 12 shows a photomicrograph of a single neuron of the 5×5 CMOS-SEED neural network. The rectangular features are the MQW modulators while the silicon circuits are visible between and beneath the modulators. The MQW modulators are approximately 70 μm × 30 μm and have optical windows, which are 18 μm × 18 μm.

Simulations of this network predict individual neuron switching speeds of less than 1 μs, which corresponds to network convergence speeds capable of providing real-time digital image halftoning. Individual component functionality and dynamic operation of the full 5×5 neural array were both experimentally characterized. Figure 13 shows CCD images of the operational 5×5 CMOS-SEED smart pixel array. The white spots show SEED MQW modulators emitting light. Figure 13a shows the fully functioning array while Figure 13b shows the 5×5 CMOS-SEED neural array under 50% gray-scale input. Here 50% of the SEED modulators are shown to be in the *on*-state demonstrating correct operation of the network and the halftoning architecture.

Both the simulations and the experimental results demonstrate that this approach to a smart pixel implementation of the error diffusion neural network provides sufficient accuracy for the digital halftoning application. The individual neuron switching speeds also demonstrate the capability for this smart pixel hardware implementation to provide real-time halftoning of video images.

Figure 11 Photomicrograph of a smart pixel implementation of a 5×5 CMOS-SEED error diffusion neural network for digital image halftoning. (Reprinted with permission from Shoop BL, *Photonic Analog-to-Digital Conversion*, Springer Series in Optical Sciences, Vol. 81 [1], Fig. 8.5, p. 221. Copyright 2001, Springer-Verlag GmbH & Co. KG.)

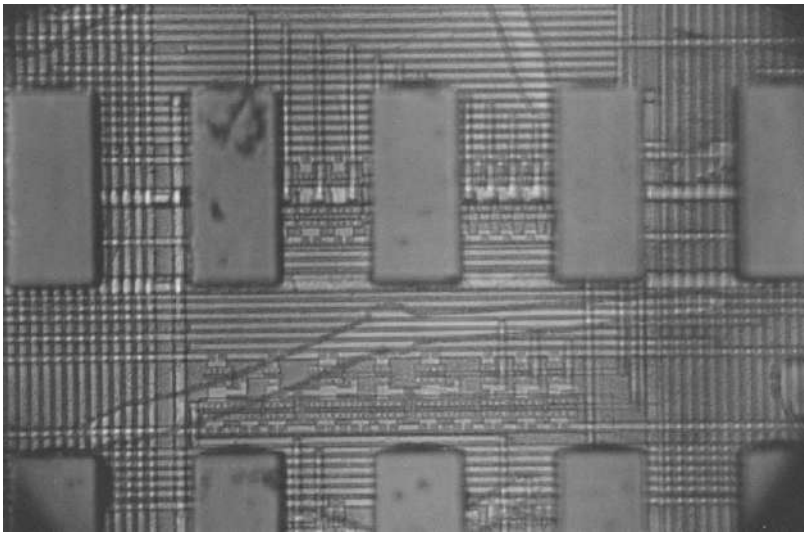

Figure 12 Photomicrograph of a single neuron of the 5 × 5 error diffusion neural network. (Reprinted with permission from Shoop BL, *Photonic Analog-to-Digital Conversion*, Springer Series in Optical Sciences, Vol. 81 [1], Fig. 8.4, p. 220. Copyright 2001, Springer-Verlag GmbH & Co. KG.)

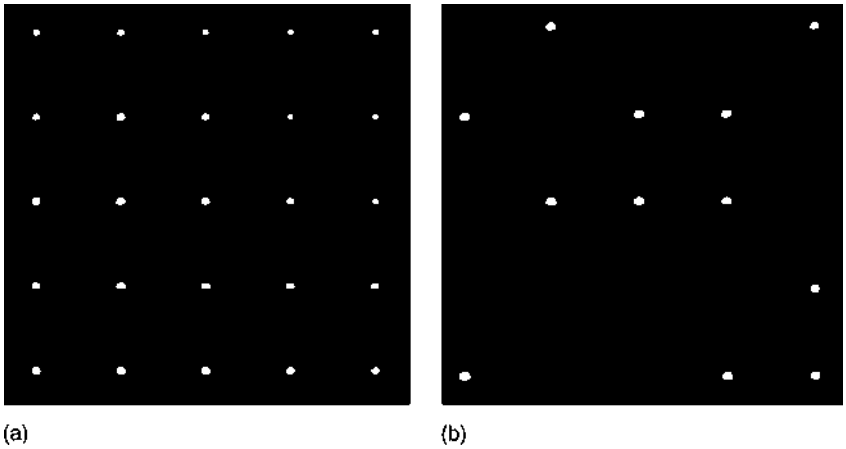

Figure 13 CCD images of the 5 × 5 CMOS-SEED array. (a) Fully-operational array, and (b) under 50% gray scale illumination. (Reprinted with permission from Shoop BL, *Photonic Analog-to-Digital Conversion*, Springer Series in Optical Sciences, Vol. 81 [1], Figs. 8.19 and 8.20, p. 236. Copyright 2001, Springer-Verlag GmbH & Co. KG.)

Image Processing Extensions

Other image processing functionality is also possible by extending the fundamental concepts of 2D, symmetric error diffusion to other promising image processing applications. One important application includes color halftoning while other extensions include edge enhancement and feature extraction. In the basic error diffusion neural network, the error diffusion filter was specifically designed to produce visually pleasing halftoned images. Care was taken to ensure that the frequency response of the filter was circularly symmetric and that the cutoff frequency was chosen in such a way as to preserve image content. An analysis of the error diffusion neural network shows that the frequency response of this filter directly shapes the frequency spectrum of the output halftone image and therefore directly controls halftone image content. Other filter designs with different spectral responses can provide other image processing features such as edge enhancement, which could lead to feature extraction and automatic target recognition applications.

See also

Detection: Smart Pixel Arrays. **Information Processing:** Optical Neural Networks. **Optical Processing Systems**.

Further Reading

Dudgeon DE and Mersereau RM (1984) *Multidimensional Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall.
Lau DL and Arce GR (2001) *Modern Digital Halftoning*. New York: Marcel Dekker.
Lim JS (1990) *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ: Prentice Hall.
Mead CA (1989) *Analog VLSI and Neural Systems*. New York: Addison-Wesley.
Poon T-C and Banerjee PP (2001) *Contemporary Optical Image Processing with Matlab*. New York: Elsevier.
Russ JC (1998) *The Image Processing Handbook*. Boca Raton, FL: CRC Press.
Shoop BL (2001) *Photonic Analog-to-Digital Conversion*. Springer Series in Applied Optics, vol. 81. Berlin: Springer-Verlag.
Shoop BL, Sayles AH and Litynski DM (2002) New devices for optoelectronics: smart pixels. In: De Cusatis (ed.) *Fiber Optic Data Communication: Technological Trends and Advances*, pp. 352–421. San Diego, CA: Academic Press.
Ulichney R (1987) *Digital Halftoning*. Cambridge, MA: The MIT Press.
Wax C (1990) *The Mezzotint: History and Technique*. New York: Harry N. Abrahms, Inc.

Optical Neural Networks

H J Caulfield, Fisk University, Nashville, TN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

An artificial neural network (ANN) is a nonlinear signal processing system based on the neural processes observed in animals. Usually they have multiple inputs and often multiple outputs also. Conventionally, each input sends its signal to many neurons, and each neuron receives signals from many inputs. The neuron forms an intermediate sum of weighted inputs and transforms that sum, according to some nonlinearity, to form an output signal. Often the output signal from one neuron is used as the input signal for another in feed forward, feedback, or mixed mode complex neural networks.

ANNs are of value in helping us understand how biological neural networks perform and in a variety of technical applications, most of which involve pattern association. An auto-associative neural network, that associates patterns with themselves, can have useful noise cleanup or pattern completion properties. A hetero-associative neural network is often used in pattern recognition to associate input patterns with preselected class labels.

Neural Networks – Natural and Artificial

The most primitive one-celled organisms have no nerve cells, but they do exhibit various chemical means for interacting with their environments, within their own cells, and among individuals of the same or different types. As organisms increased in size, a common chemical environment became impractical, and nerves arose as a means for long-distance communication within larger organisms. One thing seen in evolution is that widespread useful adaptations are often modified but seldom discarded. Thus, it should not be surprising to find that chemical interactions play a major role in the most complex thing we know of – the human brain. Chemicals play a huge role in making us who we are, setting our moods, and so forth. Realizing this, in the late twentieth century, psychiatrists started to make serious progress in treating mental diseases through the use of chemistry.

There are a variety of processes that we use as humans, but the primary means is neural. Only some animals and no plants have neurons, but soon after neurons evolved, they began to concentrate in special regions of the body that developed into primitive brains. Not only were there many neurons, but also they were densely and complexly interconnected to form neural networks.

Neural networks did not replace chemical interactions, they used and supplemented them. Chemical communication across the synaptic cleft dominates communication between networks, but the neural networks also control chemical production and distribution within the brain, and those chemicals change the behavior of the neurons that cause their secretion. The functioning of real brains is incredibly complex, possibly irreducibly complex in terms of current mathematics and logic.

These things are noted to avoid a confusion that is distressingly common. Technologists produce simplified ANNs and then try to interpret the behavior of brains in those terms. Quite often, ANNs yield useful insight into brain behavior but seldom offer a detailed account of them. Reality is far more complex than any ANN. Current mathematics provides no means

to describe continuous sets of events, each of which both causes and is caused by the others.

ANNs themselves have grown much more complicated as well. Starting out as simple, feed-forward, single-layer systems, they rapidly became bidirectional and even multilayered as well. Higher-order neural networks processed the input signals before inserting them into the neural network. Pulse coupled neural networks (PCNNs) arose that more closely modeled the pulsations in real brains. Learning was broken into learning with instruction, following example, and self-organizing, the latter leading to self organized maps and adaptive resonance theory, and many related subjects. These are but crude pointers to the wealth and variety of concepts now part of the ANN field.

The history of ANNs is both complicated and somewhat sordid. This is not the occasion to retrace the unpleasant things various members of this community have done to others in their quests for stature in the field. One result of that history is a dramatically up-and-down pattern of interest in and support for ANNs. The first boom followed the development of a powerful single-layer neural network called the Perceptron by its inventor Rosenblatt. Perceptrons were biologically motivated, fast, and reasonably effective; so interest in the new field of ANNs was high. That interest was destroyed by a book on the field arguing that Perceptrons were only linear discriminants and most interesting problems are not linearly discriminable. Ironically, the same argument has been made about optical Fourier correlators, and it is true there as well. But optical processing folk largely ignored the problem and kept working on it. Sociologists of science might enjoy asking why reactions differed so much in the two fields. Funding and interest collapsed and the field went into hibernation.

It was widely recognized that multilayer ANNs would comprise nonlinear discriminants, but training them presented a real problem. The simple rewards and punishment of weights depending on performance was easy when each output signal came from a known set of weighted signals that could be rewarded or punished jointly, according to how the output differed from the sought-after output. With a multilayer perceptron, credit assignment of inner layers is no longer simple. It was not until Werbos invented 'backward error propagation' that a solution to this problem was available and the field started to boom again. Also key to the second boom were enormously popular articles by Hopfield on rather simple auto- and hetero-associative neural networks. Basically, those networks are almost never used, but they were of great value in restarting the field. In terms of research and applications, we are still in that second boom, that began during the 1980s.

> An ANN is a signal processor comprised of a multiplicity of nodes which receive one or more input signals and generate one or more output signals in a nonlinear fashion wherein those nodes are connected with one or two way signaling channels.

Note also that biological inspiration is no longer a part of the definition. Most ANNs used for association are only vaguely related to their biological counterparts.

Optical Processors

Optics is well-suited for massive parallel interconnections, so it seems logical to explore optical ANNs. The first question to address is "Why bother to use optical processors at all?" as electronic digital computers seem to have everything:

- Between the writing of this article (on an electronic computer) and its publication, the speed of electronic computers will almost double and they will become even cheaper and faster.
- They can have arbitrary dynamic range and accuracy.
- Massively improved algorithms also improve their speed. For instance, the wonderful fast Fourier transform (FFT) that opticists have used to simulate optics for decades has been replaced by an even faster fastest Fourier transform in the West (FFTW). And for wavelets, the fast wavelet transform (FWT) is even faster.

Optical processors, on the other hand, have severely limited analog accuracy and few analog algotectures or archirithms. They are specialized systems (equivalent in electronics, not to computers but to application specific integrated circuits (ASICs)). They are usually larger and clumsier than electronic systems and always cost more.

There are several reasons why optics may still have a role in some computations. First, many things on which we wish to perform computations are inherently optical. If we can do the computations optically before detection, we can avoid some time and energy and sensitivity penalties inflicted, if we first detect the signals then process them electronically, and then (sometimes) convert them back to optics. Examples include spectroscopy and optical communication.

Second, pure optics is pure quantum mechanics. Input of data and setup of the apparatus is experiment preparation. Readout of data is classical measurement. Nature does the rest free of charge. There is a unitary transformation from input to output that

requires no human intervention and is, in fact, destroyed by human intervention. Those intermediate (virtual) computations require no energy and no time, beyond the time taken for light to propagate through the system. Speed and power consumption are limited only by the input and output. This applies for everything from Fourier optics to digital optics to logic gates. Thus, optics can have speed and power consumption advantages over electronics.

Third, if quantum computers involving entanglement ever emerge, they may well be optical. Such computers would be used for nondeterministic solutions of hard problems. That is, they would explore all possible paths at once and see which one does best. This is something we can already do electronically or optically for maze running (using classical, not quantum entanglement) means.

Fourth, most optical processors are analog, which is sometimes superior to digital. In those cases, we would still have to choose between analog optics and analog electronics, of course.

So, optical processing (like all other processing) is a niche market. But some of the niches may be quite large and important. It is not a universal solution, but neither is digital electronics.

The Major Branches of Optical Processing and Their Corresponding Neural Networks

Every branch of optical processing leads to its own special purpose optical neural networks, so those branches become a good place to start an article such as this one.

Optical Linear Algebra

Many neural network paradigms use linear algebra followed by a point-by-point nonlinearity in each of multiple stages. The linear operation is a matrix–vector multiplication, a matrix–matrix multiplication, or even a matrix–tensor multiplication. In any case, the key operations are multiplication and addition – operations very amenable to performance with analog optics. With effort, it is possible to combine simple electronics with analog optics operating on non-negative signals (the amount of light emitted from a source, the fraction of light passing through a modulator, etc.) to expand to real or even complex numbers. With even more effort, we can use multiple analog non-negative signals to encrypt a digital signal, allowing digital optical processors. Almost all optical neural networks using linear algebra use only real numbers. In most of these, the real number is encoded as a positive number and a negative number. Data are read in parallel using spatial light modulators (SLMs), source arrays, acousto-optic modulators, etc. They are operated upon by other SLMs, acousto-optic systems, and so forth. Finally, they are detected by individual detectors or detector arrays. The required nonlinearities are then done electronically. It is a theory that holds for all linear algebraic processes up to computational complexity $O(N^4)$ that they can be performed with temporal complexity $O(N^0)$ if the extra dimensions of complexity are absorbed in space, up to $O(N^2)$ through 2D arraying and in fanin/fanout, the ability of multiple beams of light to be operated upon in parallel by the same physical modulator – this has the same complexity as the 2D array of modulators, namely $O(N^2)$. Thus, we can invert a matrix, an order $O(N^4)$ for an $N \times N$ matrix, $O(N^0)$ in time, that is, independently of the size of N, as long as that matrix is accommodated by the apparatus. This offers a speed electronics cannot match.

Most of the optical neural networks of the mid-1980s had optical vector–matrix multipliers at their heart. Most were a few layers of feed forward systems, but by the 1990s feedback had been incorporated as well. These are typically N inputs connected to N outputs through an $N \times N$ matrix, with N varying from 100 to 1000. With diffuser interconnection and a pair of SLMs, it is possible to connect an $N \times N$ input with an $N \times N$ output using N^4 arbitrary weights, but that requires integrating over a time up to N^2 intervals. Most of the N^4-dimensional matrices can be approximated well by fewer than N^4 terms using singular value decomposition (SVD) and (as this version of the interconnection uses outer products), that reduces the number of integration times considerably in most cases. We have shown, for instance, that recognizing M target images can be accomplished well with only M terms in the SVD.

Coherent Optical Fourier Transformation

It took opticists several years, during the early 1960s, to realize that if Fourier mathematics was useful in describing optics, optics might be useful in performing Fourier mathematics. The first applications of optical Fourier transforms were in image filtering and particularly in pattern recognition. Such processors are attractive not only for their speed but also for their ability to locate the recognized pattern in 1D, 2D, or even 3D. It was soon obvious that a Fourier optical correlator was a special case of a single layer perceptron. Of course, it inherits the weaknesses of a single layer perceptron too. Soon, workers had added

feedback to create a winner-takes-all type of auto-associative optical neural network, a system that associated images with themselves. This proved useful in restoring partial or noisy images. It was not until 2003 that we learned how to make powerful hetero-associative networks with Fourier optical systems. This creates a nonlinear discrimination system that maintains the target location and generalizes well from a few samples of multiple classes to classify new data quite accurately.

Imaging

Imaging is the central act of optics and might seem to be of little value in optical processing. But multiple imaging (with lenslets, self-focusing fibers, holograms, and the like) has been the basis of numerous optical neural networks. Multiple imaging is a nonrandom way of distributing information from one plane to the next. Diffusers have also been used to produce fully interconnected optical neural networks. If each broad image point contains the whole image (something easily arranged with a diffuse hologram), then we have an auto-associative memory. Different such auto-associative networks can be superimposed in a single hologram. Then feedback after nonlinear image filtering can be used to make a general auto-associative memory.

Even deliberately poor imaging has been used. A blurred or out-of-focus image can accomplish the lateral signal spreading needed to facilitate a PCNN. These networks are substantially more like real brain networks than those discussed so far. It starts with a 2D array of integrate-and-fire units. The firing rate of each depends only in the incident power (how long integration to the thresholding firing takes). But if the fired signal is spread to its neighbors, the firing of a neighbor can hasten any unit's firing. This causes firings to synchronize and results in 'autowaves' of firing patterns moving across the 2D plane. Thus an incident 2D image is converted to a 3D signal – the two spatial dimensions and time. This results, for all but some very simple cases, in a chaotic signal occupying a strange attractor. It is interesting but not very revealing in itself. But suppose we 'integrate out' either the two spatial dimensions or the time dimension. If we integrate out the spatial dimensions, we are left with a time signal. Its strange attractors are remarkably useful in pattern recognition, because they are more syntactic than statistical (in pattern recognition terms). They describe the shape as similar, independently of the size, rotation, perspective, or even shading of the pattern! If we integrate out the time dimension, we restore the original image – almost. In fact, we can use this aspect of PCNNs for optimum blind segmentation of images. The autowaves can be made to 'leave tracks' so they can be reversed, which allows them to solve maze problems nondeterministically, because they can take all the paths. When the first wave exits, we simply retrace it to find the shortest path through the maze – a prototypical nondeterministic polynomial operation.

Holograms

Holograms are the most general, complex, and flexible means of light manipulation so far invented. Naturally, they have many applications in optical neural networks. One such application is the interconnection of each of an $N \times N$ array of inputs to each of an $N \times N$ array of outputs with its own unique connection strength (weight in the language of neural networks). The practical limit seems to be around $N = 1000$, yielding 10^{12} weights, roughly the same number as in the human brain! The brain, however, is not fully interconnected, so this produces an entirely different type of neural network. The resulting network would have fixed weights, but it could perform a variety of useful operations and be both forward and backward connected.

Special on-axis holograms, called diffractive optical elements, can be mass manufactured very effectively and inexpensively to perform tasks such as multiple imaging.

Photorefractive Neural Networks

A very complex thing happens when a pattern of fringes, such as those that form an interference pattern for hologram recording, strike materials such as lithium niobate, strontium barium niobate, gallium arsenide, and many others that fall into this category of 'photorefractive' materials. Of course, 'photo' refers to light and 'refractive' refers to the speed of light in a material – the speed of light in a vacuum divided by the index of refraction. These materials are photoconductive, so the light releases charge carriers in what would otherwise be an electrical insulator. The periodic charge pattern sets up a periodic space charge within the material. But the material is also electro-optic, that is, its index of refraction changes with applied electric field, thus generating a periodic index of refraction grating in the material. Finally, that index of refraction pattern (a phase hologram) diffracts the incoming light. Usually, but not always, a steady-state situation is reached after time that depends on the material and the photon density, although chaos can arise under some conditions. This makes for a dynamic

hologram, one that can change or adapt over time as the input pattern changes. Once a satisfactory photorefractive hologram is formed, it can be stabilized (fixed) if desired. Most optical neural network uses envision a dynamic situation with changing interconnections.

What has not been discussed above is that the index of refraction varies with the light's electric field vector cubed, so photorefractives provide a convenient way to accomplish such nonlinear operations as winner-takes-all.

Sometimes, we can record holograms with photorefractives that cannot be recorded by conventional holograms in conventional materials. To explain this, a hologram is a transducer between two optical wavefronts – usually called the reference and object beams. If the reference beam is incident on the hologram, it is (fully or partially) converted into the object beam. If the precisely reversed (phase conjugated) reference beam is incident on it, the precisely reversed (phase conjugated) object beam is derived from it. Likewise, the object wavefront (or its phase conjugate) produces the reference wavefront (or its phase conjugate). The hologram is made by recording the interference pattern between the two wavefronts: the reference and object beams. The ability of two wavefronts to form a temporally stable interference pattern at some point in space is called the mutual coherence of those wavefronts. Usually, both wavefronts are derived from the same laser beam to make achieving high mutual coherence possible. And, of course, both beams must be present to form an interference pattern. In some cases of photorefractive holograms, all of those conditions can be violated. The two wavefronts can derive from two sources that are different in wavelength, polarization, etc., and need not be simultaneously present. This phenomenon, called mutual phase conjugation, converts each wavefront into the phase conjugate of the other, but for effects due to any differences in wavelength and polarization. This allows new freedoms that have been exploited in optical neural networks.

Photorefractive materials also tend to be piezoelectric. It has been found that up to about 20 totally independent reflective holograms can be stored with different applied electric field in lithium niobate. Changing electric fields so changes the fringe spacings and indices of refraction, that only one of those holograms has any appreciable diffraction efficiency at any one electric field value. The result is an electrically selectable interconnection pattern. The same hologram can contain the information needed for many neural networks or multiple layers of fewer of them. There is no need, in general, to have more than an input layer, an output layer, and two intermediate or hidden layers.

Others have used photorefractive material to implement learning in neural networks. In biological neural networks the synaptic strengths (the weights between neurons) continue to adapt. With photorefractive holograms, we can achieve this in an optical ANN.

As previously discussed, a key development in the history of ANNs was a way to train multilayer neural networks and thus accomplish nonlinear discrimination. The first, and still most popular, way to do this is called backward error propagation. Several groups have been implemented in photorefractive optical neural networks.

Conclusions

Essentially every major type of neural network has been implemented optically. Every major tool of optics has found its way into these optical neural networks. Yet, it remains the case that almost every neural network used for practical applications is electronic. The mere existence proof of a use for optical neural networks may be what the field needs to move forward from the demonstration stage to the application stage. The tools exist and there are niches where optics seems more appropriate than electronics. Hopefully, the next review of the field will include a number of commercial successes. Nothing spurs science better than the funds that follow successful applications.

List of Units and Nomenclature

Adaptive resonance theory (ART)	Another self-organized neural net categorizer that works by mutual adjustment between bottom-up and top-down neural connections.
Computational complexity	The measure of how the number of required calculations scales with the problem size, usually denoted by N. The scaling is represented by the Big O form of (N), where $f(n)$ shows how it scales asymptotically (large N). Problems that scale as $O(N^P)$ are said to be of polynomial complexity. Unfortunately, some of the most important problems have exponential complexity.

Fanin/Fanout — Beams of light (unlike beams of electrons or holes) do not interfere with one another, so many of them can be modulated by the same physical elements quite independently, so long as they can be identified. This is usually done by sending the different beams in at different angles, giving rise to the words fanin and fanout, as suggested by the drawing below:

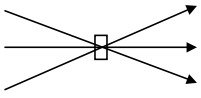

Self organized maps (Kohonen networks) — A means of self-organization of inputs into (normally 2D) maps wherein every concept falls into some category and the categories themselves develop out of application of the Kohonen's algorithm and user-selected parameters applied to the given data.

Singular value decomposition (SVD) — Any matrix can be broken into (decomposed) a weighted sum of simpler, rank-one matrices. The simplest such decomposition (in the sense of fewest number of terms) is the SVD. The weights are the singular values and are indicators of the importance of that term in the decomposition.

Spatial light modulator (SLM) — A 1D or 2D array of elements under electronic or optical control that change some property of the light incident on them. Usually, it is the index of refraction in some direction that is changed. That change can be used to modulate the amplitude, the polarization, or the phase of the light.

See also

Fourier Optics. **Holography, Techniques:** Overview. **Nonlinear Optics, Basics:** Photorefraction. **Optical Communication Systems:** Local Area Networks. **Optical Processing Systems**.

Further Reading

Abu-Mostafa YS and Psaltis D (1987) Optical neural computers. *Scientific American* 256(3): 88–95.

Anderson DZ and Erie MC (1987) Resonator memories and optical novelty filters. *Optical Engineering* 26: 434–444.

Brady D, Psaltis D and Wagner K (1988) Adaptive optical networks using photorefractive crystals. *Applied Optics* 27: 1752–1759.

Caulfield HJ, Kinser J and Rogers SJ (1987) Optical neural networks. *Proc. IEEE* 77: 1573–1583.

Chavel P, Lalanne P and Taboury J (1989) Optical inner-product implementation of neural network models. *Applied Optics* 28: 377–385.

Collins SA, Ahalt SC, Krishnamurthy AK and Stewart DF (1993) Optical implementation and applications of closest-vector selection in neural networks. *Applied Optics* 32: 1297–1303.

Farhat NH, Psaltis D, Prata A and Paek E (1985) Optical implementation of the Hopfield model. *Applied Optics* 24: 1469–1475.

Galstyan T, Frauel Y, Pauliat G, Villing A and Roosen G (1997) Topological map from a photorefractive self-organizing neural network. *Optics Communications* 135: 179–188.

Goodman JW, Dias AR and Woody JM (1978) Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms. *Optics Letters* 2: 1–3.

Johnson JL (1994) Pulse-coupled neural nets: Translation, rotation, scale, distortion, and intensity signal invariance for images. *Applied Optics* 33: 6239–6253.

Jutamulia S (1994) *Selected Papers on Optical Neural Networks*. Bellingham: SPIE Press.

Kinser JM, Caulfield HJ and Shamir J (1988) Design for a massive all-optical bidirectional associative memory: the big BAM. *Applied Optics* 27: 3442–3444.

Neifeld MA and Psaltis D (1993) Optical implementations of radial basis classifiers. *Applied Optics* 32: 1370–1379.

Owechko Y, Dunning GJ, Marom E and Soffer BH (1987) Holographic associative memory with nonlinearities in the correlation domain. *Applied Optics* 26: 1900–1910.

Paek EG and Psaltis D (1987) Optical associative memory using Fourier-transform holograms. *Optical Engineering* 26: 428–433.

Psaltis D and Farhat NH (1985) Optical information processing based on associative-memory model of neural nets with thresholding and feedback. *Optics Letters* 10: 98–100.

Soffer BH, Dunning GJ, Owechko Y and Marom E (1986) Associative holographic memory with feedback using phase-conjugate mirrors. *Optics Letters* 11(2): 118–120.

Yu FTS (1993) Optical neural networks – architecture, design and models. In: Wolf E (ed.) *Progress in Optics*, II, pp. 63–144. Amsterdam: North-Holland.

INSTRUMENTATION

Contents

Astronomical Instrumentation
Ellipsometry
Photometry
Scatterometry
Spectrometers
Telescopes

Astronomical Instrumentation

J Allington-Smith, University of Durham, Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

This article describes the requirements placed on astronomical instruments and the methods used to achieve them, making reference to generic techniques described elsewhere in this encyclopedia. It is limited in scope to electromagnetic radiation with wavelengths from the atmospheric cutoff in the ultraviolet (0.3 μm) to the point in the near-infrared where the technology changes radically (5 μm), taking in the transition from uncooled to cooled instruments at ~1.7 μm. Thus, it excludes the extreme UV and X-ray regimes, where standard optical techniques become ineffective (e.g., requiring grazing incidence optics) and all mid-infrared, sub-mm, THz, and radio-frequency instruments. Terrestrial telescopes are the main focus but space-borne telescopes are subject to similar design considerations, except that the diffraction limit of the optics may be achieved at short wavelengths. Adaptive optics (AO), of increasing importance in astronomy and almost essential for the next generation of extremely large telescopes (ELTs), is covered elsewhere in this encyclopedia.

Astronomical Data

The purpose of astronomical instruments is to measure the following properties of photons arriving on or near to the Earth and brought to a focus by an astronomical telescope:

(i) direction of arrival;
(ii) energy;
(iii) polarization state; and
(iv) time of arrival.

In practice, these properties cannot be measured simultaneously, requiring the following generic classes of instrument:

- *Imagers* to measure (i) and, if equipped with ultra-low-noise detectors, (iv);
- *Spectrographs* to measure (ii), although some can be configured to do a good job on (i) as well. *Integral field spectrographs* can measure (i) and (ii) simultaneously; and
- *Polarimeters* for (iii) plus (i) and *spectropolarimeters for* (iii) plus (ii).

In principle, energy-resolving detectors, such as arrays of superconducting tunnel-junction devices (STJs) can replace (i), (ii), and (iv).

These must interface to the telescope which provides a focal surface with an image scale which is approximately constant over the field. Currently, almost all telescopes are based on the Cassegrain or Ritchy-Chretien reflecting configurations, comprising a large curved primary mirror (M1) which defines the collecting aperture (diameter D_T) and a curved secondary mirror (M2). This feeds the Cassegrain focus via a hole in M1 in line with the central obstruction caused by M2, or the Nasmyth focus if a tertiary (usually plane) mirror is added to direct light from M2 along the elevation axis to a location in which the gravity vector changes only in rotation about the optical axis. For the Cassegrain focus, the full range of change in the gravity vector is experienced as the telescope tracks objects during long exposures or slews between targets. It is also possible to dispense with M2 and mount the

Table 1 Optical interfaces to representative types of telescope and achievable angular resolution

Aperture diameter D_T (m)	*Primary focus*			*Final focus*			*Diffraction limit $2.44\lambda/D_T$(mas)*	
	F/#	*Image scale ($\mu m/''$)*	*FOV ($'$)*	*F/#*	*Image scale ($\mu m/''$)*	*FOV ($'$)*	*$\lambda = 0.5\ \mu m$*	*$\lambda = 5.0\ \mu m$*
2.4[i]	2.3			24	279	25	101	1007
4.2[ii]	2.8	60	40	11	221	15	60	600
8.0[iii]	1.8	(70)	(45)	16	621	10	32	315
30[iv]	1–1.5	~180	~20	15–19	~2500	~20	8.4	84
100[v]	~1.4			~6.5	~3000	~10	2.5	25
			For enclosed telescopes				500–1500	
All	Obtainable angular resolution for terrestrial telescopes		Free-atmosphere seeing at good site			Best	250	160
						Median	420	270
			With adaptive optics (approx)				50–200	*Diff. limit*[a]

[a]The diffraction limit listed in the column above for $\lambda = 5.0\ \mu m$ applies.
Exemplars: (i) Hubble Space Telescope; (ii) Herschel Telescope, La Palma; (iii) Gemini telescopes, Hawaii/Chile; (iv) CELT/GSMT (USA); (v) OWL (Europe).

instrument directly at the focus of M1 (prime focus), although this imposes severe restrictions on the mass and size of the instrument. An alternative mounting for instruments is via optical fibers. The input, with any fore-optics required to couple the telescope beam into the fiber, is mounted at the appropriate focus while the output, possibly several tens of meters distant, is mounted where the instrument is not subjected to variations in the gravity vector.

The useful field of view (FOV) of the telescope is determined by how rapidly the image quality degrades with radius in the field; restrictions on the size of the Cassegrain hole and the space set aside for the star trackers required for accurate guiding and auxiliary optics involved in target acquisition, AO, etc.

Typical telescope parameters are shown in Table 1, of which the last two entries are speculative examples of ELTs. The table also lists the diffraction limit of the telescopes and the limits imposed by seeing, with and without the effects of the telescope structure and enclosure, and approximate limits for performance with AO which approaches the diffraction limit in the infrared.

Instrument Requirements

Efficiency

Almost all astronomical observations are of very faint objects or require very high signal-to-noise (SNR) to extract the desired subtle information. The former, for example, might reveal the initial phase of galaxy or star formation, while the latter could reveal the mix of elements in the big bang or uncover changes in fundamental constants on cosmological timescales ($\sim 10^{10}$ years). The term 'efficiency' has several components:

(i) Throughput of the optical system, including the efficiency of the optical components (lenses or mirrors), dispersers, filters, and also detectors and also includes losses due to transmission/reflection, diffraction, scattering, and vignetting by internal obstructions;
(ii) Noise from all sources: the detector, including time-independent (read-noise) and time-dependent (dark-current) components; thermal emission from the instrument, telescope, enclosure and sky; straylight (see above), and electronic interference;
(iii) Efficient use of the instrument and telescope to maximize the fraction of the time spent accumulating useful photons instead of, say, calibrating; reconfiguring the instrument; or acquiring targets;
(iv) Parallelizing the observations so that the same instrument records data from many objects at the same time. This is discussed further below.

Angular Resolution

The required angular resolution is summarized in Table 2 for key features in galaxies. The resolutions currently achievable with AO systems (with severe restrictions on sky coverage pending routine availability of laser beacons), or without, are shown in bold. Finer resolution is possible using interferometric techniques, but is limited to a handful of bright stars.

Table 2 Typical requirements for angular resolution and single-object field of view

Target	*Distance*[a]	*Active nucleus*		*Star cluster (10 pc)*	*Visible extent*
		Central engine (10 AU)	*Dynamical core (0.5 pc)*		
Milky Way	8 kpc	1 mas	10″	4′	
Local group galaxy (M31)	800 kpc	0.01 mas	100 mas	3″	100″
Distant galaxy	>1 Gpc	10^{-5} mas	0.1 mas	2 mas	0.5–2″
Cluster of galaxies	>1 Gpc				3′
Cosmological structure	>1 Gpc				~1°

[a]1 pc = 3.1×10^{16} m.

Field of View and Multiplexing

The FOV of an individual observation must be sufficient to cover the angular extent of the object with the minimum number of separate pointings of the telescope. As well as minimizing the total observation time, this reduces photometric errors caused by changes in the ambient conditions (e.g., flexure, thermal effects, and changes in the sky background). (Photometry is the measurement of the flux density received from an astronomical object, usually done with the aid of an imager employing filters with standardized broad passbands. Spectrophotometry is the same but in the much narrower passbands obtained by dispersing the light.) The last column of Table 2 indicates the required size of FOV. It is also highly desirable to observe many objects simultaneously since most studies require large statistically homogenous samples. The precise requirement depends on the surface number density of potential targets and the sampling strategy adopted to create the sample. Cosmological studies of the large-scale distribution of galaxies and clusters emphasize the size of the field ($\geq 0.5°$), since the targets may be sparsely sampled, while studies of clustered objects (galaxies or stars) aim to maximize the surface density of targets (≤ 1000 arcmin^{-2}) within a modest FOV ($\sim 5'$).

Spectral Resolution

The required energy resolution is specified in terms of the spectral resolving power, $R \equiv \lambda/\delta\lambda$, where $\delta\lambda$ is the resolution in wavelength defined usually in terms of the Rayleigh criterion. One of the main data products required in astrophysics is the velocity along the line of sight (radial velocity) obtained through measurement of the wavelength of a known spectral feature (emission or absorption line). This may be determined to an accuracy of:

$$\delta v = \frac{Qc}{R} \qquad [1]$$

where Q is the fraction of the width of the spectral resolution element, $\delta\lambda$, to which the centroid can be determined. This is limited by either the SNR of the observation of the line via $Q \approx 1/\mathrm{SNR}$, or by the stability of the spectrograph via:

$$Q = \frac{\Delta x}{\chi}\left(\frac{\mathrm{d}\chi}{\mathrm{d}x}\right) \qquad [2]$$

where χ is the angular width of the slit measured on the sky and $\delta\chi/\mathrm{d}x$ is the image scale at the detector. Δx is the amount of uncorrectable flexure in units of the detector pixel size. This is the part of the motion of the centroid of the line which is not susceptible to modeling with the aid of frequent wavelength calibration exposures. Typically, $\Delta x \approx 0.2$ pixels is achievable (roughly 3 μm at the detector or 10 μm at the slit for an 8 m telescope) implying $Q \approx 1/30$ for a typical $(\mathrm{d}\chi/\mathrm{d}x) = 0.1–0.2''$/pixel and $\chi = 0.5–1''$. Table 3 gives typical requirement for R and the value of Q required to achieve it, given the spectral resolution obtainable for different types of current spectrograph. Extrasolar planet studies require very high stability.

Some spectral observations require very accurate measurements of the shape of the spectral features to infer the kinematics of ensembles of objects which are individually not resolvable (e.g., stars in distant galaxies, gas in blackhole accretion disks). Other applications require accurate measurements of the relative flux of various spectral features within the same spectrum (e.g., to reveal abundances of elements or different types of star).

Optical Principles

Imaging

The simplest type of imager consists of a detector, comprising a 2D array of light-sensitive pixels (Table 4) placed at a telescope focal surface without any intervening optics except for the filters required for photometry. Although maximizing throughput, this arrangement removes flexibility in changing the image scale since the physical size of pixels (typically 15–25 μm) is fixed. Thus the Gemini telescopes

Table 3 Typical requirements for spectral resolution and stability

Target	*δv or required data product*	*Required λ/δλ*	*Spectral regime*	*Available R*	*Stability required = 1/Q*	*Simultaneous wavelength range*
Cosmological parameters	1000 km/s (line centroid)	300	*Low dispersion*	300–3000	1	≳1 octave
Galaxies						
Intra-cluster	100 km/s (line centroid)	3000			1–10	100 nm
Element abundance	Ratio of line fluxes	1000–3000				1 octave
Stellar orbits	10 km/s (line shapes)	3×10^4	*Medium*	$1–3 \times 10^4$	(1)	10 nm
Individual stars	1 km/s (line centroid)	3×10^5			10–30	10 nm
Extrasolar planets	<1 m/s (line centroid)	$>3 \times 10^8$	*High*	$\sim 10^5$	>3000	<1 nm

Table 4 Detectors in current use for astronomy

Wavelength (μm)	*Material/ type*	*Pixel size (μm)*	*Format*		*Read noise (e-/pix)*	*Dark current (e-/pix/hr)*	*QE range (%)*
			Single	*Mosaic*			
		Typical	*State-of-the-art*				
0.3–1.0	CCD	10–20	4k × 4k	12k × 8k	2	Negligible	10–90[a]
	CMOS	5–10	4k × 4k		5–30		
1–2.5 +	HgCdTe	15–20					
0.6–5	InSb	20–30	2k × 2k	4k × 4k	10	0.01–0.1	70–90

[a]Quantum efficiency varies strongly over quoted wavelength range.

($D_T = 8$ m) provides an uncorrected image scale of 25–40 mas/pixel at its sole F/16 focus. Although well-suited to high-resolution observations with AO, this is inappropriate for most observations of faint galaxies where 100–200 mas/pixel is required. This arrangement also provides limited facilities for defining the wavelength range of the observations since there is no intermediate focal surface at which to place narrowband filters, whose performance is sensitive to field angle and are, therefore, best placed at an image conjugate. There is also no image of the telescope pupil at which to place a cold stop to reject the external and instrumental thermal radiation which is the dominant noise source at $\lambda > 1.5$ μm.

The alternative is to use a focal-reducer consisting of a collimator and camera. The auxiliary components (filters, cold-stops, Fabry–Perot etalons, etc.) then have an additional image and pupil conjugate available to them and the image scale may be chosen by varying the ratio of the camera and collimator focal lengths. The extra complexity of optics inevitably reduces throughput but, through the use of modern optical materials, coatings such as Sol–Gel and the avoidance of internally obstructed reflective optics, losses may be restricted to 10–20% except at wavelength shorter than ~0.35 μm. An imaging focal reducer is a special case of the generic spectrograph described in the next section.

Spectroscopy

Basic principles

Astronomical spectrographs generally employ plane diffraction gratings. The interference between adjacent ray paths as they propagate through the medium is illustrated in Figure 1. From this, we obtain the grating equation:

$$m\rho\lambda = n_1 \sin\alpha + n_2 \sin\beta \equiv G \qquad [3]$$

where α and β are the angles of incidence and diffraction respectively, $\rho = 1/a$ is the ruling density, λ is the wavelength, and m is the spectral order. The refractive indices in which the incident and diffracted rays propagate are n_1 and n_2 respectively.

The layout of a generic spectrograph employing a plane diffraction grating in a focal reducer arrangement is shown in Figure 2. The spectrograph re-images the slit onto the detector via a collimator and camera. The disperser is placed in the collimated beam close to a conjugate of the telescope pupil.

A field lens placed near the focus is often incorporated in the collimator for this purpose.

The geometrical factor, G, defined in the equation is constrained by the basic angle of the spectrograph:

$$\psi = \alpha - \beta \qquad [4]$$

which is the (normally fixed) angle between the optical axes of the collimator and camera. $\psi = n\pi$, corresponds to the Littrow configuration (for integer n).

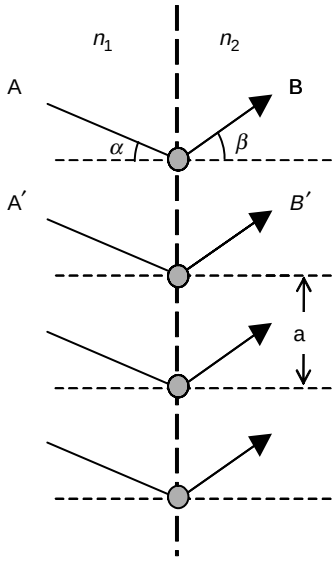

Figure 1 Derivation of the grating equation by consideration of the optical path difference between neighboring ray paths AB and A′B′.

Considering the case of a diffraction grating in vacuo, differentiation of the grating equation with respect to the diffracted angle yields the angular dispersion:

$$\frac{d\lambda}{d\beta} = \frac{\cos\beta}{m\rho} \qquad [5]$$

from which the linear dispersion is obtained by considering the projection of $d\beta$ on the detector, dx:

$$\frac{d\lambda}{dx} = \frac{d\lambda}{d\beta}\frac{d\beta}{dx} = \frac{\cos\beta}{m\rho f_2} \qquad [6]$$

where f_2 is the focal length of the camera (see Figure 2).

In the diffraction-limited case, where the slit is arbitrarily narrow, the resolving power is given simply by the total number of rulings multiplied by the spectral order:

$$R^* = m\rho W \qquad [7]$$

where w is the length of the disperser as defined below. But in astronomy, it is usually determined by the width of the image of the slit, s, projected on the detector, s'. By conservation of Etendue:

$$s' = s\frac{F_2}{F_1} \qquad [8]$$

where F_1 and F_2 are the collimator and camera focal ratios, respectively. The spectral resolution of the spectrograph is simply the width of this expressed in

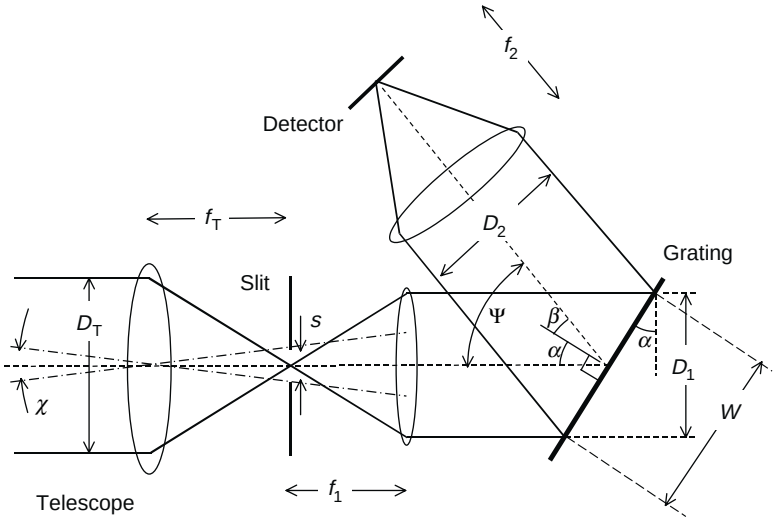

Figure 2 Generic spectrograph employing a plane diffraction grating. Note that according to the sign convention implicit in Figure 1, the diffraction angle, β, is negative in this configuration.

wavelength units:

$$\delta\lambda = \left(\frac{\mathrm{d}\lambda}{\mathrm{d}x}\right)s' = \frac{\cos\beta}{m\rho f_2}s\frac{F_2}{F_1} = \frac{sD_1\cos\beta}{m\rho D_2 f_1} \qquad [9]$$

The length of the intersection between the collimated beam and the plane of the grating (not necessarily the actual physical length of the grating) is

$$W = \frac{D_2}{\cos\beta} = \frac{D_1}{\cos\alpha} \qquad [10]$$

Using this to eliminate $\cos\beta$, the spectral resolution becomes

$$\delta\lambda = \frac{s}{m\rho F_1 W} \qquad [11]$$

and the resolving power is

$$R \equiv \frac{\lambda}{\delta\lambda} = \frac{m\rho\lambda F_1 W}{s} = G\frac{F_1 W}{s} \qquad [12]$$

where G is the quantity defined in eqn [3]. Note that R is now independent of the details of the camera. This expression is useful for the laboratory since it is expressed in terms of the parameters of the experimental apparatus; but, for astronomy, it is more useful to express the resolving power in terms of the angular slit width, χ, and the telescope aperture diameter, D_T, via:

$$s = \chi f_\mathrm{T} = \chi F_\mathrm{T} D_\mathrm{T} = \chi F_1 D_\mathrm{T} \qquad [13]$$

since the collimator and telescope focal ratios are the same if the spectrograph is directly beam-fed from the telescope. Note that even if the telescope focal surface is re-imaged onto the slit, the expression for the resolving power still holds due to the conservation of Etendue in the re-imaging optics. Thus the resolving power is

$$R = \frac{m\rho\lambda W}{\chi D_\mathrm{T}} = \frac{GW}{\chi D_\mathrm{T}} \qquad [14]$$

Note that the resolving power obtained with a nonzero slit width is always less than the theoretical maximum, $R \leq R_*$, for wavelengths:

$$\lambda < \lambda_* = \chi D_\mathrm{T} \qquad [15]$$

Thus, long-wavelength applications may approach the theoretical limit; in which case they are said to be diffraction-limited. If so, the resolving power is independent of the slit width, simplifying the interface to different telescopes. For a nondiffraction-limited spectrograph, the resolving power obtained depends on the aperture of the telescope to which it is fitted.

Optical configuration

The non-Littrow configuration shown in Figure 2 introduces anamorphism into the recorded spectra such that a monochromatic image of the slit mask is compressed in the dispersion direction by a factor:

$$A = \frac{D_2}{D_1} = \frac{\cos\beta}{\cos\alpha} \qquad [16]$$

which varies, depending on the choice of angles required to place the chosen wavelength on the detector using eqns [3] and [4].

Practical configurations that maximize R, do so by increasing $W = D_1/\cos\alpha$. This implies maximizing $\cos\beta$, since $\psi = \alpha - \beta$, which results in $A > 1$. This is known as the grating-to-camera configuration since the grating normal points more towards the camera than to the collimator. The grating-to-collimator configuration may also be used at the penalty of lower R but does permit a smaller camera. Thus the camera aperture must be oversized in the dispersion direction since $D_2 > D_1$. The Littrow configuration is rarely used with reflective designs when large FOV is required because of internal vignetting and consequent light loss. However, it can be used effectively with transmissive dispersers.

If the disperser is replaced by a plane mirror (or if the disperser is removed in a transmissive system), the system acts as a versatile imager, with $A = 1$. This also facilitates target acquisition which requires that the telescope attitude be adjusted until the targets are aligned with apertures cut in an opaque mask placed at the input of the spectrograph. These apertures (normally slits) are usually oversized in the direction perpendicular to the dispersion to allow the spectrum of the sky to be recorded directly adjacent to that of the target to permit accurate subtraction of the background.

The slit width is chosen by trading off the desire to maximize the amount of light entering the instrument from the target (but not so much as to admit excessive background light) and the need to maximize spectral resolving power, R, by reducing the slit width, χ, according to eqn [14].

Since $|G| \leq 2$, the maximum attainable resolving power, according to eqn [14], is determined chiefly by the telescope aperture and angular slitwidth. The only important parameter of the spectrograph is the illuminated grating length, W. Therefore, maintaining the same resolving power as the telescope aperture increases, requires W to increase in direct proportion. Because of limits placed on the geometry by the need to avoid obstructions, etc., this means that both the collimated beam diameter and the size of the disperser must increase. This is one of the major

problems in devising instruments for ELTs. Current instruments on 8 m telescopes feature collimated beam diameters of $D_1 \approx 150$ mm. For ELTs, the implied sizes are 0.5–2 m, requiring huge optics, accurately co-aligned mosaics of dispersing elements, and enormous support structures to provide the required stability. A simple calculation indicates that the ruling stylus for a classically ruled diffraction grating would have to cover a thousand kilometers without significant wear unless the disperser was made from smaller pieces.

Choice of dispersing element

Astronomical spectrographs mostly use surface-relief diffraction (SL) gratings. These are ruled with groove densities $50 < \rho < 3000\ \mathrm{mm}^{-1}$.

Blazing diffraction gratings. The intensity of the multibeam interference pattern from N rulings of finite width is given by

$$I = \left(\frac{\sin^2 N\phi}{\sin^2 \phi}\right)\left(\frac{\sin^2 \theta}{\theta^2}\right) \qquad [17]$$

where 2ϕ is the phase difference between the center of adjacent rulings and θ is the phase difference between the center and edge of a single ruling. To be useful for astronomy, the peak in this pattern, determined by the second term, must coincide with a useful order, such as $m = 1$, rather than zero order. This can be done by introducing a phase shift into the second term, which represents the diffraction of a single ruling, such that:

$$\theta = \frac{\pi \cos \gamma}{\rho\lambda}(\sin i - \sin r) \qquad [18]$$

where the groove angle, γ, is the angle by which the facets are tilted with respect to the plane of the grating. With the aid of Figure 3, it can be seen that $i = \alpha - \gamma$ and $r = \gamma - \beta$ so that the maximum efficiency, which occurs when $\theta = 0$, corresponds to $i = r$, which is equivalent to simple reflection from the facets, and $\alpha + \beta = 2\gamma$.

Thus, through this process of blazing, the grating eqn [3], at the blaze wavelength, λ_B, becomes:

$$\rho m \lambda_B = 2\sin \gamma \cos\frac{\psi}{2} \qquad [19]$$

where ψ, is given by eqn [4] using the identity: $\sin x + \sin y \equiv 2 \sin\left(\frac{x+y}{2}\right) \cos\left(\frac{x+y}{2}\right)$. Thus, the resolving power at blaze is obtained from

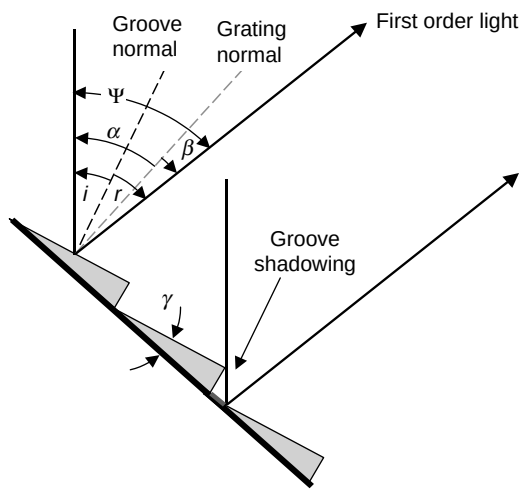

Figure 3 Blazing of ruled diffraction grating.

eqns [10], [14], and [19] as:

$$R_B = \frac{2D_1}{\chi D_T}\frac{\sin\gamma \cos(\psi/2)}{\cos(\gamma + \psi/2)} \qquad [20]$$

Use of grisms. An alternative configuration uses transmission gratings. The advantage here is that the collimator and camera can share the same optical axis in a straight-through configuration with unit anamorphism, $A = 1$, which does not require oversizing of the camera. The same phase-shift can be applied as before by making the facets into prisms with the required groove angle. However, this would only allow zero-order to propagate undeviated into the camera so an additional prism is required with vertex angle $\phi = \gamma$ to allow first-order light to pass undeviated into the camera. This composite of blazed transmission grating and prism is called a grism (Figure 4).

The grating equation (eqn [3]) is modified for a grism in the blaze condition to $m\rho\lambda_B = (n-1)\sin\phi$ since $\alpha = -\beta = \phi$ and $n_1 \equiv n, n_2 = 1$. Noting that $\psi = 0$, the resolving power at blaze is then:

$$R_B = \frac{D_1}{\chi D_T}(n-1)\tan\phi \qquad [21]$$

where n is the refractive index of the prism and grating material. Due to problems such as groove shadowing (illustrated for reflection gratings in Figure 3), these lack efficiency when $\phi \gtrsim 30°$ restricting their use to low-resolution spectroscopy.

Use of volume phase holographic gratings. A newer alternative to surface-relief (SR) dispersers are volume phase holographic (VPH) gratings in which

the interference condition is provided by a variation of refractive index of a material such as dichromated gelatine, n_g, which depends harmonically on position inside the grating material as:

$$n_g(x,z) = \bar{n}_g + \Delta n_g \cos[2\pi\rho_g(x\sin\gamma + z\cos\gamma)] \qquad [22]$$

where ρ_g is the density of the lines of constant n_g and $\bar{n}_g$ and Δn_g are the mean refractive index and its modulation amplitude, respectively (Figure 5). These are described by the same grating (eqn [3]), as before leading to an identical expression for the resolving power at blaze, where $m\rho\lambda_B = 2\sin\alpha$, to that of a blazed plane diffraction grating in the Littrow configuration (eqn [20]):

$$R_B = \frac{D_1}{\chi D_T} 2\tan\alpha \qquad [23]$$

Unlike SR gratings, the blaze condition of VPH gratings may be varied by changing the basic angle of the spectrograph, $\psi = 2(\gamma + \alpha)$, although this is often mechanically inconvenient. They may also be sandwiched between prisms to provide the necessary incident and diffraction angles while retaining the advantage of a straight-through configuration, to form an analogue of a grism, sometimes termed a vrism. Unlike grisms, these are not constrained by groove shadowing and so can be efficient at relatively high dispersion, subject to the constraint imposed by the size of the device.

Use of echelle gratings. Another option to increase the resolving power is to use a coarse grating in high order via the use of an echelle format. The resolving power in the near-Littrow configuration in which it is usually operated, is given by eqn [20] with $\psi = 0$.

$$R_B = \frac{D_1}{\chi D_T} 2\tan\gamma \qquad [24]$$

where γ is the groove angle.

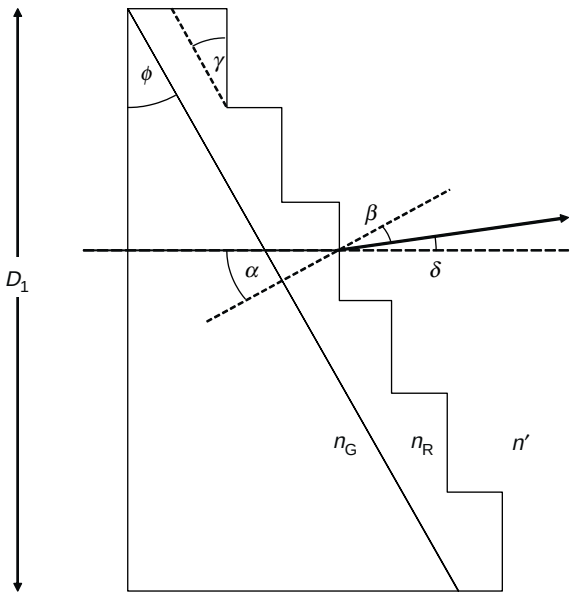

Figure 4 Typical configuration of a grism. The prism vertex and groove angles are chosen for maximum efficiency on the assumption that refractive indices, $n_G = n_R = n$ and $n' = 1$. The blaze condition occurs when $\delta = 0$. In this configuration, $\phi = \gamma = \alpha$.

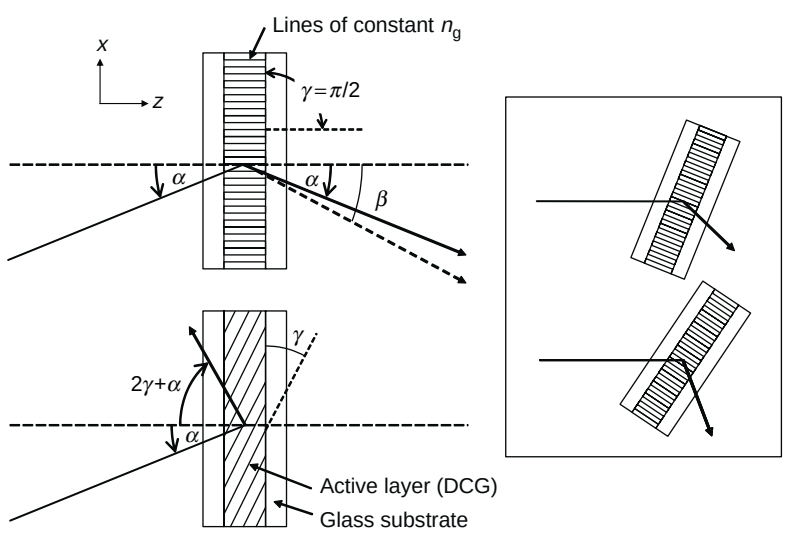

Figure 5 Various configurations of volume phase holographic gratings. The blaze condition is when $\beta = \alpha$. The inset shows how the blaze condition may be altered by changing the basic angle of the spectrograph.

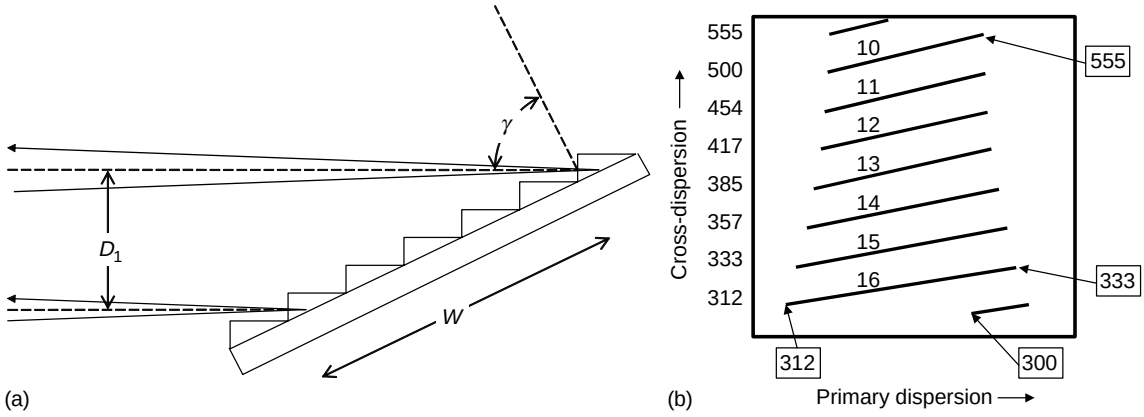

Figure 6 (a) Basic layout of an Echelle grating used in a near-Littrow configuration. (b) Illustrative cross-dispersed spectrogram showing a simplified layout on the detector. The numbers 10–16 label the different spectral orders. The numbers labeling the vertical axis are the wavelength (nm) at the lowest end of each complete order. Other wavelengths are labeled for clarity. For simplicity the orders are shown evenly spaced in cross-dispersion.

In the configuration shown in Figure 6, R is maximized by increasing W. This also means that $G = m\rho\lambda$ is large. In order to avoid excessively high ruling densities, the grating may be used in high orders. However, this presents the problem of order overlap since the wavelength in order m occurs in the same place as wavelength $\lambda_n = (m/n)\lambda_m$ in order n. Limiting the observation to a single order via the use of order-sorting filters is one solution while another is to use cross-dispersion via a chain of prisms or a grating with dispersion perpendicular to that of the echelle to separate the orders out. This option severely curtails the use of such spectral formats in multi-object spectroscopy.

Increasing resolving power using immersed gratings. All the methods of using diffraction gratings discussed so far are subject to various geometric constraints which ultimately limit the maximum obtainable resolving power by placing limits on W (e.g., in eqn [14]). These may be partly overcome by optically coupling the dispersers to immersing prisms. This can lead to a doubling of the resolving power in some cases.

Use of prisms. The final method of dispersion is to use a prism. This has the advantage that, since it does not work by multibeam interference, the light is not split into separate orders which removes the problem of order-overlap and improves efficiency. However, it does not produce the high spectral resolution required for some applications and the dependency of dispersion on λ is often markedly nonlinear.

From a consideration of Fermat's principle, it can be shown that the resolving power of a prism is:

$$R = \frac{\lambda t}{\chi D_{\mathrm{T}}}\left(\frac{\mathrm{d}n}{\mathrm{d}\lambda}\right) \qquad [25]$$

where t is the baselength of the prism. Using modern materials, $R \lesssim 300$ may be obtained. The problem of the nonlinearity in dispersion can be alleviated by the use of composite materials with opposite signs of dispersion.

Multiobject spectroscopy (MOS)

As discussed above, there is great value in increasing the multiplex gain. This can be done in two ways:

(i) *Increasing the number of apertures in the slit mask (Multislits).* This requires that each slit is carefully cut in the mask because, although the telescope attitude (translation and rotation) may be adjusted to suit one aperture, errors in the relative positions of slits cannot be compensated since the image scale is fixed. Each aperture produces a spectrum on the detector and the mask designer must ensure that the targets and their matching slits are chosen to avoid overlaps between spectra and orders. This has the effect of limiting the surface density of targets which can be addressed in one observation. Passband filters can be used to limit the wavelength range and reduce the overlap problem. This option requires superior optics able to accommodate both a wide FOV and large disperser.

(ii) *Using optical fibers (Multifibers).* A number of fibers may be deployed at the telescope focus to

direct light from the various targets to a remote spectrograph whose input focus consists of one or more continuous pseudoslits. These are made up of the fiber outputs arranged in a line. The continuous nature of the slit means that spectral overlap is avoided without restricting the surface density of addressable targets; although there is a limit imposed by the distance of closest approach of the fibers. The method of deploying the fiber inputs may be a plugplate consisting of pre-cut holes into which encapsulated fibers are manually plugged, or a pick-and-place robot which serially positions the fiber inputs at the correct location on a magnetized field plate. The latter is highly versatile but mechanically complex with a significant configuration time that may erode the actual on-target integration time.

The two systems have contrasting capabilities (Figure 7). The multislit approach provides generally better SNR since the light feeds into the spectrograph directly, but is compromised in terms of addressable surface density of targets, by the need to avoid spectral overlap and in FOV by the difficulty of the wide-field optics. A current example of this type of instrument, GMOS, is shown in Figure 8 and described in Table 5. The multifiber approach is limited in SNR by modal noise in the fibers and attendant calibration uncertainties and the lack of contiguous estimates of the sky background. However, it is easier to adapt to large fields since the spectrograph field can be much smaller than the field over which the fiber inputs are distributed. In summary, multislit systems are best for narrow-but-deep surveys while the multifiber systems excels at wide-but-shallow surveys. Fiber-fed instruments may further prove their worth in ELTs where technical problems may require that these bulky instruments are mounted off the telescope (see below for further discussion).

Of paramount importance in MOS is the quality of the background subtraction as discussed above. Traditionally, this requires slits which sample the sky background directly adjacent to the object. An alternative is nod-and-shuffle (va-et-vient) in which nearby blank regions are observed alternately with the main field using the same slit mask by moving the telescope (the nod). In the case of CCDs, the photo-generated charge from the interleaved exposures is stored temporarily on the detector by moving it to an unilluminated portion of the device (the shuffle). After many repeats on a timescale less than that of variations in the sky background (a few minutes), the accumulated charge is read out, incurring the read-noise penalty only once. Although requiring an effective doubling of the exposure time and an

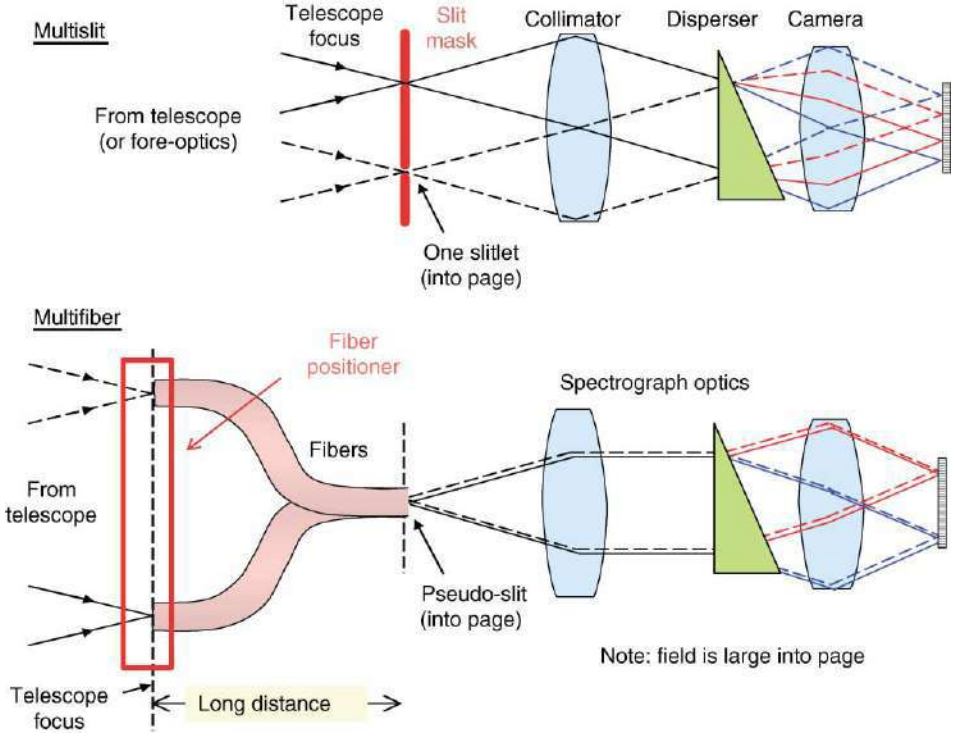

Figure 7 Illustration of the difference between the multislit and multifiber approaches to multi-object spectroscopy.

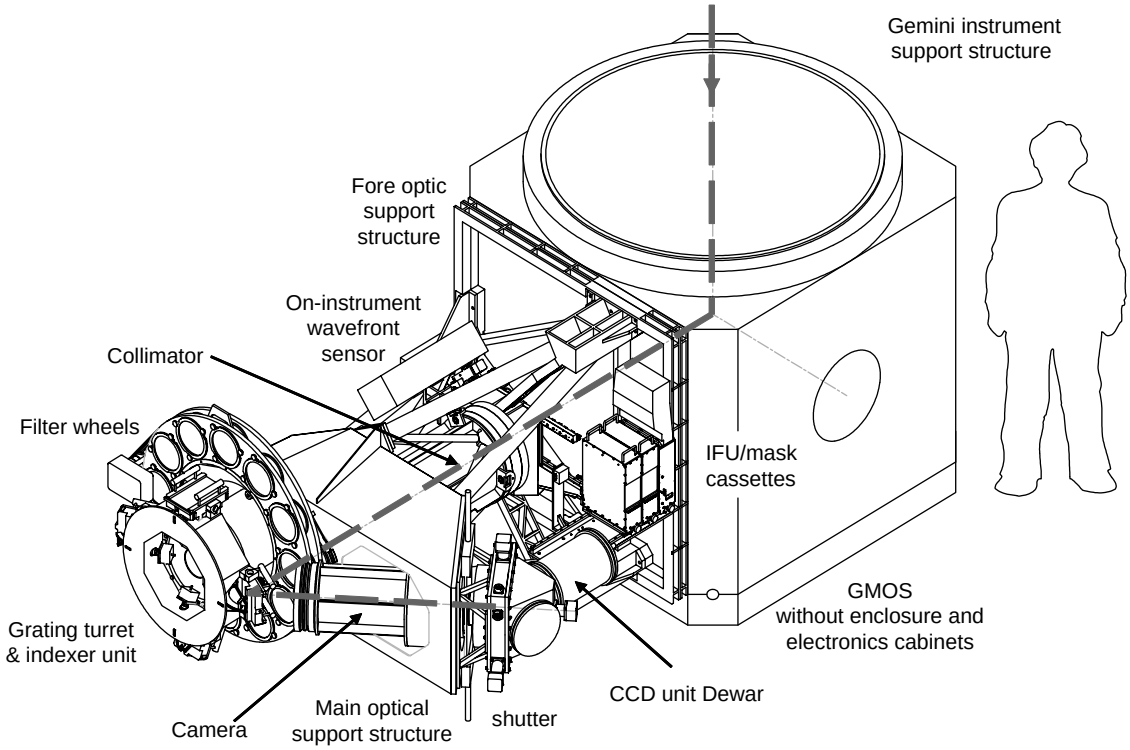

Figure 8 The Gemini Multiobject Spectrograph (GMOS). One of two built for the two Gemini 8-m telescopes by a UK–Canadian consortium. It includes an integral field capability provided by a fiber-lenslet module built by the University of Durham. The optical configuration is the same as shown in **Figure 2**. It is shown mounted on an instrument support unit which includes a mirror to direct light from the telescope into different instruments. The light path is shown by the red dashed line. The slit masks are cut by a Nd-YAG laser in 3-ply carbon fiber sheets. See **Table 5** for the specification.

Table 5 Main characteristics of the Gemini multiobject spectrographs

Image scale	72 mas/pixel	FOV	$5.5' \times 5.5'$
Detector	CCD with 13.5 μm pixels Format: $3 \times (4608 \times 2048)$	Slit/mask configuration	$\leq$ few $\times$ 100 slits of width $\geq 0.2''$
Wavelength range	Total: 0.4–1.0 μm Simultaneous: $\leq$ 1 octave	Integral field unit	$1500 \times 0.2''$ samples in 50 sq.$''$
Spectral resolving power	$R \leq 5000$ with $0.5''$ slit	Dispersion options	3 gratings + mirror for imaging

increase in the size of the detector, this technique allows the length of the slits to be reduced since no contiguous sky sample is required, thereby greatly increasing the attainable multiplex gain if the number density of potential targets is sufficiently high.

Integral field spectroscopy (IFS)

IFS provides a spectrum of each spatial sample simultaneously within a contiguous FOV. Other approaches provide the same combination of imaging and spectroscopy but require a series of nonsimultaneous observations. Examples include imaging through a number of narrow passband filters with different effective wavelength and spectroscopy with a single slit which is stepped across the object. Other such techniques are Fabry–Perot, Fourier-transform and Hadamard spectroscopy. Since the temporal variation of the sky background is a major limitation in astronomy of faint objects, IFS is the preferred technique for terrestrial observations of faint objects, but nonsimultaneous techniques are preferable in certain niche areas, and relatively more important in space where the sky background is reduced.

The main techniques (**Figure 9**) are as follows:

(i) *Lenslets*. The field is subdivided by placing an array of lenslets at the telescope focus (or its conjugate following re-imaging fore-optics) to form a corresponding array of micropupils. These are re-imaged on the detector by a focal

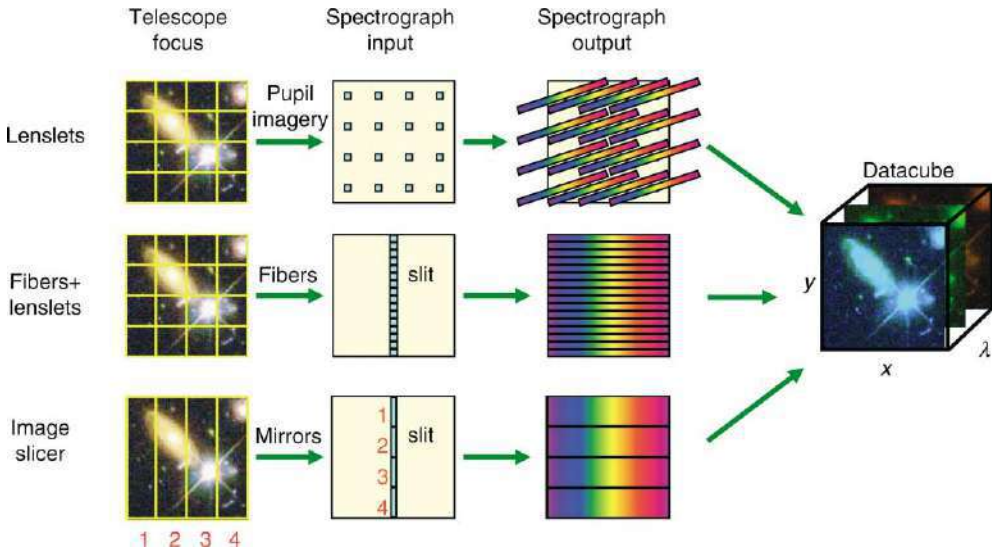

Figure 9 Main techniques of integral field spectroscopy.

reducer using a conventional disperser. The spectra are, therefore, arranged in the same pattern as that of the lenslet array. Spectral overlap is reduced by angling the dispersion direction away from the symmetry axes of the lenslet array and by the fact that the pupil images are smaller than the aperture of the corresponding lenslet.

(ii) *Fibers + lenslets*. The field is subdivided as in (i) but the pupil images are relayed to a remote spectrograph using optical fibers which reformat the field into linear pseudoslits. This avoids the problem of overlap, allowing a greater length of spectrum than in (i), but is more complex since the arrays of fibers and lenslets must be precisely co-aligned and are subject to modal noise in the fibers.

(iii) *Image slicer*. The field is subdivided in only one dimension by a stack of thin slicing mirrors placed at a conjugate of the telescope focus (Figure 10). Each mirror is angled so as to direct light to its own pupil mirror which re-images the field to form part of a linear pseudoslit. Thus the slices into which the image is divided are rearranged end-to-end to form a continuous slit at the spectrograph input. Unlike the other techniques, this retains spatial information along the length of each slice. This is dispersed by conventional means via a focal-reducer. An additional optic is often required at the image of each slice at the slit to reimage the micropupil images produced by the slicing mirrors onto a common pupil inside the spectrograph. In principle this is the most efficient of the three methods since the one-dimensional slicing produces fewer diffraction losses in the spectrograph than the two-dimensional division used by the others, and minimizes the fraction of the detector surface which must be left unilluminated in order to avoid cross-talk between noncontiguous parts of the field. However the micromirrors are difficult to make since they require diamond-turning or grinding in metal or glass with a very fine surface finish (typically with RMS ~ 1 nm for the optical and ~ 10 nm in the infrared). This sort of system is, however, well-matched to cold temperatures, since the optical surfaces and mounts may be fabricated from the same material (e.g., Al) or from materials with similar thermal properties.

The data are processed into a datacube whose dimensions are given by the two spatial coordinates, x, y, plus wavelength. The datacube may be sliced up in ways analogous to tomography to understand the physical process operating in the object.

Spectropolarimetry and Polarimetry

Spectropolarimetry and polarimetry are analogous to spectroscopy and imaging, where the polarization state of the light is measured instead of the total intensity. Like IFS, spectropolarimetry is a photon-starved (i.e., limited in SNR by photon noise) area which benefits from the large aperture of current and projected telescopes.

The Stokes parameters of interest are I, Q, and U. V is generally very small for astronomical objects.

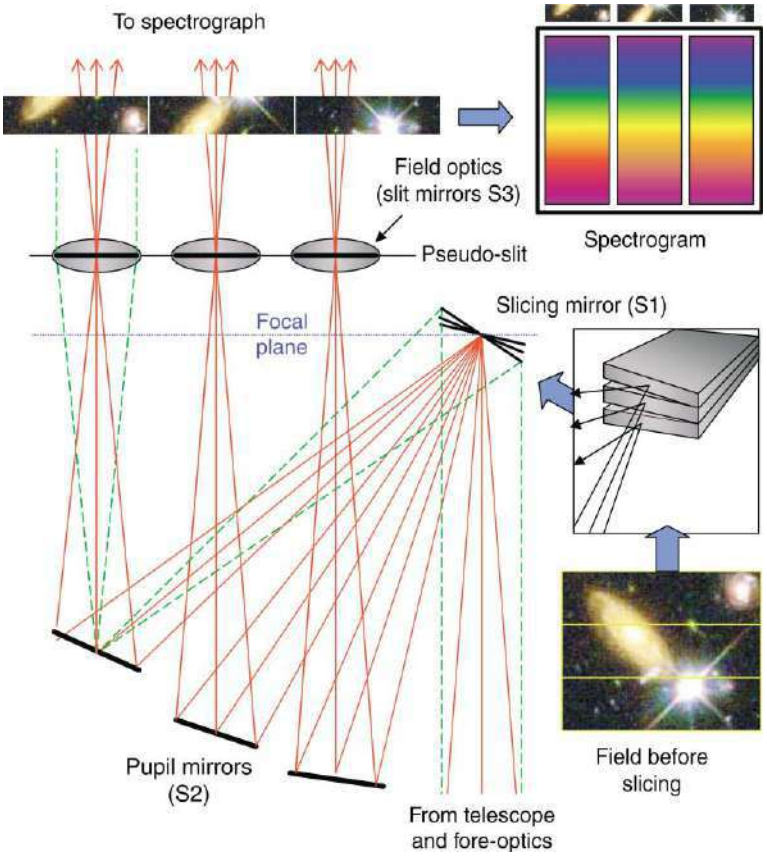

Figure 10 Principle of the advanced image slicer. Only three slices are shown here. Real devices have many more, e.g., the IFU for the *Gemini Near-Infrared Spectrograph* has 21 to provide a $5'' \times 3''$ field with $0.15'' \times 0.15''$ sampling. Reproduced from Content R (1997) A new design for integral field spectroscopy with 8 m telescopes. *Proceedings SPIE* 2871: 1295–1305.

From these, the degree and azimuthal angle of linear polarization are obtained as:

$$\pi_{\mathrm{L}} = \left(\frac{Q^2 + U^2}{I^2}\right) \qquad \psi = \frac{1}{2}\arctan\left(\frac{U}{Q}\right) \qquad [26]$$

The Stokes parameters may be estimated using a modified spectrograph with a rotatable achromatic half-wave retarder, characterized by the angle of rotation about the optical axis, θ, placed before the instrument and a polarizing beamsplitter which separates the incident beam into orthogonal polarization states (Figure 11). The two states are recorded simultaneously on different regions of the detector. The separation is achieved usually through the angular divergence produced by a Wollaston prism placed before the disperser or through the linear offset produced by a calcite block placed before the slit. The intensity as a function of the waveplate rotation angle, $S(\theta)$, is recorded for each wavelength and position in the spectrum or image. The Stokes parameters for each sample are then given by

$$\begin{aligned}
\frac{Q}{I} &= \frac{S(0) - S(\pi/2)}{S(0) + S(\pi/2)} \\
&= \left\langle \frac{A_1 - \omega_{AB}B_1}{A_1 + \omega_{AB}B_1}, \frac{A_2 - \omega_{AB}B_2}{A_2 + \omega_{AB}B_2} \right\rangle \\
\frac{U}{I} &= \frac{S(\pi/4) - S(3\pi/4)}{S(\pi/4) + S(3\pi/4)} \\
&= \left\langle \frac{C_1 - \omega_{CD}D_1}{C_1 + \omega_{CD}D_1}, \frac{C_2 - \omega_{CD}D_2}{C_2 + \omega_{CD}D_2} \right\rangle \\
I &= S(0) + S(\pi/2) \\
&= \left\langle \frac{A_1 + \omega_{AB}B_1}{\eta_{A1}}, \frac{A_2 + \omega_{AB}B_2}{\eta_{A2}} \right\rangle \text{ etc.}
\end{aligned} \qquad [27]$$

where A and B are the signals (functions of position and/or λ) recorded with the waveplate at $\theta = 0$ and $\theta = \pi/2$ and the subscripts 1 and 2 refer to the two polarization states produced by the beamsplitter.

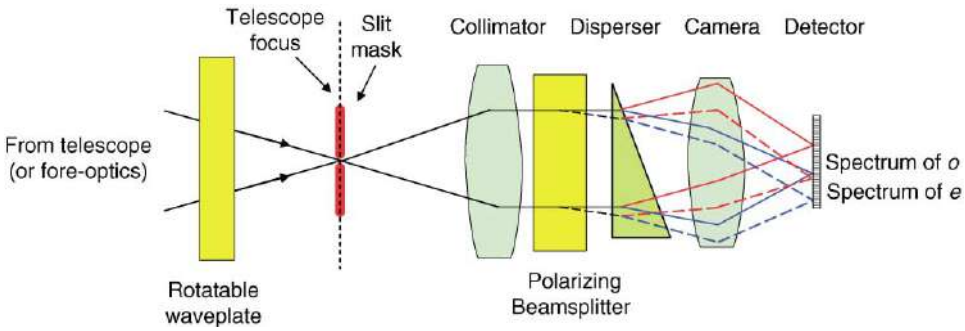

Figure 11 Schematic layout of spectropolarimeter using a polarizing beamsplitter in the collimated beam (e.g., a Wollaston prism producing *e*- and *o*-states) to produce separate spectra for orthogonal polarization states on the detector. Alternative configurations in which the two polarizations have a linear offset (e.g., a calcite block) may be used in which case the beamsplitter is placed before the slit.

Likewise, C and D are a pair of observations taken with $\theta = \pi/4$ and $\theta = 3\pi/4$. The angled brackets indicate a suitable method of averaging the quantities inside. The factors ω_{AB} and ω_{CD} are estimated as

$$\omega_{AB} = \sqrt{\frac{A_1 A_2}{B_1 B_2}} \qquad \omega_{CD} = \sqrt{\frac{C_1 C_2}{D_1 D_2}} \qquad [28]$$

and η_{A1} is a single calibration factor for polarization state 1 of observation A, etc. One problem is instrumental polarization caused by oblique reflections from telescope mirrors upstream of the polarizing waveplate, such as the fold mirror external to GMOS (see Figure 8). This must be carefully accounted for by taking observations of stars with accurately known polarization. The dispersing element is also likely to be very strongly dependent on the polarization state of the incident light with widely different efficiency as a function of wavelength. Although this effect is cancelled out using the procedure described above, it introduces a heavy toll in terms of SNR reduction.

Technology Issues

Use of Optical Fibers in Astronomy

As discussed above, optical fibers are commonly used in astronomy for coupling spectrographs to wide, sparsely sampled FOVs in the case of MOS, or small contiguously sampled FOVs in the case of IFS. For both applications, the most important characteristics are:

(i) *Throughput*. Near-perfect transmission is required for wavelength intervals of ~1 octave within the overall range of 0.3–5 μm. For MOS, a prime focus fiber feed to an off-telescope spectrograph implies fiber runs of several tens of meters for 8 m telescopes, increasing proportionally for ELTs. For IFS, runs of only ~1 m are required for self-contained par-focal IFUs. For $\lambda < 2.5$ μm, fused silica is a suitable material, although it is not currently possible to go below ~0.35 μm, depending on fiber length. Newer techniques and materials may improve the situation. For $\lambda > 2.5$ μm, alternative materials are required, such as ZrF4 and, at longer wavelengths, chalcogenide glasses, but these are much more expensive and difficult to use than fused silica.

(ii) *Conservation of Etendue*. Fibers suffer from focal ratio degradation, a form of modal diffusion, which results in a speeding up of the output beam with respect to the input. Viewed as an increase in entropy, it results in a loss of information. In practice, this implies either reduced throughput, as the fiber output is vignetted by fixed-size spectrograph optics, or a loss in spectral resolution if the spectrograph is oversized to account for the faster beam from the fibers. This effect may be severe at the slow focal ratios pr oduced by many telescopes ($F > 10$) but is relatively small for fast beams ($F < 5$). Thus, its effect may be mitigated by speeding up the input beam by attaching lenslets directly to the fibers, either individually (for MOS) or in a close-packed array (for IFS). If the spectrograph operates in both fiber- and beam-fed modes (e.g., Figure 8), it is also necessary to use lenslets at the slit to convert the beam back to the original speed.

(iii) *Efficient coupling to the telescope*. This requires that the fibers are relatively thick to match the physical size of the sampling at the telescope focus which is typically $\chi_s = 0.1''$ for IFS and $\chi_s = 2''$ for MOS. By conservation of Etendue, the physical size of the fiber aperture is

$$d_f = \chi_s D_T F_f \qquad [29]$$

where F_f is the focal ratio of the light entering the fiber and D_T is the diameter of the telescope aperture. Using the above values with $F_f = 4$, as required to maintain Etendue, implies $15 < d_f < 300$ μm for an 8 m or $60 < d_f < 1200$ μm for a 30 m telescope. Except at the lower end, this implies the use of multimode (step-index) fibers. However, the need in IFS to oversize the fibers to account for manufacturing errors and current requirements of $\chi_s \geq 0.2''$ to produce sufficient SNR in a single spatial sample gives a practical limit of $d_f > 50$ μm which is in the multimode regime.

Recently, the use of photonic crystal fibers in astronomy has been investigated. These, together with other improvements in material technology may improve the performance of optical fibers, which will be of special relevance to ELTs where, for example, there will be a need to couple very fast beams into fibers.

Cooling for the Near-Infrared

As shown in Figure 12, it is necessary to cool instruments for use in the infrared. This is to prevent thermal emission from the instrument becoming the dominant noise source. The minimum requirement is to place a cold stop at a conjugate of the telescope pupil. However, much of the rest of the instrument requires cooling because of the nonzero emissivity of the optics. The telescope also requires careful design. For example, the Gemini telescopes are optimized for the NIR by undersizing M2 so that the only stray light entering the final focal surface is from the cold night sky.

Irrespective of the thermal background and wavelength, all detectors (see Table 4), require cooling to reduce internally generated noise to acceptable levels. For the CCDs used in the optical, the cryogen is liquid nitrogen, but for the infrared, cooling with liquid helium is required.

Structure

As discussed above, the necessity for instruments to scale in size with the telescope aperture cannot be achieved by simply devising a rigid structure, since the required materials are generally not affordable (instruments on 8 m telescopes cost roughly €/$ 5–15 M, including labor). Furthermore, the stability also depends critically on the mounting of the individual optical elements, where the choice of material is constrained by the need for compliance with the optical materials. The solution adopted for GMOS was to design the collimator and camera mounting so that variations in the gravity vector induce a translation in each component orthogonal to the optical axis without inducing tilts or defocus. Therefore, the only effect of instability is a translation of the image of the slit on the detector. Since the slit area is rigidly attached to the telescope interface, the image of a celestial object does not move with respect

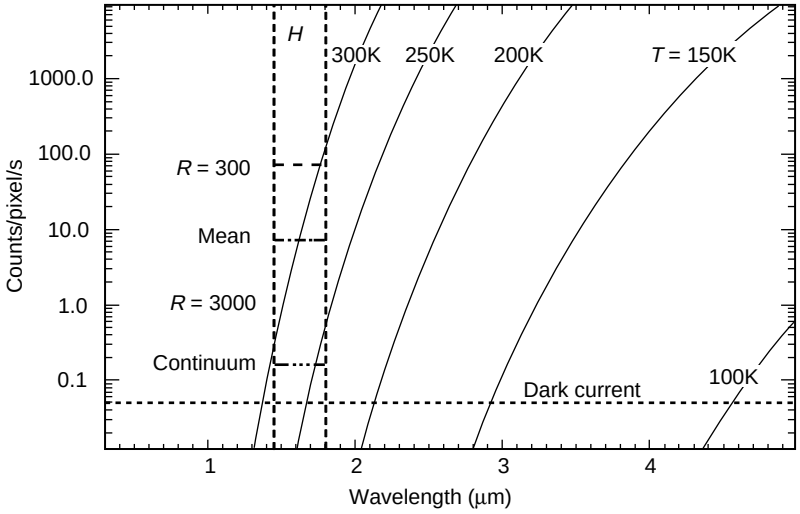

Figure 12 Illustration of the need for cooling in NIR instruments. This is modeled for the case of a spectrograph on an 8 m telescope with $0.2'' \times 0.2''$ sampling with both throughput and total emissivity of 50%. The curves indicate the cumulative thermal background for wavelengths shortward of that shown on the scale. For comparison, a typical dark current is shown for NIR detectors (see Table 4). Also shown is the background from the night sky for the case of the *H*-band, for two values of spectral resolution, *R*. The mean sky background level is shown for both values of *R*. For $R = 3000$, the continuum level found between the strong, narrow OH-lines which make up most of the signal is also shown. To observe in the *H*-band, cooling to $-40\,^{\circ}$C is required for high spectral resolution, but is unnecessary at low resolution. At longer wavelengths, fully cryogenic cooling with liquid nitrogen is generally required.

to the slit providing that the telescope correctly tracks the target. The motion of the image of the slit on the detector is then corrected by moving the detector orthogonal to the optical axis to compensate. Taking account of nonrepeatable motion in the stucture and the detector mounting, it proved possible to attain the desired stability in open-loop by measuring the flexure experienced with the control system turned off, making a look-up table to predict the required detector position for each attitude setting (elevation and attitude) and applying the correction by counting pulses sent to the stepper motors which control the detector translation.

Some instruments require much greater stability than GMOS does. The flexure-control outlined above may be augmented by operation in closed-loop so that nonrepeatable movements may be accounted for. However, this ideally requires a reference light source which illuminates the optics in the same way as the light from the astronomical target and which is recorded by the science detector without adversely impacting the observation. Alternatively, partial feedback may be supplied by repeated metrology of key optical components such as fold-mirrors.

One strategy to measure radial velocities to very small uncertainities (a few meters per second) is to introduce a material (e.g., iodine) into the optical path, which is present throughout the observations. This absorbs light from the observed object's continuum at one very well-defined wavelength. Thus, instrument instability can be removed by measuring the centroid of the desired spectral line in the object relative to the fixed reference produced by the absorption cell. However, flexure must still be carefully controlled to avoid blurring the line during the course of a single exposure.

Mounting of instruments via a fiber-feed remotely from the telescope, at a location where they are not subjected to variations in the gravity vector, is a solution for applications where great stability is required. But, even here, care must be taken with the fore-optics and pickoff system attached to the telescope, and to account for modal noise induced by changes in the fiber configuration as the telescope tracks.

Finally, the instrument structure has other functions. For cooled instruments, a cryostat is required inside which most of the optical components are mounted. For uncooled instruments, careful control of temperature is needed to avoid instrument motion due to thermal expansion. This requires the use of an enclosure which not only blocks out extraneous light, but provides a thermal buffer against changes in the ambient temperature during observations and reduces thermal gradients by forced circulation of air.

See also

Diffraction: Diffraction Gratings. **Fiber and Guided Wave Optics**: Fabrication of Optical Fiber. **Imaging**: Adaptive Optics. **Instrumentation**: Telescopes; Spectrometers. **Spectroscopy**: Fourier Transform Spectroscopy; Hadamard Spectroscopy and Imaging.

Further Reading

Allington-Smith JR, Murray G, Content C, *et al.* (2002) Integral field spectroscopy with the Gemini Multiobject Spectrograph. *Publications of the Astronomical Society of the Pacific* 114: 892.

Bjarklev A, Broeng J and Sanchez Bjarklev AS (2003) *Photonic Crystal Fibers*. Kluwer.

Bowen IS (1938) The Image-Slicer: a device for reducing loss of light at slit of stellar spectrograph. *Astrophysical Journal* 88: 113.

Carrasco E and Parry I (1994) A method for determining the focal ratio degradation of optical fibres for astronomy. *Monthly Notices of the Royal Astronomical Society* 271: 1.

Courtes G (1982) *Instrumentation for Astronomy with Large Optical Telescope, IAU Colloq, 67*. In: Humphries CM (ed.) Astrophysics and Space Science Library, vol. 92, pp. 123. Dordrecht: Reidel.

Glazebrook K and Bland-Hawthorn J (2001) Microslit nod-shuffle spectrocopy: a technique for achieving very high densities of spectra. *Publications of the Astronomical Society of the Pacific* 113: 197.

Lee D, Haynes R, Ren D and Allington-Smith JR (2001) Characterisation of microlens arrays for astronomical spectroscopy. *PASP* 113: 1406.

McLean I (1997) *Electronic Imaging in Astronomy*. Chichester, UK: John Wiley and Sons Ltd.

Palmer C (2000) *Diffraction Grating Handbook*. Richardson Grating Laboratory, 4th edn. Rochester, NY.

Ridgway SH and Brault JW (1984) Astronomical Fourier transform spectroscopy revisited. *Ann. Rev. Astron. Astrophys.* 22.

Saha SK (2002) Modern optical astronomy: technology and impact of interferometry. *Review of Modern Physics* 74: 551.

Schroeder D (2000) *Astronomical Optics*, 2nd edn. San Diego, CA: Academic Press.

van Breugel W and Bland-Hawthorn J (eds) *Imaging the Universe in Three Dimensions*, ASP Conference Series, vol. 195.

Weitzel L, Krabbe A, Kroker H, *et al.* (1996) 3D: The next generation near-infrared imaging spectrometer. *Astron. & Astrophys. Suppl.* 119: 531.

Ellipsometry

J N Hilfiker, J A Woollam & Co., Inc., Lincoln, NE, USA

J A Woollam, University of Nebraska, Lincoln, NE, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Ellipsometry measures a change in polarization as light reflects from or transmits through a material structure. The polarization-change is represented as an amplitude ratio, Ψ, and a phase difference, Δ. The measured response is dependent on optical properties and thickness of each material. Thus, ellipsometry is primarily used to determine film thickness and optical constants. However, it is also applied to the characterization of composition, crystallinity, roughness, doping concentration, and other material properties associated with a change in optical response.

Interest in ellipsometry has grown steadily since the 1960s as it provided the sensitivity necessary to measure nanometer-scale layers used in microelectronics. Today, the range of applications has spread to basic research in physical sciences, semiconductor, data storage, flat panel display, communication, biosensor, and optical coating industries. This widespread use is due to increased dependence on thin films in many areas and the flexibility of ellipsometry to measure most material types: dielectrics, semiconductors, metals, superconductors, organics, biological coatings, and composites of materials.

This article provides a fundamental description of ellipsometry measurements along with the typical data analysis procedures. The primary applications of ellipsometry are also surveyed.

Light and Materials

Ellipsometry measurements involve the interaction between light and material.

Light

Light can be described as an electromagnetic wave traveling through space. For ellipsometry, it is adequate to discuss the electric field behavior in space and time, also known as polarization. The electric field of a wave is always orthogonal to the propagation direction. Therefore, a wave traveling along the z-direction can be described by its x- and y-components. If the light has completely random orientation and phase, it is considered unpolarized. For ellipsometry, we are interested in the case where the electric field follows a specific path and traces out a distinct shape at any point. This is known as polarized light. When two orthogonal light waves are in-phase, the resulting light will be linearly polarized (Figure 1a). The relative amplitudes determine the resulting orientation. If the orthogonal waves are 90° out-of-phase and equal in amplitude, the resultant light is circularly polarized (Figure 1b). The most general polarization is 'elliptical', which combines orthogonal waves of arbitrary amplitude and phase (Figure 1c). This is how ellipsometry gets its name.

Materials

Two values are used to describe optical properties, which determine how light interacts with a material.

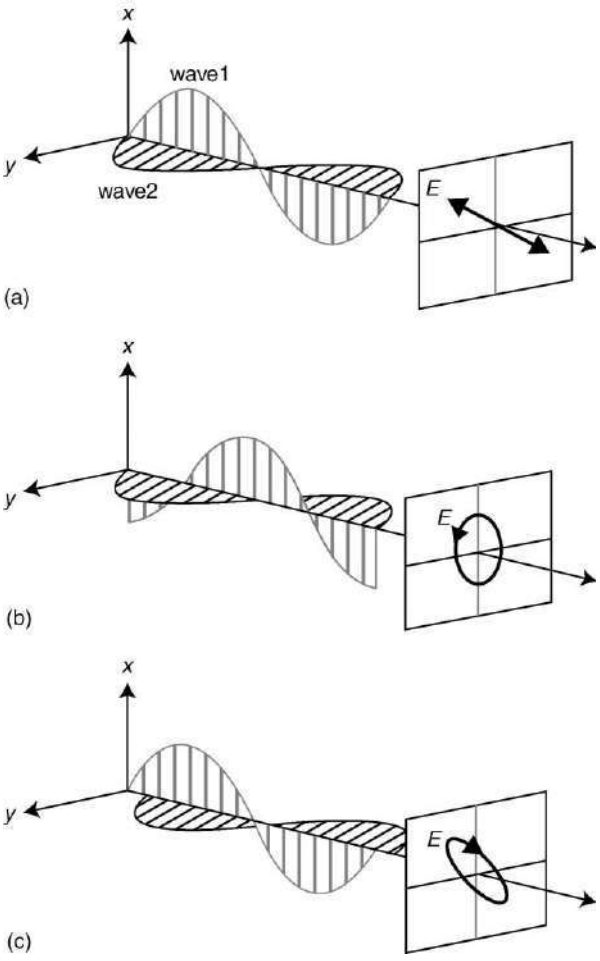

Figure 1 Orthogonal waves combine to demonstrate polarization: (a) linear; (b) circular; and (c) elliptical.

They are generally represented as a complex number. The complex refractive index ($\tilde{n}$) consists of the index (n) and extinction coefficient (k):

$$\tilde{n} = n + \mathrm{i}k \qquad [1]$$

Alternatively, the optical properties can be represented as the complex dielectric function:

$$\tilde{\varepsilon} = \varepsilon_1 + \mathrm{i}\varepsilon_2 \qquad [2]$$

with the following relation between conventions:

$$\tilde{\varepsilon} = \tilde{n}^2 \qquad [3]$$

The index describes the phase velocity for light as it travels through a material compared to the speed of light in vacuum, c:

$$v = \frac{c}{n} \qquad [4]$$

Light slows as it enters a material with higher index. Because frequency remains constant, the wavelength will shorten. The extinction coefficient describes the loss of wave energy to the material. It is related to the absorption coefficient, α, as:

$$\alpha = \frac{4\pi k}{\lambda} \qquad [5]$$

Light loses intensity in an absorbing material, according to Beer's Law:

$$I(z) = I(0)\,\mathrm{e}^{-\mathrm{i}\alpha z} \qquad [6]$$

Thus, the extinction coefficient relates how quickly light vanishes in a material. These concepts are demonstrated in Figure 2, where a light wave travels from air into two different materials of varying properties.

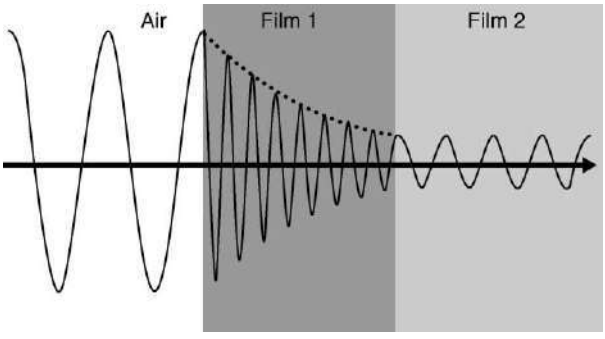

Figure 2 Wave travels from air into absorbing Film 1 and then transparent Film 2. The phase velocity and wavelength change in each material depending on index of refraction (Film 1: $n = 4$, Film 2: $n = 2$).

The optical constants for TiO_2 are shown in Figure 3 from the ultraviolet (UV) to the infrared (IR). The optical constants are wavelength dependent with absorption ($k > 0$) occurring in both UV and IR, due to different mechanisms that take energy from the light wave. IR absorption is commonly caused by molecular vibration, phonon vibration, or free-carriers. UV absorption is generally due to electronic transitions, where light provides energy to excite an electron to an elevated state. A closer look at the optical constants in Figure 3 shows that real and imaginary optical constants are not independent. Their shapes are mathematically coupled together through Kramers–Krönig consistency. Further details are covered later in this article.

Interaction Between Light and Materials

Maxwell's equations must remain satisfied when light interacts with a material, which leads to boundary conditions at the interface. Incident light will reflect and refract at the interface, as shown in Figure 4. The angle between the incident ray and sample normal

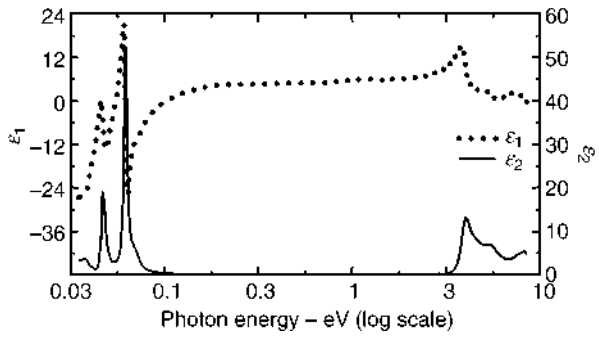

Figure 3 Complex dielectric function for TiO_2 film covering wavelengths from the infrared (small eV) to the ultraviolet (high eV).

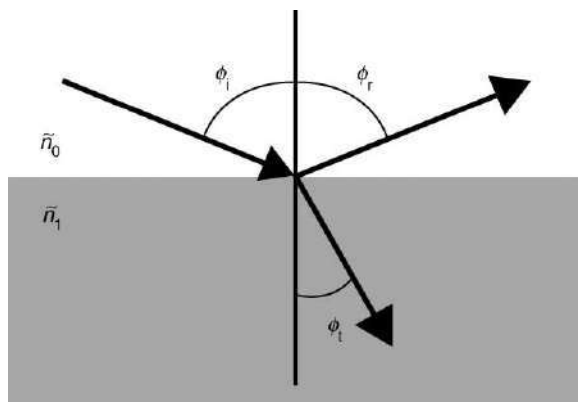

Figure 4 Light reflects and refracts according to Snell's law.

(ϕ_i) will be equal to the reflected angle, (ϕ_r). Light entering the material is refracted to an angle (ϕ_t) given by:

$$n_0 \sin(\phi_i) = n_1 \sin(\phi_t) \quad [7]$$

The same occurs at every interface where a portion reflects and the remainder transmits at the refracted angle. This is illustrated in Figure 5. The boundary conditions provide different solutions for electric fields parallel and perpendicular to the sample surface. Therefore, light can be separated into orthogonal components with relation to the plane of incidence. Electric fields parallel and perpendicular to the plane of incidence are considered p- and s-polarized, respectively. These two components are independent for isotropic materials and can be calculated separately. Fresnel described the amount of light reflected and transmitted at an interface between materials:

$$r_s = \left(\frac{E_{0r}}{E_{0i}}\right)_s = \frac{n_i \cos\theta_i - n_t \cos\theta_t}{n_i \cos\theta_i + n_t \cos\theta_t} \quad [8a]$$

$$r_p = \left(\frac{E_{0r}}{E_{0i}}\right)_p = \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_i \cos\theta_t + n_t \cos\theta_i} \quad [8b]$$

$$t_s = \left(\frac{E_{0t}}{E_{0i}}\right)_s = \frac{2n_i \cos\theta_i}{n_i \cos\theta_i + n_t \cos\theta_t} \quad [8c]$$

$$t_p = \left(\frac{E_{0t}}{E_{0i}}\right)_p = \frac{2n_i \cos\theta_i}{n_i \cos\theta_t + n_t \cos\theta_i} \quad [8d]$$

Thin film and multilayer structures involve multiple interfaces, with Fresnel reflection and transmission coefficients applicable at each. It is important to track the relative phase of each light component to correctly determine the overall reflected or transmitted beam. For this purpose, we define the film phase thickness as:

$$\beta = 2\pi\left(\frac{t_1}{\lambda}\right) n_1 \cos\theta_1 \quad [9]$$

The superposition of multiple light waves introduces interference that is dependent on the relative phase of each light wave. Figure 5 illustrates the combination of light waves in the reflected beam and their corresponding Fresnel calculations.

Ellipsometry Measurements

For ellipsometry, primary interest is measurement of how p- and s-components change upon reflection or transmission relative to each other. In this manner, the reference beam is part of the experiment. A known polarization is reflected or transmitted from the sample and the output polarization is measured. The change in polarization is the ellipsometry measurement, commonly written as:

$$\rho = \tan(\Psi)\, e^{i\Delta} \quad [10]$$

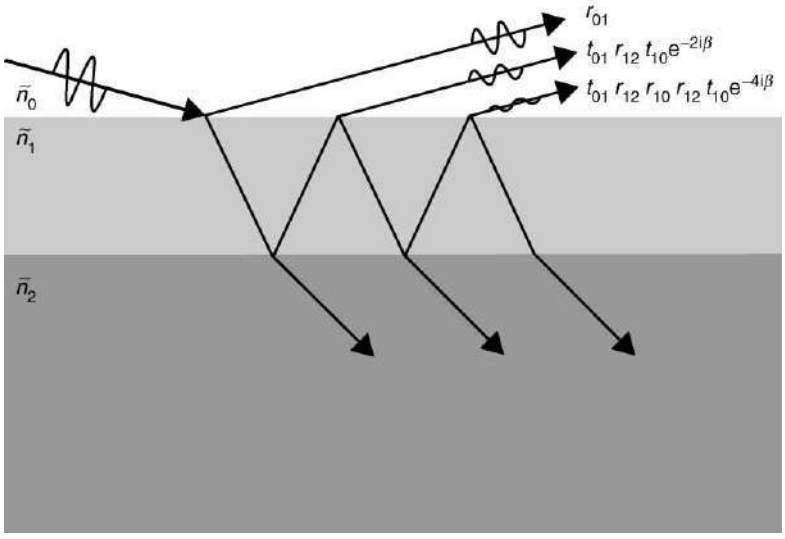

Figure 5 Light reflects and refracts at each interface, which leads to multiple beams in a thin film. Interference between beams depends on relative phase and amplitude of the electric fields. Fresnel reflection and transmission coefficients can be used to calculate the response from each contributing beam.

An example ellipsometry measurement is shown in Figure 6. The incident light is linear with both p- and s-components. The reflected light has undergone amplitude and phase changes for both p- and s-polarized light, and ellipsometry measures their changes.

The primary methods of measuring ellipsometry data all consist of the following components: light source, polarization generator, sample, polarization analyzer, and detector. The polarization generator and analyzer are constructed of optical components that manipulate the polarization: polarizers, compensators, and phase modulators. Common ellipsometer configurations include rotating analyzer (RAE), rotating polarizer (RPE), rotating compensator (RCE), and phase modulation (PME).

The RAE configuration is shown in Figure 7. A light source produces unpolarized light, which is sent through a polarizer. The polarizer passes light of a preferred electric field orientation. The polarizer axis is oriented between the p- and s-planes, such that both arrive at the sample surface. The linearly polarized light reflects from the sample surface, becoming elliptically polarized and travels through a continuously rotating polarizer (referred to as the analyzer). The amount of light allowed to pass will depend on the polarizer orientation relative to the electric field 'ellipse' coming from the sample. The detector converts light from photons to electronic signal to determine the reflected polarization. This information is used with the known input polarization to determine the polarization change caused by the sample reflection. This is the ellipsometry measurement of Ψ and Δ.

Single Wavelength Ellipsometry (SWE)

SWE uses a single frequency of light to probe the sample. This results in a pair of data, Ψ and Δ, used to determine up to two material properties. The optical design can be simple, low-cost, and accurate. Lasers are an ideal light source with well-known wavelength and significant intensity. The optical elements can be optimized to the single wavelength. However, there is relatively low information content (two measurement values). SWE is excellent for quick measurement of

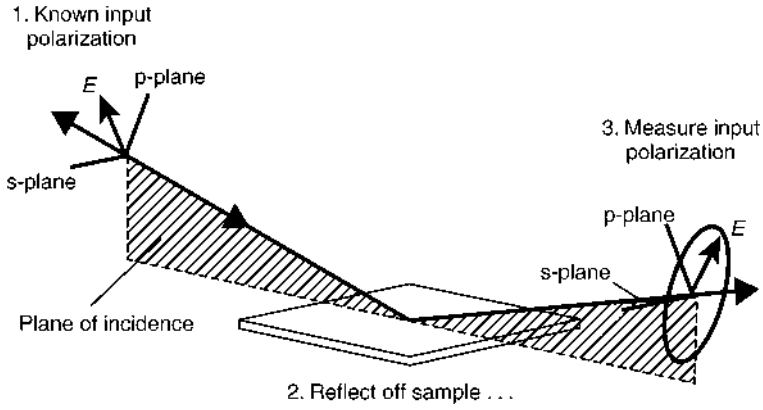

Figure 6 Typical ellipsometry configuration, where linearly polarized light is reflected from the sample surface and the polarization change is measured to determine the sample response.

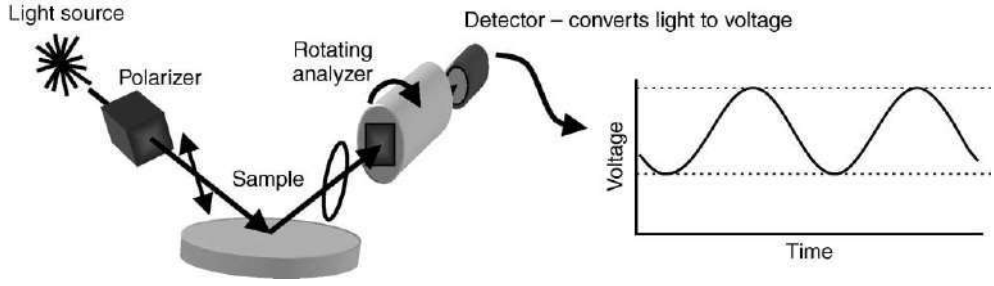

Figure 7 Rotating analyzer ellipsometer configuration uses a polarizer to define the incoming polarization and a rotating polarizer after the sample to analyze the outgoing light. The detector converts light to a voltage with the time-dependence leading to measurement of the reflected polarization.

nominally known films, like SiO_2 on Si. Care must be taken when interpreting unknown films, as multiple solutions occur for different film thickness. The data from a transparent film will cycle through the same values as the thickness increases. From Figure 5, it can be seen that this is related to interference between the multiple light beams. The refracted light may travel through different thicknesses to emerge in-phase with the first reflection, but is the returning light delayed by one wavelength or multiple wavelengths? Mathematically, the thickness cycle can be determined as:

$$D_\phi = \frac{\lambda}{2\sqrt{\tilde{n}_1^2 - \tilde{n}_0^2 \sin^2\phi}} \qquad [11]$$

This is demonstrated in Figure 8, where the complete thickness cycle is shown for SiO_2 on Si at 75° angle of incidence and 500 nm wavelength. A bare silicon substrate would produce data at the position of the star. As the film thickness increases, the data will move around this circle in the direction of the arrow. From eqn [11], the thickness cycle is calculated as 226 nm. Thus the cycle is completed and returns to the star when the film thickness reaches 226 nm. Any data point along the cycle represents a series of thicknesses which depend on how many times the cycle has been completely encircled. For example, the square represents a data point for 50 nm thickness. However, the same data point will represent 276 and 502 nm thicknesses. A second data point at a new wavelength or angle would help determine the correct thickness.

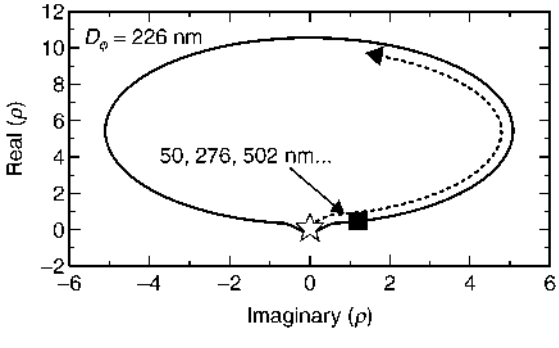

Figure 8 Data from a single wavelength will cycle through the same points as the film thickness increases. The star represents the starting point, with no SiO_2 film on Si. As the thickness increases, the data follow the cycle in the direction of the arrow. After the cycle is complete (thickness = D_ϕ), the data repeat the cycle. Thus, any point along the curve represents multiple possible thicknesses – example shown by a square at 50 nm thickness, which is also equal to 276 and 502 nm.

Spectroscopic Ellipsometry (SE)

Spectroscopic measurements solve the 'period' problem discussed for SWE. Data at surrounding wavelengths insure the correct thickness is determined, even as there remain multiple solutions at one wavelength in the spectrum. This is demonstrated in Figure 9, showing the spectroscopic data for the first three thickness results. The SE data oscillations clearly distinguish each thickness. Thus, a single thickness solution remains to match data at multiple wavelengths.

SE provides additional information content for each new wavelength. While the film thickness will remain constant regardless of wavelength, the optical constants will change across the spectrum. The optical constant shape contains information regarding the material microstructure, composition, and more. Different wavelength regions will provide the best information for each different material property. For this reason, SE systems have been developed to cover very wide spectral regions from the UV to the IR.

Spectroscopic measurements require an additional component, a method of wavelength selection. The most common methods include the use of monochromators, Fourier-transform spectrometers, and gratings or prisms with detection on a diode array. Monochromators are typically slow as they sequentially scan through wavelengths. Spectrometers and diode arrays allow simultaneous measurement at multiple wavelengths. This has become popular with the desire for high-speed SE measurements.

Variable Angle Ellipsometry

Ellipsometry measurements are typically performed at oblique angles, where the largest changes in polarization occur. The Brewster angle, defined as:

$$\tan(\phi_B) = \frac{n_1}{n_0} \qquad [12]$$

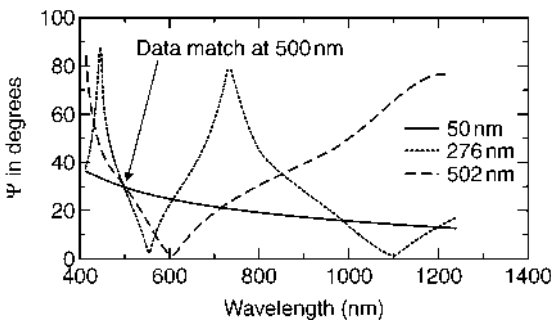

Figure 9 Three spectroscopic ellipsometry measurements that match at 500 nm wavelength, but are easily distinguishable as a function of wavelength.

gives the angle where reflection of p-polarized light goes through a minimum. Ellipsometry measurements are usually near this angle, which can vary from 55° for low-index dielectrics to 75° or 80° for semiconductors and metals.

It is also common to measure at multiple angles of incidence in ellipsometry. This allows additional data to be collected for the same material structure under different conditions. The most important change introduced by varying the angle is the change in path length through the film as the light refracts at a different angle. Multiple angles do not always add new information about a material structure, but the extra data builds confidence in the final answer.

In Situ

Within the last decade, it has become common to employ optical diagnostics during processing. *In situ* ellipsometry allows the optical response to be monitored in real-time. This has also led to feedback control for many processes. While *in situ* is generally restricted to a single angle of incidence, there is a distinct advantage to collecting data at different 'times' during processing to get unique 'glimpses' of the sample structure as it changes.

In situ SE is commonly applied to semiconductors. Conventional SE systems measure from the UV to NIR, which matches the spectral region where semiconductors absorb due to electronic transitions. The shape and position of the absorption can be very sensitive to temperature, composition, crystallinity, and surface quality, which allows SE to closely monitor these properties. *In situ* SE has also been used to monitor numerous material types, including metals and dielectrics. A promising *in situ* application is for multilayer optical coatings, where thickness and index can be determined in real-time to allow correction of future layers, to optimize optical performance.

Data Analysis

Ellipsometry measures changes in polarization. However, it is used to determine the material properties of interest, such as film thickness and optical constants. In the case of a bulk material, the equations derived for a single reflection can be directly inverted to provide the 'pseudo' optical constants from the ellipsometry measurement, ρ:

$$\langle\tilde{\varepsilon}\rangle = \sin^2(\phi)\left[1 + \tan^2(\phi)\left(\frac{1-\rho}{1+\rho}\right)\right] \qquad [13]$$

This equation assumes there are no surface layers of any type. However, there is typically a surface oxide or roughness for any bulk material and the direct inversion would incorporate these into the bulk optical constants. The more common procedure used to deduce material properties from ellipsometry measurements follows the flowchart in Figure 10. Regression analysis is required because an exact equation cannot be written. Often the answer is overdetermined with hundreds of experimental data points for a few unknowns. Regression analysis allows all of the measured data to be included when determining the solution.

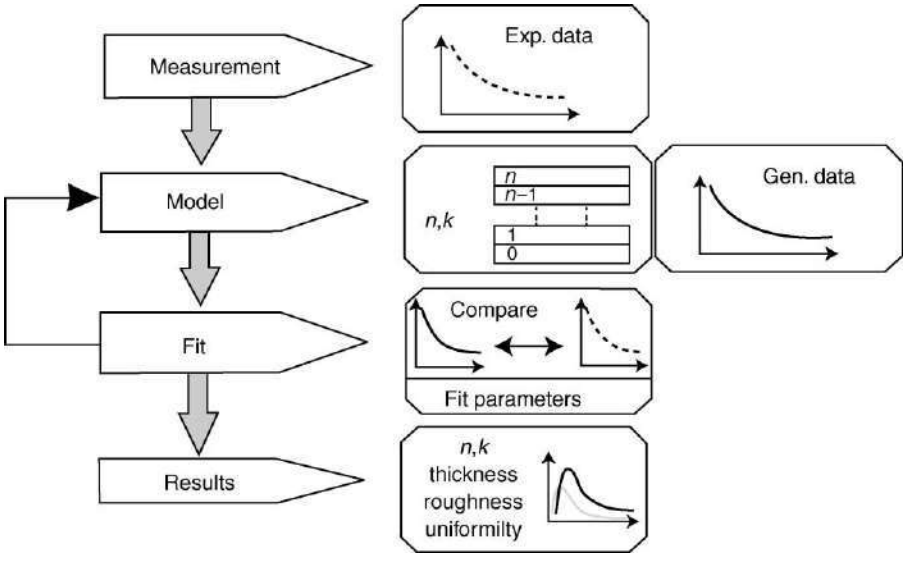

Figure 10 Flowchart for ellipsometry data analysis.

Data analysis proceeds as follows. After the sample is measured, a model is constructed to describe the sample. The model is used to calculate the predicted response from Fresnel's equations. Thus, each material must be described with a thickness and optical constants. If these values are not known, an estimate is given for the purpose of the preliminary calculation. The calculated values are compared to experimental data. Any unknown material properties can then be varied to improve the match between experiment and calculation. The number of unknown properties should not exceed the information content contained in the experimental data. For example, a single-wavelength ellipsometer produces two data points (Ψ,Δ) which allows a maximum of two material properties to be determined. Finding the best match between the model and experiment is typically done through regression. An estimator, like the mean squared error (MSE), is used to quantify the difference between curves. The unknown parameters are allowed to vary until the minimum MSE is reached.

Care must be taken to ensure the best answer is found, corresponding to the lowest MSE. For example, Figure 11a shows the MSE curve versus film thickness for a transparent film on silicon. There are multiple 'local' minima, but the lowest MSE value occurs at a thickness of 749 nm. This corresponds to the correct film thickness. It is possible that the regression algorithm will mistakenly fall into a 'local' minima, depending on the starting thickness and the MSE structural conditions. Comparing the results by eye for the lowest MSE and a local minima easily distinguish the true minima (see Figures 11b and c).

Ellipsometry Characterization

The two most common material properties studied by ellipsometry are film thickness and optical constants. In addition, ellipsometry can characterize material properties that affect the optical constants. Crystallinity, composition, resistivity, temperature, and molecular orientation can all affect the optical constants and in turn be measured by ellipsometry. This section details many of the primary applications important to ellipsometry.

Film Thickness

The film thickness is determined by interference between light reflecting from the surface and light traveling through the film. Depending on the relative phase of the rejoining light to the surface reflection, there can be constructive or destructive interference.

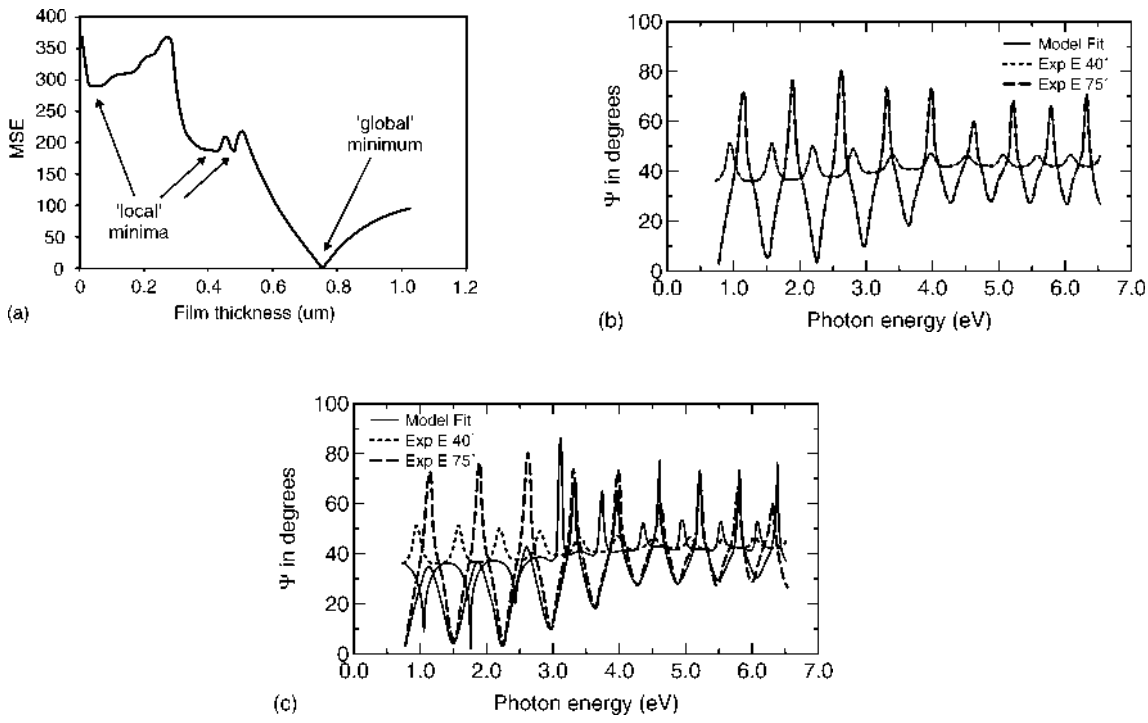

Figure 11 (a) MSE curve versus thickness shows the 'global' minimum where the best match between model and experiment occurs, and 'local' minima that may be found by the regression algorithm, but do not give the final result. (b) The experimental data and corresponding curves generated for the model at the 'global' minimum. (c) Similar curves at the 'local' minimum near 0.45 microns thickness is easily distinguishable as an incorrect result.

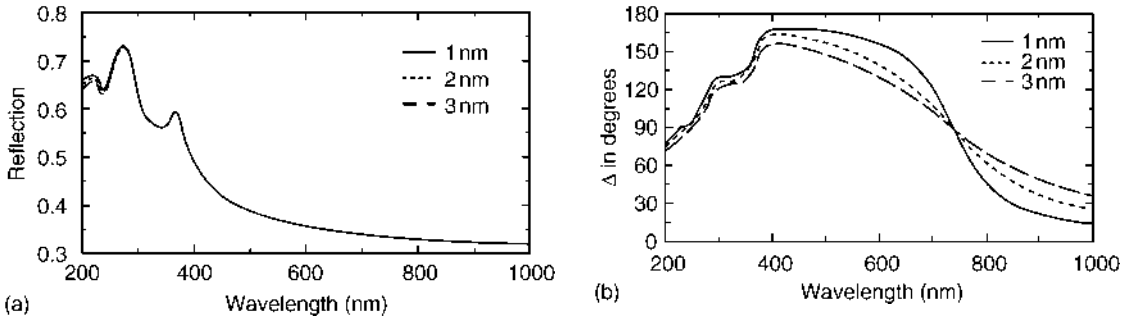

Figure 12 (a) Reflected intensity and (b) ellipsometric delta for three thin oxides on silicon show the high sensitivity of Δ to nanometer-scale films not available from the intensity measurement.

The interference involves both amplitude and phase information. The phase information from Δ is very sensitive to films down to submonolayer thickness. Figure 12 compares reflected intensity and ellipsometry for the same series of thin SiO_2 layers on Si. There are large variations in Δ, while the reflectance for each film is nearly the same.

Ellipsometry is typically used for films with thickness ranging from sub-nanometers to a few microns. As films become greater than several tens of microns thick, it becomes increasingly difficult to resolve the interference oscillations, except with longer infrared wavelengths, and other characterization techniques become preferred.

Thickness measurements also require a portion of the light to travel through the entire film and return to the surface. If the material is absorbing, thickness measurements by optical instruments will be limited to thin, semi-opaque layers. This limitation can be circumvented by measuring in a spectral region where there is lower absorption. For example, an organic film may strongly absorb UV and IR light, but remain transparent at mid-visible wavelengths. For metals, which strongly absorb at all wavelengths, the maximum layer for thickness determination is typically ~100 nm.

Optical Constants

Thickness measurements are not independent of the optical constants. The film thickness affects the path length of light traveling through the film, but the index determines the phase velocity and refracted angle. Thus, both contribute to the delay between surface reflection and light traveling through the film. Both n and k must be known or determined along with the thickness to get the correct results from an optical measurement.

The optical constants for a material will vary for different wavelengths and must be described at all wavelengths probed with the ellipsometer. A table of optical constants can be used to predict the response at each wavelength. However, it is less convenient to adjust unknown optical constants on a wavelength-by-wavelength basis. It is more advantageous to use all wavelengths simultaneously. A dispersion relationship often solves this problem, by describing the optical constant shape versus wavelength. The adjustable parameters of the dispersion relationship allow the overall optical constant shape to match the experimental results. This greatly reduces the number of unknown 'free' parameters compared to fitting individual n, k values at every wavelength.

For transparent materials, the index is often described using the Cauchy or Sellmeier relationship. The Cauchy relationship is typically given as:

$$n(\lambda) = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} \qquad [14]$$

where the three terms are adjusted to match the refractive index for the material. The Sellmeier relationship enforces Kramers–Krönig (KK) consistency, which ensures the optical dispersion retains a realistic shape. The Cauchy is not constrained by KK consistency and can produce unphysical dispersion. The Sellmeier relationship can be written as:

$$\varepsilon_1 = \frac{A\lambda^2\lambda_0^2}{(\lambda^2 - \lambda_0^2)} \qquad [15]$$

Absorbing materials will often have a transparent wavelength region that can be modeled with the Cauchy or Sellmeier relationships. However, the absorbing region must account for both real and imaginary optical constants. Many dispersion relationships use oscillator theory to describe the absorption for materials, which include the Lorentz, Harmonic, and Gaussian oscillators. They all share similar attributes, where the absorption features are described with an amplitude, broadening, and center energy (related to frequency of light).

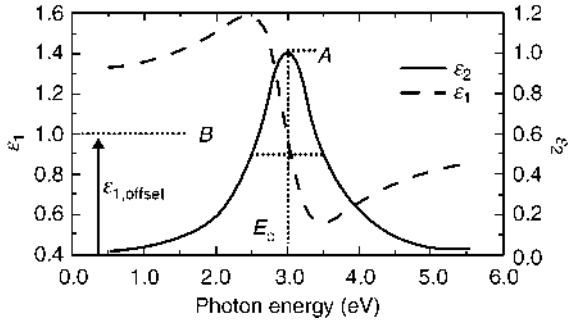

Figure 13 Lorentz oscillator illustrating the primary oscillator parameters: amplitude (A), broadening (B), and center energy (E_c) to describe the imaginary dielectric function shape and an offset ($\varepsilon_{1,\text{offset}}$) to help match the real component after KK transformation has defined its shape.

Kramers–Krönig consistency is used to calculate the shape of the real component after the imaginary behavior is described by the oscillator. An offset to the real component is added to account for extra absorption outside the measured spectral region. The Lorentz oscillator can be written as:

$$\tilde{\varepsilon} = \varepsilon_{1,\text{offset}} + \frac{AE_c}{E_c^2 - E^2 - \mathrm{i}BE} \qquad [16]$$

where the parameters for amplitude (A), broadening (B), center energy (E_c), and offset ($\varepsilon_{1,\text{offset}}$) are also shown in Figure 13 for a typical Lorentz oscillator. The energy, E, is related to the frequency of a wave, ν:

$$E = h\nu \cong \frac{1240}{\lambda_{\text{nm}}} \qquad [17]$$

where h is Planck's constant and the wavelength, λ, is given in nanometers. More advanced dispersion models, like the Tauc–Lorentz and Cody–Lorentz, will include terms to describe the bandgap energy.

Mixing Materials

When two or more materials are mixed on a microscopic level, an effective medium approximation (EMA) may be used to determine the resulting optical constants. In the case of the Bruggemann EMA, which is commonly used, the mixture optical constants ($\tilde{\varepsilon}_{\text{eff}}$) relate to those of the individual materials, as:

$$f_a \frac{\varepsilon_a - \varepsilon_{\text{eff}}}{\varepsilon_a + 2\varepsilon_{\text{eff}}} + f_b \frac{\varepsilon_b - \varepsilon_{\text{eff}}}{\varepsilon_b + 2\varepsilon_{\text{eff}}} = 0 \qquad [18]$$

This can be interpreted as small particles of material A suspended in host material B. The length-scale for mixed particles must satisfy certain electromagnetic equations: typically smaller than one tenth the wavelength of light. In practice, EMA theory is useful for studying very thin surface roughness or interfacial intermixing of materials. These cases are both generally approximated by mixing the two surrounding materials in equal portions. In the case of surface roughness, the second material is void ($n = 1$). EMA theory has also been extended for application to porous materials where the directional dependence of the inclusions is handled mathematically.

Crystallinity

Semiconductors such as Si, are widely used materials, but their properties depend strongly on crystallinity. The UV absorption features in crystalline silicon are broadened and shifted as the material becomes more amorphous. This change in optical properties related to the degree of crystallinity has been used to advantage with ellipsometry measurements to monitor semiconductors and other films. Polysilicon films are used in both the display and semiconductor industry. The degree of crystallinity varies for different process conditions and can be monitored optically to ensure consistent material properties.

Composition

The composition of many alloys will affect the optical constants. The strongest changes often occur in the absorbing region, with shifts in position and amplitude of absorption features. For example, the electronic transitions in $Hg_{1-x}Cd_xTe$ move to higher energy with Cd concentration increase (Figure 14). This material is used for IR detectors that require precise control of composition. Spectroscopic ellipsometry performs this task in real-time with instant feedback to correct the composition during processing. Other applications include AlGaN and InGaN for optoelectronics.

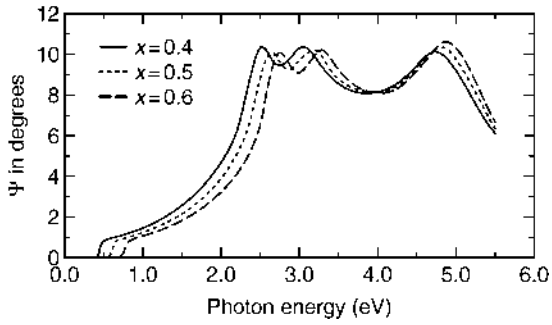

Figure 14 Optical properties for $Hg_{1-x}Cd_xTe$ vary with changes in composition.

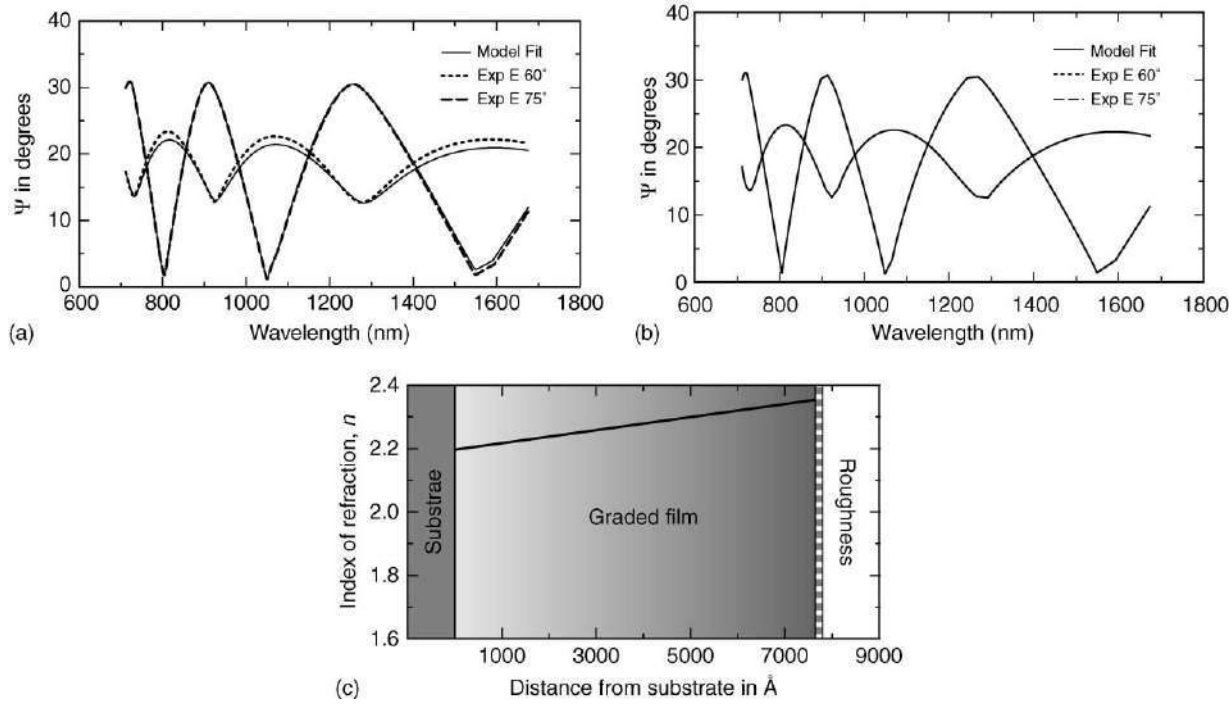

Figure 15 Experimental ellipsometry data and corresponding fit when the model is described as (a) a homogeneous single-layer and (b) a graded film with index variation through the film. (c) The best model to match subtleties in experimental data allows the index to increase toward the surface of the film, with a thin roughness layer on the surface.

Doping Concentration

Dopants in a material will introduce absorption in the infrared due to free-carriers. Spectroscopic ellipsometry measurements at long wavelengths can characterize this optical absorption, thus characterizing the doping concentration. This is common for highly doped semiconductors and transparent conductors such as indium tin oxide (ITO).

Optical Variation (Grading)

Many thin film properties change vertically throughout the film (along the direction perpendicular to the surface). This is most often induced by processing conditions, whether intentional or unintentional. Figure 15a shows the fit to experimental spectroscopic ellipsometry data taken from a single-layer film when it is modeled as a homogeneous layer. To improve the agreement between experimental and model-generated curves, the index was allowed to vary in a linear manner throughout the film. The best fit is shown in Figure 15b, where the model includes the index variation and a thin surface roughness layer. The sample structure is demonstrated in Figure 15c.

Optical Anisotropy

Many materials are optically anisotropic; i.e., their optical properties vary in different film directions. Typical ellipsometry measurements assume no cross-coupling between p- and s-polarizations. This cannot be assumed with anisotropic materials, which has led to 'generalized' ellipsometry measurements. Generalized ellipsometry measures additional information regarding the p- to s- and s- to p- conversion upon reflection. This allows characterization of anisotropic materials, which contain directionally dependent optical constants.

Conclusions

Ellipsometry is a common optical technique for measuring thin films and bulk materials. It relies on the polarization changes due to reflection or transmission from a material structure to deduce material properties, like film thickness and optical constants. This technique continues to develop as the requirement for thin film characterization increases.

See also

Geometrical Optics: Lenses and Mirrors; Prisms. **Optical Coatings:** Thin-Film Optical Coatings. **Optical Materials:** Measurement of Optical Properties of Solids.

Polarization: Introduction; Matrix Analysis. **Semiconductor Physics:** Band Structure and Optical Properties. **Spectroscopy:** Fourier Transform Spectroscopy.

Further Reading

Aspnes DE (1985) The accurate determination of optical properties by ellipsometry. In: Palik ED (ed.) *Handbook of Optical Constants of Solids*, pp. 89–112. Orlando FL: Academic Press.

Azzam RMA and Bashara NM (1987) *Ellipsometry and Polarized Light*. Amsterdam, The Netherlands: Elsevier Science B.V.

Boccara AC, Pickering C and Rivory J (eds) (1993) *Spectroscopic Ellipsometry*. Amsterdam: Elsevier Publishing.

Collins RW, Aspnes DE and Irene EA (eds) (1998) *Proceedings from the Second International Conference on Spectroscopic Ellipsometry*. Thin Solid Films, vols. 313–314.

Gottesfeld S, Kim YT and Redondo A (1995) Recent applications of ellipsometry and spectroellipsometry in electrochemical systems. In: Rubinstein I (ed.) *Physical Electrochemistry: Principles, Methods, and Applications*. New York: Marcel Dekker.

Herman IP (1996) *Optical Diagnostics for Thin Film Processing*, pp. 425–479. San Diego, CA: Academic Press.

Johs B, Woollam JA, Herzinger CM, *et al.* (1999) Overview of variable angle spectroscopic ellipsometry (VASE), Part II: Advanced applications. *Optical Metrology*, vol. CR72, pp. 29–58. Bellingham, Washington: SPIE, Bellingham.

Johs B, Hale J, Ianno NJ, *et al.* (2001) Recent developments in spectroscopic ellipsometry for *in situ* applications. In: Duparré A and Singh B (eds) *Optical Metrology Roadmap for the Semiconductor, Optical, and Data Storage Industries II*, vol. 4449, pp. 41–57. Bellingham, Washington: SPIE.

Roseler A (1990) *Infrared Spectroscopic Ellipsometry*. Berlin: Akademie-Verlag.

Rossow U and Richter W (1996) Spectroscopic ellipsometry. In: Bauer G and Richter W (eds) *Optical Characterization of Epitaxial Semiconductor Layers*, pp. 68–128. Berlin: Springer-Verlag.

Tompkins HG (1993) *A User's Guide to Ellipsometry*. San Diego, CA: Academic Press.

Tompkins HG and Irene EA (eds) (in press) *Handbook of Ellipsometry*. New York: William Andrew Publishing.

Tompkins HG and McGahan WA (1999) *Spectroscopic Ellipsometry and Reflectometry*. New York: John Wiley & Sons, Inc.

Woollam JA (2000) Ellipsometry, variable angle spectroscopic. In: Webster JG (ed.) *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 109–116. New York: John Wiley & Sons.

Woollam JA and Snyder PG (1992) Variable angle spectroscopic ellipsometry. In: Brundle CR, Evans CA and Wilson S (eds) *Encyclopedia of Materials Characterization: Surfaces, Interfaces, Thin Films*, pp. 401–411. Boston, MA: Butterworth-Heinemann.

Woollam JA, Johs B, Herzinger CM, *et al.* (1999) Overview of variable angle spectroscopic ellipsometry (VASE), Part I: Basic theory and typical applications. *Optical Metrology*, vol. CR72, pp. 3–28. Washington: SPIE, Bellingham.

Photometry

J Schanda, University of Veszprém, Veszprém, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Fundamentals of Photometry

Photometry is defined by the International Commission on Illumination (internationally known as CIE from the abbreviation of its French name: Commission Internationale de l'Eclairage) as 'measurement of quantities referring to radiation as evaluated according to a given spectral efficiency function, e.g., $V(\lambda)$ or $V'(\lambda)$.' A note to the above definition states that in many languages it is used in a broader sense, covering the science of optical radiation measurement. We will restrict our treatise to the above fundamental meaning of photometry.

Under spectral efficiency function we understand the spectral luminous efficiency function of the human visual system. The internationally agreed symbols are $V(\lambda)$, for photopic vision (daylight conditions) and $V'(\lambda)$ for scotopic vision (nighttime conditions). CIE standardized the $V(\lambda)$ function in 1924 and the $V'(\lambda)$ function in 1951; their spectral distribution is shown in Figure 1. The $V(\lambda)$ function was determined using mainly the so-called flicker photometric technique, where the light of a reference stimulus and of a test stimulus of varying wavelengths are alternatively shown to the human observer at a frequency at which the observer is already unable to perceive the difference in color (due to the different wavelength of the two radiations), but still perceives a

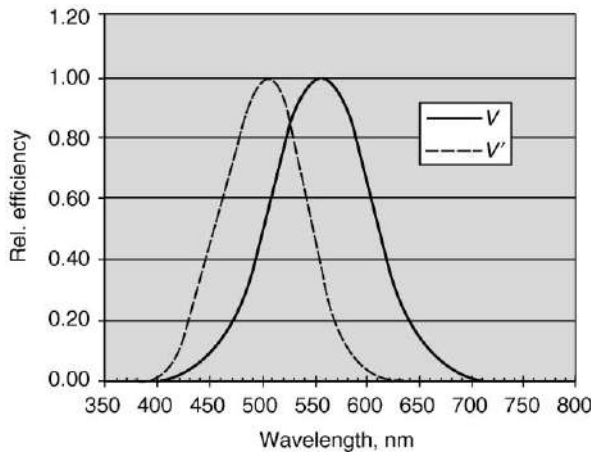

Figure 1 Spectral efficiency functions of the human eye under photopic $V(\lambda)$ and scotopic $V'(\lambda)$ conditions (see CIE The basis of physical photometry. CIE 18.2:1983).

flicker sensation if the luminances of the two stimuli are different. Adjusting the radiance of the test stimulus, we can reach the situation of minimum flicker sensation. In this case we state that we set the two stimuli to equal luminance. Luminance is the quantity that the human observer perceives as brightness (in the case of near white stimuli, see below); radiance is the physical counterpart measured in $\mathrm{W\,m^{-2}\,sr^{-1}}$. For the $V'(\lambda)$ function one can just project the two stimuli side by side, as under scotopic adaptation we cannot distinguish colors, only brightness differences are perceived, thus one can set equal scotopic luminance for the two stimuli by adjusting the brightness of the test stimulus until it agrees with that of the reference stimulus.

Illuminating engineering was mainly interested in describing the effect of near white light, and therefore a psychophysical correlation was selected that described the perceived brightness sensation reasonably well. For near white stimuli, additivity holds, i.e., if A, B, C, and D are four stimuli, and the A stimulus matches the B stimulus and the C stimulus matches the D stimulus, then the superposition of the A and C stimuli matches the superposition of the B and D stimuli. The 'match' word is used here to describe stimuli that produce the same perception. It could be shown that for these stimuli, if the radiance is weighted with the $V(\lambda)$ function, the constructed luminances will be equal, i.e:

if $\quad L_A = L_B$ and $L_C = L_D$

then $\quad L_A + L_C = L_B + L_D$, or $L_A + L_D = L_B + L_C$

where

$$L_x = K_m \int_{380\ \mathrm{nm}}^{780\ \mathrm{nm}} S_{\lambda x} V(\lambda) \mathrm{d}\lambda \qquad [1]$$

and x refers to A, or B, or C, or D, and $S_{\lambda x}$ is the spectral radiance distribution that produces the stimulus

$$S_\lambda = \frac{\mathrm{d}S(\lambda)}{\mathrm{d}\lambda}$$

where $S(\lambda)$ is the spectral radiance. K_m is the maximum value of the luminous efficacy of radiation, its value is $683\ \mathrm{lm\,W^{-1}}$ (see the discussion of luminous flux in the sub-section on photometric quantities). To be precise the integration should go from 0 nm to infinity, but it is usual to define the lower and upper wavelength limits of the visible spectrum as 380 nm and 780 nm. $V(\lambda)$ is defined between 360 nm and 830 nm, and $V'(\lambda)$ between 380 nm and 780 nm, see Figure 1.

We have to stress that the concept of luminance – and the entire system of the present-day photometry – is not to quantify brightness perception; it is only a reasonable approximation for near white stimuli. For colored lights a brightness–luminance discrepancy exists (called the Helmhotz–Kohlrausch effect: saturated colors look brighter than predicted by luminance). Luminance is, however, a good description for the visibility of fine details, thus it is a good concept for lighting calculations.

Photometry has a unique situation in the SI system of units: the candela (cd) is a base unit of the SI system, the only one that is connected to psychophysical phenomena. By 1979, in the definition of the 16th General Conference of Weights and Measures, the candela was traced back to radiation quantities; nevertheless it was kept as the base unit of photometry:

> The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} Hz and that has a radiant intensity in that direction of $1/683\ \mathrm{W\,sr^{-1}}$. ($540 \times 10^{12}$ Hz corresponds approximately to 555 nm.)

To be able to calculate a photometric quantity of radiation of other wavelengths, one of the psychophysical functions $V(\lambda)$ or $V'(\lambda)$ has to be used and eqn [1] applied. We will see later that besides $V(\lambda)$ and $V'(\lambda)$ in modern applications, a number of further spectral luminous efficiency functions might be used.

Photometric Quantities

As discussed in the previous section, from the point of view of vision, the most important quantity is

luminance. Looking, however, at the definition of the base unit, it is obvious that from the physical point of view the definition of a quantity corresponding to radiant power, measured in watts, can help to bridge the gap between photometry and radiometry: luminous flux, measured in lumens (lm), with the symbol Φ, is defined by eqn [1], with $S_\lambda(\lambda)$ inserted in W m^{-1}.

Based on this more practical quantity of luminous flux and its unit, lm, the different quantities used in photometry can be built up as follows:

$$\textit{Luminous flux}: \quad \Phi = K_m \int_0^\infty \frac{\mathrm{d}\Phi_e(\lambda)}{\mathrm{d}\lambda} V(\lambda)\mathrm{d}\lambda \qquad [2]$$

where $\Phi_e(\lambda)$ is the radiant flux measured in W,

$$\Phi_{e,\lambda} = \frac{\mathrm{d}\Phi_e(\lambda)}{\mathrm{d}\lambda}$$

is the spectral distribution (or spectral concentration) of the radiant flux (W m^{-1}). $K_m = 683$ lm W^{-1}, is the maximum value of the luminous efficacy of radiation for $\lambda_m \approx 555$ nm.

Similar equations can be written for scotopic luminous flux, with $V(\lambda)$ exchanged with $V'(\lambda)$, where $K'_m = 1700$ lm W^{-1} for $\lambda_m \approx 507$ nm. All further quantities can be defined both for photopic and scotopic conditions. Here we write them only as photopic quantities.

$$\textit{Luminous intensity}: \quad I = \frac{\mathrm{d}\Phi}{\mathrm{d}\Omega} \qquad [3]$$

where $\mathrm{d}\Phi$ is the luminous flux traveling in an elementary solid angle $\mathrm{d}\Omega$, assuming a point source (see Figure 2). The unit of luminous intensity is the candela (cd = lm sr^{-1}).

$$\textit{Luminance}: \quad L = \frac{\partial^2\Phi}{\partial A \cos\Theta\, \partial\Omega} \qquad [4]$$

where $\mathrm{d}\Phi$ is the luminous flux traveling in an elementary solid angle $\mathrm{d}\Omega$, $\mathrm{d}A$ is the elementary surface area emitting the radiation, and Θ is the angle between the normal of $\mathrm{d}A$ and the direction of luminance measurement (see Figure 3). The unit of luminance is cd m^{-2} (in some older literature called nit).

$$\textit{Illuminance}: \quad E = \frac{\mathrm{d}\Phi}{\mathrm{d}A} \qquad [5]$$

where $\mathrm{d}\Phi$ is the luminous flux incident on the $\mathrm{d}A$ element of a surface (see Figure 4). The unit of illuminance is the lux (lx = lm m^{-2}).

A remark on the use of SI units and related quantities

No prefixes can be added to the SI units, thus irrespective whether one measures luminance based

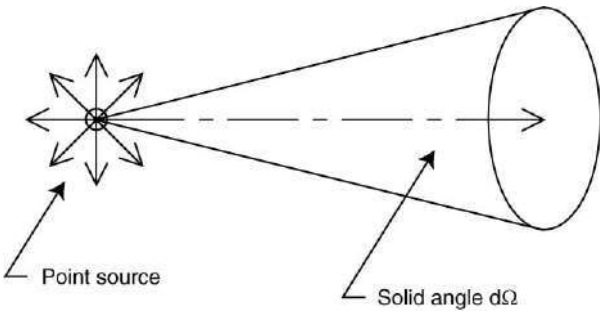

Figure 2 Concept of point source, solid angle, and luminous intensity.

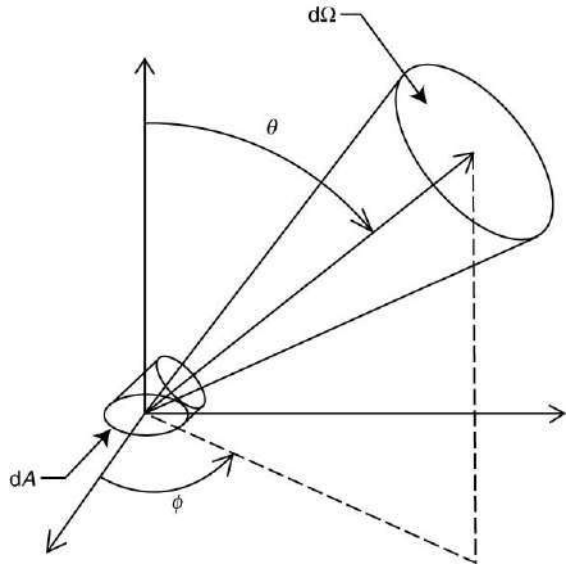

Figure 3 Geometry for the definition of luminance.

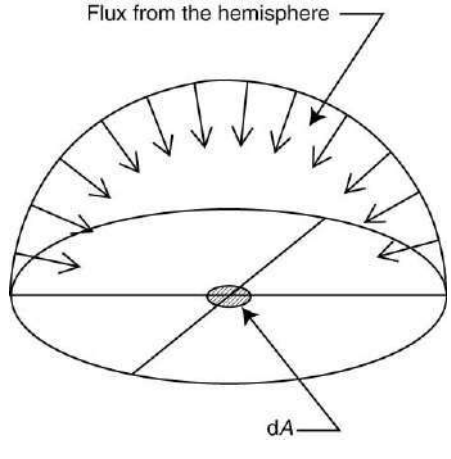

Figure 4 Illuminance is the total luminous flux per unit area incident at a point coming from a hemispherical solid angle.

on the photopic $V(\lambda)$ function, or the scotopic $V'(\lambda)$ function, the so-determined photopic luminance or scotopic luminance is measured in cd m^{-2} (no photopic or scotopic lumen, candela, etc., exist!). This often creates confusion, because only for a 555 nm monochromatic radiation will a 1 cd m^{-2} photopic or scotopic luminance produce equal visual sensation, for every other wavelength the two are not commensurable. One can only relate photopic measurement results to photopic ones, and scotopic measurement results to scotopic ones.

Concepts of Advanced Photometry

The quantities and units, as described in the previous section, are the quantities and units internationally agreed by the Meter Convention and the International Standards Organization (ISO). Modern photometric applications need, however, some further quantities and weighting functions. Thus CIE, the international organization for the development of standards in the field of optical radiation measurements, has defined a series of further weighting functions and quantities. The most important ones are the following:

$V_M(\lambda)$ function

In the blue part of the spectrum (below 460 nm) the values of the $V(\lambda)$ function turned out to be too low. For decades this was only of concern for the vision research community, but with the introduction of blue LEDs and other short wavelength emitting sources (e.g., blue channel of displays) this shortcoming of the $V(\lambda)$ function became of practical importance. The 'CIE 1988 2° spectral luminous efficiency function for photopic vision' corrects this anomaly (see Figure 5).

$V_{10}(\lambda)$ function

The official $V(\lambda)$ function is valid only for foveal vision (i.e., for targets that are smaller than 4° of visual angle). The foveal area of the retina is covered with a yellow pigmented layer (macula lutea) that absorbs in the blue part of the spectrum, at larger visual angles this screening is not effective anymore. The $V_{10}(\lambda)$ function was determined for a visual angle of 10° (see Figure 5). Its international recommendation is still under consideration at the time of writing this article and if accepted, it will be recommended for targets seen at approximately 10° off-axis, e.g., for measuring the photometric properties of traffic signs and signals, where the driver has to observe the information in the periphery of his or her eye.

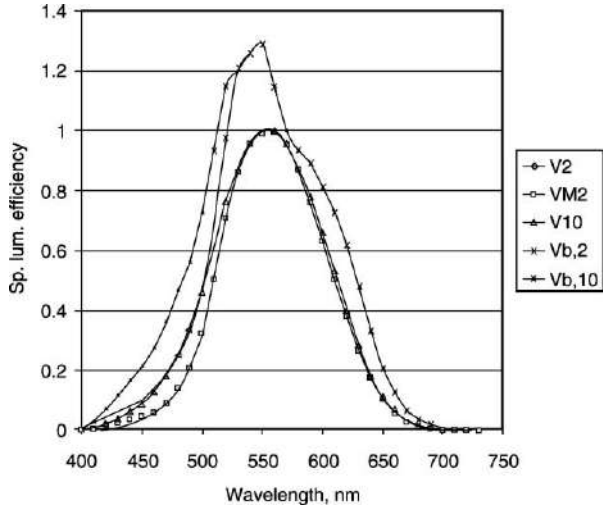

Figure 5 Spectral luminous efficiency (SPL) functions defined or under consideration for international adoption: V2: standard $V(\lambda)$ function; VM2: CIE 1988 $V_M(\lambda)$ function, this is equivalent to the brightness SPL for a point source; V10: 10° visual field SPL, Vb,2: 2° visual field brightness SPL, $V_{b,2}(\lambda)$; Vb,10: 10° visual field brightness SPL, $V_{b,10}(\lambda)$.

Brightness matching functions

As already mentioned in the Introduction, luminance is not a good correlate of brightness, which is a human perception. To find a better correlation of brightness the problem to be addressed is that brightness is a nonadditive phenomenon, i.e., in eqn [1] one cannot add (integrate) the monochromatic radiations to get a brightness correlation for a nonmonochromatic radiation. Brightness evaluating spectral luminous efficiency functions can be used only to compare monochromatic radiations. The CIE has compiled such functions for point sources, 2° and 10° visual field sizes. Figure 5 also shows these functions.

Due to the fact that the brightness perception is nonadditive in respect of the stimuli that produce it, no brightness photometry can be built that uses equations, as shown in eqn [1]. For brightness description we have to rely on the concept of equivalent luminance, a term the definition of which has recently been updated.

Equivalent luminance

Of a field of given size and shape, for a radiation of arbitrary relative spectral distribution L_{eq} : Luminance of a comparison field in which monochromatic radiation of frequency 540×10^{12} Hz has the same brightness as the field considered under the specified photometric conditions of measurement; the comparison field must have a specified size and shape

which may be different from that of the field considered.

To build an instrument that measures this quantity is a real challenge for the future.

Advanced Use of Photometry

Based on above newly defined quantities, several attempts are under way to extend the usefulness of photometry in designing the human visual environment. The eye is an optical system and as in every such system, the depth of focus and the different aberrations of the system will decrease with decreasing pupil size. Pupil size will decrease with increasing illumination, and in the case of constant luminance with higher content of short wavelength radiation. Thus, there exists a school of researchers who advocate that increased blue content in the light has beneficial effects on vision, and one should extend the classical photopic-based photometry with a scotopic-based one to properly describe the visual effect of lighting.

Other investigations are concerned about the visibility at low light levels, the range used in street lighting (from about a few thousands of a candela per square meter to about a few candelas per square meter, according to one definitions: 10^{-3}cd m^{-2}–3 cd m^{-2}). In this mesopic range both rods and cones are contributing to vision, and this changes with lighting level and direction of view (for foveal vision, i.e., looking straight ahead, photopic photometry seems to hold even at low light levels). For peripheral visual angles brightness perception and the perception of an object (a signal, sign or obstacle in a nighttime driving situation) seem to have different spectral responsivity. In driving situations the necessary reaction time of the driver is an important parameter, thus experiments are going on to define a photometric system based on reaction time investigations.

In indoor situations apart from the necessary level of illumination, the observed glare is a contributor whether an environment will be accepted as pleasing or annoying. Illuminating engineering distinguishes between two types of glare: disability glare reduces visibility, discomfort glare is just an annoying experience without influencing the short-term task performance. An interesting question is the spectral sensitivity to discomfort glare, as it can influence not only indoor but also outdoor activity. Preliminary experiments seem to show that luminance sensitivity and discomfort glare sensitivity have different spectral distribution; glare sensitivity seems to peak at shorter wavelengths.

The above might be related to a further question, lying already at the boundaries of photometry, but that has to be considered in photometric design and measurement: the human daily and yearly rhythm (circadian and seasonal rhythm) of human activity coupled to hormone levels. They are influenced by light, as, for example, the hormone melatonin production is influenced by light exposure. Physiological investigations showed that melatonin production suppression has a maximum around 460 nm and might be coupled to a radiation sensitive ganglion cell in the retina. Whether discomfort sensation is mediated via the same neural pathway or via a visual one has not yet been decided. But photometry has to take these also into consideration and in the future, measurement methods and instruments to determine them, will have to be developed.

Advances in Photometric Measurements

Primary Standards

The main concern in photometry is that the uncertainty of photometric measurements is still much higher than that in other branches of physics. This is partly due to the higher uncertainty in radiometry and partly to the increase in uncertainty within the chain of uncertainty propagation from the National Laboratory to the workshop floor measurement.

In National Standards Laboratories, very sophisticated systems are used to determine the power of the incoming radiation and then elaborated spectroradiometric techniques are used to evaluate the radiation in the form of light, i.e., perform photometric measurements (e.g., NIST CIRCUS equipment, where multiple laser sources are used as power sources, and highly sophisticated methods to produce a homogeneous nonpolarized radiation field for the calibration of secondary photometric detectors).

There is, however, also a method to supply end users with absolute detectors for the visible part of the spectrum. Modern high-end Si photovoltaic cells have internal quantum efficiencies in the visible part of the spectrum of over 99.9%. Reflection losses at the silicon surface are minimized by using three or more detectors arranged in a trap configuration, where the light reflected from one detector is fed to the second one, from there to the third one, and eventually to some further ones. In a three detector configuration, as shown in Figure 6, the light from the third detector is reflected back to the second and from there to the first one. As every detector reflects only a small amount of radiation, by the fifth reflection practically

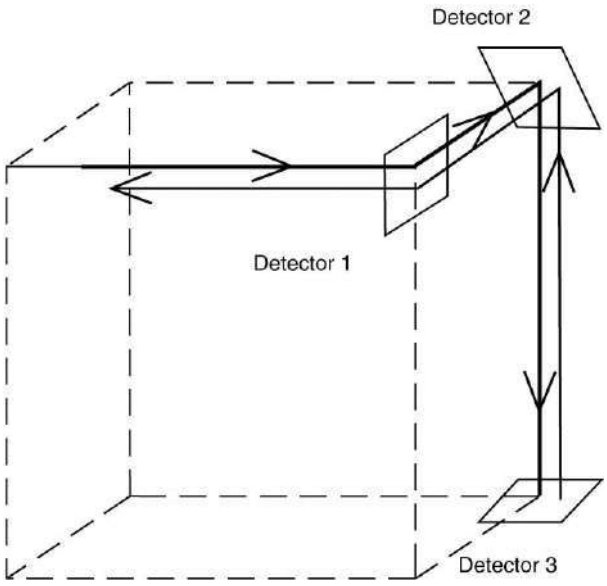

Figure 6 Schematic layout of a three Si-cell trap detector: light comes in from the left, is first partly absorbed, partly reflected on Detector 1, then on Detector 2, then on Detector 3, from where it is reflected back to Detector 2 and 1.

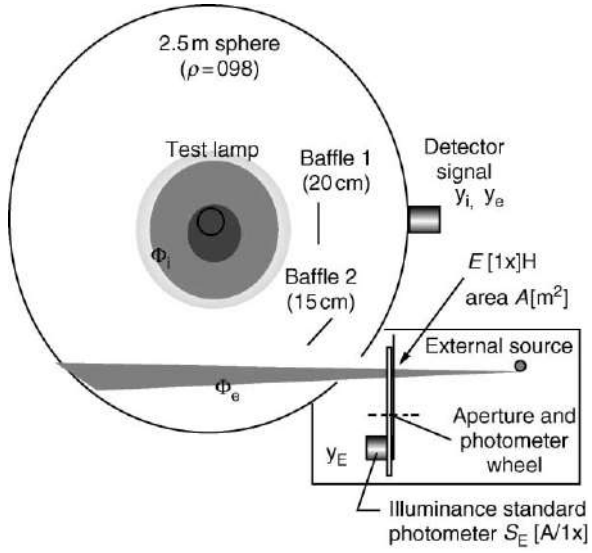

Figure 7 Arrangement of the absolute integrating sphere system developed at NIST for the detector-based total luminous flux calibration. By permission of IESNA from Ohno Y and Bergman R (2003) Detector-referenced integrating sphere photometry for industry. *J. IES Summer* 21–26.

all the radiation is absorbed and contributes to the electric signal. Such trap detectors have an almost 100% quantum efficiency in the visible part of the spectrum, and can be used as photometric detectors if a well designed color correcting filter is applied in front of the detector.

Secondary Type Measurements

In practical photometry the three most important quantities to be measured are the total luminous flux of different lamps, the illuminance in a plane and the luminance.

Light source measurement

Total luminous flux. The two methods to measure the total luminous flux is to use a goniophotometer or a photometer (Ulbicht) sphere. In goniophotometry, recent years have not brought major breakthroughs, the automation of the systems got better, but the principles are unchanged.

The integrating sphere photometer (a sphere with inner white diffuse coating, where the lamp is in the middle of the sphere) used to be a simple piece of equipment to compare total luminous flux lamps against flux standards. In recent years a new technique has been developed at NIST–USA. This enables the absolute measurement of luminous flux from illuminance measurement, the fundamentals of this new arrangement being shown in Figure 7: the test lamp is as usual in the middle of the sphere, but now light from an external source is introduced into the sphere. An illuminance meter measures the flux entering from this source. The sphere detector compares the two signals (y_i and y_e). Knowing the absolute characteristics of the sphere (a difficult measurement) one can determine the total luminous flux of the test lamp using the absolute illuminance value. As illuminance is easily determined from luminous intensity (from eqns [3] and [5] one gets with $d\Omega = dA/r^2$ that $E = I/r^2$, where r is the distance between the source and the illuminated surface, supposed to be perpendicular to the direction to the source), this technique permits us to derive the total luminous flux scale from illuminance or luminous intensity measurement using an integrating sphere.

Luminous intensity of LEDs. An other major break-through achieved during the past years was the unified measurement of LED intensity. Light-emitting diodes became, in recent years, important light sources for large scale signaling and signing, and it is foreseen that they will become important contributors in every field of light production (from car headlamps to general illumination). The most fundamental parameter of the light of an LED is its luminous intensity. The spatial power distribution of LEDs is usually collimated, but the LEDs often squint, as seen in Figure 8. In the past, some manufacturers measured the luminous intensity in

the direction of maximum emission, others used the direction of the optical axis for this quantity. The highly collimated character of the radiation made measurements in far field rather difficult. Therefore, CIE recommended a new term and measuring geometry: average LED intensity can be measured under two measuring conditions, as shown in Figure 9. The detector has to be set in the direction of the LED mechanical axis (discussions are still going on as to what the reference direction should be with modern surface mounted LEDs, as with those the mechanical axis is ill-defined, the normal to the base-plane could be a better reference direction). The detector has to have an exactly 1.00 cm^2 circular aperture, and the distance between this aperture and the tip of the LED is for condition A, $d = 0.316$ m, and for Condition B, $d = 0.100$ m (these two distances with the 1.00 cm^2 detector area provide 0.001 sr (steradians) and 0.01 sr opening angles). Recent international round-robins have shown that, based on the new recommendations, agreement between different laboratories decreased from the 10 to 20% level to 1 to 2%. The remaining difference is mainly due to the fact that the LEDs emit in narrow wavelength bands, and the transfer of the calibration value for the 100 mm^2 detector from a white (CIE Standard Illuminant A color temperature) incandescent lamp to the narrow band LED emission is still uncertain, mainly due to stray light effects in the spectral responsivity and emission measurement.

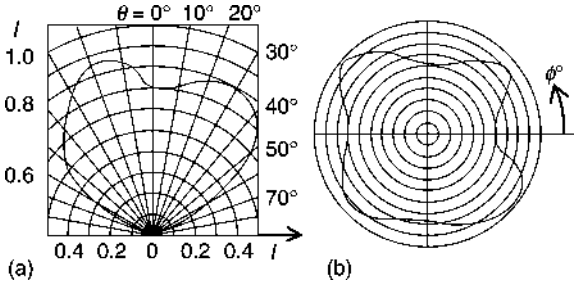

Figure 8 Spatial light distribution of an LED, Figure 8a shows the distribution of a 'squinting' LED in a plane including the optical axis, Figure 8b shows light distribution in a plane perpendicular to the optical axis. By permission of the Commission Internationale de l'Eclairage, from the Publication "Measurement of LEDs" CIE 127-1997; CIE Publications are obtainable from the CIE Central Bureau: Kegelgasse 27, A-1033 Wien, Austria.

Luminance distribution measurement

The human observer sees luminance (and color) differences. Thus for every illuminating engineering design task the luminance distribution in the environment is of utmost importance. Traditionally this was measured using a spot-luminance meter, aiming the device into a few critical directions. The recent development of charge coupled device (CCD) two-dimensionally sensitive arrays (and other, e.g., MOSFET, CMOS, Charge injection device (CID), charge imaging matrix (CIM) systems: as for the time being the CCD technology provides best performance, we will refer to two-dimensional electronic image capture devices as to CCD cameras) opened the possibility of using an image-capturing camera for luminance distribution measurements. Such measurements are badly needed in display calibration, near-field photometry (photometry in planes nearer than in which the inverse square law holds), testing of car headlamp light distribution, glare, and homogeneity studies indoors and outdoors, etc.

Solid-state cameras have the big advantage over older type vacuum-tube image capturing devices, that the geometric position and alignment of the single pixels is well-defined and stays constant. Nowadays, CCD image detector chips are mass produced, and one can get a variety of such devices and cameras starting with some small resolution

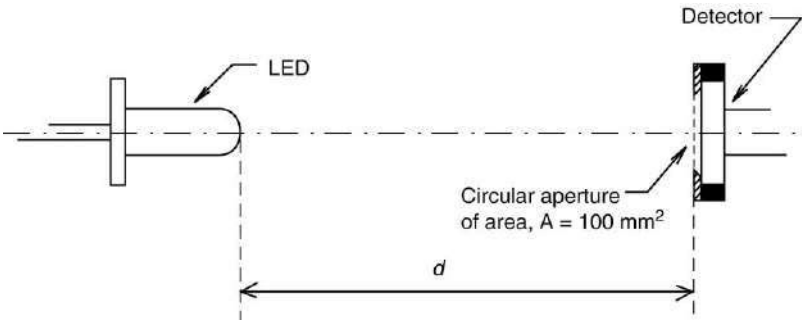

Figure 9 Schematic diagram of CIE Standard Conditions for the measurement of Average LED Intensity. Distance $d = 0.316$ m for Condition A and $d = 0.100$ m for Condition B. By permission of the Commission Internationale de l'Eclairage, from the Publication "Measurement of LEDs" CIE 127-1997; CIE Publications are obtainable from the CIE Central Bureau: Kegelgasse 27, A-1033 Wien, Austria.

(few thousand pixels) devices up to cameras with tens of mega pixel resolution. Detectors are now available with internal intensification enabling measurements down to a few photons per second intensity levels.

Main problems with these detectors are:

- spectral and absolute nonuniformities of the single pixels (see Figure 10), where a spatial homogeneity map of a CCD two-dimensional array detector is shown; the irregular 3% sensitivity change on the surface of the detector is negligible for imaging purposes, but has to be corrected in cases of photometric measurements. Even if the receptor chip would have an absolutely homogeneous sensitivity, there would be a drop in response from the middle of the imaging area to the boarders: Light reaching the edges of the detector reach the detector at an oblique angle and this produces a decrease of sensitivity with α^4, where α is the angle of incidence, measured from the middle of the lens to the given pixel of the detector and the surface normal of the detector.
- aliasing effects if the pixel resolution is not large enough to show straight lines as such when they are not in line with a pixel row or column.
- nonlinearity and cross-talk among the adjacent pixels. Figure 11 shows the so-called 'inverse gamma' characteristic of a CCD camera. The digital electronic output of the camera shows a $Y = E^{-\gamma}$ type function, where Y is the output DAC (digital-analog converter) values, E is the irradiance of the pixel, and γ is the exponent (this description comes from the film industry, where the film density depends exponentially on the irradiation; display devices have usually a nonlinear input (DAC value) – output (luminance) characteristic, and the camera inverse gamma value corrects for this output device characteristic. This is, however, not required if the camera is used for photometric measurements; this built-in nonlinearity has to be corrected in the evaluating software if the camera is intended for photometric measurements.
- to be able to perform photometric measurements the camera has to have a spectral responsivity corresponding to the CIE $V(\lambda)$-function. Many cameras have built-in filters to do this – eventually also red and blue filters to be able to capture color – but the color correction of cameras, where the correction is made by small adjacent filter chips, is usually very poor. Figure 12 shows the spectral

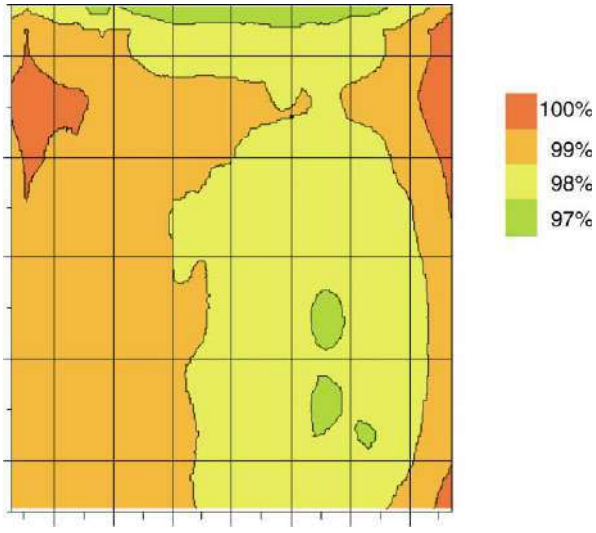

Figure 10 Spatial homogeneity of a two-dimensional CCD array.

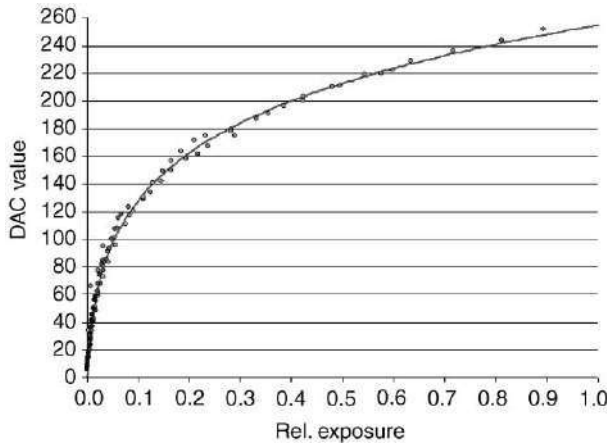

Figure 11 'Inverse gamma' characteristic of a commercial digital photographic camera, measurement points are shown at different speed settings, curve is a model function representative of the camera response.

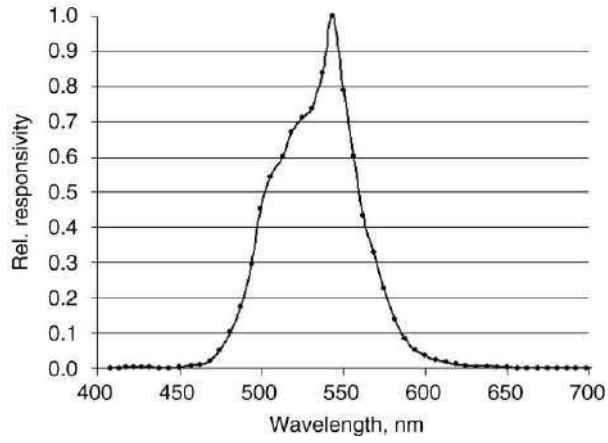

Figure 12 Spectral sensitivity of a commercial digital photographic camera.

sensitivity curve of a digital photographic camera; the output signal is produced by an internal matrix transformation of the signals produced by adjacent pixels equipped with different color filters. Figure 13 shows the spectral sensitivity of a CCD camera specially designed for photometric measurements. Naturally, meaningful photometric measurements can be made only with such a camera. Very often, however, an approximate luminance distribution is enough, and then a picture captured by a digital photographic camera, plus one luminance measurement of a representative object for absolute calibration, suffices.

Above nonspectral systematic errors can be corrected by appropriate soft-ware, so a CCD camera photometer, as shown schematically in Figure 14, is well suited to measure display characteristics, indoor and outdoor luminance distributions. The challenge for illuminating engineering is at present how the many millions of luminance values can be evaluated to get to meaningful light measurement data.

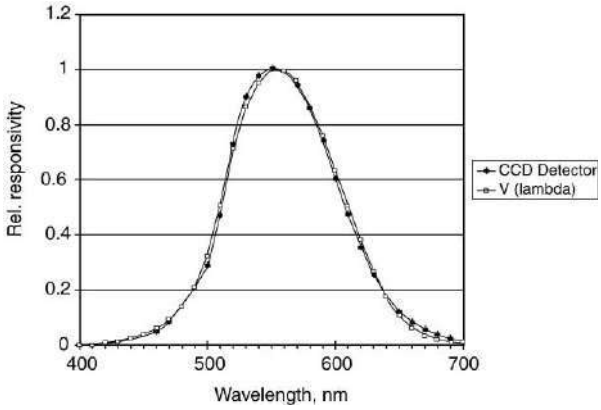

Figure 13 Spectral sensitivity of a CCD camera specially designed for photometric measurements. Kindly supplied by INPHORA Inc.

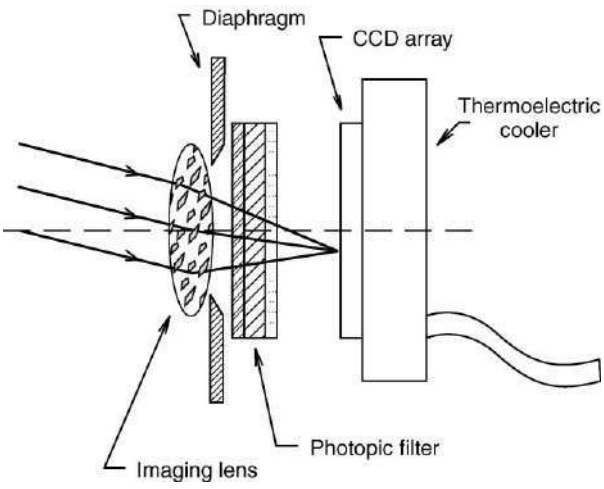

Figure 14 Cross-section of a photometric CCD camera.

The real challenge will come if visual science provides better hints how the human visual system evaluates the illuminance distribution on the retina, and instrument manufacturers will be able to capture signals corresponding to those produced by the receptor cells and provide the necessary algorithms our brain uses to get to brightness, lightness, color, luminance contrast, and glare type of output information.

Concluding Remarks

Advances in optical instrumentation, both in the field of light sources and detectors – coupled with the possibilities modern digital computation (eventually in the future increasing the use of neural networks) – provide already many new measurement technical solutions and further ones are certainly underway.

The use of LEDs needs the rethinking of many classical illuminating engineering concepts, from visibility and glare evaluation, evenness of illumination to color rendering. All of them are coupled with problems in basic photometry. Thus, there is a need to re-evaluate concepts used in design techniques. The new area-sensitive detectors provide methods of determining classical photometric quantities of entire visual fields in one shot, but already foreshadow the development of new quantities that correlate better with visual perceptions.

List of Units and Nomenclature

Terms with an $*$ refer to definitions published by the CIE in the International Lighting Vocabulary, CIE 17.4:1986, where further terms and definitions related to light and lighting are to be found.

Candela*: SI unit of luminous intensity: The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt per steradian. (16th General Conference of Weights and Measures, 1979).

$$1\ \mathrm{cd} = 1\ \mathrm{lm\ sr}^{-1}$$

Equivalent luminance*: Luminance of a comparison field in which monochromatic radiation of frequency 540×10^{12} Hz has the same brightness as the field considered under the specified photometric conditions of measurement; the comparison field must have a specified size and shape which may be

different from that of the field considered.

$$\text{unit} : \text{cd m}^{-2}$$

Notes:

1. Radiation at a frequency of 540×10^{12} Hz has a wavelength in standard air of 555.016 nm.
2. A comparison field may also be used in which the radiation has any relative spectral distribution, if the equivalent luminance of this field is known under the same conditions of measurement.

Far field photometry: Photometry where the inverse square law is valid.

Flicker photometer*: Visual photometer in which the observer sees either an undivided field illuminated successively, or two adjacent fields illuminated alternately, by two sources to be compared, the frequency of alteration being conveniently chosen so that it is above the fusion frequency for colours but below the fusion frequency for brightnesses.

Fovea*: Central part of the retina, thin and depressed, which contains almost exclusively cones and forming the site of most distinct vision.

Note: The fovea subtends an angle of about 0.026 rad (1.5 degree) in the visual field.

Goniophotometer*: Photometer for measuring the directional light distribution characteristics of sources, luminaires, media or surfaces.

Illuminance*: Quotient of the luminous flux $d\Phi_v$ incident on an element of the surface containing the point, by the area dA of that element.

Equivalent definition. Integral, taken over the hemisphere visible from the given point, of the expression $L_v \cos\theta \, d\Omega$, where L_v is the luminance at the given point in the various directions of the incident elementary beams of solid angle $d\Omega$, and θ is the angle between any of these beams and the normal to the surface at the given point.

$$E_v = \frac{d\Phi_v}{dA} = \int_{2\pi sr} L_v \cos\theta \, d\Omega$$

$$\text{unit} : \text{lx} = \text{lm m}^{-2}$$

Inverse square law: The illumination at a point on a surface varies directly with the luminous intensity of the source, and inversely as the square of the distance between the source and the point if the source is seen as a point source.

Lumen*: SI unit of luminous flux: Luminous flux emitted in unit solid angle (steradian) by a uniform point source having a luminous intensity of 1 candela. (9th General Conference of Weights and Measures, 1948).

Equivalent definition. Luminous flux of a beam of monochromatic radiation whose frequency is 540×10^{12} hertz and whose radiant flux is 1/683 watt.

Luminance*: Quantity defined by the formula

$$L = \frac{\partial^2 \Phi}{\partial A \cos\theta \, \partial\Omega}$$

where $\partial\Phi$ is the luminous flux transmitted by an elementary beam passing through the given point and propagating in the solid angle $d\Omega$ containing the given direction; dA is the area of a section of that beam containing the given point, θ is the angle between the normal to that section and the direction of the beam.

$$\text{unit} : \text{cd m}^{-2}$$

Luminous intensity*: Quotient of the luminous flux $d\Phi_v$ leaving the source and propagated in the element of solid angle $d\Omega$ containing the given direction, by the element of solid angle:

$$I_v = \frac{d\Phi_v}{d\Omega}$$

$$\text{unit} : \text{cd} = \text{lm sr}^{-1}$$

Lux*: SI unit of illuminance: Illuminance produced on a surface of area 1 square meter by a luminous flux of 1 lumen uniformly distributed over that surface.

$$1 \text{ lx} = 1 \text{ lm m}^{-2}$$

Note: Non-metric unit: **lumen per square foot** (lm ft^{-2}) or **footcandle** (fc) (USA) = 10.764 lx.

Near-filed photometry: photometry made in the vicinity of an extended source, so that the inverse square law is not valid.

Photometer (or Ulbicht) sphere*: A hollow sphere, whitened inside. Owing to the internal reflexions in the sphere, the illumination on any part of the sphere's inside surface is proportional to the luminous flux entering the sphere, or produced inside the sphere by a lamp. The illuminance of the internal sphere wall is measured via a small window.

Pixel: The individual picture elements of an image or elements in a display that can be addressed individually.

Radiance*: Quantity defined by the formula

$$L = \frac{\partial^2 \Phi}{\partial A \cos\theta \, \partial\Omega}$$

where $\partial\Phi$ is the radiant flux transmitted by an elementary beam passing through the given point and propagating in the solid angle $d\Omega$ containing the given direction; dA is the area of a section of that beam containing the given point, θ is the angle between the normal to that section and the direction of the beam.

$$\text{unit} : \text{W m}^{-2} \text{ srad}$$

Retina*: Membrane situated inside the back of the eye that is sensitive to light stimuli; it contains photoreceptors, the cones and the rods, and nerve cells that interconnect and transmit to the optic nerve the signals resulting from stimulation of the photoreceptors.

Spectral*: An adjective that, when applied to a quantity X pertaining to electromagnetic radiation, indicates:

- either that X is a function of the wavelength λ, symbol: $X(\lambda)$,
- or that the quantity referred to is the spectral concentration of X, symbol: $X_\lambda \equiv dX/d\lambda$.

X_λ is also a function of λ and in order to stress this may be written $X_\lambda(\lambda)$ without any change of meaning.

Spectral luminous efficiency function: for photopic vision, $V(\lambda)$; for scotopic vision $V'(\lambda)$*: Ratio of the radiant flux at wavelength λ_m to that at wavelength λ such that both radiations produce equally intense luminous sensations under specified photometric conditions and λ_m is chosen so that the maximum value of this ratio is equal to 1.

Unless otherwise indicated, the values used for the spectral luminous efficiency in photopic vision are the values agreed internationally in 1924 by the CIE (*Compte Rendu* 6^e session, p.67), completed by interpolation and extrapolation (*Publications CIE* No.18 (1970), p.43 and No. 15 (1971), p.93), and recommended by the International Committee of Weights and Measures (CIPM) in 1972. For *scotopic vision*, the CIE in 1951 adopted, for young observers, the values published in *Compte Rendu* 12^e session, Vol. 3, p. 37, and ratified by the CIPM in 1976. These values define respectively the $V(\lambda)$ or $V'(\lambda)$ functions.

Total luminous flux: luminous flux of a source emitted into 4π steradians.

Trap detector: Detector array prepared from detectors of high internal quantum efficiency, where the reflected radiation of the first detector is directed to the second one, and so on so that practically all the radiation is absorbed by one of the detectors (the radiation is "trapped" in the detector system).

See also

Displays. **Incoherent sources:** Lamps.

Further Reading

Commission Internationale de l'Eclairage: *The basis of physical photometry* (1983) CIE 18.2.

Commission Internationale de l'Eclairage: Proceedings of the CIE Symposium 2002 on *Temporal and Spatial Aspects of Light and Colour Perception and Measurement*, August 2002: Publ.CIE x025:2003.

Cromer CL, Eppledauer G, Hardis JE, Larason TC and Parr AC (1993) National Institute of Standards and Technology detector-based photometric scale. *Applied Optics* 32: 2936–2948.

Decusatis C (ed.) (1997) *Handbook of Applied Photometry*. Woodbury, New York: American Institute of Physics.

Rea MS (ed.) (2000) Illuminating Engineering Society of North America. *The IESNA Lighting Handbook: Reference and Application*, 9th edn.

Kaiser PK and Boynton RM (1966) *Human Color Vision*, 2nd edn. Washington, DC: Optical Society of America.

McCluney WR (1994) *Introduction to radiometry and photometry*. Boston, London: Artech House.

Ohno Y and Bergman R (2003) Detector-referenced integrating sphere photometry for industry. *J IES Summer* 21–26.

Walsh JWT (1953) *Photometry*. London: Constable.

Scatterometry

J C Stover, The Scatter Works, Inc., Tucson, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Scattered light is a limiting source of optical noise in many advanced optical systems, but it can also be a sensitive indicator of optical component quality. Consider the simple case of a telescope successfully used to image a dim star against a dark background; however, if light from a bright source (such as the moon located well out of the field of view) enters the telescope, it will scatter from the interior walls and the imaging optics themselves. Some of this light eventually reaches the detector and creates a dim background haze that washes out the image of the distant star. A good telescope design accounts for these effects by limiting potential scatter propagation paths and by requiring that critical elements in the optical system meet scatter specifications. This means doing a careful system analysis and a means to quantify the scattering properties of the telescope components. This article discusses modern techniques for quantifying, measuring, and analyzing scattered light, and reviews their development.

Like many scientific advances moving scatterometry from an art to a reliable metrology was done in a series of small hops (not always in the same direction), rather than a single leap. It started in 1961 when a paper by Hal Bennett and Jim Porteous reported measurements made by gathering most of the light scattered from front surface mirrors and normalizing this signal by the much larger specular reflection. They defined this ratio as the total integrated scatter (TIS), and using a scalar diffraction theory result drawn from the radar literature, related it to the surface root mean square (rms) roughness. By the mid-1970s, several angle-resolved scatterometers had been built as research tools in university, government, and industry labs. Unfortunately, instrument operation and data manipulation were generally poor, and meaningful comparison measurements were virtually impossible due to instrument differences, sample contamination, and confusion over what parameters should be compared. Analysis of scatter data, to characterize sample surface roughness, was the subject of many publications. A derivation of what is commonly called 'BRDF' (bidirectional reflectance distribution function) was published by Nicodemus and co-workers at the National Bureau of Standards (now the National Institute of Science and Technology or NIST) in 1970, but did not gain common acceptance as a way to quantify scatter measurements until the late 1980s when the advent of small computers, combined with inspection requirements for defense-related optics, dramatically stimulated the development of scatter metrology. Commercial laboratory instrumentation became available that could measure and analyze as many as 50 to 100 samples a day, and the number (and sophistication) of measurement facilities increased dramatically. The first ASTM Standards were published (TIS in 1987 and BRDF in 1991), but it was still several years before most publications correctly used these quantifying terms. Government defense funding decreased dramatically in the early 1990s, following the end of the Cold War, but the economic advantages of using scatter metrology and analysis for space applications and in the rapidly advancing semiconductor industry, continued state-of-the-art advancements.

The following sections detail how scatter is quantified when related to area (roughness) and local (pit/particle) generating sources. Instrumentation and the use of scattering models are also briefly reviewed.

Quantifying Scattered Light

Scatter signals can be easily quantified as scattered light power per unit solid angle (in watts per steradian); however, in order to make the results more meaningful, these signals are usually normalized, in some fashion, by the light incident on the scatter source. The three ways commonly employed to do this are defined below.

If the scattering feature in question is uniformly distributed across the illuminated spot on the sample (such as surface roughness), then it makes sense to normalize the collected scattered power in watts/steradian by the incident power. This simple ratio, which has units of inverse steradians, was commonly referred to as 'the scattering function.' Although this term is occasionally still found in the literature, it has been generally replaced by the closely related BRDF, which is defined by the differential ratio of the sample radiance normalized by its irradiance. After some simplifying assumptions are made, this reduces to the original scattering function with a cosine of the polar scattering angle in the denominator. The BRDF, defined in this manner, has become the standard way to report angle-resolved scatter from features that uniformly fill the illuminated spot. The cosine term results from the fact that NIST used radiometric terms to define BRDF:

$$\mathrm{BRDF} = \frac{\mathrm{Ps}/\Omega}{\mathrm{Pi}\cos\theta_s} \qquad [1]$$

The scatter function is often referred to as the 'cosine corrected BRDF' and is simply equal to the BRDF multiplied by the cosine of the polar scattering angle. Figure 1 gives the geometry for the situation, and defines the polar and azimuthal angles (θ_s and ϕ_s), as well as the solid collection angle (Ω). Other common abbreviations are BSDF, for the more generic bidirectional scatter distribution function, and BTDF for quantifying transmissive scatter.

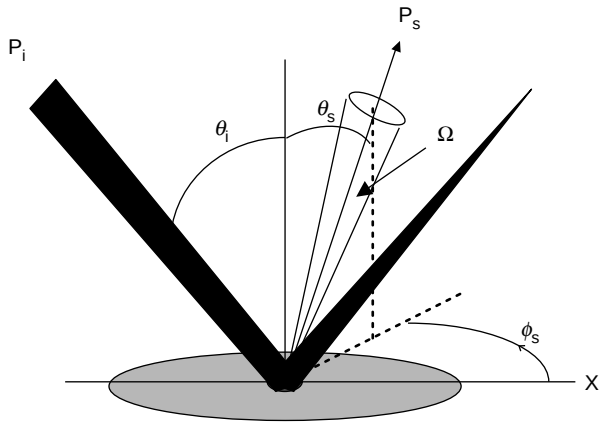

Figure 1 Scatter analysis uses standard spherical coordinates to define terms.

Integration of the scatter signal over much of the scattering hemisphere allows calculation of TIS, as the ratio of the scatter signal to the reflected specular power. This integration is usually carried out experimentally in such a way that both the incident beam and reflected specular beam are excluded. In the most common TIS situation, the beam is incident at a small angle near surface normal, and the integration is done from small values of θ_s to almost 90 degrees. If the fraction of light scattered from the specular reflection is small and if the scatter is caused by surface roughness, then it can be related to the rms surface roughness of the reflecting surface. As a ratio of powers, the TIS is a dimensionless quantity. The normalization is by Pr (instead of Pi) because reductions in scatter caused by low reflectance do not influence the roughness calculation. The pertinent relationships are given below, where σ is the rms roughness and λ is the light wavelength:

$$\mathrm{TIS} = \mathrm{Ps/Pr} \cong (4\pi\sigma/\lambda)^2 \qquad [2]$$

Of course, all scatter measurements are integrations over a detector collection aperture, but the TIS designation is reserved for situations where the aim is to gather as much scattered light as possible, while 'angle resolved' designs are created to gain information from the distribution of the scattered light. Notice that TIS values become very large when measured from a diffuse surface, where the specular reflection is very small. Although TIS can be measured for any surface, the diffuse reflectance (equal to Ps/Pi) would often be more appropriate for diffuse surfaces. The various restrictions associated with relating TIS to rms roughness are detailed below.

Scatter from discrete features, such as particles and pits, which do not completely fill the illuminated spot, must be treated differently. This is because changes in spot size, with no corresponding change in total incident power, will change the incident intensity (watts/unit area) at the feature and thus also change the scatter signal (and BRDF) without any corresponding changes in the scattering feature. Clearly this is unacceptable if the object is to characterize the defect with scatter measurements. The solution is to define another quantification term, known as the differential scattering cross-section (DSC), where the normalization is the incident intensity at the feature (the units for DSC are area/steradian). Because this was not done in terms of radiometric units at the time it was defined, the cosine of the polar scattering angle is not in the definition. The same geometrical definitions, found in Figure 1, also apply for the DSC:

$$\mathrm{DSC} = \frac{\mathrm{Ps}/\Omega}{Ii} \qquad [3]$$

If the DSC is integrated over the solid angle associated with a collection aperture then the value has units of area. Because relatively small area focused laser beams are often used as a source, area is most commonly given in micrometers squared.

These three scatter parameters, the BRDF, the TIS, and the DSC, are obviously functions of system variables such as geometry, scatter direction (both in and out of the incident plane), incident wavelength and polarization, as well as feature characteristics. It is the dependence of the scatter signal on these system parameters that makes the scatter models useful for optimizing instrument designs. It is their dependence on feature characteristics that makes scatter measurement a useful metrology tool.

A key point needs to be stressed. When applied appropriately, TIS, BRDF, and DSC are absolute terms, not relative terms. The DSC of a 100 nm PSL in a given direction for a given source is a fixed value, which can be repeatedly measured and even accurately calculated from models. The same is true for TIS and BRDF values associated with surface roughness of known statistics. Scatter measuring instruments, such as particle scanners or lab scatterometers, can be calibrated in terms of these quantities. As has already been pointed out, the user of a scanner will almost always be more interested in characterizing defects than in the resulting scatter values, but the underlying instrument calibration can always be expressed in terms of these three quantities. This is true even though designers and users may find it convenient to use other metrics (such as polysterene latex (PSL) spheres) as a way to relate calibration.

Angle Resolved Scatterometers

The diagram in Figure 2 shows the most common scatterometer configuration. The source is fixed and the sample is rotated to the desired incident angle. The receiver is then rotated about the sample during scatter measurement. Most commonly, scatterometers operate just in the plane of incidence; however, instruments capable of measuring at virtually any location in either the reflective or transmissive hemispheres have been built. Although dozens of instruments have been built following this general design, other configurations are in use. For example, the source and receiver may be fixed and the sample rotated so that the scatter pattern moves past the receiver. This is easier mechanically than

moving the receiver at the end of an arm, but complicates analysis because the incident angle and the observation angle change simultaneously. Another combination is to fix the source and sample together, at constant incident angle, and rotate this unit (about the point of illumination on the sample) so that the scatter pattern moves past a fixed receiver. This has the advantage that a long receiver/sample distance can be used without motorizing a long (heavy) receiver arm. It has the disadvantage that heavy (or multiple) sources are difficult to deal with. Other configurations, with everything fixed, have been designed that employ several receivers to merely sample the BSDF and display a curve fit of the resulting data. This is an economical solution if the BSDF is relatively uniform without isolated diffraction peaks. The goniometer section of a real instrument, similar to that of Figure 2, is shown in Figure 3.

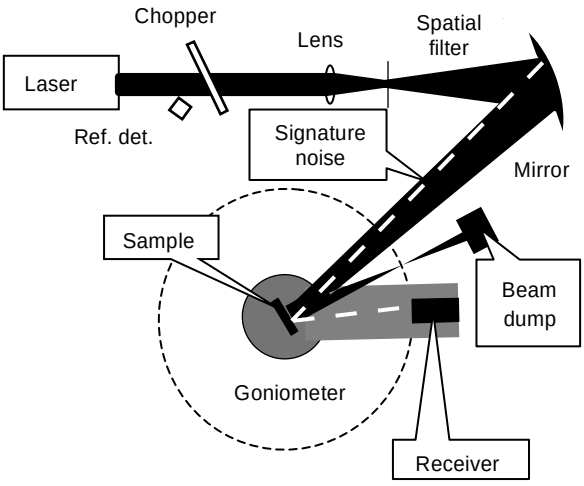

Figure 2 Basic elements of an incident plane scatterometer are shown.

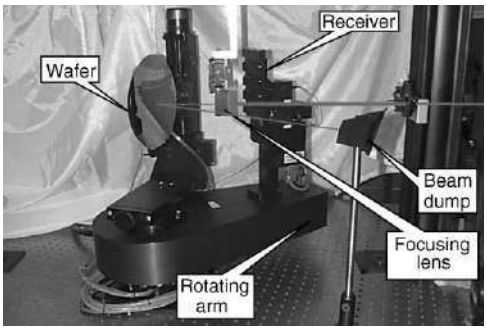

Figure 3 The author's scatterometer, which is similar to the diagram of Figure 2 is shown. In this case a final focusing lens is introduced to produce a very small illuminated spot on the silicon wafer sample. The white background was introduced to make the instrument easier to view.

Computer control of the measurement is essential to maximize versatility and minimize measurement time. The software required to control the measurement plus the display and analysis of the data can be expected to be a significant portion of total instrument development cost. The following reviews typical design features (and issues) associated with the source, sample mount and receiver components.

The source in Figure 2 is formed by a laser beam that is chopped, spatially filtered, expanded, and finally brought to a focus on the receiver path. The beam is chopped to reduce both optical and electronic noise. This is usually accomplished through the use of lock-in detection in the electronics package which suppresses all signals except those at the chopping frequency. Low noise, programmable gain electronics are essential to reducing system noise. The reference detector is used to allow the computer to ratio out laser power fluctuations and, in some cases, to provide the necessary timing signal to the lock-in electronics. Polarizers, wave plates, and neutral density filters are also commonly placed prior to the spatial filter. The spatial filter removes source scatter from the laser beam and presents a point source which is imaged by the final focusing element, in this case a mirror, to the detector zero position. Focusing the beam at this location allows near specular scatter to be more easily measured. Lasers are convenient sources, but are not necessary. Broadband sources are often required to meet a particular application or to simulate the environment where a sample will be used. Monochromators and filters can be used to provide scatterometer sources of arbitrary wavelength. The noise floors with these tunable incoherent sources increases as the spectral bandpass is narrowed, but they have the advantage that the scatter pattern does not contain laser speckle.

The sample mount can be very simple or very complex. In principal, six degrees of mechanical freedom are required to fully adjust the sample. The order in which these stages are mounted affects the ease of use (and cost) of the sample holder. In practice, it often proves convenient to either eliminate, or occasionally duplicate, some of these degrees of freedom. In addition, some of these axes may be motorized to allow the sample area to be raster-scanned to automate sample alignment or to measure reference samples. As a general rule, the scatter pattern is insensitive to small changes in incident angle but very sensitive to small angular deviations from specular. Instrumentation should be configured to allow location of the specular reflection (or transmission) very accurately. Receiver designs vary, but changeable entrance apertures, bandpass

filters, lenses, and field stops are generally positioned in front of the detector.

A serious measurement problem is getting light scattered by the instrument, called instrument signature, confused with light scattered by the sample. An example of instrument signature is shown by the dotted line in Figure 2, which represents scatter from the final mirror. The signature is often measured in the straight through (transmission) direction, multiplied by the measured specular reflectance and then compared to the measured sample BRDF. Another issue is the fact that the measured BRDF is really the convolution of the receiver aperture with the actual (incremental) BRDF. When the scatter signal varies slowly across the aperture the measurement is virtually identical to the true BRDF. Near the specular reflection, or at diffraction peaks, the differences between the measurement (or convolution) and the actual (incremental) BRDF can be huge. Measurements made using invisible sources and measurements of curved samples present additional problems. These problems and the issues of calibration and accuracy are covered in the Further Reading section at the end of this article.

TIS Instruments

The two common methods of making TIS measurements are shown in Figures 4 and 5. The first one is based on a hemispherical mirror (or Coblentz sphere) to gather scattered light from the sample and image it onto the scatter detector. The specular beam enters and leaves the hemisphere through a small circular hole. The diameter of that hole defines the near specular limit of the instrument. The reflected beam (not the incident beam) should be centered in the hole

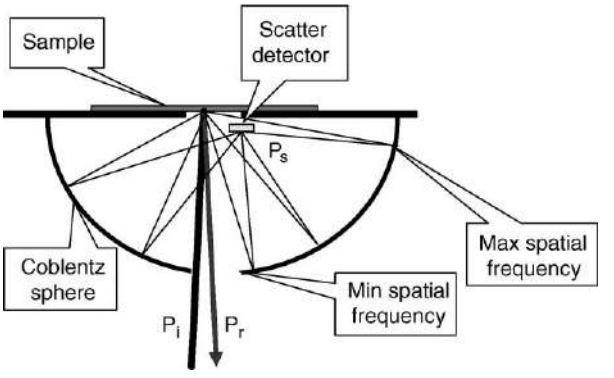

Figure 4 A diagram showing the Coblentz sphere approach to TIS measurements used in the early development of scatter instrumentation.

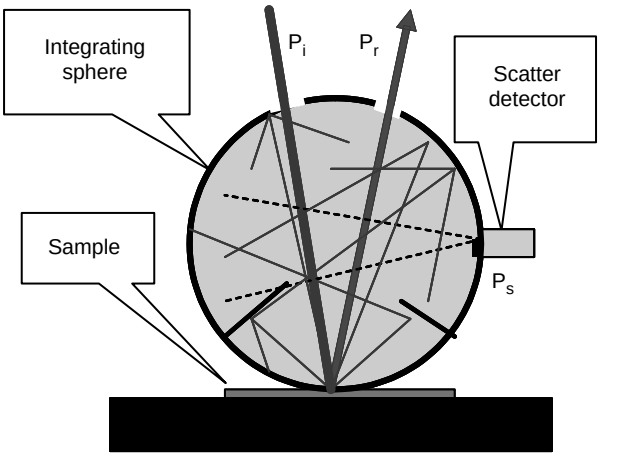

Figure 5 More modern TIS instruments make use of an integrating sphere approach which is easier to align and does not suffer from problems associated with measuring high-angle scatter from the sample.

because the BRDF will be symmetrical about it. Alignment of the hemispherical mirror is critical, and not trivial, in this approach. The second approach involves the use of an integrating sphere. A section of the sphere is viewed by a recessed detector. If the detector field of view (FOV) is limited to a section of the sphere that is not directly illuminated by scatter from the sample, then the signal will be proportional to total scatter from the sample. Again, the reflected beam should be centered on the exit hole. The Coblentz sphere method presents more signal to the detector; however, some of this signal is incident on the detector at very high angles. Thus, this approach tends to discriminate against high-angle scatter (which is not a problem for many samples). The integrating sphere is easier to align, but has a lower signal to noise ratio (less signal on the detector) and is more difficult to build in the IR where uniform diffuse surfaces are harder to obtain. A common mistake with TIS measurements is to assume that for near normal incidence, the orientation between source polarization and sample orientation is not an issue. TIS measurements made with a linearly polarized source on a grating at different orientations will quickly demonstrate this dependence.

TIS measurements can be made very near the specular reflection by utilizing a diffusely reflecting plate with a small hole in it. A converging beam is reflected off the sample and through the hole. Scatter is diffusely reflected from the diffuse plate to a receiver designed to uniformly view the plate. The reflected power is measured by moving the plate so the specular beam misses the hole and then taking that measurement. The ratio of the two scatter measurements gives the TIS. Measurements starting

closer than 0.1 degrees from specular can be made in this manner and it is an excellent way to check incoming optics or freshly coated optics for low scatter.

Analyzing Scatter from Surface Roughness

The preceding sections have concentrated on obtaining and quantifying accurate scatter data, but that leaves the question of what to do with it once you have it. In rare situations you may be given a scatter (BRDF) specification – such as, the BRDF from the mirror must be less than 10^{-4} sr^{-1} 10 degrees from specular when measured at a wavelength of 633 nm incident at 5 degrees with an S polarized source. Unfortunately this is very uncommon. If the issue is limiting scatter as a noise source, you will probably have to generate your own specification based on specific system requirements. More difficult, and often of more economic value, is the situation where scatter measurements are being used as a metrology to learn something about the sample characteristics – like roughness or defect size and/or type.

The relationship between the measured BRDF and reflector roughness statistics was a subject of intense interest from the mid-1970s through the early 1990s. Dozens of papers, and even some books, have been written on the subject, and it can only be outlined here. The relatively easy case of scatter from roughness on a clean, optically smooth, front surface reflector was first published in 1975; however it was several years later before confirming experiments were completed. The deceptively simple relationship, based on vector perturbation theory, is shown below.

$$\mathrm{BRDF} = (16\pi^2/\lambda^4)\cos\theta_i \cos\theta_s\, QS(f_x, f_y) \qquad [4]$$

Q is the polarization factor and is determined by the material constants of the reflector, as well as the system geometry. In many cases, it is numerically about equal to the specular reflectance and this approximation is often justified. Exact expressions are available in the literature. $S(f_x, f_y)$ is the surface power spectral density function (or PSD). It may be thought of as roughness power (surface height variations squared) per unit spatial frequency (undulations per unit distance instead of per unit time). Integration of the PSD over spatial frequency space results in the mean square roughness over that band of frequencies. Taking the square root gives the root mean square roughness (or rms). Frequencies in both the x and y directions on the surface are involved and they are defined by the well-known grating equations as:

$$f_x = (\sin\theta_s \cos\phi_s - \sin\theta_i)/\lambda \qquad \text{and}$$

$$f_y = (\sin\theta_s \sin\phi_s)/\lambda \qquad [5]$$

Thus eqn [4] becomes a model for surface roughness that allows BRDF measurement to be used to find and/or verify surface roughness specifications. Of course, there are exceptions. If the 'clean, optically smooth, front surface reflector' limitations are violated, then the surface will have more than just roughness as a source of scatter and the PSD found from eqn [4] will be too large. Obtaining the same PSD from BRDF measurements made at different wavelengths, or different polarizations, is an indication that the surface is scattering 'topographically' and the PSD can be found using this technique. Because some scatter measurements can be made very rapidly, there are industry situations where scatter metrology offers a very fast means of monitoring surface quality.

A little study of eqns [4] and [5] makes it clear that ranges of spatial frequencies in the PSD correspond directly to angular ranges of the BRDF. Scatter from a single spatial frequency corresponds to a single scatter direction.

Following the above discussion, it becomes clear why the pioneering integrated (TIS) scatter measurements could be used to produce surface rms values as indicated by eqn [2]. Unfortunately, the result of eqn [4] was not available in 1961. Instead eqn [2] was derived for the special case of a smooth surface with Gaussian statistics. Nobody was thinking about spatial bandwidths and angular limits. When striking differences were found between roughness measurements made by TIS and profilometer, the Gaussian assumption became the 'whipping boy,' and TIS scatter measurements took an underserved hit. In fact, integration of eqn [4] results in the TIS result given in eqn [2] under the generally true small angle assumption that most of the scatter is close to the specular reflection. The differences make sense when the concept of spatial frequency bandwidths (or appropriate angular limits) is introduced. It becomes clear that different wavelengths, incident angles and scatter collection angles will also generate different rms values (for the same surface) and why so much confusion resulted following the definition of TIS in terms of rms roughness. There is no such thing as a unique rms roughness for a surface, anymore than there is a single spatial bandwidth for the PSD, or a single set of angles over which to integrate the BRDF. TIS measurements and rms measurements

should always be given with enough information to determine bandwidth limits.

Measuring and Analyzing Scatter from Isolated Surface Features

Understanding scatter from discrete surface features has led to big changes in the entertainment business (CDs, DVDs, digitized music, and films, etc.) as well as providing an important source of metrology for the semiconductor industry as they develop smaller faster chips for a variety of modern uses. Thus, just about everybody in the modern world utilizes our understanding of scatter from discrete surface features.

On the metrology side, the roughness signals described in the last section are a serious source of background noise that limits the size of the smallest defects that can be found. Discrete surface features come in a dazzling array of types, sizes, materials, and shapes – and they all scatter differently. A 100 nm silicon particle scatters a lot differently than a 100 nm silicon oxide particle (even if they have the same shape), and a 100 nm diameter surface pit will have a different scatter pattern. Models describing scatter from a variety of discrete surface features have been developed. Although many of these are kept confidential for competitive reasons, NIST offers some models publicly through a web site.

Confirming a model requires knowing exactly what is scattering the light. In order to accomplish this, depositions of PSLs of known size are made on the surface. Scatter is measured from one or more spheres. A second measurement of background scatter is then subtracted and the net BRDF is converted to DSC units using the known (measured) illuminated spot size. Measurements of this type have been used to confirm discrete feature scatter models. The model is then used to calculate scatter from different diameters and materials. Combined with the model for surface roughness to evaluate system noise, this capability allows signal to noise evaluation of different defect scanners designs.

Practical Industrial Instrumentation

Surface defects and particles, smaller than 100 nm are now routinely found on silicon wafers using scatter instruments known as particle scanners. Thousands of these instruments (with price tags approaching a million dollars each depending on type and use) are in daily use. These instruments are truly amazing; they inspect 200 mm wafers at a rate of one every thirty seconds reporting feature location, approximate size, and in some cases even type (pit or particle) by analyzing scatter signals that last only about 100 nanoseconds. Similar systems are now starting to be used in the computer disk industry and flat panel display inspection of surface features will follow.

Scanners are calibrated with PSLs of different sizes. Because it is impossible to identify all defect types and because diameter has little meaning for irregularly shaped objects, feature 'size' is reported in 'PSL equivalent diameters.' This leads to confusing situations for multiple detector systems, which (without some software help) would report different sizes for real defects that scatter much differently than PSLs. These difficulties have caused industry confusion similar to the 'Gaussian statistics/bandwidth limited' issues encountered before roughness scatter was understood. The publication of international standards relating to scanner calibration has reduced the level of confusion.

Scatter Related Standards

Early scatter related standards for BRDF, TIS, and PSD calculations were written for ASTM (the American Society for Testing Materials), but as the industrial need for these documents moved to the semiconductor industry, there was pressure to move the documents to SEMI (Semiconductor Equipment and Materials Inc.), which is an international organization. By 2004, the process of rewriting the ASTM documents in SEMI format was well underway in the Silicon Wafer Committee. Topics covered include: surface defect specification (M35), defect capture rate (M50), scanner specifications (M52), scanner calibration (M53), particle deposition testing (M58), BRDF measurement (ME1392), and PSD calculation (MF1811). This body of literature is probably the only place where all aspects of these related problems are brought together in one place.

Conclusion

The bottom line is that scatter measurement and analysis has moved from an art to a working metrology. Dozens of labs around the world can now take the same sample and get about the same measured BRDF from it. Thousands of industrial surface scanners employing scattered light signals are in use every day on every continent. In virtually every

house in the modern world there is at least one entertainment device that depends on scatter signals. In short – scatter works.

See also

Scattering: Scattering Theory.

Further Reading

Bennett JM and Mattsson L (1995) *Introduction to Surface Roughness and Scattering*, 2nd edn. Washington, DC: Optical Society of America.

Bennett HE and Porteus JO (1961) Relation between surface roughness and specular reflectance at normal incidence. *Journal of the Optical Society of America* 51: 123.

Cady FM, Bjork DR, Rifkin J and Stover JC (1989) BRDF error analysis. *Proceedings of SPIE* 1165: 154–164.

Church EL and Zavada JM (1975) Residual surface roughness of diamond-turned optics. *Applied Optics* 14: 1788.

Doicu A, Eremin Y and Wriedt T (2000) *Acoustic & Electromagnetic Scattering Analysis Using Discrete Sources*. San Diego, CA: Academic Press.

Gu ZH, Dummer RS, Maradudin AA and McGurn AR (1989) Experimental study of the opposition effect in the scattering of light from a randomly rough metal surface. *Applied Optics* 28(3): 537.

Nicodemus FE, Richmond JC, Hsia JJ, Ginsberg IW and Limperis T (1977) *Geometric Considerations and Nomenclature for Reflectance*. NBS Monograph 160, US Dept. of Commerce.

Scheer CA, Stover JC and Ivakhnenko VI (1998) Comparison of models and measurements of scatter from surface-bound particles. *SPIE Proceedings* 3275.

Schiff TF, Stover JC and Bjork DR (1992) Mueller matrix measurements of scattered light. *Proceedings of SPIE* 1753: 269–277.

Stover JC (1975) Roughness characterization of smooth machined surfaces by light scattering. *Applied Optics* 14(8): 1796.

Stover JC (1995) *Optical Scattering: Measurement and Analysis*, 2nd edn. Bellingham, WA: SPIE.

Stover JC (2001) Calibration of particle detection systems. In: Alain Diebold (ed.) *Handbook of Silicon Semiconductor Metrology*, chap. 21. New York: Marcel Dekker.

Wolfe WL and Bartell FO (1982) Description and limitations of an automated scatterometer. *Proceedings of SPIE* 362-30.

Spectrometers

K A More, SenSyTech, Inc. Imaging Group, Ann Arbor, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Sp7ectrometers were developed after the discovery that glass prisms disperse light. Later, it was discovered that diffraction from multiple, equally spaced, wires or fibers also dispersed light. About 100 years ago, Huygens proposed his wave theory of light and Fraunhofer developed diffraction theory, which allowed scientific development of diffraction gratings.

These discoveries then led to the development of spectrometers. Light theory was sufficiently developed such that spectrometer designs, developed over 100 years ago, are still being used today. These theories and designs are briefly described along with comments on how current technology has improved upon these designs. This is followed by some examples of imaging spectrometers which have wide spectral coverage from 450 nm to 14 μm and produce images of more than 256 spectral bands.

The basic elements of a spectroscopic instrument are shown in Figure 1. The source, or more usually an image of the source, fills an entrance slit and the radiation is collimated by either a lens or mirror. The radiation is then dispersed, by either a prism or a grating, so that the direction of propagation of the radiation depends upon its wavelength. It is then brought to a focus by a second lens or mirror and the spectrum consists of a series of monochromatic images of the entrance slit. The focused radiation is detected, either by an image detector such as a photographic plate, or by a flux detector such as a photomultiplier, in which case the area over which the flux is detected is limited by an exit slit. In some cases the radiation is not detected at this stage, but passes through the exit slit to be used in some other optical system. As the exit slit behaves as a monochromatic source, the instrument can be regarded as a wavelength filter and is then referred to as a monochromator.

Prisms

The wavelength dependence of the index of refraction is used in prism spectrometers. Such an optical

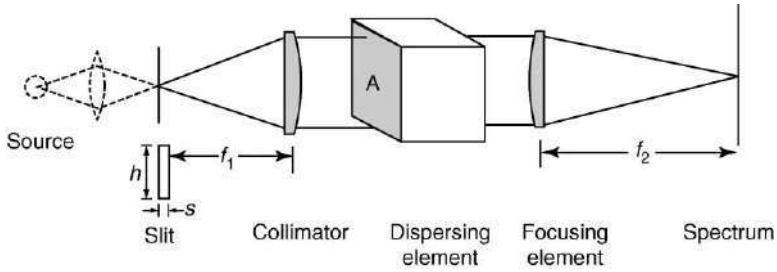

Figure 1 The basic elements of a spectroscopic instrument. With permission from Hutley MC (1982) *Diffraction Gratings*, pp. 57–232. London: Elsevier.

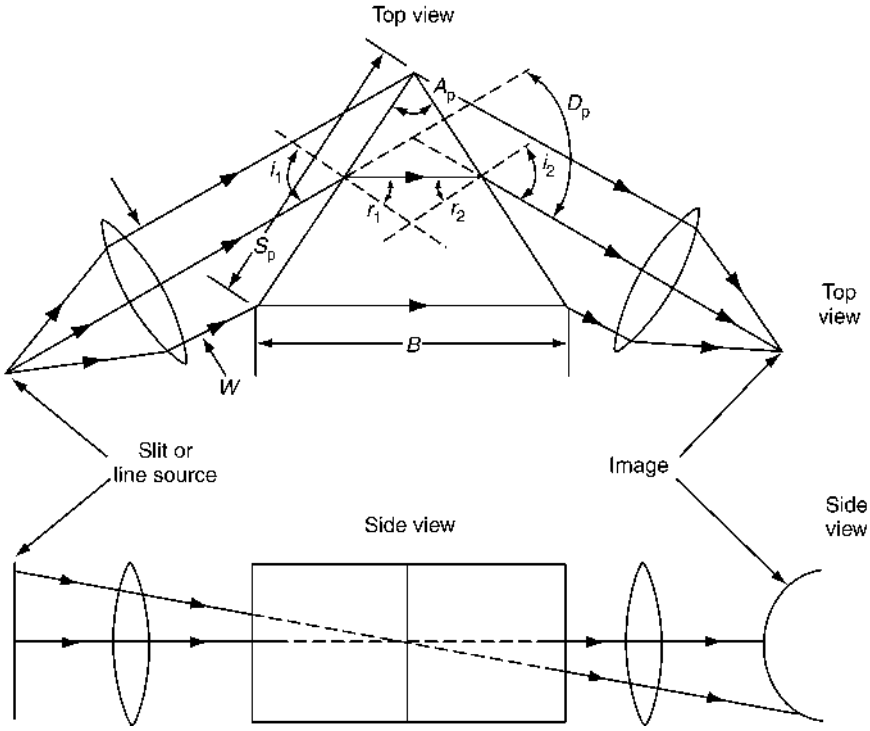

Figure 2 Elementary prism spectrometer schematic. W is the width of the entrance beam; S_p is the length of the prism face; and B is the prism base length. Reproduced with permission from *The Infrared Handbook* (1985). Ann Arbor, MI: Infrared Information Analysis Center.

element disperses parallel rays or collimated radiation into different angles from the prism according to wavelength. Distortion of the image of the entrance slit is minimized by the use of planewave illumination. Even with planewave illumination, the image of the slit is curved because not all of the rays from the entrance slit can traverse the prism in its principal plane. The prism is shown in the position of minimum angular deviation of the incoming rays in Figure 2. At minimum angular deviation, maximum power can pass through the prism. For a prism adjusted to the position of minimum deviation:

$$r_1 = r_2 = A_p/2 \qquad [1]$$

and

$$i_1 = i_2 = (D_p + A_p)/2 \qquad [2]$$

where

D_p = angle of deviation
A_p = angle of prism
r_1 and r_2 = internal angles of refraction
i_1 and i_2 = angles of entry and exit.

The angle of deviation, D_p, varies with wavelength. The resulting angular dispersion is defined as $dD_p/d\lambda$, while the linear dispersion is $dx/d\lambda = F\, dD_p/d\lambda$, where F is the focal length of the camera or imaging lens and x is the distance across the image plane.

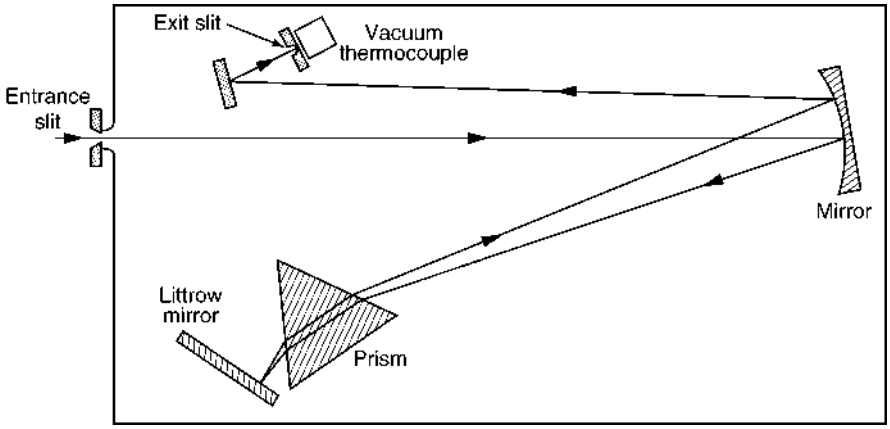

Figure 3 Infrared spectrograph of the Littrow-type mount with a rock salt prism. Reproduced with permission from *The Infrared Handbook* (1985) Ann Arbor, MI: Infrared Information Analysis Center.

It can be shown that:

$$dD_p/d\lambda = [B/W][dn/d\lambda] = [dD_p/dn][dn/d\lambda] \quad [3]$$

where

B = base length of the prism,
W = width of the illumination beam
n = index of refraction

while

$$dx/d\lambda = F[B/W][dn/d\lambda] \quad [4]$$

One may define the resolving power, RP, of an instrument as the smallest resolvable wavelength difference, according to the Rayleigh criterion, divided into the average wavelength in that spectral region. Thus:

$$RP = \lambda/\Delta\lambda = [\lambda/dD_p][dD_p/d\lambda]$$

$$= [\lambda/dD_p][B/W][dn/d\lambda] \quad [5]$$

The limiting resolution is set by diffraction due to the finite beamwidth, or effective aperture of the prism, which is rectangular. Thus:

$$RP = [\lambda/(\lambda/W)][B/W][dn/d\lambda] \quad [6]$$

so that:

$$RP = B[dn/d\lambda] \quad [7]$$

If the entire prism face is not illuminated, then only the illuminated base length must be used for B.

Littrow showed that aberrations would be minimized by making the angle of incidence equal to the angle of refraction β (also known as the Littrow Configuration). Littrow used a plane mirror behind the prism for a double pass through the prism, as shown in Figure 3.

Gratings

Rowland is credited with the development of a grating mount that reduced aberrations in the spectrogram. He found that a grating ruled on a concave surface of radius R and locating the entrance and exit slits on the same radius would give the least aberrations, as shown in Figure 4. $r = R\cos\alpha$ is the distance to the entrance slit *(r)* and $r_1 = R\cos\beta$ is the distance to the focal point for the exit slit (r_1). Rowland showed that:

$$\cos\alpha/R - \cos^2\alpha/r + \cos\beta/R - \cos^2\beta/r_1 = 0 \quad [8]$$

One solution to this equation is for α and β to each be zero and

$$r = R\cos\alpha$$

and

$$r_1 = R\cos\beta$$

This condition is met if r and r_1 lie on the Rowland circle.

There are various ways in which the Rowland circle condition may be satisfied and some of them are shown in Figure 5. The simplest mounting of all is that due to Paschen and Runge (Figure 5a) in which the entrance slit is positioned on the Rowland circle and a photographic plate (or plates) is constrained to fit the Rowland circle. Alternatively, for photoelectric detection, a series of exit slits is arranged around the Rowland circle each having its own detector. In the latter case, the whole spectrum is not recorded, only a series of predetermined wavelengths, but when used in this 'polychromator' form, it is very rugged and

convenient for applications such as the routine analysis of samples of metals and alloys. In this case, slits and detectors are set up to measure the light in various spectral lines, each characteristic of a particular component or trace element.

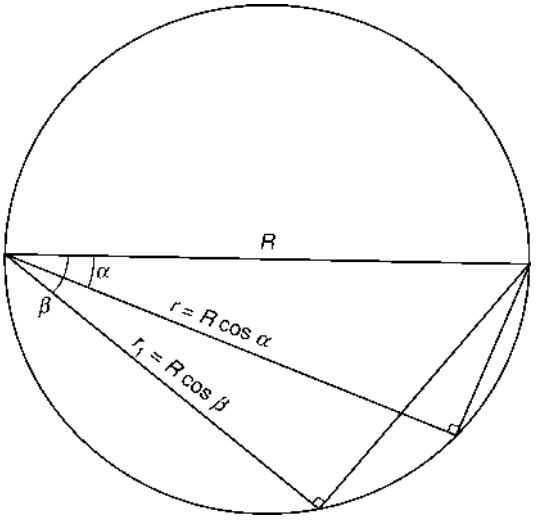

Figure 4 The construction of the Rowland Circle. With permission from Hutley MC (1982) *Diffraction Gratings*, pp. 57–232. London: Elsevier.

The oldest concave grating mount is that designed by Rowland himself and which bears his name (Figure 5b). In this case, the grating and photographic plates are fixed at opposite ends of the diameter of the Rowland circle by a moveable rigid beam. The entrance slit remains fixed above the intersection of two rails at right angles to each other and along which the grating plate holder (or the exit slit) is free to move. In this way, this entrance slit, grating, and plate holder are constrained always to lie on the Rowland circle and it has the advantage that the dispersion is linear, which is useful in the accurate determination of wavelengths. Unfortunately, this mounting is rather sensitive to small errors in the position of the entrance slit and in the orthogonality of the rails, and is now very rarely used.

A variation on this mounting was devised by Abney, who again mounted the grating and the plate holder on a rigid bar at opposite ends of the diameter of the Rowland circle (Figure 5c). The entrance slit is mounted on a bar of length equal to the radius of the Rowland circle. In this way, the slit always lays on the Rowland circle, but it had to rotate about its axis in order that the jaws should remain perpendicular to the line from the grating to

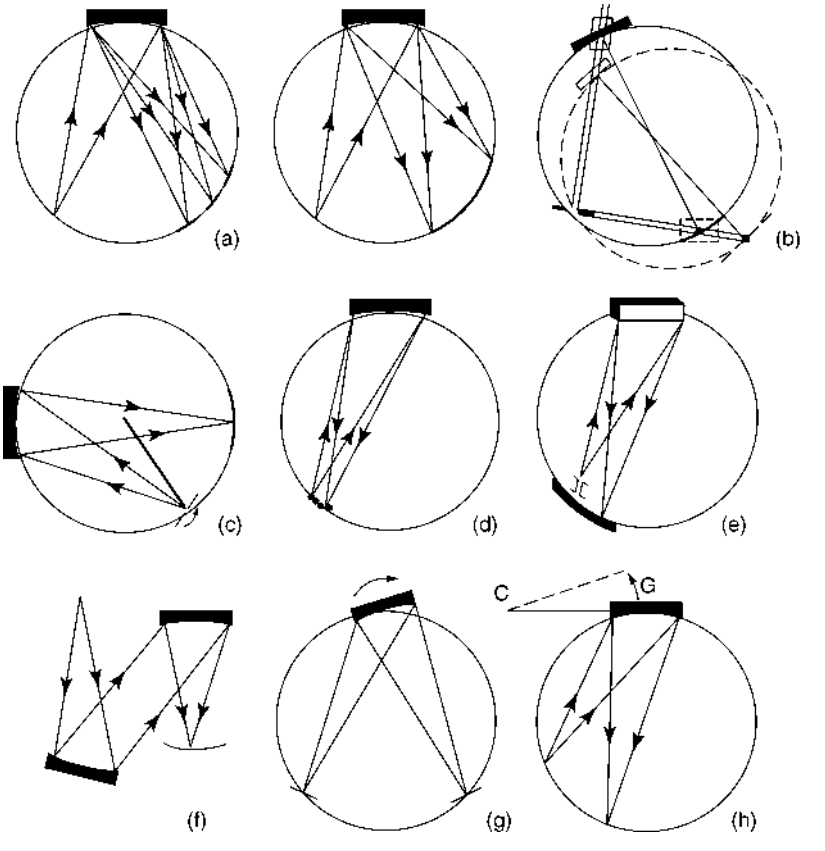

Figure 5 Mountings of the concave grating: (a) Paschen–Runge; (b) Rowland; (c) Abney; (d) and (e) Eagle; (f) Wadsworth; (g) Seya–Namioka; (h) Johnson–Onaka. With permission from Hutley MC (1982) *Diffraction Gratings*, pp. 57–232. London: Elsevier.

the slit. It also had the disadvantage that the source must move with the exit slit, which could be inconvenient.

In the Eagle mounting (Figures 5d and 5e), the angles of incidence and diffraction are made equal, or very nearly so, as with Littrow mounting for plane gratings. Optically, this system has the advantage that the astigmatism is generally less than that of the Paschen–Runge or Rowland mounting, but on the other hand, the dispersion is nonlinear. From a mechanical point of view, it has the disadvantage that it is necessary with great precision both to rotate the grating and to move it nearer to the slits in order to scan the spectrum. However, it does have the practical advantage that it is much more compact than other mountings, and this is of particular importance when we bear in mind the need to enclose the instrument in a vacuum tank. Ideally, the entrance slit and exit slit or photographic plate should be superimposed if we are to set $\alpha = \beta$. In practice, of course, the two are displaced either sideways, in the plane of incidence as shown in Figure 5, or out of the plane of incidence, in which case the entrance slit is positioned below the meridonal plane and the plate holder just above it, as shown in Figure 5. The out-of-plane configuration is generally used for spectrographs and the in-plane system for monochromators. The penalty incurred in going out of plane is that coma is introduced in the image, and slit curvature becomes more important. This limits the length of the ruling that can effectively be used.

The second well-known solution to the Rowland equation is the Wadsworth mounting (Figure 5f), in which the incident light is collimated, so r is set at infinity and the focal equation reduces to $r_1 = R\cos^2\beta/(\cos\alpha + \cos\beta)$. One feature of this mounting is that the astigmatism is zero when the image is formed at the center of the grating blank, i.e., when $\beta = 0$. It is a particularly useful mounting for applications in which the incident light is naturally collimated (for example, in rocket or satellite astronomy, spectroheliography and in work using synchrotron radiation). However, if the light is not naturally collimated, the Wadsworth mount requires a collimating mirror, so one has to pay the penalty of the extra losses of light at this mirror. The distance from the grating to the image is about half that for a Rowland circle mounting which makes the instrument more compact and, since the grating subtends approximately four times the solid angle, there is a corresponding increase in the brightness of the spectral image.

Not all concave grating mountings are solutions to the Rowland equation. In some cases, other advantages may compensate for a certain defect of focus. A particularly important example of this is the Seya–Namioka mounting (Figure 5g), in which the entrance slit and exit slit are kept fixed and the spectrum is scanned by a simple rotation of the grating (Figure 5g). In order to achieve the optimum conditions for this mounting, we set $\alpha = \varphi + \theta$ and $\beta = \theta - \varphi$, where 2φ is the angle subtended at the grating by the entrance and exit slit and θ is the angle through which the grating is turned. The amount of defocus is given by:

$$F(\theta, \varphi, r, r_1) = [\cos^2(\theta + \varphi)/r] + [\cos(\theta + \varphi)/R] + [\cos^2(\theta - \varphi)/r_1] + [\cos(\theta - \varphi)/R] \quad [9]$$

and the optimum conditions are those for which $F(\theta,\varphi,r,r_1)$ remains as small as possible as θ is varied over the required range. Seya set F and three derivatives of F with respect to θ equal to zero for $\theta = 0$ and obtained the result:

$$\varphi = \sin^{-1}(1/\sqrt{3}) = 35°15' \quad [10]$$

and

$$r = r_1 = R\cos\varphi \quad [11]$$

which corresponds to the Rowland circle either in zero order or for zero wavelength. In practice, it is usual to modify the angle slightly so that the best focus is achieved in the center of the range of interest rather than at zero wavelength.

The great advantage of the Seya–Namioka mounting is its simplicity. An instrument need consist only of a fixed entrance and exit slit, and a simple rotation of the grating is all that is required to scan the spectrum. It is, in fact, simpler than instruments using plane gratings. Despite the fact that at the ends of the useful wavelength range the resolution is limited by the defect of focus and the astigmatism is particularly bad, the Seya–Namioka mounting is very widely used, particularly for medium-resolution rather than high-resolution work.

A similar simplicity is a feature of the Johnson–Onaka mounting (Figure 5h). Here again, the entrance and exit slits remain fixed, but the grating is rotated about an axis which is displaced from its center; in this way, it is possible to reduce the change of focus that occurs in the Seya–Namioka mounting. The system is set up so that at the center of the desired wavelength range, the slits and grating lie on the Rowland circle, as shown in Figure 5h. The optimum radius of rotation, i.e., the distance GC, was found by

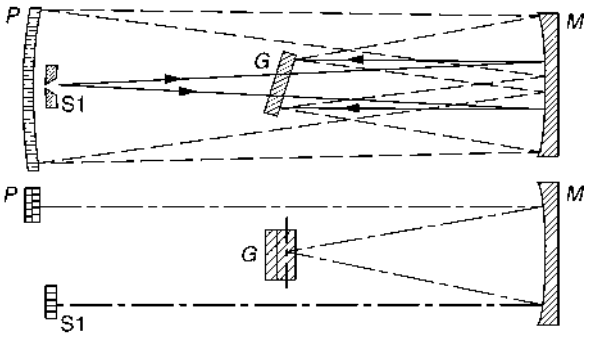

Figure 6 Ebert mounting of the plane grating designed by Fastie. 'SI' is the entrance slit; *G* is the grating; *M* is the concave mirror; and *P* is the photographic plate. The horizontal section is at the top and the vertical section is at the bottom. Reproduced with permission from *The Infrared Handbook* (1985) Ann Arbor, MI: Infrared Information Analysis Center.

Onaka to be:

$$GC_{opt} = [R \sin\tfrac{1}{2}(\alpha+\beta)]/[1 - \tfrac{1}{2}\tan(\alpha+\beta) \times (\tan\beta - \tan\alpha)] \quad [12]$$

Another non-Rowland spectrometer is the Ebert–Fastie mount, which mounts the slit, flat grating, and a detector array in the arrangement shown in Figure 6. Ebert first developed the design using two separate concave mirrors, one for collimating the incident beam and the second to focus the diffracted spectrum. Fastie used a single but larger concave mirror, which simplified the mounting structure and produced a rugged, compact spectrometer that has been used in rocket flights and space satellite observatories for astronomical and upper atmospheric applications. The Czerny–Turner mount is similar to the Ebert mount, except the flat grating is located in the same plane that contains the entrance and exit slits.

Advanced Spectrometers

While the above spectrometer designs are still used, major advances in implementation are now available. Ray tracing allows the designer to quantify the aberrations and determine solutions to remove them. Aspheric optic elements can now be fabricated to correct aberrations. Gratings can be ruled on aspherical surfaces to not only disperse the light beam, but also be an element in the optical design. Holography has been developed to etch gratings for use over a wide spectral range and to not only disperse the light, but also work as an optical element. Because holographic gratings are chemically etched, there are no machine burrs to scatter light and the hologram is free of periodic differences in groove widths that create ghosts in ruled gratings. Linear and array photodetectors have replaced film. An advantage of film was its ability to fit curved focal planes. Photodetectors are etched into flat wafers of the photodiode material. To adapt flat arrays to a curved focal plane, fiber optic face plate couplers have been ground on one side to match the focal plane curvature and the flat back side is either optically coupled to the detector or closely coupled for proximity focusing. The photodetector arrays are available in multiple materials to cover the spectral range, from soft X-rays to the thermal infrared. Often cooling is required to obtain low noise. They are now available with sensitivities never reached with film and the ease of coupling the output data to a computer for real-time analysis has taken much of the labor out of analyzing the spectrograph. Another advance is precision motor drives with precision motion sensors, some using laser interferometry, to provide feedback on automated movement of gratings, slits, and detectors as a spectrometer cycles through its wavelength range.

Imaging Spectrometers

Spectrometers that image the spectral characteristics in each pixel of the image to form a data cube, as shown in Figure 7, are named imaging spectrometers, and if the spectra have high resolution and blocks of consecutive neighbors, then the data are called hyperspectral. Imaging spectrometers have several applications that range from medical to remote sensing of land use and environment. This article will cover examples for remote sensing imaging spectrometers which require spectral coverage in all of the atmospheric windows from the UV to thermal IR.

Imaging spectrometers require a combination of spectrometers, light collecting optics, and scan mechanisms to scan the instantaneous field of view of the spectrometer over a scene. Remote sensing of the Earth applications requires an aerial platform. Either a helicopter, an aircraft, or orbital space satellites are used. The platform motion is used as part of the scanning process so that the optics image a single point on the ground so that a scanner (called a line scanner) scans a long line that is cross tracked to the platform motion. Or, the optics image a slit that is parallel to the platform track that covers many scan lines and a scanner moves the slit cross track to the platform motion. This scanner is called a whisk-broom scanner. Or, the optics image a large slit so no scan mechanism is needed other than the platform

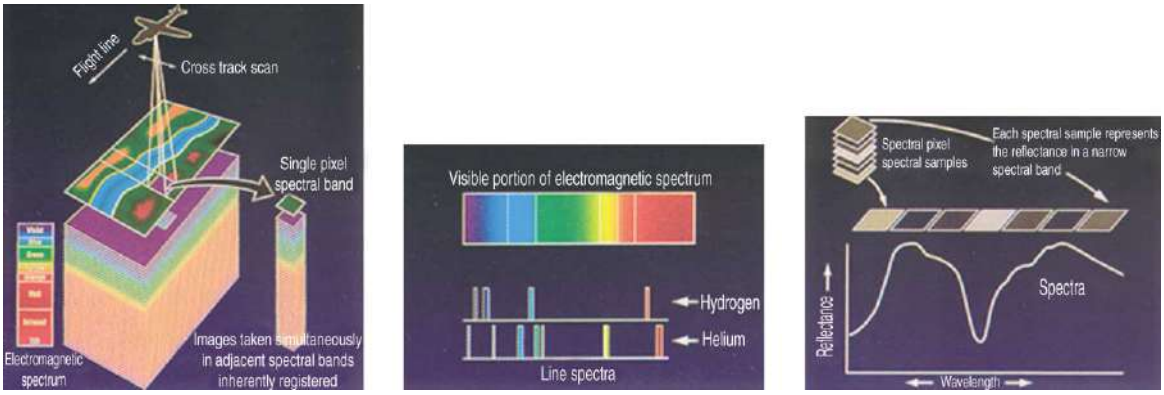

Figure 7 Hyperspectral Data Cube. Hyperspectral imagers divide the spectrum into many discrete narrow channels. This fine quantization ofn spectral information on a pixel by pixel basis enables researchers to discriminate the individual constituents in an area much more effectively. For example, the broad spectral bands of a multispectral sensor allow the user only to coarsely discriminate between areas of deciduous and coniferous forest, plowed fields, etc., whereas a hyperspectral imager provides characteristic signatures which can be correlated with specific spectral templates to help determine the individual constituents and possibly even reveal details of the natural processes which are affecting them.

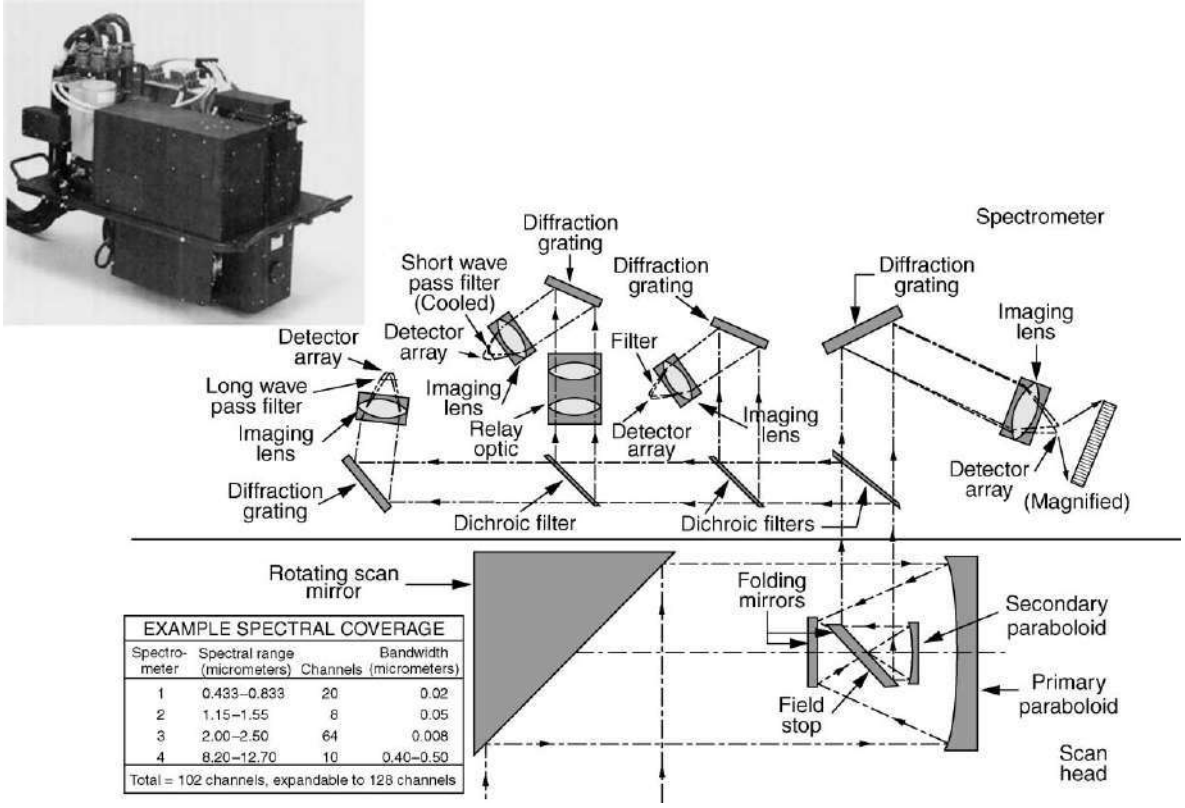

EXAMPLE SPECTRAL COVERAGE

Spectrometer	Spectral range (micrometers)	Channels	Bandwidth (micrometers)
1	0.433–0.833	20	0.02
2	1.15–1.55	8	0.05
3	2.00–2.50	64	0.008
4	8.20–12.70	10	0.40–0.50
Total = 102 channels, expandable to 128 channels			

Figure 8 Multispectral infrared and visible imaging spectrometer optical schematic.

motion to form an image. This scanner is called a pushbroom scanner.

One important requirement is that all spectral measurements of a pixel be coregistered. Most airborne imaging spectrometers use a common aperture for a line scanner, or a slit for whisk and pushbroom scanners, so that platform instability from pitch roll and yaw and from inability to fly in a straight line do not compromise the coregistration of spectral data on each pixel. The image data may require geometric correction, but the spectral data are not compromised.

There is one other type of imaging spectrometer that uses a linear variable filter (also known as a wedge filter) over a 2D array of photodetectors. Each image contains a different spectral band over each row of pixels so that each frame images a scene with a different spectral band over each row of pixels. The array is oriented so the rows of spectral bands are perpendicular to the flight track. After the platform motion moves over one ground pixel, the whole array is read out and the frame is shifted one row of pixels so the second frame adds a second spectral band to each row imaged in the first frame. This frame stepping is carefully timed to the platform velocity and is repeated until each row of pixels is imaged in all spectral bands. This type of imaging spectrometer has been flown in aircraft and satellites. In aircraft, the platform motion corrupts the coregistration of the spectrum in each pixel. Extensive ground processing is required to geometrically correct each frame to improve spectral coregistration. This is a difficult task and this type of imaging spectrometer has lost favor for airborne use. Stabilized satellites have been a better platform and the wedge type of imaging spectrometer has been used successfully in space.

As examples of airborne and ground-based imaging spectrometers, a line scanner imaging

Table 1 MIVIS physical properties

	Height		*Width*		*Depth*[a]		*Weight*	
	in	*cm*	*in*	*cm*	*in*	*cm*	*lbs*	*kg*
Scan head	26.5	67.0	20.6	52.0	28.1	71.0	220	100
Electronics	40.0	102.0	19.0	48.3	24.0	61.0		
Total system weight (approx.)							460	209

[a]Not including connectors or cable bends.

Table 2 MIVIS spectral coverage (μm)

OPTICAL PORT 1			*OPTICAL PORT 3*						*OPTICAL PORT 4*		
CH #	*Band Edges*		*CH #*	*Band Edges*		*CH #*	*Band Edges*		*CH #*	*Band Edges*	
	Lower	*Upper*		*Lower*	*Upper*		*Lower*	*Upper*		*Lower*	*Upper*
1	0.43	0.45	29	2.000	2.008	61	2.250	2.258	93	8.20	8.60
2	0.45	0.47	30	2.008	2.016	62	2.258	2.266	94	8.60	9.00
3	0.47	0.49	31	2.016	2.023	63	2.266	2.273	95	9.00	9.40
4	0.49	0.51	32	2.023	2.031	64	2.273	2.281	96	9.40	9.80
5	0.51	0.53	33	2.031	2.039	65	2.281	2.289	97	9.80	10.20
6	0.53	0.55	34	2.039	2.047	66	2.289	2.297	98	10.20	10.70
7	0.55	0.57	35	2.047	2.055	67	2.297	2.305	99	10.70	11.20
8	0.57	0.59	36	2.055	2.063	68	2.305	2.313	100	11.20	11.70
9	0.59	0.61	37	2.063	2.070	69	2.313	2.320	101	11.70	12.20
10	0.61	0.63	38	2.070	2.078	70	2.320	2.328	102	12.20	12.70
11	0.63	0.65	39	2.078	2.086	71	2.328	2.336			
12	0.65	0.67	40	2.086	2.094	72	2.336	2.344			
13	0.67	0.69	41	2.094	2.102	73	2.344	2.352			
14	0.69	0.71	42	2.102	2.109	74	2.352	2.359			
15	0.71	0.73	43	2.109	2.117	75	2.359	2.367			
16	0.73	0.75	44	2.117	2.125	76	2.367	2.375			
17	0.75	0.77	45	2.125	2.133	77	2.375	2.383			
18	0.77	0.79	46	2.133	2.141	78	2.383	2.391			
19	0.79	0.81	47	2.141	2.148	79	2.391	2.398			
20	0.81	0.83	48	2.148	2.156	80	2.398	2.406			
OPTICAL PORT 2			49	2.156	2.164	81	2.406	2.414			
21	1.15	1.20	50	2.164	2.172	82	2.414	2.422			
22	1.20	1.25	51	2.172	2.180	83	2.422	2.430			
23	1.25	1.30	52	2.180	2.188	84	2.430	2.438			
24	1.30	1.35	53	2.188	2.195	85	2.438	2.445			
25	1.35	1.40	54	2.195	2.203	86	2.445	2.453			
26	1.40	1.45	55	2.203	2.211	87	2.453	2.461			
27	1.45	1.50	56	2.211	2.219	88	2.461	2.469			
28	1.50	1.55	57	2.219	2.227	89	2.469	2.477			
			58	2.227	2.234	90	2.477	2.484			
			59	2.234	2.242	91	2.484	2.492			
			60	2.242	2.250	92	2.492	2.500			

spectrometer and a ground-based scanned push broom scanner are described below.

MIVIS (multispectral infrared and visible imaging spectrometers) hyperspectral scanner developed by SenSyTech Imaging Group (formerly Daedalus Enterprises) for the CNR (Consiglio National Researche) of Italy is a line scanner imaging spectrometer. The optical schematic is shown in Figure 8. A rotating 45° mirror scans a line of pixels on the ground. Each pixel is collimated by a parabolic mirror in a Gregorian telescope mount. The collimated beam is reflected by a Pfund assembly to an optical bench that houses four spectrometers. An aperture pinhole in the Pfund assembly defines a common instantaneous field of view for each pixel. The collimated beam is then split off with either thin metallic coated or dielectric coated dichroics to four spectrometers. The thin metallic mirrors reflect long wavelengths and transmit short wavelengths. The multiple dielectric layers cause interference such that light is reflected at short wavelengths and transmitted at long wavelengths. After splitting off four wide bands (visible, near infrared, mid-wave infrared, and thermal infrared) each band is dispersed in its own spectrometers.

The wavelength coverage for each spectrometer is based on the wavelength sensitivity of different photodetector arrays. The visible spectrometer uses a silicon photodiode array; the near infrared spectrometer, an InGaAs array, the mid-infrared a InSb array and the thermal infrared, a MCT (mercury doped cadmium teluride) photo-conductor array. Note the beam expander in spectrometer 3 for the

Table 3 LAFS technical specifications

Spatial resolution	1.0 mrad IFOV vertical and horizontal (square pixels). Optical lens for 0.5 mrad IFOV
Spatial coverage	15° TFOV horizontal, nominal. A second, 7.5° TFOV, is an option
Focus range	5′ to infinity
Spectral resolution	5.0 nm per spectral channel, sampled at 2.5 nm interval
Spectral range	400–1100 nm
Dynamic range	8 bit (1 part in 256)
Illumination	Solar illumination ranging from 10:00 a.m. to 2:00 p.m. under overcast conditions to full noon sunshine
Acquisition time	1.5 sec for standard illuminations. Option for longer times for low light level conditions
Viewfinder	Near real time video view of the scene
Calibrations	Flat fielding to compensate for CCD variations in responsivity. Special calibration
Data displays	Single band imagery, selectable Waterfall Chart (one spatial line by spectral, as acquired by the CCD). Single pixel spectral display
Data format	Convertible to format compatible with image processors
Data storage	Replaceable hard disc, 14 data cubes/disk

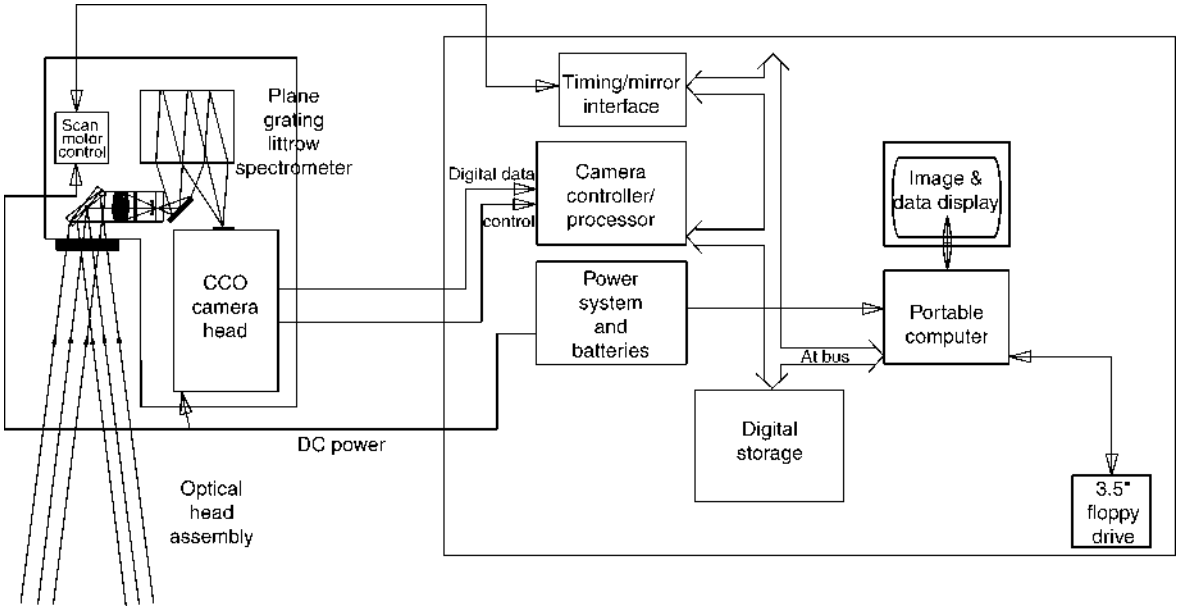

Figure 9 LAFS system block diagram.

mid-infrared (IR). Since wavelength resolution increases with the size of the illuminated area of a diffraction grating a beam expander was needed to meet the specified mid-IR resolution.

The physical characteristics of MIVIS are given in Table 1 and the spectral coverage in Table 2.

An example of a push broom imaging spectrometer is the large area fast spectrometer (LAFS), which is a field portable, tripod mounted, imaging spectrometer. LAFS uses a push broom spectrometer with a galvanometer driven mirror in front of the entrance slit. The mirror is stepped to acquire a 2D image $\times$ 256 spectral bands data cube.

LAFS was developed for the US Marine Corps under the direction of the US Navy Coastal Systems Station of the Dahlgren Division. The technical specifications are shown in Table 3. Various concepts for the imaging spectrometer were studied and the only concept that could meet the above specification was a Littrow configuration grating spectrometer that imaged one line in the scene and dispersed the spectra perpendicular to the line image onto a CCD array. A galvanometer mirror stepped the line image over the scene to generate a spectral image of the scene. A block diagram of LAFS is shown in Figure 9 and a photograph of the prototype in Figure 10. Figure 11 is a photograph of the optical head. Data from a single frame are stored in computer (RAM) memory then transferred to a replaceable hard disk for bulk storage. Data are collected in a series of 256 CCD frames and each CCD frame has a line of 256 pixels along one axis and 256 spectral samples along the second axis. The 256 CCD frames constitute a data cube as shown in Figure 12. This does not allow field viewing of an image in a single spectral band. Earth View™ software is used in the field portable computer

Figure 10 LAFS.

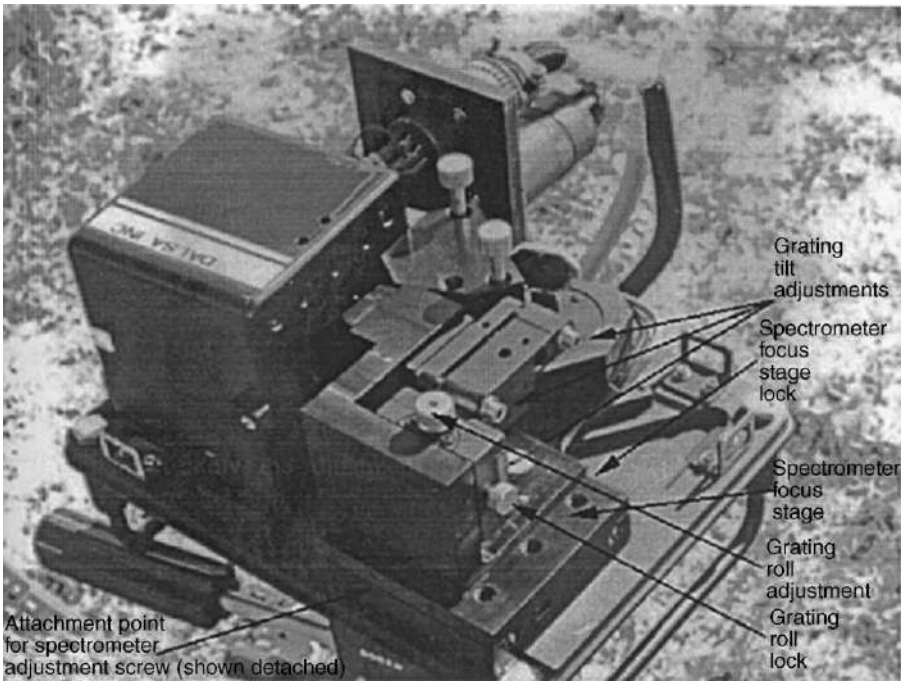

Figure 11 LAFS optical head.

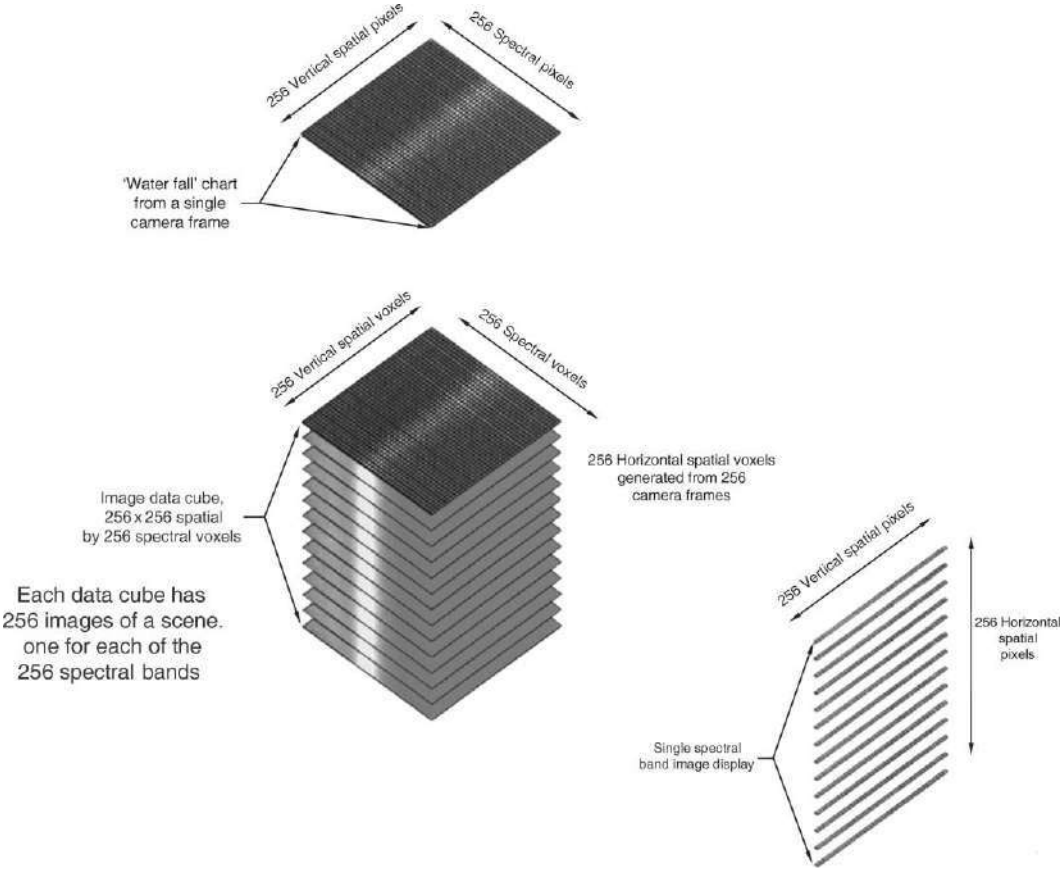

Figure 12 Arrangement of image memory and construction of image displays.

Table 4 LAFS physical properties

Dimensions and weight	*Size (inches)*	*weight (lbs)*
Optical head	7 × 9 × 8	11
Electronics	12 × 15.5 × 9	30
Battery	4.5 × 16.5 × 2	7
		48 + cables

Power: Less than 150 W at 12 VDC; Battery time – 45 min/charge.

to reorder the data cube into 256 spatial images, one for each spectral band as shown in Figure 12.

The prototype is packaged in two chassis, a compact optical head and a portable electronic chassis. The size, weight, and power are shown in Table 4.

LAFS was designed to also work as an airborne imaging spectrometer. The framing mirror can be locked so it views a scene directly below the aircraft (push broom mode) or the mirror can be programmed to sweep a 256 lines cross track to the aircraft flight path (whisk broom mode). In the whisk broom mode, the scan mirror rotation arc can be increased for a wider field of view than in the push broom mode. LAFS illustrates the decrease in size, weight, and power of a push broom imaging spectrometer that results from a 256 × 256 pixel array rather than single pixel line arrays of a line scanner. The array increases integration time of a line by a factor of 256, which allows longer detector integration dwell time on each pixel and thus smaller light collecting optics.

List of Units and Nomenclature

Data cube	Multiple images of a scene in many spectral bands. The spectra of each pixel can be obtained by plotting the spectral value on the same pixel in each spectral band image.
Pixel	An image point (or small area) defined by the instantaneous-field-of-view of either the optics, entrance aperture or slit, and the detector area, or by some combination of these components. A pixel is the smallest element in a scene that is resolved by the imaging system.
Instantaneous-field-of-view (IFOV)	The IFOV is used in describing scanners to define the size of the smallest field-of-view (usually in milliradians or microradians) that be resolved by a scanner system.
Field-of-view	The total size in angular dimensions of a scanner or imager.
Hyperspectral	A data cube with many (>48) spectral bands.
Line scanners	An optical/mechanical system that scans one pixel at a time along a line of a scene.
Whisk broom	An optical/mechanical system that scans two or more lines at a time.
Push broom	An optical system that images a complete line at a time.

See also

Diffraction: Diffraction Gratings; Fraunhofer Diffraction. **Fiber Gratings**. **Geometrical Optics:** Prisms. **Imaging:** Hyperspectral Imaging; Interferometric Imaging. **Interferometry:** Overview. **Modulators:** Acousto-Optics. **Optical Materials:** Color Filters and Absorption Glasses. **Semiconductor Materials:** Dilute Magnetic Semiconductors; Group IV Semiconductors, Si/SiGe; Large Gap II–VI Semiconductors; Modulation Spectroscopy of Semiconductors and Semiconductor Microstructures. **Spectroscopy:** Fourier Transform Spectroscopy; Raman Spectroscopy.

Further Reading

Baselow R, Silvergate P, Rappaport W, *et al.* (1992) *HYDICE (Hyperspectral Digital Imagery Collection Experiment Instrument Design*. International Symposium on Spectral Sensing Research.

Bianchi, R, Marino, CM and Pignatti, S *Airborne Hyperspectral Remote Sensing in Italy*. Pomezia (Roma), Italy: CNR, Progetto LARA, 00040.

Borough HC, Batterfield OD, Chase RP, Turner RW and Honey FR (1992) *A Modular Wide Field of View Airborne Imaging Spectrometer Preliminary Design Concept*. International Symposium on Spectral Sensing Research.

Carmer DC, Horvath R and Arnold CB (1992) *M7 Multispectral Sensor Testbed Enhancements*. International Symposium on Spectral Sensing Research.

Hackwell JA, Warren DW, Bongiovi RP, *et al.* (1996) LWIR/MWIR Imaging Hyperspectral Sensor for Airborne and Ground-Based Remote Sensing. In: Descour M and Mooney J (eds) *Imaging Spectrometry II*. Proc. SPIE 2819, pp. 102–107.

Hutley MC (1982) *Diffraction Gratings*. London: Academic Press.

James JF and Sternberg RS (1969) *The Design of Optical Spectrometers*. London: Chapman and Hall.

Kulgein NG, Richard SP, Rudolf WP and Winter EM (1992) *Airborne Chemical Vapor Detection Experiment*. International Symposium on Spectral Sensing Research.

Lucey PG, Williams T, Horton K, Hinck K and Buchney C (1994) *SMIFTS: A Cryogenically Cooled, Spatially Modulated, Imaging Fourier Transform Spectrometer for Remote Sensing Applications*. International Symposium on Spectral Sensing Research.

Mourales P and Thomas D (1998) Compact Low-Distortion Imaging Spectrometer for Remote Sensing. *SPIE Conference on Imaging Spectrometry IV* 3438: 31–37.

Pagano TS (1992) *Design of the Moderate Resolution Imaging Spectrometer – NADIR (MODIS-N)*. American Society of Photogrametry & Remote Sensing, pp. 374–385.

Stanich CG and Osterwisch FG (1994) *Advanced Operational Hyperspectral Scanners MIVIS and AHS*. First International Airborne Remote Sensing Conference and Exhibition, Strasbourg, France: Environmental Research Institute of Michigan.

Sun X and Baker JJ (2001) *A Hight Definition Hyperspectral Imaging System*. Proceedings, Fifth International Airborne Remote Sensing Conference and Exhibition, Veridian.

Torr MR, Baselow RW and Torr DG (1982) Imaging Spectrascopy of the Thermosphere from the Space Shuttle. *Applied Optics* 21: 4130.

Torr MR, Baselow RW and Mounti J (1983) An imaging spectrometric observatory for spacelab. *Astrophys. Space Science*. 92: 237.

Vane G, Green RO, Chrien TG, Enmark HT, Hansen EG and Porter WM (1993) The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of the Environment* 44: 127–143.

Warren CP, Anderson RD, Velasco A, Lin PP, Lucas JR and Speer BA (2001) *High Performance VIS and SWIR Hyperspectral Imagers for Airborne Applications*. Fifth International Conference and Exhibition, Veridian.

Wolfe WL (1997) *Introduction to Imaging Spectrometers*. SPIE Optical Engineering Press.

Wolfe WL and Zisfis GJ (eds) (1985) *The Infrared Handbook*, revised edition. Washington, DC: Office of Naval Research, Department of the Navy.

Zywicki RW, More KA, Holloway J and Witherspoon N (1994) *A Field Portable Hyperspectral Imager: The Large Area Fast Spectrometer*. Proceedings of the International Symposium on Spectral Sensing Research.

Telescopes

M M Roth, Astrophysikalisches Institut Potsdam, Potsdam, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In a broad sense, telescopes are optical instruments which provide an observer with an improved view of a distant object, where improved may be defined in terms of magnification, angular resolution, and light collecting power. Historically, the invention of the telescope was a breakthrough with an enormous impact on fundamental sciences (physics, astronomy), and, as a direct consequence, on philosophy; but also on other technological, economical, political, and military developments. It seems that the first refractive telescopes were built in the Netherlands near the end of the sixteenth century by Jan Lipperhey, Jakob Metius, and Zacharias Janssen. Based on reports of those first instruments, Galileo Galilei built his first telescope, which was later named after him, and made his famous astronomical observations of sunspots, the phases of Venus, Jupiter's moons, the rings of Saturn, and his discovery of the nature of the Milky Way as an assembly of very many stars. The so-called astronomical telescope was invented by Johannes Kepler, who in 1611 also published in *Dioptrice* the theory of this telescope. The disturbing effects of spherical and chromatic aberrations were soon realized by astronomers, opticians, and other users, but it was not before Moor Hall (1792) and John Dollond (1758) that at least the latter flaw was corrected for by introducing the achromat. Later, the achromat was significantly improved by Peter Dollond, Jesse Ramsden, and Josef Fraunhofer. The catadioptric telescope was probably introduced by Lenhard Digges (1571) or Nicolas Zucchius (1608). In 1671, Sir Isaac Newton was the first to use a reflector for astronomical observations. Wilhelm Herschel improved this technique and began in 1766 to build much larger telescopes with mirror diameters up to 1.22 m. When, in the nineteenth century, the conventional metal mirror was replaced by glass, it was Léon Foucauld (1819–1868) who applied a silver coating in order to improve the reflectivity of its surface.

Modern refractors can be built with superb apochromatic corrections, and astronomical reflectors have advanced to mirror diameters of up to 8.4 m (monolithic mirror), and 10 m (segmented mirror).

While the classical definition of a telescope involves an observer, i.e., the eye, in modern times often electronic detectors have replaced human vision. In the following, we will, therefore, use a more general definition of a telescope as an optical instrument, whose purpose is to image a distant object, either in a real focal plane, or in an afocal projection for the observation by eye.

Basic Imaging Theory

In the most simple case, a telescope may be constructed from a single lens or a single mirror,

creating a real focal plane. Let us, therefore, introduce some basic principles of imaging using optical elements with spherical surfaces.

Refraction at a Spherical Surface

Figure 1 shows an example of refraction at a spherical surface. A ray emerging from object O hits the surface between two media of refractive index n and n' at point P. After refraction, the ray continues at a deflected angle ε' towards the normal and intersects the optical axis at point O′. In the ideal case, all rays emerging from O at different angles σ are collected in O′, thus forming a real image of object point O in O′. In the paraxial approximation, we have $\sin\sigma = \tan\sigma = \sigma$, and $p \approx 0$. In triangle (O, O′, P): $\sigma' - \sigma = i - i'$. Using Snell's law $ni = n'i'$, we obtain:

$$\sigma' - \sigma = i'\frac{n' - n}{n} \qquad [1]$$

Since $\phi = \sigma' + i'$, and in the paraxial approximation:

$$\sigma' n' - \sigma n = \phi(n' - n) \qquad [2]$$

$$\frac{h}{s'}n' - \frac{h}{s}n = \frac{h}{r}(n' - n) \qquad [3]$$

we obtain finally:

$$\frac{n'}{s'} - \frac{n}{s} = \frac{n' - n}{r} \qquad [4]$$

For an object at infinity, where $s = \infty$, s' becomes the focal distance f':

$$f' = r\frac{n'}{n' - n} \quad \text{and} \quad \tilde{f} = -r\frac{n}{n' - n} \qquad [5]$$

Lenses

Combining two or more refractive surfaces, as in the preceding section, allows us to describe a single lens or more complex optical systems with several lenses in series. For a single lens in air, we just give the lens equation (without deriving it in detail):

$$s_2'(F') = \frac{1}{n-1}\frac{r_2[nr_1 - (n-1)d]}{(n-1)d + n(r_2 - r_1)} \qquad [6]$$

where d is the lens thickness, measured between the vertices of its surfaces, r_1, r_2 are the radii of the surfaces, n the index of refraction of the lens material, and $s_2'(F')$ the distance of the object image from the vertex of the last surface, facing the focal plane.

The back focal distance is given by:

$$f' = \frac{1}{n-1}\frac{nr_1r_2}{(n-1)d + n(r_2 - r_1)} \qquad [7]$$

An important special case is where the lens thickness is small compared to the radii of the lens surfaces, i.e., when $d \ll |r_2 - r_1|$. In this case we can neglect the term of $(n-1)d$ and write:

$$f' = \frac{1}{n-1}\frac{r_1r_2}{(r_2 - r_1)} \qquad [8]$$

Reflection at a Spherical Surface

Let us now consider the other simple case of imaging object O into O′ by reflection from the concave spherical surface, as depicted in Figure 2. In the paraxial case, we have:

$$\sigma = y/s, \qquad i = \phi - \sigma, \qquad \sigma' = y/s', \qquad i' = \phi - \sigma', \qquad \phi = y/r \qquad [9]$$

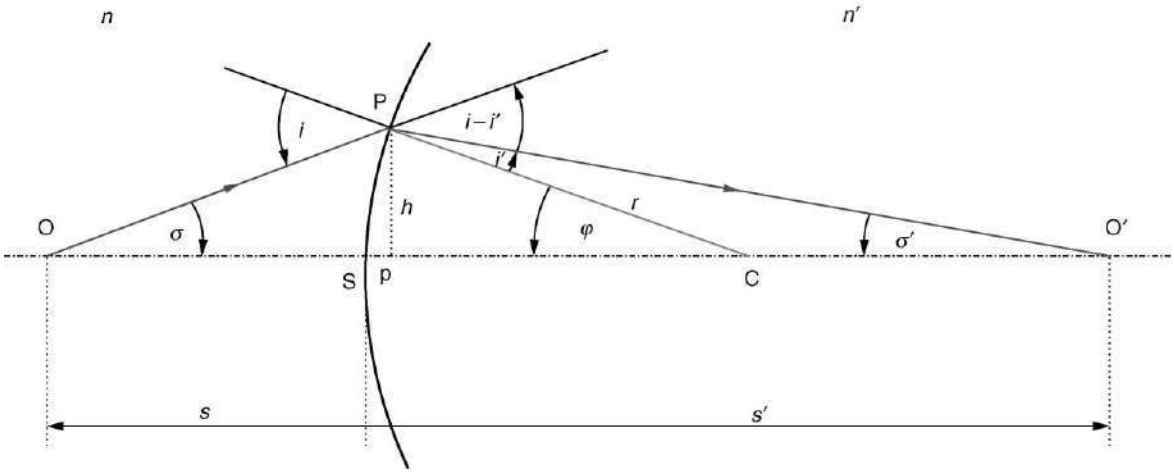

Figure 1 Refraction at a spherical surface between media with refractive indices n and n'.

With $i = -i'$, we obtain:

$$\frac{1}{s'} - \frac{1}{s} = -\frac{2}{r} \qquad [10]$$

Again, with $s = \infty$, the focal length becomes:

$$f' = -\frac{r}{2} \qquad [11]$$

In what follows, we will now assume that the distance of the object is always large (∞), which, in general, is the situation when a telescope is employed. The image is formed at a surface defined by O′, ideally a focal plane (paraxial case). We will consider deviations from the ideal case below.

Simple Telescopes

As we have observed above, a telescope is in principle nothing but an optical system, which images an object at infinity. In the most simple case, a lens with positive power or a concave mirror will satisfy this condition. The principle of a refractor and a reflecting telescope is shown in Figures 3 and 4. In both cases, parallel light coming from the distant object is entering through the entrance pupil, experiences refraction (reflection) at the lens (mirror), respectively, and is converging to form a real image in the focal plane.

Two point sources, for example two stars in the sky, separated by an angle δ, will be imaged in the focal plane as two spots with separation Δy. When f is the focal length of the telescope, we therefore have:

$$\tan\delta = \frac{\Delta y}{f} \qquad [12]$$

The so-called plate scale m of the telescope in units of arcsec/mm is given by:

$$m = 3600\,\mathrm{atan}\frac{1}{f} \quad (f \text{ in mm}) \qquad [13]$$

In the case of the refractor, we have also shown an eyepiece, illustrating that historically the instrument was invented to enhance the human vision. The eye is located at the position of the exit pupil, thus receiving parallel light, i.e., observing an object at infinity, however (i) with a flux which is increased by a factor A, given by the ratio of the area of the telescope aperture with diameter D to the area of the eye's pupil d with $A = (D/d)^2$, and (ii) with a magnification M of the apparent angle between two separate objects at infinity, which is given by the ratio of the focal lengths of the objective and the eyepiece:

$$\Gamma = \frac{f_{\mathrm{obj}}}{f_{\mathrm{ocl}}} \qquad [14]$$

The magnification can also be expressed as the ratio of the diameters of the exit pupil and the entrance pupil:

$$\Gamma = \frac{D_{\mathrm{expup}}}{D_{\mathrm{enpup}}} \qquad [15]$$

The first astronomical telescopes with mirrors were equipped with an eyepiece for visual observation. The example in Figure 4 shows a configuration which is more common nowadays for professional astronomical observations, using a direct imaging detector in the focal plane (photographic plate, electronic camera).

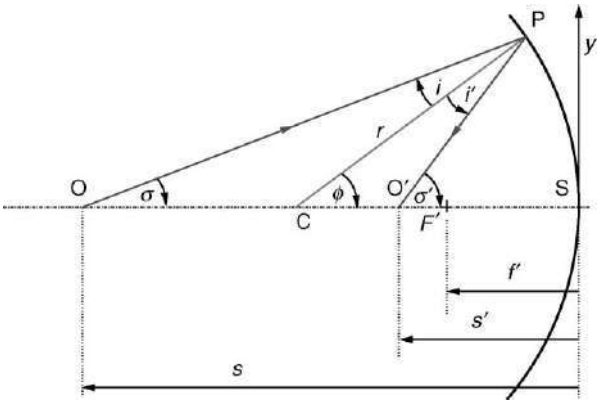

Figure 2 Reflection at a spherical surface with radius r.

Aberrations

Reflector Surfaces

We shall now study the behavior of a real telescope beyond the ideal in the case of a reflector. Figure 5 shows the situation of a mirror, imaging an on-axis

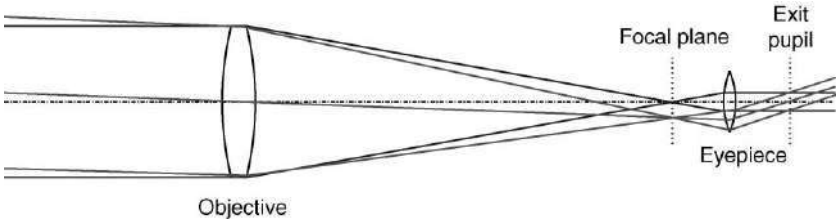

Figure 3 Refractor, astronomical telescope for visual observations with eyepiece (Kepler telescope).

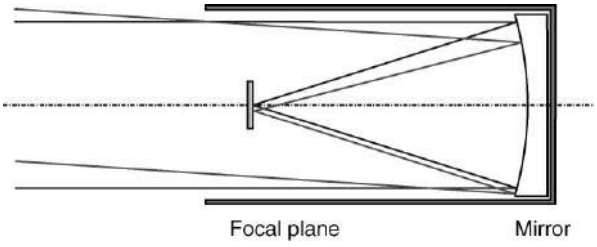

Figure 4 Reflector, astronomical telescope with prime focus imager.

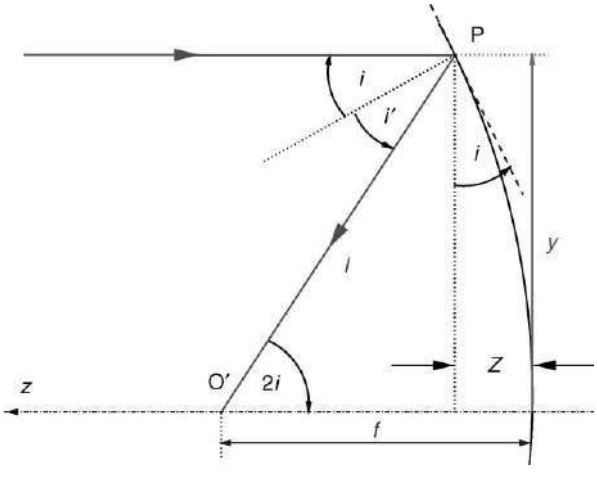

Figure 5 Imaging an object at infinity with reflector.

object at infinity into focus at point O′. With Fermat's principle, the optical paths of all rays forming the image must have identical length l. For all j, we must satisfy:

$$Z_j + f = l \qquad [16]$$

From the geometry in Figure 5, we find also:

$$y_i^2 + (f - Z_j)^2 = l^2 \qquad [17]$$

Combining eqns [16] and [17], we obtain the equation for a surface which satisfies the condition for perfect on-axis imaging:

$$y^2 = -4fz \qquad [18]$$

Obviously only a paraboloid is capable of fulfilling the condition for all rays parallel to the optical axis at whatever separation y. Also, we see that spherical mirrors are less than ideal imagers: rays at increasing distance y from the axis will suffer from increasing path length differences. The effect is called spherical aberration. We will investigate aberrations more quantitatively further below.

Elementary axisymmetrical surfaces can be expressed as follows, with parameter C being the conic constant:

$$y^2 - 2Rz + (1 + C)z^2 = 0 \qquad [19]$$

and the following characteristics:

$C > 0$	ellipsoidal, prolate
$C = 0$	spherical
$-1 < C < 0$	ellipsoidal, oblate
$C = -1$	paraboloidal
$C < -1$	hyperboloidal

Let us now leave the paraxial approximation and determine the focal length of rays which are parallel to, but further away at a distance y from the optical axis. Figure 5 shows the relevant geometrical relations. The intersection of the reflected ray with the optical axis at distance f from the vertex can be found from:

$$f = z + (f - z) \quad \text{where} \quad \frac{y}{f - z} = \tan(2i) \qquad [20]$$

We use the slope of the tangent in point P:

$$\frac{dz}{dy} = \tan i \qquad [21]$$

differentiate eqn [19] to get:

$$\frac{dz}{dy} = \frac{y}{R - (1 + C)z'} \qquad [22]$$

use the trigonometrical relation:

$$\tan 2\phi = \frac{2 \tan \phi}{1 - \tan^2 \phi} \qquad [23]$$

and substitute for $\tan(2i)$ in eqn [20] to obtain:

$$f = \frac{R}{2} + \frac{(1 - C)z}{2} - \frac{y^2}{2(R - (1 + C)z)} \qquad [24]$$

In order to express eqn [24] solely in terms of y, we rewrite eqn [19]:

$$z = \frac{R}{1 + C}\left(1 - \sqrt{1 - \frac{y^2}{R^2}(1 + C)}\right) \qquad [25]$$

Expanding z in a series and neglecting any orders higher than 6:

$$z = \frac{y^2}{2R} + (1 + C)\frac{y^4}{8R^3} + (1 - C)^3 \frac{y^6}{16R^5} \qquad [26]$$

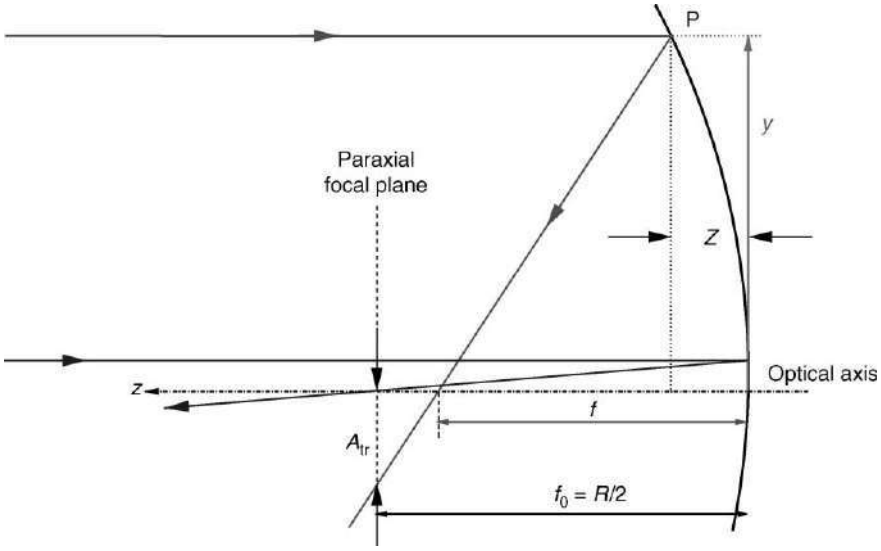

Figure 6 Transverse spherical aberration.

we obtain finally:

$$f = \frac{R}{2} - \frac{1+C}{4R}y^2 - \frac{(1+C)(3+C)}{16R^3}y^4 \qquad [27]$$

.

Transverse Spherical Aberration

Obviously, for any $C \neq -1$, rays of different distance y to the optical axis are focused at different focal lengths f (spherical aberration). The net effect for a distant point source is that the image is blurred, rather than forming a sharp spot. The lateral deviation of nonparaxial rays intersecting the nominal focal plane at $f_0 = R/2$ is described by the transverse spherical aberration A_{tr} (Figure 6):

$$A_{tr,sph} = \frac{y}{f-z}(f_0 - f) \qquad [28]$$

Using eqns [19] and [27], one obtains:

$$A_{tr,sph} = -\frac{1+C}{2R^2}y^3 + \frac{3(1+C)(3+C)}{8R^4}y^5 + \cdots \qquad [29]$$

Angular Spherical Aberration

It is sometimes more convenient to consider the angular spherical aberration, since a telescope is naturally measuring angles between objects at infinity. To this end, we compare an arbitrary reflector surface according to eqn [19] with a reference paraboloid, which is known to be free of on-axis spherical aberration (Figure 7). For any given ray, we observe how the reflected ray under study is tilted

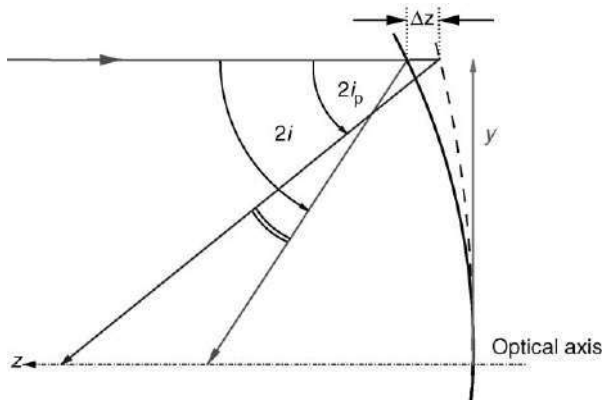

Figure 7 Angular aberration determined with reference to a paraboloid.

towards the reference ray, which would be reflected from the ideal surface of the paraboloid. While the former is measured under an exit angle of $2i$ towards the optical axis (see Figure 5), we designate the latter $2i_p$. The angular spherical aberration is defined as the difference of these two angles:

$$A_{ang,sph} = 2i - 2i_p = 2(i - i_p) \qquad [30]$$

Again, we can use the slope of the tangent to determine the change of angles $i \rightarrow i_p$:

$$A_{ang,sph} = 2\left(\frac{dz}{dy} - \frac{dz_p}{dy}\right) = 2\frac{d}{dy}\Delta z \qquad [31]$$

It is, therefore, sufficient to consider the difference in z between the surface under study and the

reference paraboloid. From eqn [26], we obtain as an approximation to third order:

$$A_{\text{ang,sph}} = -(1+C)\frac{y^3}{R^3} \qquad [32]$$

Aberrations in the Field

Since the use of a telescope exclusively on-axis would be somewhat limited, let us finally investigate the behavior under oblique illumination, i.e., when an object at infinity is observed under an angle Θ towards the optical axis of the system. This will allow us to assess aberrations in the field. For simplicity, we are considering a paraboloid, remembering that on-axis the system is free of spherical aberration. Again, we are employing an imaginary reference paraboloid which is pointing on-axis towards the object at angle Θ (Figure 8). Its coordinate system (O', z', y') is displaced and tilted with regard to the system under study (O, z, y). One can show that the offset Δz, which is used to derive the angular aberration in the same way as above, is given by:

$$2\Delta z = a_1\frac{y^3\Theta}{R^2} + a_2\frac{y^2\Theta^2}{R} + a_3 y\Theta^3 \qquad [33]$$

The angular aberration in the field is then:

$$A_{\text{ang,f}} = 3a_1\frac{y^2\Theta}{R^2} + 2a_2\frac{y\Theta^2}{R} + a_3\Theta^3 \qquad [34]$$

The coefficients are called:

a_1: coma

a_2: astigmatism

a_3: distortion

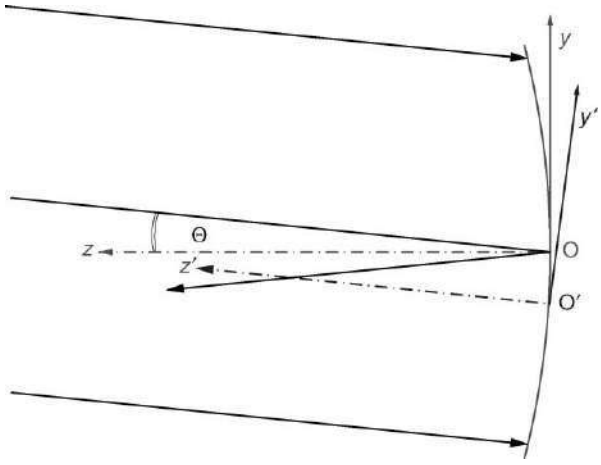

Figure 8 Aberration in the field, oblique rays.

Chromatic Aberration

So far we have only considered optical aberrations of reflecting telescopes. Spherical aberration in the case of a lens is treated in analogous ways and will not be discussed again. However, the wavelength dependence of the index of refraction in optical media (dispersion) is an important factor and gives rise to chromatic aberrations which are only encountered in refractive optical systems. We can distinguish two principal effects in the paraxial approximation: (i) longitudinal chromatic aberration, resulting in different axial image positions as a function of wavelength; and (ii) lateral chromatic aberration, which can be understood as an aberration of the principal ray as a function of wavelength, leading to a varying magnification as a function of wavelength (plate scale in the case of a telescope).

As an example, let us determine the wavelength dependence of the focal length of a single thin lens. Opticians are conventionally using discrete wavelengths of spectral line lamps of different chemical elements, e.g., the Mercury e line at 546.1 nm, Cadmium F' 480.0 nm, or Cadmium C' 643.8 nm, covering more or less the visual wavelength range of the human eye. Differentiation of eqn [8] yields for the wavelength of e:

$$\frac{f'}{\mathrm{d}n} = -\frac{1}{(n_e-1)^2}\frac{r_1 r_2}{r_2 - r_1} \qquad [35]$$

Substituting

$$\frac{r_1 r_2}{r_2 - r_1} = f'_e(n_e - 1) \qquad [36]$$

we get:

$$\mathrm{d}f' = -\frac{\mathrm{d}n}{n_e - 1}f'_e \qquad [37]$$

Replacing the differential by the differences $\Delta f' = f'_{F'} - f'_{C'}$ and $\Delta n = n_{F'} - n_{C'}$, then:

$$f'_{F'} - f'_{C'} = -\frac{n_{F'} - n_{C'}}{n_e - 1}f'_e = -\frac{f'_e}{\nu_e} \qquad [38]$$

ν_e is called Abbe's Number:

$$\nu_e = \frac{n_e - 1}{n_{F'} - n_{C'}} \qquad [39]$$

It is a tabulated characteristic value of dispersion for any kind of glass and allows one to quickly estimate the difference in focal length at the extreme wavelengths F' and C' when the nominal focal length of a lens is known as e.

Angular Resolution

In the previous sections, we have exclusively employed geometrical optics for the description of telescope properties. With geometrical optics and in the absence of aberrations, there is no limit to the angular resolution of a telescope, i.e., there is no limiting separation angle δ_{lim}, below which two distant point sources can no longer be resolved. However, real optical systems experience diffraction, leading to the finite width of the image of a point source (Airy disk), whose normalized intensity distribution in the focal plane as a function of radius r' from the centroid for the case of a circular aperture of diameter D can be conveniently described by a Bessell function (λ: wavelength, f': focal length of the objective):

$$\frac{I(r)}{I_0} = \left[\frac{2J_1(\nu)}{\nu}\right]^2 \quad \text{where} \quad \nu = \frac{\pi D r'}{\lambda f'} \qquad [40]$$

The intensity distribution is plotted in Figure 9. The first four minima of the diffraction ring pattern are listed in Table 1. Two point sources are defined to be at the resolution limit of the telescope when the maximum of one diffraction pattern happens to coincide with just the first diffraction minimum of the other one (Figure 9, right). As a rule of thumb, the angular resolution is of a telescope with diameter D is given by:

$$\delta_{\min} = 1.22\frac{\lambda}{D} \rightarrow \frac{138}{D} \qquad [41]$$

the latter in units of arcsec for a wavelength of 550 nm and D in mm.

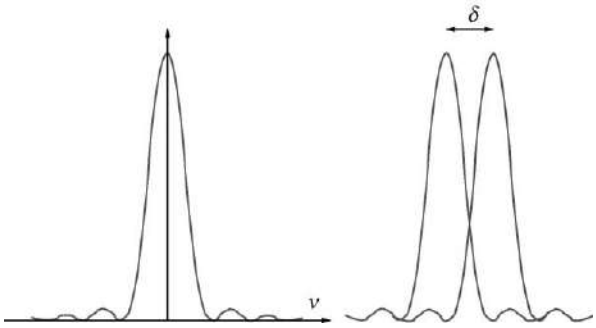

Figure 9 Left: radial intensity distribution of diffraction pattern (Airy disk). Right: two overlapping point source images with same intensity at limiting angular separation δ.

Table 1 Diffraction minima of airy disk

n	ν_n	$r'D/2f'\lambda$
1	3.83	0.61
2	7.02	1.12
3	10.2	1.62
4	13.3	2.12

Types of Telescopes

There are very many different types of telescopes which were developed for different purposes. In order to obtain an impression, we will only briefly introduce a nonexhaustive selection of refractors and reflectors, which by no means is representative of the whole. For a more complete overview, the reader is referred to the literature.

Refractor Objectives

Refractors are first of all characterized by their objectives. Figure 10 shows a sequence of lenses, indicating the progress of lens design developments from left to right. All lenses are groups 2 or 3, designed to compensate for chromatic aberration through optimized combinations of crown and flint glasses. The first lens is a classical C achromat, e.g., using BK7 and SF2 glass, which is essentially the design which was introduced by Hall and Dollond. It is still a useful objective for small telescopes, where a residual chromatic aberration is tolerable. The second lens is the Fraunhofer E type, which is characterized by an air gap between the two lenses (BK7, F2), which helps to optimize the image quality of the system, reducing spherical aberration and coma. As a disadvantage, the two additional glass–air interfaces lead to ghost images (reflected light) and Fresnell losses. The objective is critical with regard to the alignment of tip and tilt. The third example is an improved AS type achromat (KzF2, BK7), which is similar to the E type in

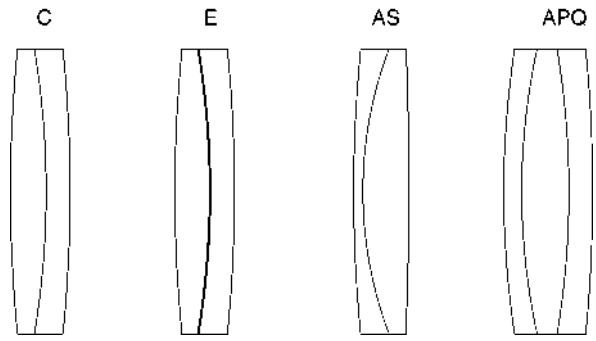

Figure 10 Different types of refractor objectives, from left to right: classical C achromat, Fraunhofer E achromat with air gap, AS achromat, APQ apochromat.

that it also possesses an air gap, however with an improved chromatic aberration correction in comparison to the Fraunhofer objective. The lenses show a more pronounced curvature and are more difficult to manufacture. The beneficial use of KzF2 glass is an example of the successful research for the development of new optical media, which was pursued near the end of the nineteenth century in a collaboration between Otto Schott and Ernst Abbé. The last example is a triplet, employing ZK2, CaF_2, and ZK2. This APQ type of objective is an apochromat, whose chromatic aberration vanishes at 3 wavelengths. It is peculiar for its use of a crystal (calcium fluoride, CaF_2) as the optical medium of the central lens. The manufacture of the CaF_2 lens is critical and expensive, but it provides for an excellent correction of chromatic and spherical aberration of this high quality objective. The triplet is also difficult to manufacture because of the coefficient of thermal expansion, which is roughly different by a factor of two between CaF_2 and the glass. In order to avoid disrupture of the cemented lens group, some manufacturers have used oil immersion to connect the triplet. Since CaF_2 not only has excellent dispersion properties to correct for chromatic aberration, but also exhibits an excellent transmission beyond the visual wavelength regime, this type of lens is particularly useful for applications in the blue and UV.

Finally, Figure 11 shows the complete layout of an astrograph, which is a telescope for use with photographic plates. The introduction of this type of instrument meant an important enhancement of astrophysical research by making available a significantly more efficient method, compared to the traditional visual observation of individual stars, one by one. Using photographic plates, it became possible to measure position and brightness of hundreds and thousands of stars, when previously one could investigate tens. Typical astrographs were built with apertures between 60 and 400 mm, and a field-of-view as large as 15 degrees. Despite the growing importance of space-based observatories for astrometry, some astrographs are still in use today.

Basic Oculars

While the general characteristics of a refractor are dominated by the objective, the eyepiece nevertheless plays an important role for the final appearance to the observer, with an effect on magnification, image quality, chromatism, field of view, and other properties of the entire system. From a large variety of different ocular types, let us briefly consider a few simple examples.

The first eyepiece in Figure 12 is the one of the Kepler telescope, which was already introduced in Figure 3. It has a quite simple layout with a single lens of positive power. The lens is located after the telescope focal plane at a distance which is equal to the ocular focal length f_oc. The telescope and the ocular focal planes are coinciding, and as a result, the outcoming beam is collimated, forming the system exit pupil after a distance f_oc. This is where ideally the observer's eye is located. The focal plane is creating a real image, which is a convenient location for a reticle. The Kepler telescope is, therefore, a good tool for measurement and alignment purposes. We note that the image is inverted, which is relatively unimportant for a measuring telescope and astronomy, but very annoying for terrestrial use.

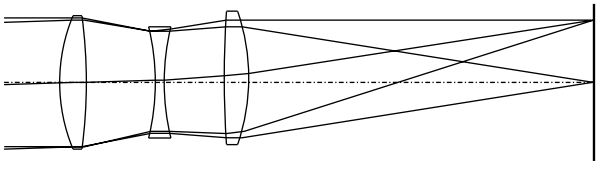

Figure 11 Astrograph.

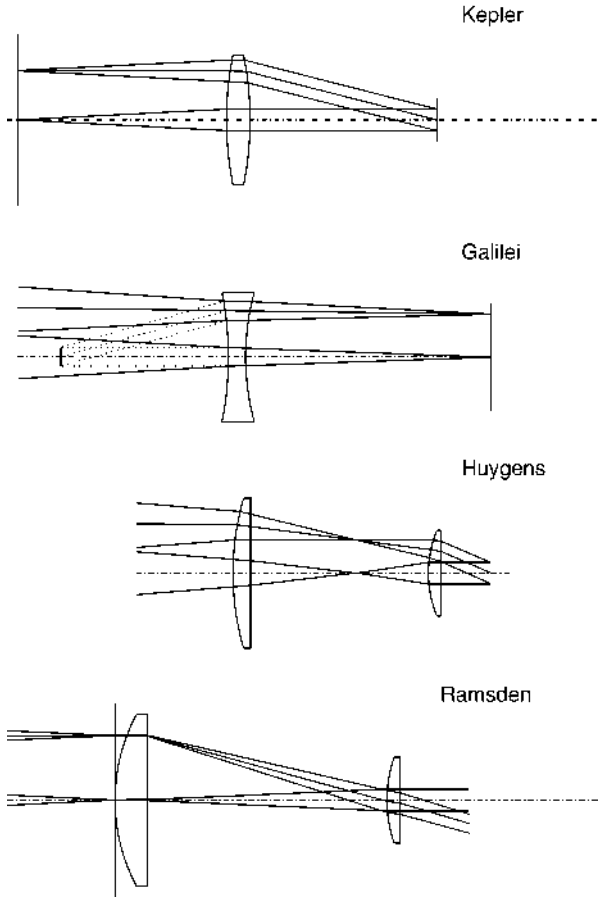

Figure 12 Elementary eyepieces after Kepler, Galilei, Huygens, and Ramdsen.

This latter property has been avoided in the next example, which is the ocular of the Galilean telescope, which was built by Galileo and used for his famous observations of planets and moons in the solar system. It uses a single lens, but contrary to the Kepler type with negative power. The lens is located in the telescope beam in front of the focal plane of the objective, which means that the exit pupil is virtual and not accessible to the observer (indicated by dotted lines in Figure 12). The overall length of the telescope is somewhat shorter than Kepler's, however at the expense of the lack of a real focal plane, so no reticle can be applied.

The third example is the 2-lens-ocular of Christian Huygens, who was able to prove with his design that it is possible to correct paraxially the lateral chromatic aberration. The eyepiece is still being built today.

Another elementary ocular is shown in Figure 12, which is the eyepiece after Ramsden. This system is also corrected for lateral chromatic aberration. The real image is formed on the front surface of the second lens, providing a location for a reticle or a scale for measuring purposes. The exit pupil is interior to the ocular. In addition to these selected basic types, there are many more advanced multilens oculars in use today, with excellent correction and large field of view (up to 60°).

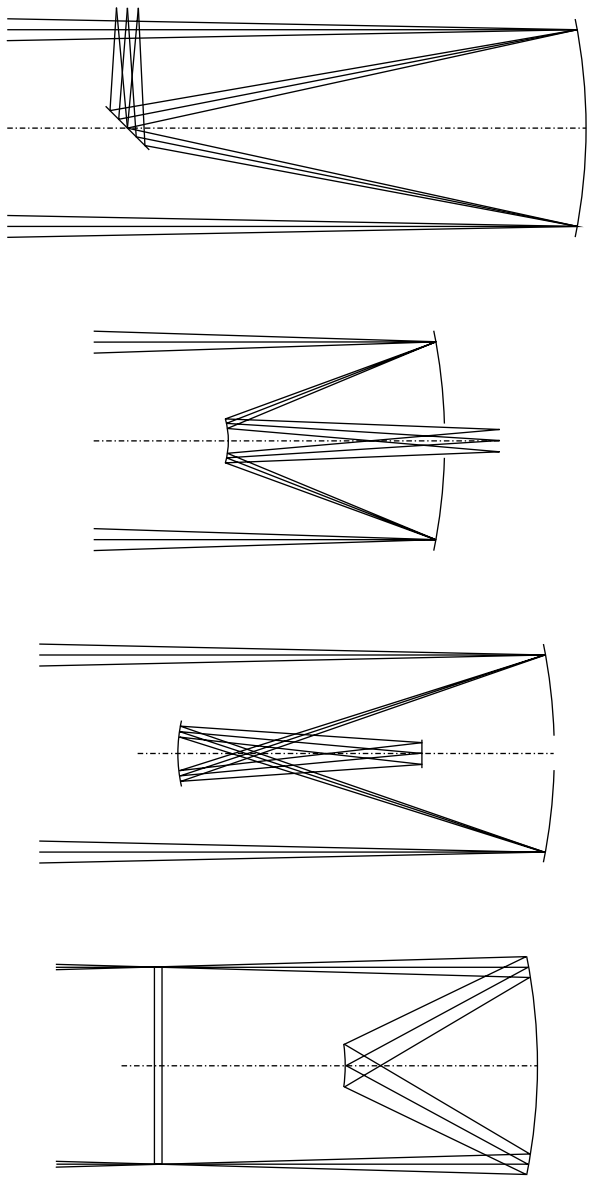

Figure 13 Reflectors, from top to bottom: Newton, Cassegrain, Gregory, Schmidt.

Reflectors

Let us now describe several basic types of reflectors (see Figure 13). The first example is the Newton reflector, which uses a flat folding mirror to create an accessible exterior focus. It is conceptually simple and still in use today for amateur astronomy. When the main mirror is a paraboloid, the spherical aberration is corrected on the axis of the system.

The second example represents an important improvement over the simple 1-mirror design of Newton's telescope, in that the combination of a parabolic first mirror (main mirror, M1) with a hyperbolic secondary mirror (M2) eliminates spherical aberration, but not astigmatism and coma. The Cassegrain telescope has a curved focal surface which is located behind the primary mirror. Due to the onset of coma and astigmatism, the telescope has a modest useful field-of-view. Besides the correction properties, the system has significant advantages in terms of mechanical size and weight, playing an important role for the construction of large telescopes. The basic type of this design has practically dominated the construction of the classical large astronomical telescopes with M1 diameters of up to 5 m, and, more recently, the modern 8–10 m class telescopes. On the practical side, the external focus of the Cassegrain on the bottom of the mirror support cell is a convenient location for mounting focal plane instruments (e.g., CCD cameras, spectrographs, and others).

The more common variant of the Cassegrain type, which was actually used for the design of these telescopes, is the layout of Ritchey (1864–1945) and Cretien (1879–1956), which combines two hyperbolic mirrors for M1 and M2 (RCC), but otherwise has the same appearance as the Cassegrain. Besides spherical aberration, it also corrects for coma, giving a useful field-of-view of up to 1°. Like in the case of the Cassegrain telescope, the RCC exhibits some curvature of the focal plane.

Our third example is the Gregory reflector, which also is a 2-mirror system, however with a concave elliptical secondary mirror. The correction of this telescope is similar, but somewhat inferior to the one of the Cassegrain. The use of a concave M2 is advantageous for manufacture and testing. Nevertheless, the excessive length of the system has precluded the use of this layout for large astronomical telescopes, except for stationary beams, for example, in solar observatories.

The last example is the design put forward by Bernhard Schmidt (1879–1935) who wanted to create a telescope with an extremely wide field-of-view. The first telescope of his new type achieved a field-of-view of 16° – something which had not been feasible with previous techniques. The basic idea is to use a fast spherical mirror, and to correct for spherical aberration by means of an aspherical corrector plate in front of the mirror, which also forms the entrance pupil of the telescope. The focal surface is strongly curved, making it necessary to employ a detector with the same radius of curvature (bent photographic plates), or to compensate the curvature by means of a field lens immediately in front of the focus.

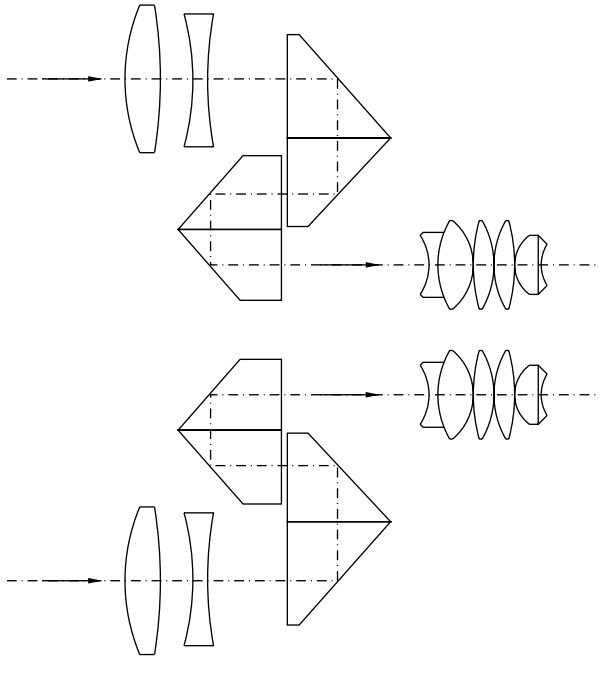

Figure 14 Principle of prism binoculars with folded beams.

The presence of the corrector plate and, if used, the field flattener, gives rise to significant chromatism. Nevertheless, the Schmidt telescope was a very successful design and dominated wide-angle survey work in astronomy for many decades.

Applications

General Purpose Telescopes

According to the original meaning of the Greek 'tele–skopein', the common conception of the telescope is the one of an optical instrument, which provides an enlarged view of distant objects. In addition, a large telescope aperture, compared to the eye, improves night vision and observations with poorly illuminated scenes. There are numerous applications in nautical, security, military, hunting, and other professional areas. For all of these applications, an intuitive vision is important, essentially ruling out the classical astronomical telescope (Kepler telescope, see Figure 3), which produces an inverted image.

Binoculars with Porro prisms are among the most popular instruments with upright vision, presenting a compact outline due to the folded beam geometry (Figure 14), but also single tube telescopes with re-imaging oculars are sometimes used.

Metrology

As we have realized in the introduction, telescopes are instruments which perform a transformation from angle of incidence in the aperture to distances from the origin in the focal plane. If the focal length of the objective is taken sufficiently long, and a high-quality eyepiece is used, very small angles down to the arcsec regime can be visually observed. Telescopes are, therefore, ideal tools to measure angles to very high precision.

In Figure 15 we see a pair of telescopes, facing each other, where one instrument is equipped with a light source, and the other one is used to observe the first one. The first telescope is also called a collimator. It possesses a lamp, a condensor system, a focal plane mask (usually a pinhole, a crosshair, or some other useful pattern imprinted on a transparent plate), and the telescope objective. The focal plane mask is projected to infinity. It can thus be observed with the

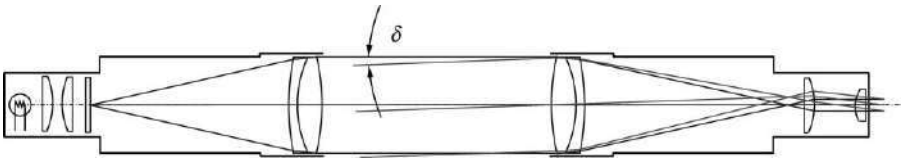

Figure 15 Collimator and telescope.

measuring telescope, such that an image of the collimator mask appears in the telescope focal plane. Also the telescope is equipped with a mask in its focal plane, usually with a precision scale or another pattern for convenient alignment with the collimator mask. If now the collimator is tilted by a small angle δ against the optical axis of the telescope, the angle of incidence in the measuring telescope changes by this same amount, giving rise to a shift of the image by $\Delta x = f \tan \delta$. Note that the measurement is completely insensitive to parallel shifts between the telescope and the collimator, since it is parallel light which is transmitted between the two instruments.

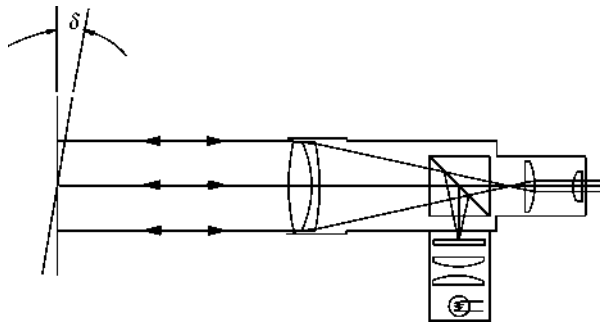

Figure 16 Autocollimator.

Another special device, which combines the former two instruments into one, is the Autocollimator. The optical principle is shown in Figure 16: in principle, the autocollimator is identical to a measuring telescope, except that a beamsplitter is inserted between the focal plane and the objective. The beamsplitter is used to inject light from the same kind of light source device as in the normal collimator. The axis of the folded light source beam and of the illuminated collimator mask are aligned to match the focal plane mask of the telescope section: the optical axis of the emerging beam is identical to the optical axis of the telescope. Under this condition, the reflected light from a test object will be exactly centered on the telescope focal plane mask, when it is perfectly aligned to normal incidence. Normally, a high-quality mirror is attached to the device under study, but sometimes also the test object itself possesses a plane surface of optical quality, which can be used for the measurement.

Figure 17 shows schematically an example for such a measurement, where a mirror is mounted on a linear stage carriage for the purpose of measuring the flexure of the stage to very high precision. With a typical length of ≈ 1 m, this task is not a trivial problem for a mechanical measurement.

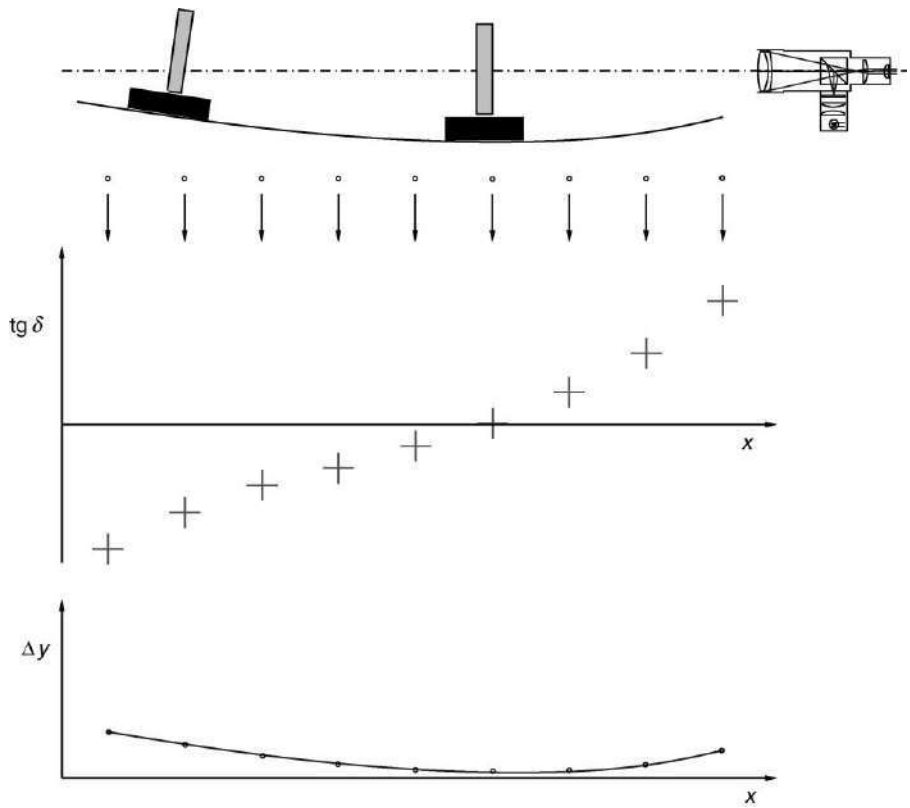

Figure 17 Using an autocollimator to measuring the flexure of a linear stage.

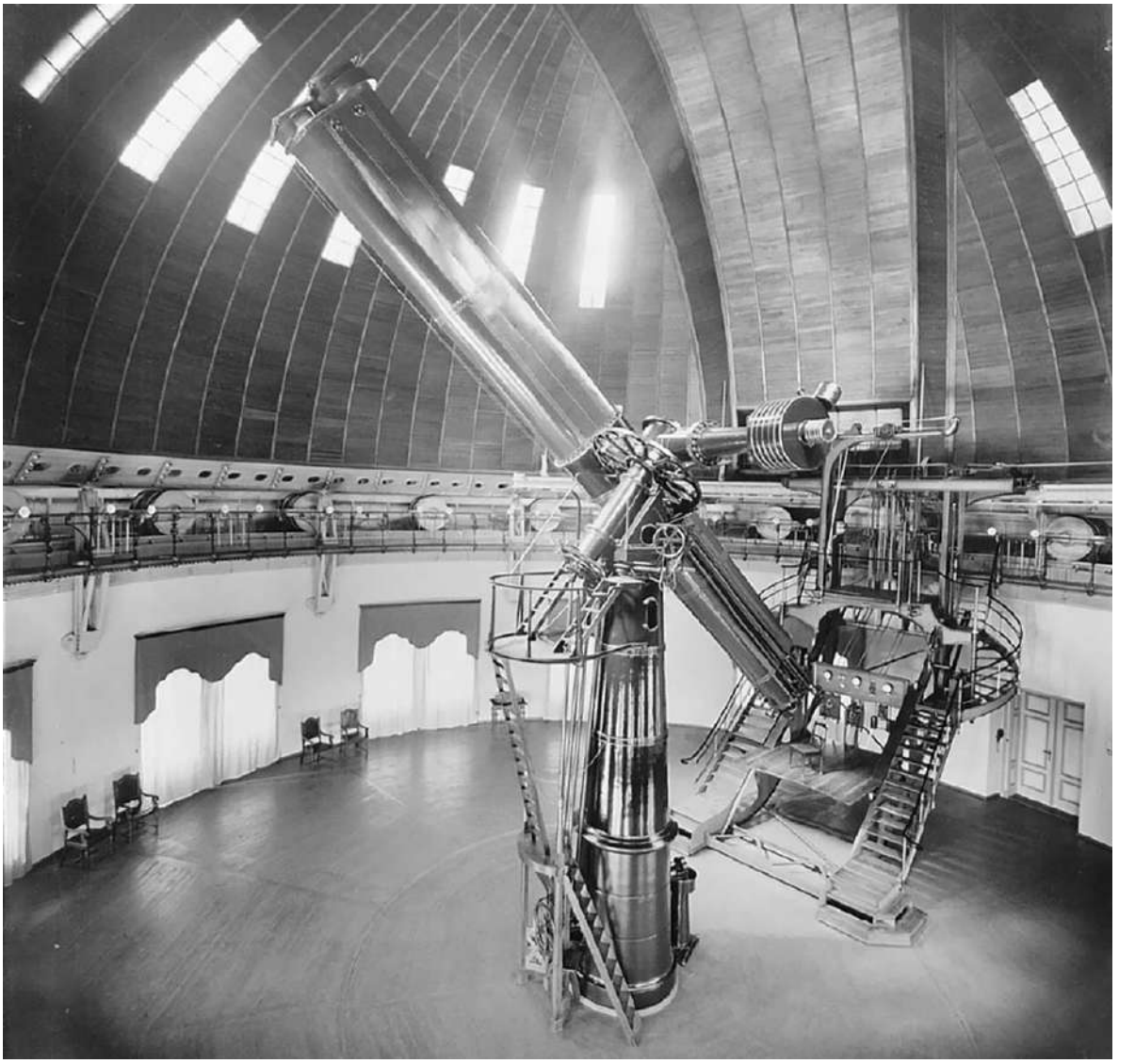

Figure 18 The 80 cm + 60 cm 'Great Refractor' at the Astrophysical Observatory Potsdam. The correction of the 80 cm objective of this telescope led Johannes Hartmann to the development of the 'Hartmann Test', which was the first quantitative method to measure aberrations. This method is still in use today, and has been further developed to become the 'Shack-Hartmann' technique of wavefront sensing. Courtesy of Astrophysikalisches Institut Potsdam.

By measurement with the autocollimator the tilt of the mirror consecutively for a sufficiently large number of points x_i along the rail, we can now plot the slope of the rail at each point x_i as tan d:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \tan\delta \qquad [42]$$

Integrating the series of slope values reproduces the shape of the rail $y = f(x)$.

Other applications include multiple autocollimation, wedge measurements, and right angle measurements.

Multiple autocollimation is achieved by inserting a fixed plane-parallel and semi-transparent mirror between the autocollimator and the mirror under study. Light coming from the test mirror will be partially transmitted through the additional mirror towards the autocollimator, and partially reflected back to the test mirror, and so forth. After k reflections, a deflection of $k\delta$ is measured, when δ is the tilt of the test mirror.

The wedge angle α of a plane-parallel plate with refractive index n_{w} is easily measured by observing the front and back surface reflections from the plate, keeping in mind that the beam bouncing back from the rear will also experience refraction in the plate. If δ is the measured angle difference between the two front and rear beams, then $\alpha = \delta/2n_{\mathrm{w}}$.

Two surfaces of an object can be tested for orientation at right angles by attaching a mirror to surface 1 and observe the mirror with the fixed autocollimator. In a second step, an auxiliary pentaprism is inserted into the beam, deflecting the light towards surface 2, where the mirror is attached next. If a precision pentaprism is used, any deviation from right angles between the two surfaces is seen as an offset between step 1 and 2.

These are only a few examples of the very many applications for alignment and angle measurements in optics and precision mechanics.

Without going into much detail, we are just mentioning that also the theodolite and sextant make use of the same principle of angular measurements, however without the need of a collimator or autocollimation.

Astronomical Telescopes

Despite its importance as a measurement and testing device for optical technology and manufacture as described in the preceeding paragraph, it is the original invention and subsequent improvement of the telescope for *Astronomy* which is probably the most fascinating application of any optical system known to the general public. The historical development of the astronomical telescope spans from the 17th century Galileian refractor with an aperture of just a few centimeters in diameter, over the largest refractors built for the Lick and Yerkes observatories with objectives of up to ≈ 1 m in diameter (see also Figures 18–20), the 4 m-class reflectors of the second half of the 20th century like e.g. the Hale Observatory 5m, the Kitt Peak and Cerro Tololo 4 m, the La Silla 3.6 m, or the La Palma 4.2 m telescopes (to name only a few), to the current state-of-the-art of 8–10 m class telescopes, prominent examples being the two Keck Observatory 10 m telescopes on Mauna Kea (Hawaii), or the ESO Very Large Observatory (Paranal, Chile), consisting of four identical 8.2 m telescopes (Figures 21–24).

This development has always been driven by the discovery of fainter and fainter celestial objects, whose quantitative study is the subject of modern astrophysics, involving many specialized disciplines such as atom/quantum/nuclear physics, chemistry, general relativity, electrodynamics, hydrodynamics, magneto-hydrodynamics, and so forth. In fact, cosmic objects are often referred to as 'laboratories'

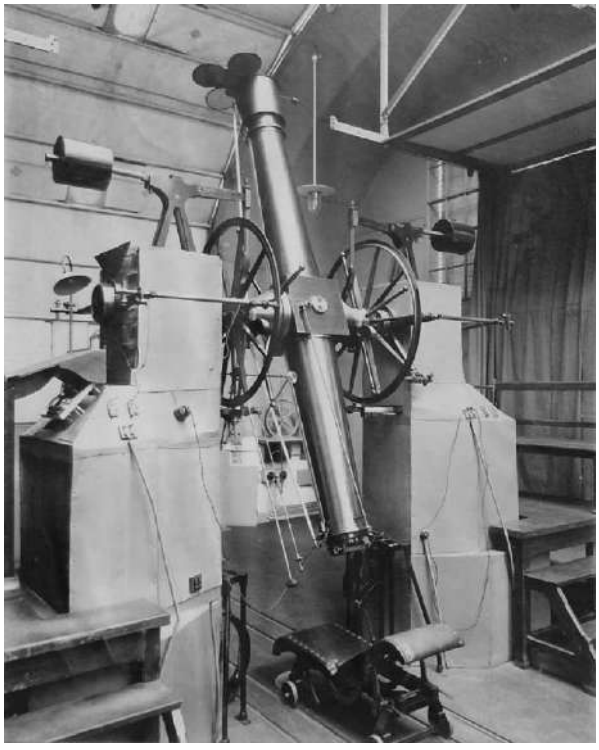

Figure 19 Pistor & Martin Meridian Circle at the former Berlin Observatory in Potsdam Babelsberg. The instrument was one of the finest art telescopes with ultra-high precision for astrometry, built in 1868. Courtesy of Astrophysikalisches Institut Potsdam.

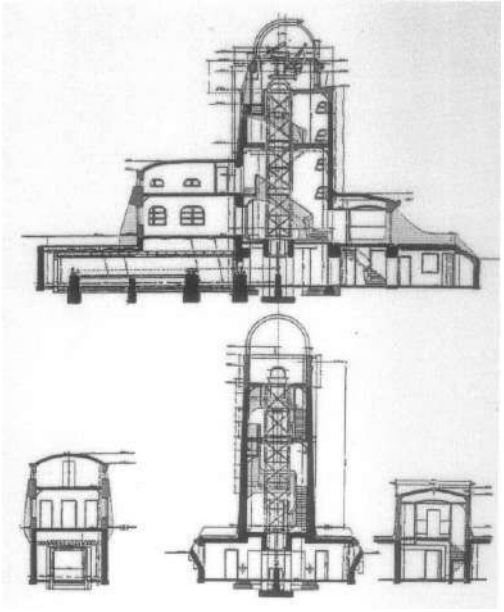

Figure 20 Architect's sketch of historical 'Einsteinturm' solar telescope, Astrophysical Observatory Potsdam. The telescope objective is mounted in the tower structure, facing upward a siderostat which is mounted inside the dome. The light is focused in the basement after deflection from a folding mirror on the slit of a bench-mounted high resolution spectrograph. The telescope represented state-of-the-art technology at the beginning of the twentieth century. Courtesy of Astrophysikalisches Institut Potsdam.

Figure 21 The Very Large Telescope Observatory (VLT) on Mt Paranal, Chile, with VLT Interferometer (VLTI). The four large enclosures are containing the active optics controlled VLT unit telescopes, which are operated either independently of each other, or in parallel to form the VLTI. The structures on the ground are part of the beam combination optics/delay lines. Courtesy of European Southern Observatory.

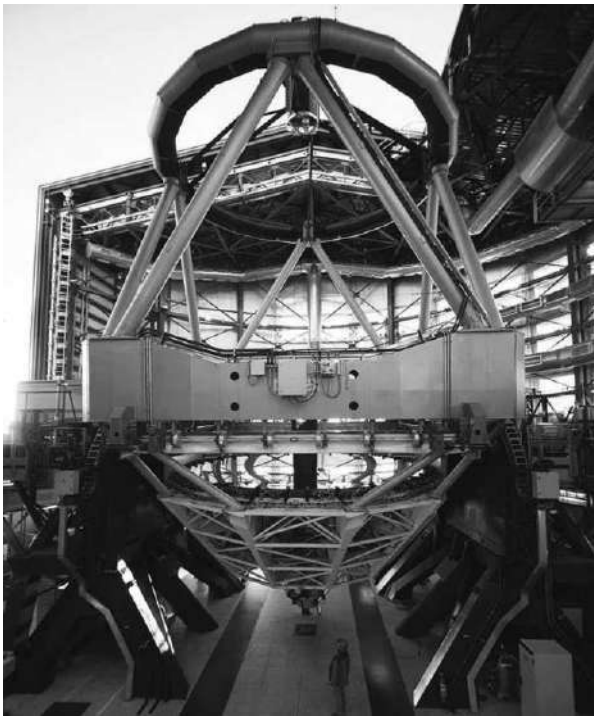

Figure 22 Total view of Kueyen, one of four unit telescopes of the VLT observatory. Courtesy of European Southern Observatory.

with extreme conditions of density, temperature, magnetic field strength, etc., which would be impossible to achieve in any laboratory on Earth. Except for particles (neutrinos, cosmic rays), gravitational

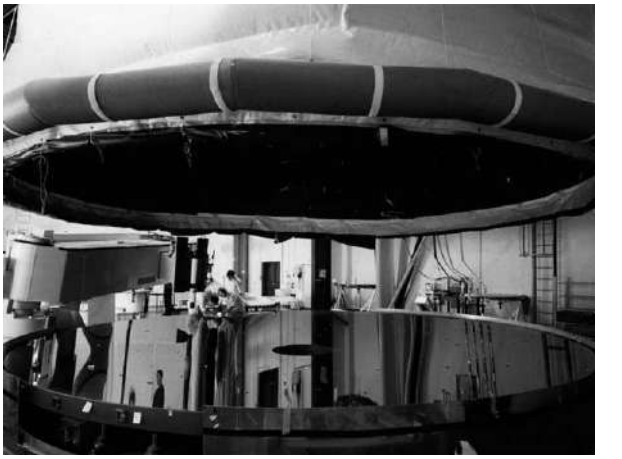

Figure 23 8.2 m thin primary mirror for the VLT. The VLT design is based on thin meniscus mirrors, made out of ZERODUR, which have low thermal inertia and which are always kept at ambient temperature, thus virtually removing the effect of thermal turbulence (mirror seeing). Seeing is the most limiting factor of ground-based optical astronomy. The optimized design of the VLT telescopes has introduced a significant leap forward to a largely improved image quality from the ground, and, consequently, improved sensitivity. The thin mirrors are constantly actively controlled in their mounting cells, using numerous piston actuators on the back surface of the mirror. Another key technology under development is 'adaptive optics', a method to compensate atmospheric wavefront deformations in real-time, thus obtaining diffraction limited images. Courtesy of European Southern Observatory.

waves, and the physical investigation of meteorites and other solar system material collected by space probes, telescopes (for the entire wavelength range of the electromagnetic spectrum) are the *only* tool to accomplish quantitative measurements of these cosmic laboratories.

In an attempt to unravel the history of the birth and evolution of the universe, to discover and to measure the first generation of stars and galaxies, and to find evidence for traces of life outside of planet Earth, plans are currently underway to develop yet another new generation of giant optical telescopes with apertures of 30–100 m in diameter (ELT: extremely large telescope). The technological challenges in terms of precision and stability are outstanding.

Along with the impressive growth of light collecting area, we must stress the importance of angular resolution, which has also experienced a significant evolution over the history of astronomical telescopes: from ≈ 5 arcsec of the Galilei telescope, which already provided an order of magnitude improvement over the naked eye, over roughly 1 arcsec of conventional telescopes in the 20th century, a few hundredths of an arcsec for the orbiting Hubble Space

Figure 24 Top: spin-cast 8.4 m honeycomb mirror #1 for the Large Binocular Telescope, the largest monolithic mirror of optical quality in the world. Shown in the process of applying a protective plastic coating to the end-polished surface. Bottom: LBT mirror #1, finally mounted in mirror cell. Courtesy of Large Binocular Telescope Observatory.

Telescope (Figure 25), down to milli-arcsec resolution of the Keck and VLT Interferometers. The development of the aforementioned ELTs is thought to be only meaningful if operated in a diffraction-limited mode, using the emerging technique of *adaptive optics*, which is a method to correct for wavefront distortions caused by atmospheric turbulence (image blurr) in real-time.

The combination of large optical/infrared telescopes with powerful detectors and focal plane instruments (*see* **Instrumentation:** Astronomical Instrumentation) has been a prerequisite to make astrophysics one of the most fascinating disciplines of fundamental sciences.

See also

Geometrical Optics: Aberrations. **Imaging:** Adaptive Optics. **Instrumentation:** Astronomical Instrumentation.

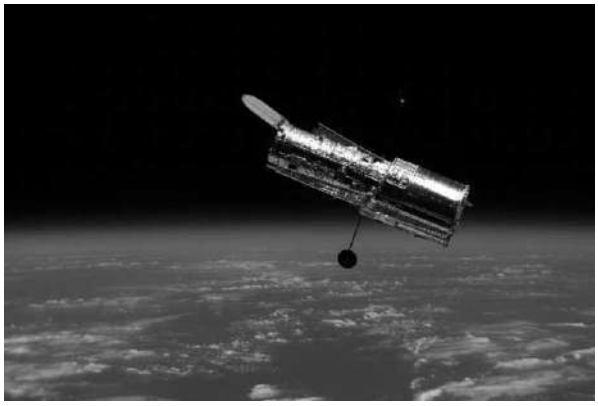

Figure 25 Hubble Space Telescope (HST) in orbit. This 2.4 m telescope is probably the most productive telescope which was ever built. Due to the absence of the atmosphere, the telescope delivers diffraction-limited, extremely sharp images, which have revealed unprecedented details of stars, star clusters, nebulae and other objects in the Milky Way and in other galaxies. Among the most spectacular results, HST has provided the deepest look into space ever, revealing light from faint galaxies that was emitted when the universe was only at 10% of its present age. Courtesy of Space Telescope Science Institute, Baltimore.

Further Reading

Bely PY (ed.) (2003) *The Design and Construction of Large Optical Telescopes*. New York: Springer.

Born M and Wolf E (1999) *Principles of Optics*, 7th ed. Cambridge: Cambridge University Press.

Haferkorn H (1994) *Optik*, 3rd ed. Leipzig: Barth Verlagsgesellschaft.

Korsch D (1991) *Reflective Optics*. San Diego: Academic Press.

Laux U (1999) *Astrooptik*, 2nd ed. Heidelberg: Verlag Sterne und Weltraum.

McCray WP (2004) *Giant Telescopes: Astronomical Ambition and the Promise of Technology*. Cambridge: Harvard University Press.

Osterbrock DE (1993) *Pauper and Prince: Richey, Hale, and Big American Telescopes*. Tucson: University of Arizona Press.

Riekher R (1990) *Fernrohre und ihre Meister*, 2nd ed. Berlin: Verlag Technik GmbH.

Rutten HGJ and van Venrooij MAM (1988) *Telescope Optics*. Richmond, VA: Willman-Bell.

Schroeder DJ (1987) *Astronomical Optics*. San Diego: Academic Press.

Welford WT (1974) *Aberrations of the Symmetrical Optical System*. London: Academic Press.

Wilson RN (1996) *Reflecting Telescope Optics I: Basic Design Theory and its Historical Development*. Berlin: Springer.

Wilson RN (1999) *Reflecting Telescope Optics II: Manufacture, Testing, Alignment, Modern Techniques*. Berlin: Springer.

INTERFEROMETRY

Contents

Overview
Gravity Wave Detection
Phase-Measurement Interferometry
White Light Interferometry

Overview

J C Wyant, University of Arizona, Tucson, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Interferometers are a powerful tool used in numerous industrial, research, and development applications, including measuring the quality of a large variety of manufactured items such as optical components and systems, hard disk drives and magnetic recording heads, lasers and optics used in CDs and DVD drives, cameras, laser printers, machined parts, components for fiber optics systems, and so forth. Interferometers can also be used to measure distance, spectra, and rotations, etc. The applications are almost endless.

There are many varieties of interferometers. Some interferometers form interference fringes using two beams and some use multiple beams. Some interferometers are common path; some use lateral shear; some use radial shear; and some use phase-shifting techniques. Examples of the various types are given below, but first we must give the basic equation for two-beam interference.

Two-Beam Interference

When two quasi-monochromatic waves interfere, the irradiance of the interference pattern is given by

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos[\phi] \qquad [1]$$

where the I's represent the irradiance of the individual beams and ϕ is the phase difference between the two interfering beams. The maximums of the interference occur when $\phi = 2m\pi$, where m is an integer. In most applications involving interferometers, ϕ is the quantity of interest since it is related to the quantity being measured. Figure 1 shows typical two-beam interference fringes.

There are many two-beam interferometers and we will now look at a few of them.

Fresnel Mirrors

Fresnel Mirrors are a very simple two beam interferometer. A beam of light reflecting off two mirrors, set at a slight angle to one another, will produce an interference pattern. The fringes are straight and equally spaced. The fringes become more finely spaced as the angle between the two mirrors increases, d increases, and D becomes smaller (Figure 2).

Plane Parallel Plate

When a monochromatic source illuminates a plane parallel plate, as shown in Figure 3, the interference

Figure 1 Typical two-beam interference fringes.

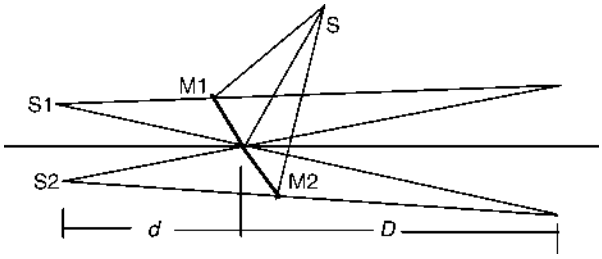

Figure 2 Fresnel mirrors.

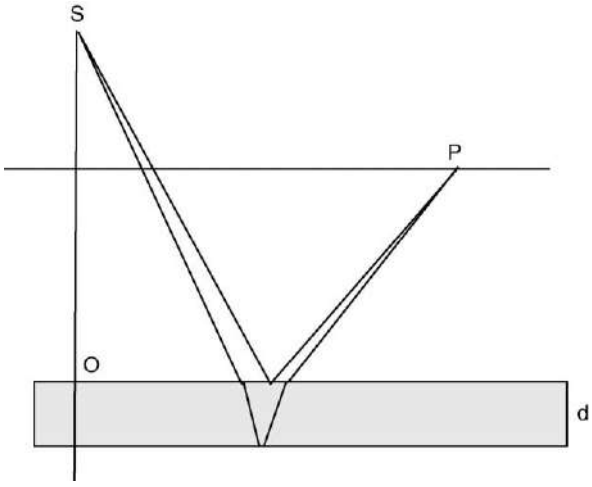

Figure 3 Plane parallel plate interferometer.

fringes are circles centered on the normal to plate. The fringes are called Haidinger fringes, or fringes of equal inclination since, for a given plate thickness, they depend on the angle of incidence. If the maximum or minimum occurs at the center, the radii of the fringes are proportional to the square root of integers. If the plate has a slight variation in thickness and it is illuminated with a collimated beam then the interference fringes are called Fizeau fringes of equal thickness and they give the thickness variations in the plate.

Michelson Interferometer

A Michelson interferometer is shown in Figure 4. Light from the extended source is split into two beams by the beamsplitter. The path difference can be viewed as the difference between the mirror M1 and the image of mirror M2, M2′. With the M1 and M2′ parallel, the fringes are circular and localized at infinity. With M1 and M2′ at a slight angle, the interference fringes are straight lines parallel to the equivalent intersection of the mirrors and localized approximately at the intersection. When the mirrors are only a few wavelengths apart, white light fringes

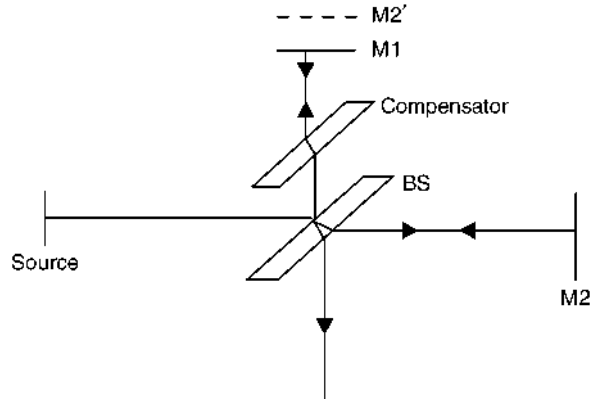

Figure 4 Michelson interferometer.

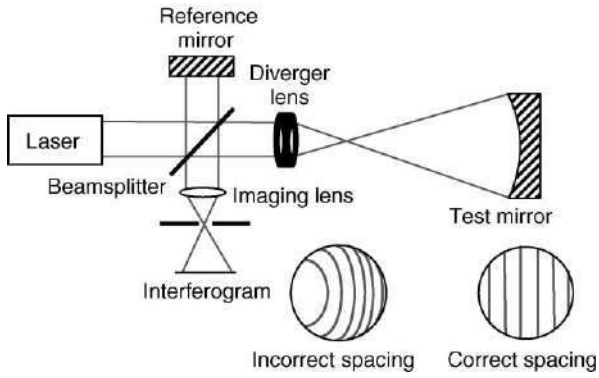

Figure 5 Twyman–Green interferometer for testing a concave spherical mirror.

appear, and can be used to determine their coincidence.

Twyman–Green Interferometer

If a point source is used in a Michelson interferometer it is generally called a Twyman–Green interferometer. Twyman–Green interferometers are often used to test optical components such as flat mirrors, curved mirrors, windows, lenses, and prisms. Figure 5 shows a Twyman–Green interferometer for the testing of a concave spherical mirror. The interferometer is aligned such that the focus of the diverger lens is at the center of curvature of the spherical mirror. If the mirror under test is perfect, straight equi-spaced fringes are obtained. Figure 6 shows two interference fringes and the relationship between surface height error for the mirror being tested and fringe deviation. λ is the wavelength of the light.

Fizeau Interferometer

The Fizeau interferometer is a simple device used for testing optical surfaces, especially flats and spheres

(Figure 7). The fringes are fringes of equal thickness. It is sometimes useful to tilt the reference surface a little to get several nearly straight interference fringes. The relationship between surface height error and fringe deviation from straightness is the same as for the Twyman–Green interferometer.

Mach–Zehnder Interferometer

A Mach–Zehnder interferometer is sometimes used to look at samples in transmission. Figure 8 shows a typical Mach–Zehnder interferometer for looking at a sample in transmission.

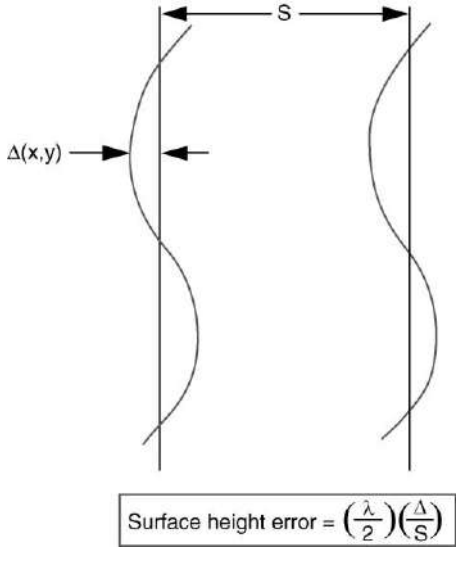

Figure 6 Relationship between surface height error and fringe deviation.

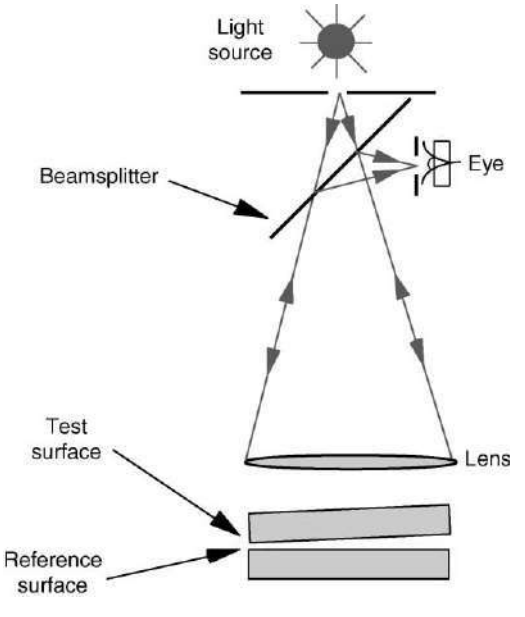

Figure 7 Fizeau interferometer.

Murty Lateral Shearing Interferometer

Lateral shearing interferometers compare a wavefront with a shifted version of itself. While there are many different lateral shear interferometers, one that works very well with a nearly collimated laser beam is a Murty shearing interferometer shown in Figure 9. The two laterally sheared beams are produced by reflecting a coherent laser beam off a plane parallel plate. The Murty interferometer can be used to measure the aberrations in the lens or it can be used to determine if the beam leaving the lens is collimated. If the beam incident upon the plane parallel plate is collimated a single fringe results, while if the beam is not perfectly collimated straight equi-spaced fringes result where the number of fringes gives the departure from collimation.

Radial Shear Interferometer

Radial shear interferometers compare a wavefront with a magnified or demagnified version of itself. While there are many different radial shear

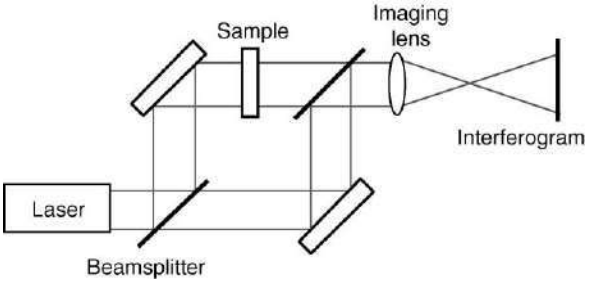

Figure 8 Mach–Zehnder interferometer.

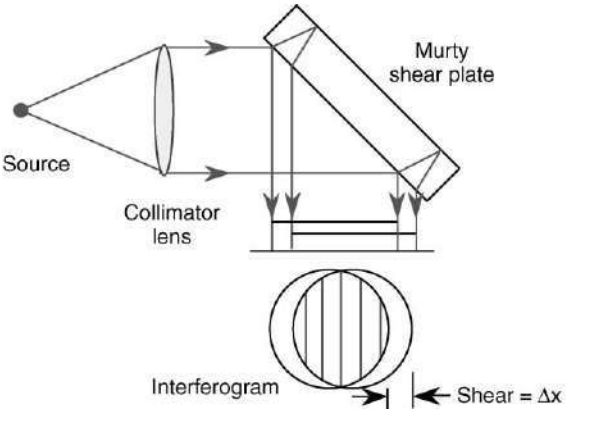

Figure 9 Murty lateral shear interferometer.

interferometers, one that works very well with a nearly collimated beam is shown in Figure 10. It is essentially a Mach–Zehnder interferometer where, in one arm, the beam is magnified and in the other arm the beam is demagnified. Interference fringes result in the region of overlap. The sensitivity of the interferometer depends upon the amount of radial shear. If the two interfering beams are approximately the same size there is little sensitivity, while if the two beams greatly differ in size there is large sensitivity. Sometimes radial shear interferometers are used test the quality of optical components.

Scatterplate Interferometer

A scatterplate interferometer, invented by Jim Burch in 1953, is one of the cleverest interferometers. Almost any light source can be used and no high-quality optics are required to do precision interferometry. The critical element in the interferometer is the scatterplate which looks like a piece of ground glass, but the important item is that the scattering points are arranged so the plate has inversion symmetry as shown in Figure 11. The light that is scattered the first time through the plate, and unscattered the second time through the plate, interferes with the light unscattered the first time and scattered the second time. The resulting interference fringes show errors in the mirror under test. This type of interferometer is insensitive to vibration because it is what is called a common path interferometer, in that the test beam (scattered–unscattered) and the reference beam (unscattered–scattered) travel along almost the same paths in the interferometer.

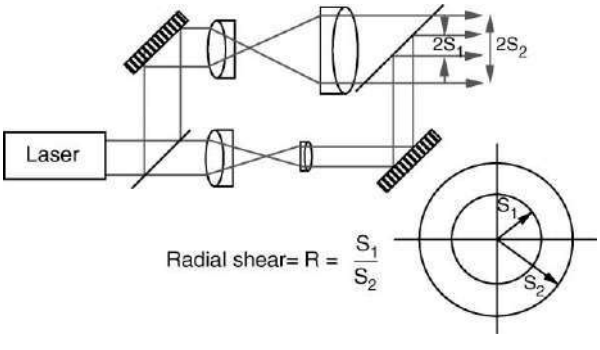

Figure 10 Radial shear interferometer.

Smartt Point Diffraction Interferometer

Another common path interferometer used for the testing of optics is the Smartt point diffraction interferometer (PDI) shown in Figure 12. In the PDI the beam of light being measured is focused onto a partially transmitting plate containing a pinhole. The pinhole removes the aberration from the light passing through it (reference beam), while the light passing through the plate (test beam) does not have the aberration removed. The interference of these two beams gives the aberration of the lens under test.

Sagnac Interferometer

The Sagnac interferometer has beams traveling in opposite directions, as shown in Figure 13. The interferometer is highly stable and easy to align. If the interferometer is rotated with an angular velocity there will be a delay between the transit times of the clockwise and the counterclockwise beams. For this reason, the Sagnac interferometer is used in laser gyros.

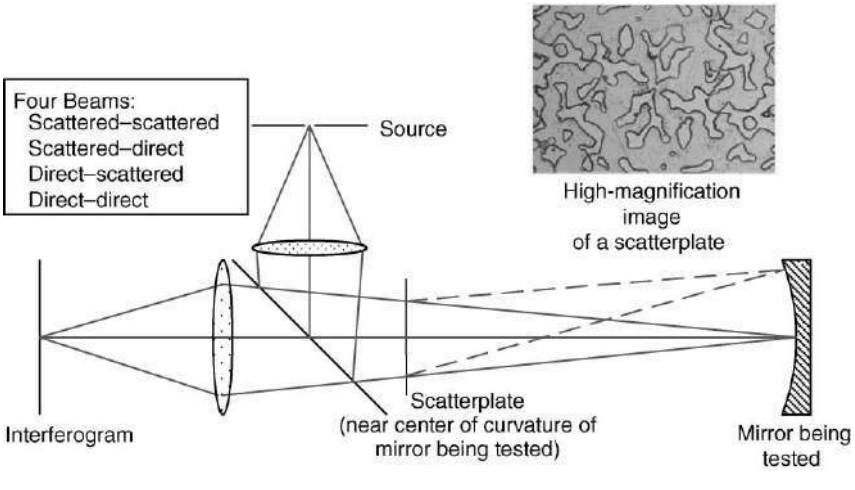

Figure 11 Scatterplate interferometer.

Fiber Interferometers

Fiber interferometers were first used for rotation sensing by replacing the ring cavity in the Sagnac interferometer with a multiloop made of a single-mode fiber. Fiber-interferometer rotation sensors are attractive as rotation sensors because they are small and low-cost.

Since the optical path in a fiber is affected by its temperature and it also changes when the fiber is stretched, fiber interferometers can be used as sensors for mechanical strains, temperature, and pressure. They can also be used for the measurement of magnetic fields by bonding the fiber to a magneto-restrictive element. Electric fields can be measured by bonding the fiber to a piezoelectric film.

Multiple Beam Interferometers

In general, multiple beam interferometers provide sharper interference fringes than a two beam interferometer. If light is reflected off a plane parallel plate there are multiple reflections. If the surfaces of the plane parallel plate have a low reflectivity, the higher-order reflections have a low intensity and the multiple reflections can be ignored and the resulting interference fringes have a sinusoidal intensity profile, but if the surfaces of the plane parallel plate have a high reflectivity, the fringes become sharper. Figure 14 shows the profile of the fringes for both the reflected and the transmitted light for different values of surface reflectivity. Multiple beam interference fringes are extremely useful for measuring the spectral content of a source. One such multiple beam interferometer, that is often used for measuring spectral distributions of a source, is a Fabry–Perot interferometer that consists of two plane-parallel plates separated by an air space, as shown in Figure 15.

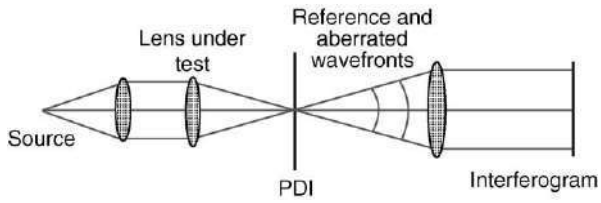

Figure 12 Smartt point diffraction interferometer.

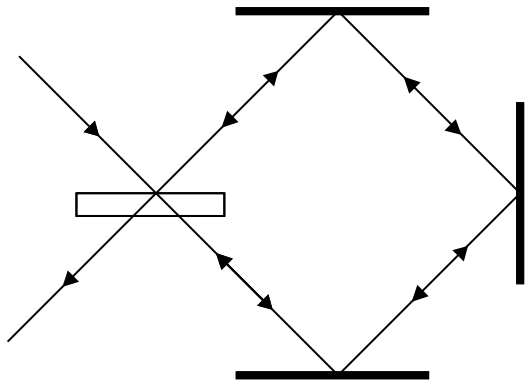

Figure 13 Sagnac interferometer.

Phase-Shifting Interferometry

The equation for the intensity resulting from two-beam interference, contains three unknowns, the two individual intensities and the phase difference between the two interfering beams. If three or more measurements are made of the intensity of the two-beam interference as the phase difference between the two beams is varied in a known manner, all three unknowns can be determined. This technique, called phase-shifting interferometry, is an extremely powerful tool for measuring phase distributions. Generally the phase difference is

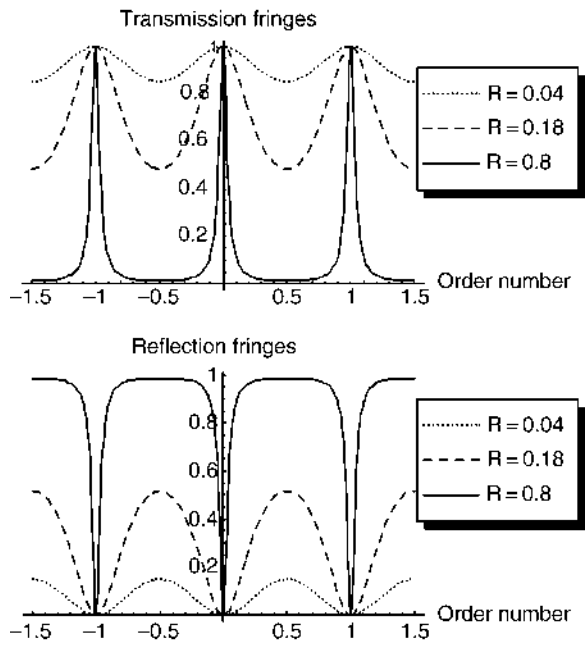

Figure 14 Multiple beam interference fringes for the transmitted and reflected light for a plane parallel plate.

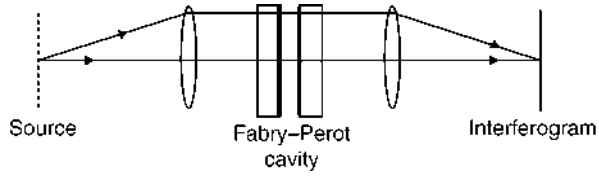

Figure 15 Fabry–Perot interferometer.

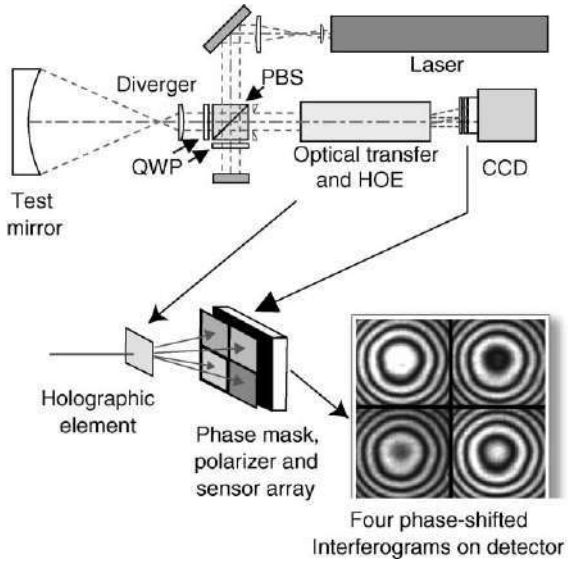

Figure 16 Vibration-insensitive phase-shifting interferometer.

changed by 90 degrees between consecutive measurements of the interference intensity in which case the three intensities can be written as

$$I_a(x,y) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\phi(x,y) - 90^\circ)$$
$$I_b(x,y) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\phi(x,y))$$
$$I_c(x,y) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\phi(x,y) + 90^\circ)$$

and the phase distribution is given by

$$\phi(x,y) = \mathrm{ArcTan}\left(\frac{I_a(x,y) - I_c(x,y)}{-I_a(x,y) + 2I_b(x,y) - I_c(x,y)}\right)$$

Phase-shifting interferometry has greatly enhanced the use of interferometry in metrology since it provides a fast low noise way of getting interferometric data into a computer.

Vibration-Insensitive Phase-Shifting Interferometer

While phase-shifting interferometry has greatly enhanced the use of interferometry in metrology, there are many applications where it cannot be used because of the environment, especially vibration. One recent phase-shifting interferometer that works well in the presence of vibration is shown in Figure 16. In this interferometer, four phase-shifted frames of data are captured simultaneously. The test and reference beams have orthogonal polarization. After the test and reference beams are combined they are passed through a holographic element to produce four identical beams. The four beams are then transmitted through wave retardation plates to cause 0, 90, 180, and 270 phase difference between the test and reference beams. After passing through a polarizer the four phase-shifted interferograms fall on the detector array. Not only can the effects of vibration be eliminated, but by making short exposures to freeze the vibration, the vibrational modes can be measured. Movies can be made showing the vibration. Likewise, flow fields can be measured.

Combining modern electronics, computers, and software with old interferometric techniques, provides for very powerful measurement capabilities.

List of Units and Nomenclature

I	irradiance
R	intensity reflectance
λ	wavelength
ϕ	phase

See also

Coherence: Speckle and Coherence. **Detection:** Fiber Sensors. **Holography, Techniques:** Computer–Generated Holograms; Digital Holography; Holographic Interferometry. **Imaging:** Adaptive Optics; Interferometric Imaging; Wavefront Sensors and Control (Imaging Through Turbulence). **Interferometry:** Phase Measurement Interferometry; White Light Interferometry; Gravity Wave Detection. **Microscopy:** Interference Microscopy. **Tomography:** Optical Coherence Tomography.

Further Reading

Candler C (1951) *Modern Interferometers*. London: Hilger and Watts.
Francon M (1966) *Optical Interferometry*. New York: Academic Press.
Hariharan P (1991) Selected papers on interferometry. *SPIE*, vol. 28.
Hariharan P (1992) *Basics of Interferometry*. San Diego: Academic Press.
Hariharan P (2003) *Optical Interferometry*. San Diego: Academic Press.
Hariharan P and Malacara D (1995) Selected papers on interference, interferometry, and interferometric metrology. *SPIE*, vol. 110.
Malacara D (1990) Selected papers on optical shop metrology. *SPIE*, vol. 18.
Malacara D (1992) *Optical Shop Testing*. New York: Wiley.
Malacara D (1998) *Interferogram Analysis for Optical Testing*. New York: Marcel Decker.
Robinson D and Reid G (1993) *Interferogram Analysis*. Brisol, UK: Institute of Physics.
Steel WH (1985) *Interferometry*. Cambridge: Cambridge University Press.
Tolansky S (1973) *An Introduction to Interferometry*. New York: Longman.

Gravity Wave Detection

N Christensen, Carleton College, Northfield, MN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In the later part of the nineteenth century Albert Michelson performed extraordinary experiments that shook the foundations of the physics world. Michelson's precise determination of the speed of light was an accomplishment that takes great skill to reproduce today. Edward Morley teamed up with Michelson to measure the velocity of the Earth with respect to the aether. The interferometer that they constructed was exquisite, and through amazing experimental techniques the existence of the aether was disproved. The results of Michelson and Morley led to a revolution in physics, and provided evidence that helped Albert Einstein to develop the general theory of relativity. Now the Michelson interferometer may soon provide dramatic confirmation of Einstein's theory through the direct detection of gravitational radiation.

An accelerating electric charge produces electromagnetic radiation – light. It should come as no surprise that an accelerating mass produces gravitational light, namely gravitational radiation. In 1888 Heinrich Hertz had the luxury to produce and detect electromagnetic radiation in his laboratory. There will be no such luck with gravitational radiation because gravity is an extremely weak force.

Albert Einstein postulated the existence of gravitational radiation in 1916, and in 1989 Joe Taylor and Joel Weisberg indirectly confirmed its existence through observations of the orbital decay of the binary pulsar 1913 + 16. Direct detection of gravity waves will be difficult. For humans to have any chance of detecting gravity waves there must be extremely massive objects accelerating up to relativistic velocities. The only practical sources are astrophysical: supernovae, pulsars, neutron star or black hole binary systems, black hole formation or even the Big Bang. The observation of these types of events would be extremely significant for contributing to knowledge in astrophysics and cosmology. Gravity waves from the Big Bang would provide information about the Universe at its earliest moments. Observations of supernovae will yield a gravitational snapshot of these extreme cataclysmic events. Pulsars are neutron stars that can spin on their axes at frequencies in the hundreds of hertz, and the signals from these objects will help to decipher their characteristics. Gravity waves from the final stages of coalescing binary neutron stars could help to accurately determine the size of these objects and the equation of state of nuclear matter. The observation of black hole formation from these binary systems would be the *coup de grace* for the debate on the existence of black holes, and the ultimate triumph for general relativity.

Electromagnetic radiation has an electric field transverse to the direction of propagation, and a charged particle interacting with the radiation will experience a force. Similarly, gravitational radiation will produce a transverse force on massive objects, a tidal force. Explained via general relativity it is more accurate to say that gravitational radiation will deform the fabric of space-time. Just like electromagnetic radiation there are two polarizations for gravity waves. Let us imagine a linearly polarized gravity wave propagating in the z-direction, $h(z,t) = h_{0+}e^{i(kz-\omega t)}$. The fabric of space is stretched due to the strain created by the gravity wave. Consider a length L_0 of space along the $\boldsymbol{x}$-axis. In the presence of the gravity wave the length oscillates like

$$L(t) = L_0 + \frac{h_{0+}L}{2}\cos\omega t$$

hence there is a change in its length of

$$\Delta L_x = \frac{h_{0+}L}{2}\cos\omega t$$

A similar length L_0 of the $\boldsymbol{y}$-axis oscillates, like

$$L(t) = L_0 - \frac{h_{0+}L}{2}\cos\omega t$$

or

$$\Delta L_y = \frac{-h_{0+}L}{2}\cos\omega t$$

One axis stretches while the perpendicular one contracts, and then vice versa, as the wave propagates through. The other polarization (h_{0x}) produces a strain on axes 45° from ($\boldsymbol{x}$,$\boldsymbol{y}$). Imagine some astrophysical event produces a gravity wave that has amplitude h_{0+} on Earth; in order to detect a small distance displacement ΔL one should have a detector that spans a large length L. A supernova within our own Galaxy might possibly produce a gravity wave of size $h \sim 10^{-18}$ with characteristic frequencies around

1 kHz, but the occurrence of an event such as this will be extremely rare. More events will come from novae in other galaxies, but due to the great distances and the fact that the magnitude of the gravity wave falls of as $1/r$, such events will be substantially diminished in magnitude.

A Michelson interferometer, with arms aligned along the x- and y- axes, can measure small phase differences between the light in the two arms. Therefore, this type of interferometer can turn the length variations of the arms produced by a gravity wave into changes in the interference pattern of the light exiting the system. This was the basis of the idea from which modern laser interferometric gravitational radiation detectors have evolved. Imagine a gravity wave of amplitude h is incident on an interferometer. The change in the arm length will be $h \sim \Delta L/L_0$, so in order to optimize the sensitivity it is advantageous to make the interferometer arm length L_0 as large as possible. Detectors coming on-line are attempting to measure distance displacements that are of order $\Delta L \sim 10^{-18}$ m or smaller, less that the size of an atomic nucleus! If the interferometers can detect such distance displacements, and directly detect gravitational radiation, it will be one of the most spectacular accomplishments in experimental physics. The first implementation of a laser interferometer to detect a gravity wave was by Forward who used earphones to listen to the motion of the interference signal. The engineering of signal extraction for modern interferometers is obviously far more complex.

Numerous collaborations are building and operating advanced interferometers in order to detect gravitational radiation. In the United States there is the Laser Interferometer Gravity Wave Observatory (LIGO), which consists of two 4-km interferometers located in Livingston, Louisiana and Hanford, Washington. In addition, there is an additional 2-km interferometer within the vacuum system at Hanford. An Italian–French collaboration (VIRGO) has a 3-km interferometer near Pisa, Italy. GEO, a German–British collaboration, is a 600-m detector near Hanover, Germany. TAMA is the Japanese 300-m interferometer in Tokyo. The Australians (AIGO) are constructing a 500-m interferometer in Western Australia. All of the detectors will be attempting to detect gravity waves with frequencies from about 50 Hz to 1 kHz.

As will be described below, there are a number of terrestrial noise sources that will inhibit the performance of the interferometric detectors. The sensitivity of detection increases linearly with interferometer arm length, which implies that there could be advantages to constructing a gravity wave detector in space. This is the goal of the Laser Interferometer Space Antenna (LISA) collaboration. The plan is to deploy three satellites in a heliocentric orbit with a separation of about 5×10^6 km. The launch for LISA is planned for around 2015, and the detector will be sensitive to gravity waves within the frequency range of 10^{-3} to 10^{-1} Hz. Due to the extremely long baseline, LISA is not strictly an interferometer, as most light will be lost as the laser beams expand while traveling such a great distance. Instead, the phase of the received light will be detected and used to lock the phase of the light that is re-emitted by another laser.

Joe Weber pioneered the field of gravitational radiation detection in the 1960s. He used a 1400-kg aluminum cylinder as an antenna. Gravitational waves would hopefully excite the lowest-order normal mode of the bar. Since the gravity wave is a strain on space-time, experimentalists win by making their detectors longer in length. This provides a long-term advantage for laser interferometers, which can scale in length relatively easily. However, as of 2002, interferometers and bars (now cooled to 4 K) have comparable sensitivities.

Interferometer Styles

The Michelson interferometer is the tool to be used to detect a gravitational wave. Figure 1 shows a basic system. The beamsplitter and the end mirrors would be suspended by wires, and effectively free to move in the plane of the interferometer. The arms have lengths L_1 and L_2 that are roughly equal on a kilometer scale. With a laser of power P and wavelength λ incident on the beamsplitter, the light

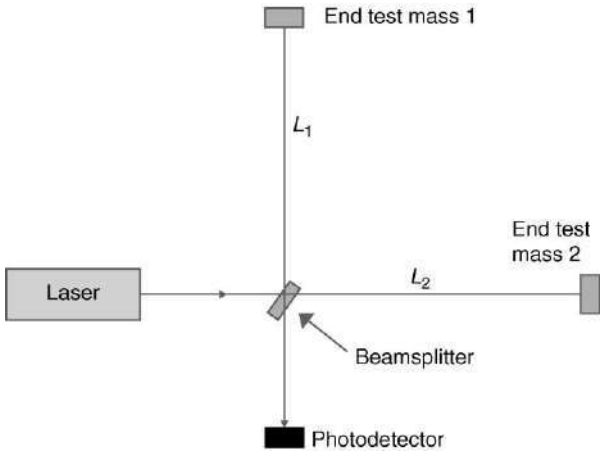

Figure 1 A basic Michelson interferometer. The photodetector receives light exiting the dark port of the interferometer and hence the signal.

exiting the dark port of the interferometer is

$$P_{\text{out}} = P \sin^2\left[\frac{2\pi}{\lambda}(L_1 - L_2)\right]$$

The interferometer operates with the condition that in the absence of excitation the light exiting the dark port is zero. This would be the case for a simple and basic interferometer. However, interferometers like those in LIGO will be more sophisticated, and will use a heterodyne detection strategy. If E_0 is the amplitude of the electric field from the laser, and assuming the use of a 50–50 beamsplitter, the electric field (neglecting unimportant common phase shifts) for the light incident on the photodetector would be

$$E_{\text{out}} = \frac{E_0}{2}\left(e^{i\delta\phi_1} - e^{i\delta\phi_2}\right) \approx i\frac{E_0}{2}(\phi_1 - \phi_2)$$
$$= iE_0\frac{2\pi}{L}(L_1 - L_2)$$

The laser light will be phase modulated at a frequency in the MHz regime. As such, the deconvolved current from the photodiode that detects light at the interferometer output will be proportional to the phase acquired by the light, namely $I \propto \frac{2\pi}{\lambda}(L_1 - L_2)$.

A gravity wave of optimal polarization normally incident upon the interferometer plane will cause one arm to decrease in length while the other increases. The Michelson interferometer acts as a gravity wave transducer; the change in arm lengths results in more light exiting the interferometer dark port. The mirrors in the interferometer are suspended via wires so that they are free to move under the influence of the gravity wave.

An interferometer's sensitivity increases with arm length, but geographical and financial constraints will limit the size of the arms. If there could be some way to bounce the light back and forth to increase the effective arm length it would increase the detector performance. Fabry–Perot cavities on resonance light have a storage time of $2L/[c(1 - \sqrt{R_1R_2})]$. Figure 2 shows the system of a Michelson interferometer with Fabry–Perot cavities. This gravity wave interferometer design was proposed and tested in the late 1970s by Ron Drever. The far mirror R_2 has a very high reflectivity ($R_2 \sim 1$) in order to ultimately direct the light back towards the beamsplitter. The front mirror reflectivity R_1 is such that LIGO's effective arm length increases to $L \sim 300$ km. The optical properties of the mirrors of the Fabry–Perot cavities must be exquisite in order to achieve success. LIGO's mirrors were tested, and the root mean squared surface uniformity is less than 1 nm, scattered light is less than 50 parts per million (ppm), absorption is less than 2 ppm, and the radii of curvature for the mirrors are matched to less than 3%. A LIGO test mass (and therefore a Fabry–Perot mirror) can be seen in Figure 3.

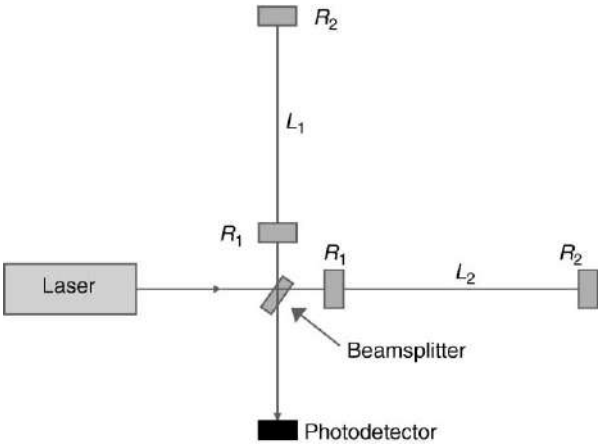

Figure 2 A Michelson interferometer with Fabry–Perot cavities in each arm. The front cavity mirrors have reflectivity R_1 while the end mirrors have $R_2 \sim 1$. By using Fabry–Perot cavities LIGO will increase the effective arm length by a factor of 75.

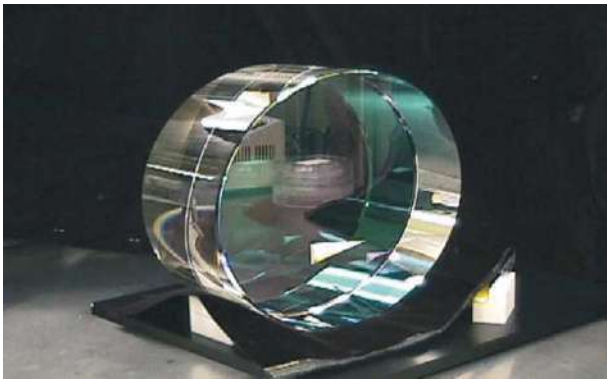

Figure 3 A picture of a mirror and test mass for LIGO. The fused silica component is 10.7 kg in mass and 25 cm in diameter. Photograph courtesy of Caltech/LIGO.

In 1888 Michelson and Morley, with their interferometer, had a sensitivity that allowed the measurement of 0.02 of a fringe, or about 0.126 radian. Prototype interferometers constructed by the LIGO science team have already demonstrated a phase noise spectral density of $\phi(f) = 10^{-10}$ radian/$\sqrt{\text{Hz}}$ for frequencies above 500 Hz. Assuming a 1 kHz signal with 1 kHz bandwidth this implies a phase sensitivity of $\Delta\phi = 3.2 \times 10^{-9}$ radian. This is about the phase sensitivity that LIGO hopes to accomplish in the 4-km Fabry–Perot system. There has been quite an evolution in interferometry since Michelson's time.

The noise sources that inhibit the interferometer performance are discussed below. However, let us

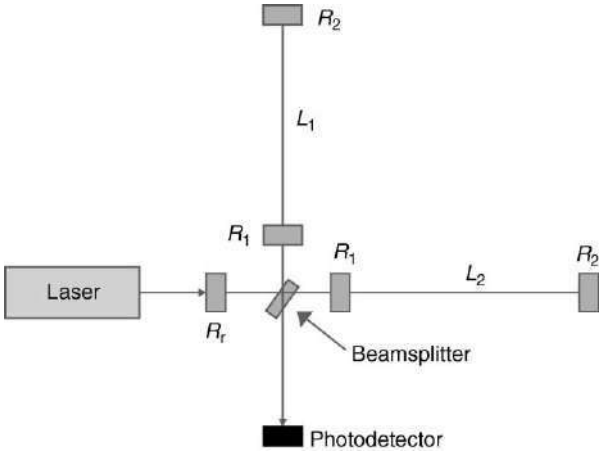

Figure 4 A power recycled Michelson interferometer with Fabry–Perot cavities in each arm. Normally light would exit the interferometer through the light port and head back to the laser. Installation of the recycling mirror with reflectivity R_r sends the light back into the system. A Fabry–Perot cavity is formed between the recycling mirror and the first mirror (R_1) of the arms. For LIGO this strategy will increase the power circulating in the interferometer by a factor of 50.

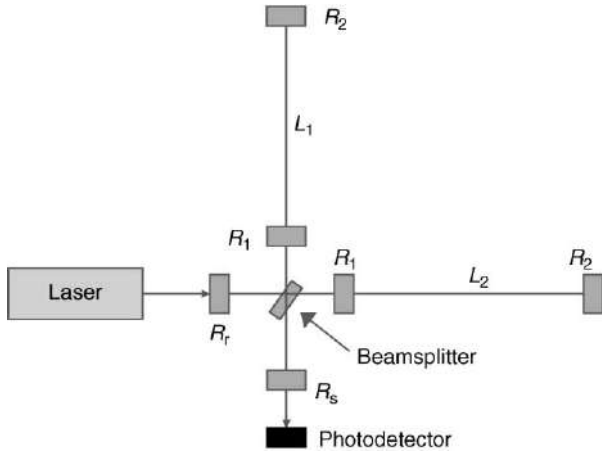

Figure 5 A signal recycled and power recycled Michelson interferometer with Fabry–Perot cavities in each arm. Normally light containing the gravity wave signal would exit the interferometer through the dark port and head to the photodetector. Installation of the signal recycling mirror with reflectivity R_s sends the light back into the system. The phase of the light acquired from the gravity wave will build up at a particular frequency determined by the reflectivity R_s.

consider one's ability to measure the relative phase between the light in the two arms. The Heisenberg uncertainty relation for light with phase ϕ and photon number N is $\Delta\phi\Delta N \sim 1$. For a measurement lasting time τ using laser power P and frequency f, the photon number is $N = P\lambda\tau/hc$, and with Poisson statistics describing the light $\Delta N = \sqrt{N} = \sqrt{P\lambda\tau/hc}$. Therefore

$$\Delta\phi\Delta N = \frac{2\pi}{\lambda}\Delta L\sqrt{P\lambda\tau/hc} = 1$$

implies that

$$\Delta L = \frac{1}{2\pi}\sqrt{\frac{hc\lambda}{P\tau}}$$

With more light power the interferometer can measure smaller distance displacements and achieve better sensitivity. LIGO will use about 10 W of laser power, and will eventually work towards 100 W. However, there is a nice trick one can use to produce more light circulating in the interferometer, namely power recycling. Figure 4 displays the power recycling interferometer design. The interferometer operates such that virtually none of the light exits the interferometer dark port, and the bulk of the light returns towards the laser. An additional mirror, R_r in Figure 4, recycles the light. For LIGO, recycling will increase the effective light power by another factor of 50. The higher circulating light power therefore improves the sensitivity.

There is one additional modification to the interferometer system that can further improve sensitivity, but only at a particular frequency. A further Fabry–Perot system can be made by installing what is called a signal recycling mirror; this would be mirror R_s in Figure 5. Imagine light in arm 1 on the interferometer that acquires phase as the arm expands due to the gravity wave. The traveling gravity wave's oscillation will subsequently cause arm 1 to contract while arm 2 expands. If the light that was in arm 1 could be sent to arm 2 while it is expanding, then the beam would acquire additional phase. This process could be repeated over and over. Mirror R_s serves this purpose, with its reflectivity defining the storage time for light in each interferometer arm. The storage time defined by the cavity formed by the signal recycling mirror, R_s, and the mirror at the front of the interferometer arm cavity, R_1, determines the resonance frequency. Signal recycling will give a substantial boost to interferometer sensitivity at a particular frequency, and will eventually be implemented in all the main ground-based interferometric detectors.

The LIGO interferometers are infinitely more complex than the relatively simple systems displayed in the figures of this paper. Figure 6 presents an aerial view of the LIGO site at Hanford, Washington. The magnitude of the 4-km system is apparent.

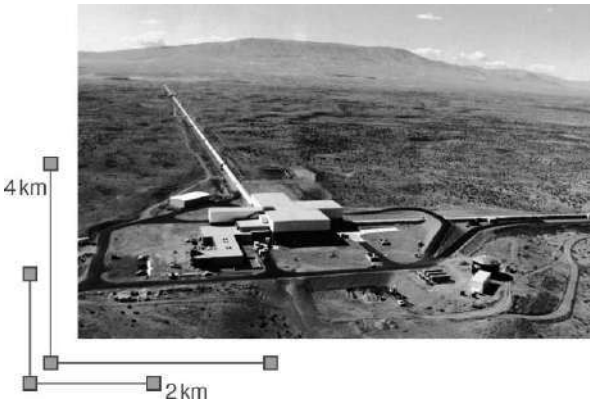

Figure 6 Arial view of the LIGO Hanford, Washington site. The vacuum enclosure at Hanford contains both 2-km and 4-km interferometers. Photograph courtesy of Caltech/ LIGO.

Noise Sources and Interferometer Sensitivity

If the interferometers are to detect distance displacements of order $\Delta L \sim 10^{-18}$ m then they must be isolated from a host of deleterious noise sources. Seismic disturbances should not shake the interferometers. Thermal excitation of components will affect the sensitivity of the detector and should be minimized. The entire interferometer must be in an adequate vacuum in order to avoid fluctuations in gas density that would cause changes in the index of refraction and hence modification of the optical path length. The laser intensity and frequency noise must be minimized. The counting statistics of photons influences accuracy. If ever there was a detector that must avoid Murphy's law this is it; little things going wrong cannot be permitted if such small distance displacements are to be detected. The expected noise sensitivity for the initial LIGO interferometers is displayed in Figure 7.

In the best of all worlds the interferometer sensitivity will be limited by the counting statistics of the photons. A proper functioning laser will have its photon number described by Poisson statistics, or shot noise; if the mean number of photons arriving per unit time is N then the uncertainty is $\Delta N = \sqrt{N}$, which as noted above implies an interferometer displacement sensitivity of

$$\Delta L = \frac{1}{2\pi}\sqrt{\frac{hc\lambda}{P\tau}}$$

or a noise spectral density of

$$\Delta L(f) = \frac{1}{2\pi}\sqrt{\frac{hc\lambda}{P}} \quad (\text{in units of m}/\sqrt{\text{Hz}})$$

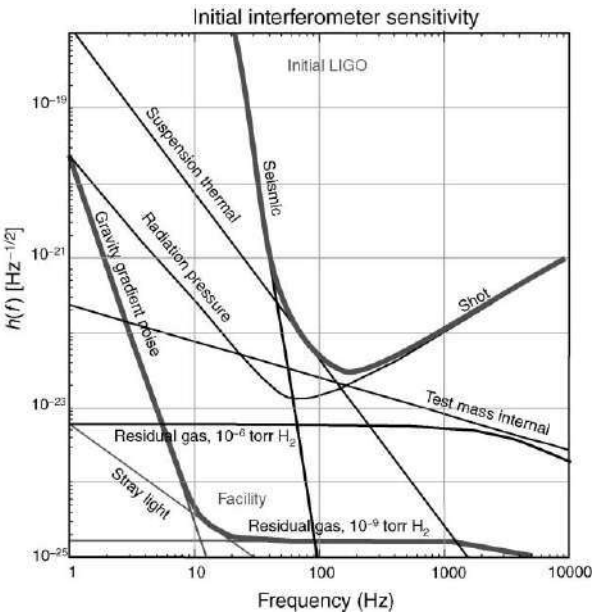

Figure 7 The target spectral density of the noise for the initial LIGO system. LIGO will be dominated by seismic noise at low frequencies (10–100 Hz), thermal noise (from the suspension system and internal modes within the mirrors) in the intermediate regime (100–300 Hz), and photon shot noise thereafter. Other sources of noise are also noted, specifically gravity gradients (gravitational excitation of masses from the seismic motion of the ground), radiation pressure of photons on the mirrors, stray (scattered) light, and index of refraction fluctuations from residual gas in the vacuum.

Note also that the sensitivity increases as the light power increases. The reason for this derives from the statistics of repeated measurements. The relative lengths of the interferometer arm lengths could be measured, once, by a photon. However, the relative positions are measured repeatedly with every photon from the laser, and the variance of the mean decreases as $\sqrt{N}$, where N is the number of measurements (or photons) involved. The uncertainty in the difference of the interferometer arm lengths is therefore inversely proportional to photon number, and hence the laser's power. In terms of strain sensitivity this would imply

$$h(f) = \frac{1}{2\pi L}\sqrt{\frac{hc\lambda}{P}} \quad (\text{in units of } 1/\sqrt{\text{Hz}})$$

This assumes the light just travels down the arm and back once. With Fabry–Perot cavities the light is stored, and the typical photon takes many trips back and forth before exiting the system. In order to maximize light power the end mirrors ($R_2 \sim 1$) and

the strain sensitivity is improved to

$$h(f) = \frac{1}{2\pi L'}\sqrt{\frac{hc\lambda}{P}}$$

where $L' = 4L/(1 - R_1)$.

As the frequency of gravity waves increases the detection sensitivity will decrease. If the gravity wave causes the interferometer arm length to increase, then decrease, while the photons are still in the arm cavity, then the phase acquired from the gravity wave will be washed away. This is the reason why interferometer sensitivity decreases as frequency increases, and explains the high-frequency behavior seen in Figure 7. Taking this into account, the strain sensitivity is

$$h(f) = \frac{1}{2\pi L'}\sqrt{\frac{hc\lambda}{P}}\left(1 + \left(\frac{2\pi L' f}{c}\right)^2\right)^{1/2}$$

where $L' = 4L/(1 - R_1)$ and f is the frequency of the gravity wave.

If the gravitational wave is to change the interferometer arm length then the mirrors that define the arm must be free to move. In systems like LIGO wires suspend the mirrors; each mirror is like a pendulum. While allowing the mirrors to move under the influence of the gravity wave is a necessary condition, the pendulum itself is the first component of an elaborate vibration isolation system. Seismic noise will be troublesome for the detector at low frequencies. The spectral density of the seismic noise is about $(10^{-7}/f^2)\mathrm{m}/\sqrt{\mathrm{Hz}}$ for frequencies above 1 Hz. A simple pendulum, by itself, acts as a motion filtering device. Above its resonance frequency the pendulum filters motion with a transfer function like $T(f) \propto (f_0/f)^2$. Detectors such as LIGO will have a pendulum with resonant frequencies of about $f_0 \approx 1$ Hz, thus providing an isolation of 10^4 when looking for signals at $f = 100$ Hz. The various gravity wave detector collaborations have different vibration isolation designs. The mirrors in these interferometers will be suspended in elaborate vibration isolation systems, which may include multiple pendulum and isolation stacks. Seismic noise will be the limiting factor for interferometers seeking to detect gravity waves in the tens of hertz range, as can be seen in the sensitivity curve presented in Figure 7.

Due to the extremely small distance displacements that these systems are trying to detect it should come as no surprise that thermal noise is a problem. This noise enters through a number of components in the system. The two most serious thermal noise sources are the wires suspending the mirrors in the pendulum, and the mirrors themselves. Consider the wires; there are a number of modes that can oscillate (i.e. violin modes). At temperature T each mode will have energy of $k_B T$, but distributed over a band of frequencies determined by the quality factor (or Q) of the material. Low-loss (or high-Q) materials work best; for the violin modes of the wires there will be much noise at particular frequencies (in the hundreds of hertz). The mirror is a cylindrical object, which will have normal modes of oscillation that can be thermally excited. The first generation of LIGO will have these masses composed of fused silica, which is typical for optical components. The Qs for the internal modes are greater than 2×10^6. A switch may eventually be made to sapphire mirrors, which have better thermal properties. The limitation to the interferometers' sensitivity due to the thermal noise internal to the mirrors can be seen in Figure 7, and will be a worrying noise source for the first-generation LIGO in the frequency band around 100–400 Hz.

The frequency noise of the laser can couple into the system to produce length displacement sensitivity noise in the interferometer. With arm lengths of 4 km, it will be impossible to hold the length of the two arms absolutely equal. The slightly differing arm spans will mean that the light sent back from each of the two Fabry–Perot cavities will have slightly differing phases. As a consequence, great effort is made to stabilize the frequency of the light entering the interferometer. The LIGO laser can be seen in Figure 8. The primary laser is a Lightwave Model 126 nonplanar ring oscillator. High power is generated from this stabilized laser through the use of optical amplifiers. The beam is sent through four optical amplifiers, and then retro-reflected back

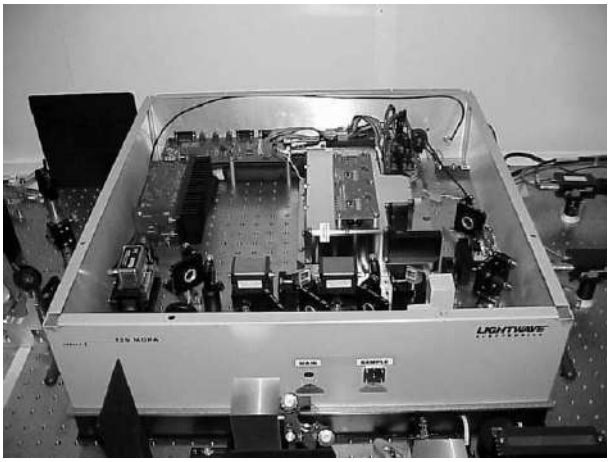

Figure 8 A picture of the Nd:YAG laser and amplifier system that produces 10 W of light for LIGO. Photograph courtesy of Caltech/LIGO.

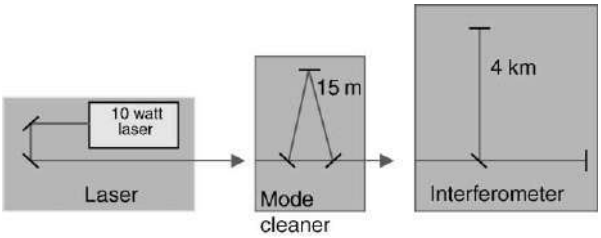

Figure 9 The optical system for the LIGO system. Light moves from the frequency-stabilized laser to a mode cleaner, and then to the interferometer. Presently 8 W of TEM_{00} laser power is delivered to the interferometer.

through the amplifiers again. For LIGO, the laser is locked and held to a specific frequency by use of signals from a reference cavity, a mode cleaner cavity, and the interferometer. For low-frequency stabilization the temperature of the ring oscillator is adjusted. At intermediate frequencies adjustment is made by signals to a piezo-electric transducer within the ring oscillator cavity. At high frequencies the noise is reduced with the use of an electro-optic crystal. The LIGO lasers currently have a frequency noise of $2 \times 10^{-2}\,\mathrm{Hz}/\sqrt{\mathrm{Hz}}$ at frequencies above 1 kHz.

It will prove important to worry about the stability of the laser power for the interferometric detectors. The hope is to be shot noise limited at frequencies above a few hundred hertz. The Nd:YAG power amplifiers used are pumped with an array of laser diodes, so the light power is controlled through feedback to the laser diodes. The LIGO requirements for the fluctuations on the power P are $\Delta P/P < 10^{-8}/\sqrt{\mathrm{Hz}}$. The spatial quality of the light is ensured through the use of a mode-cleaning cavity. LIGO uses a triangular array of mirrors separated by 15 m. LIGO's input optical system can be seen in Figure 9. The current LIGO optical system yields 8 W of 1.06 μm light in the TEM_{00} mode.

Conclusion

The attempt to measure gravitational radiation with laser interferometers could possibly be the most difficult optical experiment of our time. Over a hundred years ago Michelson succeeded in carrying off experiments of amazing difficulty as he measured the speed of light and disproved the existence of the aether. Gravity wave detection is an experiment worthy of Michelson, and there are hundreds of physicists striving to make it a reality.

Great success has already been achieved. The TAMA 300-m interferometer is operational and has achieved a sensitivity $h(f) = 5 \times 10^{-21}/\sqrt{\mathrm{Hz}}$ in the 700 Hz to 1.5 kHz frequency band. The LIGO interferometers are now operating, and scientists are presently de-bugging the system in order to achieve the target sensitivity. The first scientific data taking for LIGO commenced in 2002, and is continuing at the time of writing.

This will be a new telescope to peer into the heavens. With every new means of looking at the sky there has come unexpected discoveries. Physicists do know that there will be signals that they can predict. Binary systems of compact objects (neutron stars or black holes) will produce chirp signals that may be extracted by matched filtering techniques. A supernova will produce a burst that will hopefully rise above the noise. Pulsars, or neutron stars spinning about their axes at rates sometimes exceeding hundreds of revolutions per second, will produce continuous sinusoidal signals that can be seen by integrating for sufficient lengths of time. Gravity waves produced by the Big Bang will produce a background stochastic noise that can possibly be extracted by correlating the outputs from two or more detectors. These are exciting physics results that will come through tremendous experimental effort.

See also

Imaging: Interferometric Imaging. **Interferometry:** Overview.

Further Reading

Barish BC (1997) Gravitational wave detection. In: Tsubono K, Fujimoto MK and Kurodo K (eds) *Gravitational Wave Detection*, pp. 155–161. Tokyo: Universal Academic.

Blair DG, Taniwaki M, Zhao CN, *et al.* (1997) Progress in the development of technology for advanced laser interferometric gravitational wave detectors. In: Tsubono K, Fujimoto MK and Kurodo K (eds) *Gravitational Wave Detection*, pp. 75–94. Tokyo: Universal Academic.

Brillet A (1997) Virgo status report. In: Tsubono K, Fujimoto MK and Kurodo K (eds) *Gravitational Wave Detection*, pp. 163–173. Tokyo: Universal Academic.

Danzmann K (1997) LISA–A gravitational wave observatory in heliocentric orbit. In: Tsubono K, Fujimoto MK and Kurodo K (eds) *Gravitational Wave Detection*, pp. 111–116. Tokyo: Universal Academic.

Drever RWP, Ford GM, Hough J, *et al.* (1980) A gravity-wave detector using optical cavity sensing. In: Schmutzer E (ed.) *Proceedings of the Ninth International Conference on General Relativity and Gravitation*, pp. 265–267. Berlin: Springer.

Forward RL (1978) Wideband laser-interferometer gravitational-radiation experiment. *Physical Review* D17: 379.

Hough J and the GEO 600 Team (1997) GEO 600 Current status and some aspects of the design. In: Tsubono K, Fujimoto MK and Kurodo K (eds) *Gravitational Wave Detection*, pp. 175–182. Tokyo: Universal Academic.

Robertson NA (2000) Laser interferometric gravitational wave detectors. *Classical Quantum Gravity* 17: R19–R40.

Saulson PR (1994) Fundamentals of Interferometric Gravitational Wave Detectors. Singapore: World Scientific.

Taylor JH and Weisberg JM (1989) Further experimental tests of relativistic gravity using the binary pulsor PSR 1913 + 16. *Astrophysical Journal* 345: 434–450.

Thorne KS (1987) Gravitational radiation. In: Hawking S and Isreal W (eds) *300 Years of Gravitation*, pp. 330–458. Cambridge, UK: Cambridge University Press.

Tsubono K and the TAMA collaboration (1997) TAMA Project, Gravitational wave detection. In: Tsubono K, Fujimoto MK and Kurodo K (eds) *Gravitational Wave Detection*, pp. 183–191. Tokyo: Universal Academic.

Weiss R (1972) Electromagnetically coupled broadband gravitational antenna. *MIT Quarterly Progress Report* no. 105.

Phase-Measurement Interferometry

K Creath, Optineering and University of Arizona, Tucson, AZ, USA

J Schmit, Veeco Instruments, Inc., Tucson, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Phase-measurement interferometry is a way to measure information encoded in interference patterns generated by an interferometer. Fringe patterns (fringes) created by interfering beams of light are analyzed to extract quantitative information about an object or phenomenon. These fringes are localized somewhere in space and require a certain degree of spatial and temporal coherence of the source to be visible. Before these techniques were developed in the late 1970s and early 1980s, fringe analysis was done either by estimating fringe deviation and irregularity by eye or by manually digitizing the centers of interference fringes using a graphics tablet. Digital cameras and desktop computers have made it possible to easily obtain quantitative information from fringe patterns. The techniques described in this article are independent of the type of interferometer used.

Interferometric techniques using fringe analysis can measure features as small as a micron wide or as large as a few meters. Measurable height ranges vary from a few nanometers up to 10s of mm, depending upon the interferometric technique employed. Measurement repeatability is very consistent, and typically repeatability of 1/100 rms of a fringe is easily obtainable while 1/1000 rms is possible. Actual measurement precision depends on what is being measured and how the technique is implemented, while accuracy depends upon comparison to a sanctified standard.

There are many different types of applications for fringe analysis. For example, optical surface quality can be determined using a Twyman–Green or Fizeau interferometer. In addition, wavefront quality measurements of a source or an optical system can be made in transmission, and the index of refraction and homogeneity of optical materials can be mapped out. Many nonoptical surfaces can also be measured. Typically, surface topography information at some specific spatial frequency scale is extracted. These measurements are limited by the resolution of the optical system and the field of view of the imaging system. Lateral and vertical dimensions can also be measured. Applications of nonoptical surfaces include disk and wafer flatness, roughness measurement, distance and range sensing. Phase measurement techniques used in holographic interferometry, TV holography, speckle interferometry, moiré, grating interferometry, and fringe projection are used for nondestructive testing to measure surface structure as well as displacements due to stress and vibration. This article outlines the basics of phase measurement interferometry (PMI) techniques as well as the types of algorithms used. The bibliography lists a number of references for further reading.

Background

Basic Parts of a Phase-Measuring Interferometer

A phase-measuring interferometer consists of a light source, an illumination system (providing uniform illumination across the test surface), a beamsplitter (usually a cube or pellicle so both beams have the same optical path), a reference surface (needs to be good because these techniques measure the difference between the reference and test surface), a sample

fixture, an imaging system (images a plane in space where the surface is located onto the camera), a camera (usually a monochrome CCD), an image digitizer/frame grabber, a computer system, software (to control the measurement process and calculate the surface map), and often a spatial or temporal phase shifter to generate multiple interferograms.

Steps of the Measurement Process

To generate a phase map, a sample is placed on a sample fixture and aligned, illumination levels are adjusted, the sample image is focused onto the camera, the fringes are adjusted for maximum contrast, the phase shifter or fringe spacing is adjusted or calibrated as necessary, a number of images is obtained and stored with the appropriate phase differences, the optical path difference (OPD) is calculated as the modulo 2π phase and then is unwrapped at each pixel to determine the phase map.

To make consistent measurements, some interferometers need to be on vibration isolation systems and away from heavy airflows or possible acoustical coupling. It helps to cover any air paths that are longer than a few millimeters. Consideration needs to be made for consistency of temperature and humidity.

The human operating the interferometer is also a factor in the measurement. Does this person always follow the same procedure? Are the measurements sampled consistently? Are the test surfaces clean? Is the sample aligned the same and correctly? Many different factors can affect a measurement. To obtain repeatable measurements it is important to have a consistent procedure and regularly verify measurement consistency.

Common Interferometer Types

One of the most common interferometers used in optical testing is the Twyman–Green interferometer (Figure 1). Typically, a computer controls a mirror pushed by a piezo-electric transducer (PZT). The test surface is imaged onto the camera and the computer has a frame grabber that takes frames of fringe data.

The most common commercially available interferometer for optical testing is the Fizeau interferometer (Figure 2). This versatile instrument is very insensitive to vibrations due to the large common path for both interfering wavefronts. The reference surface (facing the test object) is moved by a PZT to provide the phase shift.

Interference microscopes (*see* **Microscopy:** Interference Microscopy. **Interferometry:** White Light Interferometry) are used for looking at surface roughness and small structures (see Figure 3). These instruments can employ Michelson, Mirau, and Linnik interference objectives with a laser or white light source, or Fizeau-type interference objectives that typically use a laser source because of the unequal paths in the arms of the interferometer. The phase shift is accomplished by moving the sample, the reference surface, or parts of the objective relative to the sample. Figure 3 shows a schematic of a Mirau-type interferometric microscope for phase measurement.

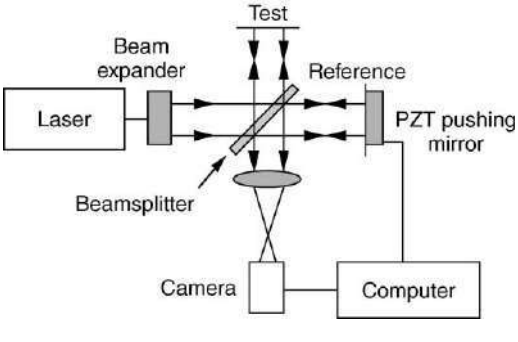

Figure 1 Twyman–Green interferometer.

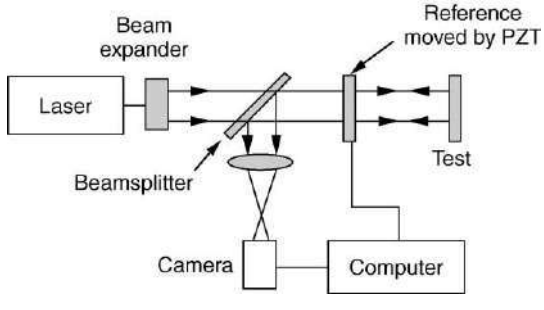

Figure 2 Fizeau interferometer.

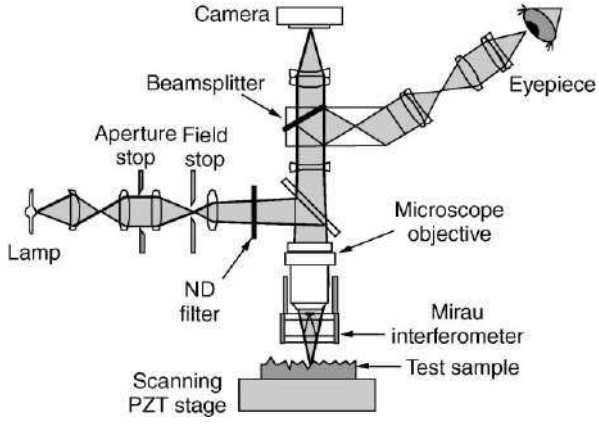

Figure 3 Mirau interference microscope.

Determination of Phase

The Interference Equation

Interference fringes from a coherent source (e.g., a laser) are theoretically sinusoidal (Figure 4a), while fringes from an incoherent source (e.g., white light) are localized in a wavepacket at a point in space (Figure 4b) where the optical paths of the arms of the interferometer are equal and marked on Figure 4 as scanner positions $z = 0$ (*see* **Interferometry:** White Light Interferometry). For generality, this analysis considers the determination of phase within the wave packet for fringes localized in space.

Interference fringes at any point in the wavepacket can be written in the following form:

$$I(x, y) = I_0\{1 + \gamma(x, y, z)\cos[\phi(x, y)]\} \qquad [1]$$

where I_0 is the dc irradiance, γ is the fringe visibility (or contrast), $2I_0\gamma$ is the modulation (irradiance amplitude or the ac part of the signal), and ϕ is the phase of the wavefront as shown in Figure 5. For simplicity, this drawing assumes that the interference fringe amplitude is constant.

Fringe visibility can be determined by calculating

$$\gamma = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \qquad [2]$$

where $I_{\max}$ is the maximum value of the irradiance for all phase values and $I_{\min}$ is the minimum value. The fringe visibility has a real value between 0 and 1 and varies with the position along the wavepacket. The fringe visibility, as defined here, is the real part of the complex degree of coherence.

Types of Phase Measurement Techniques

Phase can be determined from either a number of interference fringe patterns or from a single interferogram with appropriate fringe spacing. Temporal techniques require an applied phase shift between the test and reference beams as a function of time while multiple frames of interference fringe data are obtained. Spatial techniques can obtain data from a single interferogram that requires a carrier pattern of almost straight fringes to either compare phases of adjacent pixels or to separate orders while performing operations in the Fourier domain. Spatial techniques may also simultaneously record multiple interferograms with appropriate relative phase shift differences separated spatially in space. Multiple frame techniques require more data than single frame techniques. Temporal techniques require that the interference fringes be stable over the time period it takes to acquire the number of images. While single-frame spatial techniques require less data and can be done with a single image, they generally have a reduced amount of resolution and less precision than temporal techniques.

There are literally hundreds of algorithms and techniques for extracting phase data from interference fringe data. The references listed in the Further Reading offer more details of these techniques.

Temporal Phase Measurement

Temporal techniques use data taken as the relative phase between the test and reference beams is modulated (shifted). The phase (or OPD) is calculated at each measured point in the interferogram. As the phase shifter is moved, the phase at a single point in the interferogram changes. The effect looks like the fringes are moving across the interferogram, and because of these techniques are sometimes called fringe scanning or fringe shifting techniques. However, the fringes are not really moving; rather the irradiance at a single detector point is changing (hopefully sinusoidally) in time (see Figure 6). A 180° or π phase causes a bright fringe to become a dark fringe.

Phase Modulation Techniques

There are many ways to introduce a phase modulation (or shift). These include moving a mirror or the

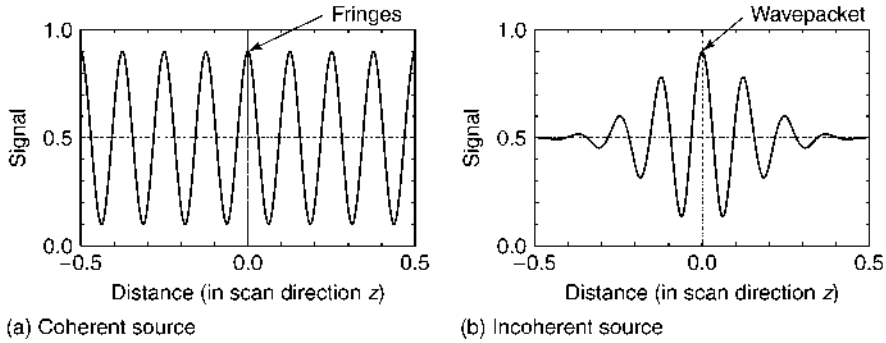

Figure 4 Interference fringes for coherent and incoherent sources as observed at point $x, y, z = 0$ corresponds to equal optical path lengths of the reference and object beams.

sample, tilting a glass plate, moving a diffraction grating, rotating a polarizer, analyzer, or half-wave plate, using a two-frequency (Zeeman) laser source, modulating the source wavelength, and switching an acousto-optic Bragg cell, or magneto-optic/electro-optic cell. While any of these techniques can be used for coherent sources, special considerations need to be made for temporally incoherent sources (*see* **Interferometry:** White Light Interferometry). Figure 7 shows how moving a mirror introduces a relative phase shift between object and reference beams in a Twyman–Green interferometer.

Extracting Phase Information

Including the phase shift, the interference equation is written as

$$I(x, y) = I_0(x, y)\{1 + \gamma_0(x, y)\cos[\phi(x, y) + \alpha(t)]\} \quad [3]$$

where $I(x, y)$ is the irradiance at a single detector point, $I_0(x, y)$ is the average (dc) irradiance, $\gamma_0(x, y)$ is the fringe visibility before detection, $\phi(x, y)$ is the phase of the wavefront being measured, and $\alpha(t)$ is the phase shift as a function of time.

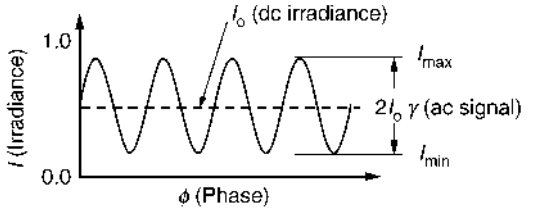

Figure 5 Variable definitions for interference fringes.

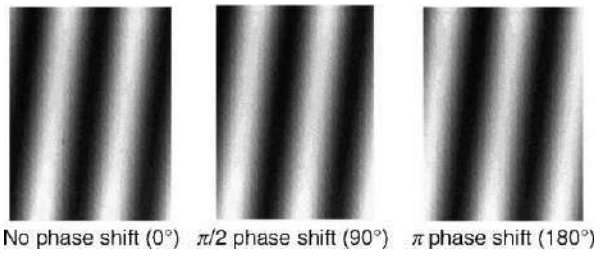

Figure 6 Interference fringe patterns corresponding to different relative phase shifts between test and reference beams.

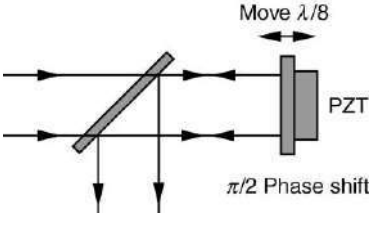

Figure 7 Moving a mirror to shift relative phase by $\lambda/4$.

Since the detector has to integrate for some finite time, the detected irradiance at a single point becomes an integral over the integration time Δ (Figure 8) where the average phase shift for the ith frame of data is α_i:

$$I_i(x, y) = \frac{1}{\Delta}\int_{\alpha_i-(\Delta/2)}^{\alpha_i+(\Delta/2)} I_0(x, y)\{1 + \gamma_0(x, y) \times \cos[\phi(x, y) + \alpha(t)]\}\mathrm{d}\alpha(t) \quad [4]$$

After integrating over $\alpha(t)$, the irradiance of the detected signal becomes

$$I_i(x, y) = I_0(x, y)\left\{1 + \gamma_0(x, y)\mathrm{sinc}\left(\frac{\Delta}{2}\right) \times \cos[\phi(x, y) + \alpha_i]\right\} \quad [5]$$

where $\mathrm{sinc}(\Delta/2) = \sin(\Delta/2)/(\Delta/2)$, which reduces the detected visibility.

Ramping Versus Stepping

There are two different ways of shifting the phase; either the phase shift can be changed in a constant and linear fashion (ramping) (see Figure 8) or it can be stepped in increments. Ramping provides a continuous smooth motion without any jerking motion. This option may be preferred if the motion does not wash out the interference fringes. However, ramping requires good synchronization of the camera/digitizer and the modulator to get the correct OPD (or phase) changes between data frames. Ramping allows faster data taking but requires electronics that are more sophisticated. In addition, it takes a finite time for a mass to move linearly. When ramping, the first frame or two of data usually needs to be discarded because the shift is not correct until the movement is linear.

The major difference between ramping and stepping the phase shift is a reduction in the modulation of the interference fringes after detection (the $\mathrm{sinc}(\Delta/2)$ term in eqn [5]). When the phase shift is stepped ($\Delta = 0$), the sinc term has a value of one. When the phase shift is ramped ($\Delta = \alpha$) for a phase

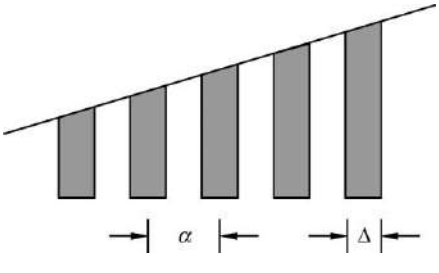

Figure 8 Change in OPD integrating over a Δ phase shift with sample separation α.

shift of $\alpha = 90° = \pi/2$, this term has a value of 0.9. Therefore, ramping slightly reduces the detected fringe visibility.

Signal Modulation

Phase measurement techniques assume that the irradiance of the interference fringes at the camera covers as much of the detector's dynamic range as possible and that the phase of the interference fringe pattern is modulating at each individual pixel as the phase between beams is modulated. If the irradiance at a single detector point does not modulate as the relative phase between beams is shifted, the height of the surface cannot be calculated. Besides the obvious reductions in irradiance modulation due to the detector sampling and pixel size or a bad detector element, scattered light within the interferometer, and defects or dirt on the test object, can also reduce signal modulation. Phase measurement techniques are designed to take into account the modulation of the signal at each pixel. If the signal does not modulate enough at a given pixel, then the data at that pixel are considered unusable, flagged as 'bad' and are often left blank. Phase values for these points may be interpolated from surrounding pixels if there are sufficient data.

Three-Frame Technique

The simplest method to determine phase uses three frames of data. With three unknowns, three sets of recorded fringe data are needed to reconstruct a wavefront providing a phase map. Using phase shifts of $\alpha_i = \pi/4$, $3\pi/4$, and $5\pi/4$, the three fringe measurements at a single point in the interferogram may be expressed as

$$I_1 = I_0\left[1 + \gamma\cos\left(\phi + \frac{\pi}{2}\right)\right] = I_0\left[1 + \frac{\sqrt{2}}{2}\gamma(\cos\phi - \sin\phi)\right] \quad [6]$$

$$I_2 = I_0\left[1 + \gamma\cos\left(\phi + \frac{3\pi}{2}\right)\right] = I_0\left[1 + \frac{\sqrt{2}}{2}\gamma(-\cos\phi - \sin\phi)\right] \quad [7]$$

$$I_3 = I_0\left[1 + \gamma\cos\left(\phi + \frac{5\pi}{2}\right)\right] = I_0\left[1 + \frac{\sqrt{2}}{2}\gamma(-\cos\phi + \sin\phi)\right] \quad [8]$$

Note that (x, y) dependencies are still implied. The choice of the specific phase shift values is to make the math simpler. The phase at each detector point is

$$\phi = \tan^{-1}\left(\frac{I_3 - I_2}{I_1 - I_2}\right) \quad [9]$$

In most fringe analysis techniques, we are basically trying to solve for terms such that we end up with the tangent function of the phase. The numerator and denominator shown above are respectively proportional to the sine and cosine of the phase. Note that the dc irradiance and fringe visibility appear in both the numerator and denominator. This means that variations in fringe visibility and average irradiance from pixel to pixel do not affect the results. As long as the fringe visibility and average irradiance at a single pixel is constant from frame to frame, the results will be good. If the different phase shifts are multiplexed onto multiple cameras, the results will be dependent upon the gain of corresponding pixels.

Bad data points with low signal modulation are determined by calculating the fringe visibility at each data point using:

$$\gamma = \frac{\sqrt{(I_3 - I_2)^2 + (I_1 - I_2)^2}}{\sqrt{2}I_0} \quad [10]$$

It is simpler to calculate the ac signal modulation $(2I_0\gamma)$ and then set the threshold on the modulation to a typical value of about 5–10% of the dynamic range. If the modulation is less than this value at any given data point, the data point is flagged as bad. Bad points are usually caused by noisy pixels and can be due to scratches, pits, dust and scattered light.

Although three frames of data are enough to determine the phase, this and other 3-frame algorithms are very sensitive to systematic errors due to nonsinusoidal fringes or nonlinear detection, phase shifter miscalibration, vibrations and noise. In general, the larger the number of frames of data used to determine the phase, the smaller the systematic errors.

Phase Unwrapping

The removal of phase ambiguities is generally called phase unwrapping, and is sometimes known as integrating the phase. The phase ambiguities owing to the modulo 2π arctangent calculation can simply be removed by comparing the phase difference between adjacent pixels. When the phase difference between adjacent pixels is greater than π, a multiple of 2π is added or subtracted to make the difference less than π. For the reliable removal of

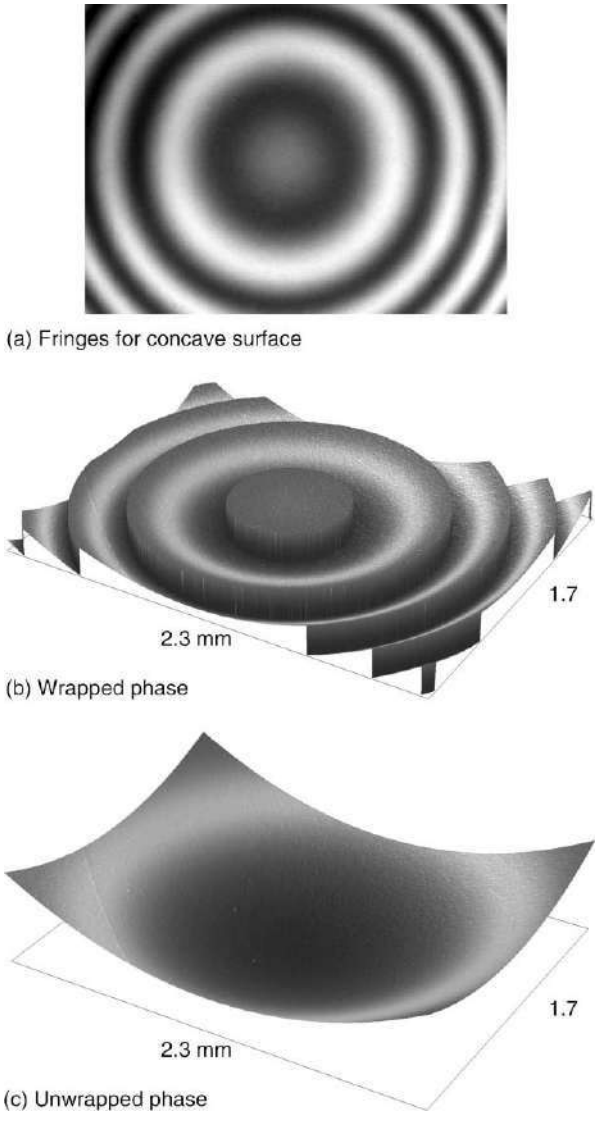

Figure 9 Fringes, wrapped, and unwrapped phase maps.

discontinuities, the phase must not change by more than π ($\lambda/2$ in optical path (OPD)) between adjacent pixels. Figure 9 shows an example of wrapped and unwrapped phase values.

Given a trouble-free wrapped phase map, it is enough to search row by row (or column by column) for phase differences of more than π between neighboring pixels. However, fringe patterns usually are not perfect and are affected by systematic errors. Some of the most frequently occurring error sources are noise, discontinuities in the phase map, violation of the sampling theorem and invalid data points (e.g., due to holes in the object or low modulation regions). Some phase maps are not easily unwrapped. In these cases different techniques like wave packet peak sensing in white light interferometry are used.

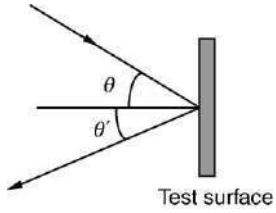

Figure 10 Angles of illumination and viewing.

From Wavefront to Surface

Once the phase of the wavefront is known, surface shape can be determined from the phase map. Surface height H of the test surface relative to the reference surface at a location (x, y) is given by

$$H(x, y) = \frac{\phi(x, y)\lambda}{2\pi(\cos\theta + \cos\theta')} \qquad [11]$$

where λ is the wavelength of illumination, and θ and θ' – the angles of illumination and viewing with respect to the surface normal – are shown in Figure 10. For interferometers (e.g., Twyman–Green or Fizeau) where the illumination and viewing angles are normal to the surface ($\theta = \theta' = 0$), the surface height is simply:

$$H(x, y) = \frac{\lambda}{4\pi}\phi(x, y) \qquad [12]$$

Since the wavefront measured represents the relative difference between the interfering reference and test wavefronts, this phase map only directly corresponds to the surface under test when the reference wavefront is perfectly flat. In practice, the shape of the reference surface needs to be accounted for by measuring it using a known test surface and subtracting this reference measurement from subsequent measurements of the test surface.

Phase Change on Reflection

Phase shifting interferometry measures the phase of reflected light to determine the shape of objects. The reflected wavefront will represent the object surface (within a scaling factor) if the object is made of a single material. If the object is comprised of multiple materials that exhibit different phase changes on reflection, the measured wavefront needs to be corrected for these phase differences (*see* **Interferometry:** White Light Interferometry).

Overview of Phase Measurement Algorithms and Techniques

There are literally hundreds of published algorithms and techniques. The optimal algorithm depends on

the application. Most users prefer fast algorithms using a minimal amount of data that are as accurate and repeatable as possible, immune to noise, adaptable, and easy to implement. In practice, there are trade-offs that must be considered when choosing a specific algorithm or technique. This section provides an overview of the types of algorithms to aid the reader in sifting through published algorithms.

Synchronous Detection

One of the first techniques for temporal phase measurement utilized methods of communication theory to perform synchronous detection. To synchronously detect the phase of a noisy sinusoidal signal, the signal is first correlated (or multiplied) with sinusoidal and cosinusoidal reference signals (signals in quadrature) of the same frequency and then averaged over many periods of oscillation. This method of synchronous detection as applied to phase measurement can be extracted from the least squares estimation result when the phase shifts are chosen such that N measurements are equally spaced over one modulation period. With phase shifts α_i such that:

$$\alpha_i = \frac{2\pi}{N} i, \qquad \text{with } i = 1, \ldots, N \qquad [13]$$

the phase can be calculated from

$$\phi(x, y) = \tan^{-1}\left[\frac{\sum I_i(x, y)\sin\alpha_i}{\sum I_i(x, y)\cos\alpha_i}\right] \qquad [14]$$

Note that N can be any number of frames (or samples). The more frames of data, the smaller the systematic errors. This technique does not take large amounts of memory for a large number of frames, because only the running sums of the fringes multiplied by the sine and cosine of the phase shift need to be remembered. The 4-frame algorithm from Table 1 is an example of a direct adaptation of synchronous detection where simple values of 1, −1 or 0 for every $\pi/2$ phase shift can be assigned to the sine and cosine functions.

Algorithm Design

In the last ten years, a lot of work has been done to generalize the derivation of fringe analysis algorithms. This work has enabled the design of algorithms for specific applications, which are insensitive to specific systematic error sources.

Most algorithms use polynomials for the numerator and denominator. Given fringe data:

$$I_i = I_0[1 + \gamma\cos(\phi + \alpha_i)] \qquad [15]$$

Table 1 Sampling function weights for a few selected algorithms

N frames	*Phase shift*	*Coefficients*
3	$\frac{\pi}{2}$	$\frac{1, -1, 0}{0, 1, -1}$
4	$\frac{\pi}{2}$	$\frac{0, -1, 0, 1}{1, 0, -1, 0}$
5	$\frac{\pi}{2}$	$\frac{0, -2, 0, 2, 0}{1, 0, -2, 0, 1}$
7	$\frac{\pi}{3}$	$\frac{\sqrt{3}[0, 1, 1, 0, -1, -1, 0]}{-1, -1, 1, 2, 1, -1, -1}$
8	$\frac{\pi}{2}$	$\frac{1, 5, -11, -15, 15, 11, -5, -1}{1, -5, -11, 15, 15, -11, -5, 1}$
12	$\frac{\pi}{3}$	$\frac{\sqrt{3}[0, -3, -3, 3, 9, 6, -6, -9, -3, 3, 3, 0]}{2, 1, -7, -11, -1, 16, 16, -1, -11, -7, 1, 2}$

the phase is calculated using

$$\phi = \tan^{-1}\left[\frac{\sum_i n_i I_i}{\sum_i d_i I_i}\right] \qquad [16]$$

The numerator and denominator of the arctangent argument are both polynomials. The numerator is a sum proportional to the sine (imaginary part) and the denominator is a sum proportional to the cosine (real part),

$$\text{num} = 2kI_0\gamma\sin\alpha_i = \sum_i n_i I_i \qquad [17]$$

$$\text{den} = 2kI_0\gamma\cos\alpha_i = \sum_i d_i I_i \qquad [18]$$

where the constant k depends on the values of coefficients. From this the fringe visibility is given by

$$\gamma = \frac{\sqrt{(\text{num})^2 + (\text{den})^2}}{2kI_0} \qquad [19]$$

The coefficient vectors for the numerator and denominator are window functions. For an algorithm such as the 4-frame technique, the weights of all samples are equal [1, 1, 1, 1]. This makes the coefficients all equal. For other algorithms, such as the 5-frame technique, the weights are larger on the middle frames than on the outer frames. The weights for the 5-frame technique are [1, 2, 2, 2, 1]. A property of the coefficient vectors is that the sum of the coefficients for each, the numerator and denominator, should be zero. Examples of coefficient vectors for a few selected algorithms are given in Table 1.

Heterodyne Interferometry

Historically, heterodyne techniques were developed and used before temporal phase-shifting techniques. These techniques generally determine the phase electronically at a single point by counting fringes and fractions of fringes. Areas are analyzed by scanning a detector.

Phase shifts are usually obtained using two slightly different frequencies in the reference and test beams. The beat frequency produced by the interference between the reference and test beams is compared to a reference sinusoidal signal, which may be produced either optically or electronically. The time delay (or distance traveled) between the crossing of the zero phase points of the test and reference sinusoidal signals is a measure of the phase. Every time the test signal passes through another zero in the same direction as the test surface is moved, another fringe can be counted. This is how fringe orders are counted. If the beam is interrupted as the detector is scanned across the interferogram, the fringe count is corrupted and the measurement needs to be started again. Frequency multiplication (harmonics) can also be used to determine fractions of fringes. Today, heterodyne techniques are used mainly in distance measuring interferometers. The precision and accuracy of distance measuring interferometers is at least on the order of 1 part in 10^6.

Fourier-Transform Technique

The Fourier-transform technique is a way to extract phase from a single interferogram. It is used a lot in nondestructive testing and stellar interferometry where it is difficult to get more than a single interferogram. The basic technique is shown schematically in Figure 11. The recorded interferogram distribution is Fourier transformed, and one order (usually the + or − first order) is either isolated and shifted to zero frequency or filtered out using a rectangular window. After an inverse Fourier transform, the result is the phase.

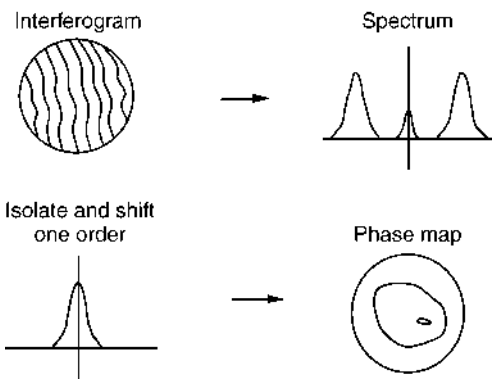

Figure 11 Fourier transform technique.

To illustrate this technique mathematically, the interference equation is rewritten as

$$I(x, y) = I_0(x, y) + c(x, y)\exp(\mathrm{i}2\pi f_0 x) + c^*(x, y)\exp(-\mathrm{i}2\pi f_0 x) \quad [20]$$

where $c(x, y) = I_0(x, y)\gamma(x, y)\exp[\mathrm{i}\phi(x, y)]$ and the $*$ indicates a complex conjugate. The term $c(x, y)$ contains the phase information we wish to extract. After performing a one-dimensional Fourier transform:

$$I(\xi, y) = I_0(\xi, y) + c(\xi - f_0, y) + c^*(\xi - f_0, y) \quad [21]$$

where ξ is the spatial frequency in the x direction, and italics indicate Fourier transforms. The next step is to filter out and isolate the second term, and then inverse Fourier transform to yield $c(x, y)$. The wavefront modulo 2π is then given by

$$\phi(x, y) = \tan^{-1}\left\{\frac{\mathrm{Im}[c(x, y)]}{\mathrm{Re}[c(x, y)]}\right\} \quad [22]$$

where Re and Im refer to the real and imaginary part of the function.

This technique has limitations. If the fringes are nonsinusoidal, there will not be a simple distribution in the frequency space; there will be many orders. Another problem is overlapping orders. There needs to be a carrier frequency present that ensures that the orders are separated in frequency space. This carrier frequency is produced by adding tilt fringes until the orders are separated. This means that the aberration (fringe deviation) has to be less than the fringe spacing. Another problem is aliasing. If the interferogram is not sampled sufficiently, there will be aliasing and it will not be possible to separate the orders in frequency space. Finally, large variations in average fringe irradiance and fringe visibility across the interferogram can also cause problems.

Spatial Carrier-Frequency Technique

This is essentially the equivalent of the Fourier transform technique but is performed in the spatial domain. It is also used when there is only one interferogram available and its major applications include nondestructive testing and measurement of large optics.

These techniques relate closely to the temporal phase-measurement methods; however, instead of using a number of interferograms they can obtain all the information from a single interferogram. As an example, let's assume that the fringes are vertical and parallel to the columns on the detector array.

The carrier frequency (i.e., the number of tilt fringes) is adjusted so that there is an α phase change from the center of one pixel to the next. As long as there is not much aberration (deviation), the phase change from pixel to pixel across the detector array will be approximately constant.

When the fringes are set up this way, the phase can be calculated using adjacent pixels (see Figure 12). If one fringe takes up 4 pixels, the phase shift α between pixels will be 90°. An algorithm such as the three-frame, four-frame, or five-algorithm can be used with adjacent pixels as the input. Therefore, 3, 4, or 5 pixels in a row will yield a single-phase point. Then, the analysis window is shifted sideways one pixel and phase is calculated at the next point. This technique assumes that the dc irradiance and fringe visibility do not change over the few pixels used to calculate each phase value.

Spatial Multichannel Phase-Shift Techniques

These techniques detect all phase maps simultaneously and multiplex the phase shift using static optical elements. This can be done by using either separate cameras as illustrated below or by using different detector areas to record each of the interferograms used to calculate the phase. The phase is usually calculated using the same techniques that are used for the temporal phase techniques.

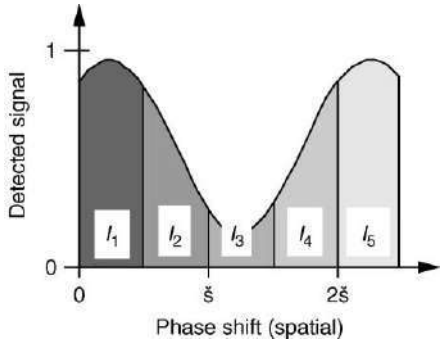

Figure 12 Spatial phase shifting. Different relative phase shifts are on adjacent detector elements.

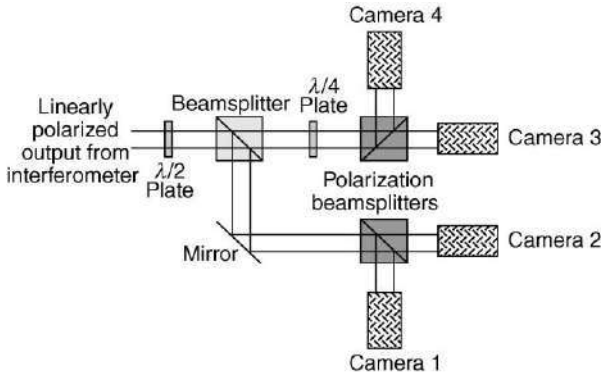

Figure 13 Spatial phase shifting using four cameras.

As an example, a four-channel interferometer can be made using the setup shown in Figure 13 to record four interferograms with 90° phase shifts between them. Camera 1 will yield fringes shifted 180° with respect to camera 2, and cameras 3 and 4 will have phase shifts of 90° and 270°. The optical system may also utilize a holographic optical element to split the beam to multiplex the four phase shifts on four quadrants of a single camera.

Signal Demodulation Techniques

The task of determining phase can be broadened by looking toward the field of signal processing. For communication via radio, radar, and optical fibers electrical engineers have developed a number of ways of compressing and encoding a signal as well as decompressing a signal and decoding it. An interference fringe pattern looks a lot like an am radio signal. Thus, it can be demodulated in similar ways. In recent years many new algorithms have been developed by drawing on techniques from communication theory and applying them to interferogram processing. Many of these use different types of transforms such as Hilbert transforms for straight fringes or circular transforms for closed fringes.

Extended Range Phase Measurement Techniques

A major limitation of phase measurement techniques is that they cannot determine surface discontinuities larger than $\lambda/4$ ($\lambda/2$ in optical path (OPD)) between adjacent pixels. One obvious solution is to use longer wavelength sources in the infrared where optically rough surfaces look smooth and their shape can be measured. An alternative is two-wavelength interferometry where two measurements at different wavelengths are taken and the measurable height limitation is now determined by the equivalent wavelength:

$$\lambda_{eq} = \frac{\lambda_1 \lambda_2}{|\lambda_1 - \lambda_2|} \qquad [23]$$

Another method that allows for measurement of smooth surfaces with large height discontinuities that is limited only by the working distance of the objective combines a white light and a phase measurement interferometric technique in one long scan. The position of the wave packet resolves 2π ambiguities that result from the arctangent function (see Phase unwrapping section above). The phase of the fringe close to the wave packet maximum is determined without ambiguity. Sometimes it is not the step height discontinuity that is a problem but rather the high slope of the measured surface. If we know that the surface is continuous, then the

unwrapping procedure can take advantage of this a priori information to look for a continuous.

Techniques for Deformation Measurement

Some of the techniques for deformation measurement have already been described earlier (see sections on Single frame techniques and Spatial multichannel phase-shift techniques). However, fringes in any interferometer can be analyzed not only in the x, y planes but also in the x, z, or y, z planes for which the carrier frequency of the fringes is introduced. Analysis of fringes in any plane has the same restrictions.

If the initial static object shape is measured using conventional phase measurement techniques, not only can the motion of each object point be determined but also the deformation of the whole object in time. If the motion of the object is periodic, then the object motion can be 'frozen' by using stroboscopic illumination of the same frequency as the object's motion. Once the motion is frozen, any temporal technique can be used to measure object shape at some phase of its motion. Changing the time offset between stroboscopic illumination and the periodic signal driving the motion, the object can be 'frozen' and thus measured at different phases of its periodic motion.

Systematic Errors

Noise Sensitivity

Measurement noise mostly arises from random fluctuations in the detector readout and electronics. This noise reduces precision and repeatability. Averaging multiple measurements can reduce effects of these random fluctuations.

Phase Shifter Errors

Phase shifter errors can be due to both miscalibration of the system and nonlinearities in the phase shift. It is possible to purchase very linear phase shifters. It is also possible to correct nonlinearities by determining the voltage signal making the phase shifter provide a linear phase shift.

Linear phase shifter errors (miscalibration) of the phase shift have an error signature that is at twice the frequency of the fringes. If there are two fringes across the field of view, the error signature will have four across the field of view. Figure 14 shows the difference between a calibrated and an uncalibrated phase shifter as well as the difference in error for two different phase measurement algorithms. Some algorithms are obviously more sensitive than others to this type of error.

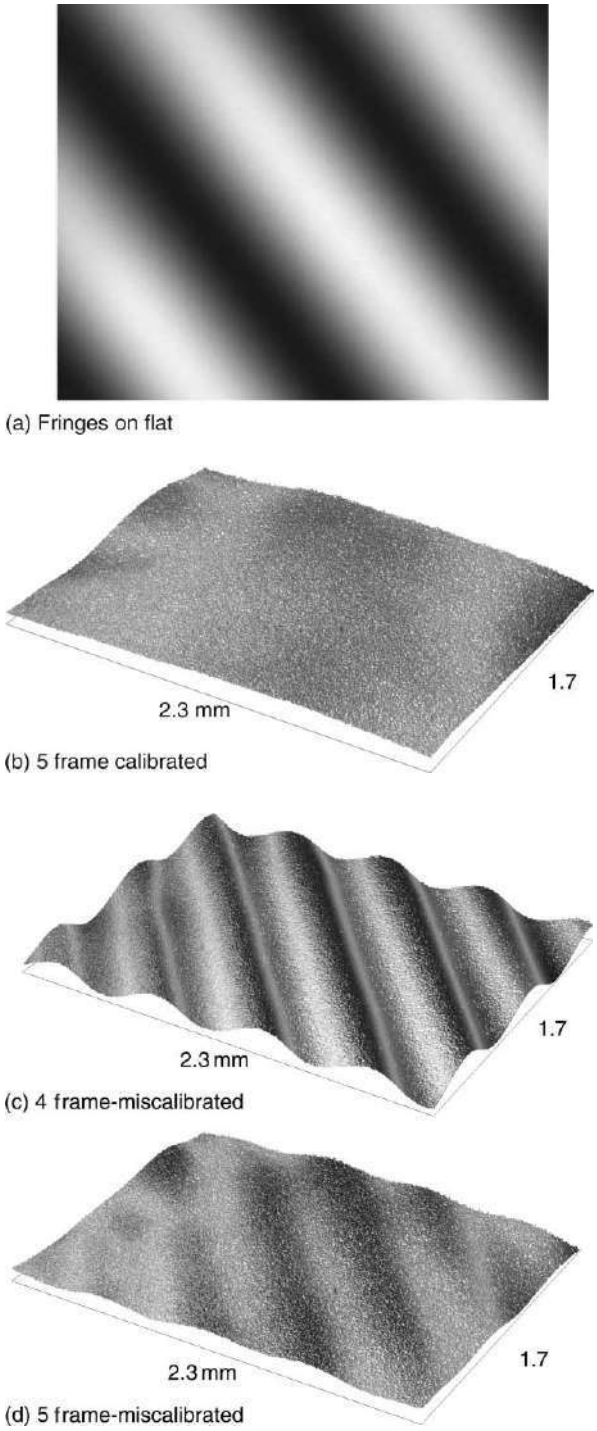

Figure 14 Comparison of phase maps for calibrated and uncalibrated phase shifts using two different phase measurement algorithms.

Other Types of Systematic Errors to Consider

Other types of errors to consider are detector nonlinearities, quantization errors due to analog-to-digital converters, and dissimilar materials (see Phase change upon reflection). For temporal phase measurement techniques errors due to vibration and

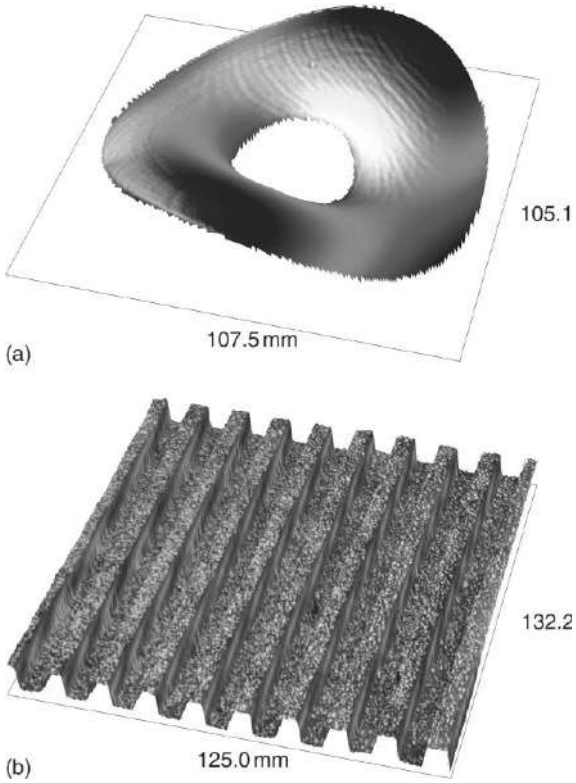

Figure 15 Measurements of (a) hard disc substrate with a vertical range is 2.2 micrometers and (b) binary grating roughness standard.

air turbulence need to be considered as well. Spatial phase measurement techniques are sensitive to miscalibrated tilt (wrong carrier frequency), unequally spaced fringes, and sampling and windowing in Fourier transform techniques.

Choosing an Algorithm or Technique

Each algorithm and type of measurement technique is sensitive to different types of systematic errors. Choosing a proper algorithm or technique for a particular type of measurement depends on the specific conditions of the test itself and reducing the systematic errors for a particular type of measurement. This is the reason that so many algorithms exist. The references in the bibliography will help the reader determine what type of algorithm will work best for a specific application.

Examples of Applications

Phase shifting interferometry can be used for measurements such as hard disk flatness, quality of optical elements, lens curvature, dimensions and quality of air-bearing surfaces of magnetic read/write heads, cantilevers, and semiconductor elements. Figure 15 shows results for measurements of a hard disk substrate and a roughness grating.

Conclusions

Phase measurement interferometry techniques have increased measurement range and precision enabling the production of more complex and more precise components. As work continues on development of interferometric techniques, phase measurement techniques will continue to become more robust and less sensitive to systematic errors. Anticipated advances will enable measurements of objects that were unimaginable 20 or 30 years ago.

See also

Interferometry: White Light Interferometry. **Microscopy:** Interference Microscopy.

Further Reading

Bruning JH (1978) Fringe scanning interferometers. In: Malacara D (ed.) *Optical Shop Testing*, pp. 409–438. New York: Wiley.

Creath K (1988) Phase-measurement interferometry techniques. In: Wolf E (ed.) *Progress in Optics*, vol. 26, pp. 349–393. Amsterdam: Elsevier.

Creath K (1993) Temporal phase measurement methods. In: Robinson DW and Reid GT (eds) *Interferogram Analysis*, pp. 94–140. Bristol: IOP Publishing.

Creath K and Schmit J (1996) N-point spatial phase-measurement techniques for non-destructive testing. *Optics and Lasers in Engineering* 24: 365–379.

Ghiglia DC and Pritt MD (1998) *Two-Dimensional Phase Unwrapping*. New York: Wiley.

Greivenkamp JE and Bruning JH (1992) Phase shifting interferometry. In: Malacara D (ed.) *Optical Shop Testing*, 2nd edn, pp. 501–598. New York: Wiley.

Hariharan P (1992) *Basics of Interferometry*. New York: Academic Press.

Kerr D (ed.) (2004) FASIG special issue on fringe analysis. *Optics and Lasers in Engineering* 41: 597–686.

Kujawinska M (1993) Spatial phase measurement methods. In: Robinson DW and Reid GT (eds). *Interferogram Analysis*, pp. 141–193. Bristol, UK: IOP Publishing.

Larkin KG (1996) Neat nonlinear algorithm for envelope detection in white light interferometry. *Journal of Optical Society of America A* 13: 832–842.

Malacara D (1992) *Optical Shop Testing*, 2nd edn. New York: Wiley.

Malacara D, Servin M and Malacara Z (1998) *Interferogram Analysis for Optical Testing*. New York: Marcel Dekker.

Robinson DW and Reid GT (1993) *Interferogram Analysis: Digital Processing Techniques for Fringe Pattern Measurement*. Bristol, UK: IOP Publishing.

Schmit J and Creath K (1996) Window function influence on phase error in phase-shifting algorithms. *Applied Optics* 35: 5642–5649.

Surrel Y (2000) Fringe analysis. *Photo-Mechanics: Topics in Applied Physics* 77: 55–102.

White Light Interferometry

J Schmit, Veeco Instruments, Inc., Tucson, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In 1960 the development of the laser injected new possibility into an old discipline, interferometry (*see* **Interferometry**: Overview). The excellent spatial and temporal coherence (ideally a single point and a single wavelength source) of the laser allowed for the formation of nonlocalized interference fringes. With lasers, researchers could easily generate good contrast fringes practically anywhere where the beams overlapped. These excellent-quality fringes from laser interferometers enabled high-resolution distance and displacement measurements on the order of meters and made possible noncontact surface probing with nanometer and subnanometer resolution. This kind of precision, which was supported by the advent of computers and the advancement of detectors, was previously unimaginable.

However, interference phenomena were used in metrology long before lasers and can be observed without any complicated device or source. On a rainy day we can observe the colorful interference patterns that have formed on a thin layer of gasoline in a puddle. The relationship between colors and the small distance between reflecting surfaces, observed as far back as 1665 by Hook, is still used to approximate layer thicknesses. The colors in interference patterns have been used to determine the spectral components of beams. By the end of the nineteenth century, Michelson and Benoit were using interference phenomena to determine distances, long before the invention of the computer and the laser.

While the application of lasers in interferometry has advanced the science, white light interferometry (WLI), which uses a spatially and temporally incoherent source and creates fringes localized over a few microns in space, has also benefited from technological advancements. The development of solid state detectors, fast computers, electronic signal processing, and precise scanning stages has allowed for incredibly fast analysis of white light interferograms. In fact, WLI is used in many applications from film thickness and surface measurement, through spectroscopy to astronomy.

This chapter focuses on the formation of white light fringes. We first examine the existence of white light fringes in the everyday world. A description of different interferometric setups and their uses follows. The bulk of this article details the formation of white light fringes and examines their different applications with an emphasis on the analysis of object structure. Suggestions for further reading, both within this encyclopedia and from outside sources, can be found at the end of this article. A basic theory of interferometry and coherence is to be found in corresponding articles of this encyclopedia.

White Light Interference in Everyday Life

The effects of interference display spectacular colors when observed in white light, such as the sun. We often see these effects in oil or gasoline spills, in soap bubbles, or between two smooth pieces of glass in contact (separated by a very thin layer of air film) – which are the simplest everyday white light interferometers. The colors we see in these displays are interference colors, and their origin is interference rather than dispersion as in the colors of a rainbow or light reflected from a CD. The picture in Figure 1 is of a gasoline spill under cloudy sky illumination; we see beautiful interference colors that are hard to find in a rainbow: iron-gray, magenta, grayish blue, whitish green, and brownish yellow. Interference colors will differ not only with the thickness of the layer but also with the layer's absorption and dispersion, the relative indices of refraction of the film, the surrounding media, and the illumination. Interference colors can be observed for layers from a fraction of a micron to a few microns thick.

Interferometers for White Light Observation

In a white light interferometer, either the colors or the intensity distribution of the fringes is typically analyzed to retrieve the necessary encoded information, such as the film thickness, birefringence, index of refraction, dispersion, spectral properties, and surface structure. White light interference can be observed only in interferometric designs where the optical paths in both arms of the interferometer are (nearly) equal and the system is (nearly) compensated for dispersion.

Interference in Thin Film

Beams of light of different wavelengths incident on a transparent thin film (such as the ones in the puddle) are partially reflected from the top air/film interface

Figure 1 Interference colors in a gasoline spill in a puddle. This picture would closely resemble a picture from an interference microscope, if not for the leaf in the top left corner.

and partially from the bottom film/substrate interface and then beams of the same wavelength interfere with each other. The optical path difference (OPD) traveled by the interfering beams is related to the film thickness, the index of refraction for the given wavelength and the angle of incidence.

Fringes of equal thickness will be observed if the film thickness (distance between interfering wavefronts) varies. These fringes represent points of equal distance between the two surfaces of the film and are formed close to the film (see Figure 2). This situation is typical in any interferometer since rarely are wavefronts perfectly plane and parallel.

For any individual wavelength, λ_i, where the OPD is equal to $m\lambda_i$, and where m is an integer, a bright fringe of the color of the wavelength, λ_i, will be observed due to constructive interference. For white light illumination, the color at a given point will be dominated by the color of the wavelength for which the interference will be constructive. This color can be used to estimate the optical thickness of the film.

An additional factor needs to be taken into account when color is used for estimating optical thickness, namely the phase change on reflection. Different colors will be observed if the ratio of indices of refraction for both film interfaces is <1 (or >1) as opposed to when the ratio of indices is <1 for one interface and >1 for the other.

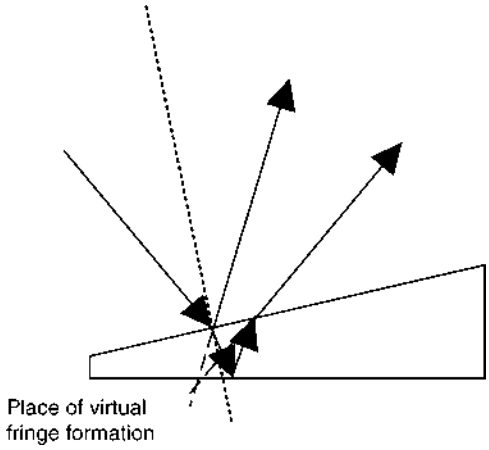

Figure 2 Formation of fringes of equal thickness for film with wedge.

Polarization Microscope

Interference colors that correspond to equal thickness fringes can represent the birefringence of an object; these colors are often observed using a polarization microscope (*see* **Microscropy:** Interference Microscopy). The polarization microscope is a conventional microscope with a set of polarizers with crossed polarization axes, one placed above and the other placed below the tested birefringent object.

Interference colors have been tabulated for many years and for many different purposes. Newton devised his color scale to describe interference fringes for two pieces of glass in contact with each other. Michel-Lévy and Lacroix in 1889, created a color scale to help recognize different rock forming minerals. For more information about colors in white light interference see sections on *White Light Interference* and *Spectral Interference* below.

Michelson Interferometer

A Michelson interferometric setup, shown in Figure 3, is often used to analyze white light interference, but the intensity of the fringes is observed as one mirror is scanned rather than the color of fringes that is usually observed.

Two plane mirrors (equivalent to the top and bottom surfaces of a thin film) return the beams to the beamsplitter, which recombines parts of the returned beams and directs them towards the detector where the interference is observed. Beam S1 travels through the parallel plate three times while beam S2 passes through the plate only once causing the system to be not well compensated (balanced) for dispersion. Thus, for observation of white light interference, an identical plate is placed in the path of beam S2. As one of the plane mirrors moves along the optical axis to change the OPD, the detector collects the irradiance, the intensity. Fringes will be observed if the system is well-compensated for dispersion and for optical path lengths. Any spectral changes or changes in optical path lengths in the interferometer affect the shape or position of the fringes, and the interferometer measures these changes.

Common-path interferometers, like the Mach–Zender or Jamin interferometers, are naturally compensated interferometers and can be used for measurement of dispersion in gases. Other interferometers, such as the Twyman–Green or Fizeau interferometers, make use of their unequal interferometer arms and their nonlocalized fringes from laser sources for testing different optical elements and systems in reflection and transition (*see* **Interferometry:** Phase Measurement Interferometry).

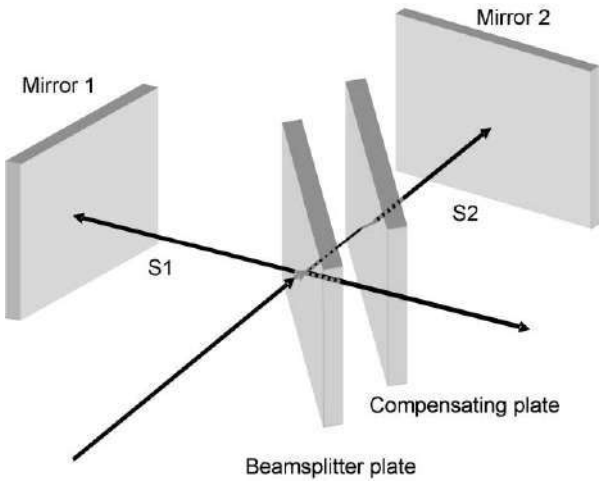

Figure 3 Michelson interferometer.

White Light Interference

A white light source consists of a wide spectrum of wavelengths in visible spectrum, from about 380 up to 750 (violet to red) nanometers. However, the principles of WLI described in this article basically apply when any low coherence source is used. Low coherence refers to a source of not only low temporal but also low spatial coherence; WLI can also be referred to as low coherence interferometry. We will be concerned mainly with temporal effect of the source, the source spectrum.

Since different wavelengths from the source spectrum are mutually incoherent, we will first look at interference between two waves of a selected monochromatic illumination with wave number $k = 2\pi/\lambda$, where λ is the wavelength. The intensity of the interference fringes at point x, y (these coordinates are omitted in all equations), as one of the mirrors is scanned along the optical axis, z can be described as

$$I(k,z) = I_1 + I_2 + 2\sqrt{I_1 I_2}|\gamma(z)|\cos(kz) \qquad [1]$$

or can be written in the form:

$$I(k,z) = I_0\left[1 + \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}|\gamma(z)|\cos(kz)\right] \qquad [2]$$

where I_1 and I_2 are intensities of each of the beams and $I_0 = I_1 + I_2$, $|\gamma(z)|$ is the modulus of the complex mutual coherence function (*see* **Coherence:** Overview) assumed here to equal 1 (perfect coherence for each wavelength). The optical path difference z equals $z_1 - z_2$ where z_1 and z_2 are the optical path lengths that the interfering waves have traveled. The difference in the traveled paths, z, corresponds to the position of the scanning mirror.

White light interference (polychromatic interference) is the overlaying of all the monochromatic fringes created for each wavelength of the source spectrum (Figure 4a). A detector observes the sum of all the fringe intensities.

In mathematical form, this interference can be described as the integral of all the fringes $I(k,z)$ for all

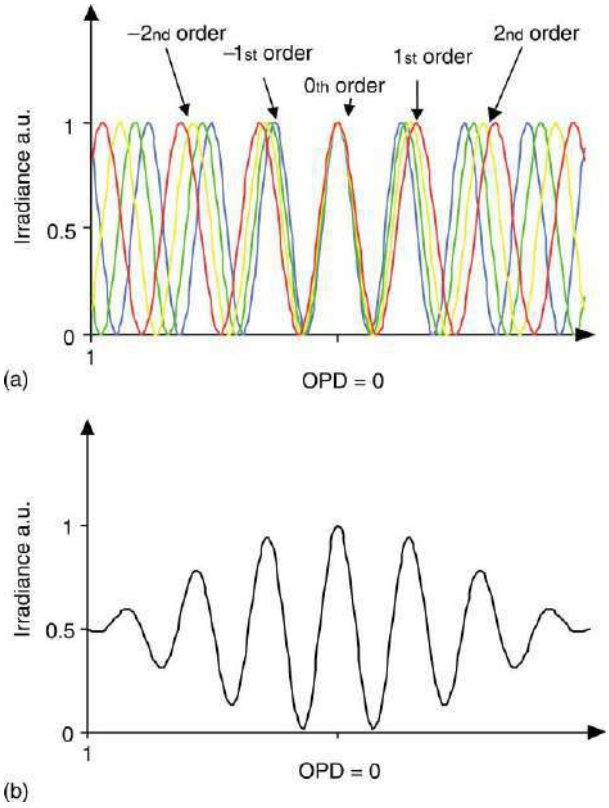

Figure 4 Formation of white light fringes: (a) fringes for individual wavelengths; (b) fringes for white light.

wavenumbers k:

$$I(z) = \int_{k_1}^{k_2} S(k)D(k)I(k,z)\mathrm{d}k \qquad [3]$$

where $S(k)$ is the spectral distribution of the light source, with $S(k) = 0$ outside of k_1 through k_2 range of wave numbers, and $D(k)$ is a spectral response of the detector. We assume that the detector's response $D(k)$ equals 1 over the whole spectrum. Because the spacing of the fringes for each wavelength emitted by the white light source is different, the maxima of fringes will align around only one point for a well-compensated interferometer, as shown in Figure 4a. This alignment around a single point, occurs because in a well-compensated interferometer there is one point for which the OPD is zero for all wavelengths. Away from this point, the observed sum of the intensities quickly falls off, as shown in Figure 4b. The maximum fringe, the fringe that marks the zero OPD, is called the zero-order fringe, and each next fringe of smaller amplitude on either side is called $+1$ and -1, $+2$ and -2 order fringe, and so on.

Because white light fringes are localized and can only be found within microns or tens of microns of the zero OPD, white light interferometers are excellent distance and 3D position sensors. However, this same characteristic makes them more difficult to align than interferometers with nonlocalized fringes.

Envelope of Fringes Due to Source Spectrum

The resultant intensity of interference from a broad spectrum source (see eqn [3]) can be described in general form as

$$I(z) = I_0[1 + V(z)\cos(k_0 z)] \qquad [4]$$

where $I_0 = I_1 + I_2$ is the background intensity, $V(z)$ is the fringe visibility function or coherence envelope, and $k_0 = 2\pi/\lambda_0$ is the central wave number for fringes under the envelope. The $V(z)$ is proportional to the modulus of the Fourier transform of the source spectrum $S(k)$. Generally, if the light source has a Gaussian spectrum $S(k)$, then the envelope of the fringes can be described also as a Gaussian function $V(z)$. If the spectrum of the source is rectangular, then the envelope of the fringes will be a sinc function. The wider that the spectrum of the source is the narrower the width of the envelope will be, as shown in Figure 5. The width of the fringe envelope determines the coherence length of the source (*see* **Coherence:** Overview); for a white light source this width is in the order of 1–2 microns.

Different white light sources, such as tungsten-halogen, incandescent, or arc lamps, have different spectra and thus create different coherence envelopes, as shown in Figure 5. The spectra of semiconductor light sources, such as LEDs and SLDs, are similar in shape to a Gaussian function.

The fact that the coherence envelope is related to the spectrum of the source by its Fourier transform is commonly used in Fourier transform spectroscopy, where the Fourier transform of the detected fringes is calculated to find the spectral components of the beams.

Position of Fringes Under Envelope Due to Reflection of Dielectric

Thus far we have assumed that in WLI the interferometer is compensated for all wavelengths, and thus, for this position the maximum of the fringes aligns with the maximum of the envelope, namely where there is a zero phase shift $\varphi_0 = 0$ between fringe and envelope maxima.

If there is an odd number of reflections from dielectric surfaces in one arm of the interferometer and an even number in the other, the fringes will be shifted by $\varphi_0 = 180°$ under the coherence envelope and the minimum of the fringe will align with the maximum of the coherence envelope. Thus eqn [4]

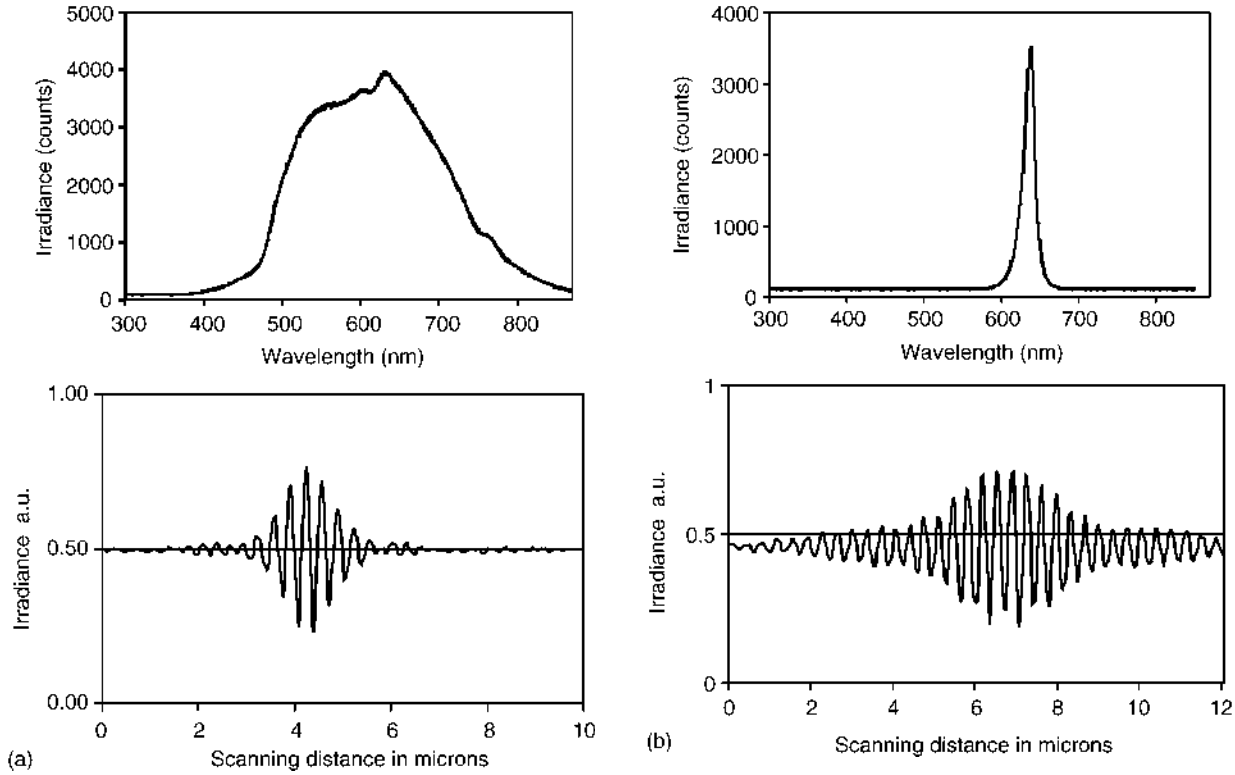

Figure 5 Spectrum and interferogram for (a) tungsten-halogen lamp and for (b) red LED sources.

can be expressed in a more general form:

$$I(z) = I_0[1 + V(z)\cos(k_0 z + \varphi_0)] \qquad [5]$$

Figure 6 shows color fringes for such a case; the dark fringe marks the zero OPD, and this fringe is surrounded by the greenish blue colors of shorter wavelengths. In contrast, the bright fringe at the maximum envelope position or marking the zero OPD would be surrounded by the reddish colors of longer wavelengths.

In a real system the fringes may be shifted with respect to the envelope by any amount of φ_0, and this shift may be due to any number of factors, such as the phase shift on reflection from nondielectric surfaces and dispersion, which we consider next.

Changes in Envelope and Fringes Due to Reflection of Nondielectric Materials

The relative position of the envelope maximum and fringe position will be different if the beam is reflected from different nondielectric materials. This difference exists because the phase change on reflection from a test surface, like metals or heavily doped semiconductors, varies with each wavelength. This variance in fringe and peak coherence position may be predicted and corrected for in surface height calculations. The linear dependence of the phase change on reflection versus wave number shifts the location of the coherence envelope peak position of fringes by different amounts. The constant phase change on reflection and higher-order terms only shift the fringes underneath the coherence envelope. As long as the change on reflection has a small dependence on the second- and higher-order of the wave number, the shape of the coherence envelope is preserved.

Changes in Envelope and Fringes Due to Dispersion

Dispersion in an interferometer that is not balanced, perhaps because a dispersive element was placed in one arm or the compensating plate has an incorrect thickness, will influence fringe formation. The phase delay between interference patterns for individual wavelengths is proportional to the product of the geometrical path and the index of refraction equal to $d \times n(k)$.

The intensity may be described as:

$$I(z) = \int_{k_1}^{k_2} \langle 1 + V(z)\cos\{kz - kd[n(k)]\}\rangle \, \mathrm{d}k \qquad [6]$$

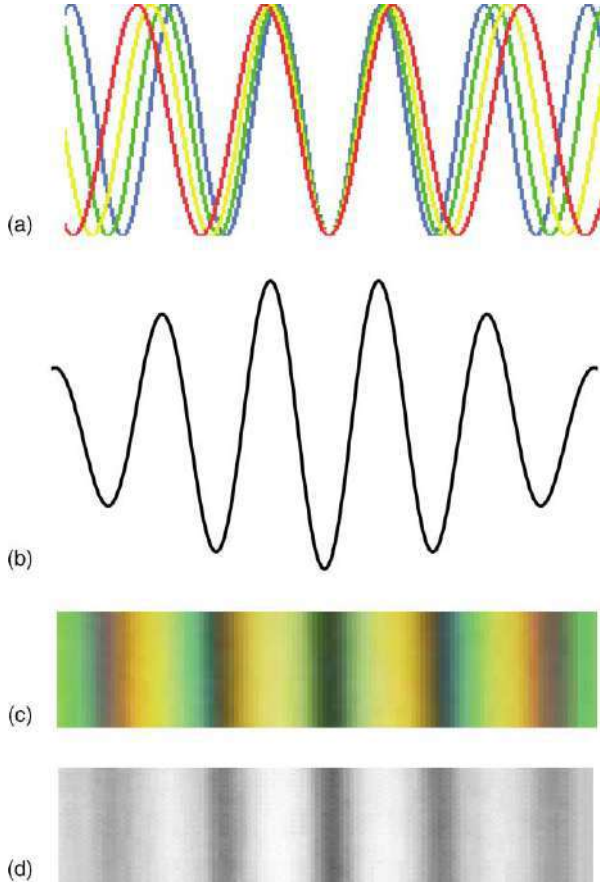

Figure 6 Formation of white light fringes with destructive interference for OPD = 0.

The dependence of the refractive index on the wave number k can be described as a linear expansion:

$$n(k) = n(k_0) + \frac{\mathrm{d}n}{\mathrm{d}k}(k - k_0) \qquad [7]$$

The linear dispersion shifts the envelope by the group index of refraction times the thickness of the dispersive element; this dispersion also shifts the fringes under the envelope slightly. In other words, the linear phase delay for a range of wavelengths causes a group delay of the whole envelope (wave packet). Higher-order dispersion, absorption of the elements, or effects due to thin films can cause the envelope to widen or even become asymmetrical, the position of the fringes under the envelope to shift, the fringes to lose contrast, and the period of the fringes to change or vary with the z position. Dispersion effects will be stronger for sources that have a wider spectrum; however, the observed changes will be different for different shapes of spectra.

The phase of the fringes under the envelope and the position of the envelope itself are parameters often used in astronomy. Highly accurate white light fringe estimation, using the optical path-length delay between the two arms of the interferometer, is a cornerstone of stellar interferometry.

Fourier Analysis of White Light Interferograms

Wavelength-dependent changes in a white light interferogram can be more easily analyzed in the spectrum rather than the spatial domain. The Fourier transform of a white light interferogram yields two symmetrical side lobes at the mean wavelength of the interferogram; analysis of just one of these side lobes is sufficient. The spectral amplitude of the side lobe contains information about the spectral components of the interfering beams, while the spectral phase in regions with appreciable spectral amplitude supplies information about any unbalanced dispersion in the interferometer, as shown in Figure 7. Fourier transform analysis is extensively used in Fourier transform spectrometry (*see* **Spectroscopy:** Fourier Transform Spectroscopy).

For a dispersion-balanced interferometer, the interferogram is symmetrical around the zero OPD position. For a symmetrical interferogram, the spectral phase will be zero if the zero OPD position of the sampled interferogram is in the middle of sampling range; otherwise a linear factor appears in the spectral phase. This linear term, while useful to surface profiling, because it determines the position of the coherence signal with respect to scanner sampling, is unwanted in Fourier spectroscopy and needs to be corrected for. Dispersion and possible thin-film effects will commonly introduce a nonlinearity in the spectral phase.

Controlled Phase Shift of Fringes Under the Envelope – Geometric Phase Shift

Many techniques in interferometry depend on shifting the phase of the interfering wavefronts. Mechanical shifters used in white light interferometry will introduce the same shift measured in nanometers for all wavelengths; however, when measured in degrees or radians, the shift will vary for different wavelengths. Geometric phase shifters (achromatic phase shifters operating on the principle of geometric phase) introduce for all wavelengths the same shift when measured in degrees and are based on polarization elements like a rotating wave plate in circularly polarized light or a rotating polarizer in a circularly polarized beam. Fringes underneath the coherence envelope shift, as shown in Figure 8, while the coherence envelope stays in place.

The advantage to these techniques is that only the phase of fringes changes, not the fringe contrast

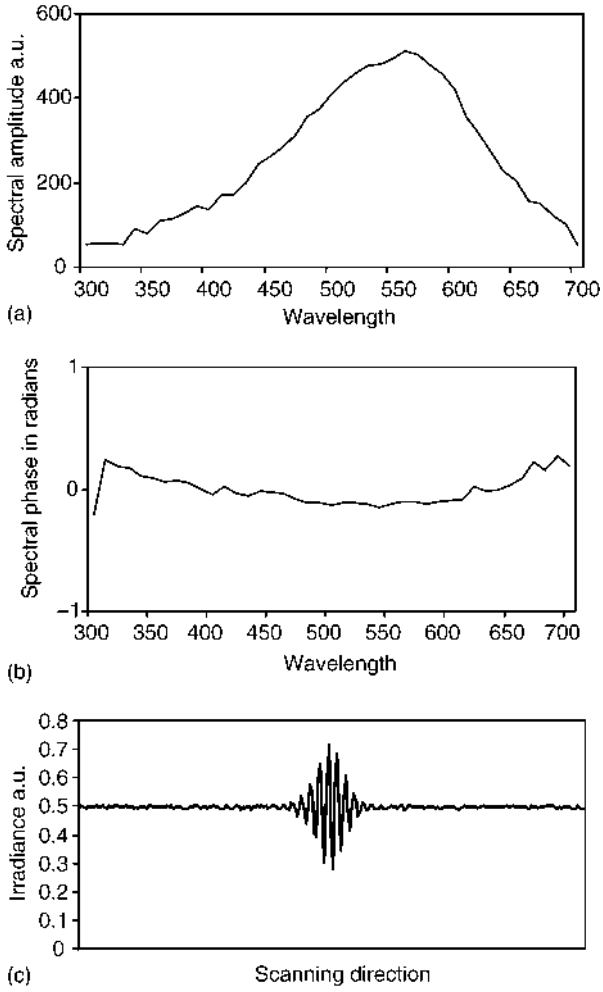

Figure 7 Spectral amplitude (a) and spectral phase (b) for input from unbalanced interferometer (c).

(*see* **Interferometry:** Phase Measurement Interferometry). A geometrical phase shifter can be very useful in polarization microscopes (Figure 9), white light shearing interferometers, or any system where the phase of the white light fringes needs to be measured.

Spectral Interference

White light fringes, because they are made up of fringes of many wavelengths, are observed only over very small path differences. If we filter only a narrow band from the white light spectrum, fringes would be visible over a much larger scan. For different wavelengths we would observe different color and frequency fringes; this is simply the reverse process of the description in the white light interference section.

If we place a spectrometer in the observation plane of the white light interferometer (Figure 10), we will observe fringes with continuously changing wavelengths in dispersed light. These fringes are called

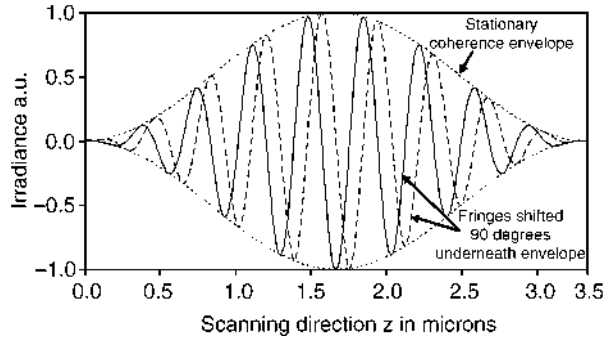

Figure 8 Fringes for two different geometric phase shifts.

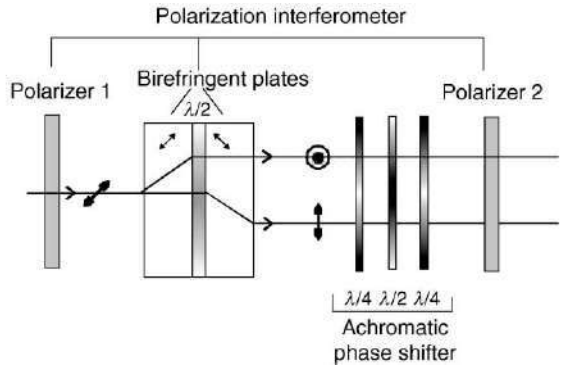

Figure 9 Polarization interferometer with geometric phase shift.

fringes of chromatic order or channeled spectrum fringes and find their application in film thickness measurement or absolute distance measurement in the range up to 1 mm. Channeled spectra were used for analysis of dispersion, thin film, and spectroscopic measurements. The number of observed fringes for a given wavelength is directly proportional to the measured optical path difference. The optical path difference can be determined if the difference in fringe numbers is determined for two well-known wavelengths (this is equivalent to two-wavelength interferometry). The optical path difference can be also quickly estimated from the frequency of fringes for a given wavelength; the larger the optical path difference, the denser the fringes.

Surface Topography and Object Structure Measurement

Although WLI has many applications, this section focuses on white light interference as applied to surface topography measurement. Interference microscopes that use white light illumination are often based on the Michelson setup shown in Figure 11.

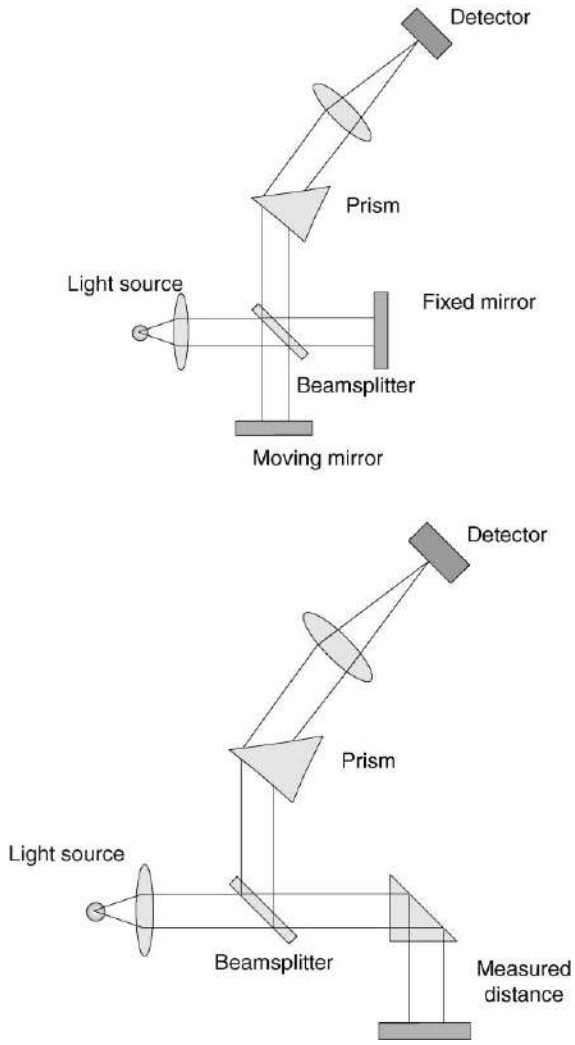

Figure 10 Interferometers with prism spectrometer to observe fringes of equal chromatic order for OPD measurement.

A beamsplitter cube is placed underneath the bright field objective, and one mirror is placed to the side at the focus plane of the objective while the other mirror is replaced by the measured surface with respect to which the whole interference objective is scanned. The narrowly localized fringes at the best focus position for each point of the surface corresponding to a pixel on the CCD camera is observed during the scan. Because of the design of the Michelson objective, in order to accommodate higher magnification objectives constraints, other interference setups like Mirau and Linnik were developed (*see* **Microscopy:** Interference Microscopy). However, these designs are still based on the Michelson interferometer, being also compensated interferometers with equal lengths and amounts of glass in each arm.

These interference microscopes typically have two complementary modes of operation; one mode uses

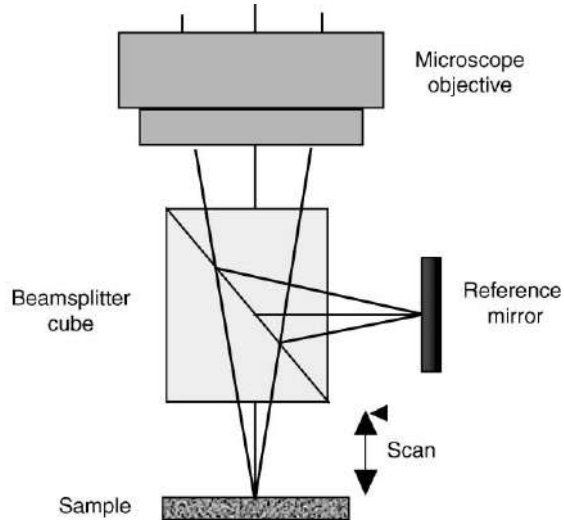

Figure 11 Michelson interferometric objective.

monochromatic illumination and the other employs a white light source. Monochromatic illumination provides excellent vertical resolution but is limited in its range. It cannot correctly assign fringe order to discontinuous surfaces larger than 160 nanometers. Figure 12a shows monochromatic fringes for the surface of the profile shown in Figure 13. We see that with the monochromatic illumination the height of the grooves remains unknown because it is impossible to assign order numbers to these fringes.

To resolve this ambiguity, white light illumination is employed because it allows for easy identification of the zero order fringe. Figure 12b shows white light fringes created for the same surface. We see the zero-order fringe for the zero OPD as well as the next orders due to the decreasing contrast of the fringes. The position of the zero-order fringe can be followed and approximate heights can be determined visually. Thus, white light illumination permits the measurement of a broader range of surfaces that are rough or have large height variations up to a few millimeters.

The white light interferometer is a great focus detector (Figure 14) at each point of the field of view. The principle behind WLI is finding these individual focus positions using localized fringes observed during the surface scan. WLI interferometers provide high lateral resolution and large vertical range in the order of hundreds of nanometers or micrometers (Figure 15). White light interferometers are commonly used to measure magnetic heads, MEMS devices, binary optic, and machined surfaces.

Signal Processing of White Light Interferograms

For WLI topographic measurements, each pixel registers an interferogram whose position varies

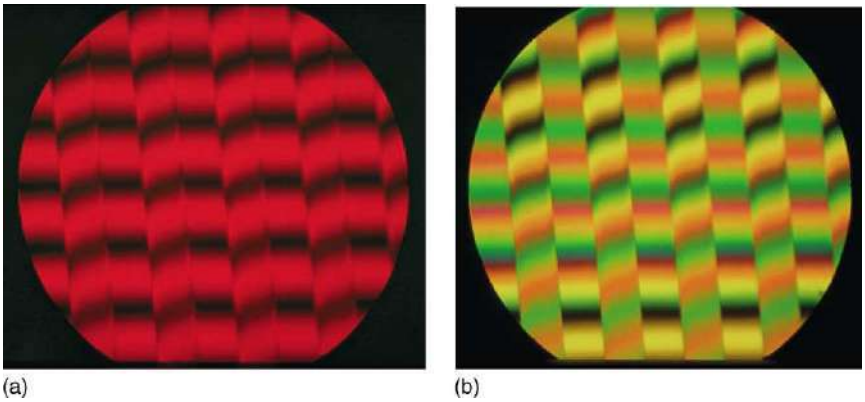

(a) (b)

Figure 12 Fringes in quasimonochromatic and white light for object similar to one as presented in Figure 13.

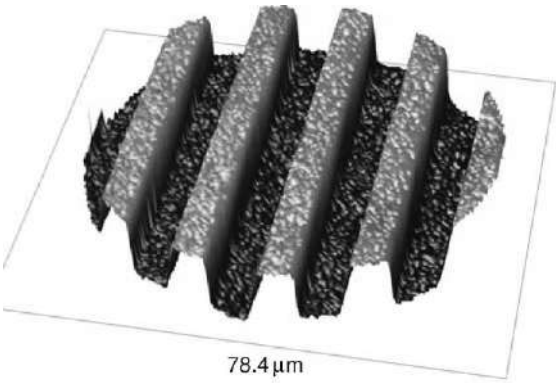

Figure 13 3D object profile – binary grating.

with the surface height. Figure 16 shows two interferograms for two positions in the field of view of a measured step height surface; the first shows the top surface of the step height and the second shows the bottom surface of the step height.

Algorithms have been developed to analyze the signal in different ways. Some focus on finding the fringe envelope position and others examine the position of the fringe underneath the envelope. These algorithms work under the assumption that white light fringes can be described by the same function over the whole field of view of the interferometer. Algorithms are applied to a signal so as to find the position of the coherence envelope for each pixel. This process is called coherence gating, and all algorithms perform coherence gating.

Some algorithms look for the position of the envelope's maximum in a process whereby first the coherence envelope of the fringes is found using signal filtering. Then, the curve is fit to a few points around the envelope's maximum, and finally the position of the envelope's maximum is found. Other algorithms calculate the center of mass of the coherence signal, as described in eqn [8]:

$$H(x,y) = \frac{\sum_{n=1}^{N-1} \left(I_i(x,y) - I_{i+1}(x,y)\right)^2 z_i}{\sum_{n=1}^{N-1} \left(I_i(x,y) - I_{i+1}(x,y)\right)^2} \qquad [8]$$

This method is very fast and computationally efficient. Center of mass calculations are equivalent to calculations of the maximum of the envelope position, but only for a symmetrical signal. For an asymmetrical signal a piston is introduced for each point; however, this piston does not affect the whole measurement. Still another method calculates the spectral phase slope in the Fourier domain to determine the position of the coherence envelope in a process that is equivalent to finding the center of mass. Finally, a different group of algorithms tracks the position of the bright or dark fringe close to the envelope's maximum.

Scanner Speed – Sampling Rate

A practical sampling rate of the white light interference signal is around four pixels per fringe. Because in most algorithms it is not the position of the fringes that we want to determine but rather the position of the envelope of the fringes, sampling the envelope of the fringes with four samples is usually sufficient. However, this sufficiency comes at the expense of lower repeatability (higher noise). The advantage to this sampling rate is short measurement time; higher sampling rates increase measurement time. Measurement speed can be increased by up to 25 times when the envelope of the fringes is widened by filtering light of a bandwidth about 20–40 nm from the white light source.

Scanner Nonlinearity

Scanner motion is assumed to be linear, but nonlinearity in this motion impairs measurement

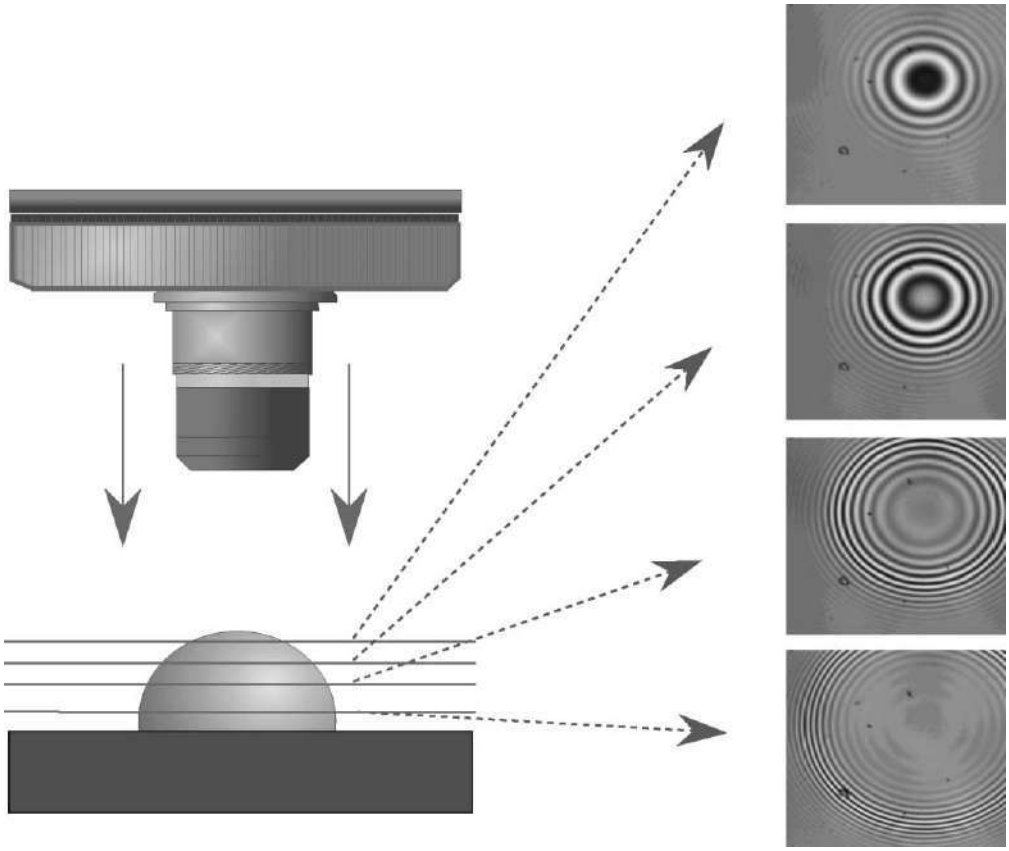

Figure 14 Operation of white light interferometer for surface topography measurement.

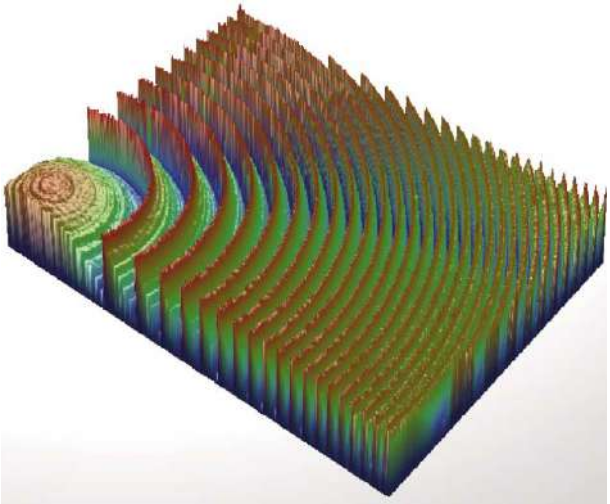

Figure 15 WLI measurement of Fresnel lens in a high-resolution mode. Courtesy of Glimmerglass.

accuracy. To account for this, simple algorithms can calculate the average steps along each interferogram in scanning direction (Figure 17) and measured steps can be used in coherence detection algorithms.

In order to have continuous information about the scan steps, the fringes should be visible somewhere in the field of view at each moment of the scan. This task may require introducing a large tilt for samples with large discontinuities such as stepped surfaces. For large discontinuities where introducing sample tilt may not be sufficient, the envelope of the fringes can be extended by increasing the spatial or temporal coherence of the light source. An alternate solution involves measuring the scanner motion with a distance-measuring interferometer or triggering the camera to collect intensity data at every equal scan step, which can be determined by using the zero-crossing technique, as is commonly done in Fourier transform spectroscopy. Equal scan steps are more suitable for techniques based on transform techniques that assume equal sampling. Other peak detection algorithms, such as center of mass calculations, make use of measured but not necessarily equal steps directly in the algorithm.

Observing the phase along the interferogram can also provide information that is important in optical fiber sensing. These observed changes, which assume that sampling rates of the interferogram are known,

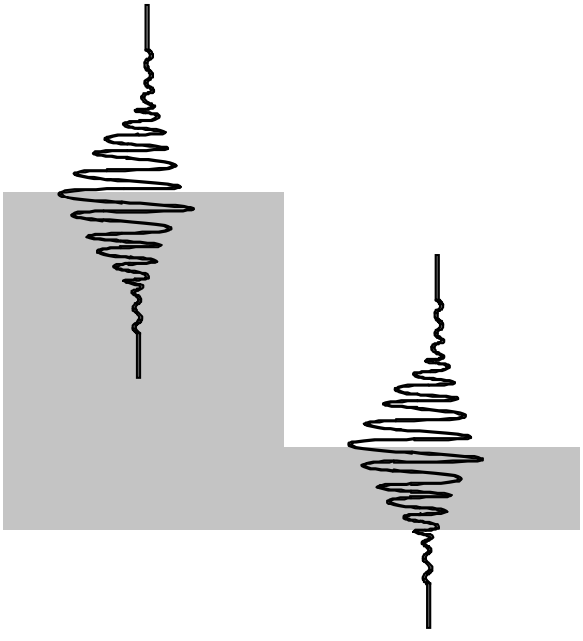

Figure 16 White light fringes as seen by two pixels during scan through focus for object in form of step.

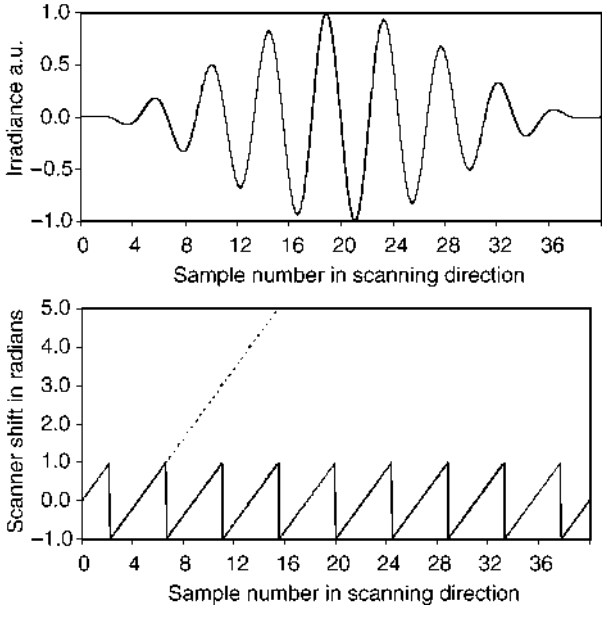

Figure 17 Phase calculated along the white light interferogram. This calculated phase can be used to determine scanner motion.

enable correction for changes in wave number value k_0 which can be due to such things as changes in the working voltage of bulb, an introduced higher-order dispersion, or a large tilt of the sample.

Increased Resolution White Light Interferometry

Interferometric methods that employ a monochromatic light source to detect the phase of the fringes

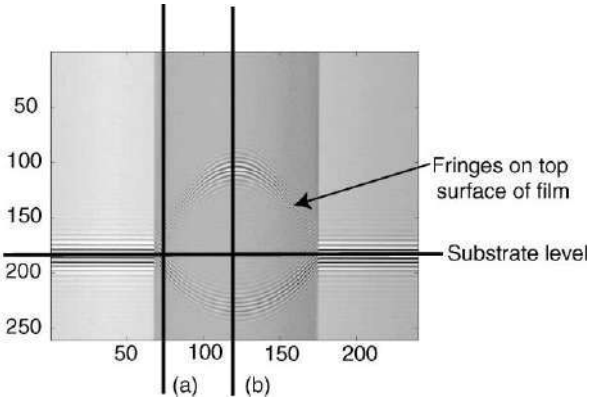

Figure 18 Fringes for (a) thin and (b) thick film.

(*see* **Interferometry:** Phase Measurement Interferometry) can achieve about 10 times better (0.3 nanometers versus 3 nanometers) vertical resolution than the WLI methods described so far. Combining coherence position detection from WLI to determine fringe order with the phase detection of phase shifting techniques allows for the measurement of samples with height discontinuities larger than 160 nanometers with the resolution and accuracy of phase shifting interferometry (PSI). This combination is particularly well-suited for determining the shape of smooth surfaces with large height differences such as binary diffractive optics or micro-electromechanical systems (MEMS) (see Figure 15).

Using this combined method we obtain both a lower-resolution map of the envelope position and a higher-resolution map of the phase (position) of the zero order fringe. These maps may differ slightly due to effects similar to those discussed in sections *Changes in Envelope and Fringes Due to Reflection of Nondielectric* and *Changes in Envelope and Fringes Due to Dispersion* above. In interference microscopes, shifts in the envelope and fringe position may be introduced by field-dependent and chromatic aberrations of the system and the system's effective numerical aperture. These shifts can vary for different points on the tested surface, but for simplicity are assumed to be constant over the field. Some correction for these effects can be applied.

Film Thickness Measurement

White light interferograms that are obtained in an optical profiler can be used to measure transparent film thicknesses because the character of the interferogram changes with the thickness of the film (Figure 18). Two different techniques, a thin or thick film technique, are used depending of the range of the film thickness. A thick film technique is used if two

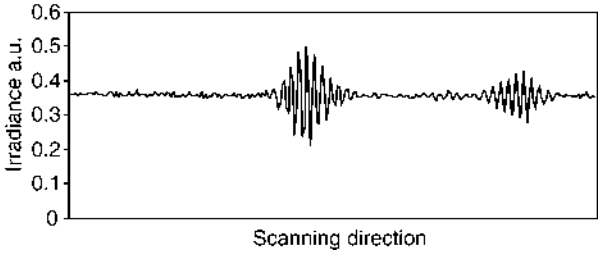

Figure 19 Thick film (see **Figure 18b**) irradiance.

sets of best contrast white light fringes from each interface are clearly separated, meaning that no interference occurs between the wavefronts reflected from the top and bottom surfaces of the film. A thin film technique is employed when the two sets of fringes overlap.

Thick film measurement

A simple technique for finding the relative position of the peaks of the fringe envelopes can be implemented to find the thickness of a film. Figure 19 shows two clearly separated sets of fringes formed for the air/film and film/substrate interfaces. Whether these two sets of fringes are separated or not depends mainly on the geometrical thickness of the film and its group index of refraction, which is determined by the dispersion of the film. The typical range of measurable film thicknesses runs from 3 to 150 microns, depending on the dispersion of the film. This measurement allows for the detection of flows on the surface and interface of the film. Similar principles based on finding the position of the coherence envelopes are used for distance sensing, thickness measurement of plates, and the cornea of the eye, in low coherence reflectometry and structure measurements in biological samples in optical coherence tomography.

Thin film measurement

For thicknesses of five microns down to tens of nanometers, the white light interferogram is created from the interference between the beams reflected from the reference mirror and the two beams reflected from the thin film layer. Once the interferogram is registered while the objective is scanned vertically with a constant velocity, the spectral phase is calculated by applying a Fourier transform to the measured signal at each pixel as described in the section *Fourier Analysis of White Light Interferograms* above. The phase slope is subtracted and the dispersion of the system needs to be known. The spectral phase for the thin film interference has the form of a polynomial; thus, the polynomial for the chosen film model (n and k) is fitted, and regression

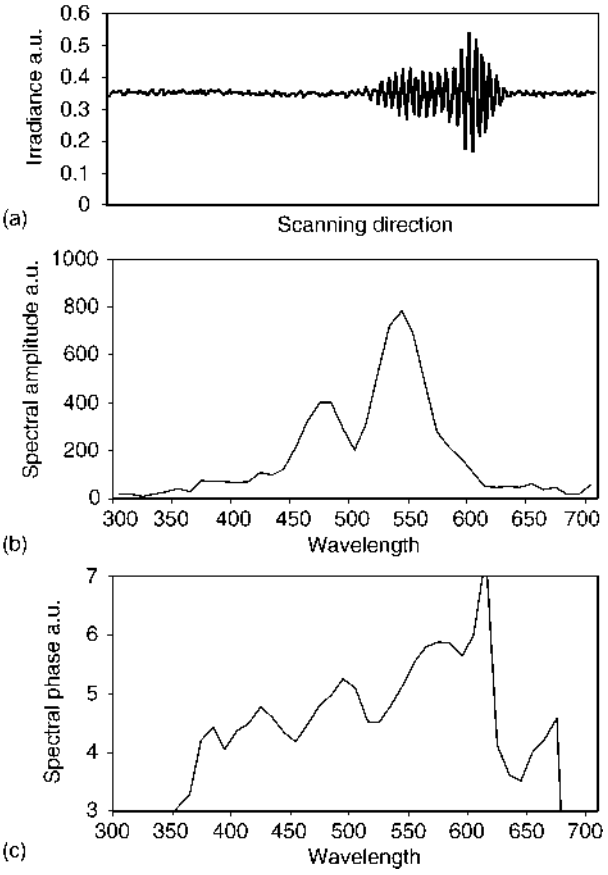

Figure 20 Thin film (see **Figure 18a**): (a) irradiance, (b) spectral amplitude and (c) spectral phase.

analysis is used to find the best fit and, therefore, the film thickness (Figure 20).

Spatial Coherence Effects in the Interference Microscope

So far we have been discussing temporal coherence effects, but everywhere that the source has a wavelength bandwidth, spatial coherence may play an important role. In an interference microscope for surface topography measurement, the size of the aperture of the condenser is on the order of the aperture of the objective that illuminates the surface with a wide range of angles. For large angles, which are determined by the numerical aperture (NA) of the objective, where NA = 0.5–0.95, the combination of the influence of the spatial and temporal coherence is clearly visible. The additional spatial effects include a reduction of the fringe envelope width and a decrease in fringe spacing corresponding to k_0. Calibration of fringe spacing is typically done on the system and accounts for spatial effects as well as for any uncertainty of k_0 due to the source spectra, the working temperature of the source, the spectral

response of the detector and the influence of other factors like the intensity distribution in the illuminating aperture. The spatial influences can be reduced by stopping down the condenser, which causes an increase in the contrast of the fringes.

Rough Surfaces

Rough surfaces are difficult to measure using interferometric techniques, but under certain coherence conditions white light interference can do the job. For rough surfaces, if the microstructure of the object is not resolved by the imaging system, speckles, rather than fringes, are observed. Each speckle has a random phase which is approximately constant in the whole speckle area. If the rough surface is scanned through focus, each individual speckle exhibits the intensity modulation that is typical for WLI. These speckles enable the measurement, but they also introduce noise proportional to the roughness of the measured surface.

Despite the noise that speckles introduce into the WLI measurement, WLI has an advantage because it rejects the light that has undergone scattering outside of a small sample volume, thus allowing precise noninvasive measurement of object structure, even in dense media.

Applications

WLI is used in many disciplines and instruments such as:

- Fourier transform spectroscopy – source and material properties;
- Michelson stellar interferometer – angular size of star, binary stars, delay measurement in optical paths of interferometer;
- Shearing interferometry – structure measurement;
- DIC Nomarski interferometry – structure measurement;
- Speckle interferometry – structure measurement;
- Holography – structure measurement;
- Optical sensors – temperature, pressure, distance;
- Optical coherence tomography – structure measurement.

See also

Coherence: Coherence and Imaging; Overview. **Holography, Techniques:** Holographic Interferometry. **Interferometry:** Phase Measurement Interferometry. **Microscopy:** Interference Microscopy. **Spectroscopy:** Fourier Transform Spectroscopy.

Further Reading

Born M and Wolf E (1997) *Principles of Optics*. New York: Cambridge University Press.

Creath K (1997) *Sampling requirements for white light interferometry*. Fringe'97 Proceedings of the 3rd International Workshop on Automatic Processing of Fringe Patterns, pp. 52–59.

Danielson BL and Boisbert CY (1991) Absolute optical ranging using low coherence interferometry. *Applied Optics* 30: 2975–2979.

Davis SP, Abrams MC and Brault JW (2001) *Fourier Transform Spectrometry*. London: Academic Press.

de Groot P and Deck L (1995) Surface profiling by analysis of white light interferograms in the spatial frequency domain. *Journal of Modern Optics* 42: 389–401.

Dresdel T, Hausler G and Venzke H (1992) Three dimensional sensing of rough surfaces by coherence radar. *Applied Optics* 31(7): 919–925.

Harasaki A, Schmit J and Wyant JC (2001) Offset envelope position due to phase change on reflection. *Applied Optics* 40: 2102–2106.

Hariharan P (1992) *Basics of Interferometry*. New York: Academic Press.

Hariharan P, Larkin KG and Roy M (1994) The geometric phase: interferometric observation with white light. *Journal of Modern Optics* 41: 663–667.

Larkin KG (1996) Neat nonlinear algorithm for envelope detection in white light interferometry. *Journal of the Optical Society of America A* 13: 832–842.

Michelson AA and Benoit JR (1895) Détermination expérimentale de la valeur du métre en longueurs d'ondes lumineuses. *Travaux et Mémoires des Bureau International des Poids et Mesures* 11: 1.

Olszak AG and Schmit JG (2003) High stability white light interferometry with reference signal for real time correction of scanning errors. *Optical Engineering* 42(1): 54–59.

Park M-C and Kim S-W (2000) Direct quadratic polynomial fitting for fringe peak detection of white light scanning interferograms. *Optical Engineering* 39: 952–959.

Pavlicek P and Soubusta J (2004) Measurement of the influence of dispersion on white light interferometry. *Applied Optics* 43: 766–770.

Schmit JG and Olszak A (2002) High precision shape measurement by white light interferometry with real time scanner error correction. *Applied Optics* 41(28): 5943–5950.

Sheppard CJR and Larkin KG (1995) Effects of numerical aperture on interference fringe spacing. *Applied Optics* 34: 4731–4734.

Sheppard CJR and Roy M (2003) Low-coherence interference microscopy. In: Torok P and Kao F-J (eds) *Optical Imaging and Microscopy*, pp. 257–273. Berlin, Heidelberg, New York: Springer-Verlag.

Steel W (1987) *Interferometry*. New York: Cambridge University Press.

Turyshew SG (2003) Analytical modeling of the white light fringe. *Applied Optics* 42(1): 71–90.

L

LASERS

Contents

Carbon Dioxide Laser
Dye Lasers
Edge Emitters
Excimer Lasers
Free Electron Lasers
Metal Vapor Lasers
Noble Gas Ion Lasers
Optical Fiber Lasers
Organic Semiconductors and Polymers
Planar Waveguide Lasers
Semiconductor Lasers
Up-Conversion Lasers

Carbon Dioxide Laser

C R Chatwin, University of Sussex, Brighton, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

This article gives a brief history of the development of the laser and goes on to describe the characteristics of the carbon dioxide laser and the molecular dynamics that permit it to operate at comparatively high power and efficiency. It is these commercially attractive features and its low cost that has led to its adoption as one of most popular industrial power beams. This outline also describes the main types of carbon dioxide laser and briefly discusses their characteristics and uses.

Brief History

In 1917 Albert Einstein developed the concept of stimulated emission which is the phenomenon used in lasers. In 1954 the MASER (Microwave Amplification by Stimulated Emission of Radiation) was the first device to use stimulated emission. In that year Townes and Schawlow suggested that stimulated emission could be used in the infrared and optical portions of the electromagnetic spectrum. The device was originally termed the optical maser, this term being dropped in favor of LASER, standing for: Light Amplification by Stimulated Emission of Radiation. Working against the wishes of his manager at Hughes Research Laboratories, the electrical engineer Ted Maiman created the first laser on the 15 May 1960. Maiman's flash lamp pumped ruby laser produced pulsed red electromagnetic radiation at a wavelength of 694.3 nm. During the most active period of laser systems discovery Bell Labs made a very

significant contribution. In 1960, Ali Javan, William Bennet and Donald Herriot produced the first Helium Neon laser, which was the first continuous wave (CW) laser operating at 1.15 μm. In 1961, Boyle and Nelson developed a continuously operating Ruby laser and in 1962, Kumar Patel, Faust, McFarlane and Bennet discovered five noble gas lasers and lasers using oxygen mixtures. In 1964, C.K.N. Patel created the high-power carbon dioxide laser operating at 10.6 μm. In 1964, J.F. Geusic and R.G. Smith produced the first Nd:Yag laser using neodymium doped yttrium aluminum garnet crystals and operating at 1.06 μm.

Characteristics

Due to its operation between low lying vibrational energy states of the CO_2 molecule, the CO_2 laser has a high quantum efficiency, ~40%, which makes it extremely attractive as a high-power industrial materials processing laser (1 to 20 kW), where energy and running costs are a major consideration. Due to the requirement for cooling to retain the population inversion, the efficiency of electrical pumping and optical losses – commercial systems have an overall efficiency of approximately 10%. Whilst this may seem low, for lasers this is still a high efficiency. The CO_2 laser is widely used in other fields, for example, surgical applications, remote sensing, and measurement. It emits infrared radiation with a wavelength that can range from 9 μm up to 11 μm. The laser transition may occur on one of two transitions: $(00^01) \rightarrow (10^00)$, $\lambda = 10.6$ μm; $(00^01) \rightarrow (02^00)$, $\lambda = 9.6$ μm, see Figure 1. The 10.6 μm transition has the maximum probability of oscillation and gives the strongest output; hence, this is the usual wavelength of operation, although for specialist applications the laser can be forced to operate on the 9.6 μm line.

Figure 1 illustrates an energy level diagram with four vibrational energy groupings that include all the significantly populated energy levels. The internal relaxation rates within these groups are considered to be infinitely fast when compared with the rate of energy transfer between these groups. In reality the internal relaxation rates are at least an order of magnitude greater than the rates between groups.

Excitation of the upper laser level is usually provided by an electrical glow discharge. However, gas dynamic lasers have been built where expanding a hot gas through a supersonic nozzle creates the population inversion; this creates a nonequilibrium region in the downstream gas stream with a large population inversion, which produces a very high-power output beam (135 kW – Avco Everett Research Lab). For some time the gas dynamic laser was seriously considered for use in the space-based Strategic Defence Initiative (SDI-USA). The gas mixture used in a CO_2 laser is usually a mixture of carbon dioxide, nitrogen, and helium. The proportions of these gases varies from one laser system to another, however, a typical mixture is 10%-CO_2;

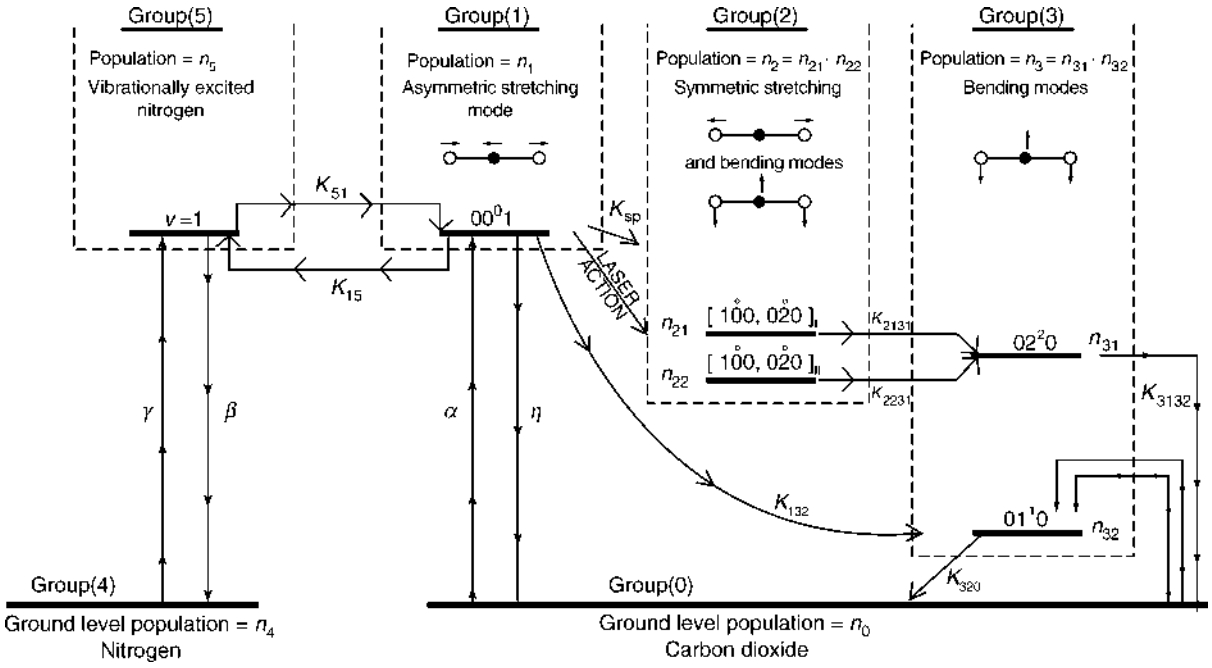

Figure 1 Six level model used for the theoretical description of CO_2 laser action.

10%-N_2; 80%-He. Helium plays a vital role in the operation of the CO_2 laser in that it maintains the population inversion by depopulating the lower laser level by nonradiative collision processes. Helium is also important for stabilization of the gas discharge; furthermore it greatly improves the thermal conductivity of the gas mixture, which assists in the removal of waste heat via heat exchangers.

Small quantities of other gases are often added to commercial systems in order to optimize particular performance characteristics or stabilize the gas discharge; for brevity we only concern ourselves here with this simple gas mixture.

Molecular Dynamics

Direct Excitation and De-excitation

It is usual for excitation to be provided by an electrical glow discharge. The direct excitation of carbon dioxide (CO_2) and nitrogen (N_2) ground state molecules proceeds via inelastic collisions with fast electrons. The rates of kinetic energy transfer are α and γ, respectively, and are given by eqns [1] and [2]:

$$\alpha = \frac{F_{CO_2} \times \mathrm{IP}}{E_{00^01} \times n_0} (\mathrm{sec}^{-1}) \qquad [1]$$

$$\gamma = \frac{F_{N_2} \times \mathrm{IP}}{E_{v=1} \times n_4} (\mathrm{sec}^{-1}) \qquad [2]$$

where:

F_{CO_2} = Fraction of the input power (IP) coupled into the excitation of the energy level E_{00^01}, n_0 is the CO_2 ground level population density;

F_{N_2} = Fraction of the input power (IP) coupled into the excitation of the energy level $E_{v=1}$; and n_4 is the N_2 ground level population density.

The reverse process of the above occurs when molecules lose energy to the electrons and the electrons gain an equal amount of kinetic energy; the direct de-excitation rates are given by η and β, eqns [3] and [4], respectively:

$$\eta = \alpha \times \exp\left(\frac{E_{00^01}}{E_e}\right) (\mathrm{sec}^{-1}) \qquad [3]$$

$$\beta = \gamma \times \exp\left(\frac{E_{v=1}}{E_e}\right) (\mathrm{sec}^{-1}) \qquad [4]$$

where E_e is the average electron energy in the discharge.

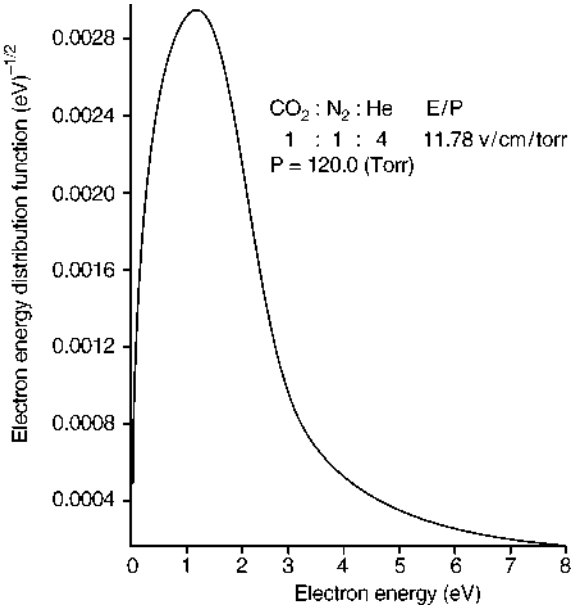

Figure 2 Electron energy distribution function.

E_e, F_{CO_2}, and F_{N_2} are obtained by solution of the Boltzmann transport equation (BTE); the average electron energy can be optimized to maximize the efficiency (F_{CO_2}, F_{N_2}) with which electrical energy is utilized to create a population inversion. Hence, the discharge conditions required to maximize efficiency can be predicted from the transport equation. Figure 2 shows one solution of the BTE for the electron energy distribution function.

Resonant Energy Transfer

Resonant energy transfer between the CO_2 (00^01) and N_2 ($v = 2$) energy levels (denoted 1 and 5 in Figure 1) proceeds via excited molecules colliding with ground state molecules. A large percentage of the excitation of the upper laser level takes place via collisions between excited N_2 molecules and ground state CO_2 molecules. The generally accepted rate of this energy transfer is given by eqns [5] and [6]:

$$K_{51} = 19\,000\, P_{CO_2}\ (\mathrm{sec}^{-1}) \qquad [5]$$

$$K_{15} = 19\,000\, P_{N_2}\ (\mathrm{sec}^{-1}) \qquad [6]$$

where P_{CO_2} and P_{N_2} are the respective gas partial pressures in Torr.

Hence, CO_2 molecules are excited into the upper laser level by both electron impact and impact with excited N_2 molecules. The contribution from N_2 molecules can be greater than 40% depending on the discharge conditions.

Collision–Induced Vibrational Relaxation of the Upper and Lower Laser Levels

The important vibrational relaxation processes are illustrated by Figure 1 and can be evaluated from eqns [7–10]; where the subscripts refer to the rate between energy levels 1 and 32; 21 and 31; 22 and 31; 32 and 0, respectively:

$$K_{132} = 367\,P_{CO_2} + 110\,P_{N_2} + 67\,P_{He}\ (\text{sec}^{-1}) \qquad [7]$$

$$K_{2131} = 6\times10^5\,P_{CO_2}\ (\text{sec}^{-1}) \qquad [8]$$

$$K_{2231} = 5.15\times10^5\,P_{CO_2}\ (\text{sec}^{-1}) \qquad [9]$$

$$K_{320} = 200\,P_{CO_2} + 215\,P_{N_2} + 3270\,P_{He}\ (\text{sec}^{-1}) \qquad [10]$$

K_{132}, K_{2131}, and K_{2231} are vibration/vibration transfer rates and K_{320} is a vibration/translation transfer rate. Note the important effect of helium on eqn [10]; helium plays a major role in depopulating the lower laser level, thus enhancing the population inversion. P_{He} is the partial pressure of helium in Torr.

Radiative Relaxation

Spontaneous radiative decay is not a major relaxation process in the CO_2 laser but it is responsible for starting laser action via spontaneous emission. The Einstein 'A' coefficient for the laser transition is given by eqn [11]:

$$A = (\tau_{sp})^{-1} = 0.213\ (\text{sec}^{-1}) \qquad [11]$$

Gain

The gain (g) is evaluated from the product of the absorption coefficient (σ) and the population inversion, eqn [12]:

$$g = \sigma\left(n_{00^01} - \frac{g_1}{g_2}n_{10^00}\right)\ \text{cm}^{-1} \qquad [12]$$

For most commercial laser systems the absorption coefficient is that for high-pressure collision-broadening ($P > 5.2$ Torr) where the intensity distribution function describing the line shape is Lorentzian. The following expression describes the absorption coefficient, eqn [13]:

$$\sigma = \frac{692.5}{Tn_{CO_2}\left(1 + 1.603\dfrac{n_{N_2}}{n_{CO_2}} + 1.4846\dfrac{n_{He}}{n_{CO_2}}\right)}\ (\text{cm}^2) \qquad [13]$$

where T is the absolute temperature and n refers to the population density of the gas designated by the subscript.

This expression takes account of the different constituent molecular velocity distributions and different collision cross-sections for $CO_2 \rightarrow CO_2$, $N_2 \rightarrow CO_2$ and $He \rightarrow CO_2$ type collisions. Equation [13] also takes account of the significant line broadening effect of helium.

Neglecting the unit change in rotational quantum number, the energy level degeneracies g_1 and g_2 may be dropped. n_{10^00} is partitioned such that $n_{10^00} = 0.1452n_2$ and eqn [12] can be re-cast as eqn [14]:

$$g = \sigma(n_1 - 0.1452n_2)\ \text{cm}^{-1} \qquad [14]$$

where n_1 and n_2 are the population densities of energy groups '1' and '2' respectively.

Stimulated Emission

Consider a laser oscillator with two plane mirrors, one placed at either end of the active gain medium, with one mirror partially transmitting (see Figure 4). Laser action is initiated by spontaneous emission that happens to produce radiation whose direction is normal to the end mirrors and falls within the resonant modes of the optical resonator. The rate of change of photon population density (I_p) within the laser cavity can be written as eqn [15]:

$$\frac{dI_P}{dt} = I_p cg - \frac{I_p}{T_0} \qquad [15]$$

where the first term on the right-hand side accounts for the effect of stimulated emission and the second term represents the number of photons that decay out of the laser cavity, T_0 is the photon decay time, given by eqn [16], and is defined as the average time a photon remains inside the laser cavity before being lost either through the laser output window or due to dispersion; if dispersion is ignored, I_p/T_0, is the laser output:

$$T_0 = \frac{2L}{c\log_e\left(\dfrac{1}{R_B R_F}\right)} \qquad [16]$$

where L is the distance between the back and the front mirrors, which have reflectivities of R_B and R_F, respectively. The dominant laser emission occurs on a rotational–vibrational P branch transition $P(22)$, that is $(J = 21) \rightarrow (J = 22)$ line of the $(00^01) \rightarrow (10^00)$, $\lambda = 10.6\ \mu\text{m}$ transition, where J is the rotational quantum number. The rotational level relaxation rate is so rapid that equilibrium is maintained between rotational levels so that they

feed all their energy through the $P(22)$ transition. This model simply assumes constant intensity, basing laser performance on the performance of an average unit volume. By introducing the stimulated emission term into the molecular rate equations, which describe the rate of transfer of molecules between the various energy levels illustrated in Figure 1, a set of molecular rate eqns [17–21] can be written that permit simulation of the performance of a carbon dioxide laser:

$$\frac{dn_1}{dt} = \alpha n_0 - \eta n_1 + K_{51}n_5 - K_{15}n_1 - K_{sp}n_1 - K_{132}\left(n_1 - \left[\frac{n_1}{n_{32}}\right]_e n_{32}\right) - I_p cg \qquad [17]$$

$$\frac{dn_2}{dt} = K_{sp}n_1 + I_p cg - K_{2131}\left(n_{21} - \left[\frac{n_{21}}{n_{31}}\right]_e n_{31}\right) - K_{2231}\left(n_{22} - \left[\frac{n_{22}}{n_{31}}\right]_e n_{31}\right) \qquad [18]$$

$$\frac{dn_3}{dt} = 2K_{2131}\left(n_{21} - \left[\frac{n_{21}}{n_{31}}\right]_e n_{31}\right) - 2K_{2231}\left(n_{22} - \left[\frac{n_{22}}{n_{31}}\right]_e n_{31}\right) + K_{132}\left(n_1 - \left[\frac{n_1}{n_{32}}\right]_e n_{32}\right) - K_{320}\left(n_{32} - \left[\frac{n_{32}}{n_0}\right]_e n_0\right) \qquad [19]$$

$$\frac{dn_5}{dt} = \gamma n_4 - \beta n_5 - K_{51}n_5 + K_{15}n_1 \qquad [20]$$

$$\frac{dI_p}{dt} = I_p cg - \frac{I_p}{T_0} \qquad [21]$$

The terms in square brackets ensure that the system maintains thermodynamic equilibrium; subscript 'e' refers to the fact that the populations in the square brackets are the values for thermodynamic equilibrium. The set of five simultaneous differential equations can be solved using a Runge–Kutta method. They can provide valuable performance prediction data that is helpful in optimizing laser design, especially when operated in the pulsed mode. Figure 3a illustrates some simulation results for the transverse flow laser shown in Figure 9. The results illustrate the effect of altering the gas mixture and how this can be used to control the gain switched spike that would result in unwelcome work piece plasma generation if allowed to become too large. Figure 3b shows an experimental laser output pulse from the high pulse repetition frequency (prf–5kHz) laser illustrated in Figure 9. This illustrates that even a quite basic physical model can give a good prediction of laser output performance.

Optical Resonator

Figure 4 shows a simple schematic of an optical resonator. This simple optical system consists of two mirrors (the full reflector is often a water cooled gold coated copper mirror), which are aligned to be orthogonal to the optical axis that runs centrally along the length of the active gain medium in which there is a population inversion. The output coupler is a partial reflector (usually dielectrically coated zinc selenide – ZnSe–that may be edge cooled) so that some of the electromagnetic radiation can escape as an output beam. The ZnSe output coupler has a natural reflectivity of about 17% at each air–solid interface. For high power lasers (2 kW) 17% is sufficient for laser operation; however, depending on the required laser performance the inside surface is often given a reflective coating. The reflectivity of the inside face depends on the balance between the gain (eqn [14]), the output power and the power stability requirements. The outside face of the partial reflector must be anti-reflection (AR) coated.

Spontaneous emission occurs within the active gain medium and radiates randomly in all directions; a fraction of this spontaneous emission will be in the same direction as the optical axis, perpendicular to the end mirrors, and will also fall into a resonant mode of the optical resonator. Spontaneous emission photons interact with CO_2 molecules in the excited upper laser level, excited state (00^01), which stimulates these molecules to give up a quanta of vibrational energy as photons via the radiative transition $(00^01) \rightarrow (10^00)$, $\lambda = 10.6\mu m$. The radiation given up will have exactly the same phase and direction as the stimulating radiation and thus will be coherent with it. The reverse process of absorption also occurs, but so long as there is a population inversion there will be a net positive output. This process is called light amplification by stimulated emission of radiation (LASER). The mirrors continue to redirect the photons parallel to the optical axis and so long as the population inversion is not depleted, more and more photons are stimulated by stimulated emission which dominates the process and also dominates the spontaneous emission, which is important to initiate laser action.

Light emitted by lasers contains several optical frequencies, which are a function of the different modes of the optical resonator; these are simply the standing wave patterns that can exist within the resonator structure. There are two types of

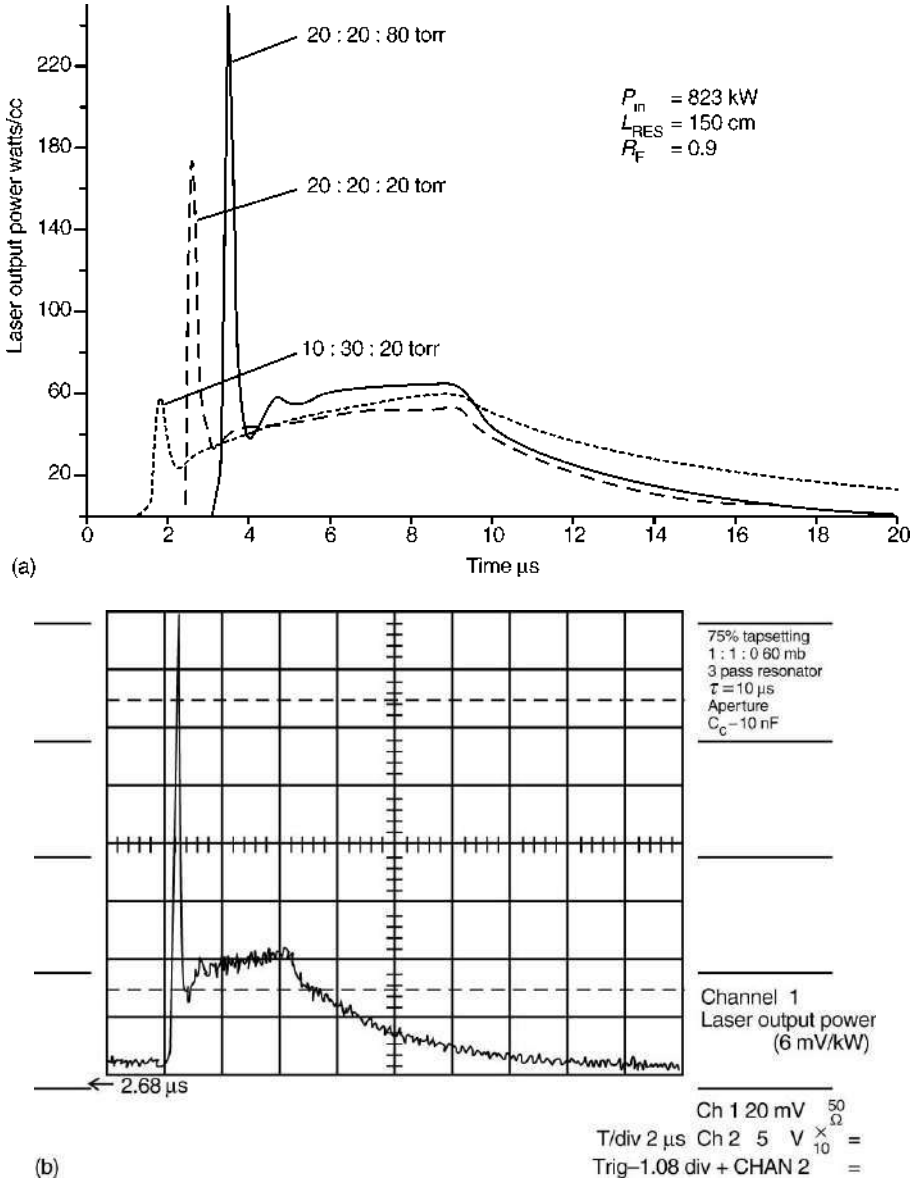

Figure 3 (a) Predicted output pulses for transverse flow CO_2 laser for different gas mixtures, (b) Experimental output pulse from transverse flow CO_2 laser.

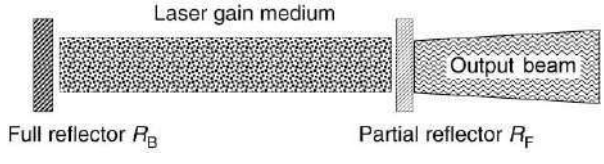

Figure 4 Optical resonator.

resonator modes: longitudinal and transverse. Longitudinal modes differ from one another in their frequency of oscillation whereas transverse modes differ from one another in their oscillation frequency and field distribution in a plane orthogonal to the direction of propagation. Typically CO_2 lasers have a large number of longitudinal modes; in CO_2 laser applications these are of less interest than the transverse modes, which determine the transverse beam intensity and the nature of the beam when focused. In cylindrical coordinates the transverse modes are labelled TEM_{pl}, where subscript 'p' is the number of radial nodes and 'l' is the number of angular nodes. The lowest order mode is the TEM_{00}, which has a Gaussian-like intensity profile with its maximum on the beam axis. A light beam emitted from an optical resonator with a Gaussian profile is said to be operating in the 'fundamental mode' or the TEM_{00} mode. The decrease in irradiance (I) with distance 'r' from the axis (I_0) of the Gaussian beam is

described by eqn [22]:

$$I(r) = I_0 \exp(2r^2/w^2) \qquad [22]$$

where w is the radial distance, where the power density is decreased to $1/e^2$ of its axial value. Ideally a commercial laser should be capable of operation in the fundamental mode as, with few exceptions, this results in the best performance in applications. Laser cutting benefits from operation in the fundamental mode; however, welding or heat treatment applications may benefit from operation with higher-order modes. Output beams are usually controlled to be linearly or circularly polarized, depending upon the requirements of the application. For materials processing applications the laser beam is usually focused via a water cooled ZnSe lens or, for very high power lasers, a parabolic gold coated mirror. Welding applications will generally use a long focal length lens and cutting applications will use a short focal length, which generates a higher irradiance at the work piece than that necessary for welding. The beam delivery optics are usually incorporated into a nozzle assembly that can deliver cooling water and assist gases for cutting and anti-oxidizing shroud gases for welding or surface engineering applications.

Laser Configuration

CO_2 lasers are available in many different configurations and tend to be classified on the basis of their physical form and the gas flow arrangement, both of which greatly affect the output power available and beam quality. The main categories are: sealed off lasers, waveguide lasers, slow axial flow, fast axial flow, diffusion cooled, transverse flow, transversely excited atmospheric lasers, and gas dynamic lasers.

Sealed-Off and Waveguide Lasers

Depopulation of the lower laser level is via collision with the walls of the discharge tube, so the attainable output power scales with the length of the discharge column and not its diameter. Output powers are in the range 5 W to 250 W. Devices may be constructed from concentric glass tubes with the inner tube providing the discharge cavity and the outer tube acting to contain water-cooling of the inner discharge tube. The inner tube walls act as a heat sink for the discharge thermal energy (see Figure 5). The DC electrical discharge is provided between a cathode and anode situated at either end of the discharge tube. A catalyst must be provided to ensure regeneration of CO_2 from CO. This may be accomplished by adding about 1% of H_2O to the gas mixture, or alternatively,

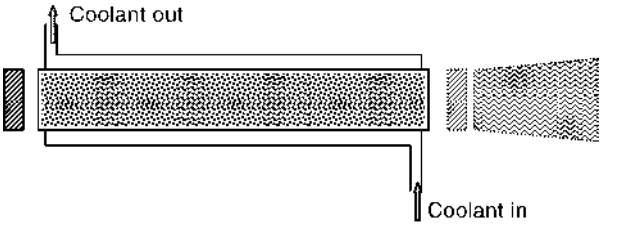

Figure 5 Schematic of sealed-off CO_2 laser, approximately 100 W per meter of gain length, gas cooled by diffusion to the wall.

recombination can be achieved via a hot (300 °C) Ni cathode, which acts as a catalyst. RF-excited all metal sealed-off tube systems can deliver lifetimes greater than 45 000 hours.

Diffusion-cooled slab laser technology will also deliver reliable sealed operation for 20 000 hrs. Excitation of the laser medium occurs via RF excitation between two water-cooled electrodes. The water-cooled electrodes dissipate (diffusion cooled) the heat generated in the gas discharge. An unstable optical resonator provides the output coupling for such a device (see Figure 6). Output powers are in the range 5 W to 300 W and can be pulsed from 0 to 100 kHz. These lasers are widely used for marking, rapid prototyping, and cutting of nonmetals (paper, glass, plastics, ceramics) and metals.

Waveguide CO_2 lasers use small bore tubes (2–4 mm) made of BeO or SiO_2 where the laser radiation is guided by the tube walls. Due to the small tube diameter, a gas total pressure of 100 to 200 Torr is necessary, hence the gain per unit length is high. This type of laser will deliver 30 W of output power from a relatively short (50 cm long) compact sealed-off design; such a system is useful for microsurgery and scientific applications. Excitation can be provided from a longitudinal DC discharge or from an RF source that is transverse to the optical axis; RF excitation avoids the requirement for an anode and cathode and results in a much lower electrode voltage.

Slow Axial Flow Lasers

In slow flow lasers the gas mixture flows slowly through the laser cavity. This is done to remove the products of dissociation that will reduce laser efficiency or prevent it from operating at all, and the main contaminant is CO. The dissociated gases (mainly CO and O_2) can be recombined using a catalyst pack and then reused via continuous recirculation. Heat is removed via diffusion through the walls of the tube containing the active gain medium. The tube is frequently made of Pyrex glass with a concentric outer tube to facilitate water-cooling of the

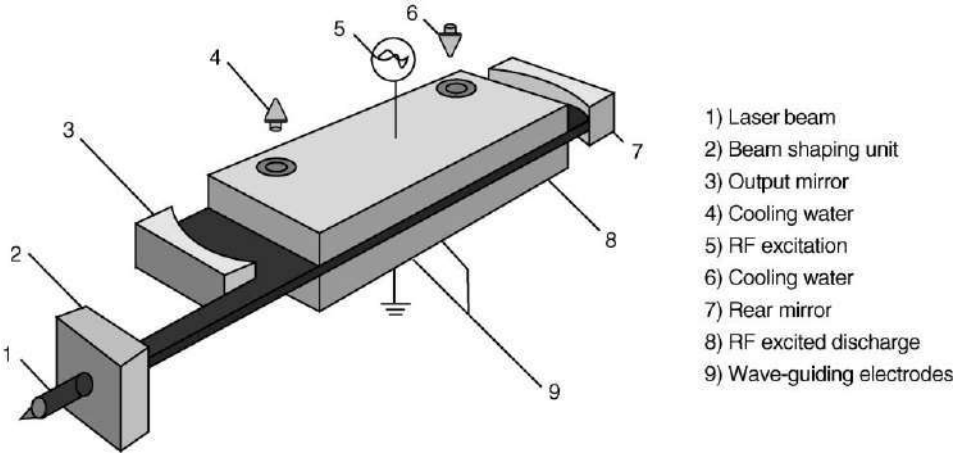

Figure 6 Schematic for sealed-off slab laser and diffusion cooled laser with RF excitation (courtesy of Rofin).

laser cavity (see Figure 7). Slow flow lasers operate in the power range 100 W to 1500 W, and tend to use a longitudinal DC electrical discharge which can be made to run continuously or pulsed if a thyratron switch is build into the power supply; alternatively, electrical power can be supplied via transverse RF excitation. The power scales with length, hence high power slow flow lasers have long cavities and require multiple cavity folds in order to reduce their physical size.

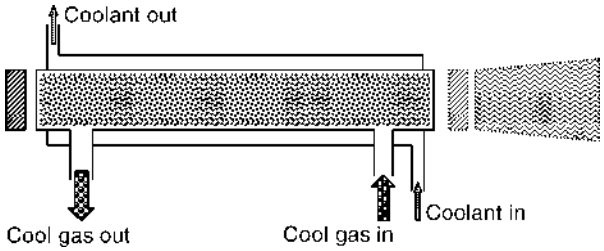

Figure 7 Slow flow CO_2 laser, approximately 100 W per meter of gain length, gas cooled by diffusion to the wall.

Fast Axial Flow Lasers

The fast axial flow laser, Figure 8, can provide output powers from 1 kW to 20 kW; it is this configuration that dominates the use of CO_2 lasers for industrial applications. Industrial lasers are usually in the power range 2–4 kW. The output power from these devices scales with mass flow, hence the gas mixture is recycled through the laser discharge region at sonic or supersonic velocities. Historically this was achieved using Rootes blowers to compress the gas upstream of the laser cavity. Rootes compressors are inherently inefficient and the more advanced laser systems utilize turbine compressors, which deliver greater efficiency and better laser stability. Rootes compressors can be a major source of vibration. With this arrangement heat exchangers are required to remove heat after the laser discharge region and also after the compressor stage, as the compression process heats up the laser gases. Catalyst packs are used to regenerate gases but some gas replacement is often required. These laser systems have short cavities and use folded stable resonator designs to achieve higher output powers with extremely high-quality beams that are particularly suitable for cutting applications. They also give excellent results when used for welding and surface treatments.

Fast axial flow lasers can be excited by a longitudinal DC discharge or a transverse RF discharge. Both types of electrical excitation are common. For materials processing applications it is often important to be able to run a laser in continuous wave (CW) mode or as a high pulse repetition rate (prf) pulsed laser and to be able to switch between CW and pulsed in real time; for instance, laser cutting of accurate internal corners is difficult using CW operation but very easy using the pulsed mode of operation. Both methods of discharge excitation can provide this facility.

Diffusion Cooled Laser

The diffusion-cooled slab laser is RF excited and gives an extremely compact design capable of delivering 4.5 kW pulsed from 8 Hz to 5 kHz prf or CW with good beam quality (see Figure 6). The optical resonator is formed by the front and rear mirrors and two parallel water cooled RF-electrodes. Diffusion cooling is provided by the RF-electrodes, removing the requirement for conventional gas recirculation via Rootes blowers or turbines.

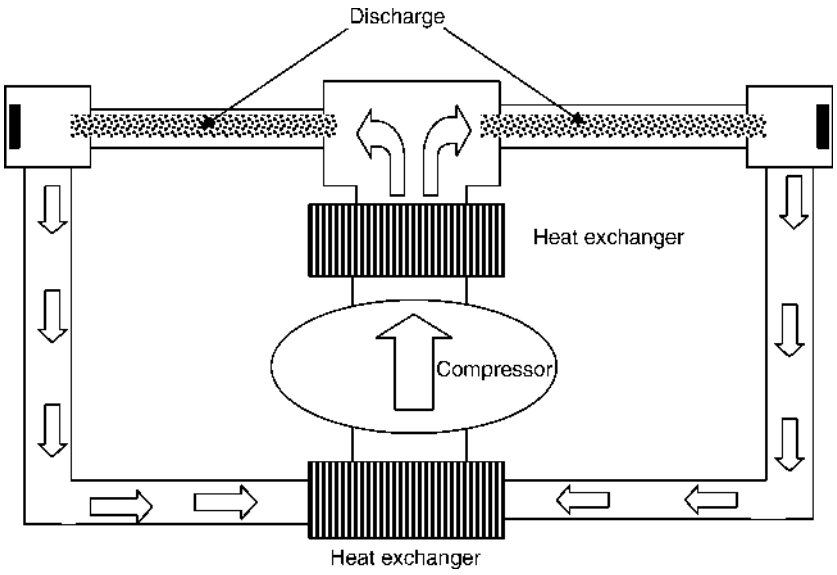

Figure 8 Fast axial flow carbon dioxide laser.

This design of laser results in a device with an extremely small footprint that has low maintenance and running costs. Applications include: cutting, welding, and surface engineering.

Fast Transverse Flow Laser

In the fast transverse flow laser (Figure 9a) the gas flow, electrical discharge, and the output beam are at right angles to each other (Figure 9b). The transverse discharge can be high voltage DC, RF, or pulsed up to 8 kHz (Figure 9c). Very high output power per unit discharge length can be obtained with an optimal total pressure (P) of ~100 Torr; systems are available delivering 10 kW of output power, CW or pulsed (see Figures 3a and b). The increase in total pressure requires a corresponding increase in the gas discharge electric field, E, as the ratio E/P must remain constant, since this determines the temperature of the discharge electrons, which have an optimum mean value (optimum energy distribution, Figure 2) for efficient excitation of the population inversion. With this high value of electric field, a longitudinal-discharge arrangement is impractical (500 kV for a 1 m discharge length); hence, the discharge is applied perpendicular to the optical axis. Fast transverse flow gas lasers provided the first multikilowatt outputs but tend to be expensive to maintain and operate. In order to obtain a reasonable beam quality, the output coupling is often obtained using a multipass unstable resonator. As the population inversion is available over a wide rectangular cross-section, this is a disadvantage of this arrangement and beam quality is not as good as that obtainable from fast axial flow designs. For this reason this type of laser is suitable for a wide range of welding and surface treatment applications.

Transversely Excited Atmospheric (TEA) Pressure

If the gas total pressure is increased above ~100 Torr it is difficult to sustain a stable glow discharge, because above this pressure instabilities degenerate into arcs within the discharge volume. This problem can be overcome by pulse excitation; using submicrosecond pulse duration, instabilities do not have sufficient time to develop; hence, the gas pressure can be increased above atmospheric pressure and the laser can be operated in a pulsed mode. In a mode locked format optical pulses shorter than 1 ns can be produced. This is called a TEA laser and with a transverse gas flow is capable of producing short high power pulses up to a few kHz repetition frequency. In order to prevent arc formation, TEA lasers usually employ ultraviolet or e-beam preionization of the gas discharge just prior to the main current pulse being applied via a thyratron switch. Output coupling is usually via an unstable resonator. TEA lasers are used for marking, remote sensing, range-finding, and scientific applications.

Conclusions

It is 40 years since Patel operated the first high power CO_2 laser. This led to the first generation of lasers which were quickly exploited for industrial laser materials processing, medical applications, defense,

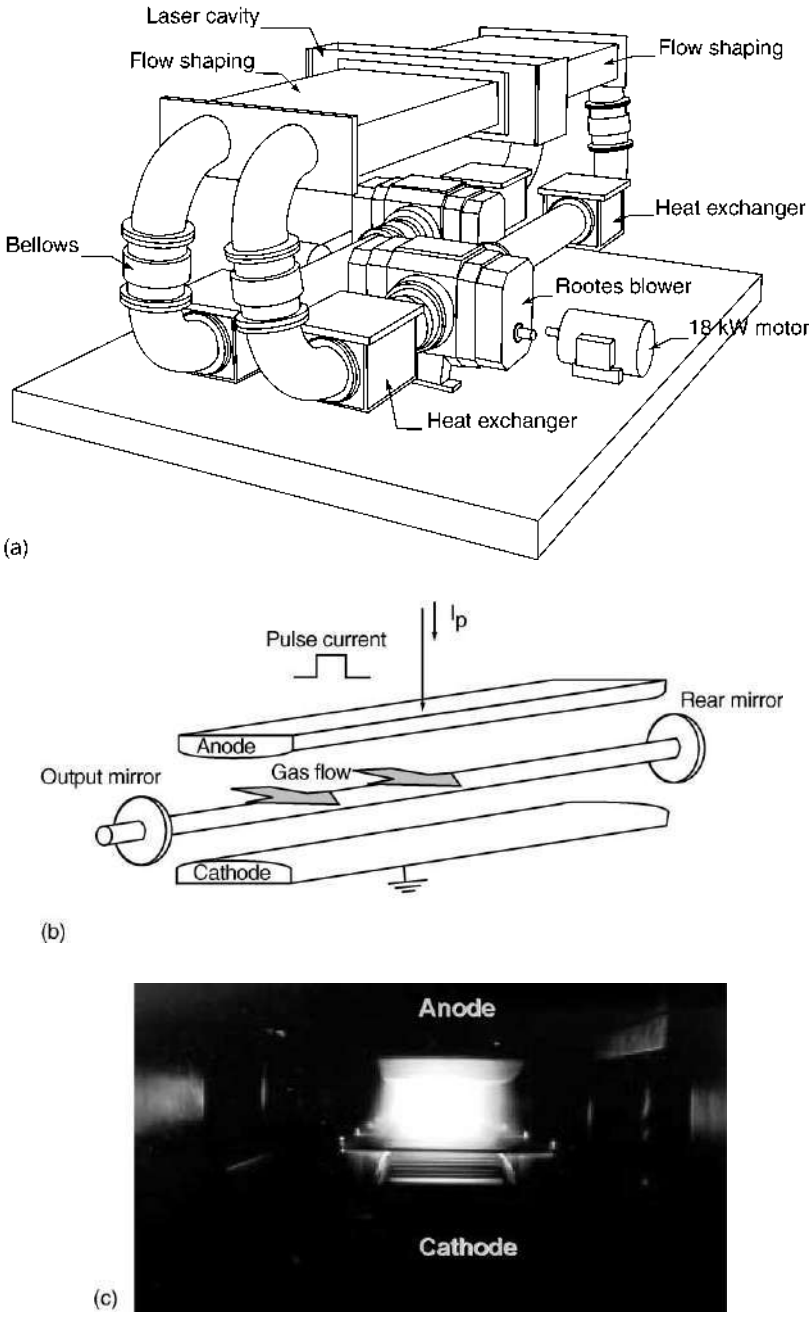

Figure 9 (a) Transverse flow carbon dioxide laser gas recirculator, (b) Transverse flow carbon dioxide electrodes, (c) Transverse flow carbon dioxide gas discharge as seen from the output window.

and scientific research applications; however, the first generation of lasers were quite unreliable and temperamental. After many design iterations, the CO_2 laser has now matured into a reliable, stable laser source available in many different geometries and power ranges. The low cost of ownership of the latest generation of CO_2 laser makes them a very attractive commercial proposition for many industrial and scientific applications. Commercial lasers incorporate many novel design features that are beyond the scope of this article and are often peculiar to the particular laser manufacturer. This includes gas additives and catalysts that may be required to stabilize the gas discharge of a particular laser design; it is this optimization of the laser design that has produced such reliable and controllable low-cost performance from the CO_2 laser.

List of Units and Nomenclature

α	The direct excitation of carbon dioxide (CO_2) ground state molecules (sec^{-1})
β	The direct de-excitation of nitrogen (N_2) (sec^{-1})
γ	The direct excitation of nitrogen (N_2) ground state molecules (sec^{-1})
η	The direct de-excitation of carbon dioxide (CO_2) (sec^{-1})
λ	wavelength (m)
σ	absorption coefficient (cm^2)
τ	Electrical current pulse length (μs)
τ_{sp}	Spontaneous emission life time of the upper laser level (sec)
$A = (\tau_{sp})^{-1} = 0.213(sec^{-1})$	The Einstein 'A' coefficient for the laser transition
c	velocity of light (cm sec^{-1})
C_c	Coupling capacitance (nF)
CO_2	Carbon dioxide
e	subscript 'e' refers to the fact that the populations in the square brackets are the values for thermodynamic equilibrium
E	Electric Field (V cm^{-1})
F_{CO_2}	Fraction of the input power (IP) coupled into the excitation of the energy level E_{00^01}
F_{N_2}	Fraction of the input power (IP) coupled into the excitation of the energy level $E_{v=1}$
g	gain (cm^{-1})
g_1 and g_2	energy level degeneracy's of levels 1 and 2
He	Helium
I	beam irradiance (W cm^{-2})
I_0	beam irradiance at the center of a Gaussian laser beam (W cm^{-2})
I_p	Photon population density (photons cm^{-3})
I_p	Input current (A)
IP	Electrical input power (W cm^{-3})
J	the rotational quantum number
K_{51}, K_{15}	Resonant energy transfer between the CO_2 (00^01) and $N_2(v = 2)$ energy levels proceeds via excited molecules colliding with ground state molecules (sec^{-1})
K_{132}, K_{2131}, K_{2231}	are vibration/vibration transfer rates (sec^{-1}) between energy levels 1 and 32; 21 and 31; 22 and 31, respectively (see Figure 1)
K_{320}	is a vibration/translation transfer rate (sec^{-1}) between energy levels 32 and 0 (see Figure 1)
K_{sp}, A	Spontaneous emission rate (sec^{-1})
L	The distance between the back and the front mirrors, which have reflectivity's of R_B and R_F (cm)
n	molecular population (molecules cm^{-3})
N_2	Nitrogen
P	Pressure (Torr)
P_{CO_2}, P_{He} and P_{N_2}	The respective gas partial pressures (Torr)
P_{in}	Electrical input power (kW)
r	radius of laser beam (cm)
R_B	Back mirror reflectivity
R_F	Front mirror reflectivity
t	time (sec)
T	Temperature (deg K)
T_0	the photon decay time (sec)
w	the radial distance where the power density is decreased to $1/e^2$ of its axial value

See also

Fiber and Guided Wave Optics: Overview. **Lasers:** Noble Gas Ion Lasers.

Further Reading

Anderson JD (1976) *Gasdynamic Lasers: An Introduction*. New York: Academic Press.

Chatwin CR, McDonald DW and Scott BF (1991) Design of a High P.R.F. Carbon Dioxide Laser for Processing High Damage Threshold Materials. Selected Papers on Laser Design, SPIE Milestone Series, pp. 425–433. Washington: SPIE Optical Engineering Press.

Cool AC (1969) Power and gain characteristics of high speed flow lasers. *Journal of Applied Physics* 40(9): 3563.

Crafer RC, Gibson AF, Kent MJ and Kimmit MF (1969) Time-dependent processes on CO_2 laser amplifiers. *British Journal of Applied Physics* 2(2): 183.

Gerry ET and Leonard AD (1966) Measurement of 10.6-μ CO_2 laser transition probability and optical broadening cross sections. *Applied Physics Letters* 8(9): 227.

Gondhalekar A, Heckenberg NR and Holzhauer E (1975) The mechanism of single-frequency operation of the hybrid CO_2 laser. *IEEE Journal of Quantum Electronics* QE-11(3): 103.

Gordiets BF, Sobolev NN and Shelepin LA (1968) Kinetics of physical processes in CO_2 lasers. *Soviet Physics JETP* 26(5): 1039.

Herzberg G (1945) *Molecular Spectra & Molecular Structure*. Infra-red and Raman Spectra of Polyatomic Molecules, vol. 2. New York: Van Nostrand.

Hoffman AL and Vlases GC (1972) A simplified model for predicting gain, saturation and pulse length for gas dynamic lasers. *IEEE Journal of Quantum Electronics* 8(2): 46.
Johnson DC (1971) Excitation of an atmospheric pressure CO_2–N_2–He laser by capacitor discharges. *IEEE Journal of Quantum Electronics* Q.E.-7(5): 185.
Koechner W (1988) *Solid State Laser Engineering*. Berlin: Springer Verlag.
Kogelnik H and Li T (1966) Laser beams and resonators. *Applied Optics* 5: 1550–1567.
Levine AK and De Maria AJ (1971). *Lasers*, vol. 3. New York: Marcel Dekkar, Chapter 3.
Moeller G and Ridgen JD (1965) *Applied Physics Letters* 7: 274.
Moore CB, Wood RE, Bei-Lok Hu and Yardley JT (1967) Vibrational energy transfer in CO_2 lasers. *Journal of Chemical Physics* 11: 4222.
Patel CKN (1964) *Physics Review Letters* 12: 588.
Siegman A (1986) *Lasers*. Mill Valley, California: University Science.
Smith K and Thomson RM (1978) *Computer Modeling of Gas Lasers*. New York and London: Plenum Press.
Sobolev NN and Sokovikov VV (1967) CO_2 lasers. *Soviet Physics USPEKHI* 10(2): 153.
Svelto O (1998) *Principles of Lasers*, 4th edn. New York: Plenum.
Tychinskii VP (1967) Powerful lasers. *Soviet Physics USPEKHI* 10(2): 131.
Vlases GC and Money WM (1972) Numerical modelling of pulsed electric CO_2 lasers. *Journal of Applied Physics* 43(4): 1840.
Wagner WG, Haus HA and Gustafson KT (1968) High rate optical amplification. *IEEE Journal of Quantum Electronics* Q.E.-4: 287.
Witteman W (1987) *The CO_2 Laser*. Springer Series in Optical Sciences, vol. 53.

Dye Lasers

F J Duarte, Eastman Kodak Company, New York, NY, USA

A Costela, Consejo Superior de Investigaciones Cientificas, Madrid, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Background

Dye lasers are the original tunable lasers. Discovered in the mid-1960s these tunable sources of coherent radiation span the electromagnetic spectrum from the near-ultraviolet to the near-infrared (Figure 1). Dye lasers spearheaded and sustained the revolution in atomic and molecular spectroscopy and have found use in many and diverse fields from medical to military applications. In addition to their extraordinary spectral versatility, dye lasers have been shown to oscillate from the femtosecond pulse domain to the continuous wave (cw) regime. For microsecond pulse emission, energies of up to hundreds of joules per pulse have been demonstrated. Further, operation at high pulsed repetition frequencies (prfs), in the multi-kHz regime, has provided average powers at kW levels. This unrivaled operational versatility is summarized in Table 1.

Dye lasers are excited by coherent optical energy from an excitation, or pump, laser or by optical energy from specially designed lamps called flashlamps. Recent advances in semiconductor laser

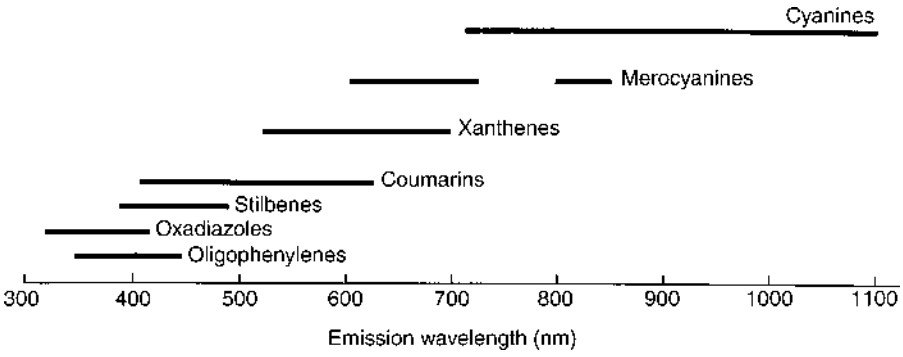

Figure 1 Approximate wavelength span from the various classes of laser dye molecules. Reproduced with permission from Duarte FJ (1995) *Tunable Laser Handbook*. New York: Academic Press.

Table 1 Emission characteristics of liquid dye lasers

Dye laser class	*Spectral coverage*[a]	*Energy per pulse*[b]	*Prf*[b]	*Power*[b]
Laser-pumped pulsed dye lasers	350–1100 nm	800 J[c]	13 kHz[d]	2.5 kW[d]
Flashlamp-pumped dye lasers	320–900 nm	400 J[e]	850 Hz[f]	1.2 kW[f]
CW dye lasers	370–1000 nm			43 W[g]

[a]Approximate range.
[b]Refers to maximum values for that particular emission parameter.
[c]Achieved with an excimer-laser pumped coumarin dye laser by Tang and colleagues, in 1987.
[d]Achieved with a multistage copper-vapor-laser pumped dye laser using rhodamine dye by Bass and colleagues, in 1992.
[e]Reported by Baltakov and colleagues, in 1974. Uses rhodamine 6G dye.
[f]Reported by Morton and Dragoo, in 1981. Uses coumarin 504 dye.
[g]Achieved with an Ar^+ laser pumped folded cavity dye laser using rhodamine 6G dye by Baving and colleagues, in 1982.

technology have made it possible to construct very compact all-solid-state excitation sources that, coupled with new solid-state dye laser materials, should bring the opportunity to build compact tunable laser systems for the visible spectrum. Further, direct diode-laser pumping of solid-state dye lasers should prove even more advantageous to enable the development of fairly inexpensive tunable narrow-linewidth solid-state dye laser systems for spectroscopy and other applications requiring low powers. Work on electrically excited organic gain media might also provide new avenues for further progress.

The literature of dye lasers is very rich and many review articles have been written describing and discussing traditional dye lasers utilizing liquid gain media. In particular, the books *Dye Lasers*, *Dye Laser Principles*, *High Power Dye Lasers*, and *Tunable Lasers Handbook* provide excellent sources of authoritative and detailed description of the physics and technology involved. In this article we offer only a survey of the operational capabilities of the dye lasers using liquid gain media in order to examine with more attention the field of solid-state dye lasers.

Brief History of Dye Lasers

1965: Quantum theory of dyes is discussed in the context of the maser (R. P. Feynman).
1966: Dye lasers are discovered (P. P. Sorokin and J. R. Lankard; F. P. Schäfer and colleagues).
1967: The flashlamp-pumped dye laser is discovered (P. P. Sorokin and J. R. Lankard; W. Schmidt and F. P. Schäfer).
1967–1968: Solid-state dye lasers are discovered (B. H. Soffer and B. B. McFarland; O. G. Peterson and B. B. Snavely).
1968: Mode-locking, using saturable absorbers, is demonstrated in dye lasers (W. Schmidt and F. P. Schäfer).
1970: The continuous-wave (cw) dye laser is discovered (O. G. Peterson, S. A. Tuccio, and B. B. Snavely).
1971: The distributed feedback dye laser is discovered (H. Kogelnik and C. V. Shank).
1971–1975: Prismatic beam expansion in dye lasers is introduced (S. A. Myers; E. D. Stokes and colleagues; D. C. Hanna and colleagues).
1972: Passive mode-locking is demonstrated in cw dye lasers (E. P. Ippen, C. V. Shank, and A. Dienes).
1972: The first pulsed narrow-linewidth tunable dye laser is introduced (T. W. Hänsch).
1973: Frequency stabilization of cw dye lasers is demonstrated (R. L. Barger, M. S. Sorem, and J. L. Hall).
1976: Colliding-pulse-mode locking is introduced (I. S. Ruddock and D. J. Bradley).
1977–1978: Grazing-incidence grating cavities are introduced (I. Shoshan and colleagues; M. G. Littman and H. J. Metcalf; S. Saikan).
1978–1980: Multiple-prism grating cavities are introduced (T. Kasuya and colleagues; G. Klauminzer; F. J. Duarte and J. A. Piper).
1981: Prism pre-expanded grazing-incidence grating oscillators are introduced (F. J. Duarte and J. A. Piper).
1982: Generalized multiple-prism dispersion theory is introduced (F. J. Duarte and J. A. Piper).
1983: Prismatic negative dispersion for pulse compression is introduced (W. Dietel, J. J. Fontaine, and J-C. Diels).
1987: Laser pulses as short as six femtoseconds are demonstrated (R. L. Fork, C. H. Brito Cruz, P. C. Becker, and C. V. Shank).
1994: First narrow-linewidth solid-state dye laser oscillator (F. J. Duarte).
1999–2000: Distributed feedback solid-state dye lasers are introduced (Wadsworth and colleagues; Zhu and colleagues).

Molecular Energy Levels

Dye molecules have large molecular weights and contain extended systems of conjugated double bonds. These molecules can be dissolved in an adequate organic solvent (such as ethanol, methanol, ethanol/water, and methanol/water) or incorporated into a solid matrix (organic, inorganic, or hybrid). These molecular gain media have a strong absorption generally in the visible and ultraviolet regions, and exhibit large fluorescence bandwidths covering the entire visible spectrum. The general energy level diagram of an organic dye is shown in Figure 2. It consists of electronic singlet and triplet states with each electronic state containing a multitude of overlapping vibrational–rotational levels giving rise to broad continuous energy bands. Absorption of visible or ultraviolet pump light excites the molecules from the ground state S_0 into some rotational–vibrational level belonging to an upper excited singlet state, from where the molecules decay nonradiatively to the lowest vibrational level of the first excited singlet state S_1 on a picosecond time-scale. From S_1 the molecules can decay radiatively, with a radiative lifetime on the nanosecond time-scale, to a higher-lying vibrational–rotational level of S_0. From this level they rapidly thermalize into the lowest vibrational–rotational levels of S_0. Alternatively, from S_1, the molecules can experience nonradiative relaxation either to the triplet state T_1 by an intersystem crossing process or to the ground state by an internal conversion process. If the intensity of the pumping radiation is high enough a population inversion between S_1 and S_0 may be attained and stimulated emission occurs. Internal conversion and intersystem crossing compete with the fluorescence decay mode of the molecule and therefore reduce the efficiency of the laser emission. The rate for internal conversion to the electronic ground state is usually negligibly small so that the most important loss process is intersystem crossing into T_1 that populates the lower metastable triplet state. Thus, absorption on the triplet–triplet allowed transitions could cause considerable losses if these absorption bands overlap the lasing band, inhibiting or even halting the lasing process. This triplet loss can be reduced by adding small quantities of appropriate chemicals that favor nonradiative transitions that shorten the effective lifetime of the T_1 level. For pulsed excitation with nanosecond pulses, the triplet–triplet absorption can be neglected because for a typical dye the intersystem crossing rate is not fast enough to build up an appreciable triplet population in the nanosecond time domain.

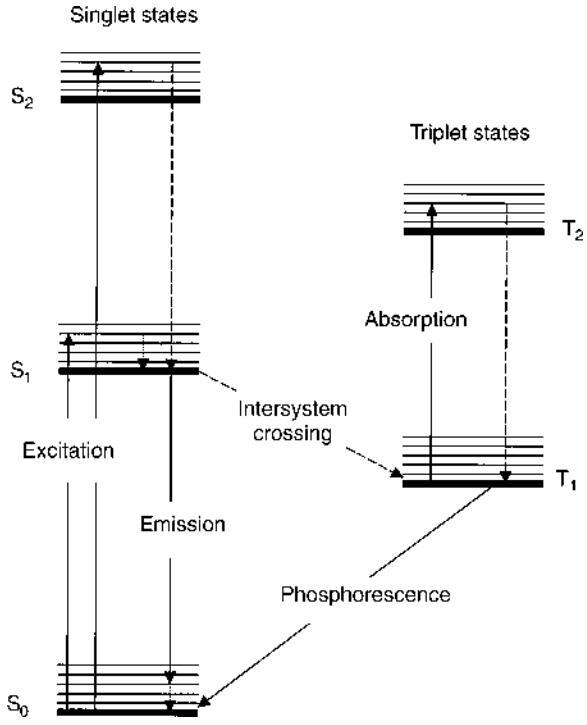

Figure 2 Schematic energy level diagram for a dye molecule. Full lines: radiative transitions; dashed lines: nonradiative transitions; dotted lines: vibrational relaxation.

Dye molecules are large (a typical molecule incorporates 50 or more atoms) and are grouped into families with similar chemical structures. A survey of the major classes of laser dyes is given later. Solid-state laser dye gain media are also considered later.

Liquid Dye Lasers

Laser-Pumped Pulsed Dye Lasers

Laser-pumped dye lasers use a shorter wavelength, or higher frequency, pulsed laser as the excitation or pump source. Typical pump lasers for dye lasers are gas lasers such as the excimer, nitrogen, or copper lasers. One of the most widely used solid-state laser pumps is the frequency doubled Nd:YAG laser which emits at 532 nm.

In a laser-pumped pulsed dye laser the active medium, or dye solution, is contained in an optical cell often made of quartz or fused silica, which provides an active region typically some 10 mm in length and a few mm in width. The active medium is then excited either longitudinally, or transversely, via a focusing lens using the pump laser. In the case of transverse excitation the pump laser is focused to a beam ~10 mm in width and ~0.1 mm in height. Longitudinal pumping requires focusing of the excitation beam to a diameter in the 0.1–0.15 mm range. For lasers operated at low prfs (a few pulses per

second), the dye solution might be static. However, for high-prf operation (a few thousand pulses per second) the dye solution must be flowed at speeds of up to a few meters per second in order to dissipate the heat. A simple broadband optically pumped dye laser can be constructed using just the pump laser, the active medium, and two mirrors to form a resonator. In order to achieve tunable, narrow-linewidth, emission, a more sophisticated resonator must be employed. This is called a dispersive tunable oscillator and is depicted in Figure 3. In a dispersive tunable oscillator the exit side of the cavity is comprised of a partial reflector, or an output coupler, and the other end of the resonator is composed of a multiple-prism grating assembly. It is the dispersive characteristics of this multiple-prism grating assembly and the dimensions of the emission beam produced at the gain medium that determine the tunability and the narrowness, or spectral purity, of the laser emission.

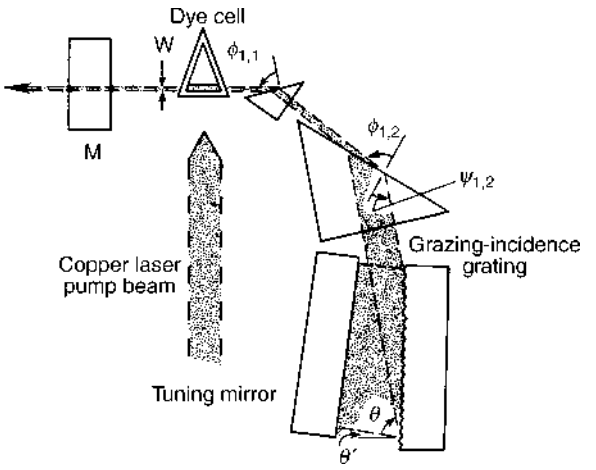

Figure 3 Copper-vapor-laser pumped hybrid multiple-prism near grazing incidence (HMPGI) grating dye laser oscillator. Adapted with permission from Duarte FJ and Piper JA (1984) Narrow linewidth high prf copper laser-pumped dye-laser oscillators. *Applied Optics* 23: 1391–1394.

In order to selectively excite a single vibrational–rotational level of a molecule such as iodine (I_2), at room temperature, one needs a laser linewidth of $\Delta\nu \approx 1.5$ GHz (or $\Delta\lambda \approx 0.0017$ nm at $\lambda = 590$ nm). The hybrid multiple-prism near-grazing-incidence (HMPGI) grating dye laser oscillator illustrated in Figure 4 yields laser linewidths in the 400 MHz $\leq \Delta\nu \leq$ 650 MHz range at 4–5% conversion efficiencies whilst excited by a copper-vapor laser operating at a prf of 10 kHz. Pulse lengths are ~10 ns at full-width half-maximum (FWHM). The narrow-linewidth emission from these oscillators is said to be single-longitudinal-mode lasing because only one electromagnetic mode is allowed to oscillate.

The emission from oscillators of this class can be amplified many times by propagating the tunable narrow-linewidth laser beam through single-pass amplifier dye cells under the excitation of pump lasers. Such amplified laser emission can reach enormous average powers. Indeed, a copper-vapor-laser pumped dye laser system at the Lawrence Livermore National Laboratory (USA), designed for the laser isotope separation program, was reported to

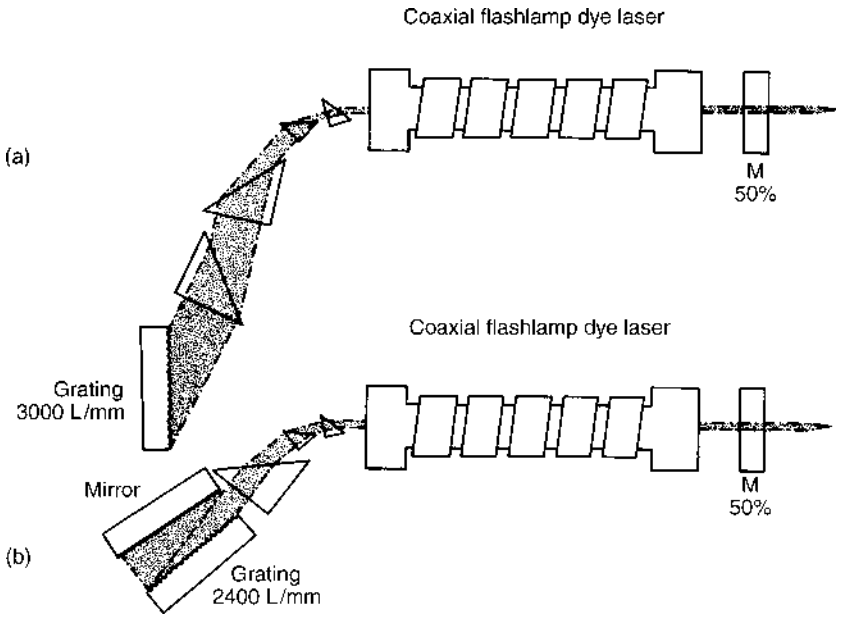

Figure 4 Flashlamp-pumped multiple-prism grating oscillators. From Duarte FJ, Davenport WE, Ehrlich JJ and Taylor TS (1991) Ruggedized narrow-linewidth dispersive dye laser oscillator. *Optics Communications* 84: 310–316. Reproduced with permission from Elsevier.

yield average powers in excess of 2.5 kW, at a prf of 13.2 kHz, at a better than 50% conversion efficiency as reported by Bass and colleagues in 1992. Besides high conversion efficiencies, excitation with copper-vapor lasers, at $\lambda = 510.554$ nm, has the advantage of inducing little photodegradation in the active medium thus allowing very long dye lifetimes.

Flashlamp-Pumped Dye Lasers

Flashlamps utilized in dye laser excitation emit at black-body temperatures in the 20 000 K range, thus yielding intense ultraviolet radiation centered around 200 nm. One further requirement for flashlamps, and their excitation circuits, is to deliver light pulses with a fast rise time. For some flashlamps this rise time can be less than a few nanoseconds.

Flashlamp-pumped dye lasers differ from laser-pumped pulsed dye lasers mainly in the pulse energies and pulse lengths attainable. This means that flashlamp-pumped dye lasers, using relatively large volumes of dye, can yield very large energy pulses. Excitation geometries use either coaxial lamps, with the dye flowing in a quartz cylinder at the center of the lamp, or two or more linear lamps arranged symmetrically around the quartz tube containing the dye solution. Using a relatively weak dye solution of rhodamine-6G (2.2×10^{-5} M), a coaxial lamp, and an active region defined by a quartz tube 6 cm in diameter and a length of 60 cm, Baltakov and colleagues, in 1974, reported energies of 400 J in pulses 25 μs long at FWHM.

Flashlamp-pumped tunable narrow-linewidth dye laser oscillators described by Duarte and colleagues, in1991, employ a cylindrical active region 6 mm in diameter and 17 cm in length. The dye solution is made of rhodamine 590 at a concentration of 1×10^{-5} M. This active region is excited by a coaxial flashlamp. Using multiple-prism grating architectures (see Figure 4) these authors achieve a diffraction limited TEM_{00} laser beam and laser linewidths of $\Delta\nu \approx 300$ MHz at pulsed energies in the 2–3 mJ range. The laser pulse duration is reported to be $\Delta t \approx 100$ ns. The laser emission from this class of multiple-prism grating oscillator is reported to be extremely stable. The tunable narrow-linewidth emission from these dispersive oscillators is either used directly in spectroscopic, or other scientific applications, or is utilized to inject large flashlamp-pumped dye laser amplifiers to obtain multi-joule pulse energies with the laser linewidth characteristics of the oscillator.

Continuous Wave Dye Lasers

CW dye lasers use dye flowing at linear speeds of up to 10 meters per second which are necessary to remove the excess heat and to quench the triplet states. In the original cavity reported by Peterson and colleagues, in 1970, a beam from an Ar^+ laser was focused on to an active region which is contained within the resonator. The resonator comprised dichroic mirrors that transmit the blue-green radiation of the pump laser and reflect the red emission from the dye molecules. Using a pump power of about 1 W, in a TEM_{00} laser beam, these authors reported a dye laser output of 30 mW. Subsequent designs replaced the dye cell with a dye jet, an introduced external mirror, and integrated dispersive elements in the cavity. Dispersive elements such as prisms and gratings are used to tune the wavelength output of the laser. Frequency-selective elements, such as etalons and other types of interferometers, are used to induce frequency narrowing of the tunable emission. Two typical cw dye laser cavity designs are described by Hollberg, in 1990, and are reproduced here in Figure 5. The first design is a linear three-mirror folded cavity. The second one is an eight-shaped ring dye laser cavity comprised of mirrors M_1, M_2, M_3, and M_4. Linear cavities exhibit the effect of *spatial hole burning* which allows the cavity to lase in more than one longitudinal mode. This problem can be overcome in ring cavities (Figure 5b) where the laser emission is in the form of a traveling wave.

Two aspects of cw dye lasers are worth emphasizing. One is the availability of relatively high powers in single longitudinal mode emission and the other is the

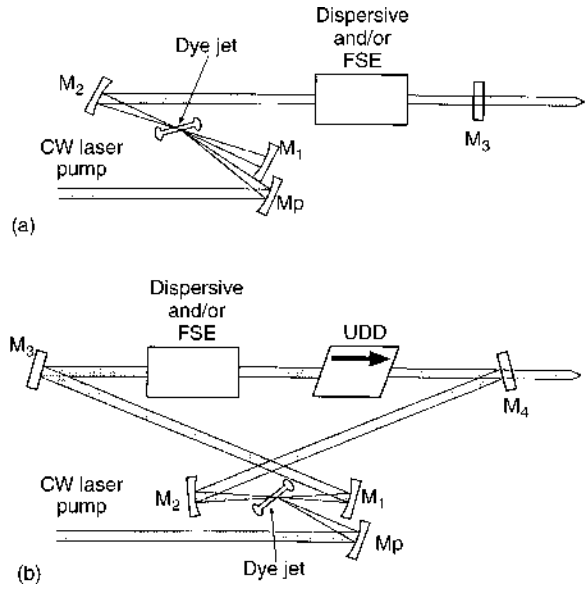

Figure 5 CW laser cavities: (a) linear cavity and (b) ring cavity. Adapted from Hollberg LW (1990) CW dye lasers. In: Duarte FJ and Hillman LW (eds) *Dye Laser Principles*, pp. 185–238. New York: Academic Press. Reproduced with permission from Elsevier.

demonstration of very stable laser oscillation. First, Johnston and colleagues, in 1982, reported 5.6 W of stabilized laser output in a single longitudinal mode at 593 nm at a conversion efficiency of 23%. In this work eleven dyes were used to span the spectrum continuously from ~400 nm to ~900 nm. In the area of laser stabilization and ultra narrow-linewidth oscillation it is worth mentioning the work of Hough and colleagues, in 1984, that achieved laser linewidths of less than 750 Hz employing an external reference cavity.

Ultrashort-Pulse Dye Lasers

Ultrashort-pulse, or femtosecond, dye lasers use the same type of technology as cw dye lasers configured to incorporate a saturable absorber region. One such configuration is the ring cavity depicted in Figure 6. In this cavity the gain region is established between mirrors M_1 and M_2 whilst the saturable absorber is deployed in a counter-propagating arrangement. This arrangement is necessary to establish a collision between two counter-propagating pulses at the saturable absorber thus yielding what is known as colliding-pulse mode locking (CPM) as reported by Ruddock and Bradley, in 1976. This has the effect of creating a transient grating, due to interference, at the absorber thus shortening the pulse. Intracavity prisms were incorporated in order to introduce negative dispersion, by Dietel and colleagues in 1983, and thus subtract dispersion from the cavity and ultimately provide the compensation needed to produce femtosecond pulses.

The shortest pulse obtained from a dye laser, 6 fs, has been reported by Fork and colleagues, in 1987, using extra-cavity compression. In that experiment, a dye laser incorporating CPM and prismatic compensation was used to generate pulses that were amplified by a copper-vapor laser at a prf of 8 kHz. The amplified pulses, of a duration of 50 fs, were then propagated through two grating pairs and a four-prism sequence for further compression.

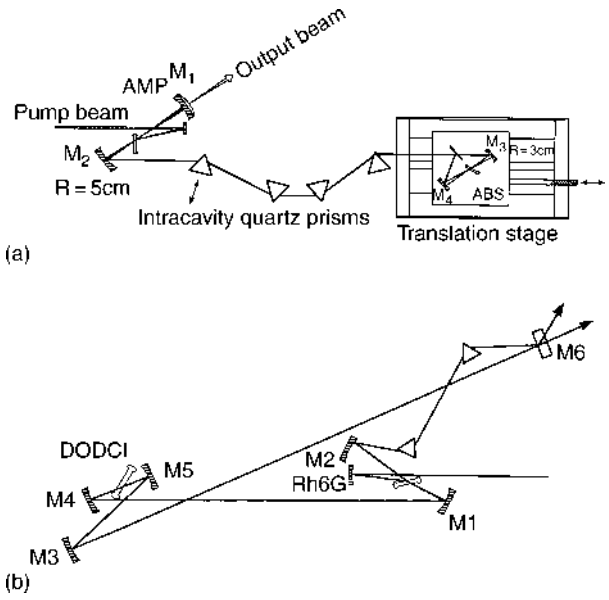

Figure 6 Femtosecond dye laser cavities: (a) linear femtosecond cavity and (b) ring femtosecond cavity. Adapted from Diels J-C (1990) Femtosecond dye lasers. In: Duarte FJ and Hillman LW (eds) *Dye Laser Principles*, pp. 41–132. New York: Academic Press. Reproduced with permission from Elsevier.

Solid-State Dye Laser Oscillators

In this section the principles of linewidth narrowing in dispersive resonators are outlined. Albeit the discussion focuses on multiple-prism grating solid-state dye laser oscillators, in particular, the physics is applicable to pulsed high-power dispersive tunable lasers in general.

Multiple-Prism Dispersion Grating Theory

The spectral linewidth in a dispersive optical system is given by

$$\Delta\lambda \approx \Delta\theta(\nabla_\lambda \theta)^{-1} \qquad [1]$$

where $\Delta\theta$ is the light beam divergence, $\nabla_\lambda = \partial/\partial\lambda$, and $\nabla_\lambda\theta$ is the overall dispersion of the optics. This identity can de derived either from the principles of geometrical optics or from the principles of generalized interferometry as described by Duarte, in 1992.

The cumulative single-pass generalized multiple-prism dispersion at the mth prism, of a multiple-prism array as illustrated in Figure 7, as given by Duarte and Piper, in 1982,

$$\nabla_\lambda \phi_{2,m} = H_{2,m}\nabla_\lambda n_m + (k_{1,m}k_{2,m})^{-1} \times \left(H_{1,m}\nabla_\lambda n_m \pm \nabla_\lambda \phi_{2,(m-1)}\right) \qquad [2]$$

In this equation

$$k_{1,m} = \cos\psi_{1,m}/\cos\phi_{1,m} \qquad [3a]$$

$$k_{2,m} = \cos\phi_{2,m}/\cos\psi_{2,m} \qquad [3b]$$

$$H_{1,m} = \tan\phi_{1,m}/n_m \qquad [4a]$$

$$H_{2,m} = \tan\phi_{2,m}/n_m \qquad [4b]$$

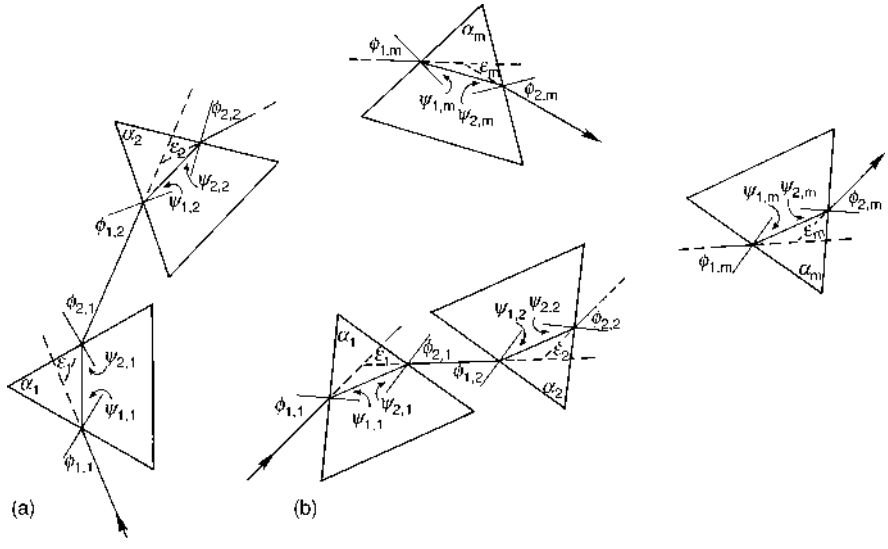

Figure 7 Generalized multiple-prism arrays in (a) additive and (b) compensating configurations. From Duarte FJ (1990) Narrow-linewidth pulsed dye laser oscillators. In: Duarte FJ and Hillman LW (eds) *Dye Laser Principles*, pp. 133–183. New York: Academic Press. Reproduced with permission from Elsevier.

Here, $k_{1,m}$ and $k_{2,m}$ represent the physical beam expansion experienced by the incident and the exit beams, respectively.

Equation [2] indicates that $\nabla_\lambda \phi_{2,m}$, the cumulative dispersion at the mth prism, is a function of the geometry of the mth prism, the position of the light beam relative to this prism, the refractive index of the prism, and the cumulative dispersion up to the previous prism $\nabla_\lambda \phi_{2,(m-1)}$.

For an array of r identical isosceles, or equilateral, prisms arranged symmetrically, in an additive configuration, so that the angles of incidence and emergence are the same, the cumulative dispersion reduces to

$$\nabla_\lambda \phi_{2,r} = r \nabla_\lambda \phi_{2,1} \qquad [5]$$

Under these circumstances the dispersions add in a simple and straightforward manner. For configurations incorporating right-angle prisms, the dispersions need to be handled mathematically in a more subtle form.

The generalized double-pass, or return-pass, dispersion for multiple-prism beam expanders was introduced by Duarte, in 1985:

$$\nabla_\lambda \Phi_P = 2M_1M_2 \sum_{m=1}^{r} (\pm 1) H_{1,m} \left(\prod_{j=m}^{r} k_{1,j} \prod_{j=m}^{r} k_{2,j} \right)^{-1} \nabla_\lambda n_m + 2 \sum_{m=1}^{r} (\pm 1) H_{2,m} \left(\prod_{j=1}^{m} k_{1,j} \prod_{j=1}^{m} k_{2,j} \right) \nabla_\lambda n_m \qquad [6]$$

Here, M_1 and M_2 are the beam magnification factors given by

$$M_1 = \prod_{m=1}^{r} k_{1,m} \qquad [7a]$$

$$M_2 = \prod_{m=1}^{r} k_{2,m} \qquad [7b]$$

For a multiple prism expander designed for an orthogonal beam exit, and Brewster's angle of incidence, eqn [6] reduces to the succinct expression given by Duarte, in 1990:

$$\nabla_\lambda \Phi_P = 2 \sum_{m=1}^{r} (\pm 1)(n_m)^{m-1} \nabla_\lambda n_m \qquad [8]$$

Equation [6] can be used to either quantify the overall dispersion of a given multiple-prism beam expander or to design a prismatic expander yielding zero dispersion, that is, $\nabla_\lambda \Phi_P = 0$, at a given wavelength.

Physics and Architecture of Solid-State Dye-Laser Oscillators

The first high-performance narrow-linewidth tunable laser was introduced by Hänsch in 1972. This laser yielded a linewidth of $\Delta\nu \approx 2.5$ GHz (or $\Delta\lambda \approx 0.003$ nm at $\lambda \approx 600$ nm) in the absence of an intracavity etalon. Hänsch demonstrated that the laser linewidth from a tunable laser was narrowed significantly when the beam incident on the tuning grating was expanded using an astronomical

telescope. The linewidth equation including the intracavity beam magnification factor can be written as

$$\Delta\lambda \approx \Delta\theta(M\nabla_\lambda\Theta_G)^{-1} \qquad [9]$$

From this equation, it can be deduced that a narrow $\Delta\lambda$ is achieved by reducing $\Delta\theta$ and increasing the overall intracavity dispersion ($M\nabla_\lambda\Theta_G$). The intracavity dispersion is optimized by expanding the size of the intracavity beam incident on the diffractive surface of the tuning grating until it is totally illuminated.

Hänsch used a two-dimensional astronomical telescope to expand the intracavity beam incident on the diffraction grating. A simpler beam expansion method consists in the use of a single-prism beam expander as disclosed by several authors (Myers, 1971; Stokes and colleagues, 1972; Hanna and colleagues, 1975). An extension and improvement of this approach was the introduction of multiple-prism beam expanders as reported by Kasuya and colleagues, in 1978, Klauminzer, in 1978, and Duarte and Piper, in 1980. The main advantages of multiple-prism beam expanders, over two-dimensional telescopes, are simplicity, compactness, and the fact that the beam expansion is reduced from two dimensions to one dimension. Physically, as explained previously, prismatic beam expanders also introduce a dispersion component that is absent in the case of the astronomical telescope. Advantages of multiple-prism beam expanders over single-prism beam expansion are higher transmission efficiency, lower amplified spontaneous emission levels, and the flexibility to either augment or reduce the prismatic dispersion.

In general, for a pulsed multiple-prism grating oscillator, Duarte and Piper, in 1984, showed that the return-pass dispersive linewidth is given by

$$\Delta\lambda = \Delta\theta_R(MR\nabla_\lambda\Theta_G + R\nabla_\lambda\Phi_P)^{-1} \qquad [10]$$

where R is the number of return-cavity passes. The grating dispersion in this equation, $\nabla_\lambda\Theta_G$, can be either from a grating in Littrow or near grazing-incidence configuration. The multiple-return-pass equation for the beam divergence was given by Duarte, in 2001:

$$\Delta\theta_R = (\lambda/\pi w)\left(1 + (L_R/B_R)^2 + (A_R L_R/B_R)^2\right)^{1/2} \qquad [11]$$

Here, $L_R = (\pi w^2/\lambda)$ is the Rayleigh length of the cavity, w is the beam waist at the gain region, while A_R and B_R are the corresponding multiple-return-pass elements derived from propagation matrices.

At present, very compact and optimized multiple-prism grating tunable laser oscillators are found in two basic cavity architectures. These are the multiple-prism Littrow (MPL) grating laser oscillator, reported by Duarte in 1999 and illustrated in Figure 8, and the hybrid multiple-prism near grazing-incidence (HMPGI) grating laser oscillator, introduced by Duarte in 1997 and depicted in Figure 9. In early MPL grating oscillators the individual prisms integrating the multiple-prism expander were deployed in an additive configuration thus adding the cumulative dispersion to that of the grating and thus contributing to the overall dispersion of the cavity. In subsequent architectures the prisms were deployed in compensating configurations so as to yield zero dispersion and thus allow the tuning characteristics of the cavity to be

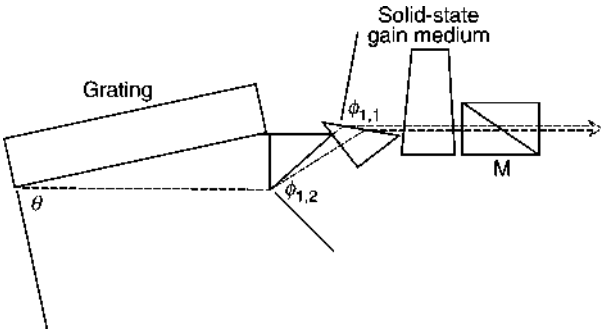

Figure 8 MPL grating solid-state dye laser oscillator: optimized architecture. The physical dimensions of the optical components of this cavity are shown to scale. The length, of the solid-state dye gain medium, along the optical axis, is 10 mm. Reproduced with permission from Duarte FJ (1999) Multiple-prism grating solid-state dye laser oscillator: optimized architecture. *Applied Optics* 38: 6347–6349.

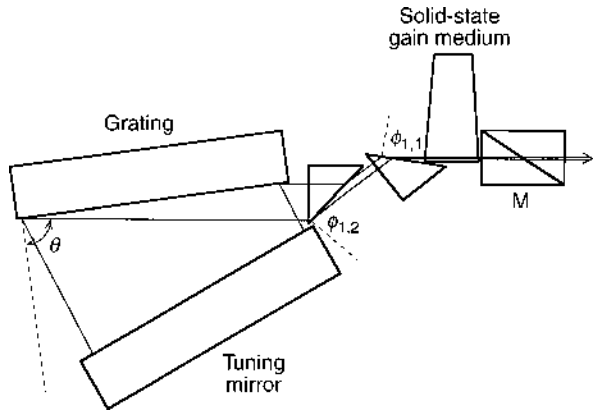

Figure 9 HMPGI grating solid-state dye laser. Schematics to scale. The length of the solid state dye gain medium, along the optical axis, is 10 mm. Reproduced with permission from Duarte FJ (1997) Multiple-prism near-grazing-incidence grating solid-state dye laser oscillator. *Optics and Laser Technology* 29: 513–516.

determined by the grating exclusively. In this approach the principal role of the multiple-prism array is to expand the beam incident on the grating thus augmenting significantly the overall dispersion of the cavity as described in eqns [9] or [10]. In this regard, it should be mentioned that beam magnification factors of up to 100, and beyond, have been reported in the literature. Using solid-state laser dye gain media, these MPL and HMPGI grating laser oscillators deliver tunable single-longitudinal-mode emission at laser linewidths 350 MHz $\leq \Delta\nu \leq$ 375 MHz and pulse lengths in the 3–7 ns (FWHM) range. Long-pulse operation of this class of multiple-prism grating oscillators has been reported by Duarte and colleagues, in 1998. In these experiments laser linewidths of $\Delta\nu \approx$ 650 MHz were achieved at pulse lengths in excess of 100 ns (FWHM) under flashlamp-pumped dye laser excitation.

The dispersive cavity architectures described here have been used with a variety of laser gain media in the gas, the liquid, and the solid state. Applications to tunable semiconductor lasers have also been reported by Zorabedian, in 1992, Duarte, in 1993, and Fox and colleagues, in 1997.

Concepts important to MPL and HMPGI grating tunable laser oscillators include the emission of a single-transverse-mode (TEM_{00}) laser beam in a compact cavity, the use of multiple-prism arrays, the expansion of the intracavity beam incident on the grating, the control of the intracavity dispersion, and the quantification of the overall dispersion of the multiple-prism grating assembly via generalized dispersion equations. Sufficiently high intracavity dispersion leads to the achievement of return-pass dispersive linewidths close to the free-spectral range of the cavity. Under these circumstances single-longitudinal-mode lasing is readily achieved as a result of multipass effects.

A Note on the Cavity Linewidth Equation

So far we have described how the cavity linewidth equation

$$\Delta\lambda \approx \Delta\theta(\nabla_\lambda \theta)^{-1}$$

and the dispersion equations can be applied to achieve highly coherent, or very narrow-linewidth, emission. It should be noted that the same physics can be applied to achieve ultrashort, or femtosecond, laser pulses. It turns out that intracavity prisms can be configured to yield negative dispersion as described by Duarte and Piper, in 1982, Dietel and colleagues, in 1983, and Fork and colleagues in 1984. This negative dispersion can reduce significantly the overall dispersion of the cavity. Under those circumstances, eqn [1] predicts broadband emission which according to the uncertainty principle, in the form of,

$$\Delta\nu\Delta t \approx 1 \qquad [12]$$

can lead to very short pulse emission since $\Delta\nu$ and $\Delta\lambda$ are related by the identities

$$\Delta\lambda \approx \lambda^2/\Delta x \qquad [13]$$

and

$$\Delta\nu \approx c/\Delta x \qquad [14]$$

Distributed-Feedback Solid-State Dye Lasers

Recently, narrow-linewidth laser emission from solid-state dye lasers has also been obtained using a distributed-feedback (DFB) configuration by Wadsworth and colleagues, in 1999, and Zhu and colleagues, in 2000. As is well known, in a DFB laser no external cavity is required, the feedback being provided by Bragg reflection from a permanently or dynamically written grating structure within the gain medium as reported by Kogelnik and Shank, in 1971. By using a DFB laser configuration where the interference of two pump beams induces the required periodic modulation in the gain medium, laser emission with linewidth in the 0.01–0.06 nm range have been reported. Specifically, Wadsworth and colleagues reported a laser linewidth of 12 GHz (0.016 nm at 616 nm) using Perylene Red doped PMMA at a conversion efficiency of 20%.

It should also be mentioned that the DFB laser is also a dye laser development that has found wide and extensive applicability in semiconductor lasers employed in the telecommunications industry.

Laser Dye Survey

Cyanines

These are red and near-infrared dyes which present long conjugated methine chains (!CH = CH!) and are useful in the spectral range longer than 800 nm, where no other dyes compete with them. An important laser dye belonging to this class is 4-dicyanomethylene-2-methyl-6-(*p*-dimethylamino-styryl)-4H-pyran (DCM, Figure 10a), with laser emission in the 600–700 nm range depending on the pump and solvent, liquid or solid, used.

Xanthenes

These dyes have a xanthene ring as the chromophore and are classified into rhodamines, incorporating

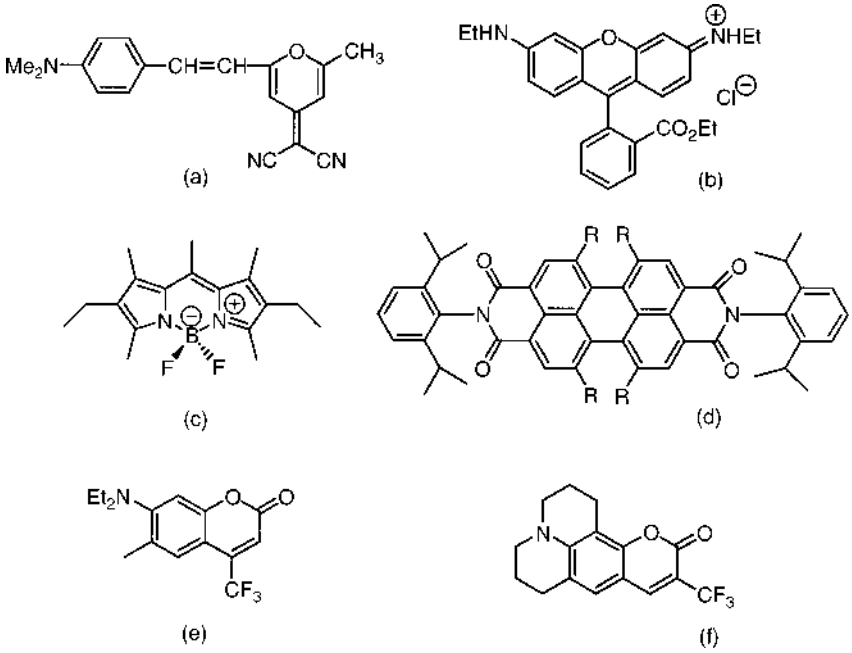

Figure 10 Molecular structures of some common laser dyes: (a) DCM; (b) Rh6G; (c) PM567; (d) Perylene Orange (R = H) and Perylene Red (R = C_4H_6O); (e) Coumarin 307 (also known as Coumarin 503); (f) Coumarin 153 (also known as Coumarin 540A).

amino radicals substituents, and fluoresceins, with hydroxyl (OH) radical substituents. They are generally very efficient and chemically stable and their emission covers the wavelength region from 500 to 700 nm.

Rhodamine dyes are the most important group of all laser materials, with rhodamine 6G (Rh6G, Figure 10b), also called rhodamine 590 chloride, being probably the best known of all laser dyes. Rh6G exhibits an absorption peak, in ethanol, at 530 nm and a fluorescence peak at 556 nm, with laser emission typically in the 550–620 nm region. It has been demonstrated to lase efficiently both in liquid and solid solutions.

Pyrromethenes

Pyrromethene.BF_2 complexes are a new class of laser dyes synthesized and characterized more recently as reported by Shah and colleagues, in 1990, Pavlopoulos and colleagues, in 1990, and Boyer and colleagues, in 1993. These laser dyes exhibit reduced triplet–triplet absorption over their fluorescence and lasing spectral region while retaining a high quantum fluorescence yield. Depending on the substituents on the chromophore, these dyes present laser emission over the spectral region from the green/yellow to the red, competing with the rhodamine dyes and have been demonstrated to lase with good performance when incorporated into solid hosts. Unfortunately, they are relatively unstable because of the aromatic amine groups in their structure, which render them vulnerable to photochemical reactions with oxygen as indicated by Rahn and colleagues, in 1997. The molecular structure of a representative and well-known dye of this class, 1,3,5,7,8-pentamethyl-2,6-diethylpyrromethene-difluoroborate complex (pyrromethene 567, PM567) is shown in Figure 10c. Recently, analogs of dye PM567 substituted at position 8 with an acetoxypolymethylene linear chain or a polymerizable methacryloyloxy polymethylene chain have been developed. These new dipyrromethene.BF_2 complexes have demonstrated improved efficiency and better photostability than the parent compound both in liquid and solid state.

Conjugated Hydrocarbons

This class of organic compounds includes perylenes, stilbenes, and *p*-terphenyl. Perylene and perylimide dyes are large, nonionic, nonpolar molecules (Figure 10d) characterized by their extreme photostability and negligible singlet–triplet transfer, as discussed by Seybold and Wagenblast and colleagues, in 1989, which have high quantum efficiency because of the absence of nonradiative relaxation. The perylene dyes exhibit limited solubility in conventional solvents but dissolve well in acetone, ethyl acetate, and methyl methacrylate, with emission wavelengths in the orange and red spectral regions.

Stilbene dyes are derivatives of uncyclic unsaturated hydrocarbons such as ethylene and butadiene, with emission wavelengths in the 400–500 nm range. Most of them end in phenyl radicals. Although they

are chemically stable, their laser performance is inferior to that of coumarins. *p*-Terphenyl is a valuable and efficient *p*-oligophenylene dye with emission in the ultraviolet.

Coumarins

Coumarin derivatives are a popular family of laser dyes with emission in the blue-green region of the spectrum. Their structure is based on the coumarin ring (Figure 10e,f) with different substituents that strongly affect its chemical characteristics and allow covering an emission spectral range between 420 and 580 nm. Some members of this class rank among the most efficient laser dyes known, but their chemical photobleaching is rapid compared with xanthene and pyrromethene dyes. An additional class of blue-green dyes is the tetramethyl derivatives of coumarin dyes introduced by Chen and colleagues, in 1988. The emission from these dyes spans the spectrum in the 453–588 nm region. These coumarin analogs exhibit improved efficiency, higher solubility, and better lifetime characteristics than the parent compounds.

Azaquinolone Derivatives

The quinolone and azaquinolone derivatives have a structure similar to that of coumarins but with their laser emission range extended toward the blue by 20–30 nm. The quinolone dyes are used when laser emission in the 400–430 nm region is required. Azaquinolone derivatives, such as 7-dimethylamino-1-methyl-4-metoxy-8-azaquinolone-2 (LD 390), exhibit laser action below 400 nm.

Oxadiazole Derivatives

These are compounds which incorporate an axodiazole ring with aryl radical substituents and which lase in the 330–450 nm spectral region. Dye 2-(4-biphenyl)-5-(4-*t*-butylphenyl)-1,3,4-oxadiazole (PBD), with laser emission in the 355–390 nm region, belongs to this family.

Solid-State Laser Dye Matrices

Although some first attempts to incorporate dye molecules into solid matrices were already made in the early days of development of dye lasers, it was not until the late 1980s that host materials with the required properties of optical quality and high damage threshold to laser radiation began to be developed. In subsequent years, the synthesis of new high-performance dyes and the implementation of new ways of incorporating the organic molecules into the solid matrix resulted in significant advances towards the development of practical tunable solid-state dye lasers. A recent detailed review of the work done in this field has been given by Costela *et al.*

Inorganic glasses, transparent polymers, and inorganic–organic hybrid matrices have been successfully used as host matrices for laser dyes. Inorganic glasses offer good thermal and optical properties but present the difficulty that the high melting temperature employed in the traditional process of glass making would destroy the organic dye molecules. It was not until the development of the low-temperature sol-gel technique for the synthesis of glasses that this limitation was overcome. The sol-gel process, based on inorganic polymerization reactions performed at about room temperature and starting with metallo-organic compounds such as alkoxides or salts, as reported by Brinker and Scherrer, in 1989, provides a safe route for the preparation of rigid, transparent, inorganic matrix materials incorporating laser dyes. Disadvantages of these materials are the possible presence of impurities embedded in the matrix or the occurrence of interactions between the dispersed dye molecules and the inorganic structure (hydrogen bonds, van der Waals forces) which, in certain cases, can have a deleterious effect on the lasing characteristics of the material.

The use of matrices based on polymeric materials to incorporate organic dyes offers some technical and economical advantages. Transparent polymers exhibit high optical homogeneity, which is extremely important for narrow-linewidth oscillators, as explained by Duarte, in 1994, good chemical compatibility with organic dyes, and allow control over structure and chemical composition making possible the modification, in a controlled way, of relevant properties of these materials such as polarity, free volume, molecular weight, or viscoelasticity. Furthermore, the polymeric materials are amenable to inexpensive fabrication techniques, which facilitate miniaturization and design of integrated optical systems. The dye molecules can be either dissolved in the polymer or linked covalently to the polymeric chains.

In the early investigations on the use of solid polymer matrices for laser dyes, the main problem to be solved was that of the low resistance to laser radiation exhibited by the then existing materials. In the late 1980s and early 1990s new modified polymeric organic materials began to be developed with a laser-radiation-damage threshold comparable to that of inorganic glasses and crystals as reported by Gromov and colleagues, in 1985, and Dyumaev and colleagues, in 1992. This, together with the above-mentioned advantages, made the use of polymers in solid-state dye lasers both attractive and competitive.

An approach that intends to bring about materials in which the advantages of both inorganic glasses and polymers are preserved but the difficulties are avoided is that of using inorganic–organic hybrid matrices. These are silicate-based materials, with an inorganic Si–O–Si backbone, prepared from organosilane precursors by sol-gel processing in combination with organic cross-linking of polymerizable monomers as reported by Novak, in 1993, Sanchez and Ribot, in 1994, and Schubert, in 1995. In one procedure, laser dyes are mixed with organic monomers, which are then incorporated into the porous structure of a sol-gel inorganic matrix by immersing the bulk in the solution containing monomer and catalyst or photoinitiator as indicated by Reisfeld and Jorgensen, in 1991, and Bosch and colleagues, in 1996. Alternatively, hybrids can be obtained from organically modified silicon alkoxides, producing the so-called ORMOCERS (organically modified ceramics) or ORMOSILS (organically modified silanes) as reported by Reisfeld and Jorgensen, in 1991. An inconvenient aspect of these materials is the appearance of optical inhomogeneities in the medium due to the difference in the refractive index between the organic and inorganic phases as well as to the difference in density between monomer and polymer which causes stresses and the formation of optical defects. As a result, spatial inhomogeneities appear in the laser beam, thereby decreasing its quality as discussed by Duarte, in 1994.

Organic Hosts

The first polymeric materials with improved resistance to damage by laser radiation were obtained by Russian workers, who incorporated rhodamine dyes into modified poly(methyl methacrylate) (MPMMA), obtained by doping PMMA with low-molecular weight additives. Some years later, in 1995, Maslyukov and colleagues demonstrated lasing efficiencies in the range 40–60% with matrices of MPMMA doped with rhodamine dyes pumped longitudinally at 532 nm, with a useful lifetime (measured as a 50% efficiency drop) of 15 000 pulses at a prf of 3.33 Hz.

In 1995, Costela and colleagues used an approach based on adjusting the viscoelastic properties of the material by modifying the internal plasticization of the polymeric medium by copolymerization with appropriate monomers. Using dye Rh6G dissolved in a copolymer of 2-hydroxyethyl methacrylate (HEMA) and methyl methacrylate (MMA) and transversal pumping at 337 nm, they demonstrated laser action with an efficiency of 21% and useful lifetime of 4500 pulses (20 GJ/mol in terms of total input energy per mole of dye molecule when the output energy is down to 50% of its initial value). The useful lifetime increased to 12 000 pulses when the Rh6G chromophore was linked covalently to the polymeric chains as reported by Costela and colleagues, in 1996.

Comparative studies on the laser performance of Rh6G incorporated either in copolymers of HEMA and MMA or in MPMMA were carried out by Giffin and colleagues, in 1999, demonstrating higher efficiency of the MPMMA materials but superior normalized photostability (up to 240 GJ/mol) of the copolymer formulation.

Laser conversion efficiencies in the range 60–70% for longitudinal pumping at 532 nm were reported by Ahmad and colleagues, in 1999, with PM567 dissolved in PMMA modified with 1,4-diazobicyclo[2,2,2] octane (DABCO) or perylene additives. When using DABCO as an additive, a useful lifetime of up to 550 000 pulses, corresponding to a normalized photostability of 270 GJ/mol, was demonstrated at a 2 Hz prf.

Much lower efficiencies and photostabilities have been obtained with dyes emitting in the blue-green spectral region. Costela and colleagues, in 1996, performed some detailed studies on dyes coumarin 540A and coumarin 503 incorporated into methacrylate homopolymers and copolymers, and demonstrated efficiencies of at most 19% with useful lifetimes of up to 1200 pulses, at a 2 Hz prf, for transversal pumping at 337 nm. In 2000, Somasundaram and Ramalingam obtained useful lifetimes of 5240 and 1120 pulses, respectively, for coumarin 1 and coumarin 490 dyes incorporated into PMMA rods modified with ethyl alcohol, under transversal pumping at 337 nm and a prf of 1 Hz.

A detailed study of photo-physical parameters of R6G-doped HEMA-MMA gain media was performed by Holzer and colleagues, in 2000. Among the parameters determined by these authors are quantum yields, lifetimes, absorption cross-sections, and emission cross-sections. A similar study on the photophysical parameters of PM567 and two PM567 analogs incorporated into solid matrices of different acrylic copolymers and in corresponding mimetic liquid solutions was performed by Bergmann and colleagues, in 2001.

Preliminary experiments with rhodamine 6G-doped MPMMA at high prfs were conducted using a frequency-doubled Nd:YAG laser at 5 kHz by Duarte and colleagues, in 1996. In those experiments, involving longitudinal excitation, the pump laser was allowed to fuse a region at the incidence window of the gain medium so that a lens was formed. This lens and the exit window of the gain medium comprised a

short unstable resonator that yielded broadband emission. In 2001, Costela and colleagues demonstrated lasing in rhodamine 6G- and pyrromethene-doped polymer matrices under copper-vapor-laser excitation at prfs in the 1.0–6.2 kHz range. In these experiments, the dye-doped solid-state matrix was rotated at 1200 rpm. Average powers of 290 mW were obtained at a prf of 1 kHz and a conversion efficiency of 37%. For a short period of time average powers of up to 1 W were recorded at a prf of 6.2 kHz. In subsequent experiments by Abedin and colleagues the rpf was increased to 10 kHz by using as pump laser a diode-pumped, *Q*-switched, frequency doubled, solid state Nd:YLF laser. Initial average powers of 560 mW and 430 mW were obtained for R6G-doped and PM567-doped polymer matrices, respectively. Lasing efficiency was 16% for R6G and the useful lifetime was 6.6 min (or about 4.0 million shots).

Inorganic and Hybrid Hosts

In 1990, Knobbe and colleagues and McKiernan and colleagues, from the University of California at Los Angeles, incorporated different organic dyes, via the sol-gel techniques, into silicate (SiO_2), aluminosilicate (Al_2O_3–SiO_2), and ORMOSIL host matrices, and investigated the lasing performance of the resulting materials under transversal pumping. Lasing efficiencies of 25% and useful lifetimes of 2700 pulses at 1 Hz repetition rate were obtained from Rh6G incorporated into aluminosilicate gel and ORMOSIL matrices, respectively. When the dye in the ORMOSIL matrices was coumarin 153, the useful lifetime was still 2250 pulses. Further studies by Altman and colleagues, in 1991, demonstrated useful lifetimes of 11 000 pulses at a 30 Hz prf when rhodamine 6G perchlorate was incorporated into ORMOSIL matrices. The useful lifetime increased to 14 500 pulses when the dye used was rhodamine B.

When incorporated into xerogel matrices, pyrromethene dyes lase with efficiencies as high as the highest obtained in organic hosts but with much higher photostability: a useful lifetime of 500 000 pulses at 20 Hz repetition rate was obtained by Faloss and colleagues, in 1997, with PM597 using a pump energy of 1 mJ. The efficiency of PM597 dropped to 40% but its useful lifetime increased to over 1 000 000 pulses when oxygen-free samples were prepared, in agreement with previous results that documented the strong dependence of laser parameters of pyrromethene dyes on the presence of oxygen. A similar effect was observed with perylene dyes: deoxygenated xerogel samples of Perylene Red exhibited a useful lifetime of 250 000 pulses at a prf of 5 Hz and 1 mJ pump energy, to be compared with a useful lifetime of only 10 000 pulses obtained with samples prepared in normal conditions. Perylene Orange incorporated into polycom glass and pumped longitudinally at 532 nm gave an efficiency of 72% and a useful lifetime of 40 000 pulses at a prf of 1 Hz as reported by Rhan and King, in 1998.

A direct comparison study of laser performance of Rh6G, PM567, Perylene Red, and Perylene Orange in organic, inorganic, and hybrid hosts was carried out by Rahn and King in 1995. They found that the nonpolar perylene dyes had better performance in partially organic hosts, whereas the ionic rhodamine and pyrromethene dyes performed best in the inorganic sol-gel glass host. The most promising combinations of dye and host for efficiency and photostability were found to be Perylene Orange in polycom glass and Rh6G in sol-gel glass.

An all-solid-state optical configuration, where the pump was a laser diode array side-pumped, Q-switched, frequency-doubled, Nd:YAG slab laser, was demonstrated by Hu and colleagues, in 1998, with dye DCM incorporated into ORMOSIL matrices. A lasing efficiency of 18% at the wavelength of 621 nm was obtained, and 27 000 pulses were needed for the output energy to decrease by 10% of its initial value at 30 Hz repetition rate. After 50 000 pulses the output energy decreased by 90% but it could be recovered after waiting for a few minutes without pumping.

Violet and ultraviolet laser dyes have also been incorporated into sol-gel silica matrices and their lasing properties evaluated under transversal pumping by a number of authors. Although reasonable efficiencies have been obtained with some of these laser dyes, the useful lifetimes are well below 1000 pulses.

More recently, Costela and colleagues have demonstrated improved conversion efficiencies in dye-doped inorganic–organic matrices using pyrromethene 567 dye. The solid matrix in this case is poly-trimethylsilyl-methacrylate cross-linked with ethylene glycol and copolymerized with methyl methacrylate. Also, Duarte and James have reported on dye-doped polymer-nanoparticle laser media where the polymer is PMMA and the nanoparticles are made of silica. These authors report TEM_{00} laser beam emission and improved dn/dT coefficients of the gain media that results in reduced $\Delta\theta$.

Dye Laser Applications

Dye lasers generated a renaissance in diverse applied fields such as isotope separation, medicine,

photochemistry, remote sensing, and spectroscopy. Dye lasers have also been used in many experiments that have advanced the frontiers of fundamental physics.

Central to most applications of dye lasers is their capability of providing coherent narrow-linewidth radiation that can be tuned continuously from the near ultraviolet, throughout the visible, to the near infrared. These unique coherent and spectral properties have made dye lasers particularly useful in the field of high-resolution atomic and molecular spectroscopy. Dye lasers have been successfully and extensively used in absorption and fluorescence spectroscopy, Raman spectroscopy, selective excitations, and nonlinear spectroscopic techniques. Well documented is their use in photochemistry, that is, the chemistry of optically excited atoms and molecules, and in biology, studying biochemical reaction kinetics of biological molecules. By using nonlinear optical techniques, such as harmonic and sum frequency and difference frequency generation, the properties of dye lasers can be extended to the ultraviolet.

Medical applications of dye lasers include cancer photodynamic therapy, treatment of vascular lesions, and lithotripsy as described by Goldman in 1990.

The ability of dye lasers to provide tunable sub-GHz linewidths in the orange-red portion of the spectrum, at kW average powers, made them particularly suited to atomic vapor laser isotope separation of uranium as reported by Bass and colleagues, in 1992, Singh and colleagues, in 1994, and Nemoto and colleagues, in 1995. In this particular case the isotopic shift between the two uranium isotopes is a few GHz. Using dye lasers yielding $\Delta\nu \leq 1$ GHz the ^{235}U can be selectively excited in a multistep process by tuning the dye laser(s) to specific wavelengths, in the orange-red spectral region, compatible with a transition sequence leading to photoionization. Also, high-prf narrow-linewidth dye lasers, as described by Duarte and Piper, in 1984, are ideally suited for guide star applications in astronomy. This particular application requires a high-average power diffraction-limited laser beam at $\lambda \approx 589$ nm to propagate for some 95 km, above the surface of the Earth, and illuminate a layer of sodium atoms. Ground detection of the fluorescence from the sodium atoms allows measurements on the turbulence of the atmosphere. These measurements are then used to control the adaptive optics of the telescope thus compensating for the atmospheric distortions.

A range of industrial applications is also amenable to the wavelength agility and high-average power output characteristics of dye lasers. Furthermore, tunability coupled with narrow-linewidth emission, at large pulsed energies, make dye lasers useful in military applications such as directed energy and damage of optical sensors.

The Future of Dye Lasers

Predicting the future is rather risky. In the early 1990s, articles and numerous advertisements in commercial laser magazines predicted the demise and even the oblivion of the dye laser in a few years. It did not turn out this way. The high power and wavelength agility available from dye lasers have ensured their continued use as scientific and research tools in physics, chemistry, and medicine. In addition, the advent of narrow-linewidth solid-state dye lasers has sustained a healthy level of research and development in the field. Today, some 20 laboratories around the world are engaged in this activity.

The future of solid-state dye lasers depends largely on synthetic and manufacturing improvements. There is a need for strict control of the conditions of fabrication and a careful purification of all the compounds involved. In the case of polymers, in particular, stringent control of thermal conditions during the polymerization step is obligatory to ensure that adequate optical uniformity of the polymer matrix is achieved, and that intrinsic anisotropy developed during polymerization is minimized. From an engineering perspective what are needed are dye-doped solid-state media with improved photostability characteristics and better thermal properties. By better thermal properties is meant better dn/dT factors. To a certain degree progress in this area has already been reported. As in the case of liquid dye lasers, the lifetime of the material can be improved by using pump lasers at compatible wavelengths. In this regard, direct diode laser pumping of dye-doped solid-state matrices should lead to very compact, long-lifetime, tunable lasers emitting throughout the visible spectrum. An example of such a device would be a green laser-diode-pumped rhodamine-6G doped MPMMA tunable laser configured in a multiple-prism grating architecture.

As far as liquid lasers are concerned, there is a need for efficient water-soluble dyes. Successful developments in this area were published in the literature with coumarin-analog dyes by Chen and colleagues, in 1988, and some encouraging results with rhodamine dyes have been reported by Ray and colleagues, in 2002. Efficient water-soluble dyes could lead to significant improvements and simplifications in high-power tunable lasers.

See also

Coherent Control: Experimental. **Nonlinear Optics, Applications:** Pulse Compression via Nonlinear Optics. **Quantum Optics:** Entanglement and Quantum Information. **Relativistic Nonlinear Optics**.

Further Reading

Costela A, García-Moreno I and Sastre R (2001) In: Nalwa HS (ed.) *Handbook of Advanced Electronic and Photonic Materials and Devices*, vol. 7, pp. 161–208. San Diego, CA: Academic Press.

Diels J-C and Rudolph W (1996) *Ultrashort Laser Pulse Phenomena*. New York: Academic Press.

Demtröder W (1995) *Laser Spectroscopy*, 2nd edn. Berlin: Springer-Verlag.

Duarte FJ (ed.) (1991) *High Power Dye Lasers*. Berlin: Springer-Verlag.

Duarte FJ (ed.) (1995) *Tunable Lasers Handbook*. New York: Academic Press.

Duarte FJ (ed.) (1995) *Tunable Laser Applications*. New York: Marcel Dekker.

Duarte FJ (2003) *Tunable Laser Optics*. New York: Elsevier Academic.

Duarte FJ and Hillman LW (eds) (1990) *Dye Laser Principles*. New York: Academic Press.

Maeda M (1984) *Laser Dyes*. New York: Academic Press.

Schäfer FP (ed.) (1990) *Dye Lasers*, 3rd edn. Berlin: Springer-Verlag.

Radziemski LJ, Solarz RW and Paisner JA (eds) (1987) *Laser Spectroscopy and its Applications*. New York: Marcel Dekker.

Weber MJ (2001) *Handbook of Lasers*. New York: CRC.

Edge Emitters

J J Coleman, University of Illinois, Urbana, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The semiconductor edge emitting diode laser is a critical component in a wide variety of applications, including fiber optics telecommunications, optical data storage, and optical remote sensing. In this section, we describe the basic structures of edge emitting diode lasers and the physical mechanisms for converting electrical current into light. Laser waveguides and cavity resonators are also outlined. The power, efficiency, gain, loss, and threshold characteristics of laser diodes are presented along with the effects of temperature and modulation. Quantum well lasers are outlined and a description of grating coupled lasers is provided.

Like all forms of laser, the edge emitting diode laser is an oscillator and has three principal components; a mechanism for converting energy into light, a medium that has positive optical gain, and a mechanism for obtaining optical feedback. The basic edge emitting laser diode is shown schematically in Figure 1. Current flow is vertical through a pn junction and light is emitted from the ends. The dimensions of the laser in Figure 1 are distorted, in proportion to typical real laser diodes, to reveal more detail. In practical laser diodes, the width of the stripe contact is actually much smaller than shown and the thickness is smaller still. Typical dimensions are (stripe width × thickness × length) $2 \times 100 \times 1000\ \mu$m. Light emission is generated only within a few micrometers of the surface, which means that 99.98% of the volume of this small device is inactive.

Energy conversion in diode lasers is provided by current flowing through a forward-biased pn junction. At the electrical junction, there is a high density of injected electrons and holes which can recombine in a direct energy gap semiconductor material to give an emitted photon. Charge neutrality requires equal numbers of injected electrons and holes. Figure 2 shows a simplified energy versus momentum (E–k) diagrams for (a) direct, and (b) indirect semiconductors. In a direct energy gap material, such as GaAs, the energy minima have the same momentum vector, so recombination of an electron in the conduction band and a hole in the valence band takes place directly. In an indirect material, such as silicon, momentum conservation requires the participation of a third particle – a phonon. This third-order process is far less efficient than direct recombination and, thus far, too inefficient to support laser action.

Optical gain in direct energy gap materials is obtained when population inversion is reached.

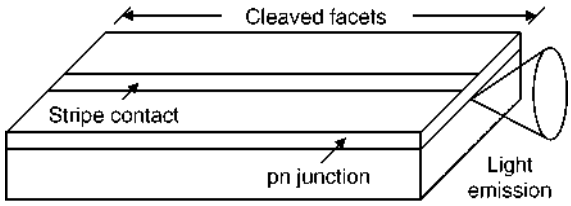

Figure 1 Schematic drawing of an edge emitting laser diode.

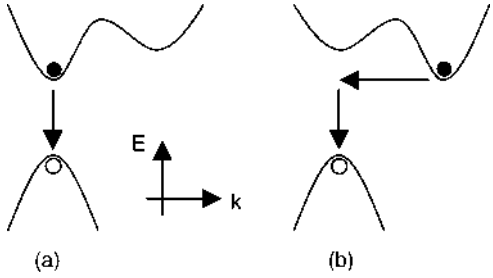

Figure 2 Energy versus momentum for (a) direct, and (b) indirect semiconductors.

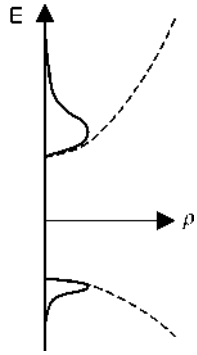

Figure 3 Density of states versus energy for conduction and valence bands.

Population inversion is the condition where the probability for stimulated emission exceeds that of absorption. The density of excited electrons n, in the conduction band is given by

$$n = \int_{E_c}^{\infty} \rho_n(E) f_n(E) \mathrm{d}E \ \mathrm{cm}^{-3} \qquad [1]$$

where $\rho_n(E)$ is the conduction band density of states function and $f_n(E)$ is the Fermi occupancy function. A similar equation can be written for holes in the valence band. Figure 3 shows the density of states ρ versus energy as a dashed line for conduction and valence bands. The solid lines in Figure 3 are the $\rho(E)f(E)$ products, and the areas under the solid curves are the carrier densities n and p.

The Fermi functions are occupation probabilities based on the quasi-Fermi levels E_{Fn} and E_{Fp}, which are, in turn, based on the injected carrier densities, δn and δp. Population inversion implies that the difference between the quasi-Fermi levels must exceed the emission energy which, for semiconductors, must be greater than the bandgap energy:

$$E_{Fn} - E_{Fp} \geq \hbar\omega \sim E_g \qquad [2]$$

Charge neutrality requires that $\delta n = \delta p$. Transparency is the point where these two conditions are just met and any additional injected current results in optical gain. The transparency current density J_o, is given by

$$J_o = \frac{q n_o L_z}{\tau_{spon}} \qquad [3]$$

where n_o is the transparency carrier density ($\delta n = \delta p = n_o$ at transparency), L_z is the thickness of the optically active layer, and τ_{spon} is the spontaneous carrier lifetime in the material (typically 5 nsec).

A resonant cavity for optical feedback is obtained simply by taking advantage of the natural crystal formation in single crystal semiconductors. Most III–V compound semiconductors form naturally into a zincblende lattice structure. With the appropriate choice of crystal planes and surfaces, this lattice structure can be cleaved such that nearly perfect plane–parallel facets can be formed at arbitrary distances from each other. Since the refractive index of these materials is much larger (~3.5) than that of air, there is internal optical reflection for normal incidence of approximately 30%. This type of optical cavity is called a Fabry–Perot resonator.

Effective lasers of all forms make use of optical waveguides to optimize the overlap of the optical field with material gain and cavity resonance. The optical waveguide in an edge emitting laser is best described by considering it in terms of a transverse waveguide component, defined by the epitaxial growth of multiple thin heterostructure layers, and a lateral waveguide component, defined by conventional semiconductor device processing methods. One of the simplest, and yet most common, heterostructure designs is shown in Figure 4. This five-layer separate confinement heterostructure (SCH) offers excellent carrier confinement in the active layer as well as a suitable transverse waveguide.

Clever choice of different materials for each layer results in the energy band structure, index of refraction

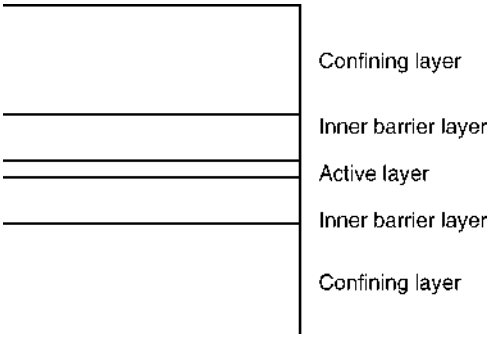

Figure 4 Schematic cross-section of a typical five-layer separate confinement heterostructure (SCH) laser.

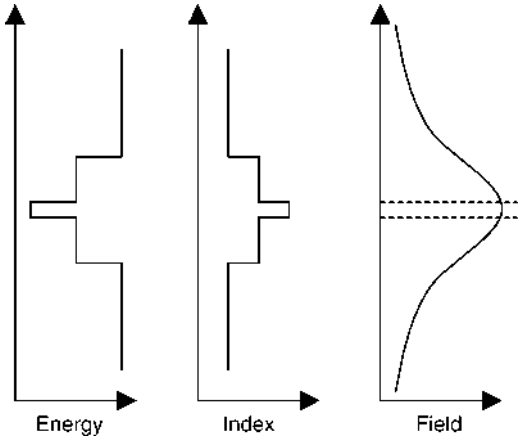

Figure 5 Energy band structure, index of refraction profile, and optical field mode profile of the SCH laser of **Figure 4**.

profile, and optical field mode profile shown in Figure 5. The outer confining layers are the thickest ($\sim 1\ \mu m$) and have the largest energy bandgap and lowest refractive index. The inner barrier layers are thinner ($\sim 0.1\ \mu m$), have intermediate bandgap energy and index of refraction, and serve both an optical role in defining the optical waveguide and an electronic role in confining carriers to the active layer. The active layer is the narrowest bandgap, highest index material, and is the only layer in the structure designed to provide optical gain. It may be a single layer or, as in the case of a multiple quantum well laser, a combination of layers. The bandgap energy, E, of this active layer plays a key role in determining the emission wavelength, λ, of the laser:

$$E = \frac{hc}{\lambda} \qquad [4]$$

where h is Planck's constant and c is the speed of light.

The bandgap energy of the SCH structure is shown in Figure 5. The lowest energy active layer collects injected electrons and holes and the carrier confinement provided by the inner barrier layers allows the high density of carriers necessary for population inversion and gain. The layers in the structure also have the refractive index profile shown in Figure 5 which forms a five-layer slab dielectric waveguide. With appropriate values for thicknesses and indices of refraction, this structure can be a very strong fundamental mode waveguide with the field profile shown. A key parameter of edge emitting diode lasers is Γ, the optical confinement factor. This parameter is the areal overlap of the optical field with the active layer, shown as dashed lines in the figure and can be as little as a few percent, even in very high performance lasers.

Lateral waveguiding in a diode laser can be accomplished in a variety of ways. A common example of an index guided laser is the ridge waveguide laser shown in Figure 6. After the appropriate transverse waveguide heterostructure sample is grown, conventional semiconductor processing methods are used to form parallel etched stripes from the surface near, but not through, the active layer. An oxide mask is patterned such that electrical contact is formed only to the center (core) region between the etched stripes, the core stripe being only a few microns wide. The lateral waveguide that results arises from the average index of refraction (effective index) in the etched regions outside the core being smaller than that of the core. Figure 7 shows the asymmetric near field emission profile of the ridge waveguide laser and cross-sectional profiles of the laser emission and effective index of refraction.

The optical spectrum of an edge emitting diode laser below and above threshold is the superposition of the material gain of the active layer and the resonances provided by the Fabry–Perot resonator. The spectral variation of material gain with injected carrier density (drive current) is shown in Figure 8. As the drive is increased, the intensity and energy of peak gain both increase.

The Fabry–Perot resonator has resonances every half wavelength, given by

$$L = \frac{m}{2}\frac{\lambda}{n} \qquad [5]$$

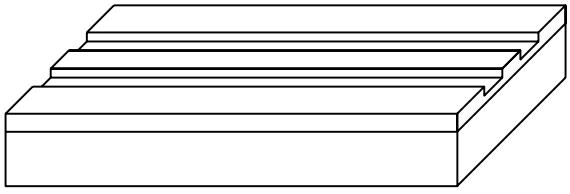

Figure 6 Schematic diagram of a real index guided ridge waveguide diode laser.

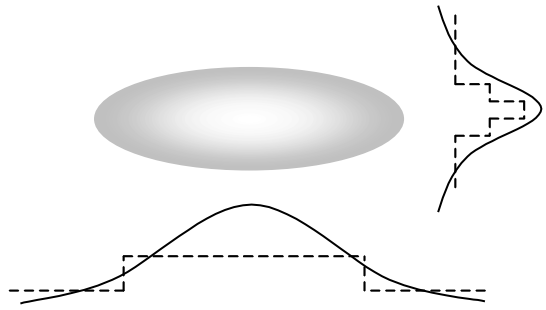

Figure 7 Near field emission pattern, cross-sectional emission profiles (solid lines), and refractive index profiles (dashed lines) from an index guided diode laser.

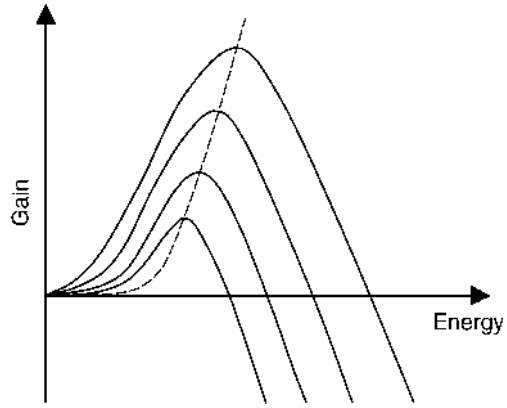

Figure 8 Gain versus emission energy for four increasing values of carrier (current) density. Peak gain is shown as a dashed line.

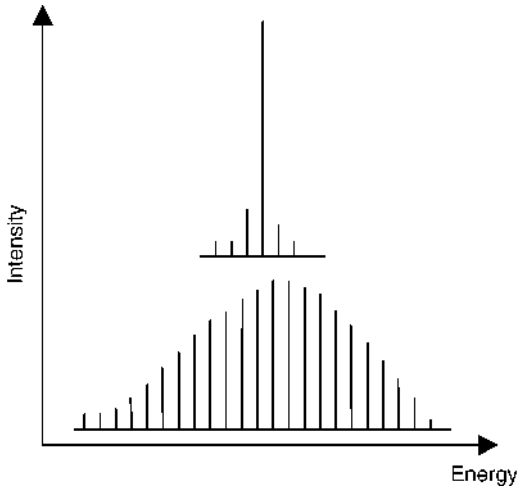

Figure 9 Emission spectra from a Fabry–Perot edge emitting laser structure at, and above, laser threshold.

where L is the cavity length, λ/n is the wavelength in the medium (n is the refractive index), and m is an integer. For normal length edge emitter cavities, these cavity modes are spaced only 1 or 2 Å apart. The result is optical spectra, near laser threshold and above, that look like the spectra of Figure 9. Many cavity modes resonate up to threshold, while above threshold, the superlinear increase in emission intensity with drive current tends to greatly favor one or two of the Fabry–Perot longitudinal cavity modes.

Laser threshold can be determined from a relatively simple analysis of the round trip gain and losses in the Fabry–Perot resonator. If the initial intensity is I_o, after one round trip the intensity will be given by

$$I = I_o R_1 R_2 e^{(g-\alpha_i)2L} \qquad [6]$$

where R_1 and R_2 are the reflectivities of the two facets, g is the optical gain, α_i are losses associated with optical absorption (typically 3–15 cm^{-1}), and L is the cavity length. At laser threshold, $g = g_{th}$ and $I = I_o$, i.e., the round trip gain exactly equals the losses. The result is

$$g_{th} = \alpha_i + \frac{1}{2L} \ln \frac{1}{R_1 R_2} \qquad [7]$$

where the second term represents the mirror losses α_m. Of course, mirror losses are desirable in the sense that they represent useable output power. Actually, the equation above does not take into account the incomplete overlap of the optical mode with the gain in the active layer. The overlap is defined by the confinement factor Γ, described above, and the equation must be modified to become

$$\Gamma g_{th} = \alpha_i + \frac{1}{2L} \ln \frac{1}{R_1 R_2} \qquad [8]$$

For quantum well lasers, the material peak gain, shown in Figure 8, as a function of drive current is given approximately by

$$g_t = \beta J_o \ln \frac{J}{J_o} \qquad [9]$$

where J_o is the transparency current density. The threshold current density J_{th}, which is the current density where the gain exactly equals all losses, is given by

$$J_{th} = \frac{J_o}{\eta_i} \exp \frac{\alpha + \frac{1}{2L} \ln \frac{1}{R_1 R_2}}{\Gamma \beta J_o} \qquad [10]$$

Note that an additional parameter η_i has appeared in this equation. Not all of the carriers injected by the drive current participate in optical processes; some are lost to other nonradiative processes. This is accounted for by including the internal quantum efficiency term η_i, which is typically at least 90%. The actual drive current, of course, must include the geometry of the device and is given by

$$I_{th} = wLJ_{th} \qquad [11]$$

where L is the cavity length, and w is the effective width of the active volume. The effective width is basically the stripe width of the core but also includes current spreading and carrier diffusion under the stripe.

The power–current characteristic (P–I) for a typical edge emitting laser diode is shown in Figure 10. As the drive current is increased

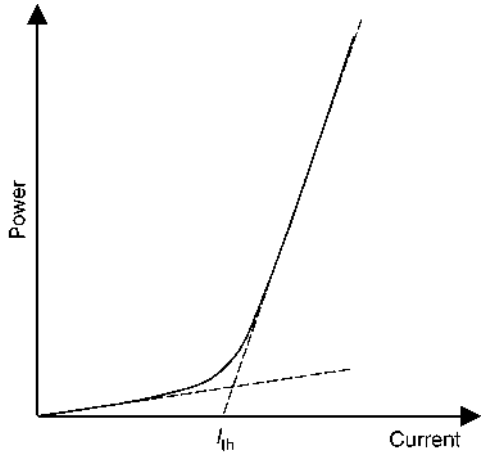

Figure 10 Power versus current (P–I) curve for a typical edge emitting laser diode.

from zero, the injected carrier densities increase and spontaneous recombination is observed. At some point the gain equals the internal absorption losses, and the material is transparent. Then, as the drive current is further increased, additional gain eventually equals the total losses, including mirror losses, and laser threshold is reached. Above threshold, nearly all additional injected carriers contribute to stimulated emission (laser output). The power generated internally is given by

$$P = wL(J - J_{th})\eta_i \frac{h\nu}{q} \qquad [12]$$

where wL is the area and $h\nu$ is the photon energy. Only a fraction of the power generated internally is extracted from the device

$$P_o = wL(J - J_{th})\eta_i \frac{h\nu}{q} \frac{\alpha_m}{\alpha_i + \alpha_m} \qquad [13]$$

Clearly a useful design goal is to minimize the internal optical loss α_i. The slope of the P–I curve of Figure 10 above threshold is described by an external differential quantum efficiency η_d, which is related to the internal quantum efficiency by

$$\eta_d = \eta_i \left[\frac{\alpha_m}{\alpha_i + \alpha_m} \right] \qquad [14]$$

Other common parameters used to characterize the efficiency of laser diodes include the power conversion efficiency η_p

$$\eta_p = \frac{P_o}{V_F I} \qquad [15]$$

and the wall plug efficiency η_w

$$\eta_w = \frac{P_o}{I} \qquad [16]$$

The power conversion efficiency η_p, recognizes that the forward bias voltage V_F is larger than the photon energy by an amount associated with the series resistance of the laser diode. The wall plug efficiency η_w, has the unusual units of WA^{-1} and is simply a shorthand term that relates the common input unit of current to the common output unit of power. In high performance low power applications, the ideal device has minimum threshold current since the current used to reach threshold contributes little to light emission. In high power applications, the threshold current becomes insignificant and the critical parameter is external quantum efficiency.

Temperature effects may be important for edge emitting lasers, depending on a particular application. Usually, concerns related to temperature arise under conditions of high temperature operation ($T > 50\,°C$), where laser thresholds rise and efficiencies fall. It became common early in the development of laser diodes to define a characteristic temperature T_o for laser threshold, which is given by

$$I_{th} = I_{th\,o} \exp\left(\frac{T}{T_o}\right) \qquad [17]$$

where a high T_o is desirable. Unfortunately, this simple expression does not describe the temperature dependence very well over a wide range of temperatures, so the appropriate temperature range must also be specified. For applications where the emission wavelength of the laser is important, the change in wavelength with temperature $d\lambda/dT$ may also be specified. For conventional Fabry–Perot cavity lasers, $d\lambda/dT$ is typically 3–5 $ÅC^{-1}$.

In order for a diode laser to carry information, some form of modulation is required. One method for obtaining this is direct modulation of the drive current in a semiconductor laser. The transient and temporal behavior of lasers is governed by rate equations for carriers and photons which are necessarily coupled equations. The rate equation for carriers is given by

$$\frac{dn}{dt} = \frac{J}{qL_z} - \frac{n}{\tau_{sp}} - (c/n)\beta(n - n_o)\varphi(E) \qquad [18]$$

where J/qL_z is the supply, n/τ_{sp} is the spontaneous emission rate, and the third term is the stimulated emission rate $(c/n)\beta(n - n_o)\varphi(E)$ where β is the gain coefficient and $\varphi(E)$ is the photon density.

The rate equation for photons is given by

$$\frac{d\varphi}{dt} = (c/n)\beta(n - n_o)\varphi + \frac{\theta n}{\tau_{sp}} - \frac{\varphi}{\tau_p} \qquad [19]$$

where θ is the fraction of the spontaneous emission that couples into the mode (a small number), and τ_p is the photon lifetime (~1 psec) in the cavity. These rate equations can be solved for specific diode laser materials and structures to yield a modulation frequency response. A typical small signal frequency response, at two output power levels, is shown in Figure 11. The response is typically flat to frequencies above 1 GHz, rises to a peak value at some characteristic resonant frequency, and quickly rolls off. The characteristic small signal resonant frequency ω_r, is given by

$$\omega_r^2 = \frac{(c/n)\beta\bar{\varphi}}{\tau_p} \qquad [20]$$

where $\bar{\varphi}$ is the average photon density. Direct modulation of edge emitting laser diodes is limited by practicality to less than 10 GHz. This is in part because of the limits imposed by the resonant frequency and in part by chirp. Chirp is the frequency modulation that arises indirectly from direct current modulation. Modulation of the current modulates the carrier densities which, in turn, results in modulation of the quasi-Fermi levels. The separation of the quasi-Fermi levels is the emission energy (wavelength). Most higher-speed lasers make use of external modulation schemes.

The discussion thus far has addressed aspects of semiconductor diode edge emitting lasers that are common to all types of diode lasers irrespective of the choice of materials or the details of the active layer structure. The materials of choice for efficient laser devices include a wide variety of III–V binary compounds and ternary or quaternary alloys. The particular choice is based on such issues as emission wavelength, a suitable heterostructure lattice match, and availability of high-quality substrates. For example, common low-power lasers for CD players ($\lambda \sim 820$ nm) are likely to be made on a GaAs substrate using a GaAs–$Al_xGa_{1-x}As$ heterostructure. Lasers for fiberoptic telecommunications systems ($\lambda \sim 1.3$–1.5 μm) are likely to be made on an InP substrate using an InP–$In_xGa_{1-x}As_yP_{1-y}$ heterostructure. Red laser diodes ($\lambda \sim 650$ nm) are likely to be made on a GaAs substrate using a GaAs–$In_xGa_{1-x}As_yP_{1-y}$ heterostructure. Blue and ultraviolet lasers, a relatively new technology compared to the others, make use of $Al_xGa_{1-x}N$–GaN–$In_yGa_{1-y}N$ heterostructures formed on sapphire or silicon carbide substrates.

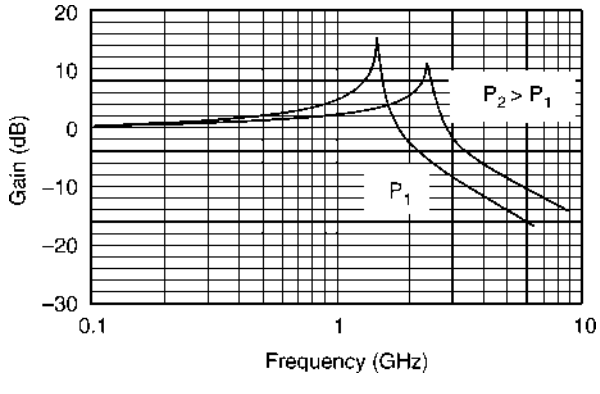

Figure 11 Direct current modulation frequency response for an edge emitting diode laser at two output power levels.

An important part of many diode laser structures is a strained layer. If a layer is thin enough, the strain arising from a modest amount of lattice mismatch can be accommodated elastically. In thicker layers, lattice mismatch results in an unacceptable number of dislocations that affect quantum efficiency, optical absorption losses, and, ultimately, long-term failure rates. Strained layer GaAs–$In_xGa_{1-x}As$ lasers are a critical component in rare-earth doped fiber amplifiers.

The structure of the active layer in edge emitting laser diodes was originally a simple double heterostructure configuration with a single active layer of 500–1000 Å in thickness. In the 1970s however, advances in the art of growing semiconductor heterostructure materials led to growth of high-quality quantum well active layers. These structures, having thicknesses that are comparable to the electron wavelength in a semiconductor (<200 Å), revolutionized semiconductor lasers, resulting in much lower threshold current densities, higher efficiencies, and a broader range of available emission wavelengths. The energy band diagram for a quantum well laser active region is shown in Figure 12.

This structure is a physical realization of the particle-in-a-box problem in elementary quantum mechanics. The quantum well yields quantum states in the conduction band at discrete energy levels, with odd and even electron wavefunctions, and breaks the degeneracy in the valence band, resulting in separate energy states for light holes and heavy holes. The primary transition for recombination is from the $n = 1$ electron state to the $h = 1$ heavy hole state and takes place at a higher energy than the bulk bandgap energy. In addition, the density of states for quantum wells becomes a step-like function and optical gain is enhanced. The advantages in practical edge

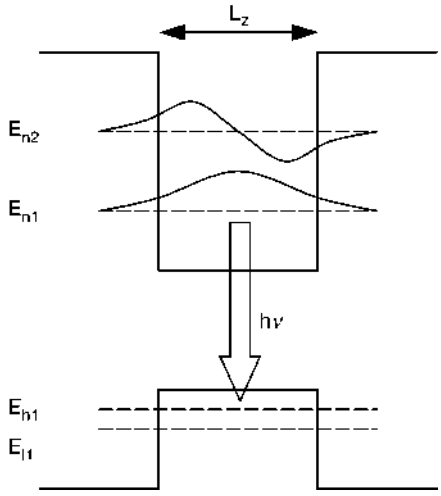

Figure 12 Energy band diagram for a quantum well heterostructure.

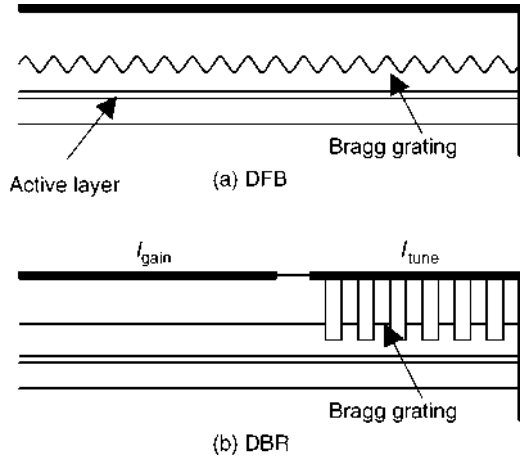

Figure 13 Schematic diagram of distributed feedback (DFB) and distributed Bragg reflector (DBR) edge emitting laser diodes.

emitting lasers that are the result of one-dimensional quantization are such that virtually all present commercial laser diodes are quantum well lasers. These kinds of advantages are driving development of laser diodes with additional degrees of quantization, including two dimensions (quantum wires) and three dimensions (quantum dots).

The Fabry–Perot cavity resonator described here is remarkably efficient and relatively easy to fabricate, which makes it the cavity of choice for many applications. There are many applications, however, where the requirements for linewidth of the laser emission or the temperature sensitivity of the emission wavelength make the Fabry–Perot resonator less desirable. An important laser technology that addresses both of these concerns involves the use of a wavelength selective grating as an integral part of the laser cavity. A notable example of a grating coupled laser is distributed feedback laser shown schematically in Figure 13a. This is similar to the SCH cross-section of Figure 4 with the addition of a Bragg grating above the active layer.

The period Λ, of the grating is chosen to fulfill the Bragg condition for mth-order coupling between forward- and backward-propagating waves, which is

$$\Lambda = \frac{m\lambda}{2n} \qquad [21]$$

where λ/n is the wavelength in the medium. The index step between the materials on either side of the grating is small and the amount of reflection is also small. Since the grating extends throughout the length of the cavity, however, the overall effective reflectivity can be large. Another variant is the distributed Bragg reflector (DBR) laser resonator, shown in Figure 13b. This DBR laser has a higher index contrast, deeply etched surface grating located at one or both ends of the otherwise conventional SCH laser heterostructure. One of the advantages of this structure is the opportunity to add a contact for tuning the Bragg grating by current injection.

Both the DBR and DFB laser structures are particularly well suited for telecommunications or other applications where a very narrow single laser emission line is required. In addition, the emission wavelength temperature dependence for this type of laser is much smaller, typically 0.5 ÅC^{-1}.

See also

Semiconductor Physics: Quantum Wells and GaAs-based Structures.

Further Reading

Agrawal GP (ed.) (1995) *Semiconductor Lasers Past, Present, and Future*. Woodbury, NY: American Institute of Physics.

Agrawal GP and Dutta NK (1993) *Semiconductor Lasers*. New York: Van Nostrand Reinhold.

Bhattacharya P (1994) *Semiconductor Optoelectronic Devices*. Upper Saddle River, NJ: Prentice-Hall.

Botez D and Scifres DR (eds) (1994) *Diode Laser Arrays*. Cambridge, UK: Cambridge University Press.

Chuang SL (1995) *Physics of Optoelectronic Devices*. New York: John Wiley & Sons.

Coleman JJ (1992) *Selected Papers on Semiconductor Diode Lasers*. Bellingham, WA: SPIE Optical Engineering Press.
Einspruch NG and Frensley WR (eds) (1994) *Heterostructures and Quantum Devices*. San Diego, CA: Academic Press.
Iga K (1994) *Fundamentals of Laser Optics*. New York: Plenum Press.
Kapon E (1999) *Semiconductor Lasers I and II*. San Diego, CA: Academic Press.
Nakamura S and Fasol G (1997) *The Blue Laser Diode*. Berlin: Springer.
Verdeyen JT (1989) *Laser Electronics*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.
Zory PS Jr (ed.) (1993) *Quantum Well Lasers*. San Diego, CA: Academic Press.

Excimer Lasers

J J Ewing, Ewing Technology Associates, Inc., Bellevue, WA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

We present a summary of the fundamental operating principles of ultraviolet, excimer lasers. After a brief discussion of the economics and application motivation, the underlying physics and technology of these devices is described. Key issues and limitations in the scaling of these lasers are presented.

Background: Why Excimer Lasers?

Excimer lasers have become the most widely used source of moderate power pulsed ultraviolet sources in laser applications. The range of manufacturing applications is also large. Production of computer chips, using excimer lasers as the illumination source for lithography, has by far the largest commercial manufacturing impact. In the medical arena, excimer lasers are used extensively in what is becoming one of the world's most common surgical procedures, the reshaping of the lens to correct vision problems. Taken together, the annual production rate for these lasers is a relatively modest number compared to the production of semiconductor lasers. Current markets are in the range of $0.4 billion per year (Figure 1). More importantly, the unique wavelength, power, and pulse energy properties of the excimer laser enable a systems and medical procedure market, based on the excimer laser, to exceed $3 billion per year. The current average sales price for these lasers is in the range of $250 000 per unit, driven primarily by the production costs and reliability requirements of the lasers for lithography for chip manufacture. Relative to semiconductor diode lasers, excimer lasers have always been, and will continue to be, more expensive per unit device by a factor of order 10^4. UV power and pulse energy, however, are considerably higher.

The fundamental properties of these lasers enable the photochemical and photophysical processes used in manufacturing and medicine. Short pulses of UV light can photo-ablate materials, being of use in medicine and micromachining. Short wavelengths enable photochemical processes. The ability to provide a narrow linewidth using appropriate resonators enables precise optical processes, such as lithography. We trace out some of the core underlying properties of these lasers that lead to the core utility. We also discuss the technological problems and

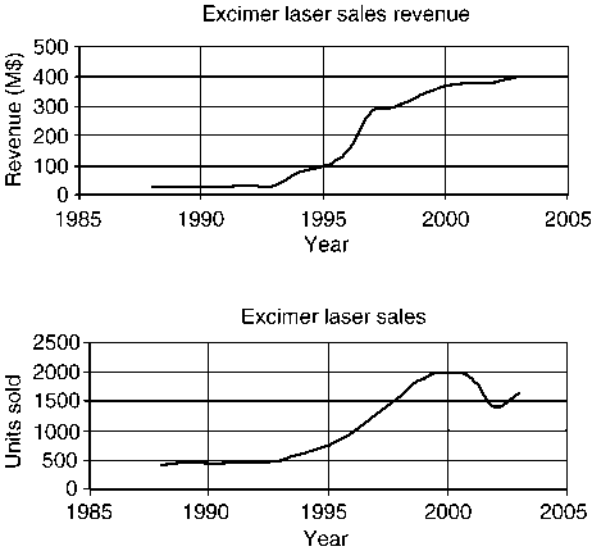

Figure 1 The sales of excimer lasers have always been measured in small numbers relative to other, less-expensive lasers. For many years since their commercial release in 1976 the sales were to R&D workers in chemistry, physics and ultimately biology and medicine. The market for these specialty but most useful R&D lasers was set to fit within a typical researcher's annual capital budget, i.e., less than $100K. As applications and procedures were developed and certified or approved, sales and average unit sales price increased considerably. Reliance on cyclical markets like semiconductor fabrication led to significant variations in production rates and annual revenues as can be seen.

solutions that have evolved to make these lasers so useful.

The data in Figure 1 provide a current answer to the question of 'why excimer lasers?' At the dawn of the excimer era, the answer to this question was quite different. In the early 1970s, there were no powerful short wavelength lasers, although IR lasers were being scaled up to significant power levels. However, visible or UV lasers could in principle provide better propagation and focus to a very small spot size at large distances. Solid state lasers of the time offered very limited pulse repetition frequency and low efficiency. Diode pumping of solid state lasers, and diode arrays to excite such lasers, was a far removed development effort. As such, short-wavelength lasers were researched over a wide range of potential media and lasing wavelengths. The excimer concept was one of those candidates.

We provide, in Figure 2, a 'positioning' map showing the range of laser parameters that can be obtained with commercial or developmental excimer lasers. The map has coordinates of pulse energy and pulse rate, with diagonal lines expressing average power. We show where both development goals and current markets lay in a map. Current applications are shown by the boxes in the lower right-hand corner. Excimer lasers have been built with clear apertures in the range of a fraction of 0.1 to 10^4 cm^2, with single pulse energies covering a range of 10^7. Lasers with large apertures require electron beam excitation. Discharge lasers correspond to the more modest pulse energy used for current applications. In general, for energies lower than ~5 J, the excitation method of choice is a self-sustained discharge, clearly providing growth potential for future uses, if required.

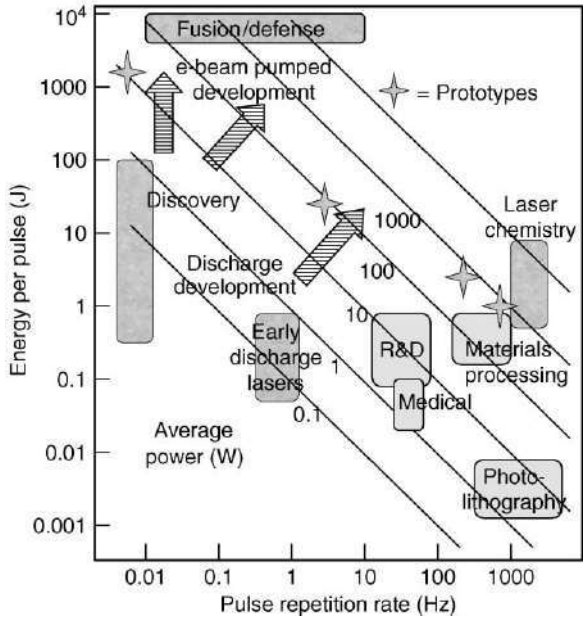

Figure 2 The typical energy and pulse rate for certain applications and technologies using excimer lasers. For energy under a few J, a discharge laser is used. The earliest excimer discharge lasers were derivatives of CO_2 TEA (transversely excited atmospheric pressure) and produced pulse energy in the range of 100 mJ per pulse. Pulse rates of order 30 Hz were all that the early pulsed power technology could provide. Early research focused on generating high energy. Current markets are at modest UV pulse energy and high pulse rates for lithography and low pulse rates for medical applications.

The applications envisaged in the early R&D days were in much higher energy uses, such as in laser weapons, laser fusion, and laser isotope separation. The perceived need for lasers for isotope separation, laser-induced chemistry, and blue-green laser-based communications from satellites, drove research in high pulse rate technology and reliability extension for discharge lasers. This research resulted in prototypes, with typical performance ranges as noted on the positioning map that well exceed current market requirements. However, the technology and underlying media physics developed in these early years made possible advances needed to serve the ultimate real market applications.

The very important application of UV lithography requires high pulse rates and high reliability. These two features differentiate the excimer laser used for lithography from the one used in the research laboratory. High pulse rates, over 2000 Hz in current practice, and the cost of stopping a computer chip production line for servicing, drive the reliability requirements toward 10^{10} shots per service interval. In contrast, the typical laboratory experiments are more often in the range of 100 Hz and, unlike a lithography production line, do not run 24 hours a day, 7 days a week. The other large market currently is in laser correction of myopia and other imperfections in the human eye. For this use the lasers are more akin to the commercial R&D style laser in terms of pulse rate and single pulse energy. Not that many shots are needed to ablate tissue to make the needed correction, and shot life is a straightforward requirement. However, the integration of these lasers into a certified and accepted medical procedure and an overall medical instrument drove the growth of this market.

The excimer concept, and the core technology used to excite these lasers, is applicable over a broad range of UV wavelengths, with utility at specific wavelength ranges from 351 nm in the near UV to 157 nm in the vacuum UV (VUV). There were many potential excimer candidates in the early R&D phase (Table 1). Some of these

Table 1 Key excimer wavelengths

Excimer emitter	*λ (nm)*	*Comment*
Ar_2	126	Very short emission wavelength, deep in the Vacuum UV; inefficient as laser due to absorption, see Xe_2.
Kr_2	146	Broad band emitter and early excimer laser with low laser efficiency. Very efficient converter of electric power into Vacuum UV light.
F_2	157	Molecule itself is not an excimer, but uses lower level dissociation; candidate for next generation lithography.
Xe_2	172	The first demonstrated excimer laser. Very efficient formation and emission but excited state absorption limits laser efficiency. Highly efficient as a lamp.
ArF	193	Workhorse for corneal surgery and lithography.
KrCl	222	Too weak compared to KrF and not as short a wavelength as ArF.
KrF	248	Best intrinsic laser efficiency; numerous early applications but found major market in lithography. Significant use in other materials processing.
XeI	254	Never a laser, but high formation efficiency and excellent for a lamp. Historically the first rare gas halide excimer whose emission was studied at high pressure.
XeBr	282	First rare gas halide to be shown as a laser; inefficient laser due to excited state absorption, but excellent fluorescent emitter; a choice for lamps.
Br_2	292	Another halogen molecule with excimer like transitions that has lased but had no practical use.
XeCl	308	Optimum medium for laser discharge excitation.
Hg_2	335	Hg vapor is very efficiently excited in discharges and forms excimers at high pressure. Subject of much early research, but as in the case of the rare gas excimers such as Xe_2 suffer from excited state absorption. Despite the high formation efficiency, this excimer was never made to lase.
I_2	342	Iodine UV molecular emission and lasing was first of the pure halogen excimer-like lasers demonstrated. This species served as kinetic prototype for the F_2 'honorary' excimer that has important use as a practical VUV source.
XeF	351, 353	This excimer laser transition terminates on a weakly bound lower level that dissociates rapidly, especially when heated. The focus for early defense-related laser development; as this wavelength propagates best of this class in the atmosphere.
XeF	480	This very broadband emission of XeF terminates on a different lower level than the UV band. Not as easily excited in a practical system, so has not had significant usage.
HgBr	502	A very efficient green laser that has many of the same kinetic features of the rare gas halide excimer lasers. But the need for moderately high temperature for the needed vapor pressure made this laser less than attractive for UV operation.
XeO	540	This is an excimer-like molecular transition on the side of the auroral lines of the O atom. The optical cross-section is quite low and as a result the laser never matured in practice.

candidates can be highly efficient, >50% relative to deposited energy, at converting electrical power into UV or visible emission, and have found a parallel utility as sources for lamps, though with differing power deposition rates and configurations. The key point in the table is that these lasers are powerful and useful sources at very short wavelengths. For lithography, the wavelength of interest has consistently shifted to shorter and shorter wavelengths as the printed feature size decreased. Though numerous candidates were pursued, as shown in Table 1, the most useful lasers are those using rare gas halide molecules. These lasers are augmented by dye laser and Raman shifting to provide a wealth of useful wavelengths in the visible and UV.

All excimer lasers utilize molecular dissociation to remove the lower laser level population. This lower level dissociation is typically from an unbound or very weakly bound ground-state molecule. However, halogen molecules also provide laser action on excimer-like transitions, that terminate on a molecular excited state that dissociates quickly. The vacuum UV transition in F_2 is the most practical method for making very short wavelength laser light at significant power. Historically, the rare gas dimer molecules, such as Xe_2, were the first to show excimer laser action, although self absorption, low gain, and poor optics in the VUV limited their utility. Other excimer-like species, such as metal halides, metal rare gas continua on the edge of metal atom resonance lines, and rare gas oxides were also studied. The key differentiators of rare gas halides from other excimer-like candidates are strong binding in the excited state, lack of self absorption, and relatively high optical cross-section for the laser transition.

Excimer Laser Fundamentals

Excimer lasers utilize lower-level dissociation to create and sustain a population inversion. For example, in the first excimer laser demonstrated, Xe_2, the lasing species is one which is comprised of two atoms that do not form a stable molecule in the ground state, as xenon dimers in the ground state are not a stable molecule. In the lowest energy state, that with neither of the atoms electronically excited, the interaction between the atoms in a collision is primarily one of repulsion, save for a very weak 'van der Waals' attraction at larger internuclear separations. This is shown in the potential energy diagram of Figure 3. If one of the atoms is excited, for example by an electric discharge, there can be a chemical binding in the electronically excited state of the molecule relative to the energy of one atom with an electron excited and one atom in the ground state. We use the shorthand notation * to denote an electronically excited atomic molecular species, e.g., Xe^*. The excited dimer, excimer for short, will recombine rapidly at high pressure from atoms into the Xe_2^* electronically excited molecule. Radiative lifetimes of the order 10 nsec to 1 μsec are typical before photon emission returns the molecule to the lower level. Collisional quenching often reduces the lifetime below the radiative rate. For Xe_2^* excimers, the emission is in the vacuum ultraviolet (VUV).

There are a number of variations on this theme, as noted in Table 1. Species, other than rare gas pairs, can exhibit broadband emission. The excited state binding energy can vary significantly, changing the fraction of excited states that will form molecules. The shape of the lower potential energy curve can vary as well. From a semantic point of view heteronuclear, diatomic molecules, such as XeO, are not excimers, as they are not made of two of the same atoms. However, the more formal name 'exciplex' did not stick in the laser world; excimer laser being preferred. Large binding energy is more efficient for capturing excited atoms into excited excimer-type molecules in the limited time they have before emission. Early measurements of the efficiency of converting electrical energy deposited into fluorescent radiation showed that up to 50% of the energy input could result in VUV light emission. Lasers were not this efficient due to absorption.

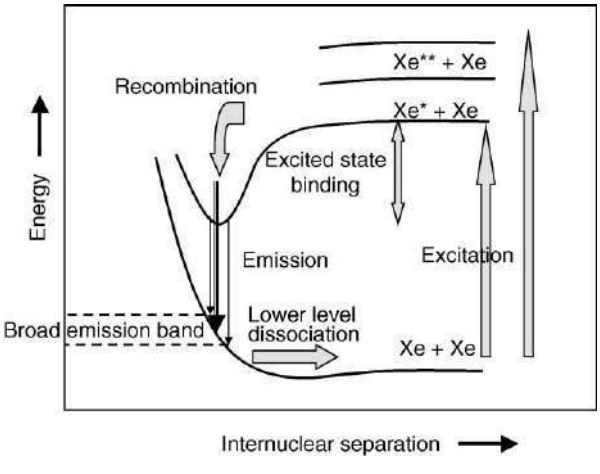

Figure 3 A schematic of the potential energy curves for an excimer species such as Xe_2. The excited molecule is bound relative to an excited atom and can radiate to a lower level that rapidly dissociates on psec timescales since the ground state is not bound, save for weak binding due to van der Waals forces. The emission band is intrinsically broad.

The diagram shown in Figure 3 is simplified because it does not show all of the excited states that exist for the excited molecule. There can be closely lying excited states that share the population and radiate at a lower rate, or perhaps absorb to higher levels, also not shown in Figure 3. Such excited state absorption may terminate on states derived from higher atomic excited states, noted as Xe^{**} in the figure, or yield photo-ionization, producing the diatomic ion-molecule plus an electron, such as:

$$Xe_2^* + h\nu \rightarrow Xe_2^+ + e \qquad [1]$$

Such excited state absorption limited the efficiency for the VUV rare gas excimers, even though they have very high formation and fluorescence efficiency.

Rare gas halide excimer lasers evolved from 'chemi-luminescence' chemical kinetics studies which were looking at the reactive quenching of rare gas metastable excited states in flowing after glow experiments. Broadband emission in reactions with halogen molecules was observed in these experiments, via reactions such as:

$$Xe^* + Cl_2 \rightarrow XeCl^* + Cl \qquad [2]$$

At low pressure the emissions from molecules, such as $XeCl^*$, are broad in the emissions bandwidth. Shortly after these initial observations, the laser community began examination of these species in high-pressure, electron-beam excited mixtures. The results differed remarkably from the low-pressure experiments and those for the earlier excimers such as Xe_2^*. The spectrum was shifted well away from the VUV emission. Moreover, the emission was much sharper at high pressure, though still a continuum with a bandwidth of the order of 4 nm. A basis for understanding is sketched in the potential energy curves of Figure 4. In this schematic we identify the rare gas atoms by Rg, excited rare gas atoms by

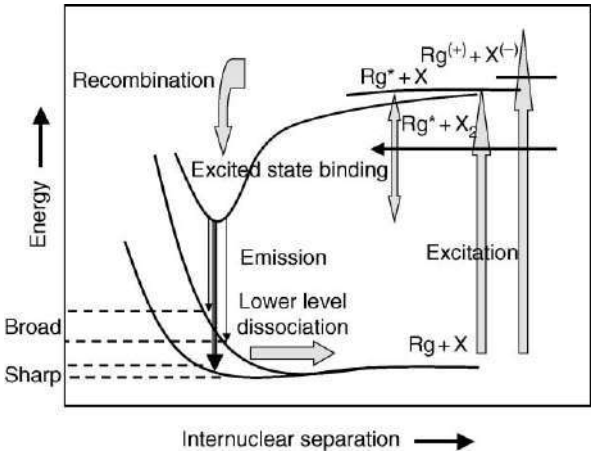

Figure 4 The potential energy curves for rare gas halides are sketched here. Both relatively sharp continuum emission is observed along with broad bands corresponding to different lower levels. The excited ion pair states consist of 3 distinct levels, two that are very close in energy and one that is shifted up from the upper laser level (the B state) by the spin orbit splitting of the rare gas ion.

Rg^*, and halogen atoms by X. Ions play a very important role in the binding and reaction chemistry, leading to excited laser molecules.

The large shift in wavelength from the VUV of rare gas dimer excimers, ~308 nm in $XeCl^*$ versus 172 nm in Xe_2^*, is due to the fact that the binding in the excited state of the rare gas halide is considerably stronger. The sharp and intense continuum at high pressure (> ~100 torr total pressure) is due to the fact that the 'sharp' transition terminates on a 'flat' portion of the lower-level potential energy curve. Indeed, in XeCl and XeF, the sharp band laser transition terminates on a slightly bound portion of the lower-level potential energy curve. Both the sharp emissions and broad bands are observed, due to the fact that some transitions from the excited levels terminate on a second, more repulsive potential energy curve.

An understanding of the spectroscopy and kinetic processes in rare gas plus halogen mixtures is based on the recognition that the excited state of a rare gas halide molecule is very similar to the ground state of the virtually isoelectronic alkali halide. The binding energy and mechanism of the excited state is very close to that of the ground state of the most similar alkali halide. Reactions of excited state rare gases are very similar to those of alkali atoms. For example, Xe^* and Cs are remarkably similar, from a chemistry and molecular physics point of view, as they have the same outer shell electron configuration and similar ionization potentials. The ionization potential of Xe^* is similar to that of cesium (Cs). Cs, and all the other alkali atoms, react rapidly with halogen molecules. Xenon metastable excited atoms react rapidly with halogen molecules. The alkali atoms all form ionically bonded ground states with halogens. The rare gas halides are effectively strongly bonded ion pairs, $Xe^{(+)}X^{(-)}$, that are very strongly bonded relative to the excited states that yield them. The difference between, for example, $XeCl^*$ and CsCl, is that $XeCl^*$ radiates in a few nanoseconds while CsCl is a stable ground-state molecule. The binding energy for these ionic bonded molecules is much greater than the more covalent type of bonding that is found in the rare gas excimer excited states. Thus XeI^*, which is only one electron different (in the I) than Xe_2^*, emits at 254 nm instead of 172 nm.

The first observed and most obvious connection to alkali atoms is in formation chemistry. Mostly, rare gas excited states react with halogen molecules rapidly. The rare gas halide laser analog reaction sequence is shown below, where some form of electric discharge excitation kicks the rare gas atom up to an excited state, so that it can react like an alkali atom, eqn [3]:

$$e_{fast} + Kr \rightarrow Kr^* + e_{slow} \qquad [3]$$

$$Kr^* + F_2 \rightarrow KrF^*(v, J)_{high} + F \qquad [4]$$

There are subtle and important cases where the neutral reaction, such as that shown in eqn [2] or [4], is not relevant as it is too slow compared to other processes. Note that the initial reaction does not yield the specific upper levels for the laser but states that are much higher up in the potential well of the excited rare gas halide excimer molecule. Further collisions with a buffer gas are needed to relax the states to the vibrational levels that are the upper levels for the laser transition. Such relaxation at high pressure can be fast, which leads the high-pressure spectrum to be much sharper than those first observed in flowing after-glow experiments.

The excited states of the rare gas halides are ion pairs bound together for their brief radiative lifetime. This opens up a totally different formation channel, indeed the one that often dominates the mechanism, ion–ion recombination. Long before excimer lasers were studied, it was well known that halogen molecules would react with 'cold' electrons (those with an effective temperature less than a few eV) in a process called attachment. The sequence leading to the upper laser level is shown below for the Kr/F_2 system, but the process is generically identical in other rare gas halide mixtures.

KrF^* formation kinetics via ion channel:

Ionization: $$e + Kr \rightarrow Kr^+ + 2e \quad [5]$$

Attachment: $$e + F_2 \rightarrow F^{(-)} + F \quad [6]$$

Ion–ion recombination:

$$Kr^+ + F^{(-)} + M \rightarrow KrF^*(v, J)_{high} + M \quad [7]$$

Relaxation:

$$KrF^*(v, J)_{high} + M \rightarrow KrF^*(v, J)_{low} + M \quad [8]$$

Finally, there is one other formation mechanism, displacement, that is unique to the rare gas halides. In this process, a heavier rare gas will displace a lighter rare gas ion in an excited state to form a lower-energy excited state:

$$ArF^* + Kr \rightarrow KrF^*(v, J)_{high} + Ar \quad [9]$$

In general, the most important excimer lasers tend to have all of these channels contributing to excited state formation for optimized mixtures.

These kinetic formation sequences express some important points regarding rare gas halide lasers. The source of the halogen atoms for the upper laser level disappears via both attachment reactions with electrons and by the reaction with excited states. The halogen atoms eventually recombine to make molecules, but at a rate too slow for sufficient continuous wave (cw) laser gain. For excimer-based lamps, however, laser gain is not a criterion and very efficient excimer lamps can be made. Another point to note is that forming the inversion requires transient species (and in some cases halogen fuels) that absorb the laser light. Net gain and extraction efficiency are a trade-off between pumping rate, an increase in which often forms more absorbers and absorption itself. Table 2 provides a listing of some of the key absorbers. Finally, the rare gas halide excited states can be rapidly quenched in a variety of processes. More often than not, the halogen molecule and the electrons in the discharge react with the excited states of the rare gas halide, removing the contributors to gain. We show, in Table 3, a sampling of some of the types of formation and quenching reactions with their kinetic rate constant parameters. Note that one category of reaction, three-body quenching, is unique to excimer lasers. Note also that the three-body recombination of ions has a rate coefficient that is effectively pressure dependent. At pressures above ~3 atm, the diffusion of ions toward each other is slowed and the effective recombination rate constant decreases.

Though the rare gas halide excited states can be made with near-unity quantum yield from the primary excitation, the intrinsic laser efficiency is not this high. The quantum efficiency is only ~25% (recall rare gas halides all emit at much longer wavelengths than rare gas excimers); there is inefficient extraction due to losses in the gas and windows. In practice there are also losses in coupling the power into the laser, and wasted excitation during the finite start-up time. The effect of losses is shown in Table 4 and Figure 5 where the pathways are charted and losses identified. In KrF, it takes ~20 eV to make a rare gas ion in an electron beam excited mixture,

Table 2 Absorbers in rare gas halide lasers

Species	*XeF*	*XeCl*	*KrF*	*ArF*
Laser λ (nm)	351, 353	308	248	193
F_2 absorption σ (cm^2)	8×10^{-21}	NA	1.5×10^{-20}	
Cl_2 absorption σ (cm^2)	NA	1.7×10^{-19}	NA	NA
HCl absorption σ (cm^2)	NA	Nil	NA	NA
$F^{(-)}$ absorption σ (cm^2)	2×10^{-18}	NA	5×10^{-18}	5×10^{-18}
$Cl^{(-)}$ absorption σ (cm^2)	NA	2×10^{-17}	NA	NA
$Rg_2^{(+)}$ diatomic ion absorption σ (cm^2)[a]	Ne: $\sim 10^{-18}$ Ar: $\sim 3 \times 10^{-17}$ Xe: $\sim 3 \times 10^{-17}$	Ne: $\sim 10^{-17}$ Ar: $\sim 4 \times 10^{-17}$ Xe: $\sim 3 \times 10^{-18}$	Ne: $\sim 2.5 \times 10^{-17}$ Ar: $\sim 2 \times 10^{-18}$ Kr $< 10^{-18}$	Ne: $\sim 10^{-18}$ Ar: NIL
Other transient absorbers	Excited states of rare gas atoms and rare gas dimer molecules $\sigma \sim 10^{-19} - 10^{-17}$	Excited states of rare gas atoms and rare gas dimer molecules $\sigma \sim 10^{-19} - 10^{-17}$	Excited states of rare gas atoms and rare gas dimer molecules $\sigma \sim 10^{-19} - 10^{-17}$	Excited states of rare gas atoms and rare gas dimer molecules $\sigma \sim 10^{-19} - 10^{-17}$
Other	Lower laser level absorbs unless heated	Weak lower level absorption $\sigma \sim 4 \times 10^{-16}$	Windows and dust can be an issue	Windows and dust can be an issue

[a]Note that the triatomic rare gas halides of the form Rg_2X will exhibit similar absorption as the diatomic rare gas ions as the triatomic rare gas halides of this form are essentially ion pairs of an $Rg_2^{(+)}$ ion and the halide ion.

Table 3 Typical formation and quenching reactions and rate coefficients

Formation	
Rare gas plus halogen molecule	$Kr^* + F_2 \rightarrow KrF^*(v, J) + F$ $k \sim 7 \times 10^{-10}$ cm^3/sec; Branching ratio to $KrF^* \sim 1$ $Ar^* + F_2 \rightarrow ArF^*(v, J) + F$ $k \sim 6 \times 10^{-10}$ cm^3/sec; Branching ratio to $ArF^* \sim 60\%$ $Xe^* + F_2 \rightarrow XeF^*(v, J) + F$ $k \sim 7 \times 10^{-10}$ cm^3/sec; Branching ratio to $XeF^* \sim 1$ $Xe^* + HCl(v = 0) \rightarrow Xe + H + Cl$ $k \sim 6 \times 10^{-10}$ cm^3/sec; Branching ratio to $XeCl^* \sim 0$ $Xe^* + HCl(v = 1) \rightarrow XeCl^* + H$ $k \sim 2 \times 10^{-10}$ cm^3/sec; Branching ratio to $XeCl^* \sim 1$
Ion–ion recombination	$Ar^{(+)} + F^{(-)} + Ar \rightarrow ArF^*(v,J)$ $k \sim 1 \times 10^{-25}$ cm^6/sec at 1 atm and below $k \sim 7.5 \times 10^{-26}$ cm^6/sec at 2 atm; Note effective 3-body rate constant decreases further as pressure goes over ~2 atm $Kr^{(+)} + F^{(-)} + Kr \rightarrow KrF^*(v, J)$ $k \sim 7 \times 10^{-26}$ cm^6/sec at 1 atm and below; rolls over at higher pressure
Displacement	$ArF^*(v, J) + Kr \rightarrow KrF^*(v, J) + Ar$ $k \sim 2 \times 10^{-10}$ cm^3/sec; Branching ratio to $KrF^* \sim 1$
Quenching	
Halogen molecule	$XeCl^* + HCl \rightarrow Xe + Cl + HCl$ $k \sim 5.6 \times 10^{-10}$ cm^3/sec; $KrF^* + F_2 \rightarrow Kr + F + F$ $k \sim 6 \times 10^{-10}$ cm^3/sec; All halogen molecule quenching have very high reaction rate constants
3-body inert gas	$ArF^* + Ar + Ar \rightarrow Ar_2F^* + Ar$ $k \sim 4 \times 10^{-31}$ cm^6/sec $KrF^* + Kr + Ar \rightarrow Kr_2F^* + Ar$ $k \sim 6 \times 10^{-31}$ cm^6/sec $XeCl^* + Xe + Ne \rightarrow Xe_2Cl^* + Ne$ $k \sim 1 \times 10^{-33}$ cm^6/sec; 4×10^{-31} cm^6/sec with Xe as 3rd body These reactions yield a triatomic excimer at lower energy that can not be recycled back to the desired laser states
Electrons	$XeCl^* + e \rightarrow Xe + Cl + e$ $k \sim 2 \times 10^{-7}$ cm^3/sec In 2-body quenching by electrons, the charge of the electron interacts with the dipole of the rare gas halide ion pair at long range; above value is typical of others
2-body inert gas	$KrF^* + Ar \rightarrow Ar + Kr + F$ $k \sim 2 \times 10^{-12}$ cm^3/sec $XeCl^* + Ne \rightarrow Ne + Xe + Cl$ $k \sim 3 \times 10^{-13}$ cm^3/sec 2-body quenching by rare gases is typically slow and less important than the reactions noted above

somewhat less in an electric discharge. The photon release is ~5 eV; the quantum efficiency is only 25%. Early laser efficiency measurements (using electron beam excitation) showed laser efficiency in the best cases slightly in excess of 10%, relative to the deposited energy. The discrepancy relative to the quantum efficiency is due to losses in the medium. For the sample estimate in Table 4 and Figure 5 we show approximate density of key species in the absorption chain, estimated on a steady-state approximation of the losses in a KrF laser for an excitation rate of order 1 MW/cc, typical of fast discharge lasers. The net small signal gain is of the order of 5%/cm. The key absorbers are the parent halogen molecule, F_2, the fluoride ion $F^{(-)}$, and the rare gas excited state atoms. A detailed pulsed kinetics code will provide slightly different results due to transient kinetic effects and the intimate coupling of laser extraction to quenching and absorber dynamics. The ratio of gain to loss is usually in the range of 7 to 20, depending on the actual mixture used and the rare gas halide wavelength. The corresponding extraction efficiency is then limited to the range of ~50%. The small signal gain, g_o, is related to the power deposition

Table 4 Typical ground and excited state densities shown for KrF at $P_{in} \sim 1$ MW/cm^3

Species	*Amount particles/cm^3*	*Absorption σ (cm)^2*	*Loss %/cm*
F_2	$\sim 1.2 \times 10^{17}$	1.5×10^{-20}	0.18
Kr	$\sim 3 \times 10^{18}$	–	–
Buffer	4×10^{19}	–	–
$F^{(-)}$	$\sim 10^{14}$	5×10^{-18}	0.05
Rg^*, Rg^{**}	1×10^{15}	$\langle 10^{-18} \rangle$ depends on ratio of Rg^{**}/Rg^*	0.1
KrF (before dissociating)	$\sim 10^{12}$	2×10^{-16}	0.02
Rg_2F^*	8×10^{14}, Note this species density decreases as flux grows	1×10^{-18}	0.08
KrF^* (B)	2.5×10^{14}	2×10^{-16}	5 (gain)

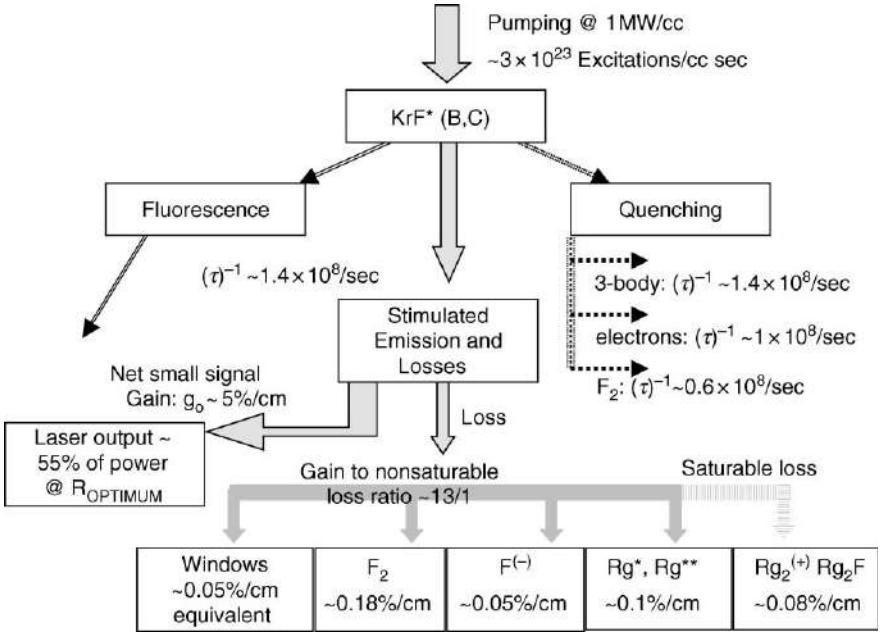

Figure 5 The major decay paths and absorbers that impact the extraction of excited states of rare gas halide excimers. Formation is via the ion–ion recombination, excited atom reaction with halogens and displacement reactions, not shown. We note for a specific mixture of KrF the approximate values of the absorption by the transient and other species along with the decay rates for the major quenching processes. Extraction efficiency greater than 50% is quite rare.

rate per unit volume, $P_{\text{deposited}}$, by eqn [10].

$$P_{\text{deposited}} = g_0 E^* (\tau_{\text{upper}} \eta_{\text{pumping}} \eta_{\text{branching}} f \sigma)^{-1} \qquad [10]$$

E^* is the energy per excitation, ~20 eV or 3.2×10^{-18} J/excitation, f is the fraction of excited molecules in the upper laser level, ~40% due to KrF^* in higher vibrational states and the slowly radiating 'C' state. The laser cross-section is σ, and the η terms are efficiencies for getting from the primary excitation down to the upper laser level. The upper state lifetime, τ_{upper}, is the sum of the inverses of radiative and quenching lifetimes. When the laser cavity flux is sufficiently high, stimulated emission decreases the flow of power into the quenching processes.

The example shown here gives a ratio of net gain to the nonsaturating component of loss of 13/1. Of the excitations that become KrF^*, only ~55% or so will be extracted in the best of cases. For other less ideal rare gas halide lasers, the gain to loss ratio, the extraction efficiency, and the intrinsic efficiency are all lower. Practical discharge lasers have lower practical wall plug efficiency due to a variety of factors, discussed below.

Discharge Technology

Practical discharge lasers use fast-pulse discharge excitation. In this scheme, an outgrowth of the CO_2 TEA laser, the gas is subjected to a rapid high-voltage pulse from a low-inductance pulse forming line or network, PFL or PFN. The rapid pulse serves to break down the gas, rendering it conductive with an electron density of $> \sim 10^{14}$ electrons/cm^3. As the

gas becomes conductive, the voltage drops and power is coupled into the medium for a short duration, typically less than 50 nsec. The discharge laser also requires some form of 'pre-ionization', usually provided by sparks or a corona-style discharge on the side of the discharge electrodes. The overall discharge process is intrinsically unstable, and pump pulse duration needs to be limited to avoid the discharge becoming an arc.

Moreover, problems with impedance matching of the PFN or PFL with the conductive gas lead to both inefficiency and extra energy that can go into post-discharge arcs, which both cause component failure and produce metal dust which can give rise to window losses. A typical discharge laser has an aperture of a few cm^2 and a gain length of order 80 cm and produces outputs in the range of 100 mJ to 1 J. The optical pulse duration tends to be ~20 nsec, though the electrical pulse is longer due to the finite time needed for the gain to build up and the stimulated emission process to turn spontaneously emitted photons into photons in the laser cavity.

The pre-ionizer provides an initial electron density on the order of 10^8 electrons/cm^3. With spark pre-ionization or corona pre-ionization, it is difficult to make a uniform discharge over an aperture greater than a few cm^2. For larger apertures, one uses X-rays for pre-ionization. X-ray pre-ionized discharge excimer lasers have been scaled to energy well over 10 J per pulse and can yield pulse durations over 100 nsec. The complexity and expense of the X-ray approach, along with the lack of market for lasers in this size range, resulted in the X-ray approach remaining in the laboratory.

Aside from the kinetic issues of the finite time needed for excitation to channel into the excimer upper levels and the time needed for the laser to 'start up' from fluorescent photons, these lasers have reduced efficiency due to poor matching of the pump source to the discharge. Ideally, one would like to have a discharge such that when it is conductive the voltage needed to sustain the discharge is less than half of that needed to break the gas down. Regrettably for the rare gas halides (and in marked distinction to the CO_2 TEA laser) the voltage that the discharge sustains in a quasi-stable mode is ~20% of the breakdown voltage, perhaps even less, depending on the specific gases and electrode shapes.

Novel circuits, which separate the breakdown phase from the fully conductive phase, can readily double the efficiency of an excimer discharge laser, though this approach is not necessary in many applications. The extra energy that is not dissipated in the useful laser discharge often results in arcing or 'hot' discharge areas after the main laser pulse, leading to electrode sputtering, dust that finds its way to windows, and the need to service the laser after ~10^8 pulses. When one considers the efficiency of the power conditioning system, along with the finite kinetics of the laser start-up process in short pulses, and the mismatch of the laser impedance to the PFN/PFL impedance, lasers that may have 10% intrinsic efficiency in the medium in e-beam excitation may only have 2% to 3% wall plug efficiency in a practical discharge laser.

A key aspect of excimer technology is the need for fast electrical circuits to drive the power into the discharge. This puts a significant strain on the switch that is used to start the process. While spark gaps, or multichannel spark gaps, can handle the required current and current rate of rise, they do not present a practical alternative for high-pulse rate operation. Thyratrons and other switching systems, such as solid state switches, are more desirable than a spark gap, but do not offer the desired needed current rate of rise. To provide the pumping rate of order 1 MW/cm^3 needed for enough gain to turn the laser on before the discharge becomes unstable, the current needs to rise to a value of the order of 20 kA in a time of ~20 nsec; the current rate of rise is ~10^{12} A/sec. This rate of rise will limit the lifetime of the switch that drives the power into the laser gas. As a response to this need, magnetic switching and magnetic compression circuits were developed so that the lifetime of the switch can be radically increased. In the simplest cases, one stage of magnetic compression is used to make the current rate of rise to be within the range of a conventional thyratron switch as used in radar systems. Improved thyratrons and a simple magnetic assist also work. For truly long operating lifetimes, multiple stages of magnetic switching and compression are used, and with a transformer one can use a solid state diode as the start switch for the discharge circuit. A schematic of an excimer pulsed power circuit, that uses a magnetic assist with a thyratron and one stage of magnetic compression, is shown in Figure 6. The magnetic switching approach can also be used to provide two separate circuits to gain high efficiency by having one circuit to break down the gas (the spiker) and a second, low-impedance 'sustainer' circuit, to provide the main power flow. In this approach, the laser itself can act as part of the circuit by holding off the sustaining voltage for a μ-sec or so during the charge cycle and before the spiker breaks down the gas. A schematic of such a circuit is shown in Figure 7. Using this type of circuit, the efficiency of a XeCl laser can be over 4% relative to the wall plug. A wide variety of technological twists on this approach have been researched.

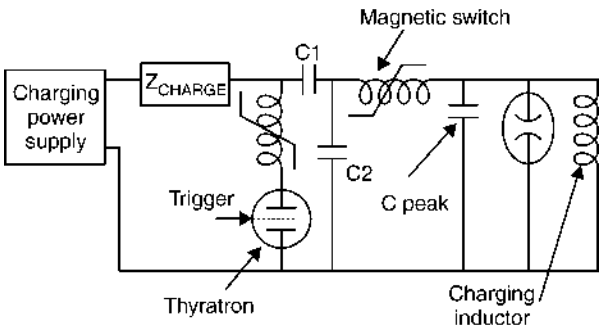

Figure 6 One of the variants on magnetic switching circuits that have been used to enhance the lifetime of the primary start switch. A magnetic switch in the form of a simple 1 turn inductor with a saturable core magnetic material such as a ferrite allows the thyratron to turn on before major current flows. Other magnetic switches may be used in addition to peak up the voltage rate of rise on the laser head or to provide significant compression. For example one may have the major current flow in the thyratron taking place on the 500 nsec time scale while the voltage pulse on the discharge is in the 50 nsec duration. A small current leaks through and is accommodated by the charging inductor.

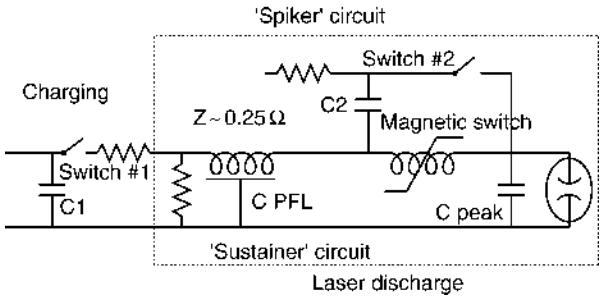

Figure 7 For the ultimate in efficiency the circuit can be arranged to provide a leading edge spike that breaks the gas down using a high impedance 'spiker' and a separate circuit that is matched to the discharge impedance when it is fully conductive. High coupling efficiency reduces waste energy that can erode circuit and laser head components via late pulse arcs. The figure shows a magnetic switch as the isolation but the scheme can be implemented with a second laser head as the isolation switch, a rail gap (not appropriate for long life) or a diode at low voltage.

Excimer lasers present some clear differences in resonator design compared to other lasers. The apertures are relatively large, so one does not build a TEM_{oo} style cavity for good beam quality and expect to extract any major portion of the energy available. The typical output is divergent and multimode. When excimer lasers are excited by a short pulse discharge, ~30 nsec, the resulting gain duration is also short, ~20 nsec, limiting the number of passes a photon in a resonator can have in the gain. The gain duration of ~20 nsec only provides ~3.5 round trips of a photon in the cavity during the laser lifetime, resulting in the typical high-order multimode beam. Even with an unstable resonator, the gain duration is not long enough to collapse the beam into the lowest-order diffraction-limited output. If line narrowing is needed, and reasonable efficiency is required, an oscillator amplifier configuration is often used with the oscillator having an appropriate tuning resonator in the cavity. If both narrow spectral linewidth and excellent beam quality are needed one uses a seeded oscillator approach where the narrowband oscillator is used to seed an unstable resonator on the main amplifier. By injecting the seed into the unstable resonator cavity, a few nsec before the gain is turned on, the output of the seeded resonator can be locked in wavelength, bandwidth, and provide good beam quality. Long pulse excimer lasers, such as e-beam excited lasers or long pulse X-ray pre-ionized discharge devices can use conventional unstable resonator technology with good results.

Excimer lasers are gas lasers that run at both high pressure and high instantaneous power deposition rates. During a discharge pulse, much of the halogen bearing 'fuel' is burned out by the attachment and reactive processes. The large energy deposition per pulse means that pressure waves are generated. All of these combine with the characteristic device dimensions to require gas flow to recharge the laser mixture between pulses. For low pulse rates, less than ~200 Hz, the flow system need not be very sophisticated. One simply needs to flush the gas from the discharge region with a flow having a velocity of the order of 5 m/sec for a typical discharge device. The flow is transverse to the optical and discharge directions. At high pulse rates, the laser designer needs to consider flow in much more detail to minimize the power required to push the gas through the laser head. Circulating pressure waves need to be damped out so that the density in the discharge region is controlled at the time of the next pulse. Devices called acoustic dampers are placed in the sides of the flow loop to remove pressure pulses. A final subtlety occurs in providing gas flow over the windows. The discharge, post pulse arcs, and reactions of the halogen source with the metal walls and impurities creates dust. When the dust coats the windows, the losses in the cavity increase, lowering efficiency. By providing a simple gas flow over the window, the dust problem can be ameliorated. For truly long life, high-reliability lasers, one needs to take special care in the selection of materials for electrodes and insulators and avoid contamination, by assembling in clean environments.

See also

Lasers: Carbon Dioxide Lasers. **Laser-Induced Damage of Optical Materials**. **Nonlinear Optics,**

Applications: Pulse Compression via Nonlinear Optics; Raman Lasers. **Scattering:** Raman Scattering.

Further Reading

Ballanti S, Di Lazzaro P, Flora F, *et al.* (1998) Ianus the 3-electrode laser. *Applied Physics B* 66: 401–406.

Brau CA (1978) Rare gas halogen lasers. In: Rhodes CK (ed.) *Excimer Lasers, Topics of Applied Physics*, vol. 30. Berlin: Springer Verlag.

Brau CA and Ewing JJ (1975) Emission spectra of XeBr, XeCl, XeF, and KrF. *Journal of Chemical Physics* 63: 4640–4647.

Ewing JJ (2000) Excimer laser technology development. *IEEE Journal of Selected Topics in Quantum Electronics* 6(6): 1061–1071.

Levatter J and Lin SC (1980) Necessary conditions for the homogeneous formation of pulsed avalanche discharges at high pressure. *Journal of Applied Physics* 51: 210–222.

Long WH, Plummer MJ and Stappaerts E (1983) Efficient discharge pumping of an XeCl laser using a high voltage prepulse. *Applied Physics Letters* 43: 735–737.

Rhodes CK (ed.) (1979) *Excimer Lasers, Topics of Applied Physics*, vol. 30. Berlin: Springer-Verlag.

Rokni M, Mangano J, Jacob J and Hsia J (1978) Rare gas fluoride lasers. *IEEE Journal of Quantum Electronics* QE-14: 464–481.

Smilanksi I, Byron S and Burkes T (1982) Electrical excitation of an XeCl laser using magnetic pulse compression. *Applied Physics Letters* 40: 547–548.

Taylor RS and Leopold KE (1994) Magnetic-spiker excitation of gas discharge lasers. *Applied Physics B, Lasers and Optics* 59: 479–509.

Free Electron Lasers

A Gover, Tel-Aviv University, Tel-Aviv, Israel

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The Free Electron Laser (FEL) is an exceptional kind of laser. Its active medium is not matter, but charged particles (electrons) accelerated to high energies, passing in vacuum through a periodic undulating magnetic field. This distinction is the main reason for the exceptional properties of FEL: operating at a wide range of wavelengths – from mm-wave to X-rays with tunability, high power, and high efficiency.

In this article we explain the physical principles of FEL operation, the underlying theory and technology of the device and various operating schemes, which have been developed to enhance performance of this device.

The term 'Free Electron Laser' was coined by John Madey in 1971, pointing out that the radiative transitions of the electrons in this device are between free space (more correctly – unbound) electron quantum states, which are therefore states of continuous energy. This is in contrast to conventional atomic and molecular lasers, in which the electron performs radiative transition between bound (and therefore of distinct energy) quantum states. Based on these theoretical observations, Madey and his colleagues in Stanford University demonstrated FEL operation first as an amplifier (at $\lambda = 10.6\ \mu m$) in 1976, and subsequently as an oscillator (at $\lambda = 3.4\ \mu m$) in 1980.

From the historical point of view, it turned out that Madey's invention was essentially an extension of a former invention in the field of microwave-tubes technology – the Ubitron. The Ubitron, a mm-wave electron tube amplifier based on a magnetic undulator, was invented and developed by Philips and Enderbry who operated it at high power levels in 1960. The early Ubitron development activity was not noticed by the FEL developers because of the disciplinary gap, and largely because its research was classified at the time. Renewed interest in high-power mm-wave radiation emission started in the 1970s, triggered by the development of pulsed-line generators of 'Intense Relativistic Beams' (IRB). This activity, led primarily by plasma physicists in the defense establishment laboratories of Russia (mostly IAP in Gorky – Nizhny Novgorod) and the US (mostly N.R.L. – DC) led to development of high-gain high-power mm-wave sources independently of the development of the optical FEL. The connection between these devices and between them to conventional microwave tubes (as Traveling Wave Tubes – TWT) and other electron beam radiation schemes, like Cenenkov and Smith-Purcell radiation that may also be considered FELs, was revealed in the mid-1970s, starting with the theoretical works of P. Spangle, A. Gover and A. Yariv who identified that all these devices satisfy the same dispersion equation as the TWT derived by John Pierce in the 1940s. Thus, the optical FEL could be conceived as a kind of immense electron tube, operating with a high-energy electron beam in the low gain regime of the Pierce TWT dispersion equation.

The extension of the low-gain FEL theory to the general 'electron-tube' theory is important because it led to development of new radiation schemes and new operating regimes of the optical FEL. This was exploited by physicists in the discipline of accelerator physics and synchrotron radiation, who identified, starting with the theoretical works of C. Pellegrini and R. Bonifacio in the early 1980s, that high-current, high-quality electron beams, attainable with further development of accelerators technology, could make it possible to operate FELs in the high-gain regime, even at short wavelengths (vacuum ultra-violet – VUV and soft X-ray) and that the high-gain FEL theory can be extended to include amplification of the incoherent synchrotron spontaneous emission (shot noise) emitted by the electrons in the undulator. These led to the important development of the 'self (synchrotron) amplified spontaneous emission (SASE) FEL', which promised to be an extremely high brightness radiation source, overcoming the fundamental obstacles of X-ray lasers development: lack of mirrors (for oscillators) and lack of high brightness radiation sources (for amplifiers).

A big boost to the development of FEL technology was given during the period of the American 'strategic defense initiative – SDI' (Star-Wars) program in the mid-1980s. The FEL was considered one of the main candidates for use in a ground-based or space-based 'directed energy weapon – DEW', that can deliver megawatts of optical power to hit attacking missiles. The program led to heavy involvement of major American defense establishment laboratories (Lawrence–Livermore National Lab, Los-Alamos National Lab) and contracting companies (TRW, Boeing). Some of the outstanding results of this effort were demonstration of the high-gain operation of an FEL amplifier in the mm-wavelength regime, utilizing an Induction Linac (Livermore, 1985), and demonstration of enhanced radiative energy extraction efficiency in FEL oscillator, using a 'tapered wiggler' in an RF-Linac driven FEL oscillator (Los-Alamos, 1983). The program has not been successful in demonstrating the potential of FELs to operate at the high average power levels needed for DEW applications. But after the cold-war period, a small part of the program continues to support research and development of medical FEL application.

Principles of FEL Operation

Figure 1 displays schematically an FEL oscillator. It is composed of three main parts: an electron accelerator, a magnetic wiggler (or undulator), and an optical resonator.

Without the mirrors, the system is simply a synchrotron undulator radiation source. The electrons in the injected beam oscillate transversely to their propagation direction z, because of the transverse magnetic Lorenz force:

$$\mathbf{F}_{\perp} = -ev_z\hat{\mathbf{e}}_z \times \mathbf{B}_{\perp} \qquad [1]$$

In a planar (linear) wiggler, the magnetic field on axis is approximately sinusoidal:

$$\mathbf{B}_{\perp} = B_{\mathrm{w}}\hat{e}_y \cos k_{\mathrm{w}}z \qquad [2]$$

In a helical wiggler:

$$\mathbf{B}_{\perp} = B_{\mathrm{w}}(\hat{\mathbf{e}}_y \cos k_{\mathrm{w}}z + \hat{\mathbf{e}}_x \sin k_{\mathrm{w}}z) \qquad [3]$$

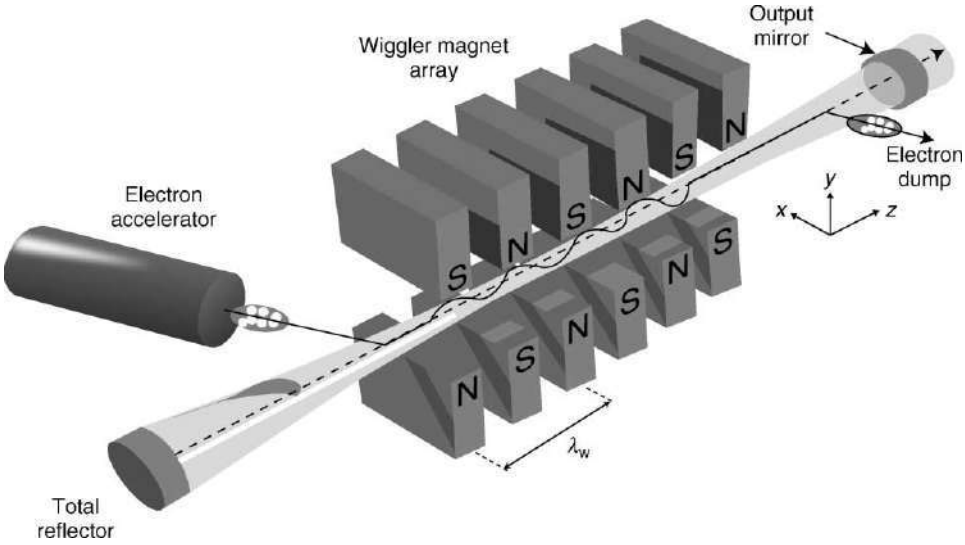

Figure 1 Components of a FEL-oscillator. (Reproduced from Benson SV (2003) Free electron lasers push into new frontiers. *Optics and Photonics News* 14: 20–25. Illustration by Jaynie Martz.)

In either case, if we assume constant (for the planar wiggler – only on the average) axial velocity, then $z = v_z t$. The frequency of the transverse force and the mechanical oscillation of the electrons, as viewed transversely in the laboratory frame of reference, is:

$$\omega_{0s} = k_w v_z = 2\pi \frac{v_z}{\lambda_w} \qquad [4]$$

where $\lambda_w = 2\pi/k_w$ is the wiggler period.

The oscillating charge emits an electromagentic radiation wavepacket. In a reference frame moving with the electrons, the angular radiation pattern looks exactly like dipole radiation, monochromatic in all directions (except for the frequency-line-broadening due to the finite oscillation time, i.e., the wiggler transit time). In the laboratory reference-frame the radiation pattern concentrates in the propagation ($+z$) direction, and the Doppler up-shifted radiation frequency depends on the observation angle Θ relative to the z-axis:

$$\omega_0 = \frac{\omega_{0s}}{1 - \beta_z \cos\Theta} \qquad [5]$$

On axis ($\Theta = 0$), the radiation frequency is:

$$\omega_0 = \frac{ck_w\beta_z}{1-\beta_z} = (1+\beta_z)\beta_z\gamma_z^2 ck_w \cong 2\gamma_z^2 ck_w \qquad [6]$$

where $\beta_z \equiv v_z/c$, $\gamma_z \equiv (1-\beta_z^2)^{-1/2}$ are the axial (average) velocity and the axial Lorenz factor, respectively, and the last part of the equation is valid only in the (common) highly relativistic limit $\gamma_z \gg 1$.

Using the relations $\beta_z^2 + \beta_\perp^2 = \beta^2$, $\beta_\perp = a_w/\gamma$, one can express γ_z:

$$\gamma_z = \frac{\gamma}{1 + a_w^2/2} \qquad [7]$$

(this is for a linear wiggler, in the case of a helical wiggler the denoninator is $1 + a_w^2$):

$$\gamma \equiv (1-\beta^2)^{-1/2} = 1 + \frac{\varepsilon_k}{mc^2} = 1 + \varepsilon_k\,[\text{MeV}]/0.511 \qquad [8]$$

and a_w – (also termed K) 'the wiggler parameter' is the normalized transverse momentum:

$$a_w = \frac{eB_w}{k_w mc} = 0.093 B_w\,[\text{KGauss}]\lambda_w\,[\text{cm}] \qquad [9]$$

Typical values of B_w in FEL wigglers (undulators) are of the order of Kgauss', and λ_w of the order of CMs, and consequently $a_w < 1$. Considering that electron beam accelerator energies are in the range of MeV to GeV, one can appreciate from eqns [6]–[8], that a significant relativistic Doppler shift factor $2\gamma_z^2$, in the range of tens to millions, is possible. It, therefore, provides incoherent synchrotron undulator radiation in the frequency range of microwave to hard X-rays.

Synchrotron undulator radiation was studied in 1951 and since then has been a common source of VUV radiation in synchrotron facilities. From the point of view of laser physics theory, this radiation can be viewed as 'spontaneous synchrotron radiation emission' in analogy to spontaneous radiation emission by electrons excited to higher bound-electron quantum levels in atoms or molecules. Alternatively, it can be regarded as the classical shot noise radiation, associated with the current fluctuations of the randomly injected discrete charges comprising the electron beam. Evidently this radiation is incoherent, and the fields it produces average in time to zero, because the wavepackets emitted by the randomly injected electrons interfere at the observation point with random phase. However, their energies sum up and can produce substantial power.

Based on fundamental quantum-electrodynamical principles or Einstein's relations, one would expect that any spontaneous emission scheme can be stimulated. This principle lies behind the concept of the FEL, which is nothing but stimulated undulator synchrotron radiation. By stimulating the electron beam to emit radiation, it is possible, as with any laser, to generate a coherent radiation wave and extract more power from the gain medium, which in this case is an electron beam, that carries an immense amount of power. There are two kinds of laser schemes which utilize stimulation of synchrotron undulator radiation:

(i) *A laser amplifier*. In this case the mirrors in the schematic configuration of **Figure 1** are not present, and an external radiation wave at frequencies within the emission range of the undulator is injected at the wiggler entrance. This requires, of course, an appropriate radiation source to be amplified and availability of sufficiently high gain in the FEL amplifier.

(ii) *A laser oscillator*. In this case an open cavity (as shown in **Figure 1**) or another (waveguide) cavity is included in the FEL configuration. As in any laser, the FEL oscillator starts building up its radiation from the spontaneous (synchrotron undulator) radiation which gets trapped in the resonator and amplified by stimulated emission along the wiggler. If the threshold condition is satisfied (having single path gain higher than the round trip losses), the oscillator arrives to saturation and steady state coherent operation after a short transient period of oscillation build-up.

Because the FEL can operate as a high-gain amplifier (with a long enough wiggler and an electron beam of high current and high quality), also a third mode of operation exists: self amplified spontaneous emission (SASE). In this case, the resonator mirrors in Figure 1 are not present and the undulator radiation generated spontaneously in the first sections of the long undulator is amplified along the wiggler and emitted at the wiggler exit at high power and high spatial coherence.

The Quantum-Theory Picture

A free electron, propagating in unlimited free space, can never emit a single photon. This can be proven by examining the conservation of energy and momentum conditions:

$$\varepsilon_{k_i} - \varepsilon_{k_f} = \hbar\omega \qquad [10]$$

$$\mathbf{k}_i - \mathbf{k}_f = \mathbf{q} \qquad [11]$$

that must be satisfied, when an electron in an initial free-space energy and momentum state $(\varepsilon_{k_i}, \hbar k_i)$ makes a transition to a final state $(\varepsilon_{k_f}, \hbar k_f)$, emitting a single photon of energy and momentum $(\hbar\omega, q)$. In free space:

$$\varepsilon_k = c\sqrt{(\hbar\mathbf{k})^2 + (mc^2)^2} \qquad [12]$$

$$\mathbf{q} = \frac{\omega}{c}\hat{\mathbf{e}}_q \qquad [13]$$

and eqns [10]–[13] have only one solution, $\omega = 0$, $\mathbf{q} = 0$. This observation is illustrated graphically in the energy–momentum diagram of Figure 2a in the framework of a one-dimensional model. It appears that if both eqns [10] and [11] can be satisfied, then the phase velocity of the emitted radiation wave $v_{\mathrm{ph}} = \omega/q$ (the slope of the chord) will equal the electron wavepacket group velocity $v_{\mathrm{g}} = v_{\mathrm{z}}$ at some intermediate point $k^* = p^*/\hbar$:

$$v_{\mathrm{ph}} = \frac{\omega}{q} = \frac{\varepsilon_{ki} - \varepsilon_{kf}}{\hbar(k_i - k_f)} = \left.\frac{\partial\varepsilon}{\partial p}\right|_{p^*} = v_{\mathrm{g}} \qquad [14]$$

For a radiation wave in free space (eqn [13]), this results in $c = v_{\mathrm{g}}$, which contradicts special relativity.

The reason for the failure to conserve both energy and momentum in the transition is that the photon momentum $\hbar q$ it too small to absorb the large momentum shift of the electron, as it recoils while releasing radiative energy $\hbar\omega$. This observation leads to ideas on how to make a radiative transition possible:

(i) *Limit the interaction length*. If the interaction length is L, the momentum conservation

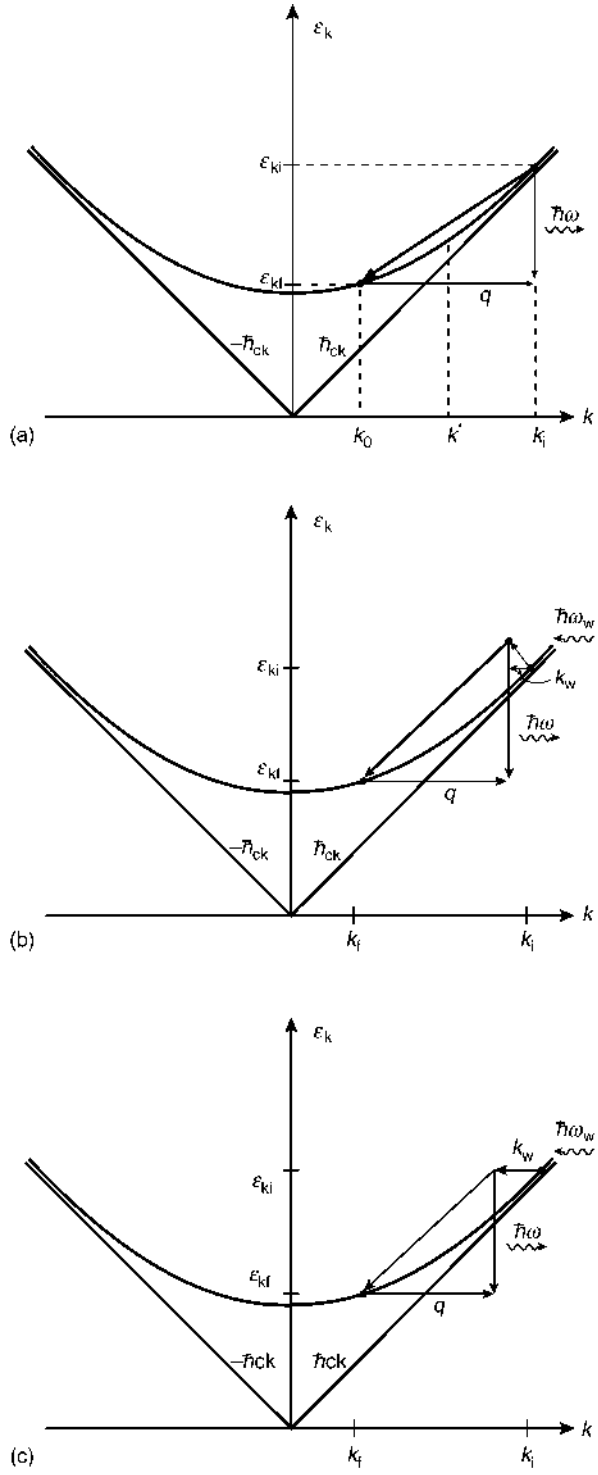

Figure 2 Conservation of energy and momentum in forward photon emission of a free electron: (a) The slope of the tangent to the curve at intermediate point k^*, $\partial\varepsilon_k(k^*)/\partial k$ may be equal to the slope of the chord $\hbar\omega/q$ which is impossible in free space. (b) electron radiative transition made possible with an electromagnetic pump (Compton Scattering). (c) The wiggler wavenumber – k_w conserves the momentum in electron radiative transition of FEL.

condition in eqn [11] must be satisfied only within an uncertainty range $\pm\pi/L$. This makes it possible to obtain radiative emission in free electron radiation effects like 'Transition Radiation' and in microwave tubes like the Klystron.

(ii) *Propagate the radiation wave* in a 'slow wave' structure, where the phase velocity of the radiation wave is smaller than the speed of light, and satisfaction of eqn [14] is possible. For example, in the Cerenkov effect, charged particles pass through a medium (gas) with index of refraction $n > 1$. Instead of eqn [13], $\mathbf{q} = n(\omega)(\omega/c)\hat{\mathbf{e}}_q$, and consequently $q_z = n(\omega)(\omega/c)\cos\Theta_q$, where we assume radiative emission at an angle Θ_q relative to the electron propagation axis z. Substitution in eqn [14] results in the Cerenkov radiation condition $v_g n(\omega)\cos\Theta_q = 1$.

Another example for radiation emission in a slow-wave structure is the Traveling Wave Tube (TWT). In this device, a periodic waveguide of periodicity λ_w permits (via the Floquet theorem) propagation of slow partial waves (space harmonics) with increased wavenumber $q_z + mk_w$ ($m = 1, 2, \ldots$), and again eqn [14] can be satisfied.

(iii) *Rely on a 'two-photon' radiative transition.* This can be 'real photon' Compton scattering of an intense radiation beam (electromagnetic pump) off an electron beam, or 'virtual photon' scattering of a static potential, as is the case in bremsstrahlung radiation and in synchrotron–undulator radiation. The latter radiation scheme may be considered as a 'magnetic brehmsstrahlung' effect or as 'zero frequency pump' Compton scattering, in which the wiggler contributes only 'crystal momentum' $\hbar k_w$, to help satisfy the momentum conservation condition in eqn [11]. The Compton scattering scheme is described schematically for the one-dimensional (back scattering) case in Figure 3, and its conservation of energy and momentum diagram is depicted in Figure 2b (a 'real photon' (ω_w, k_w) free-space pump wave is assumed with $k_w = \omega_w/c$). The analogous diagram of a static wiggler ($\omega_w = 0$, $k_w = 2\pi/\lambda_w$) is shown in Figure 2c. It is worth noting that the effect of the incident scattered wave or the wiggler is not necessarily a small perturbation. It may modify substantially the electron energy-dispersion diagram of the free electron and a more complete 'Brillouin diagram' should be used in Figure 2c. In this sense, the wiggler may be viewed as the analogue of a one-dimensional crystal, and its period λ_w analogous to the crystal lattice constant. The momentum conservation during a radiation transition, with the aid of the wiggler 'crystal momentum' $\hbar k_w$ is quite analogous to the occurrence of vertical radiative transitions in direct bandgap semiconductors, and thus the FEL has, curiously enough, some analogy to microscopic semiconductor lasers.

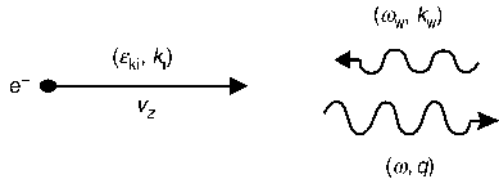

Figure 3 The scheme of backward scattering of an electromagnetic wave off an electron beam (Doppler shifted Compton scattering).

All the e-beam radiation schemes already mentioned can be turned into stimulated emission devices, and thus may be termed 'free electron lasers' in the wide sense. The theory of all of these devices is closely related, but most of the technological development was carried out on undulator radiation (Magnetic brehmsstrahlung) FELs, and the term FEL is usually reserved for this kind (though some developments of Cerenkov and Smith–Purcell FELs are still carried out).

When considering a stimulated emission device, namely enhanced generation of radiation in the presence of an external input radiation wave, one should be aware, that in addition to the emission process described by eqns [10] and [11] and made possible by one of the radiation schemes described above, there is also a stimulated absorption process. Also, this electronic transition process is governed by the conservation of energy and momentum conditions, and is described by eqns [10] and [11] with k_i and k_f exchanged.

Focusing on undulator-radiation FEL and assuming momentum conservation in the axial (z) dimension by means of the wiggler wavenumber k_w, the emission and absorption quantum transition levels and radiation frequencies are found from the solution of equations:

$$\varepsilon_{k_{zi}} - \varepsilon_{k_{zf}} = \hbar\omega_e \qquad [15a]$$

$$k_{zi} - k_{zf} = q_{ze} + k_w \qquad [15b]$$

$$\varepsilon_{k_{za}} - \varepsilon_{k_{zi}} = \hbar\omega_a \qquad [16a]$$

$$k_{za} - k_{zi} = q_{za} + k_w \qquad [16b]$$

For fixed k_w, fixed transverse momentum and given e-beam energy $\varepsilon_{k_{zi}}$ and radiation emission angle Θ_q ($q_z = (\omega/c)\cos\Theta_q$), eqns [15] and [16] have separately

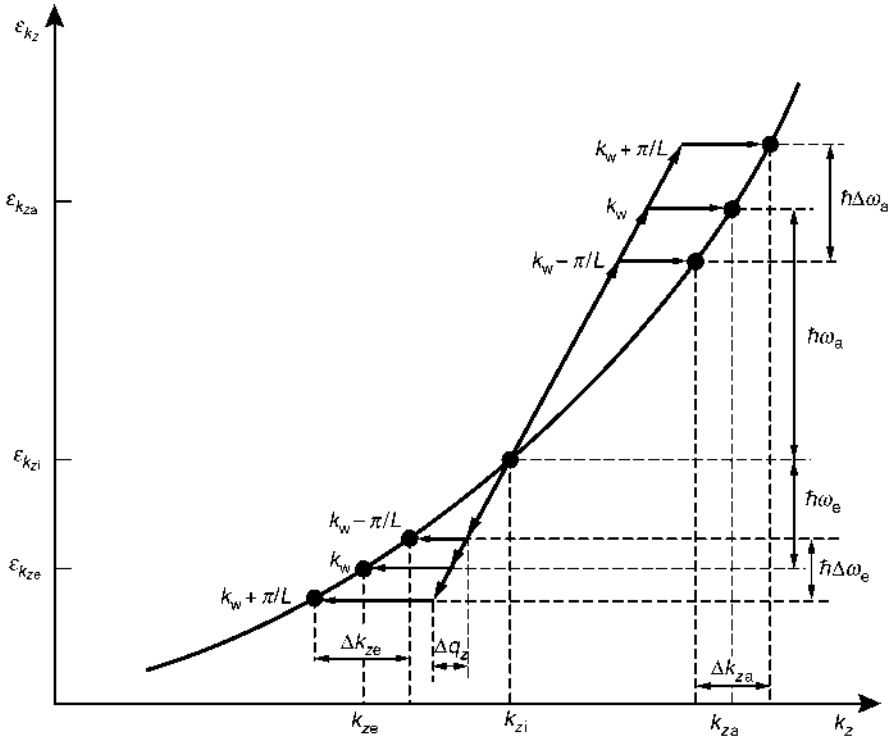

Figure 4 The figure illustrates that the origin of difference between the emission and absorption frequencies is the curvature of the energy dispersion line, and the origin of the homogeneous line broadening is momentum conservation uncertainty $\pm\pi/L$ in a finite interaction length. (Reproduced with permission from Friedman A, Gover A, Ruschin S, Kurizki G and Yariv A (1988) Spontaneous and stimulated emission from quasi-free electrons. *Reviews of Modern Physics* 60: 471–535. Copyright (1988) by the American Physical Society.)

distinct solutions, defining the electron upper and lower quantum levels for radiative emission and absorption respectively. The graphical solutions of these two set of equations are shown in Figure 4, which depicts also the 'homogeneous' frequency-line broadening $\hbar\Delta\omega_e$, $\hbar\Delta\omega_a$ of the emission and absorption lines due to the uncertainty in the momentum conservation $\pm\pi/L$ in a finite interaction length. In the quantum limit of a cold (monoenergetic) e-beam and a long interaction length L, the absorption line center ω_a is larger than the emission line center ω_e, and the linewidths $\Delta\omega_e \cong \Delta\omega_a = \Delta\omega_L$ are narrower than the emission and absorption lines spacing $\omega_a - \omega_e$, as shown in Figure 5a. The FEL then behaves as a 3-level quantum system, with electrons occupying only the central level, and the upper level is spaced apart from it more than the lower level (Figure 4).

In the classical limit $\hbar \to 0$, one can Taylor-expand ε_{kz} around k_{zi}. Using:

$$v_z = \frac{1}{\hbar}\frac{\partial\varepsilon_{kz}}{\partial k_z}, \qquad \frac{1}{\gamma_z^2\gamma m} = \frac{1}{\hbar^2}\frac{\partial^2\varepsilon_{kz}}{\partial k_z^2}$$

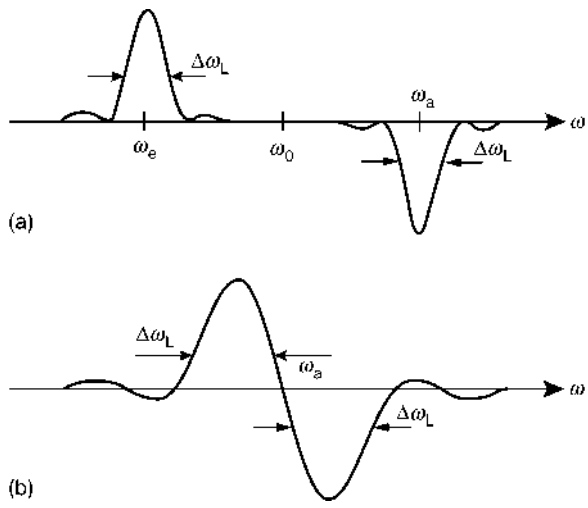

Figure 5 Net gain emission/absorption frequency lines of FEL: (a) in the quantum limit: $\omega_a - \omega_e \gg \Delta\omega_L$, (b) in the classical limit: $\omega_a - \omega_e \ll \Delta\omega_L$.

one obtains:

$$\omega_e \cong \omega_a \cong \omega_0 = v_z(q_{z0} + k_w) \qquad [17]$$

which for $q_{z0} = (\omega/c)\cos\Theta_q$ reproduces the classical synchronism condition in eqn [5]. The homogeneous

broadening linewidth is found to be:

$$\frac{\Delta\omega_L}{\omega_0} = \frac{1}{N_w} \qquad [18]$$

where $N_w = L/\lambda_w$ is the number of wiggler periods.

The classical limit condition requires that the difference between the emission and absorption line centers will be smaller than their width. This is expressed in terms of the 'recoil parameter ε' (for $\Theta_q = 0$):

$$\varepsilon = \frac{\omega_a - \omega_e}{\Delta\omega_L} = \frac{1+\beta}{\beta_z^2}\frac{\hbar\omega_0}{\gamma mc^2}N_w \ll 1 \qquad [19]$$

This condition is satisfied in all practical cases of realizable FELs. When this happens, the homogeneous line broadening dominates over the quantum-recoil effect, and the emission and absorption lines are nearly degenerate (Figure 5b). The total quantum-electrodynamic photonic emission rate expression:

$$\frac{d\nu_q}{dt} = \Gamma_{sp}[(\nu_q+1)F(\omega-\omega_e) - \nu_q F(\omega-\omega_a)] \qquad [20]$$

reduces then into:

$$\frac{d\nu_q}{dt} = \nu_q\Gamma_{sp}\varepsilon\Delta\omega_L\frac{d}{d\omega}F(\omega-\omega_e) + \Gamma_{sp}F(\omega-\omega_0) \qquad [21]$$

Here ν_q is the number of photons in radiation mode q, Γ_{sp} – the spontaneous emission rate, and $F(\omega-\omega_0)$ is the emission (absorption) lineshape function. Figure 5b depicts the transition of the net radiative emission/absorption rate into a gain curve which is proportional to the derivative of the spontaneous emission lineshape function (first term in eqn [21]). Equation [21] presents a fundamental relation between the spontaneous and stimulated emission of FELs, which was observed first by John Madey (Madey's theorem). It can be viewed as an extension of Einstein's relations to a classical radiation source.

The Classical Picture

The spontaneous emission process of FEL (synchrotron undulator radiation) is nothing but dipole radiation of the undulating electrons, which in the laboratory frame of reference is Doppler shifted to high frequency. The understanding of the stimulated emission process requires a different approach.

Consider a single electron, following a sinusoidal trajectory under the effect of a planar undulator magnetic field in eqn [2] (Figure 1):

$$v_x = v_w \cos k_w z_e(t) \qquad [22]$$

$$x = x_w \sin k_w z_e(t) \qquad [23]$$

where $v_w = ca_w/\gamma$ and $x_w = v_w/(v_z k_w)$. An electromagnetic wave $E_x(z,t) = E_0\cos(\omega t - k_z z)$ propagates collinearly with the electron. Figure 6 displays the electron and wave 'snap-shot' positions as they propagate along one wiggler period λ_w.

If the electron, moving at average axial velocity v_z, enters the interaction region $z = 0$ at $t = 0$, its axial position is $z_e(t) = v_z t$, and the electric force it experiences is $-eE_x(z_e(t),t) = -eE_0\cos(\omega - k_z v_z)t$. Clearly, this force is (at least initially at $t = 0$) opposite to the transverse velocity of the electron $v_x = v_w\cos(k_w v_z)t$ (imply in deceleration) and the power exchange rate $-ev_e\cdot E = -ev_x E_x$ corresponds to transfer of energy into the radiation field on account of the electron kinetic energy. Because the phase velocity of the radiation mode is larger than the electron velocity, $v_{ph} = \omega/k_z > v_z$, the electron phase $\varphi_e = (\omega - k_z v_z)t$ grows, and the power exchange rate $-ev_x E_x$ changes. However, if one synchronizes the electron velocity, so that while the electron traverses one wiggler period ($t = \lambda_w/v_z$), the electron phase advances by $2\pi : (\omega - k_z v_z)\cdot\lambda_w/v_z = 2\pi$, then the power exchange rate from the electron to the wave remains non-negative through the entire interaction length, because then the electron transverse velocity

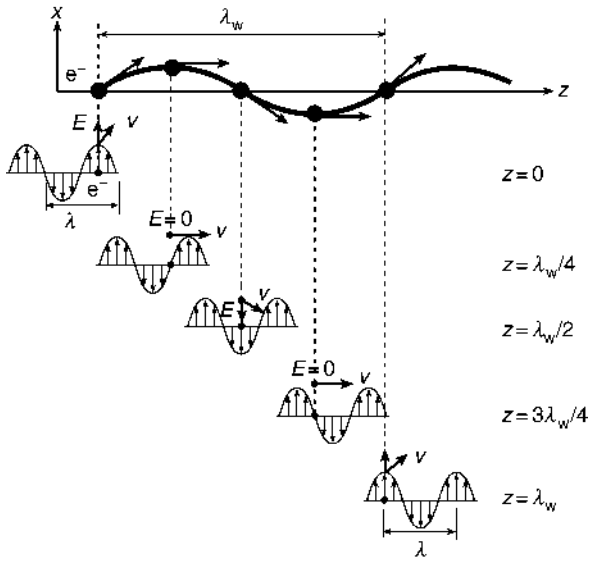

Figure 6 'Snapshots' of an electromagnetic wave period slipping relative to an undulating electron along one wiggler period λ_w. The energy transfer to the wave – $e\mathbf{v}\cdot\mathbf{E}$ remains non-negative all along.

v_x and the wave electric field E_x reverse sign, at each period exactly at the same points ($\lambda_w/4, 3\lambda_w/4$). This situation is depicted in Figure 6, which shows the slippage of the wave crests relative to the electron at five points along one wiggler period. The figure describes the synchronism condition, in which the radiation wave slips one optical period ahead of the electron, while the electron goes through one wiggle motion. In all positions along this period, $v \cdot E \geq 0$ (in a helical wiggler and a circularly polarized wave this product is constant and positive $v \cdot E > 0$ along the entire period). Substituting $\lambda_w = 2\pi/k_w$, this phase synchronism condition may be written as:

$$\frac{\omega}{v_z} = k_z + k_w \qquad [24]$$

which is the same as eqns [17] and [5].

Figure 6 shows that a single electron (or a bunch of electrons of duration smaller than an optical period) would amplify a co-propagating radiation wave, along the entire wiggler, if it satisfies the synchronism condition in eqn [24] and enters the interaction region ($z = 0$) at the right (decelerating) phase relative to the radiation field. If the electron enters at the opposite phase it accelerates (on account of the radiation field energy which is then attenuated by 'stimulated absorption'). Thus, when an electron beam is injected into a wiggler at the synchronism condition with electrons entering at random times, no net amplification or absorption of the wave is expected on the averages. Hence, some more elaboration is required, in order to understand how stimulated emission gain is possible then.

Before proceeding, it is useful to define the 'pondermotive force' wave. This force originates from the nonlinearity of the Lorenz force equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma m \mathbf{v}) = -e(\mathbf{E} \times \mathbf{v} \times \mathbf{B}) \qquad [25]$$

At zero order (in terms of the radiation fields), the only field force on the right-hand side of eqn [25] is due to the strong wiggler field (eqns [2] and [3]), which results in the transverse wiggling velocity (eqn [22] for a linear wiggler). When solving next eqn [25] to first order in terms of the radiation fields:

$$\begin{aligned} \mathbf{E}_s(r, \mathbf{t}) &= \mathrm{Re}[\tilde{\mathbf{E}}_s\, \mathrm{e}^{\mathrm{i}(k_z z - \omega t)}] \\ \mathbf{B}_s(r, \mathbf{t}) &= \mathrm{Re}[\tilde{\mathbf{B}}_s\, \mathrm{e}^{\mathrm{i}(k_z z - \omega t)}] \end{aligned} \qquad [26]$$

the cross product $\mathbf{v} \times \mathbf{B}$ between the transverse components of the velocity and the magnetic field generates a longitudinal force component:

$$\mathbf{F}_{pm}(z, t) = \mathrm{Re}[\hat{\mathbf{e}}_z \tilde{F}_{pm} \mathrm{e}^{\mathrm{i}(k_z + k_w)z - \mathrm{i}\omega t}] \qquad [27]$$

that varies with the beat wavenumber $k_s + k_w$ at slow phase velocity ($v_{ph} = \omega/(k_z + k_w) < c$). This slow force-wave is called the pomdermotive (PM) wave. Assuming the signal radiation wave in eqn [26] is polarization-matched to the wiggler (linearly polarized or circularly polarized for a linear or helical wiggler respectively), the PM force amplitude is given by:

$$|\tilde{F}_{pm}| = e|\tilde{E}_{s\perp}|a_w/\gamma\beta_z \qquad [28]$$

With large enough k_w, it is always possible to slow down the phase velocity of the pondermotive wave until it is synchronized with the electron velocity:

$$v_{ph} = \frac{\omega}{k_z + k_w} = v_z \qquad [29]$$

and can apply along the interaction length a decelerating axial force, that will cause properly phased electrons to transfer energy to the wave on account of their longitudinal kinetic energy.

This observation is of great importance. It reveals that even though the main components of the wiggler and radiation fields are transverse, the interaction is basically longitudinal. This puts the FEL on an equal footing with the slow-wave structure devices as the TWT and the Cerenkov–Smith–Purcell FELs, in which the longitudinal interaction takes place with the longitudinal electric field component of a slow TM radiation mode. The synchronism condition in eqn [29] between the pondermotive wave and the electron, which is identical with the phase-matching condition in eqn [24], is also similar to the synchronism condition between an electron and a slow electromagnetic wave (eqn [14]).

Using the pondermotive wave concept, we can now explain the achievement of gain in the FEL with a random electron beam. Figure 7 illustrates the interaction between the pondermotive wave and electrons, distributed at the entrance ($z = 0$) randomly within the wave period. Figure 7a shows 'snap-shots' of the electrons in one period of the pondermotive wave $\lambda_{pm} = 2\pi/(k_z + k_w)$ at different points along the wiggler, when it is assumed that the electron beam is perfectly synchronous with the pondermotive wave $v_{ph} = v_0$. As explained before, some electrons are slowed down, acquiring negative velocity increment Δv. However, for each such electron, there is another one, entering the wiggler at an accelerating phase of the wave, acquiring the same positive velocity increment Δv. There is then no net change in the energy of the e-beam or the wave, however, there is clearly an effect of 'velocity-bunching' (modulation), which turns along the wiggler into 'density-bunching' at the same

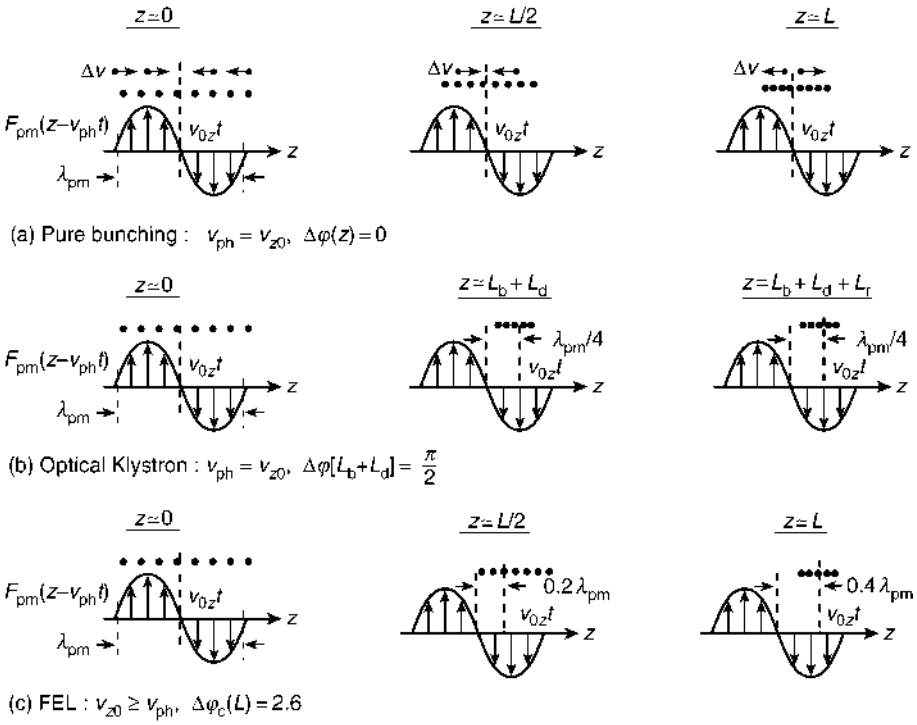

Figure 7 'Snapshots' of a pondermotive wave period interacting L with an initially uniformly distributed electron beam taking place respectively along the interaction length $0 < z < L_b$. (a) Exact synchronism in a uniform wiggler (bunching). (b) Energy bunching, density bunching and radiation are in the energy buncher, dispersive magnet $L_b < z < L_b + L_d$ and radiating wiggler $L_b + L_d < z < L_b + L_d + L_r$ sections of an Optical–Klystron. (c) Slippage from bunching phase to radiating phase at optimum detuning off synchronism in a uniform wiggler FEL.

frequency ω and wavenumber $k_z + k_w$ as the modulating pondermotive wave. The degree of density-bunching depends on the amplitude of the wave and the interaction length L. In the nonlinear limit the counter propagating (in the beam reference frame) velocity modulated electrons may over-bunch namely cross over, and debunch again.

Bunching is the principle of classical stimulated emission in electron beam radiation devices. If the e-beam had been prebunched in the first place, we would have injected it at a decelerating phase relative to the wave and obtained net radiation gain right away (super radiant emission). This is indeed the principle behind the 'Optical-Klystron' (OK) demonstrated in Figure 7b. The structure of the OK is described ahead in Figure 19. The electron beam is velocity (energy) modulated by an input electromagnetic radiation wave in the first 'bunching-wiggler section' of length L_b. It then passes through a drift-free 'energy-dispersive magnet section' (chicane) of length L_d, in which the velocity modulation turns into density bunching. The bunched electron beam is then injected back into a second 'radiating-wiggler section', where it co-propagates with the same electromagnetic wave but with a phase advance of $\pi/2 - m2\pi$, $m = 1, 2, \ldots$ (spatial lag of $\lambda_{pm}/4 - m\lambda_{pm}$ in real space) which places the entire bunch at a decelerating phase relative to the PM-wave and so amplifies the radiation wave.

The principle of stimulated-emission gain in FEL, illustrated in Figure 7c, is quite similar. Here the wiggler is uniform along the entire length L, and the displacement of the electron bunches into a decelerating phase position relative to the PM-wave is obtained by injecting the electron beam at a velocity v_{z0}, slightly higher than the wave v_{ph} (velocity detuning). The detailed calculation shows that detuning corresponding to a phase shift of $\Delta\Psi(L) = [(\omega/v_{z0}) - (k_z + k_w)]L = -2.6$ (corresponding to spatial bunch advance of $0.4\lambda_{pm}$ along the wiggler length), provides sufficient synchronism with the PM-wave in the first half of the wiggler to obtain bunching, and sufficient deceleration-phasing of the created bunches in the second part of the wiggler to obtain maximum gain.

Principles of FEL Theory

The 3D radiation field in the interaction region can be expanded in general in terms of a complete set of

free-space or waveguide modes $\{\varepsilon_q(x,y)\mathcal{H}_q(x,y)\}$:

$$E(\boldsymbol{r},t) = \mathrm{Re}\Big[\sum_q c_q(z)\varepsilon_q(x,y)\mathrm{e}^{\mathrm{i}(k_{zq}z-\omega t)}\Big] \qquad [30]$$

The mode amplitudes $C_q(z)$ may grow along the wiggler interaction length $0 < z < L$, according to the mode excitation equation:

$$\frac{\mathrm{d}}{\mathrm{d}z}C_q(z) = -\frac{1}{4\mathcal{P}}\mathrm{e}^{-\mathrm{i}k_{zq}}\int\int \mathbf{J}(x,y,z)\cdot\varepsilon_q^*(x,y)\mathrm{d}x\,\mathrm{d}y \qquad [31]$$

where $\mathcal{P} = -\frac{1}{2}\mathrm{Re}\int\int \varepsilon_q \times \mathcal{H}_q\cdot\hat{e}_z\,\mathrm{d}x\,\mathrm{d}y$ is the mode normalization power, and $\tilde{\mathbf{J}}$ is the bunching current component at frequency ω, that is phase matched to the radiation waves, and needs to be calculated consistently from the electron force equations.

The FEL Small Signal Regime

We first present the basic formulation of FEL gain in the linear (small signal) regime, namely the amplified radiation field is assumed to be proportional to the input signal radiation field, and the beam energy loss is negligible. This is done in the framework of a one-dimensional (single transverse radiation mode) model.

The electron beam charge density, current density, and velocity modulation are solved in the framework of a one-dimensional plasma equations model (kinetic or fluid equations). The longitudinal PM-force in eqn [27] modulates the electron beam velocity via the longitudinal part of the force eqn [25]. This brings about charge modulation $\rho(z,t) = \mathrm{Re}[\tilde{\rho}(k_z+k_\mathrm{w},\omega)\mathrm{e}^{\mathrm{i}(k_z+k_\mathrm{w})z-\mathrm{i}\omega t}]$ and consequently, also longitudinal space-charge field $\tilde{E}_\mathrm{sc}(k_{zq}+k_\mathrm{w},\omega)$ and longitudinal current density modulation $\tilde{J}_z(k_z+k_\mathrm{w},\omega)$, related through the Poison and continuity equations:

$$\mathrm{i}(k_z+k_\mathrm{w})\tilde{E}_\mathrm{sc}(k_z+k_\mathrm{w},\omega) = \tilde{\rho}(k_z+k_\mathrm{w},\omega)/\varepsilon \qquad [32]$$

$$(k_z+k_\mathrm{w})\tilde{J}_z(k_z+k_\mathrm{w},\omega) = \omega\tilde{\rho}(k_z+k_\mathrm{w},\omega) \qquad [33]$$

Solving the force eqn [25] for a general longitudinal force $F_z(z,t) = \mathrm{Re}[\tilde{F}_z(k_z,\omega)\mathrm{e}^{\mathrm{i}(k_z z-\omega t)}]$ results in a linear longitudinal current response relation:

$$\tilde{J}_z(k_z,\omega) = -\mathrm{i}\omega\chi_\mathrm{p}(k_z,\omega)\tilde{F}_z(k_z,\omega)/(-e) \qquad [34]$$

where $\chi_\mathrm{p}(k_z,\omega)$ is the longitudinal susceptibility of the electron beam 'plasma'.The beam charge density in the FEL may be quite high, and consequently the space charge field $\tilde{E}_\mathrm{sc}$, arising from the Poison eqn [32], may not be negligible. One should take into consideration then that the total longitudinal force $\tilde{F}_z$ is composed of both the PM-force of eqn [27] and an arising longitudinal space-charge electric force $-e\tilde{E}_\mathrm{sc}$. Thus, one should substitute in eqn [34]:

$$\tilde{F}_z(k_z+k_\mathrm{w},\omega) = -e[\tilde{E}_\mathrm{pm}(k_z+k_\mathrm{w},\omega) + \tilde{E}_\mathrm{sc}(k_z+k_\mathrm{w},\omega)] \qquad [35]$$

and solve it self-consistently with eqns [32] and [33] to obtain the 'external-force' response relation:

$$\tilde{J}_z(k_z+k_\mathrm{w},\omega) = \frac{-\mathrm{i}\omega\chi_\mathrm{p}(k_z+k_\mathrm{w},\omega)}{1+\chi_\mathrm{p}(k_z+k_\mathrm{w},\omega)/\varepsilon}\tilde{E}_\mathrm{pm}(k_z+k_\mathrm{w},\omega) \qquad [36]$$

where we defined the PM 'field': $\tilde{E}_\mathrm{pm} = \tilde{F}_\mathrm{pm}/(-e)$.

In the framework of a single-mode interaction model, we keep in the summation of eqn [30] only one mode q (usually the fundamental mode, and in free space – a Gaussian mode). The transverse current density components in eqn [31] $\tilde{\boldsymbol{J}}_\perp = \frac{1}{2}\tilde{\rho}\,\tilde{v}_w$ are found using eqns [22], [33], and [36]. Finally, substituting $C_q(z) = \tilde{C}_q\,\mathrm{e}^{\mathrm{i}\delta k z}$ (where $\delta k \equiv k_z - k_{qz}$ and $k_z \cong k_{qz}$ is the wavenumber of the radiation wave modified by the interaction with the electrons) results in the general FEL dispersion relation:

$$(k_z - k_{zq})[1+\chi_\mathrm{p}(k_z+k_\mathrm{w},\omega)/\varepsilon] = \kappa\chi_\mathrm{p}(k_z+k_\mathrm{w},\omega)/\varepsilon_0 \qquad [37]$$

Equation [37] is a general expression, valid for a wide variety of FELs, including Cerenkov–Smith–Purcell and TWT. They differ only in the expression for κ. For the conventional (magnetic wiggler) FEL:

$$\kappa = \frac{1}{4}\frac{A_\mathrm{e}}{A_\mathrm{em}}\frac{a_\mathrm{w}^2}{\gamma^2\beta_z^2}\frac{\omega}{c}A_\mathrm{JJ}^2 \qquad [38]$$

where A_e is the cross-section area of the electron beam, and $A_\mathrm{em} \equiv \mathcal{P}_q/[\frac{1}{2}\sqrt{\varepsilon_0/\mu_0}|\varepsilon_{q\perp}(0,0)|^2]$ is the effective area of the interacting radiation-mode q, and it is assumed that the electron beam, passing on axis $(x_\mathrm{e},y_\mathrm{e}) = (0,0)$, is narrow relative to the transverse mode variation $A_\mathrm{e}/A_\mathrm{em} \ll 1$. The 'Bessel-functions coefficient' A_JJ is defined for a linear wiggler only, and is given by:

$$A_\mathrm{JJ} = J_0\left[\frac{a_\mathrm{w}^2}{2(a_\mathrm{w}^2+2)}\right] - J_1\left[\frac{a_\mathrm{w}^2}{2(a_\mathrm{w}^2+2)}\right] \qquad [39]$$

In a helical wiggler, $A_\mathrm{JJ} \equiv 1$. Usually $a_\mathrm{w} \ll 1$, and therefore $A_\mathrm{JJ} \cong 1$.

The Pierce Dispersion Equation

The longitudinal plasma response susceptibility function $\chi_\mathrm{p}(k_z,\omega)$ has been calculated, in any plasma formulation, including fluid model, kinetic model, or

even quantum-mechanical theory. If the electron beam axial velocity spread is small enough (cold beam), then the fluid plasma equations can be used. The small signal longitudinal force equation derived from eqn [25], together with eqn [33] and the small signal current modulation expression:

$$\tilde{J}_z \cong \rho_0 \tilde{v}_z + v_z \tilde{\rho} \qquad [40]$$

result in:

$$\chi_{\mathrm{p}}(k_z, \omega) = -\frac{\omega_{\mathrm{p}}'^2}{(\omega - k_z v_z)^2}\varepsilon \qquad [41]$$

where $\omega_{\mathrm{p}}' = (\mathrm{e}^2 n_0/\gamma\gamma_z^2 \varepsilon m)^{1/2}$ is the longitudinal plasma frequency, n_0 is the beam electrons density ($\rho_0 = -en_0$), v_z is the average axial velocity of the beam.

In this 'cold-beam' limit, the FEL dispersion eqn [37] reduces into the well-known 'cubic dispersion equation' derived first by John Pierce in the late 1940s for the TWT:

$$\delta k(\delta k - \theta - \theta_{\mathrm{pr}})(\delta k - \theta + \theta_{\mathrm{pr}}) = Q \qquad [42]$$

where $\delta k = k_z - k_{zq}$, θ is the detuning parameter (off the synchronism condition of eqn [24]):

$$\theta \equiv \frac{\omega}{v_z} - k_{zq} - k_{\mathrm{w}} \qquad [43]$$

$$\theta_{\mathrm{p}} = \frac{\omega_{\mathrm{p}}}{v_z} \qquad [44]$$

$$Q = \kappa\theta_{\mathrm{p}}^2 \qquad [45]$$

Here $\theta_{\mathrm{pr}} = \mathrm{r_p}\theta_{\mathrm{p}}$, where $r_{\mathrm{p}} < 1$ is the plasma reduction factor. It results from the reduction of the longitudinal space-charge field $\tilde{E}_{\mathrm{sp}}$ in a beam of finite radius r_{b} due to the fringe field effect ($r_{\mathrm{p}} \to 1$ when the beam is wide relative to the longitudinal modulation wavelength: $r_{\mathrm{b}} \gg \lambda_{\mathrm{pm}} = 2\pi/(k_{zq} + k_{\mathrm{w}})$).

The cubic equation has, of course, three solutions δk_i ($i = 1, 2, 3$), and the general solution for the radiation field amplitude and power is thus:

$$C_q(z) = \sum_{j=1}^{3} A_j\, \mathrm{e}^{\mathrm{i}\delta k_j z} \qquad [46]$$

$$P(z) = |C_q(z)|^2 \mathcal{P}_q \qquad [47]$$

The coefficients A_i can be determined from three initial conditions of the radiation and e-beam parameters $C_q(0)$, $\tilde{v}(0)$, $\tilde{i}(0)$, and can be given as a linear combination of them (here $\tilde{i} = A_{\mathrm{e}}\tilde{J}_z$ is the longitudinal modulation current):

$$A_j = A_j^{\mathrm{E}}(\omega)C_q^{\omega}(0) + A_j^{v}(\omega)\tilde{v}^{\omega}(0) + A_j^{i}(\omega)\tilde{i}^{\omega}(0) \qquad [48]$$

Alternatively stated, the exit amplitude of the electromagnetic mode can in general be expressed in terms of the initial conditions:

$$C_q(L) = H^{\mathrm{E}}(\omega)C_q(\omega, 0) + H^{\mathrm{v}}(\omega)\tilde{v}_q(\omega, 0) + H^{\mathrm{i}}(\omega)\tilde{i}(\omega, 0) \qquad [49]$$

where

$$H^{(\mathrm{E,v,i})}(\omega) = \sum_{j=1}^{3} A_j^{(\mathrm{E,v,i})}(\omega)\mathrm{e}^{\mathrm{i}\delta k_j L} \qquad [50]$$

In the conventional FEL, electrons are injected in randomly, and there is no velocity prebunching ($\tilde{v}(\omega, 0) = 0$) or current prebunching ($\tilde{i}(\omega, 0) = 0$) (or equivalently $\tilde{n}(\omega, 0) = 0$). Consequently, $C_q(z)$ is proportional to $C_q(0)$ and one can define and calculate the FEL small-signal single-path gain parameter:

$$G(\omega) \equiv \frac{P(L)}{P(0)} = \frac{|C_q(\omega, L)|^2}{|C_q(\omega, 0)|^2} = |H^{\mathrm{E}}(\omega)|^2 \qquad [51]$$

The FEL Gain Regimes

At different physically meaningful operating regimes, some parameters in eqn [42] can be neglected relative to others, and simple analytic expressions can be found for δk_{i}, A_{i}, and consequently $G(\omega)$. It is convenient to normalize the FEL parameters to the wiggler length: $\bar{\theta} = \theta L$, $\bar{\theta}_{\mathrm{pr}} = \theta_{\mathrm{pr}} L$, $\bar{Q} = QL^3$. An additional figure of merit parameter is the 'thermal' spread parameter:

$$\bar{\theta}_{\mathrm{th}} = \frac{v_{z\mathrm{th}}}{v_z}\frac{\omega}{v_z}L \qquad [52]$$

where $v_{z\mathrm{th}}$ is the axial velocity spread of the e-beam (in a Gaussian velocity distribution model: $f(v_z) = \exp[(v_z - v_{z0})/v_{z\mathrm{th}}]/\sqrt{\pi}\, v_{z\mathrm{th}}$). The axial velocity spread can result out of beam energy spread or angular spread (finite 'emittance'). It should be small enough, so that the general dispersion relation of eqn [37] reduces to eqn [42] (the practical 'cold beam' regime).

Assuming now a conventional FEL ($\tilde{v}_z(0) = 0$, $\tilde{i}(0) = 0$), the single path gain in eqn [51] can be calculated. We present next this gain expression in the different regimes. The maximum values of the gain expression in the different regimes are listed in Table 1.

Low gain

This is the regime where the differential gain in a single path satisfies $G - 1 = [P(L) - P(0)]/P(0) \ll 1$. It is not useful for FEL amplifiers but most FEL oscillators operate in this regime.

Table 1 The gain regimes maximum gain expressions

	Gain regime	*Parameters domain*	*Max. gain expression*
I	Tenuous beam low-gain	$\bar{Q}$, $\bar{\theta}_{pr}$, $\bar{\theta}_{th} < \pi$	$\frac{P(L)}{P(0)} = 1 + 0.27\bar{Q}$
II	Collective low-gain	$\bar{\theta}_{pr} > \frac{\bar{Q}}{2}$, $\bar{\theta}_{th}$, π	$\frac{P(L)}{P(0)} = 1 + \bar{Q}/2\bar{\theta}_{pr}$
III	Collective high-gain	$\bar{Q}/2 > \bar{\theta}_{pr} > \bar{Q}^{1/3}$, $\bar{\theta}_{th}$, $\bar{Q} > \pi$	$\frac{P(L)}{P(0)} = \frac{1}{4}\exp(\sqrt{2\bar{Q}/\bar{\theta}_{pr}})$
IV	Strong coupling high-gain	$\bar{Q}^{1/3} > \bar{\theta}_{pr}$, $\bar{\theta}_{th}$, $\bar{Q} > \pi$	$\frac{P(L)}{P(0)} = \frac{1}{9}\exp(\sqrt{3}\bar{Q}^{1/3})$
V	Warm beam	$\bar{\theta}_{th} > \bar{\theta}_{pr}$, $\bar{Q}^{1/3}$, π	$\frac{P(L)}{P(0)} = \exp(3\bar{Q}/\bar{\theta}_{th}^2)$

The three solutions of eqn [42] – namely the terms of eqn [46] – are reminiscent of the three eigenwaves of the uncoupled system ($\kappa = Q = 0$) : the radiation mode and the two plasma (space-charge) waves of the e-beam (the slow and fast waves, corresponding respectively to the forward and backward propagating plasma-waves in the beam rest reference-frame). In the low-gain regime, all three terms in eqn [46] are significant. Calculating them to first order in κ, results in analytical gain expressions in the collective ($\bar{\theta}_{pr} \gg \pi$) and tenuous-beam ($\bar{\theta}_{pr} \ll \pi$) regimes (note that $\bar{\theta}_{pr}/2\pi = f'_{pr}L/v_z$ is the number of plasma oscillations within the wiggler transit time L/v_z).

In most practical situations the beam current density is small enough, and its energy high enough, to limit operation to the tenuous-beam regime. The gain curve function is then:

$$G(\omega) - 1 = \bar{Q}F(\bar{\theta}(\omega)) = \bar{Q}\frac{\mathrm{d}}{\mathrm{d}\bar{\theta}}\mathrm{sinc}^2(\bar{\theta}/2) \qquad [53]$$

$$\bar{\theta}(\omega) \equiv \theta(\omega)L = 2\pi\frac{\omega - \omega_0}{\Delta\omega_L} \qquad [54]$$

where $\mathrm{sinc}(u) \equiv (\sin u)/u$, and in free space (no waveguide) propagation ($k_{zq} = \omega/c$), the FWHM frequency bandwidth of the $\mathrm{sinc}^2(\bar{\theta}/2)$ function is:

$$\frac{\Delta\omega_L}{\omega_0} = \frac{1}{N_w} \qquad [55]$$

The small signal gain curve is shown in Figure 8. There is no gain at synchronism – $\omega = \omega_0$. Maximum gain – $G - 1 = 0.27\bar{Q}$, is attained at a frequency slightly smaller than ω_0, corresponding to $\bar{\theta} = -2.6$. The small gain curve bandwidth is $\Delta\omega_{SG} \cong \Delta\omega_L/2$,

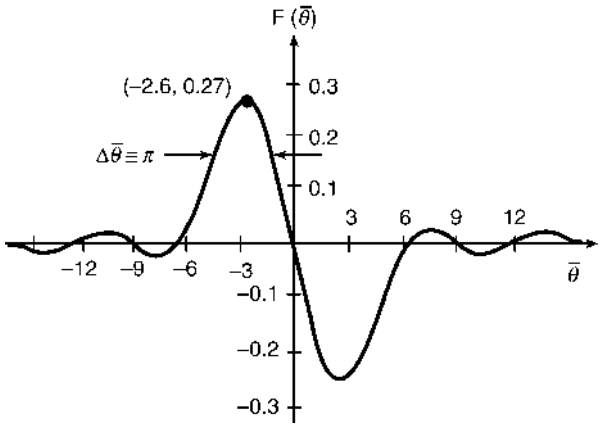

Figure 8 The low-gain cold-beam small-signal gain curve of FEL as a function of the detuning parameter $\bar{\theta}(\omega)$.

namely:

$$\frac{\Delta\omega_{SG}}{\omega_0} = \frac{1}{2N_w} \qquad [56]$$

High gain

This is the regime where the FEL gain in a single path satisfies $G = P(L)/P(0) \gg 1$. It is useful, of course, when the FEL is used as an amplifier.

Since the coefficients of the cubic eqn [42] are all real, the solutions δk_i ($i = 1, 2, 3$) must be either all real, or composed of one real solution δk_3 and two complex solutions, which are complex conjugate of each other: $\delta k_1 = \delta k_2^*$. In the first case, all terms in eqn [46] are purely oscillatory, there is no exponential growth, and the FEL operates in the low gain regime. In the second case, assuming $\mathrm{Im}(\delta k_1) < 0$, $\mathrm{Im}(\delta k_2) > 0$, the first term grows exponentially, and if L is long enough it will dominate over the other decaying ($j = 2$) and oscillatory ($j = 3$) terms, and

result in an exponential gain expression:

$$G(\omega) = \left(\frac{A_1}{A_1 + A_2 + A_3}\right)^2 e^{2(\mathrm{Im}\ \delta k_1)L} \qquad [57]$$

If we focus on the tenuous-beam strong coupling (high–gain) regime $\theta_{\mathrm{pr}} \ll |\delta k|$, then the cubic eqn [42] gets the simple form:

$$\delta k\lfloor(\delta k)^2 - \theta^2\rfloor = \Gamma^3 \qquad [58]$$

where

$$\Gamma = Q^{1/3} = \left(\pi \frac{a_{\mathrm{w}}^2}{\gamma^3\gamma_z^2\beta_z^5}\frac{\omega/c}{A_{\mathrm{em}}}\frac{I_{\mathrm{b}}}{I_{\mathrm{A}}}A_{\mathrm{JJ}}^2\right)^{1/3} \qquad [59]$$

and $I_{\mathrm{A}} = 4\pi\varepsilon_0 m_{\mathrm{e}}c^3/e \cong 17$ kA is the Alfven current. The solution of eqn [58] near synchronism ($\theta \cong 0$) is:

$$\delta k_1 = \frac{1-\sqrt{3}\mathrm{i}}{2}\Gamma, \qquad \delta k_2 = \frac{1+\sqrt{3}\mathrm{i}}{2}\Gamma, \qquad \delta k_3 = -\Gamma \qquad [60]$$

resulting in:

$$H^{\mathrm{E}}(\omega) = \frac{C_q(z)}{C_q(0)} = \frac{1}{3}\left[e^{\frac{\sqrt{3}+\mathrm{i}}{2}\Gamma z} + e^{\frac{-\sqrt{3}+\mathrm{i}}{2}\Gamma z} + e^{-\mathrm{i}\Gamma z}\right] \qquad [61]$$

and for $\Gamma L \gg 1$:

$$G \cong \frac{1}{9}e^{\sqrt{3}\Gamma L} \qquad [62]$$

The FEL gain is then exponential and can be very high. The gain exponential coefficient is characterized then by its third-order root scaling with the current, $\alpha I_{\mathrm{b}}^{1/3}$. The high-gain frequency detuning curve (found by solving eqn [58] to second order in θ) is:

$$G \cong \frac{1}{9}e^{\sqrt{3}\Gamma L}\exp(-\theta^2/\Gamma L 3^{3/2}) \equiv \frac{1}{9}e^{\sqrt{3}\Gamma L}\exp\left[-\frac{(\omega-\omega_0)^2}{\Delta\omega_{\mathrm{HG}}^2}\right] \qquad [63]$$

where $\Delta\omega_{\mathrm{HG}}$ is the 1/e half-width of the gain curve:

$$\frac{\Delta\omega_{\mathrm{HG}}}{\omega_0} = \frac{3^{3/4}}{2\pi}\frac{\lambda_{\mathrm{w}}}{(L/\Gamma)^{1/2}} \qquad [64]$$

Super-Radiance, Spontaneous-Emission and Self Amplified Spontaneous Emission (SASE)

Intense coherent radiation power can be generated in a wiggler or any other radiation scheme without any input radiation signal ($C_q(\omega, 0) = 0$) if the electron beam velocity or current (density) are prebunched. Namely, the injected e-beam has a frequency component $\tilde{v}(\omega)$ or $\tilde{i}(\omega)$ in the frequency range where the radiation device emits. In the case of pure density bunching ($\tilde{v}(\omega) = 0$), the coherent power emitted is found from eqns [46, 47, 49]:

$$(P_q)_{\mathrm{SR}} = P_q|H^i(\omega)|^2|\tilde{i}(\omega, 0)|^2 \qquad [65]$$

A 'prebunched-beam FEL' emits coherent radiation based on the process of Super-radiant Emission (in the sense of Dike). Because all electrons emit in phase radiation wavepackets into the radiation mode, the resultant field amplitude is proportional in this case to the beam current I_{b} and the radiation. By contrast, spontaneous emission from a random electron beam (no bunching) is the result of incoherent superposition of the wavepackets emitted by the electrons and its power is expected to be proportional to the current I_{b}.

When the current to radiation field transfer function $H^i(\omega)$ is known, eqn [65] can be used to calculate the superradiant power, and in the high-gain regime also the amplified-superradiant power. The latter is the amplification of the superradiant radiation in the downstream sections of a long wiggler. Such unsaturated gain is possible only when the beam is partly bunched $\tilde{i}(\omega)(I_{\mathrm{b}})$ (because the FEL gain process requires enhanced bunching).

The expressions for the current to field transfer function, in the superradiant gain and the high-gain amplified superradiance limits respectively, are:

$$|H^i(\omega)| = \frac{(P_{\mathrm{pb}}/P_q)^{1/2}}{I_{\mathrm{b}}}\mathrm{sinc}(\theta L/2) \qquad [66]$$

$$\left|H^i(\omega)\right| = \frac{(P_{\mathrm{pb}}/P_q)^{1/2}}{3\Gamma L I_{\mathrm{b}}}e^{(\sqrt{3}/2)\Gamma L}\ e^{-(\omega-\omega_0)^2/2(\Delta\omega)_{\mathrm{HG}}^2} \qquad [67]$$

where

$$P_{\mathrm{pb}} = \frac{I_{\mathrm{b}}^2 Z_q}{32}\left(\frac{a_{\mathrm{w}}}{\gamma\beta_z}\right)^2\frac{L^2}{A_{\mathrm{em}}} \qquad [68]$$

Z_q is the radiation mode impedance (in free-space $Z_q = \sqrt{\mu_0/\varepsilon_0}$). From these expressions one can calculate the power and spectral power of both coherent (superradiant) and partially coherent (spontaneous emission) radiation of FEL in the negligible gain and high gain regimes. The corresponding super-radiant power is in the negligible superradiance gain limit:

$$P_{\mathrm{SR}} = P_{\mathrm{pb}}\left|\frac{\tilde{i}(\omega)}{I_{\mathrm{b}}}\right|^2\mathrm{sinc}^2(\theta L/2) \qquad [69]$$

(proportional, as expected, to the modulation current squared) and in the high-gain amplified superradiance limit (assuming initial partial bunching $|i(\omega)/I\mathrm{b}| \ll 1$):

$$P_{\mathrm{SR}} = P_{\mathrm{pb}} \left| \frac{\tilde{i}(\omega)}{I_{\mathrm{b}}} \right|^2 \frac{1}{9(\Gamma L)^2} \mathrm{e}^{\sqrt{3}\Gamma L}\, \mathrm{e}^{-(\omega-\omega_0)^2/(\Delta\omega)^2_{\mathrm{HG}}} \qquad [70]$$

The discussion is now extended to incoherent (or partially coherent) spontaneous emission. Due to its particulate nature, every electron beam has random frequency components in the entire spectrum (shot noise). Consequently, incoherent radiation power is always emitted from an electron-beam passing through a wiggler, and its spectral-power can be calculated through the relation:

$$\frac{\mathrm{d}P_q}{\mathrm{d}\omega} = \frac{2}{\pi} \mathcal{P}_q |H^i(\omega)|^2 \frac{\langle |\tilde{i}(\omega)|^2 \rangle}{T} \qquad [71]$$

Here $\tilde{i}(\omega)$ is the Fourier transform of the current of randomly injected electrons $i(t) = -e\sum_{j=1}^{N_T} \delta(t - t_{oj})$, where N_T is the average number of electrons in a time period T, namely, the average (DC) current is $I_{\mathrm{b}} = -eN_T/T$. For a randomly distributed beam, the shot noise current is simply $\langle |i(\omega)|^2 \rangle / T = eI_{\mathrm{b}}$, and therefore the spontaneous emission power of the FEL, which is nothing but the 'synchrotron-undulator radiation', is given by (see eqn [66]):

$$\frac{\mathrm{d}P_q}{\mathrm{d}\omega} = \frac{1}{16\pi} eI_{\mathrm{b}} Z_q \frac{L^2}{A_{\mathrm{em}}} \left(\frac{a_{\mathrm{w}}}{\gamma\beta_z} \right)^2 \mathrm{sinc}^2(\theta L/2) \qquad [72]$$

If the wiggler is long enough, the spontaneous emission emitted in the first part of the wiggler can be amplified by the rest of it (SASE). In the high-gain limit (see eqn [67]), the amplified spontaneous emission power within the gain bandwidth of eqn [64] is given by:

$$P_q = \frac{2}{\pi} \mathcal{P}_q eI_{\mathrm{b}} \int_0^\infty |H^i(\omega)|^2 \mathrm{d}\omega = \frac{1}{9} P_{\mathrm{sh}}\, \mathrm{e}^{\sqrt{3}\Gamma L} \qquad [73]$$

where P_{sh} is an 'effective shot-noise input power':

$$P_{\mathrm{sh}} = \frac{2}{\sqrt{\pi}} \frac{eP_{\mathrm{pb}}}{I_0(\Gamma L)^2} (\Delta\omega)_{\mathrm{HG}} \qquad [74]$$

Saturation Regime

The FEL interaction of an electron with an harmonic electromagnetic (EM) wave is essentially described by the longitudinal component of the force in eqn [25], driven by the pondermotive force of eqns [27] and [28]:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma_i m v_{zi}) = |\tilde{F}_{\mathrm{pm}}| \cos[\omega t - (k_z + k_{\mathrm{w}}) z_i] \qquad [75]$$

$$\mathrm{d}z_i/\mathrm{d}t = v_{zi} \qquad [76]$$

As long as the interaction is weak enough (small signal regime), the change in the electron velocity is negligible – $v_{zi} \cong v_{z0}$, and the phase of the force-wave, experienced by the electron, is linear in time $\Psi_i(t) = [\omega - (k_z + k_{\mathrm{w}})v_{z0}](t - t_{0i}) + \omega t_{0i}$. Near synchronism condition $\theta \cong 0$ (eqn [24]), eqn [75] results in bunching of the beam, because different acceleration/deceleration forces are applied on each electron, depending on their initial phase $\Psi_i(0) = \omega t_{0i}(-\pi < \Psi_i(0) < \pi)$ within each optical period $2\pi/\omega$ (see Figure 7). Taylor expansion of v_{zi} around v_{z0} in eqns [75] and [76], and use of conservation of energy between the e-beam and the radiation field, would lead again to the small signal gain expression eqn [53] in the low gain regime.

When the interaction is strong enough (the non-linear or saturation regime), the electron velocities change enough to invalidate the assumption of linear time dependence of Ψ_i and the nonlinear set of eqns [75] and [76] needs to be solved exactly.

It is convenient to invert the dependence on time $z_i(t) = \int_{t_{0i}}^{t} v_{zi}(t')\mathrm{d}t'$, and turn the coordinate z to the independent variable $t_i(z) = \int_0^z \mathrm{d}z'/v_{zi}(z') + t_{0i}$. This, and direct differentiation of $\gamma_i(v_{zi})$, reduces eqns [75] and [76] into the well-known pendulum equation:

$$\frac{\mathrm{d}\theta_i}{\mathrm{d}z} = K_{\mathrm{s}}^2 \sin \Psi_i \qquad [77]$$

$$\frac{\mathrm{d}\Psi_i}{\mathrm{d}z} = \theta_i \qquad [78]$$

where

$$\Psi_i = \int_0^z (\omega/v_{zi}(z') - k_z - k_{\mathrm{w}})\, \mathrm{d}z' \qquad [79]$$

$$\theta_i = \frac{\omega}{v_{zi}} - k_z - k_{\mathrm{w}} \qquad [80]$$

are respectively the pondermotive potential phase and the detuning value of electron i at position z.

$$K_{\mathrm{s}} = \frac{k\sqrt{a_{\mathrm{w}} a_{\mathrm{s}} A_{\mathrm{JJ}}}}{\gamma_0 \gamma_{z0} \beta_{z0}^2} \qquad [81]$$

is the synchrotron oscillation wavenumber, where a_{w} is given in eqn [9], $a_{\mathrm{s}} = e|\tilde{E}|\omega/mc$, and $\gamma_0 = \gamma(0)$, $\gamma_{z0} = \gamma_z(0)$, and $\beta_{z0} = \beta_z(0)$ are the initial parameters of the assumed cold beam.

The pendulum eqns [77] and [78] can be integrated once, resulting in:

$$\tfrac{1}{2}\theta_i^2(z) - K_{\mathrm{s}}^2 \cos \Psi_i(z) = C_i \qquad [82]$$

and the integration constant is determined for each electron by its detuning and phase relative to the

pondermotive wave at the entrance point ($z = 0$): $C_i = \frac{1}{2}\theta_i^2(0) - K_s^2 \cos \Psi_i(0)$.

The $\theta(z)$, $\Psi(z)$ phase-space trajectories of eqn [82] are shown in Figure 9 for various values of C_i (corresponding to the initial conditions $\theta_i(0)$, $\Psi_i(0)$). The trajectories corresponding to $|C_i| > K_s^2$ are open; namely electrons on these trajectories, while oscillating, can slip-off out of the pondermotive–potential wave period to adjacent periods, ahead or backward, depending on the value of their detuning parameter $\bar{\theta}$. The trajectories corresponding to $|C_i| < K_s^2$ are closed, namely the electrons occupying these trajectories are 'trapped', and their phase displacement is bound to a range $|\Psi_i(z) - n\pi| < \Psi_{im} \equiv \arccos(|C_i|/K_s^2) < \pi$ within one pondermotive-wave period. The trajectory $C_i = K_s^2$ defines the 'separatrix':

$$\theta_i(z) = \pm 2K_s \cos(\Psi_i/2) \qquad [83]$$

which is sometimes referred to as the 'trap' or 'bucket'. Every electron within the separatrix stays trapped, and the ones out of it are free (untrapped). The height of the separatrix (maximum detuning swing) is $\Delta\theta_{trap} = 4K_s$. The oscillation frequency of the trapped electrons can be estimated for deeply trapped electrons ($\Psi_m \ll 2\pi$). In this case the physical pendulum eqns [77] and [78] reduce to the mathematical pendulum equation with an oscillation frequency K_s, in the z coordinate. This longitudinal oscillation, called 'synchrotron oscillation', takes place as a function of time at the 'synchrotron frequency' $\Omega_s = K_s v_z$.

Differentiation of $\theta_i(v_{zi})$ and $v_{zi}(\gamma_i)$ permits to describe the phase-space dynamics in terms of the more physical parameters $\delta v_{zi} = v_{zi} - v_{ph}$ and

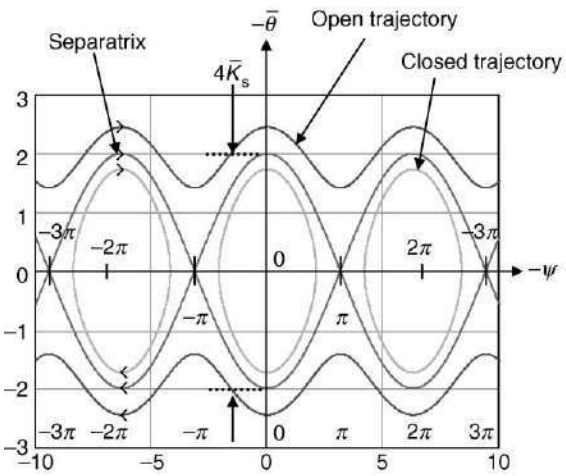

Figure 9 The (θ–Ψ) phase-space trajectories of the pendulum equation.

$\delta\gamma_i = \gamma_i - \gamma_{ph}$, where:

$$v_{ph} = \frac{\omega}{k_z + k_w} \qquad [84]$$

is the phase velocity of the pondermotive wave and $\gamma_{ph} \equiv (1 - \beta_{ph}^2)^{-1/2}$:

$$-\theta_i = \frac{\omega}{c^2\beta_{z0}^2}\delta v_{zi} = \frac{k}{\beta_{z0}^3\gamma_{z0}^2\gamma_0}\delta\gamma_i \qquad [85]$$

Figure 10 displays a typical dynamics of electron beam phase-space (γ, Ψ) evolution for the case of a cold beam of energy $\gamma(0)$ entering the interaction region at $z = 0$ with uniform phase distribution (random arrival times t_{0i}). The FEL is assumed to operate in the low-gain regime (typical situation in an FEL oscillator), and, therefore, the trap height (corresponding to $\Delta\theta_{trap} = 4k_s$):

$$\Delta\gamma_{trap} = 8\beta_z^2\gamma_z^2\gamma K_s/k \qquad [86]$$

remains constant along the interaction length. Figure 10a displays the e-beam phase-space evolution in the small signal regime. The uniform phase distribution evolves along the wiggler into a bunched distribution (compare to Figure 7c), and its average kinetic energy goes down $(\Delta E_k) = [\langle\gamma_i(L)\rangle - \gamma(0)]mc^2 < 0$, contributing this energy to the field of the interacting radiation mode, $\Delta P_q = (\Delta E_k)I_0/e$. In this case (corresponding in an FEL oscillator to the early stages of oscillation build-up), the electrons remain free (untrapped) along the entire length L.

Figure 10b displays the e-beam phase-space evolution in the large signal (saturation) regime (in the case of an oscillator – at the steady-state saturation stage). Part of the electrons are found inside the trap, immediately upon entering the interaction region ($z = 0$), and they lose energy of less than (but near) $mc^2\Delta\gamma_{trap}$ as they pass through the interaction region ($z = L$). A portion of the electrons remain outside the traps upon entrance. They follow open trajectories and lose less energy or may even become accelerated due to their interaction with the wave.

It can be appreciated from this discussion that a good design strategy in attempting to extract maximum power from the electron beam in the FEL interaction, is to set the parameters determining the synchrotron oscillation frequency K_s in eqn [81] so that only half a synchrotron oscillation period will be performed along the interaction length:

$$K_s L = \pi \qquad [87]$$

This is controlled in an amplifier by keeping the input radiation power $P_q(0)$ (and consequently a_s) small enough, so that K_s will not exceed the value set

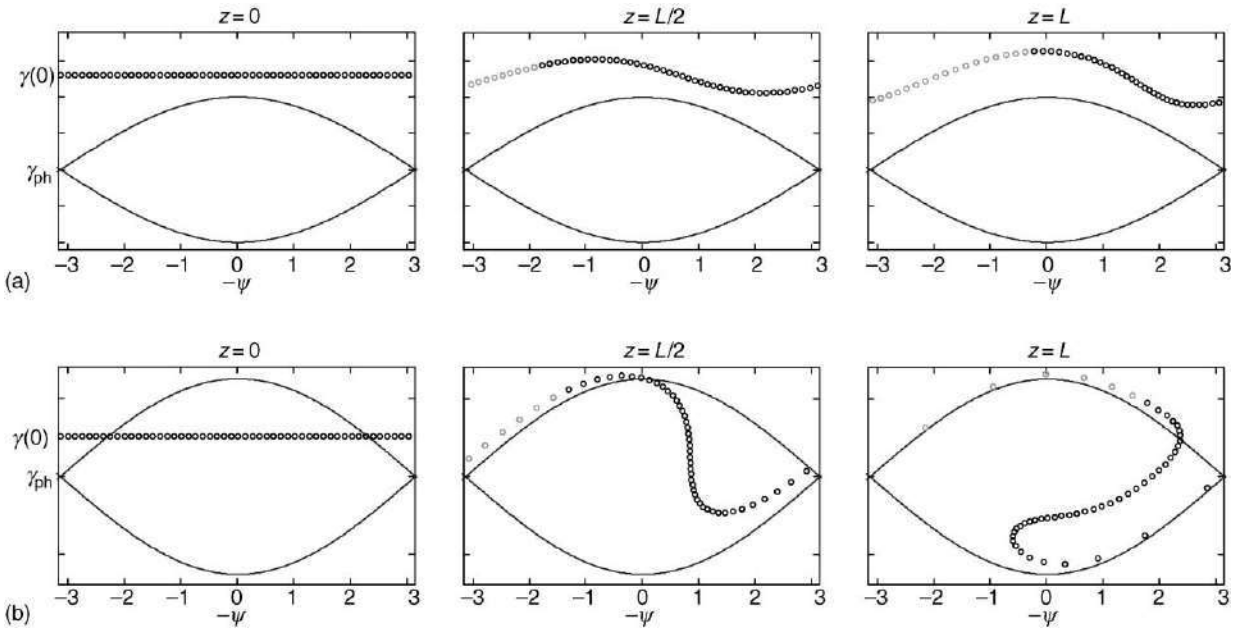

Figure 10 'Snapshots' of the (γ–Ψ) phase-space distribution of an initially uniformly distributed cold beam relative to the PM-wave trap at three points along the wiggler (a) Moderate bunching in the small-signal low gain regime. (b) Dynamics of electron beam trapping and synchrotron oscillation at steady state saturation stage of a FEL oscillator ($K_s L = \pi$).

by eqn [87]. In an oscillator, this is controlled by increasing the output mirror transmission sufficiently, so that the single path incremental small signal gain G-1 will not be much larger than the round trip loss, and the FEL will not get into deep saturation. When the FEL is over-saturated ($K_s L > \pi$), the trapped electrons begin to gain energy as they continue to rotate in their phase-space trajectories beyond the lowest energy point of the trap, and the radiative energy extraction efficiency drops down.

A practical estimate for the FEL saturation power emission and radiation extraction efficiency can be derived from the following consideration: the electron beam departs from most of its energy during the interaction with the wave, if a significant fraction of the electrons are within the trap and have positive velocity δv_{zi} relative to the wave velocity $v_{\rm ph}$ at $z = 0$, and if at the end of the interaction length ($z = L$), they complete half a pendulum swing and reverse their velocity relative to the wave $\delta v_{zi}(L) \cong -\delta v_{zi}(0)$. Correspondingly, in the energy phase-space diagram (Figure 10b) the electrons perform half a synchrotron oscillation swing and $\delta\gamma_i(L) = \gamma_i(L) - \gamma_{\rm ph} = -\delta\gamma_i(0)$. In order to include in this discussion also the FEL amplifier (in the high gain regime), we note that in this case the phase velocity of the wave $v_{\rm ph}$ in eqn [84], and correspondingly $\gamma_{\rm ph}$, are modified by the interaction contribution to the radiation wavenumber – $k_z = k_{z0} + {\rm Re}(\delta k)$, and also the electron detuning parameter (relative to the pondermotive wave) θ_i in eqn [80] differs from the beam detuning parameter θ in eqn [43]: $\theta_i = \theta - {\rm Re}(\delta k)$. Based on these considerations and eqn [85], the maximum energy extraction from the beam in the saturation process is:

$$\Delta\gamma = 2\delta\gamma_i(0) = 2\beta_{z0}^3\gamma_{z0}^2\gamma_0\frac{{\rm Re}\,\delta k - \theta}{k} \qquad [88]$$

where θ is the initial detuning parameter in eqn [43].

In an FEL oscillator, operating in general in the low-gain regime, $|{\rm Re}\,\delta k| \ll |\theta|$, oscillation will start usually at the resonator mode frequency, corresponding to the detuning parameter $\theta(\omega) = -2.6/L$, for which the small signal gain is maximal (see Figure 8). Then the maximum radiation extraction efficiency can be estimated directly from eqn [88]. It is, in the highly relativistic limit ($\beta_{z0} \cong 1$):

$$\eta_{\rm ext} = \frac{\Delta\gamma}{\gamma_0} \cong \frac{1}{2N_{\rm w}} \qquad [89]$$

In an FEL amplifier, in the high-gain regime ${\rm Re}\,\delta k = \Gamma/2 \gg |\theta|$, and consequently in the same limit:

$$\eta_{\rm ext} \cong \frac{\Gamma\lambda_{\rm w}}{4\pi} \qquad [90]$$

It may be interpreted that the effective wiggler length for saturation is $L_{\rm eff} = 2\pi/\Gamma$.

Equation [90], derived here for a coherent wave, is considered valid also for estimating the saturation

efficiency also in SASE-FEL. In this context, it is also called 'the efficiency parameter' $-\rho$.

FEL Radiation Schemes and Technologies

Contrary to conventional atomic and molecular lasers, the FEL operating frequency is not determined by natural discrete quantum energy levels of the lasing matter, but by the synchronism condition of eqn [24] that can be predetermined by the choice of wiggler period, $\lambda_w = 2\pi/k_w$, the resonator dispersion characteristics $k_{zq}(\omega)$, and the beam axial velocity v_z.

Because the FEL design parameters can be chosen at will, its operating frequency can fit any requirement, and furthermore, it can be tuned over a wide range (primarily by varying v_z). This feature of FEL led to FEL development efforts in regimes where it is hard to attain high-power tunable conventional lasers or vacuum-tube radiation sources – namely in the sub-mm (far infrared or THz) regimes, and in the VUV down to soft X-ray wavelengths.

In practice, in an attempt to develop short wavelength FELs, the choice of wiggler period λ_w is limited by an inevitable transverse decay of the magnetic field away from the wiggler magnets surface (a decay range of $\approx k_w^{-1}$) dictated by the Maxwell equations. To avoid interception of electron beam current on the walls or on the wiggler surfaces, typical wiggler periods are made longer than $\lambda_w > 1$ cm. FELs (or FEMs – free electron masers) operating in the long wavelengths regime (mm and sub-mm wavelengths) must be based on waveguide resonators to avoid excessive diffraction of the radiation beam along the interaction length (the wiggler). This determines the dispersion relation $k_{zq}(\omega) = (\omega^2 - \omega_{coq}^2)^{1/2}/c$ where ω_{coq} is the waveguide cutoff frequency of the radiation mode q. The use of this dispersion relation in eqn [24] results in an equation for the FEL synchronism frequency ω_0. Usually the fundamental mode in an overmoded waveguide is used (the waveguide is overmoded because it has to be wide enough to avoid interception of electron beam current). In this case ($\omega_0 \gg \omega_{co}$) and certainly in the case of an open resonator (common in FELs operating in the optical regime) $k_{zq} = \omega/c$, and the synchronism condition in eqn [24] simplified to the well-known FEL radiation wavelength expression in eqn [6]:

$$\lambda = (1 + \beta_z)\beta_z\gamma_z^2\lambda_w \cong 2\gamma_z^2\lambda_w \qquad [91]$$

where γ_z, a_w are defined in eqns [7]–[9].

To attain strong interaction, it is desirable to keep the wiggler parameter a_w large (eqn [38]), however, if $a_w > 1$, this will cause reduction in the operating wavelength according to eqns [7] and [91]. For this reason, and also in order to avoid harmonic frequencies emission (in case of a linear wiggler), $a_w < 1$ in common FEL design. Consequently, considering the practical limitations on λ_w, the operating wavelength eqn [91] is determined primarily by the beam relativistic Lorentz factor γ (eqn [8]).

The conclusion is that for a short wavelength FEL, one should use an electron beam accelerated to high kinetic energy E_k. Also, tuning of the FEL operating-wavelength can be done by changing the beam energy. Small-range frequency tuning can be done also by changing the spacing between the magnet poles of a linear wiggler. This varies the magnetic field experienced by the e-beam, and effects the radiation wavelength through change of a_w (see eqns [7] and [91]).

Figure 11 displays the operating wavelengths of FEL projects all over the world versus their e-beam energy. FELs were operated or planned to operate over a wide range of frequencies, from the microwave to X-ray – eight orders of magnitude. The data points fall on the theoretical FEL radiation curve eqns [7], [8], and [91].

FEL Accelerator Technologies

The kind of accelerator used is the most important factor in determining the FEL characteristics. Evidently, the higher the acceleration energy, the shorter is the FEL radiation wavelength. However, not only the acceleration beam energy determines the shortest operating wavelength of the FEL, but also the e-beam quality. If the accelerated beam has large energy spread, energy instability, or large emittance (the product of the beamwidth with its angular spread), then it may have large axial velocity spread v_{zth}. At high frequencies, this may push the detuning spread parameter $\bar{\theta}_{th}$ (eqn [52]) to the warm beam regime (see Table 1), in which the FEL gain is diminished, and FELs are usually not operated.

Other parameters of the accelerator determine different characteristics of the FEL. High current in the electron beam enables higher gain and higher power operation. The e-beam pulse shape (or CW) characteristics, affect, of course, the emitted radiation waveform, and may also affect the FEL gain and saturation characteristics. The following are the main accelerator technologies used for FEL construction. Their wavelength operating-regimes (eqn [91]) (determined primarily by their beam acceleration energies), are displayed in Figure 12.

Modulators and pulse-line accelerators

These are usually single pulse accelerators, based on high voltage power supplies and fast discharge stored

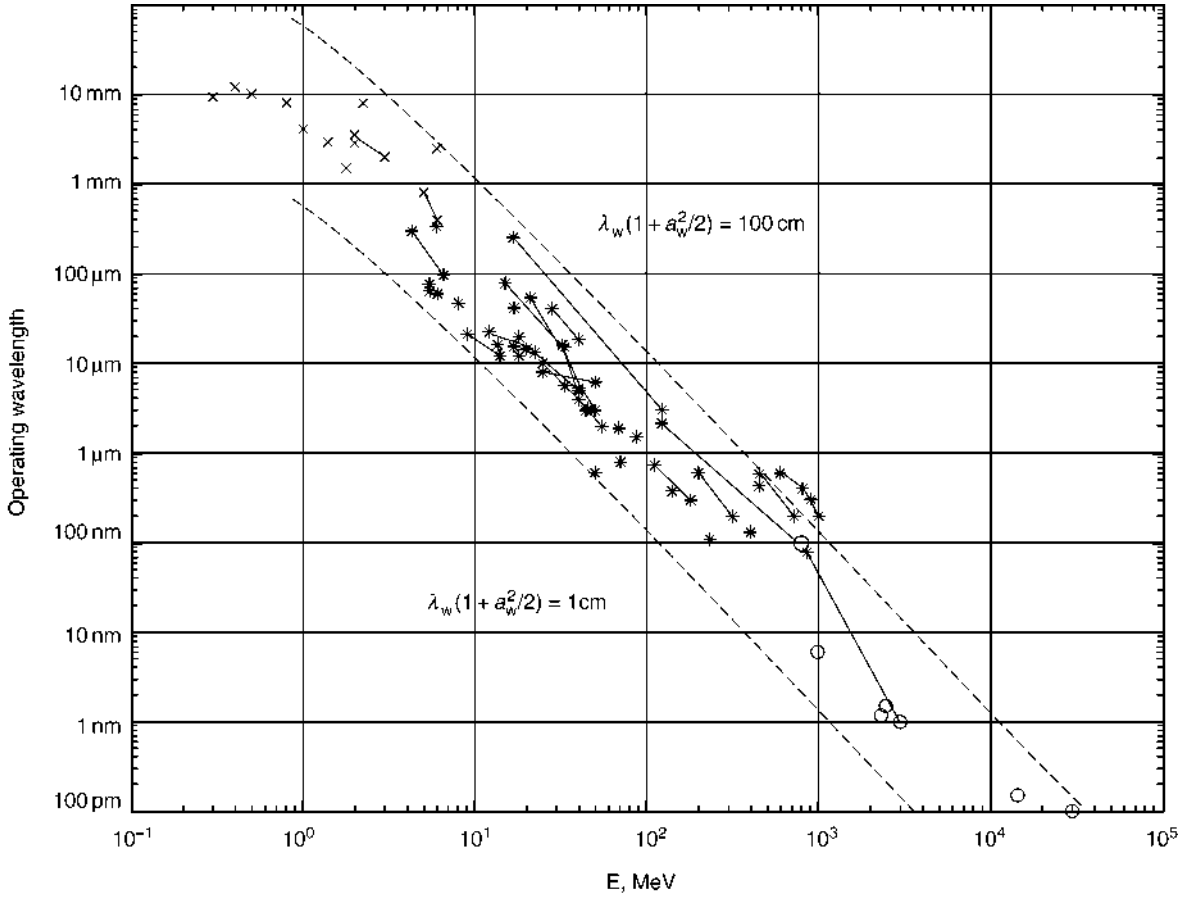

Figure 11 Operating wavelengths of FELs around the world vs. their accelerator beam energy. The data points correspond in ascending order of accelerator energy to the following experimental facilities: NRL (USA), IAP (Russia), KAERI (Korea), IAP (Russia), JINR/IAP (Russia), INP/IAP (Russia), TAU (Israel), FOM (Netherlands), KEK/JAERI (Japan/Korea), CESTA (France), ENEA (Italy), KAERI-FEL (Korea), LEENA (Japan), ENEA (Italy), FIR FEL (USA), mm Fel (USA), UCSB (USA), ILE/ILT (Japan), MIRFEL (USA), UCLA-Kurchatov (USA/Russia), FIREFLY (GB), JAERI-FEL (Japan), FELIX (Netherlands), RAFEL (USA), ISIR (Japan), UCLA-Kurchatov-LANL (USA/RU), ELSA (France), CLIO (France), SCAFEL (GB), FEL (Germany), BFEL (China), KHI-FEL (Japan), FELI4 (Japan), iFEL1 (Japan), HGHG (USA), FELI (USA), MARKIII (USA), ATF (USA), iFEL2 (Japan), VISA (USA), LEBRA (Japan), OK-4 (USA), UVFEL (USA), iFEL3 (Japan), TTF1 (Germany), NIJI-IV (Japan), APSFEL (USA), FELICITAI (Germany), FERMI (Italy), UVSOR (Japan), Super-ACO (France), TTF2 (Germany), ELETTRA (Italy), Soft X-ray (Germany), SPARX (Italy), LCLS (USA), TESLA (Germany). X, long wavelengths; ∗, short wavelengths; ○, planned short wavelengths SASE-FELs. Data based in part on H. P. Freund, V. L. Granatstein, Nucl. Inst. and Methods In Phys. Res. A249, 33 (1999), W. Colson, Proc. of the 24th Int. FEL conference, Argone, III. (ed. K. J. Kim, S. V. Milton, E. Gluskin). The data points fall close to the theoretical FEL radiation condition expression (91) drawn for two practical limits of wiggler parameters.

electric energy systems (e.g., Marx Generator), which produce short pulse (tens of nSec) Intense Relativistic Beam (IRB) of energy in the range of hundreds of keV to few MeV and high instantaneous current (order of kAmp), using explosive cathode (plasma field emission) electron guns. FELs (FEMs), based on such accelerators, operated mostly in the microwave and mm-wave regimes. Because of their poor beam quality and single pulse characteristic, these FELs were, in most cases, operated only as Self Amplified Spontaneous Emission (SASE) sources, producing intense radiation beams of low coherence at instantaneous power levels in the range of 1–100 MW. Because of the high e-beam current and low energy, these FEMs operated mostly in the collective high-gain regime (see Table 1).

Some of the early pioneering work on FEMs was done in the 1970s and 1980s in the US (NRL, Columbia Univ., MIT), Russia (IAP), and France (Echole Politechnique), based on this kind of accelerators.

Induction linacs

These too are single pulse (or low repetition rate) accelerators, based on induction of electromotive potential over an acceleration gap by means of an electric-transformer circuit. They can be cascaded to high energy, and produce short pulse (tens to hundreds

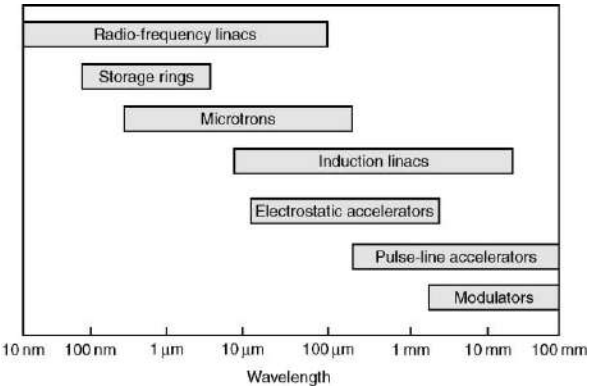

Figure 12 Approximate wavelength ranges accessible with FELs based on current accelerator and wiggler technologies.

of nSec), high current (up to 10 kA) electron beams, with relatively high energy (MeV to tens of MeV). The interest in FELs, based on this kind of accelerator technology, stemmed in the 1980s primarily from the SDI program, for the propose of development of a DEW FEL. The main development of this technology took place on a 50 MeV accelerator – ATA (for operating at 10 μm wavelength) and a 3.5 MeV accelerator – ETA (for operating at 8 mm wavelength). The latter experiment, operating in the high-gain regime, demonstrated record high power (1 GW) and energy extraction efficiency (35%).

Electrostatic accelerators

These accelerators are DC machines, in which an electron beam, generated by a thermionic electron-gun (typically 1–10 Amp) is accelerated electrostatically. The charging of the high voltage terminal can be done by mechanical charge transport (Van de Graaff) or electrodynamically (Crockford–Walton accelerator, Dynamitron). The first kind can be built at energies up to 25 MeV, and the charging current is less than mAmp. The second kind have terminal voltage less than 5 MeV, and the charging current can be hundreds of mAmps.

Because of their DC characteristics, FELs based on these kinds of accelerators can operate at arbitrary pulse shape structure and in principle – continuously (CW). However, because of the low charging current, the high electron beam current (1–10 Amp), required for FEL lasing must be transported without any interception along the entire way from the electron gun, through the acceleration tubes and the FEL wiggler, and then decelerated down to the voltage depressed beam-collector (multistage collector), closing the electric circuit back to the e-gun (current recirculation). The collector is situated at the e-gun potential, biased by moderate voltage high current power supplies, which deliver the current and power needed for circulating the e-beam and compensates for its kinetic energy loss in favor of the radiation field in the FEL cavity. This beam current recirculation is, therefore, also an 'Energy retrieval' scheme, and can make the overall energy transfer efficiency of the electrostatic-accelerator FEL very high.

In practice, high-beam transport efficiency in excess of 99.9% is needed for CW lasing, and has not been demonstrated yet. To avoid HV-terminal voltage drop during lasing, electrostatic-accelerator FELs are usually operated in a single pulse mode. Few FELs of this kind have been constructed. The first and main facility is the UCSB FEL shown in Figure 13. It operates in the wavelength range of 30 μm to 2.5 mm (with three switchable wigglers) in the framework of a dedicated radiation user facility. This FEL operates in the negatively charged terminal mode, in which the e-gun and collector are placed in the negatively charged HV-terminal inside the pressurized insulating gas tank, and the wigglers are situated externally at ground potential. An alternative operating mode of positively charged terminal internal cavity electrostatic-accelerator FEM was demonstrated in the Israeli Tandem–Accelerator FEM and the Dutch F.O.M. Fusion-FEM projects. This configuration enables operating with long pulse, high coherence, and very high average power. Linewidth of $\Delta\omega/\omega \cong 10^{-5}$ was demonstrated in the Israeli FEM and high power (730 kW over few microseconds) was demonstrated in the Dutch FEM, both at mm-wavelengths. The goal of the latter development project (which was not completed) was quasicontinuous operation at 1 MW average power for application in fusion plasma heating.

Radio-frequency (RF) accelerators

RF-accelerators are by far the most popular electron-beam sources for FELs. In RF accelerators, short electron beam bunches (bunch duration 1–10 pSec) are accelerated by the axial field of intense RF radiation (frequency about 1 GHz), which is applied in the acceleration cavities on the injected short e-beam bunches, entering with the accelerating-phase of the RF field. In microtrons, the electron bunches perform circular motion, and get incremental acceleration energy every time they re-enter the acceleration cavity. In RF-LINACs (linear accelerator), the electron bunches are accelerated in a sequence of RF cavities or a slow-wave structure, which keep an accelerating-phase synchronization of the traversing electron bunches with the RF field along a long linear acceleration length. The bunching of the electrons, prior to the acceleration step, is traditionally performed by bunching RF-cavities and a dispersive magnet (chicane) pulse compression system.

The FEL small signal gain, must be large enough to build-up the radiation field in the resonator from noise to saturation well within the macropulse duration.

RF-Linacs are essential facilities in synchrotron radiation centers, used to inject electron beam current into the synchrotron storage ring accelerator from time to time. Because of this reason, many FELs based on RF-LINACs were developed in synchrotron centers, and provide additional coherent radiation sources to the synchrotron radiation center users.

Figure 15 displays FELIX – a RF-LINAC FEL which is located in one of the most active FEL radiation user-centers in FOM – Holland.

Storage rings

Storage rings are circular accelerators in which a number of electron (or positron) beam bunches (typically of 50–500 pS pulse duration and hundreds of ampere peak current) are circulated continuously by means of a lattice of bending magnets and quadrupole lenses. Typical energies of storage ring accelerators are in the hundreds of MeV to GeVs range. As the electrons pass through the bending magnets, they lose a small amount of their energy due to emission of synchrotron radiation. This energy is replenished by a small RF acceleration cavity placed in one section of the ring. The electron beam bunch dimensions, energy spread, and emittance parameters are set in steady state by a balance between the electrons oscillations within the ring lattice and radiation damping due to the random synchrotron emission process. This produces high-quality (small emittance and energy spread) continuous train of electron beam bunches, that can be used to drive a FEL oscillator placed as an insertion device in one of the straight sections of the ring between two bending magnets.

Demonstrations of FEL oscillators, operating in a storage ring, were first reported by the French (LURE-Orsay) in 1987 (at visible wavelengths) and the Russians (VEPP-Novosibirsk) in 1988 (in the ultra-violet). The short wavelength operation of storage-ring FELs is facilitated by the high energy, low emittance and low energy spread parameters of the beam.

Since storage ring accelerators are at the heart of all synchrotron radiation centers, one could expect that FEL would be abundant in such facilities as insertion devices. There is, however, a problem of interference of the FEL operating as an insertion device in the normal operation of the ring itself. The energy spread increase, induced in the electron beam during the interaction in a saturated FEL oscillator, cannot be controlled by the synchrotron radiation damping process, if the FEL operating power is too high.

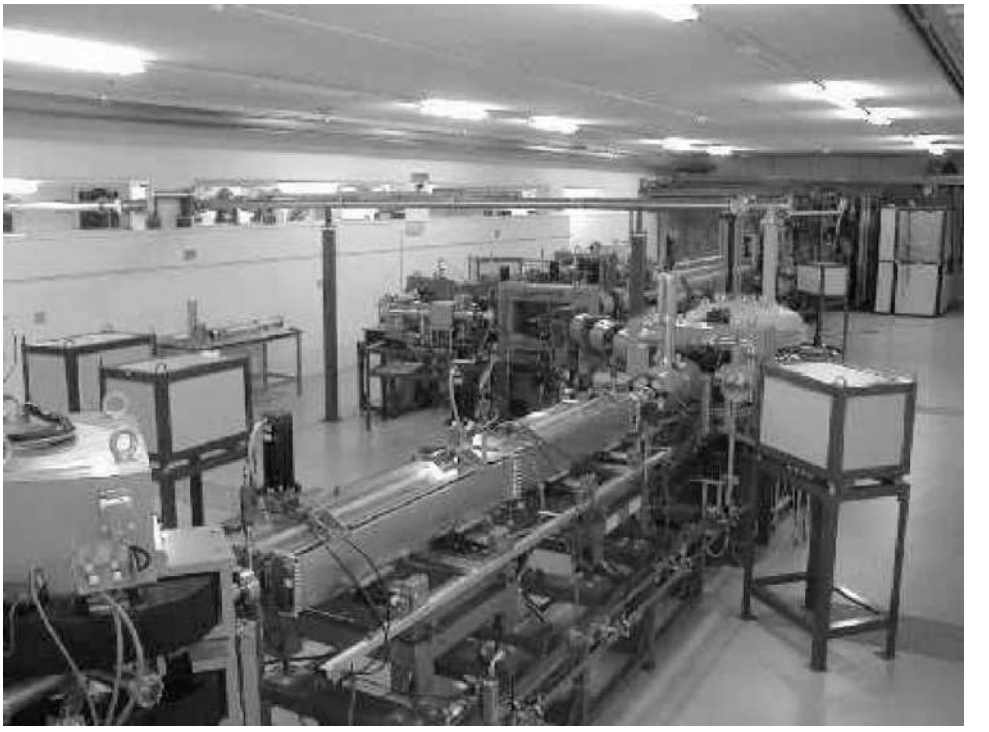

Figure 15 The FELIX RF-Linac FEL operating as a radiation users center in F.O.M. Netherlands. (Courtesy of L. van der Meer, F.O.M.)

This limits the FEL power to be kept as a fraction of the synchrotron radiation power dissipation all around the ring (the 'Renieri Limit'). The effect of the FEL on the e-beam quality, reduces the lifetime of the electrons in the storage ring, and so distrupts the normal operation of the ring in a synchrotron radiation user facility.

To avoid the interference problems, it is most desirable to operate FELs in a dedicated storage ring. This also provides the option to leave long enough straight sections in which long enough wigglers provide sufficient gain for FEL oscillation. Figure 16 displays the Duke storage ring FEL, which is used as a unique radiation user facility, providing intense coherent short wavelength radiation for applications in medicine, biology, material studies, etc.

Superconducting (SC) RF-LINACS

When the RF cavities of the accelerator are superconducting, there are very low RF power losses on the cavity walls, and it is possible to maintain continuous acceleration field in the RF accelerator with a moderate-power continuous RF source, which delivers all of its power to the electron beam kinetic energy. Combining the SC-RF-LINAC technology with an FEL oscillator, pioneered primarily by Stanford University and Thomas Jefferson Lab (TJL) in the US and JAERI Lab in Japan, gave rise to an important scheme of operating such a system in a current recirculating energy retrieval mode. This scheme revolutionized the development of FELs in the direction of high-power, high-efficiency operation, which is highly desirable, primarily for industrial applications (material processing, photochemical production, etc.).

In the recirculating SC-RF-LINAC FEL scheme the wasted beam emerging out of the wiggler after losing a fraction of only few percents (see eqn [89]) out of its kinetic energy, is not dumped into a beam-dump, as in normal cavity RF accelerators, but is re-injected, after circulation, into the SC-RF accelerator. The timing of the wasted electron bunches re-injection is such that they experience a deceleration phase along the entire length of the accelerating cavities. Usually, they are re-injected at the same cell with a fresh new electron bunch injected at an acceleration phase, and thus the accelerated fresh bunch receives its acceleration kinetic energy directly from the wasted beam bunch, that is at the same time decelerated. The decelerated wasted beam bunches are then dumped in the electron beam dump at much lower energy than without recirculation, at energies that are limited primarily just by the energy spread induced in the beam in the FEL laser-saturation process. This scheme, not only increases many folds the over-all energy transformation efficiency from wall-plug to radiation, but would solve significant heat dissipation and radioactivity activation problems in a high-power FEL design.

Figure 17 displays the TJL Infrared SC-RF-LINAC FEL oscillator, that demonstrated for the first time

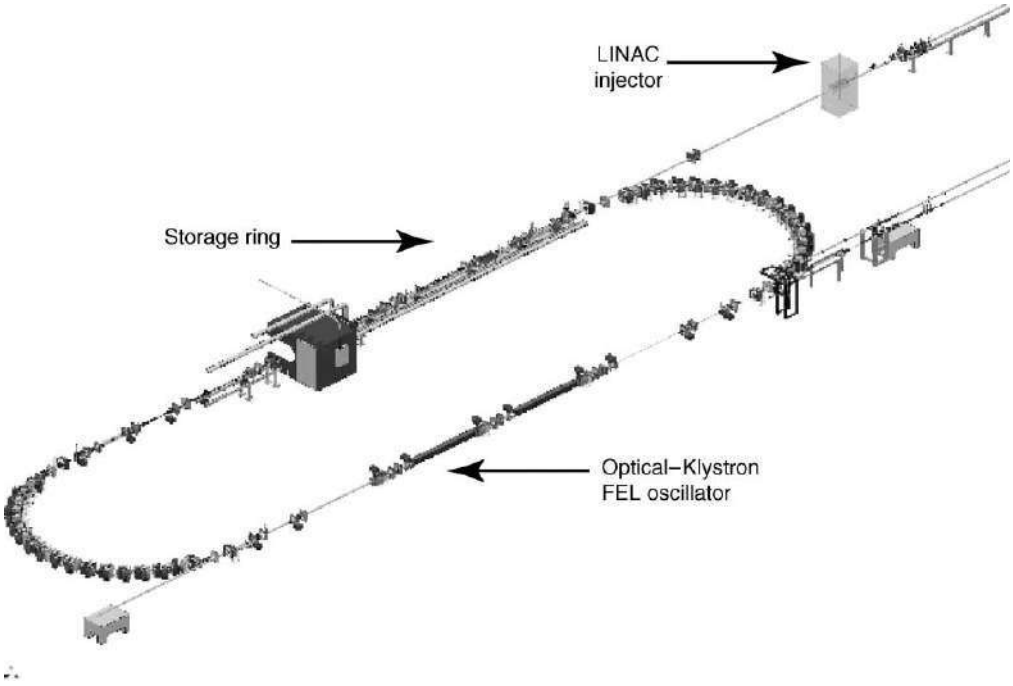

Figure 16 The Duke – University Storage Ring FEL operating as a radiation-users center in N. Carolina, USA. (Mendening: Matthew Busch, courtesy of Glenn Edwards, Duke FEL Lab.)

record high average power levels – nearly 10 kWatt at optical frequencies (1–14 μm). The facility is in upgrade development stages towards eventual operation at 100 kWatt in the IR and 1 kWatt in the UV. It operates in the framework of a laser material processing consortium and demonstrates important material processing applications, such as high-rate micromachining of hard materials (ceramics) with picoSecond laser pulses.

The e-beam current recirculation scheme of SC-RF-LINAC FEL has a significant advantage over the e-beam recirculation in a storage ring. As in electrostatic accelerators, the electrons entering the wiggler are 'fresh' cold-beam electrons from the injector, and not a wasted beam corrupted by the laser saturation process in a previous circulation through the FEL. This also makes it possible to sustain high average circulating current despite the disruptive effect of the FEL on the e-beam. This technological development has given rise to a new concept for a radiation-user facility light-source 4GLS (fourth-generation light source), which is presently in a pilot project development stage at Daresbury Lab in the UK (see Figure 18). In such a scheme, IR and UV FEL oscillators and XUV SASE-FEL can be operated together with synchrotron magnet dipole and wiggler insertion devices without disruptive interference. Such a scheme, if further developed, can give rise to new radiation-user, light-source facilities, that can provide a wider range of radiation parameters than synchrotron centers of previous generation.

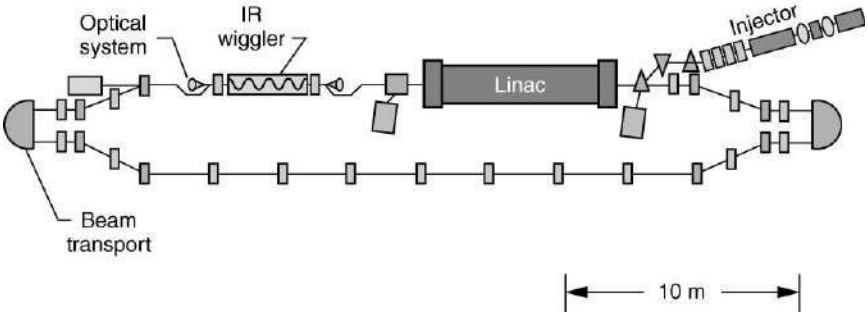

Figure 17 The Thomas Jefferson Lab. recirculating beam-current superconducting Linac FEL operating as a material processing FEL-user center in Virginia USA (Courtesy of S. Benson, Thomas Jefferson Laboratory).

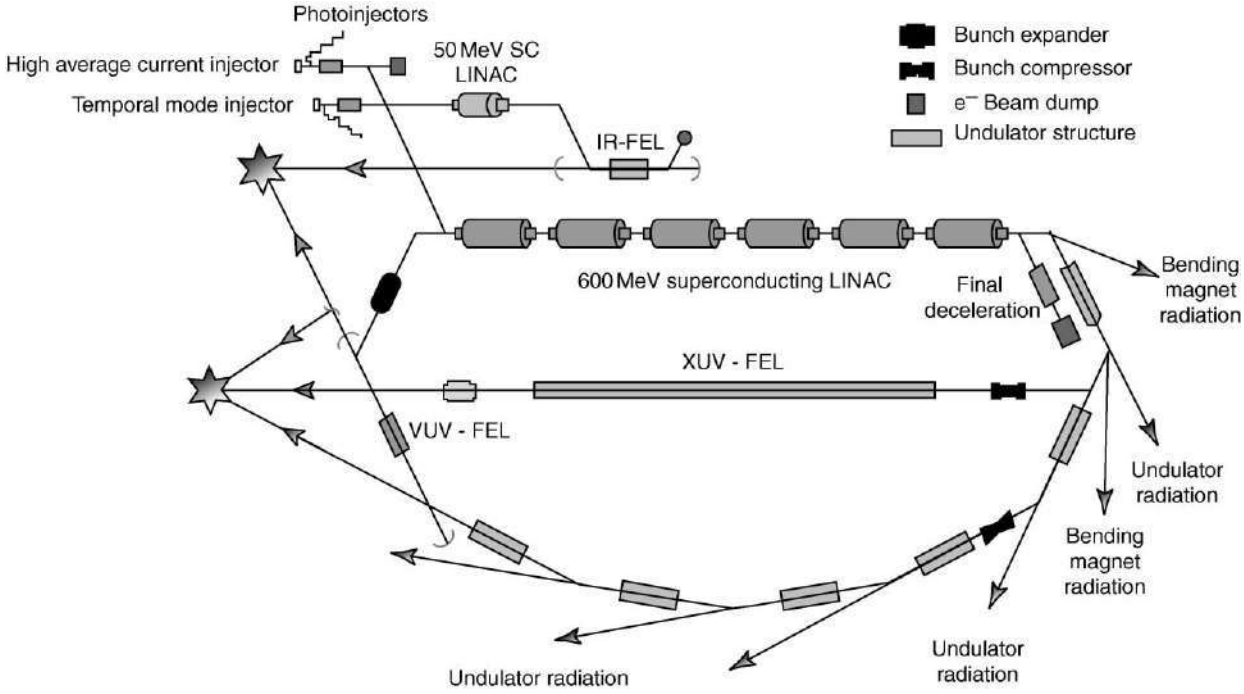

Figure 18 The Daresbury Fourth Generation Light-Source concept (4GLS). The circulating beam-current superconducting Linac includes SASE-FEL, bending magnets and wigglers as insertion devices. (Courtesy of M. Poole, Damesbury Laboratory)

Magnetic Wiggler Schemes

The optical klystron

The stimulated emission process in FEL (see Figure 7c) is based on velocity (energy) bunching of the e-beam in the first part of the wiggler, which turns into density bunching along the central part of the wiggler, and then the density-bunched electron beam performs 'negative work' on the radiation wave and emits radiative energy in the last part of the wiggler. In the OK, these steps are carried out in three separate parts of the wiggler: the energy bunching wiggler section, the dispersive magnet density buncher, and the radiating wiggler section (see Figure 7b).

A schematic of the OK is shown in Figure 19. The chicane magnetic structure in the dispersive section brings all electrons emerging from the bunching wiggler back onto the axis of the radiating wiggler, but provides variable delay $\Delta t_{\mathrm{d}i} = \int_{L_\mathrm{b}}^{L_\mathrm{b}+L_\mathrm{d}} (v_{zi}^{-1} - v_{\mathrm{ph}}^{-1})\, \mathrm{d}z = [\mathrm{d}(\Delta t_\mathrm{d})/\mathrm{d}\gamma]\delta\gamma_i$ relative to the pondermotive wave phase to different electrons, which acquired different energy modulation increments $\delta\gamma_i = \gamma_i - \gamma_{\mathrm{ph}}$ in the final section. The radiation condition is satisfied whenever the bunch-center phase satisfies $\Delta\varphi_\mathrm{d} = \omega\Delta t_\mathrm{d} = \pi/2 - 2m\pi$ (see Figure 7b). However, because the energy dispersion coefficient $\mathrm{d}(\Delta t_\mathrm{d})/\mathrm{d}\gamma$, is much larger in the chicane than in a wiggler of the same length, the density bunching amplitude, and consequently the OK gain, are much larger than in a uniform wiggler FEL of the same length.

The OK was invented by Vinokurov and Skrinsky in 1977 and first demonstrated in 1987 at visible wavelengths in the ACO storage ring of LURE in Orsay, France, and subsequently in 1988 at UV wavelengths, in the VEPP storage ring in Novosibirsk, Russia. The OK is an optimal FEL configuration, if used as an insertion device in a storage ring, because it can provide sufficient gain to exceed the high lasing threshold at the short operating wavelengths of a high-energy storage-ring FEL, and still conform with the rather short straight sections available for insertion devices in conventional synchrotron storage rings. It should be noted that the OK is equivalent to a long wiggler FEL of length L_{eff} of equal gain and therefore its axial velocity spread acceptance is small (this is determined from the cold beam limit $\bar{\theta}_{\mathrm{th}} \ll \pi$ with L_{eff} used in eqn [52]). This too is consistent with storage ring accelerators, which are characterized by small energy spread and emittance of the electron beam.

Radiation emission at harmonic frequencies

In a linear wiggler (eqn [2]), the axial velocity:

$$\beta_z = [\beta^2 - (a_\mathrm{w}/\gamma)^2 \cos^2 k_\mathrm{w} z]^{1/2} \qquad [92]$$

is not constant. It varies with spatial periodicity $\lambda_\mathrm{w}/2$, and in addition to its average value $\bar{\beta}_z = [\beta^2 - a_\mathrm{w}^2/2\gamma^2]^{1/2}$, contains Fourier components of spatial frequencies $2mk_\mathrm{w}$ $(m = 1, 2, \ldots)$. When $a_\mathrm{w} \gg 1$, the axial oscillation deforms the sinusoidal trajectory of the electrons in the wiggler (eqns [22] and [23]), and in a frame of reference moving at the average velocity $\bar{\beta}_z$, the electron trajectories in the wiggling $(x{-}z)$ plane forms an figure 8 shape, rather than a pure transverse linear motion. In the laboratory frame this leads to synchrotron undulator emission in the forward direction at all odd harmonic frequencies of ω_0, corresponding to substitution of $k_\mathrm{w} \to (2m+1)k_\mathrm{w}$ $(m = 1, 2, 3, \ldots)$ in eqn [6]:

$$\omega_{2m+1} = (2m+1)\omega_0 \cong 2\gamma_z^2 c(2m+1)k_\mathrm{w} \qquad [93]$$

All the stimulated emission gain expressions, presented earlier for the fundamental harmonic, are valid with appropriate substitution of

$$\theta_{2m+1} = \frac{\omega}{v_z} - k_z - (2m+1)k_\mathrm{w} \qquad [94]$$

instead of eqn [43], and substitution of the harmonic-weight Bessel-function coefficient of

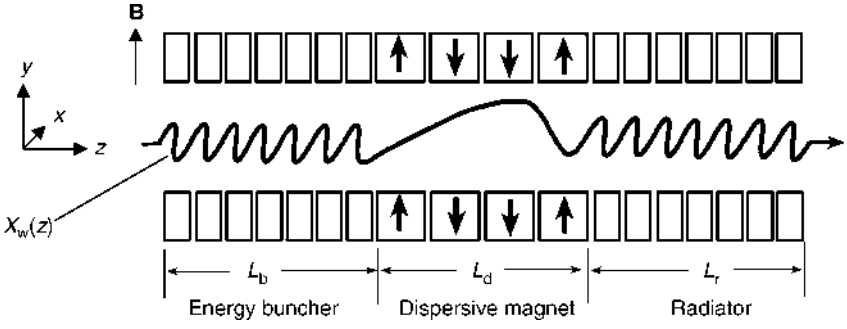

Figure 19 Schematics of the Optical–Klystron, including an energy bunching wiggler, a dispersive magnet bunching section and a radiating wiggler.

synchronism with the beam. Slowing down the PM wave can be done by the gradual increase of the wiggler wavenumber $k_w(z)$ (or decrease of its period $\lambda_w(z)$), so that eqns [29] or [91] keep being satisfied for a given frequency, even if v_z (or γ_z) goes down.

A more correct description of the nonlinear interaction dynamics of the electron beam in a saturated tapered-wiggler FEL is depicted in Figure 20: the electron trap synchronism energy $\gamma_{ph}(z)$ tapers down (by design) along the wiggler, while the trapped electrons are forced to slow down with it, releasing their excess energy by enhanced radiation. An upper limit estimate for the extraction efficiency of such a tapered wiggler FEL would be:

$$\eta_{ext} = \frac{\gamma_{ph}(0) - \gamma_{ph}(L)}{\gamma_{ph}(0)} \qquad [99]$$

and the corresponding radiative power generation would be: $\Delta P = \eta_{ext} I_b E_k / e$. In practice, the phase-space area of the tapered wiggler separatrix is reduced due to the tapering, and only a fraction of the electron beam can be trapped, which reduces correspondingly the practical enhancement in radiative extraction efficiency and power.

An alternative wiggler tapering scheme consists of tapering the wiggler field $B_w(z)$ (or wiggler parameter amplitude $a_w(z)$). If these are tapered down, the axial velocity and axial energy (eqn [7]) can still keep constant (and in synchronism with the PM wave) even if the beam energy γ goes down. Thus, in this scheme, the excess radiative energy extracted from the beam comes out of its transverse (wiggling) energy.

Efficiency and power enhancement of FEL by wiggler tapering have been demonstrated experimentally both in FEL amplifiers (first by Livermore, 1985) and oscillators (first by Los-Alamos, 1983). This elegant way to extract more power from the beam still has some limitations. It can operate efficiently only at a specified high radiation power level for which the tapering was designed. In an oscillator, a long enough untapered section must be left to permit sufficient small signal gain in the early stages of the laser oscillation build-up process.

FEL Oscillators

Most FEL devices are oscillators. As in any laser, in order to turn the FEL amplification process into an oscillation process, one provides a feedback mechanism by means of an optical resonator. In steady state saturation, $GR_{rt} = 1$, where R_{rt} is the round trip reflectivity factor of the resonator and $G = P(L)/P(0)$ is the saturated single-path gain coefficient of the FEL. To attain oscillation, the small signal (unsaturated) gain, usually given by the small gain expression in eqn [53], must satisfy the lasing threshold condition $G > 1/R_{rt}$, as in any laser.

When steady state oscillation is attained, the oscillator output power is:

$$P_{out} = \frac{T}{1 - R_{rt}} \Delta P_{ext} \qquad [100]$$

where $\Delta P_{ext} = \eta_{ext} I_0 (\gamma_0 - 1) mc^2/e$ and η_{ext} is the extraction efficiency, usually given by eqn [89] (low-gain limit).

Usually, FEL oscillators operate in the low-gain regime, in which case $1 - R_{rt} = L + T \ll 1$ (where L is the resonator internal loss factor). Consequently, then $P_{out} \cong \Delta P_{ext} T/(L + T)$, which would give a maximum value, depending on the saturation level of the oscillator. In the general case, one must solve the nonlinear force equations together with the resonator feedback relations of the oscillating radiation mode, in order to maximize the output power (eqn [100]) or efficiency by choice of optimal T for given L.

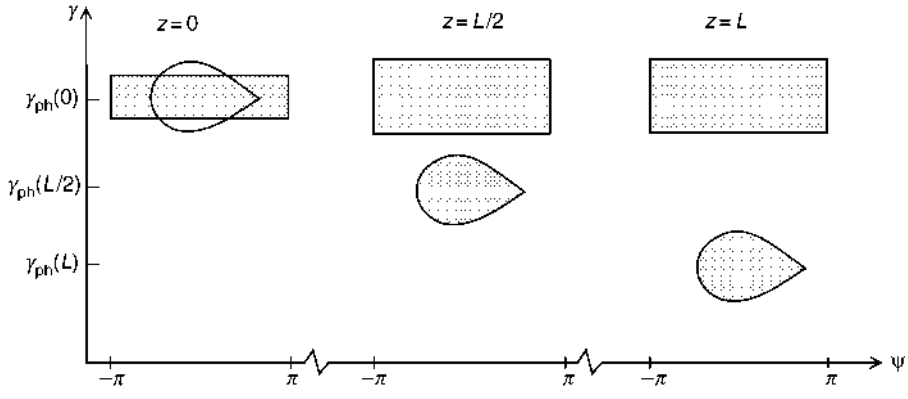

Figure 20 'Snapshots' of the trap at three locations along a tapered wiggler FEL.

In an FEL oscillator operating with periodic electron bunches (as in RF-acclerator based FEL), the solution for the FEL gain and saturation dynamics requires extension of the single frequency solution of the electron and electromagnetic field equations to the time domain. In principle, the situation is similar to that of a mode-locked laser, and the steady state laser pulse train waveform constitutes a superposition of the resonator longitudinal modes that produces a self-similar pulse shape with the highest gain (best overlap with the e-beam bunch along the interaction length). Because the e-beam velocity v_{z0} is always smaller (in an open resonator) than the group velocity of the circulating radiation wavepacket, the radiation wavepacket slips ahead of the electron bunch one optical period λ in each wiggling period (Slippage effect). This reduces the overlap between the radiation pulse and the e-beam bunch along the wiggler (see **Figure 14**) and consequently decreases the gain. Fine adjustment of the resonator mirrors (as shown in **Figure 14**) is needed to attain maximal power and optimal radiation pulse shape. The pulse-slippage gain reduction effect is negligible only if the bunch length is much longer than the slippage length $N_w\lambda$, which can be expressed as:

$$\tau_p \gg 2\pi/\Delta\omega_L \qquad [101]$$

where $\Delta\omega_L$ is the synchrotron undulator radiation frequency bandwidth (eqn [55]). This condition is usually not satisfied in RF-accelerator FELs operating in the IR or lower frequencies, and the

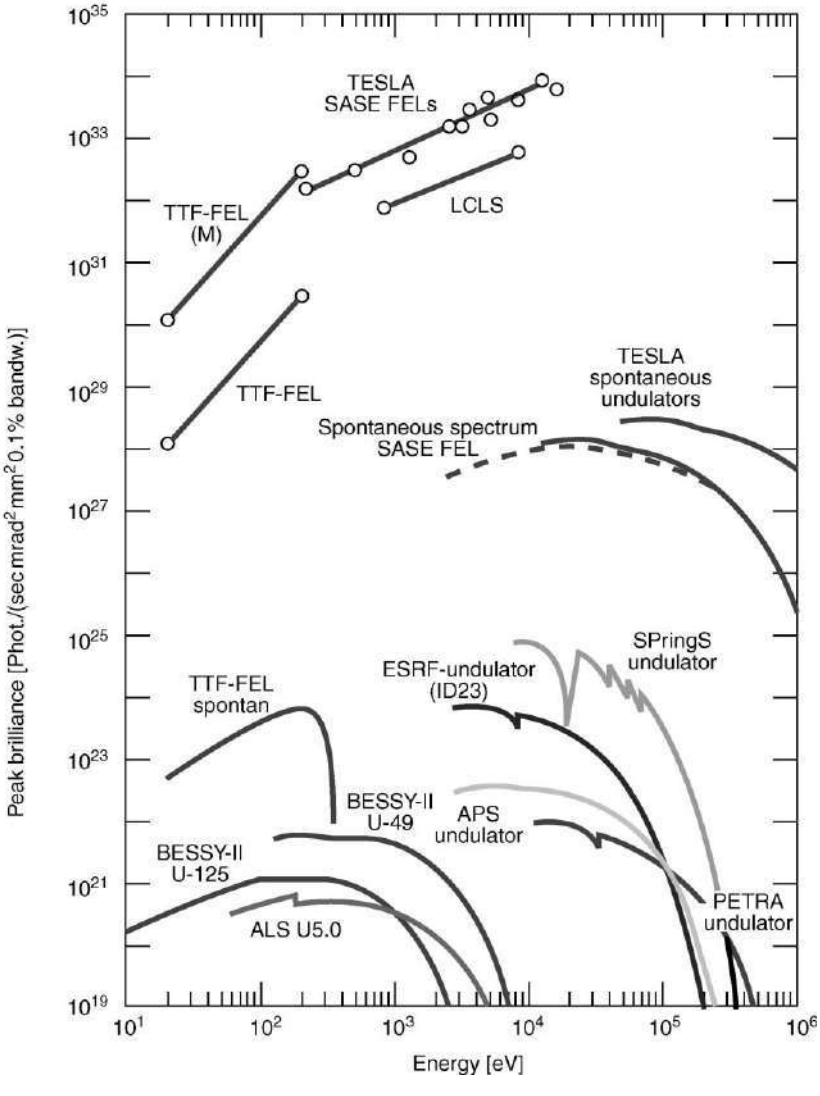

Figure 21 Anticipated peak brightness of SASE FELs (TTF-DESY, LCLS-SLAC) in comparison to the undulators in present third generation Synchrotron Radiation sources. Figure courtesy of DESY, Hamburg, Germany.

slippage effect gain reduction must be then taken into account.

An FEL operating in the cold-beam regime constitutes an 'homogeneous broadening' gain medium in the sense of conventional laser theory. Consequently, the longitudinal mode competition process that would develop in a CW FEL oscillator, leads to single-mode operation and high spectral purity (temporal coherence) of the laser radiation. The minimal (intrinsic) laser linewidth would be determined by an expression analogous to the Schawlow–Towns limit of atomic laser:

$$(\Delta f)_{\text{int}} = \frac{(\Delta f_{1/2})^2}{I_b/e} \qquad [102]$$

where $\Delta f_{1/2}$ is the spectral width of the cold resonator mode. Expression [102] predicts extremely narrow linewidth. In practice, CW operation of FEL was not yet attained, but Fourier transform limited linewidths in the range of $\Delta f/f_0 \cong 10^{-6}$ were measured in long-pulse electrostatic accelerator FELs. In an FEL oscillator, based on a train of e-beam bunches (e.g., an R.F. accelerator beam), the linewidth is very wide and is equal to the entire gain bandwidth (eqn [56]) in the slippage dominated limit, and to the Fourier transform limit $\Delta\omega \cong 2\pi/\tau_p$ in the opposite negligible-slippage limit (eqn [101]). Despite this slippage, it was observed in RF-LINAC FEL that the radiation pulses emitted by the FEL oscillator are phase corrected with each other, and therefore their total temporal coherence length may be as long as the

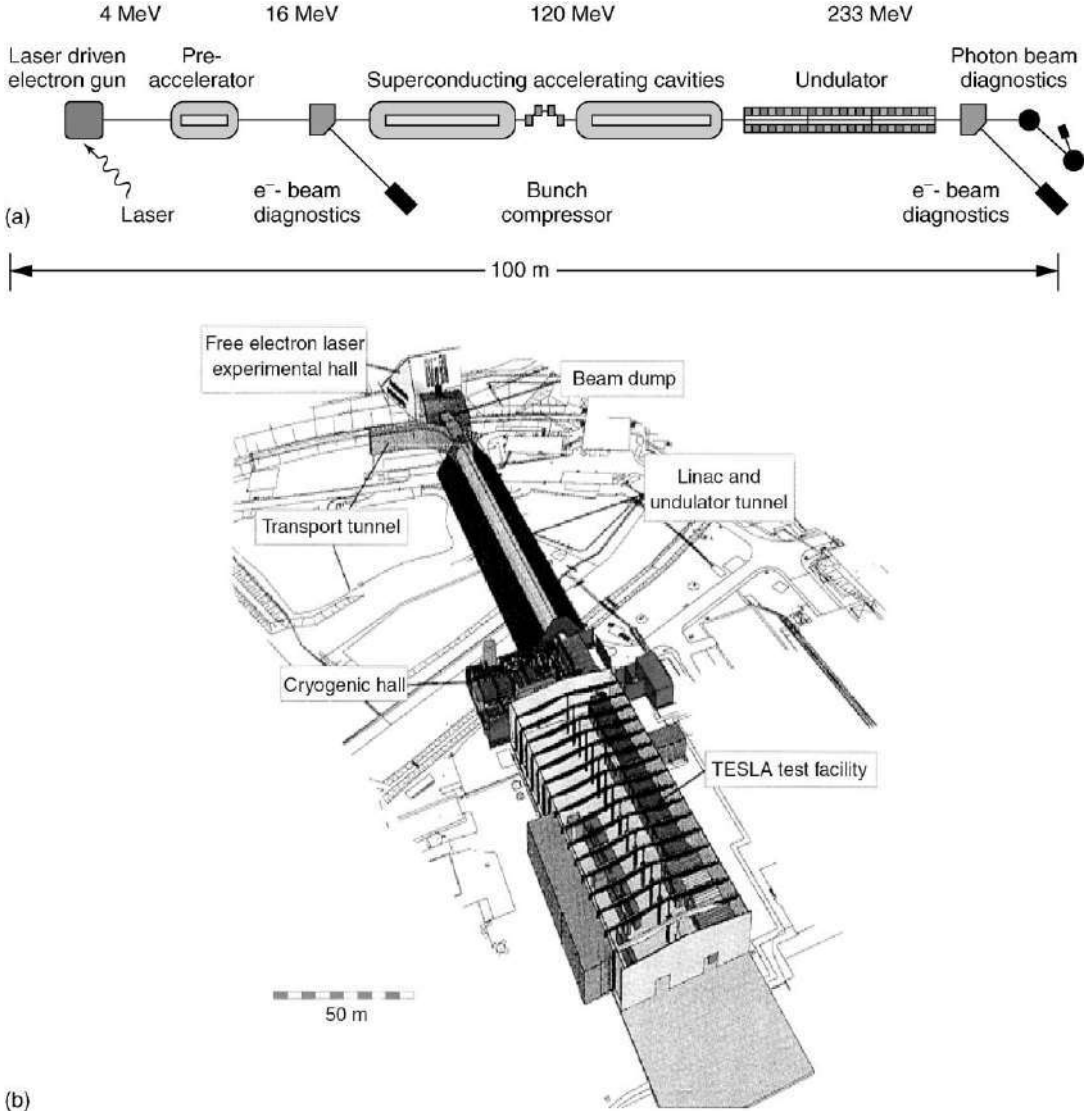

Figure 22 Phase 1 of the SASE FE L (TTF VUV-FEL1): (a) Accelerator layout scheme; (b) General view of the TESLA test facility. Figure courtesy of DESY, Hamburg, Germany.

Metal Vapor Lasers

D W Coutts, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Metal vapor lasers form a class of laser in which the active medium is a neutral or ionized metal vapor usually excited by an electric discharge. These lasers fall into two main subclasses, namely cyclic pulsed metal vapor lasers and continuous-wave metal ion lasers. Both types will be considered in this article, including basic design and construction, power supplies, operating characteristics (including principal wavelengths), and brief reference to their applications.

Self-Terminating Resonance-Metastable Pulsed Metal Vapor Lasers

The active medium in a self-terminating pulsed metal vapor laser consists of metal atoms or ions in the vapor phase usually as a minority species in an inert buffer gas such as neon or helium. Laser action occurs between a resonance upper laser level and a metastable lower laser level (Figure 1). During a fast pulsed electric discharge (typically with a pulse duration of order 100 ns) the upper laser level is preferentially excited by electron impact excitation because it is strongly optically connected to the ground state (resonance transition) and hence has a large excitation cross-section. For a sufficiently large metal atom (or ion) density, the resonance radiation becomes optically trapped, thus greatly extending the lifetime of the upper laser level such that decay from the upper laser level is channelled through the emission of laser radiation to the metastable lower laser level. Lasing terminates when the electron temperature falls to a point such that preferential pumping to the upper laser level is no longer sustained, and the build-up of population in the metastable lower laser level destroys the population inversion. Therefore after each excitation pulse, the resulting excited species in the plasma (in particular the metastable lower laser levels which are quenched by collisions with cold electrons) must be allowed sufficient time to relax and the plasma must be allowed to partially recombine before applying the next excitation pulse. The relaxation times for self-terminating metal vapor lasers correspond to operating pulse repetition frequencies from 2 kHz to 200 kHz.

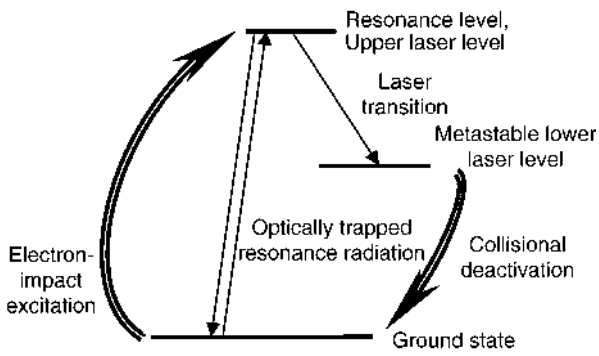

Figure 1 Resonance-metastable energy levels for self-terminating metal vapor lasers.

Many metal vapors can be made to lase in the resonance-metastable scheme and are listed together with their principal wavelengths and output powers in Table 1. The most important self-terminating pulsed metal vapor laser is the copper vapor laser and its variants which will be discussed in detail in the following sections. Of the other pulsed metal vapor lasers listed in Table 1, only the gold vapor laser (principal wavelengths 627.8 nm and 312.3 nm), and the barium vapor laser which operates in the infrared (principal wavelengths 1.5 μm and 2.55 μm) have had any commercial success. All the self-terminating pulsed metal vapor lasers have essentially the same basic design and operating characteristics as exemplified by the copper vapor laser.

Copper Vapor Lasers

Copper vapor lasers (CVLs) are by far the most widespread of all the pulsed metal vapor lasers. Figure 2 shows the energy level scheme for copper. Lasing occurs simultaneously from the $^2P_{3/2}$ level to the $^2D_{5/2}$ level (510.55 nm) and from the $^2P_{1/2}$ level to the $^2D_{3/2}$ level (578.2 nm). Commercial devices are available with combined outputs of over 100 W at 510.55 nm and 578.2 nm (typically with a green-to-yellow power ratio of 2:1).

A typical copper vapor laser tube is shown in Figure 3. High-purity copper pieces are placed at intervals along an alumina ceramic tube which typically has dimensions of 1–4 cm diameter and 1–2 m long. The alumina tube is surrounded by a solid fibrous alumina thermal insulator, and a glass or quartz vacuum envelope. Cylindrical electrodes made of copper or tantalum are located at each end of the plasma tube to provide a longitudinal discharge

Table 1 Principal self-terminating resonance-metastable metal vapor lasers

Metal	*Principal wavelengths (nm)*	*Powers*		*Total efficiencies*	*Pulse repetition frequency (kHz)*	*Technological development*
		Typical (W)	*Maximum (W)*			
Cu	510.55	2–70	2500 total	1%	4–40	Highly developed and commercially available
	578.2	1–50				
Au	627.8	1–8	20	0.13%	2–40	Commercially available
	312	0.1–0.2	1.2		1–8	
Ba	1500	2–10	12	0.5%	5–15	Have been produced commercially, but largely experimental
	2550	1	1.5			
	1130	0.5	1.0			
Pb	722.9		4.4	0.15%	10–30	Experimental
Mn	534.1 (>50%)		12 total	0.32%	~10	Experimental
	1290					

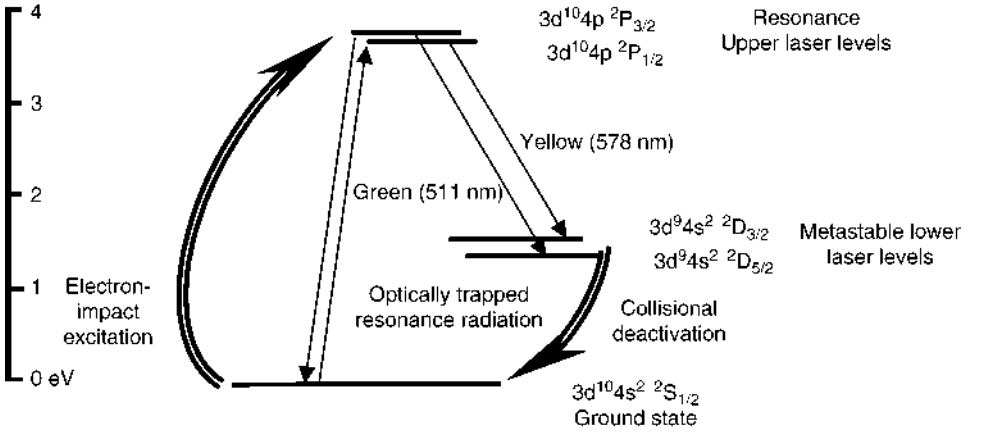

Figure 2 Partial energy level scheme for the copper vapor laser.

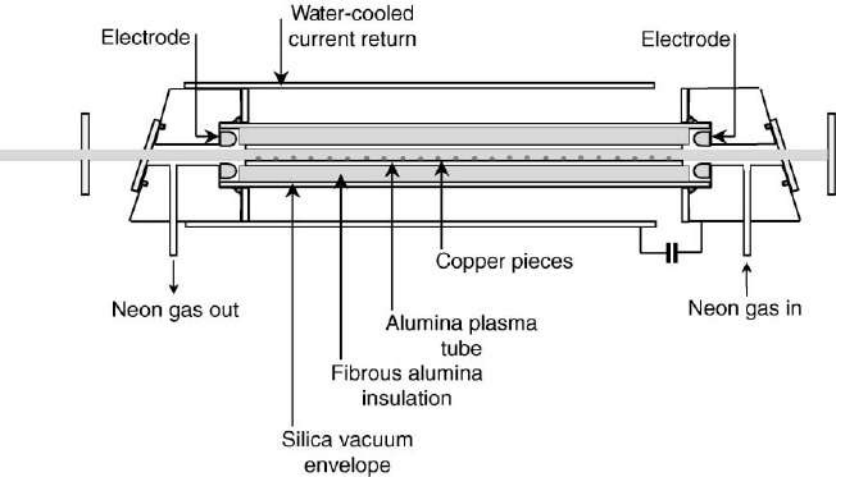

Figure 3 Copper vapor laser tube construction.

arrangement. The cylindrical electrodes and silica laser end windows are supported by water-cooled end pieces. The laser windows are usually tilted by a few degrees to prevent back reflections into the active medium. The laser head is contained within a water cooled metal tube to provide a coaxial current return for minimum laser head inductance. Typically a slow flow (~5 mbar l min^{-1}) of neon at a pressure of 20–80 mbar is used as the buffer gas with an approximately 1% H_2 additive to improve the afterglow plasma relaxation. The buffer gas provides a medium to operate the discharge when the laser is cold and slows diffusion of copper vapor out of the ends of the hot plasma tube. Typical copper fill times are of order 200–2000 hours (for 20–200 g copper load). Sealed-off units with lifetimes of order 1000 hours have been in production in Russia for many years.

During operation waste heat from the repetitively pulsed discharge heats the alumina tube up to approximately 1500 °C at which point the vapor pressure of copper is of approximately 0.5 mbar which corresponds to the approximate density required for maximum laser output power. Typical warm-up times are therefore relatively long at around one hour to full power.

One way to circumvent the requirement for high temperatures required to produce sufficient copper density by evaporation of elemental copper (and hence also reduce warm-up times) is to use a copper salt with a low boiling point located in one or more side-arms of the laser tube (Figure 4). Usually copper halides are used as the salt, with the copper bromide laser being the most successful. For the CuBr laser, a temperature of just 600 °C is sufficient to produce the required Cu density by dissociation of CuBr vapor in the discharge. With the inclusion of 1–2% H_2 in the neon buffer gas, HBr is also formed in the CuBr laser, which has the additional benefit of improving recombination in the afterglow via dissociative attachment of free electrons: $HBr + e^- \rightarrow H + Br^-$, followed by ion neutralization: $Br^- + Cu^+ \rightarrow Br + Cu^*$. As a result of the lower operating temperature and kinetic advantages of HBr, CuBr lasers are typically twice as efficient (2–3%) as their elemental counterparts. Sealed-off CuBr systems with powers of order 10–20 W are commercially produced.

An alternative technique for reducing the operating temperature of elemental CVLs is to flow a buffer gas mixture consisting of ~5% HBr in neon at approximately 50 mbar l min^{-1} and allow this to react with solid copper metal placed within the plasma tube at about 600 °C to produce CuBr vapor *in situ*. The so-called Cu HyBrID (hydrogen bromide in discharge) laser has the same advantages as the CuBr laser (e.g., up to 3% efficiency) but at the cost of requiring flowing highly toxic HBr in the buffer gas, a requirement which has so far prevented commercialization of Cu HyBrID technology.

The kinetic advantages of the hydrogen halide in the CuBr laser discharge can also be applied to a conventional elemental CVL through the addition of small partial pressure of HCl to the buffer gas in addition to the 1–2% H_2 additive. HCl is preferred to HBr as it is less likely to dissociate (the dissociation energy of HCl at 0.043 eV is less than HBr at 0.722 eV). Such kinetic enhancement leads to a doubling in average output power, a dramatic increase in beam quality through improved gain characteristics, and shifts the optimum pulse repetition frequency for kinetically enhanced CVLs (KE-CVLs) from 4–10 kHz up to 30–40 kHz.

Preferential pumping of the upper laser levels requires an electron temperature in excess of the 2 eV range, hence high voltage (10–30 kV), high current (hundreds of A), short (75–150 ns) excitation pulses are required for efficient operation of copper vapor lasers. To generate such excitation pulses CVLs are typically operated with a power supply incorporating a high-voltage thyratron switch. In the most basic configuration, the charge-transfer circuit shown

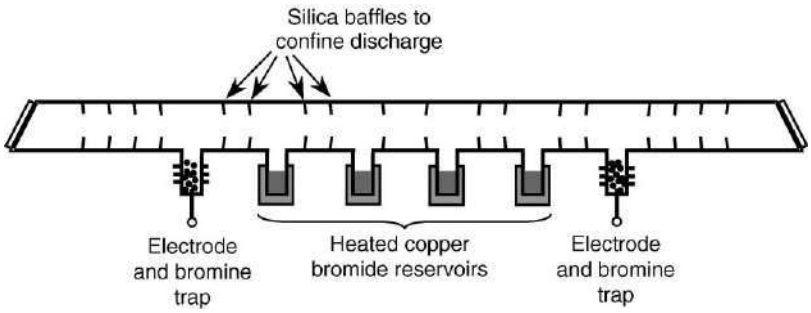

Figure 4 Copper bromide laser tube construction.

in Figure 5a, a dc high-voltage power supply resonantly charges a storage capacitor (C_S, typically a few nF) through a charging inductor L_C, a high-voltage diode and bypass inductor L_B, up to twice the supply voltage V_S in a time of order 100 μs. When the thyratron is triggered, the storage capacitor discharges through the thyratron and the laser head on a time-scale of 100 ns. Note that during the fast discharge phase, the bypass inductor in parallel with the laser head can be considered to be an open circuit. A peaking capacitor C_P ($\sim 0.5\ C_S$) is provided to increase the rate of rise of the voltage pulse across the laser head. Given the relatively high cost of thyratrons, more advanced circuits are now often used to extend the service lifetime of the thyratron to several thousand hours. In the more advanced circuit (Figure 5b) an LC inversion scheme is used in combination with magnetic pulse compression techniques and operates as follows. Storage capacitors C_S are resonantly charged in parallel to twice the dc supply voltage V_S as before. When the thyratron is switched, the charge on C_{S1} inverts through the thyratron and transfer inductor L_T. This LC inversion drags the voltage on the top of C_{S2} down to $-4V_S$. When the voltage across the first saturable inductor L_{S1} reaches a maximum ($-4V_S$) L_{S1} saturates, and allows current to flow from the storage capacitors (now charged in series) to the transfer capacitor

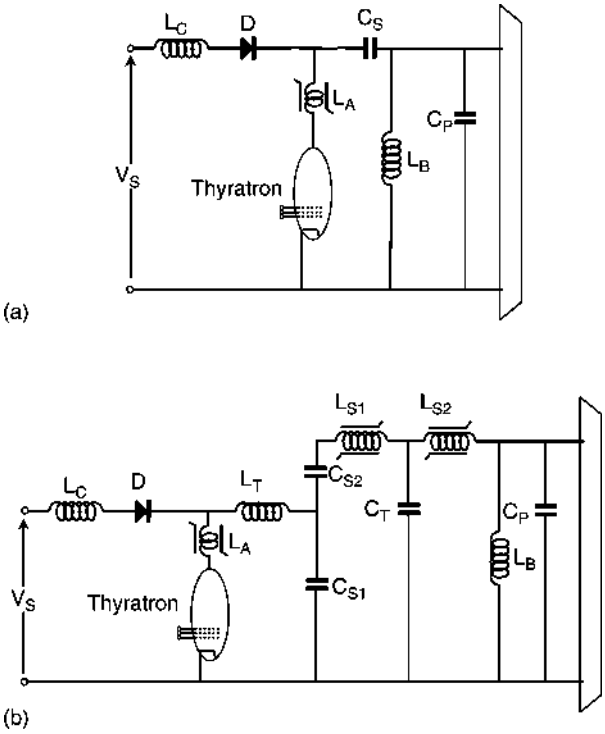

Figure 5 Copper vapor laser excitation circuits. (a) Charge transfer circuit; (b) LC inversion circuit with magnetic pulse compression.

($C_T = 0.5C_{S2}$) thereby transferring the charge from C_{S1} and C_{S2} to C_T in a time much less than the initial LC inversion time. At the moment when the charge on transfer capacitor C_T reaches a maximum (also $-4V_S$), L_{S2} saturates, and the transfer capacitor is discharged through the laser tube, again with a peaking capacitor to increase the voltage rise time. By using magnetic pulse compression the thyratron switched voltage can be reduced by 4 and the peak current similarly reduced (at the expense of increased current pulse duration) thereby greatly extending the thyratron lifetime. Note that in both circuits a 'magnetic assist' L_A saturable inductor is provided in series with the thyratron to delay the current pulse through the thyratron until after the thyratron has reached high conductivity thereby reducing power deposition in the thyratron.

Copper vapor lasers produce high average powers (2–100 W available commercially, with laboratory devices producing average powers of over 750 W) and have wall-plug efficiencies of approximately 1%. Copper vapor lasers also make excellent amplifiers due to their high gains, and the amplifiers can be chained together to produce average powers of several kW. Typical pulse repetition frequencies range from 4 to 20 kHz, with a maximum reported pulse repetition frequency of 250 kHz. An approximate scaling law states that for an elemental device with tube diameter D (mm) and length L (m) the average output power in watts will be of order $D \times L$. For example, a typical 25 W copper vapor laser will have a 25 mm diameter by 1 m long laser tube, and operate at 10 kHz corresponding to 2.5 mJ pulse energy and 50 kW peak power (50 ns pulse duration). Copper vapor lasers have very high single-pass gains (greater than 1000 for a 1 m long tube), large gain volumes and short gain durations (20–80 ns, sufficient for the intracavity laser light to make only a few round trips within the optical resonator). Maximum output power is therefore usually obtained in a highly 'multimode' (spatially incoherent) beam by using either a fully stable or a plane–plane resonator with a low reflectivity output coupler (usually Fresnel reflection from an uncoated optic is sufficient). To obtain higher beam quality a high magnification unstable resonator is required (Figure 6). Fortunately, copper vapor lasers have sufficient gain to operate efficiently with unstable resonators with magnifications ($M = R_1/R_2$) up to 100 and beyond. Resonators with such high magnifications impose very tight geometric constraints on propagation of radiation on repeated round-trips within the resonator, such that after two round-trips the divergence is typically diffraction-limited. Approximately half the stable-resonator

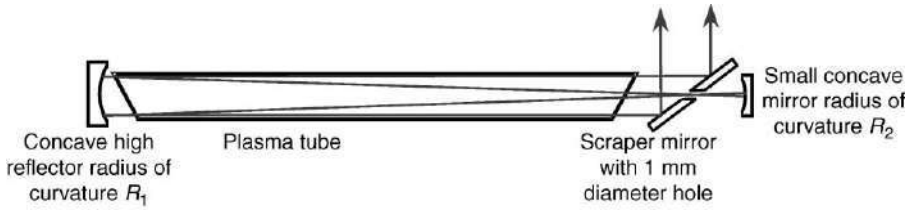

Figure 6 Unstable resonator configuration often used to obtain high beam quality from copper vapor lasers.

output power can therefore be obtained with near diffraction-limited beam quality by using an unstable resonator. Often, a small-scale oscillator is used in conjunction with a single power amplifier to produce high output power with diffraction-limited beam quality. Hyperfine splitting combined with Doppler broadening lead to an inhomogeneous linewidth for the laser transitions of order 8–10 GHz corresponding to a coherence length of order 3 cm.

The high beam quality and moderate peak power of CVLs allows efficient nonlinear frequency conversion to the UV by second harmonic generation (510.55 nm → 255.3 nm, 578.2 nm → 289.1 nm) and sum frequency generation (510.55 nm + 578.2 nm → 271.3 nm) using β-barium borate β-BaB_2O_4 (BBO) as the nonlinear medium. Typically average powers in excess of 1 W can be obtained at any of the three wavelengths from a nominally 20 W CVL. Powers up to 15 W have been obtained at 255 nm from high-power CVL master-oscillator power-amplifier systems using cesium lithium borate; CsLi B_6O_{10} (CLBO) as the nonlinear crystal.

Key applications of CVLs include pumping of dye lasers (principally for laser isotope separation) and pumping Ti:sapphire lasers. Medically, the CVL yellow output is particularly useful for treatment of skin lesions such as port wine stain birth marks. CVLs are also excellent sources of short-pulse stroboscopic illumination for high-speed imaging of fast objects and fluid flows. The high beam quality, visible wavelength and high pulse repetition rate make CVLs very suited to precision laser micromachining of metals, ceramics and other hard materials. More recently, the second harmonic at 255 nm has proved to be an excellent source for writing Bragg gratings in optical fibers.

Afterglow Recombination Metal Vapor Lasers

Recombination of an ionized plasma in the afterglow of a discharge pulse provides a mechanism for achieving a population inversion and hence laser output. The two main afterglow recombination metal vapor lasers are the strontium ion vapor laser (430.5 nm and 416.2 nm) and calcium ion vapor laser (373.7 nm and 370.6 nm) whose output in the violet and UV spectral regions extends the spectral coverage of pulsed metal vapor lasers to shorter wavelengths.

A population inversion is produced by recombination pumping where doubly ionized Sr (or Ca) recombines to form singly ionized Sr (or Ca) in an excited state: $Sr^{++} + e^- + e^- \rightarrow Sr^{+*} + e^-$. Note that recombination rates for a doubly ionized species are much faster than for singly ionized species hence recombination lasers are usually metal ion lasers. Recombination (pumping) rates are also greatest in a cool dense plasma, hence helium is usually used as the buffer gas as it is a light atom hence promotes rapid collisional cooling of the electrons. Helium also has a much higher ionization potential than the alkaline-earth metals (including Sr and Ca) which ensures preferential ionization of the metal species (up to 90% may be doubly ionized).

Recombination lasers can operate where the energy level structure of an ion species can be considered to consist of two (or more) groups of closely spaced levels. In the afterglow of a pulsed discharge electron collisional mixing within each group of levels will maintain each group in thermodynamic equilibrium yielding a Boltzmann distribution of population within each group. If the difference in energy between the two groups of levels (5 eV >> kT) is sufficiently large then thermal equilibrium between the groups cannot be maintained via collisional processes. Given that recombination yields excited singly ionized species (usually with a flux from higher to lower ion levels), it is possible to achieve a population inversion between the lowest level of the upper group and the higher levels within the lower group. This is the mechanism for inversion in the Sr^+ (and analogous Ca^+) ion laser as indicated in the partial energy level scheme for Sr^+ shown in Figure 7.

Strontium and calcium ion recombination lasers have similar construction to copper vapor lasers described above. The lower operating temperatures (500–800 °C) mean that minimal or no thermal insulation is required for self-heated devices.

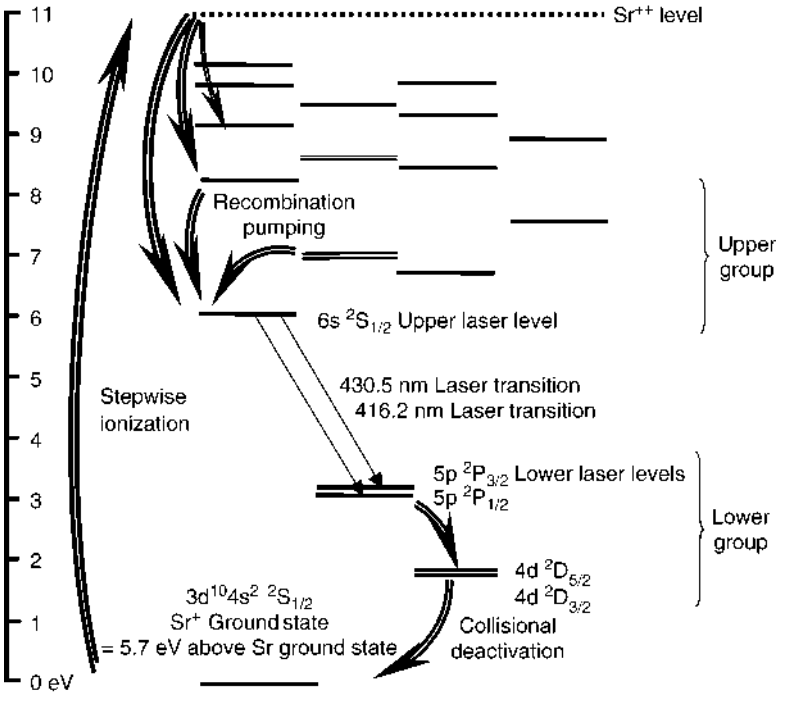

Figure 7 Partial energy level scheme for the strontium ion laser.

For good tube lifetime BeO plasma tubes are required due to the reactivity of the metal vapors. Usually helium at a pressure of up to one atmosphere is used as the buffer gas. Typical output powers at 5 kHz pulse repetition frequency are of order 1 W (~0.1% wall plug efficiency) at 430.5 nm from the He Sr^+ laser and 0.7 W at 373.7 nm from the He Ca^+ laser. For the high specific input power densities (10–15 W cm^{-3}) required for efficient lasing, overheating of the laser gas limits aperture scaling beyond 10–15 mm diameter. Slab laser geometries have been used successfully to aperture-scale strontium ion lasers. Scaling of the laser tube length beyond 0.5 m is not practical as achieving high enough excitation voltages for efficient lasing becomes problematic. Gain in strontium and calcium ion lasers is lower than in resonance-metastable metal vapor lasers such as the CVL, hence optimum output coupler reflectivity is approximately 70%. The pulse duration is also longer at around 200–500 ns resulting in moderate beam quality with plane–plane resonators.

For both strontium and calcium ion lasers the two principal transitions share upper laser levels and hence exhibit gain competition such that without wavelength-selective cavities usually only the longer wavelength of the pair is produced. With wavelength-selective cavities up to 60% (Sr^+) and 30% (Ca^+) of the normal power can be obtained at the shorter wavelength.

Many potential applications exist for strontium and calcium ion lasers, given their ultraviolet (UV)/violet wavelengths. Of particular importance is fluorescence spectroscopy in biology and forensics, treatment of neonatal jaundice, stereolithography, micromachining and exposing photoresists for integrated circuit manufacture. Despite these many potential applications, technical difficulties in power scaling mean that both strontium ion and calcium ion lasers have only limited commercial availability.

Continuous-Wave Metal Ion Lasers

In a discharge excited noble gas, there can be a large concentration of noble gas atoms in excited metastable states and noble gas ions in their ground states. These species can transfer their energy to a minority metal species (M) via two key processes: either charge transfer (Duffendack reactions) with the noble gas ions (N^+):

$$M + N^+ \rightarrow M^{+*} + N + \Delta E$$

(where M^{+*} is an excited metal ion state)

or Penning ionization with a noble gas atom in an excited metastable state (N^*):

$$M + N^* \rightarrow M^{+*} + N + e^-$$

In a continuous discharge in a mixture of a noble gas and a metal vapor, steady generation of excited metal ions via energy transfer processes can lead to a steady state population inversion on one or more pairs of levels in the metal ion and hence produce cw lasing.

Several hundred metal ion laser transitions have been observed to lase in a host of different metals. The most important such laser is the helium cadmium laser, which has by far the largest market volume by number of unit sales of all the metal vapor lasers. In a helium cadmium laser, Cd vapor is present at a concentration of about 1–2% in a helium buffer gas which is excited by a dc discharge. Excitation to the upper laser levels of Cd (Figure 8) is primarily via Penning ionization collisions with He^+ 2^3S_1 metastable ions produced in the discharge. Excitation via electron-impact excitation from the ion ground state may also play an important role in establishing a population inversion in Cd^+ lasers. Population inversion can be sustained continuously because the $^2P_{3/2}$ and $^2P_{1/2}$ lower laser levels decay via strong resonance transitions to the $^2S_{1/2}$ Cd^+ ground state, unlike in the self-terminating metal vapor lasers. Two principal wavelengths can be produced, namely 441.6 nm (blue) and 325.0 nm (UV) with cw powers up to 200 mW and 50 mW available respectively from commercial devices.

Typical laser tube construction (Figure 9) consists of a 1–3 mm diameter discharge channel typically 0.5 m long. A pin anode is used at one end of the laser tube together with a large-area cold cylindrical cathode located in a side-arm at the other end of the tube. Cadmium is transported to the main discharge tube from a heated Cd reservoir in a side-arm at 250–300 °C. With this source of atoms at the anode end of the laser a cataphoresis process (in which the positively charged Cd ions are propelled towards the cathode end of the tube by the longitudinal electric field) transports Cd into the discharge channel. Thermal insulation of the discharge tube ensures that it is kept hotter than the Cd reservoir to prevent condensation of Cd from blocking the tube bore. A large-diameter Cd condensation region is provided at the cathode end of the discharge channel. A large-volume side arm is also provided to act as a gas ballast for maintaining correct He pressure. As He is lost through sputtering and diffusion through the Pyrex glass tube walls, it is replenished from a high-pressure He reservoir by heating a permeable glass wall separating the reservoir from the ballast chamber which allows He to diffuse into the ballast chamber. Typical commercial sealed-off Cd lasers have operating lifetimes of several thousand hours. Overall laser construction is not that much more

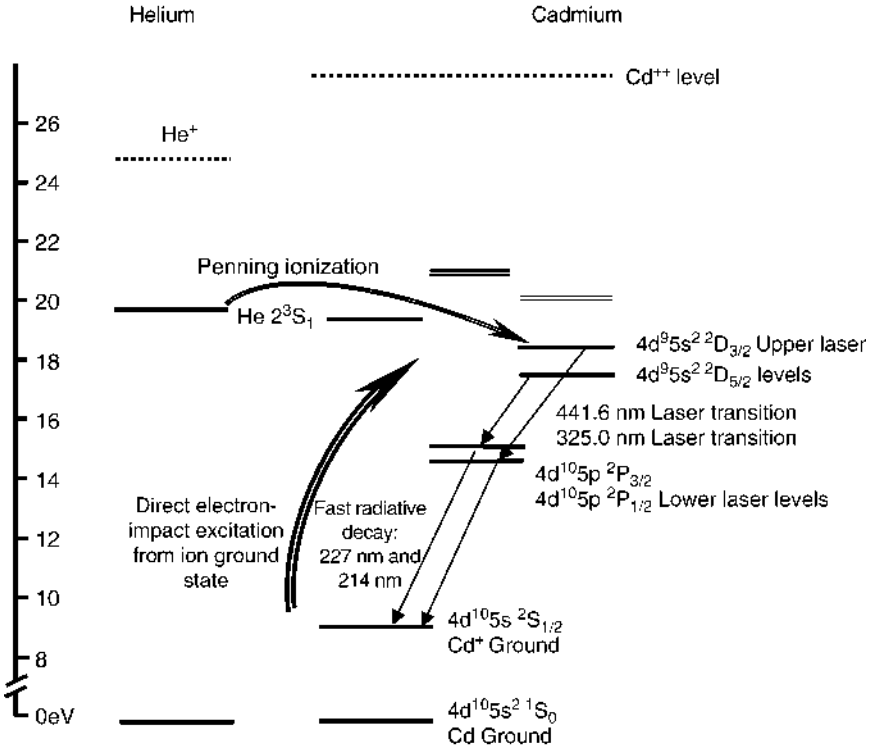

Figure 8 Partial energy level diagram for helium and cadmium giving HeCd laser transitions.

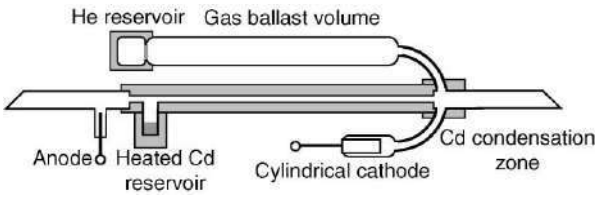

Figure 9 Helium cadmium laser tube construction.

complex than a HeNe laser, hence unit costs are considerably lower than low-power argon ion lasers which also provide output in the blue. Usually Brewster angle windows are provided together with a high Q stable resonator (97–99% reflectivity output coupler) to provide polarized output.

With their blue and UV wavelengths and relatively low cost (compared to low-power argon ion lasers), HeCd lasers have found wide application in science, medicine and industry. Of particular relevance is their application for exposing photoresists where the blue wavelength provides a good match to the peak photosensitivity of photoresist materials. A further key application is in stereolithography where the UV wavelength is used to cure an epoxy resin. By scanning the UV beam in a raster pattern across the surface of a liquid epoxy, a solid three-dimensional object may be built up in successive layers.

List of Units and Nomenclature

Boltzmann constant	[eV K^{-1}]	k
Excited metal ion		M^{+*}
Excited noble gas atom		N^{*}
Energy difference	[eV]	ΔE
Electron		e^{-}
Gas flow	[mbar l min^{-1}]	
Mirror curvatures	[m]	R_1, R_2
Metal atom		M
Noble gas atom		N
Noble gas ion		N^{+}
Pulse duration	[ns]	
Pulse repetition frequency	[kHz]	
Quality factor		Q
Temperature	[K]	T
Tube diameter	[mm]	D
Tube length	[m]	L
Wavelength	[nm], [μm]	
Unstable resonator magnification		M

See also

Nonlinear Sources: Harmonic Generation in Gases.

Further Reading

Little CE (1999) *Metal Vapor Lasers*. Chichester, UK: John Wiley.

Ivanov IG, Latush EL and Sem MF (1996) *Metal Vapor Ion Lasers, Kinetic Processes and Gas Discharges*. Chichester, UK: John Wiley.

Little CE and Sabotinov NV (eds) (1996) *Pulsed Metal Vapor Lasers*. Dordrecht, The Netherlands: Kluwer.

Lyabin NA, Chursin AD, Ugol'nikov SA, Koroleva ME and Kazaryan MA (2001) Development, production, and application of sealed-off copper and gold vapor lasers. *Quantum Electronics* 31: 191–202.

Petrash GG (ed.) (1989) *Metal Vapor and Metal Halide Vapor Lasers*. Commack, NY: Nova Science Publishers.

Silfvast WT (1996) *Laser Fundamentals*. New York: Cambridge University Press.

Noble Gas Ion Lasers

W B Bridges, California Institute of Technology, Pasadena, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

History

The argon ion laser was discovered in early 1964, and is still commercially available in 2004, with about $70 million in annual sales 40 years after this discovery. The discovery was made independently and nearly simultaneously by four different groups; for three of the four, it was an accidental result of studying the excitation mechanisms in the mercury ion laser (historically, the first ion laser), which had been announced only months before. For more on the early years of ion laser research and development, see the articles listed in the Further Reading section at the end of this article.

The discovery was made with pulsed gas discharges, producing several wavelengths in the blue and green portions of the spectrum. Within months, continuous operation was demonstrated, as well as oscillation on many visible wavelengths in ionized krypton and xenon. Within a year, over 100 wavelengths were observed to oscillate in the ions

of neon, argon, krypton, and xenon, spanning the spectrum from ultraviolet to infrared; oscillation was also obtained in the ions of other gases, for example, oxygen, nitrogen, and chlorine. A most complete listing of all wavelengths observed as gaseous ion lasers is given in the Laser Handbook cited in the Further Reading section. Despite the variety of materials and wavelengths demonstrated, however, it is the argon and krypton ion lasers that have received the most development and utilization.

Continuous ion lasers utilize high current density gas discharges, typically 50 A or more, and 2–5 mm in diameter. Gas pressures of 0.2 to 0.5 torr result in longitudinal electric fields of a few V/cm of discharge, so that the power dissipated in the discharge is typically 100 to 200 W/cm. Such high-power dissipation required major technology advances before long-lived practical lasers became available. Efficiencies have never been high, ranging from 0.01% to 0.2%. A typical modern ion laser may produce 10 W output at 20 kW input power from 440 V three-phase power lines and require 6–8 gallons/minute of cooling water. Smaller, air-cooled ion lasers, requiring 1 kW of input power from 110 V single-phase mains can produce 10–50 mW output power, albeit at even lower efficiency.

Theory of Operation

The strong blue and green lines of the argon ion laser originate from transitions between the 4p upper levels and 4s lower levels in singly ionized argon, as shown in Figure 1. The 4s levels decay radiatively to the ion ground state. The strongest of these laser lines are listed in Table 1. The notation used for the energy levels is that of the L–S coupling model. The ion ground state electron configuration is $3s^2 3p^5(^2P^o_{3/2})$. The inner ten electrons have the configuration $1s^2 2s^2 2p^6$, but this is usually omitted for brevity. The excited states shown in Figure 1 result from coupling a 4p or 4s electron to a $3s^2 3p^4(^3P)$ core, with the resulting quantum numbers S (the net spin of the electrons), L (the net angular momentum of the electrons), and J (the angular momentum resulting from coupling S to L) are represented by $^{2S+1}L_J$, where $L = 0, 1, 2, \ldots$ is denoted S, P, D, F, … The superscript 'o' denotes an odd level, while even levels omit a superscript. Note that some weaker transitions involving levels originating from the $3s^2 3p^4(^1D)$ core configuration also oscillate. Note also that the quantum mechanical selection rules for the L–S coupling model are not rigorously obeyed, although the stronger laser lines satisfy most or all of these rules. The selection rule $|\Delta J| = 1$ or 0, but not $J = 0 \rightarrow J = 0$ is always obeyed. All transitions shown in Figure 1 and Table 1 belong to the second spectrum of argon, denoted Ar II. Lines originating from transitions in the neutral atom make up the first spectrum, Ar I; lines originating from transitions in doubly ionized argon are denoted Ar III, and so forth for even more highly ionized states.

Much work has been done to determine the mechanisms by which the inverted population is formed in the argon ion laser. Reviews of this extensive research

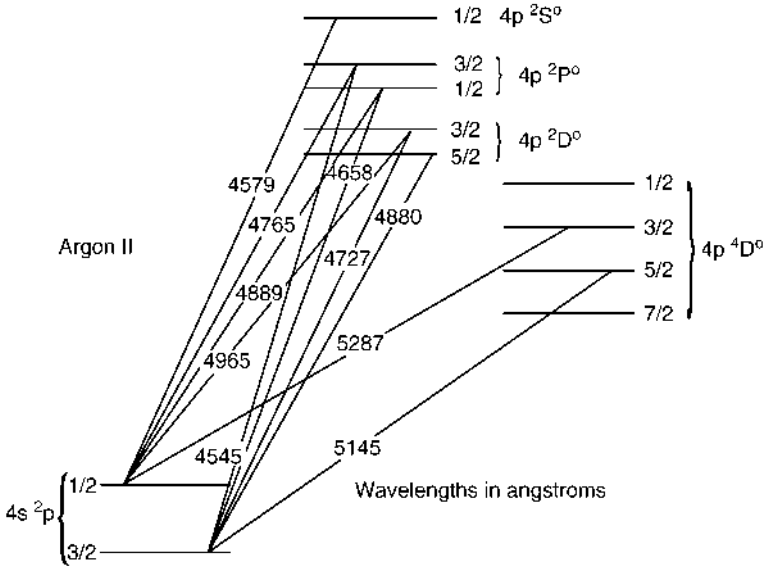

Figure 1 4p and 4s doublet levels in singly ionized argon, showing the strongest blue and green laser transitions.

Table 1 Ar II laser blue-green wavelengths

Wavelength (nanometers)	*Transition*[a] *(upper level) → (lower level)*	*Relative strength*[b] *(Watt)*
454.505	4p $^2P^o_{3/2} \rightarrow$ 4s $^2P_{3/2}$	0.8
457.935	4p $^2S^o_{1/2} \rightarrow$ 4s $^2P_{1/2}$	1.5
460.956	(^{1}D)4p $^2F^o_{7/2} \rightarrow$ (^{1}D)4s $^2D_{5/2}$	–
465.789	4p $^2P^o_{1/2} \rightarrow$ 4s $^2P_{3/2}$	0.8
472.686	4p $^2D^o_{3/2} \rightarrow$ 4s $^2P_{3/2}$	1.3
476.486	4p $^2P^o_{3/2} \rightarrow$ 4s $^2P_{1/2}$	3.0
487.986	4p $^2D^o_{5/2} \rightarrow$ 4s $^2P_{3/2}$	8.0
488.903	4p $^2P^o_{1/2} \rightarrow$ 4s $^2P_{1/2}$	[c]
496.507	4p $^2D^o_{3/2} \rightarrow$ 4s $^2P_{1/2}$	3.0
501.716	(^{1}D)4p $^2D^o_{5/2} \rightarrow$ 3d $^2D_{3/2}$	1.8
514.179	(^{1}D)4p $^2F^o_{7/2} \rightarrow$ 3d $^2D_{5/2}$	[c]
514.532	4p $^4D^o_{5/2} \rightarrow$ 4s $^2P_{3/2}$	10
528.690	4p $^4D^o_{3/2} \rightarrow$ 4s $^2P_{1/2}$	1.8

[a]All levels are denoted in $L-S$ coupling with the (^{3}P) core unless otherwise indicated. Odd parity is denoted by the superscript 'o'.
[b]Relative strengths are given as power output from a commercial Spectra-Physics model 2080-25S argon ion laser.
[c]These lines may oscillate simultaneously with the nearby strong line, but are not resolved easily in the output beam.

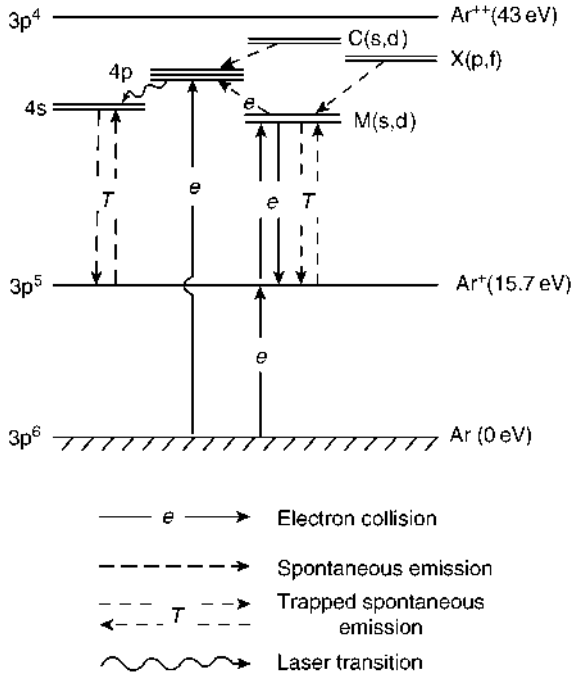

Figure 2 Schematic representation of energy levels in neutral and singly ionized argon, indicating alternative pathways for excitation and de-excitation of the argon ion laser levels.

are found in the Further Reading section. While a completely quantitative picture of argon ion laser operation is lacking to this day, the essential processes are known. Briefly, the 4p upper levels are populated by three pathways, as illustrated in Figure 2:

(i) by electron collision with the 3p^6 neutral ground state atoms. This 'sudden perturbation process' requires at least 37 eV electrons, and also singles out the 4p $^2P^o_{3/2}$ upper level, which implies that only the 476 and 455 nm lines would oscillate. This is the behavior seen in pulsed discharges at very low pressure and very high axial electric field. This pathway probably contributes little to the 4p population under ordinary continuous wave (cw) operating conditions, however.

(ii) by electron collision from the lowest-lying s and d states in the ion, denoted M(s,d) in Figure 1. This only requires 3–4 eV electrons. These states have parity-allowed transitions to the ion ground state, but are made effectively metastable by radiation trapping (that is, there is a high probability that an emitted photon is re-absorbed by another ground state ion before it escapes the discharge region), or by requiring $|\Delta J|$ to be 2 to make the transition (forbidden by quantum selection rules). Thus, these levels are both created and destroyed primarily by electron collision, causing the population of the M(s,d) states to follow the population of the singly ionized ground state, which, in turn, is approximately proportional to the discharge current. Since a second electron collision is required to get from the M(s,d) states to the 4p states, a quadratic variation with current for the laser output power would be expected, and that is what is observed over some reasonable range of currents between threshold and saturation. Note that radiative decay from higher-lying opposite-parity p and f states, denoted X(p,f), can also contribute to the population of M(s,d), but the linear variation in population of the M(s,d) with discharge current is assured by electron collision creation and destruction.

(iii) by radiative decay from higher-lying opposite-parity s and d states, denoted C(s,d). These states are populated by electron collision with the 3p^5 ion ground states, and thus have populations that also vary quadratically with discharge current. The contribution of this cascade process has been measured to be 20% to 50% of the 4p upper laser level population.

Note that it is not possible to distinguish between processes (ii) and (iii) by the variation of output power with discharge current; both give the observed quadratic dependence.

The radiative lifetimes of the 4s ^{2}P levels are sufficiently short to depopulate the lower laser levels by radiative decay. However, radiation trapping greatly lengthens this decay time, and a bottleneck

can occur. In pulsed ion lasers, this is exhibited by the laser pulse terminating before the excitation current pulse ends, which would seem to preclude continuous operation. However, the intense discharge used in continuous operation heats the ions to the order of 2300 K, thus greatly Doppler broadening the absorption linewidth and reducing the magnitude of the absorption. Additionally, the ions are attracted to the discharge tube walls, so that the absorption spectrum is further broadened by the Doppler shift due to their wall-directed velocities. The plasma wall sheath gives about 20 V drop in potential from the discharge axis to the tube wall, so most ions hit the wall with 20 eV of energy, or about ten times their thermal velocity. A typical cw argon ion laser operates at ten times the gas pressure that is optimum for a pulsed laser, and thus the radiation trapping of the $4s \rightarrow 3p^5$ transitions is so severe that it may take several milliseconds after discharge initiation for the laser oscillation to begin.

As the discharge current is increased, the intensities of the blue and green lines of Ar II eventually saturate, and then decrease with further current. At these high currents, there is a buildup in the population of doubly ionized atoms, and some lines of Ar III can be made to oscillate with the appropriate ultraviolet mirrors. Table 2 lists the strongest of these lines, those that are available in the largest commercial lasers. Again, there is no quantitative model for the performance in terms of the discharge parameters, but the upper levels are assumed to be populated by processes analogous to those of the Ar II laser. At still higher currents, lines in Ar IV can be made to oscillate as well.

Much less research has been done on neon, krypton, and xenon ion lasers, but it is a good assumption that the population and depopulation processes are the same in these lasers. Table 3 lists both the Kr II and Kr III lines that are available from the largest commercial ion lasers. Oscillation on lines in still-higher ionization states in both krypton and xenon have been observed.

Operating Characteristics

A typical variation of output power with discharge current for an argon ion laser is shown in Figure 3. This particular laser had a 4 mm diameter discharge in a water-cooled silica tube, 71 cm in length, with a 1 kG axial magnetic field. The parameter is the argon pressure in the tube before the discharge was struck. Note that no one curve is exactly quadratic, but that the envelope of the curves at different filling pressures is approximately quadratic. At such high discharge current densities (50 A in the 4 mm tube is approximately 400 A/cm^2) there is substantial pumping of gas out of the small-bore discharge region. Indeed, a return path for this pumped gas must be provided from anode to cathode ends of the discharge to keep the discharge from self-extinguishing. The axial electric field in this discharge was 3–5 V/cm, so the input power was of the order of 10 to 20 kW, yielding an efficiency of less than 0.1%, an unfortunate characteristic of all ion lasers.

Table 2 Ultraviolet argon ion laser wavelengths

Wavelength (nanometers)	*Spectrum*	*Transition*[a] *(upper level) → (lower level)*	*Relative strength*[b] *(Watt)*
275.392	III	$(^2D^o)4p\ ^1D_2 \rightarrow (^2D^o)4s\ ^1D^o_2$	0.3
275.6	?	?	0.02
300.264	III	$(^2P^o)4p\ ^1P_1 \rightarrow (^2P^o)3d\ ^1D^o_2$	0.5
302.405	III	$(^2P^o)4p\ ^3D_3 \rightarrow (^2P^o)4s\ ^3P^o_2$	0.5
305.484	III	$(^2P^o)4p\ ^3D_2 \rightarrow (^2P^o)4s\ ^3P^o_1$	0.2
333.613	III	$(^2D^o)4p\ ^3F_4 \rightarrow (^2D^o)4s\ ^3D^o_3$	0.4
334.472	III	$(^2D^o)4p\ ^3F_3 \rightarrow (^2D^o)4s\ ^3D^o_2$	0.8
335.849	III	$(^2D^o)4p\ ^3F_2 \rightarrow (^2D^o)4s\ ^3D^o_1$	0.8
350.358	III	$(^2D^o)4p\ ^3D_2 \rightarrow (^2D^o)4s\ ^3D^o_2$	0.05
350.933	III	$(^4S^o)4p\ ^3P_0 \rightarrow (^4S^o)4s\ ^3S^o_1$	0.05
351.112	III	$(^4S^o)4p\ ^3P_2 \rightarrow (^4S^o)4s\ ^3S^o_1$	2.0
351.418	III	$(^4S^o)4p\ ^3P_1 \rightarrow (^4S^o)4s\ ^3S^o_1$	0.7
363.789	III	$(^2D^o)4p\ ^1F_3 \rightarrow (^2D^o)4s\ ^1D^o_2$	2.5
379.532	III	$(^2P^o)4p\ ^3D_3 \rightarrow (^2P^o)3d\ ^3P^o_2$	0.4
385.829	III	$(^2P^o)4p\ ^3D_2 \rightarrow (^2P^o)3d\ ^3P^o_1$	0.15
390.784	III	$(^2P^o)4p\ ^3D_1 \rightarrow (^2P^o)3d\ ^3P^o_0$	0.02
408.904	IV?	?	0.04
414.671	III	$(^2D^o)4p\ ^3P_2 \rightarrow (^2P^o)4s\ ^3P^o_2$	0.02
418.298	III	$(^2D^o)4p\ ^1P_1 \rightarrow (^2D^o)4s\ ^1D^o_2$	0.08

[a]All levels are denoted in $L-S$ coupling with the core shown in ( ). Odd parity is denoted by the superscript 'o'.
[b]Relative strengths are given as power output from a commercial Spectra-Physics model 2085-25S argon ion laser.

Table 3 Krypton ion laser wavelengths

Wavelength (nanometers)	*Spectrum*	*Transition*[a] *(upper level) → (lower level)*	*Relative strength*[b] *(Watt)*
337.496	III	$(^2P^o)5p\ ^3D_3 \rightarrow (^2P^o)5s\ ^3P^o_2$	–
350.742	III	$(^4S^o)5p\ ^3P_2 \rightarrow (^4S^o)5s\ ^3S^o_1$	1.5
356.432	III	$(^4S^o)5p\ ^3P_1 \rightarrow (^4S^o)5s\ ^3S^o_1$	0.5
406.737	III	$(^2D^o)5p\ ^1F_3 \rightarrow (^2D^o)5s\ ^1D^o_2$	0.9
413.133	III	$(^4S^o)5p\ ^5P_2 \rightarrow (^4S^o)5s\ ^3S^o_1$	1.8
415.444	III	$(^2D^o)5p\ ^3F_3 \rightarrow (^2D^o)5s\ ^1D^o_2$	0.3
422.658	III	$(^2D^o)5p\ ^3F_2 \rightarrow (^2D^o)4d\ ^3D^o_1$	–
468.041	II	$(^3P)5p\ ^2S^o_{1/2} \rightarrow (^3P)5s\ ^2P_{1/2}$	0.5
476.243	II	$(^3P)5p\ ^2D^o_{3/2} \rightarrow (^3P)5s\ ^2P_{1/2}$	0.4
482.518	II	$(^3P)5p\ ^4S^o_{3/2} \rightarrow (^3P)5s\ ^2P_{1/2}$	0.4
520.832	II	$(^3P)5p\ ^4P^o_{3/2} \rightarrow (^3P)5s\ ^4P_{3/2}$	–
530.865	II	$(^3P)5p\ ^4P^o_{5/2} \rightarrow (^3P)5s\ ^4P_{3/2}$	1.5
568.188	II	$(^3P)5p\ ^4D^o_{5/2} \rightarrow (^3P)5s\ ^2P_{3/2}$	0.6
631.024	III	$(^2D^o)5p\ ^3P_2 \rightarrow (^2P^o)4d\ ^3D_1$	0.2
647.088	II	$(^3P)5p\ ^4P^o_{5/2} \rightarrow (^3P)5s\ ^2P_{3/2}$	3.0
676.442	II	$(^3P)5p\ ^4P^o_{1/2} \rightarrow (^3P)5s\ ^2P_{1/2}$	0.9
752.546	II	$(^3P)5p\ ^4P^o_{3/2} \rightarrow (^3P)5s\ ^2P_{1/2}$	1.2
793.141	II	$(^1D)5p\ ^4F^o_{7/2} \rightarrow (^3P)4d\ ^2F_{5/2}$	0.3
799.322	II	$(^3P)5p\ ^4P^o_{3/2} \rightarrow (^3P)4d\ ^4D_{1/2}$	

[a]All levels are denoted in $L-S$ coupling with the core shown in ( ). Odd parity is denoted by the superscript 'o'.
[b]Relative strengths are given as power output from a commercial Spectra-Physics model 2080RS ion laser.

Technology

With such high input powers required in a small volume to produce several watts output, ion laser performance has improved from 1964 to the present only as new discharge technologies were introduced. The earliest laboratory argon ion lasers used thin-walled ($\approx$1 mm wall thickness) fused silica discharge tubes, cooled by flowing water over the outside wall of the tube. The maximum input power per unit length of discharge was limited by thermal stresses in the silica walls caused by the temperature differential from inside to outside. Typically, ring-shaped cracks would cause the tube to fail catastrophically. Attempts to make metal–ceramic structures with alumina (Al_2O_3) discharge tubes to contain the plasma were made early on (1965) but were not successful. Such tubes invariably failed from fracture by thermal shock as the discharge was turned on. Later, successful metal–ceramic tubes were made with beryllia (BeO), which has a much higher thermal conductivity than silica or alumina and is much more resistant to thermal shock. Today, all of the lower power (less than 100 mW output) are made with BeO discharge tubes. Some ion lasers in the 0.5 to 1 W range are also made with water-cooled BeO discharge tubes.

A typical low-power air cooled argon ion laser is shown in Figure 4. The large metal can (2) on the right end of the tube contains an impregnated-oxide hot cathode, heated directly by current through ceramic feed-through insulators (3). The small ($\approx$1 mm diameter) discharge bore runs down the center of the BeO ceramic rod (1), and several smaller diameter gas return path holes run off-axis parallel to the discharge bore to provide the needed gas equalization between cathode can and anode region. A copper honeycomb cooler is brazed to the cathode can, two more to the outer wall of the BeO cylinder, and one to the anode (4) at the left end of the tube. The laser mirrors (5) are glass-fritted to the ends of the tube, forming a good vacuum seal. Note that in this small laser, the active discharge bore length is less than half of the overall length.

The very early argon ion lasers were made with simple smooth dielectric tubes, allowing the continuous variation in voltage along the length of the tube required by the discharge longitudinal electric field. However, this variation can be step-wise at the discharge walls, and still be more or less smooth along the axis. Thus, the idea arose of using metal tube segments, insulated one from another, to form the discharge tube walls. The first version of this idea (1965) used short ($\approx$1 cm) metal cylinders supported by metal disks and stacked inside a large diameter silica envelope, with each metal cylinder electrically isolated from the others. The tubes and disks were made of molybdenum, and were allowed to heat to incandescence, thus radiating several kilowatts of heat through the silica vacuum envelope to a water-cooled collector outside. While this eliminated the problems of thermal shock and poor thermal conductivity inherent in dielectric wall discharges, it made another

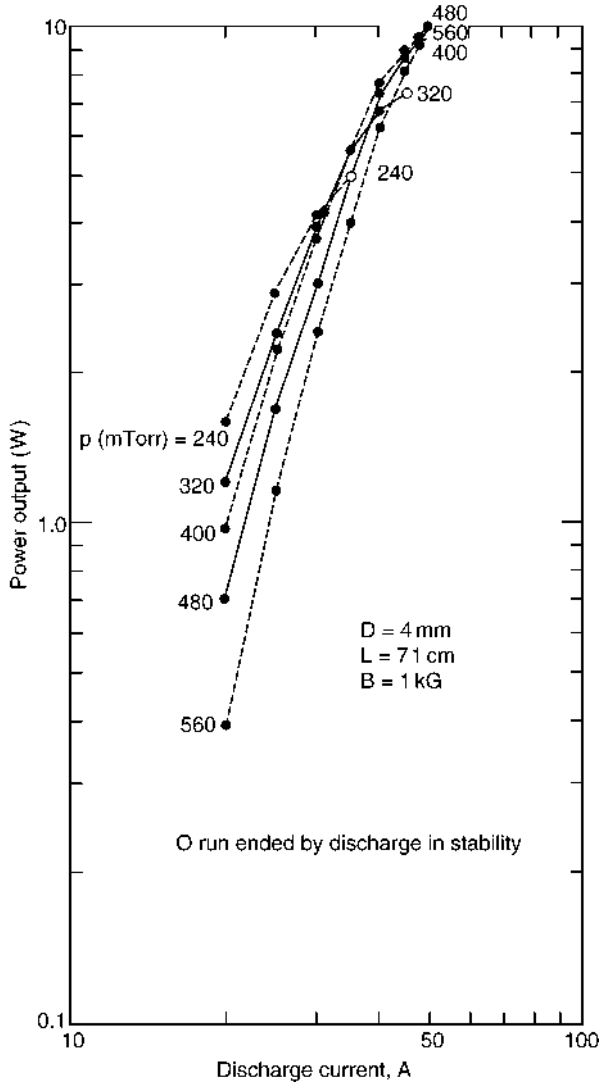

Figure 3 Laser output power (summed over all the blue and green laser lines) versus discharge current for a laser discharge 4 mm in diameter and 71 cm long, with a 1 kilogauss longitudinal magnetic field. The parameter shown is the argon fill pressure in mTorr prior to striking the discharge. The envelope of the curves exhibits the quadratic variation of output power with discharge current.

problem painfully evident. The intense ion bombardment of the metal tube walls sputtered the wall material, eventually eroding the shape of the metal cylinders and depositing metal films on the insulating wall material, thus shorting one segment to another.

Many different combinations of materials and configurations were investigated in the 1960s and 1970s to find a structure that would offer good laser performance and long life. It was found that simple thin metal disks with a central hole would effectively confine the discharge to a small diameter (on the order of the hole size) if a longitudinal d-c magnetic field of the order of 1 kiloGauss were used. The spacing between disks can be as large as 2–4 discharge diameters. Of the metals, tungsten has the lowest sputtering yield for argon ions in the 20 eV range (which is approximately the energy they gain in falling to the wall across the discharge sheath potential difference). An even lower sputtering yield is exhibited by carbon, and graphite cylinders contained within a larger diameter silica or alumina tube were popular for a while for ion laser discharges. Unfortunately, graphite has a tendency to flake or powder, so such laser discharge tubes became contaminated with 'dust' which could eventually find its way to the optical windows of the tube. Beryllia and silica also sputter under argon ion bombardment, but with still lower yields than metals or carbon; however, their smaller thermal conductivities limit them to lower-power applications.

The material/configuration combination that has evolved for higher-power argon ion lasers today is to use a stack of thin tungsten disks with 2–3 mm diameter holes for the discharge. These disks, typically 1 cm in diameter, are brazed coaxially to a larger copper annulus (actually, a drawn cup with a 1 cm hole on its axis). A stack of these copper/tungsten structures is, in turn, brazed to the inside wall of a large diameter alumina vacuum envelope. The tungsten disks are exposed to and confine the discharge, while the copper cups conduct heat radially outward to the alumina tube wall, which, in turn, is cooled by fluid flow over its exterior. Thus, the discharge is in contact only with a low-sputtering material (tungsten) while the heat is removed by a high thermal conductivity material (copper). Details differ among manufacturers, but this 'cool disk' technology seems to have won out in the end. A photo of a half-sectioned disk/cup stacked assembly from a Coherent Innova™ ion laser is shown in Figure 5. The coiled impregnated-tungsten cathode is also shown.

Sputtering of the discharge tube walls is not uniform along the length of the gas discharge. The small diameter region where the laser gain occurs is always joined to larger diameter regions containing cathode and anode electrodes (as shown in Figure 5, for example). A plasma double sheath (that is, a localized increase in potential) forms across the discharge in the transition region between large and small diameter regions (the discharge 'throats,' which may be abrupt or tapered). This sheath is required to satisfy the boundary conditions between the plasmas of different temperatures in the different regions. Such a double sheath imparts additional energy to the ions as they cross the sheath, perhaps

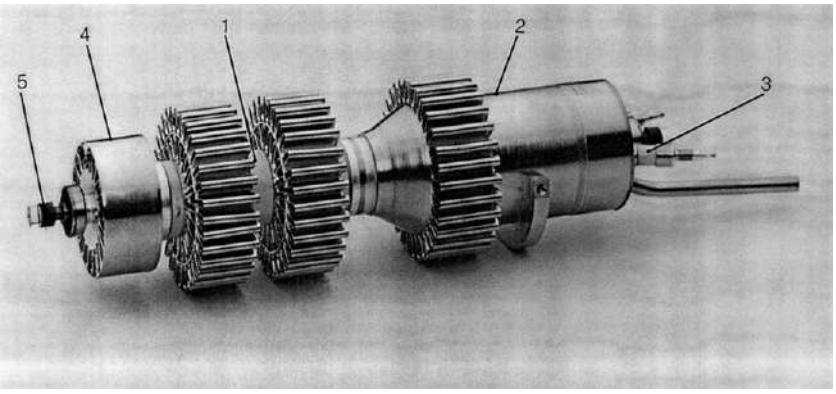

Figure 4 A typical commercial low-power, air-cooled argon ion laser. The cathode can (2) is at the right and the beryllia bore (1) and anode (4) is at the left. The cathode current is supplied through ceramic vacuum feed-throughs (3). Mirrors (5) are glass fritted onto the ends of vacuum envelope. (Photo courtesy of JDS Uniphase Corp.)

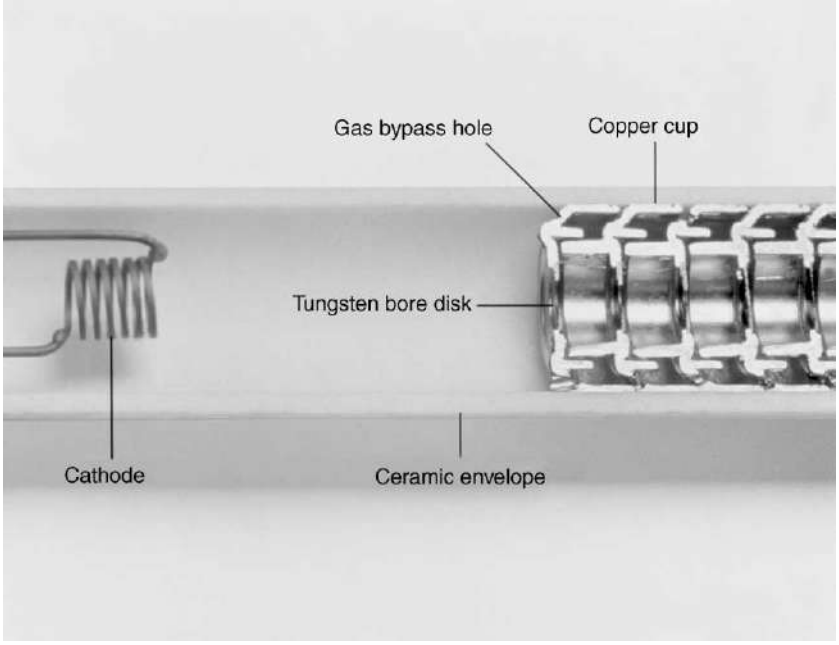

Figure 5 Discharge bore structure of a typical commercial high-power argon ion laser, the Coherent Innova™. Copper cups are brazed to the inner wall of a ceramic envelope, which is cooled by liquid flow over its outer surface. A thin tungsten disk with a small hole defining the discharge path is brazed over a larger hole in the bottom of the copper cup. Additional small holes in the copper cup near the ceramic envelope provide a gas return path from cathode to anode. Also shown is the hot oxide-impregnated tungsten cathode. (Photo courtesy of Coherent, Inc.)

an additional 20 eV. When these ions eventually hit the discharge walls near the location of the sheath, they have 40 eV of energy rather than the 20 eV from the normal wall sheath elsewhere in the small diameter discharge. Sputtering yield (number of sputtered wall atoms per incident ion) is exponentially dependent on ion energy in this low ion energy region, so the damage done to the wall in the vicinity of the 'throat' where the double sheath forms may be more than ten times that elsewhere in the small diameter bore region. This was the downfall of high-power operation of dielectric discharge bores, even BeO; while sputtering was acceptable elsewhere in the discharge, the amount of material removed in a small region near the discharge throat would cause catastrophic bore failure at that point. This localized increase in sputtering in the discharge throat is common to all

ion lasers, including modern cooled tungsten disk tubes. Eventually, disks near the throat are eroded to larger diameters, and more material is deposited on the walls nearby. Attempts to minimize localized sputtering by tapering the throat walls or the confining magnetic field have proven unsuccessful; a localized double sheath always forms somewhere in the throat. Because of the asymmetry caused by ion flow, the double sheath is larger in amplitude in the cathode throat than the anode throat, so that the localized wall damage is larger at the cathode end of the discharge than at the anode end.

Another undesirable feature of sputtering is that the sputtered material 'buries' some argon atoms when it is deposited on a wall. This is the basis of the well-known Vac-Ion® vacuum pump. Thus, the operating pressure in a sealed laser discharge tube will drop during the course of operation. In low-power ion lasers, this problem is usually solved by making the gas reservoir volume large enough to satisfy the desired operating life (for example, the large cathode can in Figure 4). In high-power ion lasers, the gas loss would result in unacceptable operating life even with a large reservoir at the fill pressure. Thus, most high-power ion lasers have a gas pressure measurement system and dual-valve arrangement connected to a small high-pressure reservoir to 'burp' gas into the active discharge periodically to keep the pressure within the operating range. Unfortunately, if a well-used ion laser is left inoperative for months, some of the 'buried' argon tends to leak back into the tube, with the gas pressure becoming higher than optimum. The pressure will gradually decrease to its optimum value with further operation of the discharge (in perhaps tens of hours). In the extreme case, the gas pressure can rise far enough so that the gas discharge will not strike, even at the maximum power supply voltage. Such a situation requires an external vacuum pump to remove the excess gas, usually a factory repair.

In addition to simple dc discharges, various other techniques have been used to excite ion lasers, primarily in a search for higher power, improved efficiency and longer operating life. Articles in the Further Reading section give references to these attempts. Radio-frequency excitation at 41 MHz was used in an inductively coupled discharge, with the laser bore and its gas return path forming a rectangular single turn of an air-core transformer. A commercial product using this technique was sold for a few years in the late 1960s. Since this was an 'electrode-less' discharge, ion laser lines in reactive gases such as chlorine could be made to oscillate, as well as the noble gases, without 'poisoning' the hot cathode used in conventional dc discharge lasers. A similar electrode-less discharge was demonstrated as a quasi-cw laser by using iron transformer cores and exciting the discharge with a 2.5 kHz square wave. Various microwave excitation configurations at 2.45 GHz and 9 GHz also resulted in ion laser oscillation. Techniques common to plasma fusion research were also studied. Argon ion laser oscillation was produced in Z-pinch and Θ-pinch discharges and also by high-energy (10–45 keV) electron beams. However, none of these latter techniques resulted in a commercial product, and all had efficiencies worse than the simple dc discharge lasers.

It is interesting that the highest-power output demonstrations were made in less than ten years after the discovery. Before 1970, 100 W output on the argon blue-green lines was demonstrated with dc discharges two meters in length. In 1970, a group in the Soviet Union reported 500 W blue-green output from a two-meter discharge with 250 kW of dc power input, an efficiency of 0.2%. Today, the highest output power argon ion laser offered for sale is 50 W output.

Manufacturers

More than 40 companies have manufactured ion lasers for sale over the past four decades. This field has now (2004), narrowed to the following:

Coherent, Inc.	http://www.coherentinc.com
INVERSion Ltd.	http://inversion.iae.nsk.su
JDS Uniphase	http://www.jdsu.com
Laser Physics, Inc.	http://www.laserphysics.com
Laser Technologies GmbH	http://www.lg-lasertechnologies.com
LASOS Lasertechnik GmbH	http://www.LASOS.com
Lexel Laser, Inc.	http://www.lexellaser.com
Melles Griot	http://lasers.mellesgriot.com
Spectra-Physics, Inc.	http://www.spectraphysics.com

Further Reading

Bridges WB (1979) Atomic and ionic gas lasers. In: Marton L and Tang CL (eds) *Methods of Experimental Physics, Vol 15: Quantum Electronics, Part A*, pp. 31–166. New York: Academic Press.

Bridges WB (1982) Ionized gas lasers. In: Weber MJ (ed.) *Handbook of Laser Science and Technology, Vol. II, Gas Lasers*, pp. 171–269. Boca Raton, FL: CRC Press.

Bridges WB (2000) Ion lasers – the early years. *IEEE Journal of Selected Topics in Quantum Electronics* 6: 885–898.

Davis CC and King TA (1975) Gaseous ion lasers. In: Goodwin DW (ed.) *Advances in Quantum Electronics*, vol. 3, pp. 169–469. New York: Academic Press.

Dunn MH and Ross JN (1976) The argon ion laser. In: Sanders JH and Stenholm S (eds) *Progress in Quantum Electronics*, vol. 4, pp. 233–269. New York: Pergamon.

Weber MJ (2000) *Handbook of Laser Wavelengths*. Boca Raton, FL: CRC Press.

Optical Fiber Lasers

G E Town, Macquarie University, NSW, Australia

N N Akhmediev, Australian National University, ACT, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

Optical fiber lasers were first demonstrated in the 1960s, and since then have developed to become versatile optical sources with many desirable properties. Aided by developments in associated technologies, such as fiber design and fabrication methods, semiconductor pump diode technology, and fiber-coupled and in-fiber components such as Bragg grating filters, optical fiber lasers now compete with other laser technologies in many applications, from telecommunications to materials processing.

An optical fiber laser is fundamentally an optical oscillator, which converts input pump power to coherent optical output power at one or more well-defined wavelengths. Optical oscillators require two basic elements: optical gain, and optical feedback. For sustained oscillation to occur the round-trip gain in the laser cavity must be unity, and the round-trip phase a multiple of 2π. In optical fiber lasers the optical gain is provided by an optical fiber amplifier using one or a combination of fundamental physical processes, e.g., stimulated emission, stimulated scattering, or nonlinear parametric processes. Early fiber amplifiers used stimulated Raman scattering to produce optical gain; however, rare-earth-doped fiber amplifiers, in which gain is provided by stimulated emission, are now more common. The reader is referred to **Optical Amplifiers:** Erbrium Doped Fiber Amplifiers for Lightwave Systems, for more details on doped fiber amplifiers, and to **Scattering:** Stimulated Scattering, **Nonlinear Optics, Applications:** Raman Lasers, **Optical Parametric Devices:** Optical Parametric Oscillators (Pulsed), and **Scattering:** Raman Scattering, for more details on stimulated scattering and parametric gain processes. Optical feedback may be provided in two fundamental ways, e.g., by using a closed ring of fiber, or from reflections from nonuniformities and discontinuities in the waveguide such as in Bragg grating filters or at fiber ends. The reader is referred to **Fiber Gratings** for more details on fiber Bragg gratings.

The main advantages of optical fiber lasers are derived from the confinement of the pump and signal in a small optical waveguide. In contrast with bulk lasers, the pump intensity in the fiber waveguide is largely independent of the laser length, resulting in large amplifier gain and low laser threshold, even for gain media with small absorption and emission cross-sections. The large gain enables lossy elements such as optical fiber-coupled devices and bulk optical elements to be incorporated into the laser cavity, providing additional control over the optical signal being generated. To obtain a large gain optical fiber amplifiers must usually be relatively long (i.e., from several centimeters to many meters), and so the linear and nonlinear properties of the fiber waveguide can have a significant influence on the optical signal being generated, resulting in some interesting and useful phenomena, e.g., as in soliton fiber lasers. The reader is referred to **Fiber and Guided Wave Optics:** Optical Fiber Cables, for fiber-based components, to **Fiber and Guided Wave Optics**: Dispersion, Light Propagation, and Nonlinear Effects (Basics), for a review of the linear and nonlinear properties of optical fibers, and to **Solitons:** Soliton Communication Systems, and Temporal Solitons, for the theory and applications of temporal solitons.

Whilst optical fibers may be fabricated in a range of materials, including polymer and crystalline materials, most fibers and fiber lasers are currently made of silica containing rare-earth ion dopants – see **Fiber and Guided Wave Optics:** Fabrication of Optical Fiber, for details. Lasers in silica optical fiber have much in common with other glass-host lasers, including a wide range of potential pump and lasing wavelengths, and broadband absorption and gain due to homogeneous and inhomogeneous broadening of the lasing energy levels in the amorphous glass host. For example, erbium-doped amplifiers in alumino-silicate glass fibers at room temperature have lasing transitions with homogeneous and

Table 1 Commonly used pump and signal transitions in rare-earth doped fiber lasers

Rare earth dopant	*Pump wavelengths [nm]*	*Signal wavelengths [nm]*	*Excited state lifetime [ms]*	*Energy levels*
Praseodymium	480, 585	885, 1080	0.1	4
Neodymium	590, 800	920 (900–950)	0.5	3
		1060 (1055–1140)		4
Samarium	488	651	1.5	4
Holmium	455, 650, 1150	2040	0.6	3
Erbium	800, 975, 1480	1550 (1530–1610)	10	3
Thulium	790, 1210 (1060–1320)	1480 (1460–1520)		
		1850 (1650–2050)	0.3	3
Ytterbium	920 (840–980)	975	0.8	3
		1040 (1010–1160)		4

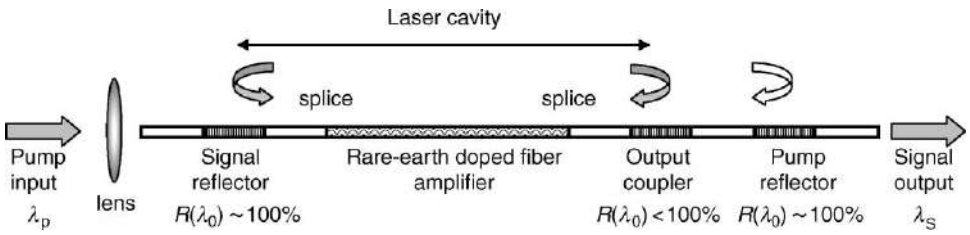

Figure 1 Example of a simple all-fiber optical oscillator. The rare-earth doped fiber provides optical gain, and the Bragg gratings provide narrowband optical feedback.

inhomogeneous linewidths of several nanometers, and can be optimized to provide around 30 dB gain over an 80 nm bandwidth centered around 1570 nm. The large gain-bandwidth is useful for achieving either wide tunability of the lasing wavelength and/or ultrashort pulse generation. Table 1 summarizes the most commonly used pump and signal transitions for a range of rare-earth dopants in silicate optical fibers.

Other desirable properties of silica fiber lasers include high efficiency, excellent thermal dissipation, substantial energy storage, high power capability, and compatibility with a wide range of optical fiber devices and systems. The characteristics of optical fiber lasers can be optimized for some very different applications, ranging from narrow linewidth and low-noise lasers to broadband pulsed lasers with high energy and/or high peak power. In the following sections the fundamentals of laser theory as applied to fiber lasers will be reviewed, highlighting the reasons underlying such versatility.

Fiber Laser Fundamentals

An example of a simple all-fiber laser is shown schematically in Figure 1. The laser contains the basic elements of optical gain (e.g., a rare-earth doped fiber amplifier) and optical feedback (e.g., a pair of Bragg grating filters). In the laser shown, the two Bragg gratings provide optical feedback only over a narrow range of wavelengths, which further restricts the range of possible lasing wavelengths. One grating must have a reflectivity less than unity to allow a proportion of the lasing signal to be coupled out of the laser cavity. At threshold, the gain in the amplifier exactly compensates for the loss in the output coupler.

The optical pump, typically from a semiconductor laser diode, is absorbed as is propagates along the fiber amplifier, e.g., by rare-earth ion dopants, which are raised to an excited state. The absorbed energy is stored and eventually emitted at a longer wavelength by either spontaneous or stimulated emission. Spontaneous emission produces unwanted noise and is associated with a finite excited state lifetime, whilst stimulated emission produces optical gain provided the gain medium is 'inverted', i.e., with more active ions in the upper lasing energy level than the lower lasing energy level. In practice some energy is also lost to nonradiative (thermal) emissions, e.g., in the transfer of energy between the pump and upper lasing energy levels. In an ideal gain medium the nonradiative transfer rate between the pump and upper lasing energy levels is much larger than the spontaneous or stimulated emission rates, so the inversion may usually be assumed to be independent of nonradiative transfer rates. Energy level diagrams for three- and four-level atoms are shown in Figure 2.

The local optical gain coefficient (i.e. power gain per unit length, in nepers[2]) in a fiber amplifier is proportional to the local inversion, i.e., $g(z) \sim \sigma_S \Delta N(z)$, where σ_S is the emission cross-section at the

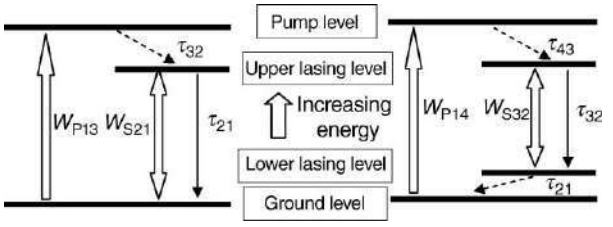

Figure 2 Energy level diagrams for three- and four-level laser transitions. *W* represents the probability of an optical transition per unit time associated with stimulated absorption or emission, and $1/\tau$ is the rate of spontaneous emission. Dotted lines represent nonradiative transitions (e.g., between the pump and upper lasing level).

signal wavelength, and $\Delta N(z)$ the difference in population density between the upper and lower lasing energy levels ($\Delta N > 0$ if the gain medium is inverted). The local inversion can be calculated by solving the rate equations including all stimulated and spontaneous transitions between the laser energy levels. The end-to-end amplifier power gain, $G = P_S(L_A)/P_S(0)$, in length L_A of amplifying fiber is then

$$G = \exp\left[\int_0^{L_A} g(z)\,\mathrm{d}z\right] \qquad [1]$$

which in decibels is

$$G = 4.34\int_0^{L_A} g(z)\,dz$$

A large gain coefficient is desirable for amplification with minimal added noise (i.e., low noise figure), and a large gain per unit pump power is desirable for low threshold lasing. For a given gain medium the only factor usually accessible to control the inversion and hence gain is the pump rate, i.e., the rate at which pump energy is absorbed by the gain medium, which for an unbleached amplifier is proportional to pump intensity. For example, in an ideal four-level unsaturated laser medium with uniform pump intensity, the small signal gain is approximately

$$\int_0^{L_A} g(z)\,\mathrm{d}z \approx \sigma_S \frac{P_{\mathrm{Pabs}}\tau}{h\nu_P A} \qquad [2]$$

in which τ is the lifetime of the upper lasing level, P_{Pabs} is the pump power absorbed in the fiber amplifier with cross-sectional area A, h is Planck's constant, and ν_P is the pump frequency.

For a three-level laser under the same conditions, the small signal gain is

$$\int_0^{L_A} g(z)\,\mathrm{d}z \approx -\sigma_S N L_A + \sigma_S \frac{2P_{\mathrm{Pabs}}\tau}{h\nu_P A} \qquad [3]$$

in which the first term corresponds to ground state absorption of the signal, in which N is the dopant ion density, and the factor of 2 in the second term is due to the fact that in an ideal three-level system every absorbed pump photon increases the inversion, ΔN, by two. Equation [3] can be approximated by multiplying [2] by the factor $(\mathrm{P_{Pabs}} - \mathrm{P_{Psat}})/(\mathrm{P_{Pabs}} + \mathrm{P_{Psat}})$, where $P_{\mathrm{Psat}} = h\nu_P A/(\sigma_P\tau)$ is the pump power which must be absorbed to reduce the pump absorption coefficient to half its unpumped value (or the signal absorption coefficient to zero), and σ_P is the pump absorption cross-section. Unlike four-level lasers in which gain is available as soon as pump power is applied, three-level lasers require half the active ions to be excited before the inversion becomes positive and gain is produced.

The main advantages of optical fiber amplifiers are now clear; in an optical fiber waveguide the light is confined in a very small cross-sectional area, typically less than ten microns diameter. Consequently even a small coupled pump power can have a large intensity and produce a large gain coefficient, or low laser threshold. Even if the pump absorbed per unit length in the fiber amplifier is small (e.g., due to low dopant concentration or low absorption coefficient), this can often be compensated by using a long length of fiber with small intrinsic loss. Furthermore, rare-earth dopants have relatively long metastable state lifetimes, τ, which further assists in producing a large inversion throughout the fiber, and hence large overall gain or low lasing threshold, even at low pump powers. For example, a typical four-level neodymium doped fiber amplifier providing gain at $\lambda_S = 1.06\ \mu\mathrm{m}$, pumped at $\lambda_P = 0.8\ \mu\mathrm{m}$, with core diameter $7\ \mu\mathrm{m}$, $\tau = 0.5$ ms, $\sigma_S = 1.4 \times 10^{-20}\ \mathrm{cm}^2$, and $\sigma_P = 23 \times 10^{-21}\ \mathrm{cm}^2$, could theoretically provide a small signal gain of 0.3 dB for every milliwatt of absorbed pump power.

Whilst the above analysis highlights the main advantages of fiber amplifiers, it is approximate in a number of respects. For example, eqns [2] and [3] assume both pump and signal have a uniform intensity distribution across the gain medium, and completely overlap. In single mode optical fibers the intensity distribution across the fiber core is not uniform but approximately Gaussian, with a width and effective area that depends on the numerical aperture of the waveguide and the wavelength. Consequently the overlap of the pump and signal beams with the doped fiber core and with each other is less than 100%, and the pump and signal intensities are higher in the center of the fiber than at the core-cladding boundary. The nonuniform distribution of pump and signal beams slightly modifies both the gain efficiency and gain saturation behavior

(i.e., the dependence of gain on signal power) in fiber amplifiers. More accurate analysis would also take into account factors such as the rare-earth dopant distribution (which may not be uniform), the variation of pump and signal intensities along the fiber, spectral variation in the absorption and emission cross-sections, spatial and spectral hole-burning in the gain medium, degeneracies in the energy levels, excited state absorption and other loss mechanisms, temperature, and spontaneous emission noise.

Continuous Wave Fiber Lasers

The ideal continuous wave (cw) laser converts input pump power with low coherence to a highly coherent optical output signal which is constant in amplitude and wavelength, with the spectral purity of the output (i.e., linewidth) limited only by cavity losses. Practical continuous wave lasers may be characterized by four parameters: the pump power required for the onset of oscillation (i.e., threshold); the conversion efficiency of pump to signal power above threshold; the peak wavelength of the optical output; and the spectral width of the optical output. Continuous wave fiber lasers often have a low threshold, high efficiency, and narrow linewidth relative to other types of laser. Output powers approaching 100 watts have been achieved using specialized techniques.

The threshold power is determined by the total cavity loss and gain efficiency (i.e., gain per unit of absorbed pump power) of the fiber amplifier. For example, for the laser shown in Figure 1 the internal losses are minimal, and the pump power required to reach threshold may be calculated by setting the product of the round-trip small-signal gain, and the output mirror reflectivity, G^2R, equal to unity. For the laser configuration shown in Figure 1, and using eqns [1] and [2] for a four-level laser, the pump power which must be absorbed in the amplifier for the laser to reach threshold is

$$P_{\mathrm{th}} = \alpha_{\mathrm{cav}} L_{\mathrm{C}} \frac{h\nu_{\mathrm{P}} A}{\sigma_{\mathrm{S}} \tau} \qquad [4]$$

in which L_{C} is the cavity length, and $\alpha_{\mathrm{cav}} = \alpha_{\mathrm{int}} + \frac{1}{2} L_{\mathrm{C}} \ln(1/R)$ is the total cavity loss per unit length, comprising both internal losses, α_{int}, and outcoupling losses through a mirror with reflectivity R. For example, if the internal losses were negligible and the transmittance, $T = 1 - R$, of the output reflector in the laser shown in Figure 1 was 50% (i.e. 3 dB), and the other fiber parameters were the same as above (i.e. $G = 0.3$ dB/mW), then the laser threshold would be 5 mW. In three-level fiber lasers ground state absorption of the signal is usually the dominant loss mechanism which must be overcome before a net gain is produced, and hence the threshold of three-level lasers is given approximately by eqn [4] with $\alpha_{\mathrm{int}} = \sigma_{\mathrm{S}} N L_{\mathrm{A}} 2 L_{\mathrm{C}}$, and increases with amplifier length. In either case the pump power required at the input of the laser to reach threshold, $P_{\mathrm{P}}(0)$, may be approximated using the relation $P_{\mathrm{Pabs}} = P_{\mathrm{P}}(0) - P_{\mathrm{P}}(L_{\mathrm{A}}) \approx P_{\mathrm{P}}(0)[1 - \exp(-\sigma_{\mathrm{P}} N L_{\mathrm{A}})]$.

For either three- or four-level lasers pumped above threshold, the inversion of the gain medium remains at its threshold level as the internal round-trip gain must remain unity; however a coherent optical output signal at the peak emission wavelength grows from noise until its amplitude is limited by saturation of the amplifier gain. Provided the output coupling loss is not too large (e.g., $T < 60\%$) such that the total signal intensity is constant along the whole length of the fiber amplifier, it can be shown that changes in pump power above threshold cause a proportional change in the signal power coupled out of the Fabry–Perot cavity in one direction, P^{+}_{Sout}, i.e.

$$P^{+}_{\mathrm{Sout}} = \frac{(1-R)}{\delta_0} \frac{\nu_{\mathrm{S}}}{\nu_{\mathrm{P}}} (P_{\mathrm{Pabs}} - P_{\mathrm{th}}) \qquad [5]$$

in which $\delta_0 = 2\alpha_{\mathrm{int}} L + \ln 1/R$ represents the total round-trip loss at threshold. The relationship between absorbed pump power and output signal power is shown schematically in Figure 3. The constant of proportionality is known as the slope or conversion efficiency, defined as $\eta_{\mathrm{S}} = P_{\mathrm{Sout}}/(P_{\mathrm{Pabs}} - P_{\mathrm{th}})$. In both three- and four-level fiber lasers with small intrinsic losses $\alpha_{\mathrm{int}} \approx 0$, and small output coupling (i.e., $R \approx 1$), the slope efficiency can approach the intrinsic quantum efficiency, $\nu_{\mathrm{S}}/\nu_{\mathrm{P}}$.

Sustained oscillation can only occur if the optical path length around the cavity is an integral number of wavelengths. Wavelengths satisfying this condition are called 'modes' of the cavity, determined by

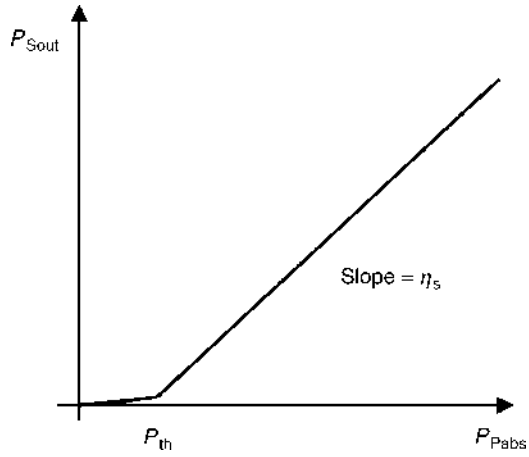

Figure 3 Representative plot of laser output power versus absorbed pump power.

$2nL_C = m\lambda_0$, where n is the effective refractive index seen by light propagating in the cavity, L_C is the cavity length, m is an integer, and λ_0 is the free-space wavelength. In Figure 1, the cavity length comprises the amplifying fiber, and two Bragg grating filters which may be regarded as having an effective length (i.e., less than the actual grating length) which depends on the grating characteristics. In fiber lasers the cavity is typically between several centimeters and many meters in length, hence there are usually very many potential lasing modes within the gain-bandwidth of the amplifying medium. For example, in the Fabry–Perot cavity of Figure 1, if the cavity length was 5 meters, the mode spacing would be $\nu_{m+1} - \nu_m = c/(2nL_C) = 20$ MHz, i.e. orders of magnitude less than the bandwidth of the optical amplifier.

Ideally the first mode to reach threshold would determine the laser output wavelength and the spectral width, or linewidth, of the laser output would be $\Delta\nu_L = h\nu/(2\pi\tau_C^2 P_{Sout})$, in which τ_C is the lifetime of photons in the cavity and determined by cavity losses, including outcoupling, and given by $\tau_C = t_R/\varepsilon$, where t_R is the cavity round trip time and ε the fractional energy loss in the cavity per round trip. In practice, the linewidth is usually significantly larger than the latter theoretical limit due to transient lasing of different modes within the bandwidth of the optical amplifier and environmental perturbations to the fiber cavity. Fiber amplifiers typically have a large homogeneously broadened gain-bandwidth, and being long and flexible are susceptible to acoustic and thermal perturbations, hence care must be taken to physically stabilize the laser cavity to minimize both mode-hopping and the output linewidth.

Whilst the CW laser shown in Figure 1 has the advantage of simple construction, it would not be ideal for narrow linewidth CW generation. The reason is that counterpropagating waves in the Fabry–Perot cavity form a standing wave, i.e., in which the local signal intensity varies due to interference between forward and backward propagating waves. The standing wave can cause spatial hole-burning in the gain medium (i.e., a spatial variation in gain, linked to the spatial variation in intensity) which in turn reduces the average gain for the lasing mode and promotes mode-hopping and hence spectral broadening of the output.

A preferable arrangement for many fiber lasers is a traveling wave laser, which can be realized using a ring configuration, as shown in Figure 4. In this configuration the optical isolator ensures unidirectional lasing and avoidance of spatial hole-burning. Additional components in the cavity can include a wavelength selective coupler (WSC), i.e., to couple the pump wavelength into the laser cavity whilst not coupling the signal wavelength out, and a narrowband filter (e.g. a small Fabry–Perot resonator) to further stabilize and/or narrow the lasing linewidth. Rare-earth doped fiber lasers with similar cavity configurations have been realized with linewidths of approximately 10 kHz, limited by environmental (e.g., acoustic) perturbations.

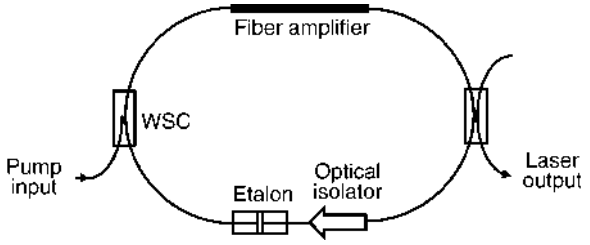

Figure 4 Schematic of a traveling wave laser, useful for low noise continuous wave oscillation. The wavelength selective coupler (WSC) is required to couple pump power into the laser cavity without coupling signal out.

Pulsed Fiber Lasers

There are two main types of fiber laser useful for generating short high-power pulses: Q-switched lasers, and mode-locked lasers. Q-switched fiber lasers are useful for generating large energy pulses (e.g., microjoules) with very high peak power (e.g., kilowatts) with relatively long pulse duration (e.g., tens of nanoseconds), whilst mode-locked fiber lasers are typically capable of generating ultrashort pulses (e.g., sub-picosecond) with moderate energy (e.g., nanojoules) and moderate-to-high peak power (e.g., tens of watts). Pulsed lasers typically produce less average output power than CW lasers (e.g., up to about 10 W). Both types of pulsed laser typically contain an element for loss modulation within the cavity, however the pulse generation mechanisms are very different.

Q-switched lasers operate by rapid switching of the cavity loss. Whilst the intracavity loss is high and the laser below threshold, energy is transferred from the pump to the lasing medium. The long excited state lifetime and small emission cross-sections of rare-earth dopants assists greatly in the latter process, so that a significant amount of energy can be stored in long fiber amplifiers before amplification of spontaneous emission noise begins to deplete the inversion. After the gain medium is fully inverted the cavity loss is suddenly reduced, and the laser is taken well above threshold, resulting in a rapid build-up of noise and the formation of a pulse which ideally extracts all the stored energy within a few round-trips in the laser cavity. The situation is shown schematically in Figure 5.

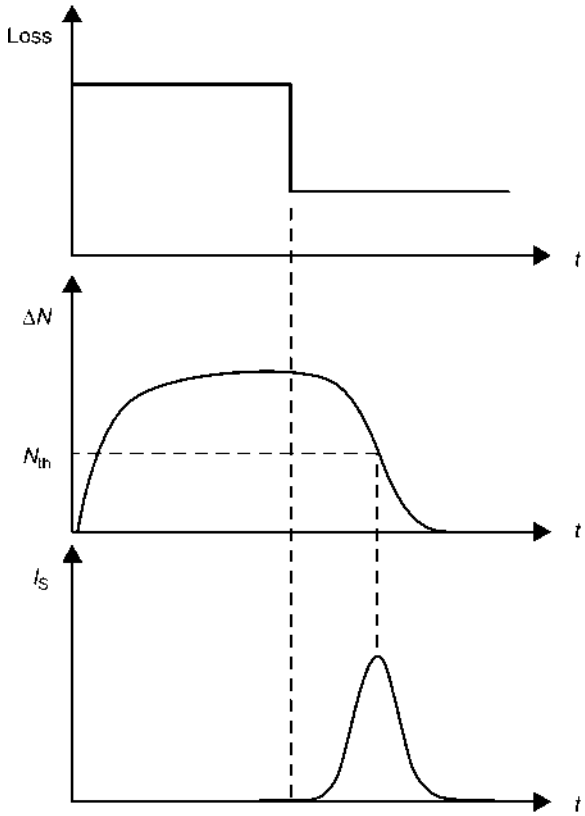

Figure 5 Pulse generation in a *Q*-switched laser. Initially the intracavity loss is high, the laser remains below threshold, and the inversion of the gain medium increases. When the intracavity loss is suddenly reduced, the laser goes well above threshold and spontaneous emission noise is rapidly amplified into a large pulse which extracts much of the energy previously stored in the gain medium.

An upper limit on the pulse energy obtainable from a *Q*-switched laser may be determined by assuming all active ions in the laser cavity are excited before the pulse is generated (i.e., at the time the cavity Q is switched high), that no ions remain excited immediately after the pulse is generated, and that the pulse generated is much shorter than the time required to pump all the ions into their excited state. The energy stored in the gain medium, and hence the maximum conceivable pulse energy, would then be $E = h\nu_S NV$, where N is the density of active ions and V is the cavity volume. For example, a typical erbium-doped fiber laser of length 10 m and dopant density $N = 10^{19}$ ions/cm in a single-mode fiber with core diameter 7 μm could ideally store up to 500 μJ of energy.

Two things limit the maximum inversion and energy storage achievable in practice: spontaneous emission noise (due to the finite excited-state lifetime) which is amplified and depletes the inversion, and unwanted feedback (e.g., due to the finite extinction ratio of the Q-switching element) which results in cw lasing which limits the inversion. Following from eqns [1] and [2], the unidirectional gain per unit of stored energy in a fiber laser may be expressed as $G = 4.34/(A_{\text{core}}F_{\text{sat}}) = 4.34/E_{\text{sat}}$ decibels per joule, where $F_{\text{sat}} = h\nu_S/\sigma_S$ is the saturation fluence, and $E_{\text{sat}} = A_{\text{sat}}F_{\text{sat}}$ is the saturation energy. For example, if the maximum gain achievable before the onset of cw lasing was 30 dB (i.e., even when the Q-switch introduces a large loss), then the maximum energy storable in the gain medium would be $E = 6.9E_{\text{sat}}$. For the same erbium-doped fiber parameters as used in the previous paragraph, the maximum storable energy would be $E \approx 70$ μJ.

Only a proportion of the energy stored in the gain medium may usually be converted into a Q-switched output pulse. The actual Q-switched pulse energy, duration, and peak power can be calculated by solving the rate equations for the photon flux and the population difference between the lasing energy levels in the gain medium. The equations must usually be solved numerically, however assuming the Q changes much more rapidly than the pulse build-up time, which is in turn much shorter than the time taken for a significant change in inversion to either pumping or spontaneous emission, the pulse parameters may be approximated analytically and are found to be determined by only two parameters, i.e., the population inversion (or stored energy) just before Q-switching, and the lifetime of photons within the cavity, τ_C, defined previously.

Under the conditions described above, standard analysis gives the Q-switched pulse duration as

$$T_P \approx \frac{\tau_C(r-1)}{(r-1-\ln r)} \qquad [6]$$

where r is the ratio of the inversion just before Q-switching (low Q) to the threshold inversion for cw lasing after Q-switching (Q high). In fiber lasers the gain can be very high, hence it is common for $r \gg 1$ and almost all available energy stored in the gain medium is extracted as a Q-switched laser pulse. The Q-switched pulse duration is then determined primarily by the cavity lifetime; shorter pulses can be obtained only by shortening the laser cavity and/or increasing the cavity loss. For example, for a 10 m long ring cavity in which the sum of cavity and outcoupling losses is 10 dB and $r \gg 1$, the pulse duration would be $T_P\tau_C = 22$ ns.

The exact proportion of energy stored in the gain medium that is converted to a Q-switched pulse depends on the number of energy levels in the lasing medium, the degeneracy of those levels, and the rate

of nonradiative energy transfer between those levels. For example, for an ideal three-level gain medium (e.g., erbium dopant) with no internal cavity losses the maximum available Q-switched pulse energy is half the stored energy, or $E_{\text{pulse}} = E/2$. It can be shown that the peak of the output pulse power occurs as the inversion decreases through threshold, and may be approximated by dividing the pulse energy by the pulse duration. Even for moderate pulse energies the peak power is potentially very high. For the erbium-doped fiber laser parameters used previously, the peak pulse power would be approximately 1600 W, i.e., sufficient to cause significant spectral broadening of the output wavelength due to stimulated Raman scattering and other nonlinear effects within the optical fiber.

Mode-locked lasers generate optical pulses by forcing many cavity modes to be locked in phase. As per Fourier theory, the pulse repetition rate is related to the spacing of the lasing modes, and the more modes that add coherently, the greater the pulse bandwidth and the shorter the transform-limited pulse duration. For example, the relationship between mode spacing and pulse repetition rate, and between the spectral and temporal envelopes of a train of Gaussian pulses is shown in Figure 6.

In mode-locked lasers a nonlinear or time-varying element must be present which synchronously modulates the optical signal circulating in the cavity in amplitude or phase at a frequency which is a harmonic of the cavity mode spacing. There are two main methods of mode locking:

(i) active mode locking (e.g., by an electro-optic modulator), in which amplitude or phase modulation is driven by an external source; and
(ii) passive mode locking (e.g., by a fast or slow saturable absorber), in which an element with nonlinear gain or loss responds to the optical signal in the laser cavity itself.

In both cases the modulation provided by the mode-locking element distributes energy between neighboring modes, effectively injection-locking them in phase. A variation of mode locking techniques often present in fiber lasers is soliton mode-locking, in which the active or passive modulation is weak, and a balance between nonlinearity and dispersion is the dominant process underlying pulse formation. Hybrid mode-locking techniques can also be used to combine the best aspects of different pulse-generation methods.

Whilst the latter frequency-domain view of mode locking is helpful, nonlinear effects are almost always significant in mode-locked fiber lasers, and hence the linear concept of a mode is not appropriate for detailed analysis. Hence it is usually preferable to describe pulse generation in mode-locked fiber lasers in the time domain; for stable operation the pulse at any point in the cavity must self-replicate after each round-trip, with the net effects of temporal and spectral loss, gain, dispersion and nonlinearity all canceling.

An optical pulse propagating in a fiber laser is usually described by a slowly varying complex envelope, $\Psi(z,t)$, where $|\Psi(z,t)|^2$ is the instantaneous power in a frame of reference moving at the pulse group velocity, and $\arg[\Psi(z,t)]$ is the instantaneous phase relative to the optical carrier. The various elements in the laser cavity, such as the passive fiber, optical fiber amplifier, filters, and mode-locking devices, each modify the complex pulse envelope as it propagates around the laser cavity.

For example, in the absence of other effects, group velocity dispersion (GVD) causes different wavelengths to travel at different velocities along the fiber, resulting in 'chirping' and temporal broadening of optical pulses by an amount proportional to the distance propagated. The effect can be described mathematically as the accumulation of a phase shift which is quadratic in frequency, i.e., $\Psi(z,\omega) = \Psi(0,\omega)\exp[\mathrm{i}\beta_2\omega^2 z/2]$. Taking the derivative with respect to propagation distance, and using $\partial/\partial t = -\mathrm{i}\omega$, the incremental change in the complex envelope

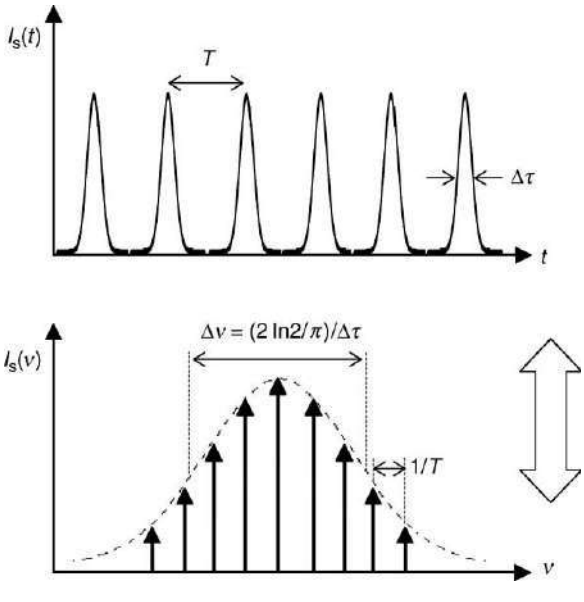

Figure 6 Schematic showing the Fourier-transform relationship between an infinite train of Gaussian pulses and the lasing modes in a mode-locked laser. The temporal and spectral widths are measured full width at half the maximum intensity.

with propagation distance is

$$\frac{\partial \Psi}{\partial z} = -\mathrm{i}\frac{\beta_2}{2}\frac{\partial^2 \Psi}{\partial t^2} \qquad [7]$$

If the GVD parameter, β_2, is positive then long wavelengths propagate faster than shorter wavelengths, and the dispersion is called 'normal'. If the reverse is true, the dispersion is 'anomalous'. In standard single-mode step index optical fibers the dispersion passes through zero at wavelengths around 1300 nm, and is anomalous for longer wavelengths. It is often necessary to control the net dispersion in a fiber laser, and this can be done by incorporating specially designed fibers with wavelength-shifted or otherwise modified dispersion characteristics, or by using chirped Bragg grating reflectors.

Whilst silica is not commonly regarded as a nonlinear medium (i.e., in which the optical properties are a function of the intensity of light), in ultrashort pulsed fiber lasers the peak powers are usually such that nonlinear effects cannot be ignored. In silica the main nonlinearity is associated with the Kerr effect, and is readily observed as an intensity-dependent refractive index, i.e. $n = n_0 + n_2 I$, where the nonlinear index, n_2, is typically about 3×10^{-20} m^2/W in silica fibers. Other nonlinear effects related to the Kerr nonlinearity which can sometimes be important are cross-phase-modulation and four-wave mixing. In the absence of other effects, the nonlinear index results in self-phase modulation (SPM), or a phase shift in the optical carrier which is proportional to the instantaneous intensity of the signal. Mathematically, $\Psi(z,t) = \Psi(0,t)\exp[\mathrm{i}\gamma|\Psi(0,t)|^2 z]$, in which $\gamma = n_2\omega_0/(cA)$ is the nonlinear coefficient of the fiber, ω_0 is the radian frequency of the optical carrier, and A is the effective core area for the guided mode. The incremental change per unit length in the complex pulse envelope induced by self-phase modulation is then

$$\frac{\partial \Psi}{\partial z} = \mathrm{i}\gamma|\Psi|^2\Psi \qquad [8]$$

For example, self-phase-modulation is used to realize a nonlinear switching element in the passively mode-locked laser cavity shown schematically in Figure 7. For obvious reasons the laser is called a 'figure-8' laser, though it is actually a ring laser into which a device known as a nonlinear amplifying loop-mirror (NALM) has been incorporated. The NALM acts as a fast saturable absorber, with transmittance dependent on the instantaneous optical intensity. The NALM operates as follows: light propagating through the optical isolator is divided equally into co- and counter-propagating waves in the loop-mirror on the right-hand side. The light propagating in a clockwise direction is amplified immediately, and propagates around the loop with high intensity, whilst the counter-clockwise propagating signal has a low intensity for most of its time in the loop. The difference in intensity means that the co- and counter-propagating waves experience different nonlinear phase shifts. When they are recombined at the coupler, high-intensity light is transmitted back into the ring, whilst low-intensity light is reflected back to the isolator, and lost. In this way the NALM provides both gain and a nonlinear transmittance which depends on the instantaneous power of the input, described by

$$T \approx G\left\{1 - \frac{1}{2}\left[1 + \cos\left(\frac{1}{2}\gamma L P_{\mathrm{in}} G\right)\right]\right\} \qquad [9]$$

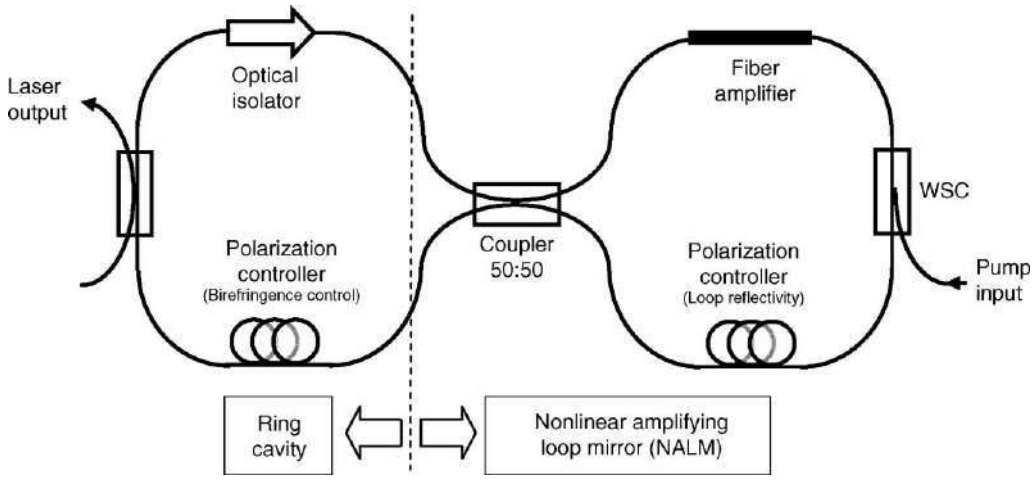

Figure 7 Schematic of a 'figure-8' fiber laser. The laser is passively mode-locked by the nonlinear amplifying loop mirror on the right, which acts as a fast saturable absorber, transmitting and amplifying high-power optical signals whilst reflecting low optical power signals.

in which G is the linear gain of the amplifier, and L the loop length, P_{in} is the input power, and T the ratio of transmitted power to input power. The nonlinear transmittance of the NALM promotes pulse formation and shortening in the laser cavity, and hence passively mode-locks the laser. The polarization controllers are to compensate for unwanted birefringence in the optical fiber, and to control the transmittance of the loop-mirror at low input powers. The fiber dispersion and nonlinearity, together with other factors, further influence the pulses generated in the laser, as described in the following paragraphs.

The combined effects of dispersion and self-phase-modulation can give rise to some interesting and useful physical phenomena. In lossless optical fibers, or fiber lasers in which loss is balanced by gain, pulse propagation is described by the nonlinear Schrödinger equation (NLSE), i.e., a nonlinear wave equation including the effects of both group velocity dispersion and intensity-dependent refractive index. The NLSE may be derived by combining eqns [8] and [9], i.e.

$$\mathrm{i}\frac{\partial \Psi}{\partial z} - \frac{\beta_2}{2}\frac{\partial^2 \Psi}{\partial t^2} + \gamma|\Psi|^2\Psi = 0 \qquad [10]$$

It can be shown that if the dispersion is anomalous ($\beta_2 < 0$) then for a particular shape of pulse envelope the distributed effects of dispersion and nonlinearity can exactly cancel, so $\partial\Psi/\partial z = 0$, and the pulse propagates without changing shape. Such pulses are called fundamental solitons, the solution for which is $\Psi(z,t) = \Psi_0\,\mathrm{sech}(t/T_0)\exp[\mathrm{i}z/(2L_D)]$, in which the peak amplitude is $\Psi_0 = 1/\sqrt{\gamma L_D}$, T_0 is the pulse width parameter, and $L_D = T_0^2/\beta_2$ is known as the dispersion length. For example, in a standard silica fiber at $\lambda = 1.55\ \mu\mathrm{m}$, $\beta_2 \approx -20\ \mathrm{ps}^2/\mathrm{km}$ and $\gamma \approx 10\ \mathrm{W}^{-1}\mathrm{km}^{-1}$, hence a soliton with $T_0 = 1$ ps would have peak power 2 W and energy $E_0 = 2|\beta_2|/(\gamma T_0) = 4$ pJ. The energy of a soliton is quantized, in that pulses with initial energy between 0.5 and 1.5 times E_0 can evolve into a fundamental soliton during propagation. Consequently solitons behave somewhat like particles, and are robust even in the presence of significant perturbations. The pulses which form spontaneously in the figure-8 fiber laser have many of the properties of solitons.

Mode-locked fiber lasers, such as that shown in Figure 7, can often be described by a generalized nonlinear Schrödinger equation, in which extra terms are included in eqn [10] to account for the average effects of an excess linear gain or loss, nonlinear gain and loss (e.g., due to a saturable absorber), spectral filtering, and temporal modulation, etc. Assuming the effects of the added elements can be regarded as distributed over the length of the cavity rather than lumped, such that the envelope of any pulse generated in the laser does not change shape significantly as it propagates around the cavity, then the master differential equation describing propagation of an optical signal within a fiber laser which is passively mode-locked by a fast saturable absorber can be written

$$\mathrm{i}\frac{\partial U}{\partial \zeta} + \left(\frac{D}{2} - \mathrm{i}\beta\right)\frac{\partial^2 U}{\partial \tau^2} + (1 - \mathrm{i}\varepsilon)|U|^2 U = \mathrm{i}\delta U - (\nu - \mathrm{i}\mu)|U|^4 U \qquad [11]$$

Equation [11] is known as the Ginzburg–Landau equation, which describes propagation of the normalized pulse envelope, $U(\zeta,\tau)$, as a function of delayed time, τ, and normalized distance, ζ, and under the combined influence of group velocity dispersion ($D = \pm 1$ for anomalous or normal group velocity dispersion, respectively), dispersion due to band-limiting by amplification in a homogeneously broadened gain medium ($\beta > 0$), nonlinear refractive index, nonlinear gain or loss (ε), linear gain (δ), and quintic terms, μ and ν, which if less than zero relate to saturation of the nonlinear gain and refractive index, respectively. Note that if the right-hand side of the equation is set to zero, and $\varepsilon = \beta = 0$, then the equation reduces to the normalized form of the nonlinear Schrödinger equation. Stable pulses generated within the laser are solutions of the above equation with $\partial U/\partial\zeta = 0$, and hence the effect of all terms must balance. Consequently it is possible for stable pulses to be generated even with normal dispersion in the cavity, though the pulses are then generally strongly chirped, and have an envelope which is approximately Gaussian shaped rather than the hyperbolic secant shape expected with anomalous dispersion. Stable soliton-like pulses can be found by solving the Ginzburg–Landau equation for specific ranges of parameters. There are many numerical solutions, and some analytical solutions, but in general the dynamics of pulse formation and propagation in passively mode-locked fiber lasers can be quite complex, and for some sets of parameters remarkable and unexpected numerical solutions can be found. One such example which has been observed experimentally is the 'exploding' soliton, which unlike the usual soliton is periodically unstable, as shown in the plot of pulse evolution in Figure 8.

Other Fiber Lasers

The lasers discussed in the previous sections are indicative of the main types of fiber laser in terms of their output characteristics; however it should be

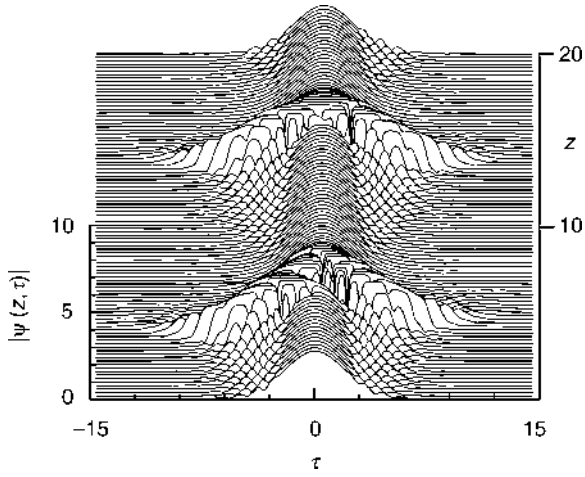

Figure 8 Two periods of the evolution of an exploding soliton found by numerical solution of the quintic Ginzburg–Landau equation with normalized parameters $\varepsilon = 1.0$, $\delta = 0.1$, $\beta = 0.125$, $\mu = 0.1$, and $\nu = 0.6$. The process never repeats itself exactly in successive 'periods'; however the pulse always returns to the same shape. Such solutions cannot be found in analytic form; nevertheless they are as common as stationary pulse solutions and exist for a wide range of parameters.

noted that there is a wide range of other fiber laser configurations with similar behaviors incorporating a variety of other fiber and/or bulk elements. For example, specially designed fibers in which the pump and/or signal has a large cross-sectional area can be used to realize fiber lasers with a high output power, comparable to that in many bulk laser systems. Linear and/or nonlinear processing of the output of fiber lasers can be used to increase their peak output power and/or pulse energy, or to change their output wavelength (i.e., by harmonic generation) from that achievable in the laser cavity itself. In mode-locked fiber lasers it is now common to include passive semiconductor devices which provide a slow saturable absorber action to initiate and stabilize soliton mode-locking.

It should also be noted that there are a number of fiber lasers which generate optical outputs using different or modified physical processes compared to those described in previous sections. For example, it is possible to realize up-conversion fiber lasers in which the output wavelength is shorter than the pump wavelength, multiwavelength fiber lasers (pulsed and cw) which produce outputs at several wavelengths simultaneously, and lasers with limited output coherence such as chaotic fiber lasers and broadband superfluorescent sources, etc. Lastly, whilst the rare-earth doped fiber lasers dealt with in previous sections have some very desirable properties, it should be recognized that optical fiber lasers based on other optical gain mechanisms, such as stimulated scattering and parametric effects, can also be realized (e.g., as in fiber Raman lasers, and fiber Brillouin lasers). In the latter cases the lasing wavelength is limited only by available pump wavelengths. The reader is referred to the references in the Further Reading for information on the wide range of fiber lasers that are possible in practice.

List of Units and Nomenclature

Absorption cross-section	σ_a	[m^{-2}]
Amplifier length	L_A	[m]
Area	A	[m^2]
Cavity length	L_C	[m]
Complex amplitude	Ψ	[W$^{0.5}$]
Dispersion length	L_D	[m]
Distance	z	[m]
Emission cross-section	σ_e	[m^{-2}]
Frequency	ν	[Hz]
Gain coefficient	g	[m^{-1}]
Group velocity dispersion	β_2	[s^{2}m^{-1}]
Inversion ratio	r	
Lifetime	τ	[s]
Loss coefficient	α	[m^{-1}]
Nonlinear coefficient	γ	[m^{-1}W^{-1}]
Nonlinear refractive index	n_2	[m^{2}W^{-1}]
Number density	N	[m^{-3}]
Power gain	G	
Pump power	P_P	[W]
Pump wavelength	λ_P	[m]
Reflectivity	R	
Saturation energy	E_{sat}	[J]
Saturation fluence	F_{sat}	[J.m^{-2}]
Signal power	P_S	[W]
Signal wavelength	λ_S	[m]
Threshold power	P_{th}	[W]

See also

Fiber and Guided Wave Optics: Dispersion; Fabrication of Optical Fiber; Light Propagation; Nonlinear Effects (Basics); Optical Fiber Cables. **Fiber Gratings**. **Lasers:** Up Conversion Lasers. **Nonlinear Optics, Applications:** Raman Lasers. **Optical Amplifiers:** Basic Concepts; Erbrium Doped Fiber Amplifiers for Lightwave Systems. **Optical Parametric Devices:** Optical Parametric Oscillators (Pulsed). **Solitons:** Temporal Solitons. **Scattering:** Raman Scattering; Stimulated Scattering. **Ultra-Fast Laser Techniques:** Generation of Femtosecond Pulses.

Further Reading

Agrawal GP (2001) *Nonlinear Fiber Optics*, 3rd edn. San Diego, CA: Academic Press.

Agrawal GP (2001) *Applications of Nonlinear Fiber Optics*. San Diego, CA: Academic Press.
Akhmediev NN and Ankiewicz A (1997) *Solitons*. London: Chapman and Hall.
Becker PC, Olsson NA and Simpson JR (1999) *Erbium-Doped Fiber Amplifiers: Fundamentals and Technology*. San Diego, CA: Academic Press.
Desurvire E (1994) *Erbium-Doped Fiber Amplifiers: Principles and Applications*. New York: Wiley.
Digonnet MJF (ed.) (2001) *Rare-Earth-Doped Fiber Lasers and Amplifiers*, 2nd edn. New York: Marcel Dekker.
Duling IN (ed.) (1995) *Compact Sources of Ultrashort Pulses*. Cambridge, UK: Cambridge University Press.
France PW (ed.) (1991) *Optical Fibre Lasers and Amplifiers*. Glasgow, UK: Blackie.
Jeunhomme LB (1990) *Single-Mode Fiber Optics: Principles and Applications*, 2nd edn. New York: Dekker.
Kashyap R (1999) *Fiber Bragg Gratings*. San Diego, CA: Academic Press.
Siegman AE (1986) *Lasers*. Oxford, UK: Oxford University Press.
Weber MJ (ed.) (1995) *CRC Handbook of Laser Science and Technology*. Boca Raton, FL: CRC Press.

Organic Semiconductors and Polymers

G A Turnbull, University of St Andrews, Fife, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Organic semiconductors, and conjugated polymers in particular, are emerging as an interesting new class of visible laser media. The polymers are electroluminescent, broadband visible emitters and exhibit large optical gain coefficients. As plastics, they are amenable to simple processing, and can be shaped into complex structures. This combination of photonic, electronic, and processing properties make these materials extremely well suited to forming novel microlasers, with future prospects as inexpensive, plastic diode lasers operating throughout the visible spectrum.

Conjugated Polymers

Plastics are widely used in low-cost optical components such as lenses, diffraction gratings, optical fibers, and adhesives. Their common use derives from a combination of qualities: good optical transmission, a wide range of mechanical properties and simple fabrication via molding: extrusion or solution processing. Functional optical polymers, including liquid crystals and photorefractives, have also been widely applied to polarization control and for nonlinear- and electro-optics. Conjugated polymers are a remarkable class of functional plastics that exhibit some unusual properties. Firstly, they are plastics that can conduct electricity; and in fact are electrically semiconducting. Secondly, they are plastics that can be stimulated either optically or electrically to emit visible light.

These two unusual properties derive from a particular chemical structure specific to this family of plastics. In all conjugated polymers, the backbone of the molecule is a chain of carbon atoms joined by alternating single and double covalent bonds. The simplest example is illustrated in Figure 1a, which shows the chemical structure of the polymer poly(acetylene). Structural integrity of the polymer is provided by a string of single σ-bonds between adjacent carbon atoms. The additional bonding electrons in the double bonds are located in π-orbitals oriented perpendicular to the polymer chain. Adjacent π-orbitals overlap, and their electrons can be widely delocalized along the polymer backbone. The delocalized electrons can move relatively freely along the chain, giving the electronic properties of the material. The π-orbitals can be transformed between bonding and anti-bonding states by the absorption or emission of light, leading to strong optical transitions. While poly(acetylene) is a very inefficient light emitter, other conjugated polymers, notably those that include phenyl rings in the backbone (Figure 1b–d), can be highly emissive. Photoluminescence quantum

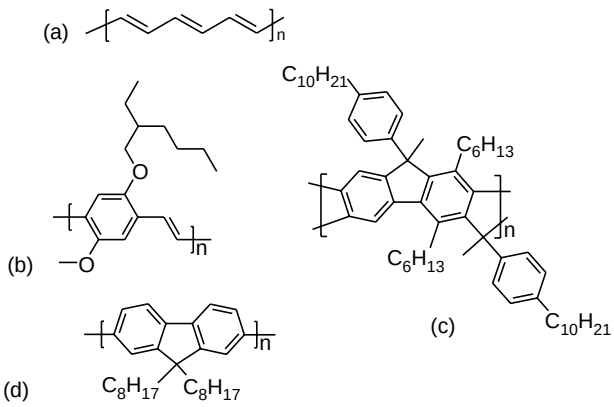

Figure 1 Chemical structures of several conjugated polymers. (a) poly[acetylene]; (b) poly[2-methoxy-5-(2′-ethylhexyloxy)-1,4-phenylene vinylene], (MEH-PPV); (c) ladder-type poly[1,4-phenylene], (MeLPPP); (d) poly[9,9-dioctylfluorene], (PFO).

efficiencies, approaching unity in solution and as high as 60% in the semiconducting solid state, have been realized.

Organic Semiconductor Lasers

While organic dye lasers have been widely used since the mid-1960s, the history of organic semiconductor lasers is much more recent. Other than a few early observations of lasing in molecular crystals in the 1970s by Karl and others, the field of organic semiconductor lasers began in earnest following the related discovery in 1990 by Burroughes *et al.* of electroluminescence in a conjugated polymer. In 1992, Moses demonstrated lasing for the first time in a solution of the conjugated polymer MEH-PPV (Figure 1b). However, it was another four years – once material synthesis had been sufficiently refined – before stimulated emission and then true laser action were observed in polymers in the semiconducting solid state. In the following year the first demonstrations were made of dye-doped small molecular semiconductor lasers. Since that time, there has been a great growth in interest in the field with an average of ~30 research papers per year. Further details of the development of the field of organic semiconductor lasers can be found in the Further Reading.

So what are the particular properties of conjugated polymers that have stimulated interest in developing organic semiconductor lasers? Firstly these materials are broadband visible emitters. Through appropriate chemical design, their emission can be tuned throughout the visible spectrum, from 400 nm to 700 nm (Figure 2). The polymers exhibit large optical gain coefficients, because the optical transitions are dipole-allowed. Compared with small molecular dyes, polymers exhibit a relatively low degree of self-absorption of the emitted light, and much reduced quenching of the luminescence with increased concentration. This means that it is possible to work with much higher chromophore densities than conventional laser dyes, which is advantageous for achieving very low threshold lasing in extremely compact resonators. Furthermore, because these materials are semiconducting, they have the potential for direct electrical excitation. This could lead to semiconductor diode lasers operating throughout the visible spectrum and, in particular, in the blue and green spectral bands where currently there are few commercially viable inorganic semiconductor sources. Finally, these materials are plastics and so are amenable to simple processing, and can be shaped and structured into complex feedback resonators.

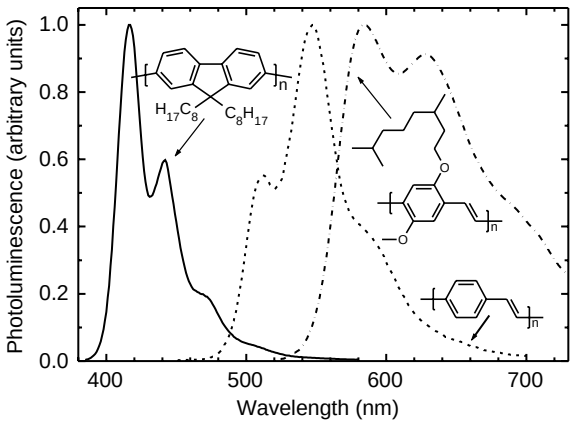

Figure 2 Photoluminescence spectra of three commonly used conjugated polymers; chemical structures inset.

Organic Semiconductor Gain Materials

A great advantage of organic laser media is that a very wide range of materials can be readily synthesized, and through minor changes in the molecular structure one can relatively simply tune the emission properties of the material. Many different organic semiconductors have been applied as lasers, though they fall into three major categories. These are conjugated polymers, dye-doped small molecular semiconductors, and molecular single crystals. Conjugated polymers make up the most diverse materials set, most notably including derivatives of poly(paraphenylene-vinylene), poly(paraphenylene), and poly(fluorene) (Figure 1b–d). The second category comprises small molecular semiconductors that are co-evaporated under vacuum with a dopant laser dye. In this blend of materials, the organic semiconductor functions simply as an energy transfer host for the emitting dye molecules. In the third category are organic single crystals, including materials such as polyacenes and thiophene oligomers. These materials have the advantage of having the highest electrical mobilities of any organic semiconductors, which is important for electrical excitation. However, they can be demanding to grow into high-quality single crystals, and so lack the key processing advantages of the other families of organic semiconductors.

Figure 3a shows a typical exciton energy level structure of a semiconducting polymer. Similar to a conventional laser dye, the energy levels comprise a manifold of singlet states and a manifold of triplet states, each with vibrational sublevels. The laser transition occurs from the bottom of the first excited singlet state S_1 to the vibronic overtones of the ground

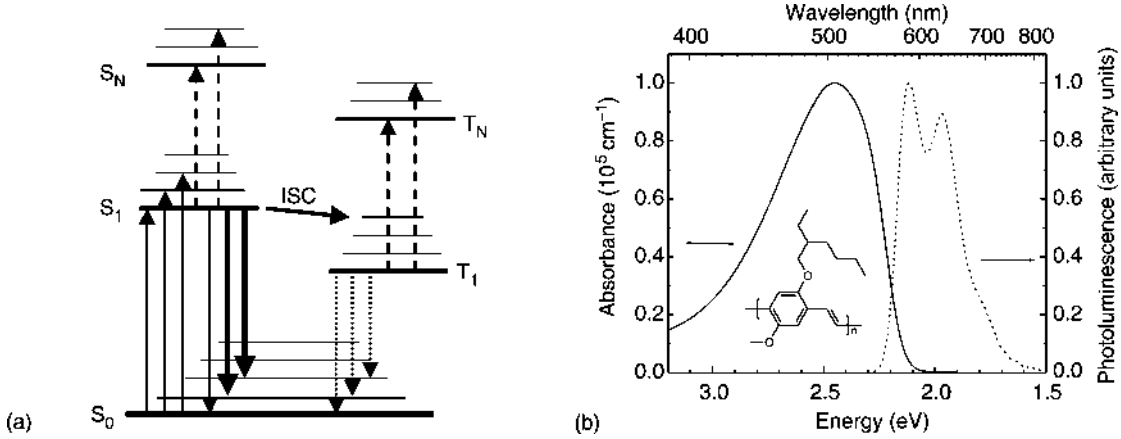

Figure 3 (a) Generic energy level structure and optical transitions of a conjugated polymer. S_N refers to the *N*th singlet state; T_N refers to the *N*th triplet state; ISC: intersystem crossing. (b) Absorption and photoluminescence spectra of the polymer MEH-PPV.

state S_0, so forming a classic 4-level laser system. The optical transition is dipole-allowed and so has a high oscillator strength. As a result, conjugated polymers typically exhibit simulated emission cross-sections in the range 10^{-16} cm^2 to 10^{-15} cm^2, while the population inversion lifetime is as short as a few hundred picoseconds.

There are, however, several other transitions that can compete with the stimulated emission process. There can be excited-state absorption from the S_1 singlet state into higher-lying singlet states. Additionally intersystem crossing from the S_1 state to the lowest excited triplet state T_1 can subsequently lead to excited triplet absorption. Other processes in the polymer can also impede the stimulated emission. Firstly, charged excitations can exhibit absorption bands that overlap with the emission. Secondly, at high excitation densities, exciton–exciton annihilation can significantly deplete the excited states available for stimulated emission. Ultimately, singlet–singlet annihilation is the limiting factor for the maximum excitation density possible, and while polymer films can have a repeat-unit density of $\sim 10^{21}$ cm^{-3}, the excited state density is typically limited to around 10^{18} cm^{-3}. In many materials this is not restrictive on achieving optical gain because stimulated emission can occur at densities of 10^{17} cm^{-3} and below.

An important factor that distinguishes the photophysics of semiconducting polymers from conventional laser dyes is an internal process of exciton migration between the absorption and emission of a photon. Excitations are confined to short straight sections of the polymer between kinks or twists in the chain. These straight sections, or 'sites', are typically a few repeat units in length, though there is a statistical spread of site lengths in any given material. When formed, excitons tend to hop to the longest, lowest-energy, sites before emitting a photon. The exciton migration process has two effects. First, it means the emission takes place from a subset of the chain segments, resulting in a lower degree of inhomogeneous broadening of the emission spectrum (vibronic peaks are better resolved than in absorption (see Figure 3b)). Second, the peaks of the absorption and emission spectra tend to be widely separated in wavelength, reducing the level of self-absorption at the lasing wavelength.

As mentioned above, blending materials is a particularly attractive approach towards further reducing self-absorption in organic semiconductor lasers. In such systems, a narrow bandgap dopant chromophore is blended with a concentration of 1–2% in a wider bandgap semiconducting host. Emission from the host is strongly quenched via nonradiative Förster energy transfer to the guest molecule. The blended material exhibits an absorption profile that is dominated by the host absorption, while the emission profile is almost completely that of the guest dopant. By strongly separating the emission peak from the absorption edge of the host, it is possible to achieve a substantially reduced (by perhaps an order of magnitude or more) residual absorption coefficient at the lasing wavelength. In doing so, the combined material acts much more like an ideal 4-level laser medium and can thereby have substantially lower pump thresholds for stimulated emission.

Measuring Gain

The dynamics of the excited state population in semiconducting polymers have been widely studied using femtosecond transient absorption experiments.

This is a pump–probe technique in which the transmission of the broadband probe pulse is compared in the presence and absence of a pump pulse that excites molecules into the first excited state. Such measurements have the time resolution of the pump pulses used and so can typically be ~100 fs. By systematically delaying the arrival of the broadband probe pulse, it is possible to map out the evolution of the gain dynamics across the emission spectrum of the material with very high time resolution.

Figure 4 shows a series of snapshots of the absorption induced by a pumping pulse in a poly(thiophene) derivative. This figure shows the changing absorption of the sample between 200 fs and 400 ps after excitation. Within the absorption band of the material, one finds that the level of absorption is slightly reduced because the pump pulse has partly depleted the ground state. In the emission band there is also a reduction in the absorption that mimics in shape the photoluminescence spectrum of the polymer. This negative absorption – or optical gain – shows direct evidence for stimulated emission through much of the emission band. In the low energy tail of the emission, however, there is an increase in absorption that arises due to the excited-state absorption process mentioned earlier. Clearly, for a good laser material, the spectra of excited-state and charge-induced absorptions should overlap as little as possible with the photoluminescence.

Many other investigations of gain in semiconducting polymer films have used a phenomenon known as spectral line narrowing to explore indirectly the amplifying properties of the medium. The experimental procedure for such studies is as follows. A thin film of the semiconducting material, of typically 100 nm thickness, is excited by a pulsed pump laser focused to a stripe of dimensions ~100 μm wide by a few millimeters in length. A significant fraction of the light emitted by the material is trapped in the film by total internal reflection at the polymer–air and polymer–substrate interfaces, and is thereby waveguided along the length of the excitation stripe. This waveguided spontaneous emission can be amplified by stimulated emission before being emitted out the edge of the film.

One finds that, above a particular threshold of pump energy, there is a superlinear increase in the output energy with pump intensity. Associated with this is a dramatic narrowing of the broad emission spectrum to a linewidth of typically less than 10 nm, as illustrated in Figure 5. These two features are signatures of gain narrowing via the process of amplified spontaneous emission. The line-narrowed feature usually appears at the first vibronic overtone of the optical transition, where the gain exceeds the losses in the material by the greatest amount. This particular wavelength experiences the greatest amplification and, above a particular pumping density, undergoes a runaway effect in which most of the light is stimulated to emit at that wavelength at the expense of the rest of the spectrum.

In the case of amplified spontaneous emission or mirrorless lasing as it is also known, the spontaneous emission guided along the length of the stripe acts as

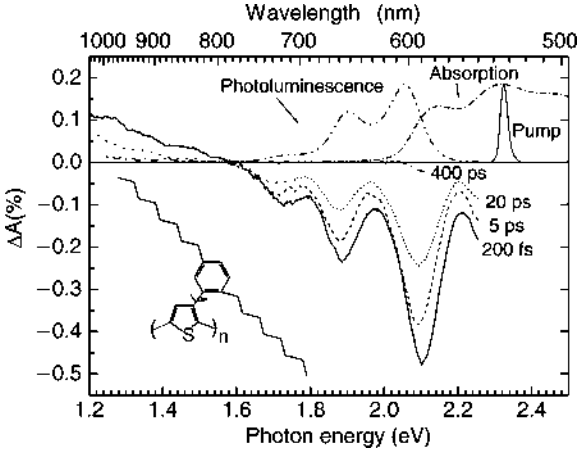

Figure 4 Transient absorption spectra of poly[3-(2,5-dioctylphenyl)-thiophene]; chemical structure inset. (Reproduced from Ruseckas A, Theander M, Valkunas L, Andersson MR, Ingänas O and Sundström V (1998) Energy transfer in a conjugated polymer with reduced inter-chain coupling. *Journal of Luminescence* 76 & 77: 474–477. Copyright (1998), with permission from Elsevier.)

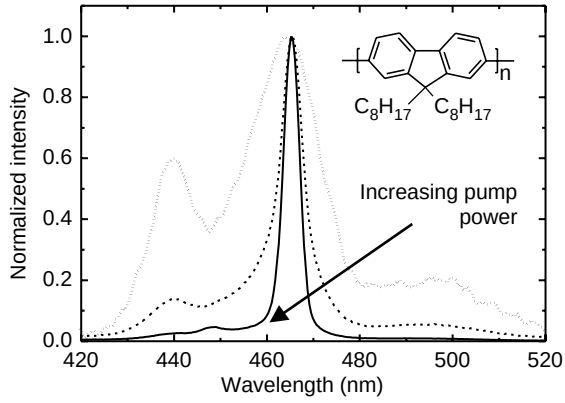

Figure 5 The change in spectral shape with increasing excitation density (from 10 to 133 μJ cm^{-2}) of the edge emission from a PFO film – the normalized spectra show a significant spectral line narrowing at high excitation densities. (Adapted from Heliotis G, Bradley DDC, Turnbull GA and Samuel IDW (2002) Light amplification and gain in polyfluorene waveguides. *Applied Physics Letters* 81: 415–417. Copyright (2002), with permission from American Institute of Physics.)

the 'probe pulse'. In these experiments it is therefore necessary to extract information indirectly about the net gain in the material. By integrating the spontaneous and stimulated emission along the length of the stripe, one finds that the wavelength-dependent output intensity $I(\lambda)$ of the spectral line narrowed light is given by the relationship

$$I(\lambda) \propto \frac{1}{g(\lambda)}[\exp[g(\lambda)l] - 1] \qquad [1]$$

where I_p is pumping intensity, l is the length of the stripe, and $g(\lambda)$ is the net gain coefficient. Therefore, by monitoring the intensity of the line-narrowed emission as a function of the length of the pump stripe, one can extract the value of the net gain for the material. Similarly, by progressively moving the stripe further and further away from the edge of the film, it is possible to evaluate the waveguide losses of light propagating through an unpumped region of the waveguide. Waveguide losses in conjugated polymers typically lie in the range of 3–50 cm^{-1}. Net gains can be as high as 60 cm^{-1} or 260 dB cm^{-1} at modest pumping densities of 4 kW cm^{-2}. These substantial gains, in very small propagation lengths, therefore highlight the attraction of waveguide-based resonators for polymer lasers. In the next section we will discuss the most commonly used feedback structures that have been used for polymer lasers.

Polymer Laser Resonators

The first demonstration of cavity lasing in a conjugated polymer was made by Moses in 1992. The resonator consisted simply of two flat dielectric mirrors on either side of a cuvette containing the polymer in solution. Lasing occurred at the peak wavelength of the gain profile. Subsequent developments in solid-state polymer lasers have involved more sophisticated resonators that exploit the capacity of these materials for very simple processing. For example, solution-processing methods such as spin-casting or dip-coating, make it possible to fabricate very high-quality optical waveguides. Wavelength-scale microstructures can also be readily formed, allowing the straightforward fabrication of laser microresonators.

Many different resonators have been explored for solid-state polymer lasers. Some of the key types are shown in Figure 6, including different geometries of both microcavities and distributed feedback structures. The first resonator used, by Tessler *et al.*, was a planar microcavity (Figure 6a), in which the polymer film was sandwiched between two mirrors. The structure supported standing-wave resonances perpendicular to the plane of the film, at three wavelengths in the emission band of the polymer. Above laser threshold, light was preferentially stimulated into only one of these modes. Microcavities of other geometries have also been studied, including annular and spherical microresonators (Figure 6b,c), in which the optical field is confined in waveguide modes, whispering gallery modes, or a superposition of the two. Such structures provide a longer interaction length in the gain medium for stimulated emission; though tend to lase on many longitudinal modes and have no directional output beam.

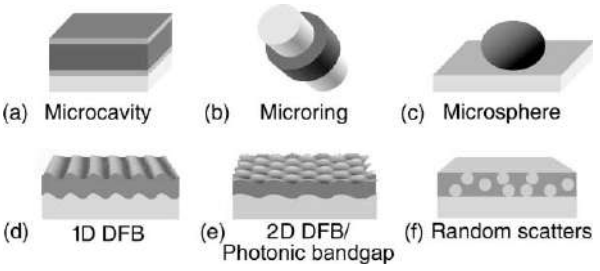

Figure 6 Resonators used for conjugated polymer lasers. (a) Planar microcavity; (b) annular microcavity; (c) spherical microcavity; (d) distributed feedback resonator; (e) 2D DFB/ photonic crystal resonator; (f) random laser.

The final key group of resonators are distributed feedback (DFB) structures, which researchers have progressively favored to attain the low pump threshold densities that will be required to realize electrically pumped lasing. Figure 6d shows a typical DFB laser structure. The device is based on a thin polymer film that is typically 100 nm in thickness deposited on top of a silica substrate. The silica–polymer–air structure forms an asymmetric slab waveguide that supports only the lowest order transverse electric mode TE_0. To provide the feedback, the substrate is modulated with a wavelength-scale periodic grating structure. These corrugations give rise to Bragg scattering of the waveguide mode. Through careful selection of the grating period, one may arrange that optical wavelengths close to the peak of the polymer gain will be Bragg scattered from the propagating TE_0 mode to the counter-propagating TE_0 mode. This is achieved when the Bragg equation

$$2n_{\mathrm{eff}}\Lambda = m\lambda_{\mathrm{B}} \qquad [2]$$

is satisfied, where Λ is the grating period, n_{eff} is the effective refractive index of the waveguide, m is an integer, and λ_{B} is the Bragg wavelength. In so doing, one may arrange that two counter-propagating waveguide modes can be coupled together, thereby providing the distributed resonant feedback for the laser process. Commonly in semiconducting polymer

lasers, a second-order (i.e., $m = 2$) Bragg grating has been favored to provide the feedback for the laser, so that the first-order scattering provides a surface-emitted output coupling, perpendicular to the plane of the guide. This surface-emitting geometry is particularly attractive in polymer lasers because it is relatively difficult to achieve a high optical quality edge to the waveguide. DFB lasers have the advantage over simple microcavity resonators in that they can combine long interaction lengths, limited by the scattering length of the Bragg grating, with excellent wavelength selection, which is controlled by the grating period.

In addition to conventional 1D DFB gratings, a number of more exotic 2D and 3D DFB, and photonic crystal resonators have been explored in semiconducting polymers. Even randomly located scattering sites in a polymer film have been used to provide a weak closed-loop cavity feedback. Examples of 2D gratings are given in Figure 7, which shows atomic force micrographs of an eggbox structure, formed from two orthogonal gratings, and a circular DFB grating. Such 2D DFB structures can significantly improve the operation of surface-emitting polymer lasers compared with 1D gratings. Improved feedback yields lower oscillation thresholds, combined with higher slope efficiencies and significantly improved beam quality. Indeed, such surface emitting DFB lasers can produce nearly diffraction limited laser beams.

Typical operating characteristics of an egg-box DFB polymer laser are shown in Figure 8. This laser, which is based on the material MEH-PPV, is transversely pumped with a pulsed pump laser focused onto a small region of the waveguide; up to 90% of the incident pump light is absorbed in the 100 nm thick film. The laser has a threshold pump energy of 4 nJ and an external slope efficiency of $>7\%$, and operates on a single frequency close to the Bragg wavelength of the periodic waveguide. Lasing does not occur exactly at the Bragg wavelength, because at this wavelength the propagation of light is suppressed by the periodic structure. This leads to a photonic stopband within which the emission is reduced. At the edges of this stopband the microstructure can support standing wave electric field patterns of the same spatial period. At these wavelengths there can be a strong interaction between the electric field and excited states in the gain medium, which leads to low-threshold band edge lasing as observed in the figure.

Although DFB resonators have advantages over microcavity devices, the complicated lithographic and etching steps involved in their fabrication can detract from the key material advantage of simple processing. To address this issue, several different techniques for simply replicating the complicated microstructures have been explored. These include UV embossing of flexible plastic substrates, and hot embossing and micromolding of the active

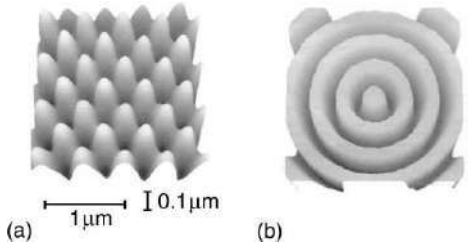

Figure 7 Atomic force microscope images of (a) 'egg-box' DFB grating, and (b) circular DFB grating; grating period is 400 nm in each structure.

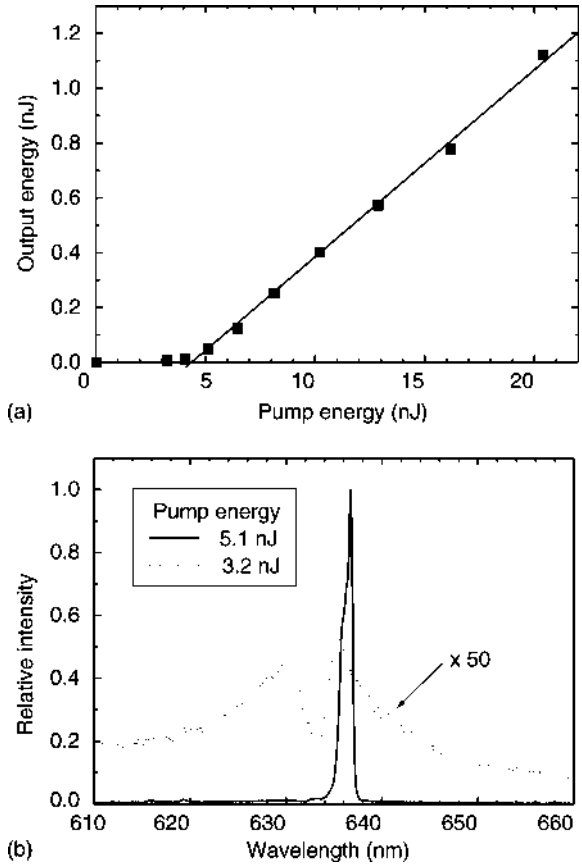

Figure 8 (a) Energy characteristic of an egg-box DFB laser; (b) normalized emission spectra above and below lasing threshold, showing lasing in the edge of the photonic stopband. (Reprinted from Turnbull GA, Andrew P, Barnes WL and Samuel IDW (2002) Operating characteristics of a semiconducting polymer laser pumped by a microchip laser. *Applied Physics Letters* 82: 313–315. Copyright (2003). With permission from American Institute of Physics.)

polymer itself. Such techniques can very readily reproduce structures with feature sizes as small as 100 nm and could allow future mass production of photonic microstructures in these materials.

Towards Plastic Diode Lasers

A typical experimental configuration for characterising optically pumped polymer lasers is shown in Figure 9. The laser is mounted in a vacuum chamber in order to isolate the polymer from oxygen and water. A pulsed visible or ultraviolet pump laser is attenuated and then focused onto the waveguide to transversely pump the polymer laser. The emission is collected using a CCD spectrograph for spectral measurements and energy meter and beam profilers for other characteristics. Typical pulsed pump lasers, which have been used, include nitrogen laser pumped dye lasers, and spectral harmonics of regeneratively amplified Ti^{3+}-sapphire lasers or flashlamp pumped Nd^{3+}:YAG lasers. Recent advances have shown the thresholds to be low enough to be successfully pumped with a modest Nd^{3+} pulsed microchip laser.

So while optically pumped polymer lasers are now proving to be compact, functional light sources, a major motivation for research in this field remains the prospect of electrically driven plastic diode lasers. Such lasers could be very low-cost, flexible light sources that would access any wavelength throughout the visible spectrum. In particular, there are exciting prospects of producing blue and green diode lasers in spectral regions currently difficult to access with inorganic semiconductors. The lowest threshold densities reported for optically pumped organic semiconductor lasers are of the order of 100 W cm^{-2}. To achieve similar excitation densities electrically, one requires pump current densities of ~200 A cm^{-2}. This is substantially more than the few mA cm^{-2} commonly used in polymer LEDs for display applications. Presently, such high current densities are infeasible with continuous excitation, but can be exceeded with short current pulses of a few tens of nanoseconds duration.

Electroluminescence in conjugated polymers was first achieved in 1990 by Burroughes and co-workers. Since that time there has been remarkable progress made in the science and technology of organic LEDs. Commercial products are now on the market, and the feasibility of both flexible and full-color light-emitting displays, in which the individual pixels have been ink-jet printed, have been demonstrated in the laboratory. Despite such rapid progress in the field of organic LEDs, there have been no successful demonstrations to date of electrically pumped lasing in any organic semiconductor. In this section we highlight some of the challenges that must be faced if electrically pumped lasing is to be achieved. To understand the problems associated with electrical driving of organic semiconductor lasers, it is useful first to review the generic structure of organic light-emitting diodes.

A generic organic LED structure is shown in Figure 10. The device consists of one or more organic layers sandwiched between two electrical contacts. Usually a transparent anode is used, commonly made from ITO, which allows light to be emitted from the front face of the LED. A low work-function metal, such as calcium or aluminum, is used as the cathode. The organic part of the device may consist of several layers. These include one or more charge transport layers, plus a light-emitting layer in which the electrons and holes combine to form excitons before emitting light. Typically, the combined thickness of these organic layers is as little as 100 nm. Such thin devices are necessary to keep operating voltages low, to overcome potential barriers for injection, and because charge mobilities in organic semiconductors are very low (10^{-5} to 10^{-2} cm^2 V^{-1} s^{-1}).

The inclusion of electrical contacts into organic semiconductor laser structures can be problematic since they need to be situated very close to the emitting region. Metallic contacts can quench the luminescence and introduce losses to the waveguide mode, thereby increasing the oscillation thresholds. Furthermore, the ITO anode has a high refractive index and so can reduce the spatial overlap of the

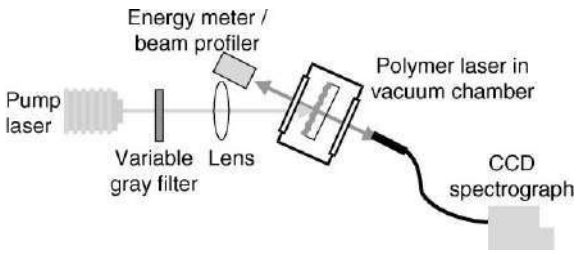

Figure 9 Schematic of a typical experimental configuration for characterizing polymer lasers.

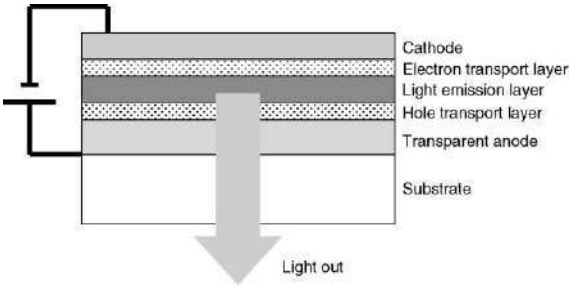

Figure 10 Schematic structure of a generic organic light-emitting diode.

guided mode with the active medium. Such issues indicate that the design of organic LEDs needs to be somewhat modified for achieving electrically pumped lasing. In particular, there needs to be a trade-off between the optimum electronic and photonic designs. To address these issues, various device structures have been proposed that distance the optical mode from metallic contacts, while reducing the thickness of the ITO layer to optimize the mode confinement in the active material.

A more demanding obstacle to electrically pumped lasing is the additional excited state absorption associated with the electrical pumping scheme. As many as three-quarters of the generated excitons in an organic LED are formed in the triplet state which is nonemissive and can lead to excited-state absorption, albeit principally at lower energies than the lasing wavelength. Additionally, injected charges that have not yet formed excitons, exhibit strong absorption, which overlaps significantly with the peak of the material gain. Indeed, this charge-induced absorption has so far prevented the demonstration of injection lasing in conjugated polymers, even in optimized device structures mentioned above. A possible alternative route for efficient pumping of polymer lasers could be through indirect electrical pumping. For example, an inorganic diode laser or LED, or even an organic LED, could be used as an electrically efficient optical pump for the semiconducting polymer laser. Recent advances in GaN-based LEDs and lasers provide interesting prospects for such alternative systems.

Research on semiconducting polymer lasers continues, both to reduce the excitation densities needed for threshold and to investigate new material systems that could overcome the obstacle of charge-induced absorption. By building on recent developments in resonator geometries, material blends, and compact pumping schemes, semiconducting polymer laser systems, whether optically or electrically pumped, should soon become practical, visible light sources for a range of applications.

List of Units and Nomenclature

Charge mobility	[$cm^2\ V^{-1}\ s^{-1}$]
Current density	[$A\ cm^{-2}$]
Energy density	[$\mu J\ cm^{-2}$]
Excited state densities	[cm^{-3}]
Gain coefficient α ($\exp(\alpha z)$)	[cm^{-1}]
Intensity/power density	[$kW\ cm^{-2}$]
Optical gain	[$dB\ cm^{-1}$]
Stimulated emission cross-section	[cm^2]
Waveguide propagation loss coefficient γ ($\exp(-\gamma z)$)	[cm^{-1}]
Wavelength	[nm]

See also

Diffractive Systems: Diffractive Laser Resonators. **Lasers:** Dye Lasers; Planar Waveguide Lasers. **Optical Amplifiers:** Basic Concepts. **Optical Materials:** Plastics. **Photonic Crystals:** Photonic Crystal Lasers, Cavities and Waveguide.

Further Reading

Burroughes JH, Bradley DDC, Brown AR, *et al.* (1990) Light-emitting-diodes based on conjugated polymers. *Nature* 347: 539–541.

Friend RH, Gymer RW, Holmes AB, *et al.* (1999) Electroluminescence in conjugated polymers. *Nature* 397: 121–128.

Karl N (1972) Laser emission from an organic molecular crystal. *Physica Status Solidi (a)* 13: 651–655.

Kranzelbinder G and Leising G (2000) Organic solid-state lasers. *Reports on Progress in Physics* 63: 729–762.

McGehee MD and Heeger AJ (2000) Semiconducting (conjugated) polymers as materials for solid-state lasers. *Advanced Materials* 12: 1655–1668.

Miyata S and Nalwa HS (eds) (1997) *Organic Electroluminescent Materials and Devices*. Amsterdam: OPA.

Moses D (1992) High quantum efficiency luminescence from a conducting polymer in solution – a novel polymer laser-dye. *Applied Physics Letters* 60: 3215–3216.

Pope M and Swenberg CE (1999) *Electronic Processes in Organic Crystals and Polymers*. New York: Oxford University Press.

Samuel IDW (2000) Polymer electronics. *Philosophical Transactions of the Royal Society of London A* 358: 193–210.

Scherf U, Riechel S, Lemmer U and Mahrt RF (2001) Conjugated polymers: lasing and stimulated emission. *Current Opinion Solid State and Materials Science* 5: 143–154.

Tessler N (1999) Lasers based on semiconducting organic materials. *Advanced Materials* 11: 363–370.

Tessler N, Denton GJ and Friend RH (1996) Lasing from conjugated-polymer microcavities. *Nature* 382: 695–697.

Planar Waveguide Lasers

S Bhandarkar, Alfred University, Alfred, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Lasers have evolved to become true marvels of technology. From their invention in the early 1960s, laser light has changed the way society operates, its application stretching to virtually every aspect of human activity. The versatility of lasers is derived from the fact that there are several different types, each with a unique set of advantages. Solid-state lasers, made from specific semiconducting materials, are commonly used in the optical networks that underlie the greatest machine on Earth – the Internet. High power gas lasers, such as CO_2 lasers, are used in industrial machining, whereas rare-earth doped and dye lasers have found applications ranging from eye surgery to tattoo removal.

Planar waveguide lasers are a relatively new class of lasers made possible by advances in the field of optical waveguides. Over the past decade, with the aim to miniaturize optical components, planar waveguide technology has been developed to a high degree of sophistication. Likewise, Er and other rare-earth doping was studied extensively owing to its importance in all-optical amplification. These elements turned out to be the building blocks for the development of fiber lasers and subsequently planar waveguide lasers.

This chapter will discuss the basic concepts underlying this category of lasers, the variations therein, and the main characteristics. It is important to understand the basic features of waveguides and lasers, which is included as a prelude. We will then discuss planar waveguide lasers with emphasis on the essential elements: the laser materials, the various designs, their performance, and the resultant advantages as well as the applications.

Waveguides

A waveguide is a general conduit for efficient propagation of an electromagnetic field. It is usually defined by making a region of higher refractive index within a dielectric medium. This behaves as an attractive potential well that confines light in the form of bound modes, much like a potential well that confines electrons to a bound state. The propagating field is referred to as the lightwave and is governed by Maxwell's equation. For a weakly guiding approximation, as is the case when the refractive index contrast (Δn) is small, Maxwell's equation can be reduced to a scalar wave equation for the transverse electric field. For a given wavelength (λ) and polarization, the scalar wave equation in a nonmagnetic medium is

$$\nabla^2 E(x) + \frac{4\pi^2}{\lambda^2} n(x)^2 E(x) = 0 \qquad [1]$$

where E is the electric field. For a waveguide aligned in the z direction, the field is the sum of the two propagating transverse modes, each with the form $E(x) = \Phi(x, y)e^{ik_z z}$ where $\Phi(x, y)$ is the field distribution and k_z is the propagation constant. The wave equation for the transverse mode reduces to

$$\frac{\lambda^2}{4\pi^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\Phi(x, y) + n(x, y)^2\Phi(x, y) = (\bar{n})^2\Phi(x, y) \qquad [2]$$

Here, $\bar{n}$ is referred to as the effective refractive index and is given by $\bar{n} = k_z\lambda/2\pi$. The effective index dictates the occurrence of bound modes – these states occur when $\bar{n}$ is between n_{core} and n_{cladding}. For smaller values of $\bar{n}$, there is a continuum of unbound and radiation modes. These are leaky modes that are dispersed quickly over a short distance. Only bound modes are guided in the core of the waveguide structure.

Planar Waveguides

Waveguides formed on a flat substrate are called planar waveguides. These are typically made by stepwise deposition of films of dielectric materials (typically glass). The waveguide core is defined by one of several methods, the most common of which uses lithography and etching. In this case, a film of an appropriately higher index material is deposited on an 'under-cladding' film and then selectively etched to define a core. This is generally covered with a suitable 'upper-cladding' layer so that the index contrast is controlled in all directions. Other methods of forming the core include patterned ion exchange or ion implantation or increasing the index of the medium using a femtosecond laser, etc. In all these cases, the end result is the formation of an embedded higher-index region whose locus defines the optical path for the lightwave. A circuit made from these waveguides is known as planar lightwave circuits (PLC).

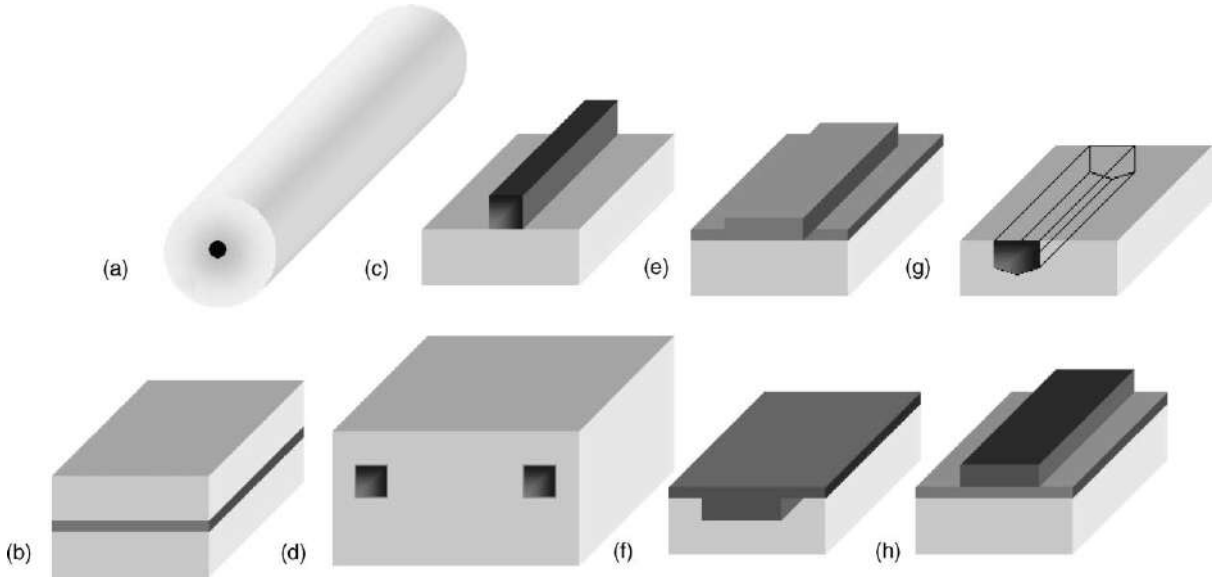

Figure 1 Range of different waveguide designs, with the core represented by the darker regions: (a) Optical fiber; (b) slab waveguide; (c) channel waveguide; (d) embedded channel waveguides with a layer of the overcladding; (e) rib waveguide; (f) inverted rib; (g) ion diffused; and (h)strip loaded, comprising of a strip of a different higher index material (that serves to confine light) on a core layer.

Waveguide cores may be designed in several different configurations, as shown in Figure 1. These designs allow for different fabrication processes as well as the control over the modal properties of the lightwave. For a given index contrast, the number of modes traveling in the waveguide is primarily determined by the dimensions of the waveguide. The aspect ratio of the waveguide affects the propagation of the transverse electric (TE) and the transverse magnetic (TM) modes, due to different propagation constant k_z for each mode.

Lasers

The generation of laser light requires four basic conditions to be satisfied:

- a source capable of light emission;
- an energy supply source – typically electrical, optical, or chemical;
- population inversion to the excited state in the lasing medium; and
- absence of parasitic effects that would absorb the emitted light.

Laser light is often a result of either electronic or vibrational excitation. This is because these excitations involve transfer between energy states that coincide with the radiation of UV, visible, or infrared light. Population inversion is an important requirement in lasers since the emission is stimulated. This involves having a greater fraction of the atoms or molecules in the excited state so that a stimulating photon can cause a radiative decay to the lower energy level. The stimulated light from most lasing media is weak, which requires unacceptably long lengths to generate sufficient power. Instead, as shown in Figure 2, lasers are made with two parallel mirrors with an axial light-generating source in between. This configuration is known as a resonant cavity, a resonator, or an oscillator. The back mirror is usually fully reflective but a partial front mirror allows a portion of the light in the cavity to pass through. The two mirrors cause the light to bounce to and fro, forcing it into making multiple passes through the cavity. Light emanating from this structure is thus amplified and is also coherent (that is the lightwaves are in phase

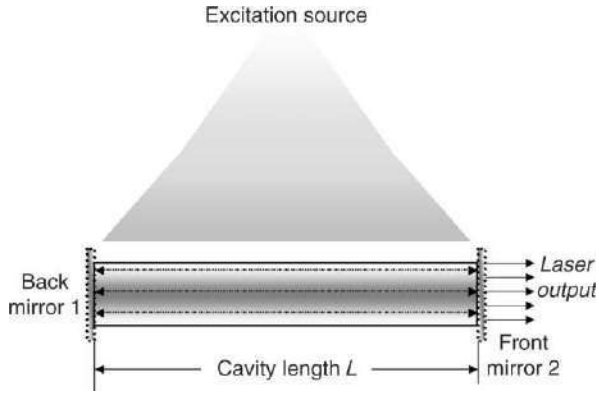

Figure 2 Conceptual representation of a simple Fabry–Perot laser.

with each other spatially and temporally). This has come to represent the signature output of a laser.

Planar Waveguide Lasers (PWL)

By way of an example, let us consider a solid state, rare-earth doped laser, such as an Nd-doped YAG laser. All rare-earth ions are able to fluoresce, some (e.g., Nd) more efficiently than others, and hence can be light emitters. These ions have to be dissolved (doped) into a solid host, say, a glass or a single crystal such as YAG (**Y**ttrium **A**luminum **G**arnet). The excitation of these ions is done optically with light of a suitably higher energy (or lower wavelength). When this 'pump' light is made incident on a rod of the rare earth doped material, it is absorbed by the rare earth ions and used to reach an excited electronic state. Amongst all the possible higher energy states, only certain ones are sustainable because the electron resides in that state for an acceptably long period of time (called the lifetime). Hence, population inversion can be more readily achieved in these longer lifetime states and stimulated processes such as amplification and lasing are readily possible. The laser light generated in the doped rod is emitted across its entire cross-section.

Planar waveguide lasers are designed to miniaturize the kinds of stand-alone solid state lasers describe above. In principle, a PWL is a laser where the light emitting resonant cavity is a planar waveguide. This topic of our interest is a relatively recent development that renders some unique advantages. These can be broadly classified into two categories: one relating to making efficient high-power lasers and the other, to the integration of lasers with other optical components.

An example of the first type is a Nd-YAG laser configured into a planar geometry so that the laser light is now contained within the waveguide. One design that has evolved to achieve this well is shown in Figure 3. The light is guided by the waveguiding structure setup by the sapphire planes. Since both the excitation (or pump) and the lasing modes are well confined, high modal overlap is achieved. These PWLs can be edge-pumped or face-pumped and high powers (measured in the order of Watts) can be achieved.

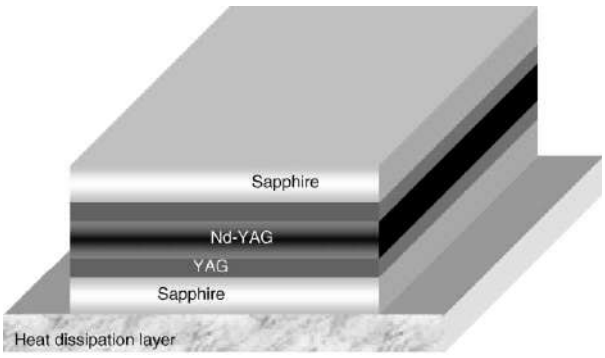

Figure 3 End view of a planar waveguide form of a Nd-YAG laser made using sapphire (single crystal Al_2O_3) for cladding.

In the above design, the cavity is defined by two parallel reflectors bracketing the light-generating medium and is known as a Fabry–Perot etalon. The distance between the two mirrors sets up a wavelength filtering mechanism whereby only resonant wavelengths are sustained. These are related to the intra-mirror distance by the simple equation shown below:

$$L_c = \frac{a\lambda}{2n} \qquad [3]$$

where L_c is the length of the cavity, a is an integer, n is the refractive index of the medium, and λ is the wavelength in vacuum. As the cavity length increases, the number of possible resonant wavelengths increases and the spacing between wavelengths is reduced. These are called the longitudinal modes (Figure 4a). Likewise, lateral modes (Figure 4b) can be generated in the cavity as well, tending to make the operation of the laser multimoded, as shown in Figure 5. This is characteristic of a Fabry–Perot laser, which can be overcome by embellishing the basic etalon design.

As stated above, another big advantage here is that the platform for fabricating PWLs can be chosen to be the same as to make PLCs. Waveguide comprizing the laser can be continued further and configured into other desirable components, such as splitters, couplers, etc. Thus the laser does not have to be separately coupled to the waveguide. This is a major advantage as this connecting process is time-consuming and incurs loss of light. In the field of photonics, the ability to integrate active components with passive ones onto the same chip is considered to be the first step toward photonic chips, analogous to the integrated circuit chip that has revolutionalized our world. This is the driver behind continued research and development of active elements such as PWLs that can be coupled with other actives, such as thin film electro-optic modulators or acousto-optic filters, etc.

Waveguide Laser Materials

The waveguiding structure is comprised of materials that are either dielectrics or semiconductors, both inorganic and organic. For their homogeneity and broad transmittance, high silica glasses are ideal

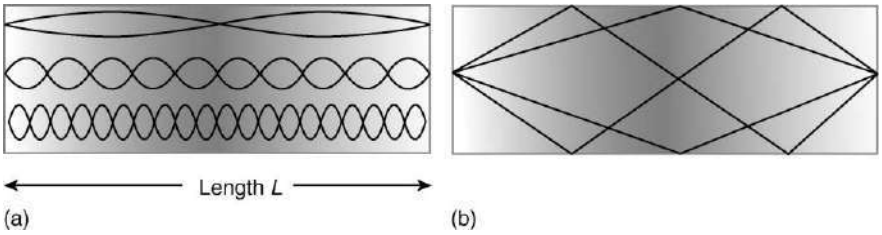

Figure 4 Generation of longitudinal (a) and lateral modes (b) in a Fabry–Perot laser.

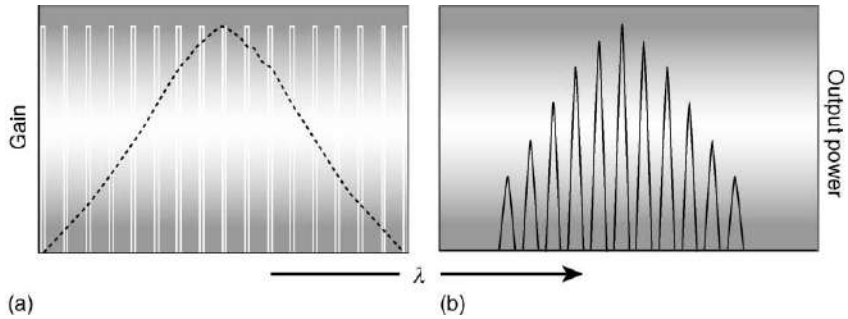

Figure 5 The output of a Fabry–Perot laser: (b) is governed by the number of modes within the gain spectrum of the lasing medium shown by the dotted lines (a). The spacing between the modes can be altered changing the length of the cavity. The smallest mode in output spectrum (b) needs to have enough gain to overcome the lasing threshold.

materials for waveguiding. The chief advantages of high silica waveguides are easy index and mode matching to standard optical fibers, low propagation loss, and good durability. When crystalline materials are used, they have to be single-crystalline to avoid grain boundary scattering (for instance, Ti-diffused waveguide in single-crystal $LiNbO_3$). If semiconductors are used, the bandgap of the material needs to be sufficiently high so as to avoid absorption of the photons. Silicon waveguide (bandgap = 1.1 eV) on a silica cladding (SOI) is commonly used system in PLCs used for telecommunication applications.

The PWLs described above are fabricated in rare-earth doped glass waveguides. Waveguides have to be much shorter than fiber, so the challenge of incorporating the rare-earth ions at high concentration in the host material has to be overcome. Planar waveguide lasers are a compressed version fiber lasers that have the unique advantage that they can be directly integrated with other components on the same chip using very large-scale integration (VLSI)-style processing. The concentration of Er (or the lasing ion in general) in a PWL needs to be many times that in fiber amplifiers (~100 ppm). The need for high gain in a short length requires high pump powers and high lasing ion concentration, both of which lead to nonideal behavior. Since rare earth ions tend to have multiple excited electronic states, there can be absorption at the excited state level as well. When a higher energy state is separated from the upper emission level by the energy in one laser photon, the ions can undergo excited state absorption or ESA (Figure 6). Rare-earth ions in close proximity tend to undergo cooperative upconversion processes, where two excited ions transfer energy between each other so that neither ion may fluoresce at the desired wavelength. These processes can negatively affect the

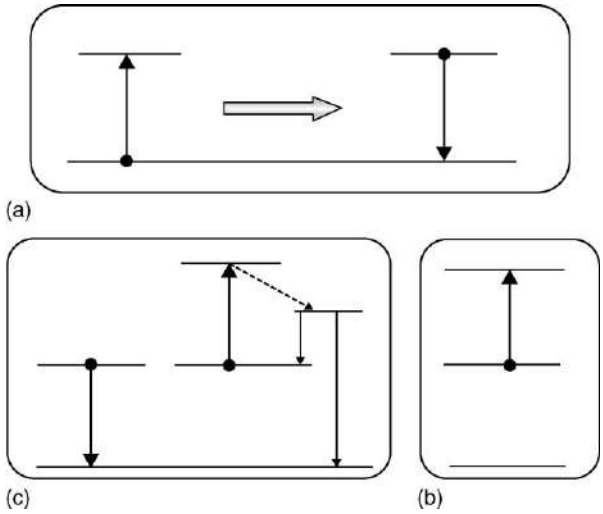

Figure 6 The energy level diagram for a representative two level system. The inset (a) shows the excitation and emission process. Excited state absorption is shown in (b) and cooperative upconversion is shown in (c).

useful gain of the system. Ensuring high spatial dispersion of the Er atoms is critical to minimize this effect. Special glass compositions, such as phosphate glasses, are used here to ensure that the Er is not phase separated or clustered. These glasses also provide a high induced cross-section for gain. Glasses with low phonon energy are also desirable to minimize phonon-mediated nonradiative decays in the rare earths. Light emitting semiconductors and polymers containing fluorescent dyes have also been used to make PWLs.

Now that we have gained an overall perspective of PWLs, we can discuss different manifestations. The following discussion uses a Er doped system as a model since there is a substantial body of work on this system.

Distributed Bragg Reflector (DBR) PWL

In planar waveguide systems, it is difficult to construct efficient mirrors or reflective facets perpendicular to the waveguide. However, mirrors can be replaced by Bragg gratings, which be fabricated relatively easily. In-line Bragg gratings result in wavelength specific reflection as per eqn [4]:

$$\Lambda = \frac{m\lambda}{2\bar{n}} \qquad [4]$$

where Λ is the grating pitch, m is the order of the grating, and λ is the central Bragg wavelength. Gratings have another important benefit in that they ensure that unwanted frequencies outside the limited reflection spectrum are dispersed before they reach the lasing threshold.

A DBR laser is similar in principle to a Fabry–Perot design but uses a Bragg grating on either end of planar Er-doped waveguide. This is shown schematically in Figure 7. Gratings can be made in a number of different ways. One popular way is to exploit the photosensitivity of glasses using 193 nm or 244 nm UV light to periodically alter the refractive index of a glass waveguide. Photosensitivity occurs by a combination of molecular changes (such as the formation of oxygen deficiency sites or color centers) as well as changes in the material strain in the region exposed to the UV light (Figure 7c). Alternatively, gratings can also be formed using etching to make periodic grooves in the waveguide. These are called corrugated or surface relief gratings (Figure 7a and 6b).

The back Grating 1 has a high reflectance of almost 100%, whereas Grating 2 is a weaker grating. The reflectance strength of Grating 2 is determined by taking into account the gain coefficient of the material and the output power required from the laser. The reflection intensity can be altered according to the following equation:

$$R = \tanh^2\left[\frac{\pi L_G \Delta n_{\text{eff}} \eta(V)}{\lambda_B}\right] \qquad [5]$$

where L_G is the grating length, Δn_{eff} is the index modulation of the grating, λ_B is the Bragg wavelength, $\eta(V)$ is the confinement factor, a function of the waveguide parameter V that represents the fraction of the integrated mode intensity contained in the core. Increasing the reflectance strength also widens the linewidth of reflected light. The overlap of the spectral width between the back (Grating 1) and the front grating (Grating 2) becomes the effective output window within the gain spectrum of the lasing material (Figure 8).

Not unexpectedly, a DBR structure generates longitudinal and lateral modes. These modes are, however, limited to the overlapping section of the reflection spectrum of the two gratings, as opposed to the entire gain spectrum as in Fabry–Perot lasers (Figure 9). Yet, this can lead to problems such as spectral hole burning and mode hopping, whereby the dominant mode hops from one wavelength to another. The separation between the modes needs to

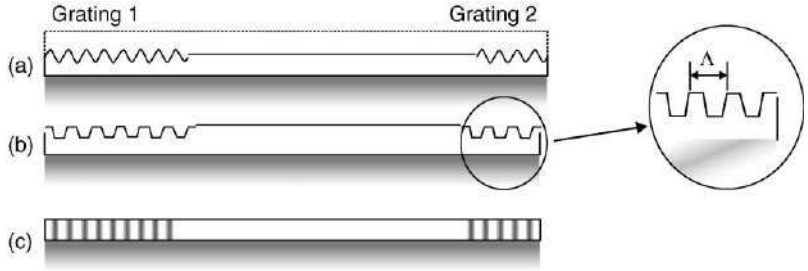

Figure 7 DBR planar waveguide laser. The PWLs shown in (a) and (b) have sinusoidal and trapezoidal surface-relief gratings respectively etched in their cores. An in-line grating of the kind seen in (c) could be fabricated using the photosensitivity of the glass. The top cladding is shown only in (a).

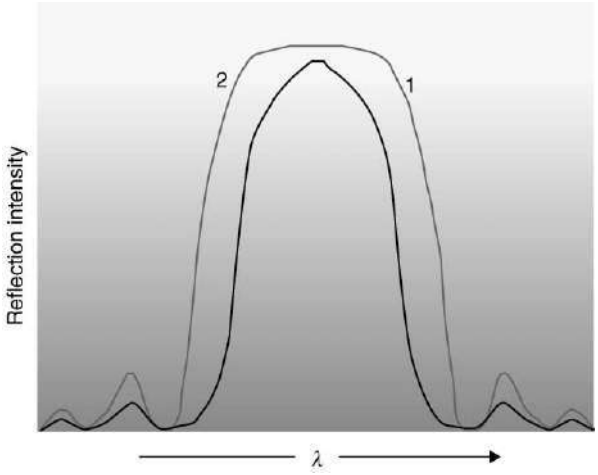

Figure 8 The reflection spectrum of the back (1) and the front (2) gratings. As the intensity of the reflected light is made to increase, so does the linewidth. Hence, the narrower spectrum of partially reflecting grating 2 tends to dictate the operating wavelength range of a DBR.

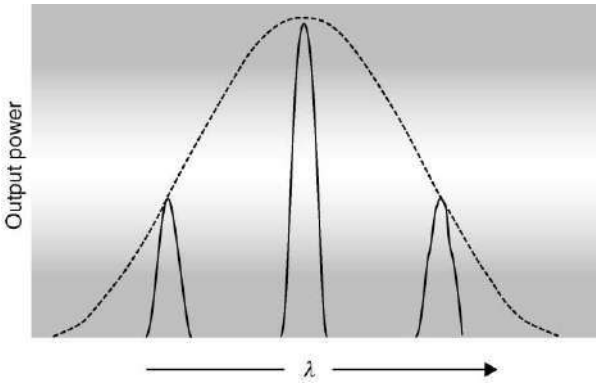

Figure 9 A sample output from a DBR laser, shown in this case with 3 modes. The dotted line shows the effective overlap between the two gratings. The existence of the modes depends on several factors, in particular the pump intensity.

be made sufficiently high compared to the spectral overlap so that single-mode operation can be achieved. But in practice the length of the cavity (L_c in eqn [3]) is determined by factors such as power and space and may not be independently altered to achieve this. However, the reflection spectrum of the back and the front gratings could be slightly shifted in opposite directions to realize a smaller overlap within which only one or a few modes may exist (Figure 10). The occurrence of a single mode also depends on the gain profile of the lasing material and the optical pump power. In Er-doped PWLs, modes other than the primary one can be suppressed by using a pump power below a threshold limit and stable operation can be thus achieved.

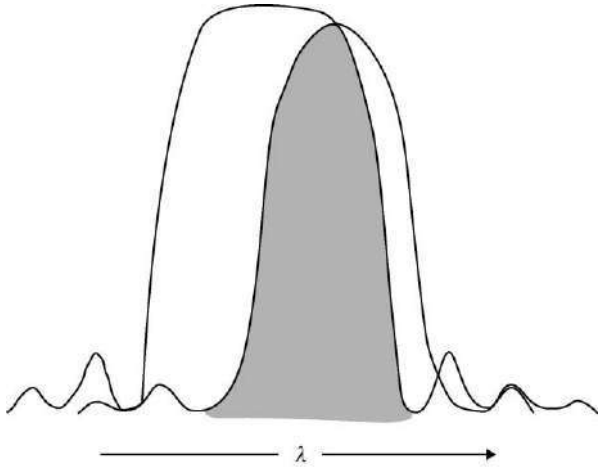

Figure 10 Shifting the front and back gratings is one means of controlling the width of the overlap region, shown in gray. This way a single mode operation can be achieved even for a longer cavity with many modes.

Distributed Feedback (DFB) PWL

The modal problems listed above can be addressed by having a single, distributed grating that is placed along the entire length of the lasing waveguide itself (Figure 11). Alternatively, the gratings may be also placed in the upper or the lower claddings to minimize loss of optical power due to scattering because of the longer grating. In this case, it is the evanescent portion of the light that interacts with the grating. In all cases, the periodicity of the grating leads to a condition for a single lasing wavelength based on constructive interference. The strongest mode occurs when the period Λ of the Bragg grating satisfies the primary Bragg condition (for first order grating or $m = 1$) in eqn [4]. The device will also function if the grating pitch is equal to small integer multiples of $(\lambda/2\bar{n})$.

The grating needs to be made strong enough to generate sufficient feedback (reflections), which also widens the linewidth of the spectrum. To ensure a single narrow linewidth output, a quarter wave phase shift is introduced in the grating, as seen in Figure 11. This phase shift creates a transmission fringe in the spectrum (Figure 12) that significantly narrows the linewidth of the output signal. This design is considered to be standard for the DFB lasers of today. Figure 13 shows a typical output of a DFB laser.

Any of the geometries seen in Figure 1 can be used to make the waveguide for the laser. Due to processing constraints, rib (c) and ion diffused (d) configurations are commonly used. Waveguide losses, characteristically due to scattering from sidewall roughness, have to be low. Otherwise, this adds a

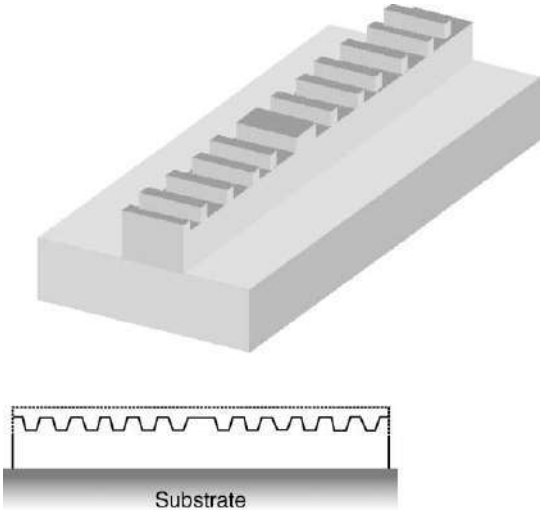

Figure 11 A DFB planar waveguide laser.

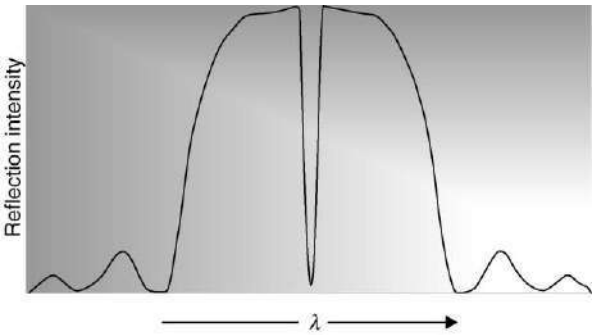

Figure 12 The reflection spectrum of a grating with a central $\lambda/4$ shift in the pitch, showing the transmission fringe.

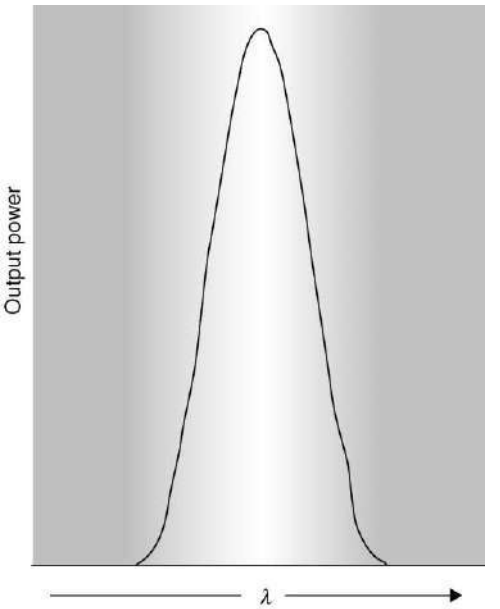

Figure 13 The typical output of a DFB PWL. The linewidth can be made to be as narrow as a few hundred kHz.

significant overhead onto the pump power and lowers the output of the PWL. Birefringence caused by thermal expansion coefficient differences and excessive sidewall roughness also need to be minimized as they can substantially alter the loss of the TE mode compared to TM.

Though their spectral properties are ideal as sources for long haul transmission, DFB lasers tend to be less powerful than the corresponding DBR designs. But the DBR PWLs are very sensitive to reflections that can cause a broadening of the laser linewidth or can act as a source of noise. In both cases, single mode lasers with very narrow linewidths in the kHz range with low radiance and noise, have been demonstrated. But neither is known to have been commercialized yet.

PWL Arrays

One of the unique advantages of the PWLs is that they are often processed using VLSI techniques. One important implication of this is the ability to make a PWL array where multiple lasers are formed simultaneously on the same chip, each with a unique output (Figure 14). Such arrays are likely to be very useful in future dense wave division multiplexing (DWDM) optical network systems where as many as 128 channels (and more) are planned. Since packaging of individual lasers can be up to 75% of the total cost, making chips with multiple lasers is very attractive. The lasers may be DBRs or DFBs whose output wavelength is controlled by changing the grating parameters. To vary the individual wavelengths either the effective index and/or the pitch of the grating can be changed. Thus, any output spectrum can be designed and executed merely by making the appropriate lithography mask.

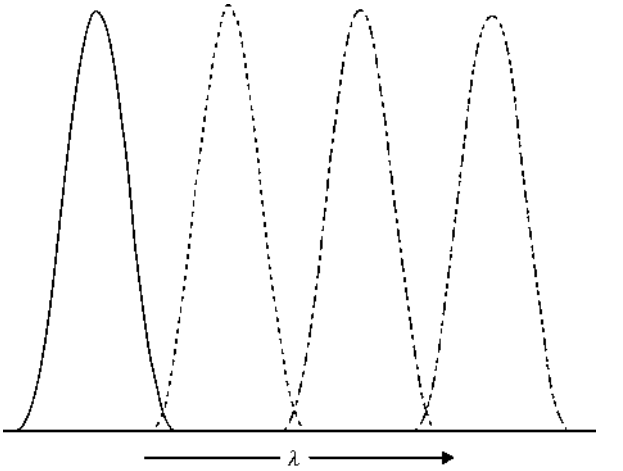

Figure 14 Idealized output from a 4-PWL array.

The gratings are often written holographically, where the wafer is exposed to an interference pattern caused by two coherent light beams. The wafer underneath can be sensitized with a photosensitive glass layer or a film a photoresist. It is not efficient to do multiple holographic exposures to individually change each of the gratings in the array. One way of doing this with a single holographic exposure is by constructing the waveguides at different relative angles. Then the component of the index modulation, resolved along the axis of the waveguide, would vary for each waveguide depending on the relative waveguide angle (Figure 15); this would lead to differences in $\bar{n}$. Alternatively, the waveguide array can be made of the same height but of different widths. This is easily done since waveguide height is usually deposited uniformly but the width is a function of the mask layout. Here the waveguide and the grating are orthogonal to each other and the variation in wavelengths is achieved by the fact that the effective propagation constant for each of the waveguides is different.

Arrays of PWLs have been made on a silicon substrate using deposition of an Er containing glass that is subsequently patterned into waveguide, and ultimately laser, arrays. Another approach has been to start with a phosphate glass as the host and the substrate. A phosphate glass is a superior host for rare-earth ions such as Er since the sensitization efficiency is nearly unity and much high doping levels are possible before concentration quenching effects set in. Waveguides are created by ion-exchange by immersion in a hot bath containing suitable ions (Na for Li, for instance). When starting with a uniformly doped glass, one issue in the assimilation of the laser array with other components is the fact that Er-doped region does not start and end with the waveguide laser section. Recently, the Er has been deposited selectively in the laser waveguide in phosphate glass using methods such as ion implantation, bringing the concept of integration closer to reality.

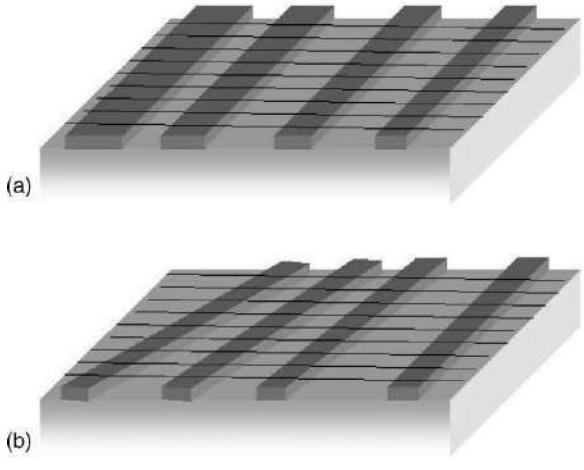

Figure 15 Two different means of making planar waveguide laser array with a single holographic exposure to write the gratings. The darker rectangular blocks represent the core channels or ribs. In (a) the width and hence the effective index of the waveguide is varied whereas the relative angle between the waveguides is altered in (b). The lines represent the direction of the gratings, which is orthogonal only to the left most waveguide in (b).

Pumping

One of the key issues here is the optical pumping of the rare-earth PWLs to generate acceptably high output power. Individual diode lasers or an array of diode lasers all on a single chip (called a diode bar) have been utilized as pumps. Both single and multimode diode lasers are available and have been used. There are pros and cons for either choice: whereas single mode lasers are less powerful (in the range of few 100 mW), it is difficult to convert the multimoded output of their counterparts to useful power. The choice is determined by the waveguide design, which needs to be optimized for high overlap of the lasing mode with the pump. Carefully tapered waveguides can be used distribute the pump power from a diode bar to the various lasers in an array without significant loss.

PWL Stability

Stable operation of a PWL is one of the big challenges. In DBR PWLs, it is seen that there is a threshold for pump power only below which single moded operation can be achieved. Additionally, any PWL die needs to be made immune to ambient temperature changes. These changes can affect many of the critical output parameters of the PWL: wavelength, linewidth, power and in the case of an array, the wavelength separation. This occurs primarily due to the thermal shift in the grating pitch as well as the thermo-optic effect. It is overcome by supplanting the laser die onto a heat spreader chip made from conducting material such as Cu or BeO, followed by a thermoelectric cooler. Ironically, the laser's output wavelength and linewidth are fine-tuned and 'fixed' using temperature to overcome any drifts due to processing errors or other reasons.

Reliability is another important issue that needs to be addressed, especially for telecommunication applications where these types of components have to pass high quality standards such as the Telecordia qualification in the USA.

Other PWL Systems

Many of the PWLs to date have been developed using Er doping, which emits light in the range of interest for telecommunications (the C-band and even the L-band). Here, Yb can be added as a co-dopant to improve power conversion efficiency. Other fluorescing ions have also been used: Nd, Tm, Ho, and Cr. Several research teams have successfully demonstrated Nd-YAG PWLs, in particular. Ridge waveguide lasers have been made using semiconductor materials InGaAs/InP heterojunctions in both Fabry–Perot and DFB configurations. These systems are electrically pumped. DFB PWLs have also been made using dye doped polymers.

It must be mentioned that there are other ways to use waveguides to make lasers. The primary example is a configuration where the waveguide resonator is external to the active region. Designs, such as ring resonator lasers and an external cavity lasers, fall in this category.

Future research and development of PWLs would undoubtedly utilize photonic crystal structures. These structures have a periodic fluctuation in the index of refraction in two or three dimensions. Owing to the resultant Bragg reflections, specific allowed and forbidden states are created which can be used to control the confinement of light. Hitherto, these structures have been used to form mirrors or hexagonal ring resonators. These more sophisticated laser designs are likely to enable new functions such as tunability of the output spectrum.

Summary

Planar waveguide lasers are a special class of laser where light is confined to a waveguide. They have distinctive advantages that include high optical gains, low laser thresholds, narrow linewidths in the kHz range, and optimal thermal management. Significantly, they afford a platform that can be readily expanded to include an array of lasers, each with an unique output, and can ultimately be combined with multiple optical components on the same chip. Such compact integrated devices are bound to take photonics to a new high level of capability.

List of Units and Nomenclature

a	integer denoting number of laser modes
E	electric field
k_z	propagation constant for travel in z direction
L_c	length of laser cavity
L_G	length of grating
m	order of the grating, an integer
n	refractive index of the medium
$\bar{n}$	modal effective refractive index
R	reflectance
V	normalized frequency or the V parameter of the waveguide
Δn_{eff}	modulation of the effective index of the grating
η	confinement factor or fraction of the light confined in the region of interest
λ	wavelength of the lightwave
λ_B	resonant wavelength of a Bragg grating or Bragg wavelength
Λ	grating pitch
Φ	electric field distribution

See also

Lasers: Carbon Dioxide Lasers; Dye Lasers. **Optical Amplifiers:** Erbrium Doped Fiber Amplifiers for Lightwave Systems.

Further Reading

Carroll JE, Whiteaway JEA, Plumb RGS and Plumb D (1998) *Distributed Feedback Semiconductor Lasers*. London: IEE Publishing.

Ghafouri-Shiraz H (2003) *Principles of Semiconductor Laser Diodes and Amplifiers: Analysis and Transmission Line Laser*. London: Imperial College.

Hecht J (2001) *Understanding Lasers: An Entry Level Guide*. New York: Wiley Interscience.

Lee JR, Baker HJ, Friel JG, Hilton GJ and Hall DR (2002) High-average-power Nd:YAG planar waveguide laser, face-pumped by 10 laser diode bars. *Optical Letters* 27: 524–526.

Milonni PW and Eberly JH (1988) *Lasers*. New York: John Wiley and Sons.

Saleh BEA and Teich MC (1991) *Fundamentals of Photonics*. New York: John Wiley & Sons.

Shepherd DP, Hettrick SJ, Li C, *et al.* (2001) High-power planar dielectric waveguide lasers. *Journal of Physics D: Applied Physics* 34: 2420–2432.

Siegman AE (1986) *Lasers*. Harendon, Virginia: University Science Books.

Svelto O (ed.) (1998) *The Principles of Lasers*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Veasey DL, Funk DS and Sanford A (1999) Arrays of distributed-Bragg-reflector waveguide lasers at 1536 nm in Yb/Er codoped phosphate glass. *Applied Physics Letters* 74(6): 789–791.

Young M (2001) *Optics and Lasers: Including Fibers and Optical Waveguides*. NewYork: Springer-Verlag New York, Inc.

Semiconductor Lasers

S W Koch, Philipps-University, Marburg, Germany

M R Hofmann, Ruhr-Universität Bochum, Bochum, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

This article discusses the principle of operation and the basic material engineering aspects for the design of semiconductor lasers. The physics of the gain medium is analyzed in detail and the role of many-body effects due to the Coulomb interaction of the carriers for the optical properties of the active semiconductor material is demonstrated. The two main types of laser resonators for diode lasers, Fabry–Perot resonators and vertical cavity structures, are described. Finally, the dynamics of semiconductor lasers, including the principles of short pulse generation and nonequilibrium gain dynamics, is addressed.

Basic Principles

The optical emission from semiconductor lasers arises from the radiative recombination of charge carrier pairs, i.e., electrons and holes in the active area of the device. The conduction-band electron fills the valence-band hole by simultaneously transferring the energy difference to the light field. This is called radiative recombination. Hence, the frequency of the laser light is mainly determined by the bandgap of the semiconductor laser material. In order to achieve lasing, one needs carrier inversion, i.e., a sufficient number of electrons in the conduction band so that the probability of light absorption is lower than the probability of emission.

It has been shown that a large variety of semiconductor materials is suited for semiconductor lasers. The most stringent requirement for the material is that it has a direct bandgap. In other words, the maximum of the valence band and the minimum of the conduction band have to be at the same position in k-space (in the center of the Brillouin zone). The elemental semiconductors silicon and germanium do not fulfill this condition and therefore cannot be used for semiconductor lasers.

In order to achieve carrier inversion, it is necessary to pump the semiconductor laser, i.e., to excite a sufficient number of electrons from the valence band into the conduction band. Even though this pumping in semiconductors, as in other solid state laser materials, can be done optically, the large technological impact of semiconductor lasers is at least in part due to the possibility of electrical pumping with a few tens of milli-amperes at low voltages.

The original version of such a semiconductor laser device is shown in Figure 1, The laser structure consists of a GaAs-based p-n junction which is biased in the forward direction. Electrons are injected from the n region and holes from the p region, respectively. Due to the diode characteristics of the p-n junction, semiconductor lasers are often also called laser diodes. In the device shown in Figure 1, the laser mirrors are formed by the cleaved facets of the semiconductor crystal. Because of the high refractive index of GaAs, the semiconductor–air interface provides a reflectivity of about 32% which is sufficient for laser emission.

However, the early type of laser, the so-called homojunction device, was not very efficient because of the low carrier density in the active area and because of the weak overlap between the inverted region and the optical mode. Considerable improvement towards room-temperature continuous wave (cw) operation was achieved by the introduction of the so-called double heterostructure. Its principle is shown in Figure 2.

The basic idea behind the double heterostructure laser is that the active material (e.g., GaAs) is embedded in another semiconductor with a slightly larger bandgap (e.g., (AlGa)As). This causes a potential well in which electrons and holes are confined in order to achieve high carrier densities in the active area. Moreover, the smaller refractive index of the surrounding high-bandgap material helps to guide the optical field within the region of highest inversion. Additional ridge-shaped lateral structuring is frequently used to form a complete optical waveguide.

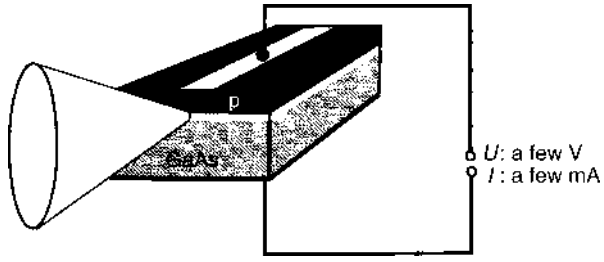

Figure 1 Principle of operation of a semiconductor laser.

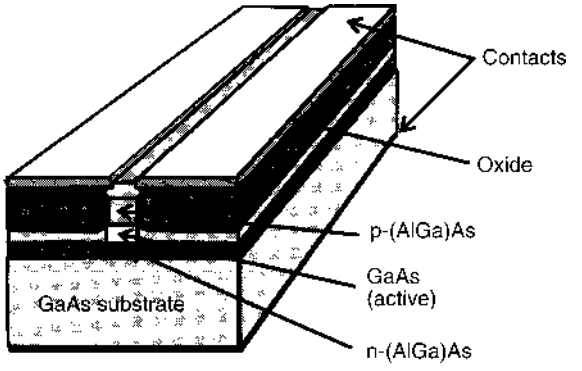

Figure 2 Double heterostructure laser diode.

Nowadays, even more advanced laser structures can be fabricated, as modern crystal growth techniques such as molecular beam epitaxy (MBE) and metal organic chemical vapor deposition (MOCVD) have become available. These techniques allow the grower not only to determine the composition of the semiconductors with remarkable precision, but also to define the shape virtually on an atomic scale. In particular, it is now possible to grow microstructures so small that their electronic and optical properties deviate substantially from those of the corresponding bulk materials. So-called quantum-well (QW) structures, in particular, turned out to be superior to bulk materials in many respects and are currently the active medium in most commercial semiconductor lasers. The gain medium in QW laser structures consists of one or more films with a thickness of a few atomic monolayers. These are embedded in a barrier material with a larger bandgap that serves to confine the carriers to the wells. In QWs the quantum confinement restricts the free carrier motion to the two dimensions within the plane of the films, since the well thickness is comparable to or less than the thermal wavelength of the carriers. Hence, one speaks of quasi-two-dimensional structures where no free motion perpendicular to the films is possible. The corresponding energy states of the carriers are quantized, i.e., only discrete values are allowed for that part of the carrier kinetic energy which is related to the confined direction.

A major consequence of this restriction of the carrier motion to two dimensions is a change in the (single particle) density of states. This density of states is altered in QWs in comparison to bulk material. The density of states for bulk material is zero at the bandgap energy and rises with the square root of the energy for higher energies. In contrast, the density of states for a two-dimensional carrier system in a QW is a step-like function with a density of states different from zero at the QW-well bandgap (the bandgap energy of the active material plus the quantization energies of electrons and holes in the well). This increase of the density of states at the bandgap promises important advantages of QW structures over bulk material. Most importantly, the number of states that have to be occupied before inversion is obtained is considerably reduced in QW structures. This reduces the threshold current and increases the modulation speed of the laser structures. Accordingly, most of the modern laser structures are QW structures.

These arguments have encouraged many researchers to develop structures with further increased carrier confinement in two dimensions (quantum wires) and three dimensions (quantum dots). The single-particle density of states is even more favorable in quantum wires and quantum dots than in QWs: in both cases the single-particle density of states peaks at the effective bandgap energy. This argument seems to promise even better performance of quantum wire and quantum dot lasers as compared to QW devices. However, the single-particle density of states description completely neglects the Coulomb interaction of the carriers which strongly influences the optical and electronic properties of semiconductor structures. The absorption is modified by excitonic effects that lead to additional resonances below the bandgap and to changes of the continuum absorption. These changes are often quantified by the so-called Sommerfeld factor, which depends on the dimensionality of the system, i.e., it differs for bulk, QW, quantum wire and quantum dot structures. For example, the Sommerfeld factor completely suppresses the singularity of the density of states at the bandgap of quantum wire structures. Accordingly, with inclusion of Coulomb effects, the benefit of the quantum confinement is much less than expected from the single-particle density of states.

The use of quantum wire and quantum dot structures as a semiconductor laser gain medium is still in its infancy even though they can be grown with various degrees of precision. However, quantum dot structures, especially, hold great promise for the future, since they can be grown in large numbers using techniques similar to those already used for QWs. The possibility of layer growth under strained conditions, such as QWs between barrier material with a slightly different atomic lattice constant, also results in a greater flexibility to design the laser gain medium with the desired emission wavelength.

In principle, the emission wavelength of a semiconductor laser can be roughly designed by the choice of the material. Different semiconductors have different bandgaps between the valence band and conduction band and the bandgap energy determines the spectral range of the emission. One can tune the bandgap and thus the spectral range of

the emission by fabricating semiconductor alloys as, e.g., (GaIn)As or (AlGa)As. Presently, a large variety of different materials is commercially available covering, for example, the red ((AlGaIn)P), the near infrared ((AlGa)As, (GaIn)As), the telecom ((GaIn)(AsP)) and even the blue–ultraviolet range ((GaIn)N).

Great care has to be taken that the crystal used for a semiconductor laser is of almost perfect quality. Dislocations in the crystal, for example, have to be avoided as well as any kinds of defects since they act as nonradiative recombination centers. Nonradiative recombination competes with the radiative recombination that is essential for the laser process. Nonradiative recombination effects include Auger recombination and recombination via defects. Auger recombination is the process where an electron and a hole recombine and completely transfer the energy difference to another carrier (electron or hole) which is excited to a higher state. This is an intrinsic effect that depends mainly on the band structure and the relevant values of the transition probabilities (transition matrix elements). The probability for Auger recombination is temperature- and carrier-density-dependent and can be minimized, e.g., by properly cooling the device and by keeping the carrier densities as low as possible.

Nonradiative recombination via defects is an extrinsic effect that depends on the material quality and can thus be minimized by optimized growth. However, trying to avoid dislocations limits the choice of materials for laser structures severely. The growth processes require 'host' crystals, so-called substrates, on which the real structure can be grown. GaAs, InP, SiC and sapphire wafers are available in sufficient quality as standard substrates for optoelectronic applications. The epitaxially grown material adapts to the crystal structure of the substrate which implies that the lattice constant of the substrate and structure material should be very similar. If the lattice constants do not agree, i.e., if there is a structural mismatch, a certain amount of strain develops which increases with increasing lattice mismatch. Too much strain leads to the formation of dislocations which are detrimental to laser applications. Control of the dislocation density is currently one of the critical issues for the wide-gap III–V materials, such as GaN, because no substrate is available that matches the GaN lattice constant. However, a small amount of strain is often even desirable as long as it does not cause dislocations. Strain influences the band structure of the material in a predictable way and can therefore be used to tailor the laser emission properties.

Thus laser materials for a certain desired wavelength range have to fulfill two important requirements: first, the bandgap of the material has to fit the desired photon energy; and second, and more crucial, an appropriate substrate with the right lattice constant has to be available. This second condition excludes many material compositions from laser applications. In many cases, both conditions can only be simultaneously met with quaternary semiconductor alloys (for example (GaIn)(AsP) on InP substrates).

Physics of the Gain Medium

From a material physics point of view, the important structural aspects of a certain laser material are summarized in the band structure of the gain medium, i.e., in the quantum mechanically allowed energy states of the material electrons. The calculation of the band structure of a particular gain medium is therefore a crucial aspect of semiconductor laser modelling. Additionally, however, one has to understand the relevant properties of the excited electron–hole plasma in the active region of the laser.

In its ground state, i.e., without excitation and at very low temperatures, a perfect semiconductor is an insulator where no electrons are in the conduction-band states. The presence of a population inversion, i.e., occupied conduction band and empty valence band states in a certain range of frequencies, is, however, a requirement for obtaining optical gain and light amplification from a semiconductor. Sufficiently strong pumping leads to the generation of an electron–hole plasma, where the term 'electron' specifically refers to conduction-band electrons and 'holes' to the missing electrons in the originally full valence bands. Holes are also quasiparticles, just as the electrons. Electrons in a solid are quasiparticles with an effective mass that is considerably different from the free-electron mass in vacuum and that is determined by the band structure, i.e., the interaction with all the ions of the solid. The properties of the missing electrons, such as a positive charge for the missing negative electron charge, a spin opposite to that of the missing electron, etc., are attributed to the holes. As a consequence of their electrical charges, the excited electrons and holes interact via the Coulomb potential which is repulsive for equal charges (electron–electron, hole–hole) and attractive for opposite charges (electron–hole). Furthermore, electrons and holes obey Fermi–Dirac statistics and the Pauli exclusion principle that forbids two Fermions from occupying the same quantum state. The consequences of the Pauli principle are often referred to as 'band-filling effects' in the physics of semiconductor lasers.

The theory therefore has to account not only for the band-structure but also for this band filling, i.e., the detailed population of the electron and hole states. Taking into account only the band structure and band filling leads to a so-called free-carrier theory. But the electron–hole plasma in a semiconductor gain medium is an interacting many-body system because of the carrier interactions and the exclusion principle. In a many-body system the properties of any one carrier are influenced by all the other carriers in their respective quantum states. Generally, advanced many-body techniques are needed to systematically analyze such an interacting electron–hole plasma and to compute the resulting optical properties, such as gain, absorption or refractive index, all of which are changing in a characteristic way with the changing electron–hole density.

It is often helpful to investigate the characteristic time-scales and the most direct consequences of the various interaction effects even though the consequences of the different many-body effects are in general not additive. The fastest interaction, under typical laser conditions, is the carrier–carrier Coulomb scattering that acts on scales of femto- to picoseconds to establish Fermi–Dirac distributions of the carriers within their bands. These distributions are often referred to as 'quasi-equilibrium distributions' since they describe electrons and holes in the conduction and valence bands which may be characterized by an 'electronic temperature' or 'plasma temperature.' The plasma temperature is a measure for the mean kinetic energy of the respective carrier ensemble and relaxes towards the lattice temperature as a consequence of electron–phonon (= quantized lattice vibration) coupling, i.e., the carriers can emit or absorb phonons. The coupling to longitudinal optical phonons in polar materials is especially strong with scattering times in the picosecond range. The corresponding times for the electron–acoustic phonon scattering are in the nanosecond range. Generally, the carrier–phonon interaction provides an exchange of kinetic energy of the carriers with the crystal lattice in addition to the relaxation of the initial nonequilibrium distribution. Deviations between plasma and lattice temperature may occur, especially when lasers are subject to strong pumping or rapid dynamic changes.

The Coulomb interaction not only causes rapid carrier scattering but also directly influences the laser gain spectrum. Gain, i.e., negative absorption, occurs in the spectral region of population inversion, in other words, where the net probability for a test photon to be absorbed is lower than the probability to be amplified by stimulated electron–hole recombination. Energetically, the gain region is bounded from below by the effective (or renormalized) semiconductor bandgap and from above by the electron–hole chemical potential. Here, the chemical potential is the energy at which neither absorption nor amplification occurs, i.e., where the semiconductor medium is effectively transparent.

The effective bandgap energy for elevated carrier densities is significantly below the fundamental absorption edge of an unexcited semiconductor. This 'bandgap renormalization' is a consequence of the fact that the Pauli exclusion principle reduces the repulsive Coulomb interaction between carriers of equal charge. Furthermore, the Coulomb screening, i.e., the density-dependent weakening of the effective Coulomb interaction potential, contributes to the reduction of the bandgap.

Another many-body effect originates in the Coulomb attraction between an electron and a hole, and leads to an excitonic or Coulomb enhancement of the interband transition probability. All these effects contribute to the detailed characteristics of the optical spectra of a semicondcutor gain medium. Figure 3 shows an example of calculated spectra in comparison to experimental data.

The example demonstrates that the microscopic theory correctly accounts for the changes in the gain spectrum with changing carrier density by means of a consistent treatment of band structure, band-filling, and many-body Coulomb effects.

The Resonator

In addition to a gain medium, a laser needs a positive feedback process for the amplified stimulated emission. Hence, the inverted medium has to be

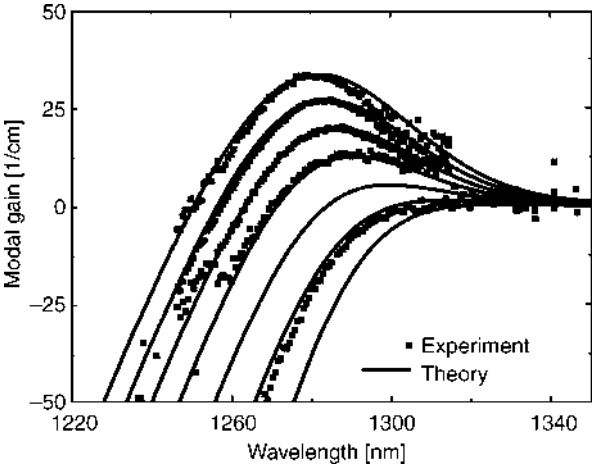

Figure 3 Comparison of experimental and theoretical gain spectra of a (GaIn)(Nas)/GaAs laser.

embedded into a resonator. One of the conceptually simplest versions of such a resonator is a so-called Fabry–Perot resonator which consists of two parallel plane mirrors. A double heterostructure Fabry–Perot laser is shown in Figure 2. Such Fabry–Perot resonators are very easy to fabricate for semiconductor lasers: the semiconductor crystal is cleaved perpendicular to the optical waveguide. The cleaved edges provide a reflectivity of about 32% due to the high refractive index of the semiconductor. This is sufficient to provide enough feedback for laser operation.

However, despite its conceptual simplicity, the cleaved-edge Fabry–Perot concept also introduces a number of problems. First, the light output occurs on two sides of the laser, whereas for most applications all the output power should be emitted out of one facet only. For practical applications, this problem can be solved by depositing coatings onto the cleaved facets of the structure. For example, deposition of a 10% reflectivity coating on one facet and of a 90% reflectivity coating on the other facet concentrates the laser emission almost completely to the facet of lower reflectivity. Antireflection coatings with reflectivities below 10^{-4} are used when laser diodes are coupled to external cavities, e.g., in tunable lasers.

A second problem with Fabry–Perot lasers is that the emission wavelength is often not well defined. A Fabry–Perot resonator of length L (optical length nL) has transmission maxima – so-called modes – at all multiples of $c/2nL$ where c is the speed of light and n the refractive index. That means that for typical devices with a length of 500 μm, hundreds of modes are present even within the limited gain bandwidth of a semiconductor. These modes often compete with each other resulting in multimode emission which may become particularly complex and uncontrollable when the laser is modulated for high-speed operation.

This problem can be overcome in so-called distributed feedback (DFB) lasers. In these devices, the refractive index is periodically modulated along the waveguide. In a simplified picture, this periodic pattern causes multiple reflections at different locations within the waveguide. For certain wavelengths, these multiple reflections interfere constructively to provide enough 'distributed' feedback for laser operation. With the DFB concept, single-mode emission can be realized even for high-speed modulation of the lasers. Alternatively, the region of distributed feedback can be separated from the region of amplification in so-called distributed Bragg reflector (DBR) lasers. These are devices with multiple sections where one section is responsible for the optical gain and another section (the DBR section) is responsible for the optical feedback.

So far, we have discussed so-called edge emitting lasers where the light emission is in the plane of the active area of the device. The lengths of these lasers are typically hundreds of microns and thus correspond to a few hundreds of optical wavelengths. This causes one of the major disadvantages of edge emitters – the multimode emission. Another disadvantage is the poor elliptic beam profile due to diffraction at the small rectangular output aperture. Finally, the fabrication procedure of edge emitters is complex since it requires multiple steps of cleaving before it is even possible to test the laser.

A novel alternative semiconductor laser concept was realized in the early 1990s: the vertical-cavity surface-emitting laser (VCSEL). A typical VCSEL structure is shown in Figure 4. In VCSELs, the light emission is parallel to the growth direction, i.e. perpendicular to the epitaxial layers of the structure. The cavity length of a VCSEL is in the order of a few (1–5) multiples of half the wavelength of the emitted light in the material. These small dimensions imply that the interaction length between the active medium and the light field in the resonator is very short. Therefore, the mirrors of a VCSEL have to be of very good quality; reflectivities of more than 99% are required in most configurations.

Very high reflectivities can be achieved with dielectric multilayer structures, so-called Bragg reflectors. Bragg reflectors consist of a series of pairs of layers of different refractive indices, where each layer has the thickness of a quarter of the emission wavelength. Most preferably, Bragg reflectors are epitaxially grown together with the active area of the device in only one growth process.

Small VCSEL resonators with Bragg reflectors at both ends of the cavity are called microcavities. They are designed in such a way that only one longitudinal mode is present in the region of the semiconductor gain spectrum. Accordingly, VCSELs are well suited to provide single-mode characteristics. Moreover, a round aperture for light emission can be realized with a diameter of a few micrometers so that

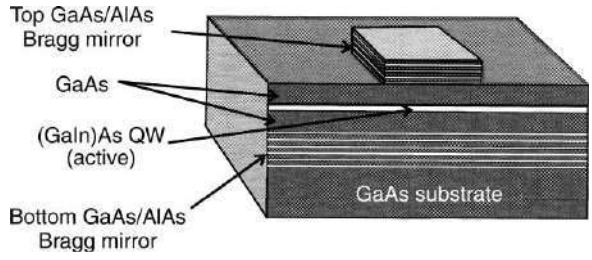

Figure 4 Vertical-cavity surface-emitting laser (VCSEL).

diffraction is weak and the beam profile is almost perfect.

VCSELs can be tested on wafer without further processing and can be arranged in two-dimensional arrays. A major disadvantage of VCSELs is their temperature dependence. The bandgap of the active material and the cavity mode depend on temperature in such a way that the cavity mode shifts away from the region of maximum gain when the temperature is varied. This severely limits the temperature range of operation for VCSELs.

Semiconductor Laser Dynamics

The dynamical behavior of semiconductor lasers is rather complex because of the coupled electron–hole and light dynamics in the active material. Like other laser materials, semiconductor lasers exhibit a typical threshold behavior when the pump intensity is increased. For weak pumping, the emission is low and spectrally broad; the device operates below threshold and shows (amplified) spontaneous emission. At threshold, the optical gain compensates the losses in the resonator (i.e., internal losses and mirror losses) and lasing action starts. Ideally the output power increases linearly with the injection current above threshold. The derivative of this curve is proportional to the differential quantum efficiency. The differential quantum efficiency can reach values close to unity and even the overall efficiency of the diode laser, i.e., the conversion rate from electrical power into optical power, can be as high as 60% which is far higher than for any other laser material.

Ideally, the laser emission is single mode above threshold. However, as discussed above, edge-emitting diode lasers tend to emit on multiple longitudinal modes and the competition of these modes can lead to complex dynamical behavior which is significantly enhanced as soon as the laser receives small amounts of optical feedback. Semiconductor lasers are extremely sensitive to feedback and tend to show dynamically unstable, often even chaotic, behavior under certain feedback conditions.

Like other lasers, semiconductor lasers respond with relaxation oscillations when they are suddenly switched on or disturbed in stationary operation. Relaxation oscillations are the damped periodic response of the total carrier density and the optical output intensity as a dynamically coupled system. Both are important for many aspects of semiconductor laser dynamics. The relaxation oscillation frequency determines the maximum modulation frequency of diode lasers. This frequency is strongly governed by the differential gain in the active material, i.e., by the variation of the optical gain with carrier density.

An important application of relaxation oscillations is short pulse generation by gain switching of laser diodes. The idea of gain switching is to inject a short and intense current pulse or an optical pump pulse into the laser structure. This pulse initiates relaxation oscillations but it is so short that already the second oscillation maximum is not amplified. As a consequence, only one intense pulse is emitted. The shortest pulse widths achieved with this scheme so far are in the few picoseconds range. For short pulse generation, a further complexity of the diode laser comes into play. The emitted pulses are observed to be strongly chirped, i.e., the center frequency changes during the pulse. This complex self-phase-modulation is a result of the strong coupling of the real and the imaginary parts of the susceptibility in the semiconductor. If the gain (which is basically given by the imaginary part of the susceptibility) changes during a pulse emission, the refractive index (basically the real part of the susceptibilty) also changes considerably. This leads to a self-phase-modulation if a strong optical pulse is emitted. This strong phase–amplitude coupling is also responsible for the high feedback sensitivity of diode lasers, for complex spatio-temporal dynamics (e.g., self-focusing, filamentation), and for a considerable broadening of the emission linewidth. The phase–amplitude coupling is often quantified with the so-called linewidth enhancement or α factor. This parameter, however, is a function of carrier density, photon wavelength, and temperature.

Diode lasers are also suited for the generation or amplification of extremely short pulses with durations below 1 picosecond. The concept used for such extremely short pulse generation is called mode-locking. The idea of mode-locking is to synchronize the longitudinal modes of the laser so that their superposition is a sequence of short pulses. The most successful concept to achieve mode-locking of diode lasers is to introduce a saturable absorber into the laser cavity and to additionally modulate the injection current synchronously to the cavity roundtrip frequency. A saturable absorber exhibits nonlinear absorption characteristics: the absorption is high for weak intensities and weak for high intensities so that high peak power pulses are supported. Such absorbers can be integrated as a separate section into multiple-section diode lasers which have already been successfully used for subpicosecond pulse generation. Furthermore, electro-absorption modulators, passive waveguide sections, and tunable DBR segments for wavelength tuning can also be integrated into such multiple-section devices.

However, when subpicosecond pulses are generated or amplified in semiconductor laser amplifiers the dynamics of the pulse generation or pulse amplification is strongly governed by complex nonequilibrium carrier dynamics in the semiconductor. The most prominent effects are spectral hole burning and carrier heating. Spectral hole burning occurs if an intense laser field removes carriers from a certain area in *k*-space faster than carrier–carrier scattering can compensate for. This leads to nonthermal carrier distributions and a gain suppression in certain energy regimes. As mentioned above, a nonthermal carrier distribution relaxes quickly (100 femtoseconds) into a thermal (quasi-equilibrium) distribution if the intense laser field is switched off, but this carrier distribution may have a plasma temperature much higher than the lattice temperature. This effect, which is referred to as carrier heating, also leads to a transient reduction of the gain on a few picoseconds timescale until the carrier distributions have cooled down to the lattice temperature by interactions with phonons. These ultrafast effects are of course particularly important for amplification and generation of short pulses but they generally influence the whole dynamical behavior of diode lasers including, e.g., the modulation behavior.

See also

Coherent Transients: Ultrafast Studies of Semiconductors. **Quantum Optics:** Entanglement and Quantum Information.

Further Reading

Chow WW and Koch SW (1999) *Semiconductor-Laser Fundamentals: Physics of the Gain Materials*. Berlin: Springer-Verlag.

Chuang SL (1995) *Physics of Optoelectronic Devices*. New York: Wiley.

Delfyett PJ, Dienes A, Heritage JP, Hong MY and Chang YH (1994) Femtosecond hybrid mode-locked semiconductor laser and amplifier dynamics. *Applied Physics* B 58: 183–195.

Haug H and Koch SW (2004) *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 4th edn. Singapore: World Scientific Publisher.

Hofmann M (1999) Gain and emission dynamics of semiconductor lasers. *Recent Research Developments, in Applied Physics* 2: 269–290.

Koch SW, Jahnke F and Chow WW (1995) Physics of semiconductor microcavity lasers, review article. *Semiconductor Science and Technology* 10: 739–751.

Osinski M and Buus J (1987) Linewidth broadening factor in semiconductor lasers – an overview. *IEEE Journal of Quantum Electronics* QE-23: 9–29.

Up-Conversion Lasers

A Brenier, University of Lyon, Villeurbanne, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Under photoluminescence excitation, a luminescent center located inside a transparent material usually emits at a longer wavelength than the excitation. This is because of the existence of de-excitation losses inside the material. Thus, efficient down-conversion lasers can be built, for example, the commercial Nd^{3+}:$Y_3Al_5O_{12}$ laser which works at 1064 nm, under usual pumping near 800 nm. More surprisingly, it has been demonstrated that fluorescence can also be emitted at shorter wavelength than the excitation. This is possible if the losses are reduced and overcome by so-called 'up conversion mechanisms'. Generally speaking, a laser emitting at a shorter wavelength than the excitation is classified as an up-conversion laser.

These lasers are able to generate coherent light in the red, green, blue, and uv spectral ranges. A point of practical importance is that some of them can be pumped in the near-infrared range (around 800 nm or 980 nm) by commercially available laser diodes. They can be based on all-solid-state technology and share its advantages: compactness of devices, low maintenance, and efficiency. They are needed for a variety of applications: color displays, high density optical data storage, laser printing, biotechnology, submarine communications, gas-laser replacement, etc.

Because the emitted photons have higher energy than the excitation ones, more than one pump photon is needed to generate a single laser photon, so the physical processes involved in the up-conversion laser are essentially nonlinear. Two kinds of laser devices can be considered. In the first one, the excitation of the luminescent center up-migrates through the electronic levels and reaches the initial laser level. This step is due to some luminescence mechanisms. Then, in a second step, lasing at a short wavelength occurs through a transition towards the final laser level. In the second kind of device, the pump excitation activates a down-conversion lasing, but

the laser wave is maintained inside the cavity, cannot escape, and is up-converted in short wavelengths by the second-order nonlinear susceptibility of the crystal host. This requires bifunctional nonlinear optical and laser crystals. The device is thus a self-frequency doubling laser or a self-sum frequency mixing laser. If the nonlinearity is not an intrinsic property of the laser crystal but is due to a second nonlinear optical crystal located inside or outside the cavity, the device operates intracavity or extracavity up conversion.

Up-Conversion from Luminescence Mechanisms

The numerous $^{2S+1}L_J$ energy levels ($4f^n$ configuration) of the trivalent rare earth ions, inserted in crystals or glasses, offer many possibilities for up conversion, particularly from long-lived intermediate levels that can be populated with infrared pumping. Crystal field induced transitions between them are comprised of sharp lines, because the $4f^n$ electrons are protected from external electric fields by the $5s^2p^6$ electrons. Their fluorescence spectra can exhibit cross-sections high enough (10^{-19} cm^2) to lead to laser operation in the visible range. For up-conversion purposes, the most popular ions are Er^{3+}, Pr^{3+}, Tm^{3+}, Ho^{3+}, and Nd^{3+}. Up conversion in Er^{3+} was used for the first time in 1959, to detect infrared radiation.

Energy Transfers

Two ions in close proximity interact electrostatically (Coulomb interaction) and can exchange energy. The ion which gives energy is called the sensitizer S or donor, and the energy receiving ion is the activator A or acceptor. Two examples discussed below are shown in Figure 1. The most typical case of such nonradiative energy transfer in trivalent rare earth ions in insulating materials is due to electrical dipole–dipole interaction. For the latter, the transition probability $W(\mathrm{s}^{-1})$ takes the form:

$$W = \frac{3\hbar^4 c^4}{4\pi n^4}\frac{1}{R^6}\frac{Q_A}{\tau_{rad}}\int \frac{f_A(E)f_S(E)}{E^4}\mathrm{d}E \qquad [1]$$

where R is the S–A distance, E is the photon energy of the transition operated by each ion, n the refractive index, and Q_A is the integrated absorption cross-section of the A-transition:

$$Q_A = \int \sigma_A(E)\mathrm{d}E, \qquad f_A = \frac{\sigma_A}{Q_A} \qquad [2]$$

In eqn [1], τ_{rad} is the radiative emission probability of S for the involved transition calculated from the emission probability A_S:

$$\frac{1}{\tau_{rad}} = \int A_S(E)\mathrm{d}E, \qquad f_S = \tau_{rad}A_S \qquad [3]$$

The integral in eqn [1] is called the overlap integral and shows that the transfer is efficient if the luminescence spectrum of S is resonant with the absorption spectrum of A (corresponding to the S and A transitions involved in the nonradiative transfer). The transitions are often not in perfect resonance but remain efficient enough to be used in practice. Then the energy mismatch is compensated by phonons, and the energy transfer is said to be phonon-assisted.

The energy transfer visualized in Figure 1a is an up-conversion cross-relaxation. It is often met in the rare earth ions cited above when they are sufficiently concentrated in the material. The S and A ions can be of the same or of different chemical species. Successive energy transfers can promote the activator from its ground state towards higher energy levels, up to the one from which lasing can occur. This is the so-called 'Addition de Photons par Transferts d'Energie', sometimes appointed 'APTE'. Yb^{3+} is the most used ion as sensitizer and leads to the following transfers in examples of Er^{3+} or Ho^{3+} co-doping:

$$\begin{aligned} {}^2F_{5/2}(Yb) + {}^4I_{15/2}(Er) &\rightarrow {}^2F_{7/2}(Yb) + {}^4I_{11/2}(Er) \\ {}^2F_{5/2}(Yb) + {}^4I_{11/2}(Er) &\rightarrow {}^2F_{7/2}(Yb) + {}^4F_{7/2}(Er) \end{aligned} \qquad [4]$$

$$\begin{aligned} {}^2F_{5/2}(Yb) + {}^5I_8(Ho) &\rightarrow {}^2F_{7/2}(Yb) + {}^5I_6(Ho) \\ {}^2F_{5/2}(Yb) + {}^5I_6(Ho) &\rightarrow {}^2F_{7/2}(Yb) + {}^5S_2(Ho) \end{aligned} \qquad [5]$$

If several sensitizers give their energy simultaneously to a single activator, promoting it directly from its ground state towards a high energy level without intermediate levels, the sensitization is said to be 'cooperative'.

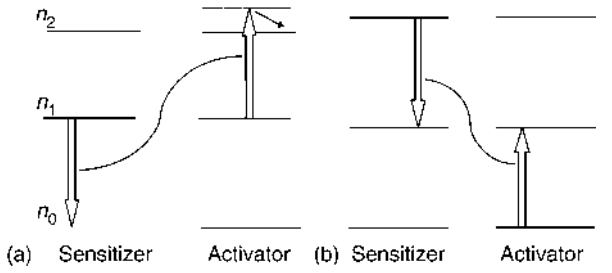

Figure 1 (a) Up-conversion cross-relaxation energy transfer. (b) Down-conversion cross-relaxation energy transfer. The example in (a) is phonon assisted, the example in (b) is resonant.

The energy transfer shown in Figure 1b is a down-conversion cross-relaxation. It is the inverse of the previous one. At first sight it is, of course, undesirable in an up-conversion laser. However, we notice that one ion in the upper level will lead to two ions in the intermediate level. This fact can be exploited in special conditions (see below) and be beneficial for up conversion.

The dynamical study of the energy transfers, starting from the microscopic point of view contained in eqn [1], is very complicated because of the random locations of the ions or because of other processes, such as energy migration among sensitizers and back transfers. A popular method leading to useful predictions in practice is the rate equation analysis dealing with average parameters. The time evolution of the population densities n_i of the various levels i are described by first-order coupled equations, solved with adequate initial conditions and including the pumping rate. As an example, the rate equations for levels 1 and 2 in Figure 1a take the form:

$$\frac{dn_1}{dt} = -\frac{n_1}{\tau_1} - 2kn_1^2 + \frac{\beta_{21}}{\tau_2}n_2 + W \qquad [6]$$

$$\frac{dn_2}{dt} = -\frac{n_2}{\tau_2} + kn_1^2 \qquad [7]$$

$$n_0 + n_1 + n_2 = 1 \qquad [8]$$

where $\tau_{1,2}$ are the lifetimes of levels 1 and 2, k is the up-conversion transfer rate, β_{21} is the branching ratio for the $2 \rightarrow 1$ transition, and W is the pump rate (not shown in Figure 1).

Sequential Two-Photon Absorption

Pumping the ground state of a luminescent ion (Figure 2a) with a wave at angular frequency ω, leads to population with density n_1 of an excited state in a first step. Because the rare earth ions have numerous energy levels, it is possible for the pump to also be resonant with the transition from level 1 towards a higher energy level 2, and is further absorbed. Such a process leading to n_2 up-conversion population is a sequential two-photon absorption. If there is no $1 \rightarrow 2$ resonant transition for the ω frequency, it is possible to use a second pump wave at a different angular frequency ω' resonant with the $1 \rightarrow 2$ transition (Figure 2b). Of course, in practice, it is advantageous when a single pump beam is required for up conversion.

Let us show the relevant terms of the rate equation analysis describing the process in Figure 2a:

$$\frac{dn_1}{dt} = -\frac{n_1}{\tau_1} + I\sigma_0 n_0 - I\sigma_1 n_1 + \frac{\beta_{21}}{\tau_2}n_2 \qquad [9]$$

$$\frac{dn_2}{dt} = -\frac{n_2}{\tau_2} + I\sigma_1 n_1 \qquad [10]$$

$$n_0 + n_1 + n_2 = 1 \qquad [11]$$

where I is the photon density of the pump (photons $s^{-1}\,cm^{-2}$) and σ_0 and σ_1 are the absorption cross-sections (cm^2) of the $0 \rightarrow 1$ and $1 \rightarrow 2$ transitions, respectively. The n_2 population density shows a nonlinear I dependence, most often close to I^2.

Photon Avalanche

In rare earth doped materials, an up-conversion emission, intense enough for visible lasing, can result from a more complex mechanism. In this case, absorption of the pump is not resonant with a transition from the ground state (weak ground-state absorption), but on the contrary, the photon pump energy matches the gap between two excited states (high excited state absorption). These levels are labeled 1 and 2 in Figure 3. Level 1 generally has a rather long lifetime in order to accumulate

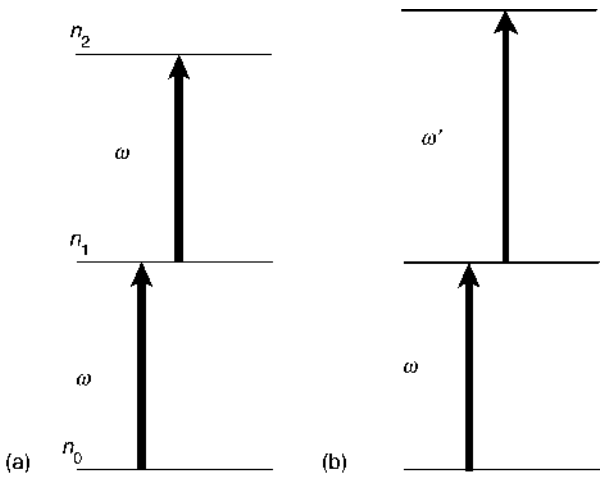

Figure 2 Two-photon absorption up-conversion.

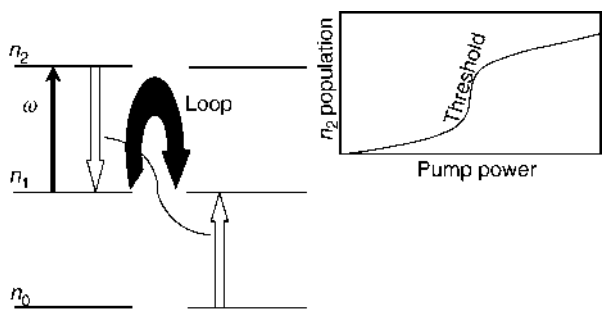

Figure 3 Photon avalanche up-conversion.

population and level 2 is the initial up-conversion laser level. When the pump light is turned on, the populations of levels 1 and 2 first grow slowly, and then can accelerate exponentially. This is generally referred to as 'photon avalanche'. The term 'looping mechanism' also applies, because we can see in Figure 3 that the excited state absorption $1 \rightarrow 2$ is followed by a cross-relaxation energy transfer (quantum efficiency up to two) that returns the system back to level 1. So after one loop, the n_1 population density has been amplified, and is further increased after successive loops. This pumping scheme was discovered in 1979, with $LaCl_3:Pr^{3+}$ crystal used as an infrared photon counter.

The rate equation analysis applied to the three-level system in Figure 3 leads to the following time evolutions:

$$\frac{dn_1}{dt} = -\frac{n_1}{\tau_1} + I\sigma_0 n_0 - I\sigma_1 n_1 + \frac{\beta_{21}}{\tau_2} n_2 + 2kn_0 n_2 \qquad [12]$$

$$\frac{dn_2}{dt} = -\frac{n_2}{\tau_2} + I\sigma_1 n_1 - kn_0 n_2 \qquad [13]$$

$$n_0 + n_1 + n_2 = 1 \qquad [14]$$

where the parameters have the same meaning as in eqns [6]–[11].

The steady state limit of the populations at infinite time, upon continuous excitation starting at $t = 0$, can be obtained from canceling the time derivative in the left-hand side of the equation. Let us point out the interesting result. Two different dynamical regimes and analytical expressions for $n_2(\infty)$ are obtained. In the usual case of low pumping rate in the ground state and if the parameters of the systems satisfy: $k < \frac{1-\beta_{21}}{\tau_2}$, then the first regime is observed, whatever the pumping rate $I\sigma_1$ in the excited state.

If the parameters satisfy:

$$k > \frac{1 - \beta_{21}}{\tau_2} \qquad [15]$$

the first regime is still observed at low pumping rate $I\sigma_1$ in the excited state, more precisely at pumping rate low enough such that

$$I\sigma_1 < \frac{\left(k + \frac{1}{\tau_2}\right)\frac{1}{\tau_1}}{k - \frac{(1 - b_{21})}{\tau_2}}$$

But interestingly, the second regime, the so-called photon avalanche, is obtained at high pumping rate $I\sigma_1$ in the excited state, that is to say if:

$$I\sigma_1 > \frac{\left(k + \frac{1}{\tau_2}\right)\frac{1}{\tau_1}}{k - \frac{(1 - b_{21})}{\tau_2}} \qquad [16]$$

So the right-hand side of eqn [16] appears to be a pump threshold separating two dynamical regimes when the condition in eqn [15] is satisfied, as the insert in Figure 3 shows.

Another way to distinguish between the two regimes is given by linearization of the system of eqns [12]–[14] ($n_0 \cong 1$). Then the description will only be valid close to $t = 0$. We know from standard mathematical methods that the solutions of eqns [12]–[14] are then a combination of $\exp(\alpha t)$ terms. If the threshold condition in eqn [16] is satisfied, a parameter α in the combination becomes positive, so the population of level 2 increases exponentially (photon avalanche) at early times and the dynamics of the system is unstable.

The physical meaning of eqns [15] and [16] is clarified by introducing the yield η_{CR} of the cross-relaxation process:

$$\eta_{CR} = \frac{k}{k + 1/\tau_2} \qquad [17]$$

and the yield η of the $2 \rightarrow 1$ de-excitation:

$$\eta = \frac{b_{21}/\tau_2}{k + 1/\tau_2} \qquad [18]$$

Then, the threshold condition in eqn [16] can be written as:

$$\left(\frac{I\sigma_1}{I\sigma_1 + 1/\tau_1}\right)(2\eta_{CR} + \eta) > 1 \qquad [19]$$

Starting with one ion in level 1, the first parenthesis in eqn [19] is the yield with which it will be promoted to level 2, and the second parenthesis is the yield of the backward path $2 \rightarrow 1$. So, the left-hand side of eqn [19] is the new number of ions after a looping path $1 \rightarrow 2 \rightarrow 1$, which has to be higher than 1 for the avalanche to occur. If we consider a lot of successive loops, the origin of the exponential behavior is clear. The role of the first ion in level 1 (and the weak pumping rate in the ground state $I\sigma_0$) is limited to initialize the process. Note that the value of the first parenthesis in eqn [19] is less than 1, so the second parenthesis has to be higher than 1. This condition is equivalent to eqn [15].

Materials: The Energy Gap Law

As we can see from the above mechanisms, up conversion needs intermediate energy levels and is not efficient if these cannot accumulate populations due to a short lifetime. The spontaneous rate of de-excitation of an electronic level is the sum of a radiative part W^{rad} and of a nonradiative part W^{nrad}. The radiative part can be obtained through the absorption cross-section corresponding to the transition or from the Judd–Ofelt theory which also exploits the absorption spectrum. The nonradiative rate is then obtained by subtracting W^{rad} from the measured fluorescence decay rate of the level. Based on the measurements in many hosts for the different trivalent rare earth ions, at different temperatures, the nonradiative decay rate of a J-level towards the next lower J′-level was found to have the form:

$$W_{\mathrm{JJ'}}^{\mathrm{nrad}} = B\exp(-\beta\Delta E_{\mathrm{JJ'}})\left(1-\exp\left(-\frac{\omega_{\max}}{kT}\right)\right)^{-p} \quad [20]$$

where $\Delta E_{\mathrm{JJ'}}$ is the energy gap between the two levels, $p = \Delta E_{\mathrm{JJ'}}/\omega_{\max}$, $\omega_{\max}$ being the energy of an effective optical phonon of the host. The values of B, β and $\omega_{\max}$ are found by fitting with experimental data. An example of such a fit is represented in Figure 4.

Equation [20] is the well-known and familiar 'energy gap law'. We have represented it for several oxide and nonoxide hosts at $T = 300$ K in Figure 5. It is clear that chloride, bromide, fluoride will be favorable hosts for up conversion due to their low phonon energy ($LaBr_3$: $\omega_{\max} = 175\ \mathrm{cm}^{-1}$, $LiYF_4$: $\omega_{\max} = 400\ \mathrm{cm}^{-1}$, $Y_3Al_5O_{12}$: $\omega_{\max} = 700\ \mathrm{cm}^{-1}$, silicate glass: $\omega_{\max} = 1100\ \mathrm{cm}^{-1}$). Let us mention also a fluorozirconate glass, ZBLAN, widely used in up-conversion fiber lasers.

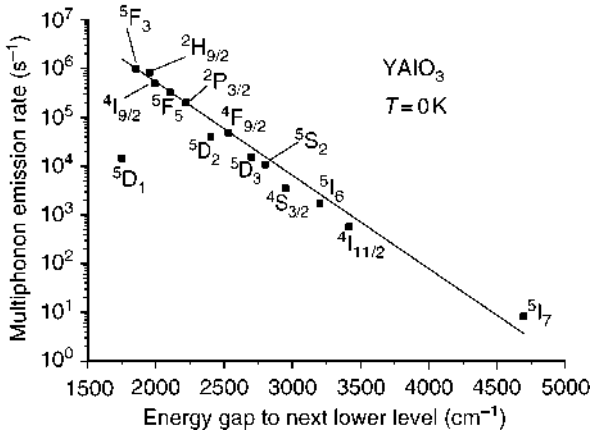

Figure 4 Energy-gap dependence of nonradiative decay rates in $YAlO_3$.

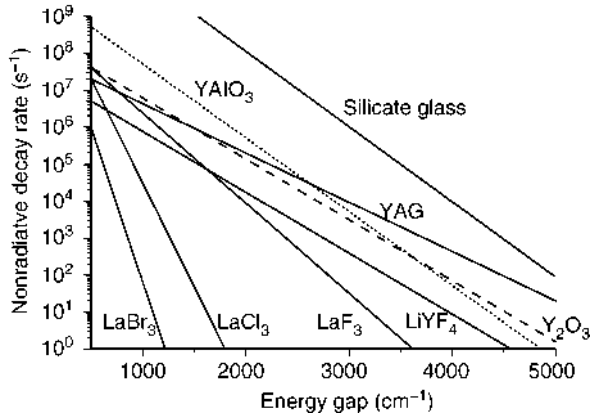

Figure 5 Energy-gap law for different hosts at 300 K.

Up-Conversion from Second-Order Optical Nonlinearity

The inter-atomic electric field acting on the electrons inside a medium has a magnitude of $\sim 10^8$ V/cm and derives from a potential that is anharmonic. The electric field **E** of an electromagnetic wave propagating through the medium drives the electrons beyond the quadratic region of the potential in the case of high **E**. So, the electron response and the associated polarization **P** take the form of a function of **E**:

$$\mathbf{P}(\mathbf{E}) = \mathscr{E}_0(\chi^{(1)} : \mathbf{E} + \chi^{(2)} : \mathbf{E}^2 + \chi^{(3)} : \mathbf{E}^3 + \cdots) \quad [21]$$

The different terms in eqn [21] are responsible for many optical phenomena because in Maxwell electromagnetic theory, the polarization is the source of the waves. We have selected in Table 1 the terms relevant for linear effects and for the purpose of frequency up conversions: second-harmonic generation (SHG) and sum-frequency mixing, due to second-order terms. It should be noticed that the latter have a nonzero value only in noncentrosymmetric materials.

Figure 6 visualizes the up conversion from SHG: two photons of the fundamental field at angular frequency ω_1 are annihilated while one photon at twice the angular frequency is created. A similar picture could be drawn for sum-frequency up conversion: SHG is the degenerate case where $\omega_1 = \omega_2$.

The second-order nonlinear processes have an efficiency given by an effective coefficient d_{eff} given hereafter:

$$d_{\mathrm{eff}} = \sum_{ijk} e_i(\omega_3) d_{ijk} e_j(\omega_1) e_k(\omega_2) \quad [22]$$

where $e_i(\omega_l)$ is the ith component of the unit vector of the electric field of the wave l at angular frequency ω_l and $d_{ijk} = (1/2)\chi_{ijk}$.

Table 1 Relevant terms for frequency up-conversion

Optical process	*Term*	*New wave*
Linear		
Refractive index, absorption, stimulated emission	$\chi^{(1)}(-\omega;\omega)$	
Second order nonlinear		
Second harmonic generation	$\chi^{(2)}(-\omega_3;\omega_1,\omega_1)$	$\omega_3 = 2\omega_1$
Sum-frequency mixing	$\chi^{(2)}(-\omega_3;\omega_1, \pm\omega_2)$	$\omega_3 = \omega_1 + \omega_2$

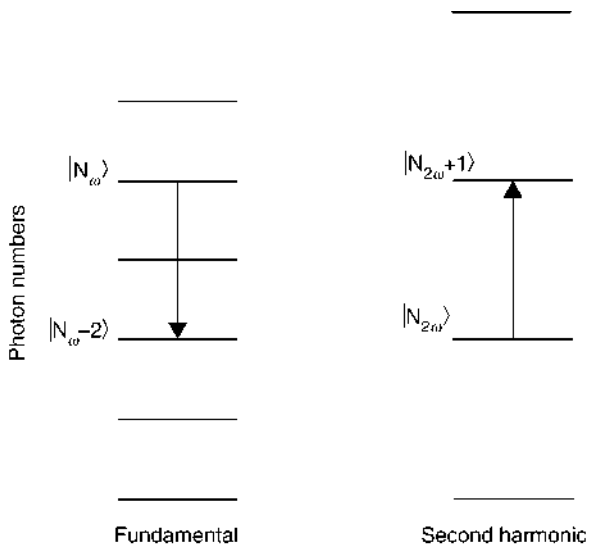

Figure 6 Up-conversion from second harmonic generation.

Moreover, the wave propagation equations in the slow varying envelope approximation, given hereafter for the E_3 electric field of the sum-frequency mixing wave:

$$\frac{\mathrm{d}E_3}{\mathrm{d}z} = \frac{\mathrm{i}\omega_3 d_{\mathrm{eff}}}{n_3 c} E_1 E_2 \quad \exp(\mathrm{i}(k_1 + k_2 - k_3)z) \qquad [23]$$

$$(\text{with } k_j = \omega_j n_j / c)$$

impose that the frequency conversion is efficient in practice, only if the following phase matching condition is satisfied:

$$\omega_1 n_1 \uparrow (\theta, \varphi) + \omega_2 n_2 \uparrow (\theta, \varphi) = \omega_3 n_3 \downarrow (\theta, \varphi) \qquad [24]$$

where $n_i \uparrow$ and $n_i \downarrow$ are respectively the upper and lower refractive index in the direction of propagation (θ, φ) (n_i in eqn [23] are the same with a simplified notation). Equation [24] is restricted here for simplicity to collinear type I (waves 1 and 2 and the same polarization) and is the expression of photon momentum conservation. It is usually achieved from the birefringence of the crystals. When this is not possible, another technique can be used, namely quasi-phase matching, which involves reversing periodically the sign of the nonlinear optical coefficient.

In bifunctional crystals, the laser effect and the $\chi^{(2)}$ interaction occur simultaneously inside the same crystal. In the case of the self-frequency doubling laser, the angular frequency $\omega_1 = \omega_2$ in eqn [24] is that of the fundamental laser wave. In the case of the self-sum frequency mixing laser, ω_1 corresponds to the fundamental laser wave and ω_2 to the pump wave. As we can see from eqns [22]–[24], the polarization of the laser emission is crucial in order for the device to work. So, the laser stimulated emission cross-section σ for propagation in the phase matching direction (θ, ϕ) in the adequate polarization has to be evaluated. This can be performed with the formula:

$$\sigma(\theta, \varphi) = \frac{\sigma_X \sigma_Y \sigma_Z}{\sqrt{\sigma_Y^2 \sigma_Z^2 e_X^2 + \sigma_X^2 \sigma_Z^2 e_Y^2 + \sigma_X^2 \sigma_Y^2 e_Z^2}} \qquad [25]$$

where $\sigma_{X,Y,Z}$ are the emission cross-sections for X, Y, Z-polarizations and $e_{X,Y,Z}$ are the components of the unit vector or the electric field of the laser wave.

To summarize, the conditions necessary for a crystal to be a bi-functional material are: (i) it must be noncentrosymmetric; (ii) it must accept fluorescent doping (usually with Nd^{3+} or Yb^{3+}); and (iii) it must be phase matchable for its laser emission. In practice, only a few crystals satisfy these requirements and have been tried with some success. The main ones are $LiNbO_3$, $LaBGeO_5$, $Ba_2NaNb_5O_{15}$, β'-$Gd_2(MoO_4)_3$, $YAl_3(BO_3)_4$, $GdAl_3(BO_3)_4$, $CaY_4(BO_3)_3O$, and $CaGd_4(BO_3)_3O$.

Up-Conversion Lasers Based on Luminescence Mechanisms in Materials

Energy Transfers

Up-conversion lasers from energy transfers have the advantage of using a single pump laser. This latter populates firstly a long-lived intermediate level. A typical example, based on $LiYF_4$:Er^{3+} crystal, is provided by Figure 7.

The pump at 802 nm from a GaAlAs diode laser, or at 969 nm, populates efficiently at the 8 ms lifetime $^4I_{11/2}$ level. The green laser, at 551 nm, corresponds to the $^4S_{3/2} \rightarrow {}^4I_{15/2}$ transition and it can be shown that energy transfer is an essential part of the mechanism by stopping abruptly the pump laser: green lasing continues to occur for several hundred microseconds because the $^4S_{3/2}$ level is still fed (an excited-state

absorption would stop immediately). The involved energy transfer is:

$$^4I_{11/2} + {}^4I_{11/2} \rightarrow {}^4F_{7/2} + {}^4I_{15/2} \qquad [26]$$

and leads to a 20 mW laser threshold and 2.3 mW green output power at 50 K temperature upon 95 mW of incident power.

By using mirrors with high transmission in the green range but having high reflectivity at blue wavelengths, blue laser emission at 470 nm can be sustained, corresponding to the $^2P_{3/2} \rightarrow {}^4I_{11/2}$ transition. It originates from a mechanism involving three up-conversion energy transfers:

$$^4I_{11/2} + {}^4I_{11/2} \rightarrow {}^4F_{7/2} + {}^4I_{15/2}$$
$$^4I_{11/2} + {}^4I_{13/2} \rightarrow {}^4F_{9/2} + {}^4I_{15/2} \qquad [27]$$
$$^4S_{3/2} + {}^4F_{9/2} \rightarrow {}^4I_{15/2} + {}^2K_{13/2}$$

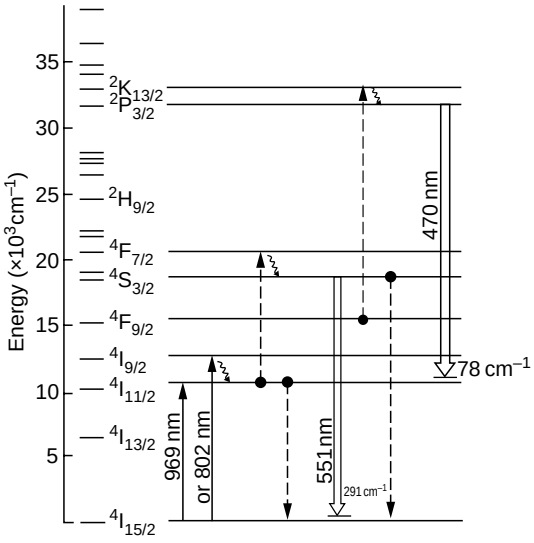

Figure 7 Energy level diagram of $LiYF_4:Er^{3+}$. The dashed lines indicate the energy transfers responsible for the green and blue up-conversion lasing. Reproduced with permission of American Institute of Physics from Macfarlane RM, Tong F, Silversmith AJ and Lenth W (1988) Violet CW neodymium upconversion lasers. *Applied Physics Letters* 52(16): 1300.

The illustration in Figure 7 shows up-conversion energy transfers between ions of the same chemical species, but co-doping the laser host with two ions of different chemical species provides the opportunity of separating their roles: one species is devoted to pump absorption (this is the sensitizer) and the other one is devoted to lasing (this is the activator). Due to the development of efficient semiconductor InGaAs diodes emitting near 980 nm, the most widely used sensitizer is Yb^{3+}. Table 2 provides a list of several up-conversion lasers based on energy transfer mechanisms, operating at room temperature. We can see that among the most efficient lasers, are those based on rare-earth ions doped glass fibers. The reason is that in fibers, the pump and laser waves remain confined inside a core (typically 5 μm diameter) and have high energy densities over a long length (tenths of cms), which is favorable for all up-conversion mechanisms.

Sequential Two-Photon Absorption

An example of an up-conversion laser pumped by a two-photon absorption mechanism is represented in Figure 8. The material is $LaF_3:Nd^{3+}$ and violet lasing at 380 nm corresponds to the $^4D_{3/2} \rightarrow {}^4I_{11/2}$ transition.

Transition from the ground state up to the $^4F_{5/2}$ level is firstly obtained from an infrared beam at 790 nm in order to feed the $^4F_{3/2}$ intermediate level after a fast nonradiative de-excitation. The $^4F_{3/2}$ level has a rather long lifetime: 700 μs, so it accumulates

Table 2 Main room temperature up-conversion lasers operating from energy transfers

Laser material	*Laser wavelength (nm)*	*Pump wavelength (nm)*	*Output power*
$BaY_{1.4}Yb_{0.59}Ho_{0.01}F_8$	670	1540 + 1054	
$BaYYb_{0.998}Tm_{0.002}F_8$	649	1054	1%
$BaYYb_{0.99}Tm_{0.01}F_8$	799–649–510–455	960	
$BaYYb_{0.99}Tm_{0.01}F_8$	348	960	
$LiYF_4$:Er(1%):Yb(3%)	551	966	37 mW
$LiY_{0.89}Yb_{0.1}Tm_{0.01}F_4$	810–792	969	80 mW
$LiY_{0.89}Yb_{0.1}Tm_{0.01}F_4$	650	969	5 mW
$LiKYF_5$:Er (1%)	550	808	150 mW
Fiber:Yb:Pr	635	849	20 mW
Fiber:Yb:Pr	635	1016	6.2 mW
Fiber:Yb:Pr	521	833	0.7 mW
Fiber:Yb:Pr	635	860	4 mW

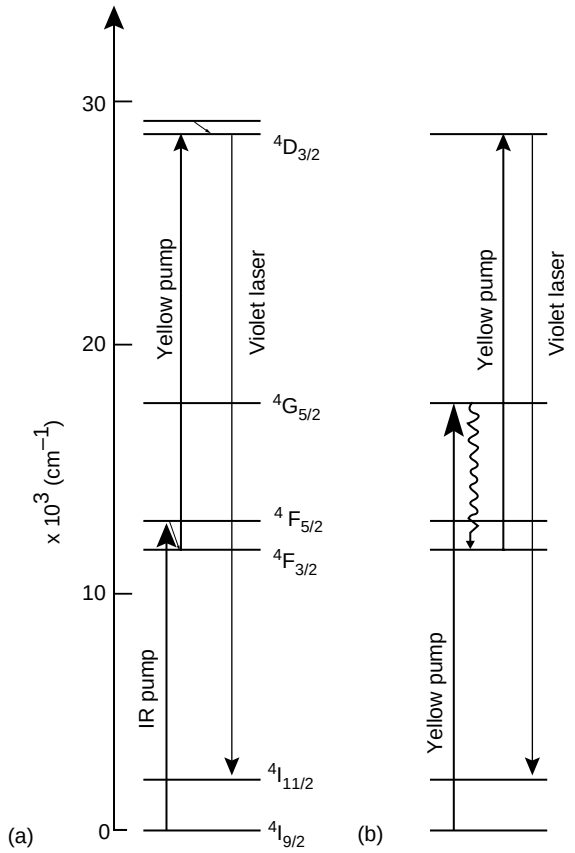

Figure 8 Simplified energy level diagram of LaF_3:Nd^{3+} and up-conversion lasing from two-photon absorption. (a) Two different pump wavelengths, (b) doubly resonant single pump wavelength. Reproduced with permission of American Institute of Physics from Lenth W, Silversmith AJ and Macfarlane RM (1988) Green infrared pumped erbium up conversion lasers. In: TAM AC, Gole JL and Stwalley WC (eds) *Advances in Laser Science – III*. AIP Conference Proceedings n°172: 8–12.

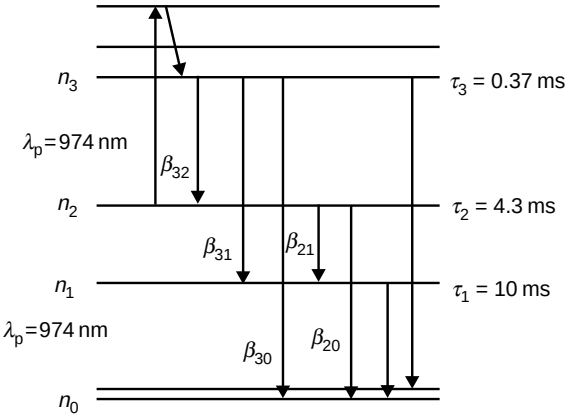

Figure 9 Simplified energy level diagram of $LiYF_4$:Er^{3+} and up-conversion lasing from two-photon absorption (doubly resonant). Reproduced with permission of Institute of Physics Publishing from Huber G (1999) Visible cw solid-state lasers. *Advances in Lasers and Applications*, Bristol and Philadelphia: Scottish Universities Summer School in Physics & Institute of Physics Publishing, 19.

enough population that a second yellow pumping at 591 nm (Figure 8a) can efficiently populate the $^4D_{3/2}$ initial laser level. With 110 mW of infrared pump and 300 mW of yellow pump, 12 mW of violet output were obtained at 20 K temperature.

A simpler device is obtained if a single-pump beam is used. This is the case in Figure 8b, where a yellow pump at 578 nm provides both ground state and excited state absorption. This doubly resonant scheme is less efficient in LaF_3:Nd^{3+} than the previous one but in Figure 9, we show another example: $LiYF_4$:Er^{3+}, working at room temperature. The lasing $^4S_{3/2} \rightarrow {}^4I_{15/2}$ transition at 551 nm delivers 20 mW upon 1600 mW pumping at 974 nm.

The simplified Er^{3+} level scheme of Figure 9 contains four main levels with population densities: $n_0 = [^4I_{15/2}]$, $n_1 = [^4I_{13/2}]$, $n_2 = [^4I_{11/2}]$, $n_3 = [^4S_{3/2}]$. The rate equations of populations of this system can

Table 3 Main room temperature up-conversion lasers operating from two-photon absorption

Laser material	*Laser wavelength (nm)*	*Pump wavelength (nm)*	*Output power*
$Y_3Al_5O_{12}$:Er 1%	561	647 + 810	
$LiYF_4$:Er 1%	551	647 + 810	0.95 mJ
$LiYF_4$:Er 1%	551	810	40 mW
$LiYF_4$:Er 1%	551	974	45 mW
$LiYF_4$:Er	551	974	20 mW
KYF_4:Er 1%	562	647 + 810	0.95 mJ
$LiLuF_4$:Er	552	974	70 mW
$LiLuF_4$:Er	552	970	213 mW
$LiYF_4$:Tm 1%	453–450	781 + 649	0.2 mJ
Fiber:Tm	480	1120	57 mW
Fiber:Tm	480	1100–1180	45 mW
Fiber:Tm	455	645 + 1064	3 mW
Fiber:Tm	803–816	1064	1.2 W
Fiber:Tm	482	1123	120 mW
Fiber:Yb:Tm	650	1120	
Fiber:Ho	540–553	643	38 mW
Fiber:Ho	544–549	645	20 mW
Fiber:Er	548	800	15 mW
Fiber:Er	546	801	23 mW
Fiber:Er	544	971	12 mW
Fiber:Er	543	800	
Fiber:Pr	635	1010 + 835	180 mW
Fiber:Pr	605	1010 + 835	30 mW
Fiber:Pr	635	1020 + 840	54 mW
Fiber:Pr	520	1020 + 840	20 mW
Fiber:Pr	491	1020 + 840	7 mW
Fiber:Nd	381	590	74 μW
Fiber:Nd	412	590	500 μW

be written and solved because all the implied parameters (lifetimes, branching ratios, ground, and excited states absorption cross-sections) have been measured in this material. Predictive models of the up-conversion green laser are successful, matching experimental measurements, and confirm that the doubly resonant mechanism is realistic at low doping concentration.

Table 3 summarizes the performances of the main up-conversion lasers based on sequential photon absorption and working at room temperature.

Photon Avalanche

Photon avalanche up-conversion was observed mainly in Nd^{3+}, Tm^{3+}, Er^{3+}, and Pr^{3+} doped materials. As an illustration of this mechanism, let us consider the case of the continuous wave Pr^{3+}–Yb^{3+}-doped fluorozirconate fiber laser (3000 ppm Pr, 20 000 ppm Yb). With two Ti:sapphire pump lasers operating at 852 and 826 nm (Figure 10), an output power at 635 nm as high as 1.02 W was obtained with an incident pump power of 5.51 W (19% slope efficiency).

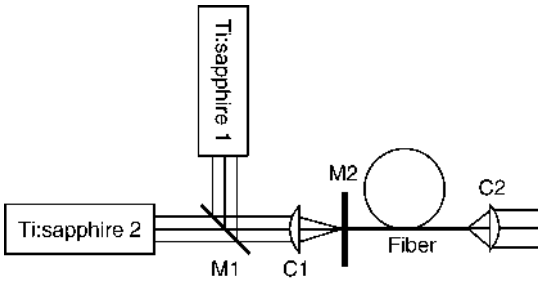

Figure 10 Experimental set-up for Yb–Pr-doped fiber up-conversion laser. M1: dichroic mirror, C1, C2: lenses, M2 dichroic input mirror. Reproduced with permission of Optical Society of America from Sandrock T, Scheife H, Heumann E and Huber G (1997) High power continuous wave up-conversion fiber laser at room temperature. *Optics Letters* 22(11): 809.

This remarkable result can be explained quantitatively by modeling the photon avalanche with a five-level diagram and restricted in Figure 11 to a single laser pump at 850 nm. The laser emission corresponds to the $^3P_0 \rightarrow {}^3F_2$ transition. The five relevant levels have the population densities: $n_0 = [^3H_4(Pr)]$, $n_1 = [^1G_4(Pr)]$, $n_2 = [^3P_0(Pr)]$, $n_3 = [^2F_{7/2}(Yb)]$, $n_4 = [^2F_{5/2}(Yb)]$.

The two components of the mechanism are:

(i) the excited state absorption of the pump by the spin allowed transition: $^1G_4 \rightarrow {}^1I_6$
(ii) the process leading to the doubling of the $^1G_4$ population, in this instance through two energy transfers:

$$\begin{aligned} &^1I_6(Pr) + {}^2F_{7/2}(Yb) \rightarrow {}^1G_4(Pr) + {}^2F_{5/2}(Yb) \\ &^2F_{5/2}(Yb) + {}^3H_4(Pr) \rightarrow {}^2F_{7/2}(Yb) + {}^1G_4(Pr) \end{aligned} \qquad [28]$$

The rate equation analysis predicts that the avalanche threshold occurs at about 1 W pump and the long fluorescence rise time is somewhat shortened at higher pump power, as observed experimentally.

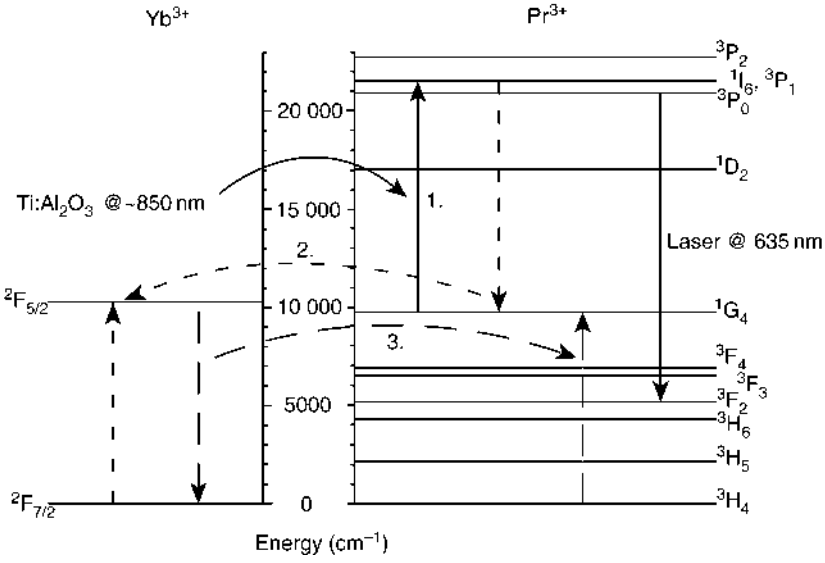

Figure 11 Energy level diagram of fluorozirconate: Yb^{3+}:Pr^{3+} fiber. The dashed lines indicate the energy transfers. Reproduced with permission of the Optical Society of America from Sandrock T, Scheife H, Heumann E and Huber G (1997) High power continuous wave up-conversion fiber laser at room temperature. *Optics Letters* 22(11): 809.

Table 4 Main room temperature up-conversion lasers operating from photon avalanche

Laser material	*Laser wavelength (nm)*	*Pump wavelength (nm)*	*Output power*
BaY_2F_8:Yb:Pr	607.5	822	55 mW
BaY_2F_8:Yb:Pr	638.7	841	26 mW
$LiY_{0.89}Yb_{0.1}Pr_{0.01}F_4$	720	830	1%
$LiY_{0.89}Yb_{0.1}Pr_{0.01}F_4$	639.5	830	
Fiber:Yb:Pr	635–637	780–880	300 mW
Fiber:Yb:Pr	605–622	780–880	44 mW
Fiber:Yb:Pr	517–540	780–820	20 mW
Fiber:Yb:Pr	491–493	780–880	4 mW
Fiber:Yb:Pr	635	850	1.02 W
Fiber:Yb:Pr	635	850	2.06 W
Fiber:Yb:Pr	520	850	0.3 W

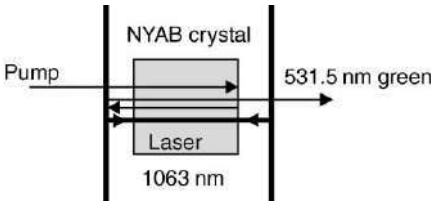

Figure 12 Scheme of a self-frequency doubling laser. Note that in reality the three beams are superimposed.

Table 4 summarizes the performances of the main up-conversion lasers based on photon avalanche mechanisms.

Up-Conversion Lasers Based on Bi-Functional Crystals

The laser stimulated emission in a bi-functional crystal was first obtained in 1969 with Tm^{3+} in $LiNbO_3$ but the most used ion is, of course, Nd^{3+} working in the two channels:

$$^4F_{3/2} \rightarrow {}^4I_{11/2} \qquad [29]$$

$$^4F_{3/2} \rightarrow {}^4I_{13/2} \qquad [30]$$

and more recently Yb^{3+} working in the channel:

$$^2F_{5/2} \rightarrow {}^2F_{7/2} \qquad [31]$$

Channels in eqns [29] and [31] operate around 1060 nm wavelength and the channel in eqn [30] near 1338 nm.

The Self-Frequency Doubling Laser

A typical laser scheme is shown in Figure 12.

Table 5 Main results of Nd^{3+} and Yb^{3+} green self-frequency doubling lasers near 530 nm (eqns [29] and [31])

Crystal	*Input power (mW)*	*Output power (mW)*	*Pumping*
$LiNbO_3$:MgO:Nd	215	1	Dye laser
$YAl_3(BO_3)_4$:Nd	870	10	Laser diode
$YAl_3(BO_3)_4$:Nd	280	3	Laser diode
$YAl_3(BO_3)_4$:Nd	400	69	Laser diode
$YAl_3(BO_3)_4$:Nd	1380	51	Laser diode
$YAl_3(BO_3)_4$:Nd	369	35	Laser diode
$LiNbO_3$:MgO:Nd	850	18	Dye laser
$LiNbO_3$:Sc_2O_3:Nd	65	0.14	Ti:sapphire laser
$LiNbO_3$:MgO:Nd	100	0.2	Laser diode
$YAl_3(BO_3)_4$:Nd	1600	225	Laser diode
$YAl_3(BO_3)_4$:Nd	2200	450	Ti:sapphire laser
(Y,Lu)$Al_3(BO_3)_4$:Nd	880	24	Laser diode
$LiNbO_3$:ZnO:Nd	430	0.65	Ti:sapphire laser
$Ba_2NaNb_5O_{15}$:Nd	270	46	Ti:sapphire laser
$CaY_4(BO_3)_3O$:Nd	900	62	Laser diode
$CaGd_4(BO_3)_3O$:Nd	1600	192	Ti:sapphire laser
$CaGd_4(BO_3)_3O$:Nd	1250	115	Laser diode
$CaGd_4(BO_3)_3O$:Nd	1560	225	Ti:sapphire laser
$YAl_3(BO_3)_4$:Yb	11 000	1100	Laser diode
$GdAl_3(BO_3)_4$:Nd	2.8 mJ/pulse	0.12 mJ/pulse	Pulsed dye laser

The laser beam cannot escape from the cavity and dichroic mirrors are used. The input mirror has high transmission at the pump wavelength and is highly reflective at laser- and second-harmonic wavelengths. The output mirror is highly reflective at laser wavelength and has high transmission for second harmonic. Channels in eqns [29] and [31] generate green light near 530 nm and the channel in eqn [30] generates red light near 669 nm.

The most exploited bi-functional crystal is $YAl_3(BO_3)_4$:Nd^{3+}, well-known as NYAB. It is a negative uniaxial trigonal crystal (extraordinary index lower than the ordinary one) and its laser emission is easily observed in ordinary polarization with a high cross-section: $2 \times 10^{-19}\,cm^2$ at 1063 nm. Type I phase matching occurs at a polar angle $\theta = 30.7°$ and 26.8° for channels in eqns [29] and [30] respectively. The effective nonlinear optical coefficient d_{eff} is close to 1.4 pm V^{-1} at azimuthal angle $\phi = 0°$.

The Yb^{3+} ion has some advantages over the Nd^{3+} ion due to its very simple energy level scheme: there is no excited state absorption at the laser wavelength, no up-conversion losses, no concentration quenching, and no absorption in the green. The small Stokes shift between pump and laser emission reduces the thermal loading of the material during laser operation. A self-doubling laser based on $YAl_3(BO_3)_4$:Yb^{3+} of crystal has produced 1.1 W green power upon 11 W diode pumping.

The drawback of $YAl_3(BO_3)_4$ is that it is rather difficult to grow because it is not congruent. This difficulty was overcome by the discovery at Ecole Nationale Supérieure de Chimie de Paris of the rare-earth calcium oxoborate $CaGd_4(BO_3)_3O$ which can be grown to a large size by the Czokhralski method. It is a monoclinic biaxial crystal and doped with Nd^{3+}, it was firstly exploited (channel in eqn [29]) in the XY principal plane at $\theta = 90°$, $\phi = 46°$ with ($d_{eff} = 0.5$ pm V^{-1}). It was soon recognized that the optimum phase matching direction ($d_{eff} = 1.68$ pm $\times$ V^{-1}) occurred out of the principal planes, in the direction $\theta = 66.8°$, $\phi = 132.6°$, and high green power can be obtained: 225 mW under 1.56 W pump.

Table 5 summarizes the main self-frequency doubling results obtained with the channel in eqn [29].

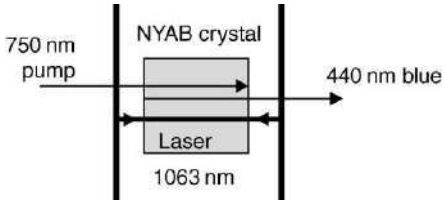

Figure 13 Scheme of a self-sum frequency mixing laser. Note that in reality the three beams are superimposed.

The Self-Sum Frequency Mixing Laser

The example in Figure 13 gives the main features of such a laser. The input mirror has high transmission at the pump wavelength and both mirrors are highly reflective at laser wavelength. The output mirror has high transmission at sum frequency mixing wavelength.

The laser wave at angular frequency ω_1 in Table 1 corresponds to the channel in eqns [29] or [30] in the case of Nd^{3+}. The wave at angular frequency ω_2 in Table 1 has two different roles: first it excites the Nd^{3+} laser center and secondly its nonabsorbed part is up converted by the second-order nonlinear process. ω_2 (corresponding to the wavelength λ_2) is then chosen to match the main Nd^{3+} absorption lines:

$$\begin{aligned} &{}^4I_{9/2} \rightarrow {}^4G_{5/2} - {}^2G_{7/2}\ (\lambda_2 = 590\ \text{nm}) \\ &{}^4I_{9/2} \rightarrow {}^4F_{7/2} - {}^4S_{3/2}\ (\lambda_2 = 750\ \text{nm}) \\ &{}^4I_{9/2} \rightarrow {}^4F_{5/2} - {}^2H_{9/2}\ (\lambda = 800\ \text{nm}) \\ &{}^4I_{9/2} \rightarrow {}^4F_{3/2}\ (\lambda_2 = 880\ \text{nm}) \end{aligned} \qquad [32]$$

The most efficient self-frequency mixing lasers based on channels in eqns [29]–[31] are gathered in Table 6.

See also

Materials Characterization Techniques: $\chi^{(2)}$. **Nonlinear Optics, Applications:** Phase Matching. **Nonlinear Optics, Basics:** $\chi^{(2)}$–Harmonic Generation.

Table 6 Main results of Nd^{3+} self-sum frequency mixing lasers based on eqn [29]

Crystal	*Pump wavelength (nm)*	*Generated wavelength (nm)*	*Output power*
$YAl_3(BO_3)_4$:Nd	740–760	436–443	0.16 mJ/pulse
$GdAl_3(BO_3)_4$:Nd	740–760	436–443	0.403 mJ/pulse
$CaGd_4(BO_3)_3O$:Nd	811	465	1 mW
$YAl_3(BO_3)_4$:Nd	585–600	377–383	0.25 mJ/pulse
$GdAl_3(BO_3)_4$:Nd	587–597	378–382	0.105 mJ/pulse
$YAl_3(BO_3)_4$:Nd	488, 515	330, 380	0.2 mW

Further Reading

Brenier A (2000) The self-doubling and summing lasers: overview and modeling. *Journal of Luminescence* 91: 121–132.

Butcher PN and Cotter D (1990) *The Elements of Nonlinear Optics*. Cambridge, UK: Cambridge University Press.

Huber G (1999) Visible cw solid-state lasers. *Advances in Lasers and Applications*. Bristol and Philadelphia: Scottish Universities Summer School in Physics & Institute of Physics Publishing.

Joubert MF (1999) Photon avalanche up-conversion in rare earth laser materials. *Optical Materials* 11: 181–203.

Kaminskii AA (1996) *Crystalline Lasers: Physical Processes and Operating Schemes*. CRC Press, Inc.

Lenth W and Macfarlane RM (1992) Upconversion lasers. *Optics & Photonics News* 3: 8–15.

Powell RC (1998) *Physics of Solid-State Laser Materials*. New York: Springer-Verlag.

Risk WP, Gosnell T and Nurmikko AV (2003) *Compact Blue-Green Lasers*. Cambridge, UK: Cambridge University Press.

Weber MJ (2001) *Handbook of Lasers*. Boca Raton, FL: CRC Press.

Yariv A (1995) *Quantum Electronics,* Third Edition. John Wiley & Sons.

LASER-INDUCED DAMAGE OF OPTICAL MATERIALS

A J Glass and A H Guenther, The University of New Mexico, Albuquerque, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Shortly after the invention of the laser, in the early 1960s, it became apparent that laser performance was limited by the ability of optical materials to withstand high energy and power densities. Broadly speaking, laser damage arises from two causes: (i) thermal effects, which are associated with high average power; and (ii) dielectric breakdown, which is associated with high values of the local electric field. Early researchers quickly recognized that most of the observed damage effects arose from extrinsic causes, such as impurities and inclusions in the bulk material, or particles or defects on exposed optical surfaces. At that time, optical materials had not been developed for high-power use, and lacked the quality needed for laser applications. In fact, in many situations, the laser provided the inspection light source used to assess the properties of the materials, such as absorption, defects, scattering centers, surface topography, and dynamic behavior. Thus, the field of laser damage studies was broadened to include many aspects of material preparation and characterization. At high-power densities, nonlinear effects become important, often limiting performance without creating any irreversible change in the material's properties. The most general definition of laser-induced damage includes all effects arising in optical materials and components that limit or degrade the performance of laser systems.

The pursuit of improved performance at high values of power and energy densities was motivated by several factors and applications. In general, improved tolerance to high flux and fluence leads to smaller, lighter, and more efficient systems. These are particularly important for military and energy applications. Since the 1970s, the laser fusion program alone has produced an extensive body of research on laser-induced damage in high-power nanosecond-scale systems, which has enabled orders of magnitude improvement in the damage resistance and extraction efficiency of glass lasers. More recently, with the development of fiber laser systems producing and transporting a kilowatt of average power in a fiber of less than 100 μm diameter, laser damage phenomena on an entirely different spatial scale have become important.

Since the early 1970s, the literature of laser damage research has been collected in a series of annual publications of the Boulder Damage Symposium, initiated by the authors of this article, which was first held in 1969. Recently the papers presented at the Damage Symposium in the years 1969 to 1998 were collected and published in CD form by the SPIE. A CD containing the proceedings of the next five years (1999–2003) will be released by SPIE this year. Also in the late 1960s, in the Soviet Union, Academician Aleksei Bonch-Bruevich launched a comprehensive research project on the nonresonant interaction of laser radiation with matter, opening up a new branch of photonics research, known in the Soviet Union as power optics. In 1969, he founded the All-Union Conference on Interaction of Optical Radiation with Matter (AURICOM), which was held every two to three years until 1990. After 1990, the conference was renamed Nonresonant Laser-Matter Interaction (NLMI), and chaired by Professor Mikhail Libenson. NLMI was held in St Petersburg in 1996, 2000, and

2003. The proceedings of the 1996 AURICOM and subsequent NLMI conferences have been published by the SPIE. We are saddened to note that Professor Libenson died in February 2004.

The Boulder Damage Symposium grew out of a one-day Symposium convened in Boulder in 1969 under the auspices of the ASTM Subcommittee on Lasers, to establish standards for laser materials. In their summary of the first Symposium, the conference organizers remarked:

> The entire question of the nature of the standards remains to be addressed. Should they take the form of energy density at which catastrophic damage is likely to occur, or should they be specified in terms of a mean number of shots at a given level of energy density before certain degradation of performance is measured? The latter would seem to express the kind of information that the buyer of laser materials would find most useful. Once a standard is agreed upon, a meaningful test configuration must be established. It is likely that this will be a well-controlled oscillator amplifier chain with the test specimen an active element of the system, but this is not certain. Another point to be determined is that of the accuracy to which its specifications should be written. This depends in part on the kind of quality control the producers of laser glass believe is feasible.

The questions raised in this first Symposium remain at the core of laser damage studies today. How do we define damage, in terms of the physical effect on the material or in terms of the degradation of performance of the system? How do we measure damage thresholds or values? How do we specify the properties and preparation of the sample, including the relevant experimental and material variables? How can we ensure that damage values are reproducible, unless we make the measurements in a very well controlled and characterized system? With what precision should damage levels be stated, in order to be useful? Extensive research was required in order to answer these questions, and to elucidate the fundamental nature of damage mechanisms. Many of the results of this research are documented in the 35 year record of the Boulder Damage Symposium and the corresponding Russian-based conferences.

From the beginning of this area of research, investigators devoted a significant effort to establishing the 'first causes' of damage phenomena. However, they quickly realized that in most cases, damage occurred at surfaces and material interfaces, or was mediated by impurities or inclusions in bulk materials. As material fabrication processes improved, and higher-quality materials became available, observed damage thresholds approached the intrinsic limits of the pure material. This was particularly true of silicate glass, metal reflectors, and in some cases, polymers. In fact, silicate glass has become one of the purest noncrystalline materials available in bulk quantities. However, for practical purposes, most laser systems are limited by damage at surfaces, coatings, or impurities. Impurities aggregate at grain boundaries in crystalline materials, and coalesce into inclusions in glasses, resulting in absorption, scattering, and field enhancement.

The coherent nature of laser light makes any scattering phenomena potentially damaging, because the scattered light interferes constructively with the laser beam, leading to the formation of intensity hot spots. Every interface in an optical system is potentially a source of constructive interference between the reflected and incident waves. Multilayer dielectric coatings are particularly vulnerable to damage, due to the presence of residual stress as well as field enhancement from Fresnel reflection. This has led to attempts to mitigate the problem by use of specific designs, such as avoiding AR coatings by using Brewster-angle surfaces, or use of nonquarter wave layers to move the peak field away from the vicinity of boundaries where impurities tend to collect. Other damage-sensitive design considerations can be found in the literature. Multiple reflections are the basis of photonic crystal optics, and applications of these devices to high-power density systems will undoubtedly be limited by field-enhanced damage phenomena.

In CW operation, of course, thermal effects dominate. They may be irreversible, creating permanent damage sites in the medium, or reversible, such as color centers that can be bleached out. As laser pulse durations vary from nanosecond to picosecond and femtosecond durations, the nature of the dominant damage mechanisms changes as well. As a rule of thumb, the scaling of damage thresholds as $\tau^{1/2}$, where τ is the pulse duration, was established empirically, and was found to be valid over a wide range of pulse durations. This scaling implies the presence of diffusion effects, either of heat or of electrons, in the vicinity of a damage center. As might be expected, this scaling does not apply for very short pulses (femtoseconds or less), where the extremely high field values can lead to instantaneous breakdown of the material.

Whether the damage is due to field enhancement or is impurity-mediated, it is difficult to establish a reproducible value for a damage threshold, unless the sample is macroscopic in extent, and the sampling laser beam is reproducibly of high spatial and temporal quality, preferably truly single-mode. Damage measurements on microscopic areas are unlikely to be representative of the entire sample. Damage will occur first at a point where the field is strongest or the material is weakest, and a sufficient

area must be tested to ensure that it includes such sites. Also, unless the beam quality of the laser used for testing is well controlled and well characterized, it will not be possible to reproduce the test results. It is equally important to characterize the samples, including their method of preparation, cleaning, and test environment. For damage research, as opposed to simple threshold testing, postmortem examination is also required, to elucidate the mechanism of the observed damage effect. However, we now have several techniques, such as photothermal deflection, that can provide useful pre-catastrophic indicators of the onset of laser damage, and thus are appropriate for quality control.

Professor Roger Wood had published a series of reviews and textbooks on laser damage in optical materials, which provide an overview of the field, with appropriate emphasis on measurement techniques, characterization of the test setup and sample, and implications for system design. He has also edited a collection of seminal publications in the field, published by the SPIE in 1990.

It is gratifying to those working in this field to observe how much of the research carried out to elucidate and, if possible, avoid the deleterious effects of laser damage has been applied to the constructive purpose of laser-assisted material processing and manufacturing. The introduction of low-cost, reliable femtosecond lasers has stimulated activity in this field, since it enables the experimenter to control to exquisite precision the power and energy deposited on a material. With femtosecond pulses, one can ablate material from the surface of an object while minimizing any collateral damage to the interior material. Surface layers of dirt can be removed from priceless art works such as oil paintings, without concern for damage to the object itself. Objects with arbitrary shape can be built up from metal particles, using laser additive manufacturing techniques. Thin films and photonic crystal structures can be created using laser-assisted deposition. Laser irradiation provides a power tool for surface cleaning, conditioning, and annealing. Laser interaction with materials and laser-assisted manufacturing techniques comprise a vital area of research in the materials science community.

In both the US and Russia, study of the interaction of intense coherent light with optical materials began with a disarmingly simple objective; to understand the fundamental mechanisms of interaction. In the review and summary of the 1969 ASTM Symposium, the editors wrote:

> In view of the number of problems remaining to be resolved, it is suggested that another Symposium on laser damage be held in 1970. Hopefully, at that time a better understanding of the nature of damage in laser glass will have been obtained. Higher threshold values will have been reproducibly achieved, and some agreement can be arrived at regarding useful and realistic standards.

Thirty-five years later, damage levels in optical materials have been significantly improved, and realistic and useful standards have been developed under the auspices of the International Organization for Standardization (ISO), and are now accepted globally. However, as the operating regimes of lasers have moved to femtosecond temporal scale and micrometer dimensions, new phenomena have arisen that require further research. An entirely new technology of laser-enabled manufacturing and materials conditioning has developed, based on the same physics as laser damage. Combining the disciplines of materials science and applied optics, the interaction of intense light with materials remains a scientifically challenging and economically important field of research.

See also

Optical Coatings: Laser Damage in Thin Film Coatings.

Further Reading

Bonch-Bruevich AM, Konov VI and Libenson MN (eds) (1990) *Optical Radiation Interaction with Matter*. SPIE Proceedings, vol. 1440.

Glass AJ and Guenther AH (eds) (1969) *Damage in Laser Glass Symposium: Review and Summary*. ASTM Special Technical Publication 469.

ISO Online Catalogue, Section 31.260, http://www.iso.ch/iso/en/CatalogueListPage.CatalogueList

Laser-Induced Damage in Optical Materials, Collected Papers 1969–1998, (CD-ROM), SPIE, Bellingham, Washington (1999).

Laser-Induced Damage in Optical Materials, Collected Papers 1999–2003, (CD-ROM), SPIE, Bellingham, Washington (2004) (To be published).

Libenson MN (ed.) (1997) *Non-Resonant Laser-Matter Interaction (NLMI-9)*. SPIE Proceedings, vol. 3093.

Libenson MN (ed.) (2001) *Non-Resonant Laser-Matter Interaction (NLMI-10)*. SPIE Proceedings, vol. 4423.

Libenson MN (ed.) (2004) *Non-Resonant Laser-Matter Interaction (NLMI-11)*. SPIE Proceedings, vol. 5506, (To be published).

Wood RM (ed.) (1986) *Laser Damage in Optical Materials*. Bristol, UK: Adam Hilger.

Wood RM (ed.) (1990) *Selected Papers on Laser Damage in Optical Material*. SPIE Milestone Series, vol. MS24. SPIE Optical Engineering Press.

Wood RM (ed.) (2003) *Laser-Induced Damage in Optical Materials*. Bristol, UK: Institute of Physics.

Wood RM (ed.) (2003) *The Power- and Energy-Handling Capability of Optical Materials, Components, and Systems*. SPIE Press Tutorial Texts in Optical Engineering, vol. TT60.

LIGHT EMITTING DIODES

J Schanda, University of Veszprém, Veszprém, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Short Historic Overview

During the second half of the nineteenth century, electric lighting became possible. In 1879, Edison demonstrated his incandescent lamp, and although some electric arc lamps competed with the incandescent lamp bulb for a few years, the first 50 years of electrical lighting was dominated by the incandescent lamp. Gas discharge lamps (both low- and high-pressure lamps) became more widely distributed only by the end of the third decade of the twentieth century, and widespread distribution of the fluorescent lamp, that still dominates our offices and many other lighting applications today, was generally accepted only after the Second World War.

The early incandescent lamps had an efficacy of a few lm/W, but by 1920 this could be increased by up to 10–15 lm/W. Gas discharge lamps started with approximately 30 lm/W, but by 1960 their efficacy increased to about 70 lm/W. Fluorescent lamps still provide only 70 to 100 lm/W; high-pressure lamp efficacy could be increased above this value. The different lamp types can be used, however, in different applications, depending on the size and the unit power of the lamp type. Nonincandescent light generation in solid-state materials dates back to the 1920s: a faint glow of a SiC crystal was observed when current passed through it at a point contact. In the 1930s, electroluminescence of ZnS powder layers was detected (Destriaux effect). By the mid-1960s, it became clear that by using the then available technology, good luminous efficiency could be expected only in III–V compounds (GaAs, GaP, and their ternary and quaternary compounds using Al, In, and N in the compounds). Pure GaAs has a bandgap in the infrared (IR), thus one can build good IR-emitters using this material, but no visible light emitter. However, GaAs proved to be an excellent coherent light-emitting material and enabled the construction of IR-emitting semiconductor lasers.

First visible light-emitting diodes were fabricated using a GaAsP alloy. Figure 1 shows how the efficacy of III–V compound LEDs can be increased by introducing new compositions, and later, new technologies and better light extraction techniques. In the figure we also show the efficacy values of average incandescent, high-pressure Hg- and fluorescent lamps. The first LEDs emitted in the red part of the spectrum, and only by 1975 was the green LED invented. By 1985, practically any color between red and green became possible, and the efficacy of these lamps reached a level where for monochromatic light applications, their efficacy could compete favorably with filtered incandescent light. LEDs emit in a relatively narrow wavelength band (20 nm–30 nm bandwidth), and if one has to filter from the continuous spectrum of an incandescent lamp such a narrowband, the loss becomes considerable. LEDs needed only a few volts to function and it became general practice to build LEDs with approximately a 20 mA current load. These LEDs became very popular for interior signaling (on instruments, household appliances, and in similar applications). The goal was, however, to be able to produce white light, but the blue-emitting LED was still missing. Only after 1990 could the problem of producing p–n junctions from GaN, that had a bandgap large enough to emit blue light, be solved. Different ternary and quaternary compounds, as shown in Figure 1, enabled efficacies comparable to incandescent lamp efficacy, and the race for producing white light began.

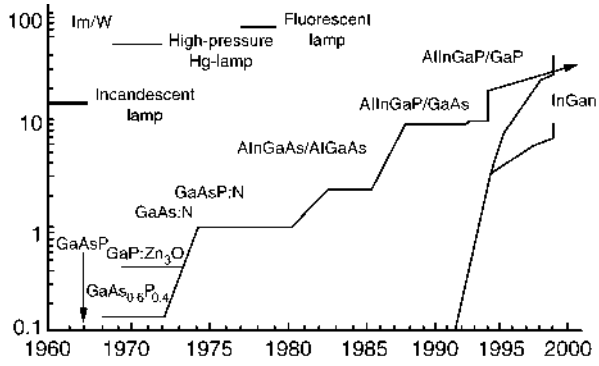

Figure 1 Increase of luminous efficacy of LEDs during the past 40 years. Courtesy of Lumileds Lighting.

Physical Fundamentals

Figure 2 shows a forward-biased p–n junction. As injected free electrons and holes are driven towards the junction, they can radiatively recombine, emitting a photon. The probability of photon emission versus nonradiative recombination will depend on the band

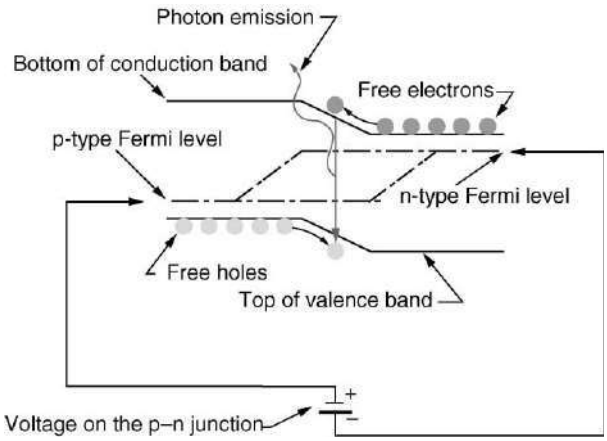

Figure 2 Forward biased p–n junction, light generation via recombination of free charge carriers.

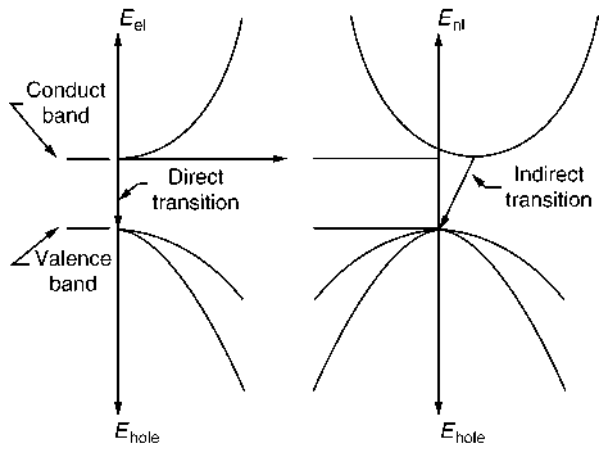

Figure 3 Band to band recombination in direct and indirect band semiconductor.

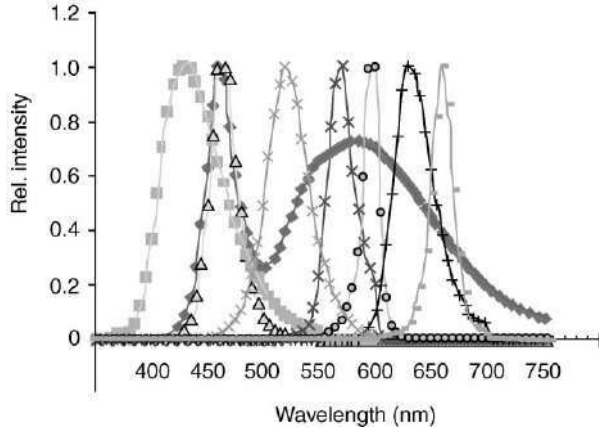

Figure 4 Relative spectral power distribution of some LED structures.

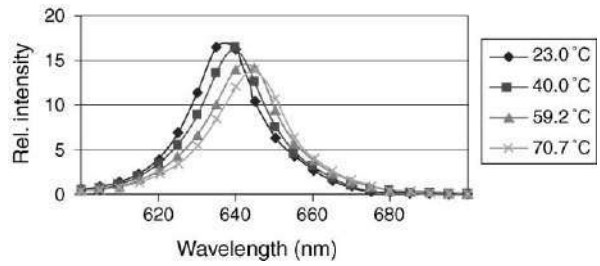

Figure 5 Spectral power distribution of a red LED at different temperatures.

structure of the semiconductor. Figure 3 shows the two possibilities of a direct and an indirect bandgap semiconductor. GaAs is, for example, a direct bandgap material. In this material free electrons at the bottom of the conduction band can recombine with free holes at the top of the valence band, as the momentum of the two 'particles' is the same and a photon emission takes care of the principle of energy conservation. GaP is an indirect bandgap material where the momentum of the free electrons at the bottom of the conduction band differs from the momentum of the free holes at the top of the valence band, and thus the recombination needs not only the fulfilment of the conservation of energy, but also of momentum. Thus, besides photon emission, phonon emission or absorption has to take place. Thus, the probability of the transition will decrease and competing nonradiative processes will become important. Using N doping produces very shallow donor states and from these an efficient recombination can take place. This enabled the production of early relatively efficient green-emitting LEDs.

GaN is ideal for producing LEDs, as it is a direct gap material and has a high bandgap. The technological difficulties are, however, enormous. Modern high-efficiency LEDs use (Al)InGaN-based materials for shorter wavelength LEDs, and AlInGaP-based materials for longer wavelength LEDs. Figure 4 shows a number of LED spectra of different wavelengths including a white LED. Its spectrum consists of a blue maximum produced by the LED emission and a yellow broader band produced by the luminescent material excited by the blue emission of the LED.

LEDs are semiconductor materials, and as with every semiconductor the LED characteristics are also temperature dependent. Not only the efficiency of the current carrier–photon emission process depends on the temperature, but the bandgap of the base material also depends on the temperature, and thus both the efficacy and the relative spectral distribution of the emission change with temperature. Figure 5 shows the spectral power distribution of a red LED at different temperatures.

Modern High Brightness LEDs

Besides the high quantum efficiency of the semiconductor junction, two further requirements to achieve high luminous efficacy are the efficient light extraction from the junction and keeping the temperature of the semiconductor chip low.

All semiconductors for producing LEDs have a high refractive index, thus light reflection at the semiconductor–air interface is relatively high, and for obliquely impinging rays the critical angle for total internal reflection is soon reached. Different manufacturers shaped the semiconductor chip in different forms to secure maximum light extraction. Figure 6 shows the cross-section of an LED, where the side planes have been cut obliquely to secure good light extraction.

The other problem that has to be solved is good heat conduction from the chip to a heat sink. At the time of writing, highest single chip power loads were in the order of 5 W. The structure of such a construction is seen in Figure 7. The GaN-based LED is grown on a visible light transparent sapphire (Al_2O_3) substrate. In modern LEDs the structure itself is quite complicated; in the vicinity of the p–n junction some nanometer thick layers are deposited forming quantum wells (QW(s)). The p-type original upper layer is metalized and soldered to the submount serving as heat-sink. The n-type layer gets a second soldered heat conducting current leading to the submount. As seen, the original structure grown on the sapphire substrate is inverted, and now the light can be extracted through the sapphire layer. The metalization on the p-type layer serves at the same time as a mirror. High brightness LEDs are now available in package sizes of over 100 lumen.

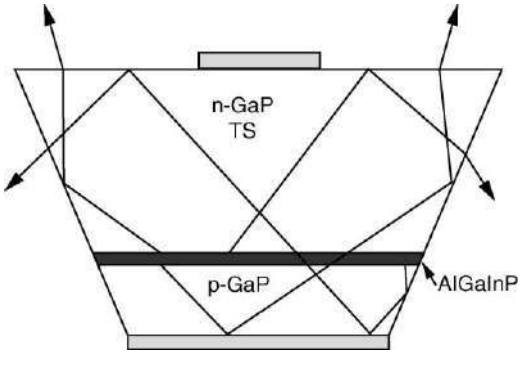

Figure 6 Truncated pyramid structure to secure good light extraction from the LED chip. Courtesy of Lumileds Lighting.

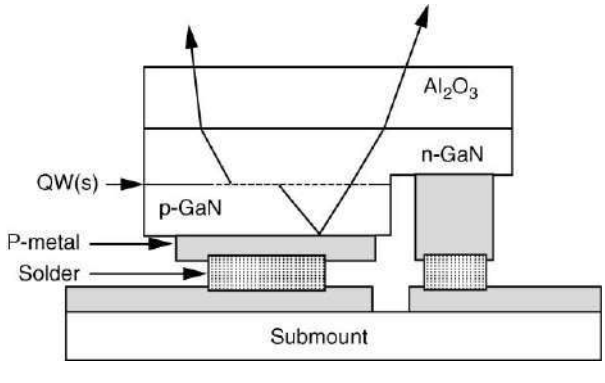

Figure 7 Inverted, flip-chip structure for good heat-sinking of the LED. Courtesy of Lumileds Lighting.

White LED Structures

The advantages of LEDs (low drive voltage, extreme long lifetime of 50000 to 100000 hours, robust shock-resistant construction) can open up a new era of lighting engineering, if white light emission can be obtained. The LED structure in itself is a narrowband emitter, thus – with inorganic LEDs – it is not possible to produce white light within one chip. Organic LEDs (OLEDs) promise the production of broadband and white light emission, but these are still in an experimental phase. At present they are used in smaller, portable displays, such as those used on portable phones, cam-coders, etc. To get further information, go to the Further Reading section at the end of this article.

At present three techniques to generate white light get thorough investigation: first, mixing the light of a red, a green, and a blue LED. This should theoretically give the highest efficiency; the problem is, however, that the three LEDs have different temperature dependence and aging characteristics, and thus it is difficult to keep the chromaticity of the mixed color constant (the three colored LEDs can be used in LED-based displays to produce full-color reproduction). At present the following technique is widely used. A blue LED excites a yellow phosphor, so that the mixture of the blue and yellow lights produce a white appearance. The color rendering of such white LEDs is not too good, and therefore alternative techniques using a green and a red phosphor are receiving increasing attention. The third method uses the same principle, but with an LED emitting in the near ultraviolet. This has the advantage that it provides a greater selection of phosphors, but as the energy difference between the UV photon and the photon emitted by the phosphor is larger than in the case of a blue-emitting LED, the theoretical efficiency is lower.

The phosphor layer is deposited on the chip surface, and for blue-emitting diodes special care has to be taken that the LED light and the phosphor

light are properly mixed so that the emitted light shows no directional color shift. Early white LEDs suffered much from the fact that the spectral distribution of the light changed considerably with viewing angle.

The problems with phosphor-coated white LEDs are how a suitable correlated color temperature and color rendering can be obtained. Early white LEDs had very bluish chromaticity, their correlated color temperature was very high, above 7000 K. Preferred white light has different chromaticity in different cultural regions, in the Far East 6000 K to 6500 K is preferred, in Europe and North America for general lighting a correlated color temperature of 4000 K is well accepted and for social and home lighting even lower correlated color temperatures, 2800 K to 3200 K are preferred. Recently, white LEDs based on blue chips and special phosphor blends have been produced that meet these requirements.

A further problem is the color rendering of the white LEDs. The Commission Internationale d'Eclairage (CIE), the International Commission on Illumination, provides standards, recommendations, and technical reports in the field of light and lighting, including vision, color, and image processing. Photometry and colorimetry are based on the standards and recommendations of the CIE.

The CIE established a method of calculating a color-rendering index in 1965 and updated it in 1974. Despite several attempts, the lighting community has been unable to agree on a revision of this outdated document since the 1970s. Thus, color-rendering indices are still calculated based on this method, although several investigations have shown that the calculated indices do not correspond with the visual observation. CIE, therefore, established a technical committee (TC) dealing only with the question of LED color rendering. It is hoped that this TC will be able to come up with some recommendations soon and, therefore, that LED manufacturers can tailor the emission spectra of their LEDs for optimal efficacy and color rendering. Until this happens, however, the published color rendering indices, especially the General Color Rendering Index (R_a) have to be handled carefully. The color appearance of objects under LED illumination might deviate considerably from what one expects based on the R_a index.

Application of LEDs

LEDs are the solution for practically every signaling application. Figure 8 shows an LED traffic light. Here it is also shown that, despite the fact that LEDs have extremely long life expectancy, the failure of the electronic circuitry can lead to unpleasant dark spots and rows in the display.

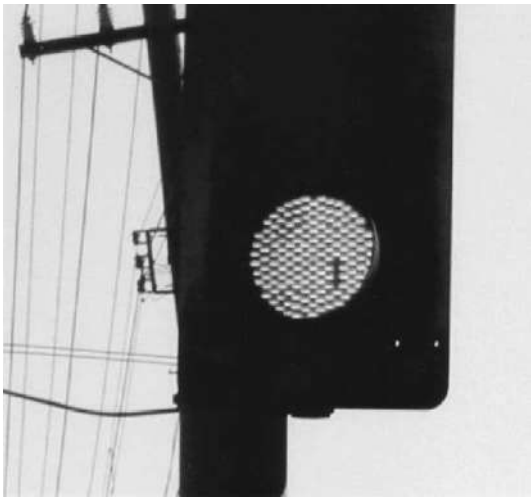

Figure 8 Green LED traffic light. Due to an electronic failure some LEDs stay dark.

LEDs are increasingly used in large active posters and displays. In such applications the highly saturated color of LED light is an advantage: LED displays can be very bright and emit saturated colors.

LEDs have an advantage in every situation where colored light has to be produced. Thus, they become the preferred automotive stoplights, dashboard lights, and are widely used in other signaling applications on the roads, at airfields, and in other traffic situations; much decorative lighting is now using LEDs.

Experiments are underway to find the best applications for white LEDs. Their drawback at the moment is still that their single unit power is below ten watts, as we often need ten to a hundred times larger units. For such applications, some LED clusters have been built, but these are still not competitive with other light sources. At places where the ruggedness of the LEDs is of an advantage, white LED lights are used – for example, in aircraft reading lights. One area where we will probably soon find higher-intensity white LEDs is car headlamps. Both the efficacy and the luminous flux output of the single LED units is increasing continuously. Applications, unbelievable only a few years ago, can now be realized efficiently with LEDs. Thus it is very difficult to foresee in which applications – beyond signaling – LEDs will first supersede other light sources, but it is certain that within the next decade we will see many new applications of LED sources.

List of Units and Nomenclature

Chromatic adaptation	Adaptation by stimuli in which the dominant effect is that of different relative spectral distributions.
Chromaticity	Partial description of the color stimulus, describing only chromatic aspects and not the 'intensity' (the luminance) of the stimulus. In light sources, chromaticity is an important descriptor, as the level of illumination can be changed by changing the distance between the source and the illuminated object. The luminance of the object will depend on the illumination and the reflection (transmission) characteristics of the object.
Color	In color science one distinguishes between color (perception), characteristic of visual perception that can be described by attributes of hue, brightness (or lightness) and colorfulness (or saturation or chroma), and color stimulus, a specification of the optical stimulus in terms of operationally defined values, such as three tristimulus values, or luminance, and chromaticity (or dominant wavelength and saturation).
Color rendering	Measure of the degree to which the psychophysical color of an object illuminated by the test illuminant conforms to that of the same object illuminated by the reference illuminant, suitable allowance having been made for the state of chromatic adaptation.
Correlated color temperature	The temperature of the Planckian radiator whose perceived color most closely resembles that of a given stimulus at the same brightness and under specified viewing conditions.
Efficacy (luminous efficacy of a source)	Quotient of the luminous flux emitted by the power consumed by the source.
Luminous efficiency	Ratio of radiant flux weighted according to $V(\lambda)$, the spectral luminous efficiency of the human eye, to the corresponding radiant flux.
Spectral luminous efficiency (of the human observer)	Ratio of the radiant flux at wavelength λ_m to that at wavelength λ, such that both radiations produce equally intense luminous sensations under specified photometric conditions and λ_m is chosen so that the maximum value of this ratio is equal to 1.

See also

Incoherent Sources: Lamps. **Lasers:** Semiconductor Lasers.

Further Reading

Fukuda M (1991) *Reliability and Degradation of Semiconductor Lasers and LEDs*. Boston, MA: Artech House.

Rudden MN and Wilson J (1993) *Elements of Solid State Physics*. Chichester, UK: Wiley.

Sze SM (1985) *Semiconductor Devices: Physics and Technology*. New York: Wiley.

Wenckebach WT (1999) *Essentials of Semiconductor Physics*. Chichester, UK: Wiley.

Žukauskas A, Shur MS and Gaska R (2002) *Introduction to Solid-State Lighting*. New York: Wiley.

M

MAGNETO-OPTICS

Contents

Faraday Rotation, CARS, ODMR, ODSR, Optical Pumping
Interband Magnetoabsorption, Cyclotron Resonance, Spin Flip Raman Scattering

Faraday Rotation, CARS, ODMR, ODSR, Optical Pumping

H Pascher, Universität Bayreuth, Bayreuth, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

A complete understanding of the band structure of a semiconductor close to the Fermi energy requires information on the fundamental energy gap, the effective masses and the effective g-factors of the valence and conduction band. Magneto-optical methods in principle allow one to determine all these quantities; they are particularly well suited to measure the spin splittings of the energy bands.

Faraday rotation is the rotation of the plane of polarization of linearly polarized light propagating through a sample parallel to an external magnetic field. The magneto-optical Kerr effect is the analog for reflection. Though in general it yields much smaller angles of rotation than the Faraday effect, it has to be applied in spectral ranges where the investigated sample is not transparent.

If a semiconductor is excited by circularly polarized light with photon energies not too high above the energy gap a spin polarization in the conduction band can be obtained (optical pumping). A small transverse magnetic field reduces the degree of this spin polarization (the Hanle effect). Owing to the spin polarization of the photo-excited conduction electrons the photoluminescence radiation is partly circularly polarized. Thus mounting the sample in a microwave resonator and applying a longitudinal magnetic field, electron spin resonance transitions can be detected optically by monitoring the degree of polarization of the photoluminescence (ODCESR: optically detected conduction electron spin resonance). Via a 'nuclear Hanle effect' mediated by the hyperfine interaction, nuclear magnetic resonance transitions can also be detected optically (ODNMR). ODCESR and ODNMR are summarized under ODMR (optically detected magnetic resonance).

In modern optics nonlinear optical effects have gained particular importance. Since electron spin resonances in semiconductors are Raman-active transitions, coherent Raman methods such as CARS (coherent anti-Stokes Raman scattering) or CSRS (coherent Stokes Raman scattering) are of great importance for the determination of spin splittings of energy bands and the investigation of magnetic structures in semiconductors.

In the present contribution the physical basis for these methods will be discussed and typical examples for their application will be presented.

For Faraday rotation and CARS the systems from which these examples will be taken are magnetic or dilute magnetic semiconductors. In these materials paramagnetic ions like Mn^{++} or Eu^{++} are incorporated. There are two types of exchange interactions: (i) between the magnetic ions which in most cases causes an antiferromagnetic coupling; and (ii) between the ions and free carriers which causes a

strong modification of the spin splittings of the energy bands with respect to nonmagnetic materials. Thus the Faraday effect or CARS are particularly well-suited methods for the investigation of such systems.

As examples of optical pumping and ODMR, results obtained with III–V semiconductors will be presented, which show the most important applications of these experimental methods.

Faraday Rotation and Magneto-optical Kerr Effect

Basic Principles

As already pointed out in the Introduction, the Faraday effect is the rotation of the plane of polarization of linearly polarized light traversing a sample parallel to a magnetic field. The angle of rotation Θ_{Far} is proportional to the length l of the sample and to the strength B of the magnetic field:

$$\Theta_{\mathrm{Far}} = VlB \qquad [1]$$

where V is called Verdet's constant.

For a phenomenological understanding of the effect one can consider plane polarized light as a superposition of right- and left-hand circularly polarized components. If the indices of refraction are different for both components, then at the end of the sample, when the two components are recombined, a phase shift occurs which manifests itself as a rotation of the polarization plane. The angle resulting from this consideration is

$$\Theta_{\mathrm{Far}} = \frac{l\omega}{2c}(n_- - n_+) \qquad [2]$$

where ω is the frequency, c the velocity of light, and n_- and n_+ are the refractive indices for left- and right-hand polarized light, respectively. If one neglects ellipticity of the light behind the sample only the real parts of the complex refractive indices $\tilde{n}_\pm$ are important,

$$\tilde{n}_\pm = n_\pm - \mathrm{i}\kappa_\pm \qquad [3]$$

The imaginary parts, however, are needed to understand the magneto-optical Kerr effect which occurs in the light reflected from the sample surface. According to Fresnel's formula for vertical incidence the field amplitude of the reflected beam is:

$$\tilde{E}^{\mathrm{r}}_\pm = \frac{1-\tilde{n}_\pm}{1+\tilde{n}_\pm}\tilde{E}^{\mathrm{i}}_\pm = \tilde{r}_\pm \tilde{E}^{\mathrm{i}}_\pm \qquad [4]$$

where the superscript 'r' refers to the reflected beam and 'i' to the incident beam. The complex reflection coefficient can be split into an amplitude and a phase factor:

$$\tilde{r}_\pm = r_\pm \exp(\mathrm{i}\psi_\pm) \qquad [5]$$

By comparing eqns [4] and [5] the phase shifts during reflection can be expressed in terms of the refractive index:

$$\psi_\pm = \arctan\frac{2\kappa_\pm}{1-n_\pm^2-\kappa_\pm^2} \qquad [6]$$

Finally the Kerr angle of rotation is obtained from the phase shifts in an analogous way to the Faraday rotation:

$$\Theta_{\mathrm{Kerr}} = \tfrac{1}{2}(\psi_- - \psi_+) \qquad [7]$$

It is seen that both effects can be understood in terms of the refractive index. In a simple oscillator model the refractive index has the form given in eqn [8]:

$$\tilde{n}_\pm^2 = \tilde{\varepsilon}_\pm(\omega,B) = \varepsilon_\infty + \sum_{i,j}\frac{F_{i,j}^\pm}{\hbar\omega_{i,j}^\pm - \hbar\omega - \frac{1}{2}\mathrm{i}\Gamma_{i,j}^\pm} \qquad [8]$$

with $\hbar\omega_{i,j}^\pm$ representing the transition energies between the energy levels i and j in σ^+ and σ^- polarization, respectively, and $F_{i,j}^\pm$ and $\Gamma_{i,j}^\pm$ are the respective oscillator strength and linewidth parameter.

From eqn [8] it is seen that when tuning the frequency, ω, resonant enhancements of the rotation angles are expected whenever the frequency corresponds to an allowed electronic transition, e.g., an exciton in a wide-gap semiconductor. This is the reason why the Faraday and magneto-optical Kerr effect can yield information on interband transition energies. In the case of such a resonance, the refractive index will be dominated by the particular term in the sum, corresponding to that transition. By a Taylor expansion one can then show that the difference of the refractive indices for σ^+ and σ^- polarization and thus the angles of rotation (see eqns [2] and [7]) are proportional to the difference of the transition energies for right- and left-hand circularly polarized light, and this difference depends on the spin splittings of the valence and conduction band. So we understand why the most important information obtained from Faraday or magneto-optical Kerr effect studies is on the spin splittings of the energy bands.

A quantum mechanical description of the relation between the material properties and the magneto-optical effects beyond the simple oscillator model given here is published by Boswarva *et al.* and the reader is referred to this paper for

further studies. A matrix formalism to describe the effects for a multilayer system has been developed by Nies and Kessler.

Examples of Experimental Results

The straightforward method to measure Faraday rotation is to put the sample between polarizers and sweep the magnetic field. Then the transmitted intensity is proportional to a $\cos^2$ function. Subsequent maxima demonstrate a rotation by 180°. There have been numerous successful investigations done in this way, but the possibilities of detecting small angles of rotation are very limited in this method. Therefore differential methods have been developed either using a polarizing beam splitter, or a photo-elastic modulator which periodically switches the state of polarization. Rotation angles of the order of 10^{-4} rad can be detected in this way.

Figure 1 shows the Faraday rotation of an epitaxial film of $Pb_{0.933}Eu_{0.067}Te$ (thickness 11μm) for two temperatures. The plot demonstrates behavior typical for a diluted magnetic semiconductor in its paramagnetic phase. The spin splittings of the valence and conduction band and accordingly the Faraday rotation are proportional to the magnetization of the sample which behaves like a Brillouin function. (The lines without noise are fitted Brillouin functions.)

Examples of the application of the magneto-optical Kerr effect to the investigation of semiconductors, particularly multilayer systems, are published, e.g., by Pohlt *et al.* and Krenn *et al.*

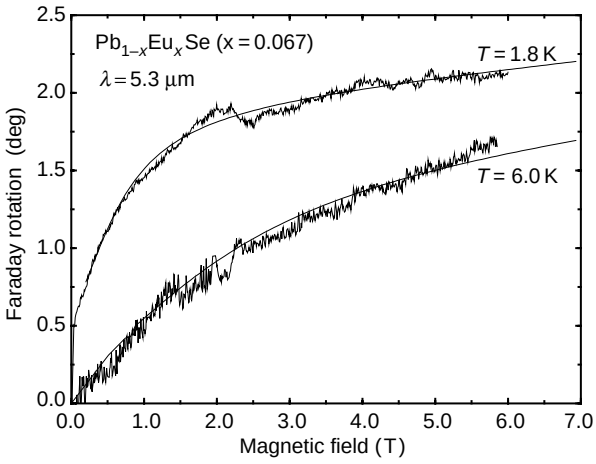

Figure 1 Faraday rotation as a function of magnetic field of a PbEuSe epitaxial film for two temperatures. Lines with noise: experimental recordings; lines without noise: fitted Brillouin functions.

Coherent Raman Scattering

Basic Principles

A more direct method for the determination of spin splittings of energy bands is provided by Raman scattering. There is a huge number of investigations of spin transitions in semiconductors using spontaneous Raman scattering (for a review of spontaneous Raman scattering in dilute magnetic semiconductors see the article by Ramdas and Rodriguez). However, there are several advantages in the spectroscopic application of the nonlinear Raman techniques compared to spontaneous Raman scattering:

(i) The scattered radiation is emitted as a collimated beam which is more easily detectable than the light spontaneously scattered into the entire solid angle. This advantage is particularly important for the investigation of narrow-gap semiconductors which have to be investigated by infrared light, a spectral range where the detectors are less sensitive than in the visible.
(ii) Fluorescence can be strongly suppressed by spatial filtering.
(iii) The spectral resolution is determined by lasers instead of a classical spectrometer.
(iv) Raman scattering cross-sections can be calculated from the intensity of a collimated beam without taking care of the aperture of the detecting system.

The coherent Raman techniques are special cases of optical four-wave mixing where two laser beams with frequencies ω_L and ω_S ($\omega_L > \omega_S$) are superimposed in the sample. Due to nonlinear interactions radiation with new frequencies is generated, among them:

$$\omega_{CARS} = 2\omega_L - \omega_S \qquad [9]$$

If one identifies ω_L with the frequency of the pump laser and ω_S with that of the scattered Stokes radiation of a spontaneous Raman experiment, then ω_{CARS} is the frequency of the anti-Stokes radiation. For the intensity of the radiation generated at ω_{CARS} one finds:

$$I_{CARS} = \frac{9}{16c_0^4\varepsilon_0^2}\frac{\omega_{CARS}^2}{n_{CARS}n_S n_L^2}\left|\chi^{(3)}\right|^2 l^2 I_S I_L^2 \times\left[\frac{\sin(\overrightarrow{\Delta k}\,\vec{l}/2)}{\overrightarrow{\Delta k}\,\vec{l}/2}\right]^2 \qquad [10]$$

where c_0 is the velocity of light, l is the length of the sample, I denotes the intensity, and n the refractive index at the respective frequency. $\chi^{(3)}_{kjkj}$ is the third-order nonlinear susceptibility of the scattering medium and

$$\overrightarrow{\Delta k} = 2\vec{k}_{\mathrm{L}} - \vec{k}_{\mathrm{S}} - \vec{k}_{\mathrm{CARS}} \qquad [11]$$

is the phase mismatch. $\overrightarrow{\Delta k} = 0$ can be achieved by noncollinear alignment of the laser beams. However, for very thin epitaxial layers with $l \ll \pi/\Delta k$ the phase factor in eqn [10] can be approximated by 1.

The information on the investigated material is contained in the nonlinear susceptibility. It can be calculated by time-dependent perturbation theory taking into account a ground state $|a>$ and intermediate states $|b>$, $|c>$ and $|d>$. It turns out that $\chi^{(3)}$ is proportional to a sum over all states of the system as intermediate states; in the sum 24 terms like the one given in eqn [12] have to be added:

$$\frac{\mu_{ijkl}}{(\omega_{ba} - \omega_{\mathrm{L}} - \mathrm{i}\Gamma_{ba})(\omega_{ca} - \omega_{\mathrm{L}} + \omega_{\mathrm{S}} - \mathrm{i}\Gamma_{ca})(\omega_{da} - \omega_{\mathrm{CARS}} - \mathrm{i}\Gamma_{da})} \qquad [12]$$

μ_{ijkl} is a product of four dipole matrix elements between states $|a>:|b>$; $|b>:|c>$; $|c>:|d>$ and $|d>:|a>$ with field polarizations as given by the indices. $\omega_{\alpha\beta} = (E_\alpha - E_\beta)/\hbar$ and $\Gamma_{\alpha\beta}$ is the linewidth of the respective transition. Resonances of the susceptibility are found by inspection of the denominators. $\chi^{(3)}$ exhibits Raman-like resonances whenever $\hbar(\omega_{\mathrm{L}} - \omega_{\mathrm{S}})$ is equal to the excitation energy of a Raman allowed transition (see the second factor in the denominator of eqn [12]). If in this case the laser beams are polarized according to the Raman selection rules then μ_{ijkl} is not zero. For spin transitions in a semiconductor in a magnetic field (free electron spin flip as well as transitions within the magnetic ions) this is the case if in the Voigt configuration ($\vec{k}_{\mathrm{L,S}} \perp \vec{B}$) the beams with frequencies ω_{L} and ω_{S} are polarized perpendicular to one another. In the case of a spin transition $\omega_{ca} = g^* \mu_{\mathrm{B}} B$ (g^* being the effective Landé-factor and μ_{B} the Bohr magneton). Thus holding the frequencies ω_{L} and ω_{S} fixed and tuning the magnetic field a resonance in I_{CARS} will be observed from the position of which the effective g-factor can be obtained. Using tunable dye lasers, of course, the laser frequencies can be tuned as well.

Close to the resonance the real part of the susceptibility becomes small and the imaginary part is not negligible. Furthermore, one-photon resonances may occur if the laser frequency ω_{L} corresponds to the energy of a real electronic transition, e.g., an allowed interband transition. If one-photon and Raman-like resonances occur simultaneously one talks about resonance CARS.

As discussed in the literature $\chi^{(3)}$ can be split into a resonant part χ_{r}, summing up all the resonant terms, and a nonresonant part χ_{nr}. In a semiconductor there are contributions to χ_{nr} due to free carriers and a background due to bound valence electrons. The background contribution in general is real. As can be seen from eqn [10],

$$I_{\mathrm{CARS}} \propto \left|\chi^{(3)}_{\mathrm{CARS}}\right|^2 = |\chi_{\mathrm{nr}}|^2 + |\chi'|^2 + |\chi''|^2 + 2 \cdot \chi_{\mathrm{nr}}\chi' \qquad [13]$$

χ' and χ'' being the real and imaginary parts of the resonant contribution. Depending on the relative strength of resonant and nonresonant contributions to $\chi^{(3)}$ the last term in eqn [13] may cause complicated lineshapes which make it difficult to find the correct transition energy in spectroscopic applications of CARS. If more than one resonance is not completely resolved the lineshape is even more complicated and a careful analysis has to be performed to get the correct resonance positions and linewidths of the transitions.

Examples of Experimental Results

Figure 2 shows a scheme of a CARS setup suitable to measure spin transitions in a semiconductor. The scheme is drawn for backscattering geometry; forward detection is accomplished by directing the incoming beams onto the sample from the opposite side to that plotted.

Two dye lasers (frequencies ω_{L} and ω_{S}) are pumped by a frequency-doubled Nd:YAG laser. A Glan–Thomson prism combines the two beams. The monochromator is set to $2\omega_{\mathrm{L}} - \omega_{\mathrm{S}}$. The iris and the polarizer PZ3 reduce stray light. The experiment is performed either at fixed magnetic field and tuning one of the dye lasers (with the monochromator swept accordingly) or by keeping the laser frequencies fixed and sweeping the magnetic field. The latter method is also applicable in the mid-infrared using step-tunable CO or CO_2 lasers.

As an example, in Figure 3 a spectrum obtained from a CdMnTe alloy crystal is reproduced. There are two peaks observed due to the paramagnetic resonance of the Mn 3d electrons and due to the spin resonance of photo-excited free electrons in the conduction band. The inset shows the paramagnetic resonance on an extended scale. The linewidth of 0.5 cm^{-1} is not limited by the experimental resolution, which in the present case is about a factor of 10 better.

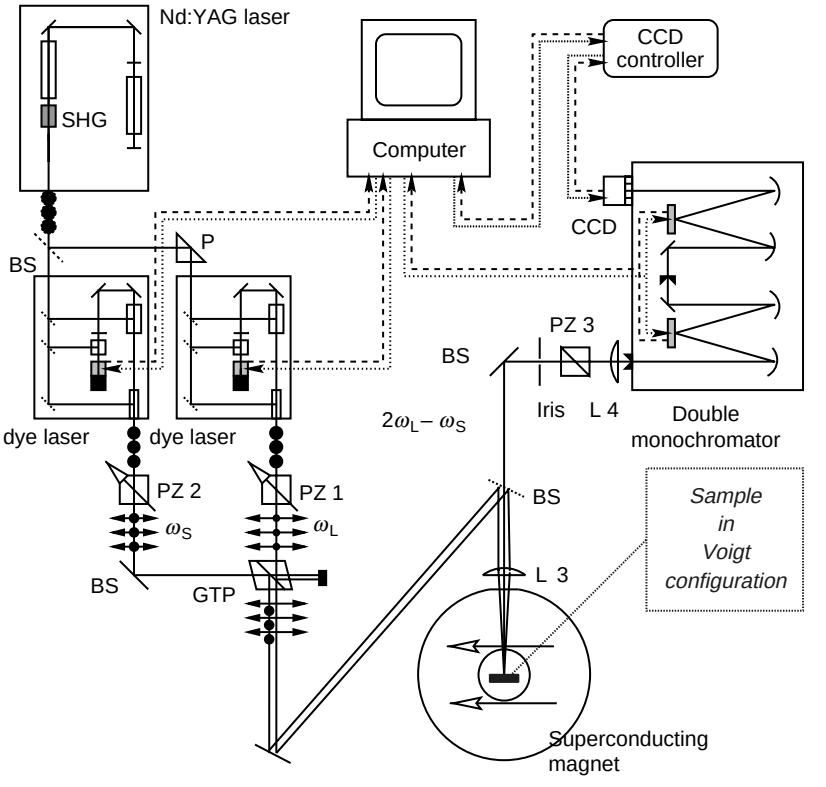

Figure 2 Scheme of the experimental arrangement for CARS. Bold full lines: light beams; dashed and dotted lines: data connections; BS: beam splitter; PZ1,2: Pockels cells for rotation of linear polarization; PZ3: polarizer; GTP: Glan–Thomson prism for superposition of beams polarized orthogonally to each other; L: lenses.

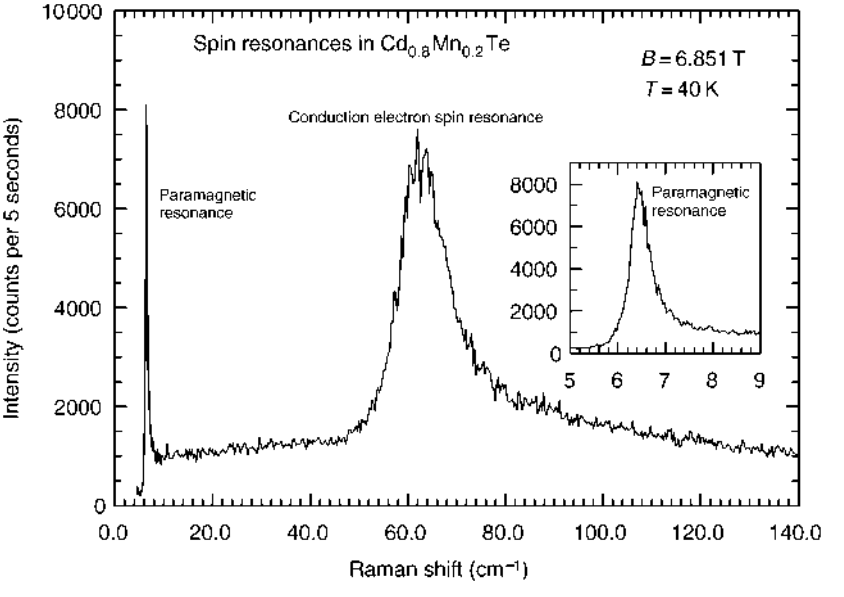

Figure 3 Coherent Raman intensity of a CdMnTe bulk crystal as a function of frequency difference $\omega_L - \omega_S$. Inset: Region of paramagnetic resonance in an extended scale. Adapted from Rupprecht R, Müller B, Pascher H, *et al.* (1998) Coherent Raman scattering from magnetic excitations in diluted magnetic semiconductors: Bulk crystals of CdMnTe. *Physical Review B* 58: 16123. Copyright (1998) by the American Physical Society.

Optical Pumping

The basic principles of optical pumping have been developed by Kastler. A material (gas or solid) is excited by absorption of circularly polarized light. Lampel was the first to demonstrate the possibility of optical spin orientation of free electrons in a semiconductor. The method has the most important applications in zincblende materials. Thus it is explained here with the example of GaAs.

Principles of Optical Spin Orientation

Most of the III–V and II–VI semiconductors have direct gaps at the center of the Brillouin zone. The conduction band (cb) is doubly degenerate with spin. The valence band consists of the heavy hole band (hh; $m_j = \pm 3/2$), the light hole band (lh; $m_j = \pm 1/2$) and the split-off band (so; $m_j = \pm 1/2$), each doubly degenerate. In bulk material hh and lh are degenerate at $k = 0$. The transition probabilities for interband transitions between hh and cb are three times larger than for transitions between lh and cb. Due to the selection rules, e.g., for the absorption of σ^+ circularly polarized light, transitions $-3/2 \rightarrow -1/2$ and $-1/2 \rightarrow +1/2$ are allowed, the first having three times higher probability. The electronic spin polarization P is defined as:

$$P = \frac{n_+ - n_-}{n_+ + n_-} \qquad [14]$$

where n_+ and n_- are the densities of electrons with the two spin orientations with respect to the direction of propagation of light. These considerations yield $P_\mathrm{i} = -0.5$ as the initial spin polarization under excitation with σ^+ circularly polarized light with a photon energy between the fundamental gap and the energy difference from the spin–orbit split valence band to the conduction band. Due to relaxation processes the actual electronic steady state polarization P is smaller than P_i:

$$P = P_\mathrm{i} \frac{T_1}{T_1 + \tau} \qquad [15]$$

with T_1 being the spin relaxation time, which in zero magnetic field is equal to the transverse relaxation time T_2, and τ is is the lifetime of conduction electrons.

The spin polarization of the electrons can be detected by analyzing the degree of circular polarization of the luminescence light which is defined as:

$$\Pi = \frac{I_+ - I_-}{I_+ + I_-} \qquad [16]$$

where I_+ and I_- are the intensities of σ^+ and σ^- polarized light, respectively. Since for emission due to free-electron/free-hole or free-exciton recombination the same transition probabilities as for absorption are obtained, one finds:

$$\Pi = 0.5\,P = 0.25\,\frac{T_1}{T_1 + \tau} \qquad [17]$$

In the case of quantum wells, the lh-hh degeneracy is lifted due to different confinement energies. Then the maximum polarization of cb $\rightarrow$ hh as well as for cb $\rightarrow$ lh luminescence light is 0.5.

Of course these considerations are only true for intrinsic or p-type materials where the number of photo-excited electrons is high compared to the equilibrium number of electrons.

Hyperfine Interaction of Spin-Polarized Electrons and Atomic Nuclei

In GaAs three isotopes exist with spin 3/2, namely ^{69}Ga, ^{71}Ga and ^{75}As. Owing to the small energy separation of the nuclear spin levels in moderate magnetic fields, in thermal equilibrium no significant nuclear spin polarization exists even at a temperature of 2 K. However, the hyperfine interaction between electrons and nuclei allows so-called flip-flop processes, that is a simultaneous reversal of electronic and nuclear spins. This causes a considerable dynamic spin polarization of the nuclei if the electronic spin system is polarized due to optical pumping. Under most experimental conditions and when nuclear spin relaxation mechanisms other than via the hyperfine interaction are negligible, the mean nuclear spin is approximately:

$$\langle I \rangle = \frac{I(I+1)}{S(S+1)}\,\frac{P}{2} \qquad [18]$$

where I and S are the operators for nuclear and electron spin. These dynamically polarized nuclei act on the electrons and shift their spin resonance magnetic field by an effective nuclear field (the Overhauser shift). This shift is proportional to $\langle I \rangle$ and to the probability of finding the electrons at the nuclear sites, and is inversely proportional to g^*. In GaAs due to the small value of g^* this field can reach several tesla.

Effect of a Constant Transverse Magnetic Field – The Hanle Effect

In a transverse (with respect to the direction of propagation of light) magnetic field the spin-oriented electrons precess around the field with an

angular frequency $\Omega = g^* \mu_B B/\hbar$. If the spin lifetime T_S ($1/T_S = 1/T_2 + 1/\tau$) is long enough, this precession diminishes the average spin polarization and consequently the degree of circular polarization of the photoluminescence according to:

$$\Pi(B) = \Pi(0)\frac{1}{1+\Omega^2 T_S^2} \qquad [19]$$

This is a Lorentzian with a width at half height of $\Delta B = \hbar/(g^* \mu_B T_S)$.

It is seen that by measuring $\Pi(B)$ the spin lifetime can be determined if g^* is known. Together with eqn [17] the spin and interband lifetime can be determined in a steady-state experiment. If, on the other hand, T_S is known from a time-resolved experiment, g^* can be found from a Hanle effect measurement.

Optically Detected Conduction Electron Spin Resonance (ODCESR)

Optical spin orientation of conduction electrons is the presupposition for ODCESR. Suppose now, a longitudinal magnetic field $\mathbf{B}_0$ is applied on the sample ($\mathbf{B}_0$ parallel to $\mathbf{k}$, the direction of propagation of light, which is parallel to the direction of the optically oriented spins). If the field is not too high the considerations of optical spin orientation remain unchanged. In higher fields the spin polarization may be modified due to the energy splitting of the spin levels. In this case the spin polarization is enhanced or diminished, depending on the direction of the field parallel or antiparallel to $\mathbf{k}$. If now an additional transverse microwave field B_1 of frequency ω_0, which fulfills the resonance condition $\hbar\omega_0 = g^* \mu_B B_0$, is applied, transitions between the two spin levels will occur which resonantly decrease the spin polarization. This decrease manifests itself in a decrease of the degree of circular polarization of the photoluminescence.

The method has several advantages compared to conventional ESR. Microwave photons with frequency of about 10^{10} Hz change the state of polarization of optical photons with frequency of about 10^{14} Hz. Thus the sensitivity is that of the optical detection which is much higher than in conventional ESR. Even electrons in a single quantum well or heterostructure can be detected by the optical method. There is also the possibility to distinguish between electrons in different quantized levels in heterostructures due to the different spectral position of their photoluminescence. Another advantage is that the electrons are photo-excited; the sample does not need to be n-doped. High-purity samples can be investigated which exhibit narrower linewidths.

Optically Detected Nuclear Magnetic Resonance (ODNMR)

With the presuppositions provided in the preceeding sections the basic principle of ODNMR is easily understood. As in ODCESR we assume that an external magnetic field $\mathbf{B}_0$ is applied parallel to $\mathbf{k}$. Perpendicular to it a radio-frequency field $\mathbf{B}_2$ with frequency ω is applied. The NMR resonance frequency is $\omega_0 = \gamma_n \cdot B_0$ with γ_n being the gyromagnetic ratio of the nucleus under consideration. In the frame rotating with ω around the direction of $\mathbf{B}_0$ an effective field $\mathbf{B}_{\text{eff}} = \mathbf{B}_0 - \boldsymbol{\omega}/\gamma_n + \mathbf{B}_2$ acts on the nuclei. The nuclear magnetization is oriented parallel to $\mathbf{B}_{\text{eff}}$. It is seen that in resonance $\mathbf{B}_{\text{eff}}$ is parallel to $\mathbf{B}_2$, which means perpendicular to $\mathbf{k}$. Via the hyperfine interaction this nuclear field acts on the electrons, diminishing their spin polarization via the Hanle effect. So the nuclear magnetic resonance can be detected as a reduction of the degree of circular polarization of the photoluminescence.

The simplified model given above does not take into account several effects which influence the lineshape, position and intensity of the observed resonance lines.

Examples of Experimental Results for Optical Pumping and Optically Detected Magnetic Resonances

These experiments can be performed at liquid helium temperature in a standard photoluminescence setup. In addition one has to provide the possibility of applying transverse (Hanle effect) or longitudinal (ODCESR, ODNMR) magnetic fields. For ODCESR the sample is mounted in a microwave resonator with optical access; for ODNMR a radio-frequency coil is wound around the sample. The state of circular polarization of the exciting laser light has to be controlled. The measured quantity is the sign and degree of circular polarization of the luminescence light at a certain wavelength. The sensitivity can be enhanced by periodically σ^+/σ^--modulating either the exciting light or the detection. Which one of these possibilities is used depends on whether or not a high nuclear spin polarization is wanted.

As pointed out above the dynamically polarized nuclei, particularly in GaAs, strongly affect the

ODCESR resonant field. To get reliable values for the effective g-factor, the nuclear spin polarization must be completely suppressed. This can be achieved by σ^+/σ^--modulation of the exciting light with a modulation period much shorter than the nuclear spin relaxation time. But due to imperfections, e.g., if the intensities of σ^+ and σ^- light or their time intervals are not equal, residual optical fields may remain. To get precise values for g^* in addition the NMR has to be saturated by a suitable radio-frequency field. On the other hand, the strong Overhauser shift of the CESR can be useful for the determination of nuclear spin relaxation times. To perform such experiments the nuclei are first polarized by optical pumping with fixed circular polarization and afterwards, for the detection, σ^+/σ^--modulation is switched on.

For the optical detection of NMR a high nuclear spin polarization is favorable. Thus for those experiments one uses fixed polarization of the exciting beam and modulates the detected light.

Figure 4 shows examples of results obtained with a p-channel AlGaAs/GaAs single heterostructure. In this sample due to the doping profile a triangular potential well at the interface in the valence band of the GaAs layer is built, in which quantized hole levels exist. The luminescence band shown in Figure 4a between 821 nm and 825 nm is due to recombination of electrons with holes in these quantized states. For the experiments shown in the other panels of the figure the spectrometer is set at a position within this band. The other lines are due to excitons and impurities and are not interesting for what follows.

In Figure 4b the results of a Hanle experiment are reproduced. The line without noise is the fitted Lorentzian. The fit yielded $\tau = 8.1$ ns and $T_1 = 4.7$ ns.

In Figure 4c a typical ODNMR spectrum of the same sample is plotted. One observes resonances due to ^{75}As at 1.36 MHz, ^{69}Ga at 1.9 MHz and ^{71}Ga at 2.4 MHz. In the inset the ^{75}As resonance is plotted in an extended scale. A quadrupole splitting into three components is clearly resolved. This result demonstrates that the tetrahedral symmetry at the nuclear site is broken either due to local defects or due to the (rather small) lattice mismatch close to the AlGaAs/GaAs interface.

Finally Figure 4d shows the ODCESR spectrum. For the experiment the exciting beam was σ^+/σ^--modulated and in addition during the sweep the NMR of all three nuclei was permanently saturated to make sure that no Overhauser shift

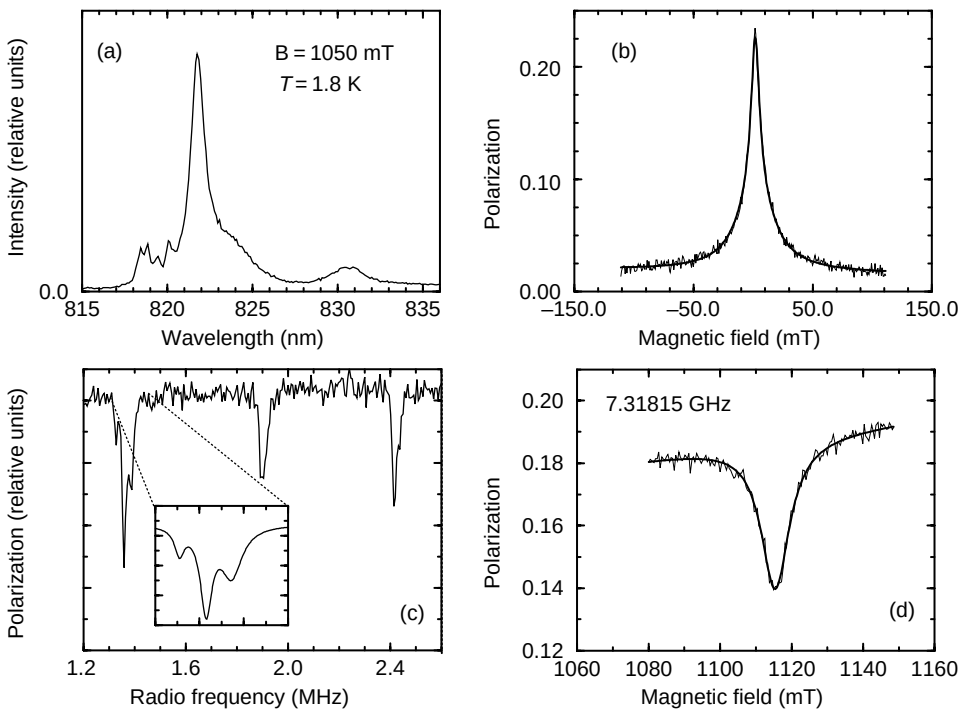

Figure 4 (a) Photoluminescence spectrum of a p-channel AlGaAs/GaAs heterostructure; (b) Hanle curve (luminescence polarization as a function of transverse magnetic field); (c) ODNMR recording; (d) ODCESR recording. Adapted from Schreiner M, Pascher H, Denninger G, *et al.* (1997) Nuclear spin relaxation in AlGaAs/GaAs heterostructures observed via optically detected magnetic resonance (ODMR) experiments. *Solid State Communications* 102: 715. Copyright (1997), with permission from Elsevier.

affects the result. The g-factor calculated from this plot is $|g^*| = 0.468$, which is larger than that found for bulk GaAs. This value was also used for fitting the Hanle curve.

List of Units and Nomenclature

Angle	[rad]
Angular frequency ω	[rad s^{-1}]
Bohr magneton μ_B	9.2742×10^{-24} J T^{-1}
Degree of circular polarization of light Π	
Electric field E	[V m^{-1}]
Gyromagnetic ratio γ	[T^{-1} s^{-1}]
Imaginary unit i	
Intensity I	[W m^{-2}]
Magnetic displacement B	[T]
Permittivity of free space ε_0	8.8542×10^{12} V s A^{-1} m^{-1}
Planck's constant $\hbar$	1.0546×10^{-34} J s
Refractive index n	
Speed of light in vacuum c_o	2.9979×10^{8} m s^{-1}
Spin polarization P	
Susceptibility χ	

CARS	coherent anti-Stokes Raman scattering
CSRS	coherent Stokes Raman scattering
ODCESR	optically detected conduction electron spin resonance
ODMR	optically detected magnetic resonance
ODNMR	optically detected nuclear magnetic resonance

Acknowledgments

The author thanks W. Herbst, B. Müller, R. Rupprecht and M. Schreiner for making unpublished results available.

See also

Magneto-Optics: Interband Magnetoabsorption, Cyclotron Resonance, Spin Flip Raman Scattering. **Semiconductor Physics:** Light Scattering; Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Boswarva IM, Howard RE and Lidiard AB (1964) Faraday effect in semiconductors. *Proceedings of the Royal Society* A269: 125–141.

Dyakonov MI and Perel VI (1984) Theory of optical spin orientation of electrons and nuclei in semiconductors. In: Meier F and Zakharchenya BP (eds) *Optical Orientation*, pp. 11–71. Amsterdam: North-Holland.

Eesley GL (1981) *Coherent Raman Spectroscopy*. Oxford: Pergamon Press.

Furdyna JK (1988) Diluted magnetic semiconductors. *Journal of Applied Physics* 64: R29–R64.

Geist F and Pascher H (1996) Resonance CARS in narrow gap semiconductors. *Journal of Raman Spectroscopy* 27: 289–295.

Geist F, Herbst W, Mejia-Garcia C, *et al.* (1997) Magneto-optical investigations of Eu-based diluted magnetic lead chalcogenides. *Physical Review B* 56: 13042.

Hermann C and Weisbuch C (1984) Optical detection of conduction electron spin resonance in semiconductors and its application to $\mathbf{k}\cdot\mathbf{p}$ perturbation theory. In: Meier F and Zakharchenya BP (eds) *Optical Orientation*, pp. 463–508. Amsterdam: North-Holland.

Krenn H, Herbst W, Pascher H, *et al.* (1999) Interband Faraday and Kerr rotation and magnetization of $Pb_{1-x}Eu_xTe$ in the concentration range $0 \times \sim 1$. *Physical Review B* 60: 8117.

Levenson MD and Kano SS (1987) *Introduction to Nonlinear Laser Spectroscopy*. San Diego: Academic Press.

Nies R and Kessler FR (1989) Coherent and incoherent magnetooptical reflection and transmission effects of multilayer structures. *Phys. Stat. Sol. (a)* 111: 639–654.

Oh E, Bartholomew DU, Ramdas AK, Furdyna JK and Debeska U (1988) Interband Faraday effect in CdMnSe. *Physical Review B* 38(18): 13183–13190.

Paget D and Berkovits VL (1984) Optical investigation of hyperfine coupling between electronic and nuclear spins. In: Meier F and Zakharchenya BP (eds) *Optical Orientation*, pp. 404ff. Amsterdam: North-Holland.

Pascher H, Geist F, Springholz G, *et al.* (1997) Coherent anti-Stokes Raman scattering in diluted magnetic IV–VI epilayers and superlattices. In: Shen SC, Tang DY, Zheng GZ and Bauer G (eds) *Narrow Gap Semiconductors*, pp. 269–278. Singapore: World Scientific.

Pohlt M, Herbst W, Pascher H, Faschinger W and Bauer G (1998) Magneto-optical Kerr effect in short-period CdTe/MnTe superlattices. *Physical Review B* 57(16): 9988–9994.

Ramdas AK and Rodriguez S (1988) Raman scattering in diluted magnetic semiconductors. In: Furdyna JK and Kossut J (eds) *Diluted Magnetic Semiconductors, Semiconductors and Semimetals*, vol. 25, pp. 345–412. London: Academic Press.

Schreiner M, Pascher H, Denninger G, *et al.* (1997) Nuclear spin relaxation in AlGaAs/GaAs heterostructures observed via optically detected resonance (ODMR) experiments. *Solid State Communications* 102: 715–720.

Interband Magnetoabsorption, Cyclotron Resonance, Spin Flip Raman Scattering

C R Pidgeon, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The addition of a magnetic field to an optical study of a semiconductor generally improves the specificity of the experiment, but in particular affects primarily the electrons rather than the lattice so that the magneto-optical properties are determined to a large degree by the electrons alone. The major interest has been in the determination of the electronic energy band structure of semiconductors, both in their bulk and low-dimensional (two-, one-, and zero-dimensional) quantum forms. The magneto-optical experiments fall mainly into two categories according to whether the effects are associated with intraband (within a single conduction or valence band) or interband (valence to conduction band) transitions; there are also hybrid (i.e., higher-order) effects such as spin-flip Raman scattering or two-photon absorption where both inter- and intraband transitions are involved in a two-stage process. Interband transitions occur typically in the visible to mid-infrared (MIR) region, and intraband transitions in the MIR to far-infrared (FIR) region.

Theoretical Background

The most common band structure involved in semiconductor magneto-optics is the so-called quasi-Ge model, which gives an accurate description of group IV, and zincblende (III–V and II–VI) semiconductors, so it is convenient to describe the theory in terms of this model. In this model the small terms associated with inversion asymmetry are neglected to a first approximation. At the Brillouin zone center (i.e., at $k = 0$) the conduction band has s-like symmetry and the triply degenerate valence band p-like symmetry. In the presence of strong spin–orbit coupling, part of this degeneracy is raised resulting in the doubly degenerate (at $k = 0$) heavy and light mass valence bands and a so-called spin–orbit split-off band. The reader should go to the general reviews quoted under 'Further Reading' for the equivalent coverage of other common systems such as the wurzite II–VI and the lead salt systems.

The dynamics of a free electron subject to an external magnetic field have been given by a variety of authors, but the most universally used treatment is that of Landau. For the motion of an electron in a semiconductor we treat first the case of a simple parabolic energy band. For the case of the magnetic field, B, along the z-direction, k_z remains a good quantum number, but k_x and k_y are subsumed into the new magnetic orbital (Landau) quantum number, n. The original three-dimensional energy band, $E = \hbar k^2/2m^*$, is coelesced into a series of one-dimensional Landau sub-bands in the k_z-direction (where the motion along k_z remains unchanged):

$$E_n^{\pm}(k_z) = (n + 1/2)\,\hbar\omega_c + \frac{\hbar k_z^2}{2m^*} \qquad [1]$$

The cyclotron frequency is $\omega_c = eB/m^*$ and μ_B is the Bohr magneton. The resulting electron energy eigenvalues at $k_z = 0$ in an external magnetic field (i.e., Landau sub-band extrema) are referred to as Landau levels, LLs. In practice each state is doubly degenerate with spin which results in an additional term $\pm g^*\mu_B B$ on the right-hand side, where g^* is the effective g-factor. (Note that in the presence of strong spin–orbit coupling and small m^*, there is a corresponding g^* which becomes negative and substantially different from the free-electron g-factor; the largest value for a binary elemental compound is that for InSb, where $g^* \approx -50$.) The situation for a simple parabolic conduction band near $k = 0$ is shown in Figure 1a,b. Intraband optical absorption associated with direct transitions between successive Landau sub-bands is referred to as cyclotron resonance (CR) absorption. There are corresponding Landau sub-bands in the valence band, and direct transitions between valence and conduction sub-bands give rise to the so-called interband magneto-optical (IMO) effects.

Expressions for the density of states (DOS) in three, two, and one dimensions in the absence of a magnetic field are given elsewhere ('semiconductor basics' – Pidgeon). In the presence of a magnetic field the three-dimensional energy band is reduced to one-dimensional sub-bands so that the energy dependence of the DOS has the one-dimensional form with the energy given by eqn [1]. Performing this substitution and multiplying by a term for the degeneracy of the Landau levels (LLs) gives the following expression for the total DOS per unit volume:

$$g(E) = \frac{(2m^*)^{1/2}}{4\pi^2\hbar}\,\frac{eB}{\hbar}\sum_n \left[E - (n + 1/2)\,\hbar\omega_c\right]^{-1/2} \qquad [2]$$

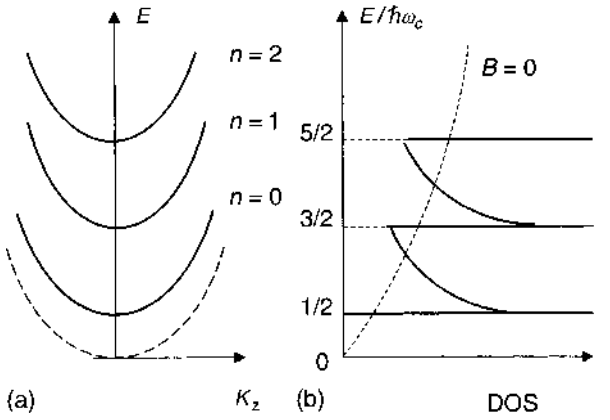

Figure 1 (a) Landau sub-bands according to eqn [1] with the magnetic field applied in the z-direction. The dashed line shows the parabolic conduction band for k along k_z and $B = 0$. (b) Schematic outline of the density of states (DOS) for free electrons in a magnetic field. It comprises the sum of identical contributions from each Landau sub-band. The density of states with $B = 0$ is shown dashed.

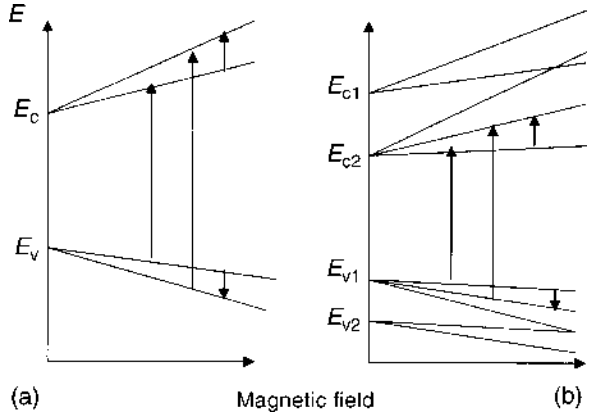

Figure 2 Fan plots of the conduction and valence band Landau levels versus magnetic field for simple parabolic bands: (a) a three-dimensional semiconductor; (b) quasi-two-dimensional systems. Some CR and interband magneto-optical transitions are shown by vertical arrows.

Thus, the DOS for Landau sub-bands is quite different from the zero-field DOS, and is a series of square-root singularities with divergences at the energies $(n + 1/2)\hbar\omega_c$. The two cases are illustrated schematically in Figure 1b. Since the optical absorption strength for both interband and intraband transitions is proportional to the joint density of states for the initial and final states, this result shows the important advantage of the use of a magnetic field in semiconductor spectroscopy because it means that the absorption lines for CR or IMO transitions are strongly peaked at the sub-band extrema (i.e., at $k_z = 0$). This allows accurate measurement of LL energy differences for transitions observed in a single experiment, and hence, by comparison to theoretical models, leads to the best determination of fundamental energy-band parameters. The basic LL selection rule can be shown straightforwardly from time-dependent perturbation theory to be $\Delta n = 0$, $\Delta m_s = 0$ for IMO transitions and $\Delta n = 1$, $\Delta m_s = 0$ for CR absorption, where m_s is the spin quantum number (i.e., $\pm\frac{1}{2}$, for spin up and spin down). The cyclotron resonance absorption occurs only for one sense of circularly polarized light, but the IMO effect can occur for all polarizations. The choice then of magnetic field direction, crystal orientation, and polarization of the optical radiation allows extremely detailed determination of the energy band parameters for particular crystal symmetries.

The data are often summarized in a so-called fan diagram of the energy levels, or transition energies, versus magnetic field. Such a plot is shown for hypothetical conduction and valence band LLs in Figure 2a, together with interband and CR transitions. The states are labeled by the Landau quantum number, with selection rules $\Delta n = 0$ and $\Delta n = 1$ for interband and CR absorption transitions, respectively (or $\Delta n = -1$ for CR emission). If the energies of the transitions themselves are plotted versus the magnetic field, then the intercept of the fan plots at $B = 0$ gives the zero-field energy gap. Note that in the presence of excitonic effects (see below) the intercept will give the energy gap plus the exciton binding energy.

For a strictly two-dimensional system the DOS per unit area is independent of energy: $g(E) = m^*/\pi\hbar^2$. In the case of quasi two-dimensional confinement in the z-direction the motion in the z-direction is quantized, and the total energy becomes (*see* **Semiconductor Physics:** Quantum Wells and GaAs-Based Structures):

$$E = E_i + \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2) \qquad [3]$$

Equation [3] describes a set of confinement sub-bands (parabolas induced by geometric, as opposed to magnetic, confinement), the bottoms of which are at the energies E_i, where i is the sub-band index. The energies, E_i, depend on the shape of the confining potential. The DOS per unit area is in this case a series of step functions of height $m^*/\pi\hbar^2$. From the point of view of magneto-optics, the main effect of the quantum confinement is to suppress the k_z motion, so that in the presence of an external magnetic field along the z-direction, the motion is completely quantized. The energy eigenvalues are given by:

$$E_{n,i} = E_i + (n + 1/2)\,\hbar\omega_c \qquad [4]$$

The total DOS becomes:

$$g(E) = \frac{m^* \omega_c}{\pi \hbar} \delta[E - (n + 1/2)\,\hbar\omega_c] \qquad [5]$$

a series of delta-functions located at the energy positions of the LLs. The DOS is obviously more singular in two dimensions than in three dimensions, and leads to substantially sharper magneto-optical and magneto-transport effects. For confined structures the selection rules for interband and intersub-band transitions are identical to the bulk case for the electromagnetic (em) radiation polarized in the xy plane provided that the external magnetic field is along the confinement (z-) direction. The only difference is that the conservation of momentum in the z-direction ($\Delta k_z = 0$) is replaced by a selection rule on the sub-band index, $\Delta i = 0$. A fan plot for this configuration is shown schematically in Figure 2b. For each confinement sub-band transition there is a set of LL transitions. When the energy positions are plotted against B, this leads to a set of Landau 'fans', each of which extrapolates to the sum of a valence plus conduction confinement sub-band energy at $B = 0$. Allowed transitions are shown for the lowest two confinement sub-bands in Figure 2b. Again this picture may be complicated in the presence of excitonic effects at low temperature.

Turning again to the bulk semiconductor picture, eqns [1]–[5] apply for a single parabolic energy band and are valid for the conduction band in the quasi-Ge model. In actual semiconductors further complications arise. Most importantly the triply degenerate p-like valence band (heavy hole, light hole, and split-off band) must be treated by so-called degenerate perturbation theory. In the quasi-Ge model the conduction and three valence bands are treated together as one coupled band system using the so-called $\mathbf{k} \cdot \mathbf{p}$ method (described elsewhere by Zawdzki). With spin this results in an 8×8 matrix equation which has to be diagonalized for the eigenvalues and eigenvectors of the system. The spin–orbit coupling results in so-called complex, mixed-spin valence bands and anomalously spaced low quantum number LLs (the so-called quantum effects). The overall energy band scheme for InSb, near the fundamental direct gap in the center of the Brillouin zone for conduction and (heavy and light mass) valence bands is shown schematically in Figure 3, for no magnetic field and for an applied magnetic field. Each LL is doubly degenerate with spin, the a-set representing the so-called spin-up and the b-set the spin-down states.

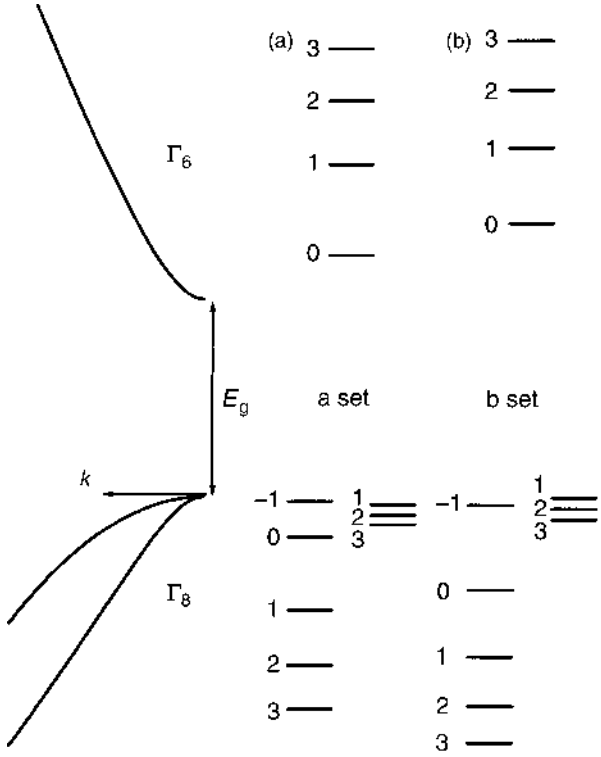

Figure 3 (a) Schematic diagram of the conduction and valence bands for InSb (symmetry labels Γ_6 and Γ_8 respectively) at $B = 0$. (b) Computed Landau levels for the conduction and valence bands of InSb at 4 K ($E_g = 0.23$ eV) with $B = 5$ T along the [100] direction, grouped according to the so-called a- and b-spin sets (see text).

Experimental

The first serious use of magneto-optics to study effective mass and energy band structure of semiconductors was in the microwave cyclotron resonance (i.e., intraband) investigations of Ge and Si in the early 1950s. Classically it can be shown that the ratio of the radiation power absorbed, P, to power absorbed in zero magnetic field, P_0, is given by:

$$\frac{P}{P_0} = \frac{1 + (\omega^2 + \omega_c^2)\tau^2}{[1 + (\omega_c^2 - \omega^2)\tau^2]^2 + 4\omega^2\tau^2} \qquad [6]$$

where τ is the collision time of the carriers. This expression for P/P_0 describes cyclotron resonance absorption and is plotted as a function of frequency in Figure 4 for different values of the parameter $\omega_c\tau$. Clearly the peak absorption occurs at $\omega = \omega_c$, namely at cyclotron resonance. Also evident is the very important requirement for observation of a clear resonance that $\omega_c\tau > 1$; this condition corresponds physically to a travel time between collisions of the particle of at least $1/2\pi$ of a revolution so that the particle can interact for a sufficiently long time

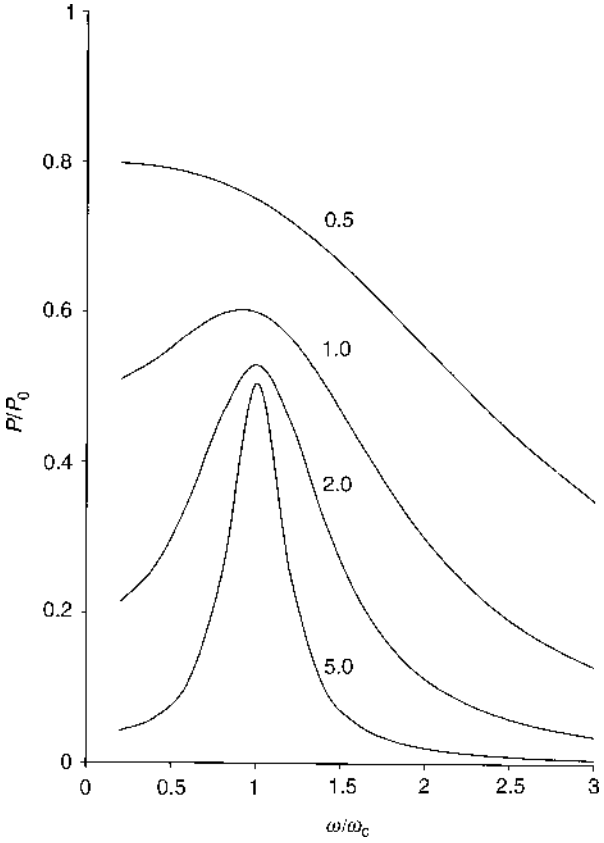

Figure 4 Ratio of the radiation power absorbed to the power absorbed at zero magnetic field, P/P_0, plotted as a function of ω/ω_c according to eqn [6], for different values of $\omega_c\tau$ from 0.5 to 5.0.

with the electromagnetic wave to absorb energy. For pure materials and low temperatures (where the condition $\omega_c\tau > 1$ is achievable) this is the best technique for obtaining the details of the electron and hole band parameters of a semiconductor near $k = 0$. The most thorough and successful studies of this type have been achieved in pure Si, Ge, and InSb where (with the aid of optical excitation to populate the levels shown in Figure 3) all possible electron, light and heavy hole cyclotron resonances have been resolved. In this way the valence band quantum effects, described above, have been observed in detail. The technique is equally applicable in low-dimensional systems where CR of both electrons and holes has been resolved in a large number of n- and p-type semiconductor heterostructures and quantum wells.

In interband – either absorption or photoluminescence (PL) – spectroscopy at low temperatures, the photo-excited electrons and holes interact and form excitons (*see* **Semiconductor Physics:** Excitons). In many cases the excitonic state may be described in the so-called envelope function approximation as a hydrogenic system, in which the electron and hole are bound by their mutual Coulomb attraction. The energy spectrum of such a hydrogenic exciton is given by

$$E_n = \frac{\mu e^4}{2\hbar^2\varepsilon^2 n^2} \qquad [7]$$

where n is the principal quantum number ($n = 1, 2, 3, \ldots$, but note that the symbol n is also used for the Landau quantum number above), μ is the electron–hole reduced effective mass, and ε is the background dielectric constant. In low-dimensional (i.e., confined) semiconductor systems of the type described above the degeneracy of the valence bands at $k = 0$ is removed, and two types of excitons are observed; these are referred to as 'heavy' and 'light' hole excitons in the literature, where heavy and light refer to the effective masses along the confinement direction. These excitons can move freely (in the remaining two unconfined directions) or be bound to an impurity or defect. In these systems exciton absorption may be observed involving not only the ground-state excitons, but also excitons associated with higher confinement sub-bands. This is shown for the so-called asymmetric quantum well system $Al_yGa_{1-y}As/In_xGa_{1-x}As/GaAs$ in Figure 5a. The data are taken both by the direct PL method (where the exciting beam has an energy much larger than the bandgap and the electrons and holes de-excite instantaneously to the bottom of the bands before recombining as exiton recombination radiation), and so-called PL excitation spectroscopy (where the exciting beam is a tunable laser, tuned into resonance with the successive exciton resonances).

The interband Landau transitions can be observed either in PL or as excitations in a magneto-absorption or magnetoreflection experiment. Excitons are affected differently by the application of an external magnetic field depending whether the ratio $\gamma = \hbar\omega_c/2E_B$ (where E_B is the exciton binding energy with $n = 1$ above) is less than or greater than one. In the so-called magnetic limit where $\gamma \gg 1$, the excitonic ground state moves approximately as $\frac{1}{2}\hbar\omega_c$, i.e., linearly with B. In this regime the magnetic field energy is the dominant term in the exciton Hamiltonian. The various states of the exciton (ground and excited states) become associated with the interband LL transition, and exhibit an energy dependence that is nearly identical to the free carrier LL transitions. The energies of the various excitonic transitions are shifted down with respect to the corresponding LL transitions by an amount equal to the binding energy of the excitonic state to that LL. Typical PL and PLE (photoluminescence excitation)

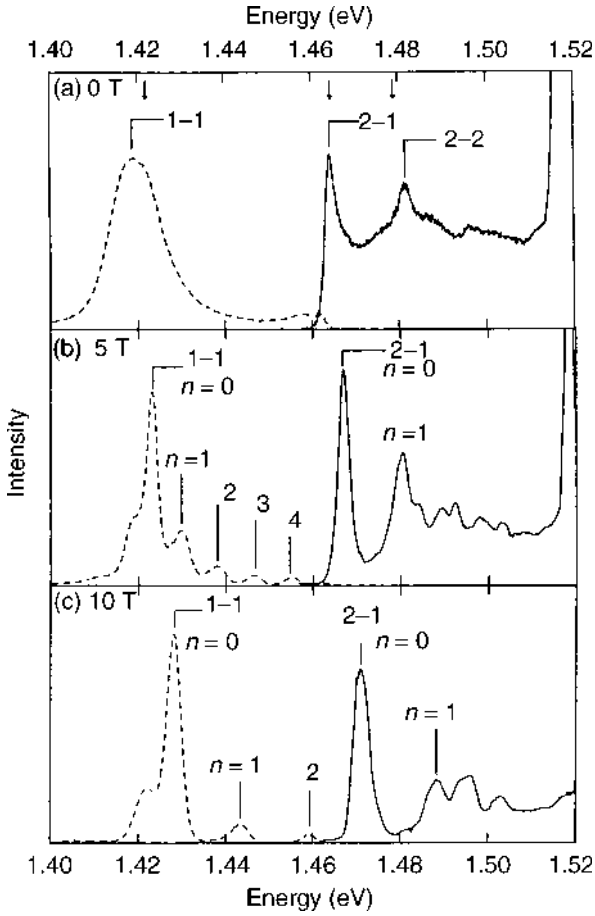

Figure 5 Interband transition data obtained in zero magnetic field (a) and fields of 5 T (b) and 10 T (c) from asymmetric quantum wells. The dashed lines indicate data obtained from PL and the solid lines data from PLE. The arrows in (a) indicate the calculated interband transition energies $1 \rightarrow 1$, $2 \rightarrow 1$, and $2 \rightarrow 2$, the latter being clearly resolved in PLE (the asymmetric wells mean that transitions with $\Delta i \neq 0$ are allowed). In (b) and (c) the development of the spectra into LL transitions is clearly seen, although beyond 1.48 eV the analysis is obviously extremely complicated by the overlap of LL transitions associated with all three of the above sub-bands.

spectra for interband magneto-optical transitions are shown in Figure 5b,c for a fixed applied magnetic field in the asymmetric quantum well system $Al_yGa_{1-y}As/In_xGa_{1-x}As/GaAs$.

Spin Flip Raman Scattering

In addition to the interband transitions discussed above, Raman scattering of intersub-band transitions between conduction or valence band LLs can also be studied. It provides a versatile means of probing band edge parameters of electrons and holes in semiconductors since the excitation (normally a laser pump) is not confined to the resonance frequencies under investigation, i.e., any pumping energy below the band edge will do. Electronic Raman scattering is then allowed between LLs via virtual interband transitions. In the presence of spin–orbit coupling strong spin-flip processes via electric dipole transitions become allowed giving rise to the so-called spin-flip Raman (SFR) cross-section. This is a single-particle process in which electrons or holes interacting with the radiation change their spin state. This process can be made to go stimulated particularly easily when the pump radiation is resonant with the bandgap of the semiconductor, forming the spin-flip Raman laser. The most important practical case has probably been that of InSb, resonantly pumped near 5 μm with the CO laser. In this case the small effective mass of a narrow-gap semiconductor in the presence of substantial spin–orbit coupling leads to both an extremely large anomalous effective g-factor ($g^* \sim -50$) – giving a large tuning range, $\Delta E = \hbar\omega_{SF} = g^*\beta B$ – and a strong cross-section which is larger than that of the free electron by a factor $(m_0/m^*)^2$. Efficient cw operation is possible with threshold powers less than 50 mW.

SFR scattering has enabled the observation of a great variety of excitations in nonmagnetic and dilute magnetic semiconductors, including: free electrons and holes, electrons bound to donors and holes bound to acceptors, scattering from bound magnetic polarons, and scattering due to SF transitions within the Zeeman multiplets of paramagnetic ions. In addition to pure SF transitions, combined resonances with combinations such as $\omega_{CR} \pm \omega_{SF}$ have also been detected. The process has also been used in a similar manner to electron spin resonance to study coherent processes, including Raman photon echo spectroscopy in semiconductors.

See also

Magneto-Optics: Faraday Rotation, CARS, ODMR, ODSR, Optical Pumping. **Semiconductor Physics:** Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Häfele HG (1991) Spin-flip raman scattering. In: Landwehr G and Rashba WI (eds) *Landau Level Spectroscopy in Modern Problems in Condensed Matter Sciences*, 27.1, 207–275.

Petrou A and McCombe BD (1991) Magnetospectroscopy of confined semiconductor systems. In: Landwehra G and Rashba WI (eds) *Landau Level Spectroscopy in Modern Problems in Condensed Matter Sciences*, 27.1, 679–775.

Pidgeon CR (1980) Free carrier optical properties of semiconductors. In: Balkanski M (ed.) *Optical Properties of Solids. Handbook of Semiconductors*, pp. 223–328. Amsterdam, The Netherlands: North-Holland.

Zawadzki W (1991) Intraband and interband magneto-optical transitions in semiconductors. In: Landwehr G and Rashba WI (eds) *Landau Level Spectroscopy in Modern Problems in Condensed Matter Sciences*, 27.1, 483–512.

MATERIALS CHARACTERIZATION TECHNIQUES

Contents

$\chi^{(2)}$
$\chi^{(3)}$

$\chi^{(2)}$

R C Eckardt, Cleveland Crystals, Inc., Highland Heights, OH, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Techniques used to measure second-order susceptibility $\chi^{(2)}$ or nonlinear optical coefficients are described in this article, and some measured values are quoted. These are the parameters that describe the property of a material that is responsible for second-harmonic generation and other three-wave nonlinear optical interactions. These other interactions include sum- and difference-frequency generation in which two waves combine to generate a third wave of a frequency that is the sum or difference of the first two waves. Also included are optical parametric generation and optical parametric oscillation, in which a single pump wave generates two resulting waves with frequencies that sum to the frequency of the pump wave. The measurement techniques involve measuring the efficiency of nonlinear frequency conversion by these processes.

The measurements require a great deal of care and precision, and it is advisable to perform a comparison with previous measurements and obtain correlation with accepted standards whenever possible. A great deal of effort continues to be expended by researchers to establish correlations between measurements and provide additional values that are accepted as standards. A number of techniques have been used for these measurements sometimes yielding results that disagree. There are, however, values derived from early measurements that have proven accurate. In these measurements it is necessary to accurately control and characterize the polarization, power, spectral distribution, and spatial and temporal intensity distribution of the incident radiation and accurately measure the frequency-converted light, which can be many orders of magnitude smaller.

The development and availability of single-longitudinal-mode and nearly diffraction-limited lasers has greatly reduced the ambiguity and simplified increased accuracy of measurements of second-order nonlinear optical coefficients. The improved optical quality of nonlinear-optical crystals has also been an important component of the increased accuracy of measurement. Measurement techniques have also improved, providing absolute measurements of optical intensity over wide dynamic ranges at different wavelengths. Until the early 1990s, there was a range of measured values for nonlinear optical coefficients of common nonlinear materials, and numerous definitions were in use. Caution is required when consulting early publications on this topic, and some of this ambiguity is carried over to present time. Different definitions vary by factors of 2 and the permittivity of free space, $\varepsilon_0 \cong 8.854 \times 10^{-12}$ C/N $\times$ m^2. It is important that the definition of a nonlinear optical coefficient is clear before a value is used. This can be done through a statement of the nonlinear electric polarization or the coupled equations for nonlinear frequency conversion.

The standard techniques that have been used to measure nonlinear optical coefficients are covered in detail in publications listed in the Further Reading at the end of this article. Brief descriptions of some of these techniques, such as Maker-fringe and wedge,

optical parametric fluorescence, and phase-matched second-harmonic generation, will be presented in this article. Each of these techniques has a unique set of advantages and disadvantages. There was some lively controversy over the technique that provided the most accurate measurements and those measurements that should be dismissed. Measurements based on phase-matched second-harmonic generation had the advantage of being most similar to practical application of nonlinear optical frequency conversion. However, the large divergence and multiple-longitudinal-mode operation of early lasers made this method unreliable. Even relative comparisons of nonlinear optical coefficients between different materials were inaccurate because highly structured laser pulses had different conversion characteristics due to differences in bandwidth acceptance, group velocity walkoff, and birefringent walkoff in different materials.

Nonlinear optical coefficients of improved accuracy were required for analysis of optical harmonic conversion in the large, well-characterized laser systems used in inertial-confinement fusion research. The development of these systems provided a standardization on the value of $d_{36} = 0.39$ pm V^{-1} (1 pm $= 10^{-12}$ m) for second-harmonic generation of neodymium laser radiation in the material potassium dihydrogen phosphate (KH_2PO_4 or KDP). The development of injection-seeded Q-switched lasers provided single-longitudinal-mode laser output that allowed the phase-matched second-harmonic techniques to be extended to provide more accurate absolute measurements and relative comparisons on KDP and a number of other materials. Correlation was achieved on a larger number of relative and absolute measurements of nonlinear optical coefficients to provide some systematic agreement. It is not well understood why the earlier measurements referenced to optical parametric fluorescence produced values that were higher by about 60%. Measurement techniques have been further refined, and more exacting measurements are leading agreement, using different methods of measurement on an increasing number of nonlinear optical materials.

Definitions

The definition of the nonlinear optical coefficient can be specified in a number of ways. Examples given here are how the coefficient relates to the nonlinear electric polarization in a material, the coupled equations for nonlinear frequency conversion, and expressions describing second-harmonic generation with a monochromatic plane-wave fundamental.

The second-order optical nonlinearity is reported as both the second-order susceptibility $\chi^{(2)}$ and the second-order nonlinear optical coefficient d, which usually differ in definition by a factor of two. These are third-rank tensors that relate products of components of the electric field vector E_j and E_k to the vector components of the nonlinear polarization P_i^{NL}. The second-order susceptibility typically is used to express instantaneous values:

$$P_i^{NL}(t) = \varepsilon_0 \sum_{j,k} \chi_{ijk}^{(2)} E_j(t) E_k(t) \qquad [1]$$

whereas the nonlinear optical coefficient more often is used to relate the vector amplitudes of frequency components of these quantities. The electric field at the fundamental angular frequency ω can be expressed as

$$E_j(\omega, \mathbf{r}, t) = \frac{1}{2}\{E_j(\omega) \exp i(\mathbf{k}_\omega \cdot \mathbf{r} - \omega t) + c.c.\} \qquad [2]$$

and the nonlinear polarization at the harmonic frequency 2ω can be expressed as

$$P_i^{NL}(2\omega, \mathbf{r}, t) = \frac{1}{2}\left\{P_i^{NL}(2\omega) \exp i(2\mathbf{k}_\omega \cdot \mathbf{r} - 2\omega t) + c.c.\right\} \qquad [3]$$

The intensity is given by $I = (nc\varepsilon_0/2)|E|^2$. The relationship between the complex amplitudes expressed in terms on the nonlinear optical coefficient tensor is

$$P_i^{NL}(2\omega) = \varepsilon_0 \sum_{j,k} d_{ijk}(-2\omega; \omega, \omega) E_j(\omega) E_k(\omega) \qquad [4]$$

When waves of three different frequencies are involved, the expression becomes

$$P_i^{NL}(\omega_3) = 2\varepsilon_0 \sum_{j,k} d_{ijk}(-\omega_3; \omega_2, \omega_1) E_j(\omega_2) E_k(\omega_1) \qquad [5]$$

Expressing the coefficient as $d_{ijk}(-\omega_3; \omega_1, \omega_2)$ allows for dispersion and indicates that fields at frequencies ω_1 and ω_2 interact to produce a nonlinear polarization at frequency ω_3, the sum of the first two frequencies. The relationship between the second-order electrical susceptibility and the second-order nonlinear optical coefficient is

$$\chi_{ijk}^{(2)} = 2d_{ijk} \qquad [6]$$

The dispersion of nonlinear optical coefficients is approximated by the use of Miller's delta:

$$\delta_{ijk} = \frac{d_{ijk}(-\omega_3; \omega_2, \omega_1)}{(n_i^2(\omega_3) - 1)(n_j^2(\omega_2) - 1)(n_k^2(\omega_1) - 1)} \qquad [7]$$

The indices of refraction in the direction of the electric field components $E_m(\omega_n)$ are $n_m(\omega_n)$ for $m=i,j,k$ and $n=1,2,3$. There is some theoretical and experimental justification for treating the components of Miller's delta as constants. Earlier definitions of Miller's delta included multiplication by ε_0, the permittivity of free space.

The nonlinear optical coefficient is usually reduced to a scalar quantity, the effective nonlinear coefficient or d_{eff}, that relates only the specific vector components of the fields involved in the interaction. The effective nonlinear optical coefficient will be discussed in more detail later. The coupled equations for second-harmonic generation using d_{eff} are

$$\frac{\mathrm{d}E_{2\omega}(z)}{\mathrm{d}z} = \mathrm{i}\frac{\omega d_{\text{eff}}}{n_{2\omega}c}E_\omega(z)E_\omega(z)\exp(-\mathrm{i}\,\Delta kz) \quad [8]$$

$$\frac{\mathrm{d}E_{\omega}(z)}{\mathrm{d}z} = \mathrm{i}\frac{\omega d_{\text{eff}}}{n_{\omega}c}E_\omega(z)E^*_\omega(z)\exp(\mathrm{i}\,\Delta kz) \quad [9]$$

Here ω is the angular frequency of the fundamental wave; z is distance in the direction of propagation, $n_{2\omega}$ and n_ω are the refractive indices of the harmonic and fundamental respectively; and c is the speed of light in free space. The complex electric field amplitudes of the harmonic and fundamental, $E_{2\omega}(z)$ and $E_\omega(z)$, respectively, are assumed to be slowly varying functions of position compared with the oscillation of the carrier waves $\exp \mathrm{i}(\mathbf{k}_{2\omega}\cdot\mathbf{r} - 2\omega t)$ and $\exp \mathrm{i}(\mathbf{k}_\omega\cdot\mathbf{r} - \omega t)$, with $\mathbf{k}_{2\omega}$ and $\mathbf{k}_\omega$ the wavevectors of harmonic and fundamental; the asterisk indicates complex conjugate. The quantity Δk is the wavevector mismatch and is given by $\Delta k = k_{2\omega} - 2k_\omega$. These are simplified expressions that apply to nearly monochromatic plane waves. Two solutions to the coupled equations are

$$I_{2\omega}(z) = \frac{2\omega^2 d_{\text{eff}}^2 l^2 I_\omega^2(0)}{n_{2\omega}n_\omega^2 c^3 \varepsilon_0}\frac{\sin^2(\Delta kz/2)}{(\Delta kz/2)^2} \quad [10]$$

for insignificant depletion of the fundamental amplitude and

$$I_{2\omega}(z) = I_\omega(0)\tanh^2\sqrt{\frac{2\omega^2 d_{\text{eff}}^2 z^2 I_\omega(0)}{n_{2\omega}n_\omega^2 c^3 \varepsilon_0}} \quad [11]$$

for the case of perfect phase matching $\Delta k = 0$ and allowing depletion of the fundamental. The quantity $I_{2\omega}(z)$ is the harmonic intensity after propagating a distance z in the nonlinear material, and $I_\omega(0)$ is the fundamental intensity at the input surface of the nonlinear material where $I_{2\omega}(0) = 0$.

The above relationships for nonlinear polarization, the coupled equations for second-harmonic generation, or harmonic conversion of a monochromatic plane wave, each determine the definition of the second-order nonlinear optical coefficient.

Symmetry

Symmetry has an important role in the measurement of nonlinear optical coefficients. Indeed the second-order nonlinear optical coefficient must be identically zero if the material has inversion symmetry. This restricts materials to noncentrosymmetric crystals or isotropic materials in which inversion symmetry is removed by such techniques as application of an electric field or polymeric materials in which alignment of molecules is achieved by some poling process. The discussion here is limited to noncentrosymmetric crystals.

The nonlinear optical coefficient is a tensor of rank 3. Representation of the tensor in a rotated coordinate system is described by

$$d'_{ijk} = \sum_{l,m,n} R_{il}R_{jm}R_{kn}d_{lmn} \quad [12]$$

where R_{il} is the matrix that describes rotation of the initial coordinate system xyz to the new coordinate system $x'y'z'$. The representation of a vector r_l in the original coordinate system becomes $r'_{\mathrm{i}} = \sum_l R_{il}r_l$ in the new coordinate system. A required symmetry property of the nonlinear optical tensor, that exchange of the second two indices of a component does not change the value of the component $d_{ijk} = d_{ikj}$, allows the third-rank tensor to be written as a 3 by 6 matrix:

$$\begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{pmatrix} \quad [13]$$

with the components of the matrix related to the tensor components by

$$\begin{aligned} &d_{i1} = d_{i11} \quad d_{i4} = d_{i23} = d_{i32} \\ &d_{i2} = d_{i22} \quad d_{i5} = d_{i13} = d_{i31},\ i = 1,2,3 \\ &d_{i3} = d_{i33} \quad d_{i6} = d_{i12} = d_{i21} \end{aligned} \quad [14]$$

A further property, tensor components with any permutation of the indices are equal $d_{ijk} = d_{kij}$, called Kleinman symmetry or Kleinman's conjecture, is often applied if there is no absorption at any of the wavelengths involved. Kleinman symmetry assures that the amount of energy depleted from the

fundamental is equal to the amount of energy gained by the second harmonic. This reduces the number of independent components of the nonlinear optical matrix to 10: d_{11}; $d_{12} = d_{26}$; $d_{13} = d_{35}$; $d_{14} = d_{25} = d_{36}$; $d_{15} = d_{31}$; $d_{16} = d_{21}$; d_{22}; $d_{23} = d_{34}$; $d_{24} = d_{32}$; and d_{33}. Sometimes components equal, according to Kleinman symmetry, are measured individually, either as a check of the measurement or to check the validity of the conjecture.

Crystal symmetry places additional constraints on the components of the nonlinear optical matrix. These constraints determine the components that must be zero and determine additional relationships between the nonzero components. Matrices for selected crystal point groups are given in Table 1. Effective nonlinear optical coefficients can be found by a coordinate frame transformation in which the rotated axes align with the direction of propagation and the eigen polarizations. Effective nonlinear coefficients for a few crystal point groups are given in Table 2. More complete listings of both the reduced matrices and effective nonlinear optical coefficients are given elsewhere.

Table 1 Nonlinear coefficient matrices for selected crystal point groups

Orthorhombic *mm*2

$$\begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & 0 \\ d_{31} & d_{32} & d_{33} & 0 & 0 & 0 \end{pmatrix}$$

Hexagonal 6

$$\begin{pmatrix} 0 & 0 & 0 & d_{14} & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & -d_{14} & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{pmatrix}$$

Tetragonal $\bar{4}2m$

$$\begin{pmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{36} \end{pmatrix}$$

Hexagonal 6*mm*

$$\begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{pmatrix}$$

Trigonal 3*m* $m \perp$ X

$$\begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & -d_{22} \\ -d_{22} & d_{22} & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{pmatrix}$$

Cubic $\bar{4}3m$

$$\begin{pmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{14} \end{pmatrix}$$

Measurements

Usually the symmetry of a nonlinear crystal is determined by techniques such as X-ray diffraction or surface morphology. Crystal symmetry determines the nonzero components of the second-order nonlinear optical tensor. Nonphase-matched second-harmonic-generation techniques, such as Maker-fringe and wedge measurements, allow selection of polarizations and propagation directions relative to the crystal lattice that isolates individual components of the tensor. Phase-matched measurement techniques usually require propagation in a direction for which an effective nonlinear optical coefficient applies and can involve more than one tensor component. Parametric fluorescence and sum- and difference-frequency-generation measurements are phase-matched processes that involve three wavelengths. All of the measurement techniques have been used with pulsed and continuous-wave laser output. In all of these methods, except parametric fluorescence, it is necessary to accurately characterize the incident radiation. Measurements relative to a standard material avoid some of the concerns of characterization of the fundamental radiation. Relative measurements are important in that they allow correlation between different sets of measurements. Measurement sets, that include both relative and absolute measurements on a number of materials, are useful in developing correlations that achieve more accurate compilations of values.

It is desirable to have an incident beam with a Gaussian-like transverse intensity distribution and close to diffraction limited propagation. Gaussian beams have advantages that they are most easily characterized and their propagation is described with relatively simple analytical methods. Accurate temporal characterization of the incident light is also essential. Here it is desirable to have stable single-mode laser operation to produce narrow bandwidth light that is well within phase-matching acceptance bandwidths and that has temporal structure slow enough to be easily characterized and slow enough

Table 2 Effective nonlinear optical coefficient for a few symmetries

*mm*2, a point group with biaxial birefringence		
$\phi = 0$, two waves $E \parallel Y$, one wave $E \parallel XZ$	$d_{\text{eff}} = d_{32} \sin\theta$	
$\phi = 90°$, two waves $E \parallel XY$, one wave $E \parallel Z$	$d_{\text{eff}} = d_{31} \cos^2\phi + d_{32} \sin^2\theta$	
$\phi = 90°$, two waves $E \parallel X$, one wave $E \parallel YZ$	$d_{\text{eff}} = d_{31} \sin\theta$	
Some crystal symmetry groups with uniaxial birefringence		
point group	two e rays and one o ray	two o rays and one e ray
$\bar{4}2m$	$d_{\text{eff}} = d_{14} \sin^2 2\theta \cos 2\phi$	$d_{\text{eff}} = -d_{14} \sin\theta \sin 2\phi$
3*m*	$d_{\text{eff}} = d_{22} \cos^2\theta \cos 3\phi$	$d_{\text{eff}} = d_{15} \sin\theta - d_{22} \cos\theta \sin 3\phi$
6 and 6*mm*	$d_{\text{eff}} = 0$	$d_{\text{eff}} = d_{15} \sin\theta$

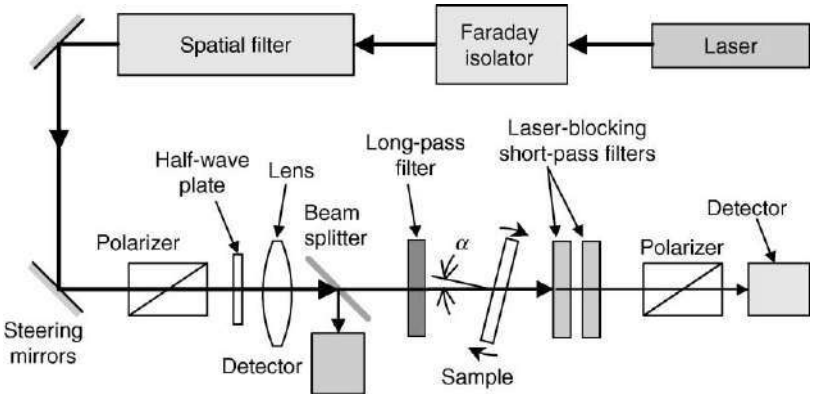

Figure 1 Schematic illustration of an experimental setup to record Maker-fringe measurements.

that group-velocity mismatch will be of no concern. Generating such laser output usually requires special techniques, such as injection-seeded pulse operation or stabilized single-frequency cw operation. Often spatial filters are required to improve the beam quality on the laser output. Even though these topics are only mentioned here, the techniques are important and demand much effort.

The analysis of measurements can also be complicated. Consider the Maker-fringe measurement technique (Figure 1). In this method a focused fundamental beam is incident on a thin slab of oriented nonlinear optical material for nonphase-matched second-harmonic generation. The thickness of the slab is large enough that the function $\sin(\Delta kz/2)$ has gone through many cycles. As the crystal is rotated both the path length in the material l and the wave vector mismatch Δk change, and $\sin^2(\Delta k \cdot l/2)$ oscillates between maxima and minima (Figure 2). The coherence length $L_c = \pi/\Delta k$ is normally small enough that conversion remains in the region of small depletion of the fundamental. Knowledge of the material dispersion and birefringence, required to calculate Δk, or L_c, can be obtained from the fringe spacing. Ideally, the incident fundamental beam will be focused to a spot size w_0, small enough to provide measurable harmonic conversion but large enough that the near-field approximation, in which diffraction is not significant, can be used for analysis of harmonic generation. In this case integration over the incident fundamental intensity distribution $I_\omega(r, z = 0) = I_{\omega,0} \exp(-2r^2/w_0^2)$ yields an expression for the harmonic power:

$$W_{2\omega}(\alpha) = \frac{8\omega^2 d_{\text{eff}}^2(\alpha) W_\omega^2 L_c^2(\alpha)}{\pi^3 w_0^2 n^3 c^3 \varepsilon_0} \sin^2\left(\frac{\pi l(\alpha)}{2L_c(\alpha)}\right) \times \text{correction factors} \quad [15]$$

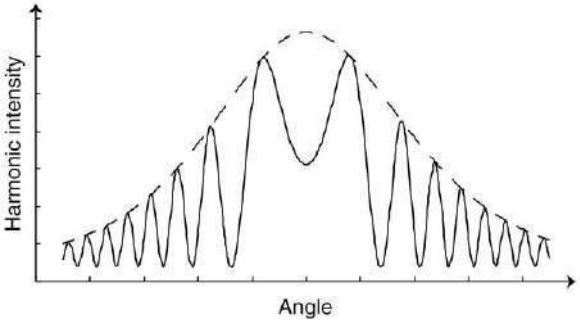

Figure 2 The envelope and spacing of the fringes allow determination of the second-order nonlinear optical coefficient and the coherence length.

where α is the angle of incidence of the fundamental radiation, $W_{2\omega}(\alpha)$ is the second harmonic power, and W_ω is the incident fundamental power. This equation applies to cw radiation. It is necessary to integrate over the temporal distributions for pulsed fundamental light. The effective nonlinear coefficient $d_{\text{eff}}(\alpha)$, the path length in the crystal $l(\alpha)$, and the coherence length $L_c(\alpha)$ change with the angle of incidence α. The correction factors include Fresnel losses at the crystal surfaces, harmonic generation at the surfaces, multiple reflections between the surfaces, absorption at either fundamental or harmonic wavelengths, and walkoff due to non-normal incidence and birefringence.

The wedge technique (Figures 3 and 4) also uses nonphase-matched second-harmonic generation. A wedged sample is translated through a focused fundamental beam, and the above equation applies with the angle α constant. The crystal length l varies with transverse position. It is important that the wedge angle and the fundamental beam spot size are chosen so that the entire beam fits well within a maximum of conversion and the beam is large enough to avoid divergence due to diffraction. Both the Maker-fringe

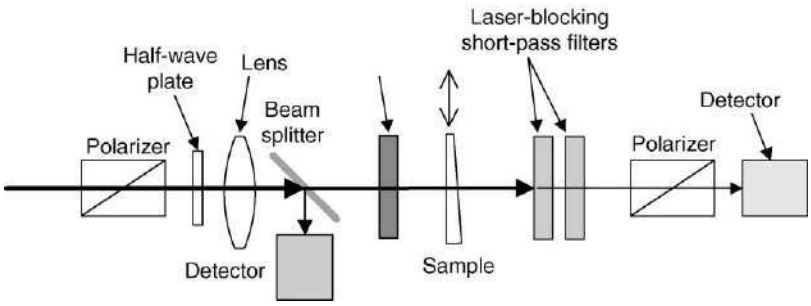

Figure 3 The setup for wedge measurements is similar to that used for Maker-fringe measurements but with a wedged sample translated in the fundamental beam to change path length.

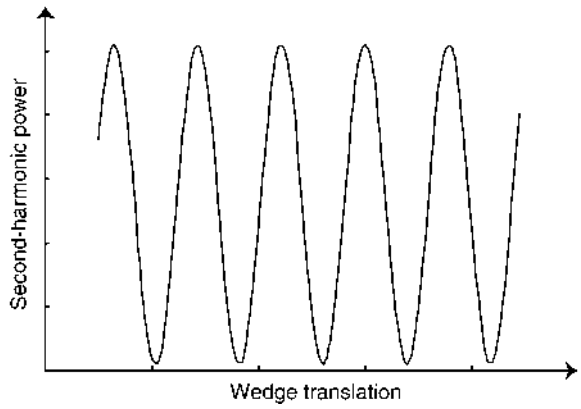

Figure 4 The harmonic intensity cycles between a minimum and maximum when the path is changed by one coherence length.

and wedge measurements have the advantage of being able to isolate individual components of the nonlinear optical tensor. Harmonic power can be low with cw beams requiring chopped, lock-in detection methods or even photon-counting techniques. Even with pulsed fundamental radiation, peak harmonic power may only be at the level of milliwatts. It is necessary to isolate the harmonic radiation from the fundamental radiation that could have peak power in the range of 10^5 W for a Q-switch laser.

Phase-matched second harmonic generation (Figure 5) offers a measurement technique that is similar to practical applications. Harmonic signals are large, thus simplifying detection. However, it is helpful to keep conversion levels below a few percent to allow the use of small depletion approximations in data analysis. Good optical quality of the nonlinear optical material and good quality and accurate characterization of the fundamental radiation are important in this measurement technique. The analysis is simplified for low levels of conversion, spectral distribution well within the phase-matching bandwidth, no absorption or other losses, and no significant diffraction or walkoff. If the fundamental intensity is a pulse with distribution of $I_\omega(r,t) = I_{\omega,0} \exp(-t^2/\tau^2)\exp(-2r^2/w_0^2)$, integration yields second-harmonic energy:

$$U_{2\omega} = \int_{-\infty}^{\infty} \mathrm{d}t \int_0^{\infty} 2\pi r\, \mathrm{d}r I_{2\omega}(r, z = l)$$
$$\approx \frac{2\sqrt{2}\omega^2 d_{\mathrm{eff}}^2 l^2 U_\omega^2}{n^3 c^3 \varepsilon_0 \pi^{3/2} \tau w_0^2} \frac{\sin^2(\Delta k l/2)}{(\Delta k l/2)^2} \qquad [16]$$

where $U_{2\omega}$ is the second harmonic energy and U_ω is the fundamental energy, r is the radial coordinate, t is time, and l is the length of the material. Harmonic conversion can be measured as Δk and varied, either by crystal rotation or temperature change, to confirm that optical quality of the nonlinear material and incident fundamental radiation are a level of tuning that is accurately predicted by theory (Figure 6). Even this simplified equation shows the importance of accurate characterization of the fundamental radiation and second-harmonic conversion in obtaining values of the nonlinear optical coefficient. Conversion efficiency is calculated more accurately with detailed analytical techniques or numerical methods, provided accurate and complete characterization of the fundamental radiation is available.

Parametric fluorescence, in principle, offers an experimental simplification over second-harmonic generation techniques. Parametric fluorescence (Figure 7) is a phase-matched technique in which a pump beam of angular frequency ω_p produces a signal beam and an idler beam of angular frequencies $\omega_\mathrm{s} + \omega_\mathrm{i} = \omega_\mathrm{p}$ by the process of spontaneous parametric generation. It is only necessary to measure the ratio of the generated signal power W_s and the incident pump power W_p. The signal power is given by

$$W_\mathrm{s} = \frac{2(h/2\pi)\omega_\mathrm{s}^4 \omega_\mathrm{i} d_{\mathrm{eff}}^2 n_\mathrm{s} l W_\mathrm{p} \Omega}{(2\pi)^2 \varepsilon_0 c^5 n_\mathrm{i} n_\mathrm{p} |\partial \Delta k/\partial \omega_\mathrm{s}|_{\omega p}} \qquad [17]$$

where h is Planck's constant, Ω is the solid angle of detection (inside the crystal), and n_s, n_i, and n_p are

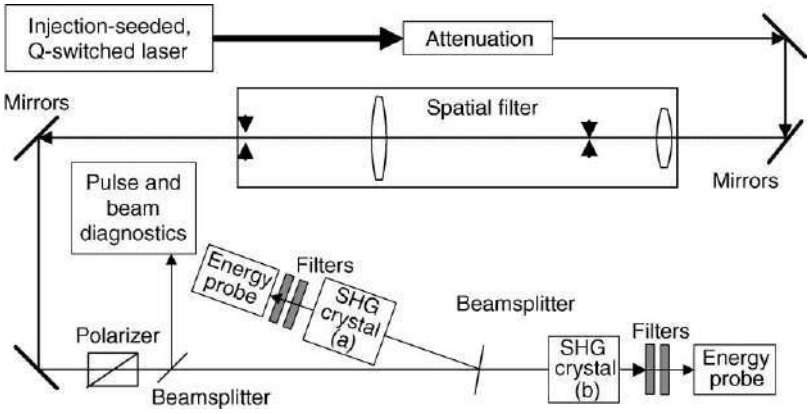

Figure 5 A setup for a relative measurement of second-order nonlinear optical coefficient based on phase-matched second-harmonic generation is shown.

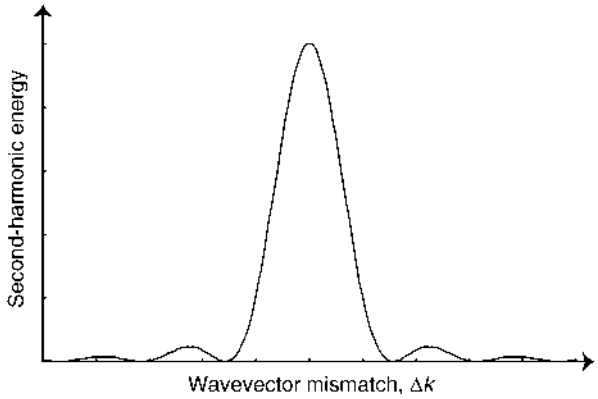

Figure 6 Observation of a theoretically predicted phase-matching tuning curve confirms required optical quality.

the indices of refraction of the signal, idler, and pump respectively. The absolute value of the partial derivative of $\Delta k = k_p - k_s - k_i$, with respect to signal frequency with the pump frequency held constant, is

$$|\partial \Delta k/\partial \omega_s|_{\omega p} = \frac{1}{c}\left| n_s - n_i + \omega_s \frac{dn_s}{d\omega_s} - \omega_i \frac{dn_i}{d\omega_i} \right| \quad [18]$$

It is still not well understood why the carefully performed earlier measurements by this method produced nonlinear optical coefficient values larger than those obtained by other techniques.

Tabulations of measurements of nonlinear optical coefficients are included in handbooks. The Internet posting 'SNLO' listed in the Further Reading is a good source of up-to-date values. Table 3 lists some nonlinear optical coefficients obtained from several sources. Signs of these coefficients are not always well determined, and they are reproduced where given and may only indicate relative polarity. Point group symmetry is given with some comments about axis assignments for the *mm*2 point group. The fundamental wavelength is given for second-harmonic measurements and twice the shortest wavelength is given for three-wave processes. Many of the values obtained by Shoji *et al.* are based on carefully performed wedge measurements with single frequency cw lasers and taking detailed account of multiple reflections in the analysis. Their measurements for congruent $LiNbO_3$ reported agreement between Maker-fringe and wedge measurements, and phase-matched difference-frequency generation measurements, and finally agreement with parametric fluorescence. The measurements of Alford and Smith combined optical-parametric amplification and second-harmonic generation at a number of different wavelengths and correlation with values previously reported in the literature to examine the validity of using a constant Miller's delta, and reported best-fit values for the second-order nonlinear coefficient. Constant Miller's delta provided reasonably accurate wavelength scaling, with good agreement for most materials in recent measurements.

The measurements of the nonlinear optical coefficients of $KNbO_3$, listed in Table 3 are of particular interest. These measurements were performed by the separated-beam wedge technique (Figure 8). With this technique, a large wedge angle is used in a relatively long crystal and with a larger fundamental beam than in the conventional wedge technique. The beam is large enough that many fringes are present in the second-harmonic at the exit surface. The fringe structure causes focused harmonic radiation to diffract into two clearly resolved spots. Integration of an earlier equation, with constant fundamental field amplitude $E_\omega(0)$, provides a simplified expression of the second-harmonic field $E_{2\omega}(z,t)$ in terms of a free wave with propagation constant

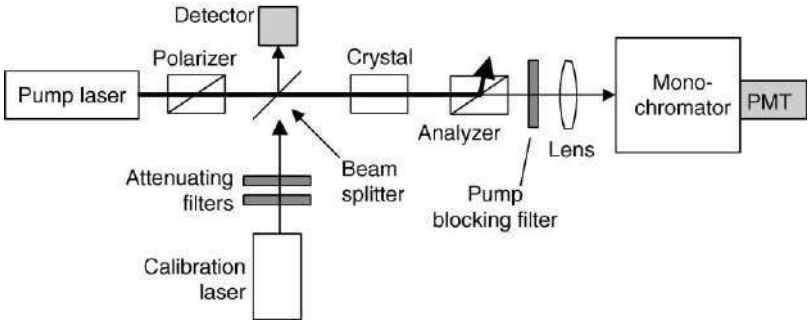

Figure 7 Spontaneous optical parametric fluorescence requires measurement of the ratio of the power of the signal fluorescent to the power of the pump radiation. The calibration laser is used to measure the signal transmission of the filtering system used to block the pump radiation from reaching the photomultiplier tube (PMT).

Table 3 Selected nonlinear optical coefficients

Material	*Point group*	*Nonlinear optical coefficient (pm V^{-1})*		*Wavelength*	*Source*
$AgGaS_2$	$\bar{4}2m$	$d_{14} = d_{24} = d_{36}$	11.2	10.6 μm	*a*
		$d_{14} = d_{24} = d_{36}$	17.5	1064 nm	*a*
$AgGaSe_2$	$\bar{4}2m$	$d_{14} = d_{24} = d_{36}$	33	10.6 μm	*a*
BaB_2O_4 (BBO)	$3m$	d_{22}	2.3	1064 nm	*a–f*
		$d_{16} = d_{21} = -d_{22}$	−2.3	1064 nm	*a*
		$d_{15} = d_{31} = d_{24} = d_{32}$	0.04	1064 nm	*d*
		d_{33}	0.04	1064 nm	*d*
CdS	$6mm$	d_{33}	32	10.6 μm	*a*
		$d_{15} = d_{31} = d_{24} = d_{32}$	−16	10.6 μm	*a*
		d_{33}	19.1	1064 nm	*e*
		$d_{15} = d_{31}$	10.1	1064 nm	*e*
		$d_{24} = d_{32}$	10.7	1064 nm	*e*
CdSe	$6mm$	d_{33}	36	10.6 μm	*a*
		$d_{15} = d_{31}\, d_{24} = d_{32}$	−18	10.6 μm	*a*
CdTe	$\bar{4}3m$	$d_{14} = d_{24} = d_{36}$	109	1064 nm	*e*
		$d_{14} = d_{24} = d_{36}$	168	10.6 μm	*c*
$CsLiB_6O_{10}$	$\bar{4}2m$	$d_{14} = d_{24} = d_{36}$	0.74	1064 nm	*f*
GaAs	$\bar{4}3m$	$d_{14} = d_{24} = d_{36}$	83	10.6 μm	*a*
			170	1064 nm	*e*
GaP	$\bar{4}3m$	$d_{14} = d_{24} = d_{36}$	37	10.6 μm	*a*
			70.6	1064 nm	*e*
KD_2PO_4 (KD^*P)	$\bar{4}2m$	d_{36}	0.37	1064 nm	*a,b*
KH_2PO_4 (KDP)	$\bar{4}2m$	d_{36}	0.39	1064 nm	*a,b,e,g,h*
$KNbO_3$	$mm2$	d_{33}	21.9	1064 nm	*h*
$Z = Z_L = x = l$		d_{15}	13	1064 nm	*h*
$X = Y_L = z = s$		d_{31}	12.4	1064 nm	*h*
$Y = X_L = -y = m$		d_{24}	9.2	1064 nm	*h*
		d_{32}	8.9	1064 nm	*h*
$KTiOAO_4$ (KTA)	$mm2$	$d_{15} = d_{31}$	4.2	1064 nm	*e*
$Z = Z_L = z = m$		d_{15}	2.91	2128 nm	*g*
$X = Y_L = -y = s$		$d_{24} = d_{32}$	2.8	1064 nm	*e*
$Y = X_L = x = l$		d_{33}	16.2	1064 nm	*e*
$KTiOPO_4$ (KTP)	$mm2$	d_{33}	14.6	1064 nm	*e*
$Z = Z_L = z = m$		d_{15}	3.7	1064 nm	*e*
$X = Y_L = -y = s$		d_{31}	3.7	1064 nm	*e*
$Y = X_L = x = l$		d_{24}	1.9	1064 nm	*e*
		d_{32}	2.2	1064 nm	*e*
LiB_3O_5 (LBO)	$mm2$	d_{33}	±0.04	1064 nm	*a,b*
$Z = Z_L = y = s$		$d_{15} = d_{31}$	±0.85	1064 nm	*a,b*
$X = Y_L = z = m$		d_{15}	1.04	1064 nm	*g*
$Y = X_L = x = l$		$d_{24} = d_{32}$	∓0.67	1064 nm	*a,b*
$LiIO_3$	6	d_{33}	4.5	1064 nm	*a,b*
		$d_{15} = d_{31}$	4.4	1064 nm	*a,b*
congruent	$3m$	d_{33}	−27.2	1064 nm	*a,b*
$LiNbO_3$		$d_{31} = d_{15} = d_{32} = d_{24}$	−4.35	1064 nm	*a,b*
		$d_{22} = -d_{21} = -d_{16}$	2.1	1064 nm	*a,b*

Table 3 (*Continued*)

Material	*Point group*	*Nonlinear optical coefficient (pm V^{-1})*		*Wavelength*	*Source*
Agreement of measurements of d_{31} by Maker fringe and wedge SHG (4.6 pm V^{-1}) and parametric fluorescence and parametric amplification (4.3 pm V^{-1})					
				1064 nm	e
5%MgO:$LiNbO_3$	3*m*	d_{33}	25	1064 nm	e
		d_{31}	4.4	1064 nm	e
$LiTaO_3$	3*m*	d_{33}	13.8	1064 nm	e
		d_{31}	0.85	1064 nm	e
$NH_4H_2PO_4$ (ADP)	$\bar{4}2m$	$d_{14} = d_{24} = d_{36}$	0.47	1064 nm	a,e
αSiO_2	32	d_{11}	0.3	1064 nm	
$ZnGeP_2$	$\bar{4}2m$	$d_{14} = d_{24} = d_{36}$	69	10.6 μm	a,b

[a]Roberts DA (1992) Simplified characterization of uniaxial and biaxial nonlinear optical crystals: a plea for standardization of nomenclature and conventions. *IEEE Journal of Quantum Electronics* 28: 2057–2074.
[b]Dmitriev VG (1997) *Handbook of Nonlinear Optical Crystals*, 2nd edn. Berlin: Springer-Verlag.
[c]Sutherland RL (2003) *Handbook of Nonlinear Optics*, 2nd edn, chs 4 and 5. New York: Marcel Dekker.
[d]Shoji I, Nakamura H, Ohdaira K, *et al.* (1999) Absolute measurement of second-order nonlinear-optical coefficients of β-BaB_2O_4 for visible to ultraviolet second-harmonic wavelengths. *Journal of the Optical Society of America B* 16: 620–624.
[e]Shoji I (1997) Absolute scale of second-order nonlinear-optical coefficients. *Journal of the Optical Society of America B* 14: 2268–2294.
[f]Shoji I, Nakamura H, Ito R, *et al.* (2001) Absolute measurement of second-order nonlinear-optical coefficients of $CsLiB_6O_{10}$ for visible-to-ultraviolet second-harmonic wavelengths. *Journal of the Optical Society of America B* 18: 302–307.
[g]Alford WJ and Smith AV (2001) Wavelength variation of the second-order nonlinear coefficients of $KNbO_3$, $KTiOPO_4$, $KTiOAsO_4$, $LiNbO_3$, $LiIO_3$, β-BaB_2O_4, KH_2PO_4, and LiB_3O_5 crystals, a test of Miller wavelength scaling. *Journal of the Optical Society of America B* 18: 524–533.
[h]Pack MV, Armstrong DJ and Smith AV (2003) Measurement of the $\chi^{(2)}$ tensor of the potassium niobate crystal. *Journal of the Optical Society of America B* 20: 2109–2116.

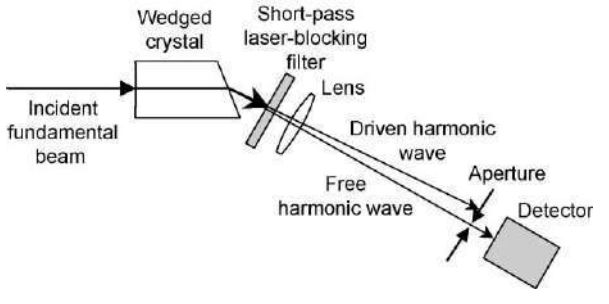

Figure 8 In the separated-beam wedge measurement technique, a fundamental beam of large transverse size generates harmonic radiation that is separated into a free harmonic wave and a driven harmonic wave after exiting the crystal.

$k_{2\omega}$ and a driven wave with propagation constant $2k_\omega$:

$$E_{2\omega}(z,t) = \frac{\omega d_{\text{eff}} E_\omega^2(0)}{n_{2\omega} c} \frac{1}{\Delta k} \{\exp i(k_{2\omega} z - 2\omega t) + c.c.\} - \frac{\omega d_{\text{eff}} E_\omega^2(0)}{n_{2\omega} c} \frac{1}{\Delta k} \{\exp i(2k_\omega z - 2\omega t) + c.c.\} \quad [19]$$

The two beams with different propagation constants are refracted at different angles at the wedged exit surface. The detailed analysis must include losses due to absorption and reflections at the surfaces and also nonlinear effects at the surfaces. Precision again is required in characterizing the fundamental beam and incident and generated intensities or energies as well as the geometry of the measurement. The large wedge eliminates problems of multiple reflections, and it is no longer necessary to carefully choose the beam spot size to fit within one fringe maximum, yet be large enough to allow use of the near field approximation. The accuracy of this measurement is demonstrated when the coefficients are expressed as Miller-delta values: $\delta_{24} = 0.153\,\text{pm V}^{-1}$, $\delta_{32} = 0.150\,\text{pm V}^{-1}$, $\delta_{15} = 0.195\,\text{pm V}^{-1}$, and $\delta_{31} = 0.192\,\text{pm V}^{-1}$. These values satisfy Kleinman symmetry, $\delta_{24} = \delta_{32}$ and $\delta_{15} = \delta_{31}$, well within the reproducibility of the measurement, which was about $\pm 3\%$.

The nonlinear optical coefficients of BBO $d_{22}(1064\ \text{nm}) = 2.3\ \text{pm V}^{-1}$ and KDP $d_{36}(1064\ \text{nm}) = 0.39\ \text{pm V}^{-1}$ are gaining agreement from a number of independent measurements. The apparent disparity in the values for GaAs $d_{36}(10.6\ \mu\text{m}) = 83\ \text{pm V}^{-1}$ and $d_{36}(1064\ \text{nm}) = 170\ \text{pm V}^{-1}$, is resolved when the coefficients are converted to Miller's delta values: $\delta_{36} = 0.089\ \text{pm V}^{-1}$ and $\delta_{36} = 0.086\ \text{pm V}^{-1}$, respectively. The same thing can be done for GaP, for which $d_{36}(10.6\ \mu\text{m}) = 37\ \text{pm V}^{-1}$ and $d_{36}(1064\ \text{nm}) = 70.6\ \text{pm V}^{-1}$ convert to $\delta_{36} = 0.085\ \text{pm V}^{-1}$ and $\delta_{36} = 0.077\ \text{pm V}^{-1}$, respectively. However, the older and newer values for other materials such as CdTe and CdS do not achieve this agreement. Other materials such as CLBO and LBO need additional investigation to resolve differences. The measurement of nonlinear optical coefficients is a continuing process.

Other techniques for measurement involving different nonlinear-optical processes are possible. Parametric amplification was mentioned above. Generally more complicated processes are less well suited for measurement of fundamental properties. Integrated-optical devices such as waveguides are even more complex. Often nonlinearity is measured in fractional conversion per unit of pump power to characterize overall device performance. There are survey techniques for evaluating materials such as grinding the material to a powder, illuminating with intense laser radiation, and looking for harmonic conversion. A surprising amount of information, such as a rough estimate of nonlinearity and the possibility of birefringent phase matching, can be gleaned by sorting the power and measuring the scattered harmonic radiation. Second-harmonic reflections from surfaces and thin film can be used in some instances. Electric fields can be use to orient molecules in polymers and solute molecules in solution to measure the hyperpolarizability of individual molecules.

Reporting Frames

Standardized axis specifications avoid confusion, but nonstandard axes can prove convenient in some applications provided care is used. The crystal $KNbO_3$ has $mm2$ point group; it could also be $m2m$ or $2mm$ depending on the choice of axes to describe the crystal properties. The axis assignment for the orthorhombic class $mm2$, as it is used in Table 3, follows the IEEE/ANSI Std. 176-1987: Z is the polar axis, X is parallel to the shorter of the remaining unit cell dimensions, and Y is chosen to complete a right-handed coordinate frame. Polarity of the axes can be assigned by pyroelectric properties and piezoelectric properties of the crystal. The common assignments of the axes in the literature are given by X_L, Y_L, Z_L in the table. Reference to the biaxial indices of refraction are given by the lower case x, y, z with $n_x < n_y < n_z$. Finally the axes are identified by the unit cell dimensions $s < m < l$. Either a plus or minus rotation about the Y axis will bring the crystal into an xyz frame. However, a $+90$ degree rotation will give coefficients that are opposite in sign from a -90 degree rotation. The signs of the nonlinear optical coefficients are often not well determined because sign is not obtained from conversion efficiency in a single crystal. However, coefficient signs can be compared in the same crystal or between two crystals to determine if they are the same or opposite.

The crystals KTP or KTA are used here as an example. A 90° rotation about the Z-axis gives a new coordinate system $X'Y'Z'$ aligned with the xyz frame determined by the indices of refraction. In the new frame, the component of the nonlinear optical tensor, in terms of the standard frame, are $d_{zxx} = d_{32}$ and $d_{zyy} = d_{31}$. All that is needed here is to exchange d_{32} and d_{31} to change the standard frame to the one commonly used in the literature, X_L, Y_L, Z_L. When the largest index of refraction is not aligned with the polar axis, as is the case for LBO and $KNbO_3$, the specification of tensor components can become confusing if a standard frame is not used.

Often crystallographers and crystal manufacturers can identify axes by surface morphology or by poling processes of ferroelectric crystals. With other crystal symmetries a change in the reporting frame that does not change linear optical properties may change the sign of some of the nonlinear optical tensor components while leaving others unchanged. For a crystal of $3m$ point group, a 180° rotation about Z will change the sign of d_{22}, d_{21}, and d_{16} and leave the other coefficients unchanged. The xyz frame with $n_x < n_y < n_z$ is convenient for calculations of optical propagation, and the standard XYZ frame is unambiguous for reporting of tensor properties.

Conclusion

There is still a range of values for nonlinear optical coefficients of many materials. Standardization is evolving on more commonly used materials. Greater precision in measurement and extension of measurement techniques are providing better accuracy. However, caution is still required when accessing values. It is necessary to scrutinize the sources of values to obtain the highest accuracy. Accurate measurements need to be extended to difficult spectral ranges of the ultraviolet and infrared. The ambiguity of the sign of tensor components needs to be resolved. In recent measurements, carefully executed wedge measurements are producing more reliable results. The separated beam wedge technique appears to be particularly promising. Phase-matched second-harmonic generation and other practical applications of nonlinear optical materials will continue to serve as techniques to verify accuracy. Current analytical and numerical techniques provide a method for precise modeling of nonlinear optical processes and the need for more accurate material characterization.

See also

Fiber and Guided Wave Optics: Nonlinear Optics.
Nonlinear Optics, Basics: $\chi^{(2)}$-Harmonic Generation.

Further Reading

Alford WJ and Smith AV (2001) Wavelength variation of the second-order nonlinear coefficients of $KNbO_3$, $KTiOPO_4$, $KTiOAsO_4$, $LiNbO_3$, $LiIO_3$, β-BaB_2O_4, KH_2PO_4, and LiB_3O_5 crystals, a test of Miller wavelength scaling. *Journal of the Optical Society of America B* 18: 524–533.

Dmitriev VG (1997) *Handbook of Nonlinear Optical Crystals*, 2nd edn. Berlin: Springer-Verlag.

Gehr RJ (1998) Separated-beam non-phase-matched second-harmonic method of characterizing nonlinear optical crystals. *Journal of the Optical Society of America B* 15: 2109–2116.

Kurtz SK (1975) Measurement of nonlinear optical susceptibilities. In: Rabin H and Tang CL (eds) *Quantum Electronics: A Treatise, vol. 1, part A*, pp. 209–281. New York: Academic Press.

Pack MV (2003) Measurement of the $\chi^{(2)}$ tensor of the potassium niobate crystal. *Journal of the Optical Society of America B* 20: 2109–2116.

Roberts DA (1992) Simplified characterization of uniaxial and biaxial nonlinear optical crystals: a plea for standardization of nomenclature and conventions. *IEEE Journal of Quantum Electronics* 28: 2057–2074.

Shoji I (1997) Absolute scale of second-order nonlinear-optical coefficients. *Journal of the Optical Society of America B* 14: 2268–2294.

SNLO nonlinear optics code available from Smith AV Sandia National Laboratories, Albuquerque, NM 87185-1423. (an Internet publication) http://www.sandia.gov/imrl/XWEB1128/xxtal.htm.

Sutherland RL (2003) *Handbook of Nonlinear Optics*, 2nd edn, chs 4 and 5. New York: Marcel Dekker.

$\chi^{(3)}$

P P Banerjee, University of Dayton, Dayton, OH, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Over the past 30 years, several methods have been used to measure and characterize optical nonlinearities. With increasing interest in nonlinear optics beyond the traditional harmonic generation, third-order nonlinear processes have been shown to be responsible for self-focusing and defocusing of optical beams, phase conjugation, optical bistability, etc. Third-order nonlinear processes in Kerr materials arise from the cubic dependence of the polarization $P = P^{L} + P^{NL}$, $P^{NL} = \chi^{(3)}{:}E{\cdot}E{\cdot}E$ in a material on the optical field E, through the nonlinear susceptibility tensor $\chi^{(3)}$. The cubic nonlinearity in turn results in a nonlinear refractive index coefficient n_2 defined through:

$$n = n_0 + n_2|E_p^{\omega_0}|^2 \qquad [1]$$

where $E_p^{\omega_0}$ is the phasor part of $E : E = Re[E_p^{\omega_0}\exp \times j\omega_0 t]$. The relation between $\chi^{(3)}$ and n_2 can be expressed as:

$$n_2 = (3/8n_0)\chi^{(3)}_{1111}(-\omega_0; \omega_0, \omega_0, -\omega_0) \qquad [2]$$

In some materials, an effective $\chi^{(3)}$ or an effective n_2 can be defined as well. For instance, in a second-order nonlinear material, an effective $\chi^{(3)}$ is established through cascading of $\chi^{(2)}$s, and in the presence of phase mismatch between the fundamental and the second harmonic. In other materials, such as liquid crystals, an effective n_2 can be defined due to the nonlinear dependence of the reorientation of the director axis on the optical field. In photorefractive materials, the induced nonlinearity stems from the nonlinear dependence of the induced space charge distribution in the material on the optical field. Some of the techniques used to measure the third-order nonlinearity in materials include degenerate four-wave mixing (DFWM) or phase conjugation, interferometry, ellipse rotation, and beam self-refraction and/or distortion and bending.

For instance, in DFWM, two contradirectional propagating pumps and the probe beam having the same frequency ω_0 interact to give rise to the conjugate, whose amplitude depends on the magnitude of the cubic nonlinearity coefficient. The conservation of energy and momentum describes a parametric interaction of the form $\omega_0 + \omega_0 - \omega_0 \rightarrow \omega_0$ (in frequency space) and, $\mathbf{k}_{\text{probe}} + \mathbf{k}_{\text{pump1}} - \mathbf{k}_{\text{pump2}} \rightarrow \mathbf{k}_{\text{conj}}$ (in $\mathbf{k}$-space). The conjugate beam is clearly distinguishable by its spatial separation from the other interacting waves. However, alignment of the three waves is critical to obtain a good phase conjugate; furthermore, the sign of the nonlinearity cannot be determined.

In the interferometric method, a laser beam is split into two parts, one serving as the reference, and the other being incident on the nonlinear sample. As in the case of interferometers, if the two beams are then recombined and introduced through a lens, an

interference pattern forms at the focal plane of the lens. While the information about the magnitude of the nonlinearity can be found from the fringes and their change when the incident power is varied, information of the sign of the nonlinearity is still not readily inferable.

In this article, we will concentrate on self-refraction, particularly in light of the fact that it results in a sensitive and at the same time simple method called the *z*-scan©, requiring the propagation of a single beam through the nonlinear sample. We comment that self-diffraction does also cause self-bending of an asymmetric beam due to the asymmetric phase change induced through the cubic nonlinearity; however, extracting the information of the sign and magnitude of the nonlinearity can be effectively done only if the sample is thin, implying that there is negligible linear diffraction of the beam as it passes through the sample. In the rest of the article, we will therefore concentrate on the *z*-scan method, since it is an elegant and simple single (symmetrical) beam technique that yields information on the magnitude and sign of the cubic nonlinearity, as well as its real and imaginary parts. We will describe in detail how the *z*-scan can measure the real part of the cubic nonlinearity, for both thin and thick samples, as well as showing its application in determining the reorientational nonlinearity of liquid crystals and the nonlinearity of photorefractive materials, where photovoltaic effect and diffusion of charge carriers contribute to the effective nonlinearity.

Since the mid-1990s *z*-scan has evolved as a standard method for determining effective nonlinear refractive index coefficients of nonlinear materials. The name *z*-scan is derived from the fact that the nonlinear test sample is scanned along the direction of propagation (z) of a beam about the back focal plane of an external lens illuminated by an incident Gaussian beam from a laser. The far-field diffraction pattern yields information about the nonlinearity of the test material. The *z*-scan technique has been used to measure and characterize the nonlinear refractive index coefficient and nonlinear absorption coefficient in materials such as CS_2, ZnSe, Cs atoms, multiphoton absorption in GaN, organic dyes, thermal nonlinearities in liquids, liquid crystals, photorefractive materials, etc., to name a few. *z*-scan measurements have been made using non-Gaussian incident beams, and *z*-scans have been performed using both CW and pulsed laser sources.

The first extensive series of work on the *z*-scan technique, including analysis and application to the measurement of the cubic nonlinearity, was performed by the group of Van Stryland at the Center of Electro-Optics and Lasers (CREOL) at the University of Central Florida. Other investigators later performed similar measurements and characterization during the study of nonlinear optical properties of a solution of chlorophyll in acetone, and Chinese tea.

Theory of the Thin Sample *z*-Scan

We will first analyze propagation of a (circularly symmetric) Gaussian beam which is initially focused by an external lens and is then incident on a 'thin' slice of a cubically nonlinear Kerr medium. This case is important because, as explained below, it provides a basis for the measurement of the nonlinear refractive index coefficient n_2.

Figure 1 shows a simple ray diagram to demonstrate what happens if the cubically nonlinear material is thin. The propagating Gaussian beam induces its own lens, which is positive for a material having a positive n_2, for reasons that will become clear to the reader later on in this section. Note that if the position of the induced lens is at the back focal plane of the external lens, the former does not affect the propagation of the beam behind the nonlinear material. The situation is thus the same as the propagation of the externally focused Gaussian beam in a linear diffraction-limited environment. However, if the separation between the external lens and the sample is less than the focal length f of the external lens, the beam in the far field is wider than the linear diffraction-limited case. Conversely, if the lens-sample separation is larger than f, the beam width is smaller. This is the basis of the *z*-scan: for a material with a positive n_2, the scanning of a thin nonlinear sample through the back focal plane of the external lens, starting from a position to the left of the focus, yields far-field profiles which are initially larger and eventually smaller than the width in the linear diffraction-limited case. The opposite is true for materials with a negative n_2. For very large displacements of the sample from the back focal plane of the

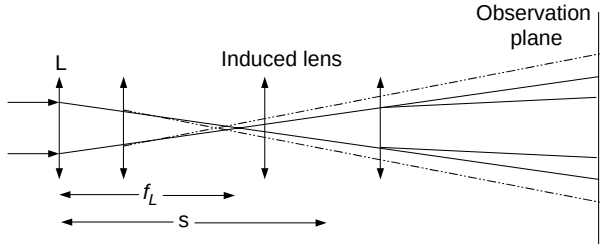

Figure 1 Geometrical optics picture of *z*-scan setup for nonlinear refractive index n_2. Reproduced with permission from Banerjee PP (2004) *Nonlinear Optics: Theory, Numerical Modeling, and Applications.* New York: Marcel Dekker.

external lens, the beam power is small enough that nonlinear lensing is negligible, and hence the beam widths return to the linear diffraction-limited case. The variation of the on-axis intensity or transmittance as a function of scan distance is complementary to the variation of the width, due to conservation of power, to be explained further below. A sketch of the on-axis transmittance during a z-scan for a positive n_2 material, is shown in Figure 2.

The simpler ray picture described above can be augmented by using a q parameter approach to Gaussian beam propagation. For the sake of simplicity, we prefer to introduce the q parameter approach here, instead of the more rigorous Gaussian decomposition of the induced phase as computed by Van Stryland's group. The q of a Gaussian beam in air is defined through the relation $1/q = 1/R - \mathrm{i}(2/k_0 w^2)$, where R and w denote the radius of curvature and the beamwidth, respectively, and k_0 is the propagation constant in air. Assume a Gaussian beam with a waist w_0 (implying initially planewave fronts) incident on the external lens. The Gaussian can be described by $q = q_0 = \mathrm{i}z_{R0}$, where $z_{R0} = k_0 w_0^2/2$ is called the Rayleigh length of the

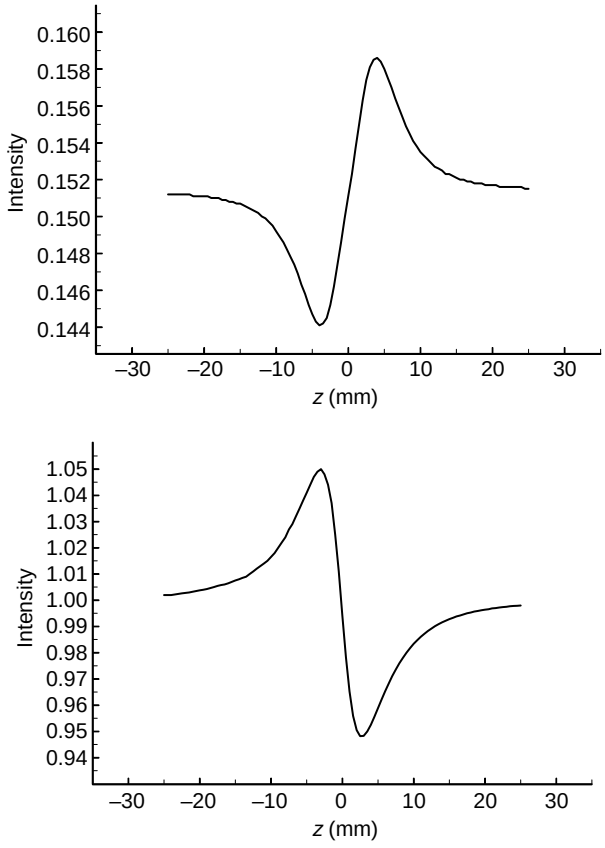

Figure 2 Typical z-scan plots for positive (top) and negative (bottom) cubic nonlinearities. The vertical axis is in arbitrary units.

incident Gaussian. The laws of q transformation of a Gaussian beam are:

(i) for translation through a distance d, the q changes to $q + d$; and
(ii) for a Gaussian with q-parameter q incident on a lens, the q immediately behind the lens, is given by $(1/q - 1/f)^{-1}$.

Using these laws, the $q = q_{\mathrm{obs}}$ of the Gaussian at a distance d from the nonlinearly induced lens of the focal length f_{ind}, which is located at a distance $s = f + \Delta s$ behind the external lens, can be found as:

$$q_{\mathrm{obs}} = \frac{f_{\mathrm{ind}}[q_0 f + (f - q_0)(f + \Delta)]}{f_{\mathrm{ind}}(f - q_0) - [q_0 f + (f - q_0)s]} + d \qquad [3]$$

In deriving eqn [3], we have used the laws of q transformation on the initial Gaussian beam of q-parameter q_0 passing through the external lens, being translated a distance $f + \Delta s$, passing through the induced lens, and finally being translated a distance d to the observation plane.

To determine $f_{\mathrm{ind}}(\Delta s)$, assume a Gaussian of the form:

$$|E_{\mathrm{p}}^{\omega_0}(x, y, z)| = E(z)\exp(-(x^2 + y^2)/w^2(z)) \qquad [4]$$

Since

$$n = n_0 + n_2|E_{\mathrm{p}}^{\omega_0}|^2 \approx n_0 - 2n_2 E^2(z)(x^2 + y^2)/w^2(z) \qquad [5]$$

the nonlinearly induced phase change is

$$\Delta\phi = k_0 n \Delta z \approx k_0[n_0 - 2n_2 E^2(z)(x^2 + y^2)/w^2(z)]\Delta z \qquad [6]$$

so that

$$f_{\mathrm{ind}}(z) = n_0 w^2(z)/4n_2 a^2(z)\Delta z \qquad [7]$$

Note that $f_{\mathrm{ind}}(z)$ is inversely proportional to $n_2 E^2(z)$ and Δz. The longer the sample length, the more the nonlinear induced phase, implying a stronger lensing (smaller f_{ind}). Also, the stronger the nonlinearity parameter n_2 and/or on-axis intensity ($E^2(z)$), the stronger the focusing since the change in n is larger.

We must point out that the focal length f_{ind} of the induced lens is also a function of Δs, since it depends on the beam intensity at the location of the nonlinear sample. In fact, the focal length f_{ind} of the nonlinearly induced lens can be expressed in terms of the beam power P, the beam width w at the position Δs, f (the focal length of the external lens), the linear refractive index n_0, the nonlinear refractive index coefficient n_2 defined through eqn [1] where $|E_{\mathrm{p}}^{\omega_0}| = E$ is the optical field in the nonlinear material of length L, assumed 'thin' for now. The term 'thin' means that the sample length is smaller than the depth of focus of the

externally focused Gaussian beam. In this case, we can assume that the beam width is a constant through the sample, and the induced focal length is then only directly dependent on the total thickness L of the sample. The focal length of a nonlinearly induced 'thin' Kerr lens can be written as

$$f_{\text{ind}}(\Delta s) = \frac{n_0 w^2(\Delta s)}{4n_2|E|^2 L} \approx \frac{\pi f^4 n_0}{k_0^4 w_0^4 \eta_0 P L n_2} \qquad [8]$$

for small Δs, where η_0 is the (linear) characteristic impedance of the nonlinear medium. Note that in eqn [8], the beam width w is given by

$$w^2(\Delta s) = w_{\text{f}}^2\left[1 + \frac{4(\Delta s)^2}{k_0 w_{\text{f}}^4}\right] \text{ and } w_{\text{f}} \cong \frac{2f}{k_0 w_0} \qquad [9]$$

In eqn [9], w_{f} denotes the waist size of the Gaussian beam at the back focal plane of the external lens. In deriving eqn [8], one has to use the power conservation relation $|E|^2 w^2 = (4/\pi)\eta P$ to re-express the dependence of the induced focal length on the field amplitude in terms of the corresponding width at the same point.

Our final interest is to find the imaginary part of $1/q_{\text{obs}}(= -2/k_0 w_{\text{obs}}^2)$, in order to ascertain the width w_{obs} of the Gaussian beam on the observation plane. Simple but tedious algebra leads us to the relation:

$$\frac{k_0 w_{\text{obs}}^2}{2} = \frac{[(ff_{\text{ind}}d - f^2 d + f^2 f_{\text{ind}}) + f(f_{\text{ind}} - d)\Delta s]^2 + z_{R0}^2[(f_{\text{ind}} - d)\Delta s + f_{\text{ind}} d]^2}{z_{R0} f^2 f_{\text{ind}}^2} \qquad [10]$$

It can be shown simply that when $\Delta s = 0$, and for $|f_{\text{ind}}| \gg f$, $w_{\text{obs}} = w_0 d/f$ for large d. In other words, in the linear case, i.e., when $f_{\text{ind}} = \infty$, as well as in the nonlinear case when $\Delta s = 0$, the far-field width of the Gaussian beam is what is predicted by geometrical optics. Furthermore, from eqn [10], the rate of change of the width with Δs around $\Delta s = 0$, (assuming $z_{R0} \gg f$, and for large d) can be found as

$$\frac{\mathrm{d}w_{\text{obs}}}{\mathrm{d}\Delta s} = \frac{-2d^2 z_{R0}}{k_0 w_{\text{obs}} f^2 f_{\text{ind}}} = -\frac{w_0 d}{f f_{\text{ind}}} \qquad [11]$$

provided we invoke the relation for w_{obs} derived above. In eqn [11], f_{ind} refers to the value of the induced focal length (eqn [8]) with $\Delta s = 0$. Equation [11] may be written as

$$\frac{\mathrm{d}w_{\text{obs}}}{\mathrm{d}\Delta s} = -\frac{k_0^4 w_0^5 P L n_2}{2\pi n_0 f^5} \qquad [12]$$

This means that in a material with a positive (negative) n_2, implying a positive (negative) induced lens, the rate of change of the observed width with the position of the nonlinear sample about the focal plane of the external lens will be negative (positive), in agreement with the results from geometrical optics. Furthermore, the rate of change is proportional to the angle of convergence w_0/f of the externally focused Gaussian, and to the position d of the observation plane.

In z-scan, it is often important to relate the on-axis transmittance in the far field to the scan distance s for measurement purposes. Since for power conservation, the intensity is inversely proportional to the width of the beam in the far field, a positive n_2 would imply a positive slope as a function of scan distance with increasing s, and a negative slope for negative n_2. The value of the slope of the on-axis transmittance could give the information of the n_2 of the material. Note that the z-scan graph in this case is an odd function of the scan distance, measured about the back focal plane of the external lens.

In practice, a detector with a finite aperture area is placed on-axis, and the detected power monitored as a function of scan distance. The transmitted power through the aperture is obtained by

$$\begin{aligned} P_{\text{T}} &= \frac{1}{2}\int_0^{2\pi}\int_0^{r_{\text{a}}} \frac{E_{\text{a}} E_{\text{a}}^*}{\eta_0}\,\mathrm{d}A \\ &= \frac{\pi w_0}{2}\left(\frac{E_0^2}{2\eta_0}\right)\left(1 - \exp\left(-\frac{2r_{\text{a}}^2}{w_{\text{obs}}^2}\right)\right) \\ &= P\left(1 - \exp\left(-\frac{2r_{\text{a}}^2}{w_{\text{obs}}^2}\right)\right) \end{aligned} \qquad [13]$$

where r_{a} is the radius of aperture. The rate of change of P_{T} with Δs can be found as

$$\frac{\mathrm{d}P_{\text{T}}}{\mathrm{d}\Delta s} = \frac{-4r_{\text{a}}^2 P}{w_{\text{obs}}^3}\exp\left(-\frac{2r_{\text{a}}^2}{w_{\text{obs}}^2}\right)\frac{\mathrm{d}w_{\text{obs}}}{\mathrm{d}\Delta s} \qquad [14]$$

Substituting eqn [12] into eqn [14] and using the approximation $w_{\text{obs}} = w_0 d/f$, eqn [14] can be rewritten as

$$\frac{\mathrm{d}P_{\text{T}}}{\mathrm{d}\Delta s} \cong \frac{4r_{\text{a}}^2 f^2 P}{f_{\text{ind}} d^2 w_0^2}\exp\left(-\frac{2r_{\text{a}}^2 f^2}{d^2 w_0^2}\right) \qquad [15]$$

from which

$$\frac{\mathrm{d}P_{\text{T}}}{\mathrm{d}\Delta s}(\Delta s \approx 0) \cong \frac{4k_0^4 w_0^2 \eta_0 r_{\text{a}}^2 P^2 L}{\pi f^2 d^2 n_0}\exp\left(-\frac{2r_{\text{a}}^2 f^2}{d^2 w_0^2}\right) n_2 \qquad [16]$$

When r_a approaches 0;

$$\frac{dP_T}{d\Delta s}(\Delta s \approx 0) = 4\left(\frac{k_0^2 w_0 r_a P}{fd}\right)^2 \frac{\eta_0 L}{\pi n_0} n_2 \qquad [17]$$

Therefore, n_2 can be estimated as

$$n_2 = \left(\frac{\pi n_0}{\eta_0 L}\right)\left(\frac{d}{2k_0^2 w_0 P}\right)^2 \left(\frac{f}{r_a}\right)^2 \frac{dP_T}{d\Delta s} \qquad [18]$$

Thus, if one knows the slope of the z-scan graph around $\Delta s \approx 0$, the value of n_2 can be found. Another way to state this result is to say that the n_2 is related to the peak-to-valley change in the on-axis transmittance ΔT_{p-v}, which is proportional to $dP_T/d\Delta s$.

In passing, we comment that the z-scan method can be also adapted to determine the nonlinear absorption coefficient α_2 of a material, defined through the relation $\alpha = \alpha_0 + \alpha_2 I$, where α is the total absorption coefficient and I denotes the intensity. In order to achieve this, the aperture is assumed to be completely open, in order to collect all the light behind the nonlinear absorptive sample. For $\alpha_2 > 0$, maximum absorption is experienced when the nonlinear sample, assumed thin, is at the back focal plane of the lens since the optical intensity is a maximum due to the external focusing. The z-scan graph is an even function of scan distance, measured about the back focal plane of the external lens.

Thick Sample *z*-Scan

Contrary to the thin sample case, a thick sample is defined as one whose thickness may be comparable to the depth of focus of the externally focused Gaussian beam. In this case, the beamwidth changes during propagation through the sample, so that the induced focal length also changes with propagation. This makes the problem substantially more involved. Indeed, to track the far-field profile in this case, one needs to describe the propagation of the Gaussian beam inside the nonlinear sample using the q parameter formalism. Since both the width and the radius of curvature of the beam at the exit face of the sample are required to predict the far-field behavior, one can write a simple differential equation describing the variation of the q through the nonlinear sample. This equation can be composed from the laws of q-transformation described earlier, and the effects of propagation and nonlinear lensing can be added assuming each is small over an infinitesimal distance of propagation, much like the numerical method called the split step beam propagation method (BPM),

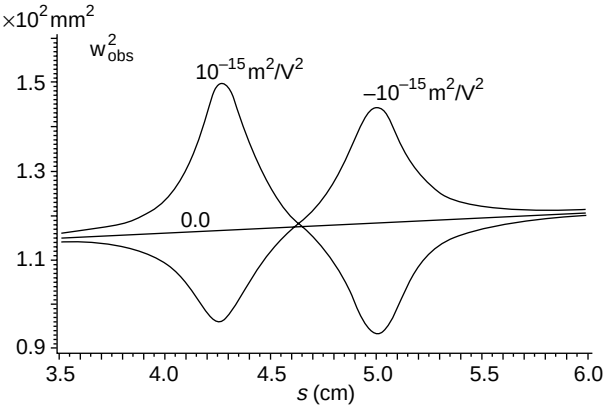

Figure 3 Plot of width of the beam in the far field as a function of the scan distance s. The light wavelength is 640 nm. Reproduced with permission from Banerjee PP (2004) *Nonlinear Optics: Theory, Numerical Modeling, and Applications*. New York: Marcel Dekker.

often used to analyze such problems. We will not present details of the computation, rather show calculation results for a typical thick sample case, and point out differences with the thin sample predictions for z-scan.

Figure 3 shows the beamwidth as a function of scan distance for the case where the observation plane is at a distance $d = 1$ m from the exit face of a 1 cm thick. Assume that a hypothetical sample with $n_0 = 1$, $n_2 = \pm 10^{-15}$ m^2/V^2 is scanned about the back focal plane of an external lens with focal length $f = 5$ cm. For comparison, the case of a linear sample is also superposed on the same graph. Note that with the variation of s, the far-field width initially increases, and then decreases before starting to increase again, for a material with a positive n_2. Note that around the point where the variations intersect the graph for the linear case, the slope of the curves are negative and positive for materials with positive and negative n_2, respectively. The graphs for the variation of on-axis transmittance are opposite to the variation of the widths, as mentioned earlier. This is also in agreement with the results predicted from geometrical optics and the q formalism for a thin sample, described earlier.

How do thick and thin sample z-scans differ? Important differences are: (i) the position of the points where the graphs for positive and negative, n_2s, intersect the graph for the linear case; and (ii) the slope of the graphs, predicted from thin and thick sample calculations. For instance, for the thin sample approximation, the value of the slope around $s = f$ is about -4 for a material with a $n_2 = 10^{-15}$ m^2/V^2. Recall from the previous section, the z-scan plot for a thin sample always intersects the graph for the linear case at the back focus of the external lens,

and therefore, it is easy to determine the slope of the z-scan at this point. From the graph in the thick sample case, it is perhaps difficult to ascertain at exactly which point one must determine the slope. A figure of merit would be the minimum value, which occurs for a value of s slightly smaller than $f = 5$ cm. A conservative estimate yields a slope of approximately -1 for this particular example. Note, however, that in the thick lens analysis, we have taken a paraxial approximation to the Gaussian at every stage of its propagation through the nonlinear sample. Strictly speaking, this is not true, and to alleviate this problem, Van Stryland's group included a correction factor with the width of the Gaussian beam.

We would like to mention that values of $n_2 \approx -10^{-15}$ m^2/V^2 can indeed be measured for thermal nonlinearities in liquids. In fact, this is a typical value that we measured for our solution of chlorophyll in ethanol during our first experiments on z-scan. Values of effective $n_2 \approx 10^{-14}$ m^2/V^2 can be measured for liquid crystals with orientational nonlinearities, and will be described below.

Example of z-Scan: Effective Cubic Nonlinearity of Liquid Crystals

The induced refractive index in a nematic liquid crystal, due to reorientational nonlinearity, is a function of position in the crystal and depends on the intensity of the incident optical plane wave, its angle with respect to the direction of propagation, and on the applied voltage. Because of the fact that the nonlinearity depends on the direction of propagation of the planewave and the director axis of the crystal, modeling Gaussian beam propagation is rather complicated. Moreover, the director axis has a certain distribution in the material due to the applied voltage, and hence, the electrostatic field in the material, as well as the optical field. This poses a level of complexity a degree higher than that discussed above for the simple thick sample case. Since the induced refractive index is not a constant, and more importantly, a function of the spatial frequency of the optical beam (which determines the angle between a planewave component of the Gaussian beam spectrum and the director axis distribution under external field), one is compelled to analyze optical propagation using split-step BPM. A typical set of simulated z-scan plots is shown in Figure 4, assuming one transverse dimension for simplicity, and the same parameters as in the experiment described below. Actual experimental results from a z-scan of a biased liquid crystal sample is shown in Figure 5 for transmitted power versus

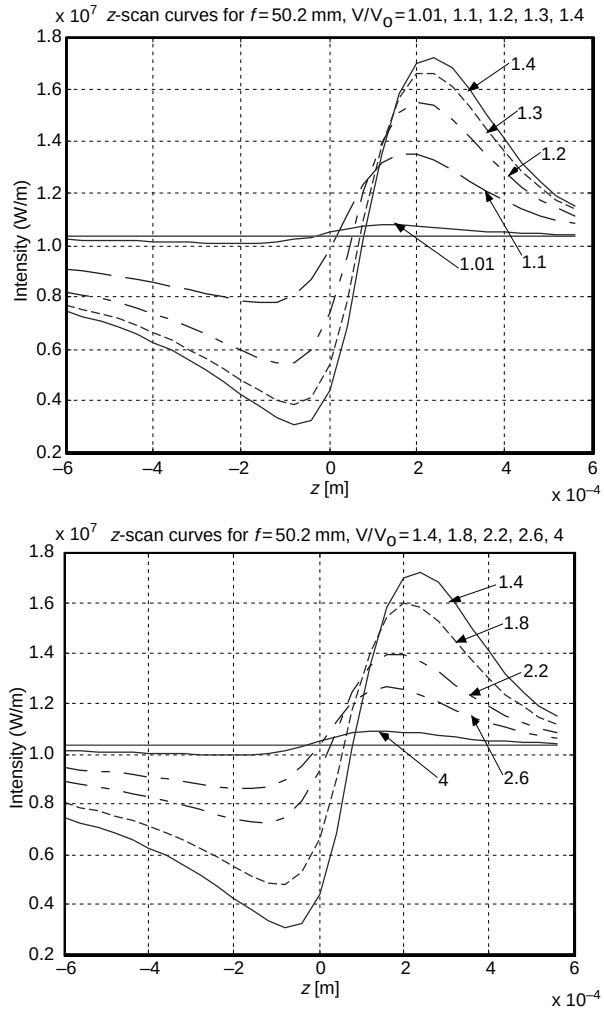

Figure 4 z-scan simulation results for $f = 50.2$ mm and $V/V_o = 1.01$, 1.1, 1.2, 1.3, 1.4 with $E_{op} = 2 \times 10^5$ V/m. z-scan simulation results for $f = 50.2$ mm and $V/V_o = 1.4$, 1.8, 2.2, 2.6, 4 with $E_{op} = 2 \times 10^5$ V/m. Reproduced with permission from Banerjee PP (2004) *Nonlinear Optics: Theory, Numerical Modeling, and Applications.* New York: Marcel Dekker.

sample position in a z-scan experiment for a liquid crystal sample for applied voltages 0.8, 1.0, 1.2, 1.8, and 2.2 V. In this case, the focal length of the external lens is approximately 50 mm and aperture diameter on the observation plane is 1 mm. The scan range is 0.15 inches or approximately 4 mm. The scan step, ΔZ, is varied from about 0.002 inches to about 0.01 inch, depending on the position of the sample. This is done so as to get the maximum information around the back focal plane of the external lens where the measurement values are to be used to find the slope of the z-scan graph, and hence the effective nonlinear refractive index coefficient.

The plots of the effective $n_2 = n_{2\text{eff}}$ (simulated and measured) are shown in Figure 6a and b, respectively. As seen from both plots, the $n_{2\text{eff}}$ increases and

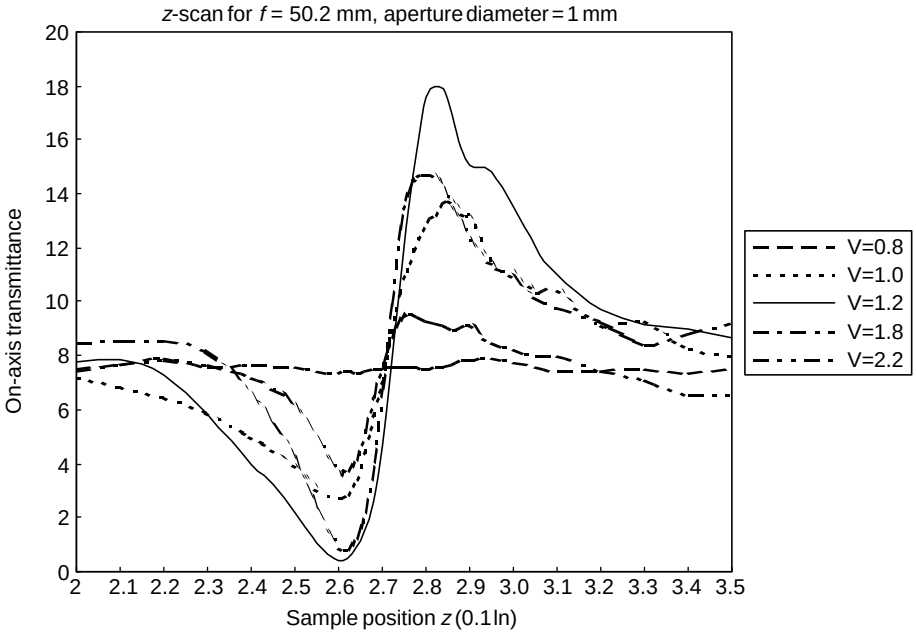

Figure 5 Measured z-scan graph for $f = 50.2$ mm and aperture diameter $= 1$ mm. Reproduced with permission from Banerjee PP (2004) *Nonlinear Optics: Theory, Numerical Modeling, and Applications.* New York: Marcel Dekker.

reaches a maximum at $V/V_o = 1.4$, before starting to decrease. Simulated and experimental $n_{2\text{eff}}$ have similar trends. Furthermore, there is reasonable agreement between the calculated and measured values of $n_{2\text{eff}}$.

Discussion: Effective n_2 of Other Materials

The z-scan is a powerful tool in the measurement and characterization of effective cubic nonlinearities of different materials. The method is simple and easy to set up, and can be readily automated. We provide two other examples of z-scans below, as examples. For instance, photorefractive materials are commonly used for dynamic holography, in order to record and reconstruct volume holograms. Photorefractive materials are doped light-sensitive electro-optic materials in which optical radiation creates mobile charge carriers which redistribute themselves to create an induced space-charge field. This electrostatic field in turn modulates the refractive index which phase modulates the optical beam. A figure of merit for the effectiveness of such a material, e.g., lithium niobate, for holographic storage, is the donor to acceptor ratio in such materials. It turns out that this ratio, along with the photovoltaic coefficient, can be related to an effective n_2 in lithium niobate. This is also responsible in producing an elliptic beam in the far field by starting out from a circularly symmetric Gaussian beam focused onto the sample. The ellipticity stems from the tensor nature of the electro-optic coefficients. Typical simulated z-scan plots for lithium niobate are shown in Figure 7, and carries the signature variation of the scan in the variation of the beam ellipticity with scan distance. From this plot, we are able to determine the effective n_2 and hence, the donor to acceptor concentration ratio.

In other photorefractive materials, such as barium titanate, in which diffusion of charge carriers plays a dominant effect, the induced change in the refractive index is proportional to the gradient of the intensity, rather than the intensity itself, as was the case in all previous examples. In this case, as seen in Figure 8, it can be shown that the z-scan plot for on-axis transmittance with a small finite aperture on the observation plane is an even function of scan distance. This, however, should not be confused with the open aperture z-scan that is also even, but which yields information of the nonlinear absorption coefficient, as described earlier. Experiments with a new material, manganese aluminum oxide, which is thought to have both photovoltaic and diffusive contribution to space charge field distribution, yields a z-scan that has both odd and even components. Through careful analysis of the z-scan results, one is able to determine the relative contributions of each of these effects. Care should be taken in the interpretation of the results in this case, since in an actual experiment, the aperture size cannot be infinitely small, and thus, an optimum aperture size must be determined that allows for sufficient detected power,

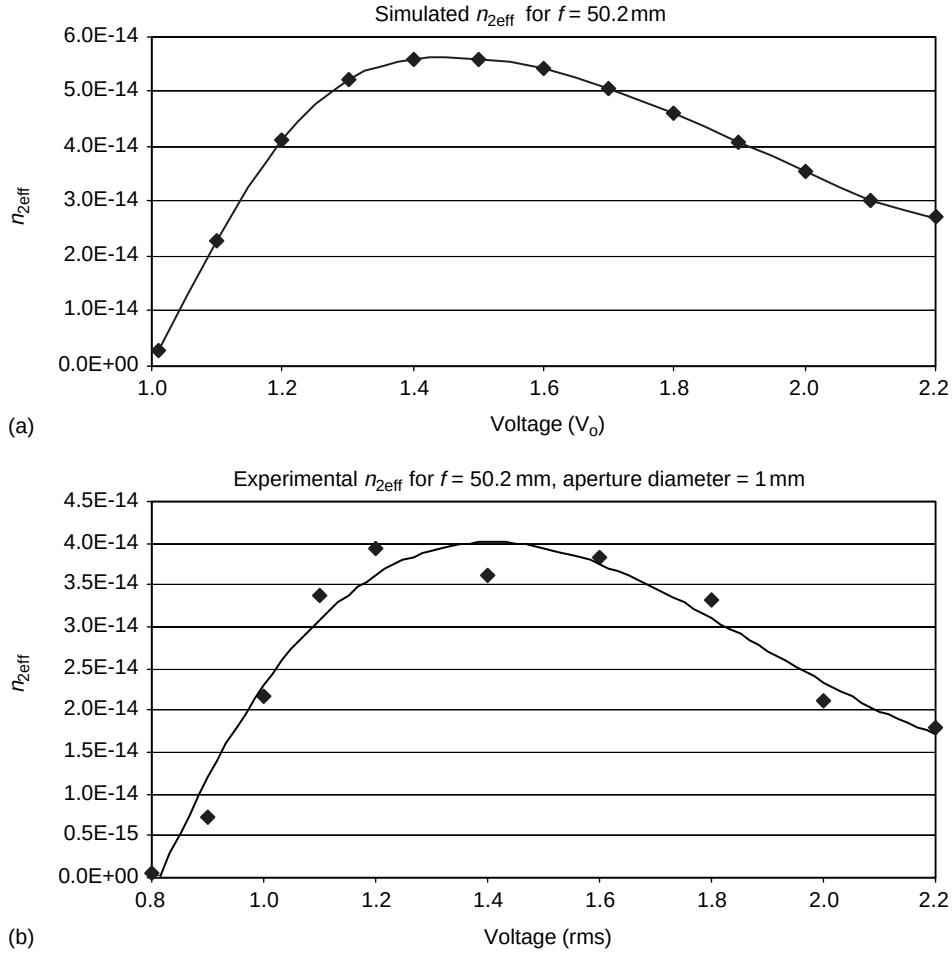

Figure 6 (a) Calculated n_{2eff} for $f = 50.2$ mm with $E_{op} = 2 \times 10^5$ V/m; (b) Experimental n_{2eff} for $f = 50.2$ mm with aperture diameter = 1 mm. Reproduced with permission from Banerjee PP (2004) *Nonlinear Optics: Theory, Numerical Modeling, and Applications.* New York: Marcel Dekker.

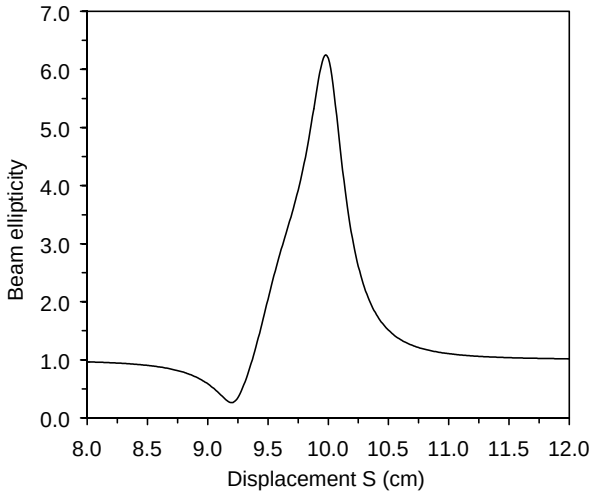

Figure 7 Typical simulated *z*-scan plot showing variation of beam ellipticity as a function of scan length. Reproduced with permission from Banerjee PP (2004) *Nonlinear Optics: Theory, Numerical Modeling, and Applications.* New York: Marcel Dekker.

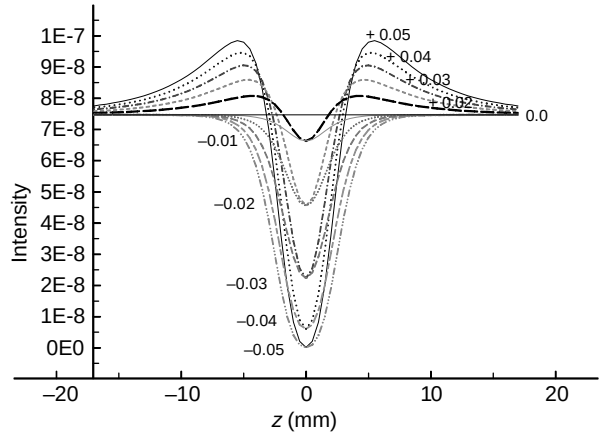

Figure 8 Typical *z*-scan plots of a diffusion dominated photorefractive material, with $\Delta n = C\nabla I$. Different values of C are written on the graph.

but at the same time allows one to neglect any effects of nonlinear absorption.

In summary, it is hoped that this article illuminates readers on the history of the z-scan, its application to thin and thick samples, its effectiveness in measurement and characterization of effective third-order optical nonlinearities for a wide class of nonlinear materials, such as $\chi^{(3)}$ materials, materials with cascaded second-order nonlinearities, liquid crystals, photorefractives, etc., and its potential for automation, while at the same time pointing out areas where care should be taken in the interpretation of the measured data. With appropriate modification in the theory, it can also be applied to characterization of linear lenses, possibly voltage controllable lenses and lenslet arrays, suitable for dynamic imaging in biomedical applications.

See also

Materials for Nonlinear Optics: Liquid Crystals for NLO. **Nonlinear Optics, Basics:** Four-Wave Mixing; $\chi^{(3)}$–Third-Harmonic Generation.

Further Reading

Banerjee PP (2004) *Nonlinear Optics: Theory, Numerical Modeling, and Applications*. New York: Marcel Dekker.

Banerjee PP, Danilieko A, Hudson T and McMillen D (1998) P-scan analysis of induced inhomogeneous optical linearities. *Journal of the Optical Society of America* 15: 2446.

Banerjee PP, Misra RM and Maghraoui M (1991) Theoretical and experimental studies of propagation of beams through a finite sample of a cubically nonlinear material. *Journal of the Optical Society of America* 8: 1072.

Boyd R (2002) *Nonlinear Optics*, 2nd edn. New York: Academic Press.

Butcher P and Cotter D (1990) *The Elements of Nonlinear Optics*. New York: Cambridge University Press.

Chapple PB, Staromlynska J and McDuff RG (1994) *Z*-scan studies in the thin- and the thick-sample limits. *Journal of the Optical Society of America B* 11: 975.

Noginov M, Helzer SW, Loutts GB, Banerjee PP, Morrisey M and Kim Y (2003) Study of photorefraction response and diffraction efficiency in Mn:$YalO_3$ crystals. *Journal of the Optical Society of America B* 20: 1233.

Sheik-Bahae M, Said AA and Van Stryland EW (1989) High-sensitivity, single beam n_2 measurements. *Optics Letters* 14: 955.

Sheik-Bahae M, Said AA, Wei TH, Hagan DJ and Van Stryland EW (1990) Sensitive measurement of optical nonlinearities using a single beam. *IEEE Journal of Quantum Electronics* 26: 760.

Sutherland RL (1996) *Handbook of Nonlinear Optics*. New York: Marcel Dekker.

Yeh P (1993) *Introduction to Photorefractive Nonlinear Optics*. New York: Wiley.

MATERIALS FOR NONLINEAR OPTICS

Contents

Liquid Crystals for NLO
Organic Nonlinear Materials

Liquid Crystals for NLO

I C Khoo, Pennsylvania State University, University Park, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Liquid crystals are complex organic molecular systems that exhibit various mesophases characterized by the degree of order. In this review, we shall focus mainly on thermotropic liquid crystals which exhibit various mesophases in the temperature range between the crystalline and the isotropic liquid states. In particular, we shall discuss the extraordinarily large optical nonlinearities of nematic liquid crystals, as well as the electronic nonlinearities of isotropic-phase liquid crystals. Recent studies have shown that these materials are promising candidates for advanced photonic applications.

Nematic liquid crystals (NLC) possess directional ordering, i.e., when placed in thin cells with surface

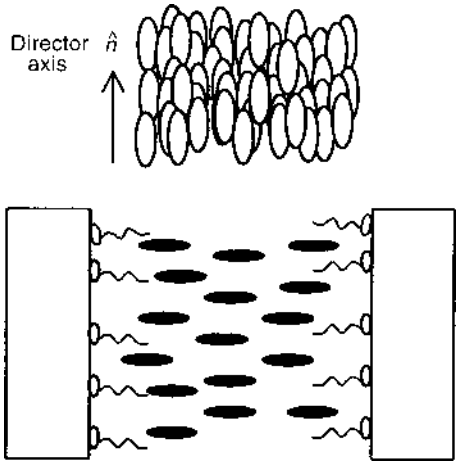

Figure 1 Nematic liquid crystals (NLC) which exhibit directional ordering characterized by the director axis. The lower figure shows how the NLC molecules are aligned in a typical cell with surfactant coated windows.

treated boundaries, the NLC molecules will self-assemble into a birefringent crystal characterized by a so-called director axis $\hat{n}$, cf. Figure 1. For light polarized parallel to $\hat{n}$, the refractive index is n_e, while for light polarized perpendicular to $\hat{n}$, the refractive index is n_0. Typically n_e is ~1.7 and n_0 is ~1.5. This large birefringence $\Delta n = n_e - n_0$ of NLC holds throughout the near UV to the far infrared spectral regime, cf. Figure 2. Perhaps the most important characteristics of NLCs is the easy susceptibility of the birefringent director (crystalline) axis orientation to externally applied fields, which has led to their ever increasing widespread use in various opto-electronics information and image display and processing devices. In analogy to dc and ac field effects, the director can easily be reoriented by an optical field, and thus cause a large optically induced index of refraction change. By doping the NLC with photosensitive or photo-charge producing agents, this process of refractive index change can be enhanced considerably. The refractive indices of NLC are also very strongly temperature dependent.

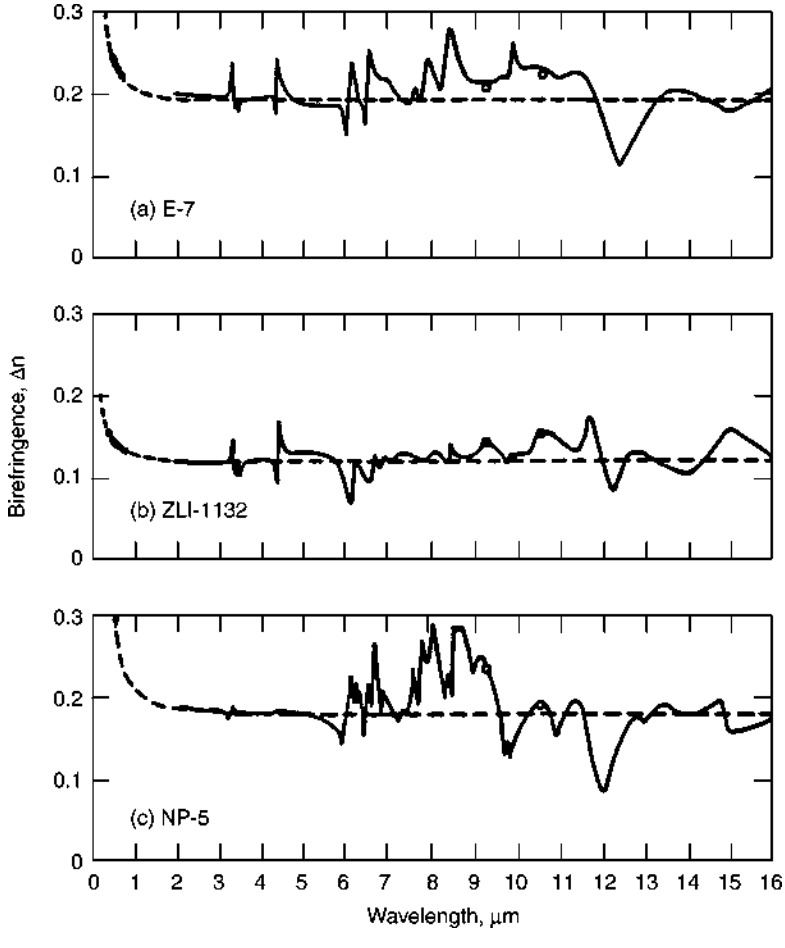

Figure 2 Broadband birefringence of nematic liquid crystals from the near ultraviolet to far infrared.

Accordingly, a large refractive index change can also be generated with laser absorption induced density and order parameter changes. In the following sections, we will broadly classify these nonlinear optical processes originating from the collective response of NLC to external fields as *nonelectronic nonlinearities*.

On the other hand, individual NLC molecules also respond to an optical field through the electronic dipolar interaction. A typical electronic energy level structure of liquid crystal molecules is shown in Figure 3. The dipolar interaction gives rise to the usual nonlinear polarization characterized by the second, third and higher-order nonlinear susceptibilities $\chi^{(2)}, \chi^{(3)}, \ldots$ etc. The complex imaginary parts of these linear and nonlinear susceptibilities account for the photonic absorption processes. Most liquid crystalline molecules possess small linear absorption in the visible region (400–700 nm), and begin to absorb near the UV region (<350 nm), cf. Figure 4. This absorption spectrum translates into a (usually) large nonlinear two-photon absorption spectrum in the visible. In some organic molecules, the intermediate absorption cross-section σ_i or the excited state absorption cross-section σ_e is much larger than the ground state linear absorption σ_g, giving rise to reverse saturable absorption (RSA) and excited state

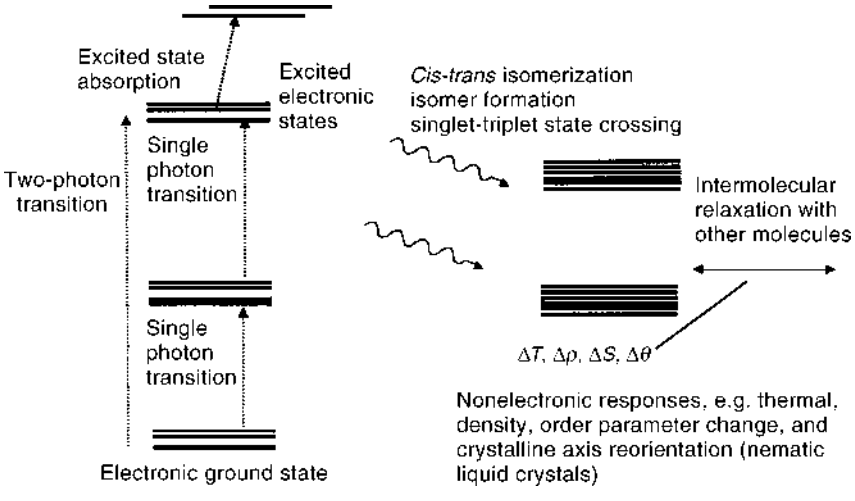

Figure 3 Schematic depiction of the typical energy level structure of liquid crystal molecules and various linear and nonlinear optical processes.

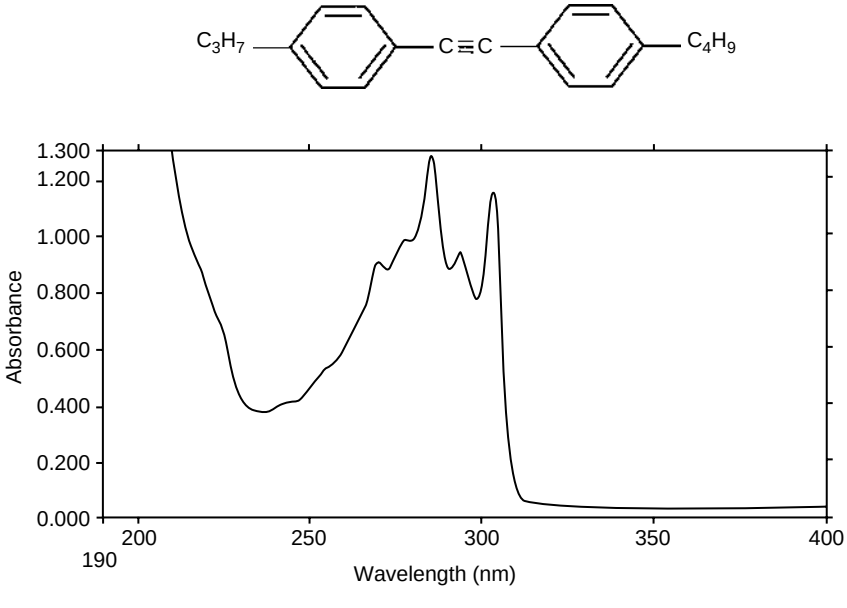

Figure 4 Molecular structure and absorption spectrum of typical isotropic liquid crystal molecules.

absorption (ESA), respectively. As we will see presently, these so-called *electronic nonlinearities* are quite useful for optical limiting applications.

Following photo-absorption, the liquid crystal molecules could undergo intersystem crossing, *cis-trans* or other conformational changes, and form isomers or charge transfer complexes. These processes will contribute to the nonelectronic nonlinearities such as reorientational, order parameter changes and thermal effects.

Orientational Nonlinearities

Purely Optically Induced Orientation

Figure 5 depicts schematically various processes that could occur in an aligned nematic film upon irradiation by an optical field. The reorientation of the director axis $\hat{n}$ is governed by the system's tendency to minimize the overall free energy, which consists of several bulk and surface interaction terms, and the interaction energy with the applied field. In purely dielectric nonabsorbing nematic liquid crystals, director axis reorientation is caused by the torque Γ_{opt} between the anisotropic nematic and the optical electric field,

$$\Gamma_{\text{opt}} = \left(\frac{\Delta\varepsilon}{8\pi}\right)(\hat{n}\cdot\mathbf{E}^*)(\hat{n}\times\mathbf{E}) \qquad [1]$$

which competes with the elastic restoring torque Γ_{el} of the surrounding NLC molecules:

$$\Gamma_{\text{el}} = \frac{(K_1\sin^2\theta + K_3\cos^2\theta)\mathrm{d}^2\theta}{\mathrm{d}z^2} + \left[(K_1 - K_3)\sin\theta\cos\theta\right]\left(\frac{\mathrm{d}\theta}{\mathrm{d}z}\right)^2 \qquad [2]$$

where the K's are the elastic constants, and $\Delta\varepsilon$ is the (optical) dielectric anisotropy.

As a result of the director axis reorientation, an optical wave propagating through the medium will 'see' a refractive index change. A frequently employed method for measuring the refractive index change in nematic liquid crystal films is the grating diffraction setup, as shown in Figure 6a. Two polarized coherent writing beams are overlapped on a nematic film, which is probed by a polarized reading beam. If the writing beams are copolarized, they impart an optical intensity grating onto the nematic film $I = I_0(1 + \cos qx)$, where q is the grating wavevector, cf. Figure 6b. On the other hand, if the writing beams are orthogonally polarized, the resulting intensity is uniform over the illuminated area, and instead a polarization grating is created, cf. Figure 6c. The polarization state of the light in the overlapped region varies from circular, through elliptical to linear in accordance to the phase difference between the two coherent writing beams. Using a linearly polarized probe beam, this setup allows one to distinguish various intensity and reorientation dependent effects, and assess their contribution to the induced refractive index change Δn.

The magnitude of the induced reorientation angle, and therefore the extraordinary index change, depends on whether the interaction is transient or steady state. In the steady state case, to the first order of approximation, Δn is proportional to the optical intensity I, i.e., $\Delta n = n_2^{\text{SS}} I$. Under some experimental conditions, one can also show that the transient nonlinear index coefficient $n_2\ (\tau_{\text{p}})$ is given by $n_2\ (\tau_{\text{p}}) \sim n_2^{\text{SS}}\tau_{\text{p}}/\tau_{\text{r}}$, where $\tau_{\text{r}} = \gamma d^2/K\pi^2$ is the orientational relaxation time constant. For $d = 100\ \mu\text{m}$, $\gamma = 0.1$ poise and $K = 10^{-6}$ dynes, we have $\tau_{\text{r}} \sim 1$ s. Typical observed values of $n_2^{\text{SS}}(\theta)$ are

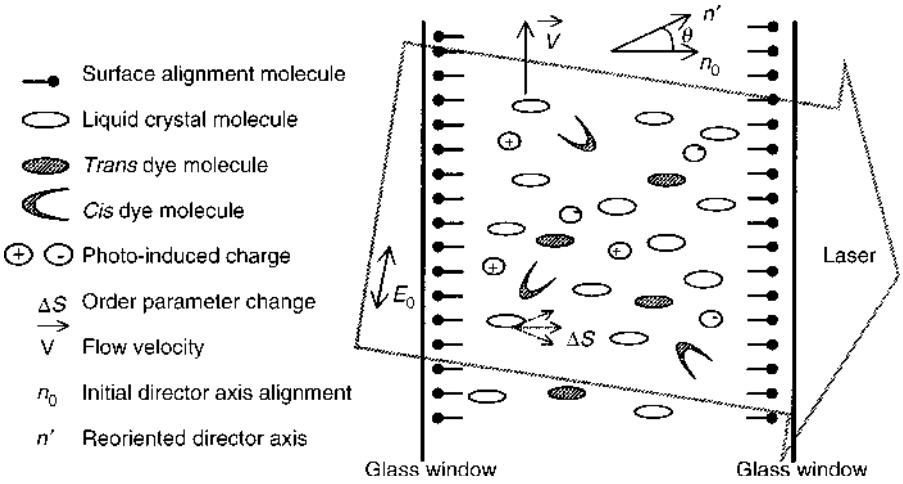

Figure 5 Schematic depiction of various laser-induced optical processes in a nematic liquid crystal film.

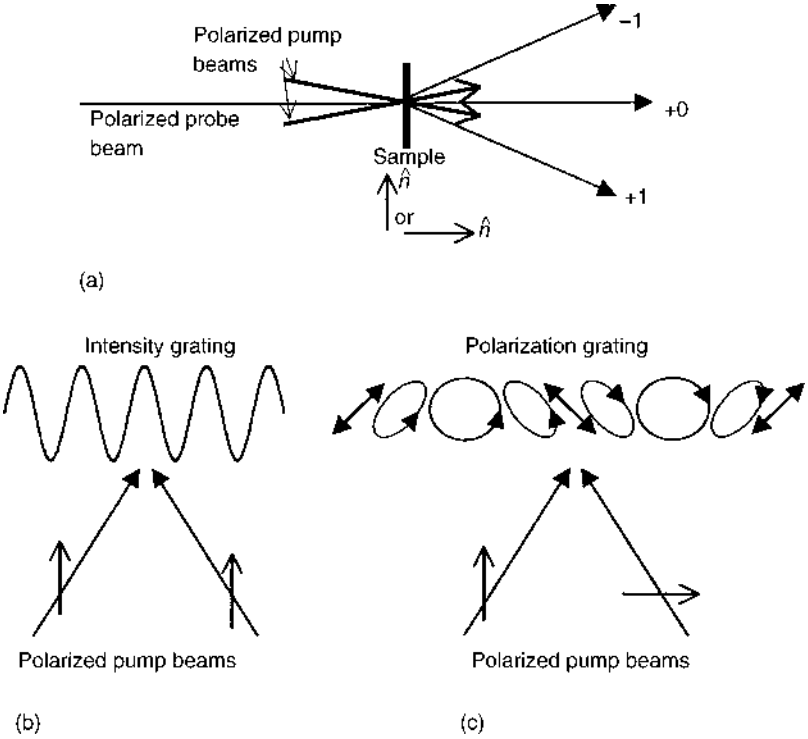

Figure 6 (a) Schematic depiction of grating formation and probe diffraction from the index grating induced in a nematic film. (b) Optical intensity grating formed by interfering two coherent copolarized writing beams. (c) Optical polarization grating formed by interfering two coherent crossed-polarized writing beams.

in the range of 10^{-5}–10^{-4} cm^2 W^{-1}. In the transient case, if $\tau_p = 1$ ns, then n_2 $(\tau_p) \sim 10^{-12}$ cm W^{-1} [$\chi^{(3)} \sim 2 \times 10^{-11}$ esu].

Excited Dye-Dopant Induced Intermolecular Torque

Studies of nematic liquid crystals doped with anthraquinone or dichroic dyes have shown that the excited dye molecules could exert an intermolecular torque Γ_{mole} that could be much larger in magnitude than Γ_{op}. By considering the interaction energies between ground and excited dyes with the surrounding LC molecules, one can show that Γ_{mole} may be written as:

$$\Gamma_{mole} = \zeta \Gamma_{op} \qquad [3]$$

where the proportionality constant ζ can be much larger than unity. The nonlinear index coefficients n_2 obtained in these films are on the order of 10^{-3} cm^2 W^{-1}. The response times for the director axis reorientation and relaxation are similar to the optically induced reorientation process discussed in the preceding section.

Photorefractive Effect

Another mechanism, namely, orientational photorefractivity, could also result in a very large optical nonlinearity. This effect occurs in nematic liquid crystals doped with photo-charge producing agents such as fullerene C_{60} or rhodamine 6G dye, in conjunction with a small applied dc field. For a 25 μm thick film with a dc bias voltage of ~1 V, the observed nonlinear coefficient n_2 is in the range of 10^{-3}–10^{-2} cm^2 W^{-1}. Figure 7 illustrates the orientational photorefractive effect. An incident optical field excites the doped liquid crystals to form charge transfer complexes (CTCs). The CTCs subsequently dissociate and create dc space charge fields through ionic diffusion, migration and other electrodynamical processes. The space charge fields, in combination with the applied dc field, create a dc field induced torque of the form:

$$\Gamma_{dc} = \left(\frac{\Delta\varepsilon_{dc}}{8\pi}\right)(\hat{n}\cdot\mathbf{E}^*_{int})(\hat{n} \times \mathbf{E}_{int}) \qquad [4]$$

Here the total internal dc field $\mathbf{E}_{int}$ contains the applied dc field $\mathbf{E}_{dc}$ and the generated dc space charge fields $\mathbf{E}_{sc}$. This causes director axis reorientation and refractive index change.

The space charge fields consist of one photorefractive and two anisotropy components. For an incident optical intensity grating function of the form $I_{op} = I_0(1 + m\cos q\xi)$, the photorefractive space charge

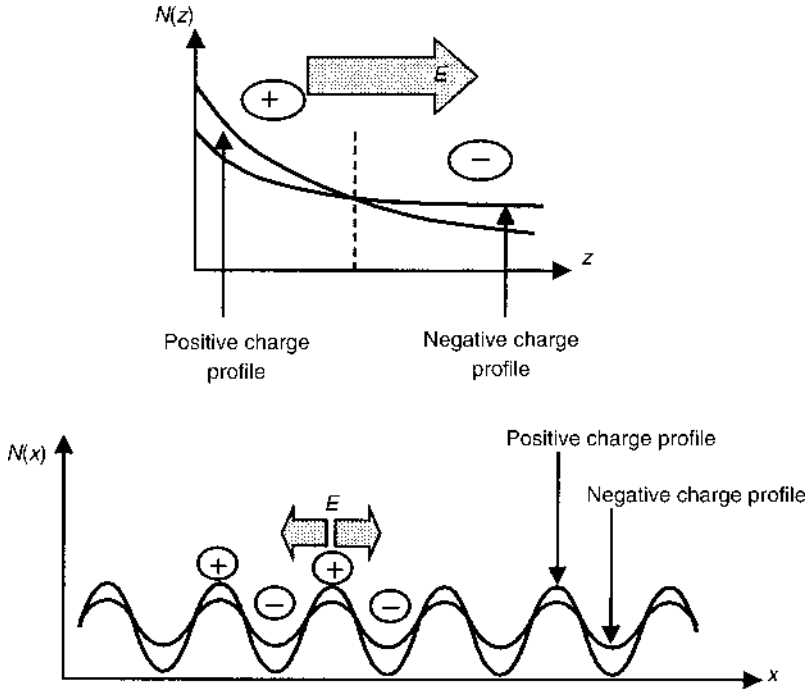

Figure 7 Schematic depiction of photo-induced space charge field formation in a nematic liquid crystal illuminated by a monotonic (upper curve) or a sinusoidal (lower curve) laser intensity function.

field is given by:

$$E_{\mathrm{ph}} = q\nu\, mk_{\mathrm{B}}T\left(\frac{\sigma - \sigma_{\mathrm{d}}}{2e\sigma}\right)\cos\left(\frac{q\xi - \pi}{2}\right) \qquad [5a]$$

where k_{B} is the Boltzmann constant, σ is the illuminated conductivity, σ_{d} is the dark state conductivity, $\nu = (D_+ - D_-)/(D_+ + D_-)$, where D_+ and D_- are the diffusion constants for the positively and negatively charged ions, respectively; m is the optical intensity modulation factor, $q = 2\pi/\Lambda$ is the magnitude of the grating wavevector, and ξ is the coordinate along q.

Two other sources of space charge fields are also at work, and they arise from the conductivity and dielectric anisotropies of the nematic under the action of the dc bias field E_{dc}:

$$E_{\Delta\sigma} = E_{\mathrm{dc}}\left(\frac{(\sigma_{//} - \sigma_{\perp})\sin\theta\cos\theta}{\sigma_{//}\sin 2\theta + \sigma_{\perp}\cos 2\theta}\right) \qquad [5b]$$

$$E_{\Delta\varepsilon} = E_{\mathrm{dc}}\left(\frac{(\varepsilon_{//} - \varepsilon_{\perp})\sin\theta\cos\theta}{\varepsilon_{//}\sin 2\theta + \varepsilon_{\perp}\cos 2\theta}\right) \qquad [5c]$$

where $(\sigma_{//} - \sigma_{\perp})$ is the low-frequency **conductivity anisotropy** and $(\varepsilon_{//} - \varepsilon_{\perp})$ the **dielectric anisotropy** of the liquid crystal, and θ is the reorientation angle of the director axis.

In addition to these space charge field induced effects, the **flows** of the charged ions and complexes under the applied dc field and the generated space charge fields also contribute significantly to reorienting the director axis. For an interaction geometry similar to Figure 5, the flow creates a shear torque Γ_{shear} of magnitude proportional to the velocity gradient

$$\Gamma_{\mathrm{shear}} = \frac{\mathrm{d}\mathbf{v}(\zeta)}{\mathrm{d}\zeta}(\alpha_2\sin^2\theta - \alpha_3\cos^2\theta) \qquad [6]$$

where the α's are the viscosity coefficients, and $\mathbf{v}(\zeta)$ is the flow velocity. The combined action of the above space charge fields and the flow effects on director axis reorientation have been quantitatively formulated and shown to describe experimental results very well, including, in particular, the observation that maximal diffraction is obtained for a grating constant $\lambda_{\mathrm{g}} \sim 2d$. Furthermore, the flow-reorientation and adsorption of the excited dopant molecules on the surfaces are believed to be responsible for quasi-permanent and permanent director axis reorientation in rhodamine 6G- and C_{60}-doped nematic films, respectively.

Supra-Nonlinear Methyl-Red Doped NLC – Observed Phenomena

Recent studies in methyl-red doped nematic liquid crystals (MRNLCs) have ushered in the era of supra-nonlinearity. Unlike photorefractive nematic liquid

crystals, no external bias field is needed. The observed n_2 values could be as large as $10\,\mathrm{cm}^2\,\mathrm{W}^{-1}$. In MRNLC, a large photovoltage drop across the sample is also observed upon irradiation. The polarity of the photovoltage is reversed if the light impinges from the opposite cell window. The extremely nonlinear optical response of MRNLC is attributed to nematic director axis reorientation by the optically induced space charge fields, similar to the effect discussed in the previous section. The unusually large response is possibly due to several factors, including the higher photo-charge producing abilities of methyl red, and the large difference in the positive and negative photo-ion diffusion constants. Another cause of director axis reorientation is the intermolecular torque exerted by the photo-excited dye molecules on the NLC. This is particularly evident in experiments using an input polarization grating on planar aligned samples. Because the laser excited dye molecules are adsorbed and strongly 'anchored' on the cell walls, they cause permanent realignment of the nematic liquid crystal. This process usually occurs upon prolonged illumination of the liquid crystal cells with a low-power laser, and can be sped up with higher laser power.

Trans–Cis Isomerization and Order Parameter Change

Some nematic liquid crystals are azo-compounds, i.e., their molecular structures are analogous to the typical one depicted in Figure 8. An azo-compound when irradiated at its absorption wavelength will change from the *trans* to the *cis* form, cf. Figure 8. *Trans–cis* isomerization of these azo-nematic liquid crystals or azo-dopants in NLC's could also give rise to a refractive index change through modification of the NLC's order parameter and/or director axis reorientation. The *trans* species are oblong in shape, and align parallel to the director axis. On the other hand, the 'bent' structure of the *cis* species is incongruous to the nematic order, and thus causes an order parameter change, and therefore a refractive index modification. Such refractive index changes owing to order parameter modification are similar to thermal effects discussed in the next section. The response [on-] time for these *trans–cis* isomerization induced order-parameter changes can be quite fast (as fast as nanoseconds in some cases). The relaxation time (typically in the millisecond regime) is decided by the director axis relaxation rate as discussed earlier.

The other index-changing mechanism is the tendency of the *trans* molecule to align with its transition moment perpendicular to the optical polarization in order to minimize the interaction (absorption) energies. This in turn causes the NLC director axis to realign through the intermolecular torque effect. Both mechanisms are characterized by a nonlinear index coefficient n_2 on the order of $10^{-3}\,\mathrm{cm}^2\,\mathrm{W}^{-1}$.

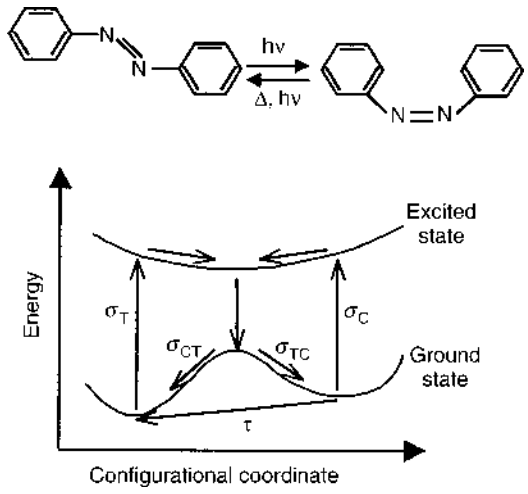

Figure 8 *Trans–cis* isomerization of azo-molecules upon optical illumination. The upper diagram depicts molecular structural changes; the lower diagram shows the excitation energy versus the molecular coordinate associated with these photo-excited processes.

Thermal and Density Effect

Laser induced temperature and density changes are described by time-dependent coupled hydrodynamics equations. In grating diffraction experiments, cf. Figure 6, the problem can be simplified to a one-dimensional one, i.e., along the grating wavevector direction, if the grating spacing is smaller than the nematic film thickness and the beam size. In that case, the thermal and Brillouin (sound wave) decay time constants are given by, respectively:

$$\tau_T = \frac{\rho_0 C_v}{\lambda_T q^2} \qquad [7a]$$

$$\tau_B = \frac{2\rho_0}{\eta q^2} \qquad [7b]$$

where $q = 2\pi/\lambda_q$ is the grating wavevector, ρ_0 the unperturbed density, C_v the specific heat, λ_T the thermal conductivity, and η the viscosity. The values of these parameters depend on the director axis alignment relative to the optical fields, thermal diffusion and sound propagation directions. For typical values of $\eta = 7 \times 10^{-2}\,\mathrm{kg\,m^{-1}\,s^{-1}}$, $\rho_0 = 10^3\,\mathrm{kg\,m^{-3}}$, $D = \lambda_T/\rho_0$, $C_v \sim 1.6 \times 10^{-7}\,\mathrm{m^2\,s^{-1}}$ and a grating spacing $\lambda_q = 20\,\mu\mathrm{m}$, we have $\tau_T \sim 50\,\mu\mathrm{s}$, whereas $\tau_B \sim 200\,\mathrm{ns}$. Such a vast difference between the relaxation times allows one to distinguish the contribution from thermal and density origins.

Table 1 Nonelectronic nonlinearities of nematic liquid crystals

Nonlinear processes	*Typical magnitude*	*Time-scale*	*Liquid crystal*
Orientational nonlinearities			
Purely optically induced n_2	10^{-4} cm^2 W^{-1}	ms	pure 5CB
Excited dopant assisted n_2	10^{-3} cm^2 W^{-1}	ms	dyed 5CB
Photorefractive	10^{-3} cm^2 W^{-1}	ms	5CB
Photorefractive + Methyl-red doped 5CB	10 cm^2 W^{-1}	ms	
Trans–Cis and order parameter modification	10^{-4} cm^2 W^{-1}	ms	dyed 5CB
Thermal nonlinearities			
n_2	10^{-6} cm^2 W^{-1}	ms	E7
n_2	10^{-11} cm^2 W^{-1}	ns	E7
Density effect			
electrostrictive n_2	10^{-11} cm^2 W^{-1}	ns	E7

In the steady state, i.e., when the laser–nematic interaction time is much longer than τ_T and τ_B, the nonlinear index coefficient is given by:

$$n_2^{SS}(\Delta T) = \left(\frac{dn}{dT}\right)\frac{\alpha\tau_r}{\rho_0 C_V} \qquad [8]$$

where dn/dT is the appropriate thermal index coefficient. The absorption constant α of liquid crystals varies widely in the visible–infrared spectrum. Near the phase transition temperature, $dn_{//,\perp}/dT$ also changes dramatically. The magnitude of $n_2^{SS}(\Delta T)$ could therefore vary by several orders of magnitude. For $\alpha = 100\ \text{cm}^{-1}$, $dn_{//,\perp}/dT \sim 10^{-3}\ \text{K}^{-1}$, and $\lambda_q = 20\ \mu\text{m}$, we have $n_2^{SS}(\Delta T) \sim 10^{-6}\ \text{cm}^2\ \text{W}^{-1}$. This is the typical value observed in wave mixing experiments with infrared [CO_2] or visible lasers.

In the transient regime ($\tau_p \ll \tau_r, \tau_B$), the nonlinear index coefficient is proportional to the laser pulse duration:

$$n_2(\tau_p) = n_2^{SS}(\Delta T)\tau_p/\tau_r \qquad [9]$$

With $\tau_r \sim 0.5 \times 10^{-4}$ s and $\tau_p = 1$ ns, $n_2(\tau_p) \sim 5 \times 10^{-11}\ \text{cm}^2\ \text{W}^{-1}$. Such nonlinearities have been observed in transient dynamic grating diffraction studies involving nanosecond laser pulses. In these studies, it is found that the contribution from the electrostrictive effect (movement of the nematic due to the optical intensity gradient) to the refractive index change can be as large as the absorptive thermal component.

In Table 1, we give a concise summary of these nonelectronic optical nonlinearities observed in nematic liquid crystals.

Electronic Nonlinearities

Second and Third Order Nonlinear Susceptibilities

Electronic interaction involves the perturbation of the electronic wavefunctions by the impinging optical fields. The induced dipole moment of a molecule is given by:

$$\mathbf{d} = \boldsymbol{\alpha}:\mathbf{E} + \boldsymbol{\beta}:\mathbf{EE} + \boldsymbol{\gamma}:\mathbf{EEE} + \cdots \qquad [10]$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are the linear, second-order and third-order nonlinear polarizability tensors, $\mathbf{E}$ is the optical electric field and : denotes tensorial operation. The macroscopic polarization $\mathbf{P}$ (dipole moment per unit volume) is

$$\mathbf{P} = \varepsilon_0\boldsymbol{\chi}^{(1)}:\mathbf{E} + \boldsymbol{\chi}^{(2)}:\mathbf{EE} + \boldsymbol{\chi}^{(3)}:\mathbf{EEE} + \cdots \qquad [11]$$

where $\boldsymbol{\chi}^{(1)}$, $\boldsymbol{\chi}^{(2)}$, and $\boldsymbol{\chi}^{(3)}$ are the linear, second and third-order nonlinear susceptibilities, respectively. The macroscopic parameters $\boldsymbol{\chi}$ are related to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by the local field correction factors.

A well-known process associated with the second-order nonlinear polarization is second harmonic generation, in which two incident photons of frequency ω are converted into a photon of frequency 2ω. Symmetry rules require that a material should be non-centrosymmetric in order to exhibit second-order nonlinear polarization. In centrosymmetric nematic liquid crystals, an applied dc field or discontinuity at a surface or interface, or director orientational curvature, could provide the required symmetry-breaking mechanism.

In ferroelectric or smectic C* liquid crystals, the molecules possess permanent polarization, i.e., the centrosymmetry is broken, and it is possible to observe second harmonic generation associated with the second-order nonlinear polarization. Typically, the observed nonlinear coefficients β_{2HG} are on the order of $\sim 10^{-14}\ \text{mV}^{-1}$. In contrast to inorganic crystals of much larger dimensions, these thin liquid crystal films are not useful for practical devices. Nonetheless, because of the sensitivity of second harmonic generation to surface and interface conditions, second harmonic generation has been proven to be an effective spectroscopic tool.

In dc field induced second harmonic generation, the nonlinear polarization involved is actually the

Table 2 Electronic optical nonlinearities of nematic liquid crystals

Nonlinear optical phenomena and nonlinearities	*Magnitude*	*Time-scale (pulse duration of laser used)*
Second harmonic generation (dc field induced),		
$\chi_{//}^{(3)}$	2.7×10^{-12} esu	ns
$\chi_{\perp}^{(3)}$	0.55×10^{-12} esu	ns
Second harmonic generation β_{2HG}	10^{-12} mV^{-1}	ps (ferroelectric)
Two-photon absorption, β	5 cm GW^{-1}	80 ps
Effective nonlinear absorption, β_{eff}	200 cm GW^{-1}	10 ns
Degenerate four-wave mixing		
electronic $\chi^{(3)}(\omega; \omega, -\omega, \omega)$	10^{-12} esu	ns

third-order nonlinear polarization given by $P(2\omega) = \boldsymbol{\chi}^{(3)}E(0)E(\omega)E(\omega)$. Typically measured values of $\boldsymbol{\chi}^{(3)}$ are $\sim 10^{-12}$ esu, expressed in the commonly used cgs units. The conversion factor between cgs and SI unit is given by

$$\boldsymbol{\chi}^{(3)}[\text{esu}] = 3^4 \times 10^{17} \boldsymbol{\chi}^{(3)}[\text{SI unit}] \qquad [12]$$

There is relatively little work done on measuring or studying the third-order nonlinear electronic polarizabilities of nematic liquid crystals. Most of the studies are centered on the molecular correlation effects associated with the nematic–isotropic phase transition. The typical magnitude of electronic $\boldsymbol{\chi}^{(3)}$ is 10^{-12} esu.

Nonlinear Absorption and Limiting

Two-photon absorption occurs where the molecule is excited to a real electronic excited state by simultaneous absorption of two photons from the incident light. Since the absorption is proportional to the square of the optical intensity, the incident light will experience an increasing absorption effect, and its transmission at high intensity will be 'quenched' to an almost constant level. Two-photon absorption is usually characterized by the two-photon absorption coefficient β, which is related to the imaginary part of $\boldsymbol{\chi}^{(3)}$. For liquid crystals, they are on the order of 5×10^{-12}–10^{-11} mW^{-1} (equivalent to 0.5–5 cm GW^{-1} in the more commonly used units).

Liquid crystalline materials in their (less scattering) isotropic phase have been extensively studied in the last few years for application in optical limiting devices as a result of their large nonlinear absorption constant. In the nanosecond time regime, the RSA and ESA effects, cf. Figure 3, act to give rise to a very large effective nonlinear absorption coefficient β on the order of several hundreds of cm GW^{-1}. In combination with special device structure such as a nonlinear fiber array, it has become possible to clamp the transmitted nanosecond laser pulse energy to below the retina damage level (below 0.5 μJ).

In Table 2, we give a concise summary of these electronic optical nonlinearities of nematic liquid crystals.

Concluding Remarks

We have presented a concise review of the nonlinear optics of liquid crystals in their nematic and isotropic phases. Nematic liquid crystals in their various pure and doped forms possess many interesting nonlinear optical responses, with nonlinearities ranging over 16 orders of magnitude. Some of the extraordinarily large nonlinearities enable the performance of several all-optical switching and limiting, image modulation and sensing processes at unprecedented low threshold power. In particular, self-defocusing and optical limiting effects with nanowatt-power lasers, and incoherent–coherent image conversion at μW cm^{-2} optical intensity have been demonstrated. Liquid crystal films have been used in selective attenuation of glares and blinding laser light. Stimulated scattering that used to be the domain of high-power lasers can now be performed in thin nematic films with mW-power cw laser. Readers can consult the Further Reading section for some of these advanced photonic devices and applications based on nonlinear optics of liquid crystals. Current active studies by several research groups worldwide are likely to result in even more interesting nonlinear optical effects and useful applications in the near future.

List of Units and Nomenclature

Absorption coefficient	α [cm^{-1}]
Density	ρ_0 [kg m^{-3}]
Eleastic constant	K [dynes]
Intensity (optical)	I [W cm^{-2}]
Nonlinear susceptibility	$\chi^{(3)}$ [esu]
Second harmonic generation coefficient	β_{2HG} [mV^{-1}]
Specific heat	C_V [m^2 s^{-1}]
The nonlinear index coefficients	n_2 [cm^2 W^{-1}]
Time constants	τ [s]

Two-photon absorption constant	β [cm GW^{-1}]
Viscosity	η [kg m^{-1} s^{-1}]
Viscosity coefficient	γ [poise]

See also

Nonlinear Optics, Basics: Cascading; Kramers–Krönig Relations in Nonlinear Optics; Nomenclature and Units; Photorefraction.

Further Reading

Khoo IC (1995) *Liquid Crystals: Physical Properties and Nonlinear Optical Phenomena*. New York: Wiley Interscience.

Khoo IC (1996) Orientational photorefractive effects in nematic liquid crystal films. *IEEE Journal of Quantum Electronics* 32: 525–534.

Khoo IC and Liang Y (2000) Stimulated orientational and thermal scatterings and self-starting optical phase conjugation with nematic liquid crystals. *Physical Review* E62: 6722–6733.

Khoo IC and Wu ST (1993) *Optics and Nonlinear Optics of Liquid Crystals*. Singapore: World Scientific.

Khoo IC, Chen PH, Wood MV and Shih MY (1999) Molecular photonics of a highly nonlinear organic fiber core liquid for picosecond–nanosecond optical limiting effect. *Chemical Physics* 245: 517–531.

Khoo IC, Wood MV, Shih MY and Chen PH (1999) Extremely nonlinear photosensitive liquid crystals for image sensing and sensor protection. *Optics Express* 4(11): 431–442.

Khoo IC, Shih MY, Wood MV, *et al.* (1999) Dye-doped photorefractive liquid crystals for dynamic and storage holographic grating formation and spatial light modulation. *IEEE Proceedings* 87: 1897–1911.

Marrucci L and Papara D (1997) Photoinduced molecular reorientation of absorbing liquid crystals. *Physical Review* E56: 1765–1772.

Rau H (1990) Photoisomerization of Azobenzenes. In: Rabeck FJ (ed.) *Photochemistry and Photophysics*, vol. 2, chap. 4, pp. 119–141. Boca Raton, FL: CRC Press.

Shen YR (1990) *Principles of Nonlinear Optics*. New York: Wiley Interscience.

Organic Nonlinear Materials

F Kajzar and I Rau*, Commissariat à l'Energie Atomique, Gif sur Yvette, France

© 2005, Elsevier Ltd. All Rights Reserved.

Nonlinear Medium Polarization

Macroscopic

Photons are neutral, noninteracting bosons. Interaction is possible only through a material medium. Indeed, photons can be absorbed by electrons leading to excited states. A reverse transition may go through radiative channels, with emission of photons, usually with a different energy than the excitation energy. Interaction of photons with matter may also lead to other types of excitation, such as vibronic, rotational, libronic, etc., or just to a change in the medium polarization. This change may be expanded in the power series of the electric forcing field strength with ith component given by:

$$\begin{aligned} P_I(\omega_\sigma) = P_{0I}(\omega_\sigma) + \varepsilon_0[&K_1\chi^{(1)}_{IJ}(-\omega_\sigma;\omega_\sigma)E_J^{\omega_\sigma} \\ &+ K_2\chi^{(2)}_{IJK}(-\omega_\sigma;\omega_1,\omega_2)E_J^{\omega_1}E_K^{\omega_2} \\ &+ K_3\chi^{(3)}_{IJKL}(-\omega_\sigma;\omega_1,\omega_2,\omega_3)E_J^{\omega_1}E_K^{\omega_2}E_L^{\omega_3} + \cdots] \end{aligned} \quad [1]$$

where the expansion coefficients $\chi^{(n)}$ are $(n+1)$ rank three-dimensional tensors, describing linear $(n=1)$ and nonlinear $(n>1)$ responses of the medium, $P_{0i}(\omega_\sigma)$ is its permanent polarization (at frequency ω_σ) and the coefficients K_i $(i=1,2,3,\ldots)$ take account of the conventions for the Fourier transform of electric field and polarization vectors used, as well as the degeneracy of nonlinear optical processes. For historical reasons two conventions in quantitative description of NLO processes are used:

$$\mathrm{E}(\mathrm{r},t) = \tfrac{1}{2}[\mathrm{E}(\mathrm{r})\mathrm{e}^{\mathrm{i}\omega t} + \mathrm{c.c.}] \qquad \text{(Convention I)} \qquad [2]$$

and

$$\mathrm{E}(\mathrm{r},t) = \mathrm{E}(\mathrm{r})\mathrm{e}^{\mathrm{i}\omega t} + \mathrm{c.c.} \qquad \text{(Convention II)} \qquad [3]$$

and similarly for the polarization field. As researchers often use different conventions, it is important to bear this in mind when comparing NLO properties of materials coming from different sources. The tabulated data and the formulas used here are within convention I, with the permutation factors included explicitly in coefficients K_i (eqn [1]). In centrosymmetric media and in dipolar approximation, the static polarization P_0, as well as all even order NLO susceptibilities, are equal to zero, as it follows from time reversal symmetry.

*On leave from POLITECHNICA University of Bucharest, Romania.

Relations Between Macroscopic and Microscopic Polarizations

Microscopic Polarization

Similar expansions to eqn [1] also obey the conventions on a molecular level. Similarly, as for bulk polarization, the ith component of the molecule dipole moment can also be expanded into the series of the local field strength giving:

$$\begin{aligned}\mu_i(\omega_\sigma) &= \mu_{0i}(\omega_\sigma) + K_1\alpha_{ij}(-\omega_\sigma;\omega_\sigma)E_j^{\omega_\sigma} \\ &\quad + K_2\beta_{ijk}(-\omega_\sigma;\omega_1,\omega_2)E_j^{\omega_1}E_k^{\omega_2} \\ &\quad + K_3\gamma_{ijkl}(-\omega_\sigma;\omega_1,\omega_2,\omega_3)E_j^{\omega_1}E_k^{\omega_2}E_l^{\omega_3} + \cdots \\ &= \mu_{0i}(\omega_\sigma) + \mu_{\text{ind}}(\omega_\sigma,\omega_1,\omega_2,\ldots,\omega_n,\alpha,\beta\ldots,\zeta, \\ &\quad E^{\omega_1},E^{\omega_2},\ldots,E^{\omega_n})\end{aligned} \quad [4]$$

where α is the linear molecular polarizability, β is the first order and γ is the second molecular hyperpolarizability. μ_{in} in eqn [4] denotes the induced dipole moment. The expansions [1] and [4] may be also considered as the Tamm–Dancoff perturbation development as functions of an increasing number of interacting photons. In this case, the second term on the right-hand sides of eqns [1] and [4] describes the interaction of two photons, the third term that of three photons, etc. The electric field in eqn [4] is the effective field experienced by the molecule and is given by

$$E = fE \quad [5]$$

where f is the local field factor, depending on the form of molecule. In eqns [1] and [4] Einstein's notation is used.

For a single crystal and for noninteracting molecules there exists a straightforward relationship between the macroscopic and microscopic quantities given by:

$$\begin{aligned}\chi^{(2)}_{IJK}(-\omega_\sigma;\omega_1,\omega_2) &= \sum_n f_n^{\omega_\sigma}f_n^{\omega_1}f_n^{\omega_2}N_n \\ &\quad \times \sum_{ijk} a_{Ii}a_{Jj}a_{Kk}\beta^n_{ijk}(-\omega_\sigma;\omega_1,\omega_2)\end{aligned} \quad [6]$$

for the second-order NLO susceptibility and

$$\begin{aligned}\chi^{(3)}_{IJKL}(-\omega_\sigma;\omega_1,\omega_2,\omega_3) &= \sum_n f_n^{\omega_\sigma}f_n^{\omega_1}f_n^{\omega_2}f_n^{\omega_3}N_n \\ &\quad \times \sum_{ijkl} a_{Ii}a_{Jj}a_{Kk}a_{Ll}\gamma^n_{ijkl}(-\omega_\sigma;\omega_1,\omega_2,\omega_3)\end{aligned} \quad [7]$$

for the third-order NLO susceptibility, respectively. The coefficients a_{Ij} in eqns [6] and [7] are Wigner rotation matrix elements, N_n denoting the number density of species (n). The subscripts ($ijkl$) refer to the molecule, and ($IJKL$) to the laboratory reference frames, respectively.

For disordered (or partly ordered) materials, the macroscopic NLO susceptibility depends on the orientation average of nonzero molecular hyperpolarizability components:

$$\begin{aligned}\chi^{(2)}_{IJK}(-\omega_\sigma;\omega_1,\omega_2) &= \sum_n f_n^{\omega_\sigma}f_n^{\omega_1}f_n^{\omega_2}N_n \\ &\quad \times \sum_{ijk} < \beta^n_{ijk}(-\omega_\sigma;\omega_1,\omega_2) >_{IJK}\end{aligned} \quad [8]$$

and similarly for $\chi^{(3)}$ susceptibility. The average in eqn [8] is taken over all NLO molecule orientations and all β tensor components. Fortunately, the number of the nonzero β tensor components (21 for β tensor) is limited by the molecular symmetry and in the case of charge transfer (CT) molecules, only the component in the charge transfer direction β_{xxx} may be retained. The others are usually significantly smaller and can be neglected with respect to this one.

Second-Order Nonlinear Optical Materials

The research on nonlinear optical materials is boosted by the potentiality of their applications in photonics, and particularly in optical signal processing, image recognition, frequency conversion, and light amplification. New fields of applications open in biology, medicine (photodynamic therapy), material processing (optical engineering), etc. Depending on the targeted application, the materials showing enhanced second-, third-, and/or higher-order nonlinear optical properties are developed and optimized. The optimization is performed on both molecular and macroscopic levels. Indeed, the bulk nonlinear optical response depends not only on the individual hyperpolarizabilities but also on the way the molecules are assembled into the bulk material. As already mentioned, only the noncentrosymmetric materials exhibit the second-order nonlinear optical response.

Molecules

The most promising molecules for the application in second-order NLO devices are intramolecular CT compounds, composed from an electron donating (D), an electron accepting (A) part, linked by a π conjugated backbone (Figure 1). The CT molecules exhibit a large first-order susceptibility β tensor component β_{xxx}, enhanced in the charge transfer direction x. The value of β_{xxx} depends on the strength of electron donating and electron accepting groups (amount of the charge transfer)

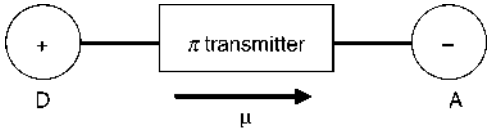

Figure 1 Schematic representation of a charge transfer molecule.

Table 1 Chemical structures of commonly used electron donating and electron accepting groups

Donors	*Acceptors*
$-N(CH_3)_2$	$-NO$
$-NH_2$	$-NO_2$
$-OCH_3$	$-CHO$
$-OH$	$-CN$
$-CH_3$	$-SO_2CH_3$
$-SCH_3$	$-SO_2CF_3$
$-N(CH_3)_2$	$=C(CN)_2$

and on the conjugation length of the transmitter. Table 1 shows the chemical structures of typical electron donating and electron accepting groups. Chemical structures of selected CT molecules are given in Table 2, together with the values of the β_{xxx} tensor component, as determined by the electric field induced second-harmonic generation technique (EFISH).

The dependence of β_{xxx} on the conjugation length L of the conjugated transmitter is given by the scaling law:

$$\beta_{xxx} \propto L^{\alpha_s} \qquad [9]$$

with $\alpha_s > 1$. Usually the conjugated backbone is composed of a number of alternate single/double (or single/triple) bonds and in this case, the scaling law is tested as a function of the number of double (or triple) bonds N:

$$\beta_{xxx} \propto N^{\alpha_\beta} \qquad [10]$$

For a number of CT oligomers with the same electron donating and electron accepting groups, differing only in the length of the conjugated transmitter, the exponent $\alpha_\beta \sim 2.5$. The β_{xxx} value saturates for a larger number of double bonds ($N > 4$, in the case of phenyl oligomers used as transmitter).

A number of studies are in progress on octupolar molecules which exhibit large β hyperpolarizabilities. However, their processing into efficient, noncentrosymmetric bulk materials is still at its initial stage and up to now no practical application has been demonstrated.

Bulk Materials

In order to obtain noncentrosymmetric materials, different approaches are used such as:

(i) single crystals growth;
(ii) epitaxy or heteroepitaxy;
(iii) Langmuir–Blodgett films;
(iv) self-assemblies;
(v) intermolecular charge transfer structures;
(vi) poled polymers; and
(vii) chiral structures.

The growth of good-quality optical single crystals is costly and difficult, especially as the charge transfer molecules usually exhibit a large ground state dipole moment (Table 2). The strong dipole–dipole interaction usually leads to their antiparallel alignment, and as a consequence, a centrosymmetric bulk structure. Only a few molecules crystallize in a noncentrosymmetric structure. One of the approaches used consists of design and synthesis of molecules with a negligible ground state dipole moment such as POM (Table 3). Therefore, a lot of effort has been put into forcing molecules to build into noncentrosymmetric bulk materials (ii)–(vi). Owing to the fact that most of applications are targeted in waveguiding configuration, a significant effort has been made to grow noncentrosymmetric thin films (ii)–(vi). While the chiral molecules form spontaneously noncentrosymmetric films, in other techniques, this is obtained more easily by application of an external electric field (poled polymers), by use of an adequate substrate (epitaxy, pseudoepitaxy, self assembly), intramolecular charge transfer in ordered layered structures (v), and interaction between molecules and molecules–substrate (Langmuir–Blodgett films). In the last case, three types of well-ordered structures, with controllable thickness are obtained: X, Y, Z (Figure 2). Only X and Z type structures give noncentrosymmetric thin films exhibiting second-order NLO properties.

Poled polymers

Up to now, the best results in getting efficient materials for second-order NLO application are obtained with poled polymers. The materials consist of a polymer matrix, exhibiting excellent optical properties, functionalized with active NLO chromophores (CT molecules). Four kinds of structures, shown schematically in Figure 3, are realized:

(i) guest–host systems;
(ii) side chain polymers;
(iii) main chain polymers; and
(iv) thermally or photo-crosslinked polymers.

Table 2 Molecular first hyperpolarizabilities and dipole moments of selected CT molecules

Molecule	*Chemical structure*	λ_{max} *[nm]*	*Solvent*	*Ground state dipole moment,* μ *[D]*	*Product* $\mu\beta$ *[10^{-77} Cm^5 V^{-1}]*	β *[10^{-40} m^4 V^{-1}]*	λ_{meas} *[nm]*	β_0 *[10^{-40} m^4 V^{-1}]*
p-NA (*para*-nitroaniline)	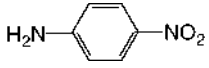	358	1,4 dioxane	5.5	1.30 0.74	71.0 40.2	1064 1907	35
MNA (2-methyl-4-nitroaniline)	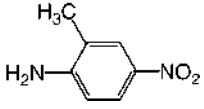	361	1,4 dioxane	7.4	1.72 0.98	69.9 39.8	1064 1907	33
NPP (*N*-(4-nitrophenyl)-(L)-prolinol)	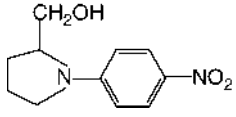	397	acetone	7.3	4.28	176	1064	67
POM (3-methyl-4-nitropyridine-1-oxide)	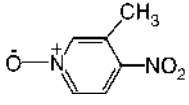	320	from crystal data	0	0	36	1064	21
PNP (1-(*N*-prolinol)-5-nitropyridine)	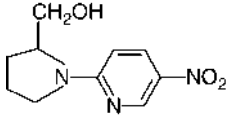	376	1,4 dioxane	7.2	2.28 1.41	95 59	1064 1907	38
MMONS (3-Methyl-4-methoxy-4′ nitrostilbene)	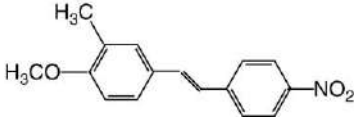	366	1,4 dioxane	5.2	2.14	123.9	1907	90
DR1 (4-[*N*-ethyl-*N*-(2-hydroxyethyl)]amino-4′-nitroazobenzene (dispersed red 1))	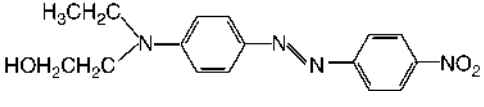	509 455	DMSO 1,4 dioxane			525 234.1	1356 1907	197 150
DADAB (4-(*N*,*N*-diethylamino)-4′-tricyanovinylazo benzene)	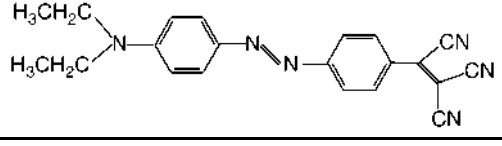	582	DMSO			1640	1580	648

$1D = (1/3) \times 10^{-39}$ Cm.

Table 3 Bulk second-order nonlinear optical properties of selected organic materials

Material	*Chemical structure*	*Form*	*Point group*	λ_{meas} *[nm]*	*Refractive index*	*SHG susceptibility* $\chi^{(2)}_{ijk}(-2\omega;\omega,\omega)$ *[pm V^{-1}]*	*Electrooptical coefficient, r, [pm V^{-1}]*
MNA (2-methyl-4-nitroaniline)	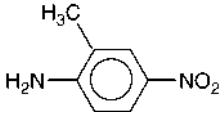	Single crystal	*m*	514.5 1064	2.0	$\chi^{(2)}_{111} = 500$ $\chi^{(2)}_{122} = 76$	70 ± 14
POM (3-methyl-4-nitropyridine-1-oxide)	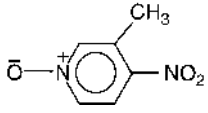	Single crystal	222	514.5 600 10 600 1064	1.712 1.919 1.638	$\chi^{(2)}_{123} = 19.2$ $\chi^{(2)}_{123} = \chi^{(2)}_{251} = \chi^{(2)}_{312}$	4.0 5.1 3.7
COANP (2 cyclooctylamino-5 nitropyridine)	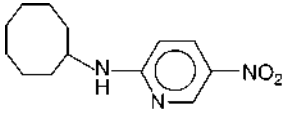	Single crystal	*mm*2	514.5 600 10 600 1064	1.672 1.781 1.647	$\chi^{(2)}_{322} = 108$ $\chi^{(2)}_{333} = 48$	9.3 ± 1.5 15.4 ± 3.2
PNP (2-(N-prolinol)-5-nitropyridine)	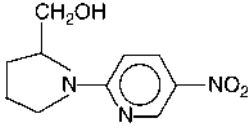	Single crystal	2	514.5 600 10 600 1064	1.990 1.788 1.467	$\chi^{(2)}_{211} = 170$ $\chi^{(2)}_{223} = 55$	8.0 ± 1.4 13.0 ± 2.2
MNBA (4′-nitrobenzylidene-3-acetamino-4-methoxyaniline)	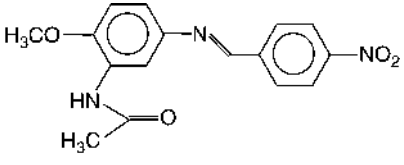	Single crystal	*m*	514.5 600 10 600 1064	2.024 1.585 1.652	$\chi^{(2)}_{111} = 437$	–
MMONS (3-Methyl-4-methoxy-4′-nitrostilbene)	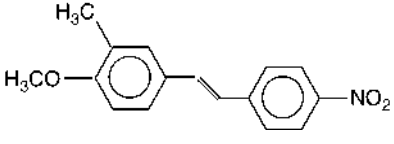	Single crystal	*mm*2	514.5 600 10 600 1064	1.569 1.693 2.129	$\chi^{(2)}_{333} = 368^a$ $\chi^{(2)}_{223} = 142^a$	19.3 ± 4 39.9 ± 7

PMMA + DR 1 (Poly(methyl methacrilate) + 4-[N-ethyl-N-(2-hydroxyethyl)] amino-4′-nitroazobenzene (Dispersed red 1))	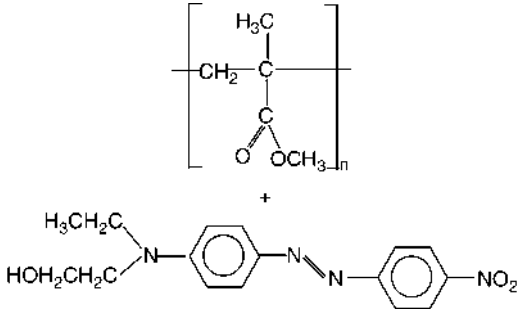	PP/GH	1580	$\chi^{(2)}_{333} = 5$	
			633		$r_{33} = 2.5$
PMMA + DCV (Poly(methyl metacrilate) + 4-dicyanovinyl-4′-[N-ethyl-(2 hydroxyethyl)] azobenzene)	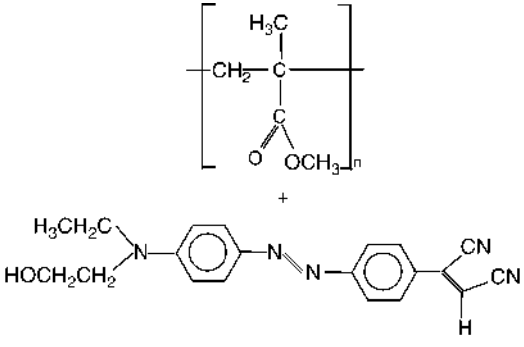	PP/SC	1580	$\chi^{(2)}_{333} = 42$	
			799		$r_{33} = 18$
PMMA + 3RDCVXY (Poly(methyl metacrilate) + 4-dicyanovinyl-4′ (diethylamino) diazobenzene)	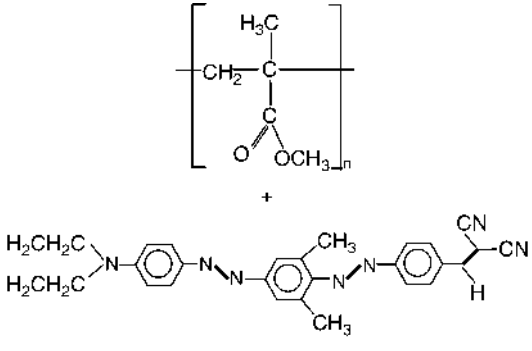	PP/SC	1064	$\chi^{(2)}_{333} = 840$	
			633		$r_{33} = 40$

Table 3 (*Continued*)

Material	*Chemical structure*	*Form*	*Point group*	λ_{meas} *[nm]*	*Refractive index*	*SHG susceptibility* $\chi^{(2)}_{ijk}(-2\omega;\omega,\omega)$ *[pm V^{-1}]*	*Electrooptical coefficient, r, [pm V^{-1}]*
NNDN + NAN (N,N-diglycidyl-4-nitroaniline + N-(2-aminophenyl)-4-nitroaniline)	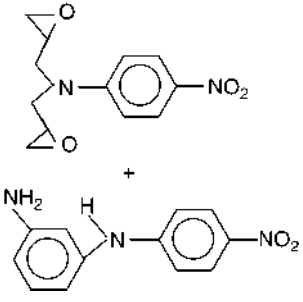	PP/CL		1064		$\chi^{(2)}_{311} = 28$	
				530.9			$r_{13} = 6.5$
DCANP	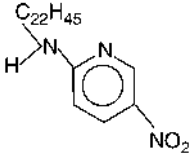	LB film		632.8	1.598 (TE_0-mode)		
				1064		$\chi^{(2)}_{333} = 19.5 \pm 2.5$	
FA06	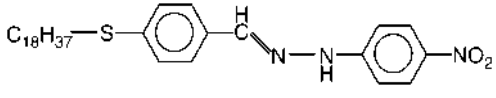	LB film		1064		$\chi^{(2)}_{322} = 20$	

PP, poled polymer; GH, guest host system; SC, side chain polymer; CL, cross linked polymer; LB, Langmuir–Blodgett film.
[a] Calibrated with KTP single crystal ($\chi^{(2)}_{333} = 27.4$ pm/V).

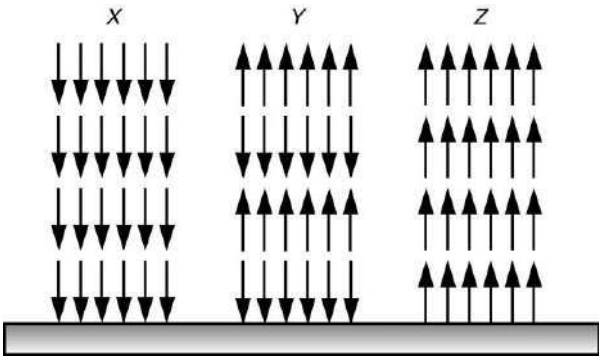

Figure 2 Schematic representation of *X, Y, Z* type Langmuir–Blodgett films. The arrows symbolize the direction of dipole moments, which usually are tilted with respect to the normal to film surface.

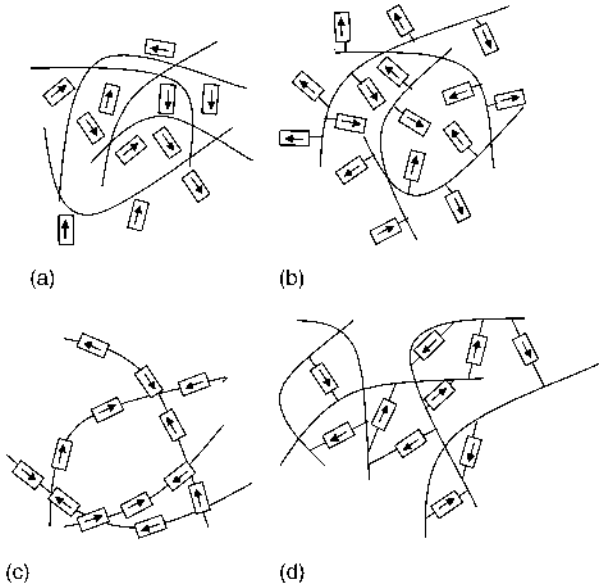

Figure 3 Schematic representation of functionalized polymers for second-order NLO effects: guest–host systems (a), side chain polymers (b), main chain polymers (c), photo and thermally cross-linking polymers (d). Arrows represent the NLO chromophores.

The guest–host systems are just solid solutions of NLO chromophores in an inert matrix. The concentration of NLO chromophores is limited because of already-mentioned strong dipole–dipole interaction leading to antiparallel alignment, resulting in an increase in propagation losses and decrease of $\chi^{(2)}$ susceptibility. The polar order is obtained by applying an external field. In order to increase the rotational mobility of chromophores, the material is heated to the glass transition temperature of polymer. The interaction of dipole moments with the external field forces them to orient in its direction. The orientation is frozen by cooling the material (usually a thin film) to room temperature. However, the stability induced in this way results in a polar order that is poor. Better stability and higher concentration of chromophores is obtained by attaching chemically the NLO active molecules to the polymer backbone (side-chain polymers). However, it requires a significant effort in chemical synthesis. The stability of induced order depends strongly on the difference between the glass transition and the operation temperatures. The larger the difference, the less relaxation is observed. Therefore, a lot of effort is going into developing functionalized high glass transition temperature polymers such as polyimides, polyurethanes, polyvinyl carbazoles, etc.

The main-chain polymers (iii) attract a relatively weak interest. This is due, not only to the difficulties encountered with chemical synthesis, but also with the efficient chromophore orientation. More promising is the last class of functionalized polymers (iv). Here the active chromophores are used to link covalently the polymer chains through a thermal or photo-crosslinking process. However, the poling process is very delicate as the glass transition temperature is increasing with the poling time, so the time process has to be carefully monitored. Also the cross-linking may lead to shrinkage and crackling of thin films and as a consequence, to the increase of propagation losses.

A special class of functionalized materials, similar to poled polymers, represent sol–gels. Here the active chromophores either fill the pores of sol–gel or are chemically bonded to the matrix. The poling process is very similar to that applied to the thermally and photo cross-linking polymers and also has to be carefully monitored. The main advantage of these materials consists of excellent optical properties, and the disadvantage on a relatively low chromophore content, limiting the bulk $\chi^{(2)}$ susceptibility value. The representative values of second-order nonlinear optical susceptibility $\chi^{(2)}_{ijk}$ for different materials for second-order NLO optics, are listed in Table 3.

Poling efficiency

The DC poling of functionalized polymers leads to birefringent noncentrosymmetric materials with two nonzero $\chi^{(2)}$ tensor components: the diagonal:

$$\chi^{(2)}_{ZZZ}(-\omega_0;\omega_1,\omega_2) = NF\beta_{xxx}(-\omega_\sigma;\omega_1,\omega_2) \times \langle\cos^3\Theta\rangle \qquad [11]$$

and the off-diagonal:

$$\chi^{(2)}_{XXZ}(-\omega_\sigma;\omega_1,\omega_2) = \frac{1}{2}NF\beta_{xxx}(-\omega_\sigma;\omega_1,\omega_2) \times \langle\sin^2\Theta\cos\Theta\rangle \qquad [12]$$

where we assumed for simplicity, that only one nonlinear species, as is usually the case. Z, in eqns [11] and [12], is the direction of the applied external field and Θ is the angle between the electric field and the molecule dipole moment. The efficiency of poling is measured by the ratio:

$$a = \frac{\chi^{(2)}_{ZZZ}(-\omega_\sigma; \omega_1, \omega_2)}{\chi^{(2)}_{XXZ}(-\omega_\sigma; \omega_1, \omega_2)} \qquad [13]$$

It varies between 1 and ∞ (perfect alignment). For isotropic poled polymers $a = 3 \div 6$ and for nematic liquid crystal side-chain polymers, values of $a \propto 18$ were obtained.

Usually, the second-order NLO susceptibility is measured by the second-harmonic generation technique ($\omega_\sigma = 2\omega$; $\omega_1 = \omega_2 = \omega$), giving the electronic part of $\chi^{(2)}$ susceptibility, with ultrafast response time as the generated harmonic wave has to follow the fundamental one. This susceptibility is important, not only for frequency conversion through the second-harmonic generation, but also for sum and difference frequency generation as well as for optical parametric oscillation (OPO) and amplification (OPA). Another important effect described by $\chi^{(2)}(-\omega; \omega, 0)$ susceptibility is the linear electro-optic effect (Pockels effect). It finds application in electro-optic modulation for signal transmission. Traditionally, this effect is described by the electro-optic coefficient r_{ijk}, which is a third-rank three-dimensional tensor. The variation of the medium refractive index under the applied electric field is given by the following formula:

$$\Delta n_i = -\frac{2}{n_i^3} r_{ij} E_j \qquad [14]$$

The electro-optic coefficient is directly linked to the $\chi^{(2)}(-\omega; \omega, 0)$ susceptibility:

$$r_{ijk} = -\frac{2\chi^{(2)}_{ijk}(-\omega; \omega, 0)}{(n_i n_j)^2} \qquad [15]$$

The measured values of electro-optic coefficient for some selected organic single crystals and thin films are also given in Table 3.

Third-Order Nonlinear Optical Materials

There is no symmetry restriction on molecules and materials for third- (or higher-odd-) order nonlinear optical effects. These materials are used in significantly larger classes of applications than the materials for second-order NLO effects, particularly in logic systems, all optical switching, frequency conversion (e.g., frequency tripling in third-harmonic generation), light amplification, optical bistability, all optical switching, etc. In the case of organic materials, it was shown that the conjugated quasi 1D π electron systems, such as semiconducting polymers (Figure 4) exhibit an enhanced third-order NLO response.

Molecules

Similarly, as in the case of second-order NLO materials, the NLO response depends strongly on the conjugation length. In the case of conjugated quasi 1D molecules, such as polymers with a chemical structure as shown in Figure 4, the γ tensor component in the conjugation direction (largest hyperpolarizability) is strongly enhanced and described by a scaling law:

$$\gamma_{xxxx} \propto L^{\alpha_t} \qquad [16]$$

where L is the conjugation length and the exponent $\alpha_t > 1$. The scaling law, eqn [16], can also be presented as dependent on γ on a number of π conjugated double or triple bonds N:

$$\gamma_{xxxx} \propto N^{\alpha_\gamma} \qquad [17]$$

It was extensively studied for a series of conjugated oligomers polyenes, oligothiophenes, etc., and the exponential coefficients are listed in Table 4, together with the calculated values. For a larger value of double (or triple) bonds, a saturation in $\chi^{(3)}$ susceptibility is observed.

The molecular second hyperpolarizability measurements are derived from solution measurements. Some typical values, obtained by the third-harmonic generation technique (THG) are listed in Table 5. Although other NLO property determination techniques were used, the THG method gives the fast, electronic part of molecular hyperpolarizability.

Bulk Materials

Importance of Order

Similarly, as in the case of materials for second-order NLO effects for practical applications the value of $\chi^{(3)}$ susceptibility is important, which on the other hand depends on the molecular second hyperpolarizability γ. Although the noncentrosymmetry is not required in the case of third-order NLO materials, the values of the bulk $\chi^{(3)}$ susceptibility depend on the order, as it follows from eqn [7]. This is particularly clear in the case of conjugated molecules, with enhanced γ_{xxxx} component. In the case of parallel polarizations

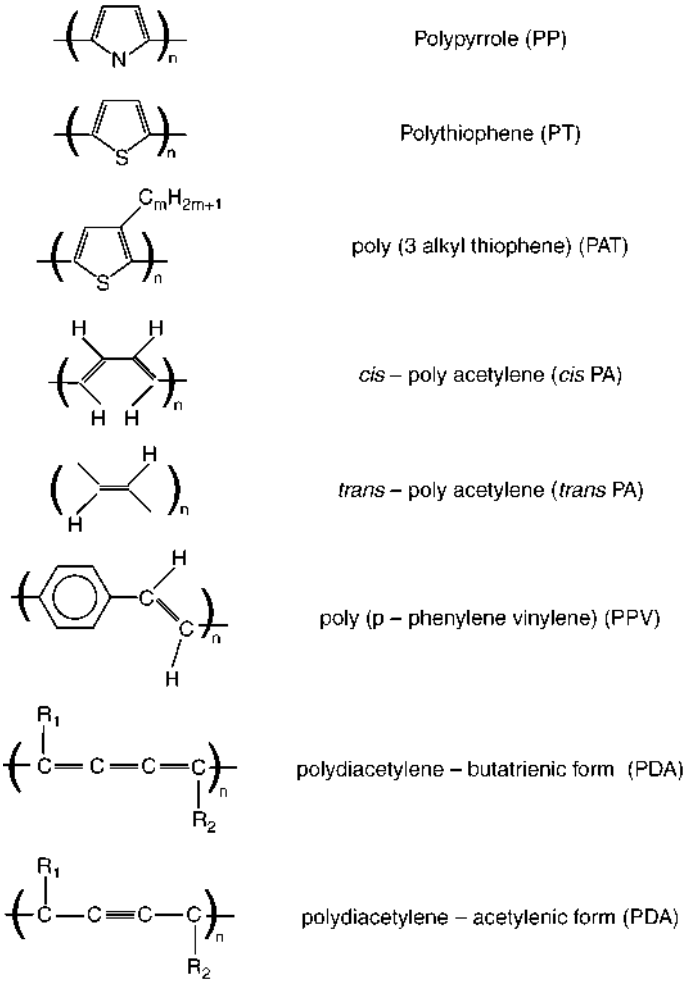

Figure 4 Chemical structures of selected conjugated quasi 1D organic polymers.

Table 4 Calculated (calc) and measured (meas) values of the scaling low exponent $\gamma \propto N^{\alpha_\gamma}$ for different conjugated oligomers

Oligomers	$\alpha_{\gamma\,calc}$	*Remark*	$\alpha_{\gamma\,meas}$	*Remark*
Polyene oligomers	5	free electron model	2.82	CT polyenes ($\omega = 0$)
		INDO ($\omega = 0$)	3.5	$\lambda_\omega = 1.907$ μm
	3	SCF/STO-3G	2.3	Symmetric A-A
	3.27	SCF/3-21G		THG 1.907 μm
	3.14	$N = 2$	3.3	Asymmetric D-A
	3.67	$N = 6$, decrease		THG 1.907 μm
	3.98	For $N > 6$ DVB-PPP	4.4	Symmetric D-D
				EFISH 1.34 μm
				no saturation up to $N = 14$
	3.8	neutral solitons		
	3.85	charged polarons		
	6.32	*trans* form		
	5.4	*cis* form		
	4.7			
Thiophene oligomers	4.5	CNDO $N < 6$	4.05	DFWM in solutions
		Saturation $N > 6$	4.54	EFISH in solid solutions $N = 3, 5, 6$
			3.4	THG 1.907 μm

of $\omega_\sigma, \omega_1, \omega_2, \omega_3$ fields, the bulk susceptibility is given by:

$$\chi^{(3)}(-\omega_\sigma; \omega_1, \omega_2, \omega_3) = NF\gamma_{xxxx}(-\omega_\sigma; \omega_1, \omega_2, \omega_3) \times \langle \cos^4 \Phi \rangle \quad [18]$$

where F is the total field factor and Φ is the angle between the polarization direction of $\omega_\sigma, \omega_1, \omega_2, \omega_3$ fields and a chosen direction in the laboratory reference frame. The value of $\langle \cos^4 \Phi \rangle$ varies between 0.2 for an isotropic distribution of polymer chains and 1 for perfectly aligned, respectively. It gives a factor of 5 difference in the macroscopic $\chi^{(3)}$ susceptibility. Some examples of third-order NLO susceptibilities for representative conjugated polymers are given in Table 6. These were obtained by third-harmonic generation technique which measures the fast, electronic contribution to $\chi^{(3)}$.

Table 5 Molecular off resonance second hyperpolarizability γ for selected organic molecules

Molecule	$\gamma(-3\omega; \omega, \omega, \omega)$ 10^{-50} m^5/V^2	λ_{meas} *nm*
Carbon tetrachloride	5.1	1064
	4.7	1907
Chloroform	3.8	1064
	3.2	1907
Acetone	2.5	1064
	2.6	1907
Methanol	1.16	1064
	1.21	1907
Ethanol	1.8	1064
	2.1	1907
Propanol	2.4	1064
	2.5	1907
Water	0.46	1064
Nitrobenzene	9.2	1907
Chlorobenzene	5.8	1907

Calibrated with silica value: $\chi^{(3)}(-3\omega; \omega, \omega, \omega) = 3.9 \times 10^{-22}$ m^2/V^2 at 1907 nm.

Two-Photon Absorbing Materials

Two-photon absorbing (TPA) materials are a special class of third-order NLO materials exhibiting a complex Kerr susceptibility $\chi^{(3)}(-\omega; \omega, -\omega, \omega)$. In this case, the nonlinear refractive index, derived from eqn [1] is complex too.

$$n = n_0 + n_2 I = n_0 + (n_2^r + \kappa_2) I \quad [19]$$

where I is the light intensity, n_2^r is the real part and κ_2 is the imaginary part (nonlinear extinction coefficient) of the nonlinear index of refraction, respectively. The nonlinear index of refraction is directly related to the Kerr susceptibility $\chi^{(3)}(-\omega; \omega, -\omega, \omega)$:

$$n_2 = \frac{3\chi^{(3)}(-\omega; \omega, -\omega, \omega)}{4\varepsilon_0 c n_0^2} \quad [20]$$

where c is the light velocity. If $\chi^{(3)}(-\omega; \omega, -\omega, \omega)$ is complex (close to a resonance) then n_2 is complex too and the light beam propagating in such a medium in direction z is attenuated through the two-photon absorption process:

$$I(z) = I_0 \, e^{-(\alpha_L + \alpha_{TPA} I) z} \quad [21]$$

Table 6 Resonant and nonresonant values of cubic susceptibility $\chi^{(3)}$ for selected organic materials

Material	*Form*	λ_{meas} *nm*	$\chi^{(3)}(-3\omega; \omega, \omega, \omega)$ 10^{-22} m^2/V^2	*Remark*
PDA-PTS	Single crystal	1890	$119\ (70) \times 10^3$	∥ chain, 3ph
			<1400	⊥ chain
		2620	$22.4\ (14) \times 10^3$	∥ chain
			<280	⊥ chain
PDA-TCDU	Single crystal	1890	$9.8\ (7) \times 10^3$	∥ chain
			<70	⊥ chain
		2620	$51.8\ (20) \times 10^2$	∥ chain
			<40	⊥ chain
DEANS	Single crystal	1900	1.4×10^4	∥ (mol. axis)
			127	⊥ (mol. axis)
PDA-AFA	LB film (blue)	1350	$7.8\ (1.4) \times 10^3$	2ph
PDA-AFA	LB film (blue)	1907	$11.6\ (1.2) \times 10^3$	3ph
PA	Isotropic oriented	1907	$78.4\ (22) \times 10^4$	3ph
			2.4×10^6	3ph
PDA-C_4UC_4	Oriented film	1907	2.9×10^4	∥ chain, 3ph
			3×10^2	⊥ chain
C_{60}	Isotropic film	1064	$11.5\ (1.1) \times 10^3$	3ph
C_{70}	Isotropic film	1420	$13\ (1.4) \times 10^3$	3ph
PPV	Isotropic film	1450	1.96×10^4	3ph

Numbers in brackets are standard deviations. Data calibrated with silica: $\chi^{(3)}(-3\omega;\omega,\omega,\omega) = 3.9 \times 10^{-22}$ m^2/V^2; 2ph, two photon resonance enhancement; 3ph, three photon resonance enhancement; for chemical structures see Figure 4.

Table 7 Two photon absorption cross-sections σ_{TPA}, maximum TPA wavelength (λ_{max}), measurement wavelength (λ_{meas}), pulse duration and measurements technique for a series of organic molecules

Molecule/material	*Chemical structure*	*Solvent*	λ_{max}^{TPA} *[nm]*	σ_{TPA} *[GM]*	*Pulse duration*	λ_{meas} *[nm]*	*Measurement technique*
BDBAS (Bis(di-*n*-butylamino)stilbene)	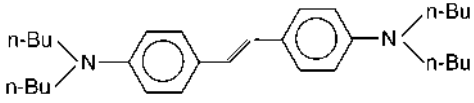	Toluene	600	310	4 ps	600	NLT
				9300	5 ns	600	
				1000	5 ns	700	
			600	210	5 ns	600	TPF
				6	5 ns	698	
				110	100 fs	620	
		Acetone		17 700	5 ns	600	NLT
	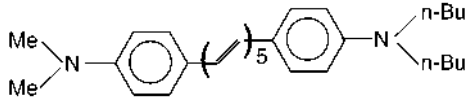	Toluene	755	10 000	5 ns	700	NLT
				17 000	5 ns	755	
				290	4 ps	700	
			730	300	5 ns	698	TPF
				790	5 ns	730	
BDPAS (Bis(diphenylamino)stilbene)	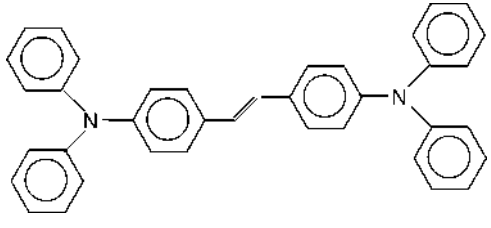	Toluene		13 000	5 ns	700	NLT
				110	4 ps	700	
Trans-stilbene	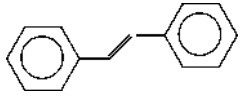	Toluene	514	12	5 ns	514	TPF
	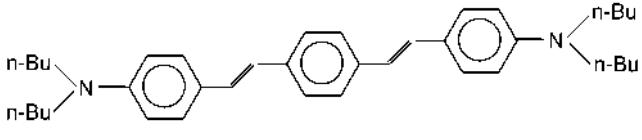	Toluene	730	995	5 ns	730	TPF
			~725	635	100 fs	725	
	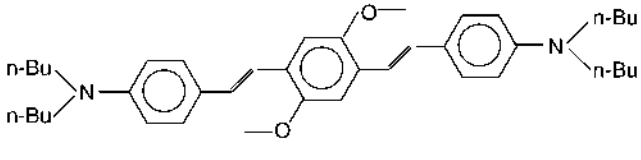	Toluene	730	900	5 ns	730	TPF
			~725	680	100 fs	725	

Table 7 (*Continued*)

Molecule/material	*Chemical structure*	*Solvent*	λ_{max}^{TPA} *[nm]*	σ_{TPA} *[GM]*	*Pulse duration*	λ_{meas} *[nm]*	*Measurement technique*
	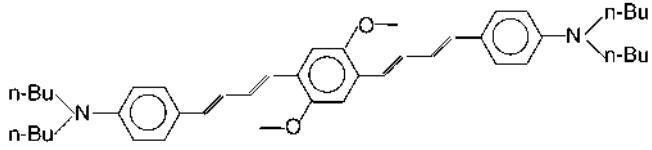	Toluene	775 ~750	1250 1270	5 ns 100 fs	775 750	TPF
	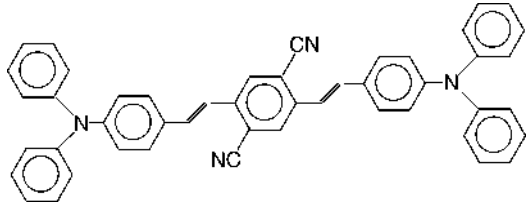	Toluene	835 810	1940 3670	5 ns 100 fs	835 810	TPF
	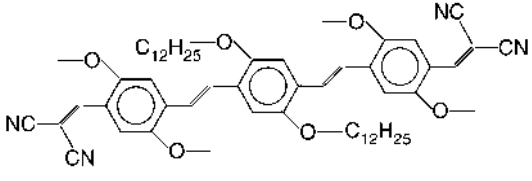	Toluene	825 940 815 910	480 620 650 470	5 ns 100 fs	825 940 815 910	TPF
	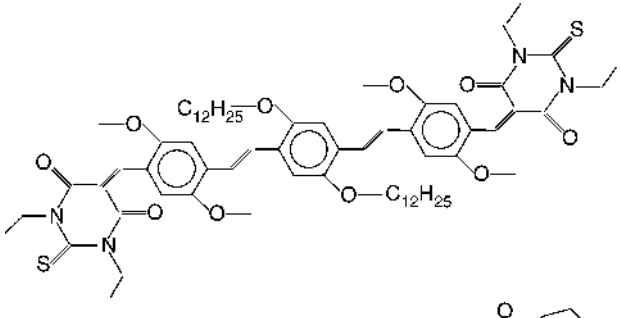	Toluene	970	1750	5 ns	970	TPF
	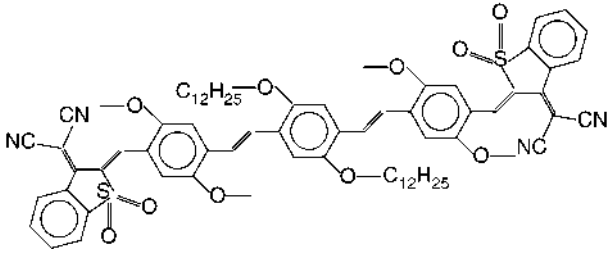	Toluene	975 945	4400 3700	5 ns 100 fs	975 945	TPF

	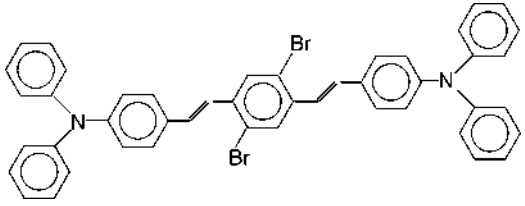	Toluene	~800	450	5 ns	800	TPF
CZ-DTT-CZ (2,6 -Bis [2-(9-ethylcarbazol-3-yl) vinyl] thiopheno [3′,2′-2,3] thiopheno [4,5-b] thiophene)	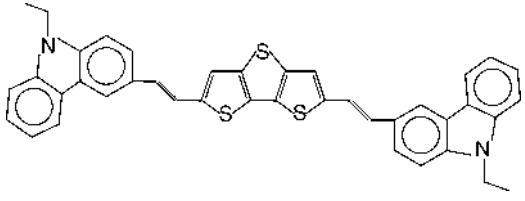	1,1,2,2-Tetra chloro-ethane		105 000	8 ns	810	NLT
CZ-DTT-OX (2-[2-(4-(5-(4-*tert*-Buthyl phenyl-1,3,4-oxadiazol-2-yl) phenyl) vinyl)-6-[2-(9-ethyl carbazol-3-yl) vinyl] thiopheno [3′,2′-2,3] thiopheno [4,5-b] thiophene)	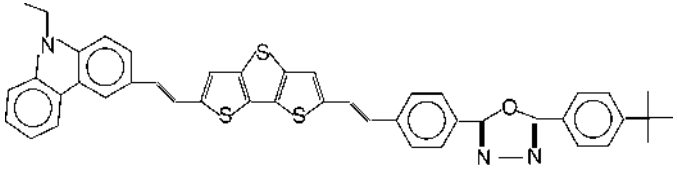	1,1,2,2-Tetra chloro-ethane		33 500	8 ns	810	NLT
TP-DTT-TP ([4-(2-(6-(2-(4-Diphenylamino) phenyl)vinyl) thiopheno [2′,3′-4,5] thiopheno [3,2-b] thiophen 2-yl-vinyl) phenyl] diphenylamine)	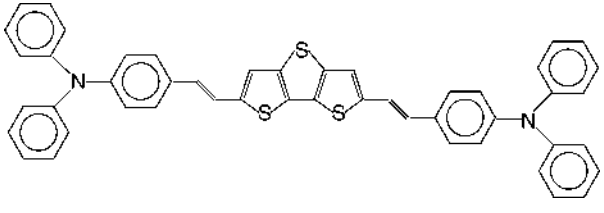	1,1,2,2-Tetra chloro-ethane		199 000	8 ns	810	NLT
TP-DTT-OX ([4-(2-(6-(2-(4-(5-(4-*tert*-Buthylphenyl-1,3,4-oxadiazol-2-yl)) phenyl) vinyl) thiopheno [2′,3′-4,5] thiopheno [3,2-b] thiophen 2-yl)-vinyl) phenyl] diphenylamine)	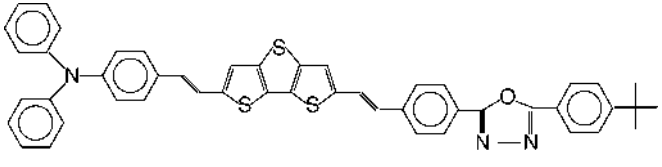	1,1,2,2-Tetra chloro-ethane		119 000	8 ns	810	NLT
Biphenyl	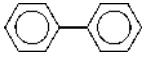			80			Theoretical calculus
Terphenyl	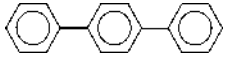			330			Theoretical calculus

Table 7 (*Continued*)

Molecule/material	*Chemical structure*	*Solvent*	λ_{max}^{TPA} *[nm]*	σ_{TPA} *[GM]*	*Pulse duration*	λ_{meas} *[nm]*	*Measurement technique*
Sexiphenyl				2090			Theoretical calculus
AF 50	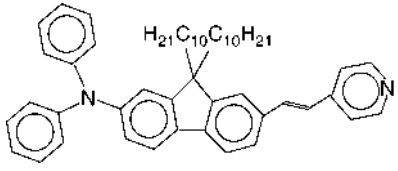	Tetrahydro-furan Benzene		11 560 19 300	8 ns	800	NLT
AF 80	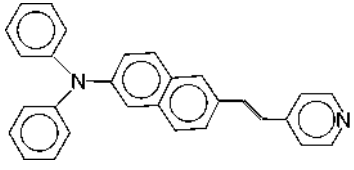	Tetrahydro-furan		6840	8 ns	800	NLT
ASPT	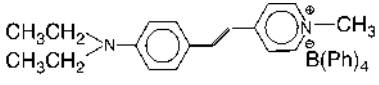	Poly hydroxyethyl metacrylate		12 000	10 ns	1064	NLT
DEANST	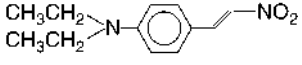	Tetrahydro-furan		200	46 ps	1064	KE
Rhodamine B	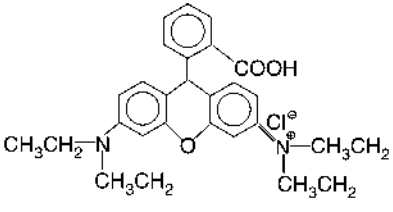	Methanol		50	100 fs	730	TPES
DR1 (4-[*N*-ethyl-*N*-(2-hydroxyethyl)] amino-4′-nitroazobenzene (dispersed red 1))	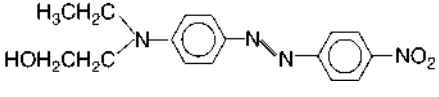	Tetrahydro-furan		102	42 ps	1064	KE

KE, optical Kerr effect; NLT, nonlinear transmission; TPF, two photon fluorescence; TPS, two photon excitation spectroscopy.
1 GM = 10^{-50} cm^4 s/photon-molecule.

Table 8 Two photon absorption coefficients for different molecules and extrapolated bulk values

Molecule	α_{TPA} *cm/GW*	λ_{max} *nm*	λ_{TPA} *nm*	*Concentration g/l*	*Solvent*	*Bulk value cm/GW*
DR 1	0.8	490	835	70	THF	11.4
DBANS	1.1	450	808	70	THF	15.7
DCNP	0.06	460	920	30	THF	2.3
DEANST	0.64	440	920	105	THF	6.1
DCM	0.1	480	950	11	DMSO	9.1
PDA-PTS	80	620	1064 not max of λ_{TPA}	Bulk	Single crystal	80
PDA-PTS	800	620	925	Bulk	Single crystal	800
Poly(3-octyl thiophene)	0.27	435 (142)	720 (94)	25	THF	9.6
Poly(3-octyl thiophene) regioregular	0.85	550 (170)	730 (70)	11.2	THF	76

λ_{max} corresponds to the maximum linear and λ_{TPA} to the maximum nonlinear absorption wavelengths, respectively, DCNP: 3-(1,1-dicyanoethenyl)-1-phenyl-4,5-dihydro-1H-pyrazole, for other molecules see **Figures 4** and **5** and preceding tables.

where α_L is the linear and α_{TPA} is the two-photon absorption coefficient, respectively. The last coefficient can be explicitly expressed by the imaginary part of the nonlinear index of refraction:

$$\alpha_{TPA} = \frac{2\pi\kappa_2}{\lambda_\omega} \qquad [22]$$

where λ_ω is the wavelength of the propagating beam. On the molecular level one also defines the two-photon absorption cross-section given by

$$\sigma_{TPA} = \frac{\hbar\omega}{N}\alpha_{TPA} \qquad [23]$$

where N is the number density of molecules. The values of TPA absorption cross-sections and two-photon absorption coefficients for selected organic molecules and solids are listed in Tables 7 and 8, respectively.

Photorefractive Materials

The photorefractive effect consists of the generation of a refractive index grating in an NLO material by its inhomogeneous illumination. This is a necessary condition but not a sufficient one. The sufficient condition consists of the possibility of creating the space charge field, which combined with the applied external field, leads to the refractive index modulation through the Pockels effect. In that case, the refractive index grating is phase mismatch with the intensity grating, leading to the transfer of photons from one beam to another, as occurs in two-beam coupling experiments, and thus applies to light amplification. Although the photorefractive effect was first observed in ferroelectric single crystals, the researchers at IBM demonstrated, at the beginning of 1990s, the photorefractive effect in organic composite materials. However, the main contribution in these materials comes from the rotational mobility of nonlinear optical (NLO) chromophores under the applied external electric field. Thus, to exhibit a photorefractive effect, the organic material has to possess the following properties:

(i) charge generation and separation capacity;
(ii) charge mobility;
(iii) linear electro-optic (Pockels) effect; and
(iv) rotational mobility.

Composite Materials

The different properties satisfying the above requirement to exhibit a photorefractive effect, are obtained by mixing together four types of molecules:

(i) photoconducting polymer;
(ii) highly second order nonlinear CT molecules;
(iii) photoensitizer for charge generation under illumination; and
(iv) plasticizer, lowering the glass transition temperature, thus increasing the rotational mobility.

Some examples of different types of molecules used are shown in Figure 5. Although these structures exhibit large exponential gain coefficients Γ and high diffraction efficiencies η, their main disadvantage consists of slow, seconds-duration response times

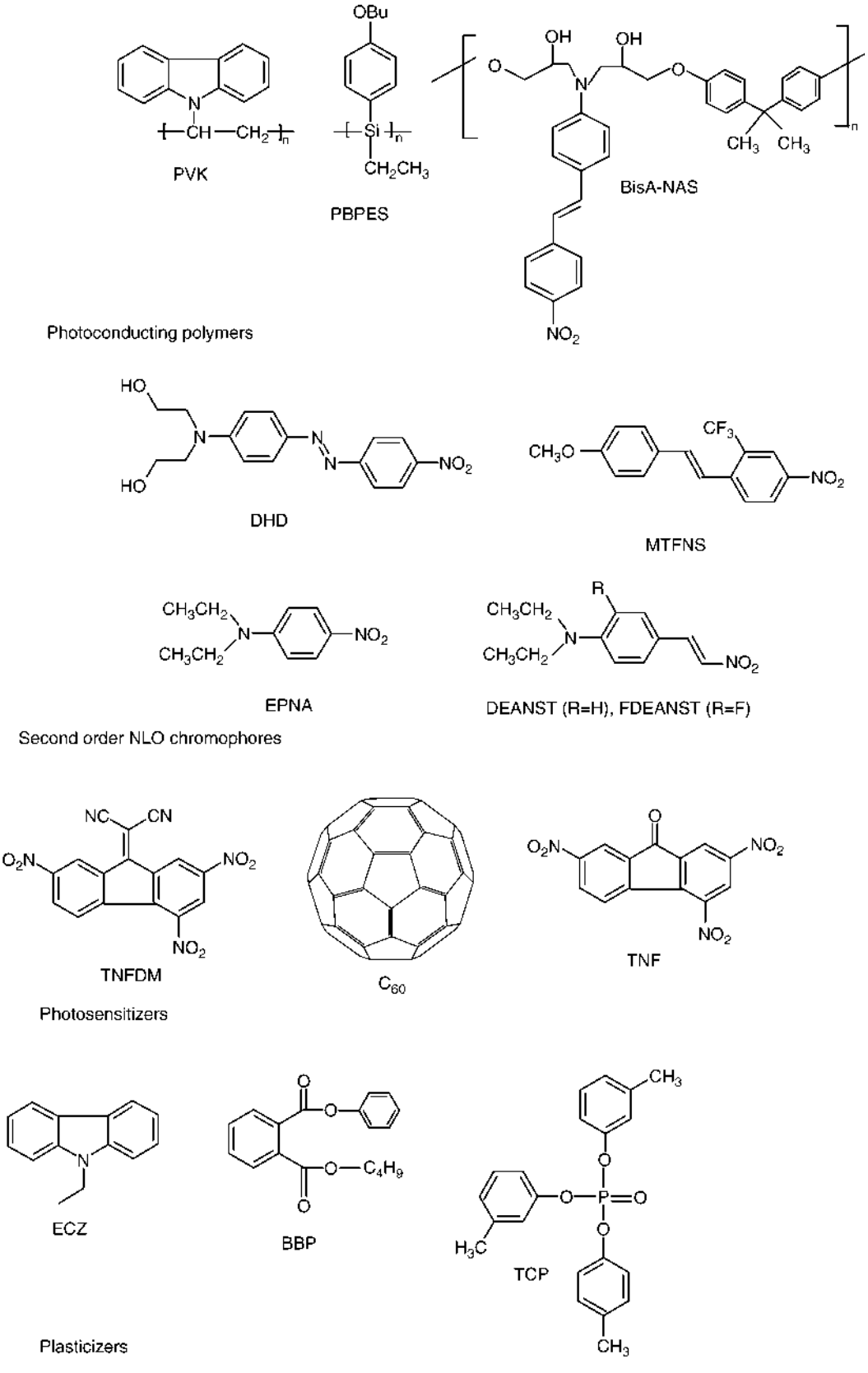

Figure 5 Chemical structures of photoconducting polymers, CT chromophores, photosensitizers and plasticizers commonly used for the fabrication of photorefractive polymers.

Table 9 Photorefractive properties of selected single crystals at zero external field, composite photorefractive polymers and liquid crystals

Structure	*Form*	*Thickness* μm	*Composition w%*	*Laser intensity* W/cm^2	*Exponential gain* $\Gamma\ cm^{-1}$	*Electric field* $V/\mu m$	*Response time s*	*Diffraction efficiency %*	λ_{meas} *nm*
COANP: TCNQ	Single crystal			0.08			100–1000 depending on grating period Λ		514.5
MNBA	Single crystal			0.1	2		10–100		514.5
DMNPAA/PVK/TNF/ECZ	Composite material	105	50: 33: 16: 1		220 (p-polarized) −40 (s-polarized)	90	1	86	514.5
Doped: 5 CB/C_{60}	Composite material	25	95: 5		2890	0.07	0.3		514.5
Doped: 8 OCB/5 CB /PER/NI	Composite material	37			640	0.04	0.04		514.5
SCLCP/E7/C_{60}	Composite material	50	49.975: 49.975: 0.05		600	0.7	3.5		632.8
PVA/E44/C_{60}	PDLC	25				40	On: 8 s Off: $>$3 days	52	
PMMA/E49/ETZ/TNF	PDLC	53	45: 33: 21: 1		41	22		8	675
POT/LC	HPCPLC	10			965	0.6	0.011	26	514.5
POT/DR1/LC	HPCPLC	10			2600	0.6	0.010	31	514.5
PVK/TNF/LC	HPCPLC	10			3700	0.9	0.010	45	514.5

LC, liquid crystals; PDLC, polymer dispersed liquid crystals; HPCPLC, hybrid photoconducting polymer–liquid crystal; COANP-2-cyclooctalamino-5 nitropyridine, TCNQ, 7,7,8,8-tetracyanoquinodimethane; MNBA4′, nitrobenzylidene-3 acetamino-4-methoxyaniline; 5-CB-4′, (*n*-octyloxy)-4-cyanobiphenyl; 8OCB-4′, (*n*-octhyloxy)-4-cyanobiphenyl; PER, perylene; NI, *N,N*′-di(*n*-octyl)1,4,5,8-naphtalenediimide; SCLCP, side chain liquid crystalline polymer (PMMA-4-cyanophenyl benzoate); E7, E44, E49, commercial LC mixtures (Merck, Germany); other molecules in **Figure 5** and **Tables 2** and **3**.

and require large external operation electric fields (~100 V/μm).

Photorefractive Liquid Crystals

In order to improve the performances of organic photorefractive materials the use of nematic liquid crystals was proposed as active media. These materials exhibit a large rotational mobility, requiring a relatively small electric field with response time of milliseconds. The reorientation of domains requires $2 \div 3$ orders of magnitude electric fields as compared with the composite materials (reorientation of individual molecules). Several types of structures of liquid crystal based photorefractive structures are realized, such as:

- composite materials in which the liquid crystals mesogens are NLO chromophores;
- polymer dispersed liquid crystals (PDLCs);
- hybrid photoconducting polymers – liquid crystals (HPCPLCs), in which the rotational mobility is separated from the charge mobility. Moreover the use of photoconducting polymer allows the modulation of the electric field on the liquid crystal layer, enhancing the effect.

The performances of different, representative, organic photorefractive structures are listed in Table 9.

See also

Nonlinear Optics, Basics: Photorefraction; $\chi^{(2)}$–Harmonic Generation. **Polarization:** Introduction.

Further Reading

Agranovich VM and La Rocca GC (eds) (2002) Organic Nanostructures: Science and Applications. *Proceed. International School of Physics "Enrico Fermi", Course CXLIX*. Amsterdam: IOS Press.

Bosshard Ch, Sutter K, Pretre Ph, Hulliger J, *et al.* (1995) In: Kajzar F and Garito AF (eds) *Organic Nonlinear Optical Materials*, *Advances in Nonlinear Optics*, vol. 1. Amsterdam: Gordon and Breach Publishers.

Bredas JL and Silbey R (eds) (1991) *Conjugated Polymers*. Dordrecht: Kluwer Academic Publishers.

Charra F, Agranovich V and Kajzar F (2003) *Organic Nanophotonics*. NATO Science Series 2. Dordrecht: Kluwer Academic Publishers.

Chemla DS and Zyss J (eds) (1987) *Nonlinear Optical Properties of Organic Molecules and Polymers*. Orlando, FL: Academic Press.

Graja A, Bulka B, Kajzar F (eds) (2002) *Molecular Low Dimensional-and Nano-Structured Materials for Advanced Applications*. NATO Science Series 2. Dordrecht: Kluwer Academic Publishers.

Hornac LA (ed.) (1992) *Polymers for Lightwave and Integrated Optics. Technology and Applications*. New York: Marcel Dekker, Publishers.

Kajzar F and Agranovich VM (eds) (2000) *Multiphoton and Light Driven Multielectron Processes in Organics: New Phenomena, Materials and Applications*. Dordrecht: Kluwer Academic Publishers.

Kajzar F and Reinisch R (eds) (1998) *Beam Shaping and Control with Nonlinear Optics*. NATO ASI Series, Series B: Physics, vol. 369. New York: Plenum Press.

Kajzar F and Swalen J (eds) (1996) *Thin Films for Waveguiding Nonlinear Optics*. Amsterdam: Gordon and Breach Scientific Publishers.

Kajzar F, Agranovich VM and Lee CY-C (1996) *Photoactive Organic Materials: Science and Application*. Dordrecht: Kluwer Academic Publishers.

Kuzyk M and Dirk C (eds) (1998) *Characterization Techniques and Tabulations for Organic Nonlinear Optical Materials*. New York: Marcel Dekker.

Lee KS and Wegner G (eds) (2003) *Polymers for Photonics Applications II: Nonlinear Optical, Photorefractive and Two-Photon Absorption Polymers*. Advances in Polymer Sciences 161. Berlin: Springer-Verlag.

Lessard R and Franke H (eds) (1997) *Materials Characterization and Optical Probe Techniques*. Critical Reviews of Optical Science and Technology, vol. CR69. Bellingham: SPIE.

Messier J, Kajzar F, Prasad P and Ulrich D (1989) *Nonlinear Optical Effects in Organic Polymers*. Dordrecht: Kluwer Academic Publishers.

Messier J, Kajzar F and Prasad P (1991) *Organic Molecules for Nonlinear Optics and Photonics*. Dordrecht: Kluwer Academic Publishers.

Miyata S and Sasabe H (eds) (1997) *Poled polymers and their Application to SHG and EO Devices*. Amsterdam: Gordon and Breach Scientific Publishers.

Prasad PN and Williams DJ (1991) *Introduction to Nonlinear Optical Effects in Molecules and Polymers*. New York: Wiley.

Richardson T (ed.) (1999) *Functional Polymers and Polymeric Materials*. Chichester, UK: John Wiley & Sons.

MICROSCOPY

Contents

Overview
Confocal Microscopy
Imaging Multiple Photon Fluorescence Microscopy
Interference Microscopy
Nonlinear Microscopy
Phase Contrast Microscopy

Overview

C J R Sheppard, National University of Singapore, Singapore

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The basic construction and imaging properties of the optical microscope, one of the most important optical instruments, are described. But the restrictive meaning of a microscope as a device that the user looks down to see a magnified image is far from the present terminology that includes scanning systems (where the image is stored in a computer), nonvisible radiation, such as electrons or X-rays, and a variety of different contrast mechanisms.

According to the *Oxford English Dictionary* a microscope is 'an instrument magnifying objects by means of lenses so as to reveal details invisible to the naked eye'. Thus, basically a microscope forms a magnified image of an object. The first microscopes used visible light and formed an image showing variations in the intensity of light scattered by the object. This intensity is in general related to the optical properties of the object in a complicated way.

Nowadays there are many different types of microscope. Our understanding of a microscope must therefore be generalized in the following ways.

- The microscope may not be a conventional imaging system using lenses or mirrors. For example, it may be a scanning imaging system. A confocal system is a combination of a conventional and a scanning system. Another important category of microscopes is that of probe microscopes, including near-field optical microscopes. Some probe microscopes, for example the atomic force microscope (AFM) do not even seem at first glance to rely on the direct use of radiation.
- The microscope may not use visible light, but other electromagnetic radiation, or even other forms of radiation.
- The image can be formed using a variety of different contrast mechanisms. Some of these, such as phase contrast, are designed to image particular optical properties of the sample. Others image the generation of a form of radiation when the object is stimulated by another form of radiation. In particular, in principle, virtually any form of spectroscopy can be the basis of building up an image by measuring the spatial variations in the signal.

Different Forms of Microscopes

The Conventional Microscope

In a simple microscope, an objective lens forms a real, magnified, and inverted image of an object. In a compound microscope, an eyepiece is added, forming a virtual image that is viewed by eye to give real image on the retina. The eye can be replaced by CCD camera to record moving or still images. In early microscopes, the objective lens and eyepiece are mounted in the two ends of a brass tube. The objective screws into the tube using an RMS (Royal Microscopical Society) thread (0.8″ Whitworth 36 threads per inch), until its mounting face is flush with the end of the tube. The length of the tube is the mechanical tube length. The distance from the mounting face of the objective to the plane of the real image is the optical tube length, which varied according to different manufacturers in the range 160–210 mm. Now most objectives have infinity tube length, so that they collimate the light from

the object. The collimated light is brought to a focus by an additional tube lens. An objective of correct tube length should always be used as otherwise spherical aberration is introduced. As the objective, tube lens, and eyepiece are designed together as a system, great care should be exercised when mixing components from different manufacturers.

Abbe showed that in order to form a perfect lateral image the objective must satisfy a different condition called the Herschel condition. Otherwise the image of an off-axis point will suffer from coma. An aberration-free system satisfying the sine condition is called an aplanatic system. A microscope produces a three-dimensional (3D) image of a 3D object. A point of the object closer to the lens appears further from the lens in the image. If the lateral magnification is M, the longitudinal magnification is approximately equal to M^2. However, it can be shown that it is impossible to devise an instrument that can produce a perfect 3D image of a 3D object, because to have a perfect transverse image the system must satisfy Abbe's sine condition, but to have a perfect axial image the system must satisfy the Herschel condition. These two conditions are not mutually compatible. Actually, we find that the longitudinal magnification is not exactly equal to M^2, and is not even constant in space. Different depths of the object are imaged differently: the microscope objective only produces a good image for one plane of focus, and other planes will exhibit spherical aberration. However, all of these problems can be overcome by bringing the part of the object under observation into focus, without refocusing the eye or other detector. A 3D image can thus be generated by scanning the object stage in the axial direction, and recording a stack of images from different depths in a computer. For modern objectives of infinite tube length, focusing can be alternatively achieved by piezoelectric scanning of the objective lens.

Resolution

For an object consisting of a single bright point object in a dark background, the image produced by a perfect microscope, according to paraxial diffraction theory, is the so-called Airy disk, consisting of a bright central spot, surrounded by a series of rings. The intensity in a meridional cross-section through the Airy disk is the same for any aperture of objective, or wavelength of light, but its width varies. The intensity of the first bright ring is 1.75% of that at the peak. The radius of the first dark ring of the Airy disk is

$$r_0 = \frac{0.61\lambda}{n \sin \alpha} \qquad [1]$$

where λ is the wavelength, n is the refractive index of the immersion medium, and α is the angle subtended at the edge of the objective aperture. The radius depends on $n \sin \alpha$, which is called the numerical aperture and should be as large as possible for good resolution. Thus, high-magnification lenses often use an immersion fluid, usually oil with a refractive index of 1.518 (ISO 8036/1).

The optical coordinate v is defined as

$$v = \frac{2\pi}{\lambda} rn \sin \alpha \qquad [2]$$

so that the dark ring of the Airy disk occurs at a value of v of $2\pi \times 0.61 = 3.83$. Resolution is sometimes expressed in terms of the full width at half maximum (FWHM) of the image of a point object, which is 3.232 in optical coordinates. A full nonparaxial theory does not give a very different figure.

Resolution of a microscope is often specified by two-point resolution, which describes whether the image of two points can be distinguished from that of a single point. According to the Rayleigh criterion of two-point resolution, two points are just resolved if the second point is placed on the first dark ring of the first. The separation is then $0.61\lambda/(n \sin \alpha)$. A cross-section through the image in an incoherent optical system of two bright point objects of equal strength and different separations is shown in **Figure 1**. The intensity in the image between the points decreases as the separation is increased. The ratio of the intensity midway between the points to that at the points for the Rayleigh criterion to be satisfied is then found to be 0.735.

The concept of resolution should be distinguished from sensitivity or precision. An object much smaller than the resolution limit can still be detected, perhaps weakly, in a microscope: this detection of weak contrast depends on the sensitivity of the microscope rather than its resolution. The size of an object can also be measured with a precision much greater than the resolution of the microscope.

Illumination

If the object is not self-luminous, in order to be imaged in a microscope it must be illuminated. A semitransparent object, such as a biological slice, is illuminated in transmission. For observation of bulk objects, or surfaces, we use illumination in the reflection geometry, called epi-illumination. Either a tungsten halogen lamp or an arc source is usually used. In critical illumination, the source is focused on to the object by a condenser lens. The disadvantage of this approach is that variations in emission of the source are imaged directly into the image.

Cheaper microscopes avoid this problem by using a diffuser. Better microscopes use instead the Köhler illumination system, where the source is placed in the front focal plane of the condenser lens. Figure 2 shows a Köhler illumination system in which the source is imaged into the front focal plane of the condenser lens by another, projector, lens. The system incorporates a separate field stop and aperture stop. In practice, the size of the field stop should be reduced to illuminate as small a region of the object as necessary in order to minimize stray light. An important part of setting up a microscope to operate properly consists in centering the aperture stop of the condenser. This is often done using a so-called Bertrand lens that allows imaging of the aperture stop.

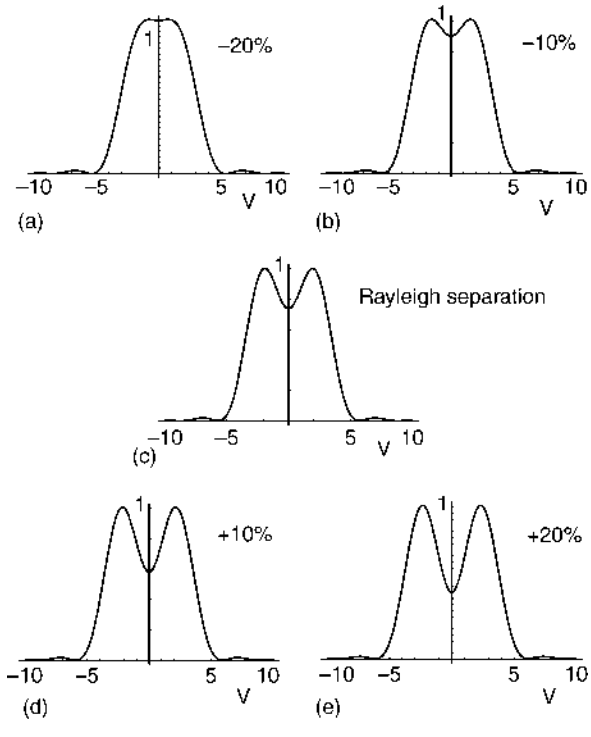

Figure 1 A cross-section through the image of two equal point objects in an incoherent microscope. The separation of the points is that for the Rayleigh criterion to be satisfied, and for changes of ±10, 20% of the Rayleigh separation.

Image formation

The Rayleigh criterion was originally specified for an incoherent optical system. It is thus applicable in fluorescence microscopy, as fluorescence is an incoherent process as two fluorescent point objects fluoresce independently. But for the image of a trans-illuminated object, the resolution depends on the coherence of the illumination. This is controlled by the numerical aperture of the condenser lens as compared with that of the objective. The ratio of these numerical apertures is called the coherence parameter, S, so that

$$S = \frac{n_c \sin \alpha_c}{n \sin \alpha} \qquad [3]$$

For $S = 0$ the illumination is purely coherent, while for $S \rightarrow \infty$ imaging becomes purely incoherent, thus corresponding to the Rayleigh criterion, so S should really be called an incoherence parameter. An important in-between case is when $S = 1$, when the apertures of objective and condenser are equal. This is termed matched, full, or complete illumination. According to the generalized Rayleigh criterion, the points are just resolved when the ratio of the intensity midway between the points to that at the points is 0.735, which is called the generalized Rayleigh criterion. We require that the distance between the points is as small as possible when this occurs. The generalized Rayleigh separation is the same for incoherent illumination and for $S = 1$, and corresponds to a distance of $2v_0 = 3.83$ in optical coordinates. Note, however, that for other separations of the points, the ratio is different for these two cases: it is not true, but often erroneously stated, that $S = 1$ corresponds to incoherent imaging. Resolution improves as the aperture of the condenser is increased ($2v_0 = 5.15$ in optical coordinates for coherent illumination) reaching a maximum when $S = 1.46$, when resolution is 9% better than for incoherent imaging ($2v_0 = 3.58$). This maximum in resolution is achieved when the numerical aperture of the condenser is larger than that of the objective, which is not

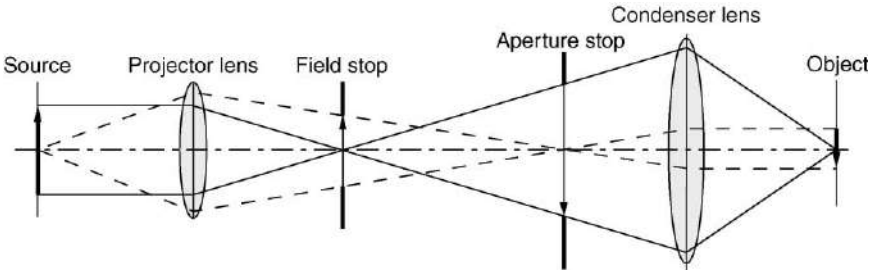

Figure 2 Geometry of the Köhler illumination system. In practice an arrangement with a smaller number of lenses is usually used.

physically achievable if the objective has the highest possible numerical aperture. Similarly, it is impossible to achieve incoherent illumination with very large objective numerical apertures.

The effect of coherence on the resolution of an image depends on the form of the object, and on the particular resolution criterion employed. In practice, contrast also depends on the aperture of the condenser. Usually, opening the condenser improves resolution, but reduces contrast, so there will be an optimum condenser aperture size.

Abbe theory of microscope imaging

Abbe argued that for coherent illumination of a grating, the grating spectra can be observed in the back focal plane of the objective. The tube lens then forms an image from the grating components. Thus, this is an early description of Fourier optics. The strength of the grating components can be altered in the back focal plane to give various optical effects, such as phase contrast. Abbe's theory was not properly appreciated at the time because it was known that in practice, resolution could be improved by using a larger condenser aperture, which does not give coherent illumination. There was much controversy about the merits of Köhler versus critical illumination, and different sizes of condenser aperture.

We now know that Köhler and critical illumination are in principle equivalent. The aberrations of the condenser lens are not important, so that the source and condenser together behave simply as a partially coherent effective source. Imaging can still be described by Fourier optics within the framework of partially coherent imaging theory.

Although Köhler and critical illumination are equivalent in principle, Köhler illumination is in practice better because of its improved illumination uniformity.

Depth of focus

The longitudinal image of a point object is found to be approximately invariant when expressed in terms of an optical coordinate u, defined as

$$u = \frac{8\pi}{\lambda} zn \sin^2 \frac{\alpha}{2} \qquad [4]$$

The axial resolution, and thus also the depth of focus, varies strongly with aperture. The FWHM for longitudinal imaging is 11.13 in optical coordinates.

Microscope objectives

Because microscope objectives often have large apertures, aberrations will be very strong, unless the objective is designed correctly. The basic design principle for a high aperture lens is the aplanatic front element (Figure 3). It is found that all rays from a point A in a sphere of refractive index n, where A is a distance r/n from its center, appear to come from a point B, distance rn, without any aberration. This principle is used by employing a front element that has one surface of the same radius of curvature as the sphere. The front surface is a sphere centered on A (Figure 3b), so that the rays are not deviated on crossing it.

A low-power achromat uses just two components with different optical dispersion to cancel chromatic aberration and correct for other aberrations (an achromatic doublet). For higher powers, two separated doublets are necessary. For the highest apertures, an aplanatic front is used. However, this can never converge the light, but only makes it less divergent. It is thus used together with an achromatic doublet. For oil immersion, the interface between the oil and the front element is not very important, so it is made planar to simplify manufacture. In practice, it is

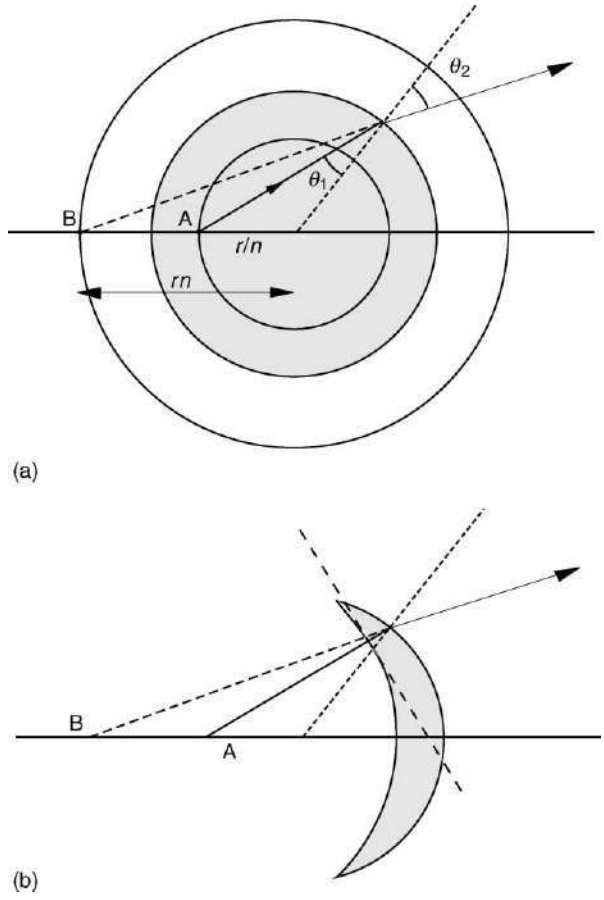

Figure 3 The principle of the aplanatic front: (a) an aplanatic surface; (b) a meniscus lens that acts as an aplanatic system.

usually made planar even for a dry lens, and the resulting aberrations cancelled out elsewhere in the objective. Apochromats, corrected for three colors, use more elements, which may include two or more stages of aplanatic front. The final objective thus consists of a number of elements that must be accurately aligned relative to one another. The objective is adjusted by the manufacturer to give a good star image (image of a point object). First a spacer is selected to optimize spherical aberration. Then a sleeve is adjusted to make the objective parfocal with others. Finally a screw is used to center the assembly to remove coma.

Conventional and Scanning Microscopes

Conventional microscope

Various different types of transmission microscope are illustrated in Figure 4. We consider image formation in a conventional microscope as illustrated schematically in Figure 4a, showing a microscope with critical illumination. A large area incoherent source is focused by the condenser lens on to the specimen, illuminating a comparatively large area of the specimen, corresponding to the whole field of view of the objective. Information from each illuminated point in the specimen is simultaneously transmitted by the objective lens to form the primary image. The objective is responsible for forming the image, with the condenser playing only a secondary role in determining the resolution of the system, through control of the coherence of the illumination. Figure 4b shows a conventional microscope in which the image is measured point by point by a detector. In practice this could be achieved by using a CCD detector.

Scanning microscope

An image can be generated by a scanning system, as illustrated in Figure 4c. A probe of light is formed by demagnification of a source, and is scanned over the object in a raster. The transmitted (or reflected) light is detected by a photodetector and thus builds up an image. The size of the probe of light limits the resolution of the system, that is, the smallest detail in the object that can be seen in the image. It should be noted that the magnification of the image is given simply by the ratio of the distance scanned in the image to the distance scanned by the probe. It is thus unrelated to the demagnification of the source. The image magnification can be altered without changing the lens and is not in any way related to the resolution.

It has been shown that the imaging properties of scanning and conventional microscopes are identical under analogous conditions, this being known as the principle of equivalence. This property is based on the principle of reciprocity, which is a very general physical law that holds even with diffraction,

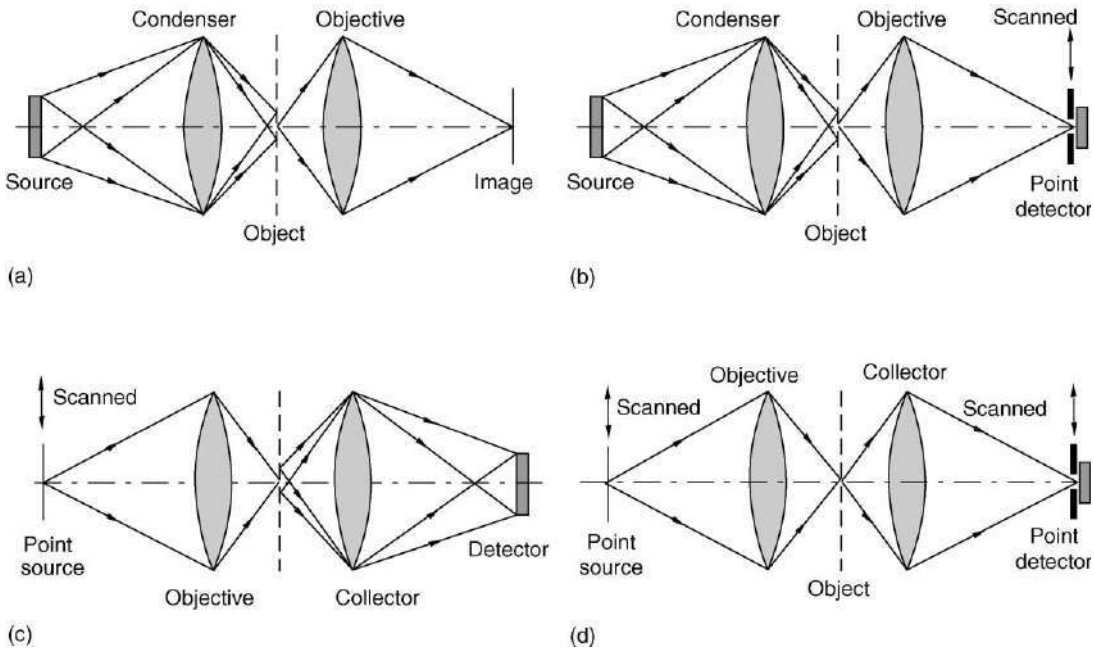

Figure 4 Different geometries of transmission microscope: (a) a conventional microscope; (b) a conventional microscope with a point detector; (c) a scanning microscope; (d) a confocal microscope.

absorption, multiple scattering, and stray light. The principle of equivalence is generally valid except if there are nonreciprocal magnetic or polarization effects, or energy losses involved (such as in fluorescence microscopy, and with inelastic scattering in electron microscopy).

Scanning systems exhibit a number of important advantages over conventional systems. Broadly, these are based on two classes of property. First, in a scanning system the image is in the form of an electronic signal. As a result, this is advantageous for quantitative measurements, as well as for image processing, including image enhancement, image restoration, and image analysis. Second, in a scanning system the object is illuminated by a focused spot, which extends the range of imaging modes available.

Confocal microscope

Finally in Figure 4d, we combine the arrangements of Figures 4b and c to give a confocal scanning optical microscope, in which a point source illuminates just a small region of the object, and a confocal point detector detects light from this illuminated region. If the point source and detector are scanned in unison, a two-dimensional image is generated. However, this system now behaves very differently from the previous ones. The confocal microscope is thus not a special case of the general partially coherent conventional imaging system. From the symmetry of Figure 4d, it is clear that the two lenses play an equal part in the imaging process. This results in an improvement in resolution. In fact, the confocal system behaves as a coherent imaging system, but with a sharper effective point spread function than in a conventional coherent microscope.

Although Figure 4 is drawn for the transmission geometry, in practice most confocal systems operate in the reflection or epi-illumination mode, in which the same objective lens is used both for illumination and detection. For a specimen placed in the focal plane, the properties of confocal transmission and reflection systems are identical. However, once the object is moved from the focal plane, certain differences arise: in particular in confocal reflection a strong optical sectioning effect occurs that allows a single section through a thick object to be imaged. This is the major advantage of the confocal microscope arrangement. Figure 4 also applies equally well to fluorescence imaging. However, in this case, because of the incoherent nature of fluorescence emissions after excitation by coherent light, imaging is then incoherent.

Confocal microscopes can be achieved by using either a point detector, in practice performed by placing a pinhole in front of a photodetector, or by using a coherent detector. Such a coherent detector, sensitive to the amplitude of the incident radiation, is not directly available for light, but is available for acoustic radiation, as in a scanning acoustic microscope. For light, a confocal effect can also be achieved by using an interferometric method to synthesize a coherent detector. Thus, in interference microscopes using the high aperture condenser and objective lenses, an optical sectioning effect arises similar to that in the confocal microscope. Another way of producing a coherent detector is to use a single-mode optical fiber.

Probe Microscopes

Scanning tunneling microscope

The scanning tunneling microscope was the first of the family of probe microscopes, relying on a fundamentally different principle from the usual forms of microscope. A physical tip in the nanometer range is brought close, also in the nanometer range, to a conducting sample. If an electric potential is applied, electrons can tunnel across between the sample and the tip. By scanning the tip mechanically across the sample, an image can be generated with a resolution in the subnanometer range. The surface topography can also be measured with subnanometer sensitivity. While interference microscopy can measure surface topography with subnanometer sensitivity, in this case the profile is averaged over the lateral resolution of the microscope, of the order of the wavelength in dimensions. The scanning tunneling microscope can image individual atoms. The signal is related to the work function of the material. Altering the bias voltage allows the band structure of the material to be investigated.

Atomic force microscope

In the atomic force microscope, the force between the surface and the tip is measured using a cantilever beam and used to build up an image. Atomic force microscopy can be performed in either a contact or noncontact mode. It can be used with insulating specimens. Optical methods, often using a position-sensitive detector, are used to measure displacements of the cantilever.

Near-field scanning optical microscope (NSOM)

In a near-field scanning optical microscope, a very small tip or aperture is scanned relative to the specimen to attain resolution greater than the classical limit set by the wavelength of the radiation. As Figure 5 shows, various different designs of near-field microscope have been proposed. It is evident that

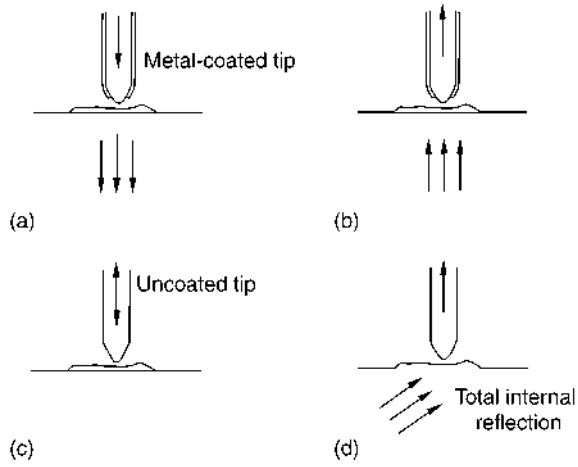

Figure 5 Different forms of near-field microscope: (a) illumination-mode NSOM; (b) collection-mode NSOM; (c) a near-field probe used for both illumination and collection in a confocal arrangement; (d) photon tunneling microscope (PTM). Reproduced with permission from Sheppard CJR (1978) The scanning optical microscope. *The Physics Teacher* 16: 648–651. © 1978 American Association of Physics Teachers.

these again form into the conventional/scanning categories described earlier. The sample can be illuminated using the near-field probe, as in Figure 5a. This is called the illumination-mode NSOM. Or the sample can be uniformly illuminated, and a signal detected using a near-field probe as in Figure 5b, giving a collection-mode NSOM. Or a confocal arrangement can be used, with a near-field probe used for both illumination and collection (c). In this case, an uncoated tip can be used; as otherwise the detected signal will be very small. Finally, Figure 5d shows the photon tunneling microscope (PTM), in which the sample is illuminated with evanescent waves produced by total internal reflection and an uncoated tip used to probe the evanescent field in the presence of the sample.

Different Types of Radiation

Electromagnetic Radiation

UV radiation and X-rays

The first microscopes used visible light, but, in fact, we do not need to use visible light, but can use any electromagnetic radiation, over a broad range of different available wavelengths. In Figure 6 we show the wavelength of the electromagnetic spectrum, illustrating how the wavelength in air varies with frequency. Instead of using visible light, shorter wavelengths allow greater resolution to be achieved. Ultra-violet light is commonly used to excite fluorescence. X-rays have also successfully been used to image biological samples.

IR and microwave radiation

In the longer wavelength region, infra-red radiation is used for observation of semiconductors or for molecular spectroscopic imaging. Imaging of emitted midinfra-red radiation can show up variations in temperature, such as hot spots in semiconductors. Microwaves have also been used, but resolution is limited by the long wavelength. For this reason microwaves (and mid-IR) have also been used for near-field microscopy.

Other Radiation

Electron microscope

The other important class of radiation that can be used for illumination is matter waves. The most common of these, of course, is electrons, the wavelength of which, as a function of acceleration voltage, is shown in Figure 7. The bend in the curve at very high voltages is caused by relativistic effects. The wavelength of the electrons is very small: even for 100 V electrons, the wavelength is only 0.1 nm. But aberrations caused by the electron lenses limit

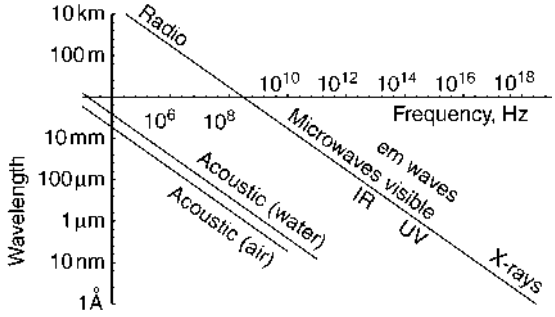

Figure 6 The wavelength of different forms of electromagnetic radiation and acoustic radiation. Reproduced with permission from Sheppard CJR (2002) The generalized microscope. In: Diaspro A (ed.) *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances*, pp. 1–18. New York: Wiley-Liss.

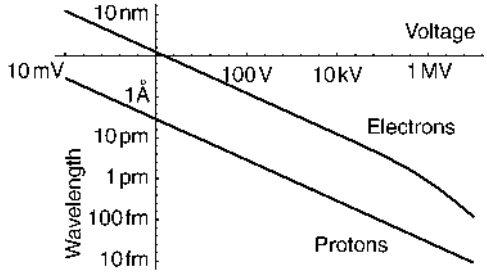

Figure 7 The wavelength of matter waves (electrons and protons) as a function of accelerating voltage. Reproduced with permission from Sheppard CJR (2002) The generalized microscope. In: Diaspro A (ed.) *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances*, pp. 1–18. New York: Wiley-Liss.

resolution so that at present atomic resolution is only just achievable, even with much higher voltages.

In a transmission electron microscope (TEM), a beam of electrons is produced by an electron gun. The specimen is illuminated via a condenser lens (or lenses). Usually these are magnetic lenses in which a current in a coil produces a magnetic field that focuses the electrons. The electrons that are transmitted through the specimen are then focused by a series of lenses to form a final image that can be viewed directly on a phosphor screen. An X-ray detector can be used to characterize the elements in the specimen. An electron-energy-loss spectrometer, together with a detector, can measure the energy of the electrons in a particular region of the image.

Other matter radiation

Instead of electrons, other particle beams can be used for illumination. Figure 7 shows the wavelength of a beam of protons, as function of the accelerating voltage. Because of the larger mass of protons, relativistic effects are not apparent for protons in the curve in Figure 7. Microscope images have also been formed using ions, and even neutral particles such as neutrons. Neutral atom microscopes are also under development.

Acoustic radiation

Instead of electromagnetic waves, acoustic radiation can be used to form a scanning acoustic microscope (SAM). Ultrasonic waves are focused into a water immersion medium. Mechanical scanning is again used to build up an image. Figure 6 shows the wavelength of the acoustic wave as a function of frequency. By using microwave frequencies, wavelengths, and therefore resolution, in the submicron range, can be generated. The SAM acts as a confocal microscope because the detector is sensitive to the amplitude of the acoustic wave. The acoustic microscope provides information on variations in the elastic properties (viscosity and elasticity) of the sample.

Different Contrast Mechanisms

The bright field microscope detects the intensity of the light transmitted through the object. A microscope with a condenser aperture of appreciable size behaves as a partially coherent system, and the image will, in general, depend on the phase of the light coming from the object as well as its intensity. Many objects are weakly scattering, so the contrast of this phase information is weak. In this case, the microscope effectively images variations in the transmittance or reflectance of the sample. Different designs of microscope can be used to image various properties of the object, thus providing new and complementary information. Phase contrast microscopy provides but one example of a different contrast mechanism. In fact, in general, any physical or chemical interaction of the illuminating radiation with the sample can be used as the basis for a contrast mechanism. These different contrast mechanisms can be performed in either a conventional or scanning arrangement, or in a confocal system.

Phase Contrast Microscopy

Phase contrast is a widely used technique in both biological microscopy (in transmission) and materials microscopy (in reflection). Its importance in biological microscopy stems, firstly, from the fact that contrast may otherwise be too weak to be visible, but secondly because the phase variations are caused by changes in optical thickness, related to physical density. In reflection microscopy, phase is related to surface height variations, but also depends on the optical material properties of the sample.

Polarization Microscopy

The polarization state of light transmitted through or reflected from a sample can also be changed. This change is detected in polarization microscopy, for example by placing the sample between crossed polarizers. The birefringence detected can be either material birefringence, or form birefringence (caused by the shape of the microstructure). Polarization microscopy is widely used in biological microscopy, and in mineralogy.

Fluorescence Microscopy

In fluorescence microscopy, electrons excited by illumination decay to the ground state and emit photons of light. Autofluorescence is natural fluorescence of the sample. Fluorescent dyes can be used as labels for particular biological or chemical constituents. Fluorescence microscopy is usually performed in the epi-illumination geometry to assist rejection of light from the source. Often it is performed in a confocal microscope, allowing 3D localization of the fluorescent labels.

Raman Microscopy

Raman spectroscopy can be performed on a microscopic scale to produce images of presence of particular molecular bonds. It is usually performed in a scanning system, often of the confocal type.

This is one example of the wider class of spectroscopic microscopies: in principle, any form of spectroscopy can be used as the basis for a microscope contrast mode.

Nonlinear Microscopy

Another class of contrast mechanisms is based on nonlinear optical interactions. Again, any nonlinear mechanism can be used as the basis for a contrast mechanism. These include two- (or multiphoton) fluorescence, second- (or higher-order) harmonic generation, sum or difference frequency generation, and coherent anti-Stokes Raman scattering (CARS). Probably other nonlinear optical mechanisms will be exploited in microscopy in the future.

Other Contrast Mechanisms

The contrast mechanisms described above, all rely on illumination and detection of light. But in many other modes, radiation of one form can be converted into another. These include cathodoluminescence microscopy (electrons to light), X-ray microanalysis (electrons to X-rays), photo-emission microscopy (light, usually UV, to electrons), photo-acoustic microscopy (light to ultrasonics), photothermal microscopy (light to thermal waves), optical or electron beam induced current (OBIC/EBIC). Again, in many of these, the illuminating or emitted radiation can be focused, or a combination can be used in a confocal arrangement.

See also

Microscopy: Confocal Microscopy; Imaging Multiple Photon Fluorescence Microscopy; Interference Microscopy; Nonlinear Microscopy; Phase Contrast Microscopy.

Further Reading

Born M and Wolf E (1999) *Principles of Optics*, 7th edn. Cambridge: Cambridge University Press.

Bradbury S and Bracegirdle B (1997) *Introduction to the Optical Microscope*. Oxford: Bios Scientific Publishers.

Inoue S and Spring KR (1997) *Video Microscopy: The Fundamentals*, 2nd edn. New York: Plenum.

Martin LC (1966) *Theory of the Microscope*. London: Blackie.

Paesler MA and Moyer P (1996) *Near-Field Optics: Theory, Instrumentation, and Applications*. New York: Wiley InterScience.

Pluta M (1988) *Advanced Light Microscopy, Volume 1: Principles and Basic Properties*. Amsterdam: Elsevier.

Sheppard CJR (1987) Scanning optical microscopy. In: Barer R and Cosslett VE (eds) *Advances in Optical and Electron Microscopy*, vol. 10, pp. 1–98. London: Academic Press.

Sheppard CJR (2002) The generalized microscope. In: Diaspro A (ed.) *Confocal and Two-Photon Microscopy: Foundations, Applications, and Advances*, pp. 1–18. New York: Wiley-Liss.

Slayter EM and Slayter HS (1992) *Light and Electron Microscopy*. Cambridge: Cambridge University Press.

Wiesendanger R (1994) *Scanning Probe Microscopy and Spectroscopy*. Cambridge: Cambridge University Press.

Wilson T and Sheppard CJR (1984) *Theory and Practice of Scanning Optical Microscopy*. London: Academic Press.

Confocal Microscopy

T Wilson, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

It is probably fair to say that the development and wide commercial availability of the confocal microscope has been one of the most significant advances in light microscopy in the recent past. The main reason for the popularity of these instruments derives from their ability to permit the structure of thick specimens of biological tissue to be investigated in three dimensions. It achieves this important goal by resorting to a scanning approach together with a novel (confocal) optical system.

The traditional wide-field conventional microscope is a parallel processing system which images the entire object field simultaneously, which is a severe requirement for the optical components. We can relax this requirement if we no longer try to image the whole object at once. The limit of this relaxation is to require a good image of only one object point at a time. The price that we have to pay is that we must scan in order to build up an image of the entire field.

A typical arrangement of a scanning confocal optical microscope is shown in Figure 1 where the system is built around a host conventional microscope. The essential components are some form of mechanism for scanning the light beam (usually from a laser) relative to the specimen and appropriate photodetectors to collect the reflected or transmitted light.

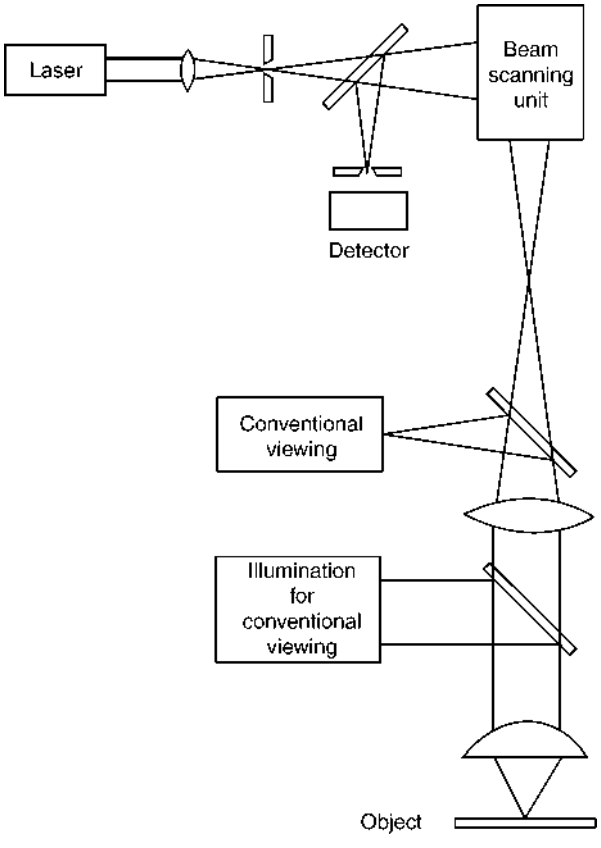

Figure 1 Schematic diagram of a confocal microscope.

Most of the early systems were analog in nature, but it is now universal, due to the serial nature of the image formation, to use a computer both to drive the microscope and to collect, process, and display the image.

In the beam scanning confocal configuration of Figure 1, the scanning is typically achieved by using vibrating galvanometer-type mirrors or acousto-optic beam deflectors. The use of the latter gives the possibility of TV-rate scanning, whereas vibrating mirrors are often relatively slow when imaging an extended region of the specimen, although significantly higher scanning speeds are achievable over smaller scan regions. We should note that other approaches to scanning may be implemented, such as specimen-scanning and lens-scanning. These methods, although not generally available commercially, do have advantages in certain specialized applications. Since this article is necessarily limited in length much additional material may be found in other sources listed in the Further Reading at the end of this article.

Image Formation in Scanning Microscopes

We will not discuss the fine detail of the optical properties of confocal systems. However, the essence is shown in Figure 1 where we see that the confocal optical system consists simply of a point source of light which is then used to probe a single point on the specimen. The strength of the reflected or fluorescence radiation from the single object point is then measured via a point, pinhole detector. The confocal – point source and point detector – optical system therefore merely produces an 'image' of a single object point and hence some form of scanning is necessary to produce an image of an extended region of the specimen. However, the use of single-point illumination and single-point detection results in novel imaging capabilities which offer significant advantages over those possessed by conventional wide-field optical microscopes. In essence, these are enhanced lateral resolution and, perhaps more importantly, result in a unique depth discrimination or optical sectioning property. It is this latter property which leads to the ability to obtain three-dimensional images of volume specimens.

The improvement in lateral resolution may at first seem implausible. However, it can be explained simply by the principle which states that resolution can be increased at the expense of field of view, which can then be increased by scanning. One way of taking advantage of this principle is to place a very small aperture extremely close to the object. The resolution is now determined by the size of the aperture rather than the radiation. In the confocal microscope we do not use a physical aperture in the focal plane but rather use the back-projected image of a point detector in conjunction with the focused point source. Figure 2 indicates the improvement in lateral resolution that may be achieved.

The confocal principle was introduced in an attempt to obtain an image of a slice within a thick specimen which was free from the distracting presence of out-of-focus information from surrounding planes. The confocal optical systems fulfills this requirement and its inherent optical sectioning or depth discrimination property has become the major

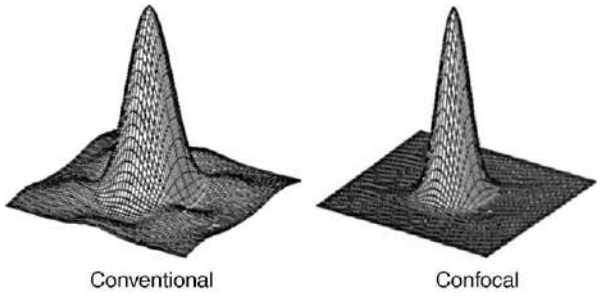

Figure 2 The point spread functions of the conventional and confocal microscopes showing the improvement in lateral resolution which may be obtained in the confocal case.

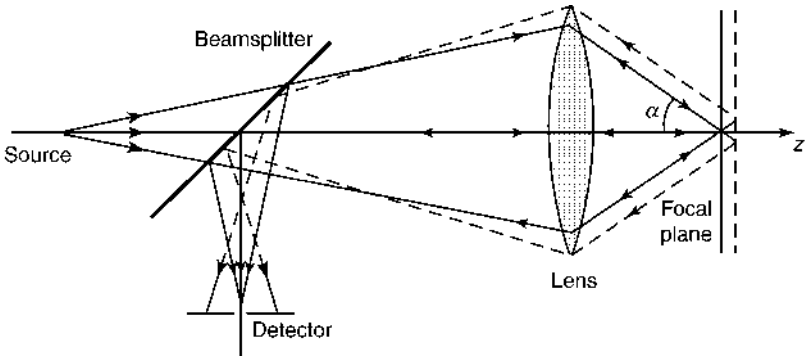

Figure 3 The origin of the optical sectioning or depth discrimination property of the confocal optical system.

motivation for using confocal microscopes, and is the basis of many of the novel imaging modes of these instruments. The origin of the depth discrimination property may be understood easily from Figure 3, where we show a reflection-mode confocal microscope and consider the imaging of a specimen with an undulating surface. The full lines show the optical path when an object feature lies in the focal plane of the lens. At a later scan position, the object surface is supposed to be located in the plane of the vertical dashed line. In this case, simple ray tracing shows that the light reflected back to the detector pinhole arrives as a defocused blur, only the central portion of which is detected and contributes to the image. In this way the system discriminates against features which do not lie within the focal region of the lens. A very simple method of both demonstrating the effect and giving a measure of its strength is to scan a perfect reflector axially through focus and measure the detected signal strength. Figure 4 shows a typical response. These responses are frequently termed the $V(z)$, by analogy with a similar technique in scanning acoustic microscopy. A simple paraxial theory models this response as

$$I(u) = \left[\frac{\sin(u/2)}{u/2} \right]^2 \qquad [1]$$

where u is a normalized axial coordinate which is related to real axial distance, z, via

$$u = \frac{8\pi}{\lambda} nz \sin^2(\alpha/2) \qquad [2]$$

where λ is the wavelength and $n \sin \alpha$ the numerical aperture. As a measure of the strength of the sectioning, we can choose the full width at half intensity of the $I(u)$ curves. Figure 5 shows this value as a function of numerical aperture for the specific case of imaging with red light from a helium neon laser. These curves were obtained using a high aperture theory which is more reliable than eqn [1]

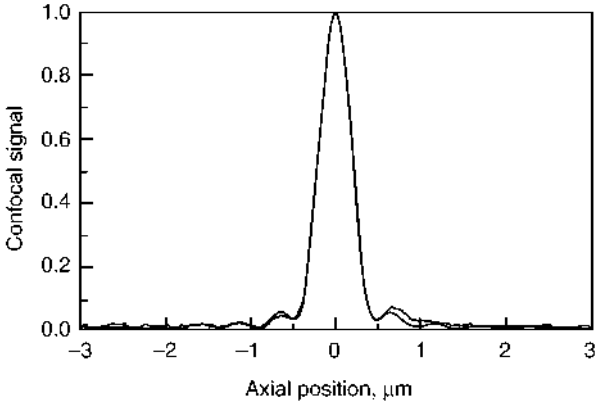

Figure 4 The variation in detected signal as a plane reflector is scanned axially through focus. The measurement was taken with a 1.3 numerical aperture objective and 633 nm radiation.

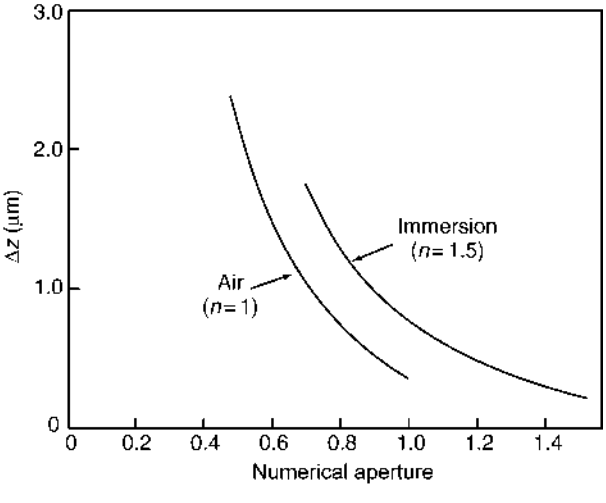

Figure 5 The optical sectioning width as a function of numerical aperture. The curves are for red light (0.6328 μm wavelength). Δz is the full width at the half-intensity points of the curves of $I(u)$ against u.

at the highest values of numerical aperture. We note, of course, that these numerical values refer to nonfluorescence imaging. The qualitative explanation of optical sectioning, of course, carries over to the fluorescence case but the actual value of the optical sectioning strength is different in the fluorescence case.

Applications of Depth Discrimination

Since this property is one of the major reasons for the popularity of confocal microscopes, it is worthwhile, at this point, to review briefly some of the novel imaging techniques which have become available with confocal microscopy.

Figure 6 illustrates the essential effect: Figure 6a shows a conventional image of a planar microcircuit which has deliberately been mounted with its normal at an angle to the optic axis. We see that only one portion of the circuit, running diagonally, is in focus. Figure 6b shows the corresponding confocal image: here the discrimination against detail outside the focal plane is clear. The areas which were out of focus in Figure 6a have been rejected. Furthermore, the confocal image appears to be in focus throughout the visible band, which illustrates that the sectioning property is stronger than the depth of focus.

This suggests that if we try to image a thick translucent specimen, we can arrange, by the choice of our focal position, to image detail exclusively from one specific region. In essence, we can section the specimen optically without having to resort to mechanical means. Figure 7 shows an idealized schematic of the process. The portion of the beehive-shaped object that we see is determined by the focus position. In this way it is possible to take a through-focus series and obtain data about the three-dimensional structure of the specimen. If we represent the volume image by $I(x, y, z)$, then by focusing at a position $z = z_1$, we obtain, ideally, the image $I(x, y, z_1)$. This, of course, is not strictly true in practice because the optical section is not infinitely thin.

It is clear that the confocal microscope allows us to form high-resolution images with a depth of focus sufficiently small that all the detail which is imaged appears in focus. This suggests immediately that we can extend the depth of focus of the microscope by adding together (integrating) the images taken at different focal settings without sacrificing the lateral resolution. Mathematically, this extended-focus image is given by

$$I_{\mathrm{EF}}(x, y) = \int I(x, y, z)\mathrm{d}z \qquad [3]$$

As an alternative to the extended-focus method, we can form an auto-focus image by scanning the object axially and, instead of integrating, selecting the focus at each picture point by recording the maximum in the detected signal. Mathematically, this might be written as

$$I_{\mathrm{AF}}(x, y) = I(x, y, z_{\max}) \qquad [4]$$

where $z_{\max}$ corresponds to the focus setting giving the maximum signal. The images obtained are somewhat similar to the extended focus and, again, substantial increases in depth of focus may be obtained. We can go one step further and turn the microscope into a noncontacting surface profilometer. Here we simply display $z_{\max}$.

It is clear by now that the confocal method gives us a convenient tool for studying three-dimensional structures in general. We essentially record the image as a series of slices and play it back in any desired fashion. Naturally, in practice it is not as simple as this, but we can, for example, display the data as an $x - z$ image rather than an $x - y$ image. This is somewhat similar to viewing the specimen from the side. As another

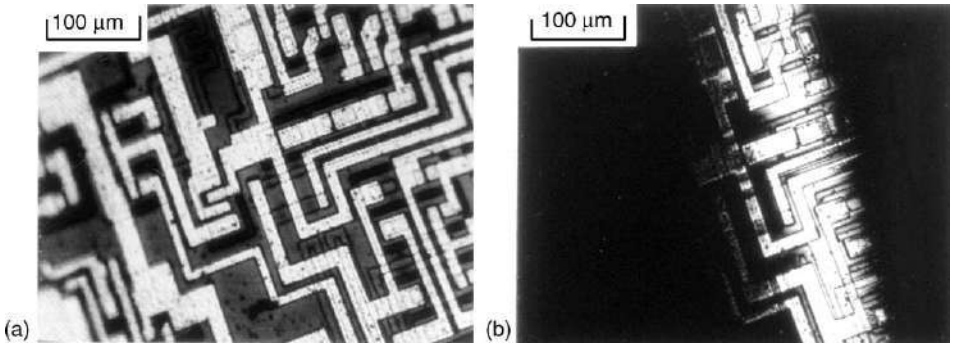

Figure 6 (a) Conventional scanning microscope image of a tilted microcircuit: the parts of the object outside the focal plane appear blurred. (b) Confocal image of the same microcircuit: only the part of the specimen within the focal region is imaged strongly.

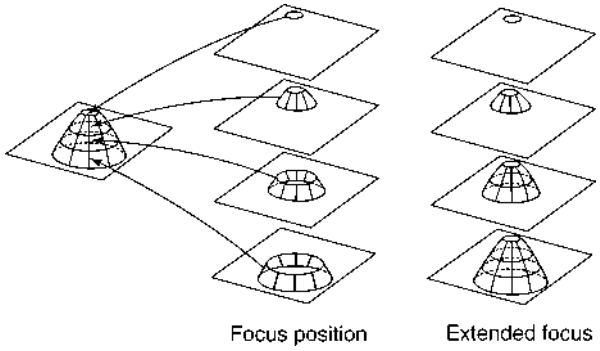

Figure 7 An idealization of the optical sectioning property showing the ability to obtain a through-focus series of images, which may then be used to reconstruct the original volume object at high resolution.

example, we might choose to recombine the data as stereo pairs by introducing a slight lateral offset to each image slice as we add them up. If we do this twice, with an offset to the left in one case and the right in another, we obtain, very simply, stereo pairs. Mathematically, we form images of the form:

$$\int I(x \pm \gamma z, y, z)\mathrm{d}z \qquad [5]$$

where γ is a constant. In practice it may not be necessary to introduce offsets in both directions to obtain an adequate stereo view.

Our discussion so far on these techniques has been by way of a simplified introduction. In particular, we have not presented any fluorescence images. The key point is that, in both brightfield and fluorescence modes, the confocal principle permits the imaging of specimens in three dimensions. The situation is, of course, more involved than we have implied. A thorough knowledge of the image formation process, together with the effects of lens aberrations and absorption, is necessary before accurate data manipulation can take place.

In conclusion to this section it is important to emphasize that the confocal microscope does not produce three-dimensional images. It essentially produces very high-quality two-dimensional images of a (thin) slice within a thick specimen. A three-dimensional rendering of the entire volume specimen may then be generated by suitably combining a number of these two-dimensional image slices from a through focus series of images.

Fluorescence Microscopy

We now turn our attention to confocal fluorescence microscopy, because this is the imaging mode which is usually employed in biological applications. Although we have introduced the confocal microscope in terms of bright-field imaging the comments we have made concerning the origin of the optical sectioning, etc., carry over directly to the fluorescence case although the numerical values describing the strength of the optical sectioning are, of course, different and we will return to this point later.

If we assume that the fluorescence in the object destroys the coherence of the illuminating radiation and produces an incoherent fluorescent field proportional to the intensity of the incident radiation, $I(v, u)$, then we may write the effective intensity point spread function, which describes image formation in the incoherent confocal fluorescence microscope, as

$$I(v, u)I\left(\frac{v}{\beta}, \frac{u}{\beta}\right) \qquad [6]$$

where the optical coordinates u and v are defined relative to the primary radiation and $\beta = \lambda_2/\lambda_1$ is the ratio of the fluorescence to the primary wavelength. We note that $v = (2\pi/\lambda)rn \sin\alpha$, where r denotes the actual radial distance.

This suggests that the imaging performance depends on the value of β. In order to illustrate this we show in Figure 8 the variation in detected signal strength as perfect fluorescent sheet scans through focus. This serves to characterize the strength of the optical sectioning in fluorescence microscopy in the same way that the mirror was used in the brightfield case. We note that the half-width of these curves is

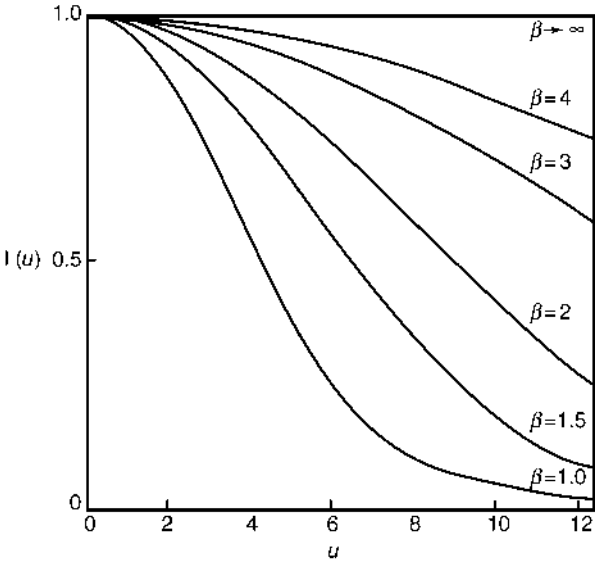

Figure 8 The detected signal as thin fluorescent sheet is scanned axially through focus for a variety of fluorescent wavelengths. We note that if we measure the sectioning by the half width of these curves, the strength of the sectioning is essentially proportional to β.

essentially proportional to β and so for optimum sectioning the wavelength ratio should be as close to unity as possible.

We have just discussed what we might call one-photon fluorescence microscopy in the sense that a fluorphore is excited by a single photon of a particular wavelength. It then returns to the ground state and emits a photon at the (slightly longer) fluorescence wavelength. It is this radiation which is detected via the confocal pinhole. Recently, however, much interest has centered on two-photon excitation fluorescence microscopy. This process relies on the simultaneous absorption of two, longer wavelength photons, following which a single fluorescence photon is emitted. The excitation wavelength is typically twice that used in the one-photon case. The beauty of the two-photon approach lies in the quadratic dependence of the fluorescence intensity on the intensity of the illumination. This leads to fluorescence emission which is always confined to the region of focus. In other words the system possesses an inherent optical sectioning property. Other benefits of two-photon fluorescence over the single photon case include the use of red or infra-red lasers to excite ultra-violet dyes, confinement of photobleaching to the focal region (the region of excitation), and the reduced effects of scattering and greater penetration. However, it should also be remembered that, compared with single photon excitation, the fluorescence yield of many fluorescent dyes under two-photon excitation is relatively low.

In order to make a theoretical comparison between the one- and two-photon modalities, we shall assume that the emission wavelength λ_{em} is the same, irrespective of the mode of excitation. Since the wavelength required for single-photon excitation is generally shorter than the emission wavelength, we may write it as $\gamma\lambda_{em}$ where $\gamma < 1$. Since two-photon excitation requires the simultaneous absorption of two photons of half the energy we will assume that the excitation wavelength may be written as $2\gamma\lambda_{em}$ which has been shown to be a reasonable approximation for many dyes.

If we now introduce optical coordinates u and v, normalized in terms of λ_{em}, we may now write the effective point spread functions in the one-photon confocal and two-photon case as

$$I_{1\text{p-conf}} = I\left(\frac{v}{\gamma}, \frac{u}{\gamma}\right)I(v,u) \qquad [7]$$

and

$$I_{2\text{p}} = I^2\left(\frac{v}{2\gamma}, \frac{u}{2\gamma}\right) \qquad [8]$$

respectively. We note that although a pinhole is not usually employed in two-photon microscopy it is possible to include one if necessary. In this confocal two-photon geometry the effective point spread function becomes

$$I_{2\text{p-conf}} = I^2\left(\frac{v}{2\gamma}, \frac{u}{2\gamma}\right)I(v,u) \qquad [9]$$

If we now look at eqns [7] and [8] in the $\gamma = 1$ limit, we find that

$$I_{1\text{p-conf}} = I^2(v,u) \qquad [10]$$

and

$$I_{2\text{p}} = I^2\left(\frac{v}{2}, \frac{u}{2}\right) \qquad [11]$$

We now see that because of the longer excitation wavelength used in two-photon microscopy the effective point spread function is twice as large as that of the one-photon confocal in both the lateral and axial directions. The situation is somewhat improved in the confocal two-photon case but it is worth remembering that the advantages of two-photon excitation microscopy are accompanied by a reduction in optical performance as compared to the single-photon case.

From the practical point of view it is worth noting that the two-photon approach has certain very important advantages over one-photon excitation, in terms of image contrast, when imaging through scattering media apart from the greater depth of penetration afforded by the longer wavelength excitation. In a single-photon confocal case it is quite possible that the desired fluorescence radiation from the focal plane may be scattered after generation in such a way that it is not detected through the confocal pinhole. Since the fluorescence is generated throughout the entire focal volume it is also possible that undesired fluorescence radiation, which was not generated within the focal region, may be scattered so as to be detected through the confocal pinhole. In either case this leads to a reduction in image contrast. The situation is, however, completely different in the two-photon case. Here the fluorescence is generated only in the focal region and not throughout the focal volume. Furthermore since all the fluorescence is detected via a large area detector – no pinhole is involved – it is not so important if further scattering events take place. This leads to high-contrast images which are less sensitive to scattering. This is particularly important for specimens which are much more scattering at the fluorescence ($\lambda/2$) wavelength than the excitation (λ) wavelength.

The Use of Structured Illumination to Achieve Optical Sectioning

An alternative approach to obtain optical sectioning is to use structured illumination in a conventional microscope. This approach is attractive since the conventional light microscope already possesses many desirable properties: real-time image capture, standard illumination, ease of alignment, etc. However, it does not produce optically sectioned images in the sense usually understood in confocal microscopy. In order to see how this deficiency may be corrected via a simple modification of the illumination system let us begin by looking at the theory of image formation in a conventional fluorescence microscope and start by asking in what way the image changes as the microscope is defocused. We know that in a confocal microscope the image signal from all object features attenuates with defocus and that this does not happen in a conventional microscope. However, when we look closely at the image formation process we find that it is only the zero spatial frequency (constant) component which does not change with defocus (Figure 9); all other spatial frequencies actually do attenuate with defocus to a greater or lesser extent. Figure 10 illustrates this by showing the image of a single spatial frequency one-dimensional bar pattern object for increasing degrees of defocus. When the specimen is imaged in focus, a good image of the bar pattern is obtained. However with increasing defocus the image becomes progressively poorer and weaker until it eventually disappears leaving a uniform gray level. This observation is the basis of a simple way to perform optical sectioning in a conventional microscope.

If we simply modify the illumination path of the microscope so as to project onto the object the image of a one-dimensional, single spatial frequency fringe pattern, then the image we see through the microscope will consist of a sharp image of those parts of the object where the fringe pattern is in focus together with an out-of-focus blurred image of the rest of the object. In order to obtain an optically sectioned image it is necessary to remove the blurred out-of-focus portion as well as the fringe pattern from the in-focus optical section. There are many ways to do this – one of the simplest involves simple processing of three images taken at three different spatial positions of the fringe pattern. The out-of-focus regions remain fairly constant between these images and the relative spatial shift of the fringe pattern allows the three images to be combined in such a way as to remove the fringes. This permits us to retrieve both an optically sectioned image as well as a conventional image in real time.

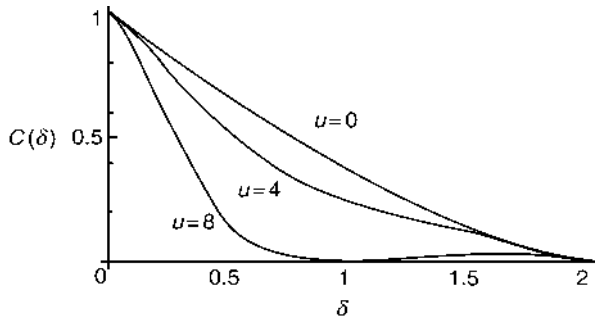

Figure 9 The transfer function for a conventional fluorescence microscope for a number of values of defocus (measured in optical units). We notice that the response for all nonzero spatial frequencies decays with defocus.

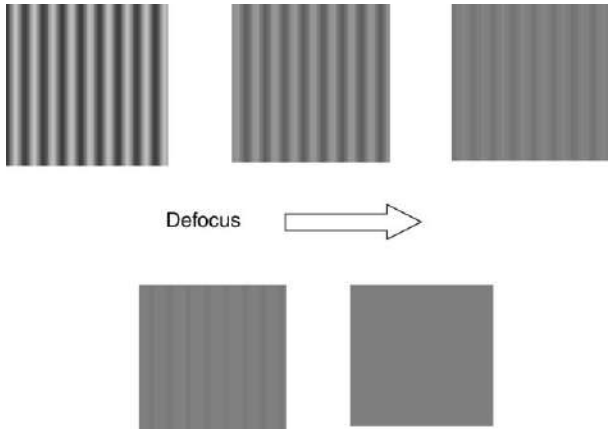

Figure 10 The image of a one-dimensional single spatial frequency bar pattern for varying degrees of defocus. We note that for sufficiently large values of defocus the bar pattern is not imaged at all.

Since the approach involves processing three conventional microscope images, the image formation is fundamentally different from that of the confocal microscope. However, the depth discrimination or optical sectioning strength is very similar and this approach, which requires very minimal modifications to the instrument, has been used to produce high-quality three-dimensional images of volume objects which are directly comparable to those obtained with confocal microscopes. A schematic of the optical arrangement is shown in Figure 11a, together with experimentally obtained axial responses in Figure 11b. It will be noted that these responses are substantially similar to those obtained in the true confocal case. As an example of the kind of images which can be obtained with this kind of microscope, we show in Figure 12 two images

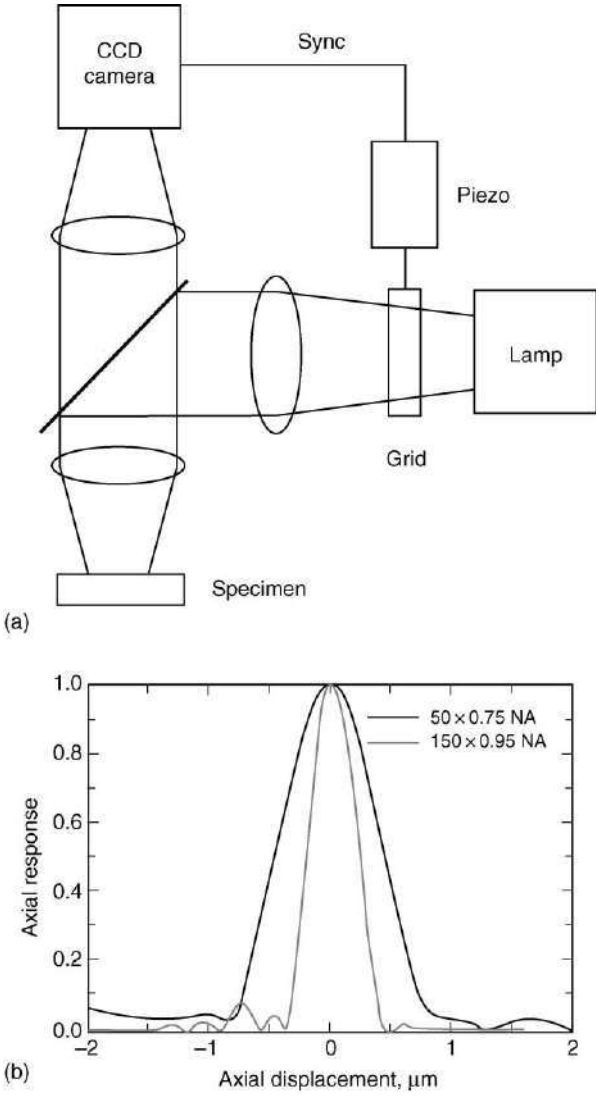

Figure 11 (a) Shows the optical system of the structured illumination microscope together with experimentally obtained axial responses in (b) which confirm the optical sectioning ability of the instrument.

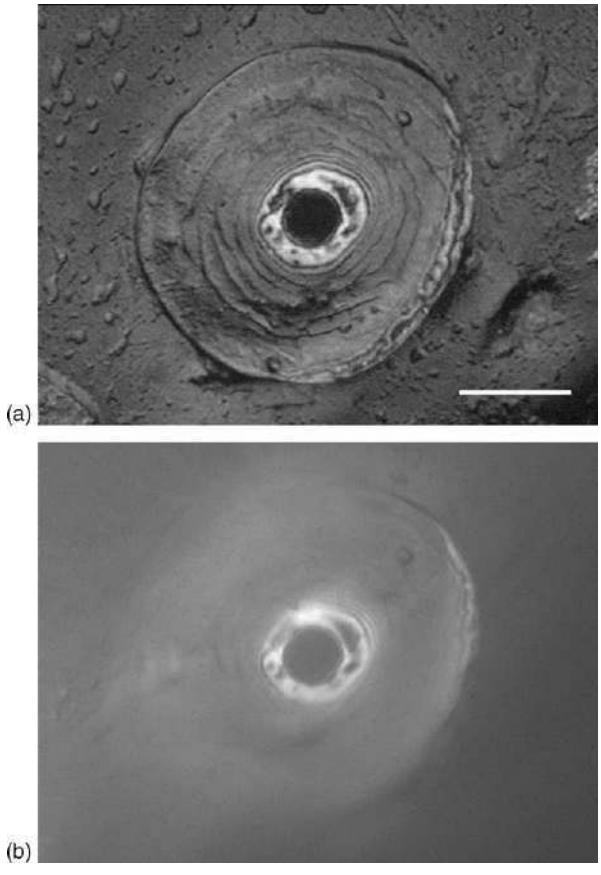

Figure 12 Two images of the region around the spiracle of a head louse. As in the previous figure, (a) is an autofocus image and (b) shows a mid-focus conventional image. The scale bar depicts a length of 10 μm.

of a spiracle of a head louse. The first is an auto-focus image of greatly extended depth of field constructed from a through-focus series of images. The second is a conventional image taken at mid-focus. The dramatic increase in depth of field is clear when compared with a mid-focus conventional image. We note that these images were taken using a standard microscope illuminator as light source. Indeed, the system is so light efficient that good-quality optical sections have been obtained or transistor specimens using simply a candle as light source!

Imaging using fluorescence light is also possible using this technique. However, an alternative approach that does not require a physical grid is possible if a laser is used as the light source. In this system the laser illumination is split into two beams which are allowed to interfere at an angle in the fluorescent specimen. This has the effect of directly 'writing' a one-dimensional fringe pattern in the specimen. Spatial shift of the fringe pattern is achieved by varying the phase of one of the interfering beams. As before, three images are taken from which both the optically sectioned image and conventional image may be obtained. The beauty of this approach is that no imaging optics are required at the illuminating wavelength and system.

Summary

We have discussed the origin of the optical sectioning property in the confocal microscope in order to introduce the range of imaging modes which this unique form of microscope leads to. A range of optical architectures have also been described. By far the most universal is that shown in Figure 1 where a confocal module is integrated around a conventional optical microscope. Other, more recent, real-time implementations have also been described. A number

of practical aspects of confocal microscopy have not been discussed. This is because they are readily available elsewhere, such as advice on the correct choice of detector pinhole size or because they are still the focus of active research, such as the development of new contrast mechanisms for achieving enhanced three-dimensional resolution such as the stimulated emission depletion method, STED or 4-Pi which are still under development but offer great promise.

See also

Microscopy: Imaging Multiple Photon Fluorescence Microscopy; Overview.

Further Reading

Corle TR and Kino GK (1996) *Confocal Scanning Optical Microscopy and Related Imaging Systems*. New York: Academic Press.

Denk W, Strikler JH and Webb WW (1990) Two photon laser scanning fluorescence microscopy. *Science* 248: 73.

Diaspro A (2002) *Confocal and Two-Photon Microscopy*. New York: Wiley-Liss.

Gu M (1996) *Principles of Three-Dimensional Imaging in Confocal Microscopes*. Singapore: World Scientific.

Hell S and Stelzer EHK (1992) Fundamental improvements of resolution with a 4Pi-confocal fluorescence microscope using two-photon excitation. *Optics Communication* 93: 277.

Hell S and Wichmann J (1994) Breaking the diffraction limit by stimulated emission. *Optics Letters* 19: 870.

Lukosz W (1966) Optical systems with resolving power exceeding the classical limit. *Journal of the Optical Society of America* 56: 1463.

Masters BR (1996) *Confocal Microscopy*. Bellingham: SPIE.

Minsky M (1961). Microscopy Apparatus. United States Patent 3,013,467, Dec 19, 1961 (Filed Nov 7, 1957).

Neil MAA, Juskaitis R and Wilson T (1997) Method of obtaining optical sectioning by using structured illumination in a conventional microscope. *Optics Letters* 22: 1905.

Pawley JB (ed.) (1995) *Handbook of Biological Confocal Microscopy*. New York: Plenum Press.

Wilson T (ed.) (1990) *Confocal Microscopy*. London: Academic Press.

Wilson T and Sheppard CJR (1984) *Theory and Practice of Scanning Optical Microscopy*. London: Academic Press.

Imaging Multiple Photon Fluorescence Microscopy

M Previte, Massachusetts Institute of Technology, Cambridge, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The field of light microscopy started with Ernst Abbe's introduction of the diffraction-limited objective lens in the late nineteenth century. Over the last 60 years, new imaging modalities, such as video, fluorescence, confocal, near field, and more recently multi-photon microscopy, have enabled the experimentalist to observe a plethora of new phenomena in the biological and physical sciences. Of these techniques, multi-photon microscopy has emerged not only as a valuable imaging technique but it also forms the technology base for emerging fields of optical memory and microfabrication.

The history of multi-photon microscopy (MPM) began with Göppert-Mayer's theoretical prediction of two-photon excitation in 1931. Although the theoretical groundwork was established, it took another 30 years for this process to be experimentally realized when Kaiser and Garrett observed two-photon absorption fluorescence emission from CaF_2:Eu^{+2}. Soon after this observation was made, Singh and Bradley witnessed higher-order multi-photon processes in 1964.

Historically, three-dimensional imaging was first achieved using confocal microscopy. In contrast to MPM, photo-interaction and fluorescence emission occurs throughout the excitation light path in confocal microscopy. Confocal microscopy selects the signal from the focal plane by the use of a conjugated pinhole aperture. In order to construct 3D images, successive planes must be imaged. As a result, the entire sample volume is exposed to the excitation light multiple times, and inevitably, photobleaching and photodamage occurs throughout the imaging volume. Furthermore, the pinhole aperture in the emission path causes additional signal reduction in optically thick specimens where the emission photon is scattered. Consequently, the utility of confocal microscopy is limited for imaging living biological specimens. The introduction of multi-photon microscopy, by Denk and Webb in 1990, addressed many of these limitations. By combining the methodology of multi-photon excitation with scanning microscopy, MPM offered advantages such as 3-D localized, femtoliter excitation volumes,

reduced scattering and absorption of the infrared excitation light by the specimen, and a large separation between the excitation and emission wavelengths. MPM can produce excellent three-dimensional images with increased signal to noise ratios.

Theoretical Basis of Multi-Photon Microscopy

Multi-photon Excitation Selection Rules

Unlike direct ground state to excited state one-photon transitions, multi-photon transitions depend on the presence of virtual electronic intermediate states to reach an excited state (see Figure 1).

Using perturbative expansion, the probability of excitation for one-photon transitions is described by the first-order solution, and multi-photon transitions can be determined by including the higher order solutions. The interaction of an electric field vector E_λ with a molecule's dipole, $\vec{r}$, is described by the respective one-photon and two-photon transition probabilities:

$$P_{1h\nu} \approx \langle \Psi_f | \vec{E}_\lambda \cdot \vec{r} | \Psi_i \rangle^2$$
$$P_{2h\nu} \approx \left| \sum_m \frac{\langle \Psi_f | \vec{E}_\lambda \cdot \vec{r} | \Psi_m \rangle \langle \Psi_m | \vec{E}_\lambda \cdot \vec{r} | \Psi_i \rangle}{\varepsilon_\lambda - \varepsilon_{m,i}} \right|^2 \quad [1]$$

where Ψ_i, Ψ_m, and Ψ_f are the initial, intermediate, and final transition states, $\vec{E}_\lambda$ is the electric field associated with the excitation light, ε_λ is the energy difference between the ground state and excited states, $\varepsilon_{m,i}$ is the energy difference between the intermediate state m and the ground state, i, and the summation is over all possible intermediate states m. Although electronic transitions, that result from the interaction of a strong electric field with a molecular dipole, are affected by vibrational mode overlap (Franck Condon factors) and anharmonic perturbative effects, the presence of intermediate transition states creates a set of general selection rules for multi-photon absorption processes.

Molecular states have eigenfunctions with a defined parity: 'gerade' describing even or symmetric states (no change in the sign of the wavefunction when the sign changes in the spatial coordinates) and 'ungerade' describing odd or antisymmetric states (the sign of the wavefunction changes when the sign changes in the spatial coordinates). Selection rules arise from the physics inherent in these even and odd eigenfunctions and the odd dipole operator that couples transitions between ground and excited states. In one-photon excitation, transitions are allowed for ground state and excited state transitions of opposite parity. In contrast, two-photon excitation occurs between ground and excited states with the same parity, by coupling through intermediate states with opposite parity. In general, ground and excited states with the same parity have a high excitation probability for an even number of photons; ground and excited states with different parity have a high transition probability for excitations involving an odd number of photons.

In the absence of major vibrational perturbations, inductive effects, or solvent relaxation, the probability of the transition depends on the magnitude of the integral overlap between the ground, intermediate, and excited states. As a result, molecule absorption cross-sections for one-photon, two-photon, three-photon, or multi-photon excitation processes can be estimated. Choosing fluorophores with higher multi-photon cross-sections allows more efficient excitation with lower excitation intensity. Finally, it should be noted that while the absorption cross-sections for one-, two-, and multi-photon excitation of a given fluorophores are different, the molecules rapidly relax to the same vibrational level in the excited electronic state. The excited state residence time (fluorescence lifetime), and the fluorescence decay processes depend only on the molecular structure and its microenvironment. Therefore, the fluorescence

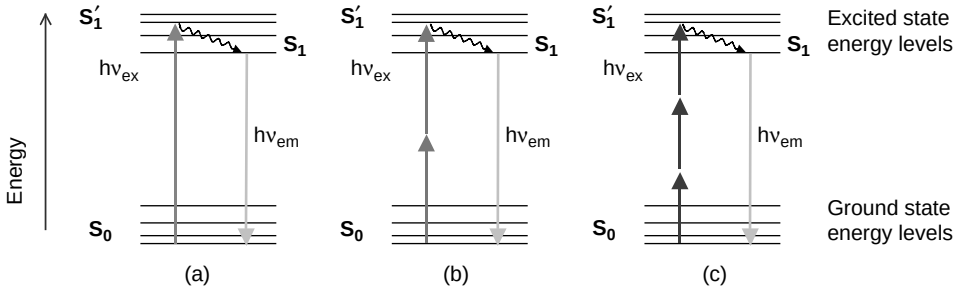

Figure 1 Jablonski diagram of a fluorescent dye molecule in one-, two- and *n*-photon excitation mode.

quantum yield and emission spectra is independent of the initial excitation process.

Enhanced Microscope Resolution Based on Multi-photon Excitation

The number of photons that are absorbed in a multi-photon process is proportional to the fluorophore concentration, $C(\vec{r},t)$, the molecule's absorption cross-section, σ_n, which depends on n, the number of photons involved in the absorption process, and the intensity of the excitation source, such that

$$N_{\text{abs}}(t) = \int \mathrm{d}\vec{r} I_0^n(\vec{r},t) C(\vec{r},t)\sigma_n \qquad [2]$$

When estimating $N_{\text{abs}}(t)$, the spatial and temporal intensity distribution of the excitation intensity are considered independently:

$$I_0(\vec{r},t) = S(\vec{r}) I_0(t) \qquad [3]$$

such that for an n-photon process, $I_0(t)$ is the temporal dependence of the excitation intensity at the maxima of the spatial distribution and $S(\vec{r})$ is the unitless spatial distribution of the excitation light. Combining eqns [2] and [3] gives

$$N_{\text{abs}}(t) = \int \mathrm{d}\vec{r} S^n(\vec{r}) I_0^n(t) C(\vec{r},t)\sigma_n \qquad [4]$$

If it is assumed that fluorophore distribution is uniform and the system does not undergo photobleaching or ground-state depletion, $C(\vec{r},t)$ can be considered a constant and the previous equation can be rewritten as

$$N_{\text{abs}}(t) = C\sigma_n I_0^n(t) \int_V S^n(\vec{r})\mathrm{d}\vec{r} \qquad [5]$$

By expanding the incident laser beam and overfilling the back aperture of the objective lens, the paraxial form of the normalized intensity point spread function (PSF, $h^2[u,v]$) determines the properties for a diffraction-limited lens. The dimensionless radial distance, v, from the optic axis and the distance from the in-focus plane, u, are given by

$$v = \frac{2\pi(\text{NA})r}{\lambda} \qquad [6]$$

$$u = \frac{2\pi(\text{NA})^2 z}{n_0\lambda} \qquad [7]$$

where $\text{NA} = n_0 \sin\theta$ (θ being the half-angle of collection for the lens and n_0 is the index of refraction the medium) is the numerical aperture. Expressing the spatial distribution function with the optical units, v, u, and θ gives the following equation:

$$S(\vec{r}) = h^2[u,v] = 2\left|\int J_0(v\theta)\mathrm{e}^{-(1/2)\mathrm{i}u\theta^2}\theta\,\mathrm{d}\theta\right|^2 \qquad [8]$$

where J_0 is the zero-order Bessel function. If an oil immersion objective with a $\text{NA} < 1.0$ is used, the paraxial approximation accurately estimates the spatial distribution, and the integral of $S(\vec{r})$ evaluates as

$$\int_{V=\infty} S^n(\vec{r})\mathrm{d}\vec{r} = a_n\left[\frac{n_0\lambda^3}{8\pi^3(\text{NA})^4}\right] \qquad [9]$$

where $a_n = \int_0^\infty 2\pi v\,\mathrm{d}v \int_{-\infty}^{\infty}\mathrm{d}u[h(u,v)]^{2n}$. Since this integral cannot be solved analytically, the values of a_n are determined using numerical integrations of $h^2(u,v)$ over all space. As a result, the volume integral of the spatial distribution is always proportional to NA^{-4} and therefore independent of the number of photons involved in the excitation process. For higher NA objectives, the paraxial form of the normalized intensity PSF may no longer be an adequate approximation. The expression for temporal part of eqn [5] can be written as

$$I_0(t) = \left(\frac{\pi\text{NA}^2}{\lambda^2}\right)P(t) \qquad [10]$$

where $P(t)$ is the excitation power and λ is the center wavelength of the pulsed excitation source. The temporal dependence can be described with the time average expression:

$$\langle I_0^n(t)\rangle = \left(\frac{\pi\text{NA}^2}{\lambda^2}\right)^n \frac{1}{T}\int_0^T P^n(t)\mathrm{d}t \qquad [11]$$

where T is an arbitrary time interval for continuous wave lasers and $T = 1/f_\text{p}$ for a pulsed laser with a repetition rate, f_p. Since multi-photon microscopy uses pulsed laser systems, the integral in eqn [11] must be evaluated accordingly. For a pulsed laser with a pulse width, τ, repetition rate, f_p, and averaged power, P_0, the pulse profile can be approximated as

$$P(t) = \frac{P_0}{f_\text{p}\tau} \qquad \textit{for } 0 < t < \tau \qquad [12]$$

$$P(t) = 0 \qquad \textit{for} \qquad \tau < t < \frac{1}{f_\text{p}} \qquad [13]$$

such that eqn [11] can be expressed as

$$\langle I_0^n(t)\rangle = \left(\frac{\pi\text{NA}^2}{\lambda^2}\right)^n \left(\frac{P_0}{f_\text{p}\tau}\right)^n \frac{1}{T}\int_0^T \mathrm{d}t \qquad [14]$$

and evaluated as

$$\langle I_0^n(t)\rangle = \left(\frac{\pi \mathrm{NA}^2}{\lambda^2}\right)^n (P_0)^n \left(\frac{1}{f_\mathrm{p}\tau}\right)^{n-1} \qquad [15]$$

Assuming no stimulated emission and no self-quenching, the quantity of emitted photons collected per unit time, $F(t)$, is

$$F(t) = \frac{1}{n}\phi\eta N_\mathrm{abs}(t) \qquad [16]$$

where η and ϕ are the fluorescence quantum efficiency of the dye and collection efficiency of the system, respectively. Since n photons are needed for each individual excitation event, $1/n$ is a normalization factor for N_abs which describes the total number of photons absorbed. Rewriting this expression in terms of the average number of absorbed photons, $\langle N_\mathrm{abs}(t)\rangle$, gives the average fluorescence intensity:

$$\langle F(t)\rangle = \frac{1}{n}\phi\eta\langle N_\mathrm{abs}(t)\rangle \qquad [17]$$

where $\langle N_\mathrm{abs}(t)\rangle$ is

$$\langle N_\mathrm{abs}(t)\rangle = C\sigma_n\langle I_0^n(t)\rangle \int_V S^n(\vec{r})\mathrm{d}\vec{r} \qquad [18]$$

such that the measured fluorescence intensity or time averaged photon flux is expressed as

$$\langle F(t)\rangle = \frac{1}{n}\phi\eta C\sigma_n\langle I_0^n(t)\rangle \int_V S^n(\vec{r})\mathrm{d}\vec{r} \qquad [19]$$

Since the time averaged fluorescence is proportional to $\langle I_0^n(t)\rangle$, eqns [9] and [15] can be substituted into eqn [19] to give a practical expression that relates average fluorescence to measurable laboratory quantities, such as P_0, τ_p, f_p, C, σ_n, n, η, and ϕ, NA, and associated numerical constants such that

$$\langle F(t)\rangle = \frac{1}{n}\phi\eta C\sigma_n\left(\frac{\pi \mathrm{NA}^2}{\lambda^2}\right)^n (P_0)^n \left(\frac{1}{f_\mathrm{p}\tau}\right)^{n-1} \times a_n\left[\frac{n_0\lambda^3}{8\pi^3(\mathrm{NA})^4}\right] \qquad [20]$$

As eqn [19] shows, the fluorescence intensity has an nth order dependence on the spatial and temporal distribution of the excitation power. Therefore, appreciable multi-photon probability occurs only at a region of high temporal and spatial concentration of photons. Confocal pinholes can be eliminated making this technique better suited for 3D biological imaging.

The higher order (>2) dependence of the fluorescence signal upon the excitation intensity distribution results in axial confinement. Experimentally, it can be shown that over 80% of fluorescence signal emanates from a 1 μm thick region about the focal point for objectives with a numerical aperture of 1.25 under two-photon excitation. This confinement is significantly different from one-photon excitation where a spatially uniform fluorescent sample contributes equal fluorescence intensities from each z-section above and below the focal plane (assuming negligible excitation attenuation). For an excitation wavelength of 960 nm, the typical point spread function of two-photon absorption has FWHM of 0.3 μm in the radial direction and 0.9 μm in the axial direction. In comparison with confocal microscopy, the localization of the volume of excitation instead of the volume of observation, as in confocal microscope, allows 3D imaging with minimal photobleaching and damaging.

Basic Imaging Applications of Multi-Photon Microscopy

Multi-photon microscopy finds an increasingly broad range of applications, especially in biological and medical areas. For example, multi-photon microscopy has been utilized to study the physiological and biological states in various highly scattering tissues such as cornea, skin, and pancreas islet. Also, multi-photon microscopy is used in embryology to study the development of sea urchin, *Caenorhabditis elegans*, hamster, and zebra fish. Neurobiological applications include qualifying calcium dynamics dendritic spines, memory and structural plasticity, hemodynamics, and plaque evolution in Alzheimer's disease. MPM has been applied to study pharmacological distribution in tissues and in the quantification of transdermal drug delivery. Due to minimally invasive capability of multi-photon microscopy, intra-vital microscopy imaging into thick tissues of living animals is possible leading to applications in the study of cancer biology.

Although the advantages of MPM are less compelling in studying cell cultures where the specimen is optically thin, a number of uniquely exciting uses have been identified. For instance, the significantly broader two-photon absorption spectra of many fluorophores allow convenient imaging of specimens labeled with multiple color fluorophores. The localized excitation volume reduces photo-interaction with fluorophores outside the excitation volume allowing better 3D images to be obtained from cellular systems labeled with easily photobleachable fluorophores. The use of three and higher photon excitation has also allowed the excitation and imaging of important biomolecules with deep ultraviolet

absorption. Another interesting application of multi-photon excitation in cellular systems is based on the ability of this method to produce localized chemical reactions. Through the photoactivation or photolysis of 'caged' molecules, their chemical structures are altered to produce biologically active molecules, fluorescent species, or other observable chemical changes. Multi-photon excitation has been used to locally uncage neurotransmitters for mapping ion channels in neurons and to uncage calcium in studying signal transduction events. In cellular systems, MPM is being used to image normally nonfluorescent biomolecules, such as serotonin in neurons, based on multi-photon induced chemical conversion and multi-photon excitation. Finally, multi-photon microscopy has also found applications in studying the conformation, aggregation, diffusion, and transport of single molecules and proteins. Two-photon fluorescence correlation spectroscopy and photon counting histograms are powerful techniques that have been applied to study protein aggregation, conformation changes, and diffusion *in vitro* with applications ranging from cell biology to high throughput drug discovery. The large separation of excitation and emission spectra of two-photon excitation also allows high sensitivity detection of fluorescence from single chromophore; however, the higher photobleaching rate at the excitation volume for some chromophores may lessen or even reverse this advantage.

Advanced Multi-Photon Microscopy Techniques

As described in the previous sections, excitation processes involving multiple numbers of photons offers considerable advantages for the imaging process in fluorescent microscopy. 4Pi Confocal Microscopy and high-speed Multifocal Multi-photon Microscopy MMM are advanced multi-photon microscopy techniques that capitalize on these advantages to improve image resolution and acquisition times in fluorescence imaging.

4Pi Confocal Microscopy

In contrast to near-field microscopy, far-field light microscopes, i.e. confocal and multi-photon microscopes, can create three-dimensional images of the interior of cells. The lateral and axial resolution of these systems is limited between 180 nm and 300 nm and 500–1000 nm, respectively. This limitation is dictated by the Abbe theory where resolution is inversely proportional to the numerical aperture of the focusing lens. When focusing with a single lens, geometric constraints limit the numerical aperture to about 1. 4Pi microscopy was developed to overcome this limitation. By utilizing two opposing objective lenses whose spherical wavefronts are coherently combined, a greater total effective numerical aperture is achieved. This results in a higher axial resolution, while the lateral resolution of the 4Pi microscope remains unchanged with respect to its confocal counterpart.

In the setup of a 4Pi microscope, the illuminating wavefronts emanate from a common point light source. The path is split by a beamsplitter, guided through two optically identical paths, and through two opposing high NA lenses that share a common focus. The fluorescence light is then collected through one lens and a confocal pinhole. In the axial direction, this arrangement yields a PSF with one main maximum lobe that is 3–4 times sharper along the optic axis than the PSF of a conventional confocal microscope, and two undesired side-lobes.

These side lobes can produce artifacts in the image and their effect needs to be suppressed for unambiguous 3D imaging. Lobe suppression can be achieved in a 4Pi system using two-photon excitation resulting in a quadratic decrease in the ratio of the side-lobe and the main lobe amplitudes. Using an excitation wavelength of 800 nm and 1.4 NA oil-immersion objectives, a FWHM along the axial direction of the main lobe is about 145 nm and side lobes are approximately 25–35% of the maximum lobe intensity. These side lobes can be further mathematically reduced by means of a linear filter after (or during) image acquisition. The filter can be applied at lobe heights below 50%, and an effective 3D PSF consisting of a single spot with a 250 nm lateral and 145 nm axial FWHM is achieved. When combined with nonlinear image restoration, 4Pi confocal microscopy achieves a 3D resolution in the 100 nm range. This resolution is similar to that obtained with near-field optical microscopy in biological applications with an added advantage of 3D imaging of the cell interior.

Since resolution improvement in 4Pi microscopy can be negated by the presence of aberrations, the usage of oil immersion lenses requires the specimen to be fixed and mounted in glycerol to minimize aberrations resulted from index mismatch. The sample thickness is further limited in the range of 20–30 μm. Due to significant index mismatches between oil and water, the imaging of live cells in aqueous medium is extremely difficult, if not impossible. In order to circumvent this difficulty, water immersion objective lenses are introduced into the 4Pi system. These lenses have a lower aperture angle $\alpha = 62°$ (NA = 1.2) than oil immersion lenses

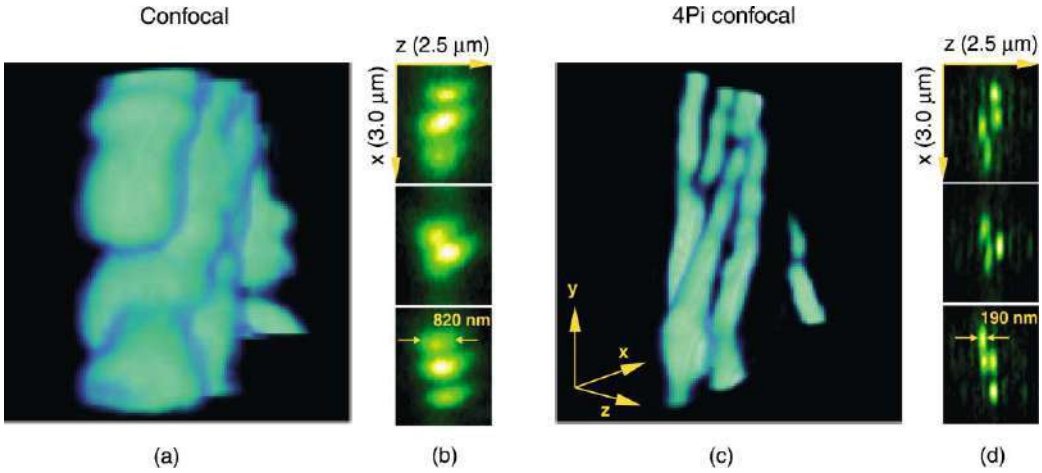

Figure 2 (a) Two-photon confocal and (c) 4Pi confocal point deconvolved images of an immunolabelled microtubule network in fixed NIH3T3 cells. The 3D rendered images in (a) and (c) were generated out of a large XZ-image-stack, partly shown in (b) and (d). As indicated in the XZ-images, the axial resolution is 820 nm and 190 nm for the two-photon confocal and the two-photon 4Pi confocal images, respectively. The images were taken with water immersion lenses (NA = 1.2). Reproduced from Bahlmann K, Jakobs S, *et al.* (2001) 4Pi-confocal microscopy of live cells. *Ultramicroscopy* 87(3): 155–164, with permission from Elsevier.

$\alpha = 68°$ (NA = 1.4) and therefore are expected to produce higher side lobes, which are reduced by 35% with careful aberration studies such that linear filtering is possible. Nonetheless, the use of water-immersion lenses solves two major problems. First, they extend the possible sample thickness to the working distance of the objective lens (~200 μm); second they allow the imaging of live cells in aqueous medium. As a result, microtubule networks (see Figure 2) and live bacteria can be imaged with a resolution of 280 nm in lateral and 190 nm in axial direction. The water immersion 4Pi microscope, utilizing two-photon excitation, is capable of the highest 3D resolution so far reported in live cell imaging to date.

Multifocal Multi-Photon Microscopy (MMM)

Improving imaging speed is critical for applying 3D microscopy in biological specimens. An MMM scans an object with an array of high-aperture foci whereas a conventional multi-photon microscope scans with a single beam. An MMM combines nonlinear fluorescence excitation with parallelized 3D imaging. The technical design consists of a densely packed microlens array; each microlens transmitting a collimated beam at a slightly different angle, overfilling the back aperture of the objective. Consequently, from the arrangement of the microlenses, an array of foci is reproduced in the image plane. The foci are scanned rapidly across the image plane to ensure full coverage of the focal plane and real-time imaging (a stack of 50 images of 30–70 μm diameter can be acquired within a few seconds). The foci's movements are either generated by rotating a microlens array that is arranged in a Lasserie pattern or by rectilinear scanning of a square microlens matrix.

MMM development was initiated because nonlinear optical damaging processes limit the usable focal peak intensity to about 200 GW/cm^2, which can be reached by a tightly focused 7–15 mW femtosecond laser beam. This power consumption makes use of only a fraction of the laser output of a typical mode-locked titanium-sapphire laser light source. The MMM better exploits the 1–2 W of the beam power by dividing the power among 25–40 foci and scanning each in parallel. With this method, typical acquisition times are cut from ~1000 ms to 10–50 ms. MMM provides an excellent method for real time 3D multi-photon imaging.

A fundamental problem in parallelized 3D microscopy is that increasing density of the foci over a certain level decreases the axial resolution. The reduction of the distance between the foci increases not only the speed or image brightness, but it also increases the interference between neighboring focal fields; especially in planes shortly before and after the focal plane. For neighboring foci in high NA objective lenses, interference effects can be significant. The relatively short laser pulses that are commonly used in multi-photon excitation provide a means to reduce this interference by using time multiplexing. Experimentally, the femtoseconds pulses of the Ti:sapphire laser, going through each lenslet, can be scheduled to arrive at the focal plane at different times by a mask of glass in front of each microlens, varying the optical path lengths of these beams. The different arrival

time eliminates interference effects. The resolution of the parallelized system remains unchanged as compared with the conventional one foci system. In conclusion, the MMM enables increased image acquisition speed by a factor of the number of foci in the optical plane without compromising image resolution.

Advanced Multi-Photon Microscopy Applications

As described by the previous section, MPM technology is undergoing rapidly development with improved resolution and increased instrument throughput. Recently, MPM has also found applications in exciting new areas of optical memory and microfabrication.

Multi-Photon 3D Optical Memory

Researchers in the field of optical memory are challenged to develop instrumentation that records three-dimensional, high density, robust bitwise information in low-cost and readily available storage materials. Initially, early optical memory techniques relied on higher energy pulsed recording beams and red shifted reading beams. Aided by the development of the Ti:sapphire laser and high numerical aperture objectives that can confine laser light to a volume of the order, λ^3, the power of the focused femtosecond pulses at the focal point can reach over 10^{18} W/cm^2 with electrical field strengths reaching 2.7×10^{10} V/cm. These fields are substantial enough to locally change the optical properties of selected materials. Thus, multi-photon processes were intertwined with the field of optical memory. Recent experiments showed that with properly chosen materials and instrumentation, high density, three-dimensional data storage was possible.

Optical memory is based on photochromic material whose optical properties are changed upon irradiation such that initial and final chemical states have different absorption spectra or indices of refraction. In 1989, Parthenopoulos and Rentzepis first showed that 3D information could be stored in a polymer film doped with photochromic molecules. Although the recorded information is lost when the photo-excited compound spontaneously reverts back to the original nonfluorescent compound, Dvornikov and Rentzepis later improved the material such that the stored information was stably recorded at room temperature.

For photochromic materials, the recording and reading mechanisms often required multiple laser sources and complex optical setups. Kawata *et al.* addressed these issues with the development of an alternative optical method that used photorefractive crystals as the storage media, a Ti:sapphire as the recording device, and a phase-contrast microscope configuration for data readout of bits separated by 20 μm. Using this method, Kawata avoided the complication of different recording and read-out laser sources, thereby simplifying the instrumentation. However, in photorefractive material, poor refractive index matching, between the storage medium and the objective, results in aberration and degrades storage densities. Recent advances in objective lens design have attempted to address these limitations. Kawata and Kawata describe many of these optical memory devices in a recent review. A recent publication by Olson *et al.* describes a method for multi-photon data storage in molecular glasses and highly cross-linked polymers. These materials prove to be highly efficient and robust for long-term data storage.

Although the field of optical memory has made great progress since 1989, further advances are required before this field becomes a widely adopted technology. For example, simpler recording and reading techniques are required to insure a cost-effective and reliable data storage technology. Consequently, further progress in the production of compact ultrafast lasers and high-speed scanning instrumentation will be needed before broader acceptance of optical memory storage devices.

Multi-Photon 3D Microfabrication

Another promising emerging application of multi-photon excitation is 3D microfabrication. 3D microfabrication processes can be performed with two-photon absorption. Maruo *et al.* and Sun *et al.* provide examples of microfabricated structures generated with femtosecond pulses to induce photopolymerization. Much like the optical memory techniques, this methodology relies on photoinitiators with large two-photon cross-sections to initiate polymerization of surrounding monomer or polymer molecules. Kawata *et al.* and Sun *et al.* showed that two-photon photopolymerization could be used successfully to create intricate three-dimensional microstructures. These systems offer lower polymerization thresholds, higher polymerization rates, and wider dynamic write ranges such that rapid fabrication and low-power writing are possible. Recently, Cumpston *et al.* described a class of π-conjugated compounds with large two-photon cross-sections, which can be effectively used for two-photon polymerization initiators in three-dimensional optical data storage and microfabrication processes.

As with most of these processes, the techniques have proven to be successful, but the mechanisms behind the multi-photon initiated chemical processes are not completely understood. With further work to optimize the reaction parameters, it is possible to design experiments that will allow new materials to be synthesized or patterned based on processes such as dielectric breakdown, microexplosions, optical damage, and pyrolytic effects.

Conclusion

Relying on femtoliter excitation volumes and large separations between excitation and emission wavelengths, multi-photon microscopy has emerged as a powerful technology that allows for increased resolution and enhanced signal-to-noise ratios. It has also enabled scientists to avoid complications in imaging biological systems due to decreased sample photodamage and in some instances, photobleaching of fluorophores. Advancements in multi-photon imaging applications, such as 4Pi Confocal microscopy and multifoci multi-photon microscopy, have further enhanced the resolution of imaging technology and improved the speed of data acquistion respectively. Furthermore, the application of multi-photon technology in the fields of microfabrication and optical memory promises to address some of the difficulties involved in designing three-dimensional high-density storage devices and microfabricated components. Overall, multi-photon microscopy provides the scientific community with an invaluable tool in the study and design of biological and physical systems.

See also

Microscopy: Confocal Microscopy; Imaging Multiple Photon Fluorescence Microscopy; Overview.

Further Reading

Bahlmann K, Jakobs S, *et al.* (2001) 4Pi-confocal microscopy of live cells. *Ultramicroscopy* 87(3): 155–164.

Bewersdorf J, Pick R, *et al.* (1998) Multifocal multiphoton microscopy. *Optical Letters* 23: 655.

Denk W, Strickler JH, *et al.* (1990) Two-photon laser scanning fluorescence microscopy. *Science* 248(4951): 73–76.

Kawata S and Kawata Y (2000) Three-dimensional optical data storage using photochromic materials. *Chem. Rev.* 100(5): 1777–1788.

Olson CE, Previte MJR, *et al.* (2002) Efficient and robust multiphoton data storage in molecular glasses and highly crosslinked polymers. *Nature Materials* 1(4): 225–228.

Pawley JB (ed.) (1995) *Handbook of Biological Confocal Microscopy.* New York: Plenum Press.

So PT, Dong CY, *et al.* (2000) Two-photon excitation fluorescence microscopy. *Annu. Rev. Biomed. Eng.* 2: 399–429.

Xu C and Webb WW (1997) Multiphoton excitation of molecular fluorophores and nonlinear laser microscopy. In: Lakowicz JR (ed.) *Topics in Fluorescence Spectroscopy.* New York: Plenum Press.

Xu C, Zipfel W, *et al.* (1996) Multiphoton fluorescence excitation: new spectral windows for biological nonlinear microscopy. *Proceedings of the National Academy of Science of the USA* 93(20): 10763–10768.

Interference Microscopy

E Novak, Veeco Instruments, Inc., Tucson, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Interference microscopy is a powerful noncontact metrology technique for reflection and transmission characterization of samples. The field can be divided roughly into two main categories, polarization-splitting interferometers and amplitude-splitting interferometers. In both cases, the light incident on a test sample is first divided into two beams and later recombined to form an interference pattern. This interference pattern is analyzed by one of a variety of methods to determine the difference in the optical paths the two beams traversed. This optical path difference may be used to quantify surface shape, refractive index variations, or thickness variations of the sample to a very high accuracy. This article describes the background of the two categories of interference microscopy, specifics of their implementations for different types of measurements, current state of the technologies, and examples of different measurements performed using these techniques.

Background

Interference fringes were first observed in a microscope by Robert Hooke in 1665. In the late 1800s, several practical versions of interference microscopes

entered use to examine transparent specimens, though none gained widespread commercial success. Then in the 1950s, differential interference contrast (DIC) microscopy was invented to increase image contrast when examining semi-transparent samples under a microscope, particularly biological specimens. As this technique does not require staining of samples, and avoids the halo effect common in phase contrast microscopy (*see* **Microscopy:** Phase Contrast Microscopy), such systems became widely used in a variety of biological applications. In the early 1960s, Krug, Rienitz and Shulz published a comprehensive book to summarize all of the various interference microscope designs that had been produced thus far, showing the large depth and breath of research to that point.

With the advent of affordable computers and CCD cameras in the 1970s and 1980s, various commercial systems were developed using a variety of interferometric configurations to quantify the measured interference patterns from small samples. Due to their noncontact nature and ability to provide area measurements, as opposed to the two-dimensional traces produced by contact stylus techniques, such systems gained rapid acceptance in a variety of industries. Performance has steadily improved such that vertical resolution is sub-nm and accuracy on the nm-level is readily achievable. Lateral resolution is limited only by the optics of the microscope, with high numerical aperture objectives achieving resolution below 0.5 μm. Today, commercial systems are used in such measurements as index variations of biological samples, roughness of paper, microfluid channel volume, MEMS properties, hard drive head dimensions, drug coating thickness, and machine part characterization.

Polarization-Division Interference Microscopes

While the concepts behind polarization division interference microscopes have been known for well over a century, they did not gain widespread use until after Georges Nomarski invented the DIC microscope, also referred to as a polarization interference contrast microscope, in 1955. A schematic of a transmission-based and reflection-based Nomarski setup is shown in Figure 1, where the displacement of the beams has been greatly exaggerated for clarity.

The illumination in Nomarski systems is generally from an unpolarized broadband source, such as a

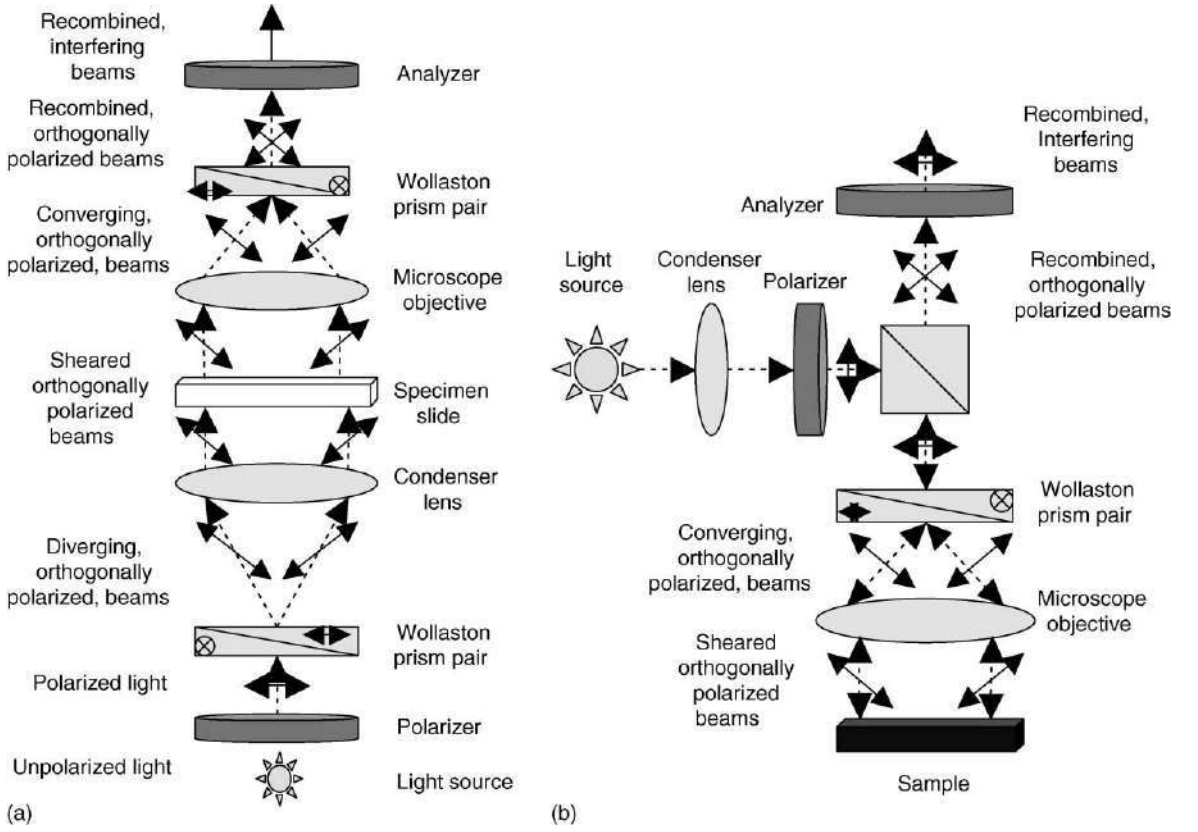

Figure 1 (a) Schematic of transmission-based Nomarski interferometer and (b) reflection-based Nomarski interferometer.

tungsten–halogen bulb. A polarizer in the illumination path is then used to set the polarization state of the light incident on a modified Wollaston prism pair. The modified Wollaston prism consists of two birefringent wedge prisms which divide the beam in two parts, each with orthogonal polarization states. The two beams travel in slightly different directions from one another. The condenser lens brings the light to illuminate the sample, with each of the two orthogonally polarized beams slightly sheared from one another. This shear is typically much less than the resolution limit of the microscope objective, typically from tens of nanometers to several hundred nanometers, such that only one image is seen by the observer. After passing through or reflecting from the sample and the imaging microscope objective, a Wollaston prism pair is used to recombine the two polarized beams. Finally, an analyzer is used to bring both beams to a common polarization state, allowing interference between the two beams.

As the system is very sensitive to the position and optical properties of the prisms, different Wollaston prism pairs are used for each objective, usually rotating into the path automatically on a turret as the objectives are changed. In a reflection system, this prism also recombines the displaced beams after they reflect from the test piece; a transmission system will usually have only one Wollaston prism pair after the sample that can handle all objectives, instead of requiring a different recombining prism for each objective.

The observed interference pattern is an image of the gradient in the sample in the direction of shear created by the Wollaston prism pairs. In order to fully characterize the surface, the sample or optics must be rotated to capture shear images in both orthogonal directions. Essentially the technique acts as a high-pass filter, emphasizing edges and lines on the test piece.

With a white-light source, dispersion effects will cause one side of a feature to appear brighter or in a specific color range whereas the other side appears darker or in different colors. This gradient effect confers a pseudo three-dimensional appearance to the specimen. The images are quite colorful, with different colors shearing differently; by adjusting the Nomarski prisms along the optical axis, one can vary the retardation between the two beams, producing more or less contrast or color variation in the observed images.

The Nomarski system has several appealing qualities for surface and index characterization over other interferometric and polarization-based techniques. First, it is a common-path interferometer; with the exception of the slight shear, both beams pass through exactly the same optics. This means that these interferometers are mostly insensitive to vibration, as both interfering beams will be affected in the same manner. In addition, the common path nature of the system reduces the quality requirements on the optics, improving measurement quality and reducing overall system cost. In addition, Nomarski techniques do not suffer the halos associated with phase contrast imaging techniques (*see* **Microscopy:** Phase Contrast Microscopy) and provide better out of focus rejection and higher light levels since they can use the entire aperture of the microscope objective.

For reflection-based systems, the observed fringes relate to the surface slope along the direction of shear; for transmission systems, index of refraction gradients or thickness variations are observed, producing much higher contrast images of biological samples than conventional bright-field techniques. The good out-of-focus rejection mentioned earlier makes possible more accurate optical sectioning of samples, where each vertical layer of the specimen is in focus while all others are out of focus. These features, combined with the fact that high contrast images can be obtained without chemical staining, have made these systems widely used in biology. In addition, the high sensitivity to slope has made DIC systems commonly used in semiconductor inspection applications and other precision surface characterization and defect inspection applications, where automatic image processing is combined with the highly sensitive measurements to quantify both surface and subsurface defects. For instance, Figure 2 shows a bright-field image of a laser diode facet as

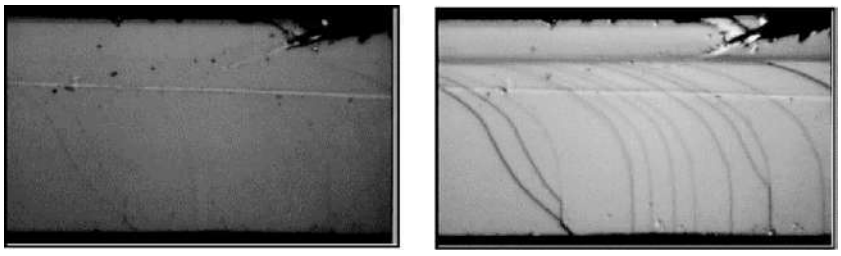

Figure 2 Bright field (left) and Nomarski (right) images of laser diode facet defects. Courtesy of Royce Instruments, Inc., Napa, CA.

well as a Nomarski image, with the facet defects greatly emphasized by the Nomarski optics.

Initial systems using differential interference contrast were strictly qualitative, with the user manually rotating the incoming polarization to produce the best contrast for a given sample and visually evaluating the resulting image. As mentioned above, visual inspection of biological specimens and defect review remain primary applications of these devices, though quantitative. By translating the Wollaston prisms laterally or rotating the analyzer in the system, one can phase shift the interference pattern. Storing multiple phase shifted images on a computer, one can obtain quantitative data about the shear using standard phase shifting interferometry equations. In addition to phase shifting through translation of the prism, quantitative data may be performed through analysis of the colors of the observed fringes or through geometric phase shifting techniques. More details on these methods can be found in the articles on **Interferometry:** Phase Measurement Interferometry; and **Interferometry:** White Light Interferometry.

Once quantitative information about the shear between the beams is obtained, the resulting data may be integrated to get back the original surface or index profile. Performing this operation on data with shear in two orthogonal directions allows the complete surface profile to be obtained. Unfortunately, since Nomarski setups are typically employed as add-on modules to standard microscope platforms rather than custom-designed metrology instrumentation, such quantitative phase shifting capability has not been widely commercialized; only a few custom or laboratory systems are in use when compared with white-light optical profilers described below. However, many research groups report nm-level surface characterization, and have developed a variety of techniques to accurately quantify thickness and index changes. The primary limitation of the quantitative techniques comes from the depth of focus of the microscope objective, which limits the vertical heights, thickness range, and index gradients that can be measured in a single measurement. Phase shifting at multiple vertical locations can be performed, but unless greatly increases the complexity of the measurement and reconstruction algorithms.

Amplitude-Division Interference Microscopes

White Light Optical Profilers

Probably the most common amplitude division interferometric microscopes are white-light interferometers, also often called optical profilers. These systems consist of three main elements: the base microscope, with illuminator and imaging optics; an interferometric element to divide and recombine the wavefront, and a scanning mechanism by which the optical path between the test and reference beams may be varied in a known manner. As the name implies, these systems typically employ spectrally broad sources, generally from a tungsten–halogen bulb. Unlike Nomarski interferometers, white light interferometers are used only in reflection-mode, as their noncommon-path nature and low-coherence source preclude interference from thick samples measured in transmission.

Figure 3 shows a schematic of a typical white light interferometer. Light from the illuminator arm, typically in a Kohler configuration, is coupled into the system with a first beamsplitter and passes through a second beamsplitter which is typically inside the microscope objective. This beamsplitter divides the beam into two parts, one of which travels to the sample and reflects from it, while the other reflects from a very high-quality reference surface. The reflected light is then recombined by the same beamsplitter and the resulting interfering light is imaged onto a camera. The static interferogram thus produced is usually by changing the optical path difference between the test and reference beams through translation of the objective with respect to the sample. The resulting signal is analyzed by one of a variety of methods as described below.

White light interferometers generally have two distinct modes of operation. In the first mode, generally referred to as phase shifting mode, a filter

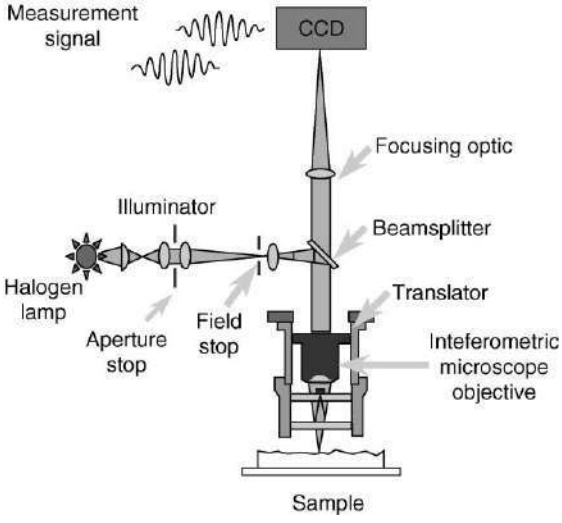

Figure 3 Schematic of white light optical profiler showing the illuminator, and translating interferometric objective.

is used to produce quasi-monochromatic light for the measurement process. This produces a large number of high-contrast interference fringes that obey the classical interference formula:

$$I_i(x, y) = I_t(x, y) + I_r(x, y) + 2\sqrt{I_t I_r}(\cos(\phi(x, y) + \alpha_i) \quad [1]$$

In the above $I_i(x, y)$ is the irradiance observed at any one point of the image, $I_t(x, y)$ and $I_r(x, y)$ are the irradiances of the test and object beams, $\phi(x, y)$ is the optical path difference between the two beams, and α_i is a phase offset introduced by the interferometric instrument. By varying α_i in a known manner and collecting at least three irradiance patterns, any of the three unknowns in the above equation can be determined; specifically, the optical path difference may be solved for every point. Assuming measurements are in a medium of known and constant refractive index, such that the only contributions to the OPD are height variations rather than index variations, surface height at each point is simply:

$$H(x, y) = \frac{\lambda}{4\pi}\phi(x, y) \quad [2]$$

Phase shifting techniques produce sub-nm repeatability and very high accuracy, but suffer from a 2π phase ambiguity from the calculation of phase from irradiance data. Thus, this technique cannot handle surfaces with point-to-point steps greater than one quarter of the wavelength of the light used. Calibration methods and further information may be found in the article on **Interferometry:** Phase Measurement Interferometry.

In order to extend interference microscopy to cover large steps, steep slopes, and vertical ranges greater than the depth-of-focus of the microscope objectives, a second technique of fringe analysis may be employed, that of coherence sensing. In this method an incoherent light source is used, such as a tungsten–halogen bulb. Due to the wide spectral range of the source, as interference patterns from different wavelengths add in the observation plane, the observed interference fringes will rapidly wash out for nonzero optical path difference between the two beams. See the article on **Interferometry:** White Light Interferometry, for further details and illustrations of this effect.

With the localization of fringes only at the point of zero optical path, the instrument scans the interference objective over the desired range and then calculates the position at which maximum fringe contrast occurs for each point. Calculating this position across all points within the field creates a height map of the surface; the interference fringes serve as a tool for depth discrimination in much the same way that the pinhole in a confocal microscope does (*see* **Microscopy:** Confocal Microscopy), another technique finding wide use in surface profiling. No phase unwrapping is required, with each pixel height calculated independently of all others, avoiding ambiguities associated with phase shifting methods. The vertical range is limited only by the scan range of the measurement and the working distance of the microscope objectives; current commercial instruments can achieve vertical ranges of over 8 mm. Figure 4 shows the fringes one would observe at a given scan location during measurement of a test piece with three steps. As the OPD between the test and reference arms varies, the fringes will

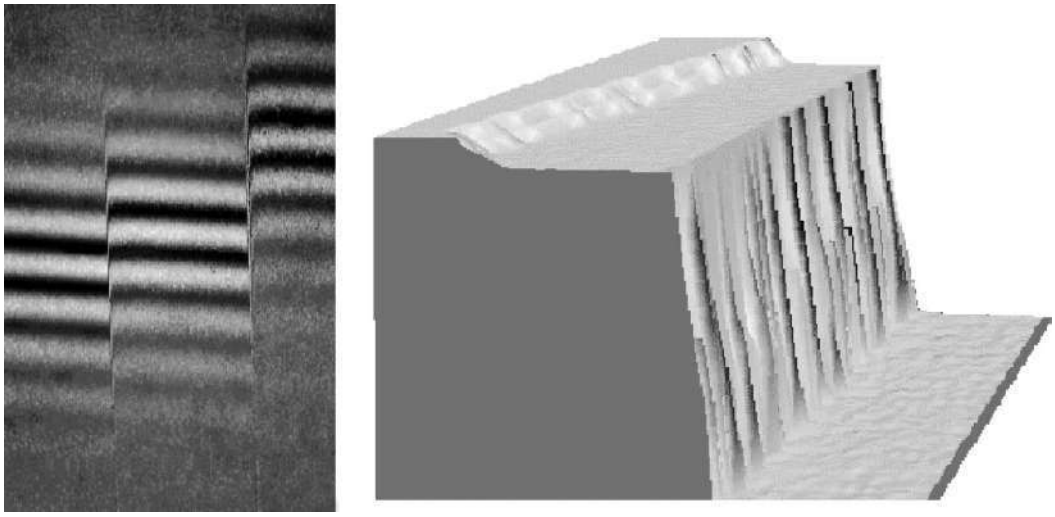

Figure 4 Localized fringes (left) of a 3-tiered surface structure (right). As the OPD is varied, the fringes will scan through the field, with the point of maximum contrast corresponding to the relative surface height of each pixel.

translate across the steps, and the surface plot at the right can be generated by calculating the scanner location of the peak contrast condition for each pixel.

There are a variety of methods for calculating the position of maximum fringe contrast using the coherence-sensing method. These include center-of-mass calculations of the peak fringe position, various frequency-space techniques and correlation of the measured fringe pattern with a theoretical pattern. Calibration of these systems generally involves repeated measurement of a known step height standard and varying the scanner motion until the step height is accurately calculated, though some systems employ separate distance-measuring interferometers to perform on-the-fly calibration for each scan. Generally speaking, coherence sensing techniques have repeatability in the range of a few nm, as opposed to the sub-nm repeatability of phase shifting techniques. This reduced sensitivity is the trade-off for avoiding the height and slope limitations of phase-shifting techniques. More details can again be found in the article on **Interferometry:** White Light Interferometry.

Recent advances in computing power have enabled an additional method of calculating surface height by combining coherence sensing with phase shifting. Coherence sensing methods are used to determine the point of best fringe contrast, and uses this information combined with the phase calculations to overcome ambiguities caused by large steps or steep slopes. Thus, one can obtain the sub-nm repeatability of phase-shifting methods without limitations on the vertical range or slope of the sample to be measured.

New fringe analysis techniques are extending the utility of these systems even further. Strobed sources are now available which effectively freeze the periodic motion for the instrument; this allows dynamic properties of samples to be measured, from surface deformation under motion to resonance frequency and switching time calculations. By analyzing the multiple interference patterns generated by transparent films deposited on a surface, film thickness, homogeneity, and differences between substrate and film surface properties may be determined.

Interferometric Microscope Objectives

One of the most critical elements in white light interferometers is the interferometric objective which splits and recombines the wavefront to form an interference pattern. Three primary interferometric configurations are used to divide and recombine the wavefront: Michelson, Mirau, and Linnik. Each of the three types of objectives is described below.

For low-magnification objectives, typically 5X or below, a Michelson interferometer configuration is typically used. A schematic of such an objective is shown in Figure 4. The illuminating beam passes through the objective and is divided by a beamsplitter placed just underneath the objective. A portion of the light travels to the test sample and the rest to a high-quality reference surface. The light reflected from each beam path is recombined by the beamsplitter and propagates to the observation plane. Note that it is important that the distance between the beamsplitter and reference surface match the distance between the beamsplitter and focus position of the sample. If the reference surface is defocused, fringe contrast will be reduced, aberrations increased, and overall system accuracy and repeatability will suffer.

This objective configuration has several advantages. First, one can add a Michelson housing to almost any standard objective with sufficient working distance for the beamsplitter, making these objectives easy to produce and relatively inexpensive. Second, the reference surface can be easily changed by the user to closely match the reflectivity of the test piece, ensuring maximum fringe contrast in the interfered beam (*see* **Interferometry:** Phase Measurement Interferometry). Third, provided there is distance between the beamsplitter and reference, one can insert a dispersion-compensating plate into the reference arm to allow measurement of parts beneath a cover glass or other dispersive medium; without such compensation, fringe contrast is generally insufficient for accurate measurements for transmissive media beyond a few hundred micrometers thickness.

As one moves towards higher magnifications, however, the working distance of the objective is typically insufficient to accommodate a large beamsplitter and still leave a reasonable distance between the objective and sample. In addition, a large cube beamsplitter will generally introduce unacceptably high aberrations in these higher numerical aperture objectives. For this reason, objectives from 10X magnification to 50X magnification are typically of the Mirau configuration, shown schematically in Figure 5.

Mirau objectives contain two additional plates compared to a standard bright-field objective: one is coated and acts as a beamsplitter, splitting the incident beam, while a second holds a small, high-quality reference surface and acts as a compensating plate so that optical paths in both arms of the interferometer are balanced (Figure 6). Such objectives are more expensive than Michelson designs, but maintain long working distances even at higher magnifications. Thermal stability becomes

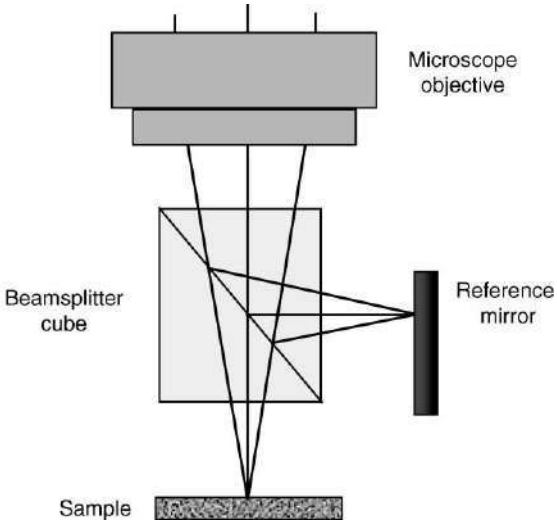

Figure 5 Schematic of a Michelson interference objective used for low-magnification applications.

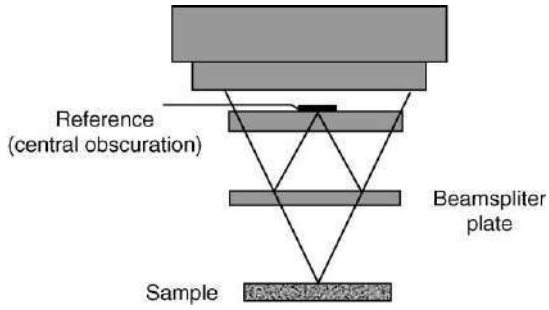

Figure 6 Schematic of a Mirau interference objective used for medium-magnification applications.

very important, as depth of focus is much reduced over lower-magnification objectives and the path from beamsplitter to reference still must be the same as from beamsplitter to in-focus test object, and most Mirau designs are therefore thermally compensated.

As one moves to even higher magnifications and numerical apertures, such as 75X or 100X, the microscope objectives do not have sufficient working distance to allow for any reasonable beamsplitter after the objective unless the numerical aperture is greatly reduced. To maintain high numerical aperture and high lateral resolution, a Linnik configuration must be used. A schematic of this configuration is shown in Figure 5. In this configuration, two separate microscope objectives are used after a beamsplitter, with one beam again reflecting from a high-quality reference and the other from the sample.

A Linnik objective is generally the most difficult to use and expensive interferometric configuration

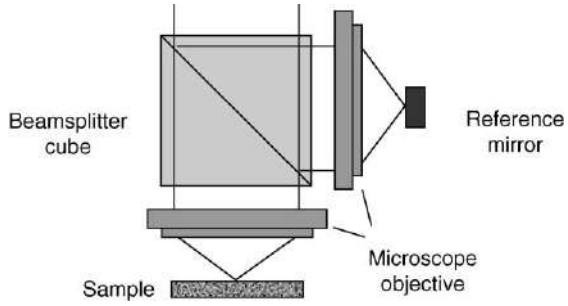

Figure 7 Schematic of a Linnik interference objective used for high-magnification applications.

(Figure 7). High numerical aperture typically limits working distance to a few hundred micrometers. In addition, depth of focus is very small, and thus these objectives must often actively adjust the reference mirror position to maintain focus over long periods of time and environmental conditions. Also, the dispersion and aberrations in the two objectives must be closely matched to ensure good fringe contrast, as there are many noncommon-path optical elements in this design; typically many objectives must be evaluated to find two that suitably match, another contributor to cost. However, these objectives do have the highest lateral resolution, and with an NA of up to 0.95, can measure the steepest slopes due to a large collection angle.

Applications

White-light interferometers are noncontact, and thus will not damage samples under test. They provide true three-dimensional surface characterization with nm-level accuracy and repeatability, and can be employed in production and laboratory environments. The various interferometric objectives allow for lateral resolutions and ranges to meet most needs; fields of view may even be stitched together to create surface maps whose extent is limited only by the range of the translation stage employed. In addition, each scan is typically quite fast, from under one second in phase-shifting mode to vertical scan rates of up to 100 μm/sec in coherence-sensing mode, with no sample preparation time. While the systems must be properly calibrated, vibration minimized, contrast maximized, and potential error sources such as dispersion and phase change on reflection effects must be managed, they remain one of the most accurate surface characterization tools available.

All of the above factors have enabled the use of white light interferometers in a variety of fields. The semiconductor and data storage industries

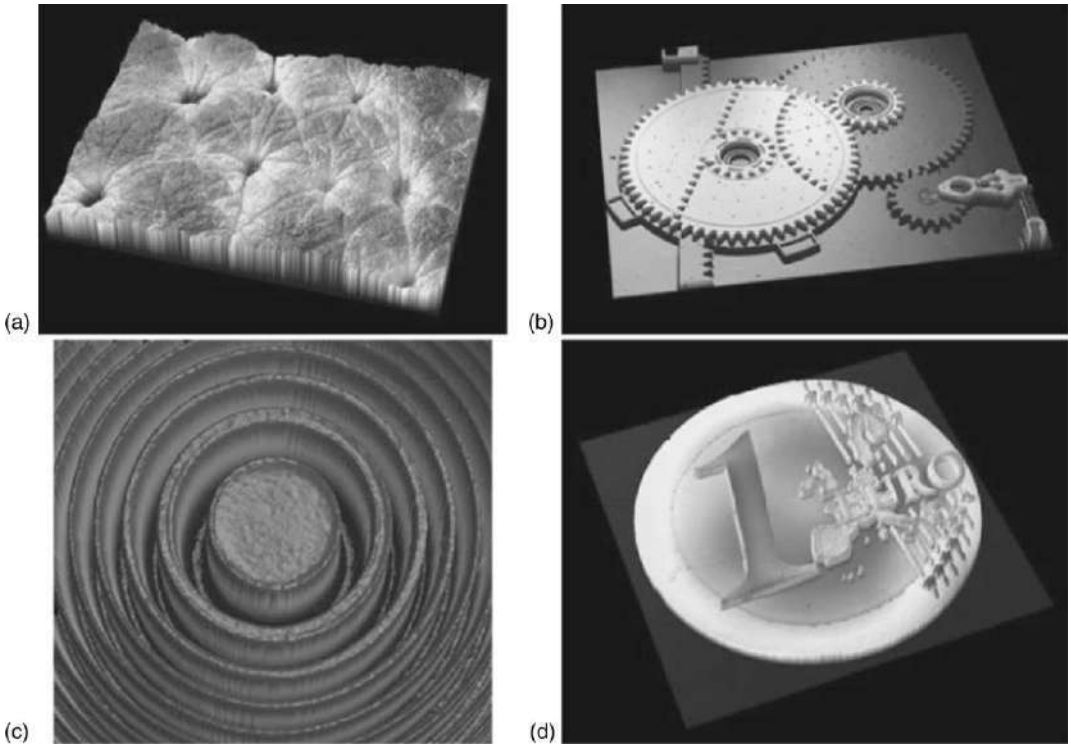

Figure 8 Surface plots from a white light optical profiler. (a) Human skin sample, (b) MEMS microgears, (c) binary optic, and (d) 1 Euro coin. Pictures courtesy of Veeco Instruments.

perform critical dimension and surface characterization of various components using these systems, some of which are configured for full wafer-handling capabilities. Systems are also used to measure precision machine parts, optics, and medical instrumentation and implants. The high sensitivity of the coherence sensing techniques allows surfaces with less than 0.1% reflectivity to be measured, so that characterization of paper, abrasives, and even different foods has been measured with these instruments. With the addition of the advanced technologies mentioned above, these systems are also now doing dynamic measurements of MEMS and thermal response characterization of precision surfaces, and also are measuring thicknesses of adhesives, drug coatings, photoresist, and a variety of other semi-transparent layers in various applications. Figure 8 shows several surface measurements, including human skin, MEMS microgears, a binary optic, and 1 Euro coin.

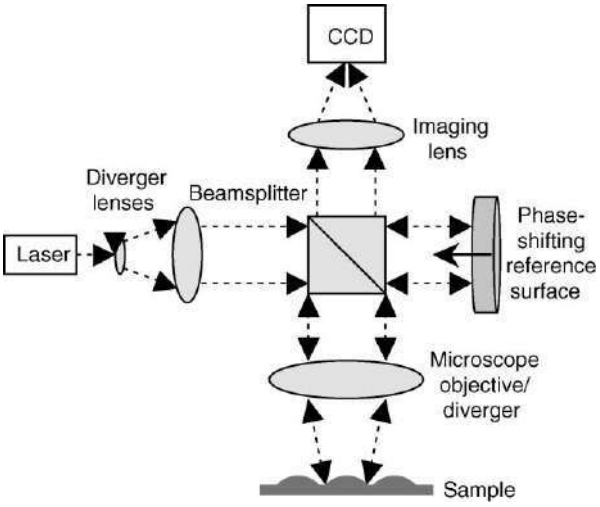

Figure 9 Schematic of a micro-LUPI interferometer operating in reflection mode.

Laser-based Interference Microscopes

Another, less common type of amplitude-division interference microscope has recently been commercialized, primarily for use in examining micro-optics. These laser-based systems employ a Tywman–Green interferometric configuration combined with a microscope, and are generally called micro-LUPI (laser unequal path interferometer) systems. A schematic of a micro-LUPI interferometer is shown in Figure 9. A beamsplitter divides the incident laser beam, half of which travels to a generally flat reference surface, while the other half typically travels through a

microscope objective that acts as a diverger lens and is incident on the test sample.

Due to the long-coherence length of the laser source, the two optical paths do not have to be equal in order to observe interference fringes, and samples may be measured either in reflection or transmission, depending on the desired test result. Collimated, diverging or converging illumination may be brought incident on the sample, depending on the optics in the test beam and their placement. The diverger elements are noncommon-path optics, and reduce the ultimate accuracy of the system to approximately lambda/10, similar to laser-based Fizeau interferometers.

Micro-LUPI-based systems have several advantages for testing small optics over other interference microscopes. First, the ability to measure both transmission and reflection properties of the optics, as well as different incident illumination schemes, means they can test the optics in their use configuration rather than inferring optical properties based solely on surface measurements. They can determine imaging properties, inhomogeneities and thickness variations in addition to characterizing the surfaces under test. Also, the curved test wavefront better matches spherical test surfaces, minimizing the number of observed fringes and making testing of small aspheres easier. Lastly, these systems can have curved reference surfaces as well as flat, and with the proper configuration aberrations in measurements due to the interferometric microscope can be minimized, ensuring high-quality measurements of the optic under test.

These systems have only phase-shifting measurement capability, as the coherence length of the source is too long for white-light interferometric techniques. Transmission tests necessitate a more complex part handling structure, with both the test piece and reflective return optic needing adjustment for tip/tilt and vertical and lateral translation. Therefore, such systems are more costly and not as versatile as a white light optical profiler, but do provide a valuable capability to the micro-optics community.

See also

Coherence: Coherence and Imaging. **Imaging:** Interferometric Imaging. **Interferometry:** Overview; Phase Measurement Interferometry; White Light Interferometry. **Microscopy:** Confocal Microscopy; Overview; Phase Contrast Microscopy.

Further Reading

Advanced Light Microscopy: Measuring Techniques; *Maksymilian Pluta*; North-Holland (1993).

Born M and Wolf E (eds) (1999) *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th edn. Cambridge University Press.

Creath K (1988) Phase-measurement interferometry techniques. In: Wolf E (ed.) *Progress in Optics*, vol. XXVI, pp. 349–393. North-Holland, Amsterdam.

Hariharan P (2003) *Optical Interferometry*, 2nd edn. Academic Press.

Hariharan P and Malacara D (eds) (1995). Selected Papers on Interference, Interferometry, and Interferometric Metrology. *SPIE Milestone*. MS 110.

Krug W, Rienitz J and Shulz G (1964) *Contributions to Interference Microscopy* (translated from the German edition of 1961 by J. Home Dickson): Hilger and Watts Ltd.

Malacara D (ed.) (1992) *Optical Shop Testing*. John Wiley & Sons.

Massie NA (ed.) (1988). Interferometric Metrology. *SPIE* 816.

Polarvarapu PL (ed.) (1997) *Principles and Applications of Polarization-Division Interferometry*. John Wiley & Sons.

Tolansky S (1970) *Multiple-beam Interferometry of Surfaces and Films*. Dover Publications.

Nonlinear Microscopy

S Lévêque-Fort and P Georges, Université Paris Sud, Orsay, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Nonlinear microscopy has become an important tool for the investigation of biological applications where high resolution and three-dimensional imaging are essential for understanding the underlying biological function. Two-photon excitation fluorescence (TPEF) microscopy has opened the way to nonlinear techniques and is now widely available. As this new technology enables noninvasive study of biological specimens with high resolution, the number of applications in biology and medicine is still increasing. By exploiting autofluorescence of cells or tissues components, or by coupling with genetically modified fluorescent indicators or small extrinsic probes, high

contrast and localization can be obtained. Different detection techniques (fluorescence lifetime imaging (FLIM), fluorescence correlation spectroscope (FCS), intensity, and spectral imaging) permits the analyze of several fluorescence properties and so obtains complementary informations.

The successful development of two-photon techniques has generated substantial interest in the scientific imaging community in further developing and exploiting optical nonlinearities for generating new image contrast. All these new developments have strongly benefited from the major advances in ultrafast lasers which have become more compact and reliable. Second- and third-harmonic generation (SHG, THG) as well as coherent anti-Stokes Raman scattering microscopy (CARS) have thus benefited from the fact that it is now easier to generate ultrashort pulses at almost any desired wavelengths. Association of these new nonlinear techniques in the same instrument (TPEF-SHG/TPEF-CARS, TPEF/THG) is commonly encountered and provides a powerful instrumental tool for probing biological samples.

Basic Principles of Two-Photon Microscopy (TPM)

Two photon microscopy is preferred over single photon method in many biological applications due to several intrinsic advantages which arise from having a nonlinear intensity depend absorption. In general, the nonlinear polarization for a material can be expressed as

$$P = \chi^{(1)}E^1 + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \cdots \qquad [1]$$

where P is induced polarization, $\chi(n)$ is the nth order nonlinear susceptibility, and E the electric field vector of the incident light. The first term describes normal absorption and reflection of light; the second term describes SHG, hyper-Rayleigh scattering, sum and difference frequency generation, and the third terms covers multiphoton absorption, third-harmonic generation and stimulated Raman processes.

In 1931 Maria Goppert Mayer predicted that photons of lesser energy can together cause an excitation 'usually' produced by the absorption of a single photon of higher energy, in a process call multiphoton excitation. The hypothesis could not be confirmed until the invention of pulsed ruby lasers, in the 1970s, and was first applied to microscopy with the work of Denk *et al*. Two-photon microscopy used the simplest version of her theory: two photons of equal energy (from the same laser) can excite a molecule, and the excitation is similar to that obtained with a single photon possessing twice the energy (Figure 1). Thus, the molecule can emit a fluorescent photon just as if it was excited with a single higher energy photon (Figures 1 and 2). Of course, the absorption of the two photons must occurs nearly simultaneously ($\sim 10^{-16}$ s), resulting in a quadratic dependence on the light intensity rather than the linear dependence of conventional one photon fluorescence.

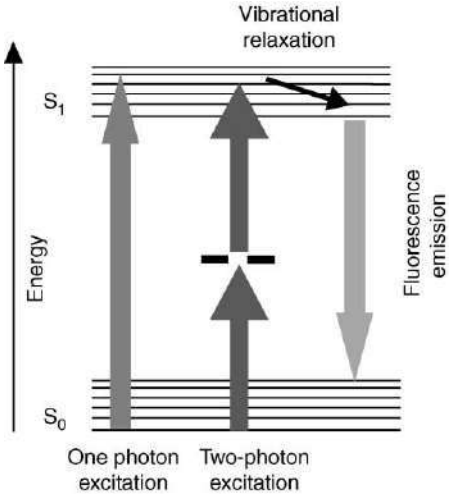

Figure 1 Jablonski diagram for one and two-photon excitation.

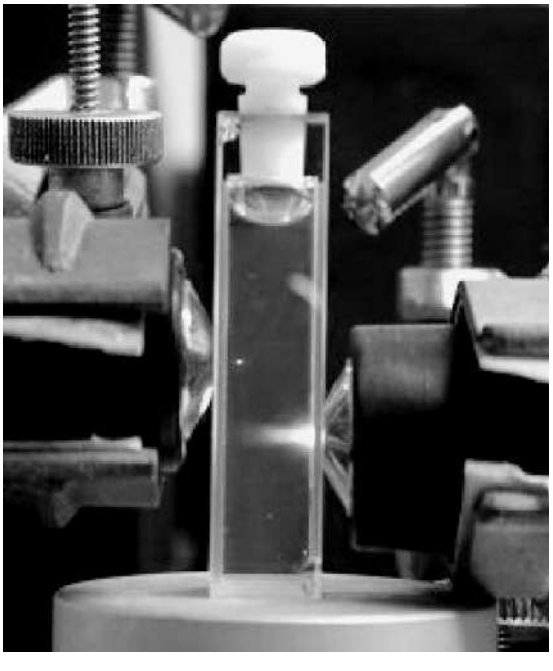

Figure 2 Localization of excitation under two photon excitation (left) compared to one photon excitation (right).

The nonlinear optical absorption property of two-photon excitation, limits the fluorochrome excitation to the point of focus. This localized excitation volume results in greatly improved axial resolution and image contrast compared with conventional microscopy, but also significantly decreases sample photobleaching and photodamage by reducing the region of photointeraction (Figure 2).

Another major advantage of TPEF is its ability to image deeper inside biological samples, owing to the reduced scattering and absorption of near infrared wavelength compared to UV and visible light. Near infrared light (700–1000 nm) is located within the 'therapeutic optical window' where the absorption of the main cells and tissue components is low.

TPEF also permits a high rejection of residual excitation light: in standard one-photon microscopy, the excitation wavelength can be spectrally close to the fluorescence emission band. Filters used to eliminate the leakage of excitation light into the detection channel often cuts off a part of the fluorescence emission. For TPEF excitation and emission band are well separated, and highly efficient filters can remove the residual excitation light without compromising the microscope sensitivity.

Under two-photon excitation, the fluorescence photon flux depends on the square of incident photon flux:

$$F \propto \sigma_{2P} C I^2 \qquad [2]$$

where σ_{2P} is the two-photon cross-section (cm^4 s) with 10^{-50} cm^4 s, called a Goppert–Mayer (GM), C the fluorophore concentration, and I^2 the incident photon flux. The two-photon cross-section is a quantitative measure of the probability of two-photon absorption and as a second-order process is low. As it is difficult to directly measure two-photon cross-sections, we usually use the cross-section of two photon-induced fluorescence (σ_{TPEF}) as the product of two photon absorption cross-section (σ_{2P}) with the fluorescence quantum yield $\phi_F \sigma_{TPEF} = \sigma_{2P}\phi_F$.

Intrinsic fluorophore, such as nicotinamide adenine dinucleotide (NADH), have low cross-sections ($<10^{-4}$ GM), usual fluorescent dye in biology ranging between 1 and 300 GM and quantum dots approaching 50 000 GM. As the probability of two-photon absorption is quiet low, high incident photon flux needs to be delivered to the sample to generate efficient absorption:

$$I \propto E \frac{1}{\Delta\tau\omega^2} \qquad [3]$$

I depends on the energy per impulsion E (mean power/pulse frequency), the pulse duration $\Delta\tau$, and the beam waist ω. The expression of I clearly demonstrates that by using ultrashort pulses and by reducing the beam waist, one can maximize the incident flux. Currently the most commonly used laser source for multiphoton microscopy is the femtosecond titanium-sapphire (Ti:Sa) system. These pulsed femtosecond lasers deliver a 100 fs pulse train at a repetition rate of ~80 MHz with a tuning range from 700 to 1000 nm. The beam waist depends on the objective numerical aperture. If we consider a Gaussian beam profile, with radial and axial parameters of σ_r and σ_z respectively and for a high numerical aperture NA > 0.8:

$$V = \left(\frac{\pi}{2}\right)^{3/2} \sigma_r^2 \sigma_z \qquad [4]$$

with $\sigma_r \approx 0.52\lambda/\sin\alpha$ and $\sigma_z \approx 0.76\lambda/(1-\cos\alpha)$. where λ is the excitation wavelength, α is the one-half of the objective angular aperture (NA $= n \sin\alpha$, n index of the immersion medium). So for a wavelength of 780 nm and an objective working with water immersion with a numerical aperture of 0.9, the excitation volume can be described as an obloidal shape with a radial axis of typically 500 nm and an axial axis of 1.90 μm. It corresponds to an excitation volume of ≈1 fL ou 1 μm^3. Higher resolution can be achieved using an immersion oil objective (NA = 1.3) at the cost of a higher cost and a shorter working distance.

Typical powers used to generate two-photon excitation without damaging the sample are between 1 and 10 mW for the mean incident power. For a laser beam of 1 mW, at a rate of 76 MHz with 100 fs pulse duration, and with an objective leading to a surface in the focal plane of 1 μm^2, the peak power is 0.13×10^9 W/cm^2 at the focal plane of the sample. Of course each samples has its own limits regarding incident intensity, and the key to use TPEF safely is always to reduce the excitation intensity in order to avoid photodamage.

Fluorescent Probes

Biological systems often possess endogenous fluorescent molecules that can be imaged, revealing important sample characteristics without the need of labeling (Figure 3a,b). Cells or tissues can be imaged through their autofluorescent, with the emission of several components, NADH, flavins, collagen fibers, etc. Pharmacology studies also exploit the natural autofluorescence of aromatic drugs, to

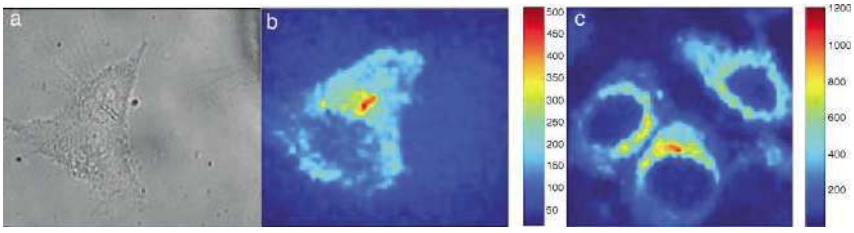

Figure 3 Applications showing various capabilities of TPEF: (a) Hela cells in transmitted image and corresponding autofluorescence image (b) under TPE (780 nm). (c) After adding 40 μM of an anti-HIV drug on similar cells, intrinsic fluorescence of the drug allows us to find its preferable localization inside the cytoplasm.

localize drug inside cells and study their reactivity without any labeling.

However, for some specific applications where the autofluorescence is too low, or if one needs to specifically identify different cells components (Figure 3g), or follow protein mobility, the samples need to be labeled your with extrinsic fluorophores. Extrinsic fluorophores are organic molecules, first developed for one-photon excitation, and are used under two-photon excitation with more or less efficiency as OPE and TPEF have different quantum mechanical selection rules, and the TPE spectrum is not always scaled to the half of OPE spectrum. Current works are focused on the development of specific markers for TPE or SHG to enhance contrast and also reduce incident laser power.

Two new generations of marker have been developed: the development of the green fluorescent protein (GFP) and its mutant introduces a convenient method for monitoring gene expression in cells and tissues; quantum dots are the latest focus in fluorescence labeling. They are promising reporters as their size (1–100 nm) induces a specific narrow emission spectrum in the visible. Absorption can occur over a wide range of short wavelengths because of the existence of multiple electronic states. Time-domain measurement has been performed with CdSe quantum dots which shows multiexponential decay with fluorescence lifetime of 3.4 ns, 16.1 ns, and 35.6 ns, much longer than autofluorescence lifetimes, hence providing greater sensitivity than intensity-based measurement. However, like GFPs, quantum dots display blinking behavior.

Two-photon Applications

Two-photon excitation has become the technique of choice for fluorescence microscopy and is applied to a variety of imaging tasks on samples ranging from single molecules to thick tissues. Owing to the high NIR penetration depth, biopsies and living tissue can be noninvasively investigated by morphological and functional fluorescence imaging of endogenous (Figure 3b) or exogenous fluorophores to depths of more than 300 μm. In genomics, multiphoton multicolour fluorescence *in situ* hybridization (FISH) are used to label various specific regions of DNA, and identified different genes and chromosome regions. In pharmacology, a higher penetration and localization of photodynamic therapy is observed, detection of low drug concentration inside cells facilitates the reactivity comprehension (Figure 3b,c). Neurobiologists investigate calcium dynamics inside brain slices and live animals, or follow neurotransmitter diffusion. Of course, these are only a few examples among the increasing number of applications, and some of them are illustrated in Figure 3.

Two-photon Instrumentation

Two-photon microscopes are now commercially available, however a large number of groups have built their own from separate components or by modifying an existing confocal microscope. Homemade microscopes present a higher configuration flexibility for excitation/detection at a much lower price.

Femtosecond, picosecond, and continuous wave (CW) laser sources have been used for TPM. By using a cw laser sources such as ArKr and NdYag, the average power increases by ~200-fold to achieve the same excitation rate obtained with a femtosecond laser, which is not highly desirable for biological samples. So as previously shown, the TPE efficiency is higher with shorter pulses, so Ti:Sa femtosecond lasers are most commonly used.

One of the drawbacks of TPM is that it is a laser scanning technique. To reconstruct an image you need either to move the sample or the laser beam. Most commonly $x-y$ mirrors driven by a galvanometer scanner are used to scan the laser at the focal plane of the sample. Lenses are used to expand the beam to overfill the objective back aperture and so obtain a diffraction limited focal volume. The excitation light enters the modified inverted epi-fluorescence microscope (Figure 4).

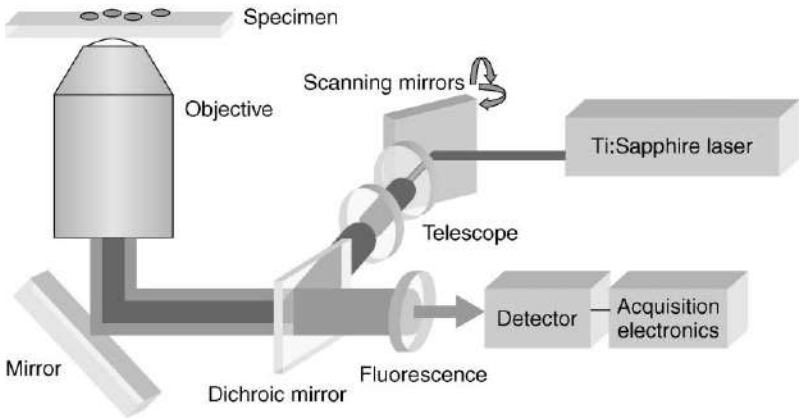

Figure 4 Schematic of a typical two-photon microscope.

A dichroic mirror transmits the excitation light and reflects fluorescence. High numerical aperture objectives are used to provide a high resolution but also maximize fluorescence collection. They are usually mounted on piezo-electric translators to provide a fast Z-scan. The fluorescence is collected by the same objective, reflected by the dichroic to the detector. A filter is placed in front of the detector to remove any residual excitation light. Typical detectors used are the photomultiplier tube (PMT), microchannel plate-PMT (MCP-PMT), avalanche photodiode or charge-coupled device (CCD) cameras. PMTs are the most commonly implemented because they are robust, low-cost, have large, active areas, and good sensitivity.

As TPE is intrinsically confined to the focal volume, two-photon microscopy unlike confocal microscopy, does not need emission pinholes and descanning optics to achieve axial depth discrimination Some authors have, however, added a pinhole to enhance resolution but at the cost of important sign al loss. Concerning the design of the detection pathway, depending on the size of the sample scanned, the objective magnification and the detector area, one may need to use a lens to collect the fluorescence onto the detector.

Fluorescence Detection

Contrast in fluorescence imaging can be provided through several complementary parameters such as steady-state intensity, spectral dependence (at absorption and emission), or fluorescence lifetime.

Intensity Imaging

Intensity images are obtained by summing all the incoming photons during an acquisition time (typically $\sim$1 to 10 ms) which strongly depends on the fluorescence quantum yield of the sample. The whole samples intensity image is obtain by scanning the beam on the focal plane with the $X-Y$ scanner, leading to an acquisition time of 25 s for a 50×50 μm image (1 μm step).

Spectral Imaging

Spectral imaging can be a powerful way to discriminate different molecules, and to identify cellular components or perform co-localization of different proteins. It can be performed by either using an interference filter in front of the detector to extract the contribution in a specific band, or more precisely with a spectrometer. However, autofluorescent spectral bands are usually broad and sometimes it is difficult to separate the different emissions. Extrinsic trackers are thus used to labels specifically, for example, the nucleus, the plasma membrane, and mitochondria, and each tracker usually emits in a specific and well-known spectral band, allowing to easily discriminate and localize the different cellular components. Quantum dots offer narrow emission spectra and allow high spatial resolution, and with the rapid development of their functionalization, they should lead to rapid progress in intracellular imaging.

Fluorescence Lifetime Imaging (FLIM)

Imaging of fluorescence lifetimes is a useful contrast-enhancing method that provides information related to the microenvironment with spatial resolution. The fluorescence lifetime τ corresponds to the average time a fluorophore stays in its excited state is given by $\tau = 1/(k_f + k_{nr})$ with k_f the radiative decay and k_{nr} the nonradiative decay rate.

The radiative decay rate is a characteristic property of a molecule and therefore may be used to contrast different chemical species or fluorophores.

The nonradiative decay rate is a function of the local environment of an excited molecule, and so fluorescence lifetime measurements can provide information about environmental changes (e.g., pH, pO_2, $[Ca^{2+}]$). Furthermore, different cellular probes or tissue components with similar emission spectra but different lifetimes can be distinguished. Another advantage of lifetime measurement is that as a relative measurement of the fluorescence intensity, it does not depend on a non-uniform intensity excitation of the sample, and also does not depend on molecular concentration, allowing one to detect very low quantity of molecules.

Following a delta excitation, the fluorescence intensity de-excitation is given by$I(t) = I_0\, e^{(-t/\tau)}$, with I_0 the initial amplitude. Fluorescence lifetime measurement has been demonstrated either in frequency or time-domains. Time-resolved measurements in the frequency domain are based on the determination of demodulation M and phase shift $\Delta\varphi$ and require modulated light sources (modulation frequency $\omega = 2\pi f$).

The excitation source $I_E(t)$ and induced fluorescence $I_F(t)$ can be written as:

$$\begin{aligned} I_E(t) &= I_{E0} + I_{E\omega}\cos(\omega t + \varphi_E) \\ I_F(t) &= I_{F0} + I_{F\omega}\cos(\omega t + \varphi_F) \end{aligned} \qquad [5]$$

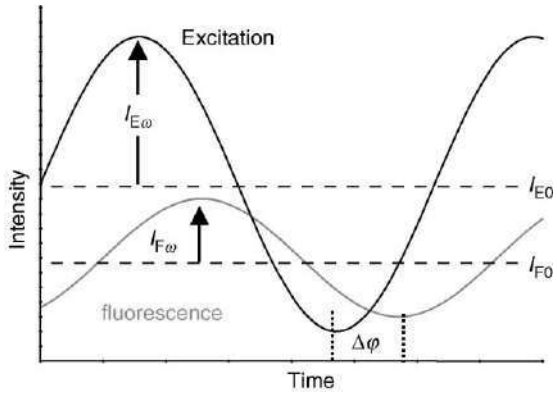

Figure 5 Time course of excitation and emission light in frequency domain fluorescence lifetime measurements techniques.

In the case of fluorescence with mono-exponential decay, the lifetime τ can deduced from the relations (Figure 5):

$$M = \frac{I_{F\omega}/I_{F0}}{I_{E\omega}/I_{E0}}. \qquad \omega\tau = \tan(\Delta\varphi) = \sqrt{\frac{1}{M^2} - 1} \qquad [6]$$

In order to translate the phase and demodulation information from the 80 MHz frequency to a lower kHz frequency, a modulated gain $G(t)$ at the frequency ω' can be applied to the detector (heterodyne detection):

$$G(t) = G_0 + G_{\omega'}\cos(\omega' t + \varphi_G) \qquad [7]$$

The detector response $S(t)$ is thus proportional to the fluorescence intensity:

$$\begin{aligned} S(t) = {} & G_0 I_{F0} + \frac{G\omega' I_{F\omega}}{2}\frac{1}{\sqrt{1+(\omega\tau)^2}} \\ & \times\cos[(\omega - \omega')t + \varphi_G - \varphi_E + \Delta\varphi] \end{aligned} \qquad [8]$$

In the time-domain, the single point detection system is called time-correlated single-photon counting (TCSPC). The principle of the TSCPC system is shown in Figure 6. For each excitation pulse a trigger signal is sent to a time-to-amplitude converter (TAC) within which a capacitor starts charging linearly with time. Upon arrival of each fluorescence photon on the detector a second trigger pulse stops the capacitor charging. The charge is proportional to the delay between the excitation pulse and the fluorescence photon. By repeating the measurement over many thousands of excitation pulse, a histogram of photon arrival can be build up. As each photon must be record individually, the incident power is reduce to avoid the detection of only photons from the beginning of the decay (pile-up) which would lead to shorter lifetimes. To match detector dead-time and TAC counting rate, the TAC is oftenly used in reverse mode: the fluorescence photon starts the TAC and the following pulse stops it.

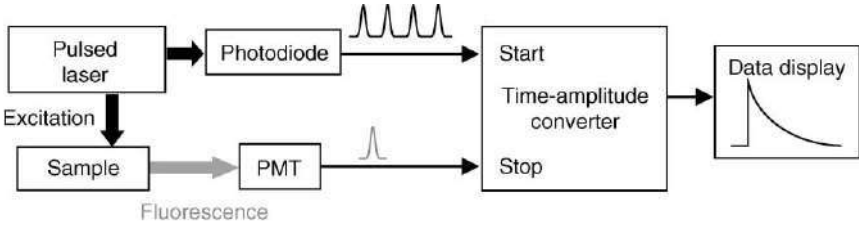

Figure 6 Schematic of time correlated single photon counting system.

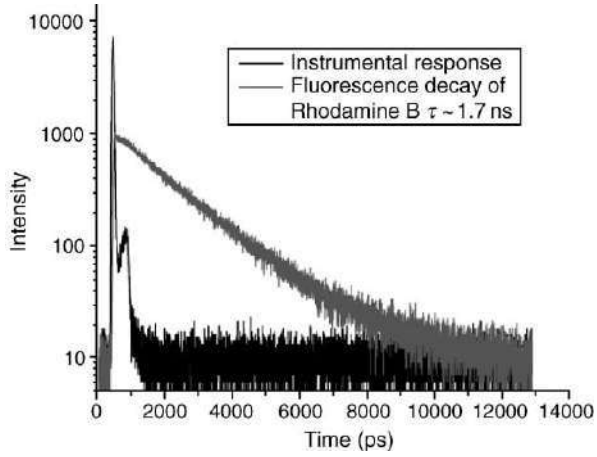

Figure 7 Typical instrumental response of MCP-PMT (black curve), and fluorescence decay curve of rhodamine B (grey curve) with $\tau \sim 1.7$ ns.

Typical detectors used are the photomultiplier tube (PMT), microchannel plate-PMT (MCP-PMT), and avalanche photodiode.

Typical instrumental response of our MCP-PMT is represented in Figure 7 (black curve) leading to a resolution of 30 ps. Depending on the fluorescence quantum yield of the sample, in order to have a good signal to noise ration, acquisition time is ranging between 1 and 30 s per point. Fluorescence decay of rhodamine B with a lifetime $\tau \sim 1.7$ ns is also presented in Figure 7.

For both detection approaches, single point lifetime measurements can be extended to lifetime imaging by synchronizing the sample scanning and the lifetime detection. However, due to the long time acquisition needed, images are only limited to small interesting areas rather than the whole sample. We will see in the recent development section below, new ways to overcome this issue.

Fluorescence Correlation Spectroscopy (FCS)

FCS experiments are performed by recording fluorescence intensity fluctuations (i.e., fluorescence bursts (Figure 8)) from the small focal volume (<1 fl). Applications include diffusion, chemical reactions, molecule concentration, or hydrodynamic flow–FCS under TPE benefits from all the advantages of TPE versus OPE.

Various experimental parameters can be extracted through temporal analysis of fluorescence intensity fluctuations by calculating the autocorrelation or cross-correlation of the fluorescence signal defined as:

$$g(\tau) = \frac{\langle I_F(t) I_F(t+\tau)\rangle}{\langle I_F^2(t)\rangle} = \frac{\int I_F(t) I_F(t+\tau)\mathrm{d}t}{\int I_F^2(t)\mathrm{d}t} \qquad [9]$$

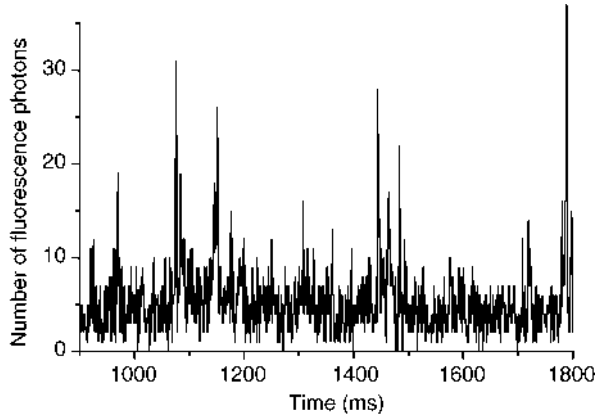

Figure 8 Fluorescence intensity raw data along the time. Bursts are observed when molecules cross the focal volume.

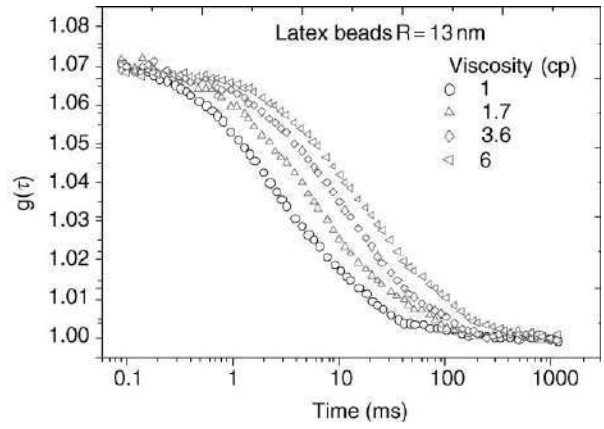

Figure 9 Correlation function observed for microspheres placed in different viscous solution.

where I_F is the time-dependent fluorescence signal. By considering that the fluorescence intensity fluctuates around a mean value:

$$I_F(t) = \langle I_F\rangle + \delta I_F(t) \qquad [10]$$

the autocorrelation function can be written as:

$$g(\tau) = 1 + \frac{\langle \delta I_F(t)\delta I_F(t+\tau)\rangle}{\langle I_F^2\rangle} \qquad [11]$$

FCS have been extensively used to extract diffusion coefficient of small fluorescent latex bead in solution, showing the influence of the bead radius or solvent properties. Figure 9 clearly exhibits the influence of the viscosity of the solvent on the diffusion of microspheres. Single molecule detection has also been reached. FCS as also been extended to biological applications where the low background associated with TPE and the reduced autofluorescence is critical. Diffusion of microbeads inside cells cytoplasm

or bacterial biofilms have been successfully demonstrated.

Recent Developments in Two-Photon Microscopy

Towards video microscopy

All the nonlinear microscopies share an important limitation: the image build up pixel-by-pixel by scanning across the sample, leading to long-time acquisition. Depending on the sample's emission and the area studied, scanning rates between 0.5 Hz and 30 Hz can be reached for intensity images. To speed up this acquisition time, a first approach is based on a line excitation, in combination with widefield CCD detection. This provides real-time image acquisition rates at the expense of a reduced axial resolution (5 μm instead of 1 μm). A more ideal way of spreading the available excitation power over the sample and speed up image buildup, is the simultaneous use of several well separated foci, each of them generating a diffraction limited spot. In combination with whole field detector, as with a CCD camera, image acquisition rates can be strongly enhanced without compromising resolution. The spacing between the foci must be carefully chosen in order to avoid out-of-focus interference. By introducing time-delay between the different foci, larger than the pulse duration, each foci of a microlens array can be position very closed without loss of resolution. An alternative way has been proposed; a special beamsplitter has been developed to create multiple beams with a higher efficiency, furthermore this setup inherently induced time-delay between each foci. Time-domain has been demonstrated in both approach, and FLIM maps on cells imaging have been obtained in only few second with 200 ps time resolution.

Sharpen the focal point

Nonlinear microscopy facilitates 3D-imaging but the resolution is still limited to 180 nm in the focal plane and 500–800 nm along the optical axis. According to the diffraction theory, the resolution of a focusing light microscope is related to the size of its focal point. In the mid-1990s, new concepts appeared to break this diffraction theory and sharpen the focal volume. In 4-pi microscopy the sample is placed between two large NA objective lenses in a single axis at their common focus. Coherent interference of beams passing through the two lenses generate a point spread function whose central maximum is 3–4 times narrower than in the normal confocal case, but which also has side lobes. This last issue can be overcome using two-photon excitation. A two-photon 4-pi microscopy can thus result in a resolution of ~100 nm in all directions, and has been demonstrated successfully on filamentous actin and immunofluorescently labeled microtubules. However, a 4-pi microscopy requires the sample to be thin and mounted between two coverslips, unless one of the lenses is a dipping lens. Recent development on sample chambers allows one to overcome this issue, and results in live mammalian cells have been shown. A compact 4-pi commercial version has recently been proposed (Leica), achieving a seven-fold-improved axial resolution (80 nm) compared to the classical two-photon microscope.

Another technique, also developed by Professor Hell's group, is called stimulated emission depletion (STED) microscopy. The fluorophore in the fluorescent state S1 is stimulated to the ground state S0 with a doughnut-shaped beam. The depletion of the S1 saturates with increased I_{STED} intensity, and therefore establishes a nonlinear relationship between the fluorescence and the intensity applied to the STED. The saturation of S1 depletion is the essential element for breaking the diffraction limit as it confines fluorescence to the center. Spot size is not limited, but in practise depends on the quality and the saturation factor limited depletion. Using STED microscopy with wavelengths of 750–800 nm leads to a lateral resolution of up to 28 nm on single molecules experiments.

STED microscope has been successfully combined to 4-pi microscopy, and lateral resolution of 30–40 nm has been demonstrated.

Others Nonlinear Imaging Techniques

The success of two-photon microscopy has opened the way for other nonlinear microscopy. As opposite to multiphoton fluorescence absorption techniques where the induced signal is incoherent fluorescence, the signal generated by harmonic generation and stimulated Raman scattering is coherent with specific directional and polarization properties, as such fluorescence and harmonic images are derived from fundamentally different contrast mechanisms. However, both techniques share many features like intrinsic three-dimensionality, reduce photobleaching or deep sectioning, and require essentially identical equipment, allowing one to combine these two techniques as discussed further below.

Second-Harmonic Generation Microscopy

SHG is a second-order nonlinear optical process that can only arise from a medium lacking a center of symmetry, for example, an anisotropic crystal or at an

interface of a membrane. SHG can be used to image tissue, muscles, cells membranes, or protein array, and has been recently reported to enhance membrane potential by using voltage sensitive dye.

SHG does not arise from an absorptive process, as the intense laser field induces a nonlinear, second-order, polarization in the assembly of molecules, resulting in the production of a coherent wave at exactly twice the incident frequency (half the wavelength). As represented in Figure 10, for an excitation at 800 nm, the SHG is observed at 400 nm and depending on the fluorescence properties of the molecule, a broadband fluorescent spectra is emitted in the visible.

Contrary to TPEF, where the emission properties depend only on the molecular properties and are independent of the laser source, the spectral and temporal characteristics of SHG are derived from the laser source. The SHG is temporally synchronous with the excitation pulse and its bandwidth is $1/\sqrt{2}$ of the laser's bandwidth. As SHG is an instantaneous process, the signal is only generated during the duration of the laser pulse (Figure 11), in contrary to TPEF where de-excitation of the molecules can occur over several nanoseconds after the pulse excitation.

A major difference between SHG and TPEF is that SHG requires a noncentrosymmetric environment to produce a signal. By examining the second term of eqn [1] $P = \chi^{(2)}E^2$ where the polarization and electric field are vector quantities and $\chi^{(2)}$ is a tensor, this condition is only satisfied for a birefringent crystal, an order array, or an interface. For example, the induced polarization from molecules in a dye cuvette (centro-symmetric sample) would be equal and opposite in all directions, leading to a vector sum zero. The SHG propagation is essentially in the forward direction, and the emission pattern is confined to twin lobes in the case of surface generation, or a cone in the case of volume generation. The off-axis of the two lobes critically depend on the tight focus of the driving field.

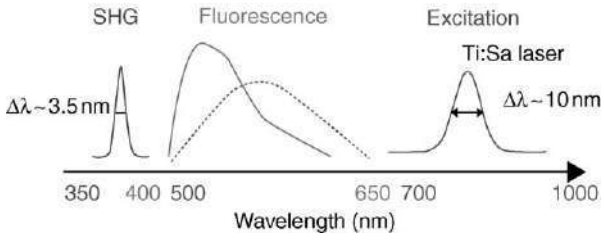

Figure 10 Wavelength domain for two-photon induced fluorescence and second-harmonic generation: excitation meets the same criteria and Ti: Sa are typically used in the range 700–1000 nm. For a typical excitation at 800 nm, fluorescence will be observed over ~450 nm, and SHG, at exactly 400 nm.

The total radiated SHG power can be derived by integrating the radiation profile from all the emission directions: $P_{\mathrm{SHG}} = \frac{1}{2}\theta_2 N^2 \sigma_2 I_0^2$ with $\sigma_2 \propto |\chi_2|^2$, I_0 the driving field intensity at the focal center, N the effective number of radiative molecules, and $\theta_2 \approx (0.1\lambda/w_0)^2$.

A schematic of a typical SHG/TPEF microscope is shown in Figure 12. The excitation source is a mode locked Ti:Sa laser which delivers ~120 fs at 76 MHz. The laser light is focused into the sample with a water immersion microscope objective (Olympus, LUMplanFl 60x) and the resultant SHG is collected in a forward direction, while TPEF is collected in the backward direction.

Figure 13 illustrates SHG and TPEF images of Di-6-ASPBS molecules under 880 nm excitation.

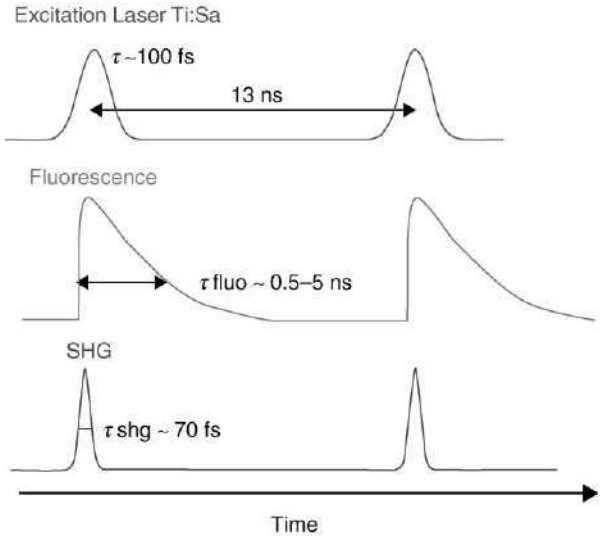

Figure 11 Typical time-response for two-photon induced fluorescence and SHG. SHG signal is an instantaneous process whereas fluorescence emission can occur several nanoseconds after the pulse excitation.

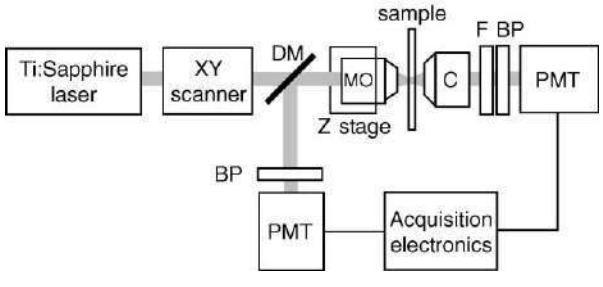

Figure 12 Schematic of SHG setup combined with TPEF extracted from (DM dichroic mirror, MO microscope objective NA 0.9 water immersion, C condenser, F colored glass filter, BP bandpass filter, PMT photomultiplier tube). Reproduced with permission from Moreaux L, Sandre O, Blanchard-Desce M and Mertz J (2000) Membrane imaging by simultaneous second-harmonic generation and two-photon microscopy. *Optics Letters* 25: 320–322, © 2000 Optical Society of America.

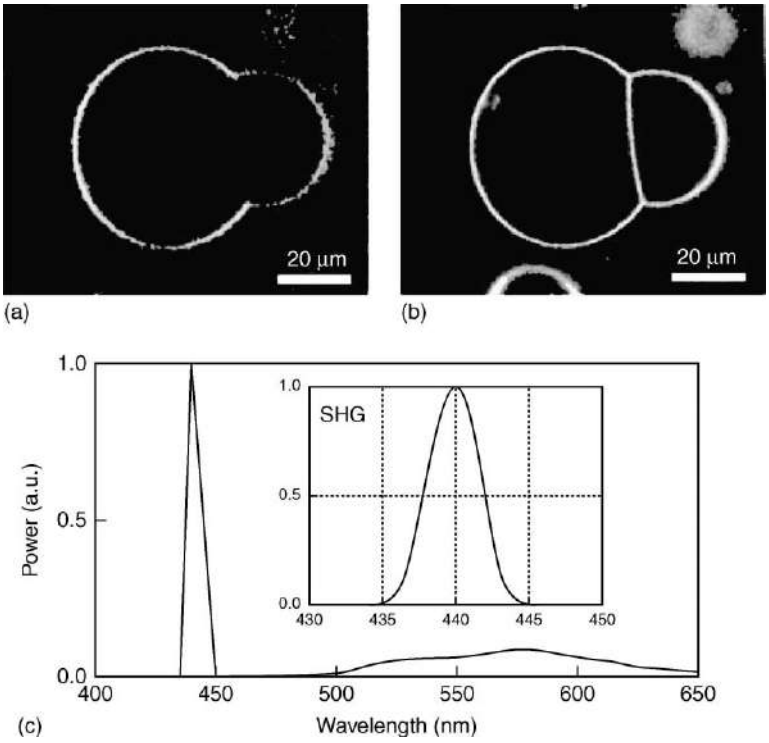

Figure 13 (a) SHG and (b) TPEF of two adhering vesicles labeled with Di-6-ASPBS (equatorial slice) excited at 880 nm. (c) The forward-detected SHG spectra (left peak) and the backward broad band fluorescence spectra (right peak). The Stokes shift in fluorescence leads to an important spectral separation between the two peaks. An expanded view of the SHG peak is presented in the inset. Reproduced with permission from Moreaux L, Sandre O, Blanchard-Desce M and Mertz J (2000) Membrane imaging by simultaneous second-harmonic generation and two-photon microscopy. *Optics Letters* 25: 320–322, © 2000 Optical Society of America.

The total acquisition time for the images was 1.5 s for an excitation power less than 1 mW. The adhesion area where the membranes are fused exhibits a centrosymmetric molecular distribution wherein TPEF is allowed, but not SHG. The corresponding forward-detected emission spectra of SHG (left peak) and TPEF (right peak) is shown in Figure 13c.

Third-Harmonic Generation (THG)

Third harmonic generation takes advantage of the discontinuity of third-order nonlinear susceptibility $\chi^{(3)}$ at material interfaces. At the focus point of an input laser beam with wavelength λ, this discontinuity gives rise to a new wave in the forward direction at the third-harmonic, $\lambda/3$. This signal can be analyzed in the transmission mode either with a single pixel detector or with a CCD camera. As with the other nonlinear techniques, THG offers a high 3D sectioning capability. Since all materials have nonvanishing third-order susceptibilities, THG microscopy can utilized as a general-purpose microscopy technique, with no need for fluorescence labeling or staining. Experimental setup for THG are very closed to SHG microscope. The need for efficient THG in the sample, together with low average laser power to avoid heating, dictates a laser source that operates with femtosecond pulses. The lasers that are often used THG microscopy are a synchronously pumped OPO at the wavelength of 1500 nm, a low repetition-rate OPA at 1200 nm and a Cr:Forsterite laser at 1240 nm. All these lasers are in the infrared above 1200 nm, thus generating THG signals in the visible range. These visible signals are compatible with the collection optics. A infrared laser undergoes low scattering, however absorption in water strongly increases.

THG have been demonstrated from subcellular to millimeter-size organisms with imaging of nerve cells, chloroplasts of a leaf, or Ca^{2+} intracellular dynamics.

Coherent Anti-Raman Stokes (CARS)

Chemical imaging by use of inherent molecular vibration signals avoids the photobleaching problem and perturbations to cell function induced by fluorophore labeling. Confocal Raman microscopy can provide 3D spatial resolution but it requires high average power due to small cross-section of Raman scattering, or long-time exposure which is not desirable in studying dynamic living system.

The above difficulties can be circumvented by multiphoton vibrational microscopy based on coherent anti-stokes Raman scattering (CARS). CARS is a four-wave mixing process in which a pump field $E_p(\omega_p)$, a Stokes field E_S (ω_S), and a probe field $E_{p'}$ $(\omega_{p'})$. Interact with a sample and generate an anti-Stokes field E_{AS}, at the frequency $\omega_{AS} = \omega_p - \omega_S + \omega_{p'}$. In practise, the pump and the probe fields derive from the same laser $\omega_p = \omega_{p'}$ and $\omega_{AS} = 2\omega_p - \omega_S$. The energy diagram of CARS is shown on Figure 14.

The CARS intensity is the square modulus of the induced polarization, $P^3 = \chi^{(3)} E_p^2 E_s$, with $\chi^{(3)}$ the third-order susceptibility.

The application of the CARS process to microscopy opens up a new method for chemical imaging. In CARS microscopy, the signal is also dependent on the square of the pump field intensity, so the temporally and spatially overlapped pump and Stokes laser pulses tightly focused into a sample, generate a signal of a small excitation volume ($<1\ \mu m^3$). As with all other nonlinear imaging techniques, the image is obtained by raster scanning the sample or the laser beams. The vibrational contrast in CARS microscopy arises from signal enhancement when $\omega_p - \omega_s$ is tuned to Raman-active vibrational frequency Ω. CARS signal generation needs to fulfill the phase-matching condition, $l < lc = \pi/|\Delta \mathbf{k}| = \pi/|\mathbf{k}_{as} - (2\mathbf{k}_p - \mathbf{k}_s)|$, where $\mathbf{k}_p$, $\mathbf{k}_s$, and $\mathbf{k}_{as}$ are wavevectors of the pump, Stoke and CARS fields, respectively, and $\Delta \mathbf{k}$ is the wavevector mismatch.

CARS presents several advantages similar to TPEF: high 3D sectioning capabilities, a low scattering and absorption of the excitation beams. CARS signal is higher in frequency than one-photon fluorescence, so it can be detected in the presence of a strong fluorescent background. Vibrational contrast is intrinsic to biological samples. Fluorescent probes are not necessary and photobleaching is thus circumvented. The coherent signal accumulation in the CARS process produces a strong and directional signal, making CARS microscopy much more sensitive than the conventional Raman microscopy. Consequently, CARS microscopy requires only a moderate average power tolerable by biological samples and has a high data acquisition speed. However, CARS microscopy is not without shortcomings. First, the intrinsically weak induced nonlinear polarizability requires sophisticated laser excitation sources with high peak power and moderate average power, and with low jitter. Second, unlike fluorescence imaging, CARS measurement is not background free, and the nonresonant signal limits the image contrast.

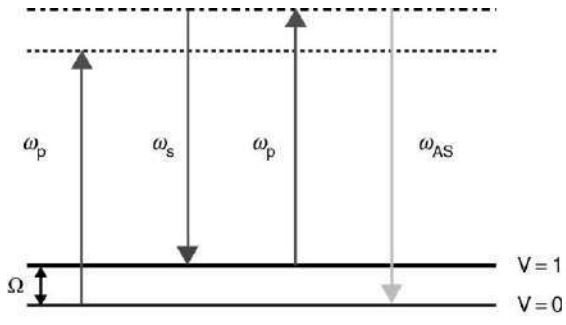

Figure 14 Energy diagram for Coherent anti-Raman spectroscopy (CARS).

The typical setup is very similar equipment compared to TPEF and SHG. The main difference arise from the excitation, two laser beams, the pump, and Stokes laser beams overlapping (in space and time) and are tightly focused into the sample using a high NA objective. The CARS signal can be detected in the forward (F-CARS) and the epi-detection (E-CARS). Unlike fluorescence emission and spontaneous Raman scattering, the radiation pattern of CARS is dependent on both the size and shape of a scatterer. The two detection geometry are thus complementary, the forward detection is suitable for imaging objects of a size comparable to or larger than the excitation wavelength.

For smaller objects, the F-CARS contrast is limited by the large nonresonant background from the solvent. Epi-detected CARS (E-CARS) microscopy provides a sensitive means of imaging objects having an axial length much smaller than the excitation wavelength because it avoids the large background from the solvent.

On the application side, CARS microscopy has been used for mapping the distribution, orientation, and diffusion of specific molecules in dynamic samples, for imaging lipids, proteins, and chromosomes in unstained live cells, and for monitoring cellular processes. CARS is a very promising technique to image small molecules such as lipids, hormones, and drug molecules, for which fluorescence labeling is prone to alter their function in cells and tissues. As for SHG, CARS microscopy is often combined with TPEF in order to obtain simultaneously the complementary contrasts.

Conclusion

Since its introduction more than a decade ago, two-photon fluorescence microscopy has quickly become a standard tool in biophotonic with applications ranging from the study of single molecules to thick tissue imaging. TPEF has opened the way to new coherent nonlinear imaging techniques such as SHG,

THG, or CARS which all shared a high spatial resolution. Based on different contrasts these different approaches are often combined for a more complete diagnosis. With developments of new low cost, compact, and reliable lasers with tuning wavelength covering the spectra of a wide range of endogenous or extrinsic markers, nonlinear imaging will continue its successful implementation outside physics labs. Development of endoscopy systems in which ultrafast pulse duration is preserved, will open a new range of *in vivo* studies. Nonlinear imaging still has more to offer and will definitely be a powerful tool for biomedical applications.

See also

Microscopy: Overview. **Nonlinear Optics, Basics:** $\chi^{(2)}$–Harmonic Generation; $\chi^{(3)}$–Third-Harmonic Generation. **Nonlinear Sources:** Harmonic Generation in Gases. **Spectroscopy:** Second Harmonic Spectroscopy.

Further Reading

Cheng J and Xie XS (2004) Coherent anti-stokes Raman scattering microscopy: instrumentation, theory and applications. *Journal of Physical chemistry B* 108: 827–840.

Gratton E, Breusegem S, Sutin J, Ruan Q and Barry N (2003) Fluorescence lifetime imaging for the two-photon microscope: time-domain and frequency-domain methods. *Journal of Biomedical Optics* 8: 381–390.

Hell SW (2003) Towards fluorescence nanoscopy. *Nature Biotechnology* 21: 1347–1355.

Konig K (2000) Multiphoton microscopy in life sciences. *Journal of microscopy* 200: 83–104.

Moreaux L, Sandre O, Blanchard-Desce M and Mertz J (2000) Membrane imaging by simultaneous second-harmonic generation and two-photon microscopy. *Optics Letters* 25: 320–322.

Schwille P, Haupts U, Maiti S and Webb WW (1999) Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophysics Journal* 77: 2251–2265.

So PTC, Dong CY, Masters BR and Berland KM (2000) Two photon micrscopy excitation fluorescence microscopy. *Annual Review of Biomedical Engineering* 399–429.

Squier J and Muller M (2001) High resolution nonlinear microscopy: a review of sources and methods for achieving optimal imaging. *Review of Science Instruments* 72: 7.

Zipfel WR, Williams RM and Webb WW (2003) Nonlinear magic: multiphoton microscopy in the biosciences. *Nature Biotechnology* 21: 1369–1377.

Zipfel WR, Williams RM, Christie R, Nikitin AY, Hyman BT and Webb WW (2003) Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation. *Proceedings of the National Academy of Science of USA* 100: 7075–7080.

Phase Contrast Microscopy

C J R Sheppard, National University of Singapore, Singapore

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Many biological specimens behave as phase objects, that is, they alter the phase of the light transmitted through them, rather than its intensity. Similarly, reflection from a surface with height variations alters the phase of the reflected light. Observation of phase objects requires special imaging techniques. These include defocusing, dark field, Zernike phase contrast, Schlieren imaging, Hoffmann modulation contrast, differential phase contrast, interference microscopy, shearing interferometry, Nomarski differential interference microscopy, and digital phase retrieval. The imaging performance of these different methods are most easily appreciated and compared by consideration of the transfer function approach. In particular, we can consider weakly scattering objects and slowly varying phase objects as special cases.

When a thin unstained biological specimen is illuminated, in general the transmitted light suffers changes in amplitude, phase, and polarization. The change in polarization is related to the birefringence of the sample, which is neglected in the following. Thus the object (complex) amplitude transmittance can be written:

$$t(x, y) = a(x, y)\mathrm{e}^{\mathrm{i}\phi(x,y)} \qquad [1]$$

where $a(x, y)$ is real and represents the modulus (also called amplitude) of the light wave on exiting from the sample, and $\phi(x, y)$, also real, represents the phase. In a perfect microscope, we observe the

intensity of the transmitted light, so the image is given by

$$I(x, y) = |a(x, y)e^{i\phi(x,y)}|^2 = a^2(x, y) \qquad [2]$$

Thus the phase information is invisible in the image. A perfect microscope does not image phase variations in the object, but results in an image that depends on the amplitude of the transmittance, called amplitude contrast. Usually the change in modulus is also small, so that the image will exhibit weak contrast, which may not be discernable. In order to render the phase variations visible it is necessary to modify the microscope in one of a number of different ways.

This conclusion is also relevant in incident light microscopy of surface structures. Then phase changes result from variations in surface height, and can be made visible by phase contrast microscopy. Quantitative measurement of the phase allows quantitative measurement of the surface profile. But care must be taken to avoid errors caused by phase change on reflection from a conducting material, or from effects of the numerical aperture of the optical system. In transmission microscopy quantitative phase measurement can be useful too, for example giving the concentration of a solute.

Some of these phase contrast methods are easily understood using the simple imaging model described above. Others are more easily appreciated by considering imaging in terms of Fourier optics, equivalent to Abbe's theory of microscope imaging. We introduce the angular spectrum of the object amplitude transmittance:

$$T(m, n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t(x, y)\exp[-i2\pi(mx + ny)]dx\,dy \qquad [3]$$

where m, n are spatial frequencies in the x, y directions. We can write for the object amplitude transmittance:

$$t(x, y) = e^{b(x,y)} \qquad [4]$$

where $b(x, y)$ is in general complex. For a weak object we have

$$t(x, y) \approx 1 + b(x, y) \qquad [5]$$

and the spectrum is

$$T(m, n) = \delta(m)\delta(n) + B(m, n) \qquad [6]$$

where δ is a Dirac delta function and $|B| \ll 1$. The problem with imaging of phase arises because a change in the imaginary part of b produces only a small change in the modulus of t.

For a coherent imaging system, with coherent transfer function $c(m, n)$, the image intensity is

$$I(x, y) = \left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c(m, n)T(m, n) \times \exp\left[\frac{i2\pi}{M}(mx + ny)\right]dm\,dn\right|^2 \qquad [7]$$

where M is the magnification. The coherent transfer function $c(m, n)$, is equal to the scaled pupil function $P(\lambda fm, \lambda fn)$, where f is the focal length of the objective. Substituting eqn [6] into eqn [7] and dropping terms of second order we obtain:

$$I(x, y) = |c(0, 0)|^2 + 2\Re\left\{c^*(0, 0)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \times c(m, n)B(m, n)\exp\left[\frac{i2\pi}{M}(mx + ny)\right]dm\,dn\right\} \qquad [8]$$

Normalizing the transfer function to unity at the origin:

$$I(x, y) = 1 + 2\Re\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c(m, n)B(m, n) \times \exp\left[\frac{i2\pi}{M}(mx + ny)\right]dm\,dn\right\} \qquad [9]$$

For an axially symmetric aberration-free system, the coherent transfer function is real and even, so that a pure phase object is not imaged (only the Hermitian component of B is imaged). Many phase contrast methods in principle rely on appropriate modification of the coherent transfer function, either by making it complex or asymmetric.

However, this theory is an approximation as microscope systems are in practice partially coherent. For a partially coherent system the image intensity is

$$I(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(m,n;p,q)T(m,n) \times T^*(p,q)\exp\left\{\frac{i2\pi}{M}[(m-p)x+(n-q)y]\right\}dm\,dn \qquad [10]$$

where $C(m,n;p,q)$ is the partially coherent transfer function, sometimes called the transmission cross-coefficient, which can be calculated as an integral over the objective and condenser pupils. The partially coherent transfer function is nonzero over a finite region of $m,n;p,q$ space, thus limiting the resolution of the system. Then eqn [8] for a weak object

becomes, normalizing so that $C(0,0;0,0)=1$:

$$I(x,y)=1+2\Re\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}C(m,n;0,0)B(m,n)\times\exp\left[\frac{\mathrm{i}2\pi}{M}(mx+ny)\right]\mathrm{d}m\,\mathrm{d}n\right\} \quad [11]$$

where $C(m,n;0,0)$ is called the weak object transfer function.

For the image of a general object that is not necessarily weak, eqn [10] is often difficult to interpret. We can gain much insight into the imaging process by considering another special form of object, this time a slowly varying object:

$$t(x,y)=a(x,y)\exp\left[\mathrm{i}\left(x\frac{\partial\phi}{\partial x}+y\frac{\partial\phi}{\partial y}\right)\right] \quad [12]$$

where a and ϕ are assumed slowly varying. Thus, the object spectrum, assumed now as slowly spatially varying, is

$$T(m,n)\approx a\delta\left(m-\frac{1}{2\pi}\frac{\partial\phi}{\partial x}\right)\delta\left(n-\frac{1}{2\pi}\frac{\partial\phi}{\partial y}\right) \quad [13]$$

Substituting in eqn [10], we then find:

$$I(x,y)=a^2(x,y)C(m,n;m,n)$$
$$m=\frac{1}{2\pi}\frac{\partial\phi}{\partial x},\ n=\frac{1}{2\pi}\frac{\partial\phi}{\partial y} \quad [14]$$

in which C is called the phase gradient transfer function, and is real. For an ordinary bright field system, C is an even function. Thus a phase gradient results in a change in image intensity, but for an axially symmetric system the sign of the change is independent of the sign of the phase change.

Defocused Imaging

Historically, the oldest method of phase contrast is simply to defocus the object. The transfer function is then complex, and phase detail is imaged. If the system is aberration-free, the contrast of the phase information reverses on the opposite side of focus. The strength of the contrast increases with defocus, until eventually artifacts are introduced when the imaginary part of the transfer function becomes negative. If the relative condenser aperture is too large, contrast is reduced.

Assume a pure weak phase object:

$$t(x,y)=1+\mathrm{i}d\cos(2\pi\nu x) \quad [15]$$

so that

$$T(m,n)=\left[\delta(m)+\mathrm{i}\frac{d}{2}\delta(m-\nu)+\mathrm{i}\frac{d}{2}\delta(m+\nu)\right]\delta(n) \quad [16]$$

Then:

$$I(x,y)=1-dC_{\mathrm{i}}(\nu,0;0,0)\cos\left(\frac{2\pi\nu x}{M}\right) \quad [17]$$

where C_{i} is the imaginary part of C. Thus the phase information is imaged by the imaginary part of the weak object transfer function.

Dark Field Microscopy

Dark field microscopy can be used to visualize weak amplitude and phase information. The constant background in eqn [5] is eliminated, so we obtain an image of b. Note that eqns [8] and [11] are no longer valid because we cannot neglect the second-order term. For a coherent system, the background can be eliminated by use of a central stop to intercept the nondiffracted light. For a partially coherent microscope, an annular condenser aperture (larger in diameter than the objective pupil) is used (**Figure 1**). In both cases, in practice a range of low spatial frequencies are eliminated, resulting in the appearance of halo artifacts.

For the pure weak phase object of eqn [12], the image in a coherent microscope is

$$I(x,y)=\frac{1}{2}d^2c^2(\nu,0)\left[1+\cos\left(\frac{4\pi\nu x}{M}\right)\right] \quad [18]$$

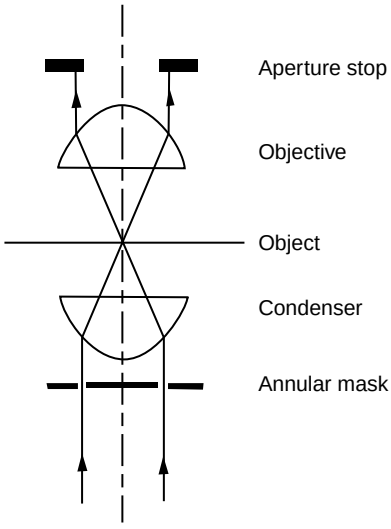

Figure 1 Schematic diagram of dark field microscopy.

so that the spatial frequency is double that in the original object. For the partially coherent microscope, we have

$$I(x,y) = \frac{1}{2}d^2\left[C(\nu,\nu;0,0) + C(\nu,-\nu;0,0)\cos\left(\frac{4\pi\nu x}{M}\right)\right] \qquad [19]$$

in which the two terms represent difference and sum frequency components. However, it can be shown that for the annular dark field system the transfer function for sum frequencies is zero, so the image intensity is a constant. The system images only difference frequencies of pairs of spatial frequencies, and not sum frequencies.

For a slowly varying phase object, zero intensity results from a region of constant phase, but the intensity rises quickly for small values of phase gradient, thus giving strong contrast. For larger phase gradients, the intensity falls off slowly.

Zernike Phase Contrast

The Zernike phase contrast method is similar to dark field except that the direct (nondiffracted) light has its relative phase changed by 90°, rather being eliminated. In practice this is achieved using an annular condenser aperture and an objective with a phase ring (Figure 2). The phase change can be ±90°, giving positive or negative phase contrast. The phase ring is usually only partially transmitting, which has the effect of enhancing the sensitivity. From eqns [8] or [11], imaging of a weak object is linear in phase. The halo artifact is present, as in dark field microscopy. Weak amplitude information is imaged as in dark field microscopy.

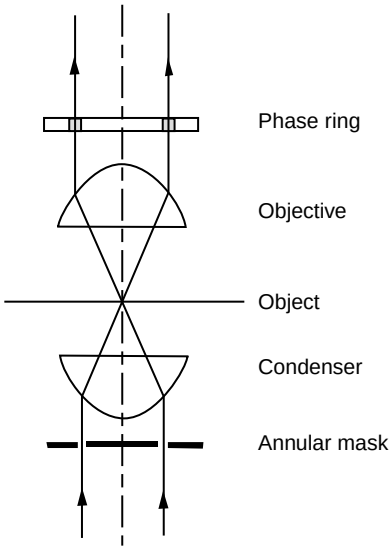

Figure 2 Schematic diagram of Zernike phase contrast microscopy.

For the weak phase object eqn [12], we then have

$$I(x,y) = 1 \pm \frac{2d}{g}C(\nu,0;0,0)\cos\left(\frac{2\pi\nu x}{M}\right) \qquad [20]$$

for positive or negative phase contrast, respectively, where g is the amplitude transmittance of the phase ring. The response of the weak object transfer function is improved as a result of the annular condenser aperture.

Schlieren Imaging, Hoffman Modulation Contrast, and Differential Phase Contrast (DPC)

In Schlieren imaging with coherent illumination, a half plane mask (a Foucault knife edge) is inserted to eliminate say the negative spatial frequencies in x. Taking now $c(0,0) = 1/2$, eqn [8] becomes, after normalization:

$$I(x,y) = 1 + 4\Re\left\{\int_{-\infty}^{\infty}\int_{0}^{\infty} c(m,n)B(m,n) \times \exp\left[\frac{i2\pi}{M}(mx+ny)\right]dm\,dn\right\} \qquad [21]$$

For the weak phase object eqn [15], we then have:

$$I(x,y) = 1 - dc(\nu)\sin\left(\frac{2\pi\nu x}{M}\right) \qquad [22]$$

The image exhibits a variation in phase quadrature to the original phase variation, that is, it is related to the x-derivative of the phase. The effect occurs because the coherent transfer function contains an odd (but real) part.

For a partially coherent microscope a similar effect can be achieved simply by offsetting the illumination system, but now the transfer function is no longer single-sided. It can be resolved into odd and even parts that result in differential phase contrast and amplitude contrast, respectively. The amplitude contrast component merely contributes to the background for a phase-only object, and can be removed simply by contrast enhancement of a digital image. The offset can be introduced in practice by inserting a half-plane mask into the condenser aperture stop. Imaging of a slowly varying object can be explained by eqn [14], in which the transfer function is not symmetrical. The image shows a pseudo-three-dimensional bas relief effect. A commercial implementation of an asymmetric transfer function is Hoffman modulation contrast, which uses offset

illumination together with an amplitude mask in the objective pupil (Figure 3). Advantages of Hoffman contrast are that it uses standard bright-field objectives and avoids problems from birefringent specimens.

A similar method can be used in scanning laser microscopy. In a scanning microscope the effect of the illumination and detection systems are interchanged. Thus DPC can be achieved in a scanning laser microscope by offsetting the detector. In practice this is conveniently done by using a split detector (Figure 4), a detector divided into two semicircular elements. If the signal of one half is subtracted from that of the other, the amplitude contrast component cancels out, so that pure differential phase contrast results. For weak phase gradients, the intensity varies linearly with phase gradient, again showing the bas relief effect. It is thus bi-directional, and the difference signal can become negative. A constant value is, therefore, usually added to it before display. The sum of the two detector elements results in a conventional bright-field image. From eqn [14], both difference and sum signals are proportional to $a^2(x, y)$, so that if the ratio of these two signals is extracted, an image of the phase gradient in the x direction results. The response to either weak phase gradients or fine detail can be enhanced by modifications using annular split detectors. Use of a quadrant detector allows the phase gradients in the x and y directions to be imaged. DPC microscopy has been demonstrated to be sensitive enough to observe monomolecular films and atomic surface steps.

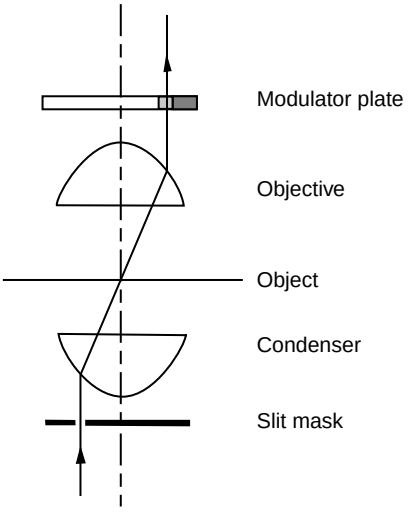

Figure 3 Schematic diagram of Hoffmann modulation contrast microscopy.

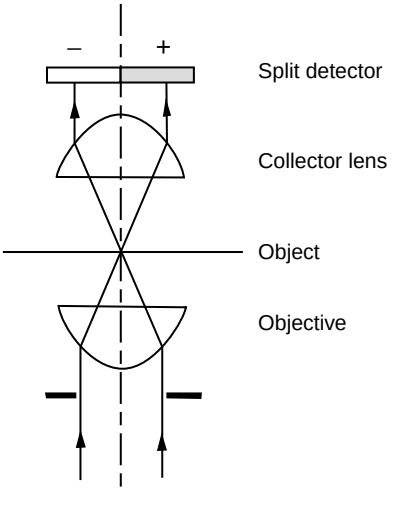

Figure 4 Schematic diagram of differential phase contrast (DPC) microscopy.

Interference Microscopy

Interference microscopy allows quantitative measurement of the object phase by interference of the object beam with a reference beam. In the transmission geometry, a Mach–Zehnder interferometer is used, which is effectively two parallel microscopes, each with matched condenser and objective lenses. In the reflection geometry we can use a Michelson, a Mirau, or a Linnik interferometer. The Michelson interferometer has an inclined beamsplitter and the reference mirror situated between the objective and object, and hence can be used only with low numerical aperture objectives. The Linnik interferometer has matched objectives in the object and reference beam paths. There is no limit to the numerical aperture that can be used, but it is not a common-path interferometer and is hence sensitive to vibrations and air currents. Probably the most practicable system for interference microscopy in the reflection geometry is based on the Mirau interferometer that uses a beamsplitter with its normal parallel to the optic axis. Until recently, interference microscopes used a small condenser aperture giving nearly coherent illumination. The image consists of three components: an object beam term (noninterference image) a reference beam term, and the interference term. The phase (and amplitude) of the object beam can be extracted using phase stepping or heterodyning techniques. The measured phase is wrapped, and requires unwrapping to obtain the absolute phase.

If the condenser aperture is opened up, the behavior is modified. The interference term is

given by

$$I_{\text{int}}(x, y) = 2\Re\left\{r^* \int\int c_{\text{int}}(m, n)T(m, n) \times \exp\left[\frac{\mathrm{i}2\pi(mx + ny)}{M}\right]\mathrm{d}m\,\mathrm{d}n\right\} \quad [23]$$

where $c_{\text{int}}(m, n)$ is the coherent transfer function for interference imaging, given by a convolution integral of the object path objective pupil with the product of the reference pupil and the source. The first thing to note is that for a high numerical aperture, the spacing of the interference fringes is increased by 20–40%, and depends on defocus, tilt, and central obscuration. Thus accurate measurement of phase requires careful calibration. The objective pupil exhibits a phase variation when defocused that reduces the modulus of the coherent transfer function, resulting in the fringes being modulated by an envelope the width of which becomes smaller as the aperture becomes larger. The envelope is similar to that of the optical sectioning effect in confocal microscopy. This envelope allows the zero-order fringe to be identified, although simulations have suggested that fringe hopping could occur at phase jumps as a result of diffraction effects.

Multiple Beam Interferometry

The topography of a surface can be investigated by using a Fizeau interferometer formed by bringing a coated optical flat near to the surface. Then if the surface and flat have high reflectivity, multiple beam interference occurs and the width of the bright fringes can be much reduced. The system can be used with a tilted flat, to give fringes that are substantially straight, or with an almost parallel flat, so that the entire field of view is occupied by a single fringe. In the latter case, the sensitivity can be in the subnanometer range. Multiple beam interference also enhances the contrast of weak amplitude information.

Shearing Interferometry

In conventional interferometry, the object beam is arranged to interfere with a constant reference beam. In shearing interferometry, on the other hand, it interferes with a shifted version of itself.

Lateral Shearing Interferometry

An approach that is suitable for quantitative optical path measurements is to use an interferometer that shears the microscope image. The Interphako system uses a Mach–Zehnder interferometer that can measure optical path differences as low as ~3 nm in either the reflection or transmission geometry. The interferometer allows the shear distance to be varied.

Differential Interference Microscopy (DIC)

The most common form of shearing interferometry is differential interference microscopy, usually associated with the name of Nomarski. DIC microscopy can be performed in either the transmission or the reflection geometry. A plane polarized illumination beam is split into two orthogonally polarized beams traveling at slightly differing angles by a Wollaston prism in the front focal plane of the condenser lens (Figure 5). The beams are thus sheared laterally in the object plane. The light from the sample travels through the objective and a second Wollaston prism, which recombines the two beams into one. In the reflection geometry, a single lens and Wollaston prism are used. An analyzer set at 45° allows the two beams to interfere. The shear is arranged to be smaller than the width of the point spread function. Different prisms are supplied for use with different magnification objectives. The system results in

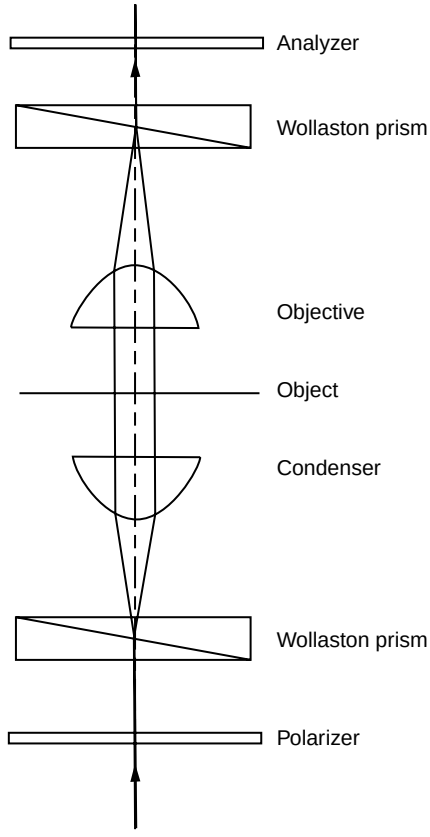

Figure 5 Nomarski differential interference contrast (DIC) microscopy.

differential phase contrast, but there are differences as compared with the DPC microscope described above.

There are two different implementations of DIC in use. In one implementation, the illumination polarizer and analyzer are crossed. The relative phase difference between the two beams (the bias retardation) is adjusted by moving the Wollaston prism laterally. It is claimed that rotation of the analyzer can be useful when observing mixed phase/amplitude objects. In the second arrangement (the Sénarmont compensator) the position of the Wollaston prism is fixed, and the analyzer rotated to change the relative phase.

A recent variation of DIC uses circularly polarized illumination. The Wollaston prism still produces two plane polarized components, but with an additional phase difference of $\pi/2$. The advantage of using circular polarization is that the prism can then be rotated to change the direction of shear.

If the shear is 2Δ, and the phase difference is $\pi - 2\phi_0$ (i.e., if $\phi_0 = 0$ then the two beams are in anti-phase), then the effective object spectrum is:

$$T_{\text{eff}}(m,n) = -2\sin(2\pi m\Delta - \phi_0)T(m,n) \qquad [24]$$

Thus the effective transfer function can be written:

$$\begin{aligned} C_{\text{eff}}(m,n;p,q) &= 4\sin(2\pi m\Delta - \phi_0)\sin(2\pi p\Delta - \phi_0) \\ &\quad \times C(m,n;p,q) \\ &= 2\{\cos[2\pi(m-p)\Delta] \\ &\quad - \cos 2\phi_0 \cos[2\pi(m+p)\Delta] \\ &\quad - \sin 2\phi_0 \sin[2\pi(m+p)\Delta]\} \\ &\quad \times C(m,n;p,q) \end{aligned} \qquad [25]$$

It is the final term of the three in braces that, because it is an odd function, results in differential phase contrast. We can now examine imaging of special forms of object. For a weak object, putting $p = 0$ as in eqn [11]:

$$\begin{aligned} C_{\text{eff}}(m,n;0,0) = 2\{&(1-\cos 2\phi_0)\cos[2\pi m\Delta] \\ &- \sin 2\phi_0 \sin[2\pi m\Delta]\}C(m,n;0,0) \end{aligned} \qquad [26]$$

The first term in braces corresponds to an amplitude contrast background component because it is real and even. The second term results in DPC, because it is real and odd. For zero bias retardation, dark-field imaging results. As the bias retardation is increased from zero, the strength of both the amplitude contrast and the DPC components increase, but that of the DPC component increases more rapidly. For visual observation we require good contrast, and hence we choose a small value of bias retardation, just large enough so that the background is large enough to avoid artifacts in the DPC image. In digital microscopy, contrast is not so important as it can be enhanced digitally, and hence we can increase ϕ_0 to a value of $\pi/4$, which maximizes the strength of the DPC component. Further increase in bias retardation decreases the DPC signal, until for $\pi/2$ pure amplitude contrast results. The actual form of the weak object transfer function depends on the magnitude of the shear. The condition $m_0\Delta = 1/4$, where m_0 is the cut-off in spatial frequency, gives optimum performance for the imaging of weak objects.

The phase gradient transfer function is:

$$C_{\text{eff}}(m,n;m,n) = 4\sin^2(2\pi m\Delta - \phi_0)C(m,n;m,n) \qquad [27]$$

For the bright-field microscope the phase gradient transfer function falls off symmetrically from its value at zero phase gradient. The effect of the pre-multiplying factor in eqn [27] is to increase the fall-off for positive phase gradients and reduce it for negative ones. A bas-relief effect occurs, with highlighting for negative phase gradients for some values of the parameters. Increasing the value of the shear increases the highlighting effect, but, unlike in DPC, the behavior is not anti-symmetric.

Because interference is dependent on the wavelength of the light, useful and beautiful effects can be seen with DIC microscopy using white light. Usually it is regarded as a qualitative, rather than a quantitative, imaging method. However, as from eqn [14], the intensity can be written:

$$I(x,y) = 2a^2(x,y)[1 - \cos(2\pi m\Delta - \phi_0)]C(m,n;m,n) \qquad [28]$$

By recording images at different bias retardations, conventional phase stepping methods can be used to extract the phase gradient $2\pi m = \partial\phi/\partial x$.

Axial Shearing Interferometry

Instead of shearing in the transverse direction, a shear in the axial direction can be used. In the reflection geometry, the complex amplitude from the point of observation is compared with the average over a defocused spot. This method has proved to be extremely sensitive, and can detect height changes smaller than 0.1 nm.

Phase Retrieval

Digital Phase Retrieval

The phase of a single two-dimensional (2D) measurement of intensity can be retrieved by an iterative

procedure, if constraints such as nonnegativity and band-limitedness can be assumed. If the modulus of both an object and its Fourier transform are known, then the phases can also be recovered. In this case, the two 2D real data sets have information content sufficient to reconstruct the 2D complex object. This approach has been extended to the case of measurement of two defocused images, using the method of phase diversity.

Imaging Using the Transport of Intensity Equation

The transport equation of intensity can be used to recover the phase of a partially coherent wavefield directly (i.e., noniteratively) from two 2D intensity measurements. If the wavefield is represented in terms of its intensity I and phase ϕ in a plane, the transport equation can be written in the paraxial approximation and in the absence of phase singularities:

$$-k\frac{\partial I}{\partial z} = \nabla_T \cdot [I\nabla_T \phi] \qquad [29]$$

where ∇_T is a 2D operator acting in the plane. This expression can be inverted to give ϕ quantitatively, if I and $\partial I/\partial z$ are known. However, it gives the phase of the image field, rather than the phase of the object, so that obtaining quantitative values for a strong object may be difficult. The transport of intensity method is valid with partially coherent illumination: if the relative condenser aperture is too great then contrast becomes weak. Further, an effect similar to the variation in fringe spacing with aperture in interference microscopy is introduced because oblique illuminating rays traverse the sample at an angle to the axis. In practice, a series of three images are usually taken at different focus positions. Then, from the quantitative phase data, various different contrast mechanisms such as DIC, Zernike phase contrast, and so on, can be simulated.

Surface Profiling Methods: Triangulation and Coherence Domain Methods

Although strictly they are not phase imaging methods, other surface profiling methods can be used to measure optical path. These could also be adapted to measure optical path in transmission. It is important also to be aware of these methods, as they could affect measurements made using other phase contrast techniques.

Apart from the phase measurement techniques described above, surface profiling methods fall into one of two categories: triangulation and coherence domain methods. An example of a technique that relies on triangulation is confocal microscopy. Confocal microscopy exhibits an optical sectioning effect similar to that described for interference microscopy. The effect is stronger for higher numerical aperture. As a reflector is scanned through focus the intensity goes through a maximum. By locating the position of the maximum, the local surface height can be obtained. In a similar way, in a confocal transmission system, introduction of an object with some optical thickness has the effect of altering the axial position of the confocal pinhole, with a resultant drop in intensity. This effect can be used to image the optical thickness of the sample.

Coherence domain methods rely on the finite coherence length of the light. In an interferometer, this results in an envelope that multiplies that arising from the nonzero relative condenser aperture. Measurement of the peak of the envelope allows local surface height to be measured. Unlike the confocal approach, the width of this coherence envelope is independent of the numerical aperture of the system.

In reflection, both the confocal and the coherence domain methods have the advantage over direct phase imaging techniques that they allow measurement of surface height changes without errors caused by the phase change on reflection from a conducting medium.

See also

Interferometry: Phase Measurement Interferometry. **Microscopy:** Interference Microscopy. **Phase Control:** Phase Conjugation and Image Correction.

Further Reading

Bennett AH, Jupnik H, Osterberg H and Richards OW (1952) *Phase Microscopy*. New York: Wiley.

Born M and Wolf E (1999) *Principles of Optics*, 7th edn. Cambridge: Cambridge University Press.

Bradbury S and Evenett PJ (1996) *Contrast Techniques in Light Microscopy*. Oxford: Bios Scientific Publishers.

Françon M (1950) *Le Contraste de Phase en Optique et en Microscopie*. Paris: Revue d'Optique.

Françon M (ed.) (1952) *Le Contraste de Phase et le Contraste par Interférences*. Paris: Revue d'Optique.

Inoue S and Spring KR (1997) *Video Microscopy: The Fundamentals*, 2nd edn. New York: Plenum.

Lessor DL, Hartman JS and Gordon RL (1979) Quantitative surface topography determination by Nomarski reflection microscopy. *Journal of the Optical Society of America* 69: 357–366.

Pluta M (1989) *Advanced Light Microscopy*. Specialized Methods, vol. 2. Amsterdam: Elsevier.
Sheppard CJR and Wilson T (1980) Fourier imaging of phase information in conventional and scanning microscopes. *Philosophical Transactions of Royal Society A* 295: 513–536.
Wilson T and Sheppard CJR (1984) *Theory and Practice of Scanning Optical Microscopy*. London: Academic Press.

MODULATORS

Contents

Acousto-Optics
Electro-Optics
Modulation and Demodulation of Optical Signals

Acousto-Optics

M Gottlieb and D Suhre, Carnegie Mellon University, Pittsburgh, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Acousto-optics deals with the interaction of light waves with sound waves and has given rise to a large number of devices related to various laser systems for display, information handling, optical signal processing, and numerous other applications requiring the spatial or temporal modulation of coherent light. The basic principles and phenomena controlling these interactions were largely understood by the mid-1930s, but had little practical significance because very high acoustic power levels were required to attain good optical efficiency and there were few good sources of well-collimated monochromatic light. During the period from 1960 to the present, several key technologies were developing rapidly, at the same time that many applications of the laser were being suggested which require high-speed, high-resolution scanning methods. These new technologies gave rise to high-efficiency, wideband acoustic transducers capable of operation to several gigahertz, high-power wideband solid-state amplifiers to drive such transducers, and the development of a number of new, synthetic acousto-optic crystals with low-drive-power requirements and low acoustic losses at high frequencies. This combination of properties makes acousto-optics feasible for many systems, and for several is the method of choice to satisfy demanding requirements. This chapter will summarize the basic features of acousto-optic interactions and the operating principles of the most common acousto-optic devices.

The Photo-Elastic Effect

The change induced in the refractive index of a transparent material by the pressure change produced by an acoustic disturbance is the underlying mechanism of all acousto-optic interactions. An acoustic wave produces periodic regions of compression and rarefaction in the material, which modulates the density. The Lorentz–Lorenz relation relates the refractive index to the density, for the simplest case of an ideal gas

$$\frac{(n^2-1)}{(n^2+2)} \propto \rho \qquad [1]$$

where n is the refractive index and ρ is the density. This relation is followed to a good approximation for most simple solid materials as well. The elastooptic coefficient is

$$\frac{\rho\,\mathrm{d}n}{\mathrm{d}\rho} = \frac{(n^2-1)/(n^2+2)}{6n} = p \qquad [2]$$

The fundamental quantity given by eqn [2], also known as the photo-elastic constant p, is related to the pressure applied by

$$p = \frac{1}{\beta}\left(\frac{\mathrm{d}n}{\mathrm{d}p}\right) \qquad [3]$$

where P is the applied pressure and β is the compressibility of the material. The photo-elastic

constant of an ideal material with refractive index of 1.5 is 0.59. The photo-elastic constants of a wide variety of materials lie in the range from about 0.1 to 0.6.

The relation in eqn [3] follows from the usual definition of the photo-elastic constant:

$$\Delta(1/\varepsilon) = \Delta(1/n^2) = pe \qquad [4]$$

where ε is the dielectric constant, $\varepsilon = n^2$, and e is the acoustic strain amplitude. The induced change in refractive index, Δn, is

$$\Delta n = -n^3 pe \qquad [5]$$

Strain amplitudes typically lie in the range of 10^{-8} to 10^{-5}, with Δn in the range of about 10^{-8} to 10^{-5} (for $n = 1.5$). Devices based upon such a small change in refractive index are capable of generating large effects because these devices are configured in a way that can produce large phase changes at optical wavelengths.

The more complete relation defining the photo-elastic interaction is more complex than the simple scalar eqn [5] in which the photoelastic constant is independent of the directional properties of the material. In fact, even for an isotropic material such as glass, longitudinal acoustic waves and transverse (shear) acoustic waves cause the photoelastic interaction to assume different parameters. A tensor relation between the dielectric properties, the elastic strain, and the photo-elastic coefficient describes the interaction, particularly for anisotropic materials. The tensor equation is

$$\Delta(1/n^2) = \sum_{kl} p_{ijkl}\, e_{kl} = \Delta(1/\varepsilon)_{ij} \qquad [6]$$

where $(1/\varepsilon)_{ij}$ is a component of the optical index ellipsoid, e_{kl} are the Cartesian strain components, and p_{ijkl} are the components of the photo-elastic tensor. The crystal symmetry of the photoelastic material places limits on the possible configurations of interaction geometry.

Diffraction by Acoustic Waves

For typical acousto-optic devices the acoustic wave acts as a diffraction grating, made up of periodic changes in optical phase, moving at sound velocity. These features determine the properties of acousto-optic diffraction. Using a quantum-based model, the light and sound may be thought of as particles, photons and phonons, undergoing collisions in which energy and momentum are conserved. Either of these descriptions may be used to obtain all the important diffraction effects, but some are more easily understood on the basis of one or the other. A description of both is given here. Consider Figure 1 in which the light wave, of frequency ω and wavelength λ, is incident into the material with an acoustic wave of frequency f and wavelength Λ. If the refractive index of the medium is $n + \Delta n$ in the presence of the acoustic wave, the phase of the optical wave will be changed by an amount

$$\Delta\phi = 2\pi(L/\lambda)\Delta n \qquad [7]$$

where L is the length of the interaction region. Some typical values of $\Delta\phi$ can be obtained by assuming $L = 2.5$ cm and $\lambda = 0.5$ μm, with Δn reaching a peak value of 10^{-5}. This yields a phase change of π rad, which is, of course, quite large. It is large because L/λ, the number of optical wavelengths, is 50 000, so that a very small Δn can still produce a sizable $\Delta\phi$. If the electric field incident on the delay line is represented by

$$E = E_0\, e^{i\omega t} \qquad [8]$$

then the field of the phase-modulated emerging light will be

$$E = E_0\, e^{(i\omega t + \Delta\phi)} = e^{i\omega t}\, e^{2\pi i(L/\lambda)\sin(ft)} \qquad [9]$$

where f is the acoustic frequency.

A detailed derivation of the resulting temporal and spatial distribution of the light field is mathematically complex, but, analogy with radio-wave modulation can be used to arrive at the resultant fields. The spectrum of a phase-modulated carrier of frequency ω consists of components separated by multiples of the modulation frequency f, as shown in Figure 2. Sidebands are produced about the carrier frequency, such that the frequency of the mth sideband is

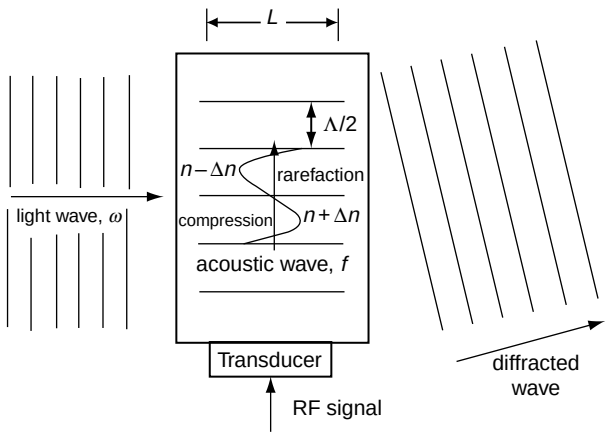

Figure 1 Interaction of light waves with acoustic waves.

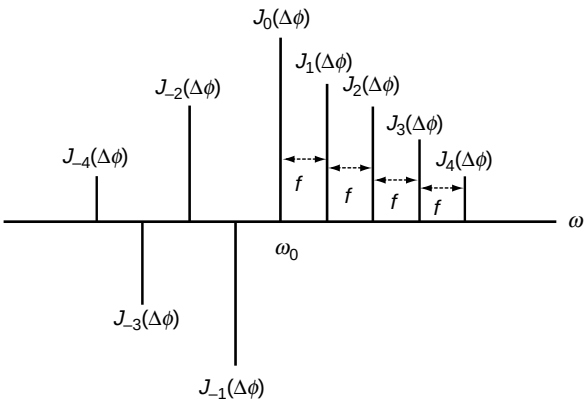

Figure 2 Spectrum of a phase-modulated wave of carrier frequency ω_0, and modulation index $\Delta\phi$.

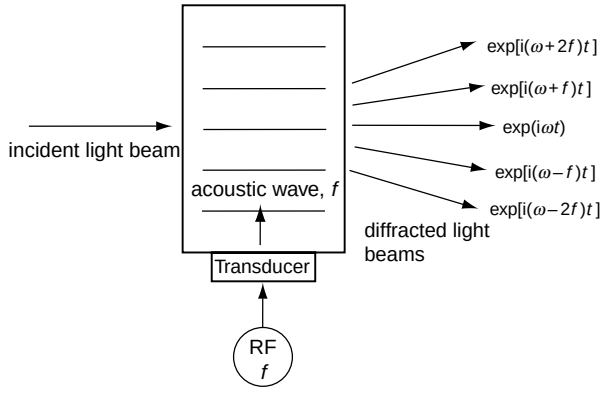

Figure 3 Diffraction of light in the Raman–Nath limit.

$\omega + mf$, where m may be positive or negative. The amplitude of each of the sidebands is proportional to the Bessel function of order equal to the sideband number, and whose argument is the modulation index $\Delta\phi$. The odd-numbered negative orders are 180° out of phase with the even-numbered ones, a feature that may be useful for certain signal processing applications. The light emerging from the delay line is composed of a number of light waves whose frequencies have been shifted by mf from the frequency ω of the incident light. The relative amplitudes will be determined by the peak change in the refractive index. The two extremes of the diffraction process are simplified in the 'thin grating' and the 'thick grating' models. The ratio of the interaction length, L, to the acoustic wavelength, Λ, will determine the character of the diffraction process. The plane wave approximation in which the acoustic energy propagates as a plane wave is valid when this ratio is very large. However, when this ratio is small, the acoustic propagation can be described in terms of a sum of plane waves, the angular spectrum of which increases as the ratio decreases. In this phenomenological model the partial wave which is propagating at an angle λ/Λ to the forward direction may diffract the light second time into an angle $2\theta = 2\lambda/\Lambda$, and the frequency of this light will once again be up-shifted, for a total frequency shift of $2f$. If the spectrum of acoustic waves contains sufficient power of still higher orders, then this rediffraction process can be repeated, so that light will be multiply diffracted m times into higher-order angles, $m\theta = m\lambda/\Lambda$ each with a frequency shift mf. A similar argument holds for the negative orders, so that a complete set of diffracted light beams will appear as shown in Figure 3, where the deflection angle corresponding to the mth order is given by $\sin\theta_m = m\lambda/\Lambda$ and the frequency of the light deflected into the mth order is $\omega_m = \omega + mf$. The intensity of the carrier wave, or zero order, will be zero when the modulation index $\Delta\phi$ is equal to 2.4. The first order will be a maximum for $\Delta\phi = 1.8$, corresponding to the maximum of the first-order Bessel function B_1, and decreasing for larger modulation indices. These phenomena were described by Debye and Sears and so are often referred to as Debye–Sears diffraction. An extensive theoretical analysis of the effect was given by Raman and Nath and is alternatively referred to as Raman–Nath diffraction. As the interaction length is increased the Raman–Nath diffraction gradually weakens. The weakening begins around an interaction length $L \sim \Lambda^2/4\lambda$. This value of L is expressed in the Q parameter

$$Q = 4L\lambda/\Lambda^2 \qquad [10]$$

which is known as the Raman–Nath parameter. A different regime of diffraction takes effect for values of the interaction length $Q \gg 1$.

The phenomena in this regime can be more easily understood in terms of the quantum description of the light and sound waves as collisions between photons and phonons. In this model, the dynamics of the collisions of the light and sound are governed by the laws of conservation of energy and momentum. The momentum vectors magnitudes of the light and sound, $\mathbf{k}$ and $\mathbf{K}$, are given by the well-known expressions

$$|\mathbf{k}| = 2\pi n/\lambda \quad \text{and} \quad |\mathbf{K}| = 2\pi/\Lambda \qquad [11]$$

where the acoustic wavelength, Λ, is related to the acoustic velocity, V, by $\Lambda = V/f$.

Conservation of momentum is expressed by the vector relation

$$\mathbf{k}_\mathrm{i} + \mathbf{K} = \mathbf{k}_\mathrm{d} \qquad [12]$$

the diagram for which is shown in Figure 3, where $\mathbf{k}_i$, and $\mathbf{k}_d$ represent the momentum of the incident photon and the diffracted photon, respectively. In this notation

$$\mathbf{k}_i = 2\pi n_i/\lambda, \qquad [13a]$$

and

$$\mathbf{k}_d = 2\pi n_d/\lambda \qquad [13b]$$

If the material is anisotropic (birefringent), n_i and n_d may be different.

Conservation of energy requires that

$$h\omega_d = h\omega_i + hf \qquad [14]$$

in which h is Planck's constant. Since ω_i lies in the optical frequency range, $\sim 10^{13}$ Hz, and f lies in the RF or microwave range, 10^6–10^9 Hz, then $\omega_d \sim \omega_i$. This results, for isotropic materials, in the magnitudes of $\mathbf{k}_i$ and $\mathbf{k}_d$ being equal and resulting in the isosceles momentum triangle of Figure 4. The angles of incidence and diffraction (with respect to the planar acoustic wavefronts), called the Bragg angles, are equal for this case, and are

$$\theta_B = (1/2)\lambda/\Lambda \qquad [15]$$

A schematic diagram for this process is shown in Figure 5. The interaction will be large only if the light is well-collimated, and incident at this angle. Unlike the Debye–Sears regime, there is only a single diffracted beam. The energy of the phonon may either be added to that of the incident photon, increasing its frequency to $\omega_d = \omega_i + f$, or the reverse, resulting in $\omega_d = \omega_i - f$. The sense of the momentum vectors determines which of these occurs. The Debye–Sears

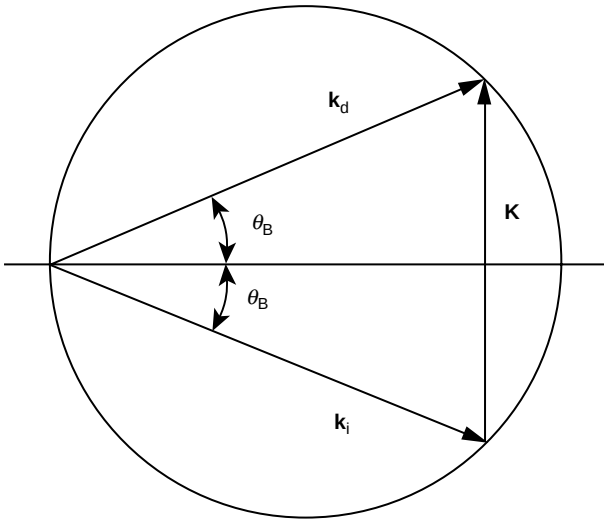

Figure 4 Momentum diagram for diffraction of light.

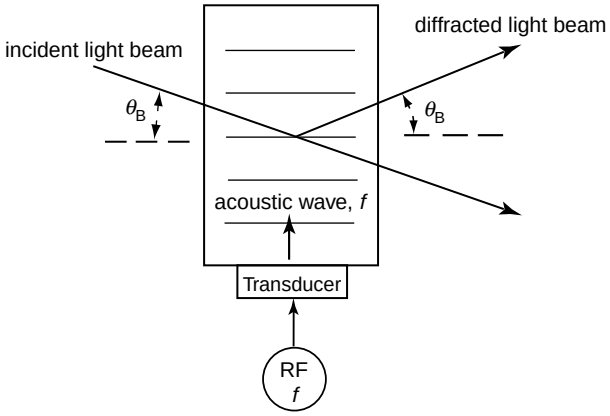

Figure 5 Acousto-optic diffraction cell.

effect and Bragg diffraction are not different phenomena, but are the limits of the same mechanism. The Raman–Nath parameter Q determines which is the appropriate limit for a given set of values λ, Λ, and L.

Anisotropic Bragg Diffraction

In optically anisotropic materials, acousto-optic Bragg diffraction can occur between an ordinary and extraordinary optical beam, and vice versa. This is generally referred to as birefringent diffraction, and it offers additional design capabilities compared to the isotropic case, such as enhancing the angular aperture, and extending the aperture–bandwidth product.

The theory of diffraction of light so far described has assumed that the optical medium is isotropic. A number of important acousto-optic devices make use of the properties of birefringent materials. It is different from diffraction in isotropic media in that the magnitude of the momentum vector of the light will in general be different for different light polarization directions. Thus, the vector diagram representing conservation of momentum will no longer be the simple isosceles triangle of Figure 4. The momentum vector for ordinary polarized light will, in general, be different from the momentum vector for light that is extraordinary polarized. To understand the effect of anisotropy on diffraction, it is necessary to mention another phenomenon which occurs when light interacts with shear acoustic waves, i.e., waves in which the displacement of matter is perpendicular to the direction of propagation of the acoustic wave. A shear acoustic wave may cause the direction of polarization of the diffracted light to be rotated by 90°. The underlying reason for this is that the shear

disturbance induces a birefringence which causes the plane of polarization to be rotated. The acousto-optic tunable filter (AOTF) is a particularly interesting birefringent device. It was first developed for the case of collinear optical and acoustic beams, and was used for wavelength selection in a dye laser.

Diffraction Efficiency

The acoustic power required to diffract light will be determined by the geometric parameters as well as by the properties of the medium. A simplified calculation leads to results that are useful. Referring to the spectrum of a phase-modulated wave shown in Figure 2, we can see that the ratio of the intensity in the first order to that in the zero order is

$$I_1/I_0 = \left[B_1(\Delta\phi)/B_0(\Delta\phi)\right]^2 \qquad [16]$$

By analogy this same result comes about for acousto-optically diffracted light. The acoustic power flow is given by

$$P_a = \frac{1}{2}cVe^2 \qquad [17]$$

where c is the elastic stiffness constant. The elastic stiffness constant is related to the bulk modulus β, the density ρ, and acoustic velocity v through the expression

$$c = \frac{1}{\beta} = \rho V^2 \qquad [18]$$

and the acoustic power density is

$$P_a = \frac{1}{2}\rho V^3 e^2 \qquad [19]$$

The optical phase modulation of the medium resulting from the change in refractive index, Δn, is

$$\Delta\phi = 2\pi(L/\lambda)\Delta n \qquad [20]$$

Using eqn [5] for Δn, the phase modulation is related to the acoustic power density by

$$\Delta\phi = -\pi(L/\lambda)\, n^3 p\, (2P_a/\rho V^3)^{1/2} \qquad [21]$$

The diffraction efficiency can now be calculated by using this value for the optical phase change in eqn [16]. The first- and second-order Bessel functions can be approximated for small modulation index by

$$B_0(\Delta\phi) \sim \cos(\Delta\phi) \sim 1 - (\Delta\phi)^2, \quad \text{and} \quad B_1(\Delta\phi) \sim \sin(\Delta\phi) \sim \Delta\phi \qquad [22]$$

The small signal approximation to the diffraction efficiency is then

$$I_1/I_0 \sim (\Delta\phi)^2 = \left(\pi^2 L^2/2\lambda^2\right)\left(n^6 p^2/\rho V^3\right)P_a \qquad [23]$$

The expression $n^6p^2/\rho V^3$ is known as the figure of merit of the material, and is designated as M_2; it is comprised entirely of intrinsic material properties.

Acousto-optic Materials

Acousto-optic device technology has matured to the point that performance is chiefly limited by material parameters, particularly the figure of merit and acoustic attenuation. Nature has arranged that materials with high figures of merit usually have high attenuation and vice versa. The widely used acousto-optic materials are fused quartz, tellurium dioxide, and lithium niobate. Development work on new infrared materials has been reported recently. A list of commonly used acousto-optic materials is given in Table 1. For materials with a low figure of merit, a higher drive power can be used to obtain the required efficiency.

Experience has indicated that the upper limit for very small devices (active area $\sim$0.1 mm^2) is a drive power density of 100–500 mW/mm^2, provided there is proper heat sinking to transfer the heat energy. For larger devices, sizes greater than $\sim$1 cm^2, the limit is closer to a few W/cm^2. At the high drive power levels, the acoustic attenuation may cause significant optical distortion.

AO Devices

Resolution, bandwidth, and speed are the important characteristics of acousto-optic scanners, shared by all types of scanning devices. Depending upon the application, only one, or all, of these characteristics may have to be optimized. Consider the acousto-optic scanner in Figure 6 with a collimated incident beam of width D, diffracted to an angle θ_B at its RF bandcenter, and whose bandwidth is Δf. If the diffracted beam is focused onto a plane by a lens, or lens combination, at the scanner, the diffraction spread of the optical beam will be

$$\Delta x = F\Delta\phi \sim F - \lambda/D \qquad [24]$$

Table 1 Selected acousto-optic materials

Material	*Transmission range* (μm) mission	*Acoustic mode & propagation direction*	*v* (cm/s × 10^5)	*Acoustic attenuation* (dB/cm GHz^2)[c]	*n*	M_2 ($s^3/g \times 10^{-18}$)
Visible–near-infrared (VIS-NIR)						
$LiNb0_3$	0.04–4.5	L[100][a]	6.57	0.15	2.20	7.0
		S[001][b]	3.59	2.6	2.29	2.92
TiO_2,	0.45–6	L[001]	10.3	0.55	2.58	1.52
α-SiO_2;	0.12–4.5	L[001]	6.32	2.1	1.54	1.48
		L[100]	5.72	3.0	1.55	2.38
TeO_2.	0.35–5	L[001]	4.20	15	2.26	34.5
		S[110]	0.616	90	2.26	793
Far IR						
Ge	2–20	L[111]	5.50	30	4.00	840
$T1_3AsS_4$ (TAS)	0.6–12	L[001]	2.5	29	2.63	510
GaAs	1–11	L[110]	5.15	30	3.37	104
ZnTe	0.55–20	L[110]	3.37	130	2.77	18
GaP	0.6–10	L[110]	6.32	60	3.31	30
$T1_3PSe_4$	0.85–9	L[100]	2.0	150	2.9	2069
Te	5–20	L[100]	2.2	60	4.8	4400
CdS	0.5–11	L[100]	4.17	90	2.44	12
GaP	0.6–10	L[110]	6.32	60	3.31	30
ZnS	0.4–12	L[001]	5.82	27	2.35	3.4
		S[001]	2.63	130	2.35	8.4

[a]Longitudinal acoustic waves propagating along the [100] crystallographic direction.
[b]Shear acoustic waves propagating along the [001] crystallographic direction.
[c]Attenuation constant at 1 GHz; the frequency dependence of the attenuation for most crystals is quadratic.

where F is the focal length of the lens. The light intensity will be distributed in the focal plane with a sinc^2 distribution by aperture diffraction. The number of resolvable spots, N, will be the angular scan range divided by the angular diffraction spread,

$$N = \frac{\Delta\theta}{\Delta\phi} \qquad [25]$$

where $\Delta\theta$ is the range of the angular scan. Differentiating the Bragg angle formula yields

$$\Delta\theta = (\lambda/V \cos\theta_0)\Delta f \qquad [26]$$

and

$$N = \Delta f(D/V \cos\theta_0) = \Delta f \tau \qquad [27]$$

where θ_0 is the Bragg angle at band center, and τ is the time that it takes the acoustic wave to cross the optical aperture. The resulting expression is the time–bandwidth product of the acousto-optic scanner, a concept applied to a variety of electronic devices as a measure of information handling capacity. The time–bandwidth product of an acousto-optic Bragg cell is equivalent to the number of bits of information which may be instantaneously processed by the system. In order to maximize the number of resolution elements, it is desirable to have as large a bandwidth as possible (i.e., large frequency range) and also as large an aperture delay time as possible. There are two factors limiting the bandwidth of an acousto-optic device: the bandwidth of the transducer structure (discussed later), and acoustic absorption in the delay medium. The acoustic absorption increases with increasing frequency; for high-purity single crystals the increase generally goes with the square of the frequency. For glassy materials, on the other hand, the attenuation will increase more slowly with frequency, often approaching a linear function. The maximum frequency is generally taken as that for which the attenuation of the acoustic wave across the optical aperture is equal to

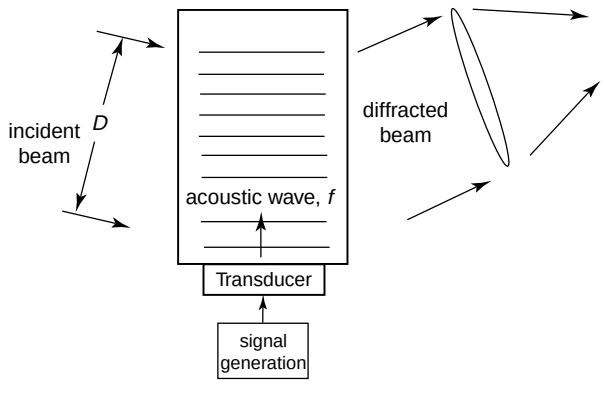

Figure 6 Acousto-optic scanner.

3 dB. A reasonable approximation of the maximum attainable bandwidth is

$$\Delta f = 0.7 f_{\max} \qquad [28]$$

Birefringent Scanners

The birefringent scanner is a significant use for anisotropic Bragg diffraction. There are a number of advantages with the birefringent scanner over the anisotropic scanner, such as a larger angular scan range along with lower frequencies. There are also disadvantages, such as lower speed due to generally lower shear wave velocities, and the particular application will dictate whether an isotropic or birefringent scanner is better. Scanners represent a fairly important aspect of acousto-optic devices due to the widespread commercial use of laser beam deflectors for displays and laser printers. For applications where the required speed is beyond that of mechanical scanners, the acousto-optic scanner is an ideal candidate. However, unlike mirrors, acousto-optic deflectors are wavelength sensitive, and can only be used with single-wavelength laser beams.

The birefringent scanner can be described with the wavevector diagram, as shown in Figure 7 for a positive uniaxial crystal where the extraordinary index of refraction is larger than the ordinary. The acoustic wavevector is tangent to the diffracted surface, which produces the largest scan angle for a given acoustic bandwidth. This is also the degenerate case, where only a single diffracted beam results, whereas two beams at two different acoustic frequencies will generally result from arbitrary input and acoustic propagation directions. The azimuthal acceptance angles (angles normal to the plane of incidence) of the acoustic and optical wavevectors can be different, although propagation and polarization directions must be selected that will produce high efficiency.

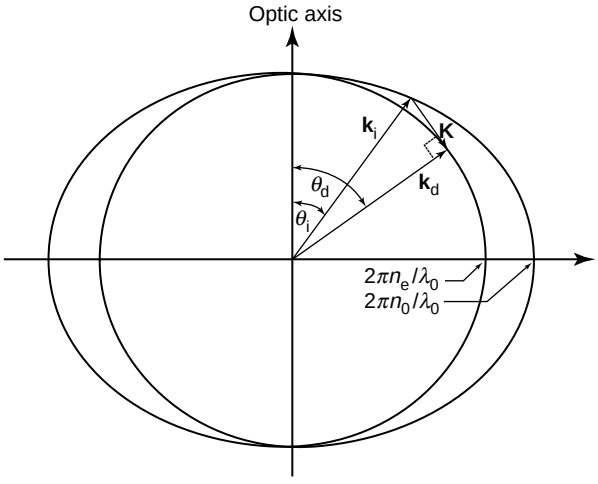

Figure 7 Wavevector diagram for birefringent scanner using a positive uniaxial crystal.

For the positive uniaxial case, the acoustic and diffracted wavevectors will be perpendicular at the design point, which allows the center frequency to be calculated from geometry as

$$f_0 = (V/\lambda)(n_i^2 - n_0^2)^{1/2} \qquad [29]$$

where n_i is the index of refraction at the incidence angle. Due to the typically low velocity of shear waves, this frequency can be much lower than an isotropic scanner, and the incidence angle can also be chosen to adjust the frequency. It is also possible to use the optical activity of certain materials such as TeO_2, along with circularly polarized light to further reduce the acoustic frequency. It is important to maintain as low a frequency as possible due to acoustic attenuation, which is especially high with the soft materials typically used for birefringent applications.

The bandwidth over which the scanner can efficiently operate is fairly large due to having the acoustic and diffracted wavevectors perpendicular, and is approximately $\Delta f = 2f_0$, assuming an octave of bandwidth for the isotropic scanner. This will produce a larger scan angle than an isotropic scanner, and more spots of resolution. The number of resolution spots is determined through diffraction by the aperture size and diffraction, along with the scan angle. Since a larger aperture requires a longer time for the acoustic waves to propagate across the aperture, the response time τ to access the scanner will increase, and the product of τ and f is related to the number of spots by eqn [27]. The advantage of the birefringent scanner is that it can operate at a lower frequency f_0 with better performance. However, since other factors such as acoustic attenuation are important in designing a scanner, for some applications it might be better to operate an isotropic scanner at a higher frequency.

AOTFs

With anisotropic Bragg diffraction, the magnitude of the diffracted wavevector $\mathbf{k}_d$ will differ from the incident wavevector $\mathbf{k}_i$, which cannot occur for the isotropic case. This is illustrated in Figure 8 for an AOTF utilizing a negative uniaxial crystal where an extraordinary input wave propagating at an incident angle θ_i relative to the crystal axis is diffracted into an ordinary output wave at an angle θ_d. This occurs through an acoustic wave propagating at an angle θ_a

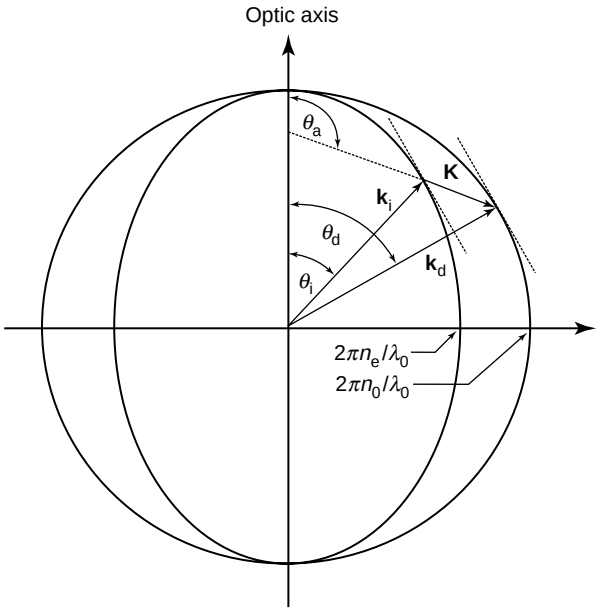

Figure 8 Wavevector diagram for an AOTF using a negative uniaxial crystal.

with wavevector $\mathbf{K}_a$, where all the wavevectors lie in a plane through the optic axis of the crystal. This diagram is identical to the index ellipsoid for the crystal, scaled by $2\pi/\lambda$, where λ is the vacuum wavelength. The acoustic wavevector is shown as adding to the incident wavevector, which increases the frequency of the optical wave by the acoustic frequency. It can also be represented in the reverse direction, which would reduce the optical frequency by the same amount.

An AOTF spectrally filters the optical input, and maintains its spectral purity over a large angular aperture. These conditions are maximized when the power flow or ray directions of the input and output beams are collinear. This produces a parallel-tangents condition, where lines drawn tangent to the wavevector surfaces and connecting the input and diffracted wavevectors are parallel. For beams at 90° to the optical axis, it is referred to as the collinear case, whereas all other parallel-tangents conditions are noncollinear. The diffracted beam is rotated by 90° during anisotropic Bragg interaction, so that crossed polarizers can be used to separate the input and diffracted beams. For the collinear case polarization separation must be used, whereas angular separation can also be used for the noncollinear case.

The AOTF design equations can be derived from the geometrical conditions imposed by the wavevector matching condition

$$\mathbf{k}_d = \mathbf{k}_i + \mathbf{K} \qquad [30]$$

along with decomposing the acoustic wave into its Fourier components which result from the finite interaction length produced by the transducer. These components allow for phase matching over a range of input angles, and they also allow for a spectral spread of the interaction. The full width half maximum (FWHM) resolution is given by

$$\Delta\lambda = \frac{1.8\lambda^2}{bL\sin 2\theta_i} \qquad [31]$$

where λ is the vacuum wavelength, b is the dispersion term (essentially $2\pi\Delta n$, where the birefringence Δn is the difference between the ordinary index n_0 and extraordinary index n_e of refraction), L is the interaction length of the input beam defined geometrically by the acoustic beam, and θ_i refers to the angle of the central or cardinal ray of the input beam, and the AOTF is designed around this angle. It is therefore a sensitive function of the incidence design angle, with the highest resolution occurring for the collinear case. The resolution is also narrower for larger birefringence.

The geometry of a noncollinear AOTF is shown schematically in **Figure 9**. The acoustic waves propagate in the correct direction and are generated by a transducer of length L_t, which is related geometrically to the interaction length. These waves are generally absorbed at the other side of the AOTF to prevent interfering reflections, and the sides may also be wedged to eliminate reflections into the interaction region. The optical beam enters the AOTF at the correct angle to the crystal optic axis, and either diffracts or passes through as the zero order. The input beam will have an angular spread, producing an acceptance angle that is a function of the interaction length, wavelength, and crystal parameters, and the external solid angular aperture

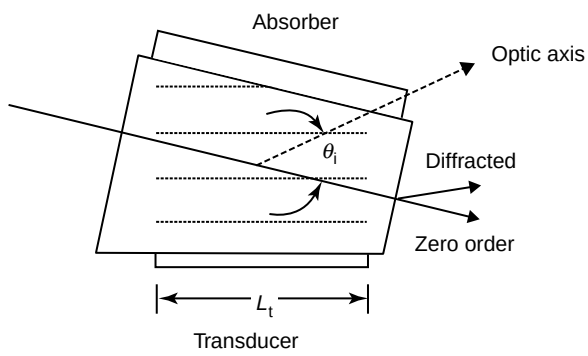

Figure 9 Noncollinear AOTF orientation of optical and acoustic beams.

is roughly given by

$$\Delta\Omega = \frac{\pi n^2 \lambda}{\Delta n L} \qquad [32]$$

where n refers to the diffracted beam index of refraction, which can be either n_0 or n_e depending upon the design. The resolution and solid angle are therefore related through the transducer length, and the product of the resolving power times the solid angle is given by

$$(\lambda/\Delta\lambda)(\Delta\Omega) = 2\pi n^2 \sin^2\theta_i / |F_\theta F_\varphi|^{1/2} \qquad [33]$$

where

$$F_\theta = 2\cos^2\theta_i - \sin^2\theta_i \qquad [34a]$$

$$F_\varphi = 2\cos^2\theta_i + \sin^2\theta_i \qquad [34b]$$

are form factors for the polar and azimuthal components of the optical field. Since all the wavectors lie in a plane containing the optic axis, the azimuthal components of the incident, diffracted, and acoustic wavevectors must always be identical under the parallel-tangents condition.

The product $(\lambda/\Delta\lambda)(\Delta\Omega)$ forms a figure of merit for spectrometers, and for the collinear case, the product is identical to a Fabry–Perot etalon having an index of n, and at other angles it is more complicated due to the angular dependence. The angular field is also symmetrical for the collinear AOTF, but nowhere else other than along the optic axis, where the resolving power becomes zero.

The great advantage of the AOTF designed under the parallel-tangents condition is the large angular aperture compared to the general case. Under this condition, the dependence of the angular aperture on resolving power is second order rather than first order, and the angular aperture can be tens of degrees for a typical resolution, which is useful for imaging applications. At the special angle of 54.7°, the second-order dependence is also zero, and the aperture can be extremely large relative to the resolving power, although this requires a high-resolution device with a corresponding narrow field, since the condition would soon be violated at larger field angles.

The efficiency of the AOTF under phase matching is given by

$$\eta = \sin^2[\pi^2 M_2 P_a L^2/(2\lambda^2)]^{1/2} \qquad [35]$$

which depends on θ_i and various material properties. Since the interaction is anisotropic, M_2 must be taken as a tensor quantity, in which both the polarization and propagation directions of the acoustic and optical waves must be accounted for. For acoustic waves, the polarization direction is the particle motion, which is perpendicular to the propagation direction for shear waves. In general, M_2 is much larger for a specific configuration of a particular material, and AOTFs are designed around this condition. The range of useful design angles is also usually limited since M_2 is generally a sensitive function of θ_i. The angular dependence on M_2 is important in designing an efficient device since it can be near zero at specific design angles for some materials, such as with TeO_2 for the collinear case.

The AOTF must be designed to optimize performance. The resolution is generally given as a system requirement, and the design angle along with the transducer length must be adjusted to optimize the throughput, or total optical power through the AOTF. This requires maximizing the efficiency, angular aperture, and aperture dimension. The maximum crystal size that can be grown ultimately limits these parameters, both in the interaction length and in the aperture size.

AOTFs have been used for a wide range of spectral filtering applications for spectroscopy and laser applications. Both collinear and noncollinear devices have electronically tuned a variety of lasers, including dye, semiconductor, and Ti:sapphire lasers. The AOTF is placed within the laser cavity, which requires high transmission on the laser line, with enough resolution to separate nearby laser transitions. A related application is the multiplexing of optical communications systems, where the AOTF is used to separate the various laser diode wavelengths. Perhaps the most widely used application is spectroscopy, and single-point detection systems have been used for a variety of biological and chemical applications in the visible and infrared. These techniques have been extended to spectral imaging, again in the visible and the infrared. Due to a variety of AOTF aberrations the image quality will be somewhat degraded. Since the AOTF is electronically tunable, it can be rapidly modulated both in amplitude and wavelength, which allows for modulation spectroscopy to detect small signals in a large background. By applying multiple frequencies, the AOTF has also been used as a rejection filter, in which all wavelengths other than that selected are allowed to pass.

Modulators

Acousto-optic modulators are used to vary and control laser beam intensity. A Bragg configuration gives a single first-order output beam, where intensity is directly linked to the power of RF control signal. The rise time of the modulator is the time required for

the acoustic wave to traverse the laser beam. For minimum rise time the laser beam will be focused down, forming a beam waist as it passes through the modulator.

A Bragg configuration gives a single first-order output beam, the intensity of which is controlled by the RF signal power, and diffaction angle controlled by the drive frequency. Two deflectors can be used in series and at right angles to each other to give full two-dimensional scanning.

One requirement in the design of acousto-optic modulators is that the diffracted beam and undiffracted beam must be well separated. For an adequate extinction ratio, the Bragg angle should be at least as large as the divergence of the optical beam. This condition puts a minimum value on the center frequency. Equating the Bragg angle

$$\theta_B = \frac{\lambda f_0}{2nV} \qquad [36]$$

and the diffraction angle of the Gaussian beam

$$\delta\theta_0 = \frac{4\lambda}{\pi(nd)} \qquad [37]$$

where d is the beam diameter at the minimum, it follows that the lower limit of the acoustic frequency is given by

$$f_0 = \frac{8}{\pi\tau} \qquad [38]$$

We may combine these expressions to determine a limit of modulator bandwidth of acousto-optic modulators, with the result

$$\Delta f \sim \frac{1}{4} f_m \qquad [39]$$

i.e., the modulation bandwidth is approximately equal to 25% of the midband acoustic frequency, f_m. In view of the present status of transducer technology, the modulation bandwidth of acousto-optic modulators is limited to several hundred MHz.

A frequency shifter uses the shift inherent in the acousto-optic interaction to up- or down-shift a laser's frequency. Two kinds of shifters can be distinguished: the fixed frequency shifter, and the variable frequency shifter. The frequency shift is equal to the signal frequency applied to the transducer. Various device configurations can be used to shift the laser beam, such as multiple travels inside the shifter to double or triple the frequency shift, or a combination of two frequency shifters in series. Acousto-optic frequency shifters can be used, for example, in optical heterodyning and interferometry, or in laser Doppler velocimetry for particle velocity analysis.

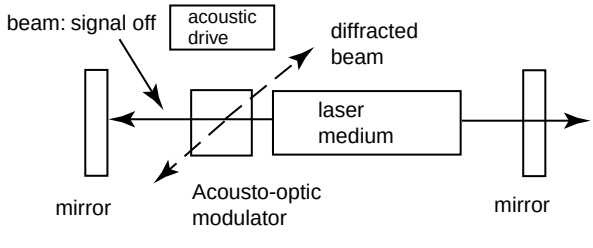

Figure 10 An acousto-optic Q-switch in a laser cavity.

Q-Switches

A Q-switch is a special modulator which introduces high repetition rate losses inside a laser cavity (typically 1 to 100 KHz). They are designed for minimum insertion loss and to be able to withstand very high laser powers. In normal use an RF signal is applied to diffract a portion of the laser cavity flux out of the cavity. This increases the cavity losses and prevents oscillation. When the RF signal is switched off, the cavity losses decrease rapidly and an intense laser pulse evolves. A diagram of a typical acousto-optic Q-switched laser cavity is shown in Figure 10. Q-switches are the preferred method of intracavity modulation where high speed, high stability, and modest cost are important design factors.

List of Units and Nomenclature

b	dispersion constant	–
c	elastic stiffness	dynes cm^{-2}
D, d	optical beam diameters	cm
e	elastic strain	–
f	acoustic frequency	hertz
F	optical focal length	cm
F_ϕ, F_θ	angular form factors	–
$\mathbf{k}$	optical momentum wavevector	cm^{-1}
$\mathbf{K}$	acoustic momentum wavevector	cm^{-1}
M_2	acousto-optic figure of merit	sec^3 gram^{-1}
n	refractive index	
p	photo-elastic constant	–
P	pressure	dynes cm^{-2}
P_a	acoustic power	watts cm^{-2}
V	acoustic velocity	cm sec^{-1}
β	compressibility	cm^2 dyne^{-1}
η	diffraction efficiency	–
ε	dielectric constant	–
ω	optical frequency	hertz
Ω	solid angle	steradians
ρ	density	grams cm^{-3}
θ	Bragg angles	radians
τ	acoustic wave travel time	sec

See also

Modulators: Acousto-Optics. **Optical Amplifiers:** Raman, Brillouin and Parametric Amplifiers. **Scattering:** Raman Scattering. **Spectroscopy:** Raman Spectroscopy.

Further Reading

Chang IC (1976) Acousto-optic devices and applications. *IEEE Transactions: Sonics and Ultrasonics* SU-23: 2–22.

Chang IC (1981) Acousto-optic tunable filter. *Optical Engineering* 20: 824–829.

Dixon RW (1967) Photo-elastic properties of selected materials and their relevance for applications to acoustic light modulators and scanners. *Journal of Applied Physics* 38: 5149–5153.

Gordon EI (1966) A review of acousto-optical deflection and modulation devices. *Proceedings of the IEEE* 54: 1391–1401.

Gottlieb M, Ireland CLM and Ley JM (1983) *Electro-optic and Acousto-optic Scanning and Deflection*. New York: Marcel Dekker.

Goutzoulis AP and Pape DR (1994) *Design and Fabrication of Acousto-optic Devices*. New York: Marcel Dekker.

Uchida N and Niizeki N (1973) Acousto-optic deflection materials and techniques. *Proceedings of the IEEE* 61: 1073–1092.

Xu J and Stroud R (1992) *Acousto-optic Devices: Principles, Design, and Applications*. New York: Wiley.

Electro-Optics

L R Dalton, University of Washington, Seattle, WA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

The electro-optic (EO) effect can be described as any one of a number of phenomena that occur when an electromagnetic wave in the optical spectrum (e.g., characterized by wavelengths in the range 200 to 2000 nm) interacts with an electric field, or with matter under the influence of an electric field. Two of the most important electro-optic phenomena are the Kerr effect (discovered by John Kerr in 1875) and the Pockels effect (discovered by Friedrich Pockels in 1893), in which birefringence is induced or modified in a liquid (the Kerr effect) or a solid (the Pockels effect). Birefringence is the difference in refractive indices for light of orthogonal line polarizations, one of which is parallel to the induced optical axis. The Kerr effect involves creation of birefringence in a liquid that is otherwise not birefringent. The degree of birefringence is quadratically proportional to the applied electric field strength. Hence, the Kerr effect is frequently referred to as the quadratic electro-optic effect while the Pockels effect involves a linear dependence on the applied electric field and is referred to as the linear electro-optic effect. In considering practical applications of the electro-optic effect, a commonly encountered term is that of 'electro-optic modulator', which can be described as a device wherein a signal control element is used to modulate a beam of light. The control element is typically an electric field with frequency components in the zero (DC) to hundreds of gigahertz (1 GHz = 10^9 Hz) range (or even tens of terahertz (1 THz = 10^{12} Hz). The modulation may be imposed on the phase, frequency, amplitude, or direction of the modulated optical beam.

Electro-optic effects are one class of second-order nonlinear optical phenomena. Other important second-order nonlinear optical phenomena include sum (e.g., second-harmonic) and difference frequency generation. Such phenomena derive from the second-order term in the power series expansion of macroscopic (material) polarization, P, in terms of applied electric fields:

$$P_i = \chi_{ij}^{(1)} E_j + \chi_{ijk}^{(2)} E_j E_k + \chi_{ijkl}^{(3)} E_j E_k E_l + \cdots \qquad [1]$$

where $\chi_{ij}^{(1)}$, $\chi_{ijk}^{(2)}$, and $\chi_{ijkl}^{(3)}$ are the linear, second-, and third-order optical susceptibilities, respectively. Second-harmonic generation (SHG) or frequency doubling (2ω) effects, where a beam at twice the frequency of the input beam is generated, can be seen by substituting a sinusoidal field, $E_\omega \cos(\omega t - kz)$; for E_j and E_k. After using a well-known trigonometric identity, the equation for macroscopic polarization becomes (through second-order and dropping the subscript i):

$$P = \chi^{(1)} E_\omega \cos(\omega t - kz) + (1/2)\chi^{(2)} E_\omega^2 \times [1 + \cos(2\omega t - 2kz)] + \cdots \qquad [2]$$

The electro-optic effect can be appreciated by considering interaction of the medium with an optical field $E_\omega \cos(\omega t - kz)$ and a low-frequency field E_0. Note that although we have used the symbol zero to

denote the frequency of the low-frequency electrical field, actual frequencies can extend to hundreds of gigahertz or even tens of terahertz. Again substituting (and now keeping terms to third order), the equation for polarization can be written as:

$$\begin{aligned} P &= \chi^{(1)}E_\omega\cos(\omega t-kz)+2\chi^{(2)}E_0E_\omega\cos(\omega t-kz) \\ &\quad +3\chi^{(3)}E_0^2E_\omega\cos(\omega t-kz)+(3/4)\chi^{(3)}E_\omega^3\cos(\omega t-kz) \\ &= \chi_{\text{eff}}E_\omega\cos(\omega t-kz) \end{aligned} \quad [3]$$

The equation for the nonlinear index of refraction, n, can be expressed as

$$\begin{aligned} n^2 &= 1+4\pi\chi_{\text{eff}} \\ &= 1+4\pi[\chi^{(1)}+2\chi^{(2)}E_0+3\chi^{(3)}E_0^2+(3/4)\chi^{(3)}E_\omega^2] \end{aligned} \quad [4]$$

Denoting the linear index of refraction as n_0, the above equation can be rewritten as

$$n^2-n_0^2=8\pi\chi^{(2)}E_0+12\pi\chi^{(3)}E_0^2+3\pi\chi^{(3)}E_\omega^2 \quad [5]$$

which in turn leads to

$$\begin{aligned} n &= n_0+(4\pi/n_0)\chi^{(2)}E_0+(6\pi/n_0)\chi^{(3)}E_0^2 \\ &\quad +(3\pi/2n_0)\chi^{(3)}E_\omega^2 \end{aligned} \quad [6]$$

The definition of light intensity in cgs units is $E_\omega^2=(8\pi/cn_0)I(\omega)$, which when substituted into the above equation gives

$$\begin{aligned} n &= n_0+(4\pi/n_0)\chi^{(2)}E_0+(6\pi/n_0)\chi^{(3)}E_0^2 \\ &\quad +(12\pi/cn_0^2)\chi^{(3)}I(\omega) \end{aligned} \quad [7]$$

The index of refraction can now be written as

$$n(\omega)=n_0(\omega)+n_1(\omega)E_0+n_2(0)E_0^2+n_3(\omega)I(\omega) \quad [8]$$

where the terms $n_1(\omega)$, $n_2(0)$, $n_3(\omega)$ correspond to the linear electro-optic effect, the quadratic electro-optic effect, and the optical Kerr effect, respectively. The above equation can also be expressed as

$$n=n_0-(1/2)n^3rE_0-(1/2)n^3sE_0^2-\cdots \quad [9]$$

where r and s are the linear and quadratic electro-optic coefficients, respectively.

Unfortunately, a great deal of confusion exists concerning the use of the terms electro-optic effect and electro-optic modulator. First of all, the terms 'electro-optic' and 'opto-electronic' are frequently confused and used interchangeably. Opto-electronic devices function as electrical-to-optical or optical-to-electrical signal transducers. Examples of commonly used opto-electronic devices include light-emitting diodes (LEDs), photodiodes (PDs), injection laser diodes (ILDs), and integrated optical circuit (IOC) elements. An electro-optic device can also function as an electrical-to-optical signal transducer (e.g., a Mach–Zehnder interferometer or a birefringent modulator employing polarizers at the input and the output); however, the mode of operation is entirely different for electro-optic and opto-electronic devices. For example, with LEDs or modulated lasers, the applied electric field produces modulation of transmission of light by altering the excited state population of the light emitting (lasing) state of the material. With electro-optic materials, no actual excited state population is generated. Electro-optic devices typically operate in regions of the optical spectrum removed from resonant transitions, i.e., regions of relatively high transparency. The applied electric field acts only to perturb the charge distribution of the material (the spatial positions of the electrons and nuclei). The altered charge distribution interacts with the optical beam transmitting the material with the result that the speed of light in the material (i.e., the index of refraction or birefringence) is altered. The response times of opto-electronic and electro-optic phenomena (and hence bandwidths of devices exploiting these phenomena) are quite different. In the former case, response will be limited by the lifetime of the relevant (emitting) excited state while in the latter case, response will be defined by the relaxation time of the material (e.g., reorientation time of a liquid or lattice relaxation time of a solid) following removal of the perturbing electric field.

An even greater confusion can arise due to the jargon used in particular technologies, such as telecommunications. A telecommunications engineer frequently refers to 'electro-optic switching' when describing opto-electronic transduction of an optical signal to an electrical signal, followed by re-routing in the electrical domain, and finally opto-electronic transduction of the electrical signal back to the optical domain. Electro-optic switching (in the sense used in this article) involves quite a different operation and would be described by that same telecommunications engineer as 'all-optical switching' – a term reserved by physicists for the optical Kerr effect (control of one light beam by another light beam).

A second point of confusion involves distinguishing between the terms 'electro-optic' and 'electro-absorptive' modulation. Again, the terms are frequently used interchangeably. However, they involve quite different physical mechanisms. An electro-absorptive

modulator is, like a modulated laser or LED, a 'resonant' device. In a device such as a gallium arsenide (GaAs) or indium phosphide (InP) quantum dot electro-absorptive modulator, the position and width of an optical absorption (resonant transition) are defined by the physical dimensions of the quantum dot. Application of an electric field shifts the spectral wavelength of the quantum dot absorption. This phenomenon can be used to dramatically change optical transmission through (or equivalently, absorption by) the material by choosing the wavelength of the propagating beam of light to be near the resonant absorption. With the electric field off, the material is reasonably transparent but becomes strongly absorbing when the electric field is applied, shifting the resonant absorption to the wavelength of the propagating beam. Obviously, the voltage required to achieve a desired change in optical transmission will depend on control of quantum dot dimensions in fabrication, which has been and continues to be a topic of research focus for electro-absorptive materials. Likewise, control of insertion losses of such devices is a concern as the wavelength used for the transmissive state of the device must be sufficiently close to the optical resonance to achieve the desired attenuation with application of a modest control voltage. Another issue to be faced with electro-absorptive devices is that of 'chirp' (optical signal distortion arising from the fact that both absorption and index of refraction – the imaginary and real parts of the optical susceptibility – change with application of the electric control field). On the positive side, electro-absorptive modulators are 'polarization-insensitive' modulators and are thus conveniently used with multimode (as well as single mode) optical transmission.

Further confusion exists because widely different types of materials can be used in electro-optic and electro-absorptive devices. As these materials are frequently competing for the same applications (signal transduction, optical switching, etc.), the advantages and disadvantages of different materials are often compared without maintaining the context that quite different phenomena are involved. Even restricting our discussion to electro-optic materials, we note that materials can range from organic liquid crystalline materials (nematic, smetic, ferroelectric, etc.) to organic electro-optic materials (both crystalline and 'macromolecular') to inorganic crystalline materials (lithium niobate, $LiNbO_3$; barium titanate, $BaTiO_3$; barium borate, BBO; potassium dihydrogen phosphate, KDP; lithium tantalate, $LiTaO_3$; zinc telluride, ZnTe, etc.). The response times of these materials to transient application of an applied electric field will be quite different, reflecting the different types of lattice motion involved. With liquid crystalline materials, particularly nematic materials, considerable molecular reorientation is involved and due to the mass that must be moved, response times are quite slow although the index change will be relatively large. With use of liquid crystalline materials, device bandwidths will typically be limited to tens of megahertz (MHz) or less. In contrast, conjugated π-electron organic materials typically exhibit response times of tens to hundreds of femtoseconds, which translates to potential device bandwidths of tens of terahertz (actual device bandwidths may be less due to factors other than material response). For conjugated π-electron organics, the response time is the phase relaxation time of the π-electron system. Because only a slight change in bond length alternation of the conjugated π-electron system occurs with application of an applied electric field and because strong electron–phonon coupling and resonance stabilization of the π-electron system act to reduce the barrier to lattice relaxation, the response times of π-electron organic materials are among the fastest observed in nature. With crystalline inorganic materials, the ionic constituents move to new locations with application of an electric field with the exact movement determined by the field strength, the charge on the ions, and the restoring force. Unequal restoring forces along the three mutually orthogonal (perpendicular) axes of the crystal lead to anisotropy in the optical properties of the material. The applied electric field will induce a change in the anisotropy (the principal refractive indices). The symmetry of the electro-optic tensor will be closely related to the symmetry of the piezoelectric tensor. The linear electro-optic effect requires that the crystal exhibit noncentrosymmetric (acentric or ferroelectric) symmetry. A centrosymmetric crystal possesses a center of symmetry defined by identical particles existing in the lattice at vectors $\mathbf{r}$ and $-\mathbf{r}$, where $\mathbf{r}$ is a position vector measured from an arbitrary origin. A centrosymmetric crystal, like an isotropic liquid or gas, can exhibit a quadratic electro-optic effect.

For the sake of completeness, it can also be noted that index of refraction changes can be induced by acoustic waves (acousto-optic modulation) and by heating (thermo-optic modulation). Also, elasto-optic and photo-elastic effects can produce index of refraction changes.

To keep this article to a manageable length, discussion is limited to solid-state electro-optic materials. Design principles being used to produce new organic electro-optic materials will be

illustrated. Since inorganic materials exist as crystals, limited chemical modification of such materials is possible, although processing techniques for fabricating thin films of such crystalline materials have been developed in a number of cases. For organic macromolecular materials, development of new materials is an on-going activity. Several representative devices being fabricated, using electro-optic materials, will be discussed. Four of the most commonly encountered types of EO devices include: (i) stripline devices; (ii) prism, cascaded prism, and superprism devices; (iii) resonated devices, such as ring microresonator and photonic bandgap devices, and (iv) waveplates. Of course, devices such as stripline devices can be configured in a variety of ways to produce polarization-sensitive and polarization-insensitive electrical-to-optical signal transducers, optical switches either using simple 1×2 or 2×2 switch architectures or multimode interference (MMI) switches, optical gyroscopes, photonically controlled phased array radar systems, spectrum analyzers, optical digital signal processors, analog-to-digital (A/D) and digital-to-analog (D/A) signal converters, electromagnetic (radiofrequency to millimeter wave) signal generators, voltage sensors, etc. Prism-type devices are typically used for spatial light modulation or laser beam deflection. Ring microresonator devices afford a rich array of applications ranging from active wavelength division multiplexing (WDM) to active optical interconnect reconfiguration, to voltage-controlled wavelength tuning of laser outputs.

Basic Relationships

The effect of the applied field is to deform the index of refraction ellipsoid, which can be represented as

$$\left(\frac{1}{n^2}\right)_1 X^2 + \left(\frac{1}{n^2}\right)_2 Y^2 + \left(\frac{1}{n^2}\right)_3 Z^2 + 2\left(\frac{1}{n^2}\right)_4 YZ + 2\left(\frac{1}{n^2}\right)_5 XZ + 2\left(\frac{1}{n^2}\right)_6 XY = 1 \qquad [10]$$

leading to the corrections

$$\Delta\left(\frac{1}{n^2}\right)_i = \sum_{j=1}^{3} r_{ij} E_{0j} \qquad [11]$$

where the electro-optic tensor components are related to the tensor components of the second-order nonlinear optical susceptibility by $r_{ij} = -(8\pi/n_{0i}^2 n_{0j}^2)\chi_{ji}^{(2)}$. The full tensorial equation for change of the 'indicatrix' (index of refraction ellipsoid) can be expressed as

$$\begin{pmatrix} \Delta\left(\frac{1}{n^2}\right)_1 \\ \Delta\left(\frac{1}{n^2}\right)_2 \\ \Delta\left(\frac{1}{n^2}\right)_3 \\ \Delta\left(\frac{1}{n^2}\right)_4 \\ \Delta\left(\frac{1}{n^2}\right)_5 \\ \Delta\left(\frac{1}{n^2}\right)_6 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{pmatrix} \begin{pmatrix} E_{0X} \\ E_{0Y} \\ E_{0Z} \end{pmatrix} \qquad [12]$$

The electro-optic tensor reduces considerably for specific materials reflecting the symmetry of the particular material. For inorganic crystalline materials, the electro-optic tensor can be reasonably complex (containing many nonzero elements). It is also helpful to note that for crystalline inorganic materials, the electric-field-induced change in crystal shape leads to strain and the orientation of the indicatrix is altered, leading to an additional contribution to the change in index coefficients:

$$\Delta(1/n^2)_i = \sum_k p_{ik} S_k + \sum_j r_{ij} E_j \qquad [13]$$

where S_k is a component of the strain and p_{ik} is the elasto-optic tensor. At high frequencies, the inertia of the crystal prevents macroscopic straining so that the first term of the above equation vanishes. At low frequencies, elasto-optic effects cannot be ignored. Because the deformation leading to strain is generally caused by the inverse piezoelectric effect, it is linearly related to the applied electric field (the same dependence as the linear electro-optic effect). This can lead to a frequency dependence of the 'effective' electro-optic coefficient for crystalline materials.

A brief comment on the photo-elastic effect (the change in index coefficients produced directly by applied stress) is warranted. This effect has the form:

$$\Delta(1/n^2)_i = \sum_l \pi_{il} \sigma_1 \qquad [14]$$

where σ_l are the components of the stress and π_{il} are the piezo-optical coefficients. Note that this effect is independent of the applied electric field.

For axially symmetric 'charge-transfer type' organic chromophores prepared by deposition or

electric field poling, only two unique tensor elements, r_{33} and r_{13}, survive:

$$\tilde{r} = \begin{pmatrix} 0 & 0 & r_{13} \\ 0 & 0 & r_{13} \\ 0 & 0 & r_{33} \\ 0 & r_{13} & 0 \\ r_{13} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad [15]$$

The electro-optic effect for organic macromolecular materials derives from molecules with conjugated π-electrons confined within molecules (chromophores). The electro-optic tensor elements can be related to molecular first hyperpolarizability, β, and the chromophore number density, N (molecules/cm^3), by

$$\begin{aligned} r_{33} &= -2Nf(0)\beta\langle\cos^3\theta\rangle/n_0e^4 \\ r_{13} &= -Nf(0)\beta\langle\sin^2\theta\cos\theta\rangle/n_{0o}^2n_{0e}^2 \end{aligned} \qquad [16]$$

The expressions in brackets are 'order parameters' describing the degree of noncentrosymmetric order. $\langle\cos^3\theta\rangle = 1$ corresponds to all 'dipolar' (charge transfer type) molecules pointing in the same direction (perfect noncentrosymmetric or ferroelectric order) while $\langle\cos^3\theta\rangle = 0$ corresponds to complete disorder and no electro-optic effect. The reason that three angles appear in the order parameter is that an angle is required to represent the principal symmetry axis for the chromophore, the principal symmetry axis of the optical field, and the principal symmetry axis of the electric control field. Thus, averaging (denoted by brackets) must be carried out over each of these three angles. The tensor element r_{33} corresponds to the electric control field applied along the principal axis of the ordered chromophores and the principal axis of the optical field vector, transverse magnetic (TM) polarized light. The tensor element r_{13} corresponds to the control electric field applied along the principal axis of the optical field, transverse electric (TE) polarized light, and orthogonal to the principal chromophore axis. The notation r_{33} is contracted notation for r_{333}. The factor $f(0)$ takes into account the dielectric nature of the medium into which the chromophores (molecules with large first hyperpolarizability) are embedded.

For organic electro-optic materials, elasto-optic effects do not appear to make significant contributions. Moreover, electro-optic activity can be systematically improved by improving β and by optimizing the product of chromophore number density and order parameter. Due to the presence of intermolecular electrostatic interactions, W, among chromophores, order parameters and number density are not independent, e.g., for materials that are prepared by electric field poling the order parameter relevant to the principal component of the electro-optic tensor can be expressed approximately as

$$\langle\cos^3\theta\rangle = L_3(\mu f(0)E_p/kT)[1 - L_1(W/kT)^2] \qquad [17]$$

where L_3, and L_1 are Langevin functions, μ is the chromophore dipole moment, E_p is the strength of the electric poling field, k is the Boltzmann constant, and T is the Kelvin poling temperature. W is a quadratic function of chromophore number density, N. The above equation indicates that a maximum will be observed in the graph of r_{33} versus N.

Materials

With inorganic and organic crystalline materials, very little can be done to optimize electro-optic activity other than discovering new EO crystalline materials or employing isotopic (e.g., deuterium for hydrogen) or ion substitution with existing materials.

As with organic liquid crystalline materials, the electro-optic activity of macromolecular organic second-order nonlinear materials can be systematically improved by molecular modification. As shown in Figure 1, quantum mechanical calculations can provide useful guidance as to which structural modifications will lead to improvements in molecular (first) hyperpolarizability and ultimately electro-optic activity. The molecular hyperpolarizability (long wavelength limit) can be increased by a factor of two by simple variation of the acceptor structure. The calculated values of wavelengths for the interband charge transfer transitions are 390 nm, 403 nm, and 430 nm. Since dipole moments do not change significantly with structure, intermolecular electrostatic interactions should be similar for these three chromophores and the improvement in β should translate to an improvement in electro-optic activity. This theoretical prediction has been experimentally verified. This is just one example of use of quantum mechanics to guide the improvement of electro-optic activity for organic materials. This figure also illustrates the typical structure of an organic electro-optic chromophore, which consists of an electron donor region (an amine donor in the example shown), a connective

conjugated π-electron bridge, and an electron deficient electron acceptor region. The theoretically predicted (by density functional theory, DFT, calculations) variation of molecular first hyperpolarizability with molecular structure has been verified for the structures shown. However, the critical point is that electro-optic activity of organic materials (currently exhibiting a maximum value of 182 pm/V at 1.3 microns wavelength) is likely to be improved for some time to come. This value is to be compared to a value of 32 pm/V for lithium niobate. Another factor influencing the value of macroscopic electro-optic activity realized for electrically poled organic materials is the effect of intermolecular electrostatic interactions on optimization of $N\langle\cos^3\theta\rangle$. Statistical mechanical calculations can provide guidance as to how this is achieved (see **Figure 2** where the role of chromophore shape is illustrated). Approximating the chromophore shape by a rigid prolate ellipsoid with embedded dipole moment yields nearly quantitative reproduction of the experimental data. If the π-electron structure (dipole moment) is left unchanged but the shape of the chromophore is altered to a spherical shape, a significant improvement in electro-optic activity is predicted. The shape effect arises because there are two components to the electronic dipole–dipole intermolecular electrostatic potential function describing the many body interaction of chromophores.

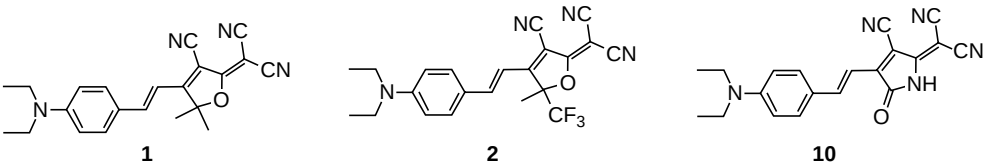

	1	2	10	
Dipole moment:	11.8	11.9	11.7	Debye
Hyperpolarizability	23.8	34.4	43.0	$*10^{-30}$ esu

Figure 1 Density functional calculations (DFT) of molecular first hyperpolarizability (β) and dipole moment (μ) for three chromophore structures.

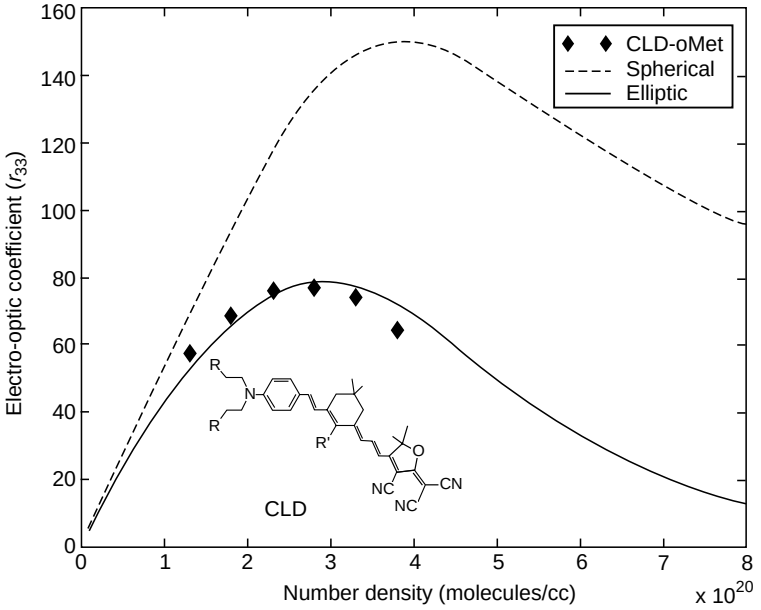

Figure 2 Two theoretical (statistical mechanical) simulations of the variation of electro-optic activity with chromophore number density are shown and compared with experimentally observed results for the CLD chromophore (structure shown) dissolved in an amorphous polycarbonate host polymer.

One component favors centrosymmetric ordering while the second component favors noncentrosymmetric ordering. Their relative weighting in defining macroscopic order depends on chromophore shape. Theoretical calculations have also shown that embedding chromophores in nanoscopic macromolecular objects (such as dendrimers and dendronized polymers) can dramatically enhance electro-optic activity. This type of engineering can also be used to control auxiliary properties such as optical loss and stability. The exceptional processability of organic electro-optic materials has permitted conformal and flexible devices to be fabricated and has permitted electro-optic circuitry to be integrated vertically on top of semiconductor VLSI (very large-scale integration) electronics. Mach–Zehnder modulators, with 3 dB operational bandwidths of 200 GHz, have been fabricated from organic electro-optic materials as have Mach–Zehnder devices with operational drive voltages of less than one volt. The major advantage that crystalline inorganic materials possess, relative to macromolecular organic EO materials, is superior thermal and photostability. A comparison of lithium niobate and the best organic EO materials, is given in Table 1. Cost is a significant factor if electro-optic materials are to compete with modulated lasers and other opto-electronic devices for applications in communications, computing, transportation, and defense. Currently, manufacturing costs for both crystalline and macromolecular materials are very high (several thousand dollars per modulator). Recently, it has been demonstrated that organic macromolecular materials can be used with soft lithography techniques to mass-produce electro-optic circuitry. This could conceivably dramatically reduce manufacturing costs. Also, the fact that organic macromolecular materials can be directly integrated with silicon photonics and VLSI semiconductor electronics, may influence future commercial utilization of electro-optic materials, which is currently dominated by lithium niobate.

Table 1 Comparison of electro-optic performance for inorganic and organic EO materials at 1.3 microns wavelength

Parameter	*$LiNbO_3$*	*NLO polymer*
Effective EO coefficient, r_{eff}, pm/V	32	182
Index of refraction (n)	2.2	1.6–1.7
Dielectric permittivity (ε)	30	3
Material Figure of Merit, $n^3(r_{eff})/\varepsilon$	10	270
Bandwidth × length product, GHz cm	7	>100
$V_\pi L$, V cm	5	1.5
Optical loss, dB/cm	0.2	0.2–1.0
Processing temperature (°C)	1000	<200
Multiple layers possible	No	Yes

Devices

Representative simple device structures are shown in Figure 3. For a simple stripline phase (birefringence) modulator (Figure 3, upper left), the phase retardation produced by application of an electric field is given by

$$\Delta\phi = 2\pi\Delta nL/\lambda = \pi n^3 rEL/\lambda = \pi n^3 rVL/\lambda d \qquad [18]$$

where L is the electrode length (the distance over which the electrical and optical fields co-propagate),

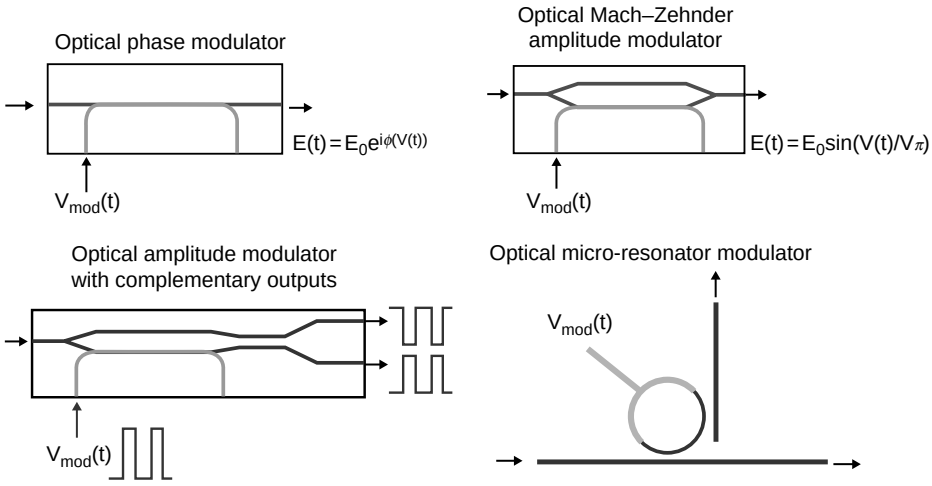

Figure 3 Schematic representation of simple electro-optic device structures are shown: Upper left (birefringent modulator), upper right (Mach–Zehnder interferometer), lower left (1 × 2 optical switch), lower right (ring microresonator).

λ is the optical wavelength, $V = E/d$ is the electric field felt by the material, and d is the electrode spacing. Electro-optic device performance is frequently characterized by the voltage, V_π, required to produce a phase shift of π radians. For a TM polarized light propagating through a birefringent modulator:

$$V_\pi = \lambda d/(n_{0e}^3 r_{33} L\Gamma) \qquad [19]$$

where Γ characterizes the overlap of the optical and electrical waves.

A simple Mach–Zehnder interferometer (electro-to-optical signal transducer) is shown in the upper right of Figure 3. When such a device is constructed from electrically poled organic electro-optic chromophores and a voltage is applied to one arm, a retardation of light propagating in that arm will be effected. This, in turn, will result in a change in the constructive/destructive interference of the beams at the output of the device. The consequence of applying an electrical signal to one arm of a Mach–Zehnder interferometer is that the applied voltage is transduced onto the optical output beam as an amplitude modulation of that beam. The voltage required to achieve full wave modulation for an optical beam of TM polarization is

$$V_\pi = \lambda d/(n_{0e}^3 r_{33} L\Gamma) \qquad [20]$$

Amplitude modulation can also be achieved using a birefringent modulator by placing polarizers at the input and output of the device. If an input polarizer is used to launch TM and TE modes, the birefringent modulator functions to produce a rotation of light, which is turned into an amplitude modulation by inserting a polarizer at the output. For such a device, V_π depends upon the difference in r_{33} and r_{13}. Note that for poled organic chromophores, $r_{33} \approx 3r_{13}$, $(V_\pi)_{\mathrm{BRM}} \approx 1.5(V_\pi)_{\mathrm{MZM}}$. For the above devices, it is clear that V_π will depend on the length over which the electrical and optical fields co-propagate in phase. For low electrical field frequencies, L will be the electrode length. The electrode spacing d, that can be used, will depend on the index of refraction difference between the core (electro-optic material) and a cladding material used to prevent the optical field experiencing the metal electrodes (an event that would dramatically increase propagation loss). For poled organic macromolecular materials and standard commercial claddings, d is typically in the order of 10 microns. Electrode lengths usually range from 1 to 3 cm. Two factors limit bandwidth in stripline devices: (i) velocity mismatch between the electrical and optical waves (which relates to the difference between n_0^2 and ε, where ε is the dielectric permittivity of the material – see Table 1); and (ii) microwave and millimeter wave loss in the metal electrodes. For organic electro-optic materials, the latter defines operational 3 dB bandwidth of devices.

Optical routing switches (see Figure 3, lower left) are more complicated to understand and typically require somewhat larger V_π voltages to effect a switching operation (e.g., $(V_\pi)_{\mathrm{simple2\times2crossbar}} \approx 1.7(V_\pi)_{\mathrm{MZM}}$).

Two conditions must be satisfied for light to be coupled into a ring microresonator device (see Figure 3, lower right): (i) the circumference of the ring must be a multiple of the wavelength of light, and (ii) the velocity of light in the ring must equal the velocity of light at the input. If a ring microresonator is fabricated from an electro-optic material, then an applied electrical voltage can be used to control coupling of light into and out of the ring microresonator. The same electrode structure, that is used to achieve voltage-controlled wavelength selective filtering, can also be used to transduce electrical information onto the optical beam while it is in the ring resonator. The quality (Q) factor of a ring microresonator defines both the wavelength selectivity and the bandwidth of the device. Ring microresonator devices also have the advantage of permitting a long interaction length to be achieved in a very compact device structure. The dimensions (radii) of ring microresonator structures are defined by bending loss and thus by the difference between core and cladding materials. Typical radii can range from hundreds of nanometers to hundreds of microns. With ring microresonators, a reduction in drive voltage requirements can be achieved by trading off bandwidth. Another price paid in the use of ring microresonator device structures is that of bending loss, which is not present in the other device structure shown. Ring microresonators can be used for active wavelength division multiplexing (WDM), voltage-controlled optical signal routing, and wavelength tuning of optical sources.

The interaction length of a prism optical beam deflector can be increased by cascading prisms together. For such a device structure, the angle of deflection becomes

$$\theta = n_0^3 r_{33}(V/d)(L/h) \qquad [21]$$

where h is the height of an individual prism and L is the length of the prism array. For currently available electro-optic materials only small deflection angles (a few degrees or less) can be achieved with practical

voltage levels. Superprism structures may permit large deflection angles to be achieved but such devices have yet to be demonstrated in a convincing manner.

See also

Interferometry: Overview. **Modulators:** Acousto-Optics. **Optical Communication Systems:** Wavelength Division Multiplexing.

Further Reading

Butcher PN and Cotter D (1990) *The Elements of Nonlinear Optics*. New York: Cambridge University Press.

Dagli N (1999) Wide-bandwidth lasers and modulators for RF photonics. *IEEE Trans. Microwave Theor. Techniques* 47: 1151–1171.

Dalton LR (2003) Rational design of organic electro-optic materials. *Journal of Physical Condensed Materials* 15: R897–R934.

Dalton L, Harper A, Ren A, *et al.* (1999) Polymeric electro-optic modulators: from chromophore design to integration with semiconductor very large scale integration electronics and silica fiber optics. *Ind. Eng. Chem. Res.* 38: 8–33.

Hornak LA (1992) *Polymers for Lightwave and Integrated Optics*. New York: Marcel Dekker.

Khazaei HR, Wang WJ, Berolo E and Ghannouchi F (1998) High speed GaAs/AlGaAs traveling wave electro-optic modulators. *Proceedings of SPIE* 3278: 94–100.

Lawetz C, Cartledge JC, Rolland C and Yu J (1997) Modulation characteristics of semiconductor mach-zehnder optical modulators. *IEEE Journal of Lightwave Technology* 15: 697–702.

Lee KS (2002). *Advances in Polymer Science*, vol. 158. Heidelberg: Springer.

Lee M, Katz HE, Erben C, *et al.* (2002) Broadband modulation of light using an electro-optic polymer. *Science* 298: 1401–1403.

Nalwa HS and Miyata S (1997) *Nonlinear Optics of Organic Molecules and Polymers*. Boca Raton, FL: CRC Press.

Noguchi K, Mitomi O and Miyazawa H (1998) Millimeter-wave Ti:$LiNbO_3$ optical modulators. *IEEE Journal of Lightwave Technology* 16: 615–619.

Shen YR (1984) *The Principles of Nonlinear Optics*. New York: John Wiley & Sons.

Shi Y, Zhang C, Zhang H, *et al.* (2000) Low (sub-1 volt) halfwave voltage polymeric electro-optic modulators achieved by controlling chromophore shape. *Science* 288: 119–122.

Wise DL, Wnek GE, Trantolo DJ, Cooper TM and Gresser JD (1998) *Electrical and Optical Polymer Systems. Fundamentals, Methods and Applications*. New York: Marcel Dekker.

Yariv A and Yeh P (1984) *Optical Waves in Crystals*. New York: John Wiley & Sons.

Modulation and Demodulation of Optical Signals

R A Minasian, University of Sydney, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Since earliest times, light has been used for communication purposes. In order to convey information, a physical variable of light, such as the optical intensity or field amplitude, must be modulated to encode a signal onto the optical carrier, and it must be demodulated or decoded after transmission. Laser sources provide the capability of extremely high modulation speeds, which result in high information-carrying capacity in optical communication systems. After transmission, semiconductor photodetectors demodulate the optical signals to recover the original information (*see* **Optical Communication Systems:** Basic Concepts).

The modulation of optical signals can be classified into two types: field modulation and intensity modulation, which is illustrated in Figure 1. In field modulation, the physical variable that is modulated is the frequency, phase, or amplitude of the optical field. These three techniques are sometimes referred to as frequency shift keying (FSK), phase shift keying (PSK) and amplitude shift keying (ASK). Figure 1b shows an example of frequency modulation of light. Field modulation requires keeping track of the phase of the carrier, which imposes strict requirements on the optical coherence (*see* **Coherent Lightwave Systems**). Because the optical carrier is at an extremely high frequency of around 2×10^{14} Hz at fiber-optic communication wavelengths of 1.5 μm, the required optical coherence is very difficult to realize, and moreover the receiver requires complex optical heterodyne methods to detect the phase of the optical field.

In intensity modulation, the physical variable that is modulated is the optical power. Figure 1c shows that it is the envelope of the light power that contains the information. The optical field oscillations at 10^{14}–10^{16} Hz in this case are not related to the

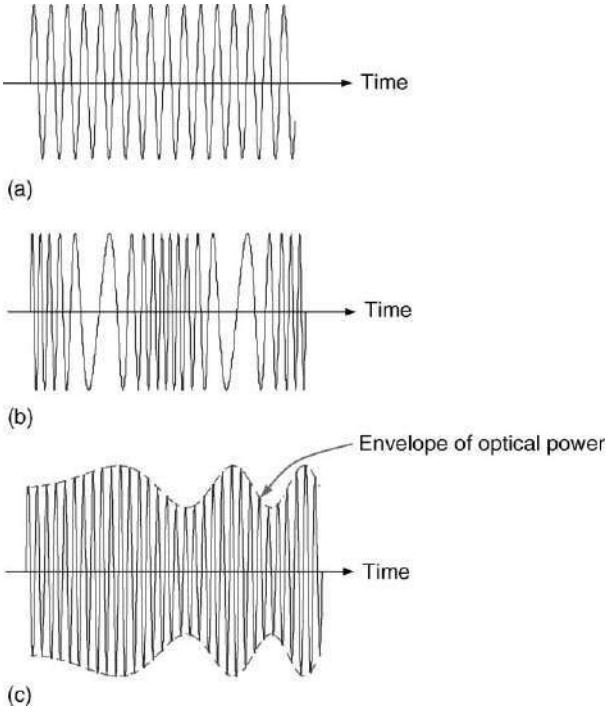

Figure 1 Types of modulation of optical signals. (a) unmodulated optical field; (b) frequency-modulated field; (c) intensity or envelope modulation of the optical power.

modulation and demodulation process, and only the optical power variation is important. This is a key advantage, and as a result intensity modulation systems are prevalent for optical communications. Here, we discuss how varying the injected current in a semiconductor laser or light emitting diode, or by operating the source with constant output and introducing an external modulator, can realize modulation of the optical power of the source. The process by which the optical power variations can be detected at the receiver with semiconductor photodiodes are also discussed, including detection noise characteristics and means of enhancing the responsivity by introducing internal gain mechanisms or wavelength selectivity. The speed capabilities of optical modulation and demodulation in semiconductor lasers and photodetectors are discussed, based on the rate equations for carrier and photon dynamics.

The Modulation of Semiconductor Lasers

Townes, Basov and Prokhorov, who shared the Nobel Prize in 1964, developed the principle of lasers. Semiconductor lasers emit light through a stimulated emission mechanism. The required optical gain for

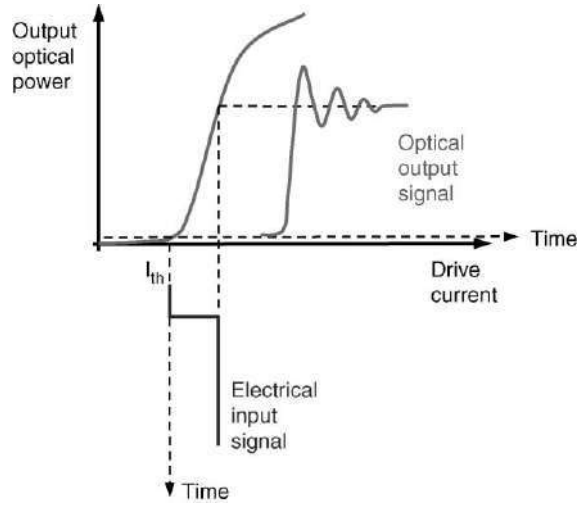

Figure 2 Light-current properties and the dynamic modulation characteristics of semiconductor lasers.

the lasing action is obtained by injecting a forward bias current into a heavily doped p-n junction, thereby injecting carriers into the active junction layer and creating a population inversion. Here we limit our discussion to the optical modulation characteristics of the laser in response to changes in the injection current.

Figure 2 shows the typical light-current characteristics of semiconductor lasers. For drive currents above the threshold value I_{th}, the emitted optical power is predominantly by stimulated emission and it increases linearly with the drive current up to moderate currents. The differential responsivity of the laser, defined by the ratio of optical power increase to electric current increase above threshold, is typically around 0.5 mW/mA. At high current levels, the light output deviates from a proportional dependence on the current because of saturation effects and the differential responsivity decreases.

The dynamic response of the semiconductor laser to modulation of the injection drive current is governed by the rate equations for the carrier density and the photon density in the active junction region. Figure 3 shows the active layer under consideration (*see* **Lasers:** Edge Emitters). The electron concentration in the active layer is denoted N and the photon concentration in the active layer is denoted S. The dynamic behavior of the electron concentration and photon density is governed by the following rate equations:

$$\frac{\partial N(x)}{\partial t} = \frac{I(x)}{q\nu_a} + D_n \frac{\partial^2 N(x)}{\partial x^2} - \frac{N(x)}{\tau_{ns}} - G(x)\cdot S(x) \qquad [1]$$

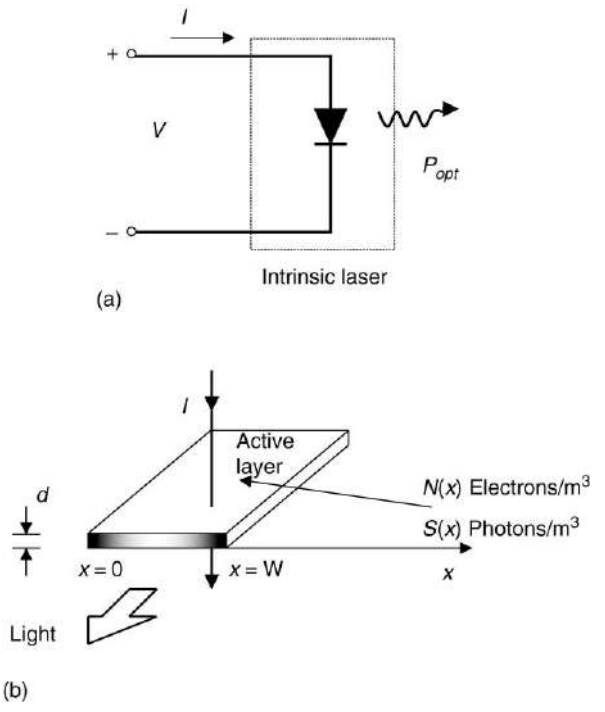

Figure 3 Semiconductor laser. (a) intrinsic device; (b) active region with injected current I, electron concentration N, and photon concentration S.

$$\frac{\partial S(x)}{\partial t} = \Gamma \cdot \beta \frac{N(x)}{\tau_{ns}} + \Gamma \cdot G(x) \cdot S(x) - \frac{S(x)}{\tau_p} \qquad [2]$$

where I is the injected current, q is the electronic charge, ν_a is the volume of the active layer, D_n is the diffusion constant, τ_{ns} is the carrier lifetime, Γ is the confinement factor (defined as the fraction of the optical power in the active region), β is the fraction of the spontaneous emission power coupled into the lasing mode, τ_p is the cavity or photon lifetime, and

$$G(x) = \gamma[N(x) - N_t] \qquad [3]$$

where G is the net rate of stimulated emission, and N_t is the electron concentration at transparency.

Equations [1] and [2] are statements of the fundamental continuity equations for electrons and photons in the active layer. Hence, eqn [1] expresses the fact that the rate of increase of electron concentration in the active layer is given by the rate of increase of electron concentration due to injected current, the rate of increase of electron concentration due to diffusion current, the rate of decrease of electron concentration due to spontaneous recombination, and the rate of decrease in electron concentration due to stimulated photon emission. Equation [2] expresses the fact that the rate of increase of photon density in the active layer is given by the rate of increase in photon density due to spontaneous emission coupled into the lasing mode, the rate of increase of photon density due to stimulated emission coupled into the active layer, and the rate of decrease of photon density due to cavity loss mechanisms.

The actual electron and photon distribution varies across the active layer. However, for simplicity, it is convenient to use average values of N and S, and the simplified rate equations in terms of position independent variables are given by:

$$\frac{dN}{dt} = \frac{I_i}{q\nu_a} - \frac{N}{\tau_{ns}} - G(1 - \varepsilon S)S \qquad [4]$$

$$\frac{dS}{dt} = \Gamma \cdot \beta \frac{N}{\tau_{ns}} + \Gamma \cdot G(1 - \varepsilon S)S - \frac{S}{\tau_p} \qquad [5]$$

where the scaling factor $(1 - \varepsilon S)$ accounts for gain saturation arising from several phenomena such as spatial hole burning, spectral hole burning, carrier heating and two-photon absorption, and ε is a nonlinear gain parameter.

The modulation response of the laser is obtained by solving the rate eqns [4] and [5] for a time dependent input current of the form

$$I(t) = I_b + I_m x(t) \qquad [6]$$

where I_b is the steady-state bias current, I_m is the modulation current amplitude and $x(t)$ represents the normalized shape of the modulation waveform.

The small-signal modulation bandwidth of the laser is obtained by considering the response to small amplitude sinusoidal current modulation at frequency ω, so that $x(t) = \sin(\omega t)$, around a steady-state bias current above lasing threshold. Under the small signal condition, the rate equations can be linearized and combined to result in a second order differential equation. The solution for the photon output exhibits a damped resonant type response. Hence the modulation response displays some peaking at the relaxation oscillation resonance frequency ω_r, and is flat for frequencies much lower than ω_r. The resonance frequency $\omega_r \propto \sqrt{S_o}$, where S_o is the photon density at the laser bias level, so the bandwidth can be increased by biasing the laser further above its threshold current. Semiconductor laser bandwidth to 30 GHz has been observed; however, it is often limited to around 10 GHz due to electrical parasitics in the current drive circuit and interface.

The large-signal modulation response of the laser is obtained by solving the differential eqns [4] and [5]. The response of the carrier-photon coupled system to

a step change of drive current has the potential for damped oscillation. This is caused by a resonant type of behavior in which, for example, an increase in photon density causes a depletion in population inversion which leads to a decrease in photon density enabling the carrier density to recover, thus producing a time lag between cause and effect resulting in damped oscillations before settling to the steady state value. These dynamic characteristics are illustrated in Figure 2. For a step modulation of the laser drive current, the optical output response exhibits a delay time, a risetime, an overshoot, and damped oscillations. The risetime of the laser output is typically in the range of 100–300 ps. The turn-on time of the laser and the overshoot are improved by ensuring that the minimum drive current does not fall below threshold. The damped oscillations in the optical output occur at the relaxation oscillation frequency of the laser.

These laser pulse waveform imperfections cause deteriorations in the eye diagram of the laser transmitter that is modulated by very high speed digital bit streams (*see* **Optical Communication Systems:** Lightwave Transmitters). Figure 4 shows the eye diagram for the optical output of a semiconductor laser directly modulated by a 5 Gbit/s pseudo-random bit stream of length $2^{15}-1$. The output waveform imperfections cause a partial closure of the eye, which ultimately limits the modulation speed that can be obtained by directly modulating the laser drive current. Although the semiconductor laser optical output does not exactly replicate the applied electrical drive waveform, the deviations are small enough so that these lasers can be used for data transmission up to around 10 Gbit/s.

Another inherent consequence of direct modulation of semiconductor lasers is the production of laser frequency chirp. This arises from the fact that modulation induced changes in the carrier density in the active region of the laser cavity cause a change in the refractive index of the active layer, which results in a chirping in the output wavelength during modulation. Hence amplitude modulation of semiconductor lasers is fundamentally accompanied by phase modulation from the basic physical mechanism of carrier induced changes in the refractive index. The resulting transient changes in the output wavelength of the laser causes a chirping phenomenon as it is being modulated. The order of chirp magnitude is about 0.1–1 GHz/mA of drive current change. During digital pulse modulation, the laser wavelength shifts to the blue side near the leading edge and towards the red side near the trailing edge of the pulse. Chirp results in a broadening of the laser linewidth. This wavelength change effect in combination with the dispersion characteristics of optical fiber (*see* **Fiber and Guided Wave Optics:** Dispersion; Light Propagation) causes an important limitation to the transmission capacity of fiber optic communication systems. The chirp in semiconductor lasers can be reduced, but not entirely eliminated, by using modulation-doped strained multiquantum-well lasers, which reduce the linewidth enhancement factor.

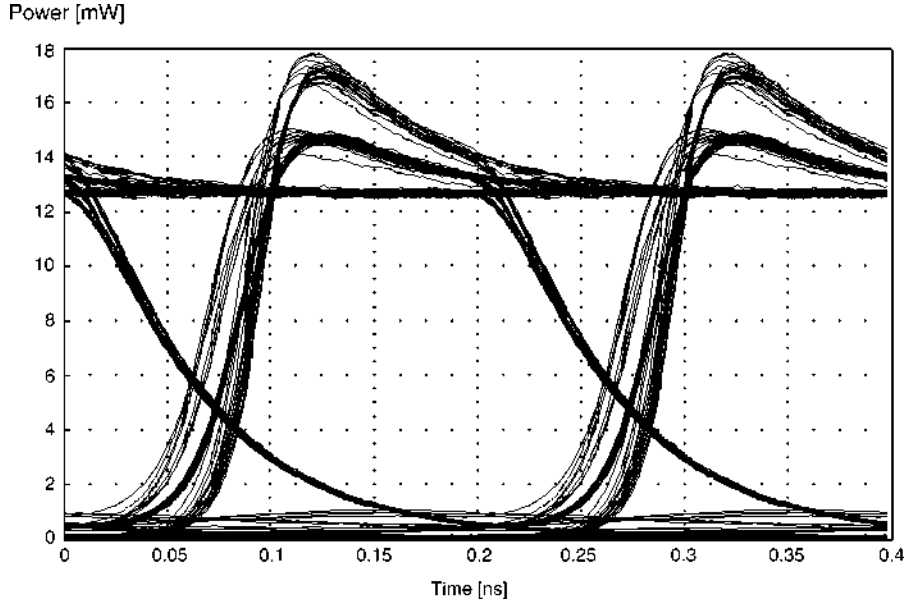

Figure 4 Eye diagram for the optical output of a semiconductor laser directly modulated by a 5 Gbit/s pseudo-random bit stream.

The Modulation of Light Emitting Diodes

The light emitting diode (LED) generates light through spontaneous emission. A semiconductor p-n junction is pumped with a forward current, so that electrons are provided in the upper energy state that can recombine radiatively to produce incoherent light by a spontaneous emission mechanism. The rate of photon generation is directly proportional to the carrier injection rate, and hence the modulating drive current.

Figure 5 shows the typical light-current properties and the modulation characteristics of a LED. The responsivity of the LED, defined as the ratio of the emitted optical power to the injected current, is typically in the range of 10 to 50 μW/mA. The optical output power generated is linearly related to the injection (drive) current up to moderate currents, and is given by

$$P = \eta \cdot h\nu \cdot \frac{I}{q} \qquad [7]$$

where η is the emission quantum efficiency (defined as the ratio of the radiative recombination rate to the total recombination rate), h is Planck's constant, ν is the optical frequency, I is the injected current, and q is the electronic charge. For larger currents, the light output deviates from a proportional dependence on the current because of saturation effects and the responsivity decreases.

The modulation response of the LED is governed by the rate equation for the carrier density N

$$\frac{dN}{dt} = \frac{I}{qv_a} - \frac{N}{\tau_c} \qquad [8]$$

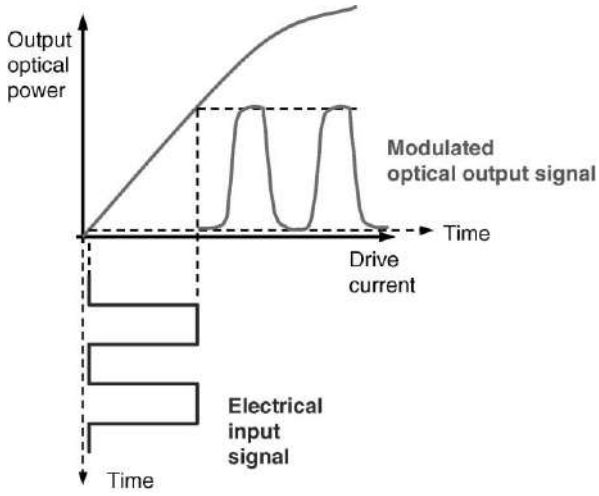

Figure 5 Light-current properties and the modulation characteristics of light emitting diodes.

where I is the injected current, q is the electronic charge, v_a is the volume of the active region, and τ_c is the carrier lifetime. Equation [8] expresses the fact that the rate of increase of carrier concentration is given by the rate of increase of carriers through current injection and the rate of carrier loss through radiative and nonradiative recombination. The small signal response arising from this first order differential equation, exhibits a simple single pole response and the resulting modulation bandwidth of the LED is given by

$$B = 1/(2\pi\tau_c) \qquad [9]$$

The pulse modulation response of the LED is shown in Figure 5. The output modulated optical waveform is not square but exhibits risetimes and falltimes determined by the carrier lifetime τ_c. This sets a limit to the maximum modulation rate that the LED can achieve. The bandwidth can be increased by decreasing the radiative lifetime, which can be controlled to a limited extent by the choice of doping level. Typical lifetimes for InGaAsP LEDs are in the range of 1–5 ns, which corresponds to attainable modulation data rates of several hundred Mbit/s.

The External Modulation of Lasers

Modulation of lasers realized by changing the injection current, as described above, is classed as direct modulation. This is the simplest method of modulation; however, it is limited by several physical phenomena. These include bandwidth limitations due to carrier dynamics in the laser active region, waveform limitations due to relaxation oscillations and overshoot effects, and chirping which causes transmission limitations in optical fibers arising from dispersion effects.

A superior modulation technique, classed as external modulation, operates the laser with continuous wave (CW) output and impresses the modulation by means of a device external to the laser. Although more complex, this approach has many modulation performance advantages. Because the laser is operated CW, this technique removes chirping effects and also enables higher power lasers to be used. Moreover, it is capable of much higher modulation bandwidths, larger on–off extinction ratios, and superior modulation spectral purity.

Electro-optic Modulation of Light

The linear electro-optic effect in materials was first described by Friedrich Pockels in 1893. The Pockels effect is the change in refractive index of the material

resulting from the application of an electric field, and is given by

$$n(E) = n - \frac{1}{2} r n^3 E \qquad [10]$$

where n is the refractive index, E is the applied electric field, and r is the electro-optic coefficient. Typical values of r lie in the range 10^{-12} to 10^{-10} m/V, and the most common electro-optic materials used include $LiNbO_3$, GaAs, InP, InGaAsP, and synthetic polymers such as polymethylmethacrylate (*see* **Modulators:** Electro-optics).

If light traverses an optical waveguide of length L, to which an electric field E is applied, it undergoes a phase shift given by

$$\Delta\phi = -\frac{\pi r n^3 E L}{\lambda} \qquad [11]$$

The electric field can be produced by applying a voltage V across the waveguide of width d, hence $E = V/d$. It is therefore possible to modulate the phase of the light by varying the voltage V that is applied to the material through which the light passes. The voltage required to produce a phase shift of π is given by

$$V_\pi = \frac{\lambda d}{n^3 r L} \qquad [12]$$

The parameter V_π, known as the half-wave voltage, is an important characteristic that expresses the applied voltage required to give a phase shift of π.

Intensity modulation of light can be obtained by placing the phase modulator in one arm of an optical interferometer. The structure, known as a Mach-Zehnder modulator, is shown in Figure 6a. The optical output is the phasor sum of two lightwaves that travel through the two arms, resulting in constructive or destructive interference depending on the optical phase difference. The electric field is applied to the optical waveguide in a direction perpendicular to the direction of light propagation using the metal electrodes, see Figure 6a. Hence the refractive index of the waveguide and thus the phase of the light can be modulated by varying the voltage V. When no voltage is applied, the optical fields in the two arms of the interferometer experience identical phase shifts, and hence interfere constructively at the output. When a voltage is applied, the waveguide index changes, and the light in that arm experiences a phase shift with respect to the light in the other arm. This destroys the constructive nature of the interference at the output, and reduces the transmitted light emitted at the output. The extreme

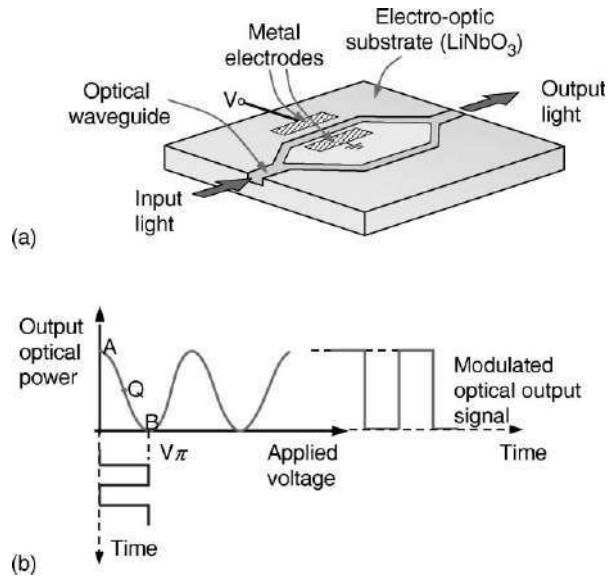

Figure 6 Mach-Zehnder electro-optic intensity modulator. (a) Structure; (b) transfer function and modulation characteristics.

case is when the phase difference between the two arms equals π, which causes complete destructive interference resulting in no light transmission.

The modulation transfer function is given by

$$P_{out} = \frac{1}{2} P_{in}[1 + \cos(\Delta\phi)] \qquad [13]$$

where $\Delta\phi$ is the phase difference encountered by light as it travels through the two arms. In terms of the applied voltage, the optical power transmitted through the modulator is given by

$$P_{out} = \frac{1}{2} P_{in}\left[1 + \cos\left(\pi \frac{V}{V_\pi}\right)\right] \qquad [14]$$

This transfer characteristic is shown in Figure 6b. The modulator may be operated as an analog optical modulator by operating in the nearly linear region around point Q. Alternately, it may be operated as a digital modulator by operating between points A and B, thus switching the light on and off as V is switched between 0 and V_π.

Improved efficiency of modulation, and hence a reduction in the value of V_π (typically a few volts) can be obtained by placing a second electrode on the other arm of the interferometer, and using a push-pull configuration to drive the interferometer. Hence the phase modulators in both arms are driven with modulating voltages of opposite polarities so the phase is advanced in one arm and is retarded in the other arm. This produces a net phase difference, which is double that produced in either arm.

The Pockels effect is an extremely fast physical mechanism, hence in principle external modulation is capable of achieving extremely high speeds. In practice, the modulation speed of an electro-optic modulator is limited by the capacitive effects at the electrodes and by the transit time of the light through the material. If the electric field varies significantly within the time it takes for light to transit through the waveguide, then the travelling optical wave will be subjected to different electric fields as it traverses the waveguide, and this results in a transit time limitation to the modulation bandwidth. A technique to avoid this limitation is to apply the voltage at one end of the waveguide and to design the electrodes to form a transmission line. By this means it is possible to make the velocity of the electrical wave match that of the optical wave to essentially eliminate the transit time effects. These external modulators have demonstrated speeds well in excess of 100 GHz.

Electro-absorption Modulation of Light

Electro-absorption external modulators are based on the principle of semiconductor bandgap decrease that occurs upon application of an electric field (the Franz-Keldysh and Stark effect). With no applied voltage, the semiconductor waveguide is transparent to light having a photon energy smaller than the bandgap, and such wavelengths are transmitted with very little attenuation. When a voltage is applied, the bandgap of the semiconductor decreases and the material absorbs the light once the photon energy exceeds the bandgap energy, and this produces optical attenuation. Hence by applying a modulation voltage to the semiconductor waveguide, the absorption property of the waveguide and hence the output light intensity can be modulated.

Improved modulation efficiency, and hence a reduction in the voltage required to decrease the bandgap energy, is obtained by using multiquantum-well semiconductor structures, which exhibit enhanced electro-absorption effects arising from the quantum confined Stark effect. The conventional technique of applying a transverse field across a multiquantum-well is to apply a reverse voltage across it. Modulator drive voltages of a few volts can be obtained. The on/off extinction ratio of modulation depends on the value of the field-dependent absorption coefficient at the operating voltage, and typical on/off extinction ratios are 10–15 dB.

The modulation bandwidth is determined by intrinsic factors related to carrier dynamics, such as exciton formation and dissociation times and the time taken to remove electrons and holes from the wells by thermionic emission and tunelling. It is also determined by extrinsic factors related to the capacitance of the junction diode. These modulators can have bandwidths of around 50 GHz. They also enable low chirp modulation to be obtained. The optical power handling capability of electro-absorption modulators is lower than for electro-optic modulators. However, a principal advantage of the electro-absorption modulator is that the semiconductor waveguide can be fabricated onto one substrate with the semiconductor laser, enabling the realization of monolithic integration on the same chip.

The Demodulation of Optical Signals

The process of demodulation converts the modulated intensity of an optical wave into an electrical signal to recover the information content. Semiconductor photodetectors are based on the principle of absorption of photons to generate mobile electron-hole pairs, and the transport of these mobile carriers via an applied electric field to create a flow of current.

Demodulation using PIN Photodiodes

The PIN photodiode comprises a P-doped/intrinsic/N-doped semiconductor structure. An applied reverse voltage bias creates a depletion region with a high electric field in the intrinsic layer. When the intrinsic layer is illuminated with light, photons are absorbed and electron-hole pairs are created, which are swept by the built-in electric field. The resulting flow of current is proportional to the incident optical power.

The quantum efficiency of the photodetector is defined as the ratio of electron generation rate to the photon incidence rate, and is given by

$$\eta = \frac{h\nu}{q} \cdot \frac{I_p}{P} \qquad [15]$$

where I_p is the photocurrent, P is the incident optical power, ν is the optical frequency, h is Planck's constant and q is the electron charge. If the width of the absorbing intrinsic region is large enough to capture all the light, quantum efficiency values approaching 100% can be achieved.

The current-voltage characteristics of the photodiode under optical illumination are shown in Figure 7a. The dark current I_d, which is the current generated in the photodetector in the absence of light originates from thermally generated electron-hole pairs, and is typically extremely small (<5 nA). Figure 7b shows the transfer characteristic of the

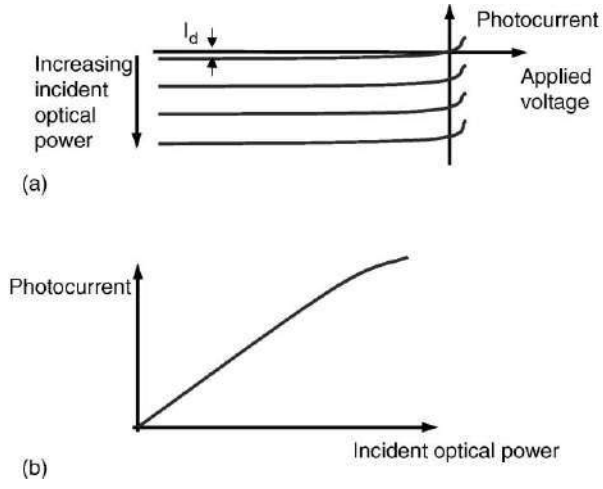

Figure 7 Semiconductor photodiode. (a) Current-light characteristics under optical illumination; (b) transfer function and demodulation photocurrent versus incident optical power.

photodetector. The current is linearly proportional to the optical power up to moderate power levels, when saturation mechanisms occur. The responsivity of the photodiode $\Re$, defined as the ratio of the generated electric current to the incident optical power, is given by

$$\Re = \frac{\eta q}{h\nu} \qquad [16]$$

The responsivity is an important parameter, and provides a measure of the photodetection efficiency of the device for demodulating optical signals. Typical responsivity values for PIN photodiodes are around 1 A/W.

The bandwidth B of demodulation of the photodetector is determined by the speed with which it responds to variations in the incident optical power. This is limited by the transit time τ_{tr} and the RC time constant τ_{RC} and is given by

$$B = \frac{2\pi}{\tau_{tr} + \tau_{\mathrm{RC}}} \qquad [17]$$

The transit time limitation arises because of the time it takes for the generated carriers to traverse the absorption region and be collected. To maximize the bandwidth, the transit time needs to be reduced, which can be achieved by decreasing the width of the absorption region. However, this requires a trade-off with the responsivity, which drops because a smaller fraction of the light can be absorbed under these conditions. The capacitance C and the resistance R of the photodetector and its circuitry form the RC time constant. To maximize the bandwidth, the photodiode capacitance needs to be reduced by decreasing the area of the photodiode. The demodulation bandwidth of PIN photodiodes can exceed 100 GHz.

Photodetection shot noise is a fundamental noise mechanism that introduces noise in the demodulation process of optical signals. This leads to fluctuations in the photocurrent, even for a constant incident optical power, because the photocurrent actually consists of a stream of electrons that are generated at random times. The photodiode current generated by a constant incident optical power is given by

$$I(t) = I_{\mathrm{avg}} + i_{\mathrm{shot}}(t) \qquad [18]$$

where I_{avg} is the average photocurrent generated, and $i_{\mathrm{shot}}(t)$ is the current fluctuation due to shot noise and which is described by a stationary random process with Poisson statistics. The mean-square noise current is given by

$$< i^2_{\mathrm{shot}}(t) >= 2qI_{\mathrm{avg}}\Delta f \qquad [19]$$

where Δf is the effective noise bandwidth of the receiver. The shot noise degrades the signal to noise ratio of the demodulated signal, and causes a fundamental limit to the sensitivity of the PIN photodetection.

Demodulation using Avalanche Photodiodes

Demodulation using avalanche photodiodes detectors (APD) has the advantage of significantly increased responsivity. This is achieved by providing an internal current gain mechanism within the device. The APD operates by generating more than one electron-hole pair per absorbed photon. This is achieved by strongly reverse biasing the diode to create a large junction electric field, so photoelectrons accelerate and acquire enough energy to produce new carriers by the process of impact ionization. These devices incorporate a modification to the PIN structure, by including an additional doped layer known as the multiplication layer, which under reverse bias provides a region of very high electric field that can produce avalanche breakdown. Hence a single primary electron generated through absorption of a photon, generates many secondary electrons and holes in the avalanche multiplication layer, and all these secondary carriers contribute to the photocurrent. As a result of this current gain mechanism, the responsivity of the APD is increased by a multiplication factor M, and is given by

$$\Re_{ADP} = \frac{M\eta q}{h\nu} \qquad [20]$$

where M can range between 5–50 depending on the bias voltage.

The bandwidth of the APD depends on the multiplication factor M. This is because the avalanche process takes additional time. The frequency response can be written

$$M(\omega) = \frac{M_0}{[1 + (\omega\tau_e M_0)^2]^{\frac{1}{2}}} \qquad [21]$$

where M_0 is the low frequency gain and τ_e is the effective transit time that depends on the ionization coefficient. The intrinsic APD bandwidth (assuming no RC limitations) is approximately given by

$$B_{\mathrm{APD}} = \frac{1}{2\pi\tau_e M_0} \qquad [22]$$

Hence the bandwidth decreases as the multiplication factor increases, and there exists a trade-off between the bandwidth and the responsivity. The demodulation characteristics of the avalanche photodiode can be expressed by means of a gain-bandwidth product $B_{\mathrm{APD}} \cdot M_0$, and this can exceed 150 GHz.

Photodetection noise in APDs has an additional contribution because the avalanche gain itself is a random variable. Hence the secondary electron-hole pairs are generated at random times and this adds a noise component to the basic shot noise associated with the generation of primary electro-hole pairs. The total shot noise is given by

$$< i^2_{\mathrm{shot,APD}}(t) >= 2qM^2F(M)I_{\mathrm{avg}}\Delta f \qquad [23]$$

where $I_{\mathrm{avg}} = \Re P$, and F is the excess noise factor expressed by

$$F = \frac{< M^2 >}{< M >^2} \qquad [24]$$

and is approximately given by $F \approx M^x$, where $0.2 < x < 1$. Hence the excess noise increases with the multiplication ratio, so there exists an optimum avalanche photodiode gain that maximizes the signal to noise ratio of the demodulated signal. Despite the fact that additional noise is introduced, because of the higher responsivity, APDs can provide higher sensitivity in demodulating optical signals than PIN photodiodes.

Demodulation using Metal-Semiconductor-Metal (MSM) Photodiodes

Demodulation using MSM photodiodes enables extremely high bandwidths to be achieved. The MSM photodiode comprises two reverse-biased metal-to-semiconductor Schottky contacts formed on a semiconductor layer that prevents flow of electrons from metal to semiconductor. Photons incident between the contacts create electron-hole pairs that are collected by the electric field to generate a photocurrent proportional to the incident optical power. The metal contacts are usually fabricated on the same side of the semiconductor to create a planar structure, which is suitable for monolithic integration. However, a drawback of the MSM device is that the electrodes block some of the light from reaching the absorption region, which results in a reduced responsivity.

The bandwidth of demodulation of the MSM photodiode is determined by the transit time of the photogenerated carriers and the RC time constant. Since the electrode spacing can be around 1 μm, the transit time is extremely small. Moreover, since in the planar structure the parasitic capacitance is inherently low, the RC time constant is also extremely small. This results in demodulation bandwidths of up to 300 GHz.

Wavelength Selective Demodulation

A technique to obtain both wavelength selective demodulation and high-speed, high-responsivity demodulation involves the use of microcavity resonant photodetectors. This is useful for very high-speed detection in wavelength multiplexed systems.

The microcavity photodiode comprises a Fabry-Perot cavity placed around the photodetector absorption region, to enhance the responsivity. The mirrors are typically formed using Bragg reflectors or dielectric mirrors implemented in the semiconductor. A resonance is built up in the cavity at the wavelengths of incoming light for which the round-trip phase shift is 2π, which corresponds to constructive interference. Hence at the resonant wavelength, the incoming light is reflected at the two mirrors and permits the absorbing layer to be reused many times by effectively recirculating the optical signal, thus enhancing the absorption and the responsivity. However, the transit length can remain extremely small, because the photogenerated carriers only have to traverse the thickness of the absorption region to be collected. This solves the problem of the fundamental responsivity-bandwidth trade-off in normal photodetectors, in which increasing the absorption layer width to increase responsivity increases the transit time thus reducing the bandwidth. Microcavity resonant photodiodes enable the simultaneous achievement of large bandwidth and high responsivity possible. Also, the resonant

behavior leads to wavelength selectivity, which can be used to advantage in some wavelength division multiplexed WDM applications (*see* **Optical Communication Systems**: Wavelength Division Multiplexing). This approach can provide demodulation of optical signals with quantum efficiencies close to 100% and with bandwidths in excess of 100 GHz.

Optical Receivers – Preamplification and Clock Recovery

The front end of an optical receiver comprises the photodiode and a preamplifier to amplify the electrical signal for further processing. In order to obtain a high demodulation bandwidth, it is important to minimize the photodetector capacitance and to minimize the input resistance of the preamplifier that acts as the load resistance. However, because the thermal noise current is inversely proportional to the load resistance, minimizing the load resistance causes excessive noise and degraded demodulation sensitivity. An alternative approach is to use a high-impedance preamplifier. This reduces the noise at the price of a lower bandwidth. Thus there is a trade-off between the bandwidth and sensitivity. The trans-impedance front-end, in which the load resistor is connected as a feedback resistor around an inverting amplifier, is often used as an optimum compromise that combines high sensitivity together with large bandwidth. This enables the use of a large resistance, which minimizes noise, but at the same time the effect of negative feedback reduces the input impedance by a factor equal to the amplifier gain, thus increasing the bandwidth.

In digital optical transmission systems the data recovery section of the optical receiver comprises a clock recovery circuit and a decision circuit. Accurate timing (clock) information about the bit slot, in order to synchronize the decision process, is obtained from the amplified front-end waveform using a timing extraction circuit (*see* **All-Optical Signal Regeneration**). Typical clock recovery circuits comprise a high-Q resonant circuit which rings at the bit rate frequency or a phase locked loop. This enables precise data time sampling in the following decision circuit within the bit intervals of the original pulse train.

See also

All-Optical Signal Regeneration. **Fiber and Guided Wave Optics:** Dispersion; Light Propagation. **Lasers:** Edge Emitters. **Modulators:** Electro-Optics. **Optical Communication Systems:** Basic Concepts; Lightwave Transmitters; Wavelength Division Multiplexing.

Further Reading

Agrawal GP (1997) *Fiber-Optic Communication Systems*. New York: Wiley.

Coldren L and Corzine S (1995) *Diode Lasers and Photonic Integrated Circuits*. New York: Wiley.

Gowar J (1993) *Optical Communication Systems*. Englewood Cliffs: Prentice Hall.

Kazovsky L, Bendetto S and Willner AE (1996) *Optical Fiber Communication Systems*. Boston: Artech House.

Keiser GE (1991) *Optical Fiber Communications*. New York: McGraw-Hill.

Miller SE and Kaminow I (eds) (1988) *Optical Fiber Telecommunications II*. New York: Academic Press.

Personick SD (1985) *Fiber Optics: Technology and Applications*. New York: Plenum Press.

Petermann K (1991) *Laser Diode Modulation and Noise*. Boston: Kluwer Academic.

Saleh BEA and Teich MC (1991) *Fundamentals of Photonics*. New York: Wiley.

Senior J (1992) *Optical Fiber Communications*. NJ: Prentice Hall.

Siegman AE (1986) *Lasers*. Sausalito, CA: University Science Books.

N

NONCLASSICAL LIGHT

H Walther, University of Munich and Max-Planck-Institute for Quantum Optics, Garching, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The pioneering experiments by Hanbury-Brown and Twiss, and the subsequent development of the quantum theory of electromagnetic radiation by Glauber and Sudarshan, raised the interest in the question of photon statistics. It followed from their theories that states of the electromagnetic field which can only exist in a quantum treatment. These nonclassical states show up in photon statistics through the phenomenon of antibunching. Later, further nonclassical states of the radiation field were found in the so-called squeezed fields, which are fields with phase-sensitive quantum fluctuations being at certain phase angles smaller than those of a perfect classical coherent field. In order to achieve squeezing of laser radiation, various forms of nonlinear optical interactions can be used.

Classical Versus Quantum Waves

In classical electrodynamics, a monochromatic wave (e.g., an extended light beam) is often described using the complex representation:

$$E(r,t) = E^{-}(r,t) + E^{+}(r,t) = \frac{1}{2}[\varepsilon(r,t)\mathrm{e}^{-\mathrm{i}\omega t} + \varepsilon^{*}(r,t)\mathrm{e}^{\mathrm{i}\omega t}] \qquad [1]$$

In classical theory, this complex description is merely a tool for the simplification of some calculations. However, there is some physical significance when using the quantum description, where ε and ε^{*} lead to the lowering and raising operators of the field. The above field can also be represented in the following way:

$$E(r,t) = \frac{\varepsilon + \varepsilon^{*}}{2}\cos\omega t + \frac{\varepsilon - \varepsilon^{*}}{2}\sin\omega t \qquad [2]$$

with real coefficients $\varepsilon_1 = (\varepsilon + \varepsilon^{*})/2$ and $\varepsilon_2 = (\varepsilon - \varepsilon^{*})/2\mathrm{i}$. This description gives the state of the field in a two-dimensional phase diagram where the sin-component is plotted versus the cos-component. In this plot, the wave above is represented by one point giving phase and amplitude of the field. These phase diagrams will play an important role in characterizing the nonclassical fields in a quantum phase diagram, as discussed below.

The properties of a field are described by the first- and second-order correlation functions. The first-order correlation function is given by

$$G^{(1)}(r_1,t_1,r_2,t_2) = \langle \varepsilon^{*}(r_1,t_1)\varepsilon(r_2,t_2)\rangle \qquad [3]$$

It is obtained by multiplying the electric field amplitudes at two spacetime points and taking the ensemble average. For $r_1 = r_2$ and $t_1 = t_2$, $G^{(1)}$ is proportional to the intensity at this point and is, therefore, called the first-order in intensity. The first-order correlation function is connected to the spectrum $I(\omega)$ of the field through its Fourier transform.

The second-order correlation function is necessary to characterize the radiation field completely. It describes the intensity fluctuations, being the classical equivalent to photon statistics, and is given in the general form:

$$G^{(2)}(r_1,t_1,r_2,t_2,r_3,t_3,r_4,t_4) = \langle \varepsilon^{*}(r_1,t_1)\varepsilon^{*}(r_2,t_2)\varepsilon(r_3,t_3)\varepsilon(r_4,t_4)\rangle \qquad [4]$$

If $G^{(2)}$ is considered for two spacetime points $r_1 = r_2$, the connection to intensity fluctuations becomes obvious:

$$G_{\mathrm{I}}^{(2)}(\tau) = 4\langle I(t_1)I(t_2)\rangle = \langle \varepsilon^{*}(t_1)\varepsilon(t_1)\varepsilon^{*}(t_2)\varepsilon(t_2)\rangle \qquad [5]$$

The subscript I indicates that we have to deal with intensity correlations; τ stands for $t_2 - t_1$.

In connection with the Hanbury-Brown, Twiss experiment, in general, the normalized intensity correlation is given by

$$g_{\mathrm{I}}^{(2)}(\tau) = \langle I(t)I(t+\tau)\rangle / \langle I(t)\rangle^2 \qquad [6]$$

$I(t)$ stands for the intensity which is given by the mean value $\langle I(t)\rangle$ plus fluctuations $\Delta I(t)$ for which the mean value is zero. The mean value of the squared intensity is then $\langle I^2\rangle = \langle I\rangle^2 + \langle \Delta I^2\rangle$. It follows for the normalized intensity correlation function:

$$g_{\mathrm{I}}^{(2)}(\tau) = \langle I(t)I(t+\tau)\rangle / \langle I(t)\rangle^2 = \left\{ \begin{array}{ll} 1 + \langle \Delta I^2\rangle / \langle I\rangle^2 & \text{for } \tau = 0 \\ 1 & \text{for } \tau \to \infty \end{array} \right\} \qquad [7]$$

For $\tau \to \infty$, the intensities are not correlated and $g_{\mathrm{I}}^{(2)}(\tau)$ becomes 1. Otherwise $\langle \Delta I^2\rangle / \langle I\rangle^2 \geq 0$; this means that $g_{\mathrm{I}}^{(2)}(\tau)$ is always equal to or larger than 1, and $g_{\mathrm{I}}^{(2)}(\tau)$ smaller than 1 is not possible classically. The photon correlation is usually measured by the classical Hanbury-Brown, Twiss setup shown in Figure 1.

The maximum of $g_{\mathrm{I}}^{(2)}(\tau)$ at $\tau = 0$ is called intensity bunching and was first observed by Hanbury-Brown and Twiss. The decrease of $g_{\mathrm{I}}^{(2)}(\tau)$ as a function of τ is a measure of the coherence length of the light source.

When the experiment in Figure 1 is considered in the photon picture there is a principal difference compared to the classical intensity. In the classical case, the intensity on the two detectors is the same if the beamsplitter is dividing the light beam in equal portions. Both detectors, therefore, measure the same fluctuations. In the quantum picture, however, a photon can only be detected in one of the two detectors; this leads to the result that values for $g_{\mathrm{I}}^{(2)}(\tau)$ smaller than 1 can be found. The effect of photon bunching was initially thought to be an inherent property of light owing to the boson nature of photons. Therefore, it came as a big surprise when Glauber showed theoretically that for the quantized radiation field $g_{\mathrm{I}}^{(2)}(\tau)$, <1 is possible. This effect, known as photon antibunching, and its experimental observation constitutes an unambiguous proof of the quantum nature of light.

In the quantum description of the electromagnetic field, the electric field amplitudes ε and ε^* are replaced by $\hat{a}$ and $\hat{a}^+$, the photon annihilation and creation operators.

In this case, the electric field operator is written as

$$\hat{E} \propto \frac{\hat{a} + \hat{a}^+}{2} \cos\omega t \frac{\hat{a} - \hat{a}^+}{2\mathrm{i}} \sin\omega t \qquad [8]$$

In analogy, the operators $\hat{a}_1 = (\hat{a} + \hat{a}^+)/2$ and $\hat{a}_2 = (\hat{a} - \hat{a}^+)/2\mathrm{i}$, can be defined describing the amplitude of the cosine and sine parts of the wave. $\hat{a}_1$ and $\hat{a}_2$ do not commute $[\hat{a}_1, \hat{a}_2] = (\mathrm{i}/2)1$, therefore, the expectation values of the two operators $\hat{a}_1$ and $\hat{a}_2$ underlie an uncertainty relation given by

$$\langle \Delta\hat{a}_1^2\rangle = \langle \hat{a}_2^2\rangle - \langle \Delta\hat{a}_i\rangle^2 \geq 1/16 \qquad [9]$$

with $\langle \Delta\hat{a}_i^2\rangle = \langle \hat{a}_i^2\rangle - \langle \Delta\hat{a}_i\rangle^2$. In the corresponding phase diagram, the expectation values for $\hat{a}_1$ and $\hat{a}_2$ are represented by a shaded area defined by the uncertainty (Figure 2). The state of the radiation field, where both phase and amplitude fluctuations are symmetrically minimized, $\langle \Delta\hat{a}_1^2\rangle = \langle \Delta\hat{a}_2^2\rangle = \frac{1}{4}$, best approximates the classical wave with well-defined phase and amplitude. The state with symmetrical minimum uncertainty is the coherent state describing laser radiation. It is an eigenstate of $\hat{a}$ with a photon number fluctuation given by a Poissonian distribution with $\langle \Delta n^2\rangle = \langle n\rangle$.

There are other minimum uncertainty states, however, with unsymmetrical uncertainty regions

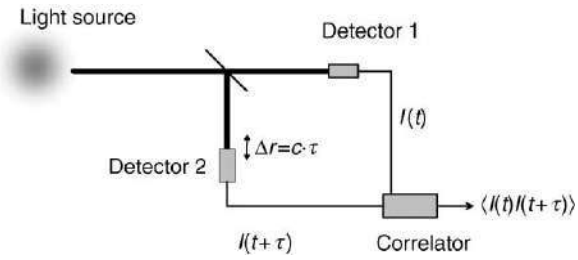

Figure 1 Experimental setup for the Hanbury-Brown, Twiss experiment.

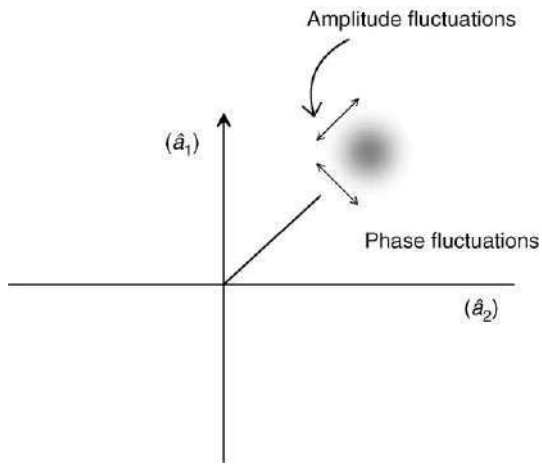

Figure 2 Phase diagram of the quantum mechanical field amplitudes $\langle \hat{a}_1\rangle$ and $\langle \hat{a}_2\rangle$ for coherent radiation.

with the uncertainty squeezed into mainly one component. These states, which are called squeezed states, are defined by $\langle \Delta \hat{a}_1^2 \rangle < 0.25$ or $\langle \Delta \hat{a}_2 \rangle < 0.25$. Squeezed state may also refer to a state which is not a minimum uncertainty state but where, nevertheless, the mean square fluctuation in one of the two components is less than 0.25.

A classical electromagnetic field consists of waves with well-defined amplitude and phase. This is not the case in a quantum treatment: fluctuations are associated with both conjugate variables. The case of a coherent state that most nearly describes a classical electromagnetic field has an equal amount of uncertainty in the two variables (when normalized to the field of a single photon). Equivalently, the field can be described in two conjugate quadrature components and the uncertainties in the two conjugate variables satisfy an uncertainty relation. The coherent state represents a minimum uncertainty state with equal uncertainties in the two quadrature components. This case is usually called the shot-noise limit; it is represented in Figure 2 as a phase diagram.

In a quantum treatment of radiation, it is also possible to generate states that are not present in the classical limit. They can show fluctuations reduced below the classical limit in one of the quadrature components, while the canonically conjugate quadrature component must display enhanced fluctuations in order to fulfill the uncertainty relation. Those states are called 'squeezed states.' An electromagnetic field with fluctuations below the standard quantum limit in one of the quadrature components has, in principle, many attractive applications, for example, in optical communication, in precision and sensitive measurements such as gravitational wave detection, or in noise-free amplification. Therefore, there has been great interest in generating squeezed radiation. Optical measurements have three characteristics that enable them to reach the quantum noise level more readily than in other fields of physics: (i) optical signals are naturally immune to external sources of noise; (ii) thermal noise at room temperature is negligible in the optical domain; and (iii) the outstanding equalities of the optical sources and detectors allow a very-low-level instrumental noise.

Squeezed states with reduced photon number fluctuations (sub-Poissonian photon statistics) correspond to a light beam in which the photons arrive more equally spaced than in a coherent beam for which $g_{\mathrm{I}}^{(2)}(0) = 1$. Therefore, sub-Poissonian number fluctuations correspond to photon antibunching. This shows that photon antibunching and squeezing are related. The state with the largest possible antibunching is the number state, represented by a circle in the phase diagram. It has large fluctuations in both components a_1 and a_2 and, therefore, is not a squeezed state. Thus, it is obvious that antibunching does not imply squeezing. On the other hand, weak squeezing of a coherent state may result in antibunching. If the state is squeezed harder the a_1 component will be better defined. The amplitude fluctuations, however, will start increasing and there will be a transition from antibunching to bunching.

The quantum calculation of the normalized second-order correlation function leads to the following result:

$$g_{\mathrm{I}}^{(2)}(0) = 1 + (\langle \Delta n^2 \rangle - \langle n \rangle / \langle n \rangle^2) \qquad [10]$$

A pure number state of the radiation field with $\langle \Delta n^2 \rangle = 0$, leads to $g_{\mathrm{I}}^{(2)}(0) < 1$, as discussed above. For a coherent state, $\langle \Delta n^2 \rangle = \langle n \rangle$ yielding $g_{\mathrm{I}}^{(2)}(0) = 1$, corresponding to the classical calculations.

It can be concluded that although antibunching and squeezing are related and nonclassical phenomena, one does not imply the other. Thus, the detection schemes necessary to determine the existence of one or the other are quite different. Antibunching or sub-Poissonian statistics can be detected via intensity correlations. Squeezing, on the other hand, is a property of either $\langle a_1 \rangle$ or $\langle a_2 \rangle$, and phase sensitive homo- or heterodyne detection is needed.

Nonclassical states of the radiation field are very sensitive to linear attenuation. Coherent and thermal light fields are the only ones that do not change their photon statistics when attenuated. This is different for nonclassical light. The change depends on the particular case and has to be carefully evaluated.

The first observation of squeezing was achieved in 1985, in an experiment of parametric generation involving four-wave mixing in sodium vapor.

Nonclassical Light from a Single Atom

Resonance fluorescence of atoms is a basic process in radiation–atom interactions, and has therefore always generated considerable interest. The methods of experimental investigation have changed continuously due to the availability of new experimental tools. A considerable step forward occurred when tunable and narrowband dye laser radiation became available. These laser sources are sufficiently intense to easily saturate an atomic transition. In addition, the lasers provide highly monochromatic light with coherence times much longer than typical natural lifetimes of excited atomic states. Excitation spectra with laser light using well-collimated atomic

beams lead to a width, being practically the natural width of the resonance transition, therefore it became possible to investigate the frequency spectrum of the fluorescence radiation with high resolution and to also study the photon statistics of the fluorescent light.

Previous experiments to investigate antibunching in resonance fluorescence have been performed by means of laser-excited collimated atomic beams. The initial results obtained in 1977 showed for the second-order correlation function $g^{(2)}(t)$, a positive slope characteristic of photon antibunching, but $g^2(0)$ was larger than $g^2(t)$ for $t \to \infty$. This was due to number fluctuations in the atomic beam and to the finite interaction time of the atoms. In 1978, the analysis of the experiment was refined and in 1982, another experiment with a longer interaction time was performed. In the latter experiment, the photon correlation was also measured for very low laser intensities. The ideal experiment to study the photon statistics in resonance fluorescence is the investigation of a trapped ion since, in such an experiment, the fluctuations of the atomic number are excluded.

The fluorescence of a single ion should also display the following property. The probability distribution of the photon number recorded in a finite time interval t is narrower than Poissonian, which means, that the variance is smaller than the mean value of the photon number. This is because the single ion can only emit a single photon. Antibunching and sub-Poissonian statistics are often associated. They are, nevertheless, distinct properties and need not necessarily be simultaneously observed, as is the case in the experiment described here. Although there is evidence of antibunching in the atomic-beam experiments, the photon counts are not sub-Poissonian as a result of fluctuations in the number of atoms. In further experiments this effect was excluded by use of a special trigger scheme for the single-atom event. In the setup with a trapped ion these precautions are not necessary since there are no fluctuations in the atomic number. Figure 3 shows the photon correlation of a single trapped ion. The measurements are performed using the $3^2S_{1/2} - 3^2P_{3/2}$ transition of the $^{24}Mg^+$-ion at a wavelength of 280 nm. The laser is tuned slightly below resonance in order to Doppler-cool the secular motion of the ion.

The investigation of the photon correlation employed the ordinary Hanbury-Brown and Twiss setup with two photomultipliers and a beamsplitter (Figure 1). The time delay τ between the photomultiplier signals was converted by a time-to-amplitude converter into a voltage amplitude proportional to the time delay. A delay line of 100 ns

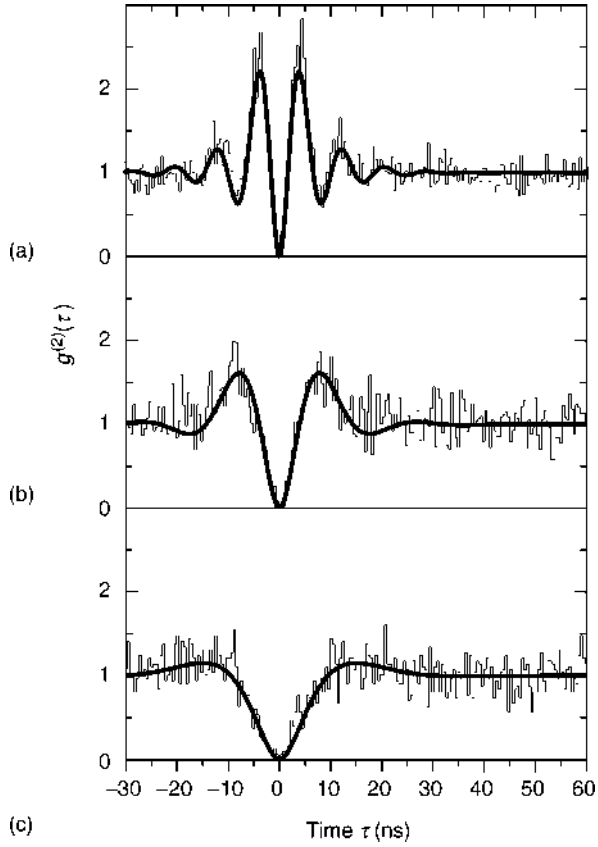

Figure 3 Antibunching signals of a single $^{24}Mg^+$-ion for different laser intensities (decreasing from (a) to (c)). At larger intensities (a) and (b) the Rabi oscillations owing to a strong coherent coupling between ground and excited state are observed.

in the stop channel allowed for the measurement of $g^{(2)}(\tau)$ for both positive and negative τ in order to check the symmetry of the measured signal. The output of the time-to-amplitude converter was accumulated by a multichannel analyzer in pulse height analyzing mode. Three typical measurements, each at small values of the laser intensity but with different detunings, are shown in Figure 3. The solid curve is a theoretical fit to the measurements. For small time delays (<20 ns), the nonclassical antibunching effect is observed, superimposed with Rabi oscillations being damped out with a time constant corresponding to the lifetime of the excited state.

Recently it also became possible to combine a trapped ion with a cavity. In this way the emission of a photon in the cavity mode can be achieved. This leads to a directed emission of the photon being more efficient than the fluorescence occurring in the entire 4π geometry. In this way, a deterministic single photon source is obtained.

Survey on the Experiments on Squeezing

In order to produce squeezed light from nonlinear optical phenomena, a variety of methods has been used, depending on the type of nonlinearity and the number of interacting modes involved. The choice of optical nonlinearity reduces to either three-wave or four-wave mixing, leading to quadratic or cubic nonlinear response functions, respectively. Each of these methods has advantages and disadvantages. Three-wave mixing requires a pump field at twice the frequency of the field-mode being squeezed. However, it has the advantage as any sidebands induced on the pump, owing to phase noise, have a large frequency offset from the squeezed fields. The other method, the much simpler four-wave mixing, does not require any frequency-doubling step. This was, therefore, the first method used to obtain squeezing with quadrature noise below the vacuum level.

Most of the early experiments on squeezing were performed with one or two cavity modes with the squeezed modes being transferred to propagating external modes via beamsplitters for measurement. This narrowband procedure is followed in more recent experiments commonly used for three-wave (parametric) squeezing. Also squeezing in waveguides was achieved leading to a relatively simple and broadband implementation of squeezing, with, mostly a four-wave mixing with short pulses or solitons being performed.

The main limiting factors that remain include control over the driving laser phase and intensity noise, absorption in the dielectric material and associated devices, and phase fluctuations caused both by low-frequency technical noise in components, and by intrinsic material properties of phonon interactions and refractive index fluctuations. It appears not impossible to reach squeezing levels of 10 dB or more below shot noise, over bandwidths of up to 1 THz. At this level of noise reduction, the main problem is the issue of improving detector technology to allow such low noise levels to be investigated.

Squeezing in Lasers

The quantum noise properties of the laser were studied intensively after its invention in the early 1960s. The conclusion of these studies is that the laser output approaches quite closely the noise properties of a coherent state, provided the laser is operated well above threshold. That is, the spectral variance of the photon number and phase fluctuations at high enough Fourier frequencies will be at the Poissonian or quantum noise limit (also referred to as the shot noise level). On the other hand, the phase of the laser is unconstrained and drifts as a function of time, so-called laser phase diffusion. As a result, the phase fluctuations show excess noise at low frequencies.

In the early 1980s, it was realized that if the pumping mechanism for the laser could be made sub-Poissonian then amplitude squeezing could be obtained at low frequency. A few years later it was shown that sub-Poissonian pumping could be realized in semi-conductor lasers by regularizing the pump current. An experimental demonstration of amplitude squeezing from semi-conductor lasers was soon achieved. Although still requiring rather specialized lasers, many groups now routinely obtain squeezing from semi-conductor lasers.

There are a number of disadvantages to semi-conductor lasers, including their poor beam quality and short coherence length. As a result, interest in other possible squeezing mechanisms continues. It has been found that intrinsic dynamical effects can also lead to amplitude squeezing in the laser output.

Laser squeezing is of a different character to that produced by more 'traditional squeezers.' For example, in parametric amplification, the de-amplification of phase fluctuations is accompanied by an equal and opposite amplification of amplitude fluctuations which preserves the uncertainty relationship between them. However, in laser squeezing the reduction in amplitude fluctuations is not linked to any increase in phase fluctuations. Thus, in laser squeezing we take advantage of the fact that the laser output is not in a minimum uncertainty state to suppress amplitude noise. This means that laser squeezing is always amplitude or intensity squeezing.

Application of Nonclassical Light

The applications of nonclassical light are numerous. The single photon sources discussed above have their main applications in quantum communication with single photons where they may be used to establish secure signal transmission. There is also a decisive advance in the field of quantum information expected from the combination of local atomic quantum processing with photonic transmission of quantum states. Conceptually, new possibilities would arise to enhance quantum computational capabilities. For example, separate local quantum processors of moderate size could be linked in a quantum network, with quantum information being exchanged by means of photons.

Although a large number of phenomena have been predicted in which the quantum nature of the squeezed vacuum plays an essential role in modifying

atomic properties, only the linear dependence of the two-photon absorption rate of a three-level ladder atom has been confirmed experimentally but classical sources show a quadratic dependence. In 1998 an important series of experiments was reported which investigated the interaction of a beam of Cs atoms with squeezed light in the Fabry–Perot cavity environment. The principal phenomena examined were the modification of the transmission spectrum of a weak, tunable probe beam due to the presence of the squeezed light, and the variation of the transmission modulation with the phase of the squeezed vacuum relative to that of an applied, saturating, coherent field.

There have been many proposed practical applications of squeezed light, and some experiments have already demonstrated its potential in precision measurements and improvement of spectroscopic sensitivity. Squeezed light may be used in optical communications to enhance signal-to-noise ratio, gravitational wave detection, and quantum nondemolition measurements.

The wider applications of squeezed light are numerous. An idea of the range and variety of applications is provided by the following selection: the protective measurement of a single squeezed harmonic oscillator state, coherent nonlinear spectroscopy using number squeezed light, the improvement of quantum nondemolition measurements using a squeezed meter input, sub-shot-noise Doppler anemometry with amplitude squeezed light, noiseless transfer of nonclassical light through bistable systems, generation of few photon states from squeezed atoms, and spin squeezing in a collection of atoms illuminated with squeezed light.

Einstein–Podolsky–Rosen (EPR) Correlations and Entanglement of Photons

In 1986, a quantum nondemolition experiment confirming the quantum correlation of the quadrature phase amplitudes of two spatially separated fields, was reported. In 1987, two-mode squeezing with two spatially separated detectors was experimentally demonstrated using parametric oscillation. With the experimental realization of such two-mode squeezing and quantum nondemolition measurements, the possibility of testing for continuous variable EPR-type correlations became apparent. In these proposals, the conjugate 'position' and 'momentum' observables would be the two orthogonal quadrature phase amplitudes of the field. The high efficiency of detectors used in the quadrature phase amplitude squeezing experiments was an important issue in this connection. Experimental tests of EPR correlations at the time were based on Bohm's version of the EPR paradox, and in fact, focused on Bell's extension of this work. These tests involved measurements performed on individual photon pairs, and used very inefficient detectors.

A specific proposal and a criterion to demonstrate the continuous variable EPR correlations for real experiments by way of a realization of the 1935 EPR paradox, was put forward in 1989. This proposal employed the two-mode squeezed state as the EPR source. The first experimental achievement, using the parametric oscillator, of the 1989 EPR criterion, for efficient measurements with continuous variable outcomes, was reported in 1992.

The EPR fields have proved significant in enabling the experimental realization of continuous variable quantum teleportation, and may have application also to quantum cryptography. Again, because of the efficient nature of the measurement of field quadrature phase amplitudes, and the intrinsic macroscopic nature of the correlated fields, continuous variable teleportation and cryptography may provide large advantages over protocols using a finite number state basis, where signals of individual photons are sent. At the moment, several groups are working in this promising direction.

Conclusions

In this contribution the generation of nonclassical light was reviewed. The generation of nonclassical light became possible after stable and reliable laser sources were available. There are many sources of nonclassical light available now. It is clear that we will see many new applications in the future. Especially in connection with optical communication, rather interesting improvements can be expected.

See also

Coherent Lightwave Systems. **Photonic Crystals:** Electromagnetic Theory. **Quantum Optics:** Atomic Coherence Effects.

Further Reading

Caves CM and Drummond PD (1994) Quantum limits on Bosonic communication rates. *Review of Modern Physics* 44: 481–537.

Cresser JD, Häger J, Leuchs G, Rateike FM and Walther H (1982) Resonance fluorescence of atoms in strong monochromatic laser fields. *Topics in Current Physics* 27: 21–59.

Dagenais M and Mandel L (1978) Investigation of two-time correlations in photon emissions from a single atom. *Physical Review A* 18: 2217–2228.

Diedrich F and Walther H (1987) Nonclassical radiation of a single stored ion. *Physical Review Letters* 58: 203–206.

Drummond PD and Ficek Z (eds) (2004) Springer Series on Atomic, Optical, and Plasma Physics, vol. 27. Berlin, Heidelberg, New York: Springer Verlag.

Gatti A, Lugiato LA and Wiedemann H (1997) In: Giacobino E, Reynaud S, and Zinn-Justin J (eds). *Quantum Fluctuations*, p. 431. Amsterdam: Elsevier.

Georgiades NP, Polzik ES, Edamatsu K, Kimble HJ and Parkins AS (1995) Nonclassical excitation for atoms in a squeezed vacuum. *Physical Review Letters* 75: 3426–3429.

Glauber R (1963) The quantum theory of optical coherence. *Physical Review* 130: 2529–2539.

Glauber R (1963) Coherent and incoherent states of the radiation field. *Physical Review* 130: 2766–2788.

Guthöhrlein GR, Keller M, Hayasaka K and Lange W (2001) A single ion as a nanoscopic probe of an optical field. *Nature* 414: 49–51.

Heidmann A, Horowicz RJ, Reynaud S, Giacobino E, Fabre C and Camy G (1987) Observation of quantum noise reduction on twin laser beams. *Physical Review Letters* 59: 2555–2557.

Höffges JT, Baldauf HW, Eichler T, Helmfrid SR and Walther H (1997) Heterodyne measurement of the fluorescent radiation of a single trapped ion. *Optical Communications* 133: 170–174.

Hood CJ, Kimble HJ, Turchette QA, Georgiades NPh and Parkins AS (1998) Squeezed excitation in cavity QED: experiment and theory. *Physical Review A* 58: 4056–4077.

Jakeman E, Pike ER, Pusey PN and Vaugham JM (1977) Effect of atomic number fluctuations on photon antibunching in resonance fluorescence. *Journal of Physics A: Mathematical and General* 10: L257–L259.

Kimble HJ, Dagenais M and Mandel L (1977) Photon antibunching in resonance fluorescence. *Physical Review Letters* 39: 691–695.

Kimble HJ, Dagenais M and Mandel L (1978) Multiatom and transit-time effects on photon-correlation measurements in resonance fluorescence. *Physical Review A* 18: 201–207.

Kimble HJ, Carnal O, Hu Z, Mabuchi H, Polzik ES, Thompson RJ and Turchette QA (1995) *Annals of the New York Academy of Sciences* 755: 87.

Levenson MD, Shelby RM, Reid MD and Walls DF (1986) Quantum nondemolition detection of optical quadrature amplitudes. *Physical Review Letters* 57: 2473–2476.

Loudon R (2000) *The Quantum Theory of Light*, 3rd edn. Oxford, New York: Oxford University Press.

Ou ZY, Pereira SF, Kimble HJ and Peng KC (1992) Realization of the Einstein–Podolsky–Rosen paradox for continuous variables. *Physical Review Letters* 68: 3663–3666.

Reid MD (1989) Demonstration of the Einstein–Podolsky–Rosen paradox using nondegenerate parametric amplification. *Physical Review A* 40: 913–923.

Short R and Mandel L (1983) Observation of sub-Poissonian photon statistics. *Physical Review Letters* 51: 384–387.

Short R and Mandel L (1984) Sub-Poissonian photon statistics in resonance fluorescence. In: Mandel L and Wolf E (eds). *Coherence and Quantum Optics*, vol. V, pp. 671–678. New York: Plenum.

Slusher RE, Hollberg LW, Yurke B, Mertz JC and Vally JF (1985) Observation of squeezed states generated by four-wave mixing in an optical cavity. *Physical Review Letters* 55: 2409–2412.

Walls DF (1979) Evidence for the quantum nature of light. *Nature* 280: 451–453.

Wu LA, Kimble HJ, Hall J and Wu H (1986) Generation of squeezed states by parametric down conversion. *Physical Review Letters* 57: 2520–2523.

Yamamoto Y and Machida S (1987) High-impedance suppression of pump fluctuation and amplitude squeezing in semiconductor lasers. *Physical Review A* 35: 5114–5130.

NONLINEAR OPTICS AT THE CRITICAL FIELD LIMIT

G A Mourou, University of Michigan, Ann Arbor, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

It is now possible to produce, with compact lasers, intensities so large that the relativistic character of the electrons dominates laser–matter interactions. This is a fundamentally new regime in optics where optics, confined to the study of eV characteristic phenomena, are now addressing phenomena with characteristic energies vastly greater in the MeV–GeV (TeV in the near future) range. With relativistic optics, the field of nonlinear optics found a number of novel applications, including fast ignition of inertially

compressed fusion target, bright sources of energetic electrons, protons, ions, positrons, pions, etc., but also nuclear physics, high energy physics, astrophysics, and general relativity.

Relative optics has been one of the most active fields of optics and it is impossible to describe all the applications here. So, this new regime will be illustrated with few examples of the key relativistic mechanisms – reminiscent of conventional nonlinear optics, such as relativistic harmonics, relativistic optical rectification (wake-field acceleration), and relativistic self-focusing in refraction and reflection. We finally show that relativistic optics offers a new and efficient gateway to attosecond physics and nonlinear quantum electrodynamics (QED).

Optics is the study of how electrons respond to light. All material optical properties are the consequence of how light interacts with electrons in materials. Because, in the visible regime the photon energy is of the order of 1 eV, optics deals with phenomena with eV characteristic energy. Light is a wave composed of coupled electric and magnetic fields oscillating in synchrony at high frequencies. The electric and magnetic field oscillate perpendicular to each other, and perpendicular to the direction of propagation, k. Because of the electric field E, electrons under the action of the Lorentz force, $F = eE$, moves a distance $x = a_1 E$ transversally to the propagation direction. In classical optics, the oscillation amplitude x is small compared to the wavelength, and the electrons oscillate and radiate at the laser frequency. The radiated field amplitude and phase depend on how the electron is bound to the atom of the material. The invention of the laser in 1960 gave access to much larger intensities and electromagnetic fields. The laser field became strong enough that the electron displacement, x, in this field could no longer be considered as strictly proportional to the laser field. This departure from linearity is addressed by the addition of new terms in the Taylor series expansion of the displacement:

$$x = a_1 E + a_2 E^2 + a_3 E^3 + \cdots \qquad [1]$$

These additional terms are responsible for all the effects that form the field of nonlinear optics. Note here that the field is still not strong enough to ionize an electron from its parent atom and is therefore below the field that could irreversibly damage the material. Thus, the typical intensity must be kept to a value of less than $\sim 10^{14}$ W/cm^2, where the quiver energy E_Q starts to challenge the ionization energy of a few eV, depending on the material:

$$E_Q = \left\langle \frac{e^2 E^2}{4m\omega^2} \right\rangle \qquad [2]$$

The bracket denotes the time average over one cycle and ω is the laser angular frequency. Above this limit, the material response is dominated by multiphoton ionization or tunneling, as shown in Figure 1. During the ionization event, high harmonics are produced. Note here again that because the electron is still bound to its nucleus during the emission, we can consider the process of high harmonic generation as bound-electron nonlinear optics. As the intensity increases from 10^{14} to 10^{18} W/cm^2 (Figure 1), we are entering the regime of relativistic intensity or relativistic nonlinear optics. This is a regime that was predicted and discussed in the 1960s but that could be accessed only after the technique of chirped pulse amplification (CPA) was demonstrated, followed by significant progress in the areas of short pulse generation and pulse focusing.

Relativistic Optics

In classical optics one always assumes that during a laser oscillation period the electron mass m_0 stays constant. It is a reasonable assumption since the quiver velocity v of the electron is small compared to the speed of light c. As the intensity is increased a point is reached where the electron quiver velocity approaches the speed of light. This has several implications; first, the mass of the electron changes to become

$$m = \gamma m_0 \qquad [3]$$

where γ is the Lorentz factor given by

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-1} \qquad [4]$$

Second, because v is close to c, the magnetic force component needs to be added and the Lorentz force becomes:

$$\vec{F} = e\vec{E} + \frac{\vec{v}}{c} \times \vec{B} \qquad [5]$$

Until recently, in classical optics and laser physics, the $v \times B$ component was totally ignored. However, in the relativistic regime, this component is as large as

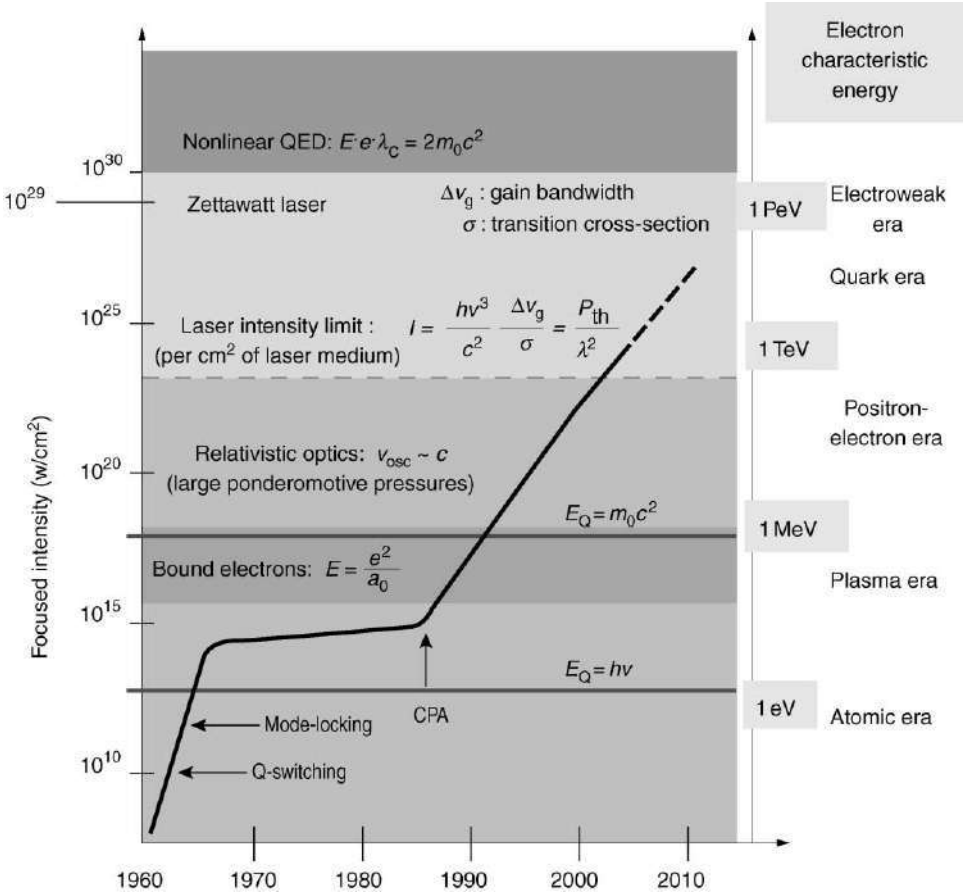

Figure 1 Laser intensity versus years. Note the very steep slope in intensities that occurs during the 1960s. This period corresponds to the discovery of most nonlinear optics effects. Since 1985 after the CPA invention followed by the successive laser short pulse improvements, we are experiencing a similar increase in intensity opening a new regime in optics at 10^{18} W/cm^2 dominated by the relativistic character of the electrons.

the electric one. It pushes the electron forward and is the tenet of the relativistic regime.

At this point, it is convenient to introduce the normalized vector potential:

$$a_0 = \frac{e\lambda E}{m_0c^2} \qquad [6]$$

which represents the energy gain by the electron over one period (λ) normalized to the electron rest mass energy. At $a_0 = 1$, corresponding to 10^{18} W/cm^2 for 1 μm wavelength light, the laser–matter interaction is strongly dominated by the relativistic character of the electrons. At these intensities, the motion of the electrons becomes relativistic and the trajectory of the electron dramatically changes. Figure 2 depicts the difference between the two regimes. It can be shown that independently of the intensity regime, the electron transverse motion is always proportional to a_0, while the longitudinal motion is always proportional to a_0^2. This means that for classical

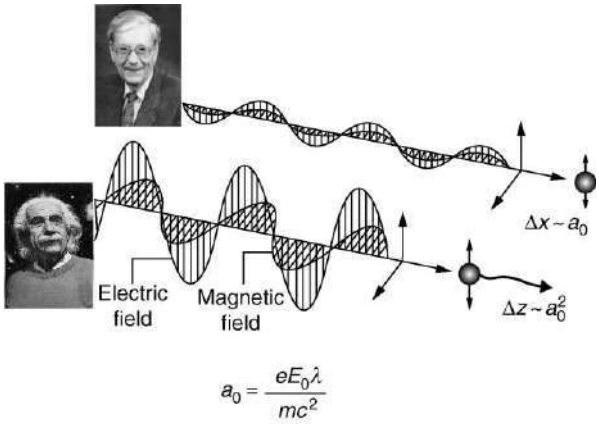

Figure 2 A comparison between classical optics and relativistic optics. The transverse electron displacement scales as a_0 and the longitudinal scales as a_0^2. (a) In classical optics $a_0 < 1$ the transverse electron displacement is proportional to a_0. It is larger than the longitudinal proportional to a_0^2. (b) In relativistic optics it is the opposite, $a_0 > 1$ the transverse displacement becomes smaller than a_0^2.

optics or even bound-electron nonlinear optics with a_0 varying from 10^{-10} to 10^{-4}, the longitudinal motion is $\sim 10^{-10}$ to 10^{-4} smaller than the transverse motion. Of course, the situation reverses completely for $a_0 > 1$, where the longitudinal motion becomes larger than the transverse motion. The departure of the electron motion from the linear regime, due to its relativistic character, gives rise to a new kind of nonlinear optical effects that constitutes the field of relativistic nonlinear optics. This area has become one of the most active in laser physics. In many cases, there are strong similarities between conventional and relativistic optics, and it is the purpose of this article to illustrate the field of relativistic optics by using some of the effects well established in bound electron nonlinear optics.

Considering the rapid progress in laser intensities that the field experienced in the 1960s, as indicated in Figure 1, scientists expected to observe the manifestation of relativistic effects in the 1970s. Formidable intensities on the order of 10^{18} W/cm^2 for 1 μm wavelength light were necessary, however, and could not be readily obtained. Such intensity levels could only be produced in the early 1980s with large-scale sources such as the Helios laser at Los Alamos National Laboratory (LANL) (Figure 3a). Work in relativistic optics virtually could not be conducted in university research laboratories, but rather only in government laboratories. The first relativistic harmonics produced by the critical surface nonlinear motion were observed in 1980 with a kJ-pulse-energy laser at mHz repetition rates (one shot/20 m).

The advent of CPA (Figure 4) revolutionized the field of relativistic optics, placing it within reach of more researchers through enormous reductions in laser size and cost. Further advances in the ultrashort-pulse duration of lasers and in laser focusing made relativistic optics more accessible, especially to universities. Relativistic intensities can be achieved with mJ energy levels at kHz repetition rates by university-size lasers (Figure 3b) instead of with large, kJ, sub-Hz-repetition-rate government-type lasers. Figures 3a and b illustrate the revolution in the relativistic laser evolution representing the Helios (LANL) system. Figure 3b depicts the relativistic λ^3 laser target chamber. The λ^3 laser reaches its relativistic intensity (10^{18} W/cm^2) with pulse energy only on the order of 1 mJ and pulse duration on the order of 10 fs, i.e., a few light periods. Its energy is concentrated on a single wavelength spot size using an f#1 paraboloid corrected by a deformable mirror. Note that all the energy is contained in a focal volume of a few λ^3, hence its name.

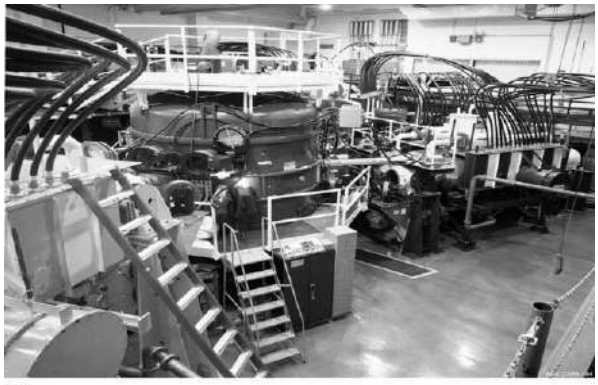

(a)

(b)

Figure 3 Two relativistic lasers. (a) Helios at LANL, was the first relativistic laser with $a_0 \sim 1$ at mHz repetition rate. (b) The λ^3 at the University of Michigan has an $a_0 \sim 1$ at kHz repetition rate.

Some Examples of Relativistic Nonlinear Optics

We will illustrate the field of relativistic optics by comparing relativistic effects with some well-known nonlinear effects.

(a) Relativistic optical rectification (laser wake-field acceleration of electrons and ions)

One well-known bound-electron optical effect coming from the second-order term in eqn [1] is optical rectification. This effect, which takes place in noncentrosymmetric crystals, gives rise to a DC field transverse to the propagation direction. In the relativistic regime, a similar effect known as wake-field acceleration will also contribute to rectification of the optical field. We have seen that in vacuum, electrons are strongly pushed forward by $v \times B$ force. In a plasma, the electromagnetic wave will also push

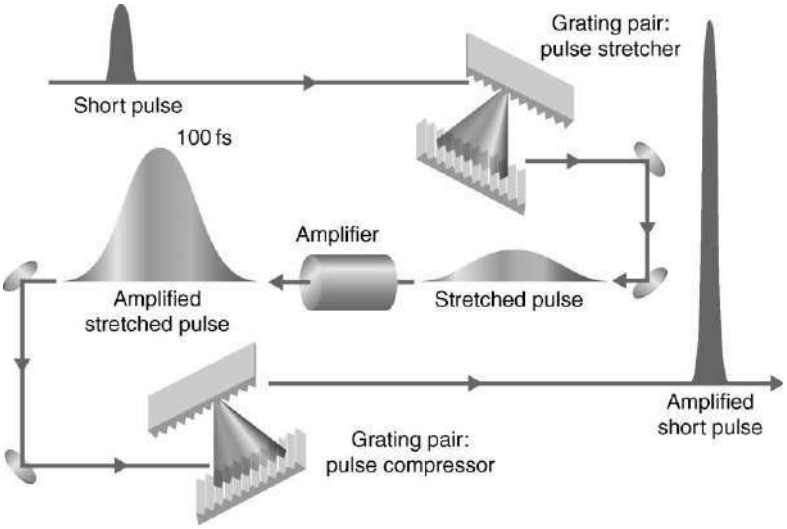

Figure 4 The concept of CPA is the key technique to get relativistic intensity. An initial short pulse is first stretched out (chirped) by a factor $\sim 10^4$, for instance by a pair of diffraction gratings. The stretched pulse intensity becomes lowered by the same factor of $\sim 10^4$. The stretched pulse intensity can then be amplified to a level 10^4 times higher. Once the laser energy has been extracted from the amplifier the stretched pulse is recompressed by the same factor ($\sim 10^4$) to its initial value. This way the laser intensity has been boosted to 10^4 times the peak intensity that the amplifier could have withstood.

forward the electrons, dragging the much heavier positive ions behind, as shown in Figures 5a and b. The charge separation that ensues will produce a strong electrostatic field that can be of the order of the transverse light electric field, which is huge. Unlike in classical optical rectification, in the relativistic regime the effect can be very efficient. Of the order of 10% or more of the transverse field may be rectified. For an intensity of 10^{18} W/cm^2, the electric field can be on the order of 10^{12} V/m (CERN on one meter), or 10^5 times larger than the field that can be obtained by more conventional technology. In addition, the rectified field is flipped along the longitudinal direction. Note also that since the field is created within a plasma, the medium cannot be broken down further and can support these extremely high fields. This effect had to wait for the relativistic-intensity lasers to become a reality before being demonstrated. Today energies up to 200 MeV have been demonstrated by using a 50 TW laser focused on a gas jet. Because the electrons drag the ions, a beam of ions is also observed concomitant with the electron beam. The generation of protons with energy up to 56 MeV has been reported by LLNL.

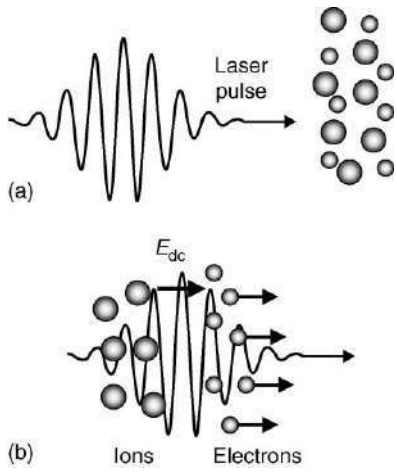

Figure 5 Relativistic rectification in plasma. (a) The high intensity pulse before entering the plasma. (b) Illustrate the $v \times B$ pushing first the plasma electrons. The electrons then drag the heavy ions behind them like a horse pulling a cart. The longitudinal electrostatic field that is generated is almost as large as the transverse laser field.

(b) Relativistic harmonic generation

Harmonic generation in the relativistic regime can be produced in two ways, as a result of the laser plasma interaction with electron density n_e, either below or above the critical density n_c. The first is for $n_e < n_c$ and corresponds to an underdense plasma where light can propagate. Harmonics generation was predicted in the late 1960s. Even-order harmonics are observed with linear and circular polarizations. This is the signature of relativistic effects since even-harmonics cannot occur in conventional nonlinear optics in centrosymmetric media.

Harmonic generation can also be produced in over-critical plasmas, where $n > n_c$. In this case, the light cannot propagate and be reflected. The strong $v \times B$ oscillatory force applied to the critical surface will produce a self-modulated reflected beam. This effect was first observed using the Helios laser. The idea of the moving-mirror model was first put forward in the early 1980s. Generation of frequencies up to the 70th harmonic has been demonstrated so this method could provide an efficient way to produce high harmonics. This technique has the advantage over conventional high harmonic generation in that it is capable of accommodating a pulse of large energy, since one can focus the beam over a small area, limited only by the laser wavelength, without fear of disrupting the plasma.

(c) Relativistic self-focusing in refraction

In bound-electron nonlinear optics, the third-order term in eqn [1] is responsible for self-focusing. A similar effect also exists in relativistic optics; it is described in Figure 6a. In the relativistic regime, the electron driven by the laser field changes its mass. In turn, this mass change will modify the index of refraction according to:

$$n^2 = 1 - \left(\frac{\omega_p}{\omega}\right)^2 \qquad [7]$$

where the plasma frequency is given by

$$\omega_p^2 = \frac{e^2}{m} n_e \qquad [8]$$

where n_e is the plasma electron density, and m is the intensity-dependent mass according to eqn [1].

For $\omega > \omega_p$, the index of refraction is positive and the pulse can propagate in the plasma. For the higher-beam-intensity part of the beam, i.e., on the beam axis, the plasma frequency will be less than the frequencies corresponding to the pulse wing. This will cause a modification of the laser wavefront and the beam to self-focus.

(d) Relativistic self-focusing in reflection and isolated attosecond pulse generation

For $\omega < \omega_p$, (eqn [7]) is negative and the pulse cannot propagate and so is reflected at the critical surface. However, the large light pressure will push the plasma critical-density surface at twice the laser frequency, making it cancel out the beam spatial intensity profile. In the relativistic λ^3 regime, the laser is focused at one wavelength. If the light is linearly polarized, the electron density will be collectively moved back and forth along the critical surface. This time-dependent deformation of the critical surface will produce a deflection of each electromagnetic half-cycle in a specific direction (Figure 7), providing a simple means for generating isolated attosecond pulses efficiently. The relativistic motion of the critical surface compresses (relativistic Doppler) the incident femtosecond pulse to the attosecond regime. To understand this compression, recall that the incident pulse travels opposite to the mirror motion formed by the critical surface (for a fraction of the cycle). As we have seen in this regime the critical surface moves also at close to the speed of light, v. Since the reflected light must also travel at c, it has no choice but to be compressed, producing an electromagnetic accumulation along the critical surface. The compression factor will

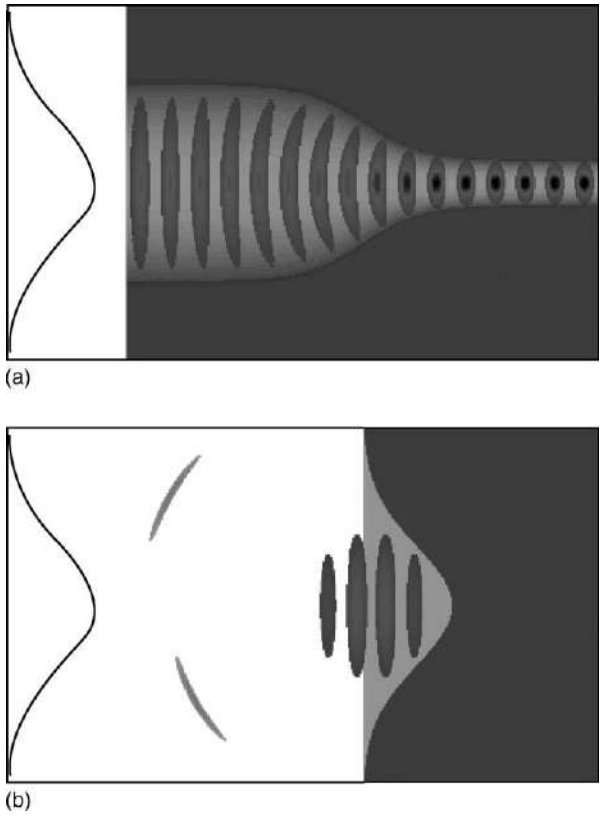

Figure 6 Relativistic self-focusing. (a) Relativistic self-focusing in refraction. Here the index of refraction is modified by the change in the electron mass. The electron mass change will in turn modify the plasma frequency across the beam and consequently the index of refraction. (b) Relativistic self-focusing in reflection showing the deformation of the critical surface by the light pressure from a strongly focused beam. The light is reflected in a series of ultrashort pulses in the attosecond domain.

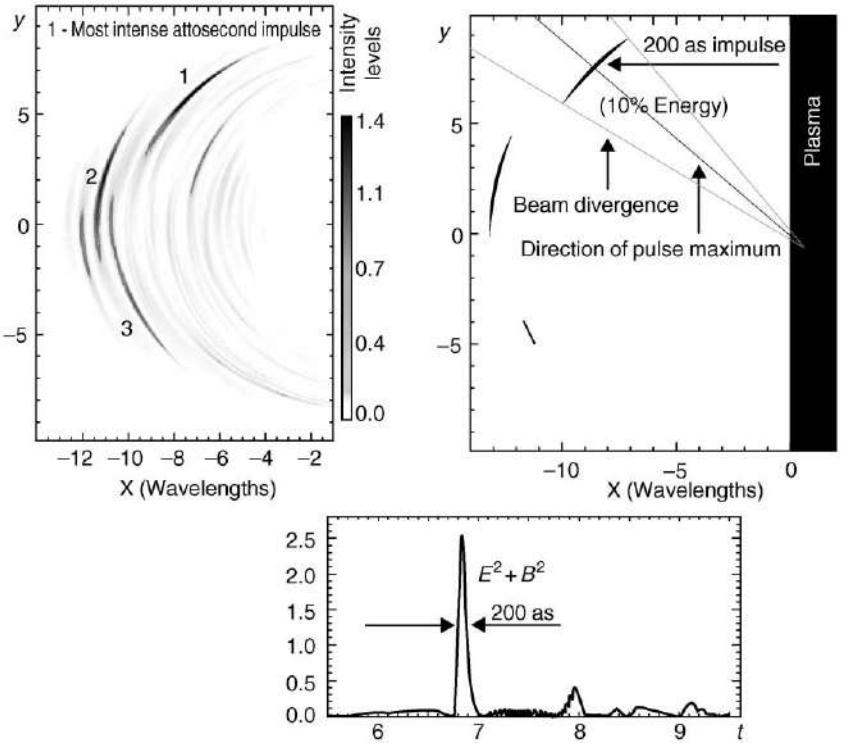

Figure 7 Compressed and deflected attosecond pulses. The combination of relativistic deflection and compression leads to 200 as isolated pulses.

depend on the critical surface velocity and is dependent on the plasma stiffness, i.e., the plasma density. Low-density plasmas are softer than high-density ones; here, n_c is the plasma critical density. The compressed pulse duration has been determined by a particles-in-cell (PIC) simulation to be given by $\tau = 600/a_0$. From the PIC simulation, it can be determined that the compressed pulse optimum duration is obtained for $n_e = a_0 n_c/2$. This technique is an interesting alternative to the one based on high-harmonic generation due to bound electrons, which has been recently demonstrated. The advantage of the relativistic optics approach is that it can be scaled up to large energies, from the mJ to the several-joule level, in contrast with demonstrated attosecond schemes that can only compress mJ pulse energies with an efficiency of 10^{-6}. For $a_0 \sim 100$, corresponding to incident intensities of $\sim 10^{22}$ W/cm^2, a 10-as pulse can be produced by relativistic compression and deflection.

(e) Nonlinear QED on a table top

Looking at Figure 1, we see that there is a regime around 10^{29} W/cm^2 where it becomes possible to produce electron–positron pairs. The condition to fulfill this is given by

$$E_s \lambda_c e = 2m_0c^2 \qquad [9]$$

Here $\lambda_c = h/m_0c$ is the Compton wavelength, E_s the Schwinger field, and h the Planck constant. This expression states that for pair creation the work of the laser field over a Compton wavelength must be equal to the electron–positron rest-mass energy. This field E_s, also known as the critical field, $E_s \sim 10^{16}$ V/cm, corresponds to a laser intensity of 10^{29} W/cm^2.

An alternative way to depict the process of electron–positron generation from vacuum is by considering the vacuum as a dielectric medium with a gap $E_G = 2m_0c^2$. The field to produce an electron–positron pair is the one necessary to bend the bands so that an electron can tunnel from the Dirac sea into the vacuum states (Figure 8). The probability for such an event is given by

$$W \propto \exp\left(-\frac{\pi E_s}{E}\right) \qquad [10]$$

Today, the intensity necessary to reach pair creation is more than six orders of magnitude higher

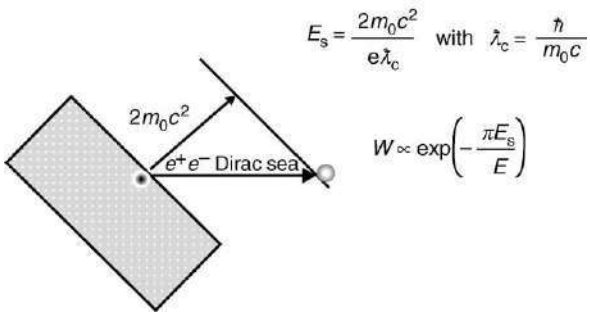

Figure 8 Pair creation. Electron tunneling from the Dirac sea into the vacuum state.

than what we could produce, $\sim 10^{22}$ W/cm^2. To bridge this enormous gap the electric field has been enhanced by colliding a short and intense pulse with a Stanford Linear Accelerator Center (SLAC) super-relativistic, 50 GeV electron beam. The Lorentz factor is $\gamma = 10^5$, and the laser field of 210^{10} V/cm (corresponding to an intensity of 10^{18} W/cm^2) can be enhanced by 10^5 to reach the critical field.

The progress in relativistic compression described above could make possible the generation of electron–positron pairs directly from the laser. We are on the verge of having a tabletop laser that could produce an intensity close to 10^{23} W/cm^2, corresponding to $a_0 > 100$. From the scaling law described previously, we should get $\tau < 6$, where the wavelength of such a compressed pulse would also be reduced by ~ 100. The focal volume would then be reduced by six orders of magnitude. If we assume an efficiency of 10^{-1}, an intensity close to the critical intensity of 10^{29} W/cm could be reached with a tabletop laser and boil vacuum.

Conclusion

Today, laser intensities are such that during light–matter interaction, the electron motion is dominated by relativistic effects responsible for a new type of nonlinear optics called relativistic nonlinear optics. Relativistic nonlinear optics is characterized by extreme field gradients (TV/m) and light pressure in the Gbar regime. Relativistic nonlinear optics offers an attractive and novel way for efficient attosecond pulse generation with intensities near the critical (Schwinger) field, opening optics to vast new horizons.

See also

Relativistic Nonlinear Optics.

Further Reading

Agostini P, Barjot G, Bonnal JF, Mainfray G and Manus C (1968) *IEEE Journal of Quantum Electronics* QE-4: 667.
Albert O, Wang H, Liu D, Chang Z and Mourou G (2000) *Optics Letters* 125: 1125.
Bloembergen N (1965) *Nonlinear Optics*. New York: Benjamin.
Boyd RW (1992) *Nonlinear Optics*. San Diego: Academic.
Brabec T and Krausz F (2000) *Review of Modern Physics* 72: 545.
Brezin E and Itzykson C (1970) *Physics Review D* 2: 1191.
Bula C, McDonald KT, Prebys EI, *et al.* (1996) *Physics Review Letters* 76: 3116.
Bulanov SV, Naumova NM and Pegoraro F (1994) *Physics Plasmas* 1: 745.
Carman RL, Benjamin RF and Rhodes CK (1981) *Physics Review A* 24: 2649.
Chen S-Y, Maksimchuk A, Esarey E and Umstadter D (2000) *Physics Review Letters* 84: 5528.
Eberly J (1969) *Progress in Optics* 7: 361.
Franken PA, Hill A, Peters C and Weinreich G (1961) *Physics Review Letters* 7: 118.
Hentschel M, Kienberger R, Spielmann Ch, *et al.* (2001) *Nature* 414: 509.
Joshi C and Katsouleas T (2003) Plasmas accelerators at the energy frontier and on tabletops. *Physics Today* 56: 47.
Litvak AG (1969) *Soviet Physics JETP* 30: 344.
Malka V, Fritzler S, Lefebvre E, *et al.* (2002) Electron acceleration by a wakefield forced by an intense ultra-short laser pulse. *Science* 22: 298.
Max C, Arons J and Langdon A (1974) *Physics Review Letters* 33: 209.
Mourou GA, Barty CPJ and Perry MD (1998) *Physics Today* 51(1): 22.
Naumova NM, Nees JA, Sokolov IV, Hou B and Mourou GA (2004) Relativistic generation of isolated attosecond pulses in a λ^3 focal volume. *Physics Review Letters* 92: 063902.
Norreys PA, *et al.* (1996) *Physics Review Letters* 76: 1832.
Sarachik E and Schappert G (1970) *Physics Review D* 1: 2738.
Shen YR (1984) *The Principles of Nonlinear Optics*. New York: Wiley.
Snavely RA, *et al.* (2000) *Physics Review Letters* 85: 2945.
Strickland D and Mourou G (1985) *Optics Communications* 55: 219 and 447.
Tajima T and Dawson JM (1979) *Physics Review Letters* 43: 262.
Tarasevitch A, Orisch A, von der Linde D, *et al.* (2000) *Physics Review A* 62: 023816.
Umstadter D, Chen S-Y, Maksimchuk A, Mourou G and Wagner R (1996) *Science* 273: 472.

NONLINEAR OPTICS, APPLICATIONS

Contents

Phase Matching
Pulse Compression via Nonlinear Optics
Raman Lasers
Self-Focusing and Related Effects (Solitons and Multiphoton Absorption)
Three-Dimensional Microfabrication

Phase Matching

A V Smith, Sandia National Laboratories, Albuquerque, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Phase matching is important in any process that coherently transfers energy from one optical mode to another over a long interaction length. We will start by illustrating the principle, by imagining that the original mode is a single transverse mode of a waveguide, and that microscopic scattering centers distributed uniformly throughout the waveguide weakly scatter light into a second mode with orthogonal polarization. Then suppose that the scattering process preserves phase so the phase of the wavelet scattered at each point into mode two is the same as that of the original wave in mode one. Such a process is shown in Figure 1 where the propagation constant of the initially excited mode is k_1 and the propagation constant of the second mode is k_2. At each z location the light scattered into the second mode has the phase $\phi = k_1 z$, and after propagating to the end of the waveguide ($z = L$) with the phase velocity characteristic of the second mode, it has the phase $\phi_z = [k_1 z + k_2(L - z)] = [k_2 L + z(k_1 - k_2)]$. Coherently summing the fields scattered at all of the z positions, gives the net output field in the second mode. Figure 2 shows how this field can be constructed graphically. The rotation angle of each phasor represents its phase, ϕ_z, while the length of each phasor represents the magnitude of the field contribution from the corresponding z segment. It is apparent that the energy transfer is maximized if the contributions from all z positions have the same phase, in which case the phasors form a straight line, and the power in the second mode is proportional to L^2. This corresponds to $(k_1 - k_2 = 0)$. When this condition is met the waves in the two

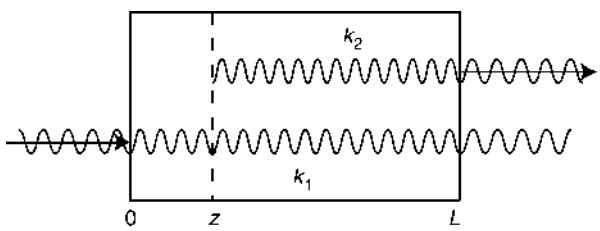

Figure 1 An input wave with propagation constant k_1 in a waveguide mode is weakly scattered into a second waveguide mode with propagation constant k_2. The phase mismatch is $\Delta k \equiv k_2 - k_1$.

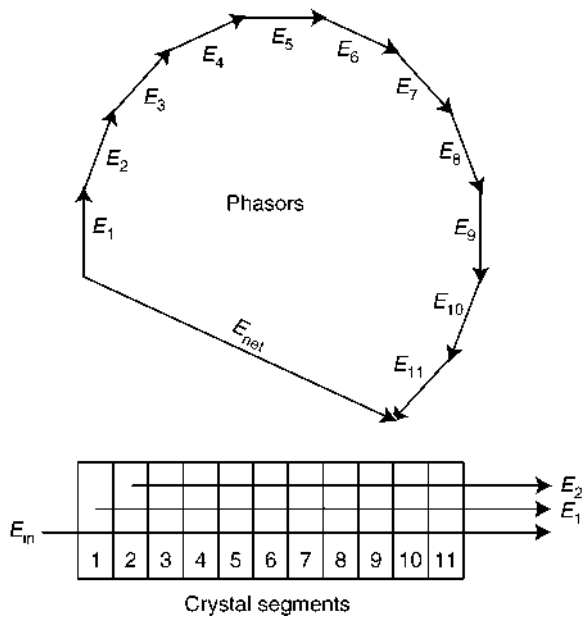

Figure 2 Phasor diagram showing the net output field contributions from segments 1–11. The net field at the output end, E_{net}, is the vector sum of the phasors. The phasors are complex quantities whose angles represent their relative phases according to $E_n = E_o \exp(i\phi_n)$ where $\phi_n = \phi_o - \Delta k(z_n - z_1)$.

modes have equal phase velocities, and the process is phase matched.

The quantity $(k_2 - k_1)$ is called Δk, the phase mismatch. If $\Delta k \neq 0$ the phasors curl to form a section of a circle, reducing the net field in mode two. If all the phasors have the same length, as they would if the wave in mode one is undepleted, the size of the mode two field at position z is

$$E(z) = Kz\frac{\sin(\Delta kz/2)}{\Delta kz/2} = Kz\,\mathrm{sinc}\left(\frac{\Delta kz}{2}\right) \qquad [1]$$

where K is a constant describing the scattering strength. The corresponding power in mode two is plotted in Figure 3 for various values of Δk. When $\Delta k = 0$ the power grows quadratically with z, but for $\Delta k \neq 0$ it is proportional to $\sin^2(\Delta kz/2)/\Delta k^2$. If we plot the normalized output power in mode two ($z = L$) versus Δk, we have the curve shown in Figure 4 which is $\mathrm{sinc}^2(\Delta kL/2)$.

When $\Delta k = 2\pi/L$ the phasors form a complete circle so the power in mode two first grows with z but then shrinks back to zero at $z = L$. For a given value of Δk the length for which the phasors form a complete circle is known as the coherence length, $L_{\mathrm{coh}} = 2\pi/\Delta k$.

Reducing the z increments to infinitesimals, we can write the output field for mode two in integral form:

$$E_{\mathrm{net}} = \int_0^L S(z)e^{-i\Delta kz}\,\mathrm{d}z \qquad [2]$$

This shows that when Δk is independent of z, the output field is the Fourier transform of a source term,

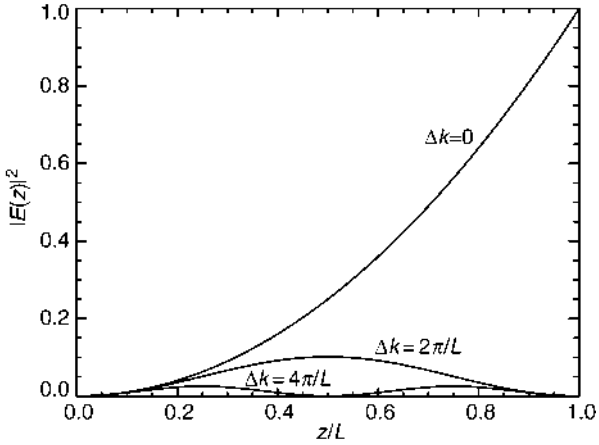

Figure 3 Plot of power in second mode versus z for different values of Δk. For $\Delta k = 0$ the power is proportional to z^2. For $\Delta k = 2\pi/\mathrm{L}$ the phasors form a closed circle so the net field is zero at $z = L$. For $\Delta k = 4\pi/\mathrm{L}$ the phasors trace two complete overlapped circles so again the field is zero at the $z = L/2$ and at $z = L$.

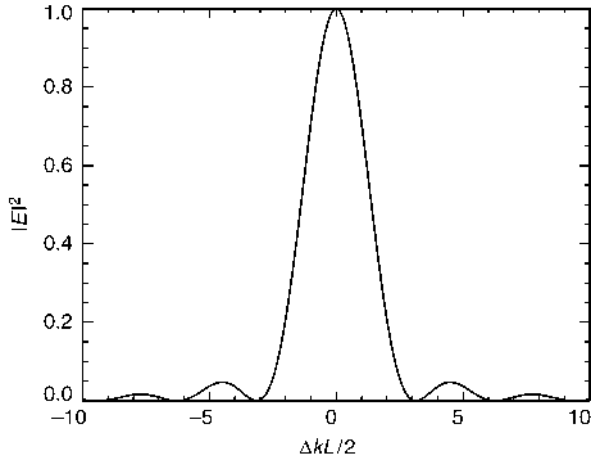

Figure 4 The plotted function indicates the relative irradiance of the output field, $|E|^2$, versus the phase mismatch Δk. The full width at half maximum is $\Delta k = 1.77\pi/L = 5.57/L$.

$S(z)$, transformed into Δk space. Where $S(z)$ has a constant value from $z = 0$ to $z = L$ and is zero outside that range, the Fourier transform has the form given by eqn [1], with $z = L$.

The $L^2\,\mathrm{sinc}^2(\Delta kL/2)$ form for the output power is ubiquitous in functions that describe the efficiency of nonlinear optical interactions, as well as acousto-optic and electro-optic modulation of light waves. However, it should be kept in mind that $L^2\,\mathrm{sinc}^2(\Delta kL/2)$ holds only when $S(z)$ is independent of z. If the scattering is strong enough to alter the phase or amplitude of the driving wave, or if the coupling strength varies with z for any other reason, $S(z)$ is not independent of z and the $L^2\,\mathrm{sinc}^2(\Delta kL/2)$ dependence is modified.

Our simple example relied on scattering in a waveguide, with a source term that was linear in the field in mode one. In practice, most optical devices that couple light between modes, rely on nonlinear interactions among multiple optical fields, or between an optical field and an acoustic or an rf electric wave. These nonlinear interactions produce a polarization wave in a suitable material at a sum or difference frequency. Each point of this polarization radiates a wavelet in all directions. These wavelets interfere constructively only in directions that are nearly phase matched. Examples include traveling wave acousto-optic filters, in which an acoustic wave traveling parallel to an input light beam of one polarization in a birefringent material, couples light into the orthogonal polarization via the elasto-optic effect. Traveling wave electro-optic modulators are similar but rely on the linear electro-optic effect for the coupling. Optical frequency mixing uses the nonlinearity of various

materials to mix multiple optical waves, creating polarization waves at sum and difference frequencies. A wide variety of such nonlinear interactions is discussed in textbooks on nonlinear optics, electro-optics, and acousto-optics. In all of these cases the phase velocity mismatch of interest is that between an induced polarization wave and a radiated mode, and the phase mismatch is $\Delta k = k_{\text{radiation}} - k_{\text{polarization}}$.

To find $S(z)$ and Δk for these nonlinear mixing processes, we must construct the polarization wave. We consider only monochromatic plane waves here, considering beams with spatial or temporal structure in a later section. The electric field of a monochromatic plane optical wave propagating in direction $\mathbf{r}_o$ can be described by

$$\begin{aligned} E &= E_o \cos(\omega_o t - \mathbf{k}_o \cdot \mathbf{r}_o + \phi_o) \\ &= \tfrac{1}{2}[E\, e^{-i(\omega_o t - \mathbf{k}_o \cdot \mathbf{r}_o)} + E^*\, e^{i(\omega_o t - \mathbf{k}_o \cdot \mathbf{r}_o)}] \end{aligned} \qquad [3]$$

The complex notation is a convenience that simplifies algebraic book-keeping. Similarly, an acoustic field propagating in direction $\mathbf{r}_a$ can be written:

$$\begin{aligned} A &= A_o \cos(\omega_a t - \mathbf{k}_a \cdot \mathbf{r}_a + \phi_a) \\ &= \tfrac{1}{2}[A\, e^{-i(\omega_a t - \mathbf{k}_a \cdot \mathbf{r}_a)} + A^*\, e^{i(\omega_a t - \mathbf{k}_a \cdot \mathbf{r}_a)}] \end{aligned} \qquad [4]$$

In a suitable material the polarization induced by the product of the optical and acoustic waves has a term of the form:

$$P = sEA \qquad [5]$$

P has components at both the sum and difference frequency. The sum frequency component is:

$$P_{\omega_o+\omega_a} = \frac{1}{4} s[EA\, e^{-i(\omega_o+\omega_a)t + i(\mathbf{k}_o+\mathbf{k}_a)\cdot\mathbf{r}} + \text{cc}] \qquad [6]$$

where 'cc' indicates the complex conjugate of the first term. The component at the difference frequency is

$$P_{\omega_o-\omega_a} = \frac{1}{4} s[EA^*\, e^{-i(\omega_o-\omega_a)t + i(\mathbf{k}_o-\mathbf{k}_a)\cdot\mathbf{r}} + \text{cc}] \qquad [7]$$

An important point is that the k vector of the sum-frequency polarization is $(\mathbf{k}_o + \mathbf{k}_a)$. Phase matching requires that the radiated sum-frequency wave has the same propagation vector. The radiated difference frequency wave must have the propagation vector $(\mathbf{k}_o - \mathbf{k}_a)$ to be phase matched. Phase matching for these two cases is shown in Figure 5. Note the general rule that if the frequency of a wave appears as $+\omega$ in the induced polarization, the associated k vector contribution to the polarization wave is $+\mathbf{k}$, while if it appears as $-\omega$, the k vector contribution is $-\mathbf{k}$. For example, the acoustic wave in $P_{\omega_o+\omega_a}$ is associated with $+\mathbf{k}_a$ contribution to the polarization wave, while in $P_{\omega_o-\omega_a}$ it is associated with a $-\mathbf{k}_a$ contribution.

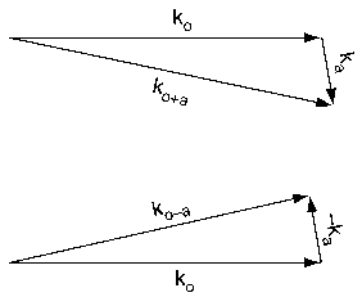

Figure 5 An optical wave with propagation vector k_o and an acoustic wave with propagation vector k_a produce a polarization wave at the sum frequency with the propagation vector k_{o+a} (top diagram), and a polarization wave at the difference frequency with the propagation vector k_{o-a} (bottom diagram).

Using the definition of k:

$$\mathbf{k} = \frac{2\pi}{\lambda}\hat{\mathbf{k}} = \frac{\omega}{v}\hat{\mathbf{k}} \qquad [8]$$

where v is the phase velocity, we can write Δk as

$$\begin{aligned} \Delta\mathbf{k} &= \mathbf{k}_{o+a} - \mathbf{k}_o - \mathbf{k}_a \\ &= \left(\frac{\omega_{o+a}}{v_{o+a}}\hat{\mathbf{k}}_{o+a} - \frac{\omega_o}{v_o}\hat{\mathbf{k}}_o - \frac{\omega_a}{v_a}\hat{\mathbf{k}}_a\right) \end{aligned} \qquad [9]$$

Clearly, if all of the waves had equal phase velocities, phase matching would be achieved for collinear interactions, where all of the propagation vectors are parallel, because the frequencies inside the parentheses sum to zero. This is seldom the case as for example, the phase velocity of an acoustic wave is much less than that of an optical wave, making an acoustic wave's k-vector disproportionately long compared with that of an optical wave. On the other hand, an rf wave usually has a disproportionately short k-vector compared with an optical wave. Even when all of the waves are optical, dispersion of the refractive index usually makes the k-vector of the bluer waves disproportionately long compared with those of the red waves. A wide variety of methods are used separately or in combinations to compensate for these effects in order to achieve phase matching and these are discussed below.

Phase Matching Techniques

Noncollinear Phase Matching

If one of the low frequency waves in a three wave mixing process, such as the acousto-optic interaction discussed above, has a lower phase velocity than the bluest wave, then noncollinear phase matching may be possible. Acousto-optic deflectors are often phase matched with a noncollinear geometry, such as that shown in Figure 6. An input optical wave of frequency ω_o reflects from a traveling acoustic wave with frequency ω_a in an optically isotropic material to produce an output optical wave of frequency $\omega_o + \omega_a$. The Bragg condition for efficient reflection is

$$\sin\theta = \frac{k_a}{2k_o} \qquad [10]$$

which may be seen from the figure to be the same as the phase matching condition, assuming k_{out} and k_{in} have the same magnitude, which is a good approximation for typical acoustic frequencies less than a few GHz.

Stimulated Brillouin scattering is another example of noncollinear mixing of acoustic and optical waves. One possible geometry has the input optical and acoustic waves traveling in opposite directions to create a reflected optical wave at the sum frequency. The phase matching condition is then

$$k_{acoustic} \approx 2k_{optical} \qquad [11]$$

which implies

$$\omega_{acoustic} = \frac{2n\omega_{optical}}{c} v_{acoustic} \qquad [12]$$

The acoustic frequency that phase matches such counter propagating Brillouin scattering is usually a few GHz in liquids and solids.

If noncollinear phase matching in an isotropic material is not possible, as is usually the case for electro-optic modulation and for all-optical sum and difference frequency mixing, it is sometimes possible to use the properties of nonisotropic media to adjust the phase velocity of one or more of the waves. There are several possibilities, such as natural birefringence, form birefringence, photonic band edge tuning, wave guide modal propagation, and various forms of quasi phase matching. Each is briefly described below.

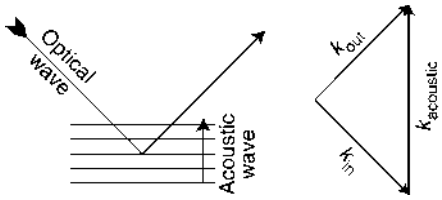

Figure 6 Phase matching in an acousto-optic deflector.

Birefringent Phase Matching

Natural birefringence

Naturally birefringent crystals can be categorized as either uniaxial or biaxial. In either case, light propagating along any direction (except along an optical axis) has two orthogonal eigen polarizations with different refractive indices. In uniaxial crystals, one of these indices is independent of the propagation direction and is called the ordinary index, n_o. The refractive index for the other polarization, known as the extraordinary polarization, is given by

$$\frac{1}{n_e^2(\theta)} = \frac{\sin^2\theta}{n_e^2} + \frac{\cos^2\theta}{n_o^2} \qquad [13]$$

where θ specifies the propagation direction measured from the z, or optical, axis, and n_e and n_o are the refractive indices for light polarized parallel and perpendicular to z. The two eigen polarizations are aligned with the lines of constant or latitude (o-wave) or constant longitude (e-wave). The variation of the two refractive indices with propagation angle θ is shown in Figure 7 for a negative uniaxial crystal which is characterized by $n_e < n_o$, and for a positive uniaxial crystal, characterized by $n_e > n_o$.

The acousto-optic tunable filter relies on natural birefringence to achieve a phase match. An optical and an acoustic wave propagate normal to the optical axis of a uniaxial crystal. If the crystal is negative uniaxial, with an input optical polarization of 'e', and an output optical polarization of 'o', the k vector of the acoustic wave must make up the difference

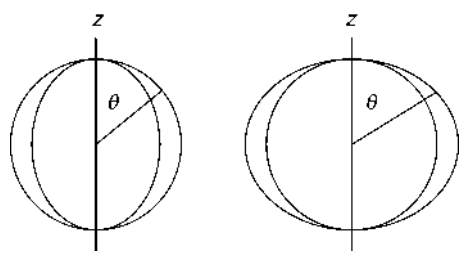

Figure 7 The circle and ellipse represent the two refractive indices for a negative uniaxial crystal (left diagram) and a positive uniaxial crystal (right diagram) for different propagation directions specified by the angles (θ, ϕ) with ϕ fixed. These curves are independent of the azimuthal angle ϕ, so in three dimensions the ordinary refractive index, n_o, forms a sphere in both cases, while the extraordinary index, n_e, forms a prolate ellipsoid for a negative uniaxial crystal, or an oblate ellipsoid for a positive uniaxial crystal.

$k_o - k_e$. This may be expressed as

$$k_a = \frac{\omega_a}{v_a} = \frac{(\omega + \omega_a)n_o}{c} - \frac{\omega n_e}{c} \qquad [14]$$

or, ignoring the difference between ω and $\omega + \omega_a$, the optical frequency that is phase matched is

$$\omega = \frac{\omega_a c}{v_a \Delta n} \qquad [15]$$

where v_a is the phase velocity of the acoustic wave, and $\Delta n = n_o - n_e$. The phase matched optical frequency is seen to be tunable with ω_a. A typical acoustic frequency for a visible filter is a few hundred MHz. The phase matching bandwidth set by the $\text{sinc}^2(\Delta kL/2)$ function can be quite narrow, with a subnanometer optical bandpass in a crystal a few cm long.

Natural birefringence is also widely used to phase match all-optical sum and difference frequency mixing. To counter normal optical dispersion, the bluest wave is polarized in the direction associated with the smaller refractive index, while at least one of the two redder waves has the opposite polarization. By adjusting the propagation angle, θ, the refractive index of the extraordinary waves can be adjusted until exact phase matching is achieved. If the two redder waves both have the high index polarization the process is sometimes called type I, whereas if they have opposite polarizations it is called type II.

Figure 8 illustrates type I phase matched second harmonic generation in a negative uniaxial crystal. The fundamental wave with frequency ω is o-polarized, while the second harmonic wave at frequency 2ω is e-polarized. The circle drawn in dashes represents the line swept out by the polarization wave vector $2k_\omega$ as the propagation direction θ is swept, while the ellipse represents the loci of $k_{2\omega}$ for the same angle. Collinear phase matching is achieved at the point of intersection of the two curves. This sets the angle θ_{pm}. For angles greater than θ_{pm}, dispersion is over-compensated and $\Delta k < 0$ so noncollinear phase matching is possible. All ϕ angles have equivalent propagation properties, so the loci of phase matched propagation directions form a cone at $\theta = \theta_{pm}$. Usually the propagation angle ϕ is chosen to maximize the nonlinear response of the crystal.

Biaxial crystals share with uniaxial crystals the property of two orthogonal eigen polarizations with different refractive indices, but the angular dependence of the indices is more complex. The two indices associated with propagation along direction (θ, ϕ) can be found by solving

$$\frac{\sin^2\theta\,\cos^2\phi}{1/n^2 - 1/n_x^2} + \frac{\sin^2\theta\,\sin^2\phi}{1/n^2 - 1/n_y^2} + \frac{\cos^2\phi}{1/n^2 - 1/n_z^2} = 0 \qquad [16]$$

for the two possible values of n. Here n_x, n_y, and n_z are the refractive indices for light polarized along the respective principal axes, where the convention is $n_x < n_y < n_z$. There are two optical axes lying in the x–z plane at an angle determined by the values of n_x, n_y, and n_z. The phase matching loci are no longer lines of constant latitude as they were for uniaxial crystals, but instead are curves that encircle either the optical axes or one of the principal axes. Figure 9 shows an

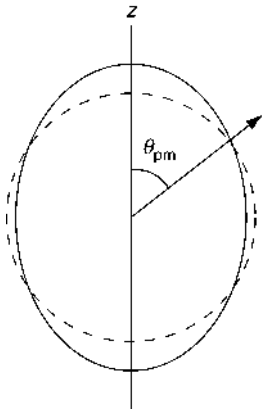

Figure 8 Collinear phase matching of second harmonic generation in a negative uniaxial crystal. The second harmonic light is e-polarized with a refractive index indicated by the solid line while the fundamental light is o-polarized as indicated by the dashed line. The crossing point of the two refractive indices determines the phase matching angle θ_{pm}.

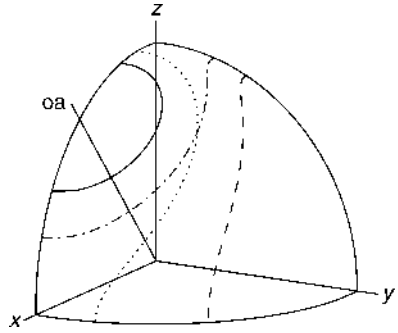

Figure 9 Calculated loci of propagation directions that achieve birefringent phase matching for type I collinear second harmonic generation of different wavelengths in the crystal CsB_3O_5. The fundamental wavelengths are 2,700 nm (dotted curve looping around the x axis), 3,000 nm (dashed curve looping around the y axis), 1,000 nm (solid curve looping around the optical axis 'oa'), and 800 nm (chained curve looping around the z axis). For clarity only a single octant of the sphere of propagation directions is shown. The loci in the other octants of the sphere are formed by reflections in the three principal planes.

example of each type of phase matching curve associated with second harmonic generation of different wavelengths in the crystal CsB_3O_5.

Note that because these curves always intersect at least one of the principal planes, if phase matching in biaxial crystals is possible at all, it is also possible in a principal plane. Biaxial crystals are usually phase matched in a principal plane, but phase matching outside the principal planes can in some cases maximize the nonlinear response, minimize the temperature sensitivity, or adjust the relative group velocities.

Form birefringence

Some materials, such as GaAs, have a strong nonlinear optical response but are not birefringent, so optical frequency mixing cannot be phase matched using natural birefringence. However, form birefringence can be engineered by layered growth, and it can be used to phase match. If many alternating layers of two materials with different refractive indices are stacked together with a layer thickness comparable to the wavelength of light, the material is birefringent for light propagating parallel to the layers. Light polarized with its E field parallel to the layers will have a higher refractive index than light polarized with its field perpendicular to the layers. The strength of birefringence, $(n_{\parallel} - n_{\perp})$, can be engineered by varying the relative thicknesses and refractive indices of the two layered materials. This has been demonstrated in GaAs/AlAs waveguide structures in which the AlAs layers were oxidized to form Al_2O_3, which has a refractive index of 1.6, compared with an index of 3.4 for GaAs. The resulting large birefringence was used to phase match infrared difference frequency mixing.

Photonic Lattice Phase Matching

The propagation constant of a light wave can be controlled by tuning its frequency near the bandgap of a Bragg structure. The Bragg structure can be a stack of layers with alternating refractive index. If the stack consists of layers of thickness and refractive index l_1, n_1 and l_2, n_2, the Bragg condition at normal incidence is

$$\phi_1 + \phi_2 = \frac{\omega_{\text{Bragg}}}{c}(n_1 l_1 + n_2 l_2) = m\pi \qquad [17]$$

where m is an integer. Frequencies near ω_{Bragg} are strongly reflected. The width of the reflective band of a structure with many layers is proportional to the refractive index contrast, $(n_1 - n_2)/(n_1 + n_2)$. At certain frequencies, on both sides of the bandgap, the reflectivity vanishes. The first transmission peak on the red side of the bandgap has an effective refractive index that is greater than the average $(n_1 + n_2)/2$, while the first peak on the blue side has an effective index less than the average. This effect can be used to adjust the effective value of Δk in the structure. As a bonus, operating on one of the transmission peaks near the bandgap enhances the optical field in the material, increasing the strength of the nonlinear interaction.

It is also possible to use a Bragg structure in reflection mode. Tuning to the red side of the band center increases the effective refractive index of the reflected wave, while tuning to the blue side decreases it. The internal fields are not enhanced in this case and the interaction length may be quite short.

Quasi-Phase Matching (QPM)

Quasi-phase matching uses spatial modulation of the polarization source with a modulation period of $L_{\text{mod}} = 2\pi/\Delta k$ to boost the efficiency of the mixing process even though $\Delta k \neq 0$. We can consider the periodic modulation as another interacting wave that is included in the creation of the polarization wave. The modulation wave has $\omega = 0$ and is stationary with period L_{mod}. Because the modulation is usually a square wave rather than sinusoidal, it is made up of multiple sinusoidal waves with periods L_{mod}/m, all of which can participate in generating a polarization wave. Like traveling waves, the stationary modulation waves can be oriented in any direction, and any Fourier component of the modulation wave with wave vector $\Lambda_m = 2\pi m/L_{\text{mod}}$ can participate in (quasi) phase matching, as shown in Figure 10.

Usually the spatial modulation is achieved by periodically reversing the sign of the nonlinear coupling coefficient. Figure 11 shows the associated phasor diagram for perfect first-order QPM. The net fields from the first two modulation periods are labeled E_1 and E_2. If the sign of the nonlinearity in the first half of the modulation period were not reversed in the second half, the phasors would form a complete circle resulting in a null net field. All following periods would also produce null net fields. However, if the sign is reversed midway through the first period, the phasors from the second half are sign reversed and the net field is E_1 as shown. The magnitude of E_1 is smaller than for a truly phase matched segment by the

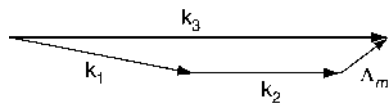

Figure 10 Quasi-phase matching vector diagram for sum frequency mixing. The mth Fourier component of the grating vector is indicated by Λm.

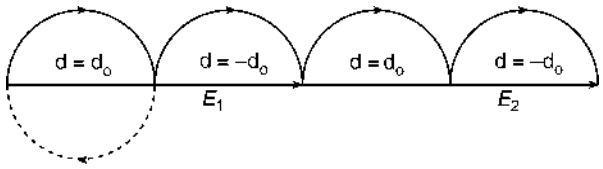

Figure 11 Phasor diagram for a quasi-phase matched interaction in a material with alternating signs of the nonlinear coefficient. The contributions from four segments, each of length $L_{coh}/2$ are shown. The net field from the first QPM period is labeled E_1, and that from the second period is E_2. The dashed line shows the path the phasors would trace if the nonlinear coefficient were not sign reversed in the even numbered segments.

ratio of the diameter of the circle to its circumference, or by $2/\pi$. Coherence lengths in crystals are typically 5–50 μm so crystals that are several mm long have a large number of periods, in which case the net phasors $E_1, E_2, \ldots$ can themselves be considered elementary phasors that form the net field for the full crystal. For example, if the modulation period L_{mod} is not exactly equal to $2\pi/\Delta k$, QPM is imperfect and phasors $E_1, E_2, \ldots$ curl up to give a net field of $\mathrm{sinc}(\Delta k' L/2)$ where $\Delta k' = \Delta k - 2\pi/L_{mod}$. For all practical purposes, a QPM crystal containing many modulation periods performs the same as a phase matched crystal with its nonlinear coupling reduced by $2/\pi$.

Referring to eqn [2], we note that periodic modulation of the source term $S(z)$ produces peaks in the phase matching curves at $\Delta k = m2\pi/L_{mod}$ where m is an integer. If $S(z)$ is a square wave of alternating sign over $0 < z < L$, its Fourier transform is the convolution of $\mathrm{sinc}(\Delta kL/2)$ with a picket fence function with spikes at $\Delta k = m2\pi/L_{mod}$. If $L_{mod} \ll L$, the crystal contains many periods of the square wave and the spikes are separated by many times the width of the $\mathrm{sinc}(\Delta kL/2)$ function. This results in well separated sinc functions centered at $\Delta k = 2\pi m/L_{mod}$, and any process with a coherence length of L_{mod}/m will be quasi-phase matched. That is, one of the Fourier peaks in Δk space will be centered on the corresponding value of Δk.

The magnitudes of the QPM peaks are maximized if the modulation of $S(z)$ is a sign reversal, as described above. If the positive and negative segments have unequal lengths, we can define the duty cycle as $D = l_+/(l_+ + l_-)$. The heights of the pickets in the picket fence are then

$$h_m = \frac{2}{m\pi}\sin(\pi m D) \qquad [18]$$

The largest magnitude peak is for $m = 1$ and $D = 0.5$ in which case $h_1 = 2/\pi$.

Unlike birefringent phase matching, QPM can, in principle, phase match any set of polarizations for the interacting waves. This makes it possible to use a polarization set that maximizes the nonlinear response of the crystal.

Ferroelectric poling

There are various ways to achieve QPM. By far the most popular and versatile is by ferroelectric poling. Ferroelectric crystals have an internal electric field that can be reversed by the application of a sufficiently strong external poling field. Reversing the internal field also reverses the sign of the nonlinearity for some light polarizations but does not alter the refractive indices. Crystals of $LiNbO_3$ and $KTiOPO_4$, with thickness up to 1 mm, can be poled with periods as short as 5 μm. This is sufficient to quasi-phase match frequency mixing of visible or near infrared light. Figure 12 shows an example of the orientation of the internal fields of patterned ferroelectric domains.

It is possible to pole a crystal aperiodically so that multiple mixing processes are quasi-phase matched simultaneously. To design the poling pattern for such a device, the Fourier transform of an initial trial pattern can be calculated and compared with the target spectrum in Δk space. Iterative algorithms can be used to modify the poling pattern, zeroing in on the target spectrum.

Ferroelectric poling can also form two-dimensional poling patterns, corresponding to a two-dimensional set of Fourier components in Δk space. It can also produce chirped gratings which have a broad phase matching peak. Chirped gratings can be used to simultaneously frequency convert and tailor the temporal shapes of the output pulse when the input pulses are frequency chirped. Poling can also yield fanned poling patterns so the poling period and thus the phase matching wavelength can be shifted by translating the grating.

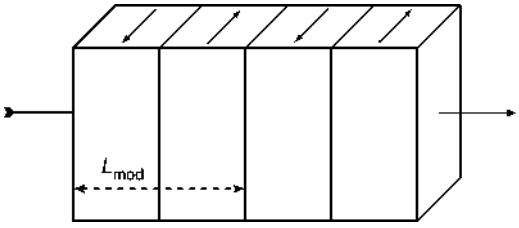

Figure 12 Diagram of a poled ferroelectric crystal. The four crystal segments represent four ferroelectric domains with internal crystal field directions indicated by the arrows. Reversal of the internal field is accompanied by reversal of the nonlinear coefficient, so this is a quasi-phase matched structure with alternating signs of d_{eff}.

Wafer stacking and patterned crystal growth

Another method of quasi-phase matching relies on stacks of thin plates oriented so the sign of the nonlinearity reverses in adjacent plates. Each plate has a thickness of $mL_{\mathrm{coh}}/2$ where m is an odd integer. If all the plates are identically prepared crystals, it is often possible to reverse the sign of their nonlinearity by reorienting them. It is sometimes possible to optically contact the plates or even to diffusion bond them, avoiding reflective losses at the interfaces. This technique is conducive to large aperture QPM crystals. However, the plates are thin so fabrication is a challenge.

Quasi-phase matched crystals of GaAs have also been epitaxially grown on patterned substrates. The patterning serves as a template for growth of zones with alternating crystal orientation and sign of the nonlinear coefficient.

Total internal reflection

Total internal reflection in a thin slab has also been exploited for QPM (Figure 13). Two effects contribute to the source modulation, one is the rephasing of the waves on total internal reflection; the other is the change in nonlinearity that may accompany the change in propagation direction. Usually plates much thicker than one coherence length are used, so the quasi-phase matching is of high order ($m \gg 1$).

Wave Guide Modal Phase Matching

In waveguides the size and shape of the guide influence the transverse field distributions and the propagation constants for the transverse modes. Generally the higher order transverse modes have higher phase velocities than the lower order modes. This can be used to counteract index dispersion in three wave mixing if the bluest wave propagates in a higher order mode than the redder waves. A problem frequently encountered in using this effect is that the transverse modes are orthogonal, or nearly so. Because the nonlinear response is proportional to the overlap integral

$$\mathrm{d}_{\mathrm{eff}} \propto \iint_A \mathrm{d}(x,y)E_1(x,y)E_2(x,y)E_3(x,y)\mathrm{d}x\,\mathrm{d}y \qquad [19]$$

where the E's are the transverse field distributions of the three waves, the nonlinear response is often quite small. If the local nonlinearity $\mathrm{d}(x,y)$ can be engineered to maximize the integral, this problem can be alleviated. This may be thought of as transverse quasi-phase matching.

Waveguides can also be designed for Cerenkov phase matching where the phase velocity of the radiated wave is less than that of the polarization wave. Figure 14 shows an example. The emission angle is

$$\cos\theta = \frac{2k_\omega}{k_{2\omega}} = \frac{v_{2\omega}}{v_\omega} \qquad [20]$$

where the v's are phase velocities. The transverse momentum Λ must be supplied by the waveguide. In other words, the source harmonic polarization must have a transverse Fourier component with spatial frequency Λ. This is another example of transverse quasi-phase matching.

Another application of waveguide phase matching is in traveling wave electro-optic modulation, where the rf wave travels along an rf waveguide parallel to the light path. For maximum modulation bandwidth, the phase velocity of the rf wave should equal that of the light wave over the desired modulation bandwidth. Because the phase velocity of the rf wave in the bulk medium is usually faster than that of the light wave, the rf guide must be designed for a slow phase velocity.

Compensated Phase Matching (CPM)

Compensated phase matching is somewhat like QPM in that multiple crystal segments are used, but it differs by requiring that sequential crystal segments have ΔkL's of equal magnitude but opposite sign. Ideally the nonlinearity would be the same in all segments. Figure 15 shows the phasor diagram for one such arrangement, in this case with unequal nonlinearity in alternating segments. CPM is realized

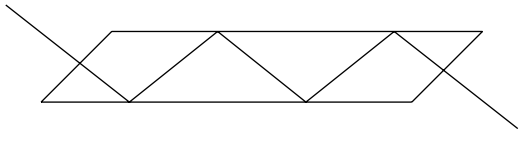

Figure 13 Diagram of quasi-phase matching based on total internal reflection. Relative phases acquired by the waves upon total internal reflection, combined with the path lengths and possible changes in the value of d_{eff} for alternating legs, can be combined to achieve quasi-phase matching.

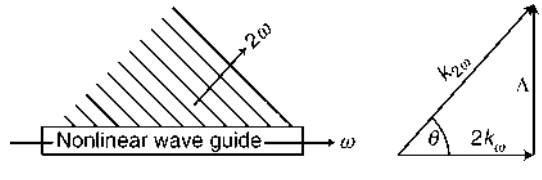

Figure 14 Cerenkov phase matched second harmonic generation. The phase velocity of the fundamental wave in the waveguide is greater than the phase velocity of the harmonic wave in the cladding.

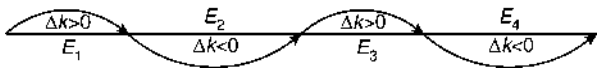

Figure 15 Phasor diagram for four segments of a device with compensated phase matching device. The magnitude of ΔkL is equal in each segment but the sign of Δk alternates. The length of the E_n field from segment n is proportional to $\mathrm{sinc}(\Delta kL/2)$ so the net field ΣE_n must also be proportional to this function.

in walkoff compensated geometries where a sequence of crystals is arranged so $\Delta k = 0$ in each for a particular monochromatic plane wave. In the case of spatial walkoff compensation the walkoff direction in alternating crystals is reversed so the walkoff is exactly compensated in pairs of adjacent crystals. Tilting the input beams by a small amount changes the values of the Δk's in such a way that CPM is achieved. In the case of temporal walkoff compensation, temporal walkoff is reversed in alternating crystals, in which case tuning the input frequency leads to CPM. The advantage of CPM is that the acceptance bands, which will be discussed in the next section, are those of the individual crystal segments which are larger than those of a single crystal of equivalent length without walkoff compensation.

Spatial and Temporal Effects

Acceptance Bands

Our discussion to this point has been based on monochromatic plane waves. Real beams have spatial and temporal structure which means they can be decomposed into sets of monochromatic plane waves with tilted propagation directions relative to the nominal direction, and with detuned frequencies relative to the nominal frequency. As we showed above for monochromatic plane waves, there is a range of angles and frequencies, called the acceptance bands, over which $\Delta kL < 2\pi$. If the interacting beams' angular spectra and frequency spectra lie within these bands, the process is well phase matched for all angular and frequency components, and efficiency is maximized. This statement is equivalent to saying that if the spatial walkoff due to beam tilts and birefringence, and the temporal walkoff due to differing group velocities are small enough that the interacting beams stay well overlapped in space and time, the efficiency of energy transfer to the output channel does not suffer. Consider, for example, second harmonic generation of short pulses. The frequency acceptance bandwidth for the fundamental wave, $\Delta\omega_1$, is found using

$$\frac{\mathrm{d}(\Delta k)}{\mathrm{d}\omega_1}\Delta\omega_1 = \left(\frac{2\mathrm{d}k_2}{\mathrm{d}\omega_2} - \frac{2\mathrm{d}k_1}{\mathrm{d}\omega_1}\right)\Delta\omega_1 = \frac{2\pi}{L} \qquad [21]$$

Using the usual definition of the group velocity:

$$v_{\mathrm{g}} = \frac{\mathrm{d}\omega}{\mathrm{d}k} \qquad [22]$$

this can be written

$$\Delta\omega_1 = \frac{\pi}{L(1/v_{\mathrm{g2}} - 1/v_{\mathrm{g1}})} = \frac{\pi}{\tau_2 - \tau_1} = \frac{\pi}{\Delta\tau} \qquad [23]$$

where τ_1 and τ_2 are the passage times through a crystal of length L for the fundamental and harmonic pulses, and $\Delta\tau = (\tau_2 - \tau_1)$ is the temporal walkoff between the two. This shows that the acceptance bandwidth is approximately equal to the transform bandwidth of a pulse whose duration is equal to the temporal walkoff $\Delta\tau$. Input pulses shorter than $\Delta\tau$ will not convert efficiently because the fundamental and harmonic pulses do not remain overlapped as they propagate through the crystal. In the frequency domain it is said that pulses shorter than the $\Delta\tau$ have spectral widths larger than the acceptance bandwidth, meaning they are not well phase matched across their full spectra, and hence do not convert efficiently. Parallel arguments apply to spatial walkoff and angular acceptance bandwidths. Beams that do not remain spatially overlapped through the crystal do not convert efficiently. Equivalently, if the angular spectra of the beams exceed the acceptance angle, efficiency suffers.

Sometimes these notions are described in terms of an effective interaction length or an effective crystal length. For short pulses the effective interaction length is that for which temporal walkoff is equal to the pulse duration. Equivalently it is the length for which the acceptance bandwidth is equal to the spectral width of the pulse. Crystals shorter than this will have mixing efficiencies nearly equal to that expected for a phase matched process, while longer crystals have efficiencies that fall short of this expectation.

Focused beams

Focused Gaussian beams are a particularly important class of spatially structured beams. In passing through a focus, a lowest order Gaussian beam acquires a phase shift relative to an unfocused beam. The phase velocity near the focus is higher for the focused beam, leading to a phase shift of π through the focus. For second harmonic generation the phase shift of the second harmonic polarization is twice this, or 2π.

A second harmonic wave also acquires a phase shift of π, but this leaves a π phase shift between the harmonic polarization source wave and a freely propagating harmonic wave. The phasors representing harmonic radiation from symmetrically located points well before and well after the focus will therefore be 180° out of phase. Point pairs closer to the focus will have less severe cancelation, but it is not surprising that the maximum efficiency is achieved, not at $\Delta k = 0$, but for a slightly negative value Δk. The phase matching function $\mathrm{sinc}(\Delta kL/2)$ is shifted and distorted for focused beams. Its exact shape depends on the nature of the mixing process. For example, the phase matching peak is different for third harmonic generation and second harmonic generation, and it is different for mixing with small birefringence and mixing with large birefringence.

Phase Conjugation

Degenerate four-wave mixing

Some wave mixing processes are automatically phase matched. One example is phase conjugation by degenerate four wave mixing. This process creates a polarization wave that is the time reversed image of one of the input waves. Figure 16 compares the reflection from a conventional mirror with that from a phase conjugate mirror.

The mixing process shown in Figure 17 is one that can act as a phase conjugate mirror. Waves f and b are strong pump waves that are phase conjugates of one another. In their presence a weak signal field at the same frequency produces a polarization given by

$$P = sE_f E_b E_s^* \, e^{-i\omega t + i(\mathbf{k}_f - \mathbf{k}_b - \mathbf{k}_s)\cdot\mathbf{r}} \qquad [24]$$

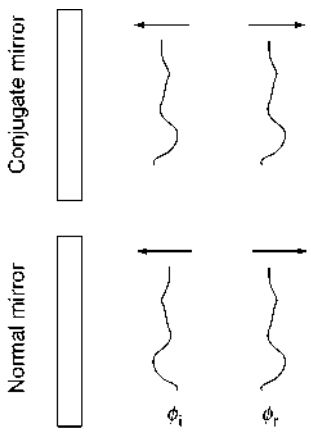

Figure 16 Comparison of a normal and a phase conjugate mirror. The wavy lines represent the phases of the incident and reflected waves.

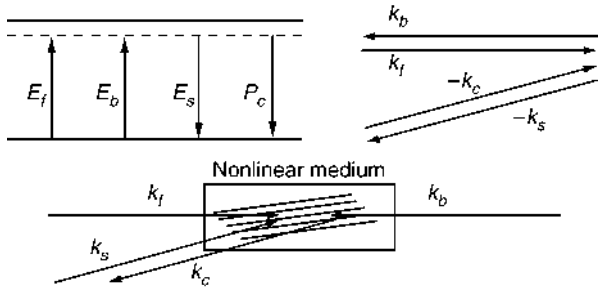

Figure 17 Phase conjugation via degenerate four wave mixing near a one photon resonance. Two counter propagating pump waves with fields E_f and E_b interact with a signal beam, E_s, to produce a polarization wave P_c where the frequencies obey $\omega_f - \omega_s + \omega_b = \omega_c$ (upper left), implying the k-vector of the polarization wave is given by $k_c = k_f - k_s + k_b$ (upper right). The lower diagram shows the grating induced by waves f and s that scatters wave b into the direction of c.

The two pump propagation vectors cancel one another leaving

$$P = sE_f E_b (E_s^* \, e^{-i\omega t - i\mathbf{k}_s\cdot\mathbf{r}}) \qquad [25]$$

The quantity in parentheses is the phase conjugate of the signal wave, so the polarization wave exactly phase matches a radiated conjugate wave.

Brillouin scattering

Other possibilities for automatic phase matching are processes that involve a stimulated wave in the nonlinear medium such as an acoustic wave (Brillouin scattering) or waves of coherent molecular vibration or rotation (Raman scattering). Phase conjugation, based on Brillouin scattering, is widely used to reduce phase distortions acquired on passing through an amplifier, for example. A single beam of light is focused into a material that has a high polarizability. Example media include high pressure gases Xe and SF_6, liquids acetone and freon, and fused silica. The conjugation process is not as simple to describe as it was in the case of four wave mixing because many acoustic waves can be generated. The key is that the phase conjugate wave has the highest gain under certain conditions and can dominate the Brillouin scattering process, so the reflected wave is nearly a phase conjugate of the incoming wave.

See also

Modulators: Acousto-Optics; Electro-Optics. **Nonlinear Optics, Basics:** Cascading; Four-Wave Mixing; Nonlinear Optical Phase Conjugation; $\chi^{(3)}$–Third-Harmonic Generation. **Nonlinear Sources**: Harmonic Generation in Gases. **Optical Parametric Devices:** Overview. **Optical Parametric Devices:** Optical Parametric Oscillators (Pulsed).

Phase Control: Phase Conjugation and Image Correction. **Photonic Crystals:** Nonlinear Optics in Photonic Crystal Fibers. **Scattering**: Raman Scattering.

Further Reading

Born M and Wolf E (1970) *Principles of Optics*. New York: Pergamon Press.
Boyd RW (1992) *Nonlinear Optics*. New York: Academic Press.
Dmitriev VG, Gurzadyan GG and Nikogosyan DN (1999) *Handbook of Nonlinear Optical Crystals*. Berlin: Springer-Verlag.
Hanna DC, Yuratich MA and Cotter D (1979) *Nonlinear Optics of Free Atoms and Molecules*. Berlin: Springer-Verlag.
Mills DL (1991) *Nonlinear Optics*. Berlin: Springer-Verlag.
Shen YR (1984) *The Principles of Nonlinear Optics*. New York: John Wiley & Sons.
Sutherland RL (1996) *Handbook of Nonlinear Optics*. New York: Marcel Dekker.
Yariv A and Yeh P (1984) *Optical Waves in Crystals*. New York: John Wiley & Sons.
Zel'dovich BY, Pilipetsky NF and Shkunov VV (1985) *Principles of Phase Conjugation*. Berlin: Springer-Verlag.
Zernike F and Midwinter JE (1973) *Applied Nonlinear Optics*. New York: John Wiley & Sons.

Pulse Compression via Nonlinear Optics

M F S Ferreira, University of Aveiro, Aveiro, Portugal

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Ultrashort pulses have a great potential in several areas of research and application, such as measurements of ultrafast physical processes, infrared time-resolved spectroscopy, and sampling systems, but they are also invaluable for communications. With increasing demands being put on the performance of optical fiber communications, the emphasis now is on producing a source of ultrashort optical pulses that can be used in high-bit rate and long-distance optical communication systems.

Several optical pulse compression techniques are described in the literature. These techniques utilize dispersion, possibly from gratings or prisms, simultaneously to compensate for chirp and achieve pulse compression. In fact, when a chirped optical pulse propagates in a linear dispersive medium, it acquires a dispersion-induced chirp. If the initial chirp is in the opposite direction to that imposed by the group-velocity dispersion (GVD), the two tend to cancel each other, resulting in the compression of the optical pulse.

Early work on optical pulse compression did not make use of any nonlinear optical effects. Only during the 1980s, when the use of single-mode silica fibers as nonlinear media became widespread, the self-phase modulation (SPM) effect was used to achieve pulse compression. It has led, in 1987, to the production of optical pulses as short as 6 fs in the 620 nm wavelength region. Such advances were possible, only after having understood the evolution of optical pulses in silica fibers.

We can consider two main categories of pulse compressors which make use of nonlinear effects in optical fibers: grating-fiber and soliton-effect compressors. In a grating-fiber compressor, the input pulse is propagated in the normal-dispersion regime of the fiber, which imposes a nearly linear, positive chirp on the pulse through a combination of SPM and GVD. The output pulse is then sent through a grating pair where it experiences anomalous GVD and gets compressed. In the case of a soliton-effect compressor, the fiber itself acts as a compressor without the need of an external grating pair. The input pulse propagates in the anomalous-GVD regime of the fiber and is compressed through an interplay between SPM and GVD. This compression mechanism is related to a fundamental property of the higher-order solitons. These solitons follow a periodic evolution pattern such that they go through an initial narrowing phase at the beginning of each period. If the fiber length is suitably chosen, the input pulses can be compressed by a factor that depends on the soliton order.

Grating-Fiber Compressors

The grating-fiber compressor is generally used at wavelengths $\lambda < 1.3\ \mu\text{m}$. At these wavelengths, optical pulses acquire a positive chirp across its entire width during their passage through conventional silica fibers. A schematic of a grating-fiber compressor in the double pass configuration is shown in Figure 1. After the propagation in the fiber, the optical pulse is sent through a grating pair, which provides the anomalous (or negative) GVD required to get the pulse compression.

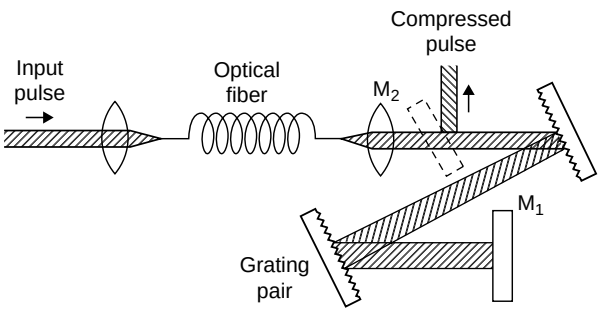

Figure 1 Schematic representation of a grating-fiber compressor in the double-pass configuration. (Reproduced with permission from Agarwal GP (2001) *Applications of Nonlinear Fiber Optics*. San Diego: Academic Press.)

Different frequency components of an optical pulse incident at one grating are diffracted at slightly different angles. As a consequence, they experience different time delays during their passage through the grating pair, the blue-shifted components arriving earlier than the red-shifted ones. In the case of an optical pulse with a positive chirp, the blue-shifted components occur near the trailing edge of the pulse and the passage through the grating pair provides its compression.

A limitation of the grating pair is that the spectral components of the pulse are dispersed not only temporally but also spatially. As a consequence, the optical beam becomes deformed, which is undesirable. This problem can be avoided simply by using a mirror to reflect the pulse back through the grating pair. Mirror M_1, in Figure 1, performs this function. Reversing the direction of propagation not only allows the beam to recover its original cross-section, but also doubles the amount of GVD, thereby reducing the grating separation by a factor of 2. The mirror M_2 in Figure 1 is used to deflect the compressed pulse out of the compressor.

The phase shift acquired by the spectral component of the pulse at the frequency ω passing through the grating pair can be written in the form:

$$\phi_g = \phi_0 + \phi_1(\omega - \omega_0) + \frac{1}{2}\phi_2(\omega - \omega_0)^2 + \frac{1}{6}\phi_3(\omega - \omega_0)^3 + \cdots \quad [1]$$

where ω_0 is the pulse center frequency, ϕ_0 is a constant, ϕ_1 is a constant delay, and ϕ_2 and ϕ_3 are parameters that take into account the GVD effects associated with the grating pairs. These parameters depend on the grating period (line spacing), as well as on the orientation and separation of the two gratings. In most cases of practical interest, the spectral width of the pulse satisfies the condition $\Delta\omega \ll \omega_0$ and the cubic and higher-order terms in eqn [1] can be neglected.

To achieve optimum performance from a grating-fiber compressor, it is necessary to optimize the fiber length, as well as the grating separation. Concerning the first aspect, the effects of both GVD and SPM during the propagation of the pulse inside the fiber must be considered. SPM alone determines a linear chirp only over the central part of an optical pulse. Since the grating pair compresses only this region, while a significant amount of energy remains in the wings, the compressed pulse is not of high quality in this case. The effect of GVD turns out to be important, since it broadens and reshapes the pulse, which develops a nearly linear chirp across its entire width. In these circumstances, the grating pair can compress most of the pulse energy into a narrow pulse.

The balanced effects of both GVD and SPM explain the existence of an optimum fiber length, z_{opt}, for maximum pulse compression. In fact, when the fiber length is less than z_{opt}, the SPM-induced chirp has not yet been linearized. On the other hand, when it is longer than z_{opt}, the SPM effects become negligible due to the GVD-induced pulse broadening.

If the input pulse is unchirped and presents a hyperbolic secant shape, the optimum fiber length for maximum pulse compression is well approximated by

$$z_{opt} \approx (6L_D L_{NL})^{1/2} \quad [2]$$

where

$$L_D = \frac{T_0^2}{|\beta_2|} \quad [3]$$

and

$$L_{NL} = \frac{1}{\gamma P_0} \quad [4]$$

are length scales associated with the GVD and the SMP effects, respectively. In eqns [3] and [4], T_0 is the input pulsewidth, β_2 is the fiber GVD parameter, γ is the nonlinearity parameter of the fiber, and P_0 is the peak power of the input pulse.

If the input pulse presents a negative chirp, this must be compensated for by the positive chirp provided by the fiber. As a result, the optimum fiber length increases. The opposite occurs in the case of an input pulse with a positive chirp.

The optimum compression factor, F_{opt}, in the case of a grating-pair compressor, is given by

$$F_{opt} = \frac{T_{FWHM}}{T_{comp}} \approx 0.625\left(\frac{L_D}{L_{NL}}\right)^{1/2} \quad [5]$$

where T_{FWHM} is the full width at half maximum (FWHM) of the input pulse ($T_{\mathrm{FWHM}} \approx 1.76T_0$ for a hyperbolic secant pulse) and T_{comp} is the FWHM of the compressed pulse. Equations [2] and [5] provide a good estimate even for pulse shapes other than a hyperbolic secant as long as $\frac{L_{\mathrm{D}}}{L_{\mathrm{NL}}} > 100$.

The previous results were obtained neglecting the higher-order nonlinear and dispersive effects, which is acceptable for pulsewidths $T_0 > 0.1$ ps. However, when shorter pulses are used at the input, their spectral width $\Delta\omega$ is large enough that the cubic term in the expansion [1] must be taken into account. In this case, the compression factor turns out to be smaller than that given by eqn [5].

The third-order dispersion (TOD) of the grating pair, resulting from cubic term in eqn [1], has been compensated for in one experiment using a combination of gratings and prisms. Performing such compensation, a reduction of the pulsewidth to about 6 fs has been achieved. Such a pulse at 620 nm consists of only three optical cycles.

Even though the compression factor $F_{\mathrm{c}} \propto \sqrt{P_0}$, the peak power of the input pulse must be kept below the Raman threshold to avoid the loss of pulse energy through the stimulated Raman scattering (SRS) process. If this condition is not then satisfied, besides the energy loss problem, the Raman pulse can interact with the input pulse through cross-phase modulation and deform the linear nature of the frequency chirp. As a result, the compression factor is reduced from the predictions of eqn [5].

Fiber Bragg Gratings

A uniform fiber Bragg grating (FBG) reflects light whose wavelength is within the stop-band centered at the Bragg wavelength. The GVD is anomalous on the high-frequency side of the stop-band and becomes normal on the low-frequency side. Near the stop-band edges, the grating exhibits large GVD. Since a FBG acts as a dispersive delay line, it can be used in place of the bulk-grating pair in a grating-fiber compressor, providing a compact all-fiber device. In fact, a typical 1 cm long fiber grating may provide as much dispersion as a bulk-grating pair with more than one meter spacing. However, the grating induced TOD can affect significantly the quality of the pulses when the optical frequency falls close to the edges of the stop-band.

The compression factor as well as the pulse quality can be significantly improved by using a FBG in which the grating optical period changes along its length – the so-called chirped fiber grating. In this case, different frequency components of the pulse are reflected from different regions of the grating. Such a device has been used since the mid-1990s to compensate for dispersion-induced broadening of pulses in fiber-optic communication systems, and compression factors above 100 have been achieved. In practice, the only disadvantage of a chirped fiber grating is that the compressed pulse is reflected rather than transmitted.

Soliton-Effect Compressors

Optical pulses at wavelengths exceeding 1.3 μm generally experience both SPM and anomalous GVD during their propagation in silica fibers. Such a fiber can act as a compressor by itself without the need of an external grating pair. In the case of a fundamental or first-order soliton, the effects of SPM cancel the effects of anomalous GVD perfectly, and the soliton propagates whilst preserving its hyperbolic secant shape in a lossless optical fiber. A higher-order soliton, on the other hand, changes its shape periodically as it propagates in a fiber. Such a soliton always experiences an initial pulse narrowing phase before recovering its original hyperbolic secant profile at integral multiples of the soliton period z_0, and this behavior is exploited to achieve pulse compression. Optimum compression is achieved when the higher-order soliton leaves the fiber just as it attains its narrowest width.

The evolution of a soliton of order N inside an optical fiber is governed by the nonlinear Schrödinger equation (NLSE), which can be written in the following form:

$$\mathrm{i}\frac{\partial q}{\partial \xi} + \frac{1}{2}\frac{\partial^2 q}{\partial \tau^2} + N^2|q|^2 q = 0 \qquad [6]$$

where q is the normalized amplitude, $\xi = z/L_{\mathrm{D}}$, $\tau = T/T_0$, and the parameter N is the soliton order, given by

$$N^2 = \frac{L_{\mathrm{D}}}{L_{\mathrm{NL}}} = \frac{\gamma P_0 T_0^2}{|\beta_2|} \qquad [7]$$

The soliton period, z_0, is given by

$$z_0 = \frac{\pi}{2} L_{\mathrm{D}} = \frac{\pi T_0^2}{2|\beta_2|} \qquad [8]$$

In writing eqn [6], the fiber losses were neglected, since fiber lengths employed in practice are relatively small. Even though higher-order solitons follow an exact periodic evolution only for integer values of N, eqn [6] can be used to describe pulse evolution for arbitrary values of N.

In practice, soliton effect compression is carried out by initially amplifying optical pulses up to the power level required for the formation of higher-order solitons. The peak optical power of the initial pulse required for the formation of the Nth-order soliton is given by

$$P_{\mathrm{N}} = 3.11\,\frac{|\beta_2|N^2}{\gamma T_{\mathrm{FWHM}}^2} \qquad [9]$$

These Nth-order solitons are then passed through the appropriate length of optical fiber to achieve a highly compressed pulse. The optimum fiber length, z_{opt}, and the optimum pulse compression factor, F_{opt}, of a soliton-effect compressor can be estimated from the following empirical relations:

$$z_{\mathrm{opt}} \approx z_0\left(\frac{0.32}{N} + \frac{1.1}{N^2}\right) \qquad [10]$$

$$F_{\mathrm{opt}} \approx 4.1N \qquad [11]$$

These relations are accurate to within a few percent for $N > 10$.

Compression factors as large as 500 have been attained using soliton-effect compressors. In fact, considering the same input pulse, the soliton-effect compressor provides a pulse compression that is larger with a fiber that is shorter than a grating-fiber compressor. However, the pulse quality is poorer, since the compressed pulse carries only a fraction of the input energy, while the remaining energy appears in the form of a broad pedestal. From a practical point of view, the pedestal is deleterious since it causes the compressed pulse to be unstable, making it unsuitable for some applications. Despite this, soliton-effect compressed pulses can still be useful because there are some techniques that can eliminate the pedestal.

One difficulty faced, when using the soliton-effect compressor, is that pulses with high peak power are required for the formation of high-order solitons in conventional fibers. As suggested by eqn [9], the use of dispersion-shifted fibers (DSFs) with small values of β_2 at the operating wavelength can reduce the peak power required for soliton generation by an order of magnitude. However, because the soliton period z_0 is inversely proportional to $|\beta_2|$ (see eqn [8]), eqn [10], indicates that longer lengths of DSFs will be required for solitons to achieve optimum compression. As a result, the total fiber loss experienced by those solitons will be larger and the loss-induced pulse broadening will have a significant impact on the compressor global performance.

Another problem introduced by the use of DSFs in soliton-effect compressors is the TOD, which generally degrades the quality of the compressed pulse. In the case of soliton-effect compressors using conventional optical fibers, the effects of TOD become significant only when the widths of the compressed pulse become very short ($\ll$1 ps). In DSFs, however, the relative importance of TOD is increased and those effects become more pronounced because the GVD parameter, β_2, is small. In this case, TOD is detrimental in the compression, even of pulses with widths of a few picoseconds, resulting in serious degradation of the optimum compression factor F_{opt}.

The result given by eqn [11] for the optimum compression factor holds only for the case of an ideal soliton-effect compressor, when high-order nonlinear and dispersive effects, such as intrapulse Raman scattering (IRS), TOD and self-steepening, are neglected. However, these effects cannot be neglected for the highly compressed pulses because of their subpicosecond widths and high intensities.

For not-too-short optical pulses (widths >50 fs) propagating not too close to the zero-dispersion wavelength of the fiber, the dominant higher-order effect is IRS, which manifests as a shift of the pulse spectrum toward the red side – the so-called soliton self-frequency shift. As a consequence, the group velocity is reduced and there is a delay of the optical pulse. This delay affects substantially the interplay between GVD and SPM that is responsible for pulse compression.

Interestingly, the IRS effect can be used to improve the quality of the compressed pulse, by removing the pedestal mentioned above. In fact, as a consequence of the change in the group velocity induced by the IRS effect, the sharp narrow spike corresponding to the compressed pulse travels more slowly than the pedestal and separates from it. Moreover, the pedestal can be removed by spectral filtering and a red-shifted, pedestal-free, and highly compressed pulse is then produced.

Compression of Fundamental Solitons

As described in the last section, the propagation of higher-order solitons provide a rapid compression method, but it suffers from the existence of residual pedestals. Some techniques can help to reduce or even eliminate those pedestals, but energy is always wasted in this process.

A less rapid technique that provides better pulse quality is adiabatic amplification of fundamental or first-order solitons, which are of primary importance in the domain of optical communications. Solitons have fixed area, so the increased energy from amplification is accommodated by an increase in

power and a decrease in width. To avoid distortion, the amplification per soliton period cannot be too great.

The same effect as adiabatic amplification can be achieved using an optical fiber with dispersion that decreases along the length of the fiber. In fact, for optical solitons, a small variation in the dispersion has a similar perturbative effect as an amplification or loss: such a variation perturbs the equilibrium between the dispersion and nonlinearity in such a way that when, for example, the dispersion decreases, the soliton pulse is compressed. It can be seen from eqn [7] that if the value of $|\beta_2|$ decreases along the fiber and to keep the soliton order N ($N = 1$ for the fundamental soliton), the pulsewidth must indeed decrease as $|\beta_2|^{1/2}$. Hence, the use of fibers with variable dispersion is viewed as a passive and effective method to control optical solitons in soliton communication systems. Dispersion-decreasing fibers (DDFs), in particular, have been recognized to be very useful for high-quality, stable, polarization-insensitive, adiabatic soliton pulse compression and soliton train generation. These fibers can be made by tapering the core diameter of a single-mode fiber during the drawing process, and hence changing the waveguide contribution to the second-order dispersion.

In a DDF of certain length, the dispersion is monotonically and smoothly decreased from an initial value to a smaller value at the end of the length according to some specified profile. Provided the dispersion variation in the DDF is sufficiently gradual, soliton compression can be an adiabatic process where an input fundamental soliton pulse can be ideally compressed as it propagates, while retaining its soliton character and conserving the energy.

The evolution of the fundamental soliton in a DDF can be described by the following NLSE:

$$\mathrm{i}\frac{\partial q}{\partial \xi} + \frac{1}{2}p(\xi)\frac{\partial^2 q}{\partial \tau^2} + |q|^2 q = 0 \qquad [12]$$

where the variable coefficient $p(\xi) = |\beta_2(\xi)/\beta_2(0)|$ takes into account the variation of dispersion along the fiber. Using the transformations $s = \int_0^\xi p(y)\mathrm{d}y$ and $U = q/\sqrt{p}$, eqn [12] assumes the form:

$$\mathrm{i}\frac{\partial U}{\partial s} + \frac{1}{2}\frac{\partial^2 U}{\partial \tau^2} + |U|^2 U = \mathrm{i}GU \qquad [13]$$

where

$$G = -\frac{1}{2p}\frac{\mathrm{d}p}{\mathrm{d}s} \qquad [14]$$

Equation [13] shows clearly that the effect of decreasing dispersion is mathematically equivalent to the effect of an optical amplifier, adding a gain term to the NLSE. The effective gain coefficient G is related to the rate at which GVD decreases along the fiber.

Since decreasing dispersion is equivalent to an effective gain, a DDF can be used in place of a conventional fiber amplifier to generate a train of ultrashort pulses. To achieve such an objective, a continuous wave (CW) beam with a weak sinusoidal modulation imposed on it is injected into the DDF. The sinusoidal modulation can be imposed, for example, by beating two optical signals. As a result of the combined effect of GVD, SPM, and decreasing GVD, the nearly CW beam is converted into a high-quality train of ultrashort solitons, whose repetition rate is governed by the frequency of the initial sinusoidal modulation.

Different approaches have been developed to determine the optimum GVD profile and its dependence on the width and peak power of input pulses. In the case of picosecond soliton pulse compression, direct numerical simulations of the NLSE show that linear, Gaussian, and exponential dispersion profiles may all be used effectively to provide ideal, adiabatic compression, where the input pulse energy is conserved and remains localized within the pulse. In this case, the final pulse compression factor at the end of a certain length L of DDF is equal to the ratio of the initial to the final second-order dispersions:

$$F_c = \frac{T_{\mathrm{FWHM}}}{T_{\mathrm{comp}}} = \frac{\beta_2(0)}{\beta_2(L)} \qquad [15]$$

Compression factors larger than 50 are possible launching input pulses with peak powers corresponding to the fundamental soliton into a DDF whose length is about one soliton period. This technique takes advantage of soliton-effect compression but requires lower peak powers and produces compressed pulses of better quality.

In the case of subpicosecond soliton pulses, the influence of higher-order nonlinear and dispersive effects must be accounted for. Some studies show that, in this case, the linear and Gaussian dispersion profiles are the most suitable to achieve high-quality, pedestal-free, adiabatic compression of fundamental solitons. However, in the presence of higher-order effects, the final compression factor is generally lower, and after reaching a maximum at a particular length of DDF, it decreases steadily. This can be explained in terms of pulse compression stabilization, which originates from a competition between the rate of dispersion decrease in the DDF and the rate of dispersion increase due to the combined effects of soliton self-frequency shift and third-order dispersion.

List of Units and Nomenclature

F_c	compression factor
F_{opt}	optimum compression factor
G	amplifier gain
L	fiber length (m)
L_D	dispersion length (m)
L_{NL}	nonlinear length (m)
N	soliton order
P_N	peak power of the Nth-order soliton (W)
P_0	pulse peak power (W)
q	normalized pulse amplitude
T_0	input pulse (s)
T_{comp}	compressed pulse FWHM (s)
T_{FWHM}	pulse full width at half maximum (s)
z_0	soliton period (m)
z_{opt}	optimum fiber length (m)
β_2	group-velocity dispersion coefficient ($s^2\,m^{-1}$)
γ	fiber nonlinearity coefficient ($W^{-1}\,m^{-1}$)
ξ	normalized distance
τ	normalized time
ϕ_g	phase shift induced by the grating pair (rad)
ω_0	pulse center frequency (rad s^{-1})
ω	optical frequency (rad s^{-1})
$\Delta\omega$	pulse spectral width (rad s^{-1})

See also

Chirped Pulse Amplification. **Fiber Gratings**. **Scattering:** Raman Scattering. **Solitons:** Soliton Communication Systems. **Ultrafast Laser Techniques:** Pulse Characterization Techniques.

Further Reading

Agrawal GP (1995) *Nonlinear Fiber Optics*, 2nd edn. San Diego: Academic Press.

Agrawal GP (1997) *Fiber-Optic Communication Systems*, 2nd edn. New York: John Wiley & Sons.

Agrawal GP (2001) *Applications of Nonlinear Fiber Optics*. San Diego: Academic Press.

Barty C, White W, Sibbett W and Trebino R (eds) (1998) Special issue on ultrafast optics. *IEEE Journal of Selected Topics on Quantum Electronics* 4: 157.

Buck JA (1995) *Fundamentals of Optical Fibers*. New York: John Wiley & Sons.

Kashyap R (1999) *Fiber Bragg Gratings*. San Diego, CA: Academic Press.

Mamyshev PV (1992) Generation and compression of femtosecond solitons in optical fibers. In: Taylor JR (ed.) *Optical Solitons – Theory and Experiment*. Cambridge, UK: Cambridge University Press.

Raman Lasers

M Santagiustina, Istituto Nazionale di Fisica della Materia, Università degli Studi di Padova, Padova, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

The Raman effect bears the name of the Indian scientist Chandrasekhara V. Raman (1888–1970) who received the Nobel Prize in the year 1930 for the discovery, published in 1928, of this scattering phenomenon.

The Raman effect is an inelastic light scattering process, occurring in gaseous, liquid, and solid media. The frequency of the scattered photon can be either down-shifted or up-shifted and the result of the process is the generation of new radiation. Frequency down-shifted and up-shifted waves are respectively called Stokes and anti-Stokes waves. The energy difference between input and output photons is either absorbed by, or released from, the atoms or molecules of the medium, which are left in different energy states; jumps can be between electronic states, but more often they are among vibrational or rotational states. The upper energy state involved in the scattering process can be either a true energy state, and this case is referred to as resonant Raman scattering, or a virtual one. The off-resonance effect has a lower probability to occur and high-power beams are needed to observe it. The emission of anti-Stokes waves requires either an inverted Raman medium or particular four-wave-mixing phase-matching conditions to be satisfied in order to occur. The second condition is referred to as coherent anti-Stokes Raman scattering. Here we will consider mainly Stokes wave generation, which is the most common in Raman lasers.

When pumping is very hard, or a seed wave is provided, stimulated Raman scattering can also occur and a high percentage of the input pump radiation can be converted into Stokes, or anti-Stokes, radiation. The stimulated process efficiency, defined as the ratio between the output Raman wave energy

and the input pump wave energy, depends on the imaginary part of the third-order nonlinear susceptibility coefficient and on the pump intensity. Stimulated Raman scattering observation dates back to 1962 during the pioneering studies with lasers sources by Woodbury and Ng; their experiment with a ruby laser was actually the first example of an intracavity Raman laser.

Since its discovery, Raman scattering has become a valuable tool for spectroscopy and a practical technique to extend laser light emission to wavelengths not directly generated by laser active materials. This fundamental property stems from the fact that Raman scattering does not require an upper energy state to occur. Hence, the frequency of the emitted radiation can be any, if a suitable incident radiation wavelength is provided. New frequencies are constrained only by the available powerful pump laser sources and by Raman medium shifts.

The frequency shifts provided by some of the materials that are most often used in Raman lasers are presented in Table 1. Finally, it is important to note that Raman scattering can also occur for the newly generated radiation through an identical process in which a Stokes photon is Raman scattered. This radiation, which is generated if the Stokes radiation intensity grows large enough, is referred to as a second-order Stokes wave. For very large pump intensity, a cascading process to several Stokes orders can take place, thus extending the range of generated frequencies to the far-infrared portion of the spectrum.

We will describe the design of Raman lasers that are most commonly used, as well as their performances, also taking into consideration the Raman medium that sets the main constraint to the laser characteristics. Continuous wave as well as pulse Raman lasers will be described and the tuning capability will be discussed.

Raman Shifters

The simplest conversion scheme is obtained when a Raman medium is placed at the output end of a laser. This single-pass setup, shown in Figure 1, provides a fairly efficient frequency shifter.

Single-pass amplification of the spontaneous scattering or of an injected seed copropagating with an intense pump is usually modeled by a couple of equations whose derivation, under a plane, continuous wave approximation, can be found in the

Table 1 Measured Raman frequency shifts and steady-state gain for some of the most important Raman media

Medium	*Frequency shift [THz]*	*Gain [cm/GW]*
Calcite ($CaCO_3$)	32.58[(1)]	5.5[(1)]
Lithium iodate ($LiIO_3$)	23.1–24.66[(2)]	4.8[(3)]
Barium nitrate ($Ba(NO_3)_2$)	31.46[(4)]	11[(5)]
Potassium gadolinium tungstate ($KGd(WO_4)_2$)	27.03 or 23[(6)]	6[(6)]
Silica fiber (SiO_2)	13.2[(7)]	0.01[(7)]
Hydrogen gas (H_2) ($p = 10$ atm, $T = 293$ K)	124.65[(8)]	1.53[(9)]
Nitrobenzene	40.35[(10)]	3[(10)]

[(1)]Pump wavelength 694 nm; from Zverev PG, Basiev TT and Prokhorov AM (1999) Stimulated Raman scattering of laser radiation in Raman crystals. *Optical Materials* 11: 335–352.
[(2)]Values refer to a propagation direction respectively perpendicular or orthogonal to the crystal axis; from Guadalberto GM and Arguello CA (1974) Raman spectra from oblique phonons in powdered samples. *Solid State Communications* 14: 911–914.
[(3)]Pump wavelength 1064 nm; from Falk J and Moshrefzadeh (1985) Oblique Raman and polariton scattering in kithium iodate. *IEEE Journal of Quantum Electronics* QE-21: 110–113.
[(4)]Pump wavelength 1064 nm; from Basiev TT, Sobol AA, Zverev PG, Ivleva LI, and Osiko VV (1999) Raman spectroscopy of crystals for stimulated Raman scattering. *Optical Materials* 11: 307–314.
[(5)]Pump wavelength 1064 nm; from Basiev TT, Voitsekhovskii VN, Zverev PG, *et al.* (1987) Conversion of tunable radiation from a laser utilizing an LiF crystal containing F_2^- color centers by stimulated Raman scattering in $Ba(NO_3)_2$ and $KGd(WO_4)_2$ crystals. *Soviet Journal of Quantum Electronics* 17: 1560–1561.
[(6)]Values refer respectively to two linear and orthogonal pump beams at the wavelength 1064 nm; from Zverev PG, Basiev TT and Prokhorov AM (1999) Stimulated Raman scattering of laser radiation in Raman crystals. *Optical Materials* 11: 335–352.
[(7)]Pump wavelength 1064 nm; from Stolen RH (1978), Fibre Raman lasers. In Ostrowsky D (ed) *Fiber and Integrated Optics*, NATO ASI Series B41. New York, Plenum Press, 157–182.
[(8)]From Bischel WK and Dyer MJ (1986) Temperature dependence of the Raman linewidth and lineshift for the Q(0)–Q(1) transition in normal and para H_2. *Physical Review* A 33: 3113–3123.
[(9)]Pump wavelength 1064 nm; from Ottusch JJ and Rockwell DA (1988) Measurements of Raman gain coefficients in hydrogen, deuterium and methane. *IEEE Journal of Quantum Electronics* 24: 2076–2080.
[(10)]Pump wavelength 694 nm; from Bloembergen N (1967) The stimulated Raman effect. *American Journal of Physics* 35: 989–1023.

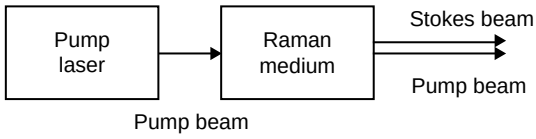

Figure 1 Experimental setup of a single-pass Raman shifter.

literature:

$$\frac{dI_S(z)}{dz} = g_R I_S(z) I_P(z) - \alpha_S I_S(z) \qquad [1]$$

I_S, I_P are the continuous wave intensities of the Stokes and pump waves, z is the propagation distance, g_R the Raman steady-state gain coefficient, and α_S, α_P are the attenuation coefficients at the Stokes and pump frequencies. The Raman gain coefficients, whose general expression as a function of the microscopic material properties can be found in many textbooks, of some of the most commonly used Raman laser media are listed in Table 1.

For short propagation lengths the medium loss can usually be neglected, and if the pump beam is well collimated, the pump intensity can be considered as a constant. For a long interaction length or very absorptive media, the evolution of the pump intensity must also be considered. Though the pump beam is depleted by the scattering process an undepleted pump approximation is often used until high conversion is reached. Under this approximation, the evolution of the pump intensity is mainly governed by diffraction spreading, by losses or by both. In bulk media, single-pass scattered radiation has a coaxial component whose divergence is dictated by the geometry of the pump beam, because Stokes radiation is close to the diffraction limit. Besides this radiation, a nonaxial component can also be observed; this phenomenon stems from the well-known effect of conical emission, which is a phase-matched parametric process. If the Raman medium is within a waveguide, losses are usually the dominating effect. In the undepleted approximation, the evolution of the pump beam intensity as a function of the propagation distance z, either dictated by diffraction or by losses, is known. Then the general solution of eqn [1] is:

$$I_S(z) = I_S(0) \exp\left[g_R \int_0^z I_P(z')dz' - \alpha_S z \right] \qquad [2]$$

where $I_S(0)$ is either the input Stokes signal intensity or an equivalent spontaneous scattering intensity. Then, the amplification gain G of the Stokes beam in the single-pass shifter, for a medium of length L, is given by

$$G = \frac{I_S(L)}{I_S(0)} = \exp\left[g_R \int_0^L I_P(z')dz' - \alpha_S L \right] \qquad [3]$$

Though single-pass Raman shifters are appealing for their simplicity, their application is hampered by the high power necessary for reaching a sufficient efficiency due to the short interaction length that can be obtained, in particular in bulk media, where diffraction dominates. Q-switched pulsed pump lasers are often required, but the propagation in the Raman medium of very powerful pulses requires special precautions to prevent other linear and nonlinear effects occurring. In fact, optical breakdown, thermal and Kerr self-focusing, four-wave mixing, and Brillouin scattering can take place, thus reducing the efficiency of the converter. Precautions may include the optimization of the pump pulse duration, of the focusing, of the interaction length, and the control of the input beam spot size. The multipass configuration has also been exploited to avoid many of the single-pass configuration shortcomings; the typical experimental setup is presented in Figure 2. Note that neither the pump nor the Stokes wave are resonant within the cavity.

The primary advantage of the multipass configuration is the lowering of the pump power required to realize an appreciable conversion. This is due to the increased total gain coefficient, that has been found to be equal to

$$G_n = G\frac{1 - R^n}{1 - R} + (n - 1)\log R \qquad [4]$$

where G is the single-pass gain of eqn [3], R the reflectivity of the mirrors at the Stokes wavelength, and n the number of passes through the Raman medium. There are two main drawbacks of multipass Raman shifters that finally limit the conversion efficiency of this setup. The first is the simultaneous generation of many undesired and uncontrolled Stokes orders, because the first-order Stokes intensity grows rapidly during propagation. The second is the

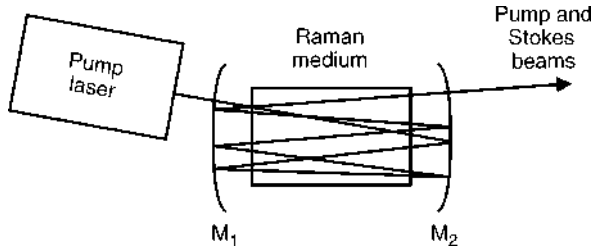

Figure 2 Experimental setup of a multipass Raman shifter. Mirrors M_1 and M_2 reflect both Stokes and pump waves.

intensity decrease associated with the dispersion of pulses, in particular if the Raman medium is a solid. These problems have limited the applications of multipass shifters to hydrogen gas.

Raman Lasers

By placing the Raman medium into an optical cavity, Stokes radiation can be selectively enhanced and Raman oscillation takes place. The two main experimental setups, either with the Raman medium placed in an external cavity or within the pump laser cavity (intracavity), are represented in Figures 3 and 4. In some cases, for the intracavity configuration, the Raman medium may also be the host of the doping laser active material.

For the external resonator Raman laser, the reflectivity of the mirrors is such that they are almost transparent for the pump but highly reflective (resonant) for the shifted radiation. The threshold condition of Raman lasers can be obtained by requiring that the net round trip gain equals the total cavity losses. The net round trip gain, under the plane wave, steady-state, undepleted pump approximation, is given by the single-pass gain G, multiplied by a factor of two to account for the forward and backward amplification. In fact, Raman scattering takes place also between the pump and the counter-propagating Stokes wave reflected by mirror M_2.

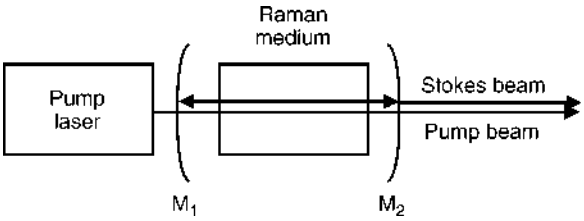

Figure 3 Experimental setup of an externally pumped Raman oscillator. Mirrors M_1 and M_2 are highly reflective for the Stokes wave but not for the pump wave.

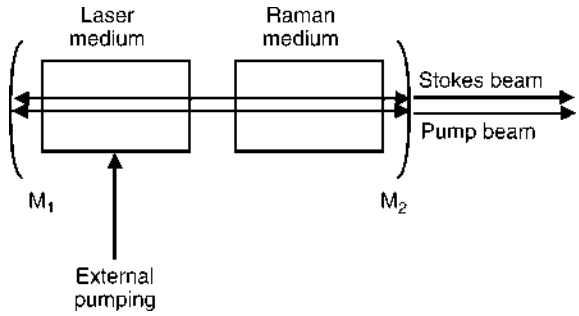

Figure 4 Experimental setup of an intracavity Raman oscillator. Mirrors M_1 and M_2 are highly reflective for both the Stokes and the pump wave.

Losses are given by the medium contribution, which is already accounted for in eqns [2] and [3], and by the energy leakage from the mirrors. Assuming that the mean reflectivities of the mirrors, at the Stokes wavelength, are respectively R_1, R_2, then the threshold condition is $GR_1R_2 = 1$. Through eqn [3] we obtain

$$g_R \int_0^L I_P(z')\mathrm{d}z' - \alpha_S L = -\log(R_1R_2) \qquad [5]$$

Equation [5] can be used to evaluate the pump intensity because at threshold low depletion occurs and the approximation taken to obtain [5] is valid. Raman steady-state gain is not very different for gases, liquids and crystals, as shown in Table 1 and oscillation thresholds depend more on the interaction length and the focusing properties of the pump and Stokes beams. Nonetheless, the type of Raman medium imposes many practical constraints for laser design, which are addressed in the next paragraphs.

With gas media it is hard to maintain a high pump intensity over the entire length of the cell, which must be used to contain the gas. Usually gas cells can be from 5 to 30 cm long; this means that Q-switched pump lasers are required to set intensities above 10^4 GW m^{-2}, that is the typical threshold value of Raman oscillation. Continuous wave oscillation of gas Raman lasers is, nonetheless, possible when the pump wavelength is near a medium resonance. In fact, the imaginary part of the third-order non-linear polarization coefficient, to which the Raman gain is proportional, is enhanced near energy transitions because of the resonance in correspondence of that photon energy. The main limitations of this setup are that pump wavelengths are dictated by the Raman medium transitions, and thus also Stokes wavelengths are fixed, and finally conversion efficiency is usually low (about 5% is reported) possibly because of the high absorption. Off-resonance, continuous wave operation can also be attained by using high finesse cavities in a doubly resonant cavity. In fact, by resonating both Stokes and pump wave with high reflectivity mirrors ($R \geq 0.999$), the threshold intensity is lowered and conversion efficiency is high (about 35% has been demonstrated). In spite of the relatively high Raman gain and the large Stokes shifts obtainable, the exploitation of gas cell Raman oscillators is mainly hampered by the need of a high-pressure-resistant vessel and a constant gas flow. The latter is needed because of the low thermal conductivity of gases that might cause local temperature variations and thermal-optical distortions that diminish the output beam uniformity and stability. Moreover the repetition rate of pump pulses must be small (below 50 Hz) so as not to impose a too strong thermal load on the gas.

Liquid media suffer basically the same problems as gases and, moreover, they show higher non-linear refractive indices that result in strong self-focusing. This lens effect results in filamentation, optical breakdown, and high beam divergence. Often Raman liquids are also toxic and very volatile; their use has been slowly but steadily decreasing.

Solid media, in particular crystals, do not show many of the problems listed above and present the advantages of solid-state technology such as compactness, low maintenance, and opportunity of integration. The high density and the restricted atom motion result in high Raman steady-state gains and minimal line broadening. Precise resonance enhancement, as well as low threshold intensities, can be obtained by choosing an adequate mirror reflectivity spectral response. Pump intensities as low as 10^3 GW m^{-2} were reported for the oscillation of an external cavity Raman laser whose medium was a 5-cm-long, barium nitrate (see Table 1) crystal, with mirror reflectivity $R = 0.49$ at the Stokes wavelength. Though the intensity threshold is about an order of magnitude smaller than for gas Raman lasers, Q-switched input pump pulses are still required. Solid-state Raman lasers can be easily forced to oscillate on a $TEM_{0,0}$ mode with divergence close to the diffraction limit even with multimode pumps. Moreover the Stokes beam is subject to a very complex clean-up process; this can be basically explained by the fact that the Stokes beam, growing from the spontaneous emission, does not retain the pump spatial characteristics. Pulse duration is equal to or slightly less than the input one. Conversion efficiency, which depends strongly on mirror reflectivity, can reach 60%, where saturation occurs due to higher-order Stokes wave generation. Solid-state Raman lasers can be easily forced to oscillate also at these higher-order Stokes frequencies if mirror reflectivity is large enough at the relative wavelengths. Second- and third-order Stokes conversion efficiencies of 30% and 20% have been reported.

The theoretical determination of the cavity pump and Stokes fields does not have an analytical solution. For continuous wave or long Q-switched pulsed lasers, eqns [1] must be extended to include all forward and backward Stokes and pump waves and boundary conditions due to the mirrors must be imposed. This problem can only be solved numerically.

The experimental setup of intracavity Raman lasers is shown in Figure 4; both the Raman and the laser medium are placed within the cavity and sometimes the two media coincide. This is, for example, the case of potassium gadolinium tungstate which can be doped with neodymium ions (see Table 1). The cavity is usually resonant for both the pump and the Stokes waves; very high conversion efficiency (up to 80%) is attained when special dichroic coatings are provided such that pump output is suppressed (high-finesse cavity for the pump). Besides the nearly diffraction-limited property of output radiation and clean output beams that are common characteristics of all Raman lasers, enhanced brightness is also observed in this setup. This effect is probably due to the high thermal loading, the consequent linear and nonlinear lens effects, and also to a short pulse generation mechanism that increases output beam intensity. This regime can be reached by reducing the Stokes cavity length with respect to the pump cavity length. A first setup is to use an additional dichroic intracavity mirror that is transparent for the pump but highly reflective for the Stokes; an alternative setup exploits a double ring cavity configuration as depicted in Figure 5.

The principle governing the compression can be explained as follows. The optical paths of the pump and Stokes pulses are made different by an amount ΔL. Then the Stokes pulse generated within the Raman medium, once re-injected into the cavity from mirror M_{S1}, will overlap with an undepleted portion of the recirculating pump pulse. Hence, another Stokes pulse is generated at each round trip, until the pump pulse is completely depleted; the Stokes output finally consists of a train of pulses of duration $\Delta L/c$, at a repetition rate equal to the cavity round-trip. This mechanism leads to pulses as short as the limit imposed by the Raman gain bandwidth. In fact, when the pulse duration becomes comparable to the inverse of the gain bandwidth, transient effects take place because of the delay in the nonlinear medium response. Ring Raman lasers have been obtained with

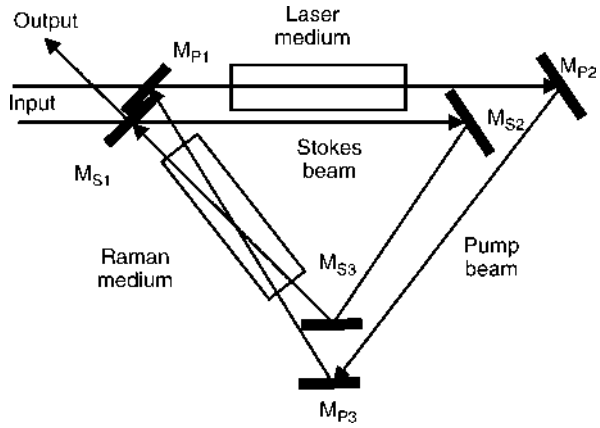

Figure 5 Experimental setup of an intracavity, double ring, Raman oscillator. Mirrors M_{S2}, M_{S3}, M_{P2}, and M_{P3} are highly reflective respectively for the Stokes and the pump wave. M_{S1}, M_{P1} are partially reflective respectively for the Stokes and pump wave.

gases (e.g., hydrogen) and crystal media (barium nitrate and lithium iodate).

Another commonly used experimental setup is that of synchronously pumped ring Raman lasers. In such devices, whose typical setup is sketched in Figure 6, a new input pump pulse is synchronized with the output Stokes pulse thus enhancing the conversion within the Raman medium.

Though electronic trigger detection is needed to lock the pump pulse time repetition rate to a multiple or submultiple of the cavity round-trip, synchronously pumped Raman lasers are relatively simple, stable, high repetition rate sources of radiation. Very compact and reliable synchronously pumped Raman lasers can be realized with optical fibers; in this case the feedback of the ring resonator can be realized by means of a fiber directional coupler. The device, besides the high integration achieved by the all-fiber technology, can also be continuously tuned over a large bandwidth.

Continuous tuning of the same laser source is a very desirable property; the Raman lasers so far described can generate radiation at wavelengths restricted to a few values. In fact, output spectral lines can be changed either by replacing the Raman medium, or the pump laser, or by exploiting cascade Raman scattering or finally resonating a different vibrational line of the same material (if there are many as for example in hydrogen gas); all these cases require substantial cavity changes. Frequency switching without setup changes is attained only in case of potassium gadolinium tungstate. For this medium there exist two independent directions, orthogonal to the crystal optical axis, for which the Raman shifts are different if the pump polarization changes. Then, the wavelength selection is just a matter of controlling the input pump polarization. Though Raman lasers can greatly expand the wavelength range of available laser sources, controlled, continuous tuning of the same Raman laser over a certain frequency band can be obtained only in a few particular cases.

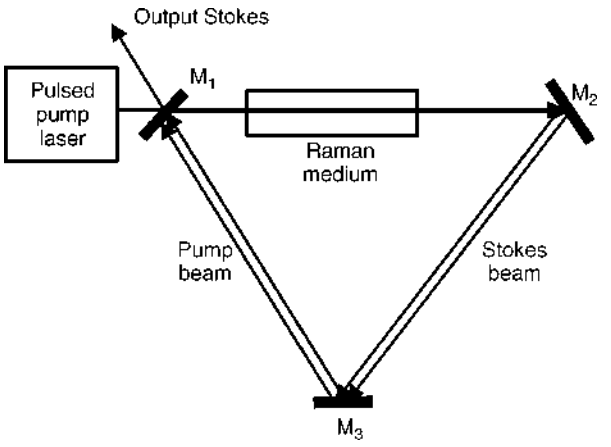

Figure 6 Experimental setup of a synchronously pumped, ring, Raman oscillator. Mirrors M_2 and M_3 are highly reflective for both the Stokes and the pump wave. M_1 is partially reflective for both the Stokes and pump wave.

Wavelength tuning can be attained by selecting the direction of propagation in polar crystals (like lithium niobate and lithium iodate). Different directions allow for wavevector matching of polariton scattering at different frequencies. Also phase matching of parametric interactions, like four-wave mixing, is a possible technique to attain wavelength tuning. However, such Raman lasers cannot operate at maximum gain, and this fact decreases the output energy, the conversion efficiency, and the stability, and increases the pump intensity threshold. Continuous tuning with an external control parameter is attained in the spin-flip Raman semiconductor laser. Semiconductors are Raman media that can be exploited to build far-infrared or even submillimeter Raman lasers, with pump intensity thresholds that are comparable with those of gas and liquid media. Though frequency shifts are fixed, under normal conditions, tuneable operation can be achieved if the semiconductor is placed in a magnetic field. Each energy level of an electron in the solid is, in fact, split into two, depending on the orientation of the electron spin with respect to the magnetic field. When there is a transition involving Raman scattering, the electron spin can change its direction. Then the wavelength of the laser can be continuously varied by changing the magnetic field applied. Tuning over a range of several micrometers in the infrared has been attained and spin-flip Raman lasers have been developed with semiconductors like InSb, $Hg_{0.77}Cd_{0.23}Te$, and $Pb_{0.88}Sn_{0.12}Te$. Use of this technique is severely limited by the need for magnetic field sources.

The simplest method to achieve tuneable radiation is through the use of a material that shows a broad Raman linewidth. Inhomogeneous broadening of the energy level is a well-known effect that takes place in noncrystalline solid media, such as glasses. Though line broadening is accompanied by a decrease in the steady-state Raman gain of about two orders of magnitude with respect to a crystalline medium, this disadvantage can be over-compensated by the increase in the Raman medium length that can be of 3 or 4 orders of magnitude. In fact, glasses are commonly exploited to make optical waveguides, in particular optical fibers, that present very low optical absorption coefficients, in the near-infrared part of the spectrum (1200–1600 nm).

Optical fiber Raman lasers based on silica telecommunication fibers are very versatile, compact,

and simple to implement. In principle the setup is not very different from the corresponding external cavity or intracavity Raman lasers so far presented; however, fiber lasers show the great advantage of an integrated, all-fiber technology.

Cavity mirrors can be replaced by fiber Bragg grating reflectors (Figure 7) that can be constructed with a very well-defined spectral reflectivity, thus allowing fine enhancement or suppression of the desired wavelengths. By splicing the Raman fiber, which can be several hundred meters long, with the gratings, the net losses are very low, finally decreasing the pump intensity threshold. The pump threshold for first-order Stokes generation can be easily as low as 200 mW; this means that diode lasers can be used for pumping. Moreover, the need for cavity alignment is eliminated, the laser stability increased and the construction simplified. This simplification allows us to build, much more easily than by using bulk optics, nested multiple resonant cavities, to generate higher-order Stokes beams (Figure 8).

Continuous-wave lasing of up to six Stokes orders has been reported for a fiber Raman laser of this type, with a total injected power of about 7 W. Another key advantage of Bragg gratings is that their reflectivity can be adjusted by means of a piezoelectric stretching of the fiber support; this allows one to control the output intensity of all waves. As previously mentioned, ring cavities can also be constructed by exploiting an all-fiber technology, if feedback is provided by a fiber directional coupler (Figure 9).

Pulsed operation of fiber Raman lasers has been obtained too; due to the long interaction length, Q-switched pump lasers are no longer needed but rather mode-locked sources can be used. This means that trains of very short laser pulses, down to the hundreds of femto-second regime, can be obtained from fiber Raman lasers, the limit being represented by the Raman bandwidth, which is as large as about 7 THz, at the half maximum, in silica fibers. Propagation of short and intense laser pulses in an optical fiber may induce several other nonlinear effects, like self-, cross-phase modulation, and four-wave mixing. Besides nonlinearity, linear dispersion and velocity mismatch become important too, due to the long propagation. All these linear and non-linear effects limit the Stokes pulse quality. In more detail: self- and cross-phase modulations lead to the spectral broadening of both pump and Stokes waves; broadening is asymmetric, because of the contemporary action of group velocity mismatch. In fact during the propagation the Raman pulse walks off the pump pulse, because of velocity mismatch, and this induces an asymmetry in gain depletion and in nonlinear index modulation. The final outcome is that Stokes waves can be asymmetric both in time and in frequency, and also highly chirped. Four-wave mixing is usually negligible, if phase-matching conditions are not imposed and Brillouin scattering is not an issue for short pulses. Very symmetric and low chirped pulses are attained as soon as the fiber Raman laser is working in the soliton regime, that is when compensation between the dispersion and the nonlinear phase shift occurs. Synchronously pumped, fiber Raman, soliton lasers are able to provide 200-fs pulses, at various wavelengths and with very high conversion efficiency (50%). Intracavity fiber Raman lasers have

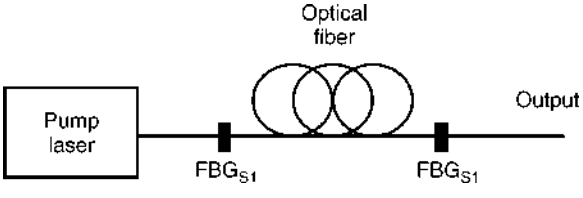

Figure 7 Experimental setup of a continuous-wave, all-fiber, Raman oscillator. FBG$_{S1}$ are fiber Bragg gratings highly reflective for both the first Stokes wave.

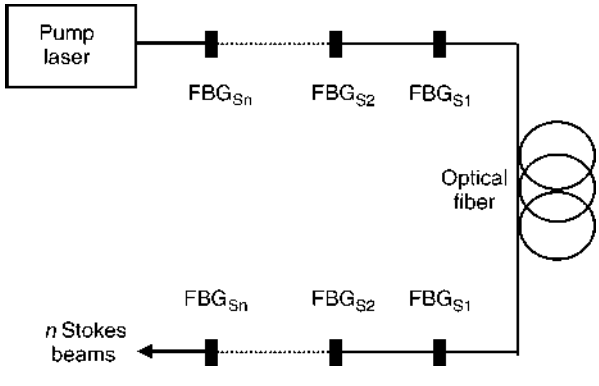

Figure 8 Experimental setup of a continuous-wave, all-fiber, *n*th-order Stokes, Raman oscillator. FBG$_{Si}$ are fiber Bragg gratings highly reflective for the *i*th Stokes wave.

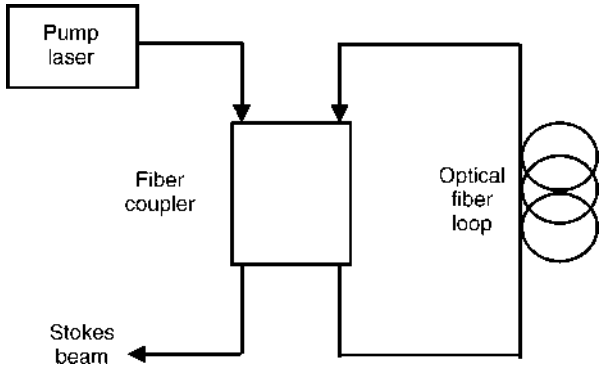

Figure 9 Experimental setup of an all-fiber, ring Raman oscillator.

also been realized by doping the fiber with neodymium, erbium, or ytterbium ions.

The extremely broad Raman spectral linewidth of the glass allows for broadband tuning of the emitted wavelength. For continuous and pulsed wave operation the change in the peak Stokes wavelength can be attained through an intracavity frequency selective element, like for example a prism or a grating. In this case tuning is not obtained in an all-fiber configuration but rather with bulk optical elements and thus additional lenses or objective are needed to collimate the Stokes wave at the fiber ends. The experimental setup is shown in Figure 10.

The laser can be continuously tuned, over about 6 THz, by adjusting the prism or grating rotation. All-fiber Raman laser tuning is possible in a double ring configuration with an inserted optical filter, whose experimental setup is shown in Figure 11. Stokes and pump waves are counterpropagating in the Raman fiber; the Stokes wave frequency is selected by the tuneable optical filter.

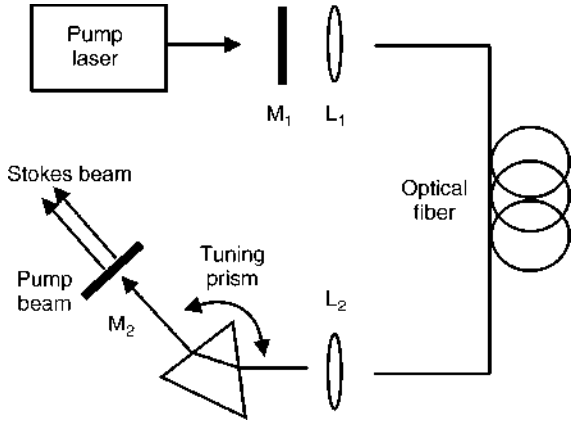

Figure 10 Experimental setup of a tuneable, fiber Raman oscillator. Mirrors M_1 and M_2 are highly reflective for the Stokes wave but not for the pump wave. L_1 and L_2 are collimating lenses or objectives. The tuning prism is rotated to control wavelength selection.

The devices of Figures 10 and 11 stop emitting Stokes radiation for frequencies at which the Raman gain has decreased to a level such that laser threshold is higher than the pump intensity. Due to the fact that Raman gain in fibers is almost constant over about 3 THz, tuning within this frequency range does not lead to any sensible decrease in conversion efficiency. For synchronously pumped ring fiber Raman lasers another technique, named time dispersion tuning, can be used. The experimental setup is shown in Figure 12.

The output fiber end is placed on a moveable mount, that controls the relative delay between the Stokes pulse and the new pump pulse at the fiber input. Since Stokes pulses are dispersed during fiber propagation, to each time slot within a pulse there corresponds a certain portion of the Stokes spectrum. By properly delaying the time of injection of the Stokes pulse back into the fiber the most powerful part of the pump pulse is overlapped to a different part of the Stokes pulse, and this causes a preferred amplification of certain frequencies with respect to others. The shift of the Stokes peak wavelength $\Delta\lambda$ is given by:

$$\Delta\lambda = \frac{\Delta l}{cLD(\lambda)} \qquad [6]$$

where Δl is the translation step, c the speed of light, L the fiber length, and $D(\lambda)$ the fiber dispersion at the Stokes wavelength λ. A total tuning range of 10 nm can be obtained with this technique that can also be used as a test measurement of fiber dispersion.

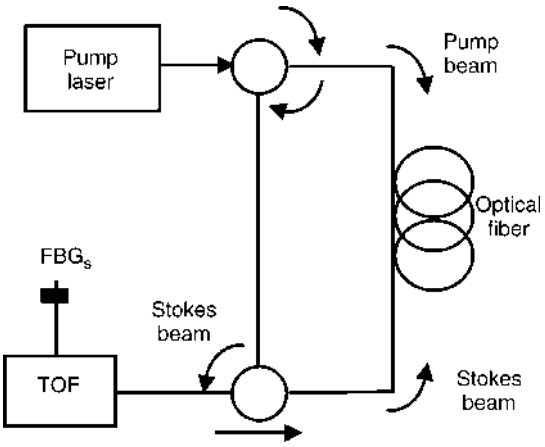

Figure 11 Experimental setup of a double-ring fiber Raman oscillator. OC_1 and OC_2 are fiber optic circulators, and TOF is a tuneable optical filter.

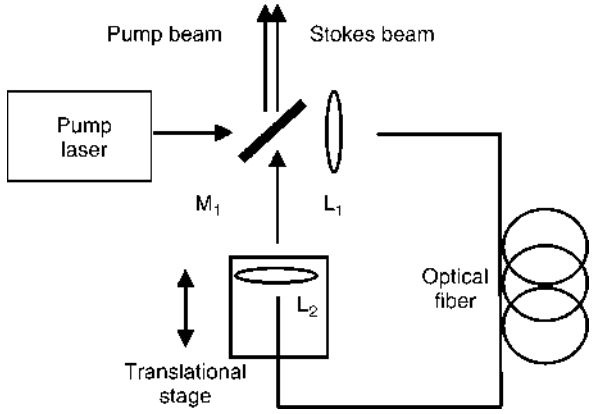

Figure 12 Experimental setup of a time dispersion tuned, ring fiber Raman oscillator. L_1 and L_2 are lenses or objectives. The output collimating lens L_2 and the fiber end are mounted on a translation stage to control wavelength selection by time delaying the re-entrant Stokes pulse.

Applications

Raman lasers have found a large variety of applications. Gas and liquid Raman lasers are fundamental tools for spectroscopic investigation. Solid-state Raman lasers have been shown to be very powerful, eye-safe sources for light detection and ranging (LIDAR) and medical applications. Fiber Raman lasers are exploited for long-haul fiber optics transmission. They can be either sources of short pulses and solitons or high-power continuous-wave sources at multiple wavelengths for subsequent link-distributed Raman broadband amplification.

List of Units and Nomenclature

Amplification gain	[dimensionless]
Efficiency	[dimensionless]
Energy	[J]
Frequency	[Hz]
Intensity	[W m^{-2}]
Length	[m]
Loss coefficient	[m^{-1}]
Power	[W]
Raman bandwidth	[Hz]
Raman gain	[m W^{-1}]
Reflectivity	[dimensionless]
Wavelength	[m]

See also

Fiber and Guided Wave Optics: Dispersion; Light Propagation; Measuring Fiber Characteristics; Nonlinear Effects (Basics); Overview. **Fiber Gratings**. **Optical Amplifiers:** Optical Amplifiers Long-Haul Transmission Systems. **Imaging:** Lidar. **Laser-Induced Damage of Optical Materials**. **Nonlinear Optics, Basics:** Four-Wave Mixing. **Magneto-Optics:** Interband Magnetoabsorption, Cyclotron Resonance, Spin Flip Raman Scattering. **Nonlinear Optics, Applications:** Pulse Compression via Nonlinear Optics; Phase Matching; Self-Focusing and Related Effects (Solitons and Multiphoton Absorption). **Optical Materials:** Optical Glasses. **Solitons:** Temporal Solitons. **Scattering:** Raman Scattering; Stimulated Scattering. **Spectroscopy:** Nonlinear Laser Spectroscopy; Raman Spectroscopy.

Further Reading

Band YB, Ackerhalt JR, Krasinski JS and Heller DF (1989) Intra-cavity Raman lasers. *IEEE Journal of Quantum Electronics* QE-25: 208–213.

Basiev TT and Powell RC (eds) (1999) Special issue on solid-state Raman lasers. *Optical Materials* 11: 301–390.

Bloembergen N (1967) The stimulated Raman effect. *American Journal of Physics* 35: 989–1023.

Eckhardt G (1966) Selection of Raman laser materials. *IEEE Journal of Quantum Electronics* QE-2: 1–8.

Grasiuk AZ and Zubarev IG (1978) High power tunable IR Raman lasers. *Applied Physics* 17: 211–232.

Kaiser W and Maier M (1972) Stimulated Rayleigh, Brillouin and Raman spectroscopy. In: Arecchi FT and Schulz-Dubois D (eds) *Laser Handbook*. Amsterdam: North Holland.

Koechner W (1999) *Solid-state Laser Engineering*. Berlin: Springer-Verlag.

Pask HM (2003) The design and operation of solid-state Raman lasers. *Progress in Quantum Electronics* 27: 3–56.

Stolen RH (1978) Fibre Raman lasers. In: Ostrowsky D (ed.) *Fiber and Integrated Optics*, NATO ASI Series B41, pp. 157–182. New York: Plenum Press.

Suto K and Nishizawa J (1994) *Semiconductor Raman Lasers*. Norwood, MA: Artech House.

Weber A (1979) *Raman Spectroscopy of Gases and Liquids*. Topics in Current Physics, vol. 11, Berlin: Springer.

Self-Focusing and Related Effects (Solitons and Multiphoton Absorption)

R L Sutherland, Science Applications International Corporation, Dayton, OH, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

When an intense beam of light with axial symmetry propagates in a medium possessing a nonlinear index of refraction, it will induce its own lens in the medium and be focused. This phenomenon, first discovered just a few years after the invention of the laser, is called self-focusing. It is one of many phenomena under the general classification of self-action, which includes such effects as self-trapping, self-bending, self-phase modulation, and self-steepening. This general class of phenomena involves action on the beam of light induced by the beam itself: light causes a change in an optical property of the medium, which

in turn produces an effect back on the light beam. We will discuss the mechanisms responsible for self-focusing in nonlinear optical media and the resultant action on the optical beam. In this we will see how the topics of self-trapping and self-phase modulation are intimately connected to self-focusing. This will lead to a discussion of related effects as well as many potential applications.

It is instructive to first review the properties of a conventional thin lens, illustrated in Figure 1. By 'thin', we mean that just before and just after the lens, the beam size and shape are approximately the same. However, the phase of the optical wave is changed on passage through the lens. We apply the paraxial (aberrationless) approximation where only rays that lie near the lens axis are considered. In this treatment the spherical surfaces of the lens are approximated as parabolic surfaces. For a symmetric double-convex lens (as shown in Figure 1), this phase change is

$$\Delta\varphi(r) = -k(n-1)\frac{r^2}{R} \equiv -\frac{kr^2}{2f}, \qquad f^{-1} = (n-1)\frac{2}{R} \qquad [1]$$

where n is the index of refraction of the lens material, and R is the radius of curvature of the lens surfaces. The light beam has a wavevector $k = \omega/c$,

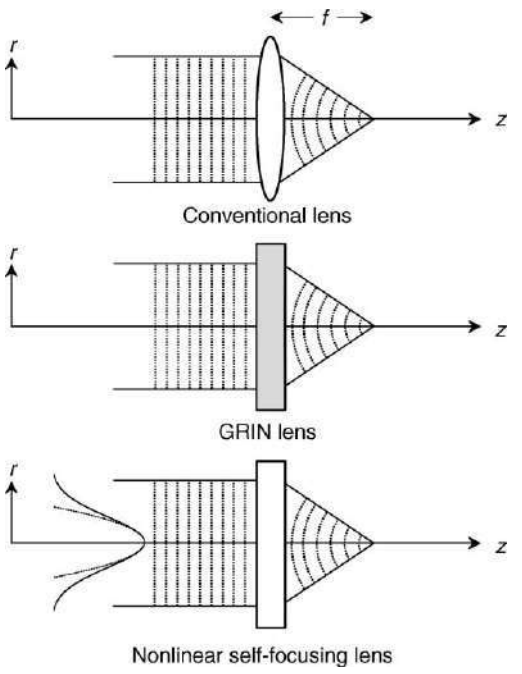

Figure 1 Focusing with a conventional lens, a graded-index (GRIN) lens, and a nonlinear self-focusing lens.

where ω is the frequency and c the speed of light in a vacuum. The phase retardation is a maximum on the lens axis (z) and decreases quadratically with the radial coordinate r. In other words, the phase advances quadratically with distance from the axis and forms a spherical, converging wave that comes to focus on the z-axis a distance f (the focal length) from the lens.

Next consider a graded-index (GRIN) lens, also illustrated in Figure 1. This is a flat slab of material with thickness L and an inhomogeneous index of refraction given by $n(r) = n_0(1 - 1/2\alpha^2 r^2)$. The index has a maximum (n_0) on the lens axis and decreases quadratically with radial distance from the axis. The constant α is a measure of the rate of the index change with r. Upon passage of light through a GRIN lens, the phase change is

$$\Delta\varphi(r) = -kn_0\alpha^2 L\frac{r^2}{2} \equiv -\frac{kr^2}{2f}, \quad f^{-1} = n_0\alpha^2 L \qquad [2]$$

Because of the parabolic nature of the index distribution, the phase retardation again decreases quadratically with r and thus forms a converging spherical wave. Notice that for both the conventional and GRIN lenses, the optical thickness of the lens is parabolic. In the conventional lens the index is constant, but the physical thickness varies continuously. In the GRIN lens, the physical thickness remains constant while the index varies continuously. Both produce the same kind of focusing action.

Finally, consider a flat slab of homogeneous material that has a nonlinear index of refraction. Light of intensity I induces a change in the refractive index given by $\Delta n = n_2 I$. The parameter n_2 is called the nonlinear refractive index coefficient and has physical units of $\mathrm{cm}^2\ \mathrm{W}^{-1}$. There are other definitions of n_2 given in the literature where the index change is given in terms of the square of the electric field, which would obviously have different physical units. We will adopt the intensity form of n_2 in this discussion. We now assume the light incident on this slab is a plane wave with a Gaussian intensity distribution given by $I(r) = I_0 \exp(-2r^2/w^2)$ as shown in Figure 1. I_0 is the peak intensity along the z-axis, and w is the radius of the beam at $1/e^2$ of the peak intensity. Consistent with the paraxial approximation, we expand the exponential function, keeping only the first two terms, to yield $I(r) \approx I_0(1 - 2r^2/w^2)$. This parabolic approximation to the intensity is sketched as the dotted curve in Figure 1. The index distribution induced in the medium by the intensity approximately follows this parabolic shape, and the resultant

phase change is

$$\Delta\varphi(r) = -k4n_2I_0\frac{L}{w^2}\frac{r^2}{2} \equiv -\frac{kr^2}{2f}, \qquad f^{-1} = \frac{4n_2I_0L}{w^2} \qquad [3]$$

Note the similarity to the GRIN lens, with the association of the peak index $n_0 \rightarrow n_2I_0$, and $\alpha^2 \rightarrow 4/w^2$. We can say that the nonlinear medium acts like a GRIN lens with an intensity-dependent focal length. However, this correspondence is only accurate within the parabolic approximation, and the nonlinear lens can introduce significant aberrations to the focal properties of the full beam.

In the paraxial approximation, the conventional, GRIN, and nonlinear lenses all have the same effect on an incident beam of light. Although we have considered only plane parallel incident waves, it should be obvious that the GRIN and nonlinear lenses will produce analogous effects on converging and diverging wavefronts as the conventional lens. The effect of a stack of nonlinear lenses in an optical train can also be analyzed in the same way as a stack of conventional lenses, with the exception that the focal length of each lens will depend on the local intensity and hence size of the beam. Finally, we note that if we change the sign of R, α^2, or n_2, we obtain a negative focal length. This is a diverging lens, and in the nonlinear optics realm the effect is called self-defocusing.

Nonlinear Index of Refraction

The concept of an intensity-dependent refractive index has its roots in the power series expression of the nonlinear dielectric polarization P in terms of the electric field E of the optical wave, which in its simplest form is

$$P = \varepsilon_0\left(\chi^{(1)}E + D^{(2)}\chi^{(2)}E^2 + D^{(3)}\chi^{(3)}E^3 + \cdots\right) = \varepsilon_0\left(\chi^{(1)} + D^{(2)}\chi^{(2)}E + D^{(3)}\chi^{(3)}E^2 + \cdots\right)E \qquad [4]$$

where $\chi^{(m)}$ $(m = 1, 2, 3\ldots)$ is called the m-th order susceptibility. The quantity $D^{(m)}$ is called the degeneracy factor and is 3 for $m = 3$ in the self-focusing case. This leads directly to the definition of a nonlinear susceptibility $\chi(E)$. All materials have first- and third-order susceptibilities. Centrosymmetric materials (e.g., gases, liquids, cubic crystals) do not possess a second-order susceptibility. Here, we will consider only centrosymmetric media and briefly discuss second-order materials later.

The refractive index is $n = [1 + \chi(E)]^{1/2}$. Consequently, the linear index (small electric field) is $n_0 = (1 + \chi^{(1)})^{1/2}$. The leading nonlinear term in $\chi(E)$ is generally small. Expanding the square root in a binomial series, the first nonlinear term in the refractive index can be approximated by $3\chi^{(3)}E^2/2n_0$. Since $I = 2\varepsilon_0 n_0 c E^2$, we see that

$$n = n_0 + n_2 I \qquad [5]$$

where $n_2 = 3\chi^{(3)}/4\varepsilon_0 n_0^2 c$. Third-order (cubic) materials that exhibit an intensity-dependent refractive index described by eqn [5] are called Kerr or Kerr-like.

Several physical mechanisms produce a refractive index of the form given in eqn [5]. A list of the significant ones is presented in Table 1. We emphasize that these mechanisms lead to eqn [5] over a limited range of intensity. For example, in materials where the main mechanism is electronic state population redistribution, when the incident intensity becomes comparable with the saturation intensity, the index begins to saturate (i.e., the change of index with

Table 1 Physical mechanisms responsible for the nonlinear index of refraction

Mechanism	*Typical n_2 ($cm^2\ W^{-1}$)*	*Typical sign*	*Typical response time (seconds)*
Electronic polarization	$\sim10^{-16}$–10^{-15}	+	$\sim10^{-15}$
Molecular orientation			
Isotropic liquids	$\sim10^{-15}$–10^{-14}	+	$\sim10^{-12}$
Liquid crystals	$\sim10^{-4}$–10^{-3}	+	$\sim10^{-3}$–10^{-2}
Electrostriction	$\sim10^{-13}$–10^{-12}	+	$\sim10^{-9}$
Population redistribution			
Atomic vapors (ω below resonance)	$\sim10^{-10}$–10^{-9}	–	$\sim10^{-8}$
Atomic vapors (ω above resonance)	$\sim10^{-10}$–10^{-9}	+	$\sim10^{-8}$
Semiconductors	$\sim10^{-6}$	+	$\sim10^{-8}$
Band filling (semiconductors)	$\sim10^{-6}$–10^{-4}	±	$\sim10^{-8}$–10^{-7}
Exciton saturation (quantum wells)	$\sim10^{-4}$	–	$\sim10^{-8}$
Thermal	$\sim10^{-8}$	±	$\sim10^{-4}$

intensity becomes sublinear). The simple expression in eqn [5] no longer applies. These types of effects can be important and will be discussed later. Finally, we mention that other mechanisms exist that produce self-focusing related effects, but they are not Kerr-like and will be briefly discussed later.

We learn from Table 1 that values of n_2 range over several orders of magnitude. High-intensity lasers are required to observe the effects of small optical nonlinearities. Hence pulsed lasers are often used in self-focusing applications. Laser pulse widths can generally range from 100 s of nanoseconds to 10 s of femtoseconds. Consequently, one must keep in mind the response time of the mechanism/material when analyzing experiments. If the pulse width is long compared to the response time, a quasi-steady-state self-focusing regime results. If the pulse is too short, transient effects will appear. Also, as one goes to longer response time mechanisms, lower intensities can be used. In some cases it is even possible to use cw lasers.

Self-Lensing

Thin nonlinear media can be employed in self-lensing applications, where self-focusing is used to modify the far field properties of a beam, just as a conventional lens would do, but with effects dependent on the beam intensity. Gaussian beams are usually employed in the paraxial approximation.

A popular technique employing this effect to measure the magnitude and sign of n_2 is the Z-scan. A typical experiment is illustrated in Figure 2. A thin nonlinear medium is scanned axially between a focusing lens and an aperture in the far field. Consider a material with positive n_2. When the sample is close to the lens the intensity is too low to induce significant self-lensing. As the sample approaches the focal point of the lens, the intensity rises and a positive self-lens is induced that focuses the beam more strongly. This displaces the focus to the left, causing the beam to spread out more in the far field and less power to pass through the aperture. As the sample passes to the right of the focal point, the positive self-lens reduces the divergence of the beam so that more power is passed through the aperture. Near the focal point the self-lens is strongest, but a thin lens at this point has no effect on the beam passing through the far-field aperture. The measured aperture transmittance as a function of scan-position Z is sketched in Figure 2. The peak and valley of this dispersion-shaped curve will be interchanged for a negative n_2. For a small aperture and fixed power of the laser, the difference between the peak and valley transmittance is proportional $|n_2|$.

The same experimental configuration is also popular for measuring small absorption coefficients or thermo-optic coefficients ($\mathrm{d}n/\mathrm{d}T < 0$ for most materials). Under steady-state conditions (cw beam), the thermal n_2 is given by

$$n_2 = \frac{\alpha\tau}{\rho C}\frac{\mathrm{d}n}{\mathrm{d}T} \qquad [6]$$

where α is the absorption coefficient, τ is the thermal diffusion time, ρ is the mass density, and C is the specific heat. This technique is extremely sensitive and can easily be observed, for example, with a ~1 mW laser pointer and a 1 cm cuvette filled with minute quantities of ink or food coloring dissolved in methanol.

This combination of self-lensing and a hard aperture has also been put to use in devices. One such device is the optical power limiter. A typical configuration is similar to the Z-scan experiment in Figure 2, with the sample fixed near the lens focal point. Thick media in the focal region have also been used. At sufficiently low optical power, the medium acts like a flat window. As the input power rises, self-lensing sets in and spreads the beam out at the aperture, limiting the transmitted power. The higher the incident power gets, the more the beam spreads. The result is a transmitted power that is clamped below a preferred level. Thermal lensing has been used for cw beams, while positive Kerr

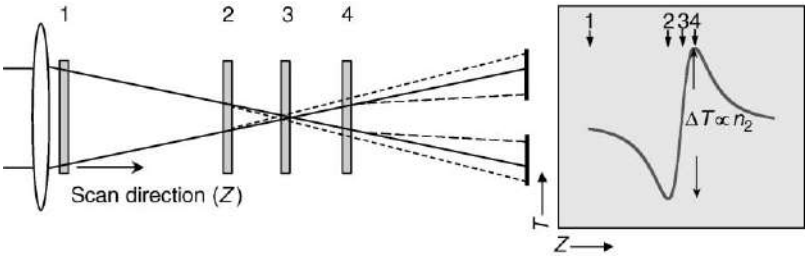

Figure 2 A Z-scan experiment. The transmittance through the aperture as a function of the position (Z) of the thin nonlinear medium is sketched at the right.

self-lensing has been employed for Q-switched lasers.

A less obvious but very important application of this configuration is for mode-locking lasers (see Figure 3). Mode-locking is a coherent phenomenon that locks the relative phase of multiple longitudinal modes in a laser having a broadband gain profile. These stabilized modes interfere to form a train of short pulses with pulsewidth inversely proportional to the gain bandwidth. A common technique for passively mode-locking a laser is to insert a saturable absorber into the laser cavity. A saturable absorber blocks low-power light, keeping the round-trip gain below the oscillation threshold, but transmits high-power light as the absorption coefficient saturates. Thus the mode-locked condition is favored for gain and oscillation since the peak power of the pulses is high. It turns out that transverse spatial modulation, due to self-focusing combined with an aperture, can produce the same type of effect in a laser cavity. The laser rod acts as both the gain medium and nonlinear lens. For example, titanium: sapphire ($Ti:Al_2O_3$) has $n_2 = 3 \times 10^{-16}\,\mathrm{cm^2\,W^{-1}}$. Thus, in a $Ti:Al_2O_3$ laser the rod produces self-lensing that modifies the beam shape propagating through an intracavity aperture. The loss due to the aperture is intensity dependent, as in the Z-scan. Therefore, the net round-trip gain of the cavity is low for low-power but high for high-power pulses, and the mode-locking condition is fulfilled. These types of lasers are called Kerr lens mode-locked (KLM) lasers. KLM lasers are generally not self-starting, but can be self-sustaining. They, therefore, require some active means of starting the mode-locking. Ultrashort pulses (<100 fs) have been obtained with a cw KLM $Ti:Al_2O_3$ laser. For pulsewidths this short, the simple quasi-steady-state self-focusing picture is not entirely adequate to describe the dynamics. Other factors, such as group-velocity dispersion (GVD) and self-phase modulation (SPM), must be considered. Moreover, since the laser rod is typically several millimeters long, it is more appropriate to model it as a thick rather than a thin nonlinear lens.

Beam Trapping

When an intense beam propagates in a thick nonlinear medium, a question arises as to the ultimate fate of the beam. A possibility that has intrigued researchers for a long time is diffractionless propagation of a beam over long distances. In other words, self-focusing exactly balances diffraction to force the beam into a nonlinear waveguiding mode. This phenomenon is called beam trapping or self-trapping. A simplified analysis of a uniform beam captures the essential physics (see Figure 4). The refractive index inside the beam is higher than outside by eqn [5]. Therefore, light incident on the 'interface' at the edge of the beam can experience total internal reflection if its angle of incidence exceeds the critical angle, dependent on the value of the nonlinear index. If the complement of the critical angle matches the diffraction angle of the beam, the beam is trapped in its own self-induced waveguide. The critical power $\mathscr{P}_{cr2}$ at which this occurs is given in Table 2. The condition for beam trapping is independent of beam size since both diffraction and self-focusing depend on beam area, and these size effects cancel.

A detailed solution of the wave equation for a Gaussian beam using the parabolic approximation yields a slightly different critical power $\mathscr{P}_{cr1}$, also given in Table 2. Numerical analyses of the wave

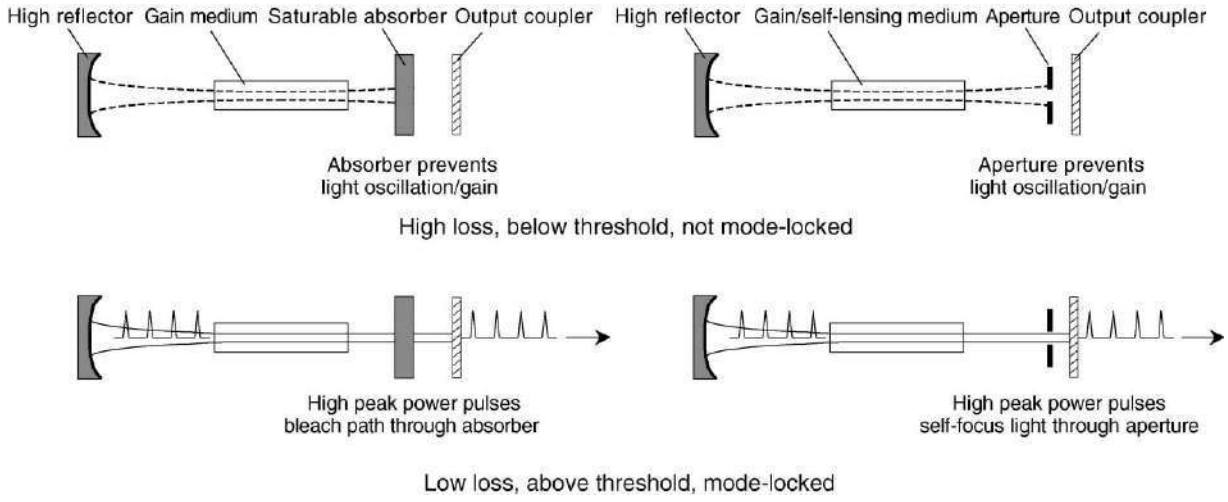

Figure 3 Laser mode-locking using a saturable absorber (left) and self-lensing in combination with an intracavity aperture (right).

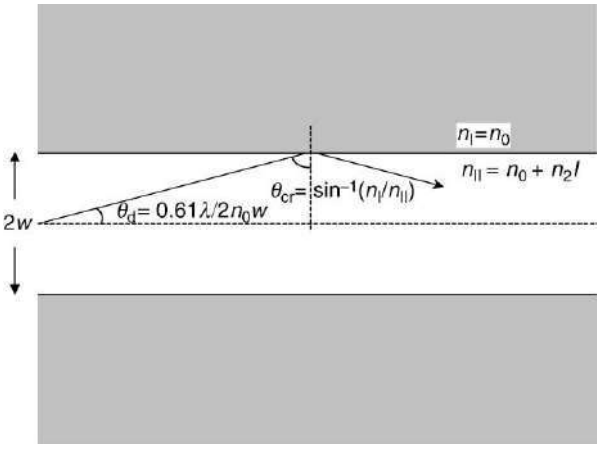

Figure 4 Self-trapping conditions for a uniform beam of intensity I and width $2w$ in a Kerr medium.

Table 2 Critical power for beam trapping (collapse)

Gaussian beam (parabolic approximation)	$\mathscr{P}_{cr1} = \frac{1}{2}\frac{\lambda^2}{4\pi n_0 n_2}$
Gaussian beam ('whole beam' collapse)	$\mathscr{P}_{cr2} \approx \frac{(1.22\pi)^2}{8}\frac{\lambda^2}{4\pi n_0 n_2}$
Townes soliton, lower bound for arbitrary beam shape	$\mathscr{P}_{crT} \approx 1.86225\frac{\lambda^2}{4\pi n_0 n_2}$
Arbitrary beam (upper bound)[a]	$\mathscr{P}_{cr3} = G[f]\frac{\lambda^2}{4\pi n_0 n_2}$

[a] $G[f]$ defined in text.

equation for a whole Gaussian beam have shown for incident power such that $\mathscr{P}_{cr1} < \mathscr{P} < \mathscr{P}_{cr2}$ the beam comes to a 'soft' focus and then diffracts (partial beam focusing), but for $\mathscr{P} > \mathscr{P}_{cr2}$ the whole beam catastrophically collapses (i.e., the beam approaches a singularity). Catastrophic self-focusing has been observed experimentally, but as the beam collapses the intensity becomes so large that additional nonlinear effects, such as multiphoton absorption and stimulated Raman scattering, come into play. These effects rob the beam of energy and terminate the collapse. At this stage the beam will then spread by diffraction.

A more rigorous analysis of this problem is obtained by examining the nonlinear wave equation in the slowly varying amplitude approximation (propagation along z), but including the transverse variation of the wave. If the electric field of the optical wave is of the form $E(\mathbf{r}) = A(\mathbf{r})\exp(ikz)$, where now $k = n_0\omega/c$, then the wave equation can be written as

$$2ik\frac{\partial A}{\partial z} + \nabla_\perp^2 A + 4\varepsilon_0 n_2 c k^2 |A|^2 A = 0 \qquad [7]$$

where the transverse Laplacian $\nabla_\perp^2$ is $\partial^2/\partial x^2 + \partial^2/\partial y^2$ in rectangular Cartesian coordinates, or $\partial^2/\partial r^2 + (1/r)\partial/\partial r$ in cylindrical coordinates (assuming azimuthal symmetry). This equation is known as the nonlinear Schrödinger equation (NLSE). In normalized coordinates $\xi = x/w$, $\eta = y/w$, $\rho = r/w$, $\zeta = z/2kw^2$, where w is the input beam radius, and $\psi = 2kw(c\varepsilon_0 n_2)^{1/2}A$, eqn [7] takes the normalized form of the NLSE:

$$i\frac{\partial\psi}{\partial\zeta} + \nabla_\perp^2\psi + |\psi|^2\psi = 0$$
$$\nabla_\perp^2 = \frac{\partial^2}{\partial\xi^2} + \frac{\partial^2}{\partial\eta^2} = \frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} \qquad [8]$$

A waveguiding (i.e., diffractionless) solution of the form $\psi(\zeta,\rho) = \exp(i\zeta)R(\rho)$ exists for eqn [8], where $R(\rho)$ is a solution of

$$\nabla_\perp^2 R - R + R^3 = 0, \quad R'(0) = 0, \quad R(\infty) = 0 \qquad [9]$$

$R(\rho)$ is known as the Townes soliton. It has a radial profile similar to a Gaussian but differing in fine detail. Since the Townes soliton exactly balances diffraction and self-focusing, it carries a critical power $\mathscr{P}_{crT}$ given by the expression in Table 2. For comparison, $(1.22\pi)^2/8 = 1.83624$. Therefore, $\mathscr{P}_{crT}$ and $\mathscr{P}_{cr2}$ differ by only about 1%. However, the Townes soliton is unstable. An analysis of the solutions of the NLSE shows that a beam will not catastrophically collapse (i.e., its intensity will not blow up) if its power is less than $\mathscr{P}_{crT}$. Thus $\mathscr{P}_{crT}$ forms a lower bound for the critical power. An upper bound can also be found from analysis of the NLSE. It depends on the input beam profile $[\psi(\rho,0) \propto f(\rho)]$ and is given as $\mathscr{P}_{cr3}$ in Table 2. The quantity $G[f]$ characterizes the input beam profile and is given by

$$G[f] = \frac{2\int |f|^2\rho d\rho \int |\nabla f|^2\rho d\rho}{\int |f|^4\rho d\rho} \qquad [10]$$

For example, a Townes soliton, a Gaussian beam, and a hyperbolic secant beam yield $G = 1.86225$, $G = 2$, and $G = 1.86257$, respectively. Thus, if the beam power is such that $\mathscr{P}_{crT} \le \mathscr{P} \le \mathscr{P}_{cr3}$, the beam may collapse, but it will definitely collapse if $\mathscr{P} > \mathscr{P}_{cr3}$. We note that if the input beam is elliptical, with ellipticity $e = b/a$ (where a and b are the semi-major and semi-minor widths, respectively), then

$$\mathscr{P}_{crT}(e) \approx \left(0.6 + 0.2\frac{e^2+1}{e}\right)\mathscr{P}_{crT} \qquad [11]$$

This expression is accurate to within 1.5% of numerical simulations of the NLSE, and also holds well for Gaussian input beams.

The conclusion is that there is no stable self-trapped $(2+1)$-D beam (two diffracting dimensions, one

propagating dimension) in a Kerr medium. The beam will diffract, blow up in a finite distance, or go through a single focusing–defocusing cycle. For an incident plane wave, the distance from the input plane to the blowup point is (assuming a Gaussian input)

$$z_f \approx \frac{0.369\pi w^2/\lambda}{(\mathscr{P}/\mathscr{P}_{cr2})^{1/2} - 0.858} \qquad [12]$$

where the approximation holds well for $\mathscr{P} > 1.2 \times \mathscr{P}_{cr2}$.

Why, then, does self-trapping appear to occur in Kerr media? Time integrated photographs show what appears to be the formation of a filament, and laser damage tracks in solids follow the shape of a self-trapped filament. This can be explained in terms of the moving focus model. In the quasi-steady-state approximation, each time slice of the pulse comes to focus independently at different times and positions. Figure 5 shows a series of snapshots of a Gaussian spatial and temporal pulse (assume axial symmetry), where the peak power of the pulse is greater than $\mathscr{P}_{cr2}$. Slices of the pulse with $\mathscr{P} > \mathscr{P}_{cr2}$ will come to a focus at a finite distance $z_f(\mathscr{P})$, which can be approximated by eqn [12]. The time at which that focus appears is determined by the time it takes for that slice of power $\mathscr{P}$ to propagate to the input plane ($z = 0$) plus propagate to $z_f(\mathscr{P})$. As illustrated in Figure 5, slice 1a takes the minimum time to come to a focus. At a later time slices 2a and 2′ come to focus simultaneously but at different positions, 2a to the left and 2′ to the right of 1a, since $\mathscr{P}(2a) > \mathscr{P}(1a)$ and $\mathscr{P}(2') < \mathscr{P}(1a)$. The solid curves in the left of Figure 5 are the trajectories of the beam rays showing where each slice will come to focus. Slice 2b, having the same power as 2a, comes to focus at the same position as 2a, but later in time, and so on. The original focus thus appears to split in two, with one branch propagating forward and the other backward. The backward branch comes to a turning point when slice 4 of maximum power in the pulse center comes to focus, then propagates forward as time slices in the back part of the pulse come to focus. The motion of the foci traces out the U-shaped curve shown to the right in Figure 5. The time-integrated picture that appears like a filament is thus just the tracks of moving foci.

The question remains, do conditions exist under which stable self-trapped beams can form? The answer is yes, and these types of beams are called spatial solitons. The $(1+1)$-dimensional NLSE admits the fundamental spatial soliton solution:

$$A(x,z) = A_0 \operatorname{sech}(x/w)\exp(iz/2kw^2)$$
$$|A_0|^2 = \frac{1}{2\varepsilon_0 c n_2 (kw)^2} \qquad [13]$$

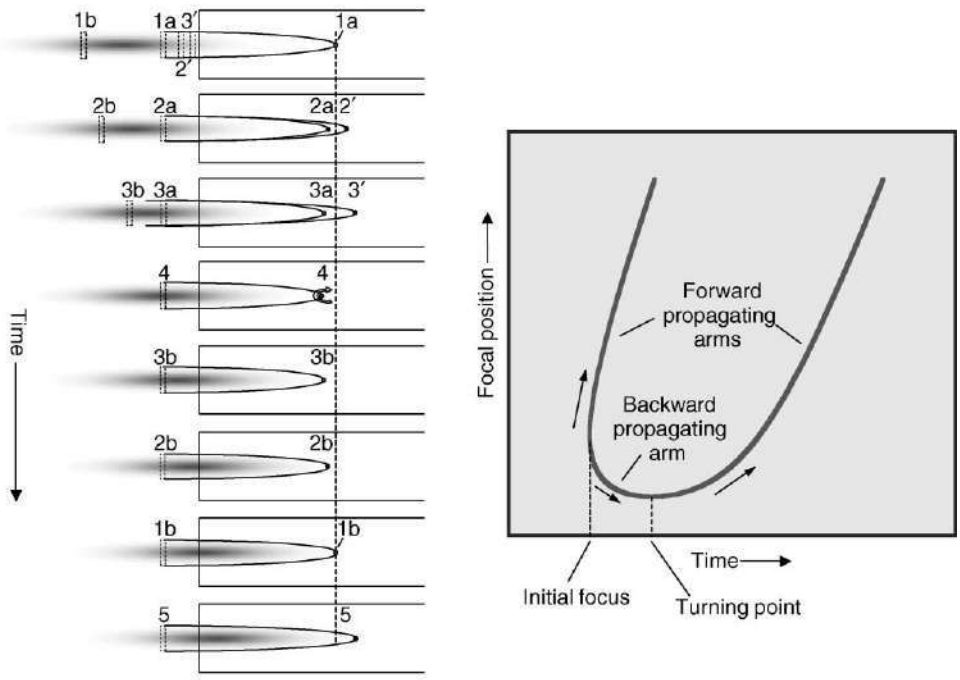

Figure 5 Moving focus model of apparent filament formation by quasi-steady-state self-focusing of a pulsed Gaussian beam. The series of snapshots on the left illustrate when and where different time slices of the pulse come to focus. The tracks of moving foci trace out the curve shown on the right.

The wave given by eqn [13] propagates along z with a constant width w along the x-axis. Such a $(1+1)$-D soliton can be formed in a slab waveguide, as illustrated in Figure 6, consisting of a nonlinear thin film deposited on a linear slab, where the wave is confined along y but can diffract along x. These types of solitons have been confirmed experimentally and shown to be quite robust. They are stable against perturbations and propagate as a self-induced waveguide. Since a soliton changes the refractive index where it propagates, it can also attract and guide other beams. The interactions or 'collisions' of solitons are an active area of current research. Intriguing potential applications include all-optical reconfigurable interconnects for telecommunications and optical computing based on the properties of soliton collisions.

Although $(2+1)$-D spatial solitons are not stable in ideal Kerr media, they can be in Kerr-like media with a saturating nonlinearity. For example, a two-level saturable absorber (e.g., an atomic vapor) at a wavelength tuned above the natural absorption frequency will exhibit a positive Kerr-like response at low intensity ($I \ll I_{\text{sat}}$), but then saturates as $(1+I/I_{\text{sat}})^{-1}$ for $I \sim I_{\text{sat}}$ or larger, where I_{sat} is the saturation intensity. This can form a stable trapped beam, as was first reported in 1974 for a cw laser in Na vapor, because as the beam self-focuses and the intensity rises, the nonlinear index levels off asymptotically until a balance between diffraction and focusing is reached. This type of situation is stable against perturbations, but may undergo almost periodic focusing–defocusing cycles.

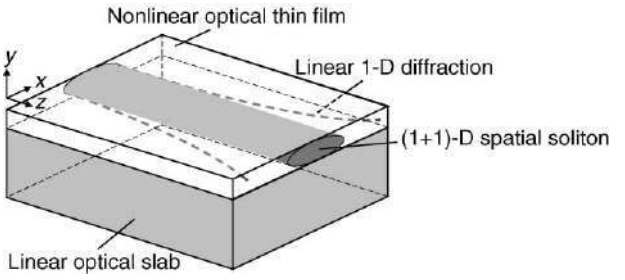

Figure 6 Spatial soliton propagation in a nonlinear slab waveguide.

Spatial solitons employing mechanisms other than the Kerr nonlinearity are also being vigorously studied. One of these is the photorefractive effect. An illustration of the generation of a slab soliton is shown in Figure 7. A Gaussian beam is focused to a line (e.g., by a cylindrical lens) and propagated as an extraordinary wave along the optic (z) axis of a uniaxial photorefractive crystal. A dc electric field is also applied along z. Charge carriers are generated in the slab beam by photo-ionization of dopant atoms and drift in the dc field. The lower resistivity in this region causes the voltage drop to be smaller and hence the field to be lower. The field induces a refractive index change by the Pockels effect ($\Delta n = -n_0^3 |r_{\text{eff}} E|/2$, where r_{eff} is the effective Pockels coefficient and n_0 is the background index). The index change is negative, but is reduced in magnitude where the field is lower. The index change depends on the local intensity, and the

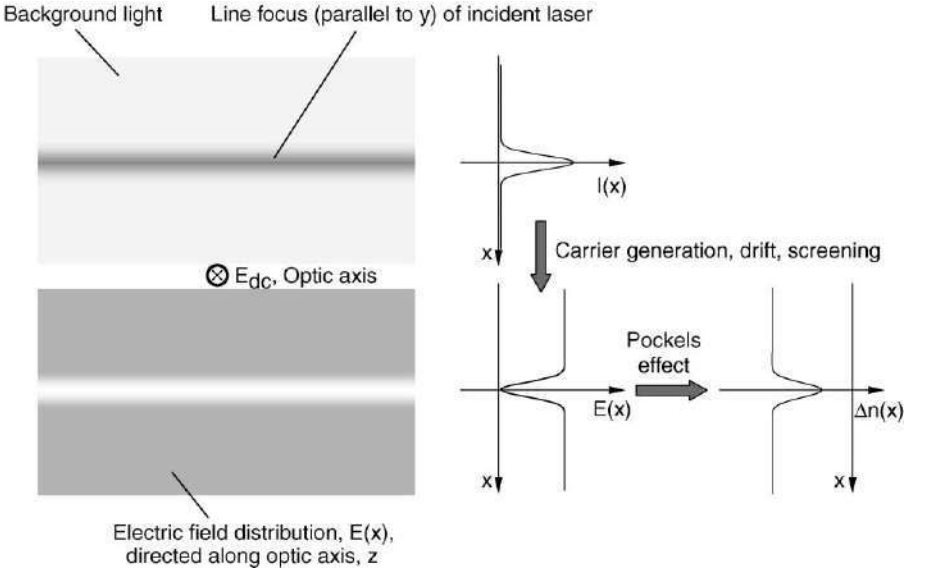

Figure 7 Mechanism of slab soliton formation in a photorefractive material.

photorefractive mechanism produces an effective saturable self-focusing nonlinearity. (There are a limited number of charge carriers, and the field cannot be reduced below zero.) This produces a waveguide that traps the slab beam, which then propagates without spreading. Thus the photorefractive soliton is stable. This effect can be observed with milliwatt-level cw lasers, but the space-charge field takes time to build up to a steady state (typically of the order of seconds). Needle (circularly symmetric) solitons have also been observed by launching a TEM_{00} Gaussian beam in a photorefractive crystal, but the theoretical description is not as simple. Other types of spatial solitons and other mechanisms have been explored using the photorefractive and photovoltaic effects. The reader is referred to the literature for a more in-depth discussion.

Another mechanism useful for soliton formation is cascaded second-order (quadratic) effects. This includes such phenomena as second-harmonic generation (SHG) and other frequency conversion processes. Consider SHG with imperfect phase matching. (The reader is referred to other articles discussing harmonic generation and phase matching.) A fundamental (F) wave generates its second harmonic (SH), but because of dispersion $n(\omega)$, the two waves travel at different phase velocities. The SH(2ω) beats with the F(ω) wave to generate, via the second-order nonlinear polarization, another wave at ω that adds to the incident F wave. This produces a net phase shift of the F wave that accumulates continuously as the waves propagate together and periodically exchange power; this is called cascading. In the nondepleted pump approximation, this phase accumulation can be expressed as

$$\Delta\varphi = -\frac{\Delta k z}{2}\left\{1 - \left[1 + \left(\frac{2\kappa\sqrt{I}}{\Delta k}\right)^2\right]^{1/2}\right\} \qquad [14]$$
$$\kappa = \frac{\omega\chi^{(2)}_{\text{eff}}}{\sqrt{2\varepsilon_0 n_0^3 c^3}}$$

where $\chi^{(2)}_{\text{eff}}$ is an effective second-order susceptibility (dependent on crystal symmetry, polarization vectors, and propagation direction), $\Delta k = 2k(\omega) - k(2\omega)$ is the phase mismatch, I is the intensity of the F wave, and n_0 is the approximate refractive index for both F and SH waves. Although not due to a refractive index change, this phase change produces a self-focusing effect for a bell-shaped intensity. For a small enough pump intensity or sufficiently large phase mismatch, an equivalent n_2 can be approximated by

$$n_2 = \frac{\omega(\chi^{(2)}_{\text{eff}})^2}{2\varepsilon_0 n_0^3 c^2 \Delta k} \qquad [15]$$

Note that the sign of n_2 depends on the sign of Δk. For $I \sim (\Delta k/2\kappa)^2$ or larger, eqn [15] is not valid and eqn [14] must be used. Note that the equivalent index change is then sublinear with respect to I, and soliton formation is stable. When pump depletion cannot be ignored, the F and SH coupled-wave equations must be solved numerically to analyze soliton dynamics. Figure 8 illustrates the difference between a beam diffracting normally and a diffractionless quadratic soliton. When an F beam is launched in a nonlinear medium, it does not immediately form a soliton, but evolves into one, shedding some energy in the process. Note that this soliton does not consist of a single frequency beam, but depends on the nonlinear coupling between the ω- and 2ω-waves that continuously exchange energy as the beams propagate. For this reason, these solitons are called simultons.

Self-Focusing in Space–Time

The moving focus model describing the self-focusing of optical pulses becomes inadequate when dispersion effects must be included. This is true in the case of ultrashort pulses (typically ~100 fs or less). To properly describe the propagation of a pulse with group velocity v_g, eqn [7] must be modified by $\partial A/\partial z \rightarrow [\partial/\partial z + (1/v_g)\partial/\partial t]A$. In addition, to include the effects of GVD, a term $-k_2 k(\partial^2 A/\partial t^2)$ is added, where $k_2 = (\partial^2 k/\partial\omega^2)$ is the GVD coefficient. For anomalous GVD, $k_2 < 0$. Taking this to be the case and defining another normalized coordinate $\tau = (t - z/v_g)/(|k_2|k)^{1/2}w$, the normalized NLSE (eqn [8]) becomes

$$i\frac{\partial\psi}{\partial\zeta} + \left(\frac{\partial^2}{\partial\xi^2} + \frac{\partial^2}{\partial\eta^2} + \frac{\partial^2}{\partial\tau^2}\right)\psi + |\psi|^2\psi = 0 \qquad [16]$$

Note the complete symmetry in eqn [16] between the transverse space (ξ, η) and time (τ) variables. There is thus a correlation between diffraction (space) and dispersion (time). The corresponding phenomenon to self-focusing is self-phase modulation (SPM). This effect denotes the additional phase retardation experienced by a pulse in a Kerr medium due to its time-dependent intensity: $\Delta\varphi(t) = -(\omega_0/c)n_2 I(t)z$, where ω_0 is the center (carrier) frequency of the pulse. SPM produces new frequencies ($d\varphi/dt$) that are red-shifted on the leading edge of the pulse and blue-shifted on the trailing edge. The frequency shifts in a self-focused filament can be quite sizeable and create a large continuum (a white light pulse) that

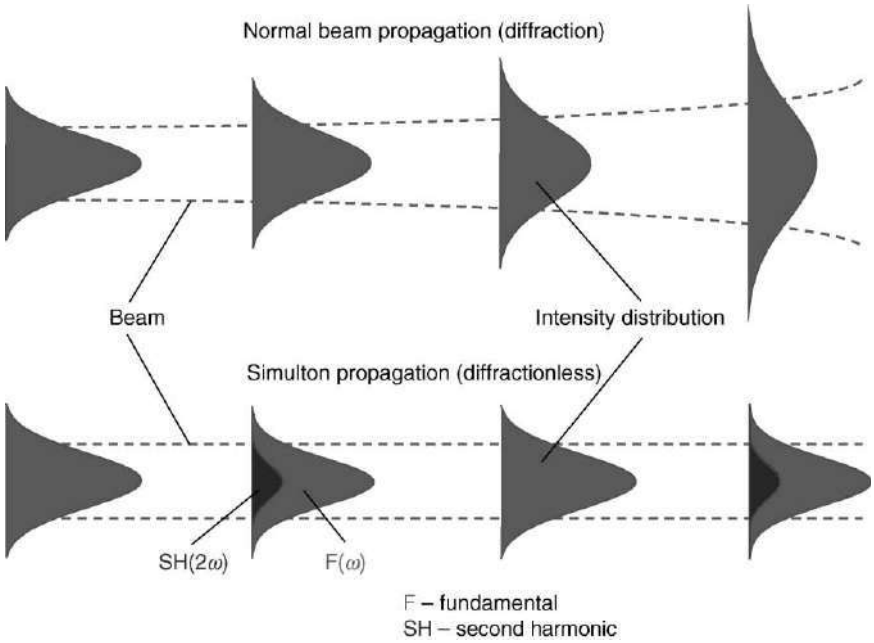

Figure 8 Comparison of normal beam propagation to simulton propagation in a quadratic nonlinear medium. In a simulton, the F and SH waves periodically exchange energy as they co-propagate in the medium.

can be used for ultrafast spectroscopy. In a medium with normal GVD the pulse spreads in time because blue frequencies in the trailing edge travel slower than red frequencies in the leading edge. However, for anomalous GVD the opposite is true, and the pulse can 'focus' or collapse in time.

The simultaneous spatiotemporal collapse of a beam of ultrashort pulses leads to the consideration of $(3+1)$-D soliton generation, where the third 'transverse' dimension is time. While the NLSE leads to unstable solitons for transverse dimensions >1, a saturating nonlinearity could possibly bound the collapse and lead to the formation of stable, diffractionless, dispersionless pulses. These types of pulses are called light bullets. Light bullets have a particle-like nature and display many interesting, exotic properties that could lead to ultrafast switching rates for optical digital logic systems in telecommunications and computing.

The saturable nature of self-focusing in quadratic nonlinear media makes them natural candidates in the search for light bullets, and $(2+1)$-D spatiotemporal simultons have been observed in crystals of lithium iodate. The drawback of ultrashort pulse simulton generation in quadratic media is the added complications of requiring anomalous GVD at both ω and 2ω, and dealing with the effects of group-velocity mismatch (GVM). The phenomenon of GVM involves the different group velocities of the F and SH pulses, which makes them walk away from one another. Without the physical overlap of these two pulses, the simulton will cease to exist. By tilting the amplitude front of the pulses with respect to their phase fronts, it is possible (over a limited distance) to compensate pulse walkoff due to GVM with spatial walkoff due to the birefringence of the quadratic medium. Accomplishing this, however, requires the beam to be in the form of a narrow stripe, and the propagation distance over which pulse overlap is good is greater for a longer stripe. This can be achieved by focusing a beam onto the medium with a cylindrical lens. The simulton then propagates without a change of width temporally or spatially (in 1-D). This is not a light bullet, however, because the confinement is not in all three dimensions.

The elliptic nature of the diffraction–dispersion operator in eqn [16] makes pulse collapse a possibility in nonlinear media with anomalous GVD. With normal GVD, the operator is hyperbolic. For such a medium, it has been predicted and observed that instead of collapsing, a self-focused pulse in a Kerr medium will split in time. Since the red-shifted frequencies generated at the front of the pulse travel faster than the blue-shifted frequencies produced at the back, the front of the pulse separates from the back part. Collapse of the pulse, as well as self-trapping, is not favored under conditions of pulse splitting.

Given this property of the NLSE, it is somewhat surprising that the reportedly first observation of light bullets was in a quadratic medium with normal GVD.

Focusing a beam of Gaussian profile in both space and time near the entrance to a lithium triborate crystal under the condition of a large positive phase mismatch, the pulses propagated without spread temporally or spatially (2D) over a distance ~20 mm (the length of the crystal). However, the pulse did not retain its shape, but evolved into a so-called 'X-wave', or X-shaped soliton. A generic X-shaped soliton (or, strictly speaking, simulton) is shown in Figure 9. The pulse has a conical (clepsydra, or hourglass-shaped) 3D structure with a central hump and slowly decaying conical tails. Interestingly, an X-shaped pulse has been shown to propagate without spreading in a linear optical medium exhibiting normal GVD. But such a pulse must be specially contrived at the input and may be impossible in many cases. In a quadratic nonlinear medium with normal GVD, the X-shaped light bullet appears to form spontaneously. The physics of this evolution lies buried in the coupled wave equations for the fundamental and second harmonic. In the strong phase-mismatched SHG process, weak perturbations in the waves at specific spatial and temporal frequencies can grow exponentially. The hyperbolic nature of the diffraction–dispersion operator with normal GVD leads to amplification of conical wave perturbations with frequencies approaching that of the X-wave. This amplification of the proper frequency components acts as a trigger that drives the evolution of the Gaussian spatiotemporal input pulse into the X-wave. This basic shape has been confirmed in experiments that outline the spatiotemporal intensity profile of the pulse exiting the rear of the crystal.

Other Related Effects (Multiphoton Nonlinearities)

At this point we should note that the susceptibility is frequency dependent. Optical waves couple principally to bound electrons, which can only occupy discrete energy levels. When the photon energy $\hbar\omega$ is nearly resonant with a difference between these energy levels, the dielectric response of the medium is enhanced. This well-known property of linear optics, accounting for absorption of light as well as normal and anomalous dispersion, is also applicable in nonlinear optics. The third-order susceptibility is actually a complex number. A sketch of its real and imaginary parts is given in Figure 10, the real part relating to n_2. The imaginary part is associated with optical loss, if positive, or gain when it is negative. In the present case the frequency dependence of the susceptibility displays a resonance (i.e., peak of the imaginary part) at a frequency $\omega = 2\omega_0$. Referring to Figure 10, when the optical frequency is significantly less than ω_0, the imaginary part of $\chi^{(3)}$ is small compared to the real part and nonlinear refraction dominates. However, when the frequency is comparable to ω_0, the resonance at $2\omega_0$ plays a major role. The real part of $\chi^{(3)}$ is resonantly enhanced, but the imaginary part signifies loss: two photons are simultaneously absorbed. This is called two-photon absorption (2PA). The intensity-dependent absorption loss per unit length is given by $\alpha_2 I$, where α_2 is called the two-photon absorption coefficient (often designated by β) and has physical units of cm W^{-1}.

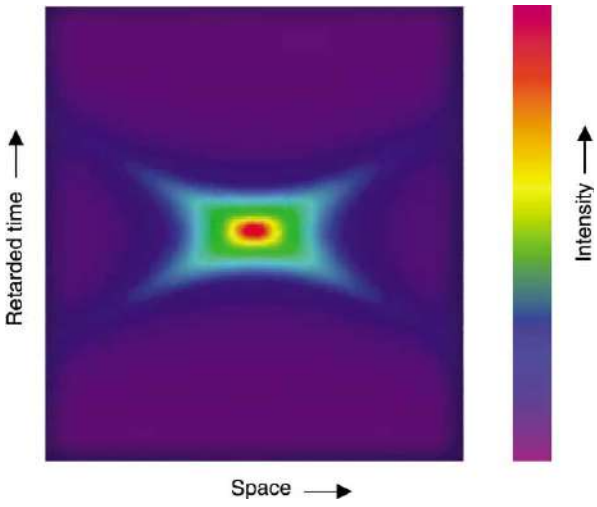

Figure 9 Generic shape of an X-wave light bullet as a function of space and retarded time ($t - z/v_g$) coordinates. The light bullet has radial symmetry.

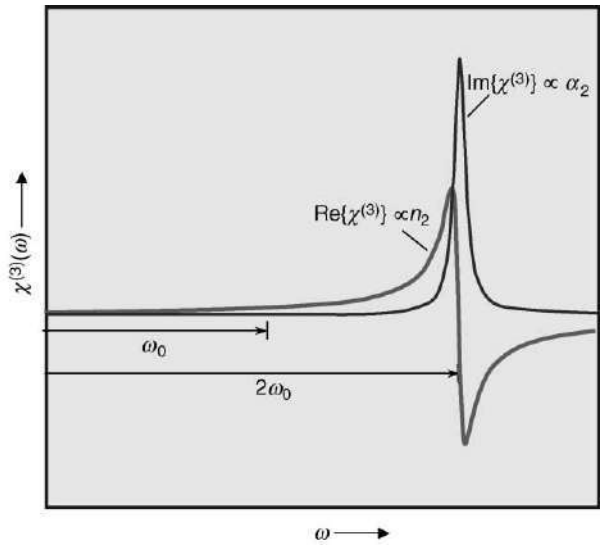

Figure 10 Sketch of the real and imaginary parts of the frequency-dependent third-order susceptibility $\chi^{(3)}$ illustrating a two-photon resonance.

In many materials at a variety of optical frequencies, 2PA will naturally accompany self-focusing. Since these two phenomena are both intensity dependent, while one decreases the intensity and the other increases it, there is an obvious interplay between the two. We illustrate the effect of 2PA in Figure 11 by comparison of the phase retardance of a Gaussian beam (in the parabolic approximation) with and without 2PA in a 'thin' medium, for which the size of an optical beam is nearly constant (diffraction ignored). Obviously, $\Delta\varphi(r)$ does not reproduce the parabolic shape when strong 2PA is present. Actually, it is possible to have multiphoton absorption (MPA) of m photons with corresponding coefficient α_m (related to $\chi^{(m)}$). The effects of 3PA and 4PA are also shown in Figure 11. 2PA, or more generally MPA, distorts the phase and modifies the focusing properties of the beam. This shows up in the Z-scan, for example, as a distortion of the dispersion-like transmittance curve: the valley is deeper and the peak is suppressed.

The reason for this phase distortion is that nonlinear absorption changes the shape of the beam: absorption is stronger where the intensity is higher. For MPA, the intensity at position z in a thin medium is

$$I(r,z) = \frac{I(r,0)}{\left[1 + (m-1)\alpha_m I^{m-1}(r,0)z\right]^{1/(m-1)}} \qquad [17]$$

The other obvious effect of MPA is that it reduces the strength of self-focusing since it robs power from the beam. The interplay of the two effects shows up more in thick media where the beam can diffract and change size. When self-focusing is dominant at the front of the medium, the intensity will start to blow up as the beam collapses, but the intensity becomes so large that MPA becomes much stronger. This robs energy from the beam and can stop or reverse the collapse. When MPA is dominant at the start, the beam will still self-focus, but MPA flattens the beam in the center causing the edges of the beam to experience dramatic diffraction effects; for example, a ring structure can develop.

The intensity of a self-focused filament can be quite high. At a level $\sim 10^{11}$ W cm^{-2} or higher, it is possible for MPA to lead to multiphoton ionization (MPI) of the medium. This is illustrated in Figure 12 where the absorption of m photons promotes an electron from the ground state to the continuum in a gas, or from the valence band to the conduction band in a solid. The ionization rate is $\sigma_m I^m$, where m depends on the optical frequency and the ionization potential of the material. For example, at a wavelength of 810 nm, $m = 11$ for nitrogen and $\sigma_{11} = 5.08 \times 10^{-144}$ s^{-1} cm^{22} W^{-11}. We note that for intensities larger than $\sim 10^{14}$ W cm^{-2}, the dominant ionization mechanism is tunneling, where a 'quivering' electron has sufficient energy to tunnel through the coulombic barrier. MPI can create a plasma whose optical properties depend on the density of free electrons. The refractive index of the plasma is nonuniform, and for $\omega > \omega_p = (N_e e^2/\varepsilon_0 m_e)^{1/2}$, the plasma frequency is <1; N_e, e, and m_e are the electron density, charge, and mass, respectively. Hence, plasma formation leads to defocusing. Consequently, self-focusing in air (which has a positive n_2) can lead to plasma formation by MPI, and the resultant defocusing can stabilize the formation of a self-trapped filament of light propagating over a distance of meters. These filaments have been called 'gas-induced solitons'. High peak power femtosecond pulses are typically used to generate these. The pulse width is short compared to the mean

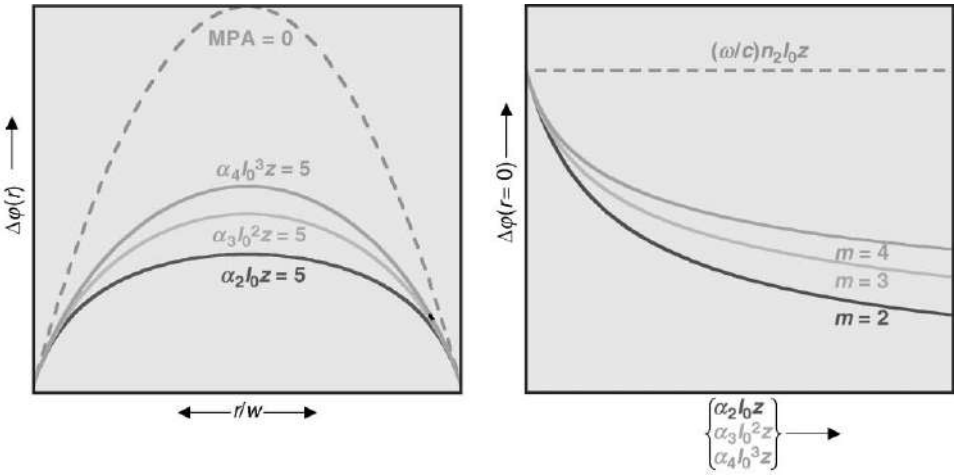

Figure 11 The effects of two-, three-, and four-photon absorption on the intensity-induced phase retardation due to a Kerr self-focusing nonlinearity. The radial dependence of $\Delta\varphi$ on MPA in the parabolic approximation is shown on the left, and the peak of $\Delta\varphi$ as a function of nonlinear absorption loss is plotted on the right.

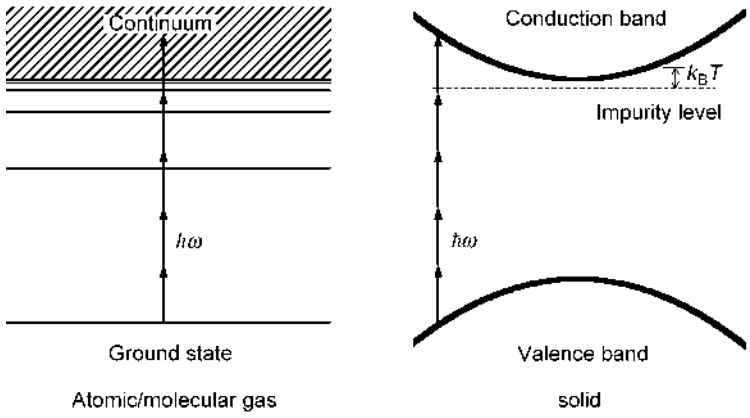

Figure 12 Multiphoton ionization in a gas (left) and a solid (right).

collision time of molecules, and the energy is low so that plasma heating is negligible.

The electron density in plasmas induced by MPI in gases is typically low enough that absorption by free electrons can be ignored. This is generally not the case in solids. MPI can provide enough 'seed' electrons to induce avalanche ionization. As illustrated in Figure 12, impurities may also supply these seed electrons by thermal ionization if they lie within an energy $\sim k_B T$ of the conduction band (k_B is the Boltzmann constant and T is the temperature). Free electrons can absorb electromagnetic energy while colliding with ions or atoms (a process called inverse bremsstrauhlung) and accelerate in a high-intensity field. They strip electrons from other atoms or molecules by impact ionization. The process multiplies free electrons and avalanche ionization ensues. A plasma forms, absorbs more energy and expands, developing a shock wave that can lead to a permanent crack in the solid. This phenomenon is called optical breakdown. Typically observed with nanosecond and picosecond pulses, permanent laser damage of the material also robs energy from the beam and generally terminates the collapse of a self-focused beam. When the electron density is sufficiently high so that $\omega_p > \omega$, the plasma absorbs and scatters light, a process that can be put to use in optical power limiters.

Many additional nonlinear phenomena play important roles when the intensity of a self-focused beam gets this large (e.g., stimulated Raman scattering, stimulated Brillouin scattering, etc.). The reader is referred to the literature on nonlinear optics for more in-depth discussions of these effects.

List of Units and Nomenclature

Critical power	[W]
Electric field	[V m^{-1}]
Four-photon absorption coefficient	[cm^5 W^{-3}]
Intensity	[W cm^{-2}]
Multiphoton (m-photon) absorption coefficient	[cm^{2m-3} W$^{-(m-1)}$]
Multiphoton (m-photon) ionization rate coefficient	[s^{-1} cm^{2m} W^{-m}]
Nonlinear refractive index coefficient	[cm^2 W^{-1}]
Pockels coefficient	[m V^{-1}]
Second-order susceptibility	[m V^{-1}]
Third-order susceptibility	[m^2 V^{-2}]
Three-photon absorption coefficient	[cm^3 W^{-2}]
Two-photon absorption coefficient	[cm W^{-1}]
Vacuum permittivity	[C V^{-1} m^{-1}]

See also

Laser-induced Damage of Optical Materials. **Materials Characterization Techniques:** $\chi^{(3)}$. **Nonlinear Optics, Applications:** Pulse Compression via Nonlinear Optics. **Nonlinear Optics, Basics:** Cascading; Kramers–Krönig Relations in Nonlinear Optics; Nomenclature and Units; Photorefraction; Ultrafast and Intense-Field Nonlinear Optics; $\chi^{(2)}$–Harmonic Generation. **Photonic Crystals**: Nonlinear Optics in Photonic Crystal Fibers. **Scattering:** Stimulated Scattering. **Solitons:** Bright Spatial Solitons.

Further Reading

Boyd RW (1992) *Nonlinear Optics*. Boston, MA: Academic Press.

Butcher PN and Cotter D (1990) *The Elements of Nonlinear Optics*. New York: Cambridge University Press.

Chiao RY, Garmire E and Townes CH (1964) Self-trapping of optical beams. *Physics Review Letters* 13: 479–482.

Crosignani B and Salamo G (2002) Photorefractive solitons. *Optics and Photonics News* 13: 38–41.
Kivshar YS and Stegeman GI (2002) Spatial optical solitons. *Optics and Photonics News* 13: 59–63.
Marburger JH (1975) Self-focusing: theory. *Progress in Quantum Electronics* 4: 35–110.
Sheik-bahae M, Said AA, Wei TH, Hagan DJ and Van Stryland EW (1989) Sensitive measurement of optical nonlinearities using a single beam. *IEEE Journal of Quantum Electronics* 26: 760–769.
Sheldon SJ, Knight LV and Thorne JM (1982) Laser-induced thermal lens effect: a new theoretical model. *Applied Optics* 21: 1663–1669.
Shen Y (1984) *The Principles of Nonlinear Optics*. New York: Wiley.
Silberberg Y (1990) Collapse of optical pulses. *Optics Letters* 15: 1282–1284.
Sutherland RL (2003) *Handbook of Nonlinear Optics*, 2nd edn. New York: Marcel Dekker.
Torner L and Sukhorukov AP (2002) Quadratic solitons. *Optics and Photonics News* 13: 42–47.
Trillo S and Torruellas WE (eds) (2001) *Spatial Solitons*. Berlin: Springer-Verlag.
Wise F and Di Trapani P (2002) The hunt for light bullets – spatiotemporal solitons. *Optics and Photonics News* 13: 28–32.

Three-Dimensional Microfabrication

S M Kuebler, University of Central Florida, Orlando, FL, USA

M Rumi, Georgia Institute of Technology, Atlanta, GA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Nonlinear optical three-dimensional microfabrication (3DM) is a photolithographic technique that enables topologically complex 3D microstructures with feature size as small as 1 μm or less to be generated in a single exposure step by nonlinear photopatterning in a material. In the broadest context, 3DM encompasses schemes based on single-beam serial patterning and parallel patterning achieved by multiple-beam interference exposure. In both cases, the photopatterning is initiated by coherent multiphoton excitation (MPE). The material may be a glass, a polymerizable resin, or even a heterogeneous composite, such as a resin containing dispersed nano-particles. The limiting resolution, or the dimensions of the smallest feature that can be generated, is determined by the optical parameters of the excitation geometry and the physical and chemical response of the material to MPE.

3DM offers great promise as a tool for generating complex microdevices, such as micro-electromechanical systems (MEMS), microfluidics, and micro-optical components. New frontiers in the technology of microdevices require that complex 3D structures can be produced in one or more materials that are optimized for a target application in terms of physical, chemical, and mechanical properties. Rapid and versatile fabrication is also essential. Extensive work has been done with conventional photolithography or electron-beam lithography, and more recently using soft-lithography approaches, such as microcontact printing. 3DM provides a new set of tools that overcomes some of the limitations inherent to other methods.

Single-Beam Three-Dimensional Microfabrication

3D Confinement

The simplest form of 3DM involves patterning a microstructure within a material using a single tightly focused laser beam that can activate the medium or one of its constituents by two-photon excitation (TPE) (Figure 1). The TPE initiates a photochemical or photophysical process (e.g., a polymerization reaction or a phase change) within the focal volume and results in an irreversible change in the material. The pattern of the target structure is impressed point by point as the focus is translated within the volume of the material. For some applications the final 3D microstructure is obtained when the unexposed material surrounding the photopatterned regions is removed. In the first report of this type of 3DM in 1990 two-photon-induced polymerization (TPIP) of an acrylate resin was used to fabricate free-standing 3D micron-scale letters on a substrate (Figure 2). The unexposed material does not have to be removed in a post-exposure process when the target structure consists of the phototransformed material supported within the matrix of the unexposed media, as in the case of 3DM of waveguide circuits embedded within a host slab.

3DM is possible because, under tight focusing, TPE and the subsequent material transformation are

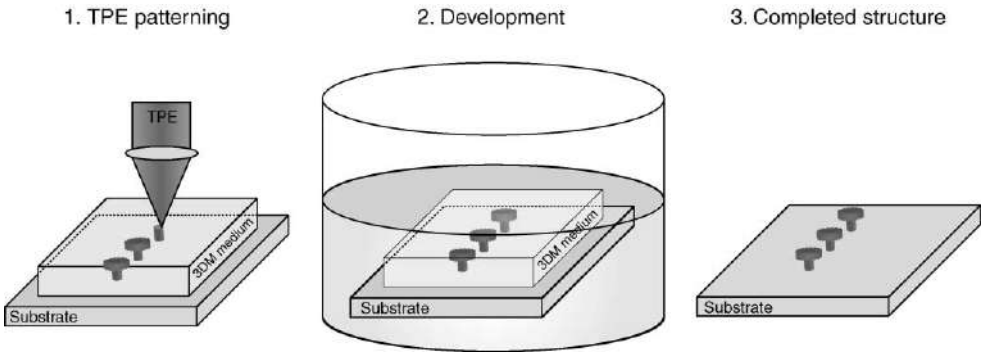

Figure 1 General scheme for single-beam 3DM based on TPE. (1. TPE patterning): The 3DM medium is supported on a substrate and translated in 3D-space relative to a tightly focused laser beam. The laser transforms the material only in the region around the focal point through a photochemical or photophysical process initiated by TPE. (2. Development): In cases where the final structure is needed free from the unexposed material, the sample is treated with a 'developer' that removes the unexposed material, for example by dissolving it in a solvent. (3. Completed structure): Following removal of the unexposed material, the final 3D structure is obtained free-standing on the substrate.

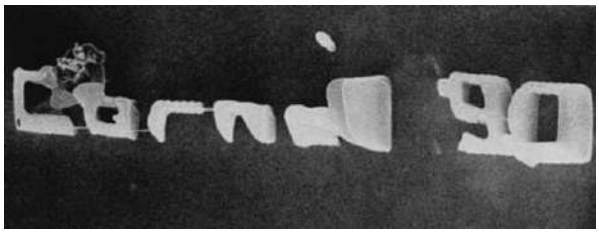

Figure 2 Scanning electron micrograph of a structure obtained by TPIP-3DM in an acrylate-ester resin. The excitation wavelength was 630 nm. The voxel dwell time was 10 ms along the letters. The thin filaments connecting some of the letters are approximately 100 nm wide. Reproduced from Strickler JH and Webb WW (1990) Two-photon excitation in laser scanning fluorscence microscopy. *Proceedings of the Society of Photo-Optical Instrumentation Engineers* 1398: 107–118. Copyright (1990), with permission of SPIE.

confined at the focus within a volume of $\sim(\lambda/n_0)^3$, where λ and n_0 are the vacuum wavelength of the exciting radiation and the refractive index of the material, respectively. Degenerate TPE is a resonant third-order nonlinear optical process in which a species is promoted to an excited electronic state by the simultaneous absorption of two photons of equal energy. Relative to conventional one-photon excitation (OPE), TPE is typically achieved using longer-wavelength radiation (often red or near infrared), for which the combined energy of two photons is sufficient to promote the species into one of its lower lying excited electronic states (Figure 3, left). The rate of TPE, that is the number of species excited per unit volume and time, $R_{2h\nu}$, is proportional to the square of the intensity, I. For a focused Gaussian beam, the on-axis intensity attains its maximum, I_0, at the focus and decreases with distance from the focal plane, z, approximately as $I \propto z^{-2}$. For a focused Gaussian beam, the on-axis intensity is given by $I = I_0/[1 + (z/z_R)^2]$, where z_R is the Rayleigh range of the focused wavefront. Thus, $I \propto z^{-2}$ in the limit of $z \gg z_R$. Consequently, the excitation rate decreases as $R_{2h\nu} \propto z^{-4}$. Similarly, the excitation rate decreases rapidly away from the propagation axis. This results in a region of high excitation, and thus material transformation, that is tightly confined both laterally and longitudinally in space about the focal point (Figure 3, right). This situation may be contrasted with that for OPE, for which the excitation rate, $R_{1h\nu}$, is linearly proportional to I, so $R_{1h\nu} \propto z^{-2}$. The weaker dependence of $R_{1h\nu}$ on z results in material being excited appreciably throughout the irradiation volume, with less longitudinal confinement at the focus than is achieved under TPE. Similar arguments can be used to describe 3DM involving higher-order excitation, for which the excitation is confined to an even smaller volume for a given λ.

The rate for molecular n-photon excitation, $R_{nh\nu}$, represents the average number of molecules that are excited per unit volume per unit time and is given by

$$R_{nh\nu} = \frac{1}{n(h\nu)^n}\sigma^{(n)}NI^n \qquad [1]$$

where n is the order of the excitation process, h is Planck's constant, ν is the frequency of the exciting radiation, $\sigma^{(n)}$ is a molecular n-photon absorption cross-section, which has units of [$\mathrm{cm}^{2n}\,\mathrm{s}^{n-1}\,\mathrm{photon}^{-(n-1)}$], N is the number density of the photo-excitable species [cm^{-3}], and I is the intensity [$\mathrm{W\,cm}^{-2}$]. The factor of $1/n$ accounts for the fact that n photons must be absorbed to promote a single molecule to an excited state. Note that $\sigma^{(n)}$ is a

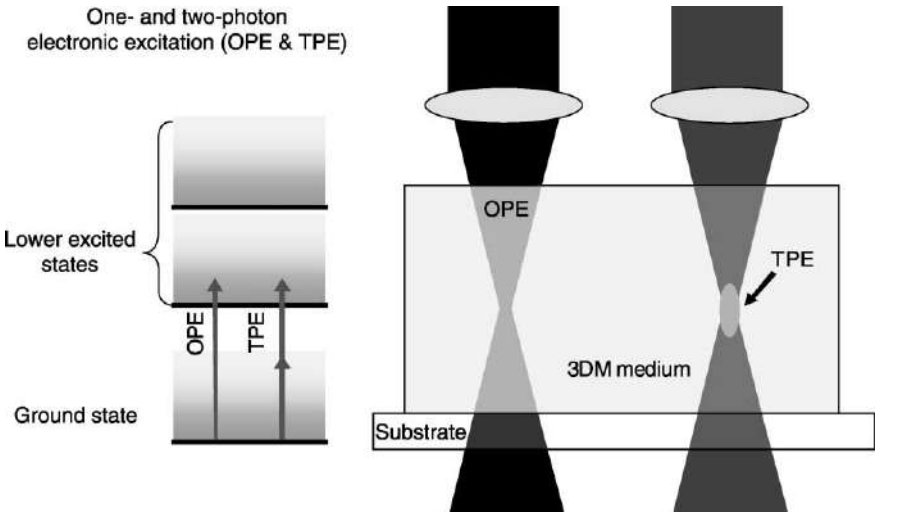

Figure 3 (Left) Jablonski representation of electronic one- and two-photon excitation (OPE and TPE, respectively). (Right) Illustration of a 3DM medium undergoing OPE and TPE using focused short- (black) and long-wavelength (dark gray) radiation, respectively. The regions that are appreciably excited are shown in light gray. Under OPE the medium is excited throughout the length of the interaction volume. Under TPE the excitation is confined in three dimensions to a small region around the focal point.

molecular parameter that is independent of the intensity. The molecular cross-sections for OPE and TPE, $\sigma^{(1)}$ and $\sigma^{(2)}$, are respectively equivalent to the parameters σ and δ that are commonly encountered in the literature.

For MPE to occur, the intensity must be sufficiently large that there is a high probability of more than one photon arriving simultaneously at the species to be excited. Typical magnitudes of the OPE and TPE cross-sections are $\sigma^{(1)} \approx 10^{-17}\,\mathrm{cm}^2$ and $\sigma^{(2)} \leq 10^{-47}\,\mathrm{cm}^4\,\mathrm{s\,photon}^{-1}$, and under conventional excitation conditions, $R_{2h\nu} \ll R_{1h\nu}$. To achieve $R2h\nu \approx R1h\nu$, $I \approx 2h\nu\sigma^{(1)}/\sigma^{(2)}$ is required. For the cross-sections given above, this condition corresponds to $I \approx 500\,\mathrm{GW\,cm}^{-2}$ at $\lambda = 800\,\mathrm{nm}$. Such peak powers can be achieved by focusing ultra-short laser pulses, so mode-locked femtosecond- or picosecond-pulse lasers are generally used for 3DM. Tight focusing in 3DM is desirable then, not only to provide a means for 3D confinement, but also to increase the intensity to a level that $R_{nh\nu}$ becomes large enough to produce a significant material change.

Opto-Mechanical Implementation

One method for implementing single-beam 3DM is illustrated in Figure 4. An excitation laser beam is directed onto a high numerical aperture (NA) objective. A beamsplitter, detector, attenuator, and shutter are placed upstream of the objective to measure and control the average power at the sample, $\langle P \rangle$. A substrate is coated with the photo-active medium and attached to a 3-axis nano/micro-positioner under the objective. The interior of the photo-active material is patterned during the exposure by translating the sample relative to the focus of the laser beam and shuttering the beam as needed. The process is automated as a computer controls the sample translation coordinates and the exposure conditions for each volume element (voxel). This configuration is well suited for patterning solid or semi-solid photo-media. An index-matching fluid must be used with high-NA objectives to achieve the smallest focal spot size. If the photo-medium is a liquid, a coverslip or other thin transparent barrier must be placed between the index matching fluid and the photomedium so that the latter does not become contaminated.

Alternatively, an inverted configuration may be used. Here, the laser is focused through the substrate into the medium. The objective and the index matching fluid are kept out of contact with the medium by the substrate. In this case, the maximum depth that can be patterned is limited by at least two conditions: (i) The working distance of high-NA objectives is small, often no more than 200 μm, so the substrate must be kept thin (~100 μm); and (ii) unless the viscosity of the medium is very high, patterning must begin at the medium/substrate interface and proceed layer by layer away from it, so that the structure remains anchored to the substrate. This could be disadvantageous, as the change in refractive index in the previously patterned layers could alter the intensity distribution

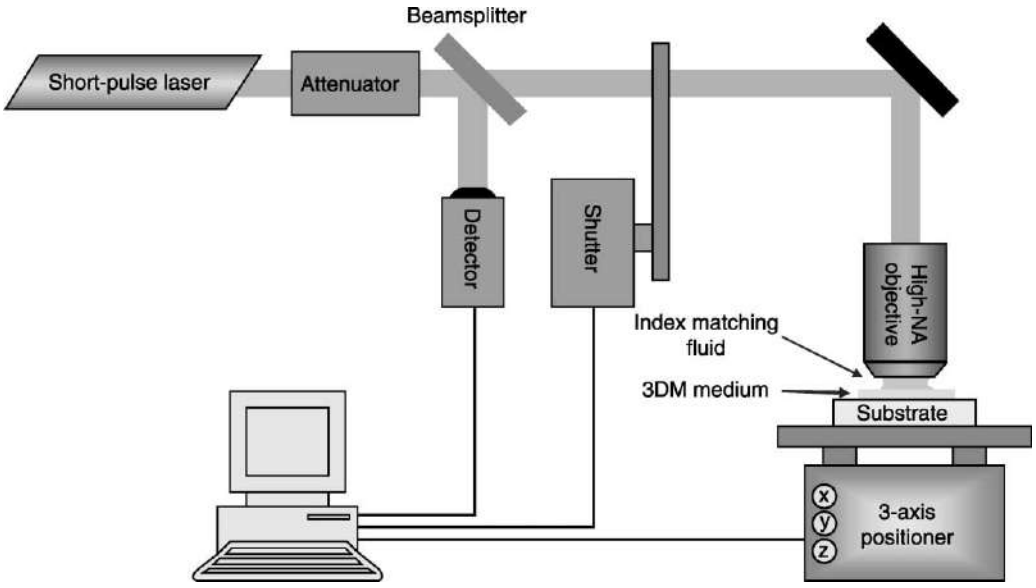

Figure 4 Optomechanical setup for single-beam 3DM.

at the focus. In all configurations, it is best to minimize the number of optical interfaces to maintain the quality of the focus.

A scanning confocal microscope system provides a convenient platform for single-beam 3DM. The nano/micro-positioner can be mounted directly on the sample stage of the microscope, and the optical components of the confocal system can be used directly for 3DM. The confocal system also enables the medium and the structure to be imaged *in situ*, using confocal reflectance or fluorescence imaging. If the material is fluorescent under MPE, a 3D image can be obtained by scanning multiphoton fluorescence microscopy. These imaging modes facilitate 3DM by providing a means for locating the focus at an absolute position within the photo-active medium.

3DM can also be realized by moving the focused beam relative to a fixed sample. One convenient implementation utilizes galvanometer-driven mirrors to scan the beam in planes parallel to the substrate (xy-plane). Motion in the z-direction is accomplished by raising and lowering the objective relative to the substrate. A galvo-based system typically enables faster fabrication, but may not afford motion as precise as that obtained using a multi-axis nano-positioner.

Excitation Sources

The most commonly used and arguably the most convenient and reliable laser for 3DM is the continuous-wave (CW) mode-locked titanium sapphire (Ti:S) laser. Ti:S lasers routinely produce sub-100 fs pulses at high repetition rates (~80 MHz) with single-pulse energies, E_p, of ~1 nJ. The Ti:S laser emission spans the range of 700–1050 nm, making it particularly well suited for two- and three-photon excitation. $\langle P \rangle$ can be as high as several watts at the peak of the gain curve ($\lambda \approx 800$ nm) with high pulse-to-pulse and time-averaged stability (RMS $\approx 1\%$). Such systems may also produce a nearly Gaussian TEM_{00} mode, which is helpful in achieving the tightest focus in the material. Note that focusing a 1 nJ, 100 fs pulse to a 1 μm-radius spot generates a focused average peak intensity of 320 GW cm^{-2}, as is needed to achieve photo-excitation rates of $R_{2h\nu} \approx R_{1h\nu}$.

For higher peak powers, amplified femtosecond (AFS) lasers can be used for 3DM. These systems can produce ~100 fs pulses with $E_p \approx 1$ mJ or higher at repetition rates of 1–100 kHz. Given that a continuous microfabricated structure should be generated from partly overlapping voxels, a low repetition rate reduces the speed with which the laser beam may be scanned within the material. An AFS laser can be used to pump an optical parametric generator to extend the output wavelength range.

Recently, some researchers have turned their attention toward identifying compact solid-state turn-key lasers for 3DM that could be more convenient and cheaper alternatives to the complex systems described above. A commercial Nd:YAG microlaser ($\lambda = 1064$ nm, $E_p = 5.4$ μJ, $\tau_p = 640$ ps, 13.6 kHz) and a mode-locked erbium-doped fiber laser ($\lambda = 780$ nm, $E_p \approx 0.5$ nJ, $\tau_p = 100$ fs, 50 MHz)

have both been shown to be satisfactory for 3DM in certain acrylate-based media.

Material Systems

Structural Materials

The most widely used materials for 3DM are based on acrylates that are patterned by multiphoton induced polymerization. The process begins when an initiator chromophore **C** is excited by n-photon absorption to $\mathbf{C}^*$, with an efficiency specified by the n-photon cross-section $\sigma^{(n)}$. $\mathbf{C}^*$ then generates an initiating radical **I·**. Acrylate monomers (**M**) can react with **I·** and undergo a free-radical chain reaction that leads to high molecular weight polymer. With sufficient exposure, the starting material is converted to less soluble high-molecular weight polymer chains or densely cross-linked polymer networks. The photo-patterned structure can then be 'developed' by immersing the sample into a solvent that removes the unexposed material, leaving behind a free-standing structure that is a replica of the photopattern.

Photo-polymerization starts when the MPE exposure generates enough **I·** to react with and consume all of the inhibitor in the local volume at point $\vec{r}$ within a homogeneous medium. In this discussion we assume that the local concentration of inhibitors is not changed by diffusion on the exposure timescale. The total number of **I·** generated during an exposure at $\vec{r}$ per unit volume, $N_{\mathrm{I}\cdot}(\vec{r},t)$, is obtained from the time-integrated MPE rate:

$$N_{\mathrm{I}\cdot}(\vec{r},t) = \frac{\phi_{\mathrm{c}}^{(n)}\sigma^{(n)}N_{\mathrm{C}}}{n(h\nu)^n}\int_{-\infty}^{t} I(\vec{r},t')^n \mathrm{d}t' \qquad [2]$$

Here, N_{C} is the number density of **C**, and $\phi_{\mathrm{c}}^{(n)}$ is the n-photon chemical quantum yield, defined as the number of **I·** generated divided by the number of $\mathbf{C}^*$ present after excitation. If we identify the initial number density of the inhibitors as N_{Inh}, then polymerization occurs when $N_{\mathrm{I}\cdot}(\vec{r},t) > N_{\mathrm{Inh}}$. In multiphoton-induced polymerization, $N_{\mathrm{I}\cdot}(\vec{r},t)$ is a nonlinear function of I, which generally involves a rapidly varying time envelope, as from a mode-locked ultrashort-pulse laser.

Several different commercial acrylate systems have been used successfully for 3DM, such as Nopcocure 800 (San Nopco), NOA 72 (urethane acrylate oligomer, Norland Products), and SCR 500 (blend of urethane acrylate monomers and oligomers, Japan Synthetic Rubber Co.). Researchers have also successfully developed custom formulations using blends of commercially available monomers, oligomers, and polymer additives to create resins with tailored physical, chemical, and optical properties. This aspect of acrylates, and more generally polymer-based systems, makes them one of the most attractive classes of materials for 3DM. Conventional one-photon UV- or blue-sensitive radical generators (e.g., benzil, benzoin methyl ether, 4,4′-bis(N,N'-dimethylamino)benzophenone, to name just a few) have most often been used, and these were activated by promoting the molecule into one of its low-lying electronic states by MPE over the range of 730–800 nm.

3DM based on epoxide polymerization has also been demonstrated using commercial resins including cross-linkable small molecular weight monomers (e.g., 4-vinyl-1-cyclohexene diepoxide) and epoxide oligomers (e.g., SU-8, first developed by IBM). Epoxide polymerization is most commonly initiated by molecules that, after excitation, generate a Brønsted acid, H^+. The propagating species is a carbocation, and bases present in the medium act as inhibitors. Relative to acrylates, epoxides shrink substantially less upon polymerization, leading to less distortion of the microstructure after development.

Acrylate chemistry has also been exploited for 3DM of functional composite materials. Acrylamide resin mixtures were used to fabricate bi-layer micro-cantilevers that deflect from the surface when illuminated with UV light. This type of microstructure could form the basis of a photo-activatable MEMS. 3D gratings were fabricated from an acrylate pre-polymer containing a dendrimer functionalized with cross-linkable acrylic L-phenylalanine groups. The dendrimer acts as a host encapsulation site that enables a laser dye to be loaded in the material at high weight-percent (see below). A variety of complex micro-structures have also been fabricated using 'Ormocer-I', an ORganically MOdified CERamic. Ormocer-I is an inorganic-organic hybrid polymer consisting of a Si–O–Si (siloxane) backbone that is functionalized with organic moieties, including cross-linkable methacryloxypropyl groups. Ormocers are extremely promising for 3DM because of the thermal stability and chemical inertness of the backbone and their attractive physical and optical properties.

3DM could be a powerful tool for fabricating biologically active structures and devices. In this context, hydrogels and Ormocers should be useful material systems as some formulations are known to be bio-compatible. Microstructures can also be fabricated from proteinaceous material. 3D microstructures were produced by patterned MPE of

solutions containing the dye rose bengal and the proteins bovine serum albumin or fibrinogen.

The material systems discussed thus far are 'negative-tone' media. This means that only the exposed regions of the material remain after the post-exposure development, so the final structure is a replica of the exposure pattern. In contrast, 'positive-tone' media are solid-state systems for which the exposed regions become soluble in a developer. The final structure is then the inverse of

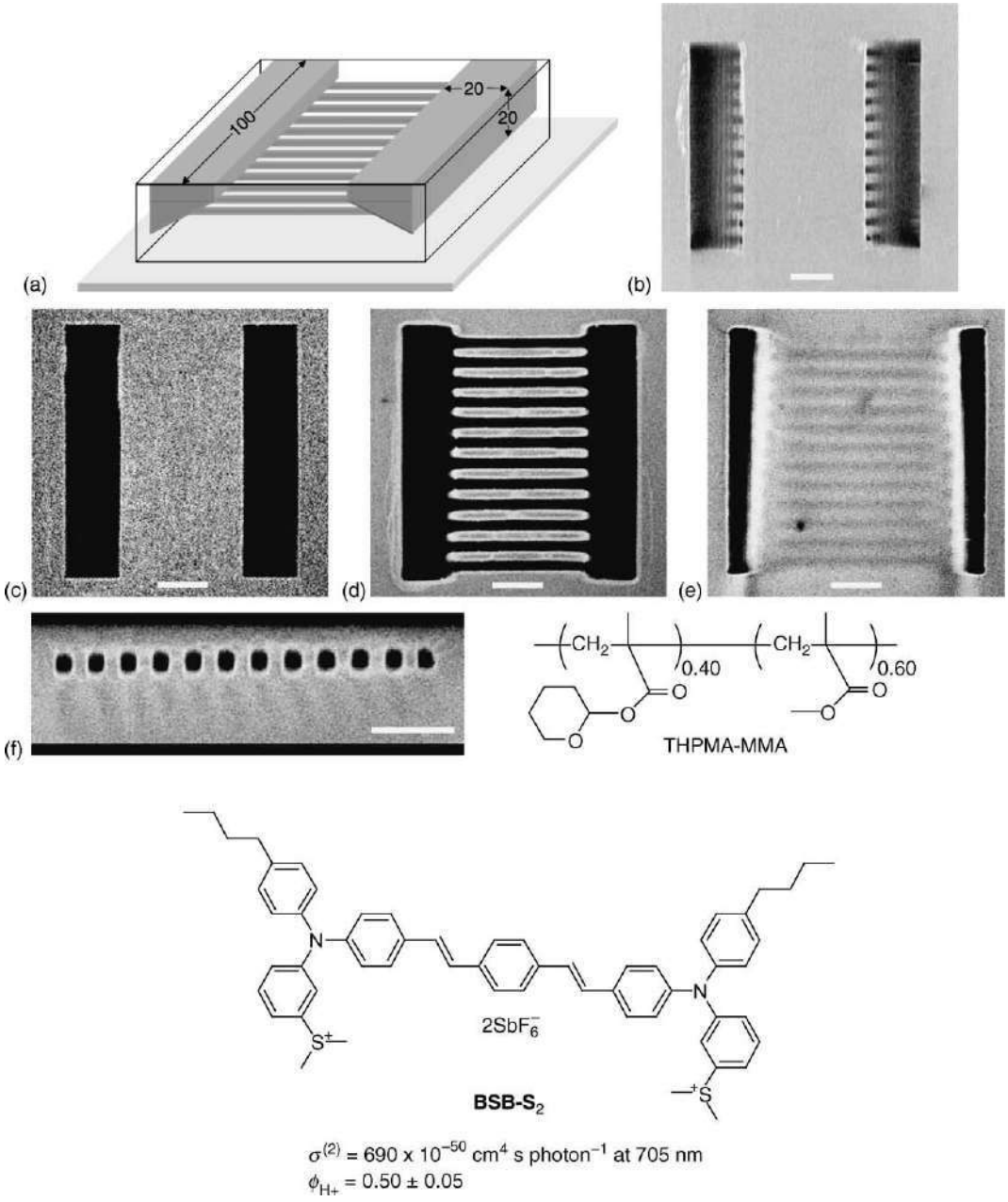

Figure 5 A 3D micro-channel structure fabricated by TPE-3DM using the positive-tone chemically amplified resist THPMA-MMA containing the two-photon PAG **BSB-S$_2$**. A 50 μm-thick film of the resist was exposed in the pattern of the target structure at 745 nm with tightly focused 80 fs pulses at $\langle P \rangle = 40$ μW and a linear scan speed of 50 μm s^{-1}. (a) Target structure consisting of two rectangular cavities (width: 100 μm, length: 20 μm, depth: 20 μm) with a sloped side-wall, that are connected by 12 channels (length: 50 μm; 4 μm by 4 μm cross-section) lying 10 μm below the surface and spaced apart by 8 μm (center-to-center). (b) Scanning electron micrograph of the final structure, viewed normal to the substrate. (c to e) Two-photon-fluorescence images of the final structure (viewed normal to the substrate): (c) at the surface of the film; (d) 10 μm below the surface; (e) 19 μm below the surface. (f) Two-photon-fluorescence cross-sectional image of the buried channels. The scale bar in b to f corresponds to 20 μm. Reproduced from Zhou W, *et al.* (2001) *Science* 296: 1106–1109. Copyright (2001), with permission of The American Association for the Advancement of Science.

the exposure pattern. Positive-tone 3DM was first demonstrated as early as 1992 with the generation of high aspect ratio trenches in thin layers of commercial photoresists. A positive-tone material system was designed for 3DM (Figure 5), is based on a chemically amplified resist and a high-sensitivity two-photon-activatable photo-acid generator (PAG) (discussed below). Using this medium, it was possible to fabricate subsurface microchannel structures and microgratings, using moderate laser scan speeds (50 μm s^{-1}) and exceptionally low $\langle P \rangle$ (40 μW).

Composites have been formulated that enable 3D-patterning of nanoscale metal features within a host matrix. A two-step process has been described that involves MPE of a Ag^+-containing sol–gel followed by thermal development of the latent image with a $AgClO_4$ solution. Using this system, 3D spirals of silver nano-particles embedded within the sol–gel matrix were generated. A $HAuCl_4$-doped poly(vinyl alcohol) composite was used to 3D-pattern gold nano-particles. It was shown that after TPE a particular sensitizer dye hosted in a poly(vinylcarbazole) matrix containing $AgBF_4$ and ligand-stabilized Ag nanoparticles, electron transfer from the dye to Ag^+ leads to nucleation of Ag^0 at the nanoparticles and the formation of conducting bulk metal features (Figure 6). Material systems like these should find wide application for generating micro- and nano-scale structures, having unusual and useful optical and electronic properties.

3DM in glasses is of great interest for photonic applications. Several groups have demonstrated that tightly focused infrared or visible amplified ultrashort laser pulses can be used to 3D-pattern pure and doped silica and other glasses. The patterned voxels can be generated as a result of photo-induced changes in the refractive index or extreme dielectric breakdown, resulting in pits within the material. The mechanisms for these two cases are not fully understood, but both are thought to be activated through an initial MPE event. The leading edge of the laser pulse generates a free-carrier plasma through MPE. The remainder of the pulse is increasingly absorbed by the plasma leading to avalanche ionization. The optical energy deposited into the local lattice results in cavitation, creating a pit at the focal point. Based on the electronic bandgap of silica, the process likely involves three- and six-photon absorption when 400- and 800-nm light is used, respectively. This approach has been used for the 3D-patterning fused silica and germanium doped glasses.

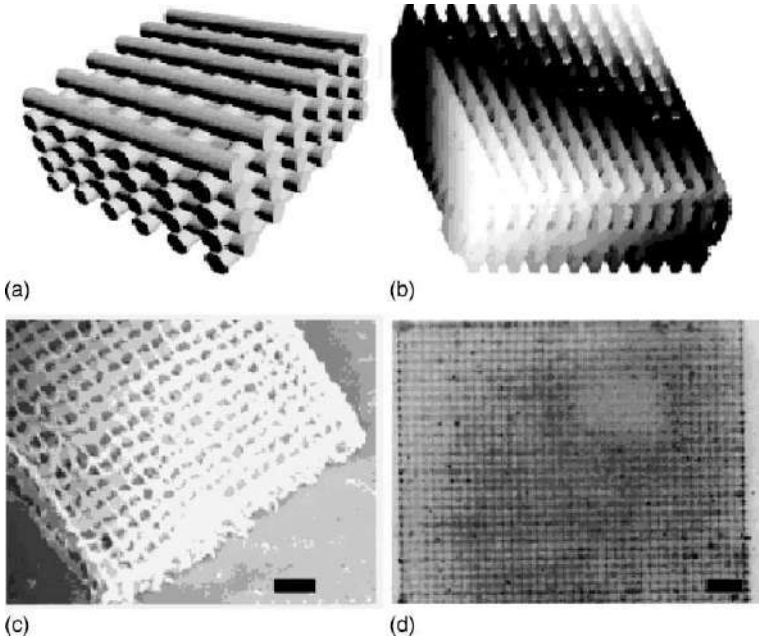

Figure 6 Metallic structures fabricated by TPE-3DM using a silver nanoparticle/polymer composite. (a) Model of the 'stack-of-logs' target structure. (b) 3D image constructed from a series of two-photon fluorescence microscopy images of the silver structure immediately following 3DM. (c) Scanning electron micrograph of the free-standing silver structure following removal of the surrounding unexposed composite. The scale bar corresponds to 10 μm. (d) Transmission optical microscopy image of the silver structure immediately following 3DM. The scale bar corresponds to 25 μm. Reproduced from Stellacci F, *et al.* (2001) *Advanced Materials* 14: 194–198. Copyright (2001), with permission of Wiley.

Multiphoton Initiators

For 3DM to develop into a widely used fabrication tool, material systems and patterning methods must be developed that enable rapid microfabrication at low cost. 'Soft materials', such as the polymer systems described above, are promising in this respect, as both the initiating species and the host resin can be engineered to optimize the multiphoton response and the physical and chemical properties of the patterned medium. It may also prove possible to create doped glasses that have physical properties close to that of neat silica and yet offer higher multiphoton sensitivity.

To develop this concept, let us consider 3DM in acrylates further. A high-sensitivity acrylate medium is one for which: (i) **C** strongly absorbs n-photons (large $\sigma^{(n)}$); (ii) $\mathbf{C}^*$ efficiently generates the initiating species **I·** (large $\phi_c^{(n)}$); and (iii) **I·** is an efficient initiator (high initiation rate). As criteria (i) and (ii) are predominantly determined by the molecular properties of **C**, it should be possible to engineer soft materials for 3DM by optimizing multiphoton initiators and incorporating these into pre-polymer resins that once patterned have the correct properties for the target application. The sensitivity of materials can be compared on the basis of the minimum average laser power, $\langle P\rangle_{\min}$, that is required to pattern a material using a fixed scan speed (or exposure time). Alternatively, we can use the product $\sigma(n)\phi c(n)$ as a multiphoton initiator figure-of-merit (FOM) that accounts for the strength of the multiphoton absorption and the efficiency with which **I·** is generated.

The chromophores that have been used most frequently for polymerization-based 3DM do efficiently generate highly active initiating species once excited; however, they were developed for one-photon response to UV and blue radiation and generally have low multiphoton FOMs. The use of xanthene dyes was explored as initiators for free-radical polymerization of acrylamides and it was found that the multiphoton sensitivity was low. As a result, 3DM using conventional initiators requires high laser power (typically $\langle P\rangle_{\min} \approx 10-100$ mW) and slow laser scan speeds ($\sim 50\ \mu\text{m s}^{-1}$). Under tight-focusing conditions, the resulting peak powers are often near the damage threshold of the materials, rendering the process unreliable from a manufacturing stand-point.

A series of **D–π–D** molecules was examined (Figure 7), where **D** is an electron-donating group and **π** is a conjugated π-electron bridge. This class of **D–π–D** chromophores was found to have large TPE cross-sections ($\sigma^{(2)} > 200 \times 10^{-50}\ \text{cm}^4\ \text{s photon}^{-1}$) and to initiate the polymerization of acrylates following TPE in the visible and near infrared, with $\langle P\rangle_{\min}$ a factor of 30–50 times lower than for conventional one-photon initiators. The sensitivity achieved with **D–π–D** initiators is so enhanced that TPIP-3DM was performed at laser scan-speeds as high as $9\ \text{cm s}^{-1}$, using $\langle P\rangle = 3$ mW, enabling a volume of $180\ \mu\text{m} \times 180\ \mu\text{m} \times 20\ \mu\text{m}$ to be polymerized in 200 s.

D–π–A molecules have also been examined for TPIP (Figure 7), where **A** is an electron accepting functionality. Such molecules can have large TPE cross-sections, and some have been shown to activate TPIP efficiently in acrylates at $\lambda \approx 800$ nm. A derivative of Michler's ketone was synthesized with an extended π-conjugated system (Figure 7). This molecule exhibited strong two-photon absorption over the broad range of 800–1100 nm and was also shown to be active toward acrylate polymerization at 1064 nm. A strategy was implemented for enhancing two-photon sensitivity based on up-converted fluorescence and used for 3DM of waveguide structures. The strongly two-photon-absorbing dye AF-183 was incorporated into a commercial acrylate photoresist containing a radical-generator that is activated by OPE at 315–450 nm. After AF-183 undergoes TPE at 800 nm, it fluoresces efficiently at ~465 nm. The up-converted emission then excites the one-photon initiator and radicals are generated.

The strategies discussed above are oriented toward enhancing multiphoton sensitivity by increasing $\sigma^{(2)}$ of the initiating chromophore. Little attention has been paid to the need of ensuring that the MPE couples to a mechanism for efficiently generating an active initiating species. This problem was addressed with the design of a two-photon PAG by covalently linking a **D–π–D** chromophore and sulfonium moieties, which can generate H^+ when reduced by photo-induced electron transfer (**BSB-S$_2$** in Figure 5). **BSB-S$_2$** features strong two-photon absorption and its quantum yield for the generation of H^+ per photo-excited molecule (ϕ_{H^+}) exceeds that of some commercial one-photon PAGs. Based on $\langle P\rangle_{\min}$, **BSB-S$_2$** was found to be 10 to 100 times more sensitive toward multiphoton epoxide polymerization than one-photon UV/blue-sensitive PAGs.

It should be understood that this discussion of multiphoton initiators is not comprehensive of all work that has been done in the field. Rather, it is intended as a highlight of the many outstanding examples of progress in this area. The use of inexpensive low-power lasers for 3DM should become more practical as high-sensitivity

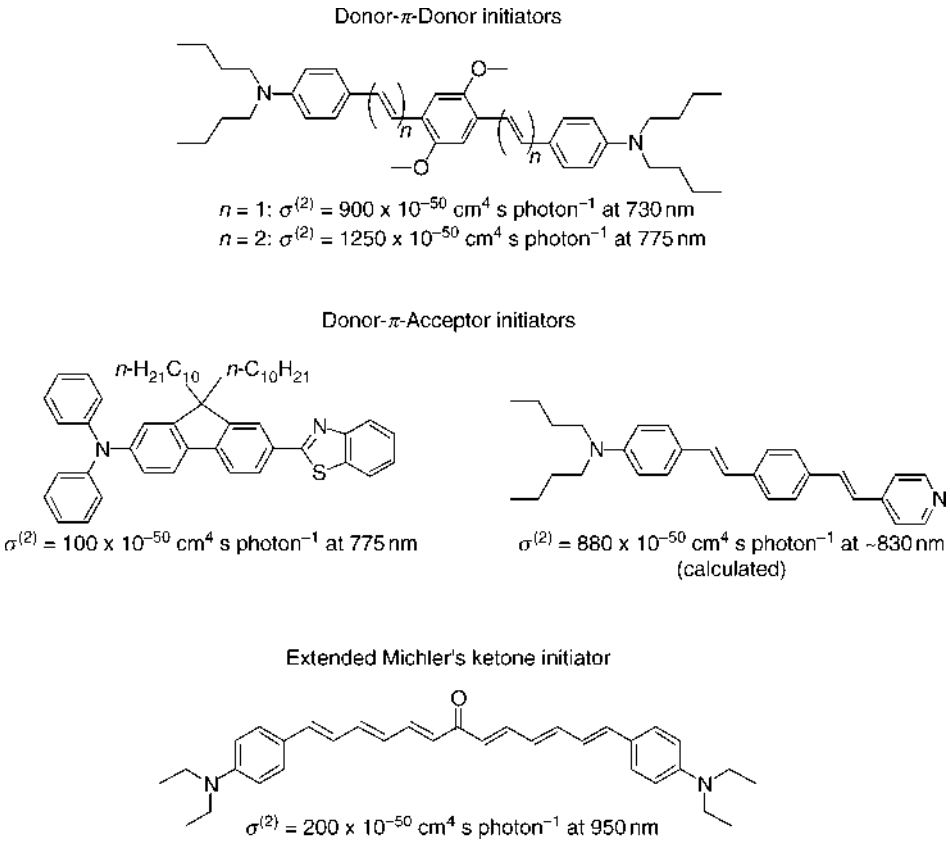

Figure 7 Examples of high-sensitivity initiators for 3DM based on TPIP of acrylates.

3DM media are developed. This should, in turn, facilitate the widespread use of 3DM as a microfabrication tool.

Resolution

The resolution of single-beam 3DM is determined by the size of the smallest voxel that can be reliably produced. Both the optical parameters and the material response under MPE affect the size of the voxel. Given the wide range of physical and chemical mechanisms that can be involved, no single set of guidelines can be outlined regarding the effect of the material response on voxel size. Additionally, this aspect of 3DM remains relatively unexplored. However, as 3DM based on TPIP of acrylates has been examined in some detail, an overview of this material response provides some valuable insight into the overall problem of resolution. The following simplified discussion is based on the local response of the material to photo-excitation.

TPIP initiated in the vicinity of the focus could, in principle, propagate throughout the entire volume of the medium, were it not for termination by inhibitors. Let us consider further the expression for $N_{\mathrm{I}\cdot}(\vec{r},t)$, expanding the intensity as:

$$I(z,r,t') = i(z,r)f(t') \qquad [3]$$

Here, $f(t')$ is a time-dependent amplitude function, $i(z,r)$ is the peak intensity, and the position vector $\vec{r}$ is expressed in cylindrical coordinates (z,r) with the origin at the focus. As before, z denotes the direction in which the beam propagates (longitudinal). The radial coordinate orthogonal to the z-axis (lateral) is denoted by r. The beam is taken to have cylindrical symmetry, so I does not depend on the angular coordinate. For a fixed exposure time $t \rightarrow \tau_{\mathrm{ex}}$, we can identify a polymerization-threshold peak intensity, I_{th}, for which $i(z,r) > I_{\mathrm{th}}$ results in $N_{\mathrm{I}\cdot}(z,r,\tau_{\mathrm{ex}}) > N_{\mathrm{Inh}}$ and polymerization occurs. The polymerized voxel consists then of all points within the volume centered at the focal point and bounded by the I_{th} isophote.

The shape of the I_{th} isophote depends upon the optical parameters of the system. The tightest focusing is achieved when the highest NA objective

is used to focus parallel incident rays that overfill the entrance aperture. This situation generates a diffraction limited intensity pattern about the focal point. Within the focal plane the lateral intensity distribution follows the Airy pattern:

$$i(0, r) = [2J_1(\rho)/\rho]^2 I_0 \qquad [4]$$

where J_1 is the first-order Bessel function:

$$\rho = 2\pi r \mathrm{NA}/\lambda \qquad [5]$$

and I_0 is the peak intensity at the origin. Near the focus, the longitudinal intensity distribution is given by

$$i(z, 0) = [\sin(\zeta/4)/(\zeta/4)]^2 I_0 \qquad [6]$$

where

$$\zeta = 2\pi \mathrm{NA}^2 z/(n_0 \lambda) \qquad [7]$$

Let us define the longitudinal and lateral widths of the focal spot, Z and R respectively, to be twice the distance along the corresponding z- and r-axes over which the peak intensity of the central bright fringe decreases to $I_0/2$. If $\langle P \rangle$ is adjusted so that $I_0/2 = I_{\mathrm{th}}$, then to first order we can expect the polymerization voxel to be a solid of rotation centered about the focus having a longitudinal and lateral width of Z and R, respectively. Numerical evaluation shows that the $I_0/2$-points occur at $\rho = \pm 1.6154$ and $\zeta = \pm 5.5661$. For $\lambda = 800$ nm, NA $= 1.4$, and $n_0 \approx 1.5$, we find $R = 0.29$ μm and $Z = 1.1$ μm.

Several groups have demonstrated that voxels generated under conditions similar to those discussed above do indeed have dimensions comparable to those calculated, based on diffraction limited propagation. As shown in Figure 8, voxels with lateral and longitudinal widths below the diffraction limit can be fabricated by steadily reducing τ_{ex} (or I_0) so that an increasingly smaller portion of the central bright fringe exceeds I_{th}. In this way, voxels having $R \approx 100$ nm and $Z < 800$ nm have been generated, which convincingly demonstrates the potential of this technique as a nano-fabrication tool.

Post-exposure material processing can also have a profound impact on the resolution. In the case of TPIP-3DM, material near the boundary of the I_{th} isophote may not have sufficient molecular weight or cross-linking density to resist removal in a development step. In this case, the final voxel will be smaller than the I_{th} isophote defined strictly in terms of a polymerization threshold. Additionally, a solvent-based developer can swell or otherwise distort the voxels. Overall, three important factors determine the final size of the voxel: (i) the exposure and focusing conditions; (ii) the response of the material to photo-excitation; and (iii) the behavior of the photoconverted material during the development process. It should be emphasized that 3D confinement is due to the combined effects of nonlinear photo-excitation and the nonlinearity inherent in the photoconversion and post-exposure development of the material. More advanced treatments have also attempted to address the effects of radical and inhibitor diffusion on the size of the voxel.

Previously, we assumed that the voxel is cylindrically symmetric. The intriguing result of some recent work has shown that the widths of the voxel in the transverse directions can actually differ by as much as 10% when linearly polarized radiation is used because the symmetry of the electric field vectors in the x- and y-directions is broken in high-NA focusing geometries. Cylindrically symmetric voxels can be obtained by using depolarized or circularly polarized radiation. Clearly, study of the combined effects of optical parameters and material response remains an important area for further research in this rapidly advancing field. Although this discussion has centered on TPIP, similar arguments can be developed for 3DM based on other photomaterials and higher-order processes.

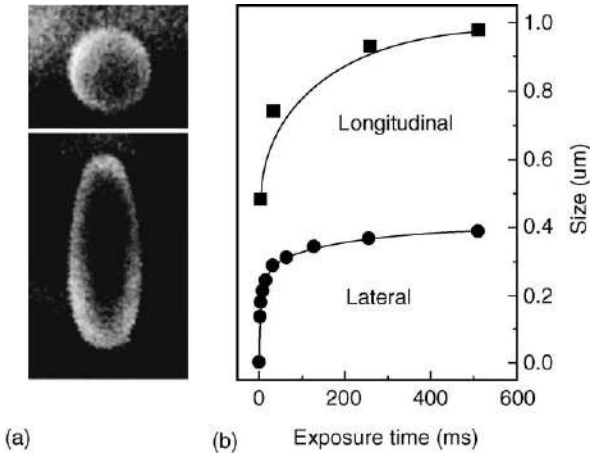

Figure 8 Single voxels produced by TPIP-3DM in an acrylate medium. (a) Scanning electron microscope (SEM) images of a single voxel showing its longitudinal (top) and lateral (bottom) width. (b) Dependence of the longitudinal and lateral widths of voxels on exposure time, as measured from SEM images following fabrication. Reproduced from Kawata S and Sun H-B (2003) Two-photon photopolymerization as a tool for making micro-devices. *Applied Surface Science* 208–209: 153–158. Copyright (2003), with permission of The American Institute of Physics.

Multibeam-Interference Three-Dimensional Microfabrication

Multibeam-interference (MBI) 3DM is an alternative approach to 3DM that enables large volumes of material to be simultaneously patterned into a 3D periodic structure. In MBI-3DM, multiple coherent laser beams interfere within a photo-active medium. A periodic structure results because the material undergoes photo-induced changes that occur maximally in the regions of peak intensity across the periodic interference pattern. Depending on the number of beams utilized and their reciprocal orientation, periodic structures in one, two, or three dimensions can be generated, and the periodicity is of the order of λ. This technique is often referred to as holographic lithography or recording.

MBI-3DM has most often been achieved using one-photon-induced polymerization; however, higher-order excitation may be used as well. In this discussion we will focus on multiphoton MBI-3DM, with the understanding that most of the concepts apply equally well to the case of linear photo-excitation. When the power of the incident beams is adjusted so that the intensity at the peaks of the interference pattern just exceeds I_{th}, polymerization only occurs in the local vicinity of these points, whereas the 3DM medium is not appreciably polymerized in regions where the intensity is below I_{th} and may be removed by dissolution in a solvent. The result is a periodic porous structure with highly interconnected parts, that extends over the whole volume where the beams overlapped.

Let us assume that m-beams are directed onto a photosensitive material. The intensity of light at a given point $\vec{r}$ in the region where the beams overlap is proportional to:

$$\begin{aligned} I(\vec{r}) \propto \vec{E}\cdot\vec{E}^* &= \left(\sum_{j=1}^{m} \vec{E}_{0j}\, \mathrm{e}^{\mathrm{i}(\vec{K}_j\cdot\vec{r}-\omega t)}\right)\cdot\left(\sum_{l=1}^{m} \vec{E}^*_{0l}\, \mathrm{e}^{-\mathrm{i}(\vec{K}_l\cdot\vec{r}-\omega t)}\right) \\ &\propto \sum_{j,l=1}^{m} \vec{E}_{0j}\cdot\vec{E}^*_{0l}\, \mathrm{e}^{\mathrm{i}(\vec{K}_j-\vec{K}_l)\cdot\vec{r}} \\ &\propto \sum_{j=1}^{m} \vec{E}_{0j}\cdot\vec{E}^*_{0j} + \sum_{j\neq l}^{m} \vec{E}_{0j}\cdot\vec{E}^*_{0l}\, \mathrm{e}^{\mathrm{i}(\vec{K}_j-\vec{K}_l)\cdot\vec{r}} \end{aligned} \qquad [8]$$

where $\vec{E}_{0j}$ and $\vec{K}_j$ are the electric field and wavevector of beam j, and ω is its angular frequency ($\omega = 2\pi\nu$). Equation [8] indicates that the intensity distribution depends on the difference between pairs of wavevectors of the interfering beams, $\vec{K}_j - \vec{K}_l$. This distribution is periodic in space and can be described as a crystal lattice with reciprocal lattice vectors equal to $\vec{K}_j - \vec{K}_l$ and with a unit cell structure determined by the pre-factors $\vec{E}_{0j}\cdot\vec{E}^*_{0l}$. By choosing the intensity of the incident beams and the exposure time appropriately, it is thus possible to generate a pattern in the medium that has the same periodicity and structure as the I_{th} isophote of the interference pattern.

The simplest implementation of MBI-3DM involves interfering two linearly polarized beams at angles $\pm\theta/2$ with respect to the normal to the film being patterned (θ is the angle between the two beams in vacuum). From eqn [8], it can be shown that the intensity is maximum if $(\vec{K}_1 - \vec{K}_2)\cdot\vec{r} = 2q\pi$, where q is an integer, or $K[\sin(\theta/2)]r' = q\pi$, where K is the magnitude of both $\vec{K}_1$ and $\vec{K}_2$, and r' is the component of $\vec{r}$ along $(\vec{K}_1 - \vec{K}_2)$. Thus, the two beams generate an interference pattern with periodicity (in one dimension) given by $\Lambda = \lambda/[2\sin(\theta/2)]$, if λ is the

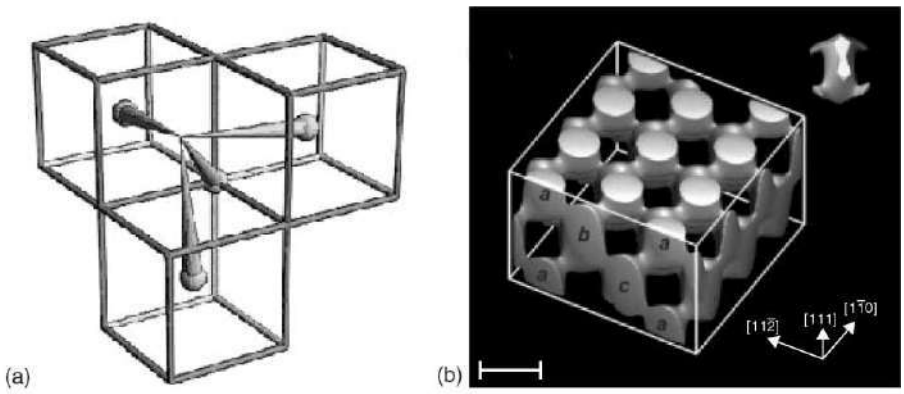

Figure 9 (a) Relative orientation of the beams in a four-beam interference scheme to obtain a face-centered cubic lattice. The beams have wave vectors along the directions [−3/2, −3/2, −3/2], [−5/2, −1/2, −1/2], [−1/2, −5/2, −1/2], and [−1/2, −1/2, −5/2]. (b) One of the isophotes generated by the interference of the four beams in (a). The inset represents the shape of the unit cell. Reproduced with permission from *Nature*, Campbell M, *et al.* (2000) *Nature* 404: 53–56. Copyright (2000), Macmillan Magazines Limited.

vacuum wavelength of the exciting beam ($\lambda = 2\pi/K$). When a film of a polymerizable material is placed at the intersection of the beams, a series of ridges is generated in the resist, spaced by Λ and with a width dependent on the beam intensity. For example, if a laser beam at 800 nm is split in two components forming an angle of 30°, $\Lambda = 1.5$ μm. The periodicity becomes smaller as the angle between the beams is increased.

By interfering three beams it is possible to obtain a two-dimensional periodic structure. For example, the interference of three beams forming an angle θ with respect to the medium normal and symmetrically arranged around it was used to generate an hexagonal lattice of rods that extends parallel to the normal. At least four beams must interfere to obtain a microstructure with 3D periodicity. The beam geometry shown in Figure 9a was used to produce a microstructure with the symmetry of a face-centered cubic lattice (Figure 9b). Beam geometries for generating interference patterns having other symmetries, such as body-centered cubic and body-centered tetragonal, have also been described.

An advantage of MBI-3DM is that it enables rapid patterning of large volumes during a single exposure. In fact, the polymerization takes place, in principle, throughout the region in which the beams overlap. When the process is activated by OPE, the beams can be attenuated significantly by absorption in the material. This can limit the thickness of the fabricated structure or result in spatial variation in the filling ratio. The use of TPIP can overcome this limitation and allow for thicker media to be patterned uniformly. Structures with the same symmetry are generated if either one- or two-photon processes are exploited to induce the material transformation. However, the structures will be different for at least two reasons: (i) if the same initiator is used in both cases, the wavelength used for TPE is approximately twice that for OPE, and

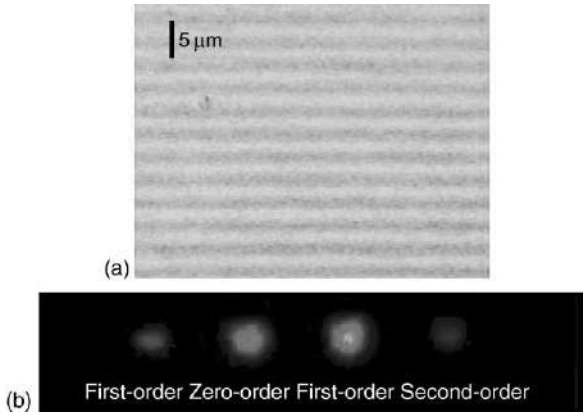

Figure 10 (a) Optical image of a diffraction grating obtained by TPIP using a two-beam interference scheme ($\lambda = 810$ nm). The spacing period is 3.4 μm and film thickness is 40 μm. The grating is made of an acrylate-based resin. (b) Diffraction pattern of the grating in (a) generated using a Helium–Neon laser (632.8 nm). The maximum diffraction efficiency was measured to be 57%. Reproduced from Guo H, *et al.* (2003) *Chemical Physics Letters* 374: 381–384. Copyright (2003), with permission from Elsevier.

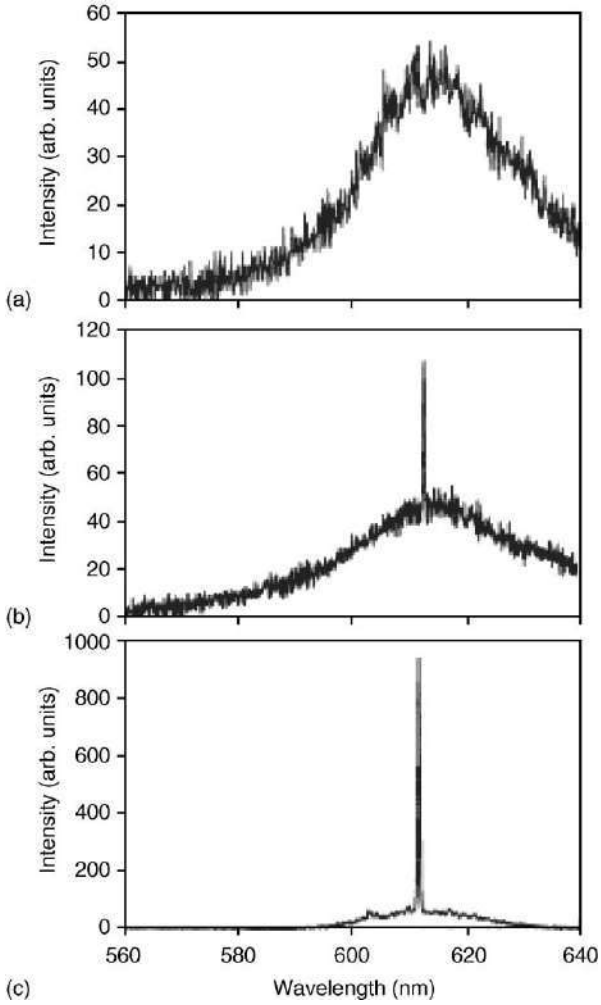

Figure 11 Emission spectra of the laser dye DCM encapsulated in a dendrimer-based resin that was patterned into a grating microstructure by single-beam 3DM. The grating was pumped at 532 nm (8 ns pulses) with excitation energies of (a) 0.03, (b) 0.05, and (c) 0.20 μJ. Starting at excitation energies as in (b), a narrow band at 612 nm appears over the spontaneous emission. The relative intensity of this lasing emission increases with excitation intensity with respect to the spontaneous emission. Reproduced from Yokoyama S, *et al.* (2003) *Thin Solid Films* 438–439: 452–456. Copyright (2003), with permission from Elsevier.

the periodicity of the structure is correspondingly different; and (ii) the I_{th} isophote, and thus the details of the polymerized structure inside each unit cell, is different under OPE and TPE because the excitation rate is proportional to I and I^2, respectively. In principle, better defined features can be obtained using two-photon or higher-order excitation because the transition between regions of low- and high-exposure is sharper than that under OPE, due to the nonlinear dependence on I. Interference fabrication techniques can enable the generation of features considerably smaller than the diffraction limit, when the beam geometry and excitation intensity are chosen appropriately.

The beams used to generate the interference pattern are usually obtained by splitting the output of a single laser source into the desired number of components. The intensity and polarization of each beam are then independently controlled. Temporal overlap of the beams at the sample can be ensured by introducing an optical delay into each of the beam paths. As fluctuations in the beam intensity can cause changes in the lattice unit cell or its filling ratio, the use of a laser with good stability is imperative in cases where exposure time extends over more than a single laser pulse.

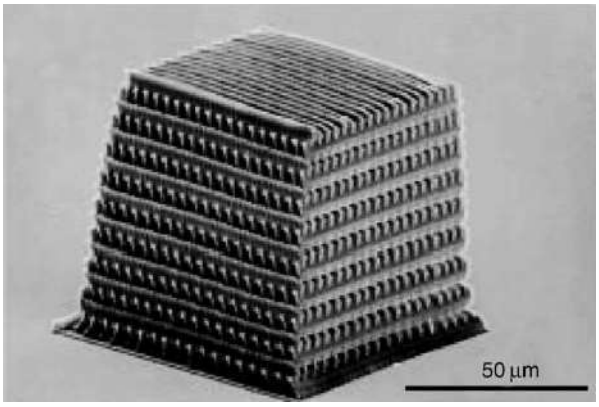

Figure 12 Example of a wood-pile PhC structure obtained by single-beam TPIP-3DM. The periodicity is 5 μm. The change in the cross-section of the structure from top to bottom, is due to shrinking of the material during development, whose effect is largest in the region farthest from the substrate. Reproduced with permission from *Nature*, Cumpston BH, Ananthavel SP, Barlow S, *et al.* (1999) Two-photon polymerization initiators for three-dimentional optical data storage and microfabrication. *Nature* 398: 51–54, Copyright (1999), Macmillan Magazines Limited.

Structures and Functional Devices

The versatility of 3DM is easily appreciated by surveying the variety of microstructures that have been fabricated over the years. These may be divided into two categories: (i) structures whose characteristics lay in their form and in the interconnectivity of their parts; and (ii) structures that incorporate an active functionality that can be exploited in a specific application. Some of these microstructures are described below.

Fabricating functional micro-optical devices is a promising application of 3DM. Figure 10 shows a grating fabricated by two-beam-interference TPIP in an acrylate with a period of 3.4 μm, whose diffraction efficiency is 28% as fabricated and 57% after annealing. The same method was used to generate a surface relief grating in a resin containing a polycationic peptide. In the presence of aqueous silicic acid, the peptide catalyzes the formation of ordered arrays of silica nanospheres in

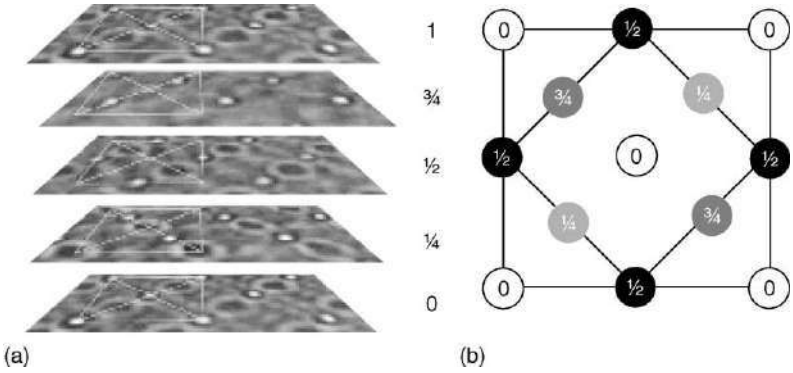

Figure 13 (a) Cross-sectional optical images of various planes of a diamond lattice fabricated in glass. The numbers to the right are the relative positions of the planes, in units of the lattice constant. (b) Diagram of the expected position of lattice points in a diamond structure. Points located in different planes have different shading. Reproduced from Mizeikis V, *et al.* (2001) *Journal of Photochemistry and Photobiology A* 145: 41–47. Copyright (2001), with permission from Elsevier.

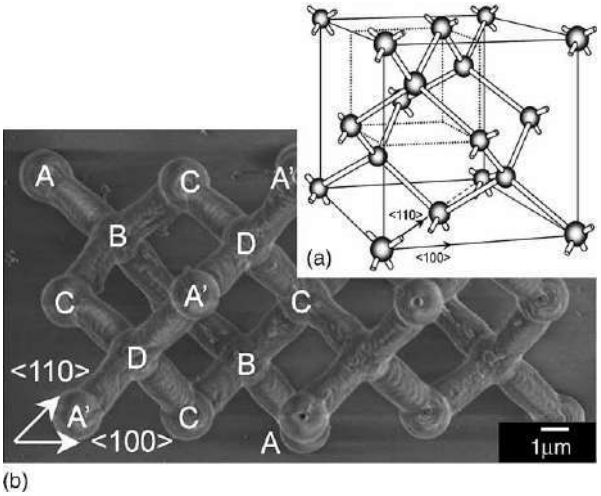

Figure 14 (a) Illustration of a diamond lattice. (b) Scanning electron micrograph of a polymeric microstructure with the symmetry of a diamond lattice, obtained by single-beam TPE-3DM. The letters A, B, C, D, A′ refer to the lattice planes at 0, 1/4, 1/2, 3/4, and 1 lattice units in the vertical direction. Reproduced from Kaneko K, *et al.* (2003) *Applied Physics Letters* 83: 2091–2093. Copyright (2003), with permission of The American Institute of Physics.

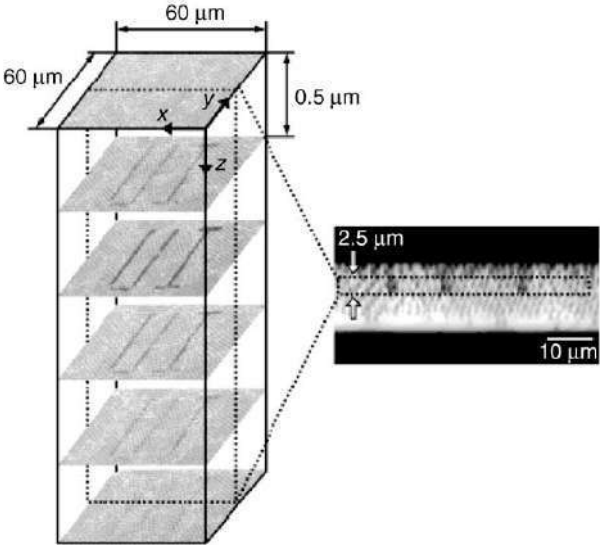

Figure 15 (Left) Images obtained by laser scanning confocal microscopy of linear defects generated by three-photon-induced polymerization of a triacrylate resin backfilled into a silica colloidal crystal. The spacing between the planes of the images is 0.5 μm. (Right) Cross-sectional image in the *xz*-plane of the same structure. Reproduced with permission from Lee W, *et al.* (2002) *Advanced Materials* 14: 271–274. Copyright (2002), with permission of Wiley.

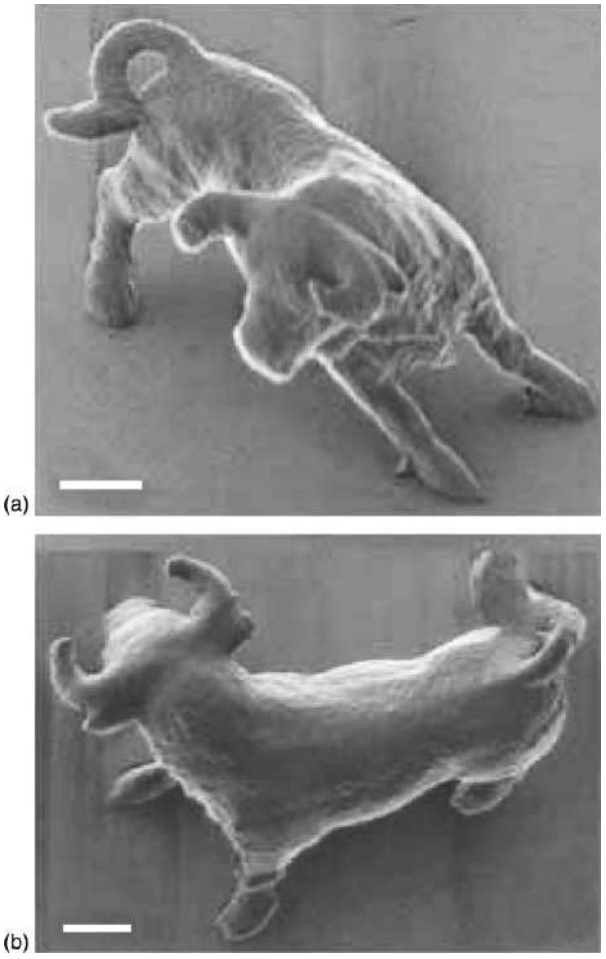

Figure 16 Microstructure of a bull fabricated by single-beam 3DM. The scale bar corresponds to 2 μm. Reproduced with permission from *Nature*, Kawata S, *et al.* (2001) *Nature* 412: 697–698. Copyright (2001), Macmillan Magazines Limited.

the troughs of the grating, which increases the diffraction efficiency 50-fold. 3DM was used to fabricate a grating structure in an acrylate resin containing the laser dye 4-(dicyanomethylene)-2-methyl-6-(4-dimethylaminostyryl)-4H-pyrane (DCM) encapsulated in a host dendrimer (see above) as an active optical medium. When pumped at 532 nm above a certain threshold, lasing emission from DCM was observed at 612 nm, with a bandwidth of 0.15 nm (Figure 11).

Several groups have explored the potential of 3DM as a tool for patterning waveguide-based optical circuitry. For example, a mixture of thermally and photocurable resins was used to generate embedded waveguides by photocross-linking within the thermally set medium. A Y-coupler was created in fused silica through the refractive index change generated by patterned MPE.

3DM provides a powerful route to 3D photonic crystals (PhC). PhCs are structures in which a basic motif, or 'unit cell', is repeated within the material with a given periodicity. PhCs have attracted great interest, because the modulation of the refractive

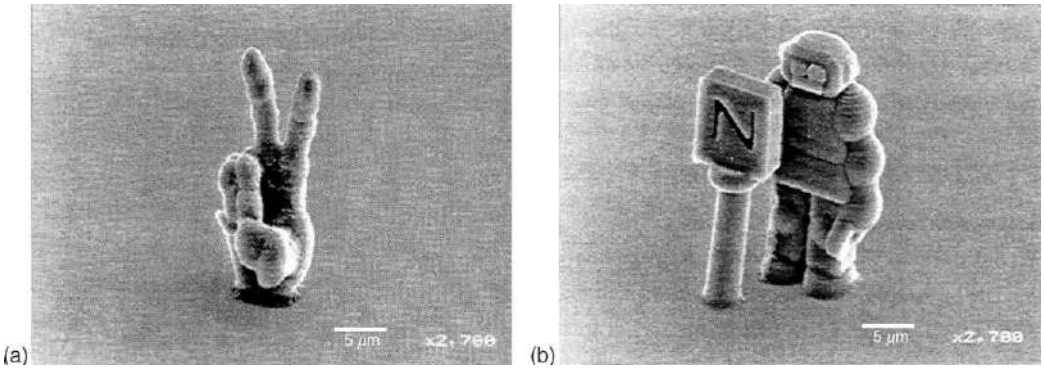

Figure 17 Scanning electron micrographs of two structures in the form of (a) a hand and (b) a robot which were fabricated by single-beam 3DM. Reproduced from Maruo S and Ikuta K (2000) *Proceedings of the Society of Photo-Optical Instrumentation Engineers* 3937: 106–112. Copyright (2000), with permission of SPIE.

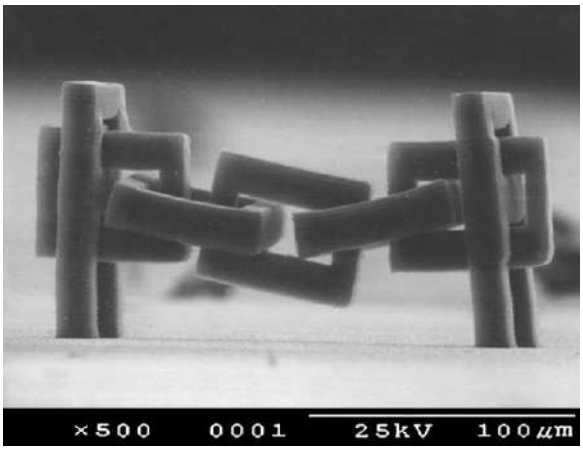

Figure 18 Scanning electron micrograph of a series of connected chain links obtained by single beam TPIP-3DM in an acrylate resin. Notice the tilt angle between the central links, indicating that each link can move with respect to the others. Reproduced from Kuebler SM, Rumi M, Watanabe T, *et al.* (2001) Optimizing two-photon initiators and exposure conditions for three-dimentional lithographic microfabrication. *Journal of Photopolymer Science and Technology* 14: 657–668. Copyright (2001), with permission of the Technical Association of Photopolymers Japan.

index between the 'filled' areas and the surrounding space provides a means for controlling the propagation of electromagnetic radiation within the structure. PhC with periodicities in the submicron and micron range exhibit their peculiar properties in the visible and near infrared spectral region. Optical wavelength PhCs can be fabricated in a straightforward way by 3DM, either by serial patterning using the single beam approach or by parallel patterning using the interference method, for which the number and orientation of the beams is dictated by the symmetry of the desired structure. Typical examples are woodpile structures (Figure 12), which consist of layers of rods of a material fabricated side by side with a fixed spacing, and alternate layers are rotated by 90° with respect to one another. If the rods in one layer are displaced by half the lattice constant with respect to the previous layer with the same orientation, the structure obtained has the symmetry of a face-center cubic lattice. PhCs with the symmetry of the diamond crystal have also been fabricated, either by inducing changes in a glass

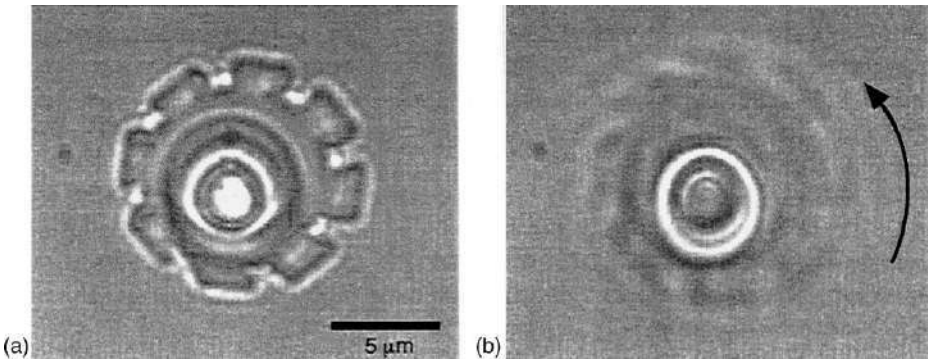

Figure 19 Optical image of a micro-wheel obtained by single-beam 3DM. (a) Structure before development. (b) When ethanol is used to remove the unpolymerized material, the wheel rotates around the axle in the solvent. Reproduced from Maruo S and Ikuta K (2000) *Proceedings of the Society of Photo-Optical Instrumentation Engineers* 3937: 106–112. With permission of the Institute of Electrical Engineers of Japan.

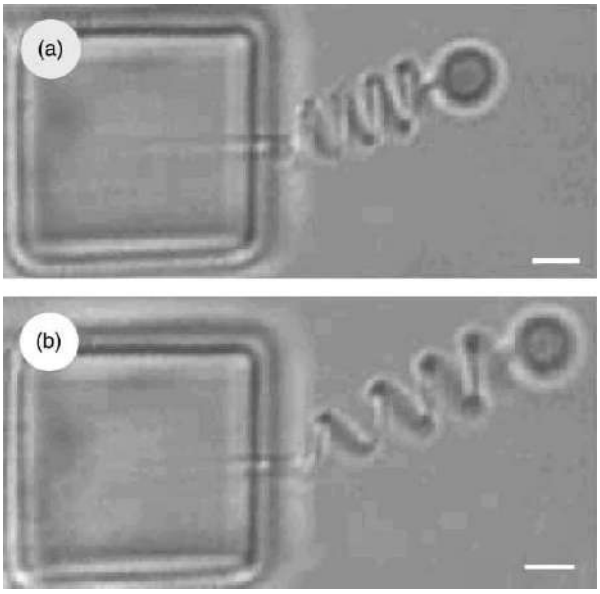

Figure 20 Optical image of a micro-spring tethered to a polymer block. The structure in (a) is at its resting point in ethanol. In (b) the spring was elongated by displacing the polymer bead at the free end using laser trapping. Reproduced with permission from *Nature*, Kawata S, *et al.* (2001) *Nature* 412: 697–698. Copyright (2001), Macmillan Magazines Limited.

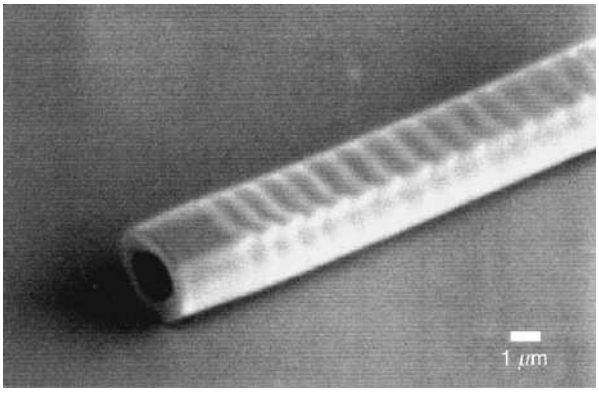

Figure 21 Hollow micro-tubule generated by single-beam TPIP-3DM in a urethane acrylate. The inner diameter is 1.8 μm. Reproduced from S Maruo and S Kawata (1998) *Journal of Microelectromechanical Systems* 7: 411–415. Copyright (1998), with permission of IEEE.

substrate at lattice points (Figure 13), or by constructing a network of polymeric rods connecting nearest neighbor points (Figure 14). It was showed that a PhC created by TPIP-3DM does exhibit a partial bandgap in the infrared spectral region. The position of the bandgap can be tuned by fabricating the structure with the appropriate periodicity.

The ability to introduce defects in PhCs in a controlled fashion is of key importance for the application of these structures in optical circuitry, as they can act as waveguides for channeling light into preferential directions. Point, line, and plane defects can be generated easily, by locally altering the fabrication pattern in the PhC. Point-by-point MPE of a Ge-doped glass was used to pattern a cubic lattice containing systematic defects that defined a Y-shaped waveguide splitter in a [100] plane. 3DM can also be used to modify a structure that is created using another fabrication method, and additional materials can be introduced into the structure in this second fabrication step. A silica colloidal crystal was back-filled with a polymerizable material and waveguides were generated by photocross-linking channels in the medium through three-photon-excitation 3DM (Figure 15).

The full potential of single-beam 3DM can be appreciated by surveying some of the structures that have been fabricated, featuring topologically complex shape, high interconnectivity, and extreme under-cut. Examples include those of Figures 16, 17, and the diamond-lattice PhC of Figure 14. Structures with freely movable parts, like those needed for MEMS, are difficult to obtain by other fabrication techniques, but can often be made by 3DM in a single exposure step. The ability to generate structures with independent but interconnected parts is exemplified by the microchain of Figure 18. A further example is the microgear in Figure 19, which is shown to spin around its axle when the sample is flooded with a liquid. Other functional microstructures have also been reported, such as the microspring in Figure 20, fabricated in a urethane acrylate resin and characterized by a core diameter of only 300 nm. When the sample is immersed in ethanol, the pendant mass can be captured and extended with optical tweezers. When the mass is released, the spring behaves like a damped oscillator.

Promising results have also been obtained in the production of structures to be used in micro-fluidics applications. Of key importance, in this case, is the ability to pattern channels, cavities, and other hollow structures that can be used to store or transfer small amounts of fluids between different parts of a device, as needed to perform, for example, analytical processes on a femtoliter scale. Examples include the hollow microtubule (Figure 21), and the submersed channels structure fabricated in a positive-tone resist (Figure 5).

List of Units and Nomenclature

A	Electron accepting molecular fragment
AFS	Amplified femtosecond
C	Chromophore
$\mathbf{C}^*$	Excited chromophore
CW	Continuous wave
D	Electron donating molecular fragment
3D	Three-dimensional
3DM	Three-dimensional microfabrication
$\vec{E}_0$	Spatial component of the electric field vector [V m^{-1}]
E_p	Laser pulse energy [J]
$f(t)$	Normalized temporal distribution of the beam intensity [dimensionless]
FOM	Figure of merit
h	Planck's constant [J s]
$i(z, r)$	Spatial distribution of the beam intensity [W cm^{-2}]
I	Intensity of a propagating beam [W cm^{-2}]
I_0	Maximum on-axis intensity for a Gaussian beam [W cm^{-2}]
I_{th}	Threshold intensity (minimum peak intensity at which polymerization occurs for a given exposure time) [W cm^{-2}]
I·	Initiating radical
j, l, q	Integers [dimensionless]
J_1	First order Bessel function [dimensionless]
K	Amplitude of a beam wave vector ($K = 2\pi/\lambda$) [m^{-1}]
$\vec{K}$	Wave vector of a beam [m^{-1}]
m	Number of beams in a multi-beam interference configuration [dimensionless]
M	Monomer
MBI	Multi-beam interference
MEMS	Micro-electromechanical systems
MPE	Multi-photon excitation
n	Number of photons involved in MPE [dimensionless]
n_0	Linear refractive index of the medium [dimensionless]
N	Number density (concentration) of a species [cm^{-3}]
NA	Numerical aperture of a lens [dimensionless]
N_C	Number density of chromophores present [cm^{-3}]
Nd:YAG	Neodymium:Yttrium aluminum garnet
$N_{I\cdot}$	Number density of initiating radicals generated [cm^{-3}]
N_{Inh}	Number density of inhibitors present [cm^{-3}]
OPE	One-photon excitation
Ormocer	Organically modified ceramic
$\langle P \rangle$	Average laser power [W]
$\langle P \rangle_{min}$	Minimum average laser power for patterning a material [W]
	Repetition rate (number of laser pulse per unit time) [Hz]
PAG	Photo-acid generator
PhC	Photonic crystal
r	Radial coordinate (distance from optical axis) [m]
r'	Component or $\vec{r}$ along a given wave vector [m]
$\vec{r}$	Position vector in three-dimensional space [m]
R	Twice the radial coordinate of the point at which the intensity is $I_0/2$ [m]
$R_{1h\nu}$	Molecular one-photon excitation rate (number of molecules excited per unit volume and time via OPE) [cm^{-3} s^{-1}]
$R_{2h\nu}$	Molecular two-photon excitation rate (number of molecules excited per unit volume and time via TPE) [cm^{-3} s^{-1}]
$R_{nh\nu}$	Molecular n-photon excitation rate (number of molecules excited per unit volume and time via n-order MPE) [cm^{-3} s^{-1}]
RMS	Root-mean-square
t	Time [s]
TEM_{00}	Lowest order transverse electromagnetic mode of a laser
Ti:S	Titanium:sapphire
TPE	Two-photon excitation
TPIP	Two-photon induced polymerization
z	Coordinate along the beam propagation direction [m]
z_R	Rayleigh range of a Gaussian beam; axial distance between the point of minimum beam waist and the point at which the waist has increased by a factor of $\sqrt{2}$ [m]
Z	Distance between the points of intensity $I_0/2$ along the coordinate z [m]
ζ	Longitudinal optical coordinate ($\zeta = 2\pi Z(\mathrm{NA})^2/n_0\lambda$) [dimensionless]
θ	Angle between a beam wave vector and the normal to a substrate [rad]
λ	Wavelength (in vacuum) [m]
Λ	Period of a one-dimensional grating [m]

ν	Frequency [s^{-1}]
$\boldsymbol{\pi}$	π-conjugated molecular fragment
ρ	Transverse optical coordinate ($\rho = 2\pi r\mathrm{NA}/\lambda$) [dimensionless]
$\sigma^{(1)}$	One-photon absorption cross-section [cm^2]
$\sigma^{(2)}$	Two-photon absorption cross-section [cm^4 s $photon^{-1}$]
$\sigma^{(n)}$	n-photon absorption cross-section [cm^{2n} s^{n-1} $photon^{-(n-1)}$]
τ_{ex}	Exposure time [s]
τ_p	Laser pulse duration [s]
$\phi_c^{(n)}$	n-photon chemical quantum yield [(# of chemical species generated)/(# of n-photon-excited species)] [dimensionless]
ϕ_{H+}	Chemical yield for the generation of H^+ (number of H^+ generated per excited molecule) [dimensionless]
ω	Angular frequency ($\omega = 2\pi\nu$) [Hz]

See also

Chemical Applications of Lasers: Nonlinear Spectroscopies. **Diffractive Systems:** Applications of Diffractive and Micro-Optics in Lithography; Design and Fabrication of Diffractive Optical Elements. **Holography, Techniques:** Overview. **Imaging:** Volume Holographic Imaging. **Laser-Induced Damage of Optical Materials**. **Materials for Nonlinear Optics:** Organic Nonlinear Materials. **Microscopy:** Confocal Microscopy; Imaging Multiple Photon Fluorescence Microscopy; Nonlinear Microscopy; Overview. **Photonic Crystals:** Electromagnetic Theory; Nonlinear Optics in Photonic Crystal Fibers; Self-Assembled and Functionalized Photonic Crystals. **Ultrafast Laser Techniques:** Generation of Femtosecond Pulses.

Further Reading

Belfield KD, Schafer KJ, Liu YU, Liu J, Ren XB and van Stryland EW (2000) Multiphoton-absorbing organic materials for microfabrication, emerging optical applications and non-destructive three-dimensional imaging. *Journal of Physical Organic Chemistry* 13: 837–849.

Born M and Wolf E (1980) *Principles of Optics*, 6th edn. Cambridge, UK: Cambridge University Press.

Cumpston BH, Ananthavel SP, Barlow S, *et al.* (1999) Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication. *Nature* 398: 51–54.

Joshi MP, Pudavar HE, Swiatkiewicz J, Prasad PN and Reinhardt BA (1999) Three-dimensional optical circuitry using two-photon-assisted polymerization. *Applied Physics Letters* 74: 170–172.

Kawata S and Sun H-B (2003) Two-photon photopolymerization as a tool for making micro-devices. *Applied Surface Science* 208–209: 153–158.

Kuebler SM, Rumi M, Watanabe T, *et al.* (2001) Optimizing two-photon initiators and exposure conditions for three-dimensional lithographic microfabrication. *Journal of Photopolymer Science and Technology* 14: 657–668.

Kuebler SM, Braun KL, Zhou W, *et al.* (2003) Design and application of high-sensitivity two-photon initiators for three-dimensional microfabrication. *Journal of Photochemistry and Photobiology A: Chemistry* 158: 163–170.

Martineau C, Anémian R, Andraud C, Wang I, Bouriau M and Baldeck PL (2002) Efficient initiators for two-photon induced polymerization in the visible range. *Chemical Physics Letters* 362: 291–295.

Odian G (1981) *Principles of Polymerization*, 2nd edn. New York: Wiley.

Pitts JD, Campagnola PJ, Epling GA and Goodman SL (2000) Submicron multiphoton free-form fabrication of proteins and polymers: studies of reaction efficiencies and applications in sustained release. *Macromolecules* 33: 1514–1523.

Ren Y, Yu X-Q, Zhang D-J, *et al.* (2002) Synthesis, structure and properties of a new two-photon photopolymerization initiator. *Journal of Materials Chemistry* 12: 3431–3437.

Strickler JH and Webb WW (1990) Two-photon excitation in laser scanning fluorescence microscopy. *Proceedings of the Society of Photo-Optical Instrumental Engineers* 1398: 107–118.

Sun H-B and Kawata S (2003) Two-photon laser precision microfabrication and its applications to micro-nano devices and systems. *Journal of Lightwave Technology* 21: 624–633.

Sun H-B, Maeda M, Takada K, Chon JWM, Gu M and Kawata S (2003) Experimental investigation of single voxels for laser nanofabrication via two-photon photopolymerization. *Applied Physics Letters* 83: 819–821.

Wu ES, Strickler JH, Harrell WR and Webb WW (1992) Two-photon lithography for microelectronic application. *Proceedings of the Society of Photo-Optical Instrumental Engineering* 1674: 776–782.

Wu P-W, Cheng W, Martini IB, Dunn B, Schwartz BJ and Yablonovitch E (2000) Two-photon photographic production of three-dimensional metallic structures within a dielectric matrix. *Advanced Materials* 12: 1438–1441.

Xu C and Webb WW (1997) Multiphoton excitation of molecular fluorophores and nonlinear laser microscopy. In: Lakowicz JR (ed.) *Topics in Fluorescence Spectroscopy*, vol. 5, pp. 471–540. New York: Plenum.

NONLINEAR OPTICS, BASICS

Contents

Cascading
$\chi^{(2)}$–Harmonic Generation
$\chi^{(3)}$–Third-Harmonic Generation
Four-Wave Mixing
Kramers–Krönig Relations in Nonlinear Optics
Nomenclature and Units
Nonlinear Optical Phase Conjugation
Photorefraction
Ultrafast and Intense-Field Nonlinear Optics

Cascading

G Assanto, University 'Roma Tre', Rome, Italy

G I Stegeman, University of Central Florida, Orlando, FL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Historically the field of nonlinear optics has evolved along two main axes, depending on the order of the nonlinearity used. The phenomena associated with the second-order nonlinearity $\chi^{(2)}$ have dealt with frequency conversion requiring wavevector matching (i.e., momentum conservation), such as second harmonic and optical parametric generation. On the other hand, effects associated with the third-order susceptibility $\chi^{(3)}$ have focused on inherently phase-matched phenomena at a single frequency, for example an intensity-dependent refractive index or degenerate four-wave mixing.

In cascading, multiple sequential nonlinear events based on $\chi^{(2)}$ can mimic some well-known $\chi^{(3)}$ phenomena. For example, it leads to nonlinear phase shifts $\Delta\phi^{\mathrm{NL}}$ reminiscent of those obtained through self-phase modulation over a distance L, i.e.,

$$\Delta\phi^{\mathrm{NL}}(L, I) = k_0 L n_2 I \qquad [1]$$

where I is the irradiance, k_0 the vacuum wavevector, and n_2 the Kerr coefficient or nonlinear refractive index. These can be used for all-optical switching, the generation of spatio-temporal solitons, etc. Some of these cascading effects have been known from the very earliest days of nonlinear optics, but their effectiveness was not realized until recently, when it proved possible to phase match second-order interactions with usefully large nonlinearities. Note that the medium must be $\chi^{(2)}$-active, i.e., noncentrosymmetric, which usually requires crystals rather than amorphous materials (such as glass).

Here we review the basic concepts, the characteristic interactions, some of the numbers involved and, finally, a few applications.

Physics of Cascading

Nonlinear Phase Shift in Collinear Second Harmonic Generation

The most important cascading effect, the nonlinear phase shift, is rooted in the fundamentals of the parametric interaction when two or three waves of different frequencies ($\omega_3 = \omega_1 \pm \omega_2$) are coupled by the second-order susceptibility $\chi^{(2)}(-\omega_3;\ \omega_1,\ \pm\omega_2)$. The simplest case is 'Type I' second harmonic generation (SHG), in which one eigenwave at the fundamental (FF) and one at the harmonic frequency (SH) interact with $\omega_1 = \omega_2 = \frac{1}{2}\omega_3$. When the waves have different phase velocities, this process involves simultaneously up-conversion ($\omega_1 + \omega_1 \rightarrow \omega_3$) and down-conversion ($\omega_3 - \omega_1 \rightarrow \omega_1$). By the latter the FF is regenerated from the harmonic after a characteristic propagation distance called the coherence length. This phenomenon is cyclic (see Figure 1a).

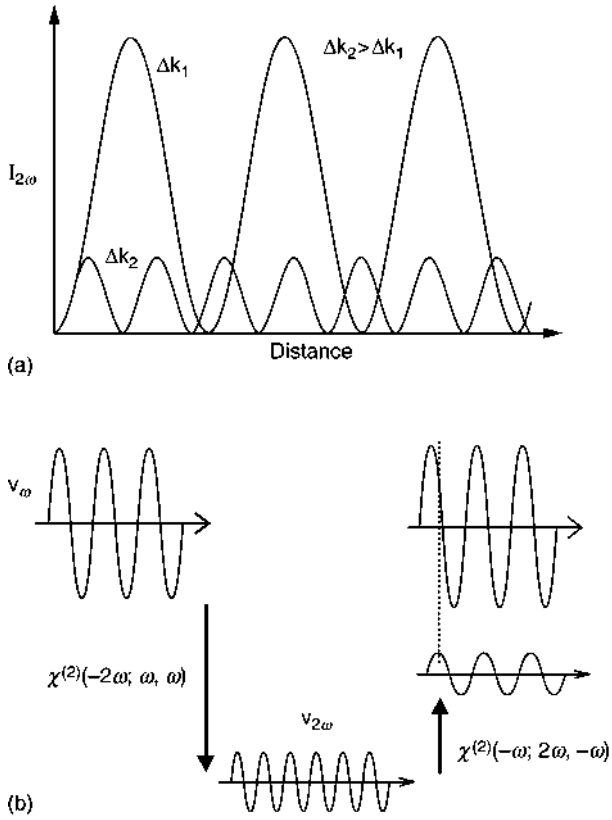

Figure 1 Phase-mismatched SHG. (a) Typical evolution of the generated second harmonic irradiance versus propagation for different mismatches. The period is twice the coherence length. (b) Intuitive picture of phase shift through cascading: the fraction of the initial energy which is first up-converted (SHG) and later down-converted (DFG), having traveled at a different phase velocity due to the wavevector mismatch, recombines with the through-FF causing an overall phase shift.

The field amplitudes in normalized form are

$$E_1(z,t) = \frac{1}{2}\sqrt{\frac{2}{c\varepsilon_0 n_1}}a_1(z)\mathrm{e}^{\mathrm{i}(\omega_1 t - k_1 z)} + \text{c.c.}$$
$$E_3(z,t) = \frac{1}{2}\sqrt{\frac{2}{c\varepsilon_0 n_3}}a_3(z)\mathrm{e}^{\mathrm{i}(\omega_3 t - k_3 z)} + \text{c.c.} \qquad [2]$$

where $|a_1|^2$ and $|a_3|^2$ are power densities, c is the speed of light in vacuum, and n_j and k_j are the refractive indices and wavevectors at angular frequencies $\omega_j\,(j = 1, 3)$, respectively. As is common in nonlinear optics, the efficiency of these processes depends on the irradiance of the different waves, and on the wavevector mismatch between them, i.e., $\Delta k = 2k_1 - k_3$ for collinear beams. Since the phase velocities of the two waves, $v_{\omega_1} = \omega_1/k_1$ and $v_{\omega_3} = \omega_3/k_3$, are generally different and $\Delta k \neq 0$, their phase fronts fall out of synchronism during propagation. This leads to a periodic oscillation with distance (coherence length $= \frac{1}{2}$period $= \pi/|\Delta k|$) in the irradiances of the fundamental and harmonic waves, conserving the total energy but giving rise to a cascading phase shift, as pictured in Figure 1b.

The equations that describe the evolution of the fields stem from coupled mode theory, and are given by

$$\frac{\mathrm{d}}{\mathrm{d}z}a_3(z) = -\mathrm{i}\Gamma a_1^2(z)\mathrm{e}^{-\mathrm{i}\Delta kz}$$
$$\frac{\mathrm{d}}{\mathrm{d}z}a_1(z) = -\mathrm{i}\Gamma a_3(z)a_1^*(z)\mathrm{e}^{\mathrm{i}\Delta kz} \qquad [3]$$

where

$$\Gamma = \frac{\omega_1 \chi_{\text{eff}}^{(2)}}{\sqrt{2n^3c^3\varepsilon_0}} \qquad [4]$$

$\chi_{\text{eff}}^{(2)}$ is the 'effective' nonlinearity (for a specific crystal orientation and field polarization) and we assumed $n = n_1 \cong n_3$. In the limit of a large mismatch, the energy conversion to the harmonic is negligibly small, and it is possible to find an analytical expression for an effective 'n_2', $n_{2,\text{eff}}$. In this limit $|a_1(z)| = |a_1(0)| =$ constant, and, in the absence of loss, solving for the small SH wave and substituting back into the equation for $\mathrm{d}a_1(z)/\mathrm{d}z$ gives

$$\frac{\mathrm{d}}{\mathrm{d}z}a_1(z) = -\mathrm{i}\frac{\Gamma^2}{\Delta k}\Big\{1 - \cos(\Delta kz) - \mathrm{i}\sin(\Delta kz)\Big\}|a_1(0)|^2 a_1(z) \qquad [5]$$

By comparing this with the corresponding equation based on n_2 ($\propto \chi^{(3)}[-\omega;\omega,-\omega,\omega]$) and describing self-phase modulation, namely

$$\frac{\mathrm{d}}{\mathrm{d}z}a_1(z) = -\mathrm{i}n_2\frac{\omega_1}{c}|a_1(z)|^2 a_1(z) \qquad [6]$$

we can define an equivalent nonlinear refractive index

$$n_2(z) = \frac{\pi\{\chi_{\text{eff}}^{(2)}\}^2}{2\varepsilon_0 c n^4 \lambda_1 \Delta k}[1 - \cos(\Delta kz)] \qquad [7]$$

which turns out to be strictly nonlocal because it relies on propagation. Since the $\cos(\Delta kz)$ term oscillates rapidly in space, the above averages to

$$n_{2,\text{eff}} = \frac{\pi\{\chi_{\text{eff}}^{(2)}\}^2}{2\varepsilon_0 c n^4 \lambda_1 \Delta k} \qquad [8]$$

In this so-called 'Kerr limit' cascading leads to a Kerr-like nonlinear response, with either a

self-focusing ($\Delta k > 0$, $n_{2,\mathrm{eff}} > 0$) or a self-defocusing nonlinearity, depending on the sign of Δk.

It is worthwhile to link the above equations to the physics of cascading with a simple argument. The evolution of the waves on propagation depends on their fields (in both amplitude and phase) at the input. When only the fundamental beam at $\omega = \omega_1$ is incident (the classical case of SHG), the harmonic is initially up-converted $\pi/2$ out of phase with the input (first of eqns [3]) via $\chi^{(2)}(-2\omega; \omega, \omega)$. If $\Delta k = 0$, the phase fronts travel at the same velocity and the relative phase is locked in at $\pi/2$. Difference frequency generation (DFG, second of eqns [3]) results in a down-converted fundamental which is an additional $\pi/2$ out of phase, i.e., exactly out of phase with the original input field. This leads to the monotonic depletion of the input and the growth of the SH with distance. If $\Delta k \neq 0$, there is only partial and periodic conversion to the harmonic, the wavefronts travel at different velocities (v_ω and $v_{2\omega}$, respectively), and the relative phase between the two waves changes with distance. As a result, the down-converted FF is no longer π out of phase with the input, thereby changing the overall phase of the total input-frequency wave by a cascaded amount $\Delta\phi^{\mathrm{NL}}$. This is the mechanism sketched in Figure 1b and buried in the better-known Maker-fringe effect represented in Figure 1a. Its magnitude depends on the irradiance and the mismatch, and its sign on that of Δk (see eqn [8]). The higher the input irradiance, the larger the harmonic conversion, and hence the larger $\Delta\phi^{\mathrm{NL}}$.

Alternatively, with a harmonic input at high irradiance and a weak seed at the FF, parametric amplification makes the fundamental grow via down-conversion. If $\Delta k = k_3 - 2k_1 \neq 0$, self-focusing (-defocusing) is obtained on the strong harmonic for $\Delta k > 0$ ($\Delta k < 0$). These phenomena can be generalized to the nondegenerate cases of sum and difference frequency generation, keeping in mind that when using three input waves their mutual phase relationship is essential in determining the details of the interaction.

The solutions to eqns [3] can be obtained numerically or written in terms of Jacobi elliptic integrals subject to boundary conditions ($a_3(0) = 0$, $a_1(0) \neq 0$) at $z = 0$, providing additional insight into the cascading phenomenon. In contrast to the Kerr case given by eqn [1], the variation in n_2 with distance implies a staircase-like increase in $\Delta\phi^{\mathrm{NL}}$ with distance for various detunings ΔkL, as shown in Figure 2a. This is consistent with our intuitive picture above, the increase in $\Delta\phi^{\mathrm{NL}}$ being maximized where the fundamental is the smallest (see Figure 2b for the oscillations in the fundamental amplitude), with the step-period corresponding to twice the coherence length. Conversely, for a fixed input irradiance, the evolution of $\Delta\phi^{\mathrm{NL}}$ exhibits two extrema with detuning ΔkL, as shown in Figure 3a. Note also that, for small ΔkL and large

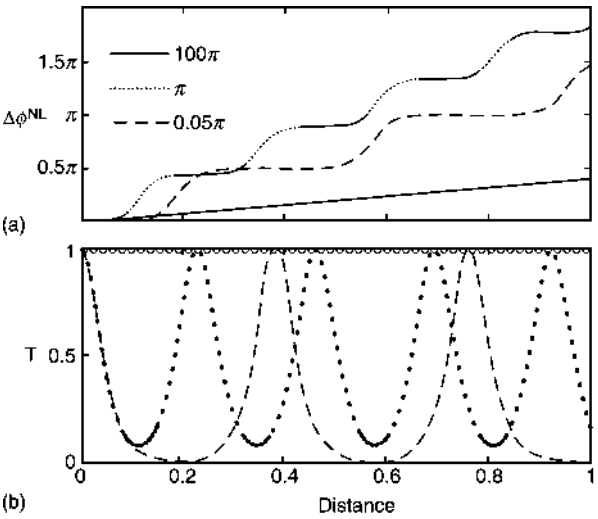

Figure 2 Calculated nonlinear phase (a) and throughput (b) of the fundamental field versus normalized propagation distance, for three values of phase mismatch.

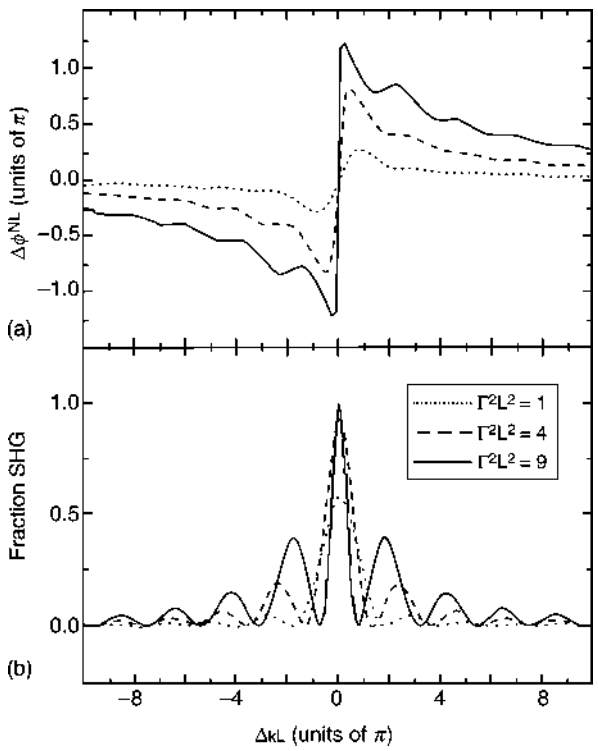

Figure 3 Calculated FF phase shift (a) and generated SH fraction (b) versus phase mismatch, i.e., the detuning curve, for three different nonlinear strengths (unitary FF irradiance).

(unrealistically high) irradiances, the increase in $\Delta\phi^{NL}$ tends to become proportional to the field amplitude, a clear indication that cascading depends on a different order of nonlinearity than $\chi^{(3)}$, namely $\chi^{(2)}$. On the other hand, for large ΔkL the growth of $\Delta\phi^{NL}$ is slower but linear with distance and irradiance, as in eqn [1].

There are trade-offs between the fundamental output and the net nonlinear phase shift. This is clearly visible in Figures 3a,b: the maxima in $\Delta\phi^{NL}$ correspond to a reduced FF throughput. Although it is possible to operate cw at the zeros of the SHG, their location changes with the irradiances required for useful phase shifts ($>\pi$), complicating the trade-offs and the response to temporal pulses (encompassing a continuous irradiance distribution). The problem can be alleviated at the expense of reducing the net phase shift for a given irradiance and propagation length, by introducing regions with a nonuniform matching condition at both ends of the sample. The goal at the input end is to first adiabatically generate the SH required for optimum $\Delta\phi^{NL}$. At the output, this harmonic is adiabatically converted back to the FF. The corresponding output for wavelengths larger than the phase-matching condition exhibits a $\Delta\phi^{NL}$ with negligible fundamental depletion. This approach has been used in temperature-tuned and quasi phase matched (QPM, where a periodic modulation of the nonlinearity provides phase matching via the additional grating momentum) lithium niobate waveguides to generate phase shifts of π with less than 10% net loss of the fundamental wave to the harmonic.

While the preceding discussion has focused on Type I SHG, it is also valid for Type II SHG (i.e., with two FF input waves), provided the two fundamental inputs have equal input energy. In fact, an input imbalance between the two fundamental polarizations could give rise to a phase shift on the weaker field even when $\Delta k = 0$, leading to all-optical modulation and transistor action, as demonstrated in potassium titanyl phosphate (KTP) crystals.

Finally, due to the nature of the quadratic nonlinearity, the phase and/or amplitude of a coherent SH seed can have a large effect on the cascaded outcome, both in the phase and amplitude of the fundamental, as shown in Figure 4 for a fixed irradiance seed 1000 times weaker than the FF input. This can be used in seed-controlled optical processors.

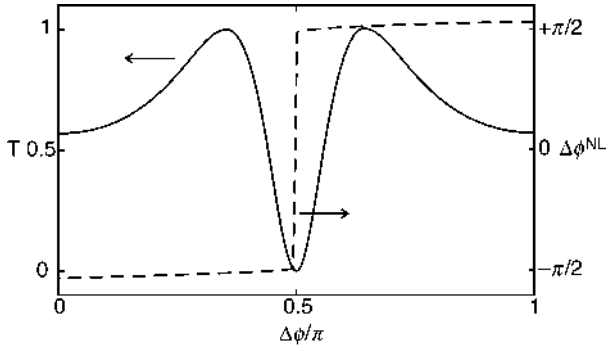

Figure 4 FF transmission (solid line) and phase shift (dashed line) versus the relative phase of a much weaker (three orders of magnitude) SH seed. Here perfect phase matching was assumed.

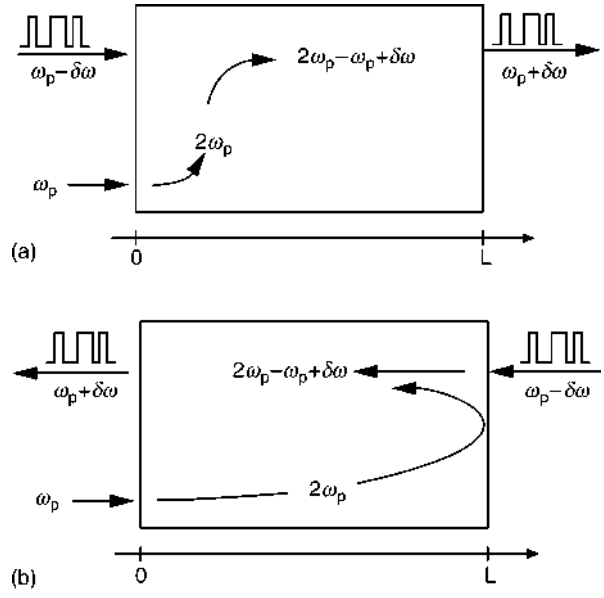

Figure 5 Wavelength shifter: (a) single-pass and (b) double-pass configurations. In (b) a dichroic mirror reflects the second harmonic of the pump.

Frequency Shifting

In the nondegenerate case of $\chi^{(2)}$ sum and difference frequency generation, two successive operations can perform the useful function of a frequency shifter (Figure 5). The idea is to shift the signal frequency ($\omega_p - \delta$) by first generating from the pump (ω_p) a second harmonic ($2\omega_p$), followed by DFG involving the latter and the signal. This leads to an output at $\omega_p + \delta$ through $\chi^{(2)}(-(\omega_p + \delta); 2\omega_p, -(\omega_p - \delta))$, i.e., a frequency shift of 2δ with an output spectrally mirror-imaged about the pump frequency. This feature is useful in WDM (wavelength division multiplexing) systems, and guarantees a protocol-insensitive wavelength shift with an additional phase conjugation useful in dispersion-compensating schemes. Notice that, since $\omega_p + \delta \cong \omega_p - \delta$ and $\delta << \omega_p$, the bandwidth is large because the interaction is near the degeneracy point.

Figure 5 shows two possible configurations for the same device length. The efficiency of the process can be increased in a double-pass geometry by using a back-propagating signal and a mirror-coated facet to reflect the SH, as sketched in Figure 5b. The best results to date were obtained at communication wavelengths with single-pass waveguides in periodically poled (QPM) $LiNbO_3$, including shift of multiple wavelength-channels with low cross-talk.

Multiple Beam Generation

Two successive $\chi^{(2)}$ processes can also be used to mimic three-wave mixing, a classic noncollinear $\chi^{(3)}$ interaction where two photons are taken out of one beam and one out of the second. In cascading, first a second harmonic is generated by one of the inputs, and then difference frequency mixing (DFM) between that SH and the second input leads to the signal of interest. In the simplest configuration, two equi-frequency (ω) beams of wavevectors $\mathbf{k}_1$ and $\mathbf{k}_2$ are incident in slightly different directions close to phase-matching and overlap inside the sample. When wave 1 is doubled (2ω) and wave 2 mixes with this harmonic via DFM, a new beam (at $2\omega - \omega \rightarrow \omega$) is generated with wavevector $2\mathbf{k}_1 - \mathbf{k}_2$. Of course the complementary output at $2\mathbf{k}_2 - \mathbf{k}_1$ is also generated via the SHG of wave 2. If these two new beams are strong enough, they can generate another pair via cascading, and so on. The generation of more than six extra beams has been demonstrated in β-barium borate with 1064 nm inputs.

This interaction can be applied to the control of the frequency chirp in pulses. For example, if wave 1 has a frequency chirp α_1 and wave 2 a chirp α_2, then the chirp of the beam scattered into the direction $2\mathbf{k}_1 - \mathbf{k}_2$ is $2\alpha_1 - \alpha_2$. This becomes especially interesting if $\alpha_2 = -\alpha_1$ so that the chirp in the new beam is $3\alpha_1$, as observed and subsequently employed in enhancing pulse compression. Note that, since a cascaded phase shift occurs on the down-conversion step back to the original fundamental (wave 1), it is not present in the scattered beam resulting in a clean transfer of the chirp.

Typical Numbers

From eqn [8] it is clear that the figure of merit (FOM) for cascaded processes is the same as for all second-order processes, namely $|\chi^{(2)}|^2/n^4$. The key is to find materials phase-matchable at the wavelengths of interest. Therefore, basically any material that is suited for frequency conversion, either SHG or parametric generation, will work. The most versatile material, which also has the largest easily usable nonlinear coefficient (18 pm/V), is lithium niobate when QPM is employed. As a reference, Table 1 gives the maximum $n_{2,\mathrm{eff}}$ for 1-cm-long samples, obtained by extrapolating eqn [8] to the ΔkL which maximizes the nonlinear phase shift. This is not strictly valid for such small ΔkL, but it is a useful estimate.

Other Implications of Cascading

A straightforward conclusion that could be drawn is that cascading with its $n_{2,\mathrm{eff}}$ and $\Delta\phi^{\mathrm{NL}}$ can generally replace the Kerr n_2. This is true for self-phase modulation and the applications stemming from it, including all-optical guided-wave switches such as directional couplers, Mach–Zehnder interferometers, etc. However, this is not necessarily true for other applications, because all the waves have to be coupled through $\chi^{(2)}$ interactions (at most three at a time) and must be nearly phase-matched for efficient effects. For example, cross-phase modulation between two input FF polarization components is feasible via Type II SHG cascading, but specific wavelengths will require different crystals, or crystal cuts. The nonlinear phase shift can also be used in all-optical devices for analog processing, such as isolators, modulators, and transistors, in laser mode-locking, etc.

In the following subsections we discuss a few additional applications. Specifically, we describe a non-reciprocal device that operates as an isolator, and the cascading of beams of finite extent in space or time. The latter can counterbalance either spatial diffraction or temporal dispersion or both and gives rise to solitons of various dimensionalities.

An Optical Diode

A nonlinear response, when combined with a device structure that is asymmetric about its midpoint, can provide nonreciprocal behavior upon excitation of the structure (crystal, waveguide) from different ends.

Table 1 Figures for cascading in various materials. The effective nonlinear coefficient is estimated under optimum conditions and a 1 cm propagation length

Material L = *1 cm*	d_{ii} *(pm/V)*	d_{ij} *(pm/V)*	$n_{2,eff}$ *(cm²/W)*
$LiNbO_3$	36		2×10^{-11}
$LiNbO_3$		5.8	5×10^{-13}
MNA(1)	165		7×10^{-10}
NPP(2)		84	2×10^{-10}
DAST(3)	600		6×10^{-9}

(1)MNA: 2-methyl-4-nitroaniline
(2)NPP: N-(4-nitrophenyl)-(L)-prolinol
(3)DAST: dimethyl amino stilbazolium tosylate

In analogy to the standard electronic element, we can refer to such device as a power-dependent optical diode. Dealing with quadratically nonlinear waveguides, simple ways to induce a directional asymmetry are variations in wavevector mismatch, or in QPM periodicity. A mismatch profile which depends on distance from one end, or a localized defect (i.e., an isolated domain in a QPM grating), will give rise to different throughputs when the waveguide is excited from opposite sides, up to complete FF depletion in one direction, and complete transmission in the opposite one. This can be explained by the fact that the output depends on the evolution and interaction of the FF field with its harmonic. The amount and phase of the latter, however, do depend on the details of the propagation path and can substantially change the FF throughput as in the case of seeded SHG.

The sketch of an all-optical diode in a QPM waveguide with its calculated response is shown in Figure 6a,b. Clearly, total depletion is obtained via SHG in backward propagation, whereas unity FF transmission is recovered by the insertion of a phase discontinuity (and seeded interaction) in the forward path. As demonstrated at 1.55 μm in lithium niobate, different degrees of isolation are available at various excitations, with a large extinction ratio at a given power when launching FF waves from opposite ends.

Quadratic Solitons

When beams propagate in a two- (planar waveguides) or three-dimensional (bulk) medium with a quadratic response, cascading induces a nonlinear phase-front distortion which, coupled to the inherent parametric gain, can effectively counteract diffraction and lead to diffractionless propagation or 'spatial solitons' (Figure 7). In contrast to the Kerr case, such solitons are multifrequency solutions or 'simultons' with field components (FF and SH in the degenerate SHG process) bound together in propagation despite possible walk-off or transverse velocity mismatch. Moreover, since the second-order nonlinearity is inherently saturable, the solitons are stable and robust even in bulk media, and provide 'clean' filtered beam profiles at FF and SH frequencies, as has been demonstrated in KTP. Just like other cascading phenomena, quadratic solitons can be controlled by a weak seed, by polarization imbalance at the input (Type II SHG), by direction of propagation (due to birefringence), by phase mismatch, etc. Employing quadratic spatial solitons, phenomena such as collisions, soliton fusion, spatial reshaping, and transverse light localization and instabilities have been studied and constitute an active field of investigation.

When employing pulses, the cascaded phase shift can balance dispersion (material, chromatic, or photonics bandgap) leading to temporal simultons with frequency components locked together despite their individual group-velocity differences and dispersion. This feature can also substantially broaden the bandwidth available for parametric generation. Finally, in analogy to the cubic case, a combination of space and time effects is expected to lead to the formation of 'light bullets' in (3 + 1) dimensions.

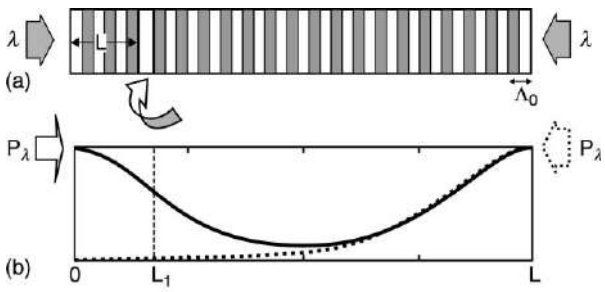

Figure 6 All-optical diode: (a) sketch of implementation with an engineered 'defect' in a QPM grating (*z*-cut periodically poled lithium niobate), and (b) example of FF transmittance *T* versus propagation for forward (solid line) and backward (dashed line) FF excitation of power P_λ.

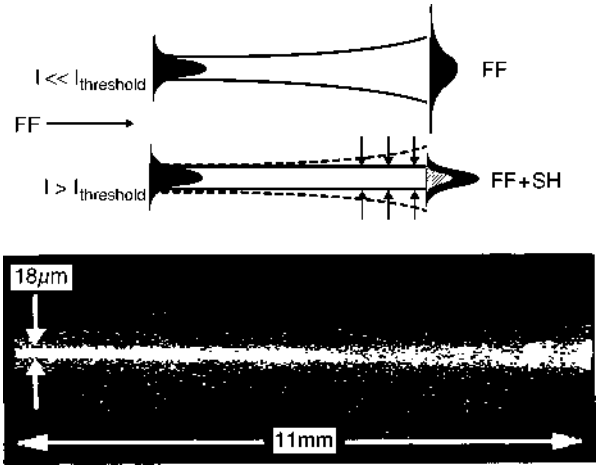

Figure 7 Top: sketch of a quadratic spatial soliton excited by an FF input: when the irradiance is high enough, linear diffraction is overcome by parametric self-trapping and a two-color simulton is obtained. Bottom: sample photograph (top view) of a quadratic soliton experimentally observed in a potassium niobate crystal excited by a 1064-nm laser beam.

List of Units and Nomenclature

Angular frequency	$[s^{-1}]$	ω
Dielectric constant of vacuum	$[FV^{-1}\,m^{-1}]$	ε_0

Electric field	[V m^{-1}]	E
Irradiance, intensity	[W m^{-2}]	I
Length	[m]	L
Refractive index	[ ]	n
Speed of light in vacuum	[m s^{-1}]	c
Second-order electronic susceptibility	[m V^{-1}]	$\chi^{(2)}$
Third-order electronic susceptibility	[m^2 V^{-2}]	$\chi^{(3)}$
Wavevector in medium	[m^{-1}]	$k = n\omega/c$
Wavevector in vacuum	[m^{-1}]	$k_0 = \omega/c$
Wavelength	[m]	$\lambda = 2\pi n/k = 2\pi/k_0$

See also

Solitons: Bright Spatial Solitons; Soliton Communication Systems. **Spectroscopy:** Second Harmonic Spectroscopy.

Further Reading

Assanto G (1995) All-optical switching in integrated structures. In: DeGiorgio V and Flytzanis C (eds) *Nonlinear Optical Materials: Principles and Applications*. Amsterdam: IOS Press.

Assanto G and Stegeman GI (2002) Simple physics of quadratic spatial solitons. *Optics Express* 10: 388–396.

Assanto G, Stegeman GI, Sheik-Bahae M and VanStryland E (1993) All optical switching devices based on large nonlinear phase shifts from second harmonic generation. *Applied Physics Letters* 62: 1323–1325.

Assanto G, Stegeman GI, Sheik-Bahae M and VanStryland EW (1995) Coherent interactions for all-optical signal processing via quadratic nonlinearities. *IEEE Journal of Quantum Electron* 31: 673.

Assanto G, Stegeman GI and Schiek R (1998) Thin film devices for all-optical switching and processing via quadratic nonlinearities. *Thin Solid Films* 331: 291–297.

Boardman AD (ed.) (1999) *Advanced Photonics with Second-Order Optically Nonlinear Processes*. Dordrecht, The Netherlands: Kluwer.

Boardman AD (ed.) (2001) *Soliton Driven Photonics*. Dordrecht, The Netherlands: Kluwer.

Gallo K and Assanto G (1999) Analysis of lithium niobate all-optical wavelength shifters for the third spectral window. *Journal of the Optical Society of America B* 16: 741–753.

Gallo K, Assanto G, Parameswaran KR and Fejer MM (2001) All-optical diode in a periodically poled lithium niobate waveguide. *Applied Physics Letters* 79: 314–316.

Kajzar F and Reinisch R (eds) (1997) *Beam Shaping and Control with Nonlinear Optics*. New York: Plenum Press.

Stegeman GI, Hagan DJ and Torner L (1996) $\chi^{(2)}$ cascading phenomena and their applications to all-optical signal processing, mode-locking, pulse compression and solitons. *Journal of Optics and Quantum Electronics* 28: 1691.

Stegeman GI and Assanto G (1999) Nonlinear integrated optical devices. In: Murphy EJ (ed.) *Design and Application of Integrated Optical Circuits and Components*. New York: Marcel Dekker.

$\chi^{(2)}$–Harmonic Generation

R C Eckardt, Cleveland Crystals, Inc., Highland Heights, OH, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical harmonic generation is nonlinear conversion of laser output to light, of a frequency that is an integer multiple of that of the fundamental. It is used over a wide range of applications, from battery-operated, pencil-sized pointers emitting less than a milliwatt of light to installations that occupy large buildings and generate thousands of Joules of ultraviolet energy in nanosecond-duration pulses. The applications also range from frequency-stable continuous-wave light to generation of pulses a few femtoseconds in duration. Laser output with specific properties may only be available in narrow spectral regions, and harmonic generation is a method to convert the laser output to spectral regions of shorter wavelength. It is customary to call light of twice the frequency of the fundamental the second harmonic; three times the third harmonic, and so on.

The conversion of light by harmonic generation is a process that is driven by the nonlinear electric polarization. The total polarization $\mathbf{P}(t)$, including the linear term, is a vector quantity related to the electric field $\mathbf{E}(t)$ by the susceptibility tensors

χ of a material:

$$\mathbf{P}(t) = \varepsilon_0\chi^{(1)} : \mathbf{E}(t) + \varepsilon_0\chi^{(2)} : \mathbf{E}(t)\mathbf{E}(t) + \varepsilon_0\chi^{(3)} : \mathbf{E}(t)\mathbf{E}(t)\mathbf{E}(t) + \ldots, \quad [1]$$

where $\varepsilon_0 \cong 8.854 \times 10^{-12}$ C/N·m^2 is the permittivity of free space. The first-order tensor describes linear optical properties on the material. The second-order tensor $\chi^{(2)}$ describes second-harmonic generation and sum- and difference-frequency generation. This is the term that is of interest here. The second-order nonlinear optical tensor, equal to half the second-order susceptibility, is used widely instead of the susceptibility:

$$d_{ijk} = \chi^{(2)}_{ijk}/2 \quad [2]$$

In calculations of harmonic generation and sum- and difference-frequency generation, the coefficient is usually reduced to a scalar quantity d_{eff}, the effective nonlinear coefficient that expresses the relationship between only the specific vector components involved in an interaction. The second-order optical nonlinearity, when present, is usually significantly larger than higher-order optical nonlinearities. When optical frequency conversion is possible, using the second-order nonlinearity, it is generally preferred. Any material that has inversion symmetry, however, will have vanishing second-order nonlinearity. Although there is continuing interest in developing other materials, such as poled polymers for future applications, materials for current practical applications are restricted to noncentrosymmetric crystals.

This article will provide a description of the process of optical harmonic generation and discuss some of the concerns in the design of harmonic generation systems. In many cases, harmonic generation can be modeled accurately by analytic or numerical techniques. Some of the theoretical development that is the basis for this modeling will be presented along with examples. There are several essential considerations that must be made in selecting a harmonic generation process. These involve both material properties and conditioning of the fundamental radiation.

General Considerations

The nonlinear optical material must be highly transmitting at the fundamental and harmonic wavelengths and must have a laser-induced-damage threshold high enough to allow optical intensities that will provide adequate conversion efficiency. Phase matching must be possible in the material. Other material characteristics must be considered for specific applications. The harmonic generation process requires that the phase velocities of the fundamental and harmonic radiation be matched. Without phase matching, power converted to the harmonic is back converted to the fundamental in a rapid oscillation, and harmonic conversion remains at low levels. The material also needs to be of good optical quality to provide optical transmission without distortion, which would destroy the phase relationship. Continuous-wave harmonic generation may involve local intensities of the order of 1 megawatt per square cm (1 MW/cm^2). Pulsed applications with Q-switched laser output of a few nanoseconds duration typically employ intensities of several hundred MW/cm^2, and harmonic-generation applications with mode-locked lasers may involve pulse durations of tens of femtoseconds with peak intensities of tens of gigawatts per square cm. Many other properties are desirable, but it is necessary to work within the limitations of available materials.

Phase matching is an essential requirement. When harmonic and fundamental waves of different eigen polarizations are coupled through the nonlinear optical tensor, it may be possible to use crystal birefringence to provide phase matching. The birefringence, if large enough, can compensate dispersion. The fundamental wave is introduced into the crystal with the polarization of the larger refractive index, and the harmonic is generated in the polarization of the smaller index. This is a description of type-I birefringent phase matching. Phase matching is achieved by selecting the direction of propagation in the crystal or by changing the temperature of the crystal. Type-II phase matching with the fundamental wave resolved into the two eigen polarizations is possible with large birefringence. Quasi phase matching involves periodically structuring the nonlinear material in a way that spatial modulation of the optical nonlinearity compensates for dispersion. Quasi phase matching is tuned by selection of the modulation period, and temperature control may provide a fine tuning. Periodic poling, one method of quasi phase matching, involves changing the orientation of ferroelectric domains after each coherence length of the harmonic generation process. These methods of phase matching are sketched in Figure 1.

The physical arrangement of the harmonic generation process depends on properties of the available fundamental light and the properties of the selected nonlinear material. The single-pass techniques, shown in Figure 1, may have the advantage of simplicity, but expanding or focusing the fundamental beam may be required. Crystal length and clear

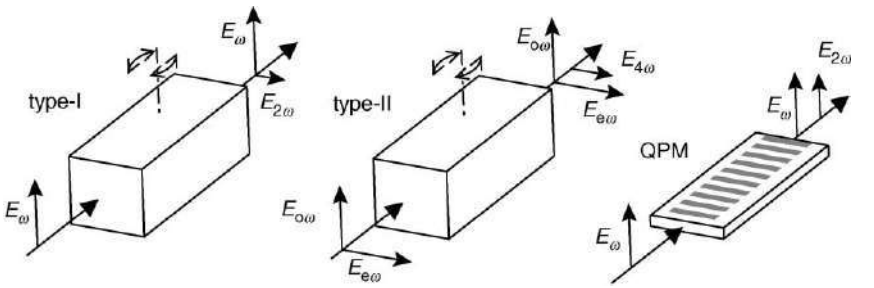

Figure 1 Different methods of phase matching optical harmonic generation are illustrated. The arrows indicate the directions of electric-field polarization and directions of propagation.

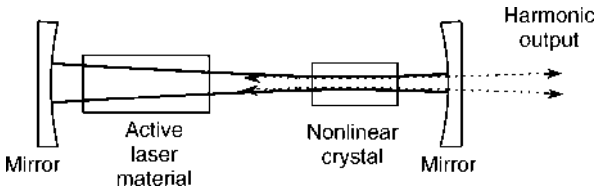

Figure 2 Placing a harmonic-generation crystal inside a laser cavity with mirrors highly reflecting at the laser wavelength is one technique of increasing the intensity of the fundamental beam.

aperture need to be selected. Intracavity harmonic generation (Figure 2), or external-resonant-cavity harmonic generation, may provide advantages in other cases. Analysis of harmonic generation provides the means to select among the options.

Basic Equations

Much of the analysis of second-harmonic generation follows from the differential equations that describe the interaction between the fundamental and harmonic fields, and it is appropriate to describe the development of these coupled equations. The wave equation:

$$\nabla^2\mathbf{E}(\mathbf{r},t) - \mu_0\sigma\frac{\partial\mathbf{E}(\mathbf{r},t)}{\partial t} - \mu_0\varepsilon_0\frac{\partial^2\mathbf{E}(\mathbf{r},t)}{\partial t^2} = \mu_0\frac{\partial^2\mathbf{P}(\mathbf{r},t)}{\partial t^2} \quad [3]$$

is obtained from Maxwell's equations. Here $\mu_0 = 4\pi\times10^{-7}$ is the permeability of free space, and σ is the conductivity, which later will be incorporated into the optical absorption coefficient $\alpha = \mu_0\sigma c/2$, and $(\varepsilon_0\mu_0)^{-1/2} = c \cong 2.998\times10^8$ m/s is the speed of light in vacuum. The electric field of a planewave, at angular frequency ω, is expressed in complex notation as the product of a slowly varying complex amplitude and a carrier wave:

$$\mathbf{E}_\omega(\mathbf{r},t) = (1/2)\{\mathbf{E}_\omega \exp i(\mathbf{k}_\omega\cdot\mathbf{r} - \omega t) + \text{c.c.}\} \quad [4]$$

The intensity of this plane wave is

$$I_\omega = (nc\varepsilon_0/2)|E_\omega|^2 \quad [5]$$

Here n is the index of refraction of the material in which E is specified. The nonlinear polarization resulting from the second-order susceptibility driven by the plane wave $\mathbf{E}_\omega(\mathbf{r},t)$ is expressed as

$$\mathbf{P}^{\text{NL}}_{2\omega}(\mathbf{r},t) = (1/2)\{\mathbf{P}^{\text{NL}}_{2\omega} \exp i(2\mathbf{k}_\omega\cdot\mathbf{r} - 2\omega t) + \text{c.c.}\} \quad [6]$$

The individual components of the polarization amplitude vector are

$$P^{\text{NL}}_{2\omega,i} = \varepsilon_0\sum_{j,k} d_{ijk}(-2\omega;\omega,\omega)E_{\omega,j}E_{\omega,k} \quad [7]$$

If the nonlinear polarization is produced by two distinguishable waves, such as for sum-frequency generation of the third harmonic by the second harmonic and the fundamental, the nonlinear polarization is

$$\mathbf{P}^{\text{NL}}_3(\mathbf{r},t) = (1/2)\{\mathbf{P}^{\text{NL}}_3 \exp i[(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{r} - (\omega_1+\omega_2)t] + \text{c.c.}\} \quad [8]$$

where

$$P^{\text{NL}}_{3,i} = 2\varepsilon_0\sum_{j,k} d_{ijk}(-\omega_3;\omega_1,\omega_2)E_{1,j}E_{2,k} \quad [9]$$

and

$$\omega_3 = \omega_1 + \omega_2 \quad [10]$$

A number of simplifications are used to arrive at the coupled equations for harmonic generation with monochromatic plane waves. All waves are assumed to be propagating collinearly in the z direction. Only the vector components involved in the

interaction are retained, and the nonlinear optical tensor is replaced by the scalar effective nonlinear coefficient d_{eff}. Finally, the slowly varying envelope approximation is invoked and the higher-order derivatives of the amplitudes are ignored, as are time derivatives of the envelopes. The wave equation simplifies to

$$\frac{\mathrm{d}E(z)}{\mathrm{d}z} + \alpha E(z) = \mathrm{i}\frac{\mu_0 c\omega}{2n}P(z) \qquad [11]$$

where n is the index of refraction.

With three distinguishable waves, as with sum-frequency generation and type-II second-harmonic generation, there are three coupled equations:

$$\frac{\mathrm{d}E_3(z)}{\mathrm{d}z} + \alpha_3 E_3(z) = \mathrm{i}\frac{\omega_3 d_{\text{eff}}}{n_3 c}E_1(z)E_2(z)\exp(-\mathrm{i}\Delta kz) \qquad [12]$$

$$\frac{\mathrm{d}E_1(z)}{\mathrm{d}z} + \alpha_1 E_1(z) = \mathrm{i}\frac{\omega_1 d_{\text{eff}}}{n_1 c}E_3(z)E_2^*(z)\exp(\mathrm{i}\Delta kz) \qquad [13]$$

$$\frac{\mathrm{d}E_2(z)}{\mathrm{d}z} + \alpha_2 E_2(z) = \mathrm{i}\frac{\omega_2 d_{\text{eff}}}{n_2 c}E_3(z)E_1^*(z)\exp(\mathrm{i}\Delta kz) \qquad [14]$$

Here the wave vector mismatch is

$$\Delta k = k_3 - k_1 - k_2 \qquad [15]$$

and

$$\omega_3 = \omega_1 + \omega_2 \qquad [16]$$

with the convention of $\omega_1 \leq \omega_2 < \omega_3$. The indices of refraction of the three waves are n_1, n_2, and n_3, respectively, and α_1, α_2, and α_3 are the respective amplitude absorption coefficients.

With type-I second-harmonic generation, the fundamental field consists of only a single eigen polarization, and there are two coupled equations:

$$\frac{\mathrm{d}E_{2\omega}(z)}{\mathrm{d}z} + \alpha_{2\omega}E_{2\omega}(z) = \mathrm{i}\frac{\omega d_{\text{eff}}}{n_{2\omega}c}E_\omega(z)E_\omega(z)\exp(-\mathrm{i}\Delta kz) \qquad [17]$$

$$\frac{\mathrm{d}E_\omega(z)}{\mathrm{d}z} + \alpha_\omega E_\omega(z) = \mathrm{i}\frac{\omega d_{\text{eff}}}{n_\omega c}E_\omega(z)E_\omega^*(z)\exp(\mathrm{i}\Delta kz) \qquad [18]$$

Indices of refraction at the harmonic and fundamental are $n_{2\omega}$ and n_ω respectively; $\alpha_{2\omega}$ and α_ω are the respective absorption coefficients; and the wavevector mismatch is

$$\Delta k = k_{2\omega} - 2k_\omega \qquad [19]$$

The equations for type-II second-harmonic generation, where $\omega_1 = \omega_2$ and $\omega_3 = 2\omega_1$, become equivalent for those for type-I second-harmonic generation, providing intensities of the two fundamental polarizations are equal, $I_1 = I_2$ and $\alpha_\omega = \alpha_{2\omega}$.

Steady-State Solutions

Three monochromatic planewave solutions for type-I second-harmonic generation are discussed. These solutions address type-I harmonic generation; first for low levels of conversion with arbitrary phase mismatch; next for high levels of conversion with perfect phase matching; and last for arbitrary levels of conversion with arbitrary amounts of phase mismatch. Near-field and focused-beam approximations also exist. Beyond that, numerical techniques, such as split-step Fourier transform methods can handle more complicated conditions. The simple approximations of monochromatic planewaves, or steady-state solutions, however, yield useful information.

The coupled equations for type-I second-harmonic generation can be solved by simple integration for low levels of fundamental depletion, where $E_\omega(z)$ is treated as constant and there is insignificant absorption. Initial conditions are $I_\omega(0)$, incident on the crystal and $I_{2\omega}(0) = 0$. Integration of the equation for the harmonic electric field yields an expression for electric field after propagating a distance of l in the crystal:

$$E_{2\omega}(l) = \frac{\omega d_{\text{eff}}E_\omega^2(0)}{n_{2\omega}c}\frac{1}{\Delta k}\left\{\exp\mathrm{i}(k_{2\omega}l) - \exp\mathrm{i}(2k_\omega l)\right\} \qquad [20]$$

Expressed as intensity the solution is

$$I_{2\omega}(l) = I_\omega(0)\Gamma^2 l^2\,\mathrm{sinc}^2(\Delta kl/2) \qquad [21]$$

where $\mathrm{sinc}(x) = \sin(x)/x$ and

$$\Gamma^2 = 2\omega^2 d_{\text{eff}}^2 I_\omega(0)/(n_{2\omega}n_\omega^2 c^3\varepsilon_0) \qquad [22]$$

When $\Delta k = 0$ and $\alpha_\omega = \alpha_{2\omega} = 0$, conditions of perfect phase matching and no absorption, the coupled equations for type-I second-harmonic generation can be combined into a single equation. This is done using conservation of energy, which requires $E_\omega^2(z) + E_{2\omega}^2(z) = E_\omega^2(0)$, in this case with no initial harmonic intensity. The equation can then be manipulated into a standard integral that yields the

inverse hyperbolic tangent. Rewritten and expressed in terms of intensity the result is

$$I_{2\omega}(l) = I_{\omega}(0)\tanh^2(\Gamma l) \qquad [23]$$

The two solutions for $I_{2\omega}(l)$ given above serve well for a number of applications. The $\mathrm{sinc}^2(\Delta k \cdot l/2)$ function is widely used to determine phase-matching acceptances. The $\tanh^2(\Gamma l)$ function is a good approximation for local conversion with large, well-collimated fundamental beams.

An exact solution to the coupled equations, however, gives insight to conditions such as high initial fundamental intensity with small amounts of phase mismatch. After extensive algebraic manipulation, a solution in the form of a Jacobian elliptic function $\mathrm{sn}(u|m)$ is obtained:

$$I_{2\omega}(l) = I_{\omega}(0)\nu_{\mathrm{b}}^2 \mathrm{sn}^2\left(\frac{\Gamma l}{\nu_{\mathrm{b}}}\middle|\nu_{\mathrm{b}}^4\right) \qquad [24]$$

where

$$1/\nu_{\mathrm{b}} = \Delta s/4 + \sqrt{1 + (\Delta s/4)^2} \qquad [25]$$

$$\Delta s = \Delta k/\Gamma$$

and Δk and Γ retain the earlier definitions. This solution can be extended to provide the phase of the fundamental and harmonic as the conversion process evolves. Expressing the complex harmonic and fundamental field amplitudes as the product of a real amplitude and a complex exponential:

$$E(z) = A(z)\exp \mathrm{i}\phi(z) \qquad [26]$$

the harmonic phase is

$$\phi_{2\omega}(z) = (\pi - \Delta k z)/2 \qquad [27]$$

The fundamental phase, however, is found to be dependent on the conversion process:

$$\phi_{\omega}(z) = (1/2)\arccos[A_{2\omega}(z)\Delta k/(2\kappa A_{\omega}^2(z))] - (\pi - \Delta k z)/z \qquad [28]$$

where $\kappa = \omega d_{\mathrm{eff}}/(n_{\omega} c)$. Small amounts of phase mismatch significantly change the phase of the fundamental at high levels of conversion. This phase change is used for switching applications in waveguide interferometers. It also makes phase-matching requirements more stringent at high levels of conversion. The three planewave solutions are illustrated in Figure 3. The narrowing of phase matching at higher conversion is illustrated in Figure 4.

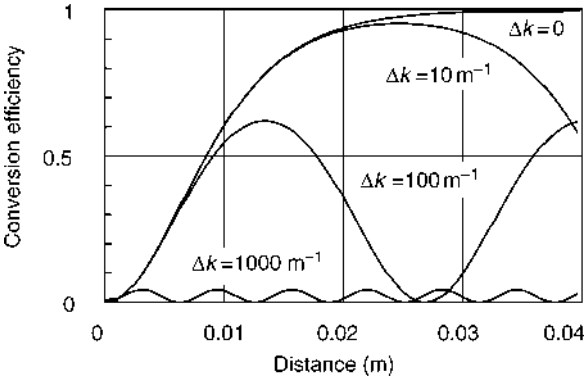

Figure 3 Monochromatic plane-wave solutions of the coupled equations for second-harmonic generation are illustrated for different amounts of wavevector mismatch. The solution takes the form $\tanh^2(\Gamma l)$ when $\Delta k = 0$ and $(\Gamma l)^2\mathrm{sinc}^2(\Delta k l/2)$ when conversion is small.

There is a Jacobian elliptic function solution for intensities with the coupled waves of type-II second-harmonic generation and sum-frequency generation. The phases of the individual waves can be recovered with elliptic integrals of the third kind. Both sets of equations, the two coupled equations for type-I phase matching and the three coupled equations for type-II phase matching, are amenable to numerical solution by Runge–Kutta methods.

Phase Matching

Phase matching is achieved in birefringent materials by balancing the dispersion of the material with the difference between the indices of refraction of the fast and slow waves. With adequate birefringence it may be possible to obtain phase matching by selecting the direction of propagation in the crystal. Equations describing the dispersion and birefringence of nonlinear crystals, necessary for calculating phase matching, along with the nonlinear optical coefficients and expressions for the effective nonlinear optical coefficients, are tabulated in handbooks. Values for a few examples are given here.

The negative uniaxial crystal potassium dihydrogen phosphate (KH_2PO_4 or KDP) is used as an illustrative example. The harmonic wave is chosen to have extraordinary polarization with the smaller index of refraction in this crystal with negative birefringence. With type-I phase matching, the fundamental consists of only an ordinary polarized wave. To have phase matching, $\Delta k = 0$, it is necessary to choose a direction of propagation that makes an angle θ with the optic axis that satisfies:

$$n_{\mathrm{e},2\omega}(\theta) = n_{\mathrm{o},\omega} \qquad [29]$$

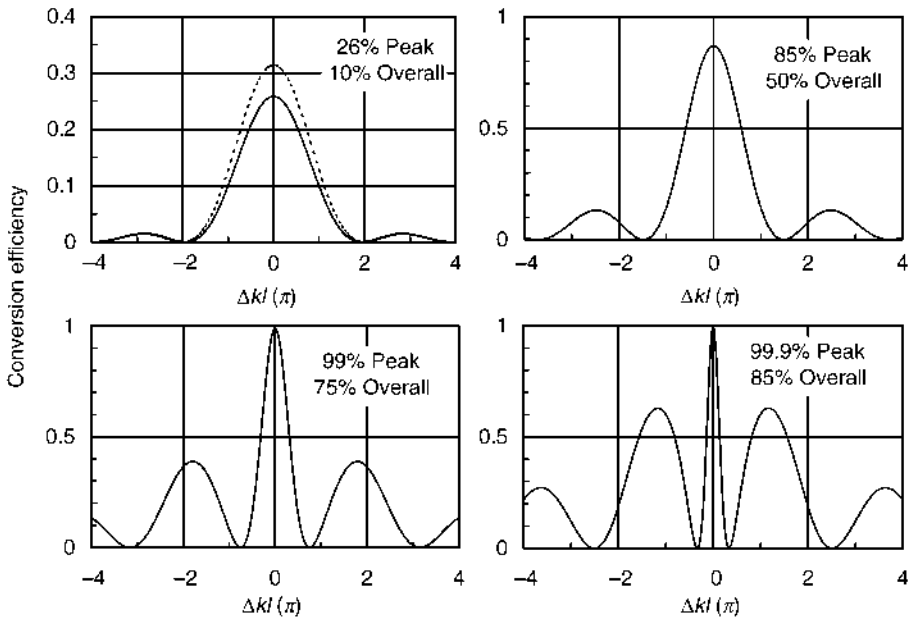

Figure 4 High levels of overall second-harmonic conversion require higher levels of conversion at peak pulse intensity. Phase-matching bandwidths narrow at the higher levels of peak conversion. The broken curve is the $(\Gamma l)^2\mathrm{sinc}^2(\Delta kl/2)$ small conversion approximation.

The extraordinary index for propagation at an angle of θ to the optical axis in a uniaxial crystal $n_e(\theta)$ is given by

$$\frac{1}{n_e^2(\theta)} = \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2} \qquad [30]$$

where n_o is the ordinary index and n_e is the extraordinary index for propagation at 90° to the optic axis. The phase-matching angle is obtained by using the identity $\cos^2\theta = 1 - \sin^2\theta$ and solving for θ:

$$\theta_{PM} = \sin^{-1}\left\{\frac{n_{e,2\omega}^2(n_{o,2\omega}^2 - n_{o,\omega}^2)}{n_{o,\omega}^2(n_{o,2\omega}^2 - n_{e,\omega}^2)}\right\}^{1/2} \qquad [31]$$

where the subscripts refer to ordinary and extraordinary indices at the fundamental and second-harmonic frequencies.

The effective nonlinear coefficient is determined by crystal symmetry, the type of phase matching, the birefringence of the crystal, and the direction of propagation. For crystals of point group $\bar{4}2m$ with negative birefringence, such as KDP, the effective nonlinear coefficient for type-I phase matching is given by

$$d_{eff} = -d_{36}\sin\theta\sin 2\phi \qquad [32]$$

The azimuthal angle ϕ is measured from the crystal x-axis. The x-axis and y-axis are specified by X-ray diffraction and piezoelectric coefficients and can be determined from the growth morphology. The azimuthal orientation of a finished crystal, however, is not obvious, and requires X-ray orientation, piezoelectric characterization, or nonlinear optical measurements, if the record of crystal orientation is lost. The nonlinear coefficient d_{36} is an abbreviation for the d_{312} and d_{321} tensor components, which are equal. For harmonic generation with a 1064 nm fundamental in KDP, $d_{36} = 0.39 \times 10^{-12}$ m/V. The unit of 10^{-12} m/V is usually expressed as picometers per volt (pm/V). The effective nonlinear coefficient for type-II phase matching for this class of crystals is

$$d_{eff} = d_{36}\sin 2\theta\cos 2\phi \qquad [33]$$

Type-I and type-II crystals cut from the same material are not interchangeable. The type-II phase-matching angle for KDP can be found by iterative solution of the equation:

$$n_{e,2\omega}(\theta) = \{n_{o,\omega} + n_{e,\omega}(\theta)\}/2 \qquad [34]$$

The birefringent walkoff angle, that is the angle the Poynting vector of an extraordinary wave makes with the wavevector, is another consideration for birefringent phase matching. There will be no further harmonic generation when the extraordinary and ordinary fundamental beams of type-II phase matching no longer overlap. For type-I phase matching, birefringent walkoff can limit the interaction length, but there can be several interaction lengths

in a crystal. The birefringent walkoff of an extraordinary wave propagating at an angle θ to the optic axis in a uniaxial crystal is given by

$$\rho = \tan^{-1}\left\{\frac{n_e^2(\theta)}{2}\left[\frac{1}{n_e^2} - \frac{1}{n_o^2}\right]\sin(2\theta)\right\} \qquad [35]$$

Refractive indices for 1064 nm to 532 nm second-harmonic generation in KDP are $n_{e,2\omega} = 1.4709$, $n_{o,2\omega} = 1.5129, n_{e,\omega} = 1.4603$, and $n_{o,\omega} = 1.4944$. The dispersion of KDP and indices at the phase-matching angles are shown in Figure 5. For type-I phase matching $d_{eff} = 0.26$ pm/V, $\theta_{PM} = 41.2°$, and $\rho_{2\omega} = 28$ milliradians (mr), and for type-II phase matching $d_{eff} = 0.34$ pm/V, $\theta_{PM} = 59.1°, \rho_{\omega} = 20$ mr, and $\rho_{2\omega} = 25$ mr. Type-II phase matching in KDP for this application offers an advantage of larger effective nonlinear coefficient. Also the angular acceptance is larger for type-II phase matching.

The angular acceptance is calculated from dispersion equations by first evaluating the derivative $\partial\Delta k/\partial\theta$ and then finding the value of $\Delta\theta$ corresponding to the full width at half maximum (FWHM) of the $\text{sinc}^2(\Delta k \cdot l/2)$ function:

$$\frac{\partial\Delta k}{\partial\theta}\Delta\theta_{FWHM} \cong 4 \times 1.39 \text{ radians}/l \qquad [36]$$

When θ is 0° or 90° for uniaxial crystals, or for propagation along one of the refractive index principal axes in a biaxial crystals, the first derivative $\partial\Delta k/\partial\theta$ becomes zero, and the angular acceptance must be calculated from the second derivative $\partial^2\Delta k/\partial\theta^2$. This case is called noncritical phase matching (NCPM), and the case for θ between 0° and 90° is called critical phase matching.

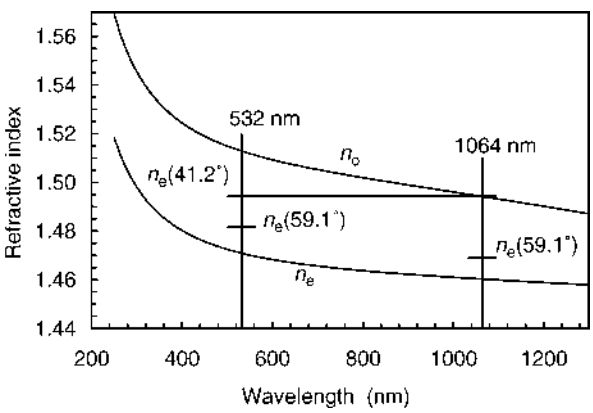

Figure 5 Dispersion of the birefringent nonlinear material KH_2PO_4 with refractive indices indicated for type-I and type-II 1064 nm to 532 nm second-harmonic generation. (Adapted from Byer RL (1977) Parametric oscillators and nonlinear materials. In: Harper PG and Wherrett BS (eds) *Nonlinear Optics*, pp. 47–159. San Francisco, CA: Academic Press.)

Second-harmonic generation of 1064 nm fundamental radiation in the crystal lithium triborate (LBO or LiB_3O_5) is an example of NCPM. At 149°C, this crystal will phase match for type-I SHG for propagation along the x-axis in the xyz frame determined by the indices of refraction of this biaxial crystal $n_x < n_y < n_z$. The following are calculated from dispersion equations for this case: $\partial\Delta k/\partial\theta = 0$; $\partial^2\Delta k/\partial\theta^2 = 2.04 \times 10^4$ (rad/cm)/rad^2; $\partial\Delta k/\partial\phi = 0$; and $\partial^2\Delta k/\partial\phi^2 = 6.45 \times 10^3$ (rad/cm)/rad^2. The FWHM angular acceptances for a 1.5 cm-long LBO crystal are $\Delta\theta_{FWHM} = 1.09°$ and $\Delta\phi_{FWHM} = 1.94°$. Angle-tuned CPM at 23°C in LBO occurs at $\theta = 90°$ and $\phi = 11°.5$. For this case $\partial\Delta k/\partial\phi = 1.32 \times 10^4$ (rad/cm)/rad, and the azimuthal angular acceptance decreases to $\Delta\phi_{FWHM} = 0.16°$. Also there is a birefringent walkoff angle of $\rho_{2\omega} = 19$ mr, which results in a displacement of 0.29 mm of the extraordinary-like harmonic beam for a 1.5 cm-long crystal. Both the larger angular acceptance and the absence of birefringent walkoff are advantages for harmonic generation. The effective nonlinear coefficient of LBO is reported as 0.85 pm/V and 1.04 pm/V by different sources.

Quasi phase matching offers the possibility, in principle, of noncritical phase matching at any wavelength in the transmission range of a material. However, it may be difficult producing the periodic structure that is required for phase matching in a particular application. The ferroelectric crystals lithium niobate ($LiNbO_3$), lithium tantalate ($LiTaO_3$), and potassium titanyl phosphate ($KTiOPO_4$ or KTP) can be processed by electric-field poling to produce domains of alternating polarity. Lithography techniques are used to apply a patterned electrode structure on the surface of a wafer or thin slab. Domains are reversed by the application of a carefully monitored and controlled electric field. Thicknesses of several mm have been achieved in periodically poled KTP (PPKTP). The higher coercive field of 21 kV/mm has limited the thicknesses that can be achieved in periodically poled lithium niobate (PPLN), particularly at shorter wavelengths for which poling periods become small. Progress is now being made with zincblende materials such as gallium arsenide. These materials have large linearity and good transmission in the infrared, but they are not birefringent. Epitaxial growth on patterned substrates is now being used to produce periodically structured GaAs on a research basis, and bonded stacks of wafers of alternating orientation also have been used. Harmonic conversion with quasi phase matching is compared with conventional phase matching and nonphase-matched harmonic generation in Figure 6.

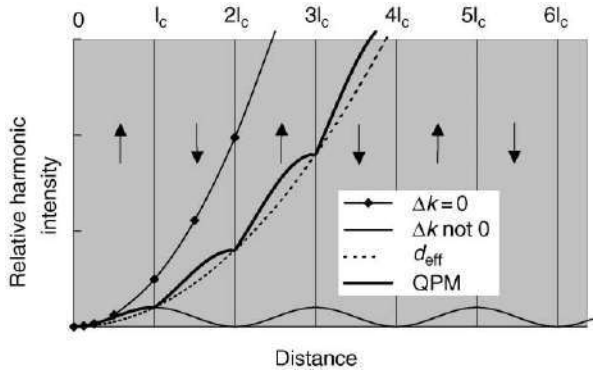

Figure 6 The growth of harmonic intensity is illustrated schematically for a quasi-phase-matched process in which the polarity of the nonlinearity is reversed after each coherence length. Conversion for $\Delta k = 0$, $\Delta k \neq 0$, and a uniform effective nonlinear coefficient are shown for comparison. (Adapted from Fejer MM (1992) Quasi-phase-matched second harmonic generation: tuning and tolerances. *IEEE Journal of Quantum Electronics* 28: 2631–2654.)

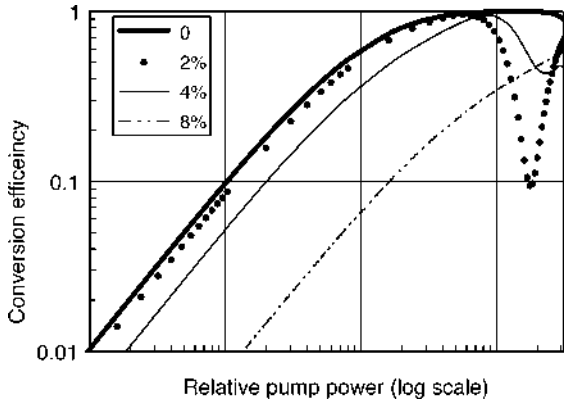

Figure 7 Random variations in domain length will lower harmonic conversion efficiency and have effect on back conversion. RMS variations in domain length are indicated in the legend for this specific example.

Quasi phase matching also provides access to tensor components that are not accessible with birefringent phase matching. This is the case for lithium niobate; the largest tensor component is $d_{333} = d_{33}$, which has a value of 25 pm/V. The effective nonlinear coefficient is the relevant Fourier component of the modulated nonlinearity:

$$d_{\text{eff}} = (2/\pi)d_{33} \approx 16 \text{ pm/V} \qquad [37]$$

This is a larger effective nonlinearity than is available with common birefringently phase matched materials in the visible and near-infrared. A more accurate treatment of second-harmonic generation involves application of the Jacobian-elliptic-function solution to each domain. After nearly complete conversion is reached, the phase perturbations on the individual domains have effect, and back conversion takes over (Figure 7). Conversion is lower and back conversion can set in sooner with random differences in the domain length. There is some experimental evidence that this type of back conversion has been observed in guided-wave harmonic conversion devices. Guided-wave nonlinear optical interaction can have long interaction lengths not limited by diffraction. Quasi phase matching is particularly useful in waveguide second-harmonic generation because the direction of propagation is fixed.

Gaussian Beams and Pulses

A radially symmetric Gaussian light beam has transverse intensity of

$$I(r) = I_0 \exp(-2r^2/w^2) \qquad [38]$$

where r is the radial coordinate and w is called the spot size. Harmonic generation in the low-conversion limit by a Gaussian beam, with a focus of spot size w_0 at the center of a crystal with optimized phase matching, is described by

$$p_{2\omega}(l)/p_\omega(0) = 2\omega^2 d_{\text{eff}}^2 p_\omega(0) lkh(B,\xi)/(\pi n^3 \varepsilon_0 c^3) \qquad [39]$$

The harmonic power after propagating distance l in the crystal is $p_{2\omega}(l)$, and $p_\omega(0)$ is the incident fundamental power. The function $h(B,\xi)$ is the Boyd and Kleinman focusing factor. It is a function of the walkoff parameter:

$$B = \rho\sqrt{lk}/2 \qquad [40]$$

and the focusing parameter:

$$\xi = l/(kw_0^2) \qquad [41]$$

The spot size w_0 of the fundamental beam at its waist is used to described the fundamental intensity distribution at the focus at the center of the crystal:

$$I_\omega(r) = (2p_\omega(0)/(\pi w_0^2))\exp(-2r^2/w_0^2) \qquad [42]$$

Values of the focusing factor for arbitrary focusing and walkoff are usually displayed in the form of a graph (Figure 8). In the case of weak focusing $\xi \ll 1$ and relatively small walkoff $w_0 > \rho l$, the focusing factor is given by

$$h(B,\xi) \approx \xi(1 - t^2/12 + t^4/120 - t^6/1344 + \ldots) \qquad [43]$$

where $t = 2B(2\xi)^{1/2}$.

An example of continuous-wave second harmonic generation is used. This calculation models a conversion experiment published in the literature. External-resonant-cavity second-harmonic generation of 6.5 W of 532 nm radiation was reported. The 1064 nm fundamental radiation with 378 W of circulating power in a resonant cavity external to the laser was focused to a 32 μm spot size in the center of a 6 mm-long LBO crystal (Figure 9). This corresponds to a focusing factor of $\xi = 0.62$ in the index $n = 1.6053$ material. The value of the focusing parameter is $h(0, 0.62) = 0.58$. The nonlinear optical coefficient for this process is reported as 0.85 pm/V and 1.04 pm/V by different sources. When these values are inserted into the earlier equation for $p_{2\omega}(l)/p_{\omega}(0)$, the higher coefficient yields a conversion efficiency of 0.0273 and 10.3 W of 532 nm output, whereas the lower coefficient yields a conversion efficiency of 0.0182 and 6.9 W of 532 nm radiation. Differences in published values of nonlinear optical coefficients are common. The reported precision of the experiment and the measured values tend to support the lower value. Practical applications are useful to check the accuracy of parameter values. With many variables and the added complication of operation at high intensity there is uncertainty that demands caution in accepting the derived parameter values. It is interesting to note that the power and beam size of the example result in a maximum intensity of 17 MW/cm^2 at the surface of this crystal. Many vendors will warranty antireflection coatings for only 1 MW/cm^2 maximum continuous-wave intensity.

For a pulsed fundamental with temporal shape $p_{\omega}(t) = p_{\omega}(0)\exp(-t^2/\tau^2)$, integration over time gives a calculated energy conversion efficiency of:

$$\eta_{\text{calc}} = U_{2\omega}(l)/U_{\omega}(0)$$
$$= \sqrt{2}\omega^2 d_{\text{eff}}^2 U_{\omega}(0) l k h(B, \xi)/(\tau\sqrt{\pi^3} n^3 \varepsilon_0 c^3) \quad [44]$$

where $U_{2\omega}(l)$ is the harmonic energy generated after propagation distance l in the crystal and $U_{\omega}(0)$ is the incident fundamental energy. This is still a low conversion approximation. An estimate of conversion efficient at arbitrary levels of depletion η_{est} is obtained with the empirical expression:

$$\eta_{\text{est}} = \eta_{\text{calc}}/(1 + \eta_{\text{calc}}) \quad [45]$$

Returning to the example 1.064 nm to 532 nm, type-II SHG in KDP and using $l = 3$ cm, $U_{\omega}(0) = 0.3$ J, and $\tau = 3.6$ ns for a 6 ns FWHM pulse duration, the above two relationships predict a conversion efficiency of 50%. These parameters give a peak intensity of 330 MW/cm^2, a value typical for efficient harmonic conversion, while avoiding laser-induced damage in a long-term, repetitive-pulse application using KDP. Single-shot laser-induced-damage thresholds for KDP are about 10 times higher.

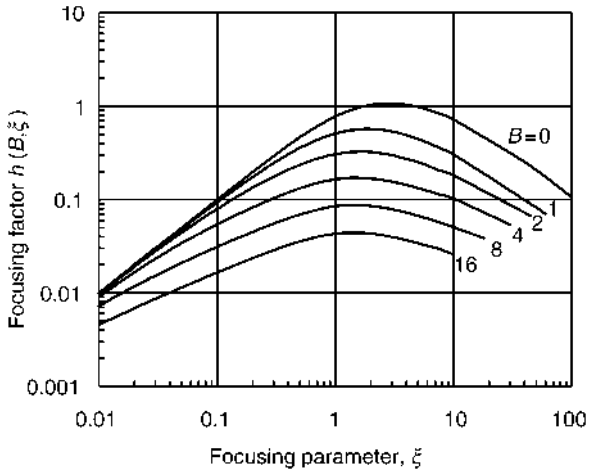

Figure 8 The Boyd and Kleinman focusing factor $h(B,\xi)$ for type-I second-harmonic generation is shown as a function of the focusing parameter ξ for different values of the walkoff parameter B. The curves shown were generated with a split-step Fourier transform calculation. (Adapted from Boyd GD (1968) Parametric interactions of focused Gaussian light beams. *Journal of Applied Physics* 39: 3597–3639.)

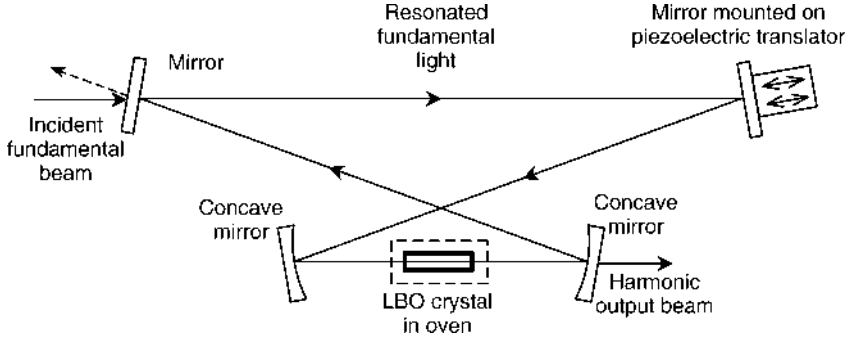

Figure 9 A bow-tie cavity for externally resonant second-harmonic generation is illustrated. The reflected fundamental beam is used to develop an error signal to drive the piezoelectric translator and hold the cavity in resonance.

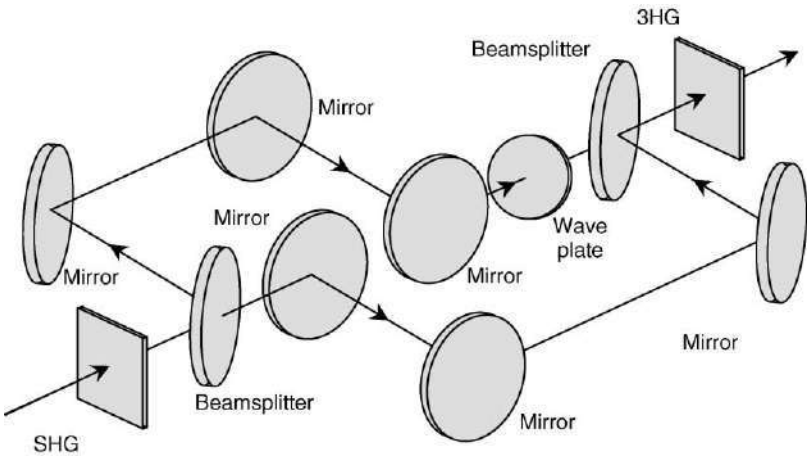

Figure 10 Short crystals less than 1 mm in length are typically used for second-harmonic generation with fundamental pulses that are only ten's of femtoseconds in duration. It is necessary to separate fundamental and second-harmonic pulses and recombine them with appropriate delay to compensate for group-velocity mismatch before sum-frequency generation to the third harmonic.

As a final example, consider second-harmonic generation of a 100 fs ($1\ \text{fs} = 10^{-15}$ s) duration pulse from a mode-locked Ti:Al_2O_3 laser operating at 800 nm. Typical laser output might consist of 10 nJ (10×10^{-9} Joule) pulses at 80 MHz repetition rate. Group velocity walkoff is a concern for this short pulse duration (Figure 10). Barium borate (BaB_2O_4 or BBO) is a common crystal for this application. The group velocity walkoff rate between the ordinary-polarization 800 nm fundamental pulse and the extraordinary-polarization 400 nm harmonic pulse is 190 fs/cm. A crystal length no longer than 0.5 mm is needed to avoid lengthening the harmonic pulse in time. Fortunately, the laser-induced-damage-threshold intensity increases approximately as the inverse of the square root of the pulse length between roughly 10 ns and 1 ps, and it is possible to focus the beam to achieve higher intensity without crystal damage and compensate for the short crystal length. However, the 68-milliradian birefringent walkoff of the harmonic will limit the size of the focus. A 20 μm spot size will allow an interaction length that roughly matches the crystal length and provides a peak fundamental intensity around 16 GW/cm^2 and a conversion efficiency of nearly 30%. Now suppose the 100 fs pulse is amplified to 100 mJ in a large chirped-pulse regenerative amplifier. A larger diameter crystal is required, and KDP is an appropriate choice. The group velocity walkoff rate in KDP is 78 fs/mm, and a crystal of 1.5 mm thickness could be used. With the beam spot size increased to 30 mm, the peak intensity will be 71 GW/cm^2, and the conversion efficiency will be about 40%. Material availability, such as the large sizes possible with KDP, often determines a choice.

Conclusion

Basic techniques for describing and modeling second-harmonic generation have been discussed. Birefringent phase matching and quasi phase matching were reviewed. Three examples were used to illustrate considerations for implementing harmonic-generation systems. There are many more parameters to be considered for other applications and for more detailed analysis. The examples presented illustrate the demands nonlinear optical frequency conversion places on materials. A few materials have been established as primary choices in common applications. But for many applications, available materials or laser system performance are limiting and it is necessary to use the materials and systems that are available. The importance of using the analysis methods is stressed. It is possible to sacrifice performance by improperly specifying a crystal. Choosing the best crystal for a harmonic-generation application with an appropriate optical design will optimize harmonic conversion.

See also

Nonlinear Optics, Applications: Phase Matching. **Nonlinear Sources:** Harmonic Generation in Gases. **Spectroscopy:** Second Harmonic Spectroscopy.

Further Reading

Akhmanov SA (1975) Optical harmonic generation and optical frequency multipliers. In: Rabin H and Tang CL (eds) *Quantum Electronics. A Treatise, vol. 1, part B*, pp. 475–583. New York: Academic Press.

Armstrong JA (1962) Interaction between lightwaves in a nonlinear dielectric. *Physics Review* 127: 1918–1962.
Boyd GD (1968) Parametric interactions of focused Gaussian light beams. *Journal of Applied Physics* 39: 3597–3639.
Boyd RW (2003) *Nonlinear Optics*, 2nd edn. New York: Academic Press.
Byer RL (1977) Parametric oscillators and nonlinear materials. In: Harper PG and Wherrett BS (eds) *Nonlinear Optics*, pp. 47–159. San Francisco, CA: Academic Press.
Dmitriev VG (1997) *Handbook of Nonlinear Optical Crystals*, 2nd edn. Berlin: Springer-Verlag.
Fejer MM (1992) Quasi-phase-matched second harmonic generation: tuning and tolerances. *IEEE Journal of Quantum Electronics* 28: 2631–2654.
Shen YR (1984) *The Principles of Nonlinear Optics*. New York: John Wiley & Sons.
Sutherland RL (2003) *Handbook of Nonlinear Optics*, 2nd edn. New York: Marcel Dekker.
Yang SL (1991) 6.5 W, 532 nm radiation by cw resonant external-cavity second-harmonic generation of an 18 W Nd:YAG laser in LiB_3O_5. *Optical Letters* 16: 1493–1495.
Yariv A (1989) *Quantum Electronics*, 3rd edn. New York: John Wiley & Sons.

$\chi^{(3)}$–Third-Harmonic Generation

B Y Soon, Thales Electro-Optics PTE Ltd., Singapore

J W Haus, University of Dayton, Dayton, OH, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The invention of the ruby laser, in 1960, provided a revolutionary new light source, which emits an intense, highly collimated polarized beam of light. The new source opened the way for developments that have enabled new technologies, thus profoundly changing our way of life. One of these new frontiers opened by the laser was the field of nonlinear optics. In 1961, Franken *et al.* were the first to report the observation of second-harmonic generation from a quartz crystal, using ruby laser light. Third-harmonic generation (THG) was soon discovered by Terhune and co-workers in 1962.

THG has found applications as a spectroscopic tool to probe optical properties of materials, but recently it has been developed as a microscopy tool and a pulse measurement technique. More efficient THG is possible by engineering the materials; for instance, one way to improve conversion efficiency is to use periodic structured dielectric materials, called photonic bandgap structures.

Polarization and the Third-Order Susceptibility

THG is a nonlinear optical phenomenon, where light of angular frequency ω is converted into light at an angular frequency 3ω that is three times higher. Light is made of particles called photons and the nonlinear medium enables the process whereby three photons of frequency ω are annihilated to create one photon of frequency 3ω. The nonlinear interactions are strongly dependent on the polarization or the dipole moment per unit volume, which is written as a series

$$P(t) = \chi^{(1)} \times E(t) + \chi^{(2)} : E^2(t) + \chi^{(3)} \vdots E^3(t) + \cdots \quad [1]$$

where χ is the susceptibility of the optical material, the superscript of χ in parentheses denotes the order of nonlinearity, t is the time variable, and E is the electric field. Both the electric field and the polarization are vectors; therefore the susceptibilities are tensors. The first-order or linear susceptibility $\chi_{ij}^{(1)}$ is a second-order tensor. This term constitutes the linear polarization of the medium denoted a P^{L}. It has nine elements, but only up to three independent elements depending on the symmetry of the lattice. The other terms represent an expansion for the nonlinear polarization of the medium, P^{NL}. Note that the superscripts 'L' and 'NL' denote linear and nonlinear polarization contributions, respectively. The third-order susceptibility $\chi_{ijkl}^{(3)}$ is a four-rank tensor, which usually has 81 elements ($3 \times 3 \times 3 \times 3 = 81$). In asymmetric crystalline solids, all 81 elements can be of nonzero value and independent. With a higher degree of symmetry in other optical materials, the 81 elements are no longer all independent, and thus the number of independent elements can be reduced to a lower number. For example, isotropic material such as glass, liquid, or vapor has only three independent elements. For illustration, we consider a general case where the applied frequencies ω are arbitrary, and the third-susceptibility is given as

$$\chi_{ijkl} \equiv \chi_{ijkl}^{(3)}(\omega_4 = \omega_1 + \omega_2 + \omega_3) \quad [2]$$

Note that the superscript has been removed. In the isotropic material, the coordinate axes must be

equivalent. Thus, the susceptibility has symmetry given by

$$\chi_{1111} = \chi_{2222} = \chi_{3333} \quad [3a]$$

$$\chi_{1122} = \chi_{1133} = \chi_{2211} = \chi_{2233} = \chi_{3311} = \chi_{3322} \quad [3b]$$

$$\chi_{1212} = \chi_{1313} = \chi_{2323} = \chi_{2121} = \chi_{3131} = \chi_{3232} \quad [3c]$$

$$\chi_{1221} = \chi_{1331} = \chi_{2112} = \chi_{2332} = \chi_{3113} = \chi_{3223} \quad [3d]$$

For the isotropic material, there are only 21 nonzero elements, as given in eqns [3a]–[3d], because these elements have a Cartesian index (1, 2, or 3) that appears an even number of times. The reason for nonexistence of the elements with an index appearing an odd number of times is because, for example, χ_{1222} would give a response in the $\hat{x}_1$ direction due to a field applied in the $\hat{x}_2$ direction. This response must vanish in an isotropic material, because there is no reason why the response should be in the $+\hat{x}_1$ direction rather than in the $-\hat{x}_1$ direction. In summary, the elements in the third-order susceptibility of the isotropic material have the relationship given by

$$\chi_{ijkl} = \chi_{1122}\delta_{ij}\delta_{kl} + \chi_{1212}\delta_{ik}\delta_{jl} + \chi_{1221}\delta_{il}\delta_{jk} \quad [4]$$

where δ is the Kronecker delta function; it has a unit value, $\delta_{ij} = 1$, if and only if $i = j$, otherwise, it is zero, $\delta_{ij} = 0$. Thus, we see from eqn [4] that there are only three independent elements in the isotropic material. Crystalline materials have more independent components and they have been determined for the 32 crystals classes. The relationships between the elements for these crystalline classes are tabulated in books by Boyd and by Butcher and Cotter.

A block diagram of a material illustrates THG and its corresponding energy-level diagram is depicted in Figure 1. Figure 1a shows that fields of frequency ω excite the χ^3 optical material to generate the third-harmonic field with the residual of the fundamental harmonic field. Figure 1b shows that three photons of frequency ω are destroyed to generate a photon of frequency 3ω in the third-harmonic generation.

For third-harmonic generation, the frequency dependence of the susceptibility is expressed as $\chi_{ijkl}(3\omega = \omega + \omega + \omega)$. With the intrinsic permutation symmetry of the nonlinear susceptibility, the elements of the susceptibility tensor has a relationship such that $\chi_{1122} = \chi_{1212} = \chi_{1221}$. So we can further reduce eqn [4] for the third-harmonic generation as

$$\chi_{ijkl}(3\omega = \omega + \omega + \omega) = \chi_{1122}(3\omega = \omega + \omega + \omega) \times (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \quad [5]$$

Therefore, we see that the third-harmonic susceptibility tensor has only one independent element.

Wave Equation for χ^3 Third-Harmonic Generation

Theoretically, we have to use the wave equation to observe how an intense laser field interacts with the nonlinear material, so that the polarization of the medium could develop new frequency components that are not present in the incident field. Thus, these new frequency components of the polarization become the sources of the new frequency components of the electromagnetic field.

To formally describe THG, we begin with Maxwell's equations (in Gaussian units), and we have

$$\nabla \times D = 4\pi\rho \quad [6a]$$

$$\nabla \times B = 0 \quad [6b]$$

$$\nabla \times E = \frac{-1}{c}\frac{\partial B}{\partial t} \quad [6c]$$

$$\nabla \times H = \frac{-1}{c}\frac{\partial D}{\partial t} + \frac{4\pi}{c}J \quad [6d]$$

where H is the magnetic field, B is the magnetic flux density, D is the electric flux density, ρ is the charge density, and J is the current density. The desired solution is in the region of space in the nonlinear material that has no free charge ($\rho = 0$), and therefore, there is no free current ($J = 0$). We assume

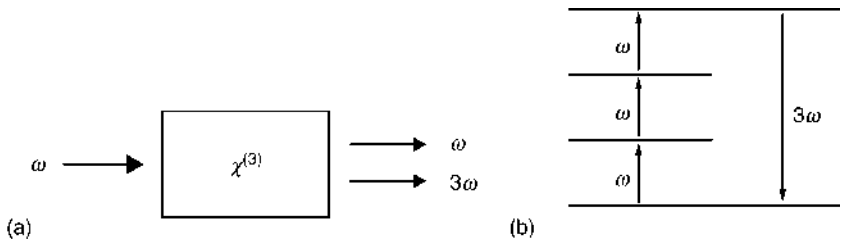

Figure 1 The third-harmonic (a) block diagram; and (b) energy level description.

that the nonlinear material is nonmagnetic, so that $B = \mu H = H$, where $\mu = 1$ in a nonmagnetic environment. In nonlinear material, the electric flux is defined as

$$D = E + 4\pi P \qquad [7]$$

so that the polarization depends nonlinearly on the strength of E. Using eqns [6c] and [6d], we have

$$\nabla \times \nabla \times E + \frac{1}{c^2}\frac{\partial^2 D}{\partial t^2} = 0 \qquad [8]$$

Substituting eqn [7] into [8], we have

$$\nabla \times \nabla \times E + \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = \frac{-4\pi}{c^2}\frac{\partial^2 P}{\partial t^2} \qquad [9]$$

Using an identity from the vector calculus, we have

$$\nabla \times \nabla \times E = \nabla(\nabla \times E) - \nabla^2 E \qquad [10]$$

In linear optics, we see from eqn [6a] that when there is no free charge in source-free isotropic material, $\nabla \times D = \nabla \times E = 0$. Thus, the first term on the right-hand side of eqn [10] has no effect. Whereas in nonlinear optics, $\nabla \times E \neq 0$ in isotropic material, but the contribution from the first term on the right-hand side of eqn [10] is negligible, therefore, we can ignore it. Thus, eqn [9] can be simplified as

$$\nabla^2 E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = \frac{4\pi}{c^2}\frac{\partial^2 P}{\partial t^2} \qquad [11]$$

For distinction of the linear and nonlinear contributions, we modified eqn [7] to get

$$D = E + 4\pi P^{\mathrm{L}} + 4\pi P^{\mathrm{NL}}$$

where $P = P^{\mathrm{L}} + P^{\mathrm{NL}}$. Let $D^{\mathrm{L}} = E + 4\pi P^{\mathrm{L}}$, and we rewrite eqn [11] as

$$\nabla^2 E - \frac{1}{c^2}\frac{\partial^2 D^{\mathrm{L}}}{\partial t^2} = \frac{4\pi}{c^2}\frac{\partial^2 P^{\mathrm{NL}}}{\partial t^2} \qquad [12]$$

Equation [12] is the most general form of the wave equation seen in nonlinear optics, and could be used as our starting point for our study of third-harmonic generation.

Methods of Third-Harmonic Generation

Since nonlinear optics is a popular research topic, there are a handful of existing methods that can be used for third-harmonic generation, and possibly new methods may be developed in the future. Here, we will discuss three methods using χ^3 material: (i) photonic crystal; (ii) defect mode using photonic crystal; and (iii) homogeneous nonlinear crystal.

Photonic Crystal

After the invention of electronic bandgap structure, the optical scientists and engineers saw the possibility to extend the operating waveband into the visible, and thus we have the photonic bandgap structure. The geometry and material choices of the photonic bandgap structure create a unique bandgap in frequency spectrum that has potential applications in laser optics, telecommunication, medical imaging, optical limiting, etc. A photonic bandgap structure is also commonly known as the photonic crystal. For a one-dimensional photonic crystal, two materials of certain thicknesses each, repeat themselves a number of times. For example, the first and second materials are of thickness *'a'* and *'b'*, respectively, so that their combined thickness is *'d'*, and have a periodicity of 10. The schematic of such a one-dimensional photonic crystal, with its corresponding dielectric function as a function of axial distance z, is depicted in Figure 2.

By using χ^3 materials for photonic crystal, and using method of multiple scales (MMS), we have the nonlinear polarization given as

$$P^{\mathrm{NL}} = \lambda \chi^{(3)} E^3 \qquad [13]$$

where λ is the perturbation parameter. With reference to Figure 2, assume that *'a'* is equal to *'b'* so that $d = 2a$, then the Fourier expansion of the dielectric

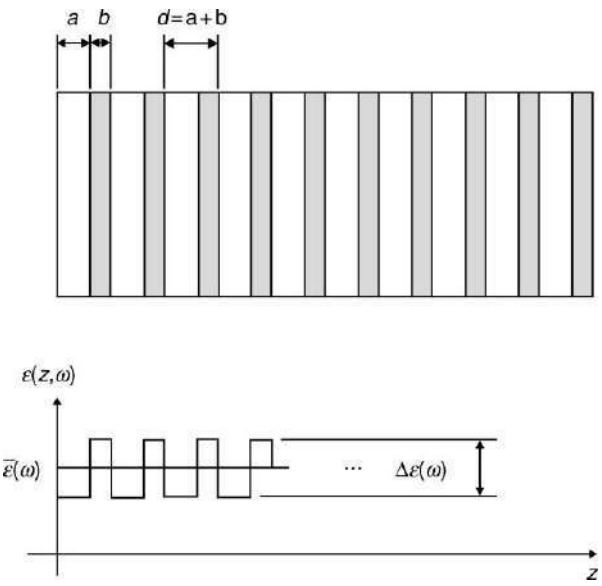

Figure 2 The schematic of the one-dimensional photonic crystal.

function is given as

$$\varepsilon(z,\omega) = \bar{\varepsilon}(\omega) + \lambda\Delta\varepsilon(\omega)\cos(2\pi z/d) - \frac{1}{3}\lambda\Delta\varepsilon(\omega)\cos(6\pi z/d) + \cdots \quad [14]$$

where $\bar{\varepsilon}(\omega)$ is the zeroth-order coefficient, and $\Delta\varepsilon(\omega)$ is first-order coefficient of the Fourier series expansion. Without going through the mathematics of MMS, we use eqns [12] and [13] to obtain the solution of the wave equation and two sets of coupled mode equations. The solution to the wave equation in eqn [12] is given by

$$E_0 = \left(\frac{1}{\sqrt{k_1}}A_{f1}(z_1,t_1)e^{ik_1z_0} + \frac{1}{\sqrt{k_1}}A_{b1}(z_1,t_1)e^{-ik_1z_0}\right)e^{-i\omega t_0} + \left(\frac{1}{\sqrt{k_3}}A_{f3}(z_1,t_1)e^{ik_3z_0} + \frac{1}{\sqrt{k_3}}A_{b3}(z_1,t_1)e^{-ik_3z_0}\right)e^{-3i\omega t_0} + \text{c.c.} \quad [15]$$

where A is the amplitude of the field, k is the wavenumber such that $k_1^2 = \omega^2\varepsilon(\omega)/c^2$, and $k_3^2 = (3\omega)^2\varepsilon(3\omega)/c^2$ where c is the speed of light, and 'c.c.' refers to the complex conjugate. Note that the subscript 'f' and 'b' refer to the forward and backward modes, respectively. Also, the subscript '0' refers to the slowest scale, and '3' refers to the fastest scale chosen in the analysis using MMS. Let $a_{f1} = NA_{f1}e^{i\pi\delta_1 z_1/d}$, $a_{b1} = NA_{b1}e^{-i\pi\delta_1 z_1/d}$, $a_{f3} = NA_{f3}e^{i\pi\delta_3 z_1/d}$, and $a_{b3} = NA_{b3}e^{-i\pi\delta_3 z_1/d}$, such that

$$N^2 = \frac{2d\sqrt{3k_3}(3\omega)^2\chi^{(3)}}{k_3c^2k_1\sqrt{k_1}}$$

The coupled mode equations for the fundamental field are given as

$$\frac{\partial a_{f1}}{\partial z_1} + \frac{d}{\pi v_{g1}}\frac{\partial a_{f1}}{\partial t_1} = \frac{id}{2\pi k_1}\nabla_\perp^2 a_{f1} + i\delta_1 a_{f1} - i\kappa_1 a_{b1} - \frac{\alpha_1}{2}a_{f1} + ia_{f1}[2|a_{b1}|^2 + |a_{f1}|^2] + ia_{f1}^{*2}a_{f3} \quad [16a]$$

$$\frac{-\partial a_{b1}}{\partial z_1} + \frac{d}{\pi v_{g1}}\frac{\partial a_{b1}}{\partial t_1} = \frac{id}{2\pi k_1}\nabla_\perp^2 a_{b1} + i\delta_1 a_{b1} - i\kappa_1 a_{f1} - \frac{\alpha_1}{2}a_{b1} + ia_{b1}[2|a_{f1}|^2 + |a_{b1}|^2] + ia_{b1}^{*2}a_{b3} \quad [16b]$$

where, at the fundamental harmonic, v_{g1} is the group velocity of the homogeneous medium, δ_1 is the laser detuning, κ_1 is the coupling coefficient, α_1 is the absorption coefficient, $\nabla_\perp^2$ is the transverse Laplacian and the whole term ($\nabla_\perp^2 a_{f1}$ and $\nabla_\perp^2 a_{b1}$) accounts for diffraction effects, and '$*$' denote the complex conjugate of the corresponding term. Similarly, the coupled mode equations for the third-harmonic field are given as

$$\frac{\partial a_{f3}}{\partial z_1} + \frac{d}{\pi v_{g3}}\frac{\partial a_{f3}}{\partial t_1} = \frac{id}{2\pi k_3}\nabla_\perp^2 a_{f3} + ia_{f3}\left[\delta_3 + 6\sqrt{\frac{3k_1}{k_3}}(|a_{f1}|^2 + |a_{b1}|^2)\right] - i\kappa_3 a_{b3} - \frac{\alpha_3}{2}a_{f3} + ia_{f1}^3 \quad [17a]$$

$$-\frac{\partial a_{b3}}{\partial z_1} + \frac{d}{\pi v_{g3}}\frac{\partial a_{b3}}{\partial t_1} = \frac{id}{2\pi k_3}\nabla_\perp^2 a_{b3} + ia_{b3}\left[\delta_3 + 6\sqrt{\frac{3k_1}{k_3}}(|a_{b1}|^2 + |a_{f1}|^2)\right] - i\kappa_3 a_{f3} - \frac{\alpha_3}{2}a_{b3} + ia_{b1}^3 \quad [17b]$$

where at the third harmonic, v_{g3} is the group velocity of the homogeneous medium, δ_3 is the laser detuning, κ_3 is the coupling coefficient, and α_3 is the absorption coefficient. Note that in the limit when $k_3 \approx 3k_1$, the coupling coefficient between the fundamental and third harmonics are related by

$$\kappa_3 = 3\left(\frac{d}{\pi}\right)\frac{\Delta\varepsilon\omega_0^2}{2k_3c^2} + 3\delta_1 = \kappa_1 \quad [18]$$

Also, note that the laser detuning between the fundamental and third harmonics are expressed as

$$\delta_3 = \left(\frac{d}{\pi}\right)\Delta k + 3\delta_1 \quad [19]$$

where $\Delta k = k_3 - 3k_1 \times$ Equation [19] described the phase matching condition, which determines the efficiency of the third-harmonic generation. When the phase is completely matched ($\Delta k = 0$), then we can achieved the most efficienct in third-harmonic generation when $\delta_3 = 3\delta_1$. Equations [16] and [17] are useful equations which are commonly used for numerically simulation of beam propagation.

Defect Mode Using Photonic Crystal

A defect mode can be generated in the photonic crystal if we introduce a defect in the geometry of the photonic crystal. The schematic of the photonic crystal with a defect of length d_d located at the center of the photonic crystal, where the subscript 'd' denotes defect, is depicted in Figure 3. For the convenience of the analysis, we divide the entire photonic crystal into three regions: (i) Region I; (ii) Region II; and (iii) Region III. Since we arbitrarily

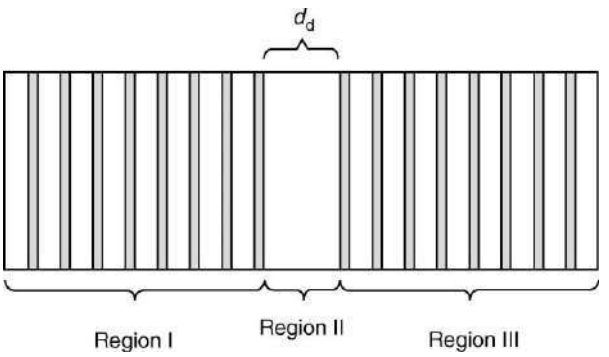

Figure 3 The schematic of the photonic crystal with a defect of length d_d.

put the defect in the center of the photonic crystal, Region I and II will be identical, thus it will be easier for us in our analysis. The strategy of our analysis is to solve the wave equation for Region I of the photonic crystal, and apply the result to the Region II analysis. Similarly, we apply the corresponding result of Region II to Region III. Thus, by breaking up the entire photonic crystal into three regions, we can then easily analyze the propagation of a field through the entire photonic crystal.

Assume that Region I and II are each of length $L/2$. Using the coupled mode equations without considering the effect of absorption, the fundamental fields in the forward and backward modes in Region I is given as

$$a_{f1(REG)}(z_1) = \left(\cos(\Delta_1 z_1) - \frac{i\delta_1}{\Delta_1}\sin(\Delta_1 z_1)\right)a_{f1(REG)}(0) + \frac{i\kappa}{\Delta_1}\sin(\Delta_1 z_1)a_{b1(REG)}(0) \quad [20a]$$

$$a_{b1(REG)}(z_1) = \left(\cos(\Delta_1 z_1) + \frac{i\delta_1}{\Delta_1}\sin(\Delta_1 z_1)\right)a_{b1(REG)}(0) - \frac{i\kappa}{\Delta_1}\sin(\Delta_1 z_1)a_{f1(REG)}(0) \quad [20b]$$

where

$$\Delta_1 = \sqrt{\delta_1^2 - \kappa^2}, \qquad \kappa = \frac{d\omega_0^2\Delta\varepsilon(\omega)}{2\pi k_1 c^2}$$

and the subscript '(REG)' denotes solution in either Region I, II, or III. The boundary conditions for the fundamental harmonic field, at the interface between Region I and II, are given as

$$a_{f1(II)}(d_d) = e^{ikd_d}a_{f1(I)}(L/2) \quad [21a]$$

$$a_{b1(II)}(d_d) = e^{-ikd_d}a_{b1(I)}(L/2) \quad [21b]$$

where $k = n\omega/c$. The field that propagates through Region II can be determined using eqns [20] and [21]. Also, the boundary conditions at the front and end faces of the Region III of the photonic crystal are given as

$$a_{f1(III)}(0) = a_{f1(II)}(d_d) \quad [22a]$$

$$a_{b1(III)}(L/2) = 0 \quad [22b]$$

Using eqns [20] and [22], we can determine the field that propagates through Region III. Similarly, we can use the same approach to analyze the third-harmonic field.

Homogeneous Nonlinear Crystal

If we examine Figure 2, we see that when $\Delta\varepsilon(\omega) = 0$, the photonic crystal becomes a homogeneous nonlinear crystal. In another word, $\kappa_3 = 0$ when we set the laser detuning to zero, thus there is no more coupling between the layers, and we are left with only the forward mode. Without going through the mathematics of four-wave mixing to describe the THG in a homogeneous nonlinear crystal, we can obtain the fundamental and third-harmonic coupled mode equations from eqns [16] and [17]. This will make it easier to make comparisons between the photonic crystal and the homogeneous nonlinear crystal, rather than comparing the coupled mode equation with the four-wave mixing equations among them. Since no backward wave is generated in the homogeneous nonlinear crystal when there is no coupling between the forward and backward modes, eqn [16] can remove the coupling term and the backward mode equation, and the modification is given as

$$\frac{\partial a_{f1}}{\partial z_1} + \frac{d}{\pi v_{g1}}\frac{\partial a_{f1}}{\partial t_1} = \frac{id}{2\pi k_1}\nabla_\perp^2 a_{f1} + i\delta_1 a_{f1} - \frac{\alpha_1}{2}a_{f1} + ia_{f1}|a_{f1}|^2 + ia_{f1}^{*2}a_{f3} \quad [23]$$

$$\frac{\partial a_{f3}}{\partial z_1} + \frac{d}{\pi v_{g3}}\frac{\partial a_{f3}}{\partial t_1} = \frac{id}{2\pi k_3}\nabla_\perp^2 a_{f3} + ia_{f3}\left[\delta_3 + 6\sqrt{\frac{3k_1}{k_3}}|a_{f1}|^2\right] - \frac{\alpha_3}{2}a_{f3} + ia_{f1}^3 \quad [24]$$

for the fundamental and third-harmonic fields, respectively.

Applications

Generation of short wavelength radiation is possible using THG. Therefore, it becomes less expensive to generate green, blue, or ultraviolet light with the

infrared and the red light. In order to improve THG conversion efficiency, a beam with a tight focus is used, which results in an electric field of highest intensity, and has the highest possible χ^3 polarization effect. Normally, due to the phenomenon of the Gouy phase shift at a tight focal point of the beam, third-harmonic generation is suppressed. However, in the presence of an interface within the focal volume, the phase is disturbed and a third-harmonic signal is generated. This phenomenon has been used to image cells in *in vivo* microscopy on a femtosecond time-scale. The ultrashort pulses have little energy and therefore little heat is deposited in the material. The cells can be continuously imaged without damage.

See also

Materials Characterization Techniques: $\chi^{(3)}$. **Photonic Crystals:** Photonic Crystal Lasers, Cavities and Waveguides.

Further Reading

Barad Y, Eisenberg H, Horovitz M and Silberberg Y (1997) Nonlinear scanning laser microscopy by third harmonic generation. *Applied Physics Letters* **70**: 922–924.

Boyd RW (1992) *Nonlinear Optics*. Boston, MA: Academic Press.

Butcher PN and Cotter D (1990) *The Elements of Nonlinear Optics*. Cambridge: Cambridge University Press.

Franken PA, Hill AE, Peters CW and Weinreich G (1961) Generation of optical harmonics. *Physics Review Letters* 7: 118.

Markowicz PP, Tiryaki H, Pudavar H, Prasad PN, Lepeshkin NN and Boyd RW (2004) Dramatic enhancement of third-harmonic generation in three-dimensional photonic crystals. *Physics Review Letters* 92: 083903.

Petrov GI, Shcheslavskiy V, Yakovlev VV, Ozerov I, Chelnokov E and Marine W (2003) Efficient third-harmonic generation in a thin nanocrystalline film of ZnO. *Applied Physics Letters* **83**: 3993–3995.

Schins J, Schrama T, Brakenhoff G, Muller M and Squier J (2002) Determination of material properties using third harmonic generation microscopy. *Journal of the Optical Society of America B* 19: 1627.

Squier J, Mueller M, Brakenhoff GJ and Wilson KR (1998) Third harmonic generation microscopy. *Optics Express* 3: 315.

Terhune RW, Maker TD and Savage CM (1962) Optical harmonic generation in calcite. *Physics Review Letters* 8: 404.

Viswanathan R (1999). Enhanced Frequency Conversion in Periodic Structures. Ph.D. dissertation, Rensselaer Polytechnic Institute.

Four-Wave Mixing

L Canioni and L Sarger, University of Bordeaux, Talence, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Light waves propagate along straight lines in a vacuum and in materials. However, their amplitude and phase can only be affected by the intrinsic properties of the sample, such as absorption and refraction index. Experimental work from the early 1960s, has deviated from this simple approach, mainly through propagating intense light beams in crystals and glasses. Among many puzzling new effects, frequency conversion and beam coupling have triggered intense studies and opened a whole new field, theoretically and experimentally. These various phenomena have been classified according to the order of their optical nonlinearity. The second order mainly occurs in frequency generation and has already been discussed in previous articles of this encyclopedia. Therefore, this article will be devoted to the higher order, in which self focusing, coupling, and mixing of light beams will be explained. We will illustrate some simple ideas with a convenient model before describing a few applications.

The framework of electromagnetic wave interaction with matter is embedded in the Lorentz force. The oscillating electric field drives the components, ions, and electrons, but due to the large mass differences, only the electron movement is strongly perturbed. In dielectric media, these carriers are connected with static bonds such that the applied electric field modulates the spatial position of equilibrium of the charges. As a result, a dipolar moment is created and oscillates quasi-adiabatically at very high frequencies (between 10^{14} and 10^{17} Hz) for an applied optical wave. These collective oscillations of the induced dipoles compose the so-called matter polarization, proportional to the driving electric field amplitude. Magnetic field interaction is usually much weaker and so can be neglected in nonmagnetic media.

As an electromagnetic wave, lasers are a coherent source of very high intensity, i.e., large electric fields. For example, in standard pulsed

lasers focused on a target, the light intensity can easily reach 10 Tw/cm^2, corresponding to an electric field of the order of magnitude of the internal bonding field (10^{10} V/m). At this level of interaction, dipolar moments are no longer strictly proportional to the applied electric field, as depicted in Figure 1. In the weak field regime, the simple perturbation approach displaces the carriers in a power series of the applied electric field. The polarization is thus developed in series, in respect to this electric field.

Formalism

Relation Between Field and Polarization: The Response Function

Although numerous works have been published with a frequency description of the material polarization, we will discuss here the study of nonlinearity in the time domain in order to follow the physical reality. In most processes, both spaces are equal and can be selected at will, as long as the entire requirements are fully understood; the electromagnetic field is classically described.

In the time domain, the polarization can be developed as

$$P(r,t) = P_{\text{Linear}}(r,t) + P_{\text{Nonlinear}}(r,t) \qquad [1]$$

where the linear part of the polarization can be written as

$$P_{\text{Linear}}(r,t) = \varepsilon_0 \int_{R^3} \int_{-\infty}^{\infty} R^{(1)}(r_1,t_1) \times E(r-r_1, t-t_1)\mathrm{d}t_1\,\mathrm{d}r_1 \qquad [2]$$

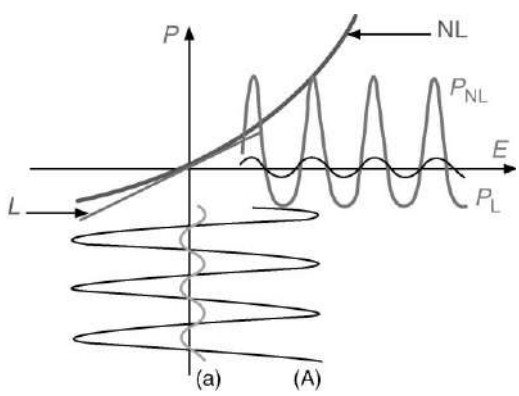

Figure 1 Polarization for various field amplitude: (a) Weak field case, linear response. (A) Strong field, distortion and nonlinearity.

This general expression is rather complex but shows that the polarization, due to collective movement of carriers in a macroscopic volume (thousands of dipoles) centered at position r, always depends on the applied electric field. This expression is also nonlocal as dipoles–dipoles interactions play an important role.

In general, for the nonlinear (NL) part of the polarization, an approximation of the local interaction is described for simplicity. For a second-order interaction, one can write:

$$P^{(2)}_{\text{NL}}(r,t) = \varepsilon_0 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} R^{(2)}(t_1,t_2) \times E(r,t-t_2)E(r,t-t_2-t_1)\mathrm{d}t_1\,\mathrm{d}t_2 \qquad [3]$$

For the other series expansion – at third order, one obtains

$$P^{(3)}_{\text{NL}}(r,t) = \varepsilon_0 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} R^{(3)}(t_1,t_2,t_3)E(r,t-t_3) \times E(r,t-t_3-t_2)E(r,t-t_3-t_2-t_1)\mathrm{d}t_1\,\mathrm{d}t_2\,\mathrm{d}t_3 \qquad [4]$$

This third-order polarization already combines three electric fields. This nonlinear polarization also radiates a fourth wave and is therefore the adopted framework for describing four-wave mixing experiments (Figure 2).

As the physical properties of materials depend only on time intervals between the different pulse interactions, eqn (4) can be rearranged as

$$P^{(3)}_{\text{NL}}(r,t) = \varepsilon_0 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} R^{(3)}(t-\tau_1, t-\tau_2, t-\tau_3) \times E(r,\tau_1)E(r,\tau_2)E(r,\tau_3)\mathrm{d}\tau_1\,\mathrm{d}\tau_2\,\mathrm{d}\tau_3 \qquad [5]$$

If the sample response time is much shorter than pulse duration, the so-called adiabatic limit (as for electronic cloud deformation excited distant from

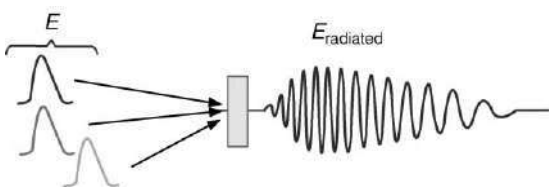

Figure 2 Three incoming beams E_1, E_2, E_3 create a polarization (nonlinear). The sample responds while radiating a fourth field E_4.

resonance) the polarization is written as

$$P_{\mathrm{NL}}^{(3)}(r,t)=\varepsilon_0\sigma^{(3)}E(r,t)E(r,t)E(r,t) \qquad [6]$$

We will see later how this expression is useful for third-harmonic generation and optical Kerr effects, etc. This expression is accurate for materials excited at a distance from resonances and care must be taken to reserve such a model for corresponding processes. In this case, the frequency approach is quite simple too and both spaces, time or Fourier, can be chosen at will.

For very long response time, such as orientational or structural molecular reorganization, one can use the Born–Oppenheimer approximation and present the nonlinear polarization as

$$P_{\mathrm{NL}}^{(3)}(r,t)=\varepsilon_0 E(r,t)R'(t) \qquad [7]$$

Similar is linear polarization, where the response function takes into account the changes of the molecular structure under the light intensity stress:

$$R'(t)=\int_{-\infty}^{+\infty} d^{(3)}(t-t_1)E(r,t_1)E(r,t_1)\mathrm{d}t_1 \qquad [8]$$

The material response will then be mostly driven by the field amplitude rather than the frequency.

Properties of Nonlinear Susceptibilities

The third-order nonlinear susceptibility $\chi^{(3)}$ is the Fourier transform of the response function R and has proved to be a most practical tool in optical nonlinear theory. The symmetry properties of the material lead to simplification on the tensor of rank 4 (fourth rank tensor), involved in four-wave mixing:

$$P_i^{(3)}(\omega_m)=\varepsilon_0\sum_{n_1,\ldots,n_p}\sum_{i_1,\ldots,i_p}\chi_{i,i_1,\ldots,i_p}^{(3)}(\omega_m;\omega_{n_1},\ldots,\omega_{n_p}) \times E_{i_1}(\omega_{n_1})\cdots E_{i_p}(\omega_{n_p}) \qquad [9]$$

The nonlinear susceptibility tensor must account for all the spatial and symmetrical order of the sample. This will eventually reduce the number of independent elements from a possible 81.

Wave Equation in Nonlinear Regime

From Maxwell's equations, with a sample without free carriers and current, one can deduce the wave equation in the time domain, as:

$$\Delta E-\frac{1}{c^2}\frac{\partial^2 E}{\partial t^2}-\frac{1}{c^2}\frac{\partial^2 P_{\mathrm{L}}}{\partial t^2}=\frac{1}{c^2}\frac{\partial^2 P_{\mathrm{NL}}}{\partial t^2} \qquad [10]$$

In the frequency domain, this equation reduces to

$$\Delta\tilde{E}(r,\omega)+k(\omega)^2\tilde{E}(r,\omega)=-\mu_0\omega^2\tilde{P}_{\mathrm{NL}}(r,\omega) \qquad [11]$$

In order to solve these (coupled) wave equations in relation to four-wave mixing, we must develop the electric field and the polarization in a slowly varying envelope approximation, using the following conventions:

$$E_i(r,t)=\frac{1}{2}\left(A_i(r,t)\mathrm{e}^{\mathrm{i}(\omega_i t-k(\omega_i)z)}+\bar{A}_i(r,t)\mathrm{e}^{-\mathrm{i}(\omega_i t-k(\omega_i)z)}\right) \qquad [12]$$

$$P_{\mathrm{NL}}(r,t)=\frac{1}{2}\left(p_{\mathrm{nl}}(r,t)\mathrm{e}^{\mathrm{i}(\omega t-k(\omega)z)}+\bar{p}_{\mathrm{nl}}(r,t)\mathrm{e}^{-\mathrm{i}(\omega t-k(\omega)z)}\right) \qquad [13]$$

Within the framework of small perturbations, diffraction can be neglected and the wave equation can be easily reduced to one dimension:

$$\frac{\partial A}{\partial z}=\frac{\mathrm{i}\omega}{2n(\omega)\varepsilon o c}p^{\mathrm{NL}}(z,\omega)\exp(-\mathrm{i}k(\omega)z) \qquad [14]$$

where ω represents the sum or difference permutations over all the frequencies ω_i, thus reflecting the law of energy conservation.

In the case of four-wave mixing, one has to solve four coupled equations for the four interacting waves. Using the polarization eqn [13], one gets:

$$\frac{\partial A_j}{\partial z}=\frac{\mathrm{i}\omega_j}{2n(\omega_j)c}\chi^3 A_1^{(*)}A_2^{(*)}A_3^{(*)} \times\exp\left(\sum_{i=1,2,3}\exp(\pm\mathrm{i}k_i(\omega_i)z)\right)\exp(-\mathrm{i}k(\omega_j)z) \qquad [15]$$

This set of generic equations respectively represent indifferently, either the three incoming beams ($j=1,2,3$), or the radiated field E_4 at ω. The algebra used here associates at each complex conjugate field (*) with the counter propagating wave ($-k$). The amplitude variation of the j field at frequency ω_j is due to the coupling of the three other fields within the sample. The accumulated field is nevertheless always strongly dependent of the phase term and will vanish except when this phase term is zero. This is known as the phase matching condition which has to be fulfilled to ensure an efficient energy conversion. A more general approach of this condition can be found in the corresponding article of this encyclopedia.

Example and Selected Applications

When inserting the full material polarization (linear and nonlinear) into the propagation equation, even

restricted at third order, a whole set of processes must be considered. All of them are described by the generic equation already presented, but only a few very important processes – with increasing complexity – will be described here.

Optical Kerr Effect

We consider one single beam incident at frequency ω on a sample. The non-linear polarization is given by

$$P_{\mathrm{NL}}^{(3)}(r,t)=\frac{\varepsilon_0}{8}\chi^{(3)}\left(E(r,t)^3+E^*(r,t)^3+3E(r,t)^2E^*(r,t)+3E(r,t)E^*(r,t)^2\right) \quad [16]$$

The polarization at frequency ω reduces to

$$P_{\mathrm{NL}}^{(3)}(r,t)=\frac{3\varepsilon_0}{4}\chi^{(3)}|E(r,t)|^2E(r,t) \quad [17]$$

This term has the same phase as the incident beam and, therefore, the phase matching is automatic. Using eqns [12] and [13], one can rewrite the wave equation for the optical Kerr effect in the thin sample approximation:

$$\frac{\partial A}{\partial z}=\frac{3\mathrm{i}\omega}{8nc}\chi^{(3)}|A|^2A \quad [18]$$

The solution of this equation is

$$A(z)=A(0)\exp\left(\frac{3\mathrm{i}}{8nc}\chi^{(3)}|A(0)|^2z\right) \quad [19]$$

Along the beam, the amplitude of the input pulse stays constant for this process but its phase is affected by a nonlinear shift proportional to power density I.

In literature, one can find the concept of n_2, the sample nonlinear index, in the following formula:

$$\psi_{\mathrm{NL}}(z)=\frac{\omega}{c}n_2Iz \quad [20]$$

In this way of describing the total phase along the beam, the index of refraction in the nonlinear regime can be expressed more conveniently as

$$n=n_0+n_2I \quad [21]$$

One of the most significant demonstrations of this effect is the self-focusing phenomenon. The usual Gaussian transverse structure of a laser beam presents high intensity at the center of the beam, therefore creating an index variation in the isotropic sample due to this optical Kerr effect. An induced lens can then be applied to this refractive index gradient, following the field distribution. As a result, the beam focuses itself along the propagation and can result in a complete breakdown.

Third Harmonic Generation

In the third harmonic generation, i.e., the third-order nonlinear process where the frequency of the generated wave is three times the frequency of the input wave at ω, only two waves are in the sample (at ω and 3ω). For the nonlinear wave, the master equation is then:

$$\frac{\partial A_{3\omega}}{\partial z}=\frac{3\mathrm{i}\omega}{8n(3\omega)c}\chi^{(3)}A_\omega^3\exp((3\mathrm{i}k(\omega)-\mathrm{i}k(3\omega))z) \quad [22]$$

Here, propagation effects have to be carefully considered as theses waves do travel at different velocities and will eventually be out of phase. Obviously this discrepancy of phase velocities precludes the co-propagation and strongly affects the global efficiency. The only way to ensure a coherent generation of the harmonic wave throughout the material is to fulfill the relation:

$$k(3\omega)=3k(\omega) \quad [23]$$

This is the so-called phase matching condition. Although difficult because the usual dispersion of materials is very large for such a frequency difference, this can be satisfied using either geometrical arrangement or, if possible, using the birefringent properties of the material. The efficiency of this method is poor and a better approach is implemented when two $\chi^{(2)}$ processes are cascaded (doubling and sum frequency mixing).

Degenerate Four-Wave Mixing and Phase Conjugation

One of the most popular and intriguing interactions is depicted in Figure 3. Where the sample is excited by three fields at the same frequency ω in this particular geometry: two-waves, named pump beams E_1 and E_2, are exactly counter-propagating with opposite wave vectors of $+k_1$ and $k_2=-k_1$, while the third wave (E_3) arrives at the sample with a given wave vector of k_3.

Of possible couplings, one contribution E_4 corresponds to a re-emitted beam along $k_4=-k_3$, i.e., opposite to the E_3 wave.

The third-order nonlinear polarization is given here by

$$P_{\mathrm{NL}}^{(3)}(r,t) = \frac{\varepsilon_0}{8}\chi^{(3)}(E_1(r,t) + E_1^*(r,t) + E_2(r,t) + E_2^*(r,t) + E_3(r,t) + E_3^*(r,t))^3 \qquad [24]$$

This E_4 wave (along $-k_3$) originates in this product from source terms including:

$$E_1(r,t)E_2(r,t)E_3^*(r,t) \qquad [25]$$

and thus present an evolution driven by:

$$\frac{\partial A_4}{\partial z} = \frac{3\mathrm{i}\omega}{4n(\omega)c}\chi^{(3)}A_1A_2A_3^* \qquad [26]$$

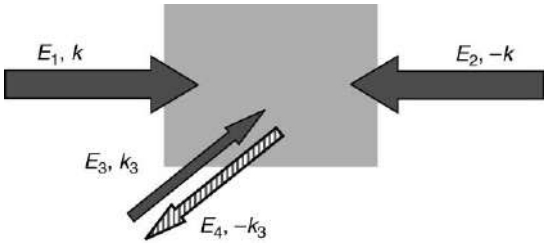

Figure 3 Two counter-propagating fields (E_1, E_2) and the third field E_3 create a polarization (nonlinear). The sample responds while radiating a fourth field E_4 opposite to the E_3 wave.

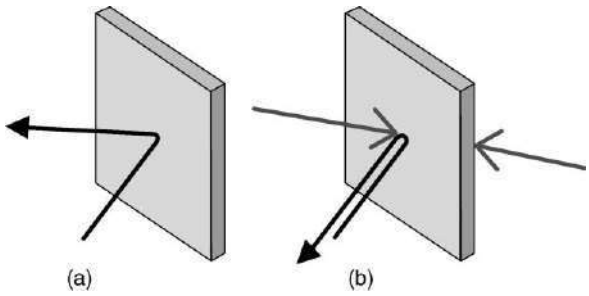

Figure 4 (a) Snell refraction in a conventional mirror; (b) Phase conjugate mirror.

where the phase matching condition is automatically fulfilled. Interesting facts can be described using this expression. First, the radiated amplitude is proportional to the complex conjugate of the E_3 field. This nonlinear process is not only able to reverse the direction of propagation but also the whole phase of an arbitrary incoming beam of light. This 'phase conjugator' can be considered as a kind of mirror with very unusual reflection properties. Unlike a conventional mirror, where a ray is redirected according to the ordinary law of reflection, a phase-conjugate mirror (PCM) retro-reflects all incoming rays back to their origin (Figure 4).

Second, the reflectivity of this mirror depends on the susceptibility and the amplitude of the first two fields, usually called pump beams. This will allow reflectivity that is much higher than the unity, unlike with conventional mirrors.

The remarkable reflection properties of the PCM have found many important applications. The most useful undoubtedly is related to distortion correction. If the image information has been distorted by the transmitting medium on its way to the PCM, then these aberrations will be corrected when the reflected signal retraces its original path through the medium. Through this amazing 'healing' property of optical phase conjugation, a high-quality optical beam can be double passed through a distorting medium, such as an imperfect imaging device or a turbulent atmosphere, without any loss of beam quality. Figure 5 shows the basic two-pass geometry for imaging and aberration correction using a PCM.

The spherical wave from the object point P is aberrated by the distorting medium. The wave-front is reversed by the PCM and a second passage through the same distorting path restores the initial field. The resulting spherical wave converges to the image point P', separated from P by the beamsplitter.

Numerous other applications for PCM and the various underlying nonlinear optical interactions have been proposed. These include effects as diverse

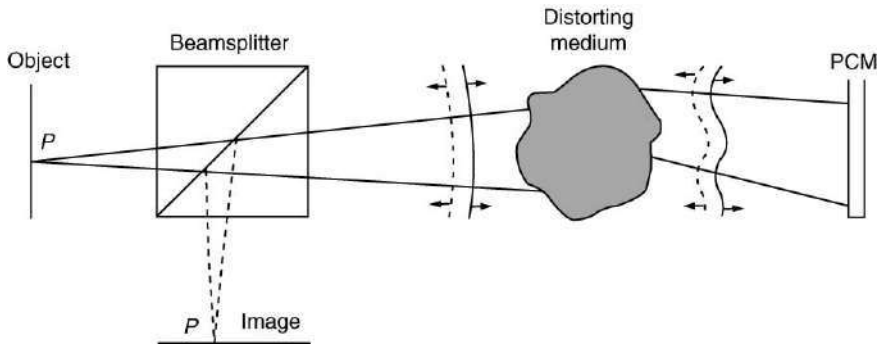

Figure 5 A basic two-pass system for imaging and distortion compensation using optical phase conjugation.

as parametric oscillation, optical tracking and pointing, spatial and temporal image processing, optical computing, optical filtering, etc. The PCM is also successfully implemented for aberration correction in high-power laser resonators.

Nondegenerate Four-Wave Mixing

If the frequencies of the impinging fields build a nonlinear polarization with an oscillating term close to absorption (or emission) frequencies of the sample, the nonlinear susceptibility increases tremendously. The radiated field amplitude changes by orders of magnitude and allows spectroscopic studies of the sample. The most popular process used to perform gas spectroscopy is CARS, an acronym for coherent anti-Stokes Raman scattering (or spectroscopy). In this case, the three incoming beams have different frequencies and the phase matching condition is eventually fulfilled in the 3D space.

Usually, two beams at the same frequency ω_P (pump beams), are mixed with a third beam at a tunable frequency ω_S, according to the energy scheme in Figure 6. As the radiated frequency ω_{CARS} is higher, the Raman spectroscopic notation is used and the subscripts 'S' hold for 'Stokes' and 'P' for 'anti-Stokes'.

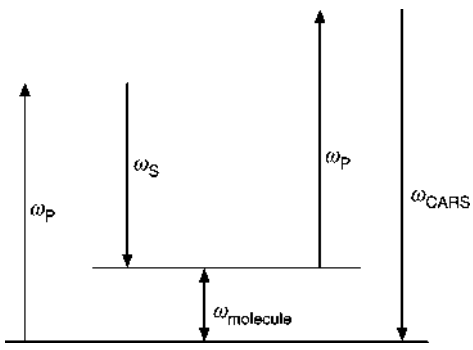

Figure 6 Energy diagram for a CARS experiment. Four-wave mixing efficiency increases when the frequency difference $\omega_P - \omega_S$ is close to a resonant frequency of the sample, $\omega_{molecule}$.

Whenever $\omega_P - \omega_S = \omega_{molecule}$, then an increase of the CARS intensity by many orders of magnitude is observed. Therefore, a Raman spectrum can be obtained by continuously changing, the Stokes laser and simultaneously recording the intensity of the CARS beam. This procedure is called scanning CARS, an interesting technique insofar as it allows taking high-resolution spectra, as well as measuring temperature of the sample while scanning over the Boltzmann distribution of the electronic ground state.

Figure 7 displays a classical geometrical arrangement where the phase matching condition fully determines the directionality of the radiated field at ω_{CARS}. Although the alignment of the setup can become tedious, this technique has been widely used, even in hostile industrial environments. The coherent aspect of this four-wave mixing process, imbedded in the phase matching condition, has a huge benefit over classical spectroscopic techniques and thus isotropic techniques, such as laser-induced fluorescence or Raman spectroscopy.

The CARS process is an interference process comparable to diffraction of a grating. The two fields at ω_P and ω_S form a laser-induced moving grating and the third field, at ω_P, undergoes a Doppler shift at ω_{CARS}. This scattered field undergoes a coherent amplitude addition in the Bragg direction.

Conclusion

In this article, we have discussed a set of optical nonlinear processes occurring at the third order of the development in series of the material polarization. At this order, wave mixing obviously shows most

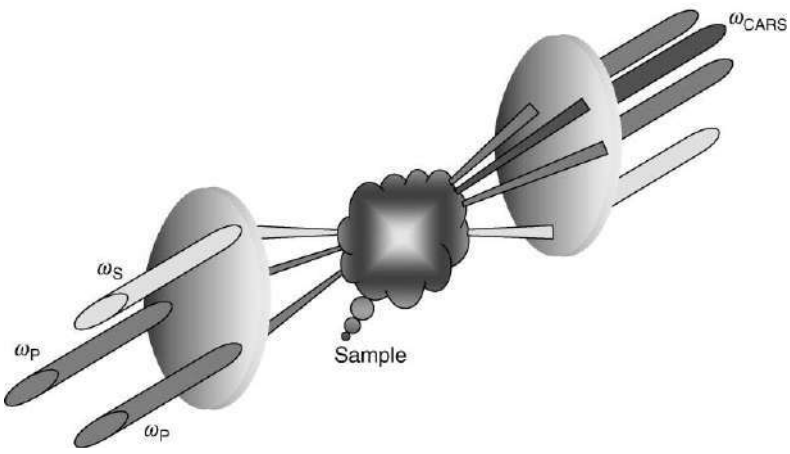

Figure 7 Layout of the CARS beams near a sample (a flame here). This particular arrangement is called folded (or 3D) Boxcars.

significant behavior and has been extensively studied and applied in many scientific areas. The general framework presented here can be further investigated using some books listed in Further Reading below.

See also

Fiber and Guided Wave Optics: Nonlinear Optics. **Nonlinear Optics, Applications:** Phase Matching. **Nonlinear Optics, Basics:** $\chi^{(3)}$–Third-Harmonic Generation. **Nonlinear Sources:** Harmonic Generation in Gases. **Optical Parametric Devices:** Optical Parametric Oscillators (Continuous Wave); Optical Parametric Oscillators (Pulsed).

Further Reading

Bloembergen N (1996) *Nonlinear Optics*. Singapore: World Scientific Pub Co.

Bloom DM and Bjorklund GC (1977) Conjugate wavefront generation and image reconstruction by four-wave mixing. *Applied Physics Letters* 31: 592–594.

Born M and Wolf E (1999) *Principles of Optics*. Cambridge, UK: Cambridge University Press.

Butcher PN and Cotter D (1991) *The Elements of Nonlinear Optics*. Cambridge, UK: Cambridge University Press.

Eaton DF (1991) Nonlinear optical materials. *Science* 253: 281–287.

Fisher RA (1984) *Optical Phase Conjugation*. San Diego, CA: Academic.

Flytzanis C (1975) Theory of nonlinear optical susceptibilities. In: Rabin H and Tang CL (eds) *Quantum Electronics*, vol. 1, pp. 9–207. New York: Academic Press.

Hellwarth RW (1977) Generation of time-reversed wave fronts by nonlinear refraction. *Journal of the Optical Society of America* 67: 1–3.

Jackson JD (1998) *Classical Electrodynamics*. New York: Wiley.

Marburger JH (1975) Self-focusing: Theory. *Progress in Quantum Electronics* 4: 35–110.

Mittra R and Habashy TM (1984) Theory of wave-front-distortion correction by phase conjugation. *Journal of the Optical Society of America A* 1: 1103–1109.

Rockwell DA (1988) A review of phase-conjugate solid-state lasers. *IEEE Journal of Quantum Electronics* 24: 1124–1140.

Shen YR (2002) *The Principles of Nonlinear Optics*. New York: Wiley.

Yariv A (1977) Compensation for atmospheric degradation of optical beam transmission. *Optical Communication* 21: 49–50.

Yariv A (1989) *Quantum Electronics*, 3rd edn. New York: Wiley.

Yariv A and Pepper DM (1977) Amplified reflection, phase conjugation, and oscillation in degenerate four-wave mixing. *Optical Letters* 1: 16–18.

Zel'dovich BYa, Pilipetsky NF and Shkunov VV (1985) *Principles of Phase Conjugation*. Berlin: Springer-Verlag.

Kramers–Krönig Relations in Nonlinear Optics

M Sheik-Bahae, The University of New Mexico, Albuquerque, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Since their introduction nearly 75 years ago, the Kramers–Krönig (KK) dispersion relations have been widely appreciated and applied in the analysis of linear optical systems. Because they are a consequence of strict causality, the KK relations apply not only to optical systems, but also to any linear, causal system such as electrical networks and particle scattering. In this article, we review the formulation and application of these relations in nonlinear optical systems. Simple logical arguments are used to derive dispersion relations that relate the nonlinear absorption coefficient to the nonlinear refraction coefficient. More general formalisms are then derived that apply to all nonlinear susceptibilities including the harmonic generating cases. Examples of recent successful application of these dispersion relations in analyzing various nonlinear materials will be presented.

The mathematical formalism of the KK dispersion relations in nonlinear optics was studied in the formative days of the field. The great usefulness of these relations was appreciated only recently, however, when they were used to derive the dispersion of the optical Kerr effect in solids from the corresponding nonlinear absorption coefficients, including two-photon absorption.

Before examining the details of KK relations in nonlinear optical systems, it is instructive to revisit the linear dispersion relations and their derivation based on the logic of causality. We will begin this task by introducing the definition of the linear as well as nonlinear susceptibilities $\chi^{(n)}$. In most nonlinear optics texts, the total material polarization (P) that

drives the wave equation for the electric field (E) is expressed as

$$\begin{aligned}P_i(t)=&\varepsilon_0\int_{-\infty}^{\infty}R_{ij}^{(1)}(t-t_1)E_j(t_1)\mathrm{d}t_1+\varepsilon_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\\&\times R_{ijk}^{(2)}(t-t_1,t-t_2)E_j(t_1)E_k(t_2)\mathrm{d}t_1\,\mathrm{d}t_2\\&+\varepsilon_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}R_{ijkl}^{(3)}(t-t_1,t-t_2,t-t_3)\\&\times E_j(t_1)E_k(t_2)E_l(t_3)\mathrm{d}t_1\,\mathrm{d}t_2\,\mathrm{d}t_3+\cdots\end{aligned}\qquad[1]$$

where $R^{(n)}$ is defined as the nth-order, time-dependent response function or time-dependent susceptibility. The subscripts are polarization indices indicating, in general, the tensor nature of the interactions. The summation over the various indices $j, k, l, \ldots$ is implied for the various tensor elements of $R^{(n)}$. Upon Fourier transformation, we obtain:

$$\begin{aligned}\vec{P}_i(\omega)=&\varepsilon_0\int_{\infty}^{-\infty}\mathrm{d}\omega_1\,\chi_{ij}^{(1)}(\omega_1)\vec{E}_j(\omega_1)\delta(\omega-\omega_1)\\&+\varepsilon_0\int_{\infty}^{-\infty}\mathrm{d}\omega_1\int_{\infty}^{-\infty}\mathrm{d}\omega_2\\&\times\chi_{ijk}^{(2)}(\omega_1,\omega_2)\vec{E}_j(\omega_1)\vec{E}_k(\omega_2)\delta(\omega-\omega_1-\omega_2)\\&+\varepsilon_0\int_{\infty}^{-\infty}\mathrm{d}\omega_1\int_{\infty}^{-\infty}\mathrm{d}\omega_2\int_{\infty}^{-\infty}\mathrm{d}\omega_3\\&\times\chi_{ijkl}^{(3)}(\omega_1,\omega_2,\omega_3)\vec{E}_j(\omega_1)\vec{E}_k(\omega_2)\vec{E}_l(\omega_3)\\&\times\delta(\omega-\omega_1-\omega_2-\omega_3),\cdots\end{aligned}\qquad[2]$$

where δ is the Dirac delta-function. Here the $E(\omega)$ are Fourier transforms of the corresponding electric field. The nth-order susceptibility is defined as the Fourier transform of the nth-order response function:

$$\begin{aligned}\chi_{ijk\ldots n}^{(n)}(\omega_1,\omega_2,\ldots,\omega_m)=&\int_{-\infty}^{+\infty}\mathrm{d}\tau_1\int_{-\infty}^{+\infty}\mathrm{d}\tau_2\cdots\int_{-\infty}^{+\infty}\\&\times\mathrm{d}\tau_n R_{ijk\ldots m}^{(n)}(\tau_1,\tau_2,\ldots,\tau_m)\mathrm{e}^{\mathrm{i}(\omega_1\tau_1+\omega_2\tau_2+\cdots+\omega_m\tau_m)}\end{aligned}\qquad[3]$$

For simplicity, we drop the polarization indices $i, j, \ldots$, and thus ignore the tensor properties of $\chi^{(n)}$ as well as the vector nature of the electric fields.

Let us for the moment concentrate on the linear polarization alone and derive the linear KK relations for the first-order susceptibility $\chi^{(1)}(\omega)$. For this, we rewrite eqn [3] for $n = 1$:

$$\chi^{(1)}(\omega)=\int_{-\infty}^{\infty}R^{(1)}(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau\qquad[4]$$

(As defined above, $\chi^{(1)}(\omega)$ and $R^{(1)}(\tau)$ are not a strict Fourier transform pair because of a missing factor of 2π.) Causality means that the effect cannot precede the cause. This can be restated mathematically as:

$$R^{(1)}(t)=R^{(1)}(t)\Theta(t)\qquad[5]$$

i.e., the response to an impulse at $t = 0$ must be zero for $t < 0$. Here $\Theta(t)$ is the Heaviside step function defined as $\Theta(t) = 1$ for $t > 0$ and $\Theta(t) = 0$ for $t < 0$. Upon Fourier transforming this equation, the product in the time domain becomes a convolution integral in frequency space

$$\begin{aligned}\chi^{(1)}(\omega)&=\chi^{(1)}(\omega)\left[\frac{\delta(\omega)}{2}+\frac{\mathrm{i}}{2\pi\omega}\right]\\&=\frac{\chi^{(1)}(\omega)}{2}+\frac{\mathrm{i}}{2\pi}\wp\int_{-\infty}^{\infty}\frac{\chi^{(1)}(\omega')}{\omega-\omega'}\mathrm{d}\omega'\\&=\frac{1}{\mathrm{i}\pi}\wp\int_{-\infty}^{\infty}\frac{\chi^{(1)}(\omega')}{\omega'-\omega}\mathrm{d}\omega'\end{aligned}\qquad[6]$$

which is the KK relation for the linear optical susceptibility. The symbol $\wp$ stands for the Cauchy principal value of the integral. The KK relation is thus a restatement of the causality condition [5] in the frequency domain. Taking the real part we have,

$$\Re e\{\chi^{(1)}(\omega)\}=\frac{1}{\pi}\wp\int_{-\infty}^{\infty}\frac{\Im m\{\chi^{(1)}(\omega')\}}{\omega'-\omega}\mathrm{d}\omega'\qquad[7]$$

Taking the imaginary part of eqn [6] leads to a similar relation relating the imaginary part to an integral involving the real part. It is conventional to write the optical dispersion relations in terms of the more familiar quantities of refractive index, $n(\omega)$, and absorption coefficient, $\alpha(\omega)$. For $|\chi^{(1)}| \ll 1$ then $n - 1 = \Re e\{\chi^{(1)}\}/2$ and $\alpha = \omega\Im m\{\chi^{(1)}\}/c$, and eqn [7] is transformed into

$$n(\omega)-1=\frac{c}{\pi}\wp\int_{0}^{\infty}\frac{\alpha(\omega')}{\omega'2-\omega^2}\mathrm{d}\omega'\qquad[8]$$

where we additionally used the reality conditions of $n(\omega) = n(-\omega)$, and $\alpha(\omega) = \alpha(-\omega)$ to change the lower integral limit to 0. More rigorous analysis shows that eqn [8] is general and valid for any value of $|\chi^{(1)}|$. Although the KK dispersion relations and the extent of their applications in linear optics are well understood, some confusion sometimes exists about their applications to nonlinear optics. Causality clearly holds for both linear and nonlinear systems. The question is: what form do the resulting dispersion relations take in a nonlinear system? The linear Kramers–Krönig relations were derived from linear system theory, so it would appear to be impossible to apply the same logic to a nonlinear system. The key insight is that one can linearize the system. This is illustrated in Figure 1 where a linear

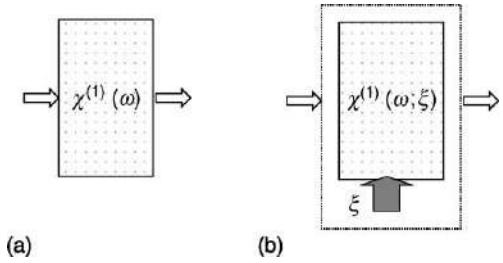

Figure 1 (a) A causal linear system obeying KK relations. (b) The system in (a) when externally perturbed by ξ. The dotted box now represents our new linear causal system whose altered $\chi^{(1)}$ obeys the KK relations.

(and of course, causal) optical material is transformed into a 'new' linear system that now contains the material and an external perturbation denoted by ξ. Although we are interested in perturbations of an optical nature, this formalism is general under any type of perturbation. It is important to appreciate the fact that our new system is causal even in the presence of the perturbation. This allows us to write down a modified form of the Kramers–Krönig relation linking the index of refraction to the absorption:

$$[n(\omega) + \Delta n(\omega;\zeta)] - 1 = \frac{c}{\pi}\wp\int_0^\infty \frac{\alpha(\omega') + \Delta\alpha(\omega';\zeta)}{\omega'2 - \omega^2}\mathrm{d}\omega' \qquad [9]$$

which, after subtracting the linear relation between n and α leaves a relation between the changes in index and absorption:

$$\Delta n(\omega;\zeta) = \frac{c}{\pi}\wp\int_0^\infty \frac{\Delta\alpha(\omega';\zeta)}{\omega'2 - \omega^2}\mathrm{d}\omega' \qquad [10]$$

where ζ denotes a general perturbation. An equivalent relation also exists whereby the change in absorption coefficient can be calculated from the change in the refractive index. It is essential that the perturbation be independent of frequency of observation, ω', in the integral (i.e., the excitation ζ must be held constant as ω' is varied).

Equation [10] has been used to determine refractive changes due to 'real' excitations such as thermal and free-carrier nonlinearities in semiconductors. In those cases, ζ denotes either ΔT (change of temperature) or ΔN (change of free-carrier density), respectively. In the former case, one calculates the refractive index change resulting from a thermally excited electron–hole plasma and the temperature shift of the band edge. For cases where an electron–hole plasma is injected (e.g., optically), the change of absorption gives the plasma contribution to the refractive index. In this case, the ζ parameter in eqn [10] is taken as the change in plasma density regardless of the mechanism of generation or the optical frequency.

Let us now extend this formalism to the case where the perturbation is virtual, occurring at an excitation frequency Ω that is below any material resonance. To the lowest order in the excitation irradiance I_Ω, we write

$$\Delta\alpha(\omega;\zeta) = \Delta\alpha(\omega;\Omega) = 2\alpha_2(\omega;\Omega)I_\Omega \qquad [11]$$

and

$$\Delta n(\omega;\zeta) = \Delta n(\omega;\Omega) = 2n_2(\omega;\Omega)I_\Omega \qquad [12]$$

where n_2 and α_2 are the nonlinear refractive index and absorption coefficients of the material, respectively. By definition, these coefficients are related to the third-order nonlinear susceptibility $\chi^{(3)}(\omega_1,\omega_2,\omega_3)$ via (*see* **Nonlinear Optics, Basics**: Nomenclature and Units)

$$n_2(\omega;\Omega) = \frac{3}{4\varepsilon_0 n_0(\omega)n_0(\Omega)c}\Re e\{\chi^{(3)}(\omega,-\Omega,\Omega)\} \qquad [13]$$

and

$$\alpha_2(\omega;\Omega) = \frac{3\omega_a}{2\varepsilon_0 n_0(\omega)n_0(\Omega)c^2}\Im m\{\chi^{(3)}(\omega,-\Omega,\Omega)\} \qquad [14]$$

We can therefore write the dispersion relations between α_2 and n_2:

$$n_2(\omega;\Omega) = \frac{c}{\pi}\wp\int_0^\infty \frac{\alpha_2(\omega';\Omega)}{\omega'2 - \Omega^2}\mathrm{d}\omega' \qquad [15]$$

Note that even when the degenerate $n_2(\omega) = n_2(\omega;\omega)$ is desired (at a given ω), the dispersion relation requires that we should know the nondegenerate absorption spectrum $\alpha_2(\omega';\omega)$ at all frequencies ω'.

Let us pause here and discuss some physical mechanisms that can be involved for a given system of interest. Consider a material characterized by an optical resonance occurring at, say ω_0 (i.e., a degenerate two-level system). For a solid, this resonance can be regarded as that of the fundamental energy gap; $\omega_0 = \omega_g = E_g/\hbar$ in a two-band system. Now, let us examine how the presence of an optical excitation at $\Omega < \omega_0$ can alter the absorption spectrum (at a variable probe ω'). In the quantum mechanical picture, this gives rise to a 'new' material whose perturbed wave functions are 'dressed' by the intensity and frequency of the applied optical field.

The lowest-order correction to the absorption is given by $\alpha_2(\omega';\Omega)$ which involves three major physical processes. Recalling that $\Omega < \omega_0$, these processes include (1) two-photon absorption (2PA) when $\omega' + \Omega \rightarrow \omega_0$ and (2) Raman-induced absorption when $\omega' - \Omega \rightarrow \omega_0$, both implying an absorption of a photon at the probe frequency ω' (i.e., $\alpha_2 > 0$). The third process can be identified as resulting from the blue-shift (for $\Omega < \omega_0$) of the resonance (known as the quadratic optical Stark effect) caused by the excitation field. For our two-level system, the latter results in a decrease followed by an increase in absorption in the vicinity of ω_0. An example of the overall absorption changes due to such processes is shown in Figure 2 where $\alpha_2(\omega';\Omega)$ is qualitatively plotted for a degenerate two-level system. We should note that the relative magnitude of each contribution as well as the width and shape of the resonances are chosen arbitrarily for the purpose of illustration. Using the KK relation in eqn [15], we can now arrive at the nonlinear index coefficient $n_2(\omega;\Omega)$. The result of this transformation is also given in Figure 2. The above simple example elucidates the key concepts involving the relationship between nonlinear absorption and refraction in materials for third-order processes. These concepts, when applied more rigorously to semiconductors, have been successful in predicting the sign, magnitude, and dispersion of n_2 due to the anharmonic motion of bound electrons. This will be briefly discussed later. Returning to the mathematical foundation of KK relations, we use eqns [13] and [14] to write eqn [15] in terms of the nonlinear susceptibility $\chi^{(3)}$:

$$\Re e\{\chi^{(3)}(\omega_1, \omega_2, -\omega_2)\} = \frac{1}{\pi}\wp\int_{-\infty}^{\infty} \frac{\Im m\{\chi^{(3)}(\omega', \omega_2, \omega_2)\}}{\omega' - \omega_1}\, d\omega' \qquad [16]$$

The above dispersion relation for $\chi^{(3)}$ was obtained using the physical and intuitive arguments that followed the linearization scheme depicted in Figure 1. General dispersion relations can be formulated following a mathematical procedure that is similar to the derivation of the linear KK relations. In this case we apply the causality condition directly to the nth-order nonlinear response $R^{(n)}$. For example, without loss of generality, we can write

$$R^{(n)}(\tau_1, \tau_2, \ldots, \tau_n) = R^{(n)}(\tau_1, \tau_2, \ldots, \tau_n)\Theta(\tau_j) \qquad [17]$$

and then calculate the Fourier transform of this equation. Here j can apply to any one of the indices $1, 2, \ldots, n$. Following the same procedure as for a linear response, we obtain

$$\chi^{(n)}(\omega_1, \omega_2, \ldots, \omega_j, \ldots, \omega_n) = \frac{-i}{\pi}\wp\int_{-\infty}^{\infty} \frac{\chi^{(n)}(\omega_1, \omega_2, \ldots, \omega', \ldots, \omega_n)}{\omega_j - \omega'}\, d\omega' \qquad [18]$$

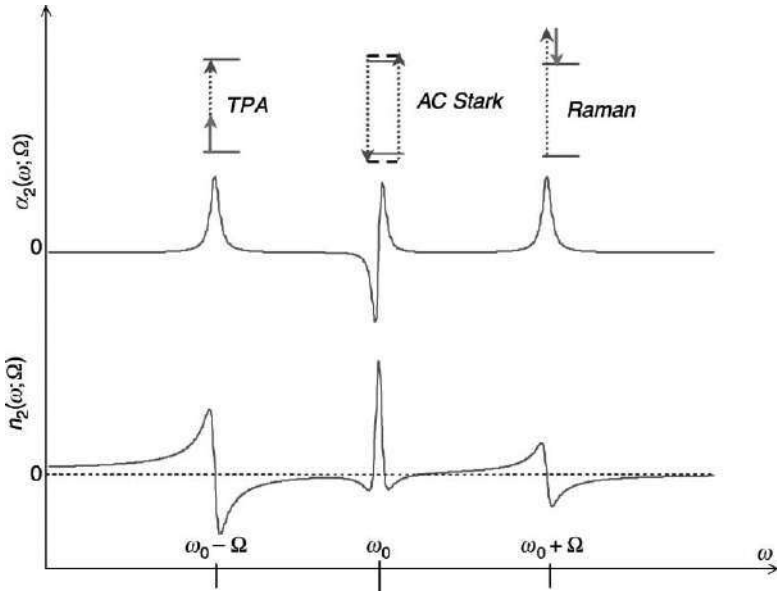

Figure 2 Upper trace: the nonlinear absorption coefficient in a fictitious 'degenerate' two-level system. Lower trace: the resulting nonlinear refractive index obtained using the KK relations. The insets show the three possible physical mechanisms involved.

By separating the real and imaginary parts of this equation, we get the generalized Kramers–Krönig relation pairs for a nondegenerate, nth-order nonlinear susceptibility:

$$\Re e\{\chi^{(n)}(\omega_1, \omega_2, \ldots, \omega_j, \ldots, \omega_n)\} = \frac{1}{\pi}\wp\int_{-\infty}^{\infty}\frac{\Im m\{\chi^{(n)}(\omega_1, \omega_2, \ldots, \omega', \ldots, \omega_n)\}}{\omega' - \omega_j}\mathrm{d}\omega' \quad [19]$$

and

$$\Im m\{\chi^{(n)}(\omega_1, \omega_2, \ldots, \omega_j, \ldots, \omega_n)\} = -\frac{1}{\pi}\wp\int_{-\infty}^{\infty}\frac{\Re e\{\chi^{(n)}(\omega_1, \omega_2, \ldots, \omega', \ldots, \omega_n)\}}{\omega' - \omega_j}\mathrm{d}\omega' \quad [20]$$

In particular, for $\chi^{(3)}$ processes having $\omega_1 = \omega_a, \omega_2 \omega_b$, and ${}_3b$, this becomes identical to eqn [16].

Note that in describing the nonlinear susceptibilities, no special attention was given to the harmonic generating susceptibility $\chi^{(N)}(N\omega) \equiv \chi^{(N)}(\omega, \omega, \ldots \omega)$, i.e., the susceptibility generating the Nth harmonic at $N\omega$. It turns out that in addition to the KK relations given by eqns [19] and [20], the real and imaginary parts of $\chi^{(N)}(N\omega)$ can also be related in different sets of dispersion integrals that involve only the degenerate forms of the susceptibilities. A more general yet simple analysis gives the most general form of KK relations for any type of $\chi^{(n)}$:

$$\chi^{(n)}(\omega_1 + p_1\omega, \omega_2 + p_2\omega, \ldots, \omega_m + p_m\omega) = \frac{1}{\mathrm{i}\pi}\wp\int_{-\infty}^{\infty}\frac{\chi^{(n)}(\omega_1 + p_1\Omega, \omega_2 + p_2\Omega, \ldots, \omega_m + p_m\Omega)}{\Omega - \omega}\mathrm{d}\Omega \quad [21]$$

for all $p_1, p_2, \ldots, p_m \geq 0$. Setting $\omega_1 = \omega_2 = \cdots = \omega_m \equiv 0$, and $p_1 = p_2 = \cdots = p_m = 1$ in eqn [21] yields an interesting form of the KK relations for the Nth-harmonic susceptibilities:

$$\Re e\{\chi^{(N)}(N\omega)\} = \frac{1}{\pi}\wp\int_{-\infty}^{\infty}\frac{\Im m\{\chi^{(N)}(N\omega')\}}{\omega' - \omega}\mathrm{d}\omega' \quad [22]$$

These dispersion relations have allowed calculations of $\chi^{(2)}(2\omega)$ and $\chi^{(3)}(3\omega)$ in semiconductors using full band structures.

At the beginning of this article, it was noted that all the KK relations for nonlinear optics were known in the early days of the field. Their application in unifying nonlinear absorption (in particular two-photon absorption) and the optical Kerr effect (n_2) in solids came only much later. More recent work demonstrated that the KK relations are a powerful analytical tool in nonlinear optics. Following the picture of a degenerate two-level system shown in Figure 2, a simple two-band model has been used to calculate the nonlinear absorption coefficient, $\alpha_2(\omega_1; \omega_2)$, resulting from three mechanisms: 2PA, the Raman absorption process, and the ac Stark effect. The optical Kerr coefficient $n_2(\omega_1; \omega_2)$ was then calculated using eqn [15]. Of particular practical interest is the degenerate case ($\omega_1 = \omega_2 = \omega$), from which the 2PA coefficient $\beta(\omega) = \alpha_2(\omega; \omega)$ can be extracted. Figure 3 depicts the calculated dispersion of n_2 and β as a function of $\hbar\omega/E_g$ where E_g is the bandgap energy of the solid. The dispersion of n_2 and its sign reversal shown in Figure 3 has been observed experimentally in many optical solids.

Finally, let us discuss a related implication of causality in nonlinear optics. The KK dispersion relations are traditionally derived in terms of internal material parameters such as susceptibility, absorption coefficient, and refractive index. Similar to the case of electrical circuits, one can obtain dispersion relations that apply to an external transfer function of the system that relates an input signal to an output signal. In this case, the dispersion of the transfer function includes system structure as well as the intrinsic dispersion of the material. As an optical (and linear) example, consider a Fabry–Perot etalon. The optical transmission of this system has well-known spectral

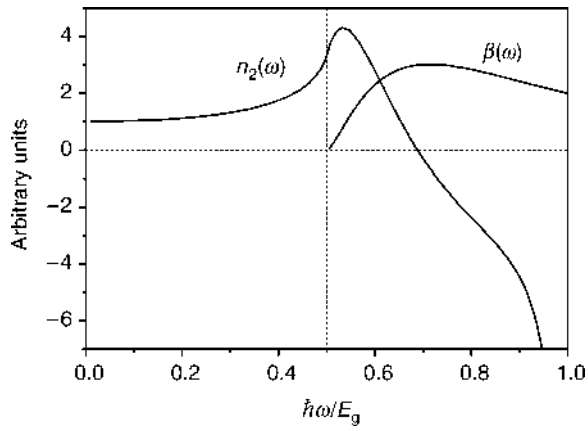

Figure 3 The two-photon absorption coefficient in semiconductors (β) calculated for a two-band model. The resultant nonlinear refractive index (n_2) obtained using a KK transformation of the calculated nondegenerate nonlinear absorption coefficient includes all major mechanisms.

features that are primarily caused by structural dispersion (i.e., interference) in addition to the intrinsic dispersion of the material. Causality still demands that the transmitted signal has a phase variation whose value and dispersion can be determined using a KK relation linking the real and imaginary parts of the transmission coefficient. In other words, the KK relations provide a spectral correlation between the real and imaginary components of the transfer function which in turn may translate to a spectral correlation between the phase and amplitude of the transmitted signal. However, the variations in phase do not necessarily imply the presence of a varying index of refraction, nor does an amplitude variation suggest the existence of material absorption (dissipation). Ultimately, this implies that any mechanism causing a variation in amplitude (including reflection, scattering, or absorption) must be accompanied by a phase variation. (One should note that the reverse of the previous statement is not necessarily true; i.e., a variation in phase does not have to be accompanied by an amplitude modulation.)

In nonlinear optics with the 'black box' approach of Figure 1, the optical perturbation ξ (with frequency Ω) can render an amplitude variation in the probe (at ω) using various frequency mixing schemes in a noncentrosymmetric material (i.e., with nonzero $\chi^{(2)}$). For instance, the probe at ω can be depleted by nonlinear conversion to $\omega_{\text{sum}} = \omega + \Omega$ via sum-frequency generation involving $\chi^{(2)}(\omega, \Omega)$ and/or to $\omega_{\text{diff}} = \omega - \Omega$ via difference-frequency generation involving $\chi^{(2)}(\omega, -\Omega)$. Such a conversion (or depletion) should be accompanied by a phase variation according to the KK dispersion relations. This type of nonlinear phase modulation is known as a $\chi^{(2)}$:$\chi^{(2)}$ cascaded nonlinearity. Such cascaded processes are routinely (and more simply) analyzed with Maxwell's equations governing the propagation of beams in a second-order nonlinear material. The KK relations, however, provide an interesting physical perspective of the process. We find that cascaded second-order nonlinearities are yet another manifestation of causality in nonlinear optics.

List of Units and Nomenclature

α	linear absorption coefficient
α_2	nonlinear absorption coefficient
β	two-photon absorption coefficient
$\chi^{(n)}$	nth-order nonlinear optical susceptibility
n	linear refractive index
n_2	nonlinear refractive index coefficient, coefficient of optical Kerr effect
$\Theta(t)$	step function
$\wp$	principal value
2PA	two-photon absorption
KK relations	Kramers–Krönig relations

See also

Materials for Nonlinear Optics: Liquid Crystals for NLO.
Nonlinear Optics, Basics: Cascading; Nomenclature and Units.

Further Reading

Bassani F and Scandolo S (1991) Dispersion-relations and sum-rules in nonlinear optics. *Physical Review B–Condensed Matter* 44: 8446–8453.

Caspers PJ (1964) Dispersion relations for nonlinear optics. *Physical Review A* 133: 1249–1251.

Hutchings DC, Sheik-Bahae M, Hagan DJ and Van Stryland EW (1992) Kramers–Krönig relations in nonlinear optics. *Optical and Quantum Electronics* 24: 1–30.

Kogan SM (1963) On the electromagnetics of weakly nonlinear media. *Soviet Physics JETP* 16: 217–219.

Nussenzveig HM (1972) Causality and Dispersion Relations. New York: Academic Press.

Price PJ (1964) Theory of quadratic response functions. *Physical Review* 130: 1792–1797.

Ridener FLJ and Good RHJ (1975) Dispersion relations for nonlinear systems of arbitrary degree. *Physical Review B* 11: 2768–2770.

Sheik-Bahae M and Van Stryland EW (1999) Optical nonlinearities in the transparency region of bulk semiconductors. In: Garmire E and Kost A (eds) *Nonlinear Optics in Semiconductors I*, 58, pp. 257–318. Academic Press.

Sheik-Bahae M, Hutchings DC, Hagan DJ and Van Stryland EW (1991) Dispersion of bound electronic nonlinear refraction in solids. *IEEE Journal of Quantum Electronics* 27: 1296–1309.

Smet F and Vangroenendael A (1979) Dispersion-relations for N-order non-linear phenomena. *Physical Review A* 19: 334–337.

Toll JS (1956) Causality and the dispersion relation: logical foundations. *Physical Review* 104: 1760–1770.

Nomenclature and Units

M P Hasselbeck, The University of New Mexico, Albuquerque, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Nomenclature Associated with the Excitation Light

Particular care must be used when characterizing the excitation beam in nonlinear optical experiments compared to linear measurements. By definition, nonlinear optical phenomena depend on the electric field to high order. The higher the order, the more sensitive the observed behavior depends on the input. Extracting a representative nonlinear coefficient from an experiment, for example, becomes progressively more difficult as the order of the optical nonlinearity increases. In other words: the errors associated with optical beam characterization get magnified by the order of the nonlinearity under study.

There is an extensive nomenclature for characterizing the light (almost always laser light) interacting with the nonlinear optical medium. Different descriptions may be used depending on whether the excitation light is pulsed, continuous, or a continuous train of pulses. When a laser beam is constant or continuous, it is often described by its 'cw power'. 'Cw' stands for 'continuous wave', an acronym taken from the nomenclature of electronics. The cw power of a laser is determined by placing a power meter in the path of a beam. Repetitively pulsed lasers can be characterized in the same way, provided the response time of the power meter is slower than the pulse separation period. This will almost certainly be the case with a continuously pumped mode-locked laser such as a dye laser, fiber laser, or Ti:sapphire laser, which produce pulses at repetition frequencies of tens of megahertz. Cw power may also be suitable for describing pulsed flashlamp lasers, gas discharge lasers, or any laser that is excited in a periodic fashion.

When the optical output can be distinguished as individual pulses, additional metrics are used. A pulse, by definition, exists during a window of time, i.e., there is a time when light is present and a time when light is absent. How one characterizes the time light is present – the pulse duration – is critically important and not always obvious. There are a myriad of ways the pulse temporal envelope can manifest itself; common examples include square pulses, triangular pulses, Gaussian pulses, and hyperbolic secant pulses. These names refer to the mathematical waveforms that map the pulse envelope as a function of time.

A pulse waveform that exhibits symmetry about the peak in its temporal envelope is commonly characterized by its 'full-width, half-maximum', abbreviated FWHM. One obtains the full-width, half-maximum by locating the two points on the pulse profile that are at half the peak value. The temporal separation of these two points is the FWHM. One uses this measure because in a strict mathematical sense, waveforms such as the Gaussian or hyperbolic-secant exist even at times $t = \pm\infty$. A less common term is the 'half-width, half-maximum' or HWHM. As the name implies, the HWHM is exactly half the FWHM value. Not all light-pulses are temporally symmetric, however, so greater detail may be needed when giving a mathematical description of asymmetric pulses.

Light pulses are also characterized by their energy and peak power. The temporal envelope recorded by a so-called square-law detector (all laboratory detectors are square-law detectors) shows the power of the pulse as a function of time, provided the detector response time is sufficiently fast. Integrating this waveform gives the pulse energy. It makes no sense to talk about the energy of a cw beam of light, unless that beam is composed of repetitive, distinguishable pulses. Each pulse in the train carries a distinct amount of optical energy.

'Ultrafast' or 'ultrashort' laser pulses generally refer to light pulses that are of such small duration they cannot be measured directly with detectors and oscilloscopes because of bandwidth limitations. To infer the pulse duration, so-called intensity autocorrelation measurements are often made. The temporal power profile is then deduced from the autocorrelation signal with the appropriate mathematical conversion factor. Equally important is the spectrum of a short pulse. The optical bandwidth determines the lower limit of temporal compression; such ideally compressed pulses are said to be 'transform limited'. If different components of the spectrum can be distinguished at specific positions on the pulse temporal profile, then the pulse is 'chirped'. Information on the methods of ultrafast laser science can be found in **Ultrafast Laser Techniques:** Generation of Femtosecond Pulses.

Nonlinear optical effects take place when light is concentrated on a target – there is a cross-section or 'footprint' of the beam on the medium.

The interaction of the light and material takes place in a region called the beam area, beam cross-section, focal area, or focal volume. One must then define the appropriate interaction area or volume.

Specification of the beam area allows one to define two important quantities used frequently in nonlinear optics. The power/area is known as irradiance or power density. This quantity is commonly called the intensity, although the strict radiometric definition of intensity is power/(solid angle). The second important quantity is energy/area, which is called the energy density or fluence.

The convention in nonlinear optics is to define the optical electric field as:

$$\mathbf{E}(\mathbf{z}, t) = \frac{1}{2} E_0 \exp[\mathrm{i}(\mathbf{k}\cdot\mathbf{z} - \omega t)] + \text{c.c.} \qquad [1]$$

where E_0 is the peak field, $\mathbf{k}$ is the propagation vector, ω is the angular frequency of the light, and c.c. stands for complex conjugate. This represents a forward- and backward-propagating (in the z-direction) infinite, transverse plane wave. Other definitions are also used. In SI/mks units and using the convention of eqn [1], the irradiance (I in units of W/m^2) inside a material of refractive index n is related to the electric field vector of the light ($\mathbf{E}$ in units of V/m) as follows:

$$I = \frac{1}{2} c n \varepsilon_0 |\mathbf{E}|^2 \qquad [2]$$

Here c is the speed of light (2.998×10^8 m/s) and ε_0 is the permittivity of free space (8.854×10^{-12} F/m). Nonlinear optics has an unfortunate tradition of mixing mks and cgs units, resulting in a lot of confusion. Irradiance is usually expressed in units of W/cm^2. One can calculate the peak electric field (in units of V/cm) of a laser beam given its irradiance (in units of W/cm^2) by using this simple formula:

$$E_0 = 38.82\sqrt{\frac{I}{n}} \left(\frac{\mathrm{V}}{\mathrm{cm}}\right) \qquad [3]$$

The root-mean-square value of the electric field is obtained by replacing the factor 38.82 by 27.45. In Gaussian/cgs units, irradiance is expressed in units of ergs/(cm^2 sec) and is related to the field as:

$$I = \frac{cn}{8\pi} |\mathbf{E}|^2 \qquad [4]$$

where the field is in units of statvolt/cm and $c = 2.998 \times 10^{10}$ cm/sec. Conversion between mks and cgs for the electric field is 3×10^4 V/m = 1 statvolt/cm; irradiance is converted using 1 erg/(cm^{2}sec) = 1×10^{-3}W/m^2. The following point must be emphasized: depending on how the electric field is defined, there can be factors of 2 or even 4 discrepancies in the irradiance values quoted by different authors. The common definitions are used here, although they are certainly not universal.

A very common realization of the spatial irradiance profile of a laser beam is the radially symmetric Gaussian function:

$$I(r) = I_0 \exp\left(-\frac{2r^2}{w^2}\right) \qquad [5]$$

where I_0 is the peak irradiance at the center of the beam ($r = 0$) and w is the 'spot size'. The spot size changes continuously as the beam propagates and the minimum spot size is known as the waist. One often hears the spot size referred to as the radius, but the radius of a Gaussian beam traditionally denotes the curvature of the phase front. By definition, the phase front radius is infinite at the waist. The factor of 2 appearing in the argument of the exponential in eqn [5] stems from the fact that the spot size w is conventionally defined for the electric field of the Gaussian beam. When the field is squared to obtain the irradiance, the factor of 2 appears. Using this function for the irradiance spatial profile leads to the common parameter '1/e^2 diameter'. The irradiance falls to 1/e^2 (0.1353) of its peak value when $r = w$.

The Gaussian spatial profile or 'Gaussian beam' results when the laser has been designed to operate in the lowest-order transverse mode, usually denoted TEM$_{00}$. The TEM$_{00}$ Gaussian beam is particularly convenient because the same relative power profile is maintained in the near- and far-field, whether or not the beam is collimated or focused.

The cross-sectional area of a TEM$_{00}$ Gaussian beam normally incident on a surface is found as follows. The power in the beam is obtained by integrating the irradiance profile (I) over the surface:

$$P = \iint_{\pm\infty} I \mathrm{d}A \qquad [6]$$

Referring to eqn [5], the integral becomes:

$$P = \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} r\,\mathrm{d}r\, I_0 \exp\left(-\frac{2r^2}{w^2}\right) = I_0 \frac{\pi w^2}{2} \qquad [7]$$

Hence the effective area of a TEM$_{00}$ Gaussian beam normally incident on a surface is:

$$A = \frac{\pi w^2}{2} \qquad [8]$$

There are many situations where the Gaussian formulation of eqn [5] is not suitable. Beams produced by unstable laser resonators, for example, are not Gaussian. A multitransverse mode beam profile obtained from a stable resonator does not usually lend itself to a simple mathematical characterization.

Nomenclature Associated with the Nonlinear Optical Medium

In nonlinear optics texts, the subject is often introduced by writing the macroscopic polarization in Maxwell's equations as a power series expansion in the electric field. This approach, proposed by Bloembergen and coworkers in the early 1960s, has been spectacularly successful for interpreting experiments, though it has also led to confusion in the definition of nonlinear optical coefficients. The confusion, which stems from arbitrary definitions and nomenclature used by different authors, cannot be resolved here. The reader should, however, be alert for these discrepancies. Comparison of results published by different laboratories requires that the fundamental equations for extracting the nonlinear coefficients from their data are known. As the discipline has matured, the nonlinear optics community largely recognized the ambiguities and confusion; precise definitions of terms and coefficients are now commonly provided in the research literature.

It should also be noted that the power-series framework may not always be the best formulation for modeling and characterizing nonlinear optical effects. A carefully designed optical limiter, for example, may exhibit an abrupt decrease of transmission at a specific 'threshold irradiance' or 'threshold fluence', perhaps at a given laser pulse duration. The physical phenomenon or phenomena driving this behavior may not readily succumb to characterization by an nth-order coefficient in a power series expansion. In this situation it may be appropriate to specify a threshold optical input parameter. Another example is a homogeneously broadened saturable absorber, in which the absorption of an optical medium decreases with increasing input irradiance. The irradiance-dependent absorption coefficient α (refer to eqn [26]) is characterized by a 'saturation irradiance' I_{sat}:

$$\alpha(I) = \frac{\alpha_0}{1 + \dfrac{I}{I_{\text{sat}}}} \qquad [9]$$

where α_0 is the linear absorption coefficient.

Nonlinear Susceptibility

The power-series expansion of the macroscopic polarization is the standard approach for modeling nonlinear optical behavior and categorizing the various phenomena. A common way to write the polarization, nonlinear in the electric field, is (mks units):

$$\mathbf{P} = \varepsilon_0\left(\chi^{(1)}\cdot\mathbf{E} + \chi^{(2)} : \mathbf{EE} + \chi^{(3)}\vdots\mathbf{EEE} + \ldots\right) \qquad [10]$$

where the polarization $\mathbf{P}$ is a time- and space-dependent vector and the terms $\chi^{(n)}$ are the various orders of the nonlinear susceptibility (the cgs equation is obtained by dropping the free-space permitivitty coefficient ε_0). A second-order effect is associated with $\chi^{(2)}$, a third-order with $\chi^{(3)}$, and so on. In general, there are distinguishable electric field vectors (in eqn [10], the fields have not been individually designated for clarity). The susceptibility terms are generally tensors, which means the medium is sensitive to the orientation of the input fields. The dots in eqn [10] indicate tensor products. The input field vectors can have different frequencies and the presence of complex conjugates in eqn [1] indicates that the frequency components will have both + and − signs, i.e. the jth electric field term will have associated frequency factors $\exp(\pm \mathrm{i}\omega_j t)$.

The nomenclature used here can be illustrated by example. In the case of the second-order nonlinear polarization (mks units):

$$\mathbf{P}_j = \varepsilon_0 \chi^{(2)}_{jkl} : \mathbf{E}_k \mathbf{E}_l \qquad [11]$$

Note that subscripts have been introduced. The indices correspond to Cartesian space vectors. This equation says that the polarization in the j-direction results from the tensor product of the appropriate $\chi^{(2)}$ with the input fields at $\mathbf{E}_k$ and $\mathbf{E}_l$. The jth component of polarization has an oscillation frequency that is determined by mixing of the input fields. In this example, the indices j, k and l can each take the value of x, y, and z. Let each input field be at one of two possible frequencies, call them ω_1 and ω_2. There are an enormous number of permutations (81 to be exact) of eqn [11]. One term is:

$$P_x(\omega_1 + \omega_2) = \varepsilon_0 \chi^{(2)}_{xzy}(\omega_1 + \omega_2) E_z(\omega_1) E_y(\omega_2) \times \exp[-\mathrm{i}(\omega_1 + \omega_2)t] \qquad [12]$$

Another possibility is:

$$P_z(\omega_2 - \omega_1) = \varepsilon_0 \chi^{(2)}_{zyx}(\omega_2 - \omega_1) E_y(\omega_2) E_x(-\omega_1) \times \exp[-\mathrm{i}(\omega_2 - \omega_1)t] \qquad [13]$$

Equation [12] corresponds to sum frequency generation ($\omega_1 + \omega_2$) and eqn [13] shows difference frequency generation ($\omega_2 - \omega_1$).

Not explicitly written on the right-hand side of the above equations is the exponential containing the mixing of the propagation vectors. To maximize the nonlinear polarization, which is the source of nonlinear behavior in Maxwell's equations, the vector sum $\sum \mathbf{k}_j$ for *all* the interacting fields should be close to zero. Arranging the propagation vectors to accomplish this is known as 'phase matching'.

It is important to point out that there are eight more terms describing the nonlinear polarization at $P_x(\omega_1 + \omega_2)$ in addition to eqn [12]; likewise for $P_z(\omega_2 - \omega_1)$. Also note that the indices k and l may be identical, hence a nonlinear polarization driven by $\chi^{(2)}_{xyy}$ for example, is allowed. Higher-order nonlinear polarizations get progressively more complicated. The general third-order nonlinear polarization is

$$\mathbf{P}_j = \varepsilon_0 \chi^{(3)}_{jklm} \vdots \mathbf{E}_k \mathbf{E}_l \mathbf{E}_m \qquad [14]$$

Sometimes the second-order nonlinear susceptibility is written with the coefficient d instead of $\chi^{(2)}$. The relation between the d-coefficient and $\chi^{(2)}$ depends arbitrarily on how it is defined; a common definition is

$$d_{jkl} = \frac{1}{2}\chi^{(2)}_{jkl} \qquad [15]$$

but the reader is cautioned that the factor $\frac{1}{2}$ is sometimes missing. In the preceding discussion it was pointed out how there can, in principle, be a large number of tensor components involved in the expansion of the nonlinear polarization. Simplification occurs when it is realized that there is no discernible physical difference between the frequency terms ω_j and $-\omega_j$ and that ordering of the fields in eqn [14] – which suggests a time order in the arrival of the fields – is irrelevant. These symmetry arguments show that $d_{jkl} = d_{jlk}$, which reduces the number of independent d-coefficients from 81 to 18. A simpler subscript notation is then introduced that uses the integers 1–6 to represent pairs of Cartesian components k and l. This reduces the number of subscripts from three to two. An example of this notation is $d_{xzz} = d_{14}$.

In certain situations, one can take advantage of crystal symmetry to further reduce the complicated summations to a single scalar coefficient for the nonlinear susceptibility, which is referred to as 'd-effective'. For these specialized cases, the nonlinear polarization is

$$P = \varepsilon_0 d_{\text{eff}} E_1 E_2 \qquad [16]$$

In SI/mks units, the polarization is in units of coul/m^2. The second-order susceptibility (i.e. $\chi^{(2)}$, d_{jkl}) must therefore have units of m/volt. The units of $\chi^{(3)}$ are m^2/volt^2, $\chi^{(4)}$ is m^3/volt^3, and so on. In Gaussian/cgs units, $\chi^{(2)}$ has dimensions cm/statvolt, $\chi^{(3)}$ will be cm^2/statvolt^2, etc. although sometimes in the Gaussian system *all* the coefficients $\chi^{(n)}$ are discussed with shorthand 'electrostatic units' or 'esu'.

Complex Quantities

In linear optics, the susceptiblity has real and imaginary parts

$$\chi^{(1)} = \chi^{(1)}_{\text{real}} + \mathrm{i}\chi^{(1)}_{\text{imaginary}} \qquad [17]$$

The complex linear index is, in turn, written as the sum of real and imaginary components, derived from the linear susceptibility as follows (mks units):

$$\sqrt{1 + \chi^{(1)}} = n_0 + \mathrm{i}\kappa \qquad [18]$$

where n_0 is the linear refractive index and κ is the imaginary term leading to absorption of light. In general, the nonlinear susceptibilty $\chi^{(n)}$ is also a complex number:

$$\chi^{(n)} = \chi^{(n)}_{\text{real}} + \mathrm{i}\chi^{(n)}_{\text{imaginary}} \qquad [19]$$

As in linear optics, the complex notation is a bookkeeping method that conveniently accounts for what is known as 'resonant enhancement'. Nonlinear optics research has revealed that the nonlinear susceptibility $\chi^{(n)}$ can be a strong function of frequency. Specifically, this quantity will be strongly enhanced when sums and/or differences of the photon energies in the interacting light beams coincide with quantum mechanical energy resonances in the material. When a resonance condition is achieved, $\chi^{(n)}$ will be dominated by its imaginary component. Far from the resonances, $\chi^{(n)}$ behaves more like a real quantity. These complex terms directly enter the wave equations describing how light propagates in a nonlinear medium and are particularly important for $\chi^{(3)}$; the complex quantity $\chi^{(3)}$ determines whether light will be refracted or absorbed and by how much. The real part of $\chi^{(3)}$ drives nonlinear refraction while the imaginary portion characterizes nonlinear absorption (e.g., two-photon absorption) and the inverse effect – gain. This rich subject – in particular the critical importance of the wavelength dependence of the nonlinear phenomena – is considered in **Nonlinear Optics, Basics:** Kramers–Krönig Relations in Nonlinear Optics.

Nonlinear Refraction

The most common manifestation of nonlinear refraction arises from the third-order nonlinear susceptibility, the so-called optical Kerr effect. In this case, the refractive index is linearly proportional to the irradiance (I) of a monochromatic light beam. The irradiance-dependent refractive index is

$$n(I) = n_0 + n_2 I \qquad [20]$$

where n_0 is the linear, irradiance-independent refractive index and n_2 is the nonlinear refractive index coefficient. Because n is a dimensionless quantity, n_2 must be in units of area/power.

The optical Kerr effect is usually written in mks units with the expression shown in eqn [20]. In cgs units, the common (but by no means universal) convention is to write:

$$n(E) = n_0 + \frac{1}{2}\tilde{n}_2|\mathbf{E}|^2 \qquad [21]$$

where the field is in units of statvolts/cm. The nonlinear coefficient $\tilde{n}_2$ (cgs) must therefore be in units of cm^2/statvolt^2, which is sometimes abbreviated to 'esu'. The conversion for n_2 between these two equations is:

$$\tilde{n}_2(\text{cgs}) = \frac{n_0 c}{40\pi} n_2(\text{mks}) \qquad [22]$$

where n_2(mks) is in units of m^2/W and c is in m/sec. Also useful is the relation between n_2 and $\chi^{(3)}$, written in mks units as:

$$\Delta n = n_2 I = \frac{\mathrm{Re}\left(\chi^{(3)}\right)}{4\varepsilon_0 n^2 c} I \qquad [23]$$

where 'Re' denotes the real part of $\chi^{(3)}$. In cgs units we have:

$$\Delta n = \frac{1}{2}\tilde{n}_2|\mathbf{E}|^2 = \frac{4\pi^2 \mathrm{Re}\left(\tilde{\chi}^{(3)}\right)}{n^2 c} I \qquad [24]$$

where I is in units of ergs/(cm^2 sec) defined by eqn [4]. The reader is again advised that different authors will show discrepancies of 2, 4, or even 8 when writing these equations. The susceptibilities $\chi^{(3)}$ are related as follows:

$$\tilde{\chi}^{(3)}(\text{cgs}) = \frac{9\times 10^8}{4\pi}\chi^{(3)}(\text{mks}) \qquad [25]$$

If the coefficients n_0 and n_2 exist, does the nomenclature imply there are terms n_1, n_3, and others? These coefficients are certainly allowed, but not often seen in discussions of nonlinear refraction. When one refers to nonlinear refraction, the common understanding is that the index depends linearly on irradiance, which is conveniently modeled by eqn [20]. But there are other physically relevant situations to consider. The coefficient n_1, for example, describes the linear electro-optic effect in which the change in index is linearly proportional to the electric field (although it is called the linear electro-optic effect, it is actually derived from the second-order nonlinear susceptibility). The electric field can be the oscillating field of a laser beam, for example, or a dc field applied to an electro-optic crystal (e.g., Pockel's cell). When other terms are added to eqns [20] or [21], the units of the coefficients must be chosen to keep the equation dimensionless.

In the preceding discussion, there is an implication that the nonlinear index is an instantaneous function of the irradiance or field. Although the material system can never respond instantaneously, this is a good approximation in many situations. Sometimes it is not. Consider a pulsed laser beam that heats the material. When the local temperature increases, the linear refractive index can change. A short, Q-switched laser can have a pulse duration far less than a microsecond, while the temperature change it induces can last many orders of magnitude longer. This means optical modification of the refractive index may persist long after the exciting pulse has vanished, i.e., when the irradiance is at zero.

Another example is when a laser beam promotes electrons from their ground state to higher energy states of the material system. These excited electrons may modify the refractive index of the material. In general, the excited electrons remain in high-energy states for some period of time before relaxing to the ground level – if this time is longer than the excitation, the change of the refractive index persists beyond the time the laser beam is present. In these situations, the model of nonlinear refraction suggested by the above equations is inaccurate. One cannot obtain the nonlinear refraction from a simple algebraic analysis. A system of time-dependent equations – describing the dynamics of excitation and relaxation – must be solved using numerical procedures. Such phenomena are sometimes loosely categorized as 'dynamic linear optical effects' or 'effective third-order nonlinearities'.

Nonlinear Absorption

Consider a single-frequency light beam passing through an optically absorbing region of length L.

For simplicity, neglect reflections caused by surfaces that may define the region of interest, i.e., ignore reflections at surfaces that may be located on the optical axis z at points $z = 0$, $z = L$, or any other point in the path. Linear absorption means that the optical power extracted from the light beam as it traverses the absorbing medium is a direct function of the power at a given point. This is described mathematically by an elementary, linear differential equation known as Beer's law:

$$\frac{\mathrm{d}}{\mathrm{d}z} I(z) = -\alpha I(z) \qquad [26]$$

The constant of proportionality is α, which is the linear absorption coefficient. Equation [23] is solved by direct integration:

$$\int_{I(0)}^{I(L)} \frac{\mathrm{d}I}{I} = -\alpha \int_0^L \mathrm{d}z \qquad [27]$$

This has the solution:

$$\frac{I(L)}{I(0)} = \exp(-\alpha L) \qquad [28]$$

which means the irradiance decreases exponentially as a function of propagation distance in a linearly absorbing medium.

In the nonlinear regime, we don't expect a classical Beer's law model to hold. By definition, the absorption will be a nonlinear function of irradiance at a given point. One makes the following power-series expansion to describe nonlinear absorption of monochromatic light:

$$\frac{\mathrm{d}}{\mathrm{d}z} I = -\alpha I - \beta I^2 - \gamma I^3 - \cdots \qquad [29]$$

The coefficient β corresponds to a two-photon absorption process and γ is the coefficient of three-photon absorption. What about four-photon and even higher-order processes? These are rarely encountered, but certainly possible. To make these distinctions, eqn [29] is sometimes written with coefficients K_i or α_i instead of the sequential Greek alphabet:

$$\frac{\mathrm{d}}{\mathrm{d}z} I = -K_1 I - K_2 I^2 - K_3 I^3 - K_4 I^4 - \cdots \qquad [30]$$

The units of the various absorption coefficients must maintain the dimensional consistency of eqns [29] and [30], which is irradiance/length or power/(length)^3. These are shown in Table 1.

If there is linear absorption, the nonlinear processes at that wavelength are often (but not always) negligibly weak. This means we have the following inequalities:

Table 1 Dimensions of optical absorption coefficients

Process	*Coefficient*	*Units*
Linear absorption	α, K_1	(length)$^{-1}$
Two-photon absorption	β, K_2	length/power
Three-photon absorption	γ, K_3	length^3/power^2
Four-photon absorption	K_4	length^5/power^3

$$K_1 I \gg K_i I^i, \qquad \text{where } i \geq 2 \qquad [31]$$

Nonlinear absorption is generally observed in the wavelength region where the medium is transparent to low-irradiance light, i.e., where linear absorption is negligible. It should be emphasized that there are situations where linear absorption is large and in fact a crucial component of the aggregate nonlinear effect. We will return to this point later in the discussion. For the moment, we neglect linear absorption. If the energy of the incident photons and energy levels of the system permit it, the lowest-order nonlinear process is two-photon absorption, described by the equation

$$\frac{\mathrm{d}}{\mathrm{d}z} I = -K_2 I^2 \qquad [32]$$

which can be solved by elementary integration:

$$\frac{I(L)}{I(0)} = \frac{1}{1 + I(0) K_2 L} \qquad [33]$$

Unfortunately, the situation is rarely this convenient. The above integration has ignored the fact that nonlinear absorption may enable significant linear absorption. This is possible because nonlinear absorption promotes electrons from low- to high-energy states in the medium. The change in excited electron density associated with absorption of light (both linearly and nonlinearly) is called 'photocarrier generation'. These photocarriers (e.g., photo-electrons) may modify the linear absorption as well as the refractive properties of the material. If two-photon absorption causes the linear absorption to increase, for example, the assumption that allowed us to ignore K_1 when writing eqn [32] ceases to be valid. While eqn [32] was appropriate at the very start of the light–matter interaction, the generation of photocarriers after the passage of time may change that condition. The behavior of the system is therefore time-dependent, i.e., it is dynamic. Simple analytic solutions are almost

always not possible when dealing with nonlinear absorption.

The time rate of change of the photocarrier population (N) is modeled by the following equation:

$$\frac{\partial}{\partial t}N = \frac{K_1 I}{\hbar\omega} + \frac{K_2 I^2}{2\hbar\omega} + \frac{K_3 I^3}{3\hbar\omega} + \frac{K_4 I^4}{4\hbar\omega} + \cdots - \text{recombination} - \text{diffusion} \quad [34]$$

where ω is the angular frequency ($\omega = 2\pi f$) of the incident light. Also included are place-holders for loss processes such as recombination and diffusion, which may themselves have complicated mathematical descriptions. The coefficients K_i have exactly the same units, dimensions, and interpretation as in eqn [30] and Table 1; in the spirit of the preceding discussion, these coefficients may be time dependent. Note that the dimensions of eqn [30] describe absorption of light (irradiance/length), while eqn [34] models the photocarrier population density (length^{-3} time^{-1}). The units of these two equations are very different.

List of Units and Nomenclature

CARS	Coherent anti-Stokes Raman scattering
c.c.	Complex conjugate
CSRS	Coherent Stokes Raman scattering
DFG	Difference frequency generation
DFM	Difference frequency mixing
DFWM, FWM	(Degenerate) Four wave mixing
DRO	Doubly resonant OPO
EIT	Electromagnetically induced transparency
ESA	Excited state absorption
GVD	Group velocity dispersion
GVM	Group velocity mismatch
NLA	Nonlinear absorption
NLR	Nonlinear refraction
OFID	Optical free-induction decay
OPA	Optical parametric amplification/amplifier
OPG	Optical parametric generation/generator
OPL	Optical power limiter
OPO	Optical parametric oscillation/oscillator
2PA, TPA	Two-photon absorption
PCM	Phase conjugate mirror
QPM	Quasi phase matching
RIKES	Raman-induced Kerr effect spectroscopy
RSA	Reverse saturable absorption
SBS	Stimulated Brillouin scattering
SFG	Sum frequency generation
SHG	Second harmonic generation
SIT	Self-induced transparency
SRO	Singly resonant OPO
SRS	Stimulated Raman scattering
SRWS	Stimulated Rayleigh wing scattering
SVEA, SVAA	Slowly varying envelope (amplitude) approximation
THG	Third harmonic generation
TPF	Two-photon fluorescence

See also

Fiber and Guided Wave Optics: Nonlinear Optics. **Nonlinear Optics, Basics:** Ultrafast and Intense-Field Nonlinear Optics. **Spectroscopy:** Nonlinear Laser Spectroscopy.

Further Reading

Bloembergen N (1992) *Nonlinear Optics*. Redwood City, CA: Addison-Wesley.

Boyd RW (1991) *Nonlinear Optics*. Boston, MA: Academic Press.

Levenson MD and Kano S (1988) *Introduction to Nonlinear Laser Spectroscopy*. Boston, MA: Academic Press.

Miller A, Miller DAB and Smith SD (1981) Dynamic nonlinear optical processes in semiconductors. *Advances in Physics* 30: 697–800.

Mills DL (1998) *Nonlinear Optics: Basic Concepts*. New York: Springer.

Newell C and Moloney JV (1992) *Nonlinear Optics*. Redwood City, CA: Addison-Wesley.

Sauter EG (1996) *Nonlinear Optics*. New York: Wiley.

Shen YR (1984) *The Principles of Nonlinear Optics*. New York: Wiley.

Sutherland RL (1996) *Handbook of Nonlinear Optics*. New York: Marcel Dekker.

Yariv A (1991) *Optical Electronics*, 3rd edn. New York: Holt, Rinehart, and Winston.

Yariv A (1989) *Quantum Electronics*. New York: Wiley.

Zernike F and Midwinter JE (1973) *Applied Nonlinear Optics*. New York: Wiley.

Nonlinear Optical Phase Conjugation

B Y Zeldovich, University of Central Florida, Orlando, FL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Phase conjugation (PC) beams and their applications are best illustrated by Figure 1, which shows the passage of a coherent planewave incident upon a transparent optical element with inhomogeneities of refractive index (Figure 1a). Reversibility of light propagation implies the existence of such an 'anti-distorted' beam, which reverts back to a planewave after reversing through the same inhomogeneities (Figure 1b). A real-valued monochromatic optical wave, $E_{\text{real}}(\mathbf{R}, t)$, may be represented by the complex amplitude $E(\mathbf{R})$ via $E_{\text{real}}(\mathbf{R}, t) = 0.5[E(\mathbf{R})\exp(-i\omega t) + E^*(\mathbf{R})\exp(i\omega t)]$, where $E^*(\mathbf{R})$ represents the complex conjugate of $E(\mathbf{R})$:

$$E(\mathbf{R}) = |E(\mathbf{R})| \exp[i\varphi(\mathbf{R})]$$
$$E^*(\mathbf{R}) = |E(\mathbf{R})| \exp[-i\varphi(\mathbf{R})] \qquad [1]$$

Mathematical expression of time-reversal, $E_{\text{real}}(\mathbf{R}, t) \rightarrow E_{\text{real}}(\mathbf{R}, -t)$, becomes the exchange $E(\mathbf{R}) \leftrightarrow E^*(\mathbf{R})$ for monochromatic waves. The reversibility of propagation means that the complex propagating field $E(\mathbf{R})$ (for example, planewave $\exp(i\mathbf{k}\cdot\mathbf{R})$), is a wave equation solution equivalent to $E^*(\mathbf{R})$ (in this example, $\exp(-i\mathbf{k}\cdot\mathbf{R})$). Conjugation of complex amplitude means reversal of the sign of the phase, and a mixed label PC has been coined and nowadays is firmly established. Terms such as 'wave front reversal' and 'generation of time-reversed replica of the beam' are also used to describe PC.

Figure 1 also illustrates one of the most important applications of PC. If the element in question is a laser-type amplifier, then double-passage allows the extraction of energy from the optically inhomogeneous laser medium in the form of a perfectly collimated beam of diffraction-limited divergence. A PC device may also serve as a mirror of a laser resonator, resulting in the same beam-correction purpose.

One of the most practical and robust methods of PC is based on stimulated Brillouin back-scattering (SBS). The incident beam illuminates the SBS-active transparent medium, for example, liquids (CS_2, CCl_4, acetone, etc.), compressed gases (CH_4, SF_6, etc.), or solids (fused or crystalline quartz, glass, etc.). This 'pump' beam, $E(\mathbf{R}) \equiv E(\mathbf{r}, z)$, must possess well-developed transverse inhomogeneities of intensity $|E(\mathbf{r}, z)|^2$. Here, $\mathbf{r} = \{x, y\}$ is the part of coordinate vector transverse with respect to the central direction z of the beam in question, and $\mathbf{R} = \{\mathbf{r}, z\}$. These inhomogeneities may constitute a narrow focal waist in the case of weakly distorted focused beams, speckle-inhomogeneities in cases of strong distortions, or a combination thereof; see Figure 2, where solid lines symbolize the 'rays' of the incident pump.

Spontaneous scattering of the pump results in a multitude of possible transverse profiles $S(\mathbf{r}, z = 0)$ of the signal, whose 'rays' are depicted via the dotted lines on Figure 2. The signal is amplified exponentially due to SBS processes. This exponent in a simplified form may be represented as $|S(\mathbf{r}, z)|^2 \propto \exp(\int g dz)$, with the gain $g(z)$ being the result of transverse overlapping of the intensity profiles of signal and pump:

$$g(z) \approx \text{const} \cdot \langle |E(\mathbf{r}, z)|^2 \cdot |S(\mathbf{r}, z)|^2 \rangle / \langle |S(\mathbf{r}, z)|^2 \rangle \qquad [2]$$

Thus, the similarity of the intensity profile of the signal to that of the pump is rewarded by the exponential preference in the output signal level. However, this is not enough to guarantee that the output backward-scattered signal $S(\mathbf{r}, z)$ is phase

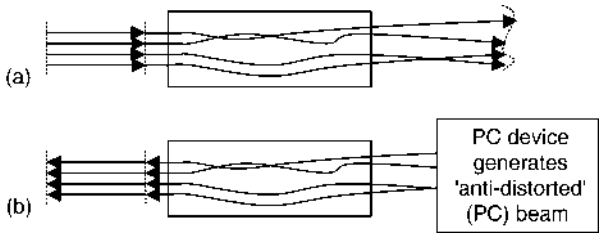

Figure 1 (a) Collimated beam is distorted by propagation through inhomogeneous medium. (b) Conjugate, a.k.a. antidistorted beam becomes collimated after backward passage through the same medium.

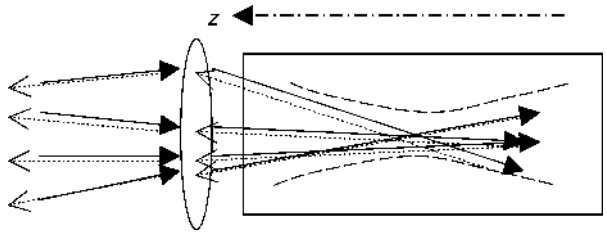

Figure 2 Stimulated Brillouin Scattering (SBS) method of PC. Solid lines symbolize the rays of incident 'beam', while dotted lines depict the rays of the signal amplified by the SBS process.

conjugate with respect to the pump. It is here that the mutual reversibility of propagation of strongly inhomogeneous signal and pump becomes crucial. Namely, good transverse overlapping is sustained along the entire interaction length z, only if $S(\mathbf{r}, z) \propto E^*(\mathbf{r}, z)$. This constitutes the 'discrimination mechanism' of SBS-PC: the 'conjugate mode' of the signal has a z-sustained advantage in exponential gain and, under appropriate conditions, suppresses all nonconjugate signal configurations in competition for pump power. Nonlinear-optical wave theory of this discrimination process has been developed further.

The holographic mechanism of PC has a different nature. It may take the form of static holography or dynamic holography; the latter essentially may be considered as a part of nonlinear optics. Here is a simplified description of one of the variants. Suppose one wants to obtain a conjugate replica of the incident monochromatic signal, whose complex profile is $S(\mathbf{R})$. A plane reference wave $A(\mathbf{R}) = \exp(i\mathbf{k}_A \cdot \mathbf{R})$ is also directed to the registering medium. If $S(\mathbf{R})$ and $A(\mathbf{R})$ are mutually coherent, then an interference pattern of intensity is produced $\{S^*(\mathbf{R})A(\mathbf{R}) + \text{compl. conj.}\}$. The medium records this pattern in the form of the modulation of refractive index and/or of the absorption coefficient, $\delta\varepsilon(\mathbf{R}) \propto (c_1 + ic_2)\{S^*(\mathbf{R})A\ (\mathbf{R}) + \text{c.c.}\}$ (Figure 3a). At the 'reconstruction' stage, the hologram is illuminated by another plane reference wave, $B(\mathbf{R}) = B_0 \exp(i\mathbf{k}_B \cdot \mathbf{R})$. In the specific case, when $\mathbf{k}_B = -\mathbf{k}_A$, the source δD_{conj} of the reconstructed wave becomes, as shown by Figure 3b:

$$\delta D_{\text{conj}} = (c_1 + ic_2)[S^*(\mathbf{R})A(\mathbf{R})] \cdot B(\mathbf{R}) \propto S^*(\mathbf{R}) \times (c_1 + ic_2)A_0 B_0 \qquad [3]$$

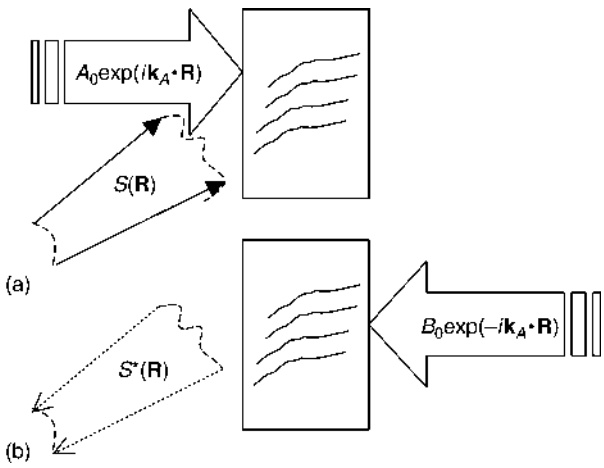

Figure 3 (a) Recording of a hologram in a photosensitive medium by a reference wave A and signal wave S. (b) Read-out of that hologram by the wave B counter-propagating to A results in generation of the conjugated signal S^*.

Reversibility of propagation laws guarantees that this source will indeed excite a conjugate wave with high efficiency. Both processes, recording of the interference pattern and reconstruction of the conjugate wave, may occur simultaneously, if one deals with the dynamic holography. The same process may also be described as nonlinear-optical four-wave mixing (FWM) via cubic nonlinearity $\chi^{(3)}$:

$$\delta D_{\text{conj}}(\mathbf{R}) = \chi^{(3)} S^*(\mathbf{R}) A(\mathbf{R}) B(\mathbf{R}) = \chi^{(3)} S^*(\mathbf{R}) A_0 B_0 \qquad [4]$$

with the four waves; A, B, S, S^*.

Important characteristics of a PC device are: fidelity of conjugation, efficiency of returning back-reflected power/energy, and reaction time or build-up time. The fidelity parameter must show how close the output of the device $E_{\text{out}}(\mathbf{r})$ is to the perfectly conjugate profile $S^*(\mathbf{R})$ of the incident signal (up to an arbitrary constant complex factor). Among other definitions, the square modulus of the normalized transverse overlapping integral of the fields is often discussed:

$$f = \frac{\left| \iint E_{\text{out}}(x, y) S(x, y) \mathrm{d}x\mathrm{d}y \right|^2}{\left(\iint |E_{\text{out}}(x, y)|^2 \mathrm{d}x\mathrm{d}y \cdot \iint |S(x, y)|^2 \mathrm{d}x\mathrm{d}y \right)} \qquad [5]$$

To achieve good (close to 1) fidelity of PC in the holographic or FWM scheme, one has to guarantee that the reference waves, $A_0 \exp(i\mathbf{k}_A \cdot \mathbf{R})$ and $B_0 \exp(i\mathbf{k}_B \cdot \mathbf{R})$, are exactly conjugate to each other, $\mathbf{k}_A = -\mathbf{k}_B$. That, in turn, requires the absence of any distortions in the holographic or nonlinear medium. To the contrary, there are very modest requirements on the phase inhomogeneities of the medium in the SBS scheme of PC, and for that reason it is usually labeled as a scheme of self-phase conjugation.

A hybrid scheme, Brillouin-assisted FWM, has been suggested and implemented, where mutually conjugate reference waves are produced via SBS-PC, while FWM, in a separate medium, exploits SBS nonlinearity. This scheme has produced extremely high, up to about 10^{10}, reflectivity, accompanied by an extremely low-noise detection of incoming signals.

A number of other PC schemes were first realized with the use of an important class of materials for nonlinear optics: photo-refractive crystals (PRC). These are crystals where ionization of the dopants by the incident light creates interference pattern of charge separation. The linear electro-optic (Pockels) effect transforms the resulting patterns of the electrostatic field into patterns of refractive index, thus creating dynamic holograms. A remarkable property of these crystals is that most of them act as very good electrical insulators. Therefore, the only source of conductivity, which tends to erase the

hologram, is photoconductivity, the value of the latter being proportional to the intensity of incident light itself. For this reason, the steady-state strength of the hologram turns out to be independent of the intensity. It is the build-up time that must be increased, if intensity is low. One can achieve various processes in PRC, including those similar to the $\chi^{(3)}$-type optical Kerr effect, stimulated scattering, etc.

The simplest use of PRC for generation of PC waves is to devise a FWM scheme (see Figure 3). However, a multitude of other, nontrivial schemes have been devised and implemented. The most impressive use of PRC for PC is based on interactions of the beams in the vicinity of a corner of a rectangular crystal, typically $BaTiO_3$, these interactions being of the stimulated scattering type. To honor J. Feinberg's cat, whose image was reconstructed in the first demonstration, this scheme is universally called 'cat conjugator'. It turned out to be a very reliable and robust scheme of self-phase-conjugation (SPC).

Another important group of the SPC scheme is called 'tail-biting'. One of the variants of the latter is shown in Figure 4. Incident 'pump' wave $A(\mathbf{R})$, whose transverse profile one wants to conjugate, enters the medium, and then is redirected by mirrors back into the same medium, to 'bite itself'. For the purpose of discussion, this redirected pump is labeled as wave $B(\mathbf{R})$ on its second arrival into the medium. Spontaneous scattering results in a seed for the amplification of the wave $C(\mathbf{R})$ via the process of stimulated scattering (SS) $A(\mathbf{R}) \rightarrow C(\mathbf{R})$.

Wave $C(\mathbf{R})$ is also redirected into the medium by the same mirrors, and is labeled as wave $D(\mathbf{R})$ on its second arrival into the medium. Both $C(\mathbf{R})$ and $D(\mathbf{R})$ are depicted in Figure 4 by white arrows. Wave $D(\mathbf{R})$ is also amplified due to the SS process $B(\mathbf{R}) \rightarrow D(\mathbf{R})$.

As it follows from the macroscopic description of SS process, volume gratings of refractive index are recorded in the medium, $\delta n(\mathbf{R}) \propto -i[A^*(\mathbf{R})C(\mathbf{R}) + B^*(\mathbf{R})D(\mathbf{R})]$. Among all the transverse profiles for the seed $C(\mathbf{R})$, the one that happens to be proportional to $B^*(\mathbf{R})$ will be reflected by the mirrors into the medium in the form $D(\mathbf{R}) \propto A^*(\mathbf{R})$. The latter case is again a consequence of the reversibility of propagation laws, where both gratings have the same profile and add coherently in the form $A^*(\mathbf{R})B^*(\mathbf{R})$. What is even more important, the grating $B^*(\mathbf{R})D(\mathbf{R}) \propto B^*(\mathbf{R})A^*(\mathbf{R})$ serves to close the feedback loop for seeding the appropriate $C(\mathbf{R})$. This, and other variants of tail-biting schemes, have also proved to be reliable and robust. Quite often a laser amplifier is inserted in the path $A \rightarrow B$ and $C \rightarrow D$, thus enhancing the feedback.

Another important scheme is traditionally called 'double PC'; albeit it may be better described as 'mutual PC'. Figure 5 should help in understanding that scheme. Consider the beam $A(\mathbf{R})$ incident to a medium active with respect to the stimulated-scattering-type process $A(\mathbf{R}) \rightarrow C(\mathbf{R})$. This process originates from a random seed and results in a fan of different waves $C(\mathbf{R})$. Such a process is characteristic of PRCs and is a consequence of recording dynamic gratings of refractive index $\delta n(\mathbf{R}) \propto -i[A^*(\mathbf{R})C(\mathbf{R})]$. Suppose another beam $B(\mathbf{R})$, typically incoherent with respect to $A(\mathbf{R})$, illuminates the medium from the other side, and also is engaged in the fanning process, this time $B(\mathbf{R}) \rightarrow D(\mathbf{R})$, with the grating of refractive index $\delta n(\mathbf{R}) \propto -i[B^*(\mathbf{R})D(\mathbf{R})]$ involved. Among all the transverse profiles for the seed grating $C(\mathbf{R})A^*(\mathbf{R})$, the one that happens to be proportional to $B^*(\mathbf{R})A^*(\mathbf{R})$ will be supported by similar profiles in the $B(\mathbf{R}) \rightarrow D(\mathbf{R})$ scattering. In this way, two fanning

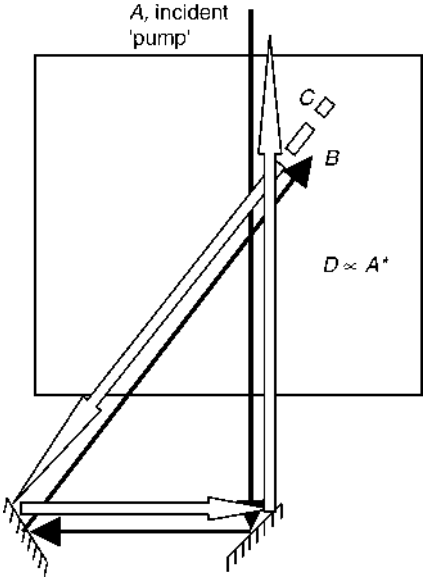

Figure 4 Tail-biting scheme of phase conjugation.

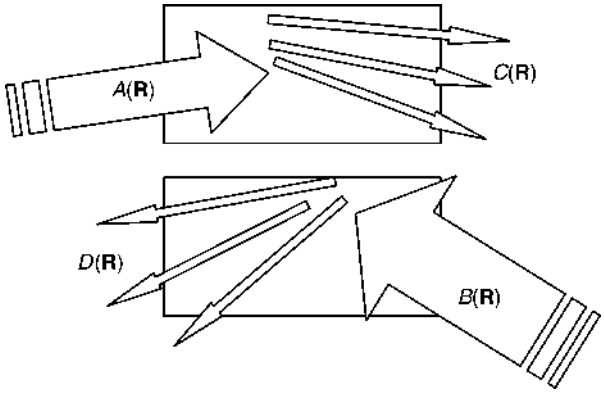

Figure 5 Towards the mechanism of double PC (mutual PC).

processes enhance each other and close the feedback loop. An important consequence is that one obtains the following output waves:

$$C(\mathbf{R}) = \text{const}_1 \times B^*(\mathbf{R})$$
$$D(\mathbf{R}) = \text{const}_2 \times A^*(\mathbf{R}) \qquad [6]$$

One can say that the output wave $C(\mathbf{R})$ presents a conjugate replica of $B(\mathbf{R})$-beam, while the output wave $D(\mathbf{R})$ presents a conjugate replica of $A(\mathbf{R})$-beam; hence the label 'mutual PC'. In some experiments the incident beams, $A(\mathbf{R})$ and $B(\mathbf{R})$, were of somewhat different colors (i.e., from different lasers), or their pulses did not overlap in time in the medium. It is the grating memory that 'informs' one beam about the presence of the other. The mechanism of cat conjugator is considered to be a combination of tail-biting and double PC schemes.

Pumped laser material with saturable gain often serves as an important nonlinear-optical medium for implementing PC. One of the evident advantages is that possible low efficiency of properly nonlinear process may be compensated by the gain in the same piece of material.

Considerable attention is paid nowadays to the chirp reversal of optical pulses that are used in fiber optical communications. Such chirped pulses, $S(t) = \exp(-i\omega_0 t + i\beta t^2/2 - t^2/\tau_1^2)$, with the value of the 'chirp' $\mathrm{d}\omega/\mathrm{d}t = -\beta$ and an increased pulse duration τ_1, arise as a result of propagation in a fiber which possesses group velocity dispersion (GVD). Pulse stretching and chirp are the consequences of different propagation times for the different frequency constituents of the pulse. If some device changes the sign of that chirp, then subsequent propagation of the pulse, through a piece of fiber with the same GVD, restores the duration of the pulse to original shorter value τ_0. A scheme has been suggested to use nonlinear optical FWM in a $\chi^{(3)}$-medium or three-wave mixing (ThWM) in a $\chi^{(2)}$-medium:

$$\delta D_{\text{conj}}(t) = \chi^{(3)} S^*(t) A(t) B(t)$$
$$\delta D_{\text{conj}}(t) = \chi^{(3)} S^*(t) C(t) \qquad [7]$$

Indeed, complex conjugation of the time dependence of the signal's field is equivalent to the change of the chirp sign. The second (ThWM) expression is written assuming that the reference wave $C(t) \propto \exp(-2i\omega_0 t)$, so that is has the frequency double that of the signal. The same process of ThWM has been shown to yield the conjugation of transverse phase profile of the beams, but for a number of reasons it was not applied.

As for the applications of PC, one of them was discussed above (**Figure 1**: double-pass scheme of a laser amplifier). Lasers with one or two PC mirrors, or with more complicated PC scheme, promise generation of high-transverse-quality beams with the use of realistic laser media, the latter inevitably possessing thermal and other distorting inhomogeneities. Another important potential application is free-space optical communications: Earth–Earth or Earth–satellite through atmosphere, or satellite–satellite communications. An 'interrogating' beam may be sent in one direction through an imperfect optical system and/or through turbulent atmosphere. Conjugation of transverse profile of the 'interrogating' beam at the other end of the communication link leads to an almost perfect redirection of the reflected beam towards the 'beacon' source, while time modulation may carry the information.

Readout of information data from a holographic storage is almost always performed in the regime of reading the conjugate wave, since it corrects for the most part of aberrations of optical systems in question.

To conclude, one may use a citation from the two consecutive *Scientific American* popular papers on PC: 'at present the number of applications would seem to be limited only by imagination.'

Reviews of various aspects of phase conjugation may be found in monographs and *Scientific American* papers listed below. The author of the present article was introduced to PC by his colleague V. V. Ragulskii, and has greatly benefited from collaboration with V. V. Ragulskii and V. V. Shkunov.

See also

Microscopy: Phase Contrast Microscopy. **Nonlinear Optics, Basics:** Four-Wave Mixing. **Phase Control:** Phase Conjugation and Image Correction. **Scattering:** Scattering Theory.

Further Reading

Bespalov VI and Pasmanik GA (1994) *Nonlinear Optics and Adaptive Laser Systems*. Commack, NY: Nova Science Publishers.

Feinberg J (1986) Self-pumped, continuous wave phase conjugator using internal reflection. *Optics Letters* 7: 486.

Fisher RA (ed.) (1983) *Optical Phase Conjugation*. New York: Academic Press.

Gower M and Proch D (eds) (1994) *Optical Phase Conjugation*. Berlin: Springer Verlag.

Hellwarth RW (1982) Optical beam conjugation by stimulated backscattering. *Optical Engineering* 21: 257.

Nosatch OYu, Popovichev VI, Ragulskii VV and Faizullov FS (1972) Compensation of phase distortions in an amplifying medium by a 'Brillouin mirror'. *JETP Letters* 16: 435.
Pepper DM (1986) Applications of optical phase conjugation. *Scientific American* 254: 74–83.
Shkunov VV and Zeldovich BYa (1985) Optical phase conjugation. *Scientific American* 253: 54–59.
Yariv A (1978) Phase conjugate optics and real-time holography. *IEEE Journal of Quantum Electronics* 14: 650.
Zeldovich BYa, Popovichev VI, Ragulskii VV and Faizullov FS (1972) On relation between wavefronts of reflected and exciting radiation in stimulated Brillouin scattering. *JETP Letters* 16: 109.
Zeldovich BYa, Shkunov VV and Pilipetsky NF (1985) *Principles of Phase Conjugation*. Berlin: Springer-Verlag.
Zeldovich BYa, Mamaev AV and Shkunov VV (1995) *Speckle-Wave Interactions in Application to Holography and Nonlinear Optics*. Boca Raton, FL: CRC Press.

Photorefraction

M Cronin-Golomb, Tufts University, Medford, MA, USA

B Kippelen, University of Arizona, Tucson, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction and Basics of the Photorefractive Effect

In 1966 Ashkin and coworkers were pursuing research on optical devices using the electro-optic material lithium niobate. They noticed that the refractive index of lithium niobate would change when it was exposed to laser light, and upset the expected operation of their devices. They called the effect optical damage. Shortly thereafter, Chen, LaMacchia, and Fraser reported on the use of the optical damage effect for holographic data storage. Thus began the field of photorefractive nonlinear optics, which has been used in various applications such as real-time holography, optical data storage, optical image amplification, nondestructive testing, distortion compensation by phase conjugation, pattern recognition, and radar signal processing. Many inorganic and organic materials have been investigated for their photorefractive effects, including ferroelectrics, semiconductors, and sensitized polymers. The most well-known inorganic materials are lithium niobate, bismuth silicon oxide, barium titanate, and strontium barium niobate. Most organic photorefractive materials are based on polymeric photoconductors such as those used in xerography that are doped with electro-active molecules, some plasticizers, and sensitizers.

While in its broadest interpretation, the photorefractive effect occurs whenever light incident on a material causes a refractive index change, one usually applies the term to electro-optic index changes resulting from optically generated space charge fields. Materials that are photorefractive in this sense share the following properties

- high transmission at the operating wavelengths;
- linear electro-optic coefficients or orientational Kerr effects;
- charge carriers that become mobile when optically excited;
- trapping centers for these charge carriers to enable spatially non-uniform redistribution of charge.

Consider two beams from the same laser crossing inside a photorefractive material such as barium titanate. The interference pattern will have bright and dark fringes. Charge carriers are excited where the light is bright, then drift and diffuse to regions of relative darkness where they preferentially recombine into trapping centers. In this way, a net excess charge develops in the dark regions, and a net deficit of charge develops in the bright regions. The spatially varying charge distribution has an electric field associated with it and this electric field causes a spatially varying refractive index profile. Because the space charge, its field, and resulting refractive index have the same spatial periodicity as the original interference pattern we have a holographic phase grating. The diffraction efficiency of the hologram can easily approach 100% for materials such as barium titanate and strontium barium niobate which have high electro-optic coefficients. Likewise, such high diffraction efficiencies are easily obtained in 100 micrometer-thick photorefractive polymer films.

The Standard Rate Equation Model

The development of photorefractive gratings can be modeled using standard semiconductor rate equations. Figure 1 shows two beams incident

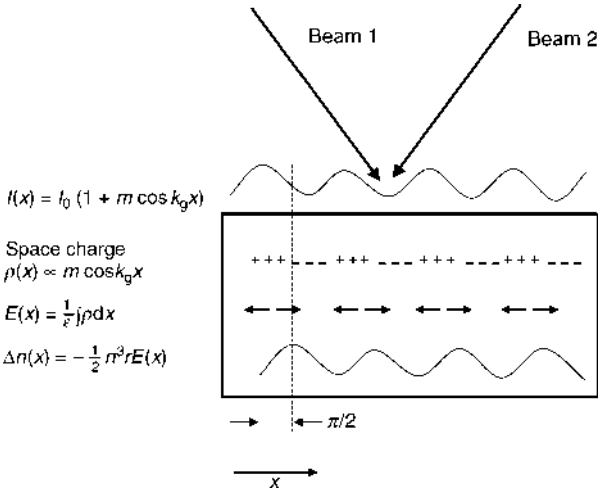

Figure 1 Diagram of two-beam coupling process showing optical interference pattern, space charge, resultant electric field, and $\pi/2$ phase shifted electro-optically induced refractive index grating.

symmetrically on a photorefractive crystal or polymer. They form an interference pattern whose intensity may be written:

$$I(x) = I_0(1 + m\cos(k_g x)) \qquad [1]$$

where x is the direction perpendicular to the interference fringes, I_0 is the average intensity, m is the modulation index, and k_g is the wavenumber of the interference pattern.

The crystal may be considered a wide bandgap semiconductor containing electron donors in the bandgap with density N_D and electron acceptors with density N_A. It is assumed that some electrons ionized from the donors permanently occupy these acceptors so that when charge in the crystal is distributed uniformly in the dark, the number density of ionized donors N_D^i is equal to N_A. Likewise, polymers contain donor and acceptor-like molecules that can be neutral or ionized. The spatially varying light distribution ionizes the donors at the following rate, assuming that $N_D^i \ll N_D$:

$$\frac{\partial N_D^i}{\partial t} = sIN_D - \gamma_R n_e N_D^i \qquad [2]$$

where s is a photoionization coefficient, γ_R is a recombination parameter, and n_e is the density of excited charge carriers, which we assume here to be electrons. The model can easily be generalized to include holes. We also use the equation of charge conservation:

$$\frac{\partial N_D^i}{\partial t} = \frac{\partial n_e}{\partial t} - \frac{1}{e}\nabla\cdot\mathbf{j} \qquad [3]$$

where e is the charge of an electron, and $\mathbf{j}$ is the current in the conduction band.

$$\mathbf{j} = \mu e n_e \mathbf{E} + k_B T \mu \nabla n_e \qquad [4]$$

where μ is the electron mobility, k_B is Boltzmann's constant, and T is the temperature. The electric field obeys Gauss's law:

$$\nabla\cdot\mathbf{E} = -e(n_e + N_A - N_D^i)/\varepsilon \qquad [5]$$

where N_A is the density of acceptors. These equations may be linearized by approximating the electron density, ionized donor density, and electric field with their first Fourier components:

$$\begin{aligned} E &= E_0 + \tfrac{1}{2}(E_1\exp(\mathrm{i}k_g x) + E_1^*\exp(-\mathrm{i}k_g x)) \\ N_D^i &= N_{D0}^i + \tfrac{1}{2}(N_{D1}^i\exp(\mathrm{i}k_g x) + N_{D1}^{i*}\exp(-\mathrm{i}k_g x)) \\ n_e &= n_{e0} + \tfrac{1}{2}(n_{e1}\exp(\mathrm{i}k_g x) + n_{e1}^*\exp(-\mathrm{i}k_g x)) \end{aligned} \qquad [6]$$

This assumption is strictly valid only when the modulation index m is much less than unity. Otherwise, a generalization to higher orders in the Fourier series is required. However, the linearized theory is sufficient to illustrate the most important features of the photorefractive effect. The solution for the space charge field E_1 for the case when the interference pattern is applied at time $t=0$ is

$$E_1 = m\frac{-\mathrm{i}E_Q(E_0 + \mathrm{i}E_D)}{E_0 + \mathrm{i}(E_D + E_Q)}(1 - \exp(-t/\tau)) \qquad [7]$$

where E_0 is an externally applied or photovoltaic field (if any), E_D is the diffusion field

$$E_D = \frac{k_B T k_g}{e} \qquad [8]$$

and E_Q is the limiting space charge field

$$E_Q = \frac{eN_A}{\varepsilon k_g} \qquad [9]$$

Some photorefractive crystals, most notably $LiNbO_3$, exhibit the photovoltaic effect in which optical illumination induces a dc field across the crystal.

The sinusoidally varying space charge field E_1 operates through the linear electro-optic effect with effective coefficient r to produce a sinusoidal variation in the refractive index n of the crystal:

$$n(x) = n_0 + \tfrac{1}{2}(n_1\exp(\mathrm{i}k_g x) + n_1^*\exp(-\mathrm{i}k_g x)) \qquad [10]$$

where

$$n_1 = -\frac{1}{2}rn_0^3E_1 \qquad [11]$$

The effective electro-optic coefficient r may be found from tensor analysis of the electro-optic tensor and

the vector space charge and optical fields. Notice that there is a 90-degree phase shift between the interference pattern and the refractive index grating when E_0 is zero. The time constant τ is given by

$$\tau = \frac{N_A}{sN_D I_0} \frac{E_0 + i(E_D + E_\mu)}{E_0 + i(E_D + E_Q)} \qquad [12]$$

where E_μ is a mobility field

$$E_\mu = \frac{\gamma_R N_A}{\mu k_g} \qquad [13]$$

When E_0 is nonzero, there is an oscillatory component to the time constant.

In contrast to the case of ordinary optical nonlinearities such as the optical Kerr effect, in which the nonlinear coupling strength is proportional to the optical intensity, the steady state strength of the photorefractive effect is independent of optical intensity while the time constant is inversely proportional to intensity in the basic model described above. The time constant varies with the photoconducting performance of a given material. In the fastest polymeric and inorganic materials it is, at the time of writing, of the order of a few milliseconds at $1\ \mathrm{W\,cm^{-2}}$ of incident optical intensity.

At low intensities (below the equivalent dark intensity, $1\ \mathrm{W\,cm^{-2}}$ in barium titanate), the above equations need to be modified to include the effect of dark conductivity. Even in the dark, there will be a few mobile charge carriers in the conduction band that tend to erase the grating. This will result in the effect appearing more Kerr-like, except still with the 90-degree phase shift between the index grating and the interference pattern.

Coupled Wave Equations

The change in refractive index n_1 can be large enough to produce substantial interactions between the writing beams. The writing beams generate a phase grating that diffracts the beams into each other. The grating influences the writing beams, which in turn influence the grating. In the cases where the diffusion field dominates, for example when the externally applied or photovoltaic field E_0 is absent, one beam is amplified by in-phase diffraction of the other beam from the grating. As shown in Figure 2, this amplification results from a 90-degree phase shift due to diffraction from a phase grating coupled with a -90-degree phase shift from the spatial phase shift between the interference pattern and the index grating. The second beam is attenuated by destructive interference with the first beam diffracted by the grating. These interactions can be modeled by standard coupled wave theory. Let the electric field amplitude associated with the jth beam be

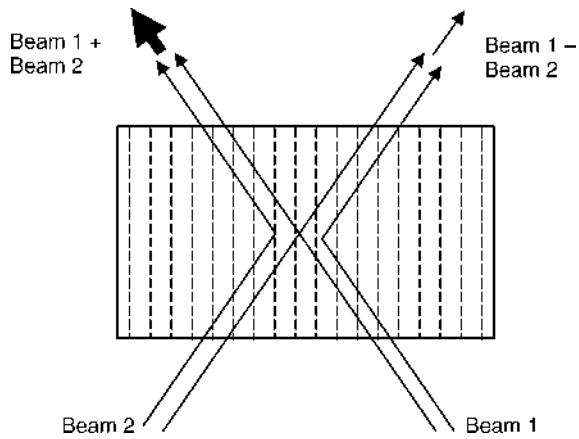

Figure 2 Two-beam coupling amplification. Beam 1 is amplified by constructive interference. Beam 2 is de-amplified by destructive interference.

$$\mathbf{E}_j(\mathbf{r}, t) = \mathbf{e}\Big[A_j(\mathbf{r}) \exp(i(\mathbf{k}_j \cdot \mathbf{r} - \omega t) + A_j^* \exp(-i(\mathbf{k}_j \cdot \mathbf{r} - \omega t))\Big] \qquad [14]$$

where $\mathbf{e}$ is the polarization unit vector. Using the scalar wave equation

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = 0 \qquad [15]$$

and the slowly varying envelope approximation

$$\left|\frac{d^2 A_j}{dz^2}\right| \ll k \left|\frac{dA_j}{dz}\right| \qquad [16]$$

we find

$$\frac{dA_1}{dz} = +\gamma \frac{A_1 |A_2|^2}{I_0}, \quad \frac{dA_2^*}{dz} = -\gamma \frac{A_2^* |A_1|^2}{I_0} \qquad [17]$$

with the coupling constant γ given by

$$\gamma = -\frac{i\omega \Delta n_m}{c \cos \vartheta} \qquad [18]$$

where ϑ is the half-angle between the writing beams and Δn_m is given by

$$\Delta n_m = -\frac{1}{2} n_0^3 r \frac{E_1}{m} \qquad [19]$$

Equation [17] shows that beam 1 is amplified and beam 2 is attenuated. That beam 1 is amplified

instead of beam 2 is a result of the choice of crystal orientation and hence the sign of γ. These nonlinear equations can be solved for normalized intensity $I_j = |A_j|^2$:

$$I_1(z) = \frac{I_0}{1 + (I_2(0)/I_1(0))\exp(+\Gamma z)}$$
$$I_2(z) = \frac{I_0}{1 + (I_1(0)/I_2(0))\exp(-\Gamma z)} \qquad [20]$$

where $\Gamma = \gamma + \gamma^*$ is the intensity coupling constant. In the limit where $I_1 \ll I_2$, the weak beam, beam 1, experiences exponential amplification. This amplification can be used to build optical oscillators with many of the same properties as conventional laser oscillators. The solution can be generalized to the case where γ has an imaginary component due to an applied or photovoltaic field E_0. Linear absorption can also be included.

Materials

Photorefractive materials may be separated into two broad classes: inorganic and organic. The first materials investigated were inorganic oxides and semiconductors. Their success led to efforts to endow more easily produced organic materials with the required photoconductivity, charge trapping centers, and electro-optic coefficients.

Inorganic Materials

The first requirement in a classic photorefractive material is a linear electro-optic coefficient, such as appears in sillenites such as bismuth silicon oxide $Bi_{12}SiO_{20}$ and bismuth germanium oxide $Bi_{12}GeO_{20}$. These were some of the first photorefractive materials used for image processing and phase conjugation applications. However, their electro-optic coefficients (a few pm/V) are not large enough to easily give rise to large diffraction efficiencies, or to photorefractive oscillators. Materials at a temperature near a phase transition generally have higher electro-optic coefficients because their crystal structures are on the verge of changing. They are extremely susceptible to the effect of any external influence such as the application of electric fields including photorefractive space charge fields. That is why ferroelectric materials are good candidates for photorefractive materials. These include barium titanate $BaTiO_3$, potassium niobate $KNbO_3$, and strontium barium niobate $Sr_xBa_{1-x}Nb_2O_6$. The mean free path of charge transport is less than that in sillenites, so they require more photons to reach a steady state. Therefore, ferroelectric materials are typically less sensitive than sillenites.

The second requirement is for photoconductivity. This requires the existence of photoexcitable charge carriers, either from valence band to conduction band or from intraband trapping centers. The latter source of photocarriers is used most often because the absorption length of light whose energy is greater than the bandgap is usually only a few micrometers. This places a severe restriction on the beams' interaction length ℓ. Thus there has been considerable research on suitable dopants, most commonly Fe^{2+} and Fe^{3+}. These dopants also act as trapping centers.

Lithium niobate is an example of a material with a large photovoltaic effect. When illuminated, this crystal develops a large dc field, which acts to enhance the grating strength. In those cases where the photorefractive effect is not wanted, such as in the design of lithium niobate waveguide devices, the photovoltaic effect can be greatly diminished by the addition of MgO to the crystal melt during growth.

For photorefractivity in the near infrared, semiconductors such as gallium arsenide and indium phosphide have been used with success. These materials also have the advantage that they can be grown in layered structures to produce, for example, multiple quantum wells that can be used to tailor the characteristics of optical excitation and charge transport.

Organic Materials

First-generation photorefractive polymers were designed to mimic the properties of their inorganic counterparts. Owing to their rich structural flexibility, organic synthetic materials with suitable charge transport, trapping, linear electro-optic effects and low optical absorption, were developed. This approach did build on the know-how developed previously in making photoconducting polymers for xerography and electro-optic polymers for optical modulation. Current polymers are based on an orientational photorefractive effect that leads to higher refractive index changes compared with traditional photorefractive materials. In materials with orientational photorefractivity, the refractive index change is produced by the field induced reorientation of anisotropic conjugated molecules that possess a permanent dipole moment and that have a high polarizability anisotropy. The photorefractive space charge field together with the applied field will periodically reorient these molecules and produce a periodic refractive index modulation through an orientational Kerr effect. The build-up and dynamics of this space-charge field are similar to

those of traditional photorefractive crystals and can be described by eqns [2]–[7]. The time constant of the hologram is a convolution of the build-up time of the space-charge and the orientational diffusion time of the dipolar molecules in the total electric field. In contrast to crystals, polymers are nearly amorphous and the transport properties are described by charge hopping processes instead of band-type transport in crystals. Consequently, the photoconducting properties of polymers are strongly field dependent and photorefractivity is mainly observed under strong applied voltages. Numerous polymer composites have been developed using the hole transport polymer poly(N-vinylcarbazole). Several materials with a refractive index modulation amplitude of the order of a percent and two-beam coupling constants $\gamma > 100\,\mathrm{cm}^{-1}$ have been reported. New polymer composites are continuously being tested. Photorefractivity is also studied in other organic materials including organic crystals, liquid crystals, nanocomposites such as polymer-dispersed liquid crystals, or hybrid materials such as sol-gels.

Applications

Holographic Data Storage

Holographic data storage takes advantage of the Bragg selectivity of thick gratings. This allows many holograms to be superimposed in the same small volume, typically of the order of one cubic centimeter. A page of data is displayed on a spatial light modulator and a laser beam passing through the modulator is holographically recorded in the crystal with a reference beam at a specific angle. Many pages of data can be superimposed by recording many holograms with angularly multiplexed reference beams. Other multiplexing schemes are implemented by changing the shape of the wavefront of the reference beam. In principle, the upper limit of recording density is determined by the wavelength of light: one bit per cubic wavelength. If the recording wavelength is 0.5 μm, then one cubic centimeter can contain 1000 gigabytes of data. In practical circumstances, when noise is taken into consideration, the capacity is more realistically of the order of one gigabyte if a 1000 × 1000 spatial light modulator is used.

Distortion Compensation by Phase Conjugation

Photorefractive materials can be used to make high-reflectivity phase conjugate mirrors. The phase conjugate of a laser beam is produced when a hologram of the beam is read by another beam traveling in the opposite direction to the original reference beam. The phase conjugate reconstruction is a time-reversed copy of the original beam. If the original beam has passed through distorting optics, then the phase conjugate beam will retrace the path of the original beam through the distortion and emerge in its undistorted original state. In the standard nomenclature of phase conjugation, the input beam is called the signal, or probe, and the two reference beams are called the pumps. The output beam at $z = 0$ is called the phase conjugate beam and has zero amplitude at its input at $z = \ell$, where ℓ is the interaction length. Applications for phase conjugation exist, for example, in signal transmission through distortions and laser cavity design. If a phase conjugate mirror is used as a cavity mirror, then the effects of intracavity distortions are removed.

Since the photorefractive gratings can be very strong, the diffraction efficiency of the counterpropagating reference beam can be so high that the phase conjugate reflectivity can exceed unity. The simplest generalization of eqn [17] to the four-wave mixing phase conjugation case is when only transmission gratings are important, as occurs in many circumstances depending on the mutual coherence properties of the beams, and the material's properties. The coupled wave equations are

$$\begin{aligned}
\frac{\mathrm{d}A_1}{\mathrm{d}z} &= \gamma\frac{(A_1A_4^* + A_2^*A_3)A_4}{I_0}\\
\frac{\mathrm{d}A_2^*}{\mathrm{d}z} &= \gamma\frac{(A_1A_4^* + A_2^*A_3)A_3^*}{I_0}\\
\frac{\mathrm{d}A_3}{\mathrm{d}z} &= -\gamma\frac{(A_1A_4^* + A_2^*A_3)A_2}{I_0}\\
\frac{\mathrm{d}A_4^*}{\mathrm{d}z} &= -\gamma\frac{(A_1A_4^* + A_2^*A_3)A_1^*}{I_0}
\end{aligned} \qquad [21]$$

In the undepleted pumps approximation, $(I_1, I_2 \gg I_3, I_4)$, the equations become linear and the solution for phase conjugate reflectivity $R = I_3(0)/I_4(0)$ is

$$R = \left|\frac{\sinh[\gamma\ell/2]}{\cosh[(\gamma\ell + \ln r)/2]}\right|^2 \qquad [22]$$

where $r = I_2/I_1$ is the ratio between the intensities of the pumping beams.

The fact that the reflectivity of the phase conjugate mirror can be greater than unity means that we can build an optical oscillator bounded by a regular mirror and a phase conjugate mirror only. Not only does it not require any additional optical gain, but it also compensates for intracavity distortions. The regular mirror can have any shape, provided that it is sufficiently reflective.

Self-pumped Phase Conjugate Mirrors

If a laser beam passes through a crystal placed inside an optical cavity formed by two facing mirrors, light scattered by imperfections in the crystal can be amplified through the photorefractive effect. The cavity provides feedback and optical oscillation can build up. The oscillation beams pump the crystal as a phase conjugate mirror for the incident beam, thus forming a self-pumped phase conjugate mirror. The feedback can even be provided by total internal reflection in the crystal in which case the crystal by itself can become a phase conjugate mirror.

Pattern Recognition and Image Filtering

Photorefractive wave mixing can be used to perform pattern recognition by matched filtering. One example would be to identify a tank in a battlefield scene, another would be to identify all of the occurrences of a particular word on a page of text. Suppose the input beams contain the corresponding pictorial information, such as might be obtained by passing the beams through an image-bearing transparency or spatial light modulator. Equation [21] shows that the source for beam 3 contains a term proportional to the product of the amplitudes of the three input beams. If lenses are placed in the input beams so that the crystal receives the Fourier transforms of those beams, then the output beam at the crystal, beam 3, will be proportional to the product of the Fourier transforms of the input beams 1, 2, and 4. A lens may then be used to perform the inverse Fourier transform of the product of the Fourier transforms of the input beams, producing an output proportional to beam 1 convolved with beam 2 correlated with beam 4. If beam 1 is a point source before its Fourier transforming lens, then it will be a plane wave at the crystal. Beam 3 after inverse Fourier transformation by its lens will be the correlation of beams 4 and 2. For example, suppose we want to find all the occurrences of a particular word, say 'optics' in a given page of text. Then we would make a transparency of the word 'optics' and place that in input beam 4. We would then place an image of the page of text in beam 2. Beam 3 would then contain a field with bright spots at the places containing the word 'optics' in the original text.

The real-time holographic recording properties of photorefractive materials can also be exploited in medical imaging applications by performing time-gated holography. In this method, a hologram is formed by the temporal overlap in the photorefractive sample of a reference pulse and the first-arriving (ballistic photons) light from a stretched image-bearing pulse that has propagated through a scattering medium. The filtering of the useful photons from the scattered ones is achieved by reconstructing the hologram formed with the ballistic photons in a four-wave mixing geometry.

Optical Limiting, the Novelty Filter, and Laser Ultrasonic Inspection

The attenuation of beam 2 in eqn [17] can be used in several applications including optical limiting and novelty filtering. If one wants to protect a sensor from high-intensity laser radiation, then one could split a small portion of the input beam directed at the sensor and use it as beam 1 in a photorefractive recording setup with the input beam acting as beam 2. If the laser intensity is above the equivalent dark intensity such that the optically excited charge density is greater than the thermally excited charge density, the photorefractive effect will be activated and the input beam will be de-amplified by destructive interference with beam 1, thus protecting the sensor. In materials with high gain–length products ($\gamma\ell > 1$), separate provision of beam 1 is unnecessary because light scattered from the input beam by imperfections in the material and other scattering centers will be greatly amplified, often to such an extent that the input beam is almost completely attenuated. This photorefractive amplification of scattered radiation is known as the fanning effect, because the amplified scattered light often appears as a fan, or brush of light, as can be seen in Figure 3. The effect can also be used to make a novelty filter, which transmits only the moving portion of a scene. The crystal is only fast enough to

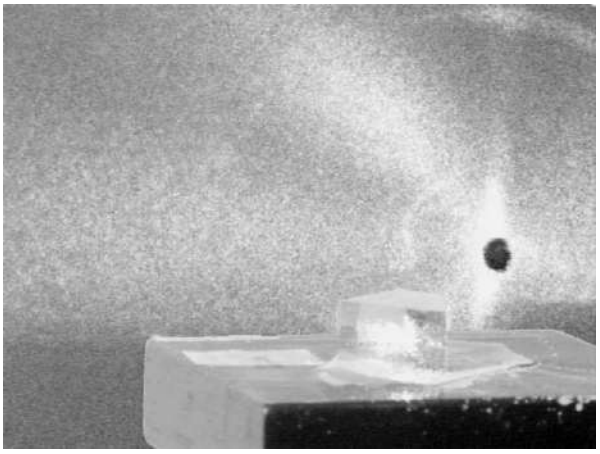

Figure 3 Photorefractive barium titanate exhibiting amplified scattering. A helium neon laser beam is incident from the lower left, passes through the crystal to a screen where it is blacked out to prevent saturation of the camera. The screen shows brushes of amplified scattering, or fanning.

respond to the slowly changing or static components of an image-bearing beam. Since the grating formed is Bragg-matched only to the slow component, the grating will only attenuate that slow component. Any rapidly changing parts of a scene pass through the crystal unattenuated. Such a filter is useful for picking out moving objects in a static cluttered background, for example a tank on a battlefield, or a micro-organism swimming against a stationary background.

A related application is in laser ultrasonic inspection. Defects in industrial material processing such as welding have characteristic ultrasonic signatures. The part under test is pinged by a pulsed laser and a probe laser is reflected from the part. As the part is shaken by the ultrasound, speckle in the reflected beam vibrates. A photorefractive recording is made of the speckle beam. Electrodes are placed on the photorefractive material, so that any optically induced currents can be detected. If the speckle pattern is not moving, or is moving more slowly than the speed of response of the photorefractive material, the photorefractive grating will be essentially at steady state: the drift currents balance the diffusion currents so there is no net current. There is no signal as the part moves through the process line. However, if the speckle beam is moving faster than the response time of the material, it will move photoexcited charge back and forth past the quasistatic grating and generate a net current for detection via the electrodes.

Adaptive Signal Processing

The relatively slow speed of photorefractive devices can be used to advantage in radio-frequency (RF) signal processing, such as signal extraction and coherent combination of signals from antenna arrays. The signal extraction application depends on grating competition in two-beam coupling. Suppose we wish to separate signals on two different RF carrier frequencies ω_1 and ω_2, respectively. The combined signal is applied to an optical carrier beam using a high-speed modulator. The resulting optical field may be represented as

$$S_1(t)\exp(\mathrm{i}\omega_1 t) + S_2(t)\exp(\mathrm{i}\omega_2 t) \qquad [23]$$

It is used to pump a unidirectional ring resonator so that grating competition allows oscillation only on the strongest component of the combined signal, say S_1. The output of the oscillator is proportional to the extracted component S_1. The effectiveness of the intersignal competition is enhanced by placing another photorefractive material in the cavity. A portion of the intracavity beam is picked off by a beamsplitter and used as a two-beam coupling pump in the second material. The crystal is oriented so that the photorefractive grating diffracts the picked-off beam back into the cavity. The return of the picked-off component is most effective for the stronger component S_1 thus decreasing its loss compared to that of the weaker component S_2. This beamsplitter/crystal combination is known as a reflexive coupler.

Photorefractive Solitons

Under favorable conditions, a single beam incident on a photorefractive crystal will induce a positive refractive index change at the center of the beam. This provides a self-focussing tendency that counteracts the beam's divergence due to diffraction. When these two effects balance each other, the beam can propagate with a constant diameter. Such a beam is known as a spatial soliton in analogy with temporal solitons in optical fibers and can occur when there is a component of the photorefractive response due to drift. The drift component of the photorefractive effect appears when a dc field E_0 is applied to the material. Potential applications are optically written waveguides and couplers.

List of Units and Nomenclature

Amplitude gain	[cm^{-1}]	γ
Applied dc field	[Vm^{-1}]	E_0
Current	[Am^{-2}]	$\mathbf{j}$
Dielectric permittivity	[Farads m^{-1}]	ε
Diffusion field	[Vm^{-1}]	E_D
Electron acceptor density	[m^{-3}]	N_A
Electron charge	[C]	e
Electron donor density	[m^{-3}]	N_D
Grating wavenumber	[m^{-1}]	k_g
Intensity gain	[cm^{-1}]	Γ
Interaction length	[cm]	ℓ
Ionized donor density	[m^{-3}]	N_D^i
Limiting space charge field	[Vm^{-1}]	E_Q
Mobility	[m^2 V^{-1} sec^{-1}]	μ
Mobility field	[Vm^{-1}]	E_μ
Normalized intensity	[V^2 m^{-2}]	I
Photo-ionization coefficient	[kg sec^{-2}]	s
Photorefractive time constant	[sec]	τ
Pump ratio	[1]	r
Radian frequency	[sec^{-1}]	ω
Recombination coefficient	[m^3 sec^{-1}]	γ_R
Slowly varying optical electric field	[Vm^{-1}]	A

See also

Holography, Applications: Holographic Recording Materials and Their Processing. **Holography, Techniques:** Holographic Interferometry. **Imaging:** Volume Holographic Imaging. **Information Processing:** Free-Space Optical Computing; Optical Neural Networks. **Materials for Nonlinear Optics:** Organic Nonlinear Materials. **Modulators:** Electro-Optics. **Nonlinear Optics, Basics:** Four-Wave Mixing; Nonlinear Optical Phase Conjugation. **Phase Control:** Phase Conjugation and Image Correction. **Scattering:** Stimulated Scattering. **Solitons:** Bright Spatial Solitons.

Further Reading

Coufal HJ, Psaltis D and Sincerbox GT (eds) (2000) *Holographic Data Storage*. Berlin: Springer.

Gunter P and Huignard J-P (eds) (1988) *Photorefractive Materials and their Applications I*. Berlin: Springer-Verlag.

Gunter P and Huignard J-P (eds) (1989) *Photorefractive Materials and their Applications II*. Berlin: Springer-Verlag.

Nalwa HS and Miyata S (eds) (1997) *Nonlinear Optics of Organic Molecules and Polymers*. Boca Raton: CRC Press.

Nolte DD (ed.) (1995) *Photorefractive Effects and Materials*. Kluwer.

Pepper DM, Feinberg J and Kukhtarev NV (1990) The photorefractive effect. *Scientific American* 263: 62.

Solymar L, Webb DJ and Grunnet-Jepsen A (1996) The Physics and Application of Photorefractive Materials. Oxford: Clarendon Press.

Yeh P (1993) Introduction To Photorefractive Nonlinear Optics. New York: Wiley.

Yu F and Yin S (eds) (2000) *Photorefractive Optics: Materials, Properties, and Applications*. San Diego: Academic Press.

Ultrafast and Intense-Field Nonlinear Optics

A L Gaeta, Cornell University, Ithaca, NY, USA

R W Boyd, University of Rochester, Rochester, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The tremendous development of high-powered femtosecond laser systems since the 1980s, has opened up new areas of research in nonlinear optics, plasma physics, atomic and molecular dynamics, and intense-field physics. For many of these applications, it is important to understand how ultrashort light pulses propagate through a medium under conditions in one or more of the processes in which nonlinear optics play an important role.

The starting point for the modeling of light propagation under these conditions is Maxwell's wave equation for the electric field $E(\mathbf{r}, t)$ which is given in Gaussian units as

$$\nabla^2 E - \frac{\partial^2 E}{\partial t^2} = 4\pi \frac{\partial^2 P}{\partial t^2} \qquad [1]$$

where the $P(\mathbf{r}, t)$ is the polarization inside the medium. Typically, the polarization is separated into a term P_l that depends linearly on the field E and into a term P_{nl} that depends nonlinearly on the applied field. The Fourier transform of the linear polarization can be expressed as $\tilde{P}_l(\mathbf{r}, \omega) = \chi^{(1)}(\omega)\tilde{E}(\mathbf{r}, \omega)$, where $\tilde{E}(\mathbf{r}, \omega)$ is the Fourier transform of electric field $E(\mathbf{r}, t)$. For the laser–matter interactions that we consider here, we assume that the linear susceptibility $\chi^{(1)}(\omega)$ is real, in which case the wave equation can be expressed as

$$\nabla^2 \tilde{E} + \left[\frac{n_l(\omega)\omega}{c} \right]^2 \tilde{E} = -\frac{4\pi\omega^2}{c^2} \tilde{P}_{\mathrm{nl}} \qquad [2]$$

where $n_l(\omega) = \sqrt{1 + 4\pi\chi^{(1)}(\omega)}$ is the frequency-dependent linear refractive index of the medium.

For light pulses that are longer than an optical period, the envelope description is valid and the electric field can be described by an amplitude envelope $A(\mathbf{r}, t)$ and a carrier frequency ω_0 such that

$$E(\mathbf{r}, t) = A(\mathbf{r}, t)\mathrm{e}^{\mathrm{i}(k_0 z - \omega_0 t)} + c.c \qquad [3]$$

where $k_0 = k(\omega_0) = n_0\omega_0/c$ is the wavevector amplitude and $n_0 = n_l(\omega_0)$. This envelope description is advantageous for performing analytical and numerical studies of pulse propagation. However, for sufficiently short laser pulses, where the shape of the envelope function does not depend on the carrier phase of the carrier wave, such a description is no longer applicable. Nevertheless the envelope description can be made valid for pulses that are nearly as short as a single cycle or, alternatively, that have spectral bandwidths that are comparable to the central frequency ω_0. To derive an equation for the spatio-temporal evolution of the pulse envelope, the relation for the electric field is substituted into the

wave equation (eqn [2]) and the following two approximations are made: i) $k_0 \partial A/\partial z \ll \partial^2 A/\partial z^2$, which signifies that the envelope varies slowly in space over a distance compared to the central wavelength; ii) $v_{\rm ph} \sim v_{\rm gr}$ where $v_{\rm ph} = c/n_0$ is the phase velocity and $v_{\rm gr} = ({\rm d}k/{\rm d}\omega)^{-1}$ is the group velocity. This latter approximation is invariably well satisfied when the frequencies contained in the laser field are highly nonresonant with any transition frequencies of the medium.

For an input pulse at $z = 0$ with a peak amplitude A_0, and characteristic widths in space and time given by w_0 and τ_p, respectively, the equation for the normalized amplitude $u(r_\perp, z, t) = A(r_\perp, z, t)/A_0$ can be expressed as

$$\frac{\partial u}{\partial \xi} = \frac{i}{4}\left(1 + \frac{i}{\omega_0 \tau_p}\frac{\partial}{\partial \tau}\right)^{-1} \nabla_\perp^2 u - i \sum_{n=2} \frac{L_{\rm ds}}{n! L_{\rm ds}^{(n)}} \frac{\partial^n u}{\partial \tau^n} + i\left(1 + \frac{i}{\omega_0 \tau_p}\frac{\partial}{\partial \tau}\right) p^{\rm nl} \qquad [4]$$

where $\tau = (t - z/v_g)/\tau_p$ is the normalized retarded time for the pulse traveling at the group velocity v_g, $L_{\rm ds}^{(n)} = \tau_p^n/\beta_n$ is the nth-order dispersion length, $\beta_n (n \geq 2)$ is the nth-order dispersion constant, $L_{\rm ds} = |L_{\rm ds}^{(2)}|$ is the dispersion length, $\zeta = z/L_{\rm df}$ is the normalized distance, $L_{\rm df} = k w_0^2/2$ is the diffraction length, $p_{\rm nl}$ is the normalized nonlinear polarization, and $\nabla_\perp^2$ is the transverse Laplacian. The presence of the operator $(1 + i\partial/\omega_0 \partial \tau)$ in the diffraction term $(\nabla_\perp^2 u)$ of eqn [4] leads to space–time focusing, while its presence in the nonlinear term results in self-steepening and both these terms arise from not making the slowly varying envelope approximation in time (i.e., $k_0 \partial A/\partial t \ll \partial^2 A/\partial t^2$) in deriving eqn [4]. For a self-consistent analysis of pulse propagation in the nonlinear regime, both the effects of space–time focusing and self-steepening must be included.

Nonlinear Refractive Index

For an isotropic medium in the highly nonresonant limit, the third-order term is the lowest-order contribution to the nonlinear susceptibility. This term gives rise to the nonlinear refractive index, that is, the index of refraction depends on the local intensity of the laser field. For many materials there are two contributions to the nonlinear refractive index: i) an instantaneous part that arises from the electronic response of the medium; and ii) a noninstantaneous contribution due to the nuclear motion of the molecules (i.e, the Raman contribution). For such a medium, the nonlinear polarization may be expressed as

$$p_{\rm nl}(\zeta, \tau) = \frac{L_{\rm ds}}{L_{\rm nl}}\left[(1 - f)|u(\zeta, \tau)|^2 + f\gamma_R \int_{-\infty}^{\tau} {\rm d}\tau' R(\tau - \tau')|u(\zeta, \tau')|^2\right] u(\zeta, \tau) \qquad [5]$$

where $L_{\rm nl} = (c/\omega_0 n_2 I_0)$ is the nonlinear length, $I_0 = n_0 c|A_0|^2/2\pi$ is the peak input intensity, f is the fraction of the Raman contribution to the nonlinear refractive index, and $R(\tau) = \left\{\left[1 + (\Omega_R \tau_R)^2\right]/\Omega_R \tau_R\right\} \exp(-\tau/\tau_R) \sin(\Omega_R \tau)$ is the Raman response function, τ_R is the characteristic Raman response time, Ω_R is the characterstic Raman frequency, and $\gamma_R = \tau_p/\tau_R$. For example, for fused silica $f = 0.15$, $\tau_R = 50$ fs, and $\Omega_R \tau_R = 4.2$. For a noble gas such as Xe, there is no Raman contribution and $f = 0$.

Self Focusing, Supercontinuum Generation, and Filamentation

The presence of the nonlinear refractive index with $n_2 > 0$ can lead to self-focusing of a laser beam. For sufficiently long pulses such that dispersion, self-steepening, and space–time focusing effects can be neglected, it is found that laser beams with input powers greater than the critical power $P_{\rm cr} = \alpha\lambda^2/4\pi n_0 n_2$ will undergo catastrophic self-focusing collapse. The dimensionless parameter $\alpha \geq 1.86$ depends on the spatial profile of the input beam and for a Gaussian input beam is given by $\alpha = 1.9$, in which case the ratio of the input power P to the critical power satisfies the relation $P/P_{\rm cr} = 1.055 L_{\rm df}/L_{\rm nl}$. Extensive studies have been made on the dynamics of laser beams undergoing self-focusing. For example, it has been shown that the shape of the collapsing beam evolves to a radially symmetric profile as it approaches the collapse point and that the total power contained in the collapsing portion always corresponds to the minimum value (i.e., $\alpha \sim 1.86$) regardless of the initial power in the beam.

For light pulses shorter than a picosecond, the role of material dispersion plays an important role and completely alters the dynamics of the self-focusing process. These dispersive effects lead to a temporal splitting of the pulse into two pulses and the arrest of its collapse. At even higher powers, other phenomena can occur, such as 'optical shock' formation at the rear edge of the pulse, due to self-steepening and space–time focusing. Shock formation leads to the emission of a broad spectrum of radiation extending from the ultraviolet to the mid-infrared, known as supercontinuum generation (SCG). This phenomenon was first observed in 1970, and since then it has been

observed in many different solids, liquids, and gases, under a wide variety of experimental conditions.

If the peak intensity of the pulse approaches intensities of $> 10^{13}$ W/cm^2, either through self-focusing or external focusing, multiphoton ionization occurs and a plasma is formed in the medium. (See section below on Plasma Nonlinearities and Relativistic Effects.) The generation of the plasma lowers the refractive index and effectively counteracts the self-focusing process, resulting in the spatial confinement of the pulse for distances far beyond what would be allowed by ordinary linear diffraction and has been observed in gases, liquids, and solids. A striking example of this apparent self-waveguiding is the observation of 'light strings' in air which can extend more than 10 km into the atmosphere. This phenomenon was first observed with 100 fs laser pulses in the near infrared ($\lambda = 800$ nm). Researchers found that pulses with energies greater than 10 mJ undergo self-focusing collapse in air and produce a highly intense ($> 10^{13}$ W/cm^2) 100-micron-diameter light filament tens of meters long.

Multiphoton Absorption

Multiphoton absorption is a process in which an atom or molecule makes a transition from a ground state to an excited state by means of the simultaneous absorption of N photons. In the lowest order of perturbation theory, such a process can be described by means of a susceptibility of order $(2N-1)$, that is, by $\chi^{(2N-1)}$. Alternatively, this process can be described in terms of an N-photon cross-section $\sigma^{(N)}$ defined such that the transition rate per atom is given by

$$R^{(N)} = \sigma^{(N)} I^N \qquad [6]$$

where I is the intensity of the laser field. Quantum mechanical expressions for the N-photon cross-sections are readily obtained. One finds, for instance, that

$$R^{(2)}_{\mathrm{ng}} = \left| \sum_m \frac{\mu_{\mathrm{nm}}\mu_{\mathrm{mg}}E^2}{\hbar^2(\omega_{\mathrm{mg}} - \omega)} \right|^2 2\pi\rho_{\mathrm{f}}(\omega_{\mathrm{ng}} - 2\omega)$$

$$R^{(3)}_{\mathrm{og}} = \left| \sum_{mn} \frac{\mu_{\mathrm{on}}\mu_{\mathrm{nm}}\mu_{\mathrm{mg}}E^3}{\hbar^3(\omega_{\mathrm{ng}} - 2\omega)(\omega_{\mathrm{mg}} - \omega)} \right|^2 2\pi\rho_{\mathrm{f}}(\omega_{\mathrm{og}} - 3\omega) \qquad [7]$$

In each of these expressions, the quantity ρ_{f} represents the density of final states, or equivalently the atomic lineshape function, evaluated at the N-photon transition frequency.

Optical Damage

High-intensity laser fields can produce unwanted damage to optical materials. As a point of reference, the threshold for laser damage to fused silica at a wavelength of 1.05 micrometers for a pulse of 30 ps duration corresponds to an intensity of 230 GW/cm^2 or a fluence of 7 J/cm^2. Over a wide range of pulselengths (approximately 1 ps to 1 μs), the threshold intensity for laser damage decreases with pulse length T as $T^{-1/2}$ and correspondingly the threshold fluence for laser damage increases with pulse length $T^{1/2}$. In this range of pulse durations, the dominant mechanism of laser damage is avalanche breakdown. In this process, free electrons are accelerated by the laser field until they acquire sufficient energy to impact-ionize other atoms in the sample. These additional electrons are similarly accelerated and create still more free electrons. The combined action of the breaking of chemical bonds and the deposition of heat energy leads to the fracturing of the optical material. For pulses shorter than 1 ps, processes such as multiphoton absorption and multiphoton dissociation contribute to the mechanism of optical damage. For laser pulses longer than approximately 1 μs (including continuous wave laser beams), the dominant damage mechanism is direct heating of the optical material by linear absorption.

High-Harmonic Generation

Let us consider how nonlinear optical effects are modified when excited by a super-intense pulse. Nonlinear optical effects are traditionally modeled using a power-series expansion, such as

$$P = \chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \cdots \qquad [8]$$

but this series is not expected to converge if the laser field strength E exceeds the atomic unit of field strength $E_{\mathrm{at}} = e/a_0^2 = 2\times10^7$ statvolt/cm $= 6\times10^9$ V/cm. This field strength corresponds to a laser intensity of $I_{\mathrm{at}} = 4\times10^{16}$ W/cm^2, which constitutes the threshold intensity for exciting nonperturbative nonlinear optical response.

One of the consequences of excitation with intensities comparable to the atomic unit of intensity I_{at}, is the occurrence of high-harmonic generation. In a typical experimental arrangement, a gas jet is irradiated by high-intensity laser radiation, and all odd harmonics of the fundamental laser frequency, up to some maximum value N_{max}, are observed. The various harmonics below N_{max} are typically emitted with approximately equal intensity; such an

observation is incompatible with a perturbative explanation of this phenomenon. Recent work has demonstrated harmonic generation with $N_{\max}$ as large as 341.

The phenomenon of high-harmonic generation can be understood in terms of a simple physical model. One imagines an atomic electron that has received kinetic energy from the laser field and is excited to a highly elliptical orbit. The positively charged atomic nucleus is at one focus of this ellipse, and each time the electron passes near the nucleus it undergoes strong acceleration and emits a short pulse of radiation. This radiation will occur in the form of a train of short pulses; the spectrum of the radiation will be the square of the Fourier transform of this pulse train, which will contain the odd harmonics of the oscillation frequency up to some maximum frequency, that is approximately the inverse of the time the electron spends near the atomic core. This argument can be made quantitative to show that the maximum harmonic number is given by

$$N_{\max}\hbar\omega = 3.17K + U_p \qquad [9]$$

where $K = e^2E^2/m\omega^2$ is the 'ponderomotive energy' (the kinetic energy of an electron in a laser field) and U_p is the ionization energy of the atom.

Plasma Nonlinearities and Relativistic Effects

The process of multiphoton ionization can liberate a sufficient number of free electrons to transform the optical medium into a plasma, that is, a fully or partially ionized gas. The process of plasma formation is described by the equation

$$\frac{\mathrm{d}N_\mathrm{e}}{\mathrm{d}t} = \frac{\mathrm{d}N_\mathrm{i}}{\mathrm{d}t} = (N_\mathrm{T} - N_\mathrm{i})\sigma^{(N)}I^N - rN_\mathrm{e}N_\mathrm{i} \qquad [10]$$

where N_e is the number density of electons, N_i is the number density of ions, N_T is the total number of atoms (both ionized and un-ionized) in the material, and r is the electron–ion recombination coefficient. The optical properties of plasmas are very different from those of typical dielectric materials; the plasma contribution to the dielectric constant is given by

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2} \qquad [11]$$

where $\omega_p = \sqrt{4\pi Ne^2/m}$ is known as the plasma frequency.

Nonlinear effects can occur in the propagation of light through a plasma. One example is the nonlinear response resulting from the relativistic change in mass of the electron due to the large velocity that it can attain in the field of an intense laser beam. Detailed consideration of this effect shows that the nonlinear change in refractive index can be described as $\Delta n = n_2 I$ where

$$n_2 = \frac{2\pi\omega_p^2 e^2}{n_0^2 m^2 c^3 \omega^4} = \frac{1}{2\pi n_0^2}\left(\frac{\omega_p}{\omega}\right)^2 1.1\times 10^{-26}\,\frac{\mathrm{cm}^2}{\mathrm{W}} \qquad [12]$$

Nonlinear Quantum Electrodynamics

One can imagine an electric field so intense that it could lead to the spontaneous creation of electron–positron pairs. Such a field would have a strength of the order of $E_\mathrm{QED} = mc^2/e\lambda\!\!\!^{-}_c$ where $\lambda\!\!\!^{-}_c = \hbar/mc$ is the reduced Compton wavelength of the electron. The intensity of a light beam with a peak field amplitude of E_QED is $I_\mathrm{QED} = 4\times 10^{29}$ W/cm^2. Detailed consideration shows that even for fields weaker than I_QED there will be a field-induced change in the dielectric tensor given by

$$\epsilon_{ik} = \delta_{ik} + \frac{e^4\hbar}{45\pi m^4 c^7}\left[2(E^2 - B^2)\delta_{ik} + 7B_iB_k\right] \qquad [13]$$

Because of the unusual tensor properties of this relation, it displays a different polarization dependence than typical optical nonlinearities. Nonetheless, to an order of magnitude one can describe the strength of this response as

$$n_2 = \frac{7}{15c}\frac{e^2}{\hbar c}\frac{1}{E_\mathrm{QED}^2} = 5.6\times 10^{-34}\ \mathrm{cm}^2/\mathrm{W} \qquad [14]$$

See also

Fiber and Guided Wave Optics: Nonlinear Optics. **Quantum Electrodynamics:** Quantum Theory of the Electromagnetic Field.

Further Reading

Agostini P, Fabre F, Mainfray G, Petite G and Rahman NK (1979) Free-free transitions following six photon ionization of xenon. *Physics Review Letters* 42: 1127.

Alfano RR (ed.) (1989) *The Supercontinuum Laser Source*. New York: Springer-Verlag.

Bloembergen N (1997) A brief history of light breakdown. *Journal of Nonlinear Optical Physics and Materials* 6: 377.

Boyd RW (2003) *Nonlinear Optics, Second Edition*, Section 12.5. Amsterdam: Academic Press.

Braun A, Korn G, Liu X, Du D, Squier J and Mourou G (1995) Self-channeling of high-peak-power femtosecond laser pulses in air. *Optical Letters* 20: 73.
Brabec T and Krausz F (2000) Intense few-cycle laser fields: Frontiers of nonlinear optics. *Reviews of Modern Physics* 72: 545.
Chang Z, Rundquist A, Wang H, Murnane MM and Kapteyn HC (1997) Generation of coherent soft X-rays at 2.7 nm using high harmonics. *Physics Review Letters* 79: 2967.
Corkum PB (1993) Plasma perspective on strong field multiphoton ionization. *Physics Review Letters* 71: 1994.
Corkum PB, Rolland C and Srinivasanrao T (1986) Supercontinum generation in gases. *Physics Review Letters* 57: 2268.
Gaeta AL (2003) Collapsing light really shines. *Science* 301: 54.
Kasparian J, Rodriguez M, Mejean G, Yu J, Salmon E, Wille H, Bourayou R, Frey S, André Y-B, Mysyrowicz A, Sauerbrey R, Wolf J-P and Woste L (2003) White light filaments for atmospheric analysis. *Science* 301: 61.
Lewenstein M, Balcou Ph, Ivanov M Yu, L'Huillier A and Corkum PB (1994) Theory of high-harmonic generation by low-frequency laser fields. *Physics Review A* 49: 2117.
Max CE, Arons J and Langdon AB (1974) Self-modulation and self-focusing of electromagnetic waves in plasmas. *Physics Review Letters* 33: 209.
Moll KD, Fibich G and Gaeta AL (2003) Self-similar wave collapse: observation of the Townes profile. *Physics Review Letters* 90: 203902.
Monot P, Auguste T, Gibbon P, Jakober F, Mainfray G, Dulieu A, Louis-Jacquet M, Malka G and Miquel JL (1995) Experimental demonstration of relativistic self-channeling of a multiterawatt laser pulse in an underdense plasma. *Physics Review Letters* 74: 2953.
Rairoux P, Schillinger H, Niedermeier S, *et al.* (2000) *Applied Physics B—Lasers Optics* 71: 573.
Ranka JK, Schirmer RW and Gaeta AL (1996) Observation of pulse splitting in nonlinear dispersive media. *Physics Review Letters* 77: 3783.
Rothenberg JE (1992) Pulse splitting during self-focusing in normally dispersive media. *Optical Letters* 17: 583.
Sprangle P, Tang C-M and Esarey E (1987) Relativistic self-focusing of short-pulse radiation beams in plasmas. *IEEE Transactions on Plasma Science* 15: 145.
Stuart BC, Feit D, Rubenchik AM, Shore BW and Perry MD (1995) Laser-induced damage in dielectrics with nanosecond to subpicosecond pulses. *Physics Review Letters* 74: 2248.
Wagner R, Chen S-Y, Maksemchak A and Umstadter D (1997) Electron acceleration by a laser wakefield in a relativistically self guided channel. *Physics Review Letters* 78: 3125.
Wood RM (1986) *Laser Damage in Optical Materials*. Bristol, UK: Adam Hilger.

NONLINEAR SOURCES

Harmonic Generation in Gases

P Villoresi, University of Padova and INFM, Padova, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The generation of laser harmonics in gases provides high-brightness ultraviolet and soft X-ray radiation induced by an intense laser pulse. This phenomenon was discovered experimentally in the late 1980s, in a simple setup as that sketched in Figure 1. The laser pulse is focused in the interaction medium, which may be a gas jet formed by a pulsed electromagnetic valve, a capillary, or a small chamber filled with the gas. The beam leaving the interaction region includes a part of the laser pulse and a number of odd harmonics.

The harmonic orders generated in this way may reach and exceed an order of 300, so that this mechanism, often referred as high-order harmonics (HHs) generation (HHG), directly up-converts low frequency radiation, usually in the near-infrared spectrum, to extreme-ultraviolet (XUV) or soft X-rays. An example of a HH spectrum is shown in Figure 2, where several tens of discrete harmonic peaks are present. The striking difference with ordinary nonlinear processes is the flatness of the photon yield versus nonlinear order, up to a sharp cut-off point.

The HHG is based on the sequence of three processes: the electric field of an intense laser pulse ionizes atoms placed near the laser focus; the liberated electrons are driven by the laser electric field in a trajectory which returns on the parent ion; and the electron recombines releasing a photon whose energy is the ionization potential plus the electron's kinetic energy. The process is periodic in time, with the period of the laser pulse, and the emitted spectrum is then expressed by discrete components, multiples of the laser frequency. The emission is in the form of a traveling wave in the direction of the laser beam and so only odd harmonics are present.

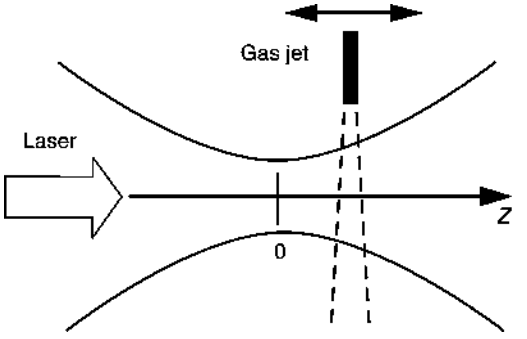

Figure 1 Scheme of the interaction between a focused laser pulse and a gas jet. *z* axis originates at laser focus.

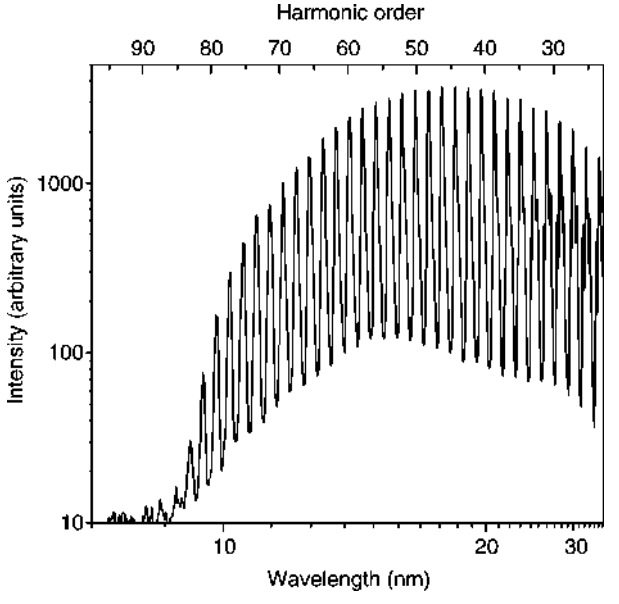

Figure 2 HHs experimental spectrum, obtained from helium gas, and a Ti:sapphire laser pulse at 790 nm, with 25 fs duration and 6×10^{14} W/cm^2 intensity.

The individual processes in HHG have highlighted the coherent response of electrons in the laser field, that show up interferences between quantum trajectories, as well as a nonlinear phase-matching issue – which have provided a simple method to investigate the interaction of single particles with the electromagnetic field, where plasma effects and bonds may be of minor influence.

Background

The interest to produce laser harmonics using gases instead of crystals was expressed by many. For the gas medium; the mass-density is three order of magnitude lower than for solids and the nonlinear susceptibility can be strongly enhanced for suitable resonance frequencies. Moreover, the phase-matching of the phase velocities may be achieved by tuning the gas pressure or by mixing diverse gases, so realizing a highly homogenous conversion region, free from walk-off and with an extension much longer than for crystals. In addition, gases have a high threshold for the intensity-induced damage, which is also non-permanent.

This research was made possible with the introduction of chirped-pulse laser amplifiers in the early 1980s, when picosecond lasers pulses reached intensities capable of directly ionizing solids and matter with their own electric field. This effect, known as optical ionization, was further enhanced when the laser pulses were reduced to subpicosecond duration, up to a few femtoseconds (10^{-15} s), close to their natural limit, which is the optical cycle. The experimental observation of HHs and of the ejection of electrons with unexpected high energy, the above threshold ionization (ATI), dramatically changed the scenario of laser matter interaction and has indicated a new regime in the coupling of laser radiation with the matter. The ordinary nonlinear optics, using the perturbative approach, was not capable of interpreting the findings and so the request was made for new models for the description of the interaction, which has been now named nonperturbative.

Typical Experimental Parameters

In the experimental investigation of the HHG, many laser wavelengths and pulse durations were used, including radiation from excimer, visible, Ti:sapphire, Nd:YAG (first and second harmonics) and CO_2 lasers, and pulses of duration from hundreds of picoseconds to a few femtoseconds. Among these, the more effective strategy in the generation of high harmonic order with high conversion efficiency, has been that of near infrared pulses with shortest duration. In order to induce optical single-ionization, the required laser intensity needs to span in the 10^{14} W/cm^2 range. Different elements are a possible choice for the gas in the interaction medium, though noble gases are used for the generation of the highest orders, since these have the higher ionization potential. Due to the very short duration of the laser pulse, the gas temperature is not relevant, while a pressure range of 10–500 mbar is usually chosen to balance, on one side, the needs of a large number of atoms that take part in the HHG and on the other side, the limit posed by the attenuation of the XUV and soft X-ray radiation after the generation, caused by the gas absorption. For the same reason, a powerful pumping system is needed, in order to reduce the ambient pressure in the apparatus below 10^{-5} mbar level.

The Models for the High Harmonics Generation

The physical processes involved in HHG are many: in the emission of the HH photons, the laser field interacts with a single atom of gas and induces the ionization, the electron's motion and its recollision with its parent ion; besides, the beam of HHs is the result of the superposition of the radiation from all the ionized atoms, as in an ordinary nonlinear medium, where the phase velocity of the laser field and of HHs is a function of the position and of the laser intensity and phase; therefore, the phase-matching issue has to be addressed in order to understand the global yield of HHG. The more relevant causes for the mismatch are the geometrical phase shift at the beam focus, which for Gaussian-profile beams is the Gouy phase, the nonlinear phase from the atomic polarization, and the effects of the plasma induced by the laser pulse itself. In addition, other effects have to be considered, such as the nonuniform density distribution of the target gas or the fact that the propagation of XUV radiation through a gas is usually influenced by a strong absorption, which is frequency dependent.

The study of HHG is first approached by introducing what is known as the simple man model, in which simplifications are introduced in order to treat the problem in a simple classical scheme, but from which the essential characteristics are derived. We then introduce the more complex quantum model based on the path integrals.

First Step: Ionization

The first step in the HHG model is the process with which the laser electric field applies to an electron, belonging to an atom or a molecule, a force that causes its detachment. This takes place only if such a field is strong enough to liberate the electron from its atomic bond. This may happen when the laser, which is taken here as linearly polarized and periodic in time, has modified the potential well so that the electrons may tunnel through it or pass over it. In fact, the sum of the atomic potential with that associated to the laser, which can be thought of as a plane that swings between a positive and negative slope at the laser frequency, gives rise to a finite sized well, through which the electron escapes by tunneling, or eventually to an aperture from which the electron is drawn away from the nucleus. In Figure 3, is shown such a well and the tunnel path for the electron.

Once freed, the electron responds to the periodic force from the laser which makes it oscillate. The average kinetic energy on such oscillations is known as the ponderomotive potential U_p, and is related to the laser intensity and wavelength, according to the following: $U_p = e^2E_0^2/4m\omega_0^2$, which reads numerically as $U_p(\text{eV}) = 9.3I_0(10^{14}\ \text{W/cm}^2)\lambda^2(\mu\text{m})$, where λ is the laser wavelength, E_0 and I_0 the laser peak electric field and intensity.

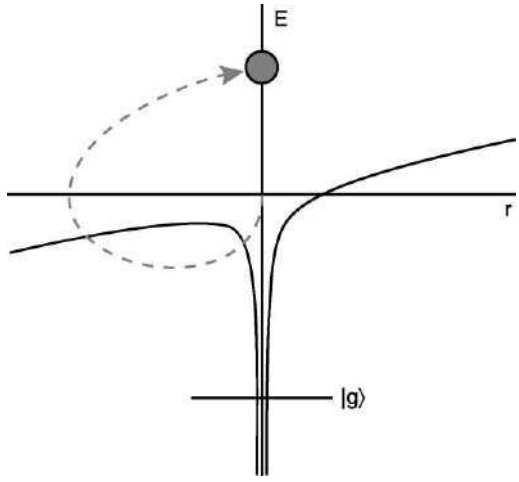

Figure 3 Electron potential when the laser pulse shines on the atom; |g⟩ indicated the level of the ground state; in red, a possible trajectory of the liberated electron after tunneling.

Second Step: Free-electron Trajectory

On the freed electron are applied both the time-varying force due to the laser field and the ionic potential. Its motion can be approximated according to the simple man model as a nonrelativistic classical-mechanics trajectory; its initial velocity is set to zero, and its initial position is by the ion; finally, the laser exerts a force much stronger than that from the ion as well as from the laser magnetic field, thus these two latter forces are cancelled out.

As mentioned above, the electron trajectories that induce HHs are those where the electron returns by the ion with strong kinetic energy. If the laser is linearly polarized, the induced trajectory takes place along a line, at which also belongs to the parent ion. Moreover, the electron final velocity is related to the actual instant of ionization, or phase of the field, with respect to the laser peak. Therefore, electron final energies span from zero to a maximum value $K_{\text{max.}}$ By considering the first semi-cycle, K_{max} occurs at about 17° of field phase angle, while for the second semi-cycle it is at 197°.

In Figure 4 the electron trajectories for different ionization in the first semi-cycle are shown: the electron displacement in the laser-field direction with

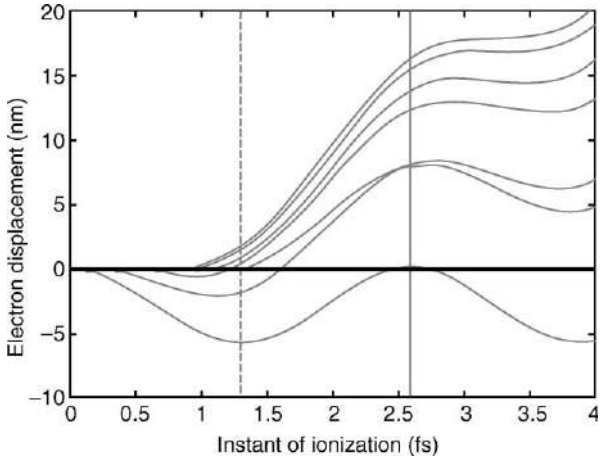

Figure 4 Electron trajectories relative to different ionization instant within the first semi-cycle; in dotted red, half-cycle, in continuous red, the whole optical cycle. Trajectories for later times do not return to the ion, do not contribute to HHG and correspond to ejected photoelectrons.

respect to the ion position, which is the initial as well as the final position, is represented versus time, where the time origin is at the laser peak, where field phase is equal to zero. The laser intensity directly influences the value of $K_{\max}$, according to the following expression: $K_{\max} = 3.17\,U_p$. In the case shown in the figure, the peak laser intensity is $I_0 = 8 \times 10^{14}$ W/cm^2 and the ponderomotive potential is $U_p = 45$ eV and $K_{\max} = 143$ eV.

Moreover, for a given final energy below $K_{\max}$, there are more than one value of the phase of the laser field where the trajectory results with the same final kinetic energy. This indicates that multiple trajectories may produce a given HH, and this is relevant in terms of the interference between them.

Third Step: Recombination

The final process which occurs in the HHG, is the recombination of the ionized electron with the parent ion. The energy conservation means that the kinetic energy and the ionization potential have to be delivered to the emitted photon. In the simple man model, the spectrum of HH then spans from I_p to the maximum values of $E_{\max} = I_p + 3.17\,K_{\max}$. In the case of Figure 2, obtained using helium gas at 6×10^{14} W/cm^2, the spectral position of cutoff can be estimated at about 9 nm, or 140 eV, which is in good agreement with the prediction of the relationship described above.

The electron recombination is not a unique process occurring at this point: the electron may be scattered by the ion and be emitted with a kinetic energy of up to 10 U_p, which is known as the rescattering plateau in ATI, or it may ionize the ion by collision, leaving a double ionized atom via sequential ionization, with a strong increase in the observed cross-section of this process.

Feynman Path-Integral Approach

The ideas in the simple man model have been expanded in order to include the quantum mechanical description of the laser-electron-ion interaction. By using the Feynman path-integral approach, the evolution of the electron driven by the laser force is associated to a probability which depends on the sum of an infinite number of amplitudes $\mathbf{f}_n$. These are complex numbers and correspond to different trajectories, or quantum orbits, $\mathbf{r}_n(t)$, followed by the electron. Their phase is proportional to the classical action of the electron motion, $S[\mathbf{r}_n(t)]$, calculated along $\mathbf{r}_n(t)$, which begins at time t_i and ends at a later time $t_f = t_i + \tau$. Here τ defines the duration of the liberated electron trajectory. In our case, the action is the sum of three terms:

$$S[\mathbf{r}_n(t)] = S_{\text{bound},n} + S_{\text{free},n} + S_{\text{final},n} \qquad [1]$$

that correspond respectively to the electron ionization:

$$S_{\text{bound},n} = -E_0 t_{i,n} \qquad [2]$$

where E_0 is the binding energy of electron in the atom ground state, $E_0 < 0$, to the free electron trajectory:

$$S_{\text{free},n} = \int_{t_i}^{t_f} [\mathbf{k} - e\mathbf{A}(t)^2]\mathrm{d}t \qquad [3]$$

where $\mathbf{k}$ is the electron momentum after ionization, directed parallel the laser electric field, and $\mathbf{A}(t)$ is the vector potential of the laser field, and to the final recombination with the parent ion:

$$S_{\text{final},n} = (\hbar\omega_q + E_0)t_{f,n} \qquad [4]$$

The condition that normally points out the trajectory made by the electron may be obtained from the principle of least action. Within the path integral approach, this principle can be used to deduce some equations which have to be obeyed by the quantum orbits, and that express the energy conservation in both the ionization and recombination process and the fulfilment of boundary conditions for electron trajectory. These are known as saddle-point equations, and reads as:

$$[\mathbf{k} - e\mathbf{A}(t_i)]^2 = 2mE_0 \qquad [5]$$

for the ionization process:

$$\mathbf{k}\tau = e\int_{t_i}^{t_f} \mathbf{A}(t)\mathrm{d}t \qquad [6]$$

which compares the initial momentum with that exchanged with the laser field during the trajectory, and

$$[\mathbf{k} - e\mathbf{A}(t_f)]^2 = 2m(\hbar\omega_q + E_0) \qquad [7]$$

for the recombination process which emits an harmonic of order q.

From eqn [5] we note that the solution for the ionization time t_i is a complex number, since the right hand side is negative. This suggests that the quantum orbits are complex as well. The reason for this is that the tunnel through the potential well is a nonclassical process.

The orbits solutions to eqns [5]–[7] have different durations, and for some of them τ is shorter than an optical cycle, while for the others the electron returns in a later cycle than the one where it was ionized. This latter solution is of minor influence, since the electron wavefunction spreads out along the trajectory, decreasing the recombination probability. Those within the initial optical cycle sum the amplitudes to give the probability of the generation of q-th harmonic, and their interference is evident from the irregular single-atom spectrum.

The time evolution of the laser pulse as well as the interaction geometry are experimental tools that influence the relevance of the different trajectories. In particular conditions it is possible to set the condition for HHG from a single trajectory, with a very simple phase structure and clear spectrum; some cases are described below.

HHs Generated by Few-Optical-Cycle Pulses

The laser electric field considered so far has been taken as stationary, relying on the hypothesis that the pulse envelope has a duration much longer than the laser optical cycle. For the analysis of the electron trajectories, that last a fraction of the laser optical cycle, this hypothesis is valid for pulses down to about 20 fs. For shorter pulses, as those generated with the hollow fiber technique, there is a noticeable variation of the electric field between a cycle and the next. The generation process is now aperiodic, and the characteristics of HH spectrum and conversion is strongly influenced by the pulse actual duration.

In order to show the striking changes in this regime, Figure 5 shows the laser electric field, the envelope and the ionization probability of helium gas in the case of a pulse of 5 fs of full width at half-maximum (FWHM) duration and a laser intensity of 9×10^{14} W/cm^2. The atom can be ionized during only three intervals, of sub-fs duration but where only one is dominant.

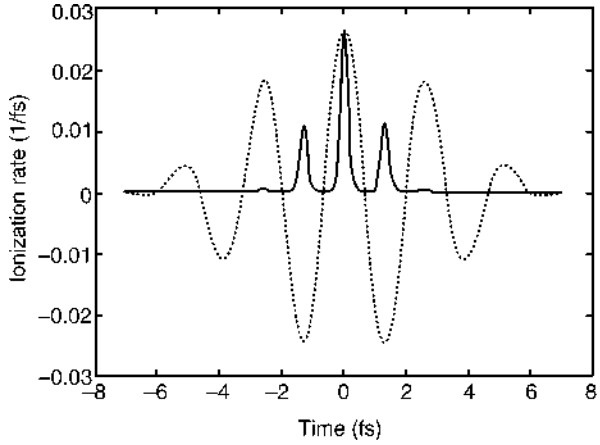

Figure 5 Ionization rate for helium gas shined by a 5 fs FWHM pulse of intensity 9×10^{14} W/cm^2. The dotted line shows the normalized evolution of electric field of the pulse. Only the three central half-cycles induce ionization.

The HHG experiment in this temporal regime have revealed that the influence of the plasma in the interaction region is progressively reduced and eventually vanishes for pulses of duration of about 5 fs. This is because the atoms are exposed directly to the maximum electric field, and so there is no pre-ionization of the gas caused by the initial part of the pulse. This effect, which is present in the case of many-cycle pulses, causes a reduction in the HHG efficiency, due to a depletion of the neutral density available at the time of maximum electric field, where the higher HHs are produced with a large dispersion, which deteriorates the phase-matching.

Moreover, the effect of the carrier-envelope-phase in the HHs spectrum has been demonstrated as an evident spectral shift of the cut-off harmonics generated with few-cycle pulses. The reason can be found in the variation of the phase of the emission from each electron after its trajectory. For the few-cycle pulses, this phase changes for every laser semi-cycle, due to the change in the actual electric field which drives the electron motion.

This portion of the spectrum is also relevant to its phase structure, since here the HHs are generated by a single electron trajectory. This condition sets a constant phase difference between the subsequent harmonic orders. In this way, similarly to the mode locking laser, this effect allows the synthesis of a pulse, from the combination of several harmonic orders, which has a duration much lower than the

individual harmonics. This process has been exploited to produce a soft X-ray pulse of duration of 650 as (1 as, attosecond = 1×10^{-18} s).

HHs Beam Characteristics

The HHG is a coherent process, in that the emitted radiation has a direct relationship with the laser pump. The emitted HH beam also reflects this fact in its good directional characteristics and in its spatial and temporal coherence, as different experiments have demonstrated.

Using a 110 fs laser pulse, the interference fringes produced in the far field of two spatially separated HH beams were measured as a function of the mutual delay. The HH coherence time t_c resulted in agreement with the expected duration of the HHs, which is a fraction of that of the laser pulse. Moreover, it has been observed at a different t_c in the central and external regions of the far field of harmonic 23rd, 50 fs and 20 fs respectively. This is because different phase-matching in the HHG contribute to the on-axis and off-axis part of the HHs beam, where the more prominent quantum orbit is also different.

The measurement of the beam divergence for different harmonic orders is relevant for both the estimate of the beam brightness and the irradiance available in the case of the HH applications. Moreover, it is useful for the understanding of the phase-matching which is realized in actual interaction geometry. The experimental measurements requires an astigmatic spectrometer, as is described later. In this kind of experiment, the focalplane images of the HH spectrum are taken for different generation conditions, usually moving the gas-jet position with respect to the laser focus, the z coordinate in Figure 1. The harmonic peaks appear as elongated spots, as shown in Figure 6 for helium gas and an intensity of 4×10^{14} W/cm^2 and duration of 20 fs, whose length in a transverse direction to the dispersion plane can be related, by a optical calculation, to the beam divergence.

A comparison between divergences for HHG driven by few- or many-optical-cycle pulses has revealed another aspect of the phase-matching condition. In this case, the laser pulse was compressed in time by means of the hollow fiber technique, and so the resulting phase front is of the Bessel-beam type, truncated at the first zero. The experiment indicates a divergence for plateau harmonics of about 3 mrad FWHM of the beam, in the case of the 7 fs pulse and larger values by more than 50% using the 20 fs pulse. In both cases, the beam divergence increase moving downstream the gas jet and with increasing harmonic order. The reduced divergence can be ascribed to a more regular wavefront of the harmonic beam, which can be related to the simpler phase structure of the few-cycle case.

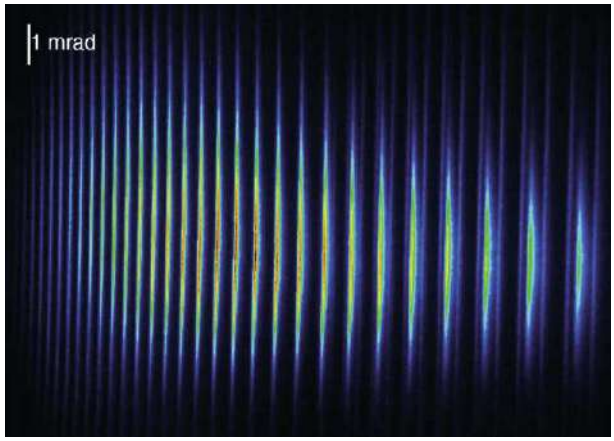

Figure 6 Focal plane image obtained from the spectrograph shown in Figure 9, relative to the HHs spectrum of helium with 20 fs and 4×10^{14} W/cm^2. The bar in upper-left corner indicates the scale-length of vertical axis corresponding to a divergence of the HH beam of 1 mrad.

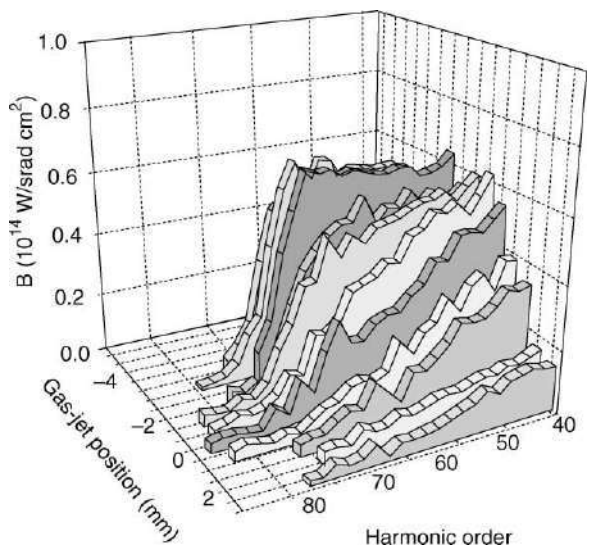

Figure 7 Peak brightness of the HHs beam determined from absolute-intensity and divergence measurements, using neon gas and a 790 nm laser pulse with 7 fs duration and 9×10^{14} W/cm^2 intensity.

By using an intensity-calibrated detector, the number of photons per harmonic order per laser shot can be measured. Combining this measure with that of the divergence, the HH beam brightness can be calculated. The source size and the HH pulse duration can be easily modeled from the experimental parameters. The result for the case of neon with 7 fs pump-pulse with intensity of 9×10^{14} W/cm^2 is shown in Figure 7, as a function of the gas-jet

position, with positive values when the jet is downstream with respect to laser focus.

The large divergence reduces the brightness of downstream positions. Maximum values of about 8×10^{13} W/srad cm^2 are found close to the focus, relatively flat with respect to harmonic order.

The conversion efficiency η_{HH} from the fundamental intensity to that of the HHs can also be derived from this kind of measurement. The more convenient parameters for HHG in the range 41–71 were found using neon gas jet from a pulsed valve and a pulse duration of 7 fs, which leads to η_{HH} in the scale of 10^{-7}, while for lower orders, from 17 to 31, a value about two orders of magnitude larger has been obtained using a short hollow fiber filled with argon gas. A remarkable intracycle optimization mechanism was introduced in this case, in order to maximize the given harmonic order of chosen. This was done by tailoring the shape of the pump pulse in spectrum and time, using an adaptive component in the laser compressor. In this case, the optimal pulse shape to maximize the efficiency in the driving of the electron on a trajectory producing the required harmonic order, is sought by means of a genetic algorithm.

A strong efficiency enhancement has also been demonstrated by manipulating the laser wavefront by means of a deformable mirror (DM). In this case, the DM was controlled by a genetic algorithm, aiming to increase the HHs intensity in a given spectral interval. The purpose of this technique is the correction of the optical aberrations on the beam, introduced by both the optical components in the beam focusing line and the phase distortion caused by nonlinear effects leading to an optimum laser front. The effect is shown in Figure 8, where the spectrum obtained with the original beam is compared with that optimized by the genetic algorithm. It can be noticed there is both a remarkable extension of the HHs spectrum, of about 20 orders, as well as a strong enhancement of the plateau intensity.

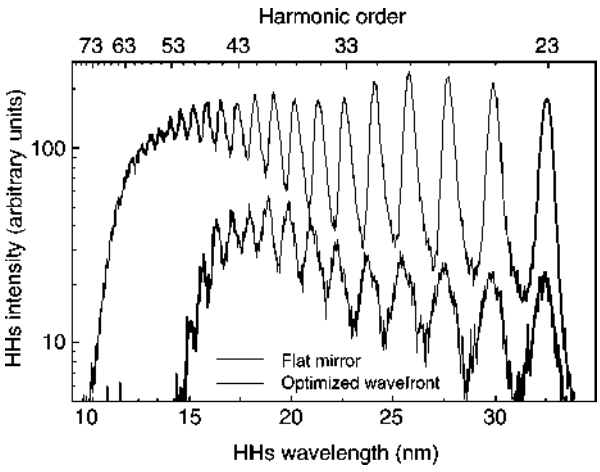

Figure 8 Comparison of the HHs spectrum obtained with and without the wavefront correction using a deformable mirror. Neon gas was used and a 790 nm laser pulse with 6 fs duration and 0.3 mJ energy.

Spectrometers and Monochromators for HHs

The extended spectrum of HHs reaches the optical domains of the extreme-ultraviolet (XUV, 10–100 nm) and soft X-ray (1–10 nm). In these regions, the optical materials have very poor or zero transmission, and so the optical design of instruments for HHs analysis uses reflective optics only. Exceptions sometimes used are the frequency-selective XUV transmission filters, whose thickness does not exceed a few tenths of a micrometer, and the diffracting zone-plates. Moreover, for the radiation of wavelength shorter than 30 nm, even the reflectivity of optics becomes poor, and this forces the use of grazing-incidence mountings or the narrow band XUV and soft X-ray multilayer mirrors.

On the other hand, HHs as optical sources have favorable characteristics, in that they are emitted in a region smaller than the laser waist focus, so usually of a few tens of micrometers, and their divergence is small, as seen above. In addition, the very short duration of the pump pulses broadens the HHs spectrum, due to time-bandwidth inequality, usually smearing details below 0.1 nm of width.

The spectrograph is then realized with the least number of components, usually combining in one or two optics of the spectral dispersion, the gathering of the HH beam and its focusing on the detector. Moreover, it is not feasible to collect with a single detector all the broad HH spectrum in once, so usually a spectral interval of interest is selected and acquired with a linear or bidimensional detector, or scanned sequentially by rotating the grating.

An example of spectrograph with 2D-detector, is shown in Figure 9, the layout of the instrument developed for the simultaneous measure of the divergence and the intensity of HHs generated from few-optical-cycle pulses.

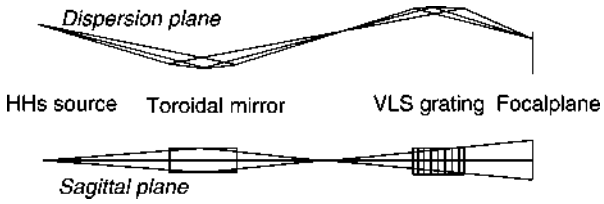

Figure 9 Optical layout of the astigmatic spectrometer for the divergence measurement.

The upper part of Figure 9 shows the optics in the dispersion plane, where the toroidal mirror gathers the HH radiation and relays it to the source point for the spherical grating. The latter has a varied line spacing (VLS) of the grooves, in order to image an erected spectrum in the focalplane. This feature is convenient in order to use a bidimensional detector. In the sagittal plane, the grating has negligible effect, and the toroidal mirror is designed to realize an astigmatic image of the source in the focalplane, from which the divergences can be extracted. The detector needs to be blind at the radiation of the laser and at the lower harmonics, and to have an extended linear response in the XUV and soft X-ray. A good solution can be obtained with an open micro-channel plate (MCP), with a suitable photocathode such as MgF_2, placed in the focalplane, coupled with a high-resolution and a high dynamic CCD camera.

For the selection of a single harmonic order or of a narrow spectral region, an XUV monochromator is used. This device is based on the spectral dispersion of the HHs beam by using a diffraction grating. The desired harmonic order is selected by scanning the wavelength which falls on the slit, usually by rotating the grating. This technique uses a single optical component, a concave grating, to reduce the reflection losses. On the other hand, the single-grating monochromator introduces a remarkable time broadening in the selected radiation. In fact, the grating diffraction is based on the optical path length difference for the rays diffracted by subsequent grating grooves. Therefore the selected beam, of wavelength λ, is composed of rays whose difference in length is $\Delta L = N_g m\lambda$, where m is the diffraction order and N_g is the number of illuminated grooves, given by the beam diameter on grating multiplied by the groove density. The time broadening of the selected pulse is then $\Delta\tau = \Delta L/c$. For a HH XUV pulse, the diffracted pulse results is 7 ps of duration, if beam size is of 3 mm and it is diffracted at 87° by a 1200 gr/mm grating.

The solution to this problem may be found in adopting a twin grating design. In this case, the dispersion induced by the first grating is used to select the desired spectral portion of the HH spectrum, and will be cancelled by the second grating. This latter grating has to be mounted in order to compensate for the different optical path length of different rays within the aperture of the HH beam. The layout of this optical scheme is reported in Figure 10. The ray tracing of this compensated monochromator indicates that for both the grazing incidence and the normal incidence regions, the residual broadening results is of the order of one femtosecond.

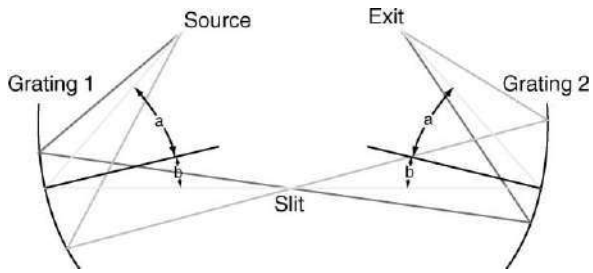

Figure 10 Scheme of the compensated XUV monochromator, for the selection of a spectral portion of the HHs spectrum without introducing temporal broadening. See also Villoresi P (1999) Compensation of optical path lengths in extreme-ultraviolet and soft X-ray monochromators for ultrafast pulses. *Applied Optics* 38: 6040.

HHs at Work: Applications and Perspectives

The advantages of HHs with respect to other XUV and soft X-ray sources, as laser-plasma, synchrotron radiation or X-ray lasers, are mainly the very short duration of their emission, their high brightness, and the source compactness and their intrinsic synchronization with a laser pulse. These aspects have been already exploited in a few experiments, based either on the detection of a photoelectron emitted in the time-dependent interaction of the matter with the HHs plus fundamental laser beams, or the observation of changes in the HHs XUV spectrum.

As examples of direct application, the interferometry of a laser-plasma has been obtained with HHs, to determine spatial distribution of the electron density. In this case, two beams of harmonic radiation were generated by splitting in equal parts the laser pulse, with one of them passing through a laser-generated plasma from a solid aluminum target. By means of a grating monochromator, a single harmonic order is selected, of both beams, and their interference pattern in the far field is acquired. From the shift in the fringes pattern, the map of the plasma electron density was estimated. Here, the broad spectrum of the HHs was exploited to operate the interference at the most convenient wavelength, far from atomic resonances and with fair fringe contrast.

The direct monitoring of a chemical reaction at the surface was achieved by pumping, with a part of the laser pulse, a platinum surface with some adsorbed molecular oxygen. By observing changes in the spectrum of the photoelectrons generated by the delayed HHs beam, the oxygen-platinum bond switch from a superoxo (O_2^-) to a peroxo (O_2^{2-}) state was observed in the time domain.

The electron bond in an atom may be dressed by a laser field, and the changes induced in its orbital are diagnosed by a HHs beam, which photo-ionizes the atom. This phenomenon have been exploited as a diagnostic of the temporal structure of a HHs beam: the sidebands in the photoelectron spectrum have been related to the phase of different harmonics. The findings indicate that HHs are locked in phase and their superposition produces a train of 250 as pulses.

By using the mixing scheme of some suitable orders of a HHs beam generated with few-optical-cycle pulses, mentioned already in section 'HHs beam characteristics,' a pulse with sub-femtosecond duration may be generated. This was used in the detailed analysis of the photoelectron spectrum in a pump-probe experiment, with time resolution of 150 as. Again using the IR and HHs radiation together, this technique has accessed the direct probe of the laser field oscillation, as demonstration by the use of HHs in the new domain of attosecond spectroscopy and metrology.

Finally, the HH radiation can also be considered as the seed for an amplifier of radiation in the soft X-ray region, as in the case of a free electron laser.

See also

Coherent Transients: Ultrafast Studies of Semiconductors. **Instrumentation:** Spectrometers. **Ultrafast Laser Techniques:** Generation of Femtosecond Pulses.

Further Reading

Bartels R, Backus S, Zeek E, *et al.* (2000) Shaped-pulse optimization of coherent emission of high-harmonic soft x-ray. *Nature (London)* 406: 164.

Bauer M, Lei C, Read K, *et al.* (2001) Direct observation of surface chemistry using ultrafast soft-X-ray pulses. *Physics Review Letters* 87: 025501.

Brabec T and Krausz F (2000) Intense few-cycle laser fields: frontiers of nonlinear optics. *Review Modern Physics* 72: 545.

Corkum PB (1993) Plasma perspective on strong-field multiphoton ionization. *Physics Review Letters* 71: 1994.

Descamps D, Lynga L, Norin J, *et al.* (2000) Extreme ultraviolet interferometry measurements with high-order harmonics. *Optic Letters* 25: 135.

Hentschel M, Kienberger R, Spielmann CH, *et al.* (2001) Attosecond metrology. *Nature (London)* 414: 509.

Kulander KC, Schafer KJ and Krause JL (1993) Dynamics of short-pulse excitation, ionization and harmonic conversion. In: Piraux B, L'Huillier A and Rzazewski K (eds) *Super-Intense Laser-Atom Physics*, pp. 95–110. New York: Plenum.

Lewenstein M, Balcou Ph, Ivanov M, *et al.* (1994) Theory of high-harmonic generation by low-frequency laser fields. *Physics Review A* 49: 2117.

L'Huillier A and Balcou P (1993) High-order harmonic generation in rare gases with a 1-ps 1053-nm laser. *Physics Review Letters* 70: 774.

McPherson A, Gibson G, Jara H, *et al.* (1987) Studies of multiphoton production of vacuum-ultraviolet radiation in the rare gasses. *Journal of the Optical Society of America B* 4: 595–601.

Nisoli M, Priori E, Sansone G, *et al.* (2002) High-brightness high-order harmonic generation by truncated bessel beams in the sub-10-fs regime. *Physics Review Letters* 88: 033902.

Nisoli M, Sansone G, Stagira S, *et al.* (2003) Effects of carrier-envelope phase differences of few-optical-cycle light pulses in single-shot high-order-harmonic spectra. *Physics Review Letters* 91: 213905.

Paul PM, Toma ES, Breger P, *et al.* (2001) Observation of a train of attosecond pulses from high harmonic generation. *Science* 292: 1689.

Poletto L, Tondello G and Villoresi P (2001) High-order laser harmonics detection in the EUV and soft x-ray spectral regions. *Review Scientific Instruments* 72: 2868.

Salières P, L'Huillier P, Antoine A, *et al.* (1999) Study of spatial and temporal coherence of high order harmonics. In: Bederson B and Walther H (eds) *Advances in Atomic, Molecular, and Optical Physics*, vol. 41, pp. 83. New York: Academic Press.

Villoresi P, Ceccherini P, Poletto L and Tondello G (2000) Spectral features and modeling of high-order harmonics generated by sub-10-fs pulses. *Physics Review Letters* 85: 2494.

Villoresi P, Bonora S, Pascolini M, *et al.* (2004) Optimization of high-order harmonic generation by adaptive control of sub-10 fs pulse wavefront. *Optic Letters* 29: 207.

O

OPTICAL AMPLIFIERS

Contents

Basic Concepts
Erbrium Doped Fiber Amplifiers for Lightwave Systems
Optical Amplifiers in Long-Haul Transmission Systems
Raman, Brillouin and Parametric Amplifiers
Semiconductor Optical Amplifiers

Basic Concepts

M F S Ferreira, University of Aveiro, Aveiro, Portugal

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The transmission distance of a fiber-optic communication system is limited by fiber loss and dispersion. For long-haul lightwave systems, the loss limitations have traditionally been overcome by periodic regeneration of the optical signals at repeaters applying conversion to an intermediate electric signal. Because of the complexity and high cost of such regenerators, the need for optical amplifiers became obvious in the mid-1980s. The optical amplifier is ideally a transparent box that provides gain and is also insensitive to the bit rate, modulation format, power and wavelength of the signal passing through it.

Several means of obtaining optical amplification has been suggested since the 1970s, including direct use of the transmission fiber as gain medium through nonlinear effects, semiconductor amplifiers, or doping optical waveguides with an active material (rare-earth ions), that could provide gain. Due to the spectacular results on erbium-doped fiber amplifiers, which are particularly suitable in the third transmission window (around 1.5 μm), an intense worldwide research activity on optical amplifiers has developed. As a consequence, the development of erbium-doped fiber amplifiers has reached an industrial level, and commercial devices are now available.

Semiconductor amplifiers, on the other hand, have the same technical basis as semiconductor lasers. Although the strong nonlinearity of semiconductor amplifiers degrades the performances of transmission systems, the state-of-art semiconductor devices seem to be the most interesting amplifiers for transmission in the second transmission window (around 1.3 μm).

Amplifier Gain and Bandwidth

In a perfect amplifier, the amplification process would be insensitive to the bit rate, modulation format, power, state of polarization, wavelength, and optical bandwidth of the signal passing through it. On the other hand, no interaction would take place if more than one signal were amplified simultaneously. In practice, however, the optical gain depends not only on the wavelength (or frequency) of the incident signal, but also on the electromagnetic field intensity at any point inside the amplifier. Details of wavelength and intensity dependence of the optical signal depend on the amplifying medium.

We consider a case in which the gain medium is modeled as a homogeneously broadened two-level system. In such medium, the gain coefficient (i.e., the gain per unit length) can be written as:

$$g(\nu, P) = \frac{g_0}{1 + \dfrac{(\nu - \nu_0)^2}{\Delta\nu_0^2} + \dfrac{P}{P_{\text{sat}}}} \qquad [1]$$

where g_0 is the peak value of the gain coefficient determined by the pumping level of the amplifier, ν the optical frequency, ν_0 the atomic transition frequency, $\Delta\nu_0$ the 3 dB local gain bandwidth, P the optical power of the signal, and P_{sat} the saturation power, which depends on the gain medium parameters. It must be emphasized that $\Delta\nu_0$ and P_{sat} refer to the local gain. However, from the communication system point of view, it is more desirable to use the related concepts of amplifier bandwidth and amplifier saturation power that will be evaluated below.

The amplifier gain G is defined as

$$G = \frac{P_{\text{out}}}{P_{\text{in}}} \qquad [2]$$

where P_{in} is the input power and P_{out} the output power of a continuous wave (CW) signal being amplified. The amplifier gain G may be found by using the relation:

$$\frac{\mathrm{d}P}{\mathrm{d}z} = g(\nu, P)P \qquad [3]$$

where $P(z)$ is the optical power at a distance z from the amplifier input end.

If the signal power obeys the condition $P \ll P_{\text{sat}}$ throughout the amplifier, the gain coefficient given by eqn [1] can be considered independent of the signal power. In such a case, the amplifier is said to be operated in the unsaturated regime and works as a linear device. The gain coefficient presents in this situation a Lorentzian profile that is characteristic of homogeneously broadened two-level systems. However, the gain spectrum of actual amplifiers can deviate significantly from the Lorentzian profile.

The solution of eqn [3] in the unsaturated regime is an exponentially growing signal power, given by

$$P(z) = P_{\text{in}} \exp(gz) \qquad [4]$$

For an amplifier length L, we then find that the linear amplifier gain is

$$G(\nu) = \exp(gL) = \exp\left[\frac{g_0 L}{1 + (\nu - \nu_0)^2/\Delta\nu_0^2}\right] \qquad [5]$$

Both the amplifier gain $G(\nu)$ and the gain coefficient $g(\nu)$ are maximum when $\nu = \nu_0$. However, $G(\nu)$ decreases much faster than $g(\nu)$ with the signal detuning $\nu - \nu_0$, because of the exponential dependence of G on g. As a consequence, the amplifier bandwidth $\Delta\nu_{\text{A}}$, which is defined as the FWHM of $G(\nu)$, is much smaller than the gain bandwidth $\Delta\nu_0$ (Figure 1). This can result in signal distortion in the case where a broadband optical signal is transmitted through the amplifier. From eqn [5] we can obtain the following relation between the amplifier bandwidth and the gain bandwidth:

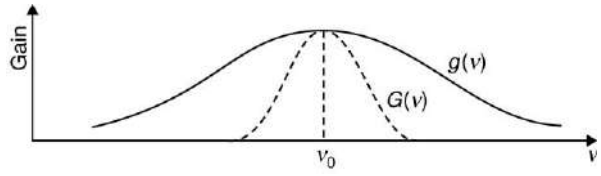

Figure 1 Gain coefficient profile $g(\nu)$ and the corresponding amplifier gain spectrum $G(\nu)$.

$$\Delta\nu_{\text{A}} = \Delta\nu_0 \sqrt{\frac{\ln 2}{g_0 L - \ln 2}} \qquad [6]$$

Gain Saturation

An important limitation of the nonideal amplifier is related with the power dependence of the gain coefficient given by eqn [1]. This property is known as gain saturation and it appears when the signal power ratio P/P_{sat} is non-negligible. Since the gain coefficient is reduced when the signal power P becomes comparable to the saturation power P_{sat}, the amplifier gain G will also decrease.

Assuming that $\nu = \nu_0$ and substituting g from eqn [1] in eqn [3] gives

$$\frac{\mathrm{d}P}{\mathrm{d}z} = \frac{g_0 P}{1 + P/P_{\text{sat}}} \qquad [7]$$

Considering the initial condition $P(0) = P_{\text{in}}$, we obtain from eqns [2] and [7] the following implicit relation for the amplifier gain:

$$(1 - G)\frac{P_{\text{in}}}{P_{\text{sat}}} = \ln\left(\frac{G}{G_0}\right) \qquad [8]$$

where $G_0 = \exp(g_0 L)$. The input saturation power P_{in}^{s} is defined as the input power for which the amplifier gain G is reduced by a factor of 2 from its unsaturated value G_0 (Figure 2). Indeed, it is obtained by using $G = G_0/2$ in eqn [8]:

$$P_{\text{in}}^{\text{s}} = \frac{2\ln(2)P_{\text{sat}}}{(G_0 - 2)} \qquad [9]$$

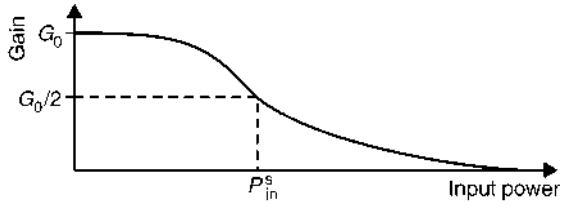

Figure 2 Saturated amplifier gain as a function of the input power.

As observed from eqn [9], the input saturation power P^s_{in} does not coincide with P_{sat}. The output saturation power is given by $P^s_{out} = G_0 P^s_{in}/2$. In practice, $G_0 \gg 2$ and P^s_{out} is found to be smaller than P_{sat}, by about 30%.

Gain saturation can be seen as a serious limitation, particularly for multichannel communication systems. However, the self-regulating effect of gain saturated amplifiers can be useful in long-haul lightwave communication systems, including many concatenated amplifiers. In fact, if the signal level in a chain of amplifiers is unexpectedly increased along the chain, the saturation effect causes a lower gain provided by the following amplifiers and vice versa for a sudden signal power decrease.

Amplifier Noise

Besides the bandwidth and gain saturation limitations, another property must be considered concerning practical optical amplifiers. In fact, optical amplifiers always add spontaneously emitted photons to the signal during the amplification process. Those photons are amplified besides the signal photons so that, at the amplifier output, an amplified spontaneous emission (ASE) noise is presented. Since spontaneous emission always takes place, ASE noise is unavoidable and does not depend on the amplifier temperature. This is one of the most important differences between optical and electrical amplifiers, where amplifier noise is of thermal origin and can be reduced by lowering the amplifier temperature.

The ASE determines a degradation of the signal-to-noise ratio (SNR). The SNR degradation is usually characterized by the amplifier noise figure, which is defined as the SNR ratio between input and output:

$$\mathrm{NF} = \frac{\mathrm{SNR_{in}}}{\mathrm{SNR_{out}}} \qquad [10]$$

The SNR is usually referred to the electrical power generated when the optical signal is converted to electrical current by using a photodetector. Therefore, the noise figure as defined in eqn [10] would usually depend on several detector parameters, which determine the shot noise and thermal noise associated with the practical detector. We will consider the case of an ideal detector, whose performance is limited by shot noise only.

The SNR of the input signal is simply determined by the detection shot noise and can be written as:

$$\mathrm{SNR_{in}} = \frac{P_{in}}{2h\nu\Delta f} \qquad [11]$$

where Δf is the detector bandwidth.

To evaluate the term $\mathrm{SNR_{out}}$, one should add the contribution of spontaneous emission to the receiver noise. The ASE noise spectral density is assumed to be constant and can be written as

$$S_{sp}(\nu) = (G-1)n_{sp}h\nu \qquad [12]$$

where G the amplifier gain and

$$n_{sp} = \frac{N_1}{N_1 - N_0} \qquad [13]$$

is known as the spontaneous emission factor or the population inversion factor. In eqn [13] N_0 and N_1 are the atomic populations for the ground and excited states, respectively.

Considering a low noise amplifier, the signal power impinging the photodetector is larger than the optical noise power and the shot noise power. As a consequence, the electrical noise, due to the signal-ASE beating, is the dominant contribution and the SNR of the amplified signal is given by:

$$\mathrm{SNR_{out}} \approx \frac{GP_{in}}{4S_{sp}\Delta f} \qquad [14]$$

Using eqns [11]–[14], the amplifier noise figure defined by eqn [10] becomes:

$$\mathrm{NF} = 2n_{sp}\frac{G-1}{G} \approx 2n_{sp} \qquad [15]$$

where the approximation holds when the gain is much higher than one. In the case of an ideal amplifier, $n_{sp} = 1$ and eqn [15] show that the SNR is degraded by 3 dB. For most practical amplifiers, NF can be as large as 6–8 dB.

Basic Amplifier Configurations

The relative importance of the different limiting factors discussed above depends on the actual

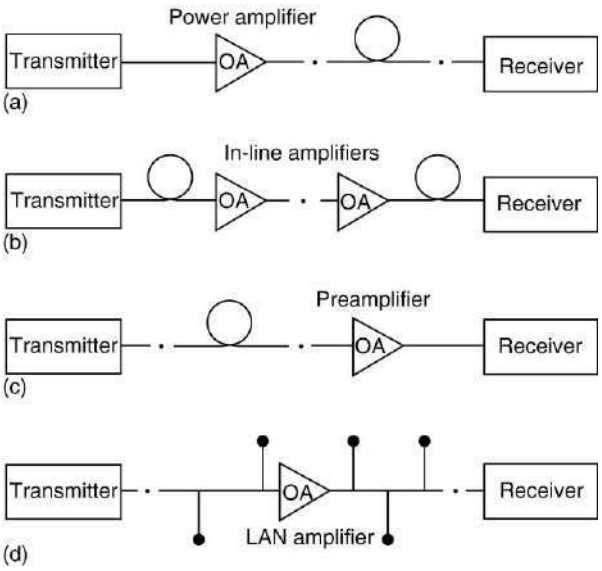

Figure 3 Four generic configurations for incorporating optical amplifiers into transmission systems: (a) as a power amplifier; (b) as in-line amplifiers; (c) as a preamplifier; (d) for compensation of distribution losses in local-area networks.

amplifier application. Figure 3 shows the four basic system configurations envisioned for the incorporation of optical amplifiers. The first configuration is to place the amplifier immediately following the laser transmitter to act as a power amplifier or booster (Figure 3a). The main purpose of such amplifiers is to boost the signal power, which can provide an increase of the transmission distance by 100 km or more. Since the signal input power is typically large (0.1–1.0 mW), the key parameter for the power amplifier will be to maximize the saturation output power and not necessarily the absolute gain.

The second configuration is to place the amplifier in-line and perhaps incorporated at one or more places along the transmission path (Figure 3b), replacing the electronic regenerators. The in-line amplifier corrects for periodic signal attenuation and may exist in a cascade form. The use of in-line optical amplifiers is particularly attractive for multichannel communication systems, since they can amplify all channels simultaneously.

The third configuration consists of using the amplifier immediately before the receiver, so it functions as a preamplifier (Figure 3c). The purpose of such an amplifier is to improve the receiver sensitivity. The main figures of merit are high gain and low amplifier noise, because the entire amplifier output is immediately detected.

In local-area networks (LANs), distribution losses often limit the number of possible nodes. The fourth application of optical amplifiers consists of using them for compensating such distribution losses (Figure 3d).

List of Units and Nomenclature

g_0	peak value of the gain coefficient at peak	m^{-1}
$g(\nu)$	gain coefficient	m^{-1}
G_0	unsaturated amplifier gain	
$G(\nu)$	amplifier gain	
L	amplifier length	m
n_{sp}	spontaneous emission factor	
P	signal power	W
P_{in}	input signal power	W
P_{in}^{s}	input saturation power	W
P_{out}	output signal power	W
P_{out}^{s}	output saturation power	W
P_{sat}	saturation power	W
$S_{sp}(\nu)$	ASE noise spectral density	J
Δf	detector bandwidth	Hz
$\Delta\nu_0$	bandwidth of the gain coefficient	Hz
ν	optical frequency	Hz
ν_0	atomic transition frequency	Hz

See also

Optical Amplifiers: Optical Amplifiers in Long-Haul Transmission Systems. **Optical Communication Systems:** Architectures of Optical Fiber Communication Systems.

Further Reading

Agrawal GP (1992) *Fiber-Optic Communication Systems*. New York: Wiley.

Bjarklev A (1993) *Optical Fiber Amplifiers: Design and System Applications*. Boston, MA: Artech House.

Desurvire E (1994) *Erbium-Doped Fiber Amplifiers: Principles and Applications*. New York: Wiley.

Green PE Jr (1993) *Fiber Optic Networks*. Englewood Cliffs, NJ: Prentice Hall.

Iannone E, Matera F, Mecozzi A and Settembre M (1998) *Nonlinear Optical Communication Networks*. New York: Wiley.

Kazovsky L, Benedetto S and Willner A (1996) *Optical Fiber Communication Systems*. Boston, MA: Artech House.

Shimada S and Ishio H (eds) (1994) *Optical Amplifiers and their Applications*. New York: Wiley.

Erbrium Doped Fiber Amplifiers for Lightwave Systems

P Bollond, JDS Uniphase Corporation, Ewing, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The telecommunications industry has undergone a revolution since the 1980s, by using glass optical fibers for the transmission of information encoded as pulses of light. A single telecommunications-grade optical fiber has been shown to support the propagation of more than 1 Tbit per second (1×10^{12} pulses per second) over distances comparable to typical city separations (>100 km). From this technology, optical fiber links have evolved to become a network on a planet-wide scale, and form the physical backbone of the information age.

Erbium-doped fiber amplifiers (EDFAs) are an enabling technology for optical fiber communication networks. They have several important properties that make them the amplification component of choice in long-distance commercial data transport networks. Erbium ions deposited in silica-based glass allow amplification in the lowest loss region of commercial-grade optical fiber (<0.25 dB/km from approximately 1530–1620 nm). Erbium-doped fiber (EDF) is manufactured in a form that allows for low-loss fusion splicing to standard communications fiber. Compact semiconductor laser diodes are available to excite the erbium ions into an amplification state. The long lifetime of the excited state of erbium provides the ability to simultaneously amplify multiple wavelength channels without significant cross-channel interference. Multiple channel systems have been deployed with more than 100 optical channels (or wavelengths) through each EDFA, and this has allowed network capacities to be dramatically increased. The low noise properties of the EDFA also allow networks to be constructed with many amplified spans before the optical signal has to be electronically regenerated. Practically unlimited transmission distance has been demonstrated using a small number of optical soliton channels through periodically amplified EDFA lightwave systems.

The development history of the EDFA can be traced back to the first optical amplifier. In 1962, a neodymium-based fiber amplifier was invented that operated at 1064 nm. During the 1980s, the need for an optical amplifier at telecommunications wavelengths initiated research at many locations throughout the world. In 1987, the University of Southampton (UK) was first to demonstrate an EDFA that had optical gain at 1550 nm, and during the following years the design of erbium-doped fiber was optimized for this application. In 1989, a practical semiconductor laser diode became available to pump EDF, and the first compact optical fiber amplifier modules soon appeared for commercial deployment. The traditional method of signal regeneration, prior to 1990, was to use electronics to detect the optical signal after each transmission span, recover the digital signal, and then retransmit using another laser diode. The EDFA allows practical wavelength division multiplexing (WDM) of multiple optical signals with all optical signal amplification, and provides significant performance and cost advantages over electronic regeneration.

EDFAs have emerged from the laboratory to be widely deployed in communication networks. EDFAs are used to boost transmitted power of the signal lasers (booster amplifier), amplify signals in transit to compensate losses sustained in the fiber (line amplifier), or amplify signals before a receiver (pre-amplifier). Typically, the amplifier module is specifically manufactured for particular systems that are mounted on electronic circuit boards. These circuit packs are then housed in central offices (local telephone exchanges), remote 'repeater huts', or even in undersea 'bottles' as part of a transoceanic cable. The high cost of network failure requires that the manufactured EDFA modules comply with stringent reliability criteria, to provide a useful operating lifetime greater than 25 years when subject to extreme environmental conditions.

Amplifiers are constructed for incorporation into either existing fiber links as part of an upgrade, or for newly planned systems. Because of the high cost of installing new fiber into the ground and securing property rights, it has become economically desirable to upgrade many existing fiber links rather than to build new systems. Transmission cables usually have many pairs of individual optical fibers, some of which will not initially be transporting data, and these 'unlit' or 'dark fibers' can be activated as consumer demand increases over time. Typically, each fiber of a pair is used to carry either 'east'- or 'west'-bound traffic. Around city areas metropolitan area networks can be expanded in this way, but for long-distance links (long haul networks with distances >1000 km) operation is designed for a larger number of channels (40 to 120 wavelengths) at higher data rates (10 or 40 GHz per channel) over specialized transmission

cables containing low numbers of fiber pairs. Typically communication systems are designed to meet certain cost targets, expressed as dollars per Gbit/s per km, for total transmission distances. The total transmission distance is limited by the optical signal to noise ratio (OSNR) degradation after each fiber-amplifier link, with a smaller permissible degradation at higher bit rates. The required gain and OSNR performance for the system is then translated to an EDFA module specification.

This article discusses EDFA design and applications, and shows the elements involved in producing reliable modules for commercial lightwave systems.

Components for EDFAs

EDFAs are comprised of passive optical components, erbium-doped fiber, and pump lasers. Passive components are chosen to meet optical and environmental specifications, while the erbium-doped fiber is selected based on the optical power, gain, and noise figure requirements. Pump lasers are a key influence on the price and performance of optical amplifiers.

The amplifier module is typically connected to the transmission fiber using fiber connectors. These polished fiber connectors have a higher insertion loss and reflectivity than fusion splicing, but allow for easy deployment in the network. Internal optical components are fusion spliced together to provide low loss, low back reflection, high strength, and high reliability joins. Fusion splicing is tailored to particular fiber types, so that optical components with dissimilar fiber types are joined with the lowest loss.

The signal and pump radiation is combined with low loss using optical components called wavelength division multiplexers (see Figure 1). These components are based on fused fiber or interference filter based technology. Fused fiber WDMs offer the lowest insertion loss (<0.1 dB is commercially available) but is restricted to widely spaced wavelengths (e.g., 980 nm pump and 1550 nm signal). Interference filter based WDMs are available for closely spaced wavelengths (e.g., 1480 nm pump and 1530 nm signal) and have a very low wavelength dependent insertion loss (flatness). Interference filters can be designed to produce sharp low-pass, high-pass, or bandpass type filters suitable for combining closely spaced wavelengths, as well as broader peaks suitable for lowering the gain in particular signal wavelength regions to produce gain flattened amplifiers.

Signal reflections can cause an amplifier to act as a laser, and this detrimental effect is eliminated in EDFAs by using optical isolators. In an isolator the signal light is coupled out of the single-mode fiber through a graded index (GRIN) lens and passes through a nonlinear crystal before being coupled back into the optical fiber. The isolator consists of a birefringent rutile (TiO_2) or Yttrium OrthoVanadate (YVO_4) wedge, followed by a Yttrium Iron Garnet (YIG) Faraday rotator, followed by another birefringent wedge. The YIG crystal is surrounded by a permanent magnet that rotates the light's polarization by 45 degrees. The 45-degree polarization rotation, coupled with the two birefringent wedges, ensures that light is efficiently coupled to the output fiber but not in the reverse direction. Commercial isolators are available with low insertion losses across the signal band, with some samples below 0.35 dB. Note that the YIG crystal works well for the 1480 nm pump and 1530–1620 nm signal bands, but currently there is no suitable isolator material that covers 980 to 1550 nm and this puts some limitations on certain EDFA designs.

The isolator design has been extended to make multiport circulators. A three-port circulator has the input into port 1 and output of port 2, light entering port 2 is directed to port 3, and light entering port 3 is blocked with an isolator. The circulator allows for adding and dropping of individual channels when a

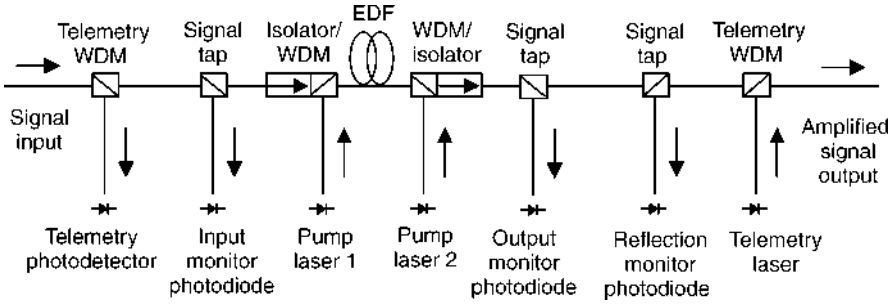

Figure 1 Single-stage EDFA with features.

narrow bandwidth reflective filter is placed on port 2. Circulators also can be used to separate co-propagating and counter-propagating traffic on a single transmission fiber before amplification is done.

Erbium-doped fiber has two strong absorption bands around 980 nm and 1480 nm that are suitable for achieving a population inversion in the erbium ion. The 980 nm wavelength allows low noise amplification, while 1480 nm lasers provide higher EDFA output. 980 nm pump lasers usually incorporate fiber Bragg gratings to stabilize the laser wavelength to match the narrow EDF absorption peak, and also have lower drive current requirements. A 1480 nm pump can provide more amplification, since there are more photons in each mW of laser power at 1480 nm than 980 nm. High-output power pump lasers incorporate thermo-electric coolers (TECs) within the pump laser package and can provide fiber coupled power greater than 500 mW. Both high-power lasers have been qualified to meet the most stringent reliability standards, with typical mean time before failure beyond 25 years. Low-cost 980 nm pump lasers, that do not have TECs, are also available in smaller packages and can supply up to ~200 mW. By using wavelength division or polarization combining techniques, it is possible to further increase the available pump power in the erbium fiber.

The EDFA module is assembled into a package that may contain passive optical components, several meters of coiled EDF, pump lasers, photo-detectors, and electronic circuit boards. In some cases, it is advantageous to thermally stabilize the components so that the amplifier performance can be maintained when exposed to extreme environmental conditions. This may occur when the central office's environmental control is compromised (e.g., air conditioning failure). In particular, EDF can exhibit undesirable spectral gain changes if the ambient temperature changes by more than 5 °C.

Pump lasers with internal TECs can dissipate more than 5 Watts of heat per laser and since cooling fans usually do not have the required reliability, passive cooling is commonly used in the module in the form of a metal heat sink. The amplifier module's size can be compact, limited by the height of optical components or pump laser diodes, or by the size of a built-in heat sink. The module is designed to survive conditions of electrostatic discharge, humidity, temperature, thermal shock, extreme vibration, and other stresses that may be inadvertently present during operation in the field. In addition, all material in the EDFA module is also qualified against problems with out-gassing (e.g., hydrogen release), combustion and chemical or biological exposure.

The Single-Stage Amplifier

A simple single-stage EDFA consists of an erbium-doped fiber spool with signal and pump combining multiplexer. The fiber is optically pumped by 980 nm and/or 1480 nm laser(s), whose light is coupled into the signal fiber by a passive component called an optical multiplexer. The pump wavelengths are readily absorbed by erbium ions embedded in silica raising them to an excited state. Amplification occurs when, stimulated by a nearby signal photon, an excited erbium ion relaxes back to the ground state producing a second, identical signal photon. The erbium ion can be approximated as a three-level atomic system that can be completely inverted by a 980 nm photon, to provide the lowest noise amplification. In contrast, the 1480 nm pump will excite the erbium ion directly into the upper laser level as a two-level system, and because of rapid spontaneous emission from this level, the maximum inversion in this case cannot exceed approximately 75%. Note that as the pump photons are absorbed, the inversion will be nonuniform along the EDF length.

There are two signal wavelength regions commonly amplified by EDFAs, the C-band (conventional band) from approximately 1528 to 1565 nm, and the L-band (long band) from approximately 1570 to 1620 nm. Amplification in the C-band readily occurs when moderate pump power is available, and relies on the erbium ion's spectral absorption and emission wavelength window that is suited to high levels of atomic inversion. L-band amplification is also achieved with moderate pump powers, but because of the lower absorption and emission cross-sections, similar gain is reached using approximately five times more EDF with a lower average inversion. The C-band amplifier is typically less costly because less EDF is used, while high-concentration erbium fibers are available specifically for L-band EDFAs.

An EDFA's most critical performance parameters are its amplified signal output power (typically stated in dBm) and its noise figure (stated in dB). Output power is mainly determined by total pump power and the amplifier internal loss. The noise figure (NF) is defined as the ratio of the signal-to-noise ratio at the input to the signal-to-noise ratio at the output.

A single-stage amplifier typically has 1 or 2 pump lasers but can have more if polarization- or wavelength-pump-combining is implemented for higher power. When the pump radiation propagates in the same direction as the signal, the amplifier is co-pumped, while counter-pumping denotes the case when the pump laser propagates against the signal. For a single pump, a co-pumping 980 nm

laser minimizes the NF (suitable for a pre-amplifier) while counter-pumped 1480 nm architecture optimizes output power at some expense to the NF (suitable for a booster amplifier). Recent semiconductor pump laser improvements have enabled 980 nm pump lasers with output powers >500 mW to be commercially available, allowing most single- or dual-stage EDFAs to be energized by one laser.

Single-stage designs can be enhanced, as shown in Figure 1. To control the network an optical signal may be used as a telemetry or network supervisory channel, and this is removed with a filter. Telemetry wavelengths are usually outside of the useful EDF amplification window, and commonly range from 1500–1520 nm and from 1620–1640 nm. An isolator may be used at the input and/or output to prevent pump laser or amplified spontaneous emission from the erbium-doped fiber 'leaking' into the transmission path. Optical taps may be included to provide information about signal spectra at the input and output sides. Their feedback can be used to control pump laser biases for tuning output power, monitoring amplifier performance, or simply to trigger alarms. In addition, a reflection monitor is sometimes placed at the output to observe backwards propagating optical signals arising from reflections. Electronics in the module will continuously monitor the pump lasers and photodetectors and, for example, can place the module in an 'eye-safe' low-output power (<10 mW) mode within milliseconds if a transmission fiber break is detected through increased back-reflected optical power.

The amplifier has a signal input power of −30 to −10 dBm (1–100 μW) for each optical channel, and has a gain of 25 dB to compensate for a typical span loss of 100 km optical fiber link. This signal level allows high powers at the receiver photodetector for high-quality signal detection, yet is low enough to avoid nonlinear propagation effects in the transmission fiber. The typical total output signal power of an 80 channel (wavelength) C-band EDFA is less than 200 mW (+23 dBm) to avoid stimulated Raman scattering (SRS) in standard transmission fiber. The use of improved low nonlinearity transmission fibers and longer transmission distances can lead to specified total EDFA output powers to be greater than 400 mW (+26 dBm). To provide the best performance, the amplifier usually will allow only a single direction of propagation, with bi-directional communications systems using one transmission fiber and circulators to separate 'east' and 'west' traffic into individual EDFAs.

Figure 2 shows the results of a numerical simulation for a single-stage amplifier as a function of the input power and EDF length. The amplifier was assumed to have EDF with peak absorption of 5.5 dB/m near 1530 nm, and was pumped with 100 mW at 980 nm. The amplifiers signal loss before and after the EDF was taken to be 0.8 and 1.2 dB, respectively. As amplifier input power increases there is a decrease in gain provided by the medium since there is a fixed amount of pump power available. The maximum gain is usually achieved near 1530 nm where the difference between the EDF's emission and absorption cross-sections is greatest. Selecting the length of EDF is critical to achieving the desired performance, and this can be examined using numerical simulation for a wide range of design options.

Undersea systems, with their long distances and large costs of network failure, place stringent design requirements on EDFAs. These systems mostly use single-stage designs with emphasis on low noise operation. This can be achieved by eliminating most of the optical components prior to the EDF, and also by co-propagating a strong 980 nm pump using a low loss fused fiber WDM. Undersea repeaters are spaced by 30 to 80 km, shorter than terrestrial networks, with each channel operated at higher power to achieve multi-thousand kilometer distances. For example, the trans-Pacific TPC-5J cable spans 8600 km from Coos Bay, Oregon (USA) to Ninomiya (Japan), using EDFAs spaced every 33 km. The tight electrical power budgets available to each repeater (powered from land) necessitate using pump lasers without TECs. During installation, the EDFA will experience large mechanical shock, as the cable and metal repeater bottles are unwound aboard ship and dropped into the ocean. Other design considerations are done for operation at the constant ambient temperature of the ocean bottom (~+2 to 4 °C) or for seasonal temperature changes on the continental shelves, where the optics and EDF will operate at 15 to 30 degrees above ambient due to heating from the electronics.

Single-stage optical amplifiers are suited for a wide range of applications such as single-channel amplifiers, simple WDM amplifiers, and low-cost amplifiers. Using high-power pump lasers or combined pump laser schemes allows such amplifiers to deliver output powers >20 dBm. However, single-stage amplifiers cannot meet the requirements of all telecommunications architectures, leading to increasing demand for multiple-stage EDFAs that are discussed below.

Multiple-Stage EDFAs

The most common implementation of a multiple-stage EDFA is to improve the noise and output power

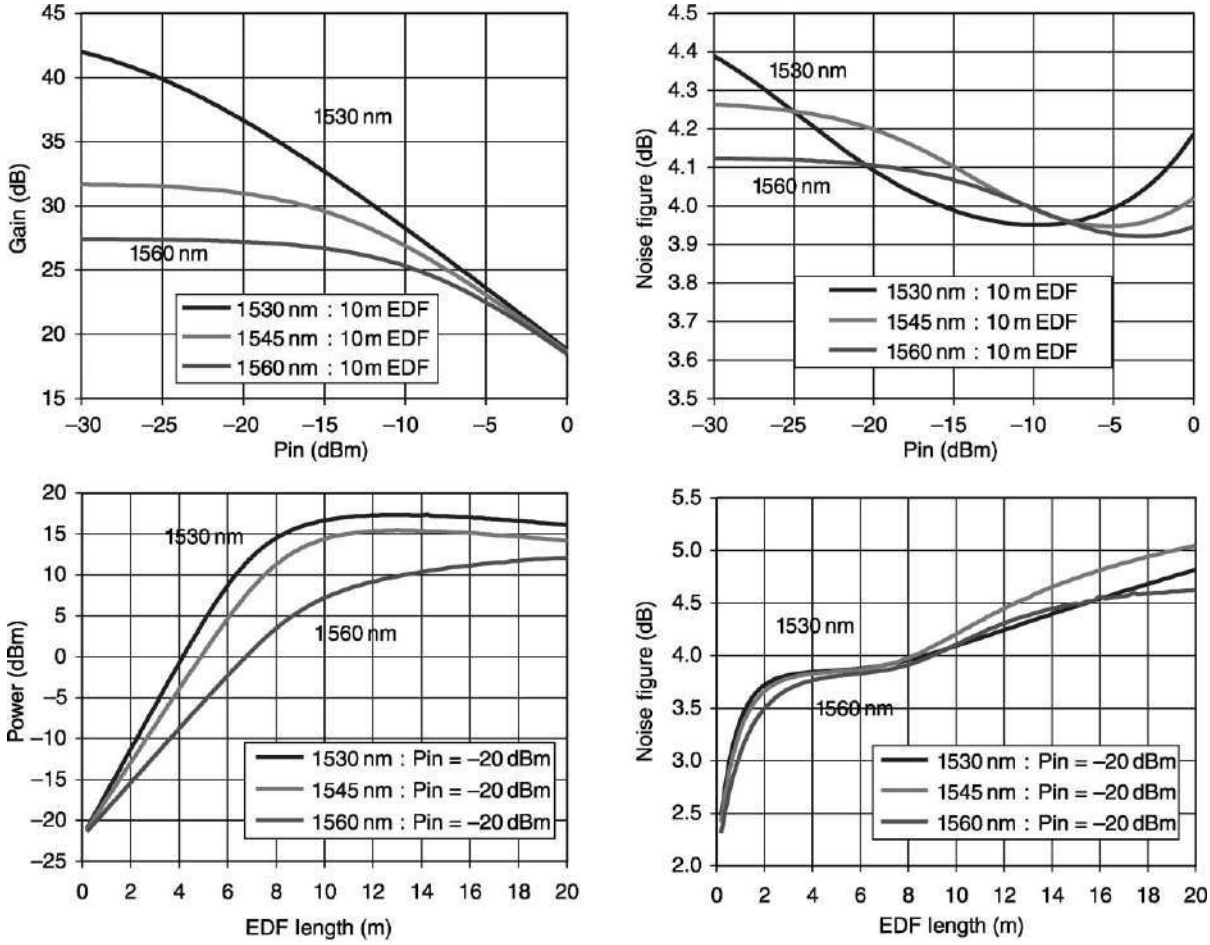

Figure 2 Typical single-channel performance of a single-stage EDFA.

characteristics by imbedding an optical isolator between two sections of EDF. The isolator blocks backward-traveling amplified spontaneous emission to improve the efficiency of the amplifier. This is shown in Figure 3 for two common single-pump designs. The first is the 'pump by-pass', design where the residual pump radiation from the first stage is redirected around the mid-stage isolator. This is often used when pumping with 980 nm since signal band optical isolators will not transmit 980 nm radiation. This design can introduce problems when using low-quality components as small signal levels can propagate around the pump by-pass fiber to create multiple path interference (MPI) effects. Although MPI effects can be eliminated by using a pump splitting coupler to directly pump each EDF section, the pump by-pass design makes the most efficient use of available pump power. The second design is called the 'pump through' design, and is used when pumping at 1480 nm since both the pump and signal band (e.g., 1550 nm) photons will transmit through commercial isolators with only small loss. Note that for cost and space constraints within the module, it is sometimes advantageous to use hybrid optical components, e.g., an isolator and WDM can be combined into one compact package.

An EDFA with multiple input signals will have a very nonuniform output gain profile. This is a consequence of the erbium ion's wavelength-dependent emission and absorption cross-sections in the host glass material. Figure 4 shows gain spectra for both C-band and L-band EDFAs with and without gain-flattening filters (GFFs). The gain spectrum is dependent on the EDF's chemical composition and EDFA design features. For an amplifier with 25 dB gain, a GFF with peak loss less than 10 dB is usually needed to correct for these gain deviations. Although many technologies are available for GFFs (e.g., thin-filter interference filters, Bragg gratings, tapered fiber filters, etc.) the basis function of these technologies is usually not identical to the EDF's gain profile, and this mismatch results in gain

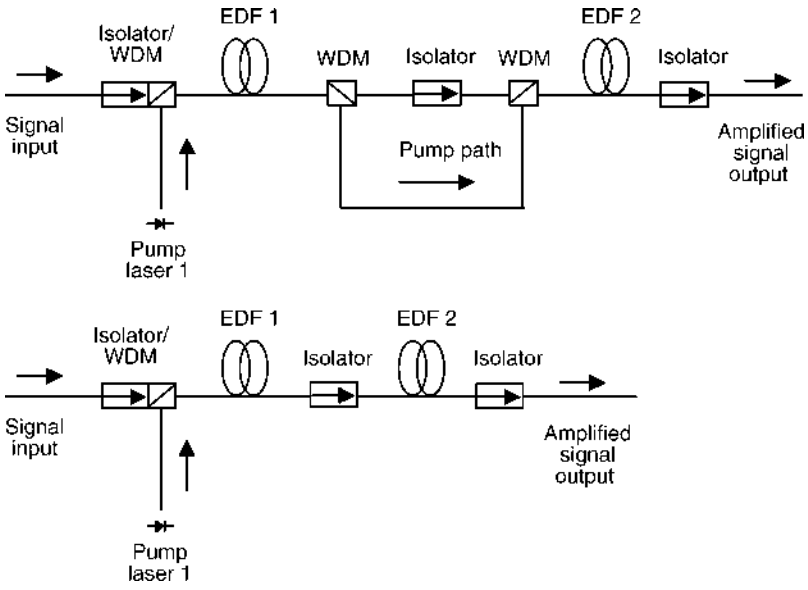

Figure 3 Imbedded Isolator EDFA designs.

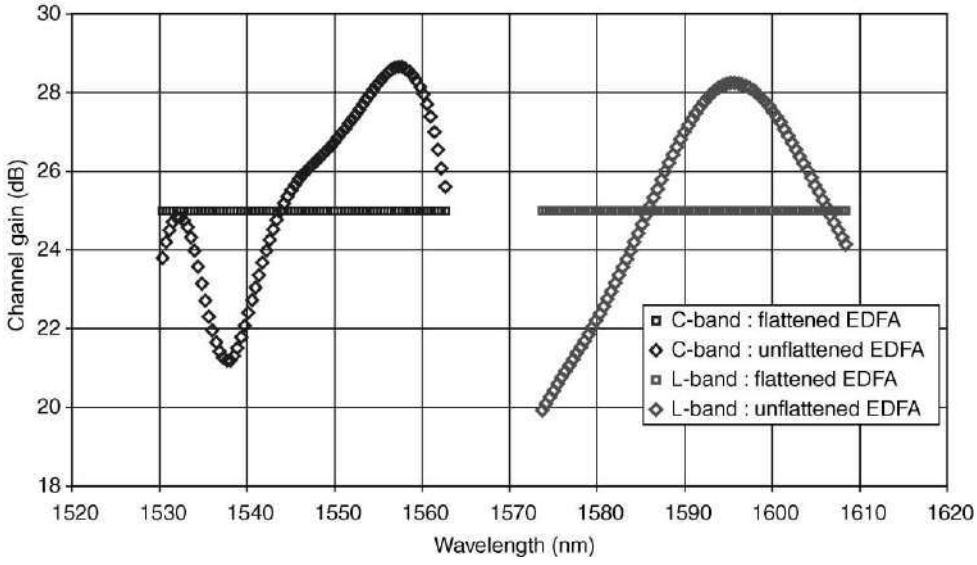

Figure 4 Gain spectrum of a multi-staged optical amplifier with and without gain flattening filters.

flatness error. However, when using these technologies, manufactured EDFAs can have a gain error (flatness) less than 1.0 dB over bandwidths greater than 35 nm.

Early EDFAs for WDM systems used 16 optical channels spaced by 200 GHz in the 1540 to 1560 nm part of the C-band. In this case, the EDF gain profile is relatively smooth and no GFFs are needed to achieve a 1.0 dB gain flatness specification. It is important to note that for a particular level of signal and pump power, a longer length of EDF in the EDFA will produce more gain at the long wavelength end of the spectrum, hence by shortening the EDF length in Figure 4 the result will be relatively flat gain from approximately 1540 to 1560 nm. As EDFA technology has matured, the EDF can be the bandwidth-limiting component in a system, and it has become necessary to use GFFs to realize useable bandwidths of up to 55 nm when using regular silica-based EDF. The most economic EDFAs operate in the C-band where relative to the L-band, pump laser efficiency is highest, EDF lengths are shortest,

and the chromatic dispersion of most installed transmission fiber limits nonlinear optical effects.

In single-stage amplifiers, the gain-flattening filter is at the amplifier output, and this results in lost output power. Wideband gain-flattened amplifiers usually have a multiple EDF stage design, with the GFF component between two EDF stages. The EDF sections that follow the attenuating GFF provide additional amplification, to give high power out of the EDFA module that has large internal losses.

In addition to GFF components inside the EDFA, an attenuator can be used to compensate for input signal power changes, which will produce a spectral gain rotation or tilt. The attenuator reduces the power on all optical channels equally and allows the amplifier to operate in a 'fixed gain' mode. In WDM systems, optical amplifiers may also need mid-stage access where the signal is fed into an external device between the amplifier stages. Reasons for doing this include monitoring, adding or dropping of individual channels, and dispersion compensation. Since additional losses up to 10 dB are introduced in the amplification path, the amplifier design has to be optimized for those losses.

Typical two-stage gain-flattened optical amplifier architecture is shown in Figure 5. Input and output couplers and telemetry WDMs can be included if required. A single pump can be split and shared between stages to save cost. When multiple pumps are needed, the most common configuration is a 980 nm pump laser co-pumping the first stage (EDF1) for low noise and a 1480 nm laser counter-pumping the second spool (EDF2) for high gain. A significant loss element, e.g., a gain-flattening filter, add/drop module or dispersion compensation module, is situated between the stages. Note that mid-stage access can be located before or after the gain flattening or between additional EDF stages. The gain spectrum of multistage gain-flattened EDFAs is shown in Figure 6, showing the gain equalization possible using a multistage amplifier design and gain-flattening filters. Using deeper GFFs can increase the usable bandwidth of the amplifiers, but this requires higher-power pump lasers to maintain the same gain and optical signal-to-noise ratio (OSNR) performance.

The two-stage EDFA shown in Figure 5 highlights a common problem for amplifier design. Given high-grade optical components and pump lasers, what

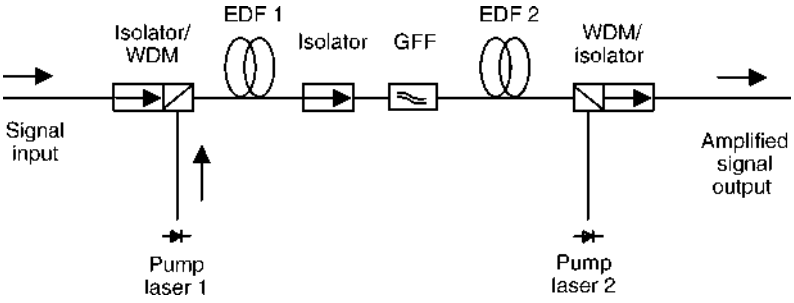

Figure 5 Typical multistage gain-flattened optical amplifier architecture.

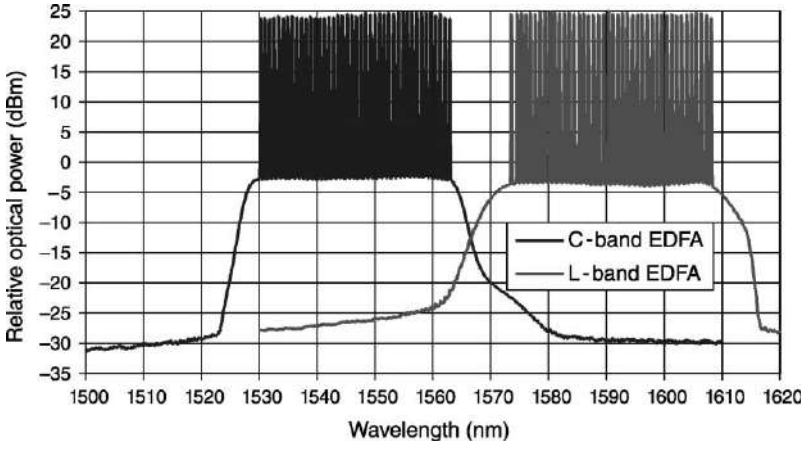

Figure 6 Measured gain and noise spectra of multistaged gain-flattened C-band and L-band EDFAs.

length of erbium-doped fiber should be used to give the best amplifier performance? This question is best answered by using the results of extensive numerical simulations of the optical amplifier.

Figure 7 shows the results of numerical simulation for a two-stage amplifier as a function of the first and second EDF stage lengths. The amplifier was assumed to have EDF with a peak absorption of 5.5 dB/m near 1530 nm, and was pumped with 130 mW at 980 nm in the first stage and 160 mW at 1480 nm in the second stage. The total input power was assumed to be −2.5 dBm for 80 channels distributed from 1530 nm to 1563 nm. Note that each of the 80 channels is separated by 100 GHz, and will generally support a long-distance communications system with each wavelength modulated at 10 GHz.

Figure 7a shows that the optical output power is largest when both EDF stages are approximately 20 meters. For this design, EDF lengths significantly longer than 20 meters will produce less output power, since the pump radiation will have been completely absorbed in the first few meters of EDF. The spectral gain flatness is shown in Figure 7b. For this design, approximately 17 meters of EDF was needed to produce a flat gain spectrum, i.e., each optical channel had the same gain. Other combinations of EDF lengths were not matched to the particular gain-flattening filter and produced large gain variations across the band. In general as the amplifier's total EDF length increases the gain spectrum exhibits a positive tilt (e.g., longer wavelength channels will experience more gain). Figures 7c, d show the average channel and maximum channel NF, respectively. From a system design perspective, a link can be limited by the optical channel with the lowest optical signal-to-noise ratio (OSNR) or highest noise figure, and this is a critical parameter of interest when designing optical amplifiers. In practice the best amplifier design is a compromise between high-gain, low-gain flatness, and low NF. From extensive numerical simulations, for the two-stage gain flattened amplifier example it was for

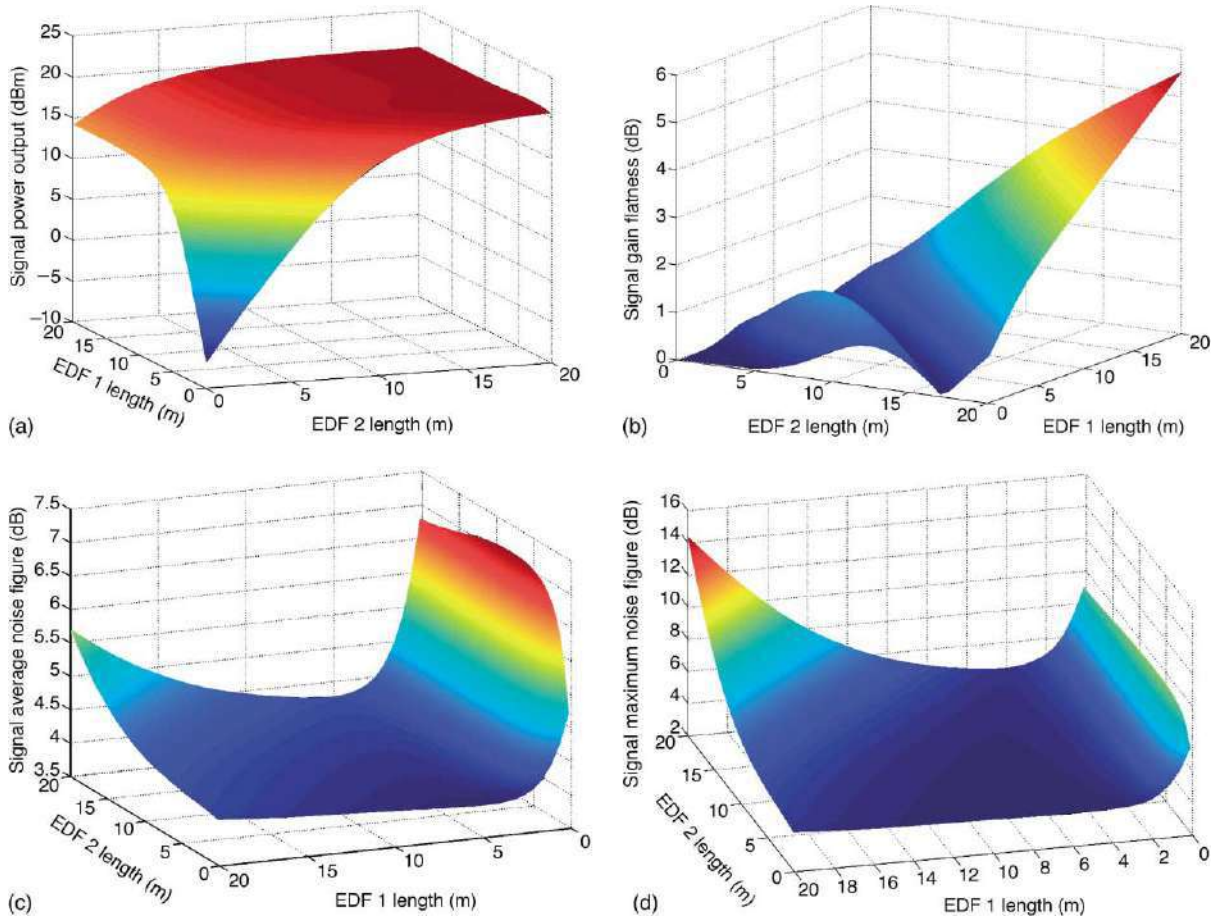

Figure 7 (a) Total output power, (b) spectral flatness, (c) average noise figure, and (d) maximum channel noise figure of a two-stage optical amplifier with a gain-flattening filter.

approximately 4 meters in EDF stage 1, and 13 meters in stage 2.

Successful commercial EDFA products are more than a single amplifier design that can accommodate all applications. In some cases an economical solution is a modular design approach, where smaller capacity, lower-cost amplifiers are initially installed in a network. Additional gain can be added to the basic amplifier for longer-distance spans or as network traffic increases over time. This 'pay as you grow' approach can be done using additional gain stages, or pump lasers, and can be upgraded without interruption of revenue-generating network traffic.

Low-Noise Design of EDFAs

A key design feature of commercial EDFAs is low-noise operation since the degradation of the OSNR limits the reach of the system. The NF of an EDFA can be defined as

$$\mathrm{NF} = \mathrm{SNR_{in}}/\mathrm{SNR_{out}} \qquad [1]$$

The NF can also be calculated from

$$\mathrm{NF} = \frac{1}{G} \times \left[\frac{P_{\mathrm{ASE}}(\nu_s, L)}{h\nu_s} + 1 \right] \qquad [2]$$

where P_{ASE} is the amplified spontaneous emission (ASE) noise power at the signal frequency ν_s for an EDFA of gain G (h = Plancks constant). An alternative equation for the NF for a single section of EDF is

$$\mathrm{NF} = \int_0^L \frac{\gamma_e(\nu_s, z) - \gamma_a(\nu_s, z)}{P_s(z)} \times P_s(0)\mathrm{d}z + 1 \qquad [3]$$

where γ_e and γ_a are the EDF emission and absorption factors at the signal frequency ν_s, at distance z along the amplifying fiber. Equation [3] shows that low NF can result from rapid signal gain in the initial fiber along the EDF. This result indicates that co-propagating pump and signal photons, both into the same end of the EDF, combined with the maximum inversion obtained by using a 980 nm pump, will result in low-noise amplification. The NF is typically expressed in dB units, and using a full quantum mechanical theory, the lowest possible NF for a fully inverted EDF with a large signal gain can be shown to be approximately 3 dB. Actual EDFAs will only approach the '3 dB quantum noise limit' when the signal is significantly lower than the available pump power (e.g., low channel count applications) so that high inversion can be maintained along the EDF length. Optical components in the EDFA prior to the first EDF will attenuate the signal but not the noise, since ASE is only generated in the EDF, and this input stage loss will directly add to the EDFA's NF.

The use of 1480 nm pump lasers can also result in a further NF penalty of 0.3 to 1.5 dB since complete inversion cannot be achieved when the erbium ions are operating as a two-level atomic system and more ASE is generated. The ASE, like the signal gain, has a wavelength-dependent spectral profile and this results in each channel having a different NF. Typically, the optical channel with the largest NF (lowest OSNR) is used to define the EDFA's noise performance. EDFAs with low-input signal power (e.g., −15 dBm) and high gain (e.g., 25 dB) may be described as low noise when the NF is <4 dB. For large total input signal power (e.g., +2.5 dBm for high channel count operation) low noise may be NF < 5 dB.

For an amplifier with multiple EDF sections, the noise figure can be calculated from

$$\mathrm{NF_{System}} = \frac{\mathrm{NF}_1}{L_1} + \frac{\mathrm{NF}_2}{L_1 G_1 L_2} + \ldots + \frac{\mathrm{NF}_N}{L_1 G_1 L_2 G_2 \ldots L_{N-1} G_{N-1}} \qquad [4]$$

where L_i is the signal loss prior to the gain G_i of the EDF stage i. From eqn [4], multistage amplifiers with large inter-stage losses will have numerically larger NFs, but the total module NF is dominated by the first EDF stage. Undersea EDFAs, with a single short EDF section, very small L_i and 980 nm pumping, can have NF <3.5 dB.

Advanced EDFA Functions

An important trend in optical amplifiers, and systems in general, is the move towards dynamic control. Several technologies are being considered and the primary idea is to dynamically control the intensity level of each optical channel or wavelength to correct for any gain/loss inequalities in the network. This is particularly important as channel counts and bit rates rise concurrently with the functionality demanded of optical communication networks. A common requirement in this category is fast amplifier response so that in a network where individual optical channels can be added or dropped there are no power distortions at other surviving wavelengths. In a typical drop event, in less than 1 microsecond, 31 of 32 wavelengths can be completely removed from the input to the EDFA by an add/drop module located elsewhere in the network. This event will cause a total input power decrease of 15 dB and without rapid pump power adjustment the remaining channel will be excessively amplified.

Erbium ions have an excited state lifetime of approximately 10 ms, and this sets the time-scale for the transient signal response. A network transient usually manifests as a sharp optical amplifier (OA) gain fluctuation lasting up to several tens of microseconds that will pulse large signal levels through the entire optical network. Dynamic network loading necessitates controlling temporal gain transients, and this may be achieved by using a microprocessor in the amplifier module. These intelligent EDFA modules can rapidly adjust the pump laser power based on feedback from internal photo detectors to control gain transients.

The need for dynamic control also exists in the frequency domain. Wideband amplifiers have a gain variation among the optical channels of typically up to 1 dB over all operating conditions, and since these EDFAs are concatenated over many spans the electronic receiver circuit at the end of the network will have to detect optical channels with widely varying power levels. This can limit the number of spans that the network can bridge before electronic regeneration becomes necessary. This problem can be corrected using a dynamic gain equalizing filter (DGEF) to actively attenuate individual optical channels levels. A DGEF can be embedded inside an amplifier to create a loss-less component, and can be used to equalize channel powers for an entire network.

The problems of accumulation of gain ripple become more apparent in systems that employ amplifiers based on stimulated Raman scattering. A conventional EDFA amplified system may have 10 spans transmitting 100 optical channels at 10 Gbit/s each over a distance of approximately 800 km before electronic regeneration becomes necessary. To span longer distances with higher capacity, Raman amplifiers are used to amplify the transmission fiber along with EDFAs. This system architecture has the advantage of a higher effective OSNR for the individual spans, and has been demonstrated for distances over 4000 km with 40 Gbit/s channels. However, the variability of the Raman amplifier gain, due to network loading and differing transmission fiber properties, necessitates the use of loss-less DGEFs.

High-capacity transmission uses individual channels that are each modulated up to typical 2.5, 10, or 40 GHz, that requires that the optical amplifiers have low polarization mode dispersion (PMD) and low chromatic dispersion (CD). Polarization mode dispersion distorts the temporal spread of an ultra-short pulse as different polarization states travel at different speeds through the transmission fiber and amplifier components. PMD is particularly a problem in the first generation of installed transmission fiber and through polarization components (e.g., isolators) that are not PMD compensated. Without PMD compensation, 40 Gbit/s transmission, for example, can be limited to very short distances (several km). Fortunately, PMD compensators have been developed for upgrading aged installed networks, and also amplifiers are available with very low PMD.

To achieve long-distance transmission, a chromatic dispersion map is produced for all the optical spans, and by using a dispersion compensating module in each amplifier, the net dispersion across the channel wavelength window is controlled. Although optical amplifiers have approximately 1000 times less optical fiber in them compared to the transmission span that it is amplifying, chromatic dispersion in the amplifier can also be of concern for very long-haul networks. The CD of the amplifier is primarily in the EDF, and is specific to the EDF's geometric and chemical composition. Furthermore, since the erbium ion is a resonant atomic system, there may be a pump and signal power contribution to the dispersion (resonant dispersion) that could degrade network performance. Characterization of these optical effects is typically done on manufactured 'field grade' amplifiers as part of a quality control process.

Conclusion

As optical networks evolve EDFAs will continue to be an enabling technology for higher capacity and more dynamic communications networks. EDFA technology has already advanced to accommodate multiple channels, to span several wavelength windows and to provide features such as dispersion compensation, gain-transient suppression, and dynamic gain flatness. As future network architectures are introduced, to incorporate Raman amplification and additional wavelength windows, the demands placed upon EDFA design will continue to expand.

See also

Optical Communication Systems: Architectures of Optical Fiber Communication Systems; Wavelength Division Multiplexing. **Scattering:** Raman Scattering.

Further Reading

Agrawal G (1989) *Nonlinear Fiber Optics*. San Diego, CA: Academic Press.

Agrawal G (1997) *Fiber Optic Communication Systems*. New York: John Wiley and Sons.

Becker PC, Olsson NA and Simpson JR (1999) *Erbium-doped Fiber Amplifiers – Fundamentals and Technology*. San Diego, CA: Academic Press.
Desurvire E (1995) *Erbium Doped Fiber Amplifiers*. New York: John Wiley and Sons.
Pedersen B, Bjarklev A, Povlsen JH, *et al.* (1991) The design of erbium-doped fiber amplifiers. *Journal of Lightwave Technology* 9(9): 1105.
Sudo S (1997) *Optical Fiber Amplifiers: Materials, Devices and Applications*. Boston, MA: Artech House.

Optical Amplifiers in Long-Haul Transmission Systems

B M Desthieux, Xtera Communications, Inc., Allen, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical amplifiers have been a key element for enabling the development of optical transmission systems, which have reached total capacity in the order of a few terabits per second. The first type of optical amplifier to see application in the telecom industry has been the erbium-doped fiber amplifier. Following its invention in the late 1980s, the mid-1990s saw its industrial application in submarine and terrestrial transmission systems.

Since then, optical transmission systems have been the predominant source for transmitting information over distances ranging from a few 100 km to more than 10 000 km, connecting cities, countries, and continents.

This article will describe the applications of optical-fiber amplifiers in long-haul transmission systems, focusing on erbium-doped fiber amplifiers and Raman amplifiers, the most popular type of optical amplifiers used in modern transmission optical systems.

Fundamentals of Long-Haul Transmission Systems

This section explains the key elements and parameters of an optical transmission system.

Key Elements

An optical transmission system consists of a transmitter, a receiver, one or more optical amplifiers, and one or more spans of transmission fiber. Figure 1 shows a simplified schematic of a bidirectional transmission system using a fiber pair.

Optical amplifiers are used to compensate for the loss of the transmission fiber and the other optical elements placed along the signal path. Boosters and pre-amplifiers refer to optical amplifiers which are located at, respectively, the input end and output end of the system. In-line amplifiers refer to optical amplifiers located between the transmission fiber spans.

The transmitter includes laser diodes to generate the signal power and optical modulators to transfer the data stream from the electrical input signals to optical signals. The optical signal is transmitted to the receiver after traveling through the transmission fiber. At the receiver are located the photodiodes, which transfer back the information into electrical output signals. In order to increase the total bandwidth of the transmission system, time-division multiplexing (TDM) and wavelength-division multiplexing (WDM) can be used. TDM techniques combine the electrical domain data stream of several electrical channels, which is carried by a single optical channel. An optical channel refers to an optical signal located at a particular wavelength and modulated with the transmission data. TDM increases, therefore, the bit rate supported by the optical channel. Typical bit rates used in modern optical transmission systems are 2.5 Gbit/s and 10 Gbit/s. With WDM, several optical channels are combined, in the optical domain, into a single optical beam, which is launched into the transmission fiber. For this technique, optical multiplexers (Mux) and de-multiplexers (Demux) are required at, respectively, the transmitter and the receiver.

Long-haul transmission systems typically operate at wavelengths around 1.5 μm, where the attenuation of fibers is the lowest. We label C-band the region located between 1530 nm and 1565 nm. Its lower bound is labeled S-band; the upper bound the L-band.

Apart from attenuation, several other effects existing in fiber can be responsible for degrading the quality of the transmission. Among those is chromatic dispersion, which distorts the optical signal as it travels along the transmission fiber. Several types of fiber have been deployed with different chromatic dispersion characteristics. The most common one is the standard single-mode fiber (SSMF), which has a zero-dispersion wavelength – wavelength for which no chromatic dispersion is observed – near 1.3 μm. The dispersion-shifted fiber (DSF) has a zero-dispersion wavelength near 1.55 μm, and the nonzero

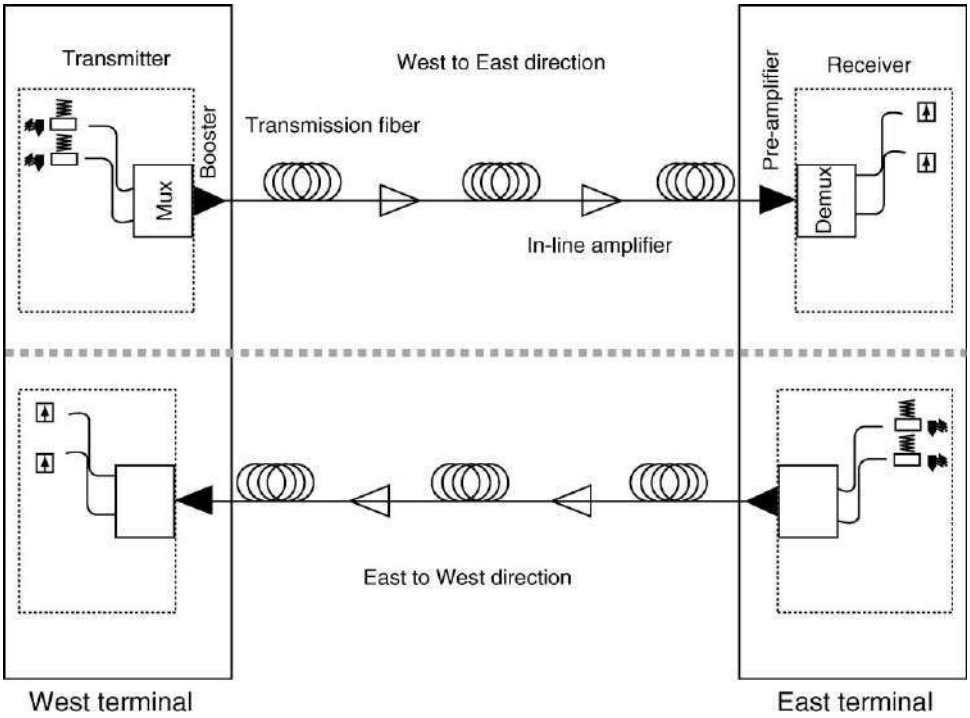

Figure 1 Schematic of a bidirectional long-haul transmission system.

dispersion shifted fiber (NZDSF) has a zero-dispersion wavelength close to 1.5 μm. To remedy this effect, dispersion compensating modules (DCM) are placed along the link. These modules – generally a special type of fiber – have chromatic dispersion with inverted characteristics with respect to those of the transmission fiber.

Key Optical-Amplifier Parameters

Optical amplifiers spontaneously generate optical noise – referred to as amplified spontaneous emission (ASE) – that is added to the modulated signal and consequently degrades the quality of the transmission. It is essential to quantify the level of ASE for a given optical transmission system.

We define the optical signal-to-noise ratio (OSNR) in the bandwidth $\Delta\nu$ as

$$\mathrm{OSNR} = \frac{P_{\mathrm{Signal}}}{P_{\mathrm{ASE}}} \qquad [1]$$

where P_{signal} is the signal power and P_{ASE} is the ASE power in the bandwidth $\Delta\nu$.

The required OSNR at a receiver, to detect the information with high quality, depends on the modulation format and on the other sources of noise or signal degradation existing in the transmission system. If we assume a nonreturn to zero (NRZ) modulation format, and signal detection limited only by ASE noise, an OSNR, in 0.1 nm bandwidth, of about 18 dB, is required at the receiver for a bit-error rate – number of erroneously transmitted bits divided by number of transmitted bits – of 10^{-12}. Introduction of other transmission degradation sources requires a higher OSNR to maintain the same bit-error rate.

The decibel unit (dB) for a given ratio, R, is defined by

$$R_{\mathrm{dB}} = 10\log_{10} R_{\mathrm{linear}} \qquad [2]$$

Similarly we defined the power decibel unit (dBm) as

$$P_{\mathrm{dBm}} = 10\log_{10} P_{\mathrm{mW}} \qquad [3]$$

where P_{mW} is the power in mW.

In order to quantify the amount of ASE generated by an optical amplifier, we define the equivalent input noise factor (n_{eq}), in a linear unit, by

$$n_{\mathrm{eq}} = \frac{P_{\mathrm{ASE}}}{h\nu\Delta\nu G} \qquad [4]$$

where P_{ASE} is the output ASE power in one polarization state in the bandwidth $\Delta\nu$, h the Planck's constant, ν the frequency of the ASE, and G the gain of the amplifier.

We also define the noise figure of the amplifier (NF) – a universal parameter to characterize

amplifier noise – as

$$\mathrm{NF} = 2n_{\mathrm{eq}} + \frac{1}{G} \qquad [5]$$

If an optical component, having a loss T_{in} ($T_{\mathrm{in}} < 1$), is inserted at the input of the amplifier, the total NF of the amplifier becomes $\mathrm{NF}_{\mathrm{total}} = \mathrm{NF}/T_{\mathrm{in}}$. If an optical component, having a loss T_{out} ($T_{\mathrm{out}} < 1$), is inserted at the output of the amplifier, $\mathrm{NF}_{\mathrm{total}} = \mathrm{NF} + 1/G(1/T_{\mathrm{out}} - 1)$. When $G \gg 1$, we have $\mathrm{NF}_{\mathrm{total}} \approx \mathrm{NF}$.

The OSNR in 0.1 nm bandwidth obtained after N fiber spans of loss, $\mathrm{Loss}_{\mathrm{Fiber}}$ (in dB), each having an amplifier, is in decibels:

$$\mathrm{OSNR} = P_{\mathrm{Signal}} - \mathrm{Loss}_{\mathrm{Fiber}} - \mathrm{NF} - 10\log_{10} N + 58 \qquad [6]$$

where P_{Signal} is the signal power at the output of the amplifier.

When the optical signal power is sufficiently high, nonlinear effects also become responsible for the degradation of the system performance. Therefore, choosing the value of the signal power launched into the transmission fiber results in a trade-off between increasing the OSNR, and reducing nonlinear effects.

Optical Fiber Amplifier

An optical fiber amplifier consists of fiber with an active medium and at least one optical pump source, which is launched simultaneously with the signal into the fiber by means of a dichroic coupler.

We define as forward and backward pumping, the case where the pump is launched, respectively, in the same direction as the signal and in the opposite direction to the signal. We define as bidirectional pumping, the case where the fiber is pumped from both ends. The most commonly used pump sources are the semiconductor laser diodes. These lasers may use different technologies to meet the various performance requirements with respect to the amplifier types and applications. The most popular pump diodes are the distributed Bragg reflector (DBR), distributed feedback (DFB), or external-grating stabilized lasers.

Erbium Fiber Amplifier

Fiber doped with the rare-earth Er^{3+} ion can amplify a signal located at 1.5 μm with the appropriate fiber host material and if pumped near 0.98 μm or near 1.48 μm. Silica is the most common material used for the fiber host. Er^{3+} has three energy levels, as described in Figure 2.

0.98 μm pumping occurs between the upper level $^4I_{11/2}$ and the ground level $^4I_{15/2}$, and 1.48 μm pumping occurs between the metastable level $^4I_{13/2}$ and the ground level.

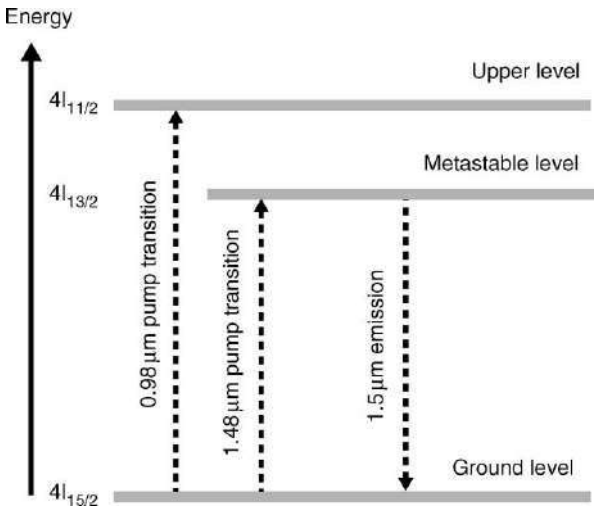

Figure 2 Energy level diagram for Er^{3+} ions.

The signal amplification takes place between the metastable level and the ground level. With 0.98 μm pumping, a rapid energy transfer occurs between the upper level and the metastable level via a phonon-assisted process. The lifetime of the upper level is much lower than the lifetime of the metastable level, about 10 ms. Therefore, with 0.98 μm pumping we can assume a two-level energy transfer, between the metastable and the ground levels, for the description of the amplification process in the Er^{3+} medium.

We define the normalized medium inversion (D) by

$$D = \frac{N_2 - N_1}{N_t} \qquad [7]$$

where N_2 is the population of the metastable level, N_1 the population of the ground level, and $N_t = N_2 + N_1$ the total population. $D = +1$ when the metastable level is fully populated and $D = -1$ when only the ground level is populated.

A simple approach for calculating the gain (G) with erbium-doped fiber amplifiers (EDFA) is given, in decibel units, by

$$G(\lambda) = \big((\alpha(\lambda) + g(\lambda)\big)\frac{D+1}{2} - \alpha(\lambda)L \qquad [8]$$

where λ is the wavelength of the signal, α the absorption coefficient in dB unit length at the signal wavelength, g the emission coefficient in dB unit length at the signal wavelength, and L the length of the erbium-doped fiber (EDF). Here D is the normalized path-average medium inversion of the EDF.

The gain per unit length, $G(\lambda)/L$, is equal to $g(\lambda)$ for $D = +1$ and to $-\alpha(\lambda)$ for $D = -1$.

The absorption and the emission coefficients depend on the EDF host material and the codoping element used for the realization of the EDF. With silicate EDF the most commonly used codoping elements are germanium, aluminum, and phosphorus.

In Figure 3 we show the dependence of the gain per unit length on the normalized path-average medium inversion, for a signal wavelength ranging from 1500 to 1560 nm.

For a more comprehensive description of the EDFA characteristics we use the standard confined-doping model. This model, which is also based on the two-level system, uses the following differential equations:

$$\frac{dP_S^+(z)}{dz} = \left(\frac{N_2}{N_t}(\alpha_S + g_S) - \alpha_s\right)P_S^+(z) \qquad [9]$$

$$\frac{dP_{ASE}^+(z)}{dz} = \left(\frac{N_2}{N_t}(\alpha_S + g_S) - \alpha_s\right) \times P_{ASE}^+(z) + \frac{N_2}{N_t} g_S 2h\nu_S\Delta\nu \qquad [10]$$

$$\frac{dP_{ASE}^-(z)}{dz} = \left(\alpha_s - \frac{N_2}{N_t}(\alpha_S + g_S)\right) \times P_{ASE}^-(z) - \frac{N_2}{N_t} g_S 2h\nu_S\Delta\nu \qquad [11]$$

$$\frac{dP_P^+(z)}{dz} = \left(\frac{N_2}{N_t}(\alpha_P + g_P) - \alpha_P\right)P_P^+(z) \qquad [12]$$

$$\frac{dP_P^-(z)}{dz} = \left(\alpha_P - \frac{N_2}{N_t}(\alpha_P + g_P)\right)P_P^-(z) \qquad [13]$$

where $P_S(z)$, $P_{ASE}(z)$, and $P_P(z)$ in W are, respectively, the signal power, the ASE power, and the pump power along the length (z-axis) of the EDF; $^+$ and $^-$ indicate, respectively, the backward-traveling and the forward-traveling directions; α_s and α_P in m^{-1} are the absorption coefficients at, respectively, the signal and pump wavelengths, g_s and g_p in m^{-1} are the emission coefficients at, respectively, the signal and pump wavelengths, ν_s and ν_p are the frequencies of, respectively, the signal and the pump beams; $\Delta\nu$ is the frequency bandwidth of the ASE; and h is the Planck's constant. N_2/N_1 is the medium inversion along the z-axis of the EDF, and is given by

$$\frac{N_2}{N_t} = \frac{\sum_k \frac{P_k(z)\alpha_k\tau}{h\nu_k B_{eff}^2 N_t}}{1 + \sum_k \frac{P_k(z)(\alpha_k + g_k)\tau}{h\nu_k B_{eff}^2 N_t}} \qquad [14]$$

where the index k represents all the beams, i.e., signal, ASE, and pump, traveling backward and forward along the EDF, τ is the lifetime of the metastable level, and B_{eff} is the effective erbium-doping area. For a typical silicate EDF, having a peak absorption coefficient of $1\ dBm^{-1}$, we have $\tau = 10$ ms, $B_{eff} = 12\ \mu m^2$, and $N_t = 5 \times 10^{23}\ m^{-3}$.

Using the standard confined-doping model we calculate the dependence of the EDFA noise figure on the normalized path-average medium inversion, as shown in Figure 4. We vary the path-average normalized medium inversion by changing the length of the EDFA, while adjusting the pump power so that the signal gain is kept constant.

The noise figure decreases when the medium inversion increases, and approaches the quantum limit value of 3 dB when the path-average normalized medium inversion is close to unity. Near full inversion

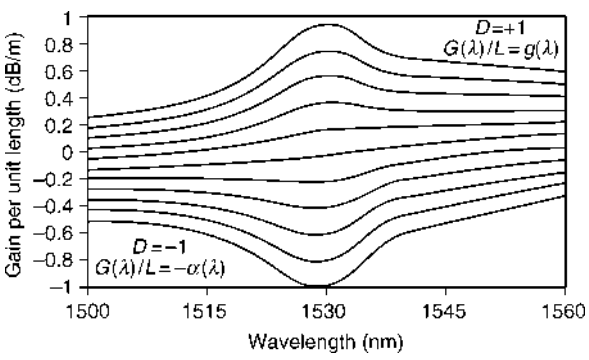

Figure 3 Dependence of the gain per unit length on the normalized path-average medium inversion, D. D ranging from −1 to +1 with a 0.2 step (alumino-silicate EDF; peak absorption of 1 dB m^{-1}; peak emission of 0.93 dB m^{-1}).

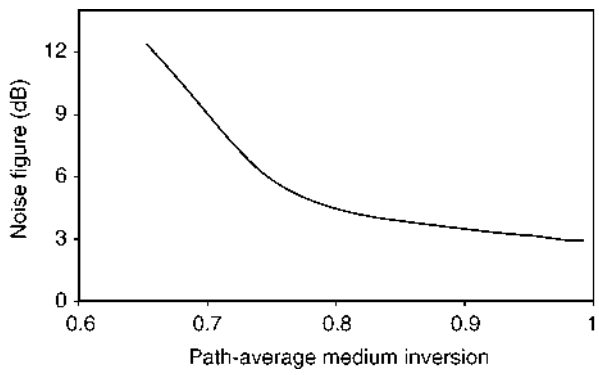

Figure 4 Dependence of the noise figure on the normalized medium inversion (alumino-silicate EDF with peak absorption of 1 dB m^{-1}; EDF lengths ranging from 16 m to 60 m; gain equals to 15 dB; input signal power of −40 dBm at 1530 nm; backward pumping at 980 nm).

may be obtained with a pump located at 0.98 μm because the emission coefficient is null at this pump wavelength. However, this is not the case at 1.48 μm where the ratio g_p/α_p is about 0.3. The maximum medium inversion that can be obtained with 1480 nm is $D = 0.8$. Therefore, higher noise figures are obtained with 1.48 μm pumping than with 0.98 μm pumping. The pumping direction also has an effect on the noise figure, in particular with 0.98 μm. Since at 0.98 μm the emission coefficient is null, when backward pumping configuration is used the 0.98 μm pump is quickly absorbed along the EDF length. This leads to a low-medium inversion at the input end of the EDF, resulting in high noise figures. However, in order to achieve forward pumping, a dichroic coupler must be placed at the input end of the EDF, thus inducing an additional loss at the input of the amplifier. This increases the noise figure and, therefore, to some extent, counterbalances the NF improvement associated with forward pumping.

Raman Fiber Amplifier

Stimulated Raman scattering, an optical-phonon based nonlinear effect present in silica fiber, is used to convert energy from a pump located at a frequency ν_p, to a signal located at a lower frequency ν_s. The efficiency of the so-called Raman amplification depends on the frequency shift $\Delta\nu_r = \nu_p - \nu_s$. Figure 5 shows the dependence of the Raman gain on the frequency shift in standard single-mode fibers. The exact Raman gain spectrum depends on the doping elements used during the fabrication process of the fiber. With standard telecommunications fibers the Raman gain is maximum for $\Delta\nu_r$ close to 13 THz.

Raman amplification can be achieved within the transmission fiber of the system – distributed amplification – or within a specific fiber, such as a dispersion compensation fiber – discrete amplification. Contrary to EDFA, Raman amplification has a very fast response time – in the order of a few hundred femtoseconds – and is strongly dependent on the polarization state of the signal and the pump, the highest gain efficiency being obtained when the polarization states of the signal and the pump are parallel. Because of the nonpolarization-maintaining nature of standard telecommunications fibers, depolarized pump sources are preferable for Raman amplification. To avoid pump fluctuation being transferred to the signal, Raman amplifiers generally use backward pumping, which virtually averages the fluctuation. Forward pumping can also be employed with Raman amplifiers but with careful control of the pump characteristics.

If the signal is sufficiently low the pump depletion can be neglected. In that case the gain of the Raman amplifier is given by

$$G = \exp\left(\frac{g_R P_P}{A_{eff}\alpha_p}(1 - \exp^{-\alpha_p L}) - \alpha_S L\right) \qquad [15]$$

where P_p is the launched pump power, L is the length of the fiber, g_R is the Raman gain coefficient, and A_{eff} is the effective area of the fiber. We assume for the simulation examples given in this article that the peak value of g_R is equal to 6×10^{-14} m/W.

A more comprehensive model for Raman amplification is given by the following set of differential equations:

$$\frac{dP_S^+(z)}{dz} = -\alpha_s P_S^+(z) + \left(\frac{g_R}{A_{eff}}\right)(P_p^+(z) + P_p^-(z))P_S^+(z) + r_S P_S^-(z) \qquad [16]$$

$$\frac{dP_S^-(z)}{dz} = \alpha_s P_S^-(z) - \left(\frac{g_R}{A_{eff}}\right)(P_p^+(z) + P_p^-(z))P_S^-(z) - r_S P_S^+(z) \qquad [17]$$

$$\frac{dP_{ASE}^+(z)}{dz} = -\alpha_s P_{ASE}^+(z) + \left(\frac{g_R}{A_{eff}}\right)(P_p^+(z) + P_p^-(z)) \times [P_{ASE}^+(z) + 2h\nu_S\Delta\nu] + r_S P_{ASE}^-(z) \qquad [18]$$

$$\frac{dP_{ASE}^-(z)}{dz} = \alpha_s P_{ASE}^-(z) - \left(\frac{g_R}{A_{eff}}\right)(P_p^+(z) + P_p^-(z)) \times [P_{ASE}^-(z) + 2h\nu_S\Delta\nu] - r_S P_{ASE}^+(z) \qquad [19]$$

$$\frac{dP_p^+(z)}{dz} = -\alpha_p P_p^+(z) - \left(\frac{\nu_p}{\nu_S}\right)\left(\frac{g_R}{A_{eff}}\right)P_p^+(z) \times (P_S^+(z) + P_S^-(z) + P_{ASE}^+(z) + P_{ASE}^-(z)) + r_P P_p^-(z) \qquad [20]$$

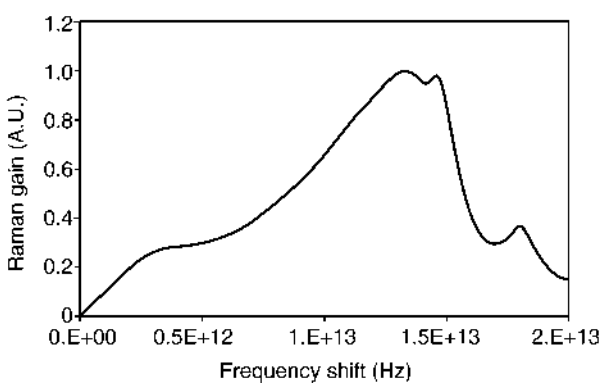

Figure 5 Raman gain spectrum in standard single mode fiber.

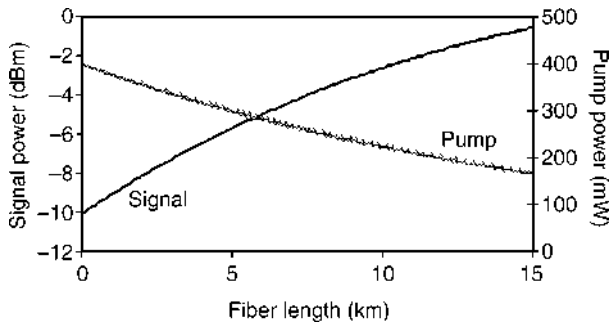

Figure 6 Evolution of the signal power and pump power along fiber length with Raman amplification (standard single-mode fiber; fiber length of 15 km; input signal power of −10 dBm at 1550 nm; forward pumping with 400 mW at 1450 nm; total amplifier gain of 9.5 dB).

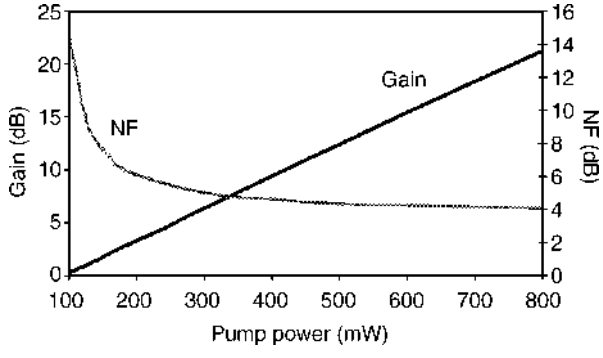

Figure 7 Dependence of the Raman amplifier gain and the noise figure on the pump power (standard single-mode fiber; fiber length of 15 km; input signal power of −10 dBm at 1550 nm; forward pumping at 1450 nm with power ranging from 100 mW to 800 mW).

$$\frac{\mathrm{d}P_{\mathrm{p}}^{-}(z)}{\mathrm{d}z} = \alpha_{\mathrm{p}} P_{\mathrm{p}}^{-}(z) + \left(\frac{\nu_{\mathrm{p}}}{\nu_{\mathrm{S}}}\right)\left(\frac{g_{\mathrm{R}}}{A_{\mathrm{eff}}}\right) P_{\mathrm{p}}^{-}(z) \times \left(P_{\mathrm{S}}^{+}(z) + P_{\mathrm{S}}^{-}(z) + P_{\mathrm{ASE}}^{+}(z) + P_{\mathrm{ASE}}^{-}(z)\right) - r_{\mathrm{P}} P_{\mathrm{p}}^{+}(z) \qquad [21]$$

where r_{S} and r_{p} are the Rayleigh backscattering coefficients at, respectively, the signal wavelength and the pump wavelength.

An example of the evolution of the signal power and pump power along fiber length is given by Figure 6.

Figure 7 shows the dependence of the Raman amplifier gain and noise figure on the pump power.

Distributed and Discrete Amplification

A system uses distributed amplification when the transmission fiber is used as the medium for generating the optical signal amplification. It uses discrete amplification when an optical amplifier is located between the transmission fiber spans.

Discrete Amplification

Both EDFA and Raman amplifiers can be used to provide discrete amplification. Discrete amplifiers can be used as booster, pre-amplifier, or in-line amplifier. They can also be used to compensate for the loss of additional elements such as DCM, which may be located between the fiber spans. In Figure 8 we show the path of the signal when a discrete amplification is used to compensate for the loss of the transmission fiber and for the loss of a DCM. In this example, we assume that the discrete amplifier is made of two amplifiers. The first-stage amplifier compensates mainly for the loss of the transmission fiber, while the second-stage amplifier compensates mainly for the loss of the DCM.

If a 'lossy' element is inserted between two amplifiers, the noise figure of the composite amplifier increases the total noise figure of the dual-stage amplifier.

The total noise figure of a dual-stage amplifier, $\mathrm{NF}_{\mathrm{total}}$, is given by

$$\mathrm{NF}_{\mathrm{total}} = \mathrm{NF}_1 - \frac{1}{G_1} + \frac{\mathrm{NF}_2}{G_1 T_{\mathrm{mid}}} \qquad [22]$$

where NF_1 is the noise figure of the first-stage amplifier, G_1 is the gain of the first-stage amplifier, NF_2 is the noise figure of the second-stage amplifier, and T_{mid} is the loss of the optical element inserted between the two amplifiers ($T_{\mathrm{mid}} < 1$).

Distributed Amplification

When fibers are deployed it is generally assumed, at least for terrestrial networks, that it can be used with different types of transmission systems, which may or

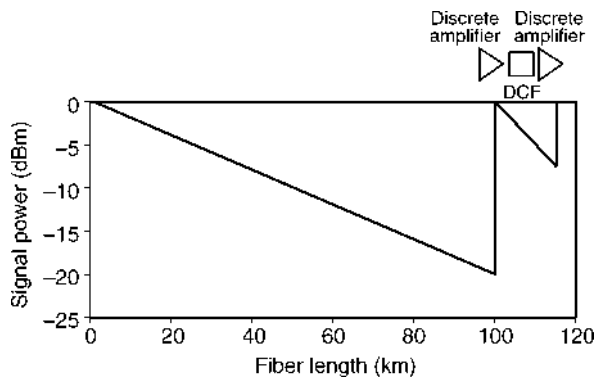

Figure 8 Schematic of a transmission system using discrete amplification to compensate for the loss of the transmission fiber and the dispersion compensating module.

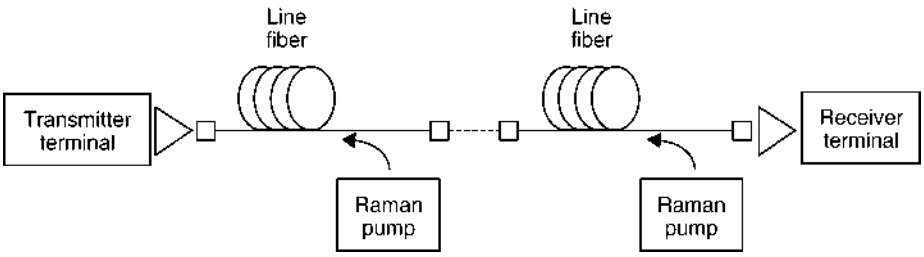

Figure 9 Schematic of a transmission system using distributed Raman amplification.

may not use distributed amplification. Doping the transmission fiber with erbium is not a practical solution since the absence of a pump would result in high fiber loss. On the other hand, Raman amplification can be achieved with conventional fibers, and is consequently the preferred solution to generate distributed amplification. Raman amplification is then obtained by launching a pump signal into the transmission fiber. Figure 9 shows the schematic of a transmission system using distributed Raman amplification with backward pumping.

Even if distributed amplification can be used to compensate fully for the loss of the transmission fiber, discrete amplifiers are generally required to compensate for the loss of the transmitter and receiver terminals. In order to compare distributed amplification with discrete amplification we define the equivalent noise figure as follows:

$$\mathrm{NF}_{\mathrm{Equ}} = \frac{1}{G_{\mathrm{On/Off}}}\left(\frac{\mathrm{Pase}}{h\nu\Delta\nu} + 1\right) \qquad [23]$$

where Pase is the total ASE power at the output end of the transmission fiber and where the on/off gain, $G_{\mathrm{On/Off}}$ is given by

$$G_{\mathrm{On/Off}} = \frac{P^{\mathrm{On}}_{\mathrm{signal}}}{P^{\mathrm{Off}}_{\mathrm{signal}}} \qquad [24]$$

where $P^{\mathrm{On}}_{\mathrm{signal}}$ and $P^{\mathrm{Off}}_{\mathrm{signal}}$ are the signal power levels at the output end of the transmission fiber when the distributed amplifier is, respectively, turned on and off.

To illustrate the meaning of the on/off gain, we show in Figure 10 the evolution of the signal power along a fiber with and without Raman pump power. In this case, the on/off gain is 20 dB.

The equivalent noise figure, which has no real physical meaning, is equal to the noise figure of an imaginary discrete amplifier that would be located at the output end of the transmission fiber and would lead to the same OSNR. Figure 11 shows the dependence of the equivalent noise figure on the Raman gain, with no Rayleigh backscattering

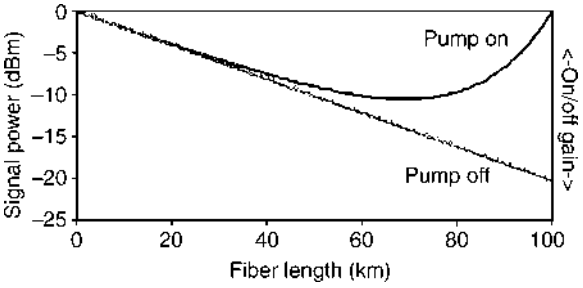

Figure 10 Evolution of signal power along the fiber length with and without pump power (standard single-mode fiber; fiber length of 100 km; fiber attenuation of 0.2 dB/km at signal wavelength; input signal power of 0 dBm at 1550 nm; backward pumping 400 mW at 1450 nm).

and with a Rayleigh backscattering level of $-74\ \mathrm{dB\ m^{-1}}$.

Once the on/off gain is higher than ~10 dB the equivalent noise figure becomes negative and it decreases when the on/off gain further increases. However, once the on/off is higher than 20 dB, Rayleigh backscattering limits the minimum equivalent noise figure that can be achieved. The lowest equivalent noise figure is observed for an on/off gain of ~33 dB. In addition, Rayleigh backscattering causes the reflected portion of the signal to be superimposed onto the transmitted signal, leading to multipath interference (MPI), which level increases with the gain. This results in an additional source of noise for the signal, and consequently degrades the system performance.

Hybrid Amplification

Discrete and distributed amplification can be used simultaneously in a same-transmission system. An example of the evolution of the signal power with hybrid amplification is represented in Figure 12, for the case where distributed Raman amplification and a dual-stage discrete amplifier are used to compensate for the loss of the transmission fiber and for the loss of a DCM.

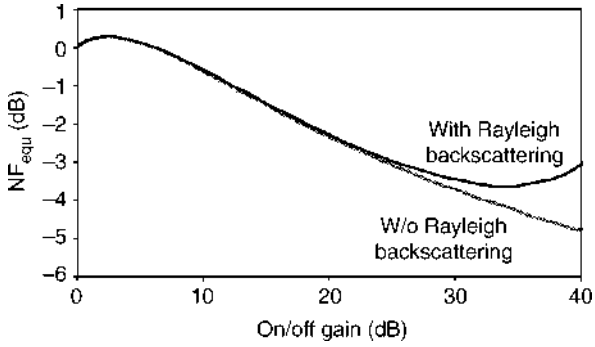

Figure 11 Equivalent noise figure versus distributed on/off gain (standard single-mode fiber; fiber length of 250 km; fiber attenuation of 0.2 dB/km at 1550 nm; input signal power of 0 dBm at 1550 nm; backward pumping ranging from 0 to 1 W at 1450 nm).

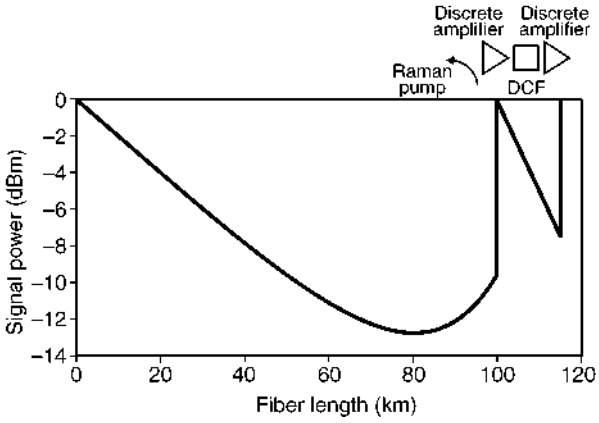

Figure 12 Evolution of the signal power using discrete and distributed amplification (standard single-mode fiber of 100 km; fiber loss of 20 dB at signal wavelength; DCM loss of 8 dB; input signal power of 0 dBm at 1550 nm; Raman backward pumping with 200 mW at 1450 nm; Raman on/off gain of 10 dB).

Similar to the case of distributed amplification, the total equivalent noise figure of a hybrid amplifier can be defined by

$$\mathrm{NF}_{\mathrm{Equ}}^{\mathrm{Total}} = \mathrm{NF}_{\mathrm{Equ}}^{\mathrm{Distributed}} + \frac{\mathrm{NF}^{\mathrm{Discrete}} - 1}{G_{\mathrm{On/Off}}} \qquad [25]$$

where $\mathrm{NF}_{\mathrm{Equ}}^{\mathrm{Distributed}}$ is the equivalent noise figure of the distributed amplifier and $\mathrm{NF}^{\mathrm{Discrete}}$ is the noise figure of the discrete amplifier.

In Figure 13 we show the dependence of the total equivalent noise figure on the on/off gain for a hybrid amplifier with fixed discrete noise figure. The dependence of the distributed equivalent noise figure on the on/off gain is also plotted on the same graph.

Since the equivalent noise figure of the distributed amplifier is relatively low, the total equivalent noise figure is lower than the noise figure of the discrete amplifier once distributed amplification is used. Moreover, as the on/off gain becomes large, the total equivalent noise figure of the hybrid amplifier converges toward the distributed equivalent noise figure.

Wavelength-Division Multiplexing

The transmission of WDM signals over long distances requires meticulous control of the spectral characteristics of the amplifier gain. The amplifier gain excursion induces a power excursion between the channels of the WDM signal, which increases with the number of spans. Consequently, the available bandwidth for WDM transmission narrows with the total distance of the link. This is illustrated in Figure 14, where the power excursion observed after an amplifier chain is plotted for different numbers of amplifiers, a fiber span being included between each amplifier. With this example, the 1 dB-bandwidth drops from 9 nm after one span to 1.5 nm after 50 spans.

For a comprehensive analysis of the spectral gain characteristic of EDFA and Raman amplifiers we generalize the single-channel model described in the amplifier section of this article to the multichannel case. The generalization of the models is obtained by adding a frequency dimension to each beam propagating in the amplifier, as described in Table 1, where each k element represents the signal, the ASE noise or the pump, λ is the wavelength of the kth beam, $\Delta\nu$ is the frequency separation between the pump beam p, and the signal beam s.

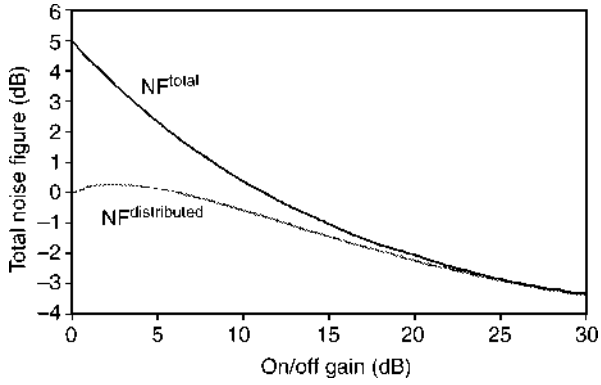

Figure 13 Total noise figure of the hybrid amplifier and noise figure of the distributed amplifier versus on/off gain (noise figure of the discrete amplifier equals to 5 dB; standard single-mode fiber of 100 km; fiber loss of 20 dB at signal wavelength; input signal power of 0 dBm at 1550 nm; Raman backward pumping ranging from 0 mW to 600 mW at 1450 nm).

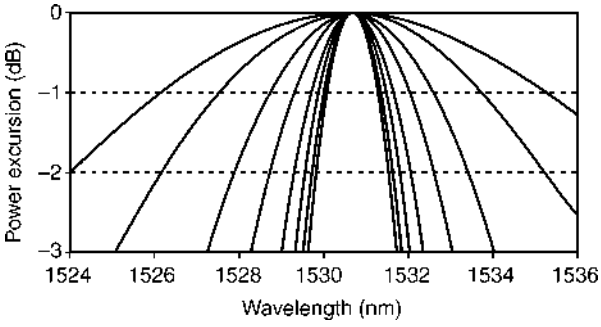

Figure 14 Dependence of the power excursion on the number of fiber spans. Number of spans equals, respectively, to 1, 2, 5, 10, 20, 30, 40, and 50 (amplifier with 15 m long alumino-silicate EDF; peak absorption of 1 dB m^{-1}; $D = 0.6$; fiber loss of 8.5 dB; flat fiber loss spectrum).

Table 1 Correspondence between parameters used for the EDFA and Raman amplifier models with the single-channel and the multichannel approaches (the multichannel approach is also used to describe the power evolution of pump beam when WDM pump scheme is used)

	Single-channel model	*Multichannel model*
Power	$P_k^{+,-}(z)$	$P_k^{+,-}(\lambda, z)$
Sum of power	$P_k^{+}(z) + P_k^{-}(z)$	$\sum_\lambda P_k^{+}(\lambda, z) + P_k^{-}(\lambda, z)$
Absorption coefficient	α_k	$\alpha(\lambda)$
Emission coefficient	g_k	$g(\lambda)$
Raman gain coefficient	g_R	$g_R(\Delta\nu_{p,s})$

With EDFAs, the main factors in the gain flatness are the path-average medium inversion, the EDF codopant composition, and fiber host material. Depending on the bandwidth and the wavelength of the WDM signal, the way the EDF is optimized may vary.

In the following example, we focus on the path-average medium inversion for the case of C-band applications. Figure 15 shows the output spectrum of a WDM signal amplified by an alumino-silicate EDFA having, respectively, a medium inversion of $D = +0.4$ and $D = 0$. The peak-to-peak gain excursion is, respectively, 2 dB and 14 dB for $D = +0.4$ and $D = 0$.

Application with C-band WDM signals requires amplifiers operating with high-medium inversion. As can be extrapolated from Figure 15, if the WDM signal is centered around 1560 nm, a lower medium inversion is more suitable. Indeed, as shown in Figure 3, when the medium inversion is close to zero the peak gain wavelength changes from 1530 nm to 1560 nm and higher; the optical window for L-band applications.

Regarding the doping elements, we show in Figure 16 the gain spectrum for a germano-silicate and an alumino-silicate EDF, respectively. The peak-to-peak gain excursion is 8 dB with the germano-silicate EDF compared to 2 dB with the alumino-silicate EDF.

With Raman amplifiers, the gain flatness can be adjusted by using wavelength-multiplexed pumps. In Figure 17 we show the power spectrum when a Raman amplifier is pumped with three wavelength-multiplexed pumps located, respectively, at 1430 nm, 1455 nm, and 1480 nm. The maximum power excursion between the channels is 2 dB.

For comparison, we also show in Figure 18 the power spectrum of the Raman amplifiers when only

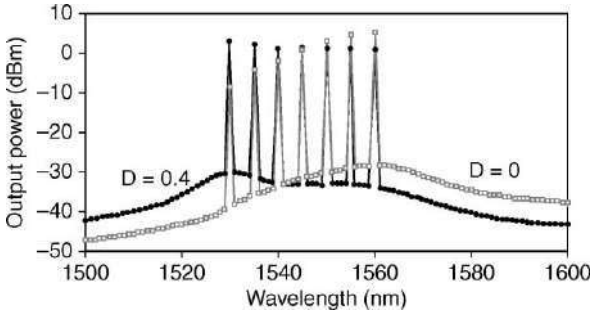

Figure 15 Gain spectrum of an alumino-silicate EDF for $D = 0.4$ and $D = 0$ (C-band WDM signal made of 7 channels equally spaced ranging from 1530 nm to 1560 nm; flat input power of −10 dBm per channel; gain of channel located at 1545 nm of 10 dB for $D = 0.4$ and $D = 0$; EDF with peak absorption of 1 dB m^{-1}; EDF length of 30 m for $D = 0.4$; EDF length of 100 m for $D = 0$).

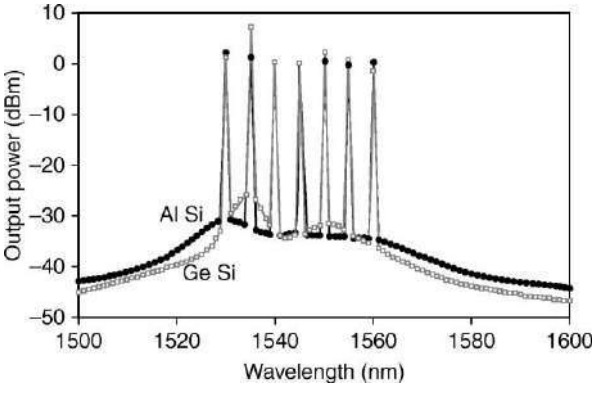

Figure 16 Gain spectrum of an alumino-silicate EDF and of a germano-silicate EDF for $D = 0.4$ (C-band WDM signal made of 7 channels equally spaced ranging from 1530 nm to 1560 nm; flat input power of −10 dBm per channel; gain of channel located at 1545 nm of 10 dB for $D = 0.4$ and $D = 0$; alumino-silicate EDF of 30 m with peak absorption of 1 dB m^{-1}; germano-silicate EDF of 40 m with peak absorption of 1 dB m^{-1}).

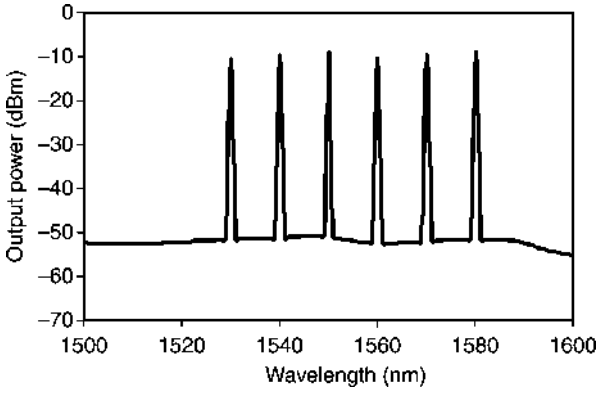

Figure 17 Output power spectrum of a Raman amplifier pumped with three wavelength-multiplexed pumps (standard single-mode fiber of 100 km; WDM signal made of 6 channels equally spaced ranging from 1530 nm to 1580 nm; pumping in forward direction with 100 mW per pump located at 1430 nm, 1455 nm, and 1480 nm).

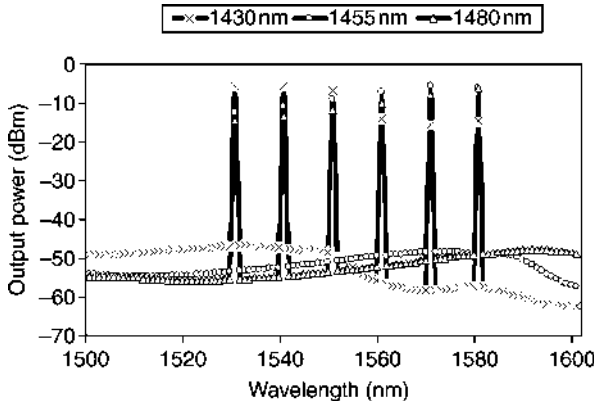

Figure 18 Output power spectrum of Raman amplifiers pumped at three different wavelengths (standard single-mode fiber of 100 km; WDM signal made of 6 channels equally spaced ranging from 1530 nm to 1580 nm; pumping in forward direction with 300 mW at, respectively, 1430 nm, 1455 nm, and 1480 nm).

one of the three multiplexed pumps is used, with the same total pump power. The power excursion becomes 10 dB for the 1430 nm pump, 7 dB for the 1455 nm pump, and 8.5 dB for the 1480 nm pump, which are much higher values than are obtained with the multiplexed pump scheme.

To further reduce the gain excursion of the amplifiers an external gain equalizing filter can be used. The filter, which is located next to the amplifier, is manufactured so that its attenuation spectrum matches the gain spectrum of amplifier, as illustrated in Figure 19. Gain equalizing filters are generally made with thin film filters or fiber Bragg gratings (FBG).

In order to obtain amplification over ultra-wide bandwidths, amplifiers of smaller bandwidths, e.g., C-band and L-band EDFAs, can be operated in parallel by means of Mux and Demux. Another option is to use a Raman amplifier with multiple pumps having their wavelengths distributed over a very large bandwidth.

Long-Haul Transmission System Types

Long-haul optical transmission systems are appropriate for different types of applications. We distinguish here three types of transmission system: the submarine systems; the terrestrial systems; and the unrepeatered systems. Each of these applications requires a careful design of the system, with particular attention to the optical amplifiers. In this section we focus on the key points of their design.

Submarine Systems

Submarine systems are used to transmit information across the oceans, i.e., over distances of up to several thousand kilometers. The transmission fiber and the in-line optical amplifiers, which are packaged in the repeater modules, are deployed undersea at the same time. The drastic constraints imposed on the reliability of the repeaters lead to their high cost, and consequently minimizing their number in a link is essential. The optimum span length, and therefore the number of in-line amplifiers, depends on the total distance of the transmission system. To illustrate this point, we derive from eqn [6] which number of spans of a particular length can be used for a given received OSNR, launched signal power, amplifier noise figure, and span loss per km. Based on the number of spans and span length, we then calculate the maximum distance that can be reached. The result is represented in Figure 20.

With the assumption made here regarding the amplifier characteristics, in order to reach distances of more than 5000 km, span lengths close to 50 km are required.

Terrestrial Systems

Terrestrial systems are used to connect traffic from cities to cities and from cities to submarine-system landing sites. Their total distances vary from a few hundred kilometers to several thousand kilometers. Generally with terrestrial systems, the deployments of the transmission fibers and the transmission systems are done at different times. During its lifetime, the same fiber plant might be used with a different generation of transmission system. Furthermore, because of the geographical constraints that can be imposed on the

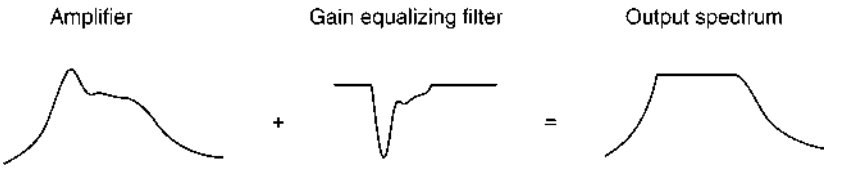

Figure 19 Combination of optical amplifier and gain equalizing filter to provide a flat gain amplifier.

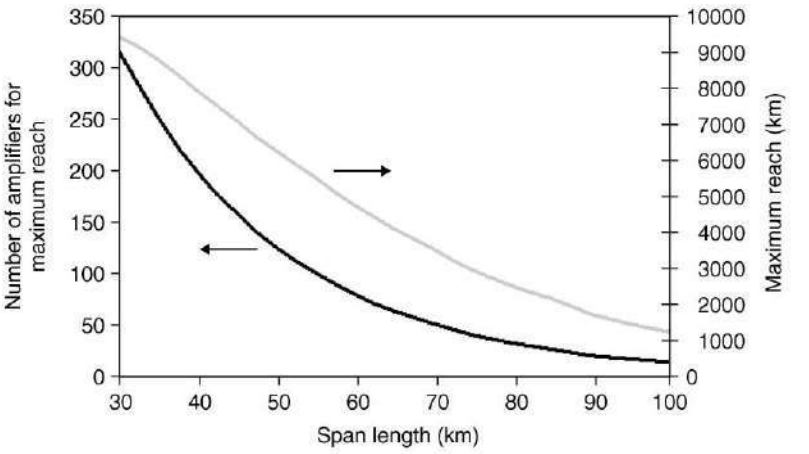

Figure 20 Dependence of the maximum distance and span number of a system on its span length (signal power at the output of the amplifier of −2 dBm; amplifier noise figure of 5 dB; fiber loss of 0.2 dB/km; required OSNR of 20 dB/0.1 nm).

locations of the terminals and the in-line optical amplifiers, the different span lengths may greatly vary in a given link. Consequently, an important characteristic of the terrestrial systems is their capability to operate with different span parameters.

For terrestrial systems using large numbers of WDM channels, maintaining the gain excursion of the optical amplifiers as little as possible under different operating conditions is crucial. Approaches to limit the gain excursion under these variations can be taken. With EDFAs of a given length, the medium inversion depends on its operating gain. The EDFA can be forced to operate at constant gain, independent of the span loss, by adding an adjustable attenuator to the amplifier, e.g., in the middle of a dual-stage amplifier, such that the sum of the attenuator loss and the span loss remains constant. With Raman amplifiers, the control of the gain excursion can be realized by adjusting the power of the pumps when a wavelength-multiplexed pumping scheme is used. Another possibility is to insert dynamically adjustable gain flattening devices along the link. The automatic adjustment of the gain flatness requires the need of a feedback generally provided by an optical channel power monitor.

Unrepeatered Systems

Unrepeatered systems are used to connect mainland sites separated by a water area, such as islands or cities located on the coast. Their main characteristic is that they do not use any electronic component along the transmission fiber; therefore optical amplifiers cannot be placed under water. The length of such systems is typically of a few hundred kilometers, meaning that, even for those systems where very low loss fiber is used, the loss of the transmission fiber is in the order of several tens of dB. To allow sufficiently high OSNR at the receiver, several options can be used. Among them are the use of very high-power boosters, distributed Raman amplification, and remotely pumped discrete amplification. Remote amplification is obtained by inserting, along the submerged fiber cable, a piece of EDF. A 1.48 μm pump is launched into the EDF from the mainland via the transmission fiber or via a dedicated fiber running along the transmission fiber. Figure 21 shows a schematic of a possible configuration for unrepeatered systems when a very high-power booster, distributed Raman amplification, and remote discrete amplification are used simultaneously.

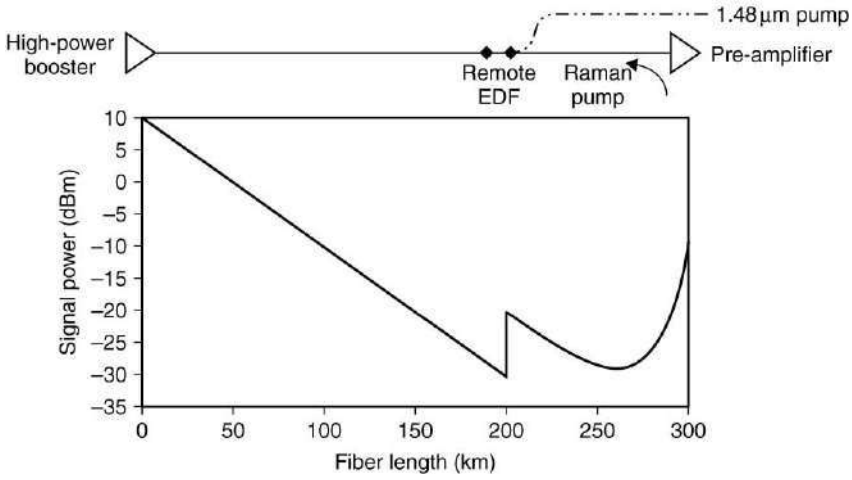

Figure 21 Schematic of unrepeatered systems using high-power booster, remote EDFA, and distributed Raman amplification.

The total equivalent noise figure, $\mathrm{NF}_{\mathrm{Equ}}^{\mathrm{Total}}$, of an unrepeatered transmission system, using remote discrete amplification and distributed amplification, is given by

$$\mathrm{NF}_{\mathrm{Equ}}^{\mathrm{Total}} = T\mathrm{NF}^{\mathrm{Remote}} + \frac{1}{G_{\mathrm{Remote}}} \times \left(\mathrm{NF}_{\mathrm{Equ}}^{\mathrm{Distributed}} + \frac{\mathrm{NF}^{\mathrm{Discrete}} - 1}{G_{\mathrm{On/Off}}} - T\right) \quad [26]$$

where $\mathrm{NF}^{\mathrm{Remote}}$ and G_{Remote} are, respectively, the noise figure and the gain of the remote amplifier; T is the loss of the transmission fiber between the remote amplifier and the pre-amplifier ($T < 1$); $\mathrm{NF}_{\mathrm{Equ}}^{\mathrm{Distributed}}$ and $G_{\mathrm{On/Off}}$ are, respectively, the equivalent noise figure and on/off gain of the distributed amplifier; and $\mathrm{NF}^{\mathrm{Discrete}}$ is the noise figure of the pre-amplifier.

Assuming that the gain and noise figures of the remote EDFA are, respectively, 10 dB and 7 dB, that the remote EDFA is located at 100 km from the pre-amplifier, that the fiber loss per km is 0.2 dB/km, that the Raman distributed amplifier provides 30 dB of on/off gain with an equivalent noise figure of −3 dB, and that the pre-amplifier noise figure is 5 dB, the equivalent total noise figure of the unrepeatered system is −10 dB.

List of Units and Nomenclature

A_{eff}	effective area
ASE	amplified spontaneous emission
B_{eff}	effective erbium doping area
D	normalized medium inversion
dB	ratio in decibel
dBm	power in decibel
DBR	distributed Bragg reflector
DCM	dispersion compensating modules
Demux	optical de-multiplexers
DFB	distributed feedback
DSF	dispersion-shifted fiber
EDF	erbium-doped fiber
EDFA	erbium-doped fiber amplifiers
FBG	fiber Bragg gratings
g_{R}	Raman gain coefficient
G	gain
$G_{\mathrm{On/Off}}$	on/off gain
h	Planck's constant
Length	meter (m)
MPI	multi-path interference
Mux	optical multiplexers
n_{eq}	equivalent input noise factor
NF	noise figure of the amplifier
NRZ	non-return to zero
NZDSF	non-zero dispersion shifter fiber
N_2/N_1	medium inversion
OSNR	optical signal-to-noise ratio
Power	watt (W)
SSMF	standard single-mode fiber
TDM	time-division multiplexing
Time	second (s)
WDM	wavelength-division multiplexing
τ	lifetime

See also

Coherent Lightwave Systems. **Dispersion Management**. **Fiber and Guided Wave Optics:** Nonlinear Effects (Basics). **Optical Amplifiers:** Basic Concepts; Erbrium Doped Fiber Amplifiers for Lightwave Systems; Raman, Brillouin and Parametric Amplifiers.

Optical Communication Systems: Architectures of Optical Fiber Communication Systems; Basic Concepts; Historical Development; Lightwave Transmitters; Optical Time Division Multiplexing; Wavelength Division Multiplexing. **Solitons:** Soliton Communication Systems.

Further Reading

Agrawal G (2001) *Nonlinear Fiber Optics*. London: Academic Press.

Chesnoy J (2002) *Undersea Fiber Communication Systems*. London: Academic Press.

Desurvire E (1994) *Erbium-Doped Fiber Amplifiers, Principles and Applications*. New York: John Wiley & Sons.

Desurvire E, Bayart D, Desthieux B and Bigo S (2002) *Erbium-Doped Fiber Amplifiers, Device and System Developments*. New York: John Wiley & Sons.

Emori Y, Kado S and Namiki S (2002) Broadband flat-gain and low-noise Raman amplifiers pumped by wavelength-multiplexed high-power laser diodes. *Optical Fiber Technology* 8: 107.

Hansen PB and Eskildsen L (1997) Remote amplification in repeaterless transmission systems. *Optical Fiber Technology* 3: 221.

Islam MN (2004) *Raman Amplifiers for Telecommunications*. New York: Springer-Verlag.

Kaminow I and Li T (2002) *Optical Fiber Telecommunications IV*. London: Academic Press.

Keiser G (2000) *Optical Fiber Communications*. New York: McGraw-Hill.

Raman, Brillouin and Parametric Amplifiers

M F S Ferreira, University of Aveiro, Aveiro, Portugal

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Glass fibers for optical communications are made of fused silica, an amorphous material, to which dopant materials of various kinds can be added to produce changes in refractive index. A number of third-order nonlinear processes can occur; these can grow to appreciable magnitudes over the long lengths available in fibers, even though the nonlinear coefficients in the materials are relatively small. The effects are particularly important in single-mode fibers, in which the small mode field dimensions result in substantially high light intensities with relatively modest input powers.

Nonlinear optical effects, such as stimulated Raman scattering (SRS), stimulated Brillouin scattering (SBS), and four-wave mixing (FWM), can cause some problems such as nonlinear loss, pulse distortion, and cross-talk between signals in wavelength division multiplexed (WDM) transmission systems. However, these effects are also useful for efficient amplification and frequency conversion of injected signals, as well as for generation of new frequencies. In this chapter the main properties of fiber Raman amplifiers (FRAs), fiber Brillouin amplifiers (FBAs), and fiber parametric amplifiers (FPAs) will be reviewed, regarding their applications in the field of fiber optic communications.

Fiber Raman Amplifiers

Stimulated Raman scattering is likely to occur in fibers, producing appreciable amplification for down-shifted signals when moderate pump power levels are used. This is largely because the process is phase-matched, so that the long interaction lengths available in fibers will enhance the process. Additionally, the small mode field cross-section yields high intensities for relatively modest power levels.

In the SRS process one has simultaneously the absorption of a photon from the pump beam at frequency ω_p and the emission of a photon at the signal (Stokes) frequency, stimulated by the incident signal beam at frequency ω_s (Figure 1a). The difference in energy is taken up by a high energy

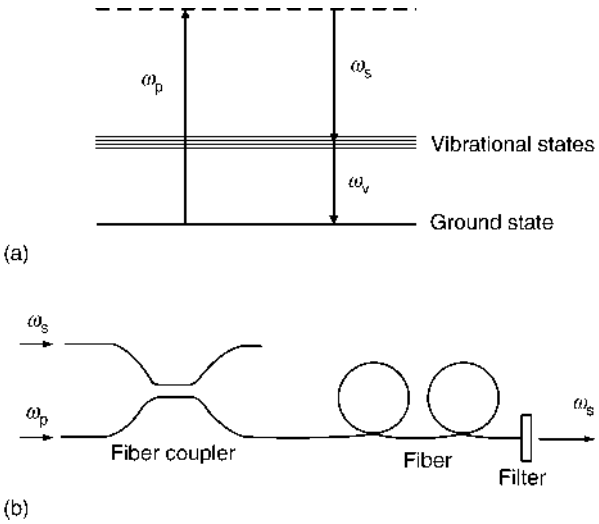

Figure 1 (a) Energy-level illustration of the process of stimulated Raman scattering and (b) schematic diagram of a fiber Raman amplifier.

phonon (molecular vibration) at frequency ω_v. Thus, SRS provides for energy gain at the signal frequency at the expense of the pump.

Figure 1b shows schematically a fiber Raman amplifier. The pump and the signal beams are injected into the fiber through a WDM fiber coupler. The case illustrated in Figure 1a shows the two beams co-propagating inside the fiber, but the counter-propagating configuration is also possible. In fact, the SRS can occur in both directions, forwards and backwards. It was confirmed experimentally that the Raman gain is almost the same in both cases.

Raman Gain and Bandwidth

Figure 2 shows the Raman gain coefficient for fused silica as derived from the spontaneous Raman spectrum. At the peak, the Raman gain coefficient for silica fibers is 9.4×10^{-14} m W^{-1} for a pumping wavelength $\lambda_p = 1.0\,\mu$m and varies as λ_p^{-1}. At this wavelength, a shift of 15 THz corresponds to 50 nm. The most significant feature of the Raman gain in silica fibers is that $\bar{g}_R$ extends over a large frequency range (up to 40 THz) with a broad dominant peak near 13 THz. This behavior is due to the amorphous nature of silica glass, whose molecular vibrational energy levels merge together to form a band.

The use of oxide glasses as dopants for fiber fabrication is suitable to enhance the Raman gain coefficient. For example, in the case of pure GeO_2, the Raman gain coefficient is about 9.2 times that for pure SiO_2 glass. In fact, doping the core of an SiO_2 Raman fiber amplifier with GeO_2 has two beneficial effects. First, the index difference Δ is raised, which increases the effective waveguiding and in turn reduces the effective cross-sectional area A_e. As a consequence, the rate of SRS is increased. Secondly, the Raman gain coefficient increases above that of pure SiO_2 in proportion to the concentration GeO_2, and is given by

$$g_R = (1 + 80\Delta)\bar{g}_R \qquad [1]$$

where $\bar{g}_R$ is the Raman coefficient for pure SiO_2.

In spite of the positive aspects mentioned above, doping with GeO_2 also increases the fiber losses, due to the concomitant rise in Rayleigh scatter and in the number of dopant-dependent loss centers. Thus, greatly increasing the GeO_2 concentration will not be worthwhile if the increased fiber losses outweigh the improvement in amplifier gain.

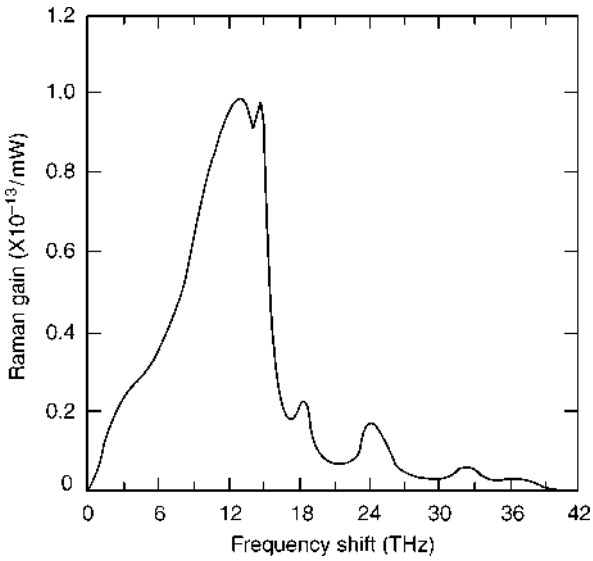

Figure 2 Measured Raman gain spectrum for fused silica at a pump wavelength $\lambda_p = 1.0\,\mu$m. The Raman gain scales inversely with λ_p. Reproduced with permission from Stolen RH and Ippen EP (1973) *Applied Physics Letters* 22: 276. © 1973 American Institute of Physics.

Gain Saturation

The interaction between the pump and Stokes waves is governed by the following coupled equations:

$$\frac{dP_s}{dz} = -\alpha_s P_s + \frac{g_R}{A_e} P_p P_s \qquad [2]$$

$$\frac{dP_p}{dz} = -\alpha_p P_p - \frac{\omega_p}{\omega_s}\frac{g_R}{A_e} P_s P_p \qquad [3]$$

where A_e is the effective cross-sectional area of the fiber mode, α_s and α_p are absorption coefficients which account for the fiber loss at the signal and pump frequencies, respectively, and the signal wave is considered to be co-propagating with the pump wave. Assuming that $\alpha_s = \alpha_p = \alpha$ (which is a reasonable approximation around the 1.5 μm wavelength region in a low loss fiber), the following approximate solutions of eqns [2] and [3] can be written when $P_p(0) \gg P_s(0)$:

$$P_p(z) = \frac{P_p(0)\exp(-\alpha z)}{1 + F(z)} \qquad [4]$$

$$P_s(z) = \frac{P_s(0)\exp\left(\Gamma(1 - e^{-\alpha z}) - \alpha z\right)}{1 + F(z)} \qquad [5]$$

where

$$\Gamma = \frac{g_R P_p(0)}{\alpha A_e} \qquad [6]$$

and

$$F(z) = \frac{\omega_p P_s(0)}{\omega_s P_p(0)} \exp\left(\Gamma(1 - e^{-\alpha z})\right) \qquad [7]$$

The dimensionless parameter $F(z)$ accounts for the effects of pump depletion and the attendant saturation of the gain seen by the signal.

Figure 3 shows the output signal power $P_s(L)$ versus the input signal power $P_s(0)$ for a Raman amplifier with length $L = 8$ km and several values of the input pump power. Typical values $g_R = 6.7 \times 10^{-14}$ m W^{-1}, $\alpha = 0.2$ dB km^{-1}, $A_e = 30 \times 10^{-12}$ m^2, $\lambda_p = 1.46$ μm and $\lambda_s = 1.55$ μm were assumed. The transfer characteristics are linear for $P_p(0) = 0.1$, 0.2, and 0.4 W, but the effects of pump depletion become discernible for input signal levels $P_s(0) > -15$ dB m when $P_p(0) = 0.6$ W. For a pump power $P_p(0) = 0.8$ W the linear behavior of the Raman amplifier cannot be observed, even for signal power levels as low as -25 dB m.

Since, in the absence of Raman amplification, the signal power at the amplifier output would be $P_s(L) = P_s(0)\exp(-\alpha L)$, the amplifier gain is given by

$$G_R = \frac{P_s(L)}{P_s(0)\exp(-\alpha L)} = \frac{\exp(g_R P_p(0) L_e / A_e)}{1 + F(L)} \qquad [8]$$

where

$$L_e = \frac{1}{\alpha}[1 - \exp(-\alpha L)] \qquad [9]$$

is the effective interaction length and L is the actual fiber length. The gain given by eqn [8] is seen to be a function of the input signal power (i.e., a saturation nonlinearity) through the term $F(L)$.

When $F(L) \gg 1$ we have from eqn [5], that

$$P_s(L) = \frac{\omega_s}{\omega_p} P_p(0)\exp(-\alpha L) \qquad [10]$$

In this case, the amplifier output reaches the pump level irrespective of the input signal level. This implies that any spontaneous Raman scatter in the fiber will be amplified up to power levels comparable to that of the pump. This is obviously inadmissible in an amplifier application, thus this operating regime must be avoided.

On the other hand, when $F(L) \ll 1$ we have, from eqn [5], that

$$P_s(L) = P_s(0)\exp(g_R L_e P_p(0)/A_e - \alpha L) \qquad [11]$$

In this case, the amplifier gain is

$$G_R = \frac{P_s(L)}{P_s(0)\exp(-\alpha L)} = \exp(g_R L_e P_0 / A_e) \qquad [12]$$

From eqn [12], the Raman gain in decibels is expected to increase linearly with the pump input power. Raman gains were experimentally confirmed to be almost the same for forwards and backwards pumping.

Figure 4 illustrates the variation of the amplifier gain (full curves, left scale) and the saturation parameter $F(L)$ (dashed curves, right scale) with the pump power $P_p(0)$ for a fixed input signal power $P_s(0) = 0$ dBm and amplifier lengths $L = 2$, 4, and 8 km. The full curve corresponding to $L = 2$ km shows a linear dependence of the gain on the pump power. However, the other two full curves show how the pump depletion limits the achievable gain. The black point on these curves marks the value of the pump power at which the depletion parameter F

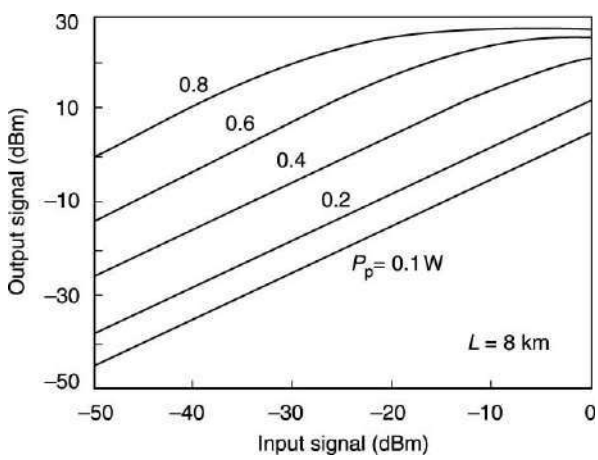

Figure 3 Transfer characteristics for a Raman amplifier with a length $L = 8$ km and several values of the input pump power.

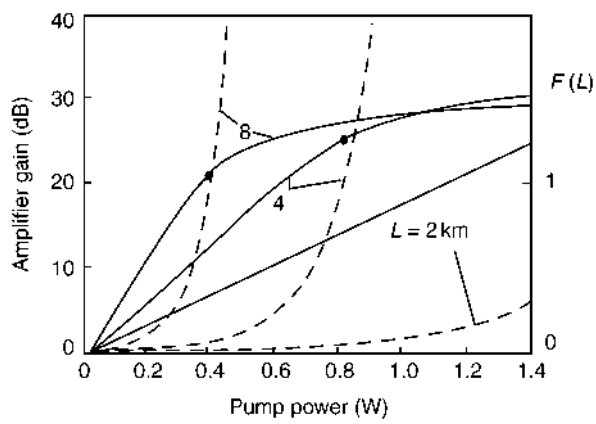

Figure 4 Raman amplifier gain (full curves, left scale) and parameter $F(L)$ (dashed curves, right scale) against input pump power for three values of the amplifier length.

becomes equal to unity. Only operating at pump powers sufficiently below each point ensures linear operation of the amplifier, whilst choosing to operate the amplifier significantly above that point will mean $F \gg 1$ and spontaneous scatter will be amplified to intolerable levels.

Noise Light Power

The dominant noise light in fiber Raman amplification is due to the amplified spontaneously Raman scattered light. In fact, a part of the pump energy is spontaneously converted into Stokes radiation extending over the entire bandwidth of the Raman gain spectrum and is amplified together with the signal. The output thus consists not only of the desired signal but also of background noise extending over a wide frequency range (~10 THz or more).

It has been shown that, for both forwards and backwards pumpings, the noise light power is equivalent to a hypothetical injection of a single-photon per unit frequency at the fiber input end for forwards pumping, and at some distance away from the fiber output for backwards pumping. Assuming that $G_R \gg 1$ and $\alpha_L \gg 1$, the noise light powers for forwards pumping, $P_{fasp}(L)$, and backwards pumping, $P_{basp}(L)$, are approximately given by

$$P_{fasp}(L) \approx h\nu_s \Delta\nu_R (G_R - 1)\exp(-\alpha L) \qquad [13]$$

$$P_{basp}(L) \approx h\nu_s \Delta\nu_R (G_R - 1)/\ln G_R \qquad [14]$$

where $\Delta\nu_R$ is the Raman gain bandwidth, h is Planck's constant, and ν_s is the signal frequency. We can observe from eqn [13] that the noise light power in the case of forward pumping decreases as the fiber length is increased. However, in the case of backwards pumping, it depends mainly on the Raman gain, being nearly independent of fiber length and loss.

FRA Performance and Applications

One main limitation affecting the operation of fiber Raman amplifiers is the requirement of a relatively high-power laser as a pump source. The experiments near 1.55 μm are usually carried out using color-center lasers for this purpose. However, such lasers are too bulky for communication applications and the use of compact high-power laser diodes (LDs) becomes desirable. This objective has recently been achieved with the advent of high-power multimode LDs. In some cases, the output lights from various LDs are combined to attain the necessary pump power.

In a Raman amplifier, SRS has to compete frequently with other nonlinear effects. For example, since the intrinsic gain coefficient for SBS is two orders of magnitude larger than that for SRS, in some cases the SBS may occur at lower pump powers and may affect significantly the SRS process. In particular, the Raman gain may not only be reduced but also become unstable. However, the SBS can be suppressed taking into account its narrow linewidth, which is typically less than 100 MHz. Therefore, if a pump laser with linewidth broader than the Brillouin bandwidth is employed, the SBS effect would not be significant. When a semiconductor laser is used as pump source, its spectral linewidth can be artificially increased by direct frequency modulation, resulting in effective SBS suppression.

Considering its broad bandwidth (>5 THz), fiber Raman amplifiers can be used to amplify several channels simultaneously in a multichannel communication system. The same characteristic also makes such amplifiers suitable for amplification of short optical pulses.

Considering a repeaterless transmission system, the FRA can be used in three different configurations: forwards pumping, backwards pumping, and bi-directional pumping. In the first case, the signal amplification occurs near the fiber input, where the signal power is relatively high. This configuration for the Raman amplification does not generally affect the transmission quality and can compensate for the transmission loss by approximately the value of the Raman gain. For example, 10 dB Raman gain provides an increase in transmission distance of 50 km when the fiber attenuation is 0.2 dB km^{-1}. In the case of backwards pumping, the signal amplification occurs near the fiber output end before photodetection, where the signal power may be very weak, comparable with the Raman noise power. This scheme for the Raman amplification can considerably reduce the minimum detectable signal power, which results in a corresponding increase of the transmission distance. However, the transmission quality is affected due to backwards noise light. The bi-directional pumping configuration uses both forwards and backwards pumping simultaneously and its performance can be estimated from those for the respective unidirectional schemes considered above.

The FRAs can also be used in nonregenerative all-optical multirepeater systems. In this case, the transmission loss is compensated for by both forwards and backwards Raman gain providing the periodically injected pump light. This scheme can be used to overcome the fiber loss in soliton-based communication systems. In this case, the pulse broadening due to chromatic dispersion of the

fiber is also eliminated, which provides the possibility of achieving ultralong distance distortionless transmission.

Fiber Brillouin Amplifiers

The process of SBS can be described as a classical three-wave interaction involving the incident (pump) wave of frequency ω_p, the Stokes wave of frequency ω_s, and an acoustic wave of frequency ω_a. The pump creates a pressure wave in the medium owing to electrostriction, which in turn causes a periodic modulation of the refractive index. This process is illustrated schematically in Figure 5. Since the acoustic wavefronts are moving away from the incident pump wave, the scattered light is shifted downward in frequency to the Stokes frequency:

$$\omega_s = \omega_p - \omega_a \qquad [15]$$

The response of the material to the interference of the pump and Stokes fields tends to increase the amplitude of the acoustic wave. Therefore, the beating of the pump wave with the acoustic wave tends to reinforce the Stokes wave, whereas the beating of the pump wave and the Stokes waves tends to reinforce the acoustic wave. This explains the appearance of the stimulated Brillouin scattering process.

In spite of the apparent similarity between SBS and SRS, these processes differ in three important aspects: i) Brillouin amplification occurs only when the pump and signal beams counter-propagate inside the fiber; ii) the Stokes shift for SBS is smaller by three orders of magnitude compared with that of SRS; and iii) the Brillouin gain spectrum is extremely narrow, with a bandwidth <100 MHz. These characteristics significantly affect the performance and the fiber Brillouin amplifiers.

Brillouin Gain and Bandwidth

Physically, each pump photon in the SBS process gives up its energy to create simultaneously a Stokes photon and an acoustic phonon. Since both the energy and the momentum must be conserved in this process, the frequencies of the three waves must satisfy eqn [15], whereas the wave vectors are related by

$$\vec{k}_p = \vec{k}_s + \vec{k}_a \qquad [16]$$

where $\vec{k}_p$ and $\vec{k}_s$ are the wave vectors of the pump and signal beams, respectively. The frequency ω_a and the wave vector $\vec{k}_a$ of the acoustic wave satisfy the dispersion relation:

$$\omega_a = |\vec{k}_a| v_a = 2 v_a |\vec{k}_p| \sin(\theta/2) \qquad [17]$$

where v_a is the acoustic velocity, θ is the angle between the pump and signal waves, and $|\vec{k}_p| \approx |\vec{k}_s|$ was used. Equation [17] shows that ω_a vanishes in the forward direction ($\theta = 0$), whereas it is maximum in the backward direction ($\theta = \pi$). The frequency shift ν_B in the backward direction is given by

$$\nu_B = \frac{\omega_a}{2\pi} = \frac{2 n v_a}{\lambda_p} \qquad [18]$$

where eqn [17] was used with $|\vec{k}_p| = 2\pi n/\lambda_p$, λ_p is the pump wavelength, and n is the refractive index. Considering $n = 1.45$ and $v_a = 5.96$ km s^{-1} as typical values for silica fibers, we obtain $\nu_B = 11.1$ GHz at $\lambda_p = 1.55$ μm.

Assuming a monochromatic pump, the spectrum of the scattered light depends on the uniformity of the fiber parameters. For a uniform medium the spectrum is Lorentzian with a width $\Delta\nu_B$ determined by the acoustic attenuation. For 1.5 μm light the width is about 17 MHz. However, both the frequency shift ν_B and the gain bandwidth $\Delta\nu_B$ can vary from fiber to fiber because of the guided nature of light and the presence of dopants in the fiber core. In particular, the inhomogeneities in the core section along the length determine an increase of the amplifier bandwidth, which can exceed 100 MHz, although typical values are 50–60 MHz for $\lambda_p \approx 1.55$ μm.

The peak value of the Brillouin gain coefficient is given by

$$g_B(\nu_B) = \frac{2\pi n^7 p_{12}^2 \gamma}{c \lambda_p^2 \rho_0 v_a \Delta\nu_B} \qquad [19]$$

where p_{12} is the longitudinal elasto-optic coefficient, ρ_0 the average material density, n the refractive index, c the vacuum speed of light, and γ is a numerical factor that depends on the polarization properties of the optical fiber, which assumes the value $\frac{1}{2}$ for long lengths of nonpolarization-preserving optical fiber. By using the parameter values typical of fused silica, g_B is estimated to be about 2.5×10^{-11} m W^{-1}, which is two orders of magnitude larger than the Raman gain coefficient at $\lambda_p = 1.55$ μm.

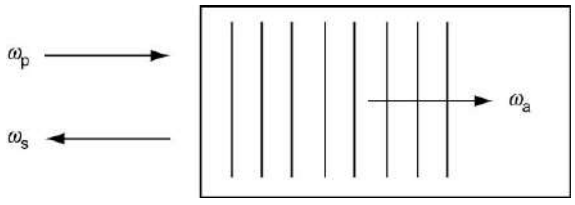

Figure 5 Schematic illustration of the stimulated Brillouin scattering process.

The expression for the Brillouin gain in eqn [19] assumes that the spectral width of the pump beam ($\Delta\nu_p$) is much narrower that the Brillouin linewidth ($\Delta\nu_B$). When such condition is not verified, the Brillouin gain is reduced according with the relation:

$$\tilde{g}_B(\nu_B) = \frac{\Delta\nu_B}{\Delta\nu_B + \Delta\nu_p} g_B(\nu_B) \qquad [20]$$

For this reason, fiber Brillouin amplifiers need to be pumped by narrow-linewidth semiconductor lasers ($\Delta\nu_p < 10$ MHz).

Gain Saturation

The slowly varying steady-state Fourier amplitudes of the Stokes (E_s) and pump (E_p) electric fields are related by the following coupled-wave equations:

$$\frac{\partial E_s}{\partial z} = \left[j\frac{K_1 K_2 |E_p|^2}{\delta - j\Gamma} + \frac{\alpha}{2} \right] E_s \qquad [21]$$

$$\frac{\partial E_p}{\partial z} = -\left[j\frac{K_1 K_2 |E_s|^2}{\delta + j\Gamma} + \frac{\alpha}{2} \right] E_p \qquad [22]$$

where Γ^{-1} is the acoustic phonon lifetime resulting in a spontaneous Brillouin scattering linewidth $\Delta\nu_B = \Gamma/\pi$, $\alpha = \alpha_s \approx \alpha_p$ is the absorption coefficient, and the coupling constants are

$$K_1 = \frac{K_2 \rho_0 n^2 \varepsilon_0}{2\nu_a}, \qquad K_2 = \frac{n^3 p_{12} \omega_s (\hat{e}_p \times \hat{e}_s)}{2c\rho_0} \qquad [23]$$

ε_0 being the free-space permittivity and $\hat{e}_p$ and $\hat{e}_s$ the unit vectors in the directions of the pump and Stokes optical fields, respectively. In eqns [21] and [22] it is assumed that the pump wave is launched at $z = 0$ and propagates in the $+z$ direction, whereas the signal wave is launched at $z = L$ and propagates in the $-z$ direction. The parameter $\delta \equiv \omega_p - \omega_s - \omega_a$ takes into account a possible detuning of the signal from the SBS gain line center.

Equations [21] and [22] can be written in terms of optical power as follows:

$$\frac{dP_s}{dz} = \alpha P_s - \frac{g_B(\delta)}{A_e} P_p P_s \qquad [24]$$

$$\frac{dP_p}{dz} = -\alpha P_p - \frac{g_B(\delta)}{A_e} P_s P_p \qquad [25]$$

where A_e is the effective core area and $g_B(\delta)$ is the Brillouin gain coefficient, given by

$$g_B(\delta) = \frac{\Gamma^2}{\delta^2 + \Gamma^2} g_B(\nu_B) \qquad [26]$$

where $g_B(\nu_B)$ is given by eqn [19].

Considering that $a \ll 1$, the following approximate solutions of eqns [24] and [25] can be written:

$$P_s(z) = P_s(0) D(z) e^{\alpha z} \qquad [27]$$

$$P_p(z) = P_p(0) D(z) H(z) e^{-\alpha z} \qquad [28]$$

where

$$D(z) = \frac{[P_p(0) - P_s(0)]}{[P_p(0) H(z) - P_s(0)]} \qquad [29]$$

and

$$H(z) = \exp\left\{ g_B [P_p(0) - P_s(0)][1 - \exp(-\alpha z)]/(A_e \alpha) \right\} \qquad [30]$$

The Brillouin amplifier gain is given from eqn [27] as

$$G_B = \frac{P_s(0)}{P_s(L) e^{-\alpha L}} = \frac{1}{D(L)} \qquad [31]$$

Figure 6 shows the output signal power $P_s(0)$ versus the input signal power $P_s(L)$ for a Brillouin amplifier with length $L = 8$ km and several values of the input pump power. Typical values $g_B(0) = 2.5 \times 10^{-11}$ m W^{-1}, $\alpha = 0.2$ dB km^{-1}, $A_e = 30 \times 10^{-12}$ m^2 were assumed. The transfer characteristics are linear for $P_p(0) = 0.2$ mW, but the effects of pump depletion become discernible

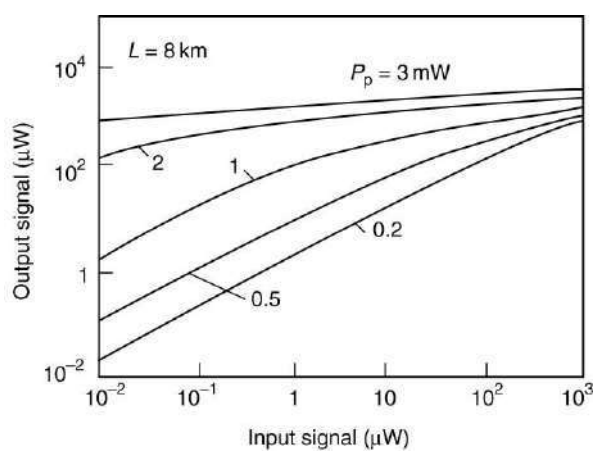

Figure 6 Transfer characteristics for a Brillouin amplifier with a length $L = 8$ km and several values of the input pump power.

for input signal levels $P_s(0) > 2\ \mu W$ when $P_p(0) = 1$ mW. For a pump power $P_p(0) = 2$ mW, linear behavior of the Brillouin amplifier cannot be observed even for signal powers as low as 10 nW.

The dependence of the Brillouin amplifier gain on the input pump power is illustrated in Figure 7 for several amounts of the signal detuning and two values of the signal power. The fiber length is $L = 5$ km and the gain bandwidth is $\Delta\nu_B = 30$ MHz, whereas the other parameter values are equal to those assumed in Figure 6. The saturated gain shows a significant dependence on the input signal power. An input signal of 1 nW, which is comparable to the original value of spontaneous emission, can be amplified by about 60 dB for a pump power $P_p(0) \approx 5$ mW. However, the amplifier gain can be considerably reduced when the signal is detuned from the gain peak. For example, for an input signal power $P_s(L) = 1$ nW and a pump power $P_p(0) \approx 3$ mW, a detuning of 15 MHz (which corresponds to a reduction of the gain coefficient to half of its maximum value) determines a reduction of about 25 dB on the amplifier gain.

Substituting eqn [28] for $P_p(z)$ in eqn [21], the following result is obtained for the Stokes field at the amplifier output:

$$E_s(0) = E_s(L)\sqrt{G_B}\exp\{-\alpha L/2 - j\phi\} \qquad [32]$$

where

$$\phi = \frac{\delta}{2\Gamma}\ln(G_B) \qquad [33]$$

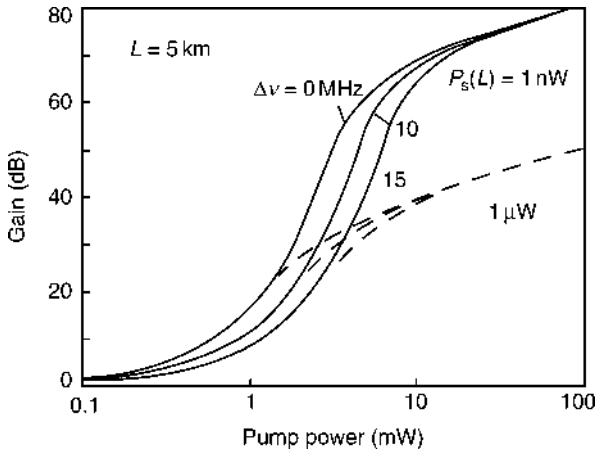

Figure 7 Brillouin amplifier gain against the input pump power for an amplifier length $L = 5$ km and different amounts of detuning from the gain center. The full curves and the dashed curves correspond to input signal powers $P_s(L)$= 1 nW and $P_s(L)$= 1 μW, respectively. Reproduced with permission from Ferreira MFS, *et al.* (1994) *Optical and Quantum Electronics* 26: 35, with permission from Kluwer Academic Publisher.

represents a nonlinear phase change, which depends on the SBS gain, input signal power, and frequency detuning. This nonlinear phase shift can impose a stringent limit on the pump- and signal-frequency stability when FBAs are used in some phase-sensitive detection schemes.

Noise Light Power

To describe the spontaneous emission noise we must add a spontaneous emission term to the equations describing the evolution of the Stokes and pump waves. The amplified spontaneous scattered power per unit frequency is then obtained in the form:

$$P_{asp}(\nu, z) = (h\nu/A_e)g_B(\delta)(N+1)G(\nu, z) \times \int_z^L P_p(z')[G(\nu, z')]^{-1}dz' \qquad [34]$$

where

$$G(\nu, z) = \exp\left\{\int_z^L [P_p(z)g_B(\delta)/A_e - \alpha]dz\right\} \qquad [35]$$

is the gain function, h is Planck's constant, and N is the thermal equilibrium number of acoustic phonons, given by

$$N = \frac{1}{\exp(h\nu_B/kT)^{-1}} \approx kT/h\nu_B \approx 500 \qquad [36]$$

where k is the Boltzmann constant and T is the absolute temperature. The total amplified signal and spontaneous powers are given by integration of $P_s(\nu, z)$ over the gain profile.

Similarly to the behavior of the gain, the noise power depends significantly on the signal magnitude in the saturation regime and becomes particularly high for low signal powers. This high noise level, which is about 20 dB above that of an ideal amplifier, imposes some limitations on the use of FBA as a receiver preamplifier.

FBA Performance and Applications

Fiber Brillouin amplifiers can provide high gains at low pump powers. However, the gain bandwidth is small and the amplified spontaneous emission noise is substantially larger than that observed with other amplifiers, leading to a noise figure that is quite large (>15 dB). This fact limits their usefulness as a preamplifier. Use as in-line amplifiers at high signal levels is possible in some configurations. Some problems in this application are concerned with the narrow bandwidth and the small saturation power of

the FBAs. In particular, the Brillouin linewidth strictly limits the bandwidth of data signals that can be amplified in a FBA.

The Brillouin gain bandwidth can be intentionally extended by more than one order of magnitude by applying frequency modulation to the pump laser. Of course, the bandwidth enhancement is accompanied by a reduction of the peak gain. Consequently, a higher pump power will be necessary to achieve the same gain.

The narrow bandwidth of the FBAs can be used to advantage in coherent lightwave systems to amplify selectively the carrier of a phase- or amplitude-modulated wave, thus allowing homodyne detection, using the amplified carrier as local oscillator. The scheme should work well at bit rates $>100\ \mathrm{Mb\ s^{-1}}$ because the modulation sidebands then fall outside the amplifier bandwidth, and the optical carrier can be amplified selectively. However, an inevitable accompaniment of this narrow-band amplification is the nonlinear phase shift induced by the pump on the signal, given by eqn [33]. This phase shift imposes the most stringent limit on pump- and signal-frequency stability when FBAs are used for self-homodyne detection. In the case of amplitude-shift keyed (ASK) signals, for example, a phase stability of about 0.1 rad for the amplified signal carrier may be required, which corresponds to a pump-signal frequency offset <100 kHz. The nonlinear phase shift can be used with advantage in some self-homodyne schemes which require some specific adjustment of the carrier phase. In the case of a self-homodyne coherent receiver for phase-shift keyed (PSK) signals, for example, a quadrature phase correction is required. Typically, this requires a detuning of only a few MHz from the SBS gain peak, well within the Brillouin gain bandwidth.

Another application of the narrow linewidth associated with the Brillouin gain, is as a tunable narrow-band filter for channel selection in a densely packed multichannel communication system. A channel can be selectively amplified through Brillouin amplification by launching a pump beam at the receiver end so that it propagates inside the fiber in a direction opposite to that of the multichannel signal. Different channels can be selectively amplified by tuning the pump laser. The adjustable bandwidth and high out-of-band rejection can be used to advantage in this case. In another approach, the fiber Brillouin amplifier can be used to selectively amplify multiple subcarrier-multiplexed (SCM) channels carried by a single optical carrier within a WDM signal, or to amplify a backward-propagating supervisory signal.

Fiber Parametric Amplifiers

In both stimulated Raman scattering and stimulated Brillouin scattering, the optical fiber plays an active role through the participation of molecular vibrations or acoustic phonons. However, in many other nonlinear phenomena, the fiber plays a passive role except for mediating the interaction among several optical waves. Since these phenomena involve the light-induced modulation of some medium parameter, they are known as parametric processes.

The parametric processes can be classified as second-order or third-order processes, depending on whether the second-order susceptibility or the third-order susceptibility is responsible for them. Concerning the second-order parametric processes, they are not significant in silica fibers, due to the small value of the second-order susceptibility. In contrast, some third-order parametric processes, namely third-harmonic generation, four-wave mixing, and parametric amplification, can exhibit relatively high conversion efficiencies. The parametric gain in optical fibers can be used to make fiber parametric amplifiers, which find some interesting applications for squeezing and optical fiber communications.

Phase-Matching in Parametric Processes

The efficiency of the four-wave mixing process depends on the relative phase among the interacting optical waves. A high efficiency requires matching of the frequencies as well as of the wave vectors. The latter requirement is often referred to as phase-matching. In contrast with SRS and SBS, where phase-matching is automatically fulfilled as a result of the active participation of the nonlinear medium, the phase-matching condition in parametric processes requires a specific choice of the frequencies and the refractive indices.

Considering the case in which two photons at energies ω_1 and ω_2 are annihilated with simultaneous creation of two photons at frequencies ω_3 and ω_4, the matching of the frequencies imposes the condition:

$$\omega_1 + \omega_2 = \omega_3 + \omega_4 \qquad [37]$$

On the other hand, the phase-matching is given by the condition $\Delta k = 0$, where

$$\begin{aligned}\Delta k &= k_3 + k_4 - k_1 - k_2 \\ &= (\tilde{n}_3\omega_3 + \tilde{n}_4\omega_4 - \tilde{n}_1\omega_1 - \tilde{n}_2\omega_2)/c \qquad [38]\end{aligned}$$

where the $\tilde{n}_i$ are the effective refractive indices of the fiber modes. It is relatively easy to satisfy the phase-matching condition when $\omega_1 = \omega_2$. In this case, a strong pump at $\omega_1 = \omega_2 \equiv \omega_p$ generates a

low-frequency sideband at ω_3 and a high-frequency sideband at ω_4, when we assume $\omega_4 > \omega_3$. These sidebands are referred as the Stokes and anti-Stokes bands, respectively, which are also often called the signal and idler bands. The frequency shift of the two sidebands is given by

$$\Omega_s = \omega_p - \omega_3 = \omega_4 - \omega_p \qquad [39]$$

The effective indices $\tilde{n}_i$ in eqn [38] can be written as

$$\tilde{n}_i = n_i + \Delta n_i \qquad [40]$$

where Δn_i is the change of index refraction due to waveguiding. As a consequence, considering the eqn [40] and the case $\omega_1 = \omega_2$, eqn [38] can be written in the form:

$$\Delta k = \Delta k_m + \Delta k_w \qquad [41]$$

where

$$\Delta k_m = (n_3\omega_3 + n_4\omega_4 - 2n_1\omega_1)/c \qquad [42]$$

results from the material dispersion and

$$\Delta k_w = (\Delta n_3\omega_3 + \Delta n_4\omega_4 - (\Delta n_1 + \Delta n_2)\omega_1)/c \qquad [43]$$

results from the waveguide dispersion. We must note that although the two pump waves have the same frequency, Δn_1 and Δn_2 are, in general, different if these waves propagate in different modes.

The material contribution Δk_m can be expressed in terms of the frequency shift Ω_s given by eqn [39], if we use an expansion of Δk_m about the pump frequency $\omega_p = \omega_1 = \omega_2$. By retaining up to terms quadratic in Ω_s in this expansion, one obtains

$$\Delta k_m = \beta_2\Omega_s^2 \qquad [44]$$

where $\beta_2 = d^2k/d\omega^2$ is the group-velocity dispersion (GVD) coefficient at the pump frequency. In the description of the evolution of the Stokes and anti-Stokes waves, an effective propagation constant appears, given by

$$\kappa = \Delta k + \Delta k_{NL} = \Delta k_m + \Delta k_w + \Delta k_{NL} \qquad [45]$$

In eqn [45], Δk_{NL} represents a nonlinear contribution, given by

$$\Delta k_{NL} = 2\gamma P_p \qquad [46]$$

where $P_p = P_1 = P_2$ is the power in each pump wave and

$$\gamma = \frac{\bar{n}_2\omega}{cA_e} \qquad [47]$$

is the nonlinearity coefficient, $\bar{n}_2$ being the nonlinear-index coefficient.

Parametric Gain

The evolution of the Stokes and anti-Stokes waves are governed by the following coupled equations:

$$\frac{dE_3}{dz} = 2j\gamma P_p \exp(-j\kappa z)E_4^* \qquad [48]$$

$$\frac{dE_4^*}{dz} = -2j\gamma P_p \exp(j\kappa z)E_3 \qquad [49]$$

The general solution of eqns [48] and [49] is given by

$$E_3(z) = (c_1\exp(g_P z) + c_2\exp(-g_P z))\exp(-j\kappa z/2) \qquad [50]$$

$$E_4^*(z) = (c_3\exp(g_P z) + c_4\exp(-g_P z))\exp(j\kappa z/2) \qquad [51]$$

where the coefficients c_i are determined from the boundary conditions and the parametric gain g_P is given by

$$g_P = \sqrt{(2\gamma P_p)^2 - \left(\frac{\kappa}{2}\right)^2} \qquad [52]$$

In the range

$$-6\gamma P_p \le \Delta k \le 2\gamma P_p \qquad [53]$$

the radicand in eqn [52] is positive, so a gain occurs. The maximum gain is achieved when $\kappa = 0$ and it is given by

$$g_{Pmax} = \Delta k_{NL} = 2\gamma P_p = \bar{g}_P\frac{2P_p}{A_e} \qquad [54]$$

where

$$\bar{g}_P = \frac{\omega_p\bar{n}_2}{c} \qquad [55]$$

Considering that $\bar{n}_2 = 3.2\times10^{-20}\ m^2\ W^{-1}$ and assuming a pump wavelength $\lambda_p = 1\ \mu m$, we find that the gain coefficient $\bar{g}_P$ is larger by a factor of 2 than the peak of the Raman gain coefficient g_R. Consequently, the threshold for the parametric process is lower than with the Raman effect. However, the Raman effect dominates in general for long fibers, since it is difficult to maintain phase matching because of fluctuations of the core diameter. One can define a coherence length, given by

$$L_{coh} = 2\pi/\kappa_{max} \qquad [56]$$

where κ_{max} is the maximum tolerable phase mismatch. Significant four-wave mixing occurs only if $L < L_{coh}$.

In the case of a monomode fiber, Δk_w is very small and can be ignored. For λ_p below the zero of dispersion $\lambda_D \approx 1.3\ \mu m$ ($\beta_2 > 0$), one has to keep both Δk_m and Δk_{NL} small by using small frequency shifts Ω_s and low pump powers. When Δk_m dominates, we have

$$L_{coh} \approx \frac{2\pi}{|\Delta k_m|} = \frac{2\pi}{\beta_2 \Omega_s^2} \qquad [57]$$

In the visible range, one has $\beta_2 \approx 50 - 60\ ps^2\ km^{-1}$, and $L_{coh} \approx 1$ km for $f_s = \Omega_s/2\pi \approx 50$ GHz. For a pump wavelength above the zero dispersion ($\beta_2 < 0$), the negative value of Δk_m must be compensated by Δk_{NL}. For phase matching one has

$$\kappa = 0 \approx \Delta k_m + \Delta k_{NL} = \beta_2 \Omega_s^2 + \Delta k_{NL} \qquad [58]$$

From this, we obtain

$$\Omega_s = \sqrt{2\gamma P_p / |\beta_2|} \qquad [59]$$

which coincides with the frequency shift characteristic of the modulation instability.

Amplifier Gain and Bandwidth

Four-wave mixing can be used, for example, in a fiber-optical parametric amplifier. In fact, the main difference between the four-wave mixing experiments and the parametric-amplification experiments is whether a signal at the phase-matched frequency is co-propagated with the pump or not. In the absence of a signal, the signal and idler waves are generated from amplifications of noise.

Let us consider the situation in which a signal wave with a power P_3 is launched at $z = 0$, and simultaneously an idler wave appears, with $P_4(z = 0) = 0$. In this case, considering the solution given by eqs [50] and [51] and taking into account eqn [52], we can write the following results for the signal and idler powers:

$$P_3(z) = P_3(0)\left[1 + \left(1 + \frac{\kappa^2}{4g_P^2}\right)\sinh^2(g_P z)\right] \qquad [60]$$

$$P_4(z) = P_3(0)\left(1 + \frac{\kappa^2}{4g_P^2}\right)\sinh^2(g_P z) = P_3(z) - P_3(0) \qquad [61]$$

The single-pass gain of the parametric amplifier becomes

$$G_P = \frac{P_3(L)}{P_3(0)} = 1 + \frac{P_4(L)}{P_3(0)} = 1 + \left(1 + \frac{\kappa^2}{4g_P^2}\right)\sinh^2(g_P L)$$
$$= 1 + \frac{(2\gamma P_p)^2}{g_P^2}\sinh^2(g_P L) \approx \left(1 + \frac{\kappa^2}{4g_P^2}\right)\sinh^2(g_P L) \qquad [62]$$

The approximation holds for $g_P L \gg 1$. According to eqn [52], amplification ($g_P > 0$) occurs only in the case:

$$|\kappa| < 4\gamma P_p \qquad [63]$$

Since $|\kappa|$ must be small, this also means good phase-matching. If one has complete phase-matching ($\kappa = 0$) and still $g_P L \gg 1$,one obtains exponential behavior for the amplifier gain:

$$G_P = \sinh^2(g_P L) \approx \frac{1}{4}\exp(2g_P L) = \frac{1}{4}\exp(4\gamma P_p L) \qquad [64]$$

From Eqs [44] and [45], we have that the phase mismatch κ and the frequency shift Ω_s are related by

$$\kappa = \beta_2 \Omega_s^2 + \Delta k_w + \Delta k_{NL} \qquad [65]$$

From eqn [65] we have

$$\frac{\Delta\kappa}{\Delta\Omega_s} = 2|\beta_2|\Omega_s \qquad [66]$$

In order to estimate the amplification bandwidth $\Delta\Omega_s$, we set

$$\Delta\kappa = \kappa_{max} = 4\gamma P_p \qquad [67]$$

From Eqs [66] and [67] we obtain

$$\Delta\Omega_s = \frac{2\gamma P_p}{|\beta_2|\Omega_s} = \frac{2}{|\beta_2|\Omega_s L_{NL}} \qquad [68]$$

where $L_{NL} = (\gamma P_p)^{-1}$ is the nonlinear length. Considering typical values $|\beta_2| = 20–60\ ps^2\ km^{-1}$, $\Omega_s/2\pi = 10–100$ THz, and $L_{NL} \sim 1$ m, we obtain a bandwidth in the range ~10–100 GHz. This bandwidth lies between that of a fiber Raman amplifier (~5 THz) and a fiber Brillouin amplifier (~100 MHz).

For $|\kappa| > 4\gamma P_p$ the parametric gain becomes imaginary and there is no longer amplification, but rather a periodical power variation of the signal and idler waves. In particular, if $|\kappa| \gg 4\gamma P_p$, then

$g_P \approx j\kappa/2$ and the parametric amplifier gain is

$$G_P = (2\gamma P_p L)^2 \frac{\sin^2(\kappa L/2)}{(\kappa L/2)^2} \qquad [69]$$

A convenient measure for the width in this case corresponds to a wavevector mismatch $\Delta\kappa = 2\pi/L$, where L is the fiber length. Using eqn [66], we then have

$$\Delta\Omega_s = \frac{\pi}{L|\beta_2|\Omega_s} \qquad [70]$$

Amplifier Performance and Applications

Parametric amplification can induce considerable crosstalk in WDM communication systems. Such crosstalk affects the system performance, particularly for channel spacings less than 10 GHz. Considering a channel spacing of 5 GHz and -20 dB allowed crosstalk, calculations show that the input power per channel should be smaller than 1 mW for a 15 km-long monomode fiber, and 0.1 mW for 100 km-long fiber. However, this effect can be avoided by an appropriate frequency separation of the channels.

Besides the limitation mentioned above, parametric amplification may be useful in the field of optical fiber communication systems. In fact, parametric amplifiers have a bandwidth of ~100 GHz, which is smaller that the bandwidth of fiber Raman amplifiers (~5 THz) but still sufficiently large for many applications. On the other hand, the frequency shift in FPAs can be as large as ~100 THz, which must be compared with the pump-signal frequency shift of ~13 THz of FRAs. These characteristics of FPAs allow a considerable flexibility in the choice of the pump source.

Parametric amplifier gains above 40 dB, with a 30-m-long fiber for input signal powers less than 1 mW and a pump power $P_p \approx 70$ W, were already measured. Such large gain values show clearly the potential of fiber parametric amplifiers provided phase-matching can be achieved. In fact, this condition puts stringent limits on the control of frequency shift Ω_s between the pump and signal waves. One method to satisfy the phase-matching requirement consists in using large birefringence of polarization-preserving fibers. Several parametric amplifiers using this kind of fiber have been demonstrated. Moreover, the possibility to realize a parametric laser was also demonstrated, by placing the fiber inside a Fabry–Perot cavity.

Fiber-optic parametric amplifiers have also attracted considerable attention in the context of a phenomenon known as squeezing. In contrast to coherent light containing equal minimum noise fluctuations in cophasal and quadrature components, in squeezed states one of these components acquires a noise fluctuation below the quantum-noise level at the expense of the other component. This phenomenon can occur in a parametric amplifier for certain values of the relative phase between the signal and idler waves. In fact, spontaneous emission at the signal and idler frequencies generates photons with random phases. Four-wave mixing increases or decreases the number of specific signal-idler photon pairs, depending on their relative phases. The noise reduction below the quantum-noise level can be verified using a phase-sensitive detection scheme in which the phase of the local oscillator is conveniently adjusted. The phase-sensitive nature of parametric amplifiers can also be usefully exploited to reduce the timing jitter in soliton communication systems.

List of Units and Nomenclature

A_e	effective fiber core area, m^2
c	vacuum speed of light, m s^{-1}
E_p	amplitude of the pump wave electric field, V m^{-1}
E_s	amplitude of the signal wave electric field, V m^{-1}
g_B	Brillouin gain coefficient, m W^{-1}
G_B	Brillouin amplifier gain
g_P	parametric gain, m^{-1}
G_P	parametric amplifier gain
g_R	Raman gain coefficient, m W^{-1}
G_R	Raman amplifier gain
$\vec{k}_a$	wave vector of the acoustic wave, rad m^{-1}
$\vec{k}_p$	wave vector of the pump wave, rad m^{-1}
$\vec{k}_s$	wave vector of the signal wave, rad m^{-1}
L	amplifier length, m
L_{coh}	coherence length, m
L_e	effective amplifier length, m
L_{NL}	nonlinear length, m
n	refractive index
$\bar{n}_2$	nonlinear-index coefficient, m^2 W^{-1}
p_{12}	longitudinal elasto-optic coefficient
P_{asp}	amplified spontaneous scattered power, W
P_p	pump power, W
P_s	signal (Stokes) power, W
v_a	acoustic velocity, m s^{-1}
ν_B	Brillouin frequency shift, s^{-1}
α_p	absorption coefficient at the pump frequency, m^{-1}

α_s	absorption coefficient at the signal frequency, m^{-1}
β_2	group-velocity dispersion coefficient, $s^2\ m^{-1}$
γ	nonlinearity coefficient, $W^{-1}\ m^{-1}$
Γ	inverse acoustic phonon lifetime, s^{-1}
δ	frequency detuning from Brillouin gain peak, s^{-1}
$\Delta\nu_B$	Brillouin gain bandwidth, s^{-1}
$\Delta\nu_p$	pump linewidth, s^{-1}
$\Delta\nu_R$	Raman gain bandwidth, s^{-1}
ε_0	free-space permittivity, $F\ m^{-1}$
ρ_0	average material density, $kg\ m^{-3}$
ϕ	nonlinear phase shift, rad
ω_a	acoustic wave frequency, $rad\ s^{-1}$
ω_p	pump wave frequency, $rad\ s^{-1}$
ω_s	signal (Stokes) wave frequency, $rad\ s^{-1}$
Ω_s	sideband frequency shift, $rad\ s^{-1}$

See also

Optical Communication Systems: Wavelength Division Multiplexing. **Nonlinear Optics, Basics:** Four-Wave Mixing. **Scattering:** Raman Scattering.

Further Reading

Agrawal GP (1992) *Fiber-Optic Communication Systems*, Ch. 8. New York: John Wiley & Sons.

Agrawal GP (1995) *Nonlinear Fiber Optics*, 2nd edn, Ch. 8–10. San Diego, CA: Academic Press.

Boyd RW (1992) *Nonlinear Optics*, Ch. 8–10. San Diego, CA: Academic Press.

Buck JA (1995) *Fundamentals of Optical Fibers*, Ch. 7 and 8. New York: John Wiley & Sons.

Cotter D (1987) Fiber nonlinearities in optical communications. *Optical Quantum Electronics* 19: 1–17.

Cotter D (1993) Nonlinearity in optical fiber communication. In: Eason RW and Miller A (eds) *Nonlinear Optics in Signal Processing*, Ch. 9. London: Chapman & Hall.

Forghieri F, Thach RW and Chraplyvy AR (1997) Fiber nonlinearities and their impact on transmission systems. In: Kaminow IP and Koch TL (eds) *Optical Fiber Telecommunications – IIIA*, Ch. 8. San Diego, CA: Academic Press.

Sauter EG (1996) *Nonlinear Optics*, Ch. 5–11. New York: John Wiley & Sons.

Schubert M and Wilhelmi B (1986) *Nonlinear Optics and Quantum Electronics*, Ch. 11 and 12. New York: John Wiley & Sons.

Shen YR (1984) *The Principles of Nonlinear Optics*, Ch. 9–11, 14, 26. New York: John Wiley & Sons.

Semiconductor Optical Amplifiers

M J Connelly, University of Limerick, Limerick, Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

There has been rapid growth in the deployment and capacity of optical fiber communication networks over the past 25 years. This growth has been made possible by the development of new optoelectronic technologies that can be utilized to exploit the enormous bandwidth of optical fiber. Today, systems are operational which operate at bit rates in excess of 100 Gb/s. Optical technology is the dominant carrier of global information. It is also central to the realization of future networks that will have the capabilities demanded by society. These capabilities include virtually unlimited bandwidth to carry communication services of almost any kind, and full transparency that allows terminal upgrades in capacity and flexible routing of channels. Many of the advances in optical networks have been made possible by the optical amplifier.

Optical amplifiers can be divided into two classes: the optical fiber amplifier and semiconductor optical amplifier (SOA). The former dominates conventional system applications such as in-line amplification to compensate for optical link losses. However, due to advances in optical semiconductor fabrication techniques and device design, the SOA is showing great promise for use in evolving optical communication networks. It can be utilized as a general gain element but also has many functional applications, including optical switching and wavelength conversion. These functions, where there is no conversion of optical signals into the electrical domain, are required in transparent optical networks.

We will review SOA basics, technology (materials and structures), signal transmission performance (pattern effects, crosstalk, and ultrashort pulse amplification), and some important functional applications (optical switching and wavelength conversion).

Basic Principles

The SOA is based on similar technology as a laser diode. Optical gain is achieved by electrically pumping a suitable semiconductor material, such that a

population inversion occurs between the material conduction and valence bands. An incoming photon can then be amplified when the resulting stimulated emission exceeds losses due to stimulated absorption and internal material losses. SOAs can be designed to operate in either the 1300 nm or 1550 nm optical communication windows. The principle of operation of an SOA with low residual reflectivities ($R \approx 0$) is shown in Figure 1. An input optical power experiences a single-pass gain $G = \exp(gL)$ after traveling through the SOA active waveguide of length L. The net gain coefficient $g = \Gamma g_m - \alpha_{int}$, where Γ is the optical confinement factor (the fraction of the propagating signal power confined to the SOA waveguide), g_m the material gain, and α_{int} the optical loss coefficient. g_m is a function of the injected carrier (electron) density and wavelength. The single-mode active waveguide can support two orthogonal polarization modes: transverse electric (TE) and transverse magnetic (TM).

The amplification process also adds broadband noise, amplified spontaneous emission (ASE) to the output signal. The amplifier noise figure (NF) is a measure of the signal-to-noise ratio (SNR) degradation of the signal after amplification. It is defined as the ratio of the SNR of the input to that of the output when the input noise is shot noise limited. Expressions for the output SNR are usually calculated assuming that the SOA is followed by a narrowband optical filter and an ideal photodetector, such that the main source of detector noise is the beat noise between the signal and the ASE falling within the filter bandwidth. In this case the SOA noise figure is given by

$$\mathrm{NF} = 2n_{sp}K \qquad [1]$$

with the excess noise factor K given by

$$K = \frac{(1 + RG)(G - 1)\Gamma g_m}{(1 - R)G(\Gamma g_m - \alpha_{int})} \qquad [2]$$

where n_{sp} is the effective population inversion parameter. To achieve a low NF, the internal losses must be small and a value of n_{sp} close to one is required. A favorable value of n_{sp} is achieved by operating the SOA at high gain.

A low R is required to prevent the SOA from oscillating at high gains. The residual reflectivity is manifest as ripples in the amplifier gain spectrum and ASE spectrum.

SOA Structures

The key parameters required for a practical SOA are:

- low reflectivity ($<10^{-4}$) to ensure low gain ripple (<0.5 dB);
- low optical loss coefficient to achieve a high gain;
- high material gain to allow low drive current operation;
- low polarization sensitivity (<0.5 dB) because the polarization state of the optical signal coming from a link fiber is usually random;
- high saturation output power (P_{sat}), defined as the output power at which the gain is reduced by 3 dB; and
- low fiber-to-chip coupling losses (<3 dB per facet).

A schematic diagram of a commercial SOA chip is shown in Figure 2. The active waveguide consists of a 0.2 μm thick InGaAsP bulk active layer sandwiched between 0.1 μm thick InGaAsP separate confinement heterostructure (SCH) layers. The central section of the active waveguide is 600 μm long with a constant width of 1.4 μm. The mode-expanding active waveguide tapers are 150 μm long with a width that changes linearly from 1.4 μm to 0.4 μm at the tip. The tapers allow optical coupling to an underlying passive waveguide enabling efficient coupling to the input and output optical fibers. The structure provides a high confinement factor because of the refractive index mismatch between the layers in the gain section. The p-n junction formed by the p-InP

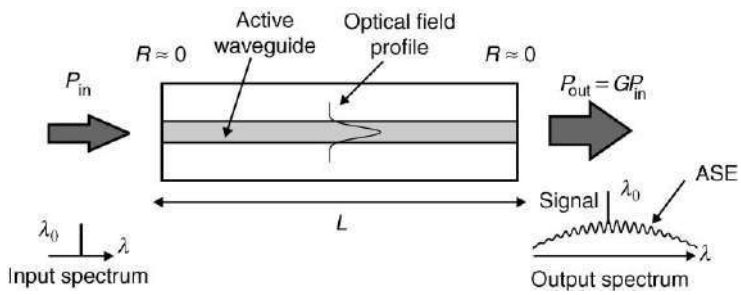

Figure 1 SOA basic structure.

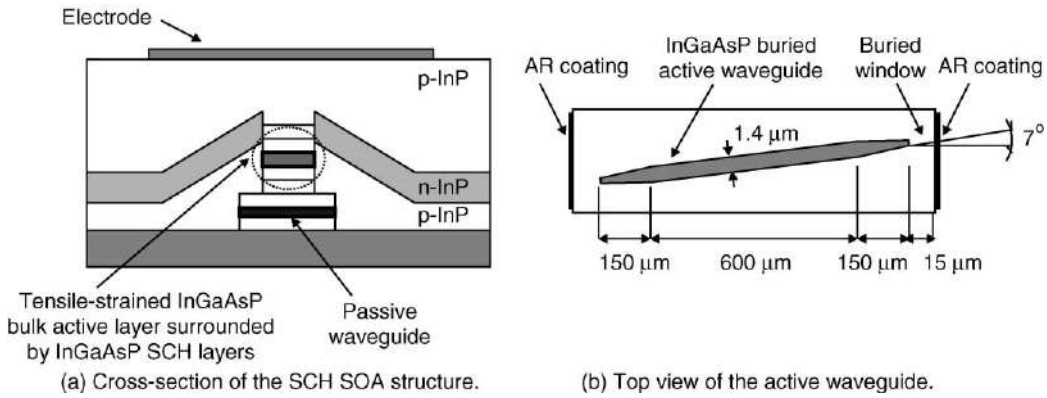

Figure 2 SCH SOA with tensile-strained bulk active layer.

Table 1 Typical specifications of a commercial SOA

Wavelength of peak gain	1550 nm
Peak fiber-to-fiber gain	22 dB
Noise figure	6.5 dB
Saturation output power	10 dBm
Polarization sensitivity	0.2 dB
Gain ripple	0.1 dB
3 dB bandwidth	40 nm
Drive current	200 mA

and n-InP layers acts as a current block, thereby providing good confinement of the injected carriers from the drive current to the active layer. Very low-reflectivity ($<10^{-6}$) is obtained by combining buried windows with antireflection-coated tilted facets (7° tilt angle).

Because of the active waveguide asymmetry, the TE confinement factor is larger than the TM confinement factor. g_m is isotropic in bulk material. The introduction of tensile strain, due to the lattice mismatch between the active layer and SCH layers, causes the bulk material band structure to change in such a way that the TM material gain is greater than the TE material gain. The introduction of the correct amount of tensile strain compensates for the difference in the TE and TM confinement factors leading to low polarization dependence.

Other SOA structures that can achieve low polarization dependency are based on active waveguides with near square cross-section having almost the same TE and TM confinement factors or the combination of compressively strained quantum wells (higher TE gain) and tensile strained quantum wells (higher TM gain). Typical specifications and characteristics of a commercial SOA are listed in Table 1 and shown in Figure 3.

As Figure 3 shows, gain saturation in conventional SOAs is manifest at output signal powers well below P_{sat}. Because the gain recovery time (carrier lifetime) in SOAs is fast (typically 0.1–1 ns), this can lead to pattern effects in single-channel systems and severe crosstalk in wavelength division multiplexed (WDM) systems. This problem can be greatly reduced by the use of a gain-clamped SOA (GC-SOA). In a GC-SOA, lasing action is produced, at a wavelength remote from the operating wavelength range, by the introduction of wavelength specific feedback. Once lasing begins, the carrier density in the SOA active layer is clamped at a fixed value. Changes in the input signal power lead to opposing changes in the lasing mode power. This has the effect of keeping the carrier density fixed (i.e., clamped), thereby making the signal gain relatively insensitive to changes in the total input power. A common method of providing this feedback is the use of distributed Bragg reflectors (DBRs) as shown in Figure 4. A typical GC-SOA gain versus output power characteristic is shown in Figure 5.

Basic Applications of SOAs in Optical Communication Systems

The three basic applications of SOAs in optical communication systems are: booster amplifier, in-line amplifier, and preamplifier. The main requirements of SOAs for such applications are listed in Table 2.

A booster amplifier is used to increase the power of a high power signal. Boosting laser power in an optical transmitter can be used to overcome external modulator losses, compensate for splitting and tap losses in optical distribution networks and to increase the power budget of optical links. The most critical requirement of a booster amplifier is a high P_{sat}, to obtain a high output signal power and minimize pattern effects. SOAs with P_{sat} in excess of 10 dBm are now available commercially.

The sensitivity of a conventional optical receiver is limited by thermal noise. The sensitivity is the

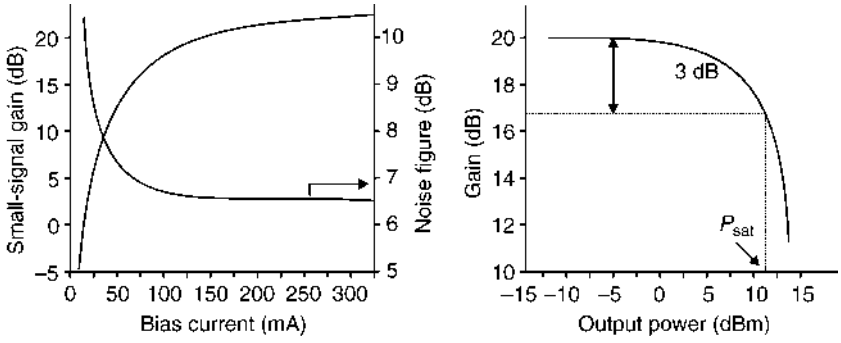

Figure 3 Typical commercial SOA small-signal gain and noise figure versus bias current and gain versus output power characteristics.

minimum signal power required at the receiver input to achieve a desired bit-error-rate (typically 10^{-9}). An optical preamplifier can be used to increase the power level of an optical data signal prior to detection and demodulation, leading to an increase in sensitivity. The performance of a preamplified optical receiver is dependent on the detector signal-to-noise ratio, $\mathrm{SNR} = i_{\mathrm{sig}}^2 / \bar{i}_{\mathrm{noise}}^2$. The signal photocurrent i_{sig} is proportional to the amplified signal power. The mean square noise current $\bar{i}_{\mathrm{noise}}^2$ includes circuit noise (thermal noise and detector dark current noise), signal shot noise, spontaneous emission shot noise, signal-spontaneous beat noise, and beat noise between the spectral components of the spontaneous emission. The best improvement in SNR occurs when the SOA is operated in the signal-spontaneous beat noise limit. In this regime, the signal power is sufficiently large such that the dominant receiver noise is the signal-spontaneous beat noise that falls within the signal bandwidth. This usually requires that the spontaneous emission from the SOA be reduced using a narrowband optical filter. The noise figure is a critical parameter in this application and should be as low as possible.

In long-haul optical transmission systems, in-line optical amplifiers can be used to compensate for link losses thereby increasing the spacing between optical regenerators. The main advantages of in-line SOAs are transparency to data rate and modulation format (unsaturated operation), bidirectionality, WDM capability, simple mode of operation, low power consumption, and compactness. The latter two advantages are important for remotely located optical components. To avoid noise accumulation in the link, it may be necessary to follow each SOA by a narrowband optical filter, but this can prevent capacity enhancements using WDM.

An example of a single channel optical transmission system utilizing booster, in-line, and preamplifier SOAs is shown in Figure 6. The transmitter comprised a 1309 nm gain-switched laser diode directly modulated with a 10 GHz sinusoid to produce a train of 40 ps wide pulses at a repetition rate of 10 GHz. At 1309 nm fiber dispersion is small and the maximum transmission distance is mainly limited by the link losses. The laser output was connected to an external modulator, driven by a 2^{31}-1 pseudo-random bit sequence (PRBS) to produce an optical data stream with an extinction ratio of 13 dB. A booster SOA was used to increase

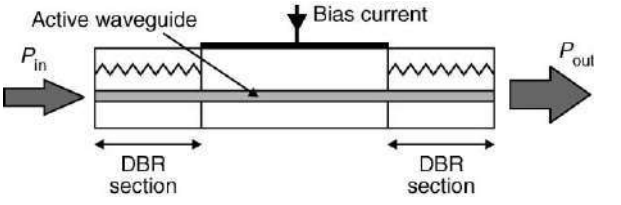

Figure 4 Schematic structure of a GC-SOA.

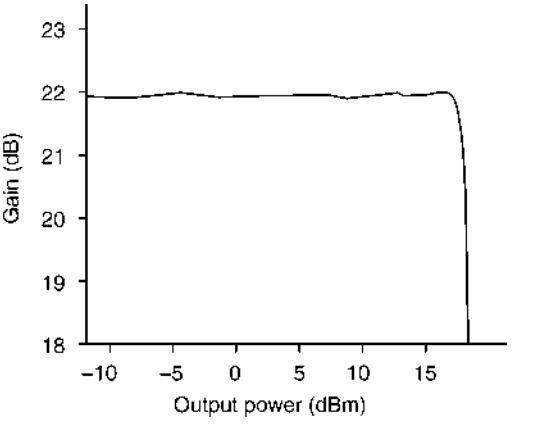

Figure 5 Typical GC-SOA gain versus output power characteristic.

Table 2 Requirements of basic SOA applications

Requirement	*Booster amplifier*	*In-line amplifier*	*Preamplifier*
High gain	Yes	Yes	Yes
High P_{sat}	Yes	Yes	Not critical
Low noise figure	Not critical	Yes	Yes
Low polarization sensitivity	Not critical	Yes	Yes
Optical filter	Not necessary	Not critical	Yes

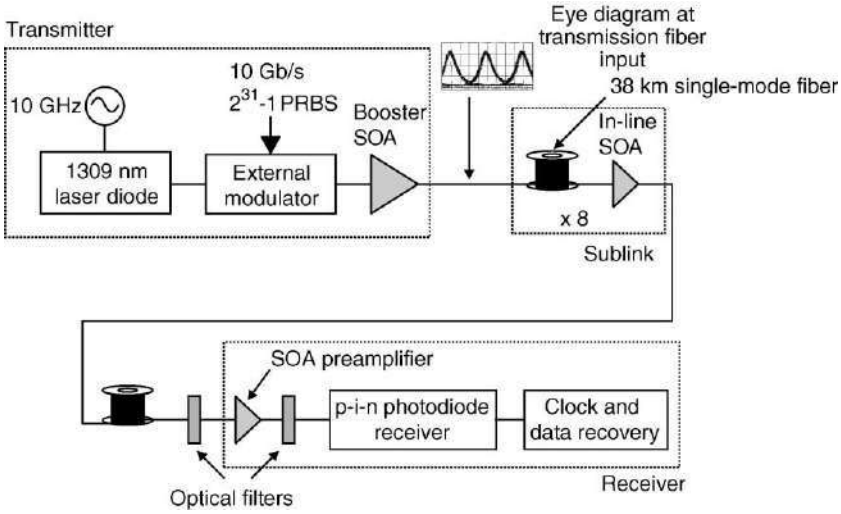

Figure 6 10 Gbit/s transmission experiment utilizing booster, in-line and preamplifier SOAs. (Adapted from Kuindersma *et al.* (1996) 10 Gbit/s RZ transmission at 1309 nm over 420 km using a chain of multiple quantum-well semiconductor optical amplifier modules at 38 km intervals. European Conference on Optical Communications 1996.)

the average transmitted power to between 0 and 2 dBm (7–9 dBm peak power). The transmission fiber length was 420 km with 12 in-line SOAs, used to compensate for fiber loss, spaced at 38 km intervals. At the receiver the signal was passed through a 1 nm bandpass filter, to reduce the accumulated spontaneous emission, amplified by an SOA preamplifier and filtered by a 1 nm bandpass filter. The signal was then detected by a p-i-n photodiode followed by an electronic clock and data recovery circuit. The SOA preamplifier and filter increased the receiver sensitivity from −14 dBm to −31 dBm for a BER of 10^{-10}. The receiver penalty after 420 km was 5 dB. In this experiment, the main limitation on transmission distance was the accumulation of spontaneous emission within the optical filter bandwidth.

Pattern Effects and Crosstalk

When an SOA is operated in the unsaturated regime, the amplifier gain is independent of the number of input signals and signal data rate. Outside this regime, the SOA will cause distortion because at high input powers, the gain saturates and compresses. Dynamic gain saturation occurs on the same time scale as the gain recovery time. This leads to pattern effects where the power of an incoming bit affects the gain experienced by subsequent bits. This is particularly important when the bit rate is of the same order as the inverse of the gain recovery time as shown in Figure 7. In WDM systems cross-gain modulation (XGM) between the channels can lead to severe interchannel crosstalk.

A further complication in WDM systems using SOAs is interchannel crosstalk caused by four-wave mixing (FWM). FWM is a coherent nonlinear process between two optical fields within the SOA, resulting in gain modulation at the beat frequency between the fields, and in the process generating new sidebands. In WDM systems with equally spaced wavelengths, the net effect of FWM is the generation of crosstalk terms interfering with the desired signals. Although the crosstalk can be relatively low in power terms, it can produce significant power penalties due to the coherent beat noise phenomena. The level of FWM crosstalk increases as the channel spacing decreases and the channel output power increases.

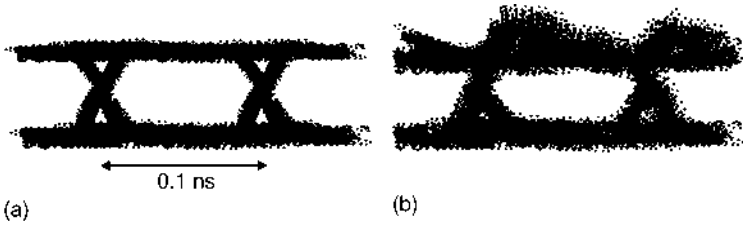

Figure 7 Typical eye diagrams of 10 Gb/s nonreturn to zero data (a) before and (b) after amplification by an SOA. ASE noise is present in the output eye diagram.

FWM crosstalk is of particular importance in dense WDM systems where the inter channel spacing is less than 100 GHz. Figure 8 shows a typical SOA output spectrum for an 8-channel multiplex with FWM-generated crosstalk signals.

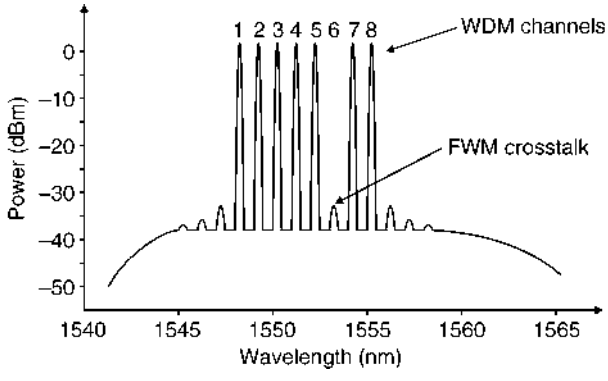

Figure 8 Typical SOA output spectrum for an 8-channel WDM multiplex with 100 GHz spacing. Channel 6 has been dropped to show the FWM crosstalk signal caused by channels 5 and 7.

Ultrashort Pulse Amplification

Optical time division multiplexing is an efficient method for increasing the bit rate of optical transmission systems. This requires the time interleaving of short optical pulses. The transmission distance is limited by the fiber group velocity dispersion, which is proportional to the pulse spectral width. Soliton transmission is also possible where the pulse shape is preserved as it propagates in the fiber.

Because an SOA has a very large bandwidth (typically 5 THz) it is capable of amplifying pulses as short as 100 fs. An input pulse can be amplified without significant distortion if the pulse energy is much less than the saturation energy E_{sat} of the SOA. Typical SOA saturation energies are of the order of a few pico Joules. As the pulse energy approaches E_{sat}, considerable spectral broadening and distortion can result. For an input pulse width τ_p (full width at half maximum) of the order of 10–100 ps spectral broadening is primarily due to self phase modulation (SPM). SPM is caused by gain saturation, which leads to intensity-dependent changes in the SOA active layer refractive index in response to carrier density variations. The degree of spectral broadening and distortion also depends on the input pulse shape. If τ_p is much less than the carrier lifetime, the output pulse power $P_{out}(t)$ and phase $\phi_{out}(t)$ are given approximately by

$$P_{out}(t) = P_{in}(t)\exp[h(t)] \qquad [3]$$

$$\phi_{out}(t) = \phi_{in}(t) - \frac{1}{2}\alpha h(t) \qquad [4]$$

with

$$h(t) = \left[1 - \left(1 - \frac{1}{G_0}\right)\exp\left(\frac{-U_{in}(t)}{E_{sat}}\right)\right]^{-1} \qquad [5]$$

$$U_{in}(t) = \int_{-\infty}^{t} P_{in}(\tau)\mathrm{d}\tau \qquad [6]$$

where $P_{in}(t)$ and $\phi_{in}(t)$ are the input pulse power and phase, respectively. G_0 and α are the amplifier unsaturated gain and linewidth enhancement factor, respectively. Typical values for α are in the range 2–10 and depend on the amplifier active region material and operating conditions. The output pulse spectrum can be obtained from

$$S(\omega) = \left|\int_{-\infty}^{\infty} [P_{out}(t)]^{1/2}\exp[i\phi_{out}(t) + i(\omega - \omega_0)t]\mathrm{d}t\right|^2 \qquad [7]$$

where $\nu_0 = \omega_0/2\pi$ is the pulse optical frequency. The output pulse chirp (frequency variation) is given by

$$\Delta\nu_{out}(t) = -\frac{1}{2\pi}\frac{\partial\phi_{out}}{\partial t} = \Delta\nu_{in}(t) + \frac{\alpha}{4\pi}\frac{\partial h}{\partial t} \qquad [8]$$

where $\Delta\nu_{in}(t)$ is the input pulse chirp. Using the above theory the shape, chirp and spectrum of an amplified zero-chirp (transform limited) Gaussian

pulse of input power

$$P_{\text{in}}(t)=\frac{E_{\text{in}}}{\tau_0\sqrt{\pi}}\exp\left(-\frac{t^2}{\tau_0^2}\right)$$

is shown in Figure 9. E_{in} is the pulse energy and $\tau_p \approx 1.665\tau_0$. The amplified pulse is asymmetric because the leading edge of the pulse experiences a larger gain than the trailing edge. The amplified pulse spectrum is broader than the input pulse and also develops a multipeak structure. This is due to the SPM-induced frequency chirp imposed on the pulse as it propagates through the amplifier. In this case the chirp is linear around the peak of the pulse power spectrum. It can be compensated by transmitting the pulse through an optical fiber with anomalous group-velocity dispersion. In practice optical pulses can be far from Gaussian and can also have an initial chirp. In this case the induced chirp and resulting pulse spectrum can be more complex and more difficult to compensate.

When the input pulse width is less than ~10 ps, the above theory is no longer adequate and other nonlinear effects within the SOA such as carrier heating and spectral hole burning must be taken into account. The resulting pulse power and spectral distortions induced by an SOA on such pulses can be very complex.

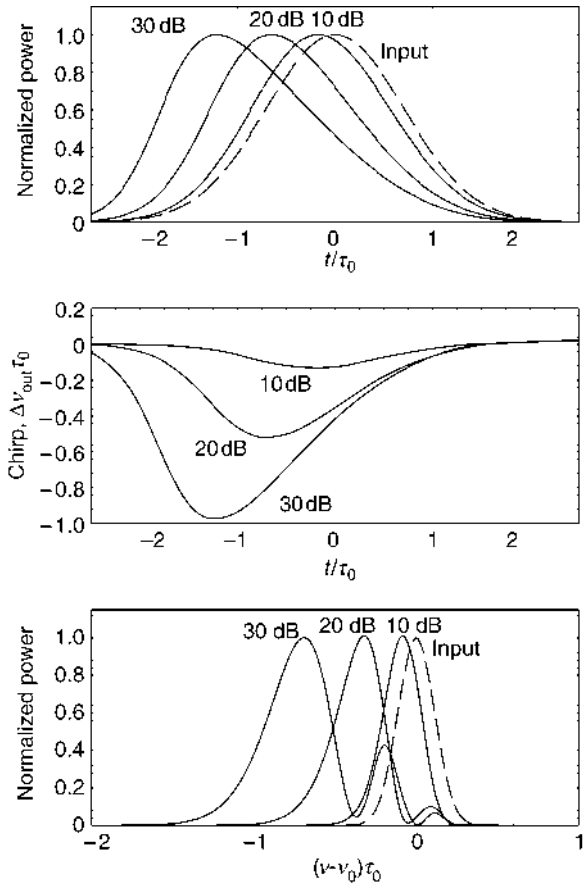

Figure 9 Amplified pulse shape, frequency chirp and spectrum for a transform-limited Gaussian input pulse with $E_{\text{in}}/E_{\text{sat}} = 0.1$ and $\alpha = 5$. ν is optical frequency. The parameter is the amplifier unsaturated gain.

Functional Applications

SOAs can be used to perform functions that are particularly useful in optically transparent networks. Recent advances in photonic integrated circuit and optoelectronic device packaging technology have made the deployment of SOA functional elements feasible. Two of the most important applications of SOAs are optical switching and wavelength conversion.

An example of a 2 × 2 SOA switch module used for optical routing is shown in Figure 10. Larger switch matrices can be constructed using this basic element. The module is based on a hybrid structure consisting of four integrated SOAs mounted on a silicon submount. The SOA-array is aligned to input and output polymer waveguides through V-grooves and alignment indentations. An incoming data packet can be routed to any output port by switching on the appropriate SOA. Switching times of the order of 1 ns are possible.

Ultrafast switching (<100 fs) can be achieved by incorporating SOAs in non-linear loop mirror structures, such as the terahertz optical asymmetric demultiplexer (TOAD) shown in Figure 11. Switching is achieved by placing an SOA offset from the center of an optical fiber loop mirror and injecting data into the loop via a 50:50 coupler. The two

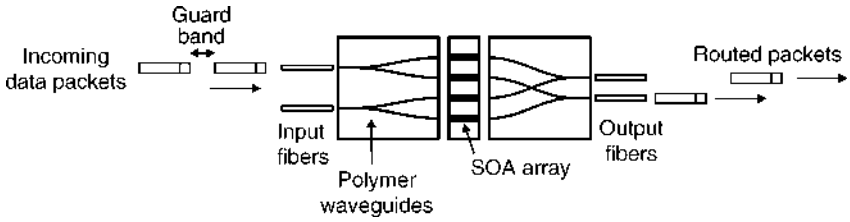

Figure 10 2 × 2 hybrid SOA switch module.

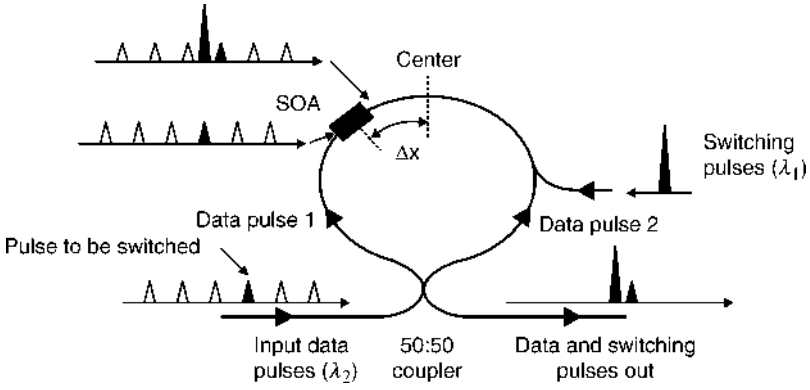

Figure 11 Ultrahigh speed optical switching using a TOAD.

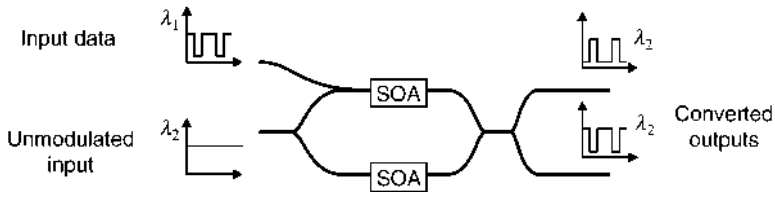

Figure 12 SOA Mach–Zehnder wavelength converter.

counter-propagating data pulse streams arrive asynchronously at the SOA. A switching pulse is timed to arrive after one data pulse but just before its replica. The switching pulse power is adjusted to impart a phase change of π radians onto the replica, so the data pulse is switched out when the two counter-propagating components interfere on their return to the coupler. The TOAD can also be used to demultiplex high-speed time division multiplexed pulse streams.

All-optical wavelength converters will play an important role in broadband optical networks. Their most important function will be to avoid wavelength blocking in optical cross-connects in WDM networks. Wavelength converters increase the flexibility and capacity of a network using a fixed set of wavelengths and can be used to centralize network management. In packet switching networks, tunable wavelength converters can be used to resolve packet contention and reduce optical buffering requirements.

Wavelength conversion in SOAs can be achieved using XGM, cross-phase modulation (XPM), or FWM. The design of an SOA wavelength converter based on XPM requires that one or more SOAs be incorporated into an interferometer, an example of which is the Mach–Zehnder interferometer (MZI) shown in Figure 12. An input data signal at wavelength λ_1 is used to modulate the upper SOA refractive index, which controls the phase shift experienced by a second unmodulated input signal at λ_2 in one of the interferometer arms. When the input data signal is low (logic 0), the arms are in phase and the λ_2 signal appears at the top output. When the input pump signal is high (logic 1) it produces an extra π phase shift between the two arms, causing them to be totally out of phase and the λ_2 signal to appear at the bottom output. Both outputs contain a copy of the original data signal imposed onto the new wavelength. Up and down conversion is possible. The lower SOA is used to equalize the gain experienced by the λ_2 signal in each of the arms. An optical filter is required at the output to remove λ_1. Such wavelength converters can operate at very high bit rates (>10 Gb/s). An important advantage of the MZI structure is that it also gives 2R regeneration (re-amplification and reshaping) of the input data signal with wavelength conversion. This is because the interferometer non-linear response increases the converted signal extinction ratio compared to the input data. All-optical 3R (2R + retiming) regenerators are of great interest in future optical communication networks to restore degraded transmission signals. SOA-based interferometric structures can be used to achieve 3R (2R + retiming) regeneration at data rates >80 Gb/s.

See also

Coherent Lightwave Systems. **Lasers:** Semiconductor Lasers. **Optical Amplifiers:** Basic Concepts; Erbrium Doped Fiber Amplifiers for Lightwave Systems; Optical

Amplifiers in Long-Haul Transmission Systems. **Optical Communication Systems:** Optical Time Division Multiplexing; Wavelength Division Multiplexing. **Semiconductor Physics:** Band Structure and Optical Properties.

Further Reading

Agrawal GP and Olsson NA (1989) Self phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers. *IEEE Journal of Quantum Electronics* 25: 2297–2306.

Connelly MJ (2002) *Semiconductor Optical Amplifiers*. Boston, MA: Kluwer Academic.

Desurvire E (1994) *Erbium-Doped Fiber Amplifiers: Principles and Applications*. New York: Wiley.

Durhuus T, Mikkelsen B, Joergensen C, Lykke Danielsen S and Stubkjaer KE (1996) All-optical wavelength conversion by semiconductor optical amplifiers. *Journal of Lightwave Technology* 14: 942–954.

Ghafouri-Shiraz H (1995) *Fundamentals of Laser Diode Amplifiers*. New York: Wiley.

Shimada S and Ishio H (1992) *Optical Amplifiers and their Applications*. New York: Wiley.

OPTICAL COATINGS

Contents

Anti-Counterfeiting and Decorative Coatings
Diamond Optical Devices and Coatings
Laser Damage in Thin Film Coatings
Optical Black Surfaces
Thin Film Optical Coatings
X-Ray Coatings

Anti-Counterfeiting and Decorative Coatings

R W Phillips and R L Bonkowski, Flex Products, Inc., Santa Rosa, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Scientists are well known for the diversity of their interests. One topic that captures the interest of virtually all scientists, and people in general, is money. Not only are people engaged in obtaining money, but they are also fascinated about the manufacture of money and the protection of its value.

The relationship between manufacturing techniques and the protection of value documents (currency, stock shares, bonds, etc.) is not accidental. Technology has always been employed to counter the contemporary threats of counterfeiters. These threats vary, depending on the skill set and resources available to the criminal, as well as the counterfeiting strategy adopted. While we refer to banknotes repeatedly in this chapter, the information applies, as well, to passports, stamps, tax banderoles, identification cards, visas, licenses, checks, and bankcards. Banknotes can be considered the most challenging application for optical anticounterfeiting technologies. Stamps, banderoles, and checks are single-use items, and do not have to be as durable as currency. Identification cards, passports, and the like are used many times but are generally in protective covers and are examined by trained personnel. Currency must endure the rigors of commerce and travel, while still being easily authenticated by the average citizen.

The ever-increasing availability of low-cost computer hardware and software, along with color photocopiers and printers, has stimulated the use of optical and electro-optical technologies that were previously applied in a limited fashion or not at all. These optical techniques include: transmitted light devices, such as watermarks and threads; luminescent materials; reflective, interference, and diffraction schemes; modulation techniques (Moiré, for example); and light polarization effects.

Each of these concepts either has been or is likely to be applied to value documents to prevent counterfeiting. Each has its own set of potential benefits and

application challenges. Also, some of these effects have been successfully applied to commercial decorative products. It is important to understand the difference between commercial and security uses. Effects aimed at commercial uses are mainly decorative or aesthetic, whereas security devices are functional and integrated into a security strategy. To ensure economic success, decorative products must be easily manufactured and widely distributed. In contrast, security products are carefully controlled in both their availability and the diffusion of manufacturing expertise.

Historical Perspective

Paper currencies (banknotes) have been printed for hundreds of years. Their history has been marked by a succession of technical changes in an effort to prevent counterfeiting. Sophisticated manufacturing techniques, such as intaglio printing and watermarks, along with special papers, made it difficult for counterfeiters to produce acceptable imitations. Intaglio is the centuries old method of printing, using incised or engraved metal plates. The tremendous pressure of the intaglio press forces the paper into an ink-coated plate engraving, forming patterns of ink on the paper.

In the eighteenth and nineteenth centuries, banknote printers often used unique engravings of guilloche patterns (intricate, overlapping curves), vignettes, and portraits as anticounterfeiting devices. These sophisticated patterns effectively discouraged early, relatively unskilled, counterfeiters. The development of photography and lithographic printing in the late nineteenth century created new opportunities for counterfeiting. Photochemical technology made it possible to produce suitable reproductions using print shop facilities throughout the world.

Photochemical techniques had significant limitations, however. The printed notes were passable but not high-quality imitations of genuine intaglio notes. Making lithographic plates required specific skill, considerable equipment, and left behind meaningful evidence (films, plates, press blankets, etc.) for police to find.

Until the 1980s, counterfeiting detection methods by governmental officials were relatively uncomplicated, as three general perpetrators could be classified in fairly predictable ways (Table 1).

Until the advent of home computers, optical scanners, color printers, and copiers, the challenge for law enforcement was difficult but focused. Amateur counterfeiting was virtually nonexistent. The effort required to make a few notes was daunting, the risk high, and potential profit meager. This meant that anticounterfeiting police efforts could be concentrated on criminal counterfeiters. Government-sponsored counterfeiting was certainly known (both the Germans and the British attempted to de-stabilize their opponents' national economies during World War II), but this was more an issue of foreign policy, diplomacy, and espionage and not the main province of national police.

The situation changed quickly and radically in the late 1980s. Color photocopiers had improved in cost, quality, and size to the point that commercial establishments were buying hundreds of machines. The 1990s introduced numerous new technologies to seduce the casual counterfeiter. These included home computers, graphics editing software, color printers, and low-cost scanners.

The amateur counterfeiter now had an array of sophisticated tools that could copy, edit, and reproduce virtually any kind of document. Banknotes made by the techniques of 'digifeiting' (a term coined by Dr. Sara Church of the US Treasury Department) were not of high quality, but passable in small quantities in low risk situations. It was now possible, economical, and relatively safe to make one or a dozen bills, especially in low denominations.

Police efforts were now severely challenged as the proliferation of amateur counterfeiters made investigation and prosecution costly. Especially in the United States, average citizens were lured by the prospect of augmenting their incomes by printing one or two notes a week. When apprehended, effective prosecution and punishment of such small-stakes criminals was expensive and unpopular.

Large-production criminal counterfeiting was augmented by new digital and electro-optical

Table 1

Class of threat	*Typical volume of notes*	*Resources and skills (1980)*
Amateur	Small (<100)	Retail paper, ink. Household chemistry
Criminal	Medium to large (1000 to 1 000 000)	Litho press, darkroom, plate making, cutting equipment. Skilled printing personnel
State-sponsored	Large (millions)	Intaglio press, currency paper. Skilled engravers and printers, Major distribution networks

breakthroughs. Lithographic plates could now be made without cameras or chemistry. Laser ablated plates made directly on the printing press (such as Heidelberg/Presstek) accurately duplicated features as small as 10 or 15 microns. Further, computer-to-plate technology with erasable plates presented an unusual opportunity for counterfeit production in small press shops.

The use of new electronic and optical tools by amateur and criminal counterfeiters coincided with a rise in state-sponsored counterfeiting. The United States, in particular, was victimized in the early 1990s by high-quality, intaglio-printed notes. One such note, dubbed the 'super $100' was allegedly produced in huge volumes in the Middle East and circulated widely in Europe.

Various means of detecting counterfeits were promoted but remain largely ineffective today. The so-called 'detector pen' applies a small amount of iodine solution onto a banknote. The iodine reacts with starch (often used to stiffen paper to give it a banknote 'feel'), turning the pen mark from light brown to black. Unfortunately, the presence or absence of starch in the paper is not a foolproof criterion. US currency paper contains no starch, while the currencies of other countries, such as Canada, have a starch component.

Concurrent with the availability of new high-tech counterfeiting tools, optical science has been hard at work to create protection schemes, utilizing the latest optical effects and manufacturing technologies. These new technologies have been largely outside the skill set and resources of amateur, criminal, and even state-sponsored counterfeiters.

An analysis of all the types of optical materials used for anticounterfeiting purposes as well as for decorative purposes shows that all are based on the control of electromagnetic radiation in the ultraviolet (UV), visible, and infrared. Exploiting various degrees of light transmission, reflectance, absorption, or scatter controls this radiation. In particular, all anticounterfeit devices fall into one or more of five basic categories: light patterning, light re-emission, light diffraction, polarization, or light interference. Characteristics of an effective anticounterfeiting solution include:

- quick verification by the average person;
- ability to function in a variety of lighting situations;
- deterrence of scanning and electronic photocopying;
- compatibility with existing banknote production schemes and equipment;
- ability to be verified by machines;
- durable and reliable despite strenuous handling conditions;
- tight control of materials and manufacturing processes.

Devices Based on Light Patterning

Watermarks

Images can be created in the paper substrate during the manufacturing process. Watermarks have been in use for at least 500 years and are produced by the arrangement of the paper fibers in the cylinder-mold paper machine or by impression in wet paper in the Fourdrinier paper machine. Because they are not on the surface of a document, watermarks are not readily visible in reflected light, but become apparent when light is transmitted through the paper substrate. Credible counterfeits, however, have been produced using images created by imprinting paper with waxy or oily substances. The fatty 'ink' is not readily detected on the surface of the document and serves to change the light transmission characteristics of the paper.

Watermarks serve chiefly to authenticate the paper substrate and to deter the practice of 'raising' notes, that is, bleaching a genuine low-denomination note and re-printing it as a higher value. The most effective watermarks are produced by the cylinder-mold process and use an image that is a duplicate of that image printed in ink (Figure 1) elsewhere on the document. This gives the examiner a reference point to compare the watermark as, for example, on all high-denomination US currency.

Watermarks are an effective counterfeiting deterrent when the general population is educated about their presence and their function. The limited opportunity to expose banknotes to a source of transmitted light (outdoors or in a dimly lit restaurant, for example) somewhat reduces their effectiveness.

Figure 1 Watermark portrait of Clara Schumann, famous pianist and wife of composer Schumann: far left – portrait in transmitted light; center – portrait in reflected and transmitted light; and right – banknote in reflection. Courtesy of Papierfabrik Louisenthal.

Threads

Thin strips of polymer or metallized polymer can be imbedded in the paper substrate during the paper-making process. Such strips are approximately 12 microns thick and 1–5 millimeters wide. These strips are known as threads because Crane & Company developed the original concept in the 1860s, using colored silk thread. Typically, threads are installed border-to-border in the narrow direction of the banknote. Because they are inside the paper, they are not seen in reflected light, but are readily visible when light is transmitted through the substrate. Such threads are effective deterrents against counterfeiting achieved by scanning or photocopying, as the scanner uses only reflected light and cannot duplicate the buried thread. Threads have evolved into several varieties:

Figure 2 Windowed thread. Reproduced with permission from De La Rue International Limited.

Figure 3 Fluorescent thread embedded into US $100 note. Reproduced with permission from Crane and Co., Inc.

- De-metallized threads: Thin polymer coated with an opaque metal. The metal is selectively removed to create text or images.
- 'Window' threads: Metallized polymer that is 'woven' in and out of the paper to permit observation in both reflected and transmitted light (Figure 2).
- Fluorescent threads: Polymer impregnated with ultraviolet inks that emit a specific color under UV irradiation (Figure 3). See the section 'Devices Based on Light Re-emission' later in this article.
- Diffractive threads: Metallized polymer strips that have been embossed with diffractive or holographic patterns and attached by hot-stamping or adhesive to the surface of the note. See the section 'Devices Based on Light Diffraction.'

Potential drawbacks to the use of threads are durability and the temptation of the general public to try to extract them from the banknote. Also, some threads can be simulated by printing or by attaching a polymer film to the surface of the banknote.

Laser Perforation

A modern 'watermark' concept is laser perforation of the banknote substrate. Precise patterns of microscopic holes (ca. 100 microns diameter) can be drilled using a high-power laser apparatus. The 200-Franc Swiss banknote features denomination numbers that are not readily visible in reflected light. However, when a light source is held behind the note, the numerals fabricated by laser perforation are clearly visible (Figure 4). In addition to being easily authenticated by non-expert observers with no special viewing apparatus, laser perforation has forensic advantages. The penetrations made by laser beams have a particular shape, spacing, and dimension, which are not easily mimicked. By clever adaptation of laser scanning software, holes of different shapes can be made and inter-mixed for additional security complexity. As a result, manual

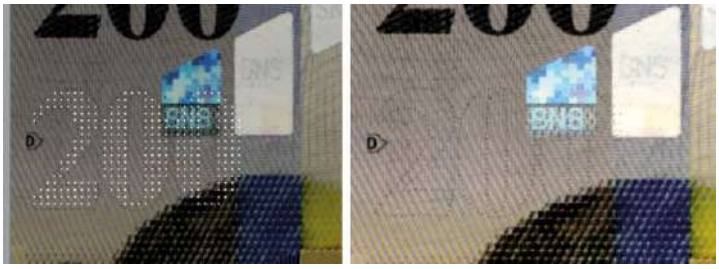

Figure 4 Microperf® laser holes in banknote. Left image in transmitted light; right image in reflected light. Reproduced with permission from Orell Füssli Security Printing Ltd.

pin pricks and mechanically pierced holes can be discriminated as forgeries.

Reflective Foils

One of the earliest defenses against photocopying and digital scanning was the use of reflective metallic foils affixed to the surface of the document. Such foils yield >80% reflection during scanning and the resultant scanned image is usually a black field.

Attachment of foils to currency paper or documents is carried out by applying a thermoset adhesive under heat and pressure and releasing the optical film from its carrier web by means of a separate polymeric release layer. This process is known as hot-stamping. Alternatively, an ultraviolet curing adhesive can be used to affix the foil to the paper substrate using a cold roll-on process.

Foiling may impede the amateur counterfeiter; since decorative foiling of stationery and packaging is well known, it presents little challenge to the professional perpetrator.

Planchettes

Planchettes are small flakes or fragments a few millimeters in diameter and 10–25 microns thick. These are often colorful or reflective materials and can feature a variety of effects including iridescence, diffraction, fluorescence, magnetic properties, or thermochromic properties. These fragments are attached to the banknote substrate by an adhesive (Figure 5). Planchettes discourage scanning and photocopying, since the copier cannot duplicate their special properties. Unfortunately, planchettes have been found to be non-durable and easy to simulate. The generally random placement makes it difficult for an examiner to authenticate a genuine note.

Thermochromic Inks

Inks that react to temperature changes have been employed on checks and identity cards. Thermochromic inks fluctuate from one color to another when the temperature of the printed image is raised. While it is possible for body temperature to activate this change, the effect is more certain if infrared heat is used. Since the process requires several seconds to activate (depending on the heat source), thermochromic materials have not been widely used on banknotes where immediate authentication is necessary.

Figure 5 Peruvian 100 Nuevo Soles note showing iridescent planchettes. Reproduced with permission from Flex Products, Inc.

Colored Fibers

Extremely fine fibers (ca. 15 microns diameter) have been incorporated into currency papers for decades. The United States, for example, uses a few hundred grams of red and blue fibers per ton of paper in making currency. While of some forensic interest, the amount and size of the fibers makes them nearly undetectable by the average citizen.

Alternatively, fibers may be impregnated with fluorescent materials, or with absorbers that selectively absorb the wavelengths of light typically used in copy machines, producing obvious forgeries when copied.

Moiré Effects

Images can be created with very fine details that produce aliasing errors when digitized. Aliasing is the creation of new and spurious patterns through the superposition of two patterns whose periodicities are unequal. It represents interaction between the spatial frequency of the genuine document and the sampling frequency of the copying device. Aliasing is commonly known as moiré in the printing industry. Typical moiré generating patterns are constructed as closely spaced concentric circles, often found around the eyes in portraits on banknotes. When scanned digitally, the resulting copy will contain spurious images (scan traps) that are easily identified as fraudulent.

For example, if a genuine document is designed in a 1000 dpi (dots per inch) printing process it can later be scanned on a 1000 dpi scanner and, theoretically, all the data can be reproduced accurately. If, however, the scanner and the document are not in synchrony, the scanner will miss pieces of data, depending on the registration error, resulting in a digitized image that does not resemble the original.

In the scanned section of a US $5 note (Figure 6), one can see patterns of curved lines intersecting in the background, whereas the authentic image consists of fine parallel lines that do not intersect. The curvature and intersections are artifacts of the moiré created by digital scanning.

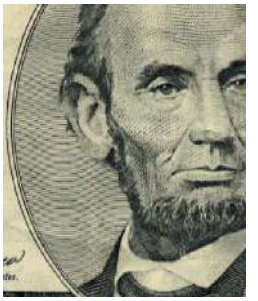

Figure 6 Scanned section of US $5 note showing moiré effect. Authentic image has no intersecting lines. Courtesy of Flex Products, Inc.

Modulation Techniques

Images produced by offset lithography are based on patterns (screens) of fine dots. By varying dot density, an array of shadings and tones can be created. More progressive modulation techniques, such as amplitude modulation (varying dot sizes) and frequency modulation (varying dot spacing) are often used to produce high-resolution images. To deter counterfeiting, banknotes can be printed using pixels, which are lines of equal length but assorted angular orientation. This technique is known as screen angle modulation (SAM).

SAM techniques can be used to defeat digital scanners by preventing the scanning of all available information due to under-sampling, or by concealing hidden patterns which are not demodulated by the human eye, but can be detected with suitable filters (Figure 7). A variant of this technology, called Scrambled Indicia®, has been used to protect US postage stamps (Figure 8).

Devices Based on Light Re-emission

Ultraviolet Fluorescent Inks

Ultraviolet (UV) inks are widely used on banknotes as semicovert deterrents. Images printed in UV inks are challenging to detect in normal lighting situations but are readily visible under UV (black light) illumination. Such printed images require a special light source for authentication and are subject to degradation over time from the handling and chemical exposure of banknotes. Typically, these products are sensitive to either short UV at 254 nanometers or long UV at 366 nanometers.

UV inks provide good protection against photocopying but can be readily simulated by more advanced counterfeiters.

Figure 7 Top image – hidden image as normally seen; Bottom image – Image decoded with visible line screen. Courtesy of Joh Enschedé Security Printing Ltd.

Figure 8 Hidden image in 32¢ US postage stamp decoded by invisible line screen (left image – normal view; right image – with decoder in place). Courtesy of Flex Products, Inc.

UV Fibers

Several countries mix minute quantities of UV luminescent fibers into the paper substrate. In most cases, these are randomly placed and invisible to the naked eye.

Devices Based on Light Diffraction

Diffractive structures range from simple periodic surface relief structures (textured surfaces) to multilevel holographic structures that form a continuum of products ranging from decorative coatings to anticounterfeit devices. Figure 9 shows a generalized structure for all known types of reflective-based relief structures.

Most devices are based on reflection; however, in a few instances where clear polymer is used, devices may also be based on diffractive light transmission. The sizes, shapes, and orientation of the grooves vary according to the exact device. The opaque aluminum layer insures that the design works in reflected light. Most of these structures exhibit a rainbow effect where the color moves from short to long wavelength as the viewing angle increases. This rainbow effect can be found for all types of structures with grooves that have sizes near that of light wavelengths. The color shifts in a direction opposite to that of a thin film interference structure. Further, the rainbow effect can only be seen in a direction perpendicular to the groove direction. In the special case where the groove dimensions are less than the wavelength of light, the

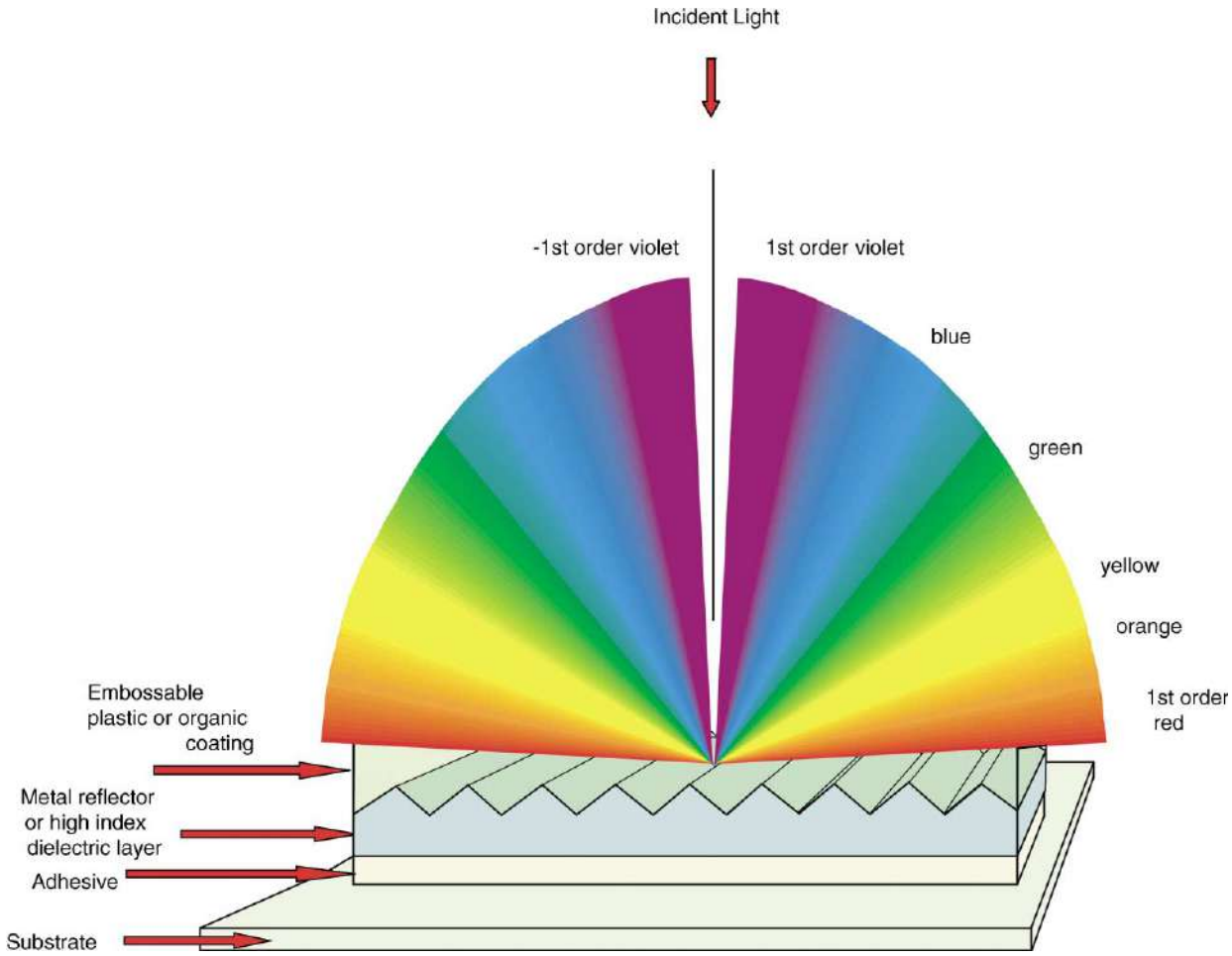

Figure 9 Cross-section of diffraction device. Courtesy of Flex Products, Inc.

rainbow effects disappear and are replaced by another type of diffraction (zero-order diffraction), where only specific colors are seen in reflection. Groove height and width for all diffraction products range from 0.1 to 10 microns.

For a unique set of discrete angles and for a given spacing d, the diffracted light from each facet is in phase with the light diffracted from any other facet, so they combine constructively:

$$Gm\lambda = \sin\alpha + \sin\beta \qquad [1]$$

where $G = 1/d$ is the groove density or pitch, α is the angle between the incident light and the normal to the grating, β is the angle between the diffracted beam and the normal to the grating and m is an integer called the diffraction order.

For $m = 0$, $\beta = -\alpha$ for all λ, the grating acts as a mirror and the wavelengths are no longer separated. This is called specular or zero-order reflection.

In contrast to grating structures (linear grooves) that only modify light intensities, holographic structures are complex surface contours that modify both the phase and intensity of the reflected light. Light reflecting off a metallized holographic structure is capable of producing an image in either two dimensions (2D) or three dimensions (3D). Heat and pressure may emboss the relief structure into a metallized polymeric film by hot-stamping. Alternatively, an embossing metal roller presses the relief image into a thermoplastic film or thin ultraviolet curable coating. The embossed structure is subsequently metallized with opaque aluminum or with a high-index dielectric material such as zinc sulfide or zirconium dioxide. The latter layers allow printed information on the substrate to show, while at the same time provide a holographic image. However, such high-index holograms are not as bright, since some of the reflective light is lost in transmission. Polymethylmethacralate (PMMA), polycarbonate (PC), polyvinyl chloride (PVC), and polyester (type G) are typical materials that can accept such an embossing.

Thick metallized embossed plastic films of ca. 50 microns are used for decorative labels. Thin metallized embossed hardcoats of ca. 7 microns, which can be released from carrier sheets or embossed thin plastic films, are used for the security field.

Holograms used for high-security applications have one characteristic that distinguishes them from ordinary holograms or other diffractive devices used for decorative purposes. This characteristic finds itself in the complication of images, multilevel images at different focal depths, different images that appear at various viewing angles, images of 3D objects superimposed with synthetic 2D images, micro-images, or images that appear to move as the device is tilted back and forth. Specific types of diffractive structures are described below. The list starts with simple holograms used in decorative markets and progresses to high-security devices.

Diffractive grating: A structure comprised of square or sinusoidal linear grooves that breaks light into its component colors. An assemblage of abutting islands, each randomly oriented and with its own parallel grooves, gives the viewer a twinkling color effect (Figure 10).

2D hologram: A relief structure that, upon viewing with a point source of light, shows a 2D image of objects, text, logos, or lines that emerge over a background. Such holograms are often referred to as 'cartoon images' since they are images of original drawings (Figure 11). Because of their simplicity and frequent use as decorative packaging, these holograms are generally not used in the security market.

Multilevel 2D hologram: A relief structure that shows a simulated 3D effect. This is accomplished by creating an image with two or more layers of flat artwork, with each layer at a slightly different focal depth (Figure 12). These images are well suited for labels, cards, and other decorative products. They have also been used on coupons, gift certificates, and tickets.

3D hologram: A relief structure that reconstructs a 3D image from a real or computer-synthesized 3D object (Figure 13). A 3D dove has been used on the VISA credit card as a security device for many years, though many now consider it more of a marketing logo than a security feature.

2D/3D hologram: A relief structure that reconstructs an image showing both a 3D and a 2D object, created from a 3D sculpture and flat art drawings (Figure 14). The artwork is created in such a way that the images appear at different depths of view.

Stereo-hologram: A relief structure that reconstructs a 3D image such that upon tilting from side to side one can see around the object, and additional features can then be viewed. At any given angle, it appears to be a 3D hologram. The image is created through multiple exposures of a 3D object during the holographic mastering process.

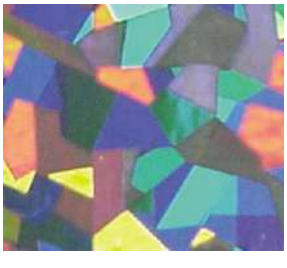

Figure 10 Picture of a typical diffractive assemblage of gratings. Reproduced with permission from Amagic Holographics Inc./K-Laser Technology, Inc.

Figure 12 Multilevel, 2D hologram. Courtesy of American Bank Note Holographics, Inc.

Figure 11 2D hologram. Courtesy of American Bank Note Holographics, Inc.

Figure 13 3D hologram. Courtesy of American Bank Note Holographics, Inc.

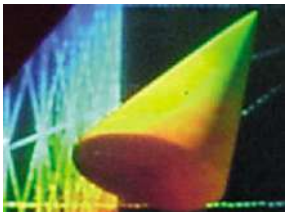

Figure 14 2D/3D hologram. Courtesy of American Bank Note Holographics, Inc.

Figure 15 True color hologram. Reproduced with permission from Light Impressions International, Inc.

Figure 16 High-definition security hologram. Number 50 appears in various colors in first angle of view. Colored shells appear at another angle of view. Reproduced with permission from DeLaRue International Limited.

Figure 17 Left and center images – Exelgram™ at two different viewing angles. Reproduced with permission from CSIRO. Right image – HiMax® OVD high security device showing portrait of Sir Isaac Newton. Reproduced with permission from Light Impressions International, Inc.

True color hologram: In contrast to the holograms described above, a true color hologram is created by diffraction angle control of the colors red, green, and blue from a 3D object or color transparency. The relief structure, when viewed in white light, reveals a picture in true color, as opposed to the rainbow effect seen in regular 2D and 3D holograms (Figure 15).

High-definition security hologram: A relief structure that produces an image composed of multiple elements, such as fine line text and complex patterns that simulate movement as the device is tilted back and forth. Different images also appear at different angles (Figure 16). These images occur because different images are recorded in the holographic mastering process at different angles. The complicated imagery and the use of multiple images is the major holographic technique in use today for anticounterfeiting.

High-definition diffraction gratings: The relief structure in Figure 17 shows a multitude of elements, each with colors that change upon tilting. It consists of a very fine line grating structure, 1000 lines or greater per millimeter. These gratings are sometimes combined with holographic images that change depictions upon tilting, thus providing a high level of security.

Kinegrams®: This grating structure, based on computer-generated graphic images, is composed of light diffracting microstructures of less than one micron, with blaze (saw-tooth) contours that when viewed display optokinematic movements. This security device was first used in 1987 on Austrian 5000 Schillings. When the orientation of the blaze grating changes across the device, two distinct images are possible, depending on the view with respect to the blaze angles. The images do, however, require an aluminum reflective layer deposited onto the relief structure in order to generate the highly reflective colored images.

A 180° rotation of the device can lead to a marked change in the diffraction properties and, consequently, the image properties. This stands in contrast to holographic symmetrically based structures that yield identical images. An example of a Kinegram® security device is shown in Figure 18. The letter 'K' on the front facet of the cube is filled with an asymmetric structure, whereas the background is realized by using the same structure rotated by 180°. When viewed from one side, the letter 'K' appears bright on a dark background. Upon rotation in its plane by 180°, the contrast between the letter 'K' and the background is reversed.

Zero-order diffraction grating: This relief structure is quite different from holographic or normal grating structures, as it is based on an abrupt change in the

Figure 18 The Kinegram® security device as seen in two views differing from each other by a rotation of 180° in its plane, while illumination and viewing conditions remain unchanged. Reproduced with permission from OVD Kinegram Corporation.

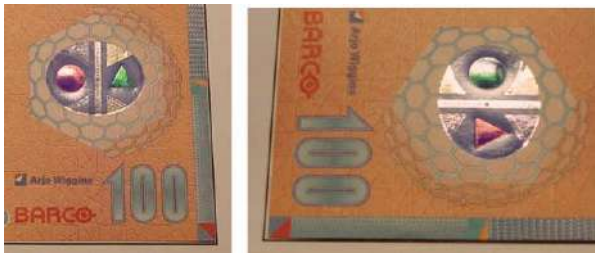

Figure 19 Zero-order diffraction images in a DID® security device: two views 180° apart in planar rotation. Reproduced with permission from Hologram Industries.

refractive index from approximately 2.3 to 1.5. Rather than employing an aluminized layer for reflection, the structure uses a thin, high-index layer ($n = 2.3$, for example) embedded in an organic polymer with a refractive index of about 1.5. The high-index layer appears as a parallel series of lines, approximately 0.2 microns wide, with another parallel series of lines beneath the first set, but oriented under the spacing of the first set of lines at a depth of about 0.2 microns.

This structure not only changes color as the device is tilted back and forth but also as it is rotated. In rotation, the color changes from red to green (Figure 19). Further, the light is polarized as it reflects. When a linear polarizer is inserted between the eye and the device, the color changes to black. The colors seen are based on specular zero-order reflection, unlike standard diffraction gratings where the reflected color is based on first-order reflection. The preeminent effect is when the device is placed over a black background. Over a white background the colors are subtler. Viewing reflection in zero order allows one to see the reflected colors in both point and diffuse lighting. The combination of polarization effects and color in rotation is ideally suited to high-security applications.

Devices Based on Light Interference

Thin-Film Foils

There are two types of thin-film interference devices. The first is based on an all-dielectric filter, and the second is based on a Fabry–Perot resonance cavity structure of the type absorber/dielectric/reflector. The all-dielectric filter uses pairs of high and low refractive index materials at a quarter-wave thickness. A typical filter is the one used initially on Canadian currency that employed the design: ZrO_2 1QW/SiO_2 3QW/ZrO_2 1QW/SiO_2 3QW/ZrO_2 1QW, where the QW (quarter-wave optical thickness) is centered at 585 nm. This design exhibits a gold-to-green color shift with viewing angle.

A Fabry–Perot filter also has a gold-to-green color shift and typically uses the design Cr 5 nm/MgF_2 4QW @ 604 nm/Al (opaque). This filter has slightly higher color saturation than the all-dielectric filter. The design is opaque and, therefore, the color observed is independent of the substrate color onto which it is affixed. This stands in contrast to the all-dielectric filter that must be laid over a black background for optical color saturation. If the all-dielectric filter is placed on a white background, the transmitted light passing through the optical filter will reflect off the white background and combine with the filter's reflection to give white light. However, with a varied colored background, the reflected color will vary depending on the combination of reflected color from the substrate and the reflected color from the filter.

The gamut of reflected colors from a metal–dielectric foil covers all four-color quadrants in the CIE Lab color space. Brightness of the foil depends on the materials used and on the design. Figure 20 shows the theoretical gamut of colors as plotted on the a^* b^* plane for the design, using the materials Cr, MgF_2 and Al.

Foils based on thin film interference layers contain a reflectance peak property moving from long wavelength to short wavelength as the viewing angle is increased. This is true across the UV, through the visible and into the near-infrared. To the human eye, the color moves from red to blue. The anti-counterfeit foil shown in Figure 21 exemplifies this effect.

Shifts from long to short wavelength occur because the path light difference (see the bold line in Figure 22) between the rays of light that reflect from the surface and the reflector are a function of the cosine of the incidence angle.

The path length difference is shown by the term $2d \cos \theta$, where θ is the angle of incidence within the

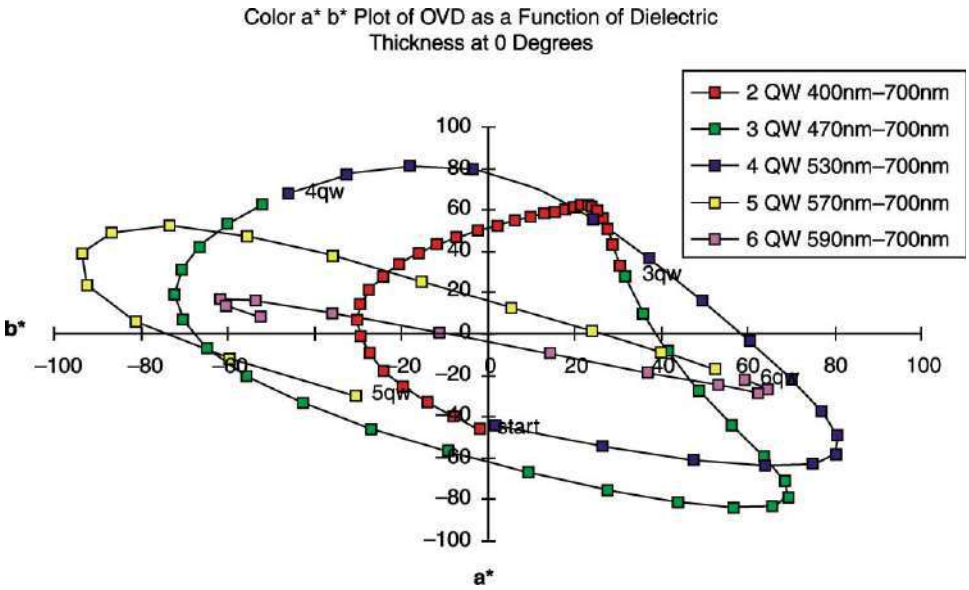

Figure 20 Gamut of colors at 0° incidence viewing under Illuminant C produced from Fabry–Perot designs as a function of quarter-wave optical thickness. Courtesy of Flex Products, Inc.

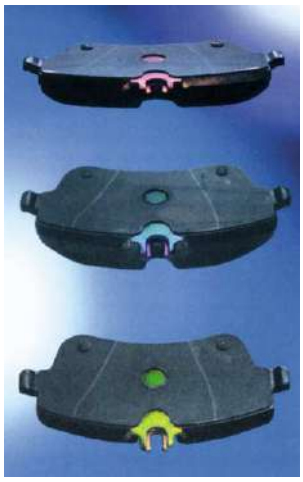

Figure 21 Anticounterfeit interference foil on brake shoes. Reproduced with permission from November AG.

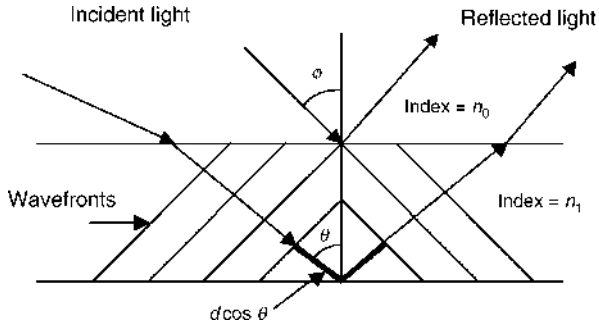

Figure 22 Diagram showing the principle of optical shift with viewing angle. Courtesy of Flex Products, Inc.

filter and d is the distance between the surface of the filter and the reflector. The angle θ and the angle ϕ are related by Snell's law: $n_0 \sin \phi = n_1 \sin \theta$.

Light interference occurs where the wavefronts superimpose on one another. The fractional number of wavelengths that can fit into this path length difference is given by $\Psi = 2n_1 d \cos(\theta)/\lambda$, where λ is the wavelength of light. Waves reflected from these two surfaces will interfere constructively when Ψ is an integral number of wavelengths and interfere destructively when Ψ is at multiples of half wavelengths. As the viewing angle (ϕ) increases, the angle θ also increases. Thus, constructive interference translates into a reflectance peak that moves toward shorter wavelengths as the angle of viewing is increased due to its cosine dependence.

The Fabry–Perot filter can use any transparent dielectric (colored or clear), a reflective metal, although optical metals are preferable since they have the highest reflectance (e.g., Ag, Al, Ni, Cu), and an absorber that is partially transmissive and partially reflective. Thin metals, absorbing dielectrics, and metal nitrides are typical materials used as an absorbing layer.

Light Interference Inks

Inks based on interference pigments have been developed over the last thirty years, starting with pigments based on high-index thin film coated micas (typically TiO_2 coated micas) and culminating with

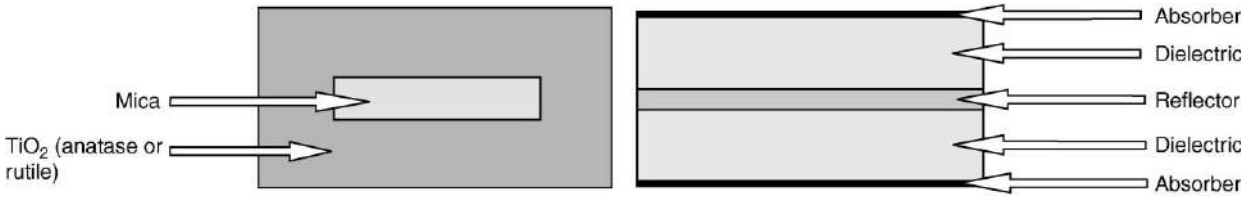

Figure 23 Left image – cross-section of interference mica; right image – cross-section of optically variable pigment (OVP®). Courtesy of Flex Products, Inc.

OVP® pigments based on the Fabry–Perot structure (Figure 23).

Inks created from micas are known as pastel or pearlescent inks, whereas inks formulated from OVP® are known as optically variable inks (OVI®). The latter inks are used as a primary defense against counterfeiting, defeating color copying and color printing from computer images. The difficulty with using interference micas for security features is that they can be readily purchased in the commercial world. In contrast, OVP® is a tightly controlled material and cannot be purchased by the general public.

OVI® security ink has been formulated into intaglio, silkscreen, flexographic and gravure applications. Its use on security documents encompasses over 75 banknotes worldwide, including the newly issued Euro notes, post-1996 $10, $20, $50, and $100 US banknotes, and banknotes for China. With the use of such inks, postage stamps, tax stamps, identification cards, and other security documents have also been effectively protected against counterfeiting. Color control for OVI® is assured because OVP® can be blended according to additive color theory to produce a desired color and color shift. Two examples of OVI® – the 100 Pula Botswana note and the new 100 Euro note – can be seen respectively in Figures 24 and 25.

OVI® inks have been extremely successful for banknote protection and might even be a *de facto* standard, because they meet the seven characteristics of an effective anticounterfeit solution as described earlier.

Figure 24 100 Pula Botswana note showing green-to-blue color shift OVI® with latent images. Reproduced with permission from SICPA.

Figure 25 100 Euro note showing magenta-to-green OVI®. Courtesy of Flex Products, Inc.

Devices Based on Light Polarization

The use of light polarization as an anticounterfeit feature has taken several avenues, one as coatings or pigments, the other as invisible images that are detected with a linear polarizer. With the exception of the zero-order diffraction device described earlier, polarization devices are based on cholesteric liquid crystals, using either nematic (planar oriented) layers or helical twisted layers.

Liquid crystal pigments are constructed from cholesteric polysilozanes that have been oriented and cross-linked in place with photopolymerization. With this cross-linking, pigments are relatively stable to temperature changes. These materials can form helical structures in either right or left twist orientation. Such materials reflect 50% of a band of wavelengths, centered at wavelength, λ_0, of the incoming light as circularly polarized light and 50% of the band centered at λ_0 as transmitted, circular polarized light, but with the opposite twist. The remaining wavelengths pass through the structure with no interaction. The amount and type of the

chiral phase (components that inherently orient into a helical structure) in the polymer determine the pitch of the helical structure. This, in turn, determines the wavelength of reflection, i.e. $\lambda_0 = nP$ where n is the average refractive index and P is the pitch length of the material (pitch length determined by 360° rotation of the helix). At tilted viewing angles, the color changes according to the formula $\lambda = \lambda_0 \cos\theta$ where θ is the viewing angle relative to the vertical, and λ_0 is the reflected wavelength seen perpendicular to the surface. By the judicious combination of left- and right-handed helical based layers, or by inserting a half-wave retarder layer between two layers having the same helical structure, the resulting pigments reflect nearly 100% at λ_0.

When these materials are properly processed and ground (particles 2–50 microns thick, with an aspect ratio of 3:1 in lateral/thickness dimensions), particles with the helical structure perpendicular to the planar surfaces can be realized. Such pigments, or liquid crystal polymeric coatings, when printed, have color-shifting properties with viewing angles that defeat color copiers and the like, since toners have no color-shifting properties (Figure 26). These pigments also have an additional security feature, namely, that when viewed through a circular polarizer having the opposite twist, the color-shifting pigments appear black.

Figure 26 Anticounterfeit label showing liquid crystal coating viewed straight-on and at tilt angles. Courtesy of Flex Products Inc.

A special technology known as linear photo-polymerization (LPP) is a process to produce photo-patterned liquid crystal polymer (LCP) $\lambda/4$ phase retarders, which, when combined with a series of layers involving a reflector, and a linear polarizer, shows positive and negative images. Without the diagnostic linear polarizer, however, the image is invisible (Figure 27).

Adding a nematic liquid crystal layer, which exhibits a helical twist configuration, extends this technology. The latter technology can produce colored images when viewed with a linear polarizer. Such devices can be coated or hot-stamped onto a substrate containing viewable information.

Combination Devices

Beginning in 2000, combinations of holographic structures with liquid crystal or interference thin films were developed, which could make counterfeiting even more difficult. Such structures emerge as a hologram floating on a color shifting background. Magnetic layers, integrated into part of the optical interference stack, allow spatial orientation and 3D effects during printing.

Decorative Coatings

Coatings Based on Diffractive Pigments

When pigments based on diffractive effects are created, individual flakes have surface contours of groove-like structures of 2000–3500 lines mm^{-1}. In spray-out coatings using such pigments, diverse interplay of subtle rainbow colors results (Figure 28). The effect is particularly striking in full sunlight as, for example, on curved surfaces found on cars and motorcycles. Some manufacturers have successfully used finely shredded diffraction gratings

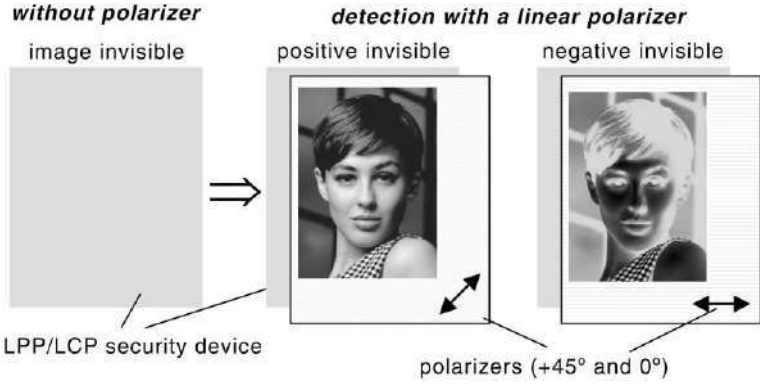

Figure 27 Reflective optical LPP/LCP security device showing a complex LPP/LCP retarder image visualized with one polarizer; rotation of the polarizer changes the image from positive to negative. Reproduced with permission from Rolic Research, Ltd.

Figure 28 Motorcycle painted with diffraction-based pigments. Courtesy of Flex Products, Inc.

(ca. 25 microns thick) as flakes mixed in paints for decorative purposes. The use of such thick flakes seriously limits the applicability and uniformity of the coating.

Coatings Based on Interference Pigments

Pigment-based interference effects are widely known and have been used for decades in the paint, plastics, and cosmetic industries. These structures are based on TiO_2 coated mica, where the mica thickness is between 300 and 600 nm with planar diameters ranging from 1 to 200 microns, depending on the application.

Depending on the thickness of the TiO_2, different reflective colors of silver, gold, red, blue, and green are achievable. For maximum color reflection, a black substrate or viewing at a particular angle is required so as to prevent dilution of the reflected color from light reflected from the substrate surface. Colored mica pigments are platelets of mica coated with colored oxides. Metallic-like colors are also formed from iron oxide coated micas. In the first instance, two colors may be seen: the platelet color (due to absorption) and the interference color at one angle. Only the interference color may be seen at another viewing angle. If the substrate is colored, the color switches from the interference color to the substrate color as the viewing angle changes. More recently, laminar pigments with an all-dielectric optical stack have been produced with a low-index central core, encapsulated with high-index layers that cause interference.

Interference pigments based on absorber/dielectric/metal/dielectric/absorber have also been introduced into the decorative marketplace. These structures are based on aluminum as the central layer, a dielectric spacer layer, and either a wavelength selective absorber layer such as iron oxide (Fe_2O_3) or a nonwavelength selective absorber such as chromium (Cr). In Fe_2O_3 based pigments, the face-on colors are mainly found in the yellows and reds whereas for the Cr-based pigments, a full color gamut is possible.

Diffractive Foils

These foils are based on diffractive and grating surfaces where the contours are generally less than 2000 lines mm^{-1}. These materials show a glittering effect, i.e., multiple points of color, as the viewing angle is changed.

Interference Polymeric Films

These films are made from multilayer extrusions of pairs of polymers of 50 or more layers. Each pair is composed of two different polymers with a slightly different refractive index. Each layer is approximately one-quarter wave in thickness. Due to the small refractive index differences between layers, the resultant film is approximately 25 microns thick. Uniform color at a given viewing angle is generally not possible due to the lack of thickness control. Hence, they tend to look like oil slicks, with each color changing as the viewing angle changes (Figure 29). Interference polymeric films have found use as decorative gift wrapping material.

Holographic Foils

As indicated earlier, holograms designed for non-security markets may range from simple 2D holograms to more complex ones that function for either decorative or brand protection purposes. Another application of holographic foils is in holographic fabrics, whereby small holograms approximating several millimeters in diameter are affixed in a pattern to the cloth (Figure 30).

Aluminum Flake Based Inks and Coatings

These inks are composed of leafing or nonleafing aluminum powders and a polymeric vehicle. The use of leafing pigments produces images and coatings of high brilliance, whereas nonleafing pigments produce a matte or satin-like finish.

Figure 29 Visual effect from multilayer extruded polymeric film. Reproduced with permission from Englehard Corporation.

Summary and Challenges for the Future

Optical scientists have shown tremendous creativity in discovering and employing clever technical devices in the fight against counterfeiting of valuable documents. Optical solutions have been created that meet the functional as well as the economic needs of the document security industry. A cadre of research institutions and manufacturing companies has accepted the challenge of protecting the instruments of our economic and physical security.

This vigilance will have to continue in the future. Nearly as fast as new defenses are developed, criminals are illicitly using new tools to defeat the protection schemes. Financing for sophisticated counterfeiting efforts is likely to increase, whether sourced by international terrorists or drug dealers. Likewise, the more we move toward an information society, the more we depend on machines to count, sort, and verify documents and money. This machine-driven trend will mandate new features, especially in banknotes and cards.

Figure 30 Fabrics showing holographic effects. Reproduced with permission from Amagic Holographics, Inc./K-Laser Technology, Inc.

It is likely that the future of optical anticounterfeiting technology will be strongly influenced by:

- Active, rather than passive, devices. That is, devices that can be applied to a document and later stimulated or interrogated by a machine to elicit a specific effect or signal.
- Schemes that may be merged with new biometric technologies, especially for individualized documents, such as passports and identification cards. In such applications, it will be necessary to confirm that a document is genuine, and that it is in the possession of its rightful owner.
- The need to clearly differentiate security applications from decorative or aesthetic uses. Effective security products must always be carefully controlled and their availability limited.
- The willingness of the public to accept regular changes in the appearance of value documents. Currency will have to undergo redesign on a three-to-six year cycle, rather than the long cycles typical in the past.

This chapter was intended only as a survey of the issues involved in anticounterfeiting protection. To learn more about the optical methods discussed here, or to learn more about the science of anticounterfeiting, please consult the "Further Reading" section and the Internet sites listed in Table 2.

Table 2 Internet sites: Banknote Design & Protection

Sponsor	*Web Address*
Amagic Holographics, Inc.	http://www.amagic-usa.com/
American Bank Note Holographics, Inc.	http://www.abholographics.com/
Bank of England	http://www.bankofengland.co.uk/banknotes/index.htm
Bundesdruckerei GmbH	http://www.bundesdruckerei.de/en/products/index.html
Crane & Co.	http://www.crane.com/specialty_papers/default.asp
Engelhard Corporation	http://www.engelhard.com/
European Central Bank	http://www.ecb.int/
Flex Products, Inc.	http://www.ocli.com/markets/flex_products.html
Giesecke & Devrient	http://www.gdm.de/eng/main/home/index.php4
Graphic Security Systems Corp.	http://www.graphicsecurity.com
Hologram Industries	http://www.hologram-industries.com/pages/index.htm
Joh Enschedé	http://www.joh-enschede.nl/divisies/jess-engels.html
K-Laser Technology, Inc.	http://www.klaser.com.tw
Light Impressions International, Inc.	http://www.lightimpressions.com/
November AG	http://www.november-ag.de
OVD Kinegram Corporation	http://www.kinegram.com/
Papierfabrik Louisenthal GmbH	http://www.louisenthal.de/
Rolic Research, Ltd.	http://www.rolic.com/
SICPA Security Inks	http://www.sicpa.com/default.asp?lang=e
Thomas De La Rue	http://www.delarue.com/
United States Bureau of Engraving	http://www.bep.treas.gov/

The authors express their gratitude to Ms. Sue Spadini for the technical editing of this work, and to the many individuals and companies who provided exemplars and illustrations. Please see the figure captions and the reference lists for their names.

List of Units and Nomenclature

2D	two-dimensional
3D	three-dimensional
CIE	Commission Internationale de l'Eclairage; International Commission on Illumination
CIELAB color space	(L^* = Lightness; $a^*.b^*$ = chroma coordinates)
CMYK	cyan-magenta-yellow-black
QW	quarter wave

See also

Holography, Techniques: Color Holography; Holographic Interferometry; Holographic Recording Materials and their Processing; Sandwich Holography and Light in Flight. **Microscopy:** Interference Microscopy. **Optical Coatings:** Thin-film Optical Coatings. **Polarization:** Introduction; Matrix Analysis.

Further Reading

Committee on Next-Generation Currency Design, National Materials Advisory Board (1993) *Counterfeit Deterrent Features for the Next Generation Currency Design*. Washington, DC: National Academic Press.

Demus D, Goodby J, Gray GW, Spiess HW and Vill V (eds) (1998) *Handbook of Liquid Crystals*. Weinheim, Germany: Wiley-VCH.

Fagan WF (1990) *Optical Security and Anticounterfeiting Systems*. Volume 1210. SPIE Conference – The International Society for Optical Engineering. Bellingham, WA.

Faris SM (1994) US Patent 5,364,557: *Aligned Cholesteric Liquid Crystal Inks*.

Glausch R, Kieser M, Maisch R, *et al.* (eds) (1998) *Special Effect Pigments*. Hanover, Germany: Vincentz Verlag.

Hutley MC (1982) *Diffraction Gratings*. London: Academic Press.

Phillips RW and Bonkowski RL (2000) *Security Enhancement of Holograms with Interference Coatings*. Volume 3973. SPIE – The International Society for Optical Engineering. Bellingham, WA.

Phillips RW and Nofi M (1999) *Colors by Chemistry or by Physics?* Society of Vacuum Coaters 42nd Annual Technical Conference Proceedings, pp. 494–499, Chicago, IL.

Phillips RW, Nofi M and Slusser R (1994) *Color Effects From Thin Films Designs*. Eighth International Conference on Vacuum Web Coating, pp. 270–284. Las Vegas, NV.

Proceedings of the International Conference on Security Documents for the 21st Century (1987) San Diego, CA. Prepared by Oak Ridge National Laboratories, CONF-87040175, Contract DE-A-C05-840R2140, Oak Ridge, TN 37831.

Rancourt JD (1987) *Optical Thin Films, Users' Handbook*. New York: Macmillan.

US House of Representatives (1984) *The Currency Design Act*. Hearing Before the Sub-Committee on Banking, Finance, and Urban Affairs, Second Session on H.R. 6005, Serial No. 98-98. Washington, DC: US Government Printing Office.

Van den Assem B, Brongers D, Rath J, *et al.* (1994) *Practical Guide for the Security Documents Printing Sector.* The Hague: SDU Uitgevererij Koninginnegracht.

Van Renesse RL (1994) *Optical Document Security.* London: Artech House.

Van Renesse RL (1996) *Optical Security and Counterfeit Deterrence Techniques*. Volume 2659. SPIE Conference – The International Society for Optical Engineering. Bellingham, WA.

Van Renesse RL (1998) *Optical Security and Counterfeit Deterrence Techniques II*. Volume 3314. SPIE Conference – The International Society for Optical Engineering. Bellingham, WA.

Van Renesse RL and Vliegenthart WA (2000) *Optical Security and Counterfeit Deterrence Techniques III*. Volume 3973. SPIE Conference – The International Society for Optical Engineering. Bellingham, WA.

Diamond Optical Devices and Coatings

D M Aslam, Michigan State University, East Lansing, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

What do a human, a plant, the photoresist used in microfabrication, and a polymer have in common? Why has carbon become a crucial part of all known living systems? The answer to these questions may be embedded in carbon's hybrid bonding structure (sp, sp^2, and sp^3 C–C bonds). Carbon has 4 electrons in its outermost shell (the valence shell), which is the second shell. At first sight, one would expect that these 4 electrons are distributed among its one 2s and three 2p orbitals. However, it is well known that, in the diamond lattice, each carbon atom bonds with its four neighbors. Are the electrons in 2s and 2p orbitals

responsible for the four bonds? One possible answer is that the 2s, $2p_x$, $2p_y$, and $2p_z$ orbitals of the carbon atom form bonds with their neighbors. There are two problems with this answer. First, these bonds will have different strengths (s and p orbitals), which is in contradiction with the experimental fact indicating an equal strength for all the four C–C bonds in the diamond. Second, the computed strength of such bonds will be lower than that found experimentally. What is then the nature of C–C bonds found in organic compounds?

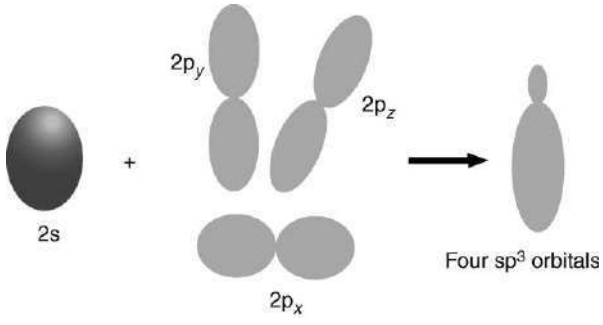

Figure 1 Hybridization of s and p orbitals in carbon to form sp^3 orbitals, which form the so-called sigma bonds responsible for the strong diamond lattice.

The answer to this difficult question was provided in 1931 by Linus Pauling (1901–1994), the winner of two Nobel prizes. With the help of quantum mechanical calculations, Pauling showed how one s orbital and three p orbitals can mix, or hybridize, to form four equivalent atomic orbitals with tetrahedral orientation. The resulting hybrid orbitals are called sp^3 orbitals as shown in Figure 1. The sp^2 or sp^1 orbital results if two or one p orbital combine with the s orbital, respectively, to form the hybrid. While sp^3 C–C bonds lead to a diamond lattice, the sp^2 C–C bonds lead to graphite, and a material which contains both sp^2 and sp^3 C–C bonds is called diamond-like carbon (DLC). The optical properties, such as the bandgap, of DLC depend on the sp^3/sp^2 ratio. If a sheet consisting of a monolayer of graphite, which consists of sp^2 bonds, is wrapped into a pipe, the resulting structure is called a single-wall nanotube (SWNT), as shown in Figure 2a. A multiwall nanotube (MWNT) has many such layers (Figure 2b–c). Other elements such as oxygen (in H_2O) and nitrogen (in NH_3) also form sp^3 orbitals but carbon seems to be only one which is capable of forming sp, sp^2, or sp^3 bonds, depending upon the bonding environment (Figure 3). This may explain why nature has selected carbon

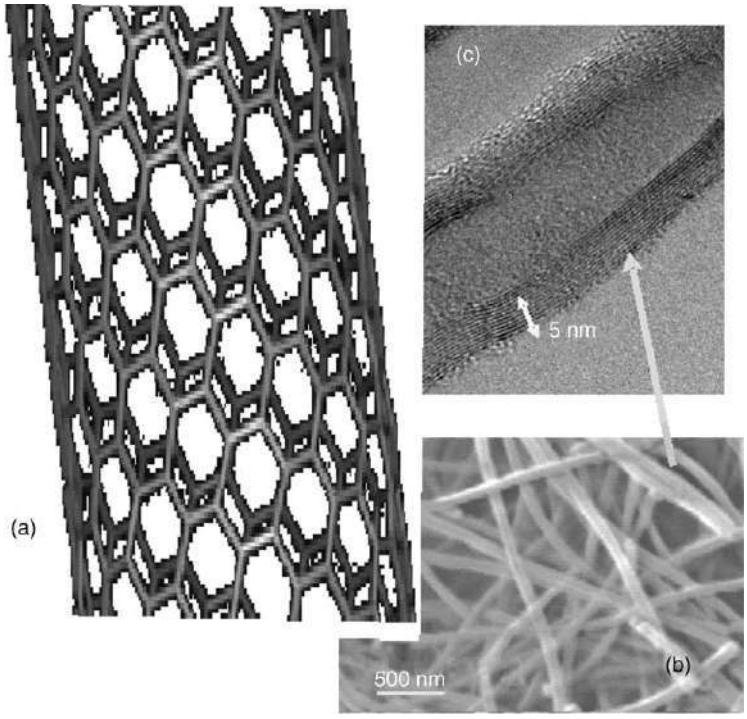

Figure 2 CNT: (a) drawing of a single wall CNT showing sp^2 bonded layer; (b) a mat of multi-wall CNTs; and (c) a TEM of MWNT showing 13 sp^2 bonded layers.

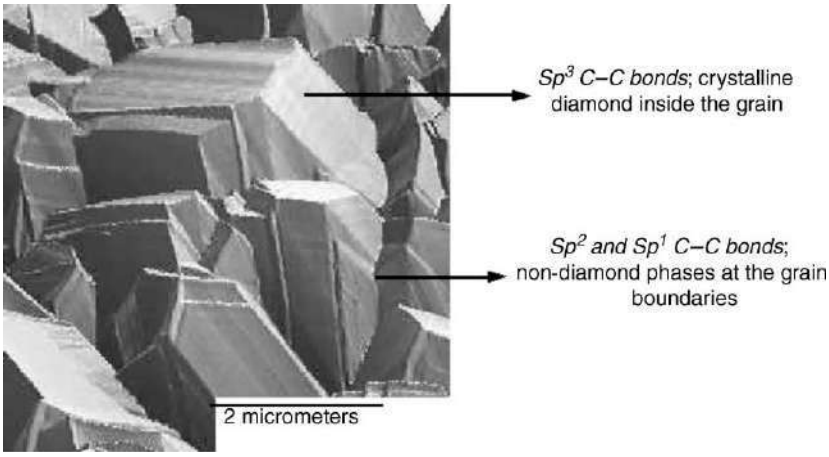

Figure 3 Polycrystalline diamond film; crystalline regions separated by grain boundaries containing non-diamond phases.

Table 1 Comparison of semiconductor properties

Property	*Diamond*		*Si*	*GaA*	*β*-SiC
	Poly-C	*Crystalline*			
Fig. of merit					
Johnson, 10^{23} W Ω s^{-2}		73 856	9	62.5	10 240
Keyes, 10^2 W cm^{-1} s^{-2} °C		444	13.8	6.3	90.3
Sat. electron vel., 10^7 cm s^{-1}		2.7	1	1	2.5
Carrier mobilities, (cm^2 V^{-1} s^{-1})					
Holes	1–165	1600	600	400	50
Electrons		2200	1500	8500	400
Resistivity, Ω cm	10^{-2}–10^8	10^{-3}–10^{13}	10^{-4}–10^8	?–10^8	150
Breakdown field, MV cm^{-1}	0.1–1	10	0.3	6	4
Band gap, eV		5.45	1.12	1.42	3
Dielectric constant	6.7	5.5	11.7	12.5	9.7
Thermal expan. coeff., 10^{-6} °C^{-1}	2.6	1.1	2.6	5.9	4.7
Thermal cond., W cm^{-1} K^{-1}	20	4–22	1.5	0.5	5
Lattice constant, Å		3.57	5.43	5.65	4.36
Density, g/cm^3		3.52	2.32	5.31	3.215
Melting point, °C		4000[a]	1412	1240	2540
Hardness, kg mm^{-2}		10 000	1000	600	3500
Poisson ratio	0.11	0.15	0.23	0.315	
Young's modulus, 10^{12} Pa	0.8–1.2	1.1–1.2	0.155	0.085	0.7

[a]Diamond may convert to graphite well below 4000 °C depending upon the ambient and it oxidizes at 650 °C to form CO_2/CO in O_2.

(perhaps due to its bonding flexibility, to form long C–C chains and rings) as the material of choice for building living systems.

The sp^3 bonding in a diamond lattice makes it a very unique material with a combination of properties not matched by any other known materials (Table 1). In fact, the C–C sp^3 bond in diamond is responsible for its remarkable properties leading to current and potential applications as shown in Figure 4. Although diamond (for centuries) and its bonding structure (for over 70 years) has been known for a long time, the growth of diamond films using chemical vapor deposition (CVD) took a long time to mature. Again, this delay relates to the difficulty with which the diamond lattice with sp^3 C–C bonds is fabricated.

Diamond Growth by CVD

First evidence of diamond growth by CVD, by Eversole in 1952–1953, led to the use of H_2 and CH_4 in the hot filament CVD (HFCVD) to grow diamond on diamond substrates (homoepitaxial growth) by Angus in 1971. The inexpensive CVD polycrystalline diamond (poly-C) was grown on non-diamond substrates by Deryagin in 1976, Spitsyn in 1981, and by Matsumoto *et al.* in 1983. Currently,

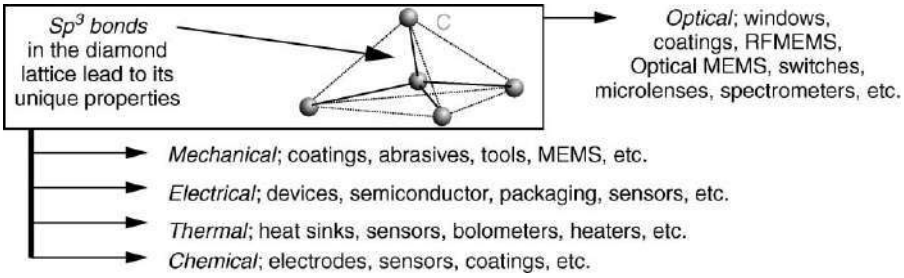

Figure 4 Sp^3 bonds in diamond make it a unique material.

Table 2 Different diamond deposition methods

Methods	*HFCVD*	*MPCVD*	*DC-arc Jet CVD*	*Combustion synthesis*	*Multiple pulsed laser*
Deposition rate (μm/hr)	0.1–10	0.1–10	30–150	4–40	3600
Substrate temperature (°C)	300–1000	300–1200	800–1100	600–1400	50
Deposition area (cm^2)	5–900	5–100	<2	<3	N/A
Advantages	Simple, large area	Quality, stability	High rate, good quality	Simple, high rate	Ultra-high rate
Disadvantages	Contamination, fragile filament	Rate	Contamination, small area	Small area	Expensive

HFCVD, hot filament CVD.

it is believed that, during the CVD of diamond, the CH_3 is responsible for deposition of C as diamond and non-diamond phases. The atomic hydrogen, present in the growth environment, removes the non-diamond phases leaving behind the diamond phase. Near the end of the 1990s, basic science of CVD diamond was well understood, and today diverse plasma and thermal techniques have been developed to produce poly-C films several mm thick and over 12 inches in diameter. Optically smooth 300-micrometer thick undoped poly-C wafers are available on the market. Although there are some reports of n-type poly-C and crystalline diamond growth, the well-established techniques exist only for *in situ* doping of p-type diamond.

A number of diamond growth techniques are currently available for the growth of diamond and carbon nanotubes. Microwave plasma CVD (MPCVD) can be used for the growth of both diamond and carbon nanotubes. As shown in Table 2, the deposition rate of diamond shows a large variation (0.1–3600 micrometers/hour) depending on the growth technique. For applications of diamond in optical coatings, optoelectronic devices and optical systems, the study of transmission, reflection and absorption of light for diamond is very important. These properties of diamond are responsible for its beauty and glamor. For applications in optoelectronic materials, devices and systems, the micro-fabrication and -machining technologies are crucial, particularly in the case of inexpensive poly-C films.

Absorption, Reflection, and Transmission in Diamond

The optical properties, which have been studied for hundreds of years, of naturally occurring single crystal diamond have led to the gem commerce. The fundamental concept involved in the tremendous historical fame of diamond is its high refractive index, n_d (in the range of 2.38–2.64 depending upon the wavelength of light). The refractive index of a material is a measure of decrease of velocity of light (which is also true for all electromagnetic waves) when the wave enters the material. Mathematically, n_d is defined by

$$n_d = V_{vac}/V_d \qquad [1]$$

where V_{vac} and V_d are velocities of light in vacuum and diamond, respectively.

When a beam of light of intensity I_0 in air is incident normally on the surface of diamond, the reflected part of the beam, characterized by the reflection coefficient R, depends on n_d, and R is given by

$$R = \left(\frac{n_a - n_d}{n_a + n_d}\right)^2 \qquad [2]$$

where n_a is refractive index of air and n_d is given by Sellmeier dispersion equation:

$$n_d = \sqrt{1 + \frac{4.3\lambda^2}{\lambda^2 - 0.0121} + \frac{0.33\lambda^2}{\lambda^2 - 0.031}} \quad [3]$$

where λ is the wavelength of light expressed in micrometers. When the beam enters diamond, part of it is absorbed in diamond. The absorption, characterized by the absorption coefficient α, is caused by: (i) impurities (such as nitrogen, boron, hydrogen, etc.); and/or (ii) intrinsic defects (such as dangling bonds, sp^2 bonds, and sp bonds). However, if the energy of light is close to or above the bandgap of diamond (5.5 eV), the absorption can be caused by electronic transitions from valence to conduction bands. The intensity $I(x)$ at a distance x below the diamond surface is given by

$$I(x) = I_0(1 - R)\exp(-\alpha x) \quad [4]$$

The fundamental expressions for α are different for indirect and direct type semiconductors. However, the modeling of α, for energies close to the bandgap of diamond, can be accomplished empirically using Urbach's rule according to the following equation:

$$\alpha = \alpha_0 \exp\left[\frac{\sigma(T)(E - E_0)}{kT}\right] \quad [5]$$

where α_0 and σ are Ubach tail parameters, and E_0 (6.5 eV) is direct gap energy of diamond. K is Boltzmann constant and T is the temperature. As at $T = 300$ K, $\alpha_0 = 4.23 \times 10^{11}$ cm^{-1} and $\sigma = 0.585$, eqn [5] can be simplified:

$$\alpha = 4.23 \times 10^{11}\, e^{22.6(E-6.5)} \quad [6]$$

where the units for α and E are in cm^{-1} and eV, respectively. Using the expression $E = 1.24/\lambda$, where E is in eV and λ is in micrometers, eqn [6] can be written as

$$\alpha = 4.23 \times 10^{11}\, e^{27.9(1/\lambda-5.24)} \quad [7]$$

where λ is in micrometers. The variations of α and n_d with wavelength and energy of light are shown in Figure 5, which are plotted using eqns [3] and [7]. Impurities and intrinsic defects in diamond strongly affect α, n_d, and its other properties.

Impurities in Diamond

A pure diamond with perfect lattice structure is transparent to ultraviolet (UV), visible, infrared (IR), millimeter and microwave regions of the electromagnetic spectrum. However, even an impurity-free diamond has some absorption in the infrared due to phonon excitations and an ultraviolet cutoff due to band-to-band absorption. The defects, impurity-related and intrinsic, in single crystal diamond have been extensively studied because they strongly affect its properties. In fact, diamonds are classified according to the type of defects that are dominant.

Impurities in diamond give rise to additional energy levels in its bandgap. If the density of these impurities is sufficiently high, they can cause absorption of electromagnetic radiation (including light) on one hand and break the lattice symmetry on the other. The breaking of the lattice symmetry can lead to one-phonon IR absorption, which is forbidden in pure diamond. The fact that impurities affect the optical properties of diamond led to its classification into types Ia, Ib, IIa, and IIb which correspond to pale yellow, golden yellow, no color, and blue, respectively. The transmission properties of these various types of diamond are depicted in Figure 6. The dominant impurities are nitrogen

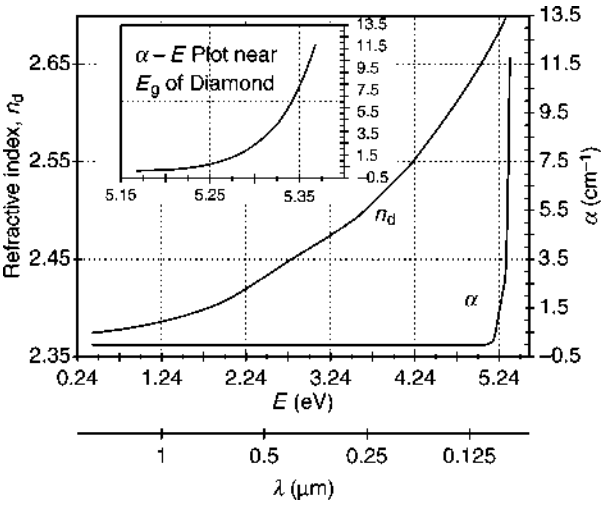

Figure 5 Wavelength and energy dependence of absorption coefficient and refractive index of diamond.

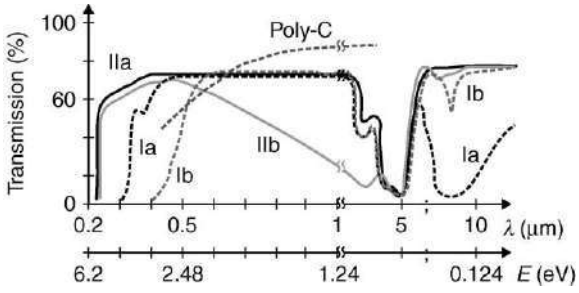

Figure 6 Transmission spectra of natural diamond and CVD polycrystalline diamond (poly-C).

(5–2500 ppm atomic) in type I and boron (10 ppm atomic) in type II. It may be pointed out that boron-doped diamond is a semiconductor with p-type conductivity measured at room temperature. Nitrogen-doped diamond does not show n-type conductivity at room temperature as the donor levels are too deep to be ionized.

Polycrystalline CVD Diamond

Due to the lack of a reliable hetero-epitaxial diamond growth technique, the crystalline diamond films are very expensive (require single crystal diamond substrates for their growth) and, consequently, their widespread commercial use is severely limited. Polycrystalline diamond (poly-C) grown by CVD, which can be as inexpensive as any other material if mass produced, is very attractive for optical and other applications (Figure 2). The properties of high-quality (with high density of sp^3 C–C bonds) poly-C films are comparable to those of crystalline films.

As shown in Figure 6, the transmission properties of poly-C are comparable to or better than those of type I and II diamonds for part of the energy range. Free-standing commercial poly-C films are typically polished and have thicknesses in the range of 50–300 micrometers (the least expensive are 300 μm thick). White poly-C films transparent to visible light are also available. Such films, with optically smooth surface, can be considered free of scattering effects. These films can be considered as a transparent plate in air and, for normal incidence of light with energy less than the bandgap, the transmission coefficient is given by

$$T = \frac{(1 - r_1^2)(1 - r_2^2)}{1 + r_1^2 r_2^2 + 2r_1 r_2 \cos(2\delta)} \qquad [8]$$

where the Fresnel coefficients are given by

$$r_1 = \frac{n_a - n_d}{n_a + n_d}, \qquad r_2 = \frac{n_d - n_a}{n_d + n_a} \qquad [9]$$

where n_a and n_d are refractive indexes of air and poly-C, respectively, and δ is phase change for a poly-C film with a thickness of t:

$$\delta = \frac{2\pi}{\lambda} n_d t \qquad [10]$$

Using $n_a = 1.0003$, $n_d = 2.4$, and $t = 100$ μm, eqn [8] can be written as

$$T = \frac{1}{1.49 - 0.493 \cos\left(\frac{3015}{\lambda}\right)} \qquad [11]$$

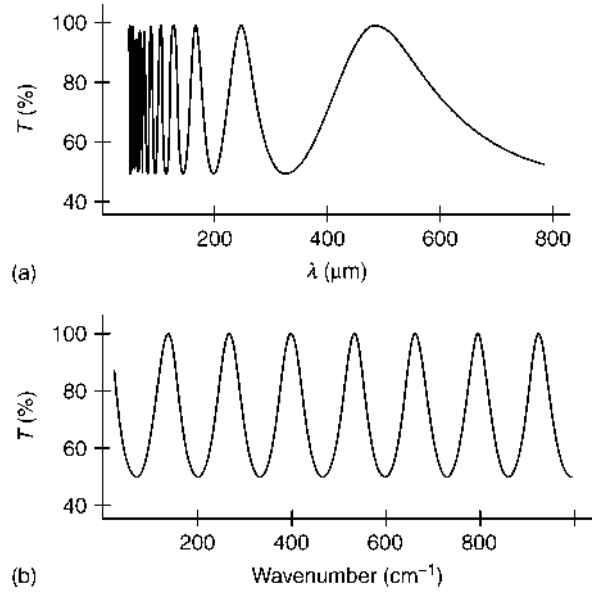

Figure 7 Transmission coefficient of a polished poly-C film with a thickness of 100 μm as a function of (a) wavelength and (b) wavenumber.

or

$$T = \frac{1}{1.49 - 0.493 \cos(0.048\, k)} \qquad [12]$$

where λ is in μm and k, the wavenumber, is in cm^{-1}. Variation of T as a function of λ and k, is shown in Figure 7. A similar behavior is observed for non-polished poly-C films if the wavelength is larger than the surface roughness of the films. For a sample with a very rough surface, scattering effects need to be included in eqn [8].

Electroluminescence in Poly-C

In a luminescence process, the electron must first be excited to higher energy, which can be accomplished by light (photoluminescence), electron beam (cathodoluminescence) or electrical energy (electroluminescence). A subsequent electron transition to lower energies leads to emission of light. A common example of electroluminescence is a forward biased p–n junction made out of direct bandgap type semiconductor such as GaAlAs. The light emitting diodes (LED) and LASERS are based on electroluminescence. Red, green, and blue LEDs are made from GaAlAs, GaP, and SiC/GaN, respectively. Indirect bandgap type semiconductors, such as diamond, typically do not exhibit electroluminescence. The fact that electroluminescence has experimentally been observed in crystalline diamond and poly-C seems to be, at first sight, a contradiction of its

indirect bandgap. In fact, defects in diamond are believed to be responsible for their luminescence properties. All types of diamond are found to have a number of defects related to vacancies and impurities. Due to a unique combination of properties of diamond (Table 1), its electroluminescence properties offer a very unique application potential in optical microelectromechanical systems (MEMS), displays and communication devices in IR, visible and UV regions.

Electroluminescence (EL) in optoelectronic devices is produced by excitation of electrons by electronic current. Yellow-green EL has been observed in Schottky diodes made of boron-doped poly-C. Blue-green and blue-violet EL has been observed in free-standing poly-C films, with thicknesses in the range of 15–400 micrometers, when the samples were subjected to applied fields in the range of 2 V/μm. The EL peaks are observed at energies in the range of 3–3.85 eV, with the main peak occurring at 3 eV.

Recently, a very interesting type of EL has been observed in poly-C and carbon nanotubes, which seems to have an application potential in displays and optical MEMS. Under applied electric fields in the range of 20–40 V/μm, electrons are known to tunnel out of a poly-C film into vacuum and the process is called field emission. EL has been observed during the field emission process, and this new effect is called field emission EL (FEEL). A typical FEEL spectra is shown in Figure 8. It is believed that FEEL results from direct type defects in poly-C, which are populated during the field emission process. The electrons in these defects, instead of being field emitted, make transitions to lower energy levels leading to light emission. The main FEEL peak is observed at 2.6 eV which is in the visible range.

Poly-C Applications

The current applications of diamond, which include optical windows, temperature sensors, heat sinks, cutting tools, surgical instruments, and optical coatings, are based on natural, synthetic, and CVD diamond. The cost of producing diamond has always been prohibitive in widespread diamond applications. Fortunately, over the past 20 years, the development of CVD diamond technologies has eradicated the fears of diamond being too expensive for its widespread applications. As the CVD poly-C can be as inexpensive as any other electronic material, widespread applications of diamond are expected for poly-C. In the area of micro- and nanostructures, the poly-C can provide a unique opportunity for both polycrystalline and crystalline microdevices. Using the present fabrication techniques, the grain size in poly-C can be varied in the range from a few tens of nanometers (nanocrystalline diamond) to hundreds of micrometers (commercial free-standing poly-C films). This leads to a whole range of device structures ranging from inter- to intra-grain devices. An intra-grain device made in poly-C will be an inexpensive single crystal device with properties better that those found in natural/synthetic diamonds, because an intra-grain poly-C device will have a higher degree of purity (because it is grown by CVD).

The applications of poly-C in microsystems (current, in-progress, and future) include optical MEMS, biological/protective/hard coatings, sensors,

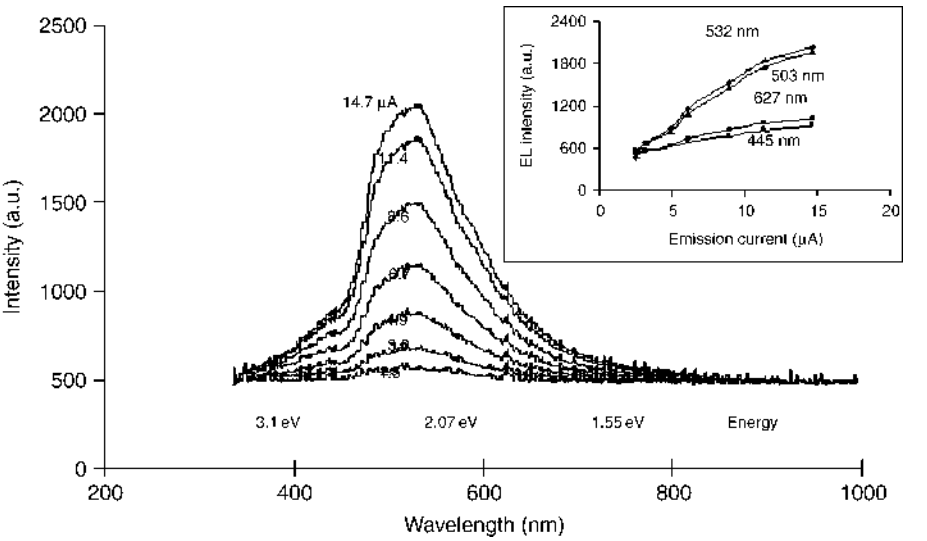

Figure 8 FEEL spectra in poly-C as function of wavelength and emission current (inset).

field emission displays, and WIMS (wireless integrated microsystems). Intra-grain microlenses, RFMEMS, heat pipes, sensors, microabrasives, and spectrometers may lead the way for unique poly-C applications.

Optical MEMS

A basic MEMS structure, with electrostatic actuation mechanism, is shown in Figure 9. Currently, most of the Optical MEMS and RFMEMS (radio frequency MEMS) devices are made from poly-Si (polycrystalline Si) using surface or bulk micro-machining fabrication techniques. The ability to integrate micro-optical elements with movable structures and micro-actuators has opened up many new opportunities for optical and optoelectronics systems, as MEMS allow a more effective manipulation of optical structures than the conventional methods. Optical MEMS have applications in displays, sensing, data storage, and communication systems. Some of the examples are optical switches, optical cross-connects, wavelength division add/drop mixers, tunable filters/lasers/detectors, dispersion components, polarization dispersion components, and spectrometers.

Material-related limitations put some constraints on the current optical MEMS technologies and applications. Diamond is an excellent material for optical MEMS and can extend the application range and reliability. The use of poly-C in optical MEMS can also remove the traditional cost constraints associated with the use of single crystal diamond. Although the poly-C optical MEMS are in their infancy, recent developments are very exciting. Figure 10 shows a poly-C resonator structure fabricated using a poly-C fabrication technology compatible with Si processing techniques. Such structures, if used as filters and switches, can lead to a superior performance as compared to any other currently used material. Poly-C on-chip spectrometers are also expected to operate in a large wavelength range.

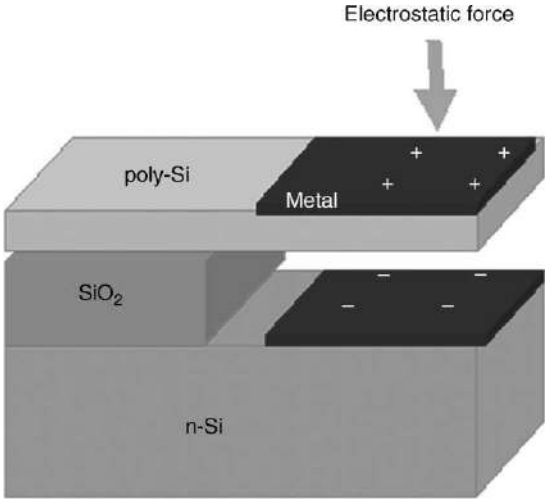

Figure 9 A simple MEMS structure with electrostatic moving of the cantilever beam.

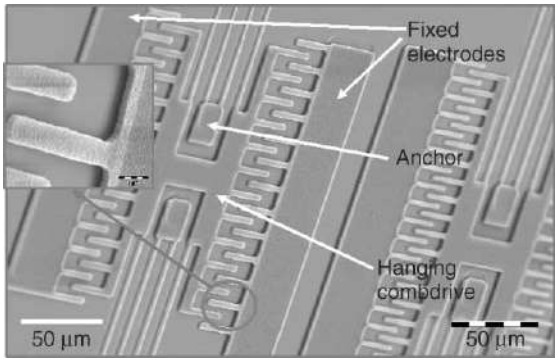

Figure 10 A poly-C MEMS combdrive for high-quality factor resonators, inertial sensors, optical MEMS, etc. Poly-C is superior to poly-Si as its Young's modulus is 4 times higher than that of poly-Si.

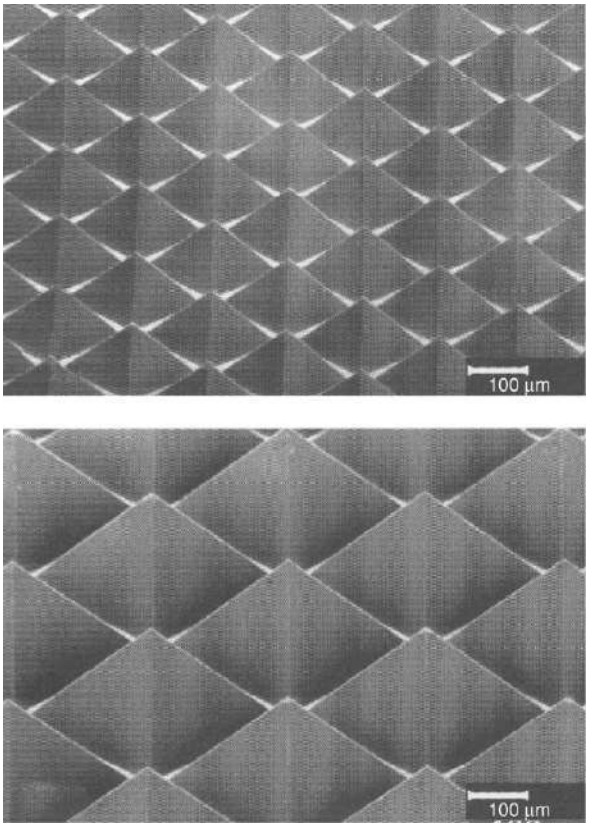

Figure 11 Poly-C microabrasive structures two different densities of pyramids. (Courtesy of Dr. H. Bjorkman of Upsala University.)

Microabrasives

Poly-C is also a very promising material for microabrasives. Micromachined poly-C MEMS structures have recently been fabricated for possible use for microabrasives. Figure 11 shows micromachined MEMS structures which are expected to enhance the life of a tool. Such abrasives by design are inherently cost-effective because they are batch fabricated.

See also

Light Emitting Diodes.

Further Reading

Asmussen J and Reinhard D (2002) *Diamond Handbook*. New York: Marcel Dekker.

Kim U and Aslam DM (2003) Field emission electroluminescence on diamond and carbon nanotube films. *Journal of Vaccum Scientific Technology B* 21(4): 1291–1296.

McMurry J (1996) *Organic Chemistry*. New York: Brooks/Cole.

Pan LS and Kania DR (1996) *Diamond: Electronic Properties and Applications*. Boston, MA: Kluwer.

Spear KE and Dismukes JP (1994) *Synthetic Diamond: Emerging CVD Science and Technology*. New York: Wiley.

Laser Damage in Thin Film Coatings

D Ristau, Laser Zentrum Hannover e.V., Hannover, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Since the very beginning of laser technology, Laser-induced Damage Thresholds (LIDT) of optical components were always an obstacle for the development of laser systems operating at high power levels. In such systems, the surface or the bulk of the optical components can be damaged by laser radiation of sufficiently high power or energy density resulting in an uneconomical standstill of the laser and its application. In the early days of high power laser development, mainly inclusions in laser rod materials were discussed as a major complication for an augmentation of the output power in solid state laser systems. Nowadays, in the course of the development of optical materials with excellent quality and power handling capability, the problem of laser-induced damage has shifted from the bulk to the surface of the optical component. The optical surface is subjected to various production steps and environmental influences, which modify its structure and composition. Especially, the thin film coating, which is deposited on the optical surface to adapt its reflectance and transmittance to the application, contributes predominantly to the reduction of the LIDT values. As a consequence, the measurement and optimization of the power handling capability of thin films is considered as one of the primary research areas in modern optics technology and is supported by an extensive scientific community.

In the following, a brief review will be given on selected fundamental damage mechanisms in thin films, considering also the scaling of LIDT values for different operation conditions. Also, current standards for the measurement of LIDT will be described, and examples illustrating some practical aspects of high power optical coatings will be presented. Finally, a summary of the present state of the art will be given and recent trends in laser technology will be discussed in respect to research in laser-induced damage.

Fundamental Damage Mechanisms in Thin Films

During thirty years of intense research in laser damage of thin films, a broad spectrum of different damage and degradation mechanisms had been discovered and studied for a variety of laser operation conditions and wavelengths. For prominent high-power laser systems often applied in industrial production fields or fundamental research, specific models for laser-induced breakdown have been developed, which are also of practical relevance. In many experiments, damage sites with the morphology illustrated in Figure 1 are observed. The coating is delaminated from the surface of the optical component in the center of the laser beam area. Obviously, mechanical stress has built up in the coating by heating, which is induced by the absorbed laser power. At a certain stress level, exceeding the adhesion strength of the coating to substrate surface, the thin film cracks and may even delaminate from the component. In other cases of this absorption-induced damage effect, the film reaches its melting point prior to the mechanical threshold and

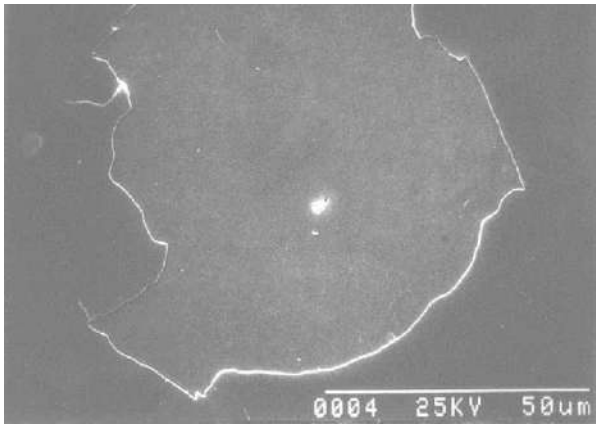

Figure 1 SEM-picture of a damage site on an alternating coating of SiO_2/HfO_2. The system was irradiated by a Nd:YAG-laser with an energy density of 45 J/cm^2. As a dominant damage mechanism, a delamination of the coating by extreme mechanical stress is suggested. Reproduced with permission from Ristau D, Ebert J and Welling H (1989) Optische Beschichtungen für Hochleistungslaser. *Laser und Optoelektronik* 21(4): S.49.

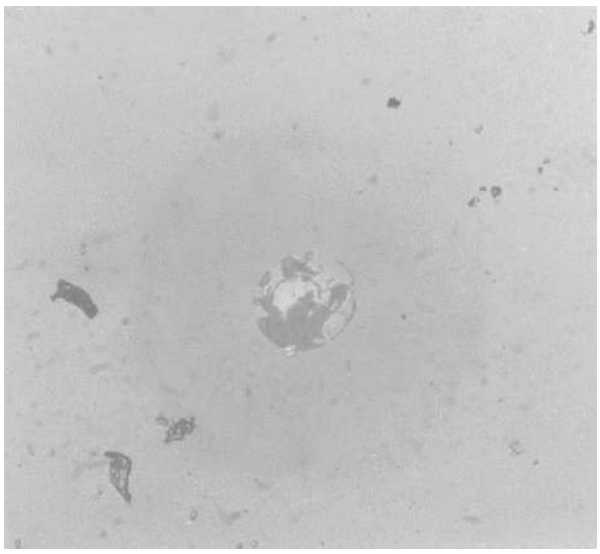

Figure 2 Nomarski micrograph (magnification 250×) of a damage site on an antireflecting coating of SiO_2/Ta_2O_5. The system was irradiated with an energy density of 51 J/cm^2 at a wavelength of 1.064 μm. A delamination of the coating in the center of the beam is surrounded by a recrystallization zone.

evaporates or changes its crystalline structure. The morphology of such damage sites exhibits a discoloring or an increased surface roughness in the center of the laser beam (see Figure 2). Absorption-induced damage is dominated by the instantaneous heating of the coating material in the area of interaction with the laser beam and can be described on the basis of the heat diffusion equation. In this approach, the source term in the heat diffusion equation is given by the heat coupled into the system by the absorbed laser power, and the boundary conditions are determined by the geometry and the structure of the layer system as well as the spatial laser beam profile. Apparently, absorption-induced damage has to be considered only for components with significant absorptance at their operation wavelength. Typical examples for coating materials and wavelength regimes, where absorption-induced damage is frequently observed, are the wavelength 10.6 μm of the CO_2-laser with ZnSe, ZnS, and fluorides as coating materials, or the DUV/VUV-wavelengths dominated by excimer lasers and fluoride coating materials. The effect of absorption-induced damage in optical materials is also of practical relevance for cw- and long-pulse operation conditions of the laser system corresponding to long thermal diffusion lengths. For cw-irradiation, the temperature rise in the component can be calculated by numerical methods involving finite elements or differences. For boundary conditions with circular symmetry and a Gaussian beam profile (beam diameter w, power P), an analytical expression can be derived for the temperature rise ΔT in the center of an irradiated circular component

$$\Delta T = \frac{2\beta_s P}{\pi^{3/2} k w} \tan^{-1}\left(\frac{16\kappa t_I}{w^2}\right)^{1/2} \qquad [1]$$

In this model the temperature rise is dependent on the thermal properties (k: thermal conductivity, κ: thermal diffusivity), on the surface absorption of the component (β_S), and on the beam diameter. For irradiation times t_I, long compared to the typical heat diffusion time w^2/κ, eqn [1] reduces to the asymptotic dependence

$$\Delta T \rightarrow \frac{\beta_s P}{\pi^{1/2} k w} \qquad t_I \rightarrow \infty \qquad [2]$$

Obviously, the temperature rise ΔT scales with the P/w for these long-time irradiation conditions compared to the short pulse regime, where the scaling of the temperature rise is proportional to P/w^2. For the application of optical components in high power cw- and long-pulse lasers, this P/w scaling law has to be considered carefully, because the power handling capability depends critically on the size of the irradiated area. Also, the onset of damage may be delayed in respect to the start of irradiation by time constants in the range of several 10 seconds.

In the course of the development of improved coating processes, optical absorptance could be reduced to very low levels in the near infrared and visible spectral region. For example, high reflecting mirrors for the wavelength of the Nd:YAG-laser with total losses below 1 ppm can be produced with

advanced ion beam sputtering processes nowadays. Therefore, another principal damage mechanism, which is based on inclusions or defects in the coatings, is often found in the near infrared spectral range. A detail of a damage site representing inclusion dominated damage induced by a pulsed Nd:YAG-laser in a coating of HfO_2/SiO_2 is shown in Figure 3. In the laser beam area numerous craters are located, which extend from circular voids with different diameters in a range below 1 μm. At the high magnification of the employed electron microscope, the multilayer structure of the coating system can be resolved indicating an origin of damage approximately in the same depth of the layer structure at the substrate surface. The underlying damage mechanism is initiated by inclusions or other defects with an absorption, which is significantly higher than for the surrounding thin film material. As a consequence, the inclusion is rapidly heated during the interaction with the laser beam and explodes removing the covering layer structure. This inclusion breakdown can be modeled by calculating the temperature behavior $T(r)$ at the radial position r of a sphere (radius g, thermal conductivity k_E, diffusivity κ_E) which is embedded in a medium with defined properties (see eqn [3]). The interaction of the sphere, which has a size in the range of the wavelength λ and a refractive index n', with the laser irradiation, is described by the Mie absorption coefficient Q. For the instant of damage, a critical temperature at the outer surface of the sphere ($r = g$) is assumed, which is often related to the melting point of the layer material

$$T(r) = \frac{3QI}{4\pi k_E g}\left[\frac{k_E}{3k} + \frac{1}{6}\left(1 - \frac{r^2}{g^2}\right) - \frac{2gb}{r\pi}\int_0^\infty \mathrm{d}y \exp\left(-\frac{y^2 t_p}{\tau_1}\right) \times \frac{(\sin y - y\cos y)(\sin ry/g)}{y^2\left[(c\sin y - y\cos y)^2 + b^2 y^2 \sin^2 y\right]}\right] \quad [3]$$

$$Q = \pi g^2\left[1 + \frac{2\exp(-8\pi n' g/\lambda)}{8\pi n' g/\lambda} + \frac{\exp(-8\pi n' g/\lambda) - 1}{(8\pi n' g/\lambda)^2}\right]$$

$$\tau_1 = \frac{g^2}{\kappa_E} \quad b = \frac{k}{k_E}\left(\frac{\sqrt{\kappa_E}}{\sqrt{\kappa}}\right) \quad c = 1 - \frac{k}{k_E}$$

Besides the thermal properties of the host material (thermal conductivity k, diffusivity κ), the properties and the statistic distribution of the inclusions have to be known for a quantitative description of inclusion breakdown.

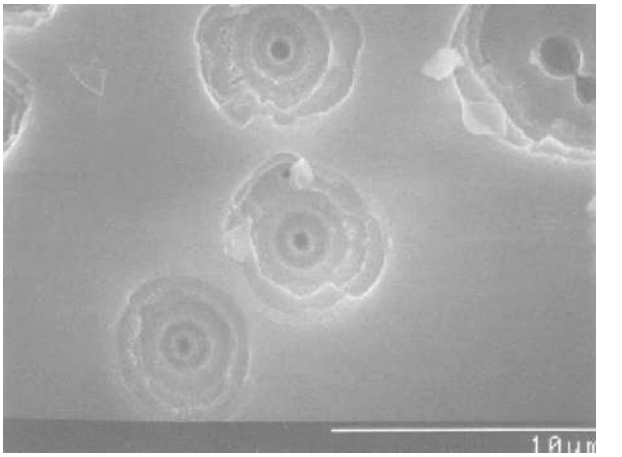

Figure 3 SEM-picture of damage site on a high reflecting coating of SiO_2/HfO_2. This inclusion-dominated breakdown mechanism was initiated at an energy density of approximately 130 J/cm^2 with a Nd:YAG-laser.

In contrast to the nearly deterministic behavior of absorption induced damage, defect dominated damage mechanisms often exhibit a statistical nature. Since the inclusions are not homogeneously distributed in the layer structure, a variation of the threshold value is observed for different damage sites on a single sample. The assessment of the inclusion parameters is extremely difficult in practice. Therefore, the inclusion model is mainly employed for the prediction of tendencies for the LIDT as a function of the beam parameters and the properties of the layer materials.

Different aspects related to the substrate polishing, cleaning, and deposition process have to be considered for an investigation in the various origins of inclusions or defects in a layer structure. Besides contamination of the substrate surface prior to the deposition process, particles may be generated during the deposition process by mechanical abrasion of material from moving parts in the plant or by delamination of material coated on the inner walls. Also, the deposition source may contribute to the formation of defects by sputtering of particles or droplets from the melt or the target. Especially in low energetic thermal deposition processes, defects are often the origin for the formation of nodules. These conical faults in the layer structure reduce the LIDT because they are less resistant against intensive laser radiation than the bulk of the coating system.

Besides models, which are based on the transformation of laser radiation into heat, direct electronic excitation has to be considered as a breakdown mechanism for short-pulse lasers. In these intrinsic breakdown models for dielectric materials, the

generation of a high electron density in the conduction band is assumed. This carrier generation can be explained by high field ionization of electrons in the valence band of the dielectric material. Many theoretical concepts introduce the avalanche effect as an important mechanism contributing to the generation of free electrons. Electrons excited by strong field ionization can interact with the electric field and accumulate excitation energy. By collisions with the lattice, this energy can be also transferred to the electrons in the valence band. Thus, when the energy of the free electrons exceed a critical value, other electrons can be excited to the conduction band, and a carrier generation like an avalanche arises. Another theoretical approach is the multiphoton model which is formulated on the basis of multiphoton absorption for the generation of free electrons. In most intrinsic models it is assumed, that laser-induced damage breakdown occurs when the electron density in the conduction band reaches a critical value of approximately 10^{21} 1/cm^3. At this high electron density, a plasma state is reached and the impinging radiation is coupled efficiently into the layer structure resulting in its destruction.

Intrinsic damage is often characterized by a deterministic damage behavior with a well defined threshold value which is also characteristic for damage in bulk materials. As a prominent example, damage induced by ultra-short pulses with a duration below 1 ps can be often attributed to intrinsic mechanisms, because the diffusion length of thermal effects can be neglected in comparison to the intensive interaction of the laser radiation with the electrons of the coating materials. In this case, the generation rate $\mathrm{d}n/\mathrm{d}t$ of electrons in the conduction band can be described on the basis of the impact ionization rate and multiphoton excitation

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \alpha n(t) q I(t) + \beta_m \left[q I(t) \right]^m \qquad [4]$$

In this model $I(t)$ is the power density of the laser radiation, and q is a correction factor representing interference effects in the coatings, which may result in a local power exceeding the power density of the radiation impinging on to the layer system. The electron excitation mechanisms are described by the impact ionization coefficient α and the m-photon absorption coefficient β_m. The relative contributions of both excitation mechanisms to laser-induced breakdown are depending on the pulse duration and the bandgap energy of the layer material.

In a rough approximation, the damage threshold increases with the bandgap energy of the materials as a consequence of the photon absorption coefficient β_m decreasing with its order m. For short pulse durations below a few 10 fs, multiphoton processes contribute significantly in the range of some 10% to the free-carrier generation. In most materials, avalanche ionization develops to the dominant generation mechanism for pulse durations above 100 fs. These tendencies could be demonstrated for single layers of selected oxide material deposited by ion beam sputtering on quartz substrates, which were tested in a pulse duration regime from 20 fs to 1 ps. According to the theoretical model, the expected increase of the damage also influencing values with the pulse duration was observed in the damage experiments.

In practice, the power handling capability of optical components is also often limited by imperfections like scratches, digs, and areas with high roughness on the optical surface. In most deposition techniques applied in optical technology, the layer system tends to replicate or even enhance surface imperfections of the substrate. In the application, additional weak points may be introduced by improper handling or contamination of the optical surfaces. If the contaminants are not removed by the impinging high-power laser radiation, they increase the surface absorption and act as initiation points for laser-induced damage. Voids, grooves, pores, or scratches reduce the power capability of the optical element, because they act as concentrators for the electric field.

Units and Scaling of Laser-Induced Damage Threshold

According to the theoretical models of laser-induced damage, the appropriate units of measurement for LIDT values are mainly given by the dominant damage mechanism and the irradiation time. For pulsed laser radiation and dielectric breakdown effects, the laser-induced damage threshold is usually reported in W/cm^2. In the case of absorption-induced damage or inclusion dominated breakdown and a pulse duration short compared to the thermal diffusion time in the layer structure, LIDT values are often scaled in J/cm^2. The unit of linear power density (W/cm) is indicated for the LIDT of long-pulse, cw-lasers or other sources, which induce a temperature increase in the optical component with rise times in the regime of seconds.

Laser-induced damage thresholds of optical thin films are critically dependent on the operation conditions of the applied laser system and on the design of the layer structure. For example, even the damage mechanism can change if the wavelength or the pulse duration is varied. Therefore, LIDT

values should be scaled with the irradiation parameters only for small intervals, in cases where the fundamental damage mechanism is clarified. In all other situations, the extrapolation of threshold values is extremely difficult and may lead to inaccurate results. Especially, an overestimation of the LIDT value holds severe dangers in practical applications, where the replacement of a damaged component and the repair of other parts impaired by the damage event causes high expenses. If the damaged components consist of toxic materials (e.g., ZnSe, GaAs, CdTe, ThF_4, chalcogenides, Be, Cr, etc.) severe health hazards may occur, and an expensive decontamination of the environment may be necessary. Therefore, the determination of the laser-induced damage threshold using the actual laser parameters under controlled environmental conditions is always recommended for unclear cases.

In the present state of research in laser-induced damage mechanisms, only few tendencies are generally accepted as confident for scaling of LIDT values. As a function of the pulse duration, laser-induced damage thresholds increase for longer pulses. For inclusion and absorption dominated breakdown, a $t_p^{1/2}$-law is often used in the pulse duration regime between 10^{-10} to 10^{-8}s. This dependency can be extended to other pulse regimes up to pulse lengths in the ms-range, if the exponent is replaced by a value between 0.5 to 2. In respect to the laser wavelength, a decrease of the LIDT value with decreasing wavelength is observed for most materials and operational conditions. Investigations in the influence of the beam diameter have been performed by many research groups, indicating a decrease of the LIDT values for increasing beam diameters. Especially for inclusion dominated breakdown, the event of damage for a certain laser irradiated site will be dependent on the distribution of inclusions at that position. If the beam size is small, the probability for interrogating a defect vulnerable to damage is low. By increasing the spot diameter, this probability will asymptotically reach unity, because at a certain beam diameter, at least one defect will always be covered by the beam. Therefore, the onset of laser-induced damage is not dependent on the beam diameter for inclusion-dominated breakdown, which is often encountered in conventional optical coating systems. Another special case is the scaling of the cw-damage threshold with the beam diameter. Since the cw-LIDT is expressed in linear power density, the power handling capability increases more slowly with the beam diameter for the cw-lasers than expected from the normal pulsed operation. For example, extrapolating the LIDT value of 100 W at a beam diameter of 1 mm for a laser mirror to a beam diameter of 10 mm, results in a maximum power load for the component of 1 kW according to the correct P/w-law. The threshold power would be extremely overestimated to approximately 3.2 kW, if the general P/w^2-dependence is applied.

Measurement of Laser-Induced Damage Thresholds

As a consequence of the complicated relation between the laser damage mechanism and a broad spectrum of thin film properties and laser parameters, laser-induced damage threshold measurements have to be performed under well-defined conditions. To investigate the comparability of LIDT measurements, an extended international round-robin experiment has been conducted on coated optics for the wavelength of 1.064 μm at the beginning of the 1980s. This experiment indicated the need for a clearly specified LIDT-measurement procedure, and conceptual work was initiated to develop a corresponding ISO-Standard series. During recent years, an International Standard (ISO 11254) has been adopted covering testing conditions relevant for most typical laser applications. In the first part of ISO 11254 1 on 1-testing of optical surfaces, in respect to laser damage, is described. The fundamental approach of the standard measurement procedure is illustrated in Figure 4. A laser source operating in transversal and longitudinal single mode is employed for the irradiation of the sample surface. The beam

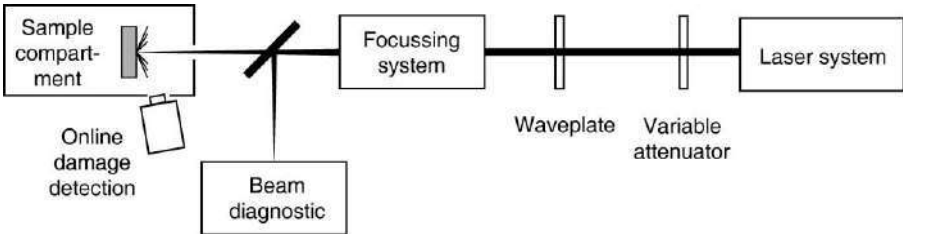

Figure 4 Fundamental setup for the measurement of laser-induced damage thresholds.

parameters of the laser are assessed by a beam diagnostic system which monitors the spatial and temporal profile as well as the energy of the laser radiation in the target plane. In order to achieve the high energy density levels necessary to destroy the surface of the specimen, a well characterized focusing system is installed. For adjustment of the laser energy in the target plane, an attenuating system is employed. In the 1 on 1 measurement protocol, each site on the sample surface is subjected to a single laser pulse once only. In a test sequence, the test surface is examined by pulses of different energies, covering low values without damage and high values causing damage. After the test, the specimen will be inspected with a Nomarski interference contrast or a darkfield microscope at a magnification of 150× or higher to identify the damaged sites. For the evaluation of the damage test, the damage probability method is recommended. In this evaluation scheme, the ratio of the number of damaged sites to the total number of sites objected to a certain energy or power level is interpreted as the damage probability. The plot of these damage probability values as a function of energy or power, which is called the survival curve of the optical component, provides an insight into the damage mechanisms involved. The damage threshold is given by the highest quantity of laser radiation for which the extrapolated probability of damage is zero. A typical example for a 1 on 1 damage test at the wavelength 1.064 μm and the extrapolated LIDT value is depicted in Figure 5.

In most catalogs of optics manufacturers, the 1 on 1 LIDT values are used to illustrate the power handling capability of their products, even though 1 on 1 data are of limited importance for practical applications, where an optical component is always subjected more than one time to a laser beam. This customary situation is covered by Part 2 of the ISO-LIDT-Standard which describes a damage test procedure for a series of pulses (S on 1-tests). For an assessment of the reliability, the concept of the characteristic damage curve is introduced by this standard. This curve is deduced directly from the S on 1 test data by plotting the energy density for a selected damage probability as a function of the number of pulses (see Figure 6). By an extrapolation of the characteristic damage curve to high pulse numbers in the order of 10^9 to 10^{12} shots, the lifetime of the optical component can be roughly estimated. For a certification of optical components in respect to their power handling capability, a third part of ISO 11254 is under development, which is concentrated on different testing protocols of a defined surface fraction at power levels expected in the application. The fundamental approach of these tests is a simulation of conditions at the upper limits of operation parameters encountered in practice.

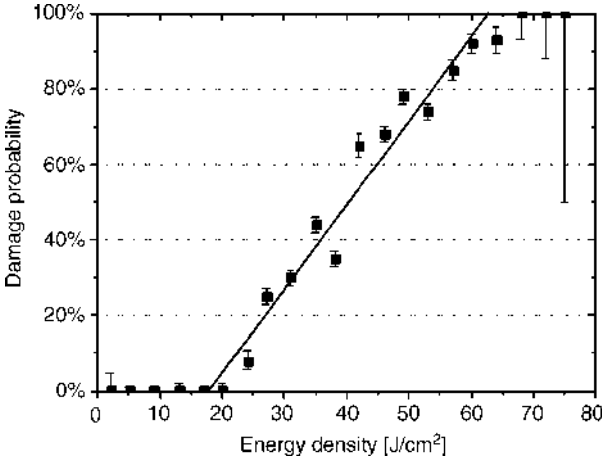

Figure 5 Survival curve of an anti-reflective layer system (two layer V-coating, SiO_2/Ta_2O_5) on fused silica substrate for the Nd:YAG-laser wavelength. The measurement was performed at a wavelength of 1.064 μm, a beam diameter of 420 μm, and a pulse duration of approximately 15 ns, respectively.

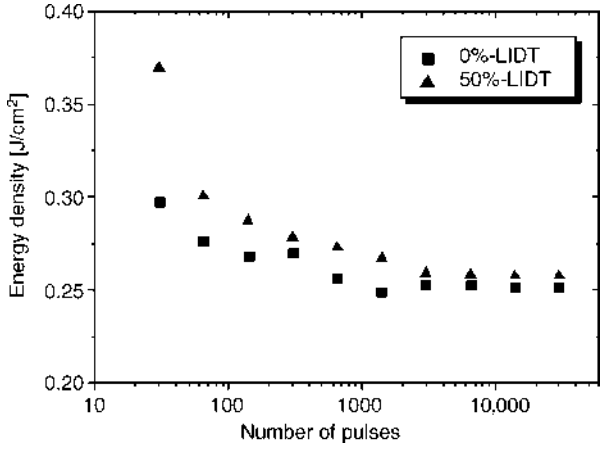

Figure 6 Characteristic damage curve of a high reflecting mirror of SiO_2/TiO_2 for fs-laser systems. The layer structure was tested with a high repetition (1 kHz repetition rate) fs laser system operating at a pulse duration of 150 fs.

Optical Coatings for High Power Lasers

For the development and application of coating systems with high LIDT values, several major aspects have to be considered. In the first approach, appropriate materials and processes have to be selected which deliver coatings with sufficient power resistance. In many studies, a correlation of the LIDT values of the constituent single layers to the power

handling capability of a layer system was observed. Therefore, single layers of potential deposition materials were investigated at the wavelengths of prominent laser systems from the VUV- to the FIR-spectral region. For specific deposition processes, often a clear ranking of the coating materials in respect to their power handling capability could be demonstrated. An example for a Nd:YAG-laser with a pulse duration of 15 ns is illustrated by Table 1. With the exception of the ZrO_2-layer, which exhibited severe inhomogeneity and a high inclusion density, a correlation can be assumed between the LIDT value and the melting point of the bulk materials under these operation conditions. This dependency can be attributed to absorption-induced damage as well as to the inclusion model, and therefore, the damage morphology has to be studied in each case to identify the dominating damage mechanism. In the fs-regime or the UV/VUV-spectral range, often a dependence of the power handling capability on the bandgap of the materials is found. According to the relationships of the avalanche and other dielectric breakdown models, materials with highest bandgaps are frequently encountered at the top of an LIDT ranking.

For the influence of the coating process on the laser-induced breakdown of optical coatings, no clear tendencies can be detected. For example, Ion Beam Sputtering (IBS), which is considered as the deposition process for coatings with extremely low losses and contamination, cannot always surpass conventional thermal deposition processes in laser stability. In general, ion or plasma assisted deposition techniques produce coatings with lower LIDT values than conventional coating processes for most operation conditions. Sol-gel processes, which were developed for the deposition of removable coatings on large optics in laser fusion, can reach superior power handling capabilities. In general, as a consequence of the complicated relation between the production parameters and laser-induced breakdown, production processes for optical coatings have to be optimized separately for different wavelength regimes and irradiation conditions.

Table 1 Laser induced damage thresholds of selected single-layer coatings in relation to the melting points of the corresponding bulk materials. In the last column, LIDT values of anti-reflective coatings (AR-Coating) composed of the high index material and SiO_2 are compiled to illustrate the correlation between single LIDT values and the damage threshold of layer systems

Material	*Melting point (°C)*	*LIDT value (J/cm^2) single layer*	*LIDT value (J/cm^2) AR-coating*
TiO_2	1775	13 ± 1	14 ± 1
Ta_2O_5	1918	28 ± 2	32 ± 5
HfO_2	2758	41 ± 3	46 ± 6
ZrO_2	2700	34 ± 4	28 ± 1
Al_2O_3	2072	39 ± 1	
SiO_2	1723	34 ± 7	

Besides the fundamental production parameters, the design of the coating system is of similar importance for the achievement of high LIDT values, because damage is directly driven by the electric field strength in the layer structure. A standing wave field pattern is depicted in Figure 7 for a high reflecting stack of Ta_2O_5 and SiO_2. The design consisting of layers with an optical thickness of 1 QWOT is typical for most standard laser mirrors. The power density reaches extreme values always at the interfaces between the layers and even exceeds the incoming irradiation power density (100%). Interfaces between the layers can be considered as weak points in a coating system, because additional contamination may occur during the switching of the material in the production process. Also, the adhesion between the adjacent layers may be reduced, and mechanical stress may be built up by the different materials. To improve damage thresholds of laser mirrors, the thickness of the first few layer pairs can be adjusted to shift the points of maximum field strength into the bulk of the layer with higher damage resistivity (see Figure 8). Also, a thick layer can be attached to the system in order to stabilize the outer layer pair in respect to thermal or mechanical stress. Another technique, which can be applied to enhance the stability of the interface is the codeposition of materials, resulting in a gradual interface with a mixing zone of materials between the layers. This region of codeposited material exhibits a higher resistance against mechanical stress resulting in an improvement of LIDT values of up to 20%.

The effect of the internal electric field strength distribution on the power handling capability of a layer system should be most apparent for intrinsic damage mechanisms, which are dominant in the ultra short pulse regime below 1 ps. According to the fundamental model (see eqn [4]), a direct relationship of the damage threshold to the maximum field strength value within the layer system is expected. As an example, an investigation in the threshold behavior of ion beam sputtered coating systems with different field strength values is illustrated in Figure 9. In this experiment, the maximum field strength in the last low index layer of SiO_2 has been adjusted to factors between 0.4 and 1.6 of the impinging field strength (see upper diagram in Figure 9) by depositing different designs on a basic 1-QWOT layer stack of TiO_2/SiO_2. In the lower part of Figure 9, the S on 1 LIDT values, measured with a fs-laser, are depicted for selected pulse numbers N. For all layer systems,

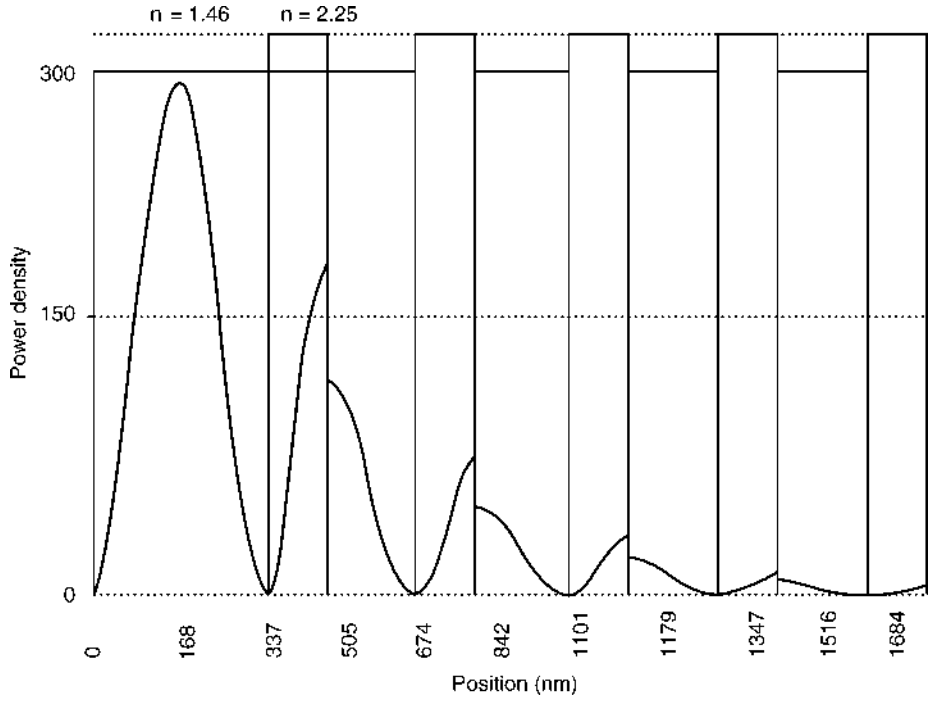

Figure 7 Distribution of the power density in a high reflecting layer stack of SiO_2/Ta_2O_5. The first nine layers of the system and a final thick stabilizing layer of SiO_2 next to the air interface are depicted. The incoming energy density is calibrated to 100%. Reproduced with permission from Ristan D, Ebert J and Welling H (1989) Optische Beschichtungen für Hochleistungslaser. *Laser und Optoelektronik* 21(4): S.53.

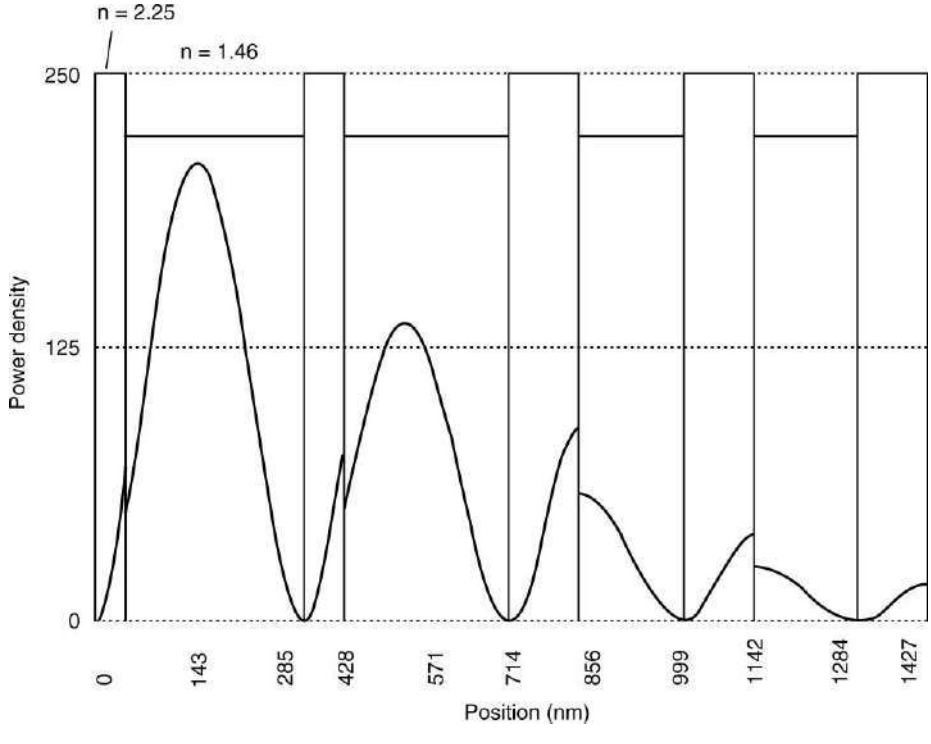

Figure 8 Distribution of power density in a high reflecting layer stack of SiO_2/Ta_2O_5. The first nine layers of the system, which is designed for reduced power density at the interface in the first layer pair are depicted. Reproduced with permission from Ristan D, Ebert J and Welling H (1989) Optische Beschichtungen für Hochleistungslaser. *Laser und Optoelektronik* 21(4): S.53.

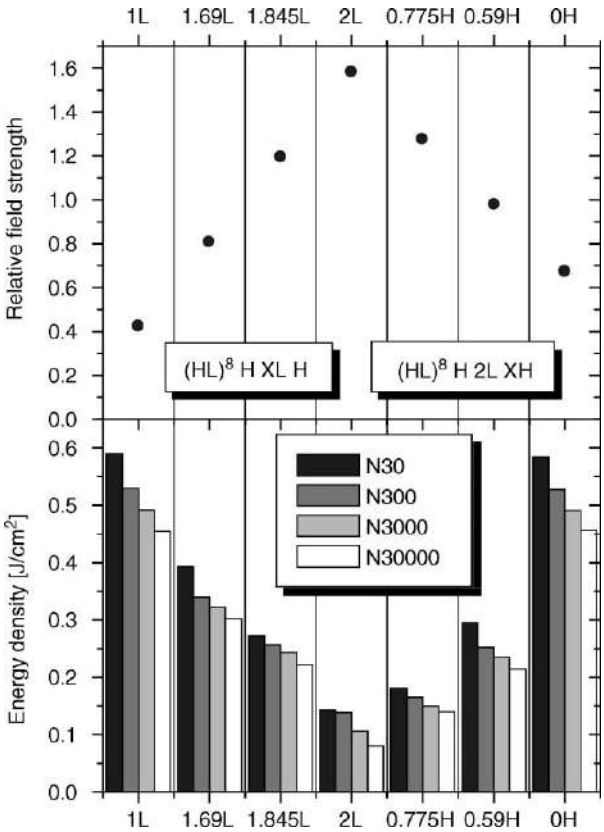

Figure 9 Damage thresholds (lower diagram) and internal field strength relative to the field strength of the impinging wave (upper diagram). The damage thresholds were measured with an ultra-short pulse laser operating at a repetition rate of 1 kHz, a pulse duration of 150 fs, and a beam diameter of 100 μm on the sample surface. Damage thresholds for selected numbers between $N = 30$ to $N = 30\,000$ of pulses are indicated by columns. The plotted maximum field strength in the last SiO_2-layer is adjusted on the basis of design variation.

which were deposited in separate deposition runs, a strong correlation between the maximum field strength and the laser threshold is observed. This experiment clearly demonstrates the role of the electric field strength distribution in high-power coatings and the potential of advanced thin film design strategies.

Besides the properties of the coatings, the quality of the substrate has to be considered in respect to surface imperfections and to the polishing procedure. For example, substrates for the production of high-quality optics with enhanced power handling capability in the VIS/NIR-spectral range should be polished to a surface roughness of less than 1 nm_{rms} with a surface imperfection value $5/1 \times 0,010$ according to ISO 10110. For an illustration of the effect of the polishing grade on the LIDT values, selected results are compiled in Table 2 for substrates of BK7-glass and high reflecting mirrors.

Table 2 Laser-induced damage threshold values of samples polished using compounds of different grain size. Besides the LIDT values of the uncoated surfaces, also data for anti-reflective coatings of selected materials on these surfaces are summarized

Coating type	*LIDT value (J/cm^2) polishing type I*	*LIDT value (J/cm^2) polishing type II*
Bare surface	76 ± 1	68 ± 12
Ta_2O_5/SiO_2	44 ± 2	38 ± 1
HfO_2/SiO_2	47 ± 16	35 ± 5
Nd_2O_3/MgF_2	34 ± 6	24 ± 8

Substrates with the polishing Type I and Type II were processed with powder of different grain diameter (2 μm and 3 μm) in conventional pitch polishing. For the LIDT values of the bare substrate and the coatings, a clear relation to the surface roughness can be observed. Especially for optical coatings with significant transmittance, subsurface damage in the substrate has to be taken into account. As a consequence of the chemical and mechanical interaction of the surface with the polishing compounds, impurities and dislocation are introduced in the surface structure resulting in a reduction of the damage threshold.

Summary

In the course of the rapid development of laser technology, a large background in high power optical coatings had been built up during the last three decades. Nowadays, the corresponding experience in the production of high power coatings is mainly located at industrial companies and a few research institutes, which are also involved in the characterization of optical coatings. For an illustration of the present state in thin film technology, damage threshold values of advanced optical components are summarized in Table 3 for laser systems and operation conditions often applied in industrial production environments.

Besides the spectacular experiments in laser fusion, isotope separation, and fundamental physics, industrial applications of lasers in material processing, medicine, information technology, and semiconductor lithography are considered as major pacemakers for the progress of high power optics. Therefore, trends in these technology fields will dominantly govern the future development of thin film technology. For example, in semiconductor lithography at a wavelength of 157 nm, which would open the way to feature sizes well below 100 nm, optical coatings with extended lifetimes are still on rank 6 of the list of challenges to achieve an effective lithographic production facility. In next generation lithography, optics

Table 3 Laser-induced damage threshold values of selected optical coating systems for laser applications (Types: HR: high reflecting mirror, AR: antireflective coating, th: thermal evaporation, IBS: ion beam sputtering)

Laser system wavelength	*Type*	*Laser-induced damage threshold ISO 11254*
193 nm, ArF-excimer	AR/th	1–2 J/cm^2 (1 on 1, 20 ns)
	HR/th	2–4 J/cm^2 (1 on 1, 20 ns)
248 nm, KrF-excimer	AR/th	10 J/cm^2 (1 on 1, 30 ns)
	HR/th	>20 J/cm^2 (1 on 1, 30 ns)
	HR/IBS	>3 J/cm^2 (1 on 1, 30 ns)
1,064 μm, Nd:YAG	AR/th	>60 J/cm^2 (12 ns, 0,25 mm)
	HR/th	>100 J/cm^2 (12 ns, 0,25 mm)
	HR/IBS	>80 J/cm^2 (12 ns, 0,25 mm)
10,6 μm, CO_2-Laser	AR/th	>20 J/cm^2 (100 ns, 1,4 mm)
		>2 kJ/cm^2 (1,2 ms, 250 μm)
		>3 kW/mm (cw, 100 μm)
	HR/th	>25 J/cm^2 (100 ns, 1,4 mm)
		>2 kJ/cm^2 (1,2 ms, 250 μm)

of even smaller wavelengths, around 13 nm, have to be optimized to achieve the 60 nm node. Ultra-short pulse lasers gain importance as innovative tools for material processing, laser medicine, and biology, as well as the analysis and control of chemical reactions. For the development of fs lasers needed in these applications, special high-power broadband coating systems are required. In addition to high LIDT values, these coatings have to fulfil demands with respect to their group delay dispersion for the compensation of dispersion effects in the laser systems. Other challenges have to be expected from applications, where a combination of the high-power handling capability with additional properties has to be achieved. As a typical example, high-power laser coatings with improved mechanical or chemical stability for applications in laser medicine can be considered.

List of Units and Nomenclature

dn/dt	generation rate of electrons in the conduction band
g	radius of a spherical inclusion; (cm)
$I(t)$	power density of laser radiation,
k	thermal conductivity; (W/(cm °C))
k_E	thermal conductivity of a spherical inclusion; (W/(cm °C))
LIDT	laser induced damage threshold (J/cm^2; W/cm^2; W/cm)
n'	refractive index of a spherical inclusion; (1)
P	output power of a laser; (W)
q	correction factor representing interference effects in coatings
Q	Mie absorption coefficient of a spherical inclusion;
QWOT	quarter wave optical thickness: unit for the thickness of the layer
t_I	irradiation time for a cw-laser; (s)
$T(r)$	temperature in a spherical inclusion at the radial position r; (°C)
w	diameter of a laser beam with Gaussian profile; (cm)
α	impact ionization coefficient; (1)
β_m	m-photon absorption coefficient; (1)
β_S	surface absorption of an optical component; (1)
ΔT	temperature rise in the center of an irradiated circular spot; (°C)
κ	thermal diffusivity; (cm^2/s)
κ_E	thermal diffusivity of a spherical inclusion; (cm^2/s)
λ	wavelength of a laser; (cm)

See also

Optical Coatings: Thin-Film Optical Coatings.

Further Reading

Bloembergen N (1973) Role of cracks, pores and absorbing inclusions on laser damage threshold of transparent dielectrics. *Applied Optics* 12(4): 661–664.

Genin FY and Stolz CJ (1996) *Morphologies of laser induced damage in hafnia-silica multilayer mirror and polarizer coatings*, Third International Workshop on Laser Beam and Optics Characterization. In: Morin M and Giesen A (eds). *Proc. SPIE 2870*, 439–448.

Guenther AH, *et al. Proceedings of the Symposium on Laser Induced Damage in Optical Materials, NBS (NIST) Special Publication 1969–1989*, SPIE vols. 1989–2003. CD-Rom version of collected papers from 1969 to 1998 available: Washington: SPIE.

Guenther KH, Ebert J, Kiesel E, *et al.* (1984) 1.06-m laser damage of thin film optical coatings: a round-robin experiment involving various pulse lengths and beam diameters. *Applied Optics* 23: 3743.

ISO 11254 (2001) *Test methods for laser induced damage threshold of optical surfaces*. Part 1: 1 on 1-test, 2000, Part 2: S on 1 test.

ISO 11254-3 (2002) *Laser and laser-related equipment – Determination of laser-induced damage threshold of optical surfaces – Part 3: Assurance of laser power handling capabilities*. Draft Standard. International Organization of Standardisation.

Jasapara J, Nampoothiri AVV, Rudolph W, Ristau D and Welsch E (2001) Femtosecond laser pulse induced breakdown in dielectric thin films. *Physical Review B* 63(4): 045117.

Puttick K, Holm R, Ristau D, *et al.* (1997) *Continuous wave CO_2 laser induced damage thresholds in optical*

components, Proceedings of the Symposium on Laser Induced Damage in Optical Materials. SPIE vol. 3244, 188–197.

Rudolph W, Mero M, Liu J, *et al.* (2003) *Femtosecond pulse damage and pre-damage behavior of dielectric thin films*, Proceedings of the 34th Annual Symposium on Optical Materials for High Power Lasers. SPIE vol. 4932, 202–215.

Walker TW, Vaidyanathan A, Guenther AH and Nielsen P (1979) *Impurity breakdown in thin films*, Proceedings of the Symposium on Laser Induced Damage in Optical Materials. NBS Special Publication, No.568, 479–495.

Wood R (1990) *Laser Damage in Optical Materials*. Bristol and New York: Adam Hilger. Reprint with corrections.

Optical Black Surfaces

S M Pompea, National Optical Astronomy Observatory, Tucson, AZ, USA

S H McCall, Stellar Optics Research International Corporation (SORIC), Thornhill, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Black surfaces play an important role in many, if not most, optical systems. Optical instruments and systems of all types often rely on black surfaces to help minimize the effect that stray or off-axis light can have in degrading optical system performance. The strategic selection and placement of appropriate black surfaces can often limit significant detrimental stray light effects in an optical systems, thereby dramatically improving system performance. The high emissivity of black surfaces also gives them an important role in the design and construction of black bodies, calibration surfaces, and radiator surfaces.

The characterization and selection of black surfaces is an important field of optics and optical system design. However, the selection of black surfaces is a specialized undertaking requiring careful study to ensure the proper selection of particular surfaces for each optical system. A black surface that works well in one application may not be at all appropriate for a different application, where different systems performance goals are desired. The wavelength of operation of the system is an important consideration since a surface which has low reflectance at one wavelength band may have different reflective properties at another wavelength or range of wavelengths. Specialized measuring devices called goniophotometers are often used to characterize the reflectance and scatter of black surfaces for a given wavelength or set of wavelengths. Goniophotometric measurements can be used to create a specialized functional description of reflectance and scatter off a surface, called the Bidirectional Reflectance Distribution Function (BRDF) or Bidirectional Scatter Distribution Function (BSDF). In practice, having the BRDF or BSDF of a surface is extremely valuable for surface selection and is more useful than other reflectance and scatter measurement measurements or descriptions. Comparing the measured BRDF or BSDF of several surfaces helps to distinguish the general optical properties of these surfaces and allows a direct comparison of the value of these different surfaces in an optical system as they contribute to overall system performance.

The BRDF distinguishes, at a particular wavelength, how a surface specularly reflects or how it scatters light in different directions when incident light from a particular direction relative to the surface (at a given angle of incidence) interacts with the surface. By understanding the BRDF of a surface, the optical designer can understand how the surface behaves in a given system location when illuminated from a specific direction. The BRDF of a black surface is also important because it can be used as a mathematical function in stray light analysis programs to predict system performance. The BRDF and other properties of surfaces can be input into stray light codes so that the propagation of light from black surfaces can be modeled through Monte Carlo-based propagation models. Using these models, the stray light can be directed through the placement of black surfaces into directions less likely to degrade system performance. Thus a measure of stray light control can be gained through absorption by black surfaces or by redirection of the stray light into less critical directions.

Black surfaces may be considered a subset of a class of surfaces known as spectrally selective surfaces. The term 'spectrally selective surface' indicates that the spectral properties of many surfaces are different when examined in various spectral regions. For applications such as the design of blackbodies, space radiators, and baffles, surfaces with unique spectral emissive and reflection properties are needed. Large databases of spectrally selective surfaces are available to allow the optical designer flexibility in choosing surfaces for applications where specific optical reflectance and optical scatter properties

are desired. Spectrally selective surfaces are important for applications that range from solar energy to telescope design. In some applications, it is desirable for a surface to be highly absorbing at one wavelength and more reflective at another, rather than to be highly absorbing of light over a wide range of wavelengths. An analysis of the spectral properties of a large number of surfaces allows the optical designer to choose surfaces optimized for a particular application.

Black surfaces come in many types, but may be generally characterized as being more shiny (specular) or more diffuse (Lambertian) in their reflective and scattering properties, depending on the form of their BRDF curve. Surfaces are often described in computer models as a combination of their respective specular and Lambertian properties for a given wavelength. Many black surfaces have significant reflectivity at some region of the electromagnetic spectrum, even if they are highly absorbing at one wavelength. Some black surfaces are constructed to be remarkably nonreflecting over a range of wavelengths; these surfaces are often used as baffle surfaces for telescopes and instruments that must operate over a wide range of wavelengths. No black surface can work well in all optical systems. Optical designers must understand the specific optical properties of a black surface to determine if the surface can work in their particular system. To characterize a surface, it is helpful to know at least the BRDF of the surface at the wavelength of interest as well as basic materials properties of the surface. The specific BRDF at the most appropriate wavelength for the system can be used with the optical system stray light model to determine if the surface can fulfill the designer's, expectations of it for the system performance. The placement and choice of black surfaces in an optical system are often critical decisions with highly significant effects on overall system performance.

For this reason, optical designers rely greatly on databases of the reflection and scatter properties of black surfaces in order to make the right choice of surface. These databases have been created to provide the optical engineer with the information needed to understand a wide variety of candidate surfaces and select the most appropriate ones. These databases not only describe the optical properties of surfaces but also elucidate material properties so that the selection process can examine the heat resistance, outgassing rates, fade resistance, durability, and a myriad of other surface or material properties that may be relevant.

The reflectance and scatter data on black surfaces are often taken using a variety of reflectance-measuring instruments and the data format often varies greatly. When published and unpublished data sets are compiled, they are often confusing and sometimes even contradictory. This has made the comparison of surfaces a difficult proposition. In recent years, there have been successful attempts to gather existing large amounts of existing data so that comparisons among black surfaces can be made in a data-rich environment so that anomalous data can be ignored. A significant amount of reflectance data on black surfaces has been created by private or governmental organizations such as NASA and NIST – the US National Institute for Standards and Technology. These data now appear in databases of BRDF and reflectance measurements. Designers are relying increasingly on databases of BRDF surfaces rather than the individual measurements that can be found in the optical physics and applied optics published literature.

Uses of Black Surfaces

Table 1 lists some of the major areas in which black surfaces or black coatings may be used in instruments and optical systems. The optical properties of desirable surfaces may be different for the specific categories in the table. Plausible surfaces for these different applications can be found in materials databases such as the SOLEXIS™ database which contains a useful compilation of data from private

Table 1 Possible uses for black surfaces and black coatings in optical systems

Apertures	Baffles	Barrels	Calibration screens	Choppers
Cold shields	Detectors	Dewars	Diffusion screens	Domes for telescopes and instruments
Enclosures and curtains for rooms and instruments	Integrating sphere enclosures	Lens edges	Light traps	Mounts
Radiators	Radiometers	Solar collector absorbers and concentrators	Standards and references of reflectance	Targets

and governmental data compilations from the US, Canada, and the former Soviet Union.

Selection of Optical Black Surfaces

The selection of appropriate black surfaces for a given application becomes progressively harder as the number of optical requirements (e.g., reflectivity, scatter, emissivity, etc.) and materials requirements (e.g., stiffness, coefficient of thermal expansion, etc.) increase. Because of this, the selection process for specialized surfaces for space applications can be quite involved, since there are a large number of criteria, each of which may be regarded as an essential requirement.

The optical system designer must consider at least the following aspects of the system in selecting a surface:

- the wavelength(s) and waveband(s) of interest;
- the position of the surface in optical system;
- the general robustness required of the surface;
- the environment of the optical system;
- the cost including inspection, handling, and assembly; and
- the production and delivery schedule, including spares.

The selection process must carefully address the system needs as well. Poor decisions are often made when only a cursory study of potential black surfaces is made or when databases and the literature on the subject are neglected. Therefore, an important part of the selection process is the management of the timing of the selection, in order to give the designer time to access available information and to influence the system design in a timely manner.

If black surfaces are used for stray light control, then the use of these surfaces must be planned at the beginning of the optical system design process. If the placement of black surfaces and the array of potential surface treatments are examined early in the program, the optimal choices can often be determined in a careful manner that enhances the system performance. If the selection is saved until the very end of the program, the choice of surfaces will be more limited and improvements in system performance may be minimal, even with the most excellent black surfaces. This is because the location of the surfaces in the system often proves to be as critical as the choice of surface. If the surface is in the wrong location, the best optical black may not be able to improve the stray light performance. For best results, the layout of the system and the placement of baffles must be considered first, before the baffle surfaces are selected.

An optimal flowchart of good system design for black surface selection is given in Figure 1. Figure 2 describes a work flowchart where black surface selection is neglected until the end of the program, often with dire consequences for system performance, cost, and schedule.

For space-based systems, the effect of atomic oxygen in a low-earth orbit environment is an important consideration. For space systems, the designer must also examine how black surfaces may be affected by or contribute to contamination and outgassing. Other considerations are the generation of particles during launch, the effect of solar

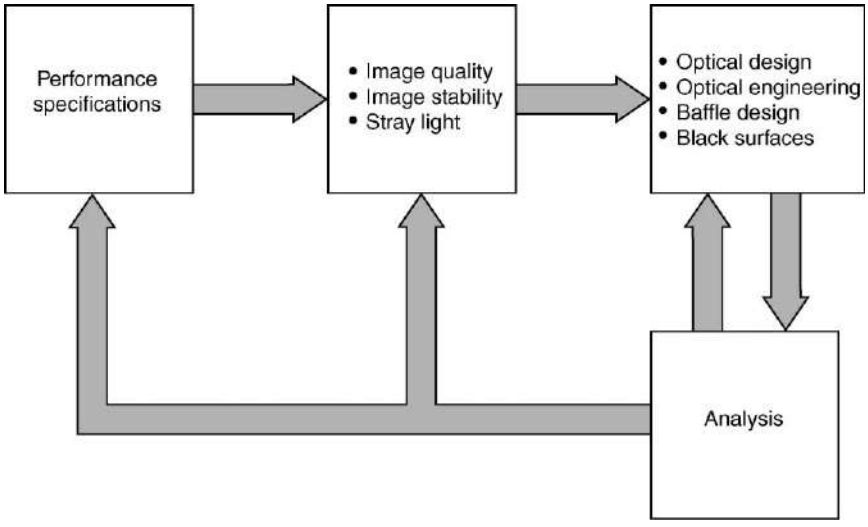

Figure 1 Diagram of the system design process which treats black surface selection as an important part of basic system design, to be done early in the design process. This gives the designer an opportunity to conduct system tradeoffs to optimize system performance.

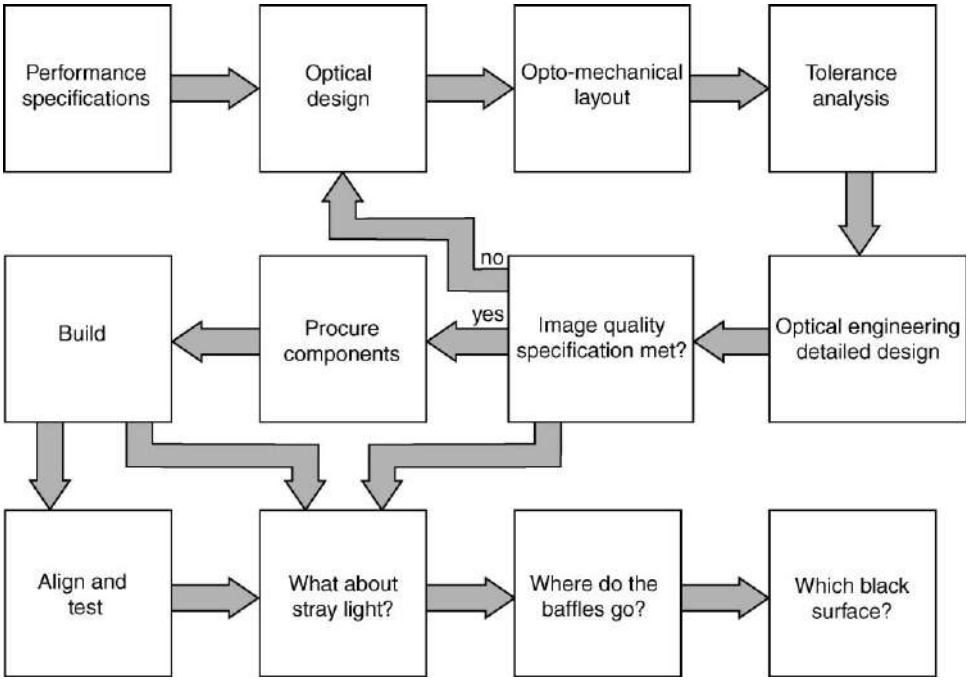

Figure 2 Diagram of a poor system design process. This process addresses the selection of black surfaces late in the system design process, where selection is significantly more difficult, expensive, and risky. Even with a late selection, there may be severe impacts on schedule.

ultraviolet radiation on black surfaces, and how spacecraft charging may affect conductive, dielectric, or insulating black surfaces. Since black surfaces may have a large surface area in an optical system, the designer must ensure that the black surfaces do not contribute significantly to the risk of system failure through some degradation process that may contaminate other optical surfaces. For space-based systems, data on surface outgassing and the past performance of a surface in a specific space environment is essential.

There are a variety of resources available to assist in the selection of black surfaces. The best databases allow the user to create 2-dimensional and 3-dimensional BRDF plots for extensive lists of potential black surfaces. For most surfaces, with no preferred directionality in the structure of the surface, a 2-dimensional BRDF plot is adequate. Most of these databases provide information on a wide variety of spectrally selective materials including black, white, and reflective surfaces for thermo-optical applications for ground and space-based instrumentation. The BRDF plots can be examined by the user or fed into existing, commercially available optical design and analysis software packages. The results of these models indicate whether the chosen surfaces give adequate performance or whether more specialized coatings are necessary to achieve the desired level of performance.

A significant problem for many designers is that the formulations of many paints and other surface treatments often change. Sometimes the name of the coating remains the same even as the optical properties change, due to a change in the compounds or methods used to prepare the coating. Sometimes the name changes with or without a change in the formulation. These often unpublished changes increase the risk of using older plots from the published literature as definitive optical modeling data. Experts in the optical blacks area continuously investigate these changes and their implications and can advise optical designers of the implications of recent changes in the paint and coating arena. These experts are often aware of situations where coating failures have occurred or where mission performance was degraded by unanticipated changes in coatings, even if reports of this incidents were never published.

In summary, the selection process, by its very nature, must be a deliberate and careful effort to examine the best surface options. Some of the better surfaces which were considered laboratory standards (e.g., 3M Black) are no longer easily available and replacements must be found. For the most systematic decision-making process, access is needed to the best optical black surface databases, often commercial databases with hundreds to thousands of BRDF curves on several hundred materials. The use of

these optical surface databases can substantially reduce the costs, time, and scope of a measurement program, as well as streamline the selection process. Additionally it allows the designer to consider materials used successfully in other programs which may not be familiar to the designer. Finally, the designer may avoid using materials that have been rendered unreliable through changes in the manufacturing process or in the coating chemistry.

Characterization of Black and Spectrally Selective Surfaces

Optical Characterization

An understanding of black surfaces relies heavily upon an understanding of the meaning and use of the BRDF. While the optical properties of surfaces can be characterized through specular and diffuse reflectance measurements, the use of BRDF measurements has become extremely important. BRDF measurements are equally adept at characterizing highly specular and highly diffuse surfaces and the myriad of surfaces with both characteristics. The BRDF is often used in general to describe the angle-dependent optical scatter from these surfaces. This is important as the angular distribution of scatter can be used to calculate if the scatter from a black surface will be a limiting factor in the resolution or noise level of the system. BRDF measurements can even be used in computer graphics visualization programs that illustrate what a surface may look like under various illumination conditions. For example, the BRDF is used in computer graphics models to render a certain color for an automobile or a countertop in a kitchen. For some black surfaces, the angular distribution of scatter can be used to calculate a variety of surface parameters such as the surface roughness, given certain assumptions about the surface. For opaque black surfaces, BRDF measurements have proved to be one of the best overall descriptors of the optical properties of the surface.

BRDF is defined as the ratio of scattered radiance (watts/(cm^2 sr)) to surface irradiance (watts/cm^2); its units are inverse steradians. Radiance is used in the definition, in order to make the BRDF independent of the parameters of the measuring instrument, such as the detector aperture and distance to the detector. BRDF measurements are made by a variety of instruments for a number of wavelengths. For example, BRDF measurements of surfaces have been made in wavelengths ranging from the vacuum ultraviolet to the far infrared. The most common BRDF measurements made on black surfaces are made at laser wavelengths of 0.6328 and 10.6 micrometers. However, these measurements may be of limited value if the system is operating at substantially different wavelengths, where the surface's BRDF is significantly different.

To make a BRDF measurement requires a light source, a sample mounted and illuminated by that source, a receiver (detector) to measure the scatter from the sample, and the computer/electronics package to accurately record the detection of the scattered light. Additionally the sample and/or light source must be able to be moved so that the sample can be illuminated from nearly every angle, and scattered light from the surface can be detected in a hemisphere around the sample. The geometry of the BRDF measurement is illustrated in Figure 3.

To make a BRDF measurement on black surfaces, appropriate sources, detectors, and low-scatter optics must be used. This becomes more difficult for wavelengths less than about 0.25 micrometers. At longer, infrared wavelengths, diffraction effects become important. Despite the instrumental difficulties, it is often better to measure the amount of scatter at a specific wavelength λ and angle θ_s, rather than try to predict it from theoretical models. This is because the extrapolation to new angles or wavelengths using scaling theories is not always accurate or even possible. Methods for wavelength scaling (extrapolating from BRDF data for one wavelength to data for any other wavelength) or angle scaling (where data taken at particular incident and scatter angles are extrapolated to other angles) will not be discussed as BRDF curves are now widely available for many surfaces at a variety of wavelengths and angles.

The definitions of BRDF are given in Table 2.

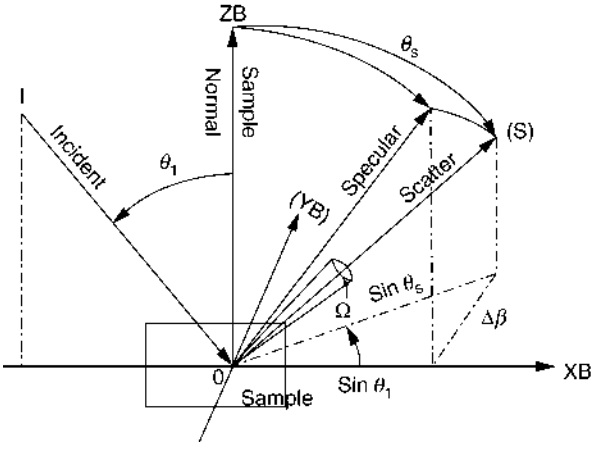

Figure 3 Geometry of BRDF measurement and the angle conventions accepted by the American Society for Testing Materials (ASTM). The plane of incidence is the I-0-ZB plane and the scatter plane is the S-0-ZB plane.

Table 2 BRDF term definitions

Bidirectional reflectance distribution function BRDF	The sample radiance divided by the sample irradiance. The procedures given in this practice are correct only if the field of view (FOV) determined by the receiver field stop is sufficiently large to include the entire illuminated area
Scatter	The radiant flux that has been redirected (by interaction with the sample) over a range of angles
Sample coordinate system	A coordinate system is fixed to the sample and used to a specify position on the sample surface for the measurements. The sample coordinate system may be application and sample specific. A Cartesian coordinate system is recommended for flat samples, with the origin at the geometric center of the sample face with the Z axis normal to the sample
Plane of incidence, PLIN	The plane defined as containing the central ray of the incident flux and the sample normal
Beam coordinate system, XB, YB, ZB	A Cartesian coordinate system defined with the origin on the central ray of the incident flux at the sample surface, the XB axis in the plane of incidence (PLIN) and the ZB axis normal to the surface. The angle of incidence, scatter angle, and incident and scatter azimuth angles are defined with respect to the beam coordinate system
Incident direction	The central ray of the incident flux specified by θ_i and φi in the beam coordinate system
Angle of incidence, θ_i	Angle between the central ray of the incident flux and the ZB axis
Incident azimuth angle, φ_i	The fixed 180° angle from the XB axis to the projection of the incident direction onto the XB–YB plane. It is convenient to use a beam coordinate system in which $\varphi_i = 180°$ since this makes φ_s the correct angle to use directly in the familiar form of the grating equation. Conversion to a sample coordinate system is straight forward, provided the sample location and rotation are known
Scatter direction	The central ray of the scattered flux specified by θ_s and φ_s in the beam coordinate system
Scatter angle, θ_s	Angle between the central ray of the scattered flux and the ZB axis
Scatter azimuth angle, φ_s	Angle from XB axis to the projection of the scatter direction onto the XB–YB plane
Specular direction	The central ray of the reflected flux that lies in the PLIN with $\theta_i = \theta_s$ and $\varphi_s = 0$
Delta theta, $\Delta\theta$	The angle between the specular direction and the scatter direction
Delta beta, $\Delta\beta$	The projection of $\Delta\theta$ onto the XB–YB plane, that is, the $\Delta\theta$ angle measured in direction cosine space. For scatter in the PLIN, $\Delta\beta = \sin\theta_s - \sin\theta_i$. For scatter out of the PLIN, the calculation of $\Delta\beta$ becomes more complicated
Receiver solid angle, Ω	The solid angle subtended by the receiver aperture stop from the sample origin
Sample radiance, L_e	A differential quantity that is the reflected radiant flux per unit projected receiver solid angle Ω per unit sample area. L_e is often an average calculated from scattered power, P_s collected by the projected receiver solid angle $\Omega \cos\theta_s$, from the illuminated area, A. The receiver aperture and distance from the sample determines Ω and the angular resolution of the instrument
Sample irradiance, E_e	The radiant flux incident on the sample surface per unit area. In practice, E_e is an average calculated from the incident power, P_i, divided by the illuminated area, A. The incident flux should arrive from a single direction; however, the acceptable degree of collimation or amount of divergence is application specific and should be reported

Table adapted from ASTM E1392-90 "Standard Practice for Angle Resolved Optical Scatter Measurements on Specular or Diffuse Surfaces," American Society for Testing Materials, 1990.

Given these definitions, BRDF, is mathematically described by

$$\mathrm{BRDF} = \frac{L_e}{E_e} = \frac{(P_s/\Omega A \cos\theta_s)}{(P_i/A)} = \frac{P_s}{P_i \Omega \cos_s} \ [\mathrm{sr}^{-1}]$$

The BRDF is a function of many measurable quantities. It depends on the angle of incidence of the incident flux and the wavelength of this radiation. It also depends on the direction in which the scattered light is measured as well as the wavelength and polarization of the incident and scattered flux. The BRDF can be calculated for a given angular configuration of incident source and detector from the average radiance divided by the average irradiance. The BRDF is therefore referenced by the incident power. The $\cos\theta_s$ term in the formula accounts for the foreshortening of A as the angle θ_s increases, so that true radiance is being used in the calculation. Because of this, the BRDF is constant for a surface with Lambertian (uniformly diffuse) properties. The detector must have a field of view so that all of A can be viewed and such that P_i includes the entire beam. In practice, P_i is measured by removing the sample from the measurement position. Then the detector is moved into the straight-through position. Often a reference detector is used to monitor the incident beam power so that any changes after the initial V_i, measurement can be recorded.

For BRDF measurements to be meaningful, the instrument signature must be examined and understood. This is particularly important in the measurement of low-scatter surfaces. The instrument signature is the BRDF when the sample is removed and there is no sample scatter. In this case the stray light and scatter signature of the device will be apparent. Angular positioning errors and electronic noise can also be important. For example, as θ_s approaches 90° the accuracy of measuring θ_s becomes critical due to the 1/cos θ_s term in the BRDF equation. Ideally the instrument signature is a small fraction of the BRDF being recorded. An understanding of the instrument signature provides the needed information to decide where the BRDF data loses its meaning.

The BRDF is most often displayed in a plot, where the BRDF value (vertical axis) and the angle of scatter (horizontal axis) are plotted on logarithmic scales. In these 'log–log' plots it is easy to examine the scatter distribution close to the specular direction which, for shiny surfaces, represents the bulk of the scattered light. For surfaces that are diffuse in nature, a log–log plot of the BRDF generally yields a straight line. The slope of this plot is characteristic of the surface of the process which is used to produce the surface. This straight-line plot is useful in comparing samples made with the same process. Deviations from the slope of the line or the value of the points on the line indicate that there may be problems with a sample. Thus the simplest scattering instrument can be used for quality control of black samples. For a fixed angle of incidence and a fixed detection angle, variations in the BRDF value may indicate changes in the materials or the processes which created their surfaces. The quality control process can be performed by looking for variations in the BRDF among the different samples.

The BRDF can be used as a measure of surface roughness – an important attribute for a diffuse black surface. As random surface roughness features increase on a surface, the scatter from the surface also increases. Because of this, the BRDF can be used to characterize surface roughness parameters. If the surface roughness is all in the form of small, micro-rough surface features, the BRDF value then is very much proportional to the surface power spectral density function (PSD), which is a measure of the power plotted against spatial frequency. The characterization of surface roughness by BRDF scatter measurements only works if the scattered energy is due to surface micro-roughness. If the scatter is from larger surface features or from surface contamination, then it will be difficult to accurately characterize the surface roughness. Thus only smooth, mirror-like black surfaces can be used for this simple leap from BRDF values to surface roughness parameters. Some sample BRDFs of black surfaces are found in Figures 4–7.

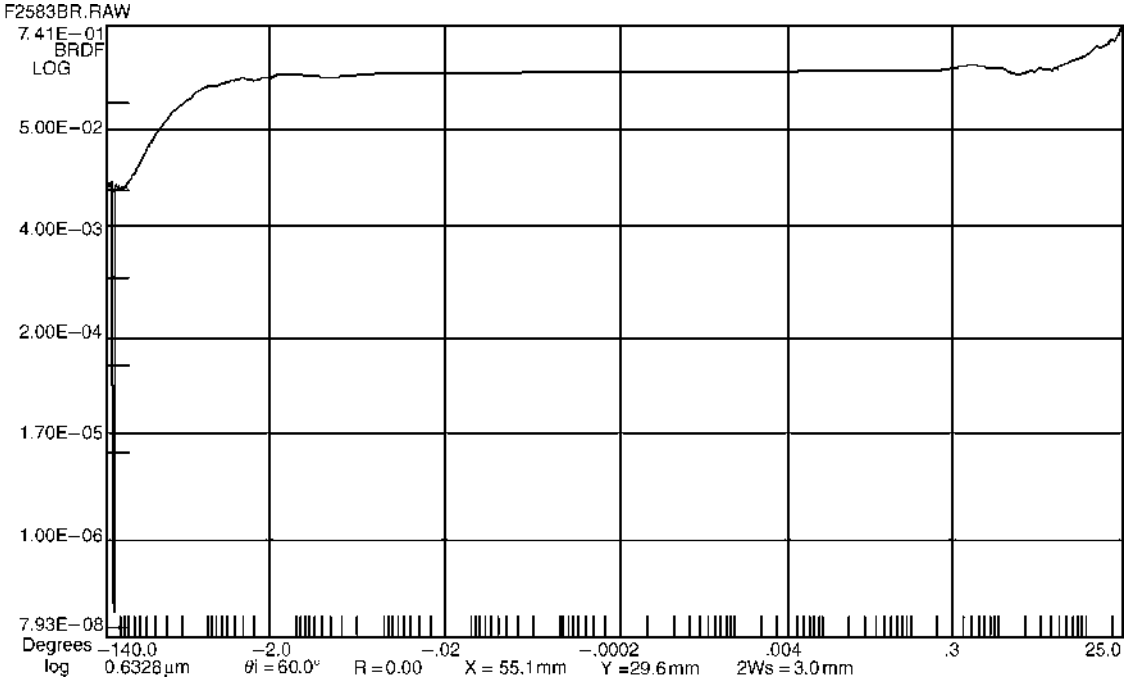

Figure 4 BRDF at wavelength of 0.633 micrometers and 60° incidence angle of black anodize sample. Plot generated by SOLEXIS using CASI software from Schmitt Measurement Systems, Portland, OR, USA.

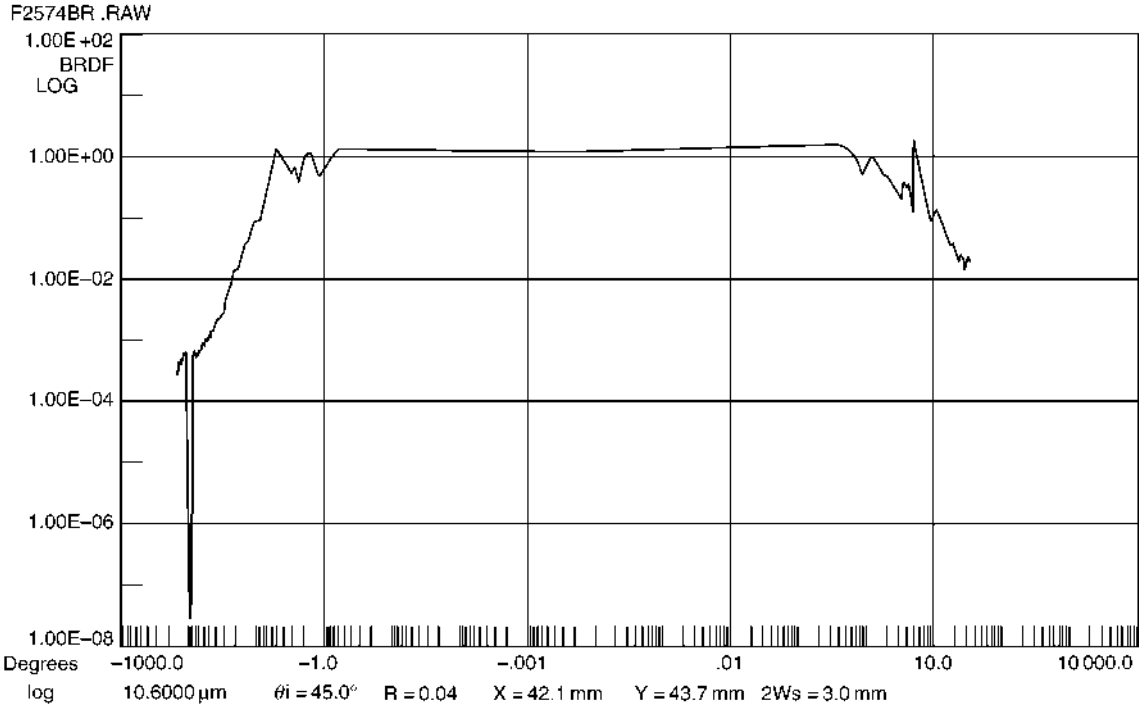

Figure 5 BRDF at wavelength of 10.6 micrometers of black anodize at 45° incidence. Plot generated by SOLEXIS using CASI software from Schmitt Measurement Systems, Portland, OR, USA.

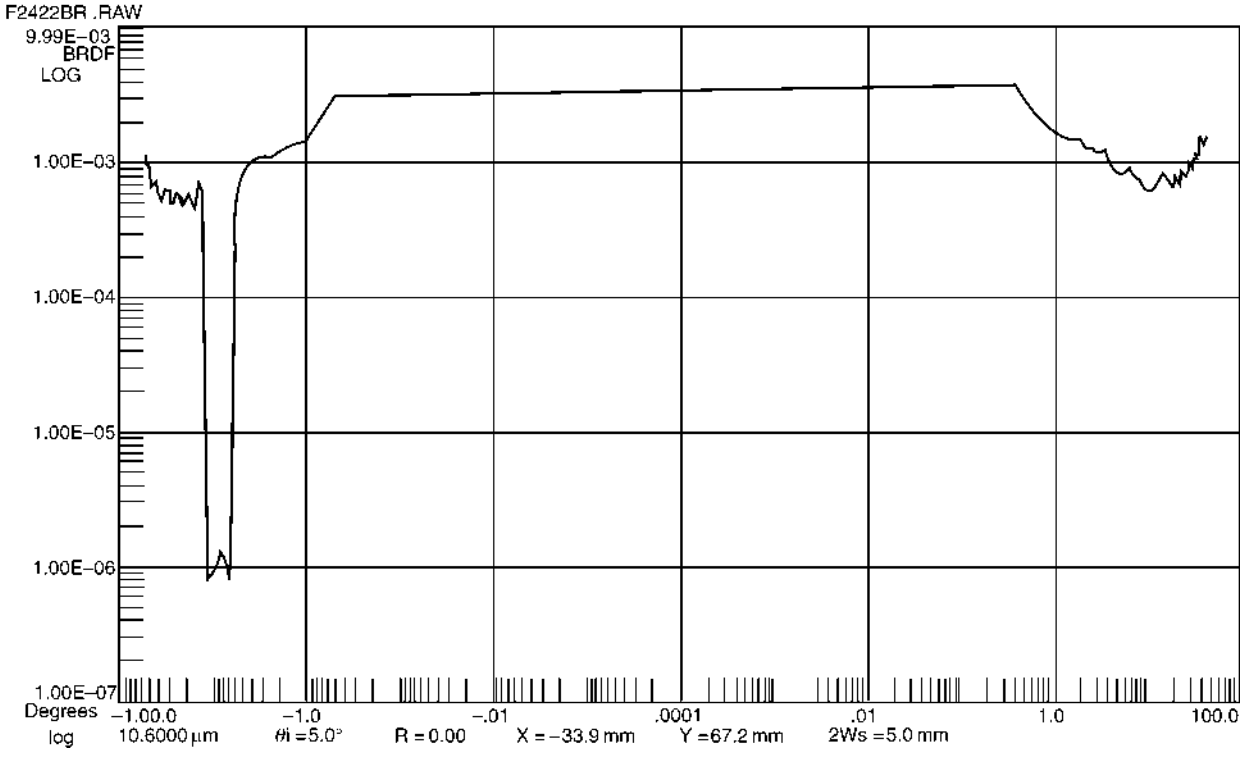

Figure 6 BRDF at a wavelength of 10.6 micrometer of Martin Optical Black at 45° incidence. Plot generated by SOLEXIS using CASI software from Schmitt Measurement Systems, Portland, OR, USA.

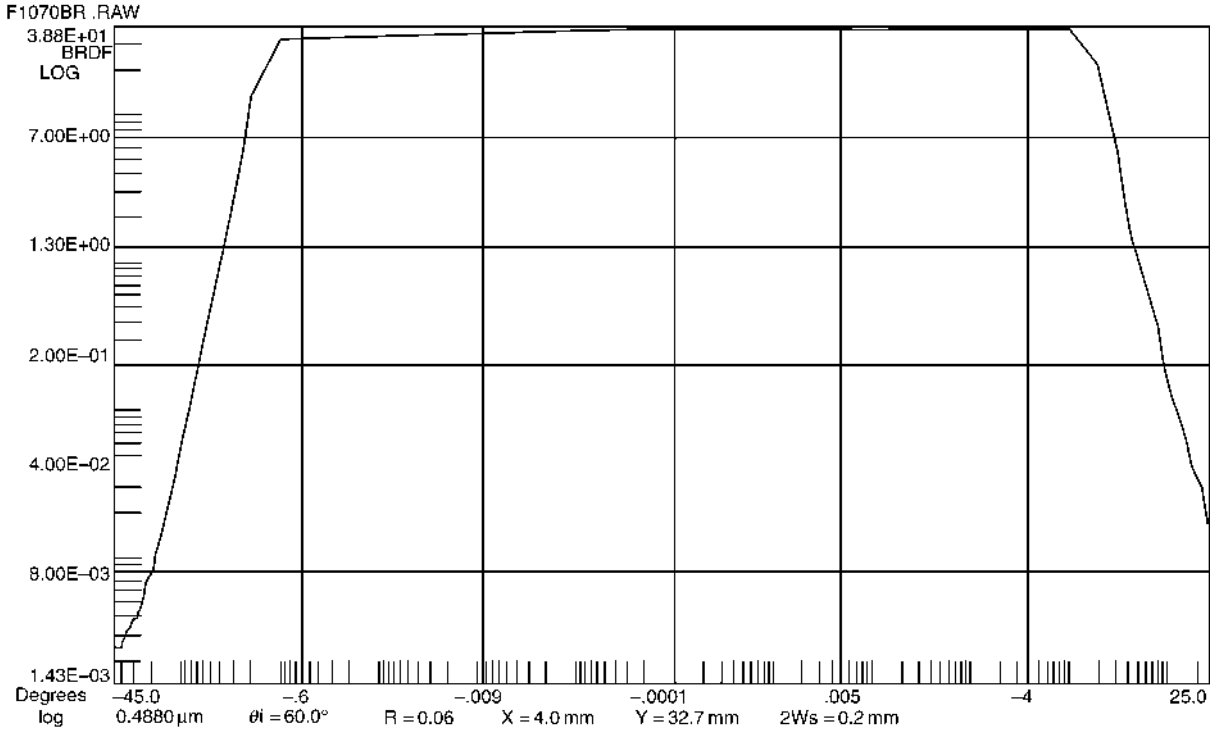

Figure 7 BRDF at a wavelength of 0.488 micrometers of a glossy black surface at 60° incidence. Plot generated by SOLEXIS using CASI software from Schmitt Measurement Systems, Portland, OR, USA.

Examples of Black Surfaces

Black surfaces appear in many forms. They may be either diffuse or specular in nature, or both. Black surfaces may be created by painting a surface, by altering a surface (such as through an anodization process), or by any number of other surface treatment effects. For example, absorbing compounds may be added to paint or to a surface to decrease reflection. Adding large (compared to the wavelength) craters or holes to a surface may result in multiple reflections within the surface leading to decreased reflection. Surfaces may be modified to increase their light-scattering ability, with the hope that reflection in a specific direction will be reduced. Sometimes surfaces may be created in such a way as to enhance thin-film interference effects, leading to reduced reflection. These and other techniques may be combined together in an attempt to reduce reflection and make the material appear blacker in a particular wavelength range of interest or in a particular direction.

Some sample black surfaces are given in Table 3, although databases are much more current on current names and manufacturers.

Many black surfaces use surface texture to enhance their light-absorbing ability. Scanning electron micrographs of surfaces reveal the surface morphology and how multiple reflections can be used for light trapping. Surface chemistry, changes in the index of refraction, and the use of particles of different sizes are also used to create better absorbers. A variety of other proprietary methods are used as well (Figures 8–11).

Future Developments

There are many new kinds of black surfaces being developed for specialized applications. Black surfaces can be created using a variety of new techniques and materials. For example, new surfaces can be created using laser ablation or chemical microetching of surfaces. Surfaces can also be created using new materials such as carbon nanotubes or by using the technology used to create flocked materials. These new kinds of surfaces generally have good absorption at a variety of wavelengths. Even with the creation of these new surfaces, the greatest gain in an optical system will probably be attained through a concerted effort to choose the appropriate black surfaces early in the program and to place them properly in the optical system. An early selection of surfaces and an analysis of their performance will lead to a higher level of system performance than a search for an exotic, ultrablack surface at a late stage

Table 3 Selected optical black surfaces

Name	*Manufacturer*	*Notes*	*Type*
Aeroglaze L300	Lord Corporation Erie, PA	Formerly called Chemglaze L300	Paint
Aeroglaze Z004	Lord Corporation Erie, PA	Formerly called Chemglaze Z004	Paint
Aeroglaze Z302	Lord Corporation Erie, PA	Formerly called Chemglaze Z302	Paint
Aeroglaze Z306	Lord Corporation Erie, PA	Formerly called Chemglaze Z306	Paint
Aeroglaze Z306 with microspheres	Lord Corporation Erie, PA	Formerly called Chemglaze Z306 with microspheres	Paint
Aeroglaze Z307	Lord Corporation Erie, PA	Formerly called Chemglaze Z307	Paint
Aeroglaze Z313	Lord Corporation Erie, PA	Formerly called Chemglaze Z313	Paint
Ames 24E (Ames 24E2)	NASA Ames Research Center Moffett Field, CA		Paint
Cardinal 6450	Cardinal Industrial Finishes South Elmonte, CA	Formerly 'Cardinal 6550'	Paint
Cornell Black	Cornell University		Paint
DeSoto Flat Black	Courtaulds		Paint
Electrically Conductive Black Optical Paint	Jet Propulsion Laboratory Pasadena, CA	Has no trade name	Paint
IITRI Bone Black D-111 (IITRI D111)	Illinois Institute of Technology Chicago, IL		Paint
LMSC Black	Lockheed Martin Palo Alto		Painted multilayer coating
MH21-I	Illinois Institute of Technology Chicago, IL		Paint
MH55	Illinois Institute of Technology Chicago, IL		Paint
MH2200	Illinois Institute of Technology Chicago, IL	Formerly 3M's ECP 2200 paint, but sold to IIT	Paint
Solarchem	Eastern Chem Lac Corporation Malden, MA		Paint
463-3-8	Akzo Coatings, Inc. Orange, CA	Formerly called 'Cat-a-lac 463-3-8' diffuse black paint	Paint
443-3-8	Akzo Coatings, Inc. Orange, CA	Formerly called 'Cat-a-lac 443-3-8' glossy black	Paint
443-3-17	Akzo Coatings, Inc. Orange, CA	Formerly called 'Sikkens 443-3-17'	Paint
Infrablack	Lockheed Martin Denver, CO	For Al substrates only	Anodization process
Martin Black	Lockheed Martin Denver, CO	For Al substrates only	Anodization process
Martin Black, Enhanced	Lockheed Martin Denver, CO	For Al substrates only	Anodization process
Infrablack	Lockheed Martin Denver, CO	For Al substrates only	Anodization process
Black Chrome Type Surfaces	Lockheed Martin Denver, CO	For many kinds of metal substrates	Electrodeposition process
Black Cobalt Type Surfaces (Cobalt Black)	Lockheed Martin Denver, CO and many more companies	For many kinds of metal substrates. References to Black Copper, Black Steel etc. are covered by Black Cobalt	Electrodeposition process, and can be followed by chemical or thermal oxidation
Black Nickel	–NIST	For many kinds of metal substrates	Deposition and etching process
(NBS Black)	–Ball Aerospace		
(Ball Black)	–Lockheed Martin and others	Ball modified the process developed by NBS	
Black Kapton™ Foil			Foil
Black Tedlar™ Foil			Foil
Boron Black	Lockheed Martin Denver, CO	For many kinds of metals	Plasma spray deposition process
Boron Carbide	Lockheed Martin Denver, CO	For Ti substrates only	Proprietary process
Silicon Carbide	Lockheed Martin Denver, CO and other companies	For many kinds of metals	Chemical vapor deposition
Textured Surfaces	–NASA Ames Research Center Moffat Field, CA –SPIRE Corporation Bedford, MA –Optics MODIL Oak Ridge National Laboratory	For many kinds of metal substrates	Either: –sputter coated –ion beam etched –sputter coated then etched

Note that finish names and manufacturers change regularly. Table adapted from SOLEXIS™ database, Stellar Optics Research International Corporation (SORIC), 78 Normark Dr., Thornhill, Ontario, Canada L3T 3R1.

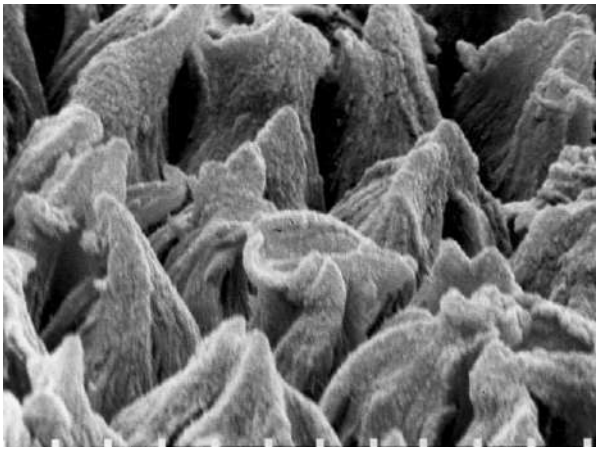

Figure 8 Scanning electron micrograph of Martin Optical Black surface on 6061T6 aluminum. This surface is a good absorber at ultraviolet, visible, and near infrared wavelengths. (Courtesy of Stephen Pompea and Donald Shepard.)

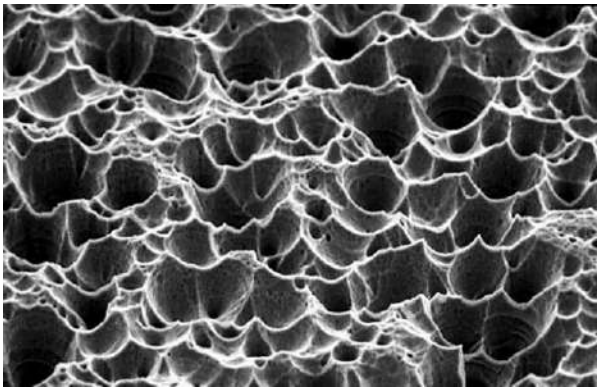

Figure 9 Scanning electron micrograph of phosphoric acid etched electroless nickel surface. This surface is a good absorber in the visible wavelength region. (Courtesy of Stephen Pompea and Donald Shepard.)

Figure 11 Scanning electron micrograph of tungsten hexafluoride surface. This surface is a good absorber at visible and near infrared wavelengths. (Courtesy of Stephen Pompea and Donald Shepard.)

in the program. Although better black surfaces are constantly being developed, the most important future development in the area of black surfaces may be a wider realization of the importance of an early and comprehensive surface selection process. The value of starting the selection process early using black surface databases has been confirmed repeatedly by optical designers.

See also

Instrumentation: Photometry; Telescopes. **Scattering:** Scattering from Surfaces and Thin Films.

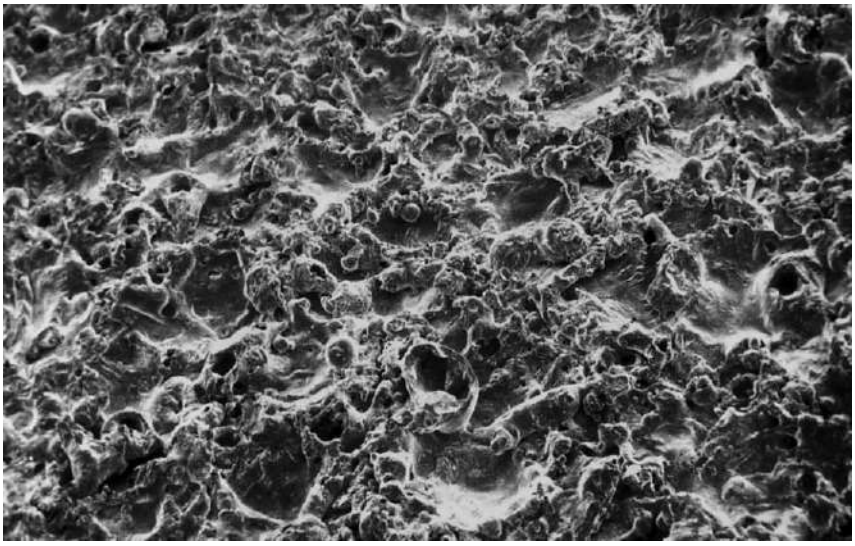

Figure 10 Scanning electron micrograph of Aeroglaze Z-306 painted surface with embedded microspheres. This surface is used primarily at visible wavelength. (Courtesy of Stephen Pompea and Donald Shepard.)

Further Reading

American Society for Testing Materials (ASTM) (1990) ASTM E1392-90. *Standard Practice for Angle Resolved Optical Scatter Measurements on Specular or Diffuse Surfaces*. West Conshohocken, PA: ASTM.

Bennett JM and Mattsson L (1989) *Introduction to Surface Roughness and Scattering*. Washington, DC: Optical Society of America.

Bohren CF and Huffman DF (1983) *Absorption and Scattering of Light by Small Particles*. New York: John Wiley and Sons.

Breault RP (1995) Control of stray light. In: *Handbook of Optics*, 2nd edn. New York: Optical Society of America.

Church EL and Takacs PZ (1995) Surface scattering. In: Bass M (chief ed.) *Handbook of Optics*, 2nd edn. New York: Optical Society of America.

Granqvist CG (1989). *Spectrally Selective Surfaces for Heating and Cooling Applications*. Volume TT 1. Bellingham, WA: SPIE Optical Engineering Text.

Gunderson JA (1977) *Goniometric Reflection Scattering Measurements and Techniques at 10.6 Micrometers*. Thesis, University of Arizona.

Henninger JH (1984) Solar absorptance and thermal emittance of some common spacecraft thermal control coatings. *NASA Reference Publication 1121*.

Jelinsky P and Jelinsky S (1987) Low reflectance EUV materials: a comparative study. *Applied Optics* 26(4): 613–615.

Persky MJ (1999) Review of black surfaces for space-borne infrared systems. *Rev. Sci. Instruments* 70(5): 2193–2217.

McCall SHCP, Pompea SM, Breault RP and Regens NL (1992) Reviews of properties of black surfaces for ground and space-based optical systems. In: Breault RP (ed.) *Proceedings of the SPIE: Stray Radiation II*, vol. 1753.

Pompea SM and Breault RP (1995) Optical black surfaces. In: *Handbook of Optics*, 2nd edn. New York: Optical society of America.

Pompea SM, McCall SHCP (1992) Outline of selection process for black buffle surfaces in optical systems. In: Breault RP (ed.) *Proceedings of the SPIE: Stray Radiation in Optical Systems* II, vol. 1753.

Pompea SM, Shepard DF and Anderson S (1988) BRDF measurements at 6328 Angstroms and 10.6 micrometers of optical black surfaces for space telescopes. In: Breault RP (ed.) *Proceedings of the SPIE: Stray Light and Contamination in Optical Systems*, vol. 967, pp. 236–248.

Smith SM (1980) Far infrared (FIR) optical black bidirectional reflectance distribution function (BRDF). *Proceedings of the SPIE: Radiation Scattering in Optical Systems* 362: 161–168.

SOLEXIS™ Database of BRDF Measurements (2003) Stellar Optics Research International Corporation (SORIC), Ontario, Canada.

Stover J (1995) *Optical Scattering: Measurement and Analysis*, 2nd edn. Bellingham, WA: SPIE Optical Engineering Press.

Wolfe WL (1978) Optical materials. In: Wolfe WL and Zissis GJ (eds) *The Infrared Handbook*. Ann Arbor, Michigan: ERIM.

Thin Film Optical Coatings

D Ristau, Laser Zentrum Hannover e.V., Hannover, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Thin-film optical coatings are employed to adapt the properties of an optical surface to its specific application. Commonly, optical coatings comprise a sequence of single layers of different transparent materials with a thickness in the range of the operation wavelength. Depending on this coating design, the spectral characteristics of an optical surface can be tuned to nearly any required transfer function including a suppression of reflection for a broad spectral range by an antireflective coating or even the achievement of highest reflectance values up to 99.999% using a high reflecting layer stack. Besides these basic functions, optical coatings can be employed to realize a variety of transfer spectra specified for multiple wavelengths or wavelength ranges to manipulate radiation according to even extremely sophisticated demands in optical systems for laser technology, communication, or fundamental research. Moreover, the desired optical properties can be combined with additional specifications, for example, concerning the mechanical or chemical stability of an optical surface by selecting adapted coating materials. As a consequence of this enormous versatility, thin-film optical coatings can be found in nearly every product and development of modern optical technology today. Aside from the large market of consumer optics including ophthalmology, camera objectives, binoculars, and optical systems for data storage, progress in optical coating technology is predominantly driven by demanding applications in many future technology fields. As one prominent pace-setter, semiconductor lithography imposes new demands on optical coatings extending the operation wavelength of stepper systems in chip production to the extreme ultraviolet (EUV)-spectral range. Other examples can be found in communication technology, where thin-film filters with a bandwidth below

0.5 nm are required, or also in laser material processing, where coatings with high power handling capability and lifetime are indispensable for the development of cost-efficient laser tools.

In this article a concentrated review will be given on the design, production, and on specific properties of optical coatings. After an introduction to the fundamental principles of operation of coating systems, basic expressions for the optical transfer function of a single layer will be derived and extended to layer structures on the basis of the matrix method. Selected conventional and ion processes for the production of coatings will be discussed with respect to the achievable optical and structural properties of the coatings. Also, the thickness control of the coatings during production will be considered introducing the crystal monitoring technique as well as optical measurement strategies at a single wavelength or in a broad spectral range. The article will conclude with a review of the present state of optical coatings and an analysis of future trends.

Design of Optical Coatings

The colorful appearance of optical coatings and microstructures, which can be observed on some butterflies and other insects, was present in nature long before mankind invented them. The beginning of the technical evolution of optical surfaces may be dated back to Greek culture, where polished metal surfaces were used as mirrors. Considering the application of transparent coatings, the investigations of Joseph Fraunhofer in the reflectivity of aged glass surfaces during the year 1819 should be mentioned as a first approach towards an antireflective coating. On the basis of controlled etching, this technique was further developed for different glass materials by Dennis Taylor in 1891 to realize the first antireflective coatings on lenses for applications in optical devices. In 1899, the Fabry–Perot theory was developed as a milestone in the theoretical description of multilayer structures. This theory describes the interference of partial waves reflected from two parallel optical surfaces, which can be considered as the fundamental arrangement of a multilayer structure (see Figure 1). In an optical thin-film system, layers of different transparent materials are composed on a stack deposited on the surface of the optical component (substrate refractive index n_T). In general, to adjust the spectral characteristics of the layer system to the demanded specifications, at least two layer materials with a high refractive index n_H and a low index n_L have to be selected. The choice of the thin-film materials is dependent on the application wavelength and the desired stability of the coating system as well

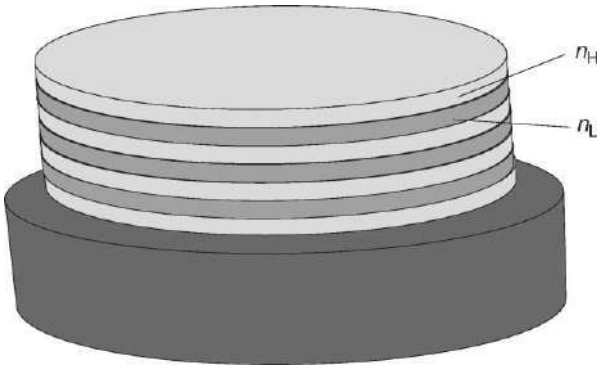

Figure 1 Basic structure of a thin-film system. Transparent layers of at least two different materials (n_H, n_L) are deposited on a substrate with refractive index n_T.

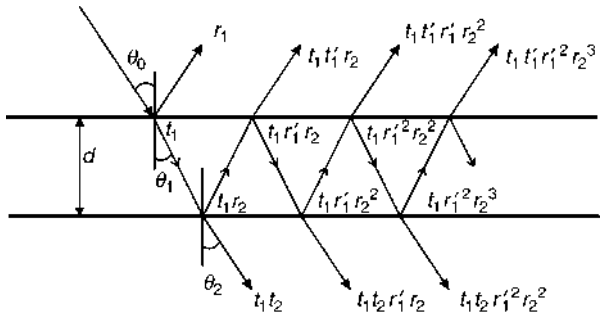

Figure 2 Schematic path of a plane wave incident onto a single layer with refractive index n_1. The reflection and transmission coefficients for the amplitudes at the interfaces $m = 1, 2$ are denoted by t_m and r_m, respectively. A prime added to the coefficients indicates a transfer of the wave opposite to the direction of incidence.

as on their optical properties essential for the specific design and the targeted optical quality. Similar to the Fabry–Perot principle, the optical function of the layer structure is based on the interference of partial waves reflected and transmitted at the interfaces between the layers. The trace of the incoming wave and the induced partial waves are depicted in Figure 2 for a single layer, which can be considered as the basic unit for coating systems even with highest complexity. For a theoretical description of the transfer function of this single layer, the contributions of the individual partial waves have to be calculated and accumulated to form the total transmitted and reflected wave. In this classic approach, the incoming wave can be described by the function of a plane wave with amplitude E_0, wavenumber $k = 2\pi/\lambda$ incorporating the wavelength λ, and frequency ω:

$$E(x, t) = E_0 \exp[-i(kx - \omega t)] \qquad [1]$$

At the first interface, a part of this plane wave is directly reflected with an amplitude of $A_0 = E_0 r_1$

(order zero), which can be calculated using the reflection coefficient r_1 given by the Fresnel formulas. After passing the first interface, a reflection and transmission of the incoming wave at the second interface has to be considered. Also a phase shift has to be taken into account, which is introduced by the path of the wave from the front to the rear interface of the single layer. This phase shift δ_i is dependent on the thickness d_i of the layer and the refractive index n_i:

$$\delta_i = \frac{2\pi n_i d_i \cos\theta_i}{\lambda} \qquad [2]$$

For oblique incidence also the angle of incidence θ_i in the layer, which has to be calculated by Snell's law, has to be taken into account. The index i has been chosen in this consideration for convenience to account for the layer number in coating systems, which will be discussed later. Following the wave reflected the first time inside the layer, a further phase shift for the path from the rear side to the front side has to be taken into account. This partial wave is then transmitted by the front surface resulting in an amplitude A_1 of the first-order wave leaving the layer:

$$A_1 = t_1 t'_1 r_2 E_0 \exp(-i2\delta) \qquad [3]$$

where t'_1 is the transmission coefficient for the amplitude of the wave passing the layer at the interface to the ambient. By following the wave further through the layer, the amplitudes A_k of each partial wave of order k can be determined:

$$A_k = t_1 t'_1 r_2 E_0 \exp(-i2\delta) * (r'_1 r_2 \exp(-i2\delta))^{k-1} \qquad [4]$$

Obviously the contributions of the partial waves follow the rule of a geometric expansion, and therefore the total amplitude of the wave reflected by the layer, which is the sum of the amplitudes of the partial waves, can be expressed in closed form:

$$A_{\text{tot}} = \left[r_1 + \frac{t_1 t'_1 r_2 \exp(-i2\delta)}{1 - r'_1 r_2 \exp(-i2\delta)} \right] E_0 \qquad [5]$$

The reflection coefficient r_S of the single layer is given by the ratio between the total amplitude of the reflected wave and the amplitude of the incoming wave:

$$r_S = \frac{r_1 + r_2 \exp(-i2\delta)}{1 + r_1 r_2 \exp(-i2\delta)} \qquad [6]$$

In this equation, the coefficients r'_1 and t'_1 have been replaced by using the relations

$$t_1 t'_1 = (1 - r'_1)(1 - r_1) = (1 - r_1)(1 + r_1) = 1 - r_1^2 \qquad [7]$$

In principle, this method of compiling the contributions of the partial waves reflected at the layer interfaces can be extended also to the calculation of systems with more than one layer. But, considering the enormously increasing number of partial beams, the resulting equations become extremely complicated for multilayer structures. Therefore, the matrix formalism, which can be deduced from the boundary conditions of the electric and magnetic field at the interfaces, is usually employed for the calculation of thin-film systems. The major advantage of this approach, which was developed in the 1940s, is the representation of each layer by a single matrix M_i containing all specific parameters of the film. The matrix M_i of a thin film at position i within the layer system relates the electric (E_{i-1}) and magnetic (H_{i-1}) field strength at the front face to the field strength values (E_i and H_i) at the rear face of the layer:

$$\begin{pmatrix} E_{i-1} \\ H_{i-1} \end{pmatrix} = \mathbf{M}_i \begin{pmatrix} E_i \\ H_i \end{pmatrix} = \begin{pmatrix} \cos\delta_i & \frac{i}{n_i}\sin\delta_i \\ i n_i \sin\delta_i & \cos\delta_i \end{pmatrix} \begin{pmatrix} E_i \\ H_i \end{pmatrix} \qquad [8]$$

By considering the boundary conditions for the field strengths at the interfaces, the matrix components given in eqn [8] can be determined. On the basis of the matrix formalism, the transfer function of a layer system can be simply accomplished by multiplication of the matrices M_i representing the constituent single layers. Thus, the transfer matrix M_S of a stack, which is formed by a number K of single layers with layer 1 located at the first interface with respect to the incoming wave, is given by

$$\mathbf{M}_S = \mathbf{M}_1 \mathbf{M}_2 \mathbf{M}_3 \ldots \mathbf{M}_i \mathbf{M}_{i+1} \ldots \mathbf{M}_K \qquad [9]$$

For calculating the reflection coefficient r_{SK} of the entire arrangement including the substrate (index of refraction n_T) and the ambient medium (index of refraction n_0), again the ratios of the amplitudes have to considered:

$$r_{SK} = \frac{n_0 \mathbf{M}_{11} + i n_0 n_T \mathbf{M}_{12} - i\mathbf{M}_{21} - n_T \mathbf{M}_{22}}{n_0 \mathbf{M}_{11} + i n_0 n_T \mathbf{M}_{12} + i\mathbf{M}_{21} + n_T \mathbf{M}_{22}} \qquad [10]$$

As a consequence of its perceptible structure, the matrix method can be readily transferred into computer codes, which cover the calculation of all important properties of even extremely complicated layer systems. In particular, considering the fundamental relations of the matrix method, also the field strength distribution within the layer structure can be determined. In software packages, which are commercially available as standard tools for optical thin-film technology, often the film thickness D_i of the layers is expressed in the units of QWOT (quarter wave optical

thickness) at the design wavelength λ_Z:

$$D_i = \frac{4n_i d_i}{\lambda_Z} \qquad [11]$$

In the case of normal incidence, the matrices M_i reduce to simple expressions for layers of integer QWOT values, because the corresponding phase shift terms are multiples of $\pi/2$. Clearly, an integer QWOT value of optical thickness also represents an extreme value in the spectral behavior at the design wavelength. This fact is often exploited for a precise thickness control by optically monitoring the reflectance or transmittance of the growing layer at the design wavelength λ_Z: by recording the transfer behavior as a function of time during production, extreme values and accordingly thickness values of multiple QWOT can be accurately determined. Thus, if a layer structure is composed only of multiple QWOT layers, switching between the layer materials can always be precisely performed at a minimum or maximum read-out of the optical monitor tuned to the design wavelength. In respect of this advantage in production, most coating designs are based on QWOT layers indicating the practical importance of the unit QWOT. For an efficient description of the designs, often a notation in capital letters is used, which indicate a 1 QWOT layer of a certain coating material. For example an alternating QWOT-stack with 11 layers of a high-index material (H, refractive index n_H) and a low-index material (L, refractive index n_L) is described by the sequence HLHLHLHLHLH or even more condensed by $(HL)^5H$. The QWOT stack is the fundamental design for dielectric high reflecting mirrors and represents the basis for a variety of functional coatings in optical thin-film technology. By applying the matrix method, an approximate expression for the reflectance R_S at the design wavelength can be derived for such a stack:

$$R_S \approx 1 - 4n_0 n_T \frac{n_L^{2N}}{n_H^{2(N+1)}} \qquad [12]$$

This expression is only valid for a large number N of QWOT layers and underlines the role of the materials for the practical production of optical coatings. Highest reflectivity with lowest number of layers can be achieved only by selecting materials with a high contrast of their refractive indices. This contrast also influences the spectral broadness of the high-reflecting region of the stack (Figure 3).

Besides highly reflecting mirrors, antireflective coatings are another important class of functional coatings for modern optics and laser technology. Antireflective coatings often involve non-QWOT layer structures of more than two materials in order to accomplish the demands of lowest residual reflectivity for a broad spectral range. Eye glasses with antireflective coatings covering the entire visible spectral range can be considered as a prominent example for consumer optics, which impose requirements on the coatings also in respect to their stability against cleaning and other environmental influences during normal use. In this context, the introduction of optical plastic materials is an additional challenge in consumer optics and ophthalmology, which has to be mastered by adapted designs and production processes. In contrast to the typical demands of consumer optics, antireflective coatings on laser windows are

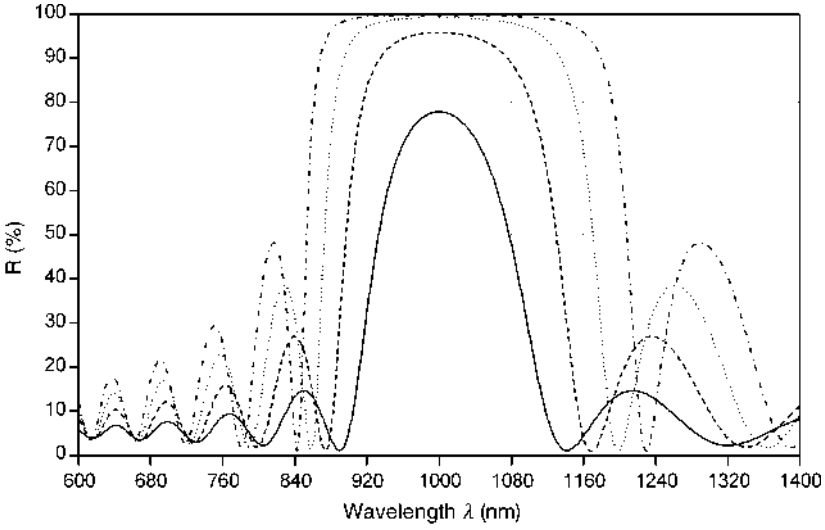

Figure 3 Spectra calculated for high-reflecting QWOT stacks (15 layers, substrate index $n_T = 1.52$) composed of a low-index material ($n_L = 1.46$) and different materials with high indices of refraction varying between $n_H = 1.7$ (solid), 1.9 (dashed), 2.1 (dotted) and 2.3 (dash-dotted). The spectral broadness and the reflectivity increase with the ratio n_H/n_L.

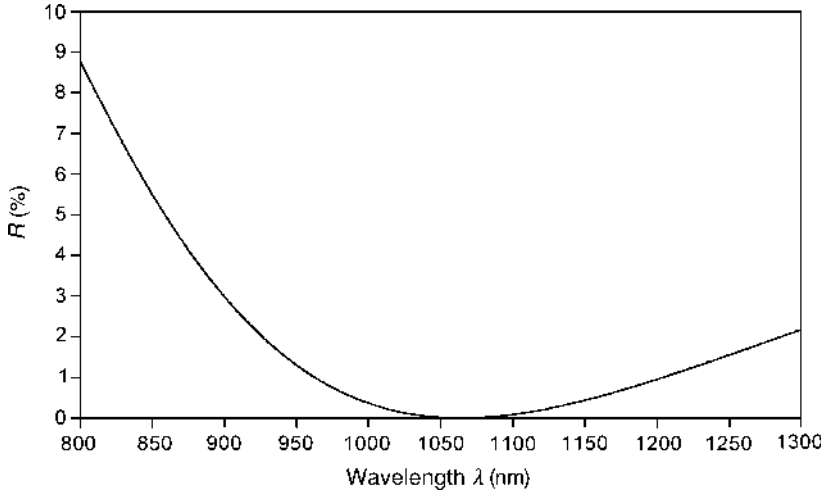

Figure 4 Spectrum of an antireflective coating for the wavelength 1.064 μm, which is designed as a V-coating with two layers of Ta_2O_5 ($n_H = 2.03$; $d_H = 0.383$ QWOT) and SiO_2 ($n_L = 1.46$; $d_L = 1.308$ QWOT) for a quartz substrate (substrate index $n_T = 1.46$).

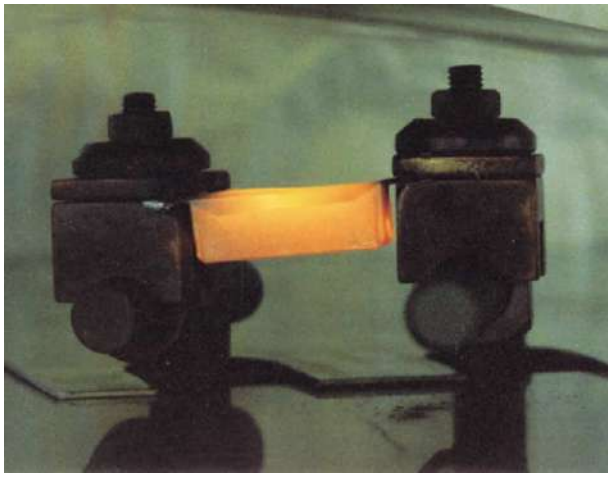

Figure 5 Boat evaporation source in operation. The deposition material is put into the boat which is made of a metal with high melting point. A high electric current is conducted through the boat for heating to temperatures up to 2000 °C.

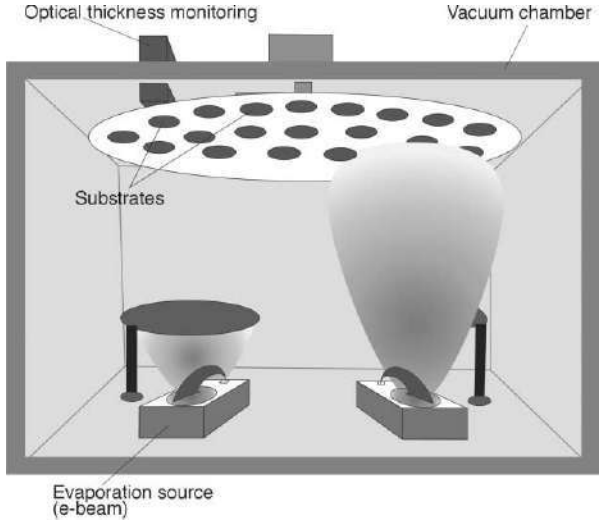

Figure 6 Outline of a modern box coater used for the production of optical coatings with e-beam deposition processes.

often restricted to one wavelength, but have to fulfill extreme requirements with respect to optical losses and power handling capability. Therefore, most antireflective coatings in laser technology are composed only of two thin layers to minimize the total thickness, which is often a direct measure of the expected losses of the layer system. With such a coating, which is often called V-coating according to its spectral characteristic (Figure 4), the reflectivity of the window surface can be suppressed ideally for one wavelength.

Production of Optical Coatings

The industrial production of optical coatings started in the 1940s on the basis of glass bell jar systems equipped with diffusion pumps and boat evaporators for the deposition of single layers of metals for mirrors and fluoride materials for antireflective coatings. Metal containers are used as boat evaporation sources. These containers are filled with the deposition material and heated by electric currents up to several 100 A (Figure 5). In the course of the rapid improvement of vacuum systems and especially the development of the electron beam evaporator over the next three decades, new materials with improved optical and mechanical quality could be combined to form more complicated dielectric coating systems. The layout of an e-beam deposition plant of these times, which can still be considered as fundamental for the advanced deposition systems of today, is illustrated in Figure 6. In such a box coater, the

substrates are installed in a calotte which is rotated to achieve a uniform distribution of the layer thickness. The deposition materials are put in water-cooled crucibles and are subjected to an electron beam operated at currents in the range between 0.1 A to more than 10 A and voltage levels of several kilovolts. As a consequence of the interaction with the energetic electrons, the deposition material can be heated to high temperatures above 5000 K allowing the evaporation of most of the desired deposition materials. These achievable high temperatures in conjunction with the good controllability are the major advantages of e-beam sources (Figure 7) in comparison with the classical boat evaporation process. In particular, e-beam evaporation opened the way for the deposition of oxide materials and other optical materials with high evaporation temperatures, which exhibit optical properties as well as a high mechanical and chemical stability superior to the material classes accessible with boat evaporation. However, the cost efficient boat evaporation technique is still applied today for the deposition of metals or temperature-sensitive materials for coating systems in the vacuum ultraviolet (VUV)/UV and mid-infrared (MIR) spectral ranges. If the shutter above the deposition source is opened, the evaporated material reaches the substrates and forms the layer by condensation. This condensation process is sensitively dependent on the environmental conditions of the substrates and the deposition sources. Sufficient microstructural properties of the growing layer can generally be achieved only under fine vacuum below 100 Pa and heating of the substrates to temperatures in the range of some 100 °C. Even under well-optimized conditions, the typical microstructure of such thermally evaporated coatings exhibits specific deficiencies (Figure 8) which influence the optical and mechanical properties of the layer systems. Often, microcrystallites with diameters in the range of a few nm are observed, which are arranged in columns of a typical size growing from the substrate interface to the outer surface of the layer. For a simulation of the growing mechanisms for this low energetic deposition process, uncomplicated models with a few input parameters already lead to satisfactory results. Assuming a kinetic energy of 0.1–0.3 eV, which is typical for thermal evaporation according to the Boltzmann distribution, the impinging atoms can move only a small distance or even adhere directly at their position of incidence on the growing layer. As a consequence of this small mobility, the atoms are not able to reach their energetically optimum positions in the layer, and they even obstruct the path of the following atoms by shadowing. On the basis of these assumptions, computer simulations of the growing process result in structures which are comparable to the columnar growth observed in practice. The packing density of such a columnar structure is significantly below the bulk values of the corresponding deposition materials, and it contains also a high number of defects and voids, which reduce the optical quality of the coating. In particular, the coating tends to

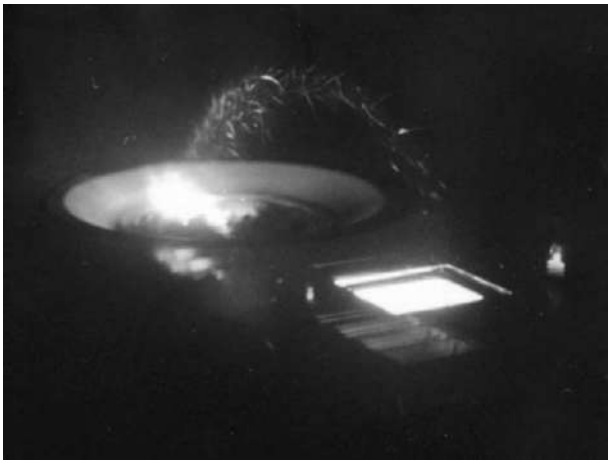

Figure 7 Electron beam source in operation. The deposition material is put in a water-cooled crucible and subjected to an e-beam. Since the zone heated by the e-beam is entirely embedded in the surrounding deposition material, very low contamination levels and temperatures above 5000 °C can be achieved. For the generation of the e-beam an arrangement similar to the system of a cathode ray tube is employed: the electrons emitted by a hot filament are focused by an aperture and accelerated by a high voltage.

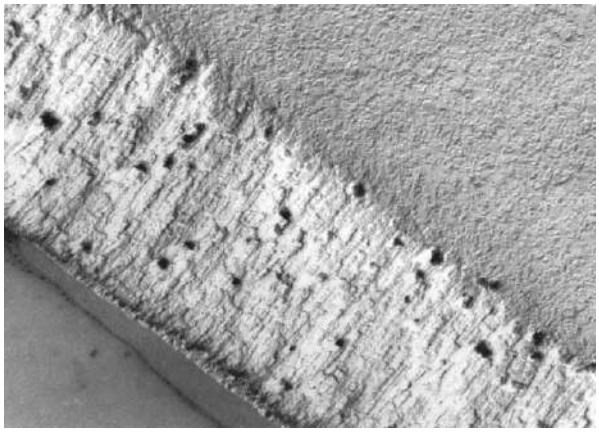

Figure 8 Microstructure of a MgF_2 single layer (geometric thickness approximately 780 nm) deposited by conventional boat evaporation. For the preparation of the TEM picture, the coating has been fractured and replicated.

adsorb water from the normal atmospheric environment into its internal structure leading to a shift of the spectral characteristics as a function of the ambient conditions. Besides this loss of spectral stability, the adsorbed water also induces an increased optical absorption of the coatings in the VUV and the MIR spectral ranges.

The increasing demand in laser technology and other optical applications for coatings far exceeds the quality level of conventionally deposited coatings and this was the driving force for the development of advanced concepts for the production of optical coatings during the last three decades. Considering the fundamental growth mechanism of conventional optical coatings, two basic approaches for an improvement of the microstructure are available. On one hand, the mobility of the atoms on the surface of the growing layer can be increased by the insertion of additional energy from other sources. This concept leads to the development of ion-assisted deposition processes, where the additional energy is transferred to the adatoms by collisions with ions generated directly in the plant by an ion source at energy levels of approximately 30–300 eV (Figure 9). Especially, by the deposition of oxide materials, which can be assisted directly with oxygen ions for balancing their stoichiometry, coating systems with high environmental stability and low optical losses can be achieved for the VIS up to the MIR spectral range. On the other hand, the energy of the adatoms can be directly increased in sputter processes which deliver deposition energies orders of magnitude higher than the thermal energy range. Besides adapted activated reactive magnetron sputter processes, particularly ion beam sputtering (IBS) was developed for a coating process delivering high-end coatings mainly for laser technology with an unsurpassed optical quality (Figure 10). For example, mirrors with total losses below one ppm can be produced for the Nd:YAG laser (wavelength 1.064 μm) with adapted IBS processes.

In addition to stable deposition processes, precise control of the layer thickness with an accuracy below 1 nm is a further important step towards the realization of complex optical coatings. Most deposition plants are equipped with quartz crystal monitors, which measure the optical thickness on the basis of the resonance frequency shift of an oscillating quartz crystal induced by the increase in mass of the deposited coating. The relative accuracy of these monitors ranges from 2% to 10% and is well suited for the regulation of the deposition rate. In order to achieve the required precision for high-quality coatings, optical monitors are often employed, which record the transmittance of the growing layer system at a single wavelength or even in a broad spectral range. Since these systems account directly for the optical properties of the deposited layer system, an efficient and accurate control of the film thickness is possible allowing the production of extremely sensitive filters with bandwidths below 1 nm for telecommunications or mirrors with defined group delay dispersion for fs-laser systems (Figure 11).

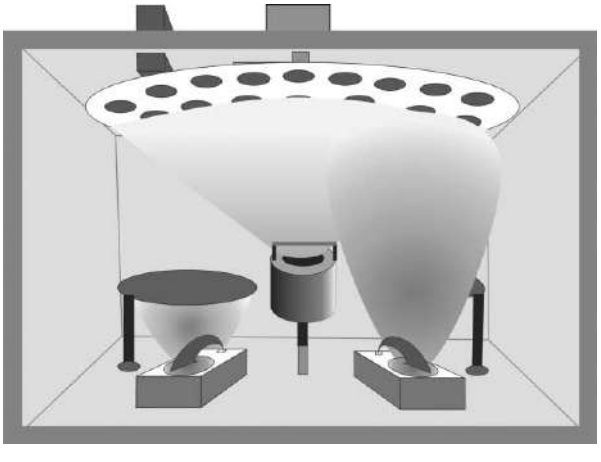

Figure 9 Ion assisted deposition. An additional ion source is installed, which delivers inert or reactive ions impinging on the growing layer. The energy transfer from the ions to the adatoms results in improved microstructure of the produced coatings with higher packing density.

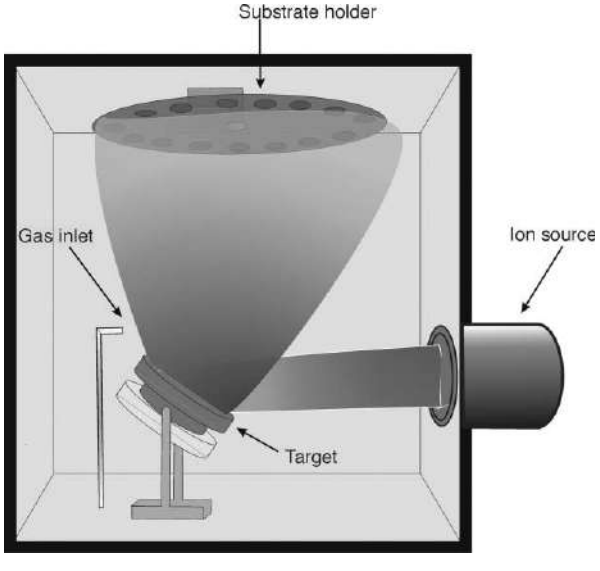

Figure 10 Principle of the ion beam sputtering (IBS) process. Inert ions with an energy in the range 1–2 KeV are generated in an ion source and directed onto the target which consists of the desired deposition material. Material is sputtered from the target and reaches the substrates. As a consequence of the high energy of the adatoms (5–30 eV) the formed layer exhibits a superior microstructure.

Quality Parameters of Optical Coatings

Besides the spectral characteristics, a variety of specific parameters has to be considered for each application of optical coating systems (Table 1). For laser systems the laser-induced damage threshold and the optical losses have to be evaluated in the context of technical and also economic requirements. Absorption is a loss channel, which transforms a fraction of the impinging radiation energy into heat, inducing a temperature increase in the coating. This generated temperature profile influences the imaging properties of the optical component via thermal expansion and temperature-induced alterations of the refractive indices of the coating as well as the substrate materials. Despite numerous technical problems, absorption always implies the loss of laser energy which is produced with high economic expense. Absorption losses are related to defects and impurities, which are introduced during the production process into the layer by a variety of contamination mechanisms. Furthermore, an unbalanced stoichiometry with a higher metal content is often observed as a cause for increased absorption in dielectric optical coatings, which are mainly composed of binary metal compounds. Absorptance values of optical components are determined by laser calorimetric techniques described in ISO 11551. In this measurement scheme, the temperature increase in the sample is monitored during irradiation with a well-characterized laser beam. After this heating phase, the laser beam is blocked and the temperature curve during the cool-down cycle is also recorded. On the basis of the measured temperature development, the absorptance can be calculated taking into account the laser power and the heat capacity of the specimen.

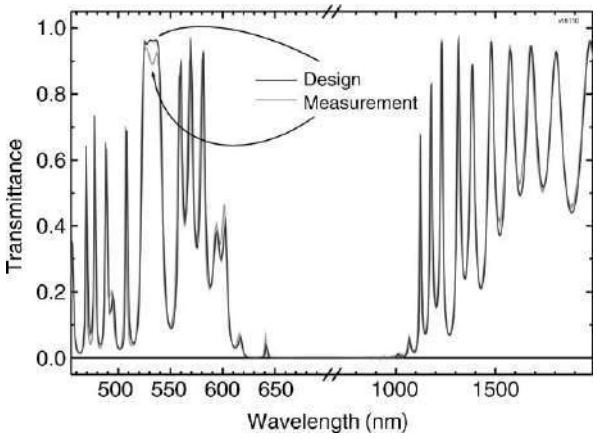

Figure 11 Example of a complicated layer system deposited under the control of an advanced optical broadband monitoring device: double chirped mirror (56 non-QWOT layers of SiO_2/TiO_2). Only small deviations between the calculated spectrum and the spectrum of the produced coating can be observed.

All mechanisms, which deflect radiation from its specular direction, contribute to optical scattering representing the second major loss channel in optical coatings. Since scatter losses are mainly attributed to microstructural imperfections and the surface roughness of the layers, a scaling with wavelength can be deduced from the corresponding theoretical models. Especially in the VUV/UV spectral range, high scatter values are apparent and may dominate the losses of the optical components. For longer wavelengths, the contribution of scattering is decreased to a point in the MIR spectral range where it can be neglected in comparison to the absorption losses. Scatter effects

Table 1 Selected quality parameters of optical coatings and surfaces in conjunction with corresponding ISO standards and measurement principles

Specification	*Parameter/Unit*		*Standard/Measurement principle*
Laser induced damage threshold (LIDT)	cw-LIDT	W/cm	ISO 11254-1: cw-laser irradiation
	1 on 1-LIDT	J/cm^2	ISO 11254-1: irradiation with single pulses
	S on 1-LIDT	J/cm^2	ISO 11254-2: repetitive irradiation with pulses
	Assurance	J/cm^2	DIS 11254-3: irradiation sequence
Optical losses	Absorptance	ppm	ISO 11551: laser calorimetry
	Total scattering	ppm	ISO 13696: integration of scattered radiation
Transfer function	Reflectance	%	CD 13697: precise laser ratiometric method
	Transmittance	%	DIS 15368: spectrophotometry
Surface quality	Scratch/Digs		ISO 10110: 13 parts containing different types of imperfections
	Roughness		
Stability	Abrasion		ISO 9211: different test methods
	Environmental Stability		ISO 9022: 21 parts containing a variety of conditioning methods

may impair the imaging properties or the resolution, and they reduce the efficiency of an optical system. Besides the economic aspect of each loss effect, scattering can also be considered as relevant for safety in high-power laser systems, where some 100 ppm of scattered radiation carry enough power to injure or blind the operator. In the ISO 13696 standard, measurement procedures based on the collection or integration of the total radiation scattered by the sample are described.

Depending on their application, coatings have to withstand also manifold mechanical, chemical, and environmental influences. Severe mechanical and chemical requirements are imposed onto coatings for ophthalmology or consumer optics, which are often cleaned with abrasive and aggressive cleaning solvents and tissues. To qualify coatings for these applications, various testing procedures have been developed and are routinely employed in industrial quality management. Coatings for laser systems and precision optics are predominantly certified according to conditioning methods described in the ISO standard series 9211. Test procedures with respect to environmental influences are extensively outlined in the more than 20 parts of ISO 9022, which is dedicated to a variety of climatic, mechanical, and chemical conditioning cycles.

Summary and Outlook

Optical coatings are the major tool for engineering the properties of functional optical surfaces for an extensive spectrum of applications. On the basis of the matrix algorithm, a comprehensive theoretical model for the spectral transfer characteristics of optical coatings is available, which has been transferred to commercial computer programs. Now, most design problems can be solved with these computer codes, whereas the reproducible and precise production of optical coatings with high quality could only be partially achieved until today (Table 2). Besides conventional thermal processes, new ion-assisted deposition processes and sputtering concepts are gaining importance in optical thin-film technology. For the quality management of optical coatings, a variety of often standardized testing procedures has been developed during the last two decades including the determination of the power handling capability, the measurement of the optical losses and a broad spectrum of conditioning methods for optical coatings.

Demanding future challenges will be imposed on optical coatings by the development of semiconductor lithography in the EUV spectral range to achieve nodes well below 100 nm in chip production. Another important trend is directed by new laser applications in material processing, medicine and communication technology, operating at extremely high repetition rates and power levels, or with ultrashort pulses in the fs-regime. Other requirements will be defined by the rapid development of optical technology towards a broad range of special products with short innovation cycles obligating significant progresses in thin-film technology in the direction towards flexible and reproducible process concepts

Table 2 Selected quality parameters of optical coating systems for laser applications

Laser, wavelength	*Type*	*Absorption ISO 11551*	*Total scattering ISO 13696*	*Laser induced damage threshold, ISO 11254*
157 nm, F_2 excimer	HR/th		1–4%	
193 nm, ArF excimer	AR/th	0.7–2.5%	0.2–0.5%	1–2 J/cm^2 (1on1, 20 ns)
	HR/th	0.4–2.0%	0.2–2.5%	2–4 J/cm^2 (1on1, 20 ns)
248 nm, KrF excimer	AR/th		<0.025%	10 J/cm^2 (1on1, 30 ns)
	HR/th	<500 ppm	<0.2%	>20 J/cm^2 (1on1, 30 ns)
	HR/IB		<0.1%	>3 J/cm^2 (1on1, 30 ns)
633 nm, HeNe laser	HR/th	<30 ppm	<30 ppm	–
	HR/IB	<5 ppm	<5 ppm	–
1.064 μm, Nd:YAG	AR/th	<20 ppm	<100 ppm	>60 J/cm^2 (12 ns, 0.25 mm)
	HR/th	<50 ppm	<100 ppm	>100 J/cm^2 (12 ns, 0.25 mm)
	HR/IB	<1 ppm	<1 ppm	>80 J/cm^2 (12 ns, 0.25 mm)
10.6 μm, CO_2 laser	AR/th	<0.16%	–	>20 J/cm^2 (100 ns, 1.4 mm) >2 kJ/cm^2 (1.2 ms, 250 μm) >3 kW/mm (cw, 100 μm)
	HR/th	<0.10%	–	>25 J/cm^2 (100 ns, 1.4 mm) >2 kJ/cm^2 (1.2 ms, 250 μm)

Types: HR: highly reflecting mirror; AR: antireflective coating; th: thermal evaporation; IB: ion beam sputtering.

for coatings with increasing complexity. In all fields of modern optics, optical thin films will represent an enabling technology which will play a key role for many future applications and products.

List of Units and Nomenclature

A_0	amplitude of partial wave of order zero
A_1	amplitude of first-order partial wave
A_k	amplitude of the partial wave of order k
d_i	thickness of layer number i
D_i	film thickness expressed in units of QWOT
E_0	amplitude of a plane wave
E_i, H_i	electric field strength and magnetic field strength at the rear front of layer number i
IAD	ion-assisted deposition
IBS	ion beam sputtering
k	wavenumber of a plane wave $k = 2\pi/\lambda$
M_i	matrix of layer number i
M_S	composite matrix of a layer stack of K single layers
n_0	index of refraction of the ambient medium
n_H	refractive index of a high-index material
n_i	refractive index of layer number i
n_L	refractive index of a low-index material
n_T	refractive index of the substrate
QWOT	quarter wave optical thickness: unit for the thickness of the layer
r_1, r_1'	reflection coefficients of the ambient–layer interface
r_2	reflection coefficient at the layer–substrate interface
r_S	reflection coefficient of a single layer
r_{SK}	reflection coefficient of a layer stack of K single layers
R_S	reflectance of a QWOT stack
t_1, t_1'	transmission coefficients of the ambient–layer interface
δ_i	phase shift of a plane wave in layer number i
λ	wavelength of a plane wave
λ_Z	design wavelength
θ_i	angle of incidence in layer number i
ω	frequency of a plane wave

See also

Optical Coatings: Laser Damage in Thin Film Coatings.

Further Reading

Guenther AH (ed.) (2002) *International Trends in Applied Optics*. Bellingham, Washington: SPIE Press.

ISO 11551 (1997) *Test Method for Absorptance of Optical Laser Components*. International Organization for Standardization.

ISO 13696 (2002) *Test Method for Radiation Scattered by Optical Components*. International Organization for Standardization.

ISO 9022 (1996, 1997) *Optics and Optical Instruments – Environmental Test Methods – Parts 1–21*. International Organization for Standardization.

ISO 9211 (1994–2001) *Optical Coatings, Parts 1–4*. International Organization for Standardization.

ISO/FDIS 15368 (2003) *Measurement of Reflectance of Plane Surfaces and Transmittance of Plane Parallel Elements*. International Organization for Standardization.

Jacobson MR (1989) *Deposition of Optical Coatings*. SPIE Milestone Series, Volume MS 6. Bellingham, Washington: SPIE Press.

Kaiser N and Pulker HK (2003) *Optical Interference Coatings*. Berlin: Springer.

Liddell HM (1980) *Computer-Aided Techniques for the Design of Multilayer Filters*. Bristol, UK: Adam Hilger.

Macleod HA (2001) *Thin Film Optical Filters*. Bristol, UK: Institute of Physics Publishing.

Rancourt JD (1996) *Optical Thin Films*. Bellingham, Washington: SPIE Press.

Thelen A (1988) *Design of Optical Interference Coatings*. New York: McGraw-Hill.

X-Ray Coatings

P Dhez, Université Paris Sud, Orsay, France

© 2005, Elsevier Ltd. All Rights Reserved.

In all spectral ranges, there exist two principal options for coating an optical element. This may be a simple layer, as for example a metallic layer is frequently used in the visible spectral region. The second option is to use a multilayer coating, making use of the optical properties of two or more materials to achieve enhanced reflectivity over a limited spectral range. In both cases, a knowledge of the refractive

index of the materials (real or complex) is required, and of course one must choose materials that can be coated in a practical deposition process. A frequent side-advantage of coating is to make it possible to choose the material of the substrate for its mechanical or cooling properties without regard to the optical properties of the coatings.

Originally (at the beginning of the twentieth century) the progress of X-ray optics was hampered by a poor understanding of the nature of these rays, and in particular by the observation that they are very weakly reflected and refracted. After the electromagnetic nature of these rays had been confirmed, the problem of producing polished surfaces that were sufficiently smooth on an atomic scale to reflect X-rays still remained. Optical coatings gave little hope of improvement due to the growth of roughness during the coating process. Thus early attempts to make X-ray mirrors produced poor results and little progress was made until the 1960s. However, in the past 30 years, much technical progress has been made, and now many optical principles used in the visible spectral region have been extended to very short wavelengths. Even refractive and diffractive focusing lenses (not treated here) have been implemented in the X-ray region.

In the first section we outline the history of the discovery of the properties of X-rays, which were achieved mainly in experiments using crystals. This allows us to understand how their electromagnetic nature was discovered, and how they interact with materials. We then describe the phenomenon of the 'total reflection' of X-rays, which occurs only at very high angles of incidence ('grazing incidence') and its main applications in X-ray optics. Finally, we outline the principles of mirrors based on interference, their applications in the X-ray region, their advantages and their limitations.

X-rays Compared to the Other Electromagnetic Spectral Regions

Infrared radiation was discovered by W. Herschel in 1800, and ultraviolet rays by J. Ritter in 1801. These were the only 'invisible rays' known before Maxwell's contribution to the understanding of the electromagnetic spectra through the development of his equations in 1869. The discovery of radio waves by H. Hertz in 1887 confirmed that our eyes are blind to most of the electromagnetic spectrum.

In 1895, W. Roentgen discovered some previously unknown 'rays' that had the surprising property of being able to penetrate low-density material opaque to visible light. This discovery was a side-effect of his studies of low-pressure electrical discharges, fashionable at the time. The nature of these unknown or 'X'-rays as Roentgen named them, remained unknown for 12 years or so. Were they electromagnetic waves or particles? Roentgen used lenses, prisms and mirrors to try to detect a reflective, refractive or focusing effect, with inconclusive results. Although the electromagnetic nature of the X-rays was favored, the magnitude of the frequency or the refractive index was unknown. We note that natural radioactivity and gamma rays were discovered a year after X-rays, and the same questions concerning their nature were raised. However, in neither case did the ignorance of the nature of the radiation impede their use in medical radiography or in cancer therapy.

In 1906, the experiments of C. Barkla provided the first evidence of the electromagnetic nature of X-rays. In scattering X-rays from blocks of a light element (in this case, carbon), he was able to show that the scattered radiation was polarized, and therefore had the nature of a transverse wave. This kind of experiment had been performed almost a century previously to demonstrate the transverse wave nature of visible light. In 1916, following a suggestion by M. von Laue that X-rays might be diffracted from a grating of period comparable to their wavelength, Friedrich and Knipping succeeded in diffracting X-rays from the atomic planes of a natural crystal. This experiment established definitively the electromagnetic nature of X-rays and also the scale of their wavelength.

In 1915 W. L. Bragg summarized the diffraction rules for crystals through his equation [1], now known as the simplified Bragg law:

$$2d_i \sin\theta_i = N\lambda_i \qquad [1]$$

This equation relates the diffracted X-rays wavelength λ_i to a set of parallel atomic planes separated by a distance d_i and to the incidence angle θ_i. N is an integer indicating the order of diffraction, as in the grating equation. In using this equation, care is needed because of two peculiarities compared to the usual mirror or grating equations in the visible range. In the Bragg equation, the angle of incidence θ is measured from a given atomic plane to the incident rays. However, inside a crystal several different atomic planes exists, not only the planes parallel to the surface as indicated in Figure 1. In addition, θ is not measured from the normal as is usual in optics; hence the angle θ is referred to as the glancing or grazing angle of incidence.

We note that the Bragg equation is not restricted to X-rays, but is a coherent scattering phenomenon which is common to all cases where waves, such as

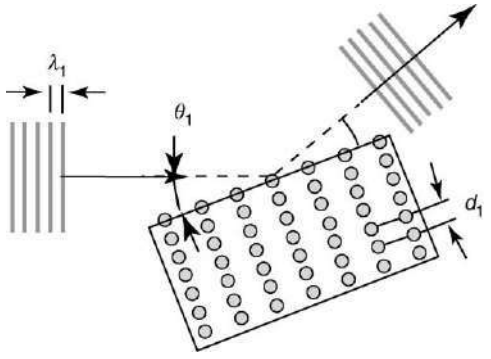

Figure 1 The geometry for the conditions leading to selective reflection on a periodic medium.

sound waves or ripples on water, are diffracted by regular arrays of scattering centers.

In order to visualize the phenomena, we refer to Figure 1 which is a schematic representation of a plane wave incident upon a two-dimensional array of scattering centers (electrons around nucleus, in the case of X-rays), shown here as dots. As the sine function has a maximum value of 1, the maximum wavelength λ that can be reflected from a spacing d is given by $\lambda = 2d_{\max}$. As the largest value of $2d$ that can be found in natural crystals is about 2 nm (in natural muscovite mica, for example), X-rays of 2 nm, with an equivalent energy of about 620 eV, are the 'softest' (least energetic) that can be reflected in this manner.

The model shown in Figure 1, similar to that proposed by M. von Laue, does not take into account the detailed interaction of the atoms with the X-rays, and therefore ignores effects such as absorption and refraction, and discrete atomic transitions. A.H. Compton attempted the first treatment of the interaction of matter with X-rays from an atomic viewpoint. In 1917, he developed a classical theory in which the electrons bound to the nucleus are treated as harmonic oscillators. In this way, he was able to derive an expression for the atomic scattering coefficients and to describe resonance effects at frequencies corresponding to specific atomic transitions. A few years later, he was able to establish a link between these 'atomic scattering factors' and the macroscopic concept of refractive index. This new formulation allowed the use of the Fresnel equations, widely used in other spectral regions, to calculate the intensity of X-rays reflected from smooth, polished surfaces. These surfaces may be amorphous as well as crystalline.

The refractive index can be written in several ways; since it is close to unity over the X-ray range, the most common form is complex:

$$n = 1 - \delta - \mathrm{i}\beta \qquad [2]$$

where δ is the refractive index decrement, and the imaginary part β the absorption index. For photon energies corresponding to X-rays, both have small negative values, typically 10^{-3}–10^{-6}, and are correcting factors to the vacuum index 1. As in other photon energy ranges, the absorption and the refraction coefficient variations are linked through the Kramers–Krönig relations. In the visible range the index values are usually larger than 1, being below 1 only around specific optical thresholds. In the UV region, the index values become smaller than unity at energies above about 30 eV, which corresponds to the energy range for which atoms can be ionized.

At an early stage, Roentgen demonstrated X-ray absorption by sheets of materials. The demonstration of refraction took much longer and was more difficult. The first direct X-ray refraction index measurement was achieved in 1924, using a prism and an X-ray tube. Figure 2 is the scheme of the experiment, which is easy to perform in the visible range to measure the refraction of a transparent material. The footprint of the direct beam on the photographic plate is made before inserting the prism. In the visible range for which $n > 1$, the beam is refracted at larger angle than the glancing incidence angle, and so below the direct beam direction as shown in Figure 2a. At the contrary, as indicated in Figure 2b, the X-ray beam is refracted toward the prism surface and such an experiment effectively demonstrated that the index was slightly inferior to unity, as suggested by A.H. Compton. We will return

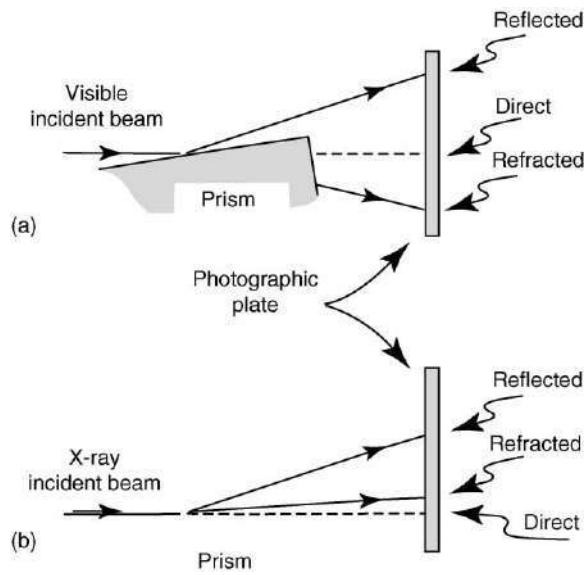

Figure 2 The geometry of the experiment used to evaluate the magnitude and the sign of the refractive index in the X-ray and the visible ranges.

to this peculiarity in the following section, when the total reflection phenomenon is described.

More details about the successive steps in our understanding and the pioneering works on X-rays can be found in standard textbooks.

Total Reflection X-Ray Mirrors

Using Fresnel's equations and tabulated values of the refractive index one can verify that, on any material, grazing incidence is necessary to reflect X-rays. Reflectivity calculations over the range from EUV to X-rays (30 eV to 10 keV), with approximations to allow calculations for rough surfaces, can be carried out using the web site of the Center for X-ray Optics of Lawrence Berkeley Laboratory http://www-cxro.lbl.gov, and curves similar to those presented below can be reproduced.

In the case of normal incidence, Fresnel's equation indicates that for any material the X-ray reflectivity $R^2 = |(n-1)/(n+1)|^2$ is very low, typically below 10^{-4}. This is a direct consequence of the X-ray index values, which are close to unity, as mentioned above.

In the visible range, normal incidence reflectivity from the surface of a transparent medium is low but can be increased just by going to glancing incidence. This is also the case in the X-ray region but in addition as the real part of the X-ray refractive index is lower than that of the vacuum, we have an 'external' total reflection phenomenon. From Figure 2 one can see how, for a sufficiently small incidence angle, the refracted wave becomes parallel to the interface. This is well known in the visible range (in which case it is the 'internal' reflection) and the phenomenon is used, for example, in fiber optics.

Figure 3 shows examples of calculated reflectivity variations versus grazing incidence angles, for perfectly smooth glass (SiO_2), nickel, and platinum surfaces. Curves are drawn for the 8 keV energy photons, corresponding to the Cu Kα line produced by conventional X-ray tubes. One observes that for this energy the reflectivity is low beyond a grazing angle around 1°, then rises quickly at low angles. This is typical of the total reflection phenomenon. The grazing angle corresponding to a 50% reflectivity is called the critical angle Θ_c, its value being roughly proportional to the refractive index decrement at the relevant photon energy. Notice that weaker absorption materials, lower Z, lead to more abrupt reflectivity variation but with smaller critical angles.

Such reflectivity variations at grazing incidence were systematically studied with X-ray tubes as early as the 1930s. The penetration depth of X-rays at grazing incidence was also studied. The evanescent wave inside the material, under total reflection conditions, is confined to a surface layer of a few tens of wavelengths. Hence, in the total reflection of X-rays, a coating having a thickness of a few tenths of a micron, on any substrate, is all that is required.

Figure 4 shows examples of the variation in nickel reflectivity calculated over the energy range 0.3–10 keV, for several grazing incidence angles. Nickel is one of the coatings commonly used for X-ray mirrors. The abrupt reflectivity variations and the dips correspond to the energies of the nickel atomic thresholds L (0.85 keV) and K (7.5 keV) for which the index tables predict a rapid change in absorption and refraction coefficients. Choosing an incidence angle provides a practical way of limiting the reflected spectral range. The energy cut-off of the

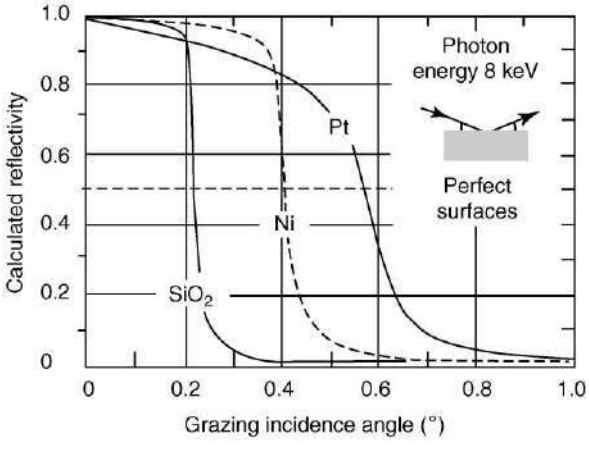

Figure 3 Examples of calculated reflectivity variations versus the incidence angle in the case of 8 keV X-rays on perfect platinum, nickel and glass surfaces.

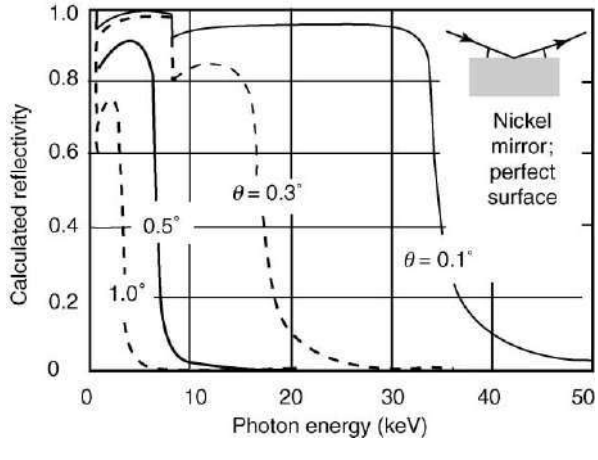

Figure 4 Calculated reflectivity variations versus photon energy, on a perfect nickel surface, for different incidence angles.

X-ray mirror acts as a filter for the higher-energy photons. That is effectively the method of suppressing the short wavelengths from an incident spectrum if they are not desired in an experiment.

When using highly divergent sources such as X-ray tubes, the restriction to operation at grazing incidence leads to a strong limitation of the collected solid angle, and thus of the beam size and angular aperture. These constraints are less stringent on highly collimated beams emitted by synchrotron sources. Thus, most synchrotron beam lines use total reflection optics in the form of curved mirrors, to refocus the beams and remove the undesired high X-ray energies. Recently, X-ray imaging optics, consisting of a series of parabolic and hyperbolic nested grazing incidence mirrors to increase the collected aperture, have been launched on satellites.

The reflectivity curves in Figures 3 and 4 are calculated for perfect surfaces. Real surfaces, even polished surfaces, have roughness created by defects over a large and continuous range of spatial frequencies, ranging from several centimeters down to the atomic scale. Coating a surface replicates the large period defects. Eventually new defects are added at near-atomic scale by nucleation of the coating material. Interferometric methods have been developed to study the range of defects from centimeter to micron spatial periods, whereas the atomic force microscope (AFM) is able to measure roughness from the micron down to the 0.1 nm scale. As waves are scattered most efficiently by asperities of a size comparable to their wavelength, scattering measurements are used in the X-ray region to probe defects of atomic scale. The prevalence of scattering explains why mirrors that reflect well in the visible range were disappointing when used in early X-ray experiments. The recent development of adaptive polishing methods has almost solved this problem.

The main effect of scattering by an optical system is not to decrease the number of photons coming out, but to modify their angular distribution. Scattering measurements are experimentally studied to test and evaluate the spatial defect distribution. In X-rays, one first characterizes the incident beam angular distribution in measuring its angular variation intensity after a scanning slit. Performing the same measurement after the optic, across the reflected beam in the case of mirrors, provides the modified angular distribution. From such a measurement, a few theoretical models adapted to the X-ray range can be used to deduce the mirror spatial defect distribution and so to check progress in polishing methods. The simplest model, coming from early crystal studies, is describing defects by a single parameter named roughness and denoted σ. Figure 5 shows an example of calculated reflectivity; on nickel surfaces versus photon energy and for increasing roughness; it shows how many photons have deviated from the symmetrical direction. Corresponding angular distributions can also be calculated.

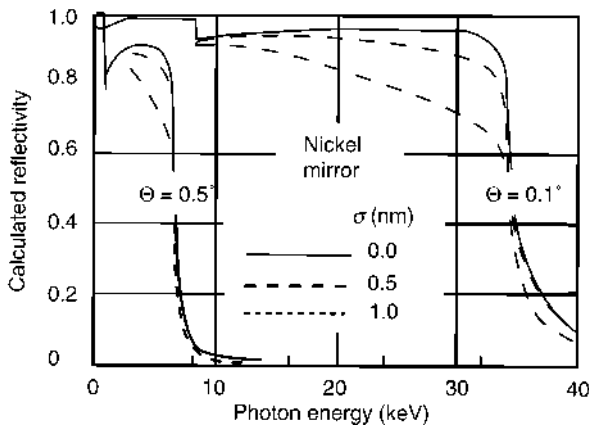

Figure 5 Calculated reflectivity variations versus photon energy, on nickel surfaces having different roughness, for different incidence angles.

The coating systems used for X-ray mirrors are material dependent. Deposition methods are similar to those used for visible optics coatings, but specific effort has been devoted to avoid nucleation growth to achieve uniform smooth layers. In general, one tries to give the evaporated particles, ions or clusters a kinetic energy lying within a specific range, which promotes mobility on the surface but is below the energy needed for implantation. In the case of electron beam deposition the energy distribution is thermal, but an auxiliary ionic bombardment can be used during deposition to provide the necessary energy to the previously deposited atomic layer (ion assisted deposition). For each material, the specific sputtering deposition and ionic bombardment conditions need to be carefully studied.

Until the 1970s, Bragg reflection by crystals and the total reflection phenomenon was the only way to reflect X-rays.

Multilayer X-Ray Mirrors

In the visible range, optics are frequently coated with stacks of dielectric multilayers obtained by periodic deposition of layers of two amorphous transparent materials. Such stacks are highly selective reflectors, or bandpass filters. They operate on the principle of

constructive interference between all the waves partially reflected by the successive interfaces. They can be regarded as analogous to two-dimensional Bragg reflectors, in which the interfaces play the same role as atomic planes in crystals. Hence the Bragg equation [1] can be applied, the relevant period d being the thickness of an individual bilayer, if the stack is truly periodic.

The concept of using artificially deposited stacks of thin films, of alternating high and low electron density, to reflect and diffract X-rays dates back at least to the 1920s. It was not until 1944, however, that DuMond and Youtz were able to demonstrate significant X-ray reflection from such an artificially layered structure. Their work used alternating layers of copper and gold, laid down by evaporation. Unfortunately these materials diffuse into one another, reducing the electron density contrast and hence the X-ray reflectivity. In 1967 Dinklage and Frerichs succeeded in making a structure of alternating layers of iron and magnesium, which do not interdiffuse quickly, and thus produced the first stable X-ray reflecting multilayer.

Since then, enormous progress has been made in the fabrication and application of such multilayer reflectors, principally because of the efforts of two pioneers. Troy Barbee, while at Stanford University, developed open loop sputtering methods for controlling the thickness of the many thin (~nm) layers required to make up such a stack, while Eberhard Spiller at IBM developed an evaporation method using *in situ* X-ray monitoring to achieve similar thickness control. Almost all of the many multilayer deposition systems in use around the world are based on one of these two methods.

As for the total reflection mirror, only recent progress in methods for deposition and controls has made this new kind of X-ray optics practical. Present applications range from the extreme ultra violet (30 eV) region to hard X-rays having energies of a few tens of keV.

Prediction of the reflectivity variations can be simulated by calculations on the same web site mentioned in the preceding section. Multilayer mirrors are based on interference, so the reflectivity coefficient and the width of the reflected bandpass are determined by the amount of absorption inside the stack and the partial reflectivity at each interface. The more efficient pairs of materials have the lowest X-ray absorption and the highest refractive index difference at the required energy. The selection of real materials is constrained by the coating conditions and the need to achieve smooth interfaces. In particular, a low interdiffusion coefficient between the successive material pairs is a requirement for temporal stability. Epitaxial deposition methods have been considered but no useful pair of materials has yet been found.

An example of a flourishing application is multilayer-coated normal incidence X-ray optics. Optical schemes used in the visible region for telescopes and microscopes have excellent optical performance and high numerical aperture but a very low reflectivity in the X-ray region if coated with a single layer. However, if coated with multilayers they can be used, providing that extremely low-roughness surfaces can be produced. The interface qualities of the deposited thin layers presently limit normal incidence optics to the soft X-ray energy range, typically below 300 eV so around the carbon edge. This corresponds to periods of about 2 nm, or a few atomic layers, and the resulting roughness seriously decreases the reflectivity.

Figure 6 is a reflectivity curve versus energy for a quasinormal incidence Mo/Si multilayer mirror optimized to get the first-order peak at 93 nm. On Mo/Si mirrors, precise control of the layer period and the quality of the interfaces allow the manufacture of mirrors with reflectivity only a few percent below the theoretical limit shown here. Considerable effort has been invested in such mirrors, with a view to applications to EUV projection lithography.

It is possible to build polarizer mirrors over the soft X-ray range. An incidence angle of 45° corresponds to the Brewster angle, because the X-ray index is close to unity. X-ray multilayer and semitransparent selective beam splitters now permit the development of Michelson interferometers for newly developed soft X-ray lasers.

Figure 7 is an example of a tungsten/carbon (W/C) reflectivity curve for a multilayer adapted to the 8 keV energy of Cu Kα, mentioned before. The curve

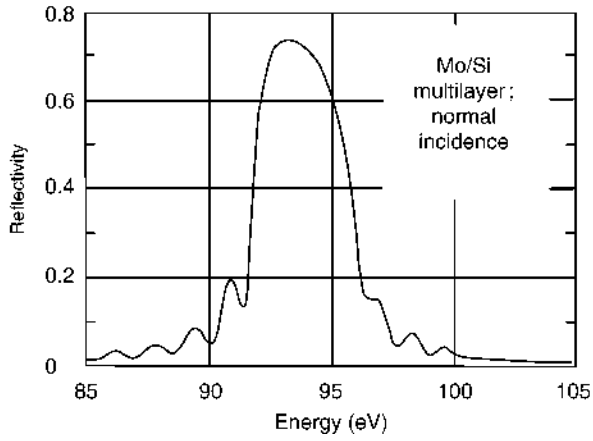

Figure 6 Calculated reflectivity versus incidence angle for a typical Mo/Si multilayer, optimized around 93 eV, under quasi-normal incidence.

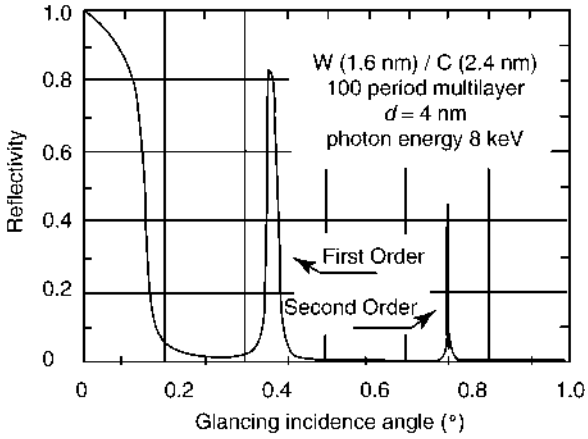

Figure 7 Calculated reflectivity versus incidence angle for a typical W/C multilayer adapted to reflect 8 keV photons.

of reflectivity versus glancing angle has been calculated from the total reflection range to the second-order peak. This kind of multilayer is being used in X-ray tubes to premonochromatize the beam and make it parallel. In such an application, the period of the coating is continuously graded laterally to satisfy locally the Bragg law at each point of the curved mirror.

As a final example of the new possibilities offered by X-ray multilayers, we mention the possibility of widening the reflected bandpass by varying continuously the period from the bottom to the top of the deposited stack. This is the equivalent of multilayer neutron 'supermirrors'. Such interference mirrors are usable at quite high X-ray energies, and give high reflectivity and a larger incidence angle than total reflection mirrors working over the same energy range.

The continuous fast development of coating technologies and X-ray multilayer applications can be followed through the proceedings of dedicated meetings at, for example, on a web site like http://PXRM.

Summary and Perspectives

After the discovery of X-rays in 1895, almost 25 years were necessary to understand their characteristics compared to the visible range. Only in 1923 did F. Holweck and R.A. Millikan demonstrate the continuity of electromagnetic radiation from X-rays to ultraviolet. The selective Bragg reflection by natural crystals is limited to energy commonly larger than about 2 keV photons. Below this energy the total reflection mirror was the only optical device to reflect and focus wavelengths shorter than the far UV until the development of adapted interferential mirrors.

During the 1960s, the EUV and X-UV ranges gained interest due to the spread of fusion research and in the astrophysics community studying the Sun and stars. For both communities, the continuum and the lines emitted at short wavelengths by ionized matter were a convenient tool to evaluate high plasma temperatures. At about the same time, the development of synchrotron radiation sources increased the need for X-ray optics. In addition, this last source offered the possibility to attempt new optical solutions with convenient intense and parallel beams from UV to X-rays. Finally, the concomitant progress in various microtechnologies afforded the necessary tools to produce new adapted optical devices for short wavelengths. At the present time, adapted optics are under development to take advantage of various applications from microscopy to interferometry over the very large electromagnetic energy X-ray range.

As a matter of fact, it is now possible to tailor a choice of mirror coatings adapted to each X-ray photon energy and application. Such coatings are specific because they must be optimized to take into account the inevitable absorption energy dependence.

Progress achieved in blank polishing and evaporation methods has led to efficient and high-quality X-ray mirrors even able to keep the spatial coherence of synchrotron sources. Such technological advances are also steps to master surfaces and interfaces at the atomic scale. They are now becoming applicable to the other spectral ranges and to various micro- and nanotechnologies.

List of Units and Nomenclature

Crystal period	d	[nm]	$1\ \text{nm} = 10^{-9}\ \text{m}$
Incidence Angle	Θ	[degree]	
Photon Energy	E	[eV]	$(h/2\pi)\omega\lambda = hc =$ 1239.842 eV.nm
Wavelength	λ	[nm]	$1\ \text{nm} = 10^{-9}\ \text{m}$

See also

Diffraction: Fresnel Diffraction. **Geometrical Optics:** Lenses and Mirrors. **Incoherent Sources:** Synchrotrons. **Optical Coatings:** Thin-film Optical Coatings.

Further Reading

Attwood D (1999) *Soft X-Rays and Extreme Ultraviolet Radiation*. Cambridge: Cambridge University Press.

Ditchburn RW (1963) *Light*. Appendix XIX, 2nd edn. London: Blackie.
Hecht E (1998) *Optics*, 3rd edn. Reading, MA: Addison-Wesley.
Henke BL, Gullikson EM and Davies JC (1993) *Atomic Data and Nuclear Data Tables* 54: 181.
James RW (1982) *The Optical Principles of the Diffraction of X-Rays*. Woodbridge, CT: Ox Bow Press.
Michette A and Pfauntsch S (1997) *X-Rays: The First Hundred Years*. New York: John Wiley.
Mould RF (1993) *A Century of X-Rays and Radioactivity in Medicine*. Bristol, UK: Institute of Physics Publishing.
Paratt LG (1954) Surfaces studies of solids by total reflection of x-rays. *Physical Review* 95: 359.
Spiller E (1972) Low-loss reflection coatings using absorbing materials. *Applied Physics Letters* 20(9): 365–367.
Underwood JH and Attwood DT (1984) The renaissance of x-rays optics. *Physics Today, April*: 44–51.

OPTICAL COMMUNICATION SYSTEMS

Contents

Basic Concepts
Historical Development
Architectures of Optical Fiber Communication Systems
Free Space Optical Communications
Lightwave Transmitters
Local Area Networks
Optical Time Division Multiplexing
Wavelength Division Multiplexing

Basic Concepts

S Lee and A E Willner, University of Southern California, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Since modern lightwave communication started in the 1970s, extensive efforts have been made to increase data rate and transmission distance. Recently, the bit rates for single-channel and multichannel systems have exceeded 40 and 2 Tb/s on single fibers respectively, for transoceanic distances. One revolutionary development in lightwave systems has been the deployment of erbium doped fiber amplifiers (EDFAs). These amplifiers facilitate transmission of an optical signal over long distances, by providing periodic analog-like amplification rather than digital-like regeneration. The wide bandwidth provided by EDFAs has made possible the increased use of wavelength division multiplexing (WDM) in which multiple lightwave signals, each having a different wavelength, are co-propagated on a single fiber. EDFAs and WDM techniques can enhance the lightwave system capacity, both in terms of obtainable transmission distance, and total number of data rates (see Figure 1).

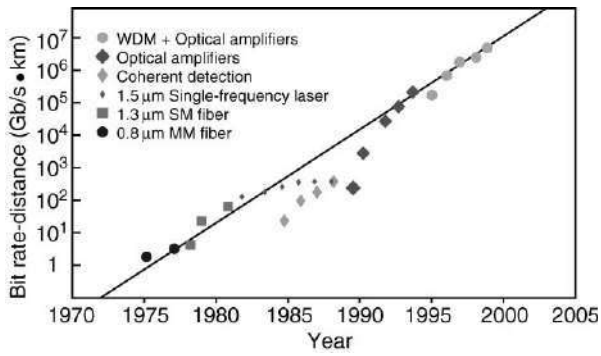

Figure 1 Bit rate product. Data from T. Li (private communication).

Optical Fiber

Most telecommunication fibers are made of silica, therefore following the attenuation of silica material. Figure 2 shows the silica fiber attenuation as a function of wavelength. The downward slope towards 1.6 μm and upward slope away from 1.6 μm, are the theoretical limits due to Rayleigh scattering and absorption of silica material, respectively. The absorption peak near 1.4 μm is due to water molecule absorption. There are two low attenuation regions: ~1.3 μm and 1.55 μm, both of which are used for communications in general, and some other wavelengths are also used for shorter distance communications. Figure 2 also shows the comparison between gain bandwidth of EDFA (~3 THz) and the low attenuation window around 1.55 μm (~25 THz). The typical values of attenuation at 1.3 μm and 1.55 μm for single-mode fiber (SMF) are ~0.35 dB/km, and 0.2 dB/km, respectively.

When an electromagnetic wave interacts with the bound electrons of a silica fiber, the medium response depends upon optical frequency (ω). This property, referred to as material dispersion, manifests itself through the frequency dependence of the refractive index $n(\omega)$. Fiber material dispersion plays a critical role in propagation of short optical pulses since different spectral components associated with the pulse travel at different speeds, given by $c/n(\omega)$. Consequently, the optical pulse at the output of the fiber is broadened in time. The commonly used system parameter is called the dispersion parameter D (the negative D value is called normal dispersion, and the positive D value is called anomalous dispersion), and is measured in ps/(nm·km); D for SMF is ~+17 ps/(nm · km) at 1.5 μm. However, because of dielectric waveguiding, the effective mode index is slightly lower than the material index $n(\omega)$, with the reduction itself being ω-dependent (waveguide dispersion). This results in a waveguide contribution that must be added to the material contribution to obtain the total dispersion (see Figure 3).

Generally, the waveguide contribution to dispersion D is to shift zero dispersion wavelength (λ_0) to longer wavelengths. Furthermore, the waveguiding can be tailored to shift the λ_0 from 1.3 μm to 1.5 μm, the wavelength at which the fiber has the minimum attenuation loss. The fiber having λ_0 in the neighborhood of 1.5 μm, is called dispersion-shifted fiber (DSF); D for DSF is usually between −2.5 and +2.5 ps/nm·km, at a wavelength of 1.5 μm. The typical dispersion parameter, D, as a function of a wavelength for both SMF and DSF, is shown in Figure 4.

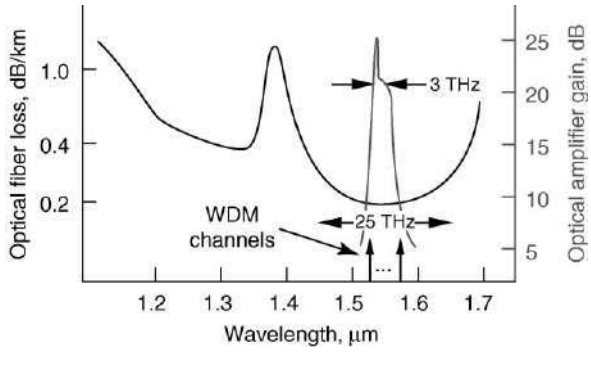

Figure 2 Silica fiber attenuation versus wavelength.

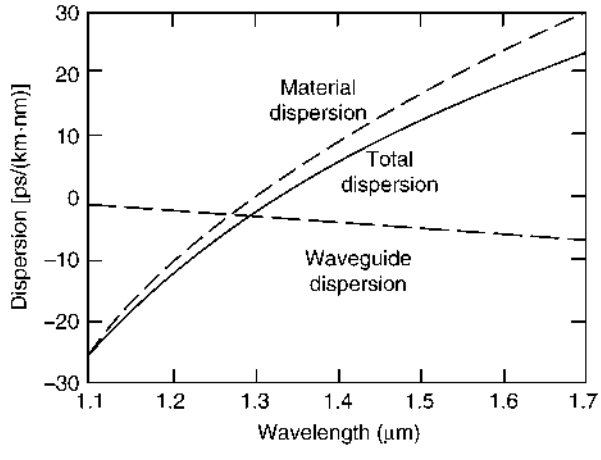

Figure 3 Total dispersion (D) and contribution of material dispersion and waveguide dispersion in conventional SMF.

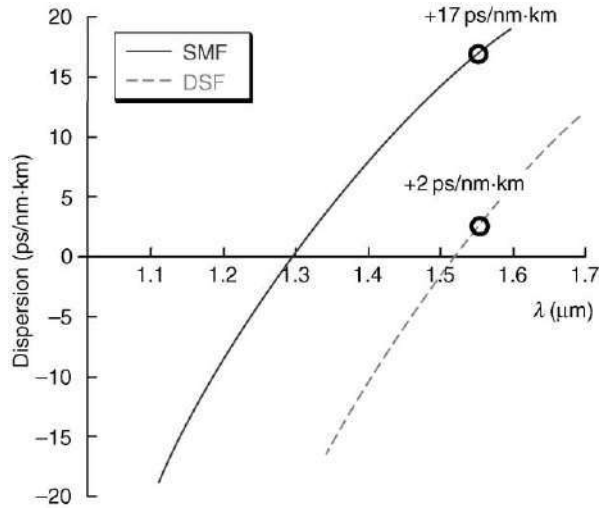

Figure 4 Dispersion versus wavelength for conventional SMF and DSF.

In addition to chromatic dispersion, optical fiber has another dispersion properties related to polarization. An optical wave of arbitrary polarization can be represented as the superposition of two orthogonally polarized modes. In an ideal fiber, these two modes are indistinguishable, and have the same propagation constants owing to the cylindrical symmetry of the waveguide. However, in real fibers there is some residual anisotropy due to unintentional circular asymmetry, usually caused by noncircular waveguide geometry or asymmetrical stress around the core, as shown in Figure 5.

In either case, the loss of circular symmetry gives rise to two distinct orthogonally-polarized modes with different propagation constants (differential phase velocity) responsible for polarization mode dispersion (PMD) in the fiber, and can be related to the difference in refractive indices (birefringence) between the two orthogonal polarization axes. The differential phase velocity indicated is accompanied by a difference in the group velocities for the two polarization modes, therefore the pulse at the end of the transmission fiber is broadened by this differential group delay (DGD). DGD is usually expressed in units of ps/km for a short length (1 m to 1 km) of birefringent fiber. However, DGD does not accumulate along a long fiber link in a linear fashion. Instead, because of random variations in the perturbations along a fiber span, the DGD in one section may either add to or subtract from another section of the fiber. As a result, average DGD in long fiber spans accumulates in a random-walk-like process that leads to a square root of transmission-length dependence. Therefore, average DGD is expressed in $ps/km^{1/2}$ in long fiber spans, referred to as the PMD of the fiber, and the typical PMD parameter has a value of 0.01 to 10 $ps/km^{1/2}$. Because of the many perturbations that act on a real-world fiber (e.g., temperature, vibration, etc.), transmission properties typically vary with time. Therefore, the PMD of the fiber link fluctuates randomly, thus causing random fluctuations in system performance.

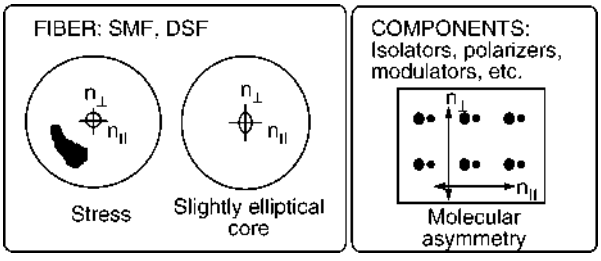

Figure 5 Origin of PMD.

It is well known that optical fibers show nonlinear behavior under conditions of high power and long interaction length. The power-times-distance products for amplified transmission systems can be large enough to make fiber nonlinear effects the dominant factor in the design of long-distance systems.

There are two categories of fundamental optical nonlinear effects: stimulated scattering effects and refractive index effects. Stimulated scattering effects arise from parametric interactions between light and acoustic or optical phonons in the fiber. Two nonlinear effects fall into this category: stimulated Brillouin scattering (SBS) and stimulated Raman scattering (SRS). The main difference between the two is that optical phonons participate in SRS, while acoustic phonons participate in SBS. In a simple quantum-mechanical picture, a photon of the incident field is annihilated to create a photon at a longer wavelength. The new photon is co-propagated and counter-propagated along the original signal in SRS, while counter-propagated in SBS. Refractive index effects are caused by modulation due to changes in light intensity. There are three types of refractive index effects: self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM). SPM introduces the change of optical phase by its own intensity, and, this leads to frequency chirp of the pulse, depending on the relative position from the peak. When this nonlinear frequency chirp interacts with the fiber dispersion, the pulse either broadens or compresses, depending on the sign and the amount of dispersion along the fiber. In XPM, the phase of the signal in one wavelength channel is modulated by the intensity fluctuation of the other wavelength channels. In a WDM system, XPM imposes far more damaging effects than SPM because XPM is stronger by a factor of two than SPM when the channel power is the same, and large number of WDM channels can contribute to XPM. FWM is the generation of modulation sideband at new frequencies, due to the phase modulation of channels between lights at different frequencies in multi-channel system. FWM causes penalties in a WDM system if the newly generated frequency is either equal to or close to the frequency of existing WDM channels.

Fiber nonlinearities impose significant degradation in optical transmission system, and limit the system on allowable number of channels, channel power, and channel spacing (see Figure 6). But, the effects can be well suppressed by carefully designing the fiber dispersion profile (i.e., dispersion map) in the transmission system, and controlling the optical launch power for each fiber span.

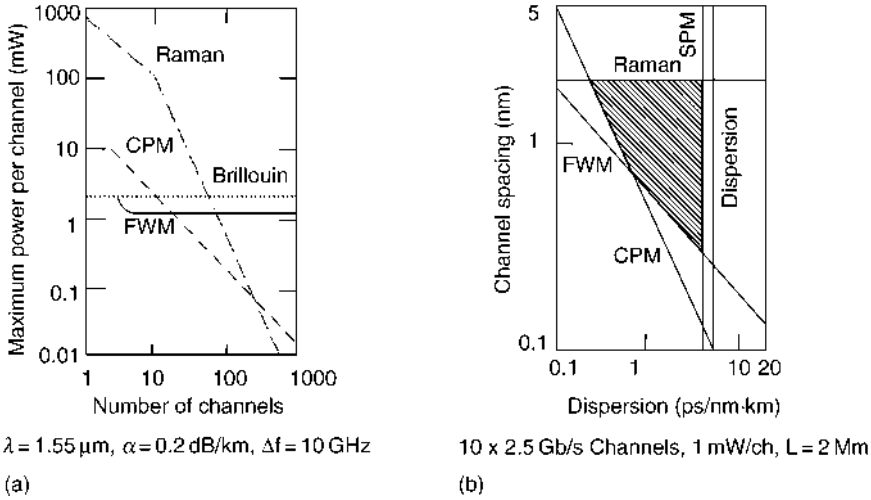

Figure 6 Limitation due to fiber nonlinearities in (a) maximum channel power and number of channels, and (b) channel spacing and dispersion. Reproduced with permission from Chraplyvy AR (1990) Limitation on lightwave communications imposed by optical-fiber nonlinearities. *IEEE Journal of Lightwave Technology* 8:1548–1557. Copyright © 1990 IEEE.

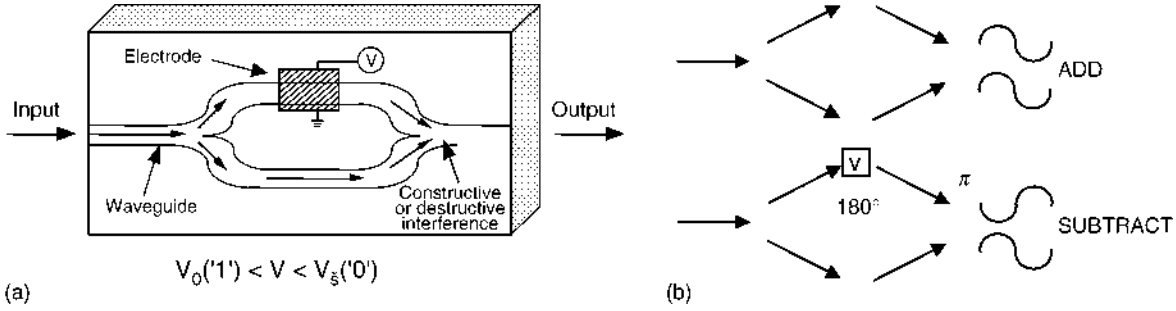

Figure 7 (a) EOM: MZ interferometer on LiNbO2, (b) operation of EOM; constructive and destructive interference resulting in intensity modulation.

Communication Components

Distributed feedback laser (DFB) is widely used as a light source for metro, long-haul, and undersea applications, due to its narrow spectral width, and wavelength stability. Fabry–Perot (FP) lasers, and vertical cavity surface emitting lasers (VCSEL) are used for local area networks (LAN), and for access applications, because these are cheaper than DFB lasers. For an application, where the cost is the most important parameter with low data modulation speed, light-emitting diode (LED) is used as a light source.

In optical data modulation, the simplest and easiest technique is turning ON or OFF the laser, depending on the binary logic '1' or '0' (direct modulation). This type of modulation is simple and cheap compared to external modulation technique. But, its application is limited by the bandwidth of modulation, and the induced frequency chirp (i.e., difference in optical frequency at the turn ON state and just before the turn OFF state of laser) that limits the transmission distance. Two different types of external modulators are widely used for digital optical data transmission. Electro absorption modulator (EAM) uses the Franz–Keldysh effect, where it is observed that the wavelength of the optical absorption edge in a semiconductor is lengthened by applying an electric field. Electro-optic modulator (EOM) utilizes the linear electro-optic effect, called the Pokels effect, which changes the refractive index of the material caused by and proportional to an applied electric field, so therefore changes the phase of the optical signal. This phase change is used to modulate the intensity of a lightwave through a Mach–Zehnder (MZ) interferometer (see Figure 7). Because the Pockels effect exists only in crystal, lithium niobate crystal ($LiNbO_2$) or electro-optic polymers are used for EOM.

EDFA is a wideband optical amplifier that has merits in that: (i) erbium ions (Er^{3+}) emit light in the 1.55 μm loss-minimum band of optical fiber, (ii) a circular fiber-based amplifier is inherently compatible with a fiber optics system; (iii) it provides

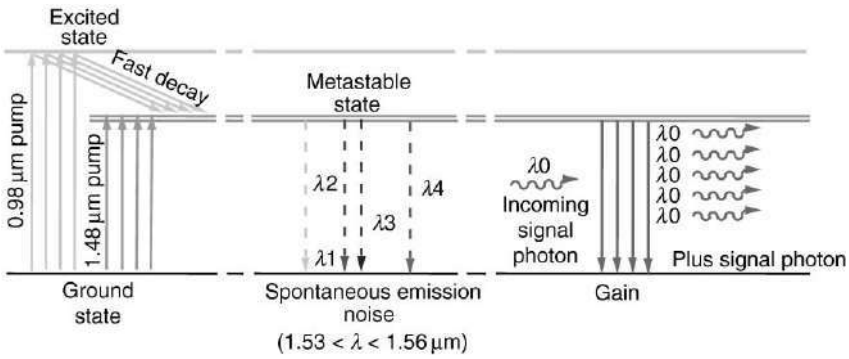

Figure 8 Energy level and transition diagram of EDFA. Reproduced with permission from Willner AE (1997) Mining the optical bandwidth for a terabit per second. *IEEE Spectrum* Apr.: 32–41. Copyright © 1997 IEEE.

amplification independent of bit-rate, modulation format, power, and wavelength; and (iv) it has low distortion and low noise during amplification. EDFA contains a gain medium (i.e., erbium-doped fiber) that must be inverted by a pump source. A signal initiates stimulated emission, resulting in gain, and spontaneous emission occurs naturally, which results in noise.

Figure 8 shows the energy level and transition diagram of EDFA. A 0.98 μm pump photon is absorbed and excites a carrier (Er^{3+} ions) into higher excited states, and the excited carrier decays rapidly to the first excited state that has a very long lifetime of ~10 ms (metastable state). In contrast, a 1.48 μm pump photon is absorbed and excites a carrier into a metastable state. This long metastable state is one of the key advantages of EDFA, and intersymbol distortion and interchannel crosstalk are negligible as a result of these slow dynamics. Depending on the external optical excitation signal, this carrier will decay in a stimulated or spontaneous fashion to the ground state and emit a photon. Both absorption and emission spectra have an associated bandwidth depending on the spread in wavelengths that can be absorbed or emitted from a given energy level. This is highly desirable because (i) a pump laser does not need to be an exact wavelength; and (ii) the signal may be at one of several wavelengths, especially in a WDM system. Figure 9 shows the bandwidth in the 1.48 μm absorption and 1.55 μm fluorescence spectrum of a typical erbium-doped amplifier.

When the noise characteristics of EDFA are considered, noise figure (NF) is an important parameter, which is defined by $NF \equiv SNR_{in}/SNR_{out}$, where SNR_{in} (SNR_{out}) is the electrical equivalent signal-to-noise ratio (SNR) of the optical wave going into (coming out) of the amplifier, if it were be detected. The typical value of NF in commercially-available EDFA is 4.5–6 dB. The importance of NF in

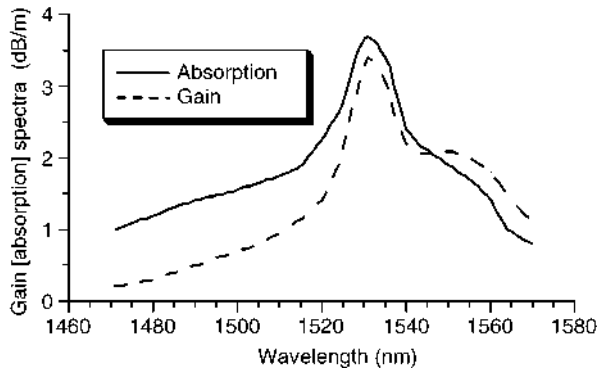

Figure 9 The absorption and fluorescence spectra for erbium near 1.5 μm. Reproduced with permission from Miniscalco WJ (1991) Erbium-doped glasses for fiber amplifiers at 1550 nm. *IEEE Journal of Lightwave Technology* 9: 234–250. Copyright © 1991 IEEE.

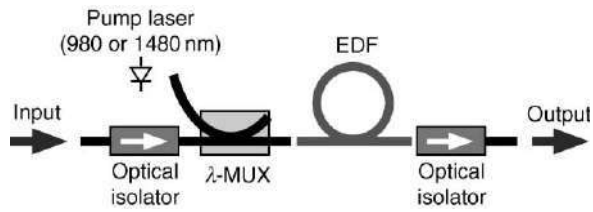

Figure 10 Typical EDFA diagram.

optical communication systems is presented later in this chapter, when the required optical SNR (OSNR) for error-free optical data transmission is considered. Figure 10 shows an example of EDFA diagram.

Recent development of EDFA in longer wavelength regimes or L-band (1570–1620 nm) has doubled the useful transmission bandwidth over the conventional band or C-band (1525–1265 nm). L-band gain of EDFA is generally achieved by proper arrangement of pump lasers through longer lengths of erbium-doped fiber than conventional C-band EDFA. Figure 11a shows the concept of wideband EDFA. The incoming

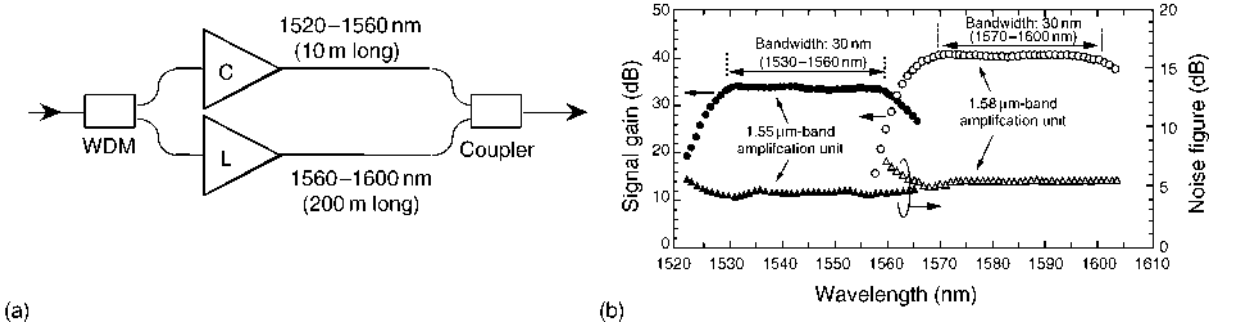

Figure 11 Wide band EDFA. Reproduced with permission from Yanada M (1997) *Electronics Letters*. Copyright © 1997 IEEE.

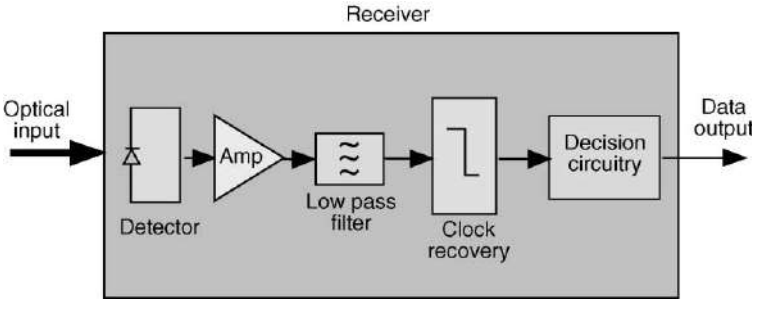

Figure 12 Block diagram of functions performed in advanced receiver packages.

WDM signal is split into two bands (C- and L-bands), amplified separately, and combined at the output. Gain and NF characteristics of this wideband EDFA is shown at Figure 11b.

Because SRS transfers the energy of a lower wavelength channel into the higher wavelength channel in the optical fiber, the transmission fiber itself can be used as a gain medium on specific channel (or channels) when the proper pump signal (or signals) are provided in the fiber. Even though Raman gain is much lower than the gain of EDFA (i.e., <0.1 dB gain/mW pump in Raman amplifier (RA) compared to a few dB gain/mW pump in EDFA), RA has several advantages over EDFA (i) RA has the capability to provide gain at any signal wavelengths, (ii) the amplification window is expandable by combining multipump wavelengths; and (iii) RA offers improved noise performance because the signal is amplified over transmission fiber (i.e., distributed amplifier). Especially improved noise performance increases system OSNR, which can be used to extend system reach or create more wavelength channels on the system. The Raman pumps are typically located ~100 nm shorter than the wavelength of the signal to be amplified, and a few 100s mW pump lasers are used to get a 10–15 dB gain.

In general, telecom receiver (RCVR) consists of photo diode (PD) with transimpedence amplifier (TIA), followed by limiting amplifier (LA) or automatic gain control amplifier (AGC), low pass filter, and clock data recovery circuit (CDR) (see Figure 12).

Vertically illuminated positive-intrinsic-negative (PIN) photodiode and avalanche photo diode (APD) are the most common PD that is used in RCVR. In PIN, incoming light is absorbed in the n^- (low doped or intrinsic) region and generates an electron-hole pair, which drifts to a depletion region and collected n^+ and p^+ regions, therefore generating electric current. APD is used for high-sensitivity detection up to 10 Gb/s. It is structured as absorption, grading, and multiplication layers. The multiplication layer of APD provides additional transit time delay compared to PIN, therefore the bandwidth of APD is smaller than PIN in general. TIA is a broadband low noise amplifier that is used to convert and amplify the weak current from PD to the desired output voltage. A low pass filter is used to suppress the noise and data distortion from high frequency components with a typical cut-off frequency ~0.7 bitrate. LA and AGC are amplifiers that limit the output voltage as a constant. LA is a nonlinear amplifier that makes a decision either to make logic '1' or '0' based on input voltage level, and limits output voltage as fixed values. AGC is a linear amplifier that limits the output voltage by adjusting the gain of the amplifier based on input voltage level. CDR consists of a phase lock loop (PLL) with a voltage controlled oscillator (VCO)

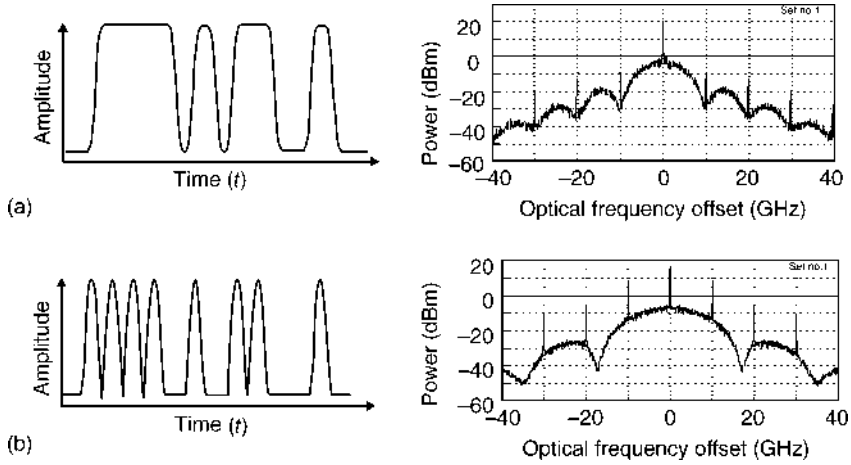

Figure 13 Amplitude modulation (a) nonreturn-to-zero (NRZ), and (b) return-to-zero (RZ).

and data flip-flop circuit. The PLL is used to synchronize the VCO to the incoming high-speed bit stream, therefore recovering the system clock. The recovered clock is used in D-FF to retime and regenerate the data.

Data Modulation Formats

In digital optical communications, numerous modulation formats have been developed and utilized, in order to achieve higher spectral efficiency (i.e., ratio of individual channel bit rate to dense WDM (DWDM) separation), and to improve system performance over fiber impairments. A simple amplitude modulation of lightwave is still the most widely used method for optical communication. The easiest amplitude modulation technique is to using a nonreturn-to-zero (NRZ) format. Figure 13a shows an example of NRZ binary sequence in time domain, and the corresponding frequency domain spectrum. In NRZ, data '1' and '0' correspond to the ON and OFF state of a lightwave transmitter, respectively, occupying entire bit time (i.e., binary symbol time). When the ON or OFF state of a lightwave transmitter does not occupy the entire bit time (T_B) (i.e., when the data pulse duration (T_d) is less then (T_B), it is called the return-to-zero (RZ). Figure 13b shows a binary sequence of RZ, and its frequency spectrum when $T_d \approx 0.5\ T_B$. The Fourier transform of the square wave is the Sinc function. Therefore, the frequency spectrum of an ideal NRZ and RZ follows the pattern of a Sinc square function. Note that the first null of frequency spectrum of RZ is about twice as high as the null of NRZ.

Even though NRZ is better in spectral efficiency than RZ, RZ has an advantage over NRZ in

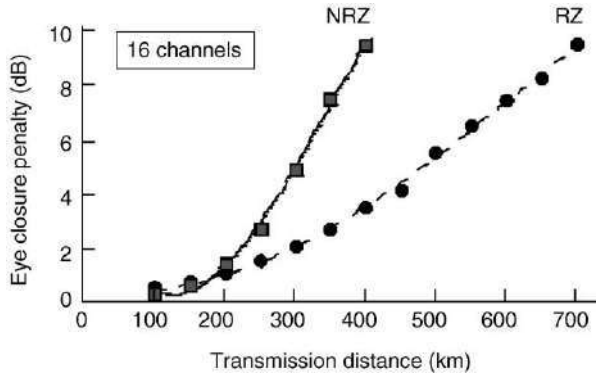

Figure 14 System performance of different modulation format (NRZ vs. RZ).

tolerance of fiber impairments. Figure 14 shows an example on eye closure penalty comparison between NRZ and RZ data formats at a 16-channel WDM system at 40 Gb/s data rate. It is clearly seen that the penalty of RZ is much less severe than the penalty of NRZ, as the transmission distance increases. (Note, nonlinearities and dispersion increases as the transmission distance increases.)

There has been much effort to increase the spectral efficiency, to improve the system performance over fiber impairments, or to achieve both simultaneously. Figure 15 shows some of examples of these. Chirped RZ (CRZ) is the modulation format inducing frequency chirp on conventional RZ. CRZ is robust on fiber impairment at the cost of losing spectral efficiency. In carrier-suppressed RZ (CS-RZ), a strong un-modulated optical carrier is removed from conventional RZ, therefore suppressing the nonlinear effects. Duo-binary is a three-level signal modulation format generated by a series of delay and added

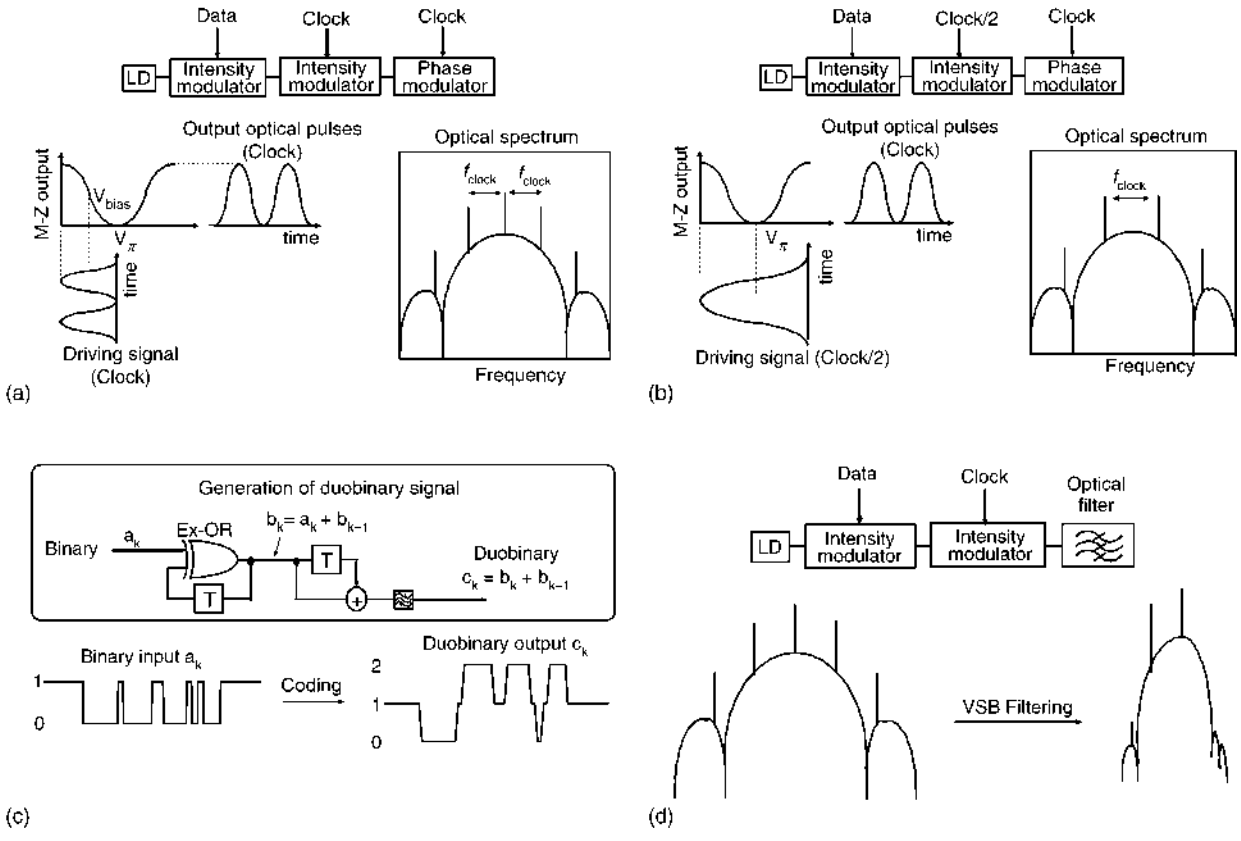

Figure 15 Examples of bandwidth efficient modulation formats. (a) Chirped RZ (CRZ) generation, (b) carrier suppressed-RZ (CS-RZ) generation, (c) duo-binary generation, and (d) vestigial single side band (VSB) RZ generation.

circuits in the transmitter. The advantage of duo-binary is that its bandwidth is reduced to one half of conventional NRZ, therefore having enhanced spectral efficiency, and has a high tolerance in chromatic dispersion, and suppression of fiber nonlinearites.

Single-sideband (SSB) or vestigial singleside band (VSB) can be generated into either NRZ or RZ, by either optically utilizing optical filtering with a sharp cut-off filter, or electrically applying tapped-delay-line filter approximation on both amplitude and phase modulators. SSB (or VSB) has about a two times higher spectral efficiency and higher tolerance on chromatic dispersion than conventional modulation formats.

Data Multiplexing

At their most basic, optical networks can imitate electrical networks in which time division multiplexing (TDM) is overwhelmingly used for digital data transmission. A fiber can carry many time-multiplexed channels, in which each channel can transmit its data in an assigned time slot. A typical TDM link is shown in Figure 16, in which N

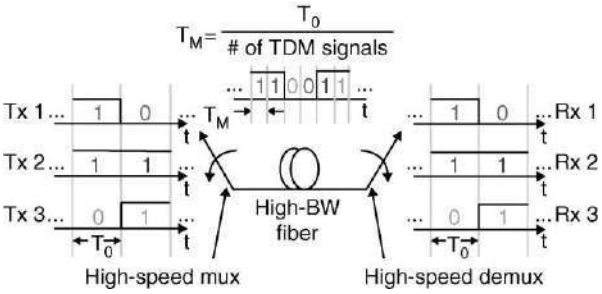

Figure 16 Concept of bit-interleaving TDM.

transmitters are sequentially polled by a fast multiplexer to transmit their data. The time-multiplexed data are sequentially and rapidly demultiplexed at the receiving node.

Time multiplexing and demultiplexing functions can be performed either electrically with an electrical time (DE)MUX switch and an optical transmitter/receiver, or optically with multiple optical transmitters/receivers and an optical time (DE)MUX switch. The major advantage of TDM is that there is no output–port contention problem (each data bit occupies its own time slot and there is only a single high-speed signal present at any given

instant). The disadvantage of TDM is that the scheme requires a ultra-high-speed switching component if the individual signals are themselves high-speed and if there are many users.

In WDM, multiple wavelength channels are transmitted through a single fiber, therefore enabling the fiber to carry more throughput. By using wavelength-selective devices for the ON and OFF ramps, independent signal routing also can be accomplished. Figure 17a shows a simple point-to-point WDM system in which several channels are multiplexed at one node, the combined signals transmitted across a distance of fiber, and the channels demultiplexed at a destination node. As shown in Figure 17b, the wavelength becomes the signature address for either the transmitter or the receivers, and the wavelength will determine the routing path through an optical network.

Many interesting challenges face the eventual implementation of WDM systems and networks. Several of these challenges involve the control and management of the data through this novel high-speed network. Figure 18 shows a small subset of critical component technologies typically required for a WDM system, including multiple-wavelength transmitters, multiport star couplers, passive and active wavelength routers, EDFAs, and tunable optical filters.

Some generic goals that a WDM-device technologist aims to achieve include; large wavelength tuning range; multi-user capability; wavelength selectivity and repeatability; low cross-talk; high extinction ratio; minimum excess losses; fast wavelength tunability; high-speed modulation bandwidth; low residual chirp; high finesse; low noise; robustness; high yield; potential low cost. Depending on system requirements, device availability and cost, WDM technologies divide into two; course-WDM (CWDM) and dense-WDM DWDM. Figure 19 shows the wavelength allocation for CWDM and

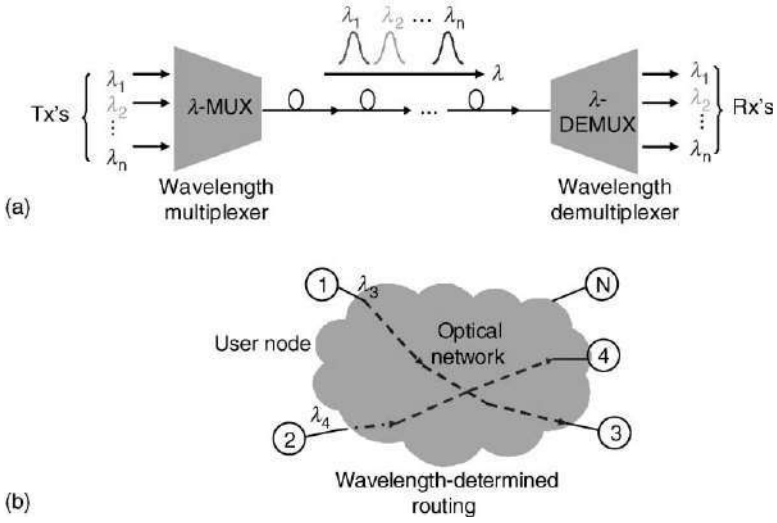

Figure 17 (a) A simple point-to-point WDM transmission system; (b) a generic multiuser network in which the communication links and routing paths are determined by the wavelengths used within the optical switching fabric.

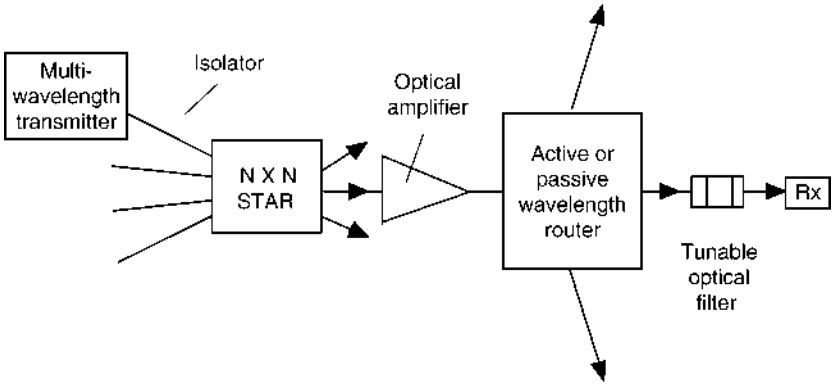

Figure 18 Schematic of a small subset of enabling device technologies for a WDM system.

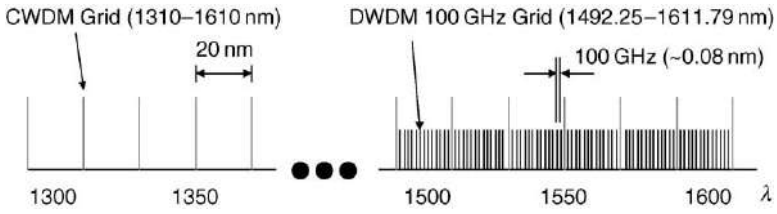

Figure 19 WDM wavelength allocation.

DWDM. CWDM technology uses an International Telecommunication Union (ITU) standard 20 nm spacing between the wavelengths from 1310 nm to 1610 nm. Also DWDM technology uses an ITU standard 100 GHz or 200 GHz between wavelengths, arranged in several bands at around 1500–1600 nm.

Dispersion Management and Compensation

As mentioned previously, chromatic dispersion is one of the most basic characteristics of fiber, although it is possible to manufacture fiber that induces zero chromatic dispersion. But such fiber is incompatible with the deployment of a WDM system since harmful nonlinear effects are generated, therefore chromatic dispersion must exist, and must be compensated for. The effect of chromatic dispersion is cumulative and increases quadratically with the data rate (see Figure 20).

The quadratic dependence of dispersion with the data rate is a result of two effects. First, a doubling of the data rate will double the Fourier-transformed frequency spectrum of the signal, thereby doubling the effect of dispersion. Second, the same doubling of the data rate makes the data pulse only half as long in time and therefore twice as sensitive to temporal spreading due to dispersion. The conventional wisdom for the maximum distance over which data can be transmitted is to consider a broadening of the pulse equal to the bit time period. For a bit period B, a dispersion value D and a spectral width $\Delta\lambda$, the dispersion-limited distance is given by $L_D = 1/(D{\cdot}B{\cdot}\Delta\lambda)$ (see Figure 21).

In theory, compensation of chromatic dispersion for high-speed or long-distance systems can be fixed in value if each link's dispersion value is known. A simple yet elegant solution is to create a dispersion 'map', in which the designer of a transmission link alternates elements that produce positive and then negative dispersion (see Figure 22). This is a very powerful concept: at each point along the fiber the dispersion has some nonzero value, effectively

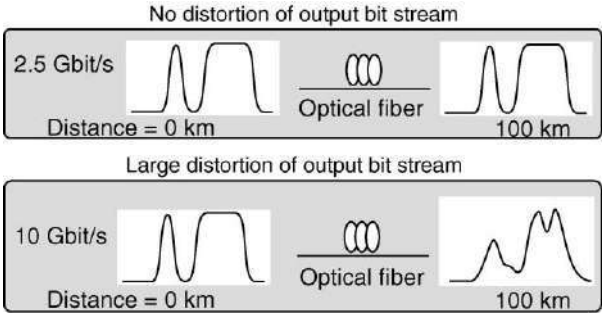

Figure 20 Pulse broadening at two different data rates (2.5 Gbit/s, and 10 Gbit/s) as a result of quadratic nature of chromatic dispersion.

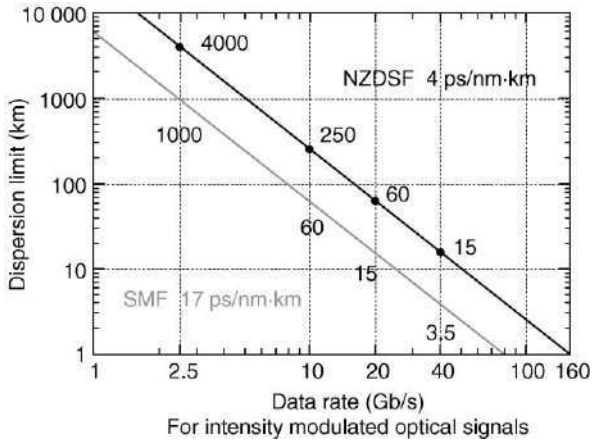

Figure 21 Dispersion-limited (uncompensated) transmission distance in single-mode fiber (SMF) as a function of data rate for intensity-modulated optical signals. Courtesy of L.D. Garrett.

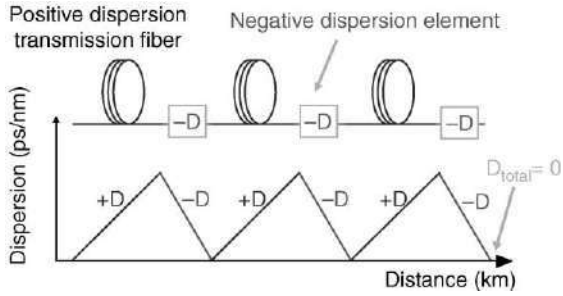

Figure 22 Typical static management of chromatic dispersion.

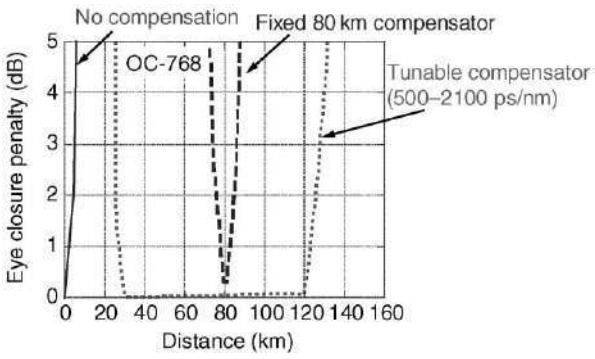

Figure 23 Effect of tunable compensation at 40 Gb/s (OC-768); it allows a wide range of transmission distance. On the contrary, the 80 km fixed compensator allows only small range of transmission around 80 km.

eliminating FWM and XPM, but the total accumulated dispersion at the end of the fiber link is zero, so that minimal pulse broadening is induced. The specific system design, as to the periodicity of management, depends on several variables, but a typical number for SMF as the embedded base is a compensation at every 80 km in a 10 Gb/s system. Dispersion-compensating fiber (DCF) has been generally used as a dispersion compensating element.

However, there are several important aspects of optical systems and networks that make tunable dispersion compensation solutions attractive, including: (i) it significantly reduces the inventory of different required types of compensation modules; (ii) it tunes to adapt to routing path changes in a reconfigurable network, (iii) it tracks dynamic changes in dispersion due to environment, and (iv) it achieves a high degree of accuracy necessary for 40 Gb/s channels (see Figure 23).

One brute-force method to achieve tunable dispersion compensation is to build a module with optical switches used to add or remove sections of fixed DCF to achieve a discrete set of dispersion compensation. Many other elegant, yet viable, approaches have been developed for tunable dispersion compensation, and these approaches include linearly chirped FBG with nonuniform heating, nonlinearly chirped Fiber Bragg grating (FBG) with a simple mechanical stretcher, virtually imaged phase array, electronic tap delay filter with weights, multiple stage all pass filter, etc.

System Performance Parameters

An eye diagram provides a simple and useful tool to visualize intersymbol interference between data bits. Figure 24a shows a perfect eye diagram. A square bit stream (i.e., series of symbol '1's and '0's) is sliced into sub-bit stream with predetermined eye intervals (i.e., several bit periods), and displayed through bit analyzing equipment (e.g., digital channel analyzer), overlapping the sliced sub-bit stream in order to obtain the eye diagram. When a perfect transmitter and receiver (i.e., infinite receiver bandwidth with zero noise characteristics and jitter), and a perfect transmission media (i.e., no dispersion and nonlinearites) are used, the received eye diagram is shaped as a perfect rectangular. In reality, the transmitter and receiver have a limited bandwidth with noise and jitter, and the transmission media (i.e., optical fiber) has dispersion and nonlinearites. Therefore, the eye diagram deviates from the perfect rectangular shape. Figure 24b shows the eye diagram close to a real situation. The shape of the eye is generally broadened and distorted (i.e., eye is closed) due to limited bandwidth and fiber impairments, and noise and timing jitter are added onto this broadened and distorted eye shape.

The Q-factor (q) is also an important system parameter widely used in long-distance optical transmission system design. It is defined as the electrical signal-to-noise ratio before the decision circuit at receiver. The Q-factor is directly related to bit-error-rate (BER: the percentage of bits that has an error relative to total number of bits received in a specific time) by: $\mathrm{BER} = 0.5 \cdot \mathrm{erfc}(q/2^{0.5})$, where $\mathrm{erfc}(x)$ is the complementary error function. The Q-factor is an unitless linear ratio, and is express in dB by $20 \cdot \log(q)$. As a conventional relationship, Q-factor of 15.6 dB (i.e., linear ratio 6) is required to achieve $\mathrm{BER} = 10^{-9}$. Figure 25 shows the relationship of eye diagram, Q-factor, and BER.

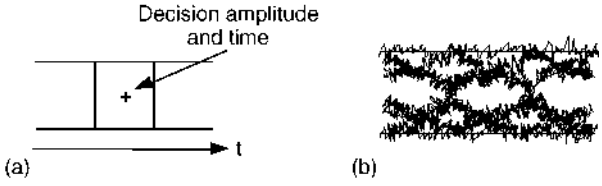

Figure 24 (a) A perfect square eye diagram, (b) a closed eye diagram due to bandwidth, fiber impairments, noise, and timing jitter.

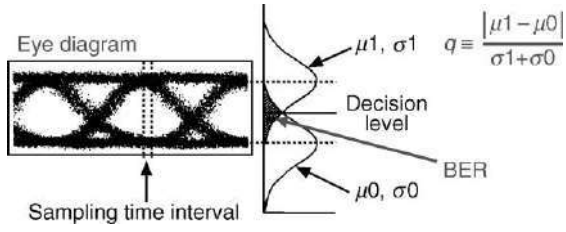

Figure 25 The relationship of eye diagram, Q-factor, and BER ($\mu 1$, $\mu 2$: mean levels of logic '1' and '0', $\sigma 1$, $\sigma 2$: standard deviation of logic '1' and '0').

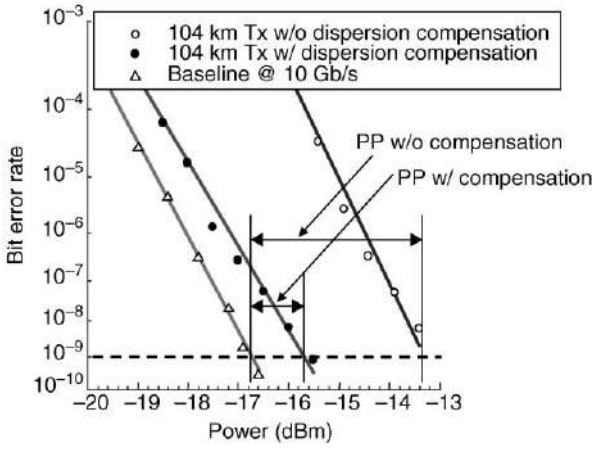

Figure 26 BER curve and power penalty.

Power penalty (PP) is one of the most important system parameters, and is defined as the received optical power difference in dB with and without signal impairments at a specified BER (conventionally 10^{-9}), from measured BER versus optical power curve. Therefore, 1 dB PP means that a system with signal impairments requires 1 dB more optical power at the receiver in order to achieve the same BER performance compared to the system without signal impairments. Figure 26 shows the typical BER curve as a function of received optical power and PP. Note that PP reduces from ~3.5 dB to ~1 dB with dispersion compensation.

In long-haul and undersea transmission system with many EDFA chains, OSNR is an important system parameter to design and characterize the optical transport system. OSNR is defined as the ratio of the power of signal channel to the power of ASE in a specified optical bandwidth (conventionally 0.1 nm). The OSNR (in dB) of a signal channel at the end of the system is approximated by: OSNR (dB) $= 58 + P_{\text{out}} - L_{\text{span}} - \text{NF} - 10\log(N_{\text{amp}})$, when the system has N_{amp} fiber spans, span loss L_{span} (in dB) followed by an optical amplifier with output power P_{out} (in dBm) per channel launched into the span and noise figure (NF) (in dB). As an example, a typical OSNR requirement for BER $= 10^{-9}$ is about 17 dB at 10 Gb/s data rate. The required OSNR should be increased by 6 dB when the data rate is increased by a factor of 4. Improving the amplifier's NF can increase OSNR, and the improved OSNR can be used to increase the system reach, reducing the channel power in order to suppress the nonlinearities, etc.

See also

Fiber and Guided Wave Optics: Light Propagation. **Imaging:** Multiplex Imaging. **Optical Amplifiers:** Erbrium Doped Fiber Amplifiers for Lightwave Systems. **Optical Communication Systems:** Wavelength Division Multiplexing.

Further Reading

Leonid K, Sergio B and Alan W (1996) *Optical Fiber Communication Systems*. Norwood, MA: Artech House Inc.

Govind PA (1995) *Nonlinear Fiber Optics*. San Diego, CA: Academic Press.

Emmanuel D (1994) *Erbium-Doped Fiber Amplifiers*. New York: Wiley.

Ivan K and Tingye L (2002) *Optical Fiber Telecommunications IV A Components*. San Diego, CA: Academic Press.

Ivan K and Tingye L (2002) *Optical Fiber Telecommunications IV B Systems and Impairments*. San Diego, CA: Academic Press.

Historical Development

G Keiser, PhotonicsComm Solutions, Inc., Newton Center, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

This article describes the development of optical fiber communication systems. After describing some of the motivations for using optical fiber communications and the advantages of this technology, the key milestones and the principal people involved in developing optical fibers and compatible light sources are presented. Following this, the article looks at the evolution of fielded systems and the use of optical fiber links in undersea applications.

One of the principal needs of people since antiquity has been to communicate. This need created interests in devising communication systems for sending messages from one distant place to another. Many forms of such systems have appeared over the years. The basic motivations behind each new one were either to improve the transmission fidelity, to increase the data rate so that more information could

be sent, or to increase the transmission distance between relay stations. Prior to the nineteenth century, all communication systems operated at a very low information rate and involved only optical or acoustical means, such as signal lamps or horns. One of the earliest known optical transmission links, for example, was the use of a fire signal by the Greeks in the eighth century BC for sending alarms, calls for help, or announcements of special events.

The invention of the telegraph by Samuel FB Morse in 1838 ushered in the era of electrical communications. In the ensuing years an increasingly larger portion of the electromagnetic spectrum, shown in Figure 1, was utilized for the conveying of information from one place to another. The reason for this trend is that, in electrical systems, the data usually are transferred over the communication channel by superimposing the information onto a sinusoidally varying electromagnetic wave, which is known as the carrier. When reaching its destination the information is removed from the carrier wave and processed as desired. Since the amount of information that can be transmitted is related directly to the frequency range over which the carrier operates, increasing the carrier frequency theoretically increases the available transmission bandwidth and, consequently, provides a larger information capacity. Thus the trend in electrical communication system developments was to employ progressively higher frequencies (shorter wavelengths), which offer corresponding increases in bandwidth or information capacity. This activity led to the birth of communication mechanisms such as radio, television, microwave, and satellite links.

Another important portion of the electromagnetic spectrum encompasses the optical region shown in Figure 1. In contrast to electrical communications, transmission of information in an optical format is carried out, not by frequency modulation of the carrier but by varying the intensity of the optical power. Similar to the radio-frequency spectrum, two classes of transmission medium can be used: an atmospheric channel or a guided-wave channel. Whereas transmission of optical signals through the atmosphere was done thousands of years ago, the use of a guided-wave optical channel such as an optical fiber is a fairly recent application.

Fiber optic communication systems have a number of inherent advantages over their copper-based and radio-transmission counterparts. Fiber optic cable can transmit at a higher capacity, thereby reducing the number of physical lines and the amount of equipment needed for a given transmission span. The lower weight and smaller size of optical fibers offer a distinct advantage over heavy, bulky copper cables in crowded underground cable ducts, in ceiling-mounted cable trays, and in mobile platforms such as aircraft and ships. Their dielectric composition is an especially important feature of optical fibers, since this ensures freedom from electromagnetic interference between adjacent fibers, eliminates ground loops, and results in extremely low fiber-to-fiber crosstalk. In addition, an optical fiber affords a high degree of data security since the optical signal is well confined within the optical waveguide.

Optical Fiber Development

Since an atmospheric channel requires a line-of-sight link and because it can be adversely affected by weather conditions, a guided-wave channel is the preferred approach in most cases for communication system applications. A challenge in using an optical fiber channel is to have a flexible, low-loss medium that transfers the optical signal over long distances

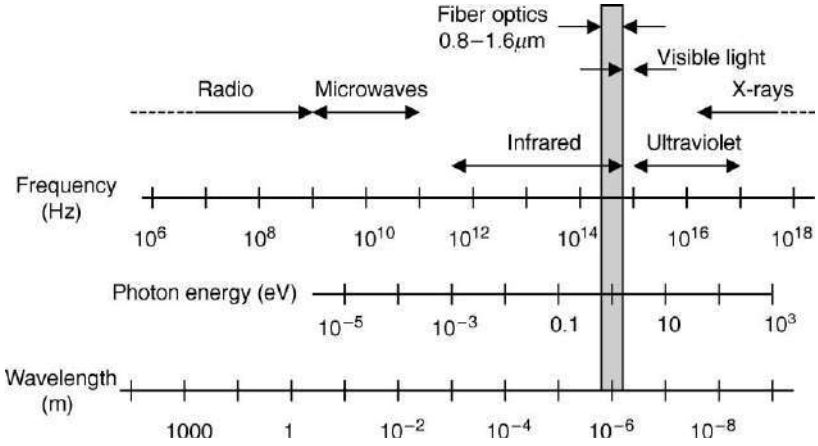

Figure 1 Electromagnetic spectrum for electrical and optical communications.

without significant attenuation and distortion. Glass is an obvious material for such applications. The earliest known glass was made around 2500 BC and glass already was drawn into fibers during the time of the Roman Empire. However, such glasses have very high attenuations and thus are not suitable for communication applications. One of the first known attempts of using optical fibers for communication purposes was a demonstration in 1930 by Heinrich Lamm of image transmission through a short bundle of optical fibers for potential medical imaging. However, no further work was done beyond the demonstration phase since the technology for producing reasonably low-loss fibers with good light confinement was not yet mature.

Further work and experiments continued on using optical fibers for image transmission and by 1960 glass-clad fibers had attenuations of about one decibel per meter. This attenuation allowed fibers to be used for medical imaging, but it was still much too high for communications. Optical fibers had attracted the attention of researchers at that time, because they were analogous in theory to plastic dielectric waveguides used in certain microwave applications. In 1961, Elias Snitzer demonstrated this similarity by drawing fibers with cores so small they carried light in only one waveguide mode. He published a classic theoretical description of single-mode fibers in May 1961. However, to be useful for communication systems, optical fibers would need to have a loss of no more than 10 or 20 decibels per kilometer.

As a reminder, decibels measure the ratio of the output power to the input power on a logarithmic scale. The abbreviation for a decibel is 'dB.' The power ratio in decibels is given by the expression '10 log (power out/power in).' As an example, a power loss of 20 dB over a 1-km distance in an optical fiber means that only 1% of the light injected into the fiber comes out of the other end. Table 1 gives some typical examples of various power ratios and their decibel equivalents.

In the early 1960s Charles Kao pursued the idea of using a clad glass fiber for an optical waveguide, building on optical waveguide research being done at the Standard Telecommunication Laboratories in England. After he and George Hockman painstakingly examined the transparency properties of various types of glass, Kao made a prediction, in 1966, that losses of no more than 20 dB/km were possible in optical fibers. In July 1966, Kao and Hockman presented a detailed analysis for achieving such a loss level. Kao then went on to actively advocate and promote the prospects of fiber communications, which generated interest in laboratories around the world to reduce fiber loss. It took four years to reach Kao's predicted goal of 20 dB/km, and the final solution was different from what many had expected.

To understand the process of making a fiber, consider the schematic of a typical fiber structure shown in Figure 2. A fiber consists of a solid dielectric (glass or plastic) cylinder of refractive index n_1 called the core. This is surrounded by a dielectric cladding which has a refractive index n_2, that is less than n_1, in order to achieve light guiding in the fiber. This structure is then encapsulated by a buffer coating to protect the fiber from abrasion and outside contaminants. The first step in making a fiber is to form the clear glass rod or tube called a preform. This normally is done by a vapor-phase oxidation process. The preform has two distinct regions that correspond to the core and cladding of the eventual fiber. Fibers are made from the preform by precision feeding it into a circular furnace that softens the end of the preform to the point where it can be drawn into a very thin filament which becomes the optical fiber.

Most researchers had tried to purify the compound glasses used for standard optics, which are easy to melt and draw into fibers. A different approach was taken at the Corning Glass Works where Robert Maurer, Donald Keck, and Peter Schultz started with fused silica. This material can be made extremely pure, but has a high melting point and a low refractive index. Silica has the approximate attenuation versus wavelength characteristic shown in Figure 3. Note that for early silica fibers there are regions of low attenuation around 850, 1310, and 1550 nm, which the literature refers to as the first, second, and third windows, respectively. The large attenuation spikes in the 1000- and 1400-nm spectral

Table 1 Examples of some optical power ratios and their decibel equivalents

Power ratio (P_{out}/P_{in})	*Decibel equivalent (dB)*
0.01	−20
0.10	−10
0.50	−3
1	0
2	3
10	10
100	20

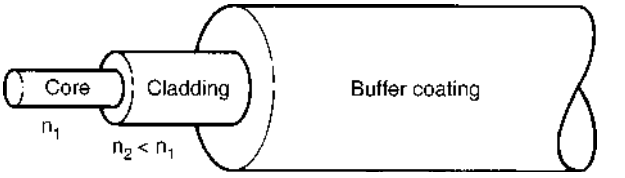

Figure 2 Typical optical fiber structure.

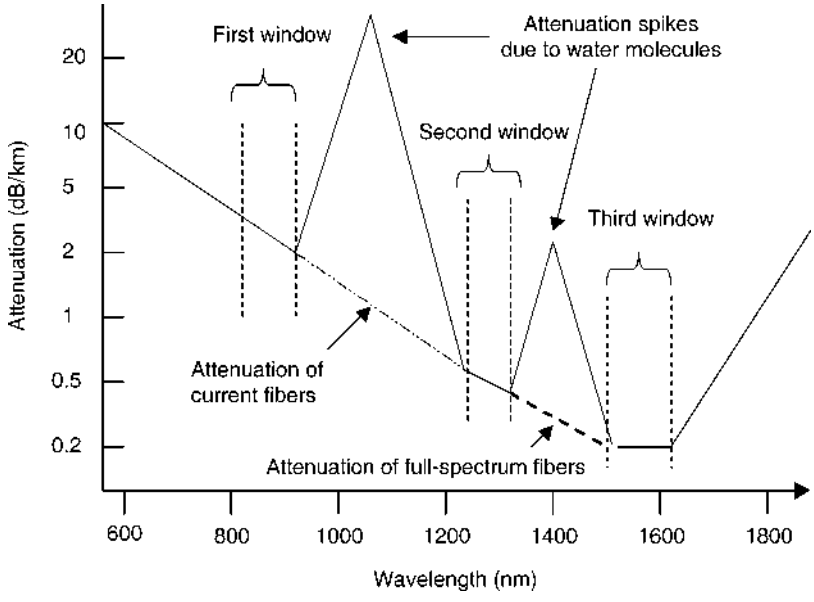

Figure 3 Attenuation versus wavelength approximation for silica fibers.

regions are due to absorption by residual water molecules in the glass. Although ultra-pure material processing techniques can eliminate these spikes to produce what are known as full-spectrum or low-water-peak fibers, most installed fibers still have a relatively large attenuation between 1350 and 1500 nm.

Note that since the attenuation spikes no longer are present in various types of fibers, the idea of operating windows has been replaced by the concept of spectral bands. The International Telecommunications Union (ITU) has designated six spectral bands for use in intermediate-range and long-distance optical fiber communications within the 1260 to 1675-nm region. The regions, which are known by the letters O, E, S, C, L, and U, are defined as follows:

- Original Band (O-Band): 1260 to 1360 nm
- Extended Band (E-Band): 1360 to 1460 nm
- Short Band (S-Band): 1460 to 1530 nm
- Conventional Band (C-Band): 1530 to 1565 nm
- Long Band (L-Band): 1565 to 1625 nm
- Ultra-long Band (U-Band): 1625 to 1675 nm

The Corning team made cylindrical preforms by depositing purified materials from the vapor phase. To produce a fiber that has light guiding properties they carefully added controlled trace levels of titanium to the core to make its refractive index slightly higher than that of the cladding without raising the attenuation significantly. In September 1970, they announced the fabrication of single-mode fibers with an attenuation of 17 dB/km at the 633-nm helium-neon line. The fibers were fragile, but independent tests at the British Post Office Research Laboratories facility in Martlesham Heath, England confirmed the low loss.

This dramatic breakthrough was the first among the many developments that opened the door to fiber optic communications. The ensuing years saw further reductions in optical fiber attenuations. By mid-1972, Maurer, Keck, and Schultz at Corning had made multimode germania-doped fibers with a 4-dB/km loss and much greater strength than the earlier brittle titania-doped fibers. In order to couple a sufficient amount of optical power into a fiber, early applications used multimode fibers with a refractive-index gradient between core and cladding of around 2% and core diameters of 50 or 62.5 micrometers.

Single-mode fibers have much smaller core diameters on the order of 9 micrometers in order to allow only one propagation mode. This type of fiber has a much higher transmission capacity since the effect of modal dispersion is eliminated. The first single-mode fibers were optimized to have a zero dispersion value at 1310 nm, since the silica material used at that time exhibited a low loss within a spectral band around this wavelength. Figure 4 shows the dispersion characteristic of this type of fiber as a function of wavelength in the S-, C-, and L-bands. As a result of its widespread use in early single-mode transmission systems, this fiber design has been standardized by the International Telecommunication Union (ITU-T) under the designation Recommendation G.652.

Standard G.652 silica fibers provide the lowest attenuation at 1550 nm, but have a much larger signal

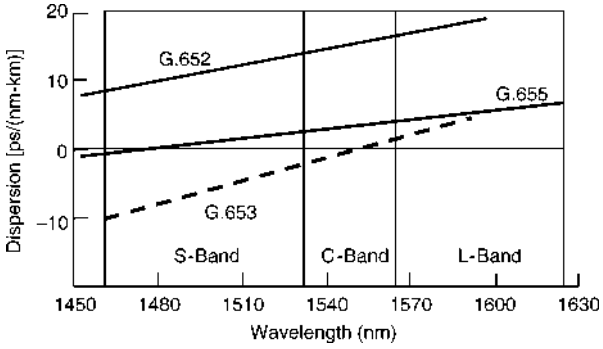

Figure 4 Dispersion characteristics of three types of standard optical fibers.

dispersion at this wavelength than at 1310 nm. Since system designers wanted to use fibers at a 1550-nm wavelength, in order to transmit high-speed data over longer distances, fiber manufacturers overcame the larger signal dispersion limitation by creating the so-called dispersion-shifted fibers. This was done through a clever manipulation of the core and cladding designs that allowed the zero-dispersion point to shift to longer wavelengths. In particular, the ITU-T created a specification for operation at 1550 nm which is designated as recommendation G.653. Figure 4 also shows the dispersion characteristic of this type of fiber as a function of wavelength.

Although the G.652 and G.653 fibers work well for single-wavelength operation, a different type of fiber, having non-zero dispersion within a broad spectral range, is needed when implementing systems that use many independent light sources simultaneously within a particular wavelength band. This led to the specification of the non-zero dispersion shifted fiber (NZDSF) in the ITU-T Recommendation G.655 that are designated to operate around 1550 nm. The main purpose of having a positive dispersion value over the entire operating spectrum is to mitigate a nonlinear optical effect called four-wave mixing (FWM), which is analogous to intermodulation distortion in electrical systems. Figure 4 shows the dispersion characteristic of the G.655 fiber in the S-, C-, and L-bands.

Light Sources

The fiber by itself is not practical unless there is a compatible optical source for launching light signals into it. The most suitable device for this is a semiconductor light-emitting diode (LED) or a laser diode. In the 1960s, a great deal of effort took place to achieve laser action in pn-junction diodes. The early devices were GaAs and GaAsP lasers that operated at a temperature of 77 K, emitted at a wavelength around 850 nm, and had high lasing threshold current densities. To make devices that were more application-friendly by operating at room temperature, structures consisting of sandwiched layers of AlGaAs and GaAs were investigated. Finally, in 1970, researchers at Bell Laboratories and a team at the Ioffe Physical Institute in Leningrad made the first semiconductor diode lasers based on a layered AlGaAs/GaAs/AlGaAs structure that were able to emit continuous-wave light in the 850-nm region at room temperature.

Major improvements in laser diode performance and reliability followed this achievement during the next decade. In addition, around 1976, room-temperature laser diodes operating at longer wavelengths, in the 1100 to 1600-nm region, started to appear. Of particular interest were GaInAsP/InP-based laser diodes emitting in the 1310-nm and 1550-nm low-loss windows of optical fibers. The progressive development of ever-improving devices during the 1980s and 1990s, included single-frequency emission with narrow linewidths under continuous operation, low levels of chirp under direct modulation, high output power, and the ability to tune specially constructed laser diodes over a wavelength range of up to 30 nm.

Fielded Systems

The bit rate-distance product $B \times L$, where B is the transmission bit rate and L is the repeater spacing, measures the transmission capacity of optical fiber links. Since the inception of optical fiber communications in the mid-1970s, the link transmission capacity has experienced a tenfold increase every four years. Several major technology advances spurred this growth. Among the technology developments are laser diodes emitting over an extremely narrow spectral band, optical amplifiers, fibers with low losses and low dispersions, and the concepts of wavelength division multiplexing.

Some of the initial telephone-system field trials in the USA were carried out in 1977 by GTE in Los Angeles and by AT&T in Chicago. These transmission links were largely for the trunking of telephone lines, which are digital links consisting of time-division-multiplexed 64-kb/s voice channels. Similar demonstrations were carried out in Europe and Japan. Applications ranged from 45 to 140 Mb/s with repeater spacings around 10 km.

With the advent of high-capacity fiber optic transmission lines in the 1980s, service providers established a standard signal format called synchronous optical network (SONET) in North America and synchronous digital hierarchy (SDH) in other

parts of the world. These standards define a synchronous frame structure for sending multiplexed digital traffic over optical fiber trunk lines. The basic building block and first level of the SONET signal hierarchy is called the synchronous transport signal–level 1 (STS-1), which has a bit rate of 51.84 Mb/s. Higher-rate SONET signals are obtained by byte-interleaving N STS-1 frames, which are then scrambled and converted to an optical carrier–level N (OC-N) signal. Thus the OC-N signal will have a line rate exactly N times that of an OC-1 signal. For SDH systems, the fundamental building block is the 155.52-Mb/s synchronous transport module–level 1 (STM-1). Again, higher-rate information streams are generated by synchronously multiplexing N different STM-1 signals to form the STM-N signal. Table 2 shows commonly used SDH and SONET signal levels.

The development of optical sources and photodetectors capable of operating at 1310 nm allowed a shift in the transmission wavelength from 850 nm to 1310 nm. This resulted in a substantial increase in the repeaterless transmission distance for long-haul telephone trunks, since optical fibers exhibit lower power loss and less signal dispersion at 1310 nm. Intercity applications first used multimode fibers, but in 1984 switched exclusively to single-mode fibers, which have a significantly larger bandwidth. Bit rates for long-haul links typically range between 155 and 622 Mb/s (OC-3 and OC-12), and in some cases up to 2.5 Gb/s (OC-48), over repeater spacings of 40 km. Both multimode and single-mode 1310-nm fibers are used in local area networks, where bit rates range from 10 Mb/s to 100 Mb/s over distances ranging from 500 m to tens of kilometers.

In the next step of system evolution, links operating in the low-loss window around 1550 nm attracted much attention for high-capacity, long-span terrestrial and undersea transmission links. These links routinely carry traffic at 2.5 Gb/s over 90-km repeaterless distances. By 1996, advances in high-quality lasers and receivers allowed single-wavelength transmission rates of 10 Gb/s (OC-192).

In 1989, the introduction of optical amplifiers gave a major boost to fiber transmission capacity. Although there are GaAlAs-based solid-state optical amplifiers for the first window and InGaAsP amplifiers for the second window, the most successful and widely used devices are erbium-doped fiber amplifiers (commonly called EDFAs) operating in the 1530 to 1560-nm range and Raman fiber amplifiers that are used for operation in the 1560 to 1600-nm region.

During the same time period, impressive demonstrations of long-distance high-capacity systems were made using optical soliton signals. A soliton is a nondispersive pulse that makes use of nonlinear dispersion properties in a fiber to cancel out chromatic dispersion effects. As an example, solitons at rates of 10 Gb/s have been sent over a 12 200-km experimental link using optical amplifiers and special modulation techniques.

Entrance of Wavelength Division Multiplexing

The use of wavelength division multiplexing (WDM) offers a further boost in fiber transmission capacity. The basis of WDM is to use multiple sources operating at slightly different wavelengths to transmit several independent information streams over the same fiber. Although researchers started looking at WDM in the 1970s, during the ensuing years it generally turned out to be easier to implement higher-speed electronic and optical devices than to invoke the greater system complexity called for in WDM. However, a dramatic surge in WDM popularity started in the early 1990s, as electronic devices neared their modulation limit and high-speed equipment became increasingly complex.

Figure 5 shows the concept of implementing many closely spaced wavelengths within a spectral band centered around 1552.524 nm. This scheme is referred to as dense WDM or DWDM. Conceptually, the DWDM scheme is the same as frequency

Table 2 Commonly used SONET and SDH transmission rates

SONET level	*Electrical level*	*Line rate (Mb/s)*	*SDH equivalent*	*Common rate name*
OC-1	STS-1	51.84	–	
OC-3	STS-3	155.52	STM-1	155 Mb/s
OC-12	STS-12	622.08	STM-4	622 Mb/s
OC-24	STS-24	1244.16	STM-8	
OC-48	STS-48	2488.32	STM-16	2.5 Gb/s
OC-96	STS-96	4976.64	STM-32	
OC-192	STS-192	9953.28	STM-64	10 Gb/s
OC-768	STS-768	39,813.12	STM-256	40 Gb/s

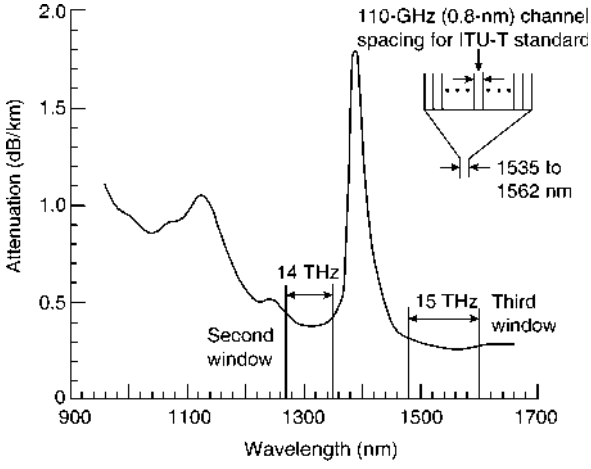

Figure 5 Wavelength division multiplexing (WDM) concept.

division multiplexing (FDM) used in microwave radio and satellite systems. Just as in FDM, the wavelengths (or optical frequencies) in a DWDM link must be properly spaced to avoid interference between channels. In an optical system this interference may arise from the fact that the center wavelength of laser diode sources and the spectral operating characteristics of other optical components in the link may drift with temperature and time, thereby giving rise to the need for a guard band between wavelength channels.

Since WDM is essentially frequency division multiplexing at optical carrier frequencies, the ITU developed DWDM standards that specify channel spacings in terms of frequency. The ITU-T Recommendation G.694.1, which is entitled 'Dense Wavelength Division Multiplexing (DWDM),' specifies WDM operation in the S-, C-, and L-bands for high-quality, high-rate metro area network (MAN) and wide area network (WAN) services. It calls out for narrow frequency spacings of 100 to 12.5 GHz (or, equivalently, 0.8 to 0.1 nm at 1550 nm). This implementation requires the use of stable, high-quality, temperature-controlled and wavelength-controlled (frequency-locked) laser diode light sources.

With the production of full-spectrum (low-water-content) fibers, the development of relatively inexpensive optical sources, and the desire to have low-cost optical links operating in metro and local area networks, came the concept of coarse WDM (CWDM). In 2002, the ITU-T released a standard aimed specifically at CWDM. This is Recommendation G.694.2, which is entitled 'Coarse Wavelength Division Multiplexing (CWDM).' The CWDM grid is made up of 18 wavelengths defined within the range 1270 nm to 1610 nm (O-through L-bands) spaced by 20 nm with wavelength-drift tolerances of ± 2 nm. This can be achieved with inexpensive light sources that are not temperature-controlled. The targeted transmission distance for CWDM is 50 km on single-mode fibers, such as those specified in ITU-T Recommendations G.652, G.653, and G.655.

Wavelength tunability of a source is an important property of WDM systems. Obviously it is not desirable or practical to maintain an inventory of dozens of lasers that emit at different wavelengths for WDM applications. The ideal tunable laser should be adjustable to emit at a specific wavelength across a broad spectral range. One such device is a distributed Bragg reflector (DBR) laser diode that can be tuned over a 10 to 20 nm spectral range. Work on perfecting such devices are still underway.

Starting in the mid-1990s, a combination of EDFAs and WDM was used to boost fiber information capacity to even higher levels and to increase the transmission distance. A major system consideration in these super-high capacity links is to ensure that there is appropriate link and equipment redundancy, so that alternate paths are available in case of disruptions in communications resulting from cable ruptures (for example, caused by errant digging from a backhoe) or equipment failures at an intermediate node. Such disruptions otherwise could have a devastating effect on a large group of users.

Undersea Optical Cable Systems

The first transoceanic fiber optic cable systems were installed in the Atlantic and Pacific Oceans in 1988 and 1989. Initially these systems operated at 280 Mb/s per fiber pair using 1310-nm lasers and single-mode fibers. The links consisted of a series of point-to-point optical fiber segments between electronic-based undersea regeneration points that were located nominally 60 km apart. Later the transmission capacity of these links was upgraded to 2.5 Gb/s and the regenerator spacing was increased to 100 km by converting the 1310-nm multiple-frequency light sources to 1550-nm single-frequency laser diodes. Later, the regenerator spacing was increased to 140 km.

Although these cable systems significantly improved the quality of the international telephone service, the optical-to-electrical conversion process at each regeneration point remained a capacity bottleneck. The introduction of erbium-doped optical fiber amplifiers (EDFA) eliminated this bottleneck from undersea lightwave systems by amplifying signals

directly in the optical domain. Since EDFAs operate over a 30-nm wavelength band, they are well suited for use with undersea WDM links, which have provided a further capacity increase. These undersea optical amplifiers are typically spaced about 45 km apart.

One example of the many worldwide installations of optically amplified WDM networks is the SEA-ME-WE-3 Cable System. This undersea network runs from Germany to Singapore, connecting more than a dozen countries in between. Hence the name SEA-ME-WE, which refers to Southeast Asia (SEA), the Middle East (ME), and Western Europe (WE). The network has two pairs of undersea fibers with a capacity of eight STM-16 wavelengths per fiber.

See also

Lasers: Optical Fiber Lasers. **Optical Communication Systems:** Architectures of Optical Fiber Communication Systems; Wavelength Division Multiplexing.

Further Reading

Bergano NS (1997) Undersea amplified lightwave system design. In: Kaminow IP and Koch TL (eds) *Optical Fiber Telecommunications*, vol. IIIA, pp. 302–335. New York: Academic.

Goralski WJ (2002) SONET/SDH, 3rd edn. New York: McGraw-Hill.

Hecht J (1999) *City of Light*. New York: Oxford University Press.

Joyner CH (1997) Semiconductor laser growth and fabrication technology. In: Kaminow IP and Koch TL (eds) *Optical Fiber Telecommunications*, vol. IIIB, pp. 163–199. New York: Academic.

Kaiser P and Keck DB (1988) Fiber types and their status. In: Miller SE and Kaminow IP (eds) *Optical Fiber Telecommunications II*, pp. 29–54. New York: Academic.

Kao KC and Hockman GA (July 1966) Dielectric-fibre surface waveguides for optical frequencies. *Proceedings IEE*, vol. 113, 1151–1158.

Keiser G (2003) *Optical Communications Essentials*. New York: McGraw-Hill.

Keiser G (2000) *Optical Fiber Communications*, 3rd edn. Burr Ridge, IL: McGraw-Hill.

Mollenauer LF, Gordon JP and Mamyshev PV (1997) Solitons in high bit rate long-distance transmission. In: Kaminow IP and Koch TL (eds) *Optical Fiber Telecommunications*, vol. IIIA, pp. 373–460. New York: Academic.

Nyman B, Farries M and Si C (Oct 2001) Technology trends in dense WDM demultiplexers. *Optical Fiber Technology* 7(4): 255–274.

Ramaswami R and Sivarajan KN (2002) *Optical Networks*, 2nd edn. San Francisco, CA: Morgan Kaufmann.

Snitzer E (May 1961) Cylindrical dielectric waveguide modes. *Journal Optical Society of America* 51: 491–498.

Trischitta PR and Marra WC (1998) Applying WDM technology to undersea cable networks. *IEEE Communication Magazine* 36(2): 62–66.

Architectures of Optical Fiber Communication Systems

G Keiser, PhotonicsComm Solutions, Inc., Newton Center, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Simple Optical Fiber Links

The most basic optical fiber communication system architecture is a point-to-point link consisting of the elements shown in Figure 1. The key sections are a transmitter consisting of a light source and its associated drive circuitry, a cable offering mechanical and environmental protection to the optical fibers contained inside, and a receiver consisting of a photodetector plus amplification and signal-restoring circuitry. Very long links may include periodic optical amplifiers for boosting the level of the optical signal. In networks that use only simple point-to-point links, the optical fiber is used principally as a transmission medium and all the switching and processing of information bits is done by electronics.

At the transmitting end of an optical fiber link, information bits consisting of electronic pulses are used to modulate the light output of the source. The resulting light pulses are coupled into an optical fiber. As they travel along the fiber these signal pulses become progressively weakened and distorted due to attenuation and dispersion mechanism that are characteristic of an optical fiber. At the end of the fiber link a photodetector converts the optical pulses back to an electrical format for switching to another link or for processing at the end station.

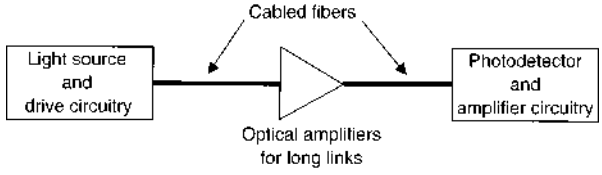

Figure 1 Basic elements of a point-to-point fiber optic link.

The cabled fiber is one of the most important elements in an optical fiber link, since it determines operational parameters such as transmission distance and achievable data rates. In addition to protecting the glass fibers during installation and service, the cable may contain copper wires for powering optical amplifiers or signal regenerators, which are needed periodically in long-distance links for amplifying and reshaping the signal.

Analogous to copper cables, the installation of optical fiber cables can be either aerial, in ducts, undersea, or buried directly in the ground. As a result of installation and/or manufacturing limitations, individual cable lengths will range from several hundred meters to several kilometers. Practical considerations such as reel size and cable weight determine the actual length of a single cable section. The shorter segments tend to be used when the cables are pulled through ducts. Longer lengths are used in aerial, direct-burial, or undersea applications. Splicing together individual cable sections forms continuous transmission lines for these long-distance links. For undersea installations, the splicing and repeater-installation functions are carried out on board a specially designed cable-laying ship.

One of the principal characteristics of an optical fiber is its attenuation as a function of wavelength, as shown in Figure 2. Early technology used the 800 to 900 nm wavelength band, since in this region the fibers made at that time exhibited a local minimum in the attenuation curve, and optical sources and photodetectors operating at these wavelengths were available. This region originally was referred to as the 'first window'. By reducing the concentration of hydroxyl ions and metallic impurities in the fiber material, in the 1980s manufacturers were able to fabricate optical fibers with very low loss in the 1100 to 1600 nm region. Two windows are defined here, these being the 'second window' centered around 1310 nm and the 'third window' centered around 1550 nm.

In 1998 a new ultra-high purifying process patented by Lucent Technologies eliminated virtually all water molecules from the glass material. The resulting fiber is designated as either a low-water-peak or a full spectrum fiber. By dramatically reducing the water-attenuation peak around 1400 nm, this process opens the transmission region between the second and third windows to provide around 100 nm more bandwidth than in conventional single-mode fibers. This particular fiber allows wider spectral tolerances on optical components, which reduces costs for moderate-distance metropolitan network applications.

Since the attenuation of low-water-peak fibers makes the designation of transmission windows obsolete, the concept of operational spectral bands arose for the 1260–1675 nm region. As shown in

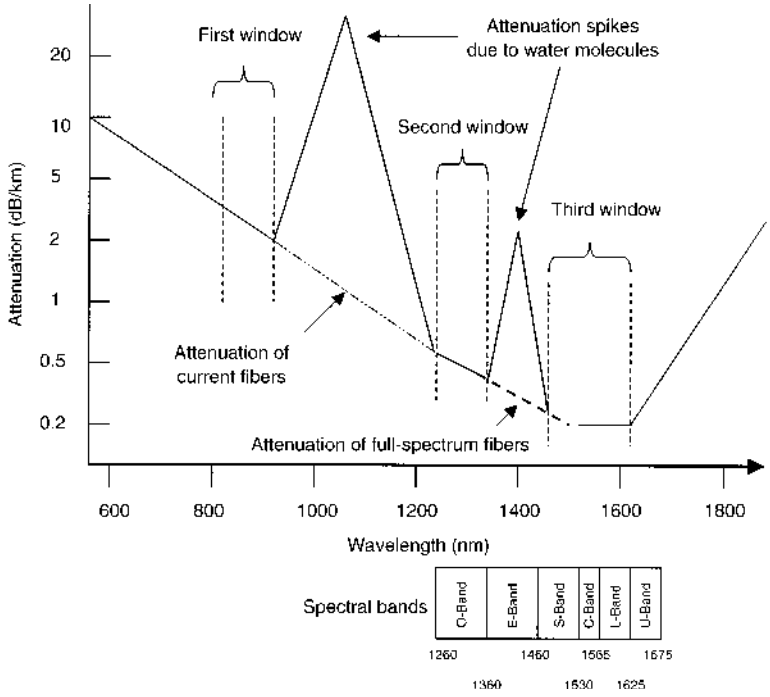

Figure 2 Attenuation versus wavelength for silica fibers and the concept of spectral bands.

Figure 2, the regions are known by the letters O, E, S, C, L, and U, which are defined as follows:

- Original band (O-Band): 1260–1360 nm;
- Extended band (E-Band): 1360–1460 nm;
- Short band (S-Band): 1460–1530 nm;
- Conventional band (C-Band): 1530–1565 nm;
- Long band (L-Band) 1565–1625 nm;
- Ultra-long band (U-Band): 1625–1675 nm.

Once the cable is installed, a light source that is dimensionally compatible with the fiber core is used to launch optical power into the fiber. Semiconductor light-emitting diodes (LEDs) and laser diodes are suitable for this purpose, since their light output can be modulated rapidly by simply varying the bias current at the desired transmission rate, thereby producing an optical signal. The electric input signals to the transmitter circuitry for the optical source can be either of an analog or digital form. For high-rate systems (nominally greater than 1 Gb/s), direct modulation of the source can lead to unacceptable signal distortion. In this case, an external modulator is used to vary the amplitude of a continuous light output from a laser diode source.

After an optical signal is launched into a fiber, it will become progressively attenuated and distorted with increasing distance because of scattering, absorption, and dispersion mechanisms in the glass material. At the receiver a photodiode will detect the weakened optical signal emerging from the fiber end and convert it to an electric current. The design of an optical receiver is inherently more complex than that of the transmitter, since it has to interpret the content of the weakened and degraded signal received by the photodetector. The principal figure of merit for a receiver is the maximum optical power necessary at the desired data rate to attain either a given error probability for digital systems or a specified signal-to-noise ratio for an analog system. The ability of a receiver to achieve a certain performance level depends on the photodetector type, the effects of noise in the system, and the characteristics of the successive amplification stages in the receiver.

Networks of Links

With the advent of fiber optic transmission lines, the next step in the evolution of the digital multiplexing scheme used by telecommunication providers was a standard signal format called synchronous optical network (SONET) in North America and synchronous digital hierarchy (SDH) in other parts of the world. In the mid-1980s, efforts started on developing a standard that would allow network engineers to interconnect fiber optic transmission equipment from various vendors through multiple-owner trunk networks. This resulted in a series of standards for SONET by the American National Standards Institute (ANSI) and a series of recommendations for SDH from the International Telecommunications Union (ITU). Examples of particular interest are the ANSI T1.105.06 standard and the ITU-T G.957 recommendation. Although there are some implementation differences between SONET and SDH, all SONET specifications conform to the SDH recommendations. A key characteristic of SONET and SDH is that they are usually configured as a ring architecture. This is done to create loop diversity for uninterrupted service protection purposes in case of link or equipment failures. The SONET/SDH rings commonly are called self-healing rings, since the traffic flowing along a certain path can be switched automatically to an alternate or standby path following failure or degradation of the link segment.

The basic building block and first level of the SONET signal hierarchy is called the Synchronous Transport Signal – Level 1 (STS-1), which has a bit rate of 51.84 Mb/s. Higher-rate SONET signals are obtained by byte-interleaving N STS-1 frames, which are then scrambled and converted to an Optical Carrier – Level N (OC-N) signal. Thus the OC-N signal will have a line rate exactly N times that of an OC-1 signal. For SDH systems the fundamental building block is the 155.52 Mb/s Synchronous Transport Module – Level 1 (STM-1). Again, higher-rate information streams are generated by synchronously multiplexing N different STM-1 signals to form the STM-N signal.

Three main features, each with two alternatives, classify all SONET/SDH rings, thus yielding eight possible combinations of ring types. First, there can be either two or four fibers running between the nodes on a ring. Second, the operating signals can travel either clockwise only (which is termed a unidirectional ring) or in both directions around the ring (which is called a bidirectional ring) Third, protection switching can be performed either via a line-switching or a path-switching scheme. Upon link failure or degradation, line switching moves all signal channels of an entire OC-N channel to a protection fiber. Conversely, path switching can move individual payload channels within an OC-N channel (e.g., an STS-1 channel within an OC-12 channel) to another path.

Of the eight possible combinations of ring types, the following two architectures have become popular for SONET and SDH networks:

- two-fiber unidirectional path-switched ring (two-fiber UPSR);
- two-fiber or four-fiber bidirectional line-switched ring (two-fiber or four-fiber BLSR).

The common abbreviations of these configurations are given in parentheses. They also are referred to as unidirectional or bidirectional self-healing rings (USHR or BSHR).

Figure 3 shows a two-fiber unidirectional path-switched ring network. By convention, in a unidirectional ring the normal working traffic travels clockwise around the ring, as indicated by the heavy arrows. For example, the connection from node 1 to node 3 uses links 1 and 2, whereas the traffic from node 3 to node 1 traverses links 3 and 4. Thus, two communicating nodes use a specific bandwidth capacity around the entire perimeter of the ring. If nodes 1 and 3 exchange information at an OC-3 rate in an OC-12 ring, then they use one-fourth of the capacity around the ring on all the primary links. In a unidirectional ring the counter-clockwise path is used as an alternate path for protection against link or node failures. To achieve this, the signal from a transmitting node is dual-fed into both the primary and protection fibers. This establishes a designated protection path on which traffic flows counter-clockwise, i.e., from node 1 to node 3 via links 5 and 6 (in that order), as shown in Figure 3b. A heavy line and a dashed line indicate the primary and protection paths, respectively.

Consequently, two identical signals from a particular node arrive at their destination from opposite directions, usually with different delays, as denoted in Figure 3b. The receiver normally selects the signal from the primary path. However, it continuously compares the fidelity of each signal and chooses the alternate signal in case of severe degradation or loss of the primary signal. Thus, each path is individually switched based on the quality of the received signal. For example, if path 2 breaks or equipment in node 2 fails, then node 3 will switch to the protection channel to receive signals from node 1.

Figure 4 illustrates the architecture of a four-fiber bidirectional line-switched ring. Here two primary fiber loops (with fiber segments labeled 1p through 8p) are used for normal bidirectional communication, and the other two secondary fiber loops are standby links for protection purposes (with fiber segments labeled 1s through 8s). In contrast to the two-fiber UPSR, the four-fiber BLSR has a capacity advantage since it uses twice as much fiber cabling and because traffic between two nodes is only sent partially around the ring.

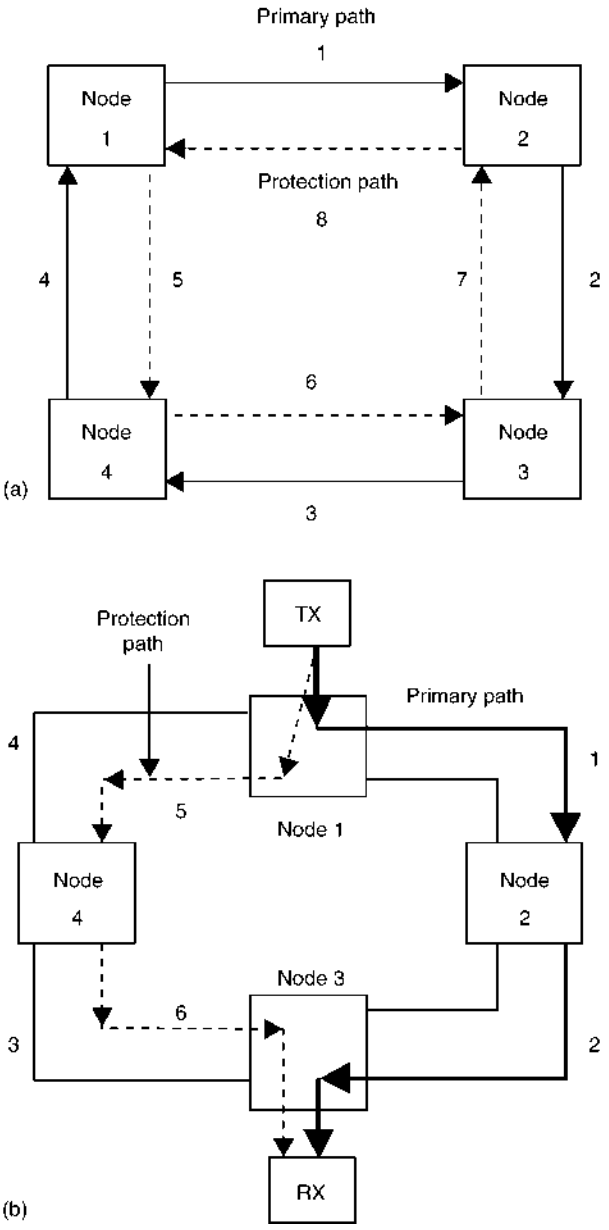

Figure 3 Schematic of a two-fiber unidirectional path-switched ring network.

To see the function and versatility of the standby links in the four-fiber BLSR, consider first the case where a transmitter or receiver circuit card used on the primary ring fails in either node 3 or 4. In this situation the affected nodes detect a loss-of-signal condition and switch both primary fibers connecting them to the secondary protection pair, as shown in Figure 5. The protection segment between nodes 3 and 4 now becomes part of the primary bidirectional loop. The exact same reconfiguration scenario will

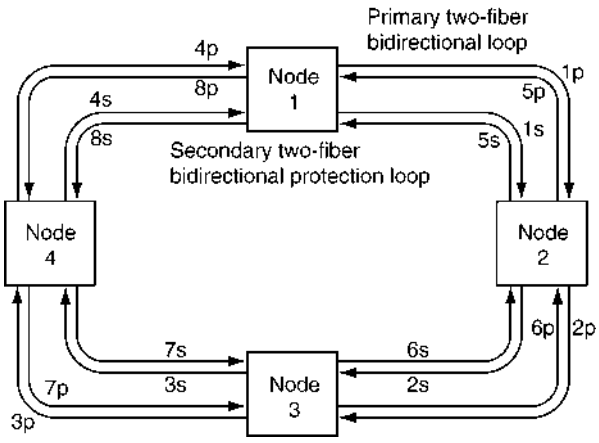

Figure 4 The architecture of a four-fiber bidirectional, line-switched ring.

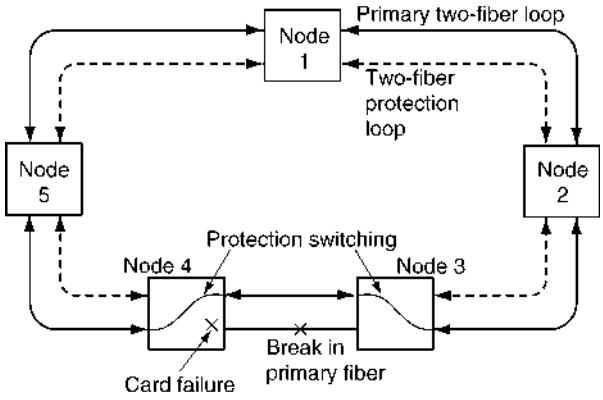

Figure 5 Reconfiguration of a four-fiber BLSR under transceiver failure.

occur when the primary fiber connecting nodes 3 and 4 breaks. Note that in either case the other links remain unaffected.

Now suppose an entire node fails, or both the primary and the protection fibers in a given span are severed, which could happen if they are in the same cable duct between two nodes. In this case the nodes on either side of the failed span internally switch the primary-path connections from their receivers and transmitters to the protection fibers, in order to loop traffic back to the previous node. This process again forms a closed ring, but now with all of the primary and protection fibers in use around the entire ring, as shown in Figure 6.

Single-Span WDM Links

An interesting and powerful aspect of an optical communication link is that many different wavelengths can be sent along a fiber simultaneously in the 1300–1600 nm spectrum. The technology of combining a number of wavelengths onto the same fiber is known as wavelength division multiplexing or WDM. Figure 7 shows the basic WDM concept. Here N independent optically formatted information streams, each transmitted at a different wavelength, are combined with an optical multiplexer and sent over the same fiber. Note that each of these streams could be at a different data rate. Each information stream maintains its individual data rate after being multiplexed with the other streams, and still operates at its unique wavelength. Conceptually, the WDM scheme is the same as frequency division multiplexing (FDM) used in microwave radio and satellite systems.

To see the potential of WDM, consider the characteristics of a high-quality optical source such as a distributed feedback (DFB) laser, which has a very narrow frequency spectrum on the order of 1 MHz, which is equivalent to a spectral linewidth of 10^{-5} nm. When using such a source, a guard band of 0.4–1.6 nm is typically employed. This is done to take into account possible drifts of the peak wavelength due to aging or temperature effects, and to give both the manufacturer and the user some leeway in specifying and choosing the precise peak emission wavelength. With such spectral band widths, simplex systems make use of only a very

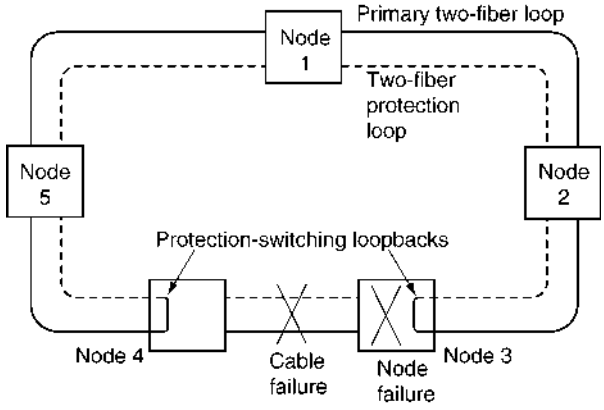

Figure 6 Reconfiguration of a four-fiber BLSR under node or fiber-cable failure.

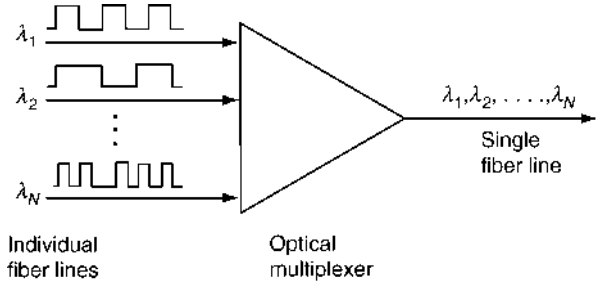

Figure 7 The basic WDM concept.

small portion of the transmission bandwidth capability of a fiber. This can be seen from Figure 2, which depicts the attenuation of light in a silica fiber as a function of wavelength. The curve shows that the two low-loss regions of a single-mode fiber extend over the wavelengths ranging from about 1260 to 1360 nm (the O-band) and from 1460 to 1625 nm (the S-, C-, and L-bands).

These regions can be viewed either in terms of spectral width (the wavelength band occupied by the light signal and its guard band) or by means of optical bandwidth (the frequency band occupied by the light signal). To find the optical bandwidth corresponding to a particular spectral width in these regions, consider the fundamental relationship $c = \lambda\nu$, which relates the wavelength λ to the carrier frequency ν, where c is the speed of light. Differentiating this yields $\Delta\nu = (c/\lambda^2)\Delta\lambda$, where the deviation in frequency $\Delta\nu$ corresponds to the wavelength deviation $\Delta\lambda$ around λ. Thus the optical bandwidth is $\Delta\nu = 14$ THz for a usable spectral band $\Delta\lambda = 80$ nm in the O-band. Similarly, $\Delta\nu = 15$ THz for a usable spectral band $\Delta\lambda = 120$ nm in the region spanning the S-, C-, and L-bands.

Since the spectral width of a high-quality source occupies only a narrow optical bandwidth, the two low-loss windows provide many additional operating regions. By using a number of light sources, each emitting at a different peak wavelength that is sufficiently spaced from its neighbor so as not to create interference, the fidelities of the independent messages from each source are maintained for subsequent conversion to electrical signals at the receiving end.

Since WDM is essentially frequency division multiplexing at optical carrier frequencies, the WDM standards developed by the ITU specify channel spacings in terms of frequency. A key reason for selecting a fixed frequency spacing, rather than a constant wavelength spacing, is that when locking a laser to a particular operating mode it is the frequency of the laser that is fixed. The ITU-T Recommendation G.692 specifies selecting the channels from a grid of frequencies referenced to 193.100 THz (1552.524 nm) and spacing them 100 GHz (0.8 nm at 1552 nm) apart. Suggested alternative spacings include 50 GHz and 200 GHz.

Historically the term dense WDM (DWDM) was used somewhat loosely and generally referred to the spacings denoted by Recommendation G.692. In 2002 the ITU-T released an updated standard aimed specifically at DWDM. This is Recommendation G.694.1, which specifies WDM operation in the S-, C-, and L-bands for high-quality, high-rate metro area network (MAN) and wide area network (WAN) services. It calls out for narrow frequency spacings of 100 to 12.5 GHz (or, equivalently, 0.8 to 0.1 nm at 1550 nm).

Alternatively, by using the larger spectral window available in low-water-content fibers, one can relax the spectral tolerances on the optical components and space the individual sources much further apart. This is the basis of coarse WDM (CWDM), which is useful for metropolitan area networks where variations in attenuation between different wavelength channels is not critical over short transmission distances. In 2002 the ITU-T released Recommendation G.694.2, which is aimed specifically at CWDM. The CWDM grid is made up of 18 wavelengths defined within the range 1270–1610 nm (O- through L-bands) spaced by 20 nm with wavelength-drift tolerances of ±2 nm.

A key feature of WDM is that the discrete wavelengths form an orthogonal set of carriers which can be separated, routed, and switched without interfering with each other. This holds as long as the total optical power intensity is kept sufficiently low to prevent nonlinear effects such as stimulated Brillouin scattering and four-wave mixing processes from degrading the link performance.

The implementation of WDM networks requires a variety of passive and/or active devices to combine, distribute, isolate, and amplify optical power at different wavelengths. Passive devices require no external control for their operation, so they are somewhat limited in their application in WDM networks. These components are mainly used to split and combine or tap off optical signals. The performance of active devices can be controlled electronically, thereby providing a large degree of network flexibility. Active WDM components include tunable optical filters, tunable sources, and optical amplifiers.

Passive Optical Networks

Another step towards realizing the full potential of optical fiber transmission capacity is the concept of all-optical WDM networks to extend the versatility of communication networks beyond architectures such as those provided by SONET. These networks can be classified as either broadcast-and-select or wavelength-routing networks. In general, broadcast-and-select techniques employing passive optical stars, buses, or wavelength routers are used for local network applications, whereas active optical components form the basis for constructing wide-area wavelength-routing networks. Broadcast-and-select networks can be categorized as single-hop or multi-hop networks. Single-hop refers to networks where

information transmitted in the form of light reaches its destination without being converted to an electrical form at any intermediate point. On the other hand, intermediate conversion can occur in a multihop network.

Figure 8 shows two alternate physical architectures for a WDM-based local network. Here N sets of transmitters and receivers are attached either to a star coupler or a passive bus. Each transmitter sends its information at a different fixed wavelength. All the transmissions from the various nodes are combined in a passive star coupler or coupled onto a bus and the result is sent out to all receivers.

Each receiver sees all wavelengths and uses a tunable filter to select the one wavelength addressed to it. In addition to point-to-point links, this configuration can also support multicast or broadcast services, where one transmitter sends the same information to several nodes. An interesting point is that these passive WDM networks are protocol transparent. This means that different sets of communicating nodes can use different information-exchange rules (protocols) without affecting the other nodes in the network. This is analogous to current time-division-multiplexed telephone lines in which voice, data, or facsimile services are sent in different time slots without interfering with each other.

Although the architectures of single-hop broadcast-and-select networks are fairly simple, there needs to be careful dynamic coordination between the nodes. For example, if the network is set up so that

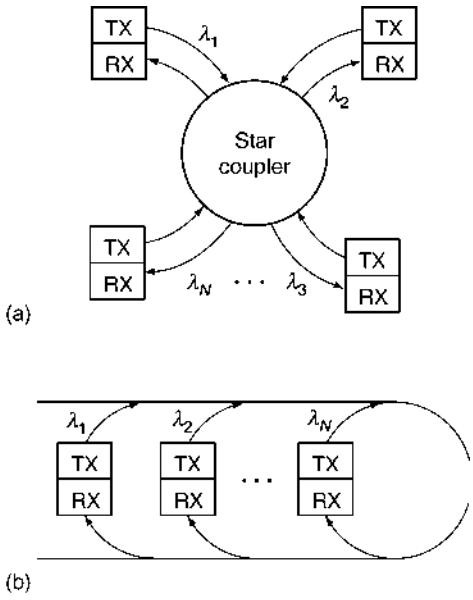

Figure 8 Two alternate physical architectures for a WDM-based local network.

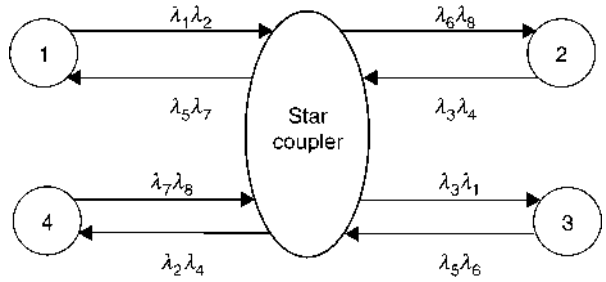

Figure 9 Example of a four-node broadcast-and-select multihop network.

a transmitter sends its information at a unique fixed wavelength, then the destined receiver needs to be informed when a message is being sent to it, so that it can tune its selective filter to that wavelength. Also conflicts need to be resolved in cases where two stations transmitting at different wavelengths want to send information to the same recipient simultaneously.

A drawback of single-hop networks is the need for rapidly tunable lasers or receiver optical filters. The designs of multihop networks avoid this need. Multihop networks generally do not have direct paths between each node pair. Each node has a small number of fixed-tuned optical transmitters and receivers. Figure 9 shows an example of a four-node broadcast-and-select multihop network where each node transmits on one set of two fixed wavelengths and receives on another set of two wavelengths. Stations can send information directly only to those nodes that have a receiver tuned to one of the two transmit wavelengths. Information destined for other nodes will have to be routed through intermediate stations.

The flow of traffic can be seen from Figure 9. If node 1 wants to send a message to node 2, it first transmits the message to node 3 using λ_1. Then node 3 forwards the message to node 2 using λ_6. In contrast to single-hop networks, with this scheme there are no destination conflicts or packet collisions in the network, since each wavelength channel is dedicated to a particular source–destination link. However, for H hops between nodes, there is a network throughput penalty of at least $1/H$.

Wavelength-Routed Optical Networks

Wavelength-routed networks overcome the limitations of passive optical networks through wavelength reuse, wavelength conversion, and optical switching. The physical topology of a wavelength-routed network consists of optical wavelength routers interconnected by pairs of point-to-point fiber links

in an arbitrary mesh configuration, as illustrated in Figure 10. Each link can carry a certain number of wavelengths, which can be directed independently to different output paths at a node. Each node may have logical connections with several other nodes in the network, where each connection uses a particular wavelength. Provided the paths taken by any two connections do not overlap, they can use the same wavelength. Thereby the number of wavelengths is greatly reduced. For example, the connection from node 1 to node 2 and from node 2 to node 3 can both be on λ_1, whereas the connection between nodes 4 and 5 requires a different wavelength (λ_2).

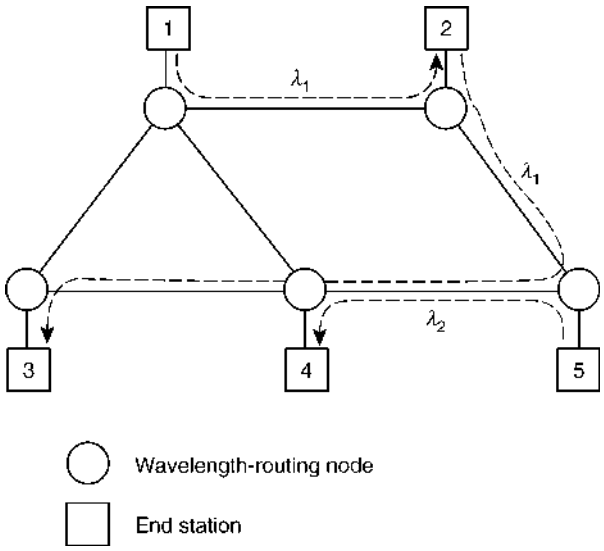

Figure 10 The physical topology of a wavelength-routed network.

A high degree of path modularity, capacity scaling, and flexibility in adding or dropping channels at a user site can be achieved by introducing the concept of an optical cross-connect architecture in the physical path structure (the so-called path layer) of an optical network. These optical cross-connects (OXCs) operate directly in the optical domain and can route very high-capacity WDM data streams over a network of interconnected optical paths.

Contentions could arise when channels having the same wavelength but traveling on different input fibers enter the OXC and need to be switched simultaneously to the same output fiber. This could be resolved by assigning a fixed wavelength to each optical path throughout the network, or by dropping one of the channels and retransmitting it at another wavelength. However, in the first case, wavelength reuse and network scalability (expandability) are reduced, and in the second case the add/drop flexibility of the OXC is lost. These blocking characteristics can be eliminated by using wavelength conversion at any output of the OXC.

As an example consider the 4×4 OXC shown in Figure 11. Here two input fibers are each carrying two wavelengths. Either wavelength can be switched to any of the four output ports. The OXC consists of three 2×2 switch elements. Suppose that λ_2 on input fiber 1 needs to be switched to output fiber 2 and that λ_1 on input fiber 2 needs to be switched to output fiber 1. This is achieved by having the first two switch elements set in the bar state (the straight-through configuration) and the third element set in the cross state, as indicated in Figure 11. Obviously without wavelength conversion there would be wavelength contention at both output ports.

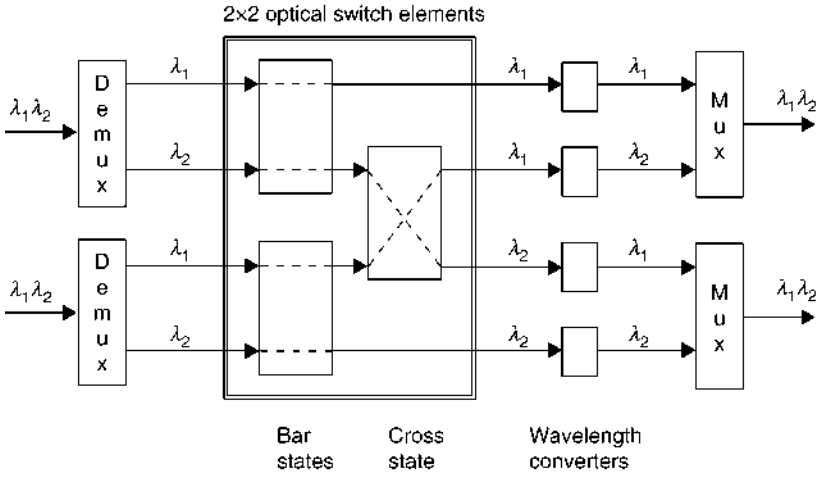

Figure 11 Example of a 4×4 optical cross-connect (OXC).

By using wavelength converters, the cross-connected channels can be converted to noncontending wavelengths.

See also

Lasers: Optical Fiber Lasers. **Optical Communication Systems:** Wavelength Division Multiplexing.

Further Reading

Blumenthal DJ, Prucnal PR and Sauer JR (1994) Photonic packet switches: architectures and experimental implementations. *Proceedings of the Institute of Electrical and Electronics Engineering* 82(11): 1650–1667.

Forouzan BA (2001) *Introduction to Data Communications and Networking*, 2nd edn. Burr Ridge, IL: McGraw-Hill.

Goralski WJ (2002) *SONET/SDH*, 3rd edn. New York: McGraw-Hill.

Gumaste A and Antony T (2003) *DWDM Network Designs and Engineering Solutions*. Upper Saddle River, NJ: Prentice Hall.

Hecht J (2002) *Understanding Fiber Optics*, 4th edn. Upper Saddle River, NJ: Prentice Hall.

Iannone E and Sabella R (1996) Optical path technologies: a comparison among different cross-connect architectures. *Journal of Lightwave Technology* 14(10): 2184–2196.

Keiser G (1999) A review of WDM technology and applications. *Optical Fiber Technology* 5(1): 3–39.

Keiser G (2000) *Optical Fiber Communications*, 3rd edn. Burr Ridge, IL: McGraw-Hill.

Keiser G (2002) *Local Area Networks*, 2nd edn. Burr Ridge, IL: McGraw-Hill.

Keiser G (2003) *Optical Communications Essentials*, New York: McGraw-Hill.

Ramaswami R and Sivarajan KN (2002) *Optical Networks*, 2nd edn. San Francisco: Morgan Kaufmann.

Special Issue (1995) *Undersea Communications Technology. AT&T Technical Journal* 74(1): 4–92.

Trischitta PR and Marra WC (1998) Applying WDM technology to undersea cable networks. *IEEE Communications Magazine* 36(2): 62–66.

Zang H, Jue JP, Sahasrabuddhe L, *et al.* (2001) Dynamic lightpath establishment in wavelength-routed WDM networks. *IEEE Communications Magazine* 39: 100–108.

Free Space Optical Communications

R Martini, Stevens Institute of Technology, Hoboken, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Free space optical (FSO) communication is the wireless transmission of data via a modulated optical beam directed through free space, without fiber optics or other optical systems guiding the light. The fundamental idea goes back to ancient times, as light (or smoke) signals were used to transmit information. From a more modern point of view, Graham Bell's patent on the photophone may mark the onset of modern FSO techniques, as it transmitted audio signals (i.e., voice) via the modulation of sunlight. A renaissance of FSO systems started with the availability of lasers, light sources with high output power and high coherence, which allowed the accurate direction of the light beams over long distances. During the 1970s and 1980s the main proposed application of FSO systems was for secure and long distance (50–1000 km) communication, mainly targeted for ground–satellite or satellite–satellite communication. This focus changed drastically over the last decade as a new market for FSO grew in the establishment of high bandwidth data link and their integration over a locally restricted area.

The main competitors in this market are the fiber-based optical network, the RF communication system, as well as the low bandwidth copper cable-based system. In comparison to the closely related wireless radio frequency transmission, the higher frequency of the optical carrier ($\sim 10^{14}$–10^{15} Hz) thereby allows for much higher transmission rates, comparable to those of typical fiber optic networks. On the other hand, the use of an optical carrier also results in much more directed beam propagation, which requires an undisturbed line-of-sight between emitter and detector. This restricts the application of most FSO systems to a range between a few hundred meters up to several kilometers, which are still favorable for distribution of high bandwidth networks over a locally restricted area. This makes FSO a highly attractive candidate for the 'last-mile' distribution of high-bandwidth Ethernet to the individual homes.

The simplicity of setup of FSO links, as well as their modularity, is thereby their biggest advantage, compared to fiber-based networks. It makes them not only highly cost efficient, but it eases the maintenance or

allows for fast and easy upgrades, as it does not require any extensive and time-consuming installation – in contrast to the installation of fiber optic cables. This advantage makes FSO highly attractive for temporary installations (emergency or short-time high broadcast situations, such as the Olympic Games), as well as to overcome geometrical restraints (river, seas, etc.). As most detector and emitter systems are typically based on the same electro-optical components used for fiber optic networks, they seamlessly integrate and expand an existing network without complications in data handling.

Standard FSO System

Figure 1 shows schematically the setup for a typical FSO system connecting two separate local networks. It consists of: an electronic data input; a small but powerful light source which can be modulated; emitter optics which shape the emitted beam to a highly directed beam; the atmosphere as transmitting device; detector optics which receive the transmitted light and focus it onto a photodetector; and an electronic amplifier which serves as data output.

Compared to a fiber optical communication system, the main difference is the use of the atmosphere as the transmission medium, in contrast to guided optical propagation. The transmission properties are thereby determined by the atmospheric conditions, which can lead to degradation and/or redirection of the beam. The main influences which have to be taken into account are atmospheric scattering, molecular absorption, turbulence effects (including beam wander and fading), as well as the impact of low-visibility weather situations like rain, snow, or fog. These depend strongly on wavelength and linewidth of the used emitter system, distance between emitter and receiver, as well as the environmental conditions. As a consequence, the characteristics of the atmospheric channel determine the layout of the other FSO components – and have to be considered first.

Atmospheric Losses

The overall losses of the atmospheric transmission are determined by the Beer–Lambert absorption law:

$$I = I_0 \exp[-\alpha L] \qquad [1]$$

where I is the intensity at the receiver, I_0 the intensity of the transmission beam after the emitter optics, and L is the range (i.e., distance between emitter and detector optics). The total extinction coefficient $\alpha = \alpha_{\text{Rayleigh}} + \alpha_{\text{Mie}} + \alpha_{\text{Absorption}}$ is based on three different contributions: Rayleigh scattering, Mie scattering, and absorption.

In contrast to absorption, the losses due to scattering originate from redirection of light and will be discussed first. Depending on the size of the scattering partner, two separate regimes have to be taken into account. For scatter smaller than the wavelength, the effect can be described based on Rayleigh's classical formula. The scattering cross-section of an element with mass m, charge q, natural oscillation wavelength λ_0, is then:

$$\sigma_s = \frac{f q^4 \lambda_0^4}{6\pi\varepsilon_0^2 m^2 c^4} \frac{1}{\lambda^4} \qquad [2]$$

where ε_0 is the dielectric constant, c the speed of light, and λ the wavelength of the incident light, leading to an extinction coefficient of $\alpha_{\text{Rayleigh}} = \sigma_s N_s$, where N_s is the concentration of the scattering elements. The λ^{-4} dependence in the scattering predicts stronger scattering for shorter wavelengths and Rayleigh scattering is therefore dominant for short and visible wavelengths (which gives the sky its blue color). For typical FSO

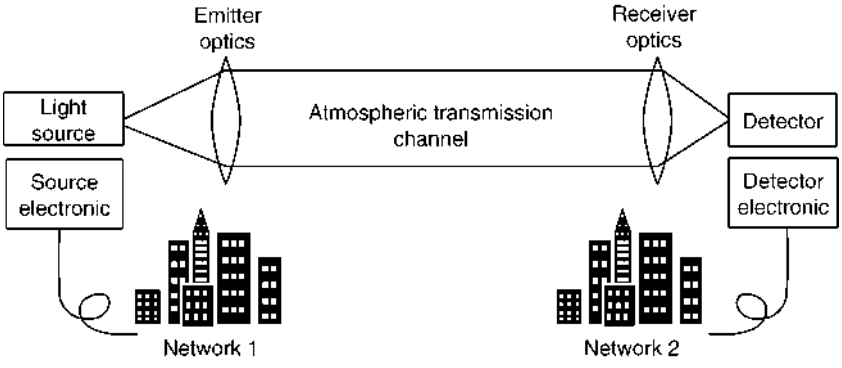

Figure 1 Typical components of a FSO system.

wavelengths in the region of near-infrared radiation (NIR, 800–2000 nm) or mid-infrared radiation (MIR, 2000–20 000 nm), the contribution of Rayleigh scattering is typically relatively low and can be neglected.

On the other hand, Mie scattering is, in most cases, the strongest contribution for FSO systems – but also the hardest to predict. It is due to resonant scattering effects, where the particle size is of the wavelength of the scattered light. For the typical FSO wavelengths these particles include water and oil droplets, dust, smog, fly ash, and aerosols (which can originate from pollution). A correct calculation of the Mie scattering coefficient is possible, but requires knowledge of the exact particle concentration as well as their size distribution. As these numbers are not only difficult to measure, but also fluctuate continuously depending on weather situation, time of day, season, temperature, climate, pollution, and other factors, a numerically correct evaluation is not possible.

Historically, most systems rely on an empirical formula, which estimates the attenuation coefficient to

$$\sigma_{\text{Mie}} = \frac{3.91}{V}\left(\frac{\lambda}{550}\right)^{-p} \qquad [3]$$

where V denotes the visibility and the value of p ranges from 1.6 for high visibility ($V > 50$ km) and 1.3 for average visibility (50 km $> V >$ 6 km) down to 0.585 $V_{1/3}$, respectively for low visibility. The formula is based on data collected in the 1950s and 1960s. More accurate measurements show that the actual scattering cross-sections differ quite strongly, especially for low-visibility cases. Especially, the predicted dependence on wavelength is not correct. The failure of such a general model is not surprising, as the actual scattering properties are strongly dependent on the weather situation and the environmental conditions. It is the focus of ongoing research to find more accurate models describing the scattering condition in multiple environments and conditions to provide the background for more accurate empirical models.

Apart from the light redirection due to scattering, absorption either by molecules or aerosols is nearly as important for the transmission calculation of a FSO system. The strongest molecular absorptions are typically due to water, carbon dioxide, as well as ozone. A typical absorption spectrum can be easily calculated using standard software together with the publicly available HITRAN database, both available from ONTAR. The upper part of Figure 2 shows a low-resolution transmission spectrum for the interesting wavelength region for a path length of 1 km.

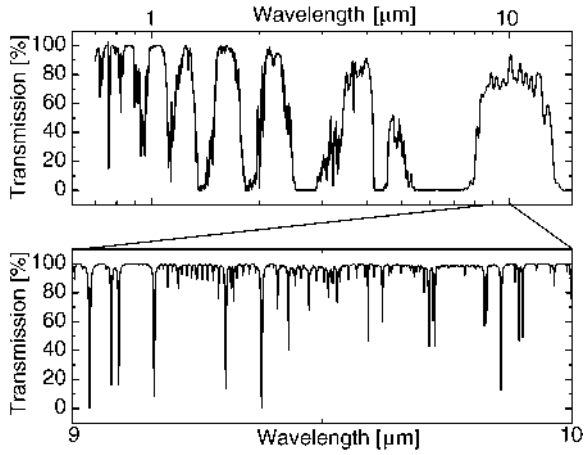

Figure 2 Upper: Low-resolution transmission spectrum of the atmosphere showing molecular absorption for a path length of 1 km. Lower: high-resolution spectrum of the atmospheric windows around 9.5 μm.

Within this spectrum, several well-known atmospheric windows exist, which are naturally preferred for FSO applications. These include the region around 830 nm, 1550 nm, and from 8 μm to 12 μm. For the exact selection of a wavelength, a closer look is needed. As an example, a high-resolution analysis of the atmospheric windows around 9.5 μm is shown in the lower part of Figure 2, which identifies additional absorption lines as well as highly transparent wavelength ranges within the window. Such substructure has to be taken carefully into account for the FSO wavelength selection, especially if lasers are employed, which have a linewidth comparable to the width of the absorption line.

Besides the molecules, aerosol particles also contribute to the absorption spectrum. Calculation of their contribution can also be performed using available software package, such as Air Force's MODTRAN and LOWTRAN program as well as OPAC from the LMU University in Munich. Figure 3 shows a typical low-resolution transmission spectrum for 50% humidity as well as 99% humidity, where by scattering as well as absorption effects have both been included in the calculation. The occurrence of sharp absorption features, as well as the increasing losses for wavelengths above 10 μm, is typically due to the aerosol absorption, whereas the scattering dominates the function for wavelengths below 3 μm. Losses due to the aforementioned Rayleigh scattering are also shown in the same diagram, demonstrating its low influence at longer wavelengths.

MODTRAN and LOWTRAN also allow the calculation of transmission spectra under rain and snow conditions, where rain and snow consist of

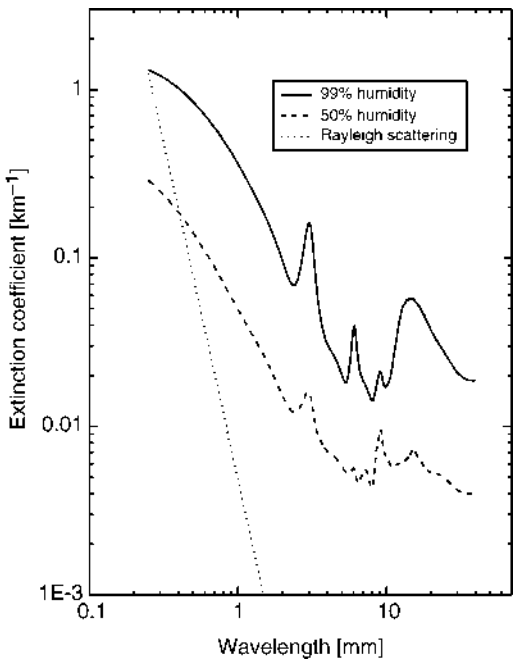

Figure 3 Typical low-resolution transmission spectrum calculated based on absorption and scattering losses from aerosols for 50% humidity (dashed line) and 99% humidity (solid line). The dotted line shows additionally the losses due to Rayleigh scattering.

clusters of water and aerosol with diameters from 200 μm to 2 mm and above. Depending on the severity of the precipitation, the losses can range from ~0.5 dB/km for light rain up to 13 dB/km for cloudbursts, requiring an additional power margin in the preparation of the FSO link. Significantly stronger is the impact of fog, as typically particles with much lower sizes are present. The losses inflicted by a moderate fog (visibility 500 m) range above 20 dB/km, surpassing the effect of a cloudburst; for thick fog the attenuation can even grow beyond 200 dB/km. However, depending on the size distribution within the fog itself, the losses can be highly wavelength dependent. Recent experiments show that long wavelength FSO systems operating in the MIR wavelength region have a much stronger stability under the influence of thick fog than one operating at NIR wavelengths.

Turbulence and Scintillation

In contrast to a static medium, the atmosphere is a dynamic mixture which undergoes continuous changes due to wind, temperature, and other influences. This creates a constant change in the parts of the atmosphere which are contributing to the FSO channel and therefore in the optical properties of the beam path. The index of refraction can differ, not only in its total value, but also within subsections of it, creating a random fluctuation in the beam paths. Such localized changes in the refractive index of the air are schematically shown in Figure 4, where turbulences of different sizes modify the beam path and disturb the beam front. The turbulence in the atmosphere can be quantified by the refractive-turbulence structure coefficient C_n^2, which can range from $10^{-13}\,\mathrm{m}^{-2/3}$ in the middle of the afternoon (highest value) to $10^{-17}\,\mathrm{m}^{-2/3}$ one hour after sunrise or sunset. As the turbulence effect is typically due to strong temperature difference, C_n^2 is strongest near hot surfaces (e.g., tar roofs or asphalt in the summer) and decreases with altitude. The effects of such turbulences on a directed light beam (such as used in FSO systems) include mainly beam wander and scintillation effects.

Beam wander is thereby based on large size turbulences which act as a lens on the propagating beam. Consequently, a narrow focused beam could be randomly steered away from the detector area, creating a fluctuation in the detected intensity.

Stronger impact for FSO links has scintillation, which is better known as the twinkling of a star or the heat shimmer of the horizon on a hot day. It is based on the continuous variation of the beam front which can lead to intensity fluctuations (like the twinkling) and loss in the image resolution (heat shimmer). Strong intensity fluctuations can thereby cause signal loss as well as receiver saturation, both resulting in downtime in the link. Such local intensity disturbances can occur up to sizes of $\sqrt{\lambda L}$, reaching, for example, 4 cm for a 1.55 μm link after 1 km. For small fluctuations, the intensity follows a classical stochastic behavior and can be described by the irradiance variation (normalized to its mean value) as

$$\sigma^2 = \exp\left[0.5\left(\frac{2\pi}{\lambda}\right)^{6/7} L^{11/6} C_n^2\right] - 1 \qquad [4]$$

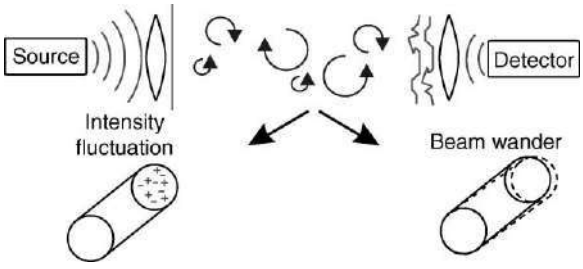

Figure 4 Schematic depiction of atmospheric turbulences with different sizes in FSO systems. Shown below are illustrations of its main effects on a directed beam causing intensity fluctuations and beam wander.

with usual values in the range from 5 to 70%. Nonetheless, it should be pointed out that actual turbulence effects can differ quite strongly from these predictions, due to local inhomogeneities in temperature distribution. Consequently, test measurements are necessary for the correct evaluation of a specific installation site.

What makes scintillation effects difficult to handle for communications systems is the relatively slow frequency of fluctuations in the order of a few hundred Hz or less. The corresponding downtimes for a link can range from 1 ms to 10 ms, or even up to seconds. This is magnitudes higher than typical downtimes in standard communications links and cannot be compensated by standard forward error correction systems. One possibility to reduce the influence of scintillation is the method of aperture averaging, whereby multiple beams originating from slightly different positions are deployed and targeted at the same receiver. Assuming that the variation in each of the separate beam paths is random and uncorrelated, the signal-to-noise ratio will decrease by the square root of the number of beams. Recent experiments have impressively demonstrated this effect and allow reduction of intensity fluctuation to an acceptable level.

Another alternative to decreasing the scintillation effects is in the application of longer wavelength FSO systems. Whereas beam wander is relatively independent from the wavelength, scintillations effects scale with $\lambda^{7/6}$, as turbulences smaller than the wavelength cannot obscure its beam front. Using a 2 times longer wavelength, for example, should have the same impact as signal averaging with 4 different beams.

Beam Spreading and Pointing

Based on optical diffraction theory, the beam leaving the emitter optics will be divergent and its beam diameter will consistently increase during transmission through the optical channel. Assuming an optimal alignment and using standard geometrical optics, the losses due to beam spreading can estimate the ratio of emitted power (P_E) to received power (P_R):

$$R_{BS} = \frac{P_R}{P_E} = \frac{D_R}{(D_E + L\theta)} \qquad [5]$$

where D_R is the diameter of the receiving optics, D_E the diameter of the emitting optics, and θ the divergence of the beam. The minimal divergence of the beam is typically determined by the light source and is given by $\theta_L = \lambda/D_L M^2$, where D_L is the initial size of the laser beam and M^2 the laser-mode structure parameter, for a typical semiconductor laser θ_L below 0.1 mrad. In many cases, the divergence of the emitted beam is commonly enlarged by slight defocusing of the emitting optics – if the power budget allows for the additional losses. The beam size at the receiver is typically several times larger than the receiving aperture, which not only eases the alignment of the system, but also reduces pointing losses. This includes the effect of beam wander due to turbulence as well as swaying of the buildings or the mount. If those effects are solely compensated for by an increased beam spread, divergence can easily reach values of 2–20 mrad for short-distance links. Obviously this is not an option for longer-distance FSO links, as the losses will increase dramatically. In these cases, active tracking systems are incorporated in the emitter optics allowing for continuously adjustment to achieve and keep an optimal alignment. Typical tracking devices are thereby based on servo-driven Gimbal mounts, steering mirrors, or recently, MEMS devices.

Laser Eye Safety

As the atmospheric transmission channel of most FSO systems is not completely isolated, there exists the possibility of a human person intercepting the beam. To allow for this, most FSO systems are designed to be eye-safe following the voluntary ANSI standard, which limits the output irradiance to a certain value depending on the deployed wavelength. As radiation below 1400 nm is still being focused on the retina, it causes a higher irradiance, which results in a lower damage threshold. For 800 nm laser systems, the minimum permissible exposure (MPE) level for a 10 s long exposure is therefore recommended not to exceed 1 mW cm^{-2}, whereas for 1.55 μm or longer wavelength systems, an MPE of 100 mW cm^{-2} is suggested. The recommendations for incoherent light sources such as LEDs are thereby less stringent as they cannot be focused quite as strongly as laser light.

The Light Source

Following the discussion above, the optimal light source for FSO system should be emitting at a wavelength away from any absorption line (or should be much broader then the absorption line), have a relatively long wavelength as well as a relative low divergence, be eye-safe, and allow for high modulation speed. In reality, most FSO systems use either GaAsAl laser diodes or LEDs emitting at 0.8 μm, or the typically more expensive 1.5 μm laser diodes

based on the InGaAsP semiconductor system. The output power of both laser systems is typically in the range of 1 mW to 100 mW, which for the longer wavelength case can be boosted to 1–10 W using erbium-doped fiber amplifiers (or even higher for specific applications). The longer wavelength system has the additional advantage of higher ANSI standards as discussed above, allowing stronger eye-safe emission power, but is typically far more expensive.

Both laser types were developed and optimized for standard fiber-based communication systems allowing direct modulation up to 10 Gbit/s. To achieve higher bandwidth in FSO systems, standard techniques known from fiber-based networks can easily be deployed, like wavelength division multiplexing which increases the bandwidth up to state-of-the-art levels of fiber-based networks.

Contrary to semiconductor lasers, LEDs provide a much less expensive alternative, but allow only for modulation up to 10–100 MHz. Their broad emission spectrum is thereby an advantage, as it is less sensitive to sharp absorption lines which – as discussed earlier – can have a drastic impact on the FSO link performance. Consequently, LEDs have been successfully implemented in several commercial FSO systems targeted at the short-range market.

Recently developed sources, which have also been applied for FSO systems, include vertical-cavity surface-emitting lasers as well as quantum cascade lasers (QCL) which emit in the mid-infrared regime. The latter devices have been predicted to have a strong impact on the FSO market, as they combine high-output power (up to 1 W) together with fundamentally high modulation bandwidth (2.5 Gbit/s recently demonstrated and higher rates predicted) at wavelengths ranging from 3.5 μm to 12 μm and beyond. They can access the long-wavelength atmospheric window atmospheric for the FSO systems (typically reserved for CO_2 lasers) with the convenience of direct semiconductor lasers like direct electrical modulation, small size, and low power consumption. Recently, first FSO experiments, employing experimental QCLs, demonstrated the predicted higher stability of the link in low-visibility situations, including strong fog as well as high dust concentration. Nonetheless, to date the high cooling requirement for QCLs restricts more general use of them, but their very fast development, recent reports of cw-operation of room-temperature QCLs as well as FSO links with TE-cooled QCLs in pulse operation, give hope for a solution to this problem.

The Detector System

Once the light reaches the receiver, it is typically collimated using a short focal length telescope or a single lens system. To optimize efficiency without sacrificing ease of alignment a compromise has to be made between a wide aperture for greater light collection and a short focal length. The concentrated light then hits the detector, which, depending on the wavelength, is either Si or InGaAs-photodiodes. Si-based devices can detect up to wavelengths of 1100 nm and InGaAs systems are typically used for the 1.5 μm region, respectively. To enhance the sensitivity of the detectors, avalanche photodiodes are preferred which internally enhance the detected signal by a factor ranging from 20 to 100. To ensure high bandwidth operation (10 MHz–10 GHz), the size of the detecting device has to be kept relatively small (20 to 100 μm) avoiding parasitical capacitance effects – but raising the requirement on the detecting optics.

The sensitivity of these NIR detectors is typically limited by the shot-noise, which is due to statistical fluctuations of the received phonons and scales with the square root of its number. For practical purposes the internal noise of the detector is quantified using the corresponding amount of incident optical power leading to a comparable signal. The values for this noise equivalent power (NEP) can range from microwatts to tens of nanowatts. Other sources for detector noise can include a background current from the detector–amplifier combination, typically referred to as Johnson noise.

Mid-infrared detectors are additionally objected to strong background radiation due to thermal emission, which peaks around the 10 μm at temperature of 300 K. The use of additional cold filters and high-resolution spectral filters can help to overcome this restriction, but still the sensitivity of those detectors is lacking compared to NIR systems.

For communication system the noise is more commonly quantified as bit error rate (BER), quantifying the average percentage of wrongly received bits (submitted '1', detected '0' or vice versa). A BER greater than 10^{-9} is usually referred to as an error-free system, as the residual errors can be corrected employing forward error correcting algorithms. In contrast, the performance of FSO systems can more easily be measured in its signal-to-noise ratio (SNR). Following standard communication theory, a peak SNR of 12 and an average SNR of 6 is required to achieve a BER of 10^{-9}, guaranteeing error-free communication. Consequently range limits for FSO links are typically based on SNR = 6.

Link Budget and FSO Range Equation

Classical link budget estimation is a commonly used tool to access the performance of FSO systems. To take into account the losses due to atmospheric absorption, scattering, turbulence, and beam spreading, as well as optical losses in the used receiver and emitter optics, they are deducted from the starting power emitted from the FSO light source. The estimated intensity at the detector is then compared with the intensity needed to establish error-free communication (for example SNR = 6), the calculated difference is the link margin of the specific system, i.e., the amount of additional losses under which the FSO link will still be working.

Comparing this number to losses from nonpredictable fluctuations (like heavy rain or fog) gives an insight under which implications the link will fail. Using weather station data, the frequency of such extremely low-visibility occurrences can be obtained and a number for the reliability of the FSO link be estimated. Typical well-installed FSO systems estimate very high reliabilities, between 99.9% and 99.999%, which have been obtained in installation and verified in long-term measurements.

Another way to estimate the performance for a given system is based on the FSO range equation, where the received power on the detector (P_R) is calculated to

$$P_R = P_E \frac{D_R}{(D_E + L\theta)^2} TK \exp[-\alpha L] \qquad [6]$$

T denotes the optical losses of the transmission and receiver system, and K denotes additional losses for incoherent lights sources and is given by the ratio of LED emitting area to detector area, and is $K = 1$ for laser light sources. This equation can also be used to estimate losses for different weather situations and obtain the maximum range L of a specific FSO system.

Recent Research and Future Considerations

Existing FSO systems have impressively demonstrated their versatility and stability, as well as their seamless integration into existing networks, on multiple occasions. A prime example is the re-establishment of communication links between Merrill–Lynch Brokerage and Wall Street in New York, in the aftermath of the terrorist attack on September 11th, which destroyed one of the central fiber optic hubs in downtown Manhattan. The installation of a commercial FSO system by Terabeam ensured a fast replacement and a stable continuation of their work.

The demonstrated performance of such links compares to standard fiber-based links. The inclusion of wavelength division multiplexing as well as other standard or nonstandard techniques in FSO systems proves the strengths of a system, which is based on identical hardware for emitter and detectors as fiber-based networks. Nonetheless, FSO allows a multitude of applications not possible in fibers, which current research projects are targeted at. This includes, for example, long wavelength FSO communication systems, continuously ground-satellite links and free-space quantum communication for extreme secure data transmission – to list only a few of the possibilities. One very obvious example is extreme wavelength division multiplexing, which uses one link operating in all atmospheric windows at the same time. Such a FSO system would not only allow very high data rates and extreme stability against weather effects, but could also employ new kinds of wavelength-based modulation and encoding schemes increasing the security of the data link. Concluding the actual state of the FSO system as well as ongoing research, free space optical communication has all the ingredients needed to be one of the mayor key ingredients of future networks – and based on the increasing demand of bandwidths they will contribute strongly.

List of Units and Nomenclature

c	speed of light
C_n	refractive-turbulence structure coefficient
D_E	diameter of the emitting optics
D_R	diameter of the receiving optics
I	intensity
L	range (i.e., distance between emitter and detector optics)
m	mass
M^2	laser-mode structure parameter
N_s	concentration
P_E	emitted power
P_R	received power
q	charge
T	optical losses of the transmission and receiver system
V	visibility
α	extinction coefficient
ε_0	dielectric constant
θ	divergence of the light beam
λ	wavelength
λ_0	natural oscillation wavelength
σ_s	scattering cross-section

See also

Environmental Measurements: Doppler Lidar; Optical Transmission and Scatter of the Atmosphere. **Imaging:** Infrared Imaging; Lidar.

Further Reading

Andrews LC and Phillips RL (1998) *Laser Beam Propagation through Random Media*. Bellingham, CA: SPIE Optical Engineering Press.

Begley DL (ed.) (1991) *Selected Papers on Free-Space Laser Communications*. SPIE Milestone Series. Bellingham, CA: SPIE Optical Engineering Press.

IEE Proceedings Optoelectronics, *Special Issue on Free Space Optical Communications*. December 1996.

Jha AR (2000) *Infrared Technologies*. New York: John Wiley & Sons.

Killinger D (2002) *Free Space Optics for Laser Communications Through the Air*. Washington: Optics and Photonics News, Optical Society of America, p. 36.

Measures R (1984) *Laser Remote Sensing*. New York: John Wiley & Sons.

Mecherle GS (ed.) (2003) *Free-Space Laser Communication Technologies XV*. Bellingham, CA: SPIE Optical Engineering Press.

Weichl H (1990) *Laser Beam Propagation in the Atmosphere*. Bellingham, CA: SPIE Optical Engineering Press.

Willebrand H and Ghuman BS (2001) *Free-Space Optics: Enabling Optical Connectivity in Today's Networks*. Indianapolis, IN: SAMS.

Wolfe WL and Zissis GJ (1978) *The Infrared Handbook*. Ann Arbor: Enviromental Research Institute of Michigan.

Zuev VE (1988) *Laser Beams in the Atmosphere*. New York: Consultant Bureau.

Lightwave Transmitters

J G McInerney, National University of Ireland-Cork, Cork, Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical fiber telecommunications has changed human society forever, providing the capacity for affordable and ubiquitous communication. It provides data rates and transport economies far in excess of those available using purely electronic means. It is doubtful whether the Internet and worldwide web, or the vast infrastructure of wireless cellular telephony, would be viable without optical fiber technology. Its rapid development and acceptance has been driven largely by contemporaneous successes in manufacturable low-loss optical fiber cables, sensitive pin-FET photoreceivers, and reliable high-performance laser diode-based transmitters. The earliest developments in silica glass fibers at Standard Telecommunication Laboratories, UK, (then part of the US-based ITT Corporation) and Corning Glass Works (USA) in the 1960s and 1970s, were pivotal in determining the directions of modern fiber technology, as were later key inventions of rare-earth doped fiber amplifiers at the University of Southampton (UK) and AT&T Bell Laboratories (USA) in the 1980s and early 1990s. However, perhaps the greatest contributors to the success of fiber optics have been semiconductor injection lasers, or laser diodes, which in various forms have made up the vast majority of all fiber optic transmitters, and the entirety of long-haul transmitters, as well as pump sources for the doped-fiber amplifiers which enable modern networks. The fascinating story of the development of fiber optics is to a large extent driven by the dramatic progress in the development of high-performance, reliable semiconductor lasers. These emitted at wavelengths near 1300 and 1550 nm, which respectively correspond to the dispersion and attenuation minima of single-mode silica optical fibers.

Semiconductor Laser Principles

Semiconductor lasers were first demonstrated in research laboratories at General Electric, IBM Corporation and the MIT Lincoln Laboratory (all USA) as early as 1962, although it took almost two decades for the basic science and engineering of materials, fabrication, reliability, and performance design issues to be developed sufficiently for their use in practical communication systems. Early successes were obtained using the GaAs/AlGaAs lattice-matched material system, in which a lightly doped GaAs or ternary compound AlGaAs active layer is sandwiched between two lattice-matched heterojunctions with n- and p-doped AlGaAs layers with higher Al fractions than the active layer, so that their refractive indices are lower and their bandgaps higher, providing optical and charge-carrier confinement. In all cases the material is epitaxially grown, meaning that the entire laser chip forms a single crystal. The gain, and hence the laser emission, occurs at the direct bandgap of the active layer, at wavelengths in the range 750–870 nm depending on the Al fraction.

Later evolutions used the quaternary compound GaInAsP clad by InP guiding layers, where the In and P fractions are chosen to match the lattice constant to the InP substrate.

In general, optical gain in a laser diode is generated by creating an electron-hole plasma in the vicinity of the forward-biased junction, a situation corresponding to population inversion in a conventional laser, and with sufficient forward current the material becomes transparent, in that the gain exactly equals the absorption and scattering losses, at a particular wavelength. When this arrangement is enclosed within a suitable optical cavity, and the forward current is increased further, the system begins to lase, that is to oscillate continuously to produce coherent optical radiation at a wavelength which satisfies two basic conditions: it must be within the set of wavelengths at which the material produces sufficient gain, and it must be close to one of the electromagnetic modes at which the optical cavity is resonant. These conditions, which define lasing threshold, are characterized by the equality of the optical gain and the total losses, that is the material losses (absorption, scattering, and free-carrier plasma) and the cavity losses which include the light output.

Early laser diodes in both the GaAs/AlGaAs and GaInAsP/InP systems were of the edge-emitting type, in which the light is emitted from the edge of the laser chip, perpendicular to the growth direction in the plane of the wafer. Although this geometry complicates laser production flow, particularly the testing function which requires that the wafer be scribed into bars to define the laser output facets, edge-emitting lasers (EELs) make up the vast majority of optical fiber transmission sources. An alternative laser diode structure is the vertical cavity surface-emitting laser (VCSEL), in which emission occurs in the growth direction perpendicular to the plane of the wafer. VCSELs offer major advantages in spectral stability, beam quality, manufacturability and cost, but their output powers are low (a few mW) and they are not yet available at the key telecommunication windows around 1300 nm and 1550 nm, due to difficulty in forming the necessary high-reflectivity Bragg mirrors in the GaInAsP/InP material system. Major initiatives are currently underway to produce long-wave VCSELs in the GaAs/AlGaAs system using InAs quantum dots, or dilute nitrides such as InGaAsN, or hybrid approaches such as InP-based gain media fused to high-reflectivity mirrors using GaAs/AlGaAs, dielectric coatings, air gaps, and others. At present, however, all long-haul optical fiber transmission is based on edge-emitting laser diodes fabricated in GaInAsP/InP.

General Structure and Requirements of Optical Fiber Communication Systems

The first optical fiber communication links were simple point-to-point affairs, consisting essentially of series-connected transmitters, fibers, and receivers. The transmitters were simple Fabry–Perot laser diodes, onto which data were encoded by direct digital modulation of the injection current at rates ~100 Mbit/s. The first commercial optical fiber link was built in 1976, a single fiber cable linking two switching centers of the Illinois Bell Telephone Co just 2.5 km apart in the Chicago metropolitan area, using 850 nm GaAs laser technology. In 1988, the first transatlantic fiber link, AT&T's TAT-8, was completed linking endpoints in New Jersey, England, and France with three fiber pairs carrying signals generated by 1300 nm GaInAsp single mode lasers. In these early systems, data were regenerated frequently by repeaters each consisting of a receiver, signal conditioner and transmitter. Each fiber in TAT-8 carried a single optical frequency modulated at 280 Mbit/s and the construction cost of the system was ~$50 k/km, resulting in an economic figure of merit of over $600 k/Mbit/s. Twenty-five years later, current systems cost the same per kilometer to build in absolute dollars despite inflation, a major effective cost reduction enabled largely by replacement of the expensive repeaters by doped-fiber optical amplifiers. Moreover, each fiber is now highly multiplexed, with the potential for hundreds of wavelengths each carrying 10 Gbit/s signals generated by 1550 nm laser transmitters, so that the unit cost has plummeted to ~$200/Mbit/s, a reduction of three and a half orders of magnitude.

In the remainder of this article, we will describe transmitters for both short- and long-haul systems where the channel spacing is only tens of GHz in the optical carrier frequency. Modern short-haul systems either feed client networks as part of bi-directional transceivers, or drive metro and local access networks connected directly by photonic switches to terminals of the long-haul systems.

DWDM Transceivers

Modern practice uses modular units containing transmitters and receivers allowing bi-directional data flow. In the send mode, data are applied to a long-haul transmitter at the local (client) source and coupled through the external fiber network to the remote long-haul receiver. In the receive mode, data arrive from a remote long-haul transmitter and are coupled (via a regenerator including signal

conditioning, clock recovery, and error correcting steps) to a local short-haul transmitter leading to the client's internal network. Each transceiver card therefore contains a long-haul transmitter and receiver set (known collectively as dense wavelength division multiplexing (DWDM) transport devices) and a short-haul transmitter and receiver set (known as client interfaces or local transport devices).

Transmitter Requirements

Short-haul transmitters may be simple Fabry–Perot edge-emitting laser diodes, or vertical cavity laser diodes, or even superluminescent LEDs for sub-Gbit/s data rates. They are generally not required to be at one of the minimum-loss telecom windows near 1550 nm, as the transmission distances are short. However, as client networks grow into large metropolitan networks, it is likely that 1300 nm sources will be used to minimize dispersion. Power requirements are modest, a few mW peak, at data rates ~1–10 Gbit/s. Single mode laser emission and polarization control are not required by present day systems, although polarization-mode dispersion may limit transmission at higher data rates. Both return-to-zero (RZ) and nonreturn-to-zero (NRZ) coding may be used, and direct modulation of the laser amplitude is the norm.

For long-haul transmitters the situation is dramatically different. Operation in one of the standard erbium-doped fiber amplifier (EDFA) bands near 1550 nm is mandatory: the C (conventional) band extends from 1530–1565 nm, the S (short-wave) band from 1460–1530 nm) and the L (long-wave) band from 1565–1625 nm. In the C-band, output powers ~10 mW are generally sufficient, and powers above ~30 mW are limited by nonlinear effects such as self-phase modulation and four-wave mixing in the fiber. In the S- and L-bands, where the available gain from EDFAs is less, lasers may need to operate close to the nonlinear limit, several tens of mW, although various schemes (differential pumping, gain-flattening filters, etc.) are being evaluated to flatten the global EDFA gain spectrum.

In current practice, a single fiber utilizing DWDM, using only the C-band, can carry up to 1 Tbit/s comprising 100 channels at 10 Gbit/s on each channel, the current standard. For such performance a channel spacing of 50 GHz in optical frequency is required. Using all three bands with similar density would enable ~3 Tbit/s, which could be doubled by using 25 GHz channel spacings. Ultimately, even higher channel densities and hence overall system data rates are possible, with a practical limit ~10 Tbit/s. The channel spacing is a key parameter in that it determines the laser design, specifically its spectral stability and linewidth, and the overall transmitter design in that laser chirp (modulation induced dynamic spectral shift) must be less than half the channel spacing. In practice, only single-wavelength and coarse wavelength division multiplexing (CWDM) systems may be modulated directly. All modern DWDM systems require external modulation, for example using electro-absorption or Mach–Zehnder interferometric devices outside the laser cavity. For the latter cases, stabilization of the laser wavelength is required for example by on-chip gratings to form distributed feedback (DFB) or distributed Bragg reflector (DBR) lasers, or by external fiber gratings.

Directly Modulated Lasers

Direct modulation of semiconductor lasers is convenient and effective: by modulating the laser injection current, one modulates the carrier density and hence the optical gain, resulting in modulation of the laser output up to ~10 Gbit/s. The actual modulation bandwidth is determined by interactions between photons and carriers, whose respective decay lifetimes, τ_s and τ_p, are determined by cavity losses and total recombination rates. The simplest form of such interactions is described by the photon and carrier rate equations for the injected carrier density N and photon density S in a single lasing mode:

$$\mathrm{d}N/\mathrm{d}t = J/ed - gNS - N/\tau_s \qquad [1]$$

$$\mathrm{d}P/\mathrm{d}t = gNS + \beta N/\tau_s - S/\tau_p \qquad [2]$$

where J is the injected current density, e the electronic charge, d is the thickness of the active region (hence J/ed is the rate of injection of carriers per unit volume), g is the gain coefficient per unit length and β is the fraction of the spontaneous emission coupled into the lasing mode, typically $\sim 10^{-2}$–10^{-5} for edge-emitting lasers (product of geometric and spectral overlap factors). These equations simply state that the rate of change of the carriers or photons is given by the rate of generation less the loss rate. Additional terms are required in the presence of optical feedback or coupling (in which case the photon density is replaced by the complex amplitude and phase of the optical electric field) or for extremely rapid modulation where the traveling wave nature of the disturbances in the photon fields is significant. These so-called traveling-wave rate equations for the photons are of the form:

$$\mathrm{d}S_+/\mathrm{d}t + c\mathrm{d}S_+/\mathrm{d}z = gNS_+ + \beta N/\tau_s \qquad [3]$$

$$dS_-/dt - cdS_-/dz = gNS_- + \beta N/\tau_s \qquad [4]$$

with the same carrier rate equation as [1] using the total photon number S made up of the forward-traveling component S_+ and the backward-traveling component $S_- : S = S_+ + S_-$.

Both sets of rate equations are based on approximations such as homogeneous gain broadening, neglect of transverse field effects, and diffusion. These are valid in most situations but care is required when analyzing lasers with large transverse apertures, or when ultrashort (picosecond) pulses are involved.

Small-Signal Modulation

Small-signal modulation may be analyzed by a linear perturbation analysis of the rate equations, leading to a conjugate pair of poles in the response function, with damped relaxation oscillations at frequency:

$$f_R = (1/2\pi)\sqrt{(gS_0/\tau_p)} \qquad [5]$$

or, in terms of the injected current density J (assuming $S \sim G(J - J_{th})/ed$ with Γ the electromagnetic confinement factor of the mode):

$$f_R = (1/2\pi)\sqrt{[\Gamma g(J - J_{th})/ed]}$$

f_R can be regarded as the resonant frequency for the interaction between the carriers and photons. As a practical matter, lasers can be modulated up to $\sim 2f_R$ but the modulation response rolls off rapidly with increasing frequency above f_R. Using explicit expressions for the threshold gain, we can write f_R as:

$$f_R = (1/2\pi)\sqrt{[(GN_T g\tau_p + 1)(J/J_{th} - 1)/\tau_s\tau_p]} \qquad [6]$$

where N_T is the transparency carrier density $\sim 10^{18}\,\mathrm{cm}^{-3}$. f_R can take values in excess of 10 GHz for a well-designed laser. It should be noted, however, that high modulation bandwidth almost always requires high differential gain dg/dN and short τ_p. Increasing dg/dN requires p-doping (which increases optical loss) and/or some special quantum-confining structure such as wells, wires, or dots. Decreasing τ_p inevitably leads to greater optical loss and thus a higher threshold.

The damping rate of the relaxation oscillations is also important in limiting modulation bandwidth, and also (when increased) in reducing the sensitivity of the laser to optical feedback. For bulk or quantum well lasers, the condition for critical damping is $A^2 = 4B$ with $A = (S/g\tau_s) - \sigma[N - N_T]/g\tau_p$ and $B = \sigma[1 + (S_0/g\tau_s) - (1 - \beta)(N - N_T)]$. For quantum dot lasers there are additional damping terms due to carrier transport and thermalization which may restrict small-signal modulation bandwidths to a few GHz.

Experimentally, small-signal modulation properties may be determined using a simple sampling oscilloscope, low-noise tunable signal generator and spectrum analyzer. A combination of these elements with automated frequency sweeping may be found in a scalar network analyzer acting as an *s*-parameter test set. The modulated laser is connected to port 1 and a high speed photodetector connected to port 2, then a swept-frequency measurement of the transfer characteristic s_{21} is performed. In addition to the limitations on modulation response due to the carrier-photon resonance, practical lasers are limited by RC parasitics in the laser chip (junction impedance) and its package. For the highest modulation speeds, packages need to be designed as microwave transmission line components with effective impedance matching and low back reflections, as characterized by the voltage standing wave ratio (VSWR).

Large-Signal Modulation

Large-signal direct modulation can be analyzed by numerical integration of the rate equations. In general, its results are beyond the scope of this review, but for effective high-speed response, digital modulation of laser diodes should be carried out with a pre-bias close to threshold. Modulation bandwidth also increases with laser power, which is limited for systems considerations by fiber nonlinearities, and for reasons of laser reliability. In pulse code modulation, care must be taken that the laser is near critical damping, to minimize thermal patterning effects due to long strings of ones (high power) or zeros (low power). Experimentally, large signal modulation is analyzed by constructing eye diagrams, in which pulse traces on a fast oscilloscope are continuously overlaid when the laser is modulated using pseudorandom binary pulses. When the signal quality is high, the high and low digital levels are easily distinguishable, resulting in an open 'eye' in the accumulated traces.

Requirements for Externally Modulated Lasers

The requirement for external modulation occurs when the laser chirp exceeds half the desired channel spacing in the communication system. Chirp occurs due to changes in the refractive index, and hence the optical phase, during modulation. This dynamic phase shift then results in an instantaneous frequency shift.

For direct modulation we have instantaneous wavelength

$$\lambda(t) = \lambda(0)(\mu(t)/\mu(0) \quad [7]$$

Where $\mu(t)$ is the refractive index at time t. μ shifts with carrier density N due to combinations of several factors (free carrier plasma dispersion, thermal bandgap shrinkage, bandgap renormalization, dynamic Burstein–Moss effects) which give $d\mu/dN \sim -3 \times 10^{-20}\,cm^3$, which for large signal digital modulation can result in tens of GHz. The actual degree of chirp depends on the so-called antiguiding or linewidth enhancement factor a given by

$$\alpha = -2k_0(d\mu/dN)/(dg/dN) \quad [8]$$

where $k_0 = 2\pi/\lambda_0$ is the free-space wavenumber. α as defined is positive, typically in the range 2–7 depending on the material, operating carrier density, and wavelength relative to the gain peak. Distributed feedback (DFB) or distributed Bragg reflector (DBR) lasers, which use integrated gratings to control the lasing wavelength, can have lower chirp by detuning from the gain peak but therefore are likely to require careful temperature stabilization. Quantum dot lasers, based on 3D nanoclustered active regions, have the potential for very low chirp, narrow linewidths, and reduced wavelength shifts, so that directly modulated DWDM transmitters may be possible.

Integration of laser diodes with electro-absorption or Mach–Zehnder type modulators is achieved in the transmitter optical subassembly and may in some cases be accomplished by monolithic integration on the same chip. Using separate modulators allows each device to be optimized independently but requires optical assembly, alignment, and retention. Mach–Zehnder modulators in X-cut lithium niobate, for example, have essentially zero chirp and can operate to 40 Gbit/s and above, with dynamic extinction ratios (ratio between fully on and fully off) of ~20 dB. Integrated semiconductor M-Z modulators have chirp an order of magnitude less than those of directly modulated lasers, so that channel spacings ~25 GHz are possible in DWDM systems with suitable temperature and wavelength controls. Although the details of such modulators are beyond the scope of this article, the laser requirements are to generate ~10 mW of continuous power with stable center wavelength and narrow linewidth. Such lasers are usually mounted in hermetic packages (e.g., the current 14-pin butterfly standard) with integrated thermo-electric cooler, power monitor photodiode, an optical isolator to suppress optical backreflections which cause instabilities and self-pulsing, and optional wavelength locking optics.

Reliability of Lasers in Fiber Optic Systems

From the early days, when laser diode lifetimes were measured in seconds even at cryogenic temperature, enormous progress has been made in achieving materials purity and reducing crystalline defects. Today's fiber optic laser transmitters have projected lifetimes of decades. In common with electronic devices, diode lasers fail at a rate given by a 'bathtub' curve, that is the failure rate $r(t)$ defined as the probability of failure per unit time at time t, has relatively high values at low t (early failures) and high t (wearout failures) and very low values in between. In terms of the population of lasers $n(t)$ we have:

$$r(t) = (-1/n(t))\, dn(t)/dt \quad [9]$$

If Δn is the number of samples which fail in time Δt, then assuming Δt begins at $t = 0$, the effective or average failure rate r_{eff} over the interval is

$$r_{eff}(t) = \Delta t \Delta n/n(0) \quad [10]$$

In reliability science it is customary to define r_{eff} in FITs (failures integrated in time) with the time-span chosen to yield statistically significant numbers. For electronic and photonic devices, it is customary to select the time interval $\Delta t = 10^9$ h (1 billion operating hours), so that if 1% of the devices fail in 10 years ($\sim 10^5$ hours) we have $r_{eff} \sim 100$ FITs, which is the order of magnitude required by modern fiber optic laser sources.

In terms of actual statistical models, failure rates can be estimated using failure probability density functions $f(t)$. This is related to $r(t)$ by $f(t) = r(t)S(t)$ where $S(t)$ is the cumulative probability of surviving until t. When $r(t)$ is nearly constant, as in early failures due to material defects or process errors, the statistics are approximately exponential:

$$f(t) + (1/\tau)\exp(-t/\tau) \quad [11]$$

or Weibull-distributed (an exponential is a Weibull distribution with $\beta = 1$):

$$f(t) = (\beta/t^{\beta})t^{b-1}\exp[-(t/\tau)^b] \quad [12]$$

but this description is not suited to wearout failures for which $f(t)$ shows a significant increase as the population ages. For such cases, the lognormal distribution is often applicable:

$$f(t) = (1/\sigma t\sqrt{(2\pi)})\exp[-(\ln t - \ln \tau)^2/2\sigma^2] \quad [13]$$

where in each case τ is approximately the mean time to failure (MTTF).

Laser diodes undergo standardized qualification tests to demonstrate sufficient reliability prior to being used in commercial systems. Operationally, failure statistics must be tabulated and the best fit obtained. Burn-in procedures are used to screen early failures. Chip and module failures should be distinguished clearly, as the latter includes many additional factors such as thermal and power management, light coupling, mechanical or chemical integrity.

Given that effective lifetimes of decades are required, it is clearly necessary to accelerate aging to produce statistically meaningful failure rates in reasonable times $\sim 10^3$ h. For the most common modes of laser chip degradation, due to recombination-induced aggregation of defects in the crystalline epitaxial material, these are thermally activated and current driven, so that it is customary to use a modified Arrhenius law at temperature T:

$$\tau \sim (1/J^n)\exp(E_a/k_B T) \qquad [14]$$

where J is the operating current density, n is an index ~ 2, E_a is the activation energy, and k_B is Boltzmann's constant. In terms of thermal acceleration, we have a factor $F = \tau(T_A)/\tau(T_B)$ with:

$$F = \exp[(E_a/k_B)(1/T_A - 1/T_B)] \qquad [15]$$

Typical values of E_a for laser diodes are ~ 1 eV so that acceleration factors $\sim 10^3$ are possible.

Finally, it should be noted that laser diodes are generally subject to catastrophic failure in the event of overdriving or static discharge, the failure mode being thermal facet damage at $\sim 1-10$ MW/cm^2 for continuous-wave operation, the exact value depending on the material, surface preparation and specifically surface state absorption and its temperature coefficient.

Conclusions

Laser diodes are ideal sources for optical fiber communication systems and have propelled the development of fiber optics from its origins in the 1960s to the present day. They are compact, rugged, efficient, and reliable sources of light at key wavelength ranges such as 1300–1310 nm (short haul high speed links) and 1500–1600 nm (long haul amplified systems). They are capable of direct modulation at gigabit rates for simple systems but require external modulation and careful wavelength control for dense wavelength division multiplexed terabit systems. Transmitter lasers for real systems must satisfy stringent reliability conditions.

See also

Lasers: Edge Emitters; Semiconductor Lasers. **Semiconductor Physics:** Quantum Wells and GaAs-based Structures.

Further Reading

Telcordia Technologies (formerly Bellcore): document family FR-796 on Reliability and Quality Generic Requirements for telecommunications equipment. Latest issue is 003 dated Jul 2003. The specific document related to laser reliability is GR-468 – Generic Reliability Assurance Requirements for Optoelectronic Devices used in Telecommunications Equipment. Standards for other network photonic, electronic and mechanical components may be found in the document family FR-2063 Network Equipment-Building System Family of Requirements, issue 003 dated Feb 2003 and specifically FR-357 – Network Equipment-Building System Family of Requirements: Component Design for Reliability.

Adams MJ and Thomas B (1977) Detailed calculations of transient effects in semiconductor injection lasers. *IEEE Journal of Quantum Electronics* QE-13: 580.

Agrawal GP (2002) *Fiber Optic Communication Systems*, 3rd edn. New York: Wiley.

Alferov ZhI, Andreev VM, Portnoi EL and Trukan MK (1970) AlAs-GaAs heterojunction injection lasers with a low room-temperature threshold. *Soviet Physics – Semiconductors* 3: 1107.

Basov NG, Krokhin ON and Popov YuM (1961) Production of negative-temperature states in p-n junctions of degenerate semiconductors. *Soviet Physics – JETP* 13: 1320.

Bernard MGA and Duraffourg G (1961) Laser conditions in semiconductors. *Physica Status Solidi* 1: 699.

Bhattacharya P (1996) *Semiconductor Optoelectronic Devices*, 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.

Cheng J and Dutta NK (eds) (2000) *Vertical Cavity Surface-Emitting Semiconductor Lasers: Technology and Applications*. London: Taylor & Francis.

Chuang S-L (1995) *Physics of Optoelectronic Devices*. New York: Wiley.

Coldren LA and Corzine SW (1995) *Laser Diodes and Photonic Integrated Circuits*. New York: Wiley.

Dupuis RD (1987) An introduction to the development of the semiconductor laser. *IEEE Journal of Quantum Electronics* QE-23: 651.

Fukuda M (1991) *Reliability and Degradation of Semiconductor Lasers and LEDs*. Norwood, MA: Artech House.

Fukuda M (1998) *Optical Semiconductor Devices*. Wiley.

Guekos G (ed.) (1998) *Photonic Devices for Telecommunications*. New York: Springer.

Hall RN, Fenner GE, Kingsley JD, Soltys TJ and Carlson RO (1962) Coherent light emission from GaAs junctions. *Physical Review Letters* 9: 366.

Hayashi I, Panish MB, Foy PW and Sumski S (1970) Junction lasers which operate continuously at room temperature. *Applied Physics Letters* 17: 109.
Holonyak N Jr and Bevacqua SF (1962) Coherent (visible) light emission from Ga(As1-xPx) junctions. *Applied Physics Letters* 1: 82.
Ito R, Nakashima H, Kishino S and Nakada O (1975) Degradation sources in GaAs-AlGaAs double-heterostructure lasers. *IEEE Journal of Quantum Electronics* QE-11: 551.
Kaminow IP and Li T (eds) (2002) *Optical Fiber Telecommunications IV (two parts)*. Amsterdam: Elsevier.
Kaminow IP and Koch TL (eds) (1997) *Optical Fiber Telecommunications III (two parts)*. London: Academic Press.
Kapon E (1998) *Semiconductor Lasers, (two parts)*. London: Academic.
Kartalopoulos SV (1999) *Introduction to DWDM Technology*. New York: Wiley.
Kartalopoulos SV (1999) *Understanding SONET/SDH and ATM*. IEEE Press.
Kasap SO (2001) *Optoelectronics and Photonics: Principles and Practices*. Englewood Cliffs, NJ: Prentice-Hall.
Kazovsky L, Benedetto S and Willner A (1996) *Optical Fiber Communication Systems*. Norwood, MA: Artech House.
Keiser G (1991) *Optical Fiber Communications*. New York: McGraw-Hill.
Kressel H and Lockwood HF (1974) A review of gradual degradation phenomena in electroluminescent diodes. *Journal de Physique* 35: C3–223.
Lasher G and Stern F (1964) Spontaneous and stimulated recombination radiation in semiconductors. *Physical Review* 133: A553.
Milonni PW and Eberly JH (1988) *Lasers*. New York: Wiley.
Morthier G and Vankwikelberge P (1997) *Handbook of Distributed Feedback Laser Diodes*. Norwood, MA: Artech House.
Nathan MI, Dumke WP, Burns G, Dill FH Jr and Lasher G (1962) Stimulated emission of radiation from GaAs p-n junctions. *Applied Physics Letters* 1: 62.
Numai T (2004) *Fundamentals of Semiconductor Lasers*. New York: Springer.
Paoli TL (1977) Changes in the optical properties of cw (AlGa) as junction lasers during accelerated aging. *IEEE Journal of Quantum Electronics* QE-13: 351.
Paoli TL and Ripper TE (1970) Direct modulation of semiconductor lasers. *Proceedings of IEEE* 58: 1457.
Pearsall TP (2003) *Photonics Essentials: An Introduction with Experiments*. New York: McGraw-Hill.
Piprek J (2003) *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*. London: Academic.
Quist TM, Rediker RH, Keyes RJ, Krag WE, Lax B, McWhorter AL and Zeiger HJ (1962) Semiconductor maser of GaAs. *Applied Physics Letters* 1: 91.
Reinhart FK, Hayashi I and Panish MB (1971) Mode reflectivity and waveguide properties of double-heterostructure injection. *Lasers* 42: 4466.
Saleh BEA and Teich MC (1991) *Fundamentals of Photonics*. New York: Wiley.
Senior JM (1992) *Optical Fiber Communications*, 2nd edn. Engelwood cliffs, NJ: Prentice-Hall.
Siegman AE (1986) *Lasers*. University Science Books.
Verdeyen JT (1994) *Laser Electronics*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.
Yariv A (1989) *Quantum Electronics*, 3rd edn. New York: Wiley.
Zachos TH and Ripper JE (1969) Resonant modes of GaAs junction lasers. *IEEE Journal of Quantum Electronics* QE-5: 29.

Local Area Networks

E Wong, University of Melbourne, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Local area networks (LANs) are privately owned networks that are used to carry internal traffic loads within an organization. A LAN may exist on many levels categorized by its application and speed. This is illustrated in Figure 1 whereby, at the simplest requirement, a low-speed LAN may be used to interconnect a cluster of personal computers and workstations. Such a scenario may exist in a department within a university campus, or a building within a technology park. If need be, these individual departmental or building LANs may be interconnected using a moderately high-speed backbone LAN. The backbone LAN may also provide interconnection to servers which support mass storage devices that are shared by the individual lower speed LANs. In larger organizations, a much higher speed LAN is linked to the backbone LAN to support bulk data transfer between mainframes, supercomputers, and other office equipment.

The common transmission mediums for traditional low-speed LANs are coaxial cable and twisted pairs. High-speed backbone LANs advocate the use of the optical fiber. The continuing progress in fiber-optic technology and devices, and the availability of appropriate standards have given rise to LANs that can span distances of hundreds of kilometers or more, and that operate at data rates in the order of hundreds to thousands of megabits per second (Mb/s).

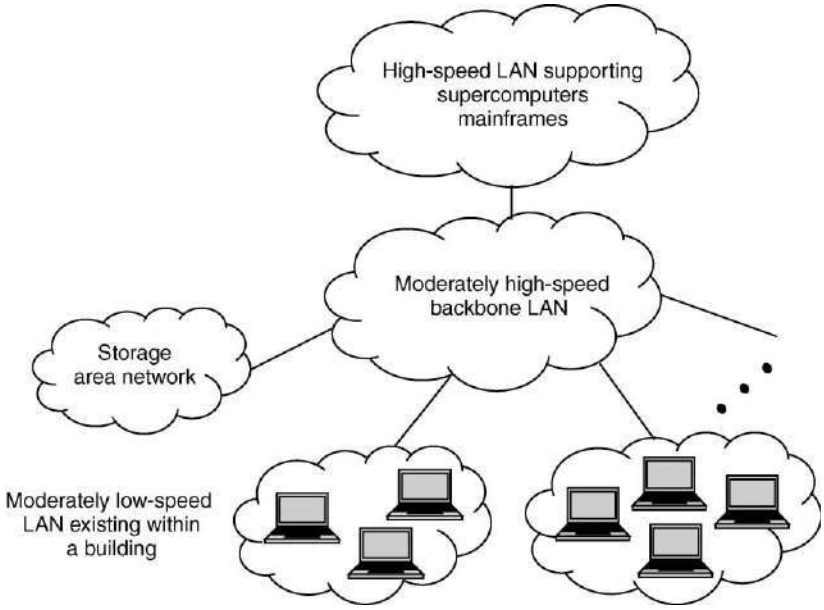

Figure 1 Interconnection of local area networks (LANs).

The increase in demand, to extend LAN capabilities towards higher data rates and across larger geographical areas, is fueled primarily by the need to provide new bandwidth intensive services such as multimedia teleconferencing, image processing, and streaming video, to a significantly higher number of end-users.

This article focuses on three existing fiber-optic LAN standards, namely Ethernet, Fiber Distributed Data Interface (FDDI), and Distributed Queue Dual Bus (DQDB). While the current uses of FDDI and DQDB are not as widespread as Ethernet, both standards have significantly contributed to the initial deployment of fiber-optics for LAN applications, with standard-based products still commercially available. In general, standards are required to ensure interoperability between components and equipment from different vendors and suppliers. Two key features are addressed by the standards: the physical topology, and the medium access control (MAC) protocol. A network's physical topology defines the way in which the users, or nodes, are interconnected to each other and to the transmission medium. Common optical LAN topologies are the star (e.g., Ethernet), bus (e.g., DQDB and Ethernet), and ring (e.g., FDDI). In comparison, the MAC protocol controls network access such that its capacity is efficiently and fairly shared amongst all nodes.

Access control can be either centralized or decentralized. In the former, a node wishing to transmit must wait until it receives permission from a controller, whereas in the latter, nodes collectively determine the transmission order. Further, capacity allocation to each node can be either synchronous or asynchronous. The synchronous approach guarantees a specific capacity to each node, e.g., using time division multiple access (TDMA), whereby nodes are allowed to transmit only during a designated time slot. In contrast, the asynchronous approach allows capacity to be dynamically allocated to each node. The asynchronous approach can be further categorized into round robin, reservation, and contention schemes. With round robin, nodes take it in turn to transmit in a cyclic manner and a node may transmit up to a maximum amount of data during its turn. With the reservation scheme, time on the medium is divided into slots, and a node wishing to transmit must reserve future slots by transmitting additional control information into the network. With the contention scheme, nodes randomly access the network. The Ethernet standard, discussed in the following section, is based on the contention scheme.

Ethernet

Overview

The IEEE (Institute of Electrical and Electronics Engineers) 802.3 standards committee first standardized 10 Mb/s Ethernet in 1985. It is now the most widely used technology for high-speed LANs, with standards extended to cater for operating speeds of 100 Mb/s, 1 Gb/s and 10 Gb/s, and encompassing the use of the optical-fiber transmission medium. Low-speed 10 Mb/s and 100 Mb/s Ethernet LANs use the

carrier-sense multiple access with collision detection (CSMA/CD) MAC protocol, whereas a switched technique is used for the 1 Gb/s and 10 Gb/s options.

In CSMA/CD, a node wishing to transmit an Ethernet frame first listens to the transmission medium by detecting the presence of signals on the channel (i.e., carrier-sense). This process allows the node to determine if another transmission is in progress. If the medium is in use, the node continues to listen until the medium is idle, and then transmits immediately. While transmitting, the node continues to listen to the medium. It may happen that two or more nodes may attempt to transmit at the same time. If this occurs, there will be a collision, and data from both transmissions will become garbled.

A collision can be indicated by a combined signal that exceeds a predefined CD threshold voltage at the carrier-sense receiver, taking into account the collision between frames from nodes that are located furthest apart. If a collision is detected during transmission (i.e., collision detection), the transmitting node ceases transmission and proceeds to transmit a brief jamming signal to notify all nodes of the collision. After transmitting the jamming signal, the node must back off before attempting to transmit the same data frame again. The back-off period is random, and is determined by a binary exponential back-off algorithm. Every time a node attempts to retransmit, the mean value of the random back-off period is doubled until the tenth attempt. After that the node is allowed to retransmit a further six times with the same mean random delay before discarding the frame and reporting an error to higher control layers.

The transmission medium remains unusable when frames collide but only until the collision is detected. As such, the minimum transmission time of a frame must be longer than the maximum propagation delay of the network so that collision is detected prior to the end of transmission. If shorter frames are used, then collision detection does not occur, and CSMA/CD thus exhibits the same performance as its less efficient predecessor protocol, CSMA. Accordingly, the CSMA/CD standards specify a minimum frame length and maximum network size, which place fundamental limitations on the practical application of a pure CSMA/CD LAN.

IEEE 802.3 10 Mb/s Ethernet

In 1993, the IEEE 802.3 committee extended the 10 Mb/s Ethernet standard to include the optical-fiber transmission medium. The standard notation of this option is 10BASE-F. Each transmission link comprises a pair of graded-index 62.5 μm core multimode optical fibers (MMF), one for each direction of transmission. The transmission wavelength is 850 nm. 10BASE-F can be further categorized into three specifications. The first, denoted 10BASE-FP (fiber passive star), specifies a broadcast star topology that can interconnect up to 33 nodes and repeaters. Each of these devices is attached to a passive star coupler via two optical-fiber links with distances up to 500 m. The passive star coupler splits the optical power of any incoming frame equally amongst its output links, allowing all nodes or repeaters in the network to receive the frame, albeit at a lower optical power level.

The second and third specifications, denoted 10BASE-FL (fiber link) and 10BASE-FB (fiber backbone) respectively, define point-to-point links over distances of up to 2 km. 10BASE-FL supports both nodes and repeaters, whereas 10BASE-FB supports only repeaters. A repeater is a device that detects the optical signals on its incoming link, performs electrical signal regeneration, and then optically retransmits the signals on its output link. Repeaters used in 10BASE-FB have an additional function in which the detected signals are retimed with a local clock before being retransmitted. This prevents the propagation of timing distortions, allowing distances of up to 15 cascaded repeater links to be implemented. The characteristics of the three specifications for 10BASE-F, along with the 100 Mb/s and 1 Gb/s specifications to be discussed in the following sections, are summarized in Table 1.

IEEE 802.3 100 Mb/s Ethernet (Fast Ethernet)

Fast Ethernet refers to the standard developed by the IEEE 802.3 committee to provide low-cost, Ethernet-compatible LANs operating at 100 Mb/s. The optical-fiber transmission medium option is denoted 100BASE-FX. Originally defined for FDDI (covered in the section on FDDI below), 100BASE-FX specifies the use of two 62.5 μm MMF links, each operating at a unidirectional rate of 100 Mb/s. While 10 Mb/s Ethernet can only operate in half-duplex mode, whereby a station can either transmit or receive a frame but cannot do both simultaneously, Fast Ethernet can operate in either half- or full-duplex modes. A node operating in full-duplex mode can simultaneously transmit and receive data frames.

In order to deploy full-duplex Fast Ethernet, all repeaters must be replaced with switching hubs. Figure 2a illustrates a repeater located in the central point in an active star topology. Each node is attached to the repeater via two optical links. The repeater retransmits frames received from any one of its input

Table 1 Summary of IEEE 802.3 physical and MAC layer specifications for 10 Mb/s, 100 Mb/s and 1 Gb/s Ethernet

	10BASE-FP	*10BASE-FL*	*10BASE-FB*	*100BASE-FX*	*1000BASE-SX*	*1000BASE-LX*
Data rate	10 Mb/s	10 Mb/s	10 Mb/s	100 Mb/s	1 Gb/s	1 Gb/s
Transmission wavelength	850 nm	850 nm	850 nm	850 nm	850 nm	1310 nm
Transmission medium	62.5 μm MMF pair	62.5 μm MMF pair	62.5 μm MMF pair	62.5 μm MMF pair	50 μm or 62.5 μm MMF pair	10 μm SMF, 50 μm or 62.5 μm MMF pair
MAC protocol	CSMA/CD (half duplex)	CSMA/CD (half duplex)	CSMA/CD (half duplex)	CSMA/CD (half duplex) Switched (full duplex)	CSMA/CD (half duplex) Switched (full duplex)	CSMA/CD (half duplex) Switched (full duplex)

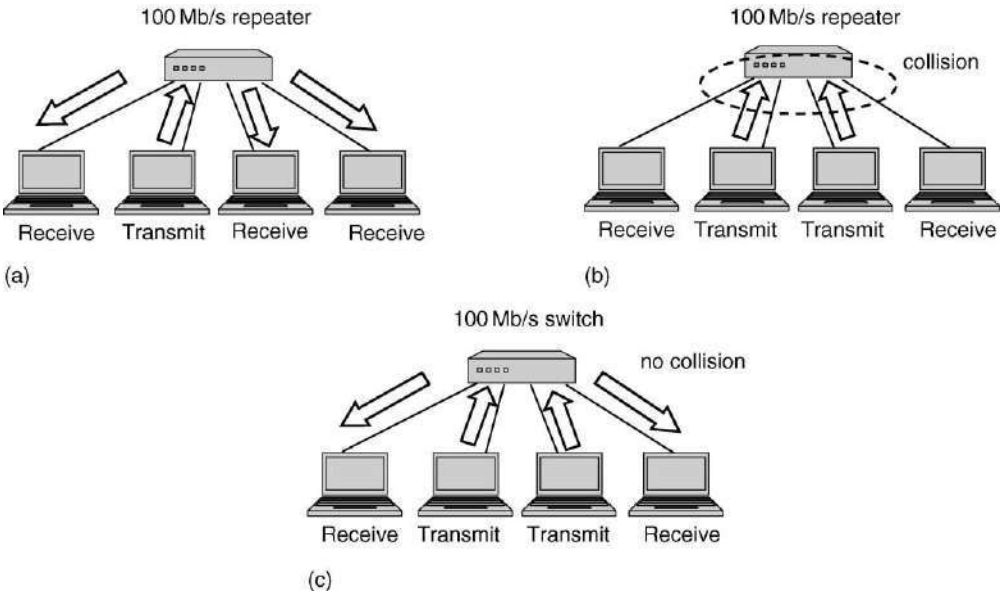

Figure 2 (a) Shared LAN repeater. (b) Collision at a shared LAN repeater due to simultaneous transmission of frames. (c) LAN switch (adapted from William S (2004) *Data and Computer Communications*, 7th edn, Ch. 15–16. Upper Saddle River, NJ: Pearson Education).

links to all of its output links. If more than one node transmits at the same time, then collisions will occur at the repeater, as illustrated in Figure 2b. A switching hub, on the other hand, switches an incoming frame from a particular input to an appropriate output line. The same switching function can be performed simultaneously on other input–output links without interference to the frames involved, as illustrated in Figure 2c. As such, each node has a dedicated capacity equal to that of the transmission rate of the LAN, and the switching hub can support mixed capacity links, e.g., 10 Mb/s and 100 Mb/s. With dedicated access to the medium and appropriate buffering at the switch to avoid contention of the same output link, there is no need to implement the CSMA/CD access algorithm. Nonetheless, the conventional Ethernet frame format is used to ensure inter-operability between existing 10 Mb/s and 100 Mb/s LANs.

IEEE 802.3 1 Gb/s Ethernet (Gigabit Ethernet)

The IEEE 802.3 1 Gb/s Ethernet standard, commonly referred to as Gigabit Ethernet, was finalized in 1998. An example of a Gigabit Ethernet backbone LAN is shown in Figure 3, whereby a Gigabit Ethernet backbone switch connects central servers to high-speed workgroup switches. In turn, each workgroup switch supports gigabit high-performance workgroup servers, as well as 100 Mb/s workstations and hubs. As in the 100 Mb/s standard, Gigabit Ethernet uses the conventional Ethernet frame format and allows both half-duplex and full-duplex operation. For half-duplex operation, the CSMA/CD access protocol is

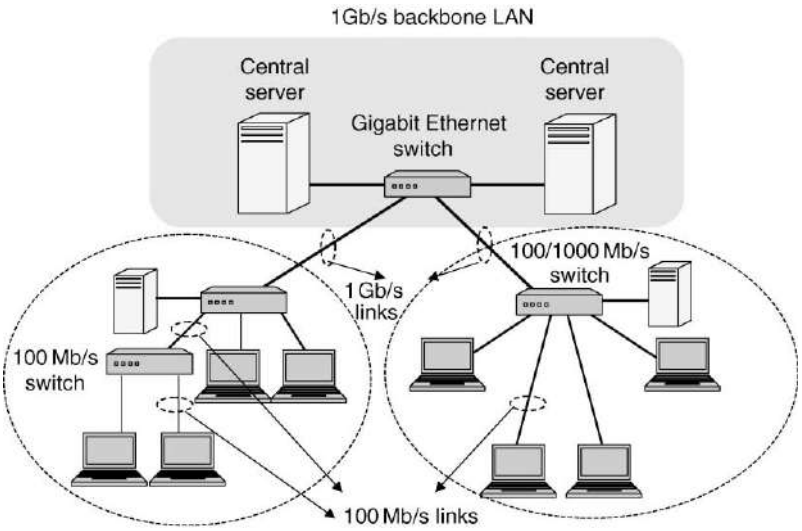

Figure 3 Gigabit Ethernet backbone network (adapted from William S (2004) *Data and Computer Communications*, 7th edn, Ch. 15–16. Upper Saddle River, NJ: Pearson Education).

equipped with two additional features. The first feature, known as carrier extension, extends the propagation time of short frames by adding nondata symbols to the end of frames. The minimum length of a frame for Gigabit Ethernet is 512 bytes as opposed to the specified minimum of 64 bytes in the preceding two standards. The second feature, known as frame bursting, allows a node to transmit multiple consecutive frames, up to a specified maximum, without having to undergo the contention process for each transmission. Both features improve the efficiency of the protocol.

There are two optical-fiber transmission medium specifications for Gigabit Ethernet. The short wavelength option, denoted 1000BASE-SX, is for 850 nm transmission. Specified maximum link distances are 275 m using 62.5 μm MMF, and 550 μm using 50 μm MMF. The long wavelength option, denoted 1000BASE-LX, is for 1310 nm transmission. Achievable duplex link distances are 550 m using 62.5 μm or 50 μm MMF, and 5 km using 10 μm SMF.

IEEE 802.3 10 Gb/s Ethernet (10GbE)

The Ethernet standard was extended to incorporate 10 Gb/s (10GbE) in 2002, fueled by ever-increasing Internet traffic and bandwidth intensive applications. 10GbE fulfills not only LAN specifications, but also metropolitan and wide-area network specifications. It is envisioned that with the emergence of 10GbE metropolitan and wide-area networks, the deployment of 10GbE LANs will enable a seamless transport of Ethernet frames across different network boundaries, thus simplifying network management and optimizing operational costs.

Unlike its predecessor standards, the physical layer options for 10GbE include only the optical-fiber transmission medium and allow only for full-duplex operation. Four specifications are defined for 10GbE. 10GBASE-S is designed for 850 nm transmission on 63.5 μm MMF, while 10BASE-L and 10BASE-LE use 10 μm single mode fiber (SMF) for 1310 nm transmission and 1550 nm transmission, respectively. The corresponding maximum achievable distances are 300 m, 10 km, and 40 km, respectively. An alternative specification, 10GBASE-LX4, is designed for wavelength division multiplexed (WDM) transmission links. The specification caters for four 2.5 Gb/s transmission channels within the 1270–1355 wavelength region on either 10 μm SMF, 50 μm MMF, or 62.5 μm MMF. Single-mode transmissions can reach distances of up to 10 km, while multimode transmissions can reach distances up to 300 m. The characteristics of these physical layer specifications are summarized in Table 2.

Fiber Distributed Data Interface (FDDI)

Overview

The Fiber Distributed Data Interface (FDDI) standard was developed by the Accredited Standards Committee X3T9.5, and approved by both ANSI (American National Standards Institute) and ISO (International Standards Organization) in 1989. While not as

Table 2 Summary of IEEE 802.3 physical and MAC layer specifications for 10 Gb/s Ethernet

	10GBASE-S	*10GBASE-L*	*10GBASE-E*	*10GBASE-LX4*
Data rate	10 Gb/s	10 Gb/s	10 Gb/s	4 × 10 Gb/s
Transmission wavelength	850 nm	1310 nm	1550 nm	1270 nm to 1355 nm
Transmission medium	62.5 μm MMF pair	10 μm SMF pair	10 μm SMF pair	10 μm SMF or 50 μm, 62.5 μm MMF pair
MAC protocol	Switched (full duplex)	Switched (full duplex)	Switched (full duplex)	Switched (full duplex)

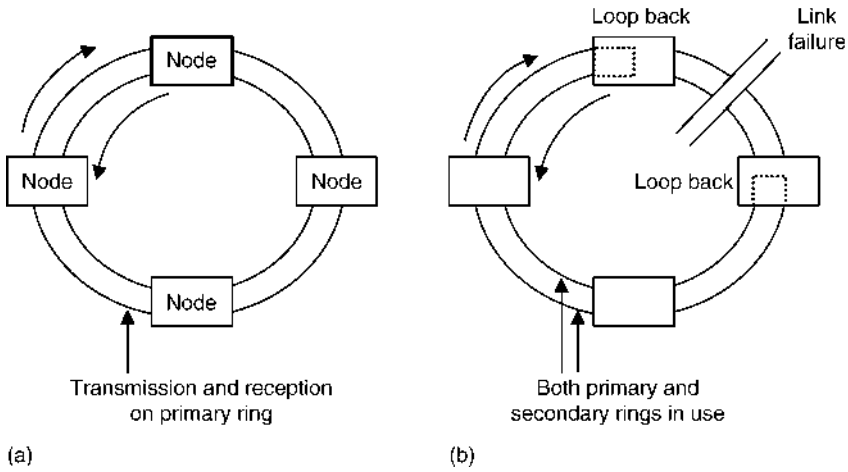

Figure 4 (a) FDDI ring network. (b) Use of secondary ring for protection against link failures.

widely deployed as the Ethernet family of LANs, due to the high cost of equipment and devices, it remains as the first to be standardized for high-speed fiber-optic LAN applications and is now mainly used to interconnect existing FDDI-compatible LANs. Extensions have been made to the FDDI standard to incorporate single-mode fiber operation for increased signal reach.

FDDI specifies a dual counter-propagating ring network as shown in Figure 4a. Each ring operates at the speed of 100 Mb/s, and can reach a maximum length of 100 km. The dual ring topology is self-healing against fiber-cuts and node failures. Under normal operation, frame transmission and reception are performed on a primary ring. The secondary ring is utilized to form a closed loop when a node or link failure occurs, as illustrated in Figure 4b. In FDDI, the interface between each node and the optical-fiber transmission medium is a repeater. Data frames, each of which carries its source and a destination address in a defined field in the frame header, are transmitted sequentially around the ring from one repeater to another. Each frame can vary in length by up to 4550 bytes. Each repeater copies the destination address field of all incoming frames and, upon recognizing the destination address as its own, additionally copies the remainder of the frame. The repeater attached to the source node also facilitates the removal of frames it has transmitted after a round-trip propagation around the ring.

FDDI employs a distributed timer-controlled, token-passing mechanism, referred to as timed token rotation protocol, to control access to the shared optical fiber ring. While Ethernet-based LANs are highly suited to support throughput sensitive services which can sustain reasonable delay in return for high throughput, FDDI provides an additional capability to support delay sensitive services such as packetised voice and video. For such services, frames must be received within a certain time interval for quality of service requirements. A node connected to an FDDI network can therefore transmit both classes of services: the former referred to as burst traffic and the latter as stream traffic. In addition, the timed token rotation protocol can limit the amount of frames that can be transmitted by a particular node for each class of service, ensuring efficient and fair network access amongst all nodes and services. FDDI is thus well-suited to carry the mixture of burst and stream traffic expected in a backbone network which supports a number of smaller LANs.

MAC Protocol

Token-passing mechanism

In FDDI, access to the shared transmission medium is governed by a timed token rotation protocol. A token is a special frame created during network (re)-initialization by a pre-assigned node. The token circulates the ring network and gives the node that holds it the right to transmit. Figure 5 gives an example of a single frame transmission from Node A to Node C, and from Node B to Node A. Node A has a frame to transmit to Node C and thus waits for a token (Figure 5a). Upon recognizing that a token frame is passing through, it aborts the token transmission without regenerating the token bits. After Node A has completely received the token, the node begins transmitting frame FA addressed to Node C, and appending a new token upon completion (Figure 5b). Meanwhile, Node B regenerates and retransmits FA as it circulates past. When FA reaches Node C, the node copies FA in addition to regenerating and retransmitting FA. Node B captures the token issued by A and transmits FB to Node A (Figure 5c). Eventually FA recirculates back to its originating node, Node A, whereby it is removed from the ring (Figure 5d). At the same time, Node B finishes transmitting FB and releases the token. Nodes C and D, with no frames to transmit or receive, regenerate and retransmit both FB and the token. When FB reaches Node A, it is copied by the node. Node A also regenerates and retransmits both FB and the token. FB is finally removed from the ring by Node B (Figure 5e), and the token remains in the ring to be received by the next node with a frame to transmit (Figure 5f).

Aside from source and destination address fields, each dataframe also contains a frame status field which comprises error detected (E), address recognized (A), and frame copied (C) indicator bits. Each node can check passing frames for errors and can set the E indicator if an error is detected using the standard cyclic redundancy check procedure. If a node detects its own address, it sets the A indicator, and upon copying the frame, the station may also set the C indicator. When a dataframe recirculates back to its source node, the indicator bits are examined. If both A and C are not set, then the destination node does not exist or is inactive. On the other hand, if the frame has not been copied by an active destination node or an error has been detected, higher control layers are notified for future retransmission.

Time-controlled mechanism

FDDI uses a time-controlled mechanism to prevent nodes from dominating the network capacity. During initialization, a target token rotation time (TTRT) is established based on the shortest allowable time between token arrivals at each node. The same TTRT value is then stored at all nodes. In addition, each node has a maximum capacity guaranteed for the transmission of its stream traffic. The total transmission time of this stream traffic and the latency of the ring (i.e., total round-trip propagation time, the transmission time of a maximum length frame and the token frame) must never exceed the TTRT. During network operation, each node measures the

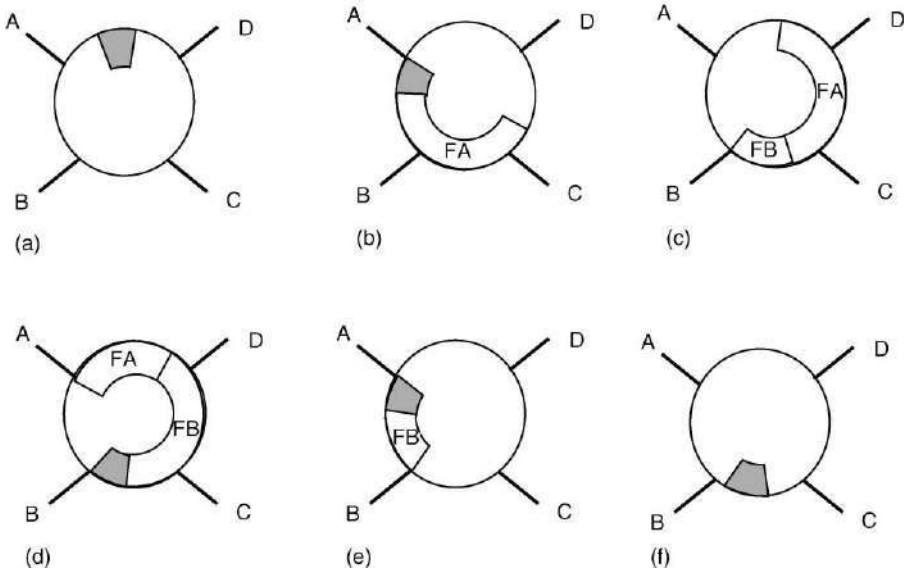

Figure 5 Fiber Distributed Data Interface (FDDI) token passing mechanism.

time interval between successive token arrivals, using a token-rotation timer (TRT). The TRT is enabled only upon a token arrival, and expires if the counter reaches the TTRT value. In turn, a late counter (LC) is incremented and TRT is reset to zero. The number of TRT expirations since a token was last received is, therefore, counted by LC. Any token that arrives at a node with a nonzero LC value is referred to as a late token. Likewise, an early token is one which arrives at a node with a zero LC value.

The action of a node when it receives a token depends on whether the token is early or late. When a late token is received, LC is reset to zero but TRT continues to run. The node then transmits for the period of time which corresponds to the predetermined bandwidth for steam traffic transmission. The node is not allowed to transmit any bursty traffic frames. In the event of an early token reception, the node transfers the current value of TRT to a second timer, the token-holding timer (THT) but does not enable it. TRT is reset to zero and is enabled to track the next rotation of the counter. The node then transmits stream traffic frames for the allocated period of time. After that, THT is enabled and bursty traffic frames are transmitted until THT reaches the TTRT value. The THT is then reset and the node releases the token into the ring. The transmission of bursty frames may be further controlled by two additional mechanisms. The restricted-token mechanism allows only two nodes in the network to use all the available bursty bandwidth, and restricts all other nodes to their stream bandwidth. This is facilitated by the use of a restricted token which is issued by one of the two nodes. The priority mechanism provides each station with the capability to transmit bursty frames according to eight priority levels. Each priority level corresponds to a different time threshold in the THT.

Note that the time-controlled mechanism of the MAC protocol guarantees a maximum response time for the ring because, in the worst case, the time between the arrival of two successive tokens will never exceed twice the value of TTRT. Most multimedia applications, such as video conferencing, real-time video, and entertainment video can be supported on FDDI with stream service and a small TTRT. However, since the TTRT cannot be less than the ring latency, applications requiring time bounds less than twice the ring latency cannot be supported by FDDI. Long and continuous isochronous services which must adhere to strict periodic access require FDDI-II.

FDDI-II

FDDI-II provides the additional circuit-switching capability to support constant bit-rate telecommunication applications. Like FDDI, the transmission bit-rate of FDDI-II is 100 Mb/s. FDDI-II can operate in either complete isochronous mode, complete packet-switching mode, or in hybrid mode in which both isochronous as well as bursty and stream services are supported. Strict periodic access required by isochronous services are accomplished with transmission opportunities repeated every 125 μs. At 100 Mb/s, time on the ring is divided into several frames, each 1562.5 bytes long. Out of the 1562.5 bytes, 1560 bytes are used for data transmission while the remaining 2.5 bytes are used as interframe gap. In turn, the 1560 bytes are divided into 16 wideband channels of 96 bytes each. Each wideband channel thus provides a bandwidth of 6.144 Mb/s, sufficiently adequate for television broadcast and multiple voice channels. For operation in the hybrid mode, some of the wideband channels are allocated for packet mode transmission. Likewise, if complete packet-switching operation is required, all wideband channels are allocated for packet switching transmission.

IEEE 802.6 Distributed Queue Dual Bus (DQDB)

Overview

The IEEE 802.6 Distributed Queue Dual Bus (DQDB) standard, finalized in 1990, specifies a dual bus network with an operating speed of 150 Mb/s. DQDB is now mainly used for MAN applications, although it can be readily deployed for LAN backbone interconnection. Figure 6 illustrates the basic architecture of a DQDB network with access nodes, Node 1 to Node N, distributed along two counter propagating buses, BUS A and BUS B. Each access node is considered to be connected to a lower speed LAN. At the head end of each bus, a slot generator continuously generates fixed-size time slots which propagate along its respective bus. The contents of each time slot are regenerated at each access node it passes, and upon reaching the end of the bus, a sink terminal receives and removes the time slots from the network. Data are transported across the network from one access node to another in the form of fixed-size segments. A node transmits a segment by changing the contents of a passing time slot. Each segment needs only be transmitted on one of the two buses, the choice depending on the locations of the transmitting and receiving nodes. For example, referring to Figure 6, a downstream transmission from Node i to Node $i+1$ requires the use of BUS A. On the contrary, an upstream transmission from Node i to Node $i-1$ uses BUS B.

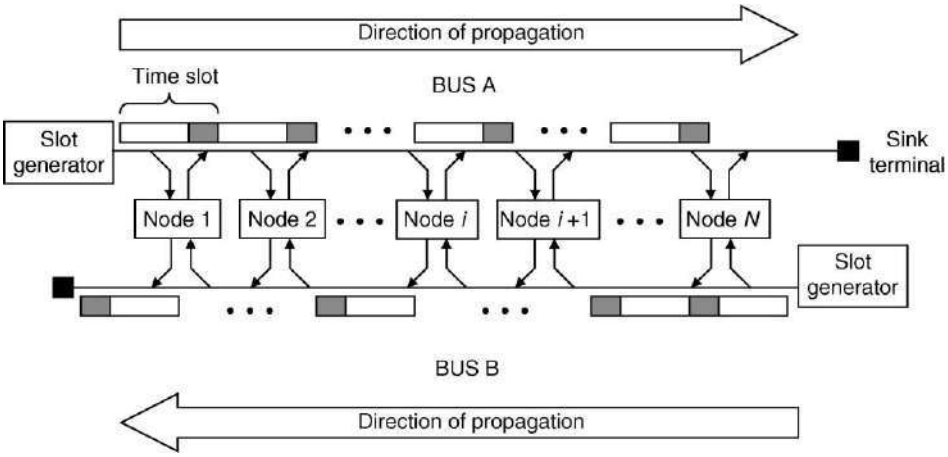

Figure 6 Distributed Queue Dual Bus (DQDB) architecture.

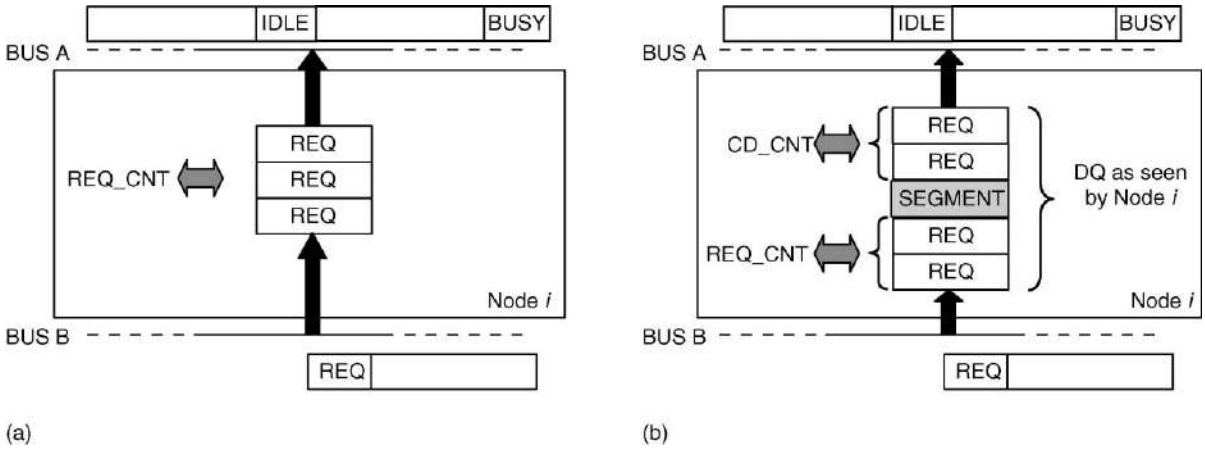

Figure 7 Behavior of Node *i* during (a) idle state; and during (b) countdown state.

MAC Protocol

In DQDB, permission to transmit segments into the network is determined by a global distributed queueing algorithm. The algorithm allows each node to continuously monitor both buses, counting the requests and available time slots that pass by. In addition to maintaining a global first-in-first-out queue (DQ) for each bus, the algorithm relies on information in the header of time slots, namely the BUSY and REQ fields, and counters at the access nodes to coordinate transmissions. A 'BUSY = 0' field represents the availability of a time slot for the transport of a segment. For example, time slots generated at the head end of each bus have default 'BUSY = 0' and 'REQ = 0' fields. The BUSY field is set to 1 when a segment is transmitted into an available time slot.

The REQ field is used in conjunction with the counters to inform nodes of an impending transmission from downstream. To simplify the explanation, consider only packets propagating on BUS A. A node wishing to transmit a segment downstream needs to queue the segment in the appropriate DQ. The node with an impending transmission must then notify all upstream nodes that it has inserted a segment in the DQ so that an available time slot can be reserved for the transmission. This is accomplished by setting the first time slot propagating in the opposite bus, BUS B, with REQ = 0 to REQ = 1. As such, the node maintains the outstanding segments in the DQ belonging to downstream nodes via the value in the REQ field.

To illustrate, Figure 7a shows an arbitrary node, Node *i*, in an idle state, whereby the node has no segments to transmit downstream on BUS A. With every 'REQ = 1' observed on BUS B, a request counter, REQ_CNT is incremented by one. However, with every empty slot, i.e., 'BUSY = 0', observed on

Table 3 Comparison of standardized local area networks

	Ethernet	*Fast Ethernet*	*Gigabit Ethernet*	*10GbE*	*FDDI*	*DQDB*
Data rate	10 Mb/s	100 Mb/s	1 Gb/s	10 Gb/s	100 Mb/s	150 Mb/s
MAC protocol	CSMA/CD	CSMA/CD or Switched technique	CSMA/CD or Switched technique	Switched technique	Timed token rotation	Distributed queue
Standard	IEEE 802.3	IEEE 802.3	IEEE 802.3	IEEE 802.3	ANSI-FDDI	IEEE 802.6

the BUS A, REQ_CNT is decremented by one. When Node *i* has a segment to transmit, it enters the countdown state as illustrated in Figure 7b. The value in REQ_CNT is transferred into an independent countdown counter, CD_CNT, and REQ_CNT is reset to zero. Node *i* also sets 'REQ = 1' in the opposite bus, BUS B. As in the idle state, each 'REQ = 1' observed on BUS B increments REQ_CNT while each 'BUSY = 1' observed on BUS A decrements it. When CD_CNT reaches zero, Node *i* transmits its segment in the first time slot with 'BUSY = 0' on BUS A and sets 'BUSY = 1'. The node returns to the idle state if there are no other segments to transmit. The algorithm allows the node to determine the number of time slots reserved by downstream nodes and the time to transmit a segment into an available time slot. Note that both the idle and countdown states are replicated for transmissions on BUS B.

The DQDB standard also specifies a limited support of isochronous services. Nodes which provide and receive such services are reserved a certain amount of time slots which are indicated with additional header information. The remaining time slots are used by all other access nodes for the transport of packet-switched traffic, and access to these time slots is managed by the global distributed queueing algorithm. A priority mechanism based on using multiple DQs, each with a different priority level, may also be implemented to guarantee access to certain segments in the network.

Summary

Due to the explosive demand for high-bandwidth applications, fiber-optics has become an essential technology in LANs. Table 3 compares the various fiber-optic LAN standards discussed in this article. The Ethernet family remains the most prevalent of LAN technologies. Its widespread use is contributed by the availability of low-cost chipsets, mature and familiar management and analysis tools. The future of fiber-optic LANs operating beyond 10 Gb/s will depend upon WDM solutions. The current usage of one transmission channel per network is impeded primarily by the limitation of the speed of electro-optic conversion, and thus does not fully exploit the massive transmission bandwidth of the optical fiber. With WDM, multiple transmission channels are concurrently transmitted on a single optical fiber, thereby greatly increasing the capacity of an optical LAN. Ongoing research efforts have been directly towards advancing WDM component technology and addressing prevailing issues such as polarization control, chromatic dispersion, and four-wave mixing among others.

List of Units and Nomenclature

CD_CNT	countdown counter
CSMA/CD	carrier-sense multiple access with collision detection
DQ	global first-in-first-out queue
DQDB	distributed queue dual bus
FDDI	fiber distributed data interface
LAN	local area network
LC	late counter
MAC	medium access control
MAN	metropolitan area network
MMF	multimode fiber
REQ_CNT	request counter
SMF	single mode fiber
TDMA	time division multiple access
THT	token-holding timer
TRT	token rotation timer
TTRT	target token rotation timer
WDM	wavelength division multiplexing

See also

Information Processing: Optical Neural Networks.

Further Reading

IEEE Std 802.6 (1990) IEEE standards for local and metropolitan area networks: distributed queue dual bus (DQDB) subnetwork of a metropolitan area network (MAN).

IEEE Std 802.3 (2002, Revision of IEEE Std 802.3, 2000 ed.) IEEE Standard for Information Technology – Telecommunications and Information Exchange Between Systems – Local and Metropolitan Area

Networks – Specific Requirements Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications.
ISO 9314-1,2 (1989) Information processing systems – Fibre Distributed Data Interface (FDDI) – Part 1: Token Ring Physical Layer Protocol (PHY) and Part 2: Token Ring Media Access Control (MAC).
Marco C, Enrico G and Luciano L (1997) *Metropolitan Area Networks*. New York: Springer-Verlag.
Mills A (1995) *Understanding FDDI*. Englewood Cliffs, NJ: Prentice Hall.
William S (2004) *Data and Computer Communications*, 7th edn, Ch. 15–16. Upper Saddle River, NJ: Pearson Education.

Optical Time Division Multiplexing

L P Barry, Dublin City University, Dublin, Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

As the demand for bandwidth continues to grow, driven by the massive increase in Internet usage, so will the necessity to have communications networks that can handle very high data rates. The use of optical fiber networks is the clear choice for such systems given the huge available bandwidth of the fiber transmission medium. However, in a basic optical communication system comprising a laser transmitter, an optical fiber transmission medium, and a receiver, the capacity is essentially limited by the speed at which light can be modulated at the transmitter. To overcome this limitation, which is basically due to the speed of available electronics, it is necessary to use optical multiplexing techniques, and one such technique, known as optical time division multiplexing (OTDM), will be the subject of this article.

Principles of Time Division Multiplexing

Time division multiplexing (TDM) has long been the traditional method for electrically combining information channels. If we take the most fundamental data rate to be that of a simple voice call at 64 kbit/s, then the transmission of data at higher bit rates is achieved by electrically multiplexing a large number of 64 kbit/s channels in the time domain. With the evolution of standards, a number of different data rates have been specified as standard transmission rates. In Europe the synchronous digital hierarchy (SDH) has a basic data rate of 155.52 Mbit/s, which is known as the synchronous transport module – Level 1 (STM-1). This particular data rate is essentially obtained by electrically multiplexing over 2000 voice calls in the time domain (with some of the capacity required for overhead information). By subsequently multiplexing a number of STM-1 channels together we can obtain transmission at the higher standard data rates of STM-4 (622 Mbit/s), STM-16 (2.48 Gbit/s), and STM-64 (9.88 Gbit/s), as outlined in Table 1.

Figure 1 illustrates how basic electrical TDM is used in standard optical communication systems. Clearly as we approach the higher data rates of STM-64, a serious level of electrical multiplexing is required, and as the data rates increase so does the cost and complexity of the electrical equipment at the transmitter and receiver. Indeed the present state of the art in electronics seems to suggest 40 Gbit/s as the limit for electrically multiplexed communication systems. However, as we stated earlier, communications traffic has been growing explosively over the last decade and will continue to do so. In order to meet this demand for capacity, and better exploit the massive available bandwidth of optical fiber, it is necessary to use optical multiplexing techniques for communications systems. The two main optical multiplexing techniques available are wavelength division multiplexing (WDM) and optical time division multiplexing (OTDM). WDM essentially involves transmitting data at a number of different wavelengths on the same fiber. Although electrical multiplexing may be limited to data transmission rates of around 40 Gbit/s using one laser, by multiplexing together N different wavelength channels each carrying 40 Gbit/s, we can achieve overall data rates up to and beyond a terabit/s.

The second of these optical multiplexing techniques, OTDM, is the subject of this article. Whereas WDM multiplexes optical data channels in the

Table 1 Standard data rates for SDH transmission systems

SDH standard	*Data rate (Mbit/s)*
STM-1	155.52
STM-4	622.08
STM-16	2488
STM-64	9953
STM-256	39 813

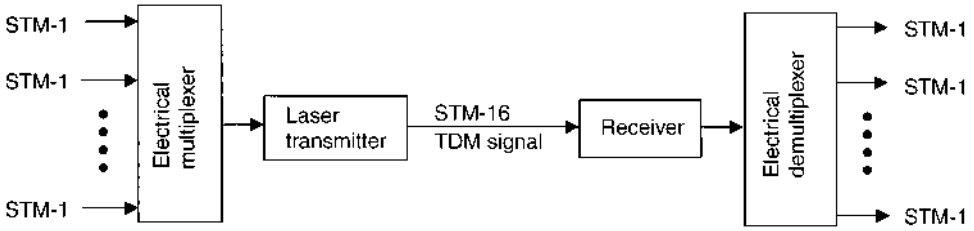

Figure 1 Electrical time division multiplexing (ETDM) in a basic optical communication system.

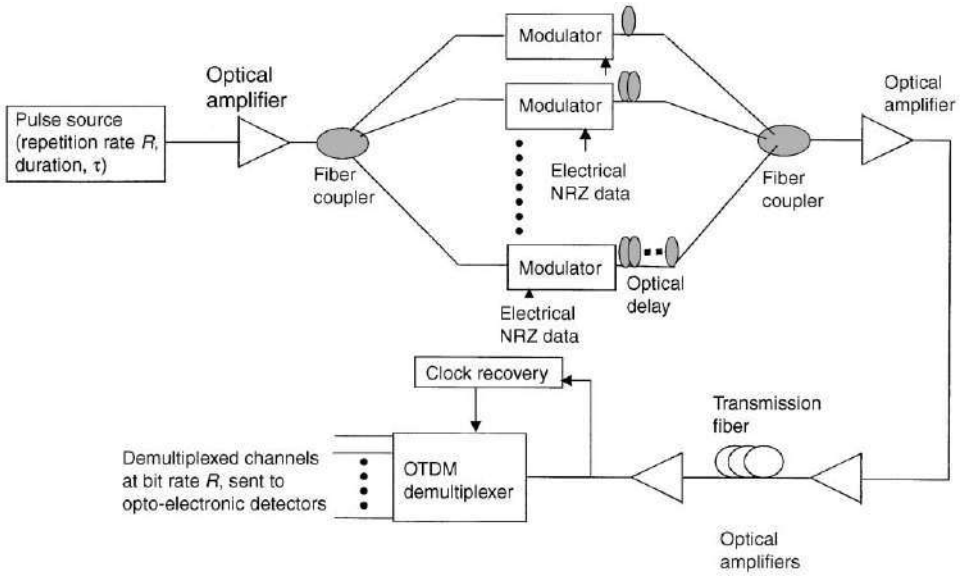

Figure 2 Basic configuration for an OTDM transmission system comprising transmitter, transmission fiber, and receiver.

wavelength domain, the basic principle of OTDM communications is to increase the system capacity by multiplexing optical data channels in the time domain. This multiplexing is usually achieved by a process known as bit-interleaving. This process can be explained by considering Figure 2, which shows a basic schematic of an OTDM based transmission system. The main component of the overall system is a source of ultrashort optical pulses (pulse duration τ) at a certain repetition rate, R. The optical pulse source is initially split into N channels using a passive fiber coupler, and each pulse train is subsequently modulated by electrical data which is at a data rate of R. The resulting output from each modulator is essentially an optical data channel where the data are represented using ultrashort optical pulses (return-to-zero data format). The data from each modulator then passes through a fixed fiber delay line which delays each channel by a time equal to $1/RN$ relative to its adjacent channel (as shown in Figure 2). The N modulated and delayed optical data channels are then recombined in another passive fiber coupler to form the OTDM data signal. We can consider that the delay lines essentially assign each data channel to a specific bit slot (of width $1/RN$) in the overall multiplexed signal. The multiplexed data signal may then be transmitted over optical fiber before arriving at the receiver which is responsible for demultiplexing the optical signal into its discrete channels.

The duration of the optical pulse source is extremely important in determining the maximum overall data rate which can be achieved. The overall data rate is basically defined by the temporal separation between channels in the multiplexor, but in order to avoid cross-channel interference, this separation must be significantly greater than the pulse duration. Thus to increase the overall data rate we must use shorter optical pulses. However, as we reduce the optical pulse width to raise the data rate we need to take into account the problems that may be encountered as this high-speed data signal propagates over optical fiber, and also the difficulty in demultiplexing a high-speed OTDM data signal. The following section will look in greater detail at

these key elements of the OTDM communication system, namely the optical pulse source, the transmission of the ultrahigh capacity signal, and the demultiplexing at the receiver.

Key Elements of an OTDM Communication System

Ultra-Short Optical Pulse Sources

As stated earlier, the optical pulse source is a key element in any OTDM-based communication system. The important characteristics of the pulse source that will affect its usefulness in an OTDM system are the pulsewidth, the spectral width, and the temporal jitter. The pulse duration clearly has to be short enough to support the desired overall transmission rate. For example, if we wish to design an OTDM system with an aggregate data rate of 100 Gbit/s (which will have a delay of 10 ps between each channel in the multiplexed signal), then the pulsewidth should normally be less the 30% of the channel spacing to avoid cross-channel interference (i.e. around 3 ps). For terabit/s OTDM systems we would require pulsewidths less than 0.5 ps. The spectral width of the pulse source is also important, as it will have a major impact on how the pulse will evolve during propagation in the fiber. A standard figure of merit which is employed is the time–bandwidth product, and ideally we require the pulse source to be transform limited, which implies that the spectral width is as small as possible for the associated pulsewidth. The impact of temporal jitter on the pulses can be easily understood by considering the above example of a 100 Gbit/s system, employing 3 ps pulses, with the multiplexed channels spaced by 10 ps. Obviously if the jitter on the pulses becomes significant in relation to the channel spacing, then this can also lead to interference between adjacent channels in the overall OTDM system. A large number of techniques have been employed to develop ultrashort pulse sources suitable for use in OTDM systems, but the three main methods which will be described below are active mode-locking, gain-switching, and external modulation of a CW light signal.

Active mode-locking of laser diodes normally involves modulating the amplitude of the optical field inside the laser cavity at a frequency which is equal to the mode spacing of the laser. This can be achieved by applying an electrical sinusoidal signal at the correct frequency, and results in the generation of optical pulses at the repetition rate of the applied signal. This technique has been successfully employed in the generation of subpicosecond pulses at repetition rates up to and beyond 40 GHz, with excellent spectral and temporal jitter characteristics. However, an inherent problem in all mode-locked pulse sources is the difficulty in synchronizing the mode-locked frequency to a specific SDH standard data rate.

Gain switching of semiconductor laser diodes is probably the simplest and most reliable technique to generate optical pulses. The technique, which is presented in Figure 3, involves applying a high-power electrical pulse (or electrical sinusoidal signal) to the laser in conjunction with a certain bias current. By ensuring that the electrical pulse signal and the bias signal have the correct level, the relaxation oscillation phenomenon of the laser results in the production of optical pulses with durations between 10 and 30 ps, at the repetition rate of the applied electrical signal. The frequency of the electrical signal applied to the laser is essentially arbitrary (provided it is not larger than the modulation bandwidth of the diode), thus making it straightforward to synchronize the optical pulse train to a SDH line-rate. The main problem with this technique is that the spectral width of the pulses generated is such that the pulses are far from transform-limited, which would affect their subsequent propagation in the fiber. In addition, temporal jitter on gain-switched pulses can also be a problem. However, by using novel arrangements such as external injection into the gain-switched laser, this difficulty can be overcome.

The third pulse generation technique mentioned above involves external modulation of a cw light signal with an electro-absorption modulator. The experimental configuration for this pulse generation technique is shown in Figure 4. By biasing the modulator around its null point, and applying an electrical sinusoidal signal to it, the cw light passing through the modulator becomes shaped into optical pulses. The optical pulse train which is generated due

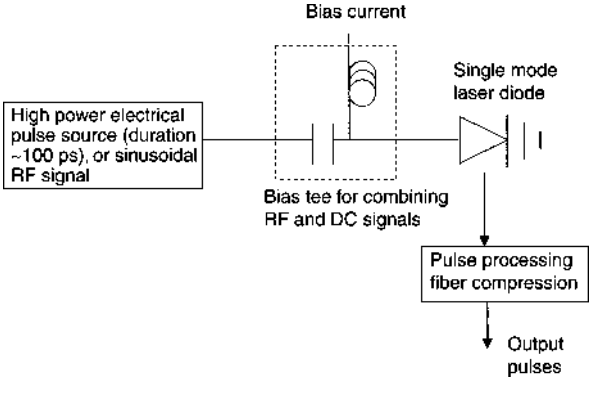

Figure 3 Pulse generation using the gain-switching technique followed by pulse compression.

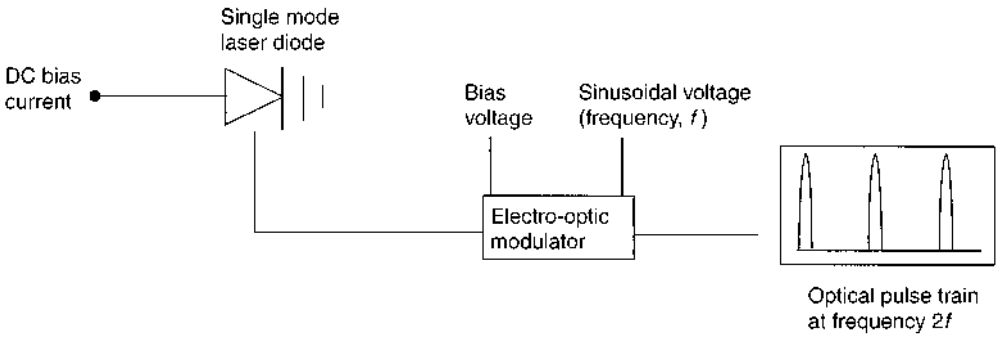

Figure 4 Pulse generation based on shaping of cw light using the nonlinear response of an external modulator biased about its null point.

the nonlinear response of the electro-absorption modulator is at a repetition rate of twice the applied electrical signal. The pulses generated are normally transform limited with extremely low temporal jitter, and the repetition rate is arbitrary as in the gain-switching technique (with the limit being ultimately determined by the modulation bandwidth of the modulator). This method may be readily used to generate pulses at repetition rates up to 40 GHz with pulsewidth around 5 ps, and it has the advantage that the optical source and the modulator can be integrated in a single device to form a compact pulse source suitable for OTDM communication systems.

It should also be noted that in addition to the various techniques that have been outlined, it is possible to use pulse compression in order to reduce the pulse width. The main issues concerned with pulse compression are the shape and spectral width of the optical pulses after compression. One of the most attractive methods of achieving pulse compression involves using nonlinear compression in dispersion-decreasing fiber, with the main advantage of this technique being the ability to maintain a transform limited pulse after compression. By employing this compression scheme, optical pulse sources at 10 GHz with pulsewidths below 200 fs have been developed, and such pulsewidths would be suitable for use in Tbit/s OTDM systems.

Transmission of an OTDM Signal over Fiber

The transmission performance of an OTDM data signal is vital in determining the distance over which the data can be transmitted successfully. The main fiber parameters, which will affect the signal transmission, are attenuation and dispersion. If we consider OTDM systems operating at a wavelength of around 1550 nm (minimum loss wavelength), to overcome the fiber attenuation and maintain a suitable optical power budget for the system, optical amplifiers are normally employed. If we thus assume that the amplifiers overcome the fiber loss problem, then the maximum transmission distance will be limited by the fiber dispersion. In a very basic OTDM system operating at 1550 nm, with transmitter and receiver linked using standard fiber (dispersion parameter of about 16 ps km^{-1} nm^{-1}), the maximum transmission distance will be limited by the overall data rate, and the pulsewidth and spectral width of the optical pulse source. For example, consider a 40 Gbit/s OTDM data signal which is formed using 8 ps optical pulses with a spectral width of 40 GHz (0.32 nm). From the spectral width we can calculate the signal broadening due to dispersion to be around 5 ps km^{-1}, and as the pulses broaden and spread into adjacent channels of the OTDM signal then this interference will make it increasingly difficult to correctly detect the signal at the receiver. In this case after propagation through 5 km of fiber, the signal pulses will have already dispersed to around 25 ps duration, the same value as the temporal bit slot into which each channel is placed. This will clearly result in serious loss of signal integrity. A straightforward possibility to increase the transmission distance of OTDM systems is to employ dispersion shifted fiber; the dispersion parameter is now around 1–2 ps km^{-1} nm^{-1}, at an operating wavelength of 1550 nm. This reduction in dispersion will obviously increase the allowed transmission distance by about an order of magnitude for the example described above. However, to develop ultrahigh-speed, long-haul OTDM communications we require more complex transmission schemes. Two possibilities for this include dispersion compensation, and soliton transmission techniques.

Dispersion compensation basically involves compensating for the dispersion that has been encountered during transmission by using some fiber with a total dispersion of opposite sign but equal magnitude to the transmission fiber. Dispersion compensation

may also be achieved using a suitably designed fiber grating. For a long-haul OTDM system, a dispersion compensator may be used every 50 or 60 km, in conjunction with the optical amplifier, thus allowing us to compensate both fiber loss and attenuation periodically along the link. The main limitation, however, with the dispersion compensation technique is caused by the dispersion slope of the fiber, as for ideal compensation it is necessary to compensate completely for the dispersion slope in addition to the overall fiber dispersion. The second technique that may be employed to greatly extend the transmission distance of high-speed OTDM communication systems is the use of soliton transmission. The basic principle of soliton transmission is to use optical data pulses with a particular shape, pulsewidth, and peak power, such that as the pulse propagates, the effects of fiber dispersion and nonlinearity counterbalance to allow the signal to propagate undistorted. By using optical amplifiers to ensure that the optical pulse peak power does not vary too much along the transmission link, OTDM transmission at data rates greater than 100 Gbit/s, over distances approaching 500 km have been demonstrated.

Demultiplexing of OTDM Signal at Receiver

For OTDM communication systems with data rates above 40 Gbit/s, it is not feasible to use electrical switching. Indeed for ultrahigh-speed systems operating at 100 Gbit/s and beyond, the only solution for demultiplexing is to use all-optical switching in which optical control signals are used to switch the OTDM data signal. All-optical switches normally use nonlinear optical effects either in optical fiber or semiconductors. A typical scenario (as presented in Figure 5) involves the injection of both the OTDM data signal and an optical control signal into the nonlinear device. The control signal consists of high-power ultrashort pulses at the repetition rate of the individual channels within the temporally multiplexed signal, and by synchronizing it with one of the OTDM channels it is possible to demultiplex this channel from the high-speed OTDM signal.

Two of the most popular all-optical demultiplexers are the nonlinear optical loop mirror (NOLM), and the terahertz optical asymmetric demultiplexer (TOAD). The NOLM is based on the nonlinear refractive index of optical fiber. It essentially consists of a 2×2 fiber coupler with its two outputs joined using a certain length of fiber. When an OTDM signal is injected into an input port of the coupler, it splits into two counterpropagating components in the fiber loop, and when these components recombine and interfere at the coupler, the overall signal is output through its initial input port. If we now inject a high-power control signal directly into the loop, such that it propagates unidirectionally, and is

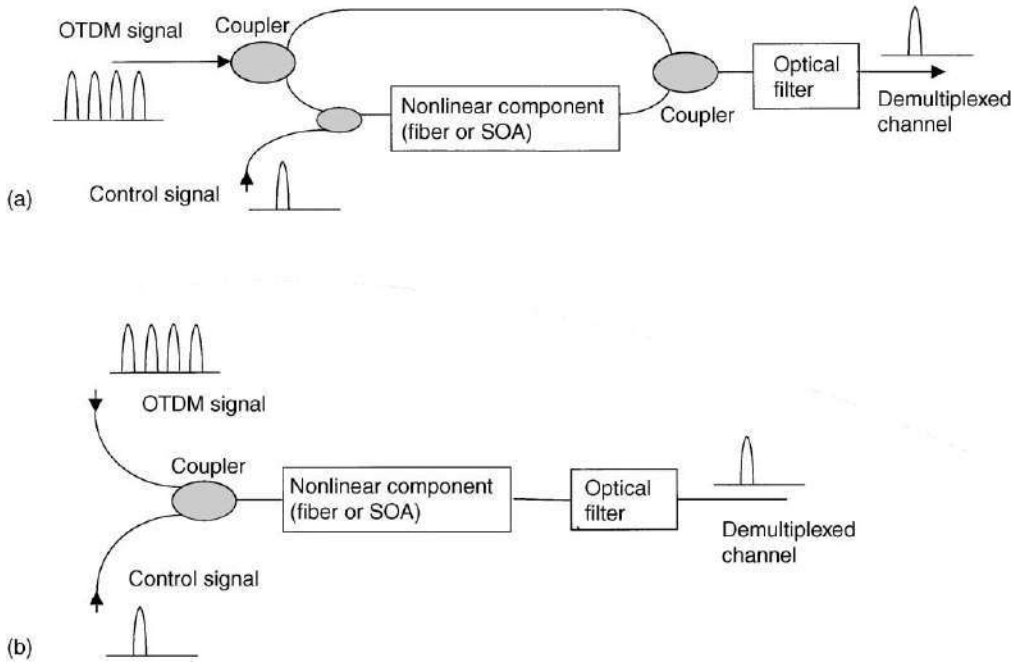

Figure 5 Configuration of OTDM demultiplexers using (a) an interferometer based on nonlinear phase shift in fiber or semiconductor optical amplifiers, and (b) four-wave mixing in fiber or semiconductor optical amplifiers.

synchronized with one of the OTDM data channels, then the phase shift induced by the control on the copropagating signal channel (via the nonlinear refractive index of the fiber) results in that particular channel being switched out to the second input port of the coupler. The remaining data channels of the OTDM signal are once again output through the same port that they entered the coupler. As this particular nonlinear effect in fiber (known as the Kerr effect) has a femtosecond response time, it should be possible to realize extremely high-speed switching with the NOLM. Demultiplexing at data rates in excess of 100 Gbit/s has already been achieved using this technique. One disadvantage of this method is that the nonlinear index coefficient of standard fiber is very small, so in order to achieve the required phase shift from the control pulse, it is necessary to use a fiber loop of around 1 km in length (depending on the power of the control pulse used). This length may affect the overall stability of the demultiplexer and its ability to be readily integrated into high-speed OTDM systems. However, it should be noted that recent developments in the design of high-nonlinearity fibers may reduce this problem.

The second all-optical switch mentioned above is known as a TOAD. Whereas the NOLM is based on the nonlinear refractive index of the fiber, the TOAD is based on a nonlinear optical effect in semiconductor optical amplifiers (SOA). The overall setup for a TOAD is very similar to that for a NOLM, except that a semiconductor amplifier, as shown in Figure 5, replaces the length of fiber in the loop. As for the NOLM, the OTDM data signal is injected into the TOAD using one input of the coupler, such that it splits into two counterpropagating signals, while the control signal propagates unidirectionally in the loop containing the SOA. By synchronizing the control with one channel in the OTDM signal, this particular channel is switched out. Although the response time of a TOAD-based switch may not be as fast as the NOLM, it does have the major advantage that it can be developed into a very compact demultiplexer, suitable for deployment in OTDM-based communication systems.

In addition to the NOLM and the TOAD structures, which are both interferometric-based switches, it is also possible to use four-wavemixing (FWM) in optical fiber or semiconductor optical amplifiers to carry out the demultiplexing of OTDM signals. In this case, the high-power control pulse is synchronized with one of the OTDM channels such that the wavelength of this channel is shifted as it passes through the nonlinear device. Optical filtering may then be used to select out the demultiplexed channel.

As we discussed in the section on demultiplexing, whatever specific all-optical switching technique is employed, we need to generate an optical clock signal for use as the control pulses. This implies that it is necessary to extract a clock at the base rate of the individual data channels from the high-speed OTDM signal. Depending on the base data rate, which may be anywhere from 2.5 to 40 Gbit/s for typical OTDM systems, different clock recovery techniques have been demonstrated. These range from basic electronic clock recovery schemes to more advanced techniques based on self-pulsating laser diodes and optical phase-locked loops.

OTDM Networking Issues

The previous section has predominantly dealt with the basic elements required to implement a high-speed OTDM transmission link. However, if we are to use OTDM technology for high-speed networking, then additional elements will be required. One of the most important of these is an add/drop multiplexer (ADM). The ADM allows us to switch out one channel from the OTDM signal, and then insert another data channel for transmission in the vacant bit slot. If we consider employing OTDM in a ring configuration network, then a typical architecture for the network will be as shown in Figure 6. The high-speed OTDM data signal propagates around the closed fiber loop, and each node in the network is responsible for switching out the data it requires, in addition to inserting the data it wishes to transmit onto the fiber ring. Add/drop multiplexing is thus imperative in developing such an OTDM ring network.

The configuration for each node in an OTDM ring network is shown in Figure 7. At each node it is initially necessary to recover the base rate clock signal. The recovered clock is then used in conjunction with the OTDM signal to switch out the required data channel from the multiplexed signal. When an add/drop multiplexer is used for this purpose we obtain two outputs, one being the demultiplexed data, and the other being the OTDM signal less the switched-out data channel. A simple fiber coupler can subsequently be used to allow this particular node to insert the data it wishes to send into the free bit slot of the overall OTDM signal. This obviously requires control of the relative delay between the OTDM signal and the local data channel to be inserted onto the fiber ring, to ensure that the local data are inserted into the vacant bit slot. A number of techniques have been used for implementing add/drop multiplexers in

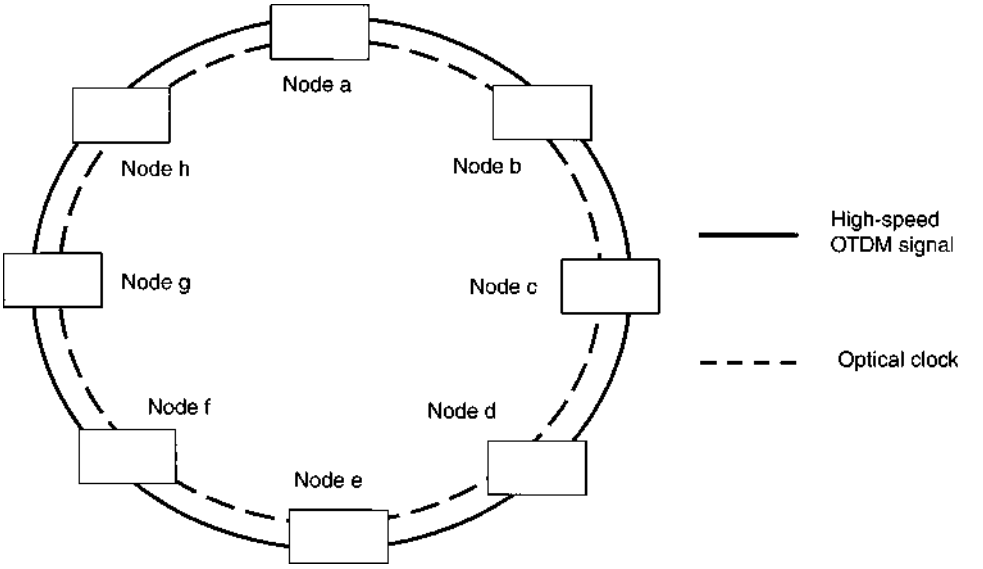

Figure 6 Schematic representation of a ring network employing OTDM with a separate fiber for clock distribution.

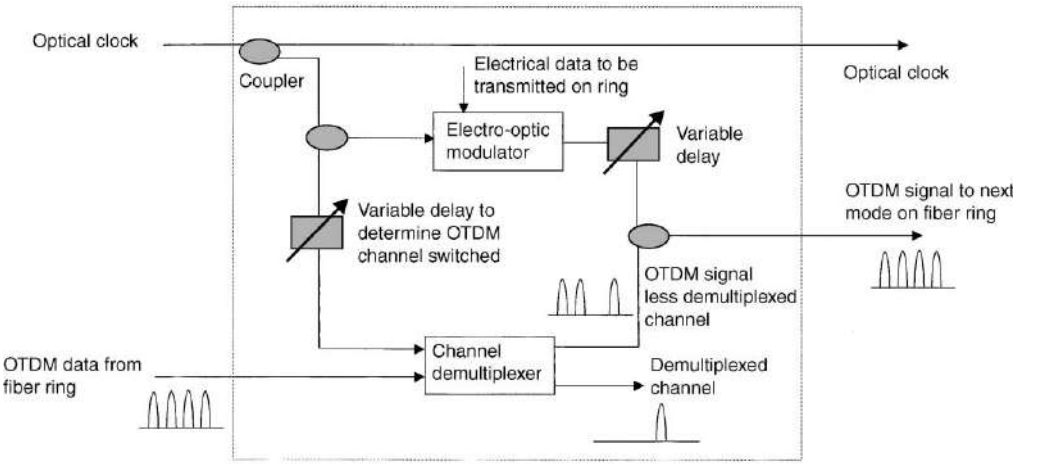

Figure 7 Configuration of an add/drop multiplexing node for use in OTDM-based communication systems.

OTDM-based networks. These have used either fiber interferometers, or interferometers using semiconductor optical amplifiers, in which we obtain both the demultiplexed data signal and the OTDM signal less the switched-out channel (to which the local data channel is then added). The continuing development and enhancement of ADM devices for OTDM systems will be vital for the future implementation of OTDM communication systems.

Conclusion

In this article we have explained the main ideas involved in building OTDM communication systems, and also introduced the principal technologies which are required to do so. OTDM is an extremely wavelength-efficient technique for delivering high-capacity data signals, and it may be used both for long-haul transmission and networking around metropolitan areas. In addition, unlike WDM, it does not require a different wavelength source for each channel in the multiplexed signal. Despite these and other advantages, OTDM is still way behind WDM when it comes to commercial maturity, with no OTDM systems currently available in the telecommunications market. However, the technologies required to implement high-speed OTDM systems (outlined in Table 2), including ultrashort pulse sources, clock extraction circuitry, and demultiplexing devices, are developing rapidly. It therefore seems

Table 2 Main elements of an OTDM communications system

Device	*Principal function*	*Technical implementation*
Optical pulse source	Generate ultrashort optical pulses suitable for transmission in high-speed OTDM systems	Gain-switching, mode-locking, or external modulation techniques
Multiplexer	Multiplex individual pulse data channels in the temporal domain using appropriate delays to achieve an OTDM signal	Passive couplers with fixed fiber delay lines, or planar waveguide circuits (for accurate control of delays)
Optical amplifier	Amplify the OTDM signal during fiber transmission to overcome fiber loss	Fiber doped with rare earth materials (e.g. erbium) in conjunction with pump laser
Transmission fiber	Transmit the OTDM data signal from transmission site to required receiver	Standard single-mode fiber, dispersion shifted fiber, or other specialty fiber may be used
Clock recovery	Extract clock (at base rate of individual channels) from OTDM data signal at receiver, for use in demultiplexing	Electronic or optical phase-locked loops, self-pulsating lasers, or mode-locked ring laser techniques
Demultiplexer	Demultiplex one of the channels from the overall OTDM signal	Interferometric or four-wave mixing techniques based on fiber or semiconductor nonlinearity
Add/drop demultiplexer	Extract one channel from OTDM data signal, and insert a local channel into the empty time slot for transmission	Similar technical implementations to demultiplexer described above

likely that in the future, OTDM technology may be used for enhancing the overall capacity of optical communication systems.

One possibility for employing OTDM systems would be for upgrading installed WDM networks. The current method for improving WDM networks involves using more wavelength channels, but it may also be feasible to increase the overall capacity by developing hybrid WDM/OTDM communications systems. In such a system, each wavelength channel may carry aggregate data rates in excess of 100 Gbit/s, achieved using OTDM. The design of these hybrid networks would permit switching of the optical data signals to be carried out at two different levels in the overall network. The WDM signals could be switched coarsely using passive filtering devices, and the OTDM data channels could be switched with fine granularity using one of the all-optical switching techniques available.

List of Units and Nomenclature

Dispersion	[ps km^{-1} nm^{-1}]
Transmission data rate	[bit s^{-1}]
ADM	Add drop multiplexer
CW	Continuous wave
FWM	Four-wave mixing
NOLM	Nonlinear optical loop mirror
OTDM	Optical time division multiplexing
SDH	Synchronous digital hierarchy
SOA	Semiconductor optical amplifier
STM	Synchronous transport module
TOAD	Terahertz optical asymmetric demultiplexer
WDM	Wavelength division multiplexing

See also

All-Optical Signal Regeneration. **Optical Communication Systems:** Basic Concepts; Wavelength Division Multiplexing.

Further Reading

Agrawal GP (1997) *Fibre Optic Communications*, 2nd edn., pp. 329–335. New York: Wiley.

Cole M (2000) *Introduction to Telecommunications: Voice, Data, and the Internet*, chapter 5. Englewood Cliffs, NJ: Prentice Hall.

Islam M (1992) *Ultrafast Fiber Switching Devices and Systems*. Cambridge, UK: Cambridge University Press.

Kawanishi S (1998) Ultrahigh-speed optical time-division-multiplexed transmission technology based on optical signal processing. *IEEE Journal of Quantum Electronics* 34: 2064–2079.

Midwinter JE and Guo YL (1992) *Optoelectronics and Lightwave Technology*, pp. 148–178. Chichester, UK: Wiley.

Mukherjee B (1997) *Optical Communications Networks*, pp. 493–506. New York: McGraw-Hill.

Nakazawa M, Kubota H, Suzuki K, Yamada E and Sahara A (2000) Ultrahigh-speed long distance TDM and WDM soliton transmission technologies. *IEEE Journal of Selected Topics in Quantum Electronics* 6: 363–396.

Prucnal PR, Santoro MA and Sehgal SK (1986) Ultrafast all-optical synchronous multiple access fiber networks.

IEEE Journal of Selected Areas of Communication 4: 1484–1493.
Spirit DM, Ellis AD and Barnsley PE (1994) Optical time division multiplexing: systems and networks. *IEEE Communications Magazine* 32(12): 56–62.
Spirit DM, Blank LC and Ellis AD (1995) Optical time division multiplexing for future high capacity network applications. In: Smith DW (ed.) *Optical Networking Technology*, pp. 54–77. London: Chapman and Hall.
Tucker RS, Einstein G and Korotky SK (1988) Optical time-division multiplexing for very high bit rate transmission. *IEEE Journal of Lightwave Technology* 6: 1737–1749.

Wavelength Division Multiplexing

J Bowers and H F Chou, University of California, Santa Barbara, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Wavelength division multiplexing (WDM) has enabled a revolution in communications technology. This article describes the technology, critical components of WDM systems, and transmission impairment and limits to transmission distance and capacity.

Multiplexing

There are three techniques used in fiber optics to multiplex the information that is transmitted from one point to another. The first is space division multiplexing (SDM), which simply consists of having multiple fibers in a cable or conduit. In a typical installation, there may be between 8 and 1000 fibers in parallel. The second approach is time division multiplexing (TDM) (*see* **Optical Communication Systems:** Optical Time Division Multiplexing), which consist of interleaving multiple data streams into one higher bit rate stream. For example, 672 voice channels (DS0) at 64 kbit/s are multiplexed into one T1 data stream at 1.544 Mbit/s. Eighty-four of these T1 channels are then multiplexed into a SONET (Synchronous Optical NETwork) OC-3 signal at 155 Mbit/s. The third multiplexing approach is wavelength division multiplexing (WDM), which consists of sending multiple signals at different wavelengths. These signals can be used to fill up the entire low loss transmission band of an optical fiber (Figure 1). The low loss transmission band extends from 1200 nm to 1600 nm, which corresponds to over 40 THz. One approach is to pack the signals densely together, as in dense wavelength division multiplexing (DWDM). This is a relatively expensive, but very high capacity approach. The other approach is to space the signals widely apart, (coarse WDM or CWDM) resulting in very low cost components, but with fairly limited transmission capacity.

Dense Wavelength Division Multiplexing (DWDM)

Most DWDM systems use multiple beams spaced at 100 GHz spacing centered at 193.1 GHz as defined by an International Telecommunications Union (ITU) standard (Figure 2a). Some higher capacity systems use finer spacing, 50, 25, or even 12.5 GHz spacing to pack more information into the fiber. The 100 GHz spacing used in DWDM systems corresponds to 0.8 nm spacing at 1.55 mm. In principle, 40 THz of information can be sent down an optical fiber. If high spectral efficiency (~1 bit/Hz) is maintained across the band, then this would correspond to 40 Tbit/s. Practical DWDM systems are limited by the bandwidth of erbium-doped fiber amplifiers (EDFAs) and by signal impairments discussed below. Consequently, most systems utilize a single band, typically the C-band extending from 1530 nm to 1565 nm. Additionally some transmission systems use multiplexers to combine other bands, typically C and L bands for additional transmission capability. The wavelength definitions are shown in Figure 1. Subbands are often used within a band to facilitate adding and dropping a group of wavelengths within a band. Typically 40 10 Gbit/s signals spaced at 100 GHz are used to fill the C-band. Higher capacity

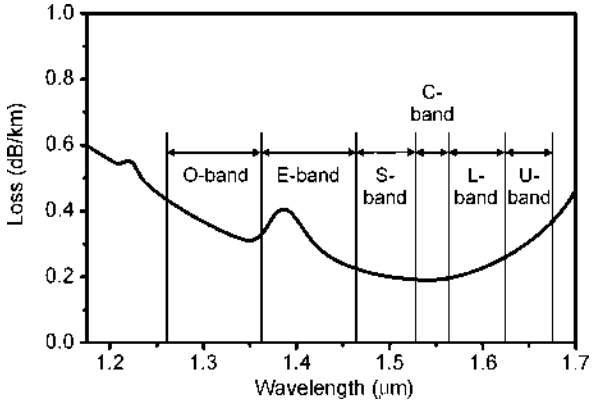

Figure 1 Transmission spectrum of optical fiber. Also shown are the standard transmission bands.

systems with 80 to 200 wavelengths have also been deployed commercially.

The historical evolution of optical transmission systems can be seen in Figure 3. The lower line shows the increasing bit rate demonstrated in experimental TDM systems, which began in 1978 with 45 Mbit/s systems and has now reached 1.28 Tbit/s. The upper line shows what has been demonstrated with DWDM with over 10 Tbit/s transmission demonstrated.

Coarse Wavelength Division Multiplexing (CWDM)

The fine frequency requirements of DWDM systems result in significant cost for transmitters, multiplexers and other components. A low-cost approach is to space the colors widely apart. The ITU standard is 20 nm spacing from 1271 to 1611 nm (Figure 2b). Note that the spacing is uniform in wavelength, but varies in frequency because the frequency spacing is related to the wavelength spacing by

$$\Delta f = -\frac{c}{\lambda^2}\Delta\lambda \qquad [1]$$

The wide spacing of signal wavelengths means that temperature stabilization of elements is not needed and the yield of components is high, resulting in a lower-cost system.

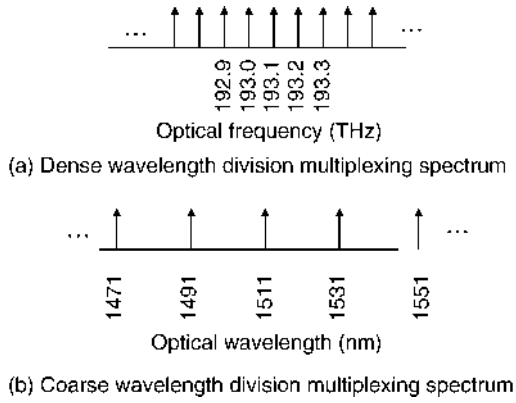

Figure 2 (a) DWDM spectrum (b) CWDM spectrum.

Point-to-Point DWDM Links

Point-to-point DWDM links were widely deployed during the 1990s. The essential elements are shown in Figure 4, namely single frequency lasers, optical multiplexers and demultiplexers, optical amplifiers, optical add drop multiplexers (OADM), and optical receivers. These will be described in more detail in subsequent sections. Essential parts of the design of these links are managing the signal-to-noise ratio and managing the pulse broadening. With periodic amplification in EDFAs and dispersion compensation, links of 1000 km are common. The widespread use of forward error correction (FEC) circuits allows error-free transmission over a link with limited signal-to-noise ratio. With careful design, ultra-long-haul

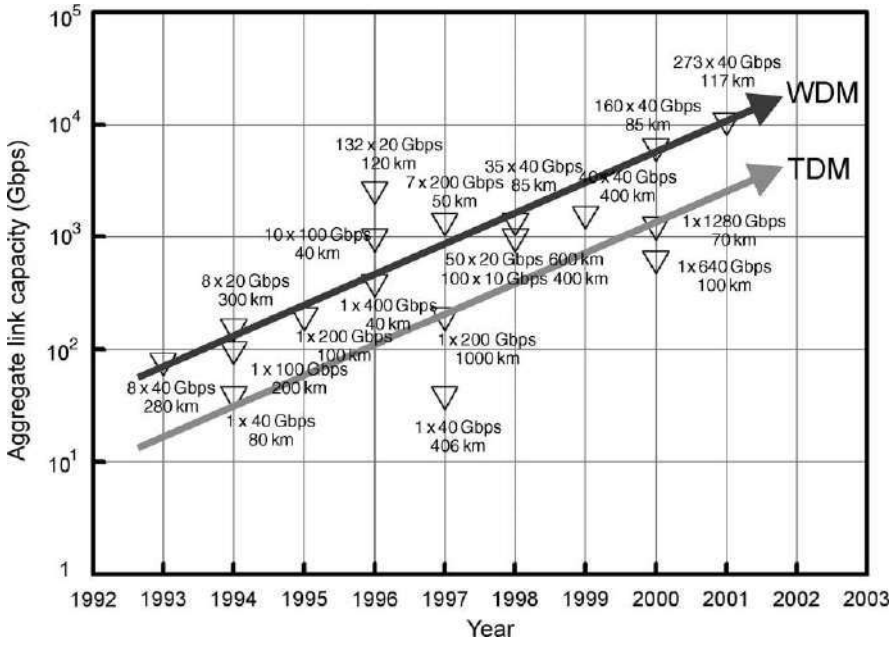

Figure 3 Historical evolution of transmission systems.

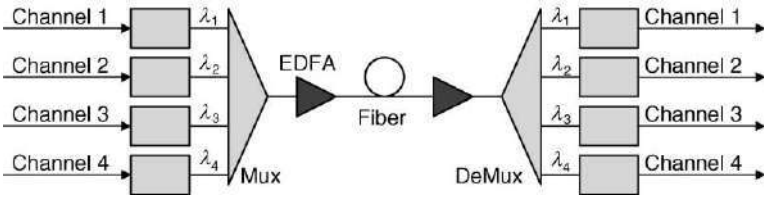

Figure 4 Point-to-point WDM link.

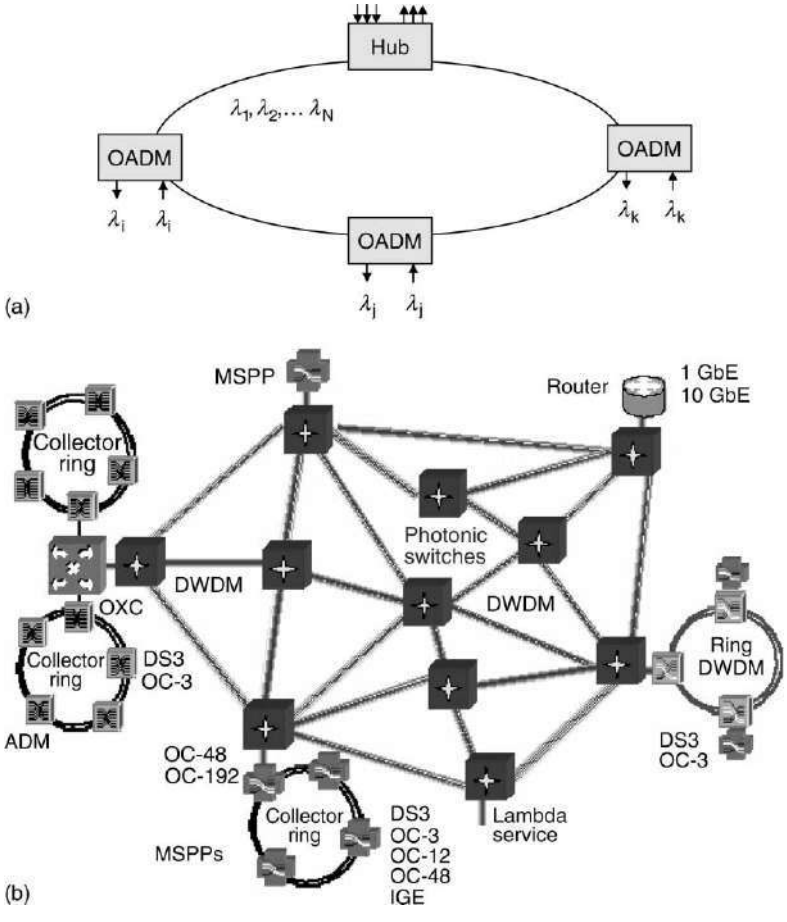

Figure 5 (a) WDM ring network. (b) DWDM mesh networks (after Basch B, Gringeri S, Goudreault R and Ravinkumar S (2002) Evolution of the photonic core within metropolitan and regional networks. *NFOEC*. Technical Proceedings).

transmission over 5000 km is possible, even with many 10 Gbit/s channels.

DWDM Networks

Point-to-point DWDM links are widely deployed in opaque networks with electronic regeneration and switching at nodes. The next stage in optical networks is the deployment of DWDM transparent ring and mesh networks. An example of a simple ring network is shown in Figure 5a where most of the light signals transit nodes transparently. The OADM elements drop particular wavelengths at each node. As more wavelengths are deployed in these ring networks, it becomes necessary to dynamically reallocate wavelengths and resources. Reconfigurable optical add/drop multiplexers (ROADMs) and optical switches have been developed to allow for dynamic transparent mesh networks (Figure 5b). This allows service providers to provision their networks remotely and quickly adapt to changing demand and restore vital connections when service interruptions occur.

Critical WDM Components

The previous sections described WDM networks. This section describes in more detail the critical components used in these networks.

Multiplexers/Demultiplexers

Multiplexers (MUXs) are used to combine multiple wavelength channels into a single fiber. They are usually wavelength selective and utilize wave phenomena such as diffraction and interference. A simple cascade of 3 dB couplers is not a practical multiplexer since the output power decreases exponentially with channel count. Due to the reciprocity of electromagnetic waves, most MUXs can also serve as a DEMUX. For high channel count systems, a multistage design may be necessary to overcome the limitations of a single MUX/DEMUX. This approach also provides a more flexible upgrading plan. In this case, wavelength channels can be grouped into bands or interleaved. Diffraction gratings and arrayed waveguide gratings (AWGs) are among the available technologies for MUX/DEMUX.

In a diffraction grating, narrow slits or micro structures are spaced evenly on the grating plane and serve as a secondary light source for the input light. The transmitted or reflected light (depending on the design) forms a diffraction pattern on the image plane. The locations of peak intensity on the image plane depend on the wavelength and the grating pitch. Therefore, wavelength channels can be separated in space and fibers can be positioned properly to couple individual channels. An AWG is a planar integrated and digitized form of diffraction grating as shown in Figure 6. It is usually fabricated on a silicon substrate with silica waveguides (silica-on-silicon). The input light is coupled into an array of waveguides with incremental difference in length. These tributaries are then combined in a free-propagating output coupler to several output waveguides. Due to the interference of these tributaries from the arrayed waveguides, different wavelength channels are focused to different output waveguides.

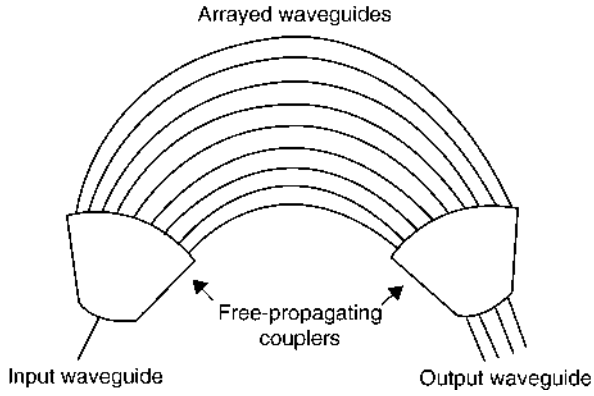

Figure 6 Schematic diagram of an AWG.

Optical Add/Drop Multiplexer

An optical add/drop multiplexer (OADM) provides critical networking capability for WDM systems. In an OADM, some of the wavelength channels are dropped from the input and are either terminated or routed. New channels at these wavelengths are then added with the through channels. This function can be done with the combination of a MUX and a DEMUX of any kind as shown in Figure 7a. However, if only one wavelength channel needs to be processed while others pass through, a simpler configuration can be used where a circulator drops the reflected wavelength from the fiber Bragg grating and a new channel is added with a coupler (Figure 7b).

The two approaches described above are termed static since the wavelength processed is fixed. To enable a more flexible network, reconfigurable OADM (ROADM) are required, where the wavelengths to be added or dropped can be dynamically specified. Figure 7c shows a configuration of ROADM based on a high port-count optical switch and a set of MUX/DEMUX. A tunable laser is highly desirable in this scenario since the wavelength of the added channel can be reconfigured to resolve wavelength contention in the network.

Optical Amplifiers

Erbium-doped fiber amplifiers (EDFA) (*see* **Optical Amplifiers:** Erbrium Doped Fiber Amplifiers for Lightwave Systems) provide simultaneous amplification of wavelength channels in the entire C-band (1530 to 1565 nm) and part of the L-band (1565 to 1605 nm). EDFAs were a key enabling technology for the deployment of WDM systems, especially for long-haul transmission where costly regenerators were eliminated. Both 980 nm and 1480 nm pumps can be utilized to achieve gain in erbium-doped fibers. The 980 nm pumps provide low-noise and high-gain operation while the 1480 nm pumps are more favorable for high-output power. One of the most significant advantages of an EDFA over a semiconductor optical amplifier (SOA) is that crosstalk among wavelength channels is much smaller in EDFA due to a millisecond-scale spontaneous emission lifetime. One critical issue for EDFAs is that the gain is not flat over the operation wavelength range and peaks around 1532 nm. However, this can be equalized by specially designed filters.

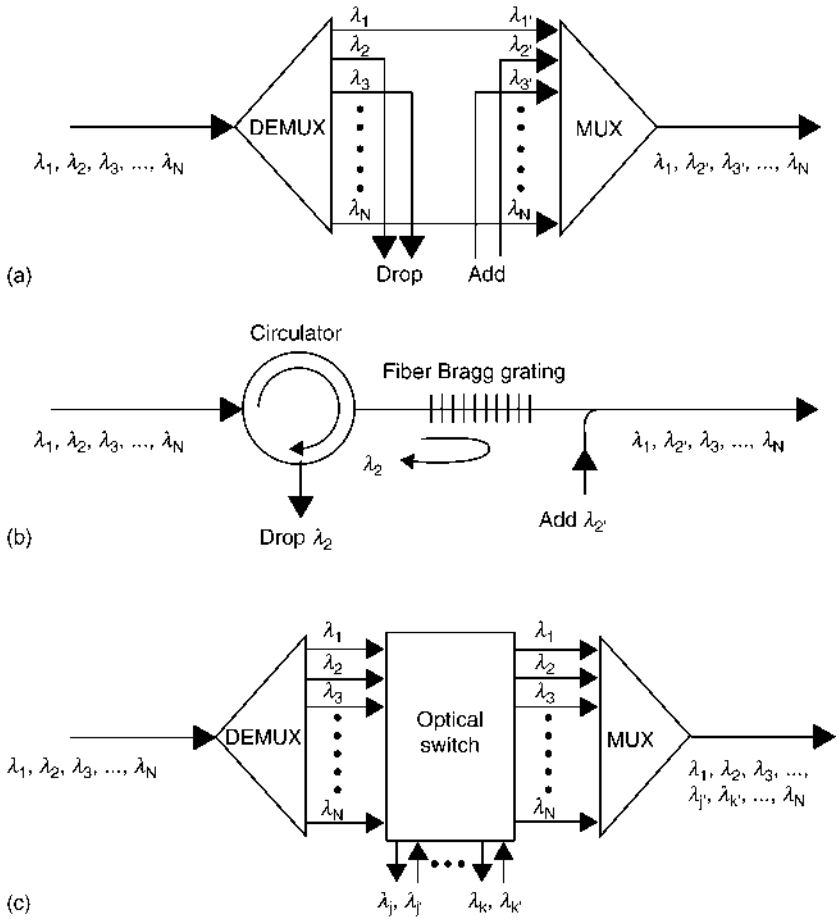

Figure 7 (a) A parallel OADM architecture which allows multiple simultaneous add's and drop's; (b) a linear OADM architecture for processing a single channel; (c) an ROADM which enables reconfigurable processing of multiple wavelength channels.

Outside the gain spectrum of EDFAs, Raman amplifiers can be used, where the gain peak is about 13 THz (100 nm) below the frequency of the pump signal. Also, Raman amplification occurs in the same fiber used for transmission so that it can provide a distributed gain over the span of the fiber, which results in a better signal-to-noise ratio. This is particularly important for ultra-long-haul transmission. A counter-propagating pump scheme is preferable to reduce the influence of pump noise and depletion. The biggest challenge for Raman amplifiers is the availability of high-power pump lasers at the required wavelengths.

Optical Filters

Filters can find many applications in WDM systems, such as in multiplexers, demultiplexers, OADM, receivers and optical amplifiers. Fiber Bragg grating, Fabry–Perot filter, and multilayer thin-film filters are among the most widely used filters and utilize the interference property of optical waves. There are some general requirements for filters in WDM systems. First, the top of the pass-band must be flat enough so that cascading of filters in the network does not lead to a significant decrease of bandwidth. In addition, the roll-off of the pass-band must be sharp so that cross-talk from other channels is minimized.

A fiber Bragg grating has a periodic perturbation of refractive index made by exposing a photosensitive fiber to interfering ultraviolet lights. These perturbations reflect the input light as it propagates. At a specific wavelength, these reflected beams add in phase, which results in a strong overall reflection. The reflections at other wavelengths do not add up and these wavelengths will eventually pass through the grating. In short, the fiber Bragg grating acts as a wavelength-selective reflector and is usually used with a circulator or isolator to manage the reflected light. A Fabry–Perot filter is composed of two partially transmitting interfaces and a cavity in between. The underlying mechanism is similar to the fiber Bragg grating but the transmission is periodic in optical frequency. The spacing is called

the free spectral range, which depends on the optical length of the cavity. The bandwidth of the passband is a strong function of the reflectivity of the partially transmitting interfaces. A multilayer thin-film filter is an extended form of Fabry–Perot filter where several layers of dielectric reflectors are formed with dielectric cavities in between. This allows for more tuning dimensions in the design of the passband shape in order to meet the requirements of WDM systems and the temperature tolerance is better than other kinds of filter.

Laser Sources

Semiconductor lasers (*see* **Lasers:** Semiconductor Lasers) are widely used in optical communication because they are compact, low cost and efficient. As a light source for transmission, single frequency operation is necessary for high-capacity WDM systems. Multilongitudinal mode lasers such as Fabry–Perot lasers have excess bandwidth which causes excessive penalty due to dispersion unless the transmission distance is short. They also result in crosstalk in DWDM systems. The most widely used single-frequency laser in WDM systems is the distributed-feedback laser (DFB), where a corrugated grating structure is fabricated on top of the gain region and only the wavelength which satisfies the Bragg condition can oscillate in the laser cavity and all other modes are suppressed.

As WDM networks advance, tunable lasers are strongly desired to reduce the excessive inventory of fixed wavelength lasers for DWDM and also to provide reconfigurability in the network. One possible structure is the distributed Bragg reflector (DBR) laser, in which the Bragg grating is decoupled from the gain region and the Bragg wavelength can be tuned independently by isolated current injection. More complicated variations of this concept such as the three-section DBR laser and the sampled grating DBR laser can provide more continuous and wider tuning in wavelength.

Photonic Switches

Photonic switches are necessary in transparent networks to allow reconfigurability in networks, and allow rapid provisioning and fault recovery. They provide connectivity between rings and between point-to-point links without requiring the DWDM signal to be broken into individual wavelengths and to be electronically regenerated. Photonic switches can be used to switch fibers, wavebands, or individual wavelengths. They are inherently transparent to bit-rate and data format. A number of different technologies have been developed.

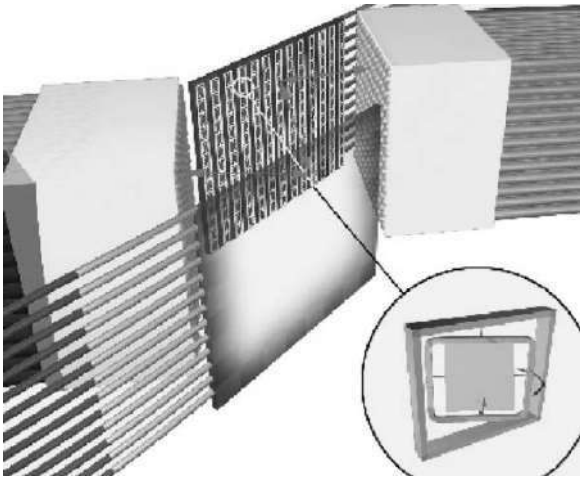

Figure 8 3D MEMS optical switch.

Microelectromechanical system (MEMS) switches are widely used for 2×2 protection switches, wavelength selective switches and for nonblocking crossbar switches with sizes from 8×8 to 1024×1024. The technology usually used for large switches is called 3D MEMS switches (Figure 8), in contrast to 2D technologies where the optical beams are confined to a two-dimensional plane.

Transmission Issues

Chromatic Dispersion

Chromatic dispersion (*see* **Fiber and Guided Wave Optics:** Dispersion) means that the group velocity, which is the propagation speed of an optical signal, changes with wavelength. Since every real pulse has a finite bandwidth, each slice of wavelength within this bandwidth propagates with a different speed and results in the distortion of the pulse as it propagates along the fiber. Depending on the sign of the dispersion and the initial phase distribution, the pulse can either be broadened or compressed. However, if the fiber is long and nonlinear effects are small, the pulses will eventually be broadened and result in intersymbol interference. Dispersion compensating fibers (DCFs), which have an opposite sign of dispersion, can be used to reverse this process and restore the original pulse shape. In general, it is advantageous to minimize the length of DCF by keeping the dispersion low. Dispersion-shifted fibers (DSF) have a zero-crossing of dispersion around 1.5 μm and hence a reduced dispersion for the C-band.

For a WDM system, a very low dispersion will result in more severe nonlinear impairments since the distance pulses can overlap with each other along the fiber is long. Nonzero-dispersion-shifted fibers

(NZ-DSF), which introduce a suitable amount of dispersion, can be used instead to solve this issue. Chirped fiber Bragg grating can also be used, where different wavelengths are reflected at different positions in the grating so as to compensate the dispersion caused by transmission. It is capable of compensating a given amount of dispersion with a much shorter length when compared to DCF. For high channel count WDM systems, the dispersion experienced by each channel may be different and dispersion compensation needs to be implemented on a per-channel basis after a demultiplexer.

Polarization Mode Dispersion

Polarization mode dispersion (PMD) means that the group velocity dispersion of the two orthogonal polarizations in the fiber are different and cannot be simultaneously compensated by a single DCF. It is usually manifested as the splitting of a pulse at the receiver even after chromatic dispersion compensation and causes intersymbol interference. PMD can originate from the imperfections in the fiber or environmental factors like temperature and mechanical pressure. Therefore, the amount of PMD can vary with time for a given link, which means that dynamic compensation is necessary. Traditional electrical equalization after the optical signal is detected can be used for compensation but the highest bit-rate achievable is limited by the speed of electronics. Optical compensation is another choice. The optical signal is split into fast and slow components and the fast component is delayed properly before being recombined with the slow component.

Nonlinear Impairments

The most important nonlinear impairments for WDM systems are the cross-phase modulation (CPM) and the four-wave mixing (FWM). They are both third-order nonlinear effects, where the nonlinear dielectric polarization is proportional to the third power of the electric field of the signal. In CPM, the phase of a wavelength channel can be modified by other channels, which increases the chirp and results in more dispersion penalty. On the other hand, FWM can generate new waves at other wavelengths, causing unwanted crosstalk if the generated wave happens to coincide with another channel. The existence of dispersion in fiber can help to reduce the effects of both impairments since the pulses at different wavelengths propagate at different speeds and walk away from each other so that nonlinear interactions are reduced. This is why NZ-DSF is preferred over DSF. Self-phase modulation (SPM) is another third-order nonlinearity, where the phase distribution of a pulse is altered by its intensity profile. It will lead to the distortion of the pulse. Other nonlinear effects include stimulated Brillouin scattering and stimulated Raman scattering. Both originate from the interactions of phonons and photons in the fiber and cause power fluctuations at different wavelengths.

Forward Error Correction

To reduce the errors of a transmission link, extra information relating to the data can be sent by the transmitter and used by the receiver to detect and correct errors in the received data. This technique is called forward error correction (FEC) and imposes redundancy to the transmission which means that the bit-rate transmitted will be higher than the actual data rate. A measure of performance for FEC is the coding gain, which is defined as the decrease in receiver sensitivity at a given bit-error-rate. Reed–Solomon codes are one of these codes. With a 7% redundancy, up to 6 dB of coding gain can be achieved. This could correspond to more than 9 orders of magnitude improvement in bit-error-rate. Therefore, the use of FEC can increase the spacing between regenerators in ultra-long-haul transmission and reduces the cost. It is also useful to overcome error floor problems due to crosstalk in a WDM system, where a further increase in received power cannot reduce the bit-error-rate. Crosstalk among wavelength channels can happen due to nonlinear impairments during fiber transmission or due to imperfect channel isolation in components like multiplexers, demultiplexers, filters, and switches. The use of FEC can greatly increase the system margin.

Conclusions

WDM technology has enabled a huge increase in the transmission capacity of optical fibers and a huge reduction in the cost of fiber optic transmission since the cost of the installed fiber is shared between the signals traveling on different wavelengths. As new components have been developed, WDM technology has expanded simple point-to-point links to transparent networks. Great progress has been made over the past 20 years to solve transmission impairments and extend the reach and bandwidth of these systems. This is likely to continue to meet the increasing communication needs of our society.

List of Units and Nomenclature

Distance (km, m, μm)
Bit rate (bit/s)
Wavelength (nm)

See also

Color and the World. **Information Processing:** All-optical Multiplexing/Demultiplexing. **Fiber and Guided Wave Optics:** Optical Fiber Cables; Passive Optical Components. **Lasers:** Edge Emitters; Semiconductor Lasers. **Modulators:** Modulation and Demodulation of Optical Signals. **Optical Amplifiers:** Basic Concepts; Optical Amplifiers in Long-Haul Transmission Systems; Semiconductor Optical Amplifiers; **Optical Communication Systems:** Architectures of Optical Fiber Communication Systems; Basic Concepts; Historical Development; Lightwave Transmitters; Optical Time Division Multiplexing. **Solitons:** Soliton Communication Systems.

Further Reading

Agrawal GP (2002) *Fiber Optic Communication Systems*, 3rd edn. New York: John Wiley & Sons Inc.

Basch B, Gringeri S, Goudreault R and Ravinkumar S (2002) Evolution of the photonic core within metropolitan and regional networks. *NFOEC*. Technical Proceedings.

Green PE (1993) *Fiber Optic Networks*. Upper Saddle River, NJ: Prentice Hall.

Ramaswami R and Sivarajan KN (2002) *Optical Networks: A Practical Perspective*, 2nd edn. San Francisco, CA: Morgan Kaufmann Publishers.

OPTICAL MATERIALS

Contents

Color Filter and Absorption Glasses
Heterogeneous Materials
Lightweight Mirrors
Measurement of Optical Properties of Solids
Optical Glasses
Plastics
Sculptured Thin Films
Smart Optical Materials
Sol-Gel Materials

Color Filter and Absorption Glasses

J E Shelby, Alfred University, Alfred, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical filter and absorbing glasses can be categorized by either the optical behavior of the filter, for example, bandpass, shortpass or longpass, neutral density, or interference filters – or by the method used to filter the light, i.e., glasses which absorb light themselves and those which serve as high quality, transparent substrates for coatings which either absorb or reflect light. The term 'filter' usually refers to glasses used in optical systems, to either transmit or prevent transmission of specific wavelengths. The term 'absorber' is usually reserved for glasses which are used for controlling light to provide protection from damage to humans, for example, for eye protection, as in sunglasses, or welder and laser goggles, or to materials exposed to light, for example, plastics and other components in the passenger compartment of automobiles. Manufacturing technology is normally based on the method used to produce the device, such as control of absorbing species in the glass or method of application of optical coatings. Terminology for optical applications is almost always based on optical behavior and not on the method used to produce that behavior.

Bandpass Filters

The first glasses used to filter light were colored by inclusion of ions from the first row of transition metals. These colored glasses still provide most of

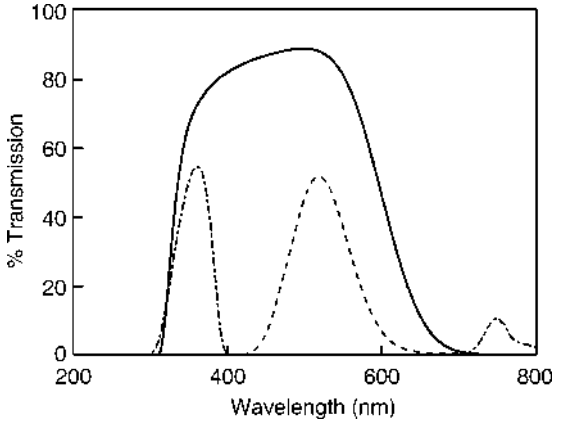

Figure 1 Transmission spectra for bandpass filters with a large bandwidth (solid line), medium bandwidth (dashed line), and narrow bandwidth (dot dash line).

the glasses used to control the wavelength of light. They are widely used as absorbers in windows in buildings and automobiles, to control the entrance of ultraviolet and infrared light into these environments. These filters allow the passage of light over a significant range of wavelengths, while absorbing wavelengths outside that range. They are designated by the peak transmission wavelength, or central wavelength, and the bandwidth, or wavelength range, over which light is transmitted with at least some designated transmission value. Filters based on absorption by transition metal or rare earth ions usually have larger bandwidths and somewhat diffuse cutoffs at the edges of the transmitted spectral region (Figure 1).

The most common bandpass filters absorb radiation in a desired wavelength region. Historically, the first bandpass filters, which are still the most common, relied on absorption of light as a result of ligand field (also known as crystal field) effects and/or charge transfer processes within a glass, to create absorption bands in the spectrum.

Ligand field effects will produce optical absorption bands in glasses which contain transition metal (usually row 3d) or rare earth (lanthanide, or 4f) ions. The differences in the electronic structures of the transition metal and rare earth ions result in very different absorption spectra. Absorption spectra, due to transition metal ions, are characterized by broad bands and are frequently accompanied by charge transfer bands, due to the same ions, while the bands due to absorption of light by rare earth ions are very sharp and are not accompanied by charge transfer absorption bands.

Iron oxide was probably the first, and is still the most widely used additive for production of glasses which primarily transmit light in the visible portion of the spectrum. A large, broad absorption band due to ferrous ions is centered at ≈1050 nm in common soda-lime-silica glasses. The tails of this band extend into the visible portion of the spectrum, allowing passage of the blue-green portion of the spectrum, but absorbing much of the red light, and to wavelengths in excess of 2000 nm, absorbing much of the near infrared spectrum.

Ferric iron produces relatively small ligand-field absorption bands in the region from 380 to 440 nm. These bands remove most of the blue part of the spectrum, allowing passage of light with longer wavelengths. Both ferrous and ferric ions also display charge transfer bands in the ultraviolet, effectively eliminating transmission of light with a wavelength <300 nm. Glasses containing enough iron oxide, to effectively eliminate both ultraviolet and near infrared transmission, exhibit various shades of green, depending on the ratio of ferrous to ferric ions present. Melting under strongly reducing conditions yields blue-green glasses, whereas melting under strongly oxidizing conditions yields yellow-green glasses. Glasses containing ≧0.5 wt% iron oxide are used in windows of automobiles to minimize the amount of ultraviolet and infrared radiation entering the passenger compartment.

Copper oxide is also used to remove a portion of the near infrared spectrum from sunlight and to produce glasses which transmit in the green and blue region. The primary ligand-field absorption band of copper is due to cupric ions and is located near 800 nm. This band in similar in shape to that due to ferrous ions, with tails extending into the visible and near infrared regions of the spectrum. The shift in band position toward the visible, however, results in less absorption in the infrared and more in the visible than for ferrous iron. As a result, glasses containing copper are more strongly colored than those containing ferrous iron and are less effective filters for infrared light. Glasses colored with cupric ions can be used to produce a transmission window from about 350 to 600 nm in soda-lime-silica glasses. Cuprous ions do not create any significant absorption bands.

Glasses containing hexavalent chromium, which must be produced in a strongly oxidizing environment, effectively remove ultraviolet light between 300 and 400 nm, but allow transmission of ultraviolet light in a narrow range centered around 300 nm. A small tail of the charge transfer band due to Cr^{6+}extends into the visible, producing a yellow-green glass. The presence of a set of much weaker ligand-field bands, centered around 650 nm due to trivalent chromium, also contributes to the green color of these glasses. Combining chromium with copper ions in a glass yields a filter with a small transmission window centered at about 520 nm.

Combination of chromium with iron results in suppression of both hexavalent chromium and ferrous iron, yielding a green glass without either the strong ultraviolet absorption of hexavalent chromium or the strong near infrared absorption of ferrous iron.

Manganese oxide can be used to produce bandpass filters which block most of the visible, but allow transmission of ultraviolet and infrared light. Addition of manganese ions to a glass results in a set of absorption bands with a maximum intensity at about 500 nm and a tail extending to around 900 nm. Since these ions do not absorb in the ultraviolet, the transmission spectra of these glasses has a window between about 300 and 400 nm, which overlaps the transmission window due hexavalent chromium. The absorption bands due to manganese ions are relatively weak, so that concentrations of manganese oxide in excess of 1 wt% are needed to effectively block visible transmission.

Cobalt oxide is the strongest colorant of the oxides of 3d transition metals. A small addition of cobalt oxide to silicate glasses creates strong absorption bands between about 450 and 700 nm. The optical window of these glasses in the ultraviolet is centered near 380 nm and is somewhat broader than that of glasses containing manganese oxide, which have similar spectra. Since cobalt oxide is such a strong absorber, black glasses are easily produced by use of cobalt oxide in combination with other oxides. If cobalt oxide is combined with nickel oxide, the resulting glass retains a bandpass in the ultraviolet, while combining cobalt oxide with iron oxide, which has a complementary absorption spectrum, yields a black glass which also effectively eliminates transmission in the ultraviolet and near infrared.

The optical spectra of glasses containing most rare earth ions are characterized by sharp absorption bands. Some rare earth ions create spectra with many absorption bands, while others display only a few bands. The band structures which result from rare earth ions are not particularly useful for creation of bandpass filters, although Yb^{3+} ions do effectively remove light around 1100 nm, without absorbing light in the rest of the spectrum.

While not commonly used for filters, glasses containing rare earth ions have been used for wavelength calibration standards for uv-vis spectrophotometers. A common wavelength standard called a 'didymium' glass contains a combination of neodymium and prasedymium oxides. Since the optical spectrum from each of these ions contains many sharp bands, these glasses provide known wavelengths for absorption bands over a wide region of the spectrum. Neodymium ions, which have absorption bands near the D line of sodium, are used in goggles for eye protection for glassblowers, to remove the intense yellow light resulting from the sodium in the glass.

Interference Filters

Interference filters are a special case of bandpass filters which are created by application of dielectric films to glass substrates. The dielectric films are deposited on a glass substrate in alternating layers of high and low refractive index materials. The combination of refractive indices and thickness of the dielectric films determines the wavelength regions where constructive and destructive interference occur, i.e., those wavelength regions which will transmit or reflect the incident light. Since the choice of materials will control the indices, production of filters for different wavelengths usually relies on deposition of layers of differing thicknesses.

These filters can be produced with very narrow bandwidths (± 10 nm) or with broader bandwidths (commonly ± 50 or ± 80 nm). The peak transmission frequently does not occur at the midpoint of the bandwidth. Since the wavelengths for constructive and destructive interference depend upon the path length through the dielectric layers, these filters are extremely dependent upon the incident light angle, with a shift toward lower wavelengths for incident angles other than normal to the filter surface. Interference filters typically reduce throughput of light as compared to absorption-based bandpass filters, but provide for much narrower transmission windows. If the dielectric layers are deposited in such a manner that their thickness varies along the length of the substrate, interference filters can be produced with a continually varying central transmission wavelength along the length of the substrate. While convenient for applications, which might otherwise require several filters of different central wavelengths, these linearly variable interference filters are much more expensive than those designed to block specific wavelengths.

Dispersion Filters

Dispersion filters are based on the passage of light through a medium consisting of a matrix material and dispersed, very small particles of a material which has a very different dispersion curve. The original version of these filters, called a Christiansen filter, consists of small glass particles with a controlled particle size suspended in a liquid. The dispersion curve for the glass is much less wavelength-dependent than that of the liquid. Transmission of light occurs at the wavelength where the refractive index of the glass and the liquid are equal, while light of other

wavelengths is scattered. Since the refractive index of a typical liquid is highly temperature sensitive, the transmitted, or peak, wavelength can be shifted over a wide region by controlling the temperature of the mixture. The bandwidth of dispersion filters can be very narrow ($<\pm 10$ nm).

The large temperature dependence of the traditional dispersion filter can be useful in that it provides an optical filter with an adjustable peak wavelength. On the other hand, this sensitivity to temperature also results in a serious control problem, since a temperature change of only a few degrees can shift the transmission window by up to 100 nm. This problem can be decreased or even eliminated by making a solid dispersion filter using two glasses of different dispersions, with one glass distributed in the other, or a glass dispersed in some other transparent solid such as a polymer with a similar thermal expansion coefficient. Since the refractive index of solids tends to vary with thermal expansion coefficient, these solid state dispersion filters can be almost temperature independent.

Neutral Density Filters

Neutral density filters are used to reduce the throughput of light to some desired value over a wide spectral region. Gray glasses are used to reduce the intensity of light from a bright source such as the sun. Black glasses are used to completely block light in the visible region. Neutral density filters can be produced using a combination of transition metal ions or by use of optical coatings. They are characterized by their percent transmission, range of wavelength where the transmission is constant, and the 'ripple', or spectral flatness, within the designated spectral region (Figure 2).

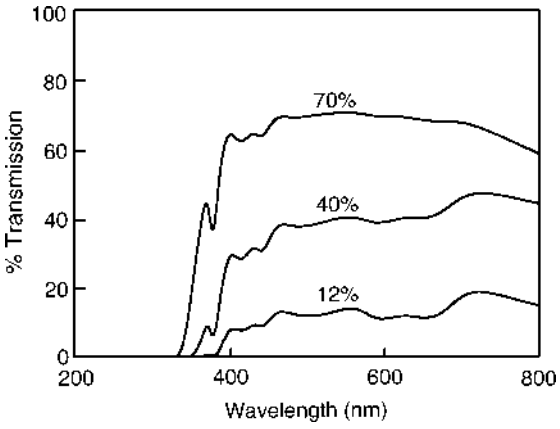

Figure 2 Transmission spectra for neutral density filters with different % transmission.

An approximation to a neutral density filter for visible light is often created by combining the absorptions in the ultraviolet and near infrared from ferrous and ferric ions, with the multiple absorption bands centered in the region of 500 to 600 nm due to divalent cobalt ions. While the resulting absorption spectrum displays small ripples throughout the region from 350 nm to beyond 1200 nm, the transmission of the glass can be controlled to within a few percent by the proper balance of ferric and ferrous ions combined with the desired concentration of CoO. These glasses are generally gray, or, if enough of iron and cobalt oxides are present, black. Combination of CoO with NiO can produce similar filters covering a smaller wavelength range which maintain transmission in the ultraviolet. Filters with different transmission values can be produced by either varying the composition of the glass or by use of different glass thicknesses. A continuously graded neutral filter can be produced from a single piece of glass by etching or polishing to form a wedge. With more effort, a step-graded filter can be produced by etching or polishing to form sharply delineated thickness steps along the length of a glass plate.

Neutral density filters can also be produced by coating a glass substrate with a metal film. Coated neutral density filters have less ripple than absorptive filters based on iron and cobalt oxides dissolved in the glass. In order to produce a constant transmission throughout the desired wavelength region, the metal used (commonly inconel) should have a constant reflectance across the desired wavelength region. Use of a metal with a varying wavelength dependence in its reflectivity, such as silver, will result in a wavelength dependent transmission. Continuously graded neutral filters can be produced by varying the thickness of the film along the length of a strip of glass.

Longpass/Shortpass Filters

Longpass filters are designed to pass light with wavelengths longer than the specified cutoff value. They are frequently used to block ultraviolet light while transmitting visible light or to block visible light while transmitting infrared light. Shortpass filters provide the opposite effect, i.e. they pass light with wavelengths shorter than the specified cutoff value, while blocking light with longer wavelengths. Shortpass filters are often used to block passage of heat (infrared light), while allowing transmission of visible light (Figure 3).

Ultraviolet blocking longpass filters can be produced by varying the basic composition of a glass or

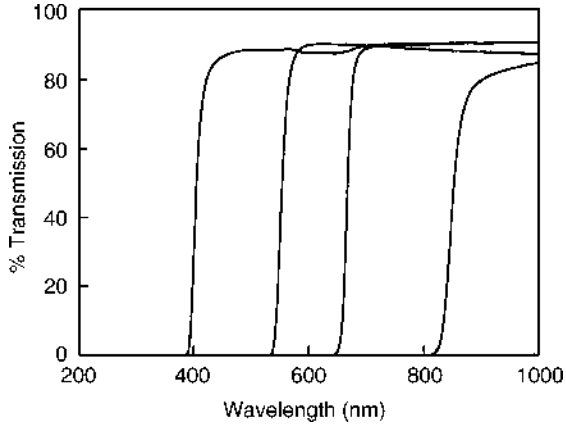

Figure 3 Transmission spectra for longpass filters with different cutoff wavelengths.

by addition of specific ions which absorb strongly in the ultraviolet. Varying the primary glassforming oxide can alter the ultraviolet cutoff frequency of a glass. Vitreous silica, for example, has an ultraviolet cutoff beyond 200 nm, so that it is a very effective transmitter of ultraviolet light. Vitreous germania, on the other hand, has an ultraviolet cutoff near 400 nm, so that it acts as an effective longpass filter for most ultraviolet light. While vitreous silica is widely used as an optical material, the poor chemical durability of vitreous germania prevents its use in most optical applications.

The ultraviolet cutoff wavelength of silicate glasses is shifted toward the visible by addition of alkali and alkaline earth oxides to silica. The addition of these oxides creates nonbridging oxygens in the vitreous network, which shift the ultraviolet cutoff toward the visible. These glasses are effectively opaque beyond their cutoff wavelength, creating longpass filter glasses with cutoff wavelengths which vary with the nonbridging oxygen concentration. Impurities such as iron or copper can also reduce the ultraviolet transmission of silicate glasses due to the creation of charge transfer bands, producing longpass filters with cutoffs in the range of 250 to 300 nm.

Longpass filters with tightly controlled ultraviolet cutoff wavelengths are often made by adding either cerium or a combination of cerium and titanium oxides to silicate glasses. Additions of cerium oxide result in a strong absorption band near 320 nm, which provides protection from ultraviolet light. Cerium also is routinely used to reduce radiation-induced coloration of glasses exposed to high energy radiation. An absorption band due to trivalent titanium also serves to cut off ultraviolet light, while producing a yellow glass. The use of a combination of cerium and titanium produces a very effective ultraviolet filter which eliminates transmission of almost all ultraviolet light.

Long bandpass absorption filters can also be based on precipitation of colloids of metals or semi-conductors in an otherwise colorless glass. The glasses are initially prepared in a colorless state, with the materials which will form the colloids dissolved as ions. The glasses are then reheated into the glass transformation region, where the desired ions are reduced to the atomic state and precipitated as colloids. Since this process is known as 'striking', these filters are known as striking filters. Colloids can be produced by striking from several elements, including silver, gold, copper, arsenic, antimony, lead, and bismuth, and from a number of compounds, including the cadmium chalcogenides and ZnS.

Formation of silver colloids in common soda-lime-silicate glasses results in a very strong optical absorption band centered near 400 nm. Glasses containing silver colloids are yellow, with a bandpass window in the ultraviolet. Since these glasses absorb so strongly at a wavelength commonly used to irradiate photosensitive polymers in the production of electrical circuit boards, they are frequently used for the production of beam masks. A pattern of silver colloids is easily produced in the surface of common float glasses by depositing metallic silver through a mask onto the surface of the glass and then heating the glass to 500 °C for a short time.

Copper can also be used to form longpass filters with a cutoff in the red portion of the spectrum by a similar process to that used to form the silver colloid filters. In this case, it is necessary to use a copper salt in place of the metallic silver, but the remainder of the process is essentially the same (Figure 4).

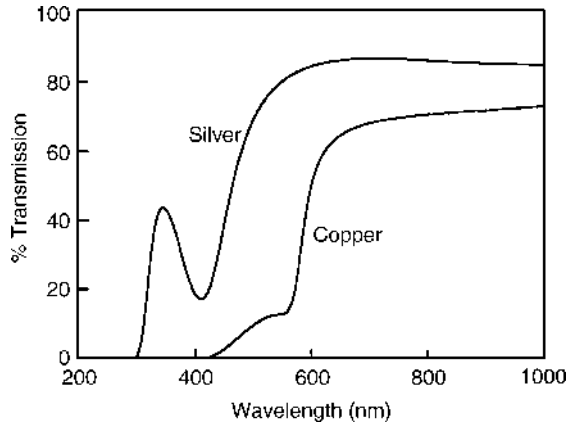

Figure 4 Transmission spectra for longpass filters containing silver or copper colloids.

Precipitation of semi-conducting colloids containing CdS, CdSe, and CdTe in silicate glasses results in sharp cutoffs in the range from the near ultraviolet to the near infrared. The cutoff is the direct result of the bandgap of these crystallites. Since these materials form a broad series of solid solutions, the desired cutoff wavelength can be obtained by use of an appropriate mixture of two of these compounds in the glass batch. Glasses containing only CdS, for example, have cutoffs in the range of 400 to 500 nm. Combining CdS with CdSe will shift the cutoff toward the infrared, with cutoffs in the range of 500 to 700 nm. Combining CdS with CdTe produces glasses with cutoffs near the edge of the visible region. Use of CdTe alone results in filters with cutoffs around 800 nm.

While the most common explanation for the shift in the cutoff in glasses containing the cadmium chalcogenides is based on solid solution arguments, it has been shown that the size of the particles also strongly affects the position of the cutoff for glasses containing CdSe. In this case, the energy gap, which determines the cutoff wavelength, varies inversely with the square of the particle size.

Glasses containing CdS and ZnS can be heat-treated to form filters with cutoffs in the range from 340 to 518 nm. Heat treatment at lower temperatures yields CdS colloids, with a cutoff around 500 nm. At higher heat treatment temperatures, the absorption edge shifts toward the ultraviolet, approaching the value of 340 nm representative of the lattice absorption of ZnS. This behavior has been interpreted as indicating that CdS crystals precipitate at lower temperatures, while at higher temperatures, solid solution crystals of ZnS and CdS are formed.

Colloids of arsenic, antimony, bismuth, and lead form less interesting filter glasses. The optical properties of these metals are such that there is no sharp cutoff wavelength. Growth of colloids produces wavelength-dependent scattering, with a long tail extending into the visible or infrared. These glasses can be produced in colors ranging from light tan to black by progressive heat treatments of the originally colorless glasses in hydrogen. The hydrogen serves to reduce the ions to the atomic state. The atoms then diffuse to form colloids. The growth process creates a sharp interface between the colored and colorless region of the glass, which can be used to produce windows for visible light of highly transmitting glass surrounded by essentially opaque glass.

Absorber Glasses

Glass filters are used to remove unwanted wavelengths from light. When the application is less demanding than those of complex optical devices or systems, the glasses are commonly called 'absorbers' rather than filters. These applications usually involve protection of humans or materials from damage by light, for example, sunglasses and goggles for welders and glassworkers, or to enhance the environment by controlling the entrance or exit of ultraviolet and infrared light. Most of these applications are variations on the filters discussed above.

The most common use of glass as an absorber for both protection from damage and enhancement of the environment is found in the soda-lime-silicate glasses used as windows in buildings and automobiles. The presence of a relatively large concentration of iron oxide provides absorption in both the ultraviolet and infrared regions of the spectrum. Since glasses containing iron do absorb some light in the visible, a compromise between total absorption of unwanted, nonvisible light and the desired visible light must be made. As a result, these glasses usually contain between 0.5 and 2 wt% iron oxide. Changes in melting conditions which alter the ratio of ferrous to ferric ions in the glass are used to tune the glass for specific applications. While these glasses are technically broad bandpass filters, their common applications are based on the favorable location and bandwidth of their transmission window, which roughly corresponds to the visible region.

Iron is also used for protection from ultraviolet and infrared radiation in welder goggles. These glasses contain much more iron (up to 8 wt% or more) than those used for windows, since the demand for protection from the intense light created in welding is much greater. The high absorption coefficient for infrared light results in considerable heating of the glass during lengthy use. A thin film of gold is often applied to the glass to reflect infrared radiation and reduce heating of the glass.

Other absorber applications are more demanding in terms of absorption wavelength. As mentioned earlier, glasses containing neodymium have a strong absorption band at the wavelength of the yellow sodium emission line. These glasses transmit visible light at other wavelengths, allowing their use by glassblowers, who need to be able to see the unheated part of the glass and their immediate surrounding area, while simultaneously reducing the intense yellow light given off by the sodium in the glasses normally used for glassworking.

Protection of the eye from laser radiation is even more demanding due to the very high intensity over very narrow wavelength ranges. Common ultraviolet longpass filters are usually satisfactory for absorption of light from ultraviolet lasers. Specific ions must be added to glasses to provide protection from many

visible and near infrared lasers, since the glass must simultaneously permit transmission of visible light of other wavelengths and total absorption at the laser wavelengths. Light from lasers which operate in the far infrared is readily absorbed by most glasses, since the infrared cutoff of oxide glasses lies around 5 micrometers. Great caution should always be used in selecting glasses for eye protection from lasers. No assumptions regarding the degree of protection should ever be made without insuring that the glass is designed for the conditions under which it will be used.

See also

Optical Coatings: Thin-Film Optical Coatings. **Optical Materials:** Optical Glasses.

Further Reading

Bamford CR (1977) *Colour Generation and Control in Glass*. Amsterdam: Elsevier Scientific Publishing Company.

Deeg EW (1986) Optical glasses. In: Boyd DC and MacDowell JF (eds). *Commercial Glasses*, pp. 1–34. Columbus, OH: The American Ceramic Society, Inc.

Edmund Industrial Optics Catalog N017 (2001) pp. 68–87. Barrington, NJ: Edmund Industrial Optics.

Hillmann E and Neuroth N (1998) Glasses for eye protection. In: Bach H and Neuroth N (eds). *The Properties of Optical Glass*, pp. 370–379. Heidelberg, Germany: Springer-Verlag.

Kolberg U (1998) Coloured glasses. In: Bach H and Neuroth N (eds). *The Properties of Optical Glass*, pp. 350–370. Heidelberg, Germany: Springer-Verlag.

Neuroth N (1998) Transmission and reflection. In: Bach H and Neuroth N (eds). *The Properties of Optical Glass*, pp. 82–95. Heidelberg, Germany: Springer-Verlag.

Paul A (1982) *Chemistry of Glasses*, pp. 204–276. London: Chapman and Hall.

Vogel W (1991) Optical properties of oxide glasses. In: Ullmann DR and Kreidl NJ (eds). *Optical Properties of Glass*, pp. 1–36. Westerville, OH: The American Ceramic Society, Inc.

Weyl WA (1951) *Coloured Glasses*. Society of Glass Technology: Sheffield, England.

Heterogeneous Materials

U Kreibig and M Quinten, Aachen University of Technology, Aachen, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The term heterogeneous materials – being a quite diffuse classification of matter with discrete building units of different kinds, irrespective of their sizes – is defined here by scaling these inhomogeneities in relation to optical light wavelengths. Matter consisting of atoms in a regular structural order (or, alternatively, in fully statistical disorder) is classified as homogeneous since the inhomogeneities are too small to effect optical fields. This applies to crystalline, liquid, and glassy or amorphous matter likewise.

All geometrical/topological and chemical heterogeneities on a larger length-scale than the atomic scale, i.e., single crystallites, organic macromolecules, domains, extended lattice defects, clusters, and nanoparticles of chemical compositions differing from their host material, etc., result in a macroscopic material that is heterogeneous, in optical properties and optical experiments. Examples are multilayer-structures, heterogeneous in one dimension, three-dimensional heterogeneous composites, i.e., isometric building units distributed in some homogeneous host (photographic systems, drugs, cosmetics, color pigments in binders, etc.), and three-dimensional heterogeneous densely packed systems (poly- or nanocrystalline materials, ceramics, or nanoceramics, etc.).

The materials most intensively investigated, in the field of nanoscience, represent a subclass of heterogeneous materials.

Most of the common materials in our everyday environment have heterogeneous structures, often on larger characteristic size-scales. More generally, heterogeneous materials are so widespread that they are much more frequent than homogeneous materials. Some more examples, relevant for their optical properties, may illustrate this: pigment colors, photovoltaic composites, aerosol-systems, hydrosol-systems, many minerals, photonic crystals, island film devices, nanosensors, special biosystems, and polymer-systems. Most interesting optical features arise when these inhomogeneities are below or of the order of light wavelengths. Recent research has concentrated on a narrow size region on the lower nanometer scale where many-atoms-systems change from molecular to solid-state behavior with increasing sizes. This 'transition' is accompanied by drastic changes of almost all physical and chemical

properties, hence these properties develop strong dependences on particle size in that size region, and materials with uncommon properties can be developed.

In contrast to the widespread applications and research perspectives, the basic theoretical understanding of heterogeneous material is generally still poor. One reason is that, since the beginning of the development of optics as a science, the homogeneous materials like glass, water, etc. were preferentially treated since they were easier to describe theoretically. Another reason is the extremely wide variability of existing heterogeneous structures usually based upon low symmetries. This variability is too large, and the resulting optical properties too complex and multifarious to enable a general and uniform description of heterogeneous matter on a quantitative level, like the one successfully developed for the much simpler topology of homogeneous crystals.

Qualitative general features, however, can be obtained from investigations of model systems, with especially simplified topological structures. In the following we describe the model system of nanoparticle/host-matter (or, synonymously, cluster-matter), which is based upon nanoparticles/clusters packed closely in a host material, thus forming macroscopic or microscopic heterogeneous many-particle systems. Even this model system exhibits broad variability of topologies and of chemical compositions, as shown in Figure 1.

The restriction in the present contribution to this topologically simple model system appears to be justified by the fact that the optical properties discussed below for this system are also qualitatively characteristic of other arbitrary heterogeneous materials. We also admit that the nanos may consist of a nucleus and a surrounding shell of different structure and/or different material.

Heterogeneous matter can be defined by its heterogeneous building units and by the borders of these units and by the host material and the borders between building units and host, i.e., the internal interfaces.

It is a general feature that structures are visible because of property changes at their terminating borders: borders create structures, and structures require borders. The more refined the structures are, the more increases the topological extent of the borders. This means that when sizes of building units, for example, in cluster matter decrease, the relative amount and the effects of surface/interface regions increase. Such interfaces in heterogeneous matter are not two-dimensional (as suggested by the term 'interface') but consist of three-dimensional

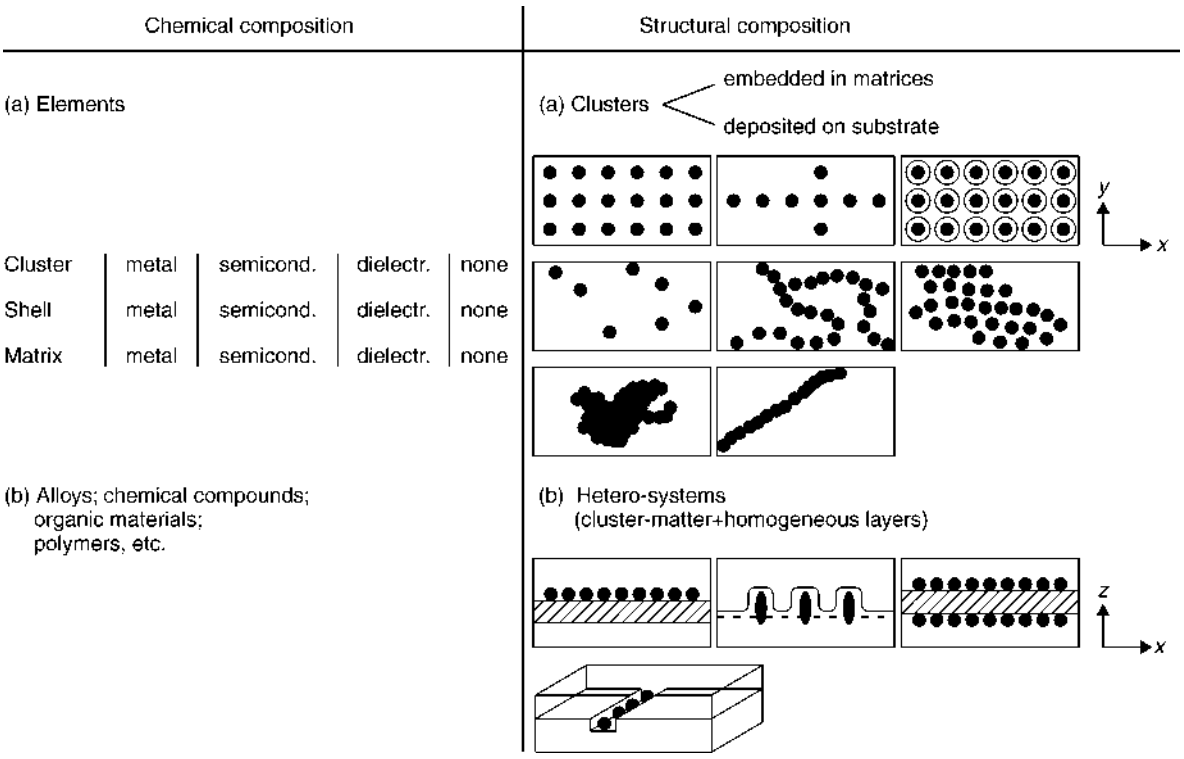

Figure 1 Examples of heterogeneous nanoparticle/host materials. Left: possible material combinations of the three building units: cluster, shell, matrix. Right: some examples of sample topologies.

interlayers, the atomic and electronic structure and the chemical composition of which are usually in complex disorder. In the case of very small particles the interface volume can easily exceed the volume of the building units themselves. There are, however, also examples where the atomic arrangement forms regular interfaces, such as at coherent grain boundaries.

Hence, heterogeneous materials are characterized by various building units, homogeneous matrices and the interface regions. To describe them, we will focus on our introduced model system of nanoparticle/host matter, where we further specialize to spherical shapes of the particles.

The investigation and description of optical properties of heterogeneous materials have to deal with the interior of the building units, e.g., characteristic optical size effects, and with the interfaces, e.g., optical interface effects. The importance of both increase with decreasing characterizing structural lengths. The role of the host materials, when present, can be different, depending on their concentration: in dilute particle systems they may act as bulk-like homogeneous matter (e.g., the electrolyte in colloidal liquid systems); at low amounts, the host may also develop size effects (e.g., the water layers in the opal).

General Optical Properties

Among the experimental characterization methods most frequently applied to heterogeneous matter, the optical ones are essential, ranging from spectroscopy in the frequency domain (from FIR to UV), to femtosecond analysis in the time domain.

Optics can mean physical optics, i.e., the science of optical material properties, including light emission (the 'active' properties) and light absorption (the 'passive' properties), and instrumental optics, i.e., special experimental techniques based upon propagation of light to analyze such samples.

The most usual experiments for the investigation of linear optical response are refraction, absorption, reflection, and elastic scattering. Also, there are the emission responses and the broad field of nonlinear optical responses.

Refraction, reflection, and absorption are typical of homogeneous matter. Elastic light scattering, on the other hand, can be used to define optical heterogeneous matter, since elastic scattering is a direct consequence of the existence of interior interfaces forming deviations from the homogeneous structure. While the grainy structure on the atomic level in homogeneous matter does not give rise to disturbances of the shape of a plane interacting optical wave, this does occur with heterogeneous matter. Each inhomogeneity gives rise to a wavefront deformation and, hence, to light scattering. Scattering spectra and intensities provide important information about the heterogeneous material, so such optical methods are frequently used for its identification.

Besides scattering, the electrical charging of the interior interfaces induced by the light field, can lead to spectacular multipolar optical absorption features, unseen in related macroscopic homogeneous materials, such as Mie resonance, discussed below.

The optical properties of our model system are composed of various contributions which all act together in a complex way, which can be described as:

1. the optical properties of the single building units. In the case of solid state units, they are characterized by the optical material properties, frequency dependent and, generally, complex dielectric function $\varepsilon(\omega)$ or, alternatively, the (generally complex) refraction index $n(\omega)$, which, in the nanometer-size region, both develop strong variations with size.
2. the optical interactions among all building units in the sample. There are different effects:
   - interference effects among the scattering radiation of the units;
   - electromagnetic coupling effects;
   - the contributions of host materials, if present; and
   - the interactions between heterogeneous building units and host.

The Optical Nanoparticle Resonances (Mie Resonances)

Metallic and semiconductor nanoparticles can exhibit special, spectrally selective absorption and scattering bands which do not have counterparts in the molecular state, in thin film structures, or in the extended bulk. They represent the most spectacular optical properties of matter with nanosized heterogeneities. Therefore they will now be treated in more detail in the frame of our model material of spherical inclusions in a homogeneous embedding medium. We begin with a comparison to thin film structures.

The interband transitions (in Ag below 325 nm) are almost equal in both spectra shown in Figure 2, of a thin Ag film and spherical Ag nanoparticles, respectively, while above 325 nm the flat relaxator spectrum of the conduction electrons in the metal film differs strongly from the oscillator spectrum of the same electrons in the clusters. The inset demonstrates the excitation of the dipolar surface oscillator mode by

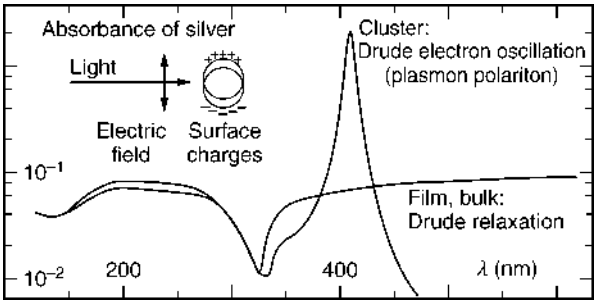

Figure 2 Comparison of absorbances of a thin Ag film and spherical small Ag nanoparticles.

incident polarized light. This figure exemplifies the above-mentioned effect that charging of interfaces by the incident electromagnetic lightwave can give rise to electromagnetic multipole resonances, the Mie-resonances.

The monotonic conduction electron absorption spectrum of the thin film, ranging from UV to zero frequency (the dc current), is thereby compressed in the nanoparticles into the narrow Mie-absorption band with peak position in the blue, containing almost the whole oscillator strength of the conduction electrons. In particular, the Ag-spheres are therefore transparent in the low frequency region, and if we assume the conductivity according to the Drude model as being definitive for a metal, we must assume that these particles are no longer metallic. The conduction electrons no longer follow the Drude relaxator behavior; instead the narrow peak indicates the properties of an oscillator.

The origin of this oscillator and its excitation can be described, being a quantum phenomenon; due to the incident electrical field, the electron plasma, the 'electronic particle', is shifted against the fixed ionic background, the 'ionic particle', causing surface charges modulated with the light frequency. The surface charges produce restoring forces to the plasma, following the periodic field strength, which is the condition required for oscillator behavior. As every free oscillator the particle develops resonances at special eigenfrequencies, and in the case of our metal clusters these are close to the peak frequency of the Mie-absorption, the oscillations of which not being free but forced by the incident field. Analogous to the mechanical case, the resulting resonance frequencies of free and forced oscillations differ slightly. To distinguish them, Mie resonances in metallic particles, coupled to the external electromagnetic field, are called plasmon polaritons.

In the frame of classical electrodynamics each single particle in the heterogeneous material acts as a spherical nano-antenna. This acts as receiving antenna by absorption and, as with every other kinds of antenna, simultaneously becomes a transmitting antenna by re-radiating elastically scattered light.

In the case of Cu, Ag, and Au, and many other metals, these absorption frequencies are in the visible, hence beautiful colors appear. However, there are a lot of metals where this oscillator is overdamped due to low electron mobility. Under particular conditions, such resonances also occur in semiconductor nanoparticles, but due to the lower charge densities, in the IR. At weak fields, the optical response of the oscillator is linear. At higher fields it becomes nonlinear, and thus second and higher harmonics can be excited.

Such unique collective excitations are not restricted to electrons but also spherical surface phonon polaritons can be excited if the particles are nonmetallic (in this case, free electron excitations do not cover the spectral region of phonon excitations). Only in the most simple case of particles which are extremely small compared to wavelength ($R/\lambda \leq 1/20$), the dipolar mode of excitation described so far (i.e., homogeneous polarization of the whole particle) comprises the whole response. For larger particle sizes, a complicated, nonhomogeneous polarization of the whole particle occurs, which can be described by an orthogonal multipole expansion, i.e., multipolar oscillations. The according size (R-) dependent multipolar contributions, which are excited simultaneously and sum up to the total optical response. Further complications arise in larger particles by retardation-induced phase shifts of the lightwaves along the particles.

A complete description of this complex multipole excitation was given by Gustav Mie on the basis of classical electrodynamics. Mie's theory is the exact solution within the frame of Maxwell theory for spherical symmetry, as are the Fresnel-formulae for the planar geometry. However Mie's theory does not describe exactly the conditions which we observe in real heterogeneous systems, hence several extensions of Mie's theory will be presented in the section entitled *Beyond Mie's Theory,* below.

Mie's theory also gives some qualitative insight into the optical behaviour of other kinds of heterogeneous matter. A highly abridged summary of this theory will be presented below.

The Theory of Mie

In the following, an outline of the Mie theory for one spherical particle is given by listing the mathematical steps:

1. Introduction of spherical coordinates r, θ, ϕ. Particle radius R;

2. Incidence of a plane, monochromatic, harmonic electromagnetic wave (circular frequency ω, wavelength λ);
3. Solution of Maxwell's equations for the incident and scattered fields in the surrounding dielectric host medium with (real) dielectric constant $\varepsilon_{\text{matrix}}$, and in the interior of an arbitrary spherical particle of (complex) dielectric function $\varepsilon(\omega)$.

   There are two solutions for transverse fields: one with the radial component of the electric field $E_r = 0$, called 'transverse electric (TE) solution', and one with the radial component of the magnetic field $H_r = 0$, called 'transverse magnetic (TM) solution'. These solutions can be obtained in terms of vector spherical harmonics which can be derived from corresponding scalar potentials Π_E and Π_H. These are solutions of the Helmholtz-equation:

$$\nabla^2\Pi_{E,H} + K^2_{\text{transverse}} \times \Pi_{E,H} = 0, \qquad K^2_{\text{transverse}} = \varepsilon_{\text{transverse}}\omega^2/c^2 \qquad [1a]$$

   In spherical coordinates, the Helmholtz equation is solved by a separation of variables ansatz and a multipole expansion, yielding an infinite number of independent partial solutions with multipolar order L, the TM partial waves or 'electric partial waves', and the TE partial waves or 'magnetic partial waves'.

   A third solution, the potential Γ, that only affects the electric field, can be found as the solution of

$$\nabla^2\Gamma = 0, \qquad \varepsilon_{\text{longitudinal}} = 0 \qquad [1b]$$

   and represents longitudinal excitations in the particle.
4. Boundary conditions at $r = R$:

   for the transverse fields:

$$E_\theta^{\text{incident}} + E_\theta^{\text{scattered}} = E_\theta^{\text{interior}} \qquad E_\phi^{\text{incident}} + E_\phi^{\text{scattered}} = E_\phi^{\text{interior}} \qquad [2a]$$

$$H_\theta^{\text{incident}} + H_\theta^{\text{scattered}} = H_\theta^{\text{interior}} \qquad H_\phi^{\text{incident}} + H_\phi^{\text{scattered}} = H_\phi^{\text{interior}} \qquad [2b]$$

   for the longitudinal fields:

$$E_r^{\text{incident}} + E_r^{\text{scattered}} = E_r^{\text{interior}} \qquad [3]$$

   (Sauter–Forstmann condition)

   The linear response function of the particle material used here, is ε, where $\varepsilon = \varepsilon_1 + i\varepsilon_2$ is averaged over the particle volume, including the surface–interface region.

   The magnetic permeability μ is set 1 for the investigated high frequency regions.
5. Computation of cross-sections for extinction, scattering, and absorption from Poynting's law:

   - Extinction, i.e., the sum of absorption and scattering losses:

$$\sigma_{\text{ext}} = \frac{2\pi}{k^2}\sum_{L=1}^{\infty}(2L+1)\text{Re}\{a_L + b_L\} \qquad [4]$$

   - Elastic scattering:

$$\sigma_{\text{sca}} = \frac{2\pi}{k^2}\sum_{L=1}^{\infty}(2L+1)(|a_L|^2 + |b_L|^2) \qquad [5]$$

   - Absorption, i.e., the energy dissipation (production of heat and thermal radiation):

$$\sigma_{\text{abs}} = \sigma_{\text{ext}} - \sigma_{\text{sca}} \qquad [6]$$

   L indicates the multipolar order of the Mie resonances.

The coefficients a_L and b_L, following from the Maxwell boundary conditions for transverse fields, are

$$a_L = \frac{m\psi_L(mx)\psi'_L(x) - \psi'_L(mx)\psi_L(x)}{m\psi_L(mx)\eta'_L(x) - \psi'_L(mx)\eta_L(x)} \qquad [7a]$$

$$b_L = \frac{\psi_L(mx)\psi'_L(x) - m\psi'_L(mx)\psi_L(x)}{\psi_L(mx)\eta'_L(x) - m\psi'_L(mx)\eta_L(x)} \qquad [7b]$$

with $m = 2x/n_{\text{Matrix}}$, where m denotes the complex-valued index of refraction of the particle $\varepsilon = n'^2$ and n_{matrix} the real-valued index of refraction of surrounding medium. k is the wavevector and $x = k{\cdot}R$, the size parameter. $\Psi_L(z)$ and $\eta_L(z)$ are Riccati–Bessel functions of the first and third kind. The prime indicates differentiation with respect to the argument in parentheses. Including longitudinal fields, eqn [7] will be slightly modified.

Applying Lambert–Beer's law, these cross-sections are related to the absorption-, scattering-, and extinction-constant $A(\omega)$, $S(\omega)$, and $E(\omega)$, respectively, and the resulting intensity loss ΔI of a many-particle system of number-density Z of the (noninteracting) particles, is

$$\Delta I = I_0(1 - \exp(-Z{\cdot}\sigma_{\text{ext}}{\cdot}d)), \qquad d = \text{sample thickness} \qquad [8]$$

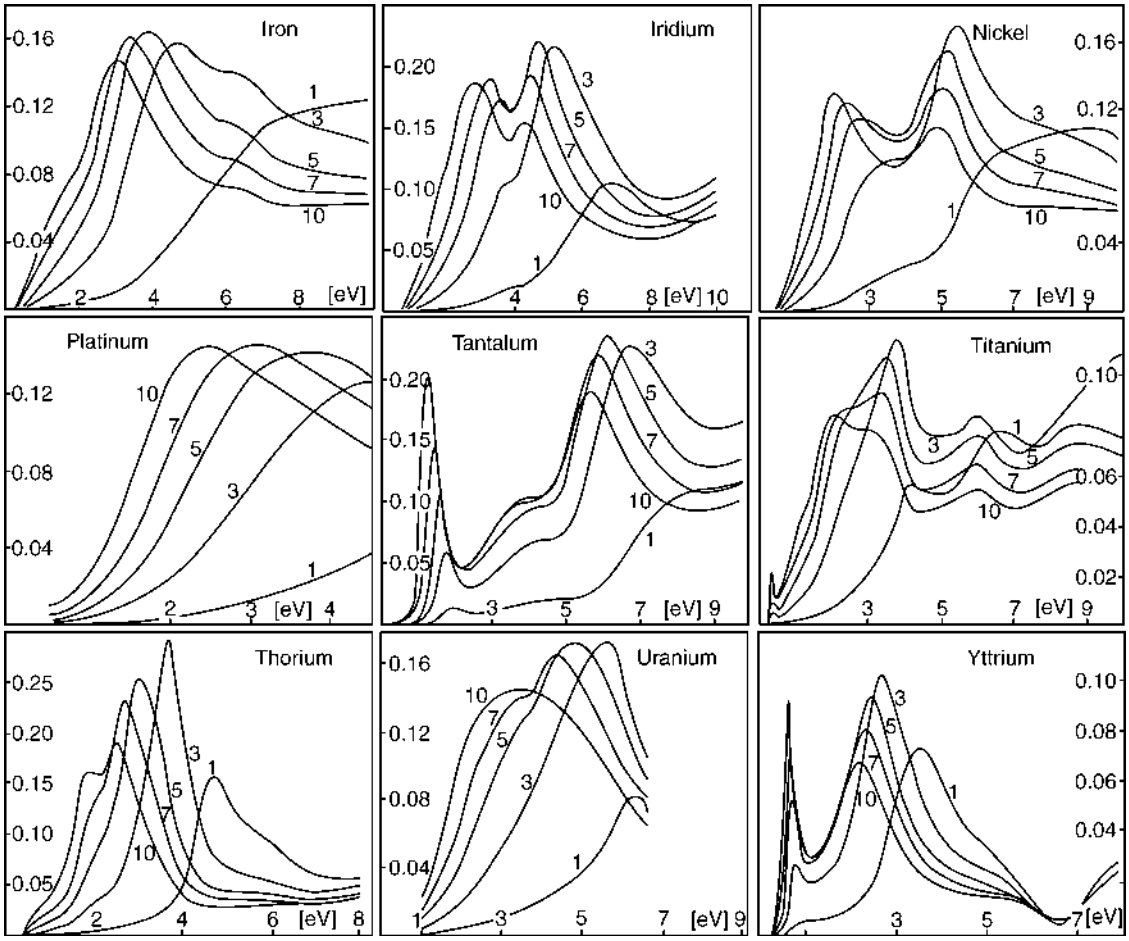

Figure 3 Mie extinction spectra of various particle materials. $2R = 10$ nm. The clusters are embedded in matrix materials with dielectric constants between 1 and 10, which are indicated for each spectrum. Obviously, the peak structure of the spectra strongly depends on the matrix.

Figure 3 gives a compilation of some spectra of general interest determined by evaluating Mie's theory.

Beyond Mie's Theory

Application of Mie's theory to experimental results is disappointing because, in most cases, quantitative coincidence is not obtained. One reason is that the simple assumptions incorporated either in Maxwell's theory or in Mie's derivation are not met by reality.

Mie's theory does not apply, if:

1. particle sizes in a diluted cluster matter sample are not uniform;
2. particle shapes differ from the sphere;
3. $\varepsilon_{\text{matrix}}$ varies with frequency, $\varepsilon_{\text{matrix}}(\omega)$
4. cluster-dependent dielectric functions differ from the according functions of the bulk;
5. the (step-function like) Maxwellian boundary conditions at the particle-matrix interface are not applicable;
6. dielectric functions of particles and/or matrix are nonlinear;
7. surrounding/embedding matrix material is absorbing;
8. particle structures are heterogeneous (core-shell structures, multi-grain clusters, etc); and
9. particles form close-packed aggregates instead of being well-separated, and hence, electromagnetic coupling among particles is important.

Of course, this list may be continued. Regarding realistic samples, it is obvious already from these deviations that quantitative correspondence between experiment and Mie theory usually cannot be expected. But, over time, each of these extra effects has been treated and included into extensions of Mie's

theory, which should therefore be applied, instead of the original theory. In the following several of these effects will be treated in some detail.

Particle Shapes Deviating From the Sphere

Even metals of high structure-symmetry, like the fcc-Au, can yield nanoparticles with low symmetry habits. Figure 4 shows Au nanoparticles with exotic shapes ranging from triangle and hexagonal platelets over fivefold symmetry multitwin-particles to rodlike shapes. The majority of particles are multigrain structures of irregular isometric shapes.

Multishell Particles

In experiments, particle surfaces often react chemically with the surrounding media, thereby forming compound shells. Core/shell structures can be clearly visible in the optical absorption spectra. Figure 5 shows, as an example, the absorption spectra of nanoparticles of Ag core/Au shell and of Au core/Ag shell, respectively with varied shell thicknesses.

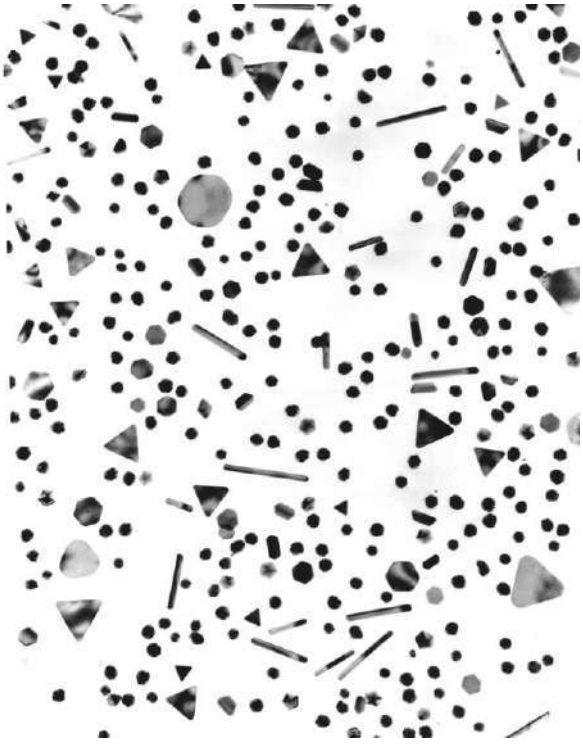

Figure 4 Chemically (Esiguondy method) produced Au clusters with extraordinary variety of different shapes. (Of special interest are the small particles of five-fold symmetry.) Mean sizes of the irregular, almost spherical particles: 36 nm.

Resonances will also occur, e.g., in metal-coated dielectric particles. An example is shown in Figure 6, with extinction spectra of silver-coated titanium-dioxide particles, computed by an extension of Mie's theory.

Aggregates of Nanoparticles with Electrodynamic Particle–Particle Coupling

The optical extinction and scattering spectra of many-particle systems with dense packing forming aggregates (see Figure 4) are strongly influenced by electromagnetic interactions: each particle feels the

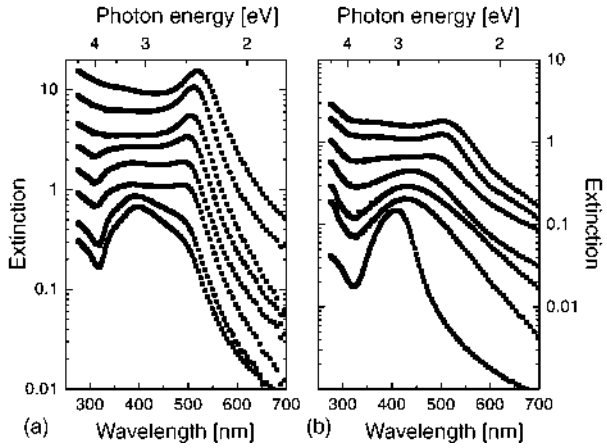

Figure 5 Measured extinction spectra of (a) silver-coated gold particles (gold core $2R = 17$ nm, shell thickness varying from $d = 0$ to $d = 3.6$ nm from top to bottom), and (b) gold-coated silver particles (silver core $2R = 18$ nm, shell thickness varying from $d = 0$ to $d = 5$ nm from bottom to top).

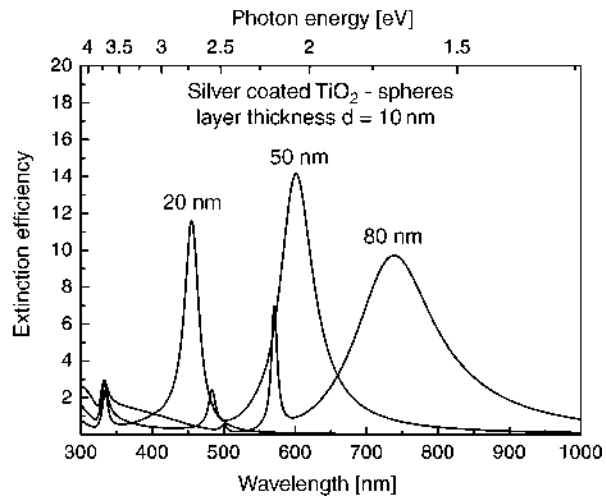

Figure 6 Calculated extinction spectra of spherical silver-coated titanium dioxide particles. The shell thickness is kept constant at $d = 10$ nm, the core sizes amount to $2R = 30$ nm, 50 nm, and 80 nm, respectively, as indicated in the graph.

scattering near fields of all its neighbors, which give rise to splitting, broadening, and shifts of resonances. This is clearly demonstrated with colloidal Ag and Au particles in Figure 7.

The splitting, broadening, and shifts can be assigned to the shape and size of the resulting many-particle aggregates. An example for the influence of the topology of the many-particle system is given in Figure 8 with extinction and scattering spectra of silver particle aggregates of $N = 5$ identical silver particles with $2R = 40$ nm, which were computed by the Generalized Mie Theory (GMT). As a rough rule-of-thumb, such electrodynamic coupling effects of the oscillators only can be ignored if the smallest next neighbor center-to-center distance d_{cc} exceeds $d_{cc} \approx 5 \cdot R$.

Size Effects

The extraordinary practical success of Mie's theory is based upon the fact that this example of classical electrodynamics enables us to compute numerically the optical response for arbitrary realistic particle materials. In contrast quantum theoretical calculations are restricted, beyond the molecule-solid state transition, to jellium and jellium-related hypothetical material.

Classical electrodynamics, being a phenomenological theory to describe light propagation, does not yield information about optical material properties. These enter via the dielectric functions inserted into the Maxwell boundary conditions. They must be taken from elsewhere, such as from experiments or from quantum solid state theory model calculations. This twofold basis of Mie's theory is illustrated in Figure 9.

There are two different sources for particle size dependences of the optical extinction spectra:

1. The electrodynamics: the larger the particles, the larger are the number of contributing multipoles and the influences of retardation effects, both influencing strongly the optical spectra;
2. The optical material properties describing the particle material which, in the nanoscale size region, exhibit size dependences. These size dependences can be treated by corrections entering the bulk material function.

In general, the dielectric function $\varepsilon(\omega)$ of a metal or semiconductor is the sum of susceptibility contributions of the lattice, of the conduction electrons, and of the interband (electron–hole) excitations:

$$\begin{aligned}\varepsilon(\omega) &= \varepsilon_1(\omega) + i\varepsilon_2(\omega) \\ &= 1 + \chi^{\text{lattice}}(\omega) + \chi^{\text{free electrons}}(\omega) \\ &\quad + \chi^{\text{interband}}(\omega)\end{aligned} \qquad [9]$$

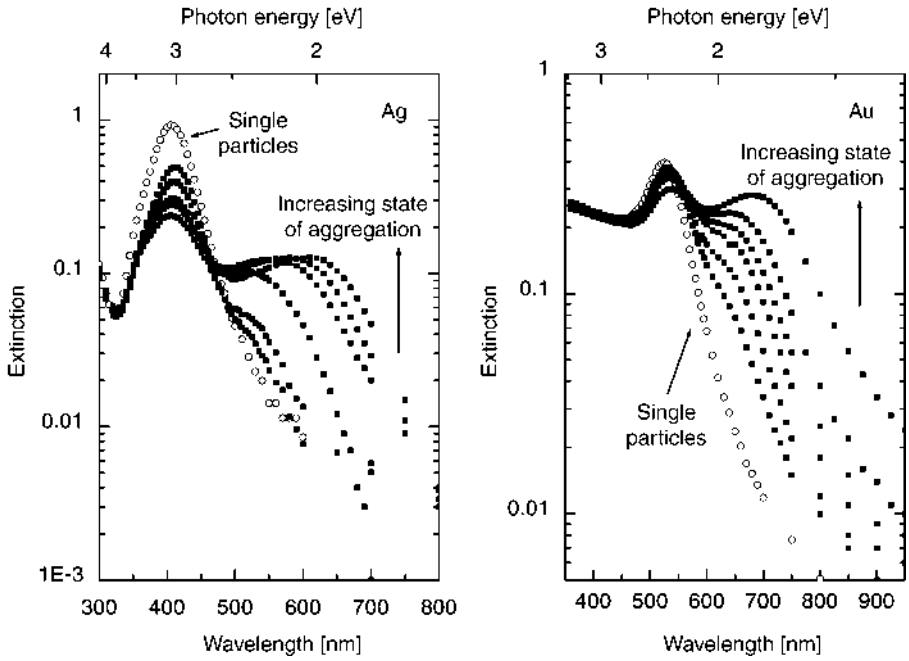

Figure 7 Measured extinction spectra of aggregated samples of (a) silver particles with $2R = 28$ nm, and (b) gold particles with $2R = 38$ nm. The samples contain random distributions of aggregates of various topologies, but with increasing state of aggregation from bottom to top.

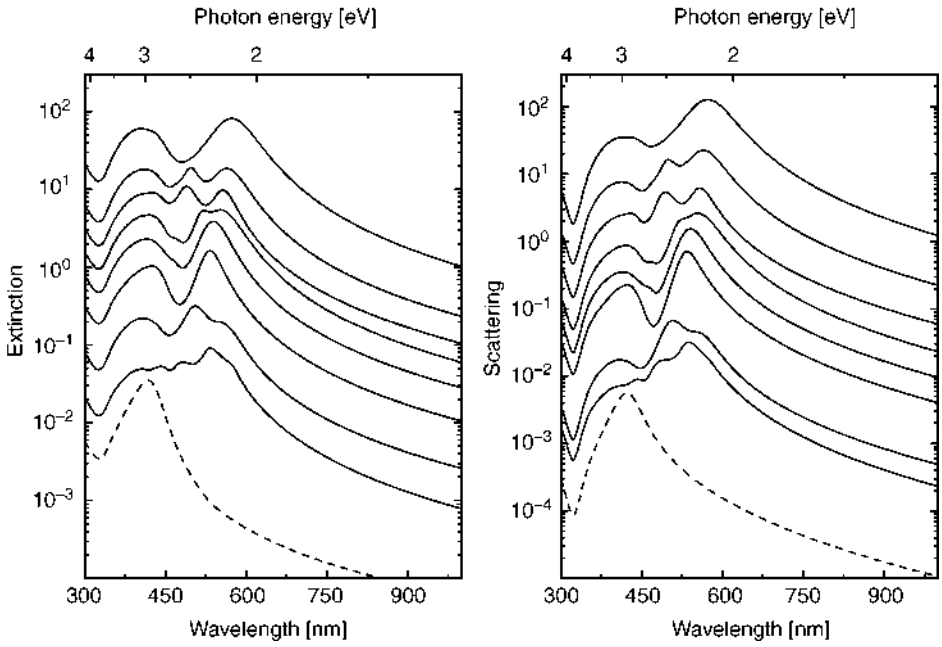

Figure 8 Calculated extinction and scattering spectra of aggregates with $N = 5$ almost touching identical silver spheres of $2R =$ 40 nm. From top to bottom, the spectra belong to the following particle arrangements in the aggregate: [illegible]. The dashed line represents the single particle spectra.

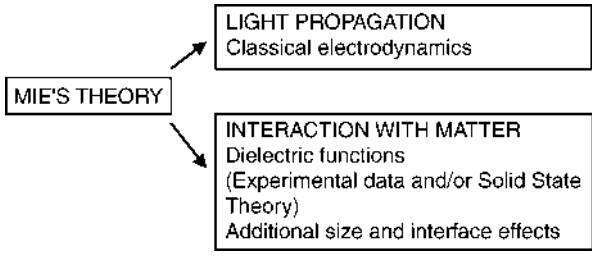

Figure 9 Scheme of the basics of Mie's theory.

A detailed expression was given by Bassani (here, the lattice contributions, mainly important in the IR, were disregarded):

$$\varepsilon(\omega) = 1 - \underbrace{\frac{(ne_0^2/\varepsilon_0 m_{\text{eff}})}{\omega^2 + i\gamma\omega}}_{\text{conduction electrons}} + \underbrace{\chi^{\text{interband}}(\omega)}_{\text{interband transitions}} \qquad [10]$$

with

$$\chi^{\text{interband}} = \frac{2\hbar^2 e^2}{\varepsilon_0 m_{\text{eff}}^2} \sum_{i,f} \int_{\text{BZ}} \frac{2d^3k}{(2\pi)^3} |\mathbf{e} \times M_{\text{if}}(\mathbf{k})|^2 \times \left\{ \frac{1}{[E_f(\mathbf{k}) - E_i(\mathbf{k})][(E_f(\mathbf{k}) - E_i(\mathbf{k}))^2 - \hbar^2\omega^2]} + i\frac{\pi}{2\hbar^2\omega^2}\delta(E_f(\mathbf{k}) - E_i(\mathbf{k}) - \hbar\omega) \right\} \qquad [11]$$

In this equation, the quantity n is the conduction electron density, $\mathbf{k}$ the wavevector of the electrons with norm k, and BZ denotes the Brillouin Zone. The vector $\mathbf{e}$ is the polarization vector of the incident light. M_{if} is the transition matrix element with i = initial and f = final states.

To which degree of approximation a single dielectric function can describe the optical response of a nanoparticle consisting of inner core and surface with differing local optical properties (i.e., polarizabilities), and how this function has to be modelled as an average, is still a problem. An easy, but only approximate answer is to identify terms in the eqns [10] and [11], which may change if the cluster size is reduced down to the nanometer scale, and to introduce correction terms.

In the case of the conduction electrons of eqn [10], these are the electron density n, the effective electron mass m_{eff}, and the relaxation frequency γ. In the terms of the interband transitions (index inter), these are the transition matrix elements M_{if} and the band structure energies of the initial and final states E_i and E_f. All of these quantities change, when the particle size is reduced, in different size regions and to different amounts, and detailed quantum theoretical investigations are required to derive proper correction terms.

The application of the continuum-based dielectric functions and of Mie's theory fail when the molecule-solid state transition region is reached.

The effects of size-dependent dielectric functions upon cluster absorption spectra are demonstrated in the series of absorption spectra of Figure 10, measured on highly diluted heterogeneous systems containing spherical Ag, Au, and Cu particles which were embedded in a glass matrix. If $\varepsilon(\omega)$ were independent of size, then the spectra should be identical, in the investigated particle size region. So, their differences reflect clearly that $\varepsilon = \varepsilon(R)$.

To describe the size effects of the conduction electrons, the relaxation frequency γ of eqn [10] has been modified introducing a parameter that is called A-parameter which also contains size effects and also particle–host interface effects. To explain the electronic size effects in detail, quantum effects have to be considered taking the discretization of the bandstructure into account, but they can also be treated by a simple classical approach: this is the mean free path effect (FPE), assuming the mean flight time of the effective conduction electrons (i.e., those with Fermi energy E_F) between subsequent surface collisions amount to

$$\tau\gamma_{FPE} = A_{FPE}\frac{v_{Fermi}}{R} \qquad [12]$$

The A-parameter is a measure of the effectiveness of the single collision in view of relaxation. Complete 'memory loss' of the drift momentum in each collision would lead to $A_{FPE} = 1$. Taking electron correlation effects and energy relaxation into account, A-values larger than 1 may also be realistic. Values larger than unity also occur in multigrain particles, due to grain boundary 'collisions'.

Various quantum mechanical models of electronic particle size effects result in relations, formally coinciding with the $1/R$-dependence in eqn [12]. Most relevant is the value of A for particles *in vacuo* with clean surface. This value has been determined theoretically and experimentally to $A \approx 0.25$ for Ag particles.

Properties of Internal Interfaces

Structures of Surfaces/Interfaces

Only mathematical interfaces in geometry can, in fact, be two-dimensional planes separating two different volume elements. This is assumed in the frame of Maxwell boundary conditions entering Mie's theory. In real nanostructured matter, however, the interface regions are three-dimensional layers formed by disturbances of atomic and/or of electronic structures, and/or of chemical compositions. In addition, due to the finite wavelengths of the conduction electrons, the according electron density 'spills out' and, as a consequence, we find electron density beyond the geometric surface. It should be kept in mind that these effects hold, with proper changes, for interfaces in nanostructured materials in general. Hence, an understanding of the interface regions is of key importance for all nano-material.

Electronic Interface States

Up to now, this discussion has been restricted to particles with uncovered, 'clean' surfaces. In most practical examples of heterogeneous materials however, the building units are embedded in a host matrix, deposited upon the surface of a substrate, or are in close contact to neighboring crystallites, etc., to stabilize the samples. Then, special electronic

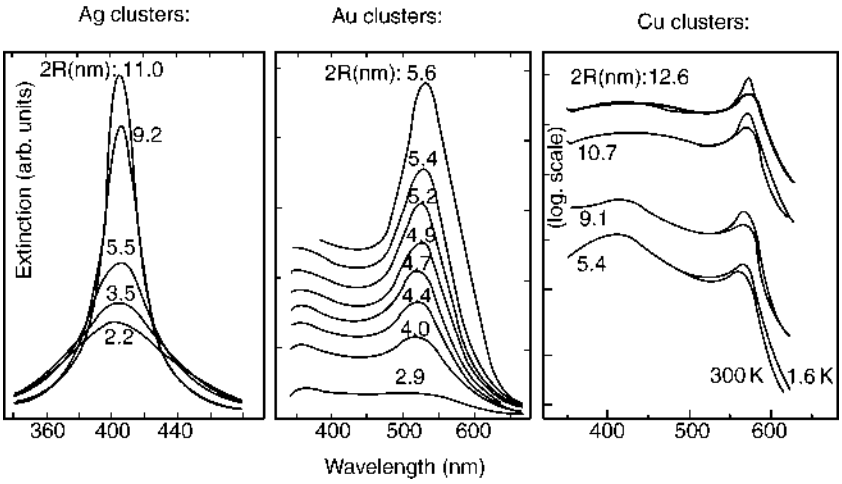

Figure 10 Measured extinction spectra of Ag, Au, and Cu clusters of various mean sizes, grown in glass matrix. In the case of Cu-particles, measurements at 1.5 K are added.

interface states are created by contributions of particle electrons and host electrons, which are (disregarding their density on the surface) similar to adsorbate states.

One might assume that the broad knowledge about these states in surface science, obtained from samples with extended crystallographic planes, should be very useful here. However, the surface of a cluster, in principle, differs from the planar surface of a single crystal by containing a multitude of different local surface structures, atomic sites, and varying atomic coordination numbers. Even if the surface structure is not irregular (as is usual in nanoparticles), the surface consists of several facets of different orientation separated by edges and corners. The coordination numbers of surface atoms, thus, vary strongly (e.g., between 5 at a corner and 9 on a plane), indicating surface sites to have different energies.

We have shown that, if the particles are metallic, the 'particle/host'-interface states give rise to two different electronic charge transfer processes of particle conduction electrons, the static and the dynamic charge transfer. This will be demonstrated qualitatively for metal particles by the example presented in Figure 11.

The Static Interface Charge Transfer and its Effects Upon Mie Resonance Positions

At the left side of Figure 11, the conduction band of the metal cluster close to its surface is shown, which is occupied with electrons up to the Fermi energy E_F. Chemisorption of a free atom (one energy level E_a of which is plotted at the right-hand side) is assumed to occur by approaching the atom to the cluster surface. The resulting energy states are also shown. Let us assume the marked electronic level E_a in the free atom is nonoccupied. During the chemisorption process, this level is shifted and broadened, developing density of nonoccupied states below the particle Fermi energy, which is localized at the three-dimensional interface region. After complete the process is, such a state will be occupied permanently by electrons of the cluster (static charge transfer), and the nanoparticle will have lost electrons, thus its electron density being reduced. The driving impact of this process is the equilibration of the chemical potential in the whole cluster-matrix system.

Now assume we embed a cluster into some surrounding material. Then all those atoms of the latter, which are situated directly at the interface, become chemisorbed at the cluster. Hence remarkable amounts of charges can be transferred into the interlayer region and an electrical double layer is created. As an example, a very strong amount of static charge transfer was observed experimentally when applying solid Fullerite (C_{60} molecules) as embedding matrix for Ag-nanoparticles: here, each C_{60} in direct contact with the Ag particle attracts one electron, forming (C_{60}^-). Hence, in this experiment with 500 atoms large particles, in total about 20% of the conduction electrons of the Ag-particle were transferred into the interlayer. Strictly speaking, such a nanoparticle of the heterogeneous material is no longer a silver particle because of the remarkable changes of electronic properties by static charge transfer.

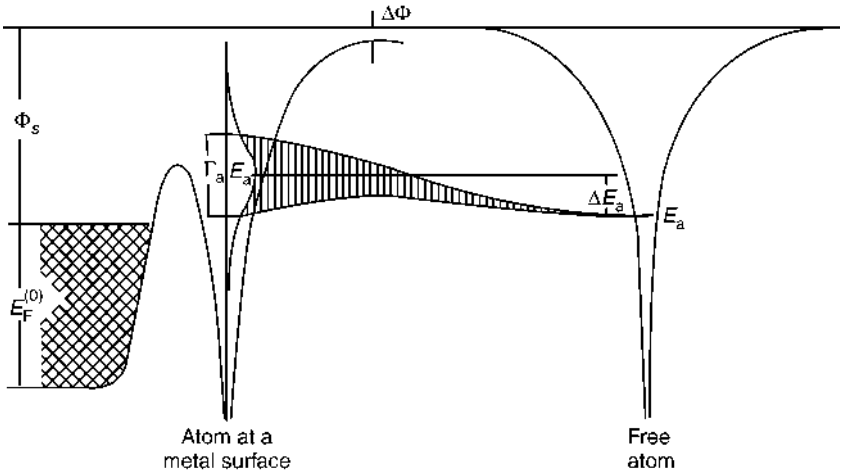

Figure 11 Electron energy scheme for atomic adsorbation at a metal particle surface. The free atom (energy spectrum: right) approaches the particle surface (left; conduction band with Fermi-level E_F). Its energy level E_a is thereby shifted and broadened (including ensemble broadening of many atoms deposited at many clusters). The atoms are assumed to be separated from the metal by a tunnel barrier (see text).

Some consequences are listed in the following:

In a first step, let us assume that the electronic bandstructure is not altered by the charge transfer. Then we expect: (i) changes of level occupation; (ii) changes of Fermi energy and Fermi velocity; (iii) changes of effective masses close to E_F; (iv) changes of surface 'spill out'; (v) additional adsorbate levels and special surface states; and (vi) changes of the optical interband transition edge.

In a second step, we may consider possible changes of electronic bandstructure (e.g., energies) and, finally, also changes of atomic structures (i.e., lattice structures and/or surface structures).

Since plasmon frequencies depend on the conduction electron density, the resulting Mie resonance can monitor and give quantitative information of this effect. We find shifted by the static charge transfer:

$$\Delta\omega_{\text{resonance}} \approx (\sqrt{n_2} - \sqrt{n_1}) \times \sqrt{\frac{e_0^2}{\varepsilon_0 m_{\text{eff}}(2\varepsilon_{\text{matrix}} + 1 + \chi_1^{\text{interband}})}} \quad [13]$$

with n_1 being the conduction electron density in the cluster *in vacuo*, and n_2 being the electron density after contact with foreign material (adsorbate, substrate, or host-matrix). This peak shift of the Mie resonance, however, is superimposed in the experiment by a much larger shift, following simply from the change of Maxwell boundary conditions by the altered dielectric constant of the particle surrounding $\varepsilon_{\text{matrix}}$, when going from vacuum to a matrix material. Hence, the experimental determination of static charge transfer processes is rendered more difficult.

The Dynamic Interface Charge Transfer and its Effects Upon Mie Resonance Widths

After the static charge transfer in Figure 11 has taken place, electrons close to E_F can fluctuate between the cluster and empty adsorbate levels, if present at E_F. The residence times in these levels, which are usually separated from the Fermi sea of the metal cluster by some tunnel barrier, vary statistically. If a plasmon is excited, i.e., mainly a collective, phase ordered motion of all conduction electrons superimposed to their statistical motion, we may assume that, after such electron transfer back and forth, the reminiscence to the primary drift momentum, and the phase of this excitation, has been lost. Thus, the phase relaxation of the plasmon is increased.

The consequences of this dynamic charge transfer effect ('chemical interface damping') have been investigated in detail. Here, we only state that the directly observable main consequence is the broadening of Mie resonances. If the particles are sufficiently small to neglect radiation damping (this is the case for Ag clusters smaller ≈ 15 nm), band broadening is only due to internal energy dissipation and the band width can be modeled quantitatively with the modified relaxation frequency γ in eqn [10] which yields an additive correction term:

$$\Gamma(A, R) \approx \Gamma_0 + \frac{2ne_0^2}{\varepsilon_0 m_{\text{eff}}\omega^3} \times \frac{A}{\sqrt{(\partial\varepsilon_1/\partial\omega)^2 + (\partial\varepsilon_2/\partial\omega)^2}} \frac{v_{\text{Fermi}}}{R} \quad [14]$$

with the A-parameter extended to the additional electronic excitations in the particle/host interlayer:

$$A = A_{\text{size effects}} + A_{\text{interface effects}} \quad [15]$$

The A-parameter of eqn [15] now expresses both the strengths of the particle size effects and of the chemical interface damping, due to statistically varying occupation of interface states near E_F.

An important feature of this model is that the dependence on the kinds of materials is described by specific material parameters ($A_{\text{size effects}}$ and $A_{\text{interface effects}}$), and the direct size dependence effects are given by an $1/R$-dependence, for both the direct size effects and the interface effects. For one fixed cluster size R, the magnitude of $A_{\text{interface effects}}$ directly characterizes the probability of the single temporal charge transfer, including the number and energies of contributing interface states; hence, important properties of the adsorption process can be investigated by determining the A-parameter.

Experimental Results

In the experiments presented in the following, we deal with the interfaces between clusters and a more or less homogeneous solid embedding medium. We selected very small silver nanoparticles of about 2 to 4 nm size with accordingly large ratio of surface- to volume-atoms, in order to optimize the sensitivity of the measurements. Silver is unique in having the most pronounced Mie resonance of all known materials.

In Figure 12 two classes of matrix materials are presented, fluorides and oxides. The absorption spectrum of the free particles with uncovered 'clean' surface (measured in a cluster beam in UHV) is added in both diagrams which is only influenced by size effect. General features are, that embedding in a host medium causes:

- drastic Mie resonance peak shifts toward IR, depending on the dielectric constant of the matrix

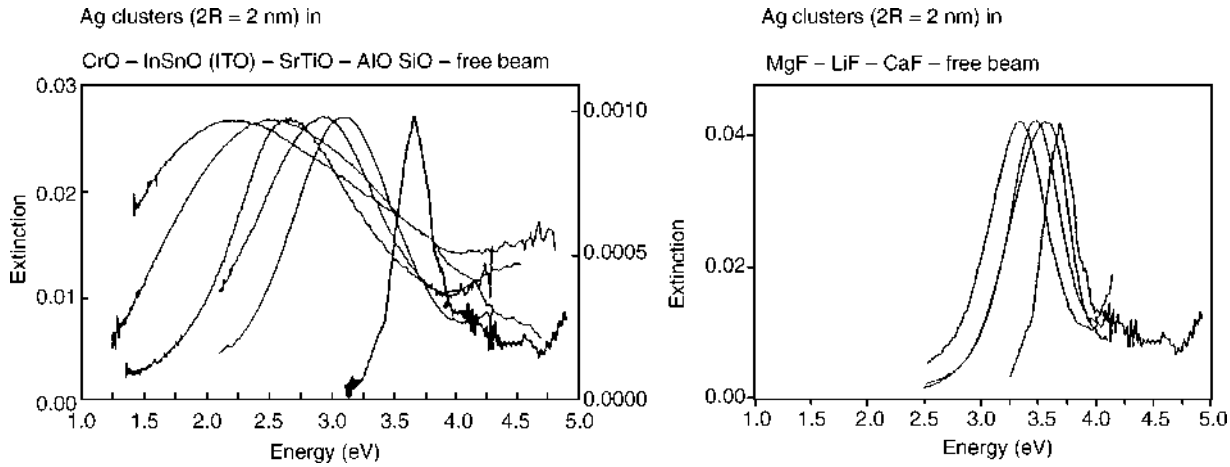

Figure 12 Measured optical extinction spectra of 2 nm Ag nanoparticles, embedded in various different embedding materials. For comparison, the spectrum of the free nanoparticle beam is added. Left side: oxidic materials, right side: fluoridic materials.

$\varepsilon_{\text{matrix}}$, which classically follow from Mie's theory and by comparingly small additional shift due to static charge transfer;

- drastic broadening (exceeding, in some cases, all known size effects), which according to eqn [14], can be described by characterizing values of the A-parameter. Recently, these parameters have been experimentally determined for Ag nanoparticles embedded in a broad variety of host materials. Their interpretation is rendered difficult, however, due to inhomogeneous broadening of varying magnitude.

We can immediately judge from Figure 12 that both shift and broadening are larger in the case of oxidic matrices than of fluoridic ones, i.e., oxides possess more interface states close to E_F of the material. In Figure 13, observed peak positions $\hbar\omega_{\max}$ and broadenings expressed in the resulting A-parameters, are summarized for a broad variety of different embedding materials.

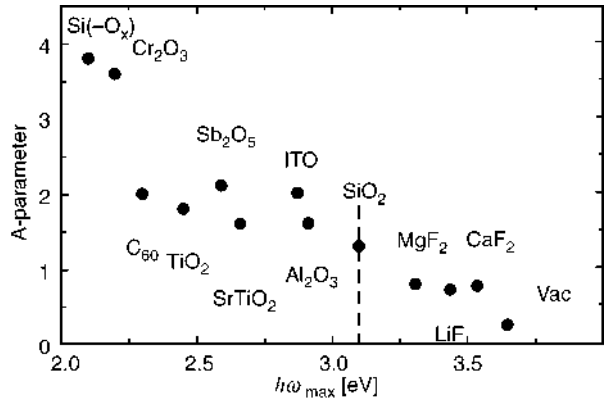

Figure 13 Silver nanoparticles ($2R = 2$ nm) embedded in various solid host matrices: measured Mie peak positions and resonance halfwidths expressed by the A-parameters.

Comparison of Plasmon Polariton Lifetimes with Femtosecond Experiments

In the last section, experimental examples were presented to show how fundamental physical properties are changed by forming heterogeneous materials. Because the Mie resonance widths are mainly a consequence of finite plasmon polariton life times τ, (in addition, there are interband contributions, in larger particles radiation damping, and dispersion influence), the latter can be evaluated from measured optical absorption spectra. This supposes the bandwidth to be essentially homogeneous. From a width analysis fitting Mie's theory numerically we obtained for Ag particles, free flying in UHV with uncontaminated free surfaces and 2 nm in size, a plasmon lifetime of $\tau = 6$ fs. Under these experimental conditions, the measured bandwidth is the homogeneous one.

Concerning the size dependences, we can extrapolate to larger particles. The longest lifetimes are expected for 15 nm particles, for which $\tau = 15$ fs was obtained by extrapolating the experimental data for 2 nm particles to this size. The same width analysis was performed for the nanoparticles embedded in matrix materials, and as a result, the lifetimes drastically decrease down to 1–2 fsec. The observed increase of the A-parameter, as shown in the previous section, is thus directly correlated to the measured decrease of the plasmon polariton lifetime. Hence, we conclude that the decrease of the lifetime is mainly due to increased phase relaxation caused by the dynamic charge transfer effect.

These results can be compared with recent femtosecond spectroscopy experiments which have recently given much impact on the experimental investigation of the dynamics of Mie resonances in silver and gold nanoparticles. In these experiments data between 6 and 10 fsec were obtained for deposited Ag and Au particles of different sizes and different substrates. Regarding the complexity of the different experiments the correspondence of the data appears convincing.

Obviously the lifetimes evaluated in the frequency domain, from our spectra of particles *in vacuo*, coincide with the lifetimes determined in the time domain. The latter, though, were measured with larger particles deposited on substrates (i.e., not from free particles but including interactions with substrates).

As a concluding remark, it may be pointed out again, that the optical properties and the heterogeneity effects presented here for the model system of nanoparticle/host-matter, are also to be found in other examples of the large field of heterogeneous materials.

List of Units and Nomenclature

Absorption bandwidth	Γ
circular frequency [s^{-1}]	ω
dielectric constant/dielectric function [–]	ε
effective electron mass [kg]	m_{eff}
electric field [$V\,m^{-1}$]	E
(electron) relaxation frequency [s^{-1}]	γ
electron state energies for initial and final state [J]	E_i, E_f
elementary charge [C]	e_0
extinction, scattering, absorption cross section [m^2]	σ_{ext}, σ_{sca}, σ_{abs}
Fermi energy [J]	E_{F}
Fermi velocity of electrons [$m\,s^{-1}$]	v_{Fermi}
intensity [$W\,m^{-2}$]	I
Kronecker energy distribution [J^{-1}]	$\delta(E)$
magnetic field [$A\,m^{-1}$]	H
norm of wave vector of electrons [m^{-1}]	k
number density of conduction electrons [m^{-3}]	n
Riccati-Bessel functions of first kind [–]	$\psi_n(z)$
particle radius [m]	R
(phase-) life time [s]	τ
Planck constant [J_s]	$\hbar$
Riccati-Bessel functions of third kind [–]	$\eta_n(z)$
scalar electromagnetic potentials [V]	Π, Γ
susceptibility [–]	χ
transition matrix element [$kg\,m\,s^{-1}$]	M_{if}
vacuum permittivity [$CV^{-1}\,m^{-1}$]	ε_0
vacuum velocity of light [$m\,s^{-1}$]	c
vacuum wavelength of light [m]	λ
wave vector of electrons [m^{-1}]	$\mathbf{k}$

See also

Scattering: Scattering Theory. **Semiconductor Physics:** Polarons.

Further Reading

Bassani F and Parravicini G (1975) *Electronic States and Optical Transitions in Solids*. Oxford: Pergamon.

Bohren CF and Huffman DR (1983) *Absorption and Scattering of Light by Small Particles*. New York: Wiley.

Born M (1960) *Optik*. Berlin: Springer-Verlag.

Duncan M (ed.) (1998) Advances in Metal and Semiconductor Clusters, vol. IV. London: JAI Press Inc.

Haberland H (ed.) (1994) *Clusters of Atoms and Molecules*. Berlin: Springer Series in Chemical Physics 56.

Nalwa HS (ed.) (2001) *Handbook of Surfaces and Interfaces of Materials*, vol. 3. San Diego: Academic Press.

Hölzl J, Schulte F and Wagner H (1979) *Solid Surface Physics*. Springer Tracts in Modern Physics 85. Heidelberg: Springer.

Hummel R and Wissmann P (eds) (1997) *Handbook of Optical Properties*, vol. II: Optics of Small Particles, Interfaces and Surfaces. Boca Raton, FL: CRS Press.

Jena P, Khanna SN, Rao BK (eds) (2000) *Cluster and Nanostructure Interfaces*. New Jersey: World Scientific.

Kerker M (1969) *The Scattering of Light and Other Electromagnetic Radiation*. New York: Academic Press.

Kreibig U and Vollmer M (1995) *Optical Properties of Metal Clusters*. Berlin: Springer Series in Material Science 25.

Nanometer Scale Science and Technology, Internatl. School of Physics 'Enrico Fermi' (2001) CXLIV. Amsterdam: IOS Press.

Quinten M (1998) *Physical and Chemical Properties of Surfaces in Nano-Particle Systems*. Aachener Beiträge zur Physik kondensierter Materie, vol. 24. Mainz, Aachen: Wissenschaftsverlag.

Ruhbahn HG (2002) *Nanophysik und Nanotechnologie*. Stuttgart: Teubner.
Schmid G (1994) *Clusters and Colloids*. Weinheim: VCH Verlag.
van de Hulst HC (1957) *Light Scattering by Small Particles*. New York: Wiley.
von Plessen G (1998) *Ultrafast Dynamics of the Surface Plasmon in Gold Particles*. Advances in Solid State Physics 253. Berlin: Vieweg.

Lightweight Mirrors

J W Bilbro, George C Marshall Space Flight Center, AL, USA

Published by Elsevier Ltd.

The desire to probe the heavens is as old as humanity; the realization that the human eye could be aided in these investigations is somewhat more recent. In May 1609, Galileo Galilei, while visiting in Venice, heard about the new 'telescope' invented by the Dutch lens-maker Hans Lipershay and decided that it would be useful in his astronomical investigations. He returned to his home in Padua and made his own telescope comprised of 4.2 cm diameter plano-convex and plano-concave lenses. With this telescope, Galileo discovered the satellites of Jupiter; got tenure; had his salary doubled – and the race was on! It did not take too long to realize that 'bigger' was 'better', and from this point forward, there has been a continuous striving to increase collecting area and resolution, thereby enabling the astronomer to 'see' farther and farther into the heavens.

The limitations of refractive optics soon prompted investigations into reflecting optics – mirrors. In 1668, Isaac Newton produced the first reflecting telescope having a 2.5 cm diameter mirror made of speculum (copper (6), tin (2), and arsenic (1)). A little over 100 years later, in 1789, William Herschel constructed his 'Great Telescope', having a 1.22 m diameter mirror made of bronze. Of course as telescopes grew, so did mirror weight. In 1845, William Parsons' 'Leviathan' was built. It had a 1.84 m diameter mirror made of speculum that weighed 3600 kg – a whopping 1354 kg m^{-2}.

Problems with corrosion in speculum and bronze mirrors led to a renewed interest in glass, and the development of highly reflective coatings soon made metal mirrors obsolete. In 1917, the mirror for the Hooker telescope arrived on Mount Wilson. Made by the Saint Gobain glassworks in France from the same material used for wine bottles, the mirror was 2.54 m in diameter, 33 cm thick and weighed over 4000 kg (789 kg m^{-2}).

The increasing weight of larger and larger mirrors produced practical problems for telescope makers and astronomers. Herschel's telescope was so unwieldy that he found it almost impossible to use and returned to smaller telescopes for most of his observations, and the 2.54 m Hooker telescope required a 100 ton mount and drive system. Obviously, if telescope apertures were to increase, steps had to be taken to make mirrors lighter. The first major success in this arena was accomplished for what was to become the Hale Telescope commissioned at the Palomar Observatory in 1948. In 1936, Corning cast a mirror made of a new borosilicate glass blend called Pyrex. The blank was lightweighted by placing hexagonal shaped obstructions in the mold which created corresponding voids in the glass. The finished mirror (Figure 1) was 5.08 m in diameter, 61 cm thick at the edge and weighed approximately 13 000 kg (641 kg m^{-2}). Even with this degree of 'lightweighting' the mounting and drive mechanisms still weighed around 182 000 kg.

Although the desire for larger ground-based telescopes provided the initial impetus for developing lightweight mirrors, the advent of the space program and the desire to put telescopes in space became the main driver in the pursuit of 'serious' lightweighting techniques. The cost of launching

Figure 1 5.08 m Pyrex™ mirror made by Corning/Caltech for the Hale Telescope. Reproduced with permission from the Palomar Observatory/California Institute of Technology.

large telescopes into space quickly becomes prohibitive as aperture size increases. Consequently, the current emphasis is on developing ultra-lightweight, deployable telescopes that permit ever-increasing apertures to be placed in orbit. Fortunately, the zero gravity environment of space permits satisfactory operation of such ultra-lightweight telescopes and therefore the extent to which a mirror can be lightweighted is primarily dependent upon the mirror material and the manufacturing techniques utilized.

Manufacturing

Mirror manufacturing can, in some sense, be divided into two general classes, one that removes material (direct generation) and one that deposits material (replication). The two classes have significantly different limitations relative to producing lightweight mirrors. Direct generation encompasses the more traditional methods of mirror manufacturing, blank fabrication, and optical surface generation, i.e., grinding and polishing. Many innovative techniques have been developed over the past decade that have not only made significant strides in improving fabrication times, optical figure, and surface finish, but have also made possible the reduction of weight to unprecedented levels.

Although extremely lightweight mirrors are being fabricated using direct generation, in the end, there are limitations in the fabrication processes that are not present in replication. This process offers the fascinating potential to create mirrors that are only a few molecules thick, thereby bringing a new meaning to the term 'lightweight optics'.

Direct Generation

Mirror blank fabrication

Casting lightweight glass mirror blanks has evolved substantially since 1936. Four 8.2 m diameter Zerodur™ mirror blanks were spun cast for the European Southern Observatory's Very Large Telescope Interferometer. The largest mirror cast to date was the 8.4 m borosilicate mirror cast in a spinning oven for the Large Binocular Telescope on Mount Graham. Although not exactly lightweight by space standards – the blank weighs 15 455 kg (266 kg m^{-2}) – it is substantially less than a conventionally cast solid blank.

It has been postulated that an areal density of around 64 kg m^{-2} probably represents the limit of lightweight glass castings. Therefore, as the demand for lighter and lighter mirrors increased, methods other than casting had to be developed. Investigations in the 1960s indicated that mirrors could be lightweighted to about 180 kg m^{-2} by first fabricating a 'honeycomb' core and then fusing plates on the front and back of the core using a high-temperature fusion process. This process was used to produce the Hubble Space Telescope mirror (Figure 2). Hubble had a 2.5 cm thick face-sheet and back-sheet and a honeycomb interior constructed of individual pieces 0.5 cm thick. These pieces were assembled and placed in a large furnace where the temperature was increased until they fused together. The resulting 2.4 m diameter mirror weighed 828 kg (183 kg m^{-2}).

During the 1980s, Frit bonding technology was developed. Unlike the fusion process, Frit allowed the bonding of glass elements at a lower temperature. This allowed the cross-sections of the glass elements to be significantly reduced. The thicknesses of face-plates were typically reduced to around 7 mm and areal densities in the 40–50 kg m^{-2} range became routine. New advances were also made in core manufacturing. Diamond tooling was developed along with computer controlled systems to machine cores from solid blanks. High-pressure water jets were developed that allowed cores to be made much faster than with diamond tooling. In addition, the water jet cutting provided the ability to create very flexible geometries optimized to minimize weight while retaining stiffness. Water jets could machine core walls that could be held to a thickness of around 1.5 mm over a core depth of 20 cm. Computer controlled diamond machining has created even thinner cores (less than 1 mm in thickness) but the process requires longer generation times.

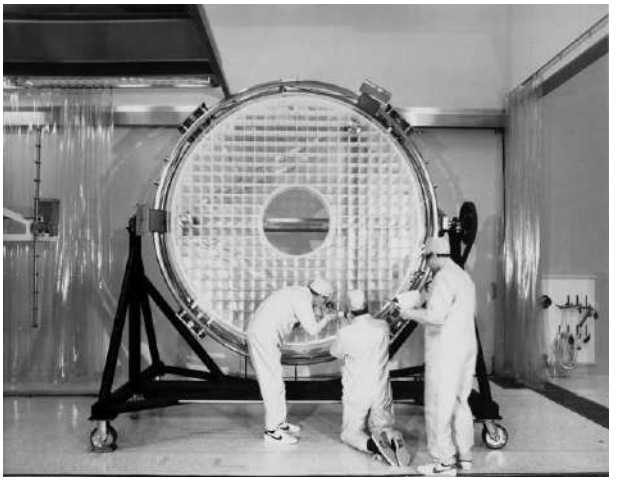

Figure 2 2.4 m ULE™ mirror made by Corning/Perkin Elmer for the Hubble Space Telescope. Photography courtesy of NASA.

Material removal

The evolution in blank manufacturing required a corresponding evolution in the process of material removal. Many new grinding and polishing techniques were developed over the years to more accurately and rapidly generate precision optical surfaces. These included adaptive force grinding, electrolytic in-process dressing (ELID), water jet figuring, computer controlled polishing, stress lap polishing, magneto-rhelogical figuring (MRF), plasma-assisted chemical etching (PACE), ion figuring, and vacuum lap figuring. Most of these processes induce localized forces on the face-sheet in the process of material removal, and for a 'honeycomb' mirror, the face-plate deflects away from the grinding/polishing tool as the tool reaches the center of the cell. This means that less material will be removed in the center of the cell than at the edge of the cell where the core wall holds the face-sheet against the tool. This creates a 'quilted' pattern in the surface that must be removed in subsequent steps. As face-sheets become thinner, quilting becomes more pronounced and the process of removing 'quilting' becomes exacerbated. For a number of years, this was the limiting factor in the development of high-precision lightweight mirrors.

Ion figuring is a noncontact process that has eliminated the problems associated with mechanical material removal. It bombards the surface of the optic with neutralized ions causing atoms to be dislodged. The process creates no deflection of the surface and therefore can remove quilting and other irregularities without generating further distortion in the surface. This has permitted 2–3 mm thick face-plates to be manufactured without concern for quilting effects. Primary limitations now being encountered are in face-plate generation and the various associated handling processes. Great care must be taken to avoid fracture. Spin casting ultra-thin face-plates is being investigated as a means of further reducing problems in the generation process. The 0.75 m mirror made for IKONOS, the high-resolution imaging satellite, represents the state-of-the-art in all-glass lightweight mirrors (28 kg m^{-2}).

Replication

Replication techniques are quite common in low-cost production mirrors and lenses for commercial purposes. However, there are also a number of techniques that hold great potential for precision mirrors as well. Chemical vapor deposition (CVD), plasma vapor deposition (PVD), plasma spray deposition, and electro-forming are all being investigated for producing lightweight precision mirrors. This class of mirror generation does not start with a mirror blank *per se*, rather it replicates a 'mandrel' that has the appropriate surface finish and the inverse of the shape desired in the final mirror. The mandrel can be made very rigid and therefore unperturbed by either gravitational forces or the localized forces generated by grinding and polishing. Extremely thin mirrors with areal densities on the order of 0.1 kg m^{-2} are currently being manufactured using replication techniques. These 'membrane' mirrors are currently being used for solar energy collection; however, recent advances in membrane quality and wavefront correction techniques hold out the promise of producing high-resolution, membrane imaging systems.

The Future?

Theoretically, single-molecule thick mirrors could be manufactured in space using self-replicating nanotechnology systems that 'grow' mirrors according to a mathematical description of the desired optical surface. These 'smart' mirrors could self-correct in the presence of physical disturbances and self-heal when damaged. They could be 'grown' to hitherto unimagined size, being limited only by the availability of raw material. Such mirrors, however, are definitely of the future and should merit re-examination 20 years hence.

Materials

A variety of materials have been developed over the years that have led to lighter and lighter mirrors. Recently, metals are coming back into play as means of fabricating stiff, lightweight mirrors. These metals include beryllium, nickel, nickel alloys, aluminum, and aluminum alloys. Composite materials are also playing a large role in lightweight mirror fabrication, both as a support structure and also as a reflecting surface (with appropriate coatings). Hybrid mirrors comprised of a combination of glass and composites have also been developed, as well as combinations of glass and beryllium. Various forms of silicon carbide have been extensively investigated and lightweight mirrors of pure silicon are being manufactured. Glass still plays a significant role in large lightweight mirrors, and various plastics are under investigation for the production of ultra-thin membrane mirrors.

Glass, Glass Ceramics, and Hybrids

The most common glass/glass ceramic materials used in large telescope mirrors are fused silica, ULE™ (Ultra-Low Expansion) glass, and Zerodur™, a glass ceramic. As was mentioned earlier, there are a variety

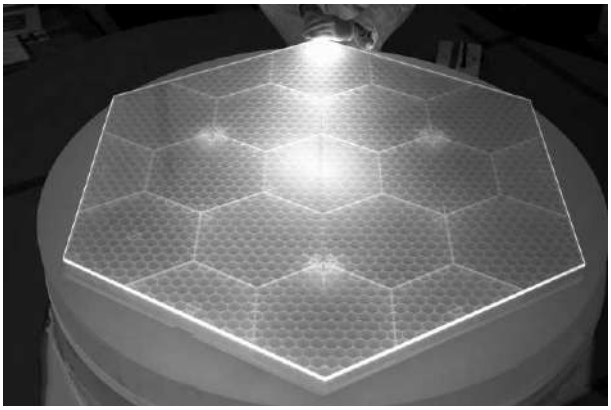

Figure 3 1.5 m ULE™ mirror made by Corning/Kodak for NASA's Next Generation Space Telescope Advanced Mirror System Demonstrator technology program. Photograph courtesy of NASA.

of methods available for the manufacture of glass mirrors depending upon the extent of lightweighting desired. Most of the current effort in lightweight mirrors is centered around hybrid mirrors. One method bonds very thin (2–3 mm) glass face-sheets to composite cores to form 'sandwich' mirrors. Another method attaches face-sheets to composite back-plane structures via actuators. In the latter case, the actuators can be used to actively control mirrors to correct for distortions. These mirrors were developed as part of the James Webb Space Telescope technology program, and typically have areal densities on the order of 15 kg m^{-2} (including back-plane structures). An all glass face-sheet and core is shown in Figure 3.

Silicon Carbide

Silicon carbide has long been a promising material for lightweight mirrors. It has many of the desirable properties of beryllium and is nontoxic. Until recently difficulties were encountered in producing mirrors larger than ~0.5 m in diameter. Of late, several manufacturers have produced large (<1 m) lightweight mirrors using different processes; sintering, liquid silicon infiltration and chemical vapor deposition.

The sintering process consists of preparing a finely ground silicon carbide powder mixed with organic binders and additives and then compacting the powder by isostatically pressing it at high pressure. The resulting compact is then machined to the desired shape and weight and the blank heated to cause agglomeration of the particles. The resulting material is known as reaction-bonded silicon carbide; it can be polished to around 30 angstroms and can be ion figured. If smoother surfaces are required, then a layer of CVD or PVD silicon carbide can be added. A 1.35 m diameter mirror has been fabricated as a demonstration for the European Space Agency's Herschel Space Observatory. The mirror was fabricated in nine segments that were then brazed together to form a unified 26 kg m^{-2} blank. This mirror was not designed to test the limits of lightweight and, consequently, there is no reason to believe that this process could not achieve an areal density comparable to the 7.8 kg m^{-2} of Cesic™ described in the following paragraph. The size of mirrors produced using this technique is currently limited to around 4 m due to the availability of the furnaces required for segment brazing.

The liquid infiltration process is used to create a biphase carbon, silicon carbide material known as Cesic™. This material is manufactured by creating a block of chopped carbon fibers embedded in a phenolic resin. The block is machined to the desired shape and then put into a furnace where the temperature is elevated until the resin is turned to carbon. The temperature is further elevated and liquid silicon infiltrated to produce silicon carbide. Bare Cesic™ can be polished to approximately 100 angstroms. In a subsequent step, a silicon carbide slurry can be applied to the blank that can be polished to around 30 angstroms. It is also possible to add a CVD or PVD silicon carbide layer to the blank that can be polished to obtain angstrom-level surfaces. A 1.04 m diameter Cesic™ mirror has been constructed for the NASA program, Solar Lite. The mirror weighs approximately 60 kg for an areal density of 56 kg m^{-2}. Again, this mirror was not designed to test the limits of lightweighting, but in a separate demonstration, an areal density of 7.8 kg m^{-2} was achieved. The size of mirrors produced using this technique is also currently limited to around 4 m due to the availability of existing furnaces required for infiltration.

The CVD process produces a SiC material that is highly polishable, yielding surface finishes on the order of angstroms. It has been used quite successfully as a coating layer, and has the potential to provide a means of replicating very lightweight mirrors by depositing a thin layer of SiC onto a mandrel. In one approach, the mandrel has the appropriate shape, but not the surface finish. In this case, once the SiC has been deposited, the mandrel is burned away and a SiC mirror blank is left. The blank is then ground and polished to produce the desired figure and finish. If the blank is too thin, this can present a problem in achieving a high-performance mirror. A second approach involves the deposition of SiC onto a figured and polished mandrel. In principle, both the figure and surface finish can be replicated. Success in this approach, however, has been somewhat limited, and although 20-angstrom surface finishes have been

replicated, they have not been repeatable. A 1 m diameter blank was manufactured using the first approach, but it too has not been repeated and the lack of success appears to be due to high stress developed in the deposition process. The residual stress in the blank causes large-scale deformation and frequently fracture.

Beryllium

The extremely high stiffness of beryllium makes it a highly desirable candidate for lightweight mirrors. Unfortunately, it is a highly toxic material and there are only a very few companies involved in the manufacture of beryllium mirrors and even fewer that produce the raw material. Blanks are manufactured by isostatically compressing beryllium powder at high temperature and pressure. The blanks are then machined and subsequently ground and polished. Early beryllium mirrors often exhibited interesting, mysterious changes in figure after seemingly benign handling events. Unpredictable changes in figure also occurred as they were cycled between ambient and cryogenic temperatures. These problems appear to have been resolved with the production of spherical powder. The older powder consisted of irregularly shaped particles and when the powder was compressed, small, irregularly shaped voids were created throughout the material. It is hypothesized that these voids changed in size and shape as the blank was stressed, thereby causing unpredictable dimensional changes in the blank. The spherical powder permits uniform, high-density packing yielding homogeneous blanks that do not appear to exhibit unpredictable dimensional changes.

The lightest-weight beryllium mirror to date ($11.8\ \mathrm{kg\,m^{-2}}$) was made for NASA's James Webb Space Telescope technology program. This technology was selected for the flight mirror. The weight reduction was achieved by machining the blank to create an open back honeycomb structure with an integral face-sheet. The ability to fabricate large beryllium mirrors is limited by the availability of facilities capable of isostatically pressing the material at the requisite pressures and temperatures. Currently, blank sizes are limited by available isostatic presses to around 1.2 m in width and 1.6 m in length. The expense of construction of larger facilities combined with limited need make it highly unlikely that any larger mirrors will be manufactured in the near term. A brazing process, however, has been demonstrated for joining segments to form a large blank. This process could, in principle, permit the fabrication of mirrors several meters in diameter. Bare beryllium can be polished to around 30 angstroms surface finishes. If smoother surfaces are required, beryllium can be vacuum deposited on the blank. Nickel may also be deposited on the blank; this has the added benefit that it can be polished without the special facilities required for the hazardous beryllium polishing operations. The drawback of using nickel is that of the bimetallic effect if the mirror is to be operated over a wide temperature range. The Vavilov Institute in Russia has developed an alternative approach to obtaining good surfaces by creating a glass that matches the coefficient of thermal expansion (CTE) of beryllium over a 50° temperature range. The glass is then fused to the beryllium blank. There have been numerous attempts over the years to produce replicated mirrors using a CVD process to deposit beryllium on a mold. So far these attempts have been unsuccessful, the process producing a material structure that no longer has high stiffness.

Silicon

Silicon as a mirror material has the advantage of being a very homogeneous material that can be polished to very smooth surfaces (angstrom level) very rapidly. A lightweight mirror has been fabricated using a silicon face-sheet brazed to a central core of silicon foam. This mirror 'sandwich' has demonstrated an areal density of $<15\ \mathrm{kg\,m^{-2}}$. Technology has been developed to braze segments together in a manner that has been shown to have no adverse effects on either polishability or cryogenic performance (i.e., no distortion across the braze). Consequently, it should be possible to manufacture lightweight silicon mirrors of 1 m diameter or larger.

Composites

There have been many attempts to manufacture mirrors from composite materials, beginning initially with graphite/epoxy and more recently with graphite/cyanate-ester. These mirror are produced by laying up graphite sheets impregnated with epoxy, or cyanate-ester on a near net shape mold and then curing the material at elevated temperatures. The optical surface is typically created by coating another mold (this time with the desired optical surface) with a passivation layer and then with epoxy and placing the near net shape composite substrate on the epoxy. When the epoxy cures, the completed mirror is separated from the mold and the epoxy surface ion milled to remove residual errors such as fiber print-through. These mirrors can be very lightweight ($\sim 5\ \mathrm{kg\,m^{-2}}$), but typically have problems with large-scale distortion due to inhomogeneities due to

the lay-up process. The epoxy surface replication can also be a problem if the mirror is subjected to temperature extremes. A 3.5 m diameter composite mirror was made as a demonstration for the FIRST program. It has suitable performance for submillimeter and infrared wavelengths, but not for shorter wavelengths. At this time, the accuracy required for large (>1 m) visible imaging systems has not yet been achieved.

Gossamer Mirrors

The term 'gossamer' refers to that category of mirrors that in and of themselves have no structural rigidity beyond spatial scales of fractions of millimeters. Several materials are currently being used for these types of materials including Kapton™ and various other polyimide films. The initial shape of the membrane is formed by a variety of techniques including vacuum forming, pressure forming, spin casting, hot forming, and others. Early mirrors were manufactured by joining together triangular-shaped segments (gores) to form the desired shape. When deployed, these mirrors require some means of applying pressure to achieve the desired shape. Some methods have involved applying force at the edge to put the mirror in tension; some involve using a clear membrane as a face-sheet and then pressurizing the combination to create a 'lens' with a reflecting inner surface. Most recently, techniques have been demonstrated involving various forms of electrostatic deflection to achieve the desired shape after deployment. At present the primary applications are in the area of solar collection, power beaming, and thermal propulsion. The mirrors are approaching the quality required for imaging telescopes, particularly for small sizes. Techniques for correcting these mirrors with 'downstream' active optics have reached the point where it is not unrealistic to imagine that membrane optics could soon reach the accuracy required for imaging. There have been two space-based demonstrations, one by the Russians – a 20 m reflector – and the other by the USA – a 14 m inflatable antenna (Figure 4).

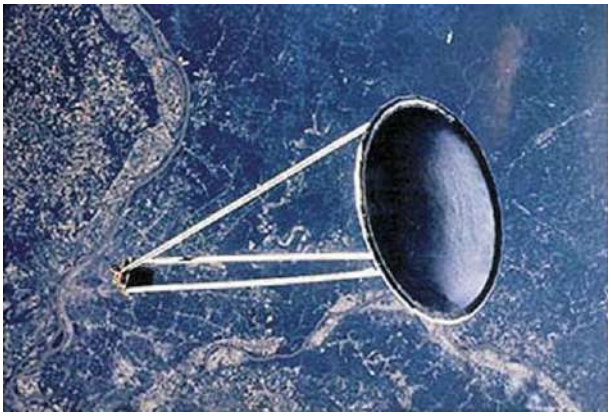

Figure 4 15 m antenna made by L'Garde for NASA's Inflatable Antenna Experiment. Photograph courtesy of NASA.

Grazing Incidence Mirrors

Broadband X-ray telescopes make use of the reflection property of materials at angles of grazing incidence. The most common design is based on the X-ray microscope objective designs of Wolter in the early part of the twentieth century. These mirrors are hyperboloids and paraboloids of revolution. This configuration provides a structural rigidity not present in normal incidence mirrors; consequently, X-ray mirrors can be made much lighter. The largest mirror of the Chandra X-ray Observatory is made from Zerodur™ (Figure 5). It is a 1.2 m diameter, 0.9 m long cylindrical mirror with a wall thickness of only 24 mm. Such a thickness in a normal incidence mirror of this size could be easily deformed.

Lightweight approaches in grazing incidence optics are almost all concentrated in various forms of replication. Recent developments in electroformed nickel alloy mirror technology have resulted in a 0.5 m diameter 0.6 m long mirror that weighs only 1.2 kg (Figure 6). A similar-sized mirror in the Chandra telescope weighs over 100 kg. Extremely lightweight cylindrical mirrors have been made from graphite/epoxy and from CVD SiC. Various investigations are underway in producing segments of cylinders that are then mounted to form a complete

Figure 5 1.2 m Zerodur™ made by Schott Glass Werken/Hughes Danbury Optical Systems for NASA's Chandra X-ray Observatory. Photograph courtesy of NASA.

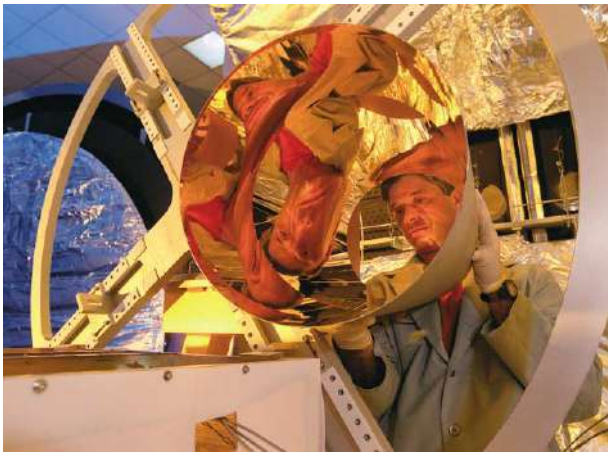

Figure 6 0.5 m electro-formed nickel–cobalt technology mirror made by NASA's George C. Marshall Space Flight Center for NASA's Constellation X-ray Mission technology program. Photography courtesy of NASA.

cylindrical mirror. These techniques involve a variety of materials including aluminum foil, and beryllium plates. In segment replication, the mandrel is passivated and then coated with the desired reflecting material of the final optic (the passivation process prevents the coating from adhering too strongly to the mandrel.) After the mandrel has been coated with the reflecting material, an epoxy coating is applied and a preformed shell segment is placed on the epoxy. After the epoxy cures, the segment can be removed from the mandrel along with the desired reflective surface (typically gold). In this case, the preformed shell does not have to have as precise a figure, or surface as the mandrel, since the epoxy will accommodate minor differences.

See also

Geometrical Optics: Lenses and Mirrors. **Imaging:** Adaptive Optics; Wavefront Sensors and Control (Imaging through Turbulence). **Instrumentation:** Astronomical Instrumentation; Telescopes. **Optical Materials**: Smart Optical Materials.

Further Reading

Bilbro J (2001) Optics in orbit. *SPIE OE Magazine* (Aug. issue).

Breckinridge JB and Jakobsen P (eds) (2000) UV, Optical, and IR Space Telescopes and Instruments. *Proceedings of SPIE* 4013.

Chen PC, Bowers CW, Content DA, *et al.* (2000) Advances in very lightweight composite mirror technology. *Optical Engineering* 39(9).

Ealey MA (ed.) (1995) Silicon carbide materials for optics and precision structures. *Proceedings of SPIE* 2543.

Engelhaupt DE, Ramsey BD, O'Dell SL, *et al.* (2000) New alloys for electroformed replicated X-ray optics. *Proceedings of SPIE* 4138: 22.

Goodman WA and Jacoby MT (2000) Dimensionally stable ultra-lightweight silicon optics for both cryogenic and high-energy laser applications. *Proceedings of SPIE* 4198: 260–270.

Jenkins C (vol. ed.) (2001) Gossamer spacecraft: membrane and inflatable structures technology for space applications. In: Zarchan P (series ed.) *Progress in Astronautics and Aeronautics* 191. Reston, VA: American Institute of Aetronautics and Astronautics.

Kendrick SE, Reed T and Streetman S (2001) In-process status of the 1.4-m beryllium semi-rigid Advanced Mirror System Demonstrator (AMSD). In: Stahl HP (ed.) *Optical Manufacturing and Testing*, vol. IV. *Proceeding of SPIE* 4451: 58–66.

Mathews G (2001) Semi-rigid active mirror technology advancements. *Proceedings of SPIE* 4451.

Müller C, Papenburg U, Goodman WA and Jacoby M (2000) C/SiC high precision lightweight components for optomechanical applications. *Proceedings of SPIE* 4198: 51–57.

Truemper JE and Aschenbach B (eds) (2000) X-Ray optics, instruments, and missions III. *Proceedings of SPIE* 4012.

Measurement of Optical Properties of Solids

P Lucas, University of Arizona, Tucson, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The interaction of light with solids takes place through different mechanisms, depending on the type of material and the range of wavelength investigated. Insulators or dielectrics are typically transparent to visible light while most semiconductors are opaque to visible light yet transparent to infrared radiations; in contrast metallic solids appear shiny because they reflect all wavelength up to the ultraviolet region. The optical properties of a solid depends on its chemical composition and its structural properties and vary for every material, though, one parameter, the complex refractive index **n**, is sufficient to characterize entirely the optical properties of a specific material.

In this article we will briefly review the physics underlining the nature of the refractive index **n**, in order to understand its relationship to the observable quantities such as absorption, reflection, and transmission routinely measured in optics experiments. A description of the experimental techniques, as well as a physical interpretation of the data, will also be presented.

Origin of Light Interaction with Matter

Light is described as a transverse electromagnetic wave consisting of an electric and magnetic field oscillating perpendicular to the direction of propagation at optical frequencies within $10^{13} \rightarrow 10^{17}$ Hz. This range of frequencies comprises the ultraviolet, visible, and infrared domain. On the other hand, solid mediums are composed of charged particles, negative electrons, and positive ions producing electric dipoles, which can be polarized under the action of an electric field. Hence, when electromagnetic radiation impinges upon a material it interacts by polarizing the molecular units, producing oscillating dipole moments. This interaction results in several observable optical phenomena such as reflection, transmission, absorption, or scattering, which will be described in more detail in a following section. The classical model of light propagation assumes that the oscillating electric field can interact with several different types of dipole oscillators within the material. Different dipoles are usually accessed by light wave from different frequency range depending on their mass.

Electronic Oscillators

Lorentz originally proposed that the electrons bound to the nucleus of an atom behave as oscillators. In this representation, electrons behave as if they were held by springs and subjected to a linear restoring force, therefore oscillating up and down (Figure 1a). This representation of an atom might seem to contradict the conventional picture of electrons orbiting around the nucleus; however, it is valid as far as light–matter interactions are concerned. The resonant frequency of such an oscillator is proportional to the inverse of its reduced mass μ defined as

$$\frac{1}{\mu} = \frac{1}{m_e} + \frac{1}{m_N} \qquad [1]$$

where m_N and m_e are the mass of the nucleus and electrons, respectively. The resonant frequency ω_0 is then related to μ according to

$$\omega_0 \propto \sqrt{\frac{1}{\mu}} \qquad [2]$$

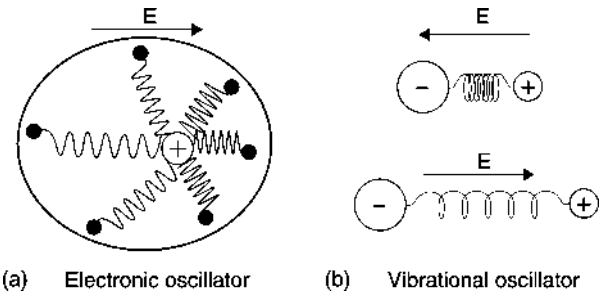

Figure 1 (a) Classical oscillator model of a polyelectronic atom. The electrons depicted as black dots are bound to the positive nucleus by springs, which represent the electrostatic restoring forces between the charged particles. The electronic oscillators are distorted under the effect of an electric field. (b) Classical model of a diatomic vibrational oscillator. The chemical bond is represented as a spring and the charged ions vibrate around their equilibrium position under the effect of an oscillating electric field.

Considering that $m_N \ll m_e$ we can assume that $\mu = m_e$ and show that the smaller mass of the electron determines the resonant frequency of the oscillator. Hence, resonant electronic oscillations are accessed at higher frequencies in the ultraviolet and visible region of the spectrum. In solids, the quantum theory shows that these resonances correspond to excitation of electrons across the gap between a full valence band and an empty conduction band.

Vibrational Oscillators

The second type of dipoles contained in the medium are based on pairs of charged atoms vibrating around their equilibrium position in the solid. Every atom has a different electronegativity, hence the electronic density on atomic pairs is distorted and result in a charge imbalance, which can couple to the electric field of light. This generates oscillating dipoles, which involves the motion of atomic nucleus (Figures 1b). The nuclear mass being several orders of magnitude larger than the electronic mass, we can show by analogy with eqns [1] and [2] that the resonant vibration will occur at lower frequencies in the infrared region. In solids, the quantum theory shows that the lattice atoms vibrate cooperatively and generate quantized lattice waves called phonons.

Free Electron Oscillators

In contrast with the bound electrons from the electronic oscillator model, free electrons can move without being subjected to a restoring force. These correspond to the conduction electrons present in metals and to a smaller extent in doped semiconductors. The free electron oscillator model is therefore

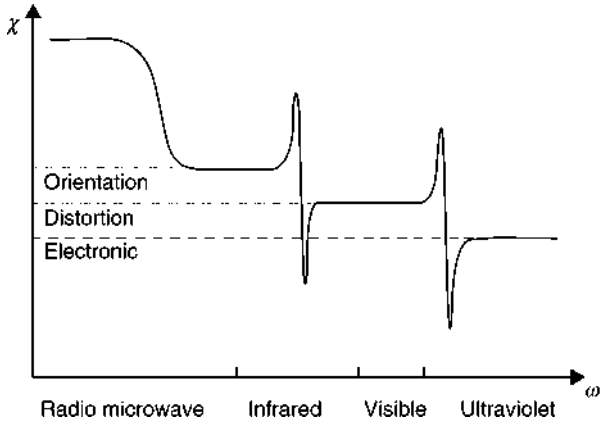

Figure 2 Contributions to the polarizability of a medium as a function of light frequency. At optical frequencies, the molecules cannot re-orient fast enough to follow the reversing electric field and the contribution of dipole orientation is lost.

principally used to describe the optical properties of metals such as reflectivity. The free electrons are well described as conventional oscillator with a natural resonant frequency $\omega_0 = 0$.

The relationship between the applied electric field E and the resulting polarization of the medium P is expressed through the dielectric susceptibility χ as

$$P = \varepsilon_0 \chi E \qquad [3]$$

where ε_0 is the permittivity of vacuum. We can generally recognize three contributions to the dielectric susceptibility or the polarizability of an optical material (Figure 2). Applied fields in the radio or microwave region oscillate slowly enough that polar molecular units can reorient and align in the direction of the field. The characteristic time for orientational motion is fairly long and this contribution is quickly lost when the molecules cannot follow the field oscillating at higher frequency. When the field reaches optical frequency, the two remaining contributions are the vibrational and electronic dipoles mentioned above. The vibrational dipoles involve a distortion of the chemical bond and are the slowest process occurring only in the infrared region. This contribution is lost at visible frequencies and leaves only the fast electronic polarization processes.

If we treat it rigorously, the polarization P is actually expressed as a function of higher order terms according to

$$P = \varepsilon_0\left(\chi E + \chi^{(2)}EE + \chi^{(3)}EEE + \ldots\right) \qquad [4]$$

This gives rise to nonlinear effects such as self-focusing and second or third harmonic generation. However, the higher order susceptibilities $\chi^{(n)}$ are very small and these effects are only significant for very high intensity light source such as lasers. In the following text we will only consider linear optical processes.

The Dipole Oscillator Model (Lorentz Oscillator)

In the Lorentz model, the bound electrons behave as harmonic oscillators and are constrained to vibrate at their natural resonant frequency ω_0. These oscillators are also submitted to a damping force corresponding to a resistance to the motion due to collisional processes. The equation of motion is then

$$m\left(\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \gamma\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2 x\right) = qE \qquad [5]$$

where x is the displacement along E, γ is the damping factor, m is the mass of the electron and q its charge. The passage of an electromagnetic wave through a medium exerts an oscillating force on the electrons causing them to vibrate up and down. In other words these electronic dipoles are submitted to a forced oscillation induced by an electric field varying periodically with time. The electric field felt by the atomic dipoles is written as

$$E = E_0 e^{i\omega t} \qquad [6]$$

where ω is the frequency of the light.

The electric field of the light wave then drives the dipole oscillations at its own frequency ω and the dipoles displacement oscillates according to

$$x = x_0 e^{i\omega t} \qquad [7]$$

Solving eqn [5] for x in terms of E gives the expression for the displacement of the electrons

$$x = \frac{q/m}{\omega_0^2 - \omega^2 + i\gamma\omega}E \qquad [8]$$

The atomic dipole moment p induced by the electrons displacement is therefore

$$p = qx = \frac{q^2/m}{\omega_0^2 - \omega^2 + i\gamma\omega}E \qquad [9]$$

And for a medium with N atoms per unit volume the total polarization P is

$$P = N\frac{q^2/m}{\omega_0^2 - \omega^2 + i\gamma\omega}E \qquad [10]$$

Our result in eqn [10] only contains one natural frequency ω_0 and therefore only accounts for one

single type of oscillator in the medium. However, normal optical mediums are composed of different kinds of atoms, each with several characteristic resonant frequencies ω_j. In our model, we can simply treat these multiple resonances by considering that each oscillator acts separately and add the contributions of all oscillators.

$$P = \left(\frac{q^2}{m} \sum_j \frac{N_j}{\omega_j^2 - \omega^2 + i\gamma_j\omega} \right) E \qquad [11]$$

Comparing eqns [3] and [11] then gives an expression for the polarizability χ of the medium

$$\chi = \frac{q^2}{\varepsilon_0 m} \sum_j \frac{N_j}{\omega_j^2 - \omega^2 + i\gamma_j\omega} \qquad [12]$$

Additionally, the polarization P is associated with Maxwell's displacement vector, D, through the relationship

$$D = \varepsilon_0 E + P \quad \text{and} \quad D = \varepsilon E \qquad [13]$$

where ε is the permittivity of the medium. The permittivity ε is normally expressed relative to the permittivity of vacuum ε_0 in terms of the dimensionless quantity $\varepsilon_r = \varepsilon/\varepsilon_0$. Where ε_r is the familiar dielectric constant of a material, which is related to the refracted index n as $n = \sqrt{\varepsilon_r}$.

Combining eqns [3] and [13] we obtain an important relationship linking χ to ε_r, and in turn χ to n.

$$\varepsilon_r = 1 + \chi \quad \text{or} \quad n^2 = 1 + \chi \qquad [14]$$

Finally, combining eqns [12] and [14] we obtain the expression for n in terms of the oscillator model.

$$n^2 = 1 + \frac{q^2}{\varepsilon_0 m} \sum_j \frac{N_j}{\omega_j^2 - \omega^2 + i\gamma_j\omega} \qquad [15]$$

The Refractive Index

Two important points should be made concerning the result of eqn [15]. The refractive index is a wavelength-dependent quantity and is a complex quantity. The complex refractive index $\mathbf{n}$ is indeed usually expressed as

$$\mathbf{n} = n + i\kappa \qquad [16]$$

where n is the real refractive index also defined as the ratio of the wave velocity in vacuum to the velocity in the medium $n = c/v$. And κ is the extinction coefficient, which is directly related to the absorption coefficient α as we shall see in the next section.

As we mentioned in the introductory section, the complex refractive index $\mathbf{n}$ is sufficient to characterize the optical properties of the solid. The real part of the index describes the change in velocity or wavelength of a wave propagating from a vacuum into a medium while the imaginary part is a measure of the dissipation rate of the wave in the medium.

By inspection of eqn [15] we can predict the form of the variation of n with wavelength. At frequencies ω below the resonant frequency ω_j, the terms ω^2 and $i\gamma_j\omega$ are much smaller than ω_j^2 and to a first approximation n is constant. When ω reaches a resonant frequency ω_j, the term $(\omega_j^2 - \omega^2)$ goes to zero and since the term $i\gamma_j\omega$ is very small, the value of n increases very sharply and exhibits a resonance line as depicted in Figure 3.

On closer inspection of eqn [15] however, we can see that n is not exactly constant in between resonance lines. As ω rises, the denominator slightly decreases and therefore n rises slowly with frequency. The refractive index is then slightly larger for wave of higher frequency. This is why a glass prism bends blue light more strongly than red light. This phenomenon is called dispersion as waves of different frequencies are 'dispersed' by the prism.

Another important point raised by eqn [15] is that because $(\omega_j^2 - \omega^2)$ goes to zero at a resonant frequency, the damping term $i\gamma_j\omega$ dominates and n becomes almost entirely imaginary. $\mathbf{n}$ is then mostly governed by the extinction coefficient κ. The damping associated with the coefficient γ_j corresponds to a friction force causing a loss of energy, which becomes the dominant effect during resonance. The resonance region is then associated with strong attenuation or

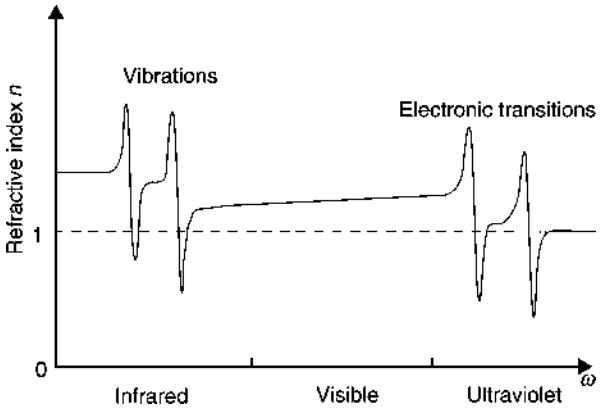

Figure 3 Variation of the refractive index of a hypothetical dielectric solid in the optical frequency range. At high frequency, ω is larger than all resonant frequencies and the refractive index n reaches unity.

absorption of the wave. In contrast, between resonant frequencies, the refractive index is almost entirely real as the imaginary part $i\gamma_j\omega$ is negligible in comparison with $(\omega_j^2 - \omega^2)$. Absorption is then very small and that region corresponds to the transparency domain of the material. In fact, we can assume that $\mathbf{n} = n$ in the transparent region, that is why tables of optical constants of transparent optical materials only list the real part of the refractive index n. However, at resonant frequencies on either side of the transparency domain, absorption becomes significant and the imaginary part of the refractive index must be taken into account so that $\mathbf{n} = n + i\kappa$.

The refractive index $\mathbf{n}$ being a complex quantity it ensues that ε_r must also be complex, since $n = \sqrt{\varepsilon_r}$. Hence we also define the complex dielectric constant as

$$\varepsilon_r = \varepsilon_1 + i\varepsilon_2 \qquad [17]$$

Both parameters describe equivalently the optical properties of a solid so that $\mathbf{n}$ or ε_r are called the optical constants of the material. Note that it is an unfortunate misnomer since they vary distinctly with wavelength (Figure 3).

Using equation [17] and $n = \sqrt{\varepsilon_r}$, we can establish the relationships between the real and imaginary parts of $\mathbf{n}$ and ε_r.

$$\varepsilon_1 = n^2 - \kappa^2 \qquad [18]$$

$$\varepsilon_2 = 2n\kappa \qquad [19]$$

We can see that $\mathbf{n}$ and ε_r are not independent variables but that the real and imaginary part of ε_r can be calculated, knowing the real and imaginary part of $\mathbf{n}$ and vice versa.

Moreover, it can be shown that the real and imaginary parts of each individual parameter are not independent either but can be computed from one another with use of the following Kramers–Kronig relations.

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi}\int_0^\infty \frac{\omega'\varepsilon_2(\omega')}{\omega' 2 + \omega^2}\,\mathrm{d}\omega' \qquad [20]$$

$$\varepsilon_2(\omega) = -\frac{2\omega}{\pi}\int_0^\infty \frac{\varepsilon_1(\omega')}{\omega' 2 + \omega^2}\,\mathrm{d}\omega' \qquad [21]$$

In these equations, only the principal value of the integral is calculated so that ε_1 can be computed if ε_2 is known over all frequencies and vice versa. The same equation connect n and κ together.

This analysis shows that it is only necessary to know one real or one imaginary part over a wide frequency range to determine all others. The interaction of light with matter is in fact characterized by a single independent real quantity.

Absorption

All electromagnetic phenomena are governed by Maxwell's equations. The form of an electromagnetic wave propagating through a medium is obtained by solving Maxwell's equation. For a wave propagating along z through a homogeneous, optically isotropic medium the form of the oscillating electric field is given by:

$$E = E_0 e^{i(kz - \omega t)} \qquad [22]$$

where k is the wave vector and ω the angular frequency of the light. For a wave traveling into a transparent medium of refractive index n, the relationship between k and ω is

$$k = \frac{n\omega}{c} \qquad [23]$$

However, for an ordinary optical material there are regions of absorption and the refractive index is therefore complex. We should then rewrite eqn [23] as

$$k = (n + i\kappa)\frac{\omega}{c} \qquad [24]$$

By substituting this value of k in eqn [22] we obtain an expression with two exponential terms

$$E = E_0 e^{-\frac{\omega\kappa z}{c}} e^{i\left(\frac{\omega n z}{c} - \omega t\right)} \qquad [25]$$

The term $e^{i\left(\frac{\omega n z}{c} - \omega t\right)}$ represents a wave traveling at a speed $v = c/n$, while the term $E_0 e^{-(\omega\kappa z/c)}$ represent the amplitude of this wave which decays exponentially with distance z. The intensity I of the wave is proportional to the square of the amplitude so that

$$I \propto e^{-\frac{2\omega\kappa z}{c}} \quad \text{or} \quad I \propto e^{-\alpha z} \qquad [26]$$

where α is the familiar absorption coefficient routinely measured by absorption spectroscopy. Hence we obtain an important relationship relating the observable quantity α to the imaginary part of the refractive index κ according to

$$\alpha = \frac{2\kappa\omega}{c} \qquad [27]$$

Local Field Correction

It should be pointed out that the derivation of our refractive index expression eqn [15] assumes that only the electric field of the light affects the polarization of the dipole oscillators. However, this

assumption is not correct in dense materials because of the contribution from the electric field of neighboring atomic dipoles. The field generated by each dipole will affect all other dipoles in the vicinity. It is possible to approximate this contribution by accounting for the effect of dipoles within a local sphere. The result of this approximation is known as the Clausius–Mossotti equation, which is accurate for isotropic medium such as glasses and cubic crystals

$$\frac{n^2-1}{n^2+2} = \frac{\chi}{3} \qquad [28]$$

Measurable Optical Parameters

Four things can happen to a beam of light propagating through a thick slab of optical material (Figure 4). Some of the light can be reflected at the surface of the solid, some can be absorbed by the sample, some can be scattered in different directions, and some of the light can be transmitted through the sample. The effect on the light beam, resulting from these phenomena, can be quantified by a number of optical coefficients, which characterize the macroscopic properties of the material.

Reflection of light at the surface of a solid is described by the reflectance R defined as the ratio of the reflected intensity I to the incident intensity I_0. For a beam falling perpendicularly on a flat surface, the reflection is called specular and is governed by the complex refractive index according to the Fresnel equation

$$R = \frac{I}{I_0} = \frac{(n-1)^2+\kappa^2}{(n+1)^2+\kappa^2} \qquad [29]$$

This provides us with a second important formula relating a measurable quantity (R) to the optical constant of the material. For measurements performed within the transparency region of a solid, the value of κ is much less than n and eqn [29] reduces to the more familiar form

$$R = \frac{(n-1)^2}{(n+1)^2} \qquad [30]$$

Absorption occurs when the light frequency reaches the natural resonance of some dipole oscillators in the medium. The energy of the electromagnetic wave is partly transferred to the material and generally dissipated in the form of heat. The light intensity is therefore attenuated as it propagates through the material and the attenuation efficiency is quantified by the absorption coefficient α. The intensity decreases exponentially with pathlength z according to Beer's law

$$I(z) = I_0 e^{-\alpha z} \qquad [31]$$

where I_0 is the incident intensity. The imaginary part κ of the optical constant can then be directly obtained from measurements of α using eqn [27].

Scattering is the process whereby light is redirected in different directions due to microscopic density fluctuation in the sample produced by defects, impurities, or structural inhomogeneities. The intensity of light propagating in the forward direction is attenuated by the scattering event and it can be quantified in a way equivalent to absorption. The intensity has an exponential dependence on pathlength z analogous to Beer's law

$$I(z) = I_0 e^{-Sz} \qquad [32]$$

where S is the scattering coefficient. When the scattering center is smaller than the wavelength of light this phenomenon is called Rayleigh scattering and the scattering coefficient S vary with the inverse fourth power of the wavelength

$$S(\lambda) \propto \frac{1}{\lambda^4} \qquad [33]$$

By measuring light attenuation we cannot tell the difference between absorption and scattering and the total attenuation is $\alpha_{\mathrm{Tot}} = \alpha + S$. However, the scattering contribution is generally much weaker than the absorption and can be neglected so that $\alpha_{\mathrm{Tot}} = \alpha$.

Transmission occurs when the light is neither reflected, absorbed, or scattered. The beam is then transmitted through the sample and the fraction of light exiting the back surface is quantified as the transmittance T. If we disregard scattering in comparison to absorption then conservation of energy

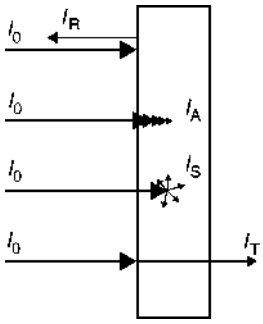

Figure 4 Representation of the four main optical processes happening to a beam as it propagates through a slab of transparent material.

require that

$$R + T + A = 1 \qquad [34]$$

where A is the fraction of light absorbed. The transmittance can then be compiled from R and α for an incident beam I_0 traveling across a sample of thickness l. In this case we must consider the reflection of the light on the front surface as well as on the back surface when it exits the transparent medium. The expression for the transmittance accounting for dual reflection is

$$T = (1 - R)^2 e^{-\alpha l} \qquad [35]$$

It should be noted that a rigorous treatment of eqn [35] should account for multiple reflections between the front and back surface in the interior of the sample. In this case, interference effects complicate the determination of **n** and a modified version of eqn [35] can be computed by summing up the intensities due to the contribution of the multiple reflections.

Transmission and reflection measurement are readily obtained using conventional spectrometers. Two types of spectrometers must be used to cover the entire optical spectral region. FTIR (Fourier Transform Infra Red) spectrometers use a glow bar as the light source, which enable it to cover the infrared and near infrared domain. The UV-VIS (ultraviolet visible) double beam spectrometers use a tungsten and deuterium lamp and cover the spectrum from the ultraviolet down to the near infrared. Using eqns [27], [29] and [35], the reflection and transmission measurements thus obtained are used to calculate the real and imaginary component of the optical constant over the entire optical frequency range.

An example of transmission measurement is shown in Figure 5, for a dielectric and a semiconductor. The two spectra have the same principal features. At short wavelength the edge of the transparency window is due to absorption of energy by the valence electrons. In classical terms this corresponds to the resonance of electronic oscillators, and in quantum mechanical terms to the excitation of electrons from the valence to the conduction band. This edge is due to a sharp increase of κ, which extends over a significant range of frequency corresponding to the domain of opacity or high absorption. The wavelength at the edge determines the minimum energy of the photon necessary to promote an electron across the bandgap E_g. Insulators have a large bandgap and appear transparent to the human eye because visible light is not energetic enough to promote an electron across E_g. In contrast, semiconductors have a smaller bandgap and appear black (opaque) because all the visible light is absorbed to induce electronic transitions. On the low frequency side the transparency window is limited by atomic vibrations in the solid structure. The wavelength of the vibrations increases with the mass of the atomic oscillators. Semiconductor compounds have higher atomic masses and therefore transmit at longer wavelength in the infrared while insulator have lower atomic mass and exhibit their vibrational absorption edge at shorter wavelength.

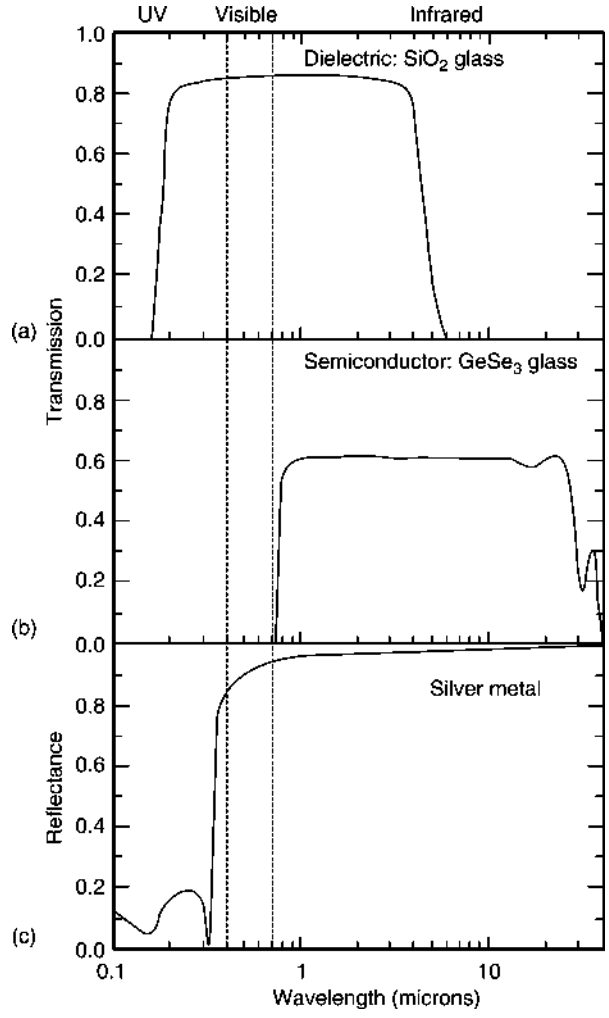

Figure 5 (a) Transmission spectrum of SiO_2 glass, a standard dielectric solid. (b) Transmission spectrum of $GeSe_3$ glass, a semiconductor ($E_g = 1.6$ eV). (c) Reflectance spectrum of silver metal.

In between the two absorption regions, κ is very small and the material is transparent. The refractive index is almost entirely governed by the real part n. Eqn [15] shows that n is proportional to the number N of electronic oscillators. The refractive index will then be higher for heavy atoms with a high electronic

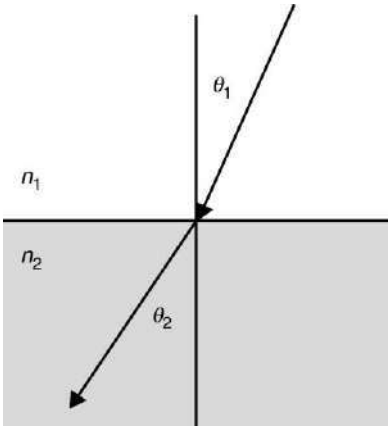

Figure 6 Refraction of a beam propagating across the interface between two mediums of different refractive index. The incident angle and the refraction angle are related to the index of refraction through Snell's Law.

density. Consequently, semiconductors exhibit a maximum transmission of only 60% because of significant surface reflectivity (eqn [35]) resulting from their higher refractive index n (eqn [30]).

Most applications of materials in optics are only concerned with the transparency domain so that the useful optical constant is reduced to the real part n. Hence a number of experimental techniques have been developed to obtain n, often by measuring angles of refraction with the use of Snell's law

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \qquad [36]$$

where θ_1 is the angle of the beam incident in the medium of index n_1, and θ_2 the angle of the beam refracted at the interface with a medium of index n_2 (Figure 6). The basic principle is to compare the sample with a standard glass of known refractive index and measure the refraction angle at their interface. The sample's index can then be obtained following eqn [36]. Since n varies with wavelength, these techniques are normally performed with monochromatic light and provide n at a single wavelength. Among these methods are: the *Abbe refractometer*, using a glass hemisphere as a standard, the *V-block refractometer*, using a V shaped prism as a standard, and the *prism goniometer* based on the relative deviation between a known glass prism and a prism shaped sample.

Ellipsometry is another technique widely used to measure the index n. This technique is normally used on thin films deposited on substrate, as it allows to simultaneously measure the thickness and the refractive index. However, the method can also be applied to bulk samples. The principle of ellipsometry is based on measuring the change in polarization of a beam reflected off the sample surface as a function of incidence angle. The reflected light is elliptically polarized to an extent depending on n. This technique is especially useful to measure n at frequency range above the absorption edge where the sample is highly absorbing.

Metals

The optical properties of metals are mainly characterized by their very high reflectivity, which causes their shiny appearance. All metals reflect light in the infrared and visible region up to a cutoff frequency in the ultraviolet. This critical frequency is called the plasma frequency ω_p. The physical significance of the plasma frequency can be understood using the Drude–Lorentz model of the free electron oscillator. As mentioned previously, the free electron can be described accurately as dipole oscillators with a resonant frequency $\omega_0 = 0$. If we consider a system with negligible damping, we can introduce $\gamma = 0$ and by assuming all free electrons to be equivalent we can consider only one type of oscillators so that eqn [15] reduces to

$$n^2 = 1 - \frac{\omega_p^2}{\omega^2} \qquad [37]$$

where

$$\omega_p^2 = \frac{Nq^2}{\varepsilon_0 m_e} \qquad [38]$$

Equation [37] means that in the low frequency domain $\omega < \omega_p$, the refractive index n must be imaginary ($n^2 < 0$) and in the high frequency domain $\omega > \omega_p$, n is a real positive number. The low frequency region is therefore dominated by the extinction coefficient κ and the reflectance R, given by eqn [29], is essentially unity. In the high frequency limit, R decreases and n is real. These features are shown in Figure 5c. Metals are opaque and highly reflective below the plasma frequency, while they become transparent in the ultraviolet.

List of Units and Nomenclature

A	absorbance
c	light velocity in vacuum
D	electric displacement
E	electric field
I	intensity of light (power per unit surface)
k	wave vector
m_e	mass of electron
m_N	mass of nucleus
$\mathbf{n}$	complex refractive index
n	real refractive index

p	dipole moment
P	polarization
q	charge of electron
R	reflectance
S	scattering coefficient
T	transmittance
v	phase velocity
α	absorption coefficient
ε_0	permittivity of vacuum
ε_r	relative dielectric constant
κ	extinction coefficient
λ	wavelength
μ	reduced mass
χ	dielectric susceptibility
ω	angular frequency
ω_0	resonant frequency
ω_p	plasma frequency

See also

Instrumentation: Ellipsometry; Spectrometers. **Scattering:** Scattering from Surfaces and Thin Films.

Further Reading

Abeles F (1972) *Optical Properties of Solids*. North Holland Publishing Company.

Born M and Wolf E (1999) *Principles of Optics*, 7th edn. Cambridge University Press.

Dragoman D and Dragoman M (2002) *Optical Characterization of Solids*. Springer.

Feynman RP (1970). *Lectures on Physics*, vol. I–III. Addison–Wesley.

Fox M (2001) *Optical Properties of Solids*. Oxford University Press.

Hecht E (1998) Optics, 3rd edn. Addison–Wesley.

Klingshirn CF (1997) *Semiconductor Optics*. Springer.

Nazarewick W, Rolland P, Dasilva E and Balkanski M (1962) ABAC chart for fast calculation of the absorption and reflection coefficients. *Applied Optics* 1(3): 369–370.

Palik ED (1985) *Handbook of Optical Constants of Solids*. Academic Press.

Savage JAR (1990) Signals radar. *Optical Materials* 1:323–419.

Simmons CJ and El-Bayoumi OH (1993) *Experimental Techniques of Glass Science*. The American Ceramic Society.

Simmons JH and Potter KS (2000) *Optical Materials*. Academic Press.

Optical Glasses

J E Shelby, Alfred University, Alfred, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

While advances continue in this field, some optical glasses have histories which can be traced to the seventeenth century. These glasses are produced by direct melting of raw materials of relatively high purity. Modern, or nontraditional, optical glasses are more commonly used to transmit light outside the visible region and are often made by nontraditional methods, including many variations on chemical vapor deposition, reactive melt processing, and sol–gel techniques. These glasses are used for ultraviolet and infrared lenses and other optical components, for optical fibers, for glass lasers, and for photonic devices such as optical isolators. Purity requirements and production demands for these glasses are significantly greater than those for traditional optical glasses.

Traditional Optical Glasses

Traditional optical glasses are primarily used as components such as lenses, prisms, or windows for optical systems operating in the visible region. These applications require glasses of high homogeneity, with precisely known refractive indices and dispersions. Glasses, as defined in this article, are nonmetallic, inorganic materials which have no long-range atomic structure and which display a temperature region of time-dependent properties.

Crown and Flint Glasses

Traditional optical glasses are designated as either crown or flint glasses. Crown glasses were originally based on the soda-lime-silica ternary system and were developed for use as window glass. The crown process involved spinning a spherical gob of molten glass to form a flattened disk which was much thicker in the center than at the edges. The edges of these disks were used for window panes, while the 'crown' in the center was often used as a crude lens. Crown glasses have low refractive indices and low to moderate dispersions.

The first flint glasses were potash-lead-silicate glasses prepared using English 'flint' as the source of silica. The presence of lead in these glasses results in

higher refractive indices and dispersions than those of crown glasses.

The designations of crown and flint have since been expanded to include glasses from many other compositional systems. Glasses with lower refractive indices and dispersions are designated as crowns, while those with higher refractive indices and dispersions are designated as flints. Modern crown glasses are based on silicate, phosphate, and borosilicate systems. Additions of fluorine are particularly useful for production of low refractive index, low dispersion glasses, and for adjusting these parameters to meet specific needs. The lowest combinations of refractive index and dispersion are found for glasses based on BeF_2, while the highest combinations of refractive index and dispersion are found in rare earth borate systems. Other common optical glasses contain barium or lanthanum, which can be used to produce either crown or flint glasses, depending on the other components of the glass.

The need for glasses with different combinations of refractive index and dispersion can be traced back to the development of the telescope. Use of lenses made of a single glass made it difficult to obtain sharp images. Newton recognized that this problem was due to divergence of light of different wavelengths while passing through a lens. He felt that this problem was inherent to lenses and could not be eliminated. Others eventually recognized that a compound lens containing a positive crown glass and a negative flint glass would result in cancellation of the divergences and produce a sharper image. While this development did not totally eliminate the colored fringes in telescope images due to the presence of the secondary spectrum, it did lead to major improvements in optical imaging. Binary compound lenses are still fashioned from a combination of crown and flint glasses, with much greater flexibility today due to development of glasses with much more varied combinations of refractive index and dispersion.

Notation for Optical Glasses

Since the refractive index and the dispersion are the two most important characteristics of an optical glass for use as a lens, a notation system has been developed to designate any glass by a six digit number. This notation system requires that the dispersion be expressed by the Abbe number, ν_d, which is defined as

$$\nu_d = (n_d - 1)/(n_F - n_c) \qquad [1]$$

where d, F, and c indicate the value of the refractive index measured at the yellow helium line at 587.6 nm, the blue hydrogen line at 486.1 nm, and the red hydrogen line at 656.3 nm, respectively. Since the Abbe number is a reciprocal dispersion, a large Abbe number indicates a low dispersion, while a small Abbe number indicates a larger dispersion. A glass is then designated by the expression

$$1000(n_d - 1)(10\nu_d) \qquad [2]$$

where each term is rounded to 3 significant digits. A common crown glass known as BK7, for example, which has a n_d of 1.516 800 and ν_d of 64.17 is designated as 517 642. A very high index flint glass, with $n_d = 1.807\,410$ and $\nu_d = 31.61$ would be designated as 807 316. This system can be used for all but the exceptional glasses with $n_d \geq 2.00$. Some tabulations are based on the green mercury line, or e line, at 546.1 nm instead of the yellow helium (d) line. The choice of the d or e line results in a small difference in the reported refractive index and Abbe number for any given glass (Figure 1).

Dispersion Formulae for Optical Glasses

The combination of a refractive index near the middle of the visible region of the spectrum and the Abbe number describe the contribution of an optical glass to the refractive power and chromatic aberration of a lens. This information was sufficient for many early applications of lenses and is still sufficient for the design of optical devices of modest expectations. More sophisticated optical devices require more information for proper design. The simple Abbe number is often insufficient for expressing the dispersion, since it is necessary to provide a

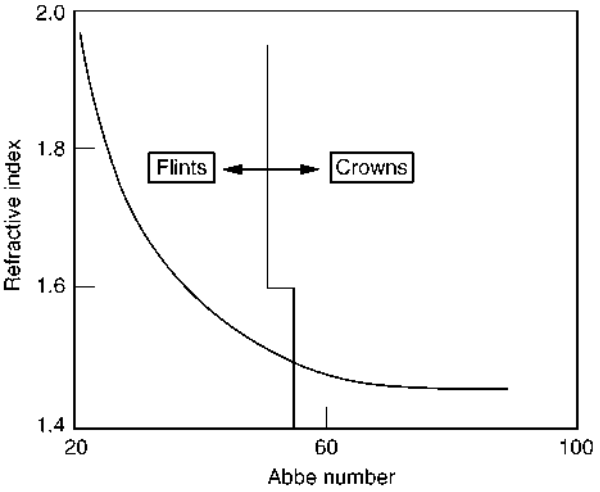

Figure 1 Relation between the refractive index and Abbe number of optical glasses. Commerical optical glasses lie in the regions labeled flint and crown above the curve.

detailed set of values of the refractive index as a function of wavelength across the entire visible region. The refractive index can be listed at a number of wavelengths, or described by an equation covering the entire wavelength range of interest. Although there are a number of expressions for the wavelength dependence of the refractive index, the most commonly used are the Cauchy dispersion formula, or

$$n = a + b\lambda^{-2} + c\lambda^{-4} \qquad [3]$$

where a, b, and c are constants derived from a best fit of the data to this expression, and variations of the Sellmeier dispersion formula, which is often simplified to the expression

$$n^2 = A_0 + A_1\lambda^2 + A_2\lambda^{-2} + A_3\lambda^{-4} + A_4\lambda^{-6} + A_5\lambda^{-8} \qquad [4]$$

where the six A_n constants are tabulated. The Cauchy dispersion formula is accurate to $\approx 10^{-4}$, while the modified Sellmeier dispersion formula is accurate to $\approx 3 \times 10^{-6}$. The validity of the latter expression can be extended further into the ultraviolet by adding a λ^{-10} term and into the infrared by adding a λ^4 term.

Annealing of Optical Glasses

Much of our knowledge of the behavior of glasses in the glass transformation region is based on studies of the annealing of optical glasses. Glass-forming melts have much higher thermal expansion coefficients than those of solid glasses of the same composition. During cooling of a melt, the density increases rapidly so long as the melt is relatively fluid. As the viscosity increases, the rate of structural rearrangement of the melt slows and eventually becomes fixed at some effective, or fictive, temperature. If the melt is cooled rapidly, the structure of the glass will be representative of a high temperature melt. If the melt is cooled more slowly, the structure will have time to equilibrate at lower temperatures and the glass formed will have the structure of a lower temperature melt and hence will be more dense. Since the refractive index of a solid is related to its density, the refractive index will be less for the glass produced by rapid cooling than for one produced by slower cooling.

There are two ramifications of the cooling effect on refractive index. First, the refractive index will be different for different cooling rates, even though the composition is identical. Statements regarding the refractive index of a glass are only applicable for a known thermal history. Second, but less obvious, the outside of a piece of glass will cool more rapidly than the inside due to the kinetics involved with removing heat from the glass. As a result, a refractive index gradient will exist in the glass, with a lower refractive index at the surface. If the glass is ground and polished to form a lens, the lens will not perform as expected if the shape is based on the assumption that the refractive index is constant through the entire lens. While cooling-induced differences in refractive index between the surface and center of a large block of glass are usually of the order of 0.000 1, they can be ≥ 0.001 for a rapidly cooled block.

Control of the cooling rate is essential for the production of a block of glass of uniform refractive index. The cooling rates used to reduce stresses to acceptable levels in the production of nonoptical glasses are much too fast for the production of optical glasses. Use of a constant cooling rate throughout the glass transformation region will not yield a block of uniform refractive index. Ideally, the cooling rate should be decreased gradually as the temperature decreases throughout the glass transformation region. The production of a fine annealed glass requires a very slow cooling process, with a gradually decreasing cooling rate with decreasing temperature within the temperature region where permanent changes in the structure of the glass can occur.

Athermal Lenses

While the thermal expansion coefficient of many glasses is relatively low, it is not zero. As a result, the dimensions and refractive index of a glass are functions of temperature. Temperature changes in an optical system therefore result in changes in optical path length. While changes in optical path length of this magnitude may be unimportant for optical systems used within a narrow temperature range, those used over ranges exceeding 10 to 20 K must be designed to minimize thermal effects. Production of athermal optical systems, i.e., systems of constant power over a range of temperatures, requires use of glasses where the thermal changes in refractive index are exactly offset by the changes in dimensions, i.e., the optical path length is independent of temperature. This condition is met if the temperature coefficients of the index and the thickness are equal and of opposite sign.

The overall effect of changes in temperature on the optical path length are tabulated in terms of the thermo-optical constant, G, given by

$$G = \alpha(n - 1) + \mathrm{d}n/\mathrm{d}T \qquad [5]$$

where α is the linear thermal expansion coefficient, n is the refractive index, and $\mathrm{d}n/\mathrm{d}T$ is the thermal

coefficient of the refractive index. Perfect compensation between changes in index and dimensions are obtained when $G = 0$. Since α is positive for most optical glasses, with values in the range of 3 to 15 ppm/K, this condition requires a negative dn/dT. Negative values of dn/dT are found primarily for glasses containing some fluorine, such as fluoride glasses and those in which oxygen is partially replaced by fluorine, such as fluorosilicates, fluoroborates, etc., or for some phosphate glasses. Silicate glasses rarely, if ever, have a negative value of dn/dT. For the same glassformer, e.g., P_2O_5, use of modifier ions with large diameters and low field strengths favor increasingly negative values of dn/dT.

The value and sign of dn/dT is determined by the competition between the decrease in refractive index induced by an increase in molar volume and the increase in the refractive index induced by an increase in polarizability of the ions in the glass with an increase in temperature. Equation [5] indicates that the absolute value of the refractive index is also important in determining the value of G. If the refractive index is high, then dn/dT must be more negative to yield $G = 0$. As a result, most athermal glasses are crown glasses. The value of dn/dT is also wavelength-dependent, generally decreasing with increasing wavelength. In particular, dn/dT becomes more positive at wavelengths near the UV edge of the glass. An athermal lens is only perfectly athermal at the design wavelength.

A few special glasses and glass-ceramics used for optical applications, e.g., vitreous silica, titanium-doped silica, and some aluminosilicate glass-ceramics, have thermal expansion coefficients which approach zero or are slightly negative. Vitreous silica has a very low refractive index and thermal expansion coefficient and can be used to produce lenses which, while not truly athermal, have relatively low temperature sensitivity. These low thermal expansion materials are often used as substrates for coated reflective optics (mirrors), where dimensional stability is far more important than any change in the refractive index of the substrate.

Production of Traditional Optical Glasses

Quality requirements for optical glasses are much greater than those for other common glasses. A constant refractive index throughout the glass requires an extremely high degree of homogeneity. The glass must be free from defects such as striae (regions of refractive index variation due to density/composition variations) and stones (undissolved material) and the bubble content must be as low as possible. The thermal history of the glass must be such that the refractive index variation, due to differing fictive temperatures, is minimized. Internal stresses must be minimized to yield a low birefringence. The glass must also be free of any species which contribute to absorption in the desired spectral region.

The quality demands of optical glasses require use of much higher quality, and thus more expensive, raw materials. Concentrations of third-row transmission metal ions must be held to rigid specifications, with maximum values usually in the ppm, or even in the ppb range, for the highest quality glasses. Defects are reduced by melting in platinum containers instead of the oxide containers used in the past. Stirring the melt, using platinum wherever possible, is essential for obtaining acceptable homogeneity. Control of the melting atmosphere may be essential for controlling the redox state of the melt. Rigorous annealing schedules must be used to produce glasses with the desired thermal history so that the problems of stress-induced birefringence and refractive index gradients are reduced to acceptable levels.

Non-Traditional Optical Glasses

While most traditional applications of optical glasses involve transmission in the visible, many modern applications require high transmission in the ultraviolet, near-infrared, or infrared regions of the spectrum. Transmission in each of these regions is determined by both the bulk composition of the glass and by the concentration of impurities present. Most nontraditional applications place even more stringent requirements on the purity of the materials used in manufacturing the glass. An extreme example of the demand for purity is found in the glasses used for telecommunication optical fibers, where concentrations of some impurities must be held to the low ppb level.

Ultraviolet-Transmitting Glasses

The ultraviolet transmission of a glass is determined by the intrinsic electronic transitions of the material, by charge transfer absorption bands due to impurities, and by surface and bulk scattering. The first of these is determined by the electronic bandgap of the material, while the other two are controlled by the care taken during production of the glass. These glasses are used in spectroscopy for windows, lenses, UV-lamps, and sample cuvets, for microlithography, and for lenses in excimer laser optical systems.

Vitreous silica is widely used as an ultraviolet transmitting glass for transmission optics at

wavelengths well below 200 nm. The lowest energy absorption band for this glass occurs at 10.2 eV, or 122 nm. Many UV–VIS spectrometers, which normally are capable of measurements extending to 185 to 190 nm, use vitreous silica as optical elements. Additions of alkali or alkaline earth oxides to silica produce nonbridging oxygens in the network. Nonbridging oxygens do not bind the excitable electrons as tightly as bridging oxygens, which shifts the absorption edge to lower energies, i.e., longer wavelenths. These glasses are incapable of high transmission beyond 5 eV, or about 250 nm, and are not used for applications requiring good ultraviolet transmission.

Addition of elements which eliminate nonbridging oxygens favors good ultraviolet transmission. Alumina and/or boric oxide are frequently used in conjunction with alkali or alkaline earth oxides to counter the formation of nonbridging oxygens and preserve the ultraviolet transparency of silicate glasses. This approach can be taken to its logical conclusion by eliminating silica from the composition and producing alkali/alkaline-earth-aluminoborate glasses with good ultraviolet transparency.

In theory, fluoride glasses should provide better ultraviolet transmission than oxide glasses due to the larger bandgaps of fluorides. This assumption is true for single crystals, where fluorides such as CaF_2 are under development for transmission to higher energies than those obtainable with oxide crystals or glasses. In practice, the superiority of fluoride glasses for ultraviolet transmitting glasses has proven difficult to achieve. Vitreous BeF_2 and related glasses have superior ultraviolet transmission when compared to silicate glasses. Unfortunately, BeF_2 is very volatile and quite toxic, which restricts production of glasses based on BeF_2. Some fluoroaluminate and fluorophosphate glasses also have excellent ultraviolet transparency, but difficulties in producing these glasses in large sizes and of optical quality has limited their application.

Impurities are particularly important in production of ultraviolet transmitting glasses. The third-row transition metals, platinum, and cerium are all known to absorb strongly in the region from 200 to 300 nm when dissolved in glasses. Ferric iron has such a strong absorption band at ≈ 230 nm that concentrations of ferric ions must be kept in the low ppm range in glasses used for ultraviolet transmission. Since iron is a very common impurity in the sands used to produce silicate glasses, special care must be taken to minimize the Fe^{3+} content of ultraviolet-transmitting silicate glasses. Since ferrous iron has lower absorption in the ultraviolet, the effect of iron can be reduced by melting under reducing conditions. Platinum can enter melts from the melting unit and can be present as either ions or metallic particles. The ionic form of platinum absorbs ultraviolet light, while the metallic form scatters the same light.

Near-Infrared-Transmitting Glasses

Most glasses have good intrinsic optical transmission in the near-infrared. Absorption in this region is primarily due to impurities. Ferrous iron is especially detrimental to transmission in this region as a result of an extremely broad ligand field absorption band centered at about 1050 nm and extending into the visible and out to over 2000 nm. Reduction of absorption due to ferrous ions is obtained by careful control of raw materials to minimize the amount of iron oxide present and by melting under oxidizing conditions, which converts the ferrous ions to ferric ions. Conversion from ferrous to ferric ions increases the ultraviolet absorption of the glass. If good transmission in both the ultraviolet and near infrared regions is required, the iron content must be held to very low levels.

Although many rare earth ions also have absorption bands in this region, these elements are rarely found as impurities in sufficient levels to impair transmission in the near-infrared region for lenses and other common optical components. A more serious problem occurs for glasses used as optical fibers, where the greater optical path length magnifies the effect of impurities which may be undetectable in the spectra of samples only a few mm thick. The same can be said of absorptions of other ions which have very weak absorption bands in this region. Even higher order overtones/combination bands of infrared vibration absorption bands due to hydroxyl become important in optical fibers. While the hydroxyl absorption in infrared-transmitting glasses used in mm thicknesses is insignificant for concentrations less than 1 ppm, the hydroxyl concentration of optical fibers must be reduced to levels of a few ppb to reduce the effect of the overtone/combination bands to acceptable levels.

Infrared-Transmitting Glasses

Infrared transparency is limited by the intrinsic vibration, or multiphonon, edge of the glass and by the presence of impurities, primarily hydroxyl and carbon dioxide, which absorb in the infrared. The position of the infrared edge is determined by the force constant, which is proportional to the bond strength, and the masses of the ions. The best infrared transmission is found for materials with weak bonds and heavy ions. Infrared transmitting glasses are used in infrared spectrometers, lenses for pyrometers,

infrared transparent domes for aircraft and missiles, and optical components for many systems using infrared lasers.

The fundamental absorptions of B_2O_3, SiO_2, P_2O_5, and GeO_2 lie in the region from 7.5 to 12 μm. These bands have such strong absorptions that the first overtone, at half the wavelength of the fundamental vibration, effectively determines the infrared edge for samples of more than a few hundred micrometers in thickness. The best infrared-transmitting oxide glasses do not contain significant amounts of B_2O_3, SiO_2, or P_2O_5. The lower fundamental vibrational frequency of GeO_2 allows production of some glasses of limited use for infrared transmission to about 5 μm. Comparable transmission is obtained for calcium aluminate glasses. Better transmission has been obtained for glasses combining the oxides of lead, bismuth, or cadmium with gallium oxide. These gallate glasses are termed heavy metal oxide, or HMO, glasses and transmit to ≈7 μm in mm thicknesses. The HMO glasses transmit into the visible, with a cutoff at 400 to 500 nm. Since these glasses are transparent to visible light, they offer advantages over many other infrared-transmitting glasses which are opaque in the visible.

Recently, heavy metal halide (HMH) glasses have been found to exist in a wide range of compositional systems. HMH glasses transmit much further into the infrared than the best oxide glasses, with cutoff wavelengths extending beyond 20 μm. These glasses are much more difficult to prepare than oxide glasses and are highly prone to crystallization, less durable in water, and mechanically weaker than the HMO glasses. Improvements in infrared transmission correlate with degradation in the other properties needed for most applications. Oxide impurities severely degrade the infrared transmission of these glasses. While much greater infrared cutoffs can be obtained in the laboratory, glasses currently produced commercially transmit to ≈8 μm.

Chalcogenide glasses contain either S, Se, or Te as the glassforming anion and are free of oxygen. Commercial compositions transmit to 16 μm, while laboratory samples transmitting beyond 20 μm can be made. These glasses have refractive indices in the infrared in the range of 2 to 3, very low glass transformation temperatures, which limit their temperature range of application, and are, with few exceptions, opaque in the visible. Toxicity is also a problem for Se and Te. Oxygen impurities must be kept below the 1 ppm level to avoid unacceptable absorption bands.

All infrared-transmitting glasses suffer from transmission interference in the region from 2.5 to 6 μm, due to hydroxyl in the glass. Since hydroxyl readily forms in any melt exposed to an atmosphere containing water vapor, great care must be used to reduce the hydroxyl content to acceptable levels. Glasses can be melted in sealed containers, under reactive atmospheres containing fluorine or chlorine, or under vacuum. Some glasses are prepared under dry conditions and then remelted under a vacuum to reduce the hydroxyl content. Similar care must be used to exclude CO_2 from HMH glasses, where a band at 4.25 μm has been assigned to the asymmetric stretching vibration of the dissolved molecules (Figure 2).

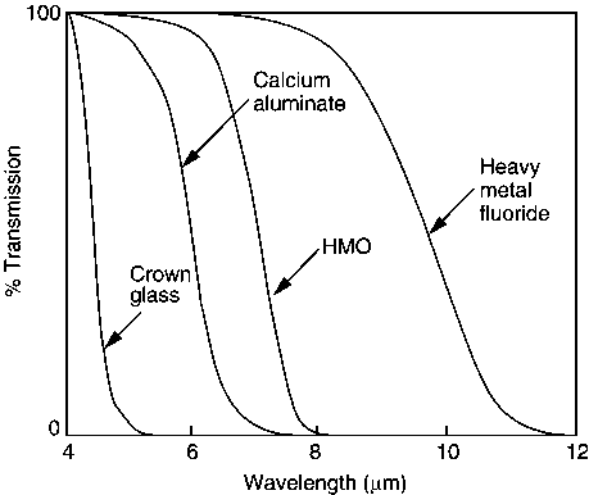

Figure 2 Infrared transmission spectra of several optical glasses showing the effect of glass composition on the infrared cutoff frequency.

Optical Isolators

Optical isolators based on the Faraday effect contain high concentrations of rare earths for paramagnetic devices and lead for diamagnetic devices. The highest concentrations of rare earth ions are obtained in aluminosilicate, aluminoborate, and aluminogermanate systems, where excellent quality glasses can be produced, while the best diamagnetic isolators are obtained using the lead bismuth gallate glasses.

See also

Optical Materials: Color Filter and Absorption Glasses. **Magneto-Optics:** Faraday Rotation, CARS, ODMR, ODSR, Optical Pumping.

Further Reading

Deeg EW (1986) Optical glasses. In: Boyd DC and MacDowell JF (eds) *Commericial Glasses*, pp. 1–34. Columbus, OH: The American Ceramic Society, Inc.

Fanderlik I (1983) *Optical Properties of Glass*. Amsterdam: Elsevier Scientific Publishing Company.
Hoffmann HJ (1998) Differential changes in the refractive index. In: Bach H and Neuroth N (eds) *The Properties of Optical Glass*, pp. 96–123. Heidelberg, Germany: Springer-Verlag.
Lentes FT (1998) Refractive index and dispersion. In: Bach H and Neuroth N (eds) *The Properties of Optical Glass*, pp. 19–58. Heidelberg, Germany: Springer-Verlag.
Liepmann MJ and Neuroth N (1998) Ultraviolet-transmitting glasses. In: Bach H and Neuroth N (eds) *The Properties of Optical Glass*, pp. 288–297. Heidelberg, Germany: Springer-Verlag.
Liepmann MJ and Neuroth N (1998) Infrared-transmitting glasses. In: Bach H and Neuroth N (eds) *The Properties of Optical Glass*, pp. 297–308. Heidelberg, Germany: Springer-Verlag.
Pfander HG (1996) *Schott Guide to Glass*. London: Chapman & Hall.
Rawson H (1980) *Properties and Applications of Glass*. Amsterdam: Elsevier Scientific Publishing Company.
Shelby J (1997) *Introduction to Glass Science and Technology*. Cambridge, UK: Royal Society of Chemistry.
Vogel W (1991) Optical properties of oxide glasses. In: Ullmann DR and Kreidl NJ (eds) *Optical Properties of Glass*, pp. 1–36. Westerville, OH: The American Ceramic Society, Inc.

Plastics

T Bauer, Wahl Optoparts GmbH, Triptis, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Injection molded precision plastic optics in high volumes were first produced during the 1960s. After the development of sophisticated measuring and manufacturing methods in the late 1960s, precise aspheric surfaces were as easy to make as spheric contours in plastics. Today plastic is a widely used low-cost option compared to glass with even more degrees of freedom for optical and component design. Polymer optical systems are used in sensor applications, visual systems, cameras, mobile phones, video-conferencing cameras, scanners, security systems, and so on.

The physical and chemical properties of plastic materials are very different from those of glass. Generally speaking, glass materials are harder, more durable, and more stable against temperature and humidity than plastics. The variety of optical glasses comprises hundreds of different materials. Compared with this the choice for plastic materials is limited only to about 10 different materials (and even less optical parameter variations). However, plastic optics offers other design freedoms that are not achievable with glass optics.

The manufacturing technologies for glass and plastic optics are totally different. Glass lenses are made by a grinding and polishing process whereas precision plastic lenses are usually made by injection molding, compression molding, or casting. Because of the material's characteristics and the manufacturing process, plastic optics have some unique advantages.

High production numbers at low costs

Injection molding is ideal for high-volume production with low-unit costs. Moderate raw material costs and multicavity mold designs (up to 32 cavities) allow large production volumes at a reasonable unit price. In spite of considerable tooling costs, these costs compared to glass design versions, can be relatively low.

Lightweight and hardiness

For a given volume glass is much heavier than plastic (by a factor of 2.3 to 4.9). However, plastic materials are relatively shatter and impact resistant. These features are important for head-mounted systems and other weight-sensitive applications (mobile phones, cameras, etc.).

Design potentials

Injection molding makes it economical to produce sophisticated optical shapes such as aspheres, diffractive optical elements, or even freeform surface structures. From the design point of view, the more sophisticated surface shapes reduce costs or obtain better performance (Fresnel structures, lens arrays, diffractive optical elements, etc.).

Optical systems and component assembly

For typical optical system designs optical components (mirrors, lenses, prisms, etc.) must be fixed in a mounting. With plastic optics it is possible to mold mounting elements, posts, or alignment notches integrally with the optical component. This can reduce part and assembly costs considerably.

Technologies adapted to plastic materials, such as ultrasonic and laser welding, gluing and integrated

snap-in structures, allow fast and cost-efficient automated and manually operated assembling solutions.

History of Transparent Plastic Materials

A Material With No Use

In 1840, the Austrian chemist Redtenbacher (1810–1870), a student of Justus Liebig, discovered during a chemical experiment a malodorous transparent liquidity, the acrylic acid. It was an unwanted byproduct which was not further analyzed at that time. Nevertheless, that discovery can be considered to be the origin of today's plastic optics.

The development of plastics began in the US in 1860 as people were searching for an alternative for ivory in the production of billiard balls.

In 1869 John Wesley Hyatt developed from cellulose nitrate and camphor the first plastic material: celluloid. At first it was used for dental plates and afterwards for shirt collars and as a film base material. Despite its flammability and its lack of light resistance, it was a first economical success.

In the following decades the first synthetical plastics were developed, for example, the material with the trade name Bakelite in 1906.

Otto Röhm's Doctorate

The basis of acrylates was rediscovered in the records of Redtenbacher in 1900. From 1901 onwards, Prof. Pechmann (1850–1902) asked his postgraduates to do more research into the acrylic acid. The acrylic acid is an acetylene compound which can be extracted from carbon, petroleum, or lime.

In 1901 Otto Röhm, a student of Pechmann, wrote his dissertation 'About polymerisation products of the acrylic acid' – the basis of the methacrylate chemistry.

Hermann Staudinger (1881–1965), head of the Institute of Chemistry in Freiburg started theoretical investigations into the structure and natural characteristics of natural and synthetical polymers in 1920. He assumed that the molecules of plastic materials consist of numerous small molecule units. Experiments as proof of the theory led to a rapid development of scientific research and to significant breakthroughs in plastic chemistry. Staudinger won the Nobel Prize for Chemistry for his work in the field of macromolecules in 1953.

In 1928 methyl methacrylate was syntheticized for the first time and patented. Acrylic resin was used as a binder in multi-layer glass.

In the 1920s and 1930s a number of plastic materials were developed, including cellulose acetate CA (photo film, synthetic fiber); polyvinyl chloride PVC (pipes, coatings, isolations); urea formaldehyde resin (dishes, electrical equipment).

Polystyrene resins were commercially produced for the first time in 1937. Polytetrafluoroethene (PTFE) was produced for the first time in 1938, and distributed as 'Teflon' since 1943. Another key development was the synthesis of Nylon, the first technical high-performance plastic material.

Plexiglas Captures the World

The industrial production of methacrylic acid methyl ester began in 1934. With its polymerization to a hard and transparent plastic material, polymethylmethacrylic (PMMA or acrylic) was developed. It was patented in 1936 and caused, under the registered trade name Plexiglas®, a sensation at the world exhibition in Paris in 1936, almost 100 years after the discovery of acrylic acid.

Plates of Plexiglas were manufactured by means of effusion and polymerization between glass plates. At first it was used for watch glass, cockpits, and the construction of street lighting and luminous advertising. People already experimented with Plexiglas injection moulding in 1935. Arthur Kingston, the founder of Combined Optical Industries Limited (COIL) had invented and patented the first plastic lens in 1933.

The First Plastic Lenses in Cameras

The first camera, incorporating plastic optical parts, to be mass produced was the 'Purma Special' introduced by RF Hunter Ltd., in 1937, using 'Perspex' (another brand name for Acrylic), for the direct-vision viewfinder.

Due to the scarcity of natural raw materials at the beginning of World War II, the plastic industry became a source of outstanding substitute materials. This development had a lasting effect far into the post-war period. Results were mainly achieved with technical plastics, such as polycarbonates, acetates, and polyamides. Other plastics were developed as a substitute for metals as well as for hardhats, for high-temperature constructions, and for products which are resistant against acids and bases.

The German chemist Karl Ziegler discovered polyethylene (Polyethen, PE) in 1953 and the Italian chemist Giulio Natta Polypropylene (PP) in 1954.

The Eastman Kodak company is credited with giving birth to the modern-day plastic Fresnel lens. In 1946 Kodak developed a tooling and manufacturing process to mass produce plastic Fresnel lenses.

The circular Fresnel lenses made at that time had spiral grooves like music records.

It took another 10 years of improvement in injection molding techniques to the first plastic lens camera – the Kodak Brownie 44A from 1959.

Between 1953 and 1959 the Polycarbonate polymer was developed by Bayer and General Electrics. Currently acrylics and polycarbonates are still the most important materials for plastic optics.

But not only thermoplastic materials were developed over this time. In May 1940 research work resulted in the monomer allyl diglycol carbonate (ADC). In the following years more than 180 different compounds of this monomer were investigated and examined.

During World War II Columbia Southern Chemical Company (a subsidiary of Pittsburgh Plate Glass) began research into ways of obtaining nonthermoplastic materials. The project was given the name 'Columbia Resins'. The 39th tested (CR 39) turned out to be the most significant due to its unique properties and was used since 1947.

At first, solely the low-refractive CR 39 was available for organic eyeglasses. Only in the mid-1980s did the development begin in the field of high-refractive nonthermoplastic materials.

Properties of Plastics for Optics

As already mentioned, the choice for plastic optical material is limited only to about 10 different types of material. Optical properties (abbe value, refractive index, transparency) as well as mechanical, thermal, and humidity boundary conditions are decisive for the material choice. Different to glass during plastic processing (injection molding, casting, etc.) the process affects not only the geometry but also inner properties like refractive index, transparency, and birefringence.

Optical Properties

Fundamental optical properties are defined by optical transmission, refractive index, and dispersion. For plastic optics birefringence is an important parameter too.

Although the total number of plastic materials has increased in recent years, the range of refractive index and dispersion characteristic is limited to almost two major groups – crown-like materials such as Acrylic (PMMA), Polyolefin (COC, COP) and flint-like materials such as Polystyrene (PS), Polycarbonate (PC) and Styrol-Acrylnitril-Copolymer (SAN).

This strongly limited variety of plastic optical materials significantly restricts the optical design freedom (Figures 1 and 2).

Physical Properties

Important physical properties are weight, impact, and abrasion resistance – and thermal properties like temperature resistance and thermal expansion. Already, during system design, both mechanical and thermal properties have to be taken into consideration. The thermal expansion of optical plastics is approximately ten times higher than that of glass materials. In an optical system this effect has to be compensated for by optical design or mounting.

Typically most optical plastics can withstand temperatures up to 90°C. The maximum service temperature of Polycarbonate and Poly-Olefin materials reaches 120°C.

The specific gravity of plastic optical materials ranges from 1 to 1.3.

Polycarbonate has the highest impact resistance of all optical plastics and so is used for windshields and crash helmets. Acrylic has the best abrasion resistance (Table 1).

Common Materials

Thermoplastics

Thermoplastics are polymers with a linear molecule structure – with or without side chains – which can be deformed reversibly without any modification of their thermoplastic characterization by means of the impact of heat. Optical elements can be manufactured by injection molding or hot embossing.

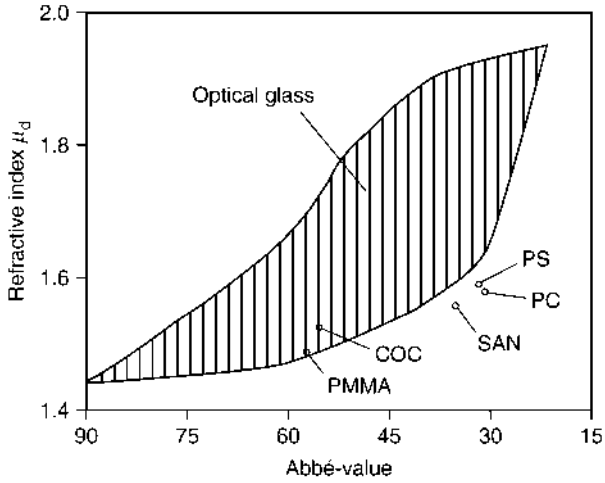

Figure 1 Abbe diagram.

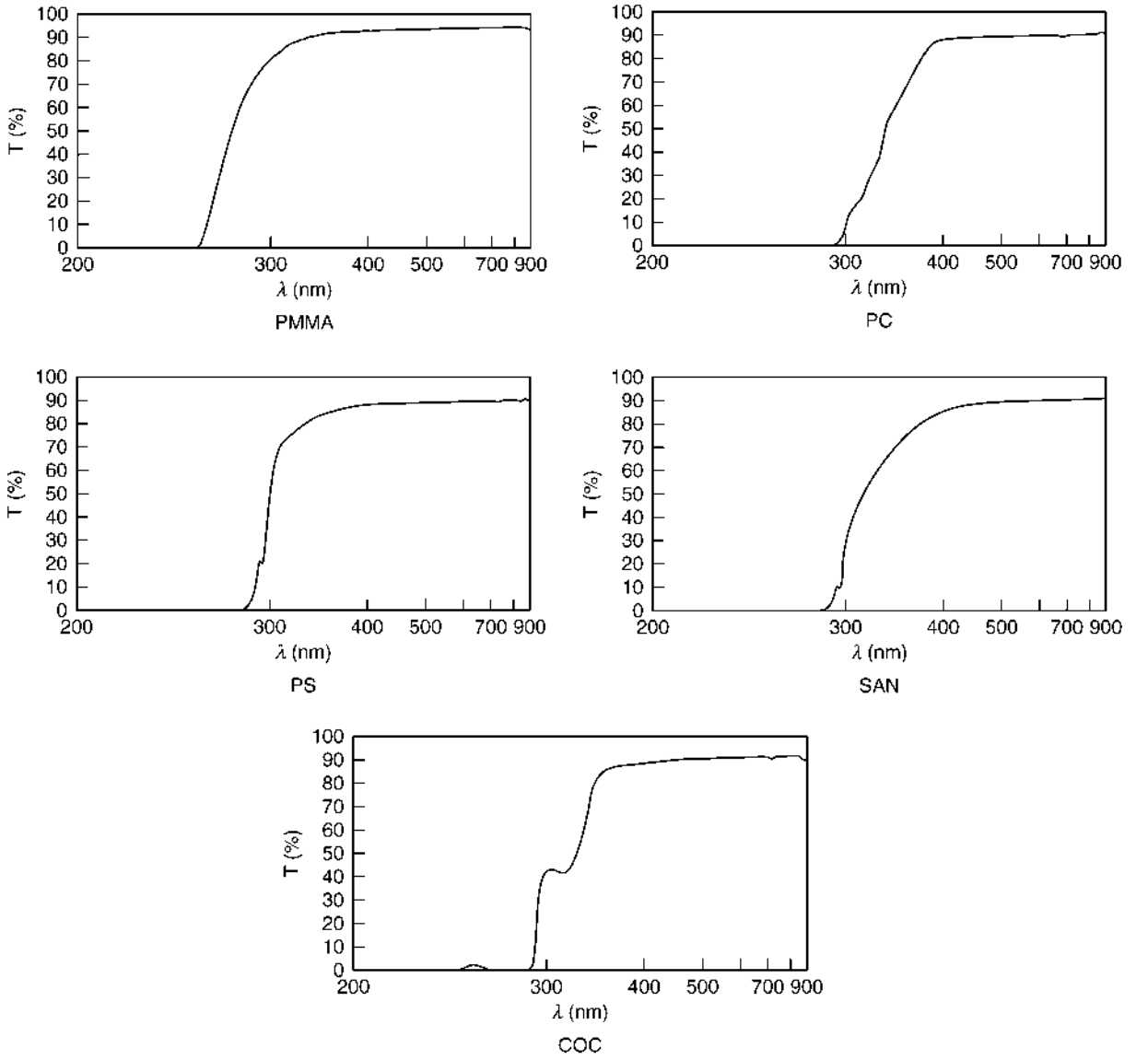

Figure 2 Transmission characteristics.

Typical optical thermoplastic polymers are listed below.

Acrylic or PMMA (Polymethyl methacrylate)

Acrylic is the most commonly used optical plastic. It is moderately priced, easy to mold, scratch resistant, and not very water absorptive. Its transparency is greater than that of most optical glasses throughout the visible range of the spectrum (organic glass). PMMA can replace glass in every application as long as the service temperature is less than 90°C and a low chemical resistance is required. Additives to acrylic (as well as to several other plastics) considerably improve its ultraviolet transmittance and stability.

PC (Polycarbonate)

Polycarbonates are linear polyesters of the carbonic acid which combine many characteristics of metal, glass, and plastics.

Polycarbonate is very similar to styrene in terms of optical properties such as transmission, refractive index, and dispersion. Polycarbonate, however, has a much broader operating temperature band of up to more than 120°C. It is used as the flint material for systems that have to withstand severe thermal conditions.

Another advantage is the high-impact resistance of polycarbonate. Safety glasses and systems requiring durability often consist of polycarbonate. Because of its high ductility polycarbonate is not easily machined.

Table 1 Material properties

Material	*Characteristics*	*Acrylic (PMMA)*	*Polystyrene (PS)*	*Polycarbonate (PC)*	*Styrol-Acrylnitril (SAN)*	*Cyclic Olefin Polymer (COP)*	*Cyclic Olefin Copolymer (COC)*	*Polyether-sulfone (PES)*	*Acrylnitril-Butadien-Styrolco-polymere (ABS)*	*CR39*	*Optical glass (B7)*
Optical	Spectral Passing Band (nm)	390–1600	400–1600	360–1600	395–1600	300–1600					
	Refractive index at 587 nm and 20 °C	1.4918	1.5905	1.5855	1.5674	1.5261		1.6600	1.5380	1.501	1.517
	Abbe value $(nD-1)/(nF-nC)$	57.2 (53)	30.8	34 (29.9)	34.8	55.8	58	19.4		58	64.4
	Transmittance (%) thickness 3.2 mm	92	87–92	85–91	88	92	92	80	85		
	Haze% thickness 3.2 mm	1.3	2–3	1.7	1.5	1.5	1.5	1.5			
Physical	Specific Gravity (g/cm)	1.18	1.06	1.25	1.07	1.01	1.03	1.37	1.05	1.32	2.53
	Max. Service Temperature (°C)	92	82	124	95	123	130	200	90	130	400
	Linear Expansion Coefficient $(1/K)$	6.8×10^{-5}	8.0×10^{-5}	6.6×10^{-5}	7.0×10^{-5}	7.0×10^{-5}	7.0×10^{-5}	5.5×10^{-5}	8.5×10^{-5}		1.1×10^{-4}
	Abrasion Resistance (1–10)	10	4	2		6	6				
Environmental	dn/dT $(\times10^{-5})$	−8.5	−12	−11.8…−14.3		−8	−10.1				
	Sensitivity to Humidity	high	low	low	medium	low		high	medium		
	Water Absorbtion (weight %) 23 °C, ISO 61	0.30	0.20	0.15	0.30	<0.01	<0.01	0.70	0.45		0
Manufacturability	Processability	excellent	good	poor	excellent	good					
	Birefrigence	low	high	high		low					
Chemical	Resitance to Alcohol	limited	good	limited				good			
Costs	approx. Material costs (EUR/kg)	3.3	2.5	4.4	4.4	27.1		21.0	3.5		25.0

PS (Polystyrene)
Polystyrene or Styrene is a low-cost material with excellent molding properties. Styrene has a higher index and a lower numerical dispersion value than other plastics. It is often used as the flint element in color-corrected plastic optical systems.

Compared with acrylic, styrene has lower transmission in the UV range of the spectrum. It has a lower resistance to UV than acrylic but is more easily scratched than acrylic.

Because its surface is less durable, styrene is more typically used in nonexposed areas of a lens system. A styrene lens, when paired with an appropriate acrylic lens, offers an effective achromatic solution.

SAN (Styrol-Acrylnitril-Copolymer)
SAN is a copolymer of certain fractions of Polystyrene and Acrylic (typically 70/30), which allows adjusting refractive indexes. It is a glassy polymer with a low tendency to stress cracks. SAN has slightly more chemical durability than PS and is a low-cost material with excellent molding properties.

ABS/MABS (Acryl-Butadien-Styrol-Copolymer)
ABS is developed from polymerization of styrene and acrylnitrit on polybutadien during the emulsion process. Process variations lead to different material characteristics. Compared to PS, the chemical resistance and temperature stability of ABS is considerably better. ABS has a natural color so it cannot be used for every optical application.

PES (Polyethersulfon)
PES shows a remarkable temperature stability, a low transmission against UV-light, a limited transmission with a natural yellow color, and cannot be used for every optical application. Because of its excellent temperature stability, PES is suitable for assembling, even on electronic motherboards by reflow soldering. This process causes temperature stress up to 220°C (Figure 3).

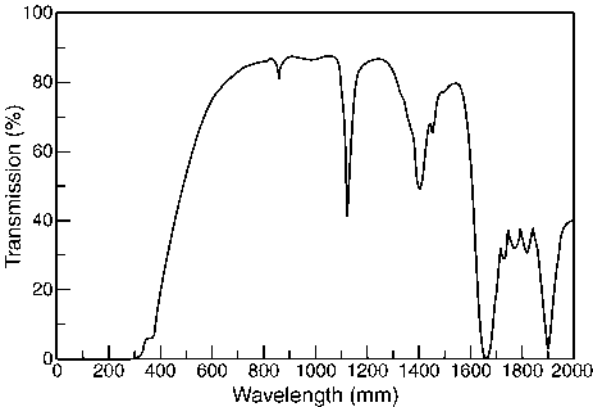

Figure 3 PES.

COP and COC (Cyclic olefin polymer and copolymer)
Cyclic olefin (co-)polymer provides a high-temperature alternative to acrylic. Its refractive index and transmittance are similar but the heat distortion temperature is about 30°C higher than for acrylic. It has a remarkably low water absorption value of less than 0.01% (compared with Polycarbonate 0.2% and PMMA 0.3%).

Cyclic Olefin Polymer (COP) is a family of amorphous plastic resins with low native stress birefringence properties. Zeonex is a resin developed by the Nippon Zeon Company – one grade, E48R, has a heat distortion temperature of about 122°C. Zeon Chemicals has another class of polymers called Zeonor, that can be provided as a lower-cost alternative to Zeonex.

Cyclic Olefin Copolymer (COC), with the trade name of Topas, is another high-temperature alternative to acrylic. Generally, it is a lower-cost alternative to COP. Particularly favorable is the high flowability (good injection moldability) of these thermoplastics.

PMP (Polymethylpenten – TPX®)
PMP has a structure similar to PE, except that only the methyl groups are replaced with isobutyl groups. The chemical resistance can be compared to PP, although it tends to stress crack under the impact of ketones or chlorinated solvents. The main advantages of PMP are its excellent transparency and its good mechanical characteristics, even at high service temperatures of up to 150°C.

PE-HD (high density polyethylene)
If polymerization is controlled catalytically, molecules with a low number of side chains will be received. Compared to PE-LD (low density), the molecules are quite compact with a higher stability and chemical resistance and service temperature can be up to 105°C. PE shows a bad transparency (milky appearance) in the visible spectral range. However, this material is permeable for infrared beams and so is frequently used for motion detectors.

Elastomere

Elastomeres are plastics with loose and networked molecules which are rubber elastic at a normal temperature. The most popular elastomeres are natural rubber and silicone rubber. These materials are optically used as flexible light guides.

Duromeres

Organic glass is a fully synthetic plastic material available in a vitreous state. It consists of macromolecular organic compounds which do not follow any principle of periodic arrangement and are hence amorphous.

Duromeres are plastics with spacial closely networked molecules which are very hard and refractory at a normal temperature. Impact of heat causes an irreversible induration. In most cases, duromers are used to produce plastic lenses made of organic glass. Once they have been thermally treated after production, their shape can no longer be changed.

The well-known plastic CR 39 is one of the organic materials used for plastic lenses.

CR39 plastic is made by polymerization of the allyl diglycol carbonate (ADC). It is a transparent, thermosetting resin which combines an exceptional range of qualities which are not available in other plastic transparent materials. The CR39 plastic is colorless and completely transparent to the visible light and almost completely opaque in the infrared and ultraviolet region of the spectrum. For these reasons, it is largely used for the production of sunglass lenses. They can be colored by surface dyeing or bulk tinting, have high abrasion resistance, and high-quality optical properties. Their weight is about half as much as glass, they keep their excellent optical properties despite long-term exposure to chemicals and resist heat distortion up to 100°C. These qualities, combined with the exceptional optical characteristics, make CR39 the best choice for applications where severe conditions of use exclude all other optical materials.

Additives and Colors

The characteristics of many plastics, particularly of thermoplastic materials, can be changed by adding so-called additives.

UV-filter materials are frequently added to improve the stability of materials in the sunlight. Sometimes nonstabilized plastics tend to yellow (transmittance decreases). Color filters can also be generated by coloring. Infrared coloring permits the production of optical elements which appear black in the visible spectral range. These materials are particularly used for sensors, optical scanners, or remote control front windows. Even characteristics, such as temperature stability, can be influenced by additives.

Manufacturing Methods

The manufacturing processes for glass and plastic optics are totally different. Glass lenses are made by a grinding and polishing process. By contrast, typical manufacturing methods for precise technical plastic optical parts are diamond turning of plastic blocks, injection, and compression molding of granulates. Other methods, such as casting, have their application in special fields like ophthalmics (eyeglasses). In any case, processing transparent plastic materials for optical purposes should take place in a clean environment.

Diamond Point Turning

Diamond point turning is an ultra-precision machining process carried out on special high-performance, ultra-precision diamond CNC-machining systems.

In combination with vibration isolation systems, digital signal processor-based machine control and integrated measuring systems with nanometer programming resolution manufacturing of optical surfaces is possible. This method can be expanded to full 3D-milling systems. With such an arrangement the generation of free-form surface profiles is possible.

Because of the long production time and the high machine costs, this technology is used for prototyping in plastics and nonferrous metals, mold inserts, and series production plastic parts – which cannot be manufactured by injection molding (because of size or precision demands).

Injection Molding

Injection molding produces several parts per shot (molding cycle) in single or multicavity tools. This production method is used for most of the plastic optic parts. Other fabrication techniques are used only when molding is inappropriate.

A plastic injection-molding machine consists of a fixed platen, a moving platen, a clamping unit, and an injection unit. Molding of optical parts requires special machine configurations and auxiliary equipment.

Compression Molding

Compression molding is used in the making of Fresnel lenses or other micro-optical structures. The material is pressed between heated platens with accurately defined temperature cycling during pressing. Mold inserts are formed by electroplated copies, replicated from master structures. These masters can be, for instance, diamond turned structures or diffraction gratings made by holographic methods.

The compression molding process allows realizing small structures with high aspect ratios and tight angular and positioning tolerances.

Manufacturing Process

Optical and System Design

On the one hand, optic design for plastic optics in principle uses the same mathematical algorithms as optic design for glass. On the other hand, designing plastic optics requires a profound understanding of the material properties and the manufacturing processes. Knowledge of production technologies, material characteristics, and assembling methods, together with design expertise, are needed to fully exhaust what precision plastic optics can be. Simply substituting the indices of refraction and re-optimizing, the design will not succeed. Expert design assistance is essential at this stage.

Designing plastic optics with modern design tools allows greater design freedom. The advantage of combining integral mounting structures with the optical surfaces to create mounting flanges, alignment, and snap features, provides the ability to automated assembly. Aspheric, cylindric, or toroidal surfaces are as easy to realize as spherical ones. Microstructures such as diffractive optical elements can be integrated too.

Prototyping

After designing a plastic optical system, lens prototypes can be made by diamond point turning in various plastic materials. The best surface finishes can be obtained with PMMA. Materials such as polycarbonate, pleximid, or Zeonex do not yield smooth surface finishes.

A major problem is the availability of semi-finished plastic blocks in various materials. For PMMA, a wide variety of bars or plates are available. For other materials or colored options, semi-finished items have to be made by injection molding. Because of the high manufacturing costs, diamond point turning is only recommended for making a limited number of prototypes to verify functionality of the optical design and to perform first tests.

In this stage optical systems are normally assembled from single elements. Housing parts often are made from aluminum. The resulting surface quality and system performance cannot be a validation of the manufacturability by injection molding. To get reliable knowledge about this, making a molded prototype from a single-cavity prototype mold is recommended.

The Injection Mold

A high-quality injection mold is obviously essential for precise plastic optic parts. The parts can never be better than the tool – but good tooling, however, does not guarantee good parts. A strong understanding of the whole manufacturing process is the key to producing precision plastic components.

Any injection mold consists of three main parts. The upper half, which is affixed to the injection-side platen, the lower mold half, which is affixed to the ejector-side platen, and the mold-ejection mechanism. Guide pins and taper locks ensure proper alignment of the mold halves.

Production volume and precision required from a tool influences the selection of mold materials and built-in maintenance features.

Single-cavity and multi-cavity molds from two to eight, 16, or even 32 cavitys are used for plastic optic parts.

Thermoplastic materials shrink during cooling in the mold. This geometric effect has to be compensated for in the injection mold. Exact shrink rates can be calculated and the tool can be modified.

For manufacturing optical plastic parts by injection molding, the optical surfaces (plano or aspherical shapes, diffractive, conical, lenticular, and cylindrical surfaces) are generated as separate inserts in the tool.

Aspheric inserts are manufactured by two steps. First a best-fit curve is generated on a stainless steel substrate. The substrate is then subjected to a nickel-plating process (electroless nickel) that deposits a thin layer of nickel (up to 500 μm). In the second step, single-point diamond turning produces the final aspheric or diffractive curve in the nickel. Because the hardness of a nickel-plated insert is less than that of a steel spherical insert, it will be more susceptible to scratch defects (Figure 4).

Pre-Production

A pre-production stage is recommended to check the manufacturing process. Typically this is done by using a single-cavity prototype mold. This mold can be used to find optimal molding conditions.

Optical and mechanical design can be verified with real molded components and design revision is possible to affordable conditions. Often the prototype molds is used to start limited production since production tooling may take much more time.

Series Production

For series production of high volumes, multicavity production molds are required. Depending on quality,

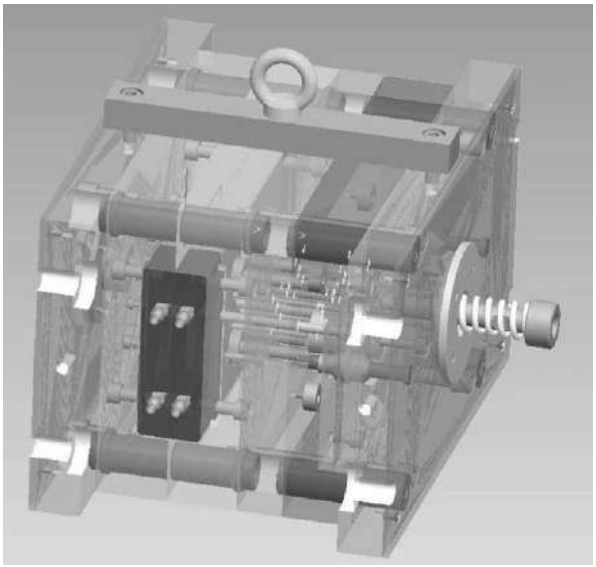

Figure 4 Injection mold.

volume, throughput, and cost, the production tool may have 2, 4, 8 or even up to 32 cavities. The production molds function for at least several hundred of thousands of injection cycles.

Coating and Component Assembly

Coating

Because of their limited temperature and UV resistance, plastic lenses must not be coated in an elevated temperature and radiation environment. During the deposition of thin films onto plastic, the coating chamber temperature is significantly lower than that for glass optics. This requires deposition techniques such as ion-assisted deposition to apply antireflective, conductive, mirror, and beamsplitter coatings.

Today, multi-layer dielectric coatings are routinely deposited on plastic components. Typical broadband antireflection coatings reduce reflection to about 0.5% per surface across the entire visible spectrum. Narrowband, multilayer antireflection coatings can achieve surface reflectances less than 0.2%. Multilayer dielectric coatings can be modified easily to scratch-resistant designs for front lenses and windows.

Several front and back surface reflector coatings are available for plastic substrates. Standard coating metals include aluminum, silver, and gold. Aluminum coatings provide surface reflectances greater than 88% across the visible spectrum and gold coatings greater than 95% for the near-infrared region.

Component Assembly

Many advantages of using plastic optics will come out in manufacturing components. The integration of optical, mechanical, and electronic elements allows building low-cost polymer optical components such as small camera lenses or scanner heads.

Plastic optical assemblies are usually carried out in a clean room environment by manual, semi-, or fully automated processes. The components have to be designed to ensure ease of assembly. Snap-on features, UV-cementing, heat-staking, and ultrasonic and laser welding, can also be employed with plastic materials.

Since most optical tolerances are additive, it is essential to establish quality check points in the manufacturing process to sort out nonconforming sub-assemblies.

Automatic in-line optical performance, monitoring such as MTF testing, can be implemented. SPC techniques should be used here to ensure the process is not drifting out of the controllable range.

Summary

For centuries, glass proved to be the material of choice for all optical applications. During the twentieth century, applications for glass elements and complex optical systems greatly expanded. Early uses of plastic molded elements included toy objectives, gauge windows, and watch crystals. Throughout the 1970s and up to the present day, high-grade polymers were developed specifically for optical applications. These advancements in materials, coupled with improved mold design have enabled plastic optics to replace glass optics in a wide and growing number of applications.

Plastic optic technology permanently expands their traditional limits and fields of application. On the one hand, plastic optic is still limited by material properties but, on the other hand, design and assembling freedoms allow new approaches.

See also

Diffraction: Fresnel Diffraction. **Geometrical Optics:** Lenses and Mirrors. **Optical Coatings:** Thin-Film Optical Coatings. **Optical Materials:** Color Filter and Absorption Glasses.

Further Reading

Corning Precision Lens Inc. (2000) *The Handbook of Plastic Optics*, 2nd edn. Cincinatti, OH: Corning Precision Lens Inc.

OPTICAL MATERIALS

Sculptured Thin Films

K Robbie, Queen's University, Kingston, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical materials, like all materials, ultimately derive their physical properties from their chemical composition and physical form. Aluminum behaves differently from copper because of a chemical difference, while graphite behaves differently from diamond because of a crystallographic difference. For materials that are not infinite perfect crystals, an important contribution to their electromagnetic (optical) response comes from the morphological form of the material. Morphological form describes the physical structure of the material: a perfect crystal is one type of morphological structure, a porous fibrous bundle is another. An example of a morphological difference is that between snow and ice. Snow and ice are chemically identical, composed of solid water, but they exhibit different optical properties because of a difference in morphology. The polished glasses and large perfect crystals that comprise most optical components have morphologies more like ice than snow; their morphologies are dense and crystalline, with few and small disordered 'defects'. Unlike ice, snow is rich with morphological structure (see Figure 1) arising from the complex growth of each flake in the clouds, and the aggregate growth of snow on the ground. The morphology of snow scatters all colors of visible light, giving it an opaque white appearance, while ice is transparent and nearly colorless.

Sculptured thin films (STFs) are materials with controlled and designed morphology that exhibit a rich behavior unlike that of the constituent materials in bulk; they are the snow to most optical materials' ice. Unlike snow, sculptured thin film morphology is designed to scatter light in a controlled manner to produce a desired optical response. A sculptured thin film could be a coating applied to an optical element where the coating converts polarizations of light, circular or linear, but is composed entirely of an optically isotropic material (which doesn't usually affect the polarization of light). The film morphology is designed to create the desired optical properties. Sculptured thin films are also useful for magnetics, where an example of a sculptured thin film is the metal film coated onto plastic tape for video, audio, and data recording. The film is deposited under conditions that impose an elongated shape onto the magnetic domains, altering the magnetic properties. The morphological form of a sculptured thin film defines the way it responds in electromagnetic, chemical, and biological interaction, and there is hope that these unique materials will someday form the core of technologies that make the world a better place.

Structure in Thin Films

As chemical composition and morphology determine the electromagnetic response of materials, it would be wonderful to have a technique where we could build materials atom by atom, placing each exactly where we wished, and varying the element at will. With this level of control it would be possible to design and fabricate materials for almost any conceivable optical application. In some sense this is what is done now to fabricate lasers and optical switches. Microelectronic and micromachining processes have produced fantastic optical devices by controllably mixing materials (mostly semiconductors) and patterning designed structures. These techniques typically employ a repeated series of thin-film depositions and photolithographic patterning. Examples of thin-film coatings include antireflection coatings on eyeglasses, hard coatings on grocery store barcode scanners,

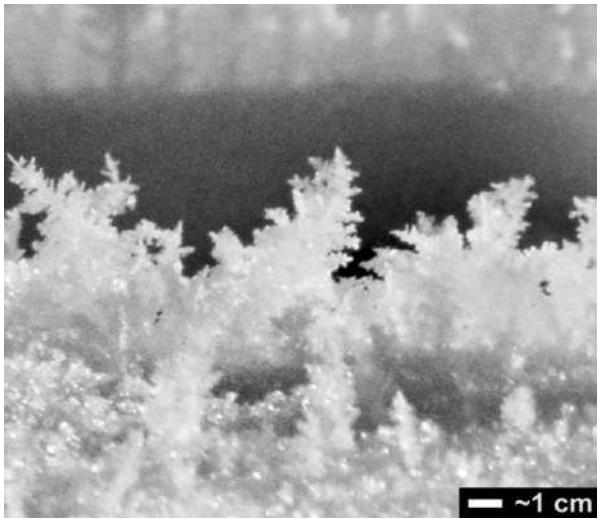

Figure 1 Photograph of snow accumulating at −20 °C. Photo courtesy of Chelsea Elliott.

ultraviolet blocking aluminum inside potato chip bags, and all integrated electronic circuits. Thin films are usually produced by vapor deposition in vacuum, and can be made with most elements and inorganic compounds (and a growing list of organics). Most thin films grow as uniform, dense, two-dimensional (2D) layers, and three-dimensional structure (3D) is produced by removing portions of the film with photolithography and etching. Photolithography is a process where ultraviolet light is projected through a mask and microscope system onto a photosensitive polymer layer atop the thin film. The light hardens areas of the polymer in the pattern of the mask, preventing them from dissolving in a chemical developer solution. Exposed portions of the thin film can then be etched away with chemical or other means. Repeating as desired, many steps can create complex 3D designed structures. Computer chips and other integrated circuits are semiconductor crystals covered with a layered structure of conductors (aluminum, copper, platinum) and insulators (metal and semiconductor oxides) patterned with thin film deposition and lithography. Semiconductor lasers and many other optical devices are made this way as well, usually employing compound semiconductors (GaAs, AlGaAs, GaAsP, InP, etc.). The micrometer-scale patterning made possible with thin-film deposition and etching techniques has produced a wealth of semiconductor optical devices.

Optical lithography, however, is limited in size scale to dimensions comparable to the wavelength of light used to expose the photomask. Modern UV lithography can pattern 100–200 nm structures. Visible light has wavelengths from 400 nm (violet) to 700 nm (red). Shrinking the dimensions of physical structures is a great benefit for optical devices as it allows light to be controlled coherently without strong random scattering. Structures with dimensions comparable to the wavelengths of light strongly scatter light waves; scratches on optical components or imperfections in materials with dimensions comparable to the wavelength are considered defects, and are usually deleterious. However, materials with morphology much smaller than the wavelength (nanometer structure, or nanostructure) are seen by propagating light as homogeneous, with polarization and dispersion effects characteristic of the underlying structure. Sculptured thin films are materials where complex nanostructure has been created by controlling growth parameters dynamically during deposition of a thin film. The simplest sculptured thin film might be one where the temperature is periodically varied during thin-film deposition (although this is a very simple example, lacking the anisotropy central to the sculptured thin-film concept). Variation in the growth process with temperature affects film density, producing a porosity variation (periodic inhomogeneity) as a function of film thickness. Varying porosity produces varying refractive index, and this film would be an optical interference filter. The largest limitation of sculptured thin films is that all attainable structures must be characteristic of the growth process. As all morphological control is accomplished by varying film growth parameters, structure that is not characteristic of thin film growth is not possible. Fortunately, thin film deposition and growth is sufficiently complex to allow the fabrication of a large range of useful structures. As with lithographic patterning, the structural control is two dimensional, but now the third dimension is constructed by continuing the deposition, not by repeating the deposition and lithography cycle. Complex structures can be produced on the scale of tens or hundreds of atoms (approximately tens of nanometers) with controlled 3D morphology. This reduction in size scale compared to lithographic patterning is a huge benefit. Materials can now be designed and created with subwavelength structure to coherently scatter light. Because the morphological structure (<100 nm) is much smaller than the wavelength of visible light (~500 nm), light traveling in a sculptured thin film sees a medium with averaged properties. The material can be designed to control refractive index through a wide range, and birefringence and circular polarization effects are controllable.

The morphological structure of thin films is a vast topic. Thin films are produced with a nonequilibrium growth process, somewhat like the growth of clouds, and the mechanics of the growth determine the structure, which is often very different from that seen in a bulk sample of the same substance. Crystal and morphological structure vary with deposition process conditions, and vary during deposition to record a history of growth much like the rings of a tree, or stratified rock. If chemical composition is varied during the deposition, that change will be reflected in layers in the film. If vapor energy or temperature is varied, variations in density and crystal structure will result. Thin film growth is an aggregation process (the film accumulates over time), and exhibits many of the features of similar processes, sometimes producing materials with structures similar to broccoli or clouds. This self-affine (self-similar) structure is observed over a large range of sizes under some film deposition conditions. Under other deposition conditions, crystallographic effects dominate growth and films become layers of faceted crystals. In sculptured thin films, these growth characteristics are controlled and exploited in a functional manner to

design material properties. For example, magnetic video and audio tapes that have the symbol ME, abbreviating 'metal evaporated', employ a structured magnetic thin film that has superior magnetic response for recording. Similar effects are observed in optical materials, and sculptured thin films appear promising for improving existing optical devices or enabling new ones to be developed.

The most studied form of sculptured thin films is that produced by an asymmetric deposition process. In deposition by vacuum evaporation, a thin film is fabricated by evaporating some source material in vacuum, and condensing a vapor onto a substrate. Eyeglasses, mirrors, television phosphor screens, etc. are coated this way. In most thin films the vapor arrives straight down (perpendicular) to the plane of a flat substrate. The film structure reflects the symmetry of the vapor, and is typically composed of a dense layer with some defects or voids oriented perpendicular to the substrate surface. A thin film exhibiting this morphology is shown in the scanning electron microscope image of Figure 2. When vapor arrives asymmetrically onto a substrate, however, that asymmetry is reflected in the growth, producing structural asymmetry in the resulting film, and altered physical properties. A second scanning electron microscope image is shown in Figure 3, of a sculptured thin film where the orientation of the substrate relative to the vapor is changed twice, producing three layers with different morphologies. The thin uniform gray at the absolute bottom of the image is the edge of the silicon crystal substrate which has been cleaved to reveal an interior section of the film. The bottom and first film layer is inclined to the right, toward the vapor source for the duration of this first part of the deposition. The middle and second film layer has a twisted structure created by slowly rotating the substrate relative to the vapor, while keeping the tilt fixed. The top layer has a vertical structure created by rapidly rotating the substrate. The entire film was produced with the vapor arriving at an angle of 45°, measured from either the plane of the substrate or its perpendicular. Varying the tilt angle has a profound effect on the morphology of thin films. The void structure in these films records the geometry of the deposition. When we cleave a film we are able to observe a 2D slice of this rich morphology. At large tilt angles, glancing angle deposition (GLAD) produces films with highly porous and complex structure. At these angles, the film structure becomes isolated columns of material where growth variations are now reflected in variations in the shape of individual columns. At less oblique vapor angles, growth variations are reflected in the shape of networks of atomic-scale voids within a largely dense film. Indeed, for the best optical devices fabricated to date (with the serial bideposition technique described later) it is difficult to distinguish any variation in structure even using a high-resolution scanning electron microscope. The film deposition process can be designed to control the morphology of the growing film, and to vary the structure as a function of time and thickness. The morphology of these films determines their optical properties; optical birefringence, second harmonic generation, filter bandwidth and attenuation, etc. can all be controlled in sculptured thin films.

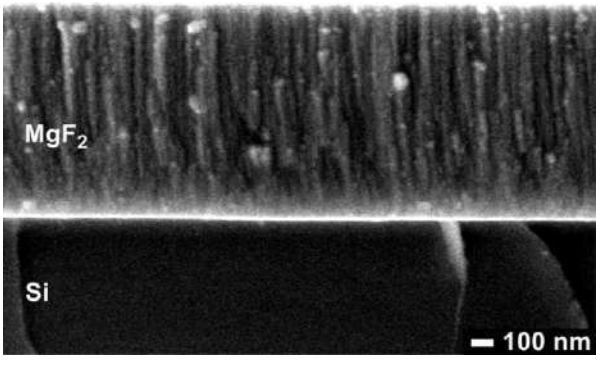

Figure 2 Micrograph of MgF_2 film deposited onto a 75 mm silicon wafer substrate, then fractured to reveal the edge shown.

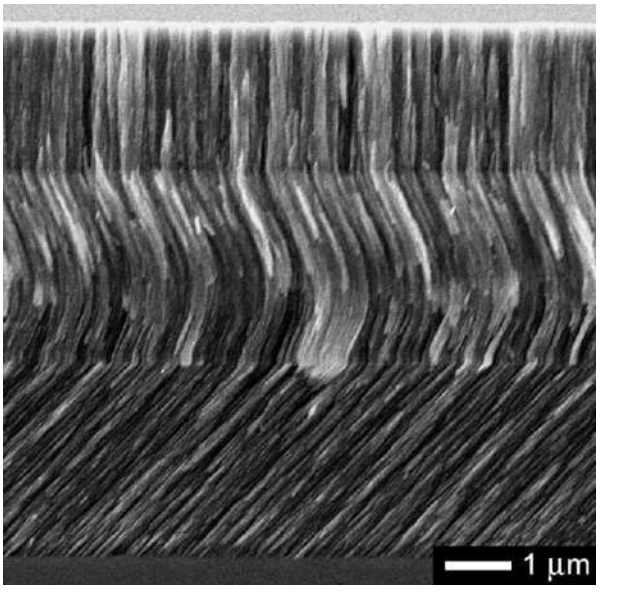

Figure 3 Micrograph of SiO film deposited with a tilt angle of 45° with one layer of no rotation, one layer with one rotation, then one layer with rapid rotation.

Optical Thin Films

Light can be described as a traveling wave with electric and magnetic fields oscillating perpendicular to each other, and to the direction of propagation. If the orientation of the electric field remains constant as the wave propagates, the light is said to be linearly polarized. If the orientation of the electric field rotates through a circular path, the light is circularly polarized, and can be left or right handed. In the most general description, the electric field traces an ellipse during propagation and the light is said to be elliptically polarized. The experimental study of light polarization is called ellipsometry. When light propagates through vacuum or air, the state of polarization remains constant. When light propagates through some types of materials, however, crystallographic or morphological asymmetry (or more correctly, anisotropy) results in different propagation conditions for different polarizations. Upon passage through these materials, the orientation and form of the state of polarization is changed, for example from left circularly polarized (LCP) to right circularly polarized (RCP). When multiple layers of anisotropic materials are combined into an optical circuit, as in the sculptured thin film concept, highly controlled optical response characteristics can be created.

The discussion of the optical properties of thin films begins with definitions of isotropic, uniaxial, and biaxial materials. Isotropic materials have optical properties that are independent of the direction that the light propagates through them. Vacuum and air are isotropic, as is glass. Rotation of the optical material relative to the light propagation direction has no effect on how the light propagates, and the state of polarization remains constant. Uniaxial materials have properties that vary with direction, but with one preferred direction (or axis) where response is independent of rotation. Biaxial materials have properties that are even more dependent on propagation direction, and require two axes to describe their optical properties. For biaxial materials there is no propagation direction where the optical response is independent of rotation. The mineral calcite is an example of a uniaxial material where the refractive index, and therefore the speed of light propagation, varies with direction relative to the crystal structure. This is a direct consequence of the trigonal crystal structure of calcite, and is an example of linear birefringence. Birefringence is the effect caused by anisotropic materials whereby light of different polarizations, even within one ray of light, travel with different speeds. When light propagates in a birefringent material, the orientation of the light's electric field relative to the material's optical axis determines the speed of propagation. Light with different polarization (orientation) propagates at different speeds, and a beam's polarization state is altered on passage through the material. Thin films can be isotropic, uniaxial, or biaxial, depending on the conditions under which the film is deposited. Sculptured thin films are films fabricated in a manner that controllably enhances the anisotropic optical response to produce materials with designed optical properties based on control of their morphological structure. As in calcite where the anisotropic crystal structure yields a uniaxial optical response, anisotropic morphological structure in sculptured thin films yields biaxial optical response. A plan view scanning electron microscope image of an anisotropic thin film is shown in Figure 4. The film was deposited with two vapor sources with the substrate placed between and the vapor arriving from both sides at an angle of 70° from the substrate perpendicular. This film is similar to those produced with the serial bideposition technique, and is optically birefringent. Polarization filters and other optical components can be constructed by designing the morphological structure of the thin film.

As discussed above, when a thin film is deposited from a vapor arriving perpendicular to a flat substrate surface, a structure composed of small (tens of nanometers diameter) columns of film material grows, with the columns oriented perpendicular to the substrate (Figure 2). This film is optically uniaxial, with the preferred axis perpendicular to the substrate, and parallel to the vapor flux. Light propagating along this optical axis (perpendicular to the substrate) travels at one speed regardless of polarization, and independent of rotations of the substrate around this axis. Light propagating along other paths will travel at a speed that depends on the polarization of the light

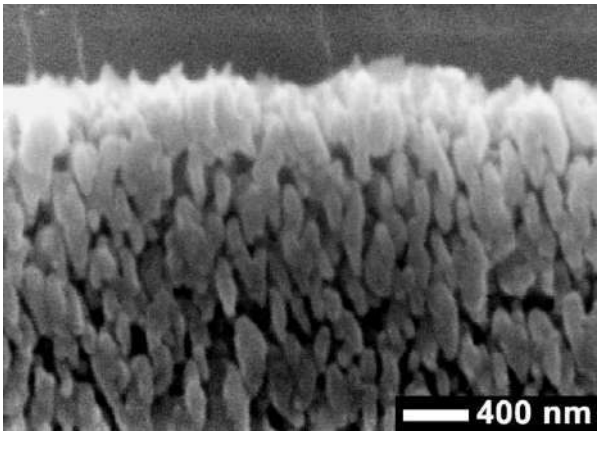

Figure 4 Micrograph of MgF_2 film, viewed from above, deposited with simultaneous growth from two vapor sources.

(orientation of the electric field relative to the propagation direction). The effect is small, and arises from the morphological structure of the film. More pronounced optical response is produced when the vapor arrives onto the substrate at an angle from the substrate perpendicular, producing optically biaxial materials, and enabling exciting polarization and filtering effects. If a thin film of magnesium fluoride is deposited onto a polished piece of glass with the vapor arriving at 45° (relative to an axis perpendicular to the substrate), a slanted columnar film is produced. The morphology of the film is tightly packed columns of polycrystalline magnesium fluoride with diameters between 50 and 200 nanometers (nm), inclined toward the vapor source, and penetrated by a network of atomic-scale voids. Because of the anisotropic nature of the deposition, the voids are anisotropic, and the film is optically biaxial. If the substrate is then held at that tilt angle and is rotated during deposition, the anisotropy in the structure rotates with the deposition, and an important type of sculptured thin film is produced. The electromagnetic description of this type of film is a 'helicoidal bianisotropic medium', describing both the helix shape of the rotation of the structure and the local bianisotropic (biaxial) physical, and optical, structure. Films with helicoidal symmetry are ideally suited for studies involving circular polarizations of light, and many demonstrations of sculptured thin films have involved these films and circularly polarized light. Like many organic molecules and some mineral crystals, these film are chiral, existing in two mirror image forms, called enantiomers. A right-handed structure rotates clockwise traveling away from an observer; a left-handed structure rotates counterclockwise. Also like some organic molecules and mineral crystals, helicoidal films are optically active – they affect the polarization of light. The observed effect depends on the wavelength of light relative to the physical helicoidal structure. If these two lengths match, a strong resonance effect is observed where the light wave coherently scatters from the anisotropic structure. Filters to produce and convert circular polarized light have been constructed with sculptured thin films of this type.

The range of structural control over the morphology of thin optical films limits the ability to tailor optical response for desired applications. While optical systems must always be treated as a whole to determine response properties, the description of some basic elements is useful. A film deposited at normal incidence will be uniaxial with the axis perpendicular to the substrate. For light propagation along this axis (common for optical systems), the material exhibits no birefringence. This structure is often called an isotropic element even though the columnar structure is slightly optically anisotropic. A truly isotropic layer could be produced by ion bombardment or elevated temperature during deposition, either of which would have the effect of eliminating the columnar structure and reducing anisotropy in the film. The second elemental structure is a slanted columnar film that is commonly described as a matchstick morphology. It is produced by changing the vapor incidence away from the substrate normal, leading to an inclination of the column structure toward the vapor source, as described above. Further structures can include: helicoidal structure where the anisotropy rotates through the thickness of the film, and periodically bent nematic structure where the anisotropy tilts in an s-shaped pattern through the film (fabricated with a complex pattern of rotations and depositions). These four elementary microstructures are shown in Figure 5. Electrical, optical, and magnetic anisotropy can also be increased by depositing obliquely from two sides instead of one, and this technique (called serial bideposition) can be used to increase the anisotropy in sculptured thin films. At extremely oblique deposition angles, a regime called glancing angle

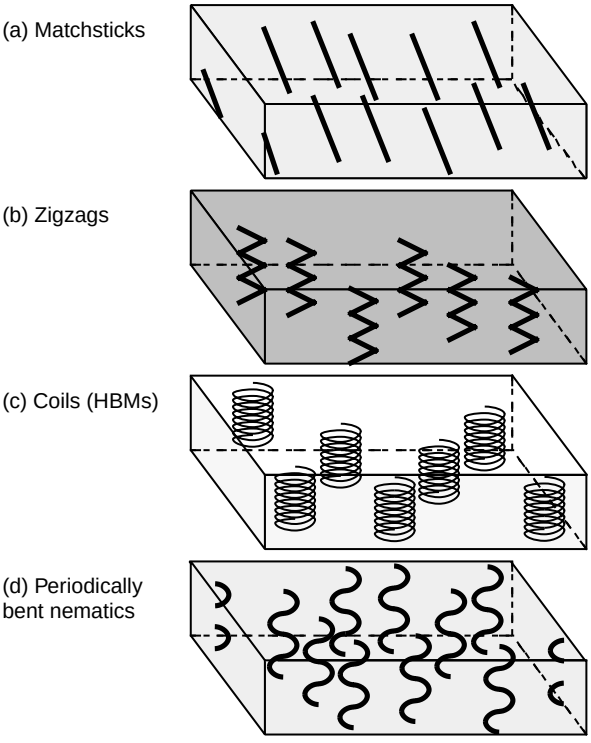

Figure 5 Elementary microstructures of sculptured thin films. From Lakhtakia A, Messier R, Brett MJ and Robbie K (1996) Sculptured thin films (STFs) for optical, chemical and biological applications. *Innovations in Materials Research* 1(2): 165–176, reproduced with permission.

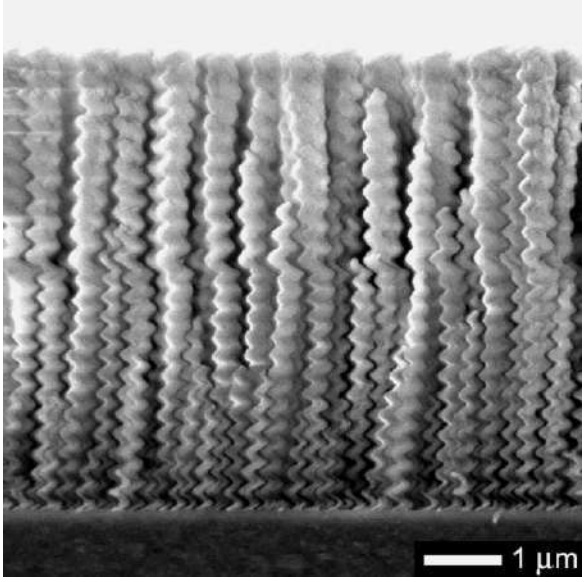

Figure 6 Micrograph of MgF_2 film deposited with a tilt angle of 85° and substrate rotation clockwise for the first half and counterclockwise for the second.

deposition, a rapid increase in porosity with approximately constant column tilt results from a competition between deposition processes. Films deposited here are highly porous and strongly scatter light, limiting their usefulness for conventional optical devices, though some application has been seen in these highly porous films when impregnated with liquid crystals, and switched similarly to liquid crystal displays. An example of a highly porous optical filter with circular polarization effects is shown in Figure 6, where two layers of helicoidal porous films are deposited with different handedness and pitch. Sculptured thin films for most optical devices are deposited at less oblique angles to reduce porosity and therefore diffuse light scatter (haze) as well as sensitivity to humidity and other environmental changes.

Combining the concepts of film growth and optical response to morphology leads to the hoped-for vision of sculptured thin films: optical circuits in a chip. Combining the elementary microstructures described above, and employing powerful computer growth modeling and optical design programs, researchers are currently defining and producing sculptured thin films with increasingly complex optical response characteristics. There are substantial steps required to move from the fairly simple optical filters fabricated today to large-scale devices with complex functionality, but continuing success and increasing interest suggest that sculptured thin films could become a dominant optical technology in years to come.

Current Research

While thin films deposited at off-perpendicular incidence angles have been studied for many years, the last decade has produced significant new insights into these materials, leading to the concept of sculptured thin films. Understanding that morphologically rich materials can be used to produce new optical effects has yielded some impressive demonstrations. Research is also underway to explore these materials for chemical, magnetic, biological, and other applications.

Significant research has been conducted in the past few years to develop and extend the concepts of sculptured thin films. Optical studies, both theoretical and experimental, have revealed a range of remarkable properties including circular birefringence and Bragg reflection, liquid crystal alignment and switching, optical interconnection, and even suggestions of photonic bandgaps. Studies of optical sensing of gases and other materials is progressing, and simulation and modeling of the growth processes is ongoing. The following are two examples of optical materials that have been fabricated with sculptured thin-film techniques; one is a polarization discriminatory light-handedness inverter and the second is a liquid crystal switching cell that can be used to electrically modulate the transmission of polarized light. The handedness inverter converts left circularly polarized light to right circularly polarized within a spectral band, and blocks right circular polarized light. It is constructed by combining a layer with vertical columnar, or matchstick, morphology together with a layer with helicoidal morphology. In keeping with the sculptured thin film concept, the films are deposited sequentially during the same deposition, and optical properties are monitored during growth. The transmission spectrum for the devices is shown in Figure 7, where T_{RL} and T_{LR} are the light transmittances of right circularly polarized light when the incident is LCP, and of LCP when the incident is RCP, respectively. LCP light is almost entirely converted to RCP upon transmission, and RCP light is blocked. The bandwidth of the devices is about 50 nm, determined by matching of the circularly polarized light with the layer with helicoidal morphology.

When films are deposited with helicoidal morphology at glancing angles, i.e., glancing angle deposition, highly porous structures are produced where the columns stand alone and are individually shaped by the deposition process. Optical devices can be constructed by filling these porous films with gases or liquids, and there have been demonstrations of display-type switching with liquid crystals, and discussions of chemical sensing possibilities.

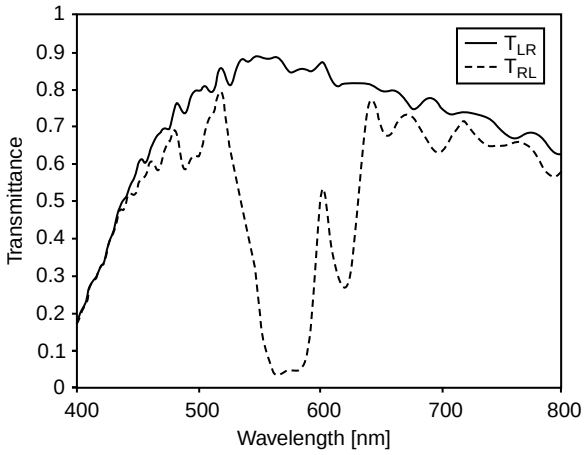

Figure 7 Transmission spectrum for visible light through a polarization-discriminatory light-handedness inverter. Adapted with permission from Hodgkinson IJ, Lakhtakia A and Wu Q-H (2000) Experimental realization of sculptured-thin-film polarization-discriminatory light-handedness inverters, *Optical Engineering*, **39**: 2831–2834.

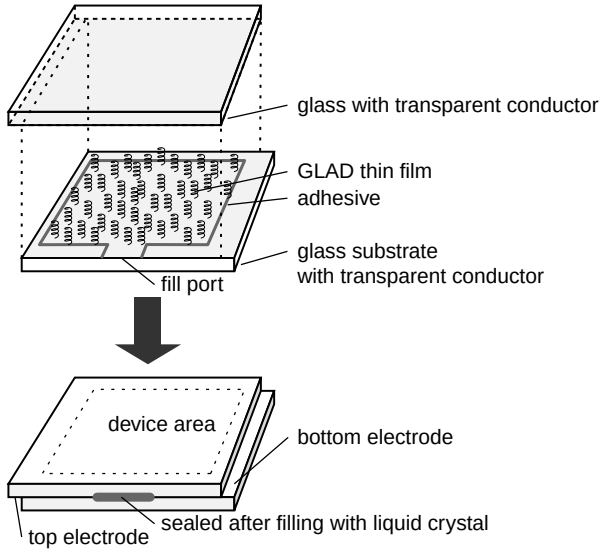

Figure 8 Liquid crystal switching cell schematic with highly porous glancing angle deposited (GLAD) film between glass plates coated with transparent electrodes. Reproduced with permission from Sit JC, Broer DJ and Brett MJ (2000) Alignment and switching of nematic liquid crystals embedded in porous chiral thin films. *Liquid Crystals* 27: 387–391, www.tandf.co.uk/journals.

The device shown in Figure 8 is a liquid crystal switching cell where the space between the electrodes, usually composed solely of liquid crystalline material, is now filled with a highly porous helical structured film. The film structure is somewhat similar to helicoidal film produced for the handedness inverter discussed above, but glancing angle deposition has produced a film of individual formed columns instead of a more dense film with a helicoidal void

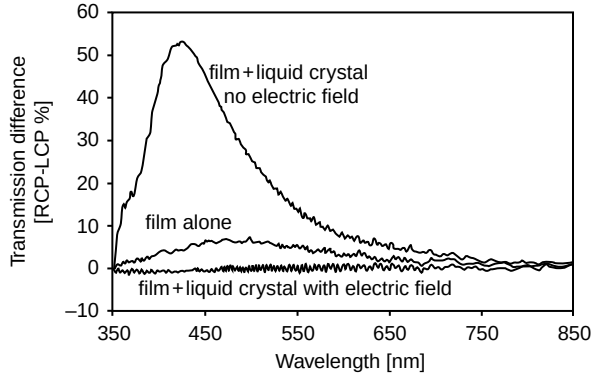

Figure 9 Transmission spectrum for visible light through a liquid crystal cell where the difference between right- and left-circularly polarized light transmission is eliminated with applied electric field. Adapted with permission from Sit JC, Broer DJ and Brett MJ (2000) Alignment and switching of nematic liquid crystals embedded in porous chiral thin films. *Liquid Crystals* 27: 387–391, www.tandf.co.uk/journals.

morphology. When the porous portions of the film are filled with simple nematic liquid crystalline material, the orientation of the rod-like molecules is strongly influenced by the helical columns. An orientational phase occurs in the liquid crystals mirroring the helical structure of the columns in the film, and producing optical properties similar to chiral nematic liquid crystals. The discrimination of left and right circularly polarized light increases relative to the unfilled film. When an electric field is applied across the liquid crystal cell, dipole moments in the liquid crystal molecules force them to align with the field, and chiral optical properties vanish (see Figure 9). Used as an electro-optic filter for circularly polarized light, this cell can switch light transmission on and off. Initial work has not shown clear benefits of this system over conventional liquid crystal display technology, but control of liquid crystal orientation throughout the cell instead of just at the glass plates appears promising. Possibly more complex optical circuits, incorporating layers of porous films filled with liquid crystals, will be useful.

Theoretical and experimental study of sculptured thin films is in its infancy, and many exciting uses for these materials are expected in coming years. Continuing study of thin film growth dynamics will improve and expand the range of structures attainable. Theoretical investigations will aid in the design and understanding of the optical response.

List of Units and Nomenclature

Aluminum (Al)	white metal used in airplanes, mirrors, and microelectronic circuits

Anisotropic	having properties that vary with orientation (rotation)
Biaxial	having anisotropic optical response characteristics that require two axes to describe
Chiral	having a shape that exists in two mirror image forms, called enantiomers
Diamond	carbon crystallographically ordered in a face-centered-cubic lattice, and used in cutting tools and jewelry
Electromagnetic waves	traveling waves of energy that compose light, x-rays, radiowaves, etc.
Epitaxy	growth of a thin film that preserves the crystal structure of the underlying film or substrate
Etching	removing material with chemical or physical atomic scale process
Graphite	carbon crystallographically ordered in hexagonal sheets, and used in pencils and lubricants
Inhomogeneity	varying with translation (position)
Isotropic	having properties that are identical in all directions
Lithography	process where patterns are produced in a sensitive layer by irradiating the layer through a mask with the desired pattern
Magnesium fluoride	ceramic (MgF_2) that is used in thin-film or single-crystal form in many optical systems
Morphology	microstructural shape, or physical arrangement of atoms in a material
Nanometers (nm)	one billionth of a meter, 10^{-9} m
Nematic liquid crystal	optical material composed of rod-shaped organic molecules that usually exhibits uniaxial optical properties
Pitch	physical period for one full rotation of a helical structure
Polarization	orientation and phase of the electric and magnetic fields in a traveling electromagnetic wave; can be linear, circular, or generally, elliptical
Refractive index	the ratio of the speed of light in vacuum to the speed of light in a material
Substrate	object that a thin film is deposited onto, often polished glass or semiconductors
Ultraviolet (UV)	electromagnetic radiation (light) with wavelength shorter than visible light and longer than X-rays
Uniaxial	having optical response characteristics that are anisotropic, but are rotationally symmetric about one axis
Vacuum	absence of matter, typically produced inside sealed glass or steel chambers by removing air with pumps
Vapor flux	moving gas-like collection of material, emitted from a source such as an evaporator or effusion cell, that will condense to form a thin film

See also

Coherence: Overview. **Coherent Lightwave Systems**. **Fiber and Guided Wave Optics:** Overview. **Imaging:** Imaging Through Scattering Media. **Information processing:** All-optical Multiplexing/Demultiplexing; Coherent Analogue Optical Processors; Optical Neural Networks. **Instrumentation:** Ellipsometry. **Optical Processing Systems**. **Optical Coatings:** Anti-Counterfeiting and Decorative Coatings; Optical Black Surfaces; Thin-Film Optical Coatings. **Optical Materials:** Measurement of Optical Properties of Solids. **Polarization:** Introduction; Matrix Analysis. **Scattering:** Scattering from Surfaces and Thin Films; Scattering Theory. **Semiconductor Materials:** Band Structure Engineering. **Terahertz Technology:** Coherent Terahertz Sources.

Further Reading

Ball P (1997) *Made to Measure*. Princeton, NJ: Princeton University Press.

Barabasi A-L and Stanley HE (1982) *Fractal Concepts in Surface Growth*. Cambridge, UK: Cambridge University Press.

Collins PJ (1990) *Liquid Crystals*. Princeton, NJ: Princeton University Press.

de Gennes PG and Prost J (1993) *The Physics of Liquid Crystals*. Oxford, UK: Clarendon Press.

Hodgkinson IJ and Wu Q-H (1997) *Birefringent Thin Films and Polarizing Elements*. Singapore: World Scientific Publishing.

Hodgkinson IJ, Lakhtakia A and Wu Q-H (2000) Experimental realization of sculptured-thin-film polarization-discriminatory light-handedness inverters. *Optical Engineering* 39: 2831–2834.

Hodgkinson IJ, Wu Q-H, Thorn KE, *et al.* (2000) Spacerless circular-polarization spectral-hole fitters using chiral sculptured thin films: theory and experiment. *Optics Communications* 184: 57–66.
Lakhtakia A, Messier R, Brett MJ and Robbie K (1996) Sculptured thin films (STFs) for optical, chemical and biological applications. *Innovations in Materials Research* 2: 165–176.
Liu F, Umlor MT, Shen L, *et al.* (1999) The growth of nanoscale structured iron films by glancing angle deposition. *Journal of Applied Physics* 85: 5486–5488.
Messier R, Venugopal VC and Sunal PD (2000) Origin and evolution of sculptured thin films. *Journal of Vacuum Science and Technology* 18: 1538–1545.
Ohring M (1992) *The Materials Science of Thin Films*. New York: Academic Press.
Robbie K, Brett MJ and Lakhtakia A (1996) Chiral sculptured thin films. *Nature* 384: 616.
Singh ON and Lakhtakia A (eds) (2000) *Electromagnetic Fields in Unconventional Materials and Structures*. New York: John Wiley.
Sit JC, Broer DJ and Brett MJ (2000) Alignment and switching of nematic liquid crystals embedded in porous chiral thin films. *Liquid Crystals* 27: 387–391.
Toader O and John S (2001) Proposed square spiral microfabrication architecture for large three-dimensional photonic band gap crystals. *Science* 292: 1133–1135.
Wang L and Clancy P (2001) Kinetic Monte Carlo simulation of the growth of polycrystalline Cu films. *Surface Science* 473: 25–38.

Smart Optical Materials

P M Martin, Pacific Northwest National Laboratory, Richland, WA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Smart optical materials are those materials whose optical properties (transmittance, reflectance, absorptance, emittance, fluorescence) can be controlled by external stimuli; either by an applied electric field or voltage, mechanical stress, incident light or electromagnetic field intensity, temperature variation, or time duration. This includes electrochromic, photochromic, thermochromic, nonlinear optical, light emitting, fluorescing, emissive, and piezochromic materials. These materials can be in various states, such as solids, gels, composites, and mesophases. Electrochromic and photochromic materials, however, are used most extensively.

The mechanism for adaptive control of light must be reversible, have a well-defined response time, and have an activation threshold, otherwise, the material is considered passive. The primary mechanisms that can cause variations in optical properties are crystal phase transitions, nonlinear effects in the polarizability, optical energy band transitions and band bending, compositional changes, defect and color center formation, and lattice strain. Crystal phase transitions change the bonding morphology of a material, which affects the energy bandgap. Most optical materials have several possible crystalline phases, depending on temperature. The optical properties of the anatase and rutile phases of TiO_2, for example, are very different, and result from different energy bandstructures (recall that the energy bandstructure depends on lattice spacing and crystalline plane orientation). The crystalline phase of several thermochromic materials changes from the martensite structure to the austenite structure with change in temperature, and an associated change in optical properties. Compositional changes are caused by chemical or physical reactions in the material, such as a metal ion bonding to an atom in the lattice, or loss of an atom in the lattice. The reaction can be caused by an energetic photon or ion driven by an electric field; electrochromic materials fit into this category. If their energy is high enough, incident photons can create defect states in the bandgap of materials, which in turn form color centers that absorb light. Modulated electric and electromagnetic fields can cause changes in the linear and nonlinear polarizability of the atoms in a lattice. Nonlinear optical and piezochromic effects are based on this mechanism.

Photochromic Glass

Photochromic glass is one of the most widely used smart optical materials, and definitely the most widely marketed. It has been around for over 30 years with hundreds of associated patents. The transmission of photochromic materials adjusts to the light level of their environment, but these materials give no output information for data processing or communications. They do, however, require no other external control mechanism other than the intensity of light (characteristics of a true smart material). Photochromic materials darken reversibly when exposed to sunlight or a specific

optical (UV–LWIR) wavelength band. Absorption in specific narrow or broad wavelength bands is due to the formation of color centers, or point defect states in the lattice. The kinetics of darkening has been described by a two-color center model with two rate constants. Energy (usually in the ultraviolet wavelength range) of the incident photons, ionizes atoms in the glass, thus forming 'F' centers. F centers are highly absorptive and the absorption generally occurs in a broad wavelength band. The earliest photochromic glass developed utilized a dispersed and sensitized colloidal silver halide as the darkening agent. Copper oxide was a typical sensitizer and AgCl is also a fairly common impurity used in PC glass. This material darkens by the equation

$$Ag^{I} + Cu^{I} \rightleftharpoons Ag^{0} + Cu^{II} \qquad [1]$$

This type of reaction is fairly typical of these materials. The defect state is created in the silver and copper, creating optical absorption. Photochromic glass can darken down to 22% transmittance in sunlight and become nearly clear indoors, transmitting 85% of visible light. Figure 1 shows the transmittance of the darkened and normal states. Temperature, however, does have an effect on the lenses. The glass will not darken completely at temperatures above 32 °C, and at cold temperatures the lens will get extremely dark in a shorter period of time. Recently, $BaO–R_2O_3–SiO_2$ (R = Al, B, La) nanoparticles have been introduced to enhance photochromicity.

Thermochromic Materials

Thermochromic materials darken and lose transmittance when heated above a critical temperature. They can be as sophisticated as a multilayer thin film optical interference coating or as mundane as a cereal bowl that changes color when heated.

Thermochromic products are used to darken windows of automobiles, for example, when the internal temperature gets above a specific value. The conductivity of these materials also decreases with darkening in many cases because additional electrons are excited into the conduction band due to the increased number of defect states in the bandgap. The most common inorganic thermochromic thin film materials are transition metal oxides, and among these, the vanadium oxide family (V_2O_5, VO_2) shows the best performance. Figure 2 shows the transmittance of a VO_2 thin film before and after heating. Here T_t is the transition temperature. When heated above T_t, the film darkens. Stoichiometric VO_2 darkens at a temperature near $T_t = 68\,°C$ due to a semiconductor to metal transition. The resistivity of VO_2 decreases by three orders of magnitude and the emittance can be increased up to 90% at the transition temperature. This material has promising applications for thermal control of space structures, where surfaces are exposed to intense solar radiation and periods of darkness at very low temperatures. The transition temperature can be engineered (increased or decreased) by doping with metals such as W and Ta.

Loss of transmission can be achieved by absorption or increased reflectance. Reflectance increases with heating are observed in substoichiometric WO_3 films and absorptance increases in stoichiometric films. The reflectance increase is attributed to a decrease in W/O ratio and an associated increase in the refractive index (n) and extinction coefficient (k).

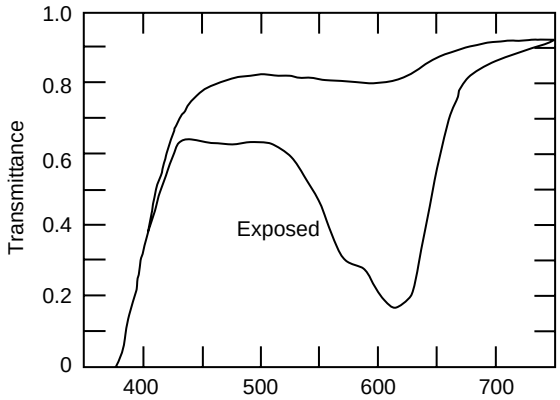

Figure 1 Transmittance of photochromic glass in darkened and normal states. (Reproduced with permission from Jorgenson GV and Lee JC (1985) Optical Materials for Energy Efficiency and Solar Energy Conversion IV, *SPIE Proceedings*, vol. 562. Bellingham, CA: SPIE.)

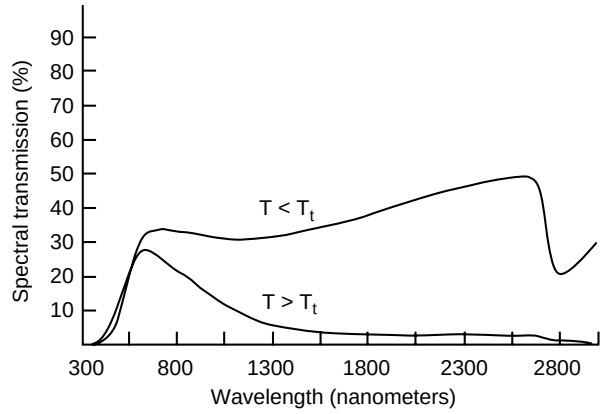

Figure 2 Change in transmittance of a thermochromic VO_2 film with heating. $T < T_t$ for normal transmission and $T > T_t$ when heated. (Reproduced with permission from Jorgenson GV and Lee JC (1985) Optical Materials for Energy Efficiency and Solar Energy Conversion IV, *SPIE Proceedings*, vol. 562. Bellingham, CA: SPIE.)

Electrochromic Materials

Electrochromic materials, also called switchable materials, are usually used in the form of thin film devices. Electrochromic windows, or smart windows, have finally reached maturity and promise to be the next major advance in energy-efficient window technology. These devices are used in architectural glazings, automotive mirrors, rear view mirrors, sunroofs, sunglasses, and other high-end applications. Electrochromic devices are well suited for both solar and thermal control. A window can be controllably darkened during periods of bright sunlight and switched back to transparency during the evening or during cloudy periods. Windows with multilayer thin film electrochromic coatings change color and darken with an applied voltage, and bleach with a reverse bias. The magnitude of the applied voltage determines the degree of coloring and bleaching. In addition to a change in visible transmission, the NIR transmission and reflectance, and the associated thermal properties, can also be controlled. The major advantages of electrochromic devices are: (1) power is used only during switching; (2) switching voltages are small (1–5 V); (3) reflectance and transmittance is always specular; (4) gray scale variations are possible; (5) polarization is minimal (reduces birefringence and distortion); and (6) the memory is adjustable in many cases.

The electrochromic effect occurs in inorganic compounds by dual injection (cathodic) or ejection (anodic) of ions (M) and electrons (e^-). A typical reaction for a cathodic coloring material is

$$WO_3(\text{colorless}) + yM^+ + ye^- \rightleftharpoons MyWO_3(\text{blue}) \quad [2]$$

where M is H^+, Li^+, Na^+, Ag^+, etc.

A typical anodic reaction is

$$Ir(OH)_x(\text{colorless}) \rightleftharpoons IrO_x(\text{black}) + XH^+ + xe^- \quad [3]$$

Movement to the right creates color or darkening, and movement to the left results in bleaching. Figure 3 shows the spectral data for the switching of a tungsten oxide film over the solar spectrum; the spectral response of the coating in the bleached and colored (or 'switched') states, and Figure 4 shows the bleached and darkened states of an actual electrochromic window.

Figure 5 shows the design of a simple electrochromic coating. The basic electrochromic coating consists of a transparent top electrode (usually indium tin oxide – ITO), which also injects electrons into the electrochromic layer beneath. By far the most studied and most promising electrochromic material is tungsten oxide (WO_3), which can have either an amorphous or crystalline microstructure. WO_3 has a high coloration efficiency and capacity. Other inorganic materials that are presently being developed for solar control applications are various forms of WO_3, NiO, $WMoO_3$, MoO_3, and IrO_x. Many of these are doped with a conduction ion such as Li^+ or H^+. The ion conductor layer is used to transport protons (H^+), which are supplied by the ion storage or counter-electrode layer, into the electrochromic material. The most promising ion conductors are certain immobile solvent polymer systems, ionic glasses, and open-channel metal oxide structures such as perovskites. Oxides of Ta and Nb also show promise. The bottom layer is another transparent conducting coating. It is possible to inject both electrons and protons into the electrochromic material, thereby inducing a strong absorption band in a given region, e.g., the visible spectrum, that results in a color change. In an

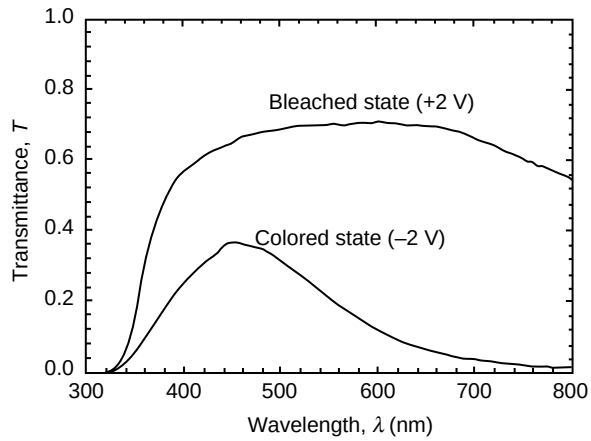

Figure 3 Spectral switching of an electrochromic window for darkened (−2V bias) and bleached (2V bias) states. (Reproduced from Kre J, Topic M, Smole F, *et al.* (2002) *Solar Energy Materials and Solar Cells* 71: 387–395, with permission from Elsevier.)

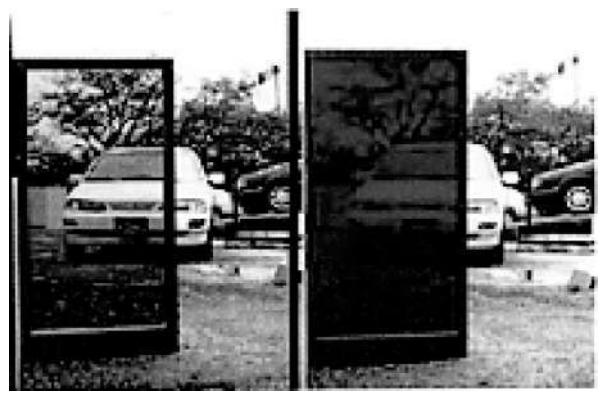

Figure 4 Bleached (left) and darkened (right) states of an electrochromic window.

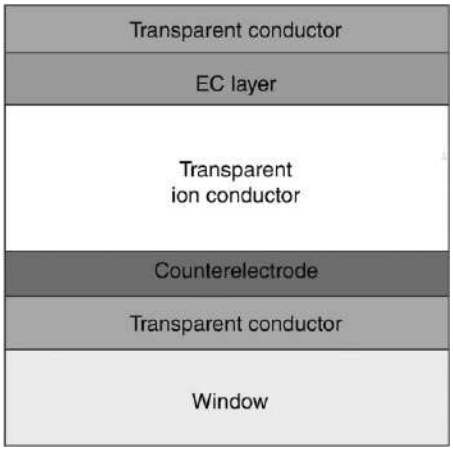

Figure 5 Basic layer design of electrochromic coating for use on a window.

Table 1 Electrochromic device structures

Electrochromic layer	*Ion conductor or electrolyte*	*Ion storage*
a-Li_xWO_3	$LiClO_4$ + PC	Redox Cple/NiO
a-WO_3	Ta_2O_5	$Ir_xSn_yO_2$:F
a-Li_xWO_3	$LiClO_4$ + PC	Prussian blue
a-H_xWO_3	Polymer	Polyaniline
a-H_xWO_3	SiO_2/metal	WO_3
Viologen	PMMA + organic	none
a-WO_3	Ta_2O_5	a-IrO_2
c-WO_3	Li–B–SiO Glass	IC/$Li_xV_2O_5$
a-WO_3	Li–PEO	CeO_x
a-Li_xWO_3	PPG–$LiClO_4$–MMA	$Li_yV_2O_5$
a-H_xWO_3	Ta_2O_5	NiO
a-WO_3	Li–PEO	NiO
NiO	a-PEO copolymer	Nb_2O_5, WO_3:Mo
a-WO_3	a-PEO copolymer	Polymer
Polyaniline	HCl	–

a, amorphous; c, crystalline; IC, ion conductor; MMA, methyl methacrylate; PC, polycarbonate; PEO, poly (ethylene oxide); PMMA, poly (methyl methacrylate); PPG, poly (propylene glycol).

electrochromic device, an externally applied electric field is used to control the injection process (voltage is applied across the two transparent electrodes). The coloration remains for some time, even after the external field is removed. The system returns to the initial state upon reversing the polarity of the external field. It is interesting to note that electrochromic windows being developed by NREL now employ solar cells to facilitate optical switching.

There are two major categories of electrochromic materials: transition metal oxides (including intercalated compounds) and organic compounds (including polymers). Table 1 shows a few of the most common electrochromic materials. Organic electrochromic materials are based on the viologens, anthraquinones, diphthalocyanines, and tetrathiafulvalenues. With organic compounds, coloration is achieved by an oxidation–reduction reaction, which may be coupled to a chemical reaction. The viologens are the most studied of the organic electrochromics. Originally, organic electrochromics tended to suffer from problems with secondary reactions during switching, but recently more stable organic systems have been developed. However, their reliability in harsh environments is questionable, and protective encapsulation may be required.

Electrochromic nickel oxide (NiO) is also being developed. NiO has been combined with manganese oxide, cobalt oxide, and niobium oxide in all alkaline devices.

Nonlinear Optical Materials

Nonlinear optical materials have evolved slowly over the past 20 years. At relatively low light intensities, the properties of most optical materials are independent of intensity of illumination. There is no interaction between the light waves in the medium. However, if the intensity of illumination gets high enough, nonlinear optical effects can occur and the optical properties become intensity dependent. When an electromagnetic field oscillating at optical frequencies ($\sim 10^{13} - \sim 10^{17}$ Hz) is applied to the lattice of a dielectric material, the electrical charges in the lattice become polarized and form electric dipoles. The positive ion cores do not move significantly, but the electrons are displaced in a harmonic motion. The polarization $\mathbf{P}$ is defined as $-Ne\mathbf{X}$, where N is the number of electric dipoles, $\mathbf{X}$ is the displacement of the electrons, and e is the charge on an electron. Here $\mathbf{X}$ is a complex number. Now, $\mathbf{P}$ is related to the electric field E by $\mathbf{P} = \varepsilon_0 \chi E$, where χ is the susceptibility and ε_0 is the permitivity of free space.

The linear dielectric constant is $\mathrm{Re}(1 + \mathbf{X})^{1/2}$, which is just the index of refraction. The imaginary part causes optical absorption.

When the intensity of illumination is high, the elastic limit of the electron's displacement is exceeded and the harmonic restoring force becomes nonlinear in $\mathbf{X}$, which adds quadratic and higher order terms to the expression for $\mathbf{X}$ and $\mathbf{P}$. In addition to the fundamental resonance, the dipoles now also oscillate at twice and thrice their fundamental resonant frequency (second and third harmonics). The polarization then is also nonlinear in displacement and can be expressed by the well-known relation

$$\mathrm{P} = \varepsilon_0(\chi^{(1)}\mathrm{E} + \chi^{(2)}\mathrm{E}^2 + \chi^{(3)}\mathrm{E}^3 + \ldots) \qquad [4]$$

where $\chi^{(2)}$ and $\chi^{(3)}$... are called the nonlinear susceptibilities. The second-order polarization $P^{(2)}$ causes optical mixing phenomena, involving sums and difference frequencies. When mixed with a DC field, the refractive index of the $\chi^{(2)}$ material can be controllably varied by varying the DC field. This is the well-known Pockels effect and is used in optical modulators. Figure 6 shows one method that is used to modulate the refractive index by the photorefractive effect. Here two laser beams (A_1 and A_2) are mixed in the $\chi^{(2)}$ material with the resulting intensity $I = I_o(1 + m\cos(kx))$. The material is polarized in phase with the mixed beams $\rho = \rho_o(1 + m\cos(kx))$, which results in a modulated electric field E_{sc}. The refractive index (Δn in the figure) is modulated in phase with the mixing signal. The $\chi^{(2)}$ materials are also used for frequency doubling.

Barium titanate ($BaTiO_3$), lithium niobate ($LiNbO_3$), KH_2PO_3 (KDP), lithium tantalite ($LiTaO_3$) and potassium niobate ($KNbO_3$) are $\chi^{(2)}$ optical switching and frequency doubling materials that have been studied for decades. Note that because they can be easily polarized, these are also excellent piezoelectric/ferroelectric smart materials.

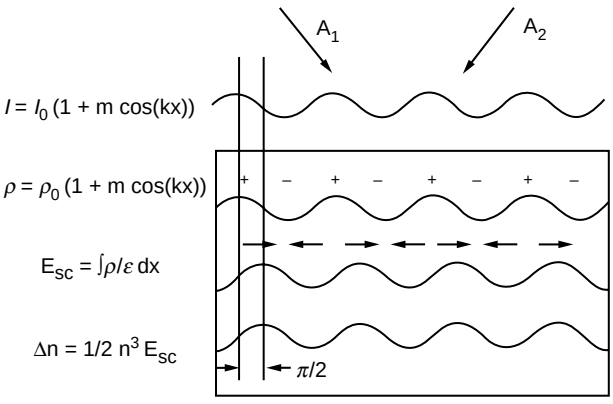

Figure 6 Modulation of the refractive index of a nonlinear optical material with incident laser modulation. From top to bottom: modulated laser intensity I, density ρ, electric field E_{sc}, refractive index Δn, and phase angle $\pi/2$. (Reproduced with permission from Bass M (ed.) (1995) *Handbook of Optics II*. New York: McGraw-Hill.)

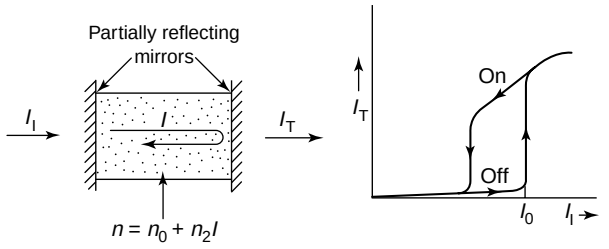

Figure 7 Optical switching behavior of a Fabry–Perot filter. I_I and I_T are the incident and transmitted intensities respectively. (Reproduced with permission from Butcher PN and Cotter D (1990) *The Elements of Nonlinear Optics*. Cambridge, UK: Cambridge University Press.)

The $\chi^{(3)}$ effects are sought after for applications such as laser eye protection, optical sensor protection, optical switching, optical limiting, and signal processing. In this case the refractive index (real and imaginary) depends on the intensity of the incident radiation. Figure 7 shows how this nonlinear effect is used in an optical switching device. It is all based on the change in the optical path length of a Fabry–Perot filter. At normal intensities, the refractive index of the spacer layer is n_0 and the filter is in the 'off' state. The refractive index becomes $n_0 + n_2 I$ and the optical thickness (nt) of the spacer layer changes at high illumination intensities, which in turn shifts the resonant wavelength (frequency) of the Fabry–Perot filter and creates the optical switch ('on' state).

There are a large number of $\chi^{(3)}$ and $\chi^{(2)}$ organic materials, including dyes, dimethylamino nitrostilbene, methyl nitroanaline, poly-BCMU, polydiacetylenes, and urea. Because they are essentially chains, many organic molecules can be easily polarized and possess higher order susceptibilities. The polarizability of organic materials is often enhanced by the mobility of delocalized π electrons in C–C bonds in aromatic rings. Hybrid nonlinear materials have even been formed by combining organic and inorganic components such as metallophthalocyanines which display very strong excited state absorptions. Most of these materials are deposited by spin casting.

Emissive Materials

Two of the newest applications of smart optical materials are emissive displays and electronic ink. This encompasses light-emitting polymers and small molecules, light-emitting diodes, and quantum cascade lasers. Organic light-emitting device (OLED) technology is leading the revolution in the flat panel display industry. This device was first invented by Tang and Van Slike and first reported in *Applied Physics Letters* in 1987. Figure 8a shows the layer structure of the OLED and Figure 8b shows a blue OLED pixel. The device is deposited onto an indium tin oxide (ITO)-coated glass or plastic substrate, with the ITO acting as the anode. The active layers can be either light-emitting polymers or small molecules. About 500 Å of a hole transport material (hole transport layer, HTL), here a tertiary amine, is deposited over the ITO. An aluminum trisquinalate (Alq), which transports electrons (electron transport layer, ETL) is deposited next. The device is topped off with

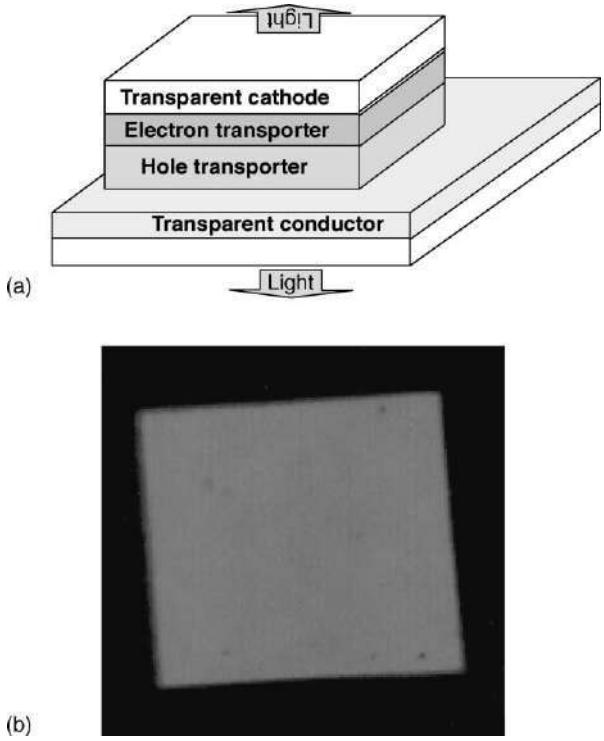

Figure 8 (a) Layer diagram of an OLED. (b) Blue OLED pixel.

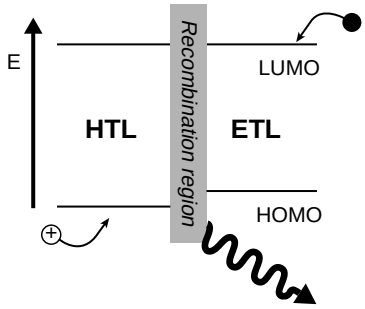

Figure 9 Simple energy level diagram of an OLED showing the relationship between the ETL and LUMO, and HTL and HOMO.

a low work function cathode, usually Ag:Mg or Ca. The entire device thickness is about 1000 Å. Even a dust particle can ruin the operation of an OLED by creating inactive regions or shorting out the electrodes.

The OLED emits light in a blue, green, or red wavelength band when activated by an electric current, i.e., electron and hole injection. The device functions by injecting holes from the ITO into the HTL. As shown in the simplified energy band diagram (Figure 9) for an OLED, the hole moves into the HOMO (highest occupied molecular orbital), which loosely corresponds to the valence band of a semiconductor. Electrons are injected from the cathode into the ETL, where it is transported into the LUMO (lowest unoccupied molecular orbital), which loosely corresponds to the conduction band of a semiconductor. When a voltage is applied between the anode and cathode, an electron flows and the holes and electrons combine in the ALQ layer and form excitons. Light is emitted when the excitons decay back a lower energy state. The color of the emitted light (green, red, or blue) can be changed by changing the HTL and ETL material composition. Light is emitted by fluorescence processes, from both singlet and triplet states if the material chemistry is correct, which results in 100% internal quantum efficiency. Twenty percent external quantum efficiency results from using all available exciton states. Photons can be lost through absorption in the layers, scattering, reflection between layers, and waveguide modes. These factors must be minimized to maximize the external quantum efficiency.

OLED video displays now have outputs near 100 cd/m (20 lm/W), which is about twice the normal video brightness. Red, green, and blue OLEDs now emit with very high efficiencies. For comparison, a 100 W light bulb emits at 17 lm/W and the best fluorescent light emits at 80 lm/W. Today, OLEDs are already sitting at intensities of 50–60 lm/W, and could easily push 150 lm/W in the near future. Another advantage of OLED displays is their wide viewing angle, up to 160°.

There are several types of OLED display configurations. Figure 10 shows a passive matrix display geometry and placement of the OLED layers. Here the OLED is sandwiched between crossed cathode and anode contacts. By injecting current between sets of crossed anode and cathode rows, the OLED will light up, each contact point forming a pixel. Displays can typically have as many as 3 million pixels. OLED displays seem brighter and livelier to the eye because the colors are additive, rather than subtractive as with LCD displays. For a full-color passive matrix display green, red, and blue pixels must be located very close together. Sony has developed an active matrix 13-inch diagonal display with 800 × 600 pixels (~150 000 functioning OLEDs).

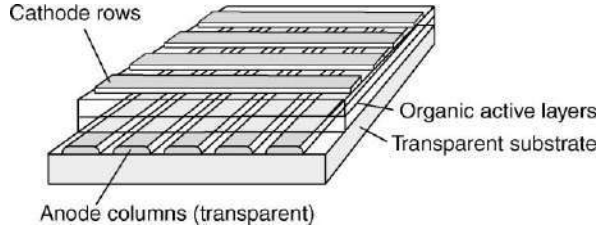

Figure 10 Geometry of a passive matrix display.

Among the more promising pixel geometries is the SOLED, the stacked OLED, where, red, green, and blue pixels are stacked on top of each other. It is mandatory that the pixels and the electrical contacts be transparent for this design, or light will not be transmitted through each layer. Very thin Mg contacts are used with a top sputtered ITO cathode. The emitted color is simply changed by applying current across the pixels, with white being obtained when all pixels are activated.

The future of OLED displays is plastics, or the flexible OLED (FOLED). Plastic substrates are lightweight, low cost, and rugged. However, they have an inherent problem of being very permeable to oxygen and water vapor, both of which severely degrade performance of the OLED by reacting with the anode layer. For displays to be practical, they must have lifetimes in excess of 10 000 hr and preferably 20 000 hr. Barrier coatings are now being developed for plastics to prevent oxygen and water ingress and extend the life of the OLED to this goal.

Electronic ink is a display medium that can be formatted using a passive matrix. It consists of millions of tiny negatively charged microcapsules, about 100 microns in diameter, floating inside a liquid medium. Each microcapsule is filled with hundreds of pigmented chips. One half of each microcapsule is white and the other half is dyed blue. When an electric charge is applied to the microcapsules, the chips are drawn either to the bottom or top of the microsphere. The viewer sees dark areas when they are drawn to the top. In another version being developed by Xerox Corp., half the chip is white and half the chip is dark. When an electric charge is placed on the chip, the dark side is oriented up and the white side down, creating dark regions where charge is placed. This can be easily erased and reformatted into text, symbols, signs, and pictures.

Future smart optical materials will simulate biological systems such as an insect's eye and will employ quantum structures to achieve optical switching, emission, and control of optical properties.

List of Units and Nomenclature

Absorptance (A)	The amount of light absorbed by an optical element divided by the incident intensity. $A = 1 - T - R$.
Active matrix display	A display geometry in which the voltage to the pixel is controlled individually by a transistor.
Candela (cd)	Formerly known as a candle. The luminous intensity equal to 1/60 of the luminous intensity per square meter of a black body radiator operating at the temperature of freezing platinum.
Exciton	An electron–hole pair, usually generated by a photon.
FOLED	Organic light emitting device on a flexible plastic substrate.
LCD	Liquid crystal diode.
Lumen (lm)	A unit of luminous flux equal to the luminous flux emitted in a solid angle of one steradian for a point source having an intensity of one candela.
NIR	The near-infrared spectral region, consisting of wavelengths between 700 and 2500 nm.
Passive matrix display	A display geometry consisting of a grid of horizontal and vertical wires. At the intersection of each grid is an active element (LCD or OLED) which constitutes a single pixel, either letting light through or blocking it.
Piezoelectric	Capable of generating a voltage when a force or stress is applied.
Polarization (coulomb/m^3)	A vector quantity equal to the electric dipole moment per unit volume.
Quantum efficiency	The number of electrons, holes, or electron–hole pairs generated per incident photon at a given optical wavelength.
Reflectance (R)	The intensity of light reflected from an optical element divided by the incident intensity. $R = 1 - T - A$.
Susceptibility (no units)	A scalar quantity equal to the ratio of the electric polarization of a material to the electric field strength.
Transmission (T)	The intensity of light transmitted through an optical element divided by the incident intensity. Note that transmission + reflection + absorption = 1 ($T + R + A = 1$).
UV	The ultraviolet spectral region consisting of wavelengths between 200 and 400 nm.
Visible	The spectral region consisting of wavelengths between 400 and 700 nm.
Wavelength (μm or nm)	The length of one oscillation of a light wave.

See also

Materials for Nonlinear Optics: Organic Nonlinear Materials. **Optical Coatings:** Thin-Film Optical Coatings. **Optical Materials:** Optical Glasses.

Further Reading

Butcher PN and Cotter D (1990) *The Elements of Nonlinear Optics*. Cambridge, UK: Cambridge University Press.

Chemla DS and Zyss J (1987) *Nonlinear Optical Properties of Organic Molecules and Crystals*, vol. 1. Orlando, FL: Academic Press.

Chemla DS and Zyss J (1987) *Nonlinear Optical Properties of Organic Molecules and Crystals*, vol. 2. Orlando, FL: Academic Press.

Clark DT (1994) Second European Conference on Smart Structures and Materials, *SPIE Proceedings*, vol. 2361. Bellingham, CA: SPIE.

Crowson A (1996) Smart Materials Technology and Biomimetics, *SPIE Proceedings*, vol. 2716. Bellingham, CA: SPIE.

Gobin P-F and Friend CM (2000) Smart Structures and Materials, *SPIE Proceedings*, vol. 4073. Bellingham, CA: SPIE.

Janocha H (1999) *Adaptronics and Smart Structures, Basics, Materials, Designs, and Applications*. Berlin: Springer.

Jorgenson GV and Lee JC (1985) Optical Materials for Energy Efficiency and Solar Energy Conversion IV, *SPIE Proceedings*, vol. 562. Bellingham, CA: SPIE.

Goff AH-L, Granqvist C-G and Lampert CM (1992) Optical Materials Technology for Energy Efficiency and Solar Energy Conversion XI: Chromogenics for Smart Windows, *SPIE Proceedings*, vol. 1728. Bellingham, CA: SPIE.

McDonuch A, Gardiner APT and McEwen KS (1994) In: Culshaw B (ed.) Second European Conference on Smart Structures and Materials, *SPIE Proceedings*, vol. 2361. Bellingham, CA: SPIE.

Zerbi G (ed.) (1993) *Organic Materials for Photonics*, European Materials Research Society Monographs, V6. Amsterdam: North-Holland.

Sol-Gel Materials

Lisa C Klein, Rutgers the State University of New Jersey, Piscataway, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Sol-gel optics represents a methodology for fabricating optical materials in fiber, thin film, and bulk geometries. Defined broadly, sol-gel optics encompasses a variety of optical devices that contain sol-gel processed components. The optical device may be based wholly on sol-gel processing or incorporate a sol-gel material, along with conventional optical materials. Principally, the optical devices prepared by sol-gel processing or assembled with sol-gel materials are traditional optics or, in some cases, opto-electronics. Among the current sol-gel optical materials, by far the most common are oxides. Further, the majority of sol-gel optics is for transparent components, which typically means glasses, and, consequently, silicates.

In fact, most sol-gel materials have been developed to duplicate conventional silica optical materials. While glass melting technology is highly advanced for making quality lenses, an equivalent material can be made using a sol-gel route. Similarly, ultralow-loss optical fiber can be made by chemical vapor deposition (CVD). Now sol-gel processing can be used to prepare comparable fibers for telecommunications. In cases where sol-gel processing produces an exact duplicate of optical materials produced other ways, the perceived advantage is that sol-gel precursors have a higher purity and better level of mixing than powder raw materials. Perhaps, a more significant advantage is that sol-gel formulations can be designed to promote a particular geometry in the final material, essentially making near-net shape materials.

While sol-gel processing is not designed to produce new optical phenomena, it does offer new ways to assemble materials to produce new optical designs, such as filters, Bragg gratings, amplifiers, and beam splitters. One noted case of sol-gel materials in nonlinear optics is the case where semiconductor particles are trapped in nanometer-sized pores. This leads to so-called quantum dots. The microporosity of the gel, no doubt, assists in this effect, but it is not the gel itself that has produced the optical effect. The optical effect is derived from the nanometer size of the particles. The gel serves as a transparent host that restricts the size of the particles and provides for their uniform distribution.

Sol-Gel Chemistry of Oxide Optical Materials

The sol-gel process for oxides is the name given to any one of a number of processes involving a solution or

sol that undergoes a sol-gel transition. Solution refers to a single-phase liquid, while a sol is a kinetically stabilized suspension of colloidal particles. At the transition, the solution or sol becomes a rigid, porous mass by destabilization, precipitation, or supersaturation. Once the solution or sol goes through a sol-gel transition, it is a rigid two-phase system of solid and solvent filled pores. The sol-gel transition is not reversible.

In using a sol-gel process, it is important to be familiar with the reagents, the intermediates and the products. The most important reagent in this case is a hydrolyzable organometallic which is called a metal alkoxide, $M(OR)_x$ where (OR) is an alkoxy group. Alkoxides are the precursors for SiO_2, GeO_2 or B_2O_3, among others. Hydrolysis occurs when the metal alkoxide and water are mixed in a mutual solvent.

The first step is choosing the right reagents. Since most of the glasses are silicates, we start with the silica precursor. Of the available silicon alkoxides, tetraethylorthosilicate (TEOS) is used most often, because it reacts slowly with water and comes to equilibrium as a complex silanol. The colorless liquid $Si(OC_2H_5)_4$ has a density of about 0.9 g/cm^3, is easy to handle safely and is extremely pure when distilled. Most metal alkoxides are prepared by the reaction of halogenated species or pure metals with alcohol.

The two other ingredients in a typical formulation are alcohol and distilled water. In the case of TEOS, ethanol serves as the mutual solvent for TEOS and water. Once TEOS is dissolved in ethanol to make it soluble in water, the chemical reactions hydrolyzation and polymerization begin. The chemical reactions are approximately:

Hydrolyzation

$$-\overset{|}{\underset{|}{Si}}-O-C_2H_5 + H_2O \rightarrow -\overset{|}{\underset{|}{Si}}-OH + C_2H_5OH \qquad [1]$$

Polymerization

$$-\overset{|}{\underset{|}{Si}}-O-C_2H_5 + Si-OH \rightarrow -\overset{|}{\underset{|}{Si}}-O-\overset{|}{\underset{|}{Si}} + C_2H_5OH \qquad [2]$$

An acid is usually added to control the rates of these reactions. The temperature is often ambient, and mixing is continued for between 1 and 3 hours. Complete hydrolysis does not occur. Instead, condensation occurs between silanol and ethoxy groups to give bridging oxygens or siloxane groups. The intermediates that remain soluble in the alcohol–water medium are silanols, ethoxysilanols, and polysiloxanes.

The equipment needed for laboratory-scale production consists of a narrow-necked glass flask with magnetic stir bar on a stirring hot plate. The reactants are added volumetrically to the flask in the order: solvent, TEOS, water, and acid. When everything is dissolved and the solution is clear, the hot plate may be turned on to accelerate reactions or a reflux condenser may be attached.

The parameters that influence the chemical reactions are temperature, pH, amount of water, solvent, and precursor. More complex compositions are easier to mix in longer chain alcohols, and longer chain organic groups give slower reaction rates. Low pH favors a polymerization scheme that gives linear molecules and the solution remains clear. High pH favors cluster growth and the solution may become cloudy. At high pH or high water, solutions produce cross-linked polymers or branched clusters. A small volume fraction of branched clusters restricts flow, when the same volume fraction of linear polymers does not. At a higher volume fraction, the linear polymers become tangled, and then low water solutions will gel. The result of these reactions under all conditions is an increase in the molecular weight of the oxide polymer. Eventually, the solution reacts to a point where the molecular structure is no longer reversible, and the sol-gel transition has been reached.

Shapes, Compositions and Precursor Concentrations

A sampling of compositions and precursors is given in Table 1. Some compositions such as SiO_2 and TiO_2–SiO_2 have been prepared in bulk, to duplicate materials made by melting for lenses and mirror blanks. Other compositions such as B_2O_3–SiO_2 have been prepared for thin-film applications. The distinguishing aspect between the solutions formulated for bulk objects and those formulated for thin films is the oxide concentration in the solution. An indication of the solution concentration is the molar ratio of water to alkoxide. The higher ratios are used when bulk objects are desired.

Table 2 gives a sampling of the geometries for sol-gel silica. For each geometry, an example of a shape is given, with an application for each shape and the approximate dimensions of the sol-gel material.

In terms of near-net shape processing, when the material goes through its sol-gel transition, it proceeds to dry and shrink. This phase is characterized either by isotropic shrinkage or anisotropic shrinkage. Monoliths fall into the first category with isotropic shrinkage following molding. Thin films on substrates fall into the second category with anisotropic shrinkage following dipping. Fibers also show

Table 1 Typical compositions of sol-gel optical materials

Oxides	*Precursors*	*Compositions in mol %*	*R**
Single oxide			
SiO_2	$Si(OC_2H_5)_4$	100 SiO_2	16
Two oxides			
$B_2O_3-SiO_2$	$Si(OC_2H_5)_4$ $B(OCH_3)_3$	80 SiO_2–20 B_2O_3	1.5
GeO_2-SiO_2	$Si(OCH_3)_4$ $Ge(OC_2H_5)_4$	90 SiO_2–10 GeO_2	4
TiO_2-SiO_2	$Si(OC_2H_5)_4$ $Ti(OC_3H_7)_4$	94 SiO_2–6 TiO_2	50
$P_2O_5-SiO_2$	$Si(OC_2H_5)_4$ H_3PO_4	90 SiO_2–10 P_2O_5	16
Three oxides			
$Li_2O-Al_2O_3-SiO_2$	$Si(OC_2H_5)_4$ $Al(OC_4H_9)_3$ $LiNO_3$	82 SiO_2–3 Al_2O_3–15 Li_2O	8
$Na_2O-B_2O_3-SiO_2$	$Si(OC_2H_5)_4$ $B(OCH_3)_3$ $NaNO_3$	82 SiO_2–12.4 B_2O_3–5.6 Na_2O	10

*Approximate ratio of water to alkoxide, e.g., $R = 16$ means 16 moles of H_2O per mole of $Si(OC_2H_5)_4$.

Table 2 Geometries and approximate dimensions for selected applications of sol-gel silica

Geometry	*Shape*	*Dimensions*	*Application*
Monolith	Flat disk	10 cm diameter 1 cm thick	Lens
	Cylinder	0.5 cm diameter 2 cm long	Rigid host for chromophores
Thin film	Film on Si wafer	500 nm thick	Antireflection
	Film on Hollow tube	~10 000 nm thick	Solar collector panel
Fiber	Rod	0.8 cm diameter 25 cm long	Preform for drawing
	Filament	150 micron diameter/100 m length	Lightguide

anisotropic shrinkage, though it depends on whether or not the fiber is drawn from a preform many times its diameter or produced from a filament that is closer to its diameter.

Sol-Gel Monolith Processing

Monoliths formed from an alkoxide solution are achieved only in dilute solutions. Therefore, alkoxide gels are somewhat difficult to dry because of their small pores (<10 nm). To avoid cracking in alkoxide gels, different treatments have been tried. Aerogels are dried in an autoclave by hypercritical techniques. That is, the solvent is removed above its critical point. The resulting gel is about 10% dense and shows no shrinkage. Xerogels are dried by natural evaporation. Xerogels are 60% dense and have reductions 40–60% in volume.

Alternatively, monoliths can be formed from colloidal solutions. These particles coalesce at gelling. Colloidal gels are easily dried, and they are less likely to crack than alkoxide xerogels. While alkoxide gels have small pores, colloidal gels have larger pores or voids between particles. For the same reason, colloidal gels densify at higher temperatures than alkoxide xerogels. The higher temperature for colloidal gels may lead to undesirable crystallization.

By either route, a serious problem for sol-gel processing is the presence of water. Especially when monoliths are used for fiber drawing, residual OH ions have a deleterious effect on optical properties. That is why non-aqueous solvents, as well as fluorination or chlorination treatments, have been investigated to reduce the OH content.

Sol-Gel Fiber Processing

While monolithic pieces need high water content, fibers can be drawn out of low water content

solutions. The sol-gel process allows one to bait and draw a string of gel about the same diameter as the desired fiber directly from the solution. This has been shown where the water to alkoxide ratio was controlled to give the property of spinnability. The time and manner in which the system loses fluidity depend on composition and catalyst. Low water solutions show a gradual increase in viscosity before losing fluidity, while high water solutions lose fluidity more abruptly from a lower viscosity. A problem with this method is the shape of the fiber, which tends not to be cylindrical.

Another approach to making near-net shape fibers from a sol-gel process involves coating a shell or concentric shells of sol-gel glass on a filament of textile fiber. This approach is referred to as the volatile host method, where a plastic filament is used as a template. Multiple dipping or adjustment of the viscosity can lead to different thicknesses. Upon exposure to atmospheric moisture, the films gel. After gelling, a rigid, continuous shell runs the length of the filament.

Once the shell has hardened, solvent and water can escape through interconnected pores. The shell shrinks in the radial dimension, but remains continuous along its axis. During heat treatment and desiccation, little if any densification occurs at temperatures well below the glass transition. The interconnected pore structure remains open until the volatile host is eliminated and undesirable gases are removed. This method is capable of producing optical fibers, but is more suitable for hollow fibers or ion exchange media.

Sol-Gel Thin Film Processing

Like sol-gel fibers, sol-gel thin films are prepared from low water content solutions. The process is exceedingly simple. A solution containing the desired oxide precursors is applied to a substrate by dipping or spinning.

At this time, the majority of sol-gel coatings are applied by dipping. This is an approach that allows the properties of the solution to control the deposition. A substrate is lowered into a vessel containing the solution. A meniscus develops at the contact of the liquid and the substrate. As the substrate is withdrawn, the meniscus generates a continuous film on the substrate. The process is able to apply a coating to the inside and outside of complex shapes simultaneously.

The equipment is inexpensive, especially in comparison to any deposition techniques that involve vacuum. Coatings can be applied to metals, plastics, and ceramics. Typically, the coatings are applied at room temperature, although most need to be calcined and densified with heating. Dip films are typically one micron, uniform over large areas and adherent. The time-to-gel is especially important when it comes to coatings because film formation, drying, and creation of pores must be rapid. Optimum film formulations correspond to those solutions that lose tackiness quickly.

In most cases, the effect of the viscosity is far greater than the surface tension. Two simple relationships are that the film thickness increases with increasing withdrawal rate and, for a given withdrawal speed, the film thickness increases with an increase in oxide content. The process is frequently a batch process, certainly on the laboratory scale. However, the batch process can be scaled up. Repeated dipping builds up a thicker film.

For one-side coating, a technique that is often used is spin coating. In this case, the substrate, usually a silicon wafer, is placed on a spinner, rotated at perhaps 200 rpm while solution is dripped on the center of the substrate. As with dip coatings, a film between 50 and 500 nm results.

After the Sol-Gel Transition

Having selected a shape and designed the formulations accordingly, there are several further steps common to monoliths, fibers, and films. First of all, the gels must be dried. Using the example of silica, TEOS is less than 30% by weight silica and the solution is even less. From the time the solution is applied to the time it gels, there is about 50% weight loss. Then, as the gel dries there is another reduction in weight by one-half. To go along with the weight loss there is about 70% volume reduction.

Films can be dried quickly in air because of the one thin dimension. It has been shown repeatedly that all shrinkage is taken up in the thin dimension and not in the plane of the substrate. Yet, the films remain adherent and continuous and maintain complete surface coverage. To go from a tacky film to a hard film may take only a few minutes.

Following drying and shrinkage comes the heat treatment. It would be inaccurate to say that glass has been formed at room temperature by simply reacting, gelling, and drying the solutions. In this state, the materials have many of the characteristics of the corresponding glass, but they are more or less porous. Water and solvent escape through interconnected pores that remain open at the surface until the gels are fired to temperatures well above 600°C. When needed to fulfill the requirements of an application, the hard gel is heated to various degrees of collapse.

The microporosity in silica xerogels is not removed entirely until 1000°C.

Applications of Sol-Gel Optical Coatings

Sol-gel processing has a special relevance to optical coatings. In most cases, the applications are substitutions of a sol-gel coating for a coating obtainable by other means. Most optical coatings achieve their effect by selective absorption and interference. By depositing 1–14 or more layers of dielectric thin films, using two or more different compositions, a variety of antireflection (AR), half-mirror, hot-mirror and cold-mirror coatings can be designed. An early design was for rear-view mirrors for cars, where the coatings consisted of titania-silica-titania interference filters that gave the effect of total reflectance and replaced conventional metallizing. Another early application was solar reflecting films for windows. These coatings consisted of Pd-containing titania films which show selective absorption.

A sampling of optical coatings is given in Table 3. The application, the mechanism, and the oxide composition are listed. Generally, these coatings can be prepared by other deposition techniques. Those applications which are specific to sol-gel coatings relate to the microporous nature of the sol-gel processed films. In Table 3, there are antireflective coatings of several types. Both borosilicates and titania silicates have been developed for this application. Changes in the composition, changes in the microstructure, and changes in the porosity can be used to grade the index of refraction. Antireflection in the ultraviolet range has been accomplished with thoria and hafnia films. Broadband antireflection has been used for laser optics. The sol-gel coating technique provides control over composition and deposition thickness. The layers, being quarter wavelength in thickness, require precise control.

Table 3 Thin-film applications

Optical application	*Mechanism*	*Typical composition*
Color filters	Selective Absorption	Cr_2O_3–SiO_2
Mirror	Interference	14-layer TiO_2: SiO_2
Antiglare	Absorption	CoO–SiO_2
Antireflection		
Narrow	Graded porosity	SiO_2
Broadband	Interference	Three-layer
	1/4 wavelength	SiO_2–TiO_2:ZrO_2: SiO_2
	Graded porosity	SiO_2
IR reflector	Interference	Three-layer TiO_2:Ag:TiO_2

Sol-Gel Rigid Hosts

An exciting way of looking at sol-gel materials is to use the gel as a host. The important point to note is that the sol-gel materials are transparent. While most of the materials are porous, the porosity is on a scale that does not scatter visible light. Therefore, most light is transmitted. The appeal of the bulk shape host is that it can be molded to near-net shape and can be cut and polished.

Hybrids

Several approaches have been taken to prepare so-called hybrid gels incorporating organic and inorganic components. Those methods that mix the components on the nanometer scale include the infiltration of previously formed oxide gels with monomers, reacting of alkoxide and organic monomers, or organic polymers with alkoxides in mutual solvents. The products of these processes are variously called ormosils (organically modified silicates), ceramers (ceramic polymers), interpenetrating networks (IPNs) and nanocomposites.

The original route to hybrids was one where a gel was formed first as a rigid matrix. Monomer was infiltrated into the open porosity of the shaped gel and polymerized inside the gel using UV curing, for example. Since silica gel is the easiest gel to prepare in monolithic form, one combination that has been investigated widely is silica gel/polymethyl methacrylate (PMMA).

Silica and PMMA, which both have good visible light transmission and similar index of refraction, have been used in hybrids, although it turns out that the match in index of refraction is less critical on the nanoscale. They are also glassy materials with similar strength. PMMA has much higher fracture energy than silica and is often used as glazing in applications where impact resistance is a priority. However, PMMA has a much lower hardness than silica and is less resistant to abrasion. Ideally, silica–PMMA hybrid nanocomposites have the strength and impact resistance of PMMA, and the hardness of silica in one. One obvious application is transparent, lightweight windows.

Encapsulation

Sol-gel hosts have been used to encapsulate a variety of organic and inorganic molecules, chromophores, and biological agents. Whatever is incorporated into the gel can act in two ways, to

give information about the gel host or to be influenced by the gel in its interaction with light. For example, spiropyrans have been incorporate for photochromics, CdS for nonlinear optics, Nd for tunable lasers, and various enzymes for sensors. Early on, silica gel hosts were used to encapsulate the electro-optically active organic compounds 2-methyl-4-nitroaniline (MNA). In comparison to conventional optical cells, MNA doped silica hosts retained optical quality at higher temperature and were more abrasion resistant. Since this discovery, silica hosts, in particular, and sol-gel processing, in general, has been used extensively for encapsulating molecules and species that cannot be heated, but need to be protected in a rigid matrix.

Future Directions

As more and more is learned about the sol-gel process, the greater the number of possibilities for useful optical devices. In the selected examples given above, a key feature is that the sol-gel material is transparent. In addition, the sol-gel material is suitable for application to a variety of substrates. Also, the process is a low-temperature process that is compatible with polymers and organic chromophores. In summary, sol-gel processing has had a number of successes in its application to optical materials, and its use is likely to expand.

See also

Optical Coatings: Thin-Film Optical Coatings.

Further Reading

Andrews MP and Najafi SI (eds) (1997) *Sol-Gel and Polymer Photonic Devices*. Bellingham, CA: SPIE Critical Reviews, CR68-03.

Brinker CJ and Scherer GW (1990) *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*. Boston, CA: Academic Press.

Dunn B and Zink JI (1991) Optical properties of sol-gel glasses doped with organic molecules. *Journal of Material Chemistry* 1: 903–913.

Klein LC (1993) Sol-gel optical materials. *Annual Reviews of Material Science* 23: 437–452.

Klein LC (ed.) (1994) *Sol-Gel Optics: Processing and Applications*. Dordrecht: Kluwer Academic.

Najafi SI (ed.) (1998) *Selected Papers on Sol-Gel for Photonics*. Bellingham, CA: SPIE Milestones Series, MS148.

Riesfeld R and Jorgenson CK (1992) Optical properties of colorants or luminescent species in sol-gel glasses. *Structure and Bonding* 77: 208–256.

Sol-Gel Optics. SPIE vols. 1328, 1758, 2288 and 3943, 1990–2000. Bellingham, CA: SPIE.

OPTICAL MICROLENSES

H Ottevaere and H Thienpont, Vrije Universiteit Brussel, Brussel, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In today's world of information processing the role of optics and opto-electronics is expected to become increasingly important as the performance of communication, processing, sensing, and display technologies is continuously evolving. Making these photonics technologies faster and smaller requires, at the same time, the introduction of massive parallelism and miniaturization. As a result, high-quality, high-precision, and low-cost microlens arrays are becoming indispensable components. Moreover, the introduction of micro-optics provides new degrees of freedom for the system design and, therefore, also for the functionality. It allows solutions which are not possible with purely conventional optics such as the integration of different micro-optical components and systems, hence reducing the additional mechanical assembly cost as required in conventional optics.

Since the beginning of the 1990s, several research groups and industrial research labs have been focusing their attention on the development of fabrication techniques for refractive and diffractive microlenses. Microlenses and microlens arrays are used for beam shaping purposes such as collimation and focusing (e.g., in combination with laserdiode arrays, detector arrays, and fibers), for illumination (e.g., in display systems and projection systems), and for imaging purposes (e.g., in photocopiers).

In this article we will mainly focus on refractive microlenses and we intend to provide an introductory guide to those basic geometrical and optical parameters that are used to characterize refractive microlenses. Indeed, some elementary principles of geometrical optics (the geometrical laws for optical propagation without the inclusion of diffraction) and wave optics (in wave optics the imaging is studied as a result of the propagation of a wavefront to the image

plane; diffraction effects can be explained on the basis of this wave theory) are required to understand the characterization and performance evaluation of these lenses. We, therefore, bring forward the basic nomenclature of refractive micro-optics and will highlight how the performance of an optical lens is affected when reducing its physical dimensions.

We start this article by introducing the imaging properties of both thick and thin lenses in the paraxial approximation. In the section on lens aberrations, we discuss the different primary aberrations that influence the imaging properties of a lens. Combined effects of aberrations and diffraction give rise to some frequently used lens performance criteria which we discuss in the section on diffraction. Different aspects have to be considered when discussing the quality of a lens, because each type of microlens application requires specific performances. It is, therefore, not possible to come up with a single figure of merit which satisfactorily evaluates the overall quality of a microlens. Moreover, one needs specific information about the characteristics of the light sources and detectors that are used in the system and also details about the imaging system geometry and the required imaging contrast. We review them in the section on quality criteria.

In the section on interference, we describe these phenomena and focus on two-beam interference that forms the basis of the photonic instrumentation that can be used to characterize discrete microlenses and microlens arrays. For a complete evaluation of a microlens it is essential to measure both its geometrical properties (diameter, sag, surface profile, etc.) and its optical performance (focal length, wave aberrations, etc.). In the final section of this article we introduce three interference-based instruments which are crucial to the full quantitative characterization of refractive microlenses: a noncontact optical profiler based on a Mirau interferometer, a transmission Mach–Zehnder interferometer, and a Twyman–Green interferometer.

Geometrical Optics

Introduction

Within the paradigm of geometrical optics or ray-optics, light travels from the source along straight lines or rays. When a light ray traverses an optical system consisting of several homogeneous media in sequence, the optical path is a chain of straight-line segments. Discontinuities in the line segments occur each time the light is refracted or reflected. The laws of geometrical optics that describe the direction of the rays are described below.

The law of refraction, also called Snell's law
The refraction of a ray at a 'well-polished' surface between two isotropic media, with indices of refraction n_1 and n_2, is given by

$$n_1 \sin \varphi_1 = n_2 \sin \varphi_2 \qquad [1]$$

where φ_1 and φ_2 are the angles between, respectively, the incident and the refracted ray and the normal to the surface.

The law of reflection
When a ray is reflected on a flat 'polished' interface dividing two uniform media, the reflected ray remains within the plane of incidence and the angle of reflection equals the angle of incidence.

Geometrical wavefronts can then be constructed from the optical rays propagating through the system. Locally they are orthogonal to the rays and the optical path lengths along all rays from one wavefront to another are equal. In regions where the wavelength is considered to be negligible, as compared to the dimensions of the relevant components of the optical system, geometrical wavefronts are close approximations of the physical wavefronts.

The Paraxial Approximation

The basic optics of image location and magnification is demonstrated by the application of paraxial or first-order optics, also called Gaussian optics. In the paraxial approximation, only those points and rays will be considered which lie in the immediate vicinity of the optical axis, so that all angles φ are small ($\varphi \leq 15°$). This assumption is true in lens systems with small object and image field sizes and small apertures of pupils. The terms involving squares and higher powers of the off-axis distances or of the angles, which the rays make with the optical axis, will not be discussed here. This leads to:

$$\sin \varphi = \tan \varphi = \varphi + O_{\geq 2}(\varphi) \approx \varphi \qquad [2]$$

With this assumption we can write the refraction law by the linear relation:

$$n_1 \varphi_1 = n_2 \varphi_2 \qquad [3]$$

Refraction at a spherical surface
In Figure 1 a spherical surface with a radius of curvature R forms an interface between two materials with indices of refraction n_1 and n_2. Two rays are shown emanating from an axial object point S. One is an axial ray, normal to the surface of the lens at its vertex V. The vertex of a lens is defined as the point on the refracting surface at the center of its free aperture. This assumes that the aperture is circular

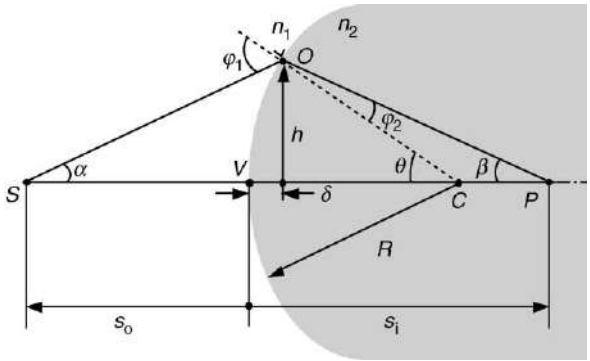

Figure 1 Refraction at a spherical surface.

and centered and the surface is spherical. More generally, if the surface is not spherical but has rotational symmetry, the vertex is the point where the axis of symmetry intersects the optical surface. It is refracted without any change in direction. The other ray is an arbitrary ray incident at O and refracting there according to Snell's Law.

We can rewrite the paraxial refraction law (eqn [3]) and derive a relation between the object and image distances, s_o and s_i, respectively.

From the triangles SOC and POC, we have:

$$\varphi_1 = \alpha + |\theta| \qquad |\theta| = |\beta| + \varphi_2 \qquad [4]$$

where $|\theta|$ and $|\beta|$ represent the absolute values of θ and β.

We also have:

$$\tan\alpha = \frac{h}{|s_o| + \delta} \qquad \tan|\beta| = \frac{h}{s_i - \delta}$$
$$\tan|\theta| = \frac{h}{R - \delta} \qquad [5]$$

As discussed above, for paraxial rays we may approximate the sine and tangent of an angle by the angle itself and one can show that the small distance δ may be ignored. In accordance with the sign convention the object distance, being to the left of the refracting surface, is to be taken as negative. Inserting eqns [4] and [5] into [3] results in:

$$\frac{n_2}{s_i} - \frac{n_1}{s_o} = \frac{n_2 - n_1}{R} \qquad [6]$$

This equation does not contain the angle α, so the image distance s_i is the same for all paraxial rays starting from S.

The spherical surface power $\wp$ is now defined as:

$$\wp \equiv \frac{n_2 - n_1}{R} \qquad [7]$$

We can relate this power of the surface $\wp$ to the object and image distance, by combining eqns [6] and [7]:

$$\frac{n_2}{s_i} - \frac{n_1}{s_o} = \wp \qquad [8]$$

The distance from the object point to the vertex of the refractive surface is called the front focal length f_{FFL}, if the image is at infinity ($s_i = \infty$). Equations [6] and [8] then give:

$$f_{FFL} = -\frac{n_1}{n_2 - n_1}R = -\frac{n_1}{\wp} \qquad [9]$$

Similarly, if the image is placed at the back focal length f_{BFL}, such that the object distance equals infinity ($s_o = \infty$), then:

$$f_{BFL} = \frac{n_2}{n_2 - n_1}R = \frac{n_2}{\wp} \qquad [10]$$

Thick lenses

A thick lens can be treated as a component consisting of two spherical refracting surfaces separated by a distance d between their vertices V_1 and V_2, as shown in Figure 2. The determination of the focal length as a function of the object and image distance and of the radii of curvature can also be solved using the matrix formalism.

The first and second focal points or the object and image foci, F_o and F_i, can be measured from the two vertices or poles and are called the front and the back focal length denoted by f_{FFL} and f_{BFL}, respectively (Figure 2). The incident and the emerging rays will virtually meet at points that belong to two curved surfaces. These surfaces, approximating planes in the paraxial region, are termed the primary and secondary principal planes, PP_1 and PP_2. Points where the primary and secondary principal planes intersect the optical axis are known as the first and second principal points, H_1 and H_2, respectively.

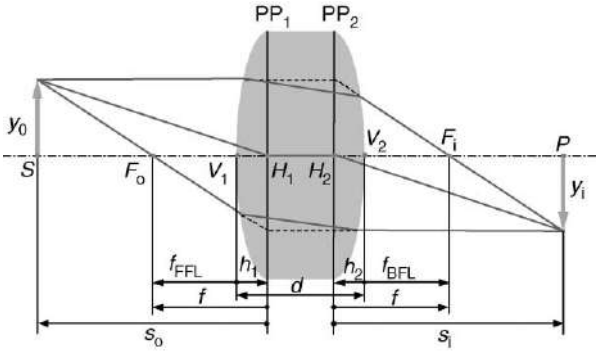

Figure 2 The thick lens geometry.

They provide a set of very useful references from which to measure several of the system parameters. To fully describe a system by its cardinal points, two more points are needed in addition to the two focal points, F_o and F_i and the two principal points, H_1 and H_2. These points are called the nodal points. However, they coincide with the principal points when the lens is surrounded on both sides by the same medium, for example, air. Note that h_1 and h_2 determine the positions of the principal points relative to the vertices V_1 and V_2, while f determines the focal point position relative to the principal points. One can show that these distances are given by:

$$h_1 = \frac{-R_1 n_m d}{n_l(R_2 - R_1) + (n_l - n_m)d}$$

$$h_2 = \frac{R_2 n_m d}{n_l(R_2 - R_1) + (n_l - n_m)d} \qquad [11]$$

$$f = \frac{n_m}{(n_l - n_m)\left[\frac{1}{R_1} - \frac{1}{R_2} + \frac{d(n_l - n_m)}{n_l R_1 R_2}\right]}$$

We now derive the Gaussian formulae that describe imaging through two surfaces which are rotationally symmetric about the same axis. We assume here that the lens, with refractive index n_l, faces the same medium of refractive index n_m on both sides (**Figure 3**). We can correctly use this hypothesis since the microlenses which we discuss throughout this work are mainly surrounded by air. Let R_1 and R_2 be the radii of curvature of the surfaces at their axial points, measured as positive when the surface is convex towards the incident light.

The image of a given object, formed by refraction at the first surface with radius of curvature R_1, becomes the object for refraction at the second surface. The image formed by the second surface with R_2 as radius is then the final image due to the action of the thick lens. We know from eqn [6] that the paraxial rays issuing from S at a distance s_{o1} from V_1 will meet at P', at a distance which we call s_{i1}. This gives:

$$\frac{n_l}{s_{i1}} - \frac{n_m}{s_{o1}} = \frac{n_l - n_m}{R_1} \qquad [12]$$

Thus, at the second spherical surface, eqn [6] becomes:

$$\frac{n_m}{s_{i2}} + \frac{n_l}{(s_{i1} - d)} = -\frac{n_m - n_l}{R_2} \qquad [13]$$

As discussed above, we have:

$$-\frac{1}{f_{FFL1}} = \frac{n_l - n_m}{n_m}\frac{1}{R_1}$$
$$\frac{1}{f_{BFL2}} = -\frac{n_m - n_l}{n_m}\frac{1}{R_2} \qquad [14]$$

where f_{FFL1} is the front focal length of the first refracting surface and f_{BFL2} the back focal length of the second refracting surface.

Eliminating s_{i1} in eqns [12] and [13] and inserting [14] yields:

$$d = \frac{n_l}{n_m\left(\frac{1}{s_{o1}} - \frac{1}{f_{FFL1}}\right)} - \frac{n_l}{n_m\left(\frac{1}{f_{BFL2}} - \frac{1}{s_{i2}}\right)} \qquad [15]$$

where s_{o1} is measured with respect to V_1 and s_{i2} is measured with respect to V_2. After some calculations this equation may also be put in the form:

$$s_{o1}s_{i2} - \frac{f_{BFL2}(n_l f_{FFL1} + dn_m)}{n_l(f_{FFL1} + f_{BFL2}) + dn_m}$$
$$s_{o1} - \frac{f_{FFL1}(n_l f_{BFL2} + dn_m)}{n_l(f_{FFL1} + f_{BFL2}) + dn_m} \qquad [16]$$
$$s_{i2} + \frac{dn_m f_{FFL1} f_{BFL2}}{n_l(f_{FFL1} + f_{BFL2}) + dn_m} = 0$$

Obviously eqn [16] is much too cumbersome to find image locations. However, we can transform and reduce the latter equation to the form:

$$\frac{1}{s_{i2} - h_2} - \frac{1}{s_{o1} - h_1} = \frac{1}{f} \qquad [17]$$

where h_1, h_2 and f are the unknowns.

Indeed, rewriting eqn [17] gives:

$$s_{o1}s_{i2} - (f + h_2)s_{o1} + (f - h_1)s_{i2} + f(h_1 - h_2) + h_1 h_2 = 0 \qquad [18]$$

Equations [16] and [18] will be identical if the coefficients of their corresponding terms are equal;

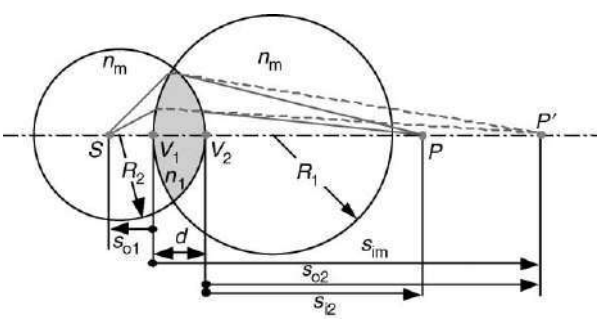

Figure 3 Refraction at both spherical surfaces of a thick lens.

that is when:

$$f + h_2 = \frac{f_{\mathrm{BFL2}}(n_{\mathrm{l}} f_{\mathrm{FFL1}} + d n_{\mathrm{m}})}{n_{\mathrm{l}}(f_{\mathrm{FFL1}} + f_{\mathrm{BFL2}}) + d n_{\mathrm{m}}}$$

$$f - h_1 = -\frac{f_{\mathrm{FFL1}}(n_{\mathrm{l}} f_{\mathrm{BFL2}} + d n_{\mathrm{m}})}{n_{\mathrm{l}}(f_{\mathrm{FFL1}} + f_{\mathrm{BFL2}}) + d n_{\mathrm{m}}} \qquad [19]$$

$$f(h_1 - h_2) + h_1 h_2 = \frac{d n_{\mathrm{m}} f_{\mathrm{FFL1}} f_{\mathrm{BFL2}}}{n_{\mathrm{l}}(f_{\mathrm{FFL1}} + f_{\mathrm{BFL2}}) + d n_{\mathrm{m}}}$$

These three equations may be solved for the three unknowns h_1, h_2, and f:

$$f = \frac{n_{\mathrm{m}}}{(n_{\mathrm{l}} - n_{\mathrm{m}})\left[\dfrac{1}{R_1} - \dfrac{1}{R_2} + \dfrac{d(n_{\mathrm{l}} - n_{\mathrm{m}})}{n_{\mathrm{l}} R_1 R_2}\right]}$$

$$h_1 = \frac{-R_1 n_{\mathrm{m}} d}{n_{\mathrm{l}}(R_2 - R_1) + (n_{\mathrm{l}} - n_{\mathrm{m}}) d} \qquad [20]$$

$$h_2 = \frac{R_2 n_{\mathrm{m}} d}{n_{\mathrm{l}}(R_2 - R_1) + (n_{\mathrm{l}} - n_{\mathrm{m}}) d}$$

Thus we see that the relation between the object position and the image position (eqn [16]) is expressed in the more convenient form of eqn [17] with the values of h_1, h_2, and f given by eqn [20]. Here h_1, h_2 and f depend on the constants of the system where, as shown in eqn [11], h_{i} represents the distance between the vertex V_{i} and the principal point H_{i}, and f is the focal length measured from a principal point on the axis and not from the vertex of the lens. If we now assume that:

$$s_{\mathrm{o}} = s_{\mathrm{o1}} - h_1, \qquad s_{\mathrm{i}} = s_{\mathrm{i2}} - h_2 \qquad [21]$$

we can write the thick lens equation as:

$$\frac{1}{s_{\mathrm{i}}} - \frac{1}{s_0} = \frac{1}{f} \qquad [22]$$

where the object and image distance, s_{o} and s_{i}, respectively, are measured from the principal points of the lens. We can conclude that the formula that describes imaging through two surfaces may be simplified to the thick lens formulation given in eqn [22] by choosing the principal points H_{i} instead of the vertices V_{i}, from which S and P are to be measured.

Thin lenses

The preceding eqn [20] takes a particularly simple form when the lens is so thin that the axial thickness of the lens d may be neglected. But obviously there exists no real lens with a thickness equal to zero. A thin lens is only a concept which is an extremely useful tool in optical system design. When a lens or an optical system has a zero thickness, the object and image calculations can be greatly simplified. In this case, the two principal points are coincident and are located where the thin lens is situated.

As a further simplification, we assume that the surrounding medium of the lens is air which means that $n_{\mathrm{m}} \approx 1$. When $n_{\mathrm{l}} = n$, we find:

$$\frac{1}{f} = \frac{1}{s_{\mathrm{i}}} - \frac{1}{s_{\mathrm{o}}} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad [23]$$

These equations are often called the thin-lens equation and the lensmaker's formula, respectively, because the latter predicts the focal length of a lens, fabricated with a given refractive index and radii of curvature. This formula is very useful in the optical design process as a first-order approach to calculate the features of the thin lenses and results in a set of lens powers (or focal lengths) for the different lenses and in the spacings between them. Once the thin lens optical system is solved and the primary aberrations are corrected for, one can design the thick lens system. The corresponding thick lens system must have the same component powers and the lenses must be spaced apart by the same distances as for the thin lens design, but the thick lens spacings must be measured from the principal points of the lenses instead of from the vertices of the lenses.

Other lens characteristics

Another basic parameter of a refractive lens is its focal number $f\#$. The f-number, also called the speed of the lens, is defined as the ratio of its effective focal length and the lens aperture diameter D (Figure 4):

$$f\# = \frac{f}{D} \qquad [24]$$

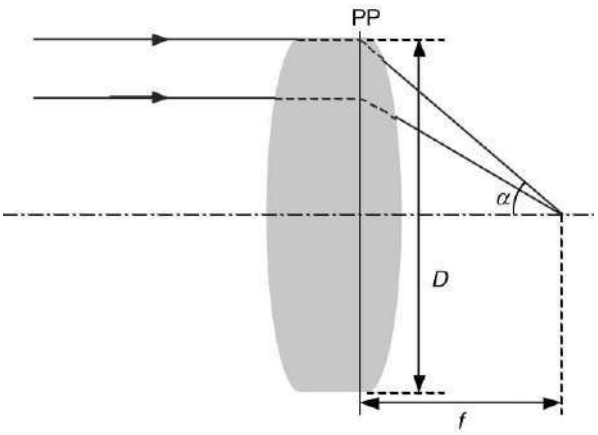

Figure 4 A lens with an aperture diameter D, a focal length f, and a semi-angle α.

The effective focal length f of a lens or often called the focal length is defined from a principal point on the axis of the lens and not from its vertex. Typical focal numbers for refractive microlenses range between 0.8 and 30.

The numerical aperture (NA), is another way of defining the same characteristic. It is described as the sine of the angular semi-aperture α in the object space multiplied by the refractive index n of the object space (Figure 4):

$$\mathrm{NA} = n \sin \alpha \qquad [25]$$

Both the focal number ($f\#$) and the NA are related in the paraxial approximation by

$$f\# = \frac{n}{2\mathrm{NA}} \qquad [26]$$

Lens Aberrations

Introduction to Third-Order Monochromatic Aberrations

An exact ray trace or a measurement on a real lens will certainly disclose inconsistencies with the corresponding paraxial description. Such deviations from the idealized Gaussian optics are known as aberrations. For monochromatic light there are five third-order or Seidel aberrations: three of them, namely spherical aberration, coma, and astigmatism, deteriorate the image by blurring it, while the remaining two, curvature of field and distortion, deform the image. An additional aberration, chromatic aberration, results from the wavelength dependence of the imaging properties of an optical system. We will not discuss the latter as this aberration can be minimized by selecting a lens material with a refractive index that only varies by a negligible amount over the wavelength range of interest.

For a brief discussion of aberrations we consider the case of a perfect lens focusing a collimated light beam. In this case, the lens transforms the incoming planewave into a spherical wave, also called the reference sphere. For a lens with aberrations, however, the real wavefront behind the lens shows deviations from the ideal spherical shape (Figure 5). The optical path length [AB] may be called the wavefront aberrations $\psi(x, y)$ and can be calculated as:

$$\psi(x, y) = W(x, y) - R(x, y) \qquad [27]$$

where $W(x, y)$ represents the spherical wavefront in the Gaussian approximation and $R(x, y)$ depicts the real wavefront. The wave aberration is given by the optical path difference along a ray between the

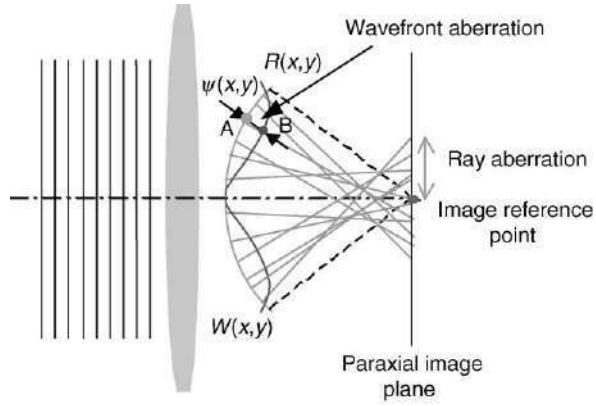

Figure 5 Relation between wavefront and ray aberration.

reference sphere and the real wavefront behind the lens. Alternatively, the aberrations can be described in terms of the ray aberrations. In this case, not all the rays intersect the paraxial image plane at the same point and the distribution of the ray intercepts around the paraxial image point can be used to measure the amount of light blur. These two ways of quantitatively describing the aberrations are apparently not independent of each other, the ray aberrations being the derivate of the wavefront aberrations.

There are several sets of suitable polynomials over which the wavefront aberrations can be expanded. The Zernike polynomials will be further used in this work (see, for example, the Mach–Zehnder and Twyman–Green interferometers in the section on Microlens testing).

For rotational symmetric optical systems, the most common approach is to expand the aberration function ψ to the fourth order as a power series in polar pupil coordinates. This corresponds to the third-order ray aberrations or primary aberrations. The different terms in the aberration polynomial ψ are to the fourth power of the object coordinate h' and the pupil coordinate r. In that case, the wavefront aberration polynomial ψ, as a function of the object coordinate h' and the polar pupil coordinates r and θ (Figure 6), is given by

$$\psi(h', r, \theta) = {}_0C_{40}r^4 + {}_1C_{31}h'r^3 \cos\theta + {}_2C_{22}h'^2r^2 \cos^2\theta + {}_2C_{20}h'^2r^2 + {}_3C_{11}h'^3 r \cos\theta \qquad [28]$$

The ${}_iC_{jk}$ coefficients are subscripted by the numbers that specify the powers of the term dependence on h', r and $\cos\theta$, respectively. These terms comprise the five monochromatic or Seidel aberrations as shown in Table 1.

Each aberration is characterized by its dependence on the deviation from the optical axis h', on the aperture of the refracting surface r, and on the

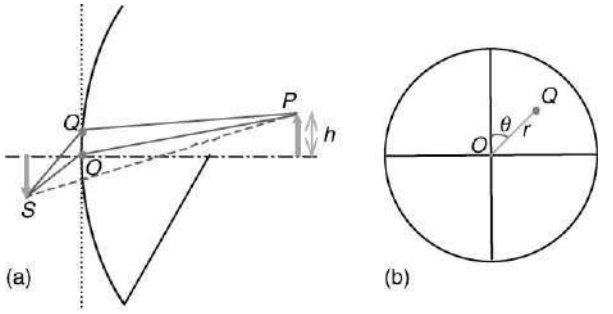

Figure 6 (a) Imaging of an off-axis point P; (b) frontal view of a portion of the wavefront.

Table 1 The five monochromatic or Seidel aberrations

r^4	Spherical aberration
$h' r^3 \cos\theta$	Coma
$h'^2 r^2 \cos^2\theta$	Astigmatism
$h'^2 r^2$	Curvature of field
$h'^3 r \cos\theta$	Distortion

symmetry around the optical axis θ. Equation [28] can be used to estimate how the aberration blurs change if the aperture or field of view of a lens is changed. In the following section we briefly describe each of these aberrations in terms of their visual effects.

Spherical Aberration

The first term in eqn [28], ${}_0C_{40}r^4$, is the only term in the third-order wave aberration that does not depend on the deviation h' from the optical axis. Thus, spherical aberration (SA) is the only monochromatic aberration that occurs on the optical axis and that exists even for axial object and image points. As shown in Figure 7a, SA results from the rays that enter via different zones of the aperture and are focused at different distances from the lens. It is evident from Figure 7a that the best focus is not at the paraxial plane. The best focus, however, is found at the location where the circular image blur is the same over the entire field of view and is called the circle of least confusion.

Because most applications consist of on-axis illuminated planoconvex microlenses, we will firstly focus our attention on how microlenses have to be positioned to minimize SA. When we consider a planoconvex lens with the curved surface towards the infinite conjugate for an incident parallel beam, both surfaces are contributing approximately equal amounts of ray bending. Figure 8a shows the ray diagram for a planoconvex lens as obtained with the ray-tracing program Solstis. Solstis is a commercial photonics design software, consisting of different

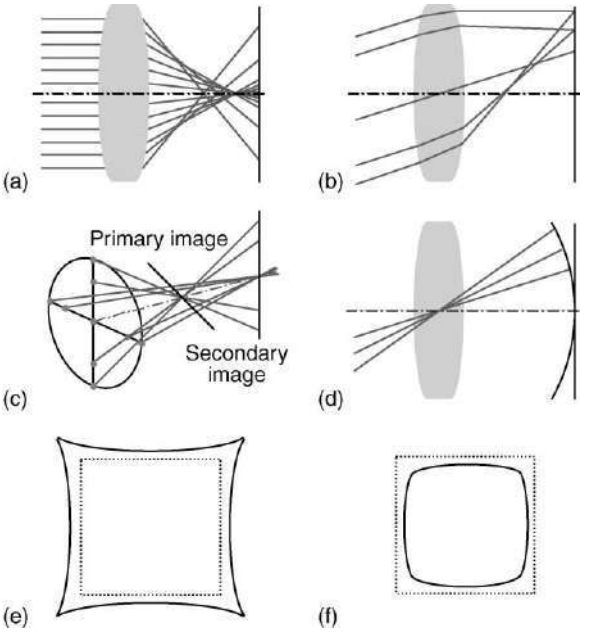

Figure 7 The primary aberrations: (a) spherical aberration – negative SA: the rays through the outer zones of the lens focus closer to the lens than the rays through the central zones; (b) coma – positive coma: the rays through the outer zones of the lens form a larger image than those going through the center; (c) astigmatism – negative astigmatism: the tangential or vertical fan of rays is focused to the left of the horizontal or sagittal fan; (d) curvature of field – negative curvature of field: images farther from the axis focus nearer to the lens than the on-axis images; (e) positive distortion – the magnification increases as the field angle increases; (f) negative distortion.

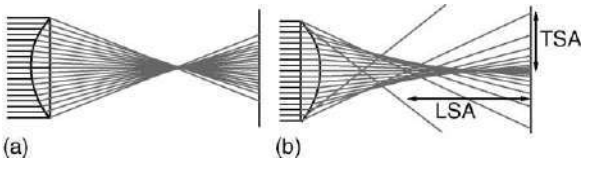

Figure 8 Ray diagram of a planoconvex lens ($D = 200$ μm; $f = 249$ μm) with planewave illumination incident on (a) the curved surface and (b) on the plane surface. LSA (longitudinal SA) represents the difference between the paraxial and the marginal focus while TSA (transversal SA) shows the radius of the spot in the paraxial focal plane.

modules, each of which is dedicated to a certain optical modeling approach. If, however, the flat surface is turned toward the incident rays, then the first surface does not bend the rays and all the ray-bending work is done by the second surface (Figure 8b). We can conclude here that the best orientation is that which distributes the ray bending equally between the two surfaces of the lens.

Because the ray blur varies with the third power of the lens aperture, lowering the aperture diameter will cause a rapid decrease in the extent of the spherical

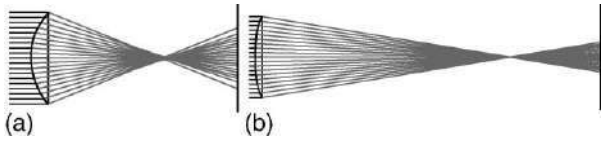

Figure 9 Ray diagram of a planoconvex lens (a) with a focal length *f* and (b) with a focal length *2f*.

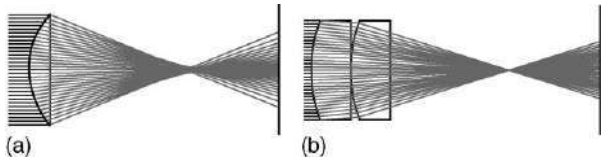

Figure 10 (a) Ray diagram for a planoconvex lens ($D = 200$ μm; $f = 249$ μm). (b) The spherical aberration is reduced with a factor 4 by splitting the optical power of the original lens and using a doublet ($D = 200$ μm; $f = 621$ μm for both lenses).

aberration blur. For a given aperture of the lens, the spherical aberration is a function of the focal length of the lens and increases with a decreasing focal length or an increasing NA (Figure 9).

A change in SA, contributing to the wave aberration, for a variation of the lens diameter D or the focal length f can be estimated with the following rules:

$$\text{Spherical aberration} \div D^4 \quad \text{for a constant } f$$

$$\text{Spherical aberration} \div 1/f^3 \quad \text{for a constant } D$$

However, by preserving the focal length and hence the optical power of the lens, the SA can be changed by varying the radii of the surfaces. This would be the same as physically bending a lens made of flexible plastic. In Figure 10 we have reduced the spherical aberration with a factor of four by splitting a lens into two elements, each with a longer focal length, keeping the total optical power of the elements the same as the original lens. This means that, in practice, high NA microlenses can be replaced by a number of microlenses reducing the SA significantly.

However, in optical systems, where only a small number of surfaces are allowed, aspherical surfaces can be used to correct the SA. In Figure 11 we show that making the lens shape elliptical while maintaining the other characteristics of the microlens ($D = 200$ μm; $R = 150$ μm) results in a reduction of the SA. The most common form of an aspheric surface is a rotationally symmetric surface with the sag, defined as:

$$z_{\text{conic}} = \frac{R^{-1}(x^2 + y^2)}{1 + \sqrt{1 - (1 + K)R^{-2}(x^2 + y^2)}} \qquad [29]$$

where R is the radius of curvature, K the conic parameter, and x and y the coordinates in the substrate plane. The meaning of K is given in Table 2.

In eqn [28] we have shown that the wavefront aberrations can be written as the sum of the Seidel or third-order aberrations. Furthermore, we have seen that only the SA contributes to the wavefront aberration when we illuminate the microlens on-axis ($h' = 0$). When we limit our experiments to on-axis illumination during the optical characterization of the microlenses, the wavefront aberrations are a measure for the SA. To prove this, we have calculated and plotted in Figure 12, the SA and the wavefront aberrations as a function of the focal length for a 200 μm lens ($n = 1.48$).

Both the SA and wavefront aberration values are obtained with the ray-tracing software Solstis. The wavefront aberration is calculated here as the optical path difference [BA] between a real wavefront surface and a spherical reference surface. From these results we can observe that for microlenses with an $f > 330$ μm or an NA lower than 0.3 only the spherical aberration contributes to the root-mean-square (RMS) wavefront aberration. The RMS or standard deviation σ of a number N of data x_i ($i = 1, \ldots, N$) with a mean value X can be written as $\sigma = \sqrt{\sum_{i=1}^{N} (X - x_i)^2 / N}$. For smaller focal length microlenses, however, we notice that higher-order aberrations appear and compensate the SA, resulting in a lower wavefront aberration compared to the SA. We can conclude that the wavefront aberration values represent the optical performance of the microlenses with a main contribution from the SA for on-axis illumination.

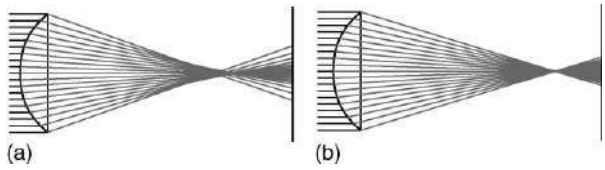

Figure 11 (a) Ray diagram of a spherical planoconvex lens ($D = 200$ μm; $R = 150$ μm; $K = 0$) with planewave illumination. (b) The spherical aberration is reduced by making the lens shape elliptical ($K = -0.57$).

Table 2 The surface type for different conic parameters K

Conic constant K	*Surface type*
$K = 0$	Sphere
$K < -1$	Hyperboloid
$K = -1$	Paraboloid
$-1 < K < 0$	Ellipsoid
$K > 0$	Oblate ellipsoid

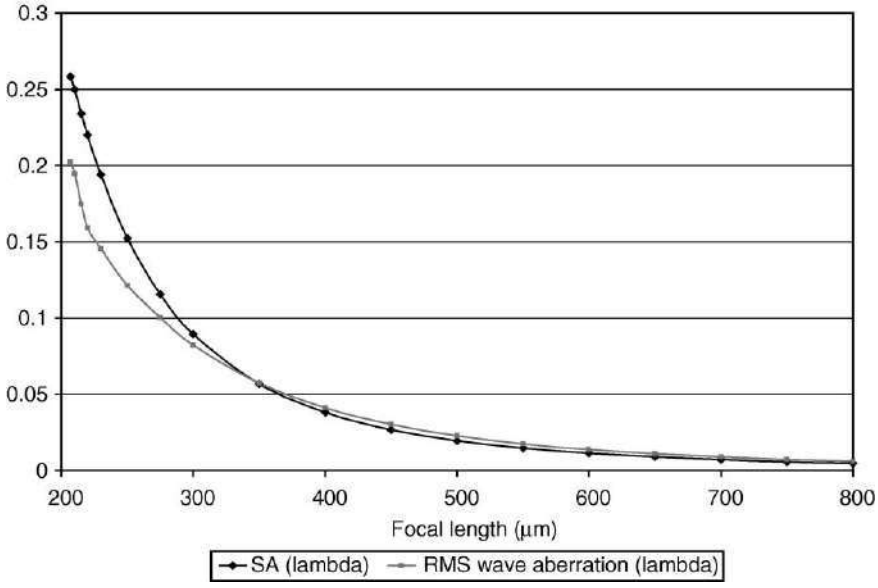

Figure 12 The spherical aberration SA and the RMS wave aberration as a function of the focal length for a 200 μm lens ($n = 1.48$).

Other Seidel Aberrations

Coma is an image-degrading aberration that increases rapidly with the lens aperture r. It indicates an off-axis aberration ($h' \neq 0$) that is nonsymmetrical around the optical axis ($\cos\theta \neq$ constant). Figure 7b illustrates the aberration due to a vertical fan of parallel rays refracted by a spherical lens. It results from a lens having different magnifications for rays passing different zones of its aperture. Each circular zone of the lens forms a circular image called the comatic circle. Every other fan of rays forms images that complete the comatic circle. The combination of all these circles causes a comet-shaped flare.

Astigmatism is the imaging of an off-axis point as two perpendicular lines at different distances. In this aberration the rays from a point object converge at a certain distance from the lens to a line which is called the primary image. This line lies in the sagittal plane and is perpendicular to the plane defined by the optical distance and the object point. At a somewhat different distance from the lens they converge to a second line, called the secondary image, which is parallel to the tangential plane. This effect is shown in Figure 7c. The elimination of astigmatism requires that the tangential and sagittal surfaces be made to coincide. The resulting surface is then called the Petzval surface. Although in that case astigmatism has been eliminated, the associated aberration, called curvature of field, remains. So off-axis images are focused on a curved saucer-shaped surface as drawn in Figure 7d, instead of on an ideal flat image surface.

The last of the five primary aberrations is distortion and causes straight lines which do not intersect the axis to be imaged as curved lines. This effect is not related to a lack of sharpness of the image but results from the fact that the magnification varies across the field and induces the image of a square object to be bowed outward (Figure 7e) or sagged inward (Figure 7f).

We have shown that three of the five Seidel aberrations (namely SA, coma, and astigmatism) are responsible for a lack of sharpness of the image. The other two (i.e., curvature of field and distortion) are related to the position and the form of the image. In general, it is impossible to design a system which is free from all the primary as well as the higher-order aberrations, and therefore, a suitable compromise as to their relative magnitude has to be made. In some cases, the effects of the Seidel aberrations are reduced by balancing them against aberrations of higher orders. In other cases, one has to eliminate certain aberrations completely, even at the price of introducing aberrations of other types. So far we have studied the aberration effects on the basis of geometrical optics only. However, if the aberrations are very small (wave aberrations of the order of a wavelength or less), diffraction starts to play an important role. That is why, in the next section, we will discuss the role of diffraction on the imaging quality of a lens.

Diffraction

Introduction

So far we have used simple geometrical optics to evaluate the effects of aberrations – deviations of the

wavefront behind the lens from the ideal shape – that influence the quality of an image generated by a lens. However, we know that even for an aberration-free lens, blur of the spot in the image plane occurs. This means that diffraction effects at the lens aperture can also influence the spot size and the point spread function (PSF), certainly because diffraction gains importance when miniaturizing optical components. Therefore, we have to introduce diffraction before we can bring forward the most important figures of merit for microlens performance and study its effects on imaging quality.

Fraunhofer and Fresnel Diffraction

In certain regions of space occupied by our optical lens system, the simple geometrical model of energy propagation is inadequate. In particular, deviations from this model can be expected in the immediate vicinity of the boundaries of shadows and in regions where a large number of rays meet. These regions are of great practical interest, as they include that part of the image space in which the optical image is situated. These deviations are manifested by the appearance of diffraction fringes. Also, the phenomena involving the superposition of coherent wavefronts cannot be treated with a geometrical model of energy propagation. Thus a more accurate treatment is necessary where the wave nature of light and diffraction are taken into account. Wave optics is such an approach with which the actual light intensity, that will be observed as a result of the propagation of a wavefront to the image plane, can be constructed. According to Huygens' construction, every point of a wavefront may be considered as a center of secondary disturbance, which gives rise to spherical wavelets. The wavefront at any later instance may be regarded as the envelope of these wavelets. Fresnel was able to account for diffraction by supplementing Huygens' construction with the postulate that the secondary wavelets mutually interfere. We discuss diffraction here, on the basis of wave theory, by the application of Huygens' construction together with the principle of interference. The latter principle, also called the Huygens–Fresnel principle, sufficiently describes the propagation of light in free space.

This Huygens–Fresnel principle may be regarded as an approximate form of the integral theorem of Helmholtz and Kirchhoff. Kirchhoff's theory makes an approximation to the diffraction of scalar waves. This scalar theory is adequate enough for the treatment of problems in instrumental optics where polarization effects are of minor importance and can be neglected. If we consider a strictly monochromatic electric field $\varepsilon(x, y, z, t) = E(x, y, z)\cdot e^{-i\omega}$ in any ideal isotropic dielectric, the space-dependent part will satisfy the homogeneous time-independent wave equation: $(\nabla^2 + k^2)E = 0$, also known as the Helmholtz equation where $k = \omega/c$. This theorem expresses the solution E of the Helmholtz equation at an arbitrary field point P, in terms of its solution E and its first derivatives $\partial E/\partial n$ at all points on an arbitrary closed surface surrounding P, provided the medium is a homogeneous dielectric:

$$E(P) = -\frac{1}{4\pi}\oiint_S \left[E\overline{\nabla}\left(\frac{e^{-iks}}{s}\right) - \frac{e^{-iks}}{s}\overline{\nabla}E \right]\bar{n}\,dS \qquad [30]$$

Here, s is the distance from point P to the point on surface S, n is the outward normal to S, and $k = \omega/c$ is the wave number in vacuum.

When we consider a monochromatic wave, from a point source P_o propagating through an aperture Ω in a plane opaque screen, and P is the point at which the light disturbance has to be calculated, the theorem can be reduced to an approximate but much simpler form (Figure 13a) on condition that we can assume that the linear dimensions of the aperture Ω, although large compared to the wavelength, are small compared to the distances of both P_o and P from the screen.

As a result of Kirchhoff's theorem we obtain:

$$E(P) = \frac{iA}{2\lambda}\iint_\Omega \frac{e^{-ik(r+s)}}{rs}[\cos(n,r) - \cos(n,s)]dS \qquad [31]$$

where A is a constant, λ the wavelength, and r and s are the distances from element dS on Ω to P_o and P, respectively. Equation [31] is known as the Fresnel–Kirchhoff diffraction formula. In case a more general field E_o is incident at the aperture Ω, replacing the point source $A\,e^{-ikr}/r$, an equivalent formulation, known as the Huygens integral, can be found:

$$E(P) = \frac{i}{\lambda}\iint_\Omega E_o(x_o, y_o, z_o)\frac{e^{-iks}}{s}\cos(n,s)dS \qquad [32]$$

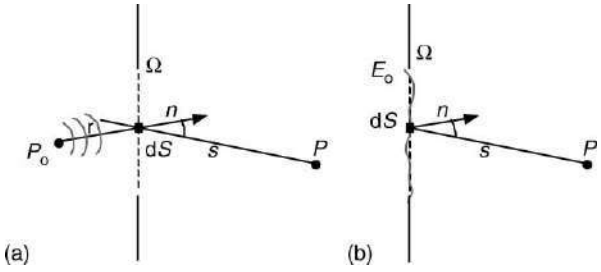

Figure 13 Geometry for evaluating (a) the Fresnel–Kirchhoff diffraction formula and (b) the Huygens' integral.

In the case of the Huygens integral an alternative Green's function is chosen, namely $G(P) = \exp(-jks)/s - \exp(-jkr)/r$ instead of $G(P) = \exp(-jks)/s$ used in the Fresnel–Kirchhoff diffraction formula. This then results in an obliquity factor $\cos(n,s)$ instead of $\cos(n,s) - \cos(n,r)$. If the distances of the points P_o and P are large compared to the dimensions of the aperture, the factor $[\cos(n,r) - \cos(n,s)]$ in eqn [31] will not vary appreciably over the aperture and may be replaced by 1. In this case, the optical beams are close to the optical axis and this is considered as the paraxial diffraction theory. In the phase factors, we can expand $r(x,y,z)$ and $s(x,y,z)$ as power series. If the quadratic and higher-order terms may be neglected in the series expansion ($r,s \gg \lambda$), one speaks of Fraunhofer diffraction, when the previous assumption is not fulfilled ($r,s \geq \lambda$), the quadratic terms cannot be neglected and one speaks of Fresnel diffraction. The regions where the Fresnel and Fraunhofer diffraction integral may be applied are called the near-field (NF) and far-field (FF) diffraction regions respectively. The Rayleigh range z_R marks the boundary between the NF and FF regions for the beam emerging from the aperture (Figure 14).

As an example, we discuss the diffraction of a uniform planewave by a circular aperture of diameter $2a$. The Rayleigh range for this circular aperture is given by

$$z_R = \frac{\pi a^2}{\lambda} \qquad [33]$$

The FF diffraction pattern ($z \gg z_R$) for a uniformly illuminated circular aperture is the well-known Airy disk pattern and is given by

$$E(r,z) \sim \frac{2J_1(2\pi a\theta/\lambda)}{(2\pi a\theta/\lambda)} \qquad [34]$$

where J_1 is the first-order Bessel function. This pattern has a single dominant central lobe surrounded by a series of increasingly weaker circular rings. The first zero of this pattern occurs at a half angle θ_1 or a radius r_1 in the FF, given by

$$\theta_1 = \frac{r_1}{z} \approx \frac{1.22\lambda}{2a} \qquad [35]$$

The NF or Fresnel diffraction patterns for a uniformly illuminated circular aperture ($z < z_R$) consist of a series of circular rings modulating a constant-amplitude background. When moving even closer to the aperture, the beam profile becomes more and more square and the frequency of the Fresnel ripples increases.

We now know that light is diffracted at the apertures of optical elements (lenses) and that this diffraction can affect its performance. Ideally, the focus of a planewave should be infinitely small, being the image of a point source located at infinity. But even in the ideal case, when no geometrical aberrations are introduced, diffraction will limit the lens performance and the focus of such diffraction-limited lenses will have a finite extension.

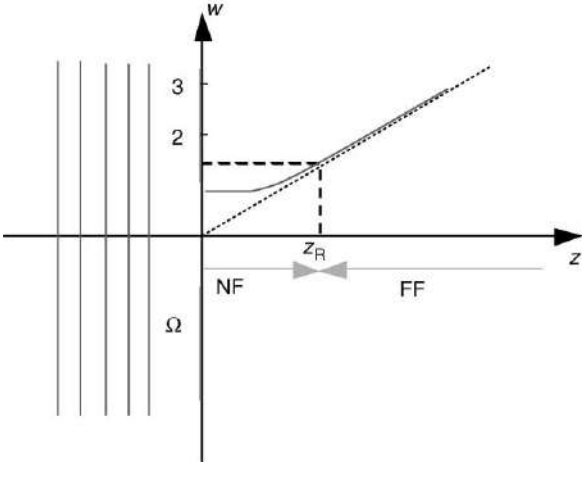

Figure 14 The Rayleigh range z_R marks the boundary between the NF and FF regions for a beam emerging from the aperture Ω.

Quality Criteria for Lens Performance

In this section we discuss some of the most important figures of merit for diffraction-limited lens performance. This survey raises the problems related to this issue. A more detailed overview can be found in the literature.

Diffraction in a Lens System: The Point Spread Function (PSF)

Diffraction at the lens aperture D causes a blur of the focus. The light distribution in the focal plane of an aberration free lens essentially arises from Fraunhofer diffraction at the aperture of the lens. From eqn [34] we know that the diffraction image for an aberration-free lens with a circular aperture D is the scaled Airy disk, given by

$$E(r,z) \sim \frac{\pi J_1(rD/\lambda f)}{(rD/\lambda f)} \qquad [36]$$

where λ denotes the wavelength of the illuminating light beam and f the focal length.

The normalized intensity of this pattern is called the diffraction-limited (PSF) of the lens:

$$\mathrm{PSF}(r,z) = |E(r,z)|^2 \qquad [37]$$

Although irradiance replaces nowadays the word intensity in optics, we will still use intensity to nominate the flow of energy per unit area per unit time. We can define the irradiance as $I \equiv \frac{1}{2}\varepsilon_0 c\langle \mathrm{E}^2\rangle$, with ε_0 and c the electric permittivity in vacuum and the speed of light, respectively. The radius of this pattern is known as the Rayleigh diffraction limit or Rayleigh resolution and is compared to the measured spot size in the image plane.

To include the effects of aberrations in the diffraction computation, we have to include both the phase profile and the amplitude of the real wavefront in the pupil of the lens. The pupil function $T(x, y)$ may be written as:

$$T(x,y) = E(x,y)\mathrm{e}^{\mathrm{i}k\psi(x,y)} \qquad [38]$$

where $E(x, y)$ represents the amplitude distribution over the exit pupil and $\psi(x, y)$ is the wavefront deformation on this pupil. Starting from the Huygens–Fresnel approach, the diffraction integral can be rewritten in the form of a Fourier transform:

$$E(r,z) \sim FT(\mathrm{e}^{\mathrm{i}k\psi(x,y)}) \qquad [39]$$

The PSF at the focal position can then be found by calculating the Fourier transform of the wavefront error ψ or the pupil function $T(x, y)$ of the lens and taking the complex square of the latter. Figure 15 shows the shape of the PSF of an ideal or diffraction-limited lens. When the lens aperture D decreases, the extension of the PSF increases proportionally, which means that it becomes easier to achieve diffraction-limited performance. Since the extension of the diffraction-limited PSF becomes larger, the constraints on the shape of the phase profile become less stringent. This has important consequences on the scaling behavior of lenses. In the extreme case, the microlens diameter becomes so small, that diffraction-limited imaging is possible without any specific phase profile in the lens pupil. This is also one of the reasons why small water droplets often form relatively good lenslets. For systems including micro-optic lenses, we therefore will have to find better criteria than the PSF of a diffraction-limited lens to evaluate or compare these optical components.

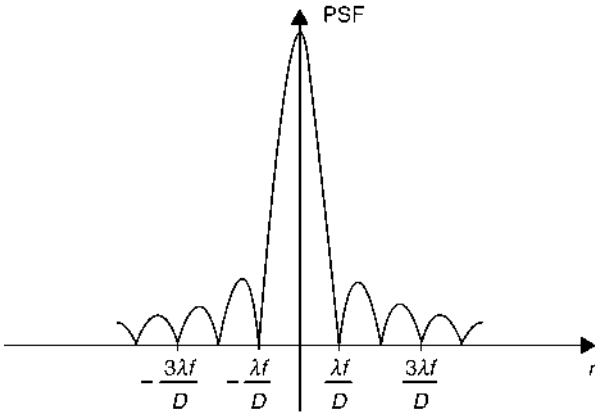

Figure 15 The 1D point spread function of an ideal lens.

The Rayleigh and Maréchal Criterion

The widely used Rayleigh criterion states that a system is diffraction-limited, as long as the maximum wavefront deviation $|\psi_{\max}(x)|$ is less than one quarter of a wavelength (Figure 16). It was shown by Rayleigh that for a system which suffers primarily from SA such that the wavefront deviation is less than a quarter wavelength, the intensity at the Gaussian focus is diminished by less than 20%; a loss of light that can usually be tolerated. Also in the presence of other commonly occurring Seidel aberrations is the quality of the image not seriously affected when the deviation is less than $\lambda/4$. This criterion, which determines the amount of aberration that can be tolerated in an image-formation system is, of course, only a rough guideline since the light distribution in the image does not only depend on the maximum deformation but also on the shape of the wavefront, and even more on the type of aberration. Moreover, the loss of light that can be tolerated depends on the particular use to which the optical component is put and more stringent tolerances have to be imposed for certain applications.

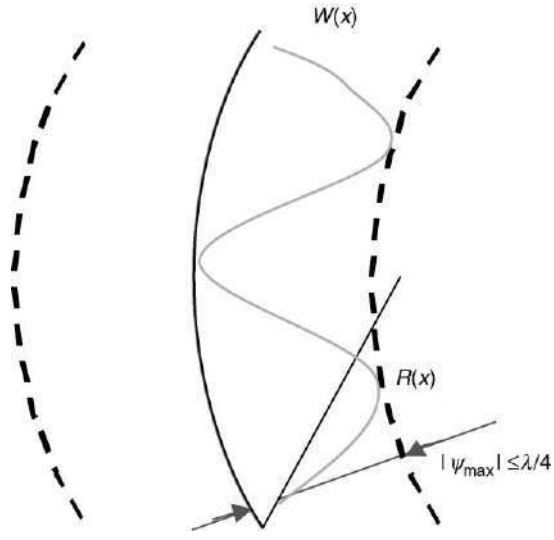

Figure 16 Illustration of the Rayleigh criterion.

Low-order aberrations merely cause a shift of light from the main lobe of the PSF to the side lobes. Higher-order aberrations additionally cause a broadening of the PSF and the correlation between the image quality and the maximum value of the wavefront aberration is usually lost. In the latter case, it seems more appropriate to formulate tolerance criteria restricting the average amount of aberration. Criteria of this type were considered by Maréchal, who used the relation that exists between the intensity at the center of the reference sphere and the RMS deviation of the wavefront from a spherical form.

This Maréchal criterion uses the following RMS wavefront aberration ψ_{RMS} to define the lens quality:

$$\psi_{\mathrm{RMS}} = \sqrt{\int |\psi(x)|^2 \, \mathrm{d}x - \left[\int |\psi(x)| \mathrm{d}x\right]^2} \leq \frac{\lambda}{14} \qquad [40]$$

This criterion is sensitive to statistical noise on the measured phase profile, while a criterion such as the Rayleigh criterion is little affected by it.

Strehl Ratio

For some applications not the imaging quality but the maximum deposited energy is the most essential feature. In such a situation, the Strehl ratio S is a suitable figure of merit. The Strehl ratio is defined as the normalized peak intensity of the PSF of the lens:

$$S = \frac{I_{\mathrm{real}}(0,0)}{I_{\mathrm{ideal}}(0,0)} = \left| \int\int \mathrm{e}^{\mathrm{i}k\psi(x,y)} \mathrm{d}x \, \mathrm{d}y \right|^2 \qquad [41]$$

$I_{\mathrm{real}}(0,0)$ and $I_{\mathrm{ideal}}(0,0)$ denote the intensities at the center of the real point image and the ideal PSF without aberrations, respectively. Even for small aberrations, which do not affect the extension of the PSF, the light intensity in the center peak can drop significantly (Figure 17). A Strehl ratio $S \geq 0.8$ is generally considered to correspond to diffraction-limited performance.

By a Taylor expansion of eqn [40], the relation between the Strehl ratio S and the Maréchal criterion can be written as:

$$S = 1 - \left(\frac{2\pi}{\lambda}\psi_{\mathrm{RMS}}\right)^2 \qquad [42]$$

If only low-order aberrations are taken into account, Rayleigh's $\lambda/4$ criterion also results in a Strehl ratio of $S = 0.8$. For large aberrations ($S \leq 0.3$), however, the behavior of the Strehl ratio becomes uncorrelated to the image quality due to interference effects.

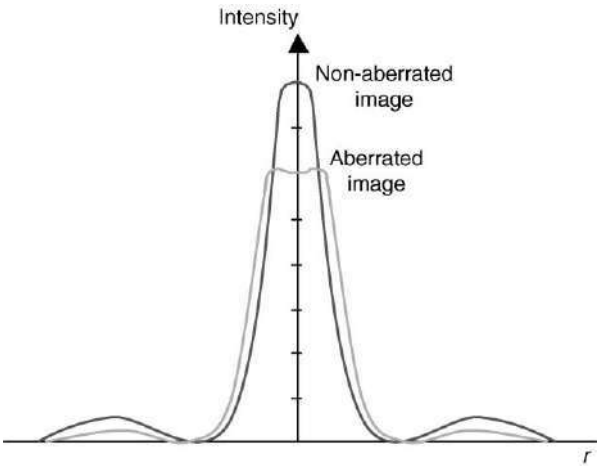

Figure 17 Strehl ratio criterion.

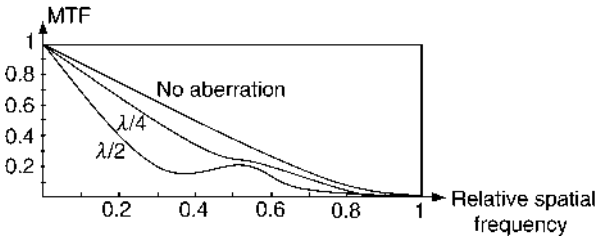

Figure 18 Modulation transfer function for a perfect lens and for small amounts of spherical aberration ($\lambda/2$ and $\lambda/4$).

Transfer Functions

Transfer functions are most commonly used for evaluating the image quality in lens design. As discussed above, the PSF is the Fourier transform of the wavefront error ψ. For simplicity we can assume that the modulus of the pupil function is 1. The PSF then follows from the Fourier transform:

$$PSF(\mu, \nu) = |\mathrm{FT}(\mathrm{e}^{\mathrm{i}k\psi(x,y)})|^2 \qquad [43]$$

The optical transfer function (OTF) is calculated from the PSF data by means of an inverse Fourier transformation:

$$\mathrm{OTF}(\vartheta, \rho) = \mathrm{FT}^{-1}[\mathrm{PSF}(\mu, \nu)] \qquad [44]$$

The OTF is a spatial frequency-dependent complex quantity whose modulus is the modulation transfer function (MTF) and whose phase is, of course, the phase transfer function (PTF). The former is a measure for the reduction in contrast from the object to the image over the whole spectrum. Figure 18 shows the normalized MTF for both a perfect lens and a lens subjected to some small amounts of spherical aberration. The PTF is, in most cases, not so conclusive concerning the optical behavior.

Therefore the wave aberrations, the PSF, the MTF, and the Strehl ratio are the characteristics that are usually provided as an output of lens design programs.

Interference

Many of the processes for the manufacturing of microlenses are not fully understood nor perfectly controlled. It occurs that one generates a microlens array and then measures its performance, hoping that by accurately repeating the processing conditions, a consistent result can be achieved. To achieve optimum conditions, it is therefore crucial to measure the geometrical and optical parameters of the microlenses. Most of the microlens measurement instruments are based on two-beam interference, the phenomena where light from a source is divided into two beams which are later on superposed again. The intensity in the region of superposition is found to vary from point to point between maxima which exceed the sum of the beam intensities and minima which may be zero.

Two-Beam Interference

We consider the interference of two waves, a reference and an object wave, represented by $\mathbf{E}_{\text{ref}}$ and $\mathbf{E}_{\text{obj}}$ respectively, where we take into account the vector properties of the electric fields. In the case of interference, both waves typically originate from a single source and recombine after traveling along different paths. For the moment we will consider only monochromatic linearly polarized waves of the form:

$$\mathbf{E}_{\text{ref}} = \mathbf{E}_{0\text{ref}}\, e^{i\Phi_{\text{ref}}(r) - i\omega t}, \quad \mathbf{E}_{\text{obj}} = \mathbf{E}_{0\text{obj}}\, e^{i\Phi_{\text{obj}}(r) - i\omega t} \qquad [45]$$

The total electric field intensity $\mathbf{E}$ at the output of the interferometer, according to the principle of superposition, is given by:

$$\mathbf{E} = \mathbf{E}_{\text{ref}} + \mathbf{E}_{\text{obj}} \qquad [46]$$

The total intensity at the output, being the time average of the square of the wave amplitude, can be written as

$$I = I_{\text{ref}} + I_{\text{obj}} + I_{\text{ref/obj}} \qquad [47]$$

with $I_{\text{ref}} \sim \langle \mathbf{E}_{\text{ref}}^2 \rangle$; $I_{\text{obj}} \sim \langle \mathbf{E}_{\text{obj}}^2 \rangle$ and $I_{\text{ref/obj}} \sim 2\langle \mathbf{E}_{\text{ref}} \cdot \mathbf{E}_{\text{obj}} \rangle$. Notice that if both $\mathbf{E}$-fields are perpendicular to one another, the interference term $I_{\text{ref/obj}}$ vanishes, while if we assume that both fields are parallel, we can restrict our discussion to scalar theory. Under the latter condition we can write the total intensity as:

$$I = I_{\text{ref}} + I_{\text{obj}} + 2\sqrt{I_{\text{ref}} I_{\text{obj}}} \cos(\Phi_{\text{obj}} - \Phi_{\text{ref}}) \qquad [48]$$

where the presence of the third term is indicative of the wave nature of light which can produce enhancement or diminution of the intensity through interference.

We can rewrite eqn [48] in Michelson's manner of writing:

$$I = I_0(1 + V\cos(\Phi_{\text{obj}} - \Phi_{\text{ref}}))$$
$$\text{where } I_0 = I_{\text{ref}} + I_{\text{obj}}, \quad V = \frac{2\sqrt{I_{\text{ref}} I_{\text{obj}}}}{I_{\text{ref}} + I_{\text{obj}}} \qquad [49]$$

with V the visibility or fringe contrast indicating the degree of modulation.

We can conclude that the measured intensity of a two-beam interferogram depends on the mean intensity I_0, the visibility V, and the phase difference $\Phi_{\text{obj}} - \Phi_{\text{ref}}$. There are a number of techniques to measure the three unknowns: the mean intensity, the visibility, and the interesting spatial phase variation. In the optical instrumentation, further discussed in this article, phase shifting interferometry (PSI) is used which requires a measurement of the intensity in each point of the aperture for at least three different reference phase values, in order to obtain three equations for the three unknowns (I_0, V, and $\Phi_{\text{obj}} - \Phi_{\text{ref}}$). In practice, the reference phase values are obtained by a piezodriven transducer (PZT) mirror and should be equally spaced over one phase cycle of 2π. It has been shown that the five-phase algorithm is the optimal choice as the calculation still remains simple and the second-order nonlinear response of the detector is eliminated. A charge coupled device (CCD) sensor then records the intensity of the resulting interference pattern for different relative phase shifts and converts the intensity data to wavefront phase data through integration.

Microlens Testing

In microlens (lens with a diameter lower than several millimeters) testing, it is essential to measure in a first step the shape or the surface profile of the microlenses or in cases where gradient-index lenses are concerned, their refractive index distribution. In addition, the paraxial parameters, such as the focal length, and the more complex properties like the wave aberrations and surface quality have to be characterized. As discussed in the introduction, massive parallelism and miniaturization ask for the fabrication of microlenses and microlens arrays. The most critical features of these lenslet arrays are the uniformity of the microlenses and their mutual positioning accuracy. As a variation of these properties deteriorates the optical

performance in most applications, dedicated test instrumentation is needed.

In the following, we discuss different measurement instruments designed to characterize all the geometrical and optical parameters of microlenses. First, we introduce an optical noncontact profiler which can be used to determine the diameter and the sag (the maximum height of the spherical lenscap) of the surface profile microlenses. In addition, it allows us to quantify the surface roughness on the top of the microlenses. With a transmission Mach–Zehnder (MZ) interferometer the wave aberrations and the focal length of the microlenses can be measured, while a Twyman–Green (TG) interferometer can be used for the measurement of the surface deviations of a microlens from an ideal spherical shape and for the measurement of the radius of curvature of the surface.

Surface Profile Measurements

Optical noncontact profiler

Principle. The surface profile of a microlens can be measured with a Wyko NT2000 noncontact optical profiler (Figure 19a). This surface profiler system uses two approaches to measure a wide range of surface heights. The PSI mode allows measuring fairly smooth and continuous surfaces (0.1 nm < heights < 160 nm) while the vertical scanning interferometry (VSI) mode can measure rough surfaces and heights ranging between 160 nm and 2 mm. We only discuss here the vertical scanning mode because the principle on which PSI is based has been explained above. As shown in Figure 19b, this noncontact profiler is based on a Mirau interference microscope where a white light beam passes through a beamsplitter which reflects half of the incident beam to a reference surface and transmits the other half onto the sample. The light reflected from the sample and from the reference surface then recombine at the beamsplitter to form interference fringes. The system measures the degree of the fringe modulation or the fringe contrast. Because white light has a short coherence length, interference fringes are present only over a very shallow depth for each focus position. The coherence length can be described here as the optical path distance (OPD) between two wavefronts of an interferometric system for which the phase remains correlated. Or in other words it is a measure of the range of heights over which the instrument will be able to obtain measurable interference fringes. Fringe contrast at a single sample point reaches a peak as the sample is translated through focus. During a measurement, the reference arm containing the interferometric objective moves vertically to scan the surface at varying heights. A linearized PZT precisely controls this motion. As the system scans downward, an interference signal for each point on the surface is recorded. Finally, a series of advanced computer algorithms is used to demodulate the envelope of the fringe signal and to extract the surface information.

The vertical resolution is approximately 3 nm RMS for a single measurement. The lateral resolution is a function of the total magnification, of the cursor position accuracy, and of the detector array size.

Measuring the geometrical lens dimensions. This profiler provides us with information on geometrical dimensions of microlenses. For each microlens we can derive from the contour plot (Figure 20) the sag and the diameter of the microlenses along the x and the y direction. As an example, we show in Figure 21, the

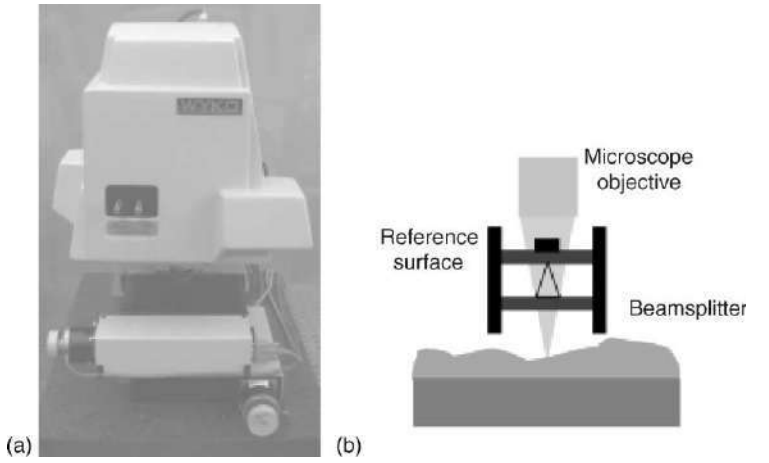

Figure 19 (a) Vertical scanning noncontact profiler (WYKO NT2000); (b) Working principle.

profiles of the cross-section of a 200 μm fused silica microlens ($NA = 0.05$; $D = 200$ μm) along the x and y directions.

The surface roughness on top of the microlens can also be derived from measurements with the noncontact profiler. As an example, we show the fringes on the vertex area (65×61 μm^2) of a 200 μm diameter lens as well as the average and RMS roughness on the vertex of the lens (Figure 22).

Measurement of the Optical Lens Performance

To investigate the optical properties of discrete microlenses one of the most accurate and reliable techniques is interferometry. It is well known that microlenses can be tested in transmitted light using instruments with a single pass geometry (e.g., MZ interferometer) as well as in reflected light interferometers using a double pass configuration (e.g., TG interferometer). The interferometers proposed in this section have been especially designed and constructed for microlens testing.

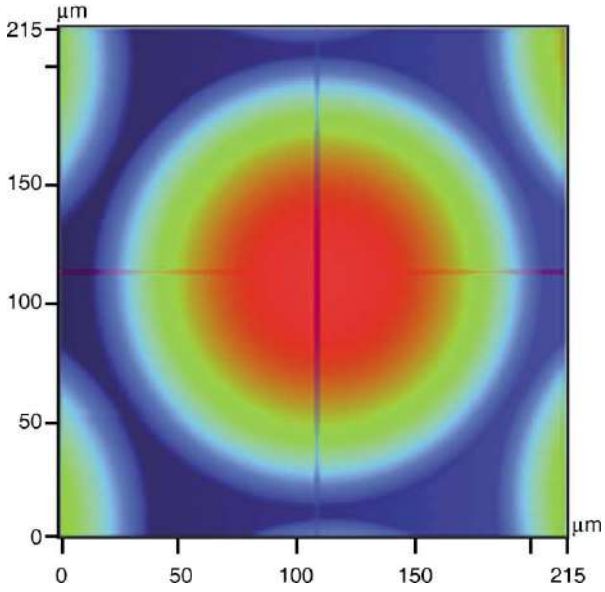

Figure 20 Contour plot of a 200 μm fused silica microlens as obtained with the noncontact profiler.

Mach–Zehnder (MZ) interferometer for transmitted light measurements

Principle. A MZ interferometer, constructed at the Erlangen–Nürnberg University, is used in this work for the measurement of microlenses in transmitted light. The interferometer is installed in a commercial Carl Zeiss Jena microscope, while fiber-optics is used for beam delivery. The scheme is shown in Figure 23. A polarizing beamsplitter (PBS) cube splits the beam of a HeNe laser while a halfwave plate at the entrance enables the balancing of the two interfering beams. The fibers are polarization preserving in order to keep the polarization in the two beams of the interferometer. The planewave of the object arm is either focused by an illumination microscope objective so that a spherical wave impinges onto the microlenses (this is the so-called spherical wave illumination) or the first microscope objective is removed so that a planewave illuminates the object (this is the so-called planewave illumination). By using a telescope consisting of the imaging microscope objective and an achromatic doublet, the expanded planewave then enters a second beamsplitter and is superimposed on a reference planewave generated by the second fiber and an expansion lens. This second beamsplitter is a nonpolarizing one. To force the two beams to interfere, the plane of polarization of the reference beam must be rotated by 90° which is achieved by twisting the fiber end accordingly. This solution has been chosen to avoid a polarizer in front of the CCD camera as it has been shown that the latter would decrease the optical quality. In addition one of the mirrors in the object

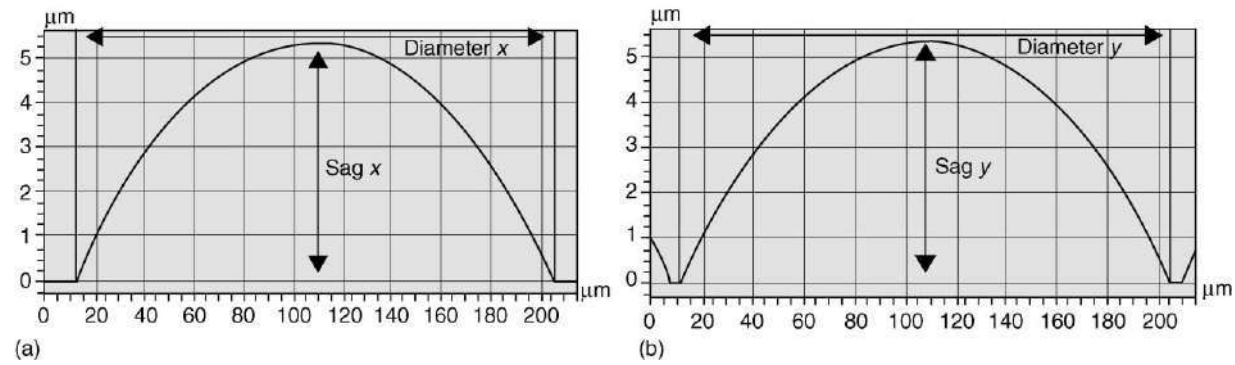

Figure 21 (a) 2D lens profile of the microlens shown in Figure 20 along the x direction and (b) along the y direction.

Roughness values	(nm)
Average roughness R_1	1.55
RMS roughness R_q	1.94

(a)

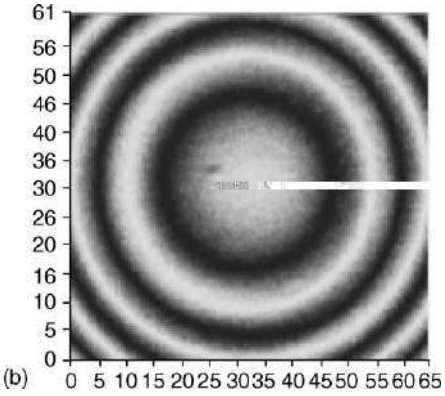

(b)

Figure 22 (a) Different roughness values on the vertex area of a 200 μm lens; (b) Fringes in this vertex area (65 × 61 μm^2).

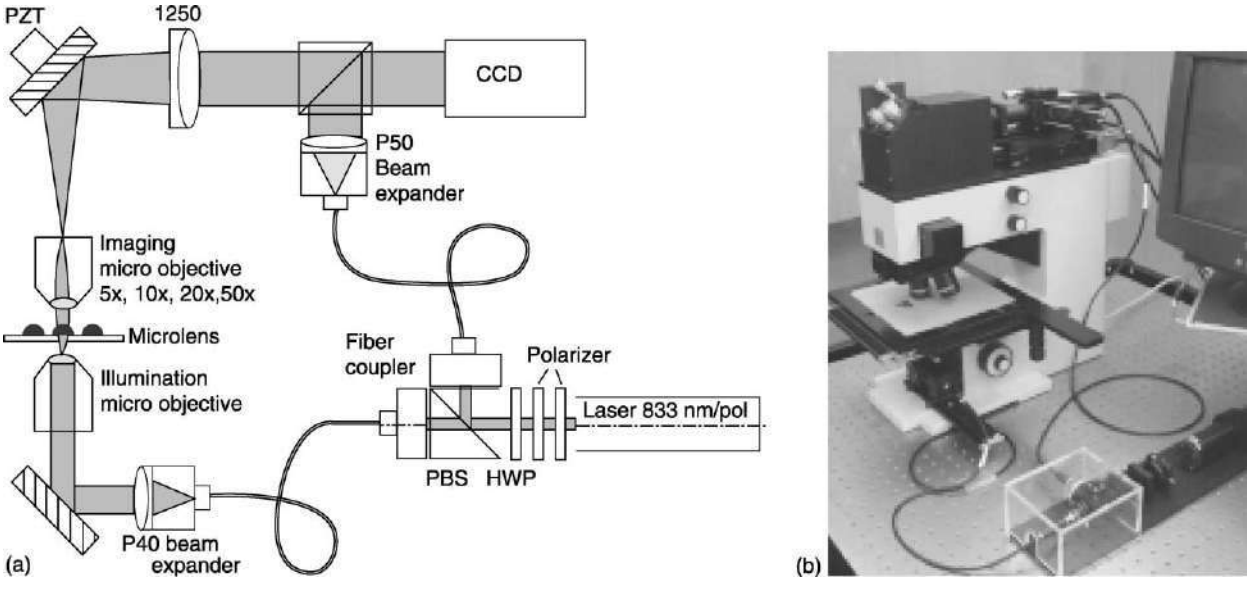

Figure 23 (a) Scheme and (b) picture of the Mach–Zehnder interferometer for measurements in transmitted light.

arm of the MZ is mounted on a PZT driver to enable phase shifting interferometry.

Focal length and array uniformity through plane-wave illumination. The MZ interferometer with planar wavefront illumination offers for slow lenses (NA < 0.1) the opportunity to measure also the total phase lag introduced by the lenses, as shown in Figure 24. This makes it possible to determine the focal length and hence the NA of the microlenses.

For certain applications of arrays of microlenses it is also important to know the uniformity of the lens array or at least of a part of the lens array. There are several different test methods (e.g., Smartt test, shearing test) which can provide a measure of the uniformity. One can even obtain the total phase retardation for a subsection of the microlens array (see Figure 25b).

Measuring the aberrations with spherical wave illumination. Wave aberrations of refractive microlenses can be determined with a spherical wavefront illumination (Figure 26). The measurement of the wave aberrations is carried out by using a PSI algorithm. This requires a measurement of the intensity in each point of the aperture for five different reference phase values. The result of a PSI evaluation is the phase distribution $\Phi(x, y)$ in the exit pupil:

$$\Phi(x, y) = \frac{2\pi}{\lambda}\psi(x, y) \qquad [50]$$

where $\psi(x, y)$ are the wave aberrations of the lens under test.

From these data, the software calculates the unwrapped phase distribution and by subtracting the tilt and defocus, the deviation from an ideal wavefront (phase/lambda) can be found. Fitting this

deviation to a fourth degree Zernike polynomial then gives the aberrations (phase/lambda). Besides the aberrations we also derive from the measured phase distribution other optical lens characteristics such as the modulation transfer function and the PSF.

Finally, we emphasise that the measurements should preferably be done in a cleanroom environment as the smallest dust particle on the lens surface decreases the optical performance.

Twyman–Green interferometer

Principle. For most applications microlenses have to feature a spherical or nearly spherical shape. The measurement of the deviations of the microlens surface from an ideal sphere is therefore very important. The basic principle comes down to a planewave which is focused by a high NA microscope objective in such a way that the focus coincides with the center of curvature of the microsphere under test. The light rays are reflected, and behind the microscope objective a nearly planewave carrying the surface deviations and misalignment errors, results. The vertex of the test surface is then imaged onto a CCD camera and the nearly plane object wave interferes with the plane reference wave. The evaluation here is also done with phase shifting interferometry by axially shifting the reference mirror with a piezo transducer.

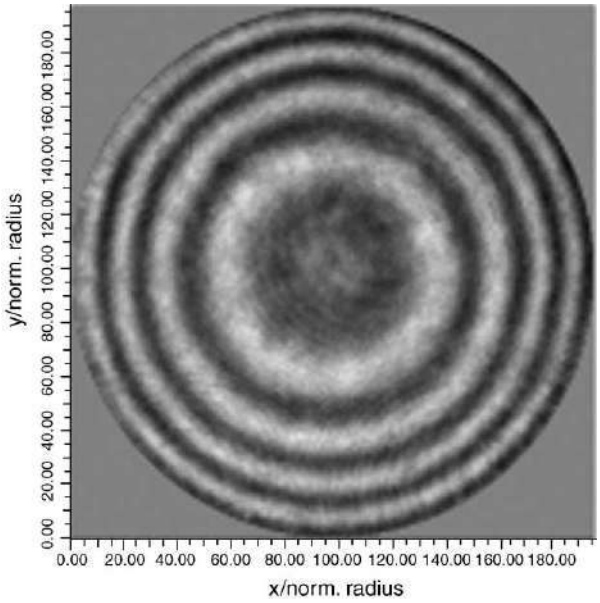

Figure 24 Interferogram of a 200 μm lens with an NA of 0.05 under planewave illumination.

The use of laser illumination allows us to obtain high contrast fringes. However, the high spatial and temporal coherence of the laser light also has important drawbacks, such as speckle noise and spurious fringes across the field. It is convenient to divide coherence into two classifications, temporal and spatial. The former relates directly to the finite bandwidth of the source, the latter to its finite extent in space. Therefore, polarization beamsplitter optics is commonly used to reduce the contribution of spurious reflections. The polarizing beamsplitter group additionally allows balancing the intensity between the object and the reference arm of the interferometer. Even so, the remaining scatter leads to poor and rough interference fringes. To further increase the fringe quality, the spatial coherence has to be reduced while the temporal coherence of the laser is preserved. This can be done by introducing a rotating scatterer in the light beam which gives a smoothing effect of the interference fringes.

Measurement of the deviations from sphericity. To determine the surface deviation of a microlens from a perfect sphere we make use of a TG

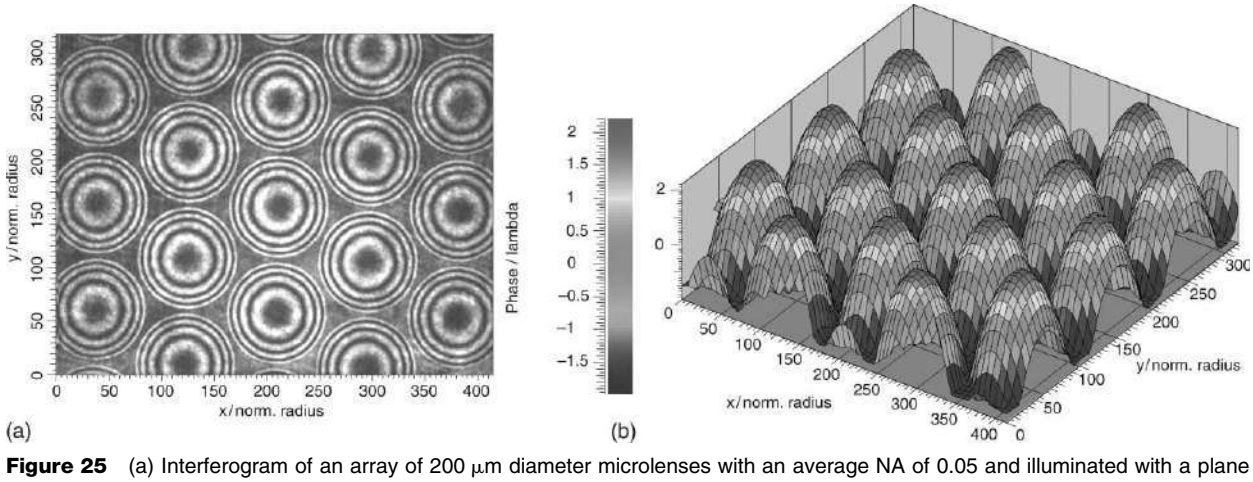

Figure 25 (a) Interferogram of an array of 200 μm diameter microlenses with an average NA of 0.05 and illuminated with a plane wavefront; (b) 3D representation of the phase lag (RMS 1.26λ, *PV* 4.37λ) for the same microlens array.

interferometer, constructed by the Erlangen-Nürnberg University (Figure 27). In this interferometer, the light of a HeNe laser is imaged onto a rotating scatterer by a microscope objective and the scattered light is collimated by an achromat. The collimated laser beam then passes a field stop which is sharply imaged by a 4f system onto the lens under test. The image of the field stop can be focused onto the surface under test by moving the field stop along the optical axis. Once the image of the field stop is sharply imaged onto the CCD array, one can translate the reference mirror until the reference image of this stop in its turn becomes sharp onto the CCD camera. Thus, two conditions have to be met to achieve high contrast fringes: the surface under test and the image of the field stop of both arms of the interferometer have to be sharp on the detector. The half wave plate in front of the PBS adjusts the splitting ratio at the PBS while the polarizer after the beamsplitter is necessary to force the two orthogonally polarized beams to interfere. Finally, the optics at the output of the interferometer gives a sharp image of the intermediate image of the lens surface onto the CCD camera. Analogous to the MZ interferometer, five phase images are grabbed and subjected to different algorithms, leading to the desired information.

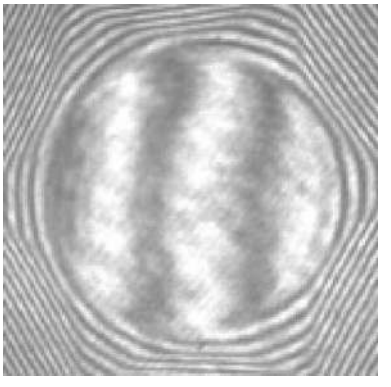

Figure 26 Interferogram of a microlens under spherical wavefront illumination: the resulting planewave is a measure for the wave aberrations.

A condensor objective transforms the radius of curvature of the spherical wavefront such that it matches the radius of curvature of the microlens. Hence, the wavefront impinges everywhere perpendicularly on the lens surface. This means that the sensitivity to surface deviations is independent on the surface position leading to high accuracy measurements. The TG interferometer allows us to test the sphericity of microlenses with diameters ranging

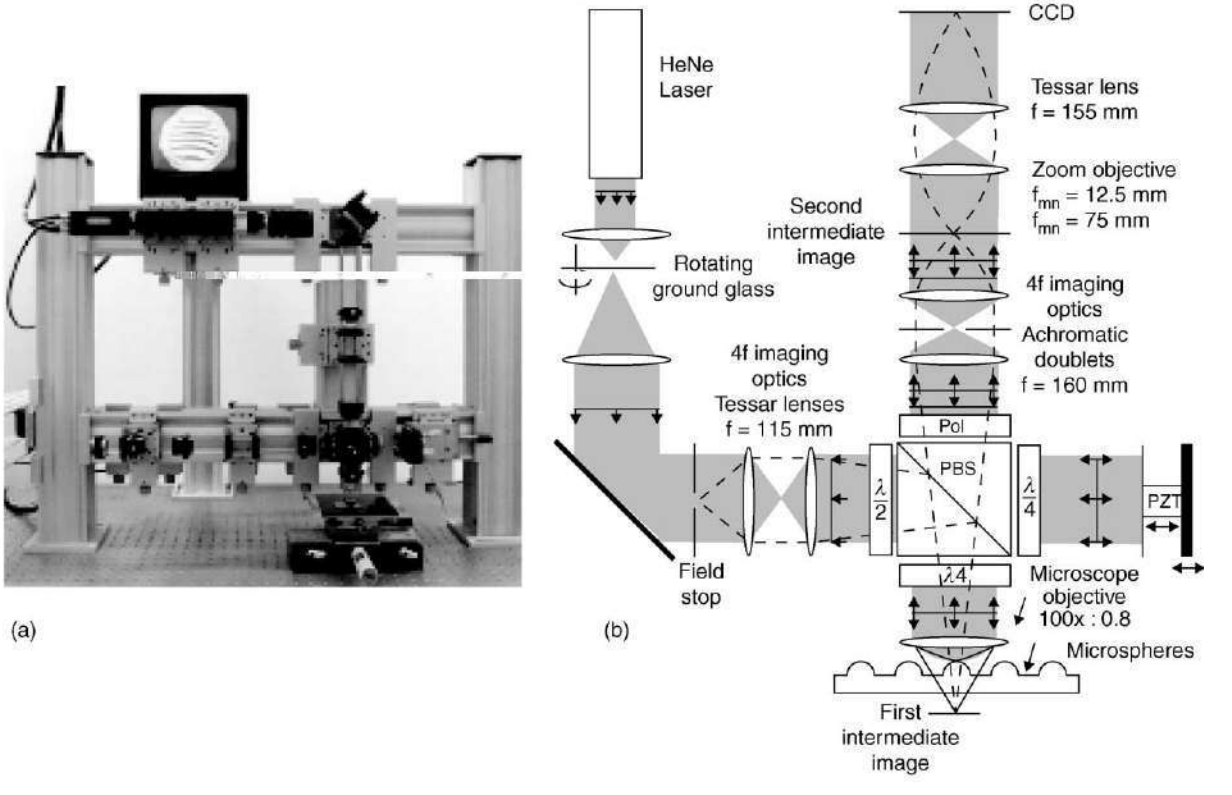

Figure 27 (a) Twyman–Green interferometer setup. (b) Detailed schematic overview of this interferometer.

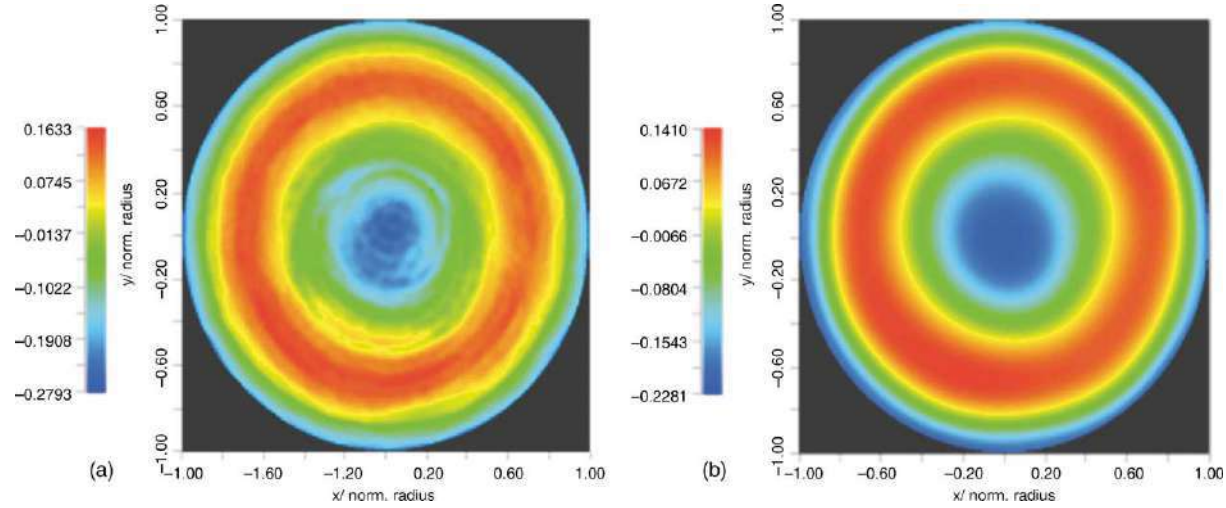

Figure 28 (a) Measured deviation from an ideal sphere (phase/lambda) for a 200 μm lens with an NA of 0.05 as obtained with a Twyman–Green interferometer. (b) Polynomial fitted surface deviation (phase/lambda) of this microlens.

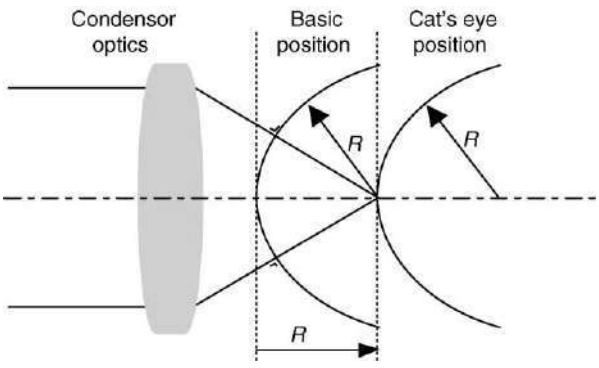

Figure 29 Necessary adjustments of the spherical microlens for radius measurements.

from 50 to 1000 μm and with a maximum radius of curvature of 1960 μm.

An example of a 200 μm fused silica microlens, with an NA of 0.05, is shown in Figure 28. The deviation from an ideal sphere can be derived from the unwrapped phase after subtracting the tilt and defocus. Fitting the deviation to a fourth degree Zernike polynomial then gives the surface deviation.

Measuring the radius of curvature. Using the TG interferometer it is also possible to determine the radius of curvature R of a microlens. To do this, the test object has to be shifted axially from the basic position, where the light rays impinge perpendicularly onto the surface and the focal length of the condensor objective coincides with the center of the sphere under test, to the so-called cat's eye position, where the focus of the impinging spherical wave is at the vertex of the sphere (Figure 29). For both positions straight, parallel, and equidistant interference fringes are obtained, only perturbed by aberrations due to the lens surface and the condensor objective. This forms the criterion for the adjustment of the two positions. The difference between these two positions then gives the radius of the microsphere. It should be mentioned that the cat's eye position causes an inversion of the object wavefront with respect to the reference wavefront, which results in a vanishing contrast with the partially coherent illumination. The latter can only be avoided by providing the necessary coherence. This implies that the rotating scatterer will have to be removed.

Summary of the Interferometric Instrumentation

Figure 30 summarizes the various instruments we discussed in this section: a noncontact optical profiler, a MZ and a TG interferometer. For each of these instruments, we highlight the measurands, the assumptions that are made, and the resulting optical lens characteristics. From this figure, it is clear that one can obtain a measure for the focal length and hence the numerical aperture and the focal number of a refractive microlens, with all three instruments. However, only the focal length obtained with the MZ interferometer is a direct measurement. The focal length obtained with the optical profiler is derived from the sag of the lens under the assumption of an ideal sphere while the radius of curvature measured with the TG interferometer allows to calculate the focal length as well. Nevertheless, the three values make an interesting comparison possible.

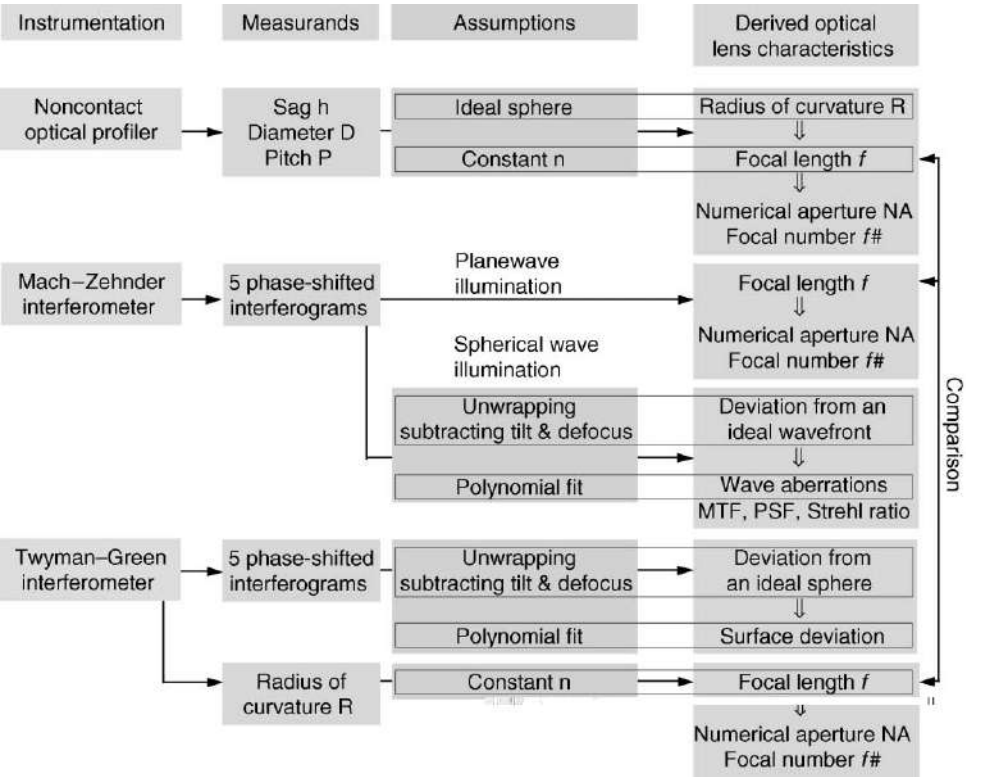

Figure 30 Overview of the instruments that we will use to quantitatively characterize the refractive microlenses: the measurands and the resulting optical lens characteristics.

Conclusions

In this article we introduced geometrical optics to discuss the imaging properties and the aberrations of refractive lenses, while wave optics was used to explain the effects of diffraction on the imaging properties of lenses. Next we put forward some commonly used quality criteria to assess lens performance and we pointed out the unique properties of microlenses. Finally, interference was briefly introduced to clarify the working principle of various instruments to quantitatively characterize refractive microlenses and microlens arrays: a noncontact profiler, a transmission MZ interferometer and a TG interferometer.

We can conclude that micro-optics is a key technology with many applications in various areas of tomorrow's information society. It enables new functionalities and applications previously unachievable with conventional optics and improves product performance, stability and robustness while reducing cost, volume, and weight. While in conventional optics discrete lenses are mounted to form optical systems using mechanical alignment equipment, for microlenses new integration approaches are available. For most of the applications, the integration of micro-optical components into fully functional systems is very important.

Thus micro-optics is a generic technology that allows the manipulation of light with micron and submicron scale structures and components. Micro-optics is therefore the technology to help provide the necessary tools to interface the macroscopic world we live in with the microscopic world of opto- and nano-electronic data processing. Micro-optic processing and fabrication technologies can be adapted to a wide variety of specialty optical materials such as plastics, polymers, semiconductors, sol-gels, and doped glasses. The application of these modified materials permits the widening of the field of photonic functionalities to virtually any region of the optical spectrum and beyond, such that a wealth of novel optical and photonic applications can be realized. Micro-optic processing technologies can be made compatible with scalable industrial mass-manufacturing and replication techniques.

Acknowledgment

This research has been made possible through the financial support of the Fund for Scientific Research-Flanders (FWO-Vlaanderen), the IAP Photon

Network, the Bilateral Scientific and Technological Cooperation Flanders-Poland, the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the IWT GBOU project 'Generic Technologies for Plastic Photonics', the Research Council (OZR) of the Vrije Universiteit Brussel and the EC Network of Excellence NEMO.

List of Units and Nomenclature

[AB]	optical path length
c	speed of light
d	lens thickness
D	lens aperture diameter
$\mathbf{E}$	electricfield
f	focal length
f_{BFL}	back focal length
f_{FFL}	front focal length
$f\#$	focal number
F_{i}	image focal point
F_o	object focal point
h_1	position of the first principal point relative to the vertex V1
h_2	position of the second principal point relative to the vertex V2
h'	deviation from the optical axis
H_1	first principal point
H_2	second principal point
I	intensity
k	wave number in vacuum
n	index of refraction
n_{i}	index of refraction of medium i
O	object point
P	arbitrary field point
P	image point
P	lens pitch
$\wp$	spherical surface power
r	aperture of the refracting surface
R	radius of curvature
$R(x, y)$	real wavefront
S	axial object point
S	Strehl ratio
s_0	object distance
s_i	image distance
V	lens vertex
V	visibility or fringe contrast
$W(x, y)$	spherical wavefront in the Gaussian approximation
y_{o}	object height
y_{i}	image height
z_{R}	Rayleigh range
α	angular semi-aperture
θ	symmetry around the optical axis
λ	wavelength of light
σ	standard deviation
φ_1	angle between the incident ray and the normal to the flat surface
φ_2	angle between the refracted ray and the normal to the flat surface
$\Phi_{\mathrm{obj}} - \Phi_{\mathrm{ref}}$	phase difference
$\Phi(x, y)$	phase distribution
$\psi(x, y)$	wavefront aberration
ω	angular frequency
Ω	aperture in a plane opaque screen
BFL	Back focal length
CCD	Charge coupled device
FF	Far-field
FFL	Front focal length
MTF	Modulation transfer function
MZ	Mach–Zehnder
NA	Numerical aperture
NF	Near-field
OPD	Optical path difference
OTF	Optical transfer function
PBS	Polarizing beamsplitter
PP	Principal plane
PSF	Point spread function
PSI	Phase shifting interferometry
PTF	Phase transfer function
PZT	Piezo-electric transducer
RMS	Root mean square
SA	Spherical aberration
VSI	Vertical scanning interferometry

See also

Diffraction: Fraunhofer Diffraction. **Diffractive Systems:** Aberration Correction with Diffractive Elements. **Geometrical Optics:** Aberrations; Lenses and Mirrors. **Interferometry:** Overview; Phase Measurement Interferometry.

Further Reading

Born M and Wolf E (1999) *Principles of Optics*, 7th edn. Cambridge, UK: Cambridge University Press.

Hecht E (1987) *Optics*. San Francisco, CA: Addison-Wesley Publishing Company.

Herzig HP (1997) *Micro-Optics: Elements, Systems and Applications*. London: Taylor & Francis.

Jenkins FA and White HE (1981) *Fundamentals of Optics*, 4th edn. New York: McGraw-Hill.

Malacara D and Malacara Z (1994) *Handbook of Lens Design*. New York: Marcel Dekker.

Malacara D, Servin M and Malacara Z (1998) *Interferogram Analysis for Optical Testing*. New York: Marcel Dekker.

Mouroulis P and Macdonald J (1997) *Geometrical Optics and Optical Design*. New York: Oxford University Press.
Pedrotti FL and Pedrotti LS (1993) *Introduction to Optics*, 2nd edn. Englewood Cliffs: Prentice-Hall International.
Shannon RR (1997) *The Art and Science of Optical Design*. Cambridge, UK: Cambridge University Press.
Siegman AE (1986) *Lasers*. Sausalito: University Science Books.
Sinzinger S and Jahns J (1999) *Microoptics*. Weinheim, Germany: Wiley-VCH GmbH.
Smith WJ (1997) *Practical Optical System Layout and Use of Stock Lenses*. New York: McGraw-Hill.
Smith W (2002) Engineering an optical system. *SPIE OE Magazine* 7: 49–54.
Young HD (1992) *Physics*, 8th edn. San Francisco, CA: Addison-Wesley Publishing Company.

OPTICAL PARAMETRIC DEVICES

Contents

Overview
Optical Parametric Oscillators (Continuous Wave)
Optical Parametric Oscillators (Pulsed)

Overview

B J Orr, Macquarie University, Sydney, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Many forms of optical parametric device (OPD) play significant roles as three-wave nonlinear-optical sources of tunable coherent light for laser-based science and technology. Any such device will fall into one of three categories, in increasing order of sophistication: optical parametric generator (OPG); optical parametric amplifier (OPA); and optical parametric oscillator (OPO). This article concentrates largely on nanosecond-pulsed OPOs, with particular emphasis on designs that facilitate their wavelength control and their optical power- or energy-efficiency. The two other important temporal manifestations of OPOs (ultra-fast pulsed and continuous-wave) are briefly outlined. OPDs serve as versatile sources of tunable, intense, coherent radiation covering the infrared, visible, or ultraviolet spectral regions. Their most obvious uses are associated with laser wavelength-conversion and/or tunability, resulting in a variety of spectroscopic applications. Four-wave nonlinear-optical parametric contributions, such as those involved in stimulated Raman scattering and highly nonlinear optical fibers, are also briefly considered.

What is an Optical Parametric Device?

An optical parametric device (OPD) is a source of coherent, laser-like radiation that is typically intense and tunable over a wide range of wavelengths. Its mechanism invariably depends on nonlinear optics, usually through a three-wave mixing process mediated by the nonlinear-optical susceptibility $\chi^{(2)}$ in a noncentrosymmetric crystalline medium. Three forms of OPD are illustrated in Figure 1, and also illustrated is a closely related (but distinct) nonlinear-optical device. Coherent light waves are represented by arrows, with their associated optical frequency ω_j and wave vector $\mathbf{k}_j$ (which are defined below). In Figure 1, input and output waves are shown as arrows on the left and right, respectively, with their breadth intended to indicate typical relative intensities.

An optical parametric generator (OPG) is the simplest form of OPD, with a single input wave (pump, P) from a laser generating two coherent output waves (signal, S; idler, I; $\omega_S \geq \omega_I$), as depicted in Figure 1a. The nonlinear-optical process itself is initiated by low-intensity parametric spontaneous emission/noise/fluorescence processes that occur naturally, effectively 'splitting' a pump photon into two new photons. Once a signal and/or idler wave has been generated, it can be coherently amplified by passing it through an optical parametric amplifier (OPA) together with input pump radiation, as depicted in Figure 1b. A further order of sophistication is reached in an optical parametric oscillator (OPO), as depicted in Figure 1c, where the

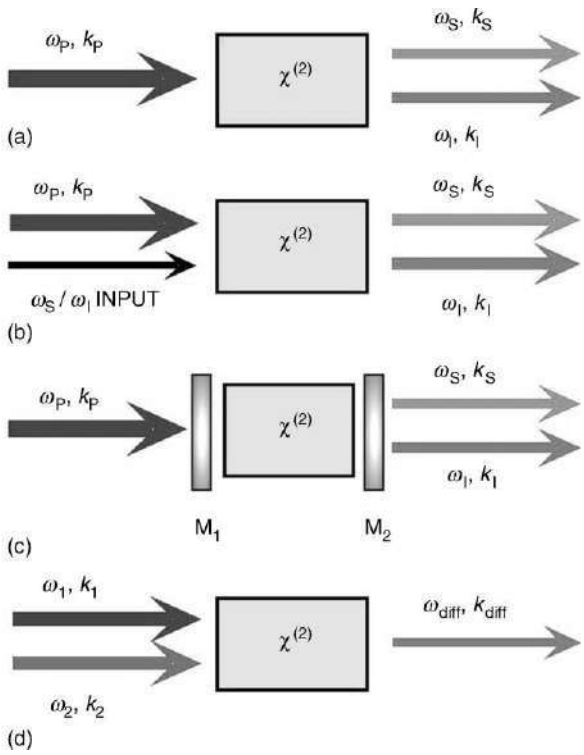

Figure 1 Schematic diagrams of three forms of optical parametric device (OPD): (a) optical parametric generator (OPG); (b) optical parametric amplifier (OPA); (c) optical parametric oscillator (OPO). Note that, by convention, optical frequencies of the signal (S) and idler (I) output waves are defined such that $\omega_S \geq \omega_I$. Also shown is a fourth closely related device: (d) difference-frequency generator (DFG). Nonlinear-optical media are denoted by their susceptibility $\chi^{(2)}$. Arrows are used to represent input and output waves, together with their optical frequencies ω_j and wave vectors $\mathbf{k}_j$. An OPO requires an optical resonator, comprising at least two aligned reflectors (M_1, M_2).

functions of an OPG and an OPA are combined by multipassing the optical waves involved inside a resonant optical cavity, formed by two or more appropriately aligned reflectors (M_1, M_2).

A difference frequency generator (DFG), as depicted in Figure 1d, is not an OPD, although the DFG source term is central to the nonlinear-optical mechanism of an OPD. In a DFG, two input waves (labeled 1 and 2) are combined coherently to generate a third output wave at the difference frequency of the two input waves: there are now two relatively high-power driving waves, rather than one, and photons are effectively 'joined', rather than 'split' as in the OPD case. Nevertheless, the outcome and utility of a DFG can be similar to that of an OPD. For instance, if coherent radiation is required at a particular infrared (IR) wavelength, it can be generated either as the idler wave of an OPD, with frequency $\omega_I = (\omega_P - \omega_S)$ *or* as the output wave of a DFG, with frequency $\omega_3 = |\omega_1 - \omega_2|$.

Many of the desirable attributes of OPDs in general, and OPOs in particular, arise from the fact that any such instrument is a nonlinear-optical device, not a laser. This allows versatile, flexible design features, such as methods of temporal and wavelength control that are not possible with lasers. The latter generally depend on population inversion of an optical gain medium, with associated optical lifetime and saturation limitations. On the other hand, optical parametric gain, oscillation and amplification are more amenable to modular system design because (as is explained below) they depend on nonlinear-optical coefficients and phase-matching conditions.

Nonlinear-Optical Origins of OPDs

Nonlinear optics involves interaction of a number ($\sigma > 2$) of optical waves in a medium with a nonlinear-optical susceptibility tensor $\chi^{(\sigma-1)}$. For inelastic optical processes, these waves (each with an angular frequency $\omega_1, \omega_2, \ldots, \omega_\sigma$) obey two conservation conditions, one for energy (or frequency):

$$\omega_1 + \omega_2 + \ldots + \omega_\sigma = 0 \qquad [1]$$

The other conservation condition (expressed in terms of wave vectors $\mathbf{k}_i$, with $i = 1, 2, \ldots, \sigma$, that have magnitudes $k_i = n_i\omega_i/c = 2\pi n_i/\lambda_i$, where n_i is the refractive index at vacuum wavelength λ_i and c is the speed of light) is for momentum:

$$\mathbf{k}_1 + \mathbf{k}_2 + \ldots + \mathbf{k}_\sigma + \Delta\mathbf{k} = 0 \qquad [2]$$

where $\Delta\mathbf{k}$ is the phase-mismatch between the σ interacting waves. Each frequency component and wavevector is ascribed a positive or negative sign, according to their phase relationships. Equation [2] defines phase-matching conditions in which $\Delta\mathbf{k}$ must be minimized to optimize the efficiency of the nonlinear-optical process of interest.

We shall consider below the two specific three-wave nonlinear-optical processes that are particularly relevant to this article, namely, those for an OPD (i.e., OPG, OPA, or OPO) or a DFG. Each of these is mediated by the nonlinear-optical susceptibility tensor $\chi^{(2)}$, which is nonzero in a crystalline medium only if it is noncentrosymmetric. Many such crystals are available. For example, lithium niobate ($LiNbO_3$) has been popular since the early days of pulsed tunable OPOs. Subsequent interest and activity in OPD technology have been stimulated by the availability of new nonlinear-optical materials such as BBO (β-barium borate, BaB_2O_4) and KTP (potassium titanyl phosphate, $KTiOPO_4$).

Additional recent impetus has come from quasi-phase-matched (QPM) nonlinear-optical media, such as periodically poled lithium niobate (PPLN) and PPKTP, tailored for specific wavelengths by periodic optical structuring of domains. QPM media offer compact, efficient, low-threshold alternatives to conventional birefringently phase-matched (BPM) media. Characteristics of BPM and QPM nonlinear-optical crystalline media are accessible *via* Arlee Smith's SNLO software package, accessible by internet at www.sandia.gov/imrl/XWEB1128/xxtal.htm.

For a three-wave OPD, which is of principal interest in this article, the energy and momentum conservation conditions of eqns [1] and [2] become:

$$\omega_P - \omega_S - \omega_I = 0, \qquad \mathbf{k}_P - \mathbf{k}_S - \mathbf{k}_I - \Delta\mathbf{k} = 0 \qquad [3]$$

where a laser input wave ('pump', frequency ω_P, wave vector $\mathbf{k}_P$) yields two coherent output waves ('signal', frequency ω_S, wave vector $\mathbf{k}_S$; 'idler', frequency ω_I, wave vector $\mathbf{k}_I$; $\omega_S \geq \omega_I$). Note that the idler frequency ω_I equals the difference $|\omega_P - \omega_S|$ between pump and signal frequencies. Equation [3] should be viewed in the context of sections (a–c) of Figure 1.

Equation [2] and the second half of eqn [3] apply strictly only to the conventional case of BPM media. In such media, the phase-matching condition $\Delta\mathbf{k} \approx 0$ is achieved by adjusting the angle and/or temperature of a birefringent nonlinear-optical crystal *via* its ordinary- and extraordinary-ray refractive indices. Such adjustments are used to optimize OPD conversion efficiency and thereby control the output signal and idler wavelengths, λ_S and λ_I. If it is assumed that the three waves are collinear and that $\Delta\mathbf{k}$ is exactly zero, then the signal frequency/wavelength is given simply in terms of the pump frequency/wavelength and the refractive indices n_i $(i = P, S, I)$ as:

$$\omega_S = \omega_P(n_P - n_I)/(n_S - n_I); \qquad \lambda_S = \lambda_P(n_S - n_I)/(n_P - n_I) \qquad [4]$$

An alternative to birefringent phase matching is the quasi-phase-matching approach first suggested by Nobel Laureate Nicolaas Bloembergen and co-workers in 1962. This approach has been realized experimentally only in the last few years, *via* QPM nonlinear-optical media such as PPLN or PPKTP. Equation [2] and the second half of eqn [3] need to be slightly modified to include QPM grating contributions.

In the corresponding case of a DFG (which, we repeat, is not an OPD), the energy and momentum conservation conditions of eqns [1] and [2] become:

$$|\omega_1 - \omega_2| - \omega_{\text{diff}} = 0, \qquad \mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_{\text{diff}} - \Delta\mathbf{k} = 0 \qquad [5]$$

where two laser input waves (frequencies ω_1, ω_2; wave vectors $\mathbf{k}_1$, $\mathbf{k}_2$) yield one coherent output wave at the difference frequency $\omega_{\text{diff}} = |\omega_1 - \omega_2|$ (with wave vector $\mathbf{k}_{\text{diff}}$), as depicted in Figure 1d. As above, phase matching is defined by $\Delta\mathbf{k} \approx 0$ for BPM media and, in the case of QPM media, there is an additional grating contribution in the second half of eqn [5].

Finally, we mention briefly two other important forms of nonlinear-optical device arising from four-wave mixing processes that are mediated by the nonlinear-optical susceptibility $\chi^{(3)}$ which can be non-zero, even in isotropic or centrosymmetric media such as gases, liquids, optical fibers, and all classes of crystal. Optical parametric processes of this type contribute to stimulated Raman scattering (SRS), involving an optical medium with Raman-active resonance frequencies ω_R that coincide with the difference between two optical frequencies. This is particularly useful in providing a relatively straightforward source of coherent radiation, Raman-shifted at discrete intervals from the frequency ω_L of an input pump laser (either tunable or fixed-wavelength). Four-wave parametric processes are involved, such as that in eqn [1] with $|\omega_1| = |\omega_3| = \omega_L, |\omega_2| = \omega_L - \omega_R$, and $|\omega_4| = |\omega_1| + |\omega_3| - |\omega_2| = \omega_L + \omega_R$, where ω_R is the characteristic Raman frequency of the nonlinear-optical medium used in the Raman shifter (e.g., high-pressure hydrogen, deuterium or methane gas). These Raman-shifted intervals, both added to (anti-Stokes) and subtracted from (Stokes) the laser frequency ω_L, are integer multiples of ω_R. Other nonlinear-optical Raman parametric processes give rise to various forms of nonlinear Raman spectroscopy, such as CARS (coherent anti-Stokes Raman spectroscopy), and to Raman fiber-optical amplifiers, used in optical telecommunications.

Another promising area of developing OPD technology entailing $\chi^{(3)}$ nonlinearities concerns OPGs, OPAs, and OPOs based on highly nonlinear optical fibers, with either pulsed or cw pump lasers. Such processes use two pump waves (P) to generate tunable signal (S) and idler (I) output waves, such that $\omega_I = 2\omega_P - \omega_S$.

Nanosecond-Pulsed OPOs – Design and Wavelength Control

Here we consider pulsed OPOs that operate on nanosecond (ns, 10^{-9} s) time-scales, and design

Table 1 Some elements of OPO applications to spectroscopy

Key properties	*Options*	*Comments on instrument/technique*
Wavelength range: what forms of spectra need to be measured?	UV/visible (0.2–0.7 μm) Near IR (0.7–4.0 μm) Longwave IR (>4.0 μm)	Many nonlinear-optical OPO materials are available, but less well developed for longwave IR applications
OPO phase matching: BPM or QPM?	Birefringent phase matching (BPM) Quasi-phase matching (QPM)	BPM is well established and preferred for high-power operation. Various low-threshold QPM media are now available for both pulsed and cw OPOs
Temporal: continuous-wave (cw) or pulsed?	Pulsed for power and timing Cw for narrowest bandwidth	Ultrafast (ps, fs) OPO output pulses are relatively broadband, owing to the Fourier transform limit
Optical bandwidth: broad or narrow?	Broadband (free-running) Single longitudinal mode	The Fourier transform limit is 44 MHz (0.0015 cm^{-1}) for an ideal 10 ns pulse Cw OPOs offer even lower $\Delta\nu$. Ultrafast-pulsed OPOs are inevitably broadband: e.g., $\Delta\nu \geq 0.44$ THz ($\geq$15 cm^{-1}) for a 1 ps pulse
Mode of recording of spectra: continuous tuning or multiplex?	Scan narrowband signal/idler OPO output wavelength Free-running OPOs operate broadband in multiplex case	Wavelength control yields continuously tuned narrowband spectra. Multiplex spectra use dispersed detection or multi-wavelength spectroscopic tailoring
Wavelength control: intracavity elements or injection seeding?	Intracavity gratings or étalons Injection seeding of signal or idler by a tunable low-power coherent light source	Intracavity-element designs yield broad tunability but can be complicated. Injection seeding facilitates narrowband, mode-hop-free spectra and tailored multi-wavelength experiments

The above options determine ways to use OPOs for spectroscopic applications, such as: (1) linear absorption (e.g., with a multipass cell); (2) cavity ringdown (CRD) absorption spectra; (3) high-resolution spectra ($\Delta\nu \approx$ MHz or kHz); (4) nonlinear-optical (e.g., coherent Raman); (5) atmospheric remote sensing (e.g., DIAL); (6) fast (μs, ns) and ultrafast (ps, fs) processes.

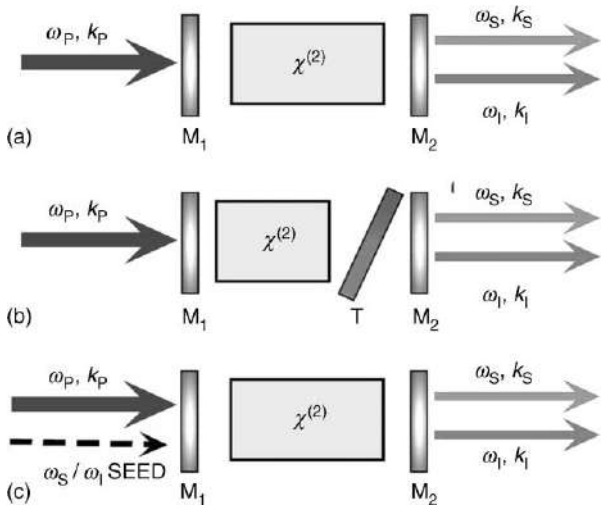

Figure 2 Schematic diagrams of three forms of optical parametric oscillator: (a) free-running OPO (with no active wavelength control), similar to **Figure 1c**; (b) OPO with an intracavity tuning element (T); (c) injection-seeded OPO.

features that make them fit for one of the principal applications of OPDs, namely, spectroscopy. Options that determine ways to use OPOs for spectroscopic applications are summarized in Table 1. Some design and wavelength-control features used in ns-pulsed OPOs are depicted schematically in Figure 2. Some of these applications and features are shared with other forms of OPO, notably the ultrafast-pulsed OPOs and cw OPOs that are considered much more briefly in later sections.

The first OPO, demonstrated by Giordmaine and Miller in 1965, was ns-pulsed; it was based on $LiNbO_3$, tunable over the range ~0.96–1.16 μm, spanning a signal- and idler-wavelength range of ±0.1 μm on either side of the degeneracy point defined by the 529 nm pump radiation (from a frequency-doubled, Q-switched Nd:$CaWO_4$ laser). This advance occurred soon after lasers were discovered and the potential of nonlinear optics had been realized through harmonic-generation processes such as frequency doubling. Pioneering work by Stephen Harris, Robert Byer and their co-workers at Stanford University soon established ns-pulsed OPOs as practical sources of tunable coherent light for significant applications such as spectroscopic sensing of chemical processes, in industrial or environmental diagnostics and in basic science. Such ns-pulsed OPOs had an impressive impact in the 1970s, followed by sporadic progress in much of the 1980s, and a dramatic resurgence of interest and activity since the late 1980s.

This irregular rate of progress is attributable to the availability of suitable high-quality nonlinear-optical crystals (as mentioned above), as well as high-performance pump lasers and advanced tunable OPO system designs.

In a typical ns-pulsed OPO, the pump laser delivers sufficient optical energy to enable parametric gain to build up from noise during the pump–pulse duration, exceeding the threshold for oscillation. To maximize gain, the parametric (signal and/or idler) waves are amplified by multipassing them during the pump pulse in an optical resonator, as depicted in Figures 1c and 2. Light travels ~3 m during a 10 ns pump pulse, so that the parametric waves can make ~15 round trips of a 10 cm linear OPO cavity.

A simple free-running OPO, comprising an optical cavity with input and output mirrors M_1 and M_2 but with no wavelength-selective elements, is depicted in Figure 2a. The output radiation from such an OPO generally has a relatively broad optical bandwidth – typically ~5 cm^{-1} (~150 GHz) or more – depending on several factors: refractivity, dispersion, and absorption of the OPO medium; wavelengths λ_P, λ_S, and λ_I; type of phase matching (BPM or QPM, whether collinear or not); crystal dimensions and orientation; cavity reflectivity and effective number of passes of the resonated wave; optical bandwidth, divergence, pulse duration and pulse energy of the pump radiation. Free-running pulsed OPOs represent one extreme of operational simplicity, yielding relatively broadband output radiation suitable for low-resolution or multiplex spectroscopy. Additional OPO wavelength-control measures are usually necessary for higher-resolution spectroscopic applications.

At the other extreme of operational complexity, intracavity wavelength-selective elements, such as gratings and/or étalons, provide a traditional way to control OPO output wavelength and to achieve narrow optical bandwidths. One such approach is depicted in Figure 2b, where a tuning element T (in this case, a tilted étalon or filter; in other designs, an intracavity diffraction grating replacing the output cavity mirror M_2) is inserted in the cavity. Such approaches were used in early pulsed $LiNbO_3$ OPO designs that were continuously tunable in the near IR with an optical bandwidth of ~0.1 cm^{-1} (~3 GHz), but these were often difficult to operate and were damage-prone (with intracavity losses from gratings and étalons causing the operating threshold to approach the damage threshold of OPO nonlinear-optical materials such as $LiNbO_3$).

Continuous narrowband tunability, preferably on a single-longitudinal-mode (SLM) basis, is a desirable performance characteristic of pulsed OPOs amenable to high-resolution spectroscopy. A commercially viable approach to this ideal was achieved by Walter Bosenberg and Dean Guyer in their advanced KTP OPO/OPA system, which includes a master OPO stage with a grazing-incidence grating. This is continuously tunable under computer control in the near IR (1.3–4 μm) with narrow optical bandwidth – ~0.02 cm^{-1} (~600 MHz) or better. Such a high-performance tunable coherent IR source has been used for various forms of laser spectroscopy (cavity ringdown, degenerate four-wave mixing, long-path absorption, etc.), including investigations of chemically reactive media, combustion diagnostics, and studies of processes in molecular beams.

A popular alternative approach to OPO wavelength control is injection seeding by a low-power tunable coherent source such as a tunable diode laser (TDL). This approach is depicted in Figure 3; in practice, a ring cavity is often used to avoid

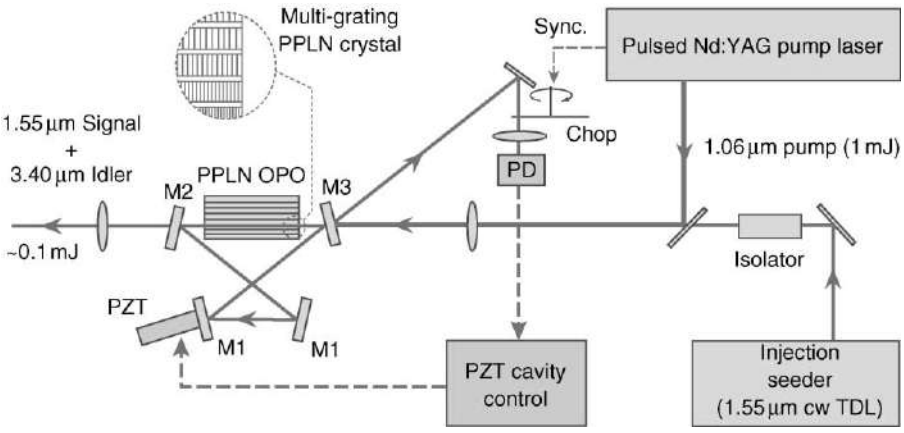

Figure 3 Schematic diagram of an injection-seeded tunable PPLN OPO system with active cavity control, developed in the author's laboratory. Legend: PD = photodetector, PZT = piezoelectric translator, TDL = tunable diode laser, M_{1-3} = cavity reflectors; the inset shows the QPM multi-grating structure of the PPLN nonlinear-optical crystal.

feedback from the OPO to the seed laser. A significant advantage of injection seeding is that OPO construction is simplified by putting the wavelength-control function into a module that is effectively separate from the optical generation and amplification functions. For example, a high-performance, TDL-seeded PPLN OPO system developed in the author's laboratory is illustrated in Figure 3. It features a multigrating QPM gain medium that is temperature-adjustable up to 250°C, and is pumped at a wavelength of 1.064 μm by a ns-pulsed, SLM Nd:YAG laser. An active cavity control scheme devised by Yabai He enables the SLM near-IR signal and idler outputs from this OPO (at ~1.5 μm and ~3.5 μm, respectively) to be tuned continuously without mode hops over optical frequency ranges of ~250 cm^{-1} (~7.5 THz) with a SLM optical bandwidth of <0.004 cm^{-1} (<120 MHz), as has been confirmed in assorted applications to molecular spectroscopy. The PPLN OPO can by itself generate only moderate output pulse energy (~0.1 mJ, approximately 70% of which is signal radiation), since the 1.064 μm Nd:YAG pump laser pulse energy is limited to ~1 mJ to avoid the optical damage threshold of the PPLN crystal. For higher-power applications, a further portion of the Nd:YAG laser energy (~200 mJ/pulse) can be used to pump an OPA stage based on BPM $LiNbO_3$, thereby generating SLM signal and idler output pulse energies of ~1.3 mJ and ~0.7 mJ, respectively.

An alternative approach to injection-seeded tuning of ns-pulsed OPOs (used extensively by the author's group in the context of BPM media such as BBO and $LiNbO_3$) entails passive control of the OPO cavity by slightly misaligning one of its reflectors. This facilitates continuous tuning of the injection-seeded OPO signal and idler outputs by decreasing the effective finesse of the OPO cavity, so that it is then not necessary to lock the OPO cavity length to the seed wavelength. This mode of operation is simpler optically and electronically than active control of the OPO cavity. The method depends on the OPO cavity having a series of high-order transverse modes to smooth out the sharp, widely separated resonances that occur when the OPO cavity is well aligned; a resulting disadvantage is that the multiple transverse modes tend to cause some degradation of output beam quality. Nevertheless, this approach has proved useful for many applications of tunable OPOs, with seeding by either pulsed dye lasers or SLM TDLs.

This passive, misaligned-cavity approach to injection seeding of ns-pulsed OPOs is particularly well suited for spectroscopic applications requiring a coherent source that simultaneously generates two or more adjustable output wavelengths. Such a TDL-seeded, dual-wavelength BBO OPO has been used in the author's laboratory for coherent-Raman spectroscopic measurements of nitrogen (N_2) in furnace air; the relative intensity of the two spectroscopic features provides an instantaneous, turbulence-immune estimate of the temperature of gas in the furnace, within a single shot of the OPO's pump laser (i.e., within ~10 ns). This is an example of so-called spectroscopic tailoring, in which an OPO generates a structured set of two or more discrete output wavelengths, each of which is set to be on- or off-resonance with characteristic features in spectra of molecular target species of interest.

Dual-wavelength injection seeding of ns-pulsed OPOs is relevant to atmospheric remote sensing by DIAL (differential absorption lidar), with TDL-seeded OPO output switched between on- and off-resonance IR wavelengths on alternate shots of the pump laser; range-resolved measurements have been made of key atmospheric species such as methane (CH_4) and water vapor (H_2O). Figure 4 illustrates a proposed extension of the spectroscopic tailoring concept, applicable to dual- and multi-wavelength remote sensing applications by long-path absorption or DIAL. This multiplex system comprises a multi-wavelength passive-cavity pulsed OPO, injection-seeded by a set of single-mode TDLs and a computer-controlled fiber-optic switch to code and decode the resulting multiplex spectroscopic signals.

A significant frontier for narrowband pulsed OPOs entails extending the spectral output range beyond the present readily attainable near-IR limit of ~4 μm (as provided by materials such as BBO, KTP, $LiNbO_3$, PPLN, and PPKTP). This extension yields high-performance tunable coherent spectroscopic sources at longwave-IR wavelengths where many of the strongest molecular fundamental IR absorption bands are located. OPOs that target the longwave IR depend on availability of suitable nonlinear-optical materials and coherent pump sources at sufficiently long wavelengths (typically >2 μm). One of the most promising longwave-IR OPO materials is zinc germanium phosphide ($ZnGeP_2$), pumped at 2–3 μm by solid-state crystal lasers such as those based on holmium or erbium. An étalon-tuned narrowband (~0.1 cm^{-1}) $ZnGeP_2$ OPO, itself pumped at 2.55 μm by a broadband (~15 cm^{-1}) Nd:YAG-pumped $LiNbO_3$ OPO, is able to cover the wavelength range 3.7–8 μm and has been used to record spectra of H_2O vapor in the 6.2 μm region. Pulsed OPOs based on gallium arsenide (GaAs) also have great potential; they can be pumped at wavelengths as short as 1 μm and

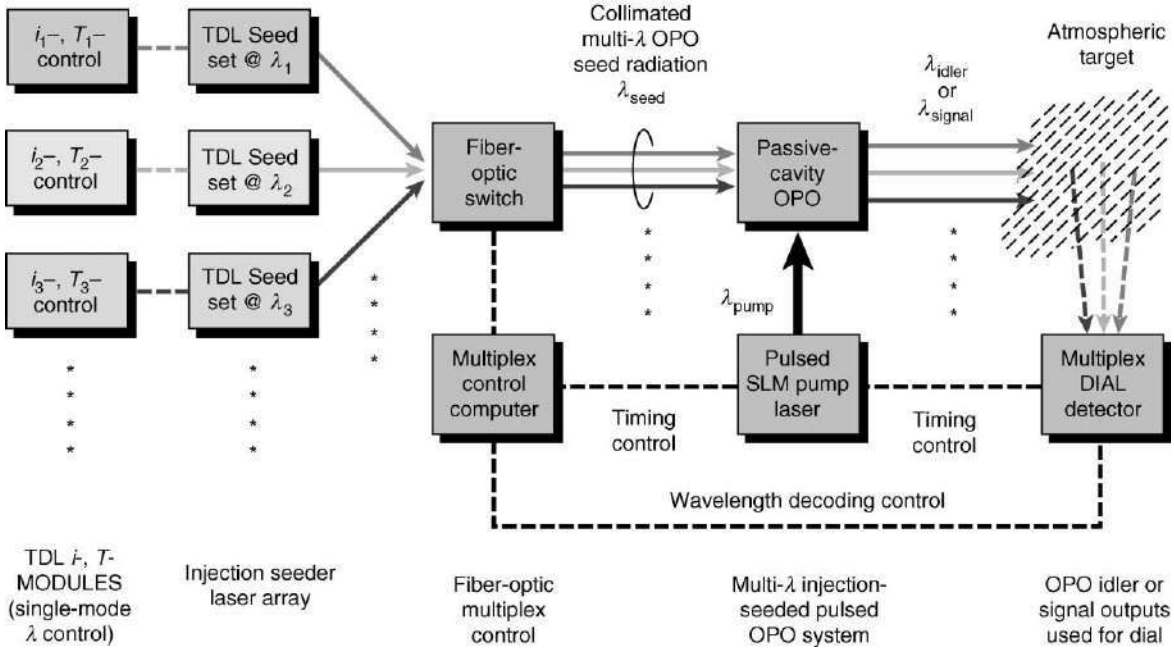

Figure 4 Schematic diagram of a multi-wavelength pulsed OPO system employing spectroscopically tailored injection seeding. This design has been proposed by the author for atmospheric remote sensing by DIAL (differential absorption lidar) or long-path absorption.

offer (so far unrealized) tunability out to ~16 μm. Despite its cubic structure and lack of birefringence, GaAs serves as a QPM nonlinear-optical medium in the form of orientation-patterned films.

Another approach to operating continuously tunable OPOs is to vary the wavelength of the pump laser, while fixing other phase-matching parameters such as temperature of the gain medium and crystal angle or QPM grating pitch. Suitable tunable pump laser sources that have been used to pump ns-pulsed OPOs include Ti:sapphire and tunable fiber lasers.

The various wavelength-control strategies outlined above illustrate some of the distinctive attributes of OPOs. It should be recognized that the signal and idler wavelengths of an OPO are distinctly different from that of the pump laser itself, thereby opening up fresh regions of the electromagnetic spectrum. Even if an OPO is injection-seeded (e.g., by a cw TDL) at its signal or idler wavelength, it still generates complementary idler or signal radiation at a new wavelength and (in the case of a pulsed OPO) it also pulse-amplifies the original seed light. Moreover, the nonlinear-optical parametric relationships between the pump, signal, and idler waves in an OPO allow various wavelength-control approaches (e.g., spectroscopic tailoring by multi-wavelength injection seeding) that are simply not attainable in tunable lasers.

Ultrafast-Pulsed OPOs

Given the relatively high peak power of very short (sub-ns) pulses of coherent laser light, it might seem quite easy to use such light to pump the gain medium of an OPD and to exceed its OPO threshold. However, such a task is not at all straightforward, because an ultrafast light pulse, on the time-scale of picoseconds (ps, 10^{-12} s) or femtoseconds (fs, 10^{-15} s), does not stay in one place long enough to enable a coherent parametric wave to build up from noise in the nonlinear-optical medium. Light travels a distance of only ~0.3 mm in 1 ps, which does not allow it to make multiple traversals of the nonlinear-optical medium and build up coherent signal and idler waves. This problem is overcome by synchronous pumping (also used in some ultrafast lasers), in which a train of many consecutive ultrafast pulses from a pump laser interact sequentially with a single signal or idler pulse circulating within the OPO cavity. The mode-locking interval of the pump laser (essentially the round-trip transit time in the laser cavity) must therefore equal the round-trip time of the down-converted (signal or idler) pulse in the OPO cavity. The situation is aggravated when sub-ps (e.g., ~100 fs) pump pulses are used, because the pump and signal (or idler) waves have sufficiently different group velocities that they undergo

'temporal walk-off', becoming separated in time after traversing a relatively short length (typically 1–10 mm for 100-fs pulses) of the nonlinear-optical medium.

Practical ultrafast OPOs were introduced around 1990 and their utility grew as improved pump lasers, nonlinear-optical materials (e.g., KTP, PPLN), and OPO cavity designs became available. Research on ps- and fs-OPOs is now a hot topic, and numerous broadly tunable commercial systems are on the market. Ultrafast OPO-based spectroscopy is becoming extremely important in applications to biology and medicine, where many key processes occur on ps- and fs-time-scales.

Continuous-Wave OPOs

The first cw OPO, reported in 1968, operated in the near-IR and was based on barium sodium niobate ($Ba_2NaNb_2O_5$ – known popularly as 'banana'); it was soon followed by visible cw OPOs based on $LiNbO_3$. However, there was not much more substantial progress on practical cw OPOs for another 20 years, until several technological advances had been made: low-threshold QPM nonlinear-optical gain media, such as $MgO:LiNbO_3$ and KTP; efficient, wavelength-selective cavity designs; all-solid-state cw pump lasers (e.g., diode-pumped crystal lasers or MOPA diode systems). More recently, QPM media (such as PPLN and PPKTP) have provided a further stimulus to efficient, compact cw OPOs.

Most ns-pulsed OPOs use singly resonant (SROPO) designs in which the cavity resonates at either the signal or idler wavelength (but not both), doubly resonant (DROPO) schemes in which both signal and idler resonate are often needed to meet the more demanding pump threshold requirements of cw OPOs. Continuous tuning of a DROPO without longitudinal mode hops needs multiparameter approaches, either dual-cavity (with separate signal- and idler-resonant arms) or concerted tuning of pump and signal/idler wavelengths. Efficient cavity designs for cw OPOs include those that are pump-enhanced (i.e., the cavity is resonant at the pump wavelength as well as the signal and/or idler) and intracavity arrangements in which the laser and OPO gain media share a common cavity. Progress is also being made on cw OPOs based on highly nonlinear optical fibers. With such developments, cw OPOs now have great potential as tunable spectroscopic sources, through their narrow optical bandwidth and their compactness.

OPGs and DFGs – Dispensing with the Optical Cavity

Most of this article has focused on OPOs, in which parametric processes are generated by pumping a nonlinear-optical medium in an optically resonant cavity. As already explained, such a cavity is used to enhance the optical parametric gain by multipassing the signal and/or idler radiation. In some cases, the cavity also contributes to wavelength control, temporal characteristics, beam quality, etc. However, not all optical parametric and related devices need a cavity, as is evident from the OPG, OPA, and DFG schematics in **Figures 1a, b,** and **d**.

The use of one or more OPA stages, to supplement the intracavity amplification already present in an OPO itself, is widespread in tunable OPO/OPA systems. Moreover, in some applications (e.g., spectroscopic measurements where optical power is not at a premium) of pulsed OPDs, it is found that a single pass of a (relatively long) nonlinear-optical crystal results in a useful OPG (without an OPO cavity). For instance, Scott Bisson, Thomas Kulp, Peter Powers and co-workers have developed and used several OPD systems of this type. A cavity-free OPG/OPA system that is pumped by a 1 kHz Nd:YAG laser, based on 'fan' PPLN (tuned by continuous transverse variation of the QPM grating pitch), and spectrally filtered by a high-finesse étalon, generates ~10 μJ/pulse of idler radiation. This OPG/OPA radiation has been used to measure narrowband ($\Delta\nu < 0.1\ \mathrm{cm}^{-1}$) cavity-ring-down spectra of CH_4 gas.

Likewise, various tunable DFG devices (closely related, but not identical, to OPDs) have been used in many spectroscopic applications. For instance, a DFG system developed by Frank Tittel and colleagues at Rice University took the near-IR output of two room-temperature TDLs (at 1.55 μm and 1.32 μm, respectively), mixed them in a BPM nonlinear-optical medium such as $AgGaSe_2$, and generated 8.7 μm mid-IR radiation for long-path absorption studies of sulfur dioxide (SO_2), an important atmospheric pollutant. The same group has also used QPM media (e.g., PPLN) for DFG mixing of TDL outputs in the 0.8 μm–1.6 μm range, yielding output at ~3.5 μm to monitor gases such as H_2O, CH_4, and formaldehyde (H_2CO); such DFG devices have been used in NASAs Lunar-Mars Life Support Test program.

Relative to OPGs and DFGs, OPOs and OPAs offer higher output power and coherence, both of which are advantageous in some applications (e.g., remote spectroscopic sensing of the atmosphere). Nevertheless, OPGs and DFGs still have a useful role to play, as indicated above.

Concluding Remarks – When to Use an OPD?

Coherent light sources based on optical parametric generation, amplification, and oscillation represent a significant application of nonlinear optics. Such devices can in turn address as many applications as lasers themselves, at least in principle. However, the fact that OPDs need to be pumped by a laser tends to limit their use to situations that take advantage of their wavelength-conversion and tunability capabilities. The most obvious areas of application of an OPO or any other OPD are therefore those in which the wavelength of the output radiation is critical (e.g., spectroscopy, as indicated in various parts of the above text).

From time to time, there have been suggestions that OPOs would largely displace established tunable lasers such as dye and Ti:sapphire lasers. In the early 1990s, one major laser manufacturer ventured to advertise its new OPO-based instrument with the slogan, "Goodbye to Ti: and Dye". However, such a projected monopoly for OPOs has not eventuated and most laser spectroscopy facilities now depend upon a balanced combination of OPOs together with various forms of tunable laser (including semiconductor lasers, which are almost ubiquitous).

See also

Fiber and Guided Wave Optics: Nonlinear Optics. **Nonlinear Optics, Applications:** Phase Matching. **Nonlinear Optics, Basics:** $\chi^{(2)}$-harmonic Generation; $\chi^{(3)}$ – Third-harmonic Generation; Four-wave Mixing. **Optical Parametric Devices:** Optical Parametric Oscillators (Continuous Wave); Optical Parametric Oscillators (Pulsed). **Scattering:** Stimulated Scattering. **Ultrafast Laser Techniques:** Generation of Femtosecond Pulses.

Further Reading

Barnes NP (1995) Optical parametric oscillators. In: Duarte FJ (ed.) *Tunable Lasers Handbook*, pp. 293–348. San Diego, CA: Academic Press.

Byer RL (1975) Optical parametric oscillators. In: Rabin H and Tang CL (eds) *Quantum Electronics: A Treatise*. Nonlinear Optics, part B, chap. 9, vol. 1, pp. 587–702. New York: Academic.

Byer RL and Herbst RL (1977) Parametric oscillation and mixing. In: Shen Y-R (ed.) *Nonlinear Infrared Generation*, pp. 81–137. New York: Springer-Verlag.

Dmitriev VG, Gurzayan GG and Nikogosyan DN (1997) *Handbook of Nonlinear Optical Crystals*, 2nd edn, pp. 339–362.

Dunn MH and Ebrahimzadeh M (1998) Optical parametric oscillators. In: Finlayson DM and Sinclair BD (eds) *Advances in Lasers and Applications*, pp. 61–82. Bristol and Philadelphia: Institute of Physics Publishing.

Ebrahimzadeh M and Dunn MH (2001) Optical parametric oscillators. In: Bass M, Enoch JM, Van Stryland EW and Wolfe WL (eds) *Handbook of Optics, Fiber Optics and Nonlinear Optics*, vol. 4, pp. 22.1–22.73. New York: McGraw-Hill.

Harris SE (1969) Tunable optical parametric oscillators. *Proceedings of the IEEE* 57: 2096–2113.

Orr BJ (2000) IR Lidar applications in air monitoring. In: Myers RA (ed.) *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation, Applications of Instrumental Methods*, vol. 3, pp. 2007–2032. Chichester: Wiley.

Orr BJ, Johnson MJ and Haub JG (1995) Spectroscopic applications of pulsed tunable optical parametric oscillators. In: Duarte FJ (ed.) *Tunable Laser Applications,* chap 2, pp. 11–82. New York: Marcel Dekker.

Sorokina IT and Vodopyanov KL (eds) (2003) *Solid-State Mid-Infrared Sources*. New York: Springer-Verlag.

Sutherland RL (1996) Optical parametric generation, amplification, and oscillation. *Handbook of Nonlinear Optics,* chap. 3, pp. 111–206. New York: Marcel Dekker.

Optical Parametric Oscillators (Continuous Wave)

S Schiller, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

A continuous-wave optical parametric oscillator (cwOPO) is a frequency conversion device in which a continuous-wave (cw) laser field (pump) of frequency ω_p is partially converted into two cw waves of lower frequencies ω_s, ω_i (historically denoted by signal and idler) inside an optical cavity. The three frequencies are related by $\omega_i + \omega_s = \omega_p$. The cavity contains a nonlinear medium with a second-order susceptibility and the mirror reflectivities are chosen such that at least one of signal and idler resonates in the cavity. The conversion efficiency from pump power to signal or idler power or both can be larger than 10%. The frequencies of signal and idler waves can be tuned by means similar to conventional lasers, i.e., by changing the spectral dependence of the cavity

loss and by tuning the cavity length. Additional tuning parameters specific to OPOs are the pump frequency and the refractive indices of the nonlinear medium for pump, signal and idler. CwOPOs provide a unique combination of features that makes them suitable sources for a variety of spectroscopic applications. Today's cwOPOs can exhibit the following features:

- spectral coverage from the near-infrared (0.7 μm) to the mid-infrared (5.2 μm); for a single device two optical octaves can be covered without change of optical components;
- high output power (up to several W);
- high pump power conversion efficiency (>30%) and good wall-plug efficiency;
- nearly diffraction-limited output beams (modes are close to lowest-order transverse electromagnetic (TEM_{00}) modes)
- single-frequency output, absence of frequency jumps (mode hops);
- narrow free-running linewidth (<6 kHz);
- high absolute frequency stability (<100 MHz drift per hour);
- continuous frequency tuning exceeding 10 GHz;
- rapid, stepwise tuning over several THz; and
- transportability.

cwOPOs complement their pulsed relatives (*see* **Optical Parametric Devices:** Optical Parametric Oscillators (Pulsed)) nanosecond and synchronously pumped pulsed parametric oscillators (*see* **Optical Parametric Devices:** Overview).

Basic Principle

The operation of a cwOPO is most simply described for a singly resonant OPO (SRO) (see Figure 1 and Figure 2a).

Without loss of generality we take the signal wave to be the resonant wave. When a higher-frequency pump wave (power P_p, angular frequency ω_p) and a lower-frequency (signal) wave (power P_s, angular frequency ω_s) traverse a $\chi^{(2)}$ medium (*see* **Fiber and Guided Wave Optics:** Nonlinear Optics), a transfer of power from the pump wave to the signal wave occurs, described by a power gain:

$$\Delta P_s = GP_s = EP_pP_s \qquad [1]$$

This parametric amplification expression is valid as long as $\Delta P_s \ll P_s, P_p$, i.e., in the limit of small gain and small pump depletion. The (unsaturated) gain coefficient E is a function of: the medium's indices of refraction, orientation and length L, the frequencies

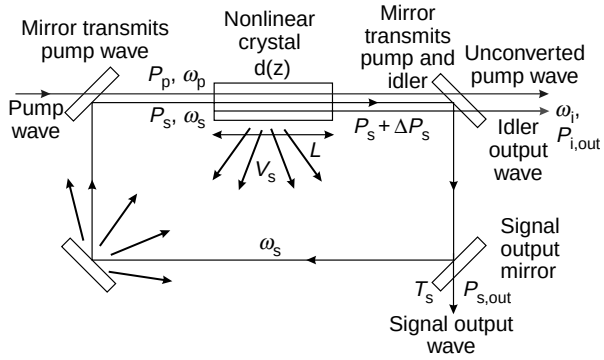

Figure 1 Schematic of a singly resonant cwOPO (SRO) in ring configuration. The signal wave (ω_s) is resonated, the idler wave (ω_i) is not. In the nonlinear crystal, the pump wave overlaps with the signal and idler waves; here they are shown displaced with respect to each other for clarity. The signal wave's attenuation coefficient per roundtrip, S_s, is the sum of loss inside the nonlinear crystal, V_s (indicated by arrows), and transmission through mirrors, here represented by a single mirror of transmission T_s. Scatter and absorption losses at mirrors (indicated by arrows) are included in V_s.

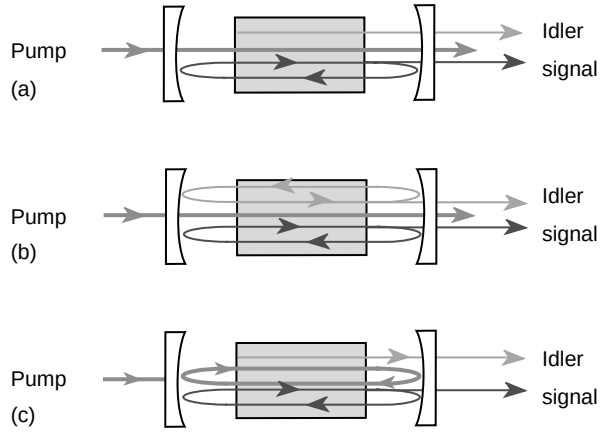

Figure 2 Common types of cwOPOs. (a) basic SRO. (b) Doubly resonant OPO (DRO). Standing-wave cavities are shown. (c) SRO with enhancement of the pump wave (PR-SRO); a common-cavity configuration is shown.

ω_s, ω_p, and the spatial modes and overlap of the two beams.

If the wave ω_s is made to circulate inside a cavity and the power attenuation per roundtrip, S_sP_s is smaller than the power gain per roundtrip (eqn [1]), a sustained oscillation is possible. This is satisfied when the input pump power P_p exceeds the threshold:

$$P_{p,in}^{th} = S_s/E \qquad [2]$$

As the pump power P_p is increased above $P_{p,in}^{th}$, conversion of pump power to idler power occurs, implying that the pump power decreases within the medium. Qualitatively, the average pump power $\langle P_p \rangle$

along the medium remains at the level $P_{\mathrm{p,in}}^{\mathrm{th}}$, so that the roundtrip fractional power gain remains constant, $G \simeq E\langle P_{\mathrm{p}}\rangle \simeq EP_{\mathrm{p}}^{\mathrm{th}} \simeq S_{\mathrm{s}}$. This ensures that the device reaches a stationary state. The circulating signal power P_{s} increases steadily with increasing pump power, and so does the signal power emitted from the OPO, $T_{\mathrm{s}}P_{\mathrm{s}}$, which is equal to the 'useable' fraction of the roundtrip power gain, $(T_{\mathrm{s}}/S_{\mathrm{s}})GP_{\mathrm{s}}$.

The parametric process that allows the signal wave to experience gain (eqn [1]), also generates an idler wave (difference frequency generation). Its power at the end of the medium is given by the Manley–Rowe relation:

$$P_{\mathrm{i}} = \frac{\omega_{\mathrm{i}}}{\omega_{\mathrm{s}}}\Delta P_{\mathrm{s}} \qquad [3]$$

An OPO also requires a process that starts the oscillation. Spontaneous parametric fluorescence provides this trigger. Due to this quantum effect, a pump wave traversing a $\chi^{(2)}$ medium will generate a flux of signal photons even if there is no signal wave present at the crystal input face. In the photon picture, a tiny fraction (on the order 10^{-9} for a 1 W pump) of the pump photons spontaneously annihilate, creating pairs of signal and idler photons (photon splitting). The idler photons have a frequency $\omega_{\mathrm{i}} = \omega_{\mathrm{p}} - \omega_{\mathrm{s}}$, so that photon energy conservation holds. The process may be thought of as being stimulated by the vacuum fluctuations of the electromagnetic field at the signal and idler frequencies. Some of the signal photons are emitted into the cavity mode, so that after a roundtrip they can serve as a seed for further (stimulated) photon splitting. The process repeats, and the signal wave builds up exponentially in time to a macroscopic level if the pump power is above threshold. The spontaneous (as well as the parametric amplification) process is strongest when momentum conservation is essentially fulfilled, $k_{\mathrm{p}} \simeq k_{\mathrm{s}} + k_{\mathrm{i}}$, where $k = \omega n(\omega)/c$.

In the photon picture, the relation (eqn [3]) is a statement that signal and idler photons are always created in pairs.

Types of OPOs

OPO types may be classified according to the number of waves that resonate within cavities. Typically, a realizable roundtrip loss coefficient is on the order of $S_{\mathrm{s}} = 1\%$ (to which the linear absorption of the nonlinear crystal and mirror transmissivities contribute) and the gain coefficient $E \sim 0.1\%/\mathrm{W}$–$1\%/\mathrm{W}$ for Gaussian waves, depending on the pump wavelength, crystal type, etc. This leads to SRO threshold powers on the order of 1 to 10 W. Until the early 1990s, solid-state pump lasers of this power level and single-frequency output were unavailable. This led to the interest in and development of cwOPOs in which more than one of the three waves is resonated (Figure 2). The corresponding external threshold powers are

$$\begin{aligned} P_{\mathrm{p,in}}^{\mathrm{th}} &\simeq S_{\mathrm{s}}S_{\mathrm{i}}/4E \text{ for a doubly-resonant OPO (DRO, Figure 2b)} \\ P_{\mathrm{p,in}}^{\mathrm{th}} &\simeq S_{\mathrm{s}}S_{\mathrm{p}}/E \text{ for a pump-resonant SRO (PR-SRO, Figure 2c, see eqn [35])} \\ P_{\mathrm{p,in}}^{\mathrm{th}} &\simeq S_{\mathrm{s}}S_{\mathrm{i}}S_{\mathrm{p}}/E \text{ for a triply-resonant OPO (TRO)} \end{aligned} \qquad [4]$$

Here $S = T + V$ is the fractional power attenuation per roundtrip, arising from the mirror transmission T and absorption/scatter loss V. While the mirror transmission maintains power in a given spatial mode, the loss V removes optical power completely and is the fundamental cause for the nonunity efficiency of the frequency conversion process. Typically, each additional resonant enhancement lowers the threshold power by one to two orders.

Examples for threshold values reported in the literature are: 1.9 W for a 920 nm-pumped SRO, 4 mW for a 0.5 μm-pumped DRO, and 140 mW for a 1 μm-pumped common-cavity PR-SRO. DROs can therefore be pumped even by low-power diode lasers. TROs (where pump, signal, and idler are resonated) are not of importance due to their higher complexity.

Output Power

As an example, the signal and idler output powers for a PR-SRO that resonates the signal are (for a derivation, see Appendix B):

$$P_{\mathrm{s,out}} \simeq T_{\mathrm{s}}P_{\mathrm{s}}(0), \qquad P_{\mathrm{i,out}} \simeq (\omega_{\mathrm{i}}/\omega_{\mathrm{s}})S_{\mathrm{s}}P_{\mathrm{s}}(0) \qquad [5]$$

where the circulating signal power is

$$P_{\mathrm{s}}(0) \simeq \frac{\omega_{\mathrm{s}}}{\omega_{\mathrm{p}}}\frac{4T_{\mathrm{p}}}{S_{\mathrm{s}}S_{\mathrm{p}}}P_{\mathrm{p,in}}^{\mathrm{th}}\left(\sqrt{\frac{P_{\mathrm{p,in}}}{P_{\mathrm{p,in}}^{\mathrm{th}}}} - 1\right) \qquad [6]$$

T_{p} is the input mirror transmission for the pump wave, and T_{s} is the output mirror transmission of the signal wave. Signal and idler output powers have the same dependence on pump input power and increase monotonically with $P_{\mathrm{p,in}}$. The conversion efficiencies reach a maximum four times above threshold, i.e., when $P_{\mathrm{p,in}} = 4P_{\mathrm{p,in}}^{\mathrm{th}}$. There, the signal and idler photon (quantum) conversion efficiencies $\omega_{\mathrm{p}}P_{\mathrm{s,out}}/P_{\mathrm{p,in}}\omega_{\mathrm{s}}$ and $\omega_{\mathrm{p}}P_{\mathrm{i,out}}/P_{\mathrm{p,in}}\omega_{\mathrm{i}}$ exceed 25% and 50%, respectively, if the mirror transmissions T_{p}, T_{s} for pump and signal

are chosen larger than the respective loss V_p, V_s. This can be achieved in practice, of course, at the expense of a higher threshold.

Nonlinear Materials

While the fundamental developments in cwOPOs were performed using birefringently phase-matched nonlinear crystals ($MgO:LiNbO_3$, LBO, KTP, $KNbO_3$) (*see* **Materials for Nonlinear Optics:** Liquid Crystals for NLO), and substantial spectral coverage is possible with these crystals, most cwOPOs are now operated with quasi-phase-matched (QPM) nonlinear crystals (*see* **Nonlinear Optics, Application:** Phase Matching). The most commonly used material is periodically poled (PP) $LiNbO_3$ (PPLN), other crystals used are PPKTP and $PPLiTaO_3$, crystals. Their use has led to a substantial widening of the spectral coverage, opening in particular the wavelength range $\lambda > 1.6$ μm, a reduction of the required pump powers thanks to their much higher nonlinearities compared to birefringently phase-matched crystals, and a simplification of (gross) tuning through the use of multigrating structures (containing several sections of differing Λ periods side by side within the same crystal). The above PP crystals exhibit a loss of 0.1%/cm at wavelength around 1 μm, and can be coated with broadband dielectric antireflection coatings, in order to maximize the overall transmission. Another important aspect is that the crystals withstand long irradiation times with high-power focused cw light. For PPLN, operation temperature above 100°C is chosen in order to prevent photorefractive effects.

Spectral Coverage

An example of particularly wide spectral coverage reported in the literature is a cwOPO pumped by a frequency-doubled $Nd:YVO_4$ laser ($\lambda_p = 532$ nm) using a multigrating PPLN crystal. With grating periods $\Lambda = 6.5–9.6$ μm, the range 660–1030 nm (signal wave) and 1100–2800 nm (idler wave) was covered. A cwOPO with a Nd:YAG pump source emitting at the fundamental wavelength ($\lambda_p = 1064$ nm) and a multigrating PPLN crystal ($\Lambda = 25.5–31.2$ μm) achieved a range of 1.45–1.99 μm (signal) and 2.3–4.0 μm (idler).

For idler wavelengths beyond 4 μm, idler absorption by the crystals types above becomes relevant, leading to a reduction in the gain G (this effect is not considered in the treatment given in the Appendix) and to an increase of the threshold. To a certain extent, this can be compensated with powerful cw pump lasers. For example, a 1064 nm pumped PPLN-SRO achieved idler emission in the range 3.6–4.7 μm at 11 W pump power. Also, an 850 nm pumped PR-SRO covered an idler range 4.3–5.3 μm when pumped with 750 mW.

In the future, cwOPOs will certainly profit from the development of novel QPM materials with wider IR transparency range.

Frequency Control

The small-signal gain $G(\omega_s)$ of a parametric interaction in a cwOPO is analogous to the homogenously broadened gain in a laser medium. In contrast to it, however, spatial hole-burning cannot occur. Therefore, mode competition is generally strong in a cwOPO, which will oscillate on a single longitudinal mode: the one for which the difference (or ratio) between 'unsaturated' parametric gain G and total loss S is largest (**Figure 3**). This is equivalent to the condition of minimum threshold.

The functions $G(\omega_s)$ and $S(\omega_s)$ are therefore important characteristics of a cwOPO. Oscillation can only occur for a frequency ω_s that is very close to a cavity mode frequency $\omega_q = 2\pi qc/OPL(\omega_q)$, where q is the integer mode number, and $OPL(\omega_s)$ is the round-trip optical path length at the frequency ω_s.

To tune a cwOPO to a particular signal/idler frequency pair, a coarse tuning is initially performed

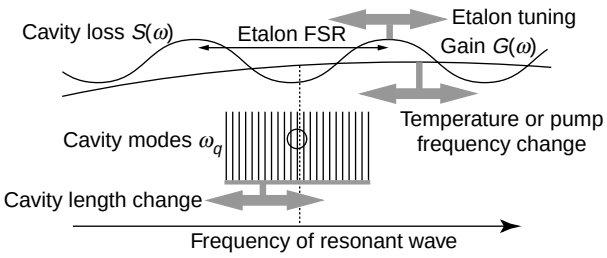

Figure 3 Frequency selection and tuning in a SRO (or PR-SRO) containing an etalon. Oscillation occurs on the signal cavity mode (circled) closest to the frequency where the difference between the small-signal gain G and the loss S is largest (dashed line). The frequency of this mode or the mode number q can be changed by changing the parameters indicated near the thick double arrows. The parameter changes shift the line(s) 'attached' to the corresponding double arrow. Changing the temperature leads to a rough setting of the oscillation frequency. Tilting the etalon allows mode-hop tuning of signal and idler frequencies. Fine tuning by a small amount can be done by changing the cavity length. The frequency of the idler is not shown here; it is always given by the difference between pump frequency and the frequency of the oscillating mode. Thus, tuning of the pump frequency generally leads to a tuning of the idler frequency. For modest pump tuning the amount of signal tuning however depends on SRO type: no signal tuning in a basic SRO and in a dual-cavity PR-SRO, while in a common-cavity PR-SRO pump and signal frequencies are correlated and thus pump tuning leads to signal tuning. FSR: free spectral range.

by tuning the corresponding phase mismatch ΔkL (L is the crystal length) approximately to zero. Taking into account a modulation of the nonlinear susceptibility of the crystal with period Λ, the quantity $\Delta k = k_p - k_s - k_i \pm 2\pi/\Lambda$ describes the deviation from momentum conservation in the photon splitting process. When using a multigrating QPM crystal, the first tuning consists in the selection of a grating of appropriate period Λ, which is moved into the pump wavepath. This is followed by tuning the photon momenta via the refractive indices of the nonlinear medium, which is usually accomplished by temperature tuning, or alternatively by angle-tuning of the crystal. These steps place the maximum of $G(\omega_s)$ to the vicinity of the desired frequency. Options for tuning the emission exactly to the desired frequency are described below for the different OPO types.

Because external perturbations change parameters that enter in $G(\omega_s)$, $S(\omega_s)$, and ω_q, the value of the oscillating frequency will also change in time. The change will be continuous over a certain range of the parameters, but if their perturbation is large enough, a mode hop can occur: the oscillation on the initial mode q will eventually cease and another mode q', for which $G(\omega_{q'}) - S(\omega_{q'})$ is larger, will oscillate instead. A variety of parameters is affected by perturbations: pump power, pump beam direction, cavity length, crystal temperature, and pump frequency.

When such perturbations are applied on intentionally, the OPO output frequencies can be tuned, scanned, or adjusted. The tolerable change in the various parameters before a mode-hop occurs, and the sensitivity of the output frequencies to changes in the parameters depends on the particular cwOPO type. One important approach to suppress mode-hops consists in adding elements into the cavity, such as an etalon, that modify the function $S(\omega)$, typically by modulating it spectrally.

SRO

In the following, we assume that the signal wave is resonated. The SRO is the simplest cwOPO type. The oscillation nominally occurs at the cavity frequency that is closest to the gain maximum, as described above. In practice, environmental disturbances will cause mode-hops on a time-scale of minutes. For this reason, an etalon is added to the cavity. Mode-hop-free oscillation, for several hours, is then achieved.

Typical nonlinear crystals exhibit parametric gain with spectral widths of hundreds of GHz. In contrast, the cavity mode spacing (free spectral range, FSR) c/OPL is on the order of 1 GHz. The tuning range achievable by a cavity length change (using, e.g., a piezo translator) is therefore, at most, one free spectral range (equal but opposite for signal and idler); then a mode-hop to the neighboring cavity mode occurs, with a signal/idler frequency jump by one FSR in the opposite direction to the previous tuning. This process repeats itself as the cavity length is tuned further. It is, therefore, not possible to easily and reliably access a large frequency range by using the cavity length as tuning parameter.

As shown in Figure 3, if an etalon of linewidth smaller than the gain bandwidth is added to the cavity, it becomes the dominant influence (etalon FSRs on the order of one hundred to a few hundred GHz have been used). This means that shifting the etalon's transmission maximum (i.e., the cavity loss S) in frequency space (by tilting a solid etalon or changing the spacing of an air-spaced etalon) as indicated by the double-headed arrows in Figure 3, forces the signal to mode-hop to an adjacent mode, with a corresponding idler frequency hop in the opposite direction. As the etalon is tuned further this occurs repeatedly. This discontinuous 'mode-hop tuning' is useful for tuning the SRO output frequencies over a significant range (e.g., 100 GHz) and allows to reach a desired frequency within one cavity free spectral range.

Continuous tuning of both signal and idler waves is obtainable, if the pump laser is not frequency tunable, by changing the cavity length and synchronously tuning the etalon. This can be achieved by a feedback or feedforward system that controls the etalon angle or spacing. Up to 38 GHz tuning for signal and idler were achieved in this way.

When a tunable pump is available, one can tune the pump frequency while keeping the cavity length and the etalon angle fixed. This keeps the signal cavity mode fixed, and thus the nonresonant wave's frequency (usually the idler) is tuned by an amount equal to the pump frequency change. Continuous tuning ranges exceeding 50 GHz have been obtained in this way. For such pump tuning ranges, the shift of the gain curve, indicated by a double-arrow in Figure 3, is usually sufficiently small that it can be neglected. The use of a widely tuneable pump laser is also possible. A (discontinuous) idler tuning of 9 THz around 3.3 μm has been demonstrated, covered in just 0.3 ms thanks to a large pump tuning rate. In such a case of large pump tuning range, the gain shift is important, causing a significant tuning also of the signal.

The advantages of the basic SRO as compared to other cwOPOs are mainly the simpler tuning behavior, the possibility of continuous tuning without necessitating a tunable pump laser, and the possibility to achieve a very wide tuning range without optics

change, since only the spectral range covered by one of the two parametric waves must be provided as a high reflectivity coating on the cavity mirrors. Moreover, in the simplest implementation, it is possible to operate without any electronic servo control of cavity elements, at the expense of a reduced power and frequency stability. In terms of spectral properties (linewidth, frequency stability, time between undesired mode-hops) the SRO does not offer particular advantages compared to other cwOPO types, even if actively stabilized (see below). The disadvantage of the basic SROs based on Gaussian modes is the high threshold, necessitating expensive high-power (several W) cw lasers. This disadvantage is likely to disappear with progress in laser development. The alternative of waveguide SROs is mentioned below.

PR-SRO

In a PR-SRO, the pump wave is also resonated, leading to a substantial reduction of threshold. For example, it is possible to generate mid-IR radiation (3 μm range), even with much more affordable 1 W-level Nd:YAG pump lasers.

One implementation is the common-cavity PR-SRO (Figure 4). The cavity length is either locked to the pump laser frequency or vice-versa. If a frequency-stable pump laser is used, the first solution is favorable, since then both signal and idler waves have a stability comparable to that of the pump laser. With an etalon (coated so as to be transparent for the pump wave) inside the cavity, mode-hop-free operation for several hours has been demonstrated.

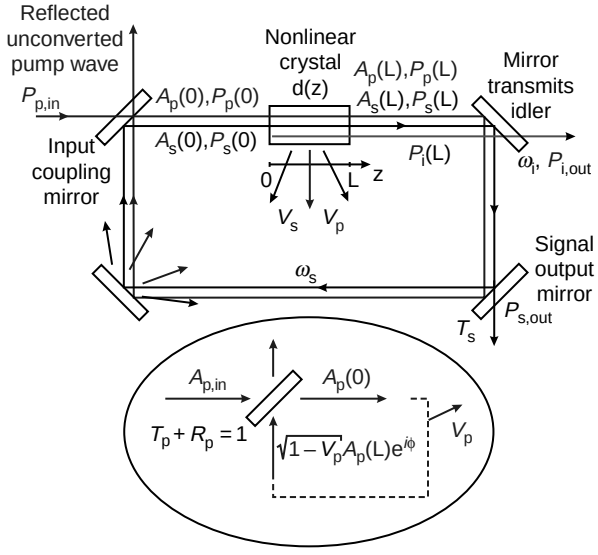

Figure 4 Schematic of a pump-resonant SRO with a common cavity for pump and signal waves. The inset shows the pump wave fields incident on and leaving the input coupling mirror. The arrow symbolizes pump wave loss during the round trip.

Mode-hop tuning in a PR-SRO with cavity-to-laser lock can be achieved:

(i) by tuning the etalon at constant pump frequency: the signal mode-hops from mode to mode, the idler tunes equally but in opposite direction;
(ii) by tuning the pump frequency at constant etalon position: the cavity length must follow and therefore the signal tunes by $\Delta\omega_s = (\omega_s/\omega_p)\Delta\omega_p$, while the idler frequency tunes by $\Delta\omega_i = (\omega_i/\omega_p)\Delta\omega_p$. This continues until the signal must mode-hop back, causing an equal but opposite frequency jump in the idler. The pump tuning is thus completely transferred to the idler.

Continuous signal and idler tuning is performed by tuning the pump frequency and synchronously adjusting the etalon angle. Thus, a limitation of the common-cavity PR-SRO concept is that continuous tunability necessitates a tunable pump laser.

In a dual-cavity, PR-SRO pump and signal waves resonate in separate cavities. The pump cavity is usually stabilized to the pump frequency and an etalon is placed into the signal cavity for better stability against mode-hops. This configuration combines the tuning simplicity of a basic SRO with a low-threshold power. For example, a threshold for a 1 μm-pumped PPLN dual-cavity PR-SRO at the 0.4 W level was obtained.

Mode-hop tuning (in opposite directions) of the frequencies of both parametric waves in steps of the cavity's FSR can be performed by changing the etalon angle (Figure 3). A range exceeding 50 GHz (at constant output power) has been achieved. Continuous signal and idler wave tuning (by equal but opposite amounts) is possible by changing the signal cavity length (Figure 3). However, the range is limited to the FSR of the cavity, typically on the order of 500 MHz, then the frequencies mode-hop back. It is expected that wider continuous tuning should be possible by synchronous etalon and cavity length scan. If a tuneable pump laser is employed, continuous tuning of only the idler frequency is possible by keeping the etalon and signal wave cavity length fixed; the pump tuning range is fully transferred to the idler.

DRO

In contrast to an SRO, this device type requires both parametrically generated waves to be resonant in the cavity. Here we consider only DROs with a common cavity for both signal and idler waves (dual-cavity DROs have also been demonstrated).

Because of the added resonance condition, mode-hops of the oscillating mode pair occur already upon a very small change (~1 nm) in the cavity length. Under normal environmental conditions such changes are unavoidable and therefore a 'free-running' DRO does not produce frequency-stable output. However, by active control of one additional parameter of the OPO system the mode-hops can be effectively suppressed. A practical solution is to actively control the length of the cavity via the position a of one of the cavity mirrors so that a particular mode pair, q_s, q_i, continues to oscillate; this is achieved by maximizing the OPO output power or by minimizing the detuning of one of the two resonant waves from its respective cavity mode frequency. This control system ensures that energy conservation and the two resonance conditions are simultaneously satisfied:

$$\begin{aligned} \omega_s + \omega_i &= \omega_p \\ \omega_s &= 2\pi c q_s/\mathrm{OPL}(d, \omega_s, T) \\ \omega_i &= 2\pi c q_i/\mathrm{OPL}(d, \omega_i, T) \end{aligned} \qquad [7]$$

Experimentally, mode-hop-free oscillation exceeding 10 h was achieved by such active stabilization.

Another system parameter may then be used to tune both signal and idler frequencies. The tuning coefficient for the crystal temperature T is small, but tuning the pump frequency is an efficient way. The tuning coefficient is approximately $\Delta\omega_s = (\omega_s/\omega_p)\Delta\omega_p$, and accordingly for ω_i.

Continuous tuning of signal and idler frequencies over 10 GHz by pump tuning was achieved. This range was limited by the available continuous tuning range of the pump laser. Mode-hop tuning can be induced by mechanical perturbations: by tapping on the DRO cavity, mode-hop tuning over 220 GHz has been demonstrated.

Linewidth and Frequency Stability

For high-resolution spectroscopy a narrow linewidth for signal or idler is desirable. One contribution to the linewidth of a cwOPO arises from the unavoidable spontaneous parametric fluorescence, similar to the Schawlow–Townes limit in lasers due to spontaneous emission. This contribution is negligible if the OPO operates at power levels of practical interest. Thus, other sources of classical noise dominate the linewidth of cwOPOs. As discussed above, the frequency of cwOPOs depends (on a fine scale) on pump frequency and cavity optical path length. Both parameters fluctuate: the pump laser has a finite linewidth as well as frequency jitter, and the cavity optical path length fluctuates due to acoustic noise affecting the mirror positions, to air pressure fluctuations, temperature fluctuations of the crystal and of the cavity structure.

With careful construction and shielding, many of these noise sources can be minimized. Furthermore, it is highly favorable to use pump lasers of narrow linewidth and high intrinsic frequency stability (e.g., diode-pumped monolithic solid-state lasers), or actively frequency-stabilized lasers. Note that this special requirement on the pump source does not exist for lasers.

In an SRO, the linewidth and frequency stability of the signal are mostly determined by the stability properties of the cavity. The idler spectral properties then follow entirely from the signal and pump spectral properties through $\omega_i = \omega_p - \omega_s$. Since the idler is usually the wave of interest, this implies that the SRO cavity and the pump laser must both have a stability appropriate to the spectroscopic application.

If the SRO cavity is not sufficiently stable (e.g., because of insufficient temperature stability of the oven containing the nonlinear crystal), the frequency fluctuation/drift of the signal and idler can be substantial, tens to hundreds of MHz within minutes being typical.

To avoid mode-hops also on long time-scales, the cavity length (or the etalon angle) can be actively stabilized so as to maximize the idler's output power. This ensures that the cavity mode position and the frequency of maximum difference between gain and loss are kept equal.

If a frequency-stable pump wave is used, it is possible to transfer these characteristics by stabilizing the SRO cavity length to the pump wave frequency. In an SRO without resonant pump one can make use of the residual reflection of the pump wave by the cavity mirrors that leads to interference with the pump wave reflected from the pump input mirror. This was implemented in a SRO pumped by a high-power diode-pumped internally doubled $Nd:YVO_4$ laser (532 nm) of high intrinsic frequency stability. Signal and idler drifts of less than 50 MHz/h were achieved. At the same time, a very small 20 kHz linewidth of the (resonant) signal radiation was obtained. Similar levels were obtained for a PR-SRO with a common cavity for pump and signal wave.

In a PR-SRO with dual-cavity configuration, the same requirements of minimization of disturbances to the parametric wave's cavity arise as in a basic SRO. Using a highly frequency-stable kHz linewidth monolithic pump laser, together with the above-mentioned idler power maximization control, frequency drift values for both parametric waves at the

level of 30 MHz/h were obtained and a short-term linewidth below 6 kHz.

In a single-cavity DRO, the linewidth and frequency stability of the parametric waves are essentially determined by the pump wave properties (see discussion in the section on DRO above). Here, too, the use of narrow-linewidth, high-frequency-stability pump lasers has led to excellent signal/idler characteristics. For example, less than 40 kHz linewidth and <40 MHz per hour drift were obtained with a DRO pumped by an externally doubled monolithic Nd:YAG laser of 10 kHz linewidth.

If further narrowing of linewidth or reduction of frequency drift is desired, the OPO output frequencies can be stabilized to a stable cavity, atomic, or molecular transition. Locking to a cavity has been demonstrated for a DRO.

Applications

CwOPOs are most interesting as sources of radiation in the IR range beyond 2 μm, especially in the 3–5 μm range, where vibrational transitions of many molecules occur. In comparison to other sources of coherent radiation (color center laser, quantum cascade laser, lead-salt diode laser, CO overtone laser, difference frequency generation), cwOPOs provide a unique combination of desirable features.

A simple application of cwOPOs would be as a source providing a power level/output wavelength/mode quality combination not available from other sources. Possible applications are vibrational spectroscopy in the condensed phase, where a narrow linewidth is not required. The excellent beam quality provides a high spatial resolution.

Concerning high-resolution spectroscopy (mostly in the gas phase), methods demonstrated in conjunction with cwOPOs include:

- Absorption spectroscopy;
- Doppler-free saturation spectroscopy;
- Photo-acoustic spectroscopy (PAS);
- Hole-burning spectroscopy;
- Coherent atomic spectroscopy;
- Cw cavity ring down spectroscopy (cw-CRDS).

Trace gas detection using PAS or cw-CRDS is a particular application where cwOPOs have reached excellent sensitivity (minimum detectable concentrations as low as 1 part in 10^{11}). The high output power and/or the narrow linewidth are specific features that make such sensitivity possible.

Quantum-Optical Properties

The pairwise production of signal and idler photons is at the heart of the parametric process. This perfect quantum correlation is to a certain extent also maintained in the output waves of a cwOPO. In particular, the intensity fluctuations of signal and idler waves emitted by a DRO are strongly (albeit never perfectly) quantum correlated if their output coupler transmissions significantly exceed their respective roundtrip losses. For these 'twin beams' the spectral density of the fluctuations of their power difference can be less than the spectral density of the fluctuations of the power of an individual wave (squeezing). Making use of this correlation permits to increase the obtainable sensitivity in spectroscopic measurements for a given power level of the wave interacting with the sample.

Device Development Trends

CwOPOs using solid-state pump lasers have become commercially available. In the near future, some likely developments will be miniaturization and optimization in order to reduce the influence of perturbations, further extension of continuous and mode-hop tuning ranges and the wider use of diode lasers as pump lasers, leading to cost reduction. To access a larger range of molecules in spectroscopy, extension of the emission range well beyond 5 μm is also of significant interest. Novel nonlinear materials and pump lasers of wavelength longer than the often employed Nd:YAG lasers will become of importance in this development.

Two specific lines of development complementary to the bulk cwOPOs described above are waveguide and intracavity cwOPOs (ICSRO). If pump, signal and idler waves are confined in a waveguide (*see* **Photonic Crystals:** Photonic Crystal Lasers, Cavities and Waveguides) the parametric gain coefficient is a few orders higher than for free-space Gaussian waves, due to the small mode cross-section and the scaling with the square of the crystal length, compared with the linear scaling in case of Gaussian beams (see eqn [36]). Values are on the order of $E \sim 1/\mathrm{W}$ at $\lambda_p = 1.5$ μm leading to low thresholds even for SROs, with pump lasers of comparatively long wavelength. For example, in a SRO consisting of a 9 cm long titanium:PPLN waveguide, a threshold below 300 mW for a 1.5 μm pump was achieved. The idler emission range was 3.1–3.4 μm. The potential for mass production and, therefore, cost minimization is clear.

An ICSRO consists of an SRO internal to the cavity of a cw laser, where the high circulating laser power is taken advantage of for pumping. An ICSRO is thus similar to a PR-SRO but there is no need to implement a control system for the frequency lock between pump laser and pump cavity nor to mode-match the pump wave to the SRO cavity mode.

Appendix A

Three-Wave Mixing in a $\chi^{(2)}$ Medium with Focused Waves

The wave equation in a nonlinear medium reads

$$\Delta\vec{E} - \mu_0\mu_r\frac{\partial^2}{\partial t^2}\vec{D}_{\mathrm{L}} = \mu_0\mu_r\frac{\partial^2}{\partial t^2}\vec{P}_{\mathrm{NL}} \qquad [8]$$

The displacement field $\vec{D}_{\mathrm{L}}$ is related to the electric field $\vec{E}$ by linear response. $\vec{P}_{\mathrm{NL}}$ is the nonlinear polarization. We deal with nonmagnetic media ($\mu_r = 1$) and choose the principal axis system as coordinate system. The Fourier transform of, for example, the x-component of the wave equation then reads:

$$\begin{aligned}\Delta E_x(\omega,\vec{r}) + \mu_0\omega^2\varepsilon_0\varepsilon_x(\omega)E_x(\omega,\vec{r}) \\ = -\mu_0\omega^2 P_{\mathrm{NL},x}(\omega,\vec{r})\end{aligned} \qquad [9]$$

In the following, we take into account only second-order nonlinear effects, so the nonlinear polarization reduces to $P_{\mathrm{NL}} = P^{(2)}$. The fields to be considered are monochromatic, propagate in the z-direction and are linearly polarized along x. The notation can then be simplified to:

$$E_x(\omega,\vec{r}) = E(\vec{r})\mathrm{e}^{\mathrm{i}kz}, \quad P_{\mathrm{NL},x}(\omega,\vec{r}) = P^{(2)}_{\omega}(\vec{r}) \qquad [10]$$

with the wavevector $k = \omega n(\omega)/c$ and the refractive index for the polarization under consideration $n_\omega^2 = n(\omega)^2 = E_x(\omega)$. The electric field is obtained as $\vec{E}(t,\vec{r}) = \frac{1}{2}\mathrm{Re}(E(\vec{r})\mathrm{e}^{\mathrm{i}(kz-\omega t)})\hat{x}$.

Since the parametric gains are small in cw conversion, the derivative of the envelope $E(\vec{r})$ changes only little over a propagation distance on the order of a wavelength $2\pi/k$. Inserting eqn [9] into eqns [6] and [8] and neglecting the term $\sim \mathrm{d}^2E(\vec{r})/\mathrm{d}z^2$, yields the paraxial or slowly varying envelope wave equation:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + 2\mathrm{i}k\frac{\partial}{\partial z}\right)E(\vec{r}) = -\mu_0\omega^2 P^{(2)}_{\omega}(\vec{r})\mathrm{e}^{-\mathrm{i}kz} \qquad [11]$$

The electric field envelope can be expanded in a complete set of Gaussian TEM_{mn} – mode functions ψ_{mn}:

$$E(\vec{r}) = \sqrt{\frac{\omega}{n_\omega}}\sum_{m,n=0}^{\infty} A_{mn}(z)\psi_{mn}(\vec{r}) \qquad [12]$$

These mode functions are defined to satisfy the paraxial wave equation in absence of nonlinearity (vanishing right-hand side in eqn [11]) and the orthonormality properties

$$\int\int_{-\infty}^{+\infty}\psi^*_{mn}(\vec{r})\psi_{m'n'}(\vec{r})\mathrm{d}x\,\mathrm{d}y = \delta_{mm'}\delta_{nn'} \qquad [13]$$

A particular mode function set is characterized by position (along z) and size of the waist, which can be chosen arbitrarily. The functions $A_{mn}(z)$ are slowly varying amplitude functions. The power in a particular mode can be calculated from:

$$P_{mn}(z) = \frac{\omega|A_{mn}(z)|^2}{2\mu_0 c} \qquad [14]$$

where c is the speed of light *in vacuo*. Inserting eqn [12] into eqn [11], multiplication with $\psi^*_{m'n'}(\vec{r})$, and integration over x, y yields:

$$\frac{\mathrm{d}A_{m'n'}(z)}{\mathrm{d}z} = \mathrm{i}\frac{\mu_0 c}{2}\sqrt{\frac{\omega}{n_\omega}}\mathrm{e}^{-\mathrm{i}kz}\int\int_{-\infty}^{\infty}P^{(2)}_{\omega}(\vec{r})\psi^*_{m'n'}(\vec{r})\mathrm{d}x\,\mathrm{d}y \qquad [15]$$

If all electric fields are linearly polarized, the Fourier amplitudes of the nonlinear polarization can be written as:

$$\begin{aligned}P^{(2)}_{\omega_\mathrm{s}}(\vec{r}) &= 2\varepsilon_0 d(z)E_{\omega_\mathrm{p}}(\vec{r})E^*_{\omega_\mathrm{i}}(\vec{r}) \\ P^{(2)}_{\omega_\mathrm{i}}(\vec{r}) &= 2\varepsilon_0 d(z)E_{\omega_\mathrm{p}}(\vec{r})E^*_{\omega_\mathrm{s}}(\vec{r}) \\ P^{(2)}_{\omega_\mathrm{p}}(\vec{r}) &= 2\varepsilon_0 d(z)E_{\omega_\mathrm{s}}(\vec{r})E_{\omega_\mathrm{i}}(\vec{r})\end{aligned} \qquad [16]$$

where d is the tensor element appropriate to the particular combination of linear polarizations of the three waves. (In eqn [16], $\omega_\mathrm{i} + \omega_\mathrm{s} = \omega_\mathrm{p}$.)

The mode amplitude equations for the three interacting waves follow from eqns [10], [15], and [16] as:

$$\frac{\mathrm{d}A_{mn}^{\left\{\begin{smallmatrix}\mathrm{p}\\ \mathrm{s}\\ \mathrm{i}\end{smallmatrix}\right\}}(z)}{\mathrm{d}z} = \mathrm{i}\frac{d(z)}{c}\begin{Bmatrix}\sqrt{\omega_\mathrm{p}/n_\mathrm{p}}\mathrm{e}^{-\mathrm{i}\Delta kz} \\ \sqrt{\omega_\mathrm{s}/n_\mathrm{s}}\mathrm{e}^{+\mathrm{i}\Delta kz} \\ \sqrt{\omega_\mathrm{i}/n_\mathrm{i}}\mathrm{e}^{+\mathrm{i}\Delta kz}\end{Bmatrix}\int\int\mathrm{d}x\,\mathrm{d}y$$
$$\times\begin{Bmatrix}\mathrm{E_s}(\vec{r})\mathrm{E_i}(\vec{r}) \\ \mathrm{E_p}(\vec{r})\mathrm{E_i}(\vec{r})^* \\ \mathrm{E_p}(\vec{r})\mathrm{E_s}(\vec{r})^*\end{Bmatrix}\psi_{mn}^{\left\{\begin{smallmatrix}\mathrm{p}\\ \mathrm{s}\\ \mathrm{i}\end{smallmatrix}\right\}}(\vec{r})^{*+\infty} \qquad [17]$$

The quantity $\Delta k = k_\mathrm{p} - k_\mathrm{s} - k_\mathrm{i}$ is the wavevector mismatch. Here, the envelope of each wave has an expansion of the form in eqn [12] with its respective (independent) set of mode functions.

Appendix B

Steady-State Description of the cwOPO

The general eqns [17] are applicable to any type of cwOPO. Below and at threshold at least two of the three waves have given modes, i.e., the corresponding function $E(\vec{r})$ is known up to a slowly z-dependent amplitude: the resonant parametric wave has a mode defined by the cavity geometry and mirrors, while the pump wave mode is defined by the used focusing optics. This knowledge allows to obtain an analytic expression for the threshold of any OPO type and for an arbitrary cavity.

Above threshold two cases can be distinguished:

(i) in a basic SRO the pump mode itself becomes a function of pump power, due to its depletion by signal and idler wave, and only a numerical solution can be given; and
(ii) if two or more waves resonate in one or more cavities (DRO, PR-SRO, TRO), an analytic solution is possible even above threshold.

In the following, the PR-SRO is treated. Figure 4 shows a schematic of a common cavity PR-SRO and the notation; the treatment below also holds for the dual cavity PR-SRO. Without loss of generality, the signal wave is taken to be the signal. For the sake of simplicity, below a further assumption about the spatial mode of the idler will be made.

The pump and signal waves are given by TEM_{00} modes of the cavity (or cavities):

$$E_p(\vec{r}) = \sqrt{\frac{\omega_p}{n_p}} A_{00p}(z)\psi_{00p}(\vec{r})$$
$$E_s(\vec{r}) = \sqrt{\frac{\omega_s}{n_l}} A_{00s}(z)\psi_{00s}(\vec{r}) \qquad [18]$$

In the case of a common-cavity PR-SRO, pump and signal resonate in the same cavity and therefore their modes have the same waist position and Rayleigh ranges $z_R = kw^2/2$, (w is the waist). In the case of a dual-cavity PR-SRO the modes are in principle independent.

Although the idler wave is in general not purely TEM_{00}, it may be approximated by a TEM_{00} mode, whose mode parameters are determined by an optimization criterion. We can then introduce the overlap function:

$$O(z) = \int\int dx\, dy (\psi_{00p}(\vec{r}))^* \psi_{00s}(\vec{r})\psi_{00i}(\vec{r})_{-\infty}^{+\infty} \qquad [19]$$

Under these assumptions, the infinite set of mode eqns [17] reduces to just three coupled equations ($A \equiv A_{00}$):

$$\frac{dA_p(z)}{dz} = -i\xi(z)A_s(z)A_i(z)e^{-i\Delta kz}O(z) \qquad [20]$$

$$\frac{dA_s(z)}{dz} = -i\xi(z)A_p(z)A_i^*(z)e^{i\Delta kz}O(z)^* \qquad [21]$$

$$\frac{dA_i(z)}{dz} = -i\xi(z)A_p(z)A_s^*(z)e^{i\Delta kz}O(z)^* \qquad [22]$$

where

$$\xi(z) = \frac{d(z)}{c}\sqrt{\frac{\omega_p\omega_s\omega_i}{n_p n_s n_i}} \qquad [23]$$

Two phasematching cases have to be considered: for birefringent phasematching, $d = \text{const}$ and ξ is independent of z. In QPM, the d-coefficient is modulated periodically along z with period Λ. It suffices to consider only the particular term of its Fourier series expansion that gives the smallest wavevector quasi-mismatch. When only the first term is relevant, this is referred to as first-order QPM and $d(z) \simeq \frac{1}{2}\hat{d}(g_1 e^{i2\pi z/\Lambda} + \text{c.c.})$. This case can be taken into account by replacing $d(z)$ in eqn [23] by $d_{eff} = \frac{1}{2}\hat{d}g_1$, and in eqns [20]–[22] the relevant wavevector mismatch becomes $\Delta k = k_p - k_s - k_i \pm 2\pi/\Lambda$, where, for usual materials, the minus sign applies.

We solve eqns [20]–[22] by expanding in powers of the nonlinearity ξ. This is appropriate due to the small gain and implies that the resonant pump and signal wave amplitudes do not change substantially (relative to the respective average amplitude) along the crystal length and around the cavity.

Integrating eqn [22] after approximating the amplitudes on the right-hand side by their values at $z = 0$ yields:

$$A_i(z) = -i\xi z A_p(0)A_s^*(0)I^*(z)$$
$$I(z) = \frac{1}{z}\int_0^z dz' O(z')e^{-i\Delta kz'} \qquad [24]$$

We insert this result into eqns [20] and [21] and approximate $A_s(z)$ and $A_p(z)$ by their values at $z = 0$. The result is

$$A_p(z) = A_p(0) - \frac{1}{2}(\xi z)^2 D^*(z)A_p(0)|A_s(0)|^2 \qquad [25]$$

$$A_s(z) = A_s(0) + \frac{1}{2}(\xi z)^2 D(z)A_s(0)|A_p(0)|^2 \qquad [26]$$

$$D(z) = \frac{2}{z^2}\int_0^z \mathrm{d}z' z' I(z') O(z')^* \, \mathrm{e}^{\mathrm{i}\Delta k z'} \qquad [27]$$

This completes the calculation of the propagation through the nonlinear medium. Now we proceed to calculate the propagation around the cavity.

Since the signal wave is resonant with the cavity, we can immediately write the self-consistency equation for the roundtrip:

$$|A_\mathrm{s}(0)| = \sqrt{R_\mathrm{s}(1 - V_\mathrm{s})}|A_\mathrm{s}(L)| \qquad [28]$$

The roundtrip cavity reflectivity and roundtrip loss for the signal wave are denoted by R_s and V_s. At the input mirror (see lower part of Figure 4), characterized by a transmission $T_\mathrm{p} = 1 - R_\mathrm{p}$, the interference of the injected pump wave $A_\mathrm{p,in}$ and the internally circulating pump wave must be considered:

$$A_\mathrm{p}(0) = \sqrt{T_\mathrm{p}}A_\mathrm{p,in} + \sqrt{R_\mathrm{p}(1 - V_\mathrm{p})}A_\mathrm{p}(L)\mathrm{e}^{\mathrm{i}\phi} \qquad [29]$$

V_p is the pump roundtrip loss. Evaluating eqn [26] at $z = L$ and comparing with eqn [28] yields:

$$|A_\mathrm{p}(0)|^2 = \left[\frac{1}{\sqrt{R_\mathrm{s}(1 - V_\mathrm{s})}} - 1\right]\frac{1}{\frac{1}{2}(\xi L)^2 \mathrm{Re}\, D(L)} \qquad [30]$$

Here the imaginary part of $D(z)$ was neglected since $|\xi L A_\mathrm{p}(0)|^2 \ll 1$. Expression [30] shows that the intracavity pump power is clamped at a fixed value, equal to the value at threshold, independent of the input power.

In eqn [29] we can assume $A_\mathrm{p,in}$ to be real without loss of generality, yielding the relation:

$$\sqrt{T_\mathrm{p}}\frac{A_\mathrm{p,in}}{A_\mathrm{p}(0)} = 1 - \sqrt{R_\mathrm{p}(1 - V_\mathrm{p})}\mathrm{e}^{\mathrm{i}\phi}\frac{A_\mathrm{p}(L)}{A_\mathrm{p}(0)} \qquad [31]$$

where $\phi = \omega_\mathrm{p}\mathrm{OPL}(\omega_\mathrm{p}, T)/c$. The pump wave is assumed to be exactly resonant; this means that $A_\mathrm{p}(0)$ is maximum. The phase ϕ must therefore be such that:

$$\mathrm{e}^{\mathrm{i}\phi}\frac{A_\mathrm{p}(L)}{|A_\mathrm{p}(0)|} \in \mathbb{R} \qquad [32]$$

Together, with eqn [25], we obtain:

$$\left|\frac{A_\mathrm{p}(L)}{A_\mathrm{p}(0)}\right|^2 = \left|1 - \frac{1}{2}(\xi L)^2 D^*(L)|A_\mathrm{s}(0)|^2\right|^2 \qquad [33]$$

The right-hand side is evaluated neglecting the contribution proportional to ξ^4; inserting the result in eqns [31] and [32] gives:

$$\sqrt{T_\mathrm{p}}\frac{A_\mathrm{p,in}}{|A_\mathrm{p}(0)|} = 1 - \sqrt{R_\mathrm{p}(1 - V_\mathrm{p})} + \sqrt{R_\mathrm{p}(1 - V_\mathrm{p})}\frac{1}{2}(\xi L)^2 \mathrm{Re}\, D|A_\mathrm{s}(0)|^2 \qquad [34]$$

Equations [30] and [34] represent the steady-state solution of the PR-SRO: they give the circulating pump and signal powers, from which all desired quantities can be computed.

The external threshold is found by setting $A_\mathrm{s}(0) = 0$ and assuming small losses, T_p, $V_\mathrm{p} \ll 1$

$$P_\mathrm{p,in}^\mathrm{th} \simeq \frac{T_\mathrm{p} + V_\mathrm{p}}{T_\mathrm{p}}\cdot(T_\mathrm{p} + V_\mathrm{p})\cdot\frac{(T_\mathrm{s} + V_\mathrm{s})}{E} \qquad [35]$$

with the nonlinearity

$$E = \frac{2\mu_0 d^2}{\pi c^2}\frac{\omega_\mathrm{s}\omega_\mathrm{i}}{n_\mathrm{p}n_\mathrm{s}n_\mathrm{i}}L^2\,\mathrm{Re}\,D(L) \qquad [36]$$

Note that $\mathrm{Re}\,D(L) = |I(L)|^2$. The second factor in eqn [35] is responsible for the reduction of threshold due to pump resonance. The last factor is the internal pump threshold and constant circulating pump power level:

$$P_\mathrm{p}(0) \simeq \frac{(T_\mathrm{s} + V_\mathrm{s})}{E} \qquad [37]$$

which is also equal to the threshold of a basic SRO. The nonlinearity may be written as

$$E \simeq \frac{4\mu_0 d^2}{\pi c^2}\frac{\omega_\mathrm{s}^2\omega_\mathrm{i}^2}{n_\mathrm{p}^2\omega_\mathrm{p}}Lh_\mathrm{PA} \qquad [38]$$

with the dimensionless focusing function h_PA. A detailed calculation shows that the value of h_PA cannot exceed 1.1 and is close to its maximum when the phase mismatch ΔkL is approximately zero and the pump and signal focusing by the cavity is such that their Rayleigh range $z_\mathrm{R} \simeq L/2$.

The OPO will select the signal cavity mode frequency ω_s and idler spatial mode that minimizes the threshold eqn [35]. Note that this is not necessarily equivalent to a maximization of E, since T_s and V_s may also be dependent on signal frequency, for example, because of the presence of an etalon inside the cavity.

The circulating signal power as derived from eqns [30] and [34], is given in eqn [6].

The signal power extracted from the cavity is $P_\mathrm{s,out} = T_\mathrm{s}P_\mathrm{s}(L) \simeq T_\mathrm{s}P_\mathrm{s}(0)$, while the (fully extracted) idler power is given by:

$$P_\mathrm{i,out} = P_\mathrm{i}(L) = \frac{\omega_\mathrm{i}}{\omega_\mathrm{s}}EP_\mathrm{s}(0)P_\mathrm{p}(0) \qquad [39]$$

See also

Nonlinear Optics, Applications: Phase Matching. **Optical Parametric Devices:** Overview.

Further Reading

Breitenbach G, Schiller S and Mlynek J (1995) 81% conversion efficiency in frequency-stable continuous-wave parametric oscillation. *Journal of the Optical Society of America B* 12: 2095–2101.

Ebrahimzadeh M and Dunn MH (2000) Optical parametric oscillators. In: Bass M, Enoch JM, van Stryland EW and Wolfe WL (eds) *Handbook of Optics IV*, pp. 2201–2272. New York: McGraw-Hill.

Ebrahimzadeh M, Turnbull GA, Edwards TJ, Stothard DJM, Lindsay ID and Dunn MH (1999) *Journal of the Optical Society of America B* 16: 1499–1511.

Gibson GM, Dunn MH and Padgett MJ (1998) Application of a continuously tunable, cw optical parametric oscillator for high-resolution spectroscopy. *Optics Letters* 22: 40–43.

Klein ME, Laue CK, Lee D-H, Boller K-J and Wallenstein R (2000) Diode-pumped singly resonant continuous-wave optical parametric oscillator with wide continuous tuning of the near-infrared idler wave. *Optics Letters* 25: 490–493.

Klein ME, Gross P, Boller K-J, *et al.* (2003) Rapidly tunable continuous-wave optical parametric oscillator pumped by a fiber laser. *Optics Letters* 28: 920–923.

Müller F, Popp A, Kühnemann F and Schiller S (2003) Transportable, highly sensitive photoacoustic spectrometer based on a continuous-wave dual-cavity optical parametric oscillator. *Optics Express* 11: 2820–2825.

Schiller S, Schneider K and Mlynek J (1999) *Journal of the Optical Society of America B* 16: 1512–1524.

Schreiber G, Hofmann D and Grundkötter W, *et al.* (2001) Nonlinear integrated optical frequency converters with periodically poled Ti:$LiNbO_3$ waveguides. *Proceeding SPIE* 4277: 144.

Special issue (1998) Continuous-wave optical parametric oscillators: materials, devices, and applications. *Applied Physics B* 66(6).

Strössner U, Meyn J-P, Wallenstein R, *et al.* (2002) Single-frequency continuous-wave optical parametric oscillator system with an ultrawide tuning range of 550 to 2830 nm. *Journal of the Optical Society of America B* 19: 1419–1424.

Yariv A and Yeh P (2002) *Optical Waves in Crystals*. Hoboken, NJ: Wiley-Interscience.

Optical Parametric Oscillators (Pulsed)

H Giessen and X P Zhang, University of Bonn, Bonn, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Advantages and Applications of Ultrashort Pulsed OPOs

The advantages of ultrashort pulsed optical parametric oscillators (OPOs) are twofold: on the one hand, it is possible to generate tunable laser radiation with a pulse duration in the femtosecond and picosecond time domain; on the other hand, the short pulses give rise to a high peak intensity of the electric field of the pump wave inside the nonlinear medium. Due to the fact that the OPO single-path gain coefficient reads as:

$$\eta = \frac{P_{\text{Signal}}}{P_{\text{Pump}}} = \frac{8\pi^2 d_{\text{eff}}^2 L^2 I_{\text{P}}}{\varepsilon_0 n_{\text{P}} n_{\text{S}} n_{\text{I}} c \lambda_{\text{S}}^2} \frac{\sin^2(\Delta k L/2)}{(\Delta k L/2)^2} \qquad [1]$$

where I_{P} is the pump intensity, Δk is the phase mismatch between pump, signal, and idler, n_{P}, n_{S}, n_{I}, are the refractive indices at the pump, signal, and idler wavelengths, respectively, and L is the length of the nonlinear medium with an effective nonlinear coefficient d_{eff}, the efficiency grows proportional to the intensity of the pump.

Ultrashort pulsed OPOs can provide widely tunable femtosecond or picosecond pulses, covering the spectral range from the visible to the mid-infrared, which have many applications, such as time-resolved spectroscopy, pump-probe measurements, semiconductor analysis, photochemistry, optical ranging, data transfer, and data processing with high carrier frequency, white-light generation, etc.

Synchronous Pumping

The key issue in operating ultrashort-pulse OPOs is synchronization between the pump pulses and the signal pulses. Consider the following sequence: a pump pulse enters the OPO cavity and generates spontaneously downconverted light at the signal wavelength (see Figure 1, inset). These signal photons travel inside the OPO cavity, entering the nonlinear medium again after exactly one cavity round trip. This light will act as seed light for another downconversion process if a second pump pulse enters the nonlinear medium at exactly the same instance (see Figure 1). The presence of the signal pulse will lead to a much more efficient stimulated downconversion process.

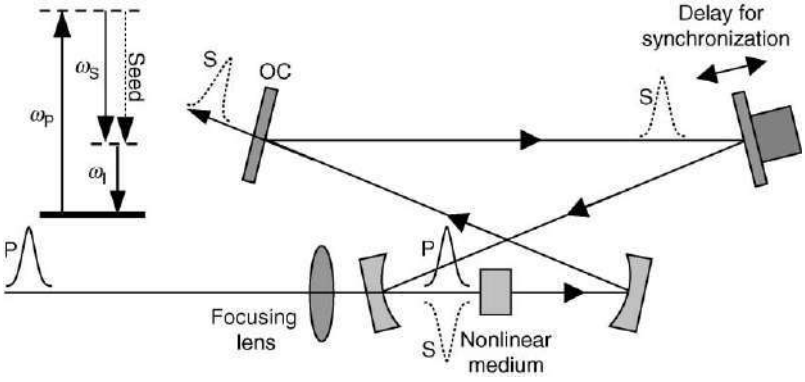

Figure 1 Synchronous pumping of a typical ring-cavity OPO: P, pump pulses; S, signal pulses; OC, output coupler; ω_P, ω_S, ω_I, frequencies of the pump, the signal, and the idler, respectively. Inset: energy level diagram for optical parametric oscillation.

This requires the repetition rate of the pump oscillator and the OPO to be synchronized precisely. Therefore, this pump scheme is called *synchronized pumping*. If the temporal separation between the pump pulses does not exactly equal the roundtrip time of the signal pulses inside the OPO cavity, the generation efficiency will decrease dramatically as the temporal pulse overlap is reduced.

Dispersion in Nonlinear Media and Optical Glasses

Maintaining synchronicity between the pump pulses and the signal pulses is a necessary, but not sufficient prerequisite for successful operation of an ultrashort-pulse OPO. The signal pulses will be temporally broadened during their repeated trip through the nonlinear medium. This will lead to problems, as the newly generated signal pulses are still rather short (as they have been generated by a short pump pulse), but the previously generated signal pulses, which should act as seed pulses and ideally have a similar pulse duration as the pump pulses, have already been broadened temporally, due to dispersion in the nonlinear medium and other intracavity optical elements. Meanwhile, temporal walk-off between the signal (idler) and the pump pulses will take place, resulting in reduced conversion efficiency and broadened signal pulses. Group velocity dispersion (GVD) and group velocity mismatch (GVM) are normally used to characterize the dispersion in nonlinear medium.

The group velocity of the laser pulses propagating in a nonlinear optical medium is defined as:

$$v_G = \frac{c}{n - \lambda \dfrac{dn}{d\lambda}} \qquad [2]$$

where c is the velocity of light in vacuum, n is the refractive index of the nonlinear medium at the center wavelength λ of the laser pulses. Then the GVD can be written as:

$$\frac{dv_G}{d\lambda} = \frac{\lambda v_G^2}{c}\frac{d^2n}{d\lambda^2} \qquad [3]$$

In most cases, we evaluate the GVD of the optical materials using the group velocity dispersion parameter (k''), which is defined as:

$$k'' = \frac{d^2k}{d\omega^2} = \frac{\lambda^3}{2\pi c^2}\frac{d^2n}{d\lambda^2} \qquad [4]$$

where $k = n\omega/c$ is the wave vector, ω is the frequency of the incident wave, and consequently:

$$\frac{dv_G}{d\lambda} = \frac{2\pi c v_G^2}{\lambda^2}\frac{d^2k}{d\omega^2} \qquad [5]$$

Due to GVD in the nonlinear medium, the spectral components in the broadband spectrum of an ultrashort pulse will travel at different group velocities and will consequently be delayed differently in the crystal, therefore the pulses will become temporally broadened or become chirped. This linear chirp induced pulse broadening effect can be characterized quantitatively by the following formula:

$$\tau_P = \tau_{P0}\sqrt{1 + \left(\frac{z}{z_d}\right)^2} \qquad [6]$$

where τ_{P0} is the incident pulse duration, τ_P is the pulse duration after the pulse propagates in the nonlinear medium over a distance of z, and z_d is called the characteristic length and defined as:

$$z_d = \frac{\tau_{P0}^2}{4\ln 2|k''|} \qquad [7]$$

In frequency conversion processes, such as second-harmonic generation (SHG), sum-frequency

generation (SFG), difference frequency generation (DFG), optical parametric oscillation (OPO), or optical parametric amplification (OPA), three waves interact with each other while propagating through the nonlinear medium. Due to dispersion of the material, different waves propagate at different group velocities. This effect is generally called group velocity mismatch. In cases of interaction on the femtosecond scale, the interacting pulses may get separated after propagating some distance in the medium, which means that there is a reduced effective interaction length. The GVM is a quantitative evaluation of this effect, which is defined as:

$$\Delta v_G = \left(\frac{1}{v_{G,i}} - \frac{1}{v_{G,j}} \right)^{-1} \qquad [8]$$

where $v_{G,i}$ and $v_{G,j}$ are the group velocities of the two interacting waves i and j. The effective interaction length L_{eff} for wave i and wave j can thus be calculated by the following formula:

$$L_{eff} = \tau_P |\Delta v_G| = \tau_P \left| \left(\frac{1}{v_{G,i}} - \frac{1}{v_{G,j}} \right)^{-1} \right| \qquad [9]$$

where τ_P is the incident pulse duration at full width at half maximum (FWHM). It is assumed that wave i and wave j have a similar pulse duration. This temporal walk-off will lead to two effects. First, the interacting pulses will get separated and the conversion efficiency will be reduced. The second effect is the broadening of the signal pulses. In the case of sub-100 fs operation, to restrict the broadening in pulse length to less than ~10%, the crystal length should be smaller than $L_{eff}/2$.

Figure 2 shows the GVD of some typical OPO crystals, including KTP, KTA, RTA, CTA, PPLN, PPKTP, and LBO. Their Sellmeier equations are as summarized in Appendix A. In addition to the wavelength of the incident beam, the GVD is also dependent on the polarization of the incident beams and the type of phase matching.

We give the GVM values between the pump and the signal (P–S), the signal and the idler (S–I), and between the pump and the idler (P–I) for a KTP crystal in Figure 3.

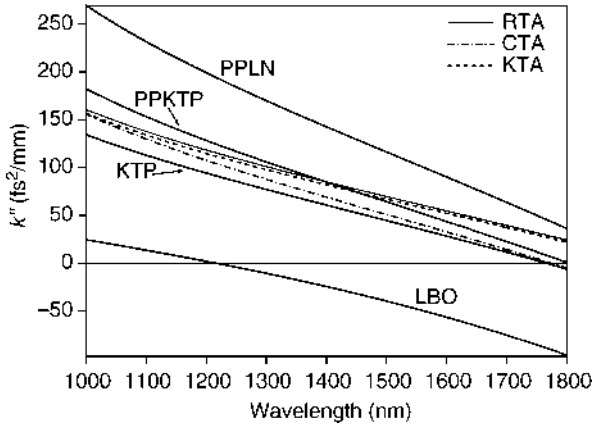

Figure 2 Group velocity dispersion characteristics of some typical OPO crystals.

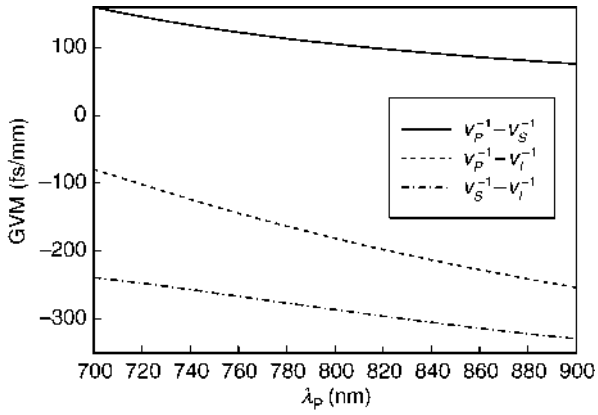

Figure 3 GVM between the interaction waves of a KTP OPO versus pump wavelength (λ_P).

Intracavity Dispersion Compensation and Extra Cavity Pulse Compression

In order to tackle the problem of intracavity dispersion, corrective elements have to be placed inside the cavity in order to compensate the chirp. There have been several solutions to this problem, namely the combination of intracavity prism pairs, chirped mirrors, chirped grating periods (for quasi-phase matching) with oppositely chirped pump pulses, and glass plate dispersion compensators. Certainly, a shorter crystal length will yield a smaller GVM and GVD, but will also give a smaller nonlinear conversion efficiency due to the shorter crystal length (see eqn [1]).

Pump Sources

As pump pulses, mode-locked femtosecond or picosecond lasers are used. The requirements on pulse durations are determined by the desired pulse durations of the OPO signal and by the phase-matching bandwidth of the correspondingly used crystal. Shorter pump pulses normally lead to higher pump power thresholds due to shorter temporal overlap between the interacting pulses. It can also lead to higher instability of the OPO operation due to its higher sensitivity to the cavity length fluctuations. Generally, larger pump bandwidth means larger bandwidth of the signal. Consequently, a potentially shorter signal pulse duration can be achieved if the

intracavity dispersion is perfectly compensated. Sometimes it is advantageous for stability to operate the OPO in the slightly negative total GVD regime. In a quasi-phase-matched (QPM) OPO, the phase-matching bandwidth is so large that only a narrow pump bandwidth is required to phase-match a large bandwidth of the signal. Therefore, a pump depletion hole can be commonly observed with the oscillation of the OPO.

Design of Ultrashort Pulsed OPOs

Design of an OPO device may involve the design of the cavity length, cavity mirrors, focal length for focusing the pump beam, intracavity dispersion compensation scheme, frequency stabilization module, crystal length, and some other design depending on the special configuration of the OPO, for example, a heating module (an oven) is required for a PPLN OPO and temperature control module is required for a temperature phase-matched LBO OPO. For an ultrashort pulsed OPO, the most important designs include the design of the crystal length, the design of intracavity dispersion compensation components, and the frequency stabilization control.

Crystal Length

Several factors have to be taken into account when designing a crystal length. The design of a crystal length should be a compromise of all of these considerations. What is special for a femtosecond or picosecond OPO is the dispersion and temporal walk-off between the interacting pulses in the crystal:

1. Effective interaction length: Pump pulse duration, estimated signal pulse duration, temporal walk-off between the pump and the signal give a limit of the crystal length that is characterized by eqn [9].
2. Pulse broadening due to GVM and GVD from the OPO crystal and other intracavity optical glass elements should be estimated. It is reasonable that shorter crystal length is always favorable for ultrashort pulse generation. The crystal length should be as short as possible while maintaining an acceptable conversion efficiency. When designing the crystal length, the available dispersion compensation device should be considered simultaneously.

Dispersion Compensation

For a crystal with a given material and length, the intracavity dispersion can be calculated. The compensation device should introduce an amount of negative GVD such that the total intracavity GVD is equal to or slightly smaller than zero at the desired center wavelength of the signal pulses. Meanwhile, the compensation should be relatively uniform (for instance having a GVD fluctuation of $<50\ \mathrm{fs}^2/\mathrm{mm}$) over the bandwidth of the expected signal spectrum. Then a compensation method (as described above) should be chosen and corresponding device should be designed.

Piezo-Driven Mirror and Frequency Stabilization

To achieve fine tuning of the cavity length, one of the cavity mirrors should be mounted onto a piezoelectric translation stage. When feeding back the signal output of the OPO to this piezo-driven mirror, forming a closed loop to compensate for the drift of the cavity length, a stabilization of the OPO repetition rate is achieved. If the signal is dispersed by a grating, a change in cavity length will also lead to a change in the output spectrum. Using a split photodiode, the difference signal of the dispersed spectrum can be utilized as the measured signal in the feedback loop.

Tuning Characteristics and Methods

Tunability is one of the most important advantages of the OPO over other ultrashort pulsed lasers. A large variety of tuning methods make the OPO more practical and versatile in many applications. For birefringently phase-matched OPOs, the tuning of the signal wavelength can be achieved through changing the pump wavelength (pump wavelength tuning), using an intracavity wavelength selection element (such as a birefringent filter), changing the temperature of the crystal when its refractive index is sensitive to the temperature change (temperature tuning), or changing the orientation of the OPO crystal (angle tuning). Some more methods can be utilized for a QPM OPO, for example through changing the poling period of the crystal (grating period tuning). A special method for an ultrashort pulsed OPO is the so-called cavity length tuning.

As has been discussed above, in synchronous pumping, the OPO must have the same cavity length as the pump laser, so that the signal pulse arrives simultaneously with the pump pulse at the crystal to get the most efficient interaction. Because of dispersion of the nonlinear crystal, the signal will be delayed differently at different wavelengths. In other words, signal pulses have different round-trip times in the cavity at different center wavelengths. When the cavity length of the OPO is changed slightly, the signal has to change its wavelength to keep constant optical path length or constant round-trip time. Cavity length tuning can be achieved easily by tuning

the voltage on a piezoelectric transducer (PZT), onto which one of the cavity mirrors is mounted. Therefore, this method is the most common and frequently used to tune the signal wavelength of a synchronously pumped OPO. This is also the basis for stabilizing the operation of a synchronously pumped OPO.

High-Repetition-Rate (HRR) Operation

Synchronously pumping an OPO at high repetition rate produces a high-repetition-rate signal pulse train which extends the application fields of the OPO. Using high-repetition-rate laser pulses (e.g., >1 GHz) as the pump is a straightforward method to run the OPO at a high repetition rate. But a high repetition rate means an extremely short cavity length. For instance, the cavity length is about 30 cm at 1 GHz repetition rate. Such a short cavity length leads to a large mode of the signal beam waist in the crystal. Mode matching between the signal and the pump beams requires that the transverse mode of the pump beam should be slightly larger than that of the signal beam within the interaction zone to maintain stable and efficient oscillation of the OPO. Large modes of the signal and the pump beams mean low interacting intensities, increasing the pump threshold significantly. Using smaller curvature radius of the curved mirrors and shorter focal length of the focusing lens might reduce the mode sizes at the focus in the nonlinear medium, but this will introduce extremely serious astigmatism when the signal beam is reflected from a strongly tilted curved mirror with a quite small curvature radius. In addition to exact synchronous pumping, there are at least two methods to realize high-repetition-rate operation of an OPO. First, pumping at low repetition rate while using a shorter cavity length of the OPO (L_{OPO}) than the pump laser (L_P) with $L_{OPO} = L_P m/N$, where m and N are both integers and $m < N$, as demonstrated in Figure 4a, which is called lower-order synchronous pumping. However, applying this technique, the signal pulses experience more round-trip losses, also resulting in a much higher pump threshold and periodically varying intensities of the output pulses. Second, pumping the OPO at high repetition rate while using a much longer cavity length of the OPO than the pump laser with $L_{OPO} = L_P N/m$, where m and N are both integers and $m < N$, is called higher-order pumping, as demonstrated in Figure 4b. In higher-order synchronous pumping, the signal pulse train has the same repetition rate as the pump, while the mode-matching problem is removed through a long OPO cavity.

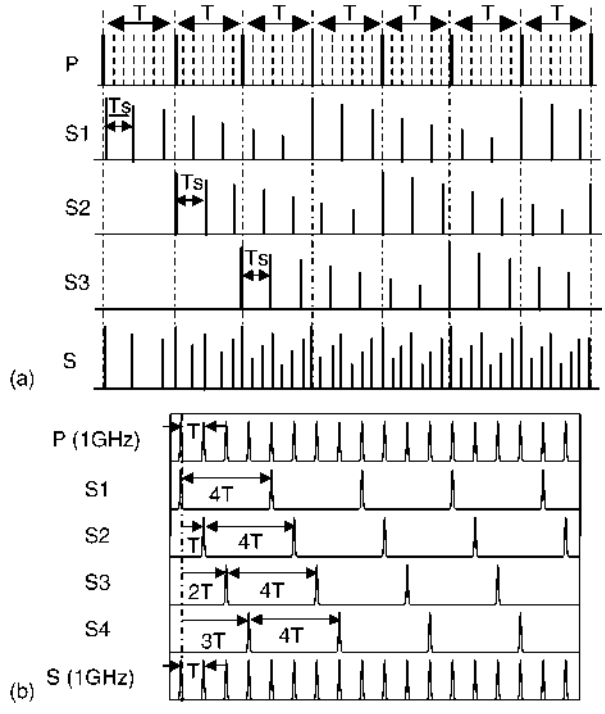

Figure 4 (a) 3/7-order synchronous pumping. Three signal pulse trains (S1–S3) are produced, each having a repetition period (Ts) 3/7 that of the pump pulse train (P): Ts = 3/7 T. The pump repetition rate is multiplied by a factor of 7 in the signal pulse train (S). (b) Fourth-order synchronous pumping at 1 GHz. The signal pulse train (S) is a combination of four pulse trains (S1–S4), each having a repetition period (4T) 4 times that of the pump pulse train (P).

See also

Optical Parametric Devices: Optical Parametric Oscillators (Continuous Wave); Overview.

Appendix A

Sellmeier Equations for some Typical Crystals and Optical Glasses

In the equations, n denotes the refractive index; λ is the wavelength of incident light in μm; $A, B, C, D, A_1, A_2, A_3, A_4, A_5, A_6, B_1, B_2, B_3, B_4, \ldots$, are the Sellmeier coefficients; the subscripts X, Y, Z, stand for the directions of the principal axes of the biaxial crystals.

KTP

$$n^2 = A + \frac{B}{1 - \left(\frac{C}{\lambda}\right)^2} - D\lambda^2$$

	A	B	C	D
n_X	2.1146	0.89188	0.20861	0.01320
n_Y	2.1518	0.87862	0.21801	0.01327
n_Z	2.3136	1.00012	0.23831	0.01679

KTA

$$n^2 = A + \frac{B}{\lambda^2 - C} - D\lambda^2$$

	A	B	C	D
n_X	3.1413	0.04683	0.04055	0.01023
n_Y	3.1593	0.04828	0.04710	0.01049
n_Z	3.4435	0.06571	0.05435	0.01460

RTA

$$n^2 = A + \frac{B}{1 - \left(\frac{C}{\lambda}\right)^2} - D\lambda^2$$

	A	B	C	D
n_X	2.04207	1.17785	0.20157	0.01035
n_Y	2.14941	1.09267	0.21462	0.01067
n_Z	2.18962	1.30103	0.22809	0.01390

CTA

$$n^2 = A + \frac{B}{1 - \left(\frac{C}{\lambda}\right)^2} - D\lambda^2$$

	A	B	C	D
n_X	2.34498	1.04863	0.22044	0.01483
n_Y	2.74440	0.70733	0.26033	0.01526
n_Z	2.53666	1.10600	0.24988	0.01711

$LiNbO_3$

$$n_e^2 = A_1 + B_1 f + \frac{A_2 + B_2 f}{\lambda^2 - (A_3 + B_3 f)^2} + \frac{A_4 + B_4 f}{\lambda^2 - A_5^2} - A_6\lambda^2$$

$$f = (T - T_0)(T + T_0 + 546) \text{ and } T_0 = 24.5\,^\circ\text{C}$$

Coefficients	*Values*
A_1	5.35583
A_2	0.100473
A_3	0.20692
A_4	100
A_5	11.34927
A_6	1.5334×10^{-2}
B_1	4.629×10^{-7}
B_2	3.862×10^{-8}
B_3	-0.89×10^{-8}
B_4	2.657×10^{-5}

LBO

$$n^2 = A + \frac{B}{\lambda^2 + C} + D\lambda^2$$

	A	B	C	D
n_X	2.4543	1.1413×10^{-2}	-9.4981×10^{-3}	-1.3900×10^{-2}
n_Y	2.5382	1.2830×10^{-2}	-1.1387×10^{-2}	-1.7034×10^{-2}
n_Z	2.5854	1.3065×10^{-2}	-1.1617×10^{-2}	-1.8146×10^{-2}

$$\frac{dn_X}{dT} = 2.0342 \times 10^{-7} - 1.9697 \times 10^{-8} T - 1.4415 \times 10^{-11} T^2$$

$$\frac{dn_Y}{dT} = -1.0748 \times 10^{-5} - 7.1034 \times 10^{-8} T - 5.7387 \times 10^{-11} T^2$$

$$\frac{dn_Z}{dT} = -8.5998 \times 10^{-7} - 1.5476 \times 10^{-7} T + 9.4675 \times 10^{-10} T^2 - 2.2375 \times 10^{-12} T^3$$

BBO

$$n_o^2 = 1.7018379 + \frac{1.0357554\lambda^2}{\lambda^2 - 0.01800344} + \frac{1.2479989\lambda^2}{\lambda^2 - 91}$$

$$n_e^2 = 1.5920433 + \frac{0.7816893\lambda^2}{\lambda^2 - 0.016067891} + \frac{0.8403893\lambda^2}{\lambda^2 - 91}$$

Sapphire

$$n^2 = 1 + \frac{A_1\lambda^2}{\lambda^2 - B_1} + \frac{A_2\lambda^2}{\lambda^2 - B_2} + \frac{A_3\lambda^2}{\lambda^2 - B_3}$$

A_1	A_2	A_3	B_1	B_2	B_3
1.023798	1.058264	5.280792	0.00377588	0.0122544	321.3616

Quartz

$$n_{o,e}^2 = A + \frac{B\lambda^2}{\lambda^2 - C} + \frac{D\lambda^2}{\lambda^2 - E}$$

	A	B	C	D	E
n_o	1.28604141	1.07044083	0.0100585997	1.10202242	100
n_e	1.28851804	1.09509924	0.0102101864	1.15662475	100

Fused Silica, BK7, SF2, SF5, SF10, SF11, SF18

$$n^2 = 1 + \frac{A_1\lambda^2}{\lambda^2 - B_1} + \frac{A_2\lambda^2}{\lambda^2 - B_2} + \frac{A_3\lambda^2}{\lambda^2 - B_3}$$

	A_1	A_2	A_3	B_1	B_2	B_3
Fused silica	0.6961663	0.4079426	0.8974794	0.0046791	0.01351206	97.9340025
BK7	1.03961212	0.23179234	1.01046945	0.0060007	0.0200179	103.560653
SF2	1.40301821	0.23176750	0.93905659	0.0105796	0.04932270	112.405955
SF5	1.46141885	0.24771302	0.94999583	0.0118261	0.05085946	112.041888
SF10	1.61625977	0.25922933	1.07762317	0.0127535	0.05819840	116.607680
SF11	1.73848403	0.31116897	1.17490871	0.0136069	0.06159605	121.922711
SF18	1.56441436	0.29141358	0.96030788	0.0121864	0.05355680	111.451201

Further Reading

Agrawal GP (1989) *Nonlinear Fiber Optics*. Boston, San Diego: Academic Press.

Albrecht TF, Sandmann JHH, Feldmann J, Stolz W, Göbel EO, Hillmer H, Lösch R and Schlapp W (1995) Design and application of a femtosecond optical parametric oscillator for time-resolved spectroscopy of semiconductor heterostructures. *Applied Physics* B 60: 459.

Arbore MA, Marco O and Fejer MM (1997) Pulse compression during second-harmonic generation in aperiodic quasi-phase-matching gratings. *Optics Letters* 22: 865–867.

Arigas D and Reid DT (2002) Efficient femtosecond optical parameteric oscillators based on aperiodically poled nonlinear crystals. *Optics Letters* 27: 851–853.

Bass M, Enoch JM, Van Stryland EW and Wolfe WL (2001) *Handbook of Optics: Volume IV, Fiber Optics and Nonlinear Optics*. New York: McGraw-Hill.

Beddard T, Ebrahimzadeh M, Reid TD and Sibbett W (2000) Synchronously pumped optical parametric oscillator driven by a femtosecond mode-locked fiber laser. *Optics Letters* 25: 1052.

Bierlein JD and Vanherzeele H (1997) Potassium titanyl phosphate: properties and new applications. *Journal of the Optical Society of America B* 6: 622.

Burr KC, Tang CL, Arbore MA and Fejer MM (1997) High-repetition-rate femtosecond optical parametric oscillator based on periodically poled lithium niobate. *Applied Physics Letters* 70: 3341.

Diels J-C and Rudolph W (1996) *Ultrashort Laser Pulse Phenomena: Fundamentals, Techniques, and Applications on a Femtosecond Time Scale*. San Diego, CA: Academic Press.

Dmitriev VG, Gurzadyan GG and Nikogosyan DN (1991) *Handbook of Nonlinear Optical Crystals*. New York: Springer-Verlag.

Dunn MH and Ebrahimzadeh M (1999) Parametric generation of tunable light from continuous-wave to femtosecond pulses. *Science* 286: 1513.

Fenimore DL, Schepler KL, Zelmon D, Kuck S, Ramabadran, Von Richter P, and Small D (1996) Rubidium titanyl arsenate difference-frequency generation and validation of new Sellmeier coefficients. *Journal of the Optical Society of America B* 13: 1935.

Ghost G (1995) Temperature dispersion of refractive indices in β-BaB_2O_4 and LiB_3O_5 crystals for nonlinear optical devices. *Journal of Applied Physics* 78: 6752.

Ghost G (1999) Dispersion-equation coefficients for the refractive index and birefringence of calcite and quartz crystals. *Optics Communications* 163: 95.

Hebling J, Giessen H, Linden S and Kuhl J (1997) Mirror-dispersion-compensated femtosecond optical parametric oscillator. *Optical Communications* 141: 229.

Jundt DH (1997) Temperature-dependent Sellmeier equation for the index of refraction, ne, in congruent lithium niobate. *Optics Letters* 22: 1553.

Kato K, Umemura N and Tanaka E (1997) 90° phase-matched mid-infrared parametric oscillation in undoped $KTiOAsO_4$. *Japanese Journal of Applied Physics 2* 36(4A): L403.

Loza-Alvarez P, Reid DT, Faller P, Ebrahimzadeh M and Sibbett W (1999) Simultaneous femtosecond-pulse compression and second-harmonic generation in aperiodically poled KTiOPO4. *Optics Letters* 24: 1071.

Matuschek N, Gallmann L, Sutter DH, Steinmeyer G and Keller U (2000) Back-side-coated chirped mirrors with ultra-smooth broadband dispersion characteristics. *Applied Physics B* 71: 509.

Mokhtari A, Fini L and Chesony J (1987) Efficient frequency mixing of a cw femtosecond laser synchronously pumped by a frequency doubled Nd:YAG laser. *Optics Communications* 61: 421.

Nebel A, Fallnich C, Beigang R and Wallenstein R (1997) Noncritically phase-matched continuous-wave mode-locked singly resonant optical parametric oscillator synchronously pumped by a Ti:sapphire laser. *Journal of the Optical Society of America B* 10: 2195.

Schubert M and Wilhelmi B (1978) *Nonlinear Optics and Quantum Electronics*. New York: John Wiley & Sons.

Seres J and Hebling J (2000) Nonstationary theory of synchronously pumped femtosecond optical parametric oscillators. *Journal of the Optical Society of America B* 17: 741.

Shen YR (1984) *The Principles of Nonlinear Optics*. New York: John Wiley & Sons.

Sutherland RL (1996) *Handbook of Nonlinear Optics*. New York: Marcel Dekker Inc.

Szipöcs R and Kohazi-Kis A (1997) Theory and design of chirped dielectric laser mirrors. *Applied Physics B* 65: 115.

Tang Y, Cui Y and Dunn MH (1995) Thermal dependence of the principal refractive indices of lithium triborate. *Journal of the Optical Society of America B* 12: 638.

Tomaru T and Petek H (2000) Femtosecond Cr4+ YAG laser with an L-fold cavity operating at a 1.2-GHz repetition rate. *Optics Letters* 25: 584.

Tomaru T and Petek H (2000) Thermal dependence of the principal refractive indices of lithium triborate. *Journal of the Optical Society of America B* 18: 388.

Velsko SP, Webb M, Davis L and Huang C (1991) Phase-matched generation in lithium triborate (LBO). *IEEE Journal of Quantum Electronics* 27: 2182.

Yariv A (1984) *Optical Waves in Crystals: Propagation and Control of Laser Radiation*. New York: John Wiley & Sons.

Yariv A (1989) *Quantum Electronics*. New York: John Wiley & Sons.

Zernike F and Midwinter JE (1973) *Applied Nonlinear Optics*. New York: John Wiley & Sons.

Zhang XP, Hebling J, Kuhl J, Rühle WW, Palfalvi L, Giessen H (2002) Femtosecond near-IR optical parametric oscillator with efficient intracavity generation of visible light. *Journal of the Optical Society of America B* 19: 2479.

Zhang XP, Kuhl J, Hebling J, Bartels A, Nau D, Rühle XP, Giessen H (2002) 1-GHz repetition-rate femtosecond optical parametric oscillator. *Applied Physics Letters* 80: 1873.

OPTICAL PROCESSING SYSTEMS

D Arbel and N S Kopeika, Ben-Gurion University of the Negev, Beer Sheva, Israel

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Modern optical systems find many applications in sensing, recording, storage, and transmission of data. These applications stimulate the development of many new optical devices, components, materials, and applied technologies. The basic element of an optical system is the light. The physical implications of the light in optical systems are very analogous to those of electricity, hence, light signals are quite similar to electric signals. High-speed spatial light components are being developed for more stringent applications in coherent optical processing systems, where they may be used in the input plane or in the Fourier transform plane. Research in this area is directed towards the design of robust and compact correlators.

Novel transforms for information processing (also known in electronics) include Fourier and wavelet transforms. Such optical transformations serve as the basic platform for linear optical system processing. Optical devices and systems for information processing include spatial light modulators, which are used to implement optical information processing systems. There are holographic applications in information processing, display systems, memory systems, optical computing systems, and imaging systems. Such devices and systems are based on applications of novel materials for information processing. Optical and digital techniques are used for optical system design, optical computing, image processing, and even for encryption and security systems.

The imaging system is an optical system and integrates hardware as well as software. Figure 1 presents a general model of an extensive imaging system. The starting point of the system is the object (or the light source). An imaging device, such as camera or optical sensor detects the energy radiated in electromagnetic waves from the light source. The image (or series of images – video) formed in the imaging device depends on the received irradiance of the image propagated through the imaging path and is correspondingly diminished from that in the object plane. The properties of the optical medium (transmission path) such as atmospheric properties, background light, motion, and vibration in both the object and the image planes, and the quality of the imaging device, all affect the quality of the images.

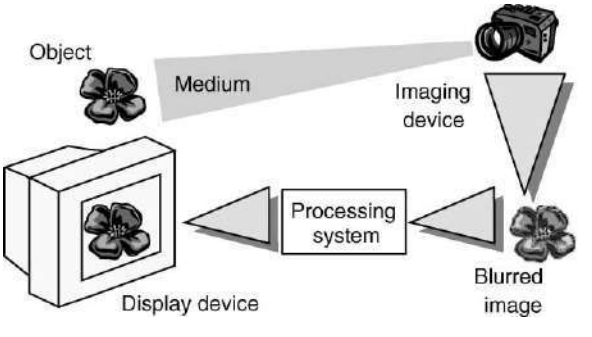

Figure 1 General description of an imaging system.

There are atmospheric phenomena that give rise to attenuation of the irradiance of the propagating image, thus reducing the contrast of the final image, as well as atmospheric phenomena that cause blurring of detail. Both types of phenomena degrade image quality.

Image blurring caused by vibration is a factor whose influence on resolution is often significant in imaging systems that involve mechanical motion. In order to obtain a well processed imaged, as similar as possible to the ideal image (the source), image processing techniques such as image enhancement, restoration, and filtering are implemented in an optical processing system which is a combination of software (algorithms) and hardware (electronic and optical components). The processing procedure involves mathematical and especially optical transformations. The output of the processing system is presented by using a display device or hardcopy. The development of imaging systems (imaging devices and displays) which include integrated optical components and processing units, assist in improving image quality at early stages, and as a result reduce the need of unique processing systems and may serve as the standalone device.

The key issues which must be addressed in the design, evaluation, and implementation of practical systems for optical processing, are provided here in a detailed review. Thus, the review is focused on imaging where considerations such as optical transformations, optical computing, effects of diffraction, scatter, absorption, and design on the performance of optical systems, imaging and the optical transfer function (OTF) for image restoration, and filtering, are all discussed in detail.

Digital Processing and Control

The Fourier transform properties of Fraunhofer diffraction permit data processing operations carried out optically that can, in many situations, be more advantageous than digital signal processing. The primary advantages of optical techniques over electronic analog and digital signal processing, are best realized when the information to be processed has two degrees of freedom. Optical systems correspondingly can have two degrees of freedom. Thus, for handling pictorial data, they can be intrinsically superior to electric or electronic systems, which have only time as an independent variable. This two-dimensional (2D) processing property can be particularly valuable in radar applications where previously velocity and position data had to be processed separately in order for sufficient accuracy in the estimation of each set of data to be obtained.

Many of the basic operations that can be performed by optical computers can also be done with electronic systems. However, if large quantities of wideband data are to be processed, the electronic techniques can often be prohibitively expensive and time consuming. The optical computer is faster than a digital computer because it operates on signals that are functions of positions, not time. The input to an optical data processor is a pattern of light irradiance that varies over some area, not light irradiance that varies as a function of time.

Thus, optical data processing can be very appropriate for such lengthy operations as the calculation of complicated Fourier transforms and the processing of side-looking radar signals. In addition to mathematical operations such as differentiation, integration, and producing Fourier transforms, the two classes of applications for which optical signal processing is most often used are filtering and correlation.

Optical Computing

Optical computing (OC) means the use of light as a primary means for carrying out numerical calculations, reasoning, artificial intelligence, etc. The history of OC is linked to that of radar systems. OC systems received a great push from the invention of laser in 1960. In the 1960s, the first schemes for all-optical digital computers were proposed. Since the 1990s, researchers have been pursuing the development of computers based on optical technology, and emphasis has shifted to optical interconnection of arrays of semiconductor smart pixels. OC provides the first in-depth review of the possibilities and limitations of optical data processing. The partial listing of those points that make OC appealing are direct image processing, massive parallelism and connectivity, immunity to electromagnetic interference (EMI), speed, size, and cost. There are three distinct trends in OC: special purpose analog optical systems, general purpose digital optical systems, and hybrid optical/electronic systems.

Innovations such as the electro-optic spatial filter, the 2D spatial light modulator, and the variable electro-optic mirror provide the fundamental building blocks for both analog optical computing (AOC) and digital optical computing (DOC) technologies. Optical techniques can provide solutions and a number of ways of extending the information processing capability of electronics. The optical processors have to be compatible with existing electronic systems. Free space digital optics is one

direction that provides some valuable solution. DOC requires the use of nonlinear optics.

All-optical processing refers to optically based processors employed nonlinear optical resources either of liquid crystal spatial light modulators (SLMs) or nonlinear interference filters (NLIFs). Electro-absorption devices are a bridge between all-optical demonstrators and optically interconnected smart pixels (logic and local on-chip interconnection is electrical whereas chip to chip interconnection is optical). Optically interconnected smart pixels have several advantages like faster data acquisition and low power consumption. A major breakthrough in nonlinear optics is essential to increase the use of optics within computing and the manufacture of optical digital computers.

An optically bistable device used for photonic switching is the self electro-optic effect device (SEED). It is the analogous component for the transistor in optical processing. A SEED consists of multiple elements, biased by an external voltage, which creates an external field that shifts the wavelength of the onset of absorption, thus varying the intensity of the transmitted light. SEEDs rely on changes in the optical transmission of a semiconductor induced by an applied electric field. SEEDs can be configured in pairs (connected electrically) so that a beam of light switching one device can cause a complementary switch in the transmission of the other. Hence, a small change in the intensity of one beam can cause a large change in intensity of the other. This kind of configuration is known as symmetric SEEDs, and acts like an electronic flip-flop and permits logic operations NAND and NOR to be carried out on pairs of optical input signals. With symmetric SEEDs, the higher the input optical power, the faster the switching speed. Cellular logic image processor (CLIP) is implemented using the symmetric SEEDs. CLIP computer architecture is designed to permit parallel information processing, in which logic operations are performed on each element of the array simultaneously.

Holography

Holography is a photographic method of recording information (two mutually coherent beams) about an object, in which one beam is the object beam and the other a reference beam, which enables the construction of objects in three dimensions. Holography relies on the encoding of the object formation in a set of complex interference fringes formed by the interaction of a plane coherent reference wave with the wavelets diffusely scattered by the object.

The microscopic interference fringes recorded on a high resolution photographic plate after development and processing become a hologram. Hologram applications assist the study of small deformations of objects, to replace conventional optical elements such as lenses and beamsplitters, for data/information storage, and use of photo-refractive materials in analog computing, object recognition, and correlation, etc.

The purpose of recording a hologram is not optical data processing but rather a lensless method of forming an image. The uniqueness of a hologram as compared to an ordinary photographic image, for example, is that the latter contains amplitude information only. A hologram, because it involves interference of the complex field amplitude (CFA) of a given scene with a reference beam, records both amplitude and phase of the CFA of the given scene as seen within its field of view (FOV) by a square-low photosensitive device such as a film.

The film is exposed within its region of linearity whereby transmission is proportional to irradiance for a given time exposure. The transparency resulting from the film is the hologram. To view the holographic image, the hologram is illuminated with a reference wave, which causes the CFA of the given scene (both amplitude and phase) to be reconstructed. Since hologram reconstruction involves the CFA of the given scene in its entirety, depth information is included as well. Furthermore, it is even possible to view the holographic scene from different physical points of view and to see around objects in the scene. This ability, however, is limited by the solid angle subtended by the object when the hologram is recorded. Hologram reconstruction cannot supply more information than that recorded in the hologram itself.

Image Processing

Figure 1 demonstrates a general behavior of an imaging process. The 'Processing System' block in Figure 1 is expanded and presented in more detail by Figure 2. Figure 2 demonstrates in a flow chart, a general process occurring within the processing system. The input to the processing system is a degraded image (or set of images or video signal) such as blurred image, noisy image, image with clutters, multi-spectral image, etc. The input depends on exterior (environmental conditions) and interior (optical, electronic, and mechanical) effects of the imaging system. It may be generated as an analog signal or digital signal. As a result, the output of the imaging system is expected to be a processed (filtered, enhanced, estimated, restored, cleared, etc.) image.

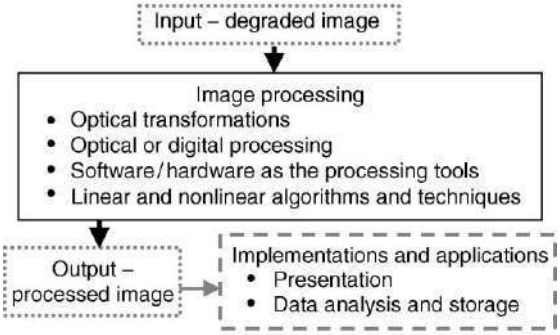

Figure 2 Image processing.

Implementations and applications depend on the output obtained by the imaging system. The output can be presented by a display device and/or hardcopy. The output can be stored for future assignments and may also serve for data analysis such as detection, recognition, and identification of patterns and other details of interest (mapping, target acquisition, defects and imperfections in the human body, meteorology, security, intelligence, etc.).

Spatial units are reciprocal to time, and are cycles/seconds or Hz. The spatial units are also reciprocal to the domain of the information being sought. Spatial signals are 2D or even 3D on occasions. 2D spatial transforms may actually exist physically in the spatial plane. Novel transforms for information processing include Fourier and wavelets transforms. Such optical transformations serve as the basic platform for linear optical system processing. Image processing techniques are based on the image spatial frequencies. In most techniques the degraded image is transformed using an optical transformation, as a primary stage before any other process is being conducted. Optical (inverse) transformations also serve as the last processing tool before the output. Practically, optical transformations are involved and integrated in most processes related to image processing.

Image processing refers to both digital and optical image processing, and also considers algorithms and techniques, which are implemented by an integration of software and hardware. Image processing is an extremely broad field. It considers digital image processing, 2D and 3D image processing and pattern recognition, real-time processing, neural networks, image enhancement/improvement, image restoration, estimation, filtering for noise removal, linear and nonlinear filtering techniques for variety of implementations, optimum filtering, etc. Image processing also include varieties of algorithms such as linear and nonlinear algorithms, algorithms for pattern recognition, linear and nonlinear techniques for pattern recognition, neural networks based algorithms, etc. Both optical and electronic implementations (hardware) are still under investigation. Nevertheless, image processing based on software presents reliable results and hence, can be transformed into hardware components, which may become an integrated part of the processing hardware.

Optical Transformations

Mathematical transformations are applied to signals to obtain further information from the signal that is not readily available in the raw signal. Most of the signals, in practice, are time-domain signals in their raw format. That is, whatever that signal is measuring, is a function of time. In many cases, the most important information is hidden in the frequency content of the signal. The frequency spectrum of a signal is basically the frequency components (spectral components) of that signal. The frequency spectrum of a signal shows what frequencies exist in the signal.

There are many transforms that are widely used, often by engineers and mathematicians, in solving problems in science and engineering. Fourier, Laplace, Hilbert, Wigner distributions, Radon, and wavelet transforms, constitute only a small portion of a huge list of transforms that are available. Every transformation technique has its own area of application, with advantages and disadvantages.

Fourier Transform

The Fourier transform (FT) is probably the most popular linear transform being used. The FT, a pervasive and versatile tool, is used in many fields of science as a mathematical or physical tool to alter a problem into one that can be more easily solved. It is used in linear systems analysis, antenna studies, optics, random process modeling, probability theory, quantum physics, and boundary-value problems and has been successfully applied to restoration of astronomical data. Some scientists understand Fourier theory as a physical phenomenon, not simply as a mathematical tool. In some branches of science, the FT of one function may yield another physical function. The FT, in essence, decomposes or separates a waveform or function into sinusoids of different frequency which sum to the original waveform. It identifies or distinguishes the different frequency sinusoids and their respective amplitudes. There are

functions for which the Fourier transform does not exist; however, most physical functions have a FT, especially if the transform represents a physical quantity. Other functions can be treated with Fourier theory as limiting cases. Many of the common theoretical functions are actually limiting cases in Fourier theory.

Wavelets

Wavelets were developed independently in the fields of mathematics, quantum physics, electrical engineering, and seismic geology. Interchanges between these fields have led to many new wavelet applications including image processing. Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. Wavelet transforms comprise an infinite set. The different wavelet families make different trade-offs between how compactly the basis functions are localized in space and how smooth they are. Within each family of wavelets are wavelet subclasses distinguished by the number of coefficients and by the level of iteration.

The wavelet transform (WT), which is also a linear operation, provides the time-frequency representation. Often a particular spectral component occurring at any instant can be of particular interest. In these cases it may be beneficial to know the time intervals at which these particular spectral components occur. WT is capable of providing the time and frequency information simultaneously, hence giving a time-frequency representation of the signal. The WT was developed as an alternative to the short-time FT (STFT). The WT was developed to overcome some resolution related problems of the STFT. The frequency and time information of a signal at some certain point in the time-frequency plane cannot be known. What spectral component exists at any given time instant is a mystery. Therefore, it is best to investigate what spectral components exist at any given interval of time. This is a problem of resolution, and it is the main reason why researchers have switched to WT from STFT. STFT gives a fixed resolution at all times, whereas WT gives a variable resolution as follows: higher frequencies are better resolved in time, and lower frequencies are better resolved in frequency. This means that a certain high frequency component can be located better in time (with less relative error) than a low frequency component. On the contrary, a low frequency component can be located better in frequency, compared to high frequency component.

Wavelet versus Fourier Transforms

Both transforms are linear and the mathematical properties of the matrices involved in the transforms are similar. The inverse transform matrix for both transforms is the transpose of the original. Wavelets have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. The most interesting dissimilarity is that individual wavelet functions are localized in space, Fourier sine and cosine functions are not. This localization feature, along with wavelets localization of frequency, makes many functions and operators using wavelets 'sparse' when transformed into the wavelet domain. This sparseness, in turn, results in a number of useful applications such as data compression, detecting features in images, and removing noise from time series. One way to see the time-frequency resolution differences between the FT and the WT is to look at the basis function coverage of the time-frequency plane. In order to isolate signal discontinuities, some very short basis functions are needed. At the same time, in order to obtain detailed frequency analysis, some very long basis functions are needed. A way to achieve this is to have short high-frequency basis functions and long low-frequency ones. This happy medium is obtained with WT. FT utilizes just the sine and cosine functions (single set of basis functions). Therefore, its new domain contains basis functions that are sines and cosines. On the other hand, WT includes an infinite set of possible basis functions. Hence, its new domain contains more complicated basis functions called wavelets, mother wavelets, or analyzing wavelets. The basis functions are localized in frequency, making mathematical tools such as power spectra and scalegrams useful at picking out frequencies and calculating power distributions. Thus wavelet analysis provides immediate access to information that can be obscured by other time-frequency methods such as Fourier analysis.

Physical Properties and Analog/Digital Mitigation Processing in Nonideal Systems

The weakest link in most imaging systems is time varying distortions, such as the Earth's atmosphere, motion, or vibrations. To obtain clear imaging through time varying distortions, it is useful to characterize the physical degradation phenomena using an OTF. It is often important to use the appropriate filter based on the OTF magnitude, modulation transfer function (MTF), to correct

(restore) the degraded image (using software or hardware). On the other hand, it may be useful to incorporate real-time phase correction into the imaging system such as adaptive optics.

Optical Transfer Function for Incoherent Imaging

In electronics, the transfer function is redefined as the Fourier transform of the impulse response. The OTF is defined similarly, but is normalized to its own maximum value which normally occurs at zero spatial frequency, that is:

$$\begin{aligned}\mathrm{OTF} &= \tau(\omega) = \tau(\omega_x, \omega_y) \\ &= \frac{\iint_{-\infty}^{\infty} s(x', y') \exp[-j(\omega_x x' + \omega_y y')] \mathrm{d}x' \, \mathrm{d}y'}{\iint_{-\infty}^{\infty} s(x', y') \mathrm{d}x' \, \mathrm{d}y'} \\ &= \frac{S(\omega_x, \omega_y)}{S(0,0)} = \mathrm{MTF} \exp(j\mathrm{PTF}) \qquad [1]\end{aligned}$$

where $S(\omega_x, \omega_y)$ is the FT of the spread function or impulse response $s(x', y')$. Ordinarily, for artificial imaging systems, the OTF is maximum at zero frequency. The magnitude of the OTF is called the MTF. The name refers to spatial modulation of light irradiance. If there is no such modulation, irradiance is uniform and there is no image. Thus, spatial modulation is related to image quality and MTF is a measure of the ability of an imaging component or system to transfer such spatial modulation from input (object plane) to output (image plane).

There is also a phase transfer function (PTF) whose importance in resolution is not nearly as widespread as that of MTF but which, nevertheless, cannot be neglected. Spatial phase determines image position and orientation, rather than amount of detail. If a target is displaced bodily in the image plane such that each part of the target image is displaced by the same amount, the target image is not distorted. However, if portions of the target image are displaced more or less than other portions, then the target image is distorted. This information is contained in the PTF.

The physical implications of the OTF are very analogous to those of electronics transfer functions. Both permit determination of output for any given input. In both cases, since the FT of the delta function is a constant (unity) amplitude, a delta function input permits indication of the frequency response of the system, that is, which frequencies pass through the system unattenuated and which are attenuated and by how much. Perfect fidelity or resemblance between input and output requires infinite bandwidth so that an impulse function is obtained at output for impulse function input. In imaging, an impulse formation at output is a point image. This means that, to obtain a point image for a point object, infinite spatial frequency bandwidth is required of the imaging system. Physically, this means that because of diffraction effects optical elements such as lenses, field stops, and mirrors must be of infinite diameter so as to eliminate diffraction at the edges. This, of course, is not attainable.

Pixel Modulation Transfer Function

To measure the point spread function (PSF), it is important to determine which size object can be considered a point object. This is affected primarily by instrumentation pixel size. The pixel is the smallest element recordable in image space. The brightness value represents average irradiance over that small portion of the image scene. Pixel size is often related to detector size. If only a portion of the detector is illuminated, the output current is equivalent to that obtained for the same total radiant power absorbed by the detector but averaged over the entire detector area. No detail smaller than a pixel can be resolved in the image. In practice, dead space between pixels violates isoplanaticism (spatial stationarity), which is a requirement for linear systems. Nevertheless, dead space is small and such effects are often neglected in image system characterization. However, it is possible to overcome distortions in MTF measurements arising from this lack of isoplanaticism by using a 'white' spatial noise random object. This gives rise to a white noise random image. The ratio of the image power spectral density to object power spectral density is equal to the square of the system MTF. In this way the lack of isoplanaticism is overcome by the spatial randomness of the object and image.

Pixel size strongly affects system MTF. If a pixel represents a point image, then pixel size and shape are minimum spread function. A best case OTF for such an imaging system is then a normalized FT of the pixel shape. For example, a square pixel yields a two-dimensional sinc MTF whose width increases as pixel size decreases.

Optical Transfer Functions for Image Motion and Vibration

Image blurring caused by vibrations is a factor whose influence on resolution is often significant in imaging systems that involve mechanical motion. A stabilized system, which obtains image motion corrections by moving optical elements so as to counteract sensor motions, has a residual error typically on the order of 2500 μrad/s involving pitch, roll, yaw, and

forward motion compensation errors. The greatest errors usually result from rotational vibration. A stabilized mount-type system has a combined error typically on the order of 100 μrad/s. Even where no motion is involved, as in large fixed-position astronomical telescopes, vibrations deriving from thermal gradients in the walls and other parts of the telescope can be the limiting factor in image resolution.

As a consequence of such image motion, in many high-resolution systems, despite the use of high-quality sensors, resolution is limited by image motion and, as a result, the high-resolution capability of the sensor may be wasted. It can be a significant waste of means to use an expensive high-resolution sensor in such a situation unless the blur can be corrected with image restoration. In imaging system design, a convenient engineering tool is OTF. The overall system OTF is limited by the OTF of the weakest link. The formulation of such image motion blur into an OTF format is thus very convenient for system design and system analysis purposes, as well as for image restoration.

OTFs can describe image quality for many forms of image motion such as linear (uniform) motion, sinusoidal vibrations at high vibration frequencies, sinusoidal vibrations at low vibration frequencies, acceleration, etc. The OTF for any type of image motion can be calculated numerically in real time provided the motion function is known. This is usually determined easily via a gyroscope or microelectronic accelerometer.

In general, the motion of interest is not the actual relative motion between object and imager, but rather the relative motion between imager and image plane. The latter motion is that observed in the image plane. Both of these motions are related to each other by the image system magnification. Hence, image plane relative motion is usually much less than the actual relative motion. However, any motion during an exposure gives rise to blurring of image plane detail. One question to be considered is how much blur can take place as a result of motion and still not be noticeable because of inherent blur deriving from the PSF of the rest of the imaging system. In other words, assuming that image motion and vibration blur cannot be removed completely, how much is tolerable before resolution and target acquisition capability are affected? This depends on focal length and system zoom.

A general method of OTF calculation is obtained by the line spread function (LSF) derived from image motion transverse to the optical axis. The MTF is derived as the modulus of the MTF or FT of the LSF, and the PTF is derived as the phase of the OTF. The LSF is the probability density function (PDF) of the histogram of the relative transverse displacement between object and sensor beginning at time t_s and ending at time $t_s + t_e$ where t_s is measured from the instant the sensor is first exposed and t_e is exposure time.

Recently a method of calculating motion OTF analytically has been developed for any type of motion. Also, methods to calculate motion OTF without a motion sensor have been derived for a sequence of consecutive images and even from a single motion-blurred image. All three methods have been used successfully in image restoration.

Optical Transfer Functions for the Atmosphere

Many properties of the atmosphere affect the quality of images propagating through it. There are atmospheric phenomena that give rise to attenuation of the irradiance of the propagating image, thus reducing the contrast of the final image. There are also atmospheric phenomena that cause blurring of detail. Both types of phenomena prevent small detail from being resolved in the final image, thus degrading image quality.

Electromagnetic (EM) wave absorption and scattering by the constituent gases and particulates of the atmosphere and airborne particulates give rise to attenuation. Scattering of photons by airborne particulates is manifested as deflections of the photons to directions other than that of the original propagation. If such scattering causes the deflected photons to miss the imaging receiver, then the scattering is manifested as attenuation. The received irradiance of the image propagated through the atmosphere is correspondingly diminished from that in the object plane. However, if the light scattering is at very small angles with respect to the original direction of the propagation, and several such small-angle scattering events take place, then forward-scattered radiation can take round-about paths and still be received by the imaging system together with the unscattered radiation. The net effect is image blurring caused by a multitude of angles of arrival at the imaging receiver of radiation emanating from the same point in the object plane. Many multiple-scattering paths give rise to a relatively large blurred-point image rather than a fine sharp-point image. Adjacent image plane points can then appear for a single object plane point, thus degrading image resolution. Such aerosol scatter blur is known as the adjacency effect.

Another effect of light scattering, particularly at large angles, is increased path radiance. The atmospheric background irradiance at wavelengths less

than 2–4 μm, consists through the day of scattered background light such as sunlight. This atmospheric background radiation is imaged over the same image space as the received irradiance from the target plane. The result is decreased contrast of the target plane scene. At wavelengths larger than 4 μm, most of the path radiance is not scattered sunlight, but rather thermal emission of the atmospheric constituents, gases in particular. This thermal path radiance decreases contrast in thermal imaging systems, particularly in the longwave infrared (8–13 μm).

Turbulence results from variations in the atmospheric refractive index, which are random in both time and space. These refractive index fluctuations are caused by local fluctuations in atmospheric temperatures, pressure, humidity, etc., all of which are affected by wind speed, which in itself is also random in time and space. Generally, the larger the temperature and humidity gradients, the more serious the image resolution degradation caused by turbulence. The lower the wind speed, the stronger the impact of turbulence. Scattering and turbulence are basically two very different physical mechanisms, each exhibiting quite different properties. Time dynamic properties such as scintillations and image dancing derive from turbulence, while aerosol scatter light changes only slowly with time.

A point that must be strongly emphasized is that there is no such thing as a purely turbulent, purely scattering, or purely absorption atmosphere. All three properties occur simultaneously in the real-world atmosphere. Modeling and interpretation of experimental results must relate to all three processes. Any attempt to pretend otherwise is not realistic. In general, blur is caused by turbulence and small-angle scatter of light by aerosols and particles. Attenuation is caused by large-angle scatter and by absorption.

Generally turbulence blur is dominant for images near the ground, where ground heating gives rise to the turbulence. Aerosol blur is usually dominant at higher elevations, even as little as several meters above ground, depending on weather conditions and wavelength.

The product of both aerosol and turbulence MTFs (in the spatial frequency domain, after FT) yields approximately atmospheric MTF. The atmospheric MTF can be predicted according to atmospheric (meteorological) parameters. Aerosol MTF calculation is based on optical depth, which includes parameters such as scattering and absorption coefficients through the atmosphere over the imaging path. Turbulence MTF calculation is based on the refractive index structure coefficient, which includes parameters such as temperature, humidity, wind speed, etc.

By referring to the edge response in a given scene the atmospheric MTF may be calculated. The edge separating the wide white and black bars is imaged from the scene. The image of the edge is not a step function because of atmospheric blur. The gradient of the edge image is the LSF, from which the overall atmospheric MTF can be obtained after the FT of the LSF is divided by the hardware MTF.

Spatial Filtering

Spatial filtering is a process by which we can alter properties of an optical image by selectively removing certain spatial frequencies that make up an object, for example, filtering video data received from satellite and space probes, or removal of raster from a television picture or scanned image. In digital image processing, the term image refers to a two-dimensional light-intensity function, where amplitude at spatial coordinates (x, y) gives the intensity (brightness) of the image at that point.

Digital image processing techniques are oriented to transform, enhance, restore, or encode an image. Digital image restoration is on the basis of the OTF or even MTF alone describing the image degradation. Resolution is limited usually not by electronics or optics, but by image motion and vibration. For long-range imaging, the image quality may be limited by atmospheric effects.

By image restoration filters we refer primarily to spatial frequency filters placed in the spatial frequency plane in the optical data processor. Although data processing is done optically, the same mathematical concept is usually carried out with a digital computer. It is also possible to do restoration in the spatial domain. The simplest image restoration filter is the inverse filter. When noise becomes a problem, there are often used three classes of minimum mean square error estimations (MMSEE), which are the noncausal Wiener filtering, causal Wiener filtering, and Kalman filtering. Kalman and Wiener filters are proposed to obtain the optimum solution for the MMSEE problem in linear systems. Wiener filters are applied in linear time invariant systems, and Kalman filters in linear systems which may be either time invariant or time variant. A novel area of application for Wiener filtering is image restoration. On the other hand, if the estimation processes are partial and an ideal image is not obtainable, a Kalman filter may present better restoration results than a Wiener filter.

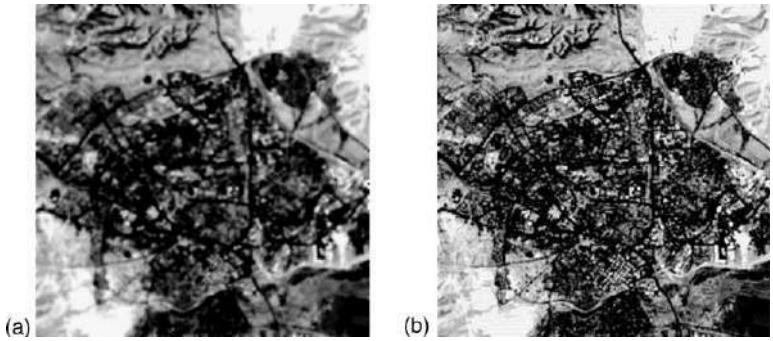

Figure 3 An example for satellite image restoration. (a) Original image of Beer-Sheva, Israel, from Landsat 7 satellite; (b) restored image after Kalman filter using aerosol and turbulence MTFs. Reproduced with permission of the American Society for Photogrammetry and Remote Sensing from Arbel D, Cohen E, Citroen M, Blumberg DG and Kopeika NS (2004) Landsat TM satellite image restoration using Kalman filters. *Photogrammetric Engineering and Remote Sensing* (PE&RS) 70(1): 91–100.

Figure 3 demonstrates the restoration result for a satellite image. Figure 3a presents an original Landsat 7 enhanced thematic mapper plus (ETM+) image of the city of Beer-Sheva, Israel. The image was recorded from 750 km satellite elevation. The image was recorded using an ETM+ sensor with 15 m ground resolution, and its wavelength range is 520–900 nm (Panchromatic band). Figure 3a is of 256×256 pixel size, which is a segment defined as a region of interest (ROI) in the original image. The scene was recorded on August 7, 1999 and was degraded (blurred) as a result of the meteorological conditions, which were measured and recorded (as data) at the same time. Figure 3b shows the restored image using the Kalman filter, which was based on the atmospheric MTF, which is obtained by using aerosol MTF with turbulence MTF. The restored image (Figure 3b) is more detailed. For example, roads and buildings were deblurred and became more accurate than in the original image (Figure 3a).

See also

Environmental Measurements: Optical Transmission and Scatter of the Atmosphere. **Fourier Optics**. **Holography, Techniques:** Computer-Generated Holograms. **Quantum Optics:** Quantum Computing with Atoms.

Further Reading

Arbel D, Greenberg S, Hadar O and Kopeika NS (2003) Criteria for satellite image restoration success. *Optical Engineering* 42(9): 2607–2613.

Arbel D, Cohen E, Citroen M, Blumberg DG and Kopeika NS (2004) Landsat TM satellite image restoration using Kalman filters. *Photogrammetric Engineering and Remote Sensing (PE&RS)* 70(1): 91–100.

Daniels A, Boreman GD and Ducharme AD (1995) Random transparency targets for modulation transfer function measurement in the visible and infrared. *Optical Engineering* 34(3): 860–868.

Gabor D (1948) A new microscope principle. *Nature* 161: 777.

Gabor D (1949) Microscopy by reconstructed wavefronts. *Proceeding of the Royal Society A* 197: 454.

Gabor D (1951) Microscopy by reconstructed wavefronts: II. *Proceeding of the Royal Society B* 64: 449.

Hadar O, Dror I and Kopeika NS (1994) Image restoration limits resulting from mechanical vibrations, part IV: real-time numerical calculation of optical transfer functions and experimental verification. *Optical Engineering* 33: 566–578.

Hadar O, Robbins M, Novogrozky Y and Kaplan D (1996) Image motion restoration from a sequence of images. *Optical Engineering* 35: 2898–2904.

Kopeika NS (1998) *A System Engineering Approach to Imaging*. Bellingham, WA: SPIE Press.

O'Donnell E (1987) Image steadiness: a critical factor in system performance. In: Wright R (ed.) *Electron Optical Imaging System Integration. Proceedings SPIE* 762: 51–56.

Stern A and Kopeika NS (1995) Analytical method to calculate optical transfer functions for image motion and vibrations using moments. *Journal of the Optical Society of America A* 14: 388–396.

Yitzhaky Y and Kopeika NS (1997) Identification of blur parameters from motion blurred images. *CVGIP: Graphical Models and Image Processing* 59(5): 321–332.

OPTICAL TWEEZERS

A Gajraj and J C Meiners, University of Michigan, Ann Arbor, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The concept of radiation pressure or pressure from the propagation of light is not new. In the early seventeenth century, Johannes Kepler proposed that radiation pressure from the sun accounted for the fact that tails of comets always point away from the sun. In 1873, James Maxwell, on the basis of electromagnetic theory, showed that pressure should result from the absorption or reflection of a beam of light. However, early studies in the beginning of the twentieth century that attempted to detect and quantify the effects of radiation pressure were unsuccessful, mainly frustrated by thermal forces from temperature gradients that tended to obscure the weak effects of radiation pressure. It was only in the 1960s, that the advent of lasers provided a light source with the power to generate sufficient radiation pressure to overcome forces due to thermal effects.

As lasers became generally available in the early 1970s, Arthur Ashkin from the Bell Labs pioneered the development of optical tweezers. He showed that the forces of radiation pressure from focused laser beams could be used to affect the dynamics of small transparent micrometer-sized particles. According to Ashkin, his early experiments were motivated by simple back of the envelope calculations of the magnitude of the radiation pressure force of light on a totally reflecting mirror. Let us consider such a calculation: a photon of frequency ω has a momentum of $\hbar\omega/c$. For a total power of I, we have a flux of $I/\hbar\omega$ photons per second incident on the mirror. When the photons are reflected, they transfer a momentum of $2I/c$ to the mirror. The force exerted on the mirror for $I = 1$ W, is approximately 10^{-9} N. If the mirror is a 1 μm size particle weighing 10^{-15} kg, this force will induce an acceleration of 10^6 g!

In the following decade, Ashkin demonstrated that by focusing laser light using a high numerical aperture microscope objective, strong intensity gradients could be created to optically trap small dielectric particles such as microspheres. Ashkin named this trapping scheme a 'single-beam gradient force trap', now more commonly called optical tweezers.

Today optical tweezers are well-established and extremely versatile tools for the physical and life sciences. Using optical tweezers, forces up to 200 pN can be applied with a resolution of 100 aN on particles with dimensions on the order of the wavelength of light. Even particles as small as 5 nm can be trapped, rotated, positioned, and arrayed. Optical tweezers have been used in single molecule studies of biopolymers (DNA, RNA), in the manipulation of various types of cells and in cell sorting, in microsurgical procedures, as well as in viscoelastic studies of membranes, and protein. Optical tweezing has also been the key tool in the understanding of the workings of molecular motors such as myosin, kinesin, RNA polymerases, etc.

Theory

Optical tweezers trap dielectric particles using radiation pressure generated by a tightly focused beam of light from a laser. Depending on the size of the particle in relation to the wavelength of the trapping beam, there are two different mechanisms responsible for trapping small particles in a single beam gradient trap. For particles in the Mie regime, where the particle size is large compared to the laser wavelength, a simple geometric-optics model serves best. Figures 1a–c, illustrate possible cases of a Mie particle in an optical trap. Here, the particle is assumed to be a sphere that is transparent with a refractive index higher than that of the surrounding medium. The particle behaves as a convergent lens refracting and deflecting the beam when its center is displaced from the laser focus. In Figure 1a, the sphere is axially displaced from the focus, away from the microscope objective. Ray A is then refracted toward the surface normal as it enters and exits the sphere, causing a change in the momentum of the ray. Due to conservation of momentum, the sphere reacts with a force F_a in the opposite direction. The situation is similar for ray B. The gradient force, F, obtained from the vectorial addition of F_a and F_b, is directed toward the light source. At the same time, the scattering force, arising from back and scattered reflections, is directed away from the light source. It can be seen that above the focal point, the scattering and gradient forces act in the same direction, pushing the particle away from the light source (Figure 1b). For a particle in the focal plane, but transversely displaced from the trap center, the gradient force acts in a restoring direction pushing the bead into the trap focus (Figure 1c). At a location just below the focus, the scattering force balances the gradient force and stable optical trapping is achieved.

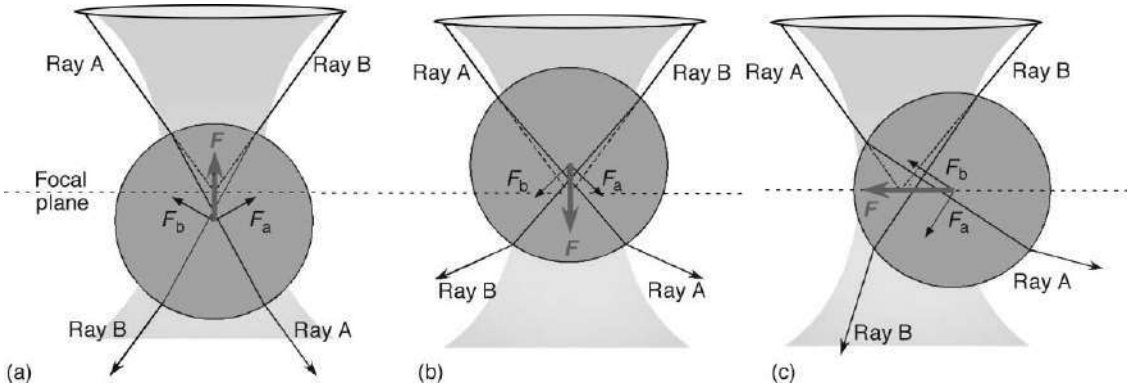

Figure 1 Ray diagram sketches showing three cases of particle trapping in the Mie regime. Rays A and B border the cone of light focused by a high numerical aperture microscope objective. The refraction of these rays gives rise to forces F_a and F_b, the resultant force F tends to restore axial and transverse displacements of the particle from the trap focus (at the intersection of dotted rays A and B).

This simple ray optics model can also be used to estimate the trapping force, an example of such a calculation is shown in Box 1.

For dielectric particles in the Rayleigh regime, where the size of the particle is small compared to the wavelength of the incident light, the particle acts as a simple dipole. The electromagnetic light wave polarizes the particle, inducing a set of oscillating dipoles. The dipoles respond to the electric field of the incident light in the direction of the gradient, resulting in the movement of the particle towards the focal point. Additionally, there is also a scattering force, acting in the direction of beam propagation, from light that is scattered or absorbed. While the scattering force is proportional to the electromagnetic field, the gradient force is proportional to the gradient of the electromagnetic field. Therefore, stable optical trapping can be achieved for a field with a gradient large enough to overcome the scattering force. This can be obtained by using a microscope objective with a high numerical aperture.

Stable trapping is achieved when the energy associated with the trapping significantly exceeds the thermal energy of the trapped particle, $\frac{1}{2}k_BT$ per degree of freedom. Since the optical gradient is weakest in the axial direction of the trap, the trap stiffness in this direction limits its stability. Furthermore, the trap is weakened in the axial direction by the scattering force, which pushes the microsphere away from the objective. To overcome this, two counterpropagating laser beams can be used, not only to increase the axial gradient of the trap, but also to provide a compensating scattering force in the opposite direction.

Force Measurement

Optical tweezers can be used as picotensiometers or highly sensitive force sensors. Measuring the displacement of a particle in a trap with known stiffness k, yields the force F given by

$$\mathbf{F} = -k \cdot x \qquad [1]$$

The displacement can be measured in different ways, ranging from video microscopy to light-scattering techniques. Some of these methods are discussed in the instrument design section below.

The trap stiffness is generally not known *a priori* as it depends on the shape and intensity of the trap, as well as the relative refractive index, shape, and position of the object in the trap. Since theoretical approximations are usually not accurate enough, the trap stiffness must be empirically determined and calibrated.

Various force calibration strategies have evolved over the years, however, they can be broadly divided into two types: (i) schemes that rely on the interaction of the optical trap with viscous forces; and (ii) those that exploit the thermal fluctuations of the trapped object.

Force Calibration by Viscous Drag

The conceptual basis for calibration by viscous drag is simple. A known force is exerted on a trapped object and the deviation in its position is measured as a function of the applied force. The trap stiffness can then be calculated from eqn [1]. Force can be exerted by either flowing liquid past the trapped object, translating the microscope stage while keeping the trap fixed, or translating the optical trap while keeping the sample fixed.

In a situation where the Reynolds number is small, inertial forces can be neglected and Stokes' Law gives the viscous drag force on a sphere of known radius r:

$$F_d = (6\pi\eta r)v = \beta v \qquad [2]$$

Box 1

Optical Tweezers in the Ray Optics Model

In a simple ray optics model, the trapped microsphere acts like a small lens on the incident laser beam. If the microsphere is not centered on the beam axis, the beam will be refracted. The corresponding change in the momentum of the laser beam gives rise to a restoring force on the microsphere.

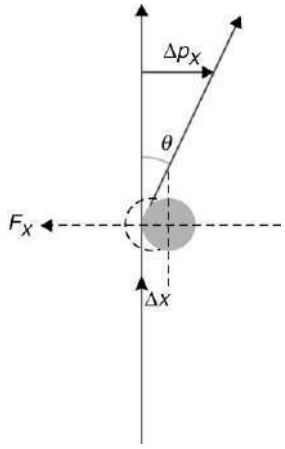

Figure B1

A photon of frequency ω in a medium of refractive index n_1 that is deflected by an angle θ, as shown in Figure B1, acquires a lateral momentum of

$$\Delta p_x = \frac{\hbar \omega n_1}{c} \sin \theta \qquad [B1]$$

Momentum conservation requires a counteracting force on the microsphere, which is for a laser beam of intensity I

$$F_x = \frac{I n_1}{c} \sin \theta \qquad [B2]$$

Assuming that the deflection angle θ is given by the deflection of a beam that is incident on biconvex lens with curvature R and refractive index n_2 at an off-axis displacement Δx, we find for the restoring force on the microsphere:

$$F_x = \frac{2I}{Rc}(n_2 - n_1)\Delta x \qquad [B3]$$

which describes a linear spring with a stiffness of the optical trap of

$$k = \frac{2I}{Rc}(n_2 - n_1) \qquad [B4]$$

Here, η, the viscosity of the sample fluid, is known and immediately allows the determination of the friction coefficient β of the sphere. The fluid velocity, v, can be measured in the case of external fluid flow by either tracking the trapped object when the trap is switched off or measuring the speed of some other marker in the sample. If the stage is moved to generate the fluid flow, the velocity of the fluid surrounding the trapped object can be estimated from the motion of the stage. In either case, the drag coefficient, from eqn [2] may need to be corrected if the trapped object is close to the walls of the sample cell. This correction is well known as Faxen's Law, which accounts for hydrodynamic interactions between the trapped particle and the wall. These effects are particularly significant when the center of the trapped particle is less than a diameter away from the wall. The advantage of these fluid-flow

calibration methods is that the trap stiffness at regions far from the trap center can be measured. The drawback is that, in general, only transverse trapping forces can be measured and the trapped objects have to be approximated as spherical or ellipsoidal.

Thermal Motion Methods

These methods measure the Brownian motion of a trapped particle to deduce the trap stiffness. A trapped particle experiences three forces: (i) the thermal forces due to collisions with surrounding solvent molecules; (ii) forces from viscous drag; and (iii) the restoring force of the trap. For low Reynolds numbers, where the inertial effects can be neglected for the microsphere, the equation of motion is

$$\beta\dot{x} + kx = F(t) \qquad [3]$$

Where β is the viscous drag, k is the trap stiffness, and $F(t)$ is a randomly fluctuating Langevin force due to thermal fluctuations. This motion of the microsphere in the trap at a temperature T is characterized by a Lorentzian power spectrum of the form:

$$x_f^2 = \frac{k_b T}{2\pi^3\beta(f_0^2 + f^2)} \qquad [4]$$

where $f_0 = k/(2\pi\beta)$ is the corner frequency, above which the motion of the particle is Brownian.

To calibrate the optical trap by the corner frequency method, the motion of a trapped particle is observed and its power spectrum, which is the Fourier transform of the autocorrelation function, is calculated. Fitting the power spectrum to a Lorentzian function will yield the corner frequency f_0, which can be directly related to the trap stiffness k, provided that the drag coefficient β can be estimated.

Alternatively, the trap stiffness can be obtained directly from the time constant of the exponential decay of the autocorrelation function of the particle motion. The spring constant is the ratio of the drag coefficient to the time constant of the trap.

Another approach known as the mean square displacement method does not require the estimation of the drag coefficient but does require a well-calibrated position detector. In this method, the mean square displacement $\langle x^2 \rangle$ of the fluctuations of a trapped particle is computed from position data and then by the Equipartition theorem:

$$\langle x^2 \rangle = k_b T/k \qquad [5]$$

This enables a simple, straightforward calculation of k, the trap stiffness.

Design

Optical tweezing stations broadly fall into three categories: (i) stand-alone systems; (ii) systems incorporating an optical microscope; and (iii) completely configured commercially available systems. Stand-alone systems feature optical tweezers built around a microscope objective and suitable optical components. These systems provide the maximum flexibility and can be easily reconfigured to adapt to different experimental needs. Furthermore, they also tend to provide the best mechanical stability, if configured properly. Figure 2a shows a picture of a stand-alone optical tweezing setup from our laboratory. This system is capable of the basic tasks of trapping, holding, and positioning microscopic objects. The most common optical tweezing stations are integrated into commercial light microscopes as sketched in Figure 3. The advantage of these hybrid systems is that the commercial microscope provides a flexible, optimized optical imaging platform, in which the optical trapping system can be integrated with only moderate difficulty. Figure 2b shows a picture of a typical optical tweezing station built around a commercial inverted microscope and mounted on a vibration–isolation table. This system includes an acousto-optic modulator that enables beam scanning for line-tweezing experiments. At the time of this writing, optical tweezing systems range from less than \$10 000 for basic demonstration stations to between \$50 000 to \$100 000 for more sophisticated hybrid systems.

Commercial fully integrated systems are available from P.A.L.M. Microlaser Technologies AG (Bernried, Germany), Cell Robotics, Inc (Albuquerque, NM, USA), and Arryx (Chicago, IL, USA). Figure 2c shows a commercial optical tweezers set-up available from P.A.L.M. Microlaser Technologies.

When designing an optical tweezing station, the first item to consider is the choice of laser system. A simple single beam tweezing system can be built around a laser with less than 100 mW of power. Even though the minimum power required for trapping is a few milliwatts in the laser focus, there is considerable power loss in the tweezing and microscope optics. Also, the trap stiffness and thus the maximum trapping force decreases linearly with intensity, therefore most optical tweezers typically use lasers with an output power of 1 W. In some special applications, when beam scanning or multiple traps are required, the power requirement can be significantly larger. The wavelength of the laser also affects the capability of the trap in that the optimum wavelength depends on the size of the objects that are to be trapped. In general, the strongest traps occur

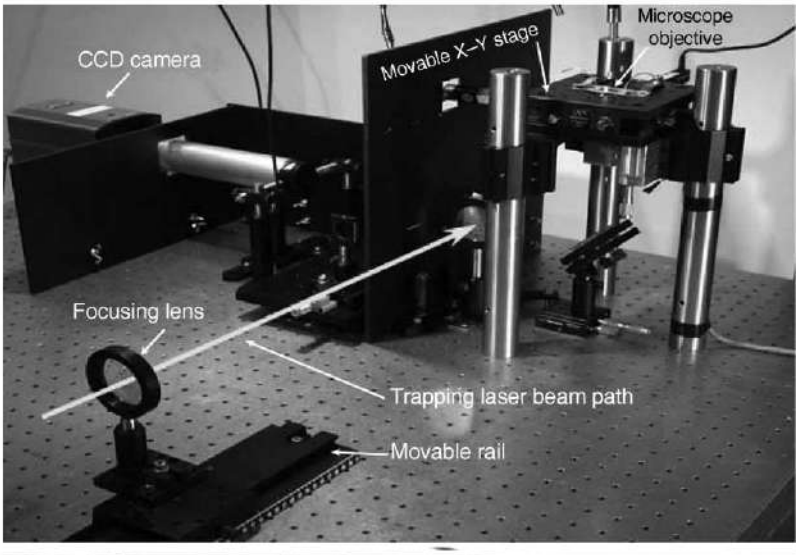
CCD camera
Movable X–Y stage
Microscope objective
Focusing lens
Trapping laser beam path
Movable rail

(a)

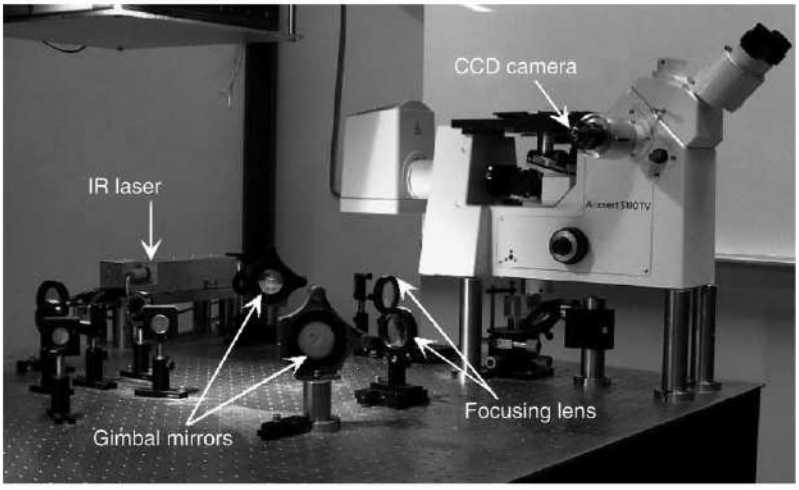
CCD camera
IR laser
Focusing lens
Gimbal mirrors

(b)

(c)

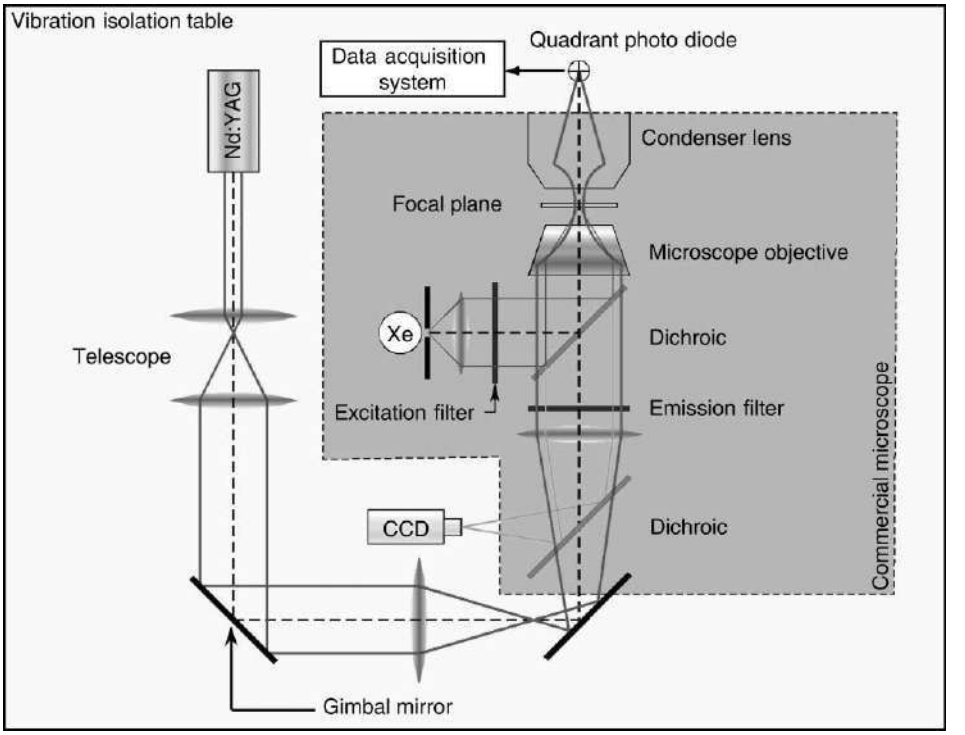

Figure 3 A sketch showing the instrumentation of a typical optical tweezing system incorporated into a commercial microscope.

for particles of size comparable to the wavelength of the laser. Another important concern in the choice of wavelength is possible damage to the sample from absorption and heating. In general, biological material is more transparent to the longer wavelengths of the IR spectrum. Studies on living cells have shown that generally, a range of wavelengths of 700 nm to 1100 nm, termed 'the optical window' of cells and tissue, is most suitable. Importantly, at higher wavelengths there is a reduced possibility of damage from multiphoton effects. DNA (260 nm) and proteins (280 nm) are strong absorbers in the ultraviolet range. Due to multiphoton effects, wavelengths between 700–800 nm have been seen to affect cell metabolism and lead to cell death within seconds. However, at wavelengths greater than 1000 nm, the simultaneous absorption of four photons required is unlikely at typical laser intensities used in trapping experiments.

The beam quality of the laser also restricts the choice of laser system. Diode lasers, in particular, have poor beam quality since the beam is often elliptical in shape and astigmatic, leading to a larger inefficient trap. Over the years, Nd-YAG lasers, at 1064 nm, have emerged as the lasers of choice, providing the best balance between price, power, and quality.

To achieve stable trapping it is essential to create an optical potential in which the gradient forces dominate the scattering force as discussed in the Theory section above. This suggests that the higher the numerical aperture (NA) of the microscope objective the stronger and more stable the trap will be. Oil or water-immersion microscope objectives with $NA > 1.2$ are commonly used. Further, this requirement also implies that the back aperture (rear entrance) of the microscope objective needs to be completely filled or, in the case of a Gaussian beam, overfilled to a certain degree to achieve optimal trapping efficiency. Since most commercial objectives are corrected for achromatic and spherical aberrations in the visual region of the spectrum, only

Figure 2 (a) A stand-alone tweezing station built around a microscope objective mounted on a vibration isolation table. (b) A typical tweezing station incorporated into a standard commercial microscope. This setup features dual-steerable optical traps and includes a CCD camera. Position detecting instrumentation can be mounted on the existing microscope stage. (c) The commercially available PALM® Combisystem utilizes an IR laser (wavelength 1064 nm) and a UV laser (wavelength 377 nm) coupled via the epi-fluorescence path to a research microscope. The system advertises automated and noncontact laser microdissection as well as trapping, sorting, and positioning. (**Figure 2:** (c), courtesy of P.A.L.M. Microlaser Technologies, Bernried, Germany.)

a small fraction of the working distance is usable for viable trapping. Normally this is limited to within 40 μm of the cover slip.

In addition to providing a beam large enough to overfill the back aperture, there are two other criteria that the external optics have to satisfy: first, the optical trap should be parfocal with the specimen plane. This enables the visualization of trapped objects. Second, the laser beam should be able to pivot around the back aperture without moving laterally to maintain the same degree of overfilling. This allows for a stable two-dimensionally movable trap. In practice, these requirements are fulfilled by using a telescopic system of lenses and Gimbal mounted mirrors. The latter are special mounts that enable a mirror to be tilted around its center position without lateral displacement. The telescope expands the incoming beam to fill the back aperture of the microscope objective and also images the back aperture onto the center of the Gimbal mounted mirror. As seen in Figure 4, since the back aperture of the microscope objective and the Gimbal mounted mirror are in conjugate planes, a small tilting of the mirror will only result in a change in the angle of the laser beam entering the microscope objective. The position of the beam will be unchanged, as will the degree of overfilling.

More sophisticated optical tweezer designs use acousto-optic modulators and galvanometer scanning mirrors to create scanning line tweezers and/or multiple traps. The basic approach is to deflect the laser beam to create multiple traps or to scan a single trap to create a one-dimensional potential well. Additionally, a new generation of holographic optical tweezers, capable of creating three-dimensional arrays of trapped particles, has emerged. Computer generated holograms that modify the phase and amplitude of the trapping laser are reverse engineered from the desired two- or three-dimensional optical trapping patterns. Furthermore, constructing such holograms from liquid crystal light modulators allows for dynamic control of the trapping pattern, yielding a fully computerized reconfigurable trapping system. An interesting example of the application of holographic optical tweezers to colloidal transport is shown in Figure 5, the first demonstration of 'kinetically locked-in states' where particles traveling across a potential mesh display a preference for certain paths.

It is to be noted that the optical tweezer is fundamentally a constant-extension technique, that is, it attempts to maintain the extension of a stretched molecule whereas the force is varied. However, it is of interest to study the action of molecular motors under constant force conditions. A basic optical tweezers setup can be modified to include an electronic feedback loop to provide constant force conditions. More recently, scanning-line optical tweezers have been adapted to implement all-optical constant force schemes that dispense with the need for electronic feedback.

An important aspect of any force measurement is the ability to accurately image and measure the position of a trapped particle. Standard digital video microscopy techniques yield a resolution of ~5 nm but have limited bandwidth and thus cannot be used for systems that employ feedback control of laser intensity or trap position. In these systems, the forward- or back-scattered trapping laser light from the trapped particle can be imaged onto a calibrated quadrant photodiode to provide sensitive position detection at high bandwidths (~1 kHz). If the photodiode is fixed in a plane conjugate to the back focal plane of the microscope objective, the trap can

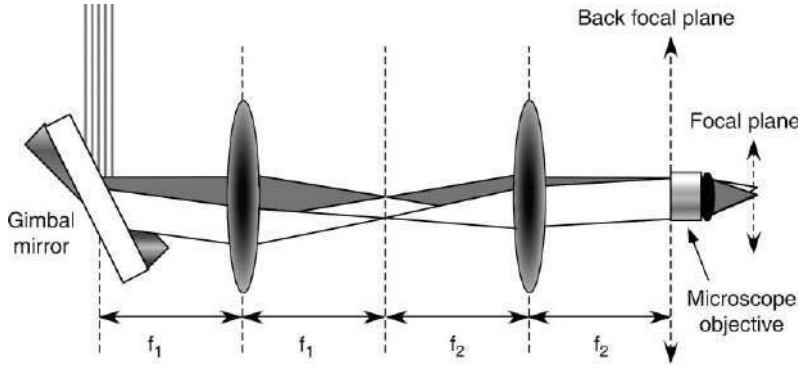

Figure 4 A sketch showing the working of a gimbal mounted mirror in two positions (shown as shaded and as unshaded). For small tilts of the mirror, the beam is seen to 'pivot' about the back aperture of the microscope objective resulting in the displacement of the optical trap in the focal plane. Note that for both beams the degree of overfilling of the back aperture of the microscope objective is unchanged, a crucial requirement for movable traps.

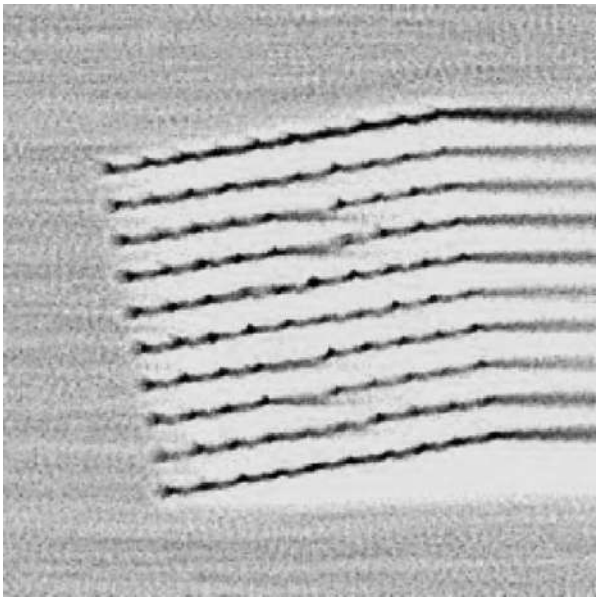

Figure 5 Data showing the superimposed paths of approximately 18 000 micron-sized particles flowing across a 10×10 grid of optical traps created by holographic optical tweezers. The particles are deflected laterally assuming a path dictated by the potential energy landscape and display kinetically locked-in states. This technique can be used to sort particles based on size, refractive index or charge. (Picture courtesy of D. Grier.)

be moved without having to move or recalibrate the photodiode detector. Using the quadrant photodiode method 2- and 3-dimensional position sensing is possible. A further improvement is found in schemes that use an additional laser dedicated to position sensing. The most sensitive method, however, for position sensing is interferometry. In this scheme, the trapping laser beam is split into two orthogonally polarized beams that are recombined after passing through the microscope objective and the sample. For a particle perfectly centered in the trap, the resulting combined beam has a circular polarization. In other cases, the bead displacement is sensitively reflected in the ellipticity of the polarization.

Applications

Colloids

Optical tweezers have been widely used to probe the interaction in colloids and other complex fluids. In fact, an optical tweezer apparatus was included in the instrumentation on the International Space Station for use with colloidal crystallization experiments in space. Various aspects of colloidal interactions have been studied using optical tweezers including mutual- and self-diffusion coefficients, the effects of hydrodynamic and electrostatic interactions, and depletion layer effects. More sophisticated scanning-line tweezing setups have been used to probe entropically driven self-assembly and interactions of colloids in polymer suspensions.

Blinking optical tweezers provide an elegant method for positioning colloidal particles in configurations of interest in a solution. When the tweezers are switched off the colloids respond to their environment, reacting to thermal forces and interacting either with other colloids or interfaces. The trajectories of the released colloids can be tracked and analyzed in 2D using digital video microscopy or 3D if combined with an evanescent illumination scheme. Switching on the tweezers again results in a return to the initial configuration, yielding a reproducible and tightly controlled experiment to examine colloidal behavior.

Biopolymers

The study of biopolymers has benefited from the development of optical tweezers. By chemically attaching a trappable latex microsphere to a biopolymer of interest, the molecule can be stretched, positioned, or moved. Measuring the response of the biopolymer to these external forces yields useful information such as elasticity, rigidity, and domain unfolding characteristics. This information obtained at a single molecule level has provided invaluable insight into the relationships between the structure and function of biopolymers. For example, in an early tweezing experiment, characterizing the intrinsic elasticity of λ-phage DNA, it was discovered that the double-stranded DNA undergoes a reversible structural transition when stretched at forces greater than 65 pN, extending to about 1.7 times its crystallographic contour length. These studies complement well-established structural techniques, such as X-ray crystallography and NMR, which provide information in a force-free environment. Furthermore, by anchoring one strand of DNA to a surface and pulling on the other strand, unzipping forces for DNA have been measured. Optical tweezers have also been combined with other mechanical tools, such as a micropipette, and used as sensitive tensiometers. Protein folding and unfolding experiments are generally harder, as their contour length is shorter than that of typical DNA fragments. Nevertheless, the folding and unfolding behavior of single giant muscle protein, titin has been probed by attaching a different latex bead to either end of the molecule. The first bead is held by a movable micropipette and the second bead is optically trapped. As the micropipette is moved, the position of the optically trapped bead provides information about the forces required to unfold the membrane domains of the protein.

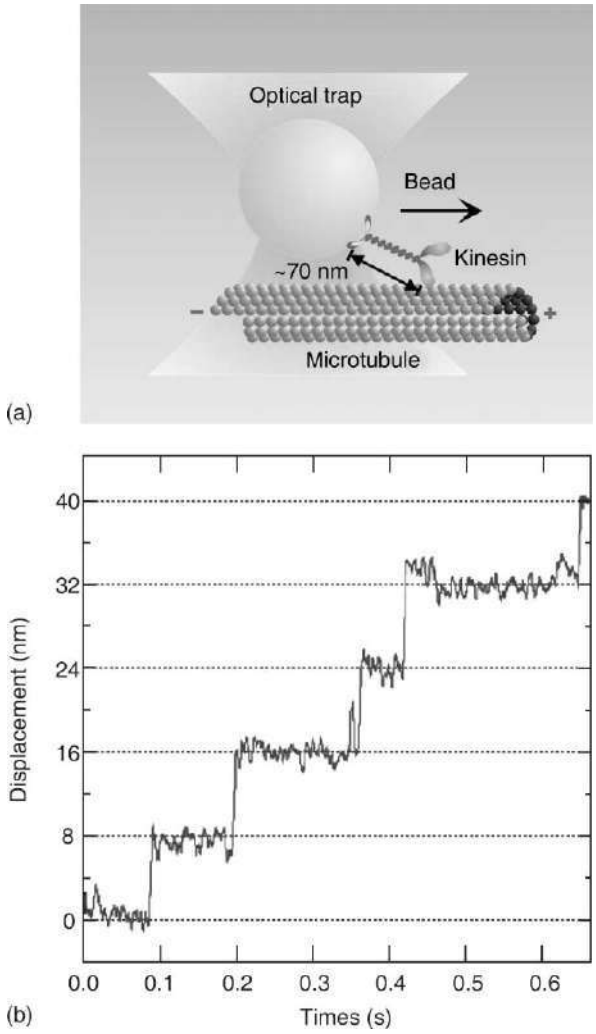

Figure 6 Observation of the stepping motion of the molecular motor kinesin. (a) A microsphere attached to a kinesin molecule is optically trapped; as the kinesin steps along a microtubule track which is immobilized onto a glass substrate, it pulls on the microsphere, whose motion is followed by an electronic feedback mechanism. (b) Motion of a single kinesin motor stepping against a 5 pN force. The data shows the discrete 8 nm steps of kinesin. (Figures courtesy of S. Block.)

Molecular Motors

Molecular motors (proteins that transduce chemical potential energy in to mechanical motion) have been the subject of intense attention since the early 1990s. Optical tweezers have played a crucial role in furthering the understanding of the mechanisms of molecular motors. The characteristics of molecular motor systems that have been studied with optical tweezers include speed, step-size, processivity, ATPs consumed per stroke, stall force, and static force. In fact, many of the advances in optical tweezers technology have been motivated by the study of molecular motors and have lead to spectacular results. An impressive example is the observance of the stepping motion of kinesin, as shown in Figure 6. Other experiments on molecular motors with optical tweezers have measured the stall force of motors as they move along their polymer track, or observed the transcription of DNA and RNA polymers, yielding interesting observations of the pauses during the transcription process.

Cells

The application of optical tweezers to cell biology was first demonstrated by Ashkin and co-workers in the late 1980s, by successfully trapping living objects: large viruses and bacteria. Since then, optical tweezers have been widely used to study various aspects of cellular biology. These include cell reproduction, growth, adhesion, mobility, and membrane elasticity. Additionally, the noninvasive nature of optical tweezers enables manipulations in the interior of an unopened object, allowing for the study of intracellular mechanisms by subcellular manipulations.

Optical tweezers have also been combined with various techniques to yield application-oriented tools for cell biologists. For instance, the combination of optical tweezers and state-of-the-art image recognition algorithms has been used to develop microfluidic cell sorters. In reproductive medicine, two of the major problems are penetration of the egg and sperm motility. A promising new solution combines laser zona drilling to pierce the egg envelope and optical tweezing to transport the sperm cells to the egg.

See also

Modulators: Acousto-Optics.

Further Reading

Ashkin A (2000) History of optical trapping and manipulation of small-neutral particle, atoms and molecules. *IEEE Journal of Selected Topics in Quantum Electronics* 6(6): 841–856.

Greulich KO (1999) *Micromanipulation by Light in Biology and Medicine: The Laser Microbeam and Optical Tweezers*. Basel; Boston; Berlin: Birkhauser.

Grier DG (2003) A revolution in optical manipulation. *Nature* 424: 810–816.

Lang MJ and Block SM (2003) Resource letter: LBOT-1: Laser-based optical tweezers. *American Journal of Physics* 71(3).

Sheetz MP (1998) *Methods in Cell Biology. Volume 55: Laser Tweezers in Cell Biology*. San Diego, CA: Academic Press.

Svoboda K and Block SM (1994) Biological applications of optical forces. *Annual Review of Biophysics Biomolecular Structure* 23: 247–285.

P

PHASE CONTROL

Contents

Phase Conjugation and Image Correction
Wavefront Coding

Phase Conjugation and Image Correction

E N Leith, University of Michigan, Ann Arbor, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Phase conjugation of a wave is a process that has long been recognized. However, the technique for producing high-quality phase conjugation is relatively new. Also, the principal use of phase conjugation, the correction of phase errors in waves that have propagated through phase-distorting media, is relatively new, having originated in holography in the 1960s. Today, phase conjugation is an important area of modern optics. This article describes the conjugation of a wavefield by the process of holography and the holographic phase conjugation method of removing phase distortions from a wavefield, thereby permitting high-quality images to be produced from phase-distorted waves.

The basic meaning of a conjugated wave is explained with the aid of Figure 1. Figure 1a shows a wavefield propagating to the right. For simplicity, we show only one wavefront, i.e., one contour of constant phase. It is understood that a wavefield consists of a sequence of such wavefronts, with each wavefront differing in phase by 2π from the adjacent one, and separated by the wavelength λ. Of the two points, A and B, on the wavefront, we say that the wavefront at A lags that at B, since when the wavefront crosses a plane normal to the direction of travel, point A of the wavefront crosses that plane behind point B. The wavefield produces, at the plane P, a field $ae^{i\phi}$, where a is the amplitude and ϕ is the phase of the wavefront at the plane P.

We place a phase conjugator in the path of the wave. The wave passes through the conjugator, emerging as a phase conjugated wave, in which the point B, which formerly led point A in phase, now lags behind it by a corresponding amount. This illustrates the basic meaning of the term 'phase conjugation'. This device could be considered as a forward phase conjugator, since the phase conjugated wave travels in the same direction as the original wave.

Next (Figure 1b) we describe a backward phase conjugator. The conjugate wave travels in a direction opposite to that of the incident wave. The conjugate wave emitted by the phase conjugator seems to be

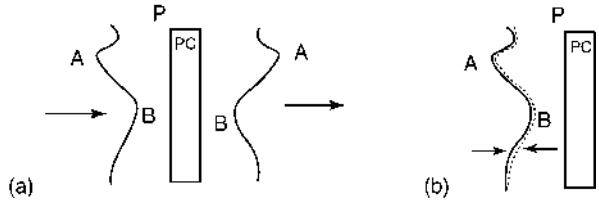

Figure 1 A phase conjugator. (a) shows a forward phase conjugator; (b) shows a backward phase conjugator. For clarity, in (b) the conjugated wave is shown with dotted lines. PC is phase conjugator, and P is the entrance plane of the conjugator.

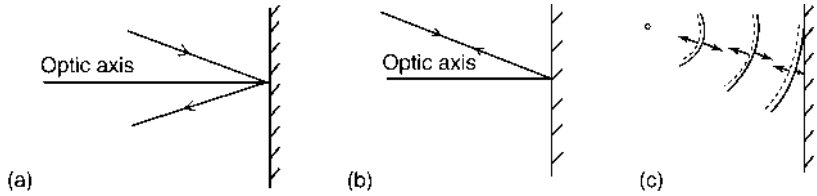

Figure 2 Light reflected from mirrors. (a) Light falling on a conventional mirror and being reflected; (b) light reflected from a conjugating mirror; (c) as in (b), except that illustration is with wavefronts instead of rays.

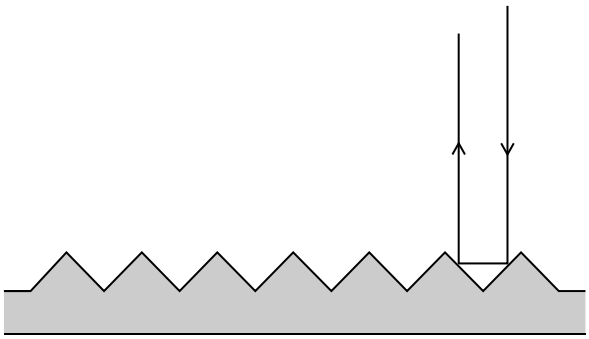

Figure 3 A ray of light reflected from a retroreflector (or corner cube).

identical to the incident wave; we show the two waves to be in coincidence, or more precisely, for purposes of clarity, almost in coincidence. The points A on the two wavefronts coincide, as do the points B. As before, on the incident wave, point A lags behind point B. On the exiting wave, since the wave is traveling in the opposite direction, point B is now lagging point A, i.e., is lagging in phase. In either case, forward or backward conjugation, the wave emerging from the conjugator has the complex amplitude $a(x, y)e^{-i\phi(x,y)}$.

A phase conjugator (backward variety) is like a mirror, in that it obeys the basic mirror law, that the angle of reflection equals the angle of incidence, but in a rather different way, as Figure 2 shows. In fact, the reflected ray travels back on itself. A simple way to generate a conjugate wave, and one that has been known for a very long time, is with the use of an array of retroreflectors, or corner cubes. As shown in Figure 3, each ray undergoes two reflections and emerges traveling in exactly the opposite direction, although with some displacement. If we considered the third dimension, there could be three reflections. The amount of displacement depends on where the ray strikes; if it strikes near the apex of the corner cube, the ray displacement will be slight; if at the edge, the displacement is greatest. Thus, the image formed in reflection is degraded. This geometrical degradation is greater for larger corner cubes. Thus, why not make the cubes very small? Then, however, the image is degraded by diffraction effects, since each corner cube constitutes a small aperture. Thus, there is an optimum size for the cubes; if too large, geometrical factors degrade the image; if too small, diffraction effects degrade the image.

Phase Conjugation by Holography

With holography, however, came a method of conjugating a wavefront that yields high-resolution imagery. The holographic process inherently produces a phase conjugate image, as was noted by Gabor in his early papers on holography. The holographic process is illustrated in Figure 4. For simplicity, we assume the object to be a planar transparency $s(x, y)$. A coherent light wave, generally a plane or spherical wave, passes through the transparency and a diffracted field u is produced at the recording plane P, as in Figure 4a. The field u is typically the Fresnel diffraction pattern of the object s, but if the propagation distance is sufficiently great, u could be the Fraunhofer pattern of s. Of course, lenses can be placed between the object and recording planes, thus modifying the diffraction pattern.

The field u is combined with a reference wave u_o, which is generally an off-axis planar or spherical wave. The two beams produce an interference pattern with intensity

$$I = |u_o + u|^2 = |u_o|^2 + |u|^2 + 2Re\, u_o^* u \qquad [1]$$

where the final term can be written as $u_o^* u + u_o u^*$; the * denotes complex conjugate. The process of forming the interference of the two beams, i.e., forming the intensity of their sum, has generated the conjugate term u^*. When the intensity is recorded, as on photographic film, for example, both terms are stored on the film. Suppose for simplicity that u_o is a plane wave $a_o e^{i2\pi f_o x}$, where a_o is a constant, and $f_o = (\sin\theta)/\lambda$, where θ is the angle that the reference wave makes with the optic axis and λ is the wavelength of the light. Also, u can be

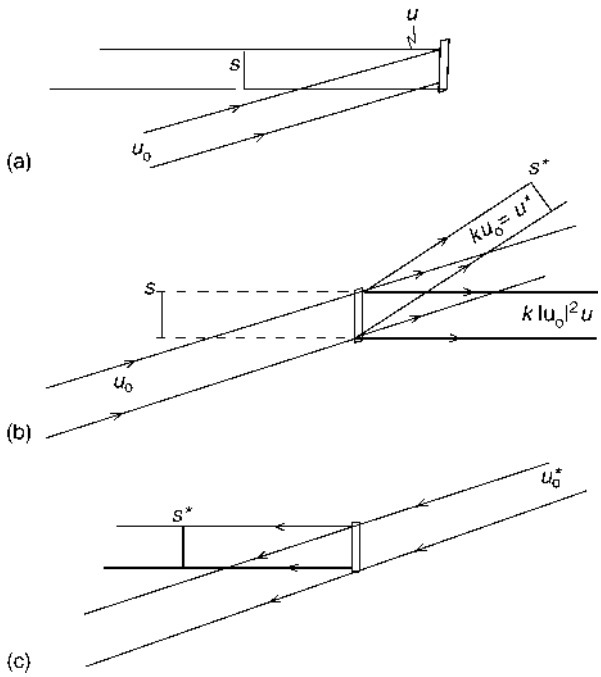

Figure 4 The holographic process. (a) Making the hologram; (b) reading out the hologram, using a readout beam that duplicates reference beam; (c) reading out with reversal of reference beam.

written as a product of an amplitude term a and a phase term $e^{i\phi}$, or $u = ae^{i\phi}$. Let the photographic plate, after exposure and development, have a transmittance t proportional to the intensity I, or $t = kI$, with k a constant; this is an oversimplification of the recording process, but preserves the essence of the holographic process. Now let the photographic plate, which after development becomes a hologram, be illuminated with a readout wave that, for simplicity, we let be u_o, a duplicate of the reference wave. Then, the field emerging from the hologram is

$$u_e = u_o k\left[|u_o|^2 + |u|^2 + u_o^* u + u_o u^*\right] \qquad [2]$$

Only the last two terms are of interest. These can be written as $k|u_o|^2 u$ and $ku_o^2 u^*$, respectively, or alternatively, as $ka_o^2 ae^{i\phi}$ and $ka_o^2 ae^{i(4\pi f_o x - \phi)}$. The term $k|u_o|^2 u$ represents a wave that is, to within a constant, identical to the object wave u; the other term represents a wave that is, to within a constant and a linear phase term, a conjugate of the original wave. The factor $e^{i4\pi f_o x}$ has the physical meaning that the wave is propagating in a direction 2θ given by $\sin 2\theta/\lambda = 2f_o$. The readout waves are shown in Figure 4b.

The so-called true wave u and the conjugate wave u^* thus propagate angularly away from each other and from the direct transmitted beam, which carries the uninteresting components $|u_o|^2$ and $|u|^2$.

However, a more suitable way to form the conjugate wave is to illuminate the hologram with a readout wave that is, again, a duplicate of the original reference wave, but propagating in the opposite direction, as in Figure 4c. The conjugate wave is now propagating along the z axis, along the path of the original object wave. It is readily shown that this counterpropagating wave is identical in shape to the original object wave recorded by the hologram, except that it travels in the opposite direction, exactly the situation portrayed in Figure 2. This wave propagates so as to form an image that is, to within a constant, identical to the original object, except for the counterpropagation, which produces a conjugate image $s^*(x, y)$. The image intensity, however, is $|s|^2$, which is identical to the original image intensity.

This conjugate wave was, in the early days of holography, a detriment, since without the use of the off-axis method (reference wave introduced at an angle to the object wave) the two images were inseparable, and the true image could be observed only against the background of the defocused conjugate image. Thus, an out-of-focus image overlaid the true image, making it noisy.

In 1962, the same year that the off-axis reference wave for separating the two images was introduced, a possible use for the conjugate image was suggested. If the original propagation path of the object wave were aberrated, as would occur if the wave propagated through an irregular medium, then the conjugate wave propagating through the medium would undergo a correction and would arrive at the original object position perfectly corrected, resulting in a sharp image, a process sometimes called wavefront healing. The idea lay dormant for 3 years, until in 1965, experimental demonstration of this idea was reported. In one case, the object was a transparency containing opaque lettering against a transparent background and the inhomogeneity was a piece of frosted glass, which resulted in extremely severe aberrations, so that the object wave, having traversed the diffuser, was no longer capable of forming an image. However, when the distorted wave was recorded holographically, conjugated, and sent back through the same diffuser, the result was a highly legible image.

The principle of wavefront healing is illustrated in Figure 5, when the inhomogeneity is an extended, or three-dimensional, structure. A point object O sends a spherical wave into the inhomogeneity. As the wavefront propagates, it becomes distorted, and loses its sphericity. When it emerges from the inhomogeneity, the wavefront is highly distorted.

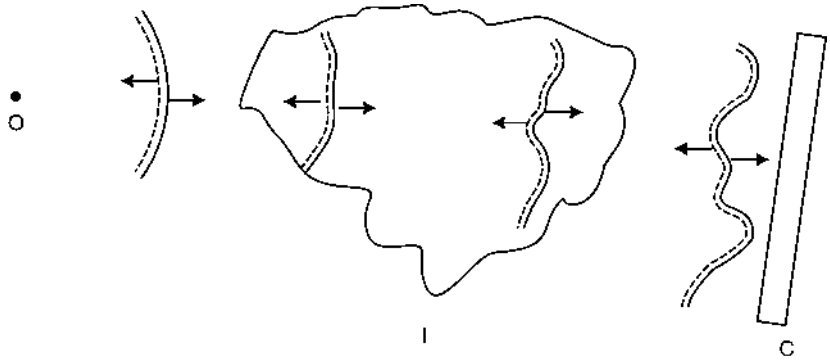

Figure 5 The wavefront healing process. C is the conjugator, I is the inhomogeneity, and O is a point object.

The wave is conjugated at the conjugator C and emerges as a back-propagating wave that retraces the original wave path through the inhomogeneity, gradually losing its aberrations and looking more like a spherical wave. When it emerges from the inhomogeneity, it is a perfect spherical wave that converges to a point image that coincides with the original object. At each plane throughout the inhomogeneity, the back-propagating conjugate wave (dashed lines) has the same shape as did the original wave when it crossed that plane (solid line).

This interesting wavefront healing property of holographic phase conjugation was the beginning of a new branch of optics, called phase conjugation, giving rise to a number of applications. However, the holographic method had the problem that it did not occur instantaneously, but had to await the processing of the photographic film on which the hologram was recorded. This delay limited the range of potential applications. Later, in the 1970s, nonlinear crystals became available that did the equivalent of holography, but in real time, or near real time – a process known as four wave mixing. There are yet other types of real-time phase conjugators with nonlinear devices.

Pseudoscopic Image

As long as the object is only two-dimensional, a function of x and y only, with no extent along the axial dimension, z, the intensity distribution of the true and conjugate images is identical. The conjugateness, being a pure phase phenomenon, disappears when the observable is the image intensity $|s|^2$, and not the complex amplitude s. Since the intensity is always the observable, whether the image is viewed by eye, or with a camera, the conjugateness generally is of no significance.

However, when the object for holography is three-dimensional, the situation is different, and the conjugate image exhibits some strange properties. These arise from conflicting cues pertaining to the three-dimensionality of the image. To explain, we review some basics about how a three-dimensional world is perceived by a viewer. First, there is the stereo effect, resulting from the viewer having two eyes, with each eye seeing an object from slightly different positions, a phenomenon called binocular disparity. The mind uses this stereo effect to perceive three-dimensionality. But there are other cues that also contribute to perception of three-dimensionality. Another is parallax; as the viewer moves his head, objects farther from the observer change their angular position more slowly than objects close by. The observer can sense the three-dimensionality of an object, even with only one eye, by moving the head, thereby inducing parallax. Also, there is interposition, whereby near object portions block the viewing of object portions lying behind them, again giving a sense of depth. There are still other cues, more subjective, such as perspective, parallel lines converging toward infinity, shadowing, etc. With these observations in mind, we consider the hologram under illumination and the two images that it forms (Figure 6). As depicted in Figure 6, these two images lie in mirror symmetrical positions relative to the hologram. Both images have the same side facing the hologram. It must be this way, because when the hologram was recorded, it was side **a** that faced the recording plane. And both images must therefore have side **a** closer to the hologram.

When a viewer observes the true image (a virtual image), he views it looking through the hologram, and he sees the image at the position where the original object was when the hologram was recorded, and it is, in its visual properties, indistinguishable from the actual object; it has all of the three-dimensional and the parallax properties that the original object would exhibit. However, when the viewer observes the conjugate image, he must view it

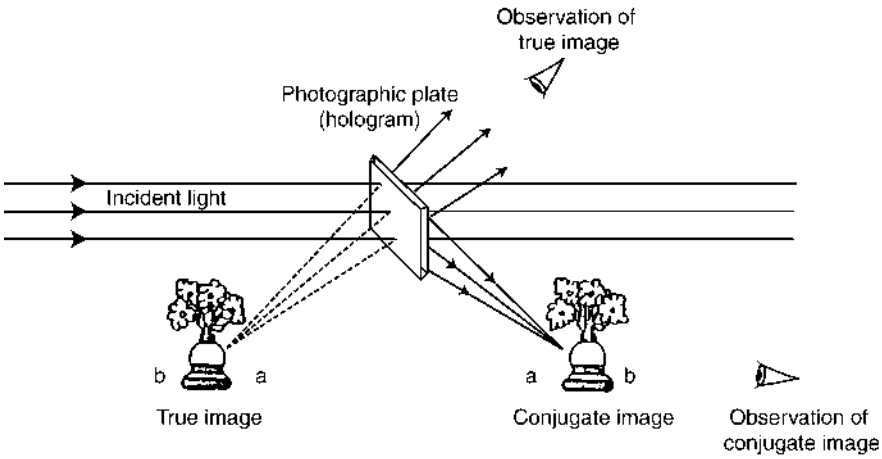

Figure 6 Formation of the true and conjugate images from a hologram.

from the side of the image opposite from the hologram, i.e., he views the image from the back side. When a viewer observes the conjugate image, he sees it from the back side (side **b**). All of the above cited cues are in agreement about the orientation of the image, viz., that side **b** is closer to the observer than is side **a**. That is, all cues but one. The conjugate image is formed as the hologram recording plate saw it, with side **a** being closer. Thus, all interpositions are formed with the same relation that the object had with respect to the recording plate. The side **a** is the part of the object that was closer to the hologram, so that when interpositions occur, side **a** blocks side **b**. Since the observer thus sees side **a** partially obscuring side **b**, he must conclude that side **a** is closer than is side **b**. However, this observation is in direct conflict with the other cues, especially the stereo and the parallax cues, which are themselves strong cues, as is the interposition cue. The conflict of these cues, with two of them interpreting the image as being formed with one depth orientation, the other with the opposite, gives a number of strange effects. For example, as the viewer moves his head so as to induce parallax, the object appears to rotate. If the image is that of a human head, for example, the observed effect is that as the viewer changes his position, the head appears to rotate in synchronism with the observer, as if the image had its gaze fixed on the viewer, an effect that is often perceived as being eerie.

Similarly, if the object were two flat plates, (Figure 7) in the situation depicted, the planes have no overlap, i.e., no interposition, as seen by the observer. And the observer unequivocally perceives plate A as being closer than plate B. However, if the viewer moves his head to a position where

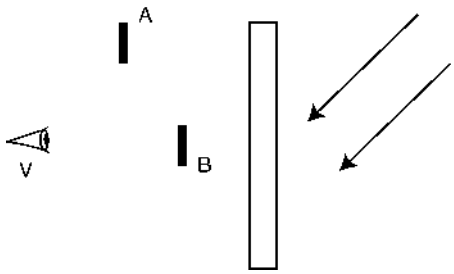

Figure 7 Pseudoscopic conjugate image formed from an object consisting of two plates. The viewer V perceives plate A as being closer than B and there are no conflicting cues, since there is no interposition. However, if the observer moves his head upward, there is partial interposition, thus conflicting cues, and the sense of three-dimensionality is lost.

interposition occurs, the conflict of cues arises, A becomes partially blocked by B, which is quite correct as observed at the position of the hologram, but is quite incorrect from the viewer's vantage point. The result of this conflict is that the viewer, who previously perceived the three-dimensionality, now sees the three-dimensionality perception suddenly collapse, and the sense of three-dimensionality is lost. Other strange perception anomalies can also be produced. Such images, with conflicting depth cues, are called pseudoscopic, as opposed to orthoscopic images, where the cues do not conflict.

One Way Phase Conjugation

A significant limitation of phase conjugation aberration correction is that the image and object are on the same side of the inhomogeneity. One would usually want them to be on opposite sides. In a modified form

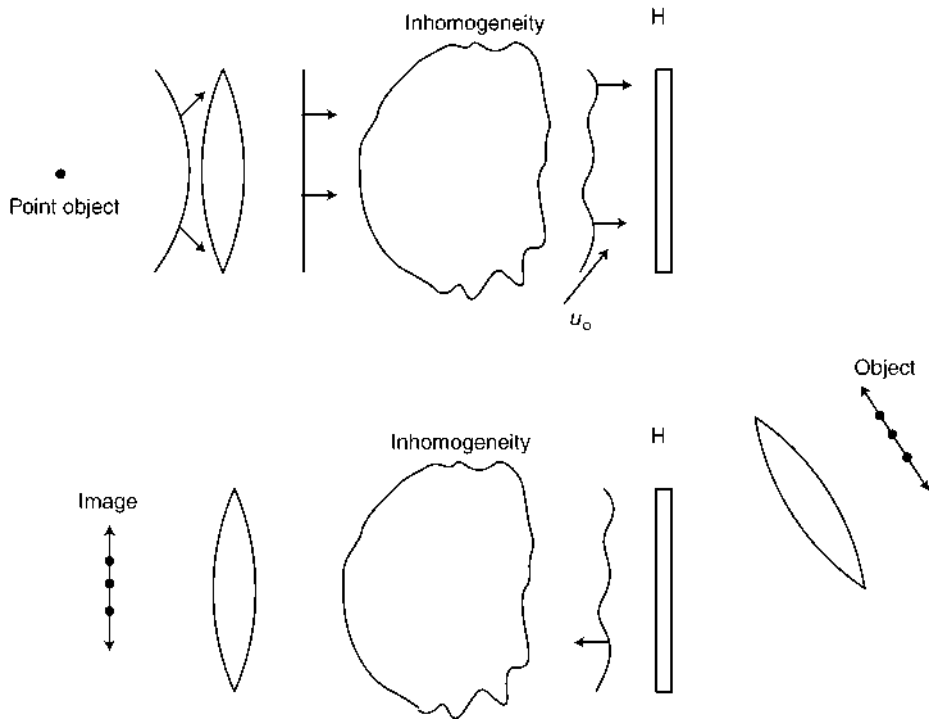

Figure 8 One-way phase conjugation. Top: Making the hologram. Bottom: The viewing process. We show three points on the object. The center point forms a well-corrected image, but the other two points produce imperfectly corrected images.

of the phase conjugation process, shown in Figure 8, this result is achieved. The first step of the process is carried out by passing a plane wave, instead of the object wave, through the inhomogeneity. This plane wave originates from a point source located at the focal plane of a collimating lens. The plane wave thus carries information about the inhomogeneity to the recording plate, where it is combined with a reference wave and a hologram of the distorted plane wave is recorded. Thus, information about the phase defect is stored on the hologram.

Next, the hologram is read out, not with a counterpropagating replica of the reference wave as before, but instead with an object-bearing wave that passes through the hologram. This wave emerges from the hologram and now has on it the recorded aberration information. As it counterpropagates along the path of the original object wave, it passes through the inhomogeneity, whereupon the aberration is cancelled and a corrected image forms at the focal plane of the collimating lens, where previously the point source had originated. The goal is achieved; the object and the corrected image are now on opposite sides of the inhomogeneity.

This correction is only approximate. To explain, we think of the object as being a collection of many points. The radiation from each point is converted into a plane wave by the lens, with different object points producing plane waves propagating at different angles. These plane waves then impinge on the hologram and emerge with each one bearing the aberration stored on the hologram. They continue on their paths and enter the inhomogeneity. The plane wave component, whose path direction matches that of the reference wave used in recording the hologram, will emerge from the inhomogeneity as a perfectly corrected wave, thus forming a well-corrected point image at the position of the original point radiator. However, another plane wave component, traveling in an incorrect direction, will be laterally displaced from the proper position for entering the inhomogeneity, and the aberration correction will be imperfect, the degree of imperfection being in direct relation to this displacement error. The displacement error is the product of the angular error and the propagation distance between hologram and inhomogeneity. If the inhomogeneity is slowly varying as a function of x and y, and if the object does not have a large extent, so that the range of angles is small, the image degradation will be small.

If the last two conditions do not apply, two options are available. First, the distance between inhomogeneity and hologram can be reduced, either

by bringing the hologram closer to the inhomogeneity, or by imaging the hologram onto the inhomogeneity. The correction is now valid over a wider extent of the image, especially if the inhomogeneity can be treated as planar rather than extending over a volume. Alternatively, a scanning process can be carried out on the object beam falling on the hologram, so that in sequence, each plane wave component assumes the proper angle to counter-propagate along the path for perfect correction, and will form a well-corrected image point. This output spot then falls on a pinhole positioned to separate it from other, imperfectly corrected points. Then, over time, all points of the object will form well-corrected image points, and the entire object will thus be well-imaged. The one way phase conjugation method was first used to correct the spherical aberration of a lens, and the result was quite successful.

Phase conjugation has become an important branch of modern optics. The process is often carried out using photographic materials. More often nowadays, it is carried out using nonlinear crystals. Some of the crystal methods are essentially holography in real time, but others, such as phase conjugation by Brillouin scattering, go beyond holography.

See also

Nonlinear Optics, Basics: Nonlinear Optical Phase Conjugation. **Phase Control:** Wavefront Coding.

Further Reading

Caulfield HJ (ed.) (1979) *Handbook of Optical Holography*. New York: Academic Press.

Goodman JW (1996) *Introduction to Fourier Optics*, 2nd edn. New York: McGraw-Hill, chap. 9. This chapter is devoted to holography, with section 9.12.3 dealing with phase conjugation aberration correction.

Kogelnik H (1965) Holographic image projection through inhomogeneous media. *Bell Systems Technology Journal* 48: 2451–2452.

Leith EN and Upatnieks J (1965) Holograms: their properties and uses. *SPIE Journal* 4: 3–6. This is the earliest publication demonstrating the principle of wavefront healing by holographic phase conjugation.

Smith HM (1975) *Principles of Holography*, 2nd edn. New York: Wiley.

Wavefront Coding

W T Cathey, University of Colorado, Boulder, CO, USA

E R Dowski, CDM Optics, Inc. Boulder, CO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

We describe a new paradigm for designing imaging systems that have performance that is otherwise unobtainable. These are hybrid systems that use the optics of the system to acquire an 'intermediate' image, which is processed to produce the final image. The optical components of the system are designed jointly with signal processing to optimize overall performance. It is possible to extend the depth of field of an imaging system, for example, by a factor of eight to ten. On the other hand, by concentrating on the increase of the depth of focus in the detector plane, it is possible to use that increase in the depth of focus to make the system invariant to focus-related aberrations. By doing this, it becomes possible to design a single-lens imaging system with small f-number and large field of view. Another example of the sort of new imaging system that can be designed is a two-lens zoom lens that images with a large field and a very short overall length.

First, the term 'Wavefront Coding' is defined as employed with imaging systems, and a description is given of how such coding is achieved. Examples of means of applying Wavefront Coding are given along with the analysis and design tools that are useful in designing hybrid optical/digital imaging systems. Finally, some potential future applications of Wavefront Coding in imaging systems are suggested.

Background and Basic Concepts

Some early suggested modifications to imaging systems can be viewed as wavefront modification, even though they were not described in that manner at the time. These include the work of Ojeda-Castañeda that described masks to block portions of the lens to extend the depth of field. In addition, it has long been known that the depth of field is slightly greater in lenses that have spherical aberration. In both of these cases, the wavefront is modified, in one case by changing the wave amplitude, and in the other by a deviation from the traditional ideal phase front. However, in both cases, the increase in the depth of

field was only a few percent. Häusler extended the depth of field by requiring that the focus be continuously changed during the exposure time of the camera. This resulted in a modulation transfer function (MTF) for the imaging system that was the incoherent superposition of a spread of MTFs for the different foci during the exposure. To form an acceptable image, signal processing was used.

The use of Wavefront Coding here refers to new optical systems arrived at by designing the modulation of the wave and signal processing to achieve previously unobtainable imaging modalities. These new imaging systems can have more than an order of magnitude increase in depth of field, for example. The coding of signals so that particular information is optimally conveyed is not new. For example, in radar, coding of the pulses is done to optimally provide information concerning a target's range. At the radar receiver, signal processing is required to extract the range information. In a similar manner, an imaging system can have a coded wavefront in order to provide range or misfocus information in a manner such that signal processing can easily extract the desired information in the scene. Alternatively, the wavefront can be modified so that the imaging system is invariant to object depth or to misfocus of the system, so that, after processing, good resolution can be obtained over a large image depth.

A primary feature of Wavefront Coding in imaging is that the optical image-gathering portion of the system can be designed in combination with the design of the signal processing. This is normally not done in the design of imaging systems. It is done, however, in tomography, coded aperture imaging, and sometimes, interferometric imaging. In 1984, a group pointed out the potential of increasing the performance of imaging systems by jointly designing the optics and the signal processing. Since that time, major advances have been made in doing this. Now, several analysis and design tools are available, and are described in the following section.

Theory and Applications of Wavefront Coding

Wavefront Coding refers to a modification of the imaging system by placing specialized phase plates (or phase and amplitude masks) in the aperture stop, or in an image of the aperture stop, of the imaging system. This could be a phase plate as a separate optical element, or, if a surface of a lens is close to the aperture stop, the Wavefront Coding surface can be superimposed onto the normal lens surface. Processing to decode the image that is formed by the detector array (the intermediate image) is related to the coding that is impressed on the wavefront. The term 'signal processing' also includes the spatial integration provided by the width and height of the pixels. The final image is hence a function of the optics that form the intermediate image and the signal processing that produces the final image.

In the following section, we discuss the Woodward function as used in analysis and design of hybrid imaging systems, and show visualizations of the effect of Wavefront Coding on the image performance of two dramatically different types of systems. Then, other design techniques are discussed for imaging systems using Wavefront Coding and several examples of systems are given where the depth of field of imaging systems is extended by at least an order of magnitude. By concentrating on the increase in the depth of focus, rather than the depth of field, very simple imaging systems can be designed where Wavefront Coding makes the imaging system invariant to focus-related aberrations, and the shape of the lens is used to minimize the other aberrations.

Wavefront Coding Theory

The Woodward function

P. M. Woodward showed that the capability of a radar system to estimate both the range and the velocity of a target can be described by a mathematical function that became known as the Ambiguity Function. The Ambiguity Function shows how well systems that use different types of radar pulses can measure the transit time of the pulse (hence the range to the target) and the Doppler shift in the frequency of the returned pulse (hence the velocity of the target). Because the parameters, range, and Doppler, are measured with different accuracy, there is ambiguity in the measurement, and the Ambiguity Function is a display of the tradeoffs that are possible, Brenner, Lohmann, and Ojeda-Castañeda showed that the same mathematical function, with differently defined variables, can represent the Optical Transfer Function (OTF) of a one-dimensional imaging system.

The Ambiguity Function for a radar system is given by

$$A(u,v) = \int P(x+u/2)P^*(x-u/2)\exp(\mathrm{i}2\pi vx)\mathrm{d}x \quad [1]$$

where $\mathrm{i} = \mathrm{sqrt}(-1)$, * denotes complex conjugate, u is time, v is frequency, x is a dummy variable, and $P(x)$ is a complex function representing the transmitted waveform of the radar system. The OTF of a one-dimensional imaging system for different values of

misfocus can be written as

$$H(u,\psi)=\int\left\{P(x+u/2)\exp(j(x+u/2)^2\psi)\right\}\times\left\{P^*(x-u/2)\exp(-j(x-u/2)^2\psi)\right\}\mathrm{d}x \qquad [2]$$

where $P(x)$ represents the complex pupil function of the imaging system. The misfocus parameter ψ can be shown to be equal to the physical parameters:

$$\psi=(\pi L^2/(4\lambda))(1/f-1/d_\mathrm{o}-1/d_\mathrm{i})=2\pi W_{20}/\lambda=kW_{20} \qquad [3]$$

where L is the one-dimensional length of the aperture, and λ is the center wavelength of the narrowband illumination. The distance d_o is measured between the object and the first principal plane of the lens, and d_i is the distance between the second principal plane of the lens and the image plane. The focal length of the lens is given by f. The wavenumber is given by k and the traditional misfocus aberration constant is given by W_{20}. A comparison of the form of the ambiguity function and the OTF shows that:

$$H(u,\psi)=A(u,u\psi/\pi) \qquad [4]$$

The mathematics of the Woodward function as used for the Ambiguity Function has been studied extensively. Hence one can apply the theory that was worked out for the Ambiguity Function to the study of imaging systems. This gives a tool that is, in many ways, more powerful than other forms of three-dimensional optical transfer functions. For example, it can be shown that the squared volume under the Woodward function is a constant. This means that whatever one does to modify the Woodward function, there is a constraint on what can be done. Another constant is the power under a vertical cut of the Woodward function at a particular value of spatial frequency. This means that there is constant power for all misfocus values at any given spatial frequency. It can be bunched up about the normal in-focus position, or it can be spread out as shown below, but the squared area does not change.

Examples of the Woodward function for different imaging systems

A display of the Woodward function for a high-quality, one-dimensional lens is shown in Figure 1a. The horizontal line and the slanted line represent in-focus and out-of-focus slices through the Woodward function. A projection onto the horizontal axis gives

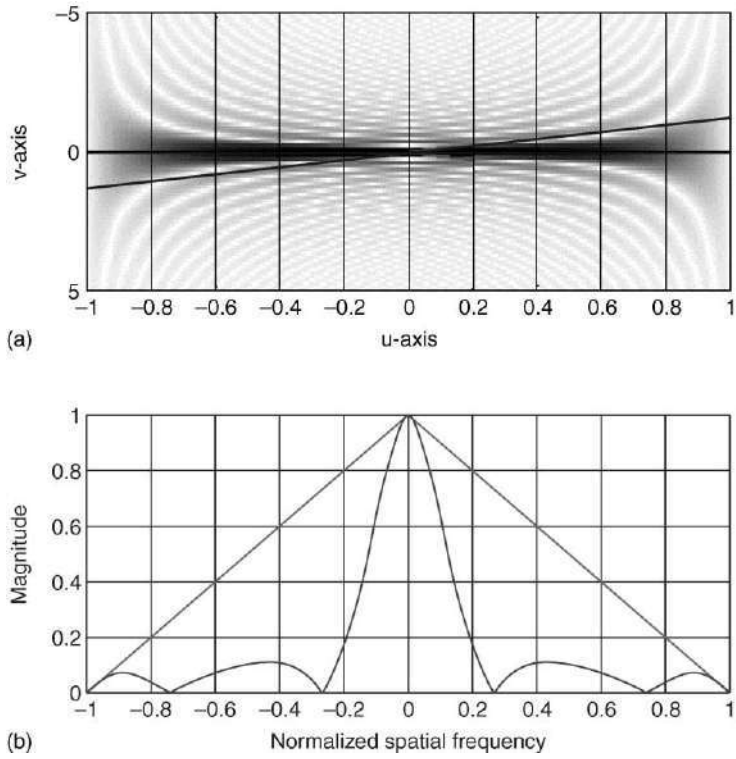

Figure 1 (a) The Woodward function for a high-quality one-dimensional lens with slices that represent in-focus and out-of-focus imaging. (b) The corresponding values of the modulation transfer function for those two cases.

the OTFs for those two focal conditions. The magnitudes of those two OTFs, the MTF, is shown in Figure 1b. Note that the out-of-focus MTF goes through a zero, where there is a phase shift, which gives a contrast reversal in the image.

Extended depth of field. Figure 2a shows the corresponding display of the Woodward function for a one-dimensional lens with a cubic phase function, $\phi = \alpha x^3$ added in the aperture stop. The appearance of the Woodward function immediately shows that this imaging system will have an increased depth of field, because the Woodward function shows that the energy is spread over a larger misfocus region. The slices in the Woodward function are taken for the same misfocus values as were used for the images in Figure 1. Notice that the MTFs of Figure 2b are very similar, and that the MTF for the misfocused case does not go through a zero. This implies that the imaging system will have an extended depth of focus.

The analogy to the Ambiguity Function of radar was key in selection of the cubic function for early demonstrations of extending the depth of field of an imaging system. Work in the radar field had shown that a radar pulse, where the carrier frequency increases quadratically with time, has an Ambiguity Function that is identical to the distribution in Figure 2a. Because a quadratic in time is a cubic in phase, the cubic phase function:

$$\phi = \alpha x^3 \qquad [5]$$

was selected as one that will provide the same Woodward function, and hence an increase in the depth of field. Many families of phase distributions have now been found that extend the depth of focus of an imaging system.

Antialiasing. When the resolution of the optics of an imaging system exceeds the resolution of the image detector array, the high spatial frequencies cannot be detected. They are not lost, however. When a spatial frequency is sampled below the Nyquist sampling rate, the high spatial frequency shows up, or aliases, as a lower spatial frequency. This aliased frequency is equal to the sampling frequency (determined by the pixel spacing) less the difference between the high frequency that has been undersampled and the sampling frequency. This results in an image with broad bands (low spatial frequencies) across the image as is often seen in television images of patterns with high spatial frequency content, such as striped shirts. Because the signal power at a given spatial

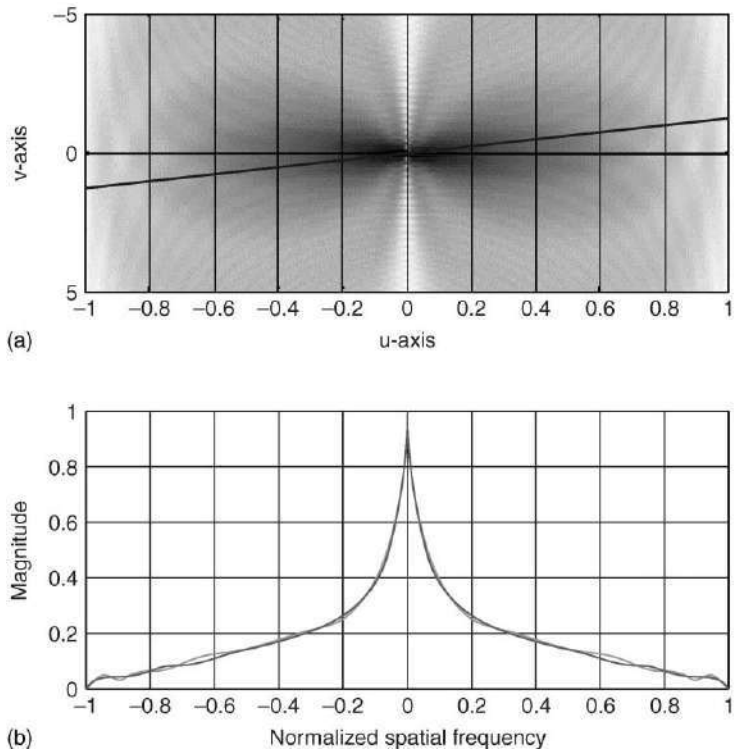

Figure 2 The Woodward function and the modulation transfer functions for a high-quality one-dimensional lens that has a phase plate with a cubic profile in the aperture stop. (a) The Woodward function showing a broadened pattern that is indicative of an increased depth of field. (b) The MTFs that are associated with the slices that represent in focus and out of focus.

frequency over all misfocus (the squared area under a vertical slice of the Woodward function at that spatial frequency) is constant, these frequency components cannot be removed. They can, however, be reduced at a given focus value by spreading the Woodward function vertically. Figure 3 shows a Woodward function that would be ideal. Below the cutoff spatial frequency, all of the signal power is in the focal plane, and beyond the cutoff frequency, the power is spread as much as possible over a large misfocus region. There is no assurance that a phase distribution can be found to give a desired Woodward function, however. Figure 4 shows the effect of one phase function, given by

$$S(r, \theta) = r^3 \cos(3\theta) \qquad [6]$$

upon the MTF of the imaging system. The dashed vertical line shows the cutoff spatial frequency that is

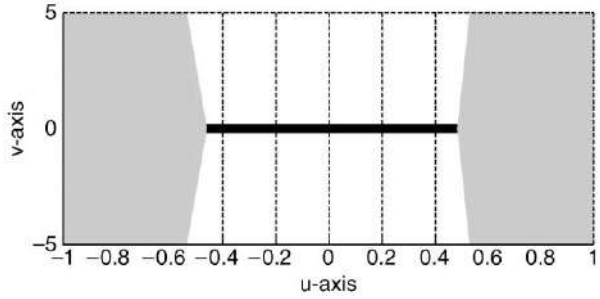

Figure 3 The Woodward function for an ideal antialiasing filter.

due to the sampling rate of the pixels. Note that in Figure 4, the magnitude of the Wavefront Coded spatial frequencies that are aliased have lower values than do those that are passed by a traditional full aperture F/2.5 system. Stopping the traditional system down to F/25 reduces the aliased MTF to an amount close to that obtained with the wide aperture and Wavefront Coding, but it also reduces the amount of light by a factor of 100.

Wavefront Coding in Imaging Systems with Extended Depth of Field or Focus

Figure 2 shows the Woodward function for one phase distribution (cubic in x and y for a two-dimensional imaging system) that increases the depth of field of an imaging system. The cubic function was selected by using analogies to the Ambiguity Function of radar. Even though it is doubtful that this phase distribution could have been found by using conventional lens design techniques, it is instructive to show the ray traces and point spread functions that are obtained when this phase distribution is used in the aperture stop.

Ray trace analysis

Figure 5 shows the ray trace from a point object through a conventional (one-dimensional) lens and through a lens that has been modified with the cubic profile of eqn [5]. Figure 5a shows the rays through a conventional lens coming to a focus, and Figure 5b shows the corresponding rays as they go through a lens with a cubic profile added. In the latter case, the

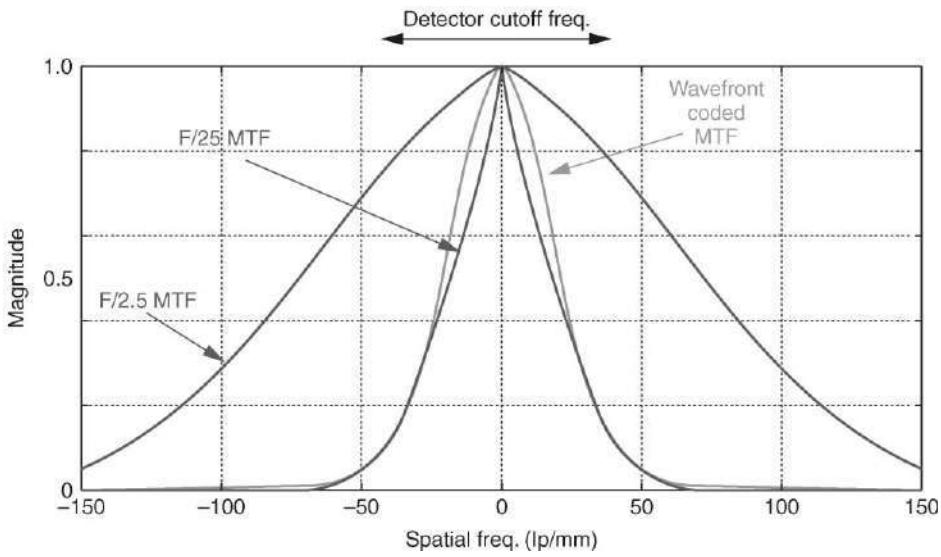

Figure 4 Normalized MTFs for traditional optics and Wavefront Coding. The full aperture F/2.5 system has a large amount of optical power beyond the detector cutoff frequency of 50 lp/mm. With Wavefront Coding the aliased optical power can be greatly reduced. The traditional system aperture needs to be reduced to F/25 to compare to the antialiasing performance of the Wavefront Coded system. This reduction reduces the light level by a factor of 100 in comparison with the F/2.5 system.

rays cross, but have a relatively constant ray density in the region of what was the focal region. Figure 6 shows experimentally acquired point spread functions (PSF) for both cases. Note that the PSF changes very little over the region of misfocus. After signal processing, the PSF can be brought back to that of the conventional imaging system when in focus. Figure 7 illustrates the effect of the signal processing upon the intermediate image that is acquired by the optics. The object is a tilted bar pattern, so that a portion is in focus, and the remainder of the bar pattern is so far out of focus that there is a contrast reversal in the image. With the system that has a coded wavefront to give an extended depth of field, the entire image appears to have the same blur, no matter which part of the tilted bar pattern is observed. This is because the same PSF is convolved with the object distribution for all regions of misfocus. Signal processing removes the effect of the modified PSF, resulting in the image of Figure 7b. Figure 7c shows a trace through the image of Figure 7b. The amount of the misfocus that can be tolerated is a design parameter, and is discussed in the next section.

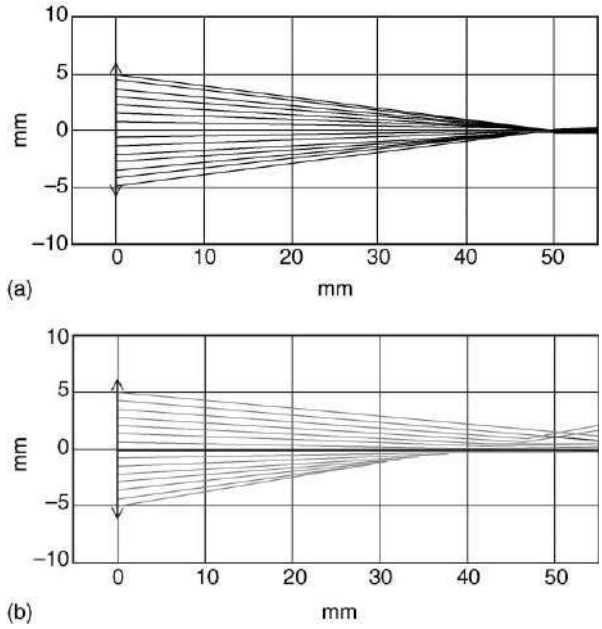

Figure 5 Ray traces from a point object through (a) a conventional one-dimensional lens, and (b) a lens that has been modified to have a cubic profile on the surface.

Signal processing requirements

It is necessary to do signal processing when using an imaging system that employs Wavefront Coding; the signal processing is needed to 'decode' the image that is formed by the detector array. This intermediate image is not a clear image. To obtain an intermediate image, the object distribution is convolved with a PSF, or impulse response, that has a shape similar to that

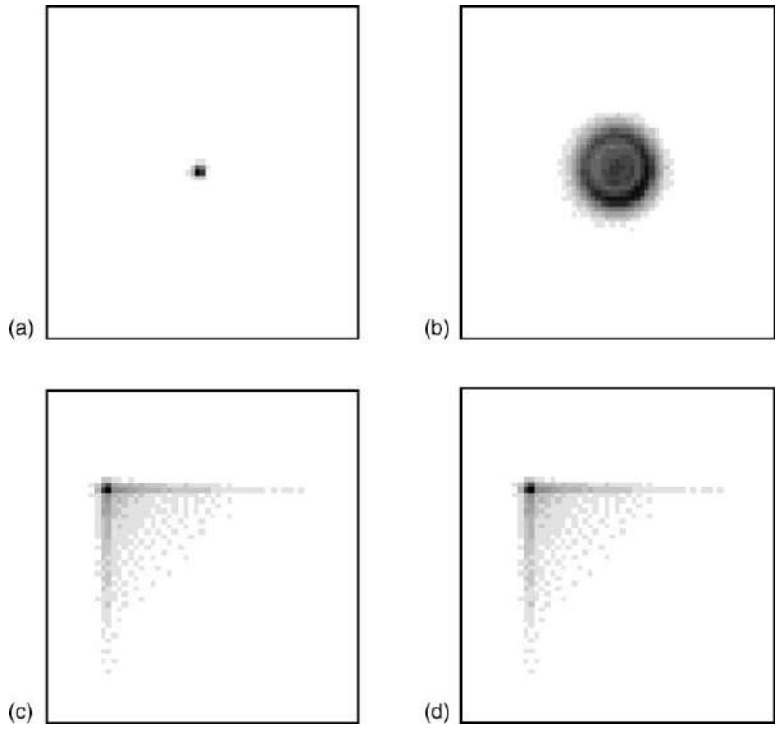

Figure 6 Point spread functions for a conventional lens in focus (a) and with large misfocus (b), and a lens that has been modified to have a cubic profile on the surface when in focus (c) and with a large misfocus (d).

of Figure 6c and d. To obtain the final image, the intermediate image must go through a deconvolution with the PSF. Because of the square law detection that occurs in the detector array, this deconvolution does not simply undo the effects of the convolution with the PSF that describes image formation. A straightforward digital convolution is required in most cases, and has been shown to run at video frame rates in software.

Parameter trade offs in imaging systems with extended depth of field

Depth of field and signal-to-noise ratio. The greater the phase modulation by the phase plate (with a cubic profile in this example), the greater the extension of the depth of field. The greater the phase modulation for extending the depth of field, the larger the PSF, and the greater the drop in the MTF, as seen in Figure 2b, which show the MTF of the imaging system. The use of the phase function to extend the depth of field causes portions of the MTF to sag as compared with the MTF of the in-focus conventional system. This reduces the signal-to-noise ratio. If there is a greater phase alteration, the MTF sags more.

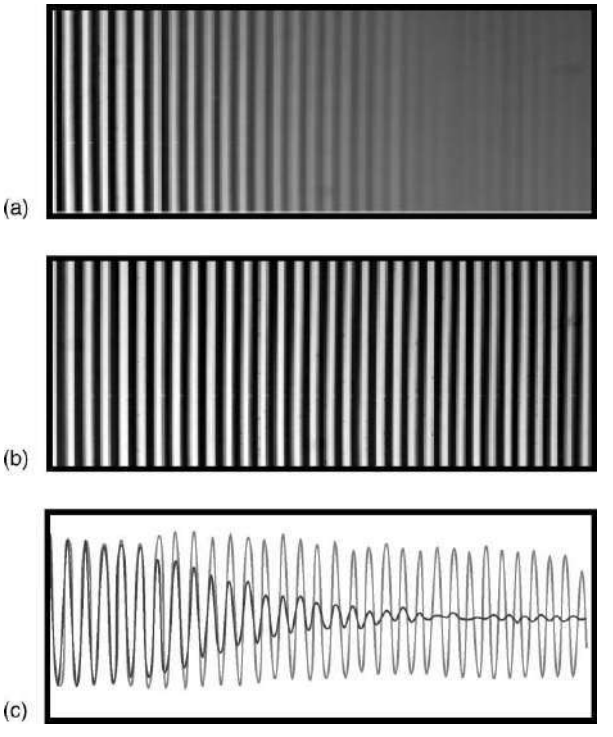

Figure 7 Image of a tilted bar pattern with an extended depth of field imaging system. (a) Image with a conventional imaging system. (b) Final image from the imaging system with Wavefront Coding. (c) Traces through the images to give quantitative comparisons of the image quality.

Focus budget and focus-related aberrations. Another set of trade offs can be seen by considering the imaging systems from the detector plane. Here, we see an increase in the depth of focus of the system. All of this increase in the focal depth can be used to increase the depth of field if the optics of the system are of high quality. Otherwise, the increase in the depth of focus can be used to compensate for focus-related aberrations. For example, chromatic aberration occurs when different colors focus to different planes. Figure 8 shows that if the depth of focus is extended, then there is a region where red, blue, and green, for example, are in focus. Figure 9 shows the red, blue, and green images of a US Air Force test pattern, as formed with a two-element imaging system with severe chromatic aberration, and with the same system when a Wavefront Coding element is used to extend the depth of focus.

Compensation can also be seen for the case of curvature of field, where the image of a plane object falls onto a curved surface. If an extension of the depth of focus causes the focal region to resemble a curved bar instead of a curved line, then it is possible to place a flat detector array in the curved region, and therefore obtain a good image after signal processing. Similar arguments can be made for other focus-related aberrations. This leads to the concept of a focus budget, where a portion of the increased depth of focus can be allocated to extend the depth of field by some amount, and other portions of the depth of focus allocated to provide invariance to focus-related aberrations.

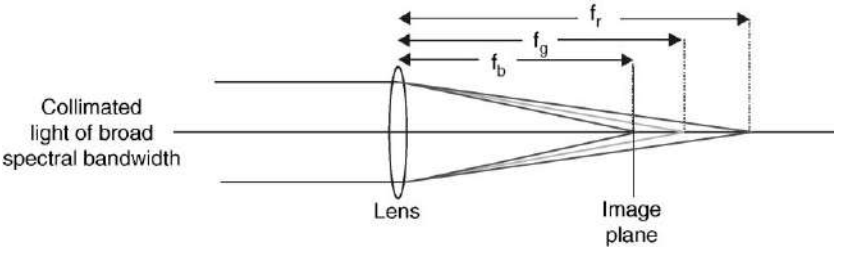

Figure 8 Ray traces showing different focal planes for red, blue, and green, and the effect of extending the depth of focus for each color.

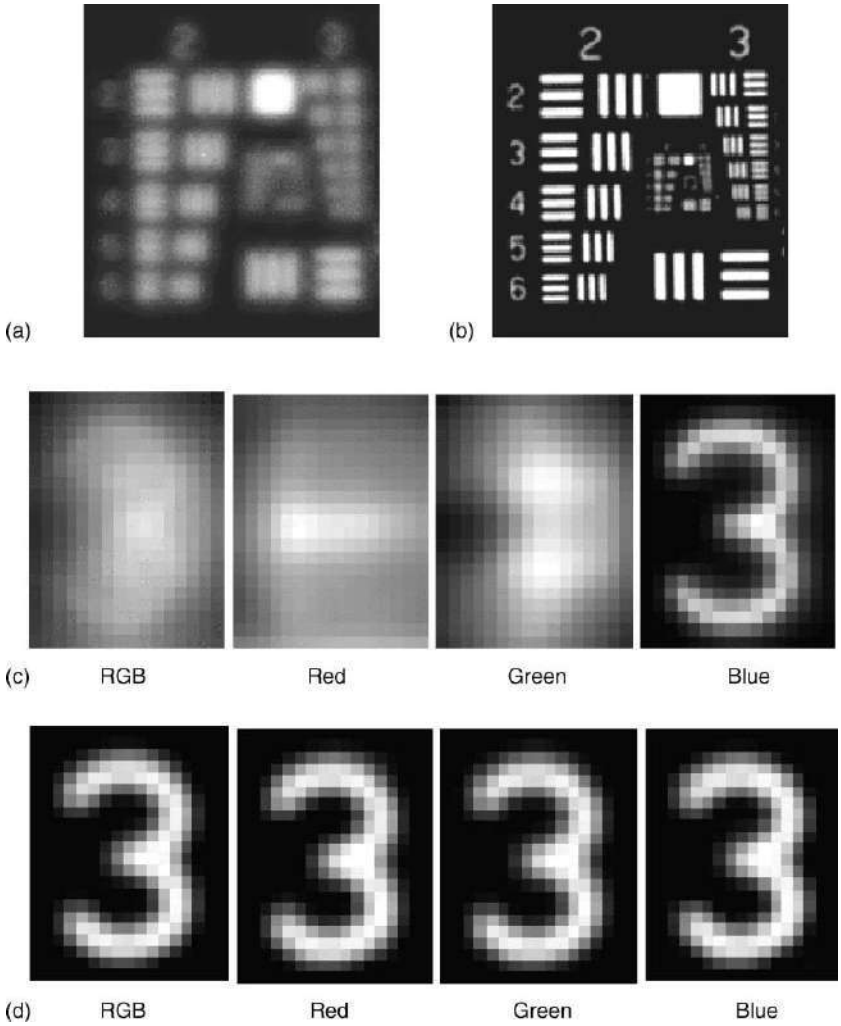

Figure 9 Color images of a US Air Force test pattern. (a) Images from a two-element lens of the center portion of the test pattern. (b) Images from an extended-depth-of-focus imaging system of the center portion of the test pattern. (c) Images from a two-element lens of a small number 'three' on the test pattern. (d) Images from an extended-depth-of-focus imaging system of a small number 'three' on the test pattern. Some residual effects remain due to different magnifications at different wavelengths.

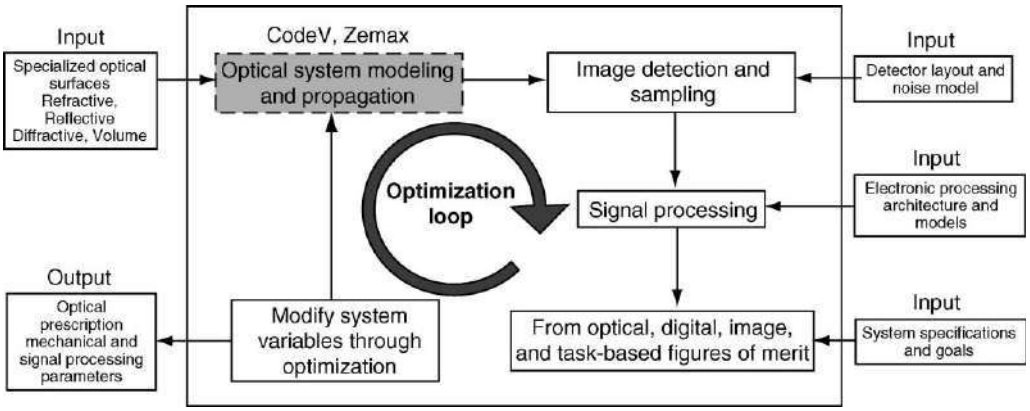

Figure 10 Schematic of a tool for designing imaging systems with Wavefront Coding. A ray trace determines the distribution in the exit aperture, and software determines the effect of the detector array and the signal processing. Then the ray trace and signal processing can be modified to optimize the entire system.

New design tools for imaging systems with wavefront coding

The Woodward function is a very good tool for analyzing imaging systems with a coded wavefront and is a good guide for someone who is expert in designing such systems. However, another approach is better for those designers who are more familiar with conventional lens design tools. If one combines a conventional ray-tracing tool with custom software, it is possible to jointly design the optical and the signal processing portions of the system. A schematic of such a system is shown in Figure 10. The ray-tracing tool is used to trace the rays and determine the distribution in the exit aperture. Various software tools then are used to take into account the effect of pixel size and spacing, and the signal processing. The criteria for optimizing the design can include the size of the digital filter, acceptable noise gain, and other parameters related to signal processing as well as the parameters of the optical elements.

Example applications of hybrid imaging systems with an extended depth of field or focus

Barcode and label readers. The images of Figure 7 show that image-based bar code reading is a good application for imaging systems that are based on Wavefront Coding. A bar code or label reader need not be positioned nearly as accurately as without the

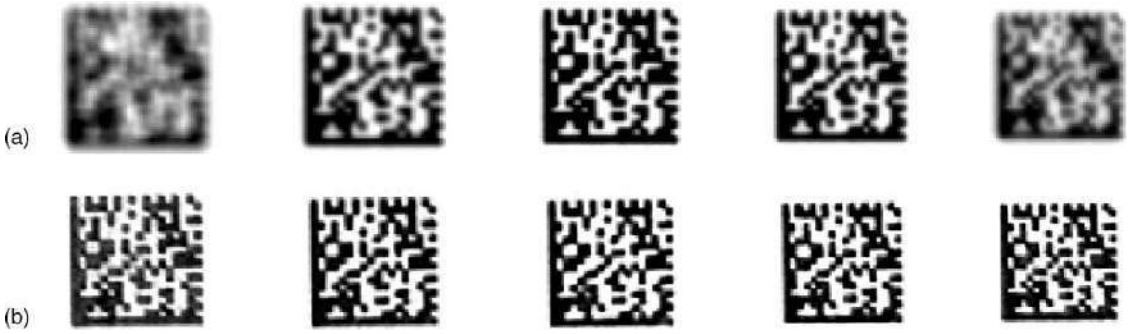

Figure 11 Images of a 2D bar code as the object position is varied for (a) conventional images and (b) Wavefront Coded images after signal processing for the same object positions as in (a).

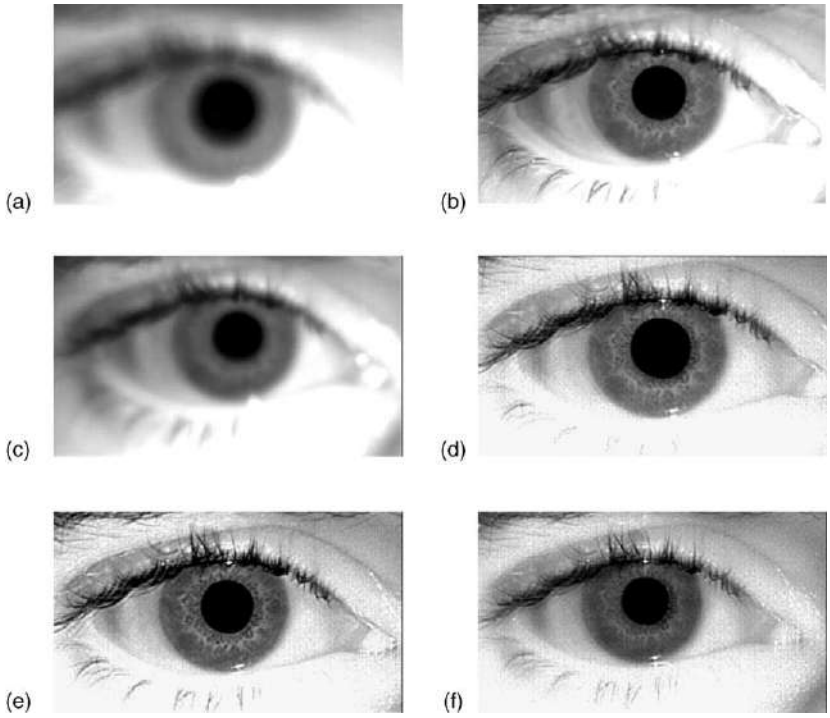

Figure 12 Images of an iris taken with infrared illumination. The top row shows images with a traditional CMOS camera. The lower row is taken with a camera using wavefront coding to extend the depth of field. The misfocus distance in the right and left columns is the same.

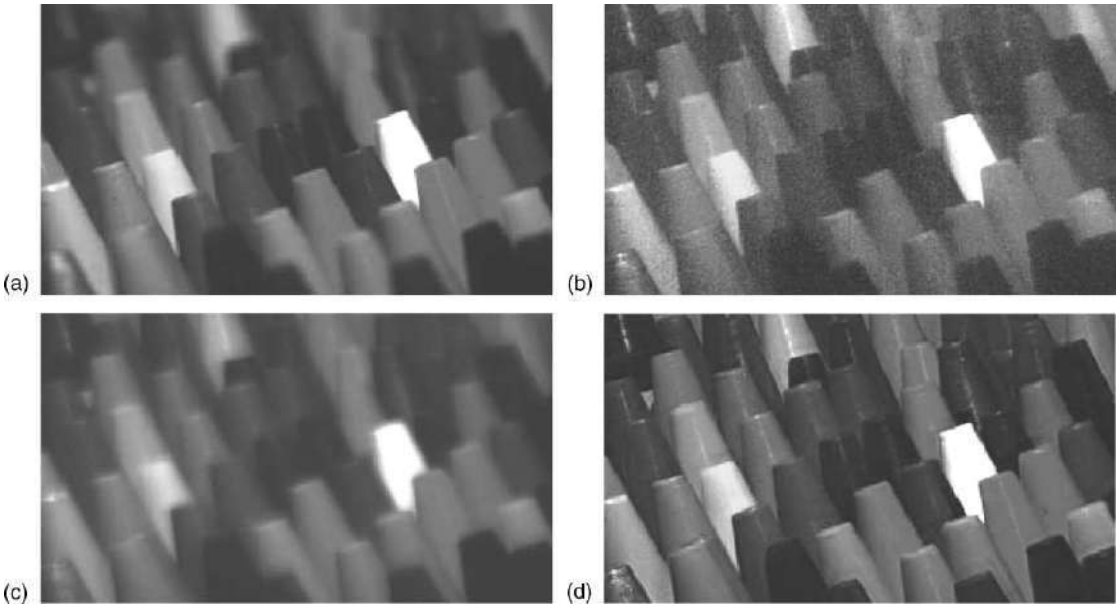

Figure 13 Images from a digital camera (a) Conventional camera with the lens wide open and focused on the white crayon. (b) Conventional camera with the lens stopped down to increase the depth of field, with the resulting increase in noise due to a lower light level. (c) Intermediate image of a camera that was modified with Wavefront Coding to have an extended depth of field. (d) Final image with the modified camera.

extended depth of field. Images of 2D bar codes imaged with and without Wavefront Coding are shown in Figure 11. Images in Figure 11a show the image of a 2D bar code formed by a conventional imaging system as the position of the object is varied. As with most imaging systems with fast optics and moderate magnification, the depth of field of the system is smaller than the application would prefer. Images in Figure 11b show the Wavefront Coded images of a 2D bar code after signal processing for the same object positions as (Figure 11a). These images are sharp and clear over a very large change in the distance to the bar code.

Biometric imaging systems. Biometric imaging applications, such as fingerprint or iris recognition, provide the means to eliminate passwords, secure E-commerce, and securely identify humans or animals. In traditional imaging technology, quite often the trade-off between light throughput and depth of field makes conventional biometric systems difficult to use or not sufficiently reliable. By using Wavefront Coding, the depth of information gathered can greatly increasing the ease of use and possibly reducing the cost of the systems.

Figures 12a,b, and c shows example images of an iris taken with near IR illumination and a traditional CMOS imaging system at three different planes of focus. At best focus (Figure 12b) the detail of the iris is sharp and clear. This iris detail is used by iris recognition algorithms to accept or reject the person.

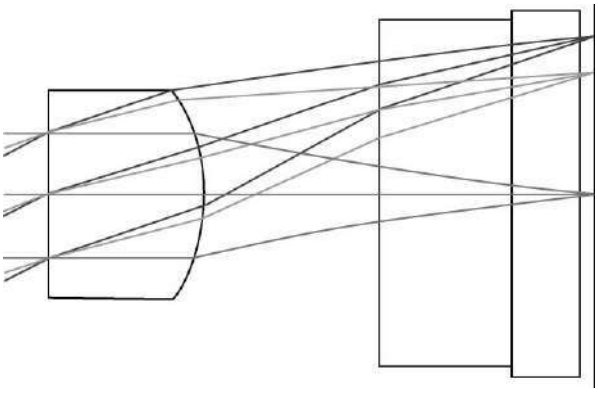

Figure 14 A schematic of a fast single-lens imaging system that has good performance over a wide field of view and the full visible range.

With slight movement towards (Figure 12a) or away (Figure 12c) from the imaging system the detail of the iris is lost. Figures 12d,e, and f show example iris images taken with the same near-IR illumination and a CMOS Wavefront Coded imaging system. At best focus (Figure 12e), the iris detail is as high as that from the traditional system. Movement towards (Figure 12d) or away (Figure 12f) from the imaging system results in images with essentially a constant level of iris detail. The size of the iris changes through magnification change. Increasing the range at which clear images can be made extends the useful image capture volume, enabling a more flexible system that is easier to use.

Color images with digital cameras. A conventional digital camera has a small depth of field as shown in Figure 13a, where the camera was focused on the white crayon. In a conventional imaging system, the only way to increase the depth of field is to stop down the lens. As seen in Figure 13b, this reduces the light that is available, and results in a very noisy image. The intermediate image of an imaging system that has an extended depth of field is shown in Figure 13c, where it is seen that the result of convolving the object distribution with the modified PSF is an intermediate image that appears to have the same blur over the entire image. After signal processing of the intermediate image to obtain the final image, a clear image is seen with good signal-to-noise ratio.

Inexpensive imaging systems with high-quality imaging characteristics. By exploiting the extension of the depth of focus that is made possible with Wavefront Coding techniques, one can design imaging systems that have fewer optical elements than traditional designs that do not use Wavefront Coding. In high-volume applications, reduction in the number of optical elements can lead to greatly reduced costs. Figure 14 shows a schematic of a fast, single-lens Wavefront Coded imaging system that is very compact and has the performance of a longer lens that traditionally requires two or more optical elements. In this example, the entire focus budget was used to compensate for focus-related aberrations, and the shape of the lens was used to compensate for distortion and coma. Even zoom systems can be greatly simplified with the new design techniques. Figure 15 shows a fast two-element Wavefront Coded zoom lens that images over a wide field with performance that is traditionally obtained only with three or more optical elements. Both of these imaging systems use non-rotationally symmetric forms of phase functions and are different from the ones of eqns [5] and [6].

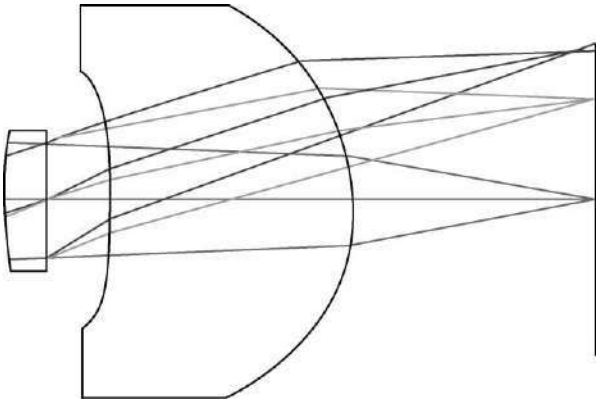

Figure 15 A schematic of a fast two-element zoom system with good performance over a 2.5 × zoom and a wide viewing angle.

Large depth of focus microscopes. A microscope can be modified to employ Wavefront Coding by either modifying the surface of one of the lenses in the objective, or by inserting a phase plate into the system at some appropriate position. One easy means of doing the modification is to place a Wavefront Coding phase plate in the slider position of a microscope, as shown in Figure 16. In this case, the optics is of high quality, and the entire focus budget was allocated to increasing the depth of field. A modified Zeiss microscope was used to obtain the images shown in Figure 17. An image of a diatom in Figure 17 shows the comparison with a conventional microscope in (a) and as with a Wavefront Coded microscope in (b). Notice that the greater depth of field gives more of an appearance of depth in the image.

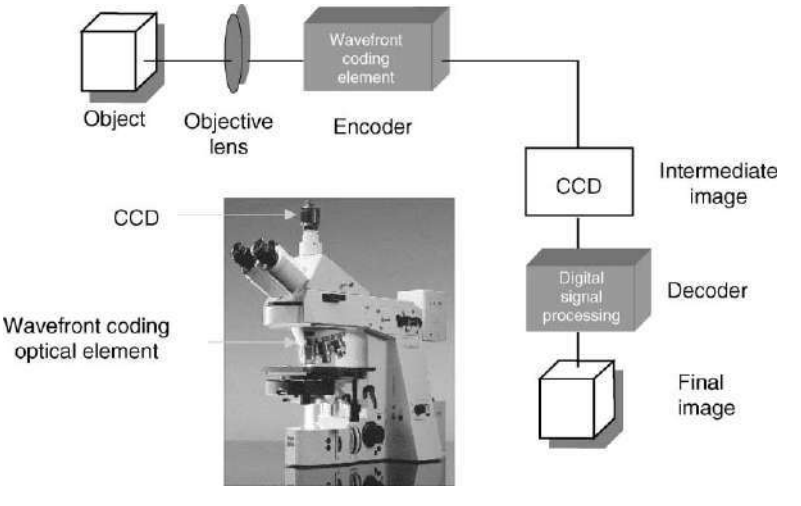

Figure 16 Schematic of a modified Zeiss microscope to employ Wavefront Coding to extend the depth of field.

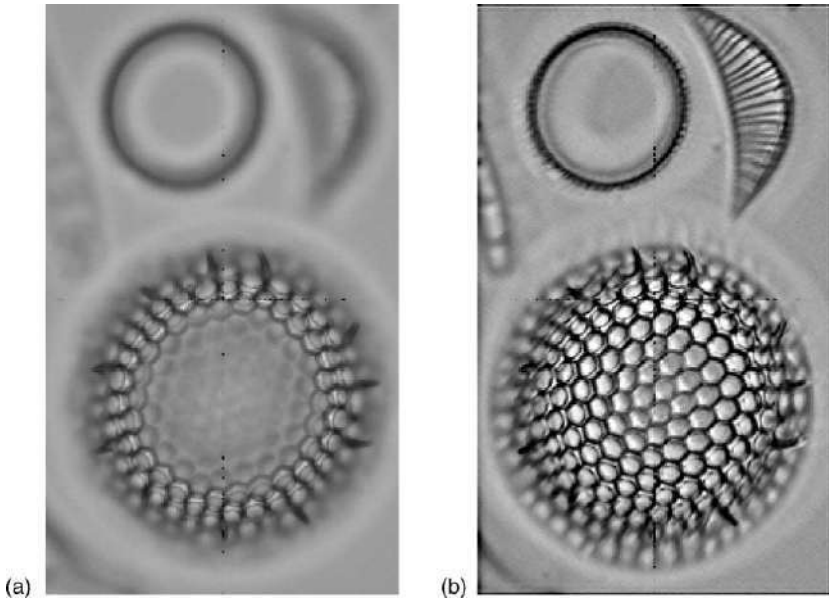

Figure 17 A 100 × image of diatoms made with (a) conventional and (b) modified Zeiss microscope to give an extended depth of field.

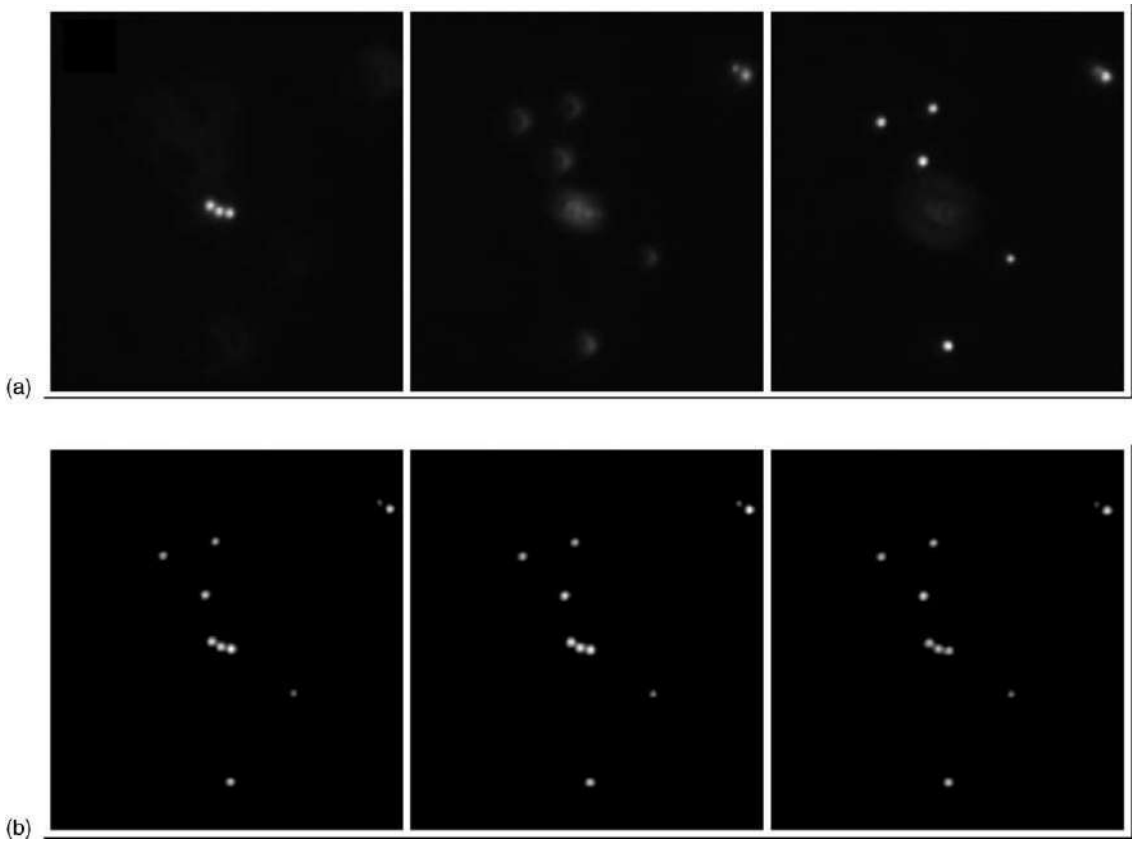

Figure 18 Imaging 0.5 micron diameter fluorescent beads at different sample distances. The images were taken with a 63×, NA = 1.4 imaging system in one micron steps. (a) Three image planes taken using a traditional system. They show that individual beads are sharp only in a single plane. The images in (b), which were taken with a Wavefront Coding system at the same image planes, show all beads are imaged clearly at each plane.

Other modes of microscopy are also possible. In fluorescence microscopy, ultraviolet light illuminates objects that have fluorescing compounds in them. The images are formed using the fluorescence that is excited by the ultraviolet illumination. Figure 18 shows images of fluorescing beads having a 0.5 micron diameter. The beads are at various depths within the sample. The traditional system, operating at 63 × with an immersion objective having a NA of 1.4, has a depth of field that is too small to image all of them. The images in Figure 18a were taken with the same microscope but with a wavefront coding phase element placed in the slider and signal processing of the captured images. All beads are clearly seen at all three of the focus positions.

Potential Future Applications

Wavefront Coding should be useful in many imaging system where signal processing is to be done anyway; in that case, optimizing the optics and the signal processing to the purpose of the imaging system. Examples where an extended depth of field may be desired are endoscopes, video cameras, and lithography. In designing an imaging system to take advantage of a large depth of focus to produce inexpensive digital cameras, high-quality images can be formed with fewer optical elements and much lighter cameras. In lithography, a larger process window is possible.

Because an increase in the depth of focus can permit wide spectral imaging, wide band infrared images can be formed. In systems that are used to provide images for machine interpretation, it is possible to design the optics and signal processing so that the desired process can be optimized globally, with the optics and the signal processing. For example, if the object is to classify cells as cancerous or noncancerous, it may be that the optics should be modified so that the information of primary importance is acquired and transmitted to the detector array. This modification would be done in conjunction with the optimization of the signal processing.

Conclusion

The mathematics and concepts that underlie a new approach to lens design were summarized. The technique has thus far been useful in enhancing the performance of some imaging systems such as microscopes, and in making simpler systems for inexpensive cameras. This new approach requires that some signal processing be done, but in digital cameras, some signal processing is being done already.

List of Units and Nomenclature

Modulation transfer function	The magnitude of the optical transfer function of an imaging system.
Optical transfer function	The response of an imaging system as a function of the spatial frequency of a sinusoidal object distribution.
Spatial frequency	A measure of spatial variation in the form of sinusoids in cycles per millimeter.

See also

Imaging: Wavefront Sensors and Control (Imaging Through Turbulence). **Microscopy:** Imaging Multiple Photon Fluorescence Microscopy; Overview.

Further Reading

Barrett HH and Swindell W (1981) *Radiological Imaging: The Theory of Image Formation, Detection, and Processing*. New York: Academic Press.

Brenner K, Lohmann A and Ojeda-Castañeda J (1983) The ambiguity function as a polar display of the OTF. *Optics Communications* 44: 323–326.

Cathey WT and Dowski ER (2002) New paradigm for imaging systems. *Applied Optics* 41: 6080–6092.

Cathey WT, Frieden BR, Rhodes WT and Rushforth CK (1984) Image gathering and processing for enhanced resolution. *Journal of the Optical Society of America A* 1: 241–249.

Cushman SC (2001) *Finite-Dimensional Matrix Treatment of Partially Coherent Optical Systems*. University of Colorado thesis, Boulder, Colorado.

Dowski ER and Cathey WT (1995) Extended depth of field through wavefront coding. *Applied Optics* 34: 1859–1866.

FitzGerrell AR (1993) *Applications of the Ambiguity Function to Optics*. Thesis, University of Colorado, Boulder, Colorado.

Häusler G (1972) A method to increase the depth of focus by two step image processing. *Optics Communications* 6: 38–42.

Ojeda-Castañeda J, Tepichin E and Diaz A (1989) Arbitrary high focal depth with quasioptimum real and positive transmittance apodizer. *Applied Optics* 28: 2666–2670.

Papoulis A (1974) Ambiguity function in Fourier optics. *Journal of the Optical Society of America* 64: 779–788.

Rihaczek AW (1969) *Principles of High Resolution Radar*. New York: McGraw-Hill.

Van der Gracht J, Dowski ER, Taylor MG and Deaver DM (1996) Broadband behavior of an optical-digital focus-invariant system. *Optics Letters* 21: 919–921.

Wach HB, Dowski ER and Cathey WT (1998) Control of chromatic focal shift through wave-front coding. *Applied Optics* 37: 5359–5367.

Woodward PM (1953) *Probability and Information Theory with Applications to Radar*. New York: Pergamon Press.

PHOTON PICTURE OF LIGHT

S J Bentley, Adelphi University, Garden City, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The photon nature of light is the central topic in the field of quantum optics, which has developed into a leading area of research. The possible technological impacts of this field to many areas, such as imaging, computing, and lithography, have greatly expanded interest into the nature of the photon. A photon description of light goes well beyond the particle nature of light, to include any optical phenomena that cannot be adequately described by classical physical optics. This article will introduce the nature of the photon through a discussion of many such phenomena, covering topics from the beginning of quantum mechanics, including black-body radiation and Compton scattering, to areas of modern interest, such as squeezed light and entangled photons.

The Origins of Quantum Optics

One of the earliest theories in the development of quantum physics was the description of black-body radiation by Max Planck (1858–1947). Using the model of a radiating body as a collection of oscillators, he postulated that the amount of energy emitted or absorbed was proportional to its frequency, ν, and that these oscillators could only have energies corresponding to integer multiples of such a packet of energy, where one unit of energy was given by

$$E = h\nu \qquad [1]$$

where $h = 6.626 \times 10^{-34}$ Js is the universal constant, now known as Planck's constant. He then assumed a Maxwell–Boltzmann distribution to describe the probability of such oscillator states being occupied:

$$P(n) = P_0 e^{-nh\nu/kT} \qquad [2]$$

where $k = 1.381 \times 10^{-23}\ \mathrm{JK}^{-1}$ is Boltzmann's constant and T is the temperature of the blackbody. The energy density in the radiation field, at a given frequency, is given by the oscillator density (at that frequency) times the average energy of the oscillators. Previous models had not only failed to quantize the oscillator energies, but in doing so also assumed equal occupation probabilities that led to a prediction of infinite energy densities at short wavelengths, the so-called ultraviolet catastrophe. The oscillator density is given by

$$n(\nu) = 8\pi\nu^2/c^3 \qquad [3]$$

where $c = 3 \times 10^8\ \mathrm{m\ s^{-1}}$ is the speed of light. This value can be found by a mode density calculation within a box, as is often done in an introductory electromagnetics text. The average energy of the oscillators is found by calculating the sum of the occupation probability times the state energy and dividing by the sum of the occupation probabilities:

$$E_{\mathrm{av}} = \sum_{n=0}^{\infty} [P(n)h\nu n] \Big/ \sum_{n=0}^{\infty} P(n) \qquad [4]$$

Evaluating the above expression for E_{av} and multiplying by $n(\nu)$, one finds the Planck radiation law:

$$I(\nu) = \frac{8h\pi\nu^3}{c^3(e^{h\nu/kT} - 1)} \qquad [5]$$

The black-body curve described by Planck's law is shown in Figure 1. By introducing quantization to solve a problem that classical physics had failed to explain, Planck paved the way for quantum physics.

Albert Einstein (1879–1955) made use of Planck's quantization for two important results that will be described here. The first is the photoelectric effect, a process by which electrons are released from a metal surface when exposed to light. The process has several traits that could not be explained by

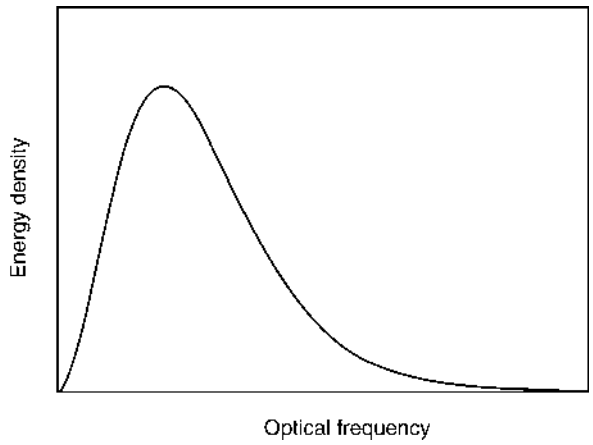

Figure 1 The spectral density emitted by a blackbody at a given temperature as given by Planck's law.

classical physics. A given metal will only release electrons if the incident light is above a certain frequency, regardless of the intensity. However, once sufficient frequency components are present, electrons are immediately released, with their number dependent on the intensity of the light and independent of its frequency. Contrarily, the maximum potential energy of the released electrons depends only on the maximum frequency of light present, independent of intensity. Einstein was able to satisfy all of these conditions by saying that the light energy comes in quantized amounts as described by Planck, and that the electrons in the metal require a minimum energy, or work function W, to be ejected. Then for each quantum of light, which we now call a photon, with energy greater than W that strikes an electron, the electron will be released with kinetic energy given by

$$K = h\nu - W \qquad [6]$$

It is historically important to note that, although Planck introduced the idea of a quantized release of radiation from oscillators, he still believed the light field itself to be continuous and classical. Planck's approach of quantizing energy levels of matter while treating light classically has since become known as the semiclassical method, and is adequate for solving many problems. Thus, it was also in quantizing light itself that Einstein took the next step, and thus can be viewed in many respects as the father of the modern photon, although the term photon was not used until two decades later.

The second area in which Einstein used the idea of energy quantization is in predicting radiative transition probabilities. A radiative transition is a process by which a photon is either absorbed or emitted to satisfy conservation of energy when an atom moves to a higher or lower energy state. Einstein assumed three such processes existed, as shown in Figure 2: absorption of a photon raising the atom to a higher energy level; spontaneous emission of a photon, where the atom emits a photon in going to a lower energy level; and stimulated emission, a process in which a photon interacts with the atom and causes it to emit a second photon (in phase with the first) and thereby go to a lower energy level. For each of these processes he defined a corresponding coefficient, now known as the Einstein A and B coefficients. They are labeled B_{12} for absorption, A_{21} for spontaneous emission, and B_{21} for stimulated emission. Note that there is no A_{12} as absorption by its very nature is a stimulated process. Einstein further stated that the rate of spontaneous emission was only dependent on his material dependent coefficient and the number of

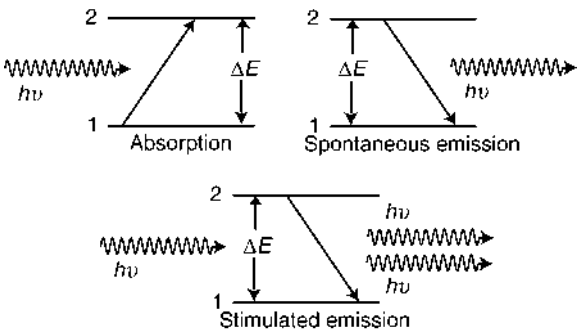

Figure 2 Energy level diagrams representing the processes of absorption, spontaneous emission, and stimulated emission.

atoms in the upper energy state, with the rate given by

$$R_{sp} = A_{21}N_2$$

However, he said that the rates of the stimulated processes were also dependent upon the strength of the optical field seen by the atoms, $I(\nu)$, and were given by

$$R_{ab} = I(\nu)B_{12}N_1 \qquad [7]$$

for absorption and

$$R_{st} = I(\nu)B_{21}N_2 \qquad [8]$$

for stimulated emission. The full absorption rate is given by R_{ab} as defined above, whereas the full emission rate is the sum of the spontaneous and stimulated rates, given by

$$R_{em} = [A_{21} + B_{21}I(\nu)]N_2 \qquad [9]$$

Einstein was primarily interested in a system in equilibrium, for which $R_{ab} = R_{em}$. If one then looks at the limit where the optical field is very large such that spontaneous emission is negligibly small and the population is equally divided between the upper and lower state, the relationship $B_{12} = B_{21}$ is found. Applying the same Maxwell–Boltzmann occupation statistics that Planck used to find his black-body expression, one can find that in thermal equilibrium the ratio of probability of finding an atom in state 1 to finding an atom in state 2 is

$$N_1/N_2 = e^{(E_2-E_1)/kT} = e^{h\nu/kT} \qquad [10]$$

Note that for a two-level system in thermal equilibrium, the population in the upper level can at most equal, never exceed, the population in the lower level, regardless of the strength of the applied field. This is important in showing that the B coefficients are equal, and will also be a crucial point when discussing the theory of the laser later in this article.

Appropriately combining the above expressions leads to

$$I(\nu) = A_{21}/[B_{21}(e^{h\nu/kT} - 1)] \qquad [11]$$

Equating this with Planck's radiation law yields the relation:

$$A_{21}/B_{21} = 8\pi h\nu^3/c^3 \qquad [12]$$

Thus, to fully specify the transition probabilities for a given system, one needs only to find A or B. It is rather straightforward using only atomic quantization to calculate B (such a derivation can be found in many introductory quantum mechanics textbooks). Direct calculation of A is more involved, requiring quantization of both the optical field as well as the atomic system.

Arthur Compton (1892–1962) used the idea of quantized radiation to explain the presence of longer wavelength components present when X-rays are scattered from matter, an effect which now bears his name. The amount of scattering is independent of the scattering material and the wavelength of the source, depending only on the angle of scattering, θ. If the scattering is treated as a collision between a particle-like photon and a massive particle (taken to be initially at rest) as shown in Figure 3, the result can then be found simply by applying conservation of energy and momentum, where the photon energy is as given above and the photon momentum is given by

$$p = E/c = h/\lambda \qquad [13]$$

Recalling that the momentum is a vector quantity, the conservation laws provide a set of three coupled equations. Solving them directly one finds that

$$\Delta\lambda = (h/c)(1 - \cos\theta)(2K)/[p^2 - (K/c)^2] \qquad [14]$$

where K and p correspond to the final kinetic energy and momentum of the massive particle. It is important to allow for the possibility of very high velocities for the massive particle, thus implying that relativistic expressions for K and p are needed. In terms of the kinetic energy, the relativistic expression for momentum is

$$p = \sqrt{2Km + (K/c)^2} \qquad [15]$$

where m is the mass of the scattering particle. Using this relation, the above expression can be simplified, giving the wavelength shift to be

$$\Delta\lambda = (1 - \cos\theta)h/mc \qquad [16]$$

The scaling factor is known as the Compton wavelength:

$$\lambda_c = h/mc \qquad [17]$$

For the case of an electron, the Compton wavelength has the value 0.02426 angstroms. For a massive particle, the Compton wavelength is often used when wanting to discuss and analyze wavelike properties. As a final note on this process, if the calculation would have been done nonrelativistically, there would have been an extra factor of $[1 - 0.25(\nu/c)^2]^{-1}$, where ν is the velocity of the particle. The appearance of the factor (ν/c) in any result implies the need for a relativistic treatment, though often (as in this case) taking the limit of $\nu/c = 0$ yields the proper result.

Another area in which the photon nature of light is revealed is a form of inelastic scattering of light now known as the spontaneous Raman effect. Discovered by CV Raman (1888–1970), this is a process in which a material illuminated by a light source scatters light into frequency components not found in the source. The frequency shift is due to the energy difference, ΔE, between two energy states in the material. There are two possible cases, each of which is represented in Figure 4. In the first, known as Stokes scattering, an atom in the lower state absorbs a source photon, raising it to the upper state via a virtual level, during which it emits a photon at the Stokes frequency, ν_s, as represented by the relation

$$h\nu_i = h\nu_s + \Delta E \qquad [18]$$

where ν_i is the frequency of the incident photon from the source. The second case, anti-Stokes scattering, has an atom in the upper state absorb the source photon and simultaneously fall to the lower state, again through the virtual level during which it emits a photon at the higher anti-Stokes frequency, ν_a,

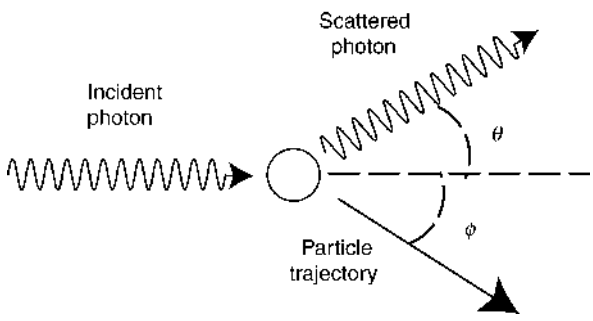

Figure 3 Schematic representation of Compton scattering.

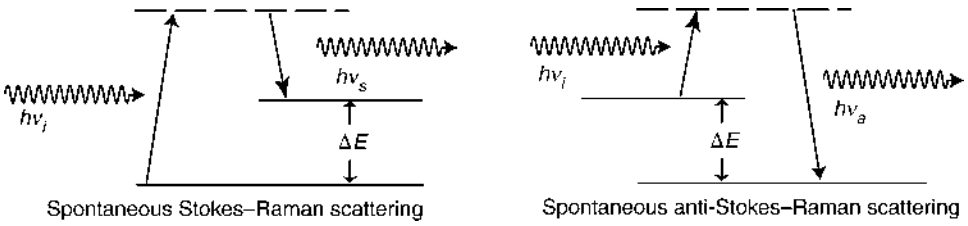

Figure 4 Energy level diagrams for Stokes and anti-Stokes Raman scattering.

given by

$$h\nu_i = h\nu_a - \Delta E \qquad [19]$$

Since the anti-Stokes scattering requires the atom to be initially in the upper state, the probability of such scattering is generally much less than that for Stokes scattering. This can be quantified by the same occupation statistics used above, giving the ratio of probabilities as

$$P_{\text{as}}/P_s = e^{-h\Delta E/kT} \qquad [20]$$

It is important to note that the energy difference can either be based on the electronic structure of an atomic system as has been discussed thus far, or other energies such as the vibrational energies in a molecular system. The ability to probe a variety of energy structures without having to tune the source to a particular wavelength has led to wide use of Raman scattering for spectroscopy. A related effect, known as stimulated Raman scattering, will be discussed later.

The Photon in Modern Physics

As an indication of the widespread importance of the results discussed thus far, they led to four Nobel prizes in physics: Planck in 1918; Einstein in 1921; Compton in 1927; and Raman in 1930. Our initial understanding of the photon comes from these results, but it was with the development of modern quantum mechanics by Erwin Schrödinger (1887–1961), Werner Heisenberg (1901–1976), and Paul Dirac (1902–1984); quantum electodynamics by Richard Feynman (1918–1988) and others; and the invention of the laser by Charles Townes (1915–) and others, that the field of quantum optics and our modern view of the photon began to fully take shape. In this section, the basics of modern quantum optics will first be introduced. A brief discussion of the theory of the laser as related to photon transitions will then be given, followed finally by examples of important areas of experimental quantum optics that have been made accessible by the laser.

It is now common to treat the photon with a very similar formalism as that originally created for massive particles in the development of quantum mechanics. That is, one describes the photon by a wave function, $|\Psi\rangle$, which contains all the information about the photon and must satisfy the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\Psi\rangle = H|\Psi\rangle \qquad [21]$$

where $\hbar = h/2\pi = 1.054\times10^{-34}$ Js and H is the Hamiltonian for the photon. Just as in classical electromagnetics, the Hamiltonian is related to an electric and magnetic field. Clearly, for this treatment to be in agreement with the well-established classical electromagnetic theory, these fields must satisfy Maxwell's equations. There are several possible representations of the Hamiltonian, but for the purposes of this discussion a convenient choice is

$$H = h\nu(\hat{n} + 0.5) \qquad [22]$$

where $\hat{n}$ is called the number operator as it returns the average number of photons in the state, $\langle\hat{n}\rangle$, represented by the operation

$$\langle\hat{n}\rangle \equiv \langle\Psi|\hat{n}|\Psi\rangle \qquad [23]$$

Explanations of the notation used can be found in any elementary quantum mechanics text. Note that in defining the Hamiltonian, a single frequency ν was used. This implies a single mode field. For a more general case, one would simply sum over the modes corresponding to various frequencies.

In addition to obeying the Schrödinger equation, the photon is also restricted by the Heisenberg uncertainties as for massive matter, given in a general form as

$$\Delta q\Delta p \geq \hbar \qquad [24]$$

where q and p are any set of canonically conjugate variables, such as position and momentum or energy and time. Note that, depending on the exact definitions used for q and p, there can be extra factors such as 0.5, and sometimes normalized

variables are used such that the inequality is with respect to 1. However, it is interesting to note that, unlike for massive particles that were considered well localized in classical theory, the idea of uncertainty for light had been accepted in some form long before the development of quantum mechanics. Neglecting the scaling factor $\hbar$, which is purely quantum in origin, an uncertainty principle can be derived for any wavelike phenomena simply using the relationship between Fourier transform pairs.

Using the basic ideas of a photon state and the uncertainty principle, some important properties of light will now be explored. There are several possible forms the wave function can take. It is often convenient to write the wave function in terms of states containing a certain number of photons. Such states are known as Fock or number states and are represented as $|n\rangle$, with n being the number of photons in the state and are said to be eigenstates of the number operator. The energy of a Fock state is given by

$$E_n = h\nu(n + 0.5) \qquad [25]$$

Notice that this predicts a background, or vacuum, energy for the electromagnetic field, even when the average number of photons in the field is zero. This vacuum energy is given by $0.5\,h\nu$. Correspondingly, due to the uncertainty principle, there must be some uncertainty in this energy as well as with any energy. The fluctuations in the vacuum energy are known as vacuum fluctuations. Many processes, such as spontaneous emission that cannot be adequately described by a simple semiclassical approach, where matter is treated quantum mechanically and light is treated classically, can be treated by the addition of vacuum fluctuations, without requiring a full quantum mechanical treatment of the electromagnetic field. Such an approach is used in many texts, but there are several important physical phenomena that do require the fully quantized approach discussed here.

The state most resembling a classical field is known as the coherent state, represented here as a sum of Fock states:

$$|\Psi\rangle = |\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \qquad [26]$$

where α is a parameter describing the field, with the average number of photons in the field given by

$$\langle \hat{n} \rangle = |\alpha|^2 \qquad [27]$$

One key property of coherent states is that they exhibit minimal uncertainty. That is, the uncertainty relations become equalities. Also, the uncertainty is equally divided between quadrature components.

Before applying the above principles to explore some of the dramatic results of quantum optics, a brief discussion of the laser will be given due to its importance in the field. The theory of operation of the laser is based directly on Einstein's transition probabilities. For the transition of interest, one wants to generate a large number of coherent photons (that is, photons in phase with one another). Recall that in the process of stimulated emission, one photon leads to two coherent photons, whereas spontaneous emission leads to photons which are not phase referenced to anything else in the system. Also, if a significant number of the photons are to escape the region where they are generated, it is important to minimize the absorption. Thus, a laser requires that the stimulated emission rate exceed both the spontaneous emission rate and the absorption rate. The first condition places a lower limit on the optical field density present at the lasing frequency. This density can be reasonably large, but presents no fundamental difficulties. In practice, mirrors are placed around the emission region to confine the photons and counteract losses, thereby building up the energy density until the above condition is satisfied. The second condition, however, can be written as

$$N_2 > N_1 \qquad [28]$$

and is the more restrictive requirement. Such a condition is referred to as inversion and, as was previously mentioned, cannot be sustained in a two-level system once equilibrium is reached. Therefore, to be able to maintain lasing, additional levels are needed. Two typical laser energy diagrams are shown in **Figure 5**. In such multilevel systems, it is possible to create an inversion in the levels of interest and thus allow lasing. Further details on lasers (both in theory and practice) are widely available, though it should be noted that the output of a laser is a good approximation of a coherent state described above.

The invention of the laser revolutionized optics, as well as the whole of modern physics, by making readily available electromagnetic fields of a vastly different nature than that from any previous source of light. This quickly led to the birth of a new field known as nonlinear optics. Nonlinear optics describes interactions between light and matter which have properties dependent on the field strength. Although some nonlinear processes can be described adequately without the use of a photon picture, most nonlinear process can be best described as some combination of annihilation (absorption) and creation (emission) of photons. Two such

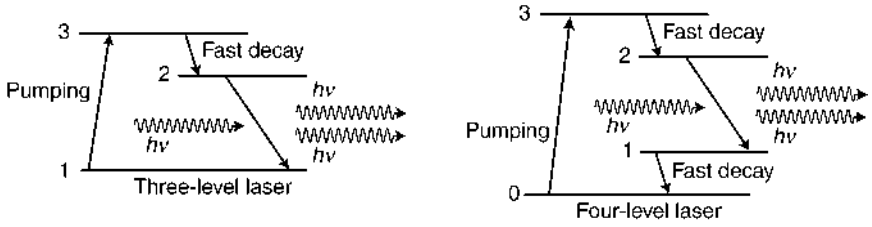

Figure 5 Energy level diagrams for three-level and four-level lasers.

processes, that are of great interest in quantum optics, are stimulated scattering and parametric down conversion.

Earlier, the process of spontaneous Raman scattering was discussed. For such a process, it is assumed that the incident field driving the process is relatively weak such that the scattering process generates very low photon densities in the various modes of the electromagnetic field. However, just as there are both stimulated and spontaneous emission processes in the radiative transitions discussed previously, in the presence of strong fields there can also be stimulated scattering. The total Raman scattering rate can be expressed as

$$R = \eta I_i (I_{\mathrm{R}} + 1) \qquad [29]$$

where η is a scaling factor, including the gain of the process for a given material, I_i is the strength of the incident optical field (generally a laser field to be sufficiently strong to cause stimulated scattering), and I_{R} is the Raman field strength. Note that for I_{R}, the rate is just linear with the applied field, representing the spontaneous process. However, with a strong pump, the generated field will quickly become large enough such that the stimulated process will dominate. The spontaneous scattering is not directional dependent, but since the stimulated process grows along the length of the scattering material in the path of the incident beam, it will tend to be generated along the path of the incident laser. Due to linear and nonlinear dispersion effects, however, the light will form in a narrow cone about the pumping laser rather than exactly collinear. Unlike the spontaneous scattering, a significant percentage of the incident photons can be converted to the Raman frequencies, making stimulated Raman scattering an extremely powerful spectroscopic tool.

Possibly the most important nonlinear optical process for quantum optics is parametric down conversion. This process is interesting both in the nature of the process itself, as well in the properties of the generated light. In parametric down conversion, shown schematically in Figure 6, one photon (known as a pump photon) of frequency ν_p is annihilated

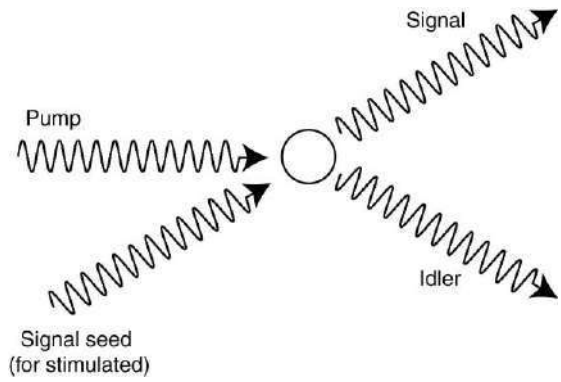

Figure 6 Schematic representation of parametric down conversion.

and two new photons (generally referred to as signal and idler) are created with frequencies ν_s and ν_i respectively, given by the conservation of energy relationship:

$$h\nu_p = h\nu_s + h\nu_i \qquad [30]$$

In the case where $\nu_s = \nu_i$, the process is referred to as degenerate. Down conversion can either be spontaneous, in which only pump photons are externally applied, or stimulated, in which both pump photons and signal photons are applied. In the stimulated case, the signal is amplified and idler photons are generated as a byproduct. Parametric down conversion is one of the leading sources of nonclassical states of light. The term nonclassical is used rather than quantum, because all states of light can be accurately described by a quantum picture, but while most can also be described very well with a classical picture the states referred to as nonclassical cannot. Two of the most common nonclassical states of light are entangled states and squeezed states.

Entangled states are multiphoton states in which the state cannot be written as a product of the states of the individual photons. For simplicity, only two-photon entangled states (commonly referred to as EPR states for Einstein, Podolsky, and Rosen) will be discussed. More extensive and general material, including discussions of higher-order entanglement,

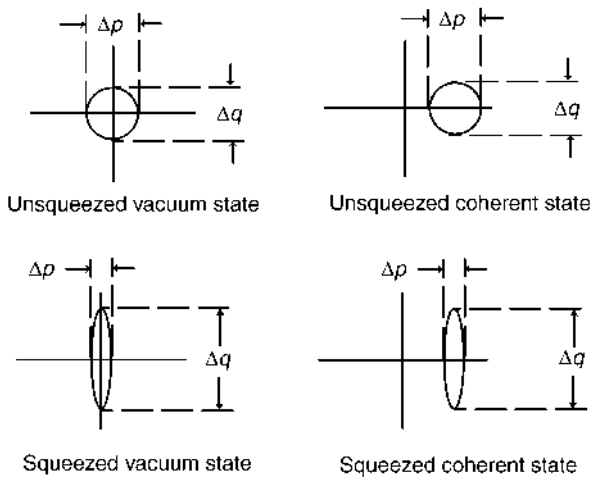

Figure 7 Uncertainty diagrams for squeezed and unsqueezed vacuum and coherent states.

such as the three-photon entangled states (commonly referred to as GHZ states for Greenberger, Horne, and Zeilinger), can be found in a variety of textbooks on modern quantum optics. One of the simplest ways to generate a two-photon entangled state is by the process of spontaneous parametric down conversion described above. If the product of the pumping intensity, generation length, and generation gain is of an appropriate level, individual pairs of photons are emitted. The states of these photon pairs are entangled, with the position, momentum, wavelength, polarization, and all other properties of the photons completely correlated and inseparable. Such states have been a powerful tool for the experimental verification of many of the predictions of quantum mechanics and quantum optics. Entangled states can also act as a source for many of the newer technologies in quantum optics, including cryptography and teleportation.

Another important form of nonclassical light is known as squeezed light. The principle behind squeezed light is that, unlike for coherent states where the uncertainty is equally distributed, the uncertainty in one quadrature can be reduced at the expense of increased uncertainty in the other quadrature. This is shown schematically in Figure 7. As opposed to entangled states which tend to be very weak, containing only a few photons, squeezed states of light can be quite intense. This makes squeezed light more desirable for many applications.

In conclusion, it should be noted that while the photon description of light can explain a wide variety of physical processes and is now leading to many exciting technological developments, much is still unknown about the photon at the fundamental level. Any accurate description of the photon must be in agreement with experimental evidence, including that discussed in this article, but one should be aware that the accepted picture of the photon nature of light is still constantly evolving and growing.

See also

Nonclassical Light. **Nonlinear Optics, Applications:** Raman Lasers. **Optical Amplifiers:** Raman, Brillouin and Parametric Amplifiers. **Quantum Optics:** Entanglement and Quantum Information. **Scattering:** Raman Scattering. **Spectroscopy:** Raman Spectroscopy.

Further Reading

Boyd RW (2002) *Nonlinear Optics*, 2nd edn. San Diego, CA: Academic Press.

Cohen-Tannoudji C, Diu B and Laloë F (1977) *Quantum Mechanics*. New York: Wiley-Interscience Publications.

Hecht E (2001) *Optics*, 4th edn. Reading, MA: Addison-Wesley Publishing.

Heisenberg W (1949) *The Physical Principles of the Quantum Theory*. New York: Dover Publications.

Lipson SG, Lipson H and Tannhauser DS (1995) *Optical Physics*, 3rd edn. Cambridge, UK: Cambridge University Press.

Loudon R (2000) *The Quantum Theory of Light*, 3rd edn. Oxford, UK: Oxford University Press.

Mandel L and Wolf E (1995) *Optical Coherence and Quantum Optics*. Cambridge, UK: Cambridge University Press.

Milonni PW and Eberly JH (1988) *Lasers*. New York: Wiley-Interscience Publications.

Roychoudhuri C and Roy R (eds) (2003) *OPN Trends – The Nature of Light: What is a Photon?* Washington, DC: Optical Society of America.

Schwinger J (ed.) (1958) *Selected Papers on Quantum Electrodynamics*. New York: Dover Publications.

Scully MO and Zubairy MS (1997) *Quantum Optics*. Cambridge, UK: Cambridge University Press.

Yariv A (1989) *Quantum Electronics*, 3rd edn. New York: John Wiley & Sons.

PHOTONIC CRYSTALS

Contents

Atomic Physics
Electromagnetic Theory
Microwave Photonic Crystals
Nonlinear Optics in Photonic Crystal Fibers
Photonic Crystal Lasers, Cavities and Waveguides
Self-Assembled and Functionalized Photonic Crystals

Atomic Physics

G Kurizki and A G Kofman, Weizmann Institute of Science, Rehovot, Israel

D Petrosyan, Institute of Electronic Structure and Laser, Heraklion, Greece

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The fabrication and study of dielectric structures whose refractive index is periodically modulated on a micron or submicron scale, known as photonic crystals (PCs), are attracting considerable interest at present. One-, two-, or three-dimensional (1D-, 2D-, or 3D-) periodic PCs exhibit photonic band gaps (PBGs), analogously to electronic band gaps in ordinary crystals, and can incorporate defects, designed to form localized narrow-linewidth (high-Q) modes at PBG frequencies. The advent of such structures opens up new perspectives in atomic physics and quantum optics, since they are expected to allow an unprecedented control over the spectral density of modes (DOM) and the spatial modulation of narrow-linewidth (high-Q) modes, in the microwave, infrared, and optical domains. An interesting situation arises when foreign atoms or ions – dopants – with transition frequencies within the PBG, are implanted in the PC. Then light near one of these frequencies resonantly interacts with the dopants and is concurrently affected by the PBG dispersion. Consequently, highly nonlinear processes with a rich variety of unusual PC-related features are anticipated. Such nonlinear optical processes, involving near-resonant transitions in PCs, undergo basic modifications as compared to the corresponding processes in free space, which are attributed to the strong suppression of the DOM within PBGs, to sharp bandedges and to intra-gap narrow lines associated with high-Q defect modes.

Results are detailed below for spontaneous emission of radiation in PCs and for the resonant interaction of atoms with the field of a single high-Q defect mode. These results stem from the failure of perturbation theory and the onset of strong field-atom coupling near sharp bandedges or narrow defect-mode lines in 2D and 3D PCs. 3D-PBGs are needed, in order to extinguish spontaneous emission in all possible directions of propagation and dipole orientations. If only one polarized atomic transition is involved in the spontaneous emission, 2D-PCs suffice for its suppression. For controlling strictly unidirectional field propagation, it is sufficient to resort to PBGs in 1D-periodic structures (Bragg reflectors or dielectric multi-layer mirrors).

As a further example of field–atom interactions in doped PCs, we will show that two ultraweak electromagnetic fields (or photons), coupled to appropriate transitions of the dopants in a PC, can mutually induce large phase shifts or drastic changes in absorption. Appreciable photon–photon correlations can then be established, which can be utilized in both classical and quantum optical communications.

Radiative Decay and Photon–Atom Binding in a PC

The interaction of a two-level atom with an arbitrary field-mode continuum can be described by the following second-quantized Hamiltonian in the

rotating-wave approximation (RWA)

$$H = \hbar\omega_a|e\rangle\langle e| + \hbar\sum_{\vec{k}}\omega_{\vec{k}}a^\dagger_{\vec{k}}a_{\vec{k}} + \hbar\sum_{\vec{k}}[\kappa(\omega_{\vec{k}})a^\dagger_{\vec{k}}|g\rangle\langle e| + \text{hermitian constant}] \quad [1]$$

Here $|e\rangle$ and $|g\rangle$ are the excited and ground atomic states, respectively, ω_a is the atomic transition frequency, $a^\dagger_{\vec{k}}$ and $a_{\vec{k}}$ are the creation and annihilation operators of a field mode with wavevector $\vec{k}$, and frequency $\omega_{\vec{k}}$, and $\hbar\kappa(\omega_{\vec{k}})$ is the resonant coupling energy of this mode with the atomic dipole. The $\vec{k}$-summation can be replaced by an integral over the continuum DOM in the frequency domain, $\sum_{\vec{k}} \rightarrow \int_0^\infty d\omega\rho(\omega)$, avoiding, for the sake of simplicity, the study of direction-(angle-)dependent DOM effects.

We consider a two-level atom that is excited at time $t = 0$ in an empty, periodic dielectric structure. The wavefunction of the combined system field + atom can be cast in the general form:

$$|\Psi(t)\rangle = \alpha(t)|e, \{0_\omega\}\rangle + \int \beta_\omega(t)|g, 1_\omega\rangle\rho(\omega)d\omega \quad [2]$$

where $\{0_\omega\}$ signifies the completeness of field modes in the vacuum state and $|1_\omega\rangle$ denotes single-photon occupation of the ω-mode. The corresponding Schrödinger equation can be solved with the initial condition $|\Psi(0)\rangle = |e, \{0_\omega\}\rangle$ by means of the continuum spectral response:

$$G(\omega) = |\kappa(\omega)|^2\rho(\omega) \quad [3]$$

The analysis of eqn [2], with $G(\omega)$ appropriate for photonic bandstructures, is aimed at revealing the prominent features of the atomic excitation decay $\alpha(t)$ and the corresponding emission spectrum.

In what follows, we consider a photonic bandstructure with several PBGs, separated by allowed bands. Each PBG is labeled by index i and has lower and upper cutoff frequencies ω_{Li} and ω_{Ui}, respectively. Then $G(\omega) = 0$ for $\omega_{Li} < \omega < \omega_{Ui}$. Naively, one might expect that an excited atom would either be stable against single-photon decay, if ω_a is within a PBG, or decay completely at $t \rightarrow \infty$, if ω_a is anywhere in an allowed band. However, both statements turn out to be inaccurate.

Incomplete decay of $\alpha(t \rightarrow \infty)$ occurs if there is a stable eigenvalue (energy level) $\hbar\omega_i$ of the total (field-atom) Hamiltonian eqn [1]. Such an energy level is possible only if $G(\omega_i) = 0$, i.e., for ω_i in a PBG. It must satisfy:

$$\omega_i = \omega_a + \Delta(\omega_i) \quad [4]$$

$$\Delta(\omega_i) = \int_0^{\omega_{Li}} \frac{G(\omega')}{\omega_i - \omega'}d\omega' + \int_{\omega_{Li}}^\infty \frac{G(\omega')}{\omega_i - \omega'}d\omega' \quad [5]$$

Here the integrals are the frequency shifts of the atomic resonance $\hbar\omega_a$ induced by the spectral parts of the reservoir situated, respectively, below and above the ith PBG (Figure 1). If eqn [4] holds, then the corresponding term in $\alpha(t)$ is proportional to $\exp(-i\omega_i t)$, and does not decay. Physically, such a stable level can be interpreted as representing the binding of the photon to the atom (photon-dressed atom), without the ability to leave the atomic vicinity, due to Bragg reflection in the PBG.

Assuming that ω_i occurs in the ith PBG, we observe that the first shift is positive whereas the second shift is negative, i.e., each part of the continuum repels the atomic level from its PBG edge. Equation [4] has a solution if ω_a falls between the minimum and maximum values assumed by its left-hand side in the ith PBG, i.e., if

$$\omega_{Li} - \Delta(\omega_{Li}) < \omega_a < \omega_{Ui} - \Delta(\omega_{Ui}) \quad [6]$$

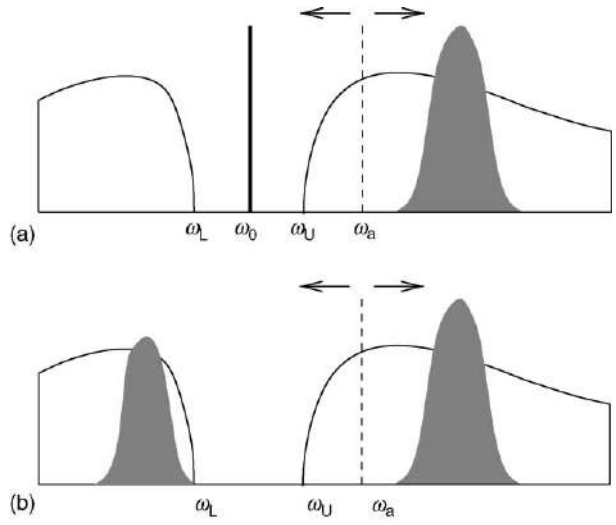

Figure 1 Frequency shifts of the atomic resonance ω_a (dashed line) due to 'repulsion' (arrows) by asymmetric spectral parts of the continuum below and above the PBG: (a) Incomplete decay for ω_a in allowed band. The interaction with the continuum splits the state into a superposition of a stable part in the PBG (solid line) and a decaying part above ω_a (shaded peak). (b) Complete decay for same ω_a, due to larger shifts that split the state into decaying parts in two allowed bands.

Incomplete decay occurs if this condition holds at least for one PBG. There may be, at most, one stable state in a single PBG (Figure 1). However, a two-level atom in a periodic structure can have several stable dressed states simultaneously, when conditions hold for several PBGs. It is quite extraordinary that the conditions (eqn [6]) for incomplete decay can be fulfilled with ω_a, not only in a PBG, but even in an allowed band, e.g., when $\Delta(\omega_{Ui})$ is negative or when $\Delta(\omega_{Li})$ is positive (Figure 1a,b).

Equally counterintuitive is the converse possibility, of complete decay for ω_a within a PBG. When there is a single PBG and one of the inequalities in eqn [6] is violated, complete decay will occur for $\omega_{Li} < \omega_a < \omega_{Li} - \Delta(\omega_{Li})$, if $\Delta(\omega_{Li}) < 0$, or $\omega_{Ui} - \Delta(\omega_{Ui}) < \omega_a < \omega_{Ui}$, if $\Delta(\omega_{Ui}) > 0$.

The possibility that one or several discrete, stable, states exist yields the corresponding wavefunction in the form:

$$|\Psi(t)\rangle = \sum_i \sqrt{c_i}|\psi_i\rangle e^{-i\omega_i t} + |\Psi_c(t)\rangle \qquad [7]$$

Here the summation is over all discrete atomic photon-dressed states with energies $\hbar\omega_i$, $|\psi_i\rangle$ is a discrete-(dressed-)state eigenfunction of the Hamiltonian (eqn [1]) normalized to unity, and weighted by the amplitude $\sqrt{c_i} = [1 + \int_0^\infty d\omega G(\omega)/(\omega - \omega_i)^2]^{-1/2}$. The dressed state $|\psi_i\rangle$ consists of an excited-state component and a photon-bound ground-state component.

The population of the excited state for long times:

$$|\alpha(t)|^2 = \sum_{i,i'} c_i c_{i'} \cos(\omega_i - \omega_{i'})t \quad \text{for} \quad t \to \infty \qquad [8]$$

is a nonzero time-constant if there is one discrete stable state, whereas in the case of several such states, it undergoes beats at the frequencies corresponding to their energy differences. The splitting of an excited state $|e\rangle$ into superposed stable states, oscillating at bandgap frequencies ω_i, whose amplitudes c_i, and eigenfrequencies ω_i are controllable by the atomic transition detuning from cutoff, constitutes spontaneous coherence control.

If, however, ω_a is far from the cutoff, then eqn [2] results in an exponentially decaying amplitude:

$$\alpha(t) \approx e^{-(\gamma_a + i\tilde{\omega}_a)t} \qquad [9]$$

where $\tilde{\omega}_a = \omega_a + \Delta_a$, Δ_a, and γ_a being the effective spectral shift and width of the decaying atom. This regime holds for a locally smooth $G(\omega \approx \omega_a)$ such that

$$|\gamma_a'|, |\Delta_a'| \ll 1; \; |\gamma_a'\Delta_a| \ll \gamma_a; \; \gamma_a, |\Delta_a| \ll |\omega_a - \omega_g| \qquad [10]$$

where ω_g is the bandedge frequency nearest to ω_a and the primes denote differentiation with respect to ω_a.

We now apply the foregoing general results to a model DOM distribution. This distribution is derived on keeping the lowest term in the Taylor expansion of the dispersion relation $\omega(\vec{k})$ near a photonic bandedge ω_U (the effects of the further PBG edge are neglected). This yields the effective mass approximation:

$$\omega \approx \omega_U + \sum_{i=x,y,z} (k - k_U)_i^2/m_i \qquad [11]$$

with $1/m_i = (1/2)(\partial^2\omega/\partial k_i^2)|_{\omega=\omega_U}$. In a structure with period L, k_U satisfies the Bragg condition $k_U = \pi/L$. The corresponding DOM in a 3D-periodic structure with an allowed symmetry may be approximated as

$$\rho(\omega) \sim (\omega - \omega_U)^{(1-D)/2}\theta(\omega - \omega_U) \qquad [12]$$

where θ is the step function and D is the dimension of the Brillouin-zone surface spanned by bandedge modes with vanishing group velocity. In realistic photonic structures $D \le 2$, $D = 2$ corresponding to completely isotropic dispersion (spherical Brillouin-zone surface) and $D = 0$ corresponding to an anisotropic three-dimensional Brillouin zone. Both cases can be represented, respectively, as the limits $\varepsilon \to 0$ and $\varepsilon \to \infty$ of the function:

$$G(\omega) = \frac{C}{\pi}\frac{\sqrt{\omega - \omega_U}}{\omega - \omega_U + \varepsilon}\theta(\omega - \omega_U) \qquad [13]$$

where ε is the 'cutoff-smoothing' parameter and C is the continuum coupling constant. Depending on whether $C^{2/3}/\varepsilon$ is greater or less than one, the continuum is close to the case $\varepsilon = 0$ ($D = 2$) or $\varepsilon \to \infty$ ($D = 0$), respectively.

Using the properties of the model DOM (eqn [13]), we can infer the criteria for the two regimes discussed above:

(i) The conditions for incomplete decay, eqn [6], are now (Figure 2) $\Delta_c \equiv \omega_a - \omega_U < C\varepsilon^{-1/2}$. Hence, the abrupt, singular cutoff of the DOM with $D = 2$ ($\varepsilon \to 0$) implies the existence of a discrete state for any ω_a, either inside or outside the PBG. The energy of the discrete state, $\hbar\omega_0$, which must lie in the PBG, is found to be a real and positive root of the equation $\omega_a - \omega_0 = C/(\sqrt{\omega_U - \omega_0} + \sqrt{\varepsilon})$.

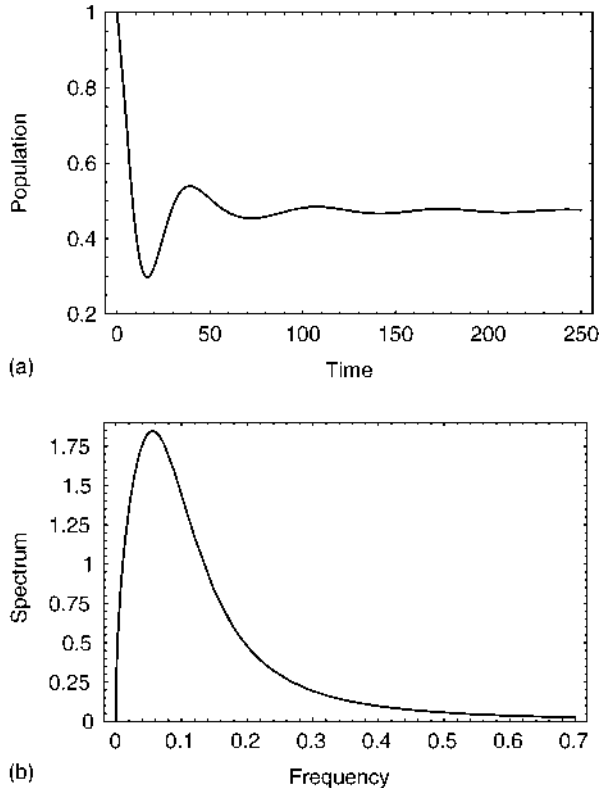

Figure 2 (a) Incomplete decay of the excited state population as a function of $\gamma_c t$, where $\gamma_c = C\varepsilon^{-1/2}$, for $\varepsilon = 10^{-3}$, at cutoff $\Delta_c = 0$. The beat period is $2\pi/(\omega_U - \omega_0)$ and the nondecaying probability is $c_0^2 = 4/9$. (b) The corresponding spectrum. The frequency is normalized to γ_c.

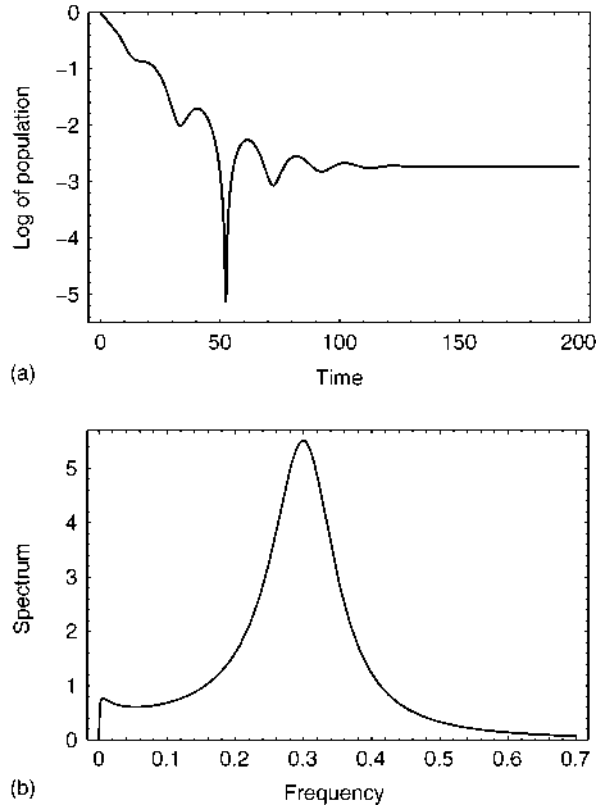

Figure 3 As in Figure 2, for a large detuning from the sharp cutoff, $\varepsilon = 10^{-3}$, $\Delta_c = 0.3$. (a) $\log_{10}|\alpha|t|^2$ is plotted as a function of i. The nearly complete decay ($c_0^2 \sim 10^{-3}$) is modulated by beats with frequency Δ_c. A power tail obtains for $t \gg \gamma_a^{-1} \approx \Delta_c^{1/2}/C$, decreasing as $t^{-1/2}$ for $t \ll \Delta_c^2/C^2$ and as $t^{-3/2}$ for $t \gg \Delta_c^2/C^2$. (b) The nearly Lorentzian spectrum, $w(\omega)$, has a small peak near ω_U, due to the sharply peaked DOM near the cutoff.

(ii) The conditions for nearly exponential decay, eqn [10], can be shown to reduce now to the requirement that ω_a be in the allowed zone sufficiently far from cutoff, $\Delta_c \gg \min\{C^{2/3}, C\varepsilon^{-1/2}\}$. Under this condition, the intermediate-time exponential decay of the excited state (Figure 3) is given to first approximation by eqn [9], with $\gamma_a = C\sqrt{\Delta_c}/(\varepsilon + \Delta_c)$ and $\Delta_a = -C\sqrt{\varepsilon}/(\varepsilon + \Delta_c)$.

The resulting atomic frequency shift Δ_a is negative in the present model, vanishing for $\varepsilon \to 0$ ($D = 2$). The long-time behavior of $\alpha(t)$ can be shown to exhibit a tail decaying as $t^{-3/2}$ (or $t^{-1/2}$ at $\Delta_c = C\varepsilon^{-1/2}$) and oscillating at the cutoff frequency ω_U (Figure 3).

With the increase of ε, the smoothing inhibits the decay more strongly for ω_a at cutoff, $\Delta_c = 0$, because $G(\omega_a \simeq \omega_U)$ is now weaker and the stable-state probability c_0 (eqn [7]) is correspondingly larger. By contrast, at a large detuning from cutoff Δ_c, the large-ε smoothing allows the decay to become complete, and the corresponding spectrum is entirely Lorentzian.

A defect in the periodic structure can produce a narrow-linewidth local mode in the PBG, whose spectral response is describable by a Lorentzian:

$$G_d(\omega) = \frac{\gamma_d}{\pi} \frac{\Gamma_d^2}{\Gamma_d^2 + (\omega - \omega_d)^2} \qquad [14]$$

where γ_d characterizes the coupling strength of the atomic dipole with the defect field, whereas ω_d and Γ_d stand for the line center and width, respectively. The presence of a nonvanishing DOM in the PBG, due to a defect, causes spontaneous emission in the PBG spectral range. This broadens the discrete state ω_0, which becomes metastable (see Figure 4). The oscillator strength of the line at ω_0 is then proportional to c_0.

EIT and Cross-Coupling of Photons in Doped PCs

Nonlinear effects, whereby one light beam influences another, require strong fields or else light confinement

in a high-Q cavity. The analysis above has shown that by choosing an appropriate detuning of an atomic transition from the PBG cutoff, the spontaneous coherence is established between the states of an initially excited atom. The ability to control this coherence, by varying the detuning, opens interesting perspectives for optical processes in PCs containing multilevel atoms with a resonant transition near a bandedge. From among such processes, we discuss here electromagnetically induced transparency (EIT) and its applicability to nonlinear photon switching or giant cross-phase modulation.

Let us examine the nonlinear coupling of two weak (single-photon) optical fields E_a and E_b with the frequencies ω_a and ω_b, respectively, propagating along the z-axis in a PC dilutely doped with identical four-level atoms (**Figure 5a**). These fields interact with the atoms via the transitions $|1\rangle \rightarrow |2\rangle$ and $|3\rangle \rightarrow |4\rangle$, respectively, while the transition $|2\rangle \rightarrow |3\rangle$ is coupled to the structured mode-continuum $\rho(\omega)$ in the PC (**Figure 5b**). Initially the atoms are in the ground state $|1\rangle$ and the continuum is in the vacuum state $\{0_\omega\}$. Then the wavefunction of the system at the position z_l of the lth atom reads:

$$|\Psi(z_l, t)\rangle = \alpha_1 |1, \{0_\omega\}\rangle + \alpha_2 |2, \{0_\omega\}\rangle + \int \beta_{3,\omega} |3, 1_\omega\rangle \rho(\omega) \mathrm{d}\omega + \int \beta_{4,\omega} |4, 1_\omega\rangle \rho(\omega) \mathrm{d}\omega \qquad [15]$$

The consecutive terms in eqn [15] denote the atom being in states $|1\rangle$, $|2\rangle$, $|3\rangle$, or $|4\rangle$, with zero $\{0_\omega\}$ or one $|1_\omega\rangle$ photon in mode ω whose DOM is $\rho(\omega)$, and α_1, α_2, $\beta_{3,\omega}$, or $\beta_{4,\omega}$ are the corresponding probability amplitudes. Upon making the weak-field linear-response approximation, we can set $\alpha_1 \simeq 1$ and solve the set of equations for the slowly varying (compared to an optical cycle) probability amplitudes α and β, using the perturbation theory. Under these conditions, we effectively obtain a free-space propagation of the E_b field. By contrast, the evolution of the E_a field in the slowly varying envelope approximation, is given by $E_a(z,t) = E_a(0, t - z/v_g)\exp(ipz)$. We thus see that the real part of the macroscopic complex polarizability p is responsible for the phase shift ϕ_a of the E_a field, $\phi_a = \mathrm{Re}(p)z$, while the probability of the absorption $\mathscr{A}$ of the field depends on the imaginary part of p, $\mathscr{A} = 1 - \exp[-2\mathrm{Im}(p)z]$. The polarizability is expressed by

$$p = a_0 \frac{i\gamma_2/2}{\gamma_2/2 - i\Delta_a + I(\Delta_a)} \qquad [16]$$

where $a_0 \equiv \sigma_0 N$ is the linear resonant absorption coefficient on the atomic transition $|1\rangle \rightarrow |2\rangle$, with σ_0 the resonant absorption cross-section and N the density of doping atoms, γ_2 is the radiative width of

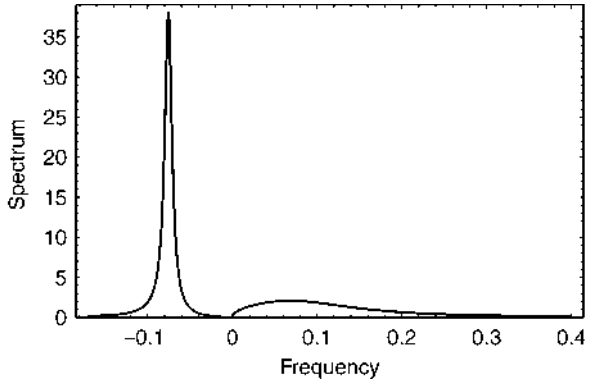

Figure 4 Spectrum w(ω)(same units as in **Figures 2** and **3**) for ω_a at cutoff, $\Delta_c = 0$, in the presence of a defect in the PBG, $\omega_d = -10$, $\gamma = 0.1$, $\Gamma_d = 3$, $\varepsilon = 10^{-3}$. The oscillator strength is now roughly equally shared between the distribution above cutoff (same as in **Figure 2b**) and the defect peak at ω_0.

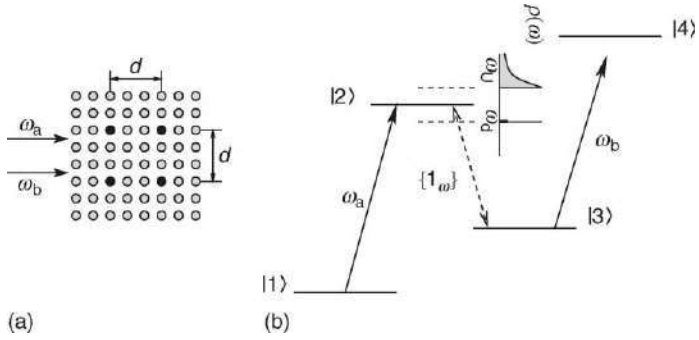

Figure 5 (a) Photonic crystal dilutely doped with atoms located at black dots. (b) Four-level atom coupled to a structured continuum $\rho(\omega)$ near the band-edge or defect mode frequencies (DOM plotted) via the intermediate transition $|2\rangle \rightarrow |3\rangle$ and interacting with two weak fields E_a and E_b at the sideband transitions $|1\rangle \rightarrow |2\rangle$ and $|3\rangle \rightarrow |4\rangle$, respectively.

state $|2\rangle$, $\Delta_a = \omega_a - \omega_{21}$ is the detuning from the atomic resonance ω_{21}, and $I(\Delta_a)$ is the integral of the saturation factor over the structured DOM. The group velocity v_g is expressed as $v_g = \partial_{k_a}\omega_a = [n_a/c + \partial_{\omega_a}\mathrm{Re}(p)]^{-1}$, where n_a is the (averaged) refraction index at the frequency ω_a.

To calculate $I(\Delta_a)$, we assume the isotropic PBG model, eqn [13] with $\varepsilon \to 0$, with the atoms doped at the positions of the local defects in the PC separated by a distance d from each other. These defect modes in the PBG are localized around each atomic site in a volume $V_d \simeq (rL)^3$ of several $(r)^3$ lattice cells L^3, with $L \simeq \pi c/\omega_{23}$, and serve as effective high-Q cavities, eqn [14] with $\Gamma_d \ll 1$. Assuming $|\Delta_b| \gg |\Delta|$, $|\Delta_a|$, γ_4, where $\Delta_b = \omega_b - \omega_{43}$ and $\Delta = \omega - \omega_{23}$, the integration leads to

$$I(\Delta_a) = \frac{B_d^2}{\gamma_{31} - i(\Delta_a - \Delta_d - s_3)} - \frac{B_U^{3/2}}{\sqrt{i\gamma_{31} + (\Delta_a - \Delta_U - s_3)}} \qquad [17]$$

where $\Delta_{d,U} = \omega_{d,U} - \omega_{23} \ll \omega_{23}$ are the detunings of the defect-mode and PBG-edge frequencies from the atomic resonance ω_{23}, γ_{31} the $|1\rangle \leftrightarrow |3\rangle$ decoherence rate, $s_3 = (\mu_{34}^2/\hbar^2\Delta_b)|E_b|^2$ is the E_b field-induced ac Stark shift of level $|3\rangle$ (μ_{ij} is the atomic dipole matrix element on the transition $|i\rangle \to |j\rangle$), and B_d and B_U are the coupling constants of the atom with the structured reservoir, whose main contributions are near ω_d and ω_U.

To illustrate the results of the foregoing analysis, we plot in Figure 6 the imaginary and real parts of the polarizability (eqn [16]). Consider first the case of one incident field $E_a(E_b = 0)$. Clearly, two frequency regions, $\Delta_a \sim \Delta_d$ and $\Delta_a \sim \Delta_U$, where the absorption vanishes and, at the same time, the dispersion slope is steep, are of particular interest. One can see that there is, however, a substantial difference between the spectra in the foregoing frequency regions, for the following physical reasons. First, in the vicinity of Δ_d, the atom interacts with the defect mode as in a high-Q cavity. This strong interaction 'dresses' the atomic states $|2\rangle$ and $|3\rangle$, thereby splitting the spectrum around $\Delta_a \simeq \Delta_d$ by the amount equal roughly to $2B_d$. Near the two-photon Raman resonance $\Delta_a = \Delta_d$, the two alternative transition paths $|1\rangle \to |2\rangle$ (the direct transition) and $|1\rangle \to |2\rangle \to |3\rangle \to |2\rangle$ (transition via state $|3\rangle$) interfere distractively with each other, cancelling thus the absorption of the E_a field and the medium becomes transparent to the radiation. This effect has been widely studied in atomic vapors, where the transition $|2\rangle \to |3\rangle$ is strongly driven by a coherent laser field, and is called electromagnetically induced transparency (EIT). The transparency window is rather broad and is given by the inverse Lorentzian (see the first term on the right-hand side of eqn [17]). Due to the steepness of the dispersion curve, the corresponding group velocity is much smaller than the speed of light, $v_g \simeq 2B_d^2/(\gamma_2 a_0) \ll c$, which leads to a large delay time $T_{del} = \zeta/v_g$ at the exit $z = \zeta$ from the medium. One has to keep in mind, however, that the absorption-free propagation time is limited by the EIT decoherence time $T_{del} < \gamma_{31}^{-1}$, which imposes a limitation on the length ζ of the active PC medium. Second, in the vicinity of Δ_U, the strong interaction of the atom with the continuum near the bandedge ω_U causes the Autler–Townes splitting of level $|2\rangle$ into a doublet with a separation equal roughly to B_U. One component of this doublet is shifted out of the PBG, while the other one remains

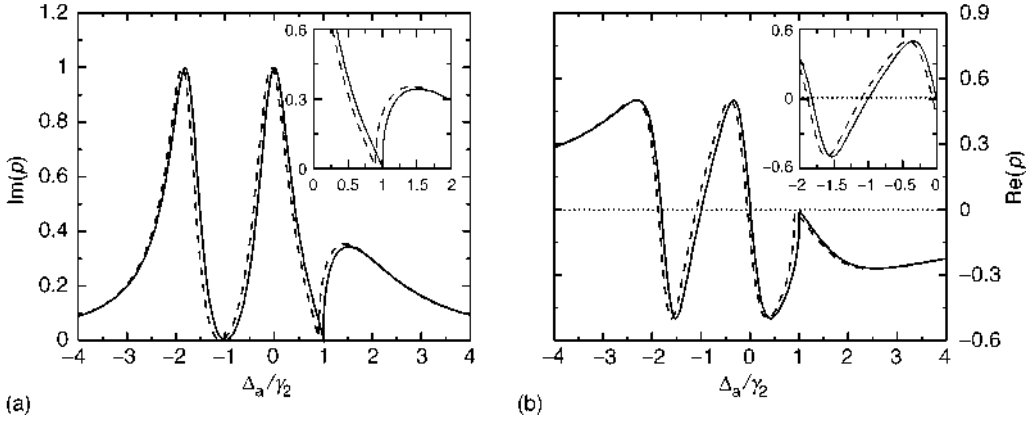

Figure 6 (a) Imaginary and (b) real part of the complex polarizability p as a function of the detuning Δ_a for the case $E_b = 0$ (solid lines) and $E_b \neq 0$ (dashed lines). The parameters (normalized by γ_2) are: $\Delta_d = -1$, $\Delta_U = 1$, $\gamma_{31} = 0.001$, $B_d = B_U = 1$, $s_3 = -0.1$, and $a_0 = 1\ \mathrm{cm}^{-1}$. The insets magnify the important frequency regions.

within the gap and forms the photon–atom bound state. Consequently, there is vanishing absorption and rapid variation of the dispersion at $\Delta_a \simeq \Delta_U$. Since the transparency region is very narrow with a width $\delta\omega \sim \gamma_{31}(\ll \gamma_2, B_U)$, for an absorption-free propagation of the E_a pulse, its temporal width τ_a should satisfy the condition $\tau_a > \pi/\delta\omega$. Simultaneously, a small deviation from the condition $\Delta_a = \Delta_U$ will lead to a strong increase in the absorption of the E_a field.

Let us now switch on the E_b field. As seen from eqn [17], its effect is merely to shift the spectrum by the amount equal to s_3 (**Figure 6**). This shift, however, will have different implications in the two frequency regions distinguished above: if $s_3 \ll B_d$, i.e., the Stark shift is smaller than the width of the EIT window at Δ_d, the medium will still remain transparent for an E_a field with the detuning $\Delta_a = \Delta_d$, but its phase will experience an appreciable nonlinear shift ϕ_a, given by $\phi_a = Re(p)z \simeq -\gamma_2 a_0 s_3 z/(2B_d^2)$. On the other hand, for an E_a field with the detuning $\Delta_a = \Delta_U$, the medium, which is transparent for $E_b = 0$, $\mathrm{Im}(p) \ll a_0$, will become highly absorptive (opaque) even for such a small frequency shift as s_3 (provided $s_3 < 0$ and $|s_3| > \Delta\omega$), $\mathrm{Im}(p) \simeq \gamma_2 a_0 \sqrt{|s_3|}/(2B_U^{3/2})$ and thus acting as an ultrasensitive, effective switch.

The remaining question is how to maximize the interaction between the E_a pulse, which propagates with a small group velocity, and the E_b pulse, which propagates with a velocity close to the speed of light. The interaction between the fields is maximized if: they enter the medium simultaneously; the transverse shapes of their envelopes overlap completely; and the pulse length l_b of the E_b field satisfies the condition $(l_b + \zeta)/c \leq \zeta/v_g$. Then the E_b pulse leaves the medium not later than the E_a pulse. The effective interaction length between the two pulses is, therefore, $z_{\mathrm{eff}} \sim l_b v_g/c \leq \zeta$, after which the two pulses slip apart. Thus, the presence of the E_b (control) field induces either strong absorption or a large phase shift of the E_a (signal) field, depending on the frequency region employed.

The effects surveyed above reveal unusual features of spontaneous emission and photon–atom binding in PCs, atomic interaction with the field of a high-Q defect mode, and nonlinear coupling of two fields via four-level dopant atoms in PCs. These effects are of fundamental interest. In addition, they can serve as the basis for highly efficient optical communications and data processing, in either the classical or the quantum domain, by providing two key elements: ultrasensitive nonlinear phase-shifters and photon switches.

List of Units and Nomenclature

Atomic or photonic eigenstates [dimensionless]	$\lvert\ldots\rangle$
Complex polarizability [cm^{-1}]	p
Coupling constant [s^{-1}]	κ, $C^{2/3}$, B
Decay rate [s^{-1}]	γ
Density of modes [s]	ρ
Frequency [s^{-1}]	ω, Δ, s
Phase [dimensionless]	ϕ
Probability amplitudes [dimensionless]	α, β, c
Reservoir spectral response [s^{-1}]	G
Resonant absorption coefficient [cm^{-1}]	a_0
Spontaneous emission spectrum [s]	W
Wave vector [m^{-1}]	k

See also

Electromagnetically Induced Transparency. **Photonic Crystals:** Photonic Crystal Lasers, Cavities and Waveguides.

Further Reading

Joannopoulos JD, Meade RD and Winn JN (1995) *Photonic Crystals: Molding the Flow of Light*. Princeton, NJ: Princeton University Press.

John S and Wang J (1991) Quantum optics of localized light in a photonic band gap. *Physics Review B* 43: 12772–12789.

Kofman AG, Kurizki G and Sherman B (1994) Spontaneous and induced atomic decay in photonic band structures. *Journal of Modern Optics* 41: 353–384.

Lambropoulos P, Nikolopoulos GM, Nielsen TR and Bay S (2000) Fundamental quantum optics in structured reservoirs. *Reports on Progress in Physics* 63: 455–503.

Loudon R (1983) *The Quantum Theory of Light*. Oxford, UK: Clarendon Press.

Meystre P and Sargent M III (1991) *Elements of Quantum Optics*. Berlin: Springer Verlag.

Petrosyan D and Kurizki G (2001) Photon–photon correlations and entanglement in doped photonic crystals. *Physics Review A* 64: 023810-1-6.

Sakoda K (2001) *Optical Properties of Photonic Crystals*. Springer Series in Optical Sciences, 80. Berlin: Springer Verlag.

Scully MO and Zubairy MS (1997) *Quantum Optics*. Cambridge, UK: Cambridge University Press.

Electromagnetic Theory

S G Johnson and J D Joannopoulos,
Massachusetts Institute of Technology, Cambridge, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Photonic crystals are periodically structured electromagnetic media, generally possessing photonic bandgaps: ranges of frequency in which light cannot propagate through the structure. This periodicity, whose lengthscale is proportional to the wavelength of light in the bandgap, is the electromagnetic analog of a crystalline atomic lattice, where the latter acts on the electron wavefunction to produce the familiar band gaps, semiconductors, etc., of solid-state physics. The study of photonic crystals is likewise governed by the Bloch–Floquet theorem, and intentionally introduced defects in the crystal (analogous to electronic dopants) give rise to localized electromagnetic states: linear waveguides and point-like cavities. The crystal can thus form a kind of perfect optical 'insulator', which can confine light around sharp bends, in lower-index media, and within wavelength-scale cavities, among other novel possibilities for control of electromagnetic phenomena. Below is introduced the basic theoretical background of photonic crystals in one, two, and three dimensions (schematically depicted in Figure 1), as well as hybrid structures that combine photonic-crystal effects in some directions with more-conventional index guiding in other directions. (Line and point defects in photonic crystals are discussed in another article.)

Electromagnetic wave propagation in periodic media was first studied by Lord Rayleigh in 1887, in connection with the peculiar reflective properties of a crystalline mineral with periodic 'twinning' planes (across which the dielectric tensor undergoes a mirror flip). These correspond to one-dimensional photonic crystals, and he identified the fact that they have a narrow bandgap prohibiting light propagation through the planes. This bandgap is angle-dependent, due to the differing periodicities experienced by light propagating at non-normal incidences, producing a reflected color that varies sharply with angle. (A similar effect is responsible for many other iridescent colors in nature, such as those of butterfly wings and abalone shells.) Although multilayer films received intensive study over the following century, it was not until 100 years later, when Yablonovitch and John, in 1987, joined the tools of classical electromagnetism and solid-state physics, that the concepts of omnidirectional photonic bandgaps in two and three dimensions was introduced. This generalization, which inspired the name 'photonic crystal', led to many subsequent developments in their fabrication, theory, and application, from integrated optics to negative refraction to optical fibers that guide light in air.

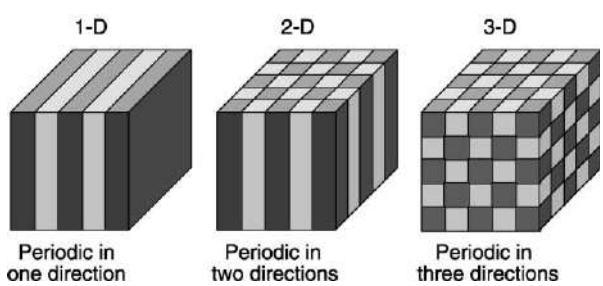

Figure 1 Schematic depiction of photonic crystals periodic in one, two, and three directions, where the periodicity is in the material (typically dielectric) structure of the crystal. Only a 3d periodicity, with a more complex topology than is shown here, can support an omnidirectional photonic bandgap.

Maxwell's Equations in Periodic Media

The study of wave propagation in three-dimensionally periodic media was pioneered by Felix Bloch in 1928, unknowingly extending an 1883 theorem in one dimension by Gaston Floquet. Bloch proved that waves in such a medium can propagate without scattering, their behavior described by a periodic envelope function multiplied by a planewave. Although Bloch studied quantum mechanics, leading to the surprising result that electrons in a conductor scatter only from imperfections and not from the periodic ions, the same techniques can be applied to electromagnetism by casting Maxwell's equations as an eigenproblem in analog with Schrödinger's equation. By combining the source-free Faraday's and Ampere's laws at a fixed (angular) frequency ω, i.e., time dependence $e^{-i\omega t}$, one can obtain an equation in only the magnetic field $\vec{H}$:

$$\vec{\nabla} \times \frac{1}{\varepsilon} \vec{\nabla} \times \vec{H} = \left(\frac{\omega}{c}\right)^2 \vec{H} \qquad [1]$$

where ε is the dielectric function $\varepsilon(x, y, z)$ and c is the speed of light. This is an eigenvalue equation, with eigenvalue $(\omega/c)^2$ and an eigen-operator $\vec{\nabla} \times (1/\varepsilon)\vec{\nabla} \times$ that is Hermitian (acts the same to the left and right)

under the inner product $\int \vec{H}^* \cdot \vec{H}'$ between two fields $\vec{H}$ and $\vec{H}'$. (The two curls correspond roughly to the 'kinetic energy' and $1/\varepsilon$ to the 'potential' compared to the Schrödinger Hamiltonian $\nabla^2 + V$.) It is sometimes more convenient to write a generalized Hermitian eigenproblem in the electric field $\vec{E}$, $\vec{\nabla} \times \vec{\nabla} \times \vec{E} = (\omega/c)^2 \varepsilon \vec{E}$, which separates the kinetic and potential terms. Electric fields that lie in higher ε, i.e., lower potential, will have lower ω; this is discussed further in the context of the variational theorem of eqn [3].

Thus, the same linear-algebraic theorems as those in quantum mechanics can be applied to the electromagnetic wave solutions. The fact that the eigen-operator is Hermitian and positive-definite (for real $\varepsilon > 0$) implies that the eigenfrequencies ω are real, for example, and also leads to orthogonality, variational formulations, and perturbation–theory relations that are discussed further below. An important difference compared to quantum mechanics is that there is a transversality constraint: one typically excludes $\vec{\nabla} \cdot \vec{H} \neq 0$ (or $\vec{\nabla} \cdot \varepsilon \vec{E} \neq 0$) eigensolutions, which lie at $\omega = 0$; i.e., static-field solutions with free magnetic (or electric) charge are forbidden.

Bloch Waves and Brillouin Zones

A photonic crystal corresponds to a periodic dielectric function $\varepsilon(\vec{x}) = \varepsilon(\vec{x} + \vec{R}_i)$ for some primitive lattice vectors $\vec{R}_i$ ($i = 1, 2, 3$ for a crystal periodic in all three dimensions). In this case, the Bloch–Floquet theorem for periodic eigenproblems states that the solutions to eqn [1] can be chosen of the form $\vec{H}(\vec{x}) = e^{i\vec{k}\cdot\vec{x}} \vec{H}_{n,\vec{k}}(\vec{x})$ with eigenvalues $\omega_n(\vec{k})$, where $\vec{H}_{n,\vec{k}}$ is a periodic envelope function satisfying

$$\left(\vec{\nabla} + i\vec{k}\right) \times \frac{1}{\varepsilon}\left(\vec{\nabla} + i\vec{k}\right) \times \vec{H}_{n,\vec{k}} = \frac{\omega_n(\vec{k})}{c}^2 \vec{H}_{n,\vec{k}} \qquad [2]$$

yielding a different Hermitian eigenproblem over the primitive cell of the lattice at each Bloch wavevector $\vec{k}$. This primitive cell is a finite domain if the structure is periodic in all directions, leading to discrete eigenvalues labeled by $n = 1, 2, \ldots$. These eigenvalues $\omega_n(\vec{k})$ are continuous functions of $\vec{k}$, forming discrete 'bands' when plotted versus the latter, in a 'band structure' or dispersion diagram – both ω and $\vec{k}$ are conserved quantities, meaning that a band diagram maps out all possible interactions in the system. (Note also that $\vec{k}$ is not required to be real; complex $\vec{k}$ gives evanescent modes that can exponentially decay from the boundaries of a finite crystal, but which cannot exist in the bulk.)

Moreover, the eigensolutions are periodic functions of $\vec{k}$ as well: the solution at $\vec{k}$ is the same as the solution at $\vec{k} + \vec{G}_j$, where $\vec{G}_j$ is a primitive reciprocal lattice vector defined by $\vec{R}_i \cdot \vec{G}_j = 2\pi\delta_{i,j}$. Thanks to this periodicity, one need only compute the eigensolutions for $\vec{k}$ within the primitive cell of this reciprocal lattice – or, more conventionally, one considers the set of inequivalent wavevectors closest to the $\vec{k} = 0$ origin, a region called the first Brillouin zone. For example, in a one-dimensional system, where $R_1 = a$ for some periodicity a and $G_1 = 2\pi/a$, the first Brillouin zone is the region $k = -\pi/a \ldots \pi/a$; all other wavevectors are equivalent to some point in this zone under translation by a multiple of G_1. Furthermore, the first Brillouin zone may itself be redundant if the crystal possesses additional symmetries such as mirror planes; by eliminating these redundant regions, one obtains the irreducible Brillouin zone, a convex polyhedron that can be found tabulated for most crystalline structures. In the preceding one-dimensional example, since most systems will have time-reversal symmetry ($k \to -k$), the irreducible Brillouin zone would be $k = 0 \ldots \pi/a$.

The familar dispersion relations of uniform waveguides arise as a special case of the Bloch formalism: such translational symmetry corresponds to a period $a \to 0$. In this case, the Brillouin zone of the wavevector k (also called β) is unbounded, and the envelope function $\vec{H}_{n,k}$ is a function only of the transverse coordinates.

The Origin of the Photonic Bandgap

A complete photonic bandgap is a range of ω in which there are no propagating (real $\vec{k}$) solutions of Maxwell's eqn [2] for any $\vec{k}$, surrounded by propagating states above and below the gap. There are also incomplete gaps, which only exist over a subset of all possible wavevectors, polarizations, and/or symmetries. Both sorts of gaps are discussed in the subsequent sections, but in either case, their origins are the same and can be understood by examining the consequences of periodicity for a simple one-dimensional system.

Consider a one-dimensional system with uniform $\varepsilon = \bar{\varepsilon}$, which has planewave eigensolutions $\omega(k) = ck$, as depicted in Figure 2(left). This ε has trivial periodicity a for any $a \geq 0$, with $a = 0$ giving the usual unbounded dispersion relation. One is free, however, to label the states in terms of Bloch envelope functions and wavevectors for some $a \neq 0$, in which case the bands for $|k| > \pi/a$ are translated (folded) into the first Brillouin zone, as shown by the dashed lines in Figure 2(left). In particular, the $k = -\pi/a$ mode in this description now lies at an equivalent wavevector to the $k = \pi/a$ mode, and at the same frequency; this accidental degeneracy is an artifact of

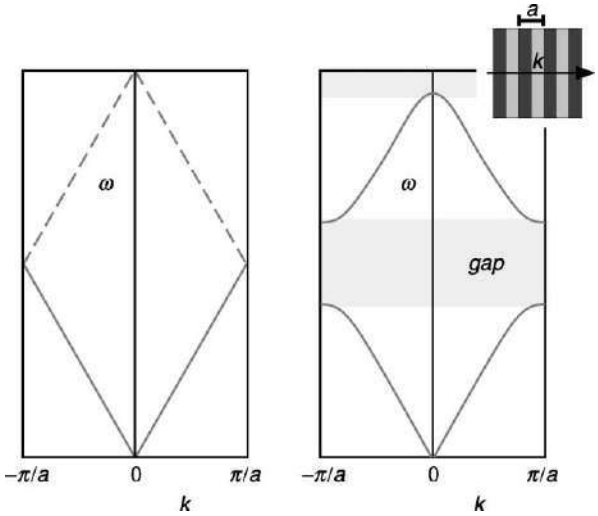

Figure 2 Left: Dispersion relation (band diagram), frequency ω versus wavenumber k, of a uniform one-dimensional medium, where the dashed lines show the 'folding' effect of applying Bloch's theorem with an artificial periodicity a. Right: Schematic effect on the bands of a physical periodic dielectric variation (inset), where a gap has been opened by splitting the degeneracy at the $k = \pm\pi/a$ Brillouin-zone boundaries (as well as a higher-order gap at $k = 0$).

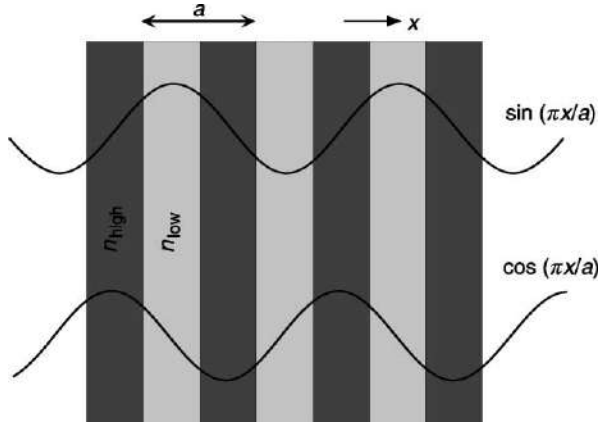

Figure 3 Schematic origin of the band gap in one dimension. The degenerate $ka = \pm\pi/a$ planewaves of a uniform medium are split into $\cos(\pi x/a)$ and $\sin(\pi x/a)$ standing waves by a dielectric periodicity, forming the lower and upper edges of the bandgap, respectively – the former has electric-field peaks in the high dielectric (n_{high}) and so will lie at a lower frequency than the latter (which peaks in the low dielectric).

the 'artificial' period that has been chosen. Instead of writing these wave solutions with electric fields $\vec{E}(x) \sim e^{\pm i\pi x/a}$, one can equivalently write linear combinations $e(x) = \cos(\pi x/a)$ and $o(x) = \sin(\pi x/a)$ as shown in Figure 3, both at $\omega = c\pi/a\sqrt{\bar{\varepsilon}}$. Now, however, suppose that one perturbs ε so that it is nontrivially periodic with period a; for example, a sinusoid $\varepsilon(x) = \bar{\varepsilon}\cdot[1 + \Delta\cdot\cos(2\pi x/a)]$, or a square wave as in the inset of Figure 2(right). In the presence of such an oscillating 'potential', the accidental degeneracy between $e(x)$ and $o(x)$ is broken: supposing $\Delta > 0$, then the field $e(x)$ is more concentrated in the higher-ε regions than $o(x)$, and so lies at a lower frequency. This opposite shifting of the bands away from the mid-gap frequency $\omega \cong c\pi/a\sqrt{\bar{\varepsilon}}$ creates a bandgap, as depicted in Figure 2(right). (In fact, from the perturbation theory described subsequently, one can show that for $\Delta \ll 1$ the bandgap, as a fraction of mid-gap frequency, is $\Delta\omega/\omega \cong \Delta/2$.) By the same arguments, it follows that any periodic dielectric variation in one dimension will lead to a bandgap, albeit a small gap for a small variation; a similar result was identified by Lord Rayleigh in 1887.

More generally, it follows immediately from the properties of Hermitian eigensystems that the eigenvalues minimize a variational problem:

$$\omega^2_{n,\vec{k}} = \min_{\vec{E}_{n,\vec{k}}} \frac{\int \left|\left(\vec{\nabla} + i\vec{k}\right)\times\vec{E}_{n,\vec{k}}\right|^2}{\int \varepsilon\left|\vec{E}_{n,\vec{k}}\right|^2}\, c^2 \qquad [3]$$

in terms of the periodic electric field envelope $\vec{E}_{n,\vec{k}}$, where the numerator minimizes the 'kinetic energy' and the denominator minimizes the 'potential energy' Here, the $n > 1$ bands are additionally constrained to be orthogonal to the lower bands:

$$\int \vec{H}^*_{m,\vec{k}}\cdot\vec{H}_{n,\vec{k}} = \int \varepsilon\vec{E}^*_{m,\vec{k}}\cdot\vec{E}_{n,\vec{k}} = 0 \qquad [4]$$

for $m < n$. Thus, at each $\vec{k}$, there will be a gap between the lower 'dielectric' bands concentrated in the high dielectric (low potential) and the upper 'air' bands that are less concentrated in the high dielectric: the air bands are forced out by the orthogonality condition, or otherwise must have fast oscillations that increase their kinetic energy. (The dielectric/air bands are analogous to the valence/conduction bands in a semiconductor.)

In order for a complete bandgap to arise in two or three dimensions, two additional hurdles must be overcome. First, although in each symmetry direction of the crystal (and each $\vec{k}$ point) there will be a bandgap by the one-dimensional argument, these bandgaps will not necessarily overlap in frequency (or even lie between the same bands). In order that they overlap, the gaps must be sufficiently large, which implies a minimum ε contrast (typically at least 4/1 in 3d). Since the 1d mid-gap frequency $\sim c\pi/a\sqrt{\bar{\varepsilon}}$ varies inversely with the period a, it is also helpful if the periodicity is nearly the same in different directions – thus, the largest gaps typically arise for hexagonal lattices in 2d and fcc lattices in 3d, which have the most nearly circular/spherical Brillouin zones. Second, one must take into account

the vectorial boundary conditions on the electric field: moving across a dielectric boundary from ε to some $\varepsilon' < \varepsilon$, the inverse 'potential' $\varepsilon|\vec{E}|^2$ will decrease discontinuously if $\vec{E}$ is parallel to the interface ($\vec{E}_\parallel$ is continuous) and will increase discontinuously if $\vec{E}$ is perpendicular to the interface ($\varepsilon\vec{E}_\perp$ is continuous). This means that, whenever the electric field lines cross a dielectric boundary, it is much harder to strongly contain the field energy within the high dielectric, and the converse is true when the field lines are parallel to a boundary. Thus, in order to obtain a large bandgap, a dielectric structure should consist of thin, continuous veins/membranes along which the electric field lines can run – this way, the lowest band(s) can be strongly confined, while the upper bands are forced to a much higher frequency because the thin veins cannot support multiple modes (except for two orthogonal polarizations). The veins must also run in all directions, so that this confinement can occur for all $\vec{k}$ and polarizations, necessitating a complex topology in the crystal.

Ultimately, however, in two or three dimensions there are only rules of thumb for the existence of a bandgap in a periodic structure, since no rigorous criteria have yet been determined. This made the design of 3d photonic crystals a trial and error process, with the first example by Ho *et al.* of a complete 3d gap coming three years after the initial 1987 concept. As is discussed by the final section below, a small number of families of 3d photonic crystals have since been identified, with many variations thereof explored for fabrication.

Computational Techniques

Because photonic crystals are generally complex, high index-contrast, two- and three-dimensional vectorial systems, numerical computations are a crucial part of most theoretical analyses. Such computations typically fall into three categories: time-domain 'numerical experiments' that model the time-evolution of the fields with arbitrary starting conditions in a discretized system (e.g., finite-difference); definite-frequency transfer matrices wherein the scattering matrices are computed in some basis to extract transmission/reflection through the structure; and frequency-domain methods to directly extract the Bloch fields and frequencies by diagonalizing the eigenoperator. The first two categories intuitively correspond to directly measurable quantities such as transmission (although they can also be used to compute e.g., eigenvalues), whereas the third is more abstract, yielding the band diagrams that provide a guide to interpretation of measurements as well as a starting-point for device design and semi-analytical methods. Moreover, several band diagrams are included in the following sections, and so the frequency-domain method used to compute them is briefly outlined here.

Any frequency–domain method begins by expanding the fields in some complete basis, $\vec{H}_{\vec{k}}(\vec{x}) = \sum_n h_n \vec{b}_n(\vec{x})$, transforming the partial differential eqn [2] into a discrete matrix eigenvalue problem for the coefficients h_n. Truncating the basis to N elements leads to $N \times N$ matrices, which could be diagonalized in $O(N^3)$ time by standard methods. This is impractical for large 3d systems, however, and is also unnecessary – typically, one only wants the few lowest eigenfrequencies, in which case one can use iterative eigensolver methods requiring only $\sim O(N)$ time. Perhaps the simplest such method is based directly on the variational theorem eqn [3]: given some starting coefficients h_n, one iteratively minimizes the variational 'Rayleigh' quotient using e.g., preconditioned conjugate–gradient descent. This yields the lowest band's eigenvalue and field, and upper bands are found by the same minimization while orthogonalizing against the lower bands ('deflation'). There is one additional difficulty, however, and that is that one must at the same time enforce the $(\vec{\nabla} + i\vec{k}) \cdot \vec{H}_{\vec{k}} = 0$ transversality constraint, which is nontrivial in three dimensions. The simplest way to maintain this constraint is to employ a basis that is already transverse, for example planewaves $\vec{h}_{\vec{G}} e^{i\vec{G}\cdot\vec{x}}$ with transverse amplitudes $\vec{h}_{\vec{G}} \cdot (\vec{G} + \vec{k}) = 0$. (In such a planewave basis, the action of the eigen-operator can be computed via a fast Fourier transform in $O(N \log N)$ time.)

Semi-analytical Methods: Perturbation Theory

As in quantum mechanics, the eigenstates can be the starting point for many analytical and semi-analytical studies. One common technique is perturbation theory, applied to small deviations from an ideal system – closely related to the variational eqn [3], perturbation theory can be exploited to consider effects such as nonlinearities, material absorption, fabrication disorder, and external tunability. Not only is perturbation theory useful in its own right, but it also illustrates both old and peculiarly new features that arise in such analyses of electromagnetism compared to scalar problems such as quantum mechanics.

Given an unperturbed eigenfield $\vec{E}_{n,\vec{k}}$ for a structure ε, the lowest-order correction $\Delta\omega_n^{(1)}$ to the eigenfrequency from a small perturbation $\Delta\varepsilon$ is given by

$$\Delta\omega_n^{(1)} = -\frac{\omega_n(\vec{k})}{2} \frac{\int \Delta\varepsilon \left|\vec{E}_{n,\vec{k}}\right|^2}{\int \varepsilon \left|\vec{E}_{n,\vec{k}}\right|^2} \qquad [5]$$

where the integral is over the primitive cell of the lattice. A Kerr nonlinearity would give $\Delta\varepsilon \sim |\vec{E}|^2$, material absorption would produce an imaginary frequency correction (decay coefficient) from a small imaginary $\Delta\varepsilon$, and so on. Similarly, one can compute the shift in frequency from a small $\Delta\vec{k}$ in order to determine the group velocity $d\omega/dk$; this variation of perturbation theory is also called $k \cdot p$ theory. All such first-order perturbation corrections are well known from quantum mechanics, and in the limit of infinitesimal perturbations give the exact Hellman–Feynman expression for the derivative of the eigenvalue. However, in the limit where $\Delta\varepsilon$ is a small shift Δh of a dielectric boundary between some ε_1 and ε_2, an important class of geometric perturbation, eqn [5] gives a surface integral of $|\vec{E}|^2$ on the interface, but this is ill-defined because the field there is discontinous. The proper derivation of perturbation theory in the face of such discontinuity requires a more careful limiting process from an anisotropically smoothed system, yielding the surface integral:

$$\Delta\omega_n^{(1)} = -\frac{\omega_n(\vec{k})}{2} \frac{\int\int \Delta h \left(\Delta\varepsilon_{12}|\vec{E}_{\parallel}|^2 - \Delta\varepsilon_{12}^{-1}|D_{\perp}|^2\right)}{\int \varepsilon |\vec{E}_{n,\vec{k}}|^2}$$

where $\Delta\varepsilon_{12} = \varepsilon_1 - \varepsilon_2, \Delta\varepsilon_{12}^{-1} = \varepsilon_1^{-1} - \varepsilon_2^{-1}$, and $\vec{E}_{\parallel}/D_{\perp}$ denotes the (continuous) interface parallel/perpendicular components of the unperturbed electric/displacement eigenfield. A similar expression is required in high index-contrast systems to employ, e.g., coupled-mode theory for slowly-varying waveguides or Green's functions for interface roughness.

Standard perturbation–theory techniques also provide expressions for higher-order corrections to the eigenvalue and eigenfield, based on an expansion in the basis of the unperturbed eigenfields. This approach, however, runs into immediate difficulty because the eigenfields are also subject to the transversality constraint, $(\vec{\nabla} + i\vec{k}) \cdot \varepsilon\vec{E}_{\vec{k}} = 0$, and this constraint varies with ε and $\vec{k}$ – the eigenfields $\vec{E}_{\vec{k}}$ are not a complete basis for the eigenfields constrained at a different ε or $\vec{k}$. For ε perturbations, this problem can be eliminated by using the $\vec{H}$ or $\vec{D}$ eigenproblems, whose constraints are independent of ε. For $\vec{k}$ perturbations, one can employ a transformation by Sipe to derive a corrected higher-order perturbation theory (for e.g., the group-velocity dispersion), based on the fact that all of the non-transverse fields lie at $\omega = 0$. Such completeness issues also arise applying the variational eqn [3], as was noted in the previous section: in order for a useful variational bound to apply, one must operate in the constrained (transverse) subspace.

Two-Dimensional Photonic Crystals

After the identification of one-dimensional bandgaps, it took a full century to add a second dimension, and three years to add the third. It should therefore come as no surprise that 2d systems exhibit most of the important characteristics of photonic crystals, from nontrivial Brillouin zones to topological sensitivity to a minimum index contrast, and can also be used to demonstrate most proposed photonic–crystal devices. The key to understanding photonic crystals in two dimensions is to realize that the fields in 2d can be divided into two polarizations by symmetry: *TM* (transverse magnetic), in which the magnetic field is in the (xy) plane and the electric field is perpendicular (z); and *TE* (transverse electric), in which the electric field is in the plane and the magnetic field is perpendicular.

Corresponding to the polarizations, there are two basic topologies for 2d photonic crystals, as depicted in Figure 4(top): high index rods surrounded by low index (top) and low-index holes in high index (bottom). Here, a hexagonal lattice is used because, as noted earlier, it gives the largest gaps. Recall that a photonic band gap requires that the electric field lines run along thin veins: thus, the rods are best suited to TM light ($\vec{E}$ parallel to the rods), and the holes are best suited to TE light ($\vec{E}$ running around the holes). This preference is reflected in the band diagrams, shown in Figure 4, in which the rods/holes (top/bottom) have a strong TM/TE band gap. For these diagrams, the rod/hole radius is chosen to be $0.2a/0.3a$, where a is the lattice constant (the nearest-neighbor periodicity) and the high/low ε is 12/1. The TM/TE bandgaps are then 47%/28% as a fraction of mid-gap frequency, but these bandgaps require a minimum ε contrast of 1.7/1 and 1.9/1, respectively. Moreover, it is conventional to give the frequencies ω in units of $2\pi c/a$, which is equivalent to a/λ (λ being the vacuum wavelength) – Maxwell's equations are scale-invariant, and the same solutions can be applied to any wavelength simply by choosing the appropriate a. For example, the TM mid-gap ω in these units is 0.36, so if one wanted this to correspond to $\lambda = 1.55\ \mu\text{m}$ one would use $a = 0.36 \cdot 1.55\ \mu\text{m} = 0.56\ \mu\text{m}$.

The Brillouin zone (a hexagon) is shown at left-center, with the irreducible Brillouin zone shaded (following the sixfold symmetry of the crystal); the corners (high symmetry points) of this zone are given canonical names, where Γ always denotes the origin $\vec{k} = 0$, K is the nearest-neighbor direction,

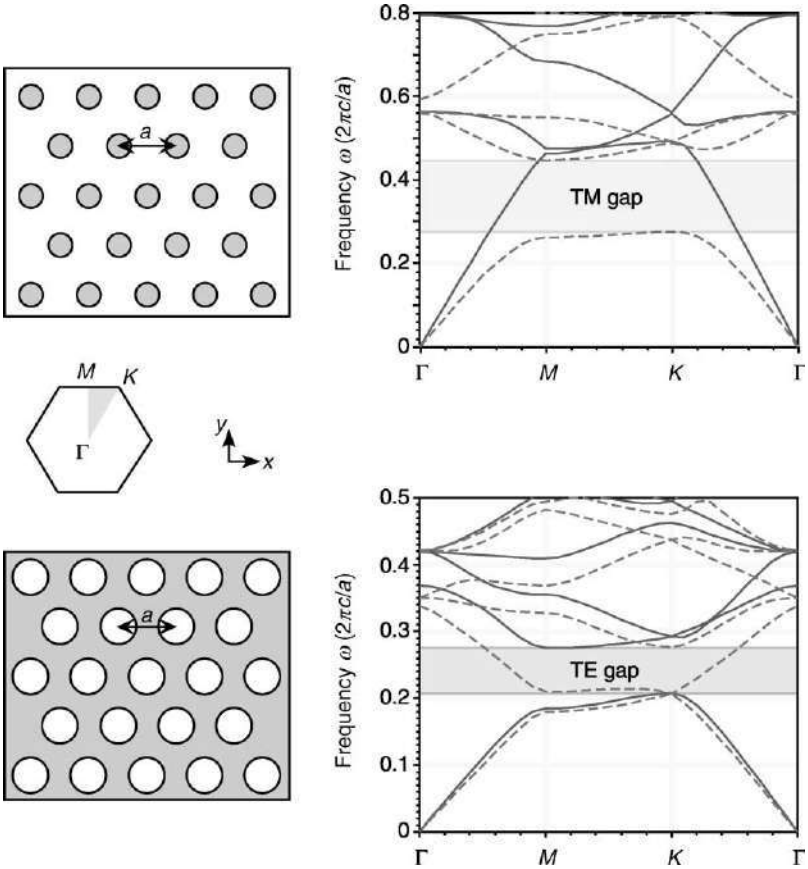

Figure 4 Band diagrams and photonic band gaps for hexagonal lattices of high dielectric rods ($\varepsilon = 12$, $r = 0.2a$) in air (top), and air holes ($r = 0.3a$) in dielectric (bottom), where a is the center–center periodicity. The frequencies are plotted around the boundary of the irreducible Brillouin zone (shaded triangle, left center), with solid/dashed lines denoting TE/TM polarization (electric field parallel/perpendicular to plane of periodicity). The rods/holes have a gap in the TM/TE bands.

and M is the next-nearest-neighbor direction. The Brillouin zone is a two-dimensional region of wavevectors, so the bands $\omega_n(\vec{k})$ are actually surfaces, but in practice the band extrema almost always occur along the boundaries of the irreducible zone (i.e., the high-symmetry directions). So, it is conventional to plot the bands only along these zone boundaries in order to identify the bandgap, as is done in Figure 4.

Actually, the hole lattice can display not only a TE gap, but a complete photonic bandgap (for both polarizations) if the holes are sufficiently large (nearly touching). In this case, the thin veins between nearest-neighbor holes induce a TE gap, while the interstices between triplets of holes form 'rod-like' regions that support a TM gap overlapping the TE gap.

Photonic-Crystal Slabs

In order to realize 2d photonic-crystal phenomena in three dimensions, the most straightforward design is to simply fabricate a 2d-periodic crystal with a finite height: a photonic-crystal slab, as depicted in Figure 5. Such a structure can confine light vertically within the slab via index guiding, a generalization of total internal reflection – this mechanism is the source of several new tradeoffs and behaviors of slab systems compared to their 2d analogs.

The key to index guiding is the fact that the 2d periodicity implies that the 2d Bloch wavevector $\vec{k}_\parallel$ is a conserved quantity, so the projected band structure – all states in the bulk substrate/superstrate (the uniform regions far below/above the slab) versus their in-plane wavevector component (projected wavevector) – creates a map of which states can radiate vertically. If the slab is suspended in air, for example, then the eigensolutions of the bulk air are $\omega = c\sqrt{|\vec{k}_\parallel|^2 + k_\perp^2}$, which when plotted versus $\vec{k}_\parallel$ forms the continuous light cone $\omega \geq c|\vec{k}_\parallel|$, shown as a shaded region in Figure 5. Because the slab has a higher ε (12) than the air (1), and frequency goes as $1/\sqrt{\varepsilon}$, discrete guided bands are 'pulled down' in frequency from this continuum – these bands, lying beneath the light cone, cannot couple to any vertically

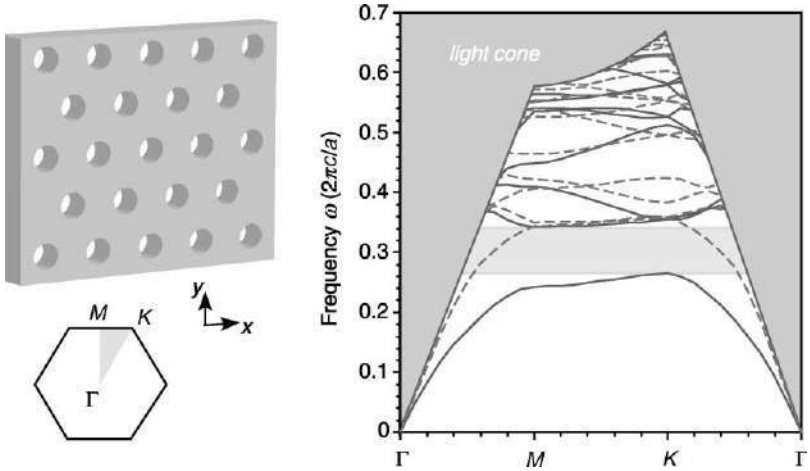

Figure 5 Projected band diagram for a finite-thickness (0.5*a*) slab of air holes in dielectric (cross section as in **Figure 4** bottom), with the irreducible Brillouin zone at lower left. The shaded region is the light cone: the projection of all states that can radiate in the air. Solid/dashed lines denote guided modes (confined to the slab) that are even/odd with respect to the horizontal mirror plane of the slab, whose polarization is TE-like/TM-like, respectively. There is a 'bandgap' (region without guided modes) in the TE-like guided modes only.

radiating mode by the conservation law and so are confined to the slab (exponentially decaying away from it). If the horizontal mid-plane of the slab is a mirror symmetry plane, then just as there were TM and TE states in 2d, here there are two categories of modes: even (TE-like) and odd (TM-like) modes under reflections through the mirror plane (which are purely TE/TM in the mirror plane itself). Because the slab here is based on the 2d hole crystal, which had a TE gap, here there is a 26% 'bandgap' in the even modes: a range of frequencies in which there are no guided modes. It is not a complete photonic bandgap, not only because of the odd modes, but also because there are radiating (light cone) modes at every ω. The presence of these radiating modes means that if all in-plane translational symmetry is broken by a localized change in the structure, say a waveguide bend or a resonant cavity, then vertical radiation losses are inevitable; there are various strategies to minimize the losses to tolerable levels, however. On the other hand, if only one direction of translational symmetry is broken, as in a linear-defect waveguide, ideally lossless guiding can be maintained.

Photonic-crystal slabs have two new critical parameters that influence the existence of a gap. First, it must have vertical mirror symmetry in order that the gaps in the even and odd modes be treatable separately – such symmetry is broken by the presence of a substrate that is not the mirror image of the superstrate, but in practice the symmetry breaking can be weak if the index contrast is sufficiently high (so that the modes are strongly confined in the slab). Second, the height of the slab must not be too small (or the modes will be weakly confined) or too large (or higher-order modes will fill the gap); the optimum height is around half a wavelength λ/n_{eff} (relative to an average/effective index n_{eff} that depends on the polarization). In **Figure 5**, a height of 0.5*a* is used, which is near the optimum (with holes of radius 0.3*a* and $\varepsilon = 12$ as in the previous section).

Three-Dimensional Photonic Crystals

Photonic-crystal slabs are one way of realizing 2d photonic-crystal effects in three dimensions; an example of another way, lifting the sacrifices imposed by the light cone, is depicted in **Figure 6**. This is a 3d-periodic crystal, formed by an alternating hole-slab/rod-slab sequence in an ABCABC stacking of bilayers – equivalently, it is an fcc (face-centered cubic) lattice of air cylinders in dielectric, stacked and oriented in the 111 direction, where each overlapping layer of cylinders forms a rod/hole bilayer simultaneously. Its band diagram is shown in **Figure 6** along the boundaries of its irreducible Brillouin zone (from a truncated octahedron, inset), and this structure has a >21% complete gap ($\Delta\omega$ as a fraction of mid-gap frequency) for $\varepsilon = 12/1$, forbidding light propagation for all wavevectors (directions) and all polarizations. Not only can this crystal confine light perfectly in 3d, but because its layers resemble 2d rod/hole crystals, it turns out that the confined modes created by defects in these layers strongly resemble the TM/TE states created by corresponding defects in two dimensions. One can therefore use this crystal to directly transfer designs from two to three dimensions while retaining omnidirectional confinement. Its fabrication, of

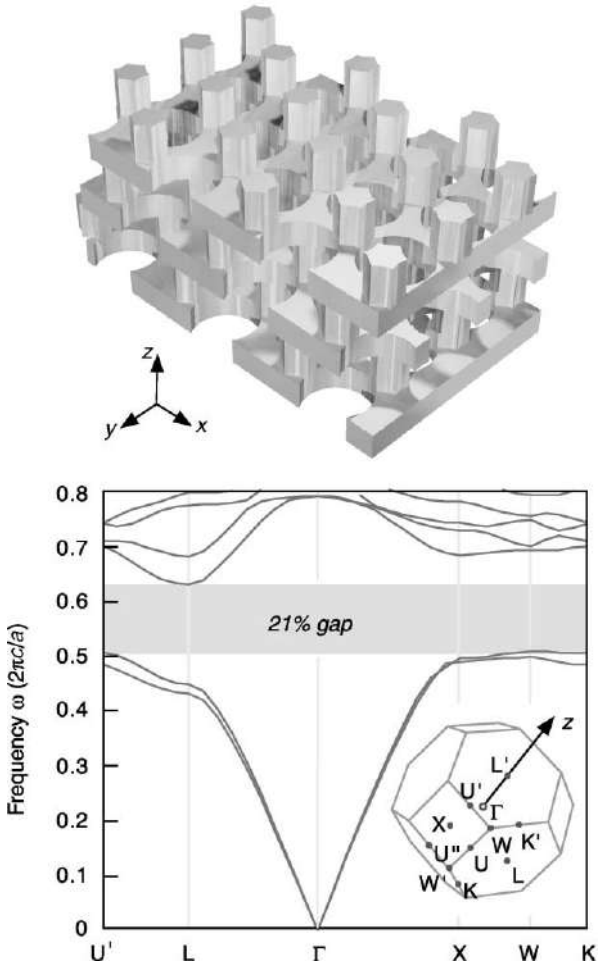

Figure 6 Band diagram (bottom) for 3d-periodic photonic crystal (top) consisting of an alternating stack of rod and hole 2d-periodic slabs (similar to Figure 4), with the corners of the irreducible Brillouin zone labeled in the inset. This structure exhibits a $\Delta\omega/\omega_{\text{midgap}} = 21\%$ omnidirectional bandgap.

course, is more complex than that of photonic-crystal slabs (with a minimum ε contrast of 4/1), but this and other 3d photonic crystal structures have been constructed even at micron (infrared) lengthscales, as described below.

There are three general dielectric topologies that have been identified to support complete 3d gaps for $\varepsilon = 12/1$ (e.g., Si:air) contrast: diamond-like arrangements of high dielectric 'atoms' surrounded by low dielectric, which can lead to $>20\%$ gaps between the 2nd and 3rd bands; fcc 'inverse opal' lattices of nearly-touching low dielectric spheres (or similar) surrounded by high dielectric, giving gaps around 10% between the 9th and 10th bands; and cubic 'scaffold' lattices of rods along the cube edges, giving $\sim$7% gaps between the 2nd and 3rd bands. It is notable that the first two topologies correspond to fcc lattices, which have the most nearly spherical Brillouin zones in accordance with the rules of thumb given above. Many variations on these topologies continue to be proposed – for example, the structure of Figure 6 is diamond/graphite-like – mainly in conjunction with different fabrication strategies, such as the following three successful approaches. First, layer-by-layer fabrication, in which individual crystal layers (typically of constant cross-section) are deposited one-by-one and etched with a 2d pattern via standard lithographic methods (giving fine control over placement of defects, etc.); Figure 6 can be constructed in this fashion (as well as other diamond-like structures with large gaps). Second, colloidal self-assembly, in which small dielectric spheres in a fluid automatically arrange themselves into close-packed (fcc) crystals by surface forces – these crystals can be back-filled with a high-index material, out of which the original spheres are dissolved in order to form inverse–opal crystals with a complete gap. Third, holographic lithography, in which a variety of 3d crystals can be formed by an interference pattern of four laser beams to harden a light-sensitive resin (which is then back-filled and dissolved, as with colloids, to achieve the requisite index contrast). The second and third techniques are notable for their ability to construct large-scale 3d crystals (thousands of periods) in a short time.

See also

Photonic Crystals: Atomic Physics; Photonic Crystal Lasers, Cavities and Waveguide.

Further Reading

Ashcroft NW and Mermin ND (1976) *Solid State Physics*. Philadelphia: Holt Saunders.

Cohen-Tannoudji C, Din B and Laloë F (1977) *Quantum Mechanics*. Paris: Hermann.

Joannopoulos JD, Meade RD and Winn JN (1995) *Photonic Crystals: Molding the Flow of Light*. Princeton: Princeton University Press.

Johnson SG and Joannopoulos JD (2002) *Photonic Crystals: The Road from Theory to Practice*. Boston: Kluwer.

Johnson SG, Ibanescu M, Skorobogatiy M, Weisberg O, Joannopoulos JD and Fink Y (2002) Perturbation theory for Maxwell's equations with shifting material boundaries. *Physical Review E* 65: 066611.

Sakoda K (2001) *Optical Properties of Photonic Crystals*. Berlin: Springer.

Sipe JE (2000) Vector $k \cdot p$ approach for photonic band structures. *Physical Review E* 62: 5672–5677.

Microwave Photonic Crystals

D F Sievenpiper, HRL Laboratories LLC, Malibu, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Photonic Crystals

Photonic crystals are materials that have a periodically varying dielectric constant that results in a bandgap, or range of frequencies where electromagnetic waves are unable to propagate. A simple one-dimensional example is the Bragg mirror, or quarter-wave stack, in which periodic layers of alternating dielectrics reflect photons within a frequency band determined by the thickness and refractive index of the layers. Photonic crystals are the extension of this concept into two or three dimensions. They are constructed as a lattice of two different dielectrics, and examples include a triangular lattice of circular rods, or a diamond lattice of spheres. Certain geometries will produce a bandgap, within which propagation is impossible in any direction. These materials are the photonic analog of semiconductors, in that their periodic dielectric function provides a bandgap for photons in much the same way that the periodic atomic potential of a crystal lattice can provide a bandgap for electrons.

While most photonic crystal research has focused on optical wavelength structures, many of the first experimental studies were performed on scale models at microwave frequencies, where fabrication and measurements were simpler. Operating at microwave frequencies also allowed a new class of materials to emerge, based on metals and lumped circuit elements. The nearly infinite dielectric contrast provided by the use of metals, and the ability to make complex subwavelength scale structures at radio frequencies, allows photonic crystals with ultra-wide bandgaps to be produced. Furthermore, the use of lumped circuit elements allows the electromagnetic properties to be easily tuned, resulting in unique microwave structures.

Wire Mesh Photonic Crystals

At optical frequencies, conductive losses preclude the use of metals for many applications. However, at microwave frequencies, metals are nearly ideal materials, since they have low loss and are easily formed into complicated geometries. The simplest metallic photonic crystal is a three-dimensional wire mesh, shown in Figure 1. This example resembles a diamond crystal lattice, in which metal wires have been substituted for atomic bonds. It can be constructed using zig-zag-shaped metal strips that fit together to form a continuous conductive network.

The electromagnetic properties of this wire mesh can be understood by considering a traditional microwave analog: an array of metal waveguides. Open tunnels permeate the mesh in various directions, and it behaves as an interpenetrating fabric of waveguides, whose walls are defined by the wires. Shorter wavelengths, that fit between the wires, can propagate through the crystal, while longer waves see the mesh as a continuous metal surface, and are reflected. As for a conventional waveguide, the cutoff frequency occurs approximately where one half wavelength fits between the rows of wires. Similarly, for a diamond lattice of wires, the cutoff frequency is

$$\omega_{\text{cutoff}} = \frac{\pi c}{na} \qquad [1]$$

where n is the refractive index of the material between the wires, and a is the length of the unit cube of the diamond lattice. Because the cutoff frequency is primarily determined by the diameter of the open tunnels, and it is largely independent of the detailed arrangement of the wires, other mesh geometries can

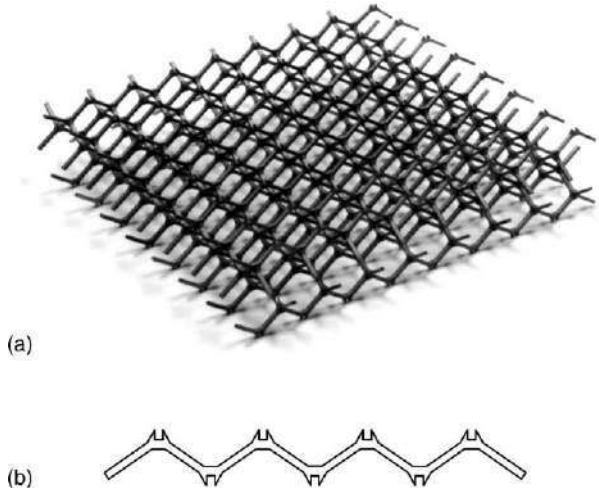

Figure 1 A wire mesh photonic crystal with diamond geometry. (a) The mesh forms continuous conductive structure that is permeated with open tunnels. (b) A diagram of one of the interlocking metal parts that fit together to form the crystal. (Reproduced with permission from Sievenpiper D, Sickmiller M and Yablonovitch E (1996) 3D wire mesh photonic crystals. *Physics Review Letters* 76: 2480. Copyright (1996) by the American Physical Society.)

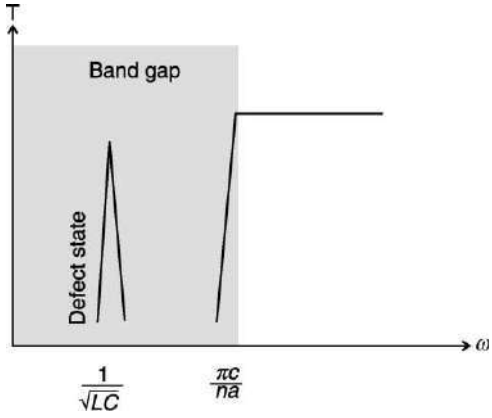

Figure 2 The transmission magnitude through a conductive wire mesh photonic crystal as a function of frequency, ω. Shorter wavelengths can propagate through the open tunnels, while longer wavelengths see the mesh as a continuous block of metal. The cutoff frequency is determined largely by the lattice constant, *a*, and the refractive index, *n*, of the material between the wires. A cut wire forms a defect state within the band gap.

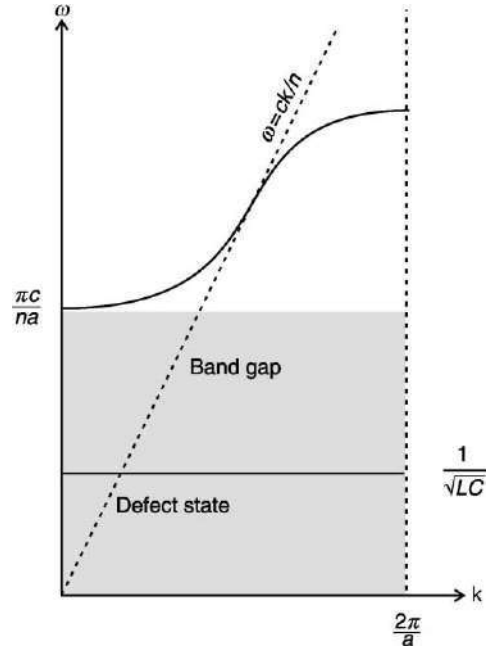

Figure 3 The dispersion diagram for a wire mesh photonic crystal, plotted as frequency, ω, versus wave vector k. The bandgap extends to zero frequency. The lower bandedge of the allowed band is determined largely by the lattice constant. Like other photonic crystals, the wire mesh also supports higher frequency bands, which are not shown. A point defect such as a cut wire creates a mode at a single frequency, which contains all wave vectors.

be expected to give similar results. However, the cutoff frequency can vary by a constant of order unity, depending on the lattice and the unit cell chosen.

The transmission magnitude, as a function of frequency for such a mesh, is depicted in Figure 2. Experimentally, it is found that a typical wire mesh structure attenuates roughly 6 dB per layer for frequencies below cutoff. However, because the exact transmission magnitude above cutoff depends on details of the measurement, such as impedance matching at the crystal boundary, the vertical scale shown here is arbitrary, and the intent is simply to give the qualitative behavior of these and similar crystal structures. Waves just below the cutoff frequency can penetrate a crystal with finite thickness, as depicted by the finite slope of the transmission edge.

We can better understand the behavior of waves within the crystal by examining the dispersion diagram, shown in Figure 3, in which we plot the allowed frequencies as a function of wave vector $k = 2\pi/\lambda$, the spatial frequency of waves inside the material. A band of allowed frequencies is bounded by upper and lower band edges, corresponding to high and low cutoff frequencies, beyond which propagation is forbidden. The reflective nature of the crystal below cutoff is shown as a bandgap that extends down to zero frequency. The slope of the dispersion curve corresponds to the group velocity of waves within the allowed band, and states at the bandedges represent standing waves. The upper edge of this band occurs where the wavelength matches the periodicity of the crystal, and for a diamond mesh, it is approximately

$$\omega_{\text{lattice}} = \frac{2\pi c}{na} \qquad [2]$$

Like traditional dielectric photonic crystals, these wire mesh structures also have higher order bands and gaps that occur at higher frequencies. The nature of the higher band structure depends on the detailed geometry of the mesh to a much greater extent than the lowest order gap, which is our primary interest for radio frequency materials.

Defect States

As in semiconductor crystals, we can alter the behavior of photonic crystals by adding defects to create localized states within the bandgap. For the wire mesh photonic crystal, a defect may be a cut wire, or a lumped circuit element inserted into the mesh. For example, replacing one of the wires with a capacitor creates a parallel resonant circuit, which acts as a small electromagnetic cavity within the mesh. Its resonance frequency is given by

$$\omega_{\text{defect}} = \frac{1}{\sqrt{LC}} \qquad [3]$$

where C is the capacitance of the defect, and L is the inductance of the surrounding wires. The inductance depends logarithmically on the wire diameter, so it is largely fixed for a given mesh lattice. However, the capacitance can be easily changed, since it is a simple lumped circuit element. Thus, by adding capacitive defects we can create tunable local states within the bandgap. The presence of a defect is seen as a narrow transmission band within the bandgap, shown in Figure 2. Because a point defect contains all wave vectors, it is shown as a horizontal line in the dispersion diagram in Figure 3.

If two defects are located in close proximity, their modes overlap, and they interact with each other. For example, two neighboring defects with identical capacitance will exhibit a split pair of resonances, corresponding to a lower frequency symmetric mode, and a higher frequency anti-symmetric mode. The modes are highly localized at the defects, and this interaction is limited to defects that are within one or two lattice constants of each other. If one member of a pair of nearby defects is tuned with respect to the other, the frequencies the two modes exhibit are an avoided crossing behavior, as is typical in many systems of coupled oscillators. The resonance frequencies for more complex arrangements of defects can be calculated by assigning an interaction energy to each neighboring pair, and solving the resulting Hamiltonian matrix to find the energy levels of the complete system. In this way, the interaction among defects in small clusters is analogous to atoms in simple molecules, and the resulting energy level structure is similar.

As more defects are added periodically throughout the mesh, they form a superlattice, and their states merge into an additional band within the original bandgap. The upper frequency of this band depends on the LC resonance frequency of the defects. The lower frequency depends on the density of the remaining wires, since waves at the lower bandedge must be able to fit through the open regions in the mesh. In the extreme case, where every wire is cut and replaced by a capacitor, the mesh is no longer a continuous conductive network. These structures have an allowed band that extends to zero frequency, and they will be explored in greater detail in the next section. Not every wire must be cut to reach this limit, and the conditions for this to occur are the subject of percolation theory.

Capacitive Mesh Photonic Crystals

In a capacitive mesh photonic crystal, small isolated metal islands are distributed on a lattice, and they are capacitively coupled to their neighbors. This represents the extreme case of the wire mesh photonic crystal, where every wire has been cut, and replaced by a capacitor. An example of such a structure, built as a stack of printed circuit boards, is shown in Figure 4. A diamond lattice can be approximated using planar circuit boards, with printed tetrapodal structures consisting of two orthogonal metal strips connected by a metal plated via. Each metallic island couples to its four nearest neighbors through their overlapping capacitor plates.

The capacitive mesh photonic crystal is transmissive at low frequencies because it is not traversed by conducting paths, like the continuous wire mesh version. The crystal supports a lower band that extends from zero frequency, up to an LC resonance frequency:

$$\omega_{\text{lower}} = \frac{1}{\sqrt{LC}} \qquad [4]$$

which marks the lower edge of the bandgap. This is determined by the capacitance between the metal islands, and the inductance associated with each island. At higher frequencies, the capacitors appear as shorts, and the crystal behaves similarly to the conductive wire mesh structure described previously, with an upper edge given by

$$\omega_{\text{upper}} = \frac{\pi c}{na} \qquad [5]$$

At this frequency, the waves can again fit through the open areas between the metallic islands. The transmission magnitude, as a function of frequency for the capacitive mesh photonic crystal, is shown in Figure 5, and the additional band is illustrated in the dispersion diagram in Figure 6. Like the conductive wire mesh, the capacitive mesh also has a rich upper band structure, that is outside the scope of this article.

With these capacitive mesh structures, it is possible to create materials with finite, but arbitrarily large bandgaps, because the upper and lower edges of the bandgap are controlled independently by the lattice constant and the capacitance, respectively. In conventional dielectric photonic crystals, the width of the bandgap depends on the dielectric contrast between the constituent materials. Waves at the upper edge primarily reside in the lower dielectric material, and waves at the lower edge inhabit the higher dielectric material. This reasoning can also be applied to capacitive mesh structures. The lower dielectric is the substrate material that fills the open spaces between the metal islands, and modes at the upper edge of the gap primarily fill these open spaces. However, for modes at the lower edge of the gap, the electric field is primarily concentrated within the capacitors. The metal regions and the capacitors can

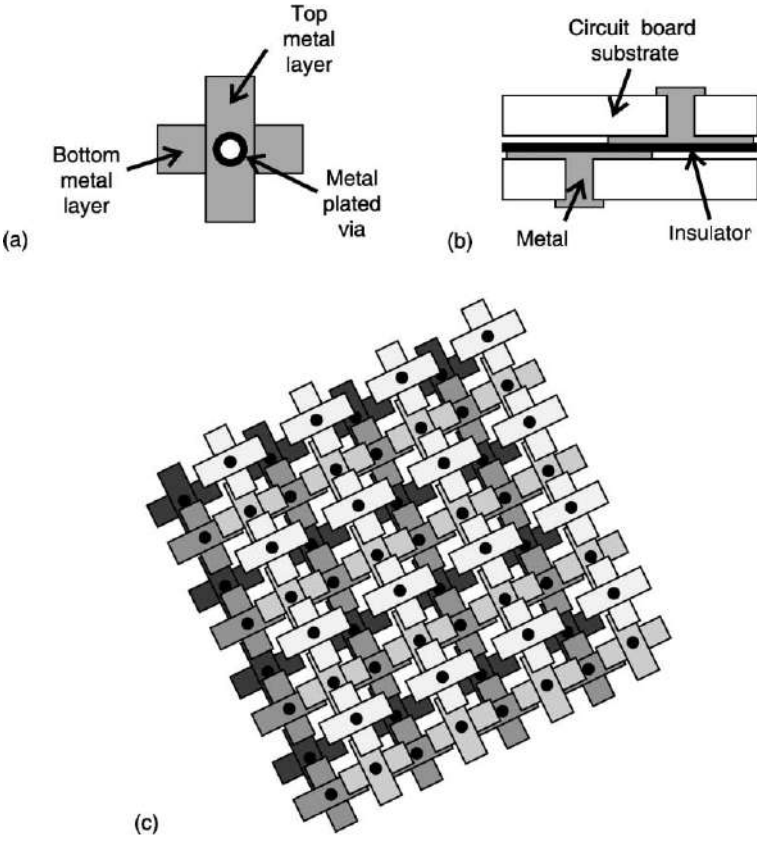

Figure 4 A capacitive mesh photonic crystal, built as a diamond lattice of capacitively coupled metallic islands. It can be built by stacking printed circuit boards. (a) One 'atom' of the crystal, which is made of two small orthogonal metal strips connected by a metal-plated via. (b) A side view, showing how the metal strips overlap to form a capacitor. (c) Several periods of the completed crystal structure, shown without the substrate. (Reproduced with permission from Sievenpiper D, Yablonovitch E, Winn J, et al. (1998) 3D metallo-dielectric photonic crystals with strong capacitive coupling between metallic islands. Physics Review Lett. 80: 2829. Copyright (1998) by the American Physical Society.)

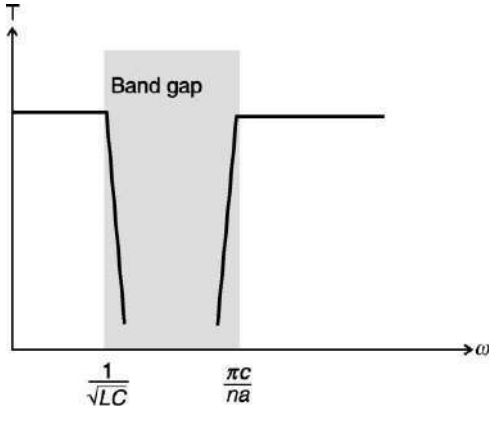

Figure 5 The transmission magnitude through a capacitive mesh photonic crystal, as a function of frequency. The crystal is transparent at low frequencies, because the conducting paths are broken by capacitors. At higher frequencies, the capacitors appear as shorts, and crystal behaves like a solid wire mesh.

be considered as an effective medium, or artificial dielectric whose properties depend on the density, arrangement, and values of the capacitors. By using large capacitors, the dielectric contrast can be made arbitrarily large, and the bandgap arbitrarily wide.

Within the bandgap, these materials can support surface waves, which are modes that are bound to the interface between the crystal and free space. Surface waves propagate along the crystal surface, and they have their own band structure. In the following sections, we examine these waves, and we will introduce periodically textured surfaces that suppress them, much like three-dimensional photonic crystals suppress internal plane waves.

Surface Waves

Electromagnetic waves that are bound to the interface between two materials are known as surface waves. They exist on ordinary metals, and at optical frequencies they are often called surface plasmons. They also exist on photonic crystals within the

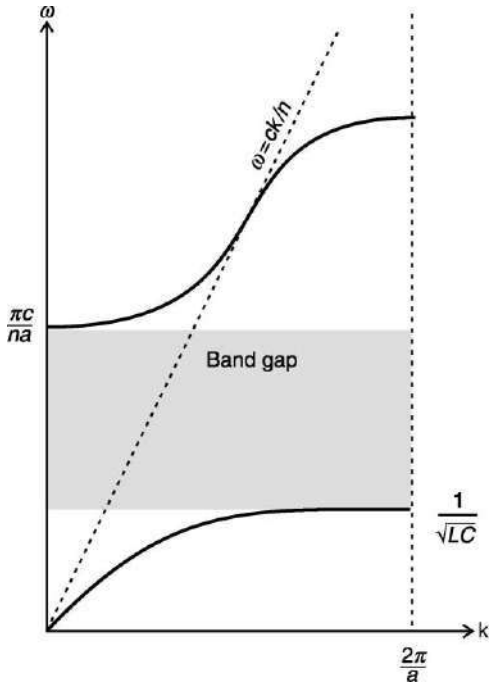

Figure 6 The dispersion diagram of the capacitive mesh photonic crystal. The two edges of the bandgap can be adjusted independently by varying the lattice constant and the capacitance.

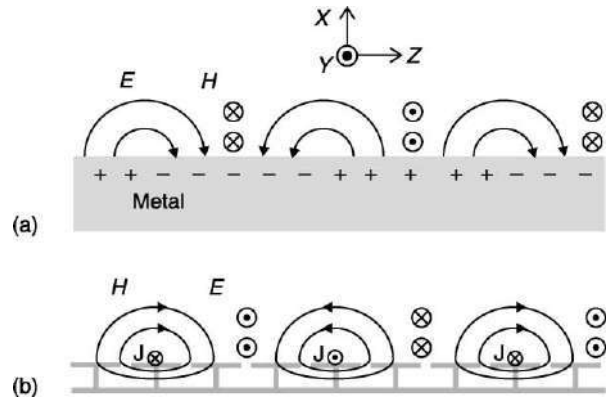

Figure 7 Surface waves can be classified into two polarizations. (a) A TM wave, shown on a metal surface; (b) A TE wave, shown on a textured surface. TM waves consist of electric charges on the metal surface that oscillate along the direction of propagation. They have nonzero electric field components both perpendicular to the surface, and parallel to the direction of propagation. TE waves are associated with transverse currents in the surface. They are similar to TM waves, but with the electric and magnetic fields exchanged.

bandgap region, as well as on other impenetrable materials. They cannot propagate into the crystal because of Bragg reflection, yet they are prevented from escaping into free space by total internal reflection. They typically interact strongly with the top layer of the crystal, and they can often be manipulated or suppressed by adjusting the geometry of that layer. At radio and microwave frequencies, these waves are usually called surface currents, because the associated fields are only loosely bound to the surface, and extend a great distance into free space. Their presence can be seen in antenna radiation patterns and other phenomena, and by suppressing them we can make a variety of unique radio frequency devices. However, we must first understand the nature of these waves.

To understand the nature of these surface currents, we begin by examining surface waves on metals, where we assume a wave confined to a dielectric interface, and then insert the complex dielectric constant of a metal. Consider a surface in the YZ plane, where $X < 0$ is filled with dielectric ε, and $X > 0$ is filled with vacuum. A wave propagates in the Z direction with propagation constant k. For transverse magnetic (TM) polarized waves, which are illustrated in Figure 7, the transverse electric field $E_y = 0$. The electric field has the form

$$E_1 = (\hat{x}E_{1x} + \hat{z}E_{1z})e^{i\omega x - ikz - \alpha x} \qquad [6]$$

for $X > 0$ and

$$E_2 = (\hat{x}E_{2x} + \hat{z}E_{2z})e^{i\omega x - ikz + \gamma x} \qquad [7]$$

for $X < 0$, where α and γ are the attenuation constants into free space and dielectric, respectively. The fields of such a wave are shown in Figure 7. We can solve the wave equation for the electric field:

$$\nabla \times \nabla \times \vec{E} = \varepsilon_r \frac{\omega^2}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} \qquad [8]$$

by matching E_x and $\varepsilon_r E_z$, at the interface to yield

$$k = \sqrt{\frac{\varepsilon}{1+\varepsilon}} \frac{\omega}{c} \qquad [9]$$

$$\alpha = \sqrt{\frac{-1}{1+\varepsilon}} \frac{\omega}{c} \qquad [10]$$

$$\gamma = \sqrt{\frac{-\varepsilon^2}{1+\varepsilon}} \frac{\omega}{c} \qquad [11]$$

A solution for transverse electric (TE) waves, where $H_y = 0$, can be found by applying the principle of duality, by exchanging the electric and magnetic fields, and substituting μ for ε.

For surface waves on a metal surface, we insert the dielectric constant of a metal, which is

$$\varepsilon = 1 - \frac{i\sigma}{\omega\varepsilon_0} \qquad [12]$$

The conductivity, λ, is

$$\sigma = \frac{nq^2\tau/m}{1 + i\omega\tau} \qquad [13]$$

where τ is the mean electron collision time, q is the electron charge, m and n are the effective mass and the density, respectively, of the conduction electrons. For frequencies much lower than $1/\tau$, which includes microwave and radio frequencies, we can substitute eqns [12] and [13] into eqn [9] to find

$$k \approx \frac{\omega}{c} \qquad [14]$$

Thus, surface waves on metals propagate at nearly the speed of light in vacuum, and they travel for many wavelengths with little attenuation. We can also solve for the decay constants into free space using eqn [10]:

$$\alpha = (1 - i)\frac{\omega}{c}\sqrt{\frac{\omega\varepsilon_0}{2\sigma}} \qquad [15]$$

and into the metal using eqn [11]:

$$\gamma \approx (1 + i)\sqrt{\frac{\omega\mu_0\sigma}{2}} = \frac{(1 + i)}{\delta} \qquad [16]$$

where δ is recognized as the skin depth. Thus, we have derived the skin depth of a metal from the penetration depth of surface waves.

From the skin depth, we can derive the impedance of a metal sheet. Assuming that E_0 is the electric field at the surface, the surface current is given by

$$J_z(x) = \sigma E_z(x) = \sigma E_0 e^{-x(1+i)/\delta} \qquad [17]$$

The magnetic field at the surface is found by integrating along a path that surrounds the surface current:

$$H_0 = \int_0^{-\infty} J_z(x)\mathrm{d}x = \frac{\sigma\delta}{1 + i}E_0 \qquad [18]$$

The surface impedance of the metal is therefore

$$Z_s = \frac{E_z}{H_y} = \frac{1 + i}{\sigma\delta} \qquad [19]$$

We can see that the resistance of a metal surface is accompanied by an equal amount of inductance, because the real and imaginary parts of the surface impedance are equal, and have the same sign.

The part of the wave that extends into vacuum is very loosely bound to the metal surface, but the currents in the metal occur in a well-defined skin depth. For example, copper has a surface impedance of $0.03(1 + i)$ ohms/square at 10 GHz. Using eqns [15] and [16], a surface wave at that frequency extends about 70 meters, or 2300 wavelengths into free space, but less than 1 micron into the metal. For this reason, surface waves are often described simply as surface currents at microwave and radio frequencies, although they are the same phenomenon as the more tightly bound surface plasmons that occur at optical frequencies. Nonetheless, their presence can be seen in a variety of phenomena at RF frequencies, such as antenna radiation patterns.

The above discussion can be generalized to a simple impedance surface, and it can be shown that TM waves are supported on inductive surfaces having an impedance of

$$Z_s(\mathrm{TM}) = \frac{i\alpha}{\omega\varepsilon} \qquad [20]$$

while TE waves are supported on capacitive surfaces, having an impedance of

$$Z_s(\mathrm{TE}) = \frac{-i\omega\mu}{\alpha} \qquad [21]$$

Thus, while ordinary metals, which are inductive, support TM waves, they do not support TE waves, which require a capacitive surface.

Surface waves on metals can be suppressed over a finite frequency band by applying various kinds of periotic textures. For example, at optical frequencies, a lattice of small bumps on a metal surface can create a surface wave bandgap. Surface waves are scattered by the bumps, and form standing waves when one-half wavelength fits between the rows of bumps. Two modes occur, in which the wave nodes lie either on the bumps or between them. These two modes have slightly different frequencies, separated by a narrow bandgap.

At microwave frequencies, more elaborate structures can be built to control the surface properties. One example is a corrugated metal surface, shown in Figure 8. The corrugations can be considered as transmission lines running into the slab, and shorted at the bottom. They are typically one-quarter wavelength deep, so that the short circuit at the bottom is transformed into an open circuit at the top. Because of these resonant structures built into the surface, it has high impedance for waves that are polarized with the electric field transverse to the corrugations, but it has low impedance for waves that are polarized in the orthogonal direction, with the electric field parallel to the metal ridges. For waves propagating transverse to or parallel to the corrugations, the surface can be described as hard or soft, respectively, by analogy to the acoustic surfaces of the same names. Both of these boundary conditions are used in antennas to control scattering properties or

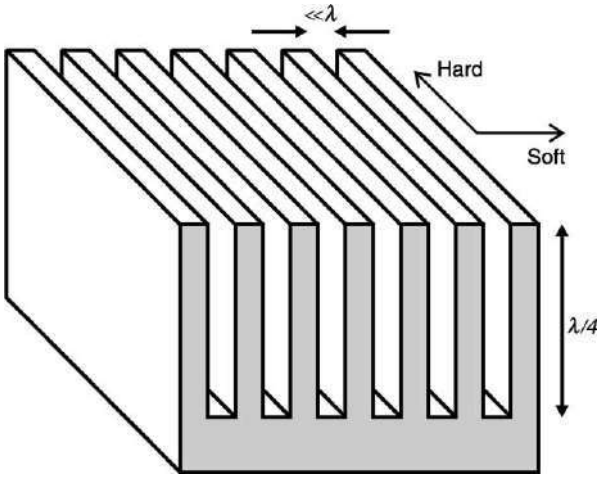

Figure 8 A corrugated metal surface containing quarter-wave resonant grooves presents a soft boundary condition to waves propagating transverse to the grooves, and a hard boundary condition to waves propagating along the grooves.

Table 1 Four basic boundary conditions in electromagnetics, and the surface impedance for each polarization

Boundary condition	$Z_s(TM)$	$Z_s(TE)$
Electric conductor	0	0
Soft surface	∞	0
Hard surface	0	∞
Magnetic conductor	∞	∞

field profiles. These two surfaces, along with the electric and magnetic conductors, describe four basic boundary conditions for electromagnetics, which are summarized in Table 1. Like the soft and hard surfaces, the magnetic conductor does not exist in nature, but it can be produced artificially over a finite frequency band, using a kind of two-dimensional photonic crystal known as a high-impedance surface.

High-Impedance Surfaces

Just as three-dimensional photonic crystals forbid internal plane waves, two-dimensional photonic crystals can be designed to suppress surface waves. Like the capacitive mesh described earlier, they can be constructed with lumped circuit elements. They can be understood as corrugated surfaces, in which the resonant corrugations have been folded up into resonant LC circuits, and distributed on a two-dimensional lattice. An example of such a structure is shown in Figure 9. It can be built using printed circuit boards, in which flat metal patches are connected to a common ground plane by metal plated vias.

When the plates are small compared to the wavelength, we may treat the surface as an effective

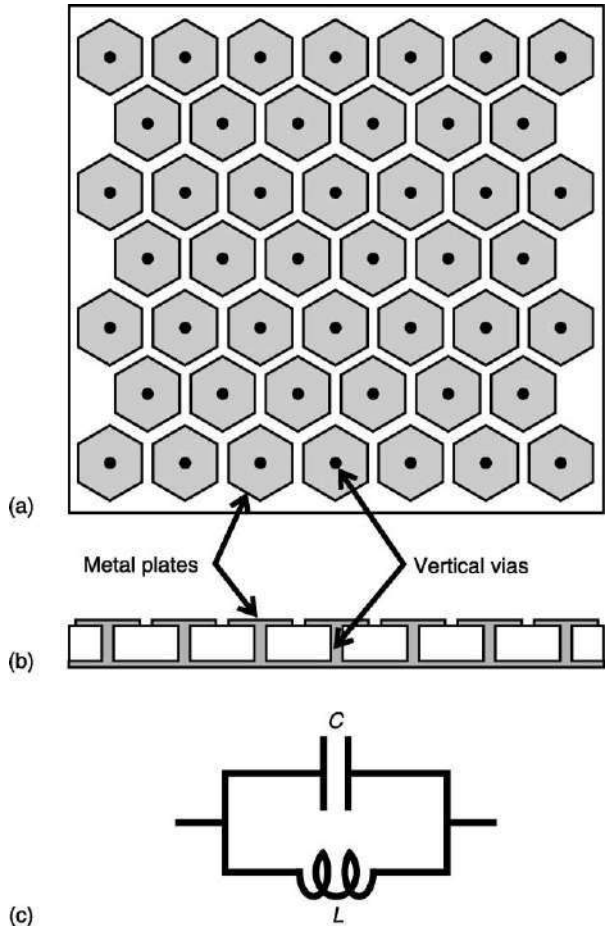

Figure 9 A high-impedance surface, constructed as a printed circuit board. (a) The top view shows an array of hexagonal metal patches; (b) the side view shows metal-plated vias that connect the patches to the ground plane; (c) the effective surface impedance can be modeled as a parallel resonant LC circuit.

medium, with a surface impedance defined by a parallel resonant LC circuit. The proximity of the neighboring metal plates provides capacitance, while the current paths, through the vias and the ground plane, provide inductance. The surface impedance is given by

$$Z_s = \frac{i\omega L}{1 - \omega^2 LC} \qquad [22]$$

and the resonance frequency is

$$\omega_0 = \frac{1}{\sqrt{LC}} \qquad [23]$$

The sheet capacitance, C, is equal to the value of the individual capacitors, multiplied by a geometrical factor that is related to their arrangement. It can be shown that for nonmagnetic materials, the sheet

inductance L is equal to

$$L = \mu_0 t \qquad [24]$$

where t is the total thickness of the high-impedance surface. The surface is inductive for frequencies below resonance, and capacitive for frequencies above resonance. Near the resonance frequency, the magnitude of the surface impedance is very high ($Z_s >> \eta$, where $\eta = 377\Omega$ is the impedance of free space).

For a normally incident wave, the reflection coefficient of the surface is

$$R = \frac{Z_s - \eta}{Z_s + \eta} \qquad [25]$$

The phase $\phi = \mathrm{Im}(Ln(R))$ of the reflection coefficient is plotted as a function of frequency in Figure 10. While a low impedance surface such as a conventional electric conductor reflects with a phase shift of π, a high-impedance surface or artificial magnetic conductor reflects with a phase shift of 0. This can be used to build low-profile antennas, because radiating elements can lie very near to a high-impedance ground plane without being shorted out. For example, a thin wire antenna can be placed within a small fraction of a wavelength of the high-impedance surface, yet retain high radiation efficiency. While image currents in a metal ground plane would normally oppose the antenna current, the currents in a high-impedance ground plane reinforce the antenna current because of the opposite reflection phase. This effect occurs when the reflection phase falls between $\pi/2$ and $-\pi/2$, and the corresponding frequency range defines the practical bandwidth of a low-profile antenna on such a surface.

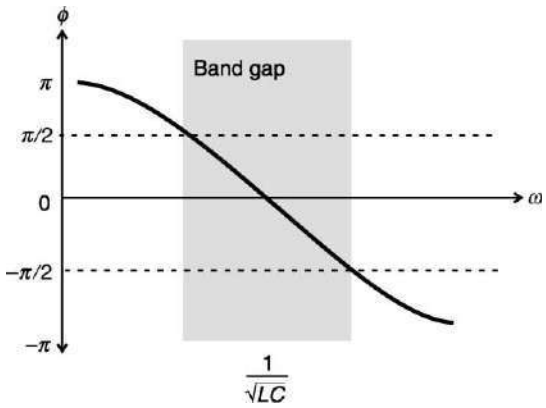

Figure 10 The reflection phase of a high-impedance surface crosses through 0 at the resonance frequency, where the surface appears as an artificial magnetic conductor. The reflection phase falls between $\pi/2$ and $-\pi/2$ for frequencies corresponding roughly to the edges of the surface wave bandgap.

Experimentally this bandwidth is found to be roughly equal to the surface wave bandgap, to be described later.

The reflection phase properties of these materials can also be used to build unique microwave reflectors. Since the phase depends on the resonance frequency, one can vary the local resonance frequency in order to tune the reflection phase as a function of position on the surface. For example, a linear phase gradient can be used for beam steering, or a parabolic phase function can be used for focusing. The resonance frequency can be adjusted by tuning the capacitance of the individual unit cells, using varactors or mechanical motion, to build actively controlled microwave reflectors.

The other important property of these textured ground planes is their surface wave bandgap, which can be used to control the radiation pattern of antennas near the surface. We examine the surface wave properties by considering a wave that decays into free space, as in eqn [6]. From Maxwell's equations, we can obtain

$$k^2 = \mu_0 \varepsilon_0 \omega^2 + \alpha^2 \qquad [26]$$

which can be combined with eqs [20] and [21] to obtain the dispersion relations for TM and TE waves on a general impedance surface, respectively. For TM waves, which occur where $\omega < \omega_0$, we have

$$k(\mathrm{TM}) = \frac{\omega}{c}\sqrt{1 - \frac{Z_s^2}{\eta^2}} \qquad [27]$$

and for TE waves, which occur where $\omega > \omega_0$ we have

$$k(\mathrm{TE}) = \frac{\omega}{c}\sqrt{1 - \frac{\eta^2}{Z_s^2}} \qquad [28]$$

These functions describe, to first order, the dispersion of waves on a high-impedance surface. However, they do not predict the bandgap, which can only be obtained by including the effects of radiation damping.

Consider an oscillating sheet of current that sees the impedance of the textured surface on one side, and the impedance of free space on the other side. At very low, or very high frequencies, the current is shunted through the surface. Near the resonance frequency, where the surface impedance is high, power is delivered to free space through radiation damping. The half power bandwidth of this radiation is defined by the frequencies where the magnitude of the surface impedance is equal to the impedance of

free space:

$$\left|\frac{i\omega L}{1-\omega^2 LC}\right| = \eta \qquad [29]$$

We solve for λ to yield the frequencies of the two bandedges:

$$\omega^2 = \frac{1}{LC} + \frac{1}{2\eta^2 C^2} \pm \frac{1}{\eta C}\sqrt{\frac{1}{LC} + \frac{1}{4\eta^2 C^2}} \qquad [30]$$

The terms in $1/\eta^2 C^2$ are typically small compared to the terms in $1/LC$, so we neglect them to obtain

$$\omega \approx \omega_0 \sqrt{1 \pm \frac{Z_0}{\eta}} \qquad [31]$$

where

$$Z_0 = \sqrt{\frac{L}{C}} \qquad [32]$$

is the characteristic impedance of the surface. For structures that are thin compared to the wavelength, the characteristic impedance is small compared to η. We can expand the square root in eqn [31], to approximate the fractional bandwidth of the radiation as

$$\frac{\Delta\omega}{\omega_0} \approx \frac{Z_0}{\eta} = \frac{\sqrt{L/C}}{\sqrt{\mu_0/\varepsilon_0}} \qquad [33]$$

Using eqn [23], this can be rearranged to obtain

$$\frac{\Delta\omega}{\omega_0} \approx t\frac{2\pi}{\lambda_0} \qquad [34]$$

where λ_0 is the free space wavelength at the resonance frequency. This defines a surface wave bandgap that is centered at the resonance frequency, between a lower-frequency TM band and a higher-frequency TE band. From eqn [34], we see that the bandwidth of a high-impedance surface, or of an antenna on that surface, is limited by its thickness with respect to its free space wavelength at resonance. The term $\lambda_0/2\pi$ is recognized as the radian length, described by Wheeler, who showed that the fractional bandwidth limit of a small resonant antenna is given by its volume with respect to a notional sphere of this radius. Equation [34] represents the equivalent bandwidth limit for a two-dimensional resonant surface.

The surface wave bandgap can be measured experimentally with a pair of small probes placed near the surface. A surface wave probe can be constructed from the end of a coaxial cable with a short section of the inner conductor exposed. Fields from the probe excite surface waves when they are placed near the surface, and the transmission magnitude between the probes can be used to analyze the surface wave band structure. For a high impedance surface, the bandgap can be seen as a range of frequencies with a significant reduction in surface wave transmission magnitude, shown in Figure 11. Within the bandgap, waves are not bound to the surface, and currents on the surface readily radiate energy into free space. The upper edge of the bandgap

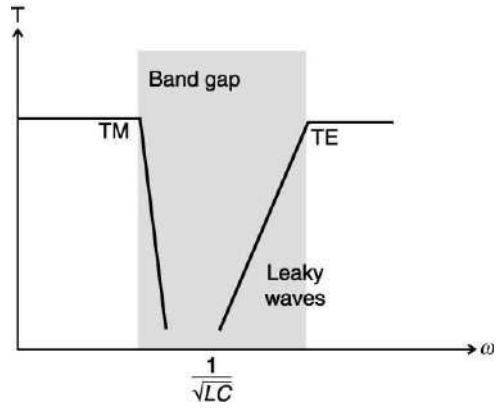

Figure 11 The surface wave transmission magnitude as a function of frequency for a high-impedance surface. TM waves are supported at low frequencies, and TE waves are supported at high frequencies. Leaky TE waves, which are damped by radiation, can occur within the bandgap.

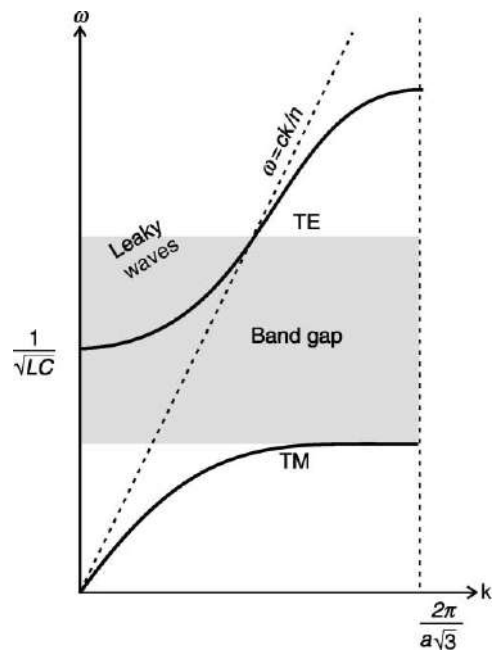

Figure 12 The dispersion diagram for surface waves on a high-impedance surface. The bandgap spans from the upper edge of the TM band, to the point where the TE band crosses the light line. Within the bandgap, electromagnetic waves are not bound to the surface. Modes to the left of the light line represent leaky waves.

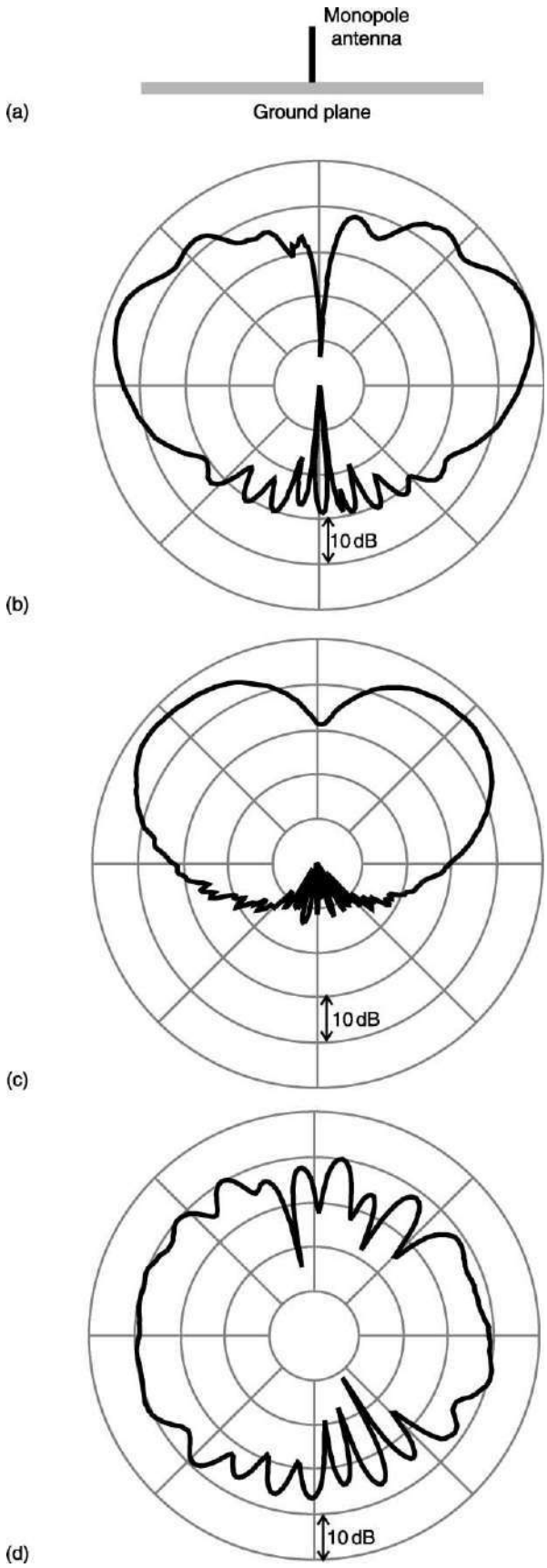

is typically less abrupt than the lower edge, due to the presence of leaky TE waves.

The dispersion diagram for surface waves on a high-impedance surface is illustrated in Figure 12. The TM band begins at zero frequency, and bends over flat below the resonance frequency. The TE band begins at the resonance frequency, and slopes upward through the light line. The bandgap, which is the region where bound surface waves are not supported, extends from the upper edge of the TM band to the point where the TE band crosses the light line. This boundary, where $\omega = ck/n$, represents waves in free space propagating parallel to the surface. TE modes, that lie to the right of the light line, are bound to the surface while those that lie left of the light line lose energy to radiation damping. They are not bound to the surface, and are described as leaky waves. The approximate dispersion functions of eqns [27] and [28] are similar to the curves shown in Figure 12, but because they neglect radiation damping, they do not describe these leaky modes. Furthermore, like other photonic crystals, high-impedance surfaces also have higher-order bands and gaps that are not explained by the effective medium model.

The suppression of propagating surface modes can be used for a variety of applications, such as antennas and other microwave structures. One example is to change the radiation pattern of an antenna by covering its ground plane with a textured surface. Figure 13 shows a simple monopole antenna, along with the radiation patterns produced by the monopole when it is placed on a flat or textured ground plane. These plots show the relative magnitude of radiation from the antenna as a function of angle. TM surface currents on the finite metal ground plane form standing waves, which scatter from the edges and corners of a finite ground plane. This causes ripples in the radiation pattern as radiation from the edges interferes with the direct radiation from the antenna. Scattering of the surface currents from the ground plane edges also allows the antenna to radiate power into the backward hemisphere, which would otherwise be shadowed by the ground plane. On the high-impedance ground plane, the suppression of surface

Figure 13 (a) A monopole antenna on a ground plane, used for testing the effect of a textured surface on the radiation pattern. (b) The E-plane radiation pattern of the monopole on a flat metal ground plane shows ripples due to standing surface waves. (c) On a high-impedance ground plane, within the bandgap, surface waves are forbidden, and the radiation pattern is smooth. (d) Outside the bandgap, in the TM band, there is a high density of surface modes.

currents within the bandgap results in a smooth radiation pattern in the forward hemisphere, and very little power in the backward hemisphere. For frequencies outside the bandgap, such as within the TM band, the high density of TM modes can be seen as fine ripples in the pattern, and significant power in the backward hemisphere.

The monopole is useful for illustrating the suppression of TM surface modes, but many other antenna designs are possible. For example, the suppression of both TM and TE surface waves can be used to build patch antennas or aperture antennas with a high degree of symmetry between the E and H planes. Because these materials behave as artificial magnetic conductors, they can also be used to design a variety of thin antennas that would be impractical on conventional metal ground planes. Finally, by incorporating active elements into the surface, tunable or steerable antennas are possible.

List of Units and Nomenclature

a	lattice constant	meters
c	speed of light in vacuum	meters/second
C	capacitance	Farads
E	electric field	Volts/meter
H	magnetic field	Amps/meter
i	unit imaginary number	–
J	current density	Amps/meter^2
k	wave vector	1/meters
L	inductance	Henrys
m	effective electron mass	kilograms
n	refractive index	–
n	electron density	number/meter^3
q	electron charge	Coulombs
r	radius	meters
R	reflection coefficient	–
t	thickness	meters
T	transmission coefficient	–
Z	surface impedance	Ohms/square
α	decay constant into free space	1/meters
γ	decay constant into material	1/meters
δ	skin depth	meters
ε	electric permittivity	Farads/meter
η	impedance of free space	Ohms
λ	wavelength	meters
μ	magnetic permeability	Henrys/meter
π	diameter/circumference ratio	–
σ	electrical conductivity	1/Ohms
τ	mean electron collision time	seconds
ω	angular frequency	radians/second
∇	vector differential operator	–
∂	scalar differential operator	–
∞	infinity	–

See also

Babinet's Principle. **Diffractive Systems**: Diffractives in Animals; Microstructure Fibers; Omnidirectional Surfaces and Fibers. **Photonic Crystals:** Atomic Physics; Electromagnetic Theory; Nonlinear Optics in Photonic Crystal Fibers; Photonic Crystal Lasers, Cavities and Waveguide; Self-Assembled and Functionalized Photonic Crystals.

Further Reading

Ashcroft N and Mermin N (1976) *Solid State Physics*. Orlando, FL: Saunders College Publishing.

Brillouin L (1953) *Wave Propagation in Periodic Structures: Electric Filters and Crystal Lattices*, 2nd edn. New York: Dover Publications.

Brown E, Parker C and Yablonovitch E (1993) Radiation properties of a planar antenna on a photonic-crystal substrate. *Journal of the Optical Society of America B* 10: 404.

Collin R (1991) *Field Theory of Guided Waves*, 2nd edn. New York: IEEE Press.

Fan S, Villeneuve P and Joannopoulos J (1996) Large omnidirectional band gaps in metallodielectric photonic crystals. *Physics Review B* 54: 11245.

Feynman R, Leighton R and Sands M (1964) *The Feynman Lectures on Physics*, vol. 3. Reading, MA: Addison-Wesley.

Joannopoulos J, Meade R and Winn J (1995) *Photonic Crystals: Molding the Flow of Light*. Princeton, NJ: Princeton University Press.

Kildal P-S (1990) Artificially soft and hard surfaces in electromagnetics. *IEEE Transactions on Antennas and Propagation* 38: 1537.

Kurizki G and Haus J (eds) (1994) Photonic band structures. *Journal of Modern Optics* 41, special issue.

Lee S, Zarrillo G and Law C (1982) Simple formulas for transmission through periodic metal grids or plates. *IEEE Transactions on Antennas and Propagation* 30: 904.

Pendry J, Holden A, Stewart W and Youngs I (1996) Extremely low frequency plasmons in metallic mesostructures. *Physics Review Letters* 76: 4773.

Ramo S, Whinnery J and Van Duzer T (1984) *Fields and Waves in Communication Electronics*, 2nd edn. New York: John Wiley and Sons.

Sievenpiper D (1999) *High-Impedance Electromagnetic Surfaces*. Ph.D. Thesis, Los Angeles: University of California.

Sievenpiper D, Sickmiller M and Yablonovitch E (1996) 3D wire mesh photonic crystals. *Physics Review Letters* 76: 2480.

Sievenpiper D, Yablonovitch E, Winn J, *et al.* (1998) 3D metallo-dielectric photonic crystals with strong capacitive coupling between metallic islands. *Physics Review Letters* 80: 2829.

Sigalas M, Chan C, Ho K-M and Soukoulis C (1995) Metallic photonic band-gap materials. *Physics Review B* 52: 11744.

Smith D, Schultz S, Kroll N, et al. (1994) Experimental and theoretical results for a two-dimensional metal photonic band-gap cavity. *Applied Physics Letters* 65: 645.

Soukoulis C (ed.) (1996) Photonic band gap materials. *Proceedings of the NATO ASI on Photonic Band Gap Materials, Elounda, Crete, Greece, June 18–30, 1995*. The Netherlands: Kluwer Academic Publishers.

Wheeler H (1947) Fundamental limitations of small antennas. *Proceedings of the Institute of Radio Engineers* 35: 1479.

Yablonovitch E (1987) Inhibited spontaneous emission in solid-state physics and electronics. *Physics Review Letters* 58: 2059.

Nonlinear Optics in Photonic Crystal Fibers

J E Sharping, Cornell University, Ithaca, NY, USA

P Kumar, Northwestern University, Evanston, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Photonic crystal fibers (PCFs) are very similar to normal optical fibers in that they consist of a core surrounded by cladding, such that light is guided within the core of the fiber. The primary difference between PCF and standard optical fibers is that PCFs feature an air–silica cross-section, whereas standard optical fibers have an all-glass cross-section. An electron micrograph of a typical PCF is shown in Figure 1. The air holes extend along the axis of the fiber for its entire length and the core of the fiber is formed by a defect, or missing hole, in the periodic structure. The core is formed of solid glass, whose refractive index is that of pure silica (or whatever other glass is chosen), and the cladding is formed by the air–glass mixture, whose effective refractive index depends on the ratio of air-to-glass, also known as the air-fill fraction, that comprises the structure. The resulting effective-index of the cladding will be lower compared with that of the core and, as such, will provide the refractive index variation necessary to support total internal reflection at the core-cladding boundary, and guide light in a manner similar to that of standard optical fibers. The fiber design (i.e., size, shape, and the air-fill fraction) dictates solutions to Maxwell's equations for light propagating within the fiber. Valid solutions are referred to as 'modes' which propagate along the fiber in a known manner, and have a well-defined shape in the transverse direction (i.e., they have a well-defined transverse mode structure).

Nonlinear-optical effects in fibers result from the interaction of optical fields with the glass via the $\chi^{(3)}$, or Kerr nonlinearity. The phenomenon of nonlinear refractive index is a manifestation of a light–material interaction mediated by $\chi^{(3)}$. The magnitudes of the components of the third-order susceptibility tensor in glass, $\chi^{(3)}$, are generally quite small compared with the analogous second-order ($\chi^{(2)}$) terms for materials exhibiting such nonlinearities (e.g., lithium niobate, beta-barium borate (BBO), etc.). The relatively small $\chi^{(3)}$ nonlinearity in optical fibers makes them ideal for wavelength-division multiplexed optical communication where light propagation subject to a minimum of nonlinear effects is critical. Nonlinearity does, however, eventually become an issue in wavelength-division multiplexed systems as the launched optical power increases and as the channel spacing decreases. On the other hand, one can utilize nonlinear-optical effects in soliton communication systems and to build useful photonic devices. Despite the weak $\chi^{(3)}$

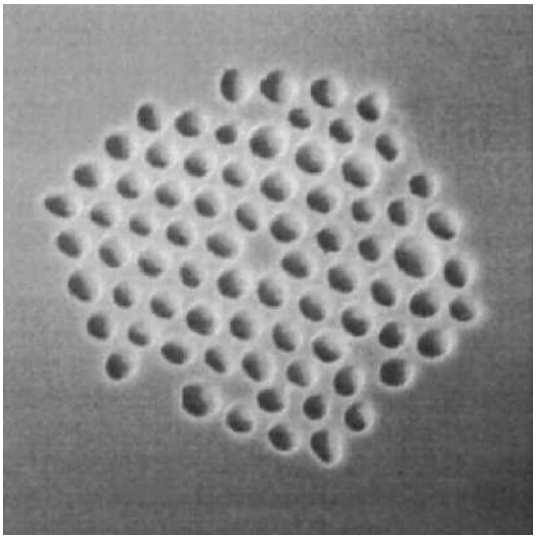

Figure 1 An electron micrograph showing the periodic microstructure of a typical PCF. The core is formed by the 'missing hole' in the center of the microstructure. (Reproduced with permission from Ranka JK, Windeler RS and Stentz AJ (2000) Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm. *Optical Letters* 25: 25–27. ©2000 Optical Society of America, courtesy of OFS.)

nonlinearity, the net nonlinear-optical effect in fibers can be large due to the ability to tightly confine intense fields within the core of an optical fiber and maintain the interaction over a long distance as the guided fields propagate through the fiber.

The study of nonlinear-fiber optics has benefited from dramatic improvements in optical fiber and fiber-optic device fabrication. The importance of understanding nonlinear-fiber optics is driven by the need to develop fiber-integrated devices, and also by the need to understand and mitigate the problems that these nonlinearities cause in optical communication systems.

This section introduces the unique linear- and nonlinear-optical properties of PCFs in order to understand the reasons why nonlinear-optical effects are often enhanced in such fibers. These discussions pertain to PCFs which are 'highly nonlinear'. It is essential to clarify that 'highly nonlinear' in this context does not mean that the $\chi^{(3)}$ is any larger than that of standard telecommunication fibers, rather that the effect of this nonlinearity is enhanced due to the fiber's very small core.

PCF Properties

Photonic crystal fibers feature a variety of interesting properties. From the standpoint of nonlinear-fiber optics there are four very useful fundamental properties of PCFs:

- a mechanically robust optical fiber can be fabricated with an extremely small core (a few μm^2);
- a fiber can be made to guide in a single transverse mode over an extremely broad wavelength range (370 nm–1600 nm);
- there are new degrees of freedom that allow one to manipulate the fiber's group-velocity dispersion (GVD) properties; and
- many, but not all, PCFs are polarization maintaining as a result of form birefringence present in the core.

The fact that small-core PCFs can be fabricated is clear from Figure 1 by taking note of the fact that the center defect region which comprises the core is about 1.7 μm in diameter. Photonic crystal fibers with even smaller cores have been fabricated.

Transverse Mode Structure

A widely accepted model used to describe the transverse modal behavior of PCFs is called the effective-index model. The effective-index model can be used to understand why some PCFs are 'endlessly single mode', meaning that the fiber guides in a single transverse mode over an exceptionally wide wavelength range (370 nm–1600 nm). In the effective-index model, the refractive index of the core, $n_{co}(\lambda)$, is that of glass, and the refractive index of the cladding, $n_{eff}(\lambda)$, assumes a value in between that of glass and air. In the context of PCFs, one makes a modification to the standard expression describing single-mode behavior in step-index fibers:

$$V_{eff} = \frac{2\pi\Lambda}{\lambda}\sqrt{n_{co}(\lambda)^2 - n_{eff}(\lambda)^2} < V_{cutoff} \qquad [1]$$

where Λ is the spacing between air holes, λ is the wavelength of light, and V_{cutoff} is the cutoff condition for the PCF. A similar expression for the V parameter is commonly used to understand the modal behavior of standard fibers where the larger the V is, the more transverse modes are supported within the fiber. In standard fibers the cutoff condition below which only a single mode can propagate within the core of a fiber is given by $V_{cutoff} < 2.405$. In the case of PCFs, a numerical method should be used to determine V_{cutoff}. Mechanically robust PCFs can be fabricated where the dispersion in $n_{eff}(\lambda)$ (i.e., the variation of n_{eff} with λ) offsets dispersion in $n_{co}(\lambda)$ and compensates for the $2\pi\Lambda/\lambda$ coefficient in eqn [1]. Therefore, the light within the fiber propagates in a single, Gaussian-like mode because for all wavelengths $V_{eff} < V_{cutoff}$. A graph of the variation of V_{eff} with Λ/λ is shown in Figure 2, where d represents the size of an air hole.

Conceptually, the effective index model can be understood by noting that at short wavelengths the mode field is confined well within the all-silica core, but as λ increases the mode field extends further into the air–glass cladding and V_{eff} and $n_{eff}(\lambda)$ both decrease.

Dispersion in PCF

Other critical differences between PCFs and standard optical fibers lie in the dispersion properties. When light propagates through a fiber its behavior depends on the light's optical frequency:

$$E(t,z) = A(t,z)e^{i[\omega t - \beta(\omega)z]} \qquad [2]$$

Equation [2] describes the mode as it propagates through the fiber. It is decomposed into a slowly varying envelope, $A(t,z)$, and a rapidly varying exponential component where ω is the frequency of the mode, t is time, z is the position along the length of the fiber, and $\beta(\omega)$ is called the mode-propagation constant. The general term used in describing the

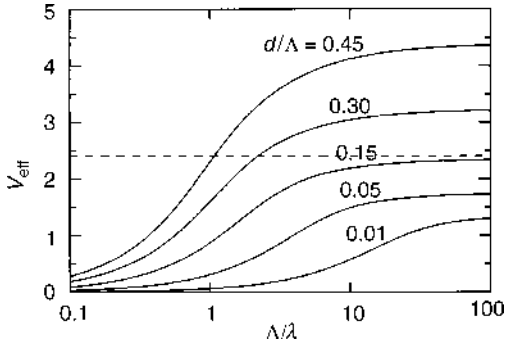

Figure 2 Variation of V_{eff} for different relative hole diameters d/Λ. The calculation assumes a fiber with an air–glass cross-section where the refractive index of air and glass was taken to be 1 and 1.45, respectively. The dashed line marks $V_{eff} = 2.405$, the cutoff value for a step-index fiber. (Reproduced with permission from Birks TA, Knight JC and Russell PStJ (1997) Endlessly single-mode photonic crystal fiber. *Optical Letters* 22: 961–963. ©1997 Optical Society of America.)

frequency dependence of β is chromatic dispersion, which includes contributions from the material as well as the waveguide. Other types of dispersion present in optical fibers include multi-modal (arising from multiple guided transverse modes) and polarization-mode dispersion.

One way to understand the chromatic dispersion of a mode propagating through an optical fiber is to study the Taylor series expansion of the mode-propagation constant, β, about the center frequency of the field, ω_0:

$$\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 + \cdots \qquad [3]$$

where $\beta_i = \mathrm{d}^i\beta/\mathrm{d}\omega^i$. The physical significance of the various β_i in eqn [3] are as follows: the phase-fronts of the electric field move at a speed given by $\omega/\beta_0 = v_p$, and the envelope, $A(t,z)$, moves at a group velocity given by $1/\beta_1$. A GVD term, which governs temporal spreading of the envelope, is given by β_2. Higher-order β terms are usually negligible for propagation of pulses of ≥1 ps duration in optical fibers and are lumped into the category of 'higher-order chromatic dispersion'.

The notation 'β_2', as defined above, is often used in the literature with dimensions of ps^2/km. However, another expression is frequently used because of its direct relationship to measured quantities. It is straightforward to measure the relative delay, T, between two pulses having different center wavelengths. Choosing a particular wavelength as a reference, one can then measure relative delay as a function of an injected pulse's center wavelength. The first derivative with respect to λ of the relative delay curve gives the GVD according to

$$D = \frac{\partial\left(\dfrac{1}{v_g}\right)}{\partial\lambda} = \frac{1}{L}\frac{\mathrm{d}T}{\mathrm{d}\lambda} = -\frac{2\pi c}{\lambda^2}\beta_2 \qquad [4]$$

where v_g is the group velocity, and L is the length of the fiber under test. The dimension commonly used for D is ps/(nm km).

Chromatic dispersion in single-mode optical fibers results from two different wavelength-dependent fiber parameters. The medium itself, glass in this case, has a wavelength-dependent refractive index. This 'material' contribution has the same magnitude regardless of the various parameters associated with the waveguide. A second contribution has to do with the design of the optical fiber. This 'waveguide' contribution to dispersion arises from the fact that the wavelength-dependent mode depends on the properties of the waveguide (i.e., the core size and refractive index contrast between the core and cladding). Empirical models can be used to describe the material, waveguide, and total GVD for standard communication fibers. Such a set of curves is given in Figure 3 where it can be seen that it is possible to have positive, negative, or zero values for D. For historical reasons, the regions where D is negative (β_2 is positive) exhibit 'normal GVD', while those where D is positive (β_2 is negative), exhibit 'anomalous GVD'. The wavelength corresponding to $D = 0$ is referred to as the zero-dispersion wavelength (λ_0), which for most silica glass fibers is about 1,300 nm.

In contrast with standard optical fibers, where the waveguide contribution to D is always less than zero, small-core PCFs can be fabricated where the waveguide contribution to GVD is positive and quite large. As such, in PCFs, λ_0 can be shifted to wavelengths shorter than the intrinsic dispersion zero of glass. Control over the GVD is essential for phase matching certain nonlinear-optical interactions involving light

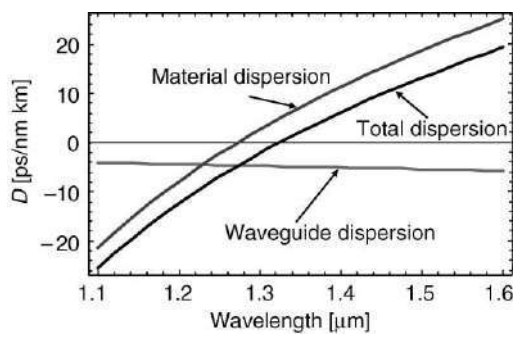

Figure 3 Plots of the theoretical dispersion coefficient, D, as a function of wavelength for a standard optical communication fiber.

of different colors co-propagating within a fiber. Indeed, several exciting applications of nonlinear optics in PCF require a fiber with a $\lambda_0 \sim 800$ nm. The GVD is also of great importance when working with pulsed light in PCF, because GVD results in temporal pulse broadening. It also governs pulse temporal walkoff effects, limiting the effective interaction length between pulses of different colors. This new flexibility to manipulate the GVD curve, by varying waveguide design parameters, is a key advantage associated with using PCFs for nonlinear optics.

Birefringence in PCF

The core of an optical fiber often exhibits some amount of anisotropy. The core may be elliptical in form (shape) which leads to a phenomenon referred to as form birefringence. Since mode propagation depends on the fiber structure, a fiber with an elliptical core will exhibit mode propagation that depends on the electric field's polarization with respect to the axes of the elliptical core. As a result of birefringence, the polarization of the mode varies as it propagates through the fiber (unless care is taken to align the polarization of the injected light with respect to a principal axis of the birefringence).

Polarization-maintaining (PM) fibers are designed to include birefringence in a particular axis of the fiber. By including a well-defined birefringence throughout the length of a fiber that is larger than that induced by external perturbations, fast and slow axes of the optical fiber are created for all guided wavelengths, giving two orthogonal 'polarization modes'. If light is injected into one of the polarization modes (i.e., with its linear polarization along one of the axes) it remains linearly polarized along that axis as it propagates along the fiber. The two polarization modes generally have different group velocities, so pulsed light in each mode will take a different amount of time to propagate through a given segment of fiber. Most PCFs exhibit strong birefringence due to a slightly elliptical core combined with a large core-cladding index difference, and so they behave similarly to PM fibers. Special care must be taken when working with PCF to be sure that the polarization of the light launched into the fiber is aligned with one of the birefringent axes.

In practice, there are a few other features of PCF that are of importance when discussing nonlinear-optic interactions:

- propagation losses are generally larger in PCFs than in standard optical fibers; and
- free-space coupling and splicing are difficult and can result in large coupling losses.

Nonlinear Phenomena

The basic principles determining nonlinear effects in PCFs are the same as those for standard optical fibers (*see* **Fiber and Guided Wave Optics:** Nonlinear Effects (Basics)). It is the new flexibility in PCFs to obtain transverse-modal and GVD behavior different from that of standard optical fibers that makes PCFs truly interesting for nonlinear optics. The relevant nonlinear-optical effects are: self-phase modulation (SPM); cross-phase modulation (CPM); third-harmonic generation (3HG); four-wave mixing (FWM); Raman scattering; and Brillouin scattering.

Self-phase modulation (also known as the optical Kerr effect) refers to the self-induced phase shift experienced by an optical field as it propagates through a fiber. It becomes particularly important for the case of pulses of light propagating through optical fibers. In small core PCFs, SPM is enhanced due to the high-intensity light propagating within the core. Self-phase modulation can lead to substantial spectral broadening of pulsed light propagating along an optical fiber.

When a pulse of light experiences normal GVD (i.e., $D < 0$) as it propagates, the longer-wavelength components travel faster than the shorter-wavelength components. Anomalous GVD (i.e., $D > 0$) leads to the opposite, short-wavelength components traveling faster than the long-wavelength components. Group-velocity dispersion generally leads to temporal broadening of pulses as they propagate along a fiber. Under ideal conditions, however, SPM in combination with anomalous GVD, leads to pulses which propagate without any temporal or spectral broadening. These self-sustaining pulses are called 'optical solitons'.

When waves of light having different wavelengths co-propagate along a fiber, CPM can occur. It can be understood as a phase shift induced on one wave, due to the presence of the other wave. Cross-phase modulation also leads to spectral broadening and solitonic pulse propagation.

In 3HG and FWM, one or more photons are destroyed and others are created. In 3HG, three 'fundamental' photons are destroyed to create one with three times the energy of the fundamental photons. In FWM, two fundamental photons are destroyed while two others are created. While it is straightforward to conserve energy in 3HG and FWM, these interactions must be 'phase-matched', meaning that the interacting waves must be made to propagate in-phase over a meaningful length. Such phase-matching conditions need to be carefully considered when studying 3HG and FWM. Nevertheless, 3HG and FWM can be used to obtain

frequency shifts and all-optical amplification. In comparison with other types of optical fibers, PCFs are particularly useful for 3HG and FWM applications. The small core of the PCF allows interactions to occur at much lower input powers, and the new flexibility associated with the GVD properties permits phase matching in cases which are not possible using standard optical fibers. Finally, the fiber's endlessly single-mode behavior permits very good transverse mode overlap between interacting waves having widely different center wavelengths.

Self-phase modulation, CPM, 3HG, and FWM are photon–photon interactions wherein no energy is exchanged with the medium itself. In contrast, Raman scattering and Brillouin scattering result from photon–phonon interactions. The differences between Raman and Brillouin scatterings lie in the energy of the phonons involved and the direction in which the interactions occur. Raman scattering is an interaction between a photon and an optical–phonon mode of the molecules making up the material. In the case of Raman scattering in glass, the energy shift, associated with molecular vibrational (Raman) modes, corresponds to frequencies of 1–12 THz. Raman scattering occurs in the forward and backward directions. With pulsed light, stimulated Raman scattering can occur when the lower-frequency spectrum of the pulse overlaps with the spectrum of the Raman resonances excited by the higher-frequency spectrum. When this happens, energy can be efficiently shifted in spectrum towards the peak of the Raman resonance. In the forward direction, this 'Raman self-frequency shift' builds up. Additionally, if the Raman self-frequency shift occurs in the presence of anomalous dispersion, a 'Raman soliton self-frequency shift' can result. As the injected power is increased, the spectral shift between the injected pulse and the resulting Raman soliton increases. The principal advantage associated with PCF is the ability to generate Raman solitons for a broader range of wavelengths than was previously possible.

For Brillouin scattering, interaction with the acoustical phonons results in frequency shifts of about 10 GHz and the interaction only occurs in the backward direction. Brillouin scattering is generally a nuisance in fiber-based devices, leading to intensity noise and other problems. The interesting feature of Brillouin scattering in PCFs is that the threshold intensity where problems begin to occur is higher for PCFs than for standard optical fibers. The higher threshold permits further optimization of fiber-based devices wherein Brillouin scattering limits the performance.

Experiment Examples

In the following subsections, a selection of experiments, demonstrating a few of the relevant nonlinear-optical phenomena, are briefly described.

Supercontinuum Generation

One of the most exciting demonstrations of nonlinear optics in PCF is that of supercontinuum generation. In a typical experiment, 100 femtosecond pulses from a mode-locked Ti:Sapphire laser operating at a wavelength of 800 nm were injected into the PCF. As the injected power was increased, a broad continuum of spectrum was generated from wavelengths of 400 nm up to 1,600 nm. Typical data showing the input and output optical spectra are given in Figure 4, where it can be seen that well over an octave of frequency spectrum is generated from an input pulse whose spectral width is only about 10 nm.

It is widely accepted that supercontinuum generation results from a combination of linear and nonlinear optical effects conspiring to generate the broad spectrum. As the pulse propagates through the PCF, SPM, FWM, and Raman scattering are all likely to occur with relative efficiencies depending heavily on the input pulse's spectral and temporal characteristics, as well as the fiber's properties. Indeed, efficient supercontinuum generation can occur in PCFs for pump wavelengths lying near the zero-dispersion point in the normal or anomalous dispersion regime of the PCF, and for fibers as short as a few millimeters in length.

Spectral broadening, due to SPM, is common in standard optical fibers, but what makes this

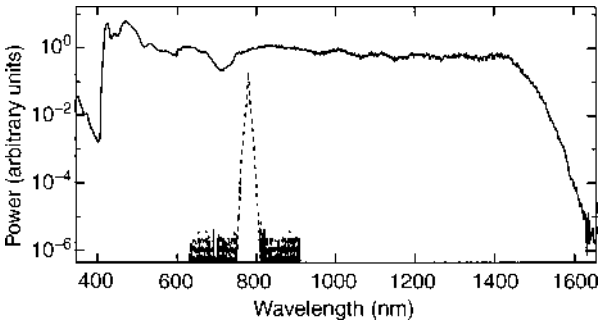

Figure 4 In supercontinuum generation one observes a broad continuum generated after short pulses of light from a Ti:Sapphire laser propagate through a 75 cm section of PCF. The spectrum of the input pulse is shown as a dashed curve, while the output is a solid curve. (Reproduced with permission from Ranka JK, Windeler RS and Strentz AJ (2000) Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm. *Optical Letters* 25: 25–27. ©2000 Optical Society of America.)

particular experiment interesting is the remarkable width of spectrum generated with a short piece of fiber (~1 m) and with comparatively small optical powers.

Optical Switching in PCF

Cross-phase modulation can be used to build an all-optical switch. Such a switch can be implemented as a three-port device where the output port for a given optical bit (the signal pulse), is determined by the presence of an optical control (the pump pulse). Switching can then be achieved by dividing the signal pulse equally on two arms of an interferometer and injecting the strong pump pulse only on one arm. Because the pump pulse co-propagates with only one of the two signal pulses, there exists a CPM-induced phase difference $\phi_{NL} = 2\gamma P_p L$ at the output of the interferometer, where P_p is the peak power of the pump pulse and L is the interaction length. By varying the intensity of the pump pulses one can vary the magnitude of this phase difference. If a π-phase shift is achieved, one can switch the interference from destructive to constructive, or vice-versa, thus realizing an all-optical switch.

Figure 5 shows an experimental setup used to observe switching near 1,550 nm (a similar apparatus using bulk-optic rather than fiber-optic components can be used to conduct experiments near 780 nm). The pump and signal are synchronous few-ps-duration pulses with a tunable wavelength separation

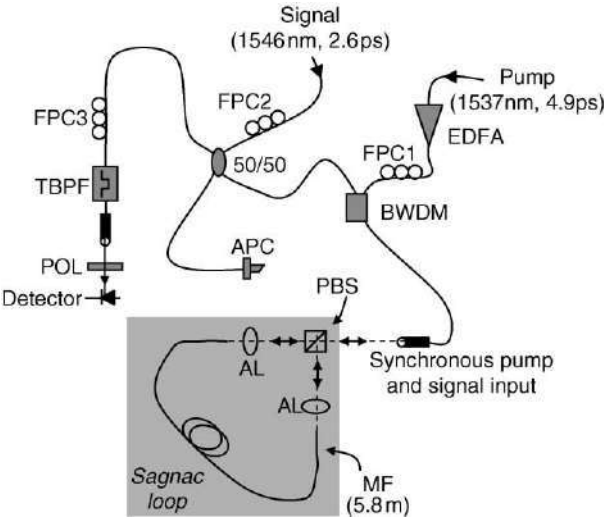

Figure 5 Experimental setup used to demonstrate all-optical switching near 1,550 nm. (EDFA, erbium-doped fiber amplifier; BWDM, bandpass wavelength-division multiplexor; PBS, polarization beamsplitter; FPC, fiber polarization controller) (Reproduced with permission from Sharping JE, Fiorentino M, Kumar P and Windeler RS (2002) All-optical switching based on cross-phase modulation in microstructure fiber. *IEEE Photonics Technology Letters* 14: 77–79. ©2002 IEEE).

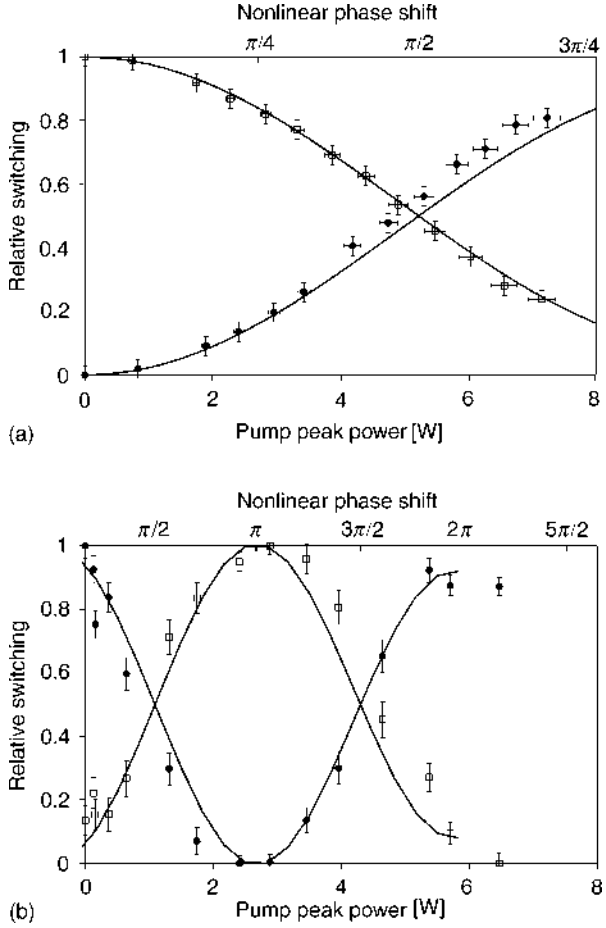

Figure 6 Switching curves (open boxes and filled circles), showing the relative power measured in each port of the switch vs. the pump peak power, for experiments conducted near (a) 1,550 nm and (b) 780 nm. The curves accompanying the data are generated from numerical solutions of coupled wave equations for CPM. (Reproduced with permission from Sharping JE, Fiorentino M, Kumar P and Windeler RS (2002) All-optical switching based on cross-phase modulation in microstructure fiber. *IEEE Photonics Technology Letters* 14: 77–79. ©2002 IEEE.)

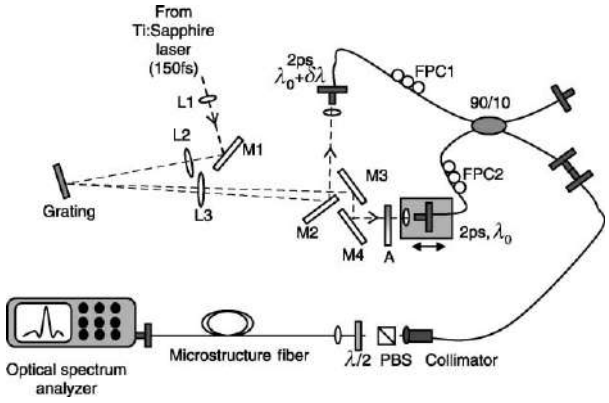

Figure 7 A schematic of the experimental setup used to investigate FWM in PCFs. (Reproduced with permission from Sharping JE, Fiorentino M, Coker A, Kumar P and Windeler RS (2001) Four-wave mixing in microstructure fiber. *Optical Letters* 26: 1048–1050. ©2001 Optical Society of America.)

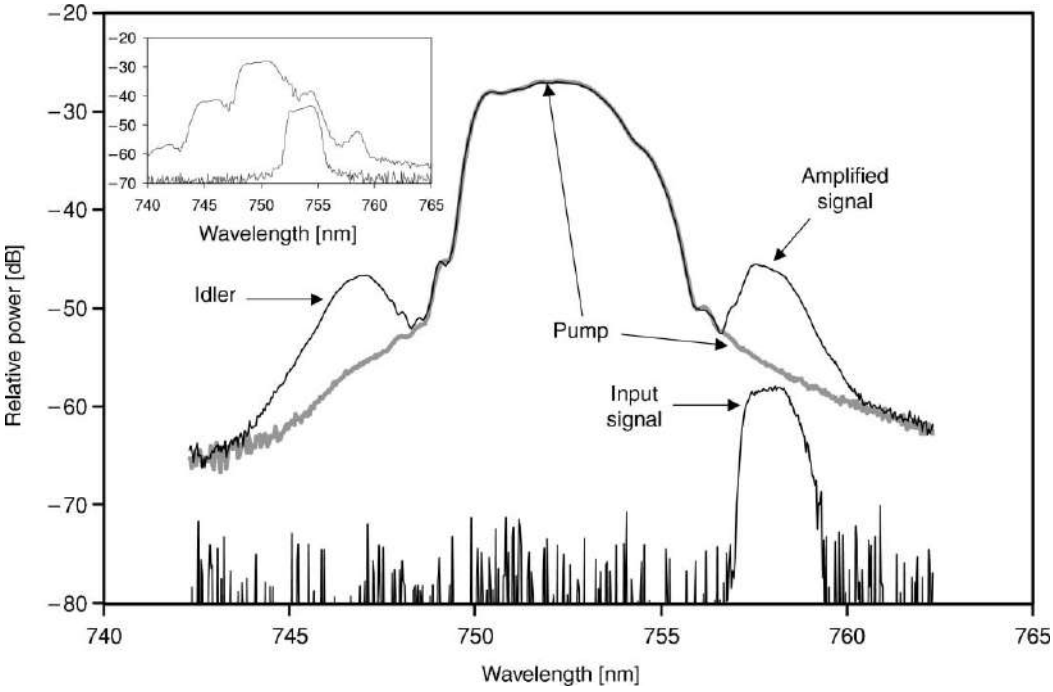

Figure 8 A typical FWM spectrum observed at the output of the microstructure fiber. The inset shows a spectrum where higher-order cascaded mixing is evident. (Reproduced with permission from Sharping JE, Fiorentino M, Coker A, Kumar P and Windeler RS (2001) Four-wave mixing in microstructure fiber. *Optical Letters* 26: 1048–1050. ©2001 Optical Society of America.)

of 5–15 nm. The switching characteristics for this implementation are shown in Figure 6. The apparatus has the advantage of requiring short fiber lengths, low switching powers, and allows switching of weak pulses. It demonstrates the feasibility of using nonlinear optics in PCF to perform essential functions in high-speed all-optical processing.

Parametric (Mixing) Processes

The first set of experiments with controlled FWM in a PCF achieved nondegenerate parametric gains over a 30 nm range of pump wavelengths near the λ_0 of the PCF. The experiments also confirmed the wavelength dependence of the GVD coefficient of the PCF near λ_0. Since the dispersion characteristics of these fibers can be adjusted during the fabrication process, the experiments demonstrate the potential for the use of PCFs in broadband parametric amplifiers, wavelength shifters, and other optical communication devices.

The experimental setup used to demonstrate phase-matched FWM in PCF is shown in Figure 7. The pump and the input signal are two synchronous pulsed beams having 3–5 nm wavelength separation with the center wavelength tunable over a 720–850 nm range. The maximum peak power of the pump pulses is ≃12 W. The two synchronous beams are then combined and injected into the PCF. The pump and signal's optical paths are adjusted to obtain temporal overlap in the PCF and their polarizations are aligned by fiber polarization controllers.

Figure 8 shows a typical FWM optical spectrum at the output of a 6.1 m long PCF. Here the strong pump beam and the weak signal beam have wavelengths of 753 nm and 758 nm, respectively. The spectrum shows the undepleted pump, the amplified signal, and the generated idler at 747 nm. The spectra in Figure 8 show that large gain is achievable for a pump-to-signal spacing of 5 nm. Gain values of more than 20 (13 dB) were obtained.

Conclusion

In summary, the advantages of using photonic crystal fibers for demonstrating nonlinear-fiber optical effects arise from four novel properties:

- the nonlinear coefficient is enhanced in small-core PCFs (core area of a few μm^2);
- PCFs can support a single transverse mode over an extremely broad wavelength range (370 nm–1600 nm);
- PCF design parameters allow one to manipulate the fiber's GVD properties; and
- nonlinear interactions are enhanced due to the polarization maintaining properties of PCFs which result from form birefringence present in the core.

These four properties combine to allow efficient interactions to occur in PCFs which are either inefficient or not possible at all in standard optical fibers.

See also

Optical Amplifiers: Raman, Brillouin and Parametric Amplifiers. **Photonic Crystals:** Electromagnetic Theory. **Scattering:** Raman Scattering.

Further Reading

Agrawal GP (2000) *Nonlinear Fiber Optics*, 3rd edn. San Diego, CA: Academic Press.

Birks TA, Knight JC and Russell PStJ (1997) Endlessly single-mode photonic crystal fiber. *Optical Letters* 22: 961–963.

Hansryd J, Andrekson PK, Westlund M, Li J and Hedekvist P (2002) Fiber-based optical parametric amplifiers and their applications. *IEEE Journal of Selected Topics in Quantum Electronics* 8: 506–520.

Monro TM, Richardson DJ, Broderick NGR and Bennett PJ (2000) Modeling large air fraction holey optical fibers. *Journal of Lightwave Technology* 18: 50–56.

Ranka JK, Windeler RS and Stentz AJ (2000) Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm. *Optical Letters* 25: 25–27.

Russell PStJ (2003) Photonic crystal fibers. *Science* 299: 358–362.

Sharping JE, Fiorentino M, Coker A, Kumar P and Windeler RS (2001) Four-wave mixing in microstructure fiber. *Optical Letters* 26: 1048–1050.

Sharping JE, Fiorentino M, Kumar P and Windeler RS (2002) All-optical switching based on cross-phase modulation in microstructure fiber. *IEEE Photonics Technology Letters* 14: 77–79.

Photonic Crystal Lasers, Cavities and Waveguides

J O'Brien and W Kuang, University of Southern California, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Photonic bandstructure engineering, in which the electromagnetic dispersion relations are intentionally modified, can be accomplished by creating one-, two-, and three-dimensionally periodic dielectric materials. These periodic materials, called photonic crystals, can be used to confine and guide light. This article describes the formation of optical resonant cavities and waveguides in these materials. Most of the discussion will focus on two-dimensional photonic crystals. One-dimensionally periodic systems are treated elsewhere in this volume. This article will begin with a short review of the relevant electromagnetic properties of photonic crystals. The formation and properties of resonant cavities will follow this, and the article will conclude with waveguides formed by linear defects in two-dimensional photonic crystals.

Electromagnetics of Photonic Crystals

A two-dimensional photonic crystal consists of a dielectric material in which the dielectric constant is periodic in two dimensions. The period of these materials is on the order of a half-wavelength of the operating optical wavelength. In this section, we briefly review the electromagnetic properties of two-dimensional photonic crystals. For more details, see the article entitled **Spectroscopy:** Raman Spectroscopy in this volume.

There are five Bravais lattices in two dimensions. Most of the research on photonic crystal resonant cavities and waveguides has focused on square and triangular lattices, and in this article we will consider examples based on the triangular lattice. Since these structures are usually defined lithographically however, there is, in principle, no need to confine the investigation to naturally occurring Bravais lattices. Due to the limitations of current nanofabrication technology, much of the research has focused on photonic crystals that have a finite thickness. The important feature of photonic crystals is that the bandgaps in the electromagnetic spectrum opened up by all of the Bragg planes overlap spatially and spectrally, so that a frequency region exists in which no electromagnetic propagation is allowed. These electromagnetic bandgaps can be used to confine optical modes in very small-volume resonant cavities and in waveguides. In high-contrast dielectric systems, such as a semiconductor/air periodic system, only a few lattice periods may be necessary to confine an electromagnetic mode. In the case of finite thickness photonic crystals, this periodicity has the effect of limiting the bandgap formed to the in-plane

directions and in the case of a two-dimensional photonic crystal in a high-index dielectric slab to the formation of a bandgap in the guided modes of the slab. In this case there are still radiation modes of the slab.

Consider a two-dimensional photonic crystal that is periodic in the $x-y$ plane. To start, we can consider materials that are uniform and infinitely extended in z. Electromagnetic fields propagating in the $x-y$ plane are Bloch states. These can be written as:

$$\vec{H}_{k_z,k_\parallel}(\vec{\rho},z,t) = \mathrm{e}^{\mathrm{i}(\omega t-(\vec{k}_\parallel\cdot\vec{\rho}+k_z z))}\vec{u}_{k_z,k_\parallel}(\vec{\rho}) \qquad [1]$$

where $\vec{\rho}$ labels the in-plane coordinates and $\vec{k}_\parallel$ labels the in-plane wavevectors. $\vec{u}_{k_z,k_\parallel}$ is a periodic function in the $x-y$ plane. Modes propagating in the $x-y$ plane in such a system can be classified as transverse electric (TE) waves or transverse magnetic (TM) waves. TE waves have nonzero E_x, E_y, and H_z field components, and TM waves have nonzero E_z, H_x, and H_y field components. A triangular lattice of air holes patterned into a high refractive index dielectric can be used to create a bandgap for both the guided TE and TM modes of the membrane for a range of hole radii. The TE and TM bandgaps that are formed can be over different spectral ranges, however. For laser applications this is unimportant since in practice the emission occurs from electron-hole recombination in a semiconductor quantum well, and for unstrained or compressively strained quantum well materials this emission is TE polarized. In most cases, waveguides have also been designed to work for a single optical polarization. It is generally true that a photonic lattice that consists of a connected high dielectric region is likely to exhibit a TE bandgap, while a lattice formed by disconnected high dielectric regions is more likely to exhibit a TM bandgap. The bandgap in the TE modes is formed between the first and second bands. The first band is called the dielectric band because the field at the Brillouin zone boundary is a standing wave with its intensity concentrated in the high dielectric regions. The second band is called the air band because at the Brillouin zone boundary this field is a standing wave with its intensity localized in the low dielectric regions. Over a reasonable range of lattice parameters, the bandwidth of the TE bandgap can be changed by changing r/a where r is the radius of the holes and a is the lattice constant of the triangular lattice. As r/a gets larger the dielectric band moves up in frequency. This can be thought of as being due to the fact that this mode has a decreasing effective index as r/a increases. The air band frequency increases as r/a increases. Since more of the electric field of the air band is located in the low dielectric regions than the field of the dielectric band, the air band increases in frequency with increasing r/a faster than the dielectric band and the bandgap therefore increases with increasing r/a. This trend holds in photonic crystals of finite thickness as well.

For photonic crystals of a finite thickness, the ability to simply classify modes as either TE or TM is lost. In the most simple finite thickness photonic crystals, that of a high index dielectric slab in which a two-dimensional photonic crystal has been patterned surrounded above and below by the same low index material which may be air of sapphire, or silicon dioxide, the modes can be classified as being either even or odd modes with respect to the mid-plane of the high index slab. In this mid-plane, however, modes can be classified as being either TE or TM, with each containing the same nonzero vector field components as in the infinite case. Away from this mid-plane, however, modes have in general six nonzero electromagnetic field components. The photonic bandgap that is created in this case is in the guided mode spectrum of the high index dielectric slab. There is no gap formed in the radiation modes of the slab. The radiation modes consist of modes with a ray vector that is not totally internally reflected at the high index slab/low index cladding interfaces. Figure 1 shows the dispersion relation, plotted over the irreducible Brillouin zone, of the even modes for dielectric slab with a refractive index of 3.4 in which a two-dimensional triangular lattice of holes has been patterned. The surrounding material is air. The figure shows an electromagnetic bandgap

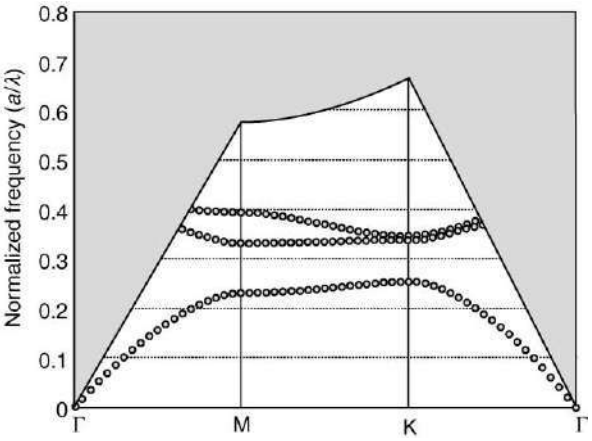

Figure 1 The photonic band diagram for a photonic crystal slab with a triangular lattice of air holes perforating a high-index dielectric slab. The slab, which assumes a refractive index $n = 3.4$, is suspended in the air with a thickness of $d/a = 0.6$. The air holes have radii of $r/a = 0.3$, where a is the lattice constant. Only the lowest three eigenmodes are shown in the plot.

formed between the first and the second guided modes of the slab. Radiation modes of the slab have real valued propagation vectors in the z-direction. These modes lie above the in-plane dispersion relation for air and exist inside the shaded region in the figure. This dividing line between the guided modes and the radiation modes is referred to as the light line.

If the upper and lower cladding layers of the high index slab are different and therefore have different indices of refraction, then the ability to classify modes as being either even or odd about the mid-plane of the slab disappears and it is possible to lose the bandgap in the guided modes altogether. In cases where the refractive index of the bottom cladding is not significantly different from that of the top cladding, such as a case where air serves as the top cladding layer and a material such as sapphire or silicon dioxide forms the bottom cladding, it is found that the even and odd modes describe the field reasonably accurately. To emphasize the approximation involved in the model, these modes are most often referred to in the literature as even-like or odd-like. Overall, there are three primary effects of having asymmetric cladding layers such as an air top cladding and sapphire bottom cladding case. These are a reduction in the effective bandgap width, an increase in the area of the radiation modes on the dispersion relation, and the loss of a rigorous bandgap in the guided modes of the slab. The first two of these effects are the most serious for device designers.

Photonic Crystal Resonant Cavities and Lasers

Photonic crystal resonant cavities can be formed at frequencies inside the bandgap or at the bandedges. Modes in the bandgap are formed as a result of a defect in the lattice. Modes at the bandedge are also sometimes used because the fields with wavevectors at Brillouin zone boundary are standing waves. These standing waves are analogous to the resonant modes that are formed in distributed feedback (DFB) laser structures. Here we will focus on modes confined in the bandgap. Allowed modes in electromagnetic bandgaps in photonic crystal slabs can be introduced by introducing one or more defects into the lattice. By perturbing a single lattice site, we can permit a localized mode or modes that have a frequency in the gap. This phenomenon is similar to the formation of deep levels in an electronic bandgap of a solid due to impurities and identical to the formation of resonant modes in vertical cavity surface emitting lasers (VCSELs) and phase-shifted DFB lasers. In each of these last cases, a defect in the periodicity of a precise thickness occurs between two one-dimensional distributed Bragg reflectors. Defect modes in photonic crystal resonant cavities can be engineered to radiate in a particular direction by trading-off the in-plane versus out-of-plane losses, and the resonant frequency is determined by the parameters of the lattice. Very small mode optical mode volumes can also be obtained in these resonant cavities leading to the possibility of modifying spontaneous emission.

Simple models exist which illustrate the formation of localized modes inside the bandgap as a result of defects in the otherwise periodic lattice. One of these was applied to the formation of deep levels in solids. This model can also be applied to the electromagnetics of defects in periodic structure. For a simple one-dimensional case, assuming that the defect exists at a single unit cell and that the dispersion of the band is dominated by nearest neighbor interactions, this can be solved exactly. Figure 2 shows the allowed frequencies of a ten unit cell chain with a single defect located at the center unit cell. Notice that one mode, called a donor mode, drops out of the band for positive values of $\Delta\varepsilon$ and one mode, called an acceptor mode, rises out of the band for negative values of $\Delta\varepsilon$. This is true in general. A perturbation of the lattice in which the defect has a higher index than the unperturbed lattice will cause a mode to drop out of the air band while a perturbation that has a lower dielectric constant than the background lattice will cause a mode to rise into the bandgap out of the dielectric band. In practice, resonant modes and frequencies in photonic crystals are modeled numerically.

A great deal of variety exists in resonant modes formed by defects in two-dimensional photonic crystal slabs. The simplest examples consist of a single missing hole in a two-dimensional square or

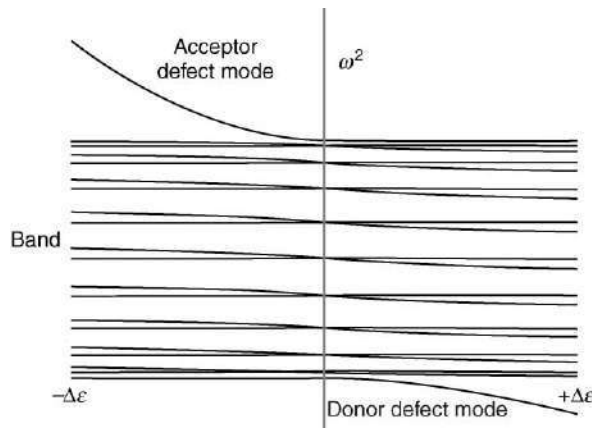

Figure 2 A one-dimensional Slater–Koster model for a single defect in a 10-unit-cell chain, showing the allowed states as a function of the defect perturbation.

triangular lattice patterned into a high-index dielectric slab. However, the resonant mode size, shape, frequency, and polarization can be engineered by tailoring the local dielectric in the region of the resonant mode. Losses in these cavities can be conceptually separated into losses in-plane through the photonic crystal and radiation losses out-of-plane. The in-plane losses are a result of having only a finite number of lattice periods surrounding the localized mode. A finite number of periods results in a finite loss due to tunneling of the fields through the lattice. This loss can in principle be made arbitrarily small by increasing the number of lattice periods surrounding the resonant mode. For a fixed number of lattice periods, the in-plane radiation loss will be reduced as the bandwidth of the photonic bandgap is increased because the magnitude of the in-plane decay constant increases with increasing bandgap width. A resonant mode formed by a defect in a truly three-dimensional photonic crystal will have losses limited by the number of achievable lattice periods and the bandwidth of the photonic crystal bandgap.

Photonic crystal cavities formed by two-dimensional photonic crystals of finite thickness suffer from an out-of-plane radiation loss. Optical confinement in the direction perpendicular to the two-dimensional photonic crystal is due to total internal reflection of the mode that occurs in the high index photonic crystal slab surrounded by lower index materials. Out-of-plane losses are due to the presence of wavevector components in the resonant mode that lie above the light line inside the radiation cone of the cladding. Figure 3 shows the Fourier transform of the magnetic field at the mid-plane of photonic crystal slab in which the resonant mode is formed by a single missing hole in a two-dimensional triangular lattice. In this cavity geometry, a doubly degenerate pair of modes is introduced into the photonic bandgap. The figure shows the Fourier transform of the field for one mode of this double degenerate pair. The boundary of the radiation cone in the figure is marked by the solid circle. Components of the mode inside this circle are not confined to the slab and contribute to out-of-plane radiation losses. These losses can be reduced by engineering the mode so that it has a smaller overlap with the radiation cone. This can be accomplished by allowing the mode to expand in real space in directions in which the wavevectors of the mode extend into the radiation cone. This reduces the spread of the Fourier transform of the mode and reduces the overlap of the mode with the radiation cone.

Symmetry can also be used to reduce the coupling to the radiation fields of the slab. Reducing the resonant frequency of the mode can also reduce out-of-plane radiation losses. This has the effect of reducing the size of the radiation cone in *k*-space. One way to accomplish this is to reduce *r*/*a* in the photonic crystal lattice. This will generally reduce frequencies of all of the photonic crystal modes. It will also, however, generally be accompanied by a reduction in the bandwidth of the photonic bandgap which will increase the in-plane losses. As a result, care must be taken in designing photonic crystal resonant cavities. Nevertheless, in the smallest mode volume photonic crystal resonant cavities, the optical losses will generally be dominated by out-of-plane radiation loss.

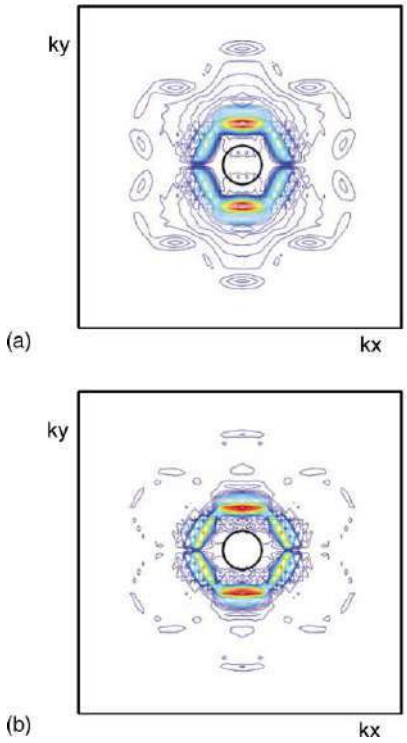

Figure 3 (a) The spatial Fourier transform of the magnetic field component, Hz, of a defect cavity mode at the mid-plane of a triangular-lattice suspended membrane single-defect photonic crystal microcavity. The inner circle indicates the light cone inside which mode components radiate vertically. (b) The spatial Fourier transform of a modified defect cavity mode. The overlap of the mode with light cone has been reduced by modifying the defect cavity.

Quality factors have been predicted to exceed 100 000 in carefully designed photonic crystal resonant cavities in which the optical mode volume is on the order of a few cubic half-wavelengths in the material. Larger resonant cavities formed by introducing multiple defects in the lattice are often less well confined in the real space lattice leading to a reduction in the components overlapping the radiation cone in *k*-space and a reduction in out-of-plane

radiation loss. These cavities also are much more likely to support multiple resonant modes for a given index perturbation.

In analyzing resonance modes of photonic crystal resonant cavities, one of the most important parameters is the quality factor Q of the cold cavity. Practically, a theoretical prediction of the quality factor requires the use of numerical methods and finite-difference time-domain and finite element techniques are commonly used. There are basically three numerical approaches that can be used to calculate the quality factor of the cavity modes. Two of these methods calculate the quality factor from the time domain fields while the third is a frequency domain approach. The first method is to calculate the slope of the exponential decay of the energy of a given cavity mode with time. This decay of energy from the cavity is described by $\exp(-t/t_{\mathrm{ph}})$ where t_{ph} is the photon lifetime which is related to the quality factor Q, by $Q = \omega t_{\mathrm{ph}}$. This method is most useful for relatively low Q modes where the slope of energy decay is visibly greater than zero.

Another method is to calculate the ratio of the full width at half magnitude (FWHM) of the cavity resonance in the frequency domain, $\Delta\omega$, to the center frequency, ω_0. However, distortion to the spectrum is often introduced because the numerical simulation terminates before the impulse response is fully evolved. This has the effect of viewing the true time-domain response through a rectangular window, which translates mathematically into the convolution of the true spectrum with a sinc function. This must be accounted for in the determination of the quality factor.

A third method calculates the ratio of cycle-averaged power absorbed in the boundary to the total energy in the cavity mode. This last method has the advantage of by being able to separate the radiation losses into different directions:

$$\frac{1}{Q} = \frac{\omega P}{U} = \frac{\omega(P_{\perp} + P_{\parallel})}{U} = \frac{1}{Q_{\perp}} + \frac{1}{Q_{\parallel}} \qquad [2]$$

in which, the effective vertical quality factor $Q_{\perp}$ is given by the ratio of power lost to the absorber at the top and bottom $P_{\perp}$ to the total cavity energy $U(t)$, and the effective in-plane quality factor $Q_{\parallel}$ is similarly given by the ratio of in-plane power loss $P_{\parallel}$ to the total cavity energy.

Photonic crystal lasers and resonant cavities have been demonstrated experimentally. The first demonstration of lasing in a two-dimensional photonic crystal laser occurred in a cavity formed by a single defect in a triangular lattice that was patterned into an InGaAsP membrane. The laser operated under optically pumped conditions at 143 K. Room temperature pulsed operation of photonic crystal defect lasers was reported shortly thereafter. Figure 4 shows an electron micrograph of such a resonant cavity. In the figure, the resonant cavity is formed in a suspended membrane in which a triangular lattice photonic crystal with a defect has been patterned. The fabrication of these devices involves pattern definition by electron beam lithography and pattern transfer by a series of wet and dry etching steps. The observed quality factors are, of course, sensitive to fabrication imperfections. Many other cavity designs and demonstrations have also been reported.

To operate a photonic crystal laser continuously at room temperature, it is necessary to design a high-Q resonant cavity that dissipates heat well. Poor heat dissipation is a major drawback of suspended membrane resonant cavities. One strategy for improving the heat dissipation in photonic crystal laser cavities is to form the two-dimensional photonic crystal membrane cavity on top of a low-index high-thermal-conductivity substrate. Figure 5 shows an electron micrograph of a photonic crystal cavity formed by leaving out 37 holes from a triangular lattice patterned into an InGaAsP membrane. This membrane is bonded to a sapphire substrate, which facilitates heat dissipation. This cavity is capable of room temperature continuous wave operation.

It is believed that these lasers are capable of electrically-pumped room temperature operation. This is based on the fact that large quality factors have been demonstrated in optically-pumped photonic crystal lasers and some quality factor reduction due to free carrier absorption and absorption at metal contacts can be tolerated and still lead to lasing. A reasonable assumption would be that the

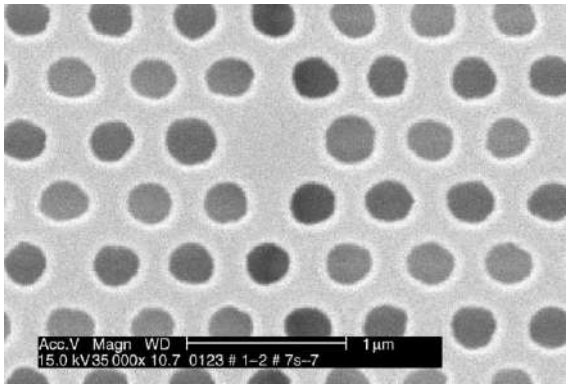

Figure 4 The scanning electron micrograph of a triangular-lattice suspended membrane photonic crystal single defect cavity.

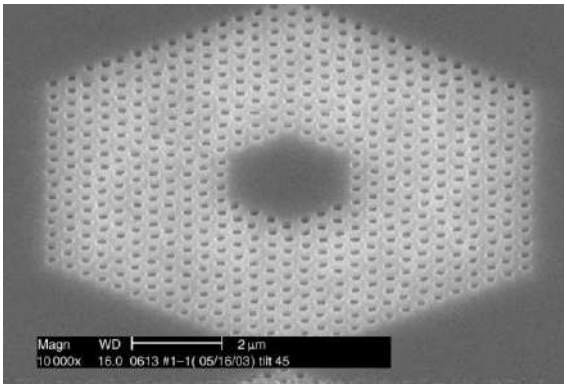

Figure 5 The scanning electron micrograph of a triangular-lattice, 37-missing-hole, photonic crystal defect cavity, bonded on sapphire which serves as lower cladding to improve heat dissipation. CW operation of the laser cavity was demonstrated.

performance of these electrically-pumped photonic crystal lasers would have much in common with the performance of VCSELs. Both VCSELs and photonic crystal cavities have small mode volumes with the cavity formed in some directions by distributed Bragg reflection. The emission direction of photonic crystal lasers can be engineered to be vertical or in-plane, but the basic resonant cavities do have important similarities. The free carrier absorption loss in photonic crystal laser cavities may be expected to be larger than in VCSELs since a photonic crystal laser mode lacks the standing wave behavior in the vertical direction that allows doping at the nodes of the standing wave to reduce the absorption loss. If free carrier absorption loss leads to larger internal losses for photonic crystal lasers than for VCSELs, then some reduction in the achievable slope efficiencies will also occur for photonic crystal lasers.

Photonic Crystal Waveguides

Another important optical element for integrated photonic circuitry is a linear waveguide. Conventional dielectric waveguides confine propagating beams by an index of refraction difference between the waveguide core and the waveguide cladding. Photonic crystal waveguides are most often formed by a linear defect, which consists of a row of modified unit cells of the lattice, patterned into a high index dielectric membrane. The guiding in this case is due to a mixture of total internal reflection at the high index membrane/low index cladding interfaces and distributed reflections from the photonic crystal in-plane. Because this confinement does not exclusively dependent on total internal reflection, the transmission loss through small bending radius waveguide branches and bends can be made very small. Predictions and experimental demonstrations exist which show that

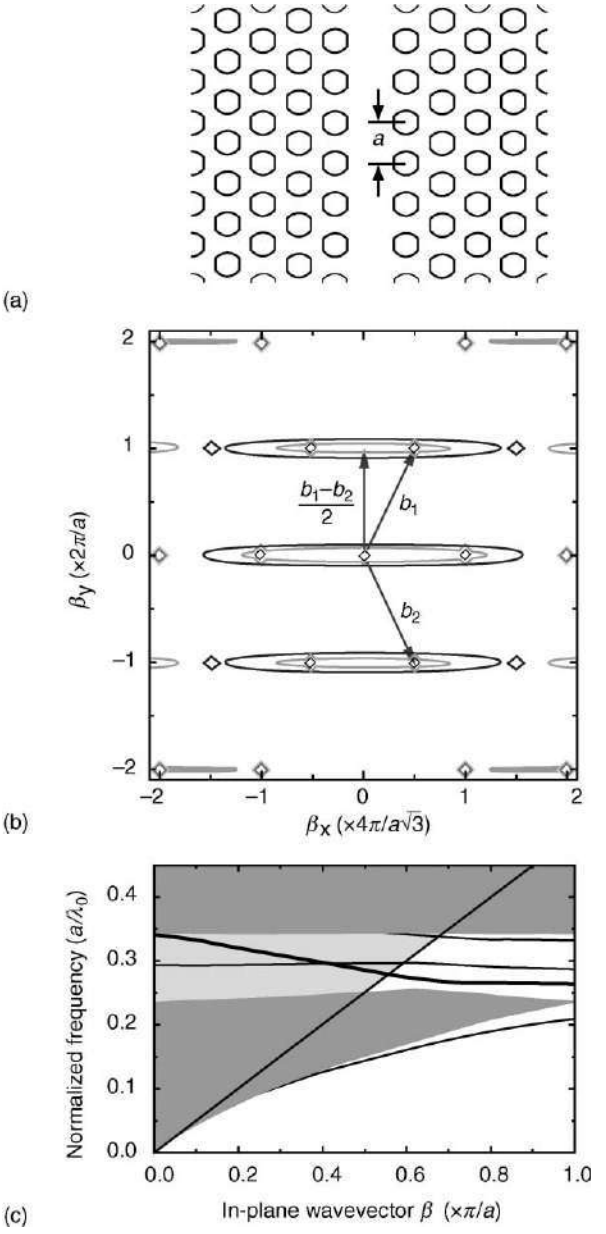

Figure 6 (a) The top view of a triangular lattice photonic crystal single-line defect waveguide with air hole radius $r/a = 0.3$. (b) The 2D spatial Fourier transform of the waveguide dielectric distribution. (c) The photonic band diagram for a suspended membrane single-line photonic crystal waveguide. The membrane has a refractive index of $n = 3.4$ and its thickness is $d/a = 0.6$. The surrounding material is air. The light gray and dark gray area indicate the regimes where field radiates through the air and photonic crystals, respectively.

photonic crystal waveguides are capable of supporting small bending radius waveguide bends with almost total transmission. This makes photonic crystal waveguides a candidate for waveguides in densely integrated photonic circuits.

Figure 6a shows an illustration of the top view of a two-dimensional photonic crystal waveguide.

This is a triangular lattice in which one row of unit cells has been modified. Shown in Figure 6b is the Fourier transform of index of refraction for the situation in which the photonic crystal consists of low index holes in a high index semiconductor. The presence of the linear defects modifies and broadens the reciprocal lattice from a simple reciprocal triangular lattice, before the creation of linear defect, to the distribution shown in the figure. This Fourier transform has components along the direction of propagation at $(\vec{b}_1 - \vec{b}_2)/2$ as shown in the figure where $\vec{b}_1$ and $\vec{b}_2$ are reciprocal lattice unit vectors. These reciprocal lattice components at $(\vec{b}_1 - \vec{b}_2)/2$ couple a planewave with wavevector β to other planewaves with wavevectors $\beta \pm (\vec{b}_1 - \vec{b}_2)/2$. The result of this coupling is that the electric field of the waveguide mode is a Bloch wave. In a photonic crystal waveguide in which the waveguide axis is the z-direction, the Bloch wavefield in the waveguide can be written in the form:

$$\vec{E}(x,y,z,t) = \sum_n c_n \vec{f}_\beta(x,z) \times \exp\left[\mathrm{i}\left(\omega t - \left[\beta + n\left(\frac{\vec{b}_1 - \vec{b}_2}{2}\right)\right]y\right)\right] \quad [3]$$

In this expression the periodic function of the Bloch wave has been explicitly expanded as a Fourier series and each term in the Fourier series is called a spatial harmonic. Note that these fields differ from that of a photonic crystal fiber because the cladding of a photonic crystal waveguide has a periodicity along the direction of a waveguide. A photonic crystal fiber does not.

Most of the realizations of photonic crystal waveguides occur in two-dimensional photonic crystals formed in high index dielectric slabs. The membrane is usually located a few microns above a high index substrate. Because the fields of the guided modes decay exponentially outside the membrane, the effect of a substrate several microns away is very nearly negligible. The photonic band structure for $r/a = 0.3$, normalized membrane thickness $d/a = 0.6$ and refractive index of $n = 3.4$ is shown in Figure 6c. Only the modes with even symmetry along the mid-plane of the membrane are shown. A bandgap opens for modes with odd symmetry only for large low index filling factors in the triangular lattice. Filling factors large enough to create a bandgap for the odd modes are often impractical so that the odd modes are rarely considered. The even and odd modes are orthogonal, so that plotting the dispersion relation for the even modes only is justified. In reality, even and odd modes may be weakly coupled as a result of fabrication imperfections. The horizontal axis of the dispersion relation covers the in-plane wavevector β along the propagation direction of the irreducible Brillouin zone. The vertical radiation light cone and transverse radiating region of the photonic crystal are mapped as light gray and dark gray areas, respectively. They correspond to the regimes in which light radiates through the air cladding and photonic crystal cladding, respectively. Waveguide dispersion relations can be calculated using two-dimensional approaches in which case the guiding due to the high index slab is accounted for by using an effective index of refraction for the mode. The effective index is the ratio of the propagation coefficient to the free space wavevector of the guided mode of the slab. The two-dimensional

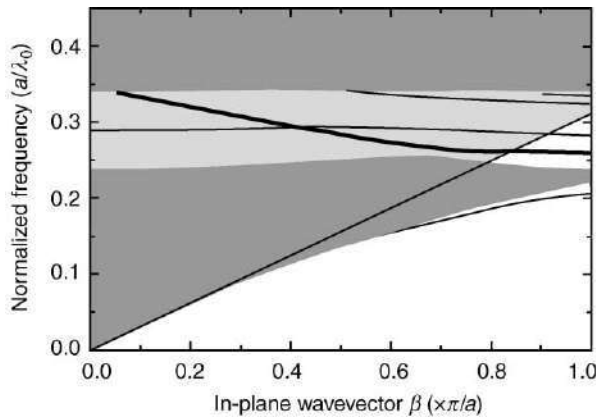

Figure 7 The photonic band diagram for a triangular-lattice sapphire-bonded single-line photonic crystal defect waveguide. The dielectric membrane has a refractive index of $n = 3.4$ and thickness of $d/a = 1.0$. The index of the sapphire bottom cladding is assumed to be 1.6.

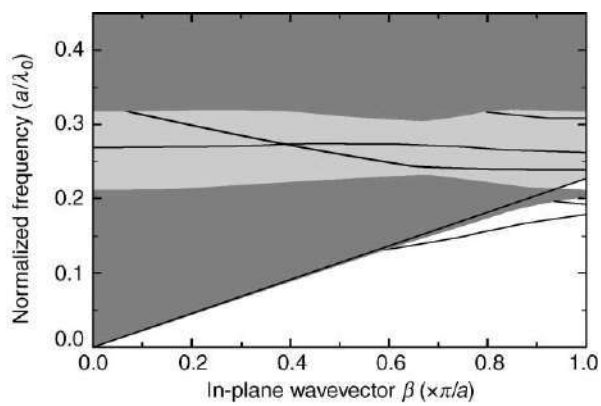

Figure 8 The photonic band diagram for a triangular-lattice deeply etched single-line photonic crystal defect waveguide. The top cladding, guiding membrane, and bottom cladding have an index of refraction of 3.0, 3.4, and 3.0, respectively. Their normalized thicknesses d/a are 1.0, 1.0, and 6.0, respectively.

calculation then uses the effective index of the mode as the high index of material in the photonic crystal. To obtain accurate results, the effective index for each mode calculated separately and the two-dimensional calculation of the photonic crystal waveguide dispersion must be repeated for each waveguide mode.

Examples of photonic crystal waveguide dispersion relations are shown in Figures 7 and 8 for waveguides formed by photonic crystals in dielectric slabs on high index lower cladding layers. Figure 7 is the dispersion relation for an oxide lower cladding layer. The area of the light gray region, which is the projection of the oxide cladding light cone onto waveguide propagation direction, increases as a result of higher index of refraction. Figure 8 shows the dispersion relation for a waveguide with an even larger lower cladding index. In this case the air holes of the two-dimensional photonic crystal are patterned all the

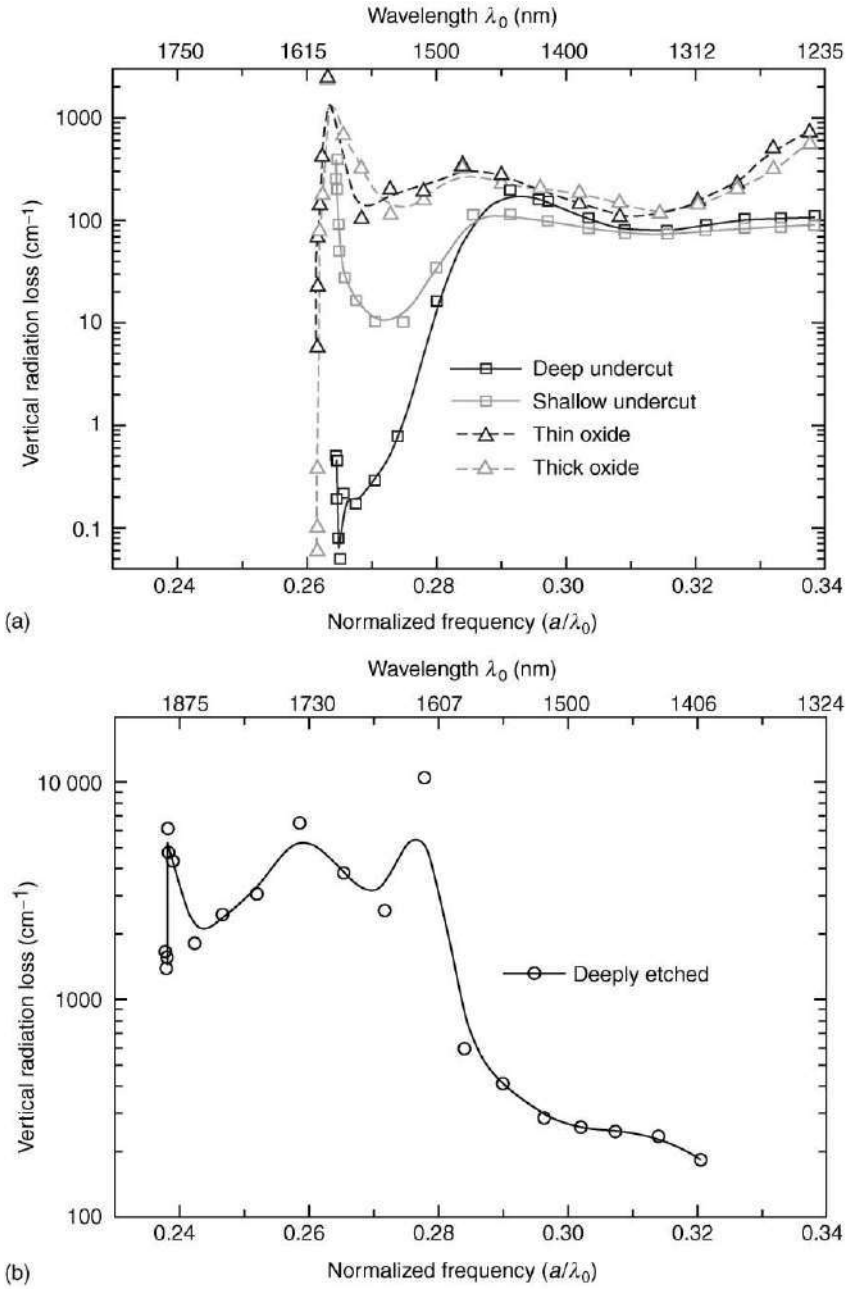

Figure 9 The out-of-plane radiation loss as a function of normalized frequency for the photonic crystal defect slab waveguides illustrated in Figures 6–8. The lattice constant $a = 450$ nm for deeply etched waveguide and 420 nm for the rest of the cases. Points are calculated values and lines are B-spline curve fits.

Table 1 Photonic crystal defect slab waveguides considered in the calculations

Waveguide structures	*Suspended membrane*		*Oxidized lower cladding*		*Deeply etched*
	Deep undercut	*Shallow undercut*	*Thin oxide*	*Thick oxide*	
Layer description: Material (normalized thickness d/a)			Air (>3.0)		
	–	–	–	–	AlGaAs (1.0)
	GaAs (0.6)	GaAs (0.6)	GaAs (0.6)	GaAs (0.6)	GaAs (1.0)
	Air (3.0)	Air (1.0)	Al_xO_y (2.0)	Al_xO_y (5.0)	AlGaAs (6.0)
			GaAs substrate (>3.0)		

Refractive index of 3.4 is assumed for GaAs, 3.0 for AlGaAs, 1.6 for Al_xO_y, and 1.0 for air.

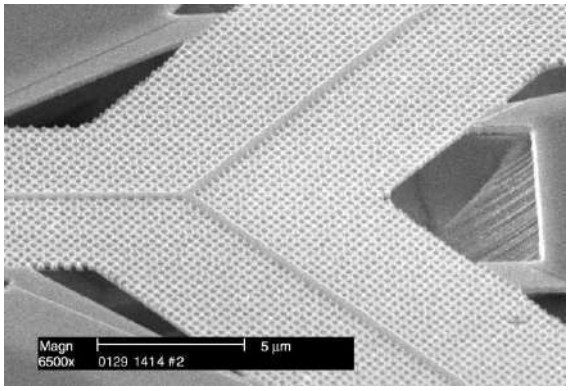

Figure 10 The scanning electron micrograph of a photonic crystal Y-branch fabricated on InP material system.

way through the lower cladding layer, with an index of 3.0, on a high index substrate. The defect modes lying within the in-plane photonic bandgap are all above the cladding light line in this case. Propagation losses are predicted to increase by over two orders of magnitude from the structure with very low index cladding layers shown in Figure 6 to the high index lower cladding structure of Figure 8.

Figure 9a and b shows the results of a fully three-dimensional calculation for the propagation loss of the three structure shown in Figures 6–8. The details of these structures are shown in Table 1. This calculated loss is due to coupling between the guided mode and the radiation modes of the high index substrate. A wavelength scale is also included using $a = 420$ nm in Figure 9a and 450 nm in Figure 9b. The suspended membrane suffers the least vertical radiation loss among all of the waveguides considered and has a minimum loss around 0.2 cm^{-1}. Low-loss waveguides formed on high index cladding layers may be possible, but care should be taken to eliminate coupling between the guided modes and the radiation modes of the cladding. In this case, the design strategy for reducing the optical loss is to reduce the magnitude of the Fourier component of the electromagnetic field inside the light cone. It should also be noted that deeply-etched waveguides formed by removing multiple rows of holes have been demonstrated with significantly lower radiation loss than is predicted for the single line defect waveguides. However, these waveguides are multi-moded at all frequencies. It is also important to remember that the nanofabricated waveguides will suffer additional losses due to fabrication imperfections that were not included in the numerical model. Figure 10 shows a nanofabricated photonic crystal waveguide. The image is an electron micrograph of a y-branch component of a Mach–Zehnder interferometer formed in a two-dimensional photonic crystal waveguide patterned into a suspended InGaAsP membrane.

Finally, it is worth noting that the group velocity of photonic crystal waveguide modes, where the group velocity is given by the gradient of the dispersion surface:

$$v_g = \nabla_{\vec{\beta}}\omega \qquad [4]$$

is very small near the zone boundary. This can be seen from the dispersion relations shown in Figures 6–8. In fact, the slope of the dispersion relations is likely to be a parameter which can be engineered by careful waveguide design. This property may allow photonic crystal waveguides to find applications in optical processing functions such as delay lines.

See also

Photonic Crystals: Atomic Physics; Electromagnetic Theory.

Further Reading

Baba T (1997) Photonic crystals and microdisk cavities based on GaInAsP–InP system. *IEEE Journal of Selected Topics in Quantum Electronics* 3: 808–830.

Joannopoulos JD, Meade RD and Winn JD (1995) *Photonic Crystals: Molding the Flow of Light*. Princeton: Princeton University Press.

Mekis A, Chen JC, Kurland I, Fan S, Villeneuve PR and Joannopalos JD (1996) High transmission through sharp bends in photonic crystal waveguides. *Physics Review Letters* 77: 3787–3790.
Painter O, Lee RK, Yariv A, *et al.* (1999) Two-dimensional photonic crystal defect laser. *Science* 284: 1819–1821.
Sakoda K (2001) *Optical Properties of Photonic Crystals*. New York: Springer.
Slater JC (1967). *Insulators, Semiconductors, and Metals, Vol. 3*. New York: McGraw-Hill Book Company.
Srinivasan K and Painter O (2002) Momentum space design of high-Q photonic crystal optical cavities. *Optics Express* 10: 670–684.
Villeneuve PR, Fan S and Joannopoulos JD (1996) Microcavities in photonic crystals: mode symmetry, tenability, and coupling efficiency. *Physics Review B* 54: 7837–7842.
Yariv A and Yeh P (1984) *Optical Waves in Crystals: Propagation and Control of Laser Radiation*. New York: John Wiley & Sons.

Self-Assembled and Functionalized Photonic Crystals

S Bhandarkar, Alfred University, Alfred, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Photonic crystals mimic the role of the atomic structure of a semiconductor in manipulating electrons, in this case for light. Hence, called semiconductors of light, these crystals are not generally found in nature and have to be designed and fabricated. Research and development in this field is taking place at a feverish pace to both understand the various nuances of this concept, as well as to evolve new methods to fabricate such crystals in the laboratory and ultimately on a commercial scale. The driving force behind this is the tremendous potential of these systems. They promise to revolutionize the way light can be used for future devices: from forming perfect, omnidirectional mirrors to the smallest ever lasers, and ultimately, complex optical chips with a wide range of functionality arranged in three-dimensional (3D) circuits.

This article seeks to elaborate on a specific technique to make photonic crystals called colloidal self assembly. Here, we will review the fundamental concepts underlying photonic crystals and see how the above process can lead to a versatile platform in making such systems. We will discuss the basic aspects of this method and look at the various innovative manifestations that have been used to fabricate many types of photonic bandgap materials.

Photonic Bandgap

A key attribute of a semiconductor is the formation of a bandgap, a band of energy states that the electrons cannot acquire. They may exist in the so-called valence band where their energy levels are less than that of the bandgap. Or they may be energized to states above the bandgap, i.e., to the conduction band, but the bandgap is the forbidden region. The continuous band of states that constitutes this bandgap is a result of the long-range order imposed by the crystal structure of the semiconductor, where the atomic lattice presents a periodic potential to a propagating electron. The tremendous usefulness of semiconductors comes from their ability to perform switching and logic functions, which in turn is a result of controlling the availability of electrons above and below the bandgap.

The propagation of light as an electromagnetic wave is influenced by the dielectric properties of the medium. For a waveguide aligned in the z direction, the field is the sum of the two propagating transverse modes, each with the form $E(x) = \Phi(x, y)e^{ik_z z}$, where $\Phi(x, y)$ is the field distribution and k_z is the propagation constant. The wave equation for the transverse mode reduces to

$$\frac{\lambda^2}{4\pi^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\Phi(x, y) + n(x, y)^2\Phi(x, y) = (\bar{n})^2\Phi(x, y) \quad [1]$$

Here, $\bar{n}$ is referred to as the effective refractive index and is given by $\bar{n} = k_z\lambda/2\pi$.

Equation [1] has the same form as the Schrödinger equation, which describes the wave functions of quantum particles such as electrons. This equivalence helps us to understand the similarity between the attractive potential well structures that bind electrons to the wells created by a local increase in the refractive index that confines light into optical modes.

The creation of a structure with a 3D periodic change in the refractive index (Figure 1) opens a gap, due to Bragg scattering off the dielectric interfaces whereby a set of frequencies of light becomes forbidden to propagate in any direction – a photonic bandgap. To understand the formation of the

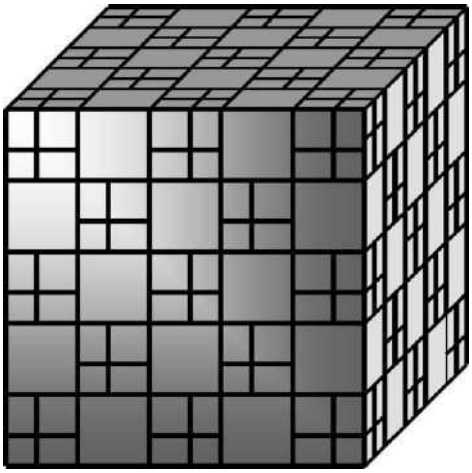

Figure 1 Schematic representation of a three-dimensional photonic crystal. The structure formed is a periodic variation in the dielectric properties of the matrix. Here, the two different dielectric regions are represented by □ and ⊞.

bandgap, one needs to consider the complex pattern of overlapping beams that results from the refractions and reflections from the numerous interfaces. This pattern is determined by the characteristics of light-waves (wavelength and direction of travel) and those of the periodic dielectric variation (pitch and magnitude of the contrast). Thus, for a select band of wavelengths, complete cancellation through destructive interference occurs in all directions.

One way to breach this restriction is by creating a defect or a disturbance in the periodicity, which would allow localized photonic states in the gap. This defect can be a point defect to form a microcavity, or have a 1D trace that behaves as a waveguide or a planar defect. Thus, a photonic crystal with a specific defect structure can be used to manipulate light in unique ways that can be used for a variety of applications, including lasers, filters, switches, waveguides, and for 3D integrated optics circuits. These structures are also interesting because they have been predicted to exhibit enhanced optical nonlinearities.

Photonic Crystals

The lattice constant for the photonic crystal must be comparable to the wavelength of light, typically in the range of 0.5 to 2 μm. This range is three orders of magnitude greater than the lattice constants found in solids and hence photonic crystals need to be synthesized. A common example of such a structure is a diffraction grating in a waveguide, which serves as a 1D photonic crystal. There are also examples of photonics crystal structures found in nature, such as the iridescent colors seen in certain butterflies and opal gemstones. The structure of opal is a 3D photonic crystal consisting of small spherical particles of hydrated silica (SiO_2). These spheres tend to be irregular in size but in rare precious opals, they are monosized. Diffraction from these spheres leads to the remarkable colors seen in opals.

Fabrication of 3D photonic crystal structures, with two materials with sufficiently different dielectric properties, has been extensively investigated over the past several years. Given the requirement of a tightly controlled periodicity, one obvious means of doing this is to use the microlithographic patterning processes that are so well developed for very large-scale integration (VLSI). Since these are planar and not volumetric techniques, several iterations of layer-by-layer deposition, lithography, and etching are necessary. A popular second method is that of colloidal self assembly.

The Self-Assembly Technique

The term 'self assembly' is usually used to describe the act of spontaneous and controlled agglomeration of some basic moieties. These basic units could be molecules, supramolecules, nanoclusters, or particles. It is similar to what is seen in living systems, where organic and inorganic materials are assembled into useful configurations mediated by various proteins. Since there is no intrinsic restriction on the number of such building blocks that can spontaneously organize into periodic structures, this process is readily scalable to large dimensions, and is the main attraction of this process.

The self-assembly technique is well suited for making periodic structures such as photonic crystals. Starting with particles of the appropriately uniform size (in the range 0.5 to 2 μm), one can conceive of organizing them into a periodic 3D array. Particles, rather than molecules or nanoclusters, are the preferred starting materials for photonic crystals because their larger size provides periodicity on the length-scale of optical wavelengths. Large block copolymer molecules have also been used instead, but they are limited in their scope.

The following sections will discuss the salient details of using self assembly to form photonic crystals. The sequence of the discussion follows the path increasing process complexity. To begin, we will examine the simplest of photonic crystals made by neatly packing uniform particles called a colloidal lattice. Given the limited range of monosized particles available, such lattices are inadequate as useful photonic crystals but can be used as templates into

which other materials can be incorporated. We shall discuss one particularly attractive incorporation technique, namely electrochemical deposition, in more detail. Finally, since the process is dominated by surface interactions, diverse classes of particles can be generated by simply functionalizing the particle surface. Looking ahead to the future, this is an important aspect that can vastly increase the capability of a colloidal assembly.

Formation of Colloidal Lattices

Colloidal crystallization has been studied for several decades but has received renewed attention due to both the availability of more uniform powders as well as new applications such as photonic crystals. Fine particles are unwieldy in their dry state due to high solid–gas interfacial energy and are typically dispersed in a liquid. Particle concentration is kept low to avoid agglomeration. Further stabilization of the particles may be achieved by either charging the particles' surface (electrostatic stabilization) or by adsorbing a neutral shielding molecule (steric stabilization). A preferred approach is to induce monosized spherical particles to settle into a tightly packed lattice. The settling rate, v, of particles of diameter d_p is given by the well-known Stokes equation:

$$v = \frac{d_p^2 a \Delta\rho}{18\mu} \qquad [2]$$

where a is the acceleration due to gravity (or a centrifugal field), $\Delta\rho$ is the density difference between the particle and the liquid, and μ is the viscosity of the liquid. Whereas settling is a popular means of self assembling, other methods such as centrifugation, filtering, pressing, or dip coating can be used (Figure 2). For charged particles, electrophoresis, where an electric field induces flow, is applicable.

The typical lattice types in these domains are hexagonal close-packed planes (hcp) as shown in Figure 3 or face-centered cubic (fcc) structures (Figure 4). In fcc crystals, the layer stacking follows the order ABCABC; in hcp it is ABAB. In both stacking sequences, hard spheres occupy a volume fraction of 0.7404 and can be shown to have little entropic differences. However, it is experimentally observed that the preferred ordering of spheres is the fcc close-packed arrangement, which is also the case in high-quality opals.

This method clearly has the advantage of being a simple, cost-effective platform to make 3D photonic crystals that can be readily scaled to various levels. There are no repetitive steps as would be the case in layer-wise lithographic pattering. However, there are still some outstanding challenges. The primary one is the difficulty in ensuring that the organization of the particles does not incur the formation of unwanted defects and domains. Also, introduction of intentional defects such as waveguides is not trivial, although this is being demonstrated in different laboratories. Another shortcoming is the

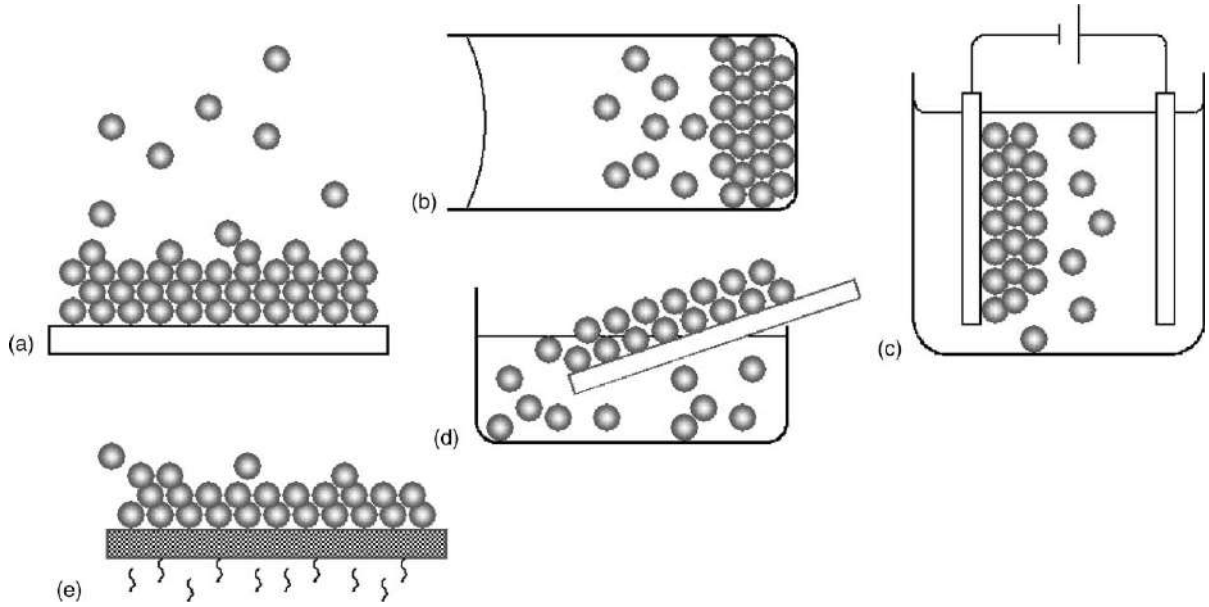

Figure 2 Diagram showing some of the techniques to use self assembly of spherical particles to make colloidal lattices. These include settling (a); centrifuging (b); electrophoresis (c); dip-coating (d); and filter-pressing (e).

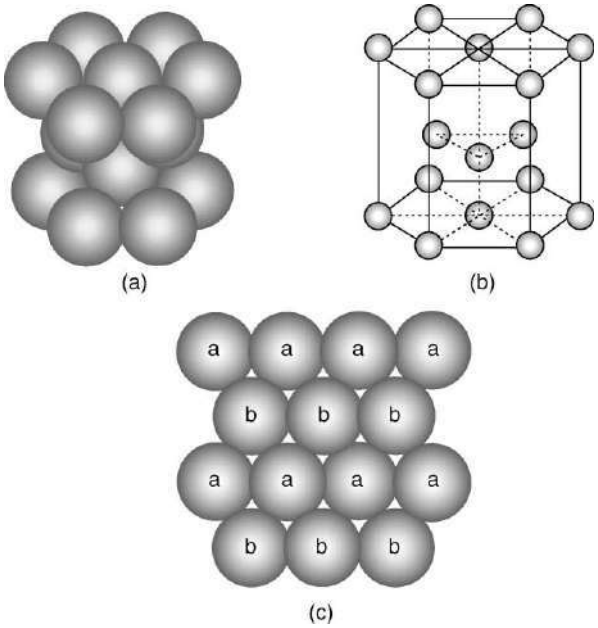

Figure 3 The arrangement of spheres in a hexagonal closed packed (hcp) arrangement (a). The unit cell is shown in (b) whereas the stacking *abab* sequence can be seen in (c).

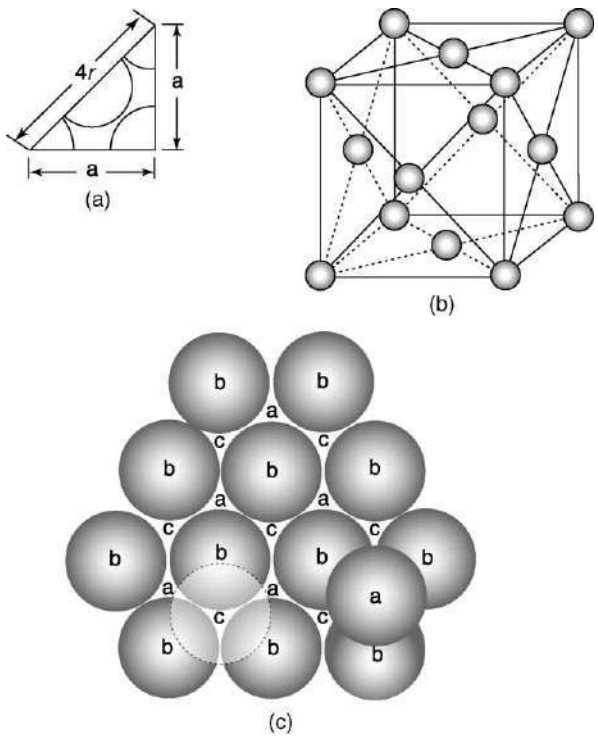

Figure 4 The arrangement of spheres in a face centered cubic (fcc) arrangement with spheres of radius *r* (a). The unit cell is shown in (b). The stacking follows an *abcabc* order. The *a* and *c* sites relative to spheres populating the *b* sites are seen in (c).

unavailability of uniformly sized particles of various different chemical compositions.

The formation of monosized spherical particles is a complex problem in its own right. Usually, such particles can be made from specific polymeric materials such as polystyrene or latex, whereas SiO_2 is the best-known inorganic system. These particles are not perfectly uniform but the typical deviations from the mean size of a few percent are acceptable for most applications. Given the attractiveness of photonic crystals for a wide array of applications in optics, research in this field is burgeoning and more breakthroughs are bound to come. Already first-generation photonic bandgap devices, such as fiber gratings or photonic crystal optical fiber, are commercially available.

Colloidal Lattices as Photonic Crystals

Due to the limited range of composition of the particles available, the particle-to-air refractive index contrast provided by straightforward colloidal lattices is small. This results in low diffraction efficiencies and hence colloidal lattices do not have photonic bandgap in all directions. However, these structures can then be used as templates for various additional modifications, as seen later.

On the other hand, a simple colloidal lattice can be more easily modeled. An analytical solution can be achieved for this system using dynamical diffraction theory for a low diffraction efficiency approximation. For a close-packed stacking of spheres in a fcc array, there exists a minimum in the transmission band at a wavelength λ_c which is given by the Bragg–Snell equation:

$$\lambda_c = 2d_s\sqrt{n_{\text{eff}}^2 - \sin^2\theta} \qquad [3]$$

where d_s is the d-spacing between the planes in the $\langle 111 \rangle$ direction, n_{eff} is the effective refractive index of the assembly and θ is the angle between the incident beam and normal to the (111) plane. This minimum in the transmission is referred to as the primary stop-band.

Photonic crystals are characterized using a Michelson or Mach Zehnder interferometer to obtain the phase and amplitude of the transmission (or reflection) as a function of the wavelength of light (Figures 5 and 6). Ideally, the whole band structure can be calculated from these data. In addition to λ_c, the close-packed structures also exhibit a transmission threshold at a wavelength λ_t, as seen in Figure 5. Both λ_c and λ_t depend on the particle diameter and increase with particle size. In a close-packed structure,

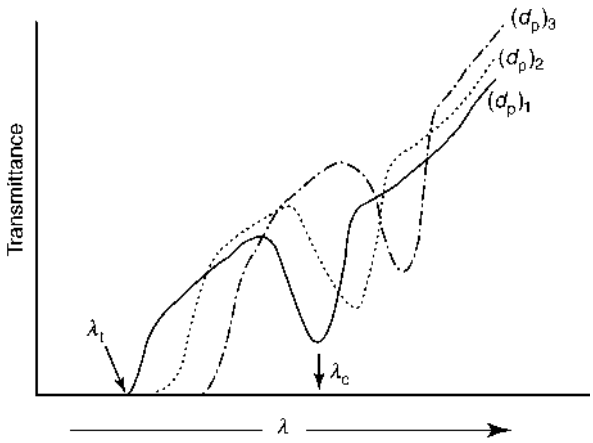

Figure 5 The typical profiles observed from transmission measurements on a photonic crystal. The stopband wavelength (λ_c) increases with d_p, where $(d_p)_1 > (d_p)_2 > (d_p)_3$. The polarization of light is fixed at *s* or *p* polarized.

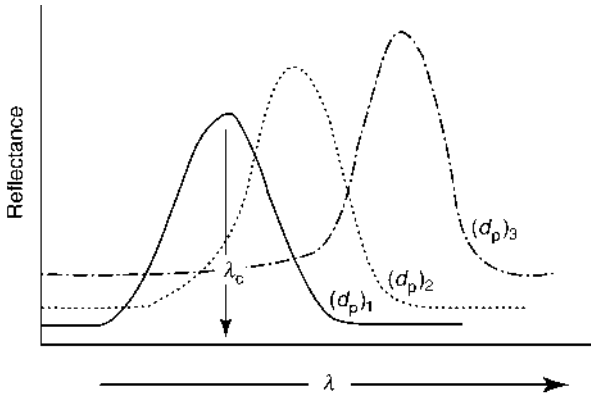

Figure 6 The typical profiles of the reflectance peaks corresponding to the transmission plot seen in Figure 5.

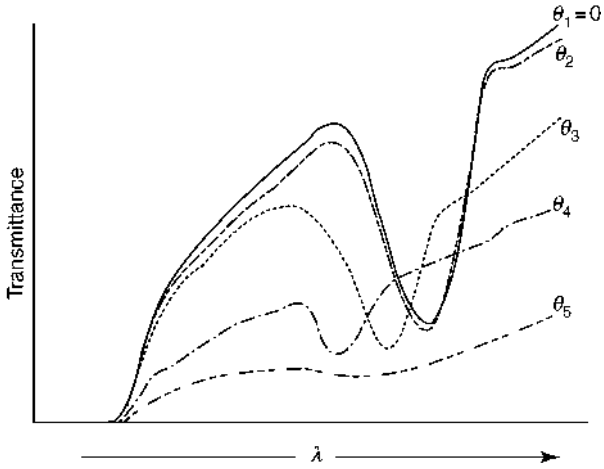

Figure 7 The transmittance profiles as a function of the angle of incidence of light of a fixed polarization. Typical values of θ would be $\theta_1 = 0$, $\theta_2 = 10°$, $\theta_3 = 30°$, $\theta_4 = 50°$, $\theta_5 = 60°$.

the lattice *d*-spacing d_s is related to the particle diameter d_p as $d_s = d_p\sqrt{2/3}$. So when the incident light is orthogonal to the structure, eqn [3] can be simplified to:

$$\lambda_c = 2\sqrt{2/3}(d_p n_{eff}) \qquad [4]$$

Thus, n_{eff} can be obtained from the slope of a plot of the stop-band wavelength λ_c as a function of the particle diameter d_p. The angular dependence of the reflectance spectra is illustrated in Figure 7. These results depend upon the polarization state of the incident light and hence need to be obtained separately for *s* and *p* polarized light.

More Complex Colloidal Lattices

While the use of SiO_2 and polymer microspheres is ubiquitous, they are optically passive materials. New methods are being developed for making uniform spherical particles from other more optically active systems. The zinc oxide (ZnO) system is one such example. The higher refractive index (~2.2) and the light emission ability of this system make it particularly interesting. Monosized particles in the size range of 100 to 600 nm can be made using a two-stage reaction method, which allows extra control over the particle formation. Other optically active systems include rare-earth doped SiO_2, dye doped PMMA spheres, or composite spheres comprised of core–shell configuration with SiO_2 or zinc sulfide (ZnS) cores and gold (Au) or SiO_2 shells, respectively.

Whereas the above discussion focuses on monosized particles, binary crystals can be made using spheres of two different sizes. The strategy here is to vary the particle size during the deposition of every alternate layer. For instance, a first layer of closely packed spheres is followed by the deposition of layers of smaller particles. The resultant crystal structure depends primarily on the concentration of the smaller particles in the dispersion. As seen in Figure 8, the second layer can vary from having three to six particles around every underlying larger particle. Thus the large-small (LS) stoichiometry of these binary crystals can be varied from LS to LS_2 to LS_3 by increasing the concentration of the smaller particles in the suspension phase. Such crystals open up interesting new avenues for complex photonic crystals.

Colloidal Lattices as Templates

By using the colloidal lattice as porous host for a second material, a greater range of photonic crystals can be synthesized. Here, the role of the colloidal array

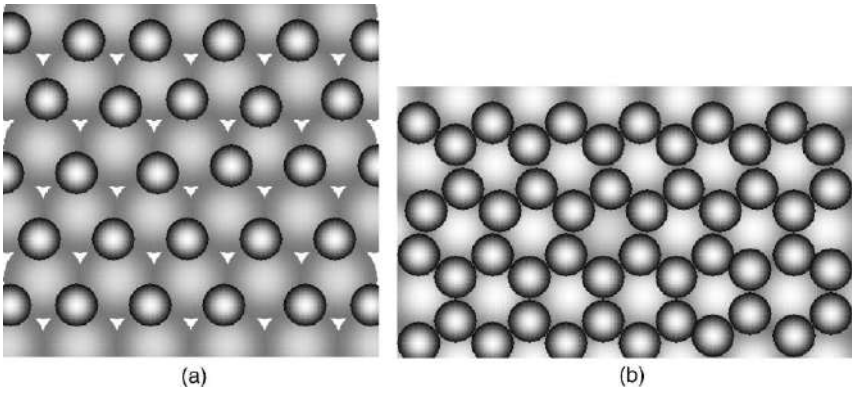

Figure 8 Binary crystals formed using spheres of two different sizes. The large-small (LS) ratio in (a) is 1:1 whereas in (b) it is 1:2.

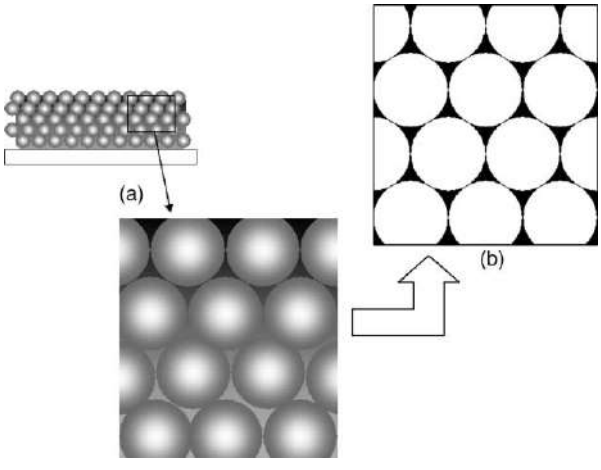

Figure 9 Schematic representation of the templating process using infiltration (a). Upon removal of the particles, a macroporous photonic crystal is formed (b). This is the inverted opal configuration.

is simply as a scaffolding structure. The second material to be infiltrated may be either a liquid (in the form of a dispersion, a solution, or a molten liquid) or a vapor. Both eventually deposit a solid material inside the template. Subsequent selective removal of the template yields a 3D photonic crystal in the form of an inverted opal (Figure 9). This is one of the best and most versatile means of achieving high index contrast, which is required for true photonic bandgap behavior.

It must be noted that to use self-assembled particle arrays as a template, the whole structure has to be often heat treated to confer greater mechanical strength. As assembled, the particles are normally attached by short-range Van der Waals forces and hydrogen bonding, which in the case of SiO_2 is silanol (Si–OH) mediated. Both of these are relatively weak bonds. By heating the assembly, a greater number of siloxane (Si–O) bonds can be formed which additionally strengthens and stabilizes the particulate network. This is important to retain the periodic structure of the assembly during the exposure to high stresses during the infiltration step, for instance. At temperatures around 750 °C, there is significant generation of siloxane bonds but minimal sintering so that the lattice does not change dimensions and retains its original open porosity. With latex particles, the annealing is done just above the glass transition temperature T_g of the polymer for no more than a critical period of time. Longer treatments can destroy the ordering of the spheres.

A template thus strengthened by heat treatment is now ready for infiltration. For example, a template comprised of polystyrene spheres can be infiltrated with a hydrolyzed sol of titanium alkoxide to ultimately form titania in the interstitial regions. Upon standing, the alkoxide sol gels in the interstitial regions leading to a polymeric network called a xerogel. This entire composite can then be further heated to high temperatures (>300 °C). All organics are removed by decomposition and combustion so as to leave behind a titania matrix with periodic voids (in place of the polystyrene spheres). The templating spheres may also be dissolved, rather than combusted, in solvents such as toluene. The refractive index of titania depends on the crystalline phase that results after heat treatment – anatase is ~2.5 while rutile is ~2.9. High indices can also be achieved with semiconducting materials such as selenium (Se) with an index of ~2.5 or cadmium selenide (CdSe), ~2.9. These chalcogen-derived materials have low absorption in the visible and near infrared, so they are suitable candidates.

When SiO_2 particles are used for templating, they are eventually removed by etching, typically with dilute, buffered hydrofluoric acid at low temperatures. Given the thermal and chemical stability of the SiO_2 particles, molten Se (or other low melting

metals) can be used to rapidly fill the interstitial space. Vapor phase deposition has also been used to fill the porous templates. Examples include the backfilling with silicon (Si) using chemical vapor deposition (CVD), or indium phosphide (InP) using metal-organic chemical vapor deposition (MOCVD).

A variety of structured, macroporous carbon matrices has been made with remarkable optical properties. The templates are artificial opals assembled using SiO_2 microspheres, which are filled with carbon by either: (a) infiltrating with phenolic resin followed by anaerobic pyrolysis of the resin to glassy carbon; (b) CVD of graphitic carbon; and (c) growing diamond-seeded CVD graphite from a plasma. The resultant carbonaceous 'inverse opals' are highly conductive, show intense opalescence due to the ordered arrays of holes, and have a photonic bandgap in the infrared region.

Electrochemical Deposition

One particularly effective backfilling technique is electrodeposition, which has been successfully used to deposit most of the different classes of materials – semiconductors, metals, and polymers. Here, a conducting substrate serves as the appropriate electrode where deposition occurs. By placing a colloidal lattice template on this electrode, the electrochemical deposit is used to fill the interstitial voids in the template (Figure 10). This method has been successfully applied for a host of different metals such as gold (Au), nickel (Ni), platinum (Pt), and alloys such as tin–cobalt (Sn–Co). Likewise, conducting polymers can be efficiently deposited by this method. Although these polymers suffer from inherently low indices, they have favorable electro-optic properties that can be tuned by controlling the composition. In the case of II–VI semiconductors, galvanostatic or potentiostatic deposition can be used: for instance, interstitial CdSe has been deposited (Figure 10) using an acidic solution of electrolytes such as cadmium sulfate ($CdSO_4$) and selenium dioxide (SeO_2).

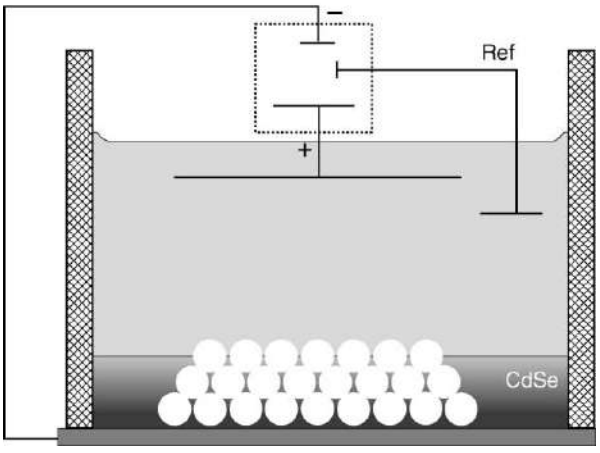

Figure 10 An electrochemical cell for deposition of CdSe from an electrolytic solution on a conducting electrode. The solution contains sulfuric acid (H_2SO_4), $CdSO_4$, and SeO_2. A potentiostat is used to generate the electric fields.

Electrochemically deposited structures have minimal porosity leading to densities close to bulk density. Hence, they have a higher refractive index and are generally more robust than those made by simple infiltration. Also, they exhibit little change in dimensions and a lower tendency toward cracking during the various physicochemical processes such as drying and sintering that occur en route to forming the photonic crystal. Hence, it is a preferred technique in many cases.

Functionalization

The self-assembly process, whether used directly or as a template to synthesize a photonic crystal, is dominated by surface interactions between particles starting from the organization process through to the thermal stabilization step. These interactions can be tailored by specially treating the surface of the particles with functional groups. Functionalized colloidal templates can enhance the infiltration of the high index material and its properties such as smoothness, density, or optical nonlinearity. Latex spheres can be surface functionalized with sulfate and carboxylic acid groups to increase their surface charging capability, which can lead to more robust self-organization. These spheres can also be coated with a layer of SiOH, which gives them a SiO_2-like appearance on the surface and allows them to be used in conjunction with SiO_2 microspheres. The infiltration of metal (such as Au) nanoparticles into a template crystal of SiO_2 microspheres can be significantly enhanced by functionalizing the SiO_2 surface. This can be achieved using long-chain organic molecules with polar groups on one end that bond with silanols on the SiO_2 surface and thiol groups on the other that strongly latch onto the gold nanoparticles. Simulations of light interaction with ordered three-dimensional arrays of these metallodielectric spheres, show that they are promising starting points for the construction of full photonic bandgap materials in the visible part of the optical spectrum. Also, in these systems, the plasmon resonance of the Au shell particles can be controlled over the visible and infrared region by varying the thickness of the Au shell.

The site-selective attachment ability of biological moieties can also be used effectively here by attaching them onto particle surfaces. DNA oligomers are well known to form conjugate specific bonds – adenine to thymine and guanine to cytosine. Thus particles functionalized with these oligomers can be made to organize in specific sequences that are differentiated, depending on the surface group. These particles can have variable sizes or compositions but their surface properties are controlled largely by the attached biomolecules. Likewise, the antigen–antibody attachment ability can be exploited to improve surface bonding.

Positioning Defects in Photonic Crystals

Much of the work on photonic crystals involves the making of a long-range periodic microstructure with high dielectric contrast because the underlying materials engineering aspects are very challenging. However, the usefulness of photonic crystals depends on the successful placement of defects in precise locations within the crystal. The placement of a point defect leads to a photonic microcavity and that of a line defect, a waveguide for propagation of a light-wave (Figure 11). However, there are certain unique advantages to this style of propagation, such as the possibility of bending light around sharp corners. Line defects in the form of air cores have been introduced in a photonic crystal using a combination of lithography and self assembly. First, ridges are defined in a photoresist on top of the substrate by optical lithography (Figure 12). Upon forming a colloidal lattice on top, the ridges are converted to linear voids by removing the photoresist using combustion or etching. Another technique is to infiltrate a photopolymerizable monomer into a colloidal crystal and preferentially polymerize the liquid by focusing a light beam inside the crystal. Thus the locus of the focused beam defines the waveguiding region that is formed upon removal of the unpolymerized liquid.

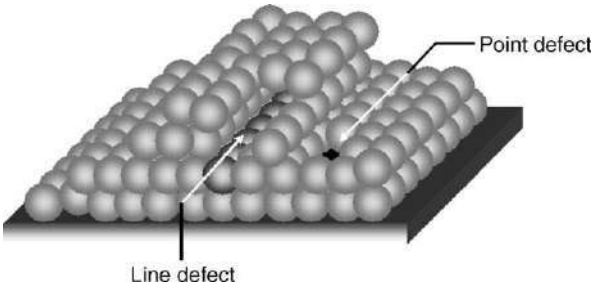

Figure 11 Two common defects in colloidal lattices.

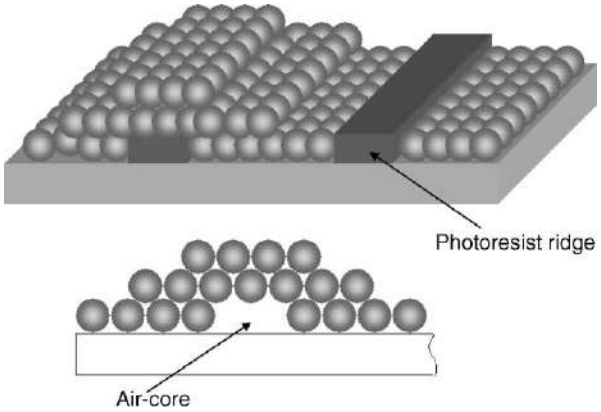

Figure 12 Formation of intentional defect using a lithographic process to define an air-core. This structure can also be used as a template for backfilling a high dielectric material.

Tunable Photonic Crystals

The holy grail in many photonic devices is tunability, and photonic crystals are no exception. Given the wide spectral range that constitutes the domain of light, the ability of devices to switch or translate their performance to another wavelength through an external stimulus is indeed very attractive. The two parameters that can be externally tuned are the dielectric contrast, which would result in a change in the n_{eff} of the crystal and the periodicity of the grating. One technique to tune the former is to form an interference pattern that organizes liquid crystals dispersed in a polymer matrix into droplets to form the photonic crystal. Applying an electric field that changes the refractive index of the droplets can therefore alter the transmission spectrum of the photonic crystal.

To control the lattice parameter, the size of or the spacing between the spheres need to be varied. This concept has also been demonstrated using self-assembled hydrogel-based nanoparticles, which are made of a thermo-responsive material such as poly-*N*-isopropylacrylamide. These particles are soft and conformable and their sizes can be tuned depending on the water content of the system. This ordered matrix is in the form of a gel that can be made to swell and compress upon changing the temperature as water is exchanged between the gel and the ambient. Wavelength tunability of one nanometer has been achieved this way. Tunability has also been sought using methods such as straining the lattice. Calculations show that a 2 or 3% shear strain ought to create a 52–73% distortion of the bandgap. In one novel approach, SiO_2 or polystyrene microspheres were strongly surface charged and ordered with a less than dense packing within a poly-acrylamide

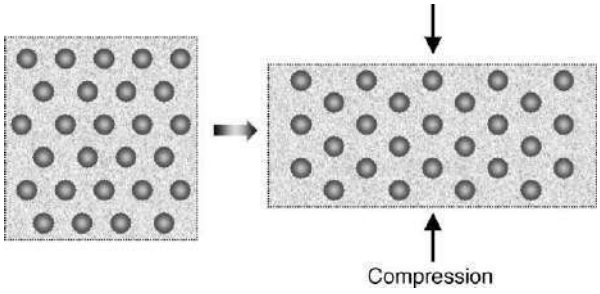

Figure 13 Tunable photonic crystal. Spherical particles are assembled in a compressible media like a hydrogel, which deforms upon applying a stress. The deformation can be used to tune the lattice parameter and hence the optical properties of the photonic crystal.

hydrogel housing (Figure 13). Due to the strong charging the arrangement tends to be body centered cubic (bcc) with the lattice spacing that depends on the concentration of the particles. When the hydrogel is deformed by applying mechanical stress, the diffraction peaks were shown to be tunable over almost the whole visible light wavelength spectrum.

Photonic crystals made from self-organized hydrogels have been used as sensors. Here, they are functionalized with chemical recognition agents that cause the gels to swell upon binding to the analyte. The resultant change in the lattice parameters causes a measurable change in the color. Specific examples include the detection of the toxic organophosphate, parathion, using acetylcholinesterase or of glucose using boronic acid and polyethylene glycol. Utilizing the same principle, even metal ions such as copper can be detected using a hydroxyquinoline-based hydrogel. These serve as a good illustration of the multifaceted potential of photonic crystals.

Summary

Photonic crystals have aroused tremendous interest with their promise to lift the field of photonics to a new level of capability. These crystals are conceptually analogous to electronic semiconductors in their function and hence are called photonic bandgap materials. The big challenge lies in synthesizing them artificially without defects. One of the few methods to synthesize these crystals is by self assembly of micron-sized building units. Self assembly leads to formation of a uniform array of monodispersed particles, typically silica or latex. This process can be further embellished by permeating a second, high-index material using the particulate assembly as a template. Useful applications of photonic crystal require them to have specifically positioned defects and tunability. These aspects, which currently constitute the state-of-the-art research in this field, promise to deliver unique devices and applications.

List of Units and Nomenclature

a	Acceleration (typically due to gravitational or applied centrifugal force)
d_p	Diameter of the particle
d_s	d-spacing or distance between lattice planes
E	Electric field
k_z	Propagation constant for travel in z direction
n	Refractive index of the medium
$\bar{n}$	Modal effective refractive index
n_{eff}	Effective refractive index of the colloidal assembly
v	Settling velocity
$\Delta\rho$	Density difference between the particle and the fluid in which it is suspended
θ	Angle between the incident beam and normal to the lattice plane in question
λ	Wavelength of the lightwave
λ_c	Stopband wavelength
λ_t	Threshold wavelength for the photonic crystal
μ	Viscosity of the suspending fluid
Φ	Electric field distribution

See also

Photonic Crystals: Atomic Physics; Electromagnetic Theory; Photonic Crystal Lasers, Cavities and Waveguide.

Further Reading

Born M and Wolf E (1981) *Principles of Optics: Electromagnetic Theory of Propagation Interference and Diffraction of Light*. Oxford, UK: Pergamon Press.

Braun PV and Wiltzius P (2002) Macroporous materials – electrochemically grown photonic crystals. *Current Opinion in Colloid and Interface Science* 7: 116–123.

Graf C and Van Blaaderen A (2002) Metallodielectric colloidal core-shell particles for photonic applications. *Langmuir* 18(2): 524–534.

Iwayama Y, Yamanaka J, Takiguchi Y, *et al.* (2003) Optically tunable gelled photonic crystal covering almost the entire visible light wavelength region. *Langmuir* 19(4): 977–980.

Joannopoulos JD, Meade RD and Winn JN (1995) *Photonic Crystals*. Princeton, NJ: Princeton University Press.

Johnson SG and Joannopoulos JD (2002) *Photonic Crystals: The Road from Theory to Practice*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Kazuaki S (2001) *Optical Properties of Photonic Crystals*. New York: Springer-Verlag.

Koenderink AF, Johnson PM, Galisteo López JF and Vos WL (2002) Three-dimensional photonic crystals as a cage for light. *Comptes Rendus Physique* 3: 67–77.
Soukoulis CM (ed.) (2001) *Photonic Crystals and Light Localization in the 21st Century*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Stein A (2001) Sphere templating methods for periodic porous solids. *Macroporous and Mesoporous Materials* 44–45: 227–239.
Velev OD and Lenhoff AM (2000) Colloidal crystals as templates for porous materials. *Current Opinion in Colloid and Interface Science* 5: 56–63.
Xia Y, Gates B, Yin Y and Lu Y (2000) Monodispersed colloidal spheres: old materials with new applications. *Advanced Materials* 12(10): 693–713.
Yablonovitch E (1987) Inhibited spontaneous emission in solid-state physics and electronics. *Physics Review Letters* 58(20): 2059–2062.

PHYSICAL APPLICATIONS OF LASERS

Contents

Free-Electron Lasers in Physics
Industrial Applications
Sum-Frequency Generation at Surfaces

Free-Electron Lasers in Physics

T Dekorsy, Forschungszentrum Rossendorf, Dresden, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The topic of this article is a survey on the main applications of free-electron lasers (FELs) in physical sciences. The merits of FELs are discussed with respect to other available laser sources. For the wavelength ranges, where FELs contributed significantly to advances in physical sciences, specific examples of the research with FELs are given.

The basic physics of FELs, their historic development, and the state of the art of these laser sources are described in detail in this Encyclopedia (*see* **Lasers:** Free Electron Lasers). The present article covers the applications of FELs in physical sciences, but excludes all phenomena involved in the physics of FELs themselves. From the application point of view, one should regard an FEL as any other laser, as a source of coherent electromagnetic radiation with high power, narrow bandwidth, and a diffraction limited beam. Thus, the radiation produced by an FEL significantly differs from that generated by synchrotrons (*see* **Incoherent Sources:** Synchrotrons). The specifications of the FEL, such as wavelength, temporal structure of the output intensity, and peak power, determine the possible applications of the laser. One might now ask, why it is worthwhile to write – and read – an article on its own devoted to the applications of FELs. The answer to this question is the uniqueness of the radiation produced by FELs, which enables certain classes of experiments and applications in specific wavelength ranges not covered by other laser sources, such as gas, solid state, or semiconductor lasers.

Since the construction and operation costs of an FEL facility are usually orders of magnitude higher than those for the equipment of a conventional laser spectroscopy laboratory, FEL facilities have been built, and are planned for the future, in different places all over the world as dedicated-user facilities. These user facilities allow researchers from universities and industry to exploit the unique properties of FEL radiation for their research and provide the basic infrastructure and support for setting up the experiments. Usually potential users apply using a standard procedure for beam time at an FEL by submitting a short proposal to the facility. The proposals are judged by a panel which distributes the available beam time to the different users. So researchers active in the field of optics should keep in mind that these facilities exist and are available to anyone who has specific needs which can be fulfilled by an FEL facility. A list of all major FEL facilities is given in Table 1.

Table 1 FEL user facilities. Facilities marked with an asterisk are under construction at the time of publication

Facility name	*Wavelength ranges*
iFEL (Osaka, Japan)	230 nm–1.2 μm 1–6 μm, 5–22 μm 20–60 μm, 50–100 μm
Duke University Free Electron Laser Laboratory (US)	193–400 nm 2–9 μm
Vanderbilt University Free Electron Laser Center (US)	2.1–9.8 μm
FELIX–FOM (Rijnhuizen, Netherlands)	4.5–250 μm
Stanford Picosecond FEL Center (US)	3–15 μm 15–65 μm
CLIO–LURE (Orsay, France)	3–90 μm
Jefferson Lab (US)	3–6.2 μm
FEL–SUT–Science University of Tokyo (Japan)	5–16 μm
UCSB Center for Terahertz Science and Technology (US)	63–338 μm 338 μm–2.5 mm
ENEA Compact FEL (Frascati, Italy)	2–3.5 mm
FELBE (Dresden, Germany)*	5–150 μm
SCSS SPring-8 Compact SASE Source (Japan)*	X-ray
Institute for Plasma Research (India)*	sub-mm
BESSY FEL (Germany)*	deep UV–X-ray
Brookhaven–NSLS SDL DUV–FEL (US)*	deep UV
University of Maryland–MIRFEL (US)*	far-infrared
University of Hawaii (US)*	UV–far-infrared
SLAC SSRL X-ray FEL–LCLS (US)*	X-ray
TESLA X-ray FEL, DESY (Germany)*	X-ray
Daresbury 4GLS (UK)*	IR, X-ray
MIT Bates Lab X-ray FEL (US)*	X-ray
INFN SPARX X-FEL (Italy)*	X-ray
Tel Aviv University (Israel)*	far-infrared
Center for Advanced Technology (India)*	far-infrared

The basic principles of an FEL are briefly summarized here, while for an in-depth description the reader is referred to the article on free electron lasers (*see* **Lasers:** Free Electron Lasers). As with any other laser, the FEL consists of a laser cavity and a gain medium. The gain media are free electrons, which travel with a velocity close to the speed of light in vacuum through a periodic array of magnets called an undulator or wiggler. The magnets are arranged in a way that the electrons feel an alternating magnetic field. Due to the Lorentz force, the electrons perform an undulating or wiggling motion perpendicular to the magnetic field, which gives rise to spontaneous electromagnetic radiation into the forward direction. The radiation, emitted in due course of the passage of the electrons through the undulator, superposes constructively. The FEL may operate in such a way that the gain from one passage of an electron bunch is high enough to produce a coherent light pulse. This regime is known as self-amplified spontaneous emission (SASE) regime which is especially desirable for short wavelength operation in the UV and X-ray region, where the manufacturing of adequate resonator optics is difficult. At longer wavelengths, a laser cavity provides the feedback for the build-up of the coherent light field. The laser starts lasing when the amplification exceeds the losses of the cavity, including the light intensity coupled out of the cavity.

One of the distinct properties of FELs, in comparison to other laser sources, is their continuous tunability. The emission wavelength of an FEL is given by the following relation (*see* **Lasers:** Free Electron Lasers):

$$\lambda = \frac{d}{2\gamma^2}\left(1 + \frac{K^2}{2}\right) \qquad [1]$$

with d the period of the magnet array, the γ-parameter is the relativistic factor related to the electron velocity u via $\gamma = 1/(1 - u^2/c^2)^{1/2}$, and the K-parameter is given by d, and the magnetic field B as $K = 0.934 \times B[\mathrm{T}] \times d[\mathrm{cm}]$. From eqn [1] it becomes obvious that the emission wavelength of the FEL can be continuously tuned via the electron energy and the magnetic field. This is in distinct contrast to other laser sources, where the emission wavelength is fixed by the energy separation of electronic levels in atoms or by the bandgap energy of semiconductors.

The main advantage of FELs concerns the high power, both average and peak, which can be obtained. In addition, an FEL provides short optical pulses due to the fact that the electron beam passing through the undulator consists of short electron bunches. The temporal width of these electron bunches can be below one picosecond (10^{-12} s), so that optical pulses in the sub-picosecond to picosecond range are obtained. These pulses are extremely adjuvant for performing nonlinear optical experiments (*see* **Spectroscopy:** Nonlinear Laser Spectroscopy) and pump-probe experiments with high temporal resolution (*see* **Chemical Applications of Lasers:** Pump and Probe Studies of Femtosecond Kinetics). Typically the optical pulses are Fourier transform limited, i.e., the product of the spectral bandwidth $\Delta\nu$ times the pulse duration $\Delta\tau$ equals a constant, which depends on the specific shape of the pulse. For a pulse of Gaussian shape in frequency and time, this so called time-bandwidth product is given as $\Delta\nu\,\Delta\tau \geq 0.441$. From this equation the minimum spectral width (temporal duration) can be calculated for a given temporal duration (spectral width) of the laser pulse. When the electron beam driving the FEL is produced by a superconducting radio frequency linear electron accelerator, the repetition rate of the pulses is in the

range of some tens of MHz. The energy per optical pulse is obtained by dividing the average power of the FEL through the repetition rate. The peak power within an optical pulse is obtained by dividing the energy per pulse through the pulse duration, which can be in the picosecond to sub-picosecond range. Typical values for FELs are energies per pulse in the μJ to mJ range, pulse lengths from some 100 femtoseconds to a few picoseconds, and accordingly peak powers in the range of kW to GW. Depending on the type of electron accelerator employed, the pulse structure of an FEL may have further peculiarities. Superconducting radio-frequency accelerators provide a quasi continuous-wave (CW) train of electron pulses, thus enabling the FEL to emit a continuous optical pulse train. Electrostatic accelerators provide macropulses of electrons at some Hz repetition rates with a duration of several μs. These macropulses consist of micropulses with MHz repetition rates and picosecond or sub-picosecond duration. Hence the FEL output has the same temporal characteristics with the FEL turning on at the beginning of each macropulse.

Another important development is the synchronization of table-top short-pulse laser systems to FELs. This is accomplished by stabilizing the repetition rate of the table-top laser system via a movable intracavity mirror accurately to the repetition rate (or a higher harmonic of the FEL) of the FEL. Using a high-frequency electronic phase-locked loop, a timing jitter, down to 500 femtoseconds between the FEL pulses and those from a mode-locked titanium-sapphire laser, have been obtained. Nonlinear frequency conversion of the near-infrared pulses from the synchronized titanium-sapphire allows one to perform multi-color pump-probe measurements in a broad spectral range in combination with the FEL.

In the following, FELs will be discussed with respect to the wavelength range which they cover. For each wavelength range alternative laser sources are given and compared to FELs. It should be noted that these alternative sources evolve very rapidly and thus pose challenges on the justification of FELs in certain wavelength and power ranges. For each wavelength region selected scientific achievements obtained from experiments performed with FELs are given.

FELs in the Far Infrared Region

The far infrared region is roughly defined as the wavelength (energy) range from 20 μm to 300 μm (62 meV to 4 meV). At the long wavelength end (wavelengths larger than 220 μm) radiation can be obtained from pure electronic devices, so-called backward wave oscillators (BWO). Furthermore, optically pumped gas lasers operate at certain discrete wavelengths in this range where there exists one fundamental problem for realizing a semiconductor laser based on transitions between the conduction and the valence bands. Auger recombination, where an electron and hole combine by transferring another electron high into the conduction band, becomes increasingly important for smaller bandgaps. Recent advancements in this field are based on quantum cascade lasers, which use inter-sub-band transitions of quantum structures for achieving lasing in the wavelength range from 105 μm to 3.4 μm. Another possibility to generate coherent radiation in this wavelength range is optical rectification or difference frequency mixing of femtosecond laser pulses at near-infrared wavelengths. However, none of these sources outperform FELs in terms of peak or average power, making FELs the most important tunable laser source in this wavelength range.

In the far-infrared, several scientific opportunities exist. In condensed matter many important elementary excitations exist in this frequency range, e.g., phonon-polaritons, plasmons, polarons, and magnons. The quantum confinement in semiconductor heterostructures leads to the formation of discrete electronic levels with energy separation typically in the far-infrared. The associated radiative transitions, so-called inter-sub-band transitions, can be arbitrarily tuned by varying the materials and the dimensions of the confining potential. In surface science, intramolecular vibration of molecules adsorbed to surfaces are of large interest. They can be studied by resonant nonlinear optical experiments such as sum-difference frequency generation or second harmonic generation, giving information about the energy transfer between the molecules and the surface. In biophysics, low-frequency modes of large bio-molecules lie in the far-infrared, whose dynamics can also be studied.

Important scientific results have been obtained with FELs in the following fields, which exclude the important fields of molecular science, chemistry, biology, and medical research.

Scanning Near-Field Optical Microscopy (SNOM)

Increasing the spatial resolution is an important task in spectroscopy. Free-space focusing of a laser beam is diffraction-limited to a spot with the dimensions of the wavelength. Especially in the far-infrared region this would allow to resolve objects larger than 10 μm only. However, the imaging of elementary excitations in solid state or of collective excitation in biomolecules on a sub-μm scale, would be highly desirable. SNOM allows to circumvent the diffraction limit. Several different SNOM techniques have been

introduced in recent years, most of them working with apertures smaller than the wavelength. Imaging is performed by scanning the aperture over the sample or vice versa. An evanescent electromagnetic field is present at the aperture which can couple to the specimen brought into the proximity of the aperture. Since the aperture is smaller than the wavelength, the light intensity drops by order of magnitudes. It is here that the high power available from FELs becomes advantageous. Usually, the end of a fiber is tapered and metalized, leaving a small aperture at the tip of the fiber. In the far-infrared region the problem exists in the absorption of the fiber material. Chalcogenide fibers are capable of transmitting in the 2–12 μm range, while hollow glass capillaries coated with metals inside work also at longer wavelengths. Recent approaches try to work without apertures, e.g., by coupling the light locally via a nanometer metal tip to the sample. Another interesting approach is to use the focus of visible laser light to create a small highly reflective spot on a material. By creating free-carriers with the visible laser light in the material, the reflectivity of the far-infrared light of the FEL can be strongly increased, thus a reflecting spot is created which is diffraction-limited by the wavelength of the visible light rather than that of the infrared light. The techniques described here are, of course, not limited to the far-infrared region but may also be applied successfully, for example, in the mid-infrared region.

Rabi Oscillations

Rabi oscillations are a fundamental phenomenon in a quantum mechanical two-level system driven by a light field. The light field promotes electrons from the lower to the upper state and back, depending on the duration and intensity of the light field, i.e., the population oscillates between the lower and upper state of the two-level system with the oscillation period depending on the intensity of the light field. This phenomenon has been studied extensively in atoms. In semiconductors, scattering processes strongly aggravate the observation of Rabi oscillations. By using the Santa Barbara FEL Rabi oscillations between the ground and excited state of a hydrogenic, impurities in GaAs could be observed. From the excited state of the impurity, electrons are transferred to the conduction band, thus changing the photoconductivity of the crystal, which is measured as a function of time delay. Overdamped oscillations with frequencies changing with the laser intensity could be clearly resolved.

Optical Nonlinearities in Semiconductors

An important concept of nonlinear optics is based on the nonlinear susceptibility $\chi^{(n)}$ of a material, where n denotes the order of the nonlinearity. The nonlinear susceptibility has tensorial character and determines, for example, the efficiency of second harmonic generation and difference frequency mixing ($\chi^{(2)}$), third harmonic generation, and the Kerr nonlinearity ($\chi^{(3)}$), etc. Hence the knowledge about the dispersion and the absolute value of $\chi^{(n)}$s is both of fundamental and practical importance. Due to the lack of tunable laser sources, these values are not well known in the far-infrared. Especially close to fundamental lattice resonances in the far-infrared, a strong dispersion is theoretically expected. Here FELs play an important role in the determination of these values. For example, in GaAs, the dispersion of the $\chi^{(2)}$ has been recently determined from second harmonic generation experiments below the fundamental optical phonon resonance at 8 THz, i.e., wavelengths larger than 37 μm, where a resonance enhancement at half the phonon frequency could be observed.

Relaxation Kinetics in Semiconductor Quantum Structures

The investigations of inter-sub-band transitions in semiconductor heterostructures have opened the pathway for the realization of the quantum cascade laser. In these devices, the radiative and nonradiative transition rates between different electronic quantum levels are of prime importance for a working device. The relevant time-scales have been obtained from pump-probe experiments with FELs. In semiconductor quantum dots, the strong coupling of phonons to inter-sublevel transitions could be observed and the polaron lifetime could be determined.

Bloch Oscillations in Semiconductor Superlattices

Bloch oscillations are an intriguing concept of solid state physics. When an electron experiences the force of an electric field in a periodic potential, the resulting dynamics is oscillatory in the absence of scattering events. However, since in real solid state the scattering rates are much higher than the expected oscillation period, such oscillations cannot be observed and the application of an electric field leads to a net electric current. However, in artificial semiconductor superlattices, the Bloch oscillation period can be tailored to be smaller than the scattering rate, leading to the observation of Bloch oscillations and the associated THz emission. The inverse effect of a Bloch oscillator, i.e., the generation of a photocurrent by tuning an FEL to the Bloch oscillation frequency, could be observed.

Cluster Physics

FELs can be an important tool for studying the dynamics of clusters and molecules in the gas phase by

infrared resonance enhanced multiphoton ionization (IR-REMPI). In IR-REMPI, the species of interest are irradiated by pulsed, tunable IR radiation from a free electron laser. When the radiation is resonant with a vibrational mode of the molecule or cluster, the absorption of many – up to several hundred – photons can take place. When the internal energy is high enough, the thermal emission of an electron can take place. The resulting ions can be detected mass selectively in a time-of-flight mass spectrometer. Monitoring the mass specific ion yield as a function of IR wavelength then yields the IR spectrum of that species. This technique has given important information on the IR absorption spectra of interstellar dust.

FELs in the Mid-Infrared Region

This wavelength (energy) range extends from 20 μm to 2 μm (62 meV to 0.62 eV). Alternative sources in this wavelength range are nonlinear optical parametric generators and amplifiers (OPGs/OPAs) which compete against FELs in terms of peak power. However, there still exists a large number of interesting phenomena where FELs are better suited, e.g., due to their higher repetition rate in the MHz range as opposed to the kHz repetition rate of OPGs/OPAs and their higher average power exceeding several W up to kW as compared to mW for OPGs/OPAs. These high-power levels find applications in material processing and recently in surgery. Since these applications cover such a wide variety of different materials, we refer the reader to publications from users of the Jefferson Laboratories FEL and the Vanderbilt University FEL (see Table 1). The advantage of FELs in this context is that their wavelength can be tuned exactly to a resonance of the material under preparation. The following significant spectroscopic application was performed with FELs.

Vibrational Relaxation of Defects in Semiconductors

The heat dissipation in submicron electronic devices is of prime importance for the failure behavior of such devices. As the device size shrinks to the nanometer scale, atomic defects play an important role. In silicon, hydrogen passivation of defects is a standard process step in the production of metal-oxide-semiconductor electronic devices. Hence detailed knowledge about the stability and heat dissipation of hydrogen defects is very important. By employing nonlinear optical methods such as four-wave mixing or a transient bleaching technique, the vibrational lifetime of hydrogen vibrational modes in silicon (around a wavelength of 5 μm to 6 μm) could be determined and important information on the dependence of the local atomic arrangement and the vibrational lifetime could be obtained.

FELs in the Near-Infrared to the UV Region

The wavelength (energy) range from 2 μm to 200 nm (0.62 eV to 6.2 eV) is the realm of solid state lasers. Especially diode pumped solid state lasers have made tremendous improvements in terms of power and stability in the past years. Nonlinear frequency conversion of ultrashort optical pulses generated from titanium-sapphire lasers and amplifiers cover almost all of this wavelength range. Hence there is no specific need for operating FELs in this wavelength range for applications in physical sciences.

FELs in the Extended UV to X-ray Region

In the wavelength (energy) range shorter than 200 nm (larger than 6.2 eV) the availability of laser sources is very scarce. In this wavelength range, solid state lasers have the problem of the reabsorption of the laser light by the crystals' hosts. Also nonlinear frequency conversion processes are hampered by the absorption of the nonlinear crystals in this wavelength range. The generation of ultrashort pulses at the wavelength of the Cu K_α line has been obtained on the basis of amplified femtosecond titanium-sapphire pulses, thus enabling the study of the lattice vibrations in the time domain via a modulation of the X-ray diffraction. Since these sources are limited in their wavelength and output power, the development and operation of FELs in this wavelength range is truly important.

Only recently the first experimental verification of the feasibility of a SASE FEL at short wavelengths was obtained at the DESY facility. The lasing wavelength was 110 nm. Since several of these sources are presently under construction worldwide, there are as yet no experimental results published. However, the future development of short-wavelength FELs will enable new types of time-resolved experiments in atoms, solid state physics, chemistry, and biology. One prominent example is protein crystallography with X-ray pulses. The high intensity of FELs in this wavelength range should allow one to obtain a diffraction pattern of a single protein or other important bio-molecules in a single shot on a subpicosecond time-scale, i.e., before the protein is fragmented. The short-pulse duration opens the pathway to study the dynamical behavior of such

structures in pump-probe experiments with solid-state lasers synchronized to the FEL. In solid state physics, such experiments could yield important insights into the dynamics of phase transitions.

Other Developments

Beyond the use of an FEL as a light source in the above described wavelength ranges, there exists the possibility of using FEL facilities in a broader way. One example is the development of a high-intensity γ-ray source which can be accomplished by Compton backscattering of FEL photons from the relativistic electrons driving the FEL. This work has been performed at a storage ring FEL (OK-4/Duke University), where 12.2 MeV γ rays were obtained by backscattering 379.4 nm free-electron laser photons from 500 MeV electrons circulating in the storage ring. The advantage of these γ rays is that they are highly monochromatic and linearly polarized, thus making them an ideal source for nuclear spectroscopy and investigations in cancer therapy.

See also

Incoherent Sources: Synchrotrons. **Lasers:** Free Electron Lasers. **Spectroscopy:** Nonlinear Laser Spectroscopy.

Further Reading

Andruszkow J, Aune B, Ayvazyan V, *et al.* (2000) First observation of self-amplified spontaneous emission in a free-electron laser at 109 nm wavelength. *Physical Review Letters* 85: 3825–3828.

Benson SV (2003) Free electron lasers push into new frontiers. *Optics and Photonics News* 14: 20–25.

Budde M, Lupke G, Chen E, *et al.* (2001) Lifetimes of hydrogen and deuterium related vibrational modes in silicon. *Physical Review Letters* 87: 145501–145503.

Cole BE, Williams JB, King BT, Sherwin MS and Stanley CR (2001) Coherent manipulation of semiconductor quantum bits with terahertz radiation. *Nature* 410: 60–62.

Colson WB, Johnson ED, Kelley MJ and Schwettman HA (2002) Putting free-electron lasers to work. *Physics Today* 55: 35–41.

Dekorsy T, Yokovlev VA, Seidel W, Helm M and Keilmann F (2003) Infrared-phonon-polariton resonance of the nonlinear susceptibility in GaAs. *Physical Review Letters* 90: 055508.

Helm M (ed.) (2000) *Long Wavelength Infrared Emitters Based on Quantum Wells and Superlattices*. Amsterdam: Gordon and Breach.

Knippels GMH, van de Pol MJ, Pellemans HPM, *et al.* (1998) Two-color facility based on a broadly tunable infrared free-electron laser and a subpicosecond-synchronized 10-fs-Ti sapphire laser. *Optical Letters* 23: 1754–1756.

Litvinenko VN, Burnham B, Emamian M, *et al.* (1997) Gamma-ray production in a storage ring free-electron laser. *Physical Review Letters* 78: 4569–4572.

Murdin BN, Knippels GMH, Vandermeer AFG, *et al.* (1994) Excite-probe determination of the intersubband lifetime in wide GaAs/AlGaAs quantum wells using a far-infrared free-electron laser. *Sem. Sci. Techn.* 9: 1554–1557.

Palanker DV, Knippels GMH, Smith TI and Schwettman HA (1998) IR microscopy with a transient photo-induced near-field probe. *Opt. Comm.* 148: 215–217.

Sauvage S, Boucaud P, Lobo RPSM, *et al.* (2002) Long polaron lifetime in InAs/GaAs self-assembled quantum dots. *Physical Review Letters* 88: 177402.

Sorokina IT and Vodopyanov KL (eds) (2003) *Solid-State Mid-Infrared Laser Sources*. Berlin: Springer.

Unterrainer K, Keay BJ, Wanke MC, *et al.* (1996) Inverse Bloch oscillator: strong terahertz-photocurrent resonances at the Bloch frequency. *Physical Review Letters* 76: 2973–2976.

von Helden G, Tielens ACGM, van Heijnsbergen D, *et al.* (2000) Titanium carbide nanocrystals in circumstellar environments. *Science* 288: 313–316.

World Wide Web Virtual Library (2004) *Free Electron Laser Research and Applications*. Available online at http://sbfel3.edu/www/UL_fel.html

Industrial Applications

I P Mercer, Imperial College of Science, Technology and Medicine, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Lasers are seeing increasing application in industry, driven by emerging market demand and continuing enabling advances in laser technology. Core laser technology advances are based on robustness, performance, and cost. A combination of gas and solid-state lasers currently see application across the spectrum, from the extreme ultraviolet (EUV) to the infrared (IR). Industrial applications continue to emerge across the spectrum and are being extended to the extreme ends of the spectrum at terahertz (THz) and X-ray frequencies (see Figure 1).

Where a step in level of performance is offered by a laser based tool, that is robust enough and cost-effective for industrial application, it can be anticipated that new applications and markets will emerge. For this reason, many established laser manufacturers and new entries in the 1980s and 1990s were driven in part by the exploitation of laser technology advances, as opposed to market application demand. As markets mature for lasers applied to materials processing and diagnostics, so the driver across industry is beginning to shift from being predominantly set by the capability of emerging laser technology, to a combination of laser technology drive and application demand.

Figure 2 shows the current and emerging picture of laser technologies across a core range of wavelengths and pulse durations. Current technologies are shown along with emerging technologies expected to see industrial application in this decade, opening up new applications and markets. Currently established technologies include: CO_2 gas lasers at 10 μm; excimer and KrF gas lasers emitting in the ultraviolet at 248 and 193 nm, respectively; direct diode lasers emitting in the near infrared and visible; and solid state lasers emitting in the near infrared and visible. Current technology developments are centered on extending the range of performance of both gas and solid-state lasers: gas lasers are being extended to the ultraviolet with gases such as F_2; diode lasers are being extended to the ultraviolet with materials such as nitride-based semiconductors and to the infrared with the incorporation of quantum cascade technology; and solid-state diode pumped lasers are being extended to the infrared and ultraviolet via nonlinear conversion. In essence, achievable laser parameter space is seeing continued expansion, not only in terms of pulse duration and wavelength, but also in terms of average power, pulse energy, beam quality, robustness, and cost. Laser technology has now arrived as a major driver for industrial and social change through this decade and into the next.

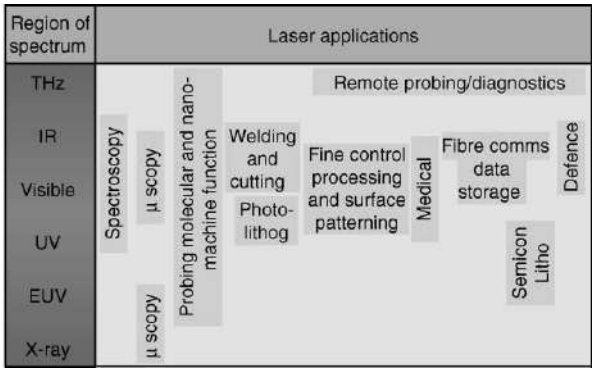

Figure 1 Industrial laser applications across the spectrum.

Infrared wavelengths directly stimulate nuclear motion, coupling to heat without necessary recourse to electronic excitation. Visible wavelengths couple to electronic transitions and at high laser intensities can ionize via multiphoton absorption. Ultraviolet wavelengths couple to electronic transitions and for wavelengths in the region of or below 200 nm, ultraviolet radiation can directly result in bond breaking. For materials processing applications, ultraviolet radiation can have advantages in reducing the laser threshold required for ablation, translating to reduced thermal load imparted to the material.

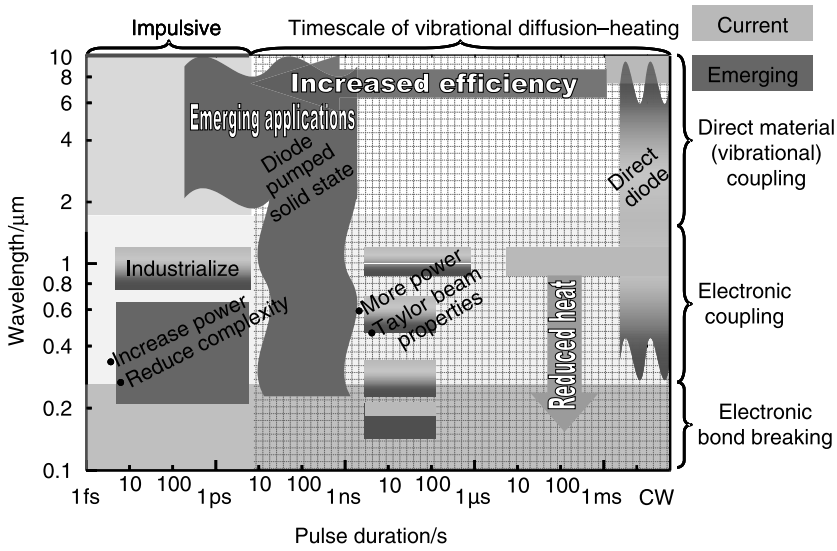

Figure 2 Current and emerging laser technologies across the spectrum as a function of laser pulse duration. Currently established technologies (light gray) are shown with technologies expected to emerge during this decade (dark gray).

Nuclear motions in materials cover a range of timescales of dynamics from 100 ps (0.3 cm^{-1}) to 10 fs (3000 cm^{-1}). As the laser pulse duration is reduced below 1 μs, the effects of thermal diffusion are mitigated in solids and as the pulse duration is further reduced below 100 ps, the interaction between radiation and matter is seen to adopt impulsive behavior. Reduction in pulse duration below the microsecond regime can lead to increased efficiency for material ablation combined with improved quality.

Materials Processing – Photochemical

Lithography in Semiconductor Manufacture

Laser technology's role in semiconductor and microelectronics fabrication is growing as manufacturers move towards producing smaller, more powerful devices. The overriding commercial driver for a generation has been to increase the number of integrated components per area of silicon. This has been achieved by reducing the feature size of a transistor on silicon and this trend is anticipated to remain in place through to 2010.

Features on silicon are chemically etched following lithographic exposure. To meet the increasing demand in resolution and throughput, lasers have been applied with a trend in development towards shorter (ultraviolet) wavelengths, with increasing power and reliability.

The semiconductor industry is currently in transition from 248 nm (XeCl gas) to 193 nm (ArF gas) ultraviolet (UV) pulsed laser technology to expose silicon wafers of diameter 30 cm. A typical laser power available at this time at 193 nm, is 40 W at a repetition rate of 4 kHz. To continue the reduction of feature sizes by a factor of two every four to five years, the industry is considering moving to 157 nm (F_2 gas) pulsed laser technology. As wavelengths reach the deep ultraviolet (DUV), transmission through gas is significantly reduced, requiring a move to vacuum-based systems. Associated technologies relating to translation (stepper), mask, and image relay optics also require significant advancement. A technology also being considered at this time is to image the mask to silicon through a medium with refractive index greater than unity, thereby reducing the minimum feature size afforded by diffraction theory, whilst maintaining the same laser wavelength and numerical aperture.

It is estimated that at a feature size of approximately 80 nm, the above technologies will be reaching their resolution limit. This will necessitate a switch to electron beam projection lithography or extreme ultraviolet (EUV, 13 nm) technology. Electron beam projection carries a disadvantage in terms of projected rates of wafer throughput and as such, emphasis has been placed by the industry on the current development of EUV technology. EUV radiation can be produced by laser interaction with a gaseous volume.

Computer to Plate Printing

A common advantage where lasers are replacing current technologies is to reduce the number of associated processes and thereby reduce the cost of process ownership. Computer to plate (CTP) printing involves the laser exposure of a printing plate without the use of a film intermediate. This technology is finding fast traction within the printing industry and is applicable to drum or flat bed platforms. CTP first gained serious attention in the 1990s and has since substantially advanced, with a wide range of technology suppliers (Figure 3).

Traditional methods have involved offset printing plates, imaged used traditional film, and high-power

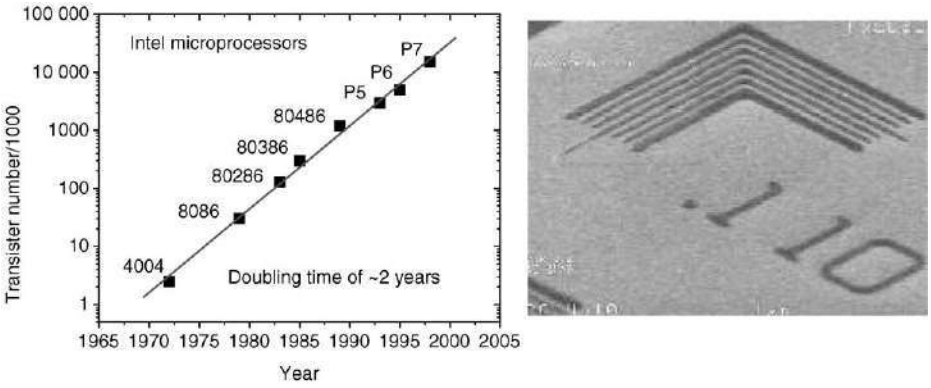

Figure 3 Moore's Law: continuing growth in computing power, showing a doubling in the number of transistors per unit area of silicon every 2 to 2.5 years. Also shown are 110 nm dense lines on silicon (state of the art in 2002). Source: International Sematech.

UV lamps. In contrast, CTP exposes the plate directly with laser radiation. The major saving reported by users of CTP is a faster turnaround time. For color printers, CTP gives almost instant color register on the press and there is a reduced requirement to stop the press to delete dust marks and scratches, as was the case with film. As the industry sees transition to reduced press run lengths, so the importance of CTP is increasing.

Digital plates generally use a high-quality, grained, and anodized aluminum substrate, as for analog plates (see Figure 4). High tolerance electrochemical graining and anodizing ensure stable ink/water balance and good-quality printed results and translates to predictable press behavior. Multiple coatings are applied for CTP, in contrast to a single coating for offset printing, and these take the form of a photopolymer or silver halide and a thermally responsive layer.

On the press, ink and water rollers are continually rolled over the plate. The grained and anodized aluminum substrate serves as a hydrophyllic surface volume to attract water. Ink is oil based and is repelled from areas of water absorbed into the substrate, but is attracted to the hydrophobic dot material.

Silver halide is rated for runs up to 350 000, proving a popular choice for commercial and newspaper applications. It is straightforward to sensitize for a range of writing wavelengths and is the fastest to write.

Thermal plates offer high resolution, equivalent to silver halide and can be baked to enable the longest application lifetimes. Typically the plates are exposed to 830 nm radiation on the external surface of a drum. This technology has served the industry well, but has struggled to find full application in the newspaper business, due to its disadvantage in writing speed.

The most common CTP technology used for newspapers uses a photopolymer coating, which is fast to write and durable. Resolution has traditionally been limited, with Agfa, for example, rating their N91 photopolymer plate to a maximum of 175 lines per inch (LPI), whilst thermal and silver plates are able to image at 200 LPI and above. However, recent advances in shorter wavelength violet diode lasers and violet sensitive polymers have improved resolution. Launched in 2000, violet technology is seeing widespread takeup. This technology is a spinoff from the DVD industry and with laser lifetimes estimated at 10 to 20 years, cost of ownership is one of the drivers. Both silver halide and photopolymers can be sensitized to violet. Current typical laser parameters are for: silver halide, 2–3 $\mu J/cm^2$ at 405 nm; photopolymer, 0.2 mJ/cm^2 at 488 or 532 nm; and thermal, 0.1 J/cm^2 at 830 nm.

The chemical composition of the CTP surface layers is balanced to determine the response to different laser wavelengths, as well as performance characteristics in different print applications. The plate coatings undergo a photochemical change during laser exposure. This is a critical stage since it is important that the plate surface is not under- or over-exposed.

Chemical post processing after exposure is generally required for CTP to date, with the exception of ablative CTP, to remove unexposed material. After imaging, thermal plates go through a pre-bake step, during which the system heats the plate to 130 °C (266 °F) for about 30 seconds. The energy from this

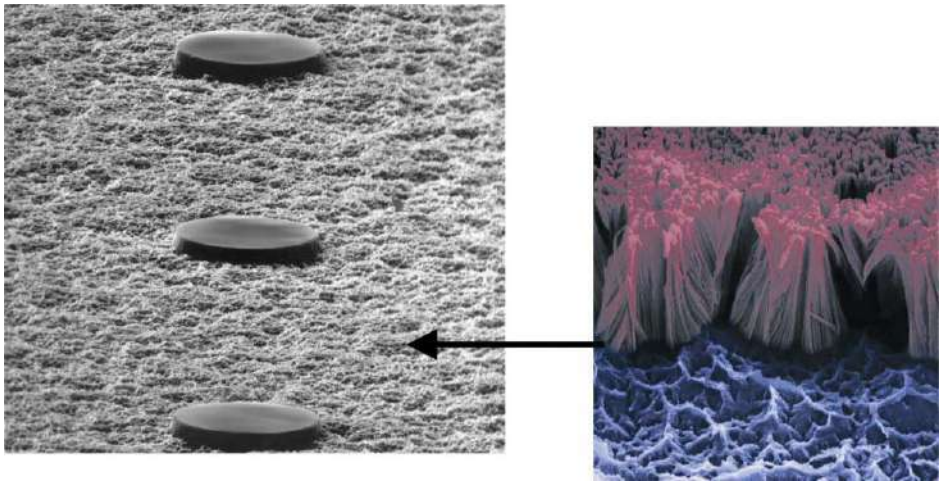

Figure 4 Three typical polymer dots of a few microns depth on an anodized aluminum substrate. Also shown is a further magnified electron microscope image of the substrate material, with the anodized layer cut away for presentation. The porous structure increases water retention. Oil-based ink is repelled from areas other than the dots. Figure courtesy of Agfa (Tony King, tony.king1@agfa.com).

heating stage completes cross-linking of polymer chains in the exposed portions of the emulsion. Chemical processing consists of processing the plate through an alkaline developer solution where the polymer that is not cross-linked washes away. Following the wash is a rinse and application of a coating. After chemical processing, some plates require baking to 200 °C (392 °F) for 35 minutes. Afterwards, the plate continues on to rinsing.

The machine tools required for these processing steps measure 30 feet long and take up to 7 minutes to process a plate after imaging. For both polymer and thermal plate processing, the temperatures are carefully controlled to prevent the aluminum substrate from deforming during the baking stages.

The effluent from thermal and other photopolymer plate processors is fairly benign and is typically directly disposed without significant environmental concern. Spent photopolymer chemistry can, however, pose problems with waste flow, it being a plastic sludge that can solidify and over time, clog plumbing systems.

Silver diffusion plate processors are relatively compact and tend to process plates in less than a minute. Environmental precautions are required with the disposal of accumulated silver particulates. Relatively simple systems are available, which neutralize the pH of the effluent and filter out silver until concentration levels are below 3 parts per million. Under these circumstances, most jurisdictions allow drain disposal.

Ablative plates do not require post processing, other than a rinse to remove residues. This is an emerging technology that has not taken hold at this time in the printing industry and requires more powerful lasers to ablate a layer from the plate. These plates use layers of ink-repellent silicon on top of an aluminum substrate to provide the non-inking portion of the plate. Variations of this technology use a laser to ablate the silicon from the substrate in the image areas of the plate, or to transfer material from a donor layer to the aluminum substrate. A weakness of this technology has been the limited durability of the plates in press. There is continuing development to produce ablative plates with other surface layers that are more durable on press.

Materials Processing – Heating

One way in which lasers can be applied to materials processing is as a directed source of heating. With material quickly raised in temperature to the point of melting and vaporization, applications lie in drilling, welding, cutting, and scintering.

The exploited characteristics of a laser source are its directionality and discrete color. Directionality enables the formation of a restricted area for heating as well as an extended working distance, whilst a discrete color enables absorption to be optimized for the material being worked. Laser technologies typical in application to materials processing through heating, range from diodes at 808 and 940 nm to flashlamp or diode pumped solid state lasers at 1 μm, and CO_2 gas lasers at 10 μm. A recent entry technology is in diode pumped fiber lasers, reaching output powers in the kilowatt regime at a wavelength of 1 μm and typically being continuous in emission or pulsed with a pulse duration in the millisecond regime.

Direct diode lasers (see Figure 5) have advantages in simplicity and robustness and with the minimum of additional assembled optical components, costs can be expected to continue to reduce, fed by Moore's law for the cost of semiconductor manufacture. Typical powers range from 1 to 10 kW and light can be fiber delivered to the work piece. A disadvantage of direct diode emission is the reduced directionality in comparison with alternative laser sources. For example, a current typical numerical aperture of emission from a diode bar array emitting a power of 60 W, is 10 and 60 degrees in directions parallel and perpendicular, respectively, to the array. These divergences can be reduced to a numerical aperture of approximately one degree with the addition of micro-optics, at a device cost increment of approximately 20%. For an array of diode bars with emitting powers in the kilowatt regime and with emitting dimensions sided 2 cm, a typical optical quality is one thousand time diffraction limited, translating to working distances are on the order of centimeters for a laser spot size at a work piece that is on the order of 1 mm.

CO_2 gas lasers and solid state lasers offer the advantage of reduced divergence and corresponding greater working distance, combined with a smaller laser spot size at the work piece for finer processing.

Figure 5 Diode bar array, with an output power exceeding 500 W. Courtesy of Rofin.

Both laser technologies can operate in the kilowatt regime, for application to sheet metal cutting and welding.

With the interaction process involving high temperatures and the ejection of material, a high-pressure gas bathes the area of illumination to remove smelt and either reduce exothermic chemical reactions, or where process acceleration is required, to utilize an assist gas to increase heating through exothermic reaction. Oxygen is typically used as an assist gas for the processing of low-alloy steels, with the process being typically known as laser flame cutting. The gas is, in general, injected coaxially with the laser beam. Laser flame cutting is used in industry for material up to 4 cm thick, albeit the width of the cut becomes significant at this thickness. Both CO_2 and Nd:YAG lasers are applied for this application, the choice of which is dependant on the cut geometry, the cycle time, and the material. Typical cutting speeds are, for example, in the region of 10 meters per minute for 1 mm, 3 meters per minute for 6 mm, and 1 meter per minute for 15 mm thick material.

In the case of high-alloy steels and aluminum, in particular, the material is bathed with an inert gas such as nitrogen or argon. The material is heated solely by the energy in the laser beam, and is typically referred to as laser fusion cutting (see Figure 6). The laser power required is higher than for laser flame cutting. Laser fusion cutting reduces oxidization to the cut edges, which is particularly important where welding is the next process step after cutting.

Laser fusion cutting is used in industry for material up to 25 mm thick, albeit the width of the cut becomes significant at this thickness.

Figure 6 Laser fusion cutting. Courtesy of Rofin.

As with laser flame cutting, both CO_2 and Nd:YAG lasers are suitable for this application, the choice of which is influenced by the geometry of the cut, the cycle time, and the material. Typical cutting speeds in steel are, for example, in the region of ~8 meters per minute for 1 mm, 4.5 meters per minute for 3 mm and 1.5 meters per minute for 8 mm thick material.

Materials Processing – Localized Heating

As a material is heated in a localized volume to the point of vaporization, so also is the surrounding material, which can melt and experience cracking or changes in phase. Such heating is mainly considered deleterious, reducing process tolerances and reducing the tensile strength of the surrounding material.

As the pulse duration is reduced for a localized laser heating source, so are the dimensions that experience heating around the interaction volume (the heat affected zone), translating to improved part quality, with the length for thermal diffusion, l_{th} given by:

$$l_{th} = \sqrt{Dt} \qquad [1]$$

where D is the thermal diffusivity and t, the time of diffusion. As the pulse duration is reduced, the pulse energy required to ablate material is also reduced, with the laser pulse energy required to raise the material temperature to the point of vaporization being proportional to the volume heated. For an example such as silver, a pulse duration of 200 fs is associated with a diffusion length of 6 nm. Extrapolating, a pulse duration of 100 ps, 100 ns, and 100 μs are associated with diffusion lengths of 0.1, 4, and 130 μm, respectively. For typical metals and dielectrics, the process quality and pulse energy threshold for ablation become decoupled from variation in pulse duration for pulse durations less than a few picoseconds. In this regime, the volume for heating is dictated more by the depth for optical absorption and further, the physical mechanism for absorption changes to that of material breakdown and multiphoton ionization (Figure 7).

Additional advantages of operating in the short-pulse regime are a sharpening of threshold to ablation, enabling the processing of features of dimension less than that of the laser spot size and an increase in the number of materials that can be processed by the same color laser source.

Given the advantages of the short-pulse regime, it should be noted that this is currently juxtaposed with the complexity and cost of the required laser sources. Solid-state laser technology is, however, advancing, significantly reducing the number of optical components required to deliver short pulses at multi-kHz

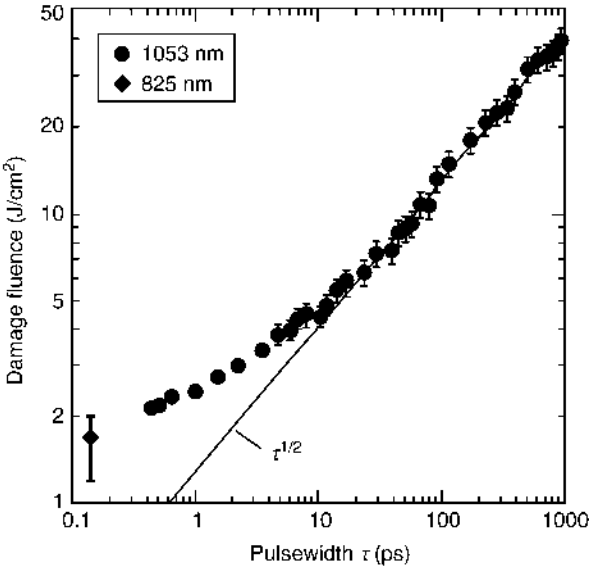

Figure 7 Pulse duration dependence of threshold damage fluence for fused silica. Reproduced with permission from Stuart BC, Feit MD, Herman S, Rubenchik AM, Shore BW and Perry MD (1996) Nanosecond-to-femtosecond laser-induced breakdown in dielectrics. *Physics Review B* 53: 1749. Copyright (1996) by The American Physical Society.

repetition rates and multi-mJ pulse energies. At this time of evolution in laser technology, solid-state, diode-pumped high repetition rate (>10 kHz), high pulse energy (>40 mJ) nanosecond (1–200 ns) lasers at a wavelength of 1 μm are seeing increasing application. Such laser sources offer a combination of robustness and cost-effectiveness, combined with a process quality improvement over sources with longer pulse duration.

In the nanosecond regime, the nature of light-matter interaction remains dependent on the wavelength of the laser source and the material properties. In particular, wavelengths in the UV exhibit reduced pulse energy threshold and can offer improved process quality. UV wavelengths less than 300 nm are currently produced by excimer and ArF gas lasers, emitting at 248 and 193 nm, respectively, or solid-state sources that are wavelength converted with nonlinear optical parametric materials to emit at 354 or 256 nm. Significant laser development effort is underway to improve optical parametric materials for nonlinear conversion, to enable greater robustness for solid-state UV sources that then offer reduced overall laser tool cost and complexity.

Semiconductor Lithography

The advantage of nanosecond and shorter pulsed lasers for materials processing extends from localization of heating. With heating being localized, a greater temperature can be achieved. This behavior is fundamental to the production of EUV radiation for application to next-generation lithographic processing.

EUV lithography is considered as an attractive candidate to succeed deep DUV-based optical lithography within this decade. EUV wavelengths around 13 nm will allow the reduction of structure sizes to less than 80 nm in semiconductor devices.

For such a high-power light source that emits at a wavelength of 13 nm, solutions are currently being pursued by multiple industry players in electrical plasma discharge and laser produced plasmas (LPP) (Figure 8). Electron beam and synchrotron sources could, in principle, satisfy the application requirements, but suffer from restriction relating to cost, size, and throughput. The choices of technologies remain to be finalized. It is, however, accepted across the industry that Moore's Law should remain intact and that a solution is required that is scalable to high throughput rates for wafer exposure.

For a laser-produced plasma, the peak wavelength of the light emitted from a plasma is determined by its temperature, T, where according to the blackbody formula:

$$T \sim E_{ph}/3 \qquad [2]$$

where E_{ph} is the photon energy associated with the wavelength of maximum emission. Liner photopolymerization (LPP) schemes use pulsed laser radiation to heat a target to the optimum temperature required for efficient EUV production. For plasma-emitted light which peaks at a wavelength of 13.5 nm, the optimum temperature is approximately 3.5×10^5 Kelvin (30 eV).

A laser is focused onto a target, transferring its energy into the target in such a way that a plasma is formed and heated. A range of targets have been investigated for use in LPP schemes ranging from gaseous and cluster targets, to liquid targets and macroscopic solid targets. The interaction begins with the production of free electrons via multiphoton ionization, which is strongly dependent on the peak intensity of the laser. As heating progresses, there is further ionization from inelastic collisions between heated electrons and ions to reach a pseudo-equilibrium state, where cooling competes with heating on nanosecond timescales. Given this time-scale for cooling and requirements on source size and temperature, the optimum laser pulse duration lies in the regime of nanoseconds and the optimum laser pulse energy is hundreds of millijoules. Laser technology

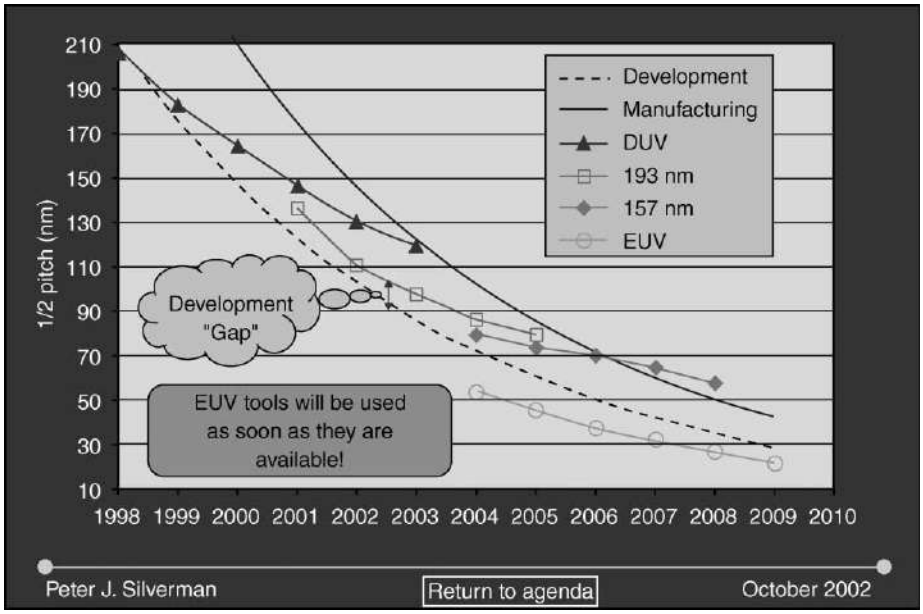

Figure 8 Light source roadmap to 2010. Courtesy of Intel Corporation.

suited to this application is diode-pumped Nd:YAG, emitting at a wavelength of 1 μm and pulsed at multikilohertz repetition rates.

Fine Control Material Removal

Localized heating afforded with the application of lasers with pulse durations in the nanosecond regime and shorter, is seeing increasing takeup to replace processes ranging from wet etching and electro-discharge machining (EDM) in applications such as scribing, drilling, and differential layer removal. Laser methods are seeing advantage in direct ablation of material, for feature sizes on the order of microns to hundreds of microns. Large areas on the order of 2 m^2 are now becoming cost-effective to process with direct laser ablation, with the advent of high average power laser sources of hundreds of watts of average power, combined with nanosecond regime or shorter pulse durations. Typical sources are gas excimer or diode-pumped solid-state.

The finest processing is advantaged by the application of pulse durations in the picosecond regime. A restriction in application of this technology has been robustness and cost of the laser sources to date; however, such technology is continuing to see rapid improvement and may be expected to see wider takeup in industry by 2010.

Also seeing application, are wavelengths in the UV regime combined with nanosecond pulse durations. UV radiation in general offers a reduced threshold of ablation, translating to reduced heating and improved process quality. High absorption at UV wavelengths in most materials, combined with photon energies that can exceed bond energies of the material, contribute to the removal of material with the minimum of heating. Excimer, KrF, and ArF gas lasers are applied here (Figure 9), but it is expected that solid-state lasers will emerge by 2010 to capture a large part of the UV market. CO_2 gas lasers are used for fine processing where the material being ablated is of low thermal diffusivity, which is typical for dielectrics or ceramics. Materials of low thermal diffusivity are associated with a low heat affected zone (see eqn [1]).

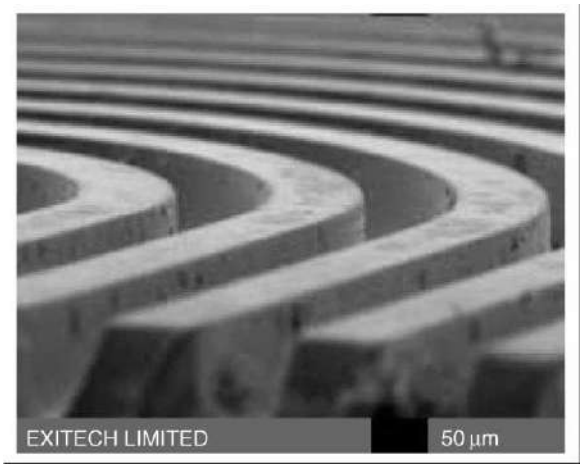

Figure 9 KrF laser micromachined microfluidic channels in polyester for application to lab-on-a-chip technologies or sensing devices. Courtesy of Exitech Ltd.

Scribing is required for dicing of wafers such as silicon and other low-*k* materials, both for computing device manufacture and for solar cell manufacture. Nanosecond and shorter pulsed lasers have seen recent application in this area, enabling higher aspect ratio scribing and a reduced material affected zone. This application has seen laser wavelengths applied from the UV to near-infrared. Short pulse laser scribing can see advantages of high throughput yet reduced or eliminated chipping, delamination, and film peeling.

Fine drilling is required for a wide variety of applications and markets ranging from constructing vias for through-layer interconnects in printed circuit board (PCB) fabrication, to ink jet nozzles (Figure 10) and high aspect ratio holes in hard materials, such as steel or alloys for application to combustion engines.

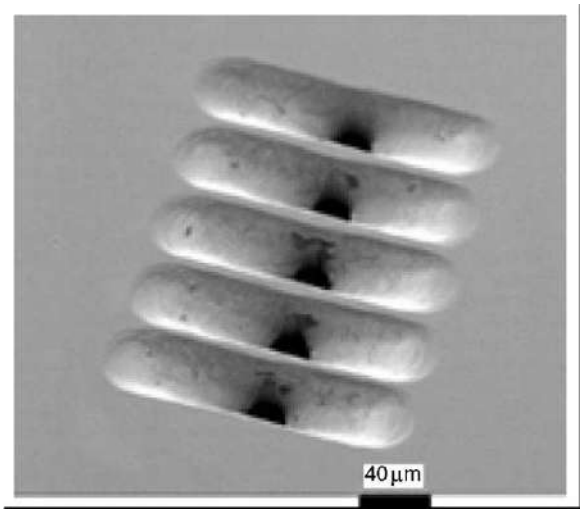

Figure 10 A laser drilled hole in steel and an array of shaped holes in polyamide for inject nozzle application. Courtesy of Exitech Ltd.

Excimer and CO_2 laser drilling has seen industrial application in the drilling of via holes for electronics interconnection packages, since early investigations in the 1980s. More recently, high repetition rate diode-pumped solid-state lasers, nonlinearly converted to emit in the UV, are seeing application. In parallel with the increasing density of transistors in silicon processors, the density of packaging of multichip modules (MCM) and PCB components has also been increasing, requiring the drilling of blind vias on the order of 100 microns in diameter to enable interconnects between multiple conducting layers (Figure 11). Also, as processor power increases, so does the number of connections required to devices. Microvias have enabled the application of two-dimensional grid array connections to the underside of devices, reducing overall package size, including processor electrical connections, by a factor of five.

Laser drilling of microvias in PCBs is expected to increase as packing density increases. PCBs remain cheaper than MCMs and traditionally have been mechanically drilled. As via sizes are reduced below 100 μm, however, it is expected that pulsed CO_2 and Nd:YAG lasers will be applied.

Thin layer ablation is required for applications ranging from surface cleaning to display manufacture and for polymers, dielectrics, and metals. For example, manufacture of plasma displays currently incorporates lithography combined with wet etching, to pattern layers of thin conducting oxide onto a substrate. Multiple process steps can be replaced by application of direct laser ablation and increased laser average power combined with pulse durations in the nanosecond regime are allowing millions of cells to be processed per minute, with the accommodation of process areas currently greater than 2.5 m^2 (see Figure 12).

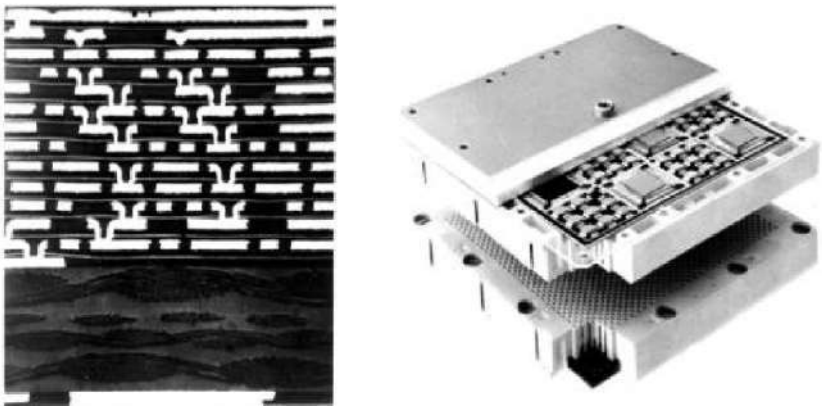

Figure 11 Cross-section of 14-layer copper plated MCM board showing 80 μm diameter microvias drilled by a KrF laser in acrylic resin-polyimide and, fully assembled MCM package. Courtesy of Siemens AG.

Laser-Based Diagnostics

With a continuing increase in reliability, performance, and range of specifications, lasers are seeing growing industrial application for process monitoring and diagnostics, ranging from gas phase detection of molecular specific concentrations to *in vitro* microscopy in the life sciences. Markets are emerging, driven in a large part by advances in laser technology. Chemical sensing with application of lasers is a vast field, which is difficult to cover here completely. Following, are some applications seeing current industrial application and growth.

Microscopy

Confocal microscopy has seen large uptake, in particular, for the application to imaging of cell and membrane function, with more than fifteen manufacturers now supplying confocal microscope tools including Zeiss, Biorad, Leica, Olympus, and Nikon. Confocal microscopy enables increases in resolution and contrast compared to standard microscopy. Techniques such as fluorescence resonant energy transfer (FRET), fluorescence recovery after photobleaching (FRAP), fluorescence loss in photobleaching (FLIP), fluorescence lifetime imaging (FLIM), or fluorescence correlation spectroscopy (FCS), where married with confocal microscopy and, in particular, where relevant molecular structures are known in advance, can offer the additional advantages of specific molecular and structural specificity through the incorporation of tracer fluorophores. This is seen as color coding of the resultant image, with sophisticated microscopes currently offering differential imaging stimulated by three laser colors and with six selectable colors for detection.

The confocal microscope operates with advantage over a standard microscope, by illuminating only a small area in the sample with dimensions of the resolution limit of the objective and re-imaging emission from that area through a pinhole spatial filter (Figures 13 and 14). Structures outside of the illuminated area or at a depth outside that of the focus, are suppressed from detection. Fluorophores (tracers) can be inserted into protein sequences at predetermined locations and are stimulated by either single or two-photon absorption of the stimulating laser light.

Two photon absorption has the additional advantages of increased resolution and detected contrast over single photon absorption. Laser sources applied are generally modelocked oscillators operating near 800 nm, with a pulse duration on the order of 100 fs. This technology is continuing to improve in robustness, with fiber and direct diode pumped solid-state sources available from multiple vendors.

FRAP and FLIP monitor diffusion or transport of fluorescent labeled molecules into or out of the laser focal volume, respectively. FCS monitors the fluctuations in detected signal from emissions within the laser focal volume, to determine the rate of diffusion with sensitivity achievable to the single molecule level (Figure 15).

FRET is seeing application for the probing of protein interactions in cells or at cell membranes. FRET involves the nonradiative transfer of excitation energy from one fluorophore to another, which subsequently emits at a different wavelength. The fluorophores are inserted into molecules, the

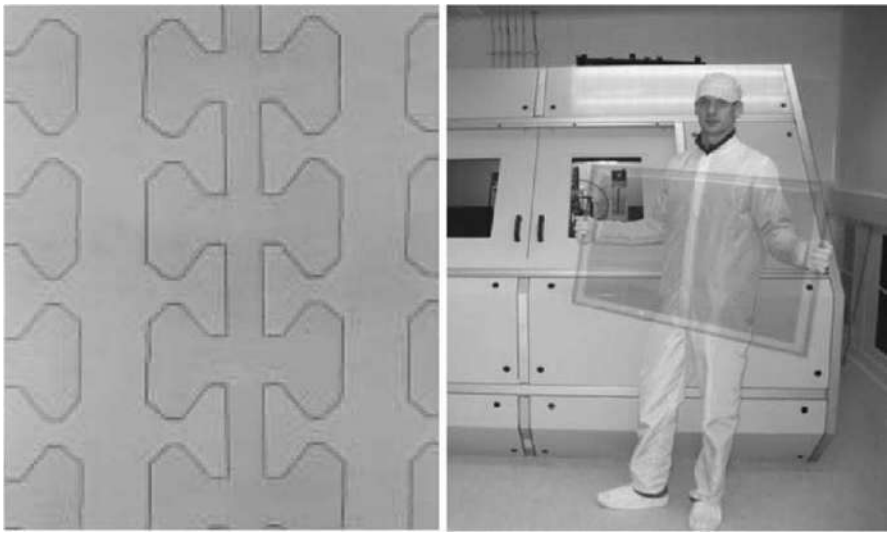

Figure 12 Plasma display manufactured by direct laser ablation of thin-film indium tin oxide (ITO). Courtesy of Exitech Ltd.

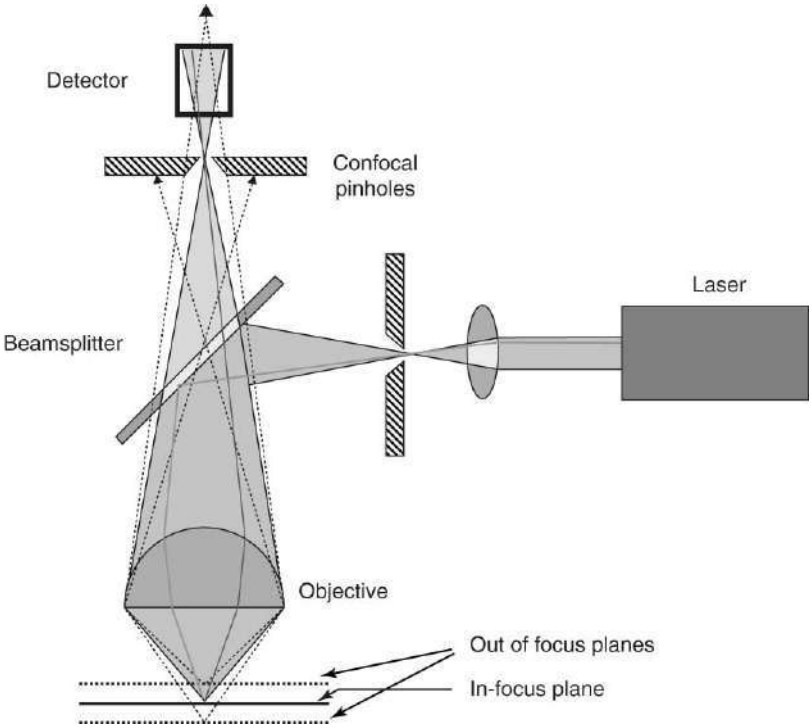

Figure 13 Confocal microscope principle of operation. Courtesy of Leica Microsystems Heidelberg GmbH.

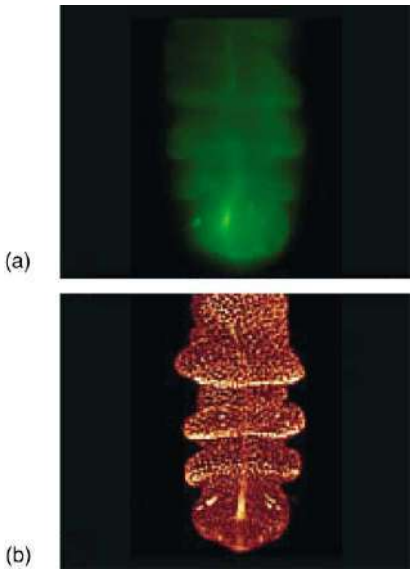

Figure 14 Image (a): *Drosophila melanogaster*, fluorescence nonconfocal configuration and Image; (b): *Drosophila melanogaster*, fluorescence confocal 3D projection. Courtesy of Leica Microsystems Heidelberg GmbH.

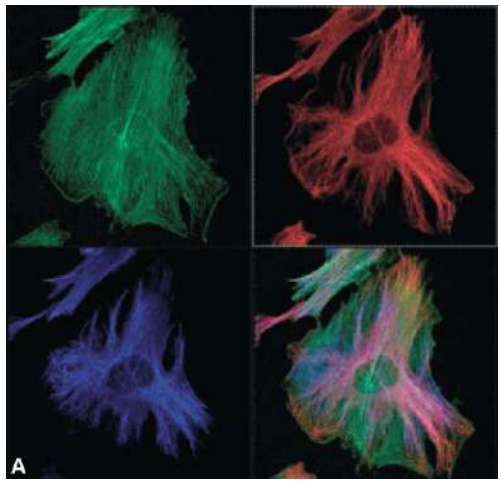

Figure 15 Spectral selective imaging with three fluorescence channels showing triple labeled mouse fibroblast: green: actin fibers, red: microtubules, blue: vimentin. Courtesy of Leica Microsystems Heidelberg GmbH.

interaction of which is being investigated. FRET effectively enables sensitivity to relative position at fluorophore separations less than 10 nm; a significant advance over the 200 nm resolution provided by a standard optical microscope. Concentrations of binding events can be monitored over time (Figure 16).

Raman Methods and Microscopy

Raman spectroscopy sees broad application in chemistry in the analysis of chemical composition and molecular structure, being used, for example, in pharmaceuticals, polymer materials science, bioclinical science, and forensic science.

The history of Raman scattering stems from theoretical predictions in the early 1920s,

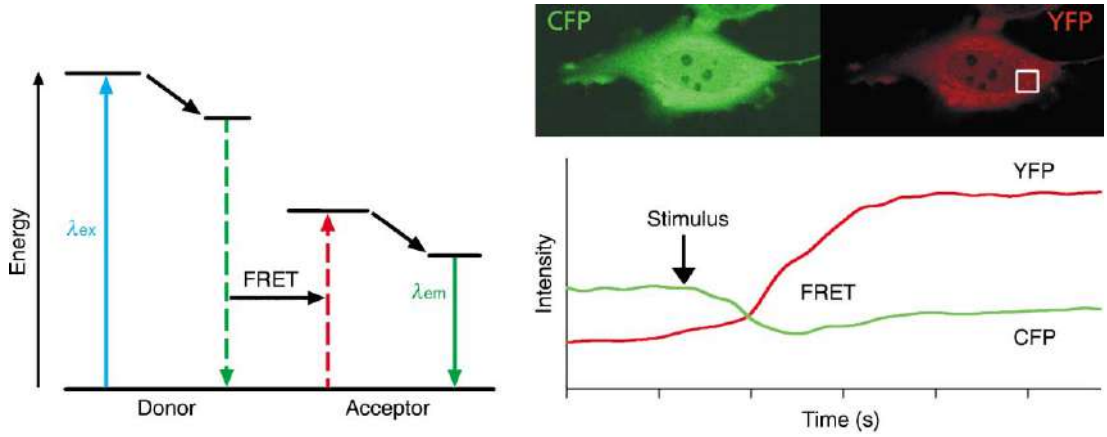

Figure 16 Fluorescence resonant energy transfer mechanism and example relating to cytoplasmic Ca^{2+} concentration. Cyan Fluorescent Protein (CFP) and Yellow Fluorescent Protein (YFP) are acceptor and donor, respectively, for the FRET. Courtesy of Carl Zeiss, Germany.

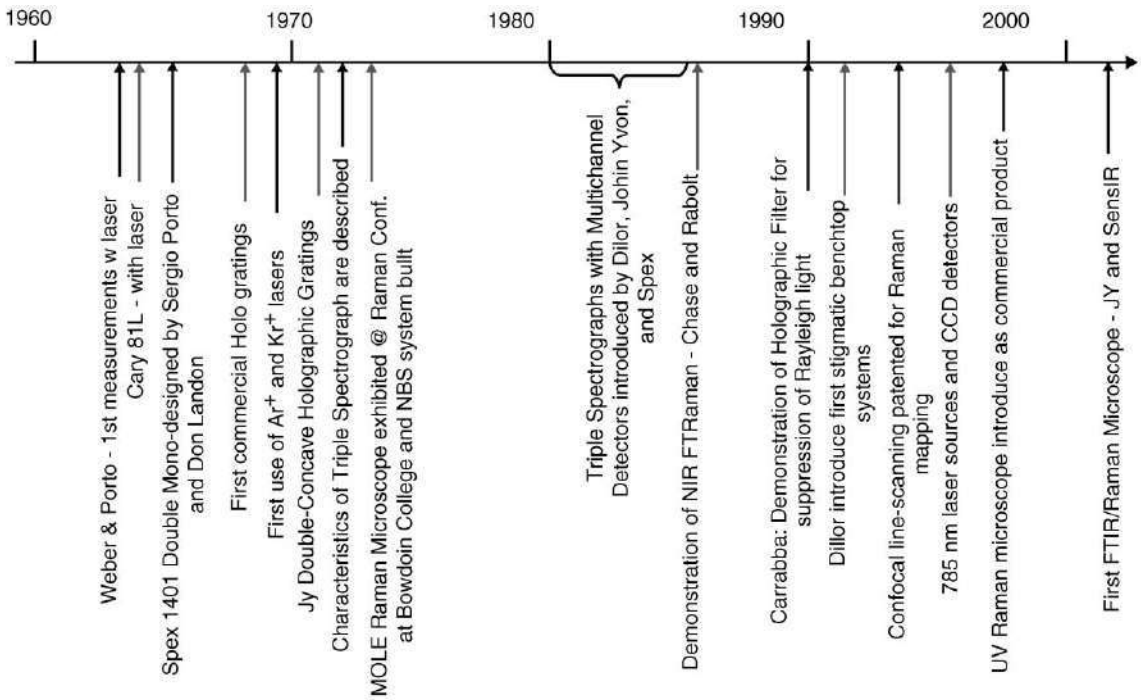

Figure 17 Timeline of Raman spectroscopy from the first incorporation of a laser source. Courtesy of Jobin Yvon.

accompanied also in the 1920s by early measurements, notable being those of C.V. Raman for which he was awarded the Nobel prize, soon followed in the 1930s by hundreds of chemistry papers describing this technique. Raman used for his early experiments filtered sunlight, a prism spectrometer, and visual observation. This was followed by application of filtered Hg emission and photographic plates. Lasers first saw application to Raman scattering in the 1960s, shortly after their innovation (see Figure 17).

In the simplest form of application of the Raman effect, UV or visible radiation illuminates a sample, coupling to excite electronic transitions near or on resonance. Spontaneous emission is shifted to a longer wavelength than that of the excitation (Stokes emission) and this is filtered and detected. In general, a broad spectrum of fluorescence is superimposed by weaker, narrow features that originate from vibration-electronic coupling (Raman emission). Inclusion of laser excitation increases the rate of

data collection and also enables nonresonant interaction of the stimulating radiation with the sample whilst maintaining data collection rates. Nonresonant interaction preferentially reduces electronic emission of fluorescence relative to that of the narrow featured Raman emission. Lasers applied to spontaneous Raman spectrometers in general emit in the visible or near-infrared running nonpulsed, such as Ar-ion, Nd:YAG or He:Ne.

Further enhancements are made to deliver spatial as well as vibration-specific information, through tight focusing of the stimulating radiation on to the sample in a confocal microscope arrangement, delivering spatial resolution down to 200 nm (Figure 18).

A limitation of standard Raman spectrometers and microscopes is the rate of signal collection, requiring concentrated samples for interrogation. The restricted signal level derives from a typical cross-section for the spontaneous emission of Raman radiation being $10^{-30}\,\mathrm{cm}^2$; 10^{10} to 10^{14} times less than that typical for spontaneous electronic fluorescence. Enhancements in the level of signal and the ratio of signal to background noise may be made through coherent laser stimulation of anti-Stokes Raman emission, known as coherent anti-stokes Raman scattering (CARS) (Figures 19 and 20).

It can be anticipated that bond-specific imaging will see wide industrial application in future, employing Raman as well as resonant infrared techniques. Application of CARS to microscopy involves two laser colors. A vibration transition is exited at the beat frequency of the two laser frequencies and the shorter laser wavelength is subsequently scattered from the sample to a yet shorter wavelength of emission. Such a coherent process has the advantage of signal scaling proportional to the square of emitters at the sample, thereby increasing the signal over that of spontaneous Raman scattering. Typical laser pump and Stokes powers range from less than 1 mW to 10s of mW, with laser repetition rates ranging from multi-kHz to greater than 100 MHz and typical pulse durations from 10s of picoseconds to 100 fs.

Infrared Methods

Infrared absorption spectroscopy provides a wealth of information on bond-specific sample concentration

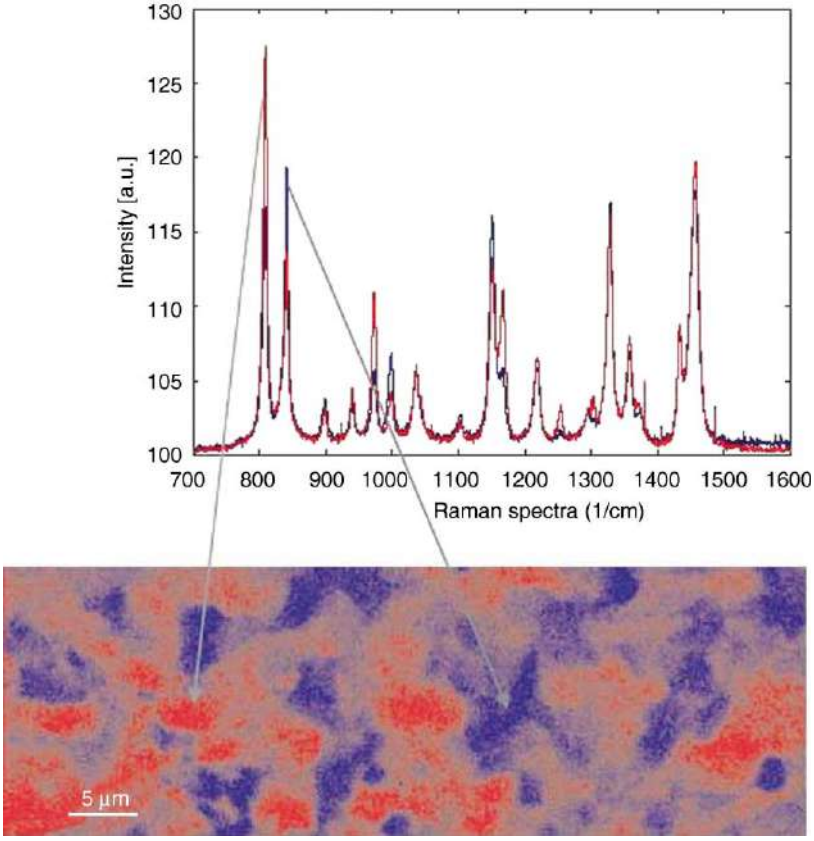

Figure 18 Confocal Raman image of semicrystalline polypropylene film. Two typical Raman spectra are shown. The ratio of the Raman lines at 809 and 841 cm increases with the degree of crystalline order, delivering a two-dimensional image (Hendra *et al.* 1995). Courtesy of Witec GmbH.

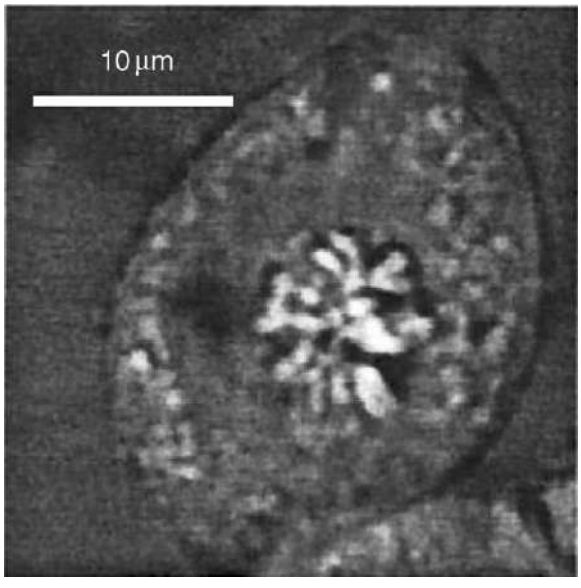

Figure 19 CARS image of a NIH 3T3 cell in metaphase, where Raman vibration stimulation is tuned to the PO_2^- symmetric stretching vibrational frequency at 1090 cm^{-1}. Reproduced from Cheng J-X, Kevin JYK, Zheng G and Xie X-S (2002) Laser-scanning coherent anti-Stokes Raman scattering microscopy and applications to cell biology. *Biophysical Journal* 83: 502–509, with permission from The Biophysical Society.

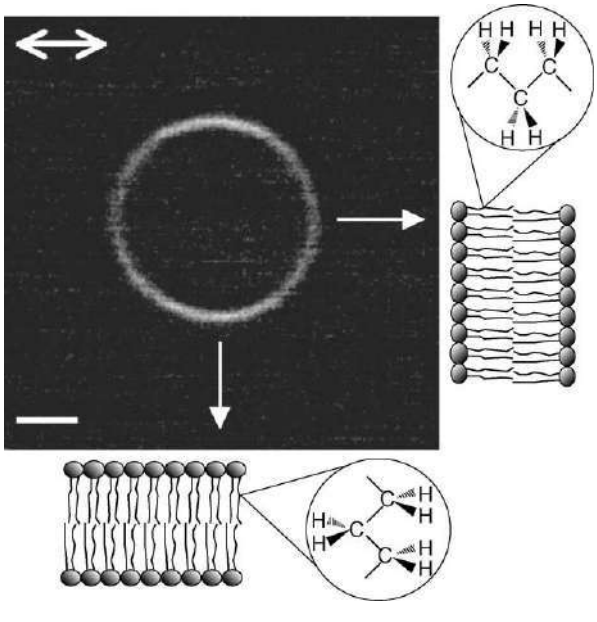

Figure 20 CARS microscope image of an erythrocyte ghost, consisting of a single lipid bilayer, with a Raman shift of 2845 cm^{-1}. The scale bar is of length 2 μm. Reproduced from Potma EO and Xie X-S (2003) Detection of single lipid bilayers with coherent anti-Stokes Raman scattering (CARS) microscopy. *Journal of Raman Spectroscopy* 36: 642–650, with permission from Wiley InterScience.

and has become a standard tool for the chemistry laboratory. Interaction between the radiation and sample is via resonant vibrational absorption and is otherwise analogous to the operation of a standard optical absorption spectrometer. Applied to a microscope arrangement, resolution on the order of 10 μm is standard. One restriction associated with infrared absorption spectroscopy applied to solution, is that absorption lineshapes are broadened and the spectra lose detail at room temperature, hindering identification of the contributing species.

A technique showing early promise for application in bond-specific molecular detection in solution, is coherent resonant infrared wavemixing. Here, cross coupling between vibrational modes are detected, promising information more detailed than that of infrared absorption. Figure 21 shows an example of triply vibrationally enhanced four-wave mixing, where two infrared laser beams mix to deliver cross peaks representing vibration mode coupling. Here, the 2D spectrum contains two sets of cross-peaks corresponding to the couplings of the two strong IR modes at 1943 and 2002 cm^{-1} of $Ni(CO)_2(PPh_3)_2$ and 1996 and 2068 cm^{-1} of $Ni(CO)_3PPh_3$. Concentrations in solution are on the order of 10 mM. Such methods provide a means sensitive to chemical structure without a requirement for fluorophores.

Current application requires complex laser equipment delivering microjoule pulse energies, tunable between 3 and 10 microns, with pulse durations on the order of 1 ps. However, following the example of other laser diagnostic techniques in application

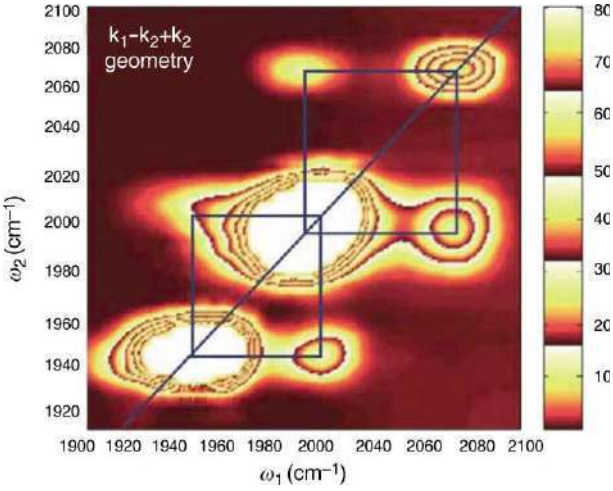

Figure 21 Triply vibrationally enhanced four-wave mixing of $Ni(CO)_2(PPh_3)_2$ and $Ni(CO)_3PPh_3$. Reproduced from Meyer KA, Besemann DM and Wright JC (2003) Coherent two dimensional spectroscopy with triply vibrationally enhanced infrared four-wave mixing. *Chemical Physics Letters* 381: 642–649, with permission of Elsevier.

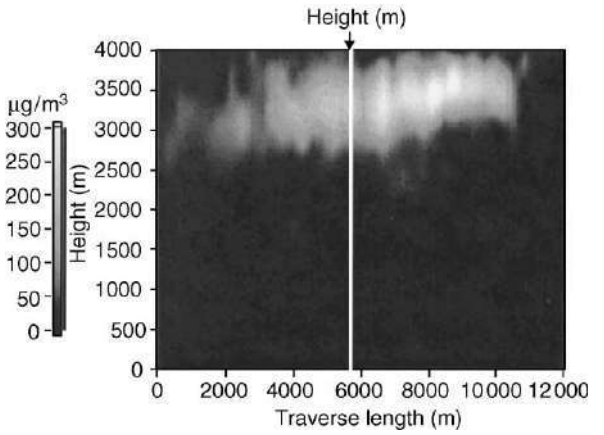

Figure 22 DIAL trace of sulfur dioxide distribution from Mount Etna. Reproduced with permission from Weibring P, Edner H and Svanberg S (2003) Versatile mobile lidar system for environmental monitoring. *Applied Optics* 42: 3583–3594. © 2003 Optical Society of America.

across industry today, as laser technology advances over this decade, we can expect such methods of probing chemical systems to see significant uptake, in particular in the bio-industries.

Laser Diagnostics and Treatment of Tumors

For clinical medicine, fluorescent detection is applied in the visualization of malignant tumors. Important natural chromophores emitting fluorescence in the blue-green region are elastin, collagen, NADH, and NAD^+, whilst administered agents such as porphyrins and phthalocyanines yield sharp peaks in the dark red.

Excitation is via UV or violet light. Where emission is monitored in both the blue-green and red regions of the spectrum and the ratio taken, an enhanced tumor differentiation is possible from that of normal tissue whilst simultaneously reducing sensitivity to geometry, illumination, and detection efficiency. Fiber optic probes are generally applied for delivery and collection of radiation.

Such detection can be combined with techniques of laser photodynamic therapy (PDT) to preferentially remove cancerous tissues. Typical laser wavelengths for photodynamic therapy are in the visible region of 600 to 700 nm with a pulse fluence of tens of joules/cm^2, a pulse duration in the nanosecond regime and a repetition rate from single shot to 10 Hz. Where combined with diagnostic detection, the PDT can be carried out by exciting the same administered sensitizer, with subsequent transfer of ground-state triplet oxygen to a toxic singlet state.

Remote Chemical Sensing

The atmosphere can be monitored by absorption and laser-induced fluorescence (LIF). Three-dimensional mapping of chemical distribution can be achieved with light detection and ranging (LIDAR), using a pulsed laser as a transmitter and an optical telescope as a receiver. Information collated includes pollution concentration, temperature, humidity, and wind velocity. Strong quenching prohibits LIF at atmospheric pressures; however, it does become practical to monitor for example, Li, Na, K, and Ca layers in the mesosphere. Pollution monitoring in the troposphere is most frequently performed with differential absorption LIDAR (DIAL). Range resolved optical transients are taken as a ratio of two laser wavelengths; on and just off resonance for the species of interest. Typical ranges for detection of SO_2, O_3, NO_2, NO, and Hg are 500 m to 5 km, making this technique suitable for urban and industrial monitoring. Sulfur dioxide is the most common pollutant emitted when fossil fuels are burned or sulfide ores are roasted and this technique sees application for the monitoring of industrial emissions. Figure 22 shows a DIAL trace recorded from Mount Etna, from which emissions of SO_2 are estimated at 75 tonnes/h. Laser wavelengths used are 300.02 and 299.30 nm for on and off resonance for absorption, respectively.

Important to the widespread application of LIDAR and LIF for atmospheric chemical monitoring, has been the advance in robustness and flexibility of the associated laser technologies. The laser technology applied for the above measurement is based on solid-sate Nd:YAG, nonlinearly converted via optical parametric amplification to wavelengths tunable between 220 nm and 4 μm, delivering up to 20 mJ in the mid-infrared at a repetition rate of 20 Hz and with bandwidths of less than 0.2 cm^{-1}, in a pulse duration of 4 ns.

See also

Imaging: Infrared Imaging. **Lasers:** Carbon Dioxide Lasers. **Microscopy:** Confocal Microscopy. **Scattering:** Raman Scattering. **Spectroscopy:** Raman Spectroscopy.

Further Reading

Allen SD, Bass M and Teisniger ML, Comparison of pulsed Nd:YAG and pulsed CO2 lasers for hole drilling in printed circuit board materials. CLEO 1982 Conference Summary.

Cheng J-X, Kevin Jia YK, Zheng G and Xie X-S (2002) Laser-scanning coherent anti-Stokes Raman scattering microscopy and applications to cell biology. *Biophysical Journal* 83: 502–509.

Cymer Inc., http://www.cymer.com

Cymer Inc., ASML, TRW, Xtreme Technologies GmbH, Fraunhoffer Institut fur Lasertechnik, Philips Extreme UV GmbH, Royal Institute of Technology Sweden, PLEX LLC.
Disco Corporation, http://www.disco.co.jp/
Exitech Ltd., http://www.exitech.co.uk/
Lamda Physik, http://www.lambdaphysik.de/
Meyer KA, Besemann DM and Wright JC (2003) Coherent two dimensional spectroscopy with triply vibrationally enhanced infrared four-wave mixing. *Chemical Physics Letters* 381: 642–649.
Perkin Elmer, www.perkinelmer.com
Phillips D (1997) Chemical mechanisms in photodynamic therapy with phthalocyanines. *Progress in Reaction Kinetics* 22(3–4): 175–300.
Potma EO and Xie X-S (2003) Detection of single lipid bilayers with coherent anti-Stokes Raman scattering (CARS) microscopy. *Journal of Raman Spectroscopy* 34: 642–650.
Powerlase Ltd., http://www.powerlase.com
Presstek Inc., Hudson, NH, http://www.presstek.com
Pronko PP, Dutta DK, Squier J, Rudd JV, Du D and Mourou G (1995) Machining of sub-micron holes using a femtosecond laser at 800 nm. *Optics Communications* 114: 106–110.
Reiser A (1989) *Photoreactive Polymers – The Science and Technology of Resists*. John Wiley & Sons.
Silverman PJ (2002) Intel Corporation. Presented at 1st Intl EUV Lithography Symposium, 15–17 October 2002, Dallas, USA.
Stuart BC, Feit MD, Herman S, Rubenchik AM, Shore BW and Perry MD (1996) Nanosecond-to-femtosecond laser-induced breakdown in dielectrics. *Physics Review B* 53(4): 1749.
Svanberg S (1996) Chemical sensing with laser spectroscopy. *Sensors and Actuators B* 33: 1–4.
Wang I, Bendsoe N, Klinteberg CAF, *et al*. (2001) Photodynamic therapy vs. cryosurgery of basal cell carcinomas: results of a phase III clinical trial. *British Journal of Dermatology* 144(4): 832–840.
Watson MN (1984) Laser drilling of printed circuit boards. *Circuit World* 11: 13.
Waynant RW (ed.) (2002) *Lasers in Medicine*. CRC Press.
Weibring P, Edner H and Svanberg S (2003) Versatile mobile lidar system for environmental monitoring. *Applied Optics* 42(18): 3583–3594.
WITEC, WITec Wissenschaftliche, Instrumente und Technologie GmbH, Hörvelsinger Weg 6, D-89081 Ulm, Germany, www.WITec.de
Yvon J, www.jobinyvon.com

Sum-Frequency Generation at Surfaces

M B Raschke and Y R Shen, University of California, Berkeley, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Under sufficiently intense illumination the optical properties of a material system depend nonlinearly on the strength of the electromagnetic field. Related to this nonlinear optical response is a large number of phenomena and fundamental processes which are discussed in various articles of this encyclopedia. This article will focus on second-harmonic generation (SHG) and sum-frequency generation (SFG) for the investigation of surfaces and interfaces. The intrinsic surface sensitivity of these techniques allows for investigations of surface properties not readily accessible by other spectroscopies. Here, the basic principles of these optical processes as well as their experimental implementation are discussed, and a summary of the applications to different material systems is given.

Among the various techniques employed for the characterization of surfaces and interfaces, those using light are particularly attractive. They are applicable *in situ* to all interfaces accessible by light, are nondestructive, and offer unprecedented time resolution. However, the penetration depth of optical radiation in matter is generally of the order of a wavelength, which makes isolation of the surface or interface contribution to the optical response from the bulk contribution difficult. In contrast, for reasons of symmetry, SHG and SFG are intrinsically surface sensitive in media with inversion symmetry, and hence the signal generated mainly originates from the topmost surface layer where the inversion symmetry is broken. By means of electronic or vibrational SHG or SFG spectroscopy, information on surface structure, chemical composition and bond or molecular orientation at solid and liquid interfaces can be deduced. To date, SFG and SHG have been well established as important tools for the investigation of surfaces and interfaces of solids ranging from metals and semiconductors to insulators and magnetic materials, liquids, polymers, biological membranes, and other systems. The studies are motivated by fundamental interests as well as applications in many areas such as heterogeneous catalysis, electrochemistry, device fabrication, epitaxial growth, and environmental science.

Theory

The nonlinear optical response of a material, in the electric-dipole approximation, can often be described by an induced polarization in the form of a power series,

$$\begin{aligned} P(\omega) = \epsilon_0[&\chi^{(1)}(\omega)E(\omega) \\ &+\chi^{(2)}(\omega=\omega_1+\omega_2)E(\omega_1)E(\omega_2) \\ &+\chi^{(3)}(\omega=\omega_1+\omega_2+\omega_3)E(\omega_1)E(\omega_2)E(\omega_3)+\ldots] \end{aligned} \quad [1]$$

where $E(\omega_i)$ is the optical field with frequency ω_i and $\chi^{(n)}$ is the nth-order linear ($n=1$) or nonlinear ($n>1$) susceptibility tensor. The second term, being the lowest-order nonlinear optical response, is responsible for sum-frequency generation and second-harmonic generation (with $\omega_1=\omega_2$). In the electric-dipole approximation, all even-order terms ($\chi^2, \chi^{(4)}$, etc.) are forbidden in media with inversion symmetry. At the surface or interface, however, the inversion symmetry is necessarily broken and hence $\chi^{(2)} \neq 0$. This makes SHG and SFG surface-sensitive and specific. While electric-quadrupole and magnetic-dipole contributions from the bulk may not be totally negligible, in many cases it has been shown that the surface contribution clearly dominates.

Surface SFG and SHG is best described by radiation from a surface polarization $P_s^{(2)}(\omega)$ induced in a thin sheet with dielectric constant ε' sandwiched between two linear media (1) and (2) as shown schematically in Figure 1. This surface polarization has an in-plane wavevector component equal to the sum of the in-plane wavevector components of the incident fields:

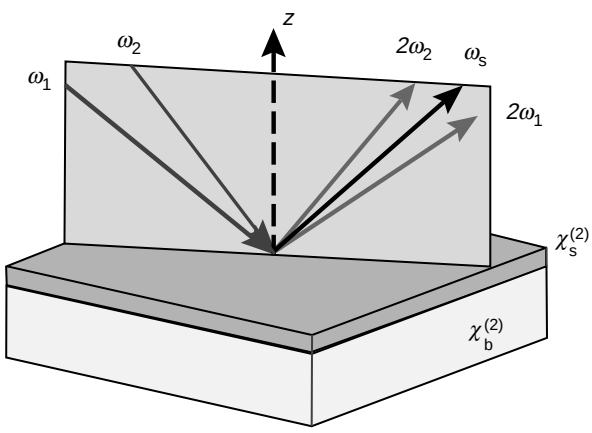

Figure 1 Schematic of SHG and SFG geometry in reflection and transmission from an interface.

$k_{\parallel}(\omega) = k_{\parallel}(\omega_1) + k_{\parallel}(\omega_2)$. The ratio of $k_{\parallel}(\omega)$ to $k(\omega)$ then determines the direction of the sum-frequency output that appears in both transmission and reflection. The reflection geometry is often used in order to minimize the bulk contribution.

With $P_s^{(2)}$ as the source term, the sum-frequency output can be obtained from the solution of the wave equation. With the proper boundary conditions the sum frequency (SF) intensity is given by:

$$\begin{aligned} I(\omega) = &\frac{8\pi^3\omega_s^2\sec^2\theta_\omega}{\hbar c^3\sqrt{\epsilon_1(\omega)\epsilon_1(\omega_1)\epsilon_1(\omega_2)}} \\ &\times |e^{\dagger}(\omega)\cdot\chi_s^{(2)} : e(\omega_1)e(\omega_2)|^2 I_1(\omega_1)I_2(\omega_2) \end{aligned} \quad [2]$$

In this expression, $\chi_s^{(2)}$ is the surface nonlinear susceptibility defined by $P_s^{(2)} = \varepsilon_0\chi_s^{(2)} : E(\omega_1)E(\omega_2)$; the quantities $e(\omega_i)$ correspond to the unit polarization vectors after the appropriate Fresnel correction, $e(\omega_i) \equiv F(\omega_i)\hat{e}(\omega_i)$, with $F(\omega_i)$ being the transmission Fresnel factor and $\hat{e}(\omega_i)$ the unit polarization vector of $E(\omega_i)$; θ_ω denotes the SF output angle with respect to the surface normal, $\varepsilon(\omega_i)$ is the dielectric constant at frequency ω_i, and I_i is the input pump intensity at ω_i.

The surface nonlinear susceptibility $\chi_s^{(2)}$ is a third-rank tensor. In Cartesian coordinates it has 27 tensor elements ($\chi_{ijk}^{(2)}$), many of which could vanish or depend on others due to the surface structural symmetry. As an example, $\chi_{ijk}^{(2)}(\omega=\omega_1+\omega_2)$ for an isotropic surface, with the z-direction defined by the surface normal, has only four independent nonvanishing elements: $\chi_{xxz}^{(2)} = \chi_{yyz}^{(2)}, \chi_{xzx}^{(2)} = \chi_{yzy}^{(2)}, \chi_{zxx}^{(2)} = \chi_{zyy}^{(2)}$, and $\chi_{zzz}^{(2)}$. Different combinations of input and output beam polarizations in SFG measurements are often used to deduce values for the non-vanishing elements. Such measurements then allow the determination of surface symmetry or surface molecular orientation. Being a third-rank tensor, $\chi_s^{(2)}$ can reflect up to three-fold rotational symmetry. $\chi_{s,ijk}^{(2)}$ is related to the molecular nonlinear polarizability or hyperpolarizability $\alpha_{\xi\eta\zeta}^{(2)}$, where $\hat{\xi}$, $\hat{\eta}$, and $\hat{\zeta}$ define the molecular coordinates, through the coordinate transformation

$$\chi_{s,ijk}^{(2)} = N_s\left\langle(\hat{i}\cdot\hat{\zeta})(\hat{j}\cdot\hat{\eta})(\hat{k}\cdot\hat{\xi})\right\rangle\alpha_{\xi\eta\zeta}^{(2)} \quad [3]$$

The angular brackets denote an average over the molecular orientational distribution, and N_s is the surface molecular density. For simplicity, microscopic local-field correction is neglected in eqn [3]. Knowing $\chi_{s,ijk}^{(2)}$ and $\alpha_{\xi\eta\zeta}^{(2)}$ thus permits deduction of information on the orientational distribution.

An explicit expression for the nonlinear optical polarizability is obtained from a second-order quantum mechanical perturbation calculation:

$$\alpha^{(2)}_{s,ijk}(\omega_s) = -\frac{e^3}{\hbar^2}\sum_{g,n,n'}\left[\frac{\langle g|r_i|n\rangle}{\omega_s - \omega_{ng} + i\Gamma_{ng}} \times \left(\frac{\langle n|r_j|n'\rangle\langle n'|r_k|g\rangle}{\omega_2 - \omega_{n'g} + i\Gamma_{n'g}} + \frac{\langle n|r_k|n'\rangle\langle n'|r_j|g\rangle}{\omega_1 - \omega_{n'g} + i\Gamma_{n'g}}\right) + \cdots\right]\rho^{(0)}_g \quad [4]$$

This expression shows how the nonlinear polarizability or susceptibility depends on material parameters such as the dipole transition moments $\langle n|r_i|g\rangle$ and energy levels. It contains a sum over eight resonant terms. The quantities ω_{ng} and Γ_{ng} are the frequencies and half-widths for the transitions between quantum states $|n\rangle$ and $|g\rangle$, and $\rho^{(0)}_g$ denotes the population in $|g\rangle$. It can be seen that $\alpha^{(2)}$, and hence the SF output, should be resonantly enhanced when the pump frequency ω_1 or ω_2 and/or the sum frequency $\omega_1 + \omega_2$ approach resonance. The resonant enhancement provides spectral information and makes SHG and SFG effective spectroscopic tools. The resonances could be electronic or vibrational, or more generally, any surface characteristic transition.

Experimental

Here we deal with basic considerations concerning the optical setup for SHG or SFG experiments in reflection from a surface as is shown schematically in Figure 2. In general, the pump radiation from pulsed laser sources is directed onto the surface. High-power tunable light pulses can be obtained from optical parametric generation and amplification (OPG/OPA) together with harmonic, sum- or difference-frequency generation, preferably pumped by picosecond or femtosecond lasers with high repetition rates. By this means radiation tunable from near UV at ~200 nm to mid IR at ~18 μm can be generated. For SFG the input pulses with frequencies ω_1 and ω_2 are directed to overlap spatially and temporally on the sample. The sample could be the surface of a liquid or solid in air, a single crystalline surface in ultrahigh vacuum, buried interfaces, etc. Polarizers and half-wave plates allow for tuning of incident power and polarizations at the sample. Together with the polarizer in the output path, different input/output polarization combinations can be chosen. In the detection system the signal has to be discriminated against reflected and scattered pump light. This is achieved by spatial and spectral filtering with apertures, dielectric mirrors, and interference and color glass filters. An optional monochromator can also be used for additional stray-light suppression. The signal is then detected by a photomultiplier tube and gated electronics are used for signal integration or photon counting.

From eqn [2] the expected signal strength for SHG/SFG can be estimated. With a typical value of $\chi^{(2)}_s$ of 10^{-21} m^2 V^{-1}, a single 1-μm pump pulse, incident at $\theta = 45°$ having pulse energy $E = I\,A\tau = 100$ μJ, beam cross-section $A = 1$ mm^2, and pulse duration $\tau = 10$ ps, can generate about 10^3 photons per pulse of SHG. By means of photon counting a minimal count rate of 10^{-3} photons/pulse

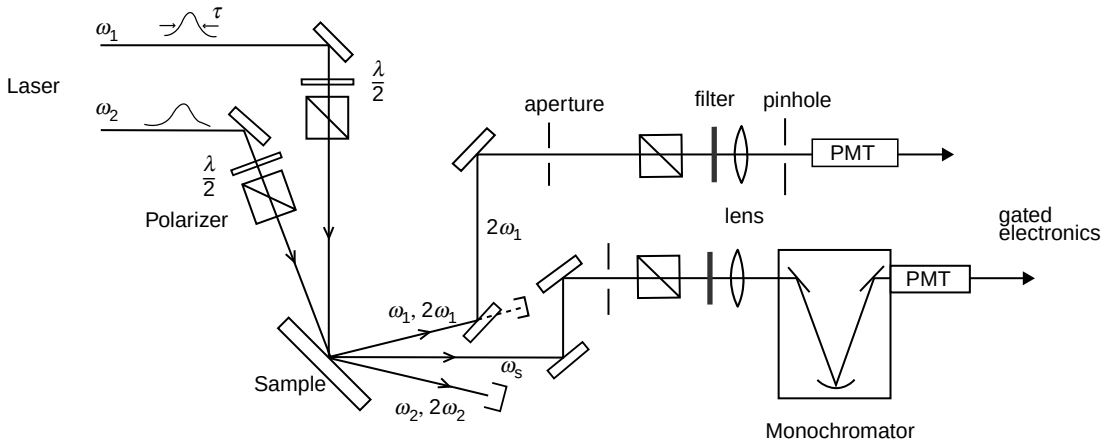

Figure 2 Schematic of experimental setup for second-harmonic and sum-frequency generation.

can be achieved which allows the detection of $\chi_s^{(2)}$ as small as 10^{-24} $m^2 V^{-1}$. The spectral resolution is limited by the spectral width of the pump pulses. In order to obtain accurate spectral data over a wide tuning range or to obtain absolute values for $\chi_{s,ijk}^{(2)}$ of the material investigated, a reference material such as *z*-cut quartz with known values of χ is to be measured for comparison.

The simplest kind of experiments are SHG at a fixed pump frequency. From the signal variation in response to surface modification one can probe kinetics and dynamics of adsorption, desorption, diffusion, surface melting, phase transitions, etc. By tuning the pump wavelength, second-harmonic spectroscopy can give information on surface electronic states. For surface vibrational spectroscopy, however, IR-vis SFG must be employed. In the latter case, tunable IR input is mixed with visible input to yield a SF signal in the visible region.

Surface Specificity

In many cases, the surface contribution to SFG or SHG from a centrosymmetric medium clearly dominates over the bulk contribution. This generally occurs when the surface or interface is composed of a polar oriented molecular layer. An example is shown in Figure 3, where the vibrational spectra in the C–H stretch region of three buried solid–liquid interfaces are presented. The spectrum for hexadecane ($C_{16}H_{34}$)/fused silica shows no C–H peaks. This is in spite of the fact that liquid hexadecane has strong infrared and Raman activity in this spectral range. The result indicates that the bulk signal contribution is negligible. In contrast, strong resonances are observed from the hexadecane/OTS/silica system, where OTS is a monolayer of octadecyltrichlorosilane ($CH_3(CH_2)_{17}SiCl_3$) chemisorbed on the fused silica plate. A similar spectrum is obtained for the CCl_4/OTS/silica interface. This clearly indicates that it is the contribution from the OTS monolayer that dominates the spectra.

Another example is described in Figure 4 where the SFG spectrum in reflection from a free liquid water surface is shown. The resonant features between 3000 and 3600 cm^{-1} are due to the O–H stretches of hydrogen-bonded OH groups. The sharp peak at 3700 cm^{-1} can be identified with the free OH bonds. Since no dangling OH bonds can exist below the surface its presence indicates that the spectrum must originate from the topmost layer of water molecules. Furthermore it can be shown that the surface water molecules are oriented with one of the OH bonds directed out of the liquid. In addition to the above examples, the surface sensitivity and specificity has been demonstrated and successfully applied to a large number of systems as will be shown below.

A note on the definition of surface or interface is in order. In SHG and SFG, the surface or interface layer refers to a thin layer between two adjacent bulk media that has a different structure from the bulk media and lacks inversion symmetry. If molecules in the surface layer are polar-ordered, then SHG and

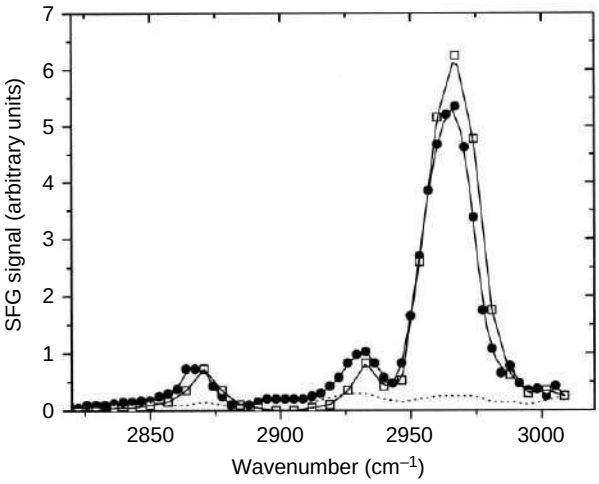

Figure 3 SFG spectra of different interfaces obtained with *ssp* polarization combination: hexadecane/silica (dashed), CCl_4/ OTS/silica (solid circles), and hexadecane/OTS/silica (open squares). OTS refers to a monolayer of octadecyltrichlorosilane.

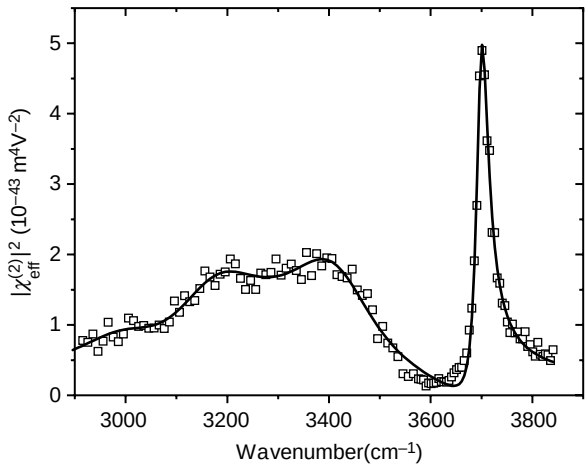

Figure 4 SFG spectrum of the liquid–vapor interface of pure water (*ssp* polarization combination). The mode at 3700 cm^{-1} is due to the dangling OH bonds at the surface.

SFG are often dominated by contributions from the surface layer. However, the bulk contribution is not strictly zero even if the media have inversion symmetry, as $\chi^{(2)}$ can arise from electric-quadrupole and magnetic-dipole contributions. Separation of surface and bulk contributions in SHG and SFG is a subtle problem in general. Experimentally, how surface modification affects the signal can be used to evaluate whether or not the surface contribution dominates.

A special case concerns materials that lack bulk inversion symmetry, for example, III–V and II–VI semiconductors such as GaAs and ZnSe. In this case, the bulk contribution to SHG and SFG is electric-dipole allowed and may be very significant. With an appropriate choice of polarization combinations and surface orientation, it can nevertheless be well suppressed, leaving the surface contribution dominant.

Applications

Both SFG and SHG have been applied to a wide variety of surface and interfacial systems. To illustrate their capabilities and versatility as surface spectroscopic tools, the focus of this summary will be on examples where the information deduced cannot readily be obtained by other techniques. (For details, readers are referred to the review articles cited in Further Reading and the original papers cited therein.)

SFG and SHG have been used for the study of adsorbates at surfaces even under ambient conditions. The adsorbate alters electronic or structural properties of the substrate and the signal is thus related to the surface density of the adsorbate and its orientation.

Elementary surface reactions, in particular adsorption, dissociation, or desorption of simple molecules such as CO, hydrogen, oxygen, or water on metal and semiconductor single-crystal surfaces under ultra-high-vacuum conditions, have been investigated. Sensitivity into the sub-percent monolayer regime is achieved and allows the identifictation of competing reaction channels and adsorbate dynamics on surfaces. Such studies can help to understand epitaxial growth or heterogeneous catalysis because the techniques are applicable at any gas pressure and surface temperature. Related to this is the study of surface diffusion, which can be probed following the temporal decay of a diffracted SHG signal off a submonolayer adsorbate grating formed by laser-induced desorption.

Making use of the spectroscopic capabilities of SHG, the surface electronic structure of selected metals and semiconductors, in particular silicon, were studied with an emphasis on the effects of temperature, crystallographic orientation and different adsorbates. Among buried interfaces, the Si/SiO_2 interface has been the subject of intense investigations and the sensitivity of SHG to static electric fields at the interface and to inhomogeneous strain was also demonstrated.

Different in-plane surface symmetries associated with different crystallographic orientations or surface reconstructions yield different SHG and SFG responses. This allowed for the investigation of surface reconstructions (e.g., $Si(111)2\times1 \rightarrow 7\times7$), order–disorder transitions (e.g., $Si(111)7\times7 \rightarrow 1\times1$, Au/Si(111)) and surface melting (e.g., single-crystal ice, Si(111), Ge(111)), providing information on latent heat, superheating, and its dynamics using pump-probe techniques. Magnetization-induced SHG was also used for probing surface and interface magnetism of ferromagnetic metals and bimetallic systems.

Many unique applications of spectroscopic SFG have been developed based on its capability to study large molecules, in particular under ambient conditions. Surface vibrational spectra with different input/output polarization combinations and sample geometries provide information on orientation, conformation and alignment of surface molecules and the composition and structure of the surface layer. The study of self-assembled monolayers (e.g., OTS on quartz) to deduce the orientation and conformation of the alkyl chains in the monolayer is an example. The effect of molecular density and temperature on the spectrum gives information on intermolecular interactions and phase transitions. Another example concerns the adsorption of liquid crystal molecules on nanostructured polymer substrates. Both the chain orientation at the polymer surface and its effect on the alignment of adsorbed liquid crystal molecules can be addressed.

Surface structures of liquids, especially those of pure liquids, are of great interest in science and technology, and SFG is a unique spectroscopic tool for probing liquid surfaces and interfaces. For example, the SFG vibrational spectrum for the water–vapor interface in Figure 4 gives information on the density and orientation of the surface water molecules at the water–vapor interface. The surface layer was found to be more ordered compared to the bulk as was also observed for various other organic liquids studied with SFG. In other cases soluble and insoluble organic molecules at a liquid–vapor,

liquid–liquid, or liquid–solid interface have also been studied. They appear as adsorbed monolayer films and their structure and phase behavior are relevant to many applications. Identification of the adsorbed molecular species at solid–liquid interfaces is also important for understanding electrochemical processes. Various studies with SFG have addressed bonding of, e.g., CO, CN^-, and SCN^- at electrodes and its variation with electrode potential.

SFG spectroscopy also finds unique applications in studies of polymer surfaces and interfaces most relevant to modern science and technology. The observed surface vibrational spectra of polymers provide information on surface composition, molecular orientation, and conformation of neat and blended polymers. The effect of environment and surface treatment on the surface structure can also be studied. Investigation of the biological functions of complex molecular systems has attracted increasing interest, and SHG and SFG have been applied successfully to selected systems ranging from the isomerization of retinal – a molecule involved in the vision process – to functional aspects of model membranes.

Both SHG and SFG can be combined with microscopy techniques for surface microscopic studies. Microscopic imaging can be achieved by rastering the probe beam or, preferably, the sample position. Near-field SFG/SHG spectroscopy has recently been developed.

Performing surface SHG and SFG experiments in the time domain can provide information on surface dynamics. Pump-probe experiments allow for measurements of energy relaxation and phase coherence of excitations such as surface electronic states and surface phonons and vibrations.

Outlook

Although SHG and SFG were established as surface analytical tools more than 15 years ago, it is still an active field of research with much potential that has not yet been explored. As an example, it has only recently been demonstrated that doubly resonant infrared–visible SFG, as a two-dimensional spectroscopy, can give access to couplings between vibrational modes and surface electronic transitions. With the commercial availability of suitable laser sources, the techniques can be extended to a much wider range of applications. This includes possible investigations of ultrafast surface dynamics, nanostructures, and biological systems.

Further Reading

Bennemann KH (ed.) (1998) *Nonlinear Optics in Metals*. Oxford, UK: Clarendon Press.

Bloembergen N (1977) *Nonlinear Optics*. New York: Benjamin.

Boyd RW (1993) *Nonlinear Optics,* 2nd edn. San Diego, CA: Academic Press.

Dumas P, Weldon MK, Chabal YJ and Williams GP (1999) Molecules at surfaces and interfaces studied using vibrational spectroscopies and related techniques. *Surface Review & Letters* 6(2): 225–255.

Guyot-Sionnest P and Harris AL (1995) Surface vibrational dynamics probed by sum frequency generation. In: Ho W and Dai HL (eds) *Laser Spectroscopy and Photochemistry of Metal Surfaces*, pp. 405–458. Singapore: World Scientific.

Guyot-Sionnest P, Superfine R, Hunt JH and Shen YR (1988) Vibrational spectroscopy of a silane monolayer at air/solid and liquid/solid interfaces using sum-frequency generation. *Chemical Physics Letter* 144: 1–5.

Heinz TF (1991) Second-order nonlinear optical effects at surfaces and interfaces. In: Stegeman GI and Ponath H-E (eds), *Nonlinear Surface Electromagnetic Phenomena*, pp. 353–416. Amsterdam, The Netherlands: North-Holland.

Matthias E and Träger F (eds) (1999) *Applied Physics B, Special Issue: Nonlinear Optics at Interfaces*, Volume 210.

McGilp JF (1999) Optical second-harmonic generation as a semiconductor surface and interface probe. *Physica Status Solidi (a)* 175: 153–167.

Miranda PB and Shen YR (1999) Liquid interfaces: A study by sum-frequency vibrational spectroscopy. *Journal of Physical Chemistry B* 103: 3292–3307.

Reider GA and Heinz TF (1995) Second-order nonlinear optical effects at surfaces and interfaces: recent advances. In: Halevi P (ed.) *Photonic Probes of Interfaces*, 413. Amsterdam, The Netherlands: Elsevier.

Shen YR (1984) The Principles of Nonlinear Optics. New York: Wiley.

Shen YR (1994) Nonlinear optical studies of polymer interfaces. *International Journal of Nonlinear Optical Physics* 3: 459–468.

Shen YR (1994) Surface spectroscopy by nonlinear optics. In: Ḧansch TW and Inguscio M (eds) *Frontiers in Laser Spectroscopy*, pp. 139–165. Amsterdam, The Netherlands: North-Holland.

Shen YR (2001) Vibrational spectroscopy of water at interfaces. In: Morra M (ed.) *Water in Biomaterials Surface Science*, Chapter 8, pp. 215–244. Chichester, UK: Wiley.

Wei X and Shen YR (2001) Motional effect in surface sum-frequency vibrational spectroscopy. *Physical Review Letter* 86: 4799–4802.

POLARIZATION

Contents

Introduction
Matrix Analysis

Introduction

J M Bennett, Michelson Laboratory, China Lake, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Polarization is one property of light waves and will be defined after a brief overview of the properties of light. Light is part of the electromagnetic spectrum of waves that have both particle-like and wave-like properties. Light waves carry energy in the form of photons which act like particles; the photon energy increases in proportion to the frequency of the wave. The particle-like properties of light and other electromagnetic waves are described by quantum mechanics. Light also acts like transverse waves that travel in straight lines in air and vacuum and can be described by classical electromagnetic theory. Polarization deals with the wave-like properties of light, and is described mathematically by Maxwell's equations, the key relations in electromagnetic theory. Polarization effects occur when light interacts with matter. In order to understand polarization, there will be a brief introduction to the subject of electromagnetic waves.

For a more in-depth treatment of the material in this article, the reader is directed to the Further Reading list at the end of this article and in particular to the two chapters on 'Polarization' and 'Polarizers' written by the author in Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. I, chapter 5 and vol. II, chapter 3. New York: McGraw-Hill Inc.

Electromagnetic waves have both electric and magnetic fields associated with them. These are vibrations in directions perpendicular to the direction the wave is traveling, i.e., the direction of propagation. $\vec{E}$ represents the vector of the electric field and $\vec{H}$, perpendicular to $\vec{E}$, represents the magnetic field vector. Both these vectors are complex and have real and imaginary parts. Polarization effects are always associated with the $\vec{E}$ vector. Specifically, the plane of polarization is defined as the plane in which the $\vec{E}$ vector is vibrating. Waves having different amplitudes, phases, or angular orientations (azimuths) of their electric or magnetic vectors can be combined by conventional vector addition methods. Also the $\vec{E}$ vector of a particular vibration can be resolved into two components in mutually perpendicular directions that are vibrating in phase.

If a light source such as the sun, a candle flame, or an electric light bulb is considered on a microscopic scale, each vibrating atom or molecule emits linearly polarized light (see the definition below). But the individual atoms or molecules do not act together, so their vibrations have no fixed phase relationships to each other and they cannot be added into a single linearly polarized beam. Thus, we call light from these sources unpolarized. In an unpolarized light beam, the $\vec{E}$ vector vibrates in all directions perpendicular to the direction of propagation. If a snapshot is taken at a particular instant of time, different parts of the beam will have $\vec{E}$ vectors vibrating with different amplitudes and phases at different angles to each other, but all in a plane perpendicular to the direction of propagation. In the most common convention used in optics, the wave travels in the $+z$ direction in a right hand coordinate system and the $\vec{E}$ vectors are all vibrating at various angles in the $x-y$ plane. The angle of vibration is measured from the positive x axis in a counterclockwise direction when the observer is looking against the direction of propagation of the light beam.

For linearly polarized or plane polarized light, if a snapshot of a light beam is taken at a particular instant, the $\vec{E}$ vector will be vibrating at a certain angle in the $x-y$ plane. As time (or position on the traveling wave) varies, the amplitude of the $\vec{E}$ vector will vary in a sinusoidal manner, but the vibration will remain at the same angle in the $x-y$ plane. As an example, the real part of the electric vector of a linearly polarized beam that is vibrating in the $+x$ and $-x$ directions and traveling in the $+z$ direction is

given by the relation:

$$\vec{E}_x(z,t) = \hat{i}E_0 \cos(kz - \omega t) \qquad [1]$$

where t is the time, $\hat{i}$ is a unit vector in the $+x$ direction, E_0 is the amplitude of the vibration, $\vec{k}$ is the propagation vector in the direction the wave is traveling, i.e., the z direction (its magnitude is $2\pi/\lambda$), and $\omega = 2\pi f$ where f is the vibration frequency of the wave. In free space, $\omega = 2\pi c/\lambda$ where c is the velocity of light and λ is the wavelength of the vibration.

Light can also be circularly or elliptically polarized. Circularly polarized light is produced by adding the electric vectors of two waves, each having the same amplitude but which are 90° out of phase; the resulting vibration sweeps out a circle in the x–y plane. When the first wave is vibrating in the $+x$ and $-x$ directions (eqn [1]), the second wave is vibrating in the $+y$ and $-y$ directions:

$$\vec{E}_y(z,t) = \hat{j}E_0 \sin(kz - \omega t) \qquad [2]$$

and is added to the first wave. The resultant wave is the sum of eqns [1] and [2]:

$$\vec{E} = E_0[\hat{i}\cos(kz - \omega t) + \hat{j}\sin(kz - \omega t)] \qquad [3]$$

The direction of vibration of this wave will rotate in a circle as either the time increases or as the distance along the wave increases, but the amplitude of the vibration will not change. As time increases, the vibration will make a circle in the clockwise direction (to negative angles). This light is defined to be right circularly polarized. If the + sign between the two parts of eqn [3] is changed to a minus sign, as time increases, the vibration will make a circle in the counterclockwise direction (to positive angles) and the light is now defined to be left circularly polarized. These definitions are for conventional (traditional) optics. However, the opposite definitions with right (left) circularly polarized light defining circles in the counterclockwise (clockwise) direction are also in use. In modern physics, there is still another convention that defines right (left) circularly polarized light as having negative (positive) helicity.

Right and left elliptically polarized light beams have the same angle conventions as for circularly polarized beams but are produced by adding two electric vectors that have different amplitudes. In eqns [1] and [2], the amplitude terms will be, for example, E_1 and E_2 instead of E_0 and eqn [3] can no longer be used. There is now a major axis and a minor axis of the ellipse. If $E_1 > E_2$, the major axis will be along the x axis; for $E_1 < E_2$, the major axis will coincide with the y axis.

The preceding discussion has dealt entirely with the electric vector of the electromagnetic field. However, one cannot observe the $\vec{E}$ vector. The irradiance (energy per unit area per unit time), $\vec{E}\vec{E}^*$, is what can be observed visually and measured by electronic detectors. Thus, measurements of the polarization are irradiance measurements. Light transmitted by a polarizer is called the transmittance of the polarizer; similarly, light reflected from a polarizer is called the reflectance of the polarizer.

One of the most important parts of the subject of polarization is how to produce linearly polarized light starting with unpolarized light. This is done by using polarizers, as discussed in the section on polarizers below. Certain materials have special properties that make them able to polarize light. Depending on the application, different kinds of polarizers are preferred. Optics textbooks by Hecht and Guenther discuss the most important polarizers and two articles by Bennett describe many kinds of polarizers including special ones (see the Further Reading at the end of this article for full references). Only the basic principles will be described here.

Sunlight scattered by air molecules in the atmosphere (Rayleigh scattering) is also partially linearly polarized. The air molecules act like tiny dipoles and vibrate when they absorb sunlight. They emit radiation that is polarized in certain directions relative to the vibration direction. When the viewer is at a 90° angle with respect to the sun, the polarization of the skylight is a maximum. Rayleigh scattering is the subject of several books and will not be further discussed here.

Retarders are used to change linearly polarized light into circularly or elliptically polarized light and compensators, which are a form of retarders, can analyze an unknown type of polarized light and determine its composition. They are discussed below.

Polarimetry and ellipsometry are closely related techniques that are used to determine the optical properties of a material by shining a known type of polarized light on the material at non-normal incidence and analyzing the polarization properties of the light after it has been reflected from the material. These subjects are treated in other articles in this encyclopedia and in several references at the end of this article.

Changes can be produced in the optical properties of some materials by applying an electric field, a magnetic field, an acoustic field, or another form of variable pressure. The materials changed in these ways are said to be electro-optic, magneto-optic, or piezo-optic. The changes in the optical properties modify some parameter of a light wave passing through a material or reflecting from it.

The amplitude, phase, frequency, polarization, or direction of the light wave can be modified. In this article, we are only concerned with modifications that produce changes in the polarization; these are discussed towards the end of this article.

Matrix methods have been developed to handle problems involving polarization and there is also a visual representation of the matrix algebra that is based on the Poincaré sphere. Both topics are covered at the end of this article.

Polarizers

Basic Relations for Polarizers

A linear polarizer is anything which, when placed in an incident unpolarized beam, produces a beam of light whose electric vector is vibrating primarily in one direction with only a small component vibrating in the direction perpendicular to it. The transmittance T of the linear polarizer is

$$T = \tfrac{1}{2}(T_1 + T_2) \qquad [4]$$

where T_1, the principal transmittance of the polarizer, is $\gg T_2$, the transmittance of the polarizer at 90° to the principal transmittance. Thus, a perfect polarizer would transmit only 50% of an incident unpolarized beam.

If a linear polarizer is placed in a linearly polarized beam and is rotated about an axis parallel to the beam direction, the transmittance will vary between a maximum value T_1 and a minimum value T_2 according to the law:

$$T = (T_1 - T_2)\cos^2\theta + T_2 \qquad [5]$$

where θ is the angle between the plane of vibration of the principal transmittance and the plane of vibration of the electric vector in the incident beam.

The extinction ratio ρ_P of a polarizer is defined as

$$\rho_P = \frac{T_2}{T_1} \qquad [6]$$

and the degree of polarization of a polarization P is

$$P = \frac{T_1 - T_2}{T_1 + T_2} \qquad [7]$$

When two identical polarizers are placed in an unpolarized beam, and the directions of their principal transmittances, T_A and T_B, are inclined at an angle θ to each other, the transmittance of the pair will be

$$T_\theta = \tfrac{1}{2}(T_1^2 + T_2^2)\cos^2\theta + T_1 T_2 \sin^2\theta \qquad [8]$$

Thus, when the directions of the principal transmittances are aligned, $T_\parallel = \frac{1}{2}(T_1^2 + T_2^2)$, and when they are perpendicular, $T_\perp = T_1 T_2$.

Birefringent Materials (Calcite)

The majority of high-quality polarizers are made from calcite. This is a birefringent (doubly refracting) crystalline material that is uniaxial (i.e., there is one preferred direction in the crystal). A birefringent material acts differently for light going in different directions through the crystal. For example, if an unpolarized light beam passes through the crystal in a certain direction, it will be split into two spatially separated beams that are parallel but are linearly polarized at right angles to each other. A uniaxial crystal has an optic axis (i.e., a certain direction in the crystal). When light rays travel parallel to the optic axis, they travel at the same velocity and there is no difference between them. When light passes through the crystal in other directions, the ray whose vibration direction is perpendicular to the optic axis is governed by the ordinary laws of geometrical optics (the same as for isotropic materials) and is called the ordinary ray. The ray whose vibration direction is parallel to the optic axis does not follow the normal geometrical optics laws and is called the extraordinary ray. One ray travels faster than the other, so there is a phase retardation for one ray relative to the other. This is the principle of a retarder or retardation plate (see the next section).

Calcite is a negative uniaxial crystal which means that the refractive index for the ordinary ray is larger than the refractive index for the extraordinary ray. When the ordinary ray enters a block of calcite from air at non-normal incidence, it is bent more than the extraordinary ray.

Calcite can be easily cleaved along three distinct planes, making it possible to produce rhombs of the form shown in Figure 1. The optic axis, going in the HI direction, makes equal angles with all three faces at point H. Any plane, such as DBFH, which contains the optic axis and is perpendicular to the two opposite faces of the rhomb ABCD and EFGH is called a principal section. If the plane of incidence of light on the rhomb coincides with a principal section, the light entering the crystal will be split into two components polarized at right angles to each other which travel in slightly different directions and leave the crystal as two beams slightly displaced but parallel to each other.

The large birefringence of calcite and its excellent transmission through the visible spectral region and into the ultraviolet and infrared regions has made it possible to make excellent high extinction ratio

polarizers with different designs. Some of these are shown in Figure 2. There are two main types: Glan types with rectangular shapes and Nicol types with rhombohedral shapes. The polarizers are made of two pieces of calcite cemented together. Glan types have their optic axes in the plane of the entrance face. In the Nicol types, the principal section is perpendicular to the entrance face, but the optic axis is neither parallel nor perpendicular to the face. The two halves of conventional polarizing prisms are cemented together with cement that has a refractive index intermediate between the ordinary and extraordinary refractive indices. This enables one ray (generally the extraordinary, or *e* ray) to be transmitted and the other to be reflected at the cut, so that only one ray exits from the prism in the direction of the incoming ray. The Glan types are also used without cement with only an air space between the two halves. They can be used at shorter wavelengths in the ultraviolet where the cement absorbs. The extinction ratio can be very high for air-spaced prisms. For example, for a Glan–Foucault prism (an air-spaced Glan–Thompson prism, Figure 2), the extinction ratio can be better than 1×10^{-5} and the prism is usable from about 0.214 μm in the ultraviolet to 2.3 μm in the infrared. However, air-spaced polarizers have very small field angles, so they are mainly used with laser sources where the beam is parallel.

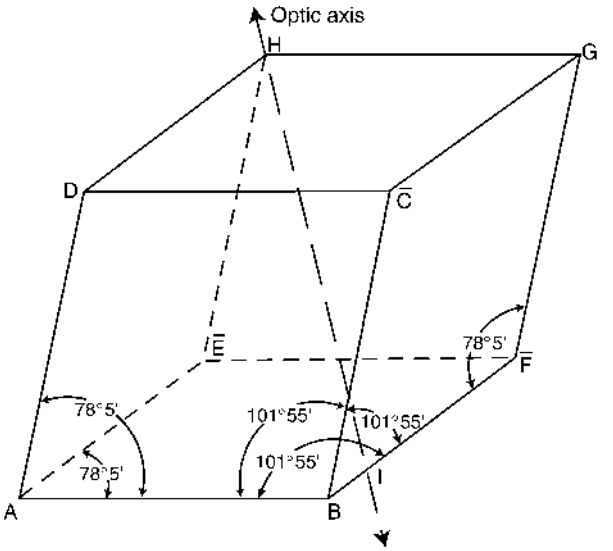

Figure 1 Schematic representation of a rhombohedral calcite crystal showing the angles between faces. The optic axis passes through corner H and point I on side BF. (Reproduced with permission from Bennett JM (1995) Polarizers. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 3. New York: McGraw-Hill, Inc.)

The extreme paths of light through a cemented Glan–Thompson prism are shown in Figure 3. This prism has a length that is three times the width of the entrance aperture (i.e., an *L*/*A* ratio of 3). The optic axis is perpendicular to the plane of incidence which is in the plane of the paper. In the first half of the polarizer, the paths of the ordinary and extraordinary rays, both of which follow the conventional law of refraction (Snell's Law, eqn [8] above) nearly coincide. Ray A is incident on the entrance face of the polarizer at an angle such that the angle of incidence on the cut is the smallest angle for which the *o* ray is totally internally reflected. Ray B is incident at an angle such that the angle of refraction in the first half of the prism is essentially equal to the cut angle, S, so that the *e* ray just passes through the cut. The field angle of the polarizer is twice the smaller of angles θ_A and θ_B. Thus, all rays having angles of incidence between rays A and B will be polarized

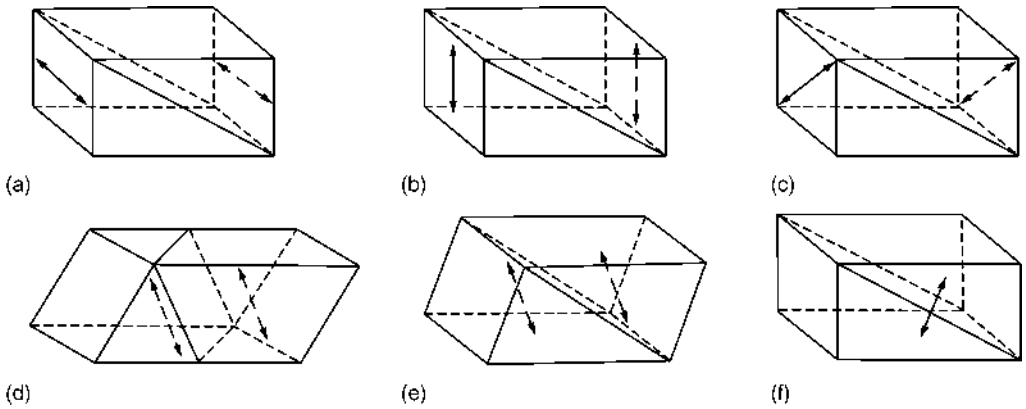

Figure 2 Types of polarizing prisms. Glan types: (a) Glan-Thompson, (b) Lippich, and (c) Frank-Ritter; Nicol types: (d) conventional Nicol, (e) Nicol, Halle form, and (f) Hartnack-Prazmowsky. The optic axes are indicated by the double-pointed arrows. (Reproduced with permission from Bennett JM (1995) Polarizers. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 3. New York: McGraw-Hill, Inc.)

when exiting the prism. The paths of similar rays can be traced through the other prism designs.

In addition to conventional polarizing prisms, there are also polarizing beamsplitter prisms and Feussner prisms. In polarizing beamsplitter prisms, the two beams that are polarized at right angles to each other both emerge from the prism but are separated spatially. Figure 4 shows a diagram of several types of polarizing beamsplitter prisms and Figure 5 shows the paths of rays through side views of these prisms.

A Feussner prism is made of isotropic material but the film separating the two halves is birefringent.

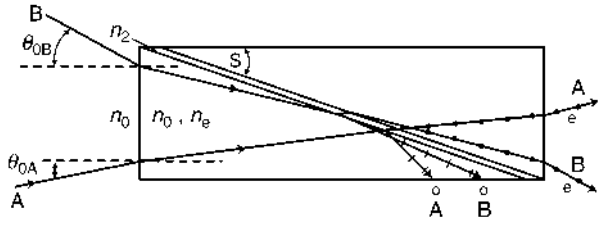

Figure 3 Extreme rays (A and B) that can pass through a cemented Glan-Thompson prism. Both rays change to ordinary and extraordinary rays in the calcite. Both ordinary rays are reflected at the boundary between the two halves of the prism and the extraordinary rays exit in the directions indicated. Rays entering the prism between rays A and B would be transmitted by the prism. The field angle of the prism is twice the smaller of the two angles of incidence (ray A).

These prisms have the advantage that much less birefringent material is required but they have a more limited wavelength range when calcite or sodium nitrite (another birefringent material) is used because the transmitted ordinary ray has a shorter transmission range than the extraordinary ray which is transmitted by the conventional and air-spaced polarizing prisms.

Dichroic Absorbers

A dichroic material is one that absorbs light polarized in one direction more strongly than light polarized at right angles to that direction. Dichroic materials are different from birefringent materials because the latter usually have negligible absorption coefficients for both the ordinary and extraordinary rays. Stretched polyvinyl alcohol sheets treated with absorbing dyes or polymeric iodine are the most common type of dichroic absorbers and are primarily sold under the tradename Polaroid. These polarizers do not have as good an extinction ratio as the calcite prism polarizers (see above), but they are inexpensive, come in large sizes, are easily rotated, and produce negligible beam deviation. Also, they are thin, lightweight, and can be made in any desired shape. One of their main advantages is that they are insensitive to the degree of collimation of the beam and can be used in strongly convergent or divergent light.

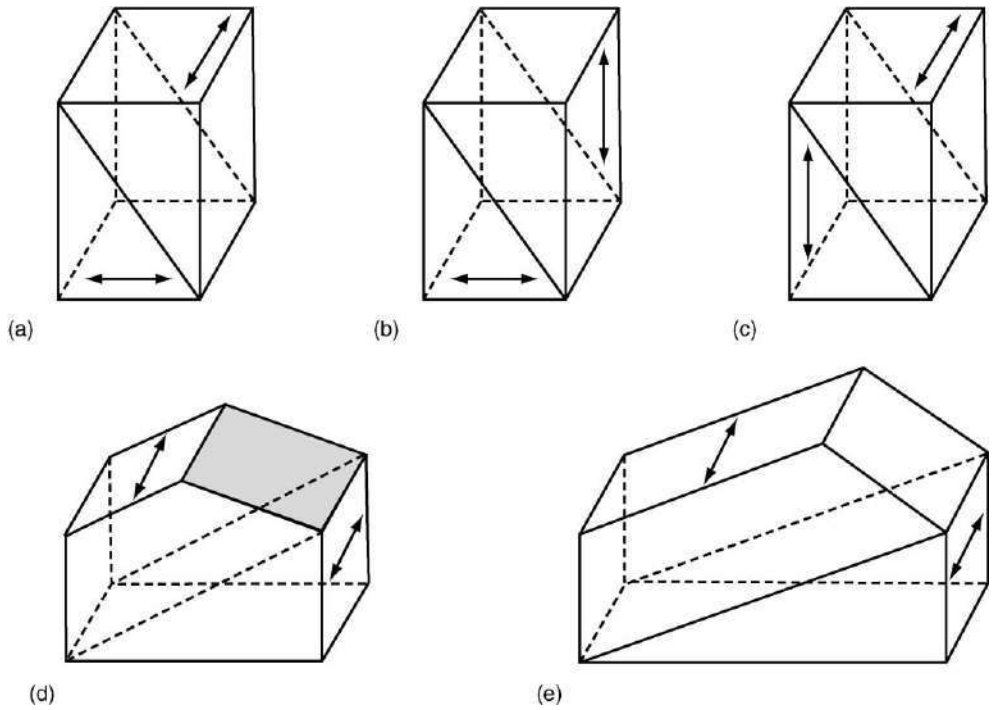

Figure 4 Three-dimensional views of various types of polarizing beamsplitter prisms: (a) Rochon; (b) Sénarmont; (c) Wollaston; (d) Foster (shaded face is silvered); and (e) beamsplitting Glan-Thompson. (Reproduced with permission from Bennett JM (1995) Polarizers. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 3. New York: McGraw-Hill, Inc.)

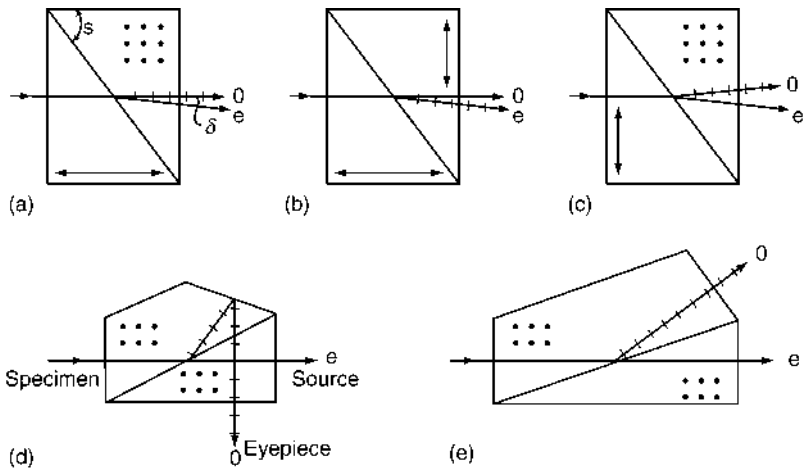

Figure 5 Side views of the polarizing beamsplitter prisms in **Figure 4**. The directions of the optic axes are indicated by the dots and the heavy double-pointed arrows. When the Foster prism is used as a microscope illuminator, the source, specimen, and eyepiece are in the positions indicated. (Reproduced with permission from Bennett JM (1995) Polarizers. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 3. New York: McGraw-Hill, Inc.)

Depending on the density and type of absorbing dye used to make the polarizer, the transmission can be maximized for the visible or near infrared spectrum. The extinction ratio of the Polaroid HN-22 sheet compares favorably with that of the Glan–Thompson prisms throughout the visible spectral region, but the transmission of the Glan–Thompson prism is superior. As the dichroic polarizer transmission increases, the extinction ratio becomes worse. Transmission and extinction ratio curves for various types of dichroic polarizers are shown in Bennett JM (1995) Polarizers. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 3. New York: McGraw-Hill, Inc.

A wire grid polarizer is another kind of dichroic absorber. It is made of a series of equally spaced conducting bars or wires that are either free standing or deposited onto a transparent substrate (backing plate). Energy that is polarized parallel to the length of the bars is absorbed out of the incoming wave by inducing oscillations in the electrons in the metal. Thus, only light polarized perpendicular to the bars will be transmitted. In order to have an appreciable degree of polarization, the wavelength must be at least twice the spacing between grids in the polarizer. Because of the difficulty of making wire grids with small enough spacings, these polarizers are limited to the mid- and long-wavelength infrared spectral region, beyond about 5 μm. The polarizer will have the best extinction ratio if the substrate has a low refractive index; if a high refractive index substrate such as silicon or germanium is used, it must be covered with an antireflection coating before the grid is deposited.

Most of the wire grid polarizers have been used with microwaves, so the theory is in the form for transmission lines. Similar transmission line theory has been applied to infrared polarizers in the form of bars and wires.

Reflection and Transmission

Light can be polarized by reflecting it from the flat surface of a material inclined at non-normal incidence to the light beam or by transmitting it through a transparent plate at non-normal incidence. These polarizers work because light has different reflectances when the electric vector is linearly polarized parallel and perpendicular to the plane of incidence. The plane of incidence contains both the incoming light beam (incident beam) and the reflected light beam. The angles for both the incident and reflected light beams are measured relative to an axis perpendicular to the surface (the surface normal). The reflection coefficients are given by the Fresnel equations and, for nonabsorbing materials, are:

$$R_s = r_s^2 = \frac{\vec{E}_s^2(\text{reflected})}{\vec{E}_s^2(\text{incident})} = \frac{\sin^2(\theta_0 - \theta_1)}{\sin^2(\theta_0 + \theta_1)} \qquad [9]$$

$$R_p = r_p^2 = \frac{\vec{E}_p^2(\text{reflected})}{\vec{E}_p^2(\text{incident})} = \frac{\tan^2(\theta_0 - \theta_1)}{\tan^2(\theta_0 + \theta_1)} \qquad [10]$$

where R_s and R_p are called the reflectances (previously called the intensity reflection coefficients), and r_s and r_p are the amplitude reflection coefficients. $\vec{E}_s$ ($\vec{E}_p$) is the incident or reflected electromagnetic

wave vibrating perpendicular (parallel) to the plane of incidence and θ_0 and θ_1 are the angles of incidence and refraction, respectively. The angle of refraction is the angle of the light beam in the material, measured from the surface normal. For a nonabsorbing material, it can be obtained in terms of the refractive index n_1 of the material from Snell's Law:

$$\frac{\sin\theta_0}{\sin\theta_1} = \frac{n_1}{n_0} \qquad [11]$$

where n_0 and n_1 are the refractive indexes of the incident medium (usually air with $n_0 = 1$) and the material. Sometimes the refractive index of the material is expressed as a ratio measured relative to the refractive index of the incident medium: $n = n_1/n_0$. The reflectances of absorbing materials are similar to eqns [9]–[11] but involve complex refractive indexes and angles of refraction.

Figure 6 shows curves for R_s and R_p as a function of angle of incidence for four nonabsorbing materials that have different refractive indexes. In all cases the reflectance is higher for the *s* component than for the *p* component except at 0° and 90° angles of incidence where they are the same. At the so-called Brewster angle, θ_B, $R_p = 0$, $\tan\theta_B = n_1/n_0$, and incident unpolarized light is now completely linearly polarized perpendicular to the plane of incidence (s-polarized light). Note that high refractive index materials produce more intense polarized beams (i.e., have higher reflectances) than low refractive index materials.

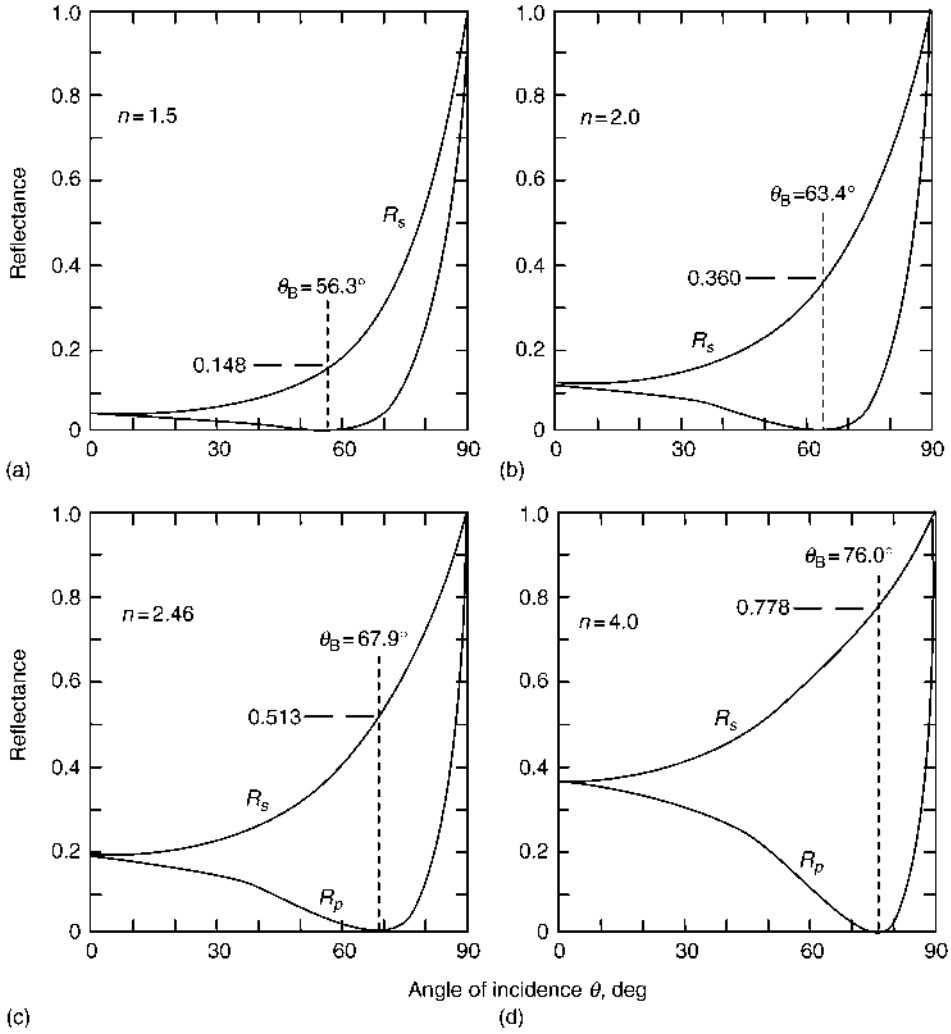

Figure 6 Reflectance of light polarized parallel R_p and perpendicular R_s to the plane of incidence from materials of different refractive index *n* as a function of angle of incidence: (a) $n = 1.5$ (alkali halides in the ultraviolet, glass (approximately) in the visible, and sheet plastics in the infrared), (b) $n = 2.0$ (AgCl in the infrared), (c) $n = 2.46$ (Se in the infrared), and (d) $n = 4.0$ (Ge in the infrared). The Brewster angle θ_B (at which R_p goes to 0) and the magnitude of R_s at θ_B are also indicated. (Reproduced with permission from Bennett JM (1995) Polarization. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. I, chap. 5. New York: McGraw-Hill, Inc.)

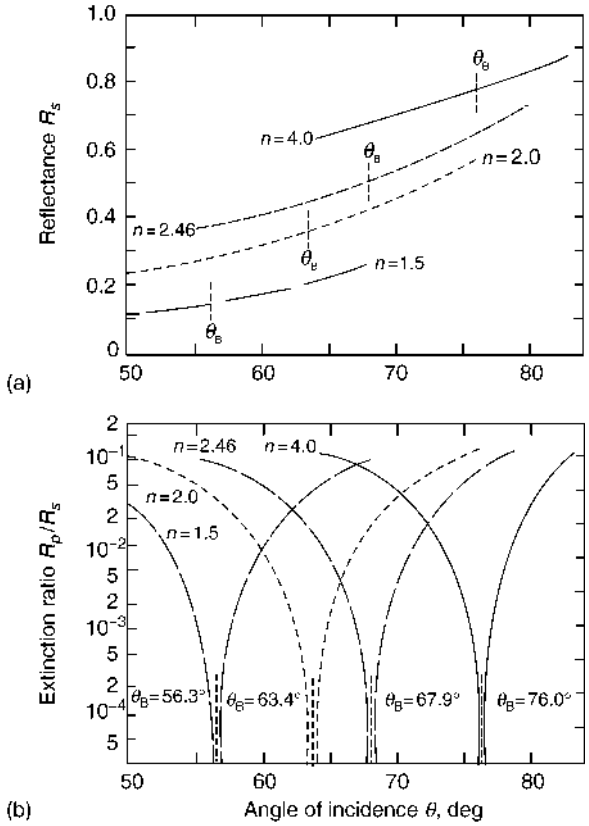

Figure 7 (a) Reflectance R_s and (b) extinction ratio R_p/R_s for materials of different refractive index at angles near the Brewster angle θ_B. A single surface of the material is assumed. (Reproduced with permission from Bennett JM (1995) Polarization. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. I, chap. 5. New York: McGraw-Hill, Inc.)

Figure 7 shows the reflectances and extinction ratios for materials having different refractive indexes. Each curve is for a single reflection. Since plates have two surfaces, there will be two reflections from an actual reflection polarizer so the reflectance and extinction ratios will increase. If the plates are thick, the reflection from the back surface of the plate will be displaced from the reflection from the front surface of the plate. A reflection polarizer is normally used close to the Brewster angle because a high degree of polarization (i.e., a very small extinction ratio) is desired. Because the extinction ratio changes rapidly for angles of incidence close to the Brewster angle, the incident beam must be well collimated to obtain a high degree of polarization. A main disadvantage of reflection polarizers is that the reflected beam is no longer parallel to the incident beam and two additional reflections are required to align the beam.

Light transmitted through a plate will only be partially linearly polarized at any angle of incidence including the Brewster angle because both s- and p-components are partially transmitted. Transmission polarizers are thus made of several plates to increase the degree of polarization. Figure 8 shows the transmittance and extinction ratio for a stack of four plane parallel plates assuming multiple incoherent reflections within each plate and no reflections between plates. At the Brewster angle the p-transmittance is theoretically 1 (assuming that there is no absorption within the material) but the extinction ratio greatly depends on the refractive index of the material. A stack of low refractive index plates ($n = 1.5$) has a poor extinction ratio, while a pile of high refractive index plates ($n = 4.0$) has a much better extinction ratio. The plates are often inclined at small angles to each other so the light multiply reflected between plates (which decreases the polarization) is removed from the transmitted beam. The sides of each plate are plane parallel to increase the polarization, but the plates are too thick for the amplitudes of the multiple internally reflected beams to add or subtract. If the amplitudes of the beams could be added, there would be interference effects and the transmittance would vary with the thickness of each plate and with the wavelength. These are so-called interference polarizers and are mentioned below.

Transmission polarizers do not have the angle deviation problem of the reflection polarizers, but the transmitted beam may be slightly displaced parallel to the incident beam if the plates are thick. The main problem with transmission polarizers is that the light is not completely polarized even with several plates in a stack. If the plates are slightly absorbing, the transmission of the polarizer is reduced.

Miscellaneous Types

There are numerous other types of polarizers that mostly use thin films. Interference effects in thin films can increase the polarization efficiency at certain wavelengths depending on the thicknesses of the films and the design of the multilayer coating. Many of the applications involve non-normal incidence designs. For example, a single high refractive index film or a multilayer coating evaporated onto a low refractive index substrate increases the degree of polarization of reflection and transmission polarizers. Polarizing beamsplitters also use multilayer dielectric coatings. Free-standing films of different thicknesses were formerly used for infrared polarizers before wire grid polarizers became available.

Dichroic absorbers have been made from two-phase lamelar eutectics of thin needles of a conducting material embedded in a transparent matrix. Thin sheets of pyrolytic graphite material also act

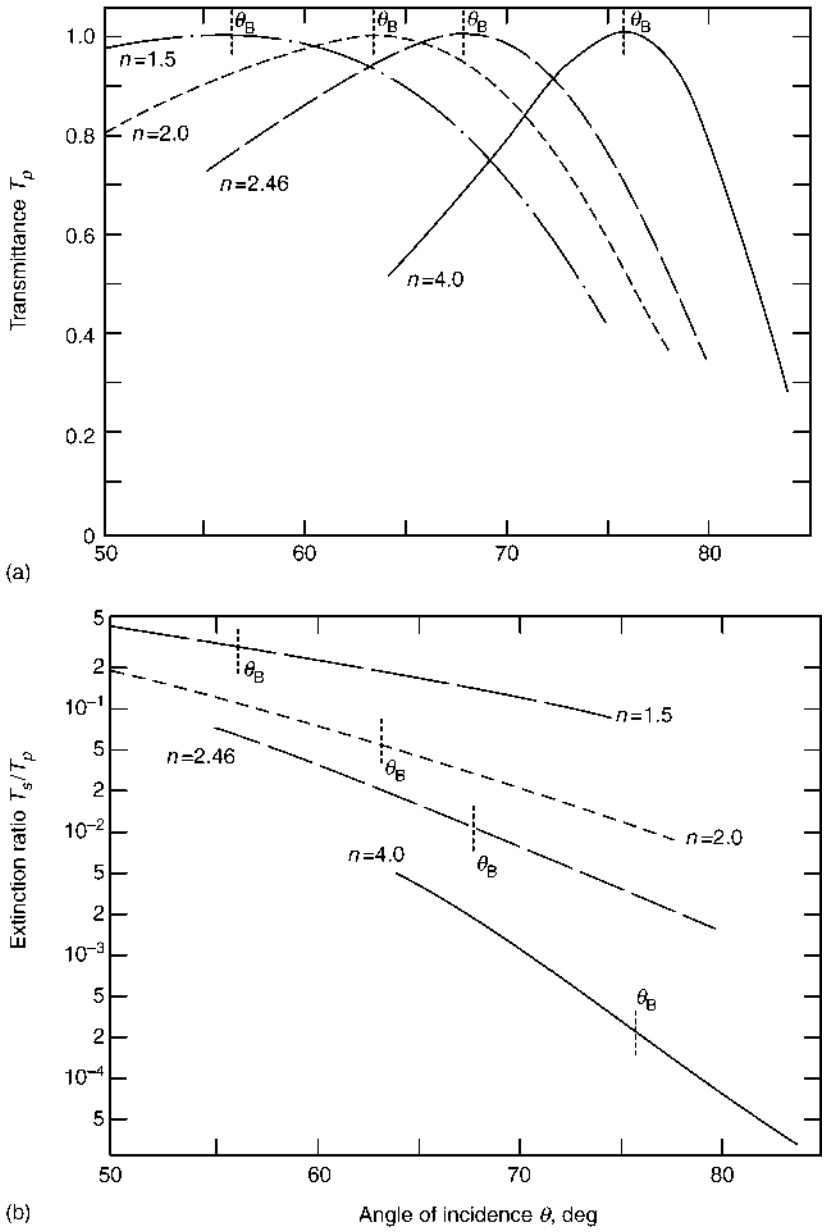

Figure 8 (a) Transmittance and (b) extinction ratio of four plane-parallel plates of refractive index n as a function of angle of incidence, for angles near the Brewster angle. Assumptions are multiple reflections but no interference within each plate and no reflections between plates. (Reproduced with permission from Bennett JM (1995) Polarization. In Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. I, chap. 5. New York: McGraw-Hill, Inc.)

as dichroic absorbers. There are also numerous miniature polarizer designs for fiber optics and nanotechnology applications.

Retarders or Retardation Plates

Retarders, or retardation plates, are devices that can change linearly polarized light into circularly or elliptically polarized light. They can also rotate the plane of polarization of linearly polarized light. A retardation plate is made from a uniaxial crystal that has an optic axis (as discussed above); the light travels in a direction perpendicular to the optic axis, as described below and shown in Figure 9. The ordinary and extraordinary rays travel at different velocities through the crystal depending on their refractive indexes: $n = c/v$, where c is the velocity of light in a vacuum and v is the velocity of light in the material. The larger the refractive index, the slower is the velocity of light in the material. The ordinary ray

with a refractive index n_o and the extraordinary ray with a refractive index n_e have mutually perpendicular vibration directions. If the ray has a vibration direction at another azimuth in the crystal, the refractive index is intermediate between n_o and n_e. For a positive uniaxial crystal $n_e > n_o$ so the extraordinary ray travels slower than the ordinary ray through the crystal. Thus, the designation 'fast axis' is often used for the ordinary ray and 'slow axis' is used for the extraordinary ray, as shown in Figure 9. In a negative uniaxial crystal, $n_e < n_o$, i.e., the velocities along the two axes are reversed.

Because there is a velocity difference between the ordinary and extraordinary rays traveling through a uniaxial crystal, they get out of phase. Depending on their velocities and the optical thickness of the material (the refractive index times the physical thickness), the addition of their amplitudes results in a wave whose vibration direction is either: (a) perpendicular to the original vibration direction; (b) rotates in a circle (right or left circularly polarized light); or (c) rotates in an ellipse (right or left elliptically polarized light). The path difference N between two rays in the crystal, measured in terms of the wavelength of the light, is $N\lambda = \pm d(n_e - n_o)$, where d is the physical thickness of the material. The corresponding phase difference δ between the two rays is $\delta = 2\pi N = \pm(2\pi d/\lambda)(n_e - n_o)$. The quantity N is important because if beams of light vibrating along the fast and slow axes of a crystal get out of step by one quarter of the wavelength of the light, the device is called a quarter wave (or $\lambda/4$) retardation plate, or simply a quarter-wave plate. There are also half-wave ($\lambda/2$) plates that rotate the plane of polarization, and full-wave (λ) plates that rotate the plane of polarization through 180°. If light passes through a full-wave plate, theoretically it is indistinguishable from the original beam. However, materials are normally temperature sensitive so that as the temperature changes, the retardation is no longer exactly one wave, but may be slightly less than or greater than one wavelength. Other thicknesses of retardation plates produce elliptically polarized light.

Figure 10 shows what happens to a beam of linearly polarized light when the axis of vibration is at 45° to the fast and slow axes of a positive

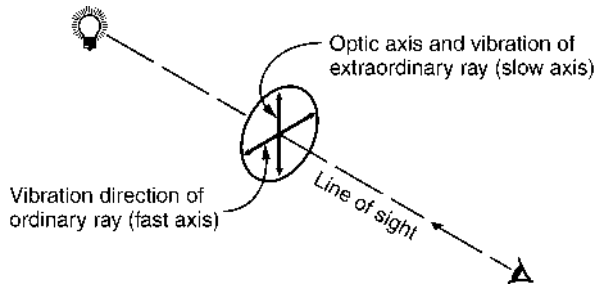

Figure 9 Light incident normally on the front surface of a retardation plate showing the vibration directions of the ordinary and extraordinary rays. In a positive uniaxial crystal, the fast and slow axes are as indicated in parentheses; in a negative uniaxial crystal, the two axes are interchanged. (Adapted with permission from Bennett JM (1995) Polarization. In Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. I, chap. 5. New York: McGraw-Hill, Inc.)

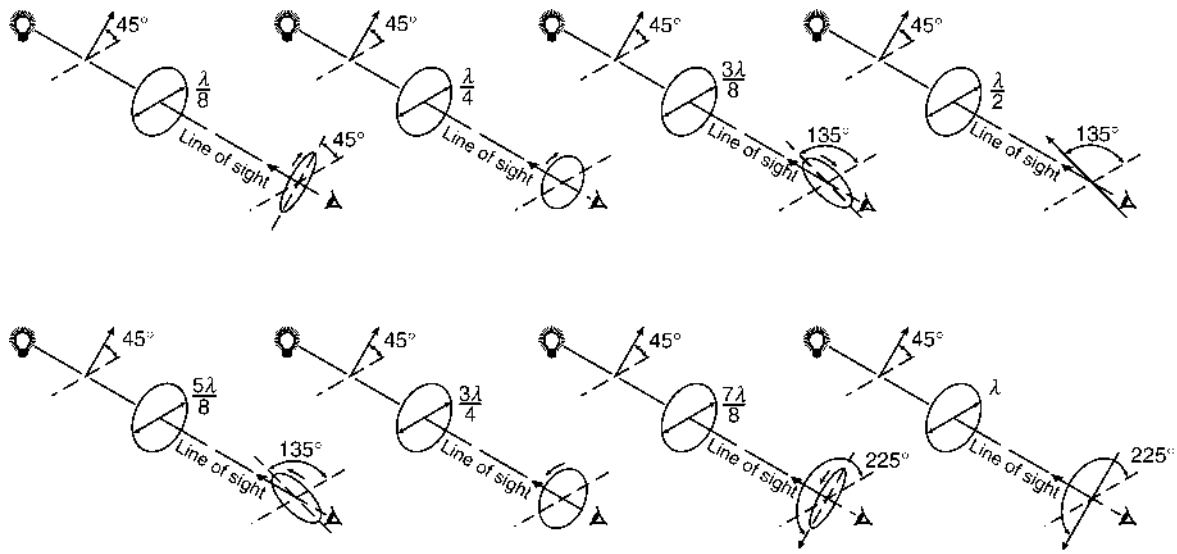

Figure 10 State of polarization of a light wave after passing through a crystal plate whose retardation is indicated in fractions of a wavelength (phase retardation $2\pi/\lambda$ times these values) and whose fast axis is indicated by the double arrow. In all cases the incident light is linearly polarized at an azimuth of 45° to the direction of the fast axis. (Adapted with permission from Bennett JM (1995) Polarization. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. I, chap. 5. New York: McGraw-Hill, Inc.)

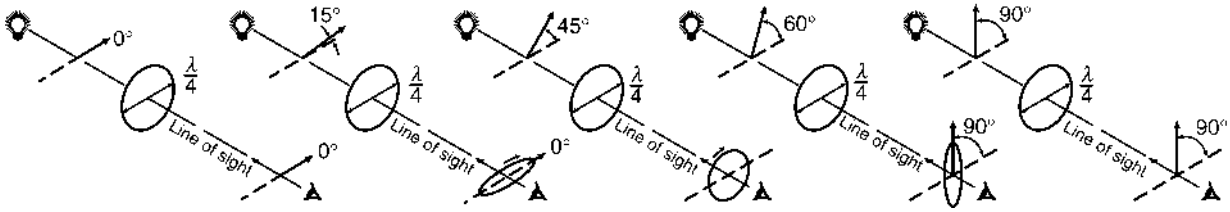

Figure 11 State of polarization of a light wave after passing through a $\lambda/4$ plate (whose fast axis is indicated by the double arrow) for different azimuths of the incident linearly polarized beam. (Adapted with permission from Bennett JM (1995) Polarization. In Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. I, chap. 5. New York: McGraw-Hill, Inc.)

uniaxial crystal. The fast axis is in the horizontal direction, the same as in Figure 9.

The main purpose of a quarter-wave plate is to change linearly polarized light into circularly polarized light. However, as Figure 11 shows, the state of polarization of the light will be quite different depending on the orientation of the plane of polarization of the incident beam relative to the fast axis of the quarter-wave plate.

A half-wave plate is used to rotate the plane of polarization of a linearly polarized beam. The plane of polarization is always rotated through an angle that is twice the angle the initial plane of polarization makes with the fast axis of the uniaxial crystal. A linearly polarized beam always remains linearly polarized.

Retardation plates are often made of mica, stretched polyvinyl alcohol, or crystal quartz, although they can also be made of other stretched plastics, sapphire, magnesium fluoride, and other materials.

Variable Retardation Plates and Compensators

Variable retardation plates are devices whose retardation can be varied in a variety of ways. They can be used to modulate or vary the phase of a beam of linearly polarized light or to analyze a beam of unknown polarization (often elliptically polarized light) such as might be produced by transmission through a birefringent material or by reflection from a metal or film-covered surface. The term compensator is often applied to a variable retardation plate since it can be used to compensate for the phase retardation produced by a material. Common types of compensators include the Babinet and Soleil compensators, in which the total thickness of a birefringent material in the light path is changed, the Sénarmont compensator which consists of a fixed quarter-wave plate and rotatable analyzer to compensate for varying amounts of ellipticity in a light beam, and tilting-plate compensators, which change the thickness of birefringent material in the light beam by changing the angle of incidence. Electro-optic and piezo-optic modulators can also be used as high-frequency variable retardation plates since their birefringence can be changed by varying the electric field or pressure (see next section).

The Babinet compensator, shown schematically in Figure 12 consists of two crystal quartz wedges, each with its axis in the plane of the face but with the axes at right angles to each other. One wedge is stationary, and the other can be moved in the direction indicated by the arrow, so that the total amount of quartz through which the light passes can be varied. The total retardation is proportional to the difference in thickness between the two wedges. This type of compensator was used extensively when the light source was a vertical slit and visual measurements were made of the state of polarization of a beam. However the bands representing different phase retardations were too narrow to be used effectively with a photoelectric detector, so the Babinet compensator was replaced by a Soleil compensator (Figure 13). This device, sometimes called a Babinet–Soleil compensator, is similar to the Babinet compensator except that the field of view has a uniform tint if the compensator is constructed correctly. This is because the ratio of the thicknesses of the two crystal quartz blocks (a movable wedge and a fixed wedge attached to a plate with the two axes perpendicular to each other) is the same over the entire field. The Soleil compensator will produce light of varying ellipticity depending on the position of the movable wedge. It is used in the same way as a Babinet compensator with the uniformly dark field of the Soleil compensator corresponding to the black zero-retardation band in the Babinet compensator.

It is sometimes necessary to accurately measure the azimuth of a beam of linearly polarized light using a photoelectric detector with a rotatable analyzer (i.e., a polarizer) directly in front it. The obvious method is to rotate the analyzer until the detector signal is a minimum and then read the analyzer angle, which equals the extinction angle (perpendicular to the azimuth of the linearly polarized beam). However, the

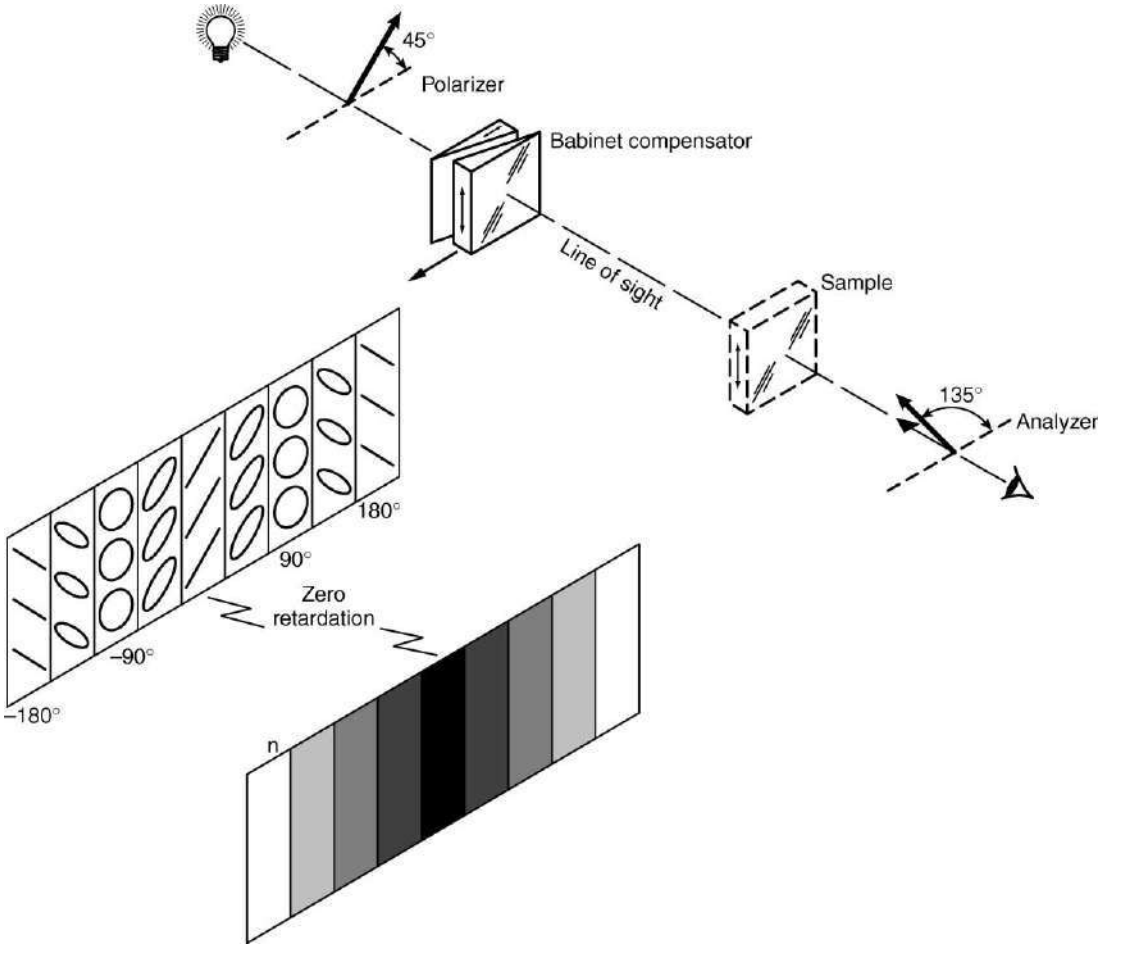

Figure 12 Arrangement of a Babinet compensator, polarizer, and analyzer for measuring the retardation of a sample. The appearance of the field after the light has passed through the compensator is shown to the left of the sample position. Retardations are indicated for alternate regions. After the beam passes through the analyzer, the field is crossed by a series of dark bands, one of which is shown to the left of the analyzer. (Adapted with permission from Bennett JM (1995) Polarizers. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 3. New York: McGraw-Hill, Inc.)

angle can be determined more precisely if the analyzer is offset by a small angle from the extinction angle and the transmittance noted. Then the analyzer is offset by a small angle on the other side of the extinction angle at the angle where the transmittance is the same. One half the difference between these two azimuthal angles gives a more accurate value of the extinction angle than can be obtained by measuring it directly.

Before the days of photoelectric detectors, half-shade devices were extensively used to determine the azimuth of a linearly polarized beam. The device consisted of two polarizers with their axes inclined at a small angle to each other. As the device was rotated, one part of the field became darker and the other part became lighter. At the match position, both parts of the field appeared equally bright. There were a variety of these devices as well as ellipticity half-shade devices that could detect very small amounts of ellipticity in a nominally linearly polarized beam and hence could verify when a compensator had completely converted elliptically polarized light into linearly polarized light.

Electro-Optic, Magneto-Optic, and Piezo-Optic Devices

The state of polarization of light can be rapidly altered by passing it through a material that has electro-optic, magneto-optic or piezo-optic properties. The voltage, magnetic field, or pressure are varied to make the material birefringent (see above). Some materials are isotropic without the applied field, i.e., the refractive index is the same in all directions. Other materials have an optic axis

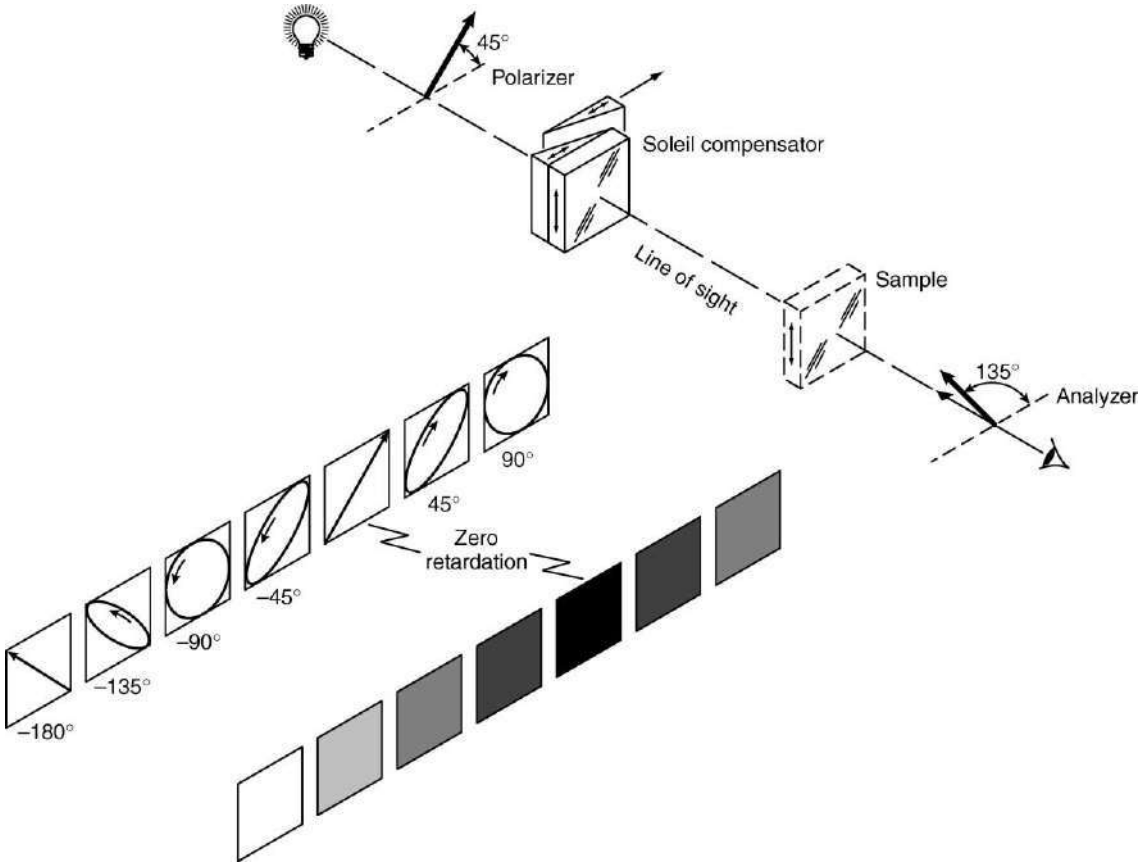

Figure 13 Arrangement of a Soleil compensator, polarizer, and analyzer for measuring the retardation of a sample. The appearance of the field after the light has passed through the compensator is shown to the left of the sample position. After the beam passes through the analyzer, the field appears as one of the shades of gray shown to the left of the analyzer. (Adapted with permission from Bennett JM (1995) Polarizers. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 3. New York: McGraw-Hill, Inc.)

(uniaxial materials) or two optic axes (biaxial materials). In these cases, the applied field creates further asymmetries in the material.

The most common electro-optic, magneto-optic and piezo-optic effects are the Pockel's effect (depending on the electric field), the Kerr effect (depending on the square of the electric field), the Faraday effect (depending on the magnetic field), the Cotton–Mouton and Voigt effects (depending on the square of the magnetic field), and stress birefringence or the photoelastic effect (depending on pressure changes). These effects are shown in Table 1 along with order of magnitude strengths needed to produce changes in the refractive index, and materials that most strongly exhibit the effects. The mathematical descriptions of the effects involve tensors and are too involved to present here. However, simple physical descriptions will be given.

If a varying electric field acts on an electro-optic material, electrons, ions, or permanent dipoles in the material are made to reorient, inducing an electric polarization. The induced polarization creates birefringence that modifies the optical polarization of a light beam passing through the material. The electric field strength determines how the polarization is changed (see Figure 10).

Linearly polarized light passing through a magneto-optic material will be rotated in a direction parallel to the direction of the applied magnetic field. This phenomenon, called the Faraday effect, or Faraday rotation, is similar to what happens in optically active materials (see the next paragraph) except that the Faraday effect depends on the direction of the magnetic field, but not on the direction the light is traveling.

There are naturally optically active materials such as camphor, nicotine, sugar solutions, and crystal quartz that can rotate the plane of linearly polarized light passing through the material. For example, the Si–O bonds in crystal quartz form a helical path around the optic axis. This crystal structure interacts with an incoming linearly polarized beam traveling parallel to the optic axis and rotates it in a clockwise direction. A 1 mm-thick piece of quartz will rotate a

Table 1 Electro-optic, magneto-optic, and piezo-optic effects

Effect	*Device*	*Effect proportional to*	*Strength* (Δn)	*Materials*
Pockels (electro-optic)	Pockels cell	E	$10^{-12}E$	$BaTiO_2$, $LiNBO_3$
Kerr	Kerr cell	E^2	$10^{-19}E^2$	Nitrobenzene, Benzonitrile
Faraday (magneto-optic)	Faraday rotation	H	$10^{-14}H$	Zns, GaAs, CS_2
Cotton–Mounton	[a]	H^2	$10^{-25}H^2$	Chloroform, acetone
Photoelastic effect; stress birefringence	[b]	Pressure	–	Fused silica, polystyrene, Ge, KDP, ruby

[a]Too small to be of technological importance.
[b]Depends on the mounting or the processing of the material.
Adapted with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley and Sons.

linearly polarized beam by 21.7°. Other materials have asymmetric structures that have mirror images. A mirror image structure cannot be obtained by simply rotating the group of atoms in space. These materials are also optically active. The amount of rotation produced by an optically active material is proportional to the thickness (optical density) of the material that the light passes through. Dextrose (a form of glucose) and levulose (also known as fructose) rotate the plane of linearly polarized light in clockwise and counterclockwise directions, respectively. Many other organic molecules have D- (right handed, dextro-rotary) and L- (left handed, levo-rotary) mirror image forms.

An isotropic material can become anisotropic when stress or an induced strain is applied because of an elasto-optic interaction known as stress birefringence, or the photoelastic effect. This effect is usually bad because it reduces the performance of optical components and devices by introducing phase distortions caused by improper mounting or unequal thermal expansion between the mounts and components. Another source of distortion is strain that has been frozen into optical components during manufacture. French curves are excellent examples that show colored strain birefringence patterns when viewed between crossed polarizers. In one application, strain birefringence has been used constructively to produce a variable phase retarder made from crystal quartz and fused silica. A block of electro-optic crystal quartz, is cemented to a block of isotropic fused silica. When a variable electric field is applied to the crystal quartz, its length changes at the resonant frequency of the block. This produces strain in the fused silica block which in turn produces a variable retardation of light passing through the strained fused silica. Depending on the magnitude of the varying electric field, the device can act like a variable quarter-wave plate, a variable half-wave plate, or have other applications.

Matrix Methods for Computing Polarization

An optical system containing various polarizing and retarding devices can be modeled using matrices. In general, there is an incident beam (in matrix form), that has some state of polarization. It interacts with a device called an instrument (also in matrix form) that alters the state of polarization, so that the exiting beam (a third matrix, the product of the first and second matrices) has another state of polarization. There can be more than one instrument matrix acting on the same incident matrix to produce the final matrix. The two most common matrix methods are the Mueller calculus and the Jones calculus.

There is also a visual representation of this process, a Poincaré sphere, on which vectors represent different states of polarization. The polarization instrument moves the vector representing the incident polarization around the sphere in a prescribed manner. The vectors are called Stokes vectors and are used in the Mueller calculus. The various states of polarization are represented on the Poincaré sphere as follows: The equator represents various forms of linear polarization, the poles represent right- and left-circular polarization, and other points on the sphere represent elliptically polarized light. Every point on the sphere corresponds to a different polarization form. The radius of the sphere represents the incident irradiance of the light beam (which is usually assumed to be unity). The effects of various polarization instruments are determined by displacements on the sphere. Partially polarized light or absorption may be dealt with approximately by ignoring the irradiance factor, since the state of polarization is generally the quantity desired. The Poincaré sphere is most useful for visualizing problems involving nonabsorbing materials, various polarization instruments including polarizers, retarders, compensators, half-shade devices, and depolarizers, and has also been applied

to ellipsometric problems and stress-optical measurements.

The Poincaré sphere is used with Stokes vectors, which are often designated S_0, S_1, S_2, and S_3. S_0 is the incident irradiance of the light beam, corresponding to the radius of the Poincaré sphere. S_1 is the difference in irradiances between the horizontal and vertical polarization components of the beam; for example, when S_1 is positive, the preference is for horizontal polarization. S_2 indicates preference for $+45°$ or $-45°$ polarization depending on whether it is positive or negative, and S_3 gives the preference for right or left circular polarization. Thus, the Stokes vectors S_1, S_2, and S_3 are simply the three Cartesian coordinates of a point on the Poincaré sphere. S_1 and S_2 are perpendicular to each other in the equatorial plane, and S_3 points toward the north pole of the sphere. Any state of polarization of a light beam can be specified by these three vectors. The irradiance vector S_0 is related to the other three by the relation $S_0^2 = S_1^2 + S_2^2 + S_3^2$ when the beam is completely polarized. If the beam is partially polarized, $S_0^2 > S_1^2 + S_2^2 + S_3^2$.

In the Mueller calculus, the incident beam is represented by the four-component Stokes vector, written as a column vector. This vector has all real elements and gives information about irradiance properties of the beam. Thus it is not able to handle problems involving phase changes or combinations of two beams that are coherent. The instrument matrix is a 4×4 matrix with all real elements.

In the Jones calculus, the Jones vector representing the incident beam is a two-component column vector that generally has complex elements. It contains information about the *amplitude* properties of the beam and hence is well suited for handling problems involving the phases of light waves. However, it cannot handle problems involving depolarization, as the Mueller calculus can. The Jones instrument matrix is a 2×2 matrix whose elements are generally complex. The Jones calculus is well suited to problems involving a large number of similar devices arranged in series in a regular manner and makes it possible to obtain a result expressed explicitly in terms of the number of such devices. The Jones instrument matrix of a train of transparent or absorbing nondepolarizing polarizers and retarders contains no redundant information. The matrix contains four elements, each of which has two parts, so that there are a total of eight constants, none of which is a function of any other. The Mueller instrument matrix of such a train contains much redundancy; there are 16 constants but only 7 of them are independent. More information about the Poincaré sphere and the Mueller and Jones calculus is available in other references.

See also

Electromagnetically Induced Transparency. **Magneto-optics:** Faraday Rotation, CARS, ODMR, ODSR, Optical Pumping; Interband Magnetoabsorption, Cyclotron Resonance, Spin Flip Raman Scattering.

Further Reading

Auton JP (1967) Infrared transmission polarizers by photolithography. *Applied Optics* 6: 1023–1027.

Azzam, Rasheed MA and Bashara NM (1977) *Ellipsometry and Polarized Light*, chaps. 1 and 2. Amsterdam: North-Holland Publishing Co.

Azzam, Rasheed MA (1995) Ellipsometry. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 27. New York: McGraw-Hill.

Bennett JM (1995) Polarization. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. I, chap. 5. New York: McGraw-Hill.

Bennett M (1995) Polarizers. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 3. New York: McGraw-Hill.

Bohren CF and Huffman DR (1983) *Absorption and Scattering of Light by Small Particles*. New York: John Wiley.

Born M and Wolf E (1999) *Principles of Optics*, 7th edn, chap. 1. Cambridge: Cambridge University Press.

Chipman RA (1995) Polarimetry. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 22. New York: McGraw-Hill.

Clark JR (1956) New calculus for the treatment of optical systems. VIII. Electromagnetic theory. *Journal of the Optical Society of America* 46: 126–131.

Guenther RD (1990) *Modern Optics*. chaps. 2, 13, and 14. New York: John Wiley and Sons.

Hass M and O'Hara M (1965) Sheet infrared transmission polarizers. *Applied Optics* 4: 1027–1031.

Hecht E and Zajac A (1997) *Optics*, 3rd edn, chap. 8. Reading, MA: Addison-Wesley Publishing Co.

Jackson JD (1999) *Classical Electrodynamics*, 3rd edn, chap. 7. New York: John Wiley and Sons, Inc.

Jasperson SN and Schnatterly SE (1969) An improved method for high reflectivity ellipsometry based on a new polarization modulation technique. *Reviews of Scientific Instruments* 40: 761–767.

Kerker M (1969) *The Scattering of Light and Other Electromagnetic Radiation*. New York: Academic Press.

Kliger DS, Lewis JW and Randall CE (1990) *Polarized Light in Optics and Spectroscopy*, chaps. 1–5. San Diego: Academic Press.

Maldonado TA (1995) Electro-optic modulators. In: Bass M (ed.) *Handbook of Optics*, 2nd edn, vol. II, chap. 13. New York: McGraw-Hill, Inc.

Markuvitz N (ed.) (1951) *Waveguide Handbook*, pp. 218–229, 285–289. New York: McGraw-Hill Book Co.

Shurcliff WA (1962) *Polarized Light, Production and Use*. Cambridge, MA: Harvard University Press.

van de Hulst HC (1957) *Light Scattering by Small Particles*. New York: John Wiley.

Matrix Analysis

B D Guenther, Duke University, Durham, NC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Hooke postulated, in the seventeenth century, that light waves must be transverse but his idea was forgotten. Young and Fresnel put forward the same idea in the nineteenth century and accompanied their postulation with a theoretical description of light based on transverse waves. Forty years later, Maxwell proved that light must be a transverse wave and that **E** and **H**, for a plane wave in an isotropic medium with no free charge and no currents, are mutually perpendicular and lie in a plane normal to the direction of propagation, **k**.

Convention requires that we use the electric vector to label the direction of the electromagnetic wave's polarization. The direction of the displacement vector is called the direction of polarization and the plane containing the direction of polarization and the propagation vector is called the plane of polarization. The selection of the electric field is not completely arbitrary, except in relativistic situations, when $v \approx c$, the interaction of the electromagnetic wave with matter will be dominated by the electric field.

Assume that a plane wave is propagating in the z-direction. In complex notation, the plane wave is given by

$$\tilde{\mathbf{E}} = \mathbf{E}_0\, e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi)} = \mathbf{E}_0\, e^{i(\omega t - kz + \phi)} \qquad [1]$$

The procedure used to decompose an arbitrary polarization into components parallel to two axes of a Cartesian coordinate system, is a technique used extensively in vector algebra to simplify mathematical calculations. According to the mathematical formalism associated with this technique, the polarization is described in terms of a set of basis vectors, $\mathbf{e}_i$. An arbitrary polarization would be expressed as

$$\mathbf{E} = \sum_{i=1}^{2} a_i \mathbf{e}_i \qquad [2]$$

The set of basis vectors, $\mathbf{e}_i$, are orthonormal, i.e.:

$$\mathbf{e}_i \mathbf{e}_j^* = \delta_{ij} = \left\{ \begin{matrix} 1 & i = j \\ 0 & i \neq j \end{matrix} \right\} \qquad [3]$$

where we have assumed that the basis vectors could be complex. We mention this mathematical formalism because an identical formalism is encountered in a description of spin.

In a Cartesian coordinate system, the $\mathbf{e}_i$'s are the unit vectors **i**, **j**, **k**. The summation in eqn [2] extends over only two terms because the electromagnetic wave is transverse, confining **E** to a plane normal to the direction of propagation (according to the coordinate convention we have selected, the **E** field is in the x, y plane).

Polarization Ellipse

Following the formalism of eqn [2], a polarized wave can be written in terms of the x and y components of $\mathbf{E}_0$

$$\tilde{\mathbf{E}} = \mathbf{E}_{0x}\, e^{i(\omega t - kz + \phi_1)}\hat{\mathbf{i}} + \mathbf{E}_{0y}\, e^{i(\omega t - kz + \phi_2)}\hat{\mathbf{j}} \qquad [4]$$

We will use only the real part of **E** for manipulation, to prevent errors. We divide each component of the electric field by its maximum value, so that the problem is reduced to one of the following two sinusoidal varying unit vectors:

$$\begin{aligned} \frac{E_x}{E_{0x}} &= \cos(\omega t - kz + \phi_1) \\ &= \cos(\omega t - kz)\cos\phi_1 - \sin(\omega t - kz)\sin\phi_1 \\ \frac{E_y}{E_{0y}} &= \cos(\omega t - kz)\cos\phi_2 - \sin(\omega t - kz)\sin\phi_2 \end{aligned} \qquad [5]$$

When these unit vectors are added together, the resulting equation will describe the path taken by the tip of the resultant vector. The path will create a Lissajous figure.

To obtain the equation for the Lissajous figure, we eliminate the dependence of the unit vectors on $(\omega t - kz)$. First, multiply the equations by $\sin\phi_2$ and $\sin\phi_1$, respectively and then subtract the resulting equations. Second, multiply the two equations by $\cos\phi_2$ and $\cos\phi_1$, respectively and then subtract the new equations. These two operations yield the following pair of equations:

$$\begin{aligned} &\frac{E_x}{E_{0x}}\sin\phi_2 - \frac{E_y}{E_{0y}}\sin\phi_1 \\ &\quad = \cos(\omega t - kz)[\cos\phi_1 \sin\phi_2 - \sin\phi_1 \cos\phi_2] \end{aligned} \qquad [6]$$

$$\frac{E_x}{E_{0x}}\cos\phi_2 - \frac{E_y}{E_{0y}}\cos\phi_1$$
$$= \sin(\omega t - kz)[\cos\phi_1 \sin\phi_2 - \sin\phi_1 \cos\phi_2] \quad [7]$$

The term in brackets can be simplified using the trig identity

$$\sin\delta = \sin(\phi_2 - \phi_1)$$
$$= \cos\phi_1 \sin\phi_2 - \sin\phi_1 \cos\phi_2 \quad [8]$$

After replacing the term in brackets by $\sin\delta$, the two equations are squared and added yielding the equation for the Lissajous figure:

$$\left(\frac{E_x}{E_{0x}}\right)^2 + \left(\frac{E_y}{E_{0y}}\right)^2 - \left(\frac{2E_xE_y}{E_{0x}E_{0y}}\right)\cos\delta = \sin^2\delta \quad [9]$$

The trig identity

$$\cos\delta = \cos(\phi_2 - \phi_1) = \cos\phi_1 \cos\phi_2 + \sin\phi_1 \sin\phi_2$$

was also used to further simplify eqn [9].

Equation [9] has the same form as the equation of a conic

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

Geometry defines the conic as an ellipse because, from eqn [9]

$$B^2 - 4AC = \frac{4}{E_{0x}^2E_{0y}^2}(\cos^2\delta - 1) < 0$$

This ellipse is called the polarization ellipse. The orientation of the ellipse, with respect to the x-axis, is

$$\tan 2\theta = \frac{B}{A - C} = \frac{2E_{0x}E_{0y}\cos\delta}{E_{0x}^2 - E_{0y}^2} \quad [10]$$

The tip of the resultant electric field vector obtained from eqn [4] traces out the polarization ellipse in the plane normal to **k**, as predicted by eqn [9]. The ratio of the length of the minor to the major axis of the ellipse is equal to the ellipticity, φ, i.e., the amount of deviation of the ellipse from a circle

$$\tan\varphi = \pm\left(\frac{E_m}{E_M}\right)$$
$$= \frac{E_{0x}\sin\phi_1 \sin\theta - E_{0y}\sin\phi_2 \cos\theta}{E_{0x}\cos\phi_1 \cos\theta + E_{0y}\cos\phi_2 \sin\theta} \quad [11]$$

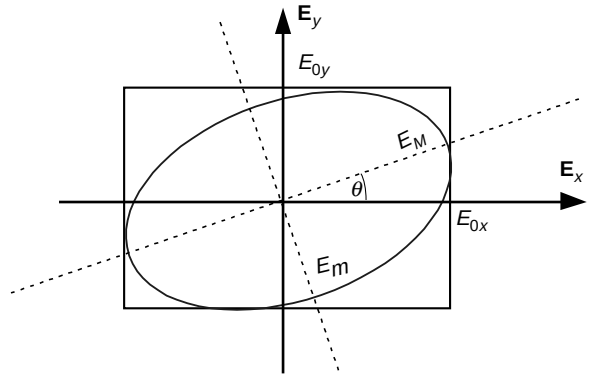

Figure 1 General form of the ellipse described by eqn [9]. Reproduced with permission from Guenther RD (1990) *Modern Optics*. New York: John Wiley and Sons.

In particle physics, the light would be said to have a negative helicity if it rotated in a clockwise direction. If we look at the source, the electric vector seems to follow the threads of a left-handed screw, agreeing with the nomenclature that left-handed quantities are negative. However, in optics the light that rotates clockwise, as we view it traveling toward us from the source, is said to be right-circularly polarized. The counterclockwise rotating light is left-circularly polarized. The association of right-circularly polarized light, with 'right-handedness' in optics, came about by looking at the path of the electric vector in space at a fixed time, then $\tan\psi = \tan(\phi - kz)$ (Figure 1).

Stokes Parameters

In the formalism associated with eqn [2], the expansion coefficients, a_i, can be used to form a 2–2 matrix which, in statistical mechanics, is called the density matrix and in optics, the coherency matrix (Table 1). The elements of the matrix are formed by the rule

$$\rho_{ij} = \mathbf{a}_i\mathbf{a}_j^* \quad [12]$$

The matrix is Hermitian, so that $\rho_{ij} = \rho_{ji}^*$. We will not develop the theory of polarization using the coherency matrix, but simply use the coherency matrix to justify the need for four independent measurements to characterize polarization. There is no unique set of measurements required by the theory, but normally measurements made are of the Stoke's parameters which are directly related to the parameters of the polarization ellipse of eqn [9].

The Stokes parameters of a light wave are measurable quantities, defined as:

$s_0 \Rightarrow$ Total flux density
$s_1 \Rightarrow$ Difference between flux density transmitted by a

Table 1 Typical Stokes vectors

Horizontal polarization	$\begin{bmatrix}1\\1\\0\\0\end{bmatrix}$	Vertical polarization	$\begin{bmatrix}1\\-1\\0\\0\end{bmatrix}$
+45° polarization	$\begin{bmatrix}1\\0\\1\\0\end{bmatrix}$	−45° polarization	$\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}$
Right circular polarization	$\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$	Left circular polarization	$\begin{bmatrix}1\\0\\0\\-1\end{bmatrix}$

linear polarizer oriented parallel to the x-axis and one oriented parallel to the y-axis. The x and y-axes are usually selected to be parallel to the horizontal and vertical directions in the laboratory.

$s_2 \Rightarrow$ Difference between flux density transmitted by a linear polarizer oriented at 45° to the x-axis and one oriented at 135°.

$s_3 \Rightarrow$ Difference between flux density transmitted by a right-circular polarizer and a left-circular polarizer.

If the Stokes parameters are to characterize the polarization of a wave, they must be related to the parameters of the polarization ellipse. It is therefore important to establish that the Stokes parameters are variables of the polarization ellipse (eqn [9]).

In its current form, eqn [9] contains no measurable quantities and thus must be modified if it is to be associated with the Stokes parameters. The time average of the Poynting vector is the quantity observed when measurements are made of light waves. We must, therefore, find the time average of eqn [9] if we wish to relate its parameters to observable quantities.

The time average eqn [9] can now be written as

$$\frac{\langle E_x^2\rangle}{E_{0x}^2} + \frac{\langle E_y^2\rangle}{E_{0y}^2} - 2\frac{\langle E_x E_y\rangle}{E_{0x}E_{0y}}\cos\delta = \sin^2\delta \qquad [13]$$

where time average is denoted by $\langle\,\rangle$

$$\langle E_x^2\rangle = \frac{1}{T}\int_{t_0}^{t_0+T} E_{0x}^2[\cos(\omega t - kz)\cos\phi_1 - \sin(\omega t - kz)\sin\phi_1]^2 \mathrm{d}t \qquad [14]$$

With the time averages, eqn [13] can be written as

$$(E_{0x}^2 + E_{0y}^2)^2 - (E_{0x}^2 - E_{0y}^2)^2 - (2E_{0x}E_{0y}\cos\delta)^2 = (2E_{0x}E_{0y}\sin\delta)^2 \qquad [15]$$

Each term in this equation can be identified with a Stokes parameter:

$$\begin{aligned} s_0 &= \langle E_{0x}^2\rangle + \langle E_{0y}^2\rangle & s_1 &= \langle E_{0x}^2\rangle - \langle E_{0y}^2\rangle \\ s_2 &= \langle 2E_{0x}E_{0y}\cos\delta\rangle & s_3 &= \langle 2E_{0x}E_{0y}\sin\delta\rangle \end{aligned} \qquad [16]$$

Equation [15] can now be written as

$$s_0^2 - s_1^2 - s_2^2 = s_3^2 \qquad [17]$$

For a polarized wave, only three of the Stokes parameters are independent. This agrees with the requirement placed upon elements of the Hermitian, coherency matrix, introduced above.

With this demonstration of the connection between the Stokes parameters and the polarization ellipse, the Stokes parameters can be written in terms of the parameters of the polarization ellipse.

$$\begin{aligned} s_1 &= s_0\cos 2\varphi\cos 2\theta \\ s_2 &= s_0\cos 2\varphi\sin 2\theta \\ s_3 &= s_0\sin 2\varphi \end{aligned} \qquad [18]$$

Mueller Calculus

Mueller pointed out that the Stokes parameters can be thought of as elements of a column matrix or a 4-vector (Table 2)

$$\mathcal{S} = \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} \qquad [19]$$

Each element of the Stokes vector represents a measurable intensity. The vector can represent not only polarized 1ight, but unpolarized or partially polarized light.

In 1929, Soleillet discovered that an optical device performed a linear transformation on the input wave and, in 1942, Perrin put this fact into a matrix formalism involving the Stokes vectors:

$$\mathcal{S}_1 = \mathcal{M}\cdot\mathcal{S}_0 \qquad [20]$$

Mueller used experimentally derived 4×4 matrices, the $\mathcal{M}$ in eqn [20], to describe the effect of an optical device on a light wave's polarization. The $\mathcal{S}$'s in eqn [20] are column matrices whose elements

Table 2 Mueller matrices for polarizers

Polarizer	*Transmission axis*	*Mueller matrix*
Linear	Horizontal	$\frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
	Vertical	$\frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
	$+45°$	$\frac{1}{2}\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
	$-45°$	$\frac{1}{2}\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Circular	Right	$\frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$
	Left	$\frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$

Table 3 Mueller matrices for retarders

Retarder	*Transmission axis*	*Mueller matrix*
Quarter-wave plate	Horizontal	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$
	Vertical	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
	$+45°$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$
	$-45°$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}$
Half-wave plate	0° or 90°	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$
	$\pm 45°$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

are the Stokes parameters (eqn [16]). The matrices are based upon an assumed linear relationship between the incident and transmitted beams (Table 3).

The analysis of the effect of a number of polarizers and retarders is made easier by the use of the Mueller–Stokes matrix calculus, coupled with the use of the Stokes vectors. To determine the Mueller matrices, one must measure the effect a device has on unpolarized, horizontally polarized, linearly polarized at $+45°$, and right-circularly polarized light. It is then an algebraic exercise to generate the elements of the matrix. A few optical devices and their Mueller matrix are listed in Tables 2 and 3.

The Mueller matrix contains 16 parameters but there are only seven independent parameters. The matrix contains no information about absolute phase but it handles partially polarized and unpolarized light without modification (Table 4).

Jones Vector

There is one other representation of polarized light, complementary to the Stokes vector, developed by Clark Jones in 1941 and called the Jones vector. It is superior to the Stokes vector in that it handles light of a known phase and amplitude with a reduced number of parameters. However, it is inferior to the Stokes vector in that, unlike the Stokes representation, which is experimentally determined, the Jones representation cannot handle unpolarized or partially polarized light. The Jones vector is a theoretical

construct that can only describe light with a well-defined phase and frequency. The density matrix formalism can be used to correct the shortcomings of the Jones vector, but then the simplicity of the Jones representation is lost.

Assuming the coordinate system is such that the electromagnetic wave is propagating along the z-axis, it was shown earlier that any polarization could be decomposed into two orthogonal $\mathbf{E}$ vectors, i.e., parallel to the x and y directions. The Jones vector is defined as a two-row, column matrix consisting of the complex components in the x and y direction:

$$\mathbf{E} = \begin{bmatrix} E_{0x}\, e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi_1)} \\ E_{0y}\, e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi_2)} \end{bmatrix} \qquad [21]$$

Table 4 Jones vectors

Horizontal polarization	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	Vertical polarization	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$
$+45°$ polarization	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$-45°$ polarization	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$
Right circular polarization	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ i \end{bmatrix}$	Left circular polarization	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -i \end{bmatrix}$

If absolute phase is not an issue, then we may normalize the vector by dividing by that number (real or complex) that simplifies the components but keeps the sum of the square of the components equal to one. For example, assume that $E_{0x} = E_{0y}$, then

$$\mathbf{E} = E_{0x}\, e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi_1)} \begin{bmatrix} 1 \\ e^{i\delta} \end{bmatrix} \qquad [22]$$

The normalized vector would be the terms contained within the bracket, each divided by $\frac{1}{\sqrt{2}}$.

The general form of the Jones vector is

$$\mathbf{E} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \qquad \mathbf{E}^* = [\mathbf{A}^*\mathbf{B}^*] \qquad [23]$$

Some examples of Jones vectors are shown in Table 4.

Jones Calculus

The Jones calculus is complementary to the Mueller calculus and operates on the Jones vector, (eqn [23]), similar to the way the Mueller matrix operates on the Stokes vector (Table 5).

The Jones matrix contains eight independent parameters with no redundancy, making it simpler than the Mueller calculus. However, the Jones calculus only applies to polarized light. The Jones calculus can be extended, using the density matrix formalism to allow manipulation of unpolarized

Table 5 Jones matrices for retarders

Retarder	*Transmission axis*	*Jones matrix*
	Horizontal	$\begin{bmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{bmatrix}$
	Vertical	$\begin{bmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}$
	$+45°$	$\frac{1}{2}\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$
Quarter-wave plate	$-45°$	$\frac{1}{2}\begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix}$
Half-wave plate	0° or 90°	$\begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$
	$\pm 45°$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Table 6 Jones matrices for polarizers

Polarizer	*Transmission axis*	*Jones matrix*
Linear	Horizontal	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
	Vertical	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$
	$+45°$	$\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
	$-45°$	$\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$
Circular	Right	$\frac{1}{2}\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}$
	Left	$\frac{1}{2}\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$

light, but with a loss of simplicity. The matrix equation for Jones calculus is

$$\mathbf{E}_{\text{out}} = \mathcal{W}\mathbf{E}_{\text{in}} \qquad [24]$$

A few optical devices and their Jones matrices, $\mathcal{W}$, are listed in Tables 5 and 6.

For every matrix in Jones calculus, there is a matrix in Mueller calculus, but the converse is not true. For example, it is possible to construct a depolarizer, by using a thick piece of opal glass in the visible or by using gold covered sandpaper in the infrared. Such a device can be described in Mueller calculus, but there is no matrix for such a device in Jones calculus (Tables 5 and 6).

See also

Polarization: Introduction.

Further Reading

Born M and Wolf E (1970) *Principles of Optics*. New York: Pergamon Press.

Clark Jones R (1941) A new calculus for the treatment of optical systems. *Journal of the Optical Society of America* 32: 488–503.

Clark Jones R (1942) A new calculus for the treatment of optical systems. *Journal of the Optical Society of America* 31: 486–493.

Guenther RD (1990) *Modern Optics*. New York: Wiley.

Hecht E (1998) *Optics*, 3rd edn. Reading, MA: Addison-Wesley.

Klein MV (1970) *Optics*. New York: Wiley.

Kliger DS, Lewis JW and Randall CE (1990) *Polarized Light in Optics and Spectroscopy*. Boston, MA: Academic Press.

McMaster WH (1954) Polarization and the Stokes parameters. *American Journal of Physics* 22: 351–362.

Mueller H (1948) The foundation of optics. *Journal of the Optical Society of America* 38: 661.

Q

QUANTUM ELECTRODYNAMICS

Contents

Quantum Theory of the Electromagnetic Field
Cavity QED
Cavity QED in Semiconductors

Quantum Theory of the Electromagnetic Field

I Bialynicki-Birula and Z Bialynicka-Birula, Polish Academy of Sciences, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

The quantization of the free electromagnetic field is explained by analogy with the quantization of a collection of harmonic oscillators.

Brief History

The first step in the formulation of the quantum theory of the electromagnetic field (EMF) was the 'quantum hypothesis' put forward in 1900 by Planck. According to this hypothesis the exchange of energy between the EMF and the cavity walls occurs in portions $h\nu$, where h – the quantum of action – is a new fundamental constant (the Planck constant) and ν is the frequency of the material oscillator in the wall. With the help of this hypothesis Planck explained the spectrum of the electromagnetic radiation in thermodynamic equilibrium with the cavity walls. The second step was made in 1905 by Einstein, who postulated that the quantum hypothesis of Planck should be applied directly to the EMF. The assumption that the energy of the electromagnetic radiation is carried by light quanta (named photons by Lewis in 1926), each of energy $h\nu$, enabled Einstein to explain the puzzling properties of the photoelectric effect, namely, the gross disparity between the tiny classical energy flux delivered to an electron and the amount of energy needed to eject the photo-electron from the metal. After the formulation of quantum mechanics, it became possible to complete the quantum theory of the EMF (Dirac 1927) by applying the general rules of quantization to the classical Maxwell theory. However, this theory ran into difficulties when it was applied to the interaction of the EMF with charged matter.

The lightest carriers of the electric charge are the electrons and their dynamics determines to a very large extent the electromagnetic properties of ordinary matter. The main principles of quantum electrodynamics (QED) – the theory unifying the quantum theory of the EMF and the relativistic theory of electrons (Dirac 1928) – have been set forth by Heisenberg and Pauli already in 1929 but the proper tool (renormalization theory) to perform highly accurate calculations was developed by Feynman and Schwinger in 1949. In 1967 QED was unified by Weinberg and Salam with the theory of weak interactions. Later, after the unification with the theory of strong interactions, QED became part of the so called 'standard model' – the present day theory of elementary ingredients of matter: quarks, leptons, gluons, intermediate bosons, and photons. Therefore, the modern quantum theory of the EMF is not a self-contained theory but part of a much more encompassing theory. Often it requires some input from a more general theory or from phenomenology. However, the quantum theory of the EMF with only a very

crude description of its coupling to the rest of the world (for example, in the form of boundary conditions or constitutive relations) is still not only a good starting point to a more elaborate theory but also a source of important information about the behavior of photons under realistic conditions. The main reason why treating photons as free particles gives a reasonably good approximation is the smallness of the Sommerfeld fine structure constant $\alpha = e^2/4\pi\varepsilon_0\hbar c \approx 1/137$ that measures the strength of the electromagnetic interactions between photons and charged particles.

Electromagnetic Field as a Collection of Harmonic Oscillators

In order to apply the same quantization procedure as in quantum mechanics, it is useful to formulate classical electrodynamics in a way that exhibits its close relationship with mechanical systems. Many properties of the electromagnetic field may be described and understood in the simplest case of the free field. Therefore, in this article we restrict ourselves to the quantum description of the free EMF.

It was discovered by Rayleigh and Jeans that the electromagnetic field may be viewed as a collection of harmonic oscillators. This analogy enables one to apply well-established methods of simple quantum mechanics to describe the much more complicated problem of the quantum field. The dynamics of a single, one-dimensional harmonic oscillator is governed by the Hamiltonian:

$$H(p,q) = \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2}$$

The equations of motion for the position and momentum of a harmonic oscillator have the same form in classical and in quantum theory:

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = \frac{p(t)}{m}, \qquad \frac{\mathrm{d}p(t)}{\mathrm{d}t} = -m\omega^2 q(t)$$

The solutions of these equations describe harmonic oscillations with the frequency ω

$$q(t) = q_0 \cos(\omega t) + (p_0/m\omega)\sin(\omega t)$$

$$p(t) = -m\omega q_0 \sin(\omega t) + p_0 \cos(\omega t)$$

A complete description of a multidimensional harmonic oscillator may be obtained in terms of normal modes. The transformation to normal modes leads to a decomposition of the Hamiltonian of the system into the sum, one term for each mode,

$$H = \sum_k \left(\frac{p_k^2}{2m} + \frac{m\omega_k^2 q_k^2}{2} \right)$$

Thus, each mode may be viewed as a separate system – a one-dimensional harmonic oscillator with characteristic frequency ω_k – noninteracting with the remaining oscillators.

The Maxwell equations for the free EMF read

$$\partial_t \mathbf{B}(\mathbf{r},t) = -\nabla \times \mathbf{D}(\mathbf{r},t)/\varepsilon, \ \nabla\cdot\mathbf{B}(\mathbf{r},t) = 0$$

$$\partial_t \mathbf{D}(\mathbf{r},t) = \nabla \times \mathbf{B}(\mathbf{r},t)/\mu, \ \nabla\cdot\mathbf{D}(\mathbf{r},t) = 0$$

In the vacuum, the constants ε and μ must be replaced by ε_0 and μ_0. The Maxwell equations lead to the wave equation for the vector $\mathbf{D}(\mathbf{r},t)$

$$\left(\frac{1}{c^2}\partial_t^2 - \Delta\right)\mathbf{D}(\mathbf{r},t) = 0$$

and to the same equation for the vector $\mathbf{B}(\mathbf{r},t)$. The solution of this equation obtained by the separation of variables has the form

$$\mathbf{D}(\mathbf{r},t) = a(t)\mathbf{u}(\mathbf{r})$$

where $a(t)$ satisfies the harmonic equation

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}a(t) + \omega^2 a(t) = 0$$

and $\mathbf{u}(\mathbf{r})$ satisfies the Helmholtz equation

$$(\Delta + (\omega/c)^2)\mathbf{u}(\mathbf{r}) = 0$$

Without going into mathematical details we shall assume that there exist a complete set of vector functions $\mathbf{u}_k(\mathbf{r})$, called the mode functions, satisfying the Helmholtz equations

$$(\Delta + (\omega_k/c)^2)\mathbf{u}_k(\mathbf{r}) = 0$$

the divergence condition $\nabla\cdot\mathbf{u}_k(\mathbf{r}) = 0$, the orthonormality conditions

$$\int \mathrm{d}^3 r \mathbf{u}_k(\mathbf{r})\cdot\mathbf{u}_l(\mathbf{r}) = \delta_{kl}$$

and also the appropriate boundary conditions. In most cases of physical interest such a set does exist.

Given the set of mode functions $\mathbf{u}_k(\mathbf{r})$ one may expand the vector field $\mathbf{D}(\mathbf{r},t)$ at each t into the series

$$\mathbf{D}(\mathbf{r},t) = -\sum_k p_k(t)\mathbf{u}_k(\mathbf{r})$$

The minus sign is purely conventional; it makes the correspondence with the harmonic oscillator a closer one. The same procedure may be repeated for the

vector $\mathbf{B}$ with a different complete set of functions $\mathbf{v}_k(\mathbf{r})$, resulting in the decomposition

$$\mathbf{B}(\mathbf{r}, t) = \sum_k q_k(t)\mathbf{v}_k(\mathbf{r})$$

The Maxwell equations are satisfied if the coefficient functions $p_k(t)$ and $q_k(t)$ obey the equations of motion for harmonic oscillators

$$\frac{\mathrm{d}q_k(t)}{\mathrm{d}t} = \frac{p_k(t)}{\varepsilon}, \quad \frac{\mathrm{d}p_k(t)}{\mathrm{d}t} = -\varepsilon\omega^2 q_k(t) \qquad [1]$$

and the functions $\mathbf{v}_k(\mathbf{r})$ are related to $\mathbf{u}_k(\mathbf{r})$ by the formulas

$$\mathbf{v}_k(\mathbf{r}) = \nabla \times \mathbf{u}_k(\mathbf{r})$$

Note that the electric constant ε plays the same role as the mass for a mechanical oscillator. The calculation of the Hamiltonian, the total field energy,

$$H = \int \mathrm{d}^3 r \left(\frac{\mathbf{D}^2}{2\varepsilon} + \frac{\mathbf{B}^2}{2\mu} \right)$$

completes the reduction of the electromagnetic field to a collection of harmonic oscillators. Upon inserting the representation of the field vectors in terms of the complete set of functions $\mathbf{u}_k(\mathbf{r})$ into the integral, one obtains the following formula

$$H = \sum_k \left(\frac{p_k^2}{2\varepsilon} + \frac{\varepsilon\omega_k^2 q_k^2}{2} \right)$$

which differs from its mechanical counterpart only by the replacement $m \to \varepsilon$. The time dependence has been omitted here because the Hamiltonian is a constant of motion.

Mode Functions

The simplest example of a complete set of mode functions is that obtained for a rectangular metallic cavity. In this case the functions $\mathbf{u}_k(\mathbf{r})$ may be expressed in terms of trigonometric functions. In order to write them down, it is convenient to replace one label k by a collection of three numbers $\boldsymbol{\kappa} = \{\kappa_x, \kappa_y, \kappa_z\}$ and an additional index λ taking on the values 1 and 2. The mode functions $\mathbf{u}_{\boldsymbol{\kappa}\lambda}(\mathbf{r})$ satisfying the proper boundary conditions have the following components

$$\mathbf{u}_{\boldsymbol{\kappa}\lambda}(\mathbf{r}) = N \begin{pmatrix} a_x^{(\lambda)} \cos(\kappa_x x)\sin(\kappa_y y)\sin(\kappa_z z) \\ a_y^{(\lambda)} \sin(\kappa_x x)\cos(\kappa_y y)\sin(\kappa_z z) \\ a_z^{(\lambda)} \sin(\kappa_x x)\sin(\kappa_y y)\cos(\kappa_z z) \end{pmatrix}$$

where $N = \sqrt{8/(L_x L_y L_z)}$, the coefficients $\{\kappa_x, \kappa_y, \kappa_z\}$ have the form $\{\kappa_x, \kappa_y, \kappa_z\} = \{\pi n_x/L_x, \pi n_y/L_y, \pi n_z/L_z\}$ and all the n's are natural numbers. The components of both vectors $\mathbf{a}^{(\lambda)}$ are subjected to the conditions $\boldsymbol{\kappa}\cdot\mathbf{a}^{(\lambda)} = 0$ (vanishing divergence) and orthonormalization $\mathbf{a}^{(\lambda)}\cdot\mathbf{a}^{(\lambda')} = \delta_{\lambda\lambda'}$. For each set $\{\kappa_x, \kappa_y, \kappa_z\}$ one must choose two independent solutions of these conditions $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$. These two vectors define two states of linear polarization. The mode function $\mathbf{u}_{\boldsymbol{\kappa}\lambda}(\mathbf{r})$ is a solution of the Helmholtz equation for the value of ω/c equal to the length of the vector $\boldsymbol{\kappa}$. The expansion of the EMF vectors into the mode functions in a rectangular cavity has the form

$$\mathbf{D}(\mathbf{r}, t) = -\sum_{\boldsymbol{\kappa}\lambda} p_{\boldsymbol{\kappa}\lambda}(t)\mathbf{u}_{\boldsymbol{\kappa}\lambda}(\mathbf{r})$$

$$\mathbf{B}(\mathbf{r}, t) = \sum_{\boldsymbol{\kappa}\lambda} q_{\boldsymbol{\kappa}\lambda}(t)\nabla \times \mathbf{u}_{\boldsymbol{\kappa}\lambda}(\mathbf{r})$$

where the summation extends over all allowed values of $\boldsymbol{\kappa}$ and two values of λ. One may explicitly write down mode functions also for a cylindrical and for a spherical cavity. They involve Bessel functions and are obtained by solving the Helmholtz equation in cylindrical and spherical coordinates, respectively.

One often uses another set of mode functions that does not correspond to any physical boundary conditions but is very helpful in the theoretical analysis. It is obtained by imposing the so-called periodic boundary conditions that were first introduced by Born in the study of crystal lattices. The periodic boundary conditions require that the electromagnetic field is periodic in space; it does not change when the coordinates are displaced by the 'period of the lattice' in all three directions. Assuming, for simplicity, that this period is the same, equal to L, in all directions, one obtains the following conditions for the mode functions

$$\mathbf{u}_k(\mathbf{r}) = \mathbf{u}_k(\mathbf{r} + L\mathbf{i}_x) = \mathbf{u}_k(\mathbf{r} + L\mathbf{i}_y) = \mathbf{u}_k(\mathbf{r} + L\mathbf{i}_z)$$

where $\mathbf{i}_i$ are the unit vectors in the directions of the coordinate axes. The complete set of solutions of the Helmholtz equation satisfying the periodic boundary conditions is labeled by the wavevector $\mathbf{k} = \{2\pi n_x/L, 2\pi n_y/L, 2\pi n_z/L\}$, where all the n's are arbitrary integers, and by the polarization index λ. The value of ω/c in the Helmholtz equation is equal in this case to the length of the vector k. Each member of the set $\mathbf{u}_{\mathbf{k}\lambda}$ has the form

$$\mathbf{u}_{\mathbf{k}\lambda}(\mathbf{r}) = \frac{1}{\sqrt{V}}\mathbf{e}_{\mathbf{k}\lambda} \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r}) \qquad [2]$$

where $V = L^3$. The polarization vectors $\mathbf{e}_{\mathbf{k}\lambda}$ must be orthogonal to the wave vector $\mathbf{k}$. It is convenient to impose the unitary orthonormality conditions $\mathbf{e}_{\mathbf{k}\lambda} \cdot \mathbf{e}^*_{\mathbf{k}\lambda} = \delta_{\lambda\lambda'}$ and also to assume that $\mathbf{e}^*_{\mathbf{k}\lambda} = \mathbf{e}_{-\mathbf{k}\lambda}$. The expansion of the EMF vectors into the mode functions obeying the periodic boundary conditions has the form

$$\mathbf{D}(\mathbf{r}, t) = -\frac{1}{\sqrt{V}} \sum_{\mathbf{k}\lambda} P_{\mathbf{k}\lambda}(t) \mathbf{e}_{\mathbf{k}\lambda} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$$
$$\mathbf{B}(\mathbf{r}, t) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}\lambda} Q_{\mathbf{k}\lambda}(t) \mathrm{i}\mathbf{k} \times \mathbf{e}_{\mathbf{k}\lambda} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \qquad [3]$$

where the sum extends again over all allowed values of the vectors $\mathbf{k}$ and over two values of λ. The expansion coefficients $P_{\mathbf{k}\lambda}$ and $Q_{\mathbf{k}\lambda}$ are now complex. In order to guarantee that the field vectors are real, they must obey the conditions

$$P^*_{\mathbf{k}\lambda} = P_{-\mathbf{k}\lambda}, \; Q^*_{\mathbf{k}\lambda} = Q_{-\mathbf{k}\lambda}$$

These conditions are satisfied if $P_{\mathbf{k}\lambda}$ and $Q_{\mathbf{k}\lambda}$ are parametrized by the real (but otherwise arbitrary) variables $p_{\mathbf{k}\lambda}$ and $q_{\mathbf{k}\lambda}$

$$P_{\mathbf{k}\lambda} = \frac{p_{\mathbf{k}\lambda} + p_{-\mathbf{k}\lambda}}{2} + \frac{\varepsilon\omega_{\mathbf{k}}(q_{\mathbf{k}\lambda} - q_{-\mathbf{k}\lambda})}{2i}$$

$$Q_{\mathbf{k}\lambda} = \frac{q_{\mathbf{k}\lambda} + q_{-\mathbf{k}\lambda}}{2} + \frac{\mathrm{i}(p_{\mathbf{k}\lambda} - p_{-\mathbf{k}\lambda})}{2\varepsilon\omega_{\mathbf{k}}}$$

The energy of the EMF splits again into a sum of independent contributions from all the modes

$$H = \sum_{\mathbf{k}\lambda} \omega_{\mathbf{k}} \left(\frac{p_{\mathbf{k}\lambda} p_{\mathbf{k}\lambda}}{2\varepsilon\omega_{\mathbf{k}}} + \frac{\varepsilon\omega_{\mathbf{k}} q_{\mathbf{k}\lambda} q_{\mathbf{k}\lambda}}{2} \right)$$

An important property of the periodic mode functions [2] is that not only the Hamiltonian but also the total momentum of the field splits into a sum of independent contributions. The total momentum $\mathbf{P}$ of the EMF is

$$\mathbf{P} = \int \mathrm{d}^3 r \, \mathbf{D} \times \mathbf{B}$$

Upon substituting into this expression the expansions [3] of the field vectors, one obtains, with the use of the orthonormality conditions,

$$\mathbf{P} = \sum_{\mathbf{k}\lambda} \mathbf{k} \left(\frac{p_{\mathbf{k}\lambda} p_{\mathbf{k}\lambda}}{2\varepsilon\omega_{\mathbf{k}}} + \frac{\varepsilon\omega_{\mathbf{k}} q_{\mathbf{k}\lambda} q_{\mathbf{k}\lambda}}{2} \right)$$

Canonical Quantization

The canonical quantization in quantum mechanics consists in replacing the c-numbers representing the classical canonical variables q_k and p_k by the operators $\hat{q}_k$ and $\hat{p}_k$. These operators are assumed to obey the canonical commutation relations

$$[\hat{q}_k, \hat{p}_l] = \mathrm{i}\hbar\delta_{kl} \qquad [4]$$

The canonical quantization of the EMF may proceed along the same lines as in quantum mechanics. All the variables q_k and p_k appearing in the decomposition into normal modes are replaced by the operators. When this is done, the field vectors $\mathbf{B}$ and $\mathbf{D}$ become the field operators $\hat{\mathbf{B}}$ and $\hat{\mathbf{D}}$ and the expression for the energy of the field becomes the Hamiltonian operator $\hat{H}$

$$\hat{H} = \sum_k \frac{\hat{p}_k^2}{2\varepsilon} + \frac{\varepsilon\omega_k^2 \hat{q}_k^2}{2}$$

The only difference between quantum mechanics and quantum field theory is that in the second case the number of canonical variables is infinite but this may cause complications only when one wants to give the quantum theory of the EMF a firm mathematical foundation. In physical applications one may think of the number of field oscillators as being very large but finite.

Time evolution of a quantum system is described by associating the time dependence either with the operators (the Heisenberg picture) or with the state vectors (the Schrödinger picture). In both pictures the time dependence is governed by the system Hamiltonian. In the Heisenberg picture the time dependence of the operators is given by the commutators with $\hat{H}$

$$\frac{\mathrm{d}\hat{q}_k(t)}{\mathrm{d}t} = \frac{1}{\mathrm{i}\hbar}[\hat{q}_k(t), \hat{H}] = \frac{\hat{p}_k(t)}{\varepsilon}$$

$$\frac{\mathrm{d}\hat{p}_k(t)}{\mathrm{d}t} = \frac{1}{\mathrm{i}\hbar}[\hat{p}_k(t), \hat{H}] = -\varepsilon\omega_k^2 \hat{q}_k(t)$$

These equations have the same form as in the classical theory (cf. eqn [1]). Therefore, in the Heisenberg picture the field operators will also obey the same equations as in the classical theory – the Maxwell equations:

$$\partial_t \hat{\mathbf{B}}(\mathbf{r}, t) = -\nabla \times \hat{\mathbf{D}}(\mathbf{r}, t)/\varepsilon, \quad \nabla \cdot \hat{\mathbf{B}}(\mathbf{r}, t) = 0$$

$$\partial_t \hat{\mathbf{D}}(\mathbf{r}, t) = \nabla \times \hat{\mathbf{B}}(\mathbf{r}, t)/\mu, \quad \nabla \cdot \hat{\mathbf{D}}(\mathbf{r}, t) = 0$$

Of course, the average values of the field operators evaluated in both pictures are the same and for every quantum state they obey the classical Maxwell equations.

The commutation relations [4] for the canonical operators and the completeness of the mode functions imply the following equal-time commutation relations for the field operators

$$[\hat{B}_i(\mathbf{r},t), \hat{D}_j(\mathbf{r}',t)] = \mathrm{i}\hbar \sum_k \varepsilon_{ijk}\partial_k \delta^{(3)}(\mathbf{r}-\mathbf{r}')$$

These relations do not depend on the choice of the mode functions. They summarize in a universal manner the canonical properties of the electromagnetic field operators.

Annihilation and Creation Operators

The canonical variables q_k and p_k are convenient for bringing home the analogy between mechanical oscillators and wave fields. However, there exists also a complementary, corpuscular aspect of the quantum theory of the EMF and that is most easily expressed in terms of photons. The best way to describe photons in quantum field theory is in terms of annihilation and creation operators. These operators were introduced by Schrödinger in the context of the quantum-mechanical harmonic oscillator but they have become really indispensable in quantum field theory. Formally, the annihilation $\hat{a}_k$ and the creation $\hat{a}_k^\dagger$ operators are defined as the following complex combinations of canonical operators

$$\hat{a}_k = \sqrt{\frac{\varepsilon\omega}{2\hbar}}\left(\hat{q}_k + \mathrm{i}\frac{\hat{p}_k}{\varepsilon\omega}\right)$$
$$\hat{a}_k^\dagger = \sqrt{\frac{\varepsilon\omega}{2\hbar}}\left(\hat{q}_k - \mathrm{i}\frac{\hat{p}_k}{\varepsilon\omega}\right) \qquad [5]$$

Annihilation and creation operators are dimensionless, one is the Hermitian adjoint of the other, and they obey very simple commutation relations:

$$[\hat{a}_k, \hat{a}_l^\dagger] = \delta_{kl} \qquad [6]$$

The Hamiltonian expressed in terms of annihilation and creation operators reads

$$\hat{H} = \sum_k \hbar\omega_k \hat{a}_k^\dagger \hat{a}_k + \frac{1}{2}\sum_k \hbar\omega_k \qquad [7]$$

This Hamiltonian determines the following time dependence of the annihilation and creation operators

$$\hat{a}_k(t) = \hat{a}_k \exp(-\mathrm{i}\omega_k t),\ \hat{a}_k^\dagger(t) = \hat{a}_k^\dagger \exp(\mathrm{i}\omega_k t)$$

The field operators expressed in terms of annihilation and creation operators corresponding to the periodic boundary conditions are

$$\hat{\mathbf{D}}(\mathbf{r},t) = \mathrm{i}\sum_{\mathbf{k}\lambda}\sqrt{\frac{\hbar\varepsilon\omega_{\mathbf{k}}}{2V}}\hat{a}_{\mathbf{k}\lambda}\mathbf{e}_{\mathbf{k}\lambda}\mathrm{e}^{-\mathrm{i}\omega_{\mathbf{k}}t+\mathrm{i}\mathbf{k}\cdot\mathbf{r}} + \mathrm{H.c.}$$

$$\hat{\mathbf{B}}(\mathbf{r},t) = \mathrm{i}\sum_{\mathbf{k}\lambda}\sqrt{\frac{\hbar}{2V\varepsilon\omega_{\mathbf{k}}}}\hat{a}_{\mathbf{k}\lambda}\mathbf{k}\times\mathbf{e}_{\mathbf{k}\lambda}\mathrm{e}^{-\mathrm{i}\omega_{\mathbf{k}}t+\mathrm{i}\mathbf{k}\cdot\mathbf{r}} + \mathrm{H.c.}$$

where H.c. stands for the Hermitian conjugate of the first term. In the limiting case for infinite empty space, when $V \to \infty$, the Fourier sum becomes the Fourier integral:

$$\hat{\mathbf{D}}(\mathbf{r},t) = \mathrm{i}\sum_\lambda \int \mathrm{d}^3k \sqrt{\frac{\hbar\varepsilon\omega_{\mathbf{k}}}{2(2\pi)^3}}(\hat{a}_\lambda(\mathbf{k})\mathbf{e}_{\mathbf{k}\lambda}\mathrm{e}^{-\mathrm{i}\omega_{\mathbf{k}}t+\mathrm{i}\mathbf{k}\cdot\mathbf{r}} - \hat{a}_\lambda^\dagger(\mathbf{k})\mathbf{e}_{\mathbf{k}\lambda}^*\mathrm{e}^{\mathrm{i}\omega_{\mathbf{k}}t-\mathrm{i}\mathbf{k}\cdot\mathbf{r}})$$

$$\hat{\mathbf{B}}(\mathbf{r},t) = \mathrm{i}\sum_\lambda \int \mathrm{d}^3k \sqrt{\frac{\hbar}{2\varepsilon\omega_{\mathbf{k}}(2\pi)^3}}(\hat{a}_\lambda(\mathbf{k})\mathbf{k}\mathbf{e}_{\mathbf{k}\lambda}\mathrm{e}^{-\mathrm{i}\omega_{\mathbf{k}}t+\mathrm{i}\mathbf{k}\cdot\mathbf{r}} - \hat{a}_\lambda^\dagger(\mathbf{k})\mathbf{k}\times\mathbf{e}_{\mathbf{k}\lambda}^*\mathrm{e}^{\mathrm{i}\omega_{\mathbf{k}}t-\mathrm{i}\mathbf{k}\cdot\mathbf{r}})$$

In this case, the commutation relations involve the (three-dimensional) Dirac delta function

$$[\hat{a}_\lambda(\mathbf{k}), \hat{a}_{\lambda'}^\dagger(\mathbf{k}')] = \delta_{\lambda\lambda'}\delta^{(3)}(\mathbf{k}-\mathbf{k}') \qquad [8]$$

The field operators may also be expressed in terms of the vector potential operator $\hat{\mathbf{A}}(\mathbf{r},t)$

$$\hat{\mathbf{D}}(\mathbf{r},t) = -\varepsilon\partial_t\hat{\mathbf{A}}(\mathbf{r},t),\ \hat{\mathbf{B}}(\mathbf{r},t) = \nabla\times\hat{\mathbf{A}}(\mathbf{r},t)$$
$$\hat{\mathbf{A}}(\mathbf{r},t) = \sum_\lambda \int \mathrm{d}^3k \sqrt{\frac{\hbar}{2\varepsilon\omega_{\mathbf{k}}(2\pi)^3}}(\hat{a}_\lambda(\mathbf{k})\mathbf{e}_{\mathbf{k}\lambda}\mathrm{e}^{-\mathrm{i}\omega_{\mathbf{k}}t+\mathrm{i}\mathbf{k}\cdot\mathbf{r}} + \hat{a}_\lambda^\dagger(\mathbf{k})\mathbf{e}_{\mathbf{k}\lambda}^*\mathrm{e}^{\mathrm{i}\omega_{\mathbf{k}}t-\mathrm{i}\mathbf{k}\cdot\mathbf{r}})$$

The vector potential is needed for the description of the interaction with charged particles.

Space of Quantum States

The standard basis in the space of quantum states consists of the state vectors of stationary states – the eigenvectors of the Hamiltonian.

Vacuum State

The lowest energy state of the EMF – the ground state of the system – is often called the vacuum state. The state vector describing the vacuum state is denoted by $|0\rangle$. The vacuum state is the common ground state of all harmonic oscillators making up the EMF. Since the ground state of a single harmonic oscillator has the energy $\hbar\omega/2$, the energy of the ground state for an infinite number of oscillators is infinite (last term in eqn [7]). This infinity has no physical consequences because only the energy differences are observable. In the presence of obstacles (mirrors, resonators, beamsplitters, etc.), owing to the change in the mode structure, there will be a change in the energy of the new ground state relative to the energy in empty space. The energy differences between various ground states do have physical significance.

Fock Basis

It follows from the commutation relations [6] and from the form [7] of the Hamiltonian that the operator $\hat{a}_k$ decreases the energy of the EMF by the Planck quantum of energy $\hbar\omega_k$. Therefore, all operators $\hat{a}_k$ must annihilate the ground state vector, $\hat{a}_k|0\rangle = 0$, because there is no state with a lower energy. Each creation operator $\hat{a}_k^\dagger$ acting on any state vector adds one excitation with the energy $\hbar\omega_k$. These excitations of the electromagnetic field with characteristic energies $\hbar\omega_k$ are interpreted as photons. The state vectors of the form $\hat{a}_k^\dagger|0\rangle$ describe one-photon states, the state vectors of the form $\hat{a}_k^\dagger\hat{a}_l^\dagger|0\rangle$ describe two-photon states, etc. All the vectors generated from the vacuum state vector by acting with the creation operators form a basis, called the Fock basis. The general state vector of this basis is characterized by a sequence of natural numbers $\{n_1, n_2, \ldots\}$ and is generated from the vacuum state according to the formula

$$|n_1, n_2, \ldots\rangle = (n_1!n_2!\cdots)^{-1/2}(\hat{a}_1^\dagger)^{n_1}(\hat{a}_2^\dagger)^{n_2}\ldots|0\rangle \qquad [9]$$

where the prefactor is needed for normalization. The number n_k tells us how many photons of type k are present in this state. The spatial structure of a field excitation created by $\hat{a}_k^\dagger$ is described by the mode function $\mathbf{u}_k(\mathbf{r})$. Thus, this mode function plays the same role as the wavefunction $\psi(\mathbf{r})$ of a massive particle and the set of indices labeling the mode functions is the analog of quantum numbers for $\psi(\mathbf{r})$. In particular, for the mode function [2], the photons have a definite momentum $\hbar\mathbf{k}$. In a cylindrical waveguide the photons may have a definite momentum only along the axis of the cylinder. In a spherical cavity the photons may have the definite value of square of the total angular momentum and its projection on a given axis. The Fock states are highly nonclassical; all average values of the EMF in these states vanish.

All states of the EMF are either pure or mixed. Pure states are described by all linear superpositions of the basis states. In particular, by superposing photon states with a range of energies, one may form nonmonochromatic photon wavepackets. All superpositions are allowed since photons do not have any absolutely conserved quantum numbers (like, e.g., charge). Mixed states are described by the density operators. Among the mixed states the prominent role is played by the thermal state describing the EMF in equilibrium with a reservoir at temperature T.

Coherent States

The one-mode coherent states $|\alpha\rangle$ are obtained from the vacuum state by the action of the unitary displacement operator $\hat{D}(\alpha, \alpha^*)$

$$|\alpha\rangle = D(\alpha, \alpha^*)|0\rangle = \mathrm{e}^{\alpha\hat{a}^\dagger - \alpha^*\hat{a}}|0\rangle = \mathrm{e}^{-|\alpha|^2/2}\mathrm{e}^{\alpha\hat{a}^\dagger}|0\rangle$$

where $\hat{a}$ and $\hat{a}^\dagger$ are the annihilation and creation operators corresponding to a given mode. Coherent states are superpositions of an infinite number of Fock states. The multimode coherent state is generated by the product of displacement operators

$$|\alpha_1, \alpha_2, \ldots\rangle = \mathrm{e}^{\sum_k(\alpha_k\hat{a}_k^\dagger - \alpha_k^*\hat{a}_k)}|0\rangle$$

In reality, the coherent states are generated by the action of an external (strong) time-dependent electric current. In that case the coefficients α_k are the projections of the current on the mode functions. The average values of the field operators $\langle\hat{B}_i(\mathbf{r}, t)\rangle$ and $\langle\hat{D}_j(\mathbf{r}, t)\rangle$ in coherent states are the solutions of the Maxwell equations in the presence of an external electric current.

Entangled States

The states of the EMF describing photons in different modes may exhibit correlations. The simplest such state is a two-photon state of the form $|\phi\rangle = 2^{-1/2}(\hat{a}_1^\dagger\hat{a}_2^\dagger + \hat{a}_3^\dagger\hat{a}_4^\dagger)|0\rangle$. In this case, when one photon is found, say, in the state 1, the other will be found with the probability 1 in the state 2. The states that exhibit correlations between photons belonging to different modes are called the entangled states.

Correlations occur also in the classical theory, but only for statistical mixtures when there is a loss of information. In quantum theory the correlations may arise even without a loss of information, i.e., for pure states. A striking feature of entangled states is that the decomposition of such a state into a sum of Fock states is not unique. For example, the state vector $|\phi\rangle$ may also be written in the form $|\phi\rangle = 2^{-1/2}(\hat{a}^\dagger_{1'}\hat{a}^\dagger_{2'} + \hat{a}^\dagger_{3'}\hat{a}^\dagger_{4'})|0\rangle$, where $\hat{a}^\dagger_{1'} = 2^{-1/2}(\hat{a}^\dagger_1 - \hat{a}^\dagger_3)$, $\hat{a}^\dagger_{2'} = 2^{-1/2}(\hat{a}^\dagger_2 - \hat{a}^\dagger_4)$, $\hat{a}^\dagger_{3'} = 2^{-1/2}(\hat{a}^\dagger_1 + \hat{a}^\dagger_3)$, $\hat{a}^\dagger_{4'} = 2^{-1/2}(\hat{a}^\dagger_2 + \hat{a}^\dagger_4)$. Entangled states play a crucial role in quantum information theory, especially in quantum cryptography.

Quantum Fluctuations and the Uncertainty Relations

It follows from the basic principles of quantum theory that there are no states of the system for which both canonical variables q and p have sharply defined values. The limits on the quantum fluctuations (variances) of these values are determined by the uncertainty relation. The most stringent form of this relation obtained from the commutation relations [4] (for a single mode) reads

$$(\Delta q)^2(\Delta p)^2 \geq \frac{\hbar^2}{4} + \frac{\langle \delta\hat{q}\delta\hat{p} + \delta\hat{p}\delta\hat{q}\rangle^2}{4} \qquad [10]$$

where $\delta\hat{q} = \hat{q} - \langle\hat{q}\rangle$, $\delta\hat{p} = \hat{p} - \langle\hat{p}\rangle$, $(\Delta q)^2 = \langle(\delta\hat{q})^2\rangle$, and $(\Delta p)^2 = \langle(\delta\hat{p})^2\rangle$. The states for which this inequality is saturated play a distinguished role in quantum theory. Since the uncertainty relation expresses the limitations on the quantum fluctuations of q and p, the best one can do in approaching the classical regime is to achieve equality in the uncertainty relation. That is why the quantum states that saturate [10] are viewed as the counterparts of the classical states. In the quantum theory of the EMF, the uncertainty relations express the complementarity between the magnetic and electric fields. The better one knows the magnetic field, the more uncertain is the electric field, and vice versa. However, the projections of the field operators $\hat{\mathbf{B}}$ and $\hat{\mathbf{D}}$ on the same direction commute (cf. eqn [5]). Therefore, there are no limitations on the magnetic and electric field values in a given direction and only the values in different directions are subjected to the quantum uncertainty relation.

In the vacuum state the variances of q and p for each mode are

$$(\Delta q)^2 = \frac{\hbar}{2\varepsilon\omega}, (\Delta p)^2 = \frac{\hbar\varepsilon\omega}{2} \qquad [11]$$

and they saturate the uncertainty relation. These fluctuations are called the shot noise. Very often one uses dimensionless quadrature operators $\hat{X}_1 = \hat{q}\sqrt{2\varepsilon\omega/\hbar}$ and $\hat{X}_2 = \hat{p}\sqrt{2/\varepsilon\omega\hbar}$, whose fluctuations in the vacuum state are equal to 1. All coherent states have the same fluctuations as the vacuum state. They may be viewed in the phase space as the vacuum state displaced from the origin (corresponding to the vacuum state) by the mean values of q and p.

Coherent states are not the only states that saturate the uncertainty relations. The other states with this property are called the squeezed states. The fluctuations of X_1 and X_2 for squeezed states are different from 1 but their product is still equal to 1, $(\Delta X_1)^2(\Delta X_2)^2 = 1$.

Fock states have well-defined energies and numbers of photons since they are the eigenstates of the Hamiltonian. However, the phases of the EMF are maximally uncertain; they are evenly spread over the whole range from 0 to 2π. That is why the average values of the field vectors vanish in all Fock states.

See also

Photon Picture of Light. **Quantum Electrodynamics:** Cavity QED.

Further Reading

Bialynicki-Birula I (1996) Photon wave function. In: Wolf E (ed.) *Progress in Optics*, vol. 36. Amsterdam, The Netherlands: Elsevier.

Bialynicki-Birula I and Bialynicka-Birula Z (1975) *Quantum Electrodynamics*. London: Pergamon.

Loudon R (1973) *The Quantum Theory of Light*. Oxford, UK: Clarendon Press.

Louisell W (1990) *Quantum Statistical Properties of Radiation*. New York: Wiley.

Mandel L and Wolf E (1995) *Optical Coherence and Quantum Optics*. Cambridge, UK: Cambridge University Press.

Power EA (1964) *Introductory Quantum Electrodynamics*. London: Longman.

Scully MO and Zubairy MS (1997) *Quantum Optics*. Cambridge, UK: Cambridge University Press.

Cavity QED

H Walther, University of Munich and Max-Planck-Institute for Quantum Optics, Garching, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Modification of the Spontaneous Transition Rate in Confined Space

In order to understand the modification of the spontaneous emission rate in an external cavity-like structure, it must be remembered that in quantum electrodynamics this rate is proportional to the density of modes of the electromagnetic field, i.e., the vacuum field fluctuations above the atomic transition frequency ω_0. As a consequence the spontaneous emission rate is increased if the atom is surrounded by a cavity tuned to the transition frequency. Conversely, the decay rate decreases when the cavity is mistuned. This fact was already recognized in the pioneering days of nuclear magnetic resonance by Purcell.

The spontaneous decay rate of the atom in the cavity, γ_c, will then be enhanced in relation to that in free space, γ_f, by a factor given by the ratio of the corresponding mode densities

$$\frac{\gamma_c}{\gamma_f} = \frac{\rho_c(\omega_0)}{\rho_f(\omega_0)} = \frac{2\pi Q}{V_c\omega_0^3} = \frac{Q\lambda_0^3}{4\pi^2 V_c}$$

where V_c is the volume of the cavity and Q is the quality factor of the cavity which expresses the 'sharpness' of the mode. For low-order cavities $V_c \approx \lambda_0^3$ this means that the spontaneous emission rate is increased by roughly a factor of Q. As mentioned already, when the cavity is detuned, the decay rate will decrease. In this case the atom cannot emit a photon, since the cavity is not able to accept it, and the energy will stay with the atom, i.e., its decay is inhibited.

To change the decay rate of an atom, in principle no resonator has to be present; any conducting surface near the radiator affects the mode density and, therefore, the spontaneous radiation rate. Parallel conducting planes can alter the emission rate somewhat but can only reduce the rate by a factor of 2 owing to the existence of modes, independent of the plate separation with an electric field perpendicular to the planes.

In order to demonstrate experimentally the modification of the spontaneous decay rate, it is not necessary to go to single-atom densities. The experiments where the spontaneous emission is inhibited can also be performed with large atom numbers. However, in the opposite case, when the increase of the spontaneous rate is observed, a large number of excited atoms may disturb the experiment by induced transitions. The first experimental work on inhibited spontaneous emission was done by Drexhage, Kuhn and Schäfer in 1974. The fluorescence of a thin dye film near a mirror was investigated. A reduction of the fluorescence decay by up to 25% was observed. Later experiments with microcavities filled with dye solutions were performed by de Martini *et al*. These experiments demonstrated that thresholdless lasing is possible in such systems.

Inhibited spontaneous emission was observed by Gabrielse and Dehmelt. In these neat experiments with a single electron stored in a Penning trap they observed that cyclotron orbits show lifetimes which are up to 10 times longer than that calculated for free space. The electrodes of the trap form a cavity which decouples the cyclotron motion from the vacuum field leading to the longer lifetime.

A new stage in experiments on cavity QED was reached when it was recognized that highly excited alkaline atoms are very suitable for these experiments. The main quantum numbers are in the range $n = 20-60$. Those so-called Rydberg atoms couple very strongly to the radiation field as will be discussed later. The transitions to neighboring states are in the microwave region. The cavities can therefore be built as low-order cavities with dimensions on the order of the wavelength of the transitions being in the mm or cm regions. The spontaneous rate of these transitions in free space is small owing to the low value of their transition frequency, therefore an enhancement is possible. Experiments with Rydberg atoms on the inhibition of spontaneous emission have been performed by Kleppner and coworkers and by Haroche and coworkers. In the latter experiment a 3.4 μm transition of the Cs atom was suppressed.

The first observation of enhanced atomic spontaneous emission in a resonant cavity was published by Haroche *et al*. This experiment was performed with Rydberg atoms of Na excited in the 23s state in a niobium superconducting cavity resonant at 340 GHz. Cavity tuning-dependent shortening of the lifetime was observed. The cooling of the cavity had the advantage of totally suppressing the black-body field. The latter effect is completely absent if optical transitions are observed, however, in this case it is more difficult to obtain low-order cavities. The first experiments on optical transitions were performed by Feld and collaborators. They succeeded

in demonstrating the enhancement of spontaneous transitions even in higher-order optical cavities.

In modern semiconductor devices both electronic and optical properties can be tailored with a high degree of precision. Therefore, electron–hole systems producing recombination radiation analogous to radiating atoms can be localized in cavity-like structures, e.g., in quantum wells. Thus optical microcavities of half or full wavelength size are obtained. Both suppression and enhancement of spontaneous emission in semiconductor microcavities were demonstrated in experiments by Yamamoto and collaborators. Yablonovitch *et al.* proposed the use of photonic band gap structures in semiconductors. In those systems spontaneous emission is forbidden in certain frequency regions owing to the special material properties.

Energy Shift in Confined Space

In the previous section we focused on changes of radiation rates of atoms near conducting walls or in cavity-like structures. Next we will discuss the more subtle phenomenon of energy shifts. While radiation rates are modified by the influence of the vacuum field on the real emission of photons the energy shifts are caused by the modification of a virtual photon interaction.

Resonant and nonresonant phenomena have to be distinguished. The resonant self-energy shift of a decaying atomic dipole in the vicinity of a conducting wall can be determined from the average polarization energy produced by the image dipole field. For distances comparable to the wavelength the near-field condition is satisfied, resulting in the z^{-3} dependence of the static dipole–dipole interaction characteristic for the van der Waals energy; under far-field conditions the distance dependence is given by z^{-1}. The polarization of the atom by the nonresonant parts of the broadband electromagnetic field causes energy shifts, of which the Lamb shift is the most prominent. In the sense of the nonrelativistic treatment by Bethe the major contribution of that shift can be described as being a result of the emission and reabsorption of virtual photons. It is plausible that just as the real emission of a photon is modified in confined space, so also is the virtual process. The latter 'real' radiation energy shift is thus a consequence of vacuum fluctuations only. It is identical with the energy shift predicted by Casimir and Polder and is analogous to the better-known result of Casimir on the force between two plane neutral conducting plates.

The question of modification of atomic energies in confined space has recently found considerable interest and many calculations of the phenomenon have been performed. Direct application to the energy shift of Rydberg atoms, which are of special interest for experimental studies, was performed by Barton in 1987. He calculated the direct electrostatic interaction with a conducting wall and the radiation induced (retarded) effects. The result is that in the case of two parallel plates the electrostatic effect is dominant when the distance L between the conducting plates is small, $L < n^3 a_0/\alpha$ (n is the principal quantum number of the Rydberg atom, a_0 the Bohr radius, and α the fine-structure constant), and the radiative effect plays the major role when large distances are used, $L > n^3 a_0 \alpha$.

Experiments to measure the van der Waals force of conducting planes have been performed by Hinds and coworkers. They could clearly demonstrate the z^{-3} dependence of the van der Waals force. The retarded effects were not detectable at large distances but at small distances from the wall a deviation from the z^{-3} dependence was found which was attributed to the retarded QED potential.

The energy shift of rubidium Rydberg atoms in confined space has been measured by Marrocco *et al.* in an experiment with extreme high-resolution using the Ramsey experiment two-field method. The atoms are excited in s-Rydberg states ($n \approx 30$) by two-photon transitions using the light of an ultrastable dye laser, the linewidth of which was less than 10 Hz. The laser intensity was enhanced in a folded cavity locked to the laser frequency. Both interaction zones necessary for the Ramsey method were enclosed in this cavity. Between the two interaction regions the atoms pass through a pair of conducting plates, the distance between which can be changed. The shift of the Ramsey interference was measured as a function of the plate distance. Using the Ramsey method has the advantage that the shift can be determined without a direct probing of the atoms in the space between the plates. A level shift on the order of about 150 Hz could be measured, in very good agreement with theory.

Maser Operation

If the rate of atoms crossing a cavity exceeds the cavity damping rate ω/Q, the photon released by each atom is stored long enough to interact with the next atom. Here ω stands for the cavity frequency and Q for the quality factor of the cavity. The atom–field coupling becomes stronger and stronger as the field builds up and evolves to a steady state. Using Rydberg atoms with a large field–atom coupling constant leads to a new kind of maser which operates with exceedingly small numbers of atoms and photons.

The photons corresponding to transitions between neighboring Rydberg atoms are in the microwave region at about 20–100 GHz. Atomic fluxes as small as a few atoms per second have generated maser action, as could be demonstrated by Walther *et al.* of the University of Munich in 1985 using a superconducting cavity. For such a low flux there is never more than a single atom in the resonator – in fact, most of the time the cavity is empty of atoms. It should also be mentioned that in the case of such a setup a single resonant photon is sufficient to saturate the maser transition.

The scheme for this one-atom maser or micromaser is shown in **Figure 1**. The setup represents the simplest system in radiation–atom interaction: a single atom is interacting with a single mode of the radiation field. This device was used to verify the complex dynamics of a single atom in a quantized field predicted by the Jaynes–Cummings model. All of the features are explicitly a consequence of the quantum nature of the electromagnetic field: the statistical and discrete nature of the photon field leads to new characteristic dynamics such as collapse and revivals in the exchange of a photon between the atom and the cavity mode. The frequency of this exchange is usually called the Rabi flopping frequency. The field in the cavity is measured through the number of atoms in the lower state of the maser transition. Due to the strong coupling between the atom and maser field both are entangled. The strong coupling can also be used to entangle subsequent atoms.

The steady-state field of the micromaser shows sub-Poisson statistics. This is in contrast to regular masers and lasers where coherent fields (Poisson statistics) are observed. The reason for nonclassical radiation being produced is that a fixed interaction time of the atoms is chosen, leading to careful control of the atom–field interaction dynamics.

Under steady-state conditions, the photon statistics $P(n)$ of the field of the micromaser are essentially determined by the pump parameter $\Theta = N_{\mathrm{ex}}^{1/2}\Omega t_{\mathrm{int}}$, denoting the angular rotation of the pseudospin vector of the interacting atom. Here, N_{ex} is the average number of atoms that enter the cavity during the decay time of the cavity τ_{cav}, Ω the vacuum Rabi flopping frequency, and t_{int} is the atom–cavity interaction time. The normalized photon number of the maser $\langle \nu \rangle = \langle n \rangle / N_{\mathrm{ex}}$ shows the following generic behavior

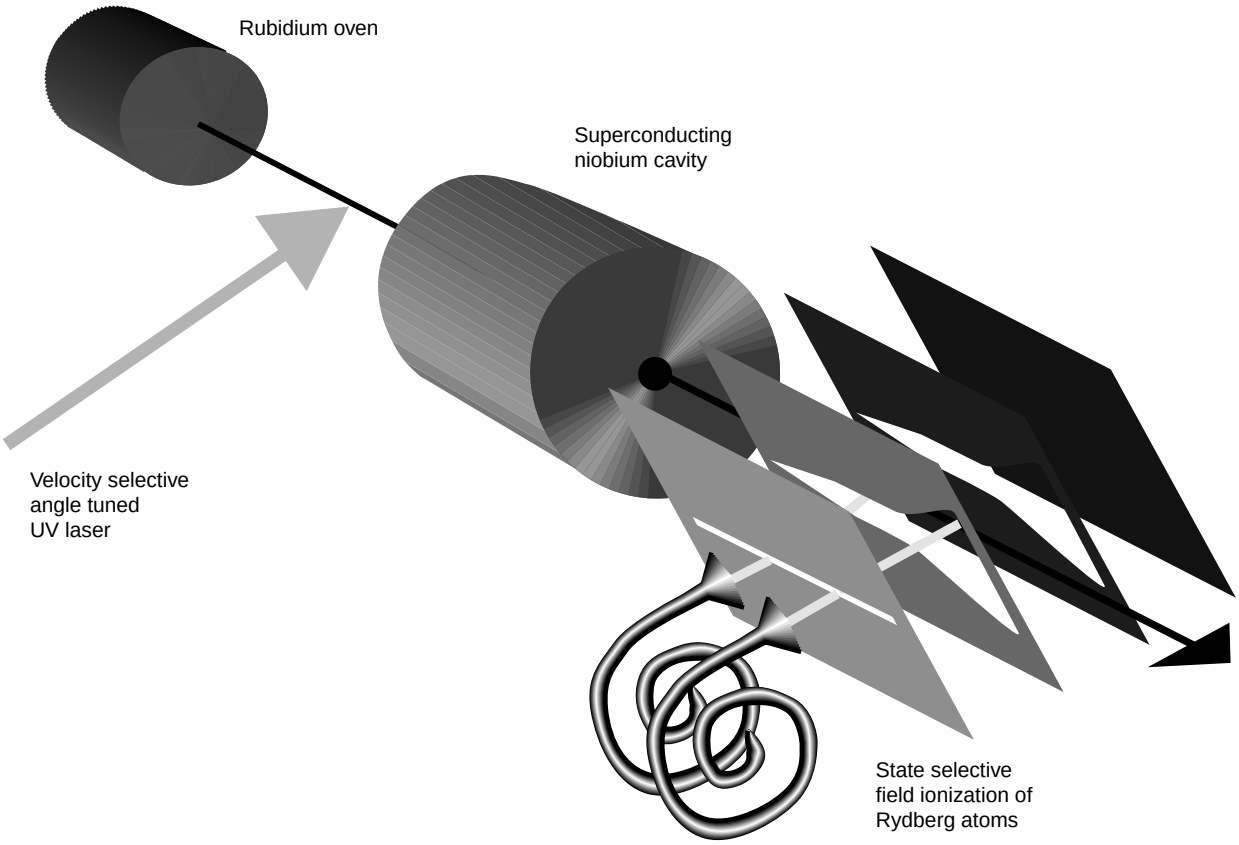

Figure 1 Micromaser setup of rubidium Rydberg atoms in the 63p state. The velocity of the atoms is controlled by exciting a velocity subgroup of atoms in the atomic beam. The atoms in the upper and lower maser levels are selectively detected by field ionization. The number of photons deposited in the cavity is determined through the number of atoms in the lower state (61d).

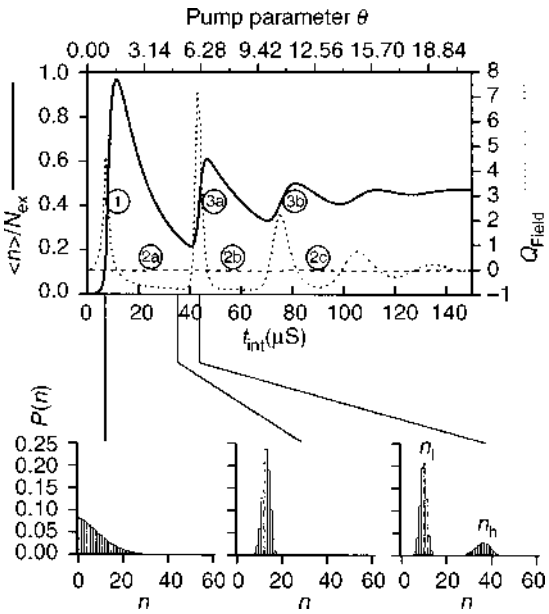

Figure 2 One-atom maser or micromaser characteristics. The upper part of the figure shows the average photon number versus interaction time (solid curve) and the photon number fluctuations represented by the Q-factor (dotted line). Both curves are determined by the photon exchange dynamics between atom and field. The pump parameter gives the angular rotation of the pseudospin vector of the interacting atom (e.g., 2π corresponds to a full rotation, i.e., the atom is again in the upper level). In the lower part of the figure the steady-state photon number distribution $P(n)$ is shown for three values of t_{int}. The distribution on the left side corresponds to the maser threshold, that on the right gives an example of the double-peaked distribution associated with the quantum jump behavior. In this situation, the atom is back in the excited state and can again emit, leading to a higher steady-state photon number n_{h}. With increasing t_{int} this part will grow and n_{l} will decrease and disappear. A new jump occurs in the region 3b.

(see Figure 2). It suddenly increases at the maser threshold value $\Theta = 1$ and reaches a maximum for $\Theta \approx 2$ (denoted by 1 in Figure 2). The behavior at threshold corresponds to the characteristics of a continuous phase transition. As Θ increases further, the normalized averaged photon number $\langle \nu \rangle / N_{\text{ex}}$ decreases to a minimum slightly below $\Theta \approx 2\pi$, and then abruptly increases to a second maximum (3a in Figure 2). This general type of behavior recurs roughly at integer multiples of 2π, but becomes less pronounced with increasing Θ. The reason for the periodic maxima of the average photon number is that for integer multiples of $\Theta = 2\pi$ the pump atoms perform an integer number of full Rabi flopping cycles, and start to flip over at a slightly larger value of Θ, thus leading to enhanced photon emission. The periodic maxima for $\Theta = 2\pi, 4\pi$, and so on may be interpreted as first-order phase transitions.

The photon statistics of the maser radiation is usually characterized by the Q-parameter introduced by Mandel:

$$Q_{\text{field}} = \left\{ \left(\langle n^2 \rangle - \langle n \rangle^2 \right) / \langle n \rangle \right\} - 1$$

It can be seen that $Q_{\text{field}} = 0$ corresponds to a Poissonian photon distribution. Q_{field} for the micromaser is plotted as a dotted line in Figure 2. The value drops below zero in the region 2a, 2b, etc. This shows the highly sub-Poissonian character of the one-atom-maser field being present over a large range of parameters.

The reason for the sub-Poissonian atomic statistics is the following. A changing flux of atoms changes the Rabi frequency via the stored photon number in the cavity. Adjusting the interaction time allows the phase of the Rabi nutation cycle to be chosen such that the probability of the atoms leaving the cavity in the upper maser level increases when the flux, and hence the photon number in the cavity, is raised. This leads to sub-Poissonian atomic statistics since the number of atoms in the lower state decreases with increasing flux and photon number in the cavity. This feedback mechanism can be neatly demonstrated when the anticorrelation of atoms leaving the cavity in the lower state is investigated. Measurements of this anticorrelation phenomenon could be made.

The fact that anticorrelation is observed shows that the atoms in the lower state are more equally spaced than expected for a Poissonian distribution of the atoms in the beam. This means, for example, that when two atoms enter the cavity close to each other, the second one performs a transition to the lower state with reduced probability.

The interaction with the cavity field thus leads to an atomic beam with atoms in the lower maser level showing number fluctuations which are up to 40% below those of a Poissonian distribution found in usual atomic beams. This is interesting because atoms in the lower level have emitted a photon to compensate for cavity losses inevitably present. Although this process is induced by dissipation giving rise to fluctuations, the atoms still obey sub-Poissonian statistics.

Generation of Number States (Fock States) of the Radiation Field

When the micromaser is operated at low temperatures, a very interesting feature is observed in the inversion of the population of the two maser levels called trapping states. They are a steady-state feature of the maser field; they occur in the micromaser as a direct consequence of field quantization in a cavity. At low cavity temperatures the number of black-body

photons, n_{th}, in the cavity mode is reduced and trapping states begin to appear; at higher temperatures they are washed out by the thermal photons. The trapping states show up when the atom–field coupling Ω and the interaction time t_{int} are chosen such that in a cavity field with n_q photons the atoms undergo an integer number k of Rabi cycles. This is summarized by the condition

$$\Omega t_{int}\sqrt{n_q + 1} = k\pi \qquad [1]$$

for the trapping state, denoted by (n_q, k). When eqn [1] is satisfied, the cavity photon number is left unchanged after interaction of an atom and hence the photon number is 'trapped'. This will occur regardless of the atomic pump rate N_{ex}. The trapping state is therefore characterized by the upper-bound photon number n_q and the number of integer multiples of full Rabi cycles k. In this situation the field is stabilized. Whenever a photon disappears, e.g., due to dissipation, the next incoming atom experiences a changed Rabi nutation frequency and emits a photon with high probability. At the trapping condition a quantum nondemolition situation is present. Through the dynamics of the Rabi nutation the field is measured, without any net change of the field.

The build-up of the cavity field can be seen in Figure 3, where the emerging atom inversion is plotted against the interaction time and pump rate. At low atomic pump rates (low N_{ex}) the maser field cannot build up and the maser exhibits Rabi oscillations due to the interaction with the vacuum field. At the positions of the trapping states, the field increases until it reaches the trapping-state condition. This is manifested as a reduced emission probability and hence as a dip in the atomic inversion. Once in a trapping state, the maser will remain there regardless of the pump rate.

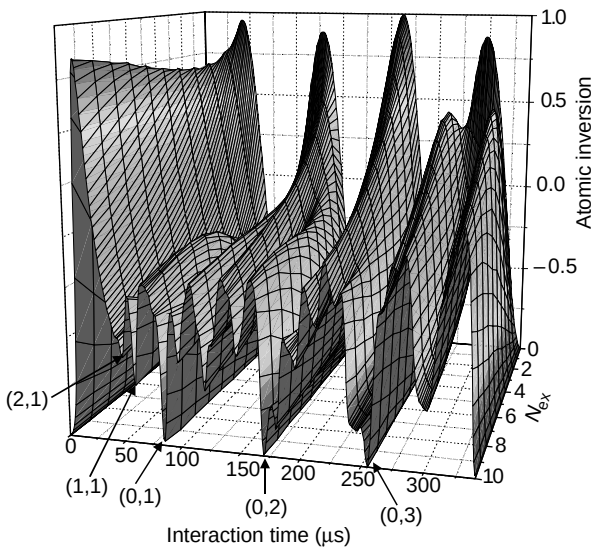

Figure 3 Behavior of the micromaser at low temperature showing the trapping states, manifested as valleys along the N_{ex} axis. For the designation of the trapping states see text.

Under ideal conditions the micromaser field in a trapping state is a Fock state, but when the micromaser is operated in a continuous wave mode, the field state is very fragile and highly sensitive to external influences and experimental parameters. However, in contrast to continuous-wave operation, in pulsed operation trapping states are more stable and more practical and can be used over a much broader parameter range than in continuous-wave operation.

To demonstrate the principle, Figure 4 shows a simulation of a sequence of 20 pulses of the pumping atoms in which an average of seven excited atoms per pulse are present. Under the trapping condition only a single emission event occurs, producing a single lower state atom which leaves a single photon in the cavity. Since the atom–cavity system is in the trapping condition, the emission probability is reduced to zero and the photon number is stabilized. Consequently, excited state atoms following the emitting atom stay in the upper maser level. The variation of the time when an emission event occurs during the atom pulses in Figure 4 is due to the Poissonian spacing of upper-state atoms entering the cavity and the stochasticity of the quantum process.

The lower part of Figure 4 shows the photon number distribution resulting from this process. It was demonstrated experimentally by Brattke *et al.* that a single photon number state could be generated with a success rate of 85%. By improving the experimental parameters one can expect to prepare single-photon Fock states in 98% of the pulses by this method.

Other Cavity Experiments

Besides the micromaser, there are also other experiments using Rydberg atoms in cavities and the strong interaction of these atoms with cavities which lead to interesting applications. One example is an experiment by S. Haroche, J. M. Raimond *et al.*, who succeeded in realizing a Schrödinger cat state in a cavity. Studies on the decoherence of this state could be conducted. The experiments are quite important in exploring the boundaries between the quantum and classical worlds. Another example is the quantum nondemolition detection of a single photon in a cavity by S. Haroche employing the dispersion level shift of probing atoms.

There is an interesting equivalence between an atom interacting with a single-mode field and a quantum

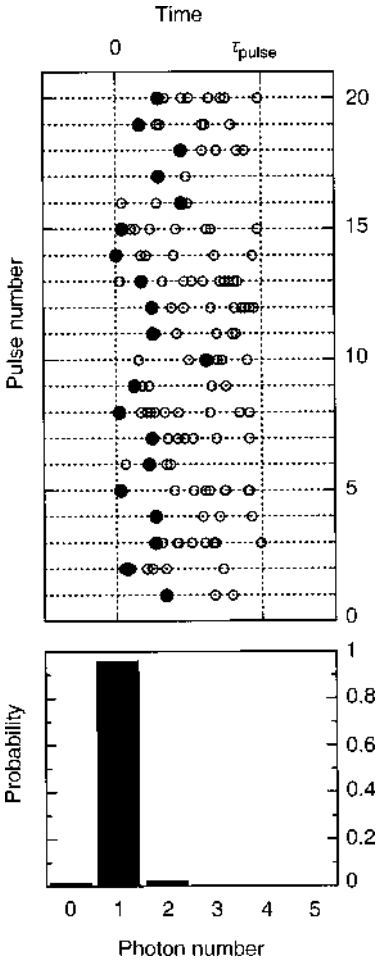

Figure 4 A simulation of a subset of 20 subsequent atom bunches generated by pulsed laser excitation with the associated probability distribution for photons in the cavity or lower-state atom production (filled circles represent lower-state atoms and open circles represent excited-state atoms). The start and end of each pulse is indicated by the vertical dotted lines marked 0 and τ_{pulse}, respectively. For details, see Battke S, Varcoe BTH and Walther H (2001) Generation of photon number states on demand via cavity, quantum electrodynamics. *Physical Review Letters* 86: 3534–3537.

particle in a harmonic potential, as was first pointed out by D. F. Walls and H. Risken; this connection results from the fact that the radiation field is quantized on the basis of the harmonic oscillator. The Jaynes–Cummings dynamics can therefore also be observed with trapped ions, as recently demonstrated in a series of beautiful experiments by D. Wineland *et al*. They also produced a Schrödinger cat state by preparing a single trapped ion in a superposition of two spatially separated wavepackets which are formed by coupling different vibrational quantum states in the excitation process.

Besides the experiments in the microwave region, a single-atom laser emitting in the visible range has also been realized by M. Feld *et al*. Furthermore, cavity quantum electrodynamic effects have been studied in the optical spectral region by J. L. Kimble *et al*.

Today's technology in microfabrication of semiconductor diode structures allows the realization of low-order cavity structures for diode lasers. In these systems the spontaneous emission is controlled in the same way as in the micromaser. Since spontaneous decay is a source of strong losses, control of this phenomenon leads to highly efficient laser systems. Quantum control of spontaneous decay thus has important consequences for technical applications. This topic will be briefly described in the next section.

Microlasers

The simplest approach to fabricating an optical microcavity is to shrink the spacing between the mirrors of a Fabry–Perot resonator to λ/n (where n stands for the refractive index) while reducing the lateral dimensions to a range of the same order of magnitude. This structure provides a single dominant longitudinal field mode that radiates into a narrow range of angles around the cavity axis.

The first optical microcavity experiments used dye molecules between high-reflectivity dielectric mirrors in the Fabry–Perot configuration. Because spontaneous emission is a major source of energy loss, speed limitations, and noise in lasers, the capability to control spontaneous emission is expected to improve laser performance. If the fraction of spontaneous emission coupled into the lasing mode is made close to one, the 'thresholdless' laser is obtained, in which the light output increases almost linearly with the pump power instead of exhibiting a sharp turn-on at the pump threshold.

Semiconductor microcavities provide high-Q Fabry–Perot cavities for both basic studies and potential applications. Molecular beam epitaxy or organometallic chemical vapor deposition techniques are used to deposit high-reflectivity mirrors consisting of alternating quarter-wavelength layers of lattice-matched semiconductors. For example, 15–20 pairs of quarter-wave layers of $Al_{0.2}Ga_{0.8}As$ and AlAs result in a reflectivity greater than 99% and Q values greater than 500. The optically active layer in such a microcavity is typically a GaAs quantum well located at the midplane of the cavity, where the field strength is maximum.

Since the microcavities have an extremely low threshold, their efficiency will also be high. Low-cost, high-density light source arrays and photonic circuits are made possible by the small size and low power consumption of such resonators. It will be possible to produce entire wafers containing millions

of microlasers with a multipole arrangement instead of having to cleave each individual semiconductor laser as at present. This improvement will lead to higher yields and lower cost per element. Another advantage of surface-emitting microcavity sources is the efficient optical coupling of their stable symmetric mode patterns into optical fibers or waveguides. These lasers are thus certain to spark a revolution in optical communication in the future.

Conclusion

This article reviewed experiments in cavity quantum electrodynamics. The experiments shed new light on our understanding of the radiation interaction of atoms. The achievable strong coupling between atoms and radiation leads to the possibility to generate entanglement between the generated radiation field and the atoms, presenting a basis for interesting applications in connection with, e.g., quantum computing and quantum information processing. Furthermore Fock states of the radiation field can be generated. The control of spontaneous decay finally leads to interesting new laser systems offering high efficiency.

See also

Quantum Electrodynamics: Cavity QED in Semiconductors; Quantum Theory of the Electromagnetic Field. **Quantum Optics:** Quantum Computing with Atoms.

Further Reading

Berman PR (1994) Cavity quantum electrodynamics. In: Bederson B and Bates D (eds). *Advances in Atomic, Molecular, and Optical Physics*, Supplement 2. Boston, MA: Academic Press.

Englert B-G, Löffler M, Benson O, *et al.* (1998) Entangled atoms in micromaser physics. *Fortschritte der Physik* 46: 897–926.

Haroche S (1992) Cavity quantum electrodynamics. In: Dalibard J, Raimond J-M and Zinn-Justin J (eds) *Fundamental Systems in Quantum Optics*. pp. 767–935. Amsterdam: North-Holland.

Haroche S and Kleppner D (1989) Cavity quantum electrodynamics. *Physics Today* 42: 24–30.

Kimble HJ, Carnal O, Georgiades N, *et al.* (1995) Quantum optics with strong coupling. In: Wineland DJ, Wieman CE and Smith SJ (eds) *Atomic Physics*, vol. 14, pp. 314–335. New York: American Institute of Physics.

Marrocco M, Weidinger M, Sang RT and Walther H (1998) Quantum electrodynamic shifts of Rydberg energy levels between parallel metal plates. *Physical Review Letters* 81: 5784–5787.

Meschede D (1992) Radiating atoms in confined space: from spontaneous emission to micromasers. *Physics Reports* 211: 201–250.

Meystre P (1992) Cavity quantum optics and the quantum measurement process. In: Wolf E (ed.) *Progress in Optics*, vol. 30. Amsterdam: North-Holland.

Walther H (1992) Experiments on cavity quantum electrodynamics. *Physics Reports* 219: 263–281.

Cavity QED in Semiconductors

M Kira, W Hoyer and S W Koch,
Philipps-University, Marburg, Germany

G Khitrova and H M Gibbs,
University of Arizona, Tucson, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

According to classical electromagnetism, propagating light fields obey Maxwell's equations under which electric and magnetic fields exchange their strength in an oscillatory manner. If a propagating field is followed at a single reference point, it is convenient to represent the electric field as a complex number $|E(t)|\exp(-\mathrm{i}\omega t)$ where $|E(t)|$ is the magnitude of the field and the phase of the field $\exp(-\mathrm{i}\omega t)$ oscillates in time t with the optical frequency ω. Figure 1 illustrates how the classical field is determined by one single point in the complex plane; consequently, the phase and amplitude of the field are known exactly. According to quantum mechanics, this simple picture has to be altered because the Heisenberg uncertainty relation states that complementary quantities like phase and amplitude cannot be determined precisely at the same time. As a result, the real and imaginary parts of the field can be determined only with accuracy $\Delta\mathrm{Re}[E]$ and $\Delta\mathrm{Im}[E]$, respectively, and the Heisenberg uncertainty principle determines the best possible accuracy $\Delta\mathrm{Re}[E]\,\Delta\mathrm{Im}[E] \geq 1$. To incorporate the fundamental inaccuracy, the electric field has to be defined by using complex-valued distributions or equivalently wavefunctions as shown in Figure 1. This quantization procedure leads to the Schrödinger equation for light analogous to that for particles. The field called quantum optics investigates the quantum electrodynamics (QED) features of light.

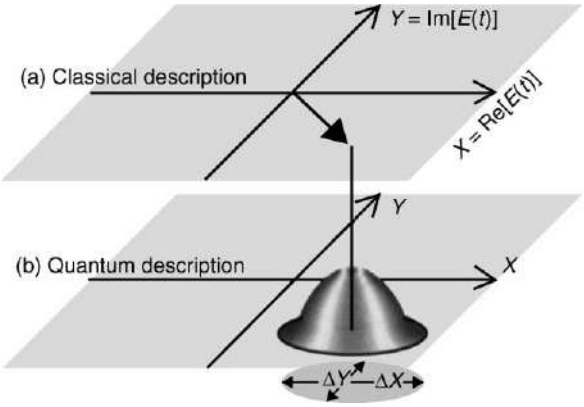

Figure 1 (a) The classically described light field $E(t)$ is a single point in the complex XY-plane; the corresponding (X, Y) vector has a definite phase and angle. (b) Quantum mechanically described light is a distribution with fluctuations ΔX and ΔY described by the shaded circle.

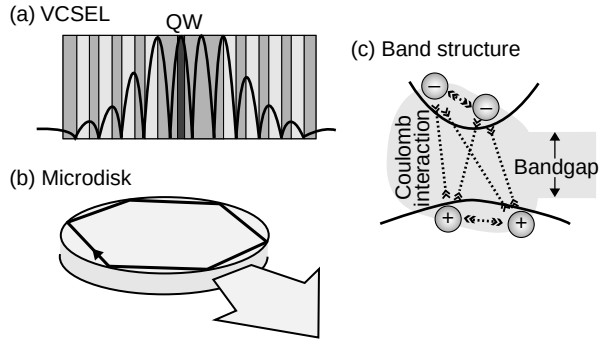

Figure 2 (a) A typical vertical-cavity surface-emitting laser (VCSEL) structure; a stack of quarter-wavelength layers with different refractive index provides a mirror such that the light intensity (solid line) is strongly confined inside the cavity. A quantum well (QW) is positioned at the field maximum. (b) In a microdisk laser, light encounters ideal total reflection when the edge of the disk is reached. The so-called whispering gallery mode emits (leaks) light from the edge. (c) A schematic illustration of semiconductor band structure, bandgap, and Coulomb interaction for an electro–hole system.

In general, light does not propagate freely in space, but it interacts strongly with the surrounding matter. For example, light can be absorbed, and its energy can be converted into excitation of the matter. As a result, the light may be slowed down, reflected, scattered, or diffracted. In order to explain the implications of the light–matter interaction in detail, one obviously has to apply a quantum mechanical description also for the matter. This approach determines the so-called eigenstates of the matter, so that the matter may occupy only certain discrete states with discrete energies. The Rydberg series of a hydrogen atom is a typical example. Similar discrete energy levels are also found for the quantized light; these levels are often referred to as photons which heuristically describe the particle properties of the light. The lowest-order light–matter interaction consists of processes where one photon is absorbed (emitted) while the matter is simultaneously excited (de-excited) from one eigenstate to another. This generic type of interaction leads to a microscopic description of the optical effects mentioned above. The magnitude of these effects depends on the strength of the interaction. By implementing different cavity configurations, the reflected light can be forced to travel across the same matter many times. As a result, the light–matter interaction is enhanced by a factor proportional to the multiple passes of the light. Thus, a cavity can be efficiently used to amplify optical phenomena. Figure 2 shows cavity configurations commonly used for semiconductors. Semiconductor cavity QED investigates quantum optical effects in semiconductor systems by enhancing them with cavity mirrors. In general, quantum optical features produce classically unexpected effects which typically stem from: (a) the discrete nature of eigenstates of light and matter; (b) the superposition principle of quantum mechanics; and (c) the Heisenberg uncertainty principle. In this article, we review both theoretical and experimental advances made so far to predict, observe, and control quantum optical effects in semiconductors with respect to effects (a)–(c).

During the last few decades, atomic quantum optics has already developed toward QED investigations, and semiconductor optics is developing rapidly into the same regime. Advanced atomic QED theories have successfully explained and guided intriguing experimental developments like laser cooling, atom condensation, and photon teleportation. However, the atomic approach can mostly be applied to describe dilute and only weakly interacting atomic gases since relatively simple models of few-level systems are used to describe the material. The elementary electronic excitations in semiconductors consist of electrons and holes (missing electrons) lifted into the conduction and valence bands, respectively. The corresponding eigenstates form continuous energy bands with the band-gap energy separation indicated in Figure 2. Since electrons and holes have opposite charges, they experience Coulomb attraction whereas bare electrons or holes repel each other. Under favorable conditions, the attractive Coulomb interaction prevails and atom-like bound electron–hole pairs (excitons) may be formed. However, since the Coulomb force has an infinite range and an electron–hole system is typically dense, excitons cannot be treated as weakly interacting quasiparticles. As a result, the atomic QED approaches cannot be used to describe quantum optics of semiconductors

in general. Thus, this article concentrates mostly on semiconductor QED theory which fully includes the interaction of charge carriers, i.e., electrons and holes, as well as quantum features of light.

The development work on quantum devices and processes is vital for advancements in several key technologies such as computers and telecommunications. The continuous decrease in component dimensions leads to a microscopic structure size (less than 1 μm); at the same time, increase in the device performance is accompanied by ultrafast operation time (faster than 1 ps). Due to these development trends, the properties of components and processes are becoming more and more quantum mechanical. Although the description of the quantum processes is complicated, the microscopic behavior offers entirely new operational functionality such as massively parallel quantum computers. These possibilities are based on the controlled manipulation of the quantum mechanical state of the light and matter. Due to rapid recent developments, optically coupled semiconductor devices have a great potential to become technologically successful and commercially viable quantum components.

Quantized States: Observation via Strong Light–Matter Interaction

In order to enhance the light–matter interaction, we choose a system where the semiconductor is placed inside a cavity in the position where the field intensity is maximum. For an empty cavity, the field intensity has a strong dependence upon frequency with a strong resonance at ω_R when the cavity length is an integer number of light wavelengths in the material. The width of the resonance is directly related to the quality of the cavity as determined by the number of back-and-forth reflections of light inside the cavity. If one chooses a vertical-cavity surface-emitting laser (VCSEL) structure, the most efficient coupling is obtained by placing a thin planar semiconductor structure at the field maximum as shown in Figure 2. If the semiconductor is planar and thin, the structure is called a quantum well because electrons and holes are confined in one direction in analogy to the standard particle-in-a-box problem of fundamental quantum mechanics. As a result, the energy levels of a quantum well are discrete in the confinement direction. If the structure is narrow enough, the system dynamics is confined to the lowest quantum-well level such that the carriers are quasi-two-dimensional due to the free in-plane motion. The optical response of such a system to a weak classical probe beam has been successfully analyzed with the so-called semiconductor Bloch equations. For the quantum-well system alone, the absorption spectrum may have a sharp resonance; this is often referred to as an excitonic resonance since it is located below the fundamental bandgap energy at a position corresponding to the exciton binding energy. Obviously, the light–matter coupling becomes large when the exciton and cavity resonances coincide. However, since these resonances are coupled, the optical response is altered; we observe a splitting into two absorption peaks instead of the original degenerate resonances. This phenomenon is commonly referred to as normal-mode coupling; in general, it is a quite common feature in quantum mechanics that degenerate states split due to an additional interaction. Figure 3 shows the first experimental observation of normal-mode coupling in semiconductors.

Since a semiconductor is a strongly interacting many-body system, several effects alter the specific character of an excitonic resonance when electrons and holes are excited. The attractive interaction between electrons and holes becomes weaker for

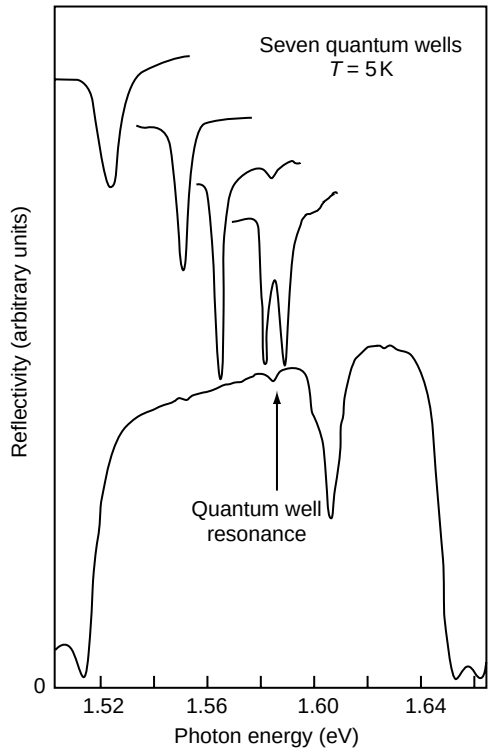

Figure 3 Probe reflectivity of a seven-quantum-well microcavity structure. The various curves correspond to the energy difference between exciton and cavity resonances. When these resonances become degenerate, the reflectivity shows two resonances corresponding to the normal-mode splitting. From Weisbuch C, Nishioka M, Ishikawa A and Arakawa Y (1992) Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity. *Physical Review Letters* 69: 3314. Copyright (2004) by the American Physical Society.

increased carrier density since the surrounding charges screen the bare Coulomb interaction. A particular exciton may also experience scattering from nearby electrons, holes, or other excitons via the Coulomb interaction. Furthermore, electrons and holes are Fermions, which requires that any given carrier state can only be occupied once. This poses fundamental limits on how many carriers can coexist in a specific volume; above a certain limit, additional occupation is prevented, and so-called Pauli blocking is observed. Since an exciton consists fundamentally of Fermionic constituents, eventually Pauli blocking effects become important even for excitons. Due to a combination of these many-body effects, the excitonic resonance is weakened by an increasing carrier density. Figure 4 shows a comparison between theory and experiment of the excitonic absorption spectrum for different carrier densities. For elevated densities, the exciton resonance is broadened and its height reduced. When the same investigations are repeated in a microcavity, we observe that the normal-mode peaks decrease in height but their separation is roughly unchanged for moderate densities. In general, the normal-mode spitting is proportional to the oscillator strength of the excitonic absorption, i.e., the area under the resonance. Thus, the observed constant splitting suggests an unchanged oscillator strength for moderate densities. The invariant oscillator strength was unexpected and can only be explained by using a theory which includes the Coulomb interaction of carriers microscopically. When the carrier density is increased even further, the exciton resonance is completely bleached, and the microcavity transmission has only a single peak, at the cavity mode energy; this is commonly referred to as the weak-coupling regime, in contrast to the nonperturbative strong coupling with two transmission peaks. The strong coupling regime provides several intriguing phenomena; for example, parametric amplification of emission has been demonstrated by applying simultaneously multiple light beams to the sample.

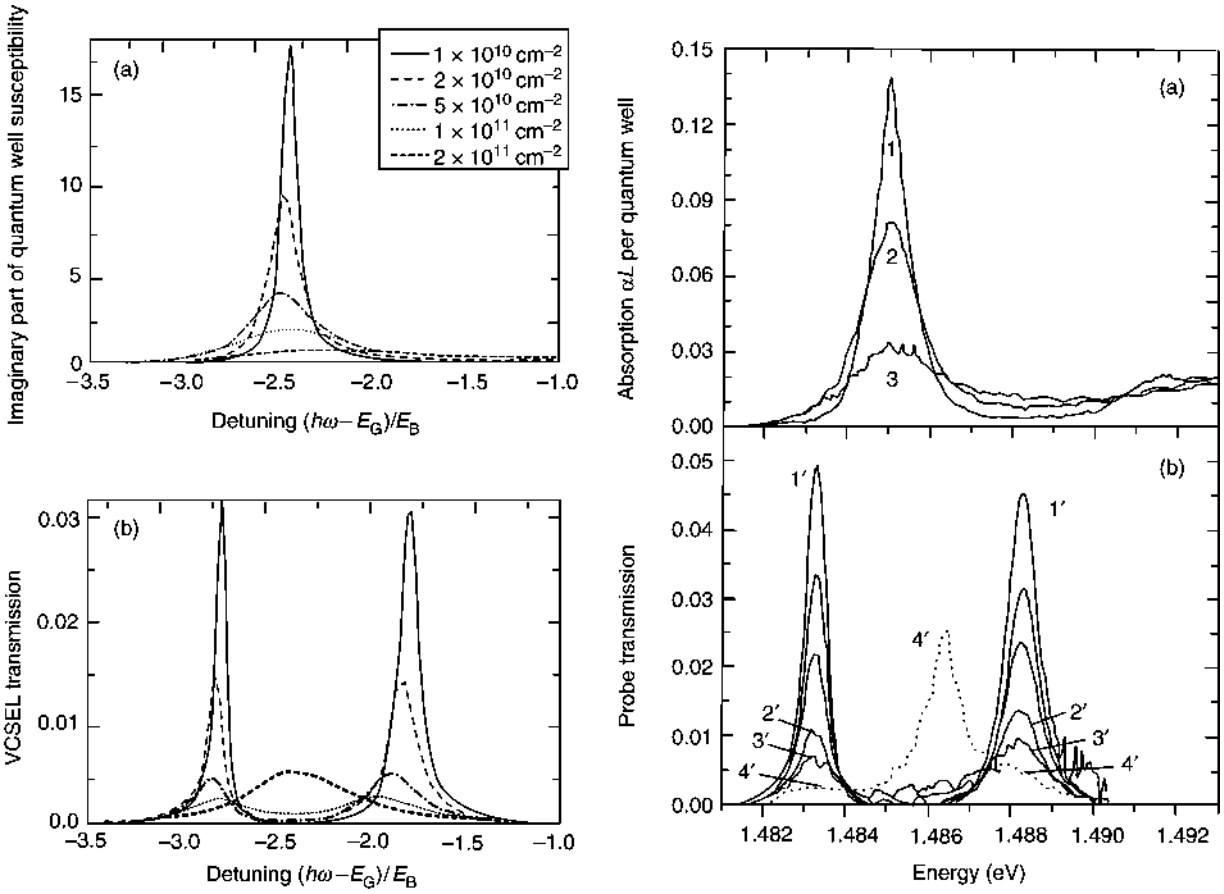

Figure 4 Left column: (a) microscopically computed bare quantum-well absorption spectrum, i.e., the imaginary part of the susceptibility, as a function of carrier density. (b) Calculated transmission of the quantum-well microcavity for corresponding densities. The right column shows experimentally determined (a) absorption and (b) microcavity transmission for the same conditions as in the calculations. From Jahnke F, Kira M, Koch SW, *et al.* (1996) Excitonic nonlinearities of semiconductor microcavities in the nonperturbative regime. *Physical Review Letters* 77: 5257. Copyright (2004) by the American Physical Society.

The ultimate cavity QED limit of normal-mode coupling follows when the light–matter interaction is so strong that even a single photon leads to splitting (genuine strong coupling). We briefly outline some aspects of the quantum statistical limit known from atomic physics. This situation can be analyzed with the so-called Jaynes–Cummings model where only two states of the atom are included together with a single light mode. In this case, the interaction between a single photon and an atom leads to a normal-mode splitting g_0. If two photons interact with one atom, the splitting increases to $2^{1/2}g_0$; more generally coupling between n photons and an atom leads to a splitting of $n^{1/2}g_0$. In the QED limit, the matter response is very nonlinear because one can detect the quantized nature of light directly from the energy splitting. This limit has already been reached in atomic cavity QED, whereas this regime is hard to approach with semiconductors because they behave more like multi-atom systems. If one has N atoms in a cavity, the first photon excites one atom, but there are N different ways of doing this, so the splitting is $N^{1/2}g_0$. If a second photon arrives, it interacts with the remaining unexcited atoms rather than the excited one, so that the splitting remains $N^{1/2}g_0$ as long as the number of photons is much smaller than N. Thus, in an N-atom system, or equivalently in semiconductors, the quantized light effects are much harder to observe than for a single atom. Currently, the normal-mode splitting in a semiconductor microcavity is explainable by the classical features of light, so that only the quantized nature of the matter is observed. In order to approach the ultimate cavity QED limit, one obviously has to reduce N and increase the light–matter coupling. In the future, this objective might be possible in quantum-dot/nanocavity systems where the semiconductor is confined in all spatial directions.

When an excited semiconductor is not under any influence of external classical light fields, it can still emit light via spontaneous emission resulting from the recombination of electron–hole pairs. The resulting light emission is called photoluminescence. Since the emission process takes place in a strongly interacting many-body system, one has to systematically include the Coulomb interaction and Fermionic features. The emission properties can be consistently described by the so-called semiconductor luminescence equations. When the microcavity photoluminescence spectrum is investigated, one observes a similar normal-mode splitting as for the transmission studies. However, this normal-mode coupling is not in the ultimate cavity QED limit even though the quantum fluctuations of the light field trigger the spontaneous emission. Nevertheless,

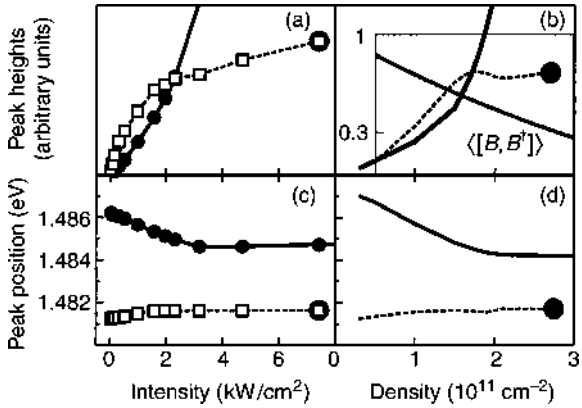

Figure 5 Measured microcavity luminescence intensities (a) and peak energies (c) versus excitation intensity for the high-energy peak (solid line) and the low-energy peak (dashed line). The results of the microscopic theory are shown in (b) and (d). From Kira M, Jahnke F, Koch SW, *et al.* (1997) Quantum theory of nonlinear semiconductor microcavity luminescence explaining "boser" experiments. *Physical Review Letters* 79: 5170. Copyright (2004) by the American Physical Society.

the luminescence shows an interesting new transition as a function of the excitation level. Figure 5 contains a comparison between theory and experiment of normal-mode peak heights and positions as a function of carrier density. For low densities, the high-energy peak is lower in height but it overtakes the low-energy peak for moderate carrier densities that are still below lasing threshold. This threshold-like overtaking was attributed to Boser action, involving exciton formation, final-state stimulation, and Bose condensation. The inset shows an estimate of just how Bosonic the excitons are, i.e., how much they actually behave as independent atoms; the value unity corresponds to the fully Bosonic situation. The nonlinear luminescence transition takes place at a density regime where the underlying Fermionic electron and hole contributions become important (commutator is roughly 0.5). A more detailed analysis with the semiconductor luminescence equations shows that Fermionic carrier nonlinearities in a detuned cavity are responsible for the experimental observations. This example of a misinterpretation based on a Bosonic analysis stresses how important it is to include the Coulomb interaction and Fermion character of carriers when analyzing the properties of semiconductors.

Superposition Principle: Observation of Quantum Optical Correlations

The above normal-mode-coupling investigations have shown interesting nonlinear effects, but they also revealed that the ultimate QED limit – demonstrating

discrete states of light directly – has not yet been reached. Alternatively, a QED investigation may concentrate on other fundamental features of quantum mechanics. One possibility is to study the implications of the superposition principle which states that the joint properties of a light–matter system can always be expressed as a superposition that combines light and matter wavefunctions. The most dramatic consequence can be observed in so-called entangled wavefunctions which cannot be factorized into light and matter parts. In a factorized wavefunction, the light and matter parts are independent whereas entangled wavefunctions conditionally connect light and matter wavefunctions. To elaborate the subtle details of entanglement, we first analyze a simple example. Assume that light can be in two different polarization states, $|\sigma^+\rangle$ or $|\sigma^-\rangle$, and the matter part is either excited $|\text{up}\rangle$ or de-excited $|\text{down}\rangle$. A wavefunction, $[|\text{up}\rangle + |\text{down}\rangle]$ $[|\sigma^+\rangle + |\sigma^-\rangle]$, is clearly a superposition of the fundamental states, and at the same time the light and matter parts are completely factorized. However, the wavefunction $[|\text{up}\rangle\ |\sigma^+\rangle + |\text{down}\rangle\,|\sigma^-\rangle]$ is entangled, since the light and matter parts cannot be separated. In the entangled state, measurement on the light state will conditionally determine a definite state of the matter, whereas the factorized state has no such conditionality. The entanglement has far-reaching consequences which cannot be understood with a classical analysis. For example, the principles of teleportation and quantum computing follow directly from the controlled manipulation of different parts of the entangled wavefunction. For atomic systems, wavefunction entanglement has been demonstrated and utilized in several experiments. For semiconductors, the direct manipulation and detection of the wavefunction seems difficult due to the overwhelming complexity of the many-body wavefunctions. Once again, the entanglement and the wavefunction are simplest for low-dimensional structures; direct entanglement effects have been demonstrated recently in a single quantum dot and between a pair of dots. Also for more complicated semiconductor systems, such as a quantum well in a microcavity, the entanglement can be observed as correlations between light and matter. In this case, the existence of QED correlations basically means that light and matter properties depend conditionally on each other. In general, the direct entanglement and QED correlation investigations have a large development potential for semiconductors.

When a semiconductor is excited with an external light pulse, entanglement-related correlations couple the classical and luminescence emission dynamics. The resulting dynamic interplay between the semiconductor Bloch and luminescence equations is mainly mediated by the correlation between photons and electron–hole densities. In the following, we analyze a microcavity configuration where a strong pump pulse generates large QED correlations. The effect of the correlations is then measured by the response of a weak probe beam. Figure 6 shows a comparison of theory and experiment of probe reflection in a configuration where the pump and probe do not have any spectral overlap. The measured probe reflection displays long-living oscillations as a function of time delay, i.e., phase difference, between the pump and the probe. Only by including the QED correlation in the theory can the oscillatory probe reflection be explained. When the QED contributions are omitted from the theory, the phase difference does not have any effect on the probe reflection. Thus, this experiment–theory example represents a direct observation of the cavity QED effects in semiconductors. Figure 6 also shows more pronounced oscillations when two phase-locked pump pulses provide the excitation; again the full QED theory explains the enhanced oscillation features.

The QED features can also alter the normal-mode coupling characteristic of a weak probe beam. Figure 7 shows a comparison of theory and experiment for a situation where the pump spectrum is

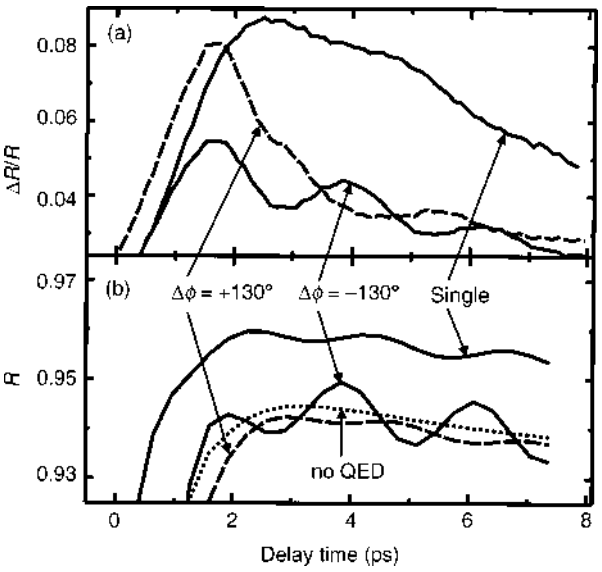

Figure 6 (a) Measured differential reflectivity and (b) computed reflection probability of the probe pulse as function of probe delay with respect to the excitation pulse. For two-pulse excitations, ±130° relative phase shifts are used. The dotted line is computed without the QED correlations. From Lee Y-S, Norris TB, Kira M, *et al.* (1999) Quantum correlations and intraband coherences in semiconductor cavity QED. *Physical Review Letters* 83: 5338. Copyright (2004) by the American Physical Society.

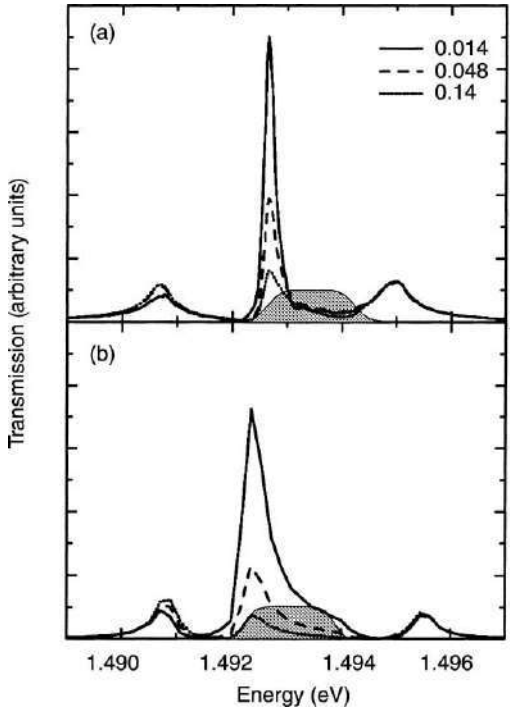

Figure 7 Dependence of probe transmission on probe intensity (nJ cm^{-2}) at a constant pump intensity of 330 nJ cm^{-2} for the microcavity. (a) Experiment, and (b) theory. The shaded regions indicate the pump spectrum. From Ell C, Brick P, Hübner M, *et al.* (2000) Quantum correlations in the nonperturbative regime of semiconductor microcavities. *Physical Review Letters* 85: 5392. Copyright (2004) by the American Physical Society.

located between the two normal-mode coupling resonances. Both the theory and experiment show a new third resonance which follows the energetic position of the pump. The third peak is a direct consequence of the QED correlations. Since the third peak is generated from the QED field–carrier correlations generated by the pump pulse, its magnitude follows the intensity of the pump. As the probe intensity is decreased the relative effect of the QED correlations is increased so that the third peak grows in the probe transmission.

In these QED investigations, the cavity plays an important role since the mirror feedback leads to a significant amplification of the QED effects through the enhanced light–matter coupling. From a practical point of view, normal-mode coupling also provides well-separated spectral features which can be used as classical-emission reference points.

Heisenberg Uncertainty Principle: Squeezing of Light Emission

Quantized light effects can also be investigated by measuring the light fluctuations of the field directly. To maximize the quantum effects, we analyze the field fluctuations in the emission directions where the classical field vanishes. This situation can be realized in planar quantum-well structures which are nearly free of disorder. In such systems, a light pulse propagating perpendicular to the structure leads to transmission and reflection of classical light only along the excitation axis. If the detection is performed at an angle, the so-called secondary emission is purely quantum mechanical. However, the classical and quantum emissions are coupled in the same way as for the QED correlation study. The resulting fluctuations of secondary emission obey the Heisenberg uncertainty principle; in the following, the special quantum features of these fluctuations are investigated.

In the full analysis, we determine the variance ΔX and ΔY of the emission as shown in Figure 8 (see also Figure 1). The Heisenberg uncertainty principle requires that the quadrature fluctuations obey $\Delta X \Delta Y \geq 1$. For a specific quadrature, the minimum uncertainty limit is usually defined to be $\Delta X = 1$ or $\Delta Y = 1$. If ΔX and ΔY are different, the observed light field is squeezed; and if the variance in one quadrature is less than one, the field is squeezed below the minimum uncertainty limit. In both cases, the field has a strong quantum nature; in particular, squeezing below unity suggests that measurements can be more accurate than the standard quantum limit for that quadrature.

To illustrate the behavior of the quantum fluctuations, we excite the quantum well resonantly with a relatively strong pulse. The emission is detected at an angle of 45° away from the excitation axis. Since the excitation pulse is relatively strong, the carrier density starts to oscillate during the pulse because Fermionic carriers can be excited only once; thus, further excitation actually leads to de-excitation. In other words, when these states become almost fully excited, the pulse can no longer excite the system, and we observe periodic de-excitation and excitation analogous to the Rabi-flopping of a strongly excited two-level system. Figure 8 shows the exciting pulse and corresponding quadrature fluctuations during the excitation process. As long as the pulse is present, the squeezing is at the 4% level and oscillates with the Rabi-flopping frequency. The quadrature fluctuations clearly show squeezing below the minimum uncertainty limit. Hence, the field has obvious quantum properties.

Similar squeezing and quantum characteristics statistics have been predicted and observed in the photon statistics of resonance fluorescence from two-level atoms subjected to an intense coherent light beam. The quantum properties of the scattered light

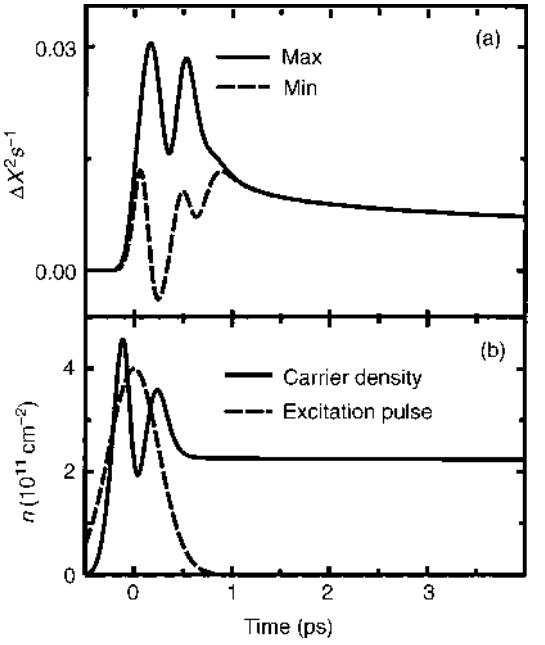

Figure 8 (a) Time-resolved maximum (solid line) and minimum (dashed line) deviations of the mode quadratures from the vacuum value of unity. (b) Corresponding time-resolved carrier density (solid line) and excitation pulse (dashed line). From Kira M, Jahnke F and Koch SW (1999) Quantum theory of secondary emission in optically excited semiconductor quantum wells. *Physical Review Letters* 82: 3544. Copyright (2004) by the American Physical Society.

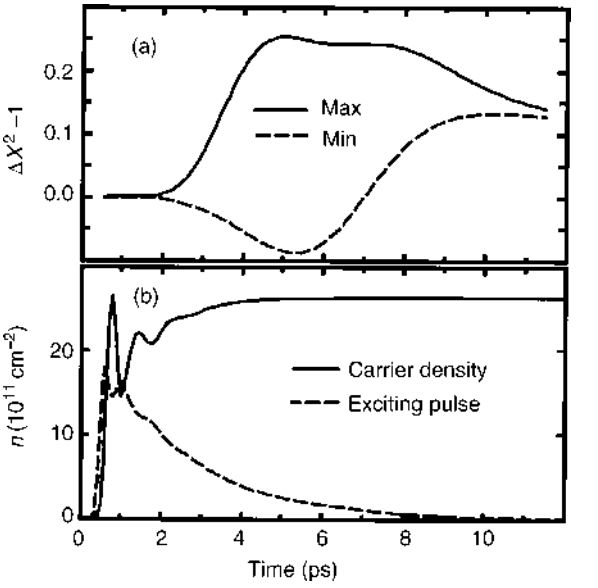

Figure 9 Squeezed light from a quantum well inside a microcavity: (a) the time evolution of maximum and minimum of the quadrature fluctuations; (b) the corresponding density and light field intensity at the quantum well.

in both quantum-well and atomic emission are enhanced when the driving field forces the system to oscillate between the excited and de-excited state. Just as described above, when the system becomes de-excited, it can no longer emit an additional photon; this inhibition manifests itself as sub-Poissonian photon statistics and squeezing in the mode quadratures. When the excitation pulse re-excites the system, such restrictions are no longer imposed. Thus, the field properties show quantum features oscillating with the Rabi-flopping frequency. Squeezing effects can also be observed with excitation by an electric current; for example, amplitude squeezing of diode laser emission has been demonstrated by controlling the electron statistics of the current flow.

The squeezing investigations demonstrate that some QED effects are expected to be seen without a cavity. Figure 9 shows the quadrature fluctuations when the cavity is added to the squeezing analysis. We observe qualitatively similar squeezing, but now the level of squeezing is enhanced up to 30% compared to the minimum uncertainty. This is once again a demonstration how QED effects can be amplified by using an increased light–matter coupling provided by the cavity.

Summary

This article overviews some of the most intriguing features of cavity quantum electrodynamics in semiconductors. Even though the related research has started only recently, several important quantum phenomena have already been predicted and measured: (a) the discrete quantum mechanical states of light and/or matter can be measured with strong coupling; (b) the consequences of the superposition principle have been detected as light–matter entanglement; and (c) the Heisenberg uncertainty principle has been tested via the squeezing of light. In all cases, the cavity enhances light–matter coupling leading to more pronounced quantum effects. Compared to atomic systems, the semiconductor has strong Coulomb correlation effects due to the relatively high density, which makes the analysis challenging. However, the semiconductor can also provide new mechanisms like the photon/carrier-density correlations which trigger new unexpected quantum phenomena.

The field of semiconductor quantum optics is very active and is developing rapidly. It benefits both from advances in computer capabilities and in semiconductor crystal growth. The increasing computer capacity allows more profound, accurate, and realistic modeling of semiconductor structures. At the same time, advances in crystal growth technology provide us with improved samples which are almost disorderless. In particular, the growth of quantum

wells with narrower exciton linewidths and quantum dots with larger dipole moments and reduced dephasing rates may be achieved, which is certain to make quantum features increasingly apparent and unavoidable.

All these research efforts eventually focus on producing devices utilizing quantum mechanical principles. One of the main objectives endeavors to develop quantum logic components for building blocks of quantum computers. Similar expanding possibilities can be expected, e.g., for accuracy of detection, device efficiency, and component design in general. Considering all of this, we are almost guaranteed to see new astonishing advancements; in this bright future, semiconductor cavity QED research will most likely be a pre-eminent element.

See also

Quantum Electrodynamics: Cavity QED.

Further Reading

Allen L and Eberly JH (1987) Optical Resonance and Two-Level Atoms, 1st edn. New York: Dover.

Banyai L and Koch SW (1993) Semiconductor Quantum Dots, 1st edn. Singapore: World Scientific.

Berman PR (ed.) (1994) Several articles in *Cavity Quantum Electrodynamics*, 1st edn. Boston, MA: Academic Press.

Cohen-Tannoudji C, Dupont-Roc J and Grynberg G (1989) Photons and Atoms, 3rd edn. New York: Wiley.

Haug H and Koch SW (2004) Quantum Theory of the Optical and Electronic Properties of Semiconductors, 4th edn. Singapore: World Scientific.

Khitrova G, Gibbs HM, Jahnke F, Kira M and Koch SW (1999) Nonlinear optics of normal-mode-coupling semiconductor microcavities. *Reviews of Modern Physics* 71: 1591–1639.

Kira M, Jahnke F, Hoyer W and Koch SW (1999) Quantum theory of spontaneous emission and coherent effects in semiconductor microstructures. *Progress in Quantum Electronics* 23: 189–279.

Scully MO and Zubairy MS (1999) Quantum Optics, 1st edn. Cambridge, UK: Cambridge University Press.

Walls DF and Milburn GJ (1994) Quantum Optics, 1st edn. New York: Springer-Verlag.

QUANTUM OPTICS

Contents

Atom Optics
Atomic Coherence Effects
Entanglement and Quantum Information
Laser Cooling of Ions
Quantum Computing with Atoms
Squeezed Phonons in Solids

Atom Optics

A D Cronin and D E Pritchard, Massachussetts Institute of Technology, Cambridge, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

In 1923 Louis de Broglie suggested that particles of matter propagate as waves with a wavelength

$$\lambda_{dB} = h/p \qquad [1]$$

where h is Planck's constant and p is the particle's momentum. This breakthrough idea motivated Schrödinger's equation, which is the equation of motion of quantum mechanics.

Schrödinger's equation is a wave propagation equation, and therefore implies the existence of a whole new type of optics – matter wave optics – in which electron waves, neutron waves, atom waves, and entire molecule waves can be manipulated coherently. Such optics enable new measurement technologies and devices that depend on matter wave interference.

Guided by a knowledge of light waves, one can understand that the de Broglie wavelength of a particle sets the scale for focusing, diffraction, and interference. However, there are differences between

matter and light. For example, the vacuum is tremendously dispersive for matter waves. In other words, the group velocity for a particle depends on its de Broglie wavelength, while for light waves the group velocity is independent of wavelength.

To illuminate the mathematical connection with light optics, we note the wave equation for light is

$$\nabla^2 \vec{E} = \mu\varepsilon \frac{\partial^2 \vec{E}}{\partial t^2} \qquad [2]$$

where $\vec{E} = \vec{E}(\vec{x}, t)$ is the optical electric field, and the parameters μ and ε may vary throughout space, describing an index of refraction. The wave equation for matter waves is Schrödinger's equation:

$$\left(-\frac{\hbar^2\nabla^2}{2m} + V\right)\psi = i\hbar\frac{\partial}{\partial t}\psi \qquad [3]$$

where $\psi = \psi(\vec{x}, t)$ is the probability amplitude, or wavefunction, for a particle of mass m. Here, the potential energy V can vary throughout space, and thus control the phase and amplitude of propagating atoms. Just as density for photons is proportional to $\vec{E}(\vec{x})^2$, the probability density for particles is given by $|\psi(\vec{x})|^2$.

In a vacuum $V = 0$ and $(\mu_0\varepsilon_0)^{-1/2} = c$, and if we assume that both ψ and E have a time dependence described by $e^{-ik\omega}$, the wave equations become

$$\left[\nabla^2 + \frac{\omega^2}{c^2}\right]\vec{E} = 0 \qquad [4]$$

and

$$\left[\nabla^2 + \frac{2m\omega}{\hbar}\right]\psi = 0 \qquad [5]$$

which look similar, but have a critical difference stemming from the single time derivative in Schrödinger's equation. To see how this makes the vacuum dispersive for matter waves, but not for light, consider a propagating plane wave of either ψ or $\vec{E}$ described by $e^{i(kx-\omega t)}$ (for either kind of wave, $k \equiv 2\pi(\lambda)^{-1}$). Using the wave equation for light waves, the dispersion relation is

$$\omega = ck \qquad [6]$$

where c is the phase (and group) velocity of light, which is independent of k. The same plane wave substituted into Schrödinger's equation for matter waves yields the dispersion relation

$$\omega = \frac{\hbar k^2}{2m} \qquad [7]$$

This is quadratic for matter waves, which means the group velocity $v_g = d\omega/dk$ is twice the phase velocity $v_p = \omega/k$, and both still depend on k.

From an engineering perspective, matter wave optics consists of lenses, mirrors, beam-splitters, and other components of an optics toolkit. The fact that electrons and neutrons can be transmitted through solid material, and that electrons can be reflected or refracted by static electromagnetic fields, has enabled optical elements to be created for these kinds of matter waves. For atoms, however, finding suitable materials for lenses or mirrors is a daunting task, since with rare exceptions (e.g., cold H on liquid He) atoms thermalize on surfaces rather than elastically bouncing off or passing through them. This was the major obstacle which delayed the blossoming of atom optics until the mid-1980s.

This obstacle has now been circumvented, but only with difficulty. Electromagnetic fields (whose spatial configurations are severely constrained by Maxwell's equations) can change the potential energy of neutral atoms, but since an atom's net charge is zero, provide only a weak and highly dispersive index of refraction based on the atom's polarizability or magnetic moment. However, these effects are tremendously enhanced if the electromagnetic fields oscillate near a transition frequency of the atom. Hence, propagating waves and standing waves of single-frequency laser light have become standard tools in atom optics. Standing light waves are often referred to as a light crystal, because the periodic potential can diffract atom beams.

An alternative source of atom optical components is based on nanofabrication. Thin sheets of material with patterned holes serve as absorptive atom optics by transmitting atom waves only through the nanometer scale openings. A variety of techniques have been used to make physical diffraction gratings, zone plates, and holograms for atom waves. Methods such as electron beam lithography and ultraviolet photolithography both rely on sophistocated etching procedures to make the final free-standing structures. Thus, in either approach – near-resonant light, or miniature physical structures – current technologies such as the tunable laser or nanofabrication techniques have been required for the development of useful atom optics.

While coherent manipulation of atom waves is often the goal, incoherent manipulations are possible also, e.g., if photons are spontaneously scattered. Dissipative processes offer an additional class of possibilities such as increasing the brightness of an atom source, and are therefore an important part of atom optics. This article excludes atom slowing and trapping, and focuses instead on atom diffraction – a coherent

process which can be used to make beam splitters for atom interferometers.

Nanofabricated Atom Optics

Because atoms stick to surfaces, mechanical structures are absorptive atom optics. Yet the de Broglie waves transmitted through slots of a mask are nearly unaffected, provided the mask is thin enough. This is because the van der Waals potential energy of interaction with nearby surfaces falls off rapidly for neutral atoms.

Consider, for example, an atom wave packet being diffracted by a mechanical grating made with physical bars and slots as shown in Figure 1. Initially, let the atom wave packet have a slowly varying spatial envelope, $f = f(x, y, z)$, and momentum only in the $\hat{x}$-direction, $\vec{p} = \hbar\vec{k}_0 = h(\lambda_{dB})^{-1}\hat{x}$, so the wavefunction can be written as

$$\psi = fe^{i(k_0x - wt)} \qquad [8]$$

Immediately after the grating, only the portions of the wavefunction passing through slots remain nonzero, as depicted in Figure 2. Here, the wavefunction is now periodic both due to the grating lattice vector $k_g\hat{y}$ and the initial wavevector $k_0\hat{x}$. According to the Schrödinger equation, a periodicity in any direction implies a momentum in that direction. Thus, variation of ψ in the $\hat{y}$ variation gives rise to a new momentum distribution.

The square wave modulating the envelope of the wave packet can be described by a Fourier sum times the incident wavefunction

$$\psi = fe^{i(k_0x - \omega t)} \sum_{n=-\infty}^{\infty} a_n \cos(nk_gy) \qquad [9]$$

where $k_g = 2\pi(d_g)^{-1}$ is the grating lattice vector. The only other physical parameter of the grating is the open fraction, γ, and this determines the Fourier amplitudes a_n

$$a_n = \frac{\sin(n\pi\gamma)}{n\pi} \qquad [10]$$

From the principle of superposition (which is applicable here because the Schrödinger equation is linear) the wavefunction can be written as a sum of traveling waves with wavevectors $\vec{k}_n$:

$$\psi = f\sum_{n=-\infty}^{\infty} a_n e^{(i\vec{k}_n\cdot\vec{r} - wt)} \qquad [11]$$

$$\vec{k}_n = \frac{k_0\hat{x} + nk_g\hat{y}}{N} \qquad [12]$$

where, just as in standard optics, the factor N serves to keep $|\vec{k}_n|$ on a spherical shell, i.e., the grating does not add energy to the atom.

Inserting this expression for the wavefunction into Schrödinger's equation we get the dispersion relation

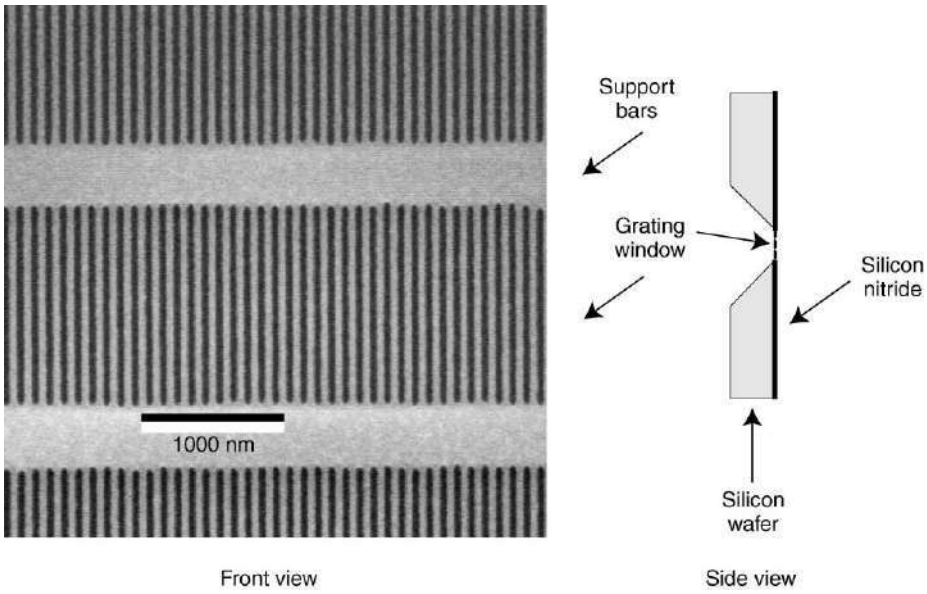

Figure 1 Scanning electron microscope image of a silicon nitride diffraction grating for atom waves. The bars (light) absorb, and the slots (dark) transmit atom waves. The grating period is 100 nm, the open fraction is 0.67, and the free-standing bars are supported every 5 μm with a thicker strip of silicon nitride material. The grating was fabricated by Tim Savas and Henry I Smith at MIT NanoStructures laboratory.

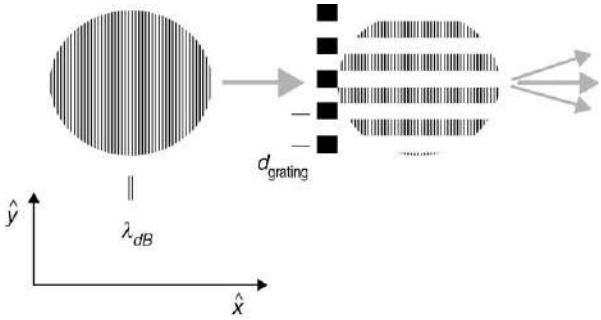

Figure 2 Schematic diagram of an atom wavepacket passing through an absorbing grating, as seen from above. The atom wavepacket is depicted before and after passing through the grating from left to right.

for matter waves and a different group velocity, v_n, for each k_n

$$\hbar\omega = \frac{\hbar^2 k_n^2}{2m} \qquad [13]$$

$$v_n = \frac{d\omega}{dk_n} = \frac{\hbar k_n}{m} = p_n/m \qquad [14]$$

Thus, diffraction through a grating puts an atom into a coherent superposition of different momenta, each separated by $\sim\hbar k_g\hat{y}$. Some time later, the atom will be in a coherent superposition of multiple locations, and the far-field probability density will be peaked at integer multiples of the diffraction angle

$$\Theta_D = \frac{k_g}{k_0} = \frac{\lambda_{dB}}{d_{\text{grating}}} \qquad [15]$$

where d_{grating} is the lattice spacing of the grating and the small-angle approximation for $\sin(\theta)$ has been used. Exactly as in light optics, these diffraction gratings are good momentum spectrometers, i.e., the diffraction angle is a measure of longitudinal momentum. (Contrary to light optics, group velocity can also be used to measure matter wave momentum.) Coherence between the diffraction orders remains until a perturbation occurs which would (even in principle) be sufficient to determine which path the atom took. This assertion has been tested with atoms in an interferometer (discussed below).

Atom flux in the nth diffraction order is given by

$$I_n = \left(\frac{\sin(n\pi\gamma)}{n\pi}\right)^2 \qquad [16]$$

To resolve the diffracted orders the initial transverse momentum distribution of the atoms must be smaller than k_g, and the time until the far-field observation must be sufficiently large given the beam width. At intermediate times, the atom waves are in the Fresnel diffraction region where the Talbot effect can be observed. This effect causes the atom waves to form a replica of the grating in free space at integer multiples of the Talbot length, $Z_T = 2d^2_{\text{grating}}\,\lambda_{dB}^{-1}$.

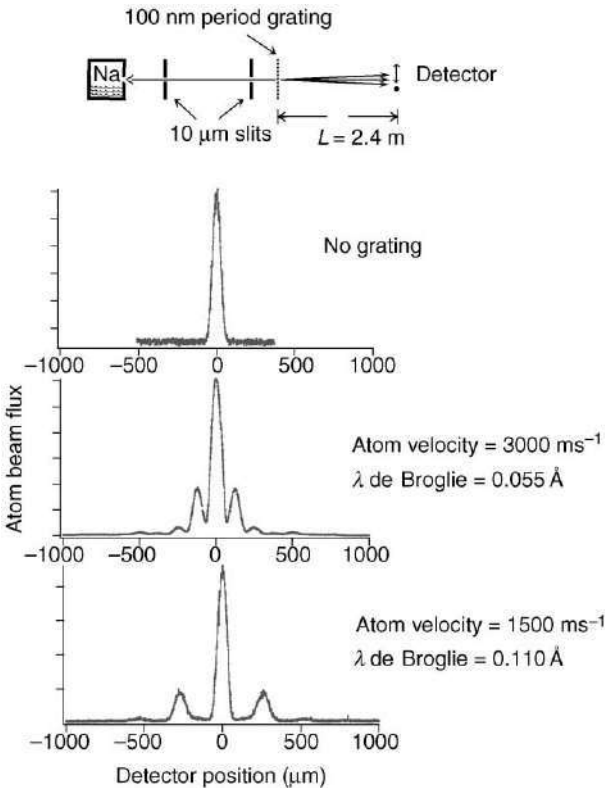

Figure 3 Diffraction data from a beam of sodium formed by passing atoms through two collimating slits followed by a nanofabricated diffraction grating. The momentum of the atoms determines their de Broglie wavelength, and thus their diffraction angle $\sin(\Theta_D) = \lambda_{dB}/d_{\text{grating}}$. These data were obtained at MIT by the authors.

Figure 3 shows the far-field diffraction pattern from a beam of sodium (^{23}Na) atoms using a material grating with a period of 100 nm. The atom beam velocity peaked at 1500 ms^{-1} corresponds to a de Broglie wavelength of 0.11 Å; thus, $\Theta_D \approx 10^{-4}$ rad. (The beam of 3000 ms^{-1} atoms diffracts at half this angle.) The spread in longitudinal velocity of $\sigma_v/v \approx 6\%$ can be deduced from the data because it slightly increases the width for higher order diffraction peaks. The open fraction of the grating, $\gamma \approx 60\%$, can also be determined from the diffraction data, because it determines the relative intensity of the orders. Effects of the atoms' finite size and interaction with the grating bars make small corrections to this simple theory.

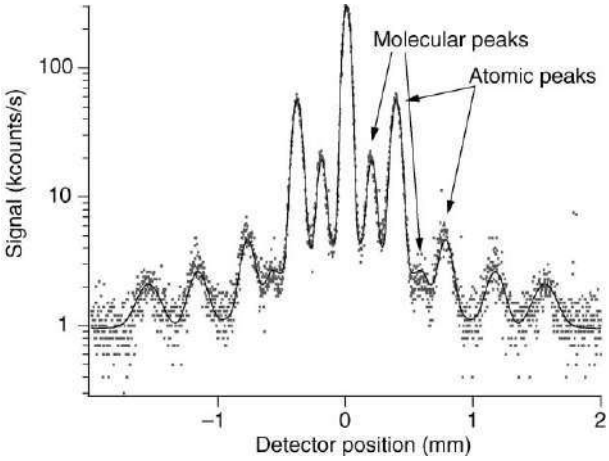

Figure 4 Diffraction data using a mixed beam of sodium atoms (Na) and sodium dimer molecules (Na_2). From this source molecules have the same speed as atoms, thus twice the momentum. Thus, molecule waves are diffracted at half the diffraction angle compared to atom waves. These data were obtained at MIT by the authors.

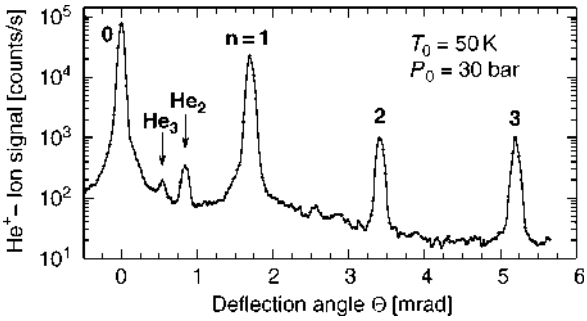

Figure 5 Diffraction of helium atoms and helium molecules through a nanofabricated grating. These data, reproduced with permission from Wieland Schoellkopf, were obtained at the Max-Planck-Institute in Göttingen.

Nanofabricated atom optics differ from those based on light forces in a number of ways: they are amplitude structures (with corresponding loss of transmission intensity), they are species-independent, their scale size can be several times smaller than attainable with light, and they can be arbitrarily patterned since they are fabricated by electron beam lithography. Insensitivity to species is demonstrated in Figure 4, where a beam of Na_2 molecules and Na atoms is diffracted and separated by a nanofabricated diffraction grating. The flexibility of electron beam lithography has allowed fabrication not only of diffraction gratings, but also of spherical and cylindrical zone plates, as well as a combination of lens and a hologram that generates a focused atom image. One stunning application laid to rest a long-standing argument concerning whether a stable bound state of the $^4He_2$ dimer exists. For this a diffraction grating was used to separate and resolve $^4He_2$ and more massive clusters (Figure 5). Subsequently a nano-sieve was used to estimate the size of the $^4He_2$ dimers to be 62 ± 10 Å.

Standing Waves of Light

In contrast to nanofabricated structures, standing waves of near-resonant light are phase gratings for atoms. The optical electric field dynamically polarizes the atoms and causes a shift to the ground state energy (the ac-Stark shift). For optical electric fields oscillating above (below) the atomic resonance frequency, the ac-Stark shift increases (decreases) potential energy for atoms and in either case perturbs the de Broglie wave phase.

In many respects, the interaction of an atom with a standing light wave is richer than the more familiar topic of light–atom interactions in a traveling light wave. Part of this richness reflects two ways in which a standing wave can be considered, either as two counter-propagating traveling waves or as a single, stationary standing wave. The standing wave–atom interaction is capable of transferring momentum in well-determined quantities, coherently splitting atomic wave packets, and generating forces much larger than possible with spontaneously scattered light. Not surprisingly, this interaction has many distinct facets that only appear in different regimes of the interaction parameters (intensity, detuning, and pulse duration) and atomic parameters (mass, initial momentum with respect to the standing wave, and excited state natural lifetime).

We begin with the Bragg scattering of atoms from a standing wave light grating. Although it can be difficult to realize the physical conditions that assure its occurrence, pure Bragg scattering is simpler than intermediate cases involving spontaneous decay (from light too close to resonance) or Kapitza–Dirac diffraction (from shorter interaction times).

Consider a standing wave light grating formed by two counter-propagating plane waves (traveling parallel to the z-axis) of equal amplitude, E_0, wavevector, k, frequency, ω, polarization vector, $\hat{e}$, and temporal envelope function, $f(t)$:

$$\vec{E}(z,t) = E_0 f(t)\sin(kz - \omega t)\hat{e} + E_0 f(t)\sin(kz + \omega t)\hat{e} \quad [17]$$

$$= 2E_0 f(t)\sin(kz)\cos(\omega t)\hat{e} \quad [18]$$

We would like to work with momentum states as our atomic basis, thus it is easiest to consider the

description of the electric field driving the transitions in terms of two counter-propagating traveling waves of definite momentum (eqn [17]), as opposed to the single standing wave they jointly form (eqn [18]).

Momentum is transferred by paired stimulated absorption and emission processes, resulting in a transfer of photons between the traveling waves. An N_B^{th} order diffraction process transfers N_B photons from one traveling wave to the counter-propagating traveling wave and changes the atomic momentum by $2N_B\hbar k\hat{z}$. Furthermore, atomic population is transferred only between $|g, -N_B\hbar k\rangle$ and $|g, +N_B\hbar k\rangle$, where $|g(e), \pm N\hbar k\rangle$ denotes a two-level atom in its ground (excited) state with momentum $\pm N\hbar k$ parallel to the standing wave axis. The excited state remains nominally unpopulated so long as the temporal envelope function, $f(t)$, does not have strong frequency components near the laser detuning

$$\delta = \omega - \omega_0 \qquad [19]$$

where ω_0 is the unperturbed frequency of the atomic transition. Furthermore, for a given initial state $(|g, -N_B\hbar k\rangle)$ the uniqueness of the final state $(|g, +N_B\hbar k\rangle)$ occurs because of the fundamental assumption that makes Bragg scattering so simple to describe; the uncertainty in the photon energy driving the transitions is small compared to the energy separation between neighboring momentum states. A quantitative discussion of the validity of this assumption will be given later.

We now calculate the probability, $P_1^B(\tau)$, of the first order $(N_B = 1)$ Bragg process. Scattering transfers an atom from $|g, -\hbar k\rangle$ to $|g, +\hbar k\rangle$ when the atoms interact with a constant light intensity for a time τ (i.e., in eqns [17] and [18] $f(t)$ is a square wave of unit amplitude and duration τ). The geometry for Bragg diffraction from a standing wave of light is shown in Figure 6, and the energy/momentum states involved in the transition are depicted in Figure 7.

Considering the electronic degree of freedom, we rewrite the Schrödinger equation as

$$(\mathcal{H}_0 + \mathcal{H}_{\text{int}})|\psi\rangle = i\hbar\frac{\partial}{\partial t}|\psi\rangle \qquad [20]$$

where the total Hamiltonian $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{\text{int}}$ is a matrix. The wavefunction is a vector of coefficients c_i in the basis $\{|g, -\hbar k\rangle, |e, 0\rangle, |g, +\hbar k\rangle\}$, with normalized population amplitudes for each component:

$$|\psi\rangle = \begin{pmatrix} c_{-1}(t) \\ c_0(t) \\ c_{+1}(t) \end{pmatrix} \qquad [21]$$

In the electric dipole approximation, the interaction Hamiltonian is $\mathcal{H}_{\text{int}}(t) = -\vec{\mu}\cdot\vec{E}(t)$, where $\mu = \langle e|q\vec{r}|g\rangle\cdot\hat{e}$ is the electric dipole matrix element connecting the ground $(|g\rangle)$ and excited $(|e\rangle)$ states

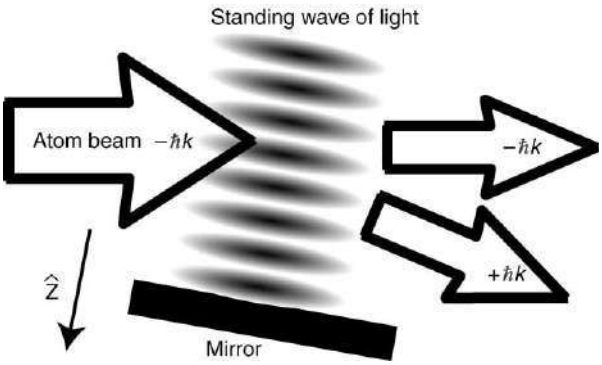

Figure 6 Two counter-propagating running waves of light superimpose to form a standing wave of light. This 'light crystal' acts as a diffraction grating for atoms, shown here in the configuration for Bragg diffraction.

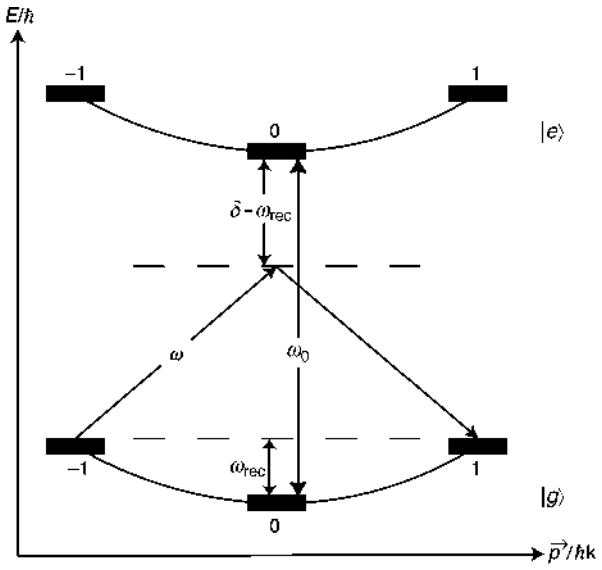

Figure 7 Energy vs. momentum diagram describing atomic states involved in Bragg diffraction. The total energy is due to the electronic state $(|g\rangle, |e\rangle)$ plus the kinetic energy associated with the recoil from photon emission (or absorption). In Bragg diffraction, atoms make a coherent transition from $|g, -\hbar k\rangle \leftrightarrow |g, +\hbar k\rangle$. Reprinted from Gupta S, *et al.* (2001) Coherent manipulation of atoms with standing light waves. *Comptes Rendus de L'Academie des Sciences – Series IV – Physics-Astrophysics* 2:479–495, copyright (2001) with permission from Elsevier.

of the atom. $\hat{e}$ is the polarization vector of the light and q is the charge of the electron. By momentum conservation, only the plane wave traveling in the $\pm\hat{z}$ direction couples the $|e,0\rangle \leftrightarrow |g,\mp\hbar k\rangle$ transitions. By using this argument, we are effectively viewing the electric field as a quantum mechanical operator. Expanding the sinusoidal variation of the electric field in terms of complex exponentials and treating the spatially dependent complex exponential terms as quantum mechanical momentum translation operators yields the interaction Hamiltonian:

$$\mathcal{H}_{\text{int}} = \frac{\hbar\omega_R}{2}\begin{pmatrix} 0 & ie^{i\omega t} & 0 \\ -ie^{-i\omega t} & 0 & ie^{-i\omega t} \\ 0 & -ie^{i\omega t} & 0 \end{pmatrix} \qquad [22]$$

For example, $e^{\pm ik\hat{z}}|g,n\hbar k\rangle = |e,(n\pm 1)\hbar k\rangle$ describes absorption, and for stimulated emission, $e^{\mp ik\hat{z}}|e,n\hbar k\rangle = |g,(n\pm 1)\hbar k\rangle$. This procedure includes the rotating wave approximation. In addition, we have neglected any frequency components associated with the sudden switch on of the fields and the finite duration of the light–atom interaction. In formulating $\mathcal{H}_{\text{int}}$, the strength of the light–atom interaction is parameterized by the single-photon Rabi frequency:

$$\omega_R = \frac{\mu E_0}{\hbar} \qquad [23]$$

Without loss of generality, we will take μ and hence ω_R to be positive, real-valued quantities.

The total Hamiltonian follows simply by including the electronic and kinetic energy terms:

$$\mathcal{H}_0 = \hbar\begin{pmatrix} \omega_{\text{rec}} & 0 & 0 \\ 0 & \omega_0 & 0 \\ 0 & 0 & \omega_{\text{rec}} \end{pmatrix} \qquad [24]$$

where the single-photon recoil energy, E_{rec}, of an atom of mass m is given by

$$E_{\text{rec}} = \hbar\omega_{\text{rec}} = \frac{\hbar^2 k^2}{2m} \qquad [25]$$

Making the *ansatz* for the solution wavefunction as

$$|\psi\rangle = \begin{pmatrix} c_{-1}(t)e^{-i\omega_{\text{rec}}t} \\ c_0(t)e^{-i\omega_0 t} \\ c_{+1}(t)e^{-i\omega_{\text{rec}}t} \end{pmatrix} \qquad [26]$$

and substituting into the Schrödinger equation yields three coupled first-order differential equations:

$$\dot{c}_{\pm 1}(t) = \mp\frac{\omega_R}{2}e^{i\Delta t}c_0(t), \qquad [27]$$

$$\dot{c}_0(t) = \frac{\omega_R}{2}e^{-i\Delta t}\big(c_{+1}(t) - c_{-1}(t)\big), \qquad [28]$$

where $\Delta = \delta + \omega_{\text{rec}} = (\omega - \omega_0) + \omega_{\text{rec}}$. Differentiating eqn [27] and substituting eqn [28] into the result yields two coupled second-order differential equations:

$$\ddot{c}_{\pm 1}(t) - i\Delta\dot{c}_{\pm 1}(t) + \frac{\omega_R^2}{4}\big(c_{\pm 1}(t) - c_{\mp 1}(t)\big) = 0 \qquad [29]$$

With the initial conditions:

$$c_{-1}(0) = 1 \qquad [30]$$

$$c_0(0) = 0 \Rightarrow \dot{c}_{\pm 1}(0) = 0 \qquad [31]$$

$$c_{+1}(0) = 0 \qquad [32]$$

and the assumption $\Delta^2 \gg \omega_R^2$, the solutions to eqn [29] are

$$c_{-1}(t) = e^{-i\frac{\omega_R^{(2)}}{2}t}\cos\left(\frac{\omega_R^{(2)}}{2}t\right) \qquad [33]$$

$$c_{+1}(t) = ie^{-i\frac{\omega_R^{(2)}}{2}t}\sin\left(\frac{\omega_R^{(2)}}{2}t\right) \qquad [34]$$

where the two-photon Rabi frequency is

$$\omega_R^{(2)} = \frac{\omega_R^2}{2\Delta} \rightarrow \frac{\omega_R^2}{2\delta}, \quad |\delta| \gg \omega_{\text{rec}} \qquad [35]$$

with both transitions driven at equal single-photon Rabi frequencies, ω_R.

Substituting the solution of eqn [33] or [34] into eqn [27] yields an expression for the excited state amplitude:

$$c_0(t) = -i\frac{\omega_R}{2\Delta}e^{-i\Delta t}e^{-i\omega_R^{(2)}t} \qquad [36]$$

This will be important in calculating the rate of spontaneous emission events later.

The solutions for $c_{-1}(t)$ and $c_{+1}(t)$ oscillate with the interaction duration, τ, yielding the result for the $|g,-\hbar k\rangle \rightarrow |g,+\hbar k\rangle$ transition probability:

$$P_1^B(\tau) = |c_{+1}(\tau)|^2 = \sin^2\left(\frac{\omega_R^{(2)}}{2}\tau\right) \qquad [37]$$

Thus, the system oscillates between the two momentum states $|g,-\hbar k\rangle$ and $|g,+\hbar k\rangle$ in a manner analogous to the Rabi oscillation of atomic population between two resonantly coupled states. This solution with

oscillatory probabilities for the two Bragg coupled states is known as the Pendellösung and has been observed for atoms, neutrons, and X-rays (Figure 8). The nice feature here is that the strength of the grating can be actually controlled by the intensity of the light.

Viewing Bragg scattering as a two-photon transition from the initial ground state to the final ground state with opposite momentum, illuminates the close connection with a Raman transition between two internal sub-states of the ground state manifold, each with its own external momentum state. The formalism describing the Raman transition is basically the same as that presented here, except the two transitions can be driven at different single-photon Rabi frequencies, ω_{R1} and ω_{R2}, so that the generic two-photon Rabi frequency is given by $\omega_R^{(2)} = \omega_{R1}\omega_{R2}/2\Delta$, where Δ is the detuning from the intermediate state.

An N_B^{th} order Bragg process (similar to a $2N_B$-photon Raman process) is a coherent succession of N_B two-photon transitions from $|g, -N_B\hbar k\rangle$ to $|g, +N_B\hbar k\rangle$ with $2N_B - 1$ intermediate states of the form:

$$|e, (-N_B + 1)\hbar k\rangle, |g, (-N_B + 2)\hbar k\rangle \ldots$$

$$\ldots |g, (N_B - 2)\hbar k\rangle, |e, (N_B - 1)\hbar k\rangle$$

Such a process is characterized by a $2N_B$-photon Rabi frequency given by

$$\omega_R^{(2N_B)} = \frac{[\omega_R]^{2N_B}}{2^{2N_B-1}\Delta_1\Delta_2\cdots\Delta_{2N_B-1}} \qquad [38]$$

where Δ_n is the detuning from the nth intermediate state. Figure 9 shows what this process would look like for an N_B^{th} order Bragg transition where the intermediate state detunings are given by

$$\Delta_n = \left\{ \begin{array}{ll} \delta + (2N_B n - n^2)\omega_{\text{rec}} : & n \text{ odd} \\ (2N_B n - n^2)\omega_{\text{rec}} : & n \text{ even} \end{array} \right\} \qquad [39]$$

Substituting these detunings into eqn [38] yields the N_B^{th} order Bragg transition $2N_B$-photon Rabi frequency, $\omega_R^{(2N_B)}$:

$$\omega_R^{(2N_B)} = \frac{[\omega_R]^{2N_B}}{2^{4N_B-3}[(N_B-1)!]^2\delta^{N_B}\omega_{\text{rec}}^{N_B-1}} \qquad [40]$$

where we have assumed $|\delta| \gg N_B^2\omega_{\text{rec}}$.

To ensure that the system truly undergoes Bragg scattering, and validate the assumption that only states of equal kinetic energy and opposite

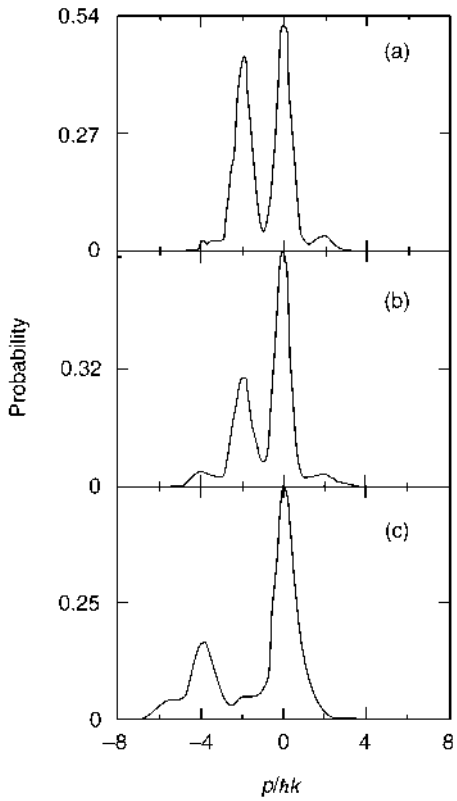

Figure 8 Bragg scattering of an atom beam from an optical standing wave. First-order Bragg scattering at (a) lower power, (b) higher power of the laser beams (Pendellösung has set in, increasing the amplitude of the undeflected peak), (c) second-order Bragg scattering. Reprinted with permission from *Physical Review Letters* 60(6): 515–518. Copyright (1998) by the American Physical Society.

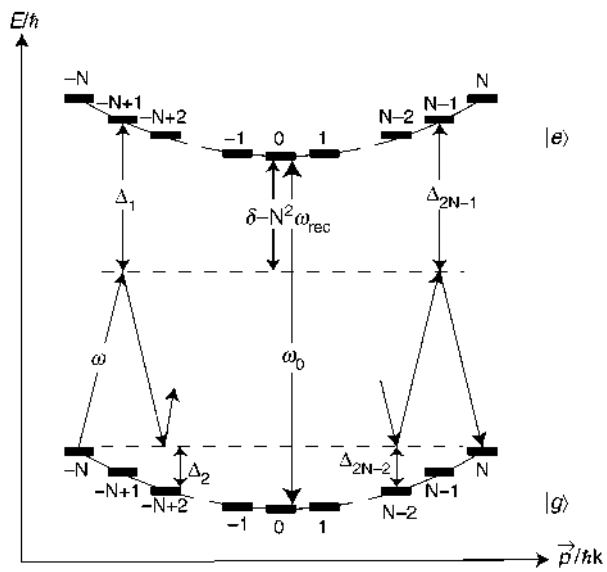

Figure 9 Energy vs. momentum diagram describing atomic states involved in higher-order Bragg diffraction. Reprinted from Gupta S, *et al.* (2001) Coherent manipulation of atoms with standing light waves. *Comptes Rendus de L'Academie des Sciences – Series IV – Physics-Astrophysics* 2:479–495, copyright (2001) with permission from Elsevier.

momentum are coupled, the overall exposure time, τ, of the atoms to the fields must be limited both from below and above.

The lower bound on τ is necessary to resolve the final momentum state $(|g, N_B\hbar k\rangle)$ from neighboring momentum states $(|g, (N_B \pm 2)\hbar k\rangle)$ two photon recoil momenta away. Avoiding population transfer into these states requires that the energy separation be resolvable, or

$$\tau(\delta E) \gg h \qquad [41]$$

For first-order Bragg scattering processes ($N_B = 1$) the nearest lying momentum states that may be mistakenly populated are $|g, \pm 3\hbar k\rangle$, for which

$$\delta E = \frac{p_{n=3}^2}{2m} - \frac{p_{n=1}^2}{2m} = 8\hbar\omega_{\text{rec}} \qquad [42]$$

Therefore

$$\tau \gg \frac{h}{8\hbar\omega_{\text{rec}}} = \frac{\pi}{4\omega_{\text{rec}}} \qquad [43]$$

For sodium atoms (^{23}Na) in 590 nm light, $\tau \gg 6\ \mu$s is the lower bound required for first-order Bragg scattering.

For all higher-order ($N_B > 1$) Bragg scattering processes, the nearest momentum states are $|g, \pm(N_B - 2)\hbar k\rangle$, which limits the interaction time to

$$\tau \gg \frac{\pi}{2(N_B - 1)\omega_{\text{rec}}} \qquad [44]$$

which for $N_B \gg 1$ reduces to

$$\tau \gg \frac{\pi}{2N_B\omega_{\text{rec}}} \qquad [45]$$

The upper bound on the interaction duration is necessary to avoid spontaneous emission. The interaction duration must be short enough so that the expected number, N_s, of spontaneous emission events per atom during the time τ is negligible. N_s is simply given by the product of the excited state fraction (eqn [36]) and the probability of spontaneous decay given that the atom is in the excited state:

$$N_s = |c_0(t)|^2\Gamma\tau = \frac{\omega_R^2}{4\Delta^2}\Gamma\tau \qquad [46]$$

where Γ is the natural decay rate of the excited state. Avoiding spontaneous emission ($N_s \ll 1$) while still having a significant probability for transitions ($\omega_R^{(2N_B)}\tau \simeq \pi$) forces one to work in the regime where $\Delta \gg \Gamma$, which is a practical requirement to avoid spontaneous emission and maintain coherence. For sodium atoms on the 590 nm transition, $\Gamma = 6.3 \times 10^7\ \text{s}^{-1}$, and a typical choice in Bragg diffraction experiments is $\Delta = 50\Gamma$, or $\Delta/2\pi = 500$ MHz.

Briefly, to relate Rabi frequency to optical intensity, the transition rate W on resonance with a Lorentzian spectral line is given by Fermi's golden rule:

$$W = \frac{\omega_R^2}{\Gamma} = \frac{\lambda_{\text{photon}}^2}{2\pi}\frac{I}{\hbar\omega} \qquad [47]$$

where the last term can be regarded as the cross section for absorption multiplied by intensity in units of photon flux. Hence

$$\omega_R^2 = \Gamma\frac{\lambda_{\text{photon}}^2}{2\pi}\frac{I}{\hbar\omega} \qquad [48]$$

For Na atoms in 12 mW(cm)$^{-2}$ resonant light, $\omega_R = \Gamma = 6.3 \times 10^7\ \text{s}^{-1}$. Both the probability of Bragg scattering and the spontaneous emission increase linearly with $\omega_R^2\tau\Gamma^{-1}$. However, spontaneous emission decreases as Δ^2 while the Bragg scattering probability decreases as Δ. Thus, as shown in Figure 10, at sufficient detuning and laser power, Bragg scattering occurs with no spontaneous emission. For reference, 12 mW(cm)$^{-2} \times 10\ \mu$s corresponds to $\omega_R^2\tau\Gamma^{-1} = 630$ for sodium atoms.

Bragg scattering of atoms from a standing light wave was first observed at MIT in 1988. A supersonic atomic beam was diffracted from a standing wave of near-resonant laser light as depicted in Figure 6. The angle between the atomic beam (of thermal wavelength λ_{dB}) and the light grating (of periodicity $\lambda_L/2$) was tuned to the appropriate Bragg angle, θ_B, where:

$$\lambda_{dB} = \lambda_L \sin(\theta_B) \qquad [49]$$

Population transfer corresponding to both first- and second-order Bragg scattering was observed (Figure 7). The experiment required a sub-recoil transverse momentum spread of the atomic beam in order to resolve the different momentum states in the far field and limit the final state to only one

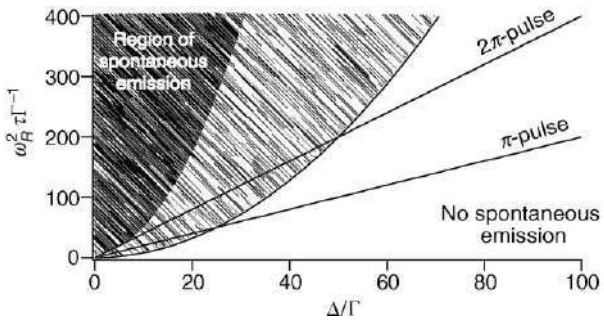

Figure 10 The conditions for which N_s >0.02(0.10) are hashed in light(dark) gray. ω_R^2/Γ is proportional to optical intensity. The line labeled π-pulse indicates the parameters which cause complete first-order Bragg diffraction. The line labeled 2π-pulse indicates conditions where population is completely transferred back to the initial state (Pendellösung oscillation).

diffracted order. The Pendellösung was observed as an oscillation in population transfer as a function of standing wave intensity, $I \propto \omega_R^{(2)}$, for a fixed interaction time, τ.

Atomic beam diffraction from an optical standing wave is a continuous-wave (CW) experiment in which the selectivity needed for the Bragg process is imposed by good angular resolution of the particle beam and a high degree of parallelism between the light crystal planes. This ensures that of the various final Bragg orders allowed by momentum conservation, only one conserves energy (energy conservation is exact in a CW experiment). For atoms scattering from a light crystal, parallelism of the crystal planes requires highly parallel photon momentum that implies a minimum width of the standing wave (the diffraction limit for the collimated photons). The transit time, τ, of the atoms across this width then exceeds the lower bound given in eqn [44].

The excellent collimation required of the atomic beam to ensure resolution of the Bragg scattered atoms reduces the intensity of the source by many orders of magnitude. A Bose–Einstein Condensate (BEC) is an attractive alternative source of atoms because its momentum spread is typically an order of magnitude below a single-photon recoil momentum. To Bragg diffract atoms initially in a stationary BEC, it is easier to move the light crystal than to accelerate the condensate. This is done by simply frequency shifting one of the traveling waves so that the resultant standing wave formed by its interference with the unshifted traveling wave moves with the proper velocity ($N_B\hbar k/m$) relative to the stationary atoms to impart the necessary momentum. The Bragg scattered atoms will then have momentum $2N_B\hbar k$ in the laboratory frame. The resonance condition thus becomes a condition on relative detuning, δ_{N_B}, between the two laser beams forming the diffraction grating. For N_B^{th}-order Bragg diffraction, the relative detuning is given by

$$\delta_{N_B} = \frac{2N_B\hbar k^2}{m} = 4N_B\omega_{\text{rec}} \qquad [50]$$

The first demonstration of Bragg scattering in a BEC was at NIST in 1999. They used Bragg scattering mainly as a tool to manipulate the momentum of the BEC, observing up to sixth-order processes. At MIT the interaction time was lengthened (to ≈100 times the lower bound of eqn [44] for first-order Bragg scattering), creating a new type of spectroscopy called Bragg Spectroscopy. It is a spectroscopic measurement of the shift of the Bragg scattering frequency due to the Doppler shift ($k \cdot v$) from the atoms' motion together with any mean field interaction. (These effects can be separated by going to higher-order Bragg scattering to enhance the Doppler shift.)

Bragg Spectroscopy was used to observe the momentum distribution of a BEC in a magnetic trap. The width of the Bragg resonance curve was primarily due to a 2 kHz Doppler-broadening that yielded the momentum distribution of the condensate. The spread in the corresponding velocity distribution was very small ($\approx 0.5\ \text{mm s}^{-1}$), even smaller than allowed by the Heisenberg uncertainty limit for a particle confined in the ground state of a harmonic trap. This reflects the increase in size of the condensate due to the mean field repulsion (^{23}Na, used in the MIT experiment, has a positive scattering length). The distribution was Heisenberg uncertainty limited at the observed size of the BEC, establishing for the first time that the coherence length of the condensate was equal to its size. In addition, the narrow Bragg resonance was shifted by the repulsive interactions within the condensate, resulting in a spectroscopic measurement of the mean-field energy.

Kapitza–Dirac Scattering

Diffraction of neutral atoms from a standing wave of near-resonant light with a short interaction time ($\tau_{KD} \ll 1/\omega_{\text{rec}}$) has come to be called Kapitza–Dirac scattering, in honor of their pioneering suggestion. In 1933, Kapitza and Dirac predicted that an electron beam propagating in a standing light wave would undergo stimulated Compton scattering and be reflected. This process has a tiny cross-section, equal to the classical electron radius squared: $\sigma_{\text{Compton}} = 8\pi/3(e^2/mc^2)^2 \approx 6 \times 10^{-25}\ \text{cm}^2$, and has only recently been observed using extremely high laser intensities. If the electrons are replaced by atoms, however, the scattering cross-section for resonant light $\sigma_{\text{atom}} = (1/2\pi)\lambda_{\text{photon}}^2 \approx 4 \times 10^{-10}\ \text{cm}^2$ is 15 orders of magnitude larger. This large cross-section, together with the ready availability of tunable lasers, has allowed stimulated scattering, both in the Kapitza–Dirac and Bragg regimes, to become the primary tool for the coherent manipulation of atoms.

In Kapitza–Dirac scattering, atomic motion during the interaction time is small compared to the characteristic dimensions of the interaction potential. This is equivalent to the eikonal approximation for scattering or the thin-lens approximation in optics. The idea is that the phase of the incident particle changes along each classical trajectory, but not the amplitude. Thus, there may be momentum transfer perpendicular to the trajectory, but the trajectory is not significantly displaced (until after the interaction is over). Mathematically, this regime

can be treated by neglecting the atomic kinetic energy term in the Hamiltonian (the Raman–Nath approximation).

The standing wave interaction may be treated by considering the standing wave (ac-Stark shift) potential resulting from the applied fields given in eqn [18]:

$$V(z,t) = \frac{\hbar\omega_R^2}{\delta} f^2(t)\sin^2(kz) \qquad [51]$$

where we have assumed $\delta \gg \Gamma$, because the constraint $N_s \ll 1$ (eqn [46]) holds in this regime as well, where N_s is the expected number of spontaneous emission events per atom during the time τ. Although we neglect the kinetic energy term in the Hamiltonian, we continue to will treat z in eqn [51] as an operator.

Given the initial atomic wavefunction in momentum space, $\psi(p_0)$, the atomic wavefunction immediately after the interaction is given by

$$\psi(p) = e^{-\frac{i}{\hbar}\int^t dt' U(z,t')}\psi(p_0) \qquad [52]$$

$$= e^{-i\frac{\omega_R^2}{2\delta}\tau} e^{i\frac{\omega_R^2}{2\delta}\tau\cos(2kz)}\psi(p_0) \qquad [53]$$

where $\tau = \int dt' f^2(t')$ and the integral is over the interaction duration. With the use of the identity for Bessel functions of the first kind, $e^{i\alpha\cos(\beta)} = \sum_{n=-\infty}^{\infty} i^n J_n(\alpha) e^{in\beta}$, the atomic wavefunction can be written as

$$\psi(p) = e^{-i\frac{\omega_R^2}{2\delta}\tau} \sum_{n=-\infty}^{\infty} i^n J_n\left(\frac{\omega_R^2}{2\delta}\tau\right) e^{i2nkz}\psi(p_0) \qquad [54]$$

$$= e^{-i\frac{\omega_R^2}{2\delta}\tau} \sum_{n=-\infty}^{\infty} i^n J_n\left(\frac{\omega_R^2}{2\delta}\tau\right)\psi(p_0 + 2n\hbar k) \qquad [55]$$

States with $2N\hbar k$ of momentum are populated with the probability:

$$P_N = J_N^2(\theta), \quad N = 0, \pm1, \pm2, \ldots \qquad [56]$$

where

$$\theta = \frac{\omega_R^2}{2\delta}\tau = \omega_R^{(2)}\tau \qquad [57]$$

is the pulse area. This leads to a transverse rms momentum of the diffracted atoms that is linearly proportional to the pulse area:

$$p_{\text{rms}} = \sum_{n=-\infty}^{\infty} (n\hbar k)^2 P_n = 2^{1/2}\theta\hbar k \qquad [58]$$

The maximum and minimum bounds on the interaction time, τ, and the amount of momentum transfer to the atoms are discussed in the next section.

Kapitza–Dirac diffraction of atoms was first observed at MIT in 1986. Diffraction of a well-collimated (subrecoil) supersonic atomic beam was observed after passage through the tightly focused waist of a near-resonant standing wave (Figure 11). Significant diffraction into momentum states $|g, \pm10\hbar k\rangle$ was observed (Figure 12).

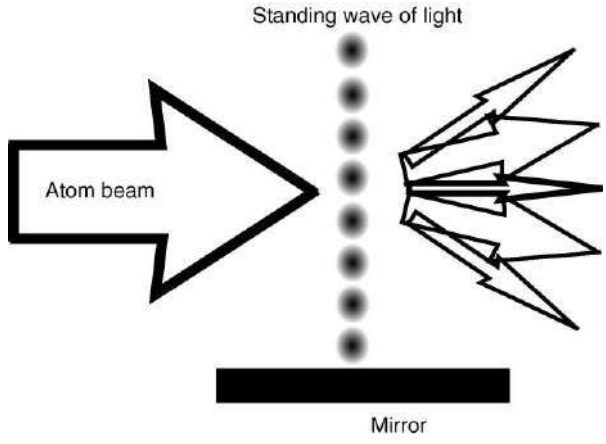

Figure 11 Two counter-propagating running waves of light superimpose to form a standing wave of light. The more narrow beam shown here represents the configuration for Kapitza–Dirac diffraction.

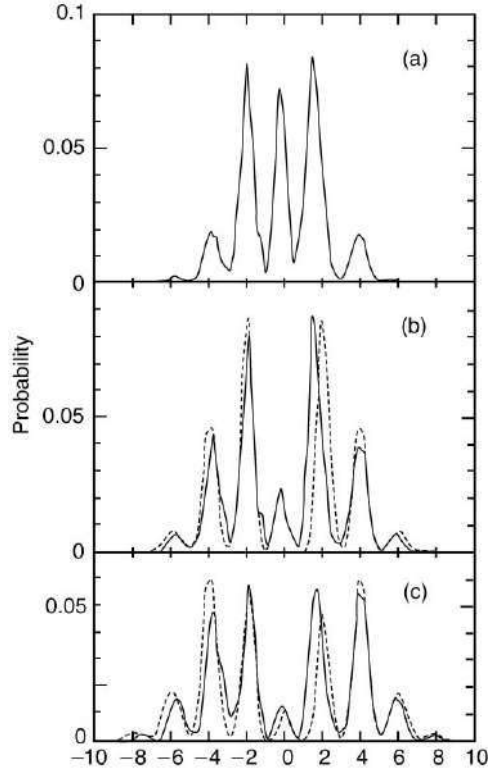

Figure 12 Kapitza–Dirac diffraction of an atomic beam from a standing wavelight grating. The solid lines are experimental data, the dashed lines are theoretical curves. (a) $\theta = 1.69$, (b) $\theta = 2.33$, (c) $\theta = 2.84$. Reprinted with permission from *Physical Review Letters* 56(8): 827–830. Copyright (1986) by the American Physical Society.

Even higher diffracted orders should be observable in the future using laser beams directed at small BECs for somewhat longer times.

Comparison of Atom-Standing Wave Interactions

The atomic motion induced by interaction with a standing light wave varies qualitatively with the interaction parameters. Figure 13 summarizes the various interaction regimes of an atom with a standing wave in terms of the two most basic parameters: the two-photon Rabi frequency, $\omega_R^{(2)} = \omega_R^2/2\delta$, and the duration of the interaction, τ. To render the plot independent of atomic species, $\omega_R^{(2)}$ is given in units of the single-photon recoil frequency, ω_{rec}, and τ is given in units of the inverse single-photon recoil frequency, ω_{rec}^{-1}. Thus, for ^{23}Na the point (1,1) on Figure 13 corresponds to (6.4 μs, $2\pi \times 25$ kHz). The product of the two-photon Rabi frequency and the duration of the interaction is the pulse area, $\theta = \omega_R^{(2)}\tau$ (eqn [57]), which is likewise the product of the pair of coordinates forming a point on the plot.

The lines KD1(KD10) on Figure 13 show where the first maxima of $P_{1,10} = J_{1,10}^2(\theta)$ (eqn [56]) occur, corresponding to the maximum possible population transfer into the first and tenth Kapitza–Dirac diffracted order respectively. Since the momentum distribution of diffracted atoms depends only on the pulse area (θ), all Kapitza–Dirac orders are parallel to each other and further offset from the origin in increasing order number.

The Kapitza–Dirac regime ends at large interaction times where the Raman–Nath approximation fails due to motion of the atoms down the slope of the standing wave potential. We show these lines dashed as the interaction time approaches the beginning of the classical oscillation regime, and terminate these lines where the oscillation produces its first focus, at the first focus line. This line corresponds to the atoms completing a quarter period of oscillation in the standing wave potential, $\tau = \tau_{\text{osc}}/4$, where the oscillation time is derived from the approximation that the potential $\langle V(z,t)\rangle_t$ (eqn [51]) is harmonic

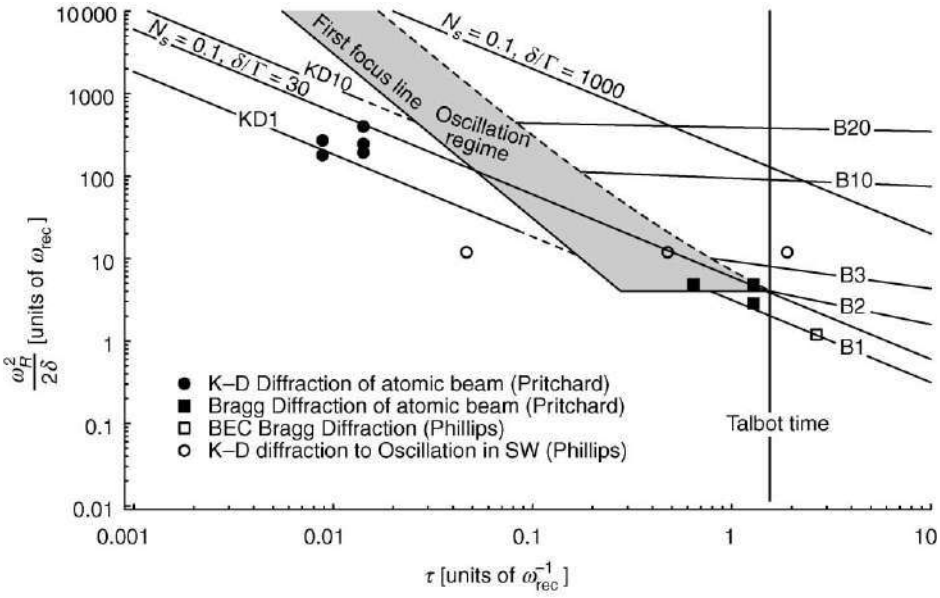

Figure 13 Atomic diffraction from a standing light wave. The vertical axis is the two-photon Rabi frequency, $\omega_R^2/2\delta$, in units of the single-photon recoil frequency, ω_{rec}, and the horizontal axis is the pulse duration, τ, in units of the inverse single-photon recoil frequency, ω_{rec}^{-1}. The scaling is chosen to eliminate the atomic species dependence of the plot. All coherent momentum transfer processes are destroyed by spontaneous decay, which occurs with probability $N_s = 0.1$ along the lines labeled accordingly and parameterized by the given ratio δ/Γ. Lines KD1 and KD10 show conditions for the maximum transfer into the first and tenth Kapitza–Dirac diffracted order respectively. As the interaction time is increased the Raman–Nath approximation is violated (termination of Kapitza–Dirac curves) and the atoms enter the oscillatory regime, executing at least a quarter period of oscillation above the first focus line (shaded area). Cures B1, B2, B3, B10, and B20 correspond to conditions that generate complete Bragg reflection in the first-, second-, third-, tenth-, and twentieth-order respectively. Experimental conditions are shown as points. Filled circles: Kapitza–Dirac diffraction of an atomic beam (Pritchard 1986). Filled squares: Bragg diffraction of an atomic beam (Pritchard 1988). Open squares: Bragg diffraction of a BEC (Phillips 1999). Open circles: transition from Kapitza-Dirac diffraction to oscillation of a BEC in a standing wave light pulse (Phillips 1999). Reprinted from Gupta S, *et al.* (2001) Coherent manipulation of atoms with standing light waves. *Comptes Rendus de L'Academie des Sciences – Series IV – Physics-Astrophysics* 2:479–495, copyright (2001) with permission from Elsevier.

about the minimum:

$$\tau_{\text{osc}} = \frac{\pi}{\omega_R}\left(\frac{|\delta|}{\omega_{\text{rec}}}\right)^{1/2} \qquad [59]$$

In the shaded region of Figure 13, the atoms will oscillate classically about the potential minimum, causing a periodic focusing of the atoms alternately in position and momentum space. However, anharmonicities of the potential away from its minimum will degrade the quality of the focusing effects.

The maximum diffracted order that can be significantly populated by the light–atom interaction is limited by energy conservation. Classically, the maximum momentum transfer to the atoms due to the sudden switch on of the standing wave is delivered to atoms that convert the full height of the standing wave potential ($\hbar\omega_R^2/\delta$), into kinetic energy. This gives

$$\frac{p_{\max}^2}{2m} \simeq \frac{\hbar\omega_R^2}{\delta} \Rightarrow p_{\max} \simeq \omega_R\left(\frac{2\hbar m}{\delta}\right)^{1/2} \qquad [60]$$

Equating the maximum absorbed momentum, $p_{\max}$, to an integer number of two-photon recoil momenta yields the maximum expected diffracted order, $N_{\max}$:

$$p_{\max} = 2N_{\max}\hbar k \Rightarrow N_{\max} \simeq \frac{\omega_R}{2}(\omega_{\text{rec}}\delta)^{-1/2} \qquad [61]$$

This is just half of the square-root of the standing wave potential height measured in units of the single-photon recoil energy:

$$N_{\max} = \frac{1}{2}\left(\frac{\frac{\omega_R^2}{\delta}}{\omega_{\text{rec}}}\right)^{1/2} = \frac{1}{2}\left(\frac{\omega_R^{(2)}}{\omega_{\text{rec}}}\right)^{1/2} \qquad [62]$$

As the two-photon Rabi frequency is reduced, $N_{\max}$ falls below unity. As a result, there is no longer time for the higher-order multi-photon processes to generate significant amplitudes in diffracted orders with $N > 1$ before dephasing due to the kinetic energy term (eqn [24]) becomes significant and the higher-order processes become negligible. Only a small population is ever transferred to the $N = \pm 1$ orders and it does not oscillate at ω_{rec}. Therefore the classical oscillation regime does not extend below $N_{\max} \simeq 1$ where it is terminated on Figure 13.

Classical oscillation would result in atoms with momentum from zero up to the maximum allowed by energy conservation. Therefore, the classical focusing of atoms must be manifest as an atomic population distribution over many of the quantized momentum states (separated by $2\hbar k$) allowed by energy conservation (eqn [61]). However, as the interaction time lengthens and extends into the Bragg regime, the populated momentum states become restricted by energy conservation until only the oscillatory Pendellösung into and back from only one final state remains. Therefore, we have ended the classical oscillation regime where the Bragg condition (eqn [44]) is satisfied. The Bragg regime presupposes a smooth light pulse shape. For a pulse with sharp edges, classically oscillatory behavior can still be observed at longer times than included in the shaded region of Figure 13, which is why we show the large τ boundary of the classical oscillation regime as dashed.

In the Bragg regime, transfer of population is restricted to (and back from) only one final momentum state. The allowable final states are restricted by limiting the frequency bandwidth (i.e., energy uncertainty) of the light fields in the atomic rest frame. This is accomplished by lengthening the interaction time and smoothing the rise and fall of the electromagnetic fields. The parameters for a first-order Bragg transition (Table 1) are typically $\delta/\Gamma \simeq 50$, $\Gamma\tau \simeq 1$, and $\omega_R/\Gamma \simeq 10$ giving $N_s \simeq 10^{-2}$ and $\omega_R^{(2)}\tau \simeq 1$. Obtaining significant population transfer with higher-order Bragg processes requires larger intensities. Various orders (1, 2, 3, 10, and 20) of Bragg diffraction are shown as lines on Figure 13 corresponding to $\omega_R^{(2N_B)}\tau = \pi$, where $\omega_R^{(2N_B)}$ is given in eqn [40]. The lines are terminated at the appropriate interaction time determined from the final momentum state resolution condition (eqn [44]). The Bragg regime extends indefinitely to larger interaction times, which might be termed the region of Bragg spectroscopy. In experiments in this regime with ^{23}Na, atomic velocity resolution below 1 mm s^{-1} was obtained at $\tau = 80\omega_{\text{rec}}^{-1}$. The study of adiabatically expanded BECs would require larger interaction times to resolve their smaller velocity spread and weaker mean field shifts.

The Kaptiza–Dirac and Bragg regimes assume a different initial atomic momentum (in the rest frame of the standing wave) parallel to the standing

Table 1 Natural parameters and typical experimental parameters involved in standing wave diffraction. The relevant frequencies together with the corresponding times ($1/\omega$) are tabulated. The parameters for ^{23}Na have been used for system dependent quantities

ω	$\omega/2\pi$	$1/\omega$
ω_0	500 THz	0.3 fs
δ (Bragg)	500 MHz	0.3 ns
ω_R (Bragg)	100 MHz	1.6 ns
ω_R (K–D)	100 MHz	1.6 ns
Γ	10 MHz	16 ns
ω_{rec}	25 kHz	6.4 μs
ω_{osc}	1 MHz	160 ns

wave axis. Kapitza–Dirac scattering assumes no component of the initial atomic momentum along this axis. The efficiency of Kapitza–Dirac diffraction falls rapidly as the initial momentum in units of the photon momentum approaches $1/\omega_{\rm rec}\tau$. To observe N_B^{th}-order Bragg scattering, the initial atomic momentum along the standing wave axis must have (nonzero) magnitude $N_B\hbar k$. Without adhering to this constraint, no final momentum state will be energetically degenerate with the initial state and the atomic sample will not respond to the presence of the standing wave, even if the interaction parameters are appropriate ($\omega_R^{(2N_B)}\tau \simeq 1$) for N_B^{th}-order Bragg scattering.

The condition $N_s = 0.1$ (eqn [46]), where N_s is the expected number of spontaneous emission events per atom during the light–atom interaction, is drawn on Figure 13 for two ratios (30 and 1000) of standing wave detuning, δ, to the atomic excited state natural lifetime, Γ. For fixed δ/Γ, we see that $N_s \propto \theta$, thus conditions with an increased pulse area (above and to the right of the $N_s = 0.1$ lines) result in a proportionally increased N_s. These $N_s = 0.1$ lines are extended across all regimes of Figure 13 because restricting spontaneous emission is required for both Kapitza–Dirac and Bragg scattering. As N_s approaches unity, the correct optical potential describing the light–atom interaction no longer is eqn [51], and scattering of atoms becomes an incoherent process which can lead to new possibilities such as a complex valued optical potential. Thus, while the standing wave fields in pure Bragg or Kapitza–Dirac scattering behaved as phase gratings for atoms, light gratings on resonance can effectively become amplitude diffraction gratings.

Atom Interferometry

In the 19th century, Fizeau (1853), Michelson (1881), Rayleigh (1881), and Fabry and Perot (Fabry 1899) exploited the interference properties of light waves to create the light interferometer which has since resulted in many beautiful experiments and precise measurements. Using technologies invented since the Second World War, the initial ideas of de Broglie and Schrödinger have evolved into construction of interferometers for neutrons, electrons, and atoms. These new interferometers are proving to be valuable tools for probing fundamental physics, for studying quantum mechanical phenomena, and for making inertial measurements.

The scientific value of interferometry with atoms and molecules has long been recognized. In fact, the concept of an atom interferometer was patented in 1973 and it has been extensively discussed since. Atom interferometry offers great richness stemming from the varied internal structure of atoms, the wide range of properties possessed by different atoms (e.g., mass, magnetic moment, absorption frequencies, and polarizability), and the great variety of interactions between atoms and their environment (e.g., static E–M fields, radiation, and other atoms).

Light interferometers are generally based either on achromatic beamsplitters such as half-silvered mirrors or on other semi-transparent membranes whose structure is small compared to the wavelength of the wave they are splitting. Lacking material structures that are either transparent to atoms or smaller than their de Broglie wavelength, diffraction gratings have been pressed into service both as beamsplitters and mirrors for atom waves. This means that atom interferometers are constrained to designs which somehow compensate for the dependence of diffraction angle on the wavelength of the individual atoms. In spite of this challenge, a surprising variety of atom and molecule interferometers have been built since 1991. A majority have used the three-grating configuration in which the first grating splits the incident beam, the second reverses the differential momenta given by the first, and the third recombines the two beams at the location where they overlap. Both material gratings and standing waves of light have been used in the Raman–Nath, Bragg, and adiabatic regimes to obtain interference fringes. Some designs render the interference fringes in position space, others in internal state space.

Figure 14 shows the MIT setup for a three-grating atom interferometer which features sufficient separation between interfering paths to accommodate an interposed metal foil. This allows different fields or media to act on atom waves only in one of the two

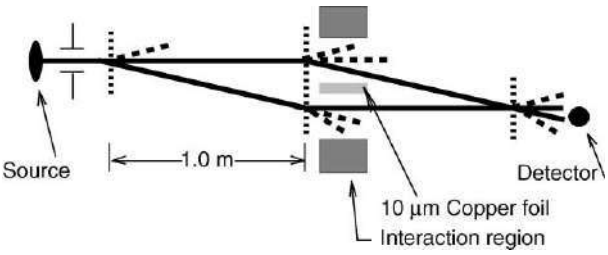

Figure 14 An interferometer for atoms built with three nanofabricated gratings. Each atom propagates in a superposition of two paths, shown in bold. The beam of atoms is well enough collimated, and the diffraction angle is large enough, that a metal barrier can be inserted between the two paths inside the interferometer. The interference fringes are observed by translating one grating transverse to the atom beam. See for example: Berman PR (ed.) (1997) *Atom Interferometry*. San Diego, CA: Academic Press.

arms of the interferometer. Such an interaction region has been used to measure phase shifts due to an electric field applied to only one of the interfering paths. The phase difference between the two paths is then given by the Feynman path integral:

$$\phi = \frac{-i}{\hbar}\int V(x)dt \qquad [63]$$

where

$$V = -\frac{1}{2}\alpha E^2 \qquad [64]$$

is the potential energy due to the electric filed E, and α is the electric polarizability of the atoms. The resulting phase shift of the interference fringes (**Figure 15**) has been used to measure the polarizability of sodium atoms to unprecedented accuracy.

Phase shifts have also been measured due a dilute target gas in one arm of the interferometer. In this case there is a complex index of refraction for the forward scattered atoms (some attenuation, and some phase shift) which depends on matter wave wavelength. Recent measurements reveal glory oscillations in the matter wave index of refraction due to passing through a dilute gas, and are a sensitive probe of interatomic potentials.

In the domain of inertial sensing, atom interferometers have been used to the local gravitational acceleration, $\vec{g}$, Newton's constant, G, gravity gradients, $\nabla\vec{g}$, and rotations Ω, each with sensitivity rivaling if not exceeding any other method. The Sagnac phase shift, due to rotations, for a two-path interferometer is

$$\phi_S = \frac{4\pi}{\lambda_{dB}v}\vec{\Omega}\cdot\vec{A} \qquad [65]$$

where v is the velocity of atoms, $\vec{A}$ is the enclosed area of the interferometer, and $\vec{\Omega}$ is the rotation rate. The two factors in the denominator, λ_{dB} and v, are both $\sim 10^5$ times larger for light than for atoms, which is why atom interferometers have the potential to be 10^{11} times more accurate than optical interferometers. In practice, atom interferometers only outperform laser gyros by a small amount because they have dramatically smaller enclosed areas and particle flux than optical interferometers.

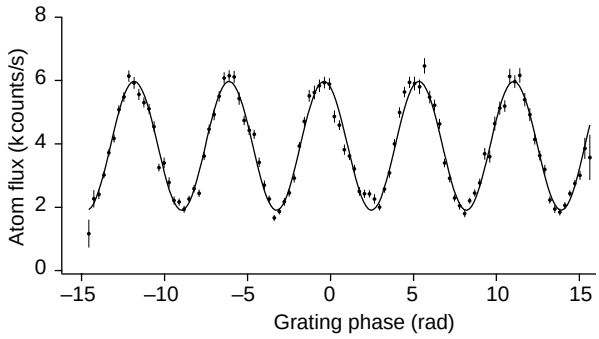

Figure 15 Interference fringes from atoms. 2π of grating phase corresponds to translating the grating one grating period, or 100 nm. These data were obtained at MIT.

Since the first atom interferometers for Na and He* began working in 1991, others have been made for Li, Ar*, Ca, Cs, K, Mg, Ne*, Rb, and Li_2, Na_2, I_2, C_{60} and C_{70} molecules, interferometers starting with trapped atoms have been made for Cs, Ca, He*, Mg, and Rb. Interferometers using Bose–Einstein condensates have been demonstrated with Na and Rb. These lists are still growing. Diffraction from a single grating has been observed for molecules as large as C_{70} buckyballs, which may someday be used in a three-grating interferometer.

Conclusion

The manipulation of atoms using light forces from standing light waves is a rich subject. The seminal suggestion of Kapitza and Dirac laid dormant for 50 years due to lack of experimental technology. However, in the 1980s this suggestion was realized with atomic sources and considerably extended both experimentally and theoretically. In the early 1990s, coherent standing wave manipulation and nonofabricated atom optics became the two major routes to making atom interferometers. As the new century begins, these techniques are being refined further with improved sources, new detection schemes, and better atom optics. Replacing thermal atomic beams by Bose–Einstein condensates (BECs), as sources of atoms, will revolutionize the field of atom optics just as lasers did in light optics.

Conversely, stimulated light scattering and nanostructures provide new ways to study atomic and molecular properties, such as Bragg spectroscopy for BEC characterization. As progress continues, exciting developments should be forthcoming, such as definitive measurements of the fine structure constant, α, atom gyroscopes that are far superior to the best laser gyroscope technology, and improved studies of BECs. Another area that is on the brink of spectacular development is the confinement of coherent matter waves in atom waveguides and the development of the scientific and technological opportunities that these represent, in analogy to fiber optic waveguides for light.

While we have described the application of nanofabrication techniques to atom optics, it is possible to imagine technology transfer in the other direction. The fundamental problem of fabricating

ever smaller structures might be tremendously advanced by atom optics because the de Broglie wavelength of atoms is so much smaller than that of light, permitting them to be focused to directly deposit much smaller features than possible with photolithography.

See also

Interferometry: Overview. **Scattering:** Scattering Theory.

Further Reading

Cohen-Tannoudji C, Diu B and Laloe F (1977) *Quantum Mechanics*. New York: Wiley and Sons.
Hecht E (1987) *Optics*. Reading, MA: Addison-Wesley.
Berman PR (ed.) (1997) *Atom Interferometry*. San Diego, CA: Academic Press.
Mandel L and Wolf EM (1995) *Optical Coherence and Quantum Optics*. London: Cambridge University Press.
Balykin VI and Letokhov VS (1995) *Atom Optics with Laser Light*. Chur, Switzerland: Harwood Academic Publishers.
Meystre P and Sargent M III (1999) *Elements of Quantum Optics*. Heidelberg, Germany: Springer Verlag.
Rauch H (2000) *Neutron Interferometry*. Oxford, UK: Oxford University Press.
Meystre P (2001) *Atom Optics*. New York: Springer-Verlag.
Campargue R (ed.) (2001) *Atomic and Molecular Beams, the State of the Art 2000*. Berlin: Springer.

Atomic Coherence Effects

A Belyanin, G R Welch and M O Scully,
Texas A&M University, College Station, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

History and Perspective

Quantum optics and laser physics have entered a new era. This started with the invention of the laser and the associated theoretical developments such as the theory of photon correlations and statistics, and the quantum theory of the laser. Later, new and novel lasers such as the free electron laser and tunable dye laser fueled many new research directions. With the development of the micromaser, quantum optics became a beautiful new test bed for the study of nonequilibrium statistical physics of photons interacting with matter in a controlled environment. The realization that a deeper understanding was needed to understand the measurement process led to the utilization of the laser as a probe of gravity, and to insight into nonclassical processes such as quantum nondemolition measurements and the production of squeezed states of light.

Effects of atomic coherence in quantum electronics have a long history. One of the first examples was the Hanle effect: excitation by polarized light creates atomic coherence that can evolve in a magnetic field, causing the change of polarization of the resulting fluorescence. Another important example is the interference of decay processes, first suggested by Fano. Modern schemes utilizing quantum coherence in atomic vapors or solid media usually involve strong coupling laser fields (laser drives). In each case, the effect of quantum coherence can be understood as interference of multiple (usually two) transition paths between quantum states of a system.

In the past decade, the concepts of atomic coherence and interference have led to many new and surprising innovations in laser physics such as electromagnetically induced transparency (EIT), refractive index enhancement, lasers without population inversion, ultra-slow light group velocity, storage of the quantum states of light, teleportation of quantum states, and nonlinear optics at an ultra-low intensity level.

Electromagnetically Induced Transparency

One of the most vivid manifestations of quantum coherence effects is the reduction of absorption in a three-level system that is driven by a coherent laser field. This effect is known as electromagnetically induced transparency, or EIT. We consider in more detail the case when the probe and drive fields are in a so-called Λ-configuration (Figure 1a) named after the appearance of the level scheme. In the conventional usage, the two lower levels are coupled to the upper level with two lasers. One of the lasers may be strong and is called the driving or coupling laser, and the other is weak and is called the probe laser.

Consider a quasi-monochromatic laser field E_d (called the 'driving' field, or 'coupling' field) and probe field E_p interacting with a three-level medium of Figure 1a. This interaction is described by a density matrix ρ_{ij}, where i, j take the values a, b, c. Within the

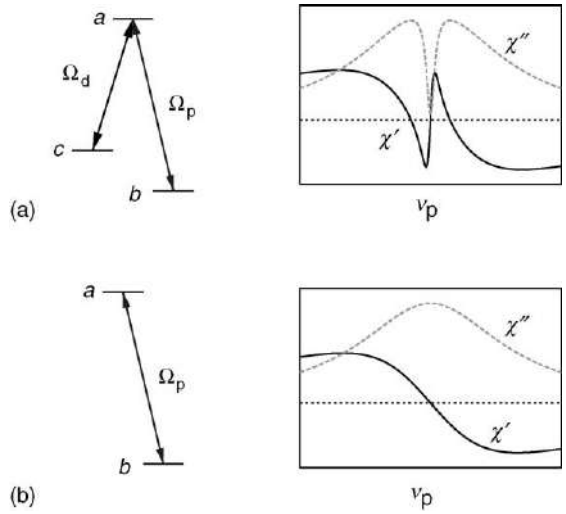

Figure 1 Real and imaginary parts of the susceptibility as functions of the frequency of the probe laser for (a) three-level Λ–type atoms, and (b) two-level atoms.

density matrix approach, quantum coherence in the system has the precise meaning of an off-diagonal element of the density matrix at the two-photon transition c–b. To have strong quantum coherence in a Λ-system, one must excite a large value of ρ_{cb}.

It is convenient to simplify the equations from the very beginning by using the rotating-wave approximation. We introduce slowly varying complex amplitudes of fields and polarizations and neglect their second derivatives. For example, for the probe field applied at the $a \rightarrow b$ transition we write

$$E_p(z,t) = \frac{1}{2}\varepsilon_p(z,t)\exp(-i\omega_p t + ik_p z) + \text{c.c.} \qquad [1]$$

After introducing the complex Rabi frequency $\Omega_p(z,t) = d_p\varepsilon_p(z,t)/2\hbar$, the wave equation for waves propagating in the z-direction can be written as

$$\frac{\partial\Omega_p}{\partial z} + \frac{n_p}{c}\frac{\partial\Omega_p}{\partial t} + \kappa_p\Omega_p = \frac{2\pi i\nu_p d_p^2 N}{\hbar c n_p^2}\sigma_{ab} \qquad [2]$$

Here d_p is the dipole moment of the probe transition a–b, N the total volume density of atoms in the active region, κ_p represents cavity losses, and $n_p = k_p c/\omega_p$ is the refraction index. The variable σ_{ab} is the slowly varying amplitude of the corresponding off-diagonal element $\rho_{ab} = \sigma_{ab}\exp(-i\nu_p t)$ of the density matrix. The same representation is assumed for the drive field E_d, with parameters n, d, ν, κ having index 'd'.

The functions σ_{ab}, σ_{ac}, and σ_{cb}, which determine the polarization response of the medium at the frequency of the external field, should be found from the density matrix equations

$$\frac{d\sigma_{cb}}{dt} + \Gamma_{cb}\sigma_{cb} = -i\Omega_p\sigma_{ac}^* + i\Omega_d^*\sigma_{ab} \qquad [3]$$

$$\frac{d\sigma_{ab}}{dt} + \Gamma_{ab}\sigma_{ab} = i\Omega_p n_{ba} + i\Omega_d\sigma_{cb} \qquad [4]$$

$$\frac{d\sigma_{ac}}{dt} + \Gamma_{ac}\sigma_{ac} = i\Omega_d n_{ca} + i\Omega_p\sigma_{cb}^* \qquad [5]$$

where

$$\begin{aligned}\Gamma_{cb} &= \gamma_{cb} + i(\nu_{cb} - \nu_p + \nu_d),\\ \Gamma_{ab} &= \gamma_{ab} + i(\nu_{ab} - \nu_p),\\ \Gamma_{ac} &= \gamma_{ac} + i(\nu_{ac} - \nu_d)\end{aligned} \qquad [6]$$

Here the γ's are relaxation rates of the polarizations at the corresponding transitions (the so-called homogeneous broadening). The population differences $n_{ik} = \rho_{ii} - \rho_{kk}$ are defined from eqns [2]–[6] together with the three equations for $\rho_{ii}, i = a, b, c$, with phenomenological rates of population relaxation and pumping:

$$\begin{aligned}\frac{d\rho_{bb}}{dt} &= -2\text{Im}[\Omega_p^*\sigma_{ab}] + r_{cb}\rho_{cc} + r_{ab}\rho_{aa} - r_{ba}\rho_{bb}\\ \frac{d\rho_{cc}}{dt} &= -2\text{Im}[\Omega_d^*\sigma_{ac}] + r_{ac}\rho_{aa} - r_{cb}\rho_{cc}\\ \frac{d\rho_{aa}}{dt} &= 2\text{Im}[\Omega_p^*\sigma_{ab}] + 2\text{Im}[\Omega_d^*\sigma_{ac}] + r_{ba}\rho_{bb}\\ &\quad - (r_{ab} + r_{ac})\rho_{aa}\end{aligned} \qquad [7]$$

Here we have included only the pumping term r_{ba} between levels b and a for simplicity. Later, we will consider the effect of incoherent pumping on gain, but in this section we put $r_{ba} = 0$. The sum of populations is evidently conserved in this model. Analysis of open systems with, e.g., flow of atoms through a cavity, leads to qualitatively similar conclusions.

Suppose that the probe field is weak, $|\Omega_p|^2 \ll \gamma_{cb}\gamma_{ac}$, and the drive-field intensity is constant in a sample. The complex susceptibility at the probe-field frequency, which is proportional to $\sigma_{ab}/\varepsilon_p$, is given by

$$\chi = \frac{id_p^2 N}{\hbar c}\left\{\frac{1}{\Gamma_{ab} + |\Omega_d|^2/\Gamma_{cb}}\left(n_{ba} - \frac{|\Omega_d|^2 n_{ca}}{\Gamma_{cb}\Gamma_{ac}^*}\right)\right\} \qquad [8]$$

This expression can be further simplified if there is no pumping to the states $|c\rangle$ and $|a\rangle$, so that we can

assume that $n_{ca} = 0$ and drop the second term in the brackets. The behavior of χ as a function of detuning from resonance $\nu_p = \nu_{ab}$ is shown in Figure 1a. In the absence of the drive field Ω_d, we retrieve the usual two-level susceptibility, shown in Figure 1b.

There are two remarkable features in the susceptibility of Figure 1a. The first one is the transparency window at the point of two-photon resonance. The absorption coefficient at resonance is

$$K'' = \frac{2\pi\nu_p d_p^2 N}{\hbar c(\gamma_{ab} + |\Omega_d|^2/\gamma_{cb})} \qquad [9]$$

Clearly, absorption can be arbitrarily small when $\gamma_{cb} \ll \gamma_{ab}$ and $|\Omega_d|^2 \gg \gamma_{cb}\gamma_{ab}$, since it is proportional to $\gamma_{cb}/|\Omega_d|^2$. For any value of $|\Omega_d|^2$, the absorption is much smaller than that resulting from simple Stark splitting of two Lorentzians. Moreover, in the most interesting case

$$\gamma_{cb}\gamma_{ab} \ll |\Omega_d|^2 \ll \gamma_{ab}^2$$

the Stark splitting is negligible while the deep and narrow EIT feature remains, with width $\gamma_{cb} + |\Omega_d|^2/\gamma_{ab}$.

Note that the intensity of the drive field required for the coherent bleaching of the medium is a factor $r_{ab}/\gamma_{cb} \gg 1$ smaller than the intensity needed for usual saturation bleaching.

EIT is closely related to the phenomenon of coherent population trapping (CPT). This can be illustrated by transforming from states $|a\rangle$, $|b\rangle$, $|c\rangle$ to a new set of states $|a\rangle$, $|B\rangle$, $|D\rangle$, where

$$|D\rangle = (\Omega_d|b\rangle - \Omega_p|c\rangle)/\sqrt{|\Omega_d|^2 + |\Omega_p|^2}$$

$$|B\rangle = (\Omega_d|b\rangle + \Omega_p|c\rangle)/\sqrt{|\Omega_d|^2 + |\Omega_p|^2}$$

It can be easily found that under the condition of two-photon resonance $\nu_p - \nu_d = \nu_{cb}$ the dipole matrix element from $|D\rangle$ to $|a\rangle$ goes to zero. This means that the population in state $|D\rangle$ does not interact with the coherent radiation field and does not contribute to absorption. That is why this state is called a non-absorbing or dark state. Under the condition $|\Omega_d|^2/\gamma_{cb}\gamma_{ab} \gg 1$ practically all the population of states $|b\rangle$ and $|c\rangle$ is optically pumped into the dark state, and only a small fraction $\propto \gamma_{cb}$ remains in the absorbing state $|A\rangle$ and contributes to absorption.

The second remarkable feature of a driven Λ-system is sharp variation of the real part of χ near the resonance, which can be used for manipulation of the group velocity of light, as discussed in the next section.

Slow, Ultra-Slow, Stored, and Frozen Light

Introduction

There are five useful definitions for the velocity of a traveling wave:

- the phase velocity, which is the speed at which the zero crossings of the carrier wave move;
- the group velocity, at which the peak of a wave packet moves;
- the energy velocity, at which energy is transported by the wave;
- the signal velocity, at which the half-maximum wave amplitude moves;
- the front velocity, at which the first appearance of a discontinuity moves.

Although these can all be different, in most cases (linear, passive, dispersive media) the last four coincide, and are usually less than the phase velocity. By using the very steep frequency dispersion in the vicinity of the narrow resonance associated with electromagnetically induced transparency (EIT), it has been shown that the group velocity of light can be reduced by approximately 100 million compared with its velocity in vacuum.

In order to understand the concept of group velocity, it is useful to consider the superposition of two waves of the same amplitude E_1 and E_2 where $E_i = E_0\cos(k_i z - \nu_i t)$ with $i = 1, 2$. Addition of these waves gives rise to the modulation shown in Figure 2:

$$E = E_0(\cos(k_1 z - \nu_1 t) + \cos(k_2 z - \nu_2 t))$$

$$= 2E_0\cos(\Delta k z - \Delta\nu t)\cos(kz - \nu t)$$

where $\Delta k = (k_1 - k_2)/2$, $\Delta\nu = (\nu_1 - \nu_2)/2$, $k = (k_1 + k_2)/2$, and $\nu = (\nu_1 + \nu_2)/2$. This superposition

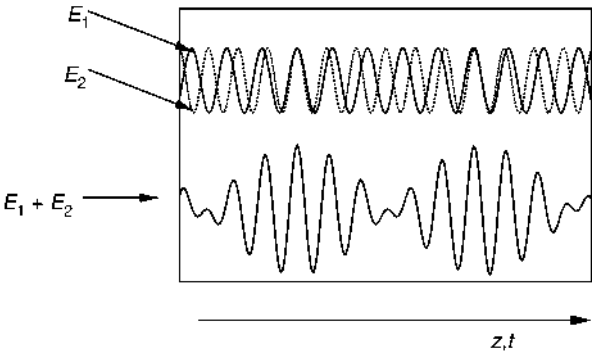

Figure 2 Interference of two monochromatic waves with different frequencies results in a wave modulated in time and space.

displays an interference pattern consisting of a rapid oscillation propagating with the so-called phase velocity,

$$v_{\text{phase}} = \frac{\nu}{k} \qquad [10]$$

and a slowly varying envelope propagating at the group velocity,

$$v_{\text{g}} = \frac{\Delta\nu}{\Delta k} \qquad [11]$$

Usually, the group velocity of wave packets consisting of more than two harmonics is given by transforming the ratio $\Delta\nu/\Delta k$ into $\mathrm{d}\nu/\mathrm{d}k$.

Slow and Ultra-Slow Light

In ordinary absorption and refraction of light, it is sufficient to consider only two atomic energy levels interacting with the light. However, in multilevel systems, coherent excitation of more than one transition can dramatically alter the dispersion and hence the group velocity. The first experiments of this effect demonstrated a group velocity of $c/13$. However, this was progressively lowered in later experiments as the steepness of the dispersion was increased, culminating in a group velocity as low as $c/300\,000\,000$ (1 m/s) in a Bose–Einstein condensate of ultra-cold sodium atoms. Slow light has been observed in a wide range of systems, from the Bose–Einstein condensates just mentioned, to hot optically dense gases of atoms and molecules, and in solids ranging from rare-earth doped crystals to solid hydrogen.

Most of these experiments occurred in a Λ system. The susceptibility of this medium has very steep dispersion without absorption, as shown in Figure 1.

The minimum possible group velocity slowed by the steep dispersion in EIT is given by

$$v_{\text{g}} = \frac{8\pi\Omega^2}{3\lambda^2 N\gamma} > 0 \qquad [12]$$

where $\gamma = \gamma_{ab} = \gamma_{ac}$. The driving field Ω should be strong enough to meet the condition $\Omega^2 \gg \gamma\gamma_{bc}$ for coherent effects. It can be shown that the group velocity approaches its minimum $v_{\text{g min}} = 8\pi\gamma_{bc}/(3\lambda^2 N)$ when $\Omega^2 \approx \gamma\gamma_{bc}$.

Bringing Light to a Halt: Frozen Light

Consider a laser pulse propagating through a medium with ultra-slow group velocity while the medium itself is moving with some velocity v in the opposite direction to the direction of laser pulse propagation. Let us denote the group velocity of the light in the frame co-moving with the atoms as $\tilde{v}_{\text{g}}$. In this frame, atoms are at rest, and hence, there is no spatial dispersion. The Galilean transformation to the laboratory frame, $k = \tilde{k}$, $\nu = \tilde{\nu} - \tilde{k}v$, where v is the atomic velocity, yields the group velocity $v_{\text{g}} = Re(\mathrm{d}\nu/\mathrm{d}k) = \tilde{v}_{\text{g}} - v$. This simple transformation shows that the light pulse is 'dragged' by the moving atoms. If the velocity of the atoms with respect to the laboratory frame is exactly equal and opposite to the group velocity of light in the frame where atoms are at rest, then we find that the group velocity in the laboratory frame is zero. We refer to this as 'freezing' of light. We may even make the group velocity negative so that the propagation direction is opposite to the wavevector (that is, that $v_{\text{g}} < 0$). The simplest example of spatial dispersion is the so-called drift dispersion corresponding to a mono-velocity atomic beam or moving sample with velocity v.

Slow light experiments in Doppler broadened systems such as warm gases allow group velocities that are much lower than the mean thermal speed (ν_{T}) of the atoms in the thermal sample (Figure 3). This means that freezing of light can occur without actually moving the atomic sample. When the experimentally obtained group velocity is less than or about equal to the mean thermal speed of the atoms, a pulse of light can be stopped in a stationary cell. In this case the delay time for a pulse passing through the medium would tend to infinity (contrary to the case of a finite atomic beam) in the sense that this pulse would never leave the cell.

This is accomplished by using a single velocity group from the Maxwellian thermal distribution of atomic velocities, and adjusting the frequency of the driving field to be resonant with that velocity group (see Figure 4). In this case it is mainly the atoms in this single velocity group that support the ultra-slow group velocity as a slow EIT polariton, and the atoms

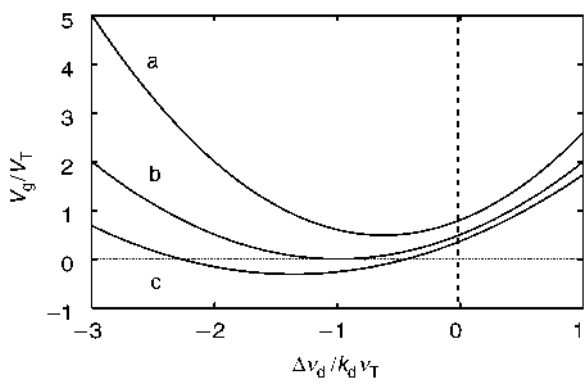

Figure 3 Ultra-slow and negative group velocity of EIT polariton versus detuning of drive laser; $\Omega = 0.25\gamma$, $k_{\text{d}}\nu_{\text{T}} = 100\gamma$, $\gamma_{cb} = 0.001\gamma$, (a) $N = 0.6\,N_{\text{cr}}$; (b) $N = N_{\text{cr}}$; (c) $N = 1.5\,N_{\text{cr}}$.

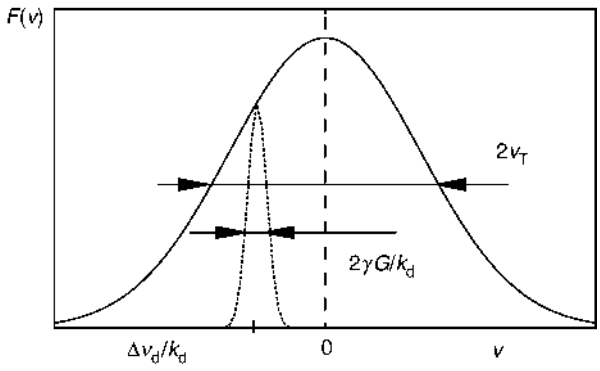

Figure 4 The velocity distribution of atoms in a cell (solid line). Effective drifting beam (dotted) selected by drive laser.

then act as an effective atomic beam. If the driving field is on resonance with the velocity group of atoms moving into the light with speed equal to $\tilde{v}$, we will have $v_g = \tilde{v}'_g - v_d$, where $\tilde{v}'_g$ is the group velocity of the light defined by the effective atom beam if those atoms were at rest. This requires that the intensity of the drive is strong enough to provide EIT for the resonant velocity group of atoms, i.e., $\Omega^2 > \gamma_{bc}\gamma$ (see Figure 4) but at the same time weak enough to avoid interaction with off-resonant atoms moving with the 'wrong' velocities (i.e. $\Omega < kv_T\sqrt{\gamma_{bc}/\gamma}$).

Storing and Retrieving Quantum Information

Photons are potentially the best particles for use in quantum computing and quantum information processing. They are the fastest, and perhaps the simplest and most robust carriers of quantum information, but their storage and manipulation is difficult without destroying their quantum state. However, for application to quantum cryptography, quantum teleportation, and quantum computation it is necessary to achieve a strong coupling between electromagnetic waves and a large nonlinear susceptibility with small losses. For this task, it may be resonant coherent media that offer the promise of success.

Recently, new coherent and reversible methods based on EIT have been found for the transfer and manipulation of photons and their quantum information. Quantum states of photons are trapped in coherently driven atomic media in which the group velocity is reduced adiabatically toward zero. The methods allow for an ideal transfer of quantum correlations between light fields and metastable states of matter. Currently, interesting applications such as quantum state memories and quantum information processing including transporting and time reversing the state of light are being investigated.

Although experiments have not yet been conducted with single quanta, proof-of-principle experiments using classical laser fields have demonstrated the technique. In these experiments control and signal light pulses propagate in a gas of three-level Λ-type atoms and excite a spatial profile of a long-lived coherence of ground spin states. This spin coherence profile contains quantum information about the 'writing' pulses. When the control field is switched back on, the signal pulse is restored via Raman scattering on the atomic coherence. Ideally, the scattered pulse is in the same quantum state as the input signal pulse, preserving entanglement, quantum statistics, etc.

Similar experiments have also shown that different reading pulses (rather than switching back on the same control pulse) can be used to obtain new tools for quantum information storage and processing. Time reversing, teleportation, color switch and multiplexing of the signal pulse of light have all been demonstrated. These new operational possibilities arise as a result of the reading of the stored information with a light pulse that is spatially separated from, has different frequency from, and propagates in the opposite direction to the signal beam.

To demonstrate these effects, one pair of control and signal pulses (the 'writing' fields) prepares a spin coherence of the lower level (Figure 5). A second control pulse (which can be thought of as a 'reading' field) scatters from this coherence resulting in the generation of a new, or 'recovered', field. This recovered field acquires some properties of the reading field because it is a Raman scattered component of the reading pulse. If the reading field pulse is centered about a frequency other than that of the writing fields, and propagates in the opposite direction to the writing fields, then the recovered field propagates in the same direction as the reading field and has a different frequency from the incident writing fields. Even if the writing and reading pulses are separated in space as well as time the scattering effect will persist. This allows the transportation of the coherence grating to another point in the atomic cell via the atomic center-of-mass motion.

In the ideal case the signal pulse profile and quantum statistics are exactly stored by the atomic coherence grating. Thus, reading by a backward propagating pulse results in a generated pulse that is also backward propagating. That is it represents an exactly time-reversed copy of the signal pulse. Similarly, atomic motion in the direction transverse to the light propagation moves the coherence grating to another spatial position. Reading the light by a spatially shifted laser results in the equivalent of

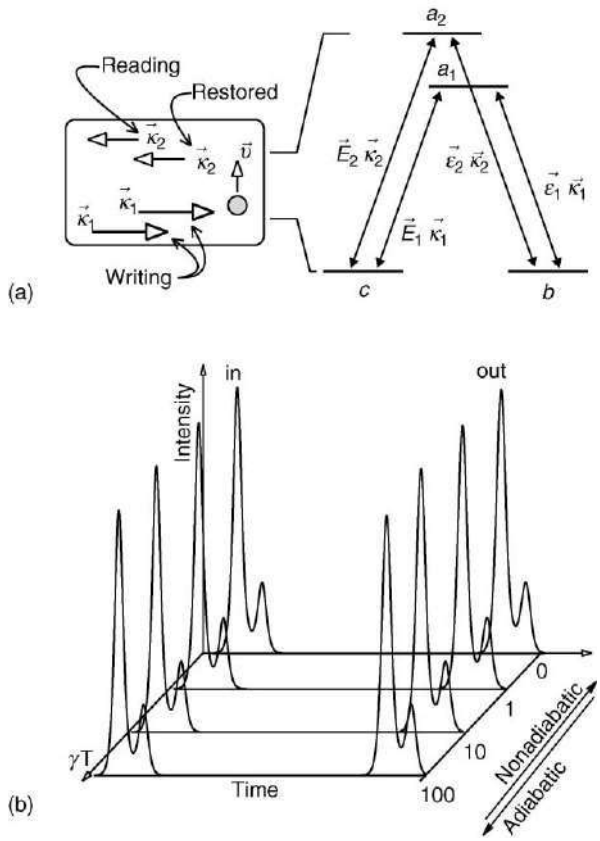

Figure 5 (a) Energy level scheme used in experiments for multiplexing, transporting, and time-reversing stored light. (b) 'Storage of light': intensity of the signal field versus time for different switching times (T) of the reading and writing fields.

'teleportation' of light by atoms in the same spirit as the teleportation of the quantum state of light by atoms that has been discussed recently. It is also possible for the frequency of the 'restored' pulse to be easily shifted by switching the frequency of the reading pulse. This means that the use of a frequency comb of reading pulses propagating in different spatial modes should result in multiplexing of the signal light.

Novel Laser Sources Based on Quantum Coherence

Introduction

In the previous sections we have seen how quantum coherence can be used to overcome resonant absorption on a transition of interest, leading to EIT. A natural question arises whether it is possible to achieve gain without population inversion at this transition, i.e. lasing without inversion (LWI). This subject has been intensively investigated since

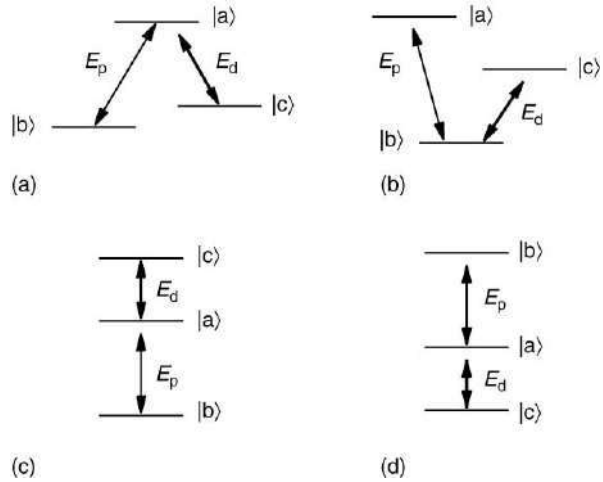

Figure 6 LWI schemes in a three-level system coherently driven by a laser field.

the early 1990s, and the first proof-of-principle experiments have been successful.

Besides the fundamental physical beauty of this phenomenon, the research effort has been motivated by potentially important applications. It is clear that LWI could be particularly useful in situations where a population inversion is difficult to maintain due to, e.g., very fast relaxation or inefficient pumping. The obvious examples are high-frequency UV, X-ray and gamma-ray lasers, or semiconductor mid/far-infrared lasers based on intersub-band transitions. Another promising application of LWI is efficient frequency up- or down-conversion. Furthermore, if resonant absorption can be eliminated, resonant nonlinear optical interactions become a reality. Since all nonlinear coefficients are maximized at resonance, highly efficient nonlinear optical generators of light can be created.

In this section we discuss the simplest possible schemes of inversionless lasing which allow one to clarify the physical origin of LWI and to see the shortcomings which have so far prevented LWI-based schemes from reaching the level of practical laser devices.

Consider a three-level system, coherently driven by a strong quasi-monochromatic field nearly resonant with one transition, and probed by a field that is nearly resonant with an adjacent transition. Some possible configurations are shown in Figure 6. We consider in detail the case when the two fields are in a Λ-configuration (Figure 6a). This is probably the most popular and well-studied scheme since it can provide probe-field gain for a very low pumping rate, when most of the population is at the lowest energy level. However, we will see that the gain

condition for the Λ-scheme is actually rather restrictive and does not allow one to realize large frequency up-conversion. For the latter purpose, ladder-type schemes (Figure 6c,d) turn out to be more favorable.

Weak-field gain in the Λ-scheme

From the general formalism presented earlier, it is straightforward to get the amplification condition

$$K'' = \mathrm{Re}\left\{\eta_\mathrm{p}\frac{1}{\Gamma_{ab}+|\Omega_\mathrm{d}|^2/\Gamma_{cb}}\left(n_{ba}-\frac{|\Omega_\mathrm{d}|^2 n_{ca}}{\Gamma_{cb}\Gamma_{ac}^*}\right)\right\}+\kappa_\mathrm{p} < 0 \qquad [13]$$

where

$$\eta_\mathrm{p} = \frac{2\pi\omega_\mathrm{p} d_\mathrm{p}^2 N}{\hbar c n_\mathrm{p}^2}$$

The physical meaning of the two terms in brackets in eqn [13], which originate from the two terms on the right-hand side of eqn [4] for σ_{ab}, can be easily understood. The first term is just the resonant one-photon absorption proportional to the population difference n_{ba}. It can be strongly reduced in the presence of a drive field Ω_d, as explained earlier. However, it can change sign and give rise to the field amplification only if $n_{ba} < 0$, meaning there is a population inversion!

The second term, originating from the product $\Omega_\mathrm{d}\sigma_{cb}$ in eqn [4], is due to the mixing of the drive field Ω_d and polarization σ_{cb} where the latter is parametrically excited by a two-photon process: $\sigma_{cb} \propto e_\mathrm{p}\sigma_{ac}^* \propto e_\mathrm{p}\Omega_\mathrm{d}^* n_{ca}$. It is this term that can overcome one-photon absorption to provide amplification without inversion in the Λ-scheme. As we see, the possibility of LWI depends crucially on our ability to excite a sufficiently large polarization σ_{cb} at the transition resonant with the beat frequency of the probe and drive fields. This conclusion is universal for all LWI schemes with a coherent drive. We see that the sometimes loosely defined term 'quantum coherence' has a precise meaning here as the polarization σ_{cb} at a two-photon transition c–b. It is clear that the effects of quantum coherence are maximized when γ_{cb} is small, i.e., the transition c–b is long-lived.

We are of course interested in the case of weak incoherent pumping. The typical dependence of gain on detuning from two-photon resonance is shown in Figure 7. At resonance, where the gain is maximized, the amplification condition $K'' < 0$ yields (in the absence of nonresonant losses):

$$r_{ac} > 2\gamma_{cb}(r_{ab}/r_{ba} - 1) \geq r_{ab} - r_{ba} \simeq r_{ab} \qquad [14]$$

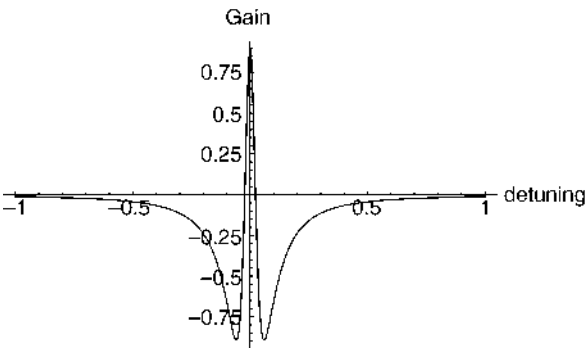

Figure 7 Gain without inversion in Λ-scheme.

because $\gamma_{cb} \geq r_{ba}/2$. Here we assumed that the incoherent pumping is weak, $r_{ba} \ll r_{ab}$, so that it does not produce population inversion at the probe transition.

Equation [14] is a simple and very important result, which is valid for all three-level laser schemes in the absence of inversion at both probe and drive transitions: the relaxation rate at the drive transition should exceed that at the probe transition. This result is an immediate consequence of the fact that the long-lived coherence σ_{cb} is inevitably destroyed by incoherent pumping, and γ_{cb} is now bounded from below. One can follow how the EIT regime is transformed into lasing as follows. With increasing incoherent pumping rate r_{ba}, EIT is actually destroyed due to the growth of $\gamma_{cb} \simeq r_{ba}/2$. In terms of CPT states, this means increase of population of the absorbing state $|B\rangle$. At the same time, the resonant Raman term in [13] also grows since n_{ac} is proportional to r_{ba}. With increase in the incoherent pumping rate, the contribution to gain overcomes the contribution to absorption if $r_{ac} > r_{ab}$.

Propagation effects

Inequality [14] has important implications for the problem of co-propagation of probe and drive fields. Indeed, to provide a significant amplification of the probe field, the absorption length $L_\mathrm{d} = \gamma_{ac}/(\eta_\mathrm{d} n_{ca})$ of the drive field should be larger than the amplification length $L_\mathrm{p} \simeq \gamma_{ac}/(\eta_\mathrm{p} n_{ca})$ of the probe. This requirement gives simply

$$\eta_\mathrm{p}/\eta_\mathrm{d} > 1 \qquad [15]$$

which in the absence of nonradiative relaxation can be rewritten as

$$\frac{r_{ab}}{r_{ac}}\frac{\nu_d^2}{\nu_p^2} > 1$$

In view of the gain condition given in eqn [14], the above requirement can be satisfied only if $\nu_d > \nu_p$, i.e., up-conversion is impossible.

An interesting solution to this problem is to realize population inversion at the driving transition. Then up-conversion can be realized with a lower relaxation rate at the drive transition as compared with probe transition. In this case in the Λ and V-schemes the two-photon term contributes to absorption at resonance, and gain is only possible far off-resonance. At the same time, in the ladder schemes shown in Figure 6c and d the gain is achieved at resonance, as discussed below in more detail.

It should also be noted that to excite laser oscillations in a high-Q cavity, large one-pass gain is not required at all, so that a Λ-scheme could still be employed for up-conversion.

Now, consider a probe field of arbitrary intensity. An important question is where the energy for amplification of the probe field comes from. It is known that for parametric processes in a transparent medium the Manley–Rowe relations are satisfied. Therefore, it might seem that the drive field is also the ultimate energy source for LWI. However, this is the case only for very large detunings of the drive and probe fields from the corresponding transition frequencies. In that case, we can neglect one-photon absorption, and the wave equations take the form typical for stimulated Raman scattering. They possess the first integral

$$\frac{|\varepsilon_d|^2}{\nu_d} + \frac{|\varepsilon_p|^2}{\nu_p} = \text{const} \qquad [16]$$

which expresses the conservation of the total number of quanta in both waves (Manley–Rowe conditions).

However, in the case of exact resonance, one can show that one-photon absorption of the drive field is always important, and the Manley–Rowe-type conservation law does not exist at all. The optimal condition for the probe field amplification is $\eta_p/\eta_d \gg 1$. In this case the stage of exponential amplification of the probe is maintained over a long distance, and the probe field intensity can grow to much larger values than the sum of initial intensities of the drive and probe fields. Apparently, the energy for the probe field amplification comes mainly from the medium, namely from the upper-level population ρ_{aa} supported by incoherent pumping.

V-Scheme

The V-scheme, shown in Figure 6b, is qualitatively similar to the Λ-scheme. An important difference is that there is now no long-lived coherence at the mixing transition c–a. Indeed, even if the transition $a \to c$ is strongly forbidden, the coherence γ_{ac} is destroyed by fast relaxations at the c–b and a–b transitions. We can immediately suggest that for the V-scheme inversionless lasing takes place at large driving field intensities exceeding the saturation value. However, one must be careful: although it appears that the reduction of absorption in the absence of pumping cannot be greater than a factor of 2, this turns out to be incorrect. Rather, almost complete transparency can be achieved, as confirmed by calculations and recent experiments. Analysis of the laser gain condition shows that the V-scheme shares with Λ-scheme the same negative feature: relaxation at the driving transition should be much faster than at the probe transition.

Schemes with Population Inversion at the Driving Transition

The above analysis illustrates the common problem of LWI schemes: the requirement of rapid relaxation at the driving transition and the resulting strong absorption of the drive field. This difficulty can be overcome in schemes with population inversion at the driving transition. Then, of course, all this makes practical sense only if the relaxation at the drive transition is much slower than that on the probe transition, so that it is easy to provide population inversion for the drive transition as compared with the probe transition. Luckily, such schemes do exist in principle, and seem to have many potential applications. The simplest possibility is the so-called ladder (or cascade) schemes shown in Figure 6c,d. LWI in ladder schemes is usually studied in the regime when there is no inversion at both probe and drive transitions. In this case gain is achieved at large detunings from two-photon resonance and under the same unfortunate condition of fast relaxation at the drive transition as compared with the probe transition. In the opposite case, population inversion at the driving transition is required for LWI at the probe transition. In ladder schemes the maximum gain of the probe field is achieved when both fields are at resonance with atomic transitions, while in V and Λ-schemes the resonance point corresponds to maximum absorption, and the gain is achieved in sidebands.

Note that population inversion at the driving transition means that the drive field can be amplified. This immediately puts forward the possibility to use a drive field which is self-generated in the same active medium. In other words, the system supports lasing with inversion at the drive frequency and LWI at the probe frequency. This situation is especially beneficial for dense gases or solids where the problems of

external drive absorption, reflection, and inhomogeneity may become otherwise insurmountable. One specific example considered in the literature is LWI on rapidly decaying UV and soft X-ray transitions in gas lasers in the presence of lasing with inversion on long-lived adjacent transitions in the visible range (as in Figure 6c). Another example is LWI on rapidly decaying mid/far-infrared intersub-band transitions in semiconductor quantum wells in the presence of lasing with inversion on the interband transition, as in Figure 6d. These interband transitions have 10^4 times longer lifetime than the intersub-band transitions.

Experiments

Experimentally, amplification without inversion has been observed in rarefied atomic vapors in the early and mid-1990s. In particular, amplification of a probe beam in the Λ-configuration was realized in atomic samarium, sodium, and cadmium, while the V-scheme was successfully implemented in atomic rubidium. Inversionless laser oscillation was observed in a V-scheme in rubidium vapor and a Λ-scheme in a sodium atomic beam. In all of these proof-of-principle experiments, the frequencies of the probe and drive fields were very close to each other, and the coherence was excited between two hyperfine states or Zeeman sublevels. One-pass gain of a few per cent was observed.

Nonlinear Optical Sources Based on Quantum Coherence

Excitation of a strong coherence has a profound impact on nonlinear optical interactions. Strong scattering on the polarization wave excited at the beat frequency of two or more laser fields allows one to achieve huge nonlinear coefficients and conversion efficiencies. A particularly impressive example is ultra-broadband collinear Raman generation in a resonant Λ-scheme, that has been observed experimentally. A pair of strong laser pulses propagating in molecular deuterium excites a strong polarization wave at a beat frequency close to a vibronic transition in deuterium. As a result, a comb of Raman sidebands from the infrared to the deep-UV region of the spectrum has been observed.

Another example is an extremely efficient simultaneous blue light and mid-infrared generation in a four-wave mixing experiment in rubidium vapor. Here, all four fields are resonant with corresponding transitions.

A final example illustrates the use of self-generated drive fields for nonlinear generation. It has been suggested that one can use two intracavity generated laser modes in a semiconductor quantum-well laser for the difference frequency, sum frequency, and second-harmonic generation in fully resonant three-wave mixing processes. This idea has already been successfully realized in two-wavelength quantum-cascade lasers, and experiments with other types of semiconductor lasers are under way.

The whole field of resonant nonlinear optics, stemming from original ideas of LWI in systems with quantum coherence, is now in explosive development.

Summary

Atomic coherence effects have already found many important applications in metrology and low-intensity nonlinear optics. Novel applications in quantum nonlinear optics, laser physics, and quantum information processing seem to be feasible. The coherence allows the storage of information about the probe light, its transportion in space, time reversing, and multiplexing the light. Moreover, the coherence allows an increase of the coupling between light fields such that it becomes possible to study the interaction between single photons. New coherent radiation sources based on the ideas of LWI and coherent nonlinear optics are at the stage of active research. There are surely many other applications ahead.

See also

Electromagnetically Induced Transparency. **Non-Classical Light**. **Nonlinear Optics, Applications:** Raman Lasers. **Nonlinear Optics, Basics:** Cascading; Four-Wave Mixing. **Quantum Optics:** Entanglement and Quantum Information.

Further Reading

Belyanin AA, Bentley C, Capasso F, *et al.* (2001) Inversionless lasing with self-generated driving field. *Physical Review A* 64: 013814.

Harris SE (1997) Electromagnetically induced transparency. *Physics Today* 50(9): 36–42.

Kocharovskaya O (1992) Amplification and lasing without inversion. *Physics Reports* 219: 175–190.

Mompart J and Corbalán R (2000) Lasing without inversion. *Journal of Optics B: Quantum and Semiclassical Optics* 2: R7–R24.

Rostovtsev Y, Kocharovskaya O, Welch GR and Scully MO (2002) Slow, ultra-slow, stored and frozen light. *Optics Photonics News*, June.

Scully M and Zubairy S (1997) Quantum Optics. Cambridge, UK: Cambridge University Press.

Entanglement and Quantum Information

P G Kwiat, University of Illinois at Urbana-Champaign, Urbana, IL, USA

D F V James, Los Alamos National Laboratory, Los Alamos, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Quantum information is an emerging field of technology that encompasses the application of fundamental quantum mechanical phenomena, such as entanglement, to tasks in information processing and communications. The importance of optical technology in quantum information is not surprising, given the success of quantum optics in the investigation of these fundamental phenomena and the pre-eminent role of optics in modern communications. This article describes some of the important advances in photonic quantum information and discusses the prospects for the future.

Basic Notions: Qubits and Entanglement

Qubits

Quantum information is usually confined to two-level quantum systems, which are referred to as quantum bits, or qubits. The qubits form the register of a quantum computer or carry the information in a quantum communications channel. Physical examples of qubits under active study include the spin degrees of freedom of an electron or spin-1/2 nucleus, two well-isolated energy levels of an atom or trapped ion, and the polarization degrees of freedom of a photon; here we concentrate on the latter.[1] The two states of each qubit are generically denoted by $|0\rangle$ and $|1\rangle$ (although, because of our concentration on optical polarization, we shall use the horizontal and vertical states $|H\rangle$ and $|V\rangle$ interchangeably with $|0\rangle$ and $|1\rangle$ when describing experiments). An arbitrary state of a single qubit, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, is specified by the complex probability amplitudes α and β associated with the two possible states, constrained by $|\alpha|^2 + |\beta|^2 = 1$. States for which α and β are deterministic quantities are referred to as pure states; when they are stochastic, the state is mixed, and must be described by a density matrix, e.g., $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$, where p_i is the probability of the pure state $|\psi_i\rangle$. Pure states are a useful but unattainable ideal, just like a perfectly coherent laser beam; in reality even the best-prepared quantum states have some level of mixedness, which can be quantified by the entropy: $S \equiv -Tr(\rho \ln\rho) = -\sum_i \lambda_i \ln\lambda_i$, where λ_i are the eigenvalues of the matrix ρ.

Multiqubit States and Entanglement

For two qubits, the possible pure states of the system are specified by four probability amplitudes:[2] $|\psi_2\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle$, with $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$, and, e.g., $|00\rangle \equiv |0\rangle \otimes |0\rangle$ denoting the state in which the first and the second qubits are both in state $|0\rangle$. In general, 2^n amplitudes are needed to characterize a pure state of n qubits. Immediately we can see one advantage of quantum systems for information storage: if each distinct probability amplitude is regarded as a data register, the size of the memory grows exponentially with the number of qubits. Further, if one were to flip just one qubit, for example the second, the state would be transformed into $|\psi'_2\rangle = \beta|00\rangle + \alpha|01\rangle + \delta|10\rangle + \gamma|11\rangle$, where all of the amplitudes have been affected. This simple example demonstrates the notion of quantum parallelism, one of the most powerful properties of quantum information processors.

If the two-particle state can be written as a product of two single-particle states, i.e., if $\alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle = (A|0\rangle + B|1\rangle) \otimes (C|0\rangle + D|1\rangle)$, then these two particles can be considered as separate, unconnected entities: the state is said to be separable.[3] When the state cannot be written in this form, it is called an entangled state. Entanglement is one of the most fascinating fundamental properties of quantum systems: Erwin Schrödinger described it as '*the* characteristic trait of quantum mechanics … that enforces its entire departure from classical lines of thought.' Consider performing a measurement on the second of the two qubits, to establish which of its two states the particle was in, leaving the first qubit

[1] Recently some attention has been devoted to quantum information based on continuous variables, such as those arising in connection with 'squeezed light'. However, here we will focus on discrete systems, i.e., qubits.

[2] Qubits are assumed to be distinguishable particles, and so the wavefunctions describing their states need not be symmetrized under exchange of particles.

[3] When mixture is also present, the state is separable if it can be written as $\rho = \sum_i p_i \rho_i^A \otimes \rho_i^B$, i.e., as a sum of product states for systems A and B.

untouched. The result would collapse the state of the first qubit into one of two possibilities: either $|\phi_0\rangle = (\alpha|0\rangle + \gamma|1\rangle)/\sqrt{P_0}$ if the outcome of the measurement were 0 (which occurs with probability $P_0 = |\alpha|^2 + |\gamma|^2$) or $|\phi_1\rangle = (\beta|0\rangle + \delta|1\rangle)/\sqrt{P_1}$ if it were 1 (with probability $P_1 = |\beta|^2 + |\delta|^2$). Thus the state of the first qubit is instantaneously projected into a specific state by performing a measurement on the second qubit, which can be in a completely different location. This example demonstrates the nonlocality of quantum mechanics, which has been confirmed experimentally in some depth by a number of elegant quantum optics experiments over the last 30 years.

The fidelity $F = |\langle\Phi|\Psi\rangle|^2$ characterizes the overlap of two states.[4] The fidelity of $|\phi_0\rangle$ and $|\phi_1\rangle$ is a way to quantify the amount of entanglement in the initial two-qubit state. If the final two states are identical, the fidelity will be unity, and the initial state is separable; conversely if the two final states are orthogonal and occur with equal probability, then in a sense the initial state is maximally entangled, since the largest possible nonlocal influence would occur due to the measurement. Mathematically, the fidelity of the two final states is $|\langle\phi_0|\phi_1\rangle|^2 = 1 - C^2/4P_0P_1$, where $C = 2|\alpha\delta - \beta\gamma|$ is called the concurrence of the state. If $C = 0$, the state is separable; if $C = 1$, the state is maximally entangled.[5]

As discussed below, entanglement forms the heart of a number of quantum information protocols, such as dense coding, teleportation, and one type of quantum key distribution (also known as quantum cryptography), and large-scale entangled states of many qubits seem to be a requirement for the more ambitious goal of practical quantum computing.

Creating Entangled States Experimentally

Entangled states can currently be created in a controlled manner using technologies such as ion traps, cavity quantum electrodynamics, and optical spontaneous parametric down-conversion (SPDC). Down-conversion is a nonlinear optical process by which an incident 'pump' photon can be split (or down-converted) into a pair of longer-wavelength daughter photons (historically called 'signal' and 'idler') in a crystal possessing a $\chi^{(2)}$ nonlinearity, such as beta-barium borate (BBO). Mathematically, the process is described using the creation and annihilation operators of the field modes $(\hat{a}_\lambda^\dagger, \hat{a}_\lambda)$, thus:

$$|\Phi_{\text{out}}\rangle = \exp\left[-i \sum_{p,s,i} c_{p,s,i}(\hat{a}_p^\dagger \hat{a}_s \hat{a}_i + \hat{a}_i^\dagger \hat{a}_s^\dagger \hat{a}_p) \right] |\Phi_{in}\rangle$$

where the triple sum is over all field modes (subscripts p, s, and i refer to pump, signal, and idler modes, respectively) and $c_{p,s,i}$ is a coefficient linearly dependent on the second-order nonlinear susceptibility $\chi^{(2)}_{ijk}$ and also on the birefringent properties of the crystal. Further, $c_{p,s,i}$ will be negligibly small unless both total photon energy and momentum are conserved (i.e., $\omega_p = \omega_s + \omega_i$ and $\vec{\kappa}_p = \vec{\kappa}_s + \vec{\kappa}_i$, where $\vec{\kappa}$ is the intracrystal momentum).

One particularly efficient method for using this phenomenon to create entangled states is as follows. For a specific geometry (type-I phase matching), the daughter photons emerge from the crystal with identical polarizations (perpendicular to the parent polarization and the crystal optic axis) on opposite sides of a cone that is centered about the pump beam (Figure 1a). Because each photon is in a definite state of polarization, the two photons are not in an entangled state. Two crystals, aligned with their axes of symmetry oriented at 90° to each other, as shown in Figure 1b, can then be used to create an entangled state.[6] With crossed crystals, two processes are possible: the parent photon can down-convert in the first crystal to yield two vertically polarized photons, or it can down-convert in the second crystal to yield two horizontally polarized photons. Because it is impossible to distinguish which of these processes has occurred, the state of the daughter photons is a coherent quantum-mechanical superposition of the states that would arise from each crystal alone; the output of the crossed crystals is photons in the maximally entangled state $|\Phi_+\rangle \equiv \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle)$.[7]

Figure 2 shows how this basic source can be adapted to produce any pure quantum state of two photons by placing rotatable half- and quarter-wave plates (which can be used to transform the

[4] The generalized fidelity between mixed states ρ_A and ρ_B is $F = |Tr\sqrt{\sqrt{\rho_A}\rho_B\sqrt{\rho_A}}|^2$.

[5] The concurrence can also be generalized to mixed states, although it has a much more complicated expression. The quantification of entanglement for mixed states with more than two subsystems is presently an active area of research.

[6] Another widely used method employs type-II phase matching in a single crystal. The down-conversion photons are emitted with perpendicular polarizations along a pair of cones. Pairs emitted along particular directions are in the maximally polarization-entangled state $|\psi_-\rangle = \frac{1}{\sqrt{2}}(|HV\rangle - |HV\rangle)$. Such a source was used in the first demonstrations of superdense coding and quantum teleportation.

[7] Note that in an arbitrary basis $|\theta\rangle \equiv \cos\theta|H\rangle + \sin\theta|V\rangle$ and $|\theta^\perp\rangle \equiv -\sin\theta|H\rangle + \cos\theta|V\rangle$, this maximally entangled state has the form $|\Phi_+\rangle = \frac{1}{\sqrt{2}}(|\theta\theta\rangle + |\theta^\perp\theta^\perp\rangle)$, demonstrating that the nonlocal correlations are present regardless of the bases used to represent the state.

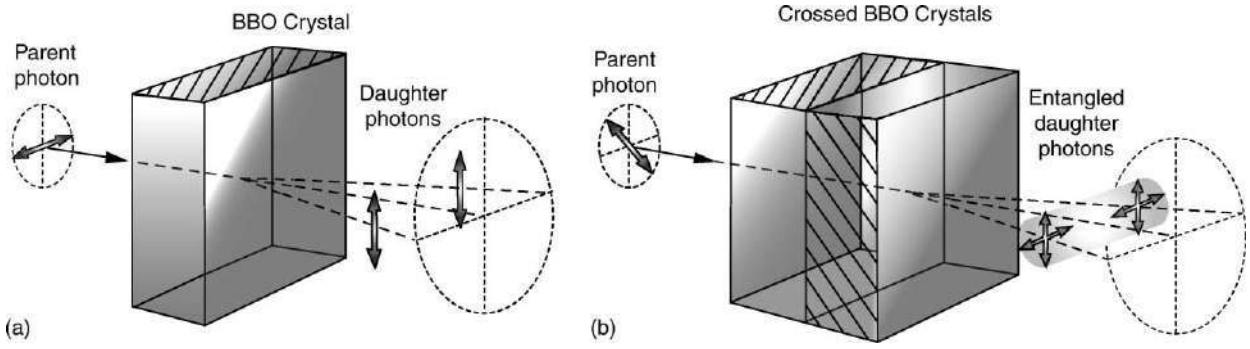

Figure 1 An entangled-photon source. (a) For a given orientation of the nonlinear crystal, a horizontally polarized parent photon produces a pair of vertically polarized daughters. The daughters emerge on opposite sides of an imaginary cone. The cone's axis is parallel to the original direction taken by the parent photon. The two daughter photons are not in an entangled state. Reorienting the BBO crystal by 90° will produce a pair of horizontally polarized daughters if a vertically polarized pump beam is used. (b) Passing a photon polarized at 45° through two crossed BBO crystals can produce two photons in an entangled state. Because of the Heisenberg uncertainty principle, there is no way to tell in which crystal the parent photon 'gave birth,' and so a coherent superposition of two possible outcomes results: a pair of vertically polarized photons or a pair of horizontally polarized photons. (Adapted with permission from James DFV and Kwiat PG (2002) Quantum state entanglement: creation, characterization, and application. *Los Alamos Science* 27: 52–67.)

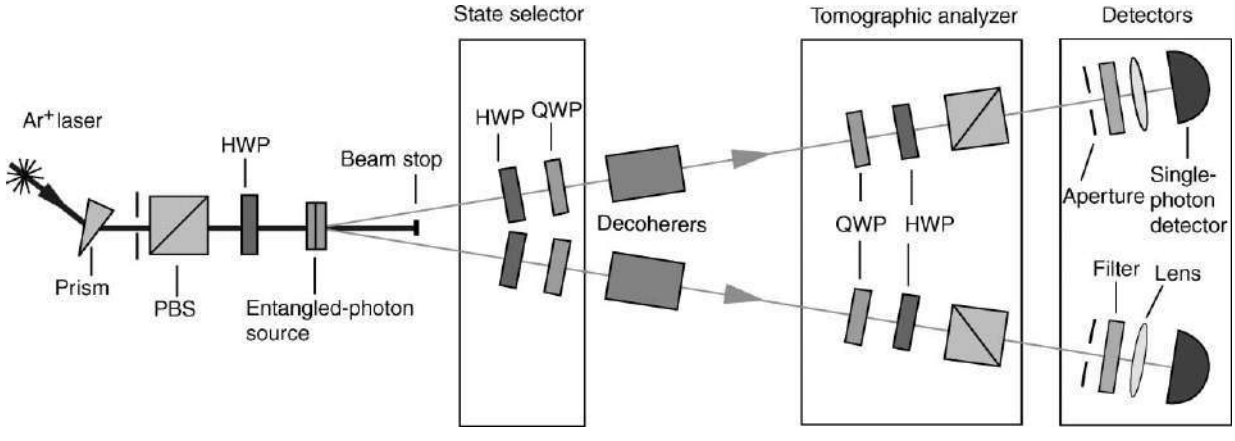

Figure 2 Creating and measuring two-photon entangled states. The 'pump' photons are created, for example, in an argon ion laser and are linearly polarized with a polarizing beamsplitter (PBS). The half-wave plate (HWP) rotates the polarization state before the photon enters the pair of nonlinear crystals that constitute the entangled-photon source; the initial angle of pump polarization controls the entanglement of the pair produced. Each photon's polarization state can be altered at will by the subsequent HWP and quarter-wave plate (QWP). The decoherers following the state selection allow the production of mixed photon states. The optical elements (QWP, HWP, and PBS) in the tomographic analyzer allow the measurement of each photon in an arbitrary basis. Coincidence measurements of photons allow the quantum state to be determined. (Adapted with permission from James DFV and Kwiat PG (2002) Quantum state entanglement: creation, characterization, and application. *Los Alamos Science* 27: 52–67.)

polarization state of a single photon) before the crystal pair and in the paths of the two daughter photons. To create mixed states, a long birefringent crystal can be used to delay one polarization component with respect to the other. If the delay is longer than the coherence time of the photons, the horizontal and vertical components are effectively decohered; that is, the phase relationship between the different states is destroyed.

The figure also shows schematically the apparatus required for measuring the quantum state. In classical optics, four Stokes parameters are required to specify the polarization of a single beam (i.e., an ensemble of uncorrelated photons); for a pair of photons, 16 projective measurements, each with different wave plate settings, is required. From these 16 measurements, all of the elements of the 4×4 density matrix describing the (in general, mixed) state of the photon pairs can be deduced. This is an example of quantum state tomography, a technique that has found application to a number of quantum optical systems.

Quantum Key Distribution

Two parties, historically known as Alice and Bob, want to have a secret conversation. A generic, classical encryption protocol would begin when Alice and Bob convert their messages to separate binary streams of 0's and 1's, which are then encrypted and decrypted with a set of secret 'keys' known only to the two. Each key is a random string of 0's and 1's that is as long as the binary string comprising each message. To encrypt, Alice (the sender) sequentially adds each bit of the key to each bit of her message, using addition modulo 2. She then sends the encrypted message over a public channel to Bob, who decrypts it by simply repeating the addition modulo 2 of the key to the message. This type of encryption, known as a one-time pad, is currently the only provably secure encryption protocol. But the one-time pad is effective only if Alice and Bob never reuse the key, and more obviously, if the key remains secret. A potential eavesdropper, Eve, cannot be allowed to glean any part of the bit stream that makes up the key. Therein lies a central problem of cryptography – how can secret keys be created and then securely distributed?

Quantum key distribution (QKD) exploits the fundamentally indivisible nature of photons to perform this task. There are a variety of QKD protocols; here we describe one that employs entangled photon pairs (see Figure 3). Alice and Bob use a source such as the one described above to produce maximally entangled photons in the state $|\Phi_+\rangle \equiv \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle) = \frac{1}{\sqrt{2}}(|45°,45°\rangle + |-45°,-45°\rangle)$. One photon goes to Alice and the other to Bob. For each pair, Alice and Bob randomly and independently analyze their respective photons (using a polarizing beamsplitter) in the H/V or the 45°/−45° basis. They record a bit value of 0 for all H or 45° results, and a 1 for all V or −45° results. After a sufficient number of measurements (dictated by the length of the key), Alice and Bob have a public discussion, e.g., over the internet. For each detected photon, they announce which basis they used for the measurement, but not the actual measurement result. Whenever they made the same basis choice (50% of the time), the correlations of the entangled state ensure their measured bit values agree. By contrast, they discard the results when they used different bases, because their measurements are completely uncorrelated (see Table 1).

An eavesdropper (Eve) cannot tap the line, as she might with conventional communications, because of the indivisibility of individual photons and the fact

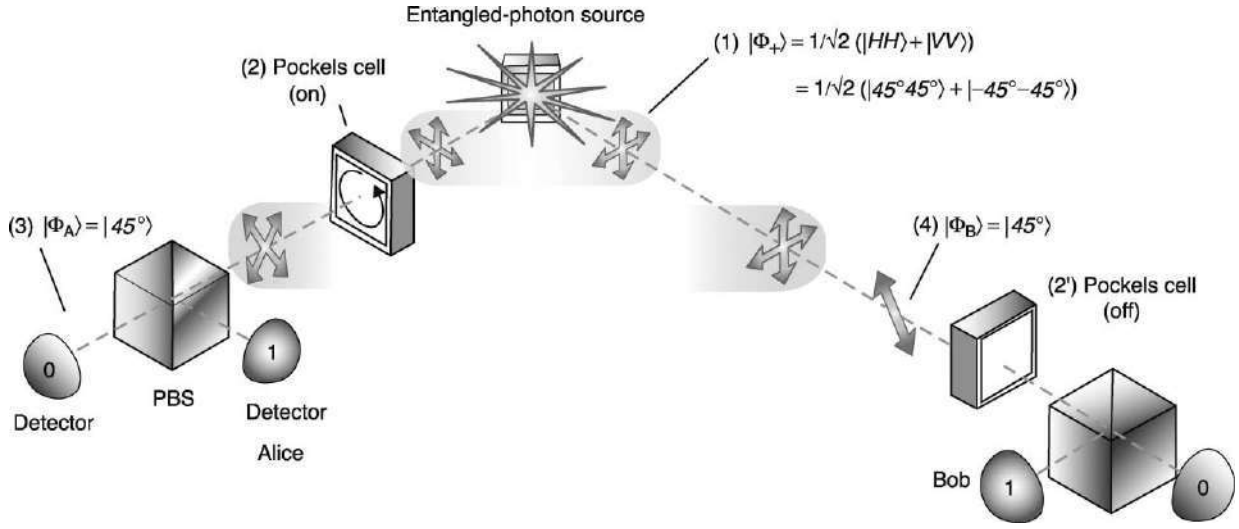

Figure 3 Quantum cryptography using entangled photons. Entangled photons can be used to create a pair of identical cryptographic keys. One member of an entangled pair (1) is sent to Alice, and the other to Bob. Alice and Bob each randomly and independently analyze their respective photon in one of two linear polarization bases with a polarizing beamsplitter (PBS); the basis can be actively chosen using a Pockels cell before the PBS, as shown, to rotate the polarization (alternatively, the 'choice' could be made by directing the photon onto a nonpolarizing 50/50 beamsplitter; in the transmitted path, one analysis basis is used, in the reflected path, the other is used). In the example shown, Alice used the 45°/−45° basis (2), and measured 45° polarization (3), thus projecting Bob's photon into the identical state (4). Since he chose the H/V analysis basis (2′), he is equally likely to detect a 0 or a 1. By subsequent public discussions Alice and Bob determine the events for which they used the same analysis basis (and discard the other events). For these events Alice and Bob will have obtained identical measurement results, which they may interpret as raw key material. After classical error correction and privacy amplification techniques are applied, the remaining string is the sought-after shared secret key. (Adapted with permission from James DFV and Kwiat PG (2002) Quantum state entanglement: creation, characterization, and application. *Los Alamos Science* 27: 52–67.)

Table 1 Polarization entanglement-based quantum cryptography protocol

Alice's analysis basis	H/V	H/V	45°/−45°	H/V	45°/−45°	45°/−45°	H/V	45°/−45°
Alice's measurement result[a]	*H*	–	*−45°*	*V*	*45°*	*45°*	*V*	*−45°*
Alice's bit value	0	–	1	1	0	0	1	1
Bob's analysis basis	H/V	45°/−45°	45°/−45°	45°/−45°	H/V	45°/−45°	H/V	HV
Bob's measurement result	*H*	*−45°*	*−45°*	*45°*	*H*	–	*V*	*H*
Bob's bit value	0	1	1	0	0	–	1	0
Public discussion: Both photons detected and same basis used?	yes	no	yes	no	no	no	yes	no
Remaining secret key	0		1				1	

[a]In some cases Alice or Bob may not detect a photon due to loss or detector inefficiency. These events simply do not contribute to the key data.

that arbitrary quantum systems cannot be accurately cloned.[8] If Eve steals Bob's photon (a 'denial-of-service' attack), the photon's information never becomes part of the key. Thus, although a wiretap would reduce the rate of the transmission, it would not jeopardize the security of the key. Eve can try to intercept the photon, measure it, and send another one to Bob. But any measurement Eve would make to determine the photon's polarization state would perturb the photon and collapse the entangled state. The photon she sends to Bob would therefore only be 'classically' correlated with Alice's photon. Consequently, Eve's intervention necessarily induces additional errors into Bob's key, which Alice and Bob can detect by publicly revealing a small subset of their actual key. Unfortunately, even with no eavesdropper, the encryption keys created by any real-world quantum cryptography system typically possess a few-percent errors. To make sure their key is secure, Alice and Bob ascribe all errors to Eve and then estimate the maximum amount of information available to the eavesdropper. They then use a classical privacy amplification protocol to reduce Eve's knowledge of the secret key to less than one bit by reducing the length of the key. It has been proven that if the initial error probability per bit is greater than ~15%, no secret bits will remain after error detection and privacy amplification.

Researchers are working to make entanglement-based quantum cryptography more practical. Also, a number of longer distance quantum cryptography demonstrations using weak pulses have been carried out, over tens of kilometers in fibers and in free space. In fact, the first commercially available systems have recently been announced (in Europe and the United States).

[8] The no-cloning proof is very simple. If we have a copying operation such that $|0\rangle|c\rangle \rightarrow |0\rangle|0\rangle$ and $|1\rangle|c\rangle \rightarrow |1\rangle|1\rangle$ ($|c\rangle$ is the initial state of the copier), then by linearity, a superposition becomes an entangled state, instead of two copies: $\frac{(|0\rangle+|1\rangle)}{\sqrt{2}}|c\rangle \rightarrow \frac{(|0\rangle|0\rangle+|1\rangle|1\rangle)}{\sqrt{2}} \neq \frac{(|0\rangle+|1\rangle)}{\sqrt{2}}\frac{(|0\rangle+|1\rangle)}{\sqrt{2}}$.

Table 2 Method to convert the initial state $|\Phi_+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ into any of the four Bell states

Alice's operation	*Polarization transformation*	*Resultant two-qubit state*
$\hat{I}$, the identity	H → H; V → V	$\lvert\Phi_+\rangle = \frac{1}{\sqrt{2}}(\lvert 00\rangle + \lvert 11\rangle)$
$\hat{\sigma}_x$	H → V; V → H	$\lvert\Psi_+\rangle = \frac{1}{\sqrt{2}}(\lvert 01\rangle + \lvert 10\rangle)$
$i\hat{\sigma}_y$	H → V; V → −H	$\lvert\Psi_-\rangle = \frac{1}{\sqrt{2}}(\lvert 01\rangle - \lvert 10\rangle)$
$\hat{\sigma}_z$	H → H; V → −V	$\lvert\Phi_-\rangle = \frac{1}{\sqrt{2}}(\lvert 00\rangle - \lvert 11\rangle)$

Quantum Superdense Coding

It is possible for Alice to send Bob two bits of classical information using a single qubit in the quantum superdense coding protocol. Suppose that Alice and Bob share one qubit each of an entangled pair in the maximally entangled state $|\Phi_+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, where we have returned to the generic labeling scheme for the qubit states. Alice has two classical bits of information, which is equivalent to one of four choices. She can encode this information by applying one of four possible transformations on her qubit; a suitable set of transformations are the three Pauli matrices and the identity (i.e., do nothing). This set of operations can be performed experimentally on photon qubits quite easily, e.g., using waveplates, as indicated in Table 2. The four resultant two-qubit states are all maximally entangled and are orthonormal. They form a special basis for the two-qubit states, the Bell basis, which is of particular use in analyzing entanglement.

Alice now sends her qubit to Bob, e.g., through an optical fiber. Bob can then perform a Bell state analysis[9] on his qubit pair, i.e., a projective measurement of the two-qubit state in the Bell basis. The result

[9] Complete discrimination of all four Bell states is presently an unsolved technical problem: for photons one needs either a nonlinear interaction (which is typically very weak) or to exploit so-called 'hyper-entanglement' involving other entangled degrees of freedom of the photons.

immediately reveals the choice of operation Alice made, and in effect, two bits of classical information have been encoded on a single qubit.

Quantum Teleportation

Another application of entanglement is quantum state teleportation (Figure 4), in which the infinite amount of information contained in an arbitrary qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is transferred by communication of two bits of classical information between Alice and Bob, if they also share two qubits in the maximally entangled state $|\Psi_-\rangle$. The three-photon initial state (i.e., the input photon plus the two entangled photons) is

$$|\psi_0\rangle = (\alpha|H\rangle + \beta|V\rangle) \otimes \tfrac{1}{\sqrt{2}}(|HV\rangle - |VH\rangle),$$

which can be rewritten with the first two photons (the input plus the first half of the entangled pair) represented by the Bell state basis:

$$\begin{aligned}|\psi_0\rangle = \ & \tfrac{1}{2}\{|\Phi_+\rangle \otimes (-\beta|H\rangle + \alpha|V\rangle) \\ & + |\Phi_-\rangle \otimes (\beta|H\rangle + \alpha|V\rangle) \\ & + |\Psi_+\rangle \otimes (-\alpha|H\rangle + \beta|V\rangle) \\ & + |\Psi_-\rangle \otimes (-\alpha|H\rangle - \beta|V\rangle)\}.\end{aligned}$$

Now suppose that a Bell state measurement is performed on the first two photons. The third photon

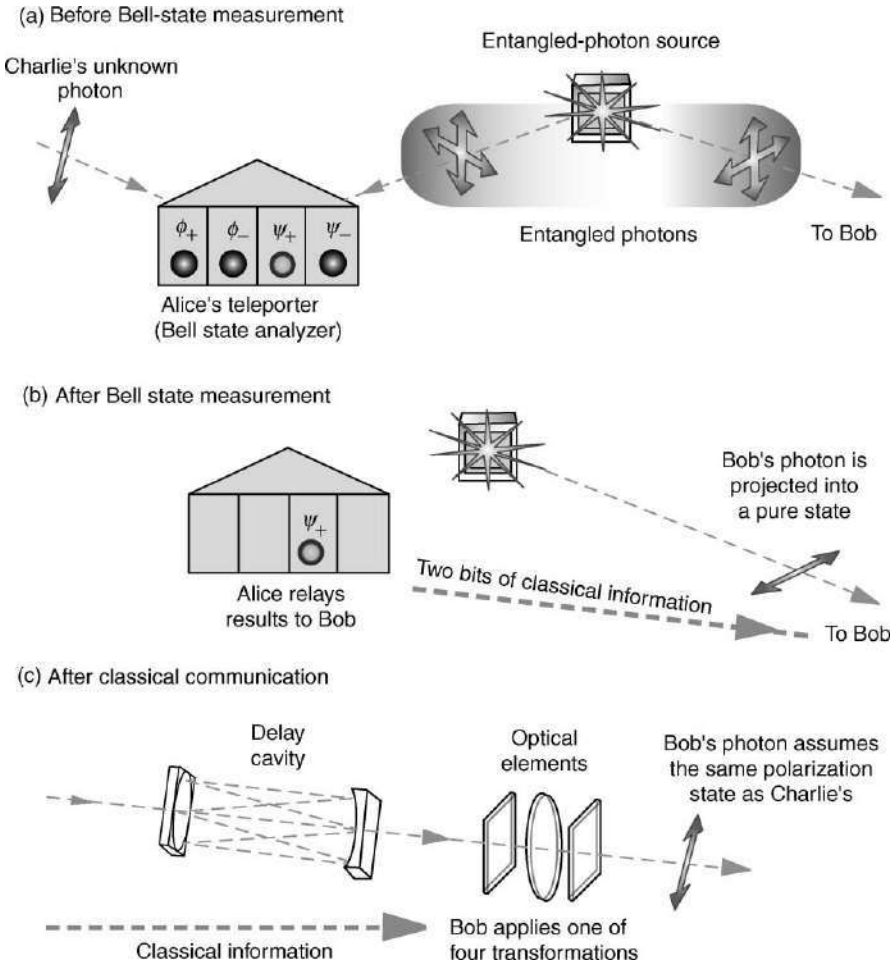

Figure 4 Quantum state teleportation. (a) Teleportation requires an entangled photon source, a Bell state analyzer and a classical communications channel. One entangled photon goes to Bob and the other to Alice, who also receives a photon of unknown polarization state. (b) Alice performs a joint polarization measurement of the two photons and relays the result to Bob using two classical bits of information. The photon going to Bob is projected into a pure state as a result of Alice's measurement. (c) Upon receiving Alice's classical information, Bob performs a simple transformation on his photon (which he has been storing), such as a rotation of the polarization vector. The resulting state of his photon is then identical to the unknown state Alice wished to teleport. (Adapted with permission from James DFV and Kwiat PG (2002) Quantum state entanglement: creation, characterization, and application. *Los Alamos Science* 27: 52–67.)

is immediately projected into one of four possible states, which can be transformed back into the state of the original input photon by a simple operation, e.g., with a waveplate. For example, if the Bell state measurement produced the result $|\Psi_+\rangle$, the third photon is immediately 'collapsed' into the pure state $|\psi_3\rangle = -\alpha|H\rangle + \beta|V\rangle$. By applying a π-phase shift to the horizontal polarization (relative to the vertical), $|\psi_3\rangle$ can be transformed into the original input state: $|\psi'_3\rangle = \alpha|H\rangle + \beta|V\rangle$. Alice (who does the Bell state measurement) need only communicate to Bob (who wants to receive the teleported photon) which of the four Bell states she measured. Note that during the entire teleportation procedure, neither Alice nor Bob can obtain any idea of the values of the parameters α and β, which specify the state. Also, because Alice's measurement collapses the unknown state, there is only a single copy at the end of the protocol.

Quantum state teleportation was first demonstrated experimentally by Zeilinger and coworkers (University of Innsbruck, Austria). The group was able to determine two of the four Bell states unambiguously (the other two states gave the same experimental signature) and proved for those cases that the state of the input photon could indeed be transferred to Bob. Other experiments have realized modified forms of quantum teleportation in different systems. For example, Kimble's group (California Institute of Technology) teleported the coherent state of an optical mode using squeezed (rather than polarization-entangled) light. Researchers have recently suggested how teleportation might form the basis of a distributed network of quantum communication channels, and how it might enable quantum computing in all-optical systems.

Other Application of Optical Entanglement

Quantum Computing

The most challenging and powerful application of quantum information is large-scale quantum computing. Two features that make quantum information processing potentially powerful are the exponentially large Hilbert space, which gives quantum registers very large capacities, and quantum parallelism, which means that data processing tasks can be performed very efficiently. One fundamental drawback, which severely constrains useful applications of quantum computers, is that the final measurement can only produce a number of classical bits equal to the number of qubits. Thus, quantum computers are limited to performing tasks in which a small amount of information is meant to be gleaned from processing a large amount of data; examples include searching an unstructured database for a specific entry (Grover's algorithm) or finding the periodicity of a function (the quantum Fourier transform). This second task is central to Shor's factor-finding algorithm, the most famous quantum computing algorithm to date.

A practicable quantum computer technology must have at least the following features, first identified by DiVincenzo:

- A set of well-characterized, distinguishable qubits to form a quantum data storage register.
- The ability to initialize the qubits of the register in a simple fiducial state.
- Decoherence times that are much longer than the time needed to perform logical operations.
- The ability to perform any single qubit operation on any qubit in the register, and the ability to perform two-qubit conditional logic gates (such as the CNOT gate: $\alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle \rightarrow \alpha|00\rangle + \beta|01\rangle + \gamma|11\rangle + \delta|10\rangle$). Together, these operations constitute a universal set of quantum gates, from which all other gates can be synthesized.
- The ability to measure each qubit.

All-optical schemes have been used to implement small quantum algorithms. However, most of the approaches are not scalable, due to the limitations of non-linear optics to perform a CNOT gate at the single-photon level. Recent proposals have suggested that very high efficiency single-photon detectors, along with sources of single photons 'on-demand,' allow scalable quantum computing with only linear optics; preliminary two-qubit gates have been experimentally demonstrated. A number of other promising candidate technologies that meet DiVincenzo's five requirements are being pursued vigorously. For example, the ability to create multi-qubit entanglement and perform reliable measurements on trapped ions cooled and manipulated by lasers has recently been demonstrated. Solid state systems offer the possibility of scalability, and a number of schemes, such as quantum dots, isolated impurities with nuclear spins, and superconducting quantum interference devices (SQUIDs), are being investigated. It is possible that the final quantum computing technology may take the form of a hybrid between various present approaches.

Lithography

Lithography, in which a pattern is optically imaged onto some photoresistive material, is the primary method of manufacturing microscale or nanoscale

electronic devices. An inherent limitation of this process is that details smaller than a wavelength of light cannot be written reliably. However, quantum state entanglement might circumvent this limitation. Under the right circumstances, the interference pattern formed by beams of entangled photons can have half the classical fringe spacing. Quantum lithography requires two beams of photons, a coherent superposition consisting of the state in which two photons are in beam A while none are in B, and the state in which no photon is in beam A while two photons are in B. Such number-entangled states can be made in the laboratory, and the predictions about fringe spacings have been verified. However, other obstacles must be overcome to surpass current classical-lithography techniques.

Two-Photon Imaging and Microscopy

At present, two-photon microscopy is widely used to produce high-resolution images, often of biological systems. However, the classical light sources (lasers) used for the imaging have random spreads in the temporal and spatial distributions of the photons, and the light intensity must be very high if two photons are to intersect within a small enough volume to cause a detectable excitation. Such high intensity can damage the system under investigation. Because the temporal and spatial correlations may be much stronger between members of an entangled photon pair, much weaker light sources could be used, which would be much less damaging to the systems being observed. The development of such systems is currently an active area of research.

List of Units and Nomenclature

Entanglement: a property of quantum systems consisting of two or more distinct subsystems, often separate particles. When transformations on one system do not affect the other, the state is said to be separable; when this is not the case, and the quantum state of the overall system cannot be resolved into separate states of the individual pieces, the system is entangled.

Pure and mixed states: when the probability amplitudes specifying a quantum state are deterministic, the state is said to be pure; when the amplitudes are random quantities, the state is mixed. The distinction is similar to that of coherent and partially coherent optical fields.

Quantum computer: an information processing device that exploits quantum mechanical phenomena to greatly enhance computational power for certain problems. Some mathematical problems thought to be intractable on conventional computers can, in theory, be performed efficiently on a quantum computer.

Quantum crytography (also known as quantum key distribution): a quantum information protocol by which two parties acquire a shared series of random numbers (a cryptographic key) by exploiting the quantum nature of light. Absolute security can be proven by virtue of the indivisibility and uncopy-ability of individual quanta.

Quantum state tomography: a means by which quantum states can be determined experimentally by performing a series of appropriate measurements on multiple identically prepared systems. From such measurements, the elements of the density matrix, which fully specifies the state, may be inferred.

Quantum superdense coding: A quantum information protocol by which two bits of classical information may be communicated by a single quantum bit, initially part of an appropriate entangled quantum state.

Quantum teleportation: a quantum information protocol by which the unknown quantum state of one particle can be transferred to another distant particle, using a pair of entangled particles, a projective measurement, and exchange of two bits of classical information.

Qubit (quantum bit): a two-level quantum mechanical system, which constitutes the building blocks of quantum information processing devices. Examples include a spin-1/2 particle, an atom with two well-isolated levels, or the polarization of a photon.

See also

Nonlinear Optics, Applications: Three-Dimensional Microfabrication. **Quantum Optics:** Quantum Computing with Atoms.

Further Reading

Berman GP, Doolen GD, Mainieri R and Tsifrinovitch V (1998) *Introduction to Quantum Computers*. Singapore: World Scientific.

Bouwmeester D, Ekert AK and Zeilinger A (eds) (2000) *The Physics of Quantum Information*. Berlin: Springer-Verlag.

Braunstein SL and Lo HK (eds) (2001) *Scalable Quantum Computers: Paving the Way to Realization*. Berlin: Wiley-VCH.

DiVincenzo DP (2000) The physical implementation of quantum computation. *Fortschrcitte der Physik* 48: 771–783.
Gisin N, Ribordy G, Tittel W and Zbinden H (2002) Quantum cryptography. *Review of Modern Physics* 74: 145–195.
Kok P, Braunstein SL and Dowling JP (2002) Quantum lithography. *Optics and Photonics News* 13(9): 24–27.
Lo HK, Popescu S and Spiller T (eds) (1998) *Introduction to Quantum Computation and Information*. Singapore: World Scientific.
Macchiavello C, Palma GM and Zeilinger A (eds) (2000) *Quantum Computation and Quantum Information Theory*. Singapore: World Scientific.
Mandel L and Wolf E (1995) *Optical Coherence and Quantum Optics*. Cambridge, UK: Cambridge University Press.
Nielsen MA and Chuang IL (2000) *Quantum Computation and Quantum Information*. Cambridge, UK: Cambridge University Press.
Williams CP and Clearwater SH (1998) *Explorations in Quantum Computing*. Santa Clara: Telos.
Yariv A (1989) *Quantum Electronics*. New York: Wiley.

Laser Cooling of Ions

H Walther, University of Munich and Max-Planck-Institute for Quantum Optics, Garching, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The fact that light can exert a mechanical force on atoms by the transfer of linear momentum was recognized as long ago as the 1930s, when Frisch demonstrated the deflection of an atomic beam using resonant light. With the advent of tunable laser light in the early 1970s, a much larger excitation rate for atoms was achieved, leading to a much stronger deflection. A little earlier it was also observed that in addition to the force resulting from photon recoil, a dipole force is also produced in a focused light beam so that even small solid particles can be trapped and manipulated. The original proposal for laser cooling of atomic gases was put forward in 1975 and later laser cooling of trapped ions was proposed. The principle of laser cooling is very simple. The ion is excited by laser light below a resonance frequency; in order that a transition can be made, the residual energy has to be taken from the kinetic energy of the particle, leading to cooling of the particle. In this article, we will discuss this cooling process in more detail.

Laser Cooling of Ions in a Radiofrequency Trap

Of the various proposed and realized methods for laser cooling of atoms or trapped ions, sideband cooling is conceptually the simplest and most efficient. It allows a trapped particle to be cooled to the quantum-mechanical ground state of a harmonic trap potential. The process can be regarded as anti-Stokes Raman scattering of the pseudomolecule formed by the ion and the trap. If the linewidth of an optical transition in this atom is smaller than the vibrational frequency of the trapped ion, both electronic levels of the transition split into a ladder of resolved vibrational levels. The absorption spectrum consists of a carrier at the resonance of the free atom and sidebands at multiples of the vibrational frequency. With a laser tuned to the m-th lower frequency sideband, each absorption of a photon will reduce the vibrational energy of the atom by m quanta. Since the subsequent spontaneous re-emission will, on average, not change the vibrational excitation, repeated photon scattering processes lead to efficient cooling of the atom. This cooling will continue until the absorption probability for the low-frequency sidebands decreases when approaching the vibrational ground state of the ion in the trap. The final temperature is limited by the heating process caused by photon recoil. However, after the ion is cooled to the vibrational ground state this heating process disappears, when the vibrational energy of the ion is larger than the recoil energy of a photon. In this case, the recoil is taken up by the entire trap setup, as it is the case in the Mössbauer effect.

Despite these obvious advantages of sideband cooling, relatively few experiments have been made using this method. This is because the regime, where the oscillation frequency ω is larger than the natural linewidth Γ of an optical transition (the so-called 'strong binding regime'), is not easily accessible. Vibrational frequencies of ions in standard electromagnetic traps are of the order of 1 MHz, whereas linewidths of electric dipole transitions in positive ions are in the range 20–50 MHz. In this 'weak binding regime' with $\Gamma > \omega$, laser cooling of trapped

ions is essentially analogous to Doppler cooling of free atoms. The minimum temperature is given by the Doppler limit $k_B T = \hbar\Gamma/2$, and the mean vibrational quantum number is $\langle n\rangle = \Gamma/2\omega > 1$. The first experiments with cooling on resolved sidebands, therefore, used a strongly forbidden optical transition or stimulated Raman transitions. In these cases, however, sideband cooling is inevitably slow and it was used only to extract a few vibrational quanta from Doppler-precooled ions. In both experiments the vibrational ground state was reached to a very good approximation (mean quantum number $\langle n\rangle \approx 0.05$). Raman sideband cooling has also recently been applied to neutral atoms in optical traps.

Laser cooling of trapped ions was motivated by their possible use for optical frequency standards where the relativistic Doppler shift, as well as the Stark shift due to the trap potential, are eliminated at low temperatures. Cooling to the ground state also opens up the possibility of studying a well-controlled mechanical quantum oscillator experimentally. This led to the realization of nonclassical states of motion, investigation of quantum decoherence phenomena, and the demonstration of a quantum logic gate for quantum computing.

Sideband Cooling

Most theoretical treatments of sideband cooling consider a static harmonic trap, whereby the time dependence of the potential of the radio-frequency (rf) ion trap is neglected. In such a model a running laser wave $\propto\exp(\mathrm{i}kr)$ of wave number k induces transitions between the eigenstates $|n\rangle$ and $|n'\rangle$ of a harmonic oscillator. The relevant transition matrix elements are

$$U_{n'n} = \langle n'|\exp(\mathrm{i}kr)|n\rangle \qquad [1]$$

$$U_{n'n} = \exp(-\eta^2/2)\sqrt{\frac{n_<!}{(n_< + \Delta n)!}}(\mathrm{i}\eta)^{\Delta n} L_{n_<}^{\Delta n}(\eta^2) \qquad [2]$$

Here $\eta = kx_0 = k\sqrt{\hbar/2m\omega}$ is the Lamb–Dicke parameter for the spatial extension x_0 of the ground state, $n_<$ the smaller of the two numbers n and n', $\Delta n = |n - n'|$, and L_n^m denotes the generalized Laguerre polynomial. In the case of high vibrational excitation $n, n' \gg 1$, $U_{n',n}$ can be approximated as

$$U_{n+\Delta n,n} \approx \mathrm{i}^{\Delta n} J_{\Delta n}(2\eta\sqrt{n}) \qquad [3]$$

with the Bessel function of n-th order J_n. The argument of the Bessel function is the product of the wave number and the classical vibration amplitude of the highly excited state. This is the result one would expect for the frequency-modulated spectrum of a classical oscillator. For arguments β among the Bessel functions $J_n(\beta)$, those of order $n \approx \beta$ have the largest modulus. Consequently, during the initial stages of cooling it is most efficient to tune the laser so that it induces transitions with $\Delta n \approx 2\eta\sqrt{n} \gg 1$.

For the final stages of cooling one can usually assume that the ion is in the Lamb–Dicke regime, i.e., that it is localized within dimensions of the optical wavelength. A detailed study of the limits of sideband cooling in a static trap was carried out in 1984. The final distribution of population over the vibrational states is thermal and can be characterized by a mean occupation number $\langle n\rangle$ or a temperature $T = \hbar\omega/k_B \ln(1 + 1/\langle n\rangle)$. The number $\langle n\rangle$ is a rational function of the four frequencies involved, i.e., the natural linewidth Γ, the oscillation frequency in the trap ω, the laser detuning δ, and the Rabi frequency ω_R. For certain limiting cases, simple expressions for the mean vibrational excitation $\langle n\rangle$ can be derived from the general result. The lowest vibrational excitation is achieved in the case of well-resolved sidebands ($\omega \gg \Gamma$) and under weak laser excitation ($\omega_R \ll \Gamma$), tuned to the first sideband ($\delta = -\omega$):

$$\langle n\rangle = \left(\alpha + \frac{1}{4}\right)\frac{\Gamma^2}{4\omega^2} \qquad [4]$$

The constant α depends on the geometry of excitation and photon re-emission and is of the order of 1. In the case of strong saturation ($\omega_R \gg \Gamma$) of the first sideband ($\delta = -\omega$), the mean vibrational quantum number is proportional to the laser intensity:

$$\langle n\rangle = \frac{\omega_R^2}{8\omega^2} \qquad [5]$$

This case is of practical relevance if one wants to detect a fluorescence signal on the cooling transition, for example, to measure the internal state of the ion in a double-resonance experiment, while keeping the ion cooled. For the choice of parameters $\omega_R = \omega = -\delta$, the ion is still predominantly in the vibrational ground state ($\langle n\rangle \approx 1/8$), but photons are scattered from the wing of the saturation-broadened carrier at a rate $P = \Gamma\omega_R^2/(\Gamma^2 + 2\omega_R^2 + 4\delta^2) \approx \Gamma/6$, which is already one-third of the maximum scattering rate. Finally, in the case of large laser detuning ($\delta \gg \Gamma$, ω, ω_R), the energy is proportional to the detuning because the lower vibrational levels are no longer depleted by the laser:

$$\langle n\rangle \approx \frac{\alpha + 1}{4}\frac{\delta}{\omega} \qquad [6]$$

If the Lamb–Dicke parameter satisfies $\eta \ll 1$, the strength of the sidebands can be calculated from a power series expansion of the matrix elements of eqn [2] and the thermal distribution of population over the oscillator levels. The $U_{n',n}$ are evaluated to first-order in η only. In this approximation, the height of the first lower-frequency sideband in the absorption spectrum relative to the carrier is $\eta^2\langle n\rangle$ and that of the first higher-frequency sideband $\eta^2(\langle n\rangle + 1)$. This information can be used to determine experimentally the vibrational quantum number from the absorption spectrum. The strengths of the higher-order sidebands contain higher powers of η^2 and therefore decrease rapidly.

Let us now briefly review some facts about the motional spectrum of an ion in a Paul trap, which is not just a static system but possesses an explicitly time-dependent potential. The equation of motion for a classical particle of charge e and mass m in the quadrupole potential of the trap $\phi(r,z) = (V_0 - U_0 \cos \Omega t)(r^2 - 2z^2)/2r_0^2$ is given by the Mathieu differential equation:

$$\frac{\delta^2 r}{\delta \tau^2} + (a_r - 2q_r \cos 2\tau)r = 0 \qquad [7]$$

with dimensionless parameters:

$$\tau = \frac{\Omega}{2}t, \qquad a_r = \frac{4eV_0}{m\Omega^2 r_0^2}, \qquad q_r = \frac{2eU_0}{m\Omega^2 r_0^2} \qquad [8]$$

Here a static voltage V_0 and an alternating voltage with amplitude U_0 and frequency Ω are applied to a Paul trap with the radius of the ring electrode r_0. The equation of motion for the z coordinate is obtained by multiplying the parameters a and q by the factor -2. The general solution of the Mathieu equation in the first stability region of the (a,q) parameter space can be written by using the Floquet ansatz:

$$r(\tau) = A\sum_{j=-\infty}^{\infty} c_j \cos(2j+\beta_r)\tau + B\sum_{j=-\infty}^{\infty} c_j \sin(2j+\beta_r)\tau \qquad [9]$$

where A and B are constants determined by the initial conditions, and the c_j and β_r depend on a and q. In the first stability region, where most of the experiments are carried out, the value of β_r is between 0 and 1. According to eqn [9] the motional spectrum of the ion has resonances at

$$\omega_{r,j} = \left(j \pm \frac{\beta_r}{2}\right)\Omega, \qquad j = 0, 1, 2\ldots \qquad [10]$$

In the adiabatic approximation, which is valid for $a, q^2 \ll 1$, the motion of the ion can be separated into an oscillation at the driving frequency Ω, the so-called micromotion, and a slower oscillation, called the secular motion, describing the oscillation in the time-averaged pseudopotential. This pseudopotential describes the effect of the ponderomotive force of the oscillating trapping field that drives the ion to the field minimum at the center of the trap. Since, in the quadrupole trap the pseudopotential is harmonic, the solution to the equation of motion can be approximated as

$$r(\tau) = r_0 \cos\beta_r\tau\left(1 + \frac{q_r}{2}\cos 2\tau\right) \qquad [11]$$

The approximate expression for β_r is $\beta_r = \sqrt{q_r^2/2 + a_r}$ and the secular frequencies are given by

$$\omega_r = \frac{\Omega}{\sqrt{8}}\sqrt{q_r^2 + 2a_r}, \qquad \omega_z = \frac{\Omega}{\sqrt{2}}\sqrt{q_r^2 - 2a_r} \qquad [12]$$

which are the lowest-order resonances $\omega_{r,z} = \beta_{r,z}\Omega/2$ in the Floquet ansatz, eqn [9]. From eqn [11] it is seen that the amplitude of motion at the resonances $\Omega \pm \omega$ is a factor $q_r/4$ smaller than those at the resonance ω. It is noteworthy that the frequency Ω of the driving field does not appear directly in the motional spectrum, but only in combination with the secular frequency ω. This changes, however, if there is an additional static force acting on the ion, displacing it from the center of the quadrupole. The approximate solution to the equation of motion then becomes:

$$r(\tau) = r_1 + r_0 \cos\beta\tau\left(1 + \frac{q_r}{2}\cos 2\tau\right) \qquad [13]$$

and a resonance at Ω appears. The force that displaces the ion by r_1 may be due to static electric stray fields or the presence of a second ion in the trap. These effects can have a strong influence on sideband laser cooling of the trapped ions. A number of theoretical studies have been published on the quantum-mechanical treatment of the Paul trap. The explicitly time-dependent potential does not allow stationary states with time-independent energy eigenvalues. One can, however, as in the case of the pseudopotential model, distinguish between two time-scales and construct wave functions of the harmonic secular oscillator that show some additional breathing motion at the trap frequency. A quantum-mechanical theory of sideband cooling in the time-dependent potential of the Paul trap thus far exists only for special cases; in 1994 there was a case where the ion was localized in the Lamb–Dicke regime at the node of a standing laser wave and time

averaged kinetic energies were calculated. Minima of the energy that agree with the results of the theory for the static harmonic trap to within a factor of 2 were found. The strongest modifications to the predictions of the cooling theory for the static trap arise in the dependence of the energy on the laser detuning. The influence of the micromotion gives rise to heating for certain regions of negative laser detuning (on the high-frequency side of a low-frequency micromotion sideband) and to cooling for some positive detunings (on the low-frequency side of a high-frequency micromotion sideband).

Trapped LASER-cooled ions have a variety of applications in spectroscopy and other fields.

In the following, a phenomenon will be discussed where observations during the cooling process of ions in traps is interpreted as phase transitions between chaotic and ordered ion structures during the cooling process.

Generation of Chaos During the Cooling Process and Phase Transitions to Ordered Configurations

Using a Paul trap, it is possible to study the physics of a few-body phase transition of ions. The ordered state of the trapped ions can be observed directly by monitoring their resonantly scattered light. The transition to an ordered or 'crystalline' configuration is expected when the Coulomb coupling constant which corresponds to the ratio of Coulomb energy to the kinetic energy of a particle, is larger than 1. The ordered structures represent configurations of maximum density for the ions, they are determined by the trap potential pushing the ions towards the trap center, and by the Coulomb repulsion between the ions acting in the opposite direction, so that minimum energy configurations are finally adopted. The system corresponds to an ordered one-component plasma or a strongly coupled plasma.

In the following, computer simulations of the phase transition in the Paul trap will be reviewed. In the trap the ions are subjected to five different forces. These are the force $\mathbf{F}^{(\mathrm{trap})}$, arising from the dynamical rf trapping field, the Coulomb interaction between the ions $\mathbf{F}^{(\mathrm{Coul})}$, the laser cooling force $\mathbf{F}^{(\mathrm{laser})}$, a random force $\mathbf{F}^{(\mathrm{rand})}$ arising from the recoil of the spontaneously emitted photons, and a parasitary force $\mathbf{F}^{(\mathrm{cont})}$, arising from a contact potential, which may be generated unintentionally by locally coating the electrodes with the trapped atoms.

The potential generated by the rf voltage applied to the hyperbolic electrodes of the trap is given by

$$\phi(\mathbf{r},t) = \varphi(t)[x^2 + y^2 - 2z^2]$$
$$\varphi(t) = \frac{U_0 + V_0\cos(\Omega t)}{r_0^2 + 2z_0^2} \qquad [14]$$

where U_0 and V_0 are the dc and ac components of the voltages applied to the trap. From eqn [1] we obtain the trapping force and the resulting equation of motion of a single ion in the trap:

$$\mathbf{F}^{(\mathrm{trap})} = m\ddot{\mathbf{r}} = \mathbf{e}\nabla\phi(\mathbf{r},t) = -2\mathbf{e}\varphi(t)[\mathbf{r} - 3z\mathbf{e}_z] \qquad [15]$$

where m, $\mathbf{r}$, and $z\mathbf{e}_z$ denote the mass of the ion, the position, and the $z\mathbf{e}_z$-component of the position vector, respectively.

The Coulomb interaction between the particles results in the force:

$$\mathbf{F}_i^{(\mathrm{Coul})} = \frac{e^2}{4\pi\varepsilon_0}, \sum_{m\neq 1} \frac{\mathbf{r}_i - \mathbf{r}_m}{[\mathbf{r}_i - \mathbf{r}_m]^3} \qquad [16]$$

In the process of laser cooling, every scattered photon changes the momentum of an ion on the average by an amount $\hbar k$.

The ensemble average of the resulting laser cooling force acting on the i-th ion is thus:

$$\mathbf{F}_i^{(\mathrm{laser})} = \hbar k R[\mathbf{r}_i(t);\, \dot{\mathbf{r}}_i(t)] \qquad [17]$$

The number $N_i(l)$ of spontaneously emitted photons from the i-th ion during the l-th cycle of period $T = 2\pi/\Omega$ is

$$N_i(l)\int_{(l-1)T}^{lT} R[\mathbf{r}_i(t);\dot{\mathbf{r}}_i(t)]\mathrm{d}t \qquad [18]$$

In the simulations, it is assumed that these N_i spontaneous photons are emitted one by one, at the end of each rf cycle, rather than emitting them one by one, according to the appropriate photon-statistics during the cycle. The orientation of the unit vector $\hat{\mathbf{u}}_j$, which points in the direction of the j-th emitted photon, is chosen at random, but weighted with a $\cos^2$ distribution (dipole characteristic) with respect to the laser polarization axis. This way, only the linearly polarized spontaneously emitted photons have been simulated in the calculations. The circularly polarized ones, which occur with a $\sin^2$ distribution, have been neglected. The random force $\mathbf{F}_i^{(\mathrm{rand})}$, is thus:

$$\mathbf{F}_i^{(\mathrm{rand})}(t) = \hbar k \sum_{l=-\infty}^{\infty} \delta(t - lT) \sum_{j=1}^{N_i} \hat{\mathbf{u}}_j \qquad [19]$$

In some of the calculations, a contact potential has been taken into account which is generated experimentally by a coating of the quadrupole electrodes by the trapped particles. Close to the center of the trap, the contact potential gives rise to an approximately homogeneous electric field. Altogether, the motion of the i-th particle is governed by the sum of the above forces, that is:

$$m\ddot{\mathbf{r}}_i = \mathbf{F}_i^{(\mathrm{trap})} + \mathbf{F}_i^{(\mathrm{Coul})} + \mathbf{F}_i^{(\mathrm{laser})} + \mathbf{F}_i^{(\mathrm{rand})} + \mathbf{F}_i^{(\mathrm{cont})} \qquad [20]$$

Given appropriate initial conditions, the eqns of motion [20] are integrated forward in time to obtain the trajectories of the n-ion system. These trajectories provide all the information necessary to find the quantities of interest, such as the occupation probability density of the i-th ion in phase space (i.e., the probability density of finding the i-th ion at a given position, and with a given velocity), the (total) kinetic energy, and the fluorescence intensity.

Some of the results of the described simulations will be summarized below. Figure 1 shows the excitation spectrum of the ions as a function of laser detuning. The experiment confirms the results of the simulations.

Owing to the Coulomb interaction of the ions at particular phases of the cooling process, chaos in the ion cloud is observed leading to a heating process of the ions by the applied field of the trap. This heating mechanism was investigated in the simulations in detail and the summary of the results is given in Figure 2.

For zero laser power and large r, there is no net heating of the ions. This is confirmed by the experiments, in which, even in the absence of a cooling laser, large clouds of ions can be stored in a Paul trap over several hours without being heated out of the trap. The ions are far apart, the Coulomb force is small, and on short time-scales the ions behave essentially like independent singly stored ions. For this reason, this part of the heating diagram is called the 'Mathieu regime' as it is well described by the Mathieu equation [7]. Turning on a small laser power, the rms radius r reduces drastically due to laser

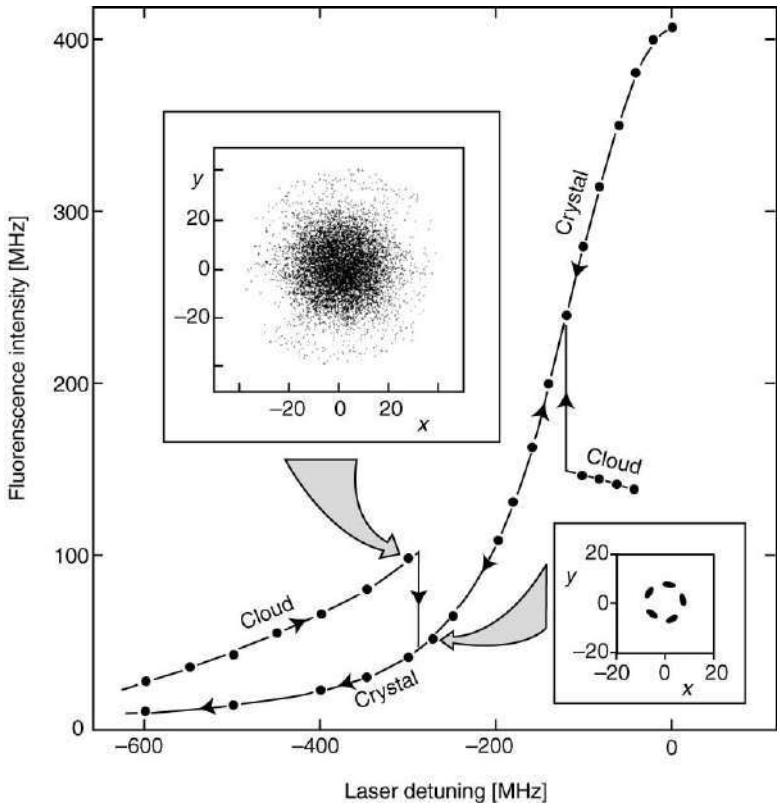

Figure 1 Excitation spectrum of five ions as a function of laser detuning. When the laser detuning is changed the cooling rate of the ion is varying. Two phase transition points (and bistability) are apparent. Insets show the results of the molecular dynamics calculations shortly before and after the jump, which demonstrates directly the existence of a transition from a cloud phase to a crystalline state. The units given at the insets are μm.

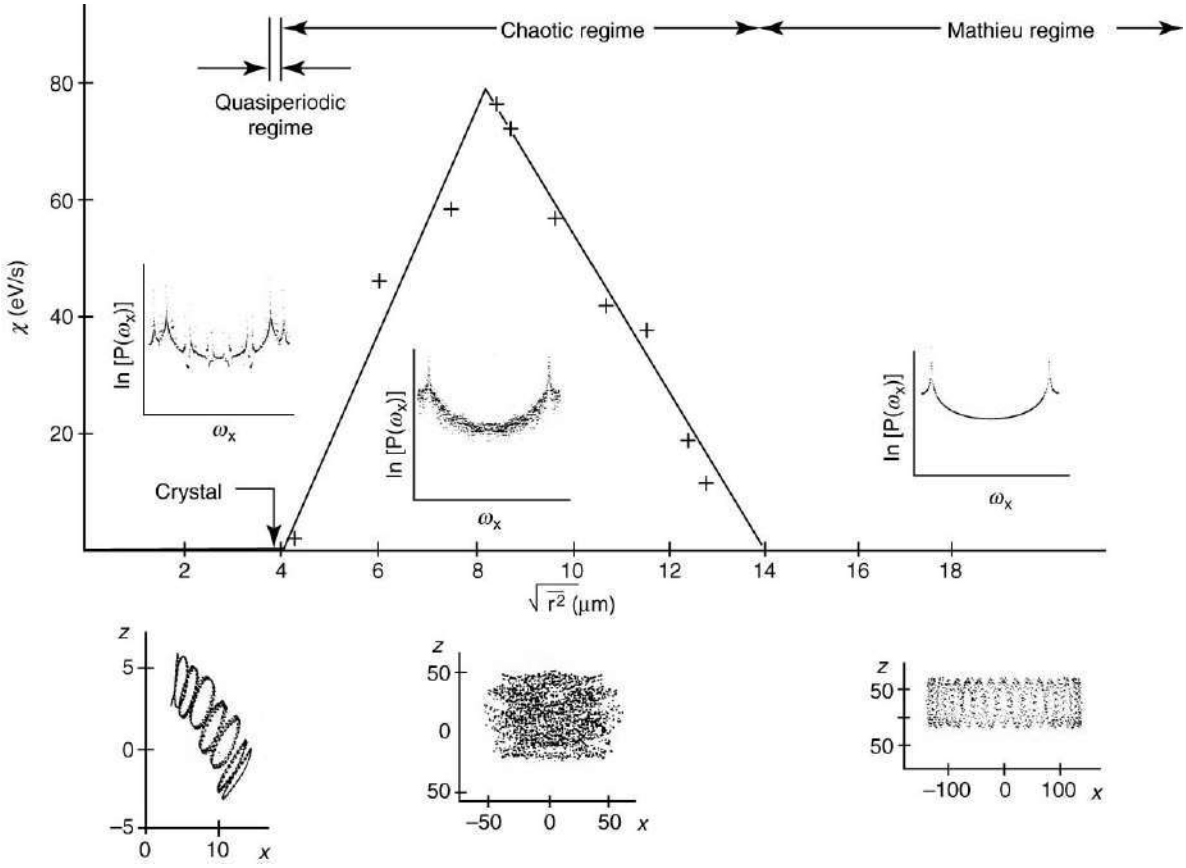

Figure 2 Average heating rate of five ions in a Paul trap versus the rms radius of the ion configuration. The insets show the power spectrum and the corresponding stroboscopic Poincaré sections in the x–z plane of relative separation for two ions in three characteristic domains: the quasiperiodic regime, the chaotic regime and the Mathieu regime. All length scales are in units of micrometers. Units gives at the x–z axis of the insets are in μm.

cooling, but comes to a halt at about 14 μm where chaotic rf heating sets in and balances the cooling power of the laser. Increasing the laser power results in an even smaller cloud. The smaller cloud produces more chaotic rf heating, as seen clearly by the negative slope of the heating curve (Figure 2) in the range $8 < r < 14$ μm.

In the range 4 μm $< r <$ 8 μm, there is still chaotic heating, but the slope of the heating curve is positive. As a consequence of the resulting triangular shape of the heating curve, at about 8 μm, the chaotic heating power can no longer balance the cooling power of the laser, and the cloud collapses into the 'crystalline' state located at $r \approx 3.8$ μm; a 'phase transition' has occurred. Due to this collapse, the behavior of the heating rate in the range 3.8 μm $< r <$ 8 μm cannot be studied by equilibrating laser cooling and rf heating. In this case, we start out from the crystal state and slightly displace the ions to explore the 'vicinity' of the crystal. We observe no heating for 3.8 μm $< r <$ 4 μm, but quasiperiodic motion, and thus name this regime the 'quasi-periodic' regime. We call the upper edge of the quasiperiodic regime ($r \approx 4$ μm) the 'chaos threshold'. An initial condition beyond the chaos threshold, i.e., satisfying $r > 4$ μm, leads to heating and expansion of the ion configuration and numerical data relevant for the shape of the heating curve can be taken during this 'explosion phase'. The laser power P is set to zero for this type of experiment. We conjecture that, apart from the trivial case of a single stored ion, the heating curve is 'universal', i.e., its qualitative shape, including the existence of the chaotic regime, does not depend on the number of simultaneously trapped ions.

For the quasiperiodic, the chaotic and the Mathieu-regime, respectively, the corresponding type of power spectrum are shown in the insets above the abscissa of Figure 2. The data were actually taken from the case of two ions, but would look similar in the five-ion case. A discrete spectrum in the quasi-periodic regime and a complicated

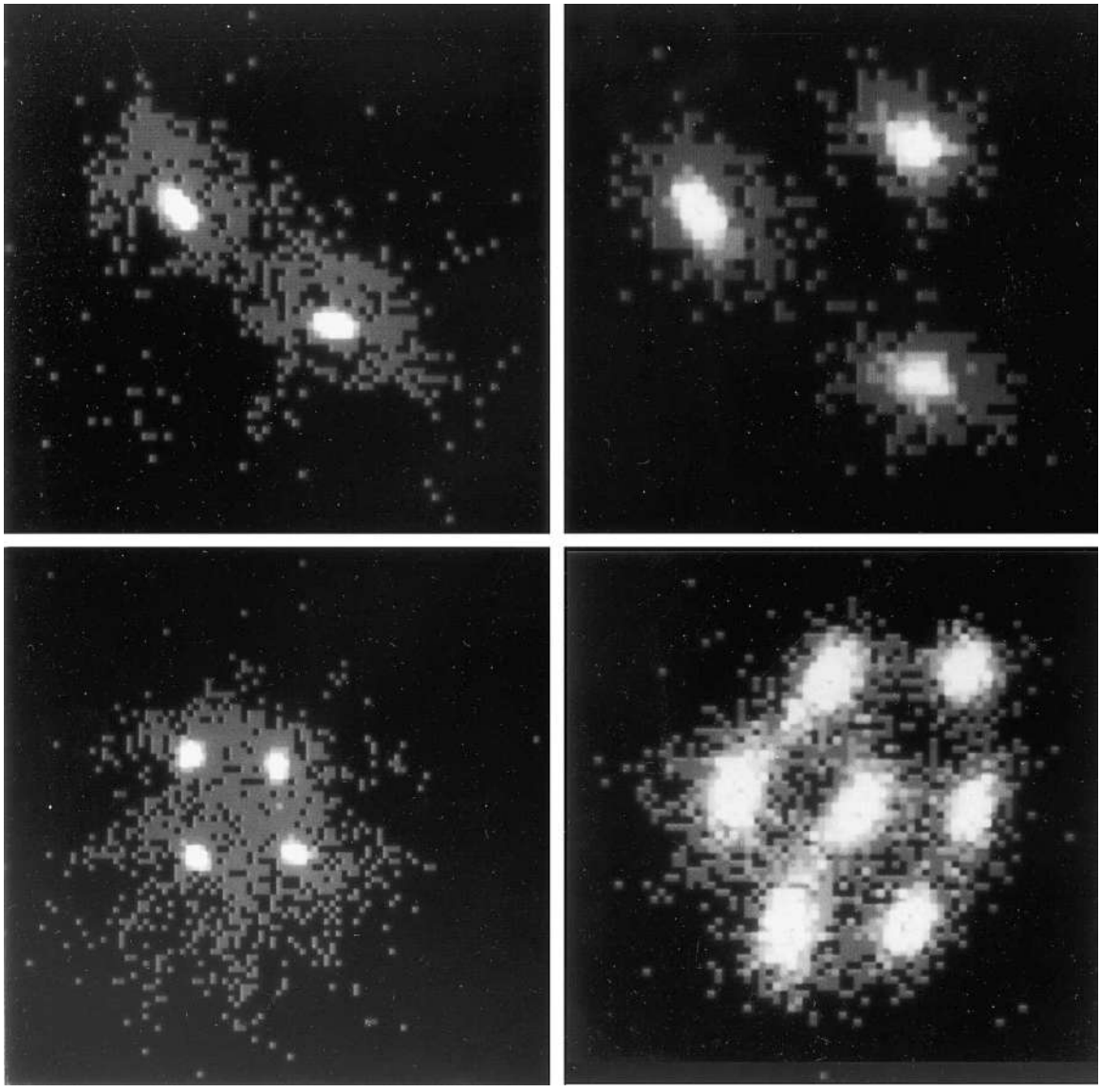

Figure 3 Ordered ion configuration of two, three, four, and seven laser-cooled Mg^+ ions confined in a Paul trap and observed using resonantly scattered light. The average ion separation is 20 μm. The structures of four and seven ions are slightly deformed owing to contact potentials in the trap.

noisy spectrum in the chaotic regime is obtained. The spectrum in the Mathieu regime is again quite simple and dominated by the secular motion frequency. Stroboscopic pictures of the locations of two ions in the $x-y$ plane of the trap characterizing the three regions, are shown on the insets below the abscissa of Figure 2.

Some of the ordered ion structures observed experimentally are shown on Figures 3 and 4. The structures in Figure 3 were obtained in a Paul trap; the others are measured in a storage ring leading to quasi-linear structures of the ion configurations, depending on the symmetry of the trap. Ordered ion structures have also been observed in Penning traps. Due to the symmetry of this trap rotationally symmetric structures are observed.

Conclusion

Laser cooling of ions in traps provides an excellent tool for numerous experiments, varying from tests of basic physics to important and new applications in quantum information processing. The linear ion structure shown on Figure 4 is presently investigated in many laboratories in order to realize gates for

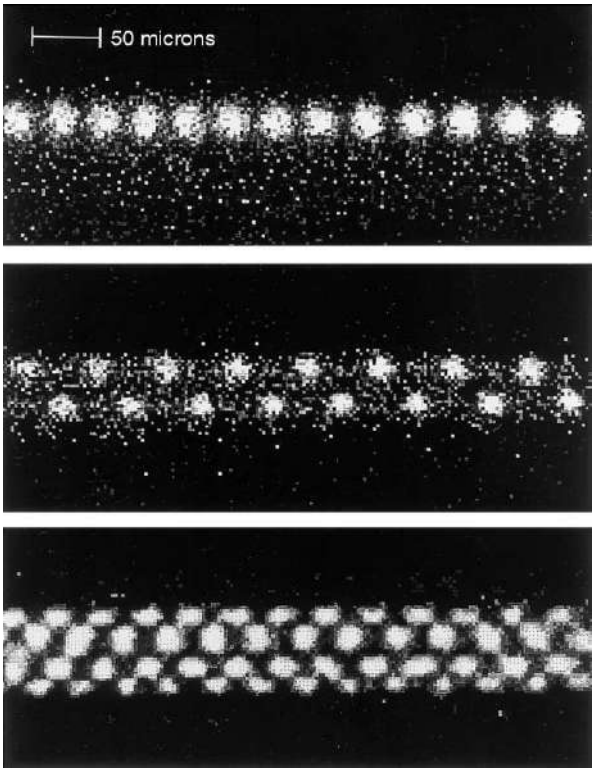

Figure 4 Crystalline structures of laser-cooled $^{24}Mg^{+}$ ions in the quadrupole storage ring. At a low ion density the ions form a string along the field axis (upper). Increasing the ion density transforms the configuration to a zigzag (middle). At still higher ion densities, the ions form ordered helical structures on the surface of a cylinder, e.g., three interwoven helices (lower). As the fluorescent light is projected onto the plane of observation in this case the inner spots are each created by two ions seated on opposite sides of the cylindrical surface, resulting in a single, bright area. (Reproduced by permission of *Nature* from Birkl G, Kassner S and Walther H (1992) Multiple shell structures of laser cooled $^{24}Mg^{+}$ ions in a quadrupole storage ring. *Nature* 357: 310–313.)

quantum computing. Single trapped ions in cavities may be used as deterministic light sources for secure single photon communication links. Furthermore, in many laboratories single laser-cooled ions are investigated as the basis for new time and frequency standards, being several orders of magnitude more accurate than our present atomic clock.

See also

Chaos in Nonlinear Optics. **Quantum Optics:** Entanglement and Quantum Information; Quantum Computing with Atoms.

Further Reading

Arimondo E, Philips WD and Strumia F (eds) (1992) Laser manipulation of atoms and ions. *Proceedings of the International School of Physics Enrico Fermi, Course CXVIII*. Amsterdam, Oxford, New York, Tokyo: North Holland.

Ashkin A (1970) Atomic-beam deflection by resonance-radiation pressure. *Physical Review Letters* 25: 1321–1324.

Birkl G, Kassner S and Walther H (1992) Multiple shell structures of laser cooled $^{24}Mg^{+}$ ions in a quadrupole storage ring. *Nature* 357: 310–313.

Blümel R, Chen JM, Peik E, Quint W, Schleich W, Shen YR and Walther H (1988) Phase transitions of stored laser-cooled ions. *Nature* 334: 309–313.

Blümel R, Kappler C, Quint W and Walther H (1989) Chaos and order of laser-cooled ions in a Paul trap. *Physical Review A* 40: 808–823.

Bollinger JJ, Kriesel JM, Mitchell TB, King LB, Jensen MJ, Itano WM and Dubin DHE (2003) Laser-cooled ion plasmas in Penning traps. *Journal of Physics B: Atomic Molecular Optical Physical* 36: 499–510.

Cirac JI, Garay LJ, Blatt R, Parkins AS and Zoller P (1994) Laser cooling of trapped ions: the influence of micro-motion. *Physical Review A* 49: 421–432.

Dehmelt H (1967 and 1969) In: Bates DR and Estermann I (eds) *Advances in Atomic and Molecular Physics*, vols. 3 and 5. New York: Academic Press.

Dehmelt HG (1983) Stored ion spectroscopy – three lectures. In: Arecchi FT, Strumia F and Walther H (eds) *Advances in Laser Spectroscopy*. New York: Plenum.

Diedrich F, Peik E, Chen JM, Quint W and Walther H (1987) Observation of a phase transition of stored laser-cooled ions. *Physical Review Letters* 59: 2931–2934.

Glauber R (1991) In: Arimondo E and Phillips W (eds) *Proceedings of the International School of Physics 'Enrico Fermi,'* vol. 118. Amsterdam: North Holland.

Hänsch TW and Schawlow AL (1975) Cooling of gases by laser radiation. *Optics Communications* 13: 68–69.

Javanainen J, Lindberg M and Stenholm S (1984) Laser cooling of trapped ions: dynamics of the final stages. *Journal of Optical Society of America B* 1: 111–115.

Leibfried D, Blatt R, Monroe C and Wineland D (2004) Quantum dynamics of single trapped ions. *Review of Modern Physics* 75: 281–324.

Peik E, Abel A, Becker Th, von Zanthier J and Walther H (1999) Sideband cooling of ions in radio-frequency traps. *Physical Review A* 60: 439–449.

Schieder R, Walther H and Wöste L (1972) Atomic beam detection by the light of a tunable dye laser. *Optics Communications* 5: 337–340.

Wineland D and Dehmelt H (1975) Proposed 10–14 laser fluorescence spectroscopy on Tl^{+} mono-ion oscillator. *Bulletin of the American Physical Society* 20: 637.

Quantum Computing with Atoms

S F Huelga, University of Hertfordshire, Hatfield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

It is a well-known fact that the speed of commercial microprocessors is doubling, roughly, every 18 months (Moore's law). This speedup is achieved mainly at the expense of reducing the size of the microchip. If this tendency is to continue, it seems unavoidable that quantum mechanical effects will eventually become important.

Remarkably, far from being a technological nuisance, the fact that information carriers are truly quantum mechanical systems offers an entirely new paradigm for computation. As opposed to classical bits, with logical values 0 and 1, quantum bits (qubits) can be prepared in arbitrary, normalized superpositions of the logic values. In addition, quantum states of multiqubit systems can be entangled, i.e., cannot be written as tensor products of states of the individual components. The theoretical description of entanglement requires a huge number of parameters which, in fact, grows exponentially with the number of particles that are involved. This exponential number of new degrees of freedom is responsible for the computational potential that quantum systems offer. Recently it has been demonstrated theoretically how we can harness this potential in the form of quantum algorithms, including Shor's factorization algorithm and Grover's protocol for data base search.

From a theoretical point of view, the implementation of quantum computation is greatly simplified by the fact that any unitary transformation involving an arbitrarily large number of qubits can be decomposed into two building blocks: single qubit operations and two-qubit controlled-NOT gates (CNOT-gates). We only need to identify two-level systems whose interaction can be mediated in a way that allows for conditional dynamics, that is, the second system undergoes certain operation (a NOT operation in our case), conditional on the state of the first qubit. However, it is easy to convince ourselves that from the experimental point of view, the challenge is going to be formidable. The following (DiVincenzo)-checklist summarizes the necessary requirements for the potential physical realization of a quantum computer:

- The system needs to provide a well-defined qubit, with two robust levels to represent the logical values 0 and 1. Moreover, the system should be scalable.
- Both the initial preparation and the final readout of the system has to be done with sufficient accuracy.
- The system has to allow the implementation of a universal set of gates (single-qubit rotations and CNOTs gates) with sufficient high fidelity. Implicit in this requirement is the ability to address individual qubits.

From the above list it is clear that the practical realization of a quantum computer relies on the ability to achieve a delicate equilibrium: the system needs to interact strongly with selected external agents, typically laser light, both during the gate operations and the preparation and readout steps, but should be shielded from any unwanted interaction if the desired dynamics is to be preserved. Perfect, selective, shielding from the environment is not possible in practice and unwanted interactions with the surroundings will unavoidably result in some form of decoherence. In this article we will review the fundamentals of the implementation of a universal set of quantum gates using (i) atomic system coupled to a resonant field sustained in a microwave cavity and (ii) trapped ions confined in an electromagnetic trap. Atomic and ionic systems provide a good means for storing quantum information in long-lived electronic states. In addition, both neutral atoms and ions can be selectively detected with high efficiency. While none of the presented schemes is expected to be scalable to a very large number of qubits, they constitute the first prototypes for quantum registers and are expected to play an important role in applications where only a moderate number of qubits are required. These applications range from the development of new frequency standards to quantum communication protocols. We will discuss in the following sections how a set of universal gates can build within these systems and will present the state of the art of the experiments performed to date.

Quantum Computation: Basic Definitions

As stated before, the basic unit of quantum information is a qubit, an arbitrary quantum mechanical two-level system. We will associate logical values 0 and 1 with quantum states $|0\rangle$ and $|1\rangle$, the

eigenvectors of an observable such as the energy, and refer to such states as the computational basis. Any unitary operation on a qubit will be called a single qubit gate. Joint operations on more than one qubit will be denoted as multi-qubit gates. Using the fact that any two-level system is isomorphic to a spin-1/2 particle (e.g., an electron), we can always think of a qubit in terms of a certain pseudo-spin whose dynamics can be represented as rotation of the associated Bloch vector within the Bloch sphere.

For instance, as depicted in Figure 1, a qubit initially prepared in state $|1\rangle$, after experiencing a single qubit gate will be represented by certain Bloch vector $\vec{S}$, whose location on the Bloch sphere is a function of the gate parameters. Note that a classical system can only assume two positions (upwards and downwards), while a quantum system can explore the whole surface of the sphere.

It is also quite standard in quantum computation to use a diagramatic representation in terms of quantum circuits. Quantum circuits are built up of wires to denote the quantum bits and boxes to represent the gates, or quantum operations. Figure 2 shows two basics elements of a quantum circuit. The element on the left is the so-called Hadamard transformation, a particular example of single-qubit gate which transform each element of the computational basis as follows:

$$|0\rangle \rightarrow \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)$$
$$|1\rangle \rightarrow \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right) \qquad [1]$$

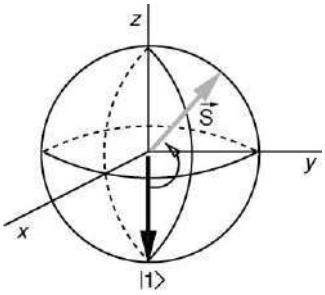

Figure 1 Geometrical representation of qubit systems using the Bloch sphere formalism. See the text for details.

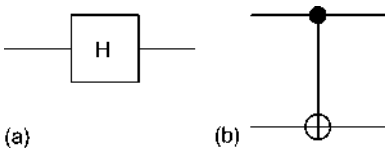

Figure 2 Examples of single qubit gates (a) Hadamard transformation and two qubit gates (b) CNOT gate as customarily represented in quantum circuits diagrams.

Part (b) of Figure 2 depicts a CNOT-gate. This is now an operation involving two qubits. The first one, the control qubit, remains unaltered after the gate operation while the second qubit, referred to as the target, undergoes a spin flip (logical NOT operation) only when the control is in the state $|1\rangle$. Specifically, the action of a CNOT-gate is defined by the following transformation rule:

$$\begin{aligned} &\text{CNOT} \\ &|00\rangle \rightarrow |00\rangle \\ &|01\rangle \rightarrow |01\rangle \\ &|10\rangle \rightarrow |11\rangle \\ &|11\rangle \rightarrow |10\rangle \end{aligned} \qquad [2]$$

We have already mentioned that single qubit rotations and CNOT gates provide a universal set of gates. Our aim now is to show how such operations can be built when the so far generic qubit are realized in terms of real atomic systems. In the two experimental scenarios we are going to analyze here, the CNOT operation will be achieved by means of a different type of gate, a controlled-phase gate supplemented by two one-qubit Hadamard gates on the target qubit. The arrangement is showed in Figure 3. The controlled-phase gate is again a two-qubit operation defined by the following transformation rule:

$$\begin{aligned} &\text{CNOT} \\ &|00\rangle \rightarrow |00\rangle \\ &|01\rangle \rightarrow |01\rangle \\ &|10\rangle \rightarrow |10\rangle \\ &|11\rangle \rightarrow -|11\rangle, \end{aligned} \qquad [3]$$

i.e., a global phase equal to π is acquired if, and only if, the two basis states are in the logical state 1. It is easy to check that when the target bit is rotated by a Hadamard gate prior and subsequently to the action

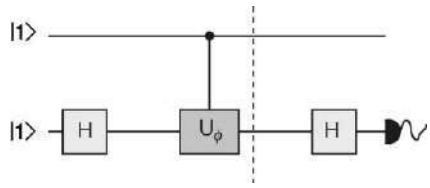

Figure 3 Controlled-phase gate with both inputs, control, and target qubits, in state $|1\rangle$. The first Hadamard gate rotates the target to state $(|0\rangle - |1\rangle)/\sqrt{2}$. The action of the controlled-phase gate with $\phi = \pi$ yields the joint state $|1\rangle + (|0\rangle - |1\rangle)/\sqrt{2}$ and a final Hadamard rotation on the target leaves the system in the product state $|1\rangle|0\rangle$.

of the joint phase gate, the resulting state is tantamount to the action of a single CNOT gate. Therefore, to prove that our atomic system allows for the implementation of a universal set of gates, it suffices to show that both single qubit rotations and controlled phase gates can be implemented with sufficient accuracy.

The next section analyzes a typical cavity QED setup and presents the basic physical ideas to understand how single and two-qubit gates can be implemented within atomic systems interacting with radiation fields. Subsequent sections will rely heavily in this analysis.

Quantum Computing with Atoms I: Cavity QED

A simplified cavity-QED arrangement for performing both single- and two-qubit gate operations is depicted in Figure 4. This configuration is well-known to quantum opticians and it is often referred to as a Ramsey interferometric setup. For concreteness, we will focus here in the experimental set up currently in operation at the ENS in Paris. There, the atomic system is a highly excited Rydberg atom initially prepared in a long-lived circular state. The relevant atomic states are represented in Figure 5, with the qubit being stored in levels $|i\rangle$ and $|g\rangle$ which will play the role of the general basis states $|0\rangle$ and $|1\rangle$.

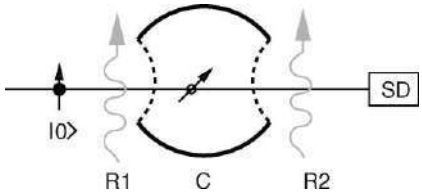

Figure 4 Cavity QED setup for single-qubit and two-qubit operations. Rotations of the atomic pseudo spin are performed via interation with classical radiation (Zone R), while the interaction with a resonant cavity field sustained in C allows the realization of a phase gate where the cavity photon acts as the control qubit.

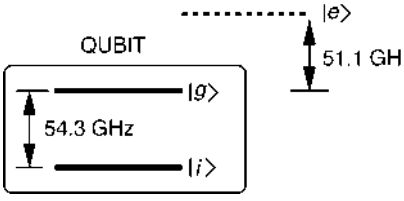

Figure 5 Circular Rydberg levels involved in the ENS experimental setup. The qubit is stored in levels $|i\rangle$ (logical value 0) and $|g\rangle$ (logical value 1). The radiative lifetime of these levels lies in the millisecond range. Dissipation can then be neglected along the 20 cm path inside the apparatus.

Regions denoted by R represent the interaction of the atomic system with classical microwave radiation. As detailed below, this interaction will allow arbitrary rotations of the atomic pseudo spin. Region C represents the interaction of the atomic system with a single mode cavity field resonant the atomic transition $g - e$. The highest Q value reported to date is of the order of 10^8, which corresponds to a photon storage time of a few milliseconds. Therefore, the relaxation time of the cavity field is much longer than the atom-cavity interaction time, which defines the two-qubit gate operation time and is of the order of a few tens of μs. The cavity field can be prepared in either of the Fock states $|0\rangle_\gamma$ or $|1\rangle_\gamma$ and will play the role of the control qubit for the realization of a phase gate, as detailed below. This set up is completed with a detection scheme (SD in the figure) that allows the selective detection of the atomic electronic state.

We will start our discussion showing how to perform an arbitrary single-qubit gate by means of irradiating the atomic system with classical radiation. We will keep the analysis in very general terms with the aim of using it later in the description of the implementation of quantum logic with trapped ions.

Single-Qubit Rotations

Let our qubit be a two-level atomic system with frequency separation ω_0. We will show that an arbitrary rotation can be achieved by means of irradiating the system with classical light, e.g., microwave radiation or a laser field, with tunable frequency ω_L. The radiation field will be specified by two controllable parameters, its Rabi frequency Ω and its phase ϕ. The atom-field Hamiltonian can be written as

$$H = \frac{\hbar\omega_0}{2}\left(|1\rangle\langle 1| - |0\rangle\langle 0|\right) + \hbar\left(g e^{i\omega_L t}|0\rangle\langle 1| + g^* e^{-i\omega_L t}|1\rangle\langle 0|\right) \qquad [4]$$

where the first term corresponds to the atomic internal energy and the second describes the coupling with the external field. The coupling constant g can be written in terms of the field parameters as

$$g = \frac{i\Omega}{2} e^{-i\phi} \qquad [5]$$

In order to write down a time-independent Hamiltonian, we will move to an interaction picture with

respect to $H_0 = \hbar\omega_L/2(|1\rangle\langle 1| - |0\rangle\langle 0|)$. In this frame,

$$H_{\text{int}} = \frac{\hbar\delta}{2}\Big(|1\rangle\langle 1| - |0\rangle\langle 0|\Big) + \hbar\Big(g|0\rangle\langle 1| + g^*|1\rangle\langle 0|\Big) \quad [6]$$

The interaction Hamiltonian can be readily expressed in terms of the Pauli spin operators

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad [7]$$

as

$$H_{\text{int}} = \frac{\hbar\delta}{2}\sigma_z + \frac{\hbar\Omega}{2}(\sin\phi\,\sigma_x - \cos\phi\,\sigma_y) \quad [8]$$

where $\delta = \omega_L - \omega_0$ denotes the detuning from the atomic transition and we have taken into account the explicit form of g as given by eqn [5]. We can now easily derive the time evolution operator by noting that H_{int} can be written in a compact form as $H_{\text{int}} = \vec{\sigma}\cdot\vec{n}$, where the vector $\vec{n}$ has cartesian coordinates $\vec{n} = (\Omega/2\sin\phi, -\Omega/2\cos\phi, \delta/2)$. Note that $|\vec{n}| \equiv \Omega_g = 1/2\sqrt{\Omega^2 + \delta^2}$. The time evolution operator can then be written as follows:

$$U = e^{-iH_{\text{int}}t} = \cos(|\vec{n}|t)1\!\!1 - i\sin(|\vec{n}|t)\vec{\sigma}\cdot\vec{n}/|\vec{n}| \quad [9]$$

In matrix form:

$$U = \begin{pmatrix} \cos\left(\frac{\Omega_g t}{2}\right) - i\frac{\delta}{\Omega_g}\sin\left(\frac{\Omega_g t}{2}\right) & \frac{\Omega}{\Omega_g}\sin\left(\frac{\Omega_g t}{2}\right)e^{-i\phi} \\ -\frac{\Omega}{\Omega_g}\sin\left(\frac{\Omega_g t}{2}\right)e^{i\phi} & \cos\left(\frac{\Omega_g t}{2}\right) + i\frac{\delta}{\Omega_g}\sin\left(\frac{\Omega_g t}{2}\right) \end{pmatrix} \quad [10]$$

In the limit of large Rabi frequency, where $\Omega \gg \delta$, we can write the following approximate expressions for the time evolution of the computational basis states

$$\begin{aligned} U|0\rangle &= \cos\left(\frac{\Omega_g t}{2}\right)|0\rangle - \sin\left(\frac{\Omega_g t}{2}\right)e^{i\phi}|1\rangle \\ U|1\rangle &= \sin\left(\frac{\Omega_g t}{2}\right)e^{-i\phi}|0\rangle + \cos\left(\frac{\Omega_g t}{2}\right)|1\rangle \end{aligned} \quad [11]$$

The transformed states by U are eigenstates of the spin operator along the direction $\vec{n}$ which is a function of the external field parameters. Therefore, by adjusting the Rabi frequency and the field's phase, it is possible to transform the basis states into spin eigenstates along any arbitrary direction, and to perform an arbitrary rotation of our qubit, we simply have to tailor appropriately the parameters of the driving field. For example, when the Rabi frequency and the interaction time are chosen such that $\Omega t/2 = \pi/4$, and the field's phase is taken equal to π, we get

$$\begin{aligned} U|0\rangle &= \frac{1}{2}\Big(|0\rangle + |1\rangle\Big) \\ U|1\rangle &= \frac{1}{2}\Big(|0\rangle - |1\rangle\Big) \end{aligned} \quad [12]$$

This relationship defines the Hadamard transformation H introduced before. Alternatively, using quantum optics jargon, we have performed a $\pi/2$-pulse. Note that using the Bloch sphere representation, we can visualize the action of a Hadamard gate on the computational basis states as a rotations that brings them to the equatorial plane of the Bloch sphere, the transformed states by H now lying in opposite directions along the x-axis.

The next step is to show that our system also allows the implementation of conditional dynamics and, in particular, allows to build a controlled-phase gate.

Conditional Dynamics: Controlled-Phase Gate

The interaction of the atomic system with the cavity field sustained by the microwave cavity is described by the Jaynes–Cummings (JC) Hamiltonian. We can see this Hamilton operator as a generalization of the operator defined by eqn [4], where the radiation field is now quantized and the terms $e^{\pm i\omega_L t}$ are replaced by the operators a and $a^\dagger$ corresponding to the annihilation and creation of a cavity photon of frequency ω. The interaction with the quantized cavity field couples levels $|e\rangle$ and $|g\rangle$, while the far detuned level $|i\rangle$ will now play the role of a passive spectator. Remember that our atomic qubit is stored in levels $|i\rangle$ and $|g\rangle$. To avoid confusion with the notation, we will write the Hamilton operator in terms of the atomic states e and g. Explicitly:

$$\begin{aligned} H_{\text{JC}} = &\frac{\hbar\omega_0}{2}\Big(|e\rangle\langle e| - |g\rangle\langle g|\Big) + \hbar\omega a^\dagger a \\ &- i\frac{\hbar\Omega}{2}\Big(a|g\rangle\langle e| - a^\dagger|g\rangle\langle e|\Big) \end{aligned} \quad [13]$$

where the coupling constant g has been taken to be $g = -i(\hbar\Omega/2)$, proportional to the Rabi frequency averaged over the mode structure, Ω. Using the same techniques employed in the previous section, we can derive the corresponding JC time evolution operator,

which yields, for exact resonance, the following transformation rules:

$$\begin{aligned} |g\rangle|0\rangle_\gamma &\rightarrow |g\rangle|0\rangle_\gamma \\ |g\rangle|1\rangle_\gamma &\rightarrow \cos\frac{\Omega t}{2}|g\rangle|1\rangle_\gamma - \sin\frac{\Omega t}{2}|e\rangle|0\rangle_\gamma \\ |e\rangle|0\rangle_\gamma &\rightarrow \cos\frac{\Omega t}{2}|e\rangle|0\rangle_\gamma + \sin\frac{\Omega t}{2}|g\rangle|1\rangle_\gamma \end{aligned} \quad [14]$$

where states $|\rangle_\gamma$ refer to the cavity field. Note that:

- The cavity field will play the role of our control qubit in the performance of a phase gate, the target qubit being the atomic system with computational states defined as before. The computational basis for the photonic qubit is provided by the Fock states $|0\rangle_\gamma$ and $|1\rangle_\gamma$.
- Strictly speaking, the time t entering in the equations above corresponds to an effective interaction time which accounts for the spatial variation of the coupling along the field mode.
- When the atom enters the cavity in state $|i\rangle$, it will acquire a global phase proportional to the ratio Ω^2/δ_i, where δ_i denotes the detuning between the atomic transition $e \leftrightarrow i$ and the mode frequency ω. In conditions of resonance, or quasi-resonance, this factor can be safely neglected, which justifies the term *passive spectator* used before.

We now have all the ingredients to understand how we can perform a controlled phase gate. Consider the situation where $\Omega t = 2\pi$, so the system experiences a full Rabi oscillation. According to the previous discussion, the basis states of two-qubit system atom-cavity field will transform as follows:

$$\begin{aligned} |i\rangle|0\rangle_\gamma &\rightarrow |i\rangle|0\rangle_\gamma \\ |i\rangle|1\rangle_\gamma &\rightarrow |g\rangle|0\rangle_\gamma \\ |g\rangle|0\rangle_\gamma &\rightarrow |g\rangle|0\rangle_\gamma \\ |g\rangle|1\rangle_\gamma &\rightarrow -|g\rangle|1\rangle_\gamma \end{aligned} \quad [15]$$

That is, whenever the atomic system enters the cavity in state $|i\rangle$, it will be left unaffected, irrespective of the photon state. However, if the atom enters the cavity in state $|g\rangle$, the system will undergo a phase shift equal to π conditional to the photon number. And that is precisely the dynamics required for a controlled phase gate as defined earlier.

A detailed account of the experiments in cavity QED, showing coherent gate operation, can be found in Further Reading. These experiments prove the working principle of a universal set of quantum logic gates with neutral atoms and provide the basic elements for the development of a cavity QED quantum register in the near future.

Quantum Computing with Atoms II: Ion Traps

Charged atomic particles, ions, provided the first atomic system where conditional dynamics was demonstrated experimentally. The theoretical proposal in 1995, by Cirac and Zoller, for performing a two-qubit controlled NOT gate between trapped ions cooled to their motional ground state, was followed a few months later by a pioneering experiment at NIST, in Colorado. While the physics involved in the description of laser cooled and trapped ions is conceptually more involved than the dynamics ruled by a purely JC Hamiltonian, we can review the rudiments of the Cirac–Zoller scheme on the basis of the techniques developed in the previous sections. Nevertheless, as discussed briefly at the end of this section, we should point out that the exact procedure followed in the actual experiments departs from the analysis presented here.

Figure 6 depicts a linear ion trap where a string of ions is confined in an essentially 1-D structure. Quantum bits are stored in two long-lived hyperfine states which can be driven with laser light of controllable Rabi frequency Ω and phase ϕ. Ideally, the separation d between ions exceeds the wavelength of the external field so that ions can be addressed individually. Detection of the atomic state relies on the well mastered quantum jump technique which provides an efficiency close to 100%. From our previous discussion, it is clear that we already have all the ingredients for performing arbitrary single qubits gates, but we still need to provide a procedure to achieve two-qubit gate operations. As detailed in the next section, this will be achieved using the motional degrees of freedom of the ions as a bus for information transfer between an arbitrary pair of ions.

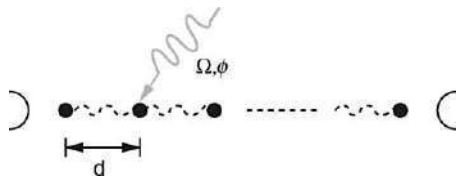

Figure 6 Schematic representation of a linear ion trap. Ions are individually addressed by external laser fields to perform single qubit operations, while two-qubit gates are mediated by the coupling to the motional degrees of freedom, depicted here as mechanical strings linking the chain of ions in the trap. The exact duration of a two-qubit gate operation and the characteristic decoherence time depend very much on the trap design, but orientative values are in the range of 50 μs and 100 − 1000 μs respectively.

The Cirac and Zoller Quantum Computer

At sufficiently low temperatures, the process of absorption and emission of photons produces a non-negligible recoil of the atomic system and the ion's motion can no longer be treated classically. In this regime, we can think of the string of ions as a system of coupled quantum harmonic oscillators whose normal mode configurations correspond to the collective motional states of the ions. The lowest energy solution corresponds to an oscillation where all ions are in phase (center of mass (CM) mode) and will be the one considered in our description. Figure 7 depicts the level scheme for ions initially cooled down to the motional ground state, illustrating both internal levels, $|e\rangle$ and $|g\rangle$, and motional (vibrational) states, denoted by $0, 1, \ldots$ Laser radiation can be tuned to be resonant with the $|g\rangle|0\rangle_{cm} \leftrightarrow |e\rangle|0\rangle_{cm}$ transition (carrier transition), but may also be tuned to drive transitions with different vibrational states (sideband transition), as represented by the external arrows in the figure. The realization of single qubit gates simply involves tuning the laser towards the carrier transition. The interaction is ruled by essentially the same Hamiltonian discussed above. Therefore, choosing appropriately the laser parameters, it is possible to induce an arbitrary rotation of the ion pseudo-spin.

To prove that the system allows the implementation of a universal set of gates, we will now show that it is possible to operate a controlled phase gate between two arbitrary ions. For that we need to couple internal and motional degrees of freedom which will be done by means of tuning the laser light to a sideband transition. The corresponding Hamiltonian will not be derived here in detail but it can be understood as describing a Jaynes–Cummings form of dynamics where the quantized electromagnetic field within a cavity is replaced by the quantum mechanical oscillator describing the vibrational states of the ion. Explicitly:

$$H = \frac{\eta\Omega}{2}\left(c|e\rangle\langle g| + c^{\dagger}|g\rangle\langle e|\right) \qquad [16]$$

where c and $c^{\dagger}$ denote the annihilation and creation operators of vibrational quanta and η is the so-called Lamb–Dicke parameter, a measure of the recoil energy. The gate operation involves the following sequence:

- Mapping the internal state of the control ion into the collective motional state of the chain,
- performing an operation on the target ion conditional to the excitation number of the vibrational state and
- reverse the initial operation restoring the vibrational state to its ground state.

Imagine ion i, which will play the role of the control, and j, the target, initially prepared in arbitrary superpositions of the corresponding internal states, with the whole system cooled down to their motional ground state $|0\rangle_{cm}$. The joint state of the two particles can be written as a product state of the form:

$$\left(\alpha_i|g\rangle_i + \beta_i|e\rangle_i\right) \otimes |0\rangle_{cm} \otimes \left(\alpha_j|g\rangle_j + \beta_j|e\rangle_j\right) \qquad [17]$$

When ion i is irradiated by a laser tuned to the first motional side band $|g\rangle|1\rangle_{cm} \leftrightarrow |e\rangle|0\rangle_{cm}$, and the timing is such that the system undergoes a π-Rabi rotation, the internal state of the ion i is mapped onto the vibrational state (Note the transformation rules given in eqn [14]. As a result, the joint state of the system is still factorizable as

$$\left(|g\rangle_i \otimes \alpha_i|0\rangle_{cm} + \beta_i|1\rangle_{cm}\right) \otimes \left(\alpha_j|g\rangle_j + \beta_j|e\rangle_j\right) \qquad [18]$$

The second step of the protocol consists of applying an operation on ion j conditional on the vibrational state of the joint system. For that we apply a laser pulse resonant with the transition $|e\rangle_j|1\rangle_{cm} \leftrightarrow |a\rangle_j|0\rangle_{cm}$ with an auxiliary level $|a\rangle_j$ so that ion j undergoes a full Rabi oscillation. This results in acquiring a phase shift equal to π only if the system is state $|e\rangle_j|1\rangle_{cm}$ (see eqn [15]). This step entangles the internal and vibrational degrees of freedom of ion j, leading to a joint state of

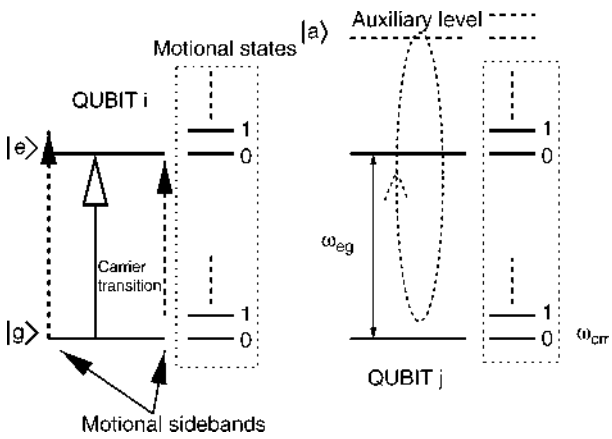

Figure 7 Relevant internal and motional atomic levels for the realization of a controlled NOT gate between two arbitrary ions within a chain. See the text for details.

the form:

$$|g\rangle_i \otimes \Big(\alpha_i|0\rangle_{cm}(\alpha_j|g\rangle_j + \beta_j|e\rangle_j) + \beta_i|1\rangle_{cm}(\alpha_j|g\rangle_j - \beta_j|e\rangle_j)\Big) \quad [19]$$

Applying a final laser pulse on ion i identical to the initial one, the state of the vibrational mode is transferred back to ion i and the atomic systems are left in the, generally, entangled state

$$\alpha_i\alpha_j|g\rangle_i|g\rangle_j + \alpha_i\beta_j|g\rangle_i|e\rangle_i + \beta_i\alpha_j|e\rangle_i|g\rangle_j - \beta_i\beta_j|e\rangle_i|e\rangle_j, \quad [20]$$

which completes the controlled-phase gate operation.

A simplified version of this scheme shortly followed the theoretical proposal by Cirac and Zoller, proving the working principle of the ions trap quantum processor. Since then, both theoretical and experimental progress have allowed the development of schemes for implementing faster gates as well as protocols which do not require cooling to the motional ground state. We will briefly revise the current state of the art and future prospects of this technology in the last section of this article.

Frequency Estimations as a Quantum Computation: Building Frequency Standards with Entangled States

Recently, it has been proposed to make use of entangled states of trapped ions for performing high-resolution spectroscopy beyond the shot noise limit. This boundary is the ultimate limit achievable when performing a frequency measurement using uncorrelated particles. Consider n atomic two-level systems with frequency separation ω_0 initially prepared in their ground state $|0\rangle$. The aim is to lock a certain oscillator of frequency ω_L to the atomic frequency. To do that, the currently employed protocols use the so-called Ramsey method of separate oscillating fields, where each atomic system undergoes two $\pi/2$-pulses (or Hadamard transformations) separated by a free time evolution of duration t, much larger than the duration of the pulse. Looking at the Hamiltonian describing the interaction between the ion and the field, eqn [1], one can easily see that the free time evolution can be understood as a controlled phase gate in the following sense. If we set $g = 0$ in eqn [1] and for simplicity re-scale the ion internal energies so that $H = \hbar\delta|1\rangle\langle 1|$, the free evolution operator can be written as $U = -\hbar\delta|1\rangle\langle 1|$, so that the atomic basis states evolve as:

$$U|0\rangle = |0\rangle \qquad U|1\rangle = e^{-i\delta t}|1\rangle \quad [21]$$

In other words, the ground state remains invariant, whereas the upper levels pick up a phase which is proportional to the detuning from the atomic transition. The state of each particle after the time evolution can then be written as

$$\frac{1}{\sqrt{2}}\Big(|0\rangle + e^{-i\delta t}|1\rangle\Big) \quad [22]$$

After the second pulse, which again mixes populations of levels $|0\rangle$ and $|1\rangle$, one measures the population of, say, level $|1\rangle$. The signal is therefore a sinusoidal function of the form

$$P_1 = \frac{1}{2}\Big(1 + \cos(\delta t)\Big) \quad [23]$$

which is symmetrical with respect to $\delta = 0$. This procedure is repeated a number of times to accumulate a sufficient data. Let us call T the total duration of the procedure, so that we accumulate $N = nT/t$ experimental data. Finally, whenever statistics in the populations are different from 50:50 at half maximum are detected, the oscillator frequency is steered to achieve exact resonance. The ultimate resolution $\Delta\omega_0$ achievable with this technique is limited simply by projection noise and can be written as

$$\Delta\omega_0 = \frac{1}{\sqrt{nTt}} \quad [24]$$

After time T, the oscillator's frequency is steered towards the resonance and new sequence starts. We will now illustrate how the use of entanglement can yield an uncertainty in the measurement beyond the shot noise limit eqn [24]. Consider the quantum network depicted in Figure 8. The n particles are now initially prepared in a maximally entangled state. This is achieved by means of subjecting the first ion to

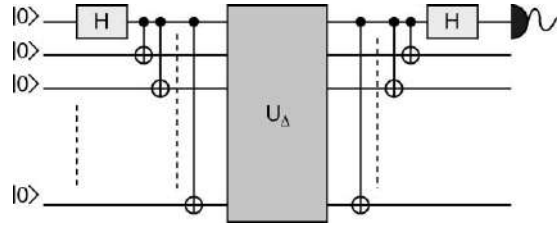

Figure 8 Quantum circuit representation for an experimental setup for a frequency estimation using maximally entangled states.

a Hadamard transformation followed by a cascade of CNOTs between the first ion, which acts as a control, and each of the remaining $(n-1)$ particles.

The resulting state is of the form

$$\frac{1}{\sqrt{2}}\Big(|0\ldots0\rangle + |1\ldots1\rangle\Big) \qquad [25]$$

The systems now evolve freely for a time t and, in light of the discussion above, we can represent this evolution in terms of a quantum gate which will leave state $|0\rangle$ invariant while introducing a phase factor proportional to the detuning if the system is in state $|1\rangle$. As a result, the joint state of the n particle after the free time evolution is given by

$$\frac{1}{\sqrt{2}}\Big(|0\ldots0\rangle + e^{-in\delta t}|1\ldots1\rangle\Big) \qquad [26]$$

with a phase factor amplified by a factor n as compared with the single particle state given by eqn [22]. The second part of the network disentangles the n particles, leaving the first ion in the state $(1/\sqrt{2})(|0\rangle + e^{-in\delta t}|1\rangle)$. If we now subject it to an additional Hadamard gate, the probability for the ion to be excited is

$$P_{1,ent} = \frac{1}{2}\Big(1 + \cos(n\delta t)\Big) \qquad [27]$$

which oscillates n times faster than the signal eqn [23]. The corresponding uncertainty for the estimation of the atomic frequency is now

$$\Delta\omega_0, ent = \frac{1}{n\sqrt{Tt}} \qquad [28]$$

This corresponds to an increase in the resolution of $\sqrt{n}$ beyond the shot noise bound. In practice, the achievable resolution using entangles states is limited by the decoherence time of the system t_{dec}, which sets the condition $T < t_{\mathrm{dec}}/2$ for the procedure to be valid. Remarkably, entanglement is still advantageous in a situation where T is chosen such that it exceeds the decoherence time. However, maximally entangled states should now be replaced by certain partially entangled ones to overcome the precision achievable with uncorrelated particles.

The application of entangled states in precision spectroscopy is a nice illustration of the potential of the ideas developed in quantum information beyond the ultimate task of building a large-scale quantum computer. The first experimental results showing the enhanced performance of entangled states have been reported by the NIST group in Boulder.

Conclusion and Perspectives

We have revised the fundamental concepts for the practical realization of quantum computation using atomic systems. We have showed how single qubit and two-qubit controlled-phase gates can be implemented using atoms interacting with the cavity field sustained by a microwave resonator and ion traps. We have illustrated how a very simple quantum computation can be used for the development of more precise frequency standards, emphasizing the potential of the ideas developed in quantum information beyond the implementation of large scale quantum computation.

Cavity QED technology has so far allowed the generation of entangled states of two and three particles and has demonstrated coherent gate operation. Ion trap technology has pushed controllable multi-particle entanglement to four particles and has showed the possibility of implementing forms of error correction. The combination of these two technologies offers interesting prospects for the near future. Surrounding traps by optical cavities and networking these devices via optical fibers is expected to allow the construction of the first prototypes for a quantum communication network.

See also

Quantum Optics: Entanglement and Quantum Information.

Further Reading

Cirac JI and Zoller P (1995) Quantum computations with cold trapped ions. *Physical Review Letters* 74: 4091.

Deutsch D, Barenco A and Ekert AK (1995) Universality in quantum computation. *Proc. Roy. Soc. London A MAT* 449: 669–677.

Grover L (1999) Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters* 79: 325–328.

Huelga SF, *et al.* (1997) Improvement of frequency standards with quantum entanglement. *Physical Review Letters* 79: 3865.

Jaynes ET and Cummings FW (1963) Comparison of quantum and semiclassical radiation theories with application to the beam maser. *Proc. IEEE* 51: 89.

Kielpinski D, Meyer V, Rowe MA, *et al.* (2001) A decoherence-free quantum memory using trapped ions. *Science* 291: 1013–1015.

Monroe C, Meekhof DM, King BE, *et al.* (1995) Demonstration of a fundamental quantum logic gate. *Physical Review Letters* 75: 4714–4717.

Raimond JM, Brune M and Haroche S (2001) *Colloquium*: manipulating quantum entanglement with atoms and

photons in a cavity. *Review Modern Physics* 73: 565 and references therein.
Sackett CA, Kielpinski D, King BE, *et al.* (2000) Experimental entanglement of four particles. *Nature* 404: 256–259.
Shor P (1994) *Proc. of the 35th Annual Symposium on the Foundations of Computer Science, IEEE Computer Society, Los Alamitos*. Santa Fe, NM: IEEE Computer Society Press.
Sorensen A and Molmer K (1999) Quantum computation with ions in thermal motion. *Physical Review Letters* 82: 1971.
Wineland DJ, Monroe C, Itano WM, Leibfried D, King BE and Meekhof DM (1998) Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl. Inst. Stand. Tech.* 103: 259 and references therein.

Squeezed Phonons in Solids

A V Bragas, Universidad de Buenos Aires, Buenos Aires, Argentina

R Merlin, University of Michigan, Ann Arbor, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The reduction of noise is crucial for measurements requiring high accuracy and, often, for the observation of new phenomena. Among noise sources, noise inherently related to the quantum nature of an object persists even after all the classical sources of noise are removed. Quantum noise is a fundamental limitation, loosely related to the Heisenberg uncertainty principle, which imposes a lower limit to the fluctuations of any pair of conjugated variables, like position and momentum. For conjugated variables A_1 and A_2 we have

$$\Delta A_1 \Delta A_2 \geq \frac{\hbar}{2} \qquad [1]$$

where

$$\Delta A_i = \sqrt{\langle A_i^2 \rangle - \langle A_i \rangle^2} \qquad i = 1, 2 \qquad [2]$$

While quantum mechanics does not forbid reducing the noise in one of the conjugated observables down to zero, the noise for the other variable will increase according to eqn [1]. Squeezing refers to the reduction of noise of a quantum variable at the expenses of its conjugated variable. The term originates from quantum optics. Nonlinear interactions of the electromagnetic field with passive and active atomic media have been successfully used to generate squeezed photons, opening new possibilities for essentially noiseless optical communications and precision measurements. The electromagnetic field is not the only quantum object that can be squeezed. By performing appropriate manipulations of the internal quantum state of a system, it is, in principle, possible to squeeze an arbitrary variable. In the past few years, vibrational states have also been experimentally squeezed in molecules and solids, by exciting the system with ultrafast optical pulses. The concept of squeezing can also be extended to an ensemble of classical objects for which the fluctuations are of thermal origin.

Squeezing applies in particular to situations where the relevant variables are the position Q and momentum P of a harmonic oscillator. Consider the Hamiltonian of a particle of mass m in a harmonic potential of frequency Ω

$$H_0 = \frac{P^2}{2m} + \frac{m}{2}\Omega^2 Q^2 \qquad [3]$$

The ground state wavefunction is

$$\varphi_0 = \frac{1}{\left[2\langle Q_0^2 \rangle \pi\right]^{1/4}} \exp\left(-\frac{Q^2}{4\langle Q_0^2 \rangle}\right) \qquad [4]$$

where $\langle Q_0^2 \rangle = \hbar/2m\Omega$. The stationary probability, $|\varphi_0|^2$, is represented in Figure 1a. This state is also a minimal uncertainty state in that $\Delta Q \Delta P \equiv \hbar/2$. $\langle Q_0^2 \rangle$ gives a measure of the quantum noise and it is known as the vacuum noise level.

Let us now discuss the behavior of a harmonic oscillator that interacts with an external source. There are two types of quantum states for which the Gaussian shape of the wavefunction is preserved after interacting with an external driving force. These are the coherent state and the squeezed state. If the interaction places the system in a state of the same form as eqn [4] but displaced with respect to the center, the wavepacket will oscillate periodically, with period $\tau = 2\pi/\Omega$, without changing its shape as shown in Figure 1b. This quantum state is known as the coherent state. It has the same uncertainty as the ground state but the expectation value $\langle Q(t) \rangle$ depends on the time following the classical trajectory in the harmonic potential. If, instead, we introduce a sudden change in the spring constant at $t = 0$, the

wavepacket will begin to 'breathe' for $t > 0$, so that its width will vary periodically with a period given by $\tau/2$. This is the squeezed state. Here the noise of one of the variables is periodically reduced below the level of the vacuum noise (Figure 1c). A state that is a combination of a coherent and a squeezed state is shown in Figure 1d. The wavepacket follows the classical trajectory and breathes at the same time.

The results in Figure 1 apply only to harmonic oscillators. While the concept of squeezing can be extended to anharmonic potentials, there are additional features to consider such as, for example, wavepacket revivals.

Dynamics of a Parametrically Excited Crystal

Harmonic oscillators describe a wide range of physical systems. In particular, the dynamics of a crystal lattice can be described in terms of a set of independent harmonic oscillators, the quantum nature of which is described in terms of phonons. In this section we discuss interactions of a harmonic oscillator with external sources that produce squeezing.

The normal modes in a crystal obey equations of the form:

$$\ddot{Q}_q + \Omega_q^2 Q_q = 0 \qquad [5]$$

where Q_q and Ω_q are the phonon coordinate and frequency, and q is the wavevector.

As briefly mentioned in the previous section, a sudden change in the spring constant produces a squeeze state. To do this, we need what is known as time-dependent parametric coupling. As in the experiments on phonon squeezing, using ultrashort optical pulses, we consider an impulsive driving force proportional to Q. The relevant equation is:

$$\ddot{Q}_q + \Omega_q^2 Q_q = \lambda_q Q_q \delta(t) \qquad [6]$$

where $\delta(t)$ is the Dirac delta function, and λ accounts for the strength of the external impulsive force. This impulsive excitation induces a sudden change in momentum, proportional to the normal coordinate Q at $t = 0$, as can be shown by integrating eqn [6]. Explicitly, $\Delta\dot{Q}_q = \Delta P_q = \lambda_q Q_q(0)$. $\Delta P_q = P_q(0^+) - P_q(0^-)$, where $0^-(0^+)$ is the time immediately before (after) the excitation. This result is illustrated in Figure 2. The main effect of the interaction is to change the width of the Gaussian packet or variance of the distribution. This leads to a time-dependence of the noise defined in eqn [2], $\Delta Q^2(t) = \langle Q^2(t)\rangle$ reflecting the 'breathing' of the wavepacket. Since P_q is the conjugated variable of Q_q, $\langle P_q^2(t)\rangle$ shows the same time-dependent behavior as $\langle Q_q^2(t)\rangle$ but in quadrature with it, so as to fulfill eqn [2] at all time. Assuming an oscillator initially in its ground state, the solution of eqn [6] is

$$\langle Q_q^2(t)\rangle = \frac{\hbar}{2m\Omega_q}\left[1 + \left(\frac{\lambda_q}{\Omega_q}\right)\sin(2\Omega_q t) + \left(\frac{\lambda_q}{\Omega_q}\right)^2 \sin^2(\Omega_q t)\right] \qquad [7]$$

It is then clear that the fluctuations of Q_q periodically fall below the vacuum quantum noise level, determined by $\langle Q_0^2\rangle$, and that the oscillations have a frequency that is twice the frequency of the oscillator.

The considerations above apply to a quantum harmonic oscillator at temperature $T = 0$. In the classical case, we need to consider an ensemble of oscillators in equilibrium with a thermal bath at temperature T. It can be shown that eqn [7] also applies to the classical case, provided $\hbar/2m\Omega$ is replaced by $k_B T/\Omega^2 m$, where k_B is the Boltzman constant. In classical squeezing, the external perturbation is applied simultaneously to the ensemble, leading to a sudden change of the momentum of all the oscillators. The jump in momentum is proportional to the position of the particle at $t = 0$. After the interaction, the noise of the ensemble will fall periodically below the thermal noise at temperature T given by $k_B T/\Omega^2 m$. A pictorial representation of classical squeezing is shown in Figure 3.

To describe a more realistic situation, it is necessary to take into account effects due to dissipation. Theoretically, this is usually accomplished by considering the coupling with a continuum of external modes simulating a thermal bath. The solution to this type of problem is nontrivial even for the simplest case of linear coupling to the bath.

The impulsive excitation of eqn [6] is a parametric excitation containing all the frequencies. A more common way to parametrically excite an oscillator is to use a monochromatic source, as described by the equation

$$\ddot{Q} + [\Omega^2 + \varepsilon\cos(\Omega_0 t)]Q = 0 \qquad [8]$$

where the parametric excitation is characterized by the amplitude ε and the modulation frequency Ω_0. Notice that the excitation induces a time-dependent frequency $\omega^2(t) = \Omega^2 + \varepsilon\cos(\Omega_0 t)$, and, hence, it

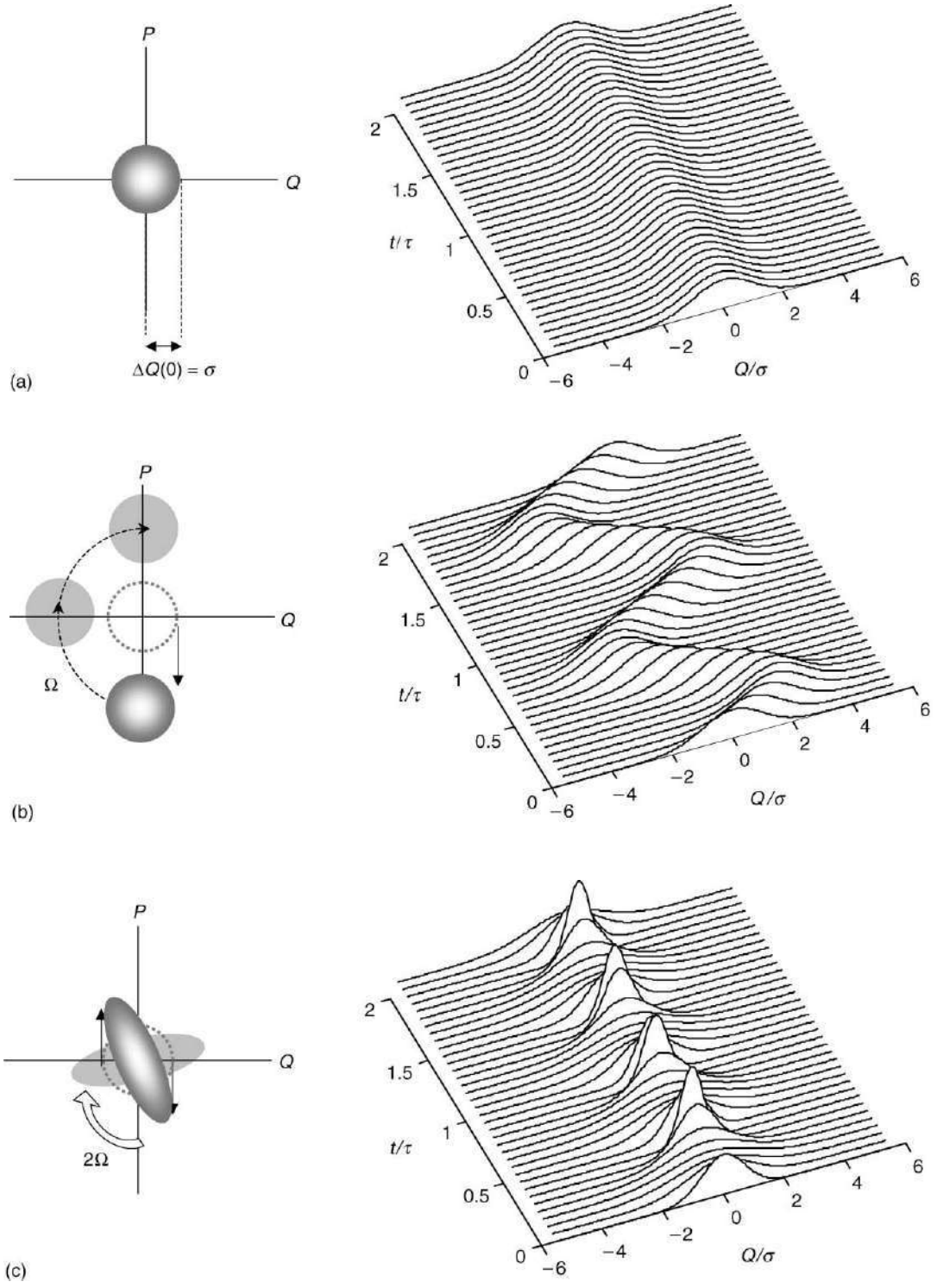

Figure 1 (a) Ground state probability density $|\varphi_0|^2$ for a quantum harmonic oscillator of frequency Ω. Q is the position, P is the momentum and σ is the standard deviation of the Gaussian wavepacket, or the square root of $\langle Q_0^2 \rangle$. The phase-space diagram on the left shows the uncertainty region for the ground state. (b) Coherent state. The uncertainty $\langle Q^2 \rangle$ does not change in time and it is always equal to $\langle Q_0^2 \rangle$, but the wavepacket oscillates with the period τ of the harmonic oscillator. The phase diagram shows the evolution of the uncertainty region, starting from the ground state (dashed line). A sudden constant change in momentum at time $t = 0$ produces a change in the equilibrium position and the wavepacket begins to oscillate. (c) Squeezed state. The wavepacket evolves by changing its width periodically with period $\pi/2$. The uncertainty, $\langle Q^2(t) \rangle$, takes values smaller than $\langle Q_0^2 \rangle$ for a fraction of the cycle. Due to the sudden change on momentum, proportional to Q, at $t = 0$, the distribution becomes ellipsoidal and rotates with frequency 2Ω. (d) Coherent-squeezed state. It is a combination of both (b) and (c). The wavepacket oscillates and breathes at the same time.

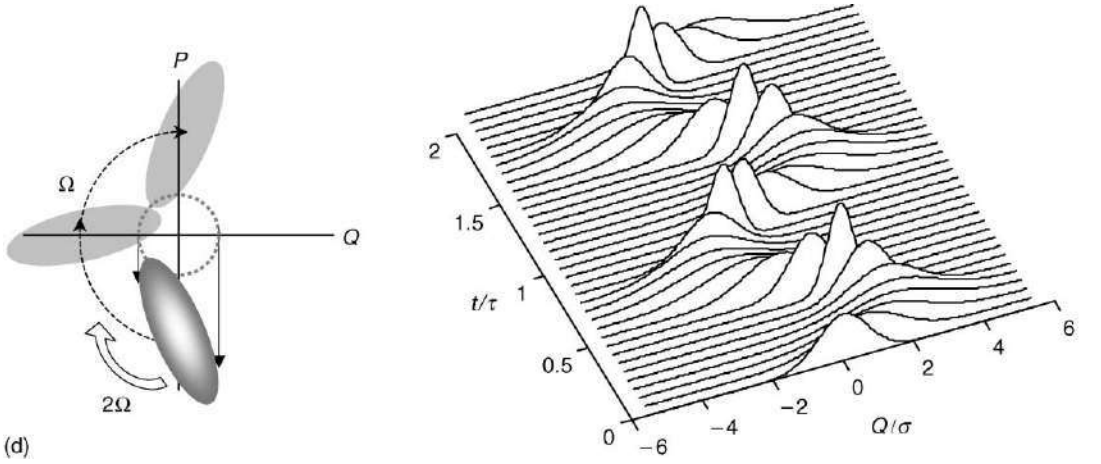

Figure 1 *Continued.*

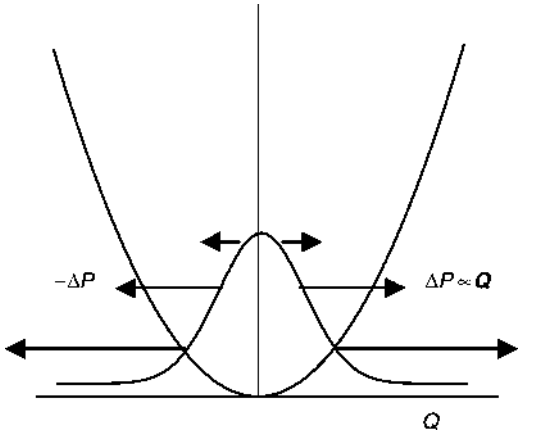

Figure 2 Schematic representation of the effect of an impulsive excitation on the ground state of the harmonic oscillator. At $t = 0$, the pulse induces a change in momentum P, proportional to the position Q. As a result, the wavepacket width in momentum space changes suddenly and starts to oscillate.

shows squeezing effects. The complete set of solutions with and without dissipation has been obtained by Zerbe and Hanggi for the quantum and classical cases.

Physical Interactions Giving Phonon Squeezing

The considerations above are rather general and specific to any physical system that behaves like a set of harmonic oscillators. Here, we consider proposals for phonon squeezing.

Equations [6] and [8] show that we need a driving force proportional to the phonon coordinate Q to achieve squeezing. One way to do this is for the external interaction to produce time-dependent correlations between two modes. Prior to the experimental achievement of phonon squeezing in 1997, using stimulated second-order Raman scattering, models based on (a) three-phonon parametric amplification and (b) polaritons had been proposed.

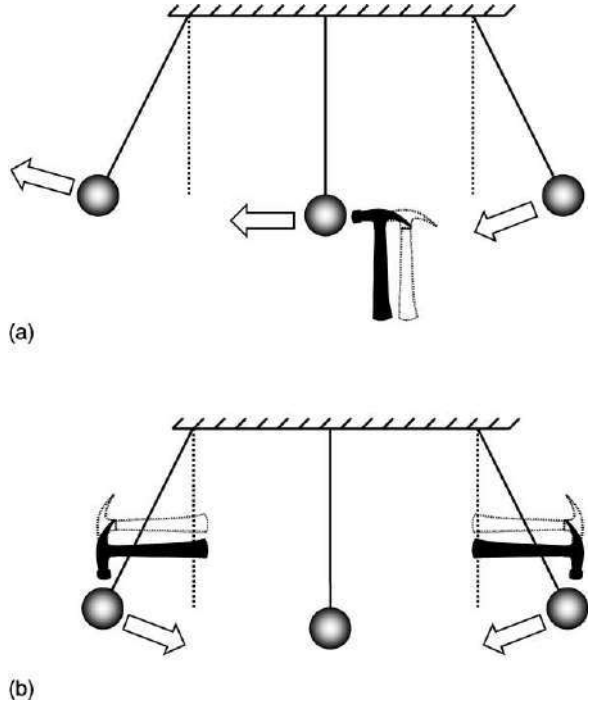

Figure 3 (a) Classical representation of a coherent phonon. An impulsive horizontal force acts on all members of a thermal distribution of pendula, displacing the group in phase space along the momentum axis. The ensemble precesses with an average position $\langle x \rangle \sim \sin(\Omega t)$. (b) Classical representation of a squeezed phonon. An impulsive vertical force produces an impulse proportional to the pendulum's displacement from equilibrium. Those pendula farthest from equilibrium receive the largest kick. The phase space distribution becomes elliptical and precesses with an ensemble averaged variance $\langle x^2 \rangle \sim \sin(2\Omega t)$.

These proposals have not yet been implemented experimentally.

The scheme proposed in (a) is based on three-phonon parametric amplification, similar to photon squeezed states that are generated by optical parametric amplification. Specifically, the proposal relies on a two-step process. First the generation by external excitation with two lasers of a 'pump' longitudinal optical phonon (LO) and, second, the decay of the optical phonon into two longitudinal acoustic phonons (LA), which will be squeezed. This is the kind of parametric oscillator described by eqn [8], for which the parametric control is driven by the coherent optical phonon field.

In case (b) it has been shown that mixed polarization modes known as phonon polaritons are intrinsically squeezed. These hybrid objects, sometimes referred to as dressed photons, are a mixture between phonons and photons that cannot be separated once they are formed.

Squeezing of vibrational states in molecules was confirmed experimentally earlier than in solids. Here, a molecule initially in the ground electronic and vibrational state is excited by an ultrafast optical pulse tuned with an electronic resonance. After the pulse arrives, the molecule is placed in a excited electronic state and the nuclei in a new (excited) potential with a new set of discrete vibrational modes. In Figure 4 we show, as an example of a diatomic molecule, the potential of a sodium dimer. For the sake of simplicity, we consider a single vibrational mode in the excited state and we disregard effects due to anharmonicity. If the optical pulsewidth is much shorter than any vibrational frequency, and the frequencies are different in the ground and excited state, the sudden change in frequency produces a vibrational squeezed state. On the other hand, even if the frequencies in the ground and excited state are equal, but the laser pulse is finite or chirped (frequency sweep on time), the final vibrational state may be squeezed. The formal demonstration of this is not trivial. Analytic solutions for the vibrational wavefunction for an arbitrary external field excitation are known for a very limited number of cases.

Squeezing of a wavepacket in an excited sodium diatomic molecule was achieved by Dunn and coworkers, measuring the time and frequency-resolved spectrum of the spontaneous emission. In this experiment, the molecule can return at any time to its electronic ground state by emitting light, the wavelength of which depends on the internuclear separation at that time. As shown in Figure 4, this is because the equilibrium position for the potential surfaces in the ground and excited states are different.

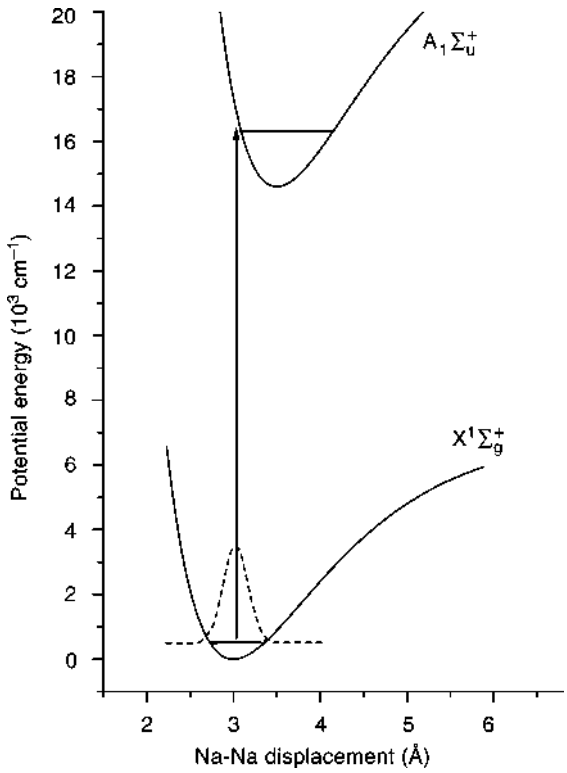

Figure 4 Potential curves for a Na_2 molecule, showing the ground and first excited states. The dashed line represents the ground state wave function. The equilibrium position of both potentials are shifted and the vibrational frequencies are different.

Thus, the time-dependent average wavelength of the emitted light oscillates as the atom vibrates, giving a signature of the dynamics of the wavepacket. Dunn observed oscillations in the signal strength that are consistent with a wavepacket expanding and contracting.

Experimental Generation and Detection of Squeezed Phonons by Second Order Raman Scattering

The first experimental demonstration of phonon squeezing was performed in 1997 by exciting a $KTaO_3$ (potassium tantalate) crystal with an ultrafast pulsed laser and coupling the light to the phonons through second order Raman Scattering. The squeezed phonon field produced by an intense optical pulse, referred to as the 'pump', is detected by a second low intensity pulse, known as the 'probe'. Phonon squeezing was also reported for $SrTiO_3$ (strontium titanate). In addition, correlations between pairs of phonons of different frequencies was reported in a closely related experiment on $KTaO_3$.

The simplest Hamiltonian describing this experiment contains the harmonic contribution from the lattice and a Raman-type interaction term. Raman scattering (RS) refers to light-crystal interaction in which the initial quantum of radiation energy, $\hbar\omega_i$, is divided into a vibrational excitation of the crystal with energy $\hbar\Omega$ and a scattered photon with energy $\hbar\omega_f$. Here, the total energy and momentum are conserved. The interaction energy is proportional to the second-order electronic susceptibility which can be expanded in terms of the atomic displacement Q. The nth term of the Taylor expansion is associated with RS of order n. Two phonons participate in the second order RS process. Although the absence of first-order RS is not required for achieving squeezing, $KTaO_3$ and SrTiO3 (above ~100 K) show no first-order Raman active modes. As a result, the leading contribution is

$$H_{\text{int}} = -\frac{1}{4}|E(t)|^2 \sum_q \chi_2(q) \qquad [9]$$

with $\chi_2(q) = (\partial^2\chi/\partial Q_q^2)Q_q^2$, where $\partial^2\chi/\partial Q_q^2$ is the second order Raman tensor, $E(t)$ is the magnitude of the light electric field, and χ is the linear electronic susceptibility. The Hamiltonian eqn [9], reflecting contributions from pairs of modes at $\pm q$, represents a change in the phonon frequency $\Delta\Omega_q(t) \approx -(1/4\Omega_q)(\partial^2\chi/\partial Q_0^2)|E(t)|^2$ that results in a time-dependent variance $\langle Q_q^2(t)\rangle$ and thus, squeezing. We note that, while individual scattering events produce a two-phonon state, the coherent ensemble of these correlated two-phonon states modulates the variance of Q. Notice that second-order RS excites phonons throughout the whole Brillouin zone instead of just the zone center at $q \approx 0$, as in first-order RS. This means that the light couples to a continuum of modes. The strength of the scattering at any particular frequency is proportional to the number of modes, i.e., the density of states, which is inversely proportional to the gradient $\nabla_q\Omega_q$ and thus, exhibits singularities at points where the gradient vanishes. These so-called critical points lead to associated singularities in the density of states referred to as van Hove singularities. $KTaO_3$ and $SrTiO_3$ show a sharp peak in the density of states at frequencies associated with certain van Hove singularities and, therefore, the continuum of modes becomes quasimonochromatic.

In the experiments, the period of the relevant phonons, $2\pi/\Omega_q$, is large compared with both the pulsewidth and the time the pulse takes for traversing the sample. Then, the time-dependent electric field can be approximated by the delta function $|E(t)|^2 \cong (4\pi I_0/nc)\delta(t)$, where I_0 is the pulse energy density, n the refractive index and c the speed of the light. Within this approximation, the dynamics of the system at $T = 0$ is described by eqn [6]. To lowest order in I_0, we have

$$\Delta Q_q = \langle Q_q^2(t)\rangle^{\frac{1}{2}} \approx \langle Q_q^2(0)\rangle^{\frac{1}{2}}\left[1 + \left(\frac{\lambda}{\Omega_q}\right)\sin(2\Omega_q t)\right] \qquad [10]$$

with $\lambda = 2\pi(I_0/nc)(\partial^2\chi/\partial Q_q^2)$.

As before, the variance drops below the quantum limit $\langle Q_q^2(0)\rangle$ for some fraction of the cycle. This zero-temperature description is only accurate for temperatures such that k_BT is very small compared with the phonon energies. For $T \neq 0$, the conditions overcoming the thermal contribution and produce quantum squeezing at low intensities are

$$n_q \ll 1 \quad \text{and} \quad \left(\frac{\lambda}{\Omega_q}\right) > 2n_q \qquad [11]$$

where $n_q = 1/[\exp(\hbar\Omega_q/k_BT) - 1]$ is the Bose factor.

Next we discuss the detection of the squeezed field. As mentioned earlier, the pump pulse creates the squeezed phonon state and the probe is affected by the changes in the material produced by the pump. The scattering of the probe by the squeezed phonon is also described by the nonlinear polarization of the material $P^{NL} = (2V)^{-1}E'(t)\sum_q\chi_2(q)$, where $E'(t)$ is the probe field, and V is the scattering volume. If the probe pulse is Gaussian-shaped of width τ_0 and centered at $\omega = \omega_0$, the total normalized change in the probe transmission ΔT is

$$\frac{\Delta T}{T}(\omega) \approx -\frac{l\pi\tau_0^2\omega(\omega-\omega_0)}{ncV}\sum_q \frac{\partial^2\chi}{\partial Q_q^2}|E'(t)|^2 \times \lambda\langle Q_q^2(0)\rangle\cos(2\Omega_q\Delta t)\exp(-2\Omega_q^2\tau_0^2) \qquad [12]$$

where l is the length of the sample. Equation [12] can be obtained from well-known results in coherent phonon detection. Comparing eqn [12] with eqn [10] it follows that the change in probe transmission is proportional to the derivative of the total time-dependent variance $\langle Q^2(t)\rangle = \sum_q\langle Q_q^2(t)\rangle$. Accordingly, the integral of $\Delta T/T$ probes $\langle Q^2(t)\rangle$.

$KTaO_3$

The perovskite $KTaO_3$ was the first material for which phonon squeezing was achieved. This compound has no first-order Raman active modes. The time-domain data in Figure 5a shows behavior

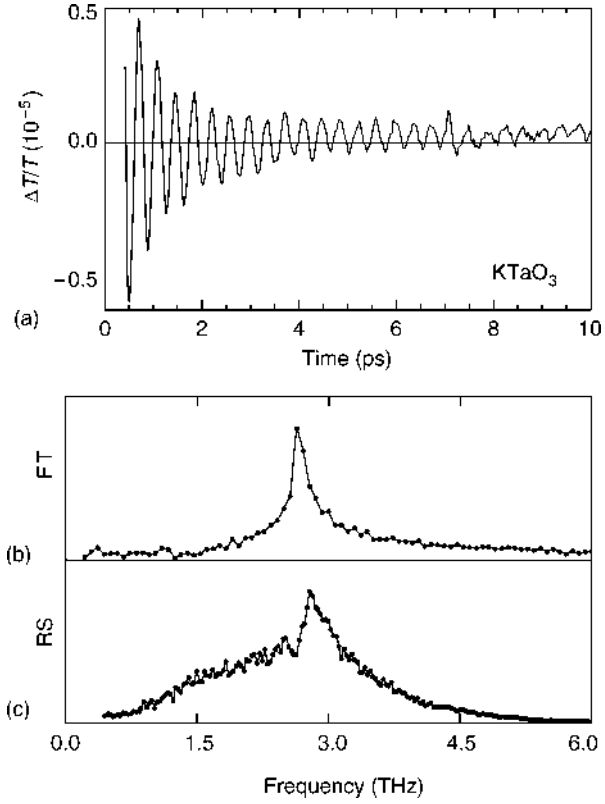

Figure 5 (a) The normalized transmitted probe intensity as a function of the probe delay for $KTaO_3$ at $T = 10$ K. Reprinted with permission from Garrett GA, Rojo AG, Sood AK, Whitaker JF and Merlin R (1997) Vacuum squeezing of solids: Macroscopic quantum states driven by light pulses. *Science*, 275(5306): 1638–1640. Copyright 1997 American Association for Advancement of Science. (b) Fourier transform of the time domain data in (a). The narrow peak appears very close to twice the frequency of the TA (transverse acoustic) mode at the X point of the Brillouin zone. (c) Second order Raman cross section.

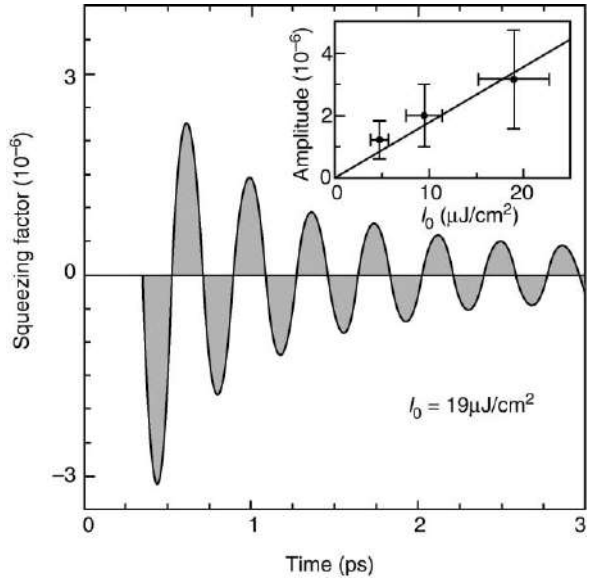

Figure 6 Experimental time dependence of the squeezing factor $S = 1-[\langle Q^2(t)\rangle/\langle Q^2(0)\rangle]^{1/2}$. I_0 is the pulse intensity. Inset: the amplitude of S as a function of the integrated intensity of the pulse, showing the linear behavior predicted by eqn [12]. Reprinted with permission from Garrett GA, Rojo AG, Sood AK, Whitaker JF and Merlin R (1997) Vacuum squeezing of solids: Macroscopic quantum states driven by light pulses. *Science* 275(5306): 1638–1640. Copyright 1997 American Association for the Advancement of Science.

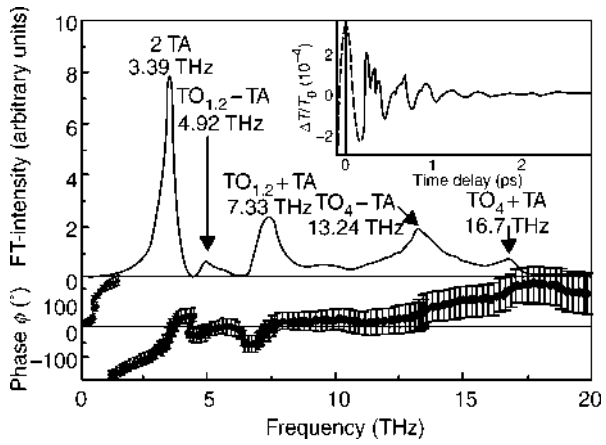

Figure 7 Fourier transform of the time domain data (inset) of the probe differential transmitivity of $KTaO_3$ at room temperature. The excited phonon combinations are indicated in the figure. Reproduced with permission from Bartels A, Dekorsy T and Kurz H (2000) Impulsive excitation of phonon-pair combination states by second-order Raman scattering. *Physical Review Letters* 84(13): 2981–2984. Copyright 2000 by the American Physical Society.

consistent with eqn [12]. The Fourier transform $\mathcal{F}_{FT}$ in Figure 5b, is dominated by a narrow peak, strongly dependent on temperature, that appears very close to twice the frequency of the TA (transverse acoustic) mode at the X point of the Brillouin zone. The comparison with the spontaneous Raman scattering data, Figure 5c, indicates that the $\mathcal{F}_{FT}$ peak is due to the 2TA overtone. The squeezing factor defined as $S = 1 - [\langle Q^2(t)\rangle/\langle Q^2(0)\rangle]^{1/2}$, is depicted in Figure 6. Notice that, whereas the squeezing factor in a quantum optics experiment can reach values as large as 70%, phonons in a solid can be squeezed by no more than, say, 0.01%. This is because larger values may lead to collapse of the solid since, according to the Lindemann criterion, solids melt when the root-mean-square displacement of an atom is a fraction of 0.2–0.25 of the lattice parameter.

In $KTaO_3$ it was also possible to induce a time-dependent covariance between two different independent modes by using second order RS to couple the two oscillators. In Figure 7 we show

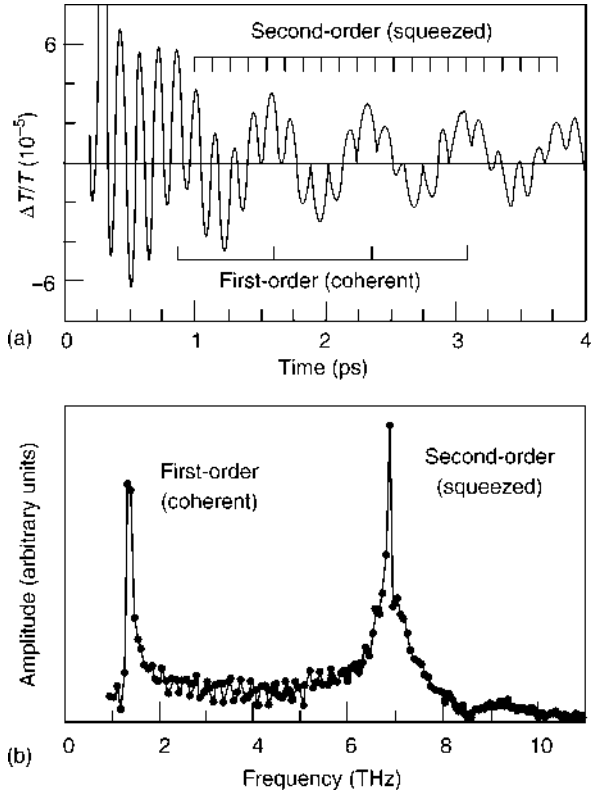

Figure 8 (a) The normalized transmitted probe intensity as a function of the probe delay for $SrTiO_3$ at $T = 7$ K. (b) Fourier transform of the time domain data. The 1.3 THz peak is the soft A_{1g}-phonon (coherent), and the 6.9 THz is the 2 TA overtone (squeezed). The second order feature is dominated by a sharp peak very close to twice the frequency of TA modes at X and M points of the Brillouin zone.

data revealing correlations between transverse optical (TO) and transverse acoustical (TA) modes.

$SrTiO_3$

This material undergoes a structural phase transition, called antiferro-distortive, at $T_C \approx 110$ K. Second-order RS is allowed at all temperatures, but first-order RS is only allowed for temperatures below T_C. Figure 8 shows data at low temperatures, revealing both first and second order features. From the comparison with spontaneous RS, the peak appearing in the $\mathcal{F}_{FT}$ at ~ 1.3 THz is due to first-order RS, while the peak at ~ 6.9 THz is due to second order RS (2TA overtone). The quantum state created by the utrafast pulse is then a mixture of a coherent and a squeezed state. Hence the phonon wavefunction will evolve qualitatively as in the picture shown in Figure 1d.

List of Units and Nomenclature

Gaussian units

I_0 energy density [$\mathrm{Jcm^{-2}}$]

E electric field [statvolt $\mathrm{cm^{-1}}$]

See also

Scattering: Stimulated Scattering; Raman Scattering. **Semiconductor Physics:** Light Scattering. **Ultrafast Laser Techniques:** Generation of Femtosecond Pulses.

Further Reading

Dunn TJ, Sweetser JN, Walmsley IA and Radzewicz C (1995) Experimental determination of the dynamics of a molecular nuclear wave-packet via the spectra of spontaneous emission. *Physical Review Letters* 74(6): 884–887.

Garraway BM and Suominen KA (1995) Wave-packet dynamics: new physics and chemistry in femto-time. *Reports on Progress in Physics* 58(4): 365–419.

Garrett GA, Rojo AG, Sood AK, Whitaker JF and Merlin R (1997) Vacuum squeezing of solids: Macroscopic quantum states driven by light pulses. *Science* 275(5306): 1638–1640.

Hayes W and Loudon R (1978) *Scattering of Light by Crystals*. New York: John Wiley & Sons, Inc.

Lindemann F (1910) Über die Berechnung molekularer Eigenfrequenzen. *Zeitschrift Für Physik* 11: 609–612.

Walls DF and Milburn GJ (1994) Generation and Applications of Squeezed Light. *Quantum Optics*, pp. 137–174. Berlin Heidelberg: Springer Verlag.

Zerbe C and Hanggi P (1995) Brownian parametric quantum oscillator with dissipation. *Physical Review E* 52(2): 1533–1543.

R

RELATIVISTIC NONLINEAR OPTICS

D P Umstadter, University of Michigan, Ann Arbor, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Immediately after the invention of the laser in 1960, light could be focused to sufficient intensity to cause nonlinear optical effects in atomic media. In conventional nonlinear optics, where the electric field of a light wave E is much smaller than an atomic field, $E_{at} = 3 \times 10^9$ V/cm, various nonlinear phenomena, such as self-focusing, harmonic generation, and Raman scattering, arise due to the anharmonic motion of electrons in the combined fields of atom and laser. Approximate analytical solutions can be obtained by means of perturbation expansion methods, using E/E_{at} as the expansion parameter. At higher light fields, when E approaches E_{at}, this method breaks down and the medium becomes photo-ionized, creating a plasma, as illustrated by Figure 1. Further increases in light intensity enabled nonlinear optical effects of even these free plasma electrons (see Figure 2). The nonlinearity arises, in this case, because the electrons oscillate at relativistic velocities in laser fields that exceed 10^{11} V/cm, resulting in relativistic mass changes exceeding the electron rest mass and the light's magnetic field becomes important. The work done by the electromagnetic field (E) on an electron ($eE\lambda$) over the distance of a laser wavelength (λ), then approaches the electron rest mass energy ($m_e c^2$), where e is the elementary charge of an electron, m_e is the electron rest mass, and c is the speed of light. Effects analogous to those studied with conventional nonlinear optics – self-focusing, self-modulation, harmonic generation, and so on – are all found, but based on this entirely different physical mechanism. Thus, a new field of nonlinear optics, that of relativistic electrons, has been launched, as illustrated by Figure 1.

One outcome of accessing this new optical regime is the generation of frequency-shifted light in a spectral region where there are no other compact sources. Another is the acceleration of other types of particles, such as positrons, ions, and neutrons. These novel radiation sources have properties (femtosecond duration, micron source size, MeV energy) that make them suitable for numerous applications in imaging and spectroscopy in basic research, as well as medical diagnostics, cancer therapy, energy production, and space propulsion. Rapid advancement is underway and new research tools, subfields, and commercial products are on the horizon, e.g., compact and ultrashort pulse duration laser-based electron accelerators and X-ray sources.

Another physical regime will be encountered at even higher intensities ($I\lambda^2 \simeq 10^{24}$ W/cm^2), when even protons will quiver relativistically, i.e., the work done on a proton over the distance of a laser wavelength approaches its rest mass energy. This might be called the nuclear regime of laser–plasma interactions, because of the fusion and fission reactions and the generation of pions, muons, and neutrinos that should occur as nuclei collide in such energetic plasmas.

The recent dramatic increase in light intensity was partly made possible by the development, in the last decade, of compact lasers that have the ability to amplify shorter light pulses. For instance, solid-state lasers use the technique of chirped-pulse amplification to generate femtosecond duration pulses. To accomplish this, an ultrashort, low-energy laser pulse is first stretched in time before it is amplified and then recompressed. Gas or dye lasers, using solid-state switches, have produced picosecond duration pulses. Present day advanced laser systems now have multiterawatt peak powers and, when focused to micron spotsizes with adaptive optics, can produce electromagnetic intensities $I\lambda^2 \simeq 10^{21}$ W/cm^2, as illustrated by Figure 3. An example of a modern ultra-high-power (0.9 PW) solid-state (Ti:Sapphire) laser

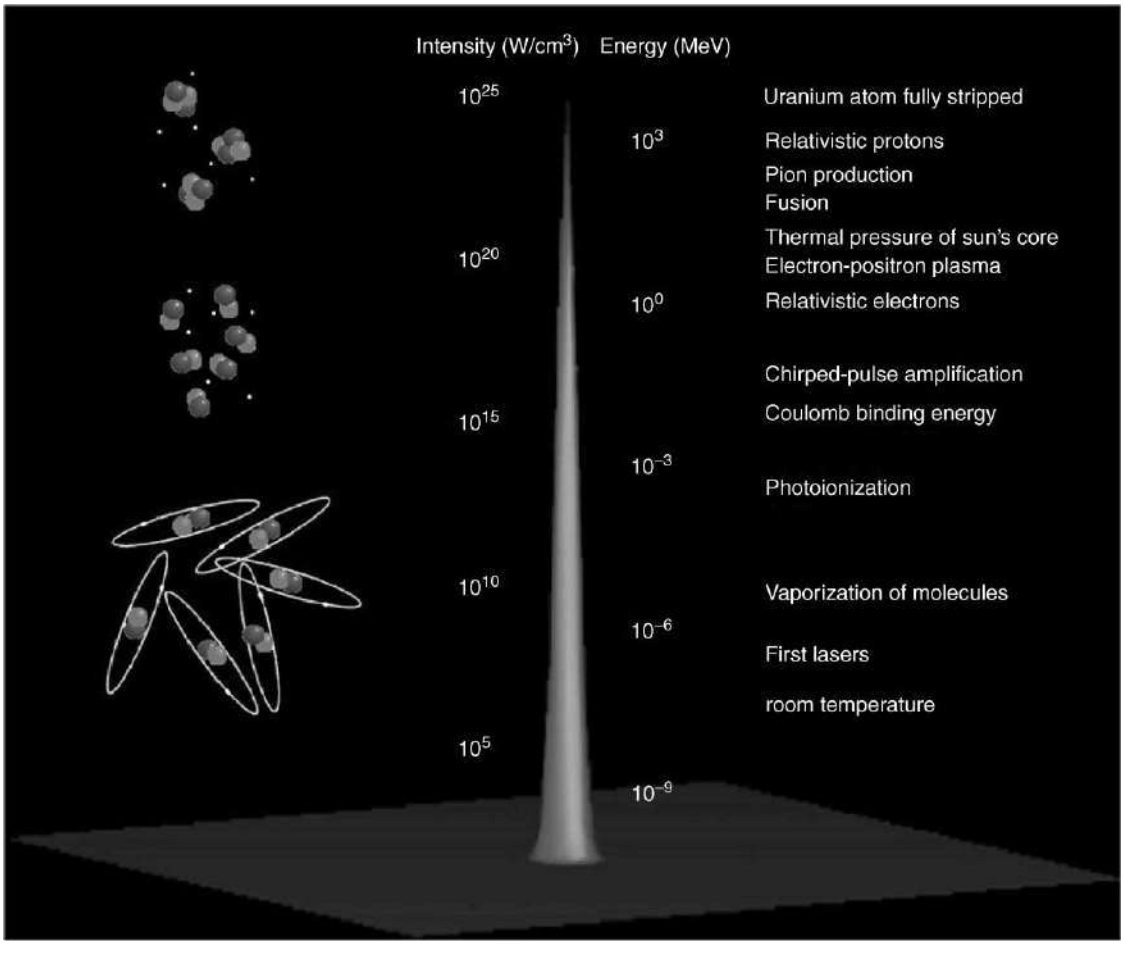

Figure 1 The various regimes of laser–matter interactions, represented by the ideal laser pulse. As the intensity of laser light increases, so does the energy of electrons accelerated in the light field and the regime of conventional nonlinear optics with electrons bound to atoms is replaced by the regime of relativistic nonlinear optics with free electrons in relativistic plasmas. At the highest intensities, even protons become relativistic, giving rise to what might be called the regime of nuclear optics, in which various nuclear processes, such as fusion, can take place. (Reproduced from Umstadter D (2003) Relativistic laser–plasma interactions. *Journal Physics D: Applied Physics* 36: R151–R165, with permission from IOP Publishing Ltd.)

system, located at the Advanced Photon Research Center, Kansai Research Establishment, Japan Atomic Energy Research Institute (JAERI), can be seen in the photograph shown in Figure 4.

This paper will discuss the fundamental concepts underlying what might be referred to more broadly as the field of high-field science. For more detailed descriptions of recent progress in experiment or theory, several review papers have been published on related topics: (i) relativistic nonlinear optics; (ii) high-intensity laser development; (iii) laser accelerators; (iv) intense laser–plasma interactions; and (v) relativistic scattering.

Definitions

For an infinite plane wave for the light wave propagating in the $\hat{z}$ direction (shown in Figure 5), the fields may be described by

$$\mathbf{E} = \hat{e}_x E_0 \exp[\mathrm{i}(kz - \omega t)] = \hat{e}_x E_0 \, \mathrm{e}^{\mathrm{i}\psi} \qquad [1]$$

and

$$\mathbf{B} = \hat{e}_y B_0 \exp[\mathrm{i}(kz - \omega t)] = \hat{e}_y B_0 \, \mathrm{e}^{\mathrm{i}\psi} \qquad [2]$$

where ω and k are the frequency and wavenumber of the laser light, respectively. For the ultrashort pulse durations of the lasers typically used to produce high fields, it is assumed that the ions do not have time to move because of their greater inertia. In this case, $m(\mathrm{d}\mathbf{v}/\mathrm{d}t) \approx -e\mathbf{E}$ and the term v/c can be used as an expansion parameter. At even higher fields, electrons can be accelerated to nearly the speed of light as they quiver in the oscillating fields of the light wave. In this case, the force due to the magnetic field can no longer be ignored and the mass of the electron

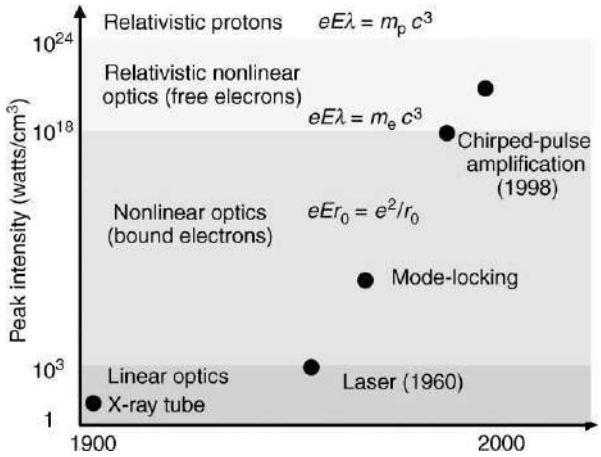

Figure 2 History of light sources over the last century. Each advance in laser power enables a new regime of optics. (Reproduced from Umstadter D (2003) Relativistic laser–plasma interactions. *Journal Physics D: Applied Physics* 36: R151–R165, with permission from IOP Publishing Ltd.)

can increase as $m = \gamma m_0$, where m_0 is the electron rest mass and the relativistic gamma factor is given by

$$\gamma = \frac{1}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \qquad [3]$$

The threshold for the relativistic regime of optics is crossed when the electron momentum p exceeds the product of the electron rest mass m_0 and c.

For problems involving charged particle motion in intense electromagnetic fields, it is often convenient to introduce the vector potential: $\mathbf{A}(\mathbf{r}, t) = \mathrm{Re}\{\mathbf{A}_0\, \mathrm{e}^{\mathrm{i}\psi}\}$. From Maxwell's equation $\nabla \cdot B = 0$, it can be seen that without loss of generality, the magnetic field can be written as the curl of a potential $B = \nabla \times A$. Further, from Faraday's law in vacuum:

$$\nabla \times E = -\frac{1}{c}\frac{\partial B}{\partial t} = -\frac{1}{c}\frac{\partial(\nabla \times A)}{\partial t} \qquad [4]$$

it is clear that $\mathbf{E} = -(1/c)\partial\mathbf{A}/\partial t$. To include the space–charge field of the plasma, as we will do in the section on collective plasma response below, we introduce the potential Φ, and the definition:

$$\mathbf{E} = -\nabla\Phi - \frac{1}{c}\frac{\partial}{\partial t}\mathbf{A} \qquad [5]$$

Assuming the Coulomb gauge:

$$\nabla \cdot \mathbf{A} = 0 \qquad [6]$$

or $A_z = 0$ in 1D, then $\mathbf{A}_\perp$ represents the laser field.

If we neglect Φ for now, then, by Fourier analysis:

$$\mathbf{E} = \mathrm{Re}\left\{\frac{\mathrm{i}\omega}{c}\mathbf{A}_0\, \mathrm{e}^{\mathrm{i}\psi}\right\} \qquad [7]$$

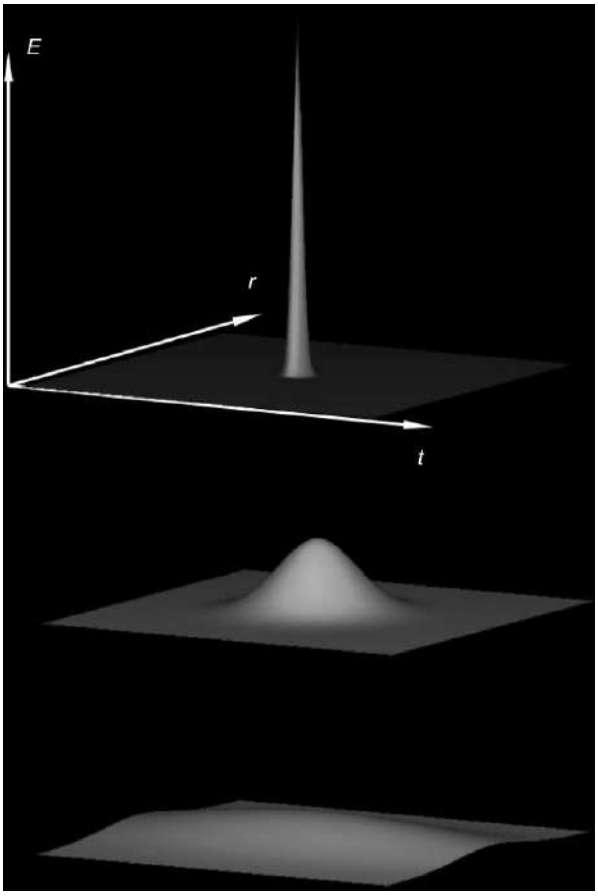

Figure 3 Laser light has become concentrated to ever-smaller regions of space (r) and time (t), dramatically increasing the peak electric field (E) at the laser focus. Prior to the development of chirped-pulse amplification (CPA), the energy of light was produced in long-duration pulses, as shown in the pulse of the bottom figure. After CPA, the pulse duration decreased dramatically as shown in the pulse in the middle. The latest improvement in laser technology has been the use of deformable mirrors, which has allowed lasers to be focused to a spatial dimension that is as small as the temporal dimension, a few laser wavelengths, as shown in the pulse on top. (Reproduced from Umstadter D (2003) Relativistic laser–plasma interactions. *Journal Physics D: Applied Physics* 36: R151–R165, with permission from IOP Publishing Ltd.)

$$\mathbf{B} = \mathrm{Re}\{\mathrm{i}k\hat{z} \times \mathbf{A}_0\, \mathrm{e}^{\mathrm{i}\psi}\} \qquad [8]$$

The light intensity is given by the time-averaged Poynting's vector, which can be written for linear (LP), $\mathbf{A}_0 = A_0\hat{e}_y$ and circular polarizations (CP), $\mathbf{A}_0 = A_0(\hat{e}_x \pm \hat{e}_y)$ as:

$$I = \langle|\mathbf{S}|\rangle = \left\langle\left|\frac{c\mathbf{E} \times \mathbf{B}}{8\pi}\right|\right\rangle = \frac{\omega k}{16\pi}A_0^2 \qquad [9]$$

It is often the combination $I\lambda^3$ that matters, which has units of power and for linear polarization can be

Figure 4 Photograph of the 33-fs duration 0.9-PW-peak power laser system at JAERI. (Reproduced with permission from Yamakawa K (2004) Table-top lasers create ultrahigh peak powers. *OYOBUTURI* 73: 186–193.)

written as

$$I\lambda^2 = \frac{\omega k \lambda^2}{16\pi} A_0^2 = \frac{\pi}{2} c A_0^2 = \left[1.37 \times 10^{18} \frac{\mathrm{W}}{\mathrm{cm}^2}\,\mu\mathrm{m}^2\right] a_0^2 \qquad [10]$$

where the normalized vector potential is defined as

$$a_0 = \frac{eA_0}{mc^2} \qquad [11]$$

Relativistic effects become significant when a_0 approaches unity, which, as can be seen from eqn [10] occurs for 1 μm wavelength light at a laser intensity of $\sim 1.37 \times 10^{18}$ W/cm^2, corresponding to a electric field strength of 6×10^9 V/cm, which is four orders of magnitude greater than E_{at}.

Single-Particle Motion

The electron orbit, subject to a linearly polarized electromagnetic wave propagating in the $+z$ direction, is governed by the Lorentz equation:

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -e\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) \qquad [12]$$

where v is the electron velocity, and E and B are the light's electric and magnetic field.

A zeroth order solution to eqn [12] is found by setting $v/c \ll 1$, which allows the term $(v/c) \times B$ to be neglected and γ from eqn [8] is set to unity. Integrating eqn [12] in this limit once over time (t) yields for the velocity:

$$\mathbf{v} = \mathrm{Re}\left\{\frac{e\mathbf{E}}{\mathrm{i}m\omega}\right\} = \frac{a_0}{c}\begin{cases} \hat{e}_y \cos\psi, & \mathrm{LP} \\ (\hat{e}_x \cos\varphi \mp \hat{e}_y \sin\psi), & \mathrm{CP} \end{cases} \qquad [13]$$

An electron in low-intensity light oscillates with this velocity in a straight line along the polarization vector ($\hat{e}_y$), which when normalized to c, is the normalized vector potential a_0. Integrating again yields for the

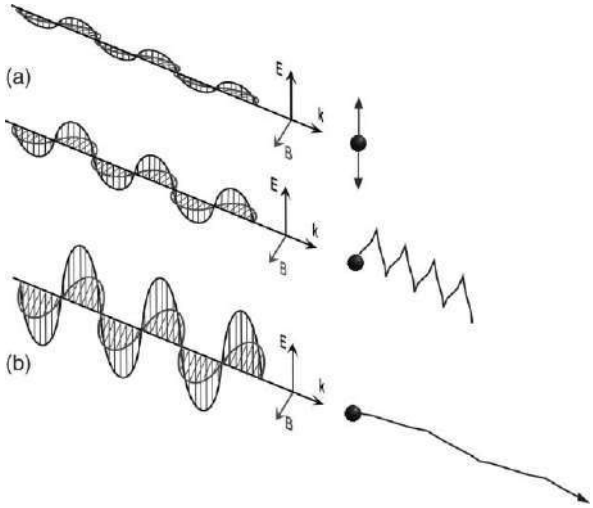

Figure 5 Classical optics versus relativistic optics. (a) In classical optics, the amplitude of the light wave is small, electrons oscillate in the direction of the electric field at the light's frequency, and there is no displacement along the light's propagation direction. Note that only the *E* field acts on the electron and the electron-oscillation velocity is very small compared with the speed of light. (b) In relativistic optics, the amplitude of the light wave is very large, the light's magnetic field becomes important and the combined action of the electric and magnetic fields push the electron forward. In this case, the electron velocity becomes close to the speed of light and its motion is a figure-eight superimposed upon a steady drift. (Reproduced from Umstadter D (2003) Relativistic laser–plasma interactions. *Journal Physics D: Applied Physics* 36: R151–R165, with permission from IOP Publishing Ltd.)

transverse displacement:

$$\mathbf{x} = \mathrm{Re}\left\{\frac{e\mathbf{E}}{m\omega^2}\right\} = \frac{a_0}{\omega c}\begin{cases}\hat{e}_y \sin\psi, & \mathrm{LP}\\ (\hat{e}_y \sin\varphi \mp \hat{e}_z \cos\psi), & \mathrm{CP}\end{cases} \qquad [14]$$

Thus, for the $a_0 \sim 1$, the electron excursion during its oscillation is approximately λ.

A first-order approximation for the electron motion can be found by substituting the zeroth order velocity eqn [13] into the $v/c \times B$ term of eqn [12]. The latter then becomes proportional to

$$E \times B \propto \frac{a_0^2}{2}[1 + \cos(2\omega_0)]\hat{e}_z \qquad [15]$$

In the frame in which the electron is on the average at rest, the relativistic motion of an electron is thus described by a figure eight, oscillating twice in the $\hat{e}_z$ or $\hat{k}$ direction for every once in the polarization direction ($\hat{e}_x$); i.e., a figure of eight motion. This originates from the fact that $v \times B \propto E \times B \propto E^2\hat{k}$, which is a product of two functions that vary sinusoidally at frequency ω and thus varies itself at frequency 2ω. In the lab frame, this transverse motion is superimposed upon a steady drift in the ($\hat{e}_z$) direction (shown in Figure 5), originating from the DC term in eqn [15]. The next order approximation would include the mass shift $m = \gamma m_0$. As the field strength increases ($a_0^2 \gg 1$), the longitudinal motion ($\propto a_0^2$) begins to dominate the transverse motion ($\propto a_0$), as shown in Figure 5. This is described more formally below.

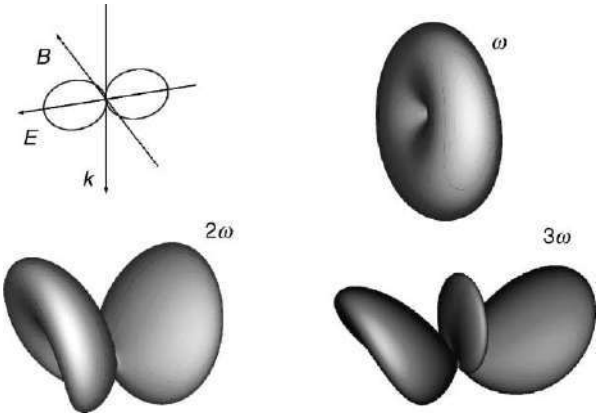

Figure 6 Harmonics driven by relativistic Thomson scattering as the electrons in high-intensity laser fields (a_0^2) undergo figure-eight motion display unique angular distributions. (Reproduced from Umstadter D (2003) Relativistic laser–plasma interactions. *Journal Physics D: Applied Physics* 36: R151–R165, with permission from IOP Publishing Ltd.)

In the regime $a_0 \lesssim 1$, electrons radiate photons at harmonics of a modified laser frequency ω_0, with each harmonic order having its own unique angular distribution, as shown in Figure 6. The radiation at the fundamental is the usual donut pattern, with a maximum in the direction perpendicular, and a minimum along, the polarization vector ($\hat{e}_x$). The second harmonic has two emission lobes with maxima pointing at an angle between $\hat{e}_z$ and $\hat{e}_x$. An additional lobe is added for each additional harmonic order. This is referred to as nonlinear or relativistic Thomson scattering. The unique angular distributions of the second and third harmonics emitted from nonlinear relativistic Thomson scattering were observed experimentally and are shown in Figure 7. As will be shown in detail in the section on radiation from relativistic electrons, below, the situation is more complex for $a_0 \gtrsim 1$, and the scattered light is no longer simply harmonic.

Constants of the Motion

Several constants of the motion can be found from an exact treatment for the motion. The starting point is the relativistically correct Lagrangian, which is written as

$$L(\mathbf{r}, \mathbf{v}, t) = -mc^2\sqrt{1 - \frac{v^2}{c^2}} + \frac{q}{c}\mathbf{v}\cdot\mathbf{A} + q\phi \qquad [16]$$

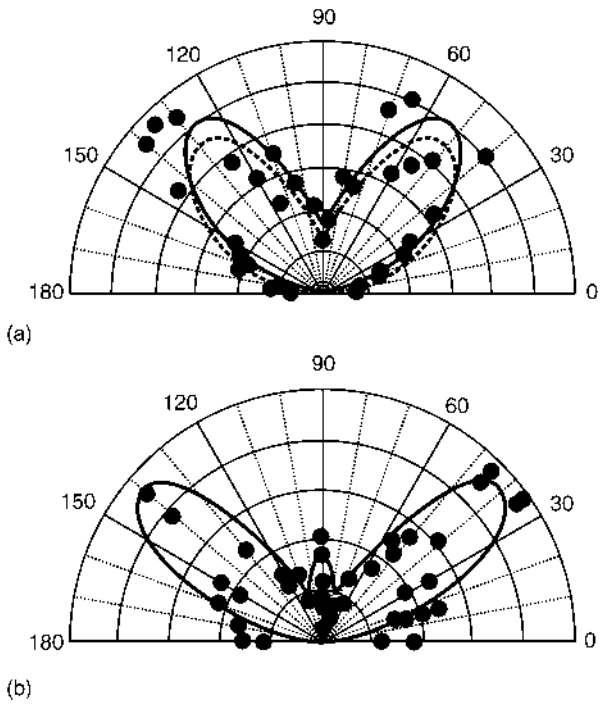

Figure 7 Angular pattern of higher-order harmonic light. Shown are polar plots of the intensity of the second-harmonic light (top) and third harmonic (bottom) as a function of azimuthal angle. Filled circles, experimental data; solid and dashed lines, theoretical results. (Reproduced with permission from Chen S-Y, Maksimchuk A and Umstadter D (1998) Experimental observation of relativistic nonlinear Thomson scattering. *Nature* 396: 653.)

where ϕ is the scalar potential associated with electrostatic fields. The Lorentz equation, eqn [12], is derived from the Euler–Lagrange equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \mathbf{v}} - \frac{\partial L}{\partial \mathbf{r}} = 0 \qquad [17]$$

For an infinite plane wave, in which the Lagrangian is independent of space in the transverse direction, eqn [17] yields the conservation of transverse canonical momentum:

$$\frac{\partial L}{\partial \mathbf{v}_\perp} = \mathbf{p}_\perp + \frac{q}{c}\mathbf{A}_\perp = \text{constant} \qquad [18]$$

The sum of the transverse momentum and field strength remain constant. The next constant follows from $\mathrm{d}H/\mathrm{d}t = -\partial L/\partial t$, yielding the relation:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\frac{\partial L}{\partial t} = c\frac{\partial L}{\partial z} = c\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial v_z} = c\frac{\mathrm{d}p_z}{\mathrm{d}t} \qquad [19]$$

where E is the time-dependent energy of the particle, which yields our second constant of the motion:

$$E - cp_z = \text{constant} \qquad [20]$$

For a particle initially at rest, the kinetic energy E_{kin} becomes

$$E_{\text{kin}} = E - mc^2 = p_z c \qquad [21]$$

which yields

$$E_{\text{kin}} = \frac{p_\perp^2}{2m} = p_x c = mc^2(\gamma - 1) \qquad [22]$$

Thus, the electron scattering angle θ is related to the transverse and longitudinal momenta or the kinetic energy by

$$\tan^2\theta = \left(\frac{p_\perp}{p_z}\right)^2 = \frac{2mE_{\text{kin}}}{(E_{\text{kin}}/c)^2} = \frac{2}{\gamma - 1} \qquad [23]$$

The angle of electrons produced by photo-ionization with intense lasers has been shown experimentally to obey the conservation of canonical momentum, as demonstrated in experiments that studied the angular distribution of relativistic electrons emitted from barrier-suppression ionization of atoms in intense laser fields (shown in Figure 8). The transverse normalized vector potential, $\mathbf{a} = e\mathbf{A}_\perp/mc^2$, is thus equal to the transverse normalized momentum:

$$\hat{\mathbf{p}}_\perp = \frac{\mathbf{p}_\perp}{mc} = \mathbf{a} = (0, a_y, a_x) \qquad [24]$$

The normalized kinetic energy can also be written in terms of a:

$$\hat{E}_{\text{kin}} = \frac{E_{\text{kin}}}{mc^2} = \gamma - 1 = \hat{p}_z = \frac{\hat{p}_\perp^2}{2} = \frac{a^2}{2} \qquad [25]$$

This can be rewritten as

$$\gamma = 1 + \frac{a^2}{2} \qquad [26]$$

From eqn [25], the normalized momenta in each direction can be found:

$$\hat{p}_z = \gamma\beta_z = \frac{\gamma}{c}\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{a^2}{2} \qquad [27]$$

$$\hat{p}_y = \gamma\beta_y = \frac{\gamma}{c}\frac{\mathrm{d}y}{\mathrm{d}t} = a_y \qquad [28]$$

$$\hat{p}_x = \gamma\beta_x = \frac{\gamma}{c}\frac{\mathrm{d}x}{\mathrm{d}t} = a_x \qquad [29]$$

Note that the momentum in the longitudinal ($\hat{e}_z$) direction is proportional to a^2 while in the two

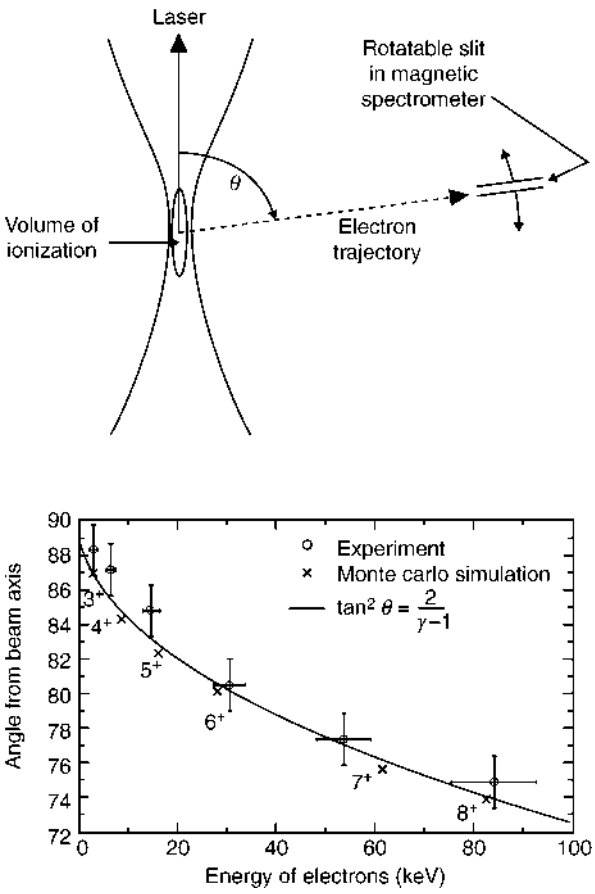

Figure 8 Top: Experiment to measure the angle of electron ejection during photo-ionization. Bottom: Kinetic energy versus angle during the production of 3^+ through 8^+ of neon. (Reproduced with permission from Moore CI, Knauer JP and Meyerhofer DD (1995) Observation of the transition from Thomson to Compton scattering in multiphoton interactions with low-energy electrons. *Physics Review Letters* 74: 2439. Copyright (1995) by the American Physical Society.)

transverse directions it is proportional to a. Equations [26]–[29] indicate that in the limit of large a, the normalized velocities are given by

$$\beta_z = \frac{a^2/2}{1+a^2/2} \to 1, \qquad \beta_y = \frac{a_y}{1+a^2/2} \to 0, \qquad \beta_x = \frac{a_x}{1+a^2/2} \to 0 \qquad [30]$$

This indicates that in the deeply relativistic regime, the electron is accelerated to near c in the longitudinal direction, while in the transverse directions its velocity goes to zero. It follows that, as a increases and the electron moves in phase with the light wave, relativistic Thomson scattering will be significantly reduced. Integrating eqn [29] in the variable $\tau = t - x(t)/c$ gives

$$\gamma\frac{\mathrm{d}}{\mathrm{d}t} = \gamma\frac{\mathrm{d}\tau}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}\tau} = \gamma\left(1 - \frac{1}{c}\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)\frac{\mathrm{d}}{\mathrm{d}\tau} = \left(1 + \frac{a^2}{2} - \frac{a^2}{2}\right)\frac{\mathrm{d}}{\mathrm{d}\tau} = \frac{\mathrm{d}}{\mathrm{d}\tau} \qquad [31]$$

which allows the equations for the displacement to be written simply as

$$\frac{\mathrm{d}z}{\mathrm{d}\tau} = c\frac{a^2}{2} \qquad [32]$$

$$\frac{\mathrm{d}y}{\mathrm{d}\tau} = ca_y \qquad [33]$$

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = ca_x \qquad [34]$$

For a box-shaped pulse, $a_x = a_0\cos(\omega\tau)$ for $0 < \tau < N(2\pi/\omega)$, $a_z = 0$, $a^2 = a_x^2\cos^2(\omega\tau)$, the trajectory for an electron initially at rest at $x = y = z = 0$ at $\tau = 0$ is then

$$x(\tau) = \frac{ca_0^2}{2}\int_0^\tau \cos^2(\omega\hat{\tau})\mathrm{d}\hat{\tau} = \frac{ca_0^2}{4}\left[\tau + \frac{1}{2\omega}\sin(2\omega\tau)\right] \qquad [35]$$

$$y(\tau) = ca_0\int_0^\tau \cos(\omega\hat{\tau})\mathrm{d}\hat{\tau} = \frac{ca_0}{4}\sin(\omega\tau) \qquad [36]$$

Thus, as we saw from the first-order appoximation above, eqn [15], the motion consists of a drift along $\hat{x}$:

$$x_\mathrm{d}(t) = \frac{a_0^2}{a_0^2+4}ct \qquad [37]$$

and a figure-eight motion in the drift frame of reference:

$$ky = a_0\sin(\omega\tau), \qquad k(z - z_\mathrm{d}) = \frac{a_0}{8}\sin(2\omega\tau) \qquad [38]$$

In the strongly relativistic regime, because the electron is strongly accelerated in the direction of the light wave, the harmonic emission from relativistic Thomson scattering loses its dependence on the laser frequency, ω, but instead depends only on the direction of observation, the amplitude a_0 and the initial phase of the electron, relative to the light wave. Electrons that do not begin their trajectories at zero or π phase with respect to the light wave can

drift in the transverse direction. For pulses with finite spatial width, such as a Gaussian pulse with field strength peaked on axis, the ponderomotive force of the light can also act to expel the electrons transversely from the axis.

For circular polarization:

$$\mathbf{a}(\mathbf{r}, t) = \mathrm{Re}\{a_0(\hat{e}_y \pm \mathrm{i}\hat{e}_x)\mathrm{e}^{-\mathrm{i}\omega\tau}\} \quad [39]$$

the electron drifts as it did in the case of linear polarization and described in eqn [37] but orbits in a helical trajectory:

$$y(\tau) = \frac{ca_0}{\sqrt{2}\omega}\sin(\omega t/\gamma) \quad [40]$$

$$x(\tau) = \mp\frac{ca_0}{\sqrt{2}\omega}\cos(\omega t/\gamma) \quad [41]$$

Rather than oscillate in amplitude, the relativistic factor remains constant:

$$\gamma = 1 + \frac{a_0^2}{4} \quad [42]$$

In either polarization case, the light pulse eventually overtakes the electron, leaving the latter at rest. However, net acceleration in vacuum can be obtained only under special circumstances, described by the Lawson–Woodward criterion, such as in the near field of metallic surfaces; with high-order focusing modes instead of infinite plane waves; or with mixed frequencies, or with optics to terminate the light but not the electrons. Regions of a tightly focused Gaussian pulse, where the accelerating field propagates at sub-luminal velocities also exist, but they are effective for only extremely large values of a and thus only very short acceleration lengths.

Role of Initial Phase

If we wish to understand the role of the electron's initial phase, the relativistic Lorentz equation (eqn [12]) may also be written, for an $\hat{x}$-polarized wave traveling in the $\hat{z}$ direction, as

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma\beta) = (\hat{x} - \hat{z}\times\beta)a\cos(t - z) \quad [43]$$

where we normalize time by $1/\omega_0$, velocity by c, and distance by c/ω_0. In eqn [43], $a = eE_0/m\omega_0 c$ is the dimensionless parameter measuring the electric field strength, $\gamma = (1 - \beta_x^2 - \beta_y^2 - \beta_z^2)^{-1/2}$ is the relativistic mass factor, and $\beta = (\beta_x, \beta_y, \beta_z)$ is the electron velocity (in units of c). The electron orbit, subject to the following general initial conditions at time $t = 0$:

$$x = 0, \qquad y = 0, \qquad z = z_{\mathrm{in}} \quad [44]$$

$$\beta_x = \beta_{x0}, \qquad \beta_y = \beta_{y0}, \qquad \beta_x = \beta_{z0} \quad [45]$$

$$\beta_{x0} = \beta_{y0} = 0 \quad [46]$$

has a closed form solution when it is expressed parametrically: $t = t(\theta)$, $r = r(\theta)$, $\beta = \beta(\theta)$, where θ is the phase of the wave, defined by

$$\theta = t - z \quad [47]$$

Note that $\beta_0 = (\beta_{x0}, \beta_{y0}, \beta_{z0})$ is the unperturbed velocity of the electron ($a = 0$ limit) and that the initial phase that the electron sees is $\theta_{\mathrm{in}} = -z_{\mathrm{in}}$ according to eqns [44] and [47]. This phase can be important in the ionization of the gas by an intense laser.

For the special case $\beta_{x0} = 0$, $\beta_{y0} = 0$, one finds $\beta_y = 0$, and the orbital equation yields the following closed form solution:

$$\gamma = \gamma_0 + \frac{a^2(\sin\theta - \sin\theta_{\mathrm{in}})^2}{2\gamma_0(1 - \beta_{z0})} \quad [48]$$

$$\gamma\beta_z = \gamma - \gamma_0(1 - \beta_{z0}) \quad [49]$$

$$\gamma\beta_x = \frac{1}{\gamma_0(1 - \beta_{z0})}a(\sin\theta - \sin\theta_{\mathrm{in}}) \quad [50]$$

$$x = \frac{a[(\cos\theta_{\mathrm{in}} - \cos\theta) - (\theta - \theta_{\mathrm{in}})\sin\theta_{\mathrm{in}}]}{\gamma_0(1 - \beta_{z0})} \quad [51]$$

$$t = \frac{(\theta - \theta_{\mathrm{in}})}{1 - \beta_{z0}}\left[1 + \frac{a^2}{2}\left(\frac{1}{2} + \sin^2\theta_{\mathrm{in}}\right)\right] + \frac{a^2(1 + \beta_{z0})}{2(1 - \beta_{z0})}\left[-\frac{\sin 2\theta}{4} + 2\cos\theta\sin\theta_{\mathrm{in}} - \frac{3\sin 2\theta_{\mathrm{in}}}{4}\right] \quad [52]$$

In eqns [48]–[52], $\gamma_0 = (1 - \beta_{z0}^2)^{-1/2}$ and β_{z0} may either be negative (counter-propagating against the laser), or zero (Thomson scattering), or positive (co-propagating with the laser). Note that the velocity components β_x and β_z are given as explicit functions of θ according to eqns [49] and [50], upon using eqn [48]. They are periodic functions of θ of period 2π. The period, T, of this periodic (**Figure 8**) motion is thus equal to the increase in t as θ increases by 2π. Thus, we obtain from eqn [53]:

$$T = \frac{2\pi}{1 - \beta_{z0}}\left[1 + \frac{a^2}{2}\left(\frac{1}{2} + \sin^2\theta_{\mathrm{in}}\right)\right] \quad [53]$$

The parametric solution for the z-coordinate of the electron orbit is given by $z = t - \theta$, in which t is given by eqn [52]. Over one orbital period, T, the electron undergoes a net displacement $r_0 = (x_0, 0, z_0)$ where x_0 is given by the increase in eqn [51] as θ increases by 2π, and z_0 is simply $T - 2\pi$:

$$z_0 = T - 2\pi = \frac{2\pi}{1-\beta_{z0}}\left[\beta_{z0} + \frac{a^3}{2}\left(\frac{1}{2} + \sin^2\theta_{\text{in}}\right)\right] \quad [54]$$

$$x_0 = \frac{-2\pi a \sin\theta_{\text{in}}}{\gamma_0(1-\beta_{z0})} \quad [55]$$

Note that the electron trajectory depends on a, β_{z0}, and θ_{in} in a rather complicated manner. Accordingly, the fundamental frequency ω_1 of the radiation spectrum depends on these three quantities. For backscattered radiation ($\hat{n} = -\hat{z}$), the dependence on electron beam and on the laser becomes decoupled.

Following is a useful formula relating the change of t with respect to θ:

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = 1 - \beta_z = \frac{\gamma_0(1-\beta_{z0})}{\gamma} \quad [56]$$

which may be verified from eqns [47] and [49]. Equation [56] is also valid for arbitrary values of β_{x0}, β_{y0}, β_{z0}, and θ_{in}, in which case γ_0 is the electron's relativistic mass factor in the absence of the laser.

Radiation from Relativistic Electrons

An electron with displacement $r(t)$ and velocity $v(t)$ carries a current density $J(r,t) = ev(t)\delta r - r(t)]$, whose Fourier transform $J_{\text{F}}(k,\omega)$ may easily be obtained. The total work done, W, in ergs, performed by the current $\mathbf{J}(\mathbf{r},t)$ on the electric field $\mathbf{E}(\mathbf{r},t)$ is given by

$$W \equiv \int \mathrm{d}\omega\,\mathrm{d}\Omega \frac{\mathrm{d}^2 W}{\mathrm{d}\omega\,\mathrm{d}\Omega} = \int \mathrm{d}\omega\,\mathrm{d}\Omega \frac{\omega^2}{c^2}\left[|\mathbf{J}_{\text{F}}(\mathbf{k},\omega)|^2 - |\mathbf{n}\cdot\mathbf{J}_{\text{F}}(\mathbf{k},\omega)|^2\right] \quad [57]$$

where the $\mathbf{k}$-space differential volume $\mathrm{d}^3\mathbf{k} = k^2\,\mathrm{d}k\,\mathrm{d}\Omega$ is expressed in terms of the solid angle (Ω) in the direction of the unit vector $\mathbf{n} = c\mathbf{k}/\omega$. We immediately obtain from eqn [56], under the far-field approximation:

$$\frac{\mathrm{d}^2 W}{\mathrm{d}\Omega\,\mathrm{d}\omega} = \frac{e^2\omega^2}{4\pi^2 c}|\mathbf{n}\times[\mathbf{n}\times\mathbf{F}(\omega)]|^2 \quad [58]$$

$$\mathbf{F}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t\,\beta(t)\mathrm{e}^{\mathrm{i}\omega[t-\mathbf{n}\cdot\mathbf{r}(t)/c]} \quad [59]$$

Equation [58] gives the energy radiated by the electron in the direction of the unit vector, per unit solid angle, per unit frequency ω. Radiation damping is ignored throughout.

Let us consider the simplest case where the electron orbit is strictly a periodic function of time with period T and a net displacement $\mathbf{r}_0$ per period. Thus, we have for all integers m (positive, negative, or zero):

$$\beta(t+mT) = \beta(t), \quad \mathbf{r}(t+mT) = m\mathbf{r}_0 + \mathbf{r}(t) \quad [60]$$

Equation [59] may then be written as, because of Thomson scattering the energy of the scattered photon is much less than the electron rest mass in the average rest frame of the electron, i.e.:

$$\frac{\hbar\omega_{\text{L}}}{mc^3} \ll 1 \quad [61]$$

$$\mathbf{F}(\omega) = \sum_{m=-\infty}^{\infty}\int_{mT}^{(m+1)T} \mathrm{d}t\,\beta(t)\mathrm{e}^{\mathrm{i}\omega[t-\mathbf{n}\cdot\mathbf{r}(t)/c]} = \sum_{m=-\infty}^{\infty} f(\omega)\mathrm{e}^{\mathrm{i}m\omega[T-\mathbf{n}\cdot\mathbf{r}_0/c]} \quad [62]$$

where

$$f(\omega) = \int_0^T \mathrm{d}t\,\beta(t)\mathrm{e}^{\mathrm{i}\omega[t-\mathbf{n}\cdot\mathbf{r}(t)/c]} \quad [63]$$

and we have used eqn [60]. Upon using $\sum_m \mathrm{e}^{\mathrm{i}mx} = \sum_m 2\pi\delta(x - 2m\pi)$ in the last infinite sum in eqn [64] and the property of the Dirac delta function, $\delta(ax) = (1/a)\delta(x)$, we obtain from eqns [62] and [63], the following expression for the spectrum:

$$\mathbf{F}(\omega) = \sum_{m=-\infty}^{\infty} \mathbf{F}_m\delta(\omega - m\omega_1) \quad [64]$$

$$\omega_1 = \frac{2\pi}{T - \mathbf{n}\cdot\mathbf{r}_0/c}, \quad [65]$$

$$\mathbf{F}_m = \frac{\omega_1}{2\pi}\int_0^T \mathrm{d}t\,\bar{\beta}(t)\mathrm{e}^{\mathrm{i}m\omega_1[t-\mathbf{n}\cdot\mathbf{r}(t)/c]} \quad [66]$$

Note from eqn [64] that the radiation spectrum is discrete, for strictly periodic motion of the electron. The base frequency of this spectrum, ω_1, depends on the periodicity (T) of the electron, on the electron's net displacement ($\mathbf{r}_0$) in one such period, and on the direction ($\hat{n}$) in which the radiation is observed. Thus, the radiation spectrum is, in general, not at the

harmonic frequency of the laser (nor at the harmonics of the electron orbital frequency, $2\pi/T$). It would thus be wrong to simply insert $\omega = n\omega_0$ in eqn [63] and to replace the electron's orbital period T there by the laser's optical period $2\pi/\omega_0$ and consider the resultant value of that integral to give the spectral amplitude of the radiation at the nth harmonic of the laser frequency. Erroneous conclusions regarding generation of high laser harmonics have appeared in the literature based on such an intuitive (but incorrect) substitution.

The power, p_m (in erg/s), radiated at the harmonic frequency $\omega = m\omega_1$ per unit solid angle in the direction of the unit vector $\hat{n}$ is then given by (from eqns [58] and [64]):

$$p_m = \frac{e^2 m^2 \omega_1^2}{4\pi^2 c} |\mathbf{n} \times \mathbf{F}_m|^2 \qquad [67]$$

where the dimensionless spectral amplitude $\mathbf{F}_m$ is given in eqn [66]. It is easy to show from eqns [60] and [65] that the integrand in eqn [66] is a periodic function of t of period T. Integrals of this type are readily evaluated by the Romberg method. The radiation spectrum observed *exactly* in the forward direction of the laser ($\hat{n} = \hat{z}$) always has only one discrete frequency, $\omega = \omega_1 = \omega_0$, which is easily shown from eqn [65] upon using the first equality of eqn [53]. This statement is true regardless of the velocity of the electron or the laser intensity, and may easily be deduced from eqn [66] for this case. However, for an energetic electron beam that is almost co-propagating with the laser, such as that produced by the laser itself, high harmonics at the laser frequency may be observed in the direction just slightly off the laser direction.

In the backscattering direction of the laser ($n = -z$), if we set $\theta_{\text{in}} = 0$, one obtains the following expressions for ω_1 and p_m, the backscatter power at $\omega = m\omega_1$ (from eqn [67]):

$$\frac{\omega_1}{\omega_0} = \left(\frac{2}{2+a^2}\right)\left(\frac{1-\beta_{z0}}{1+\beta_{z0}}\right) = \left(\frac{2}{2+a^2}\right)\gamma_0^2(1-\beta_{z0})^2 \qquad [68]$$

$$p_m = \frac{A}{\gamma_0^2(1-\beta_{z0})^2}\left(\frac{\omega_1}{\omega_0}\right)^4 s_m \qquad [69]$$

where $A = e^2\omega_0^2/4\pi^2 c = 0.69[\lambda(1\ \mu\text{m})]^{-2}$ erg/s, $s_m = 0$ for $m = 0, \pm 2, \pm 4, \ldots$, and for $m = \pm 1, \pm 3, \pm 5, \ldots$,

$$s_m = (a\pi)^2 m^2 \left[J_{\frac{(m-1)}{2}}(m\kappa) - J_{\frac{(m+1)}{2}}(m\kappa)\right]^2 \qquad [70]$$

$$\kappa = \frac{a^2}{2(a^2+2)} \qquad [71]$$

The $J_\nu(x)$ above are Bessel functions of the first kind of order ν. They also appear in the quantity '$[JJ]$' or '$F_m(K)$' in the FEL/synchrotron literature, where K is the undulator/wiggler parameter K. Setting $K = a$, one finds $s_m = \pi^2(1 + K^2/2)^2 F_m(K)$. It is easy to show that $s_m = s_{-m}$ for all odd integers m. Note that the relative spectral shape of s_m depends only on a, and is independent of the electron beam energy. The discrete spectrum for small a approaches a continuum for $a^2 \gg 1$.

The maximum values of s_m, occurring at $m = M$ with a value s_M, are shown in Figure 9. Note that the frequency component $\omega = M\omega_1$ contains the highest backscattered power. In terms of the laser frequency, the frequency component $\omega = N\omega_0$ would contain the highest backscatter power, where $N = M\omega_1/\omega_0$. The total backscatter power, P_{T} (in ergs/s), per unit solid angle in the $\mathbf{n} = -\mathbf{z}$ direction is then given by $P_{\text{T}} = \Sigma p_m$, where the sum is taken over all odd values of m:

$$P_{\text{T}} \approx \frac{\gamma_0^6(1-\beta_{z0})^6}{(\lambda/1\ \mu\text{m})^3} \times \begin{cases} 13.7a^3, & (a < 0.3) \\ 11.1/a, & (a \gg 1) \end{cases} \qquad [72]$$

To summarize, due to the rapid acceleration of electrons in the direction of the light wave and decrease of the oscillation frequency in the deeply relativistic regime, the photon energy from relativistic nonlinear scattering scales only linearly with laser field strength, $\sim a$. However, the relativistic motion of the electron results in a reduction in the angle of the scattered light, such that the harmonics are generated in a low-divergence-angle forward-propagating beam. There is also a relativistic Doppler shift ($\sim\gamma^2$). For instance, this mechanism will be used in high-energy physics experiments to cleanly make constituent particles in the gamma–gamma ($\gamma\gamma$) collider, in which gamma rays with energy 200 GeV will be generated by Compton scattering 1 eV photons from 250 GeV energy conventionally accelerated electron beams ($\gamma = 5 \times 10^{11}$).

Using a much smaller, laser-based accelerator (discussed in the following section), a 1 eV photon can be upshifted by this Doppler shift to an energy of 50 keV, corresponding to subatomic spatial resolution and of interest in medical diagnostics, by an electron beam with only $\gamma = 200$ (100 MeV). In this case, the maximum efficiency is obtained for laser fields a of order unity. When compared with conventional light sources based on cm-wavelength magnetostatic wigglers, the electromagnetic wigglers

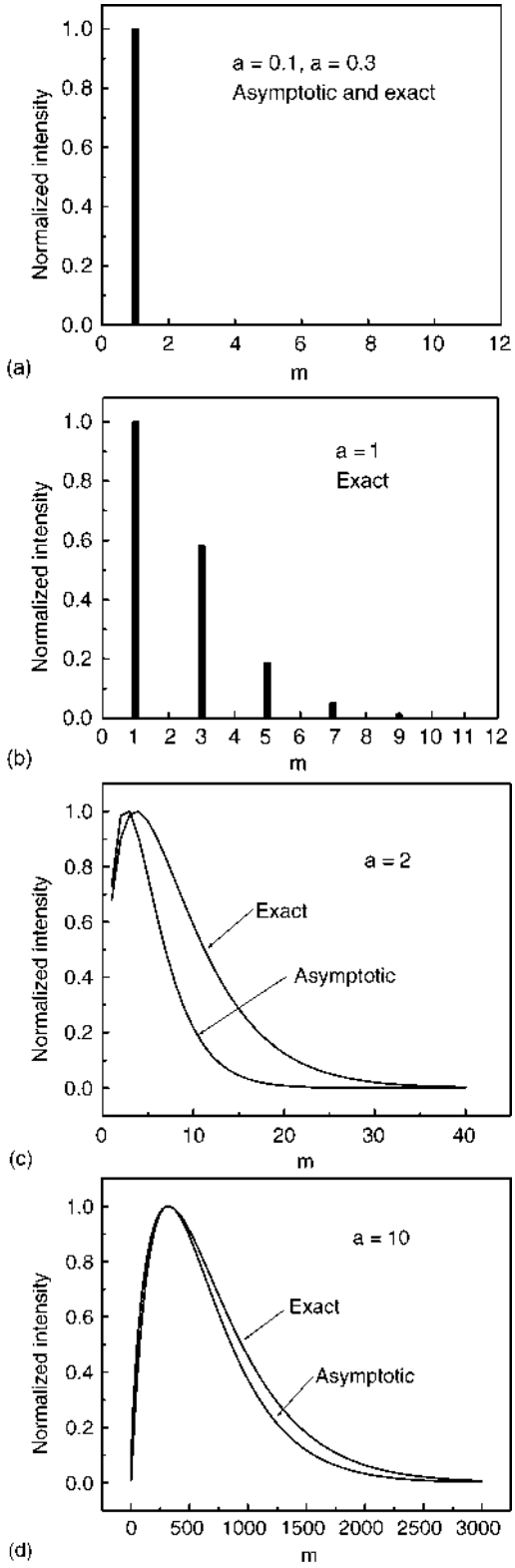

Figure 9 Normalized spectral distribution of s_m, at frequency $\omega = m\omega_1$. Here, s_m is normalized with respect to the maximum value s_M, occurring at $m = M$. (Reproduced from Lau Y, He F and Umstadter D (2003) Nonlinear Thomson scattering: A tutorial. *Physics Plasmas* 10: 2155, with permission from the American Institute of Physics.)

of such all-optical-laser-based EUV sources have ten-thousand times shorter wavelength (micron-scale). Thus, the total length of the wiggler region is correspondingly smaller (only mm in length). Another consequence of this is that the frequency upshift required to reach a given output wavelength is also ten thousand times smaller. Also, given that the required electron energy scales as the square root of the upshift, the required electron energy can be one-hundred-times lower (10–100 MeV). It follows from this, and the fact that the field gradients of these accelerators can be ten thousand times higher (1 GeV/cm) than convention RF-based accelerators, that the size of the accelerating region can, in principle, be a million times smaller (only mm in length). Besides its small size, this EUV source can produce femtosecond duration pulse and be synchronized with a relatively low jitter with another femtosecond light pulse having a different wavelength (by virtue of the possibility of deriving the two pulses from the same laser pulse); this is advantageous for the study of ultrafast pump-and-probe photo-initiated processes. The exceptionally low transverse emittance of laser-accelerated electron beams may even make it possible to generate coherent XUV radiation by means of the self-amplified spontaneous emission (SASE) free-electron lasing (FEL) mechanism.

Collective Plasma Response

The above equations apply to electron motion in vacuum. In the case of electrons in plasmas, the collective plasma electron response needs to be considered. A theory for the 1D nonlinear interaction of intense laser fields with electrons has been developed.

The variables are normalized such that

$$\mathbf{a} = e\mathbf{A}/mc^2, \qquad \phi = e\Phi/mc^2 \qquad [73]$$

$$\mathbf{u} = \mathbf{p}/mc, \qquad \beta = \mathbf{v}/c, \qquad \mathbf{u} = \gamma\beta \qquad [74]$$

$$\gamma = (1 + u^2)^{1/2} = (1 - \beta^2)^{-1/2} \qquad [75]$$

The relativistic force equation (Lorentz equation) is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u} = \nabla\phi + \frac{\partial}{\partial t}\mathbf{a} - \bar{\beta}\times(\nabla\times a) \qquad [76]$$

The energy equation ($\mathbf{u}\cdot$ force equation) is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma = \beta\cdot\left(\frac{\partial}{\partial z}\phi + \frac{\partial}{\partial t}\mathbf{a}\right) \qquad [77]$$

The transverse force equation (1D) is

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}_\perp = \left(\frac{\partial}{\partial t} + \beta_z \frac{\partial}{\partial z}\right)\mathbf{a}_\perp \qquad [78]$$

Since $\mathbf{a} = \mathbf{a}(z, t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{a}_\perp = \left(\frac{\partial}{\partial t} + \beta_z \frac{\partial}{\partial z}\right)\mathbf{a}_\perp \qquad [79]$$

hence:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{u}_\perp - \mathbf{a}_\perp) = 0 \qquad [80]$$

This is just conservation of transverse canonical momentum. Assuming $\mathbf{u}_\perp = 0$ prior to the laser interaction gives

$$\mathbf{u}_\perp = \mathbf{a}_\perp \qquad [81]$$

which is the same as eqn [24].

The electron response to the normalized scaler and vector potentials of the form $\phi = \phi(z - ct)$ and $\mathbf{a} = \mathbf{a}(z - ct)$, which are a function of only $\xi = z - ct$, is completely described by the following constants of the motion:

$$\beta_\perp - a_\perp/\gamma = 0, \qquad [82]$$

$$\gamma(1 - \beta_z) - \phi = 1, \qquad [83]$$

$$n(1 - \beta_z) = n_0, \qquad [84]$$

where $\gamma = (1 - \beta^2)^{-1/2}$ is the relativistic factor and n is the electron density. This allows the various electron quantities to be specified solely in terms of the potentials:

$$\beta_\perp = \frac{2(1 + \phi)a_\perp}{(1 + a_\perp^2) + (1 + \phi)^2}, \qquad [85]$$

$$\beta_z = \frac{(1 + a_\perp^2) - (1 + \phi)^2}{(1 + a_\perp^2) + (1 + \phi)^2}, \qquad [86]$$

$$\gamma = \frac{(1 + a_\perp^2) + (1 + \phi)^2}{2(1 + \phi)}, \qquad [87]$$

$$\frac{n}{n_0} = \frac{(1 + a_\perp^2) + (1 + \phi)^2}{2(1 + \phi)^2} \qquad [88]$$

In the single particle limit, $\phi \equiv 0$, eqns [26] and [30] are retrieved. For a long pulse interacting with a dense plasma, $\tau_\mathrm{L} \gg \omega_\mathrm{p}^{-1}$ (i.e., neglecting wakefield effects) it can be shown that $(1 + \phi) \simeq (1 + a_0^2)^{1/2}$ for circular polarization and $(1 + \phi) \simeq (1 + a_0^2/2)^{1/2}$ for linear polarization.

Propagation

To understand the propagation of high-intensity light in plasma, we need to understand how the dielectric properties of a plasma medium are affected by the relativistic electron mass change. A wave equation is found by taking the curl of Faraday's law:

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \qquad [89]$$

and using Ampere's law:

$$\nabla \times \mathbf{B} = 4\pi \mathbf{J}/c + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \qquad [90]$$

gives

$$\begin{aligned}\nabla \times \nabla \times \mathbf{E} &= -\nabla^2 \mathbf{E} + \nabla(\nabla \cdot \mathbf{E}) \\ &= \frac{1}{c^2}\left[4\pi \frac{\partial \mathbf{J}}{\partial t} - \frac{\partial^2 \mathbf{E}}{\partial t^2}\right]\end{aligned} \qquad [91]$$

where $\mathbf{J} = \sum n_i q_i v_i$ and the summation over the index i is done over all charged species of electrons and ions. Using Coulomb's law:

$$\nabla \cdot \mathbf{E} = 4\pi\rho \qquad [92]$$

where $\rho = \sum n_i q_i$, and assuming a uniform plasma, the second term on the left vanishes. Substituting eqn [13] for the velocity, we find

$$\left(-\frac{\partial^2}{\partial t^2} + c^2\nabla^2 - \omega_\mathrm{p}^2\right)\mathbf{E} = 0 \qquad [93]$$

where

$$\omega_\mathrm{p} = (4\pi n_i q_i^2/m_i)^{1/2} \qquad [94]$$

is the plasma frequency. Assuming again plane waves and Fourier analyzing, yields the well-known dispersion relation for electromagnetic waves in plasma:

$$\omega^2 = \omega_\mathrm{p}^2 + c^2 k^2 \qquad [95]$$

The index of refraction can be written fully relativistically as

$$\eta^2 = \frac{c^2 k^2}{\omega^2} = \frac{c^2}{v_\phi^2} = 1 - \frac{\omega_\mathrm{p}^2}{\omega^2} \qquad [96]$$

where v_ϕ is the phase velocity of the light wave. Assuming infinitely massive ions, the plasma frequency can be written as

$$\omega_\mathrm{p} = \frac{\omega_{\mathrm{p}0}}{\gamma^{1/2}} = \left(\frac{4\pi n_\mathrm{e} e^2}{\gamma m_0}\right)^{1/2} \qquad [97]$$

where ω_{p0} is the plasma frequency in a quiescent plasma, e is the electron charge, m_0 is the electron rest mass, and n_e is the plasma electron density. A change in mass changes the plasma frequency, which in turn modifies the index of refraction and the velocity of the light wave. The light's phase velocity then depends on the laser intensity. This can be seen clearly if we expand the phase velocity for small field strength ($a_0 \ll 1$):

$$v_\phi = \frac{c}{\eta} \sim c\left[1 + \frac{\omega_{p0}^2}{2\omega^2}\left(1 - \frac{\langle a_0^2\rangle}{2}\right)\right] \qquad [98]$$

An on-axis minimum of the phase velocity [i.e., $v_\phi(r) > v_\phi(0)$] can be created by a laser beam with an intensity profile peaked on axis, such as with a Gaussian beam, causing the wavefronts to curve inward and the laser beam to focus, as shown in Figure 10. When this focusing effect just balances the defocusing caused by diffraction, the laser pulse can propagate over a longer distance than it could in vacuum, while maintaining a small cross-section. This mechanism is referred to as relativistic self-guiding.

Relativistic Self-Focusing

The threshold for relativistic self-focusing can be obtained using the fully relativistic formalism with the vector potential, as in our previous discussion of electron motion in vacuum. In this case, we can rewrite eqn [91] as

$$\nabla^2\mathbf{a} - \frac{1}{c^2}\frac{\partial^2\mathbf{a}}{\partial t^2} = \frac{\omega_p^2}{c^2}\frac{n_e\mathbf{a}}{n_0\gamma} \qquad [99]$$

where $n_0 = n_i$ is the uniform background density. For a Gaussian beam focused in vacuum, where

$$\mathbf{a}(\mathbf{r}, t) = \mathrm{Re}\left\{\mathbf{a}_0(\mathbf{r}, t)\exp[\mathrm{i}(\mathbf{k}\cdot\mathbf{r} - \omega t)]\right\} \qquad [100]$$

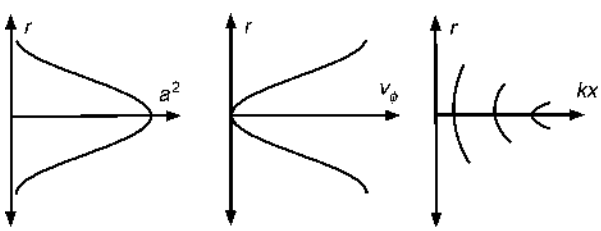

Figure 10 Relativistic self-focusing occurs when an on-axis peak in light intensity (left) produces an on-axis dip in the phase velocity (middle), which acts like a positive lens to cause the phase fronts to curve inward (right). (Reproduced from Umstadter D (2003) Relativistic laser–plasma interactions. *Journal Physics D: Applied Physics* 36: R151–R165, with permission from IOP Publishing Ltd.)

and the amplitude $\mathbf{a}(\mathbf{r}, t)$ varies much less less in r and t than does the phase:

$$|\partial a_0/\partial t| \ll |\omega a_0|, \qquad |\partial a_0/\partial z| \ll |k a_0| \qquad [101]$$

then

$$a_0(r, z) = \frac{e^{-r^2/(r_0^2(1+z^2/L_r^2))}}{\sqrt{1 + z^2/L_R^2}}\exp\left\{-\mathrm{i}\arctan\left(\frac{z}{L_R}\right) + \mathrm{i}\left(\frac{r}{r_0}\right)^2\frac{z/L_R}{1 + z^2/L_R^2}\right\} \qquad [102]$$

satisfies the envelope equation (assuming second deriviatives are small compared to first derivatives), where L_R is the Rayleigh range and r_0 is the focal radius. In plasma, eqn [99] can then be written in this envelope approximation as

$$\left[\nabla_\perp^2 + 2\mathrm{i}k\frac{\partial}{\partial z}\right]a = -\frac{\omega_p^2}{c^2}\frac{|a|^2}{4}a \qquad [103]$$

The power at which the first and last terms of eqn [103] are in balance is

$$P = \frac{\pi R^2 I_0}{2} \qquad [104]$$

An equation for the beam radius R can thus be found:

$$\frac{\mathrm{d}^2R(z)}{\mathrm{d}z^2} = \frac{4}{k^2 r_0^2}\left[1 - \frac{1}{22}\frac{\omega_p^2}{c^2}|a_0|^2 r_0^2\right] \qquad [105]$$

Self-focusing occurs when the two terms in the brackets in eqn [105] are equal, which gives for the critical power:

$$P_{crit} \simeq 17.4\cdot(\omega/\omega_p)^2\ \mathrm{GW} \qquad [106]$$

This corresponds to 1 TW for 1 μm light focused into a gas with electron density of $10^{19}/\mathrm{cm}^3$. Note that this is a power threshold, not an intensity threshold, because the tighter the focusing, the greater the diffraction (the first term in eqn [103]). Numerous recent experiments have confirmed this focusing mechanism when $P > P_{crit}$.

Any spatial variation of the laser intensity will act to push an electron to regions of lower intensity through the so-called ponderomotive force:

$$\mathbf{F} = \nabla\gamma = \nabla\sqrt{1 + a^2} = (2\gamma)^{-1}\nabla a^2 \qquad [107]$$

$$\mathbf{F}_{\text{pond}} = -\frac{m_0 c^2}{8\pi} \frac{\nabla a_0^2}{\sqrt{1+\frac{a_0^2}{2}}} \cdot \left[\left(\frac{2}{a_0^2} + 1 \right) E(2\pi, \kappa) - \frac{2}{a_0^2} F(2\pi, \kappa) \right] \quad [108]$$

where

$$\kappa = \frac{a_0^2/2}{1 + a_0^2/2} \quad [109]$$

and F and E are the elliptic integrals of the first and second kind, respectively. In the low intensity limit, i.e., when $a_0^2 \ll 1$;

$$\mathbf{F}_{\text{pond}} = -\frac{m_0 c^2}{4} \nabla a_0^2 \quad [110]$$

That is, the laser ponderomotive force is roughly proportional to the gradient of laser intensity.

A Gaussian-shaped laser's ponderomotive force will tend to expel electrons radially from the region of the axis, so called 'electron cavitation.' If the ponderomotive force is high enough for long enough, the charge displacement due to expelled electrons (in either the lateral or longitudinal direction) will eventually cause the ions to move as well through the Coulomb electrostatic force, forming a density channel. Because $n_e(0) < n_e(r)$, and thus $v_\phi(0) < v_\phi(r)$, this enhances the previously-discussed relativistic self-guiding and can itself guide a laser pulse.

For plasmas created by photo-ionization of a gas by a Gaussian laser pulse, the density will be higher on the axis than off the axis. If we instead expand the phase velocity in terms of changes in density:

$$\frac{v_\phi}{c} \sim 1 + \frac{\omega_{p0}^2}{\omega^2} \frac{\Delta n_e}{n_e} \quad [111]$$

The phase velocity will thus be higher on axis, which will tend to defocus the light and increase the self-guiding threshold. In order to avoid this, gases with low atomic number and thus fewer available electrons, such as H_3, are commonly used as targets.

Raman Scattering, Plasma Wave Excitation and Electron Acceleration

The local phase velocity, described in eqns [98] and [111], can also vary longitudinally if the intensity and/or electron density does. Local variation in the index of refraction can 'accelerate' photons, i.e., shift their frequency, resulting in photon bunching, which in turn bunches the electron density through the ponderomotive force (F), and so on. When the laser pulse duration is longer than an electron plasma period, $\tau \gg \tau_p = 2\pi/\omega_p$, this photon and electron bunching grows exponentially, leading to the stimulated Raman scattering instability. Energy and momentum must be conserved when the electromagnetic wave $(\omega_0, \mathbf{k}_0)$ decays into a plasma wave $(\omega_p, \mathbf{k}_p)$ and another light wave $(\omega_0 - \omega_p, \mathbf{k}_0 - \mathbf{k}_p)$.

From an equivalent viewpoint, the process begins with a small density perturbation, Δn_e, which, when coupled with the quiver motion, eqn [13], drives a current $J = \Delta n_e e v_e$. This current then becomes the source term for the wave equation (eqn [91]), driving the scattered light wave. The ponderomotive force, due to the beating of the incident and scattered light wave, enhances the density perturbation, creating a plasma wave and the process begins anew. In three dimensions, a plasma wave can be driven when transverse self-focusing and stimulated Raman scattering occur together, a process called the self-modulated wakefield instability.

Two conditions must be satisfied for self-modulation to occur in the plasma. First, the laser pulse must be long compared to the plasma wave, $L \gg \lambda_p$. This allows the Raman instability time to grow, and it allows for feedback from the plasma to the laser pulse to occur. Second, the laser must be intense enough for relativistic self-focusing to occur, $P > P_c$, so that the laser can be locally modified by the plasma. Under these conditions, the laser can form a large plasma wave useful for accelerating electrons.

As the long laser pulse enters the plasma, it will begin to drive a small plasma wave due to either forward Raman scattering or the laser wakefield effect from the front of the laser pulse. This small plasma wave will have regions of higher and lower density with both longitudinal and radial dependence. That is, the plasma wave will be three-dimensional in nature with a modulation along the propagation direction of the laser and a decay in the radial direction to the ambient density (see **Figure 11**). The importance of this lies with how it affects the index of refraction in the plasma. In the regions of the plasma wave where the plasma density is lower, the radial change in the index of refraction is negative, $\partial n(r)/\partial r < 0$. This means that this part of the plasma acts like a positive lens and focuses the laser. Whereas regions of the plasma wave where the density is higher, $\partial n(r)/\partial r > 0$, the opposite occurs and the laser defocuses. This has the effect of breaking up the laser pulse into a series of shorter pulses of length $\lambda_p/2$ which will be separated by the plasma period. The instability occurs because of how the plasma responds to this. Where the laser is more tightly focused, the ponderomotive force will be

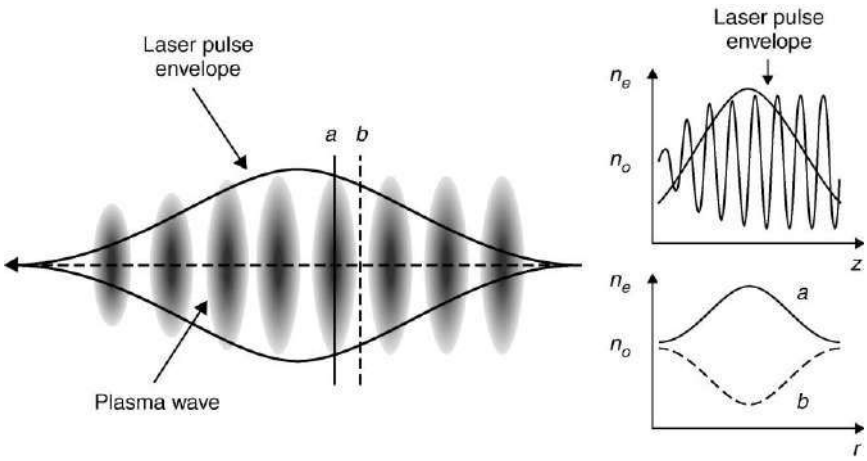

Figure 11 The plasma wave generated by a SMLWFA is three-dimensional in nature. Note that the darker regions correspond to areas of higher plasma density. The graphs to the right represent lineouts of the plasma density longitudinally and radially at the indicated points. (Reproduced with permission from Wagner R (1998) Laser–plasma electron accelerators and nonlinear relativistic optics. PhD thesis, University of Michigan.)

greater and will tend to expel more electrons. This decreases the density in these regions even further, resulting in more focusing of the laser. This feedback rapidly grows, hence the instability.

The phase velocity of the plasma wave in the case of forward scattering is equal to the group velocity of the beat wave, which for low-density plasma is close to the speed of light, as can be seen from the relation:

$$v_\phi = \omega_p/k_p = \Delta\omega/\Delta k = v_g = c\eta \sim c \qquad [112]$$

where eqn [94] and $\omega_p^2 \ll \omega^2$ were used to show that η is close to unity. Such relativistic plasma waves can also be driven by short pulses ($\tau \sim \tau_p$). In this case, the process is referred to as laser–wakefield generation, referring to the analogy with the wake driven by the bow of a boat moving through water, but the mechanism is similar (except it has the advantage that the plasma wave is driven linearly instead of as an instability).

In either case, the resulting electrostatic plasma wave can continuously accelerate relativistic electrons with enormous acceleration gradients. The gradient can be estimated from eqn [92] and the fact that because

$$\nabla \cdot \mathbf{E} \sim E\omega_p/c \propto E\sqrt{n_e} \qquad [113]$$

then

$$E \sim \sqrt{n_e}\ \text{eV/cm} \qquad [114]$$

corresponding to 1 GeV/cm for $n_e = 10^{18}\ \text{cm}^{-3}$. Because this gradient is four orders of magnitude greater than achieved by conventional accelerators (based on fields driven by radio-frequency waves pumped into metal cavities), laser-driven plasma accelerators have received considerable recent attention. They have been shown to accelerate an amount of electron charge (100 nC) comparable to that from conventional accelerators and to have superior transverse geometrical emittance (product of divergence angle and spotsize, similar to the $f/\#$ in light optics). However, their longitudinal emittance is currently much inferior, energy spreads of 100%. They have been shown to be useful for much of the same applications: radio-isotope production, radiation chemistry, as well as X-ray, proton, and neutron generation. Once the longitudinal emittance can be reduced, they may be advantageous for, among other applications, injectors (especially of short-lived unstable particles) into larger conventional accelerators for high-energy physics research and light sources, and, as discussed in the section on radiation from relativistic electrons about, as stand-alone all-optically driven ultrashort-pulse duration X-ray sources.

The SIMLAC code has been used to study wakefield generation and laser propagation in the limit $a^2 \ll 1$. It draws from nonlinear optics models and treats propagation in the group velocity frame. In this idealized model (which assumes perfect Gaussian beams), the pulse and wake are maintained over long enough propagation distances to accelerate an electron to GeV energy, as shown in Figure 12. A three-dimensional envelope equation for the laser field was derived that includes nonparaxial effects, wakefields, and relativistic nonlinearities.

The resonant wakefield has been characterized experimentally by temporal interferometry, as shown in Figure 13. However, this was done only for the tight-focusing case in which the laser spotsize is much

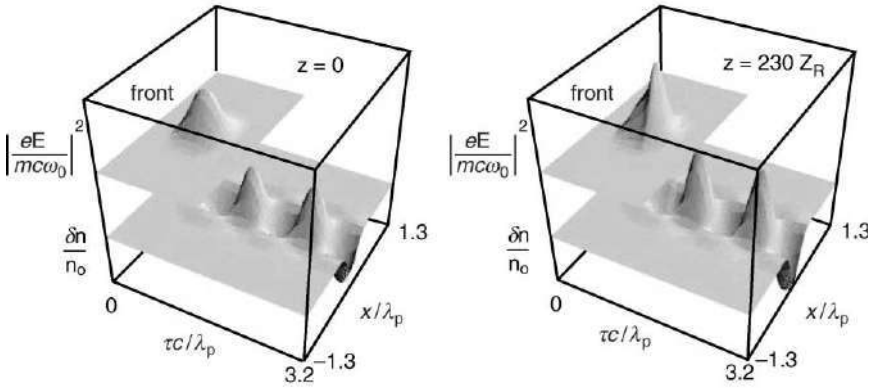

Figure 12 The 'standard' resonant wakefield simulated with SIMLAC, a code that moves at the light pulse's group velocity. (Reproduced with permission from Umstadter D (2001) Review of physics and applications of relativistic plasmas driver by ultra-intense lasers. *Physics Plasmas* 8: 1774.)

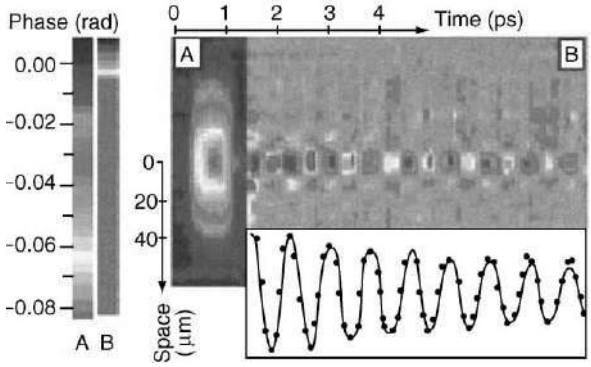

Figure 13 Typical result of a phase shift measurement to study a resonantly excited laser wakefield plasma wave by means of time-domain interferometry. Parts (A) and (B) have different color scales. The bottom graph is a line out of part (B) along the laser axis. (Reproduced with permission from Marquès JR, Dorchies F, Audebert P, Ceindre JP, Amiranoff F, Gauthier JC, Hammoniaux G, Antonetti A, Chessa P, Mora PTM and Antonson J (1997) Frequency increase and damping of nonlinear electron plasma oscillations in cylindrical symmetry. *Physics Review Letters* 78: 3463. Copyright (1995) by the American Physical Society.)

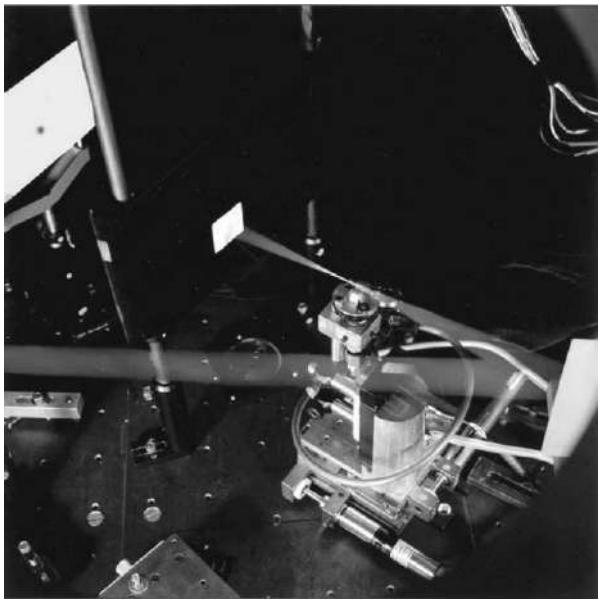

Figure 14 Artistically enhanced photograph of the acceleration of an electron beam by a laser interacting with a gas jet inside a vacuum chamber. The laser crosses the picture from left to right and is focused by a parabolic mirror (right side of the picture). The supersonic nozzle (shown in the middle of the picture) is positioned with micron accuracy with a 3-axis micropositioner. The e-beam makes a small spot on a white flourescent (LANEX) screen, shown in the upper left-hand corner of the picture. (Reproduced with permission from Umstadter D (2001) Review of physics and applications of relativistic plasmas driver by ultra-intense lasers. *Physics Plasmas* 8: 1774, with permission from the American Institute of Physics.)

smaller than the plasma wave wavelength ($r_l \ll \lambda_p$) and thus the transverse wakefield was much greater than the longitudinal wakefield.

A typical experimental setup, used to study electron acceleration, is shown in Figure 14.

Dramatic reduction of the angular divergence of a laser accelerated electron beam was observed with increasing laser power above the relativistic self-focusing threshold, as shown in Figure 15.

Interactions with Solid-Density Targets

The generation of electrons by high-intensity laser light interacting with solid targets can generate energetic X-rays, accelerate other types of particles and induce nuclear reactions, as illustrated in Figure 16. For instance, high-order harmonics have been generated by the oscillation of the critical density surface, in the so-called moving mirror model. Bright X-rays, originating from Bremstrahlung caused by electron collisions with high-Z atoms in solid targets, have created isotopes by means of photofission. Laser-accelerated electron energies and

angular distributions have been inferred from analyzing (γ, n) and $(\gamma, 2n)$ reactions in composite Pb/Cu targets and in Ta/Cu targets. Positrons were created by colliding laser-accelerated electrons with a tungsten target.

When electrons are heated to high temperatures or accelerated to high energies, they can separate from plasma ions. Such charge displacement creates an electrostatic sheath, which eventually accelerates the ions. The ions are pulled by the charge of the electrons and pushed by the other ions' unshielded charges (similar to the 'Coulomb explosion' that can occur during the ionization of atoms). When the charge displacement is driven by thermal expansion, as in long-pulse (low-power) laser–plasma experiments, the maximum ion energies are limited to less than 100 keV. However, when the charge displacement is driven by direct laser heating, as in short-pulse high-power laser–plasma experiments, multi-MeV ion energies are possible. This was first shown with gas jet targets, in which case the ions were accelerated radially into 2π, and then later with thin solid-density-films, in which case the ions were accelerated into collimated beams. In the latter case, hydrocarbons and water on the surface of the film can become ionized and provide a source of protons to be accelerated.

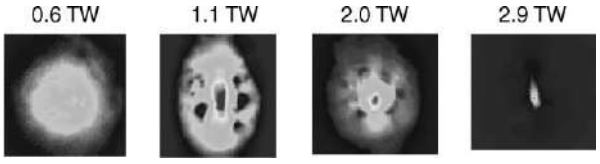

Figure 15 Images of the spatial profiles of the electron beam measured by a ccd camera imaging a LANEX screen at a distance of 15 cm from the gas jet for various laser powers. The divergence angle of the beam decreases to a value of $\Delta\theta = 1°$ at a power of 2.9 TW, corresponding to a transverse geomtrical emittance of just $\varepsilon_\perp \lesssim 0.06\pi$ mm-mrad. (Reproduced with permission from Umstadter D (2001) Review of physics and applications of relativistic plasmas driver by ultra-intense lasers. *Physics Plasmas* 8: 1774, with permission from the American Institute of Physics.)

An intense laser can ponderomotively heat electrons. If the laser contrast is high, vacuum heating can occur in the following manner. When light encounters a sharp interface between vacuum and solid density, the electromagnetic field becomes evanescent in the region above the critical density. The instantaneous '$v \times B$' force can push electrons in the direction of the light's propagation vector; it also has a frequency twice that of the pump and a magnitude proportional to the square of the normalized vector potential, a_0^3. Thus, electrons can only complete half of their figure-eight orbits, on the vacuum side, gaining relativistic energies; they move through the overdense region without the electromagnetic field to pull them back. An electrostatic sheath can thus form, which will accelerate the ions left behind. Another important heating mechanism is stochastic heating, which occurs when the light that is reflected from the critical surface beats with the incoming wave to create a standing wave. The motion of electrons in such a wave can become chaotic, resulting in a large increase in electron temperature (>100 keV).

As the heated electrons propagate through a solid, they can instantaneously field-ionize the neutral atoms of the solid. This will both modify the solid's

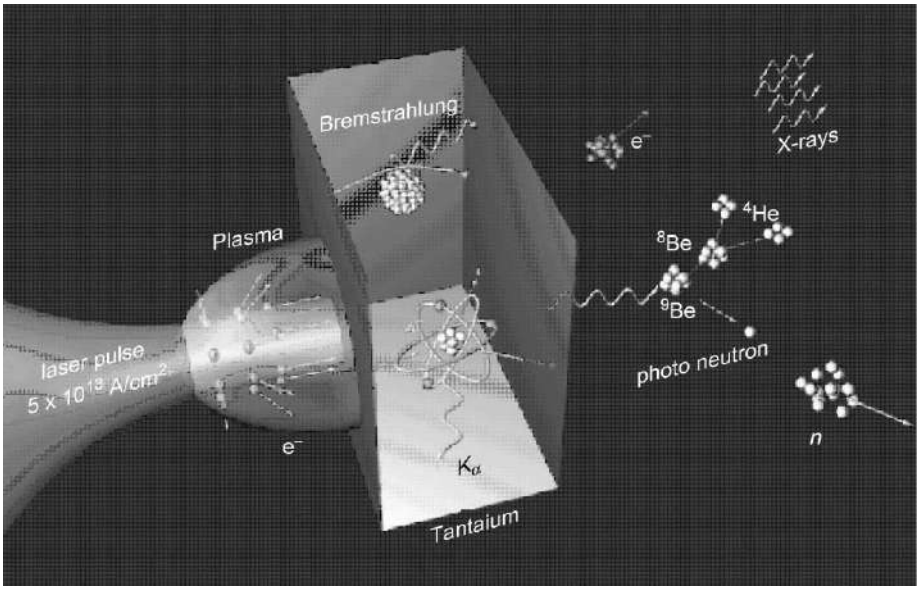

Figure 16 Illustration of the various mechanisms of particle acceleration and X-ray generation in high-intensity laser interactions with solid-density targets. (Courtesy of R. Sauerbrey.)

conductivity and provide a source of protons on the rear-side of the target. If the film is thin enough, the electrons can pass through, and create a sheath on the rear-side of the target. This latter mechanism has been dubbed the target normal sheath acceleration mechanism. The ions from thin foils have been claimed to originate from both the front and rear-side of the foil.

Several groups have reported the observation of ions originating from thin-film solid-density targets. Unlike previous long-pulse experiments, the ions were accelerated along the direction normal to the side of the target, that is opposite to that upon which the laser was incident. The ions generally originate from water or hydrocarbons on the surface of the material. The acceleration results from several different mechanisms, which may be occurring simultaneously. Charge-displacement is again common to all, with the electrons being heated ponderomotively, such as by Brunel, $J \times B$ or stochastic heating. In one case, the electrostatic sheath is formed at the backside of the ionization layer formed on the side of the target upon which the laser is incident (front side). In another case, the electrostatic sheath is formed by field ionization of the ion layer on the opposite side of the thin film target (back-side), the target normal sheath acceleration (TNSA) hypothesis. Numerical simulations show evidence for both front- and back-side acceleration.

The results of these experiments indicate that a large number of protons (10^{13} p) can be accelerated, corresponding to source current densities (10^8 A/cm^2) that are nine orders of magnitude higher than produced by cyclotrons, but with comparable, or even lower, transverse emittances ($\varepsilon_\perp \leq 1.0\pi$ mm-mrad). Proton energies up to 60 MeV have been observed in experiments at intensities exceeding 10^{20} W/cm^2 (using a petawatt power laser). The high end of the proton spectrum typically has a sharp cut-off, but is a continuum. In one experiment, protons were observed to be emitted in ring patterns, the radii of which depend on the proton energy, which was explained by self-generated magnetic fields.

A typical experimental setup used to observe the acceleration of ions is shown in Figure 17. The production of radionuclides have been used as an ion energy diagnostic. In another example of a nuclear reaction initiated by an intense laser, neutrons

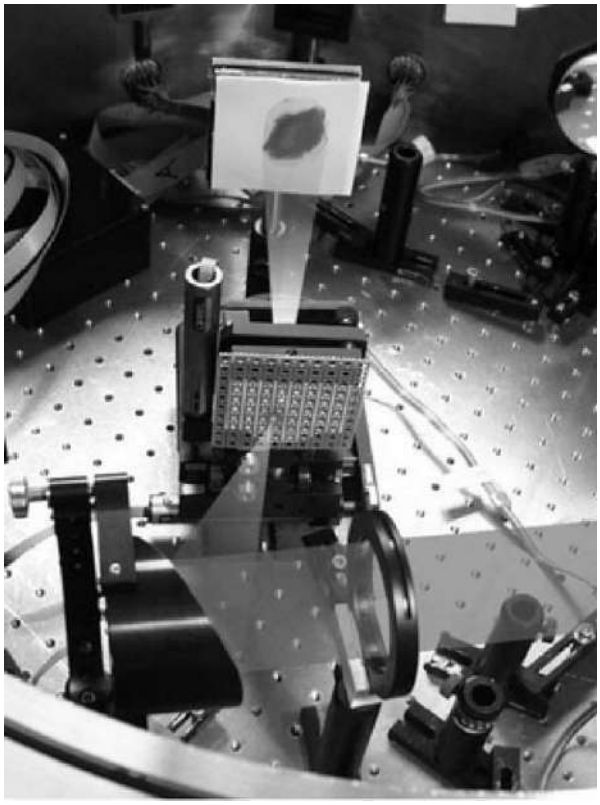

Figure 17 Artistically enhanced photograph of the typical setup used to observe the acceleration of ions. The laser (shown in the foreground) is focused with an off-axis parabola onto a thin-foil, held by a mesh that is positioned by a 3-axis micropositioner. CR-39, a nuclear track detector (shown in the background) is used to detect the ions. An actual proton-produced pattern is shown in red. (Reproduced from Umstadter D (2001) Review of physics and applications of relativistic plasmas driver by ultra-intense lasers. *Physics Plasmas* 8: 1774, with permission from the American Institute of Physics).

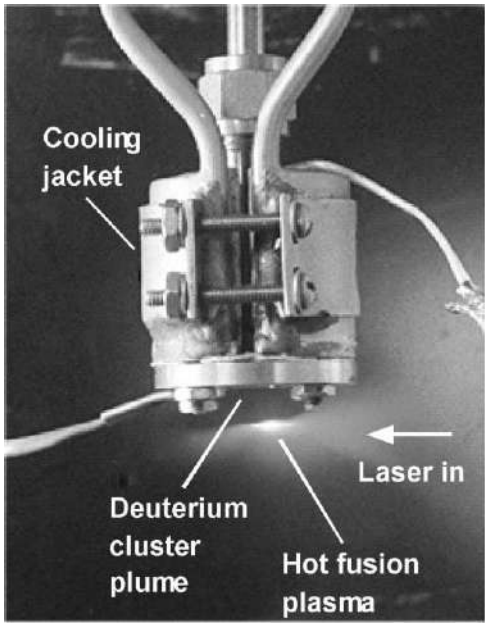

Figure 18 Photograph of a laser interacting with a deuterium cluster jet emanating from cooled gas nozzle. Fusion neutrons are produced with relatively low laser intensities. (Reproduced with permission from Ditmire T (2002) Laser fusion on a tabletop. *Optics and Photonics News* 13: 28, with permission from IEEE.)

have been produced by the He fusion reaction in the focus of 200 mJ, 160 fs Ti:sapphire laser pulses on a deuterated polyethylene target. Optimizing the fast electron and ion generation by applying a well-defined prepulse led to an average rate of 140 neutrons per shot. Neutrons have also been generated from cluster plasmas, which were produced by the cooled-nozzle depicted in Figure 18, but with significantly lower laser intensities than required with planar solid targets.

Concluding Remarks

The field of nonlinear optics, with electrons bound to atoms, has over the last forty years given rise to many scientific discoveries and technologies that are now commonplace. The relatively young field of relativistic nonlinear optics has already begun to do the same. Some of the exciting phenomena and applications that have already been identified have been discussed in this article. Others await discovery as ever higher laser intensities are reached, with no foreseeable limit.

See also

Fiber and Guided Wave Optics: Nonlinear Optics. **Scattering:** Raman Scattering.

Further Reading

Chen S-Y, Maksimchuk A and Umstadter D (1998) Experimental observation of relativistic nonlinear Thomson scattering. *Nature* 396: 653.

Cowan TE, Hunt AW, Phillips TW, Wilks SC, Perry MD, Brown C, Fountain W, Hatchett S, Johnson J, Key MH, Parnell T, Pennington DM, Snavely RA and Takahashi Y (2000) Photonuclear fission from high energy electrons from ultraintense laser–solid interactions. *Physics Review Letters* 84: 903.

Ditmire T (2002) Laser fusion on a tabletop. *Optics and Photonics News* 13: 28.

Eberly JH (1986) Fundamentals of laser interactions, vol. 229 of lecture notes in physics. *Journal Optical Society of America B* 3: 1324.

Esarey E, Sprangle P, Krall J and Ting A (1996) Overview of plasma-based accelerator concepts. *IEEE Transactions on Plasma Science* 24: 252.

Gibbon P and Förster E (1996) Short-pulse laser–plasma interactions. *Plasma Phys. Controlled Fusion* 38: 769.

Hartemann F (2001) *High-Field Electrodynamics*. Boca Raton: CRC Press.

Joshi CJ and Corkum PB (1995) Interactions of ultra-intense laser light with matter. *Physics Today* 48: 36.

Lau Y, He F and Umstadter D (2003) Nonlinear Thomson scattering: A tutorial. *Physics Plasmas* 10: 2155.

Ledingham KWD, Spencer I, McCanny T, Singhal RP, Santala MI, Clark E, Watts I, Beg FN, Zepf M, Krushelnick K, Tatarakis M, Dangor AE, Norreys PA, Allott R, Neely D, Clark RJ, Machacock AC, Wark JS, Cresswell AJ, Sanderson DCW and Magill J (2000) Photonuclear physics when a multiterawatt laser pulse interacts with solid targets. *Physics Review Letters* 84: 899.

Luther-Davies B, Gamaly EG, Wang Y, Rode AV and Tikhonchuk V (1992) Matter in ultrastrong laser fields. *Soviet Journal Quantum Electronics* 22: 289.

Marquès JR, Dorchies F, Audebert P, Ceindre JP, Amiranoff F, Gauthier JC, Hammoniaux G, Antonetti A, Chessa P, Mora PTM and Antonson J (1997) Frequency increase and damping of nonlinear electron plasma oscillations in cylindrical symmetry. *Physics Review Letters* 78: 3463.

Meyer-ter-Vehn J, Pukhov A and Sheng Z-M (2001) Relativistic laser plasma interaction. In: Batani D (ed.) *Atoms, Solids and Plasmas in Super-Intense Laser Fields*, pp. 167. New York: Kluwer Academic–Plenum.

Moore CI, Knauer JP and Meyerhofer DD (1995) Observation of the transition from Thomson to Compton scattering in multiphoton interactions with low-energy electrons. *Physics Review Letters* 74: 2439.

Mourou G and Umstadter D (1992) Development and applications of compact high-intensity lasers. *Physics Fluids B* 4: 2315.

Mourou GA and Umstadter D (2002) Extreme light. *Scientific America* 81.

Mourou GA, Barty CPJ and Perry MD (1998) Ultrahigh-intensity lasers: Physics of the extreme on a tabletop. *Physics Today* 22.

Siders CW, Blanc SPL, Fisher D, Tajima T, Downer MC, Babine A, Stepanov A and Sergeev A (1996) Laser wakefield excitation and measurement by femtosecond longitudinal interferometry. *Physics Review Letters* 76: 3570.

Sprangle P, Peñano JR, Hafizi B, Hubbard RF, Ting A, Gordon DF, Zigler A and Antonsen TM (2002) GeV acceleration in tapered plasma channels. *Physics Plasmas* 9: 2364.

Tajima T and Mourou G (2002) Zettawatt-exawatt lasers and their applications in ultrastrong-field physics. *Physics Review STAB* 5: 031301.

Umstadter D (2001) Review of physics and applications of relativistic plasmas driver by ultra-intense lasers. *Physics Plasmas* 8: 1774.

Umstadter D (2003) Relativistic laser–plasma interactions. *Journal Physics D: Applied Physics* 36: R151–R165.

Umstadter D and Norris TB (eds) (1997) Nonlinear optics with relativistic electrons. *IEEE Journal Quantum Electronics* 33: 1878.

Wagner R (1998) Laser–plasma electron accelerators and nonlinear relativistic optics. PhD thesis, University of Michigan.

Wagner R, Chen S-Y, Maksimchuk A and Umstadter D (1997) Electron acceleration by a laser wakefield in a relativistically self-guided channel. *Physics Review Letters* 78: 3125.

Yamakawa K (2004) Table-top lasers create ultrahigh peak powers. *OYOBUTURI* 73: 186–193.

S

SCATTERING

Contents

Raman Scattering
Scattering from Surfaces and Thin Films
Scattering Phenomena in Optical Fibers
Scattering Theory
Stimulated Scattering

Raman Scattering

F Kannari, Keio University, Yokohama, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

When an electromagnetic wave passes over a small elastically bound charged particle, the particle will be set into motion by the electric field. Scattering takes place for frequencies not corresponding to the natural frequencies of the particles. If the particle is bound by a force obeying Hooke's law, this vibration will have the same frequency and spatial direction as that of the electric force in the electromagnetic wave. The phase, on the other hand, of the forced vibration will differ from that of the incident wave, which causes the basis of dispersion. If the charged oscillation is bound by a force which does not obey Hooke's law, but some more complicated law, it will reradiate not only the impressed frequency, but also various combinations of this frequency with the fundamental and overtone frequencies of the oscillator.

The Raman effect was discovered by C.V. Raman (1888–1971). Ordinary Raman spectroscopy is used as a tool for studying the vibrational energy levels of molecules and of lattice optical branch vibration in crystals. A spectroscopic measurement of the scattered light reveals the existence of frequencies shifted down by increments corresponding to vibrational frequencies of the material. This down-shifted scattering is referred to as Stokes scattering. Frequencies shifted up by the vibrational frequencies are also present in the scattered radiation, which are the so-called anti-Stokes scattering.

Stimulated amplification of scattered Stokes radiation occurs through a third-order nonlinear process. To introduce stimulated Raman scattering (SRS), first we describe the steady-state stimulated scattering theory.

In general, nonlinear optical processes are described by the classical Maxwell equation with the nonlinear polarization of a material $\mathbf{P}^{\mathrm{NL}}$:

$$\nabla \times \nabla \times \mathbf{E} + \varepsilon\mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = -\mu_0 \frac{\partial^2 \mathbf{P}^{\mathrm{NL}}}{\partial t^2} \qquad [1]$$

where ε and μ_0 are the linear dielectric constant of the material and the magnetic permeability of the vacuum. We consider a relatively intense excitation radiation $\mathbf{E}_{\mathrm{L}}$ (frequency ω_{L}) and a weak signal radiation $\mathbf{E}_{\mathrm{S}}$, having the carrier frequency ω_{s} corresponding to the scattered light frequency ($\omega_{\mathrm{s}} = \omega_{\mathrm{L}} - n\omega_0$);

$$\mathbf{E}_{\mathrm{L}} = \frac{1}{2}\{\bar{\mathbf{E}}_{\mathrm{L}}(\mathbf{r}, t) \exp \mathrm{i}(\mathbf{k}_{\mathrm{L}} \times \mathbf{r} - \omega_{\mathrm{L}} t) + \mathrm{c.c.}\} \qquad [2]$$

$$\mathbf{E}_{\mathrm{S}} = \frac{1}{2}\{\bar{\mathbf{E}}_{\mathrm{S}}(\mathbf{r}, t) \exp \mathrm{i}(\mathbf{k}_{\mathrm{S}} \times \mathbf{r} - \omega_{\mathrm{S}} t) + \mathrm{c.c.}\} \qquad [3]$$

where $\bar{\mathbf{E}}_L(\mathbf{r},t)$, $\bar{\mathbf{E}}_S(\mathbf{r},t)$ are a slowly varying envelope of the complex electric fields. Assuming that the variation of the electric field during the time-scale shorter than the decay time of the induced oscillation in the material is negligible, nonlinear polarization induced by the excitation radiation is described as follows:

$$\mathbf{P}_S^{NL} = \frac{1}{2}\{\bar{\mathbf{P}}_S^{NL}(\mathbf{r},t)\exp i(\mathbf{k}_S \times \mathbf{r} - \omega_S t) + \text{c.c.}\} \quad [4]$$

$$\bar{\mathbf{P}}_S^{NL} = \varepsilon_0 \chi^{NL}(-\omega_S, -\omega_L, \omega_L, \omega_S)\bar{\mathbf{E}}_L^* \bar{\mathbf{E}}_L \bar{\mathbf{E}}_S \quad [5]$$

where ε_0 is the vacuum dielectric constant and $\chi^{NL}(-\omega_S, -\omega_L, \omega_L, \omega_S)$ is the third-order nonlinear susceptibility, of which the real part corresponds to the variation of the dielectric constant depending on the optical radiation, whereas the imaginary part describes the stimulated scattering. In steady-state approximation, substituting eqns [3]–[5] into [1], we obtain

$$\frac{d\bar{\mathbf{E}}_S}{dz} = \frac{1}{2}\left(-a + i\frac{\omega_S}{cn_S}\chi^{NL}|\bar{\mathbf{E}}_L|^2\right)\bar{\mathbf{E}}_S \quad [6]$$

where we ignore the higher-order differential of amplitude. n_S is the refractive index at ω_S, and a is the linear absorption coefficient of the signal radiation. When ignoring the decay of the excitation radiation along the z direction in the material, for making the right-hand term of eqn [6] positive, the next condition must be satisfied:

$$I_L = \frac{1}{2}c\varepsilon_0 n_L|\bar{\mathbf{E}}_L|^2 \geq I_L^{th} = \frac{c^2\varepsilon_0 n_S n_L}{2\omega_S}\frac{a}{-\text{Im}\chi^{NL}} \quad [7]$$

The signal radiation is then amplified along the z direction at an amplification constant of g,

$$G = g(I_L - I_L^{th}) \quad [8]$$

$$g = \frac{\omega_S}{c^2\varepsilon_0 n_S n_L}(-\text{Im}\,\chi^{NL}) \quad [9]$$

Stimulated Raman Scattering

In SRS, the frequency difference $\omega_L - \omega_S$ corresponds to the resonance frequency of vibrational oscillation ω_V. Therefore, the vibrational oscillation mode is excited in the material as well as the generation of scattered radiation. In the following section, we consider the SRS by optical phonon. In general, nonlinear polarization under the optical excitation is described as follows:

$$\mathbf{P}^{NL} = N_0\alpha' q\mathbf{E}, \quad [10]$$

$$\ddot{q} = \Gamma_V\dot{q} + \omega_V^2 q = \frac{1}{2}\alpha'\mathbf{E}^2 \quad [11]$$

where q is the basic coordinate of molecular vibration, Γ_V is the decay rate of the vibrational oscillation, $\alpha' = (\partial\alpha/\partial q)_0$ is the Raman polarizability of a molecule, and N_0 is the molecule density. We can obtain the steady-state nonlinear polarization from eqns [4], [5], [10], and [11] as

$$\chi^{NL} = (-\omega_S, -\omega_L, \omega_L, \omega_S) = \frac{N_0(\alpha')^2}{4\varepsilon_0}\frac{1}{\omega_V^2 - \omega^2 + i\omega\Gamma_V} \quad [12]$$

where $\omega = \omega_L - \omega_S$. Thus, the gain coefficient can be obtained from eqns [9] and [12] under the assumption of $\omega_V \gg \Gamma_V$, as follows:

$$g_R(\omega_S) = \pm\frac{N_0(\alpha')^2\omega_S}{2c^2\varepsilon_0 n_S n_L \omega_V \Gamma_V} \times \frac{(\Gamma_V/2)^2}{(\omega_L - \omega_S \mp \omega_V)^2 + (\Gamma_V/2)^2} \quad [13]$$

In general, Raman scattering is characterized by the scattering cross-section $d\sigma/d\Omega$, which is defined for one of the optical polarization directions in terms of the optical power density:

$$g_R(\omega_L - \omega_V) = \frac{16\pi^2 c^2 N_0}{\hbar n_S^2\omega_S^3\Gamma_V}\frac{d\sigma}{d\Omega}\left\{1 - \exp\left(-\frac{\hbar\omega_V}{k_B T}\right)\right\} \quad [14]$$

The value of $(d\sigma/d\Omega)\Gamma_V^{-1}$ can be characterized as a peak intensity in the measured spontaneous Raman scattering spectrum.

When the conversion to the first Stokes wave is large, we must consider both pump depletion and conversion to higher-order Stokes waves (Figure 1a). Pump depletion in first Stokes generation can be accounted for by solving the following coupling

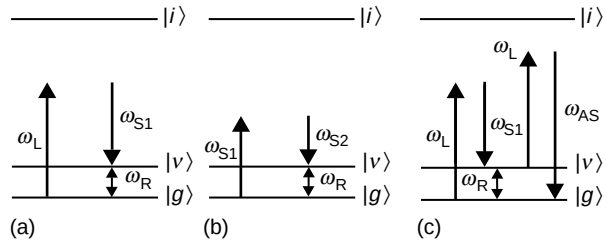

Figure 1 Energy diagrams of stimulated Raman scattering: (a) first Stokes scattering; (b) cascaded second Stokes scattering; and (c) first anti-Stokes scattering.

equations between pump power and Stokes power:

$$\frac{\partial I_L}{\partial z} + \frac{n_L}{c}\frac{\partial I_L}{\partial t} = -g_L I_L I_S \qquad [15]$$

$$\frac{\partial I_S}{\partial z} + \frac{n_S}{c}\frac{\partial I_S}{\partial t} = g_R I_L I_S \qquad [16]$$

where

$$g_L = \frac{\omega_L}{\omega_S} g_R$$

In SRS the power generated in the first Stokes serves as a pump for a second Stokes wave (Figure 1b). Cascade shifts of this type, in which the Stokes wave of order n serves as a pump for the generation of the Stokes wave order $n+1$, can result in the generation of radiation at several frequencies, each one shifted from the original pump frequency by a multiple of the vibrational mode frequency. In addition to the cascade higher-order Stokes generation, multiple Stokes orders can also be generated by four-wave mixing interactions (Figure 2), of the form:

$$\omega_{S,n+1} = 2\omega_{S,n} - \omega_{S,n-1} \qquad [17]$$

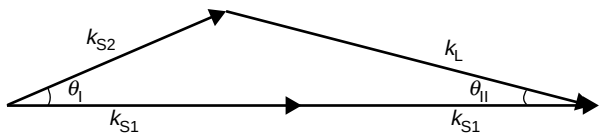

Figure 2 Vector diagram of multiple Stokes generation through four-wave mixing in positively dispersive media.

These interactions are important when the wavevector mismatch among the various waves is small. In most media exhibiting positive dispersion, the phase matching of the various interactions involves off-axis components of the generated waves. Thus, these four-wave mixing interactions are stronger for tightly focused beams than for collimated pump beams.

SRS can be realized in a variety of configurations, including single pass Raman generation, or using a variety of extra-cavity or intra-cavity Raman resonator. Variation of the Raman cross-section for the wavelength is shown in Figure 3 for various gaseous Raman media. The Raman gain coefficient simply increases with decreasing wavelength, as indicted by eqn [13]. Therefore, higher conversion efficiencies are obtainable with UV lasers such as rare-gas halide excimer lasers in a single pass configuration. Note that in some gases, such as H_2, the Raman gain becomes independent of pressure at higher pressures since the increase in linewidth, due to pressure broadening, cancels the increase in gain originated from number density. Typical SRS gains for some gas, liquid, and solid media are listed in Table 1.

Although SRS in high-pressure gases has been widely employed for frequency conversion, SRS in crystals is currently a growing area of research activity. Table 2 lists SRS frequency shifts of various crystals. Since a very high pump laser power ~1 GW/cm^2, which is already close to the laser damage threshold of many crystals, solid-state Raman lasers are

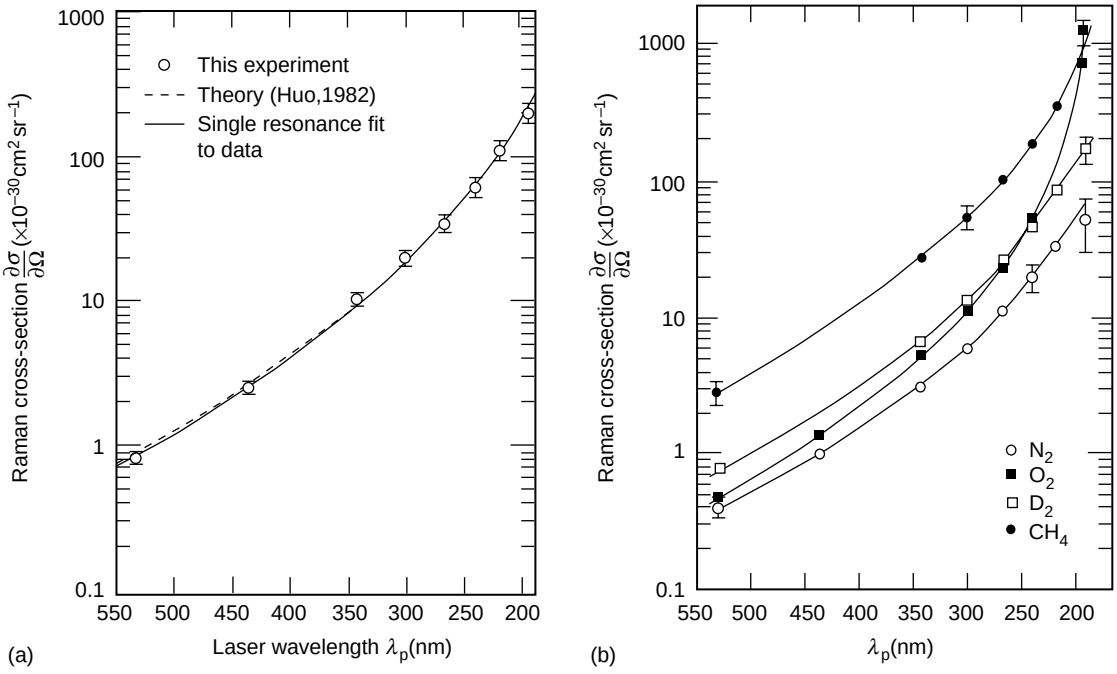

Figure 3 Variation of the Raman cross-section with pump wavelength for various gases. Source: AIP Conference Proceedings No. 100 (Excimer Laser – 1983), 181.

Table 1 Typical Raman media and their parameters

Materials	*Stokes shift (cm^{-1})*	*Raman linewidth (cm^{-1})*	$N\frac{d\sigma}{d\Omega}$ *($10^{-6}\ m^{-1}\ str^{-1}$)*	*Raman gain (10^{-5} m/MW)*
H_2 (100 atm)	4155	0.20	3.0×10^3	1.5
H_2O	3420	176	0.51	0.07
Acetone CH_3COCH_3	2921	18	5.4	1.2
Cyclohexane C_6H_{12}	2852	11	3.7	1.2
Liquid N_2	2326.5	0.067	0.29 ± 0.09	17.5 ± 5
Liquid O_2	1552	0.117	0.48 ± 0.14	14.5 ± 4
Nitrobenzene $C_6H_5NO_2$	1345	6.6	6.4	2.1
Toluene C_7H_8	1003	1.94	1.1	1.2
C_6H_5Cl	1002	1.6	1.5	1.9
C_6H5Br	1000	1.9	1.5	1.5
C_6H_6	992	2.15	3.06	2.8
CS_2	655.6	0.5	7.55	23.8
CC_{l4}	459	5	2.3	0.85
Diamond	1332	2.04	17	6.9
Calcite	1086	1.1	2.9	4.4
Si	521	0.8	305	190
Quartz	467	6.7	3.1	0.8
Li^7TaO_3	201	22	238	4.4
	215	12	167	10
Li^7NbO_3	256	23	381	8.9
	258	7	262	28.7
	637	20	231	9.4
	643	16	231	12.6
InSb ($n_e \sim 2 \times 10^{22}\ m^{-3}$)	100	2	30	16.7

designed with an intra-cavity, coupled cavities, or an extra-cavity. Efficient Raman lasers have been demonstrated using external resonator configuration when the pumping pulse is intense enough. For example, ~40% conversion efficiency to S1 was reported with a Q-switched frequency doubled Nd:YAG laser in a $Ba(NO_3)_2$ crystal using a pump fluence of 1.4 J cm^{-1} and a pulse repetition rate of 30 Hz. The resonator mirror characteristics can be designed to enhance the desired Stokes-order radiation by completely confining the lower-order Stokes radiation inside the cavity. For a lower power pump source, such as cw-pumped Q-switch lasers, a Raman crystal is placed inside a pump laser resonator to reduce the effective threshold for SRS. The resonator configuration has a significant influence on the Stokes output performance. The optimum spot size in the resonator and power density in the Raman crystal is a compromise between maximizing the conversion efficiency and avoiding optical damage. It is noteworthy that the thermal lens in the Raman crystal is caused by the energy dissipation of the phonon and thus depends on the power density of the Stokes field.

A fiber Raman laser is an optical fiber with pump light focused into one end and SRS light appearing at the output. By injecting a weak frequency-shifted signal light together with the pump light, the signal light experienced substantial amplification along the length of the fiber. This fiber Raman amplifier has been actively developed to extend the wavelength band for optical fiber communication. The fiber Raman lasers have utilized fused silica (SiO_2), germanosilicate, or phosphosilicate optical fibers for which the Raman frequency shifts are ~440–490 cm^{-1}. In contrast to Erbium-doped fiber amplifier (EDFA), since the single mode fiber itself can act as an amplifier, the Raman amplifier can be distributed over optical transmission lines in lengths of >20 km, which can keep the transmission power always nearly constant and consequently reduce the signal noise. Moreover, by preparing multipumping wavelengths, the number of signal wavelengths can increase without the limitation of an amplification gain bandwidth. High-power light sources around 1480 nm developed for EDFA can be utilized by the fiber Raman amplifier. By tailoring the power levels among multiwavelength pump lights, a flat amplification gain spectral width of ~100 nm is obtainable.

Transient Effects

In the steady-state SRS analysis, we assume that the pump pulse duration is much longer than the vibrational mode dephasing time ($\Delta\nu_p \ll \Gamma_V$). In a transient regime, when the pump pulse duration is shorter than the dephasing time ($\Delta\nu_p \gg \Gamma_V$), the effect of the transient response of the medium reduces the gain and delays the Stokes pulse relative to the pump pulse. With no pump depletion, the analytical expression for the Stokes intensity for large

Table 2 Typical solid-state Raman materials and parameters

Materials	*Stokes shift (cm^{-1})*	*Raman linewidth (cm^{-1})*	$\frac{d\sigma}{d\Omega}$ *(arb.u.)*
Diamond	1332.9	2.7	100
SO_2	464.5	7.0	2.2
Nitrates and calcite			
$Ba(NO_3)_2$	1048.6	0.4	21
$NaNO_3$	1069.2	1.0	23
$CaCO_3$	1086.4	1.2	6.0
Tungstates			
$CaWO_4$	910.7	4.8	47
$SrWO_4$	921.5	3	–
$BaWO_4$	926.5	2.2	–
$NaY(WO_4)_2$	918	15	–
$KGd(WO_4)_2$	901	5.4	54
$KGd(WO_4)_2$	901	5.4	43
$KGd(WO_4)_2$	768	6.4	19
$KGd(WO_4)_2$	768	6.4	65
$KY(WO_4)_2$	905.6	7	50
$KY(WO_4)_2$	905.6	7	45
$KY(WO_4)_2$	767.4	8.4	20
$KY(WO_4)_2$	767.4	8.4	64
$KYb(WO_4)_2$	908	7.4	48
$KYb(WO_4)_2$	908	7.4	48
$KYb(WO_4)_2$	757	15	25
$KYb(WO_4)_2$	757	15	70
Molybdates			
$CaMoO_4$	879.3	5.0	64
$SrMoO_4$	887.7	2.8	55
$BaMoO_4$	892.4	2.1	52
Iodate and niobates			
$LiIO_3$	821.6	5.0	54
$LiNO_3$	872	21.4	44
$LiNbO_3$	632	27	166
$LiNbO_3$	250	28	–
$LaNbO_4$	805	9	22
Phosphates			
$Ca_2(PO_4)_3F$	964.7	2.8	3.4
$Sr_5(PO_4)_3F$	950.3	2.8	3.4

amplification gain is given as

$$I_S(\ell) \propto I_S(0)\exp\left(-\frac{\tau_p}{T_V}\right) \times \exp\left\{2\left[I_L\tau_p\ell\frac{16\pi^2c^2N_0}{\hbar n_S^2\omega_S^3}\left(\frac{d\sigma}{d\Omega}\right)\right]^{1/2}\right\} \quad [18]$$

Here, T_V is the dephasing time ($T_V = \pi\Gamma_V^{-1}$), and τ_p is the pump pulse width. One can see that the Raman gain is proportional to the square root of the multiplication of the pump pulse energy $I_L \times \tau_p$, the crystal length ℓ, and the total integral Raman scattering cross-section ($d\sigma/d\Omega$), but does not depend on the Raman linewidth Γ_V.

Anti-Stokes Raman Scattering

From eqn [13] the gain coefficient is negative at $\omega = \omega_L + \omega_S$. However, anti-Stokes scattering (Figure 1c) is stimulated through a four-wave mixing interaction of the form:

$$\omega_A = 2\omega_L - \omega_S \quad [19]$$

The equations describing anti-Stokes Raman scattering in the absence of pump depletion are:

$$\frac{d\bar{E}_S}{dz} + \frac{a}{2}\bar{E}_S = i\frac{\omega_S^2}{2c^2\varepsilon_0 k_{Sz}}\bar{P}_S^{NL} \quad [20]$$

$$\frac{dE_A^*}{dz} + \frac{a}{2}\bar{E}_A^* = -i\frac{\omega_A^2}{2c^2\varepsilon_0 k_{Az}}\bar{P}_A^{NL*} \quad [21]$$

$$\bar{P}_S^{NL} = \varepsilon_0\chi^{NL}\left\{|\bar{E}_L|^2\bar{E}_S + \bar{E}_L^2\bar{E}_A^*\exp(i\Delta kz)\right\} \quad [22]$$

$$\bar{P}_A^{NL*} = \varepsilon_0\chi^{NL}\left\{\bar{E}_L^{*2}\bar{E}_S\exp(-i\Delta kz) + |\bar{E}_L|^2\bar{E}_A^*\right\} \quad [23]$$

Here, Δk is the wavevector mismatch, given by

$$\Delta k = k_S + k_A - 2k_L \quad [24]$$

Then, power gain per unit length is obtained to give

$$G \cong -a \pm 2\,\mathrm{Im}\left\{\left(\frac{\Delta k}{2}\right)^2 - \frac{\omega_S}{2cn_S}\chi^{NL}|\bar{E}_L|^2\Delta k\right\}^{1/2} \quad [25]$$

The interaction of the anti-Stokes wave with the other waves is stronger when the wavevector mismatch is small. The exact value of the phase-matching angle depends on the dispersion of the medium. At the tightly focused pumping condition, strong anti-Stokes radiation is generated so that the wavevector mismatch is minimum. Just as in Stokes scattering, multiple anti-Stokes orders can be generated (Figure 4).

So far, Stokes and anti-Stokes Raman scattering have been used to generate radiation over much of the UV and VUV, ranging from 138 to 400 nm. The shortest wavelength achieved in this manner is at 138 nm, generated by a seventh-order anti-Stokes shift in H_2 using 139 nm radiation of an ArF laser. In general, the pump intensity used in experiments with rare-gas halide excimer lasers was in the range of 100 MW/cm^2 to 5 GW/cm^2. The conversion efficiency and the distribution of power among the various Stokes and anti-Stokes

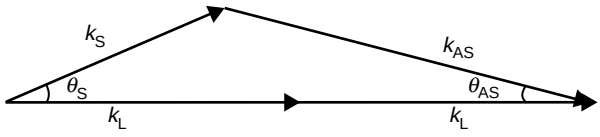

Figure 4 Vector diagram of anti-Stokes Raman scattering in positively dispersive media.

components depend on the focal geometry, and the extent of the spectrum could be varied by changing the gas pressure.

See also

Magneto-Optics: Cyclotron Resonance, Interband Magnetoabsorption, Spin Flip Raman Scattering. **Nonlinear Optics, Applications:** Raman Lasers.

Further Reading

Basiev TT, Sobol AA, Zverey PG, Ivleva LI, Osiko VV and Powell RC (1999) Raman spectroscopy of crystals for stimulated Raman scattering. *Optical Matters* 11: 307.

Bischel WK and Blach G (1983) New lines in the UV: SRS of excimer laser wavelengths. *AIP Conference Proceedings No. 100 (Excimer Lasers-1983)*, 181.

Loree TR, Sze RC, Barker DL and Scott PB (1979) *IEEE Journal of Quantum Electronics* QE-15: 337.

Reinties JF (1985) Coherent ultraviolet and vacuum ultraviolet sources. In: Laser Bass M and Stitch ML (eds) *Laser Handbook*. Amsterdam: North-Holland Physics Publishing.

Weber MJ (ed.) (2001) Raman, Brillouin and Soliton lasers. In: *Handbook of Laser*, p. 291. Boca Raton, FL: CRC Press.

Scattering from Surfaces and Thin Films

A Duparré, Fraunhofer Institut IOF, Jena, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The quality of optical components can be critically limited by light scattering. The majority of these components, such as lenses, mirrors, beamsplitters, and polarizers are covered with thin film optical coatings. To reduce scattering losses, their main sources and mechanisms must be known and, whenever possible, be controlled. Moreover, light scattering techniques have been widely recognized as a powerful tool for roughness and defect analysis as well as for general quality assessment.

Scattering can be divided into surface/interface scattering (SIS) and volume scattering (VS), attributed to the following scatter sources:

- roughness and surface defects of the substrate: SIS,
- surface/interface roughness and defects of the thin film coating: SIS,
- imperfections in the bulk substrate material: VS,
- intrinsic thin film morphology and defect structures: VS,
- surface contaminations: SIS.

Even though in particular cases and applications, VS can constitute the main scattering source, SIS largely dominates both substrate and thin film scattering losses. Studies into light scattering have widely focused on SIS effects from surface and interface micro-roughness. The theoretical outline given in the next section will also follow this concept in describing scattering from surface and interface roughness. It is emphasized, however, that volume scattering theories are accessible in the literature.

Light Scattering Models

Scattering from Rough Surfaces

A randomly rough surface can be considered as a Fourier series of sinusoidal waves with different amplitudes, periods, and phases. Following the grating equation, a single grating with spacing D causes scatter into the angle Θ according to

$$\sin\Theta = \lambda/D \qquad [1]$$

where λ is the wavelength of light. D represents one spatial wavelength in the Fourier series. Accordingly, $f = 1/D$ represents one single spatial frequency. A randomly rough surface contains many different spatial frequencies. This is quantitatively expressed by the power spectral density (PSD), giving the relative strength of each roughness component of a surface microstructure as a function of spatial frequency:

$$\mathrm{PSD}(\mathbf{f}) = \lim_{A\to\infty}\frac{1}{A}\left|\int_{\mathbf{A}}\zeta(\mathbf{r})\exp(-2\pi i\mathbf{f}\cdot\mathbf{r})\mathrm{d}\mathbf{r}\right|^2 \qquad [2]$$

where $\zeta(\mathbf{r})$ represents the height of the surface roughness profile, $\mathbf{r}$ is the position vector, and $\mathbf{f}$ is the spatial frequency vector in the x–y plane. A is the area of the measured region $\mathbf{A}$. We confine our discussion to isotropic surfaces, which represent the majority of cases in optical surface and thin film scattering studies. Thus, a PSD(f) independent of the surface direction Φ of vector $\mathbf{f}$ is obtained by averaging the two-dimensional function PSD($\mathbf{f}$) over all surface directions after transformation into polar coordinates:

$$\mathrm{PSD}(f) = \frac{1}{2\pi}\int_0^{2\pi}\mathrm{PSD}(f,\Phi)\mathrm{d}\Phi \qquad [3]$$

Well-established vector scattering theories developed, for instance, by Bousquet *et al.* or Elson provide the link between the PSD and the scattering intensity per solid angle of a surface:

$$\frac{\mathrm{d}P}{P_0\,\mathrm{d}\Omega} = F(\lambda, n, \Theta, \Phi)\mathrm{PSD}(f) \qquad [4]$$

This theory is valid for surfaces whose rms roughness is small compared to the wavelength. $\mathrm{d}P/(P_0\,\mathrm{d}\Omega)$ denotes the differential power scattered into the direction (Θ, Φ) per unit solid angle $\mathrm{d}\Omega = \sin\Theta\,\mathrm{d}\Theta\,\mathrm{d}\Phi$ divided by the incident power P_0. Θ and Φ are the polar and azimuthal angles of scattering, respectively. The optical factor F contains all information on the corresponding perfect surface (without the roughness properties), i.e., the refractive index n, wavelength, and the conditions of illumination and observation. Both backscattering and forward scattering can be expressed by eqn [4], according to the illumination and observation conditions chosen. Without loss of generality, all formulas have been written here for normal incidence. The formalism, however, allows consideration of all possible cases, including oblique angles of incidence and arbitrary polarization properties. $\mathrm{d}P/(P_0\,\mathrm{d}\Omega)$ is called angle resolved scattering (ARS) which is related to the well-established term BRDF/BTDF (bidirectional reflectance/transmittance distribution function) by multiplication with $\cos\Theta$:

$$\frac{\mathrm{d}P}{P_0\,\mathrm{d}\Omega} = \mathrm{ARS} = \mathrm{BRDF}\ (\text{or BTDF})\cdot\cos\Theta \qquad [5]$$

Total scattering (TS), which is defined as the power P scattered into the backward or forward hemisphere divided by the incident power P_0 (see also section on scattering measurement below), is obtained by integrating eqn [4] over the forward or backward hemisphere:

$$\mathrm{TS} = 2\pi\int_0^{\frac{\pi}{2}}\left(\frac{\mathrm{d}P}{P_0\,\mathrm{d}\Omega}\right)\sin\Theta\,\mathrm{d}\Theta \qquad [6]$$

If the correlation length of surface roughness is much larger than the wavelength, scalar scattering theories like the one from Carniglia can also be employed. Moreover, in this case, the well-known simple approximate formula for total backscattering can be derived from both vector and scalar theories:

$$\mathrm{TS}_{\mathrm{back}} = R_0\left(\frac{4\pi\sigma}{\lambda}\right)^2 \qquad [7]$$

where σ is the rms roughness and R_0 the specular reflectance.

It must be emphasized that this formula is only valid if the above-mentioned condition is met and as long as only single surfaces without coatings are considered.

Scattering from Dielectric Thin Films

For surfaces coated with a dielectric single layer or multilayer, ARS for a system of N layers is given by

$$\frac{\mathrm{d}P}{P_0\,\mathrm{d}\Omega} = \sum_{i=0}^{N}\sum_{j=0}^{N} F_i F_j^* \mathrm{PSD}_{ij}(f) \qquad [8]$$

where F_i is the optical factor at the i-th interface and F_j^* is the conjugate complex number of the optical factor at the j-th interface. The optical factors include both the conditions of illumination and observation and the characteristics of the ideal multilayer (refractive indices, film thickness). PSD_{ij} are the power spectral densities of the corresponding interfaces. For $i \neq j$, they describe the roughness cross-correlation between two interfaces.

From eqn [8] it becomes obvious that the scattering mechanism of coated surfaces is considerably more complex as compared to single surfaces.

The amplitudes of the scattered fields from all individual interfaces add up to the total scatter. In particular, the cross-correlation properties between the interface roughness profiles significantly determine the overall total scatter. So even for coatings having the same statistical roughness, the scatter losses can differ drastically when the cross-correlation between the interfaces is different. An example is demonstrated in a model calculation of total backscattering for $\lambda = 632.8$ nm in Figure 1.

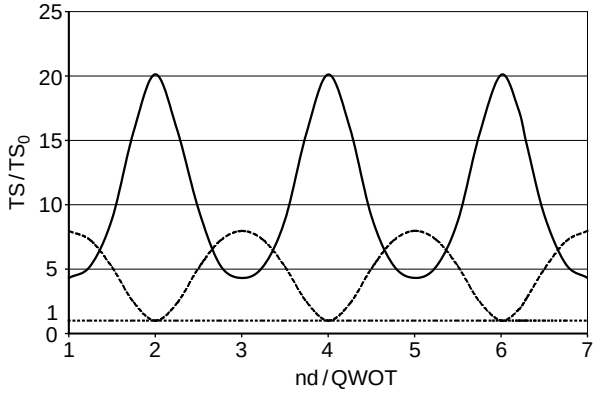

Figure 1 Calculated ratio of the total backscattering of a single high index layer ($n = 2.4$) on BK7 substrate ($n = 1.52$) to the scattering of the bare substrate as a function of optical thickness nd. (QWOT: quarter wave optical thickness). rms roughness was 1 nm, correlation length 1 μm for both the substrate–film and film–air interfaces. Full line: full interface cross-correlation. Dashed line: zero interface cross-correlation. Lower dashed line: bare substrate.

A thin film with refractive index $n = 2.4$ (typical for Titania films), and continuously increasing thickness, was modeled onto a BK 7 substrate ($n = 1.52$). Gaussian autocovariance functions with an rms roughness $\sigma = 1$ nm and a correlation length $\tau = 1$ μm were chosen for both the substrate–film and film–air interfaces. Only the cross-correlation function was varied. The two extreme cases of fully uncorrelated and fully correlated interfaces are depicted in the figure and related to the scattering of the bare substrate (TS_0). As a result of scattering interference effects, in both cases the scattering varies periodically with optical film thickness. The appearance of maxima and minima as well as the total amount of scattering crucially depends on the type of cross-correlation.

These effects have to be taken into careful deliberation whenever scattering in dielectric thin film coatings is considered. Otherwise, misinterpretation of the scatter loss origin is possible and may result in unsuitable technological attempts to minimize scatter losses.

Instrumentation for Light Scattering Measurement

Types of Measurements

Light scattering experiments on optical surfaces and thin films are, in most cases, either performed as ARS measurements using goniophotometers, or the scattering is collected over the backward and forward hemispheres yielding total integrated scattering (TIS) or TS. The latter can be measured either by using a Coblentz sphere or an integrating sphere (see schematic picture in Figure 2).

As Coblentz spheres image the scattered radiation directly onto the detector, they are much less sensitive to scattering by air particles than integrating spheres. So if super-smooth samples with scattering levels in the visible range of 1 ppm and lower shall be measured, integrating spheres require operation in vacuum or He-gas atmosphere whereas Coblentz spheres can still be used under usual clean room conditions.

Angle Resolved Scattering (ARS)

Sophisticated instruments for ARS measurements have been established in a number of laboratories. Most frequently the systems are operated at the He–Ne laser wavelength 632.8 nm, but also at 10.6 μm (CO_2 laser), 325 nm (He–Cd laser), and yet other wavelengths. These techniques are versatile and powerful, but rather unsuitable for routine measurements. For single surfaces, PSD can be calculated from such measurements. A standard procedure for ARS measurements is defined in ASTM standard E 1392 and was successfully verified in various round-robin experiments at different wavelengths. Figure 3 shows the photograph of a typical instrumentation.

The main parts of this setup are the illumination system and a double goniometer. The sample holder and the detector are each mounted on a precision goniometer. The detector head is located on the outer goniometer and can be revolved 360 degrees around the sample with an angular resolution of 0.01°. The performance of a scatterometer depends also considerably on the quality of the illumination system. Laser radiation passes several optical elements, such as a spatial filter, diaphragms, polarizers, and mirrors providing the high beam quality necessary for precision measurement.

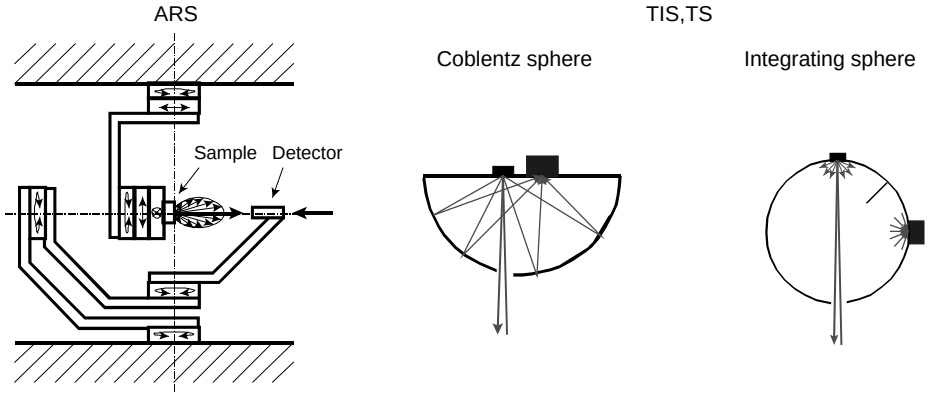

Figure 2 Schematic picture of types of scattering measurement. ARS: angle resolved scattering, TIS, TS: total integrated scattering, total scattering.

Figure 3 Instrumentation for angle resolved scattering measurement (courtesy of Fraunhofer Institute for Applied Optics and Precision Engineering, Jena, Germany).

Total Scattering (TS)

TIS measurements collect the light scattered into the backward hemisphere and are defined as the backscattered radiation divided by the total reflectance. The TIS measurement procedure is prescribed in ASTM standard F 1048 which applies to opaque reflective surfaces. Transparent or semi-transparent samples such as substrates, AR coatings, and beamsplitters cannot be measured without ambiguous additional assumption. The new international standard ISO 13696 defines TS as the backscattered and forward scattered radiation P divided by the incident radiation P_0. TS is hence equivalent to the scattering loss of the component. Opaque as well as transparent surfaces and coatings can be measured both in the backward and forward directions. TS and TIS can, however, be converted into one another, if the reflectance of the sample is known. This standard procedure was proved in an international round-robin experiment at 632.8 nm.

An example of a facility for TS measurements, in a wide range of wavelengths, is given in Figure 4. The apparatus for total backscattering and forward scattering measurements is operated from 193 nm to 10.6 μm by using several lasers. The setup is based on a Coblentz sphere imaging the light scattered into the backward or forward hemisphere within an angular range from 2° to 85° onto the detector unit. This angular range complies with the specification given in ISO 13696.

A special arrangement allows easy change from the backscatter to forward scatter operation modus. The detector unit consists of the detector (photomultiplier in the UV, VIS, NIR, and HgCdTe-element in the IR) and a small integrating sphere. The latter is used for homogeneously illuminating the detector. At all wavelengths, the light beam is modulated by a chopper and passes a beam-cleaning element. The incident radiation hits the sample surface at nearly zero degrees and the specular beam is guided back through the entrance/exit aperture of the Coblentz sphere. While performing the measurement, the sample is scanned across its surface by a positioning system, yielding one- or two-dimensional scattering diagrams. Calibration is performed with a commercial diffusing (Lambertian) standard. Filters are used for attenuating the beam during measurement of high-scatter samples and diffuse reflectance standards. Background levels smaller than 0.1 ppm at 632.8 nm were achieved without the necessity of He-gas flow or operation in vacuum.

Whereas in the past the majority of such facilities were designed for the visible, infrared, and near-ultraviolet spectral regions, the increasing demands for low-scatter optical components in the vacuum ultraviolet (VUV) region for application in photolithography has driven the development of scattering facilities for wavelengths as short as 157 nm. Because of the enhanced technical effort needed for the development of VUV scattering measurement instrumentation, only few setups exist that can be operated at 157 nm.

Applications

Fields of Applications

There exists a diversity of purposes for light-scattering investigations of surfaces and thin films, which requires proper selection of the particular method. For a number of purposes, routine measurements without detailed understanding of the scatter sources can be sufficient, while other problems need deeper insight into the scatter mechanisms and hence, require the combination of measurement with theoretical modeling. It is up to the engineer or scientist to decide in each particular case, to which level the

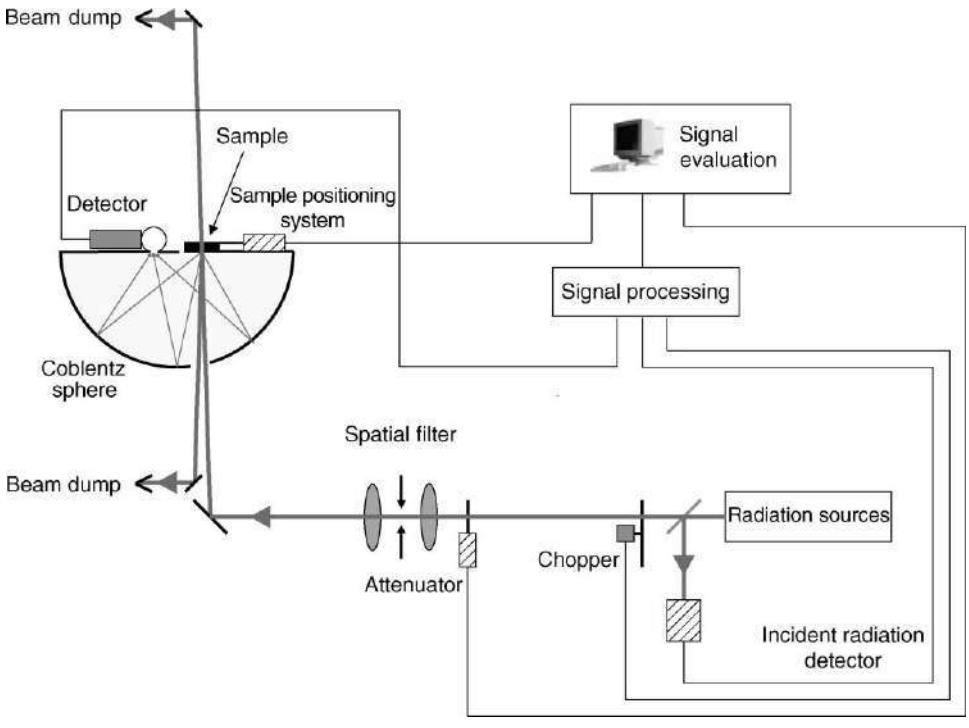

Figure 4 Schematic picture of an instrument for total scattering measurement (courtesy of Fraunhofer Institute for Applied Optics and Precision Engineering, Jena, Germany).

light-scattering study has to be extended. Examples for typical application tasks arising in industry and institutes are:

- determination of total scattering as an optical loss limiting the performance of a thin film component;
- surface finish assessment (roughness, defects) of optical substrates;
- study into the relation between scattering and the morphology and roughness of optical thin films as a precondition for optimizing deposition processes;
- assessment of cleanliness of coated components;
- detection of degradation effects in coatings and substrate materials;
- separation of interface scattering from the film and substrate;
- distinction of volume scattering from interface scattering in substrates and films.

Scattering measurements are noncontact, do not require sample preparation, enable rapid mapping of large sample areas, and are highly sensitive. This makes them usable for a large variety of applications extending from super-smooth substrates with surface roughness below 0.01 nm, thin film structures in the nanometer range, and small and large defect features up to rough engineering surfaces with micron roughness.

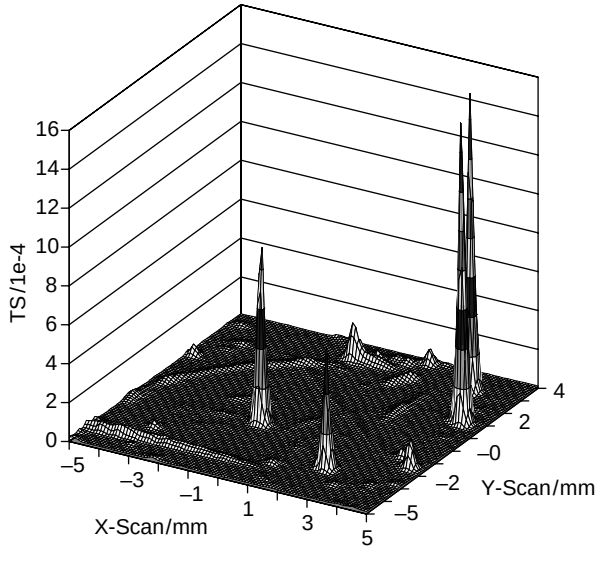

Figure 5 TS measurement (backscattering) on conventionally polished fused silica at 632.8 nm (2D-mapping) for surface quality assessment. Measurement performed with the TS instrument shown in Figure 4.

Examples of Measurements

Surface finish assessment

The result of a scattering measurement (TS, backscattering) at the He–Ne laser wavelength (632.8 nm), on conventionally polished fused silica, is shown in Figure 5. Symmetric defects, as well as

scratch-like features, can be recognized in this diagram. The minimum and averaged rms-roughness determined from the minimum (defect-free) and averaged (including defects) scattering levels are 0.51 and 0.72 nm, respectively.

Scattering from thin film components

Figure 6 shows a TS measurement (backscattering) performed on a polished CaF_2 sample. Half of the area of one surface of this sample was coated with a fluoride multilayer mirror system designed for 248 nm. The area mapping of the TS measurement at 248 nm delivers information of both the bare substrate quality and the increase in scattering after coating. According to the explanations in the first section, this increase has to be interpreted as a result of two influences: the increased reflectance (i.e., changed optical factors) and the roughness growth during deposition of the multilayer (i.e., changed PSD).

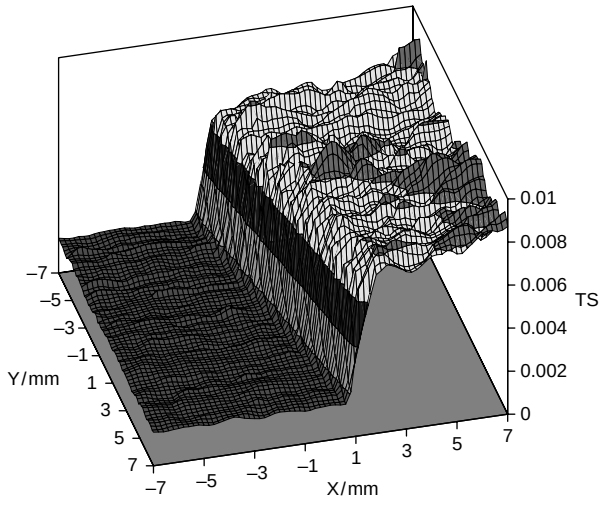

Figure 6 TS measurement (backscattering) at 248 nm on a CaF_2 substrate half of which was coated with a fluoride multilayer mirror system designed for 248 nm. Measurement performed with the TS instrument shown in Figure 4.

On the basis of such measurements, detailed interpretation can be accomplished by additional roughness analysis using scanning force microscopy, subsequent scattering modeling employing the formulas given above, and comparison of the modeled and measured results.

ARS techniques can, for example, be employed to compare the angular distribution of the scattering of a coating to that of the bare substrate. Figure 7 displays ARS measurements in the backward hemisphere at 325 nm (He–Cd laser) on a fused silica substrate before and after deposition (magnetron sputtering) of a Ta_2O_5/SiO_2 quarterwave multilayer mirror (325 nm design wavelength). As with the example discussed above, the higher scatter of the multilayer is caused by both the optical factors and increased roughness. The slight bump at large scatter angles in particular reveals high spatial frequency roughness attributed to the film morphology.

Figure 8 displays TS measurements at 157 nm. The diagram contains one-dimensional scans of forward scattering from an uncoated CaF_2 substrate and an antireflective (AR) coating on CaF_2. The background signal level is also included.

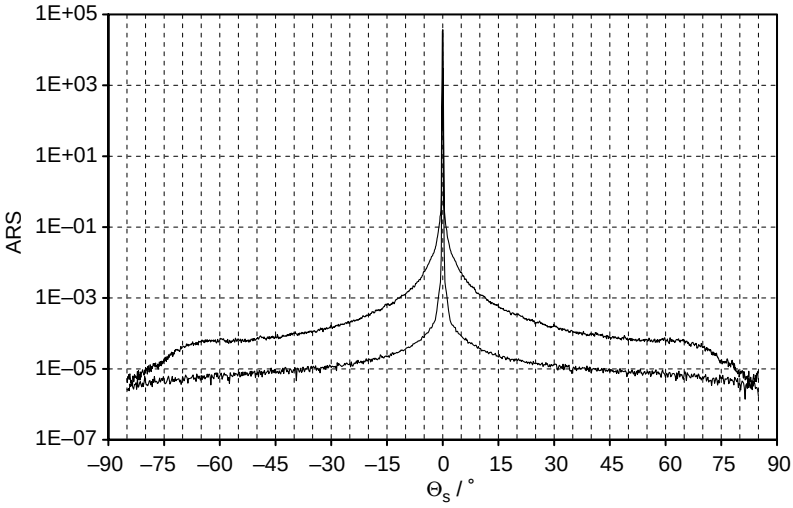

Figure 7 ARS measurement at 325 nm (backward hemisphere) on a fused silica substrate before (lower curve) and after (upper curve) deposition of a Ta_2O_5/SiO_2 quarterwave multilayer mirror with 325 nm design wavelength. Measurements performed with the ARS instrument shown in Figure 3.

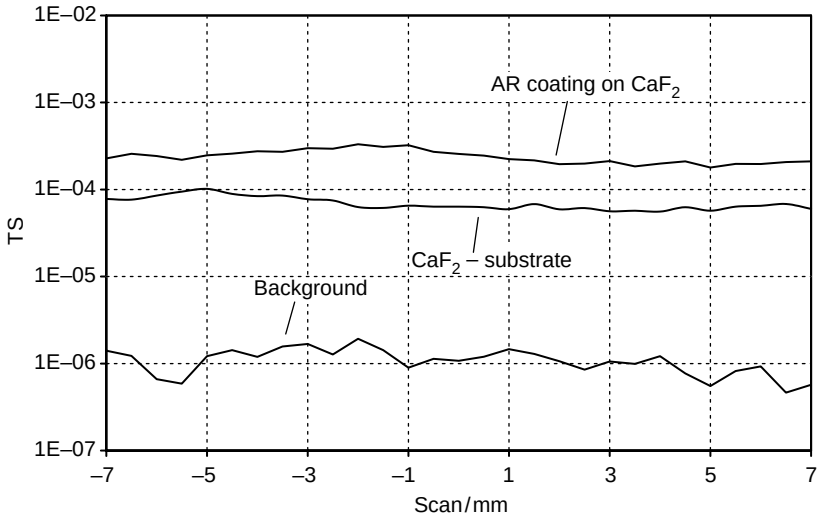

Figure 8 TS measurements at 157 nm (forward scattering): bare CaF_2 substrate, AR coating on CaF_2 and background scattering level. Measurements performed with the VUV total scattering instrumentation at the Fraunhofer Institute for Applied Optics and Precision Engineering, Jena, Germany.

List of Units and Nomenclature

Angle resolved scattering	[nondimensional]	ARS
Azimuthal scattering angle	[degree]	Φ
Correlation length	[nm]	τ
Differential scattered power	[nondimensional]	$dP/(P_0\, d\Omega)$
Grating spacing	[μm]	D
Height of surface roughness profile	[nm]	ζ
Incident power	[W]	P_0
Optical factor	[nm^{-4}]	F
Polar scattering angle	[degree]	Θ
Power spectral density	[nm^4]	PSD
Refractive index	[nondimensional]	n
Rms roughness	[nm]	σ
Spatial frequency	[μm^{-1}]	f
Specular reflectance	[nondimensional]	R_0
Total integrated scattering	[nondimensional]	TIS
Total scattering	[nondimensional]	TS
Wavelength	[nm]	λ

See also

Optical Coatings: Thin-Film Optical Coatings. **Semiconductor Physics:** Light Scattering.

Further Reading

Baeumer S, Duparré A, Herrmann T, *et al.* (1999) SLIOS – a contribution to standard procedures in stray light measurement. *Proceedings SPIE* 3739: 414–421.

Bennett JM and Mattsson L (1999) *Introduction to Surface Roughness and Scattering*, 2nd edn., pp. 28–35, 62–70. Washington, DC: Optical Society of America.

Bousquet P, Flory F and Roche P (1981) Scattering from multilayer thin films: theory and experiment. *Journal of Optical Society of America* 71: 1115–1123.

Carniglia CK (1979) Scalar scattering theory for multilayer optical coatings. *Optical Engineering* 18: 104–115.

Church L, Jenkinson HA and Zavada JM (1979) Relationship between surface scattering and microtopographic features. *Optical Engineering* 18: 125–136.

Duparré A (1995) Light scattering of thin dielectric films. In: Hummel RE and Günther KH (eds) *Thin Films for Optical Coatings*, pp. 273–303. Boca Raton, FL: CRC Press.

Elson JM (1984) Theory of light scattering from a rough surface with an inhomogeneous dielectric permittivity. *Physical Review B* 30: 5460–5480.

Elson JM and Bennett JM (1979) Vector scattering theory. *Optical Engineering* 18: 116–124.

Herman IP (1996) *Optical Diagnostics for Thin Film Processing*, pp. 522–529. London: Academic Press.

Kadkhoda P, Müller A, Ristau D, *et al.* (2000) International round-robin experiment to test the international organization for standardization total-scattering draft standard. *Applied Optics* 39: 3321–3332.

Kassam S, Duparré A, Hehl K, Bussemer P and Neubert J (1992) Light scattering from the volume of optical thin films: theory and experiment. *Applied Optics* 31: 1304–1313.
Leonhard TA and Rudolph P (1995) BRDF round robin test of ASTME 1392. *Proceedings SPIE* 1995: 285–293.
Macleod HA (2001) *Thin-Film Optical Filters*, 3rd edn., pp. 12–52. Bristol, UK and Philadelphia: Institute of Physics Publishing.
Rönnow D and Veszelei E (1994) Design review of an instrument for spectroscopic total integrated light scattering measurements in the visible wavelength region. *Review of Scientific Instruments* 65: 327–334.
Stover JC (1995) *Optical Scattering, Measurement and Analysis*, 2nd edn., pp. 133–176. Bellingham, CA: SPIE Press.
Van de Hulst HC (1957) *Light Scattering by Small Particles*, pp. 28–31, 63–84. London: Chapman and Hall Ltd.
Whitehouse DJ (1994) *Handbook of Surface Metrology*, pp. 49–60, 512–517. Bristol, UK and Philadelphia: Institute of Physics Publishing.

Scattering Phenomena in Optical Fibers

P Tchofo-Dinda and G Millot, Université de Bourgogne, Dijon, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Nonlinear phenomena that result from interaction of intense light beams with dielectric media can be broadly divided into two main categories: namely, parametric effects and scattering phenomena. Parametric effects arise whenever the state of the matter is left unchanged by the interaction, whereas scattering phenomena imply transitions between vibrational or rotational energy levels in the medium. The two dominant scattering phenomena in optical fibers are stimulated Raman scattering (SRS) and stimulated Brillouin scattering (SBS), which are nonlinear processes in which a part of the energy propagating at a given frequency is converted into one or several down-shifted Stokes beams, through interaction with the vibrational modes of the materials. Although SBS and SRS phenomena are beneficial for developing all-optical fiber lasers and amplifiers for long-haul fiber transmissions, they are detrimental to transmission of ultrashort light pulses in silica fibers. In this article we describe the fundamental aspects of SBS and SRS.

General Features

Three well-known kinds of scattering phenomena in optical fibers are Rayleigh, Raman and Brillouin scattering. Although these three phenomena manifest themselves in qualitatively different ways, all of them originate fundamentally from the effects of a light wave on the atomic and molecular charges of the dielectric medium. The Raman and Brillouin scattering phenomena involve the vibrational modes of the material that are associated with the optical and acoustic branches of the dispersion curve of the material, respectively. Such vibrational modes, called phonons, are schematically represented in Figure 1 with the help of typical dispersion curves of a one-dimensional atomic lattice.

In standard silica fibers (SiO_2), the phonons associated with the optical branch correspond to intramolecular vibrations, which give rise to the Raman scattering phenomenon, whereas those associated with the acoustic branch correspond to intermolecular vibrations, which are responsible for Brillouin scattering. In either of these two scattering processes, part of the energy of the incoming light, which propagates at a given frequency ω, is converted into downshifted (Stokes) or upshifted (anti-Stokes) light waves, at frequencies $\omega - \Omega$ and $\omega + \Omega$, respectively. The frequency shift Ω is determined by the vibrational modes involved in the scattering process. Figure 2 represents schematically

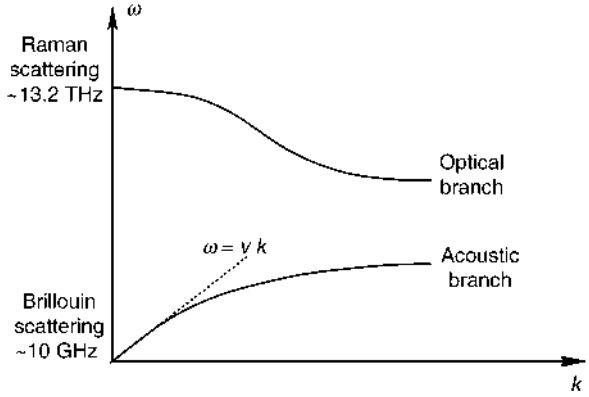

Figure 1 Schematic representation of the dispersion curves for material waves.

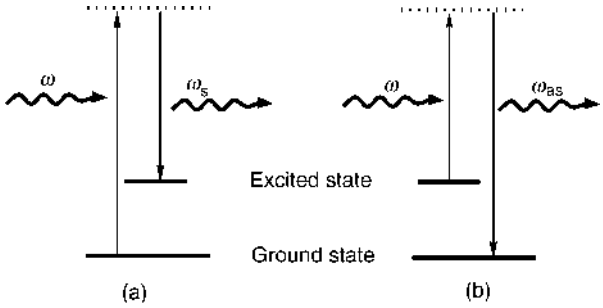

Figure 2 Schematic representation of the diagrams of energy transfer between incoming photons with frequency ω and the material, leading to (a) generation of the Stokes radiation ω_s, and (b) generation of the anti-Stokes radiation ω_{as}.

these energy conversion processes with the help of only two vibrational states, to simplify matters.

As Figure 2a shows, the material, which is initially in the ground state, absorbs a fraction $\hbar\Omega$ of the energy $\hbar\omega$ of incoming photons, to move up to the excited state. As a result, the frequency of the photons involved in this process is downshifted to $\omega_s = \omega - \Omega$. Those photons that are generated at frequency ω_s are commonly called 'Stokes radiation'. On the other hand, when a sufficiently large number of molecules is excited, they can interact with the incoming photons and return back to the ground state. As Figure 2b shows, in this process the energy is transferred from the material to the incoming photons, which thereby generates an upshifted anti-Stokes radiation at frequency $\omega_{as} = \omega + \Omega$. The Raman (Brillouin) frequency shift Ω is of the order of $\Omega_R \approx 13\,\text{THz}$ ($\Omega_B \approx 10\,\text{GHz}$).

In fact, the Raman and Brillouin effects may lead to so-called spontaneous processes, which occur in the limit of low-amplitude for all waves other than the incoming wave. For example, when a laser beam is injected in the fiber with a sufficiently low intensity, one can observe a spontaneous scattering process as schematically represented in Figure 3a.

One can clearly identify in Figure 3a the Raman and Brillouin radiation, as well as radiation scattered at the same frequency as the incoming wave, that is, Rayleigh radiation. Now, when the incoming wave propagates together with light waves whose intensity is no longer negligible with respect to that of the incoming wave, the Raman and Brillouin effects can lead to stimulated processes, in which energy is continually converted from higher to lower frequency photons. As a result, the number of Stokes photons is continually amplified at the expense of anti-Stokes photons, as schematically

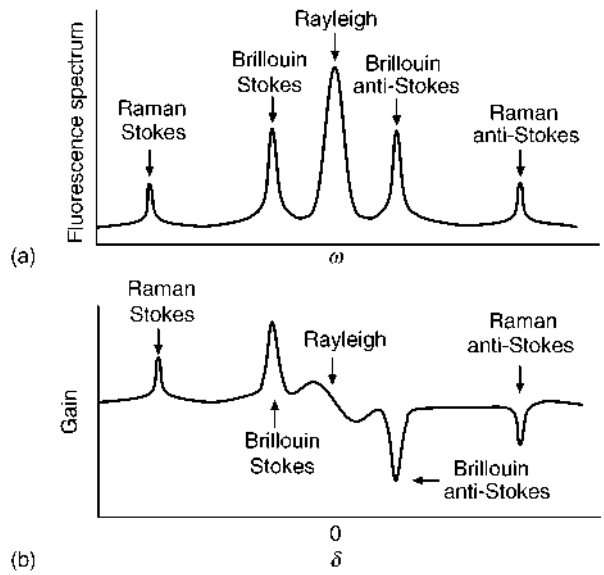

Figure 3 (a) Schematic representation of the frequency spectra for spontaneous and stimulated Raman and Brillouin scattering processes. (a) Fluorescence spectrum; (b) gain spectrum for stimulated Raman and Brillouin scattering processes. A pump and a probe waves are injected together in the fiber. δ is the frequency spacing between the pump and the probe wave.

represented in Figure 3b. The stimulated process can be obtained by either launching the incoming wave ω together with a probe wave at frequency $\omega - \Omega$ (which corresponds to the situation represented in Figure 3b), or by using a sufficiently intense incoming light beam. Indeed, above a certain power threshold (which depends strongly on the fiber length and the amount of dopants in the fiber), for which the gain corresponding to the Stokes amplification compensates for the linear fiber losses, the energy conversion from the incoming wave to the scattered waves can become sufficiently important for the scattering process to become stimulated. This process is commonly referred to as a self-stimulated process, as it does not require the use of a probe wave. Although the SRS and SBS exhibit some common general features as those mentioned above, there exist however some major differences between these two phenomena, from a fundamental point of view as well as for their practical applications to optical communication systems. We discuss separately these specific features for SRS and SBS.

Stimulated Raman Scattering

Under adiabatic following by the electrons of the optical-field-induced nuclear motions, the third-order nonlinear polarization in optical fibers may be

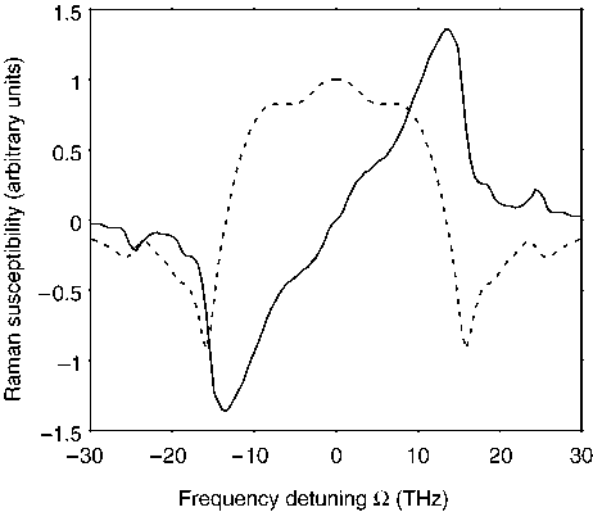

Figure 4 Schematic representation of the Raman susceptibility in an optical fiber. The solid curve represents the imaginary part of the susceptibility, that is, the Raman gain. The dashed curve represents the real part of the susceptibility.

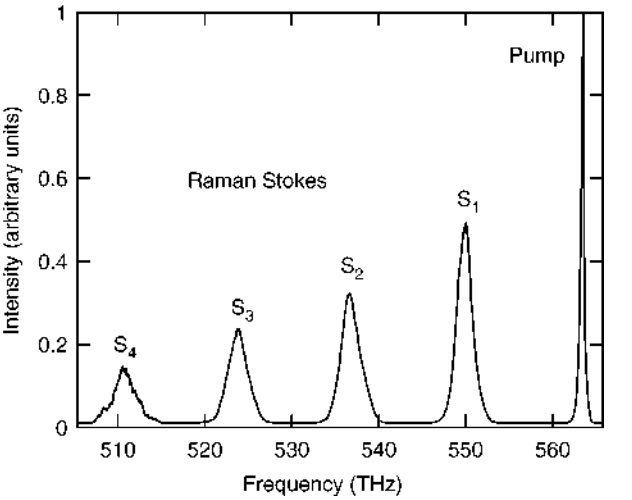

Figure 5 Self-stimulated Raman spectrum (i.e., amplified spontaneous noise), as recorded for a peak pump power $P_0 =$ 150 W and a fiber length $L = 10$ m.

written as

$$P(t) = \sigma|E(t)|^2E(t) + E(t)\int_{-\infty}^{t} \mathrm{d}s\sigma_{\mathrm{R}}(t-s)|E(s)|^2 \quad [1]$$

where σ and σ_{R} represent the instantaneous (Kerr) and noninstantaneous (Raman) parts of the fiber response, respectively. Here, we consider linearly polarized waves, for simplicity. Figure 4 shows schematically the Raman susceptibility of the fused silica, $\chi_{\mathrm{R}}(\Omega) = \sigma_{\mathrm{R}}(\Omega)/2$, which is quite close to that of standard silica fibers.

In Figure 4, the real and imaginary parts of χ_{R} are represented in solid and dashed curves, respectively. The imaginary part of the Raman susceptibility, which corresponds to the Raman gain $G_{\mathrm{R}} = \mathrm{Im}(\chi_{\mathrm{R}})$, is an experimentally measurable quantity. For example, at a pump wavelength $\lambda = 1\,\mu\mathrm{m}$, the maximum value of the Raman gain in standard silica fibers is $\approx 10^{-13}\,\mathrm{mW}^{-1}$. In fact the Raman gain $G_{\mathrm{R}}(\Omega)$ depends on the type and amount of dopants used in the fabrication of the fiber. In contrast to crystalline silica where the Raman gain occurs at a well-defined frequency of 13.2 THz, the Raman gain of silica fibers extends over a large frequency range, as can be seen in Figure 4, with a broad prominent peak around 13.2 THz. In the absence of an input signal the broadband noise is amplified within the Raman Stokes components, which are centered about the frequency detuning of 13.2 THz.

If the power of the Stokes wave becomes large enough, the Stokes wave serves as a pump to generate a second-order Stokes wave. This process of cascade SRS can generate multiple Stokes waves whose number depends on the input pump power and fiber length. Figure 5 displays a typical experimental Raman spectrum generated from a Q-switched, frequency-doubled Nd:YAG laser, emitting 5 ns duration (30 GHz spectral width) pulses at 563.63 THz (532.26 nm). The laser output was injected along one birefringence axis of a 14 m length of polarization-maintaining single-mode fiber. At an estimated injected peak power of 200 W, several higher-order Raman Stokes orders S_j $(j = 1, 4)$ (separated by 13.2 THz) are generated between 510 and 550 THz (588 and 545 nm). The spectral profile of each resulting Stokes line is near-Lorentzian with a full width at half maximum ranging from about 2 THz for the first Stokes to about 3.5 THz for the fourth Stokes. The broadening of the Raman Stokes wave, as the Stokes order increases, is due to several competing nonlinear processes.

This broadband gain is due to the amorphous nature of fused silica, in which the molecular vibrational frequencies spread out into bands which overlap and create a continuum. This outstanding feature has been fully exploited to develop wide-band amplifiers for multichannel optical communications. Recent work demonstrated the achievement of a Raman amplifier with a bandwidth of 100 nm, by use of a set of height pumps and an appropriate choice of the frequency separations between the pumps. Another important application of the Raman gain in optical fibers lies in the development of fiber-Raman lasers, which can be

tuned over a wide frequency range (≈10THz). Such lasers can be obtained by inserting a piece of single-mode fiber inside a Fabry–Perot cavity formed by two partially reflecting mirrors. A rotating output mirror and an intracavity prism allow one to select the Stokes light that provides the laser wavelength. Another example is that the Raman gain can be used to assist non-phase-matched parametric interactions.

A large power-gain enhancement was recently demonstrated for nonphase-matched waves in a three-wave mixing interaction. The Raman-assisted three-wave mixing can be used to achieve wide-bandwidth frequency conversion processes without having to satisfy the strict conditions imposed by phase-matching conditions. The above discussion clearly illustrates some useful effects of the SRS in optical fibers. On the other hand, the SRS may represent a drawback for some optical communication systems, as we discuss below.

Let us consider a pump wave injected in the fiber at frequency ω_0. The total electric field in the fiber may be written as follows:

$$E = \frac{1}{2}\mathcal{A}(z,t)\exp[\mathrm{i}(k_0 z - \omega_0 t)] \qquad [2]$$

where $\mathcal{A}$ is a slowly varying field envelope. By expanding the wavevector k_0 of the electric field in a Taylor series around ω_0, the amplitude $\mathcal{A}$ of the electric field is found to satisfy the generalized nonlinear Schrödinger equation (NLSE):

$$\begin{aligned}\frac{\partial\mathcal{A}}{\partial z} &+ \frac{\mathrm{i}}{2}\beta\frac{\partial^2\mathcal{A}}{\partial t^2}\\ &= \mathrm{i}\gamma(1-\rho)|\mathcal{A}|^2\mathcal{A} + \mathrm{i}\gamma\rho\mathcal{A}\int_0^\infty \chi_R(s)|\mathcal{A}|^2(t-s)\mathrm{d}s\\ &= \mathrm{i}\gamma|\mathcal{A}|^2\mathcal{A} + \mathrm{i}\gamma\rho C_1\mathcal{A}\frac{\partial|\mathcal{A}|^2}{\partial t} + \frac{\mathrm{i}}{2}\gamma\rho C_2\mathcal{A}\frac{\partial^2|\mathcal{A}|^2}{\partial t^2}\\ &\quad + \frac{\mathrm{i}}{6}\gamma\rho C_3\mathcal{A}\frac{\partial^3|\mathcal{A}|^2}{\partial t^3} + \cdots \qquad [3]\end{aligned}$$

In eqn [3], $C_1 \approx -7.9\times10^{-3}\,\mathrm{ps}$, $C_2 \approx -1.28\times10^{-4}\,\mathrm{ps}^2$, $C_3 \approx 9\times10^{-6}\,\mathrm{ps}^3$, and the parameter ρ, which measures the fractional contribution of Raman effects to the total nonlinearity, is ≈ 0.18 for standard silica fibers.

When the wave propagation involves only a relatively small frequency range around the pump frequency (less that 1 THz) then only the term proportional to C_1 in eqn [3], i.e., the leading term of the Raman contribution, plays a significant role. Most high-capacity transmission systems using a single channel fall into this situation, as long as picosecond pulses are used. The term proportional to C_1 induces a continuous frequency shift of the pulse frequency, as a result of a small transfer of energy from higher- to lower-frequency waves. This phenomenon, known as the soliton self-frequency shift, constitutes one of the undesirable effects of SRS in optical communications.

On the other hand, when the wave propagation involves a relatively large frequency region (of the order of a few THz), then higher-order terms that are proportional to C_2 and C_3 begin to play a significant role. In particular, the term proportional to C_3 indicates a strong transfer of energy from higher- to lower-frequency waves. In optical communications, this effect causes higher-wavelength channels of wavelength-division multiplexing systems to be amplified while depleting the lower-wavelength channels. This effect can be reduced by using filters and more amplification to compensate for the depletion at lower wavelengths (but this is achieved at the expense of a degradation of the signal-to-noise ratio).

Stimulated Brillouin Scattering

In optical fibers, the SBS occurs at input power levels that are much lower that those needed for SRS, and manifests itself through the generation of a backward-propagating Stokes wave. Although SBS is not (strictly speaking) a parametric process, the generation of the Brillouin Stokes radiation is commonly described as a parametric interaction between the pump wave, the Stokes wave and an acoustic wave (material wave). Indeed, one assumes that the annihilation of a pump photon creates simultaneously a Stokes photon and an acoustic phonon, with the following conservation laws for the energy and the momentum:

$$\Omega_\mathrm{B} = \omega_0 - \omega_\mathrm{s} \qquad [4]$$

$$K_\mathrm{B} = k_0 - k_\mathrm{s} \qquad [5]$$

where ω_0 and ω_s are the frequencies, and k_0 and k_s are the wavevectors of the incoming and Stokes waves, respectively. Then, assuming $|k_\mathrm{s}| \approx |k_0|$, the acoustic wave is found to satisfy the following dispersion relation

$$\Omega_\mathrm{B} = K_\mathrm{B}v = 2v|k_0|\sin(\theta/2) \qquad [6]$$

where v is the phase velocity of the acoustic wave, and θ is the angle between the pump and Stokes wavevectors. As eqn [6] shows, the frequency shift

Ω_B depends on the scattering angle. Ω_B is a maximum for $\theta = \pi$ (i.e., in the backward direction) and vanishes for $\theta = 0$ (forward direction). Thus, although eqn [6] predicts that Brillouin scattering should not occur in the forward direction, spontaneous Brillouin scattering can occur in the forward direction in optical fibers. But this phenomenon is quite negligible. In optical fibers, SBS occurs mainly in the backward direction with a frequency shift given by

$$\nu_B = \Omega_B/(2\pi) = 2v|k_0| = 2n_0 v/\lambda_0 \qquad [7]$$

where n_0 is the refractive index and λ_0 is the pump wavelength (incoming wave). In silica fibers, $v \approx 6\,\mathrm{km\,s^{-1}}$ and $n_0 \approx 1.45$; which leads to $\nu_B \approx 10\,\mathrm{GHz}$ at $\lambda_0 = 1.55\,\mu\mathrm{m}$. Thus, the Brillouin frequency shift is smaller by three orders of magnitude compared with the Raman frequency shift ($\approx$13 THz). The amorphous nature of fused silica leads to a Brillouin-gain coefficient $G_B(\nu)$ which extends over a frequency range of ≈ 10 MHz, with a peak value at $\nu = \nu_B$. The Brillouin gain can be approximated by a Lorentzian profile given by

$$G_B(\nu) = \frac{(\Delta\nu_B/2)^2}{(\nu - \nu_B)^2 + (\Delta\nu_B/2)^2} G_B(\nu_B), \qquad [8]$$
$$G_B(\nu_B) = \frac{2\pi n_0^7 p_{12}^2}{c d_0 \lambda_0^2 v \Delta\nu_B}$$

where $\Delta\nu_B$ is the full width at half maximum, p_{12} is the elasto-optic coefficient and d_0 is the material density.

The Brillouin gain has been exploited to develop fiber-Brillouin lasers. Such lasers are obtained by inserting a piece of fiber inside a (ring or Fabry–Perot) cavity. Another useful application of the Brillouin gain of an optical fiber lies in the amplification of a weak signal. Fiber-Brillouin amplifiers allow one to amplify signals with a frequency shift that corresponds to the Brillouin frequency shift ν_B. Fiber-Brillouin amplifiers can provide up to 40 dB gain at pump powers of a few milliwatts. But the bandwidth of such amplifiers is relatively narrow.

Conclusion

In fact scattering phenomena can be beneficial as well as detrimental for optical communication systems. If signals are transmitted in a communication channel with a power above the threshold of a given scattering phenomenon, this phenomenon will induce a transfer of energy from the signal to the Stokes radiation. The resulting signal depletion would then require more amplification. In multichannel communications, the SBS (SRS) can induce cross-talk between two bidirectional channels if their frequency separation happens to be close to the Brillouin (Raman) shift, $\approx$10 GHz ($\approx$13 THz). However, those undesirable effects can be avoided by in-line control techniques for pulse transmission in fiber-optics communication systems. On the other hand, the Raman and Brillouin effects are useful for generating light at new frequencies as well as for amplifying weak light signals. In particular Raman amplifiers are one of the key devices for high-speed telecommunication systems.

List of Units and Nomenclature

Group velocity dispersion coefficient β_2 [$\mathrm{ps^2\,km^{-1}} = 10^{-27}\mathrm{s^2\,m^{-1}}$]
Nonlinear parameter γ [$\mathrm{W^{-1}\,m^{-1}}$]
NLS (E): nonlinear Schrödinger (equation)
Optical intensity I [$\mathrm{GW\,cm^{-2}} = 10^9\,\mathrm{W\,cm^{-2}}$]
SBS: stimulated Brillouin scattering
SRS: stimulated Raman scattering
WDM: wavelength division multiplexing

See also

Fiber and Guided Wave Optics: Light Propagation; Nonlinear Effects (Basics). **Optical Amplifiers:** Raman, Brillouin and Parametric Amplifiers. **Solitons:** Soliton Communication Systems.

Further Reading

Agrawal GP (2001) Nonlinear Fiber Optics. San Diego, CA: Academic Press.
Agrawal GP (2001) Applications of Nonlinear Fiber Optics. San Diego, CA: Academic Press.
Bloembergen N (1977) Nonlinear Optics. Reading, MA: Benjamin.
Dianov EM, Mamyshev PV, Prokhorov AM and Serkin VN (1989) Nonlinear Effects in Optical Fibres. Switzerland: Harwood Academic Publishers.
Hellwarth R (1977) Third-order optical susceptibilities of liquids and solids. *Progress in Quantum Electronics* 5: 1–68.
Pocholle JP, Papuchon M, Raffy J and Desurvire E (1990) Non linearities and optical amplification in single mode fibres. *Revue Technique Thomson-CSF* 22: 187–268.
Shen YR (1984) Principles of Nonlinear Optics. New York: Wiley.

Scattering Theory

Y A Eremin, Moscow State University, Moscow, Russia

© 2005 Published by Elsevier Ltd.

Lord Rayleigh (John William Strutt) (1842–1916)

In the nineteenth century, Lord Rayleigh offered the first explanation for the sky's blue color (on a clear day!). From Rayleigh's initial theory, representing a simplification of Maxwell's electromagnetic theory for the case of particles very small compared to the wavelength of the incident light, the theory was extended to describe the scattering of light by larger macromolecules in solution.

Scattering theory has played a central role in twentieth-century mathematical physics. From Rayleigh's explanation of why the sky is blue to modern medical application of computerized tomography, has attracted both scientists and mathematicians for over a hundred years. Scattering theory is concerned with the effect an inhomogeneous medium has on an incident wave. In particular, if the total field is viewed as sum of an incident field and a scattered field, then the direct scattering problem is to determine the scattered field from a knowledge of the incident one, obstacle properties, and the equations governing the wave motion. In contrast, the inverse scattering problem consists of reconstruction of the inhomogeneity from the knowledge of asymptotic behavior of the scattered field. An everyday example is human vision: from the measurements of scattered light that reaches our retinas, our brain reconstructs a detailed three-dimensional map of the world around us. In summary, the task of direct scattering theory is to determine the relation between the incident and scattered waves. The task of inverse scattering theory is to determine properties of the obstacle.

Light propagating can be described in terms of an electromagnetic field, which is composed of two related vector fields, the electric and the magnetic fields. This means that the vectors of electric ($\mathbf{E}$) and magnetic ($\mathbf{H}$) field have values defined at each point in space and time. The vector fields $\mathbf{E}$ and $\mathbf{H}$ interact through the material parameters: electric permittivity ε and magnetic permeability μ. The behavior of electromagnetic fields can be described by Maxwell's equations.

Maxwell Equations. James Clerk Maxwell (1831–1879)

Consider the electromagnetic wave (light) propagation in a homogeneous, isotropic medium with electric permittivity $\varepsilon > 0$, electric conductivity $\sigma \geq 0$, and magnetic permeability $\mu > 0$, containing a local obstacle. We assume that $\sigma = 0$ outside an obstacle. If J is a current density, the electric field $\vec{E}(x,t)$ and magnetic field $\vec{H}(x,t)$ satisfy the Maxwell equations:

$$\begin{aligned} \nabla \times \vec{E}(x,t) + \mu \frac{\partial \vec{H}}{\partial t}(x,t) &= 0 \\ \nabla \times \vec{H}(x,t) - \varepsilon \frac{\partial \vec{E}}{\partial t}(x,t) &= \vec{J}(x,t) \end{aligned} \qquad [1]$$

Also, in an isotropic conductor, the electric field satisfies Ohm's law:

$$\sigma \vec{E}(x,t) = \vec{J}(x,t) \qquad [2]$$

The speed of wave propagation for the electric and magnetic fields is $c = 1/\sqrt{\varepsilon\mu}$ and in free space $c = 1/\sqrt{\varepsilon_0\mu_0} = 2.99792458 \times 10^8\ \mathrm{m\ s^{-1}}$. Assume that the electromagnetic field is time-harmonic, i.e., of the form:

$$\begin{aligned} \vec{E}(x,t) &= \frac{1}{\sqrt{\varepsilon_0}}\mathbf{E}(x)\mathrm{e}^{-\mathrm{j}\omega t} \\ \vec{H}(x,t) &= \frac{1}{\sqrt{\mu_0}}\mathbf{H}(x)\mathrm{e}^{-\mathrm{j}\omega t} \end{aligned} \qquad [3]$$

where $\omega > 0$ is the frequency. Many phenomenon of light scattering can be described based on monochromatic wave propagation. Let us consider a local obstacle occupying a bounded domain D. Inside the obstacle we introduce relative complex permittivity:

$$\varepsilon = \frac{1}{\varepsilon_0}\left(\varepsilon + \mathrm{j}\frac{\sigma}{\omega}\right) \qquad [4]$$

Then $\mathbf{E}(x)$ and $\mathbf{H}(x)$ obey the time-harmonic Maxwell equations:

$$\nabla \times \mathbf{E}(x) = \mathrm{j}k\mathbf{H}(x), \quad \nabla \times \mathbf{H}(x) = -\mathrm{j}k\varepsilon\mathbf{E}(x) \qquad [5]$$

where $k^2 = \varepsilon_0\mu_0\omega^2$.

Helmholtz Equation. Herman Ludwig von Helmholtz (1821–1894)

A wave equation for vector $\mathbf{E}$ can be readily derived from the Maxwell equations. We obtain:

$$(\nabla^2 + k^2)\mathbf{E}(x) = 0 \qquad [6]$$

This is known as the homogeneous Helmholtz equation. A general planewave solution of the equation can be written as

$$\mathbf{E}(x) = \mathbf{E}_0\, e^{i\vec{k}\cdot\vec{x}} \qquad [7]$$

here $\mathbf{E}_0$ is a constant vector independent of spatial coordinates and $\vec{k}\cdot\vec{x}$ is a scalar product of the vectors $\vec{x} = \{x_1, x_2, x_3\}$ and $\vec{k} = \{k_1, k_2, k_3\}$, the so called wavevector. Equations [6] and [7] yield the dispersion relation:

$$k^2 = \varepsilon_0\mu_0\omega^2 = k_1^2 + k_2^2 + k_3^2 \qquad [8]$$

For a given $\vec{k}$ vector, a constant phase front is determined by $\vec{k}\cdot\vec{x} = \text{const}$, which indicates that the front is perpendicular to the wavevector. So, the phase front is a plane and the amplitude of the electric field on the plane is a constant.

The Maxwell equations for the plane wave solution become:

$$\begin{aligned} &\vec{k}\times\mathbf{E}(x) = k\mathbf{H}(x), \quad \vec{k}\times\mathbf{H}(x) = -k\mathbf{E}(x) \\ &\Rightarrow \vec{k}\cdot\mathbf{E} = 0, \quad \vec{k}\cdot\mathbf{H} = 0 \end{aligned} \qquad [9]$$

Therefore, vectors $\mathbf{E}$ and $\mathbf{H}$ belong to the constant phase plane perpendicular to vector $\vec{k}$ and are orthogonal each other.

Poynting's Vector. John Henry Poynting (1852–1914)

For time-dependent fields, the Poynting vector is introduced as

$$\vec{S}(x,t) = \vec{E}(x,t)\times\vec{H}(x,t) \qquad [10]$$

It is interpreted as the power flow density with the dimension of watts/m^2. The time-average Poynting's vector power density is given by

$$\langle S\rangle = \frac{1}{T}\int_0^T \vec{S}(x,t)\mathrm{d}t \qquad [11]$$

For the time-dependent planewave fields:

$$\begin{aligned} \vec{E}(x,t) &= \mathbf{E}_0\cos(\vec{k}\cdot\vec{x} - \omega t) \\ \vec{H}(x,t) &= \mathbf{H}_0\cos(\vec{k}\cdot\vec{x} - \omega t) \end{aligned} \qquad [12]$$

The vector Poynting's power density shown in eqn [10], accepts the form:

$$\vec{S}(x,t) = |\mathbf{E}_0|^2\cos^2(\vec{k}\cdot\vec{x} - \omega t) \qquad [13]$$

The time-average vector power density gives

$$\langle\vec{S}(x,t)\rangle = \frac{1}{2}|\mathbf{E}_0|^2 \qquad [14]$$

which shows that Poynting's power density is also propagating in the direction of the wave vector $\vec{k}$.

Boundary Conditions

Let us consider a bound obstacle D_i with a smooth closed surface S. The Maxwell eqn [5] has been written in differential form. It must be supplemented with boundary conditions enforced at the obstacle surface. The boundary conditions can be derived from the integral form of Maxwell equations. The field vectors $\mathbf{E}$, $\mathbf{H}$ are assumed to be finite but may be discontinuous across the obstacle surface S. In cases where there is no surface current density (which is so for a perfect conductor) the boundary conditions accept the following form:

$$\vec{n}\times(\mathbf{E}_\mathrm{i} - \mathbf{E}_\mathrm{e}) = 0, \quad \vec{n}\times(\mathbf{H}_\mathrm{i} - \mathbf{H}_\mathrm{e}) = 0 \qquad [15]$$

where subscripts i and e correspond to the total fields inside and outside the obstacle, respectively, and $\vec{n}$ is unit outward normal to S. Essentially the boundary conditions state that tangential components for electromagnetic fields are continuous across the boundary.

Silver–Müller Radiation Conditions

Analysis of a direct scattering problem on a local obstacle involves Maxwell eqn [10] and boundary conditions as in eqn [15] and some infinity conditions, which provide the uniqueness of the scattering problem solutions. Let us assume that a local obstacle is excited by an incident electromagnetic field $\{\mathbf{E}_0, \mathbf{H}_0\}$, given as entire solution to the Maxwell equations. Total field outside the obstacle can be represented as

$$\mathbf{E}_\mathrm{e} = \mathbf{E}_0 + \mathbf{E}_\mathrm{s}, \quad \mathbf{H}_\mathrm{e} = \mathbf{H}_0 + \mathbf{H}_\mathrm{s} \qquad [16]$$

where scattered field $\{\mathbf{E}_\mathrm{s}, \mathbf{H}_\mathrm{s}\}$ in the exterior domain $D_\mathrm{e} = \mathbb{R}^3/\bar{D}_\mathrm{i}$ satisfies the Silver–Müller radiation conditions:

$$\begin{aligned} \frac{x}{|x|}\times\mathbf{H}_\mathrm{s} + \mathbf{E}_\mathrm{s} &= \mathrm{o}\left(\frac{1}{|x|}\right) \\ \frac{x}{|x|}\times\mathbf{E}_\mathrm{s} - \mathbf{H}_\mathrm{s} &= \mathrm{o}\left(\frac{1}{|x|}\right) \end{aligned} \quad |x|\to\infty \qquad [17]$$

uniformly for all directions $x/|x|$.

Scattering Amplitude

The following result is valid: every radiating solution $\{\mathbf{E}_s, \mathbf{H}_s\}$ to the Maxwell equations has the following asymptotic form:

$$\mathbf{E}_s(x) = \frac{e^{ik|x|}}{|x|}\left\{\mathbf{E}_\infty(\varpi) + O\left(\frac{1}{|x|^2}\right)\right\} \qquad |x| \to \infty \qquad [18]$$

$$\mathbf{H}_s(x) = \frac{e^{ik|x|}}{|x|}\left\{\mathbf{H}_\infty(\varpi) + O\left(\frac{1}{|x|^2}\right)\right\}$$

where the vector functions $\mathbf{E}_\infty$ and $\mathbf{H}_\infty$ defined on the unit sphere $\varpi \in \Omega = \{0 \le \vartheta \le \pi; 0 \le \varphi \le 2\pi\}$ are known as the electric far field pattern (scattering amplitude) and magnetic far field pattern, respectively. They satisfy:

$$\mathbf{H}_\infty = \vec{\nu} \times \mathbf{E}_\infty, \quad \vec{\nu} \cdot \mathbf{E}_\infty = \vec{\nu} \cdot \mathbf{H}_\infty = 0 \qquad [19]$$

with the unit outward normal $\vec{\nu}$ on Ω. The latter means that $\mathbf{E}_\infty$ and $\mathbf{H}_\infty$ have only tangential components at the unit sphere.

Far field pattern represents real value analytic functions on a unit sphere. Reillich's lemma establishes one-to-one correspondence between radiating electromagnetic field and their far field pattern. It means that there exists only one scattered field $\{\mathbf{E}_s, \mathbf{H}_s\}$ corresponding to a given far field pattern $\mathbf{E}_\infty$.

Reciprocity Relations

Examine the set of far field pattern corresponding to the scattering problem including: Maxwell eqn [5], boundary conditions in eqn [15], and radiation conditions in eqn [17] at the infinity for the scattered fields. Let the incident electromagnetic field is given by the planewave:

$$\mathbf{E}_0(x) = \vec{e}_0\, e^{jk\vec{x}\cdot\vec{\kappa}}, \quad \mathbf{H}_0(x) = \vec{\kappa} \times \vec{e}_0\, e^{jk\vec{x}\cdot\vec{\kappa}} \qquad [20]$$

where $\vec{\kappa} \in \Omega$ is a unit vector giving the propagation direction and $\vec{e}_0 \in \mathbb{R}^3$ is a constant vector responsible for the polarization. From the previous consideration one can see that:

$$\mathbf{E}_s(x; \vec{\kappa}, \vec{e}_0) = \frac{e^{ik|x|}}{|x|}\left\{\mathbf{E}_\infty(\varpi\vec{\kappa}, \vec{e}_0) + O\left(\frac{1}{|x|^2}\right)\right\}, \quad |x| \to \infty \qquad [21]$$

Furthermore, the following reciprocity relation for the scattering amplitude holds:

$$\vec{h}_0 \cdot \mathbf{E}_\infty(\varpi; \vec{\kappa}, \vec{e}_0) = \vec{e}_0 \cdot \mathbf{E}_\infty(-\vec{\kappa}; -\varpi, \vec{h}_0)$$

for all vectors $\varpi, \vec{\kappa} \in \Omega$ and $\forall \vec{e}_0, \vec{h}_0 \in \mathbb{R}^3$.

Optical Theorem

Far field pattern enables us to evaluate scattering cross section, which is defined as

$$\sigma_s = \int_\Omega |\mathbf{E}_\infty(\varpi)|^2 d\varpi \qquad [22]$$

For the planewave given by eqn [20] and excitation of nonabsorbing obstacle ($\sigma = 0$), the following relation holds:

$$\sigma_s = \frac{4\pi}{k} \mathrm{Im}\{\mathbf{E}_\infty(\vec{\kappa}) \cdot \vec{e}_0\} \qquad [23]$$

This is so-called optical theorem.

Huygen's Principle (Green's Theorem). Christian Huygens (1629–1695)

The Huygens' principle shows that a wavefield on the surface determines the wavefield off the surface. Let us introduce the scalar Green's function which satisfies the Helmholtz equation:

$$(\nabla^2 + k^2) g(x, y) = -\delta(x - y) \qquad [24]$$

where $\delta(x - y)$ is the Dirac delta function. For three-dimensional problems, the scalar Green's function $g(x, y)$ for isotropic media accepts the form:

$$g(x, y) = \frac{e^{jk|x-y|}}{4\pi|x - y|} \qquad [25]$$

Let electromagnetic fields $\{\mathbf{E}_s, \mathbf{H}_s\}$ be a radiating solution to the Maxwell eqn [5] in D_e. Then for any $x \in D_e$ we have the Stratton–Chu formulas:

$$\begin{aligned} \mathbf{E}_s(x) &= \nabla \times \int_S g(x, y) \vec{\nu}_y \times \mathbf{E}_s(y) ds_y \\ &\quad + \frac{j}{k} \nabla \times \nabla \times \int_S g(x, y) \vec{\nu}_y \times \mathbf{H}_s(y) ds_y \\ \mathbf{H}_s(x) &= \nabla \times \int_S g(x, y) \vec{\nu}_y \times \mathbf{H}_s(y) ds_y \\ &\quad - \frac{j}{k} \nabla \times \nabla \times \int_S g(x, y) \vec{\nu}_y \times \mathbf{E}_s(y) ds_y \end{aligned} \qquad [26]$$

here S is a smooth closed surface, and $\vec{\nu}$ is the unit normal vector to S directed into D_e. Similar relations are valid for an electromagnetic field inside a bounded domain D_i. The Stratton–Chu

formulas are extensively employed to construct boundary integral equations for scattering problems investigating.

Homogeneous Obstacle

Consider a mathematical statement of the scattering problem by a bound homogeneous obstacle. Let external excitation be a planewave $\{\mathbf{E}_0, \mathbf{H}_0\}$: (eqn [20]). In this case the boundary-value scattering problem can be formulated in the form:

$$\begin{aligned} &\nabla\times\mathbf{E}_\tau(x) = \mathrm{j}k\mathbf{H}_\tau(x), \quad \nabla\times\mathbf{H}_\tau(x) = -\mathrm{j}k\varepsilon_\tau\mathbf{E}_\tau(x), \\ &\quad x\in D_\tau, \tau = \mathrm{i}, \mathrm{e} \\ &\qquad \vec{\nu}\times(\mathbf{E}_\mathrm{i}+\mathbf{E}_\mathrm{e}) = \vec{\nu}\times\mathbf{E}_0, \\ &\qquad \vec{\nu}\times(\mathbf{H}_\mathrm{i}+\mathbf{H}_\mathrm{e}) = \vec{\nu}\times\mathbf{H}_0 \ \text{ on } S \\ &\qquad \frac{x}{|x|}\times\mathbf{H}_\mathrm{e}+\mathbf{E}_\mathrm{e} = \mathrm{o}\left(\frac{1}{|x|}\right), \quad |x|\to\infty \end{aligned} \qquad [27]$$

where $\varepsilon_\mathrm{e} = 1$, $\varepsilon_\mathrm{i} = \varepsilon$ and $\{\mathbf{E}_\mathrm{e}, \mathbf{H}_\mathrm{e}\}$ corresponds to scattered field. For a smooth surface and the case $\mathrm{Im}\,\varepsilon \geq 0$, the problem has a unique solution.

Lippman–Swinger Integral Equation

Assume we deal with the local inhomogeneous obstacle that can be described by dielectric permittivity $\varepsilon(x)$, which is a continuous complex-valued ($\mathrm{Im}\,\varepsilon(x) \geq 0$) function in D_i. The correspondent scattering problem can be formulated as below in eqn [29]. In this case, the boundary-value scattering problem can be reduced to the solution of the Lippman–Swinger integral equation

$$\begin{aligned} \mathbf{E}_\mathrm{i}(x) = {} & \mathbf{E}_0(x) + \nabla\times\nabla\times\int_{D_\mathrm{i}} g(x,y)k^2(\varepsilon(y)-1) \\ & \times\mathbf{E}_\mathrm{i}(y)\mathrm{d}\nu_y, \quad x\in D_\mathrm{i} \end{aligned} \qquad [28]$$

Volume integral eqn [28] is equivalent to the corresponding scattering problem. Once the solution is found, the associated scattered field outside the inhomogeneous obstacle can be represented as

$$\begin{aligned} \mathbf{E}_\mathrm{s}(x) = {} & \nabla\times\nabla\times\int_{D_\mathrm{i}} g(x,y)k^2(\varepsilon(y)-1) \\ & \times\mathbf{E}_\mathrm{i}(y)\mathrm{d}\nu_y, \quad x\in D_\mathrm{e} \end{aligned} \qquad [29]$$

The Lippman–Swinger integral equation plays an important role under analysis of a wave scattering by local inhomogeneous obstacle.

Born Approximation

In the case where $k^2|\varepsilon(x)-1|mesD_\mathrm{i} \ll 1$, or where the second term on the right-hand part of the Lippman–Swinger integral eqn [28] is small compared to the first term, one can write $\mathbf{E}_\mathrm{i}(x) \simeq \mathbf{E}_0(x)$. Then the scattered field can be approximately evaluated as

$$\begin{aligned} \mathbf{E}_\mathrm{s}(x) = {} & \nabla\times\nabla\times\int_{D_\mathrm{i}} g(x,y)k^2(\varepsilon(y)-1) \\ & \times\mathbf{E}_0(y)\mathrm{d}\nu_y, \quad x\in D_\mathrm{e} \end{aligned} \qquad [30]$$

The above representation for the scattered field is known as the first-order Born approximation. It is also the first-order approximation in the Neumann series for the integral eqn [28]. The Born approximation plays an important role in diffraction tomography.

Rayleigh Scattering

Rayleigh scattering characterizes the scattering of electromagnetic waves by particles much smaller then an exciting wavelength. Consider a spherical particle with constant permittivity ε and radius a deposited at the origin of a coordinate system. Let linear polarized planewave be incident upon the particle $\mathbf{E}_0(x) = \vec{e}_x\, e^{\mathrm{j}kz}$, and $\vec{e}_x$ be unit vector Cartesian coordinate system. It is polarized in x-direction and propagates along z-axis. In this case, for the scattered intensity, we get:

$$I_\mathrm{s} = \left(\frac{\varepsilon-1}{\varepsilon+2}\right)^2 \frac{k^4a^6}{r^2}\sin^2\vartheta \qquad [31]$$

where r is distance from the origin and $\vartheta \in [0, \pi]$.

Corresponding scattering cross-section is calculated as

$$\sigma_\mathrm{s} = \frac{8\pi}{3}\left(\frac{\varepsilon-1}{\varepsilon+2}\right)^2 k^4a^6 \qquad [32]$$

Thus the total scattered power is proportional to the fourth power of the wave number (minus fourth power of the wavelength – Rayleigh's Law). The scattered power is also proportional to the sixth power of the radius.

Mie Scattering

The Mie theory is a complete mathematical–physical theory of the scattering of electromagnetic wave by homogeneous spherical particles, developed by Gustav Mie in 1908. In contrast to Rayleigh scattering, the Mie theory embraces all possible ratios of the particle radius to wavelength. It relays

upon the exact solution to Maxwell equation for spherical homogeneous obstacle with planewave excitation. In the frame of Mie theory, the scattered field is represented as an infinite series of fields of spherical multipoles, which are orthogonal at a sphere surface. Mie solution plays an important role in light scattering by clouds of particles.

See also

Scattering: Raman Scattering; Scattering from Surfaces and Thin Films; Scattering Phenomena in Optical Fibers; Stimulated Scattering.

Further Reading

Boheren C and Hoffman D (1983) *Absorption and Scattering of Light by Small Particles*. New York: Wiley.
Born M and Wolf E (1970) *Principles of Optics*. New York: Pergamon Press.
Chew WC (1995) *Waves and Fields in Inhomogeneous Media*. New York: IEEE Press.
Colton D and Kress R (1992) *Inverse Acoustic and Electromagnetic Scattering Problems*. New York: Springer-Verlag.
Guenther RD (1990) *Modern Optics*. New York: Wiley.
Kong JA (2000) *Electromagnetic Wave Theory*. Cambridge: EMW Publishing.
Kong JA (2002) *Maxwell Equations*. Cambridge: EMW Publishing.

Stimulated Scattering

M Bashkansky and J Reintjes, Naval Research Laboratory, Washington, DC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Scattering of light involves conversion of an incoming light wave in one mode to an outgoing wave in a different mode. Scattering can involve changes of wavelength, direction, and polarization, and can occur in all types of materials. Without scattering we would not be able to see objects in the world around us. Scattering processes involving particular physical interactions are often given distinctive names. Familiar examples include Rayleigh scattering from the random distribution of air molecules, which makes the sky blue, and Tyndall scattering from small suspended particles, which makes the sunset red. Other types of scattering, e.g., Raman, Brillouin and Rayleigh-wing scattering, are important spectroscopic tools that provide insight into atomic and molecular properties. In their most commonly encountered form, where the intensity of the scattered light is proportional to the intensity of the incident light, the scattering processes are termed spontaneous. With the intense radiation provided by lasers, however, the intensity of the scattered light can grow exponentially, and the processes are termed stimulated scattering.

Stimulated scattering is a subset of the larger discipline of nonlinear optics. The term was first used to describe nonlinear optical interactions (e.g., Raman, Brillouin, Rayleigh) in which the intensity of the scattered light grows exponentially, in a manner similar to the exponential growth of light intensity associated with lasers. These interactions involve exchange of energy with a second mode that is an internal mode of the scattering medium, and the scattered light wave is at a different wavelength from the incoming light. The use of the term stimulated has subsequently been expanded to include a wider class of nonlinear optical interactions, such as parametric fluorescence or certain aspects of self-focusing, in which the second mode is itself a light wave, with either a different frequency or direction of propagation.

In the wider sense, scattering of light occurs through interaction of the incoming light wave with an 'idler' mode in a material. Stimulated scattering occurs when the idler mode is itself driven by interference of the incoming and scattered light waves. Under these circumstances, exponential growth of the scattered wave results. Depending on the interaction, the idler mode may be an internal mode of the material or another light wave. Most linear scattering processes have stimulated counterparts. However, there are stimulated scattering interactions that have no spontaneous analogs.

The elements of stimulated light scattering can be summarized mathematically as follows. The scattering process is modeled as an incoming laser or pump, an outgoing scattered wave and an idler mode. The interaction of the laser with the idler mode to produce the scattered light wave is described by the equation

$$\frac{\partial}{\partial z}A_{\mathrm{S}}(z,t)+\frac{n}{c}\frac{\partial}{\partial t}A_{\mathrm{S}}(z,t)=\kappa A_{\mathrm{L}}b^{*} \qquad [1]$$

where A_{S} and A_{L} are the slowly varying amplitudes of the scattered and laser light waves whose electric fields are defined as

$$E_{\mathrm{S(L)}}=\frac{1}{2}\left\{A_{\mathrm{S(L)}}\mathrm{e}^{\mathrm{i}[\omega_{\mathrm{S(L)}}t-k_{\mathrm{S(L)}}z]}+\mathrm{cc}\right\} \qquad [2]$$

$\omega_{S(L)}$ is the frequency of the scattered (laser) light, $k_{S(L)}$ is the corresponding wavevector, κ is a coupling coefficient, cc denotes complex conjugate, and b is the amplitude of the idler mode given by

$$B = \frac{1}{2}\left\{b e^{i[\omega_b t - k_b z]} + cc\right\} \quad [3]$$

The interaction of the laser and scattered wave to drive the idler mode is described by the equation

$$\mathcal{L}(b^*) = \kappa' A_L^* A_S \quad [4]$$

where $\mathcal{L}$ is a differential operator that describes the dynamics of the idler mode, e.g., a damped oscillator or a propagating wave, and κ' is a second coupling coefficient. For the class of stimulated processes that involve excitation of an internal mode of the medium, $\mathcal{L}$ has the form

$$\mathcal{L} = \frac{\partial}{\partial t} + \frac{1}{T_2} - v\frac{\partial}{\partial z} \quad [5]$$

where T_2 is the damping time of the coherence of the internal mode. When interaction times are long compared to internal response and decay times,

$$b^* \sim \kappa' A_L^* A_S$$

and the scattered wave obeys an equation of the form

$$\frac{\partial}{\partial z} A_S(z,t) = \frac{g}{2}|A_L|^2 A_S(z,t)$$

which has an exponential-gain solution for the scattered wave intensity:

$$I_S(z,t) = I_S(0,t)e^{I_L g z}$$

When the idler mode is a light wave, as in stimulated parametric fluorescence, [4] has the same form as [1] and they can be solved as a coupled pair, leading to transcendental or hyperbolic solutions. Within the framework of nonlinear optics, interactions of these types can be expressed as arising from a nonlinear optical polarization, which forms the right-hand side of [1]. The dynamics of the nonlinear polarization are then described by [4].

Stimulated processes are often associated with pumping thresholds. When these processes are configured with cavity resonators, they possess a threshold in pumping intensity just as with lasers or other oscillators. When they are operated in single or multipass scattering cells, they have no true threshold. However, they still have a practical useful threshold when the growth from quantum noise raises the scattered signal to a level comparable to the pump. By convention, this is taken to be the incident light intensity at which the scattered intensity reaches approximately 1% of the pump intensity, typically involving gains of the order of e^{23}–e^{30}. At this level, a small increase in pump intensity raises the scattered intensity to levels at which the pump starts to be depleted.

Attenuation of the scattered light can be accounted for by replacing $I_L g$ in the above expression by $I_L g - \alpha$, where α is the attenuation coefficient. Exponential gain is now experienced when the pump intensity exceeds the level given by

$$I_L > \frac{\alpha}{g}$$

This value can serve as an estimate of threshold pumping in cavity configurations, although in free propagation situations, net gain must still approach e^{23} to exhibit threshold behavior.

A list of stimulated scattering interactions is given in Table 1.

Stimulated Raman Scattering

Stimulated Raman scattering (SRS) involves conversion of an incoming light wave at one frequency, termed the laser or pump wave, into a scattered wave of different frequency, along with excitation or de-excitation of an internal mode of the medium. The frequency difference between the pump and scattered waves is equal to the frequency of the material mode and is characteristic of the material. When the frequency of the scattered wave is lower than that of the incident wave, the scattered wave is termed a Stokes wave, and the internal mode of the medium is excited in the scattering process. When the frequency of the scattered wave is higher than that of the incident wave, the scattered wave is termed an anti-Stokes wave, and the internal mode of the medium is de-excited in the scattering process.

The internal mode of the medium that is involved in Raman scattering is a nonpropagating collective mode. Many different types of internal modes in all types of materials can be involved in Raman scattering. Examples include molecular vibrations and rotations in gases, liquids, or solids, electronic states that are of the same parity as the ground state in gases or solids, optical phonons or polaritons in solids, and Langmuir waves in plasmas. Raman frequency shifts can range from tens of wavenumbers for lattice vibrations to tens of thousands of wavenumbers for electronic transitions.

Spontaneous Raman scattering was first described theoretically by Smekal in 1923 from a quantum mechanical study of light scattering, and was

Table 1 List of stimulated scattering interactions

Stimulated scattering process	*Internal material mode*	*Light mode change*
Stimulated Raman scattering	Nonpropagating internal energy mode: molecular vibrations and rotations, same parity electronic states, optical phonons and Langmuir waves	Change in wavelength, polarization, direction
Stimulated Brillouin scattering	Propagating sound waves	Change in wavelength, polarization, and direction
Stimulated Rayleigh scattering	Temperature and density fluctuations due to light absorption	Change in direction
Stimulated Rayleigh wing scattering	Molecular distribution and orientation fluctuations	Change in wavelength and direction
Stimulated concentration scattering	Fluctuations of concentration in gas mixtures	Change in direction
Photorefractive stimulated scattering	Diffusion of photo-excited carriers	Change in direction
Stimulated Thomson/Compton scattering	Momentum changes in bunched free electrons	Change in wavelength and direction
Stimulated parametric fluorescence	Nonresonant nonlinear polarization	Change in wavelength and direction
Self-focusing	Nonresonant nonlinear polarization	Change in direction

demonstrated experimentally by Raman, and shortly after by Landsberg and Mandelstamm in 1928. Stimulated Raman scattering was first observed in 1962 by Woodbury and Ng, and Eckhardt *et al.* soon after the development of pulsed lasers.

Stimulated Raman scattering can be observed in many different configurations including cavity resonators, single or multiple-pass generator cells or amplifiers. It can be observed as forward scattering, with the scattered light propagating in the same direction as the pump light, or backward scattering, with the scattered light propagating in the opposite direction to the pump light. Its most widespread use is for generation of coherent light at frequencies different from the pump light. Other applications include phase conjugation, optical gating, beam cleanup, and pulse compression.

The equations describing SRS when depletion of the pump light and diffraction of the optical beams are neglected are:

$$2ik_S\left(\frac{\partial A_S}{\partial z}\pm\frac{1}{v_{gS}}\frac{\partial A_S}{\partial t}\right)=-\frac{N}{2}\frac{\omega_S^2}{c^2}\left(\frac{\partial\alpha}{\partial Q}\right)_S Q^* A_L \quad [6]$$

$$\frac{\partial Q^*}{\partial t}+\Gamma Q^*=-\frac{i}{4\omega_o m}\left(\frac{\partial\alpha}{\partial Q}\right)_S A_L^* A_S \quad [7]$$

Here A_S and A_L are the slowly varying optical field amplitudes of the Stokes and laser waves as given in [2], $\omega_{S(L)}$ are the optical frequencies, $k_{S(L)}$ are the corresponding k vectors, $v_{S(L)}$ are the group velocities, and Q is the amplitude of the idler mode, in this case the normal mode coordinate of the material excitation, defined by:

$$\mathbf{Q}(x,y,z,t)=\frac{1}{2}(Q(x,y,z,t)e^{-i(\omega_o t-k_o z)}+cc) \quad [8]$$

$(\partial\alpha/\partial Q)_S$ is the Stokes hyperpolarizability, $\omega_0=\omega_L-\omega_S$ is the Raman transition frequency, and k_0 is the k vector of the material excitation. The phase matching relation $k_S=k_L-k_0$ is automatically satisfied for SRS because the material mode that is involved is nonpropagating. Γ is the half width at half maximum of the Raman linewidth given by $\Gamma=1/T_2$, where T_2 is the dephasing time. N is the number of molecules in the ground state and m is the effective reduced mass of the material oscillation.

When the fields are constant or vary slowly compared to the response time of the material, the Stokes intensity is given by

$$I_S(z)=I_S(0)e^{I_L g_{ss} Z} \quad [9]$$

where the intensity I is given by

$$I_{S(L)}=\frac{1}{2}(cn_{S(L)}\varepsilon_0)|A_{S(L)}|^2$$

and

$$g_{ss}=\frac{N\omega_s}{4\Gamma n_S n_L m\omega_0 c^2\varepsilon_0}\left(\frac{\partial\alpha}{\partial Q}\right)^2 \quad [10]$$

where $n_{S(L)}$ is the index of refraction of the Stokes (laser) in the material and ε_0 is the dielectric permittivity of the vacuum.

The Stokes wave grows exponentially with z at a rate that depends on the pump intensity. If the Stokes wave becomes intense enough, its power can become comparable to that of the original pump wave. In that case depletion of the pump wave must be taken into account with an equation for the pump wave amplitude similar to [6]. The Stokes intensity is now given by

$$I_S(z) = \frac{I_S(0)e^{I_L(0)gz}}{1 + \frac{\omega_L}{\omega_S}\frac{I_S(0)}{I_L(0)}e^{I_L(0)gz}} \qquad [11]$$

which is valid when $I_S(0) \ll I_L(0)$.

In the limit of strong interactions all of the laser energy can be depleted. The maximum Stokes energy is given by $I_S = I_L(\omega_S/\omega_L)$, where ω_S/ω_L is termed the Manly–Rowe ratio. The energy that is removed from the laser but not converted to Stokes radiation is transferred to the material.

For each photon produced in the scattered wave, a corresponding quantum of energy is given to the material mode and a photon is removed from the laser. In principle, 100% of the laser photons can be converted to Stokes photons, but complete conversion is not achieved in practice. Photon conversion efficiencies greater than 90% have been reported in some pulsed laser experiments.

SRS frequency shifts and gains of some materials are shown in Table 2.

When pulses are used with durations comparable to or shorter than the response time of the material, ranging from picoseconds to femtoseconds depending on the material, the interaction is transient and the effects of the time derivatives in [6] and [7] must be accounted for. Under these conditions the Stokes intensity is given as a temporal integral with Bessel function dependence on the laser intensity. When the pulse duration is much shorter than the material dephasing time and the incident Stokes functional form is the same as the pump, the amplified Stokes intensity is given by the expression

$$I_S(z,\tau) = I_S(0,\tau)I_0^2\left(\sqrt{2g_{ss}\Gamma z\int_{-\infty}^{\tau} I_L(\tau')d\tau'}\right) \qquad [12]$$

where I_0 is a modified Bessel function. In this regime, the Stokes intensity depends only on the integrated pump energy. The Stokes gain is reduced from its steady state value for pulses of constant intensity, but for pulses of constant energy the highest Raman gain is obtained in the limit of transient interactions.

One of the most important applications of Raman scattering is the generation of new coherent waves at frequencies shifted from the pump. In this case, an incident Stokes wave is not supplied as the starting term in [9] or [12]. Rather, the process is initiated by the vacuum fluctuations in the Stokes and material oscillator fields. Quantum field creation and annihilation operators must replace the classical field amplitudes of [6] and [7]. The intensity of the generated Stokes wave is given by the expectation value of the normally ordered number operator and in the extreme transient regime is given by:

$$I_S(z,d\tau) = \frac{1}{2A}\hbar\omega_S\Gamma g_{ss}I_L(d\tau)z \times\left\{I_0^2\left(\sqrt{2\Gamma g_{ss}z\int_{-\infty}^{\tau} I_L(\tau')\tau'}\right) - I_1^2\left(\sqrt{2\Gamma g_{ss}z\int_{-\infty}^{\tau} I_L(\tau')\tau'}\right)\right\} \qquad [13]$$

where A is the cross-sectional area of the interaction geometry. The Stokes light generated in this fashion has all the characteristics of amplified spontaneous emission of lasers operated below threshold until the Stokes intensity becomes high enough to deplete the laser.

The stochastic nature of the SRS quantum mechanical source is manifest in the macroscopic Stokes light when the Stokes generator is operated well below pump depletion levels. For a single spatial mode the probability density distribution of

Table 2 List of frequency shifts and gains of selected materials for stimulated Raman scattering

Material, λ_L (nm)	*Shift* (cm^{-1})	*Gain (cm/GW)*
H_2 gas (20 atm), 532	4155	2.5
D_2 gas (60 atm), 532	2987	0.45
CH_4 (115 atm), 532	2913	1.26
H_2O, 694	3290	0.14
CS_2, 694	655.6	24
$LiNbO_3$, 694	637	9.4
$Ba(NO_3)_2$, 532	1047	47

Table compiled in parts from: Kaiser W and Maier M (1972) Stimulated Rayleigh, Brillouin and Raman Spectroscopy. In: Arecchi FT and Schulz-DuBois EO (eds) *Laser Handbook*, vol. 2. pp. 1077–1150. Amsterdam, The Netherlands: North Holland and Reintjes J and Bashkansky M (2001) Stimulated Raman and Brillouin Scattering. In: Bass M (ed.) *Optical Society of America Handbook of Optics*, Ch. 18, 2nd ed., vol. IV, New York: McGraw-Hill.

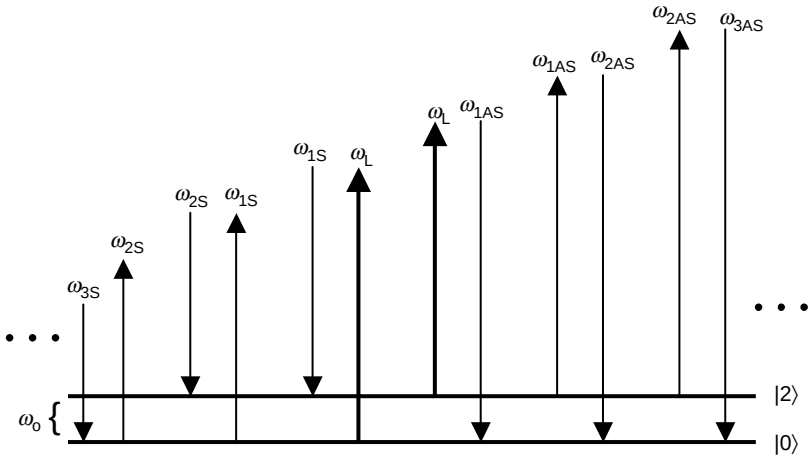

Figure 1 SRS level diagram showing generation of multiple Stokes and anti-Stokes frequencies.

the Stokes energy has the form of a negative exponential:

$$P(W) = \frac{1}{\langle W \rangle} \exp\left[-\frac{W}{\langle W \rangle} \right]$$

Statistical fluctuations in pulse energy, spatial profile, pointing, and in the spectral and temporal structure of the Stokes signal have all been observed.

If the intensity of the generated Stokes light becomes comparable to the pump intensity, it can serve as a pump wave for its own Raman interaction, producing a wave termed the second Stokes wave, shifted from the pump by frequency $2\omega_o$. If the process is driven hard enough third- and higher-order Stokes waves can be produced.

In addition to the interaction between the pump and Stokes beams described above, stimulated Raman scattering can also involve an anti-Stokes wave whose frequency is given by $\omega_{AS} = \omega_L + \omega_o$. The relationship of various Stokes and anti-Stokes waves is depicted in Figure 1. If a population inversion exists between the energy levels of the Raman process, then the anti-Stokes wave experiences exponential gain in a manner similar to that described above for the Stokes gain.

When all the population is in the ground state, which is usually the case, the anti-Stokes wave has exponential loss. The anti-Stokes wave can still grow without population inversion if the phase-matching condition

$$\Delta k = k_{AS} + k_S - 2k_L = 0$$

is approximately satisfied. Because of normal dispersion in the material this condition is usually not satisfied for collinear beams. However, if the Stokes and anti-Stokes beams propagate at small angles as shown in Figure 2, then the phase-matching condition can be met. For angles near but not exactly at phase matching the anti-Stokes wave grows exponentially as part of a mixed Stokes/anti-Stokes mode with a gain given by:

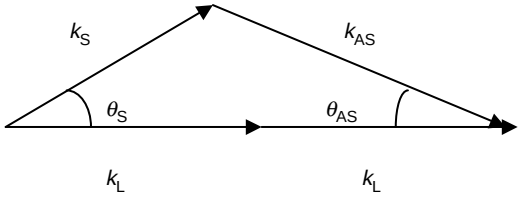

Figure 2 Phase matching diagram for Stokes–anti-Stokes coupling in a medium with positive dispersion.

$$g = \mathrm{Re}\left\{ \tfrac{1}{2}(K_3 - K_1)|A_L|^2 - \tfrac{i}{2}\sqrt{\Delta k^2 + 2i\Delta k(K_3 + K_1)|A_L|^2 - (K_1 - K_3)^2|A_L|^4} \right\} \quad [14]$$

At exact phase matching the interaction of Stokes and anti-Stokes waves serves to suppress the exponential gain of either wave.

As with Stokes waves, multiple-order anti-Stokes waves can be generated. Stokes and anti-Stokes orders up to at least 13 have been reported.

Stimulated Brillouin and Rayleigh Scattering

Light scattering can also occur through interaction with material density variations. Spontaneous

scattering through interaction with this type of internal mode produces a triplet of spectral lines. The up- and down-shifted lines are known as Brillouin anti-Stokes and Stokes lines, respectively, and are due to interaction with acoustic waves. The central peak is termed Rayleigh, or thermal Rayleigh scattering and is due to stationary density fluctuations arising from light absorption. A broad skirt around the central peak, which can be observed in anisotropic media, is due to the fluctuation of molecular distribution and orientation and is termed Rayleigh wing scattering. Each of these processes can give rise to stimulated scattering in the presence of intense radiation.

Density fluctuations are induced directly by light intensity through electrostriction, and indirectly by temperature fluctuations when optical absorption is important. A general classical model of the interaction, including both electrostriction and absorptive heating effects, is described by three equations:

The Navier-Stokes equation:

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} + \frac{v^2}{\delta}\nabla(\Delta\rho) + \frac{v^2\beta\rho_0}{\delta}\nabla(\Delta T) - \eta\nabla^2\mathbf{v} = 2\gamma\varepsilon_0\nabla(\mathbf{E}_{\mathrm{L}}\cdot\mathbf{E}^*_{\mathrm{R/B}}) - 2\varepsilon_0\left(\frac{\partial\varepsilon}{\partial T}\right)_\rho(\mathbf{E}_{\mathrm{L}}\cdot\mathbf{E}^*_{\mathrm{R/B}})\nabla(\Delta T) \quad [15]$$

The continuity equation:

$$\frac{\partial}{\partial t}\Delta\rho + \rho_0\nabla\cdot\mathbf{v} = 0 \quad [16]$$

The energy transport equation:

$$\left(\rho_0 C_v\frac{\partial}{\partial t} - \lambda_{\mathrm{T}}\nabla^2\right)\Delta T - \frac{C_v(\delta-1)}{\beta}\frac{\partial}{\partial t}\Delta\rho = 4nc\alpha\varepsilon_0(\mathbf{E}_{\mathrm{L}}\cdot\mathbf{E}^*_{\mathrm{R/B}}) + 4\varepsilon_0\left(\frac{\partial\varepsilon}{\partial T}\right)_\rho\left(T_0\mathbf{E}_{\mathrm{L}}\cdot\frac{\partial}{\partial t}\mathbf{E}^*_{\mathrm{R/B}}\right) \quad [17]$$

Here $\rho=\rho_0+\Delta\rho$, $T=T_0+\Delta T$, $\mathbf{v}$ is the volume element, v is the acoustic wave velocity, $\delta=C_p/C_v$ is the ratio of heat capacities at constant pressure and at constant volume, $\eta=(2\eta_{\mathrm{s}}+\eta_{\mathrm{d}})$ where η_{s} is the shear viscosity and η_{d} is the dilational viscosity, $\gamma=\rho_0(\partial\varepsilon/\partial\rho)_T$, λ_T and β are the thermal conductivity and thermal expansion coefficients, ε is the dielectric constant, and α is the linear absorption coefficient. $\mathbf{E}_{\mathrm{L}}$ is the pump field and $\mathbf{E}_{\mathrm{R/B}}$ is either the Rayleigh or Brillouin field.

Equations [15] and [16] can be combined to eliminate the $\mathbf{v}$ dependence:

$$\left[-\frac{\partial^2}{\partial t^2} + \left(\frac{v^2}{\delta} + \frac{\eta}{\rho_0}\frac{\partial}{\partial t}\right)\nabla^2\right]\Delta\rho + \frac{v^2\beta\rho_0}{\delta}\nabla^2(\Delta T) = 2\gamma\varepsilon_0\nabla^2(\mathbf{E}_{\mathrm{L}}\cdot\mathbf{E}^*_{\mathrm{R/B}}) - 2\varepsilon_0\left(\frac{\partial\varepsilon}{\partial T}\right)_\rho\nabla\cdot[(\mathbf{E}_{\mathrm{L}}\cdot\mathbf{E}^*_{\mathrm{R/B}})\nabla(\Delta T)] \quad [18]$$

Together with the equations describing the effect of the medium on the optical fields [19a,b], this coupled set of equations describes simultaneously stimulated Brillouin and Rayleigh scattering:

$$\left(\frac{\partial}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial}{\partial t} + \frac{1}{2}\alpha\right)A_{\mathrm{L}} = \mathrm{i}\frac{\omega_{\mathrm{L}}}{4cn_{\mathrm{L}}}\left[\frac{\gamma}{\rho_0}\Delta\rho A_{\mathrm{R/B}} + \left(\frac{\partial\varepsilon}{\partial T}\right)_\rho A_{\mathrm{R/B}}\Delta T\right] \quad [19a]$$

$$\left(\frac{\partial}{\partial z} - \frac{1}{v_{\mathrm{g}}}\frac{\partial}{\partial t} - \frac{1}{2}\alpha\right)A_{\mathrm{R/B}} = -\mathrm{i}\frac{\omega_{\mathrm{R/B}}}{4cn_{\mathrm{R/B}}}\left[\frac{\gamma}{\rho_0}\Delta\rho^* A_{\mathrm{L}} + \left(\frac{\partial\varepsilon}{\partial T}\right)_\rho A_{\mathrm{L}}\Delta T^*\right] \quad [19b]$$

where the slowly varying envelope approximation [2] is used for the waves described by A_{L}, $A_{\mathrm{R/B}}$, $\Delta\rho$, and ΔT. The nonlinear polarization term is related to density and temperature wave amplitudes in the following way:

$$P^{\mathrm{NL}} = \varepsilon_0 E\left[\left(\frac{\partial\varepsilon}{\partial\rho}\right)_T\rho + \left(\frac{\partial\varepsilon}{\partial T}\right)_\rho T\right] \quad [20]$$

Without the driving terms involving the optical fields, [17] and [18] describe broadband damped fluctuations in the variables $\Delta\rho$ and ΔT. When the optical driving terms are included, resonant behavior emerges. Steady state solutions can be obtained assuming strongly damped material excitations, allowing neglect of $\Delta\rho$ and ΔT derivatives. Separate contributions of the Rayleigh and Brillouin interactions to $\Delta\rho$ can be identified:

Rayleigh contribution:

$$\Delta\rho_{\mathrm{R}} = \left[\frac{\gamma(\Omega + \frac{1}{2}\mathrm{i}\delta\Gamma_{\mathrm{R}}) - \mathrm{i}\gamma_{\mathrm{a}}\Omega_{\mathrm{B}}}{\Omega + \frac{1}{2}\mathrm{i}\Gamma_{\mathrm{R}}}\right]\frac{\varepsilon_0}{4v^2}A_{\mathrm{L}}A^*_{\mathrm{R}} \quad [21]$$

where γ_{a} is the absorptive coupling constant given by

$$\gamma_{\mathrm{a}} = \frac{2\alpha n v^2\beta}{C_p\Omega_{\mathrm{B}}}$$

The Brillouin contribution is:

$$\Delta\rho_B = \left[\frac{-(\gamma - i\gamma_a v q/\Omega)q^2\varepsilon_0}{4(\Omega^2 + i\Omega\Gamma_B - v^2q^2)}\right]A_L A_B^* \qquad [22]$$

In [21] and [22] the frequencies and wavevectors satisfy the following relations:

$$\omega_L = \omega_{R/B} + \Omega \qquad [23a]$$

$$k_L = k_{R/B} + q \qquad [23b]$$

where ω_L is the incident laser frequency, $\omega_{R/B}$ is the frequency of the scattered Rayleigh or Brillouin wave, and k_L, $k_{R/B}$ are the corresponding wavevectors. Ω and q are the frequency and wavevector of the density fluctuation. Equations [23a,b] can also be described as energy and momentum conservation relations among incident and scattered photons and a phonon in the quantum mechanical picture.

The Rayleigh and Brillouin contributions exhibit resonant behavior at different frequency shifts. The Rayleigh contribution has its resonant frequency at $\Omega_R = 0$, resulting in typical diffusion behavior. The Rayleigh linewidth $\Gamma_R = 2\lambda_T q^2/\rho_0 C_p$ is the inverse of the decay time of the isobaric density fluctuations. The Brillouin contribution has its resonance at a frequency shift given by

$$\Omega_B = qv$$

where v is the sound velocity. Thus at resonance the density fluctuations associated with the Brillouin effect correspond to sound waves. The Brillouin linewidth

$$\Gamma_B = \frac{q^2}{\rho_0}\left[\eta + \frac{\kappa}{C_p}\left(\frac{C_p}{C_v} - 1\right)\right]$$

is the inverse of the phonon lifetime τ_B.

In general Brillouin and Rayleigh interactions involve noncollinear scattering with the outgoing wave propagating at an arbitrary angle to the incident wave. In this case [23b] is a vector relation and the interaction geometry can be represented as shown in Figure 3.

The resulting relations among the k-vectors and frequencies of the various waves can place restrictions on the values of frequency shifts that are observed under various circumstances.

The coupled equations for the optical intensities in the steady state take the following form:

$$\frac{dI_L}{dz} = -g_{R/B}I_L I_{R/B} - \alpha I_L \qquad [24a]$$

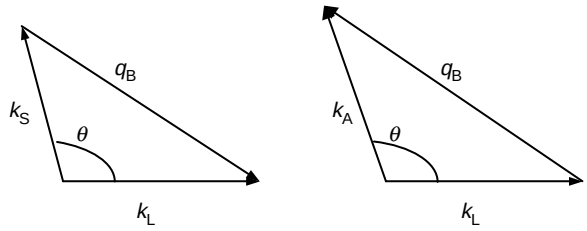

Figure 3 SBS phase matching diagram depicting relations between laser, k_L, sound wave, q_B, and Stokes and anti-Stokes, k_S, k_A k-vectors.

$$\frac{dI_{R/B}}{d\zeta} = g_{R/B}I_L I_{R/B} - \alpha I_{R/B} \qquad [24b]$$

where z is the direction of propagation of the incoming wave, ζ is the direction of propagation of the scattered wave, and $g_{R/B}$ is the gain coefficient of the Rayleigh or Brillouin effect. Equations [24a,b] describe interactions that have exponential gain. Stimulated Brillouin and Rayleigh scattering can be used in both a self-generator and an amplifier mode, but are used most often in a self-generator mode. Because the frequency shifts are small in these interactions, the starting signals for self-generators are thermal excitations of the internal modes rather than vacuum fluctuations.

Stimulated Brillouin Scattering

Stimulated Brillouin scattering (SBS) occurs through interaction with acoustic waves in solids, liquids, and gases and with ion-acoustic waves in plasmas. The Stokes or anti-Stokes frequency shifts for SBS are much smaller than for SRS, with typical values on the order of 0.1–100 GHz, depending on the excitation wavelength and interaction geometry as well as on material properties. Because the internal mode in SBS is a propagating wave, the phase matching condition shown in Figure 3 places restrictions on the frequency of the sound wave that is involved at various scattering angles. SBS is used most often for pulse compression and for correction of phase distortion in propagating light beams through phase conjugation. Although SBS is normally associated with high-power lasers in laboratory experiments, it is also a major factor in limiting the intensity of low-power lasers that can be used in optical fibers because of the long interaction lengths. SBS is also a prominent source of damage in solids and crystals caused by high-power lasers. It can also be used for generation and study of coherent hypersonic waves.

The SBS gain coefficient has both electrostrictive and absorptive contributions, $g_B = g_B^e + g_B^a$ where

$$g_B^e = \frac{\omega_B \gamma^2 \Omega_B}{2 n_B n_L c^2 v^2 \rho_0 \Gamma_B} \frac{1}{1 + (2(\Omega - \Omega_B)/\Gamma_B)^2} \quad [25]$$

and

$$g_B^a = \frac{\omega_B \gamma \gamma_a \Omega_B}{4 n_B n_L \rho_0 v^2 c^2 \Gamma_B} \frac{4(\Omega - \Omega_B)/\Gamma_B}{1 + (2(\Omega - \Omega_B)/\Gamma_B)^2} \quad [26]$$

The SBS gain from electrostriction has a maximum value of

$$g_B^e = \frac{\omega_B \gamma^2 \Omega_B}{2 n_B n_L c^2 v^2 \rho_0 \Gamma_B} \quad [27]$$

at the resonant frequency shift

$$\Omega = \Omega_B = qv \quad [28]$$

For this condition the interaction is with a sound wave in the material. The frequency of the sound wave, and the gain, depend on the angle between the incident and scattered light. Using the vector relation

$$\mathbf{k}_B = \mathbf{k}_L - \mathbf{q}$$

depicted in Figure 3 and the approximations $\Omega_B << \omega_B, \omega_L, |k_B| \approx |k_L|$ we obtain

$$q = 2|k_L| \sin(\theta/2) \quad [29]$$

The Brillouin frequency shift is then given by

$$\Omega_B = 2 n_L \omega_L (v/c) \sin(\theta/2) \quad [30]$$

and the Brillouin linewidth is given by

$$\Gamma_B = \frac{4}{\rho_0} \left[\eta + \frac{\kappa}{C_p} \left(\frac{C_p}{C_v} - 1 \right) \right] \frac{\omega_L^2 n_L^2}{c^2} \sin^2(\theta/2)$$
$$= \Gamma'' \sin^2(\theta/2) \quad [31]$$

The peak Brillouin electrostrictive gain is then given by

$$g_B^e = \frac{\omega_B \gamma^2 \omega_L}{n_B c^3 v \rho_0 \Gamma_B} \sin(\theta/2) \quad [32a]$$

$$g_B^e = \frac{\omega^2 \gamma^2}{n_B c^3 v \rho_0 \Gamma'' \sin(\theta/2)} \quad [32b]$$

where we have used the approximation $\omega_B \approx \omega_L = \omega$.

The maximum electrostrictive SBS gain is independent of the laser frequency because the Brillouin linewidth is proportional to the square of the laser frequency.

As a function of angle the maximum Brillouin gain is usually observed in the backward direction. This is consistent with the form of the gain coefficient given in [32a], which indicates that the gain is proportional to $\sin(\theta/2)$. However, when the angular variation of the Brillouin linewidth is also accounted for, the gain coefficient is actually minimum in the backward direction [32b]. Typical laboratory experiments involve pencil-like geometries in which the interaction length is significant only in the backward or forward directions. In these situations, the total gain is actually maximum in the backward direction, where the interaction length is longest. Although the formulas appear to indicate that the gain goes to infinity in the forward direction, the damping time, which is proportional to Γ_B^{-1}, also goes to infinity and the interaction becomes transient. Under these conditions the gain parameter is proportional to $\sin(\theta/2)$ and the Brillouin gain goes to zero in the forward direction. Under some geometries Brillouin scattering at 90° to the laser can also be significant. It can be a source of damage to large glass components in high-power lasers.

For thermal Brillouin scattering the gain has a dispersive shape centered on the frequency shift $\Omega = \Omega_B$, with loss occurring on the low-frequency side and gain occurring on the high-frequency side, peaking at a frequency $\Omega = \Omega_B + \Gamma_B/2$. The maximum absorptive SBS gain depends on the laser frequency as ω^{-1}.

Table 3 shows representative values of the Brillouin linewidth and gain coefficients for a selection of substances.

Brillouin-enhanced four-wave mixing is a related effect that involves both a downshifted wave and an upshifted wave in which the coherent interaction is enhanced by the resonance associated with the sound wave. Gains of the order of 10^5 can be realized. Systems based on this effect are often used for phase conjugation.

Stimulated Rayleigh Scattering

Stimulated thermal Rayleigh scattering (R) can be observed in liquids while Stimulated Rayleigh wing scattering (SRWS) is observed in liquids with anisotropic molecules. Additionally, stimulated Rayleigh scattering due to fluctuations of concentration in gas mixtures can also occur. For R the internal mode obeys a diffusion equation. The R gain is given by

$$g_R = g_R^e + g_R^a$$

Table 3 Properties of stimulated Brillouin scattering for selected materials

Substance	*Laser wavelength (nm)*	*Frequency shift (Ghz)*	$\Delta\nu$ *(MHz)*	τ_B (ns)	g_B^e *(cm/GW)*	g_B^a/α (cm/GW)*
Liquid						
Acetone	532	5.93	361	0.44	12.9	22
Benzene	532	8.33	515	0.31	12.3	24
CS_2	532	7.7	120	1.9	130	20
CCl_4	532	5.72	890	0.18	8.77	13
Chloroform	532	5.75	635	0.25	11.7	
Ethanol	532	5.91	546	0.29	12*	10
Methanol	532	5.47	325	0.49	10.6	13
Water	532	7.4	607	0.26	2.94	0.8
Gas						
Xenon (7599 torr)	532	0.654 ± 0.024	98.1 ± 8.9	$0.65\lambda_P^2P$	1.38 ± 0.19	
SF_6(20 bar)	1320	0.2		35		
N_2(100 bar)	1320	0.5		30		
Solid						
BK 7	532	34.65 ± 0.039	165.0 ± 8.6		2.15 ± 0.21	
CaF_2	532	37.164 ± 1.185	45.6 ± 8.8		4.11 ± 0.65	
Plexiglas	532	15.687 ± 0.036	253.7 ± 12.6			
SiO_2	488	35.6	156		4.482	

*Absorptive maximum gain SBS values at 694 nm, with permission from Boyd RW (1992) *Nonlinear Optics*. Academic Press: New York. Reproduced from Reintjes J and Bashkansky M (2001) Stimulated Raman and Brillouin scattering. In: Bass M (ed.) *Optical Society of America Handbook of Optics*, Ch. 18, 2nd ed., vol. IV. New York: McGraw-Hill.

where g_R^e and g_R^a are the electrostrictive and absorptive contributions given by

$$g_R^e = \frac{\omega\gamma^2(\delta-1)}{4\rho_0 n_L n_R v^2 c^2}\frac{4\Omega/\Gamma_R}{1+(2\Omega/\Gamma_R)^2} \qquad [33]$$

$$g_R^a = \frac{-\omega\gamma\gamma_a\Omega_B}{2\rho_0 n^2 v^2 c^2 \Gamma_R}\frac{4\Omega/\Gamma_R}{1+(2\Omega/\Gamma_R)^2} \qquad [34]$$

Both contributions to the STRS gain have dispersive forms. The electrostrictive contribution is maximum for a down-shifted wave with $\Omega = \frac{1}{2}\Gamma_R$. Its maximum value scales linearly with laser frequency. For the absorptive process the gain depends inversely on the laser frequency, and is maximum for an upshifted wave with $\Omega = -\frac{1}{2}\Gamma_R$. Table 4 shows representative values of the Rayleigh linewidth and gain coefficients for a selection of substances.

Table 4 Properties of stimulated Rayleigh scattering for some materials at a wavelength of 694 nm

Substance	*Calculated maximum steady state gain factor*		*Linewidth* $\delta\nu_R$ *(MHz)*
	g_R^e*(max) (cm/MW)*	g_R*(max)/*α *(cm²/MW)*	
CCl_4	2.6×10^{-4}	0.82	17
Methanol	8.4×10^{-4}	0.32	20
CS_2	6.0×10^{-4}	0.62	36
Benzene	2.2×10^{-4}	0.57	24
Aceton	2.0×10^{-4}	0.47	21
H_2O	0.02×10^{-4}	0.019	27.5
Ethanol		0.38	18

Reproduced from Kaiser W and Maier M (1972) Stimulated Rayleigh, Brillouin and Raman Spectroscopy. In: Arecchi FT and Schulz-DuBois EO (eds) *Laser Handbook*, vol. 2, pp. 1077–1150. Amsterdam, The Netherlands: North Holland.

Parametric Fluorescence

Some types of stimulated interactions do not involve exchange of energy with the material. These include parametric down-conversion and self-focusing. In parametric down-conversion an incident wave at frequency ω_L scatters into two longer-frequency light waves at signal and idler frequencies ω_s and ω_i. They satisfy the frequency and wavevector relations

$$\omega_L = \omega_s + \omega_i$$
$$k_L = k_s + k_i$$

The equations for the waves are:

$$\frac{\partial}{\partial z}A_s = -i\left(\frac{\omega_s}{2n_s c}\right)\chi^{(2)}A_L A_i^*$$

$$\frac{\partial}{\partial z}A_i = -i\frac{\omega_i}{2n_i c}\chi^{(2)}A_L A_s^*$$

where A_s and A_i are called the signal and idler fields. The solution for the signal field takes the

following form:

$$A_s(z) = A_s(0)\cosh\left(\sqrt{|\kappa|^2|A_L|^2}z\right) - i\sqrt{\frac{\omega_s n_i}{\omega_i n_s}}A_i^*(0)\sinh\left(\sqrt{|\kappa|^2|A_L|^2}z\right) \quad [35]$$

where

$$|\kappa|^2 = (\omega_s\omega_i/4n_s n_i c^2)|\chi^{(2)}|^2$$

The stimulated parametric interaction provides gain for input signal and idler waves. For large values of $\sqrt{|\kappa|^2 A_L|^2}\ z$ the gain becomes exponential in character, similar to the gain in SRS and SBS. When no incident signals are supplied, the source terms in [35] are the vacuum fluctuations at the signal and idler frequencies. Stimulated parametric fluorescence is the basis for parametric oscillators that provide tunable coherent radiation over a wide wavelength range in the visible and infrared.

Stimulated Photorefractive Scattering

Stimulated light scattering also takes place in photorefractive materials. These materials include crystals such as $LiNbO_3$, $BaTiO_3$, $Bi_{12}SiO_{20}$, and others. They possess a unique property of photo-induced index change with very low light intensities. Some of the potential applications include real-time holography, information recording, phase conjugation, parallel signal processing, and ultrafast optical gating. The nonlinearities leading to stimulated scattering are very large and can produce a significant gain in millimeter thick crystals. In order for amplification of scattered light to take place the nonlinearities have to have a nonlocal character due to diffusion or drift fields. Figure 4 shows a two-beam interaction effect in a photorefractive medium leading to stimulated scattering. Solid lines represent intensity maxima of interfering light beams while dashed lines represent diffused index change maxima. The fact that the two do not overlap allows amplification of the signal beam, with the maximum gain occurring at $\pi/2$ phase shift.

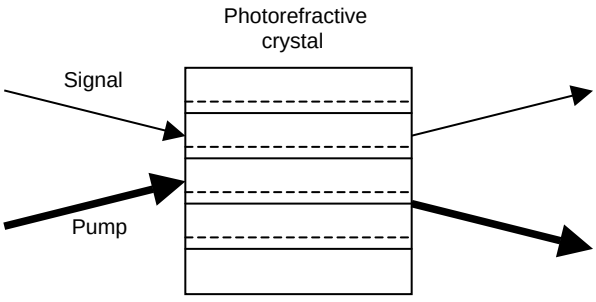

Figure 4 Stimulated photorefractive scattering diagram showing incident pump and signal beams, amplified signal beam, and an index grating formed in the photorefractive crystal.

Stimulated Compton and Thomson Scattering

Light scattering from free electrons is described by Thomson scattering in the low-energy limit, with a cross-section of $(8\pi)/(3)((e^2)/(4\pi\varepsilon_0 mc^2)^2)$, and by Compton scattering in the relativistic limit. Stimulated Thomson or Compton scattering is the source of light emitted from free electron lasers. In these interactions, the electron scatters incoming radiation in its rest frame to a backscattered wave. The ponderomotive force resulting from the interaction with the scattered waves increases the periodic electron bunching, which in turn increases the efficiency of scattering. In some cases, the incoming electromagnetic wave is provided by a microwave source, while in others it arises from a static magnetic wiggler, which appears as an incident electromagnetic wave in the electron rest frame because of a relativistic Lorentz transformation. Free electron lasers are a source of tunable coherent radiation from the microwave region to the ultraviolet.

Self-focusing

In the presence of a strong laser field with intensity I, the refractive index of many materials changes according to the relation

$$n = n_0 + n_2 I$$

where n_2 is termed the nonlinear index. This change in refractive index gives rise to several effects, depending on the temporal and spatial properties of the light. In one type of interaction, a strong forward beam couples to two weaker beams that propagate at small forward angles, providing gain for the weaker beams. One of the weak beams interferes with the strong forward beam to form a phase grating through the nonlinear index, providing gain for the second weak beam. The interference of the second weak beam with the forward beam forms a second phase grating, providing gain for the first weak beam, thereby completing the coupling

for the stimulated process. This effect is similar to Stoke/anti-Stokes Raman scattering, except that the coupling is between waves propagating in different directions rather than between waves with different frequencies.

For interactions with an instantaneous relaxation time, the gain for waves propagating at a small angle to the forward direction is given by

$$g = \frac{K_\perp}{2k}\sqrt{K_{\max}^2 - K_\perp^2}$$

where $K_\perp = \sqrt{k_x^2 + k_x^2}$ is the transverse wavevector, corresponding to the angle

$$\theta = \frac{K_\perp}{k}$$

and

$$K_{\max} = 4k^2 n^{\mathrm{L}} n_2 I$$

is the largest transverse k-vector for which gain is observed. The maximum gain is given by

$$g_{\max} = \frac{K_{\max}^2}{4k}$$

and occurs for an angle corresponding to

$$K_\perp = \frac{K_{\max}}{\sqrt{2}}$$

When the physical interaction that causes the refractive index to change with light intensity has a nonzero relaxation time, gain can be observed at angles larger than that given by $K_{\max}$.

This stimulated four-photon effect can be interpreted as the initiating stage of self-focusing. In the self-focusing effect, a laser beam with a normal mode profile (more intense in the center than at the edges) creates a positive lens in the material, focusing the beam. As the beam intensifies, the power of the lens increases, causing the beam to focus more tightly and eventually coming to a catastrophic focus if other nonlinear effects do not limit the intensity. The four-photon stimulated effect described above results in the spread of k-vectors of the forward beam, corresponding to a reduction in its diameter.

Self-focusing is responsible for breakdown and damage in many solids. Self-focusing in high-power lasers is commonly suppressed by restricting transverse k-vectors to values well below those of maximum gain.

See also

Nonlinear Optics, Applications: Self-Focusing and Related Effects (Solitons and Multiphoton Absorption). **Nonlinear Optics, Basics:** Photorefraction. **Scattering:** Scattering Theory.

Further Reading

Bloembergen N (1965) *Nonlinear Optics*. New York: Benjamin.

Bloembergen N (1967) The stimulated Raman effect. *American Journal of Physics* 35: 989–1023.

Boyd RW (1992) *Nonlinear Optics*. New York: Academic Press.

Fabellinskii IL (1968) *Molecular Scattering of Light*. New York: Plenum.

Gunter P and Huignard J-P (eds) (1989) *Photorefractive Materials and their Applications II*. Springer-Verlag Topics in Applied Physics, vol. 62. Berlin: Springer-Verlag.

Kaiser W and Maier M (1972) Stimulated Rayleigh, Brillouin and Raman spectroscopy. In: Arecchi FT and Schulz-DuBois EO (eds) *Laser Handbook*, vol. 2, pp. 1077–1150. Amsterdam, The Netherlands: North Holland.

Milanovich F (1986) Stimulated Raman scattering. In: Weber MJ (ed.) *Handbook of Laser Science and Technology. Optical Materials*, vol. III. Boca Raton, FL: CRC Press.

Penzkofer A, Laubereau A and Kaiser W (1979) High intensity Raman interactions. *Progress in Quantum Electronics* 6: 55–140.

Rabin H and Tang CL (eds) (1975) *Quantum Electronics: A Treatise*. Part A, vol. 1. New York: Academic Press.

Raymer MG and Walmsley IA (1990) The quantum coherence properties of stimulated Raman scattering. In: Wolf E (ed.) *Progress in Optics*, vol. XXVIII, p. 183. Amsterdam, The Netherlands: Elsevier.

Reintjes JF (1984) *Nonlinear Optical Parametric Processes in Liquids and Gases*. New York: Academic Press.

Reintjes JF (1995) Stimulated Raman and Brillouin scattering. Handbook of laser science and technology, supplement 2. In: Weber MJ (ed.) *Optical Materials*, pp. 334–364. Boca Raton, FL: CRC Press.

Reintjes J and Bashkansky M (2001) Stimulated Raman and Brillouin scattering. In: Bass M (ed.) *Optical Society of America Handbook of Optics*, Ch. 18, 2nd ed., vol. IV. New York: McGraw-Hill.

Shen YR (1984) *Principles of Nonlinear Optics*. New York: Wiley.

Shen YR and Bloembergen N (1965) Theory of stimulated Brillouin and Raman scattering. *Physical Review* 137: A1787.

SEMICONDUCTOR MATERIALS

Contents

Amorphous Semiconductors
Band Structure Engineering
Dilute Magnetic Semiconductors
GaAs Based Compounds
Group IV Semiconductors, Si/SiGe
III-Nitrides
Large Gap II–VI Semiconductors
Lead Salts
Mercury Cadmium Telluride
Modulation Spectroscopy of Semiconductors and Semiconductor Microstructures
Quantum Dots
Type-II Quantum Wells and Superlattices

Amorphous Semiconductors

J I B Wilson, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Amorphous semiconductors comprise several classes of materials having a range of useful properties that have made many of them commercially important. Their special characteristics are derived from the particular techniques used in their synthesis, and include the ability of some to change their structure under the influence of heat or illumination. Many are photoconductive, making them valuable for photodetectors when large area devices are required (e.g., xerography), or photovoltaic, hydrogenated amorphous silicon being widely used in thin-film solar cells.

'Amorphous' means noncrystalline, thus encompassing glassy materials and disordered covalent-bonded materials. Glasses, such as the amorphous chalcogenides (glassy sulfides, selenides, and tellurides), melt to viscous liquids when heated, whereas disordered semiconductors like amorphous silicon are metastable thin films that vaporize or crystallize when heated. Two other groups of glassy semiconductors, the covalent-bonded II–IV–V_2 compounds and ionic-bonded oxides (e.g., $V_2O_5P_2O_5$) will not be discussed here. Amorphous organic semiconductors are receiving much interest now, because of their potential use in low-cost display devices, but are more logically treated as polymer materials.

The following sections explain the main preparation methods and the structures that result, the importance of defects and the influence these have on optical and electrical behavior, and some of the most important applications.

Amorphous Structure

The common property that distinguishes all amorphous materials from crystalline ones is the absence of long-range order in their atomic structure. These disordered solids have no periodic or translational symmetry. The term is sometimes loosely applied, and it should be noted that elements and compounds that are able to form several structures might contain different phases. An example is 'hard carbon' which in thin-film form may contain tetrahedral amorphous carbon and hydrogenated amorphous carbon, as well as microcrystallites of graphite and diamond. Interestingly, the prediction in 1989 by A.Y. Liu and M.L. Cohen, that crystalline cubic-C_3N_4 would be even harder than diamond, has not been confirmed by experiment. Only a hard amorphous C_xN_y has been produced, even by nitrogen ion beam assisted sputtering of graphite, which it was thought would

have provided sufficient energy to form the close C–N bonds.

When the common crystalline semiconductors, Si and Ge, melt they become liquid metallic conductors. This is because the solid is formed by directional covalent bonds that produce tetrahedral packing of the atoms. In the liquid the bonds are weakened and a closer packed structure found in metals results. However a less ordered atomic arrangement results from condensing vaporized Si or Ge on to a cold substrate. In this case the atoms cannot move into their preferred positions for either tetrahedral or close packing arrays and instead are frozen almost wherever they land. Nonetheless, the arrangement of atoms is closer to that of the tetrahedral crystalline semiconductor than that of the metallic liquid. The radial distribution function of an amorphous semiconductor provides the numbers of atoms versus distance from an arbitrary atom. It shows that the nearest-neighbor distance is the same as in the equivalent crystalline material, although with some spread of values, but that the separation of atoms further away has a less well-defined value. Thus pure amorphous Si and Ge retain the directional covalent bonds between nearest neighbors but have distorted bond angles that make next nearest neighbors lie in different positions from the equivalent crystalline solid. Consequently there is only short-range order and the atoms form a continuous random network, a concept introduced by W.H. Zachariasen in 1932 for glasses. The deposition conditions generally lead to further disruptive effects that will incorporate small voids where the distortion is too large to allow a bond between atoms. These voids together with broken bonds give a defective network that is far from the ideal random tetrahedral 'lattice'.

Amorphous Semiconductor Energy Bands

A frequently asked question is 'Why can amorphous materials show semiconducting behavior?' since they do not have the periodic structure that leads to the allowed energy band scheme in crystalline semiconductors. However, note that metals do not become insulators when they melt, demonstrating that electrical conductivity does not require a periodic atomic structure. In fact, electron energy levels in solids may also be derived by considering chemical bonds when discrete atoms are brought closer together. The valence band in a semiconductor arises from the electron bonding orbitals and the conduction band from the antibonding electron orbitals. The forbidden energy gap (bandgap) is the separation between these two bands, which are only partially filled in semiconductors at room temperature. Thus amorphous semiconductors will have bonding and antibonding energy bands, and a more or less well-defined bandgap according to the defect content.

In fact, pure amorphous Si and Ge are extremely poor semiconductors due to an ill-defined bandgap and the large concentration of defects. Distorted bonds produce allowed electron energy levels that lie adjacent to the well-defined bands in crystals, producing tails of energy levels extending into the bandgap. Electrons that take up these 'tail state' energies are not entirely free to move through the solid but can only hop from one defective site to another. Additional defects arise from unsatisfied covalent bonds that 'dangle' in the spaces between the atoms. These produce electron energy levels within the forbidden gap (Figure 1). Indeed, their concentration is so high in amorphous silicon produced from a condensed silicon vapor, that they fix the position of the Fermi level at approximately the middle of the bandgap and the semiconductor has a high resistivity. The usual option of reducing a semiconductor's resistivity by adding small amounts of dopants from adjacent columns of the periodic table has little effect in this case. A donor atom from group V (e.g., phosphorus) may have all five valence electrons satisfied by a locally distorted arrangement of silicon atoms, instead of having only four bonded electrons and a loosely bound electron as in crystalline silicon. A similar picture exists for an acceptor atom from group III (e.g., boron).

Semiconductors containing group VI elements, S, Se, or Te, have two-fold coordination instead of the tetrahedral arrangement of the group IV elements, Si and Ge. The bonding between atoms in the solid is formed from four of the six p-electrons, which give

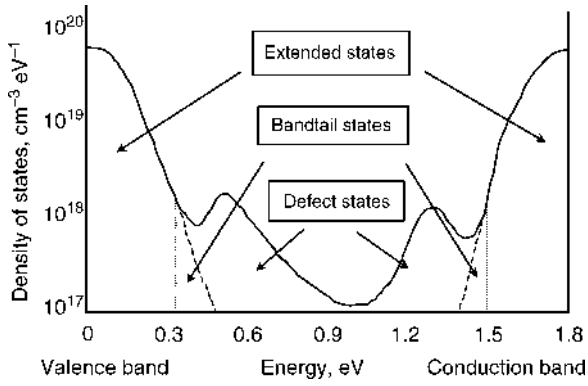

Figure 1 Schematic density of states for hydrogenated amorphous silicon.

rise to the bonding and antibonding energy bands of the solid, and the two remaining p-electrons from each atom form a non-bonding 'lone pair'. The lone pair energy band lies between the bonding and antibonding bands, and is the highest filled band in the solid, thus being the valence band.

Just as in amorphous Si or Ge, dangling bond defects will have energies between the valence and conduction bands, but in lone-pair semiconductors such disturbed energy states will now produce energy levels that fall within the lone pair band instead of within the bandgap. This makes it difficult to dope chalcogenide semiconductors, especially by donors, since the additional energy levels often lie within an existing energy band. Similar considerations apply to ternary chalcogenide alloys: these will have lone-pair bands arising from the group VI element, with additional energy levels from the other components (e.g., from group IV or V). However, amorphous chalcogenides will also contain three-fold coordinated chalcogenide atoms and these defects can associate in pairs or may interact with an adjacent two-fold coordinated atom. Defect centers with two oppositely charged chalcogens may form a 'valence alternation pair' (VAP).

So it may be seen that the semiconducting behavior of solids arises from the nearest-neighbor interactions of the atoms giving rise to allowed electron energy bands, and that long-range order is not necessary to produce the semiconductor's forbidden energy gap. Defects then produce localized energy levels that usually lie within the forbidden gap.

Preparation of Amorphous Silicon

Despite the early use of selenium as a photoconductive semiconductor in xerography, most amorphous semiconductors might have stayed as curious but technologically impractical materials, if an alternative preparation method to physical vapor deposition had not been discovered.

In the 1950s the preparation of high-purity single-crystal silicon was not yet established, and work at Standard Telecommunication Laboratories, UK, was attempting to use radio-frequency heating to produce silicon from low-pressure silane gas (SiH_4). It was noticed that a side effect was a deposit of resistive, noncrystalline material on cool parts of the container, close to unwanted plasma 'glow discharges'. In following up this observation, the technique was developed to produce a variety of other useful compounds from electrical discharges in appropriate gas mixtures, particularly silicon nitrides and oxides. With new scientific interest in amorphous semiconductors, the silicon coatings from silane were examined in more detail, and were shown to be significantly different from amorphous silicon produced by evaporating and condensing silicon. This new form was very photoconductive and could be doped n-type by adding phosphine to the silane. A more detailed study of doping led to the seminal papers by W.E. Spear and P.G. LeComber of Dundee University, UK, and to an explosive growth in the whole subject, once the potential commercial applications became apparent.

Despite many alternative synthesis methods, plasma chemical vapor deposition (CVD) remains the most widely used, and produces material that has not been surpassed in quality. The key features are a vacuum chamber into which a controlled flow of each of the gaseous sources is fed, and a heated plate on which the substrate is placed for coating. The decomposition of the gases is driven by an electrical discharge between two electrodes, which enables the gas temperature to be much lower than in conventional pyrolysis (thermal CVD). An important aspect of the process is the incorporation of hydrogen into the deposited material, which is thus more accurately known as hydrogenated amorphous silicon, a-Si:H. In the best material, hydrogen is mainly bonded singly as Si–H, but if the preparative conditions are not optimum then there may also be large amounts of $Si–H_2$ (as may be seen from the infrared absorption spectra). Alternative source gases usually contain hydrogen (e.g., Si_2H_6), and hydrogen may be added as a diluting agent for the doping gases (e.g., B_2H_6 or PH_3). Other precursors include fluorinated gases, but these have failed to give the predicted improvements in performance of the semiconductor in such applications as solar cells, and have greater chemical reactivity.

Since silane is spontaneously flammable in air, and the dopant gases are not only flammable but also very toxic, efforts have been made to produce a-Si:H by other means. Sputtering from undoped or doped silicon targets has been a popular choice, but it is still necessary to incorporate hydrogen into the deposited material, for instance by adding it to the usual argon sputter gas. This is because hydrogen serves to passivate unsatisfied 'dangling' Si bonds and to reduce the stresses that would occur in a pure silicon network. Photo-CVD, using ultraviolet light to drive the reaction without raising the temperature, can produce high-quality material but has the problem of unwanted silicon deposits on the illuminated reaction chamber window. Although conventional silane pyrolysis is not

suitable for producing a-Si:H, because the temperatures needed for useful deposition rates also remove much of the hydrogen if the substrate is too hot, it is possible to heat the gas separately from the substrate by means of a hot filament and so produce good-quality films.

Other hydrogenated silicon alloy semiconductors are synthesized by additions to silane of ammonia (giving amorphous silicon nitride, a-Si:N:H), or a hydrocarbon gas such as methane (giving a-Si:C:H), or germane (giving a-Si:Ge:H).

Preparation of Amorphous Chalcogenides

Unlike the tetrahedral group IV elements, useful members of this class of amorphous semiconductors may be produced by vacuum evaporation and condensation on to a relatively cool substrate. Compounds, such as As_2Se_3, do not evaporate congruently and so the deposit will change composition as the evaporation proceeds. More uniform compositions may be obtained by sputtering with argon, and glassy oxides may also be obtained in this way with oxygen added to the inert gas. Chalcogenide glasses may also be prepared by rapid cooling from a melt, or by chemical vapor deposition from heated vapor precursors. Since glasses are thermodynamically unstable, they will tend to transform to a more stable, sometimes crystalline structure. The rate for this process depends on the temperature and may be effectively too slow to observe.

Optical Properties of Amorphous Semiconductors

Optical absorption spectra of amorphous semiconductors clearly show a forbidden energy gap. By plotting a suitable function of the absorption coefficient, α, versus photon energy, $h\nu$, usually that given by J. Tauc, $(\alpha h\nu)^{1/2}$, the 'optical gap' may be determined from the intercept of the linear portion of the plot with the x-axis. Unlike crystalline semiconductors, amorphous ones have no k-selection rules to give indirect and direct transitions across the bandgap, thus a-Si:H behaves as a direct gap material. This 'optical gap' is greater than the bandgap determined from electrical conductivity measurements, since the latter will have contributions from both localized bandtail states (low electrical mobility) and extended states (high electrical mobility) (Figure 2).

The absorption spectra also have a portion at lower photon energies having an exponential dependence on photon energy, known as the Urbach tail. This is related to the structural disorder in the material and the tail states. Transitions involving defect bandgap states lie at even lower photon energy, and the density of defects can be determined from the weak absorption in this region of the spectrum, perhaps by using photothermal deflection spectroscopy to detect the weak optical absorption.

Photoluminescence also provides a probe of the defect content of many materials. Interpretation and identification of the actual defect (e.g., chemical impurity or structural imperfection) is aided by

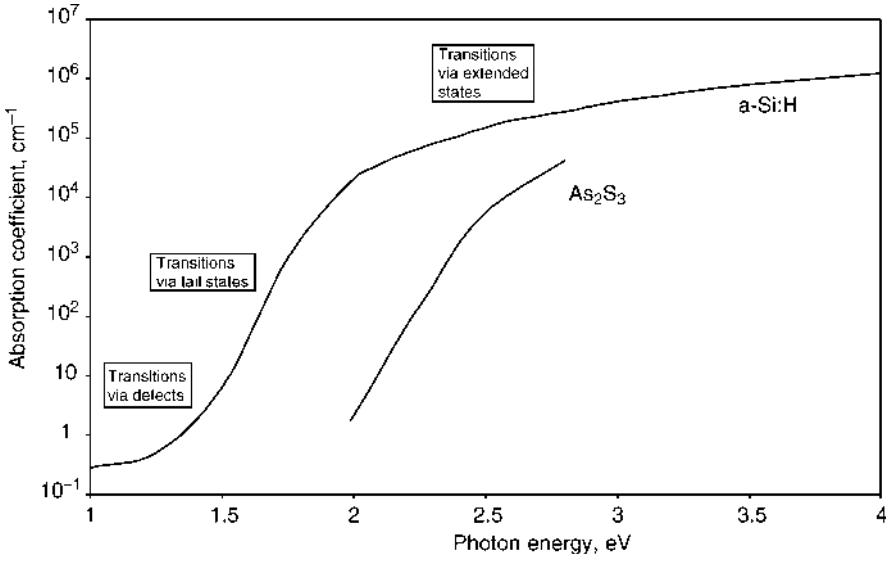

Figure 2 Typical optical absorption spectra of hydrogenated amorphous silicon and arsenic sulfide.

reference to the crystalline form. Both crystalline and amorphous forms of chalcogenide have a wealth of special effects, including reversible photoinduced bleaching and darkening, and laser-induced crystallization and amorphization. In addition the creation and dispersal of defects by appropriate illumination of some amorphous chalcogenides is accompanied by changes in the electron spin resonance (ESR) signal. Some of these effects have been the basis for applications in optical memories. More prosaically, chalcogenide glasses are important infrared transparent materials and amorphous chalcogenide thin films have had historical importance as the photoconductive layer in xerography, although here they are now being supplanted by less toxic amorphous silicon alloys.

Electrical Properties of Amorphous Semiconductors

The electrical conductivity of amorphous semiconductors is often controlled by the defects that are introduced by impurities or by structural flaws. For a-Si:H this provides alternative conduction mechanisms, via extended state electron and hole conduction (as in crystalline silicon), and via hopping of electrons or holes through localized energy levels within the bandtails or within the bandgap. The relative importance of each route depends on the temperature and on the density of states (DOS) in each energy range, as well as on the mobility of the charge carriers.

These different conduction mechanisms are clearly seen in the temperature dependence of the conductivity of a-Si:H and of many amorphous chalcogenides, since a semilog plot of conductivity versus $1000/T$ K will have portions with different slopes as each mechanism dominates a particular temperature range. Near room temperature, good-quality a-Si:H films in the dark exhibit a well-defined Arrhenius relationship with an activation energy that is the energy separation of the Fermi level from the conduction band extended states, electrons dominating the conductivity. The conduction band extended states do not have the high electron mobility found in crystalline silicon, due to the disruption of long-range order and the consequent increased electron scattering. Typical electron mobilities in a-Si:H are $\sim 1\ \mathrm{cm^2\,v^{-1}\,s^{-1}}$, and hole mobilities are even lower.

The band tail states have even worse carrier mobilities, since movement is by hopping to vacant sites that might lie at some distance, and hopping is unlikely between sites that have a large energy difference. (The process has an inverse exponential dependence on the energy difference between the initial and final states.) Hopping of carriers between states that lie within the bandgap is a significant contribution to the conductivity only at low temperatures, unless the density of such defects is high. In materials having weakly localized states, nearest-neighbor hopping is supplanted by variable range hopping, as carriers prefer to hop to a more distant site if it has a smaller energy difference from the starting site than do the neighboring sites. As the temperature is lowered, this hopping distance will increase and the conductivity has N.F. Mott's characteristic $\exp(-B/T^{1/4})$ dependence on temperature, T. Lower-quality films deposited by plasma CVD on to unheated substrates tend to be p-type due to hopping conductivity through the high density of defect states. The link between substrate deposition temperature and material quality is caused by the way in which hydrogen is incorporated into the growing material: high temperatures allow only singly bonded hydrogen, which passivates dangling bonds thus removing bandgap defect energy levels. Low temperatures also allow hydrogen to be incorporated as $Si-H_2$ and this disrupts the Si–Si lattice bonding, producing more defect levels. Confirmation of this effect is obtained from sputtered amorphous silicon, with and without hydrogen added to the argon.

When a-Si:H is illuminated it may show strong photoconductivity, depending on the preparative conditions. Free carriers are generated by the illumination but may fall into defect states (energies within the band gap) where they can be released or can recombine with the oppositely charged carriers. The detailed response to light (wavelength, intensity, duration) and temperature thus depends on the distribution and concentration of defects. Luminescence effects are similarly influenced. Most amorphous chalcogenides are only weakly photoconductive due to their high density of gap states, those near the middle of the gap (e.g., VAPs) being highly effective recombination centers.

The electrical properties of other hydrogenated amorphous tetrahedral semiconductors based on silicon are determined by their defect populations, and in general they are less well behaved. a-Si:Ge:H and a-Si:Sn:H have reduced bandgaps and a-Si:C:H and a-Si:N:H have increased bandgaps, which offer many possibilities for multilayer device design, provided that their normally high defect contents can be reduced or controlled.

The study of glassy semiconductors was stimulated in the 1960s by the discovery of electrical switching in sputtered multicomponent thin films.

These simple devices changed from a nonconducting to a conducting state when the applied electric field exceeded a certain threshold, and returned to their initial state when the current fell below a certain value. A variety of other switching types was discovered, especially under pulsed operation, including those having a memory effect based on the reversible crystallization and vitrification of a narrow filament between the contacts. A theoretical model of these devices comprises both electronic and thermal effects, which are responsible for initiating the switching event and for memory effects. A practical problem has been the variable jitter in the switching time.

Applications

Amorphous silicon alloy films are valuable as the active layers in thin-film photovoltaic cells, two-dimensional optical position detectors, linear image sensors (optical scanners), and thin-film transistors used in liquid crystal display panels. They also have uses as antireflection coatings and planar optical waveguides. Amorphous chalcogenide films have been used as electrical switches and memory devices, for optical data storage, and as dry photoresists and electron beam resists. Both types of amorphous material have been used in xerography and medical X-ray imaging detectors.

Amorphous silicon solar cells have power conversion efficiencies of ~12% for the most complicated structures. These are tandem cells that use different alloys (including a-Si:C:H) for the various layers, in order to enhance effective absorption of the solar spectrum. A serious drawback of using amorphous semiconductors in electronic devices is their short carrier diffusion length. In solar cells this is circumvented by ensuring that there is a built-in field to separate photogenerated electron–hole pairs before they can recombine. Because doped amorphous silicon alloys have high defect densities, it is difficult to make the effective pn junctions that provide such built-in fields in conventional crystalline silicon cells. Instead, amorphous silicon cells use pin structures, where the i-layer is effectively undoped and provides an extended electric field between the p-i and i-n junctions. Long periods of illumination increase the dark current in these devices, as additional defects are generated. This 'Staebler–Wronski effect' is believed to be associated with hydrogen movement at a passivated dangling bond (Si–H) and a nearby weak Si–Si bond. It is a cause of slow degradation in unstabilized solar cells although the change may be reversed by thermal annealing.

In contrast to 'minority carrier' photovoltaic devices, thin-film transistors necessitate the control of majority carriers alone. To address each pixel in a liquid crystal display, a switching element is needed and simple thin-film transistors fulfill this role. Although low carrier mobilities in amorphous semiconductors restrict the current passed and switching speeds, the requirement for coating large area glass panels at low temperatures make a-Si:H with a-Si:N:H gate dielectric almost the only options. Improved performance is possible by using pulsed laser irradiation to change the amorphous semiconductor to polycrystalline material.

Further Reading

Brodsky MH (ed.) (1985) *Amorphous Semiconductors*, 2nd edn. Berlin: Springer-Verlag.

Elliot SR (1984) *Physics of Amorphous Materials*. London: Longman.

EMIS Datareviews Series No.1 (1989) *Properties of Amorphous Silicon*, 2nd edn. London: INSPEC.

Joannopoulos JD and Lucovsky G (eds) (1984) *The Physics of Hydrogenated Amorphous Silicon*. Topics in Applied Physics, vols 55 and 56. Berlin: Springer-Verlag.

Journal of Non-Crystalline Solids (2000) *Proceedings of the 18th International Conference on Amorphous and Microcrystalline Semiconductors (1999)*, vol. 266–269. This journal has published the proceedings of most of these well-established annual meetings dealing with all aspects of amorphous semiconductors.

Kramer B and MacKinnon A (1993) Localization – theory and experiment. *Reports on Progress in Physics* 56: 1469–1564.

Luft W and Tsuo YS (1993) *Hydrogenated Amorphous Silicon Alloy Deposition Processes*. New York: Marcel Dekker.

Madan A and Shaw MP (1988) *The Physics and Applications of Amorphous Semiconductors*. San Diego, CA: Academic Press.

Mott NF (1985) Electrons in non-crystalline materials – the last twenty five years. *Contemporary Physics* 26: 203–215.

Pankove JI (volume ed.) (1984) Hydrogenated amorphous silicon. In: Willardson RK and Beer AC (series eds) *Semiconductors and Semimetals*, vol. 21, Parts A–D. Florida: Academic Press.

Zallen R (1983) *The Physics of Amorphous Solids*. New York: Wiley.

Band Structure Engineering

C R Pidgeon, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Direct gap semiconductors have many unique properties which are advantageous for optical and electronic applications, and in particular for detectors and emitters of infrared radiation associated with interband optical transitions. Gallium arsenide, GaAs, based structures have principally been used in the near-infrared, and so-called narrow gap semiconductors in the mid-infrared (MIR) spectral regions. More recently systems based on intersub-band transitions have shown great promise for MIR devices, with the obvious advantage for some applications that they can be constructed from large gap, better controlled materials (mainly InP- and GaAs-based to date) and have a peak operating frequency which is determined simply by the design of the particular quantum structure. Both approaches have been substantially advanced by band structure engineering techniques, reducing Auger losses and phonon losses, respectively, which is the subject of the present work.

The classic examples of narrow gap materials are the III–V semiconductors (InSb, InAs, and their related alloys), II–VI semiconductors (mainly $Hg_{1-x}Cd_xTe$), and the lead salt alloys. As a direct result of the small energy gap, these materials have a very low electron effective mass, very high electron mobility, even at room temperature, and a high saturation velocity. In addition they have a high ratio of stimulated to spontaneous transition probabilities and a large nonlinear susceptibility. However, they have the disadvantage of a high intrinsic carrier density at room temperature and thermal generation rate, so that a fundamental limitation to the performance of devices at elevated temperatures (preferably room temperature) is nonradiative losses resulting from Auger recombination processes. One solution is to cool these systems to 80 K or below which involves expensive and bulky cooling methods such as Stirling-cycle coolers, negating their advantages over other, otherwise inferior, devices. In the past 10 years there has been an intensive effort to circumvent this limitation by quantum structure and so-called 'band structure engineering' techniques made possible by epitaxial growth techniques, whereby the Auger processes are suppressed. Since the emphasis of the band structure engineering work has been mainly towards the optimization of III–V semiconductor lasers rather than detectors, we reflect that emphasis here. Although developed more recently than the bipolar diode lasers, the unipolar quantum cascade (QC) lasers provide perhaps the best example of systems that can be conveniently optimized by band structure engineering techniques. In this case, phonon processes dominate so that the design is to do with phonon suppression. Thus, we are concerned principally with suppression of Auger processes for interband devices, and phonon processes for intraband devices.

Nonradiative Processes and Band Structure Engineering

Suppression of Auger Recombination

Electron recombination processes are dealt with in detail elsewhere in the Encyclopedia (*see* **Semiconductor Physics:** Recombination Processes) but we briefly summarize the main points here. It is well known that there are three principal interband recombination mechanisms by which thermodynamic equilibrium is established in semiconductors: Shockley–Read recombination, radiative recombination, and Auger recombination. These are shown schematically in Figure 1. The Shockley–Read recombination occurs via extrinsic energy levels produced by lattice defects and impurities, Figure 1a, and can in principle be controlled by the procedure used to grow the material; hence it does not provide a fundamental limit to the

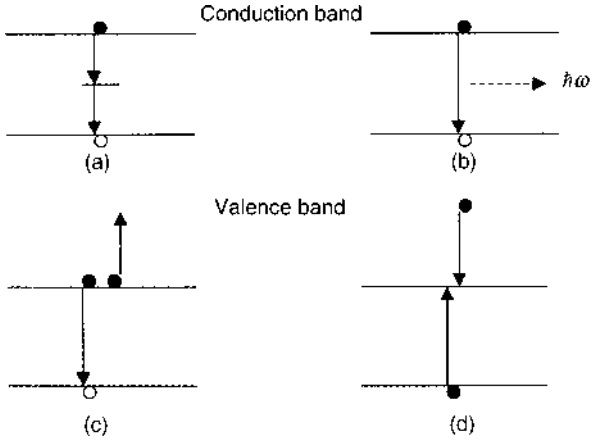

Figure 1 Electronic processes in semiconductors: (a) Shockley–Read recombination; (b) radiative recombination; (c) Auger recombination; (d) impact ionization.

carrier lifetime. The radiative and Auger mechanisms (Figure 1b,c, respectively) are, however, fundamental and are determined by the electronic structure of the material. The former is central to the operation of light-emitting diodes and is approximately inversely proportional to the electron and hole concentration. Optimum detector operation is achieved in materials that are limited by the radiative lifetime mechanism. Thus, in both cases it is desirable to minimize Auger processes.

By contrast with the two-particle radiative recombination process, Auger recombination is a three-particle interaction. For example in one so-called CHCC (or Auger-1) process an electron collides with a second electron; the first electron recombines with a hole in the valence band and the energy produced by the electron–electron collision excites the second electron higher up into the conduction band. This second electron eventually de-excites by phonon emission, and can cause observable heating of carriers by an Auger type of up-conversion. However, for the present article we are only interested in the first part of the process, so we do not consider the heating further here. The Auger process is the inverse of impact ionization where an energetic conduction electron collides with a valence electron, thereby creating an electron–hole pair as shown in Figure 1d. The final state consists of two low-energy electrons and one hole. Analogously, in Auger recombination the initial state has three particles (e.g., two electrons and one hole) and the final state one (e.g., one electron). Due to this three-body nature, the collision probability (or transition rate) per unit volume, R, varies with the cube of the carrier concentration and (if $n = p$) may be written $R = Cn^3$, where n is the electron density and C is referred to as the Auger coefficient. Note that the Auger rate is sometimes defined per carrier as $R' = Cn^2 = 1/\tau_A$, where τ_A is the Auger lifetime.

Narrow gap III–V semiconductors have the familiar three-band (so-called Kane energy band structure – (*see* **Semiconductor Physics:** Band Structure and Optical Properties) consisting of an s-like conduction band separated by a small energy gap from a triply degenerate p-like valence band, part of whose degeneracy is raised by spin–orbit splitting as shown in Figure 2. Thus, a large number of permutations of Auger processes become possible for n- and p-type semiconductors. Those which are initiated by a collision between two electrons are referred to as conduction, or n-n, Auger processes (e.g., the CHCC process), and those by a collision between two holes as valence, or p-p, Auger processes. It can be shown that the n-n process (so-called CHCC or Auger-1) illustrated in Figure 3a is dominant in bulk n-type and the p-p process

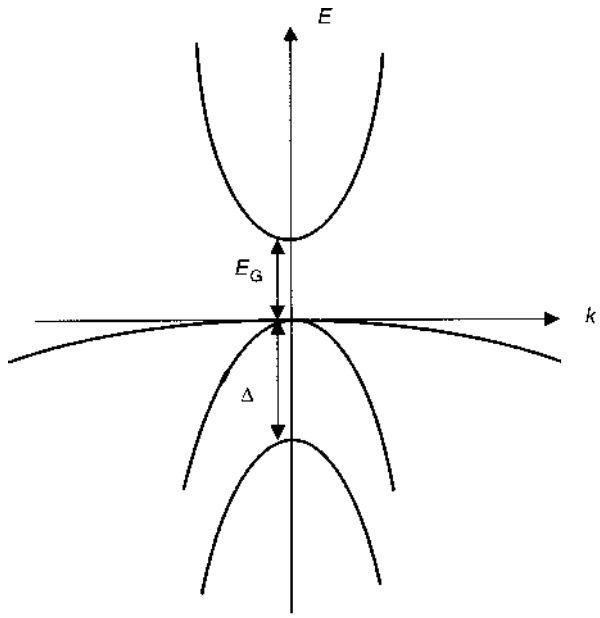

Figure 2 Schematic diagram of the conduction and valence band structure of InSb showing the energy bands around the Γ-point ($k = 0$) that are involved in Auger transitions.

(so-called CHLH or Auger-7) shown in Figure 3c is dominant in bulk p-type semiconductors, and so these have received the most attention. For larger gap (shorter wavelength) semiconductors there is sometimes a significant resonance between the bandgap and the spin–orbit splitting and the p-p process (CHSH) shown in Figure 3b becomes important. Clearly these simple descriptions require extension when the more complex band structure of strained layer superlattice quantum systems is considered.

With certain approximations, some analytic solutions for the Auger recombination rates are possible. This model is, while not appropriate for higher levels of band structure engineering, very important for giving an insight into how Auger effects should be studied, and how they themselves might be engineered. The approximations generally used are parabolic bands and Boltzmann statistics, but some further analytic results may also be obtained when these restrictions are relaxed. Bloch functions and Fermi's golden rule are utilized (*see* **Semiconductor Physics:** Recombination Processes) to describe the particle states, so obtaining the general expression for the Auger rate per unit volume. Energy and momentum conservation provide strong constraints on the resulting integral, and give rise to a considerable reduction in the Auger transition rate. However if, for example, the momentum conservation constraint is relaxed due to occupied levels existing with a broad range of momenta, such as at some band extremum with a

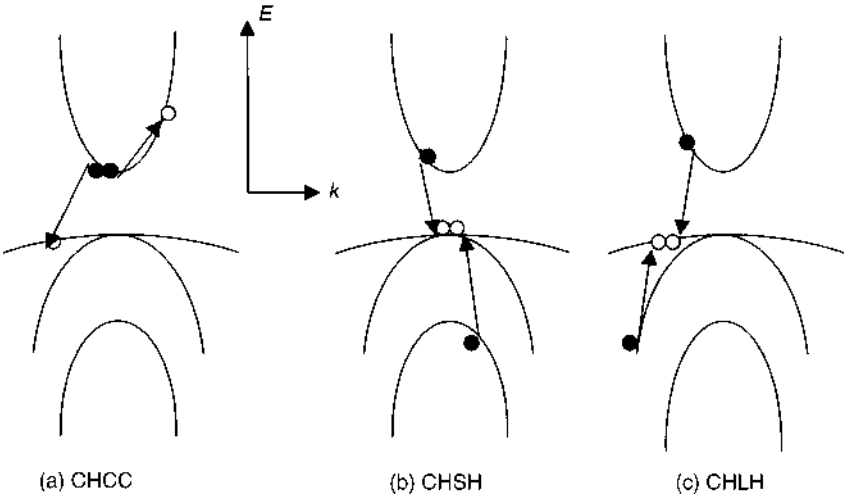

Figure 3 Important Auger recombination mechanisms for mid-infrared (MIR) semiconductor detectors and lasers: (a) CHCC or Auger-1; (b) CHSH; (c) CHLH or Auger-7.

large effective mass, then the Auger rate will be strongly enhanced. Or to put it the other way round, the Auger rate can be reduced by engineering the band structures such that the heavy hole effective mass is reduced (e.g., by stress – see below) to approach that of the conduction band. Furthermore, since we are concerned with nonradiative recombination across the bandgap with energy E_g, any energy levels (again band extrema in particular) separated from the conduction or valence band edges by E_g are particularly susceptible to the excitation part of the Auger process. The Auger rate may therefore be reduced by designing the structure so that there are no band extrema at energies E_g from the conduction and valence band edges.

Interband (Bipolar) Lasers

The technology of band structure engineering, which includes superlattices, quantum confinement, and intentional incorporation of strain, has now become sufficiently advanced that it can be employed in a routine manner. As originally proposed, the modified band structure of strained III–V quantum-well structures can lead to significant benefits for diode laser performance, including reduced threshold current density, improved efficiency, and enhanced dynamic response and speed. Predicted advantages have been demonstrated, although for near-infrared devices these have not been quite as substantial as originally hoped because of the presence of phonon-assisted Auger processes which are not so susceptible to band structure engineering techniques. Strained layer lasers are now commercially available at a variety of MIR wavelengths, where the dominant nonradiative loss is via the direct interband Auger scattering routes.

From the point of view of the present article the main contribution of 'designed' strain splittings and quantum confinement is in the reduction of the Auger coefficient itself, by removing states that easily satisfy the momentum and energy conservation conditions required for the process to take place. However, in addition to this there can be a substantial reduction in the threshold carrier injection level required for population inversion in a laser. In real III–V semiconductors the heavy hole valence band mass, m_h, is so much greater than the electron mass, m_e, that the usual laser picture of degenerate electron and hole distributions does not apply. At normal carrier injection levels the quasi-Fermi level of the holes is above the top of the valence band resulting in a classical heavy hole distribution. Thus, the lower laser levels are almost completely filled with electrons, by contrast with the ideal situation of two mirror bands. This leads to an injection level for the laser threshold of greater than 2:1 required for the actual case compared to the ideal case. Since the Auger recombination rate per unit volume is approximately proportional to the cube of n, this factor of ~ 2 reduction in injection level can produce a factor of almost an order of magnitude reduction in this loss mechanism. Large biaxial strains, either compressive or tensile, can lead to a considerable reduction in the hole mass and hence in the density of states at the top of the valence band, so that the laser characteristics may then approach those of an ideal semiconductor, with $m_e = m_h$.

A further important loss mechanism is free hole absorption of the heavy holes resulting from intervalence band transitions. This can be effectively

eliminated, and other improvements made to laser performance, by the introduction of a strained layer structure which raises the degeneracy of the light and heavy hole at $k = 0$ and reduces the mass of the heavy hole. One example of this comes just from the symmetry of the problem, where clearly the cubic symmetry of the p-like valence band is not matched to the uniaxial symmetry of the laser beam. Application of biaxial or uniaxial strain removes the cubic symmetry, changing the relative energies of the p_x, p_y, and p_z valence states, and making it possible for almost all of the injected holes to be in the states of the correct symmetry to take part in the lasing action. The one parasitic phenomenon in semiconductor lasers which is hardly affected by band structure engineering is spontaneous emission, which will always make some contribution to the current density required at threshold.

The band structure of a bulk unstrained direct gap tetrahedral semiconductor is shown again for comparison in Figure 2. The effect of biaxial tension and compression is shown respectively in Figure 4a,b. In each case the lowest conduction band is approximately parabolic near the zone center, but becomes anisotropic with the electron dispersion at small k. The axial strain breaks the cubic symmetry of the semiconductor, introducing a tetragonal distortion

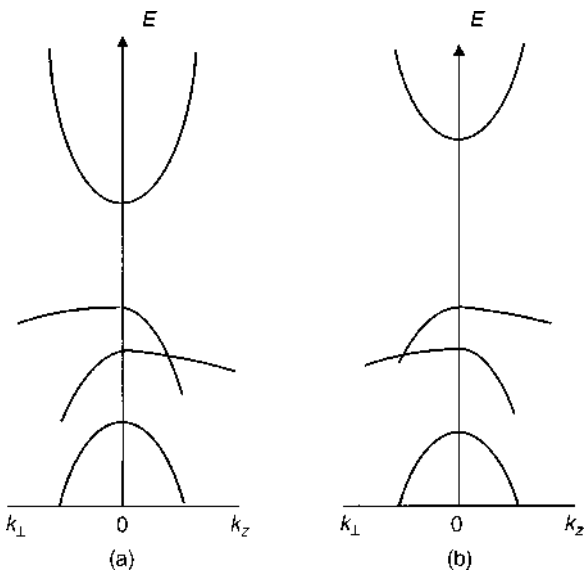

Figure 4 Schematic representation of the band structure of a direct gap tetrahedral semiconductor: (a) under biaxial tension the hydrostatic component of the tension reduces the mean bandgap, while the axial component splits the degeneracy of the valence band maximum and introduces an anisotropic valence band structure (the highest band being light along the strain axis, k_z and comparatively heavy perpendicular to that axis, $k_\perp$); (b) under biaxial compression the mean bandgap increases and the valence splitting is reversed.

which splits the degeneracy of the light and heavy hole states at the valence maximum, Γ, typically by about 60–80 meV for a 1% lattice mismatch. The resulting valence band structure is highly anisotropic (Figure 4a,b), with the band which is heavy along the strain axis, wavevector $k_\perp$, being comparatively light in the plane, along $k_\parallel$, and vice versa. The in-plane mass, $m_\parallel$, determines the density of states, and thus, under compression (Figure 4b), can be reduced according to the principles described above to improve conditions towards those for ideal laser operation: viz, (1) that there is only one band at the valence band maximum; (2) that it has an effective mass as close to that of the electron as possible; and (3) that above the Bernard–Duraffourg transparency condition, the polarization of the gain is anisotropic, with spontaneous emission and gain being suppressed along all directions except along the axis contributing to the laser beam.

Since the Auger recombination involves three carriers, the Auger current J_{NR} in a device with an undoped active region varies approximately as $J_{NR}(T) \propto C(T)n_{th}^3$, where $C(T)$ is the temperature-dependent Auger coefficient per unit volume and nth is the threshold carrier density. The influence of strain on Auger recombination has two aspects, as discussed above. First, J_{NR} is very sensitive to any reduction in n_{th} brought about by strain either through a decrease in h (heavy hole) mass or an increase in the optical transition strength. Secondly, strain may change the magnitude of $C(T)$ for the interband Auger processes described earlier (Figure 3). A simple calculation of Auger recombination in a quantum well heterostructure, assuming parabolic bands and Boltzmann statistics, shows that the coefficient increases exponentially with temperature as $C(T) = C_o\exp(-E_a/kT)$, where the activation energy, E_a, is approximately dependent on $(m_h)^{-1}$. Thus, both of these should be reduced by several orders of magnitude due to a reduction in the heavy hole mass, m_h, resulting from the strain effects discussed above.

Phonon Scattering: QWIPs and QC Lasers

Sub-band transitions are useful for MIR detectors and emitters because they make it possible to cover a broad range of MIR wavelengths using the same combination of large gap, InP-based and GaAs-based state of the art semiconductor alloys that are used for high-speed electronics and optical communications. They are the most advanced of the low-dimensional solids from the point of view of demonstrated detector performance. This superior control of materials results in high uniformity and has thus

allowed the demonstration of large 128 × 128 arrays with excellent imaging performance. Detectivities $D^* > 10^{10}\,\mathrm{cm\,Hz^{1/2}\,W^{-1}}$, low noise equivalent temperature differences, and excellent infrared video imaging can be achieved at 10 μm for temperatures $T = 10–80$ K.

Although the QWIP (quantum well infrared photodetectors) detectors showed promise for many years as alternatives to HgCdTe systems, the first intersub-band laser was not realized until the demonstration of the quantum cascade (QC) laser. Intersub-band (QC) lasers differ fundamentally from bipolar diode lasers in that they rely on only one type of carrier and on electronic transitions between conduction sub-band states arising from size quantization in semiconductor quantum heterostructures. These in-plane sub-bands have the same curvature (by contrast with the interband devices which have conduction and valence bands of opposite curvature), so that the joint density of states and corresponding gain spectrum (collision limited) is much narrower and approximately symmetric. Thus, the arguments which show that the Kane-type band structure is particularly disadvantageous for the overall Auger loss mechanism in interband diode lasers do not apply; the problems resulting from having $m_e \ll m_h$ are removed and the inversion is much less sensitive to thermal broadening of the electron distribution. In addition, there is no free hole absorption, and, since the component semiconductor materials are of comparatively wide bandgap, there is negligible intrinsic carrier concentration. However, the threshold current density tends to be rather large owing to the rapid nonradiative phonon relaxation of the population inversion.

In the original QC laser structure the electrons streamed down a potential staircase, sequentially emitting photons at the steps, consisting of coupled quantum wells in which the population inversion between discrete conduction band excited states was achieved by control of tunneling. The laser structure was grown with the $Al_{0.48}In_{0.52}As/Ga_{0.47}In_{0.53}As$ heterojunction material system lattice matched to InP. Electrical pumping was achieved by alternating 25 undoped coupled-well active regions with compositionally graded layers. The coupled-well region was essentially a four-level laser system, where a population inversion was achieved between the two excited states, $n = 3$ and $n = 2$. While the nonradiative phonon emission lifetime is short ($\tau_{32} \sim 4.3$ ps at the electric fields used in the tunneling process), strong coupling to an adjacent GaInAs well gives an even shorter lower state lifetime ($\tau_{21} \sim 0.6$ ps) from optical phonons with nearly zero momentum transfer between the strongly overlapped and closely spaced $n = 2$ and $n = 1$ states. This enables the achievement of a population inversion between these states. (Note that the separation between the $n = 2$ and $n = 1$ states, E_{21}, is designed to be approximately equal to the optical phonon energy.)

The radiative efficiency of these QC devices is unavoidably low ($<10^{-3}$) because most of the electrons in the upper sub-band relax within ~1 ps through nonradiative optical phonon emission before radiative transitions with a lifetime in the nanosecond range have a chance to occur. An alternative configuration has been proposed, the type II interband cascade laser (T2ICL), which retains the advantages of cascaded tunneling injection and wavelength tuneability but eliminates the nonradiative phonon path between valence and conduction bands of opposite curvature. In addition, band structure engineering ideas discussed above are incorporated to suppress the unwanted Auger processes; the design results in a light in-plane hole mass and the removal of resonances between the energy gap and intervalence transitions.

See also

Semiconductor Materials: Group IV Semiconductors, Si/SiGe; Type-II Quantum Wells and Superlattices. **Semiconductor Physics:** Band Structure and Optical Properties; Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures; Recombination Processes.

Further Reading

Faist J, Capasso F, Sirtori C, Sivco DL, Hutchinson AL and Cho AL (1994) Quantum cascade laser. *Science* 264: 553–556.

O'Reilly EP and Adams A (1994) Band structure engineering in strained semiconductor lasers. *IEEE Journal of Quantum Electronics* 30: 366–379.

Pidgeon CR, Ciesla CM and Murdin B (1997) Suppression of non-radiative processes in semiconductor mid-infrared emitters and detectors. *Progress in Quantum Electronics* 21: 361–419.

Yablonovitch E and Kane EO (1988) Band structure engineering of semiconductor lasers for optical communications. *Journal of Lightwave Technology* 6: 1292–1299.

Youngdale ER, Meyer JR, Hoffman CA, Bartoli FJ, Grein CH, Young PM, Ehrenreich H, Miles RH and Chow DH (1994) Auger lifetime enhancement in InAs-GaInSb superlattices. *Applied Physics Letter* 64: 3160–3162.

Dilute Magnetic Semiconductors

R R Gałązka, Polish Academy of Sciences, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction, General Remarks

Semimagnetic semiconductors (SMSs) are a group of solids at the interface between semiconductors and magnetic materials. Semimagnetic semiconductors, also referred to as dilute magnetic semiconductors, are semiconductor-based solid solutions where some of the cations are replaced by transition metals or rare earth elements. The crystallographic structure of the semiconductor is conserved; the lattice constant is a function of composition. From a magnetic point of view SMSs are disordered magnetic materials, since magnetic atoms are randomly distributed in the cation sublattice of the semiconductor compound. Table 1 presents the most complete list of bulk ternary SMSs. Quaternary SMSs are also investigated (e.g., PbSnMnTe or CdMnSeTe).

Generally speaking, in SMSs there coexist two interrelated and interacting subsystems: mobile delocalized charge carriers and localized magnetic moments connected with paramagnetic ions. Electronic (mainly optical) properties of SMSs have been the subject of intensive studies since early in 1970. Due to the strong spin exchange interaction between mobile carriers and localized magnetic moments (the exchange constant of the sp-d interaction is of the order of 1 eV for II–VI SMSs) significant changes of the band structure and behavior of the carriers were observed. A number of new physical phenomena were discovered, such as giant Faraday rotation (now important in magnetic field sensors and other applications in, for example, the automobile industry based on this strong magneto-optical effect), the magnetic field induced metal–insulator transition, the bound magnetic polaron and quantum thermomagnetic oscillations. Investigations of the influence of the electron subsystem on the magnetic properties of SMSs are connected with the observation of a ferromagnetic phase in PbSnMnTe induced by carrier concentration.

Molecular beam epitaxy (MBE) grown GaAs with Mn opened a new chapter of III–V SMSs showing ferromagnetic behavior with the highest transition temperature $T_c = 110\,\mathrm{K}$. In 1987 the first paper devoted to layered structure and magnetic properties of low-dimensional (LD) SMSs was published starting intensive studies of superlattices and other LD structures made of SMSs.

Energy Band Structure, Optical and Magneto-optical Properties of SMSs

Replacing cations such as Cd or Hg with a paramagnetic ion such as Mn in the same crystallographic structure does not markedly disturb the semiconductor properties of the material. The energy gap changes (increases) but the conduction and valence bands conserve their symmetry and character as in nonmagnetic semiconductor mixed crystals. The spin momentum of paramagnetic ions is connected with the 3d or 4f shell, for transition metals or rare earth elements, respectively. The energy level of 3d or 4f electrons lies below the top of the valence band and thus has negligible influence on its shape or that of the bottom of the conduction band. Thus all basic semiconductor optical properties connected with the band symmetry and topology are the same as for a typical semiconductor. These are described, for example, in articles on the lead salts (MS641) involving direct optical transitions at the L-point of the Brillouin zone, and on mercury cadmium telluride (MS639) with direct optical transitions at the Γ-point. The only influence of the magnetic subsystems on electrons comes from the spin exchange interaction between mobile carriers and localized magnetic ions. This interaction can be represented by a Heisenberg term

$$H_{\mathrm{ex}} = \bar{\sigma} \sum_i \bar{S}_i J(\mathbf{r} - \mathbf{R}_i) \qquad [1]$$

where $\bar{\sigma}$ and $\bar{S}$ are spin operators of the band electron and magnetic ion, respectively, the summation is over all lattice sites occupied by the magnetic ions, and J is the exchange constant.

Assuming mean field and virtual crystal approximations this term can be rewritten in the form which is periodic with a lattice constant:

$$H_{\mathrm{ex}} = x\langle\bar{S}\rangle_{\mathrm{av}}\bar{\sigma} \sum_R J(\mathbf{r} - \mathbf{R}) \qquad [2]$$

where x is the molar fraction of magnetic ions, $\langle\bar{S}\rangle_{\mathrm{av}}$ is the average over all magnetic ions and is directly

Table 1 Composition and crystal structure of ternary bulk SMSs

Material	*Crystal structure*	*Composition range*
II–VI (Mn)		
$Zn_{1-x}Mn_xS$	ZB	$0 < x \leq 0.10$
	W	$0.10 < x \leq 0.45$
$Zn_{1-x}Mn_xSe$	ZB	$0 < x \leq 0.30$
	W	$0.30 < x \leq 0.57$
$Zn_{1-x}Mn_xTe$	ZB	$0 < x \leq 0.86$
$Cd_{1-x}Mn_xS$	W	$0 < x \leq 0.45$
$Cd_{1-x}Mn_xSe$	W	$0 < x \leq 0.50$
$Cd_{1-x}Mn_xTe$	ZB	$0 < x \leq 0.77$
$Hg_{1-x}Mn_xS$	ZB	$0 < x \leq 0.37$
$Hg_{1-x}Mn_xSe$	ZB	$0 < x \leq 0.38$
$Hg_{1-x}Mn_xTe$	ZB	$0 < x \leq 0.35$
II–V (Mn)		
$(Cd_{1-x}Mn_x)_3As_2$	tetr.	$0 < x \leq 0.18$
$(Zn_{1-x}Mn_x)_3As_2$	tetr.	$0 < x \leq 0.15$
IV–VI (Mn, Eu, Gd)		
$Pb_{1-x}Mn_xS$	RS	$0 < x \leq 0.05$
$Pb_{1-x}Mn_xSe$	RS	$0 < x \leq 0.17$
$Pb_{1-x}Mn_xTe$	RS	$0 < x \leq 0.12$
$Sn_{1-x}Mn_xTe$	RS	$0 < x \leq 0.40$
$Ge_{1-x}Mn_xTe$	RH	$0 < x \leq 0.18$
	RS	$0.18 < x \leq 0.50$
$Pb_{1-x}Eu_xSe$	RS	$0 < x \leq 0.04^*$
$Pb_{1-x}Eu_xTe$	RS	$0 < x \leq 0.32^*$
$Sn_{1-x}Eu_xTe$	RS	$0 < x \leq 0.013$
$Pb_{1-x}Gd_xTe$	RS	$0 < x \leq 0.11^*$
$Sn_{1-x}Gd_xTe$	RS	$0 < x \leq 0.09^*$
II–VI (Fe, Co, Cr)		
$Zn_{1-x}Fe_xS$	ZB	$0 < x \leq 0.26^*$
$Zn_{1-x}Fe_xSe$	ZB	$0 < x \leq 0.22^*$
$Zn_{1-x}Fe_xTe$	ZB	$0 < x \leq 0.01^*$
$Cd_{1-x}Fe_xS$	W	$0 < x < 0.09^*$
$Cd_{1-x}Fe_xSe$	ZB	$0 < x < 0.20$
$Cd_{1-x}Fe_xTe$	ZB	$0 < x \leq 0.06^*$
$Hg_{1-x}Fe_xS$	ZB	$0 < x \leq 0.037^*$
$Hg_{1-x}Fe_xSe$	ZB	$0 < x \leq 0.20^*$
$Hg_{1-x}Fe_xTe$	ZB	$0 < x \leq 0.02^*$
$Zn_{1-x}Co_xS$	ZB	$0 < x \leq 0.15^*$
$Zn_{1-x}Co_xSe$	ZB	$0 < x \leq 0.10^*$
$Zn_{1-x}Co_xTe$	ZB	$0 < x < 0.06^*$
$Cd_{1-x}Co_xS$	W	$0 < x \leq 0.064^*$
$Cd_{1-x}Co_xSe$	W	$0 < x \leq 0.082^*$
$Cd_{1-x}Co_xTe$	ZB	$0 < x \leq 0.04$
$Hg_{1-x}Co_xS$	ZB	$0 < x \leq 0.02$
$Hg_{1-x}Co_xSe$	ZB	$0 < x \leq 0.047^*$
$Zn_{1-x}Cr_xSe$	ZB	$0 < x \leq 0.005^*$
$Zn_{1-x}Cr_xTe$	ZB	$0 < x \leq 0.003^*$
$Cd_{1-x}Cr_xS$	W	$0 < x \leq 0.002^*$

ZB – zinc blende, W – wurtzite, RS – rock salt, RH – rombohedral. Asterisks denote SMSs with the highest reported *x* value; tetr, tetragonal structure.
Reprinted in part from Gałązka RR (1995) Influence of electron subsystem on magnetic properties of semimagnetic semiconductors. In: Tróc R, Morkowski J and Szymczak M (eds) Proceedings of the International Conference of Magnetism, Warsaw 1994. *Journal of Magnetism and Magnetic Materials*, 113–116. Copyright Elsevier Science.

related to the magnetization of the sample:

$$M = N_0 g \mu \langle \bar{S} \rangle_{\mathrm{av}} \qquad [3]$$

N_0 is the number of unit cells per unit volume, g is the Lande factor for magnetic ions, and μ is the Bohr magneton.

The Heisenberg term must be added to the effective-mass Hamiltonian to solve the energy eigenvalue problem for SMSs. Realizing this was the turning point in understanding their properties and provided a basis for this group of alloys to be distinguished from other semiconductor mixed crystals. Since that time SMSs have become the subject of intensive studies in Europe and elsewhere.

The exchange interaction depends on the ion spin and the exchange constant. The spin of Mn is the highest for transition metals and equals five Bohr magnetons, and whilst a typical value for the exchange coupling between magnetic ions $J_{\mathrm{d-d}}$ is 10^{-3} eV, the exchange constant for electrons and holes with magnetic ions $J_{\mathrm{sp-d}}$ is about 1 eV in SMSs.

The exchange interaction is also strongly temperature and magnetic field dependent. As the macroscopic magnetization of a sample is proportional to the thermodynamic average value of the magnetic ion spin, to a good approximation we can replace the ion spin operators in the Hamiltonian by their average value, calculated or taken from measurements of the magnetization. From typical magnetic functions we can thus obtain information on the electron behavior.

In semiconductors, an external magnetic field acts on both the orbital motion and the spin of electrons producing Landau quantization and spin splitting, to an extent dependent on the effective mass of the carriers. The exchange interaction acts on their spin only and is, in effect, mass independent. Because of this, and also the large value of $J_{\mathrm{sp-d}}$, the band structure in turn changes drastically under the influence of an external magnetic field and depends strongly on temperature. From Figure 1 we can see that the very typical structure of the degeneracy of the light and heavy hole valence bands disappears in SMS when a magnetic field is applied. The change is accompanied by a drastic increase of hole mobility like that observed in low-dimensional structures. In SMS such as $Hg_{1-x}Mn_xTe$ or $Hg_{1-x}Mn_xSe$ (for $x \leq 0.07$) (Figure 2), the magnetic field produces an overlap of the valence and conduction bands, an effect never observed in nonmagnetic semiconductors.

Impurity levels are also influenced by exchange interactions: the ionization energy of an acceptor

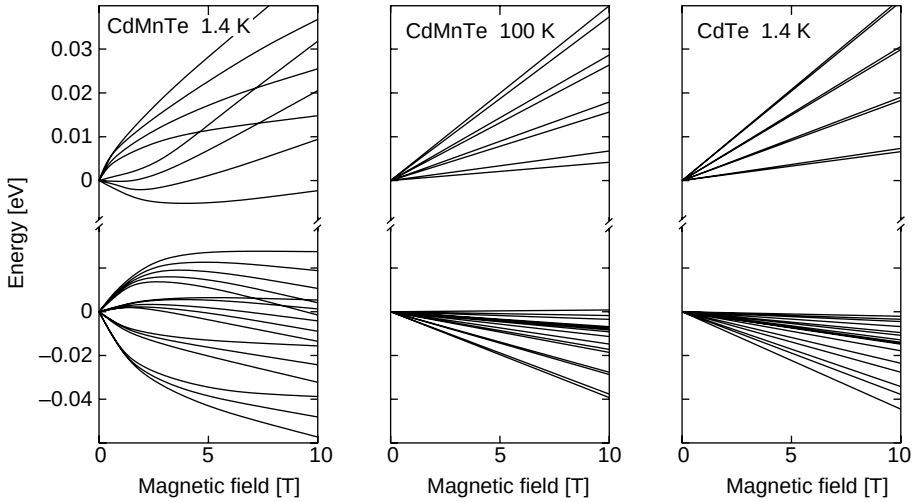

Figure 1 Conduction and valence band quantization in an external magnetic field for $Cd_{0.95}Mn_{0.05}Te$ and CdTe. Spin splitting, which is very weak in CdTe, is the main effect in CdMnTe. Notice the almost equal splitting of the heavy- and light-hole bands in CdMnTe. The influence of exchange interaction is visible even at 100 K. The picture of CdTe over this range of temperature is practically unchanged. Reproduced from Gałązka RR (1987) Semimagnetic semiconductors. *Europhysics News* 18(6): 90–92.

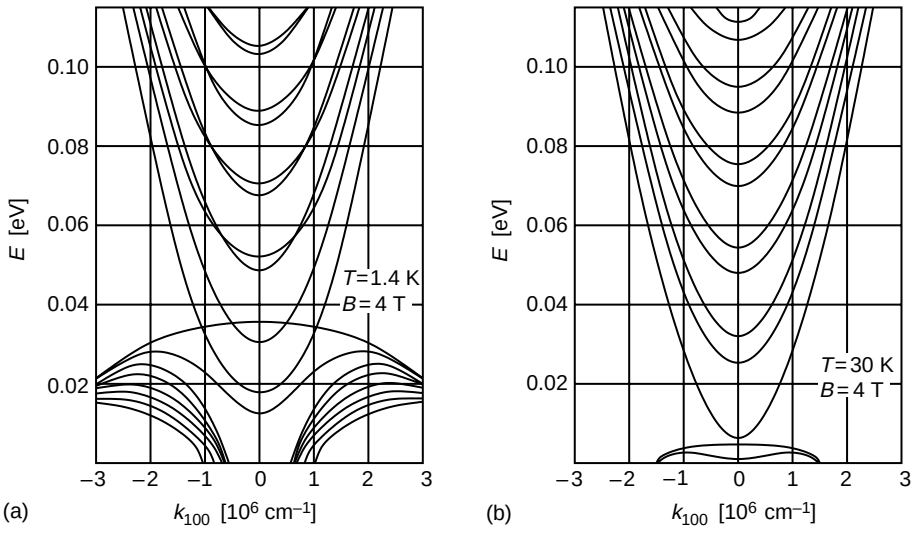

Figure 2 Energy band structure of $Hg_{1-x}Mn_xTe$, $x = 0.02$, calculated at 14 K (a) and 30 K (b) in a magnetic field of 4 T. After Gałązka RR (1978) Semimagnetic semiconductors based on HgMnTe and CdMnTe. Reproduced from Wilson BLM (ed) ICPS 14th, Edinburgh Institute of Physics, Conference series no. 43, 1978, p.133–140.

decreases and its wavefunction becomes anisotropic under the influence of a magnetic field. Both effects produce a giant negative magnetoresistance and a field-induced insulator–metal transition. Indeed all (particularly optical) properties susceptible to being changed by a magnetic field are very different in SMSs from those of standard semiconductors: the Faraday effect., Shubnikov–de Haas oscillations, interband magnetoabsorption, luminescence and spin flip Raman scattering (see below) become strongly dependent on temperature and magnetic field.

In the absence of a magnetic field, the average value of the ion spins is zero, the magnetization is zero too, and SMSs should behave as typical non-magnetic semiconductors. Whereas this is true for delocalized band electrons, when an electron is localized around an impurity, a bound magnetic

polaron can be observed. This concept was first introduced to explain the behavior of magnetic semiconductors (Eu chalcogenides) and then later was applied also to SMSs. A localized electron can produce a spontaneous magnetization within the range of its wavefunction via the exchange interaction. Local ferromagnetic ordering in this range produces a Stokes shift in spin-flip Raman scattering even in the absence of an external magnetic field and other related phenomena. Local fluctuations of magnetization and a detailed knowledge of electronic states are essential for a theoretical description of this effect. The bound magnetic polaron is a subtle example of the feedback between electronic and magnetic subsystems present in SMS.

Magnetic Properties

Magnetic properties of SMS connected with the exchange interaction between paramagnetic ions are similar to those of other dilute magnetic materials. The broad range of possible magnetic ion concentrations (up to $x = 0.80$ in $Zn_{1-x}Mn_xTe$ and $x = 0.70$ for $Cd_{1-x}Mn_xTe$) allows us to observe the evolution of magnetic interactions in the same crystallographic structure as a function of magnetic ion concentration. A continuous transition from diamagnetic behavior (for $x \leq 0.005$) of the host semiconductor, to paramagnetism, spin-glass (for $x \geq 0.02$) and finally antiferromagnetic order has been observed. Type III antiferromagnetic ordering is observed for $x > 0.6$ indicating that the magnetic elementary cell contains two elementary cells of the cation sublattice. From a crystallographic point of view the magnetic subsystem is disordered; the distribution of magnetic atoms on the cation sublattice is random.

A spin-glass-like state is observed in SMSs in the broadest range of paramagnetic ion concentrations ($0.02 \leq x \leq 0.60$) below and above the percolation threshold (Figure 3). Indeed the temperature dependence of the magnetic susceptibility below and above the transition temperature and their dynamic properties at low and high magnetic fields suggest rather the existence of two spin-glass phases arising from the competition between spin-glass and clustering behavior. Below the percolation threshold ($x \leq 0.17$ for fcc lattice) the mechanism responsible for a spin-glass phase is even less clear, although, for such a dilute system it must be long-range to produce freezing of the spins.

The antiferromagnetic order has been inferred from specific heat and magnetic susceptibility measurements, and additional information concerning this phase has come from neutron diffraction studies: only a certain fraction of the total number of magnetic ions (e.g., about 50% for CdMnTe, $x = 0.65$) is well ordered; the rest remain in a disordered spin-glass phase. Thus a mixed phase (antiferromagnetic and spin-glass coexisting together) rather than a truly antiferromagnetic phase is observed.

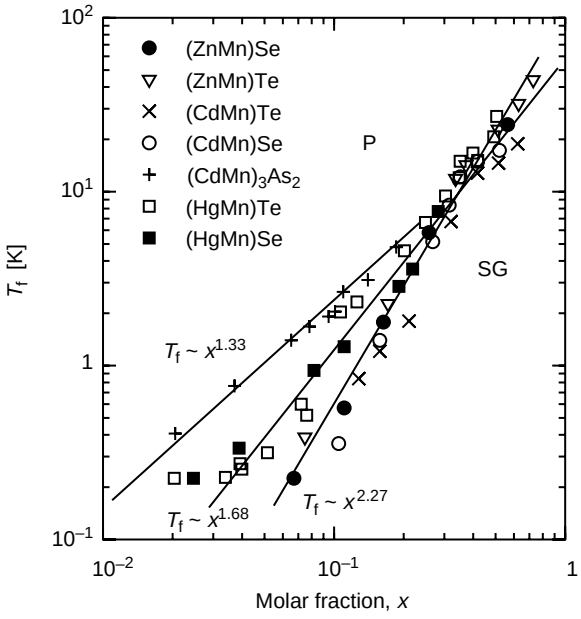

Figure 3 Magnetic phase diagram for the listed SMSs. P indicates the paramagnetic region, SG indicates the spin-glass region. Reprinted in part from Gałązka RR (1995) Influence of electron subsystem on magnetic properties of semimagnetic semiconductors. Reproduced with permission from Tróc R, Morkowski J and Szymczak M (eds) Proceedings of the International Conference of Magnetism, Warsaw 1994. *Journal of Magnetism and Magnetic Materials*, 113–116. Copyright Elsevier Science.

The influence of carrier concentration on magnetic properties has been extensively investigated in metallic alloys. Several IV–VI semiconductors like SnTe can exhibit quite high metallic-like carrier concentrations. In the quarternary alloy PbSnMnTe, the concentration of holes can be varied by annealing in the range 10^{20} to 5×10^{21} cm^{-3}. In $Pb_{0.25}Sn_{0.72}Mn_{0.03}Te$ an abrupt transition from the paramagnetic to ferromagnetic phase has been found for $p > 3 \times 10^{20}$ cm^{-3}. The Curie temperature is a function of hole concentration, but the Mn ions conserve their magnetic moment of five Bohr magnetons. The magnetic phase diagram for this first ferromagnetic SMS (Figure 4) is thus three dimensional (T,x,n) in contrast to the

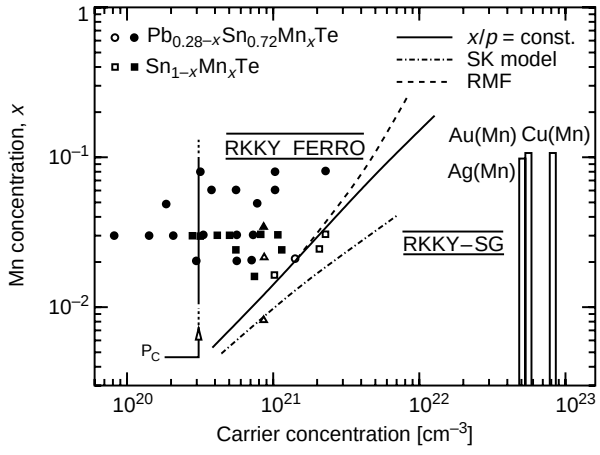

Figure 4 Magnetic phase diagram for PbSnMnTe (•) and SnMnTe (▪). The ferromagnetic phase is observed in the region above the critical carrier concentration p_c and the straight line defined by the equation $x/p =$ const. Away from this region a spin-glass state is observed. Calculated phase boundaries between the ferromagnetic and spin-glass state are also shown for two theoretical models: SK: Sherrington–Kirkpatrick model and RMF: random mean field model. Canonical metallic spin-glasses are also shown. Reproduced from de Jonge WJM, Story T, Swagten HJM and Eggerkamp PJT (1992) Carrier-induced breakdown of ferromagnetism. *Europhysics Letters* 17: 631–636.

two-dimensional diagram (T,x) usually presented for magnetic alloys. Even a low concentration of spin polarized carriers when optically pumped can change the magnetization of HgMnTe owing to the strong exchange interaction of carriers with magnetic ions.

Low-Dimensional (LD) SMSs: Magneto-optical and Spintronic Effects

Modern technology such as MBE makes it possible to obtain quasi-3D SMSs in the whole range of x from 0 to 1. It also gives broad possibilities to create different combinations of semiconductor, SMS, and magnetic materials. In addition the dimensionality of LD structures influences the properties of both electronic and magnetic subsystems. Further there is almost always some strain in the system which also influences the properties of superlattices and other LD structures. Studies of LD SMSs not only extend the material base but open a new area of dimension-dependent phenomena and interactions in SMS systems.

Among many effects observed in LD SMS structures we will mention only a few phenomena, where exchange interaction is essential and which cannot be observed in nonmagnetic LD semiconductors. In all cases the giant (magneto-optical) Zeeman splitting caused by the exchange interaction plays a key role.

In superlattices made of ZnSe/ZnMnSe with a small amount of Mn, the band offset of conduction and valence bands is negligible at zero external magnetic field. In the presence of an external field the conduction and valence bands of ZnMnSe split creating a spin-superlattice because the spin-split states allow only one spin configuration. The possibility of the existence of such a spin superlattice was earlier suggested for SMSs and later experimentally observed.

If the superlattice, quantum well or heterostructure is made of SMS and non-SMS material (e.g. ZnTe/CdMnSe or CdTe/CdMnTe) one can exploit the drastic differences in the Zeeman splitting occurring in different layers to pinpoint the localization in space of specific electronic states. If one (hole or electron) or both states involved in optical transitions originate from the SMS layer, this state will show a very strong Zeeman splitting in the external magnetic field. Thus the SMS component can be used as a 'marker' in order to bring out the general properties of the wavefunction distribution.

Excitonic Zeeman splitting in non-magnetic quantum wells with SMS barriers can be applied to studies of the shape of interface profile and the influence of the growth conditions on the shape of the interface. If the interface is not completely flat, even a small fraction of magnetic atoms in the well close to the interface can significantly enhance the Zeeman exciton splitting in the quantum well. The sensitivity of this method allows one to detect the sharpness of an interface with one monolayer accuracy.

The MBE technique opens up possibilities of creating hybrid structures with SMSs, useful for spintronic devices such as spin filters, spin transistors and tunneling magnetoresistance sensors. Very importantly, it has been realized that SMSs are highly suitable to form the ferromagnetic spin injecting electrodes in such systems where simple metallic electrodes fail because of the large electrical resistance mismatch between the electrode and the semiconductor in which the spin is being processed. A further possibility is to fabricate hybrid metallic ferromagnetic–SMS structures. In such a case one can make use of the magnetic field exerted by the metallic ferromagnet positioned very close (on an atomic scale) to the SMS layer or quantum structure, thus affecting the spin splitting in the latter. One could use such a fringe field to shape the band edges in such a fashion that an additional localization of the carriers in a two-dimensional quantum well is induced

leading to formation of quantum wires or dots. In other applications a spin-coherence storage effect has been obtained for core electron polarization in Mn-doped ZnSe quantum wells. The optically pumped conduction electron spin polarization generates a nonequilibrium spin polarization of the paramagnetic core electron spins(3d in the case of Mn) through the core–conduction-electron coupling. This core electron spin polarization persists long after the conduction electron spin polarization has decayed away and presents an interesting avenue towards information storage.

The physics of SMSs is currently a well-established part of solid state physics. In the investigations of bulk crystals and LD structures many problems are solved and new problems are waiting further research. Particularly important seems to be investigation of layers and LD structures because of the broad technological possibilities and because of the new important factor (i.e., dimensionality) whereby spin-dependent phenomena can be intentionally regulated.

List of Units and Nomenclature

g	Lande factor for paramagnetic ion
H_{ex}	Heisenberg spin exchange Hamiltonian
J_{d-d}	exchange constant for coupling between paramagnetic ions (transition metals)
$J(\bar{r} - \bar{R})$	exchange constant
J_{sp-d}	exchange constant for electron or hole coupling with paramagnetic ions
LD	low dimensional
M	magnetization
MBE	molecular beam epitaxy
N_0	number of unit cells per unit volume
SMS	semimagnetic (diluted magnetic) semiconductor
$\langle \bar{S} \rangle_{av}$	average value of spin operator
T_c	Curie temperature
V_h	hybridization energy
x	molar fraction of paramagnetic ions
$\boldsymbol{\mu}$	Bohr magneton
$\bar{\boldsymbol{\sigma}}, \bar{S}$	spin operators

See also

Magneto-Optics: Interband Magnetoabsorption, Cyclotron Resonance, Spin Flip Raman Scattering. **Semiconductor Materials:** Lead Salts; Mercury Cadmium Telluride. **Semiconductor Physics:** Spin Transport and Relaxation in Semiconductors; Spintronics.

Further Reading

Bauer G, Pascher H and Zawadzki W (1992) Magneto-optical properties of semimagnetic lead chalcogenides. *Semiconductor Science and Technology* 7: 703–723.

Dietl T (1994) Magnetic semiconductors. Moss TS (ed.) *Handbook of Semiconductors*, vol. 3b, pp. 1251–1342. Amsterdam, The Netherland: North-Holland.

Dietl T (2002) Ferromagnetic semiconductors. *Semiconductor Science and Technology* 17: 377–392. (Special issue: Semiconductor Spintronics, guest ed Hideo Ohno).

Dobrowolski W, Kossut J and Story T (2003) Magnetic semiconductors–new bulk materials and low-dimensional quantum structures. In: Buschow KHJ (ed.) *Handbook of Magnetic Materials, II–VI and IV–VI*. vol. 15, pp. 289–378. Amsterdam: Elsevier North-Holland.

Furdyna JK (1991) In: Averous M and Balkanski M (eds) *Semimagnetic Semiconductors and Diluted Magnetic Semiconductors*. New York: Plenum.

Furdyna JK and Kossut J (eds) (1988) Semiconductors and Semimetals. *Diluted Magnetic Semiconductors*, vol. 25, 470 pp. San Diego, CA: Academic Press.

Gałązka RR and Kossut J (1980) Narrow-gap semimagnetic semiconductors. In: Zawadzki W (ed.) *Lecture Notes in Physics*, vol. 133, pp. 245–265. Berlin: Springer Verlag.

Gałązka RR and Kossut J (1982) Semimagnetic semiconductors. In: Madelung O, Schulz M and Weiss H (eds) *Landolt-Börnstein New Series*, vol. III/17b, pp. 302–318. Berlin: Springer Verlag.

Gałązka RR, Kossut J and Story T (1999) Semimagnetic semiconductors. In: Rössler U (ed.) *Landolt-Börnstein New Series*, vol. III/41B, pp. 650–721. Berlin: Springer Verlag (Update to chapter 5, Vol. III/17b).

Kossut J and Dobrowolski W (1993) Diluted magnetic semiconductors. In: Buschow KHJ (ed.) *Handbook of Magnetic Materials*, vol. 7, pp. 231–305. Amsterdam, The Netherlands: North Holland.

Kossut J and Dobrowolski W (1997) Properties of diluted magnetic semiconductors. In: Capper P (ed.) *Narrow-Gap II–VI Compounds for Optoelectronics and Electromagnetic Applications*, pp. 401–429. London: Chapman and Hall.

Kossut J (2001) Low dimensional structure of diluted magnetic (semimagnetic) semiconductors – a subjective review. *Acta Physica Polonica, A* 100 (supplement): 111–138.

Ohno H, Matsukura F and Ohno Y (2002) Semiconductor spin electronics. *Japan Society of Applied Physics (JSAP International)* 5: 4–13.

Story T (2003) Optoelectronic properties. In: Monasreh MO (series ed.) *Optoelectronic Properties of Semiconductors and Superlattices* vol. 18, Khokholov D. (volume ed.) *Lead Chalcogenides: Physics and Applications, Semimagnetic Semiconductors Based on Lead Chalcogenides*, pp. 385–425. New York, London: Taylor and Francis Books Inc.

GaAs Based Compounds

J J Finley and J P R David, University of Sheffield, Sheffield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Background

Gallium Arsenide (GaAs) is an important semiconductor that has come to dominate the field of optoelectronics by virtue of its favorable electro-optical properties and the ease by which it can be controllably modified by extrinsic means; combining it with its large family of related alloys ($Al_xGa_{1-x}As$, $In_xGa_{1-x}As$, $GaAs_xP_{(1-x)}$, $(In_xGa_{(1-x)})_yAsN_{1-y}$) and via the growth of hetero-structures with reduced dimensionality (Quantum Wells-2D, Wires-1D, and Dots-0D). In such low dimensional semiconductor nanostructures, quantum mechanical effects are directly exploited to customize and dramatically enhance electro-optical properties. Such investigations of the fundamental electro-optical properties of GaAs-based compounds has resulted in the emergence of several new device concepts such as the Quantum Cascade laser, led to the discovery of new physical phenomena like the fractional Quantum Hall effect, and may pave the way for an entirely new class of opto-electronic devices with true quantum functionality.

In the following, we explore the fundamental physical and electro-optical properties of GaAs and its related alloys and illustrate the considerable impact this remarkable family of materials has had on modern semiconductor devices.

Fundamental Physical Properties

The crystal structure of GaAs (Figure 1 – inset) consists of two interpenetrating face centred cubic lattices that accommodate the group III and V atoms respectively. The sub-lattices of Ga and As are shifted from each other by $a/4$ along the [111] body diagonal, where a is the lattice constant, forming the well-known zincblende crystal structure. For stochiometric GaAs, the lattice constant is $a = 0.56536(1)$ nm at room temperature which translates to a Ga–As nearest neighbor separation (bond length) of 0.2308 nm.

Figure 1 summarizes the variation of the lattice constant for a number of GaAs related III–V semiconductor alloys, plotted as a function of their fundamental electronic bandgap. Particularly relevant is AlAs, which also crystallizes in the zincblende form and has a lattice constant (a(AlAs) = 0.5660(2) nm) almost identical to that of GaAs ($\Delta a < 1\%$). The sole fact that the $Al_xGa_{(1-x)}As$ alloy can be almost perfectly lattice matched to GaAs over its entire compositional range, whilst exhibiting near continuously variable electronic and optical properties, is the principal reason for the key role GaAs-related materials play in semiconductor opto-electronic devices. We continue by investigating the electronic and optical properties of the $Al_xGa_{(1-x)}As$ family of materials and discover why they are particularly suited to opto-electronic device applications.

Summary of Electronic Properties of GaAs and $Al_xGa_{1-x}As$

The optical properties of semiconductors are determined principally by inter-band radiative transitions involving electronic states in the vicinity of the conduction (CB) and valence band (VB) extrema. Consequently, in the following we focus on electronic properties of GaAs close to these points and substantiate our discussion using experimental data obtained from optical spectroscopy.

GaAs Bandstructure

GaAs is a direct-gap semiconductor ($E_g = 1.42$ eV at 300 K), both the conduction band minimum and

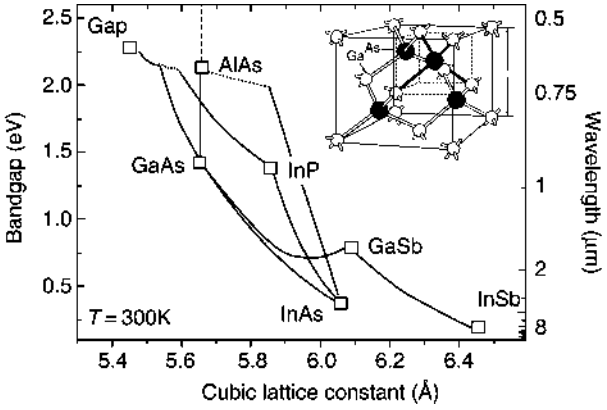

Figure 1 Fundamental bandgap of selected compounds within the III–V semiconductor family plotted as a function of the cubic lattice constant – a. The inset depicts the zincblende crystal structure within the GaAs unit cell. Data obtained from various sources, see for example Brozel MR and Stillman GE (eds) (1996) *Properties of Gallium Arsenide*, INSPEC, emis Datareviews Series No.16 and Loffe Institute, on NSM semiconductor data repository at http://www.ioffe.rssi.ru

valence band maximum occurring at the center of the Brillouin zone (Γ-point). Figure 2 summarizes our discussion of the electronic structure of GaAs, depicting the bandstructure schematically within the first Brillouin zone. The valence electron configurations of Ga and As atoms result in the band edge CB states being two-fold degenerate due to the electron spin and characterized by an almost isotropic effective mass of $m_\Gamma^* = 0.068\, m_o$. For GaAs heterostructures with reduced dimensionality, the very weak anisotropy of the conduction band states can generally be ignored when calculating the lowest energy confined electron states but has a significant impact on higher quantized levels and must, therefore, be included. In addition to the fundamental CB minimum at the zone center or Γ-point of the Brillouin zone, satellite minima occur at the six equivalent X-points ($k = (1/a, 0, 0)$) and eight equivalent L-points ($k = (1/a, 1/a, 1/a)$). The X-point and L-point minima lie $\Delta E_{\Gamma X} = 0.49$ eV and $\Delta E_{\Gamma L} = 0.32$ eV above the Γ-minimum at low temperatures ($T = 4.2$ K) and are characterized by strongly direction dependent effective masses parallel (‖) or perpendicular (⊥) to the principal axes. Electronic states in the vicinity of these minima are termed the X and L-valleys respectively and have a significant impact on high-energy electron transport in devices such as HEMTs and in avalanche carrier multiplication processes, but do not generally play a significant role in the optical properties of the binary – GaAs. However, satellite CB minima can strongly influence the optical properties of GaAs–AlAs superlattices and QWs that can exhibit a type-II

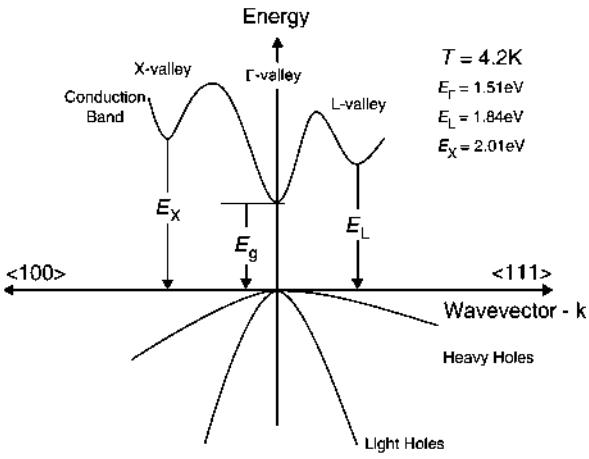

Figure 2 Schematic representation of the bandstructure of GaAs showing just a single conduction band over whole Brillouin zone in Γ-L and Γ-X directions only and neglecting the split off band that does not significantly influence electrical or optical properties.

Table 1 Low temperature GaAs effective masses for electrons (upper panel) and holes (lower panel). Values marked with * are inferred from measurements on GaAs based heterostructures and values marked ** are calculated

	Crystallographic axis	*Effective Mass (m_o)*
CB-Γ minimum	*Isotropic*	0.068(2)
CB-X minimum*	‖ [100]	1.9
	⊥ [100]	0.19
CB-L minimum**	‖ [111]	1.9
	⊥ [111]	0.075
VB-HH	[111]	0.72(3)
	[100]	0.33(6)
VB-LH	[111]	0.08(3)
	[100]	0.09(4)

band alignment for GaAs layer thicknesses < 6 nm. In this case, the electrons (holes) are localized in the AlAs (GaAs) layer and the nature of the AlAs X-states dominates the optical properties.

The valence band (VB) states of GaAs are four-fold degenerate at the bandedge and, characteristic of most zincblende semiconductors the valence bandstructure, is significantly more strongly anisotropic than the CB states discussed above. Away from the Γ-point, the four-fold degeneracy of the valence band states is lifted into 2 doubly degenerate sets of bands termed the heavy (HH) and light hole (LH) bands by virtue of their zone center effective masses (Table 1). The HH band is considerably more strongly anisotropic than the LH band and is characterized by a much larger effective mass along the [100] direction when compared with [001], $m_h^*[111] > m_h^*[001]$. The anisotropy of the LH states is much weaker than for the HH band, with approximately equivalent effective masses along [100] and [111] crystallographic directions. Typical values for the CB and VB effective masses are summarized in Table 1, obtained using a variety of experimental techniques at low temperature.

Temperature Dependence

Quantitative values given above relate specifically to the bandstructure at liquid helium temperatures ($T = 4.2$ K). Increasing temperature results in a narrowing of many of the conduction–valence band gaps discussed above and lowering of the effective masses. This temperature dependence of the fundamental bandstructure arises from the anharmonic dilation of the lattice and from interactions with phonons, the population of which becomes significant at elevated temperature. In the simplest picture, one can visualize the effect of increasing temperature

Table 2 Varshnii coefficients for the Γ, L, and X CB minima in GaAs relative to the top of the VB. For the Γ minimum a number of experimental works are available and the value cited is the average

	α(10^{-4} eV/K)	β (K)
Γ point	5.45	210
X point	7.20	205
L point	6.05	204

as a shrinking of the first Brillouin zone as the lattice constant increases, and with a subsequent increase of the band curvature leading to ~5–10% lighter effective masses at room temperature than at 4.2 K. In addition, the dramatic increase of the phonon population at elevated temperatures results in a narrowing of the fundamental bandgaps. Phenomenologically, the temperature dependence of the fundamental bandgaps can be well described by the well-known Varshnii relation

$$E_g(T) = E_g^{T=0\ \mathrm{K}} - \frac{\alpha T^2}{(T+\beta)} \qquad [1]$$

where α and β are experimentally determined constants which are sensitive to the bandgap of interest (e.g. Γ, X, or L, etc.) as summarized in Table 2.

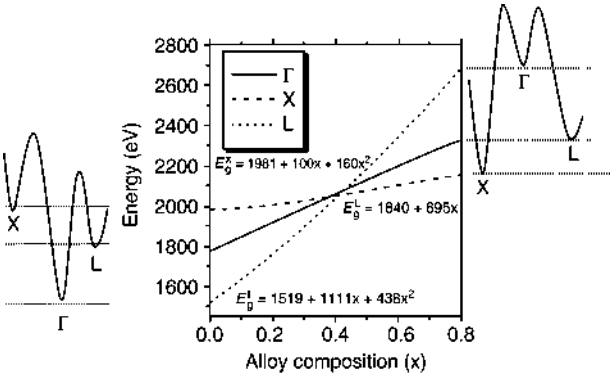

Figure 3 The absolute energy (at 4.2 K) of the Γ, X and L CB minima for $Al_xGa_{(1-x)}As$ as a function of aluminum composition up to $x \sim 0.8$. Higher compositions are not given due to the well-known unreliability of experimental data due to instability of high Al content $Al_xGa_{(1-x)}As$ due to oxidation. Polynominals describing each curve are also plotted. Data taken from the following sources: Γ-minimum, Bosio C, Staehli JL, Guzzi M, Burri G and Logan RA (1988) Direct-energy-gap dependence on Al concentration in $Al_xGa_{1-x}As$. *Physical Review* 38: 3263–3268, X-minimum, Guzzi M, Grilli E, Oggioni S, Staehli JL, Bosio C and Pavesi L (1992) Indirect-energy-gap dependence on Al concentration in $Al_xGa_{1-x}As$ alloys. *Physical Review* B45: 10951, and L-minimum, Henning JCM, Ansens JPM and Roksnoer PJ (1986) Spectroscopic determination of L6 conduction-band minima in $Al_xGa_{1-x}As$. *Journal of Physical Chemistry*, L335.

GaAs/$Al_xGa_{(1-x)}As$ – The Most Significant Alloy

Incorporation of aluminum into the GaAs crystal to form $Al_xGa_{(1-x)}As$ results in a number of profound and important modifications of the electronic properties compared with pure GaAs. Aluminum is more strongly electronegative than Gallium and consequently the electron charge distribution along the Al–As bonds are more strongly localized towards the group-III site than for Ga–As bonds which qualitatively results in a widening of the fundamental bandgap with increasing aluminum composition (x).

The variation of the $Al_xGa_{(1-x)}As$ energy gaps for the Γ, X, and L conduction band minima at low temperature are summarized in Figure 3 as a function of the alloy composition up to $x \sim 0.8$. The data of Figure 3 reflect the results of optical emission and absorption measurements and are most reliable for more dilute compositions since crystalline AlAs and high Al content $Al_xGa_{(1-x)}As$ oxidize rapidly under atmospheric conditions.[1] With increasing Al content, the Γ-minimum shifts rapidly to higher energy whilst the shift of the L and X-minima is significantly less pronounced. At $x \sim 0.4$ the Γ and X minima intersect and for $x > 0.4$, the lowest energy conduction band minimum shifts to the X-point and the fundamental bandgap becomes indirect. This effect is responsible for the short wavelength cutoff of $Al_xGa_{(1-x)}As$/GaAs based multi-quantum well LEDs at $\lambda \sim 630$ nm since the transition to an indirect bandgap is accompanied by a drastic reduction of the radiative efficiency. Shorter wavelength LED operation region of the spectrum is possible by moving to other related alloys in the III–V family such as $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ (520–630 nm) and $(GaAs)_{0.6}P_{0.4}$ (~600 nm). We return to discuss the optical properties of $Al_xGa_{(1-x)}As$ below in Section 4.

As expected, the pronounced modifications of the bandgap induced by Al incorporation are accompanied by modifications of the carrier effective masses and temperature dependence of the bandgap. These compositional dependence of the Γ, X, and L conduction band and HH and LH valence band effective masses are summarized in Table 3.

It should be stressed that the relationships given in Table 3 should be considered to be accurate at the

[1] The ease by which high Al content $Al_xGa_{(1-x)}As$ can be oxidized is of great technological benefit for device fabrication. Examples include the formation of current apertures in injection lasers, $Al_xGa_{(1-x)}As$ etch stops etc. For a summary Brozel MR and Stillman GE (eds) (1996) *Properties of Gallium Arsenide*, INSPEC, emis Datareviews Series No.16.

Table 3 Compositional variation of the electron and hole effective masses for $Al_xGa_{(1-x)}As$ taken from various sources: Data drawn from the following sources: Lautenschlager Γ-P, Garriga M, Logothetidis S and Cardona M (1997) *Physical Review* B35: 9174 and Grilli E , Guzzi M and Zamboni R (1992) *Physical Review* B45: 1638. Thurmond CD (1975) L-minimum. *Journal Electrochemical Society*, 122: 1133 and Aspines DE (1976) L-Minimum. *Physical Review* B14: 5331

	Crystallographic Axis	*$Al_xGa_{(1-x)}As$ Effective Mass (m^0)*
CB-Γ minimum	*Isotropic*	$0.0682 + 0.0835x$
CB-X minimum	∥ [100]	$1.9 - 1.02x$
	⊥ [100]	$0.19 + 0.06x$
CB-L minimum	∥ [111]	$1.9 - 0.58x$
	⊥ [111]	$0.075 + 0.075x$
VB-HH	DOS	$0.62 + 0.14x$
VB-LH	DOS	$0.087 + 0.09x$

$\sim 5-10\%$ level, since reliable experimental data only exists for the GaAs binary endpoint and more dilute alloy compositions ($x < 0.3$).

The Varshnii coefficients of the temperature dependence of the bandgap vary with $x(0 < x < 0.5)$ according to $\alpha(x) = 5.5 + 3.35x(10^{-4}\ \text{eV/K})$ and $\beta(x) = 225 + 88x(\text{K})$ respectively. Recent photoreflectance measurements have indicated that these simple linear relationships do not hold well over the entire compositional range with evidence for some bowing of $\alpha(x)$ and $\beta(x)$ closer to the AlAs binary endpoint.

GaAs/$Al_xGa_{(1-x)}As$ Hetero-interfaces

Modern epitaxial growth techniques now enable the routine fabrication of semiconductor heterostructures consisting of a well-defined sequence of atomically precise layers of different materials, which together manipulate the opto-electronic properties to suit a particular device or application. Of particular technological relevance is the GaAs/$Al_xGa_{(1-x)}As$ hetero-interface since both materials are lattice matched for $0 < x < 1$ (Figure 1), and multiple layers of GaAs and $Al_xGa_{(1-x)}As$ can be readily deposited without introducing defects into the crystal which seriously degrade optical performance. In this section, we summarize the most salient features of the GaAs/$Al_xGa_{(1-x)}As$ hetero-interface, with particular emphasis on the electronic properties before continuing to discuss the optical properties of $Al_xGa_{(1-x)}As$ in Section 4.

The most important parameters that characterize a semiconductor hetero-interface are the magnitude and sign of the conduction (ΔE_c) and valence band (ΔE_v) offsets. These parameters determine how effectively electrons and hole are localized within specific epilayers and are related to the total difference in the bandgap ($\Delta E_g = E_g^{AlGaAs} - E_g^{GaAs}$) via $\Delta E_g = |\Delta E_c + \Delta E_v|$. Taking the mean of 21 separate optical and electrical investigations between 1985 and 1995, the currently accepted value for the band-offset ratio is $\Delta E_c/\Delta E_v \approx 64/36$, with an associated variance over the body of the data of approximately $\pm 8\%$. It should be stressed that the majority of the experimental investigations of the band offsets at the GaAs–$Al_xGa_{(1-x)}As$ hetero-interface have been performed for alloy compositions $x < 0.4$ and much less experimental data exists regarding the variation of $\Delta E_c : \Delta E_v$ with higher Al composition. A rather old review (*c.* 1992), collating data available, at the time has indicated that the absolute energies of the conduction and valence band offsets vary with composition as $\Delta E_c = (800 \pm 30)x$ meV for $(x < 0.4)$ and $\Delta E_c = (510 \pm 40)x\,(0 < x < 1)$. It is widely assumed that the band offsets do not change appreciably with temperature.

Optical Properties of GaAs and Significant Alloys

Optical Absorption

Monochromatic light of energy $E_p = \hbar\omega$ incident on the surface of GaAs may stimulate interband optical transitions in which a photon is absorbed and an electron from the valence band is promoted into the conduction band. In processes involving a single photon, the photon momentum is entirely negligible compared with the electron and optical transitions are vertical in k-space.

Following excitation, the electron–hole (e–h) pairs relax extremely rapidly ($\tau_{LO} \sim 150$ fs) by sequential emission of multiple longitudinal optical (LO) phonons until their energy becomes close to the bandedge and further LO-phonon emission is forbidden by energy conservation requirements. This initial phase of the carrier relaxation can be extremely rapid, the LO-phonon scattering time in GaAs being due to the semi-polar nature of the crystal. After this initial

rapid relaxation phase, the electron and hole populations relax principally via acoustic phonon emission ($\Delta E < 2$ meV) over much longer timescales ($\sim 10-300$ ps), which depend upon the lattice temperature, carrier type and density. Finally, the electron and hole populations form thermal distributions via carrier–carrier and Auger scattering over longer timescales up ~ 500 ps. During these final relaxation and thermalization phases, e–h pairs are depleted from the conduction and valence bands by interband radiative recombination.

For energies less than the bandgap, the semiconductor normally appears transparent, although the presence of impurities and defect levels can result in a weak absorption 'tail' below the gap. Optical absorption by impurities results in only electrons in the conduction band or holes in the valence band. The wavelength of light that just causes the creation of an electron–hole pair is given by $\lambda = hc/E_g$, where h is Planck's constant, c the speed of light and E_g the fundamental bandgap. As described in Section 3.1, this corresponds to approximately $\lambda = 870$ nm at room temperature, depending on the type and level of the doping. Light with wavelength much shorter than this, corresponding to energies much larger than E_g, will create electron–hole pairs higher up the band structure with higher energy and be strongly absorbed. These carriers, however, will rapidly lose this excess energy by interacting with the lattice and relax to the minima of the conduction and valence bands.

The optical absorption is usually quantified by the absorption coefficient, α, which is a function of the photon energy or wavelength and is shown for GaAs in Figure 4. The fraction of the light intensity that is absorbed in the semiconductor after passing through a thickness x is given by $f = (1 - \exp(-\alpha x))$ and the fraction of light intensity that is transmitted is then simply $\exp(-\alpha x)$. α is usually expressed in units of cm^{-1} and at energies above the bandgap has values typically between 10^4–10^5 cm^{-1}. Light with wavelength corresponding to energies above the bandgap is usually absorbed very efficiently (>90%) within a few microns. Increasing the bandgap, say by adding aluminum, would have the effect of making the material more transparent at a given wavelength.

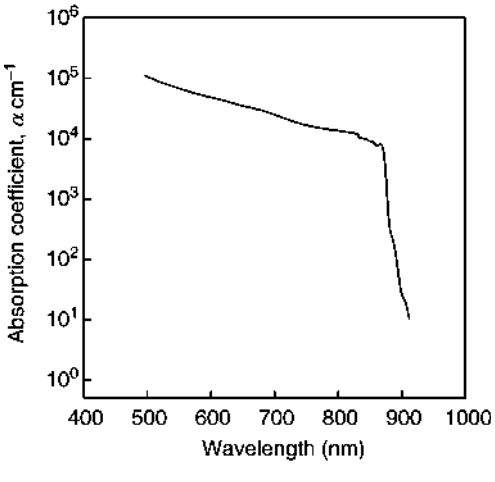

Figure 4 Variation of the absorption coefficient, α, as a function of wavelength in GaAs at room temperature.

Optical absorption is an important process as it allows light to be converted to free carriers, and hence a current in the semiconductor. This forms the basis of optical detection by most types of semiconductor photodetectors.

Refractive Index

The speed (more precisely the phase velocity) of light, of all wavelengths, in a vacuum is constant at $\sim 3.10^8$ m s^{-1}. The speed of light in a medium such as a semiconductor is slower due to the microscopic interaction of the electromagnetic wave with the crystal. The refractive index, R_I, is defined as the ratio of the speed in a vacuum to the speed in the semiconductor. Increasing the density of the medium decreases the speed. Changing the semiconductor, for example by adding aluminum, not only increases the bandgap as discussed above, but also decreases R_I for a given wavelength.

At wavelengths corresponding to an energy just below the bandgap of the semiconductor, the material is essentially transparent, i.e., the absorption coefficient is low.

Light can be made to travel in a waveguide comprising this semiconductor, provided that it was surrounded by semiconductor material of a lower R_I. Light is totally internally refracted at the interface between the GaAs and $Al_xGa_{1-x}As$ and so cannot escape. The semiconductors with a smaller R_I usually have a larger bandgap, so are even more transparent at the operating wavelength and help to physically confine the electrons and holes within the GaAs.

The R_I can be changed by injecting charge (electrons or holes) into the semiconductor, by changing the temperature (which changes the bandgap), and by applying an electric field. This forms the basis of operation of devices such as optical modulators. Unfortunately it also can cause 'chirping', i.e., the unwanted change in light

intensity or wavelength as the charge or temperature changes.

The R_I which is a function of the wavelength, is sometimes expressed as a complex number:

$$R_I = n + ik \quad [2]$$

where n is the real part and k is the imaginary part called the extinction coefficient, which is related to the absorption coefficient. Typical room temperature values of n for GaAs vary from about 3.37 at ~1.55 μm to 3.87 at 620 nm.

Common GaAs Based Alloys for Novel Technological Applications

GaInP/AlGaInP – 'Red, Orange and Green Emission'

The alloy $(Al_xGa_{1-x})_{0.52}In_{0.48}P$ is becoming technologically important as it has the same lattice parameter as GaAs for all values of x, so can grow lattice matched on GaAs. This alloy system offers the largest direct bandgap of virtually any III–V (excluding the Nitride family), so has attracted considerable interest for visible lasers and LEDs in the wavelength range 530 nm–670 nm. At room temperature the direct Γ bandgap increases almost linearly from ~1.85 eV for $Ga_{0.52}In_{0.48}P$ to ~2.3 for $(Al_{0.5}Ga_{0.5})_{0.52}In_{0.48}P$. For values of $x > 0.5$, the bandgap becomes indirect as in $Al_xGa_{1-x}As$ ($x > 0.4$) when the X-level indirect bandgap becomes lower than the Γ-level direct bandgap. Earlier work in this alloy suggested that the point at which it became indirect was at $x = 0.66$. The X-level bandgap appears to be relatively insensitive to the aluminum composition and has a value of ~2.3 eV at room temperature.

GaAs/GaInAsN – '1.3 μm and 1.55 μm for telecommunication applications'

One limitation of GaAs is that most of the compounds that are lattice matched to it, such as $Al_xGa_{1-x}As$ or $(Al_xGa_{1-x})_{0.52}In_{0.48}P$, have bandgap energies that are large, making them unsuitable for telecommunication applications which require narrow bandgap materials capable of operating in the range 1.3 μm to 1.55 μm. Adding indium (In) to GaAs can decrease the bandgap but also causes strain as the lattice parameter increases, which limits the amount of indium and total thickness of the structure.

Recently, it has been shown that while adding In to GaAs increases the lattice parameter, adding N to GaAs decreases it. Consequently GaInAsN can be made to be lattice matched to GaAs by controlling the In and N content. Unlike conventional alloy semiconductors, adding In or N both result in a reduction of the bandgap energy. Furthermore, GaInAsN is found to be a direct bandgap material with a type I band lineup when combined with $Al_xGa_{1-x}As$, which is a wide bandgap material also lattice matched to GaAs. These properties make GaInAsN ideally suited for lasers and there have been several reports of devices operating at 1.3 μm in the literature.

See also

Semiconductor Physics: Quantum Wells and GaAs-based Structures.

References

Bosio C, Staehli JL, Guzzi M, Burri G and Logan RA (1988) Direct-energy-gap dependence on Al concentration in $Al_xGa_{1-x}As$. *Physical Review* 38: 3263–3268.

Brozel MR and Stillman GE (eds) (1996) *Properties of Gallium Arsenide*, INSPEC, emis Datareviews Series No.16.

Data drawn from the following sources: Lautenschlager Γ-P, Garriga M, Logothetidis S and Cardona M (1997) *Physical Review* B35: 9174 and Grilli E, Guzzi M and Zamboni R (1992) *Physical Review* B45: 1638. Thurmond CD (1975) L-minimum. *Journal Electrochemical Society*, 122: 1133 and Aspines DE (1976) L-Minimum. *Physical Review* B14: 5331.

Guzzi M, Grilli E, Oggioni S, Staehli JL, Bosio C and Pavesi L (1992) Indirect-energy-gap dependence on Al concentration in $Al_xGa_{1-x}As$ alloys. *Physical Review* B45: 10951.

Henning JCM, Ansens JPM and Roksnoer PJ (1986) Spectroscopic determination of L6 conduction-band minima in $Al_xGa_{1-x}As$. *Journal of Physical Chemistry*, L335.

Ioffe Institute, on NSM semiconductor data repository at http://www.ioffe.rssi.ru/SVA/NSM/Semicond/index.html

Landolt-Börnstein (1986) *Numerical data and functional relationships in science and technology*. In: Madelung O (ed.) *Semiconductors*, vol 22. Berlin: Springer Verlag.

Nakwaski W (1995) InAs, AlAs and their ternary compounds. *Physica* B210: 1–25.

Varshni YP (1967) Temperature dependence of the energy gap in semiconductors. *Physica* B34: 149.

Group IV Semiconductors, Si/SiGe

D J Paul, University of Cambridge, Cambridge, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

SiGe is now a mature materials system with electronic circuits on the market place. While no optical devices are available yet, there is a substantial amount of research in the field. In this article, the fundamental materials properties will be reviewed. As SiGe is a strained system, we shall start with the basic material properties as this determines the Ge content and layer thicknesses which are available for use. The band structure and electronic behavior will then be reviewed before the optical properties of the material.

The main drive for SiGe research in optical systems is the ability to use the extensive Si processing knowledge and to integrate a number of components onto a Si chip. Conventional Si transistors may also be used as driver circuits for the optical components, reducing chip counts in systems and therefore reducing costs. As Si processing is substantially cheaper than other semiconductor materials, if SiGe devices can be made on Si processing lines, they have the potential to be substantially cheaper than comparable III–V or II–VI devices. The SiGe material properties that will be reviewed in this article should be considered as an expansion of bulk silicon.

Material Properties and Strain

Ge has a 4.2% larger lattice constant that Si (Table 1). When growing epitaxial films of germanium or the alloy $Si_{1-x}Ge_x$ $(0 < x < 1)$ on a silicon substrate, there exists a maximum thickness called the critical thickness above which it costs too much energy to strain additional layers of material into coherence with the substrate. Instead defects appear, in this case misfit dislocations, which act to relieve the strain in the epitaxial film. Relaxed Si, Ge or SiGe films form diamond, diamond and zincblende lattices respectively while strained heterolayers below the critical thickness form a tetragonal symmetry arrangement. The strain therefore will substantially modify the electronic and optical properties of the material as will be discussed later.

Most SiGe films are produced by either chemical vapor deposition or molecular beam epitaxy. Both these techniques allow the deposition of films at low temperatures (typically 670–970 K) which is essential to prevent strained layers relaxing and forming defects. Chemical vapor deposition uses a number of different gases such as silane, disilane, dichlorosilane, and germane (occasionally with a hydrogen carrier gas) as the main species, while phosphine or arsine are used as n-type dopants and diborane is the typical p-type dopant. With the high melting point of both silicon and germanium, molecular beam epitaxy requires electron beam evaporators for the host materials silicon and germanium while effusion cells of antimony (occasionally phosphorus) and boron are used for n- and p-type dopants, respectively.

There are a number of models used to calculate the equilibrium critical thickness. The most often quoted is the Matthews and Blakeslee model (Figure 1). Experimentally it has been shown that defect-free films with thicknesses above the critical thickness may be grown (Figure 1) although these layers are metastable and will relax if temperatures above the

Table 1 Materials properties of Si, Ge and relaxed $Si_{1-x}Ge_x$

Element		*Si*	*Ge*	*Bulk relaxed $Si_{1-x}Ge_x$*
Lattice		Diamond	Diamond	Zincblende
Lattice constant	a_o (nm)	0.5431	0.5657	$0.5431 + 0.0226x$
Lattice mismatch	η_o (%)	0	4.17	$0.04115x - 0.02500x^2 + 0.05826x^3 - 0.03379x^4$
Density	(kg m^{-3})	2329	5323	$2329 + 2994x$
Sound velocity	m s^{-1} (300 K)	8440	4900	$8440 - 3540x$
Poisson s ratio	ν	0.280	0.273	$0.280 - 0.007x$
Melting point	(K)	1685	1210	
Elastic moduli	c_{11} (GPa)	165.8	128.5	
Elastic moduli	c_{12} (GPa)	63.9	48.3	
Elastic moduli	c_{44} (GPa)	79.6	66.8	
Thermal conductivity	κ (W m^{-1} K^{-1}) (300 K)	140	60	
Dielectric constant	ε	11.9	16.2	$11.9 + 4.1x$

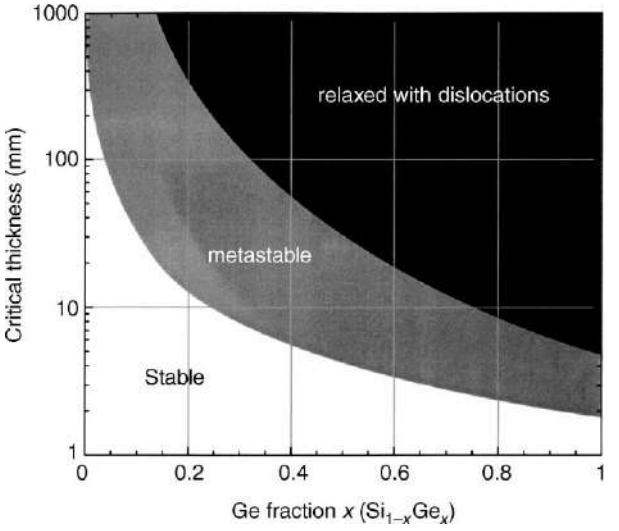

Figure 1 The critical thicknesses for growing a strained $Si_{1-x}Ge_x$ epitaxial layer on a bulk silicon substrate. The white area corresponds to the region below the equilibrium critical thickness as calculated with the Matthews and Blakeslee model while the gray region corresponds to a metastable layer which may relax after thermal treatments. The black area will have dislocations and strain relaxation using any growth technique or temperature.

growth temperature are applied to the heterolayers. For any manufacture, the layers cannot be metastable as high-temperature processes are normally employed in processing. The Matthews and Blakeslee equilibrium critical thickness, h_c for strained-$Si_{1-x}Ge_x$ on bulk Si is given by

$$h_c = \frac{0.55}{x}\ln[10h_c] \qquad [1]$$

while an approximate value can be obtained from

$$h_c \approx 1.7793x^{-1.2371}(\text{nm}) \qquad [2]$$

Figure 1 clearly demonstrates that for thick SiGe films, low Ge contents must be used.

There is also a substantial interest in producing low-defect relaxed $Si_{1-y}Ge_y$ films or strained-Si on a relaxed $Si_{1-y}Ge_y$ substrate which can provide bandgaps and band discontinuities not available to pseudomorphically grown alloy layers. Ideally the relaxed $Si_{1-y}Ge_y$ substrate is grown on a Si substrate as this allows integration with conventional complementary metal oxide semiconductor (CMOS) electronics. To date no high-quality bulk single-crystal $Si_{1-y}Ge_y$ substrates are available due to the strong segregation of Si and Ge phases as liquid SiGe crystallizes. Growing a thick $Si_{1-x}Ge_x$ layer on Si well above the critical thickness so that it relaxes produces a threading dislocation density of about 10^{12} cm^{-2} which is far too large to produce device-quality material for electronics or optics. By growing a linearly graded $Si_{1-y}Ge_y$ with the grading rate sufficiently low (typically 10% per μm), the strain relaxation mechanism changes from a nucleation relaxation mechanism to a glide relaxation mechanism resulting in dislocations with comparatively long misfit segments. The defect densities are typically 10^5–10^6 cm^{-2} (Figure 2) although recent thermal processes and other techniques have now reduced this to 100 cm^{-2}.

Substantial work has appeared in the literature on self-organized growth of islands. Such islands are of interest in many materials systems as the quantum confinement they may offer can potentially enhance luminescence properties in numerous opto-electronic devices. The dots are formed by the Stranski–Kranstanow growth mode in a strained system. Planar epitaxial layers grown under stress conditions can lower their elastic energy by morphological changes. For the growth of strained-Ge or $Si_{1-x}Ge_x$ on bulk Si, this corresponds to the formation of islands when the growth parameters in the system are correctly set. While the islands may have well-defined shapes, the formation of the dots relies on random nucleation followed by ripening and coalescence. These mechanisms have produced large distribution of sizes (up to 30% of the mean size), densities and positions. It is quite common to have distributions of two or more sizes of islands and the densities are typically of order of 10^8–10^9 islands per cm^2. While the heights of these islands may be as low as 3 nm, the

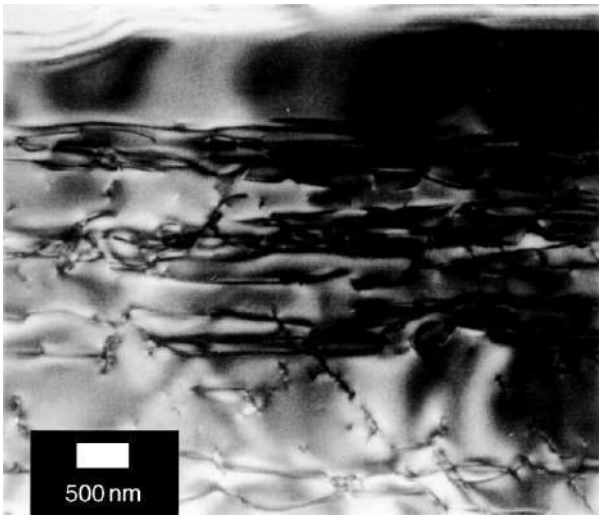

Figure 2 A transmission electron micrograph of a strain relaxation buffer grown by DERA, Malvern. The dense region of dislocations corresponds to an approximately 4 μm thick $Si_{1-y}Ge_y$ layer which is graded from $y = 0$ to $y = 0.23$. Above this is a 1 μm thick $Si_{0.77}Ge_{0.23}$ buffer which has no dislocation segments threading to the surface.

widths of the islands are typically 20 nm or greater, which is not yet small enough to produce the quantum confinement properties useful for significant improvements in opto-electronic devices.

Electronic Properties and Band Structure

As many of the optical properties and responses are derived or result from the band structure of materials, this section will concentrate on the band structure and electronic properties of SiGe layers. The bandgap of SiGe whether strained or unstrained is always below that of bulk Si. This potentially allows applications at longer wavelengths than comparable Si devices. Figure 3 shows the appropriate bandgaps for both the unstrained and strained $Si_{1-x}Ge_x$ grown on relaxed Si. For the unstrained case with a Ge fraction $x < 0.85$, the band structure is Si-like with six Δ-valleys forming the conduction band (Figure 4a). For $x > 0.85$ the band structure is Ge-like with four L-valleys. When strain is applied, both the conduction band and the valence band are split. Figure 4b shows the splitting of the conduction band with the Δ_4 valleys lower in energy and the Δ_2 valleys higher. In addition the valence band is split with the heavy-hole band being higher in energy than the light-hole band (Figure 3). The split-off band is also moved further away from the light and heavy hole bands as the strain in the system is increased. The indirect energy gap for the strained $Si_{1-x}Ge_x$ at 4.2 K as measured by photoluminescence is given by

$$E_g = 1.155 - 0.43x + 0.0206x^2\,[\text{eV}] \text{ for } x < 0.85 \quad [3]$$

$$E_g = 2.010 - 1.27x\,[\text{eV}] \text{ for } x > 0.85 \quad [4]$$

while the direct bandgap is given by

$$E_g = 3.395 - 1.287x + 0.153x(1-x)\,[\text{eV}]\,(300\,\text{K}) \quad [5]$$

Due to the critical thickness, measurements of the bandgap of strained $Si_{1-x}Ge_x$ grown pseudomorphically on bulk Si substrates has been limited to low Ge concentrations. Using photoluminescence measurements at 4.2 K, Robbins has obtained a value for the exciton bandgap for strained $Si_{1-x}Ge_x$ layers for $x < 0.25$ to be

$$E_g = 1.155 - 0.874x + 0.376x^2\,[\text{eV}] \quad [6]$$

The bandgap for any strained $Si_{1-x}Ge_x$ layer grown on a relaxed $Si_{1-y}Ge_y$ substrate is shown in Figure 5. If the active strained $Si_{1-x}Ge_x$ layer is compressively strained then the conduction band is split into Δ_4 valleys being lower in energy and the Δ_2 valleys being higher (Figure 4b). For tensile strain (such as strained Si on a relaxed $Si_{1-y}Ge_y$ substrate), the Δ_2 valleys are lower in energy and the Δ_4 valleys are higher (Figure 4b).

The band alignments for typical heterolayer structures are shown in Figure 6. Pseudomorphically grown strained $Si_{1-x}Ge_x$ layers are only useful for forming quantum wells in the valence band (Figure 6a). Photoluminescence measurements do suggest a type I band alignment for these structures (quantum wells in both the conduction and valence bands) although it was later shown that the material is type II but due to charges producing band bending, the photoluminescence wrongly suggests a type I transition. For a quantum well for electrons, strained Si layers are grown on a relaxed $Si_{1-y}Ge_y$ substrate. In addition, quantum wells in the valence band can be produced for strained $Si_{1-x}Ge_x$ layers with high Ge content grown on relaxed $Si_{1-y}Ge_y$ substrates for $x > y$. In almost all cases the band alignments are type II although a small region exists where theory suggests type I is possible but this has not been demonstrated experimentally at present.

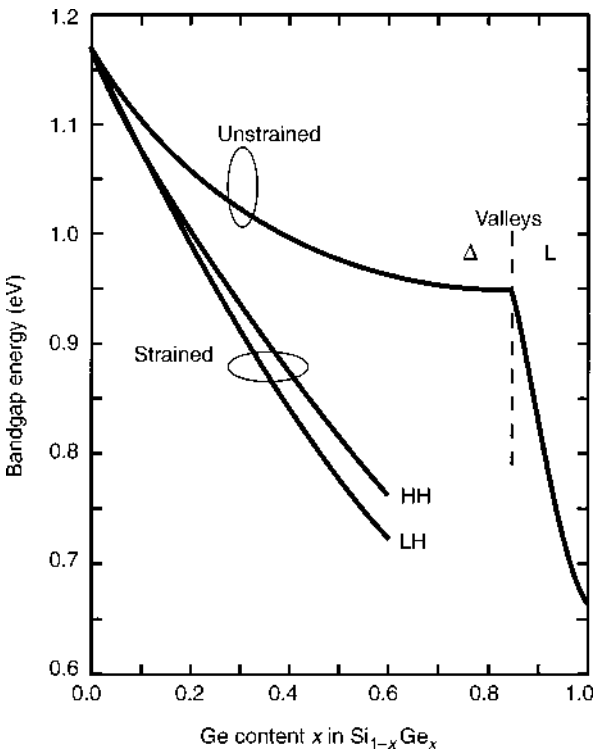

Figure 3 The bandgap energy for strained $Si_{1-x}Ge_x$ on a relaxed Si substrate and unstrained $Si_{1-x}Ge_x$. For the unstrained there is a discontinuity at $x \sim 0.85$ where the band structure changes from Si-like with Δ-valleys to Ge-like with L valleys (see Figure 4).

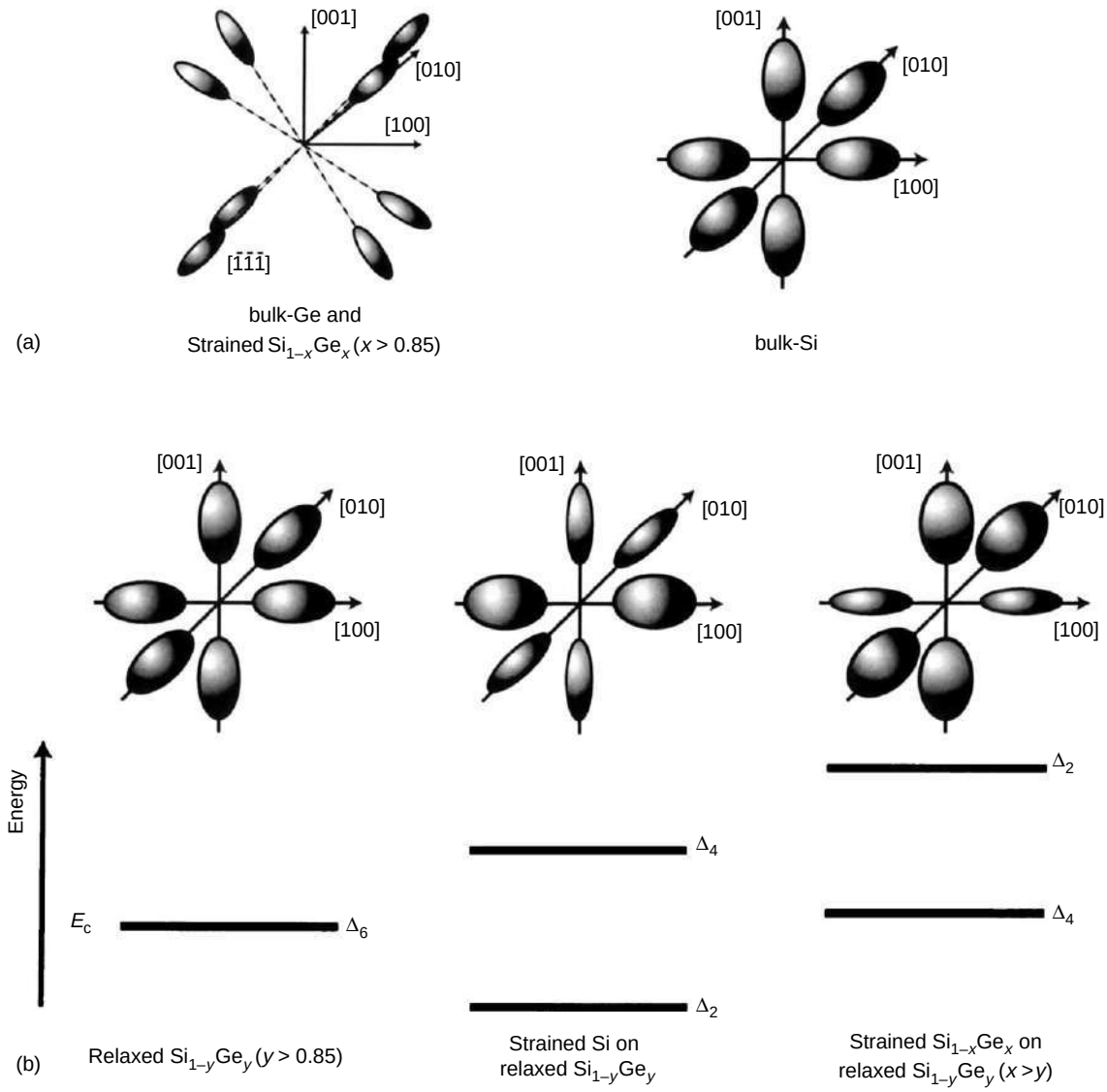

Figure 4 (a) A schematic diagram of the valleys for Ge-like and Si-like material. (b) A schematic diagram of the valley structures and energy levels of the conduction bands for relaxed, tensile strained (strained Si on relaxed $Si_{1-y}Ge_y$) and compressively strained (strained $Si_{1-x}Ge_x$ on relaxed $Si_{1-y}Ge_y$ ($x > y$)) material.

The valence band discontinuity is given by

$$\Delta E_v = (0.74 - 0.06y)(x - y)\,[\text{eV}] \qquad [7]$$

while the conduction band discontinuity for strained Si on a relaxed $Si_{1-y}Ge_y$ substrate is

$$\Delta E_c = 0.6y\,[\text{eV}] \qquad \text{for } x < 0.6 \qquad [8]$$

The complete set of ΔE_c and ΔE_v values for all x and y are shown in Figures 7 and 8 respectively. Table 2 gives some of the electronic data on Si, Ge, and relaxed $Si_{1-x}Ge_x$ including mobilities, bandgaps, and effective masses.

The mobilities for electrons and holes in relaxed $Si_{1-x}Ge_x$ layers are below the bulk Si and Ge layers due to alloy scattering (Table 2). For strained layers, mobility enhancements have been demonstrated in a number of different systems. Electron mobility enhancements result from the splitting of the valleys reducing intervalley scattering. Hole mobility enhancements result from the splitting of the valence band and reduction of effective masses. The hole mobilities for pseudomorphically grown strained $Si_{1-x}Ge_x$ quantum wells have shown values as high as 17 000 $cm^2\,V^{-1}\,s^{-1}$ at a carrier density of $2 \times 10^{11}\,cm^{-2}$ for $Si_{0.87}Ge_{0.13}$ layers. Higher Ge contents reduce the mobility due to alloy scattering. At room temperature, only modest mobility improvements of about 20% have been observed giving 220 $cm^2\,V^{-1}\,s^{-1}$ at the high carrier

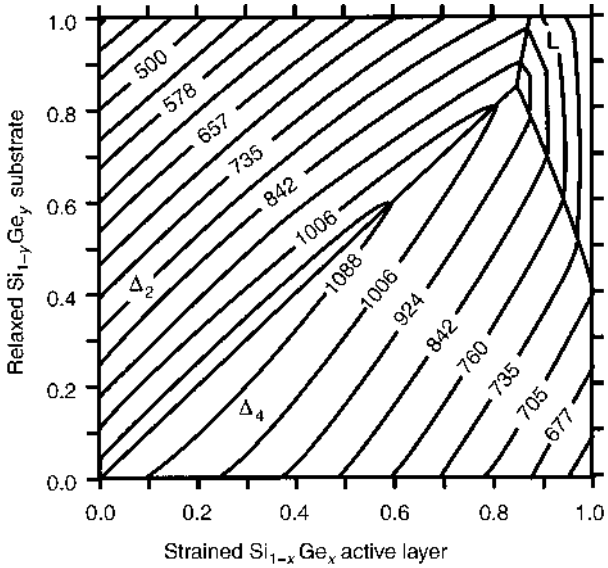

Figure 5 The bandgap energy (in meV) for strained $Si_{1-x}Ge_x$ grown on a relaxed $Si_{1-y}Ge_y$ substrate. Theoretical curves from Riger and Vogl have been calibrated to the experimental data from pseudomorphic layers.

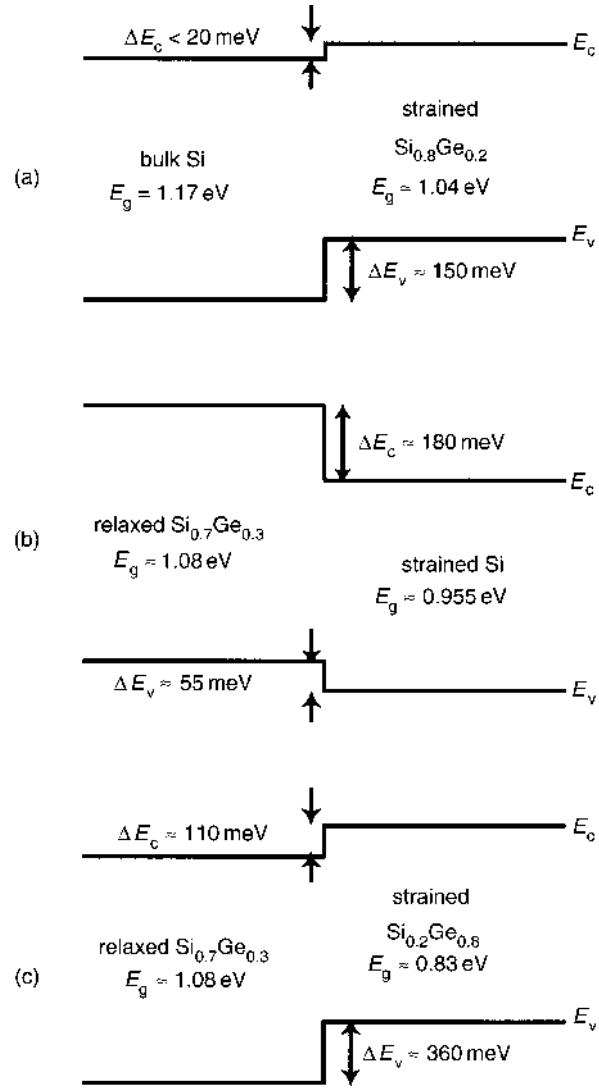

Figure 6 A schematic diagram of some typical SiGe heterolayer structures showing the bandgap, conduction band discontinuity, ΔE_c, and the valence band discontinuity, ΔE_v. (a) shows a quantum well in the valence band using a pseudomorphic strained $Si_{0.8}Ge_{0.2}$ layer; (b) shows a quantum well in the conduction band using a strained Si layer on a relaxed $Si_{0.7}Ge_{0.3}$ substrate; and (c) shows a quantum well in the valence band using a high Ge content layer on a relaxed substrate with lower Ge content.

densities required for electronic devices of about 2×10^{12} cm^{-2}. To remove alloy scattering, pure strained Ge channels grown on relaxed $Si_{1-y}Ge_y$ substrates have shown low-temperature mobilities of 55 000 cm^2 V^{-1} s^{-1} for densities of 5×10^{11} cm^{-2} with the room-temperature values reaching 1665 cm^2 V^{-1} s^{-1}. For electrons, a strained Si layer on relaxed $Si_{1-y}Ge_y$ substrates ($0.2 < y < 0.3$) also reduces the alloy scattering in the channel and produces much higher low-temperature mobilities of 390 000 cm^2 V^{-1} s^{-1} for densities of 4×10^{11} cm^{-2} while room-temperature values have peaked at 2830 cm^2 V^{-1} s^{-1}. All these values correspond to electrons or holes at heterointerfaces. Enhanced mobilities have also been observed for strained Si metal oxide semiconductor field effect transistors (MOSFET) where electron mobility enhancements of 75% compared to the bulk Si devices has been reported on relaxed $Si_{0.8}Ge_{0.2}$ although intrinsic transconductances are only increased by 60%. For holes, a strained Si layer on a $Si_{0.8}Ge_{0.2}$ relaxed substrate has demonstrated a 50% mobility enhancement over a bulk Si MOSFET.

The effective mass for electrons in strained Si layers on relaxed $Si_{1-y}Ge_y$ substrates for $0.1 < y < 0.4$ has produced a very small spread of effective masses with values of 0.19 m_0 (where m_0 is the free electron mass) for the longitudinal mass as measured by cyclotron resonance and the temperature dependence of Shubnikov–de Haas oscillations. The electron effective masses are very similar to the bulk values for strained Si quantum wells. Hole effective masses in pseudomorphically strained $Si_{1-x}Ge_x$ layers have produced a large spread of data even for similar alloy and strain values mainly due to the large nonparabolicity of the valence band (Table 3) and many of the measured samples having metastable $Si_{1-x}Ge_x$ quantum wells. The in-plane hole mass is theoretically expected to decrease as the strain in the system is increased but to date the experimental values have been much lower than theoretically calculated values. Experimental data (Table 3)

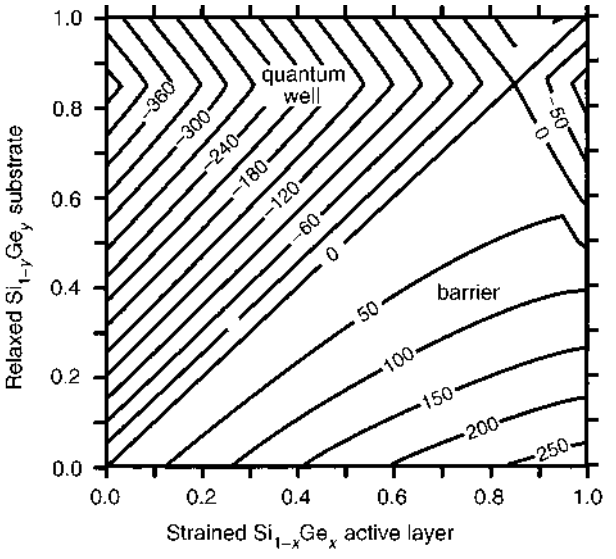

Figure 7 The conduction band discontinuity (in meV) for strained $Si_{1-x}Ge_x$ grown on a relaxed $Si_{1-y}Ge_y$ substrate. Positive numbers correspond to potential barriers to electrons while negative values represent quantum wells. The general trends and unknown experimental regimes are from the theoretical work of Rieger and Vogl while the quantum well data for $0.1 < y < 0.4$ and $x = 0$ has been recalibrated from experimental work. No experimental data are known to allow calibration of the barrier regime.

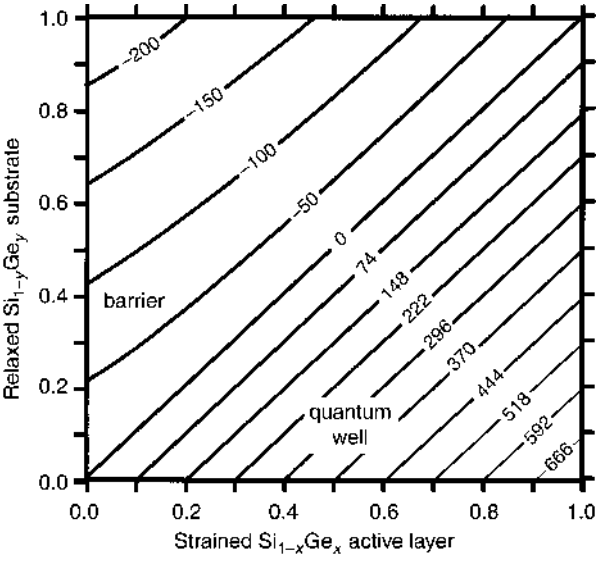

Figure 8 The valence band discontinuity (in meV) for strained $Si_{1-x}Ge_x$ grown on a relaxed $Si_{1-y}Ge_y$ substrate. Positive numbers correspond to quantum wells for holes while negative values represent potential barriers. The values from theory in the quantum well region agree well with experimental values but little experimental data exist to compare the barrier data from the theorertical values (plot from Rieger and Vogl) with experiment.

suggests that strained $Si_{0.88}Ge_{0.12}$ layers have a heavy effective mass of about $0.3\,m_0$ compared to the bulk value of $0.55\,m_0$.

Optical Properties

Due to the problems of producing high-quality single crystals of $Si_{1-x}Ge_x$, there are relatively little data on the optical properties of relaxed $Si_{1-x}Ge_x$ in the literature. The situation is even worse for strained $Si_{1-x}Ge_x$ as the layer thicknesses are limited to below 30 nm for Ge contents above 0.1, a value much smaller than the penetration depths in the system at most optical frequencies. There is, however, substantial interest in the optical properties, particularly the refractive index, as they suggest that SiGe can be used as a waveguide on Si substrates.

For the relaxed $Si_{1-x}Ge_x$ layers, most of the published work in the field has been from crystals grown by liquid phase epitaxy using melts of indium or tin. The samples had free hole concentrations of $10^{17}\,cm^{-3}$ along with indium or tin impurities and details can be found in the reviews by Humlícek. Some of the data are also from thick epitaxial layers grown well above the critical thickness where the threading dislocation density is likely to be about $10^{12}\,cm^{-2}$. The data below 2 eV were obtained by absorption measurements, those between 2 and 5.6 eV have been obtained using ellipsometry and the values between 6 and 12 eV were obtained by reflectance measurements.

Figures 9 and 10 show the real and imaginary parts of the complex dielectric function defined as $\varepsilon = \varepsilon_1 + i\varepsilon_2$ where $i = \sqrt{-1}$ for a number of different Ge contents. Figures 11 and 12 show the real and imaginary parts of the complex refractive index defined by $N = n + ik = \sqrt{\varepsilon}$. The peak starting at 3.395 eV and decreasing as the Ge content is increased corresponds to the E_1 transition across the direct gap and may be used to find the Ge content using the expression

$$x = 4.707 - \sqrt{6.538E_1 - 0.0397} \qquad [9]$$

The refractive index is high owing to the strong electronic interband absorption in the visible and ultraviolet regions of the spectrum.

The other important optical property of the material is the absorption. The absorption coefficient is defined as $K = 4\pi k/\lambda$ where k is the extinction coefficient for a wavelength λ. The intensity of a wave traversing a distance d in a homogeneous medium of extinction coefficient, k, is attenuated by a factor $\exp(-Kd)$. Therefore $1/K$ is the penetration depth. The absorption coefficient is shown in Figure 13.

Table 2 Electronic properties of Si, Ge and relaxed $Si_{1-x}Ge_x$

Element		*Si*	*Ge*	*Bulk unstrained $Si_{1-x}Ge_x$*
Direct bandgap	(eV) at 300 K	3.4	0.80	$3.395 - 1.287x + 0.153x\,(1 - x)$
Indirect bandgap	(eV) at 300 K	1.11 (D)	0.664 (L)	$1.155 - 0.43x + 0.0206x^2\ (x < 0.85)$ $2.010 - 1.27x\ (x > 0.85)$
Spin-orbit splitting	Δ_{so} (meV)	44	290	Discontinuity at $x \sim 0.85$
Electron mobility	μ_e ($cm^2\,V^{-1}\,s^{-1}$) at 300 K	1450	3900	$\mu_e = 1466.2 - 6425.9x + 6601.8x^2 + 7716.5x^3 - 10944x^4 (x < 0.6)$ $\mu_e = 2.9235 \times 10^{-28} e^{71.105x} + 4790.1 - 14246x + 11117x^2 (x > 0.6)$
Hole mobility	μ_h ($cm^2\,V^{-1}\,s^1$) at 300 K	505	1800	$\mu_h = 188.42 - 935.7x + 5455.5x^2 - 8553.7x^3 + 5751.7x^4$(all x) ($N_A = 1 - 50 \times 10^{16}\ cm^{-3}$)
Longitudinal electron mass	m_l^*	0.1905	0.082	Discontinuity at $x \sim 0.85$
Transverse electron mass	m_t^*	0.9163	1.58	Discontinuity at $x \sim 0.85$
Heavy-hole mass	m_{hh}^*	0.537	0.284	Discontinuity at $x \sim 0.85$
Light-hole mass	m_{lh}^*	0.153	0.044	Discontinuity at $x \sim 0.85$
Spin–orbit mass	m_{so}^*	0.234	0.095	Discontinuity at $x \sim 0.85$

Table 3 A comparison of effective masses in modulation-doped samples as measured by the temperature dependence of the Shubnikov–de Haas oscillations (SdH) or by cyclotron resonance (CR) at liquid helium temperatures. All samples have strained $Si_{0.88}Ge_{0.12}$ or $Si_{0.87}Ge_{0.13}$ quantum wells grown on bulk Si substrates. Therefore only the heavy-hole band should be populated at these low temperatures and strain values. ℓ is the mean free path which is the mean scattering distance of a hole in the system and is a quantitative measure of disorder

	Mobility ($cm^2\,V^{-1}\,s^{-1}$)	*p ($\times 10^{11}\,cm^{-2}$)*	*ℓ (nm)*	*m^* (m_0) SdH*	*m^* (m_0) CR*
DERA	6560	4.4	71	0.304	0.26
NRC	6800	5.9	86	0.31	–
IBM	6000	3.7	60	0.34	0.18
Warwick	9600	2.5	78	0.27	0.23, 0.29
Warwick	10500	2.1	78	0.27	0.24, 0.30

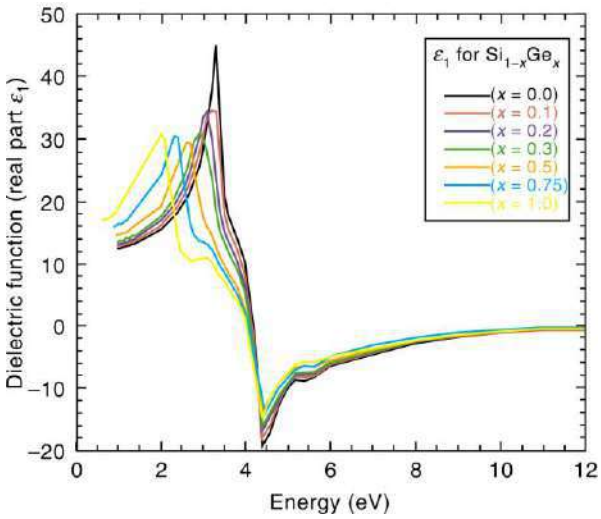

Figure 9 The real part (ε_1) of the complex dielectric function $\varepsilon = \varepsilon_1 + i\varepsilon_2$.

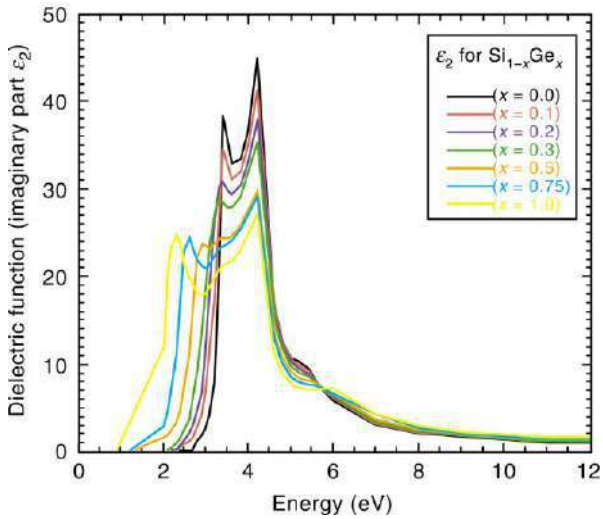

Figure 10 The imaginary part (ε_2) of the complex dielectric function $\varepsilon = \varepsilon_1 + i\varepsilon_2$.

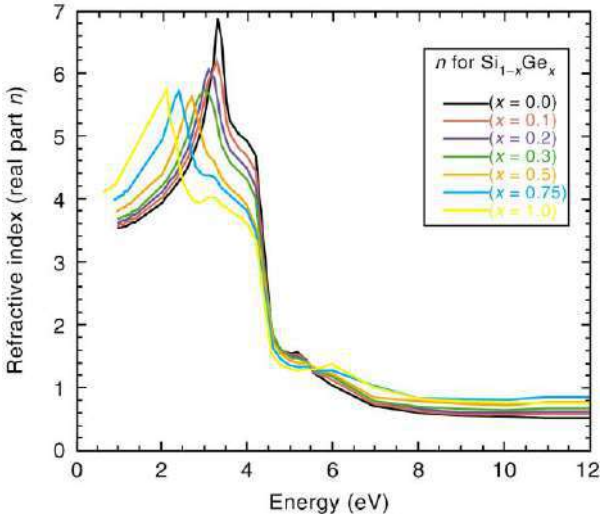

Figure 11 The real part (n) of the complex refractive index $N = n + ik$.

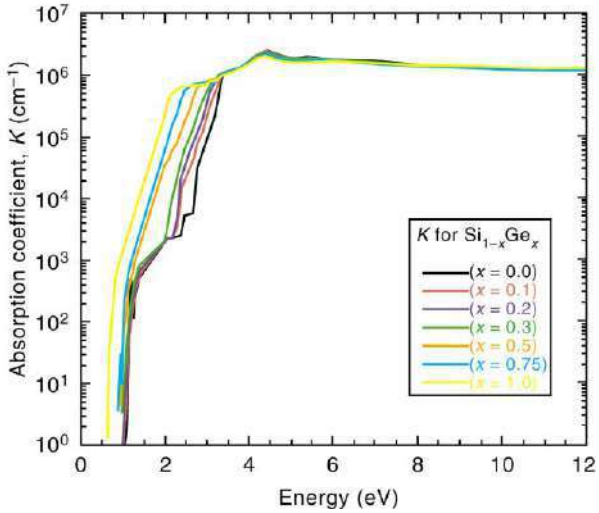

Figure 13 The absorption coefficient K.

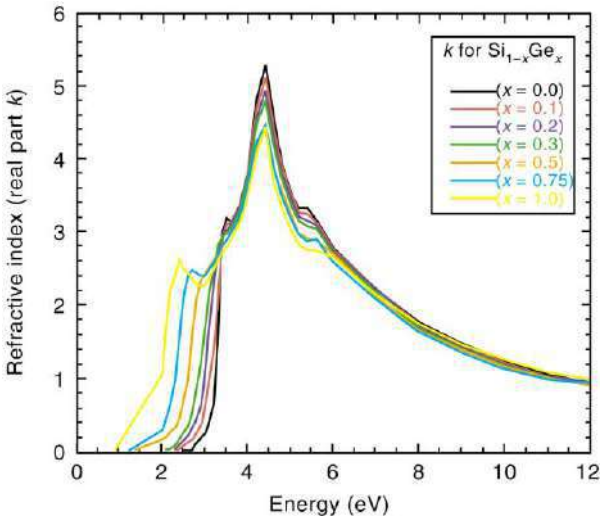

Figure 12 The imaginary part (k) of the complex refractive index $N = n + ik$.

Determining the optical properties of strained $Si_{1-x}Ge_x$ layers has been substantially more difficult. The major problem is that the absorption coefficient is about 10–20 cm^{-1} around the fundamental bandgap (Figure 6) and therefore any shift due to strain effects becomes very difficult to observe using absorption techniques in thin SiGe layers. Additional complexity results as the substrate material has a fundamental bandgap close to the strained layer. Therefore photoluminescence is one of the few methods for extracting experimentally the optical changes due to strain.

The splitting of the conduction (Figure 8) and valence bands (Figure 9) has already been discussed. The photoluminescence signals will also shift with strain for the lowest lying bandgap energy. Almost all the experimental data for strained layers have been obtained for pseudomorphically grown strained $Si_{1-x}Ge_x$ alloys on bulk Si substrates and the data agree well with the bandgaps shown in Figures 3 and 5. The quantum well photoluminescence signal is found to shift by −1.4 meV per kbar of pressure. After removal of nonhydrostatic components, a value of −1.52 meV per kbar for the pure hydrostatic shift has been reported.

Conclusions

A review has been presented of the fundamental properties of relaxed and strained $Si_{1-x}Ge_x$ alloys. The materials properties were first reviewed and demonstrate the limitations imposed by the critical thickness. The electronic properties including the bandstructure were then discussed and demonstrate the lowered bandgap for $Si_{1-x}Ge_x$ alloy compositions compared to bulk Si. This property has been used in $Si_{1-x}Ge_x$ p-i-n photodetectors to move the absorption to longer wavelengths in an attempt to compete with III–V photodetectors for the important 1.3 μm and 1.55 μm fiber optic frequencies. In the final section the optical properties of relaxed and strained $Si_{1-x}Ge_x$ material were reviewed. The difference in refractive indices has opened research into SiGe on Si waveguides.

List of Units and Nomenclature

Absorption coefficient (K)	[cm^{-1}]
Band gap energy (E_g)	[eV]
Conduction band discontinuity (ΔE_c)	[meV]
Critical thickness	[nm]
Density	[$kg\ m^{-3}$]
Dielectric constant (ε)	
Effective mass [m_0 where $m_0 = 9.11 \times 10^{-31}\ kg$]	
Elastic moduli (c_{ij})	[GPa]
Extinction coefficient (k)	[cm^{-1}]
Lattice constant (a_0)	[nm]
Mean free path (l)	[nm]
Melting point	[K]
Mobility (μ)	[$cm^2\ V^{-1}\ s^{-1}$]
Pressure	[kbar]
Refractive index (N)	
Sheet carrier densities	[cm^{-2}]
Sound velocity	[$m\ s^{-1}$]
Temperature	[K]
Thermal conductivity (κ)	[$W\ m^{-1}\ K^{-1}$]
Valence band discontinuity (ΔE_v)	[meV]
Wavelength (λ)	[nm]

See also

Semiconductor Physics: Band Structure and Optical Properties; Outline of Basic Electronic Properties; Recombination Processes.

Further Reading

Humlícek J, Garriga M, Alonso MI and Cardona M (1989) Optical spectra of $Si_{1-x}Ge_x$ alloys. *Journal of Applied Physics* 65(7): 2827–2832.

Humlícek J, Lukes F and Schmidt E (1991) Silicon-germanium alloys (Si_xGe_{1-x}). In: Palik ED (ed.) *Handbook of Optical Constants of Solids*, vol. II, pp. 607–636. San Diego, CA: Academic Press.

Kasper E and Lyutovich K (eds) *Properties of Silicon Germanium and SiGe:Carbon*. London: INSPEC IEE.

Paul DJ (1999) Silicon-germanium strained layer materials in microelectronics. *Advanced Materials* 11(3): 191–204.

Rieger MM and Vogl P (1993) Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates. *Physical Review B* 48(19): 14276–14287.

Robbins DJ, Canham LT, Barnett SJ, Pitt AD and Calcott P (1992) Near band gap photoluminescence from pseudomorphic $Si_{1-x}Ge_x$ single layers on silicon. *Journal of Applied Physics* 71(3): 1407–1414.

III-Nitrides

K P O'Donnell, University of Strathclyde, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction to InGaN

The optical properties of wurtzite GaN are of interest mainly to band-structure theorists. Luminescence from pure InN has probably never been detected. It is therefore surprising that the combination of these binary III–N compounds produces a solid solution, InGaN, which has revolutionized blue–green optoelectronics in the last 10 years, and may form the basis of a new lighting technology to replace the tungsten lamp in the first half of the twenty-first century.

InGaN is also the material basis of the blue laser diode, which will be used for future generations of optical data storage and for many other applications that require a compact source of short-wavelength coherent light. The market for nitride light emitters has been growing at a rate of ~20% a year since the introduction of the first blue nitride light-emitting diode (LED) by Nichia Chemical in 1993. The first commercial lasers, with lifetimes in excess of 10 000 hours, were released in 1998.

So what is InGaN? The naïve response–that it is a semiconductor alloy–ignores the theoretical demonstration, based on simple equilibrium thermodynamics, that InN is insoluble in GaN at the temperatures usual for growth (using metalorganic vapor phase epitaxy, MOVPE). If we imagine a cation lattice, on which we distribute Ga and In atoms, in a fixed numerical proportion, in order to form a solid solution of given composition, it is thermodynamically probable that the In atoms will associate to form agglomerations in certain regions of the lattice. The size, shape, and properties of the resulting concentration fluctuations have been a topic of intense speculation.

Practically speaking, InGaN layers are grown either as rather thin (<3 nm) quantum wells (QWs) immersed in GaN or AlGaN or as thicker epilayers (10–500 nm) on top of thick GaN templates. Only the QWs are of use in devices, whereas the composition and structure of the thick epilayers are easier to measure in research applications. It is dangerous to suppose that the material retains

the same properties when grown in thicker layers. Recently, a third class of sample has emerged, in the form of self-organized 'quantum boxes' (originally French *bôites*), which we shall call QBs. These samples, grown at low temperatures by molecular beam epitaxy (MBE) comprise pancake-shaped disks of uncertain composition, about 50 nm across and 3–5 nm high.

Optical Spectra of Devices

There have been several 'generations' of nitride LED. We refer here to the Nichia Blue, Extrablue (actually turquoise) and Green diodes which set the scene in 1996, and the Amber diode which first appeared in 1998, although not yet available commercially. The spectral properties of these devices have been stable while their efficiencies have improved in the intervening years. The basic design appears to incorporate a single quantum well. Up to a point, the wavelength tuning depends upon the incorporation of different amounts of indium in the well. (Green and amber devices may have the same amount.)

In what follows, we prefer to specify composition by quoting the indium nitride content, x, according to the formula unit

$$\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N} = x\mathrm{InN} + (1-x)\mathrm{GaN} \qquad [1]$$

Electroluminescence (EL) spectra of blue and green diodes are shown in Figure 1. Spectra are broad (compared to $k_\mathrm{B}T$ at room temperature) with exponential tails to high and low photon energies. The linewidth tends to increase as the device peak wavelength shifts further to the red.

Whereas it is difficult to measure directly the optical absorption spectrum of an ultrathin layer of material, an equivalent measurement can be made on an LED by monitoring the photocurrent (PC) generated by incident monochromatic light. PC spectra of the blue and green diode are shown in Figure 1, together with the absorption and photoluminescence (PL) spectra of a thick violet-light-emitting epilayer.

The spectra of Figure 1 transform into each other by 'stretching' the x-axis, using the bandgap energy of GaN as an anchor point. Comparison of many pairs of spectra produces evidence of a *universal scaling behavior*, which relates the details of each spectrum to a single controlling parameter. In the case of epilayers, this is undoubtedly the InN fraction, as we shall see later.

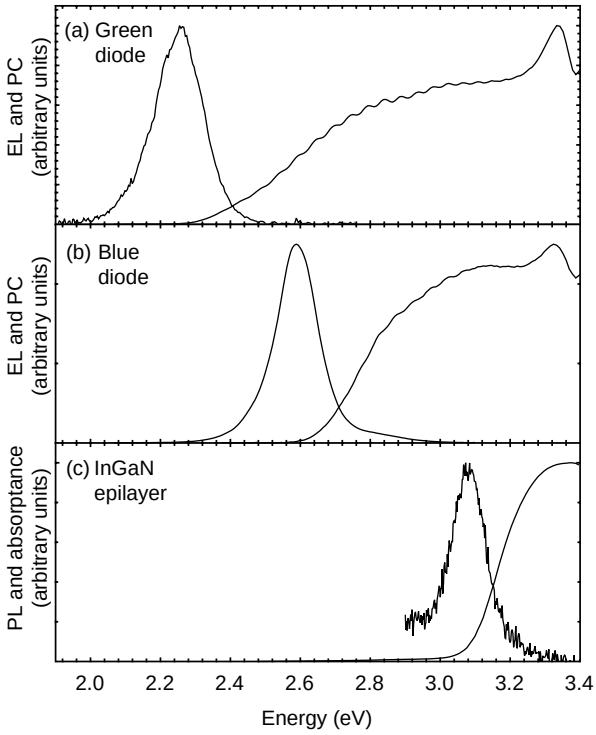

Figure 1 EL and PC (quasi-absorption) spectra of nitride LEDs, compared to PL and absorption spectra of a thick InGaN epilayer.

In order to describe the form of the absorption/excitation curve, we introduce the formal expression

$$\alpha = \frac{\alpha_0}{1 + \exp\left(\dfrac{E_\mathrm{g} - E}{\Delta E}\right)} \qquad [2]$$

where α_0 is the plateau absorption, observed experimentally, but not yet theoretically explained, E_g is the (effective) bandgap energy of the sample, E is the photon energy, and ΔE is an energy broadening term, similar to the well-known Urbach parameter.

Optical Energy Relationships

The spectral universality pointed out in the previous section requires the existence of certain relationships between the optical energies. We list the defining energies as the bandgap, E_g, the peak energy of EL or PL, E_p, and the difference between these energies, measured for a particular sample, which we call the Stokes shift:

$$\mathrm{SS} = E_\mathrm{g} - E_\mathrm{p} \qquad [3]$$

(While the energy broadening tends to increase in a regular fashion with the Stokes shift, there are irregularities in both ΔE and the PL linewidths of samples from different sources. Hence ΔE and PL

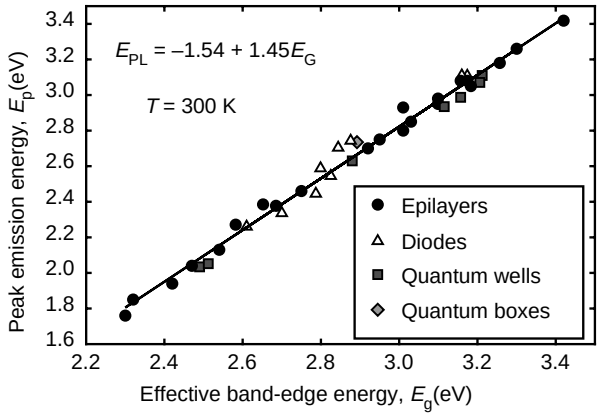

Figure 2 Linear dependence of luminescence peak energy on effective bandgap.

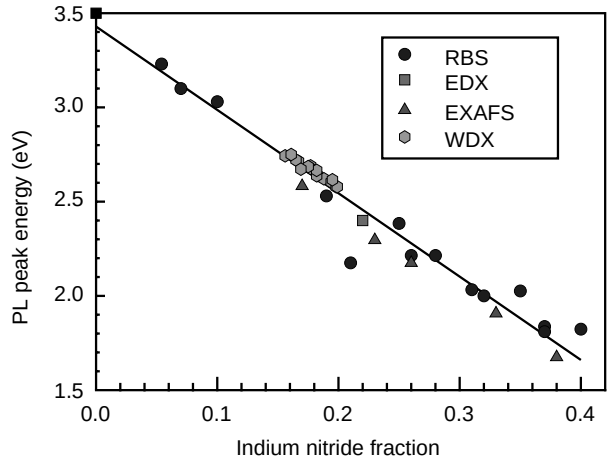

Figure 3 Linear dependence of the PL peak energy on the InN content of InGaN epilayers.

linewidth appear to be local rather than universal parameters.)

The relationship between E_g and E_p for a wide range of emitters is found to be linear (Figure 2):

$$E_p = 1.45E_g - 1.54 \qquad [4]$$

The bandgap energy and luminescence peak energy coincide at $1.54/(1.45 - 1) = 3.42$ eV, which corresponds to the energy bandgap of pure GaN. The simple linear dependence of E_p upon E_g implies a linear increase of the Stokes shift energy with increasing detuning from the GaN band edge.

Dependence of Optical Energies on *x*, the InN Fraction, for Thick Epilayers

As stated above, it is usual to measure the composition of thick layers, rather than quantum wells, for reasons of experimental convenience. Several techniques can be used to this end, including measurement of the lattice constants by X-ray diffraction (XRD) (one asymmetric and one symmetric reflections are required to take into account the strain state of the layer), Rutherford backscattering spectrometry (RBS), wavelength-dispersive (preferable) or energy-dispersive X-ray fluorescence (XRF), sputtered ion mass spectroscopy (SIMS), and extended X-ray absorption fine structure (EXAFS).

All usable techniques provide comparable results if carried out with sufficient care. We plot the peak energy of PL emission against the measured indium nitride fraction for a range of epilayer samples in Figure 3. Once more, the dependence appears to be linear in a restricted composition range. The best-fit line to a linear function $E_p(x)$ has an x-intercept which is close to the bandgap of GaN. The dependence of the optical energies on the indium nitride fraction can now be summarized by three interdependent linear relations, which are valid in the composition range $0 < x < 0.4$:

$$E_p = 3.4 - 4.4x$$
$$E_g = 3.4 - 3.0x \qquad [5]$$
$$SS = 1.4x$$

The measurements described above rely on separate determinations of composition and an optical energy for each sample. Since all samples are to some extent inhomogeneous, it would be preferable to make such measurements at one and the same time. This can be done by using an electron probe microanalyzer, modified to allow simultaneous cathodoluminescence (CL) spectroscopy. CL is analogous to PL. A recent study of samples with $0.05 < x < 0.25$ has yielded the result:

$$E_c = 3.40(1) - 3.91(5)x \qquad [6]$$

with error bars (shown) that are about ten times smaller than those of the previous work.

Bandgap of InN

It is dangerous to extrapolate experimentally determined relationships beyond the range of measurements used to derive them.

The bandgap of an alloy is best represented by a quadratic function of x, with the deviation from linearity described by a bowing parameter. The above measurements suggest a small bowing parameter for the InGaN alloy system. If we ignore bowing and equate E_p with E_c, the InN bandgap obtained by setting $x = 1$ is calculated to be 0.71 eV. This is certainly an underestimate, but is similar to recent values obtained experimentally for layers of

pure InN grown by Molecular Beam Epitaxy. Before the year 2000, the generally accepted value of the InN bandgap was close to 2 eV.

Origin of the Stokes Shift

Stokes shifts originate in systematic energy loss during the excitation–luminescence cycle. While the bandgap, as defined above, is the mean energy of electron–hole pairs across the bandgap, the PL peak is the modal energy of electron–hole recombination. The large Stokes shift in the InGaN system is probably due to the influence of giant internal electric fields that separate the carriers prior to recombination.

In the absence of a microscopic model, it is difficult to provide a quantitative description of energy loss in the excitation–emission cycle. Experiment shows that the shift can be as large as ~ 0.6 eV in an epilayer with $x = 0.4$. If such a layer were 100 nm thick and uniform in composition, it would be strained by about 4% when pseudomorphic to a GaN template, with an estimated vertical electric field of 6×10^6 V m^{-1}. Unit strain would correspond to a 15-eV shift (equivalent to a piezoelectric deformation potential). While these numbers are large, they seem reasonable in the light of sophisticated theoretical calculations of the magnitude of the piezoelectric effect in InGaN. But it is an unfortunate experimental fact that layers grown with different thicknesses, and presumably under different states of strain, may show the same value of the Stokes shift.

If the material has composition fluctuations, accompanied by strain variations, the recombination energy will be much more sensitive to the local environment. One can imagine a situation in which recombination always takes place at the lowest available energy, as carriers diffuse under the action of local strain/electric fields. This situation would be rather similar to that which pertains in quantum wells with spatial energy disorder, caused by well width fluctuations. A theoretical treatment of localization in three dimensions would be very difficult.

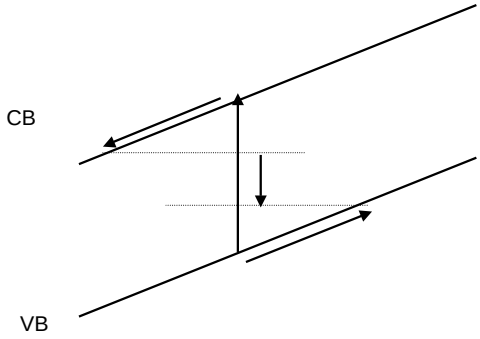

Figure 4 Excitation–emission cycle in the presence of an electric field. CB and VB are conduction band and valence band, respectively.

In any case, the physical situation of excitation and recombination in the presence of an electric field is that sketched in Figure 4.

Quantum Wells and Quantum Boxes

The incorporation of quantum wells has several advantages for light-emitting devices. In nitrides, the expected increase of oscillator strength for radiative transitions, due to the reduced dimensionality, is mitigated to some extent by the opposing effect of the internal electric fields, but the quantum well still acts as an energy sink for mobile excitations and as a site for carrier localization.

Quantum confinement can also act as a wavelength tuning mechanism in QW devices. The effect in blue-emitting nitride devices is rather small until the QW width falls below 3 nm. It should be clear that a reduction in the photon energy of recombination with increasing well width will accompany vertical separation of carriers in a piezoelectric model. Usually, as mentioned above, wavelength tuning in nitride devices is accomplished by changing the composition. What should we make of the fact that amber and green diodes apparently have the same composition? It would appear that the well widths are different. On the other hand, the amber emitters have anomalously narrow linewidths, which may imply that some other mechanism is in force. Here, we must let the matter rest, in the absence of detailed structural information.

Quantum boxes should be distinguished from quantum dots (QDs). The former are mesoscopic objects that form spontaneously under certain growth conditions, which usually involve the influence of strain. The best-characterized example in semiconductor physics is the Stranski–Krastanow growth of InGaAs structures on GaAs. While GaN in AlGaN forms recognizable nanocrystals, InGaN forms less well-defined structures of indeterminate shape, some 30–50 nm in lateral extent and 3–10 nm high. QDs, on the other hand, are truly zero-dimensional objects, in the sense that carrier motion is quantized simultaneously in three spatial dimensions. The utilization of reduced dimensionality to enhance recombination probabilities in active devices is only valid when QB structures meet the criteria of low dimensionality: the carrier motion must be restricted to a distance smaller than the appropriate Bohr radius.

Structure of InGaN Epilayers

The state of strain of InGaN epilayers is usually determined by XRD. Both symmetric and asymmetric

reflections are required to calculate the actual values of the out-of-plane and in-plane lattice constants, c and a, respectively. These can then be compared with those predicted for a relaxed alloy by Vegard's law, for example:

$$c(\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}) = xc_{\mathrm{InGaN}} + (1-x)c_{\mathrm{GaN}} \qquad [7]$$

The composition itself, as well as the state of strain, can be determined from measurements of c and a with one additional assumption: that the Poisson ratio of the alloy also obeys Vegard's law. If the composition is known from an independent chemical measurement, this assumption can be tested.

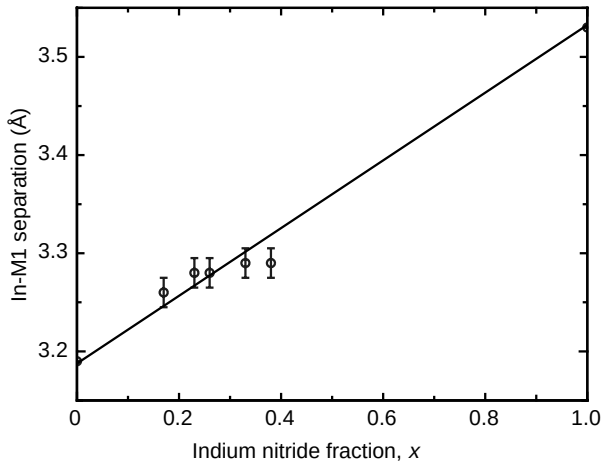

Figure 5 EXAFS-determined local structure parameters of thick InGan films.

InGaN epilayers are usually grown on thick GaN formers. In pseudomorphic growth, the constraint, $a(\text{layer}) = a_{\mathrm{GaN}}$ applies. This is sometimes assumed to be the case when layers are thinner than about 50 nm, but nothing should be taken for granted when dealing with this material. Each layer is different. A similar warning applies to the assumption that thick InGaN layers will be relaxed. It has recently been shown that partially relaxed and pseudomorphic regions may actually coexist within the same InGaN film, which nevertheless has a composition that is uniform throughout. Other films may show a concentration gradient due to compositional pulling during growth.

The local structure of InGaN films has been measured using In-edge EXAFS. Using a tetrahedral two-shell model that does not distinguish the wurtzite and cubic crystal structures, we find that the In–N1 bond length is more or less independent of the alloy composition in the range $0.17 < x < 0.38$ of samples measured, having a value of 2.11(1) Å (compared to 2.16 Å in pure InN). The M1 shell of mixed In and Ga atoms is located on average at about 3.28 Å, as shown in Figure 5.

The straight line on the figure connects the values of the a lattice parameters (of GaN at $x = 0$ and InN at $x = 1$) and hence represents the predictions of Vegard's law for a relaxed or bulk alloy. The measured In–M1 separations are less dependent on x than Vegard predicts.

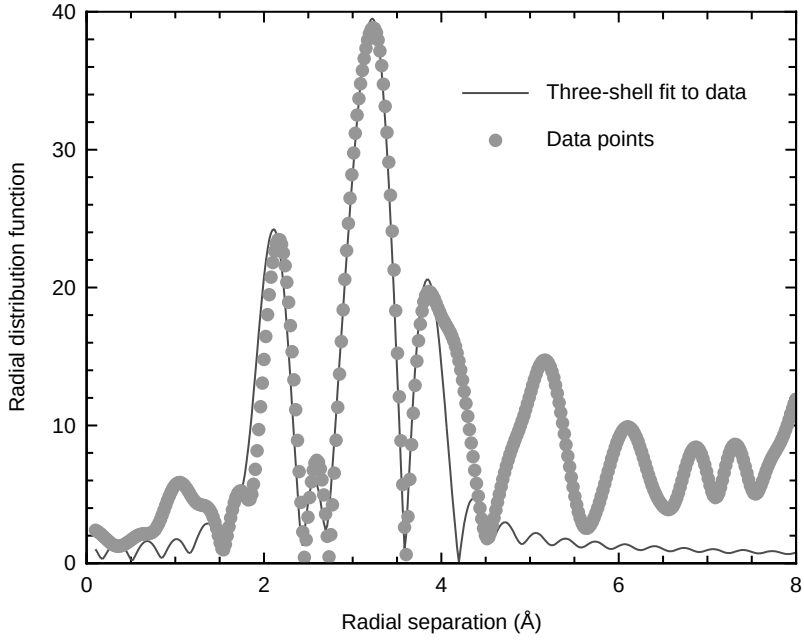

Figure 6 Fourier transform of the EXAFS spectrum yields the radial distribution function of an InGaN QB sample.

Local Structure of InGaN QW and QB

There has been a great deal of speculation in the literature concerning the local structure of InGaN alloys in the active layers of light-emitting devices. Much of this discussion paddles in the shallows of invalid comparison, as researchers extrapolate their findings from the results of measurements and calculations that can be made to those that cannot. One example is the large theoretical effort devoted to the defect physics of GaN, which is supposed somehow to 'explain' the optical properties of InGaN. ('Let us first suppose that all the horses in the race are spherical.')

We will try to avoid invalid comparisons in the brief discussion that follows. Recently, a Strathclyde University/Daresbury Laboratories/CHEA-CNRS Valbonne collaboration has succeeded in measuring the local structure of an InGaN QB sample, grown on a thick GaN template by MBE. The radial distribution function (RDF) around the target In probe atom is shown in Figure 6, together with a three-shell fit to the data.

The first shell of neighbors comprises four N atoms at 2.10(2) Å, the second shell is pure Ga at 3.22(3) Å, and the third shell pure In at 3.56(3) Å. These results suggest that there are two distinct In environments in such samples. One resembles the InGaN alloy, encountered in the epilayer samples, while the other is an environment of pure InN.

On the basis of these results, it is extremely likely that InGaN phase-separates in QB growth to form a two-component mixture comprising an In-poor and an In-rich phase. The manner in which these phases interpenetrate is a matter for speculation.

See also

Semiconductor Materials: Quantum Dots. **Semiconductor Physics:** Infrared Lattice Properties; Light Scattering; Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Gil B (ed.) (1998) *Group III Nitride Semiconductor Compounds, Physics and Applications*. Series on Semiconductor Science and Technology 6. Oxford, UK: Oxford Science Publications.

Gil B (ed.) (2002) *Low Dimensional Nitride Semiconductors*. Oxford, UK: Oxford Science Publications.

Nakamura S and Fasol G (2000) *The Blue Laser Diode*, 2nd edn. Berlin: Springer-Verlag.

Large Gap II–VI Semiconductors

J F Donegan, Trinity College Dublin, Dublin, Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

Background

The chemical reaction of group IIA (Be, Mg), IIB(Zn, Cd) and group VI (O, S, Se, Te) atoms produces solid crystalline materials which are classified as semiconductors. The archetypal compound is ZnSe and this material will be used as the main example in this article. These are termed II–VI binary (involving two elements) compounds. It is also possible to grow alloy materials where more than one atom from each group is used; an example is $Zn_xCd_{1-x}Se$ where x is the proportion of Zn in the material. In general, there is a random arrangement of the Zn and Cd atoms within the crystal structure. Both binary and alloy semiconductor materials are single crystals so that the atoms are arranged in a periodic three-dimensional system called a lattice. The lattice may be represented by a unit cell; simply by repeating the pattern of the unit cell, we can generate the entire crystal lattice.

At the atomic level, the zinc (selenium) atom is bonded to 4 selenium (zinc) atoms which are arranged at the corners of a tetrahedron. The unit cell formed by this tetrahedral arrangement is termed the zincblende structure and is illustrated in Figure 1a. The dimension shown is called the lattice constant a and is characteristic of a particular II–VI semiconductor material. Some II–VI materials crystallize with a different unit cell which is termed the wurtzite structure, though again the atoms are bonded to four others in a tetrahedral arrangement as shown in Figure 1b. In this case, two lattice constants denoted a and c are required to characterize the unit cell. Some of the II–VI materials can occur in either of these two structures depending on how they were grown. The lattice constants of several II–VI compounds are given in Table 1.

II–VI Materials Growth

All of the II–VI family of semiconductors have high melting points due to the strong bonding that exists between the atomic species; values are given in

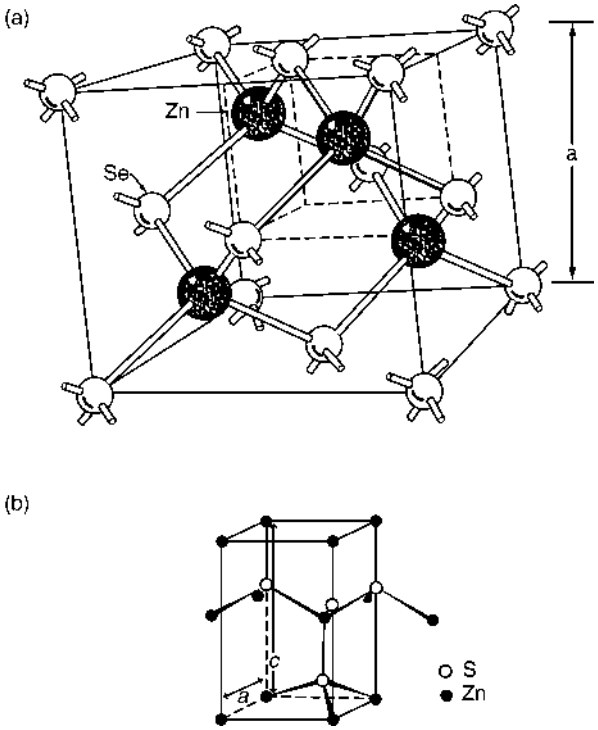

Figure 1 (a) Zincblende lattice; (b) wurtzite lattice. *a* and *c* are the lattice constants as given in Table 1.

Table 1. The crystals can be grown either by cooling a liquid (often called a melt) or by condensation from a gas (vapor). Crystal growth is a very precise and difficult process requiring the control of a large number of variables most notably temperature and pressure. As many of the II–VI materials can exist in either the zincblende or wurtzite structures, producing single crystals with only one structure is often quite difficult. Many of the crystals are 'twinned'; in a zincblende structure, a twin produces a wurtzite structure in one plane. The photograph in Figure 2 shows a ZnSe single crystal. This crystal took several days to grow indicating that very stable growth conditions must be maintained over long periods of time. The vapor phase technique takes advantage of the high vapor pressure of the II–VI compounds that allows for the growth of large single crystals at temperatures much below the melting point of these materials. Vapor growth is helped also by the congruent evaporation of the two constituents into the vapor state in a dissociated form. These two vapor constituents also recombine stoichiometrically into the solid crystal. The melting points and vapor pressures at the melting points of some II–VI semiconductors are given in Table 1.

It is also possible to grow materials in a layer-by-layer fashion onto an existing substrate; this is termed epitaxial growth. Recent developments have allowed the growth to proceed one atomic layer after the other. As a result, materials can be grown that are less than 100 Å in size. At these small sizes, the fundamental electronic and optical behavior of these materials is altered due to quantum mechanical effects. These materials are termed quantum wells. Growth temperatures in these cases are much below the melting point of these materials and allow for uniform single crystals without any 'twin' formation. In molecular beam epitaxy (MBE), atomic beams of the constituents impinge onto a heated substrate where they combine to form the crystalline material. Growth of ZnSe onto III–V substrates such as GaAs is well developed. The quality or purity of the single crystals can often be deduced from their optical properties as we will see later.

The addition of small amounts of other atoms into the crystal during the growth procedure is called doping. Semiconductors are materials whose electrical conductivity can be drastically altered by the addition of small amounts of dopants. This ability to control the electrical conductivity is the key to their use in the high-speed devices used in modern optical communications systems (*see* **Semiconductor Physics:** Impurities and Defects). In ZnSe, doping with Cl produces n-type material while doping with N produces p-type material. The ability to form a pn junction (a thin region between n and p type material) is the precursor for the production of optical devices such as light-emitting diodes (LEDs) or laser diodes (LDs) based on II–VI semiconductors. Doping also changes the optical properties of II–VI materials.

All materials are defective to some degree, and these II–VI materials are no exception. The number and type of defects are determined by thermodynamic and purity considerations during the growth process. There are native defects such as vacancies (missing atoms) and dislocations (extra or missing layers of atoms). There are also impurities present; foreign atoms due to the impurities in the starting growth materials. These impurities have an important effect on the ability of the material to emit light. Dopants and impurities may help to increase the amount of light emitted while dislocations usually reduce the emission efficiency in II–VI materials.

Optical Properties of II–VI Semiconductors: Basic Concepts

Introduction

When light is incident on a solid material, its electric field interacts with the electric field within the atoms comprising the solid. This interaction leads to a

Table 1 Structural, thermodynamic and optical properties of II–VI compound semiconductors

Material	*Lattice constant (Å)*	*Crystal structure*	*Melting point (mp) (°C)*	*Vapor pressure at mp (atm)*	*Refractive index*	*Bandgap (0 K)*	*Bandgap (300 K)*
ZnO	*a* 3.250	W	1975	22	2.0	3.44	3.37
	c 5.207						
ZnS	*a* 3.820	W	1718–1850	3.7	2.8	3.91	3.80
	c 6.260						
	a 5.410	Z			2.4	3.83	3.66
ZnSe	*a* 5.668	Z	1520	1.0	2.8	2.82	2.67
ZnTe	*a* 6.104	Z	1300	0.6	3.1	2.39	2.25
CdS	*a* 4.133	W	1405–1475	3.8	2.5	2.58	2.42
	c 6.710						
	a 5.832	Z			–	–	–
CdSe	*a* 4.299	W	1239–1264	0.4	2.6	1.84	1.74
	c 7.015						
	a 6.057	Z			–	1.77	–
CdTe	*a* 6.477	Z	–	–	2.75	1.60	1.49
MgS	*a* 5.89	Z	–	–	2.6	–	4.5
MgSe	*a* 5.62	Z	–	–	2.6	–	3.6
BeTe	*a* 5.627	Z	–	–	–	–	2.8

Data taken from many sources mentioned in the bibliography but primarily from Bhargava, R. (1997) (ed.) *Wide Bandgap II–VI Semiconductors*, The Institute of Electrical Engineers, London, Table 1 of Chapter 5.4 with permission.

Figure 2 ZnSe single crystal grown by vapor transport. Photograph courtesy of NASA.

number of effects that we can observe. The light will be reflected or scattered from the surface of the solid, depending on whether the surface is smooth or rough on a length-scale similar to the wavelength of light. The materials of interest here are high-quality crystals with a smooth reflective surface. When light enters a material, it is refracted, i.e., it changes direction within the material. The simplest example of this is when a stick is placed in a jug of water, it appears to be bent. It is of course the light that is bent due to its different interaction with the air and with water. Snell's law states that:

$$\frac{\sin \alpha_i}{\sin \alpha_{tr}} = n \qquad [1]$$

where α_i is the angle of incidence of the light and α_{tr} is the angle of refraction. n is known as the refractive index of the material and is characteristic of it. As well as being refracted, any light that enters the crystal is either absorbed within the material or transmitted through the material to emerge on the other side. If the light is absorbed the material will look colored if the absorption takes place in the visible region where our eyes are sensitive. The ZnSe crystal in Figure 2 is such a material; the orange color is due to the absorption of blue light from the white light shining on the material. Other materials look transparent; ZnO is such an example. Following absorption, some materials produce light of their own; the color of the light is characteristic of the material and this process is termed luminescence. Light may also be scattered within a crystal by imperfections and by vibrations of the atoms in the crystal.

When light enters a solid material, the electric field polarizes the atoms within the structure. The positive and negative charges within the atoms move slightly from the equilibrium positions. All of the optical effects mentioned above are related to the induced polarization. As the constituent atoms in the structure are unique, the induced polarization and the resulting optical properties are specific to particular materials, and this makes optical

characterization a very important tool for the investigation of semiconductor materials.

Reflection at the Air-to-Material Interface

Our experience of looking at reflection from glass tells us that these are weakly reflecting materials; the same is true of single crystal II–VI materials such as that in Figure 2. The interface between air and an optical material is a fundamental aspect of optical physics. The amount of light reflected can be determined from Maxwell's equations. The simplest equation is found where light is incident perpendicular to the surface of the material:

$$R = \frac{(n-1)^2}{(n+1)^2} \qquad [2]$$

where R is the reflection coefficient and n is the refractive index of the material. The value of n is characteristic of the particular material. Values of n for several II–VI materials are given in Table 1. The refractive index n for all these II–VI materials is not constant but varies slowly with the wavelength of the light impinging on the material. In Figure 3, we show the wavelength dependence of the refractive index of ZnSe.

Transmission and Absorption of Light within II–VI Materials

Once the light has passed through the interface with the air, it can either be transmitted through the material to emerge on the other side at a second material–air interface or be gradually absorbed within the body of the material. The ZnSe sample in Figure 2 is illuminated with white light, but it appears orange in color to our eyes, i.e., the orange component within the white light is transmitted through the material to our eye. The other components in the light, mostly green and blue, are absorbed within the material. To obtain more quantitative information, we take a spectrum of the absorption within the material. We illuminate the sample with light of different wavelengths and determine how much is transmitted or absorbed at each wavelength. The transmission spectrum of ZnSe is shown in Figure 4.

The visible spectral region extends from 0.4 to 0.7 microns and we see that there is a large change in transmission around 0.5 microns that is responsible for the color of the ZnSe crystal as seen in Figure 2. This spectral region around the onset of a large reduction in transmission at 0.5 microns is termed the bandgap of the material. As in the case of the refractive index, the value is unique to the material. II–VI materials are termed wide bandgap materials as they possess larger values of the bandgap energy than for Si, Ge and the III–V semiconductor materials. If we shine light of wavelength 0.5 microns onto ZnSe material, the transmission is very low. Under these conditions, the material would look black to our eyes. Once the light has been

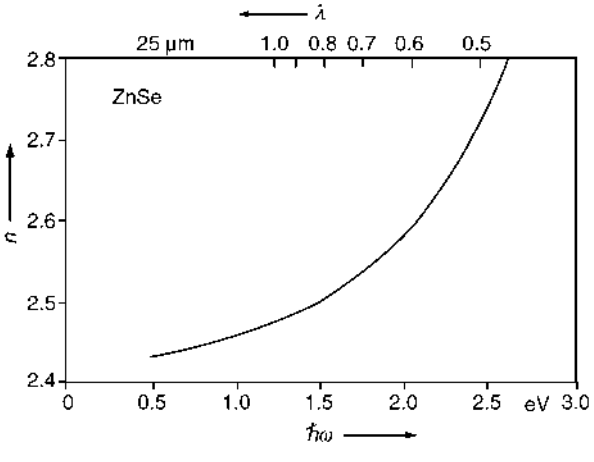

Figure 3 Refractive index n of ZnSe versus wavelength and photon energy at $T = 300\,K$. Adapted from Madelung E (ed.) (1982) *Numerical Data and Functional Relationships in Science and Technology, New Series, Group III: Crystal and Solid State Physics, vol. 17b: Physics of II–VI and II–VII Compounds, Semimagnetic Semiconductors*, Berlin: Springer-Verlag, Figure 52, with permission.

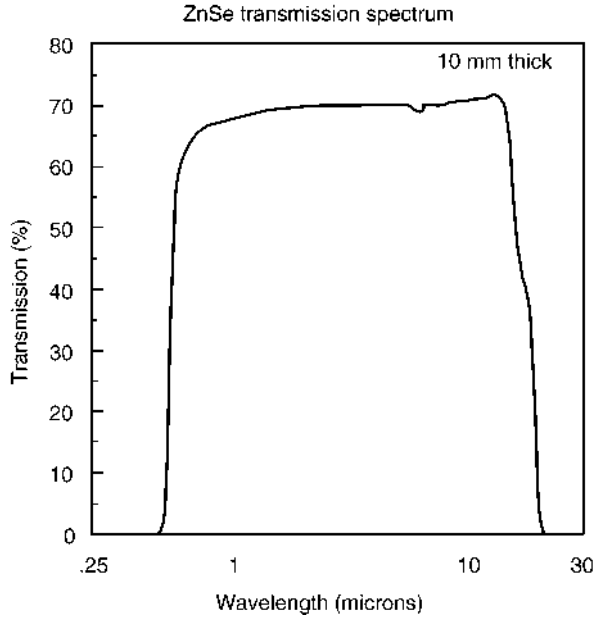

Figure 4 ZnSe transmission spectrum versus wavelength. Courtesy of II–VI Incorporated.

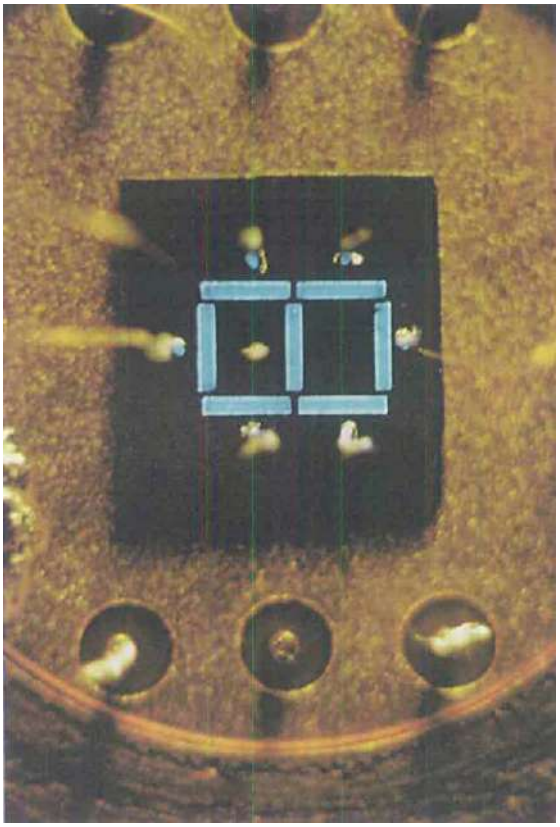

Figure 5 A seven-segment II–VI LED with blue emission taken at room temperature. Reproduced with permission from Hagerott M, Jeon H, Nurmikko AV, *et al.* (1992) Indium tin oxide as transparent electrode material for ZnSe-based blue quantum well light emitters. *Applied Physics Letters* 60: 2825–2827.

absorbed by the II–VI semiconductor, it can in turn produce light of its own. This light, known as luminescence, is again characteristic of the particular material and is the basis for the use of such materials in light-emitting diodes (LEDs) and laser diodes (LDs). In Figure 5 we show the luminescence from a II–VI LED. We see also from Figure 4 that at very long wavelengths, the ZnSe material also exhibits a large reduction in light transmission. This spectral region is termed the far infrared and the light absorbed at this spectral region produces vibrations of the atoms within the material, referred to as lattice absorption.

Light Scattering

Within the bulk of the material, small areas may have a different arrangement of atoms or may contain a substantial number of foreign atoms known as impurities. A large grouping of these impurities is known as an inclusion. In these regions, the optical properties will be different from the bulk of the material. Differences in both absorption coefficients and refractive indices can be found. These different refractive indices cause the light passing through the material to be scattered. Poor-quality material will often contain many inclusions. Even to the eye, these crystals will not look highly transparent. Changes in the refractive indices can also occur in highly pure materials due to vibrations of the atoms in the crystals. The lattice vibrations change the distance between the atoms in the material which alters the electric fields within the material and thereby changes the optical properties. Light can also be scattered due to these vibrations; experimental measurements of light scattering can provide useful information on the vibrations in these materials.

Optical Properties of II–VI Semiconductors: More Advanced Concepts

Light–Semiconductor Interactions

Light is an oscillating electromagnetic field whose interaction with semiconductor materials is in terms of the complex dielectric constant $\varepsilon_c = n_c^2$ where n_c is the complex refractive index defined as $n_c = n - ik$. k is known as the extinction coefficient and the value of k determines the amount of absorption in a material. The optical processes of reflection and absorption in semiconductors are governed by this complex dielectric constant. It turns out that there is a strong relationship between these processes. If the absorption spectrum is determined at all frequencies, then we can determine the refractive index spectrum at all frequencies by what is known as the Kramers–Kronig relations. From this description, we can now refine eqn [2] presented above:

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \qquad [3]$$

If the material is transparent at a particular wavelength, then $k = 0$ and we can determine the refractive index n. This is the case for values given in the table. The technique of spectroscopic ellipsometry can be used to determine the optical response of materials over a large wavelength range. This technique relies upon the change from linear polarization to elliptical polarization for light reflected from a material. From these measurements, the dielectric response can be determined.

If we look again at Figure 3, we see that there is a strong change in sample transmission around 0.5 μm. This corresponds to an energy of 2.18 eV

and is the bandgap energy of ZnSe. According to the details presented above, transitions are now being made from the full valence to the empty conduction bands of the semiconductor material. Once we have exceeded the bandgap energy, we can make further transitions between the bands. The spectrum shown in the figure is a broad band, reflecting the change from sharp atomic transitions to broad transitions in the solid state. Light with a wavelength greater than 0.5 μm does not have sufficient energy to effect the transitions between the bands. These wavelengths are not absorbed in the material. The material is transparent and the value of the extinction coefficient k is zero. The orange color of the crystal in Figure 2 corresponds to this transparent spectral region.

Cooling samples to low temperatures close to absolute zero has a profound effect on the optical properties of materials, particularly the absorption spectrum. In Figure 6, we show the absorption spectrum of a ZnSe sample at a temperature of 2 K. The absorption spectrum is determined from the measured transmission by the following equation

$$\alpha = \frac{1}{L} \ln \frac{I_0}{I_t} \qquad [4]$$

where α is the absorption coefficient, L the length of the sample, I_0 the light intensity incident on the sample and I_t the light intensity transmitted through the sample. In addition to the broad band spectrum there are now a number of sharp spectral lines near the bandgap energy.

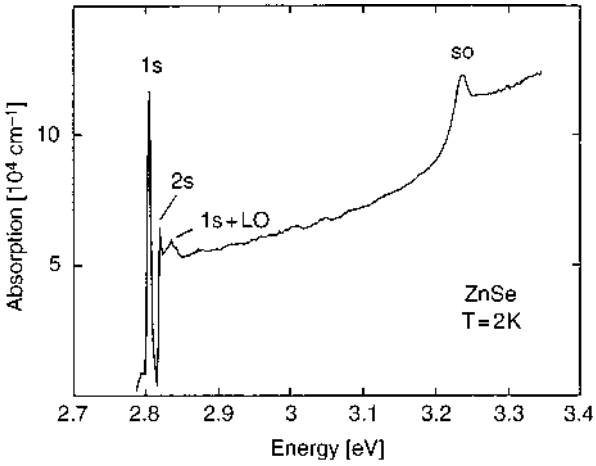

Figure 6 Low-temperature optical absorption spectrum of a ZnSe layer showing the 1S and 2S exciton features, the 1S + 1LO-phonon transition and the split-off (SO) exciton. Reproduced with permission from Bhargava R (ed.) (1997) *Wide Bandgap II–VI Semiconductors*, The Institute of Electrical Engineers: London, Figure 2 of Chapter 4.1.

These lines are due to exciton formation. An exciton is an electron–hole pair bound together by their mutual Coulomb interaction. It is the solid state equivalent of the Hydrogen atom (*see* **Semiconductor Physics:** Excitons). The bandgap energy at 2 K has increased from its value at 300 K. This is due to the thermal changes in the interatomic distance and the vibrational properties of the materials. Values of the bandgap energies determined from the position of the exciton lines in absorption spectra at low and at room temperatures are given in Table 1. At low temperatures, the luminescence is dominated by excitons, and we observe sharp line transitions. The width of the exciton transition is an excellent indicator of the quality of the material, and is often used by crystal growers to determine the quantity of defects in their materials.

It is possible for excitons and for free carriers (those that are not bound together) to recombine without giving out light. This nonradiative process usually involves the creation of local vibrations of the crystal lattice in the vicinity of defects in the material. We can define the radiative efficiency η_r as follows

$$\eta_r = \frac{R_r}{R_r + R_{nr}} \qquad [5]$$

where R_r is the radiative transition rate and R_{nr} is the nonradiative transition rate. In order to develop new optical devices based on II–VI semiconductors, it is essential that the value of R_{nr} be as small as possible. As R_{nr} is directly related to the concentration of defects in these materials, one can see why it is essential to make semiconductors with as few defects as possible. We can study these effects using a pulsed light source. Carriers are created during the pulse of light and recombine after the pulse has stopped. If we measure the luminescence intensity versus time, we can deduce the lifetime, which is given as

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}} \qquad [6]$$

where τ_r is the radiative and τ_{nr} the nonradiative lifetime, respectively. By varying the carrier density we can separately determine the contribution to the lifetime from radiative and nonradiative effects. In II–VI materials at room temperature, lifetimes are of the order of 1 ns.

Quantum Wells, High Carrier Density Effects, and Lasers

Recent developments in the growth of semiconductor materials has allowed the production of samples with monolayer accuracy (3–5 Å). If we grow a layered structure with materials of high bandgap on either side of a material of lower bandgap, we can observe quantum mechanical effects. These materials are called quantum wells (*see* **Semiconductor Physics:** Quantum Wells and GaAs-Based Structures). As an example, a 75 Å ZnSe quantum well can be grown between layers of ZnSSe. Due to the small size of the active material, quantum mechanical effects come into play, most notably an increase in the bandgap energy as the width of the quantum well decreases. The effect of placing this material between ZnSSe layers that have a larger bandgap is to confine the electrons and holes to the small region of the well. In this small region, they are less likely to encounter defects and so quantum well materials usually have a higher radiative efficiency than bulk II–VI materials which are grown by standard techniques. In Figure 7, we show the exciton absorption and luminescence from a ZnSe quantum well at low temperature.

If we increase the excitation level in these materials, we can eventually cause them to operate as lasers. At high carrier densities, nonradiative transitions become saturated and radiative efficiency increases. As the refractive indices in II–VI materials are quite large, the reflectivity of the two ends of the sample are usually sufficient to act as the mirrors for the laser cavity. All the II–VI semiconductor materials have exhibited laser action due to their direct bandgap. Most of the early studies of lasing were carried out at low temperature. In recent times, lasing action has been observed at room temperature in CdZnSe quantum well materials and there are attempts at present to develop commercial II–VI lasers for use in optical storage devices and flat panel displays. In Figure 8 we show the transition from luminescence through gain and into lasing action for a CdZnSe quantum well material.

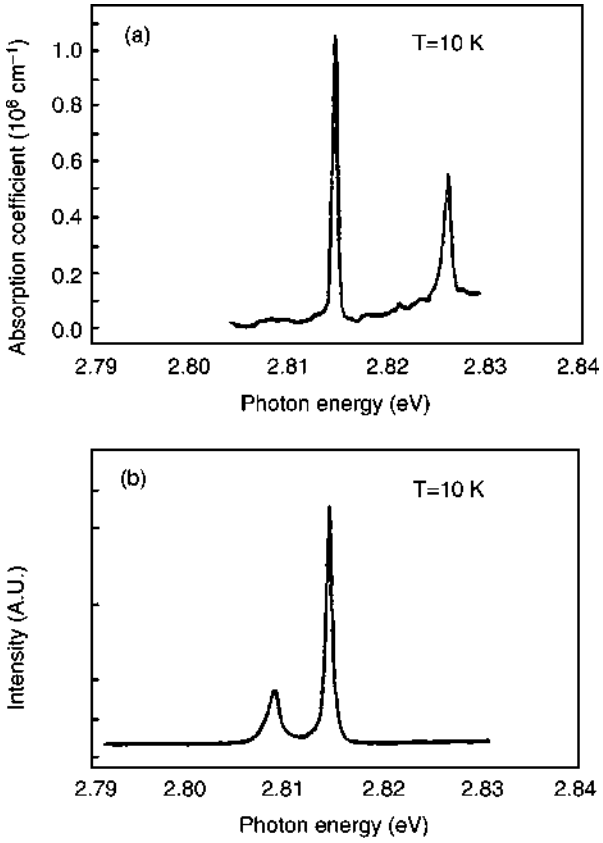

Figure 7 (a) Absorption and (b) luminescence from a single ZnSe quantum well at $T = 10$ K in the vicinity of the $n = 1$ quantum well exciton transition. Reprinted from Nurmikko AV, Jeon H, Gunshor RL and Han J (1996) II–VI lasers – directions. *Journal of Crystal Growth*, 159: 644–651 with permission from Excerpta Media Inc.

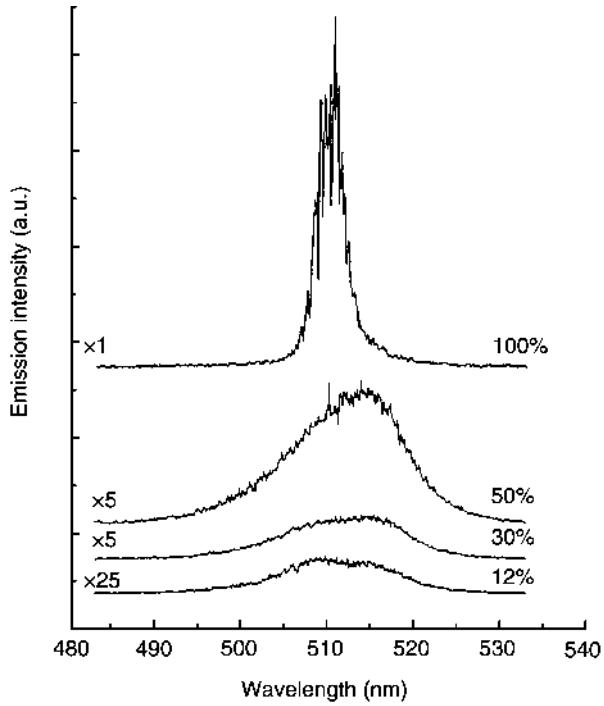

Figure 8 Emission spectra above and below lasing for a 40 Å $Cd_{0.28}/Zn_{0.72}Se/ZnSe$ system at room temperature. 100% corresponds to 1.1 times lasing threshold. Reprinted from Donegan JF, Jordan C, Rees P, Logue F, Heffernan JF and Hegarty J (1996) A study of internal loss in CdZnSe/ZnSe multiple quantum well materials. *Journal of Crystal Growth*, 159: 653–656 with permission from Excerpta Media Inc.

Dopants and Impurities

The key to the large-scale use of semiconductor materials is the large change in the electrical conductivity of the materials with the incorporation of small amounts of other elements. This process is known as doping and the foreign elements are known as dopants. The recent development of II–VI laser diodes in ZnSe materials is based on the p-n diode junction. Electrons injected into the n-type material recombine with holes injected into the p-type material at the junction between these two materials. A particular type of luminescence transition unique to semiconductor materials is the donor–acceptor pair (DAP) recombination. As donor and acceptor atoms are located in specific crystalline sites with respect to one another, we observe that the recombination energy is dependent on separation. Pairs that are close together often exhibit sharp lines while those far apart merge into a broad band. In Figure 9 we show the DAP recombination in ZnSe at 1.6 K.

It is also possible to incorporate foreign atoms in II–VI materials that have a small effect on the electrical properties but significantly change the optical properties. Such foreign atoms are termed impurities to distinguish them from dopants. An example is Mn^{2+} in ZnS. The Mn ions take the place of a small number of Zn atoms in the lattice. The luminescence from such a material is a broad emission in the visible region due to optical transitions on the Mn ions. ZnS:Mn materials are used as phosphors in flat-panel displays.

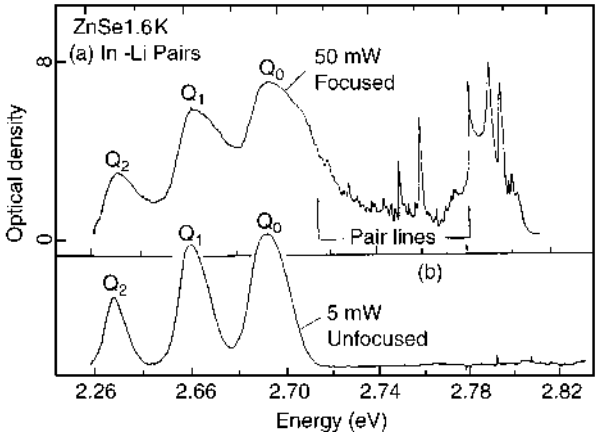

Figure 9 DAP recombination spectra of In-Li pairs in ZnSe showing pair lines, pair band (Q_0) and LO-phonon replicas (Q_1 and Q_2) at two different intensities (a) and (b). Adapted with permission from Merz JL, Nassan K and Shiever JW (1973) Pair spectra and the shallow acceptors in ZnSe. *Physical Review B* 8: 1444–1452.

Effects of Temperature, Pressure, Electric, and Magnetic Fields

The optical properties of II–VI materials are subject to variation due to temperature changes, external pressure and fields. Increasing temperature is found to lead to a decrease in the bandgap of II–VI materials. This decrease is due to the combined effects of an increase in the lattice spacing with temperature and the electron–phonon interaction. As the population of phonons is temperature dependent, this effect increases strongly as temperature increases and is the major contributor to the change in bandgap with temperature. A second effect of increasing the temperature is the change in the luminescence properties. In particular, it is found that the luminescence efficiency decreases and this is reflected in a reduced luminescence lifetime at room temperature compared with low-temperature values. Looking again at eqn [6], it is the nonradiative lifetime that changes rapidly with temperature and results in the decrease in the luminescence efficiency.

Pressure changes may be hydrostatic or uniaxial. In the hydrostatic case, the atoms are all brought closer together, and for the II–VI materials, this results in an increase in the bandgap energy. On the other hand, a uniaxial stress results in a decrease in the bandgap in the direction of the stress and an increase in the transverse directions. In II–VI materials with a wurtzite structure, the situation is more complex with two longitudinal and four transverse deformations possible.

External electric and magnetic fields can also be applied to II–VI semiconductors and thereby the optical properties can be changed. Electric fields polarize the materials and can lead to a reduction in the optical strength of transitions due to this polarization-induced carrier separation. Magnetic fields, on the other hand, lead to a splitting of the conduction and valence bands into Landau levels due to the circular motion of the carriers around the direction of the field.

See also

Semiconductor Physics: Band Structure and Optical Properties; Excitons; Impurities and Defects; Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Adachi S and Taguchi T (1991) Optical Properties of ZnSe. *Physical Review B* 43: 9569–9577.

Bhargava R (ed.) (1997) *Wide Bandgap II-VI Semiconductors*. London: Institute of Electrical Engineers.
Madelung E (ed.) (1982) *Numerical Data and Functional Relationships in Science and Technology, New Series, Group III: Crystal and Solid State Physics, vol. 17b: Physics of II-VI and I-VII Compounds, Semimagnetic Semiconductors*. Berlin: Springer-Verlag.
Madelung E (ed.) (1987) *Numerical Data and Functional Relationships in Science and Technology, New Series, Group III: Crystal and Solid State Physics, vol. 22a: Intrinsic Properties of Group IV Elements and III-V, II-VI and I-VII Compounds*. Berlin: Springer-Verlag.
Mar H (1992) Optical properties of wide bandgap II-VI semiconductors. In: Ruda HE (ed.) *Widegap II-VI Compounds for Opto-electronic Applications*, pp. 167–188. London: Chapman and Hall.
Nurmikko AV and Gunshor RL (1993) Physics and application of II-VI quantum confined structures. In: Henneberger F, Schmitt-Rink S and Göbel EO (eds) *Optics of Semiconductor Nanostructures*, pp. 229–262. Berlin: Akademie Verlag.
Pankove J (1975) *Optical Processes in Semiconductors*, 1st edn. New York: Dover Publications.
Stanley RP and Hegarty J (1993) Physics of excitons in wide bandgap II-VI quantum wells. In: Henneberger F, Schmitt-Rink S and Göbel EO (eds) *Optics of Semiconductor Nanostructures*, pp. 263–290. Berlin: Akademie Verlag.
Yu PY and Cadona M (1996) *Fundamentals of Semiconductors*. Berlin: Springer Verlag.
II–VI Compounds 1999. Ninth International Conference. Elsevier, Netherlands. *Journal of Crystal Growth* vol. 214–215, 2000.
II–VI Compounds 1997. Eighth International Conference. Elsevier, Netherlands. *Journal of Crystal Growth* vol. 184–185, 1998.
II–VI Compounds 1995. Seventh International Conference. Elsevier, Netherlands. *Journal of Crystal Growth* vol. 159, 1996.

Lead Salts

G Bauer and G Springholz, Johannes Kepler Universität, Linz, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The binary IV–VI compounds formed from Pb and group-VI elements like S, Se, and Te are among the oldest known semiconducting materials, and have been used for many years for electronic and opto-electronic device applications. The first ever reported solid state diode was made from single crystalline PbS and its rectifying properties were exploited in early radio receivers. Later on, interest in IV–VI semiconductors shifted to mid-infrared opto-electronic device applications such as photon detectors operating in the 3–14 μm wavelength range, taking advantage of the narrow energy gap between the valence and conduction band of the IV–VI compounds. Soon after the demonstration of lasing in III–V semiconductor diode structures, the first mid-infrared p-n junction laser was reported using $Pb_{1-x}Sn_xTe$, and since then efficient mid- and far-infrared IV–VI compound diode lasers have been fabricated, finding their main applications for remote sensing of gaseous pollutants in trace gas sensing devices, for toxic gas analysis systems, for human breath analysis in medical diagnostics, and for fabrication process control. Recently, IV–VI multiquantum-well structures have also attracted a lot of attention for their potential as efficient thermoelectric devices.

Basic Properties

In lead salt compounds, both the valence band maximum and the conduction band minimum occur at the L-points of the Brillouin zone. Thus these many-valley semiconductors have a direct gap, but not in the center of the Brillouin zone. In addition the small values of the energy gaps ($E_g < 410$ meV at $T = 300$ K) lead to a strong $\mathbf{k}\cdot\mathbf{p}$ interaction which in turn results in a strong nonparabolicity of the $E(k)$ dispersion relations. The small energy gaps also lead to small effective masses of electrons and holes, and the strong spin–orbit interaction (Pb is a very heavy element) leads to large values of the g-factors.

The bandgaps of the lead compounds have a positive temperature coefficient ($dE_g/dT > 0$) and a negative pressure coefficient ($dE_g/dp < 0$), i.e., the gaps increase with temperature and decrease with hydrostatic pressure. These anomalous properties are related to peculiarities of the band structure, among them the fact that in the lead compounds the Pb 6s levels are below the top of the valence band, and thus are occupied, in contrast to the group IV, III–V and II–VI compounds.

Almost all optical device applications of the IV–VI semiconductors are based on the materials subgroup of cubic lead chalcogenides PbTe, PbSe, and PbS, and their quasibinary alloys with other chalcogenide such as compounds, like PbSnTe, PbSnSe, or ternary alloys like PbEuTe, PbSrTe, PbEuSe, PbSrSe, etc. These compounds all crystallize in the rocksalt structure

with sixfold coordination and lattice constants in the 5.9 to 6.5 Å range. Mechanically, the lead salt compounds are much softer than the tetrahedrally bonded semiconductors, which has significant implications for device fabrication processes.

A further unusual property of the lead compounds is their huge static dielectric constant ε_0. In particular PbTe is close to a structural phase transition from the NaCl structure to a rhombohedral modification with ferroelectric properties. GeTe and SnTe are indeed ferroelectric with Curie temperatures of about 700 K and 145 K respectively, and also PbGeTe and PbSnTe become ferroelectric when the Ge or Sn content exceeds 1% or 20%, respectively. PbTe exhibits a softening of the transverse optic phonon mode with decreasing temperature. Although PbTe remains cubic, the temperature dependence of the static dielectric constant $\varepsilon_0 \propto C/(T-\theta)$ can be associated with an extrapolated negative Curie temperature θ of -75 K. As a consequence, the static dielectric constant increases strongly with decreasing temperature, reaching a value of $\varepsilon_0 = 1350$ at 4 K. This is the reason for the strong increase of the static dielectric constant at low temperatures as shown in Table 1, in which important properties of PbTe, PbSe, and PbS are summarized.

Also, the high-frequency dielectric constants ε_∞ of the lead salts are extremely high, which leads to refractive indices of about 4.4 for PbS, to 5.0 for PbSe and to about 6.0 for PbTe at $T = 300$ K, for photon energies just below the respective energy gaps.

As a result of the huge static dielectric constants, the hydrogen-like impurity binding energies are in the range of μeV, and the carrier scattering at ionized impurities is effectively screened. Consequently, the low-temperature carrier mobilities are high, exceeding 10^6 cm^2/Vs in PbTe. For doping of the lead salts two alternative methods are used, namely (1) adjustment of the stoichiometry of the compounds, or (2) doping with various kinds of impurities. Because the electronic states of these vacancies are resonant within the conduction or valence bands, each metal (Pb or Sn) vacancy acts as a doubly charged acceptor, and each chalcogen (Te, Se, or S) vacancy acts as a doubly charged donor. As a result, background doping levels below 10^{17} cm^{-3} are difficult. For doping with impurities, atoms from group IIIa (In, Ga, Tl), group Va (As, Sb, Bi), as well as group Ib elements (Cu, Ag) are used. These dopants are amphoteric p- or n-type, depending on whether they are incorporated substitutionally on metal or chalcogen lattice sites.

Other important aspects of the lead salts are their rather large linear thermal expansion coefficients of around 20×10^{-6} K^{-1} which are about a factor of three larger than that of GaAs, and six times larger than that of Si as shown in Table 2. The thermal conductivity is much lower than that of other semiconductors, which is favorable for their use in thermoelectric devices but causes problems in the removal of the dissipated heat in IV–VI compound diode lasers.

Table 1 Physical properties of some of binary IV–VI semiconductors

Parameter	*PbTe*	*PbSe*	*PbS*
Lattice constant a_0 (Å)	6.462	6.124	5.936
Melting point (°C)	930	1080	1113
Energy gap E_g			
at 300 K (meV)	319	278	410
at 77 K (meV)	217	176	307
at 4 K (meV)	190	147	286
Static dielectric constant ε_0			
at 300 K	414	210	169
at 77 K	1000	227	181
at 4 K	1350	280	–
High-frequency dielectric constant ε_∞	34	21	17
TO phonon mode frequency	18 (4 K) 33 (300 K)	37	65

Table 2 Expansion coefficients (10^{-6} K^{-1})

Compound	*Expansion coefficient*
PbS	20.3
PbSe	19.4
PbTe	19.8
GaAs	6
Si	2.6
BaF_2	18.8

Band Structure

A main property of the electronic band structure of the lead salt compounds is the small and direct energy gap of less than 410 meV (see Table 1). Since the maxima of the valence band and the minima of the conduction band are located at the same wavevector in reciprocal space, optical photons can be directly absorbed or emitted at the band edges. This makes the lead salts very well suited for opto-electronic applications such as infrared photon detectors or emitters. The band extrema are located at the edges in the $<111>$ direction of the Brillouin zone (L-points). In contrast to most other semiconductors, the conduction and valence bands are nearly mirror symmetric with almost equal effective masses for the electrons and holes. Because of the narrow energy gaps, the energy bands are strongly nonparabolic and the effective masses of electrons as well as holes are rather small (0.02–0.08 m_0 for the transverse masses,

Table 3 Band edge values of effective masses (in units of m_0) and g-factors (absolute values), $T = 2$ K

Parameter	*PbTe*	*PbSe*	*PbS*
m_t^{CB}	0.0207	0.037	0.081
m_l^{CB}	0.0213	0.0687	0.108
m_t^{VB}	0.0241	0.036	0.076
m_l^{VB}	0.263	0.066	0.108
g_t^{CB}	16.7	32.5	–
g_l^{CB}	66.1	44.1	12
g_t^{VB}	13.5	30.6	–
g_l^{VB}	65.6	–	13

Table 4 Pressure and temperature coefficients of the energy gaps

	PbTe	*PbSe*	*PbS*
dE_g/dp(eV Pa^{-1})	-7.5×10^{-11}	-9.1×10^{-11}	-9.1×10^{-11}
dE_g/dT (meV K^{-1})	0.457	0.457	0.462

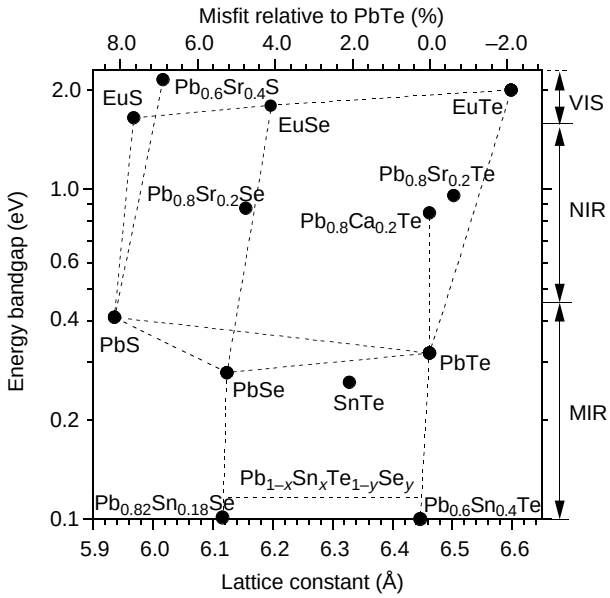

Figure 1 Energy bandgap versus lattice constant for several IV–VI compounds and alloys.

see Table 3). In addition, the bands are anisotropic, i.e., the Fermi surfaces are elongated ellipsoids of revolution around the $<111>$ axes, characterized by a longitudinal and transverse effective mass m_l and m_t parallel and perpendicular to the four-fold $<111>$ directions.

For PbTe the effective mass anisotropy is very large with a ratio of about 10 between longitudinal and transverse masses, whereas for PbSe and PbS this ratio is only 1.8 and 1.2, respectively (Table 2). Furthermore, in the alloy system $PbSe_{1-x}Te_x$ the energy gap shows a bowing, but not with a minimum value but rather a maximum gap for x at about 0.7. This is in contrast to the behavior found in direct-gap zincblende semiconductors: the bandgap $E_g(x)$ of an $A_{1-x}B_x$ alloy can generally be represented by $E_g(x) = (1-x)E_g(A) + xE_g(B) - bx(1-x)$ where b denotes the bowing coefficient. For III–V and II–VI compounds $b > 0$; whereas, for PbSeTe alloys $b < 0$.

Apart from the above-mentioned negative pressure coefficients of the direct gaps and their positive temperature coefficient which are tabulated in Table 4, there is a further intriguing anomaly in the chemical trends of the fundamental gap at the L-points: it decreases from PbS to PbSe, but increases from PbSe to PbTe, i.e., the direct gaps do not decrease monotonically with increasing atomic number of the anions. This decrease of the energy bandgap with decreasing temperature is substantial: i.e. it decreases from room temperature to 4 K by about 130 meV (see Table 1). This represents a large relative change with respect to the absolute value of the bandgap, and thus the emission of lead salt based diode lasers can be tuned over a broad wavelength range just by changing the operation temperature. This effect is actually utilized for spectroscopic applications of these lasers.

For adjustment of the fundamental absorption edge of infrared detectors as well as the adjustment of the wavelength emission of lead salt diode lasers, alloying of the lead salt compounds with other chalcogenide compounds is used. As illustrated in Figure 1, for this purpose many different ternary and quaternary alloys have been used. Important alloy systems for far-infrared applications are the lead–tin chalcogenide alloys. The tin chalcogenides (SnS, SnSe, SnTe) form single-phase pseudobinary alloy systems with the lead salts for all compositions. They have similar band structures as compared to the lead salt compounds; however, the arrangement of valence and conduction bands is exchanged. As a result, with increasing Sn content the energy bandgap of the ternary lead–tin chalcogenide alloys decreases and a zero bandgap semiconductor is reached for a certain ternary composition. For an operation temperature of 77 K, the bandgap is zero for a Sn content of $x_{Sn} \sim 40\%$ in the case of $Pb_{1-x}Sn_xTe$, and of $x_{Sn} \sim 20\%$ in the $Pb_{1-x}Sn_xSe$ case.

Optical Properties

The strong fundamental absorption in the lead salt compounds reflects the particularly high joint density of states at the direct gap at the L-points of the Brillouin zone. The main advantage of the lead

compounds with respect to other direct narrow-gap semiconductors like InSb, or HgCdTe alloys, with band extremes at the center, i.e., the Γ-point of the Brillouin zone, is related to the differences in the effective masses. Whereas the transverse effective masses of the IV–VI compounds have comparable values to the effective masses of the narrow-gap III–V and II–VI compounds, the longitudinal effective masses of the lead compounds are much larger. Thus the mass anisotropy together with the four-fold valley degeneracy at the L-point of the Brillouin zone leads to the substantially higher joint density of states and corresponding large absorption constants. Due to the rapid change of $\alpha(\omega)$ in the fundamental absorption range, a peak appears in the refractive index, $n(\omega)$, as shown in Figure 2 for PbTe, PbSe and for PbS. The shift of the peak with increasing temperature to higher photon energies reflects the increasing energy gaps.

In two-dimensional structures the changes of the absorption constant with frequency are particularly steep for energies which correspond to the onset of absorption between quantum confined electric subbands in the valence and conduction band. Such step-like features have indeed been observed in the absorption of PbTe/PbEuTe multiquantum-well (MQW) structures by Yuan *et al.* In Figure 3 the absorption constants and refractive indices are shown for a series of three MQW samples with different PbTe well widths. The steps in the absorption coefficients for the interband transitions associated to the longitudinal valleys with their main axis parallel to the [111] growth direction, denoted by 'l' as well as those associated by the oblique valleys, denoted by 'o' are accompanied by corresponding cusp-like enhancements of the refractive index. These cusps are of importance for the proper design of MQW laser structures.

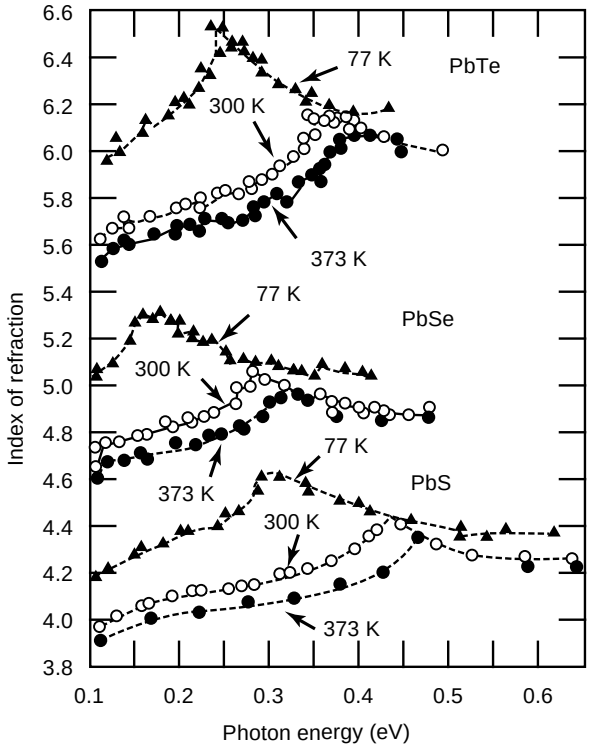

Figure 2 Refractive indices of PbTe, PbSe, and PbS as a function of photon energy at 77, 300, and 373 K. Reproduced with permission from Zemel JN, Jensen JD and Schoolar RB (1965) Electrical and optical properties of epitaxial films of PbS, PbSe, PbTe and SnTe. *Physical Review* 140: A330–A337. Copyright 1965, American Physical Society.

Because of the extremely high values of the refractive indices, the lead salts can easily be combined with materials of much lower refractive indices like EuTe, EuSe, or BaF_2 for efficient multilayer Bragg mirror structures. The high reflectivity of Bragg mirrors is caused by constructive interference of electromagnetic waves reflected at the subsequent interfaces of a multilayer structure. To obtain constructive interference, all interfaces have to be parallel, and, depending on the refractive indices of the media inside the Bragg mirror, the distance between subsequent interfaces should be an even or odd multiple of one quarter of the optical wavelength λ. In practice, stacks of two alternating layers with different refractive indices n_1 and n_2 and with a thickness equal to a quarter optical wavelength are used.

If for these λ/4 layers $Pb_{0.93}Eu_{0.07}Te$ and EuTe are used, a high index contrast of 68% is obtained and therefore very high reflectivities can be achieved already by a very small number of layer pairs. It was shown that a Bragg mirror with 3.5 periods of EuTe/$Pb_{0.93}Eu_{0.07}Te$ layer pairs for a target wavelength of 3.8 μm has only a total thickness as small as 2.55 μm. For comparison with results obtained for other material combinations it is useful to relate the stop bandwidth $\Delta\lambda$ to the center wavelength λ_T of the Bragg mirror. For EuTe/$Pb_{0.93}Eu_{0.07}Te$ this relative stop bandwidth $\Delta\lambda/\lambda_T$ is as large as 50%. An even wider stop bandwidth of 59% was demonstrated for a mirror with PbTe and EuTe λ/4 layers. This value represents, to the best of our knowledge, the highest relative stop bandwidth obtained for molecular beam epitaxy (MBE) grown Bragg reflectors.

Infrared Lasers

Among the infrared diode lasers, those based on IV–VI compounds play a particularly important role. They cover a wavelength range from 2.5 μm to about 30 μm. In fact, up to now most commercially

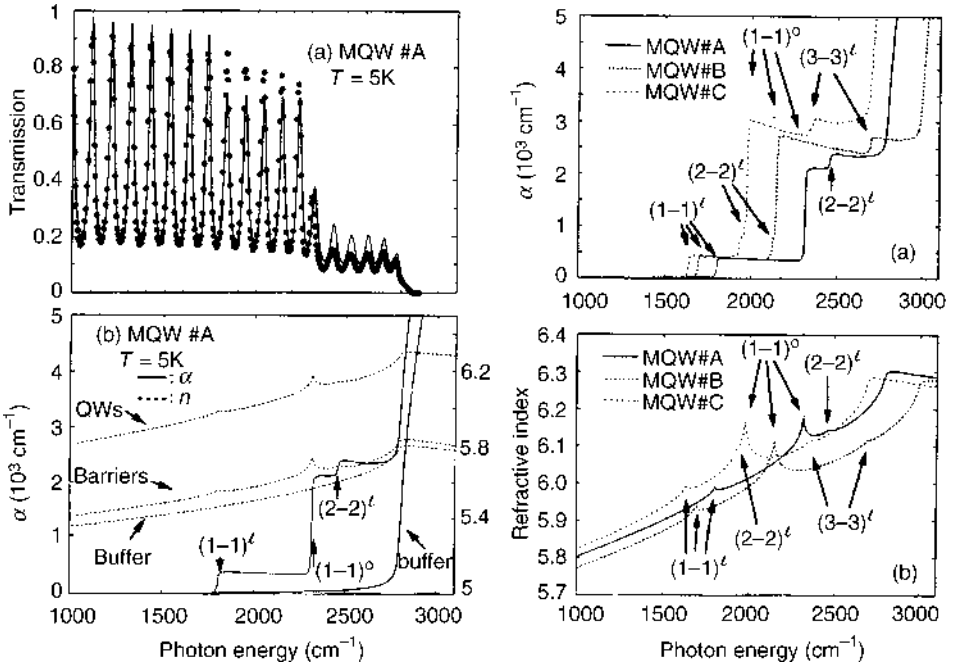

Figure 3 Left panel: (a) transmission versus energy for a $PbTe/Pb_{1-x}Eu_xTe$ multiquantum-well sample A (6.2 nm/62.1 nm, $x = 0.026$, 46 periods). Dots: experimental data; full line: calculation based on an envelope function model. (b) Frequency dependence of the absorption constant (full line) and of the refractive index (dotted lines). The arrows indicate interband transitions between electric sub-bands associated with valleys with main axis parallel to the [111] growth direction (l) and oblique to it (o). Right panel: (a) Comparison of the absorption constant of three PbTe/PbEuTe MQW samples (B: 9.3 nm/55.4 nm, $x = 0.03$, 30 periods; C: 11.8 nm/48.6 nm, $x = 0.024$, 30 periods). (b) Refractive index versus energy. Steps in the absorption constant due to onset of interband transitions are associated with cusps in the refractive index. Reproduced with permission from Yuan S, Springholz G, Bauer G and Kriechbaum M (1994) Electronic and optical properties of $PbTe/Pb_{1-x}Eu_xTe$ multiple-quantum-well structures. *Physical Review B* 49: 5476–5489, Copyright 1994, American Physical Society.

available mid- and far-infrared laser diodes are made from IV–VI compounds. The main material properties which are advantageous in this respect are (i) the electronic band structure with mirror-like bands at the L-point of the Brillouin zone and the resulting high joint density of states; (ii) the lower nonradiative Auger recombination rate in IV–VI compounds as compared to III–V and II–VI materials, and (iii) the high dielectric constant and the resulting weak detrimental influence of defects on nonradiative recombination losses. At wavelengths longer than 3 μm, IV–VI interband diode lasers are superior to their III–V counterparts for reaching high operating temperatures. So far IV–VI compound lasers hold extreme high cw operating temperature at wavelengths larger than 3 μm. With PbSe/PbSrSe diode lasers pulsed operation up to 60 °C was reported, whereas cw operation up to 223 K was achieved with separate confinement buried PbTe/PbEuSeTe heterostructure laser. Embedding PbTe quantum wells as the active medium in PbEuTe barriers between two dielectric Bragg mirrors optically pumped pulsed lasing was observed up to 65 °C. The recently developed III–V based quantum cascade lasers are rapidly improving and are becoming important competitors for IV–VI lasers. The commercial IV–VI lasers, however, are still based on conventional designs like homojunction, heterojunction, or single and multiple heterojunction lasers.

A unique property of lead salt laser diodes is the ease of tuning of the emission wavelength due to the strong change of bandgap of the IV–VI materials with increasing temperature. This, together with their narrow linewidths, makes these lasers ideally suited for spectroscopic applications, including research in molecular spectroscopy of ions or radicals in heterodyne and acoustic spectroscopy, or Doppler spectroscopy of molecular beams. So far the main applications are in time resolved combustion analysis in the automotive industry, in the monitoring of emissions of power plants, in exact monitoring of gas compositions in chemical plants, and spatially and time resolved *in situ* analysis.

Apart from the significant advances with IV–VI edge emitting diode lasers, vertical cavity surface

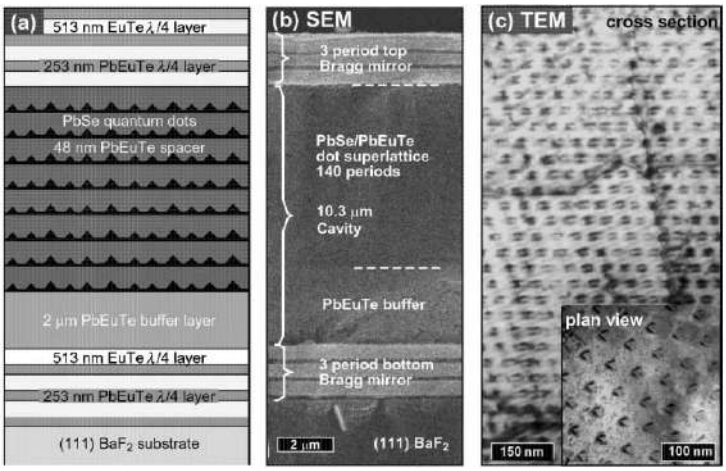

Figure 4 (a) Schematic cross-section of a PbSe quantum-dot VCSEL structure; (b) corresponding cross-sectional scanning electron micrograph; and (c) cross-sectional and plan-view transmission electron micrographs of a PbSe/PbEuTe dot superlattice sample reference sample. Reproduced with permission from Springholz G, Schwarzl T, Heiss W, *et al.* (2001) Midinfrared surface emitting PbSe/PbEuTe quantum dot lasers. *Applied Physics Letters* 79: 1225–1227. Copyright 2001, American Institute of Physics.

emitting lasers (VCSELs) have been successfully produced with optically pumped pulsed mode operation up to 50 °C. These lasers have attractive features like low-divergence circular output beam, low active volumes, and low thresholds.

High-finesse vertical cavity structures have also made it possible to realize mid-infrared surface-emitting PbSe/PbEuTe quantum dot lasers. As shown in Figure 4 in this case a self-organized PbSe/PbEuTe quantum dot superlattice was introduced into the active cavity region between Bragg mirrors consisting of three $\lambda/4$ layer pairs of 513 nm EuTe alternating with 253 nm $Pb_{0.94}Eu_{006}Te$. Optically pumped VCSEL emission at $\lambda = 4.2–3.9\ \mu m$ was achieved up to 150 K. The PbSe dots exhibit a face-centered cubic three-dimensional stacking in the superlattice. Lasing occurs simultaneously at the $m = 28^{th}$ and 29th order cavity modes, which shifts with increasing temperature to the $m = 29^{th}$ and $m = 30^{th}$ mode as shown in Figure 5. This simultaneous stimulated two-mode laser operation is a result of the inhomogeneously broadened quantum-dot gain spectrum, due to PbSe dot size fluctuations.

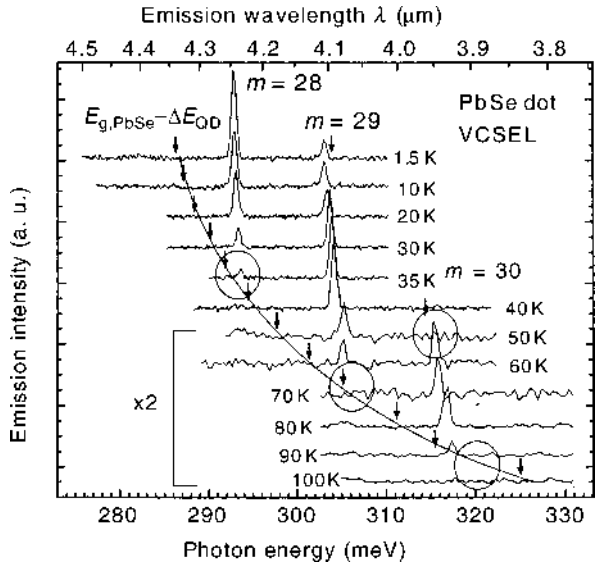

Figure 5 PbSe dot VCSEL emission spectra at various temperatures between 1.5 and 100 K for a pump power twice the threshold, showing the switching of the laser emission to higher cavity modes as the temperature increases. The arrows and the dashed line indicate the low-energy edge of the quantum-dot gain spectrum given by $E_{g,PbSe}(T) + \Delta E_{QD}$, where $\Delta E_{QD} = 141$ meV. Reproduced with permission from Springholz G, Schwarzl T, Heiss W, *et al.* (2001) Midinfrared surface emitting PbSe/PbEuTe quantum dot lasers. *Applied Physics Letters* 79: 1225–1227. Copyright 2001, American Institute of Physics.

Infrared Detectors

The photoconductive properties of PbS and PbSe have been used for the detection of infrared radiation already since about 1930. These detectors are still mass produced either by vacuum deposition or precipitation from aqueous solution in the presence of an oxidizing agent. The polycrystalline films deposited on glass or similar substrates are usually about 1 μm thick with grain sizes of about 1 μm.

Apart from polycrystalline thin-film detectors also single crystalline films have been used for infrared detection. In particular, the photoconductive response

of a sequence of n- and p-doped PbTe multilayers (so-called nipi structures) was shown to come close to that for the theoretical limit for the detectivity D^* of an ideal photodetector at temperatures between 77 and 90 K. Recently, lead selenide detectors have been used in advanced focal plane array (FPA) detectors and are commercially available as linear or bilinear FPAs with up to 526 detector elements. These elements are coupled to CMOS multiplexer readout circuits. The PbSe detector elements are thermoelectrically cooled, with entire power requirement for the FPAs and the cooler of less than 5 W. These structures are a low-cost alternative to the more elaborate FPAs based on HgCdTe technology.

The importance of thermal imaging by using focal plane arrays of many detector elements has recently stimulated further work with IV–VI epitaxial layers. In particular, the possibility of using Si substrates, as first shown by Zogg *et al.*, has brought a new impetus to the field of IV–VI infrared devices. With the use of Si wafers it is possible to integrate the detector of an infrared camera which consists of thousands of pixels to the read-out electronics which is usually based on standard silicon CMOS devices. In this case, the active lead salt infrared detector elements are grown by molecular beam epitaxy on CaF_2/BaF_2 buffer layers deposited on the Si (111) substrates after the fabrication of the electronic circuitry for the read-out. Conventional HgCdTe focal plane technology relies on hybrid structures with In bump bonds between the infrared devices and the Si substrate for each of the pixels. Due to the rather high dielectric constants of the IV–VI compounds as compared to HgCdTe, the high-frequency operation of IV–VI detector structures is limited to about 100 MHz. This upper frequency limit is, however, sufficient for infrared focal plane arrays. With IV–VI active epilayer thicknesses of few μm internal quantum efficiencies close to 100% and external efficiencies of about 60% have been achieved. Using materials like PbS, PbTe, PbEuSe, PbEuTe, or PbSnSe, the cutoff wavelength of the detector array can be varied from 4 to 14 μm at $T = 77$ K. The sensitivity of these detector structures is limited by generation-recombination noise for $T < 100$ K, whereas for higher temperatures these sensors are diffusion limited. The variation of quantum efficiencies between different pixels was shown to be less than about 3% for a 256 pixel device. Since all growth steps for these infrared sensors are performed at temperatures below 450 °C, a processing compatibility exists with standard Si-VLSI technology.

Summary

The peculiar structural, electronic, and optical properties of IV–VI semiconductors as compared to other semiconductor materials are a consequence of the ten valence electrons per atomic pair instead of the eight valence electrons typical for the tetrahedrally bonded group IV, III–V, and II–VI semiconductors. The most important group of IV–VI materials is the so-called lead chalcogenide (lead salt) compounds and their ternary and quaternary alloys. Their electronic band structure with the narrow direct minimum gap not at the center but at the edges of the Brillouin zone (L-points) exhibits a number of unique properties. The minimum energy gap increases with increasing temperature and decreases with hydrostatic pressure. The refractive indices are particularly large. As compared to narrow-gap III–V or II–VI semiconductors the Auger recombination rates are by two orders of magnitude smaller. All these properties provide the lead salts with great advantages for achieving infrared p-n junction lasing operation at comparatively high temperatures. Furthermore, efficient microcavity designs with extremely high reflectivities of 99% were realized with only three layer periods of combination like $PbSrTe/BaF_2$ or PbEuTe/EuTe due to their extremely high refractive index contrast. Further applications of the lead compounds are high sensitive infrared detectors, including focal plane array detectors which can be fabricated with comparatively low cost technologies, as well as thermoelectric devices.

List of Units and Nomenclature

Absorption *constant*	α
Absorption index	κ
Composition	x
Damping parameter	Γ_D
Detectivity	D^*
Dielectric function	ε
Energy	E
Energy gap	E_g
Free electron mass	m_0
High frequency dielectric constant	ε_∞
Kramers Kronig relations	KKR
Longitudinal effective mass	m_l
Longitudinal g-factor	g_l
Pressure	p
Refractive index	n
Static dielectric constant	ε_0
Temperature	T
Transverse effective mass	m_t
Transverse g-factor	g_t

See also

Nonlinear Optics, Basics: Kramers–Krönig Relations in Nonlinear Optics.

Further Reading

Bauer G (1980) Magnetooptical properties of IV–VI compounds. In: Zawadzki W (ed.) *Narrow Gap Semiconductors Physics and Applications*, pp. 470–494. Berlin: Springer Verlag.

Bauer G and Krenn H (1985) Lead selenide, lead telluride. In: Palik ED (ed.) *Handbook of Optical Constants*, pp. 517–523 and 535–546. New York: Academic Press.

Felix CL, Bewley WW, Vurgaftman I, *et al.* (2001) Low-threshold optically pumped $\lambda = 4.4$ μm vertical cavity surface emitting laser with a PbS quantum well active region. *Applied Physics Letters* 78: 3770–3772.

Holloway G (1980) Thin film IV–VI semiconductor photodiodes. In: Hass G and Francombe MF (eds) *Physics of Thin Films*, vol. 11, pp. 105–203. New York: Academic Press.

Jantsch W (1983) Dielectric properties and soft modes in semiconducting (PbSnGe)Te. *Springer Tracts in Modern Physics* 99: 1–50.

Katzir A, Rosman R, Shani Y, *et al.* (1989) Tuneable lead salt lasers. In: Cheo PK (ed.) *Handbook of Solid State Lasers*, pp. 227–347. New York: Marcel Dekker.

Khoklov D (2003) Lead chalcogenioles: Physics and applications. In: Manasreh MO (ed.) *Optoelectronic Properties of Semiconductors and Superlattices*, vol. 18. New York: Taylor and Francis.

Lucovsky G, Martin RM and Burstein E (1974) Structure, bonding and lattice dynamical properties of the IV–VI compounds. In: Rabii S (ed.) *Physics of IV–VI Compounds and Alloys*, pp. 93–102. New York: Gordon and Breach.

Nimtz G (1993) IV–VI compounds. In: Hellwege K and Madelung O (eds) *Numerical Data and Functional Relationships in Science and Technology*. Landolt-Börnstein, new series, group III, vol. 17, Part F, pp. 152–180. Berlin: Springer Verlag.

Partin DL and Heremans J (1994) Growth of narrow bandgap semiconductors. In: Moss TS and Mahajan S (eds) *Handbook on Semiconductors*, vol. 3A, pp. 369–450. Amsterdam: North-Holland.

Preier H (1990) Physics and applications of IV–VI compound semiconductor lasers. *Semiconductor Science and Technology* 5: S12–S20.

Ravich YI, Efimova BA and Smirnov IA (1970) Semiconducting Lead Chalcogenides. New York: Plenum.

Schwarzl T, Heiss W and Springholz G (1999) Ultra-high-finesse IV–VI microcavities for the midinfared. *Applied Physics Letters* 75: 1246–1248.

Springholz G, Shi Z and Zogg H (1998) Molecular beam epitaxy of IV–VI semiconductors. In: Liu AWK and Santos MB (eds) *Heteroepitaxial Thin Film Systems*, pp. 621–688. Singapore: World Scientific.

Springholz G, Schwarzl T, Heiss W, *et al.* (2001) Mid-infrared surface-emitting PbSe/PbEuTe quantum dot-lasers. *Applied Physics Letters* 79–81: 1225.

Tacke M (1995) New developments and applications of tunable IR lead salt lasers. *Infrared Physics and Technology* 36: 447–463.

Yuan S, Springholz G, Bauer G and Kriechbaum M (1994) Electronic and optical properties of PBTe/$PB_{1-x}Eu_x$Te multiple quantum well structures. *Physical Review* B49: 5476–5489.

Zemel JN, Jensen JD and Schoolar RB (1965) Electrical and optical properties of epitaxial films of PbS, PbSe, PbTe and SnTe. *Physical Review* 140: A330–A337.

Zogg H, Fach A, John J, *et al.* (1995) Photovoltaic IV–VI on Si infrared sensor arrrays for thermal imaging. *Optical Engineering* 34: 1964–1969.

Mercury Cadmium Telluride

M B Reine, BAE SYSTEMS, Lexington, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Mercury-cadmium telluride (HgCdTe or $Hg_{1-x}Cd_x$Te) is a pseudobinary alloy semiconductor that crystallizes in the zincblende structure. The energy bandgap $E_G(x,T)$ varies continuously, and nearly linearly, with alloy composition parameter x, ranging from 1.6 eV for the wide-gap semiconductor CdTe to −0.3 eV for the semimetal HgTe. The bandgap passes through zero at an alloy composition of $x = 0.16$ (for $T = 0$ K). This broad range of energy bandgap provides a continuum of materials whose absorption edge can be tailored to provide high-performance quantum detectors for wavelengths over the 1–30 μm spectral range. Direct allowed valence-to-conduction band transitions are responsible for large values of the optical absorption coefficient, which enable quantum efficiencies approaching 100% to be achieved in practical device designs. Long minority carrier lifetimes result in low thermal noise, and permit high performance to be achieved at operating temperatures that are the highest reported among infrared detectors of comparable cutoff wavelengths. HgCdTe is well-established today as

the most broadly applicable high-performance infrared detector material.

This article summarizes the fundamental properties of this versatile alloy semiconductor, and relates the material properties to its success as an infrared photoconductive and photovoltaic detector material.

Historical Background

It is noteworthy that HgCdTe is probably the first example of a semiconductor material that was intentionally created to satisfy well-defined theoretical device requirements. The importance of the infrared atmospheric transmission window at 8–12 μm was well known for thermal imaging systems, which enable night vision by imaging emitted infrared radiation from the scene rather than reflected visible radiation. The only available infrared detectors with the required sensitivity at 8–12 μm were extrinsic photoconductors such as copper-doped germanium, in which the photoinduced transitions were between a donor or acceptor level and the adjacent band extremum. However, such extrinsic detectors required cooling to prohibitively low temperatures, such as 10–15 K, which made thermal imaging systems bulky and impractical.

Theory showed that the reasons for such a low operating temperature were fundamental to the extrinsic detector material itself – essentially a short carrier lifetime that resulted in a large thermal generation rate and a correspondingly large thermal noise. Theory also showed that an intrinsic detector, in which the photoinduced transitions were direct (vertical) transitions between the valence band and the conduction band, could have significantly longer carrier lifetimes. This meant that an intrinsic detector, with a bandgap of 0.1 eV that matched the 8–12 μm band, could achieve the required sensitivity at much higher operating temperatures, as high as 77 K, that of liquid nitrogen. These theoretical and practical considerations motivated the search in the late 1950s for a semiconductor with a direct bandgap of 0.1 eV.

The first report of the synthesis of the semimetal HgTe and the wide-bandgap semiconductor CdTe to form the semiconductor alloy HgCdTe was published in 1959 by the Royal Radar Establishment in Malvern, England. This seminal paper reported both photoconductive and photovoltaic response at wavelengths extending to 12 μm, and made the understated observation that this material showed promise for intrinsic infrared detectors.

Soon thereafter, working under a US Air Force contract with the objective of devising an 8–12 μm background-limited semiconductor infrared detector that would operate at temperatures as high as 77 K, the group led by Paul Kruse at the Honeywell Corporate Research Center in Hopkins, Minnesota developed a modified Bridgman crystal growth technique for HgCdTe. In 1962 they reported both photoconductive and photovoltaic detection in rudimentary HgCdTe devices.

Fundamental Material Properties

Three fundamental properties determine the degree to which a semiconductor will make a successful infrared detector. First, the energy bandgap must be approximately equal to the smallest photon energy to be detected. There can be severe cooling penalties if the bandgap is significantly smaller than the smallest photon energy of interest. Second, the optical absorption coefficient must be large enough so that nearly all incident photons of interest are converted to electron–hole pairs for practical values of device thickness. Third, the inherent recombination mechanisms must allow long carrier lifetimes to be realized so that thermal generation rates, and hence thermal noise, can be minimized for practical devices and operating conditions.

The three key features that make HgCdTe the nearly ideal infrared detector material – tailorable energy band gap over the 1–30 μm range, large optical absorption coefficients that enable high quantum efficiency, and favorable inherent recombination mechanisms that lead to high operating temperature – are direct consequences of the energy band structure of this zincblende semiconductor. They apply no matter whether the device embodiment is a photoconductor, a photodiode or a metal–insulator–semiconductor detector. In addition, HgCdTe has many other favorable material properties, summarized in this article, that are technologically important for realizing practical infrared detectors. Tables 1 and 2 summarize various material properties for $Hg_{1-x}Cd_xTe$. **Table 1** summarizes material properties for $Hg_{1-x}Cd_xTe$ that vary with alloy composition x and temperature T: lattice constant a, energy band gap E_G, cutoff wavelength λ_{CO} (= 1.24/E_G), intrinsic carrier concentration n_i, conduction band effective mass ratio m_C/m_0, conduction band g-factor g_C, static and high frequency dielectric constants $\varepsilon_S/\varepsilon_0$ and $\varepsilon_\infty/\varepsilon_0$, index of refraction $n_R = (\varepsilon_\infty/\varepsilon_0)^{1/2}$, electron and heavy hole mobilities μ_e and μ_{hh}, mobility ratio b, radiative and Auger-1 lifetimes τ_{RAD} and $\tau_{Auger\text{-}1}$, and typical lifetimes $\tau_{typical}$ observed in n-type HgCdTe of low carrier concentration ($<1\times10^{15}\,cm^{-3}$).

Table 1 Material properties for selected compositions of $Hg_{1-x}Cd_xTe$, including the binary components HgTe and CdTe, and several technologically important alloy compositions. The radiative and Auger-1 lifetimes τ_{RAD} and $\tau_{Auger\text{-}1}$ are calculated for n-type HgCdTe with $N_D = 1 \times 10^{15}$ cm^{-3}

Property	*HgTe*	*$Hg_{1-x}Cd_xTe$*						*CdTe*
x	0	0.194	0.205	0.225	0.31	0.44	0.62	1.0
a (Å)	6.461	6.464	6.464	6.464	6.465	6.468	6.472	6.481
T (K)	77	77	77	77	140	200	250	300
E_G (eV)	−0.261	0.073	0.091	0.123	0.272	0.474	0.749	1.490
λ_{CO} (μm)		16.9	13.6	10.1	4.6	2.6	1.7	0.8
n_i (cm^{-3})		1.9×10^{14}	5.8×10^{13}	6.3×10^{12}	3.7×10^{12}	7.1×10^{11}	3.1×10^{10}	4.1×10^5
m_C/m_0		0.006	0.007	0.010	0.021	0.035	0.053	0.102
g_C		−150	−118	−84	−33	−15	−7	−1.2
$\varepsilon_S/\varepsilon_0$	20.0	18.2	18.1	17.9	17.1	15.9	14.2	10.6
$\varepsilon_\infty/\varepsilon_0$	14.4	12.8	12.7	12.5	11.9	10.8	9.3	6.2
n_R	3.79	3.58	3.57	3.54	3.44	3.29	3.06	2.50
μ_e (cm^2/V-s)		4.5×10^5	3.0×10^5	1.0×10^5				
μ_{hh} (cm^2/V-s)		450	450	450				
$b = \mu_e/\mu_{hh}$		1000	667	222				
τ_{RAD} (μs)		16.5	13.9	10.4	11.3	11.2	10.6	
$\tau_{Auger\text{-}1}$ (μs)		0.45	0.85	1.8	40	450	4.8×10^3	
$\tau_{typical}$ (μs)		0.4	0.8	1	7			

Table 2 Material properties for $Hg_{1-x}Cd_xTe$ that are independent of or relatively insensitive to alloy composition

Property	*Symbol*	*Value*
Kane matrix element	E_P	19 eV
Split-off band energy	Δ	0.93 eV
Heavy-hole effective mass ratio	m_{hh}/m_0	0.40–0.53
Valence band offset	E_v(HgTe) − E_v(CdTe)	0.35–0.55 eV

Energy Band Gap and Band Structure

The energy band gap $E_G(x, T)$ of $Hg_{1-x}Cd_xTe$ varies continuously, and nearly linearly, with alloy composition parameter x, ranging from 1.6 eV for the wide-gap semiconductor CdTe to −0.3 eV for the semimetal HgTe, as shown in Figure 1a. Also plotted in Figure 1a is the cutoff wavelength $\lambda_{CO}(x, T)$ ($\approx hc/E_G(x, T)$, where h is Planck's constant and c is the speed of light), defined as that wavelength at which the photoresponse falls to 50% of its peak value. The energy bandgap $E_G(x, T)$, defined at the difference between the Γ_6 and Γ_8 band extrema at $\Gamma = 0$, passes through zero at an alloy composition of $x = 0.16$ (for $T = 0$ K), as illustrated in Figure 1b. This broad range of energy bandgap provides a continuum of materials whose absorption edge can be tailored to provide high-performance quantum detectors for wavelengths over the 1–30 μm spectral range.

Optical Absorption Coefficient

Optical absorption coefficient data for several alloy compositions of $Hg_{1-x}Cd_xTe$ are plotted versus wavelength in Figure 2. Several trends are evident. The absorption strength generally decreases as the gap becomes smaller due both to the decrease in the conduction band effective mass and to the $\lambda^{-1/2}$ dependence of the absorption coefficient on wavelength λ. There is an exponential tail at energies just below the bandgap energy.

Carrier Lifetime Mechanisms

There are two fundamental mechanisms that determine the minority carrier lifetime in defect-free HgCdTe: radiative recombination and Auger recombination (Auger-1 in n-type and Auger-7 in p-type). Because they involve only interactions among electrons and holes in the valence and conduction bands, and do not require the intermediary of defect states, these mechanisms are regarded as inherent to the material itself, depending only on the band structure, the donor or acceptor concentration, and the temperature. The important advantage that HgCdTe enjoys as a high-performance detector material is that these fundamental mechanisms provide, for practically achievable

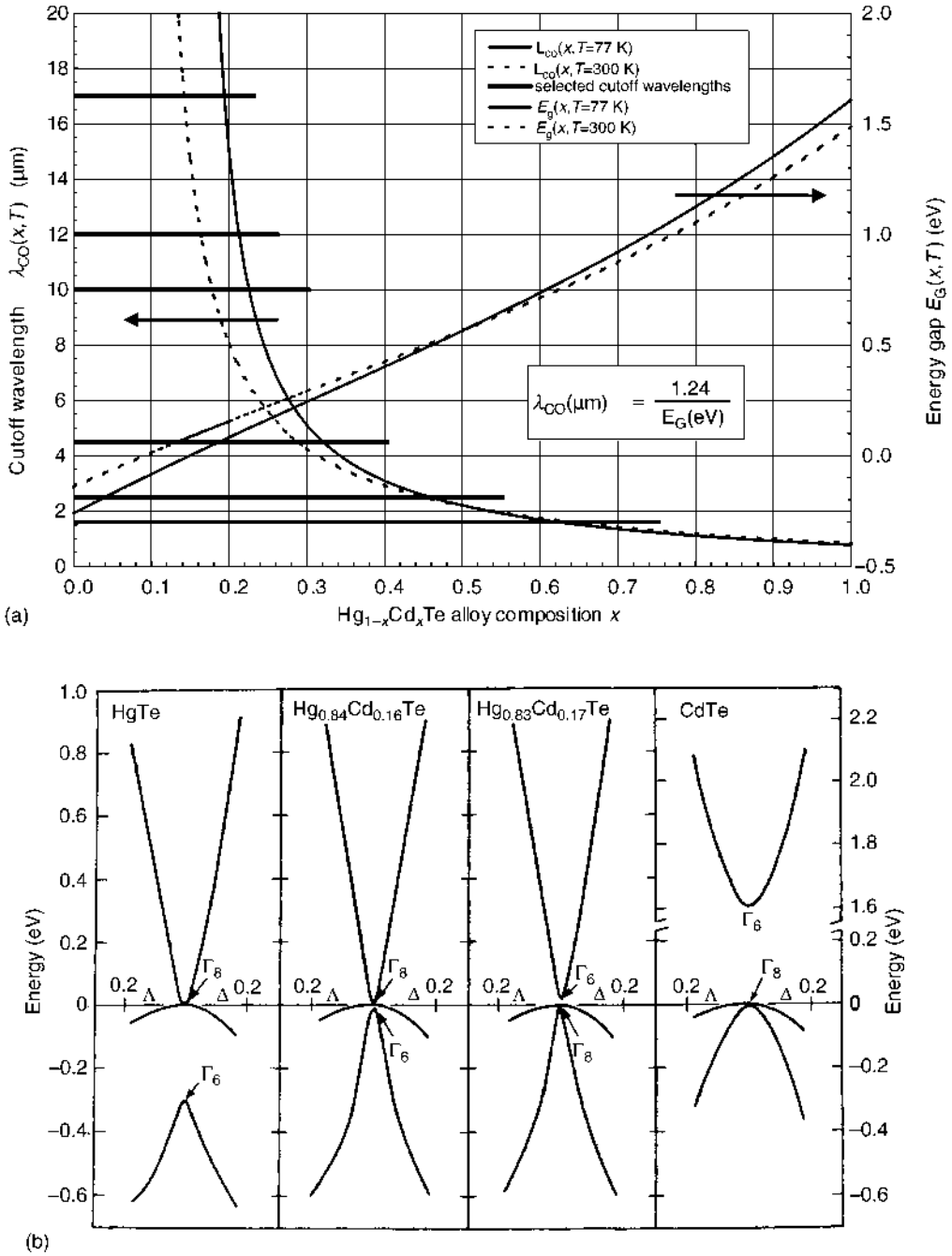

Figure 1 (a) Energy bandgap $E_G(x, T)$ and cutoff wavelength $\lambda_{CO}(x, T)$ for $Hg_{1-x}Cd_xTe$, plotted versus alloy composition parameter x, for temperatures of 77 K (solid curves) and 300 K (dashed curves). Horizontal bars indicate important cutoff wavelengths for HgCdTe detector technology. (b) Band diagrams near the Γ point calculated for HgTe, for two compositions of $Hg_{1-x}Cd_xTe$, and for CdTe, illustrating the transition from negative to positive energy bandgap. The energy bandgap is defined at the difference between the Γ_6 and Γ_8 band extrema at $\Gamma = 0$. Part (b) reproduced with permission from Chadi DJ and Cohen ML (1973) Electronic structure of $Hg_{1-x}Cd_xTe$ alloys and charge-density calculations using representative K points. *Physical Review B* 7: 692–699.

carrier concentrations, lifetimes that are long enough (and hence thermal generation rates that are low enough) that background-limited sensitivity is achieved at temperatures substantially higher than for other classes of infrared detectors, such as extrinsic detectors (As-doped Si, Cu-doped Ge) and quantum-well infrared photodetectors (QWIPs).

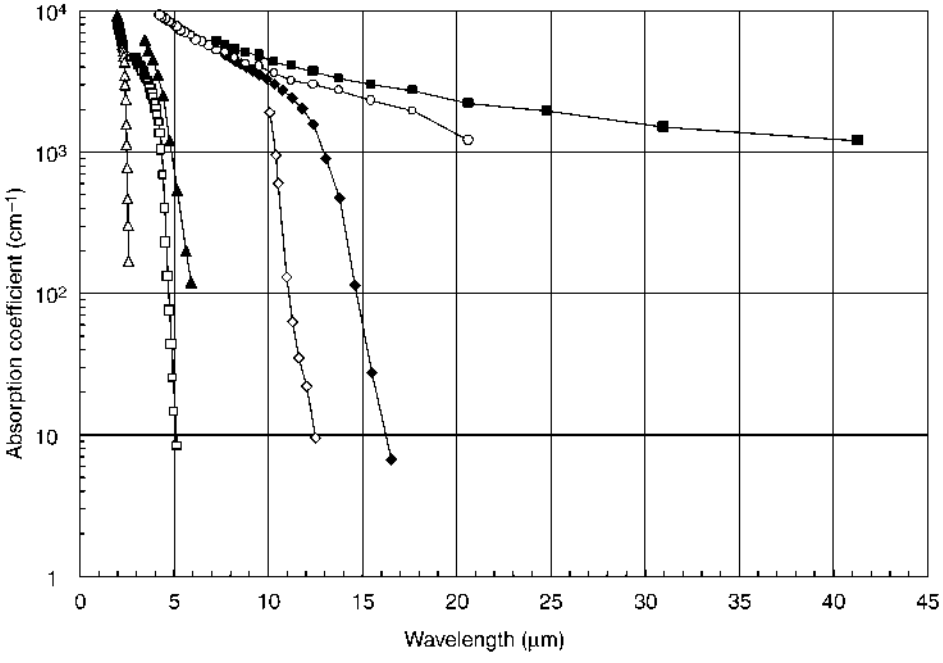

Figure 2 Optical absorption coefficient data for several $Hg_{1-x}Cd_xTe$ alloy compositions, for photon energies near the fundamental absorption edge, plotted versus wavelength.

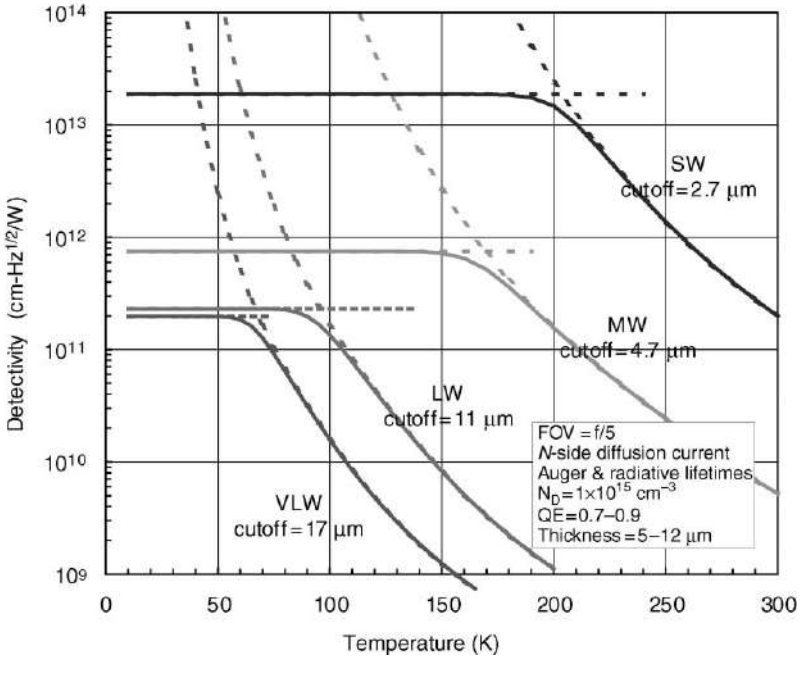

Figure 3 Calculated detectivity (D^*) for $Hg_{1-x}Cd_xTe$ infrared detectors for four technologically important wavelength regions, plotted versus operating temperature.

Figure 3 illustrates the sensitivities and operating temperatures that can be achieved for HgCdTe infrared detectors for four important wavelength regions. Detectivity, D^*, is the normalized signal-to-noise ratio that is the key figure of merit for the sensitivity of an infrared detector. At lower temperatures, detector thermal noise is negligible, and D^* is limited by detector noise due to

fluctuations in the arrival rate of photons from the thermal room-background radiation. This limit is referred to as the BLIP (background-limited infrared photodetector) limit, and is indicated in Figure 3 by the horizontal dashed lines. As detector temperature increases, the detector thermal noise increases exponentially, and eventually overcomes the background noise, causing D^* to decrease exponentially for further increases in temperature. Detector thermal noise is proportional to the thermal generation rate, which is inversely proportional to the carrier lifetime.

n-Type and p-Type Doping

HgCdTe can be made n-type or p-type by a number of relatively convenient methods, at carrier concentrations required for high-performance n-type photoconductor and p-n junction photodiode architectures. Common donors are indium and iodine. Arsenic has become the most commonly used acceptor dopant, although copper and gold are used in some cases. Some HgCdTe photodiode designs still employ native metal-vacancy point defects as acceptors, although the clear trend is toward the use of extrinsic acceptor doping to avoid the strong Shockley–Read recombination associated with the Hg vacancy. Residual electrically active impurity concentrations are generally less than $1 \times 10^{14}\,\mathrm{cm}^{-3}$, which allows controllable doping at low-$10^{14}\,\mathrm{cm}^{-3}$ concentrations.

Crystal Growth Methods

The first growth methods for HgCdTe were bulk growth. Liquid phase epitaxy (LPE) began to be developed around 1975, followed by metalorganic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE) in the early 1980s. Bulk growth, LPE, MOCVD, and MBE techniques have been refined through the years, and each is still in use today. Each technique has found application for certain types of HgCdTe devices. Bulk growth methods are used for production quantities of n-type wafers, with diameters of 12–20 mm, for LW and VLW photoconductive arrays, which are more tolerant of dislocations. LPE is in use today for production quantities of photovoltaic (PV) HgCdTe arrays, and has also been used for certain photoconductive HgCdTe arrays. Vapor-phase epitaxy (VPE) methods such as MBE and MOVPE are in use for engineering and prototype quantities. Both run-to-run reproducibility and control as well as intra-wafer uniformity of HgCdTe alloy composition are well in hand, meeting or exceeding requirements.

LPE-based one-layer and two-layer photodiode technology is well established today, offering the lowest defect densities and highest operabilities for production quantities of high performance FPAs. The VPE methods, MBE and MOVPE, will probably not entirely replace LPE in the near future, but will continue to improve, particularly with respect to lower defect densities, offering *in situ* growth of advanced bandgap-engineered detector designs that are unwieldy or impossible with LPE, such as dual-band detectors, avalanche photodiodes, and planar buried junctions with *in situ* CdTe passivation. MBE has shown the potential for rapidly switching from growth of one HgCdTe alloy composition to another in successive growth runs, without the necessity of preparatory calibration growth runs. This compositional agility, important for the rapid and cost-effective adjustment from product to product, will be further enhanced by the ongoing efforts to implement and improve *in situ* real-time monitors and feedback controls.

Substrates for Epitaxial Growth

The lattice mismatch between HgTe and CdTe is small, approximately 0.3%. This allows epitaxial growth of high-quality HgCdTe films on IR-transparent CdTe or nearly-lattice-matched IR-transparent $Cd_{1-z}Zn_zTe$ ($z \approx 0.04$) substrates, with dislocation densities in the mid-$10^4\,\mathrm{cm}^{-2}$ range. It also allows the *in situ* growth of various isotype and anisotype heterojunctions with tolerably low or negligible densities of misfit dislocations. This has led to an increasing number of bandgap-engineered HgCdTe photodiode structures, such as dual-band or two-color detector arrays and Auger-suppressed photodiodes with the potential for greatly increased operating temperature.

Sapphire and silicon are IR-transparent substrates that are less costly, that are available in much larger areas, and that are more rugged than CdTe and CdZnTe. Both sapphire and silicon can be used for epitaxial growth of HgCdTe films with dislocation densities that are acceptably low (mid-$10^6\,\mathrm{cm}^{-2}$) for many important photodiode applications, such as for the MW and SW spectral ranges, and for the LW spectral range for high background photon fluxes. LPE growth of HgCdTe($x = 0.4$) is done on 3-inch diameter sapphire substrates with MOCVD-grown CdTe buffer layers. MBE growth of HgCdTe($x = 0.3$) on 4-inch diameter (211) silicon substrates, with ZnTe/CdTe buffer layers, has recently been reported.

Dielectric Constant

The relatively low dielectric constant ($\varepsilon_S = 18\varepsilon_0$) of HgCdTe, in contrast to those of PbTe and PbSnTe, permits low junction capacitance. This is important for fast response in laser pulse detectors where small $R-C$ time constants are needed. It is also important for suppressing preamplifier noise below the detector noise, and thereby achieving high detectivity (D^*), in photodiodes operating at low temperature and low background fluxes, where the junction resistance is sufficiently high that the junction capacitance becomes an important contribution to the junction impedance at frequencies of interest.

Surface Passivation

Passivation of the exposed HgCdTe surfaces is critical for both device performance and long-term stability. The favorable surface properties of HgCdTe have enabled practical surface passivation schemes to be developed for both n-type photoconductors and p-n junction photodiodes of both planar and mesa configurations.

n-type HgCdTe photoconductors are commonly passivated with a native oxide that accumulates the surface, resulting in surface recombination velocities as low as 50 cm s^{-1}, along with low $1/f$ noise (knee frequencies less than 50 Hz). The surface accumulation layer also provides a layer of surface electrons with high mobility, though lower than the bulk electron mobility, such that the shunt conductance of this layer is tolerably low compared to the bulk conductance of the detector itself. Several methods have been used for growing this native oxide, including plasma deposition and electrochemical (anodic) oxidation.

HgCdTe photovoltaic detectors require quite different passivation because usually both n-type and p-type surfaces are exposed as well as the depletion region. The CdTe/HgCdTe heterostructure is an important passivation for HgCdTe photovoltaic detectors and arrays. CdTe passivation has enabled arrays of HgCdTe photodiodes with negligible $1/f$ noise (knee frequencies less than 1 Hz), radiation hardness well in excess of the 20–40 kRad(Si) usually encountered in Earth-resource space missions, and excellent stability to thermal bake and to exposure to moisture.

CdTe has a number of important advantages as a passivant for HgCdTe junctions. It is a 'native' passivant that is chemically compatible with HgCdTe. Adhesion is excellent. The valence band edge of CdTe is approximately 0.4 eV below that of HgTe, thereby allowing repulsive barriers to form in the underlying HgCdTe for both electrons in the conduction band and holes in the valence band. Both mesa and planar photodiodes have been successfully passivated with CdTe. A wide variety of methods has been reported for deposition of CdTe passivating layers, including MBE, MOVPE, e-beam evaporation, sputtering, and hot wall epitaxy.

Electron-to-Hole Mobility Ratio

The small values for the conduction band effective mass ratio lead to large values for the electron mobility, as large as 4.5×10^5 cm^2/V-s for $x = 0.195$ at 77 K. The heavy hole effective mass ratio is quite large, approximately 0.5 for all alloy compositions, leading to low values for the heavy hole mobility, on the order of 500 cm^2/V-s. This gives large values for the electron-to-heavy hole mobility ratio, b, as high as 1000 for $x = 0.194$ at 77 K.

These large values of b were of great benefit to n-type HgCdTe photoconductors. Most such devices have small active areas, on the order of 50×50 μm^2, and the drift length for minority carriers (holes) becomes comparable to the interelectrode spacing at modest bias voltages. In this case, when the contacts have a high recombination velocity, the photoconductive gain saturates to the value $b/2$. High values of b in HgCdTe allowed sufficiently high photoconductive gains to be achieved to raise the detector noise above the preamp noise, an essential condition for achieving BLIP sensitivity.

HgCdTe Infrared Detector Configurations

Three different types of infrared detectors can be realized in the $Hg_{1-x}Cd_xTe$ alloy semiconductor: the photoconductive (PC) detector, the photodiode, also referred to as the photovoltaic (PV) detector, and the metal–insulator–semiconductor (MIS) detector. These are illustrated in Figure 4.

In all three types of HgCdTe detector – PC, PV, and MIS – excess electron–hole pairs are photogenerated by the same mechanism: valence-to-conduction band absorption of infrared radiation with photon energy greater than the bandgap energy E_G. All three types can achieve high quantum efficiencies, approaching 100% for well-designed devices. All three types have the same fundamental recombination mechanisms (Auger-1 and radiative) that determine the highest possible operating temperature to achieve a given sensitivity (signal-to-noise ratio).

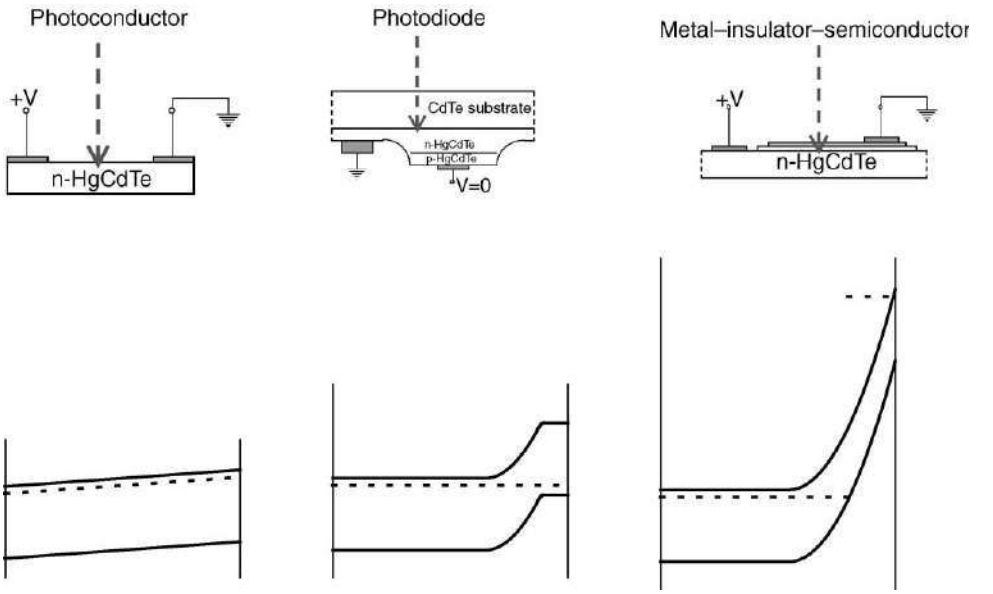

Figure 4 Three different types of infrared detectors have been realized in the $Hg_{1-x}Cd_xTe$ alloy semiconductor: the photoconductive (PC) detector, the photodiode or the photovoltaic (PV) detector, and the metal–insulator–semiconductor (MIS) detector. HgCdTe PC and PV detector arrays are being manufactured for a wide variety of applications. The MIS detector was abandoned in the mid-1980s because it placed unachievable demands on material quality.

The differences among PC, PV, and MIS HgCdTe detectors are due to the way in which the photogenerated electron–hole pairs are manifested electrically in the terminal characteristics of the device. In the PC detector, the excess electron–hole pairs are sensed as a small increase in the conductivity of what is basically a two-terminal resistor. In the PV detector, the excess electron–hole pairs are sensed as a photocurrent in a short-circuited p-n junction photodiode. In the MIS detector, the excess electron–hole pairs are sensed as a small change in the voltage due to minority photocarriers filling a transient charge storage well that has been biased to deep depletion.

PC HgCdTe Detectors

The HgCdTe PC detector is a low-resistance two-terminal device. Resistances are typically 25–100 ohms per square. A dc bias current is applied to convert the conductivity change to an observable voltage change. The device operates under near-equilibrium conditions, with small dc bias voltages typically on the order of 0.1 V. The dc bias electric field within the photoconductor is quite small, generally in the 20–50 $V\,cm^{-1}$ range. The low-resistance device requires a low-noise bipolar preamplifier, usually external to the dewar. Each PC detector element requires one lead through the dewar wall, thus limiting the practical number of PC elements in an array to about 200–300.

The basic HgCdTe photoconductor is a rectangular area, photolithographically defined in an n-type HgCdTe layer approximately 8–15 μm thick, with two ohmic contacts on opposite edges. Surface passivation is straightforward: a thin native oxide strongly accumulates the surface of n-type HgCdTe, thereby reducing the surface recombination velocity to negligibly small values.

Many useful variations on the simple rectangular geometry are possible because of the favorable material properties of HgCdTe. Several contact geometries have been devised to minimize the recombination of photocarriers at the contacts. Serpentine designs have been used to increase detector resistance and reduce bias power dissipation, especially important issues for large-area and very long-wavelength detectors. The SPRITE (signal processing in the element) design reduces the number of electrical connections and dewar leads for scanned arrays by performing both detection and integration within an elongated HgCdTe bar.

PV HgCdTe Detectors

The HgCdTe photodiode is a more complicated device than the photoconductor, requiring both n-type and p-type layers, and having a surface

depletion region that makes stringent demands on surface passivation technology.

The PV detector is a high-resistance p-n junction device. The photodiode is operated very close to equilibrium conditions, either at zero bias voltage or at a small reverse bias voltage whose magnitude is several kT, typically 20–30 mV. The absorber layer is usually n-type HgCdTe about 5–15 μm thick, with the thicker values required for longer-wavelength radiation. A p-n junction is formed by a thin p-type layer. A wide variety of junction formation methods are used. Wide-gap-p on narrow-gap-n heterojunctions are formed by two-layer LPE. Planar p-on-n or n-on-p junctions are formed by arsenic or boron implantation into n-type or p-type layers grown by LPE or MBE. Vertical-geometry p-on-n junctions are formed by a damage mechanism during ion beam milling. Both single-junction and multijunction devices are grown *in situ* by vapor phase epitaxial methods such as MBE and MOVPE. Surface passivation is generally accomplished by CdTe. The high-resistance PV device can match well with low-noise silicon CMOS preamplifiers that can be integrated with two-dimensional arrays to form large back-illuminated hybrid focal plane arrays (FPAs). Typical array formats for HgCdTe FPAs are 256 × 256 and 480 × 640, and can be as large as 1024 × 1024 and 2048 × 2048.

MIS HgCdTe Detectors

The metal–insulator–semiconductor (MIS) detector, also referred to as a photocapacitor, is usually formed on an n-type HgCdTe absorber layer. The insulator of choice is a thin native oxide. The gate electrode is a thin semitransparent metal film.

In contrast to the PC and PV detectors, the MIS detector operates in a strongly nonequilibrium mode.

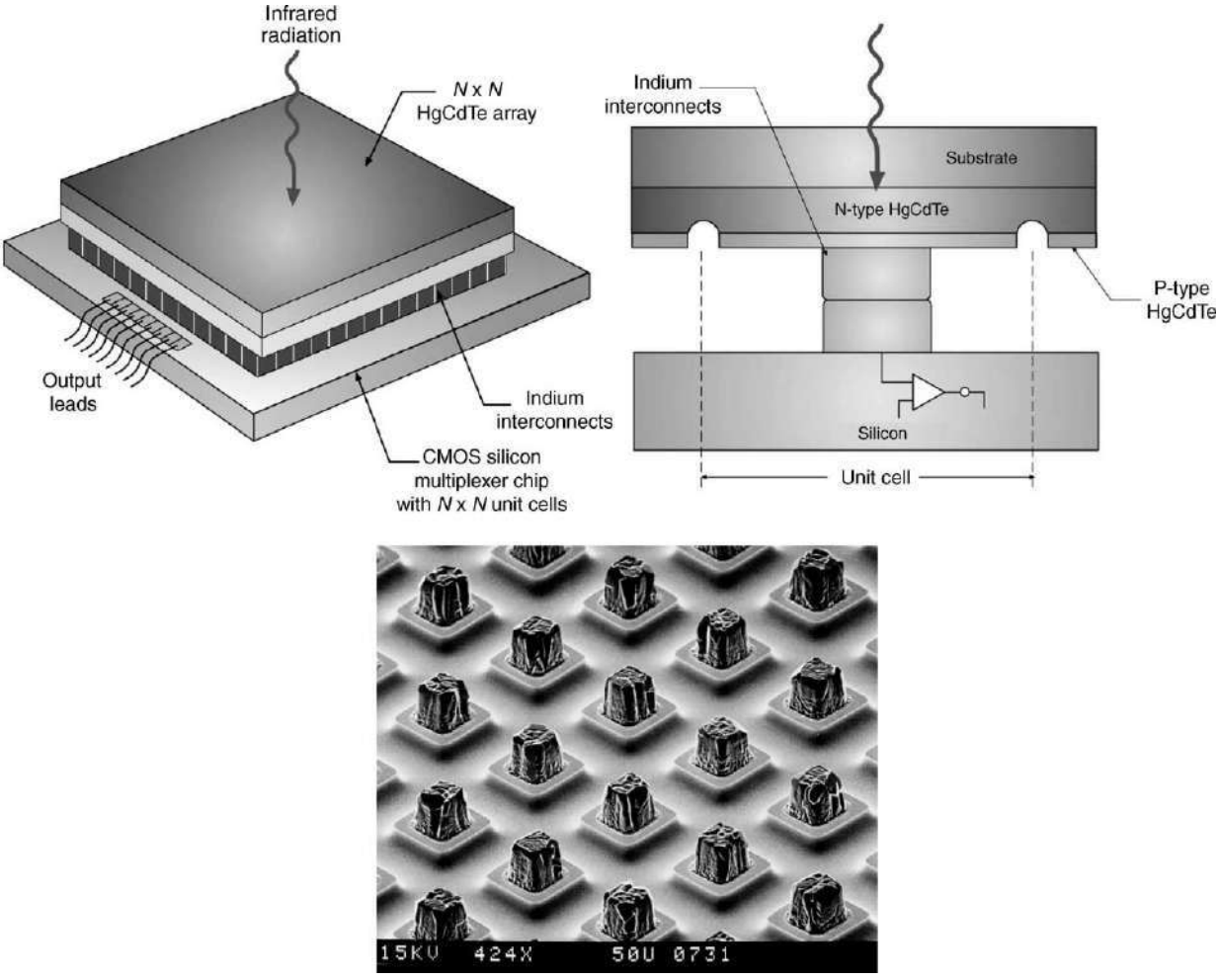

Figure 5a Back-illuminated bump-interconnected HgCdTe FPA architecture. The SEM photo shows an LPE P-on-n HgCdTe mesa photodiode array with 60 × 60 μm^2 unit cells and 33 × 33 μm^2 mesa (junction) areas.

A bias voltage pulse of several volts is applied across the capacitor to drive the HgCdTe surface into deep depletion. Excess electron–hole pairs are photogenerated in the absorber layer and diffuse to the edge of the depletion region, where the holes are collected in the charge well. After a period of time, called the integration time, the voltage across the capacitor is sensed, giving a measure of how much charge was collected, and the cycle is repeated.

Because of the nonequilibrium operation of the MIS detector, much larger electric fields are set up in the depletion region than in the p-n junction, resulting in defect-related tunneling dark current that is orders of magnitude larger than the fundamental dark current. The MIS detector requires much higher material quality than PV or PC detectors, which still has not been achieved. For this reason, all development of the HgCdTe MIS detector was abandoned in the mid-1980s.

Two-Dimensional HgCdTe Infrared Focal Plane Arrays

Among the most important applications of HgCdTe are large two-dimensional electronically scanned hybrid arrays, referred to as focal plane arrays (FPAs). A hybrid HgCdTe FPA consists of a two-dimensional HgCdTe photovoltaic detector array that is interfaced electrically, thermally and mechanically with a matching two-dimensional array of input circuits in a silicon CMOS ROIC chip. Each HgCdTe detector element in the array has one electrical connection to its matching input circuit node in the silicon ROIC.

The thermal expansion coefficient of HgCdTe is sufficiently close to that of silicon to allow several technologically viable hybrid arrangements of HgCdTe detector arrays and silicon multiplexer chips.

There are two main types of hybrid HgCdTe FPAs being developed and manufactured today: the back-illuminated bump-interconnected configuration, shown in Figure 5a, and the front-illuminated or 'loophole' configuration shown in Figure 5b. Each configuration successfully solves the problem of the thermal expansion mismatch between silicon and HgCdTe in different ways, and each requires fundamentally different HgCdTe photodiode designs and processing.

In the back-illuminated configuration, illustrated in Figure 5a, the HgCdTe detector array is bump-mounted onto the silicon ROIC by cold-welded or thermally reflowed indium interconnects. Incident radiation reaches the detectors through an infrared-transparent substrate such as CdTe, CdZnTe, sapphire, or silicon.

In the front-illuminated (or 'loophole') FPA architecture, shown in Figure 5b, the thin HgCdTe layer is epoxied to the silicon ROIC chip, and undergoes elastic deformation when the FPA is cooled. This allows large arrays to be made without need for engineering the thermal expansion of the silicon ROIC chip. Low-temperature processing techniques

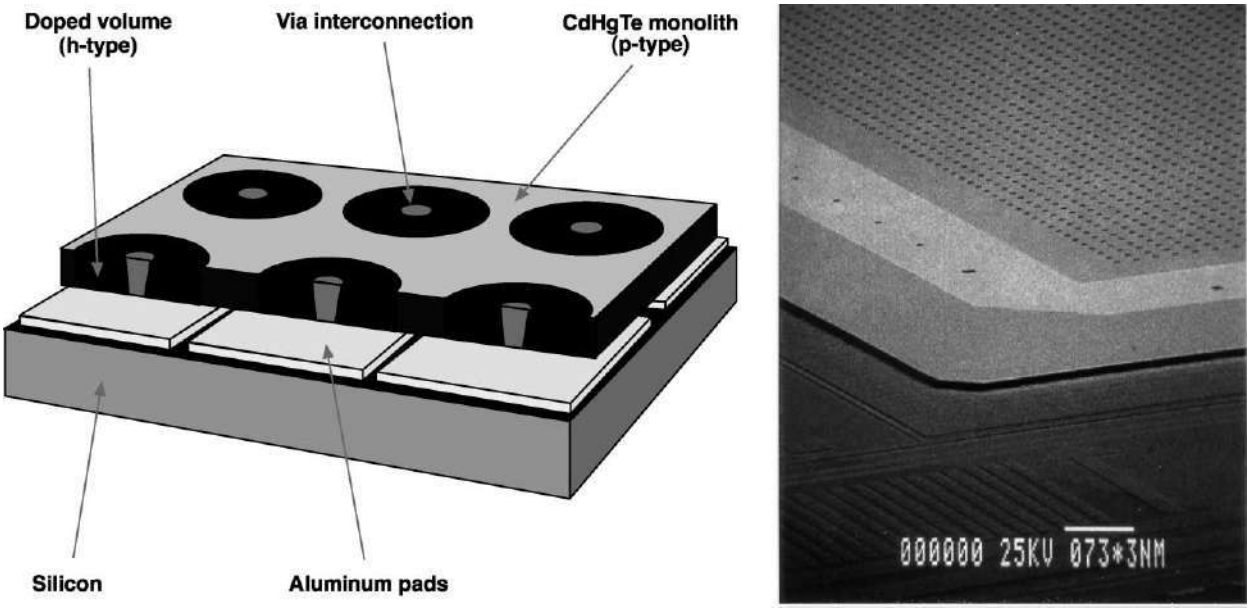

Figure 5b Front-illuminated loophole HgCdTe FPA architecture. The SEM photo shows one corner of a FPA. The metalized via holes are on 20 μm centers.

for junction formation and passivation are required because of the presence of the epoxy.

Conclusions and Trends

There has been enormous progress in HgCdTe material science and device technology over the past 45 years. This progress has been fueled by the many military and space applications for which HgCdTe PC and PV infrared detectors provide nearly ideal solutions. It is important to recognize that the progress so far in HgCdTe detector technology has been possible only through advances in basic materials growth and processing technology.

Further progress can be expected. The fundamental performance limits of HgCdTe photodiodes have not yet been reached. Continued reduction of material defects, both grown-in and process-induced, will reduce thermal generation rates, increase junction impedance, and allow higher operating temperatures for a number of important applications. Continued development of the *in situ* VPE growth methods – MBE and MOVPE – will allow bandgap-engineered heterojunction devices of increasing quality and complexity. Continued development of VPE growth on alternative substrates such as silicon will reduce the cost of two-dimensional arrays, and will make larger arrays practical.

List of Units and Nomenclature

BLIP	Background limited infrared photodetector
CdTe	Cadmium telluride
CdZnTe	Cadmium zinc telluride
CMOS	Complementary metal-oxide-semiconductor
FPA	Focal plane array
HgCdTe	Mercury cadmium telluride
$Hg_{1-x}Cd_xTe$	Mercury cadmium telluride
IR	Infrared
LPE	Liquid phase epitaxy
LW	Long wavelength
LWIR	Long wavelength infrared
MBE	Molecular beam epitaxy
MCT	Mercury cadmium telluride
MIS	Metal-insulator-semiconductor
MOCVD	Metal-organic chemical vapor epitaxy
MOVPE	Metal-organic vapor phase epitaxy
MW	Medium wavelength
MWIR	Medium wavelength infrared
PC	Photoconductive
PV	Photovoltaic
QWIP	Quantum well infrared photodetector
ROIC	Readout integrated circuit
SPRITE	Signal processing in the element
SW	Short wavelength
SWIR	Short wavelength infrared
VLW	Very long wavelength
VLWIR	Very long wavelength infrared
VPE	Vapor phase epitaxy

See also

Semiconductor Physics: Band Structure and Optical Properties; Impurities and Defects; Outline of Basic Electronic Properties.

Further Reading

Capper P (ed.) (1994) *Properties of Narrow Gap Cadmium-based Compounds*. EMIS Datareviews Series No. 10. London: INSPEC, Institution of Electrical Engineers. Reviews and compilations of fundamental and technological properties of HgCdTe and CdTe/CdZnTe/CdTeSe.

Capper P (ed.) (1997) *Narrow-Gap II–VI Compounds for Optoelectronic and Electromagnetic Applications*. London: Chapman and Hall. Twenty review chapters on growth techniques, materials characterization, and device applications of HgCdTe.

Capper P and Elliott CT (eds) (2000) *Infrared Detectors and Emitters: Materials and Devices*. Boston, MA: Kluwer Academic Publishers. Review chapters on HgCdTe materials, photoconductive and nonequilibrium HgCdTe detectors, photovoltaic HgCdTe detectors, and HgTe-CdTe superlattice detectors.

Rogalski A, Adamiec K and Rutkowski J (2000) *Narrow Gap Semiconductor Photodiodes*. Washington, DC: SPIE Press.

Willardson RK and Beer AC (eds) (1981) *Defects, (Hg,Cd)Se, (Hg,Cd)Te*. Semiconductors and Semimetals, vol. 16. New York: Academic Press. Review chapters on magneto-optical properties of HgCdTe and nonlinear optical effects in HgCdTe.

Willardson RK and Beer AC (eds) (1981) *Mercury Cadmium Telluride*. Semiconductors and Semimetals, vol. 18. New York: Academic Press. Seven review chapters: history of HgCdTe, preparation of high-purity Hg, Cd and Te, crystal growth of HgCdTe, Auger recombination in HgCdTe, and photoconductive, photovoltaic and metal-insulator-semiconductor HgCdTe detectors.

Modulation Spectroscopy of Semiconductors and Semiconductor Microstructures

Y-S Huang, National Taiwan University of Science and Engineering, Taipei, Taiwan

F H Pollak, Brooklyn College of the City University of New York, Brooklyn, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Modulation spectroscopy is a versatile and powerful optical technique for obtaining valuable information relevant to a large variety of semiconductor systems including bulk/thin films, surfaces/interfaces [semiconductor/air (vacuum), semiconductor/electrolyte, semiconductor/metal, semiconductor/semiconductor (homo- and heterojunctions)], micro- and nanostructures (quantum wells, multiple quantum wells, superlattices, quantum wires, quantum dots), the effects of growth/processing, as well as the characterization of actual device structures [heterojunction bipolar transistors (HBTs), pseudomorphic high electron mobility transistors (PHEMTs), edge-emitting and vertical-cavity surface-emitting (VCSEL) quantum well lasers, etc.].

Modulation spectroscopy deals with the measurement and interpretation of changes in the optical response of a sample which are effected by the modification of the measurement conditions. This can easily be accomplished by periodically modulating either (1) the measurement conditions themselves ('internal' modulation) or (2) some parameters applied to the sample ('external' modulation). The observed normalized changes are usually small so that the difference signals are closely related to a derivative of the absolute spectrum with respect to the modifying parameters. The derivative nature of modulation spectroscopy emphasizes structure localized in the photon energy region of the interband (intersub-band) transitions of semiconductors (semiconductor microstructures) and suppresses uninteresting background effects. As a result, weak features that may not be detected in the absolute spectra are often enhanced. Because of this derivative-like nature a large number of sharp spectral features may be observed, even at room temperature.

The ability to perform a lineshape fit is one of the great advantages of modulation spectroscopy. Since for the modulated signal the features are localized in photon energy it is possible to account for the lineshapes to yield accurate values of energies and broadening parameters of the associated interband transitions. For example, the energies of various interband (intersub-band) transitions may be evaluated to within a few meV, even at room temperature. Thus, the effects of static external perturbations such as electric and magnetic fields, temperature, hydrostatic pressure, uniaxial stress, etc., can be conveniently studied.

A particularly useful form of modulation spectroscopy is electromodulation (EM) since it is sensitive to surface/interface electric fields and can be performed in contactless modes that require no special mounting of the sample. Under appropriate conditions the EM spectrum may exhibit above bandgap oscillatory features, called Franz–Keldysh oscillations (FKOs), which are a direct measure of the relevant electric field (built-in and/or applied).

Two of the most widely used forms of EM are photoreflectance (PR) and contactless electroreflectance (CER). In PR, modulation of the built-in electric field is caused by photo-excited electron–hole pairs created by a pump source of ~3–5 mW power of either (1) an internally modulated laser diode or (2) a mechanically chopped dc laser. The modulating frequency is typically ~100–200 Hz. CER utilizes a condenser-like system consisting of a front wire grid electrode with a second metal electrode separated from the first electrode by insulating spacers, which are ~0.1 mm larger than the sample thickness. The sample is placed between these two capacitor plates. Thus, there is nothing in direct contact with the front surface of the sample. The probe beam is incident through the front wire grid. EM is achieved by apply an ac voltage (~1 kV peak to peak at ~200 Hz) across the electrodes.

Instrumentation

Shown in Figure 1 is a schematic drawing of the experimental arrangement for a general 'external' modulated reflectance experiment (except for PR). Light from an appropriate light source (xenon arc or tungsten halogen lamp) passes through a monochromator. The exit intensity at wavelength λ, $I_o(\lambda)$, is focused onto the sample by means of a lens (or mirror). The modulation (electric field, stress, temperature) is applied to the sample at frequency Ω_m. The reflected light is collected by a second lens (mirror) and is focused onto an appropriate detector

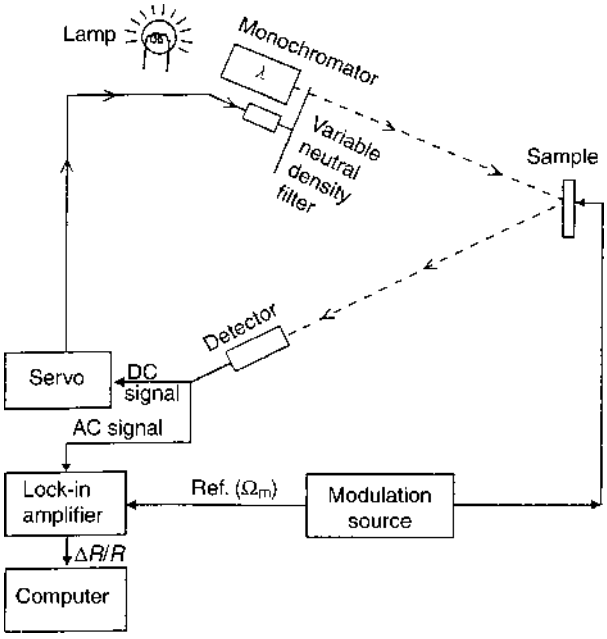

Figure 1 Schematic representation of external modulation experimental apparatus.

(photomultiplier or photodiode). For the sake of simplicity the two lenses (mirrors) are not shown.

The light striking the detector contains two signals: the dc is given by $I_o(\lambda)R(\lambda)$, where $R(\lambda)$ is the dc reflectance of the material while the modulated value (at frequency Ω_m) is $I_o(\lambda)\Delta R(\lambda)$, where $\Delta R(\lambda)$ is the change in reflectance produced by the modulation source. The ac signal from the detector, proportional to $I_o(\lambda)\Delta R(\lambda)$, is measured by a lock-in amplifier. Typically $I_o\Delta R$ is 10^{-4}–10^{-6} of I_oR. In order to evaluate the quantity of interest, i.e., the relative change in reflectance $\Delta R/R$, a normalization procedure must be used to eliminate the uninteresting common feature $I_o(\lambda)$. This can be done in several ways, one of which is shown in Figure 1, where the normalization is performed by a variable neutral density filter (VNDF) connected to a servo mechanism. The dc signal from the detector, which is proportional to $I_o(\lambda)R(\lambda)$, is fed into the servo which moves the VNDF in such a manner as to maintain a constant $I_o(\lambda)R(\lambda)$, i.e., $I_o(\lambda)R(\lambda) = C$. Under these conditions the ac signal $I_o(\lambda)\Delta R(\lambda) = C\Delta R(\lambda)/R(\lambda)$. Subsequently, the signal to the lock-in amplifier is proportional to the quantity of interest, i.e., $\Delta R(\lambda)/R(\lambda)$.

Bulk/Thin-Film Material

Modulation spectroscopy has been used to study the properties of bulk/epitaxial layered semiconductors such as ordering in $GaInP_2$, piezoelectric fields in strained GaN, the influence of isotopes, the properties of low-temperature grown GaAs, the space charge region, carrier concentration, the nature of the band bending (carrier type), impurity effects, deep levels, alloy composition, the influence of static external or internal perturbations such as temperature, external uniaxial stress, lattice-mismatch strain, hydrostatic pressure, the effects of crystal growth/processing/annealing procedures, and amorphous and microcrystalline materials.

The dashed lines in Figure 2 are the piezoreflectance (PZR) spectra in the region of the direct gap of ReS_2 in the temperature range $25\,K < T < 450\,K$. The data at low temperatures exhibit two well-resolved features labeled E_1^{ex} and E_2^{ex}. The solid lines are lineshape fits to the first derivative of a Lorentzian profile which yields the energies indicated by arrows and also the related broadening parameters. The solid/open squares and circles in Figure 3 are the temperature dependence of E_1^{ex}/E_2^{ex} of ReS_2 and $ReSe_2$, respectively. The solid, dot-dashed, and dashed lines in the figure are fits to the Varshni, O'Donnel–Chen, and Bose–Einstein expressions, respectively. The Varshni equation is:

$$E_i^{ex}(T) = \frac{E_i^{ex}(0) - \alpha_i T^2}{\beta_i + T} \qquad [1]$$

where $i = 1$ or 2, $E_i^{ex}(0)$ is the excitonic transition energy at 0 K and α, β are the Varshni coefficients. The O'Donnel–Chen expression is:

$$E_i^{ex}(T) = E_{i0} - S_i\langle\hbar\Omega_i\rangle\left(\frac{\coth\langle\hbar\Omega_i\rangle}{2kT} - 1\right) \qquad [2]$$

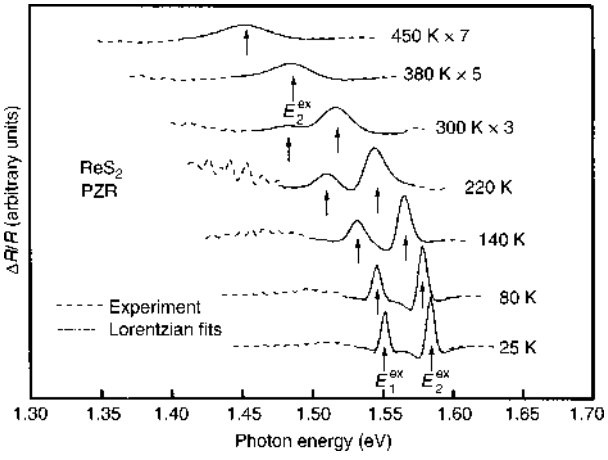

Figure 2 The PZR spectra (dashed lines) in the region of the direct gap of ReS2 in the temperature range $25\,K < T < 450\,K$. The solid lines are lineshape fits yielding the energies indicated by the arrows.

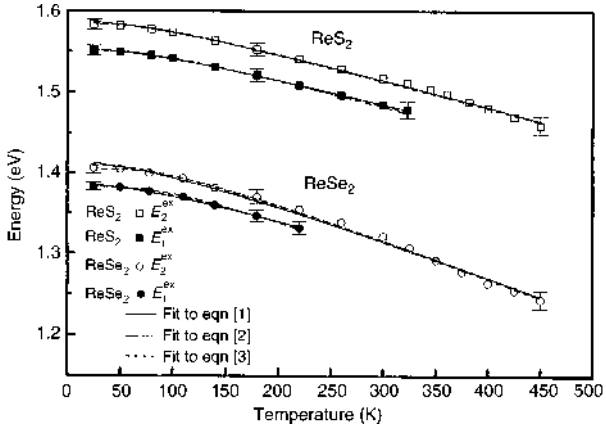

Figure 3 The temperature dependence of the excitonic transition energies of ReS_2 and $ReSe_2$. The solid, dot-dashed, and dashed lines in the figure are fits to the Varshni, O'Donnel–Chen, and Bose–Einstein expressions, respectively.

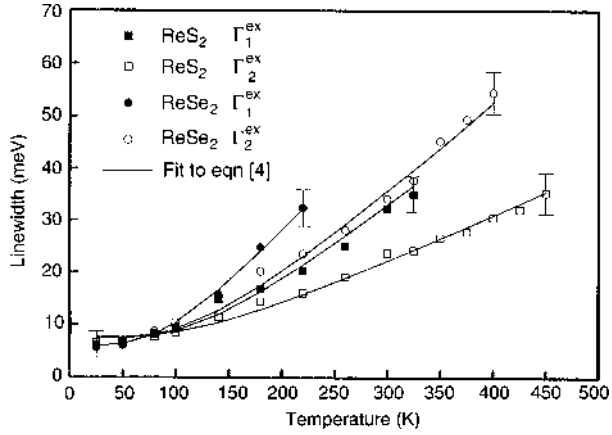

Figure 4 The solid/open squares and circles are the temperature dependence of the broadening parameters of E_1^{ex}/E_2^{ex} of ReS_2 and $ReSe_2$, respectively. The solid lines are fits to the Bose–Einstein-type expression.

where E_{i0} is the excitonic transition energy at 0 K, S_i is a dimensionless coupling constant related to the strength of the electron–phonon interaction, and $\langle \hbar\Omega_i \rangle$ is the average phonon energy. The Bose–Einstein-type expression can be written as

$$E_i^{ex}(T) = E_{iB} - a_{iB}\left[1 + \frac{2}{\exp(E_{ap}/kT) - 1}\right] \quad [3]$$

where a_{iB} represents the strength of the electron (exciton)–phonon interaction, and E_{ap} corresponds to the average phonon energy.

The temperature dependence of the broadening parameters, $\Gamma(T)$, is due only to the electron–phonon interaction and can also be described by a Bose–Einstein-type equation. For the lowest lying direct gap only the longitudinal optic (LO) phonon is involved. In Figure 4 the solid/open squares and circles are the temperature dependence of the broadening parameters of E_1^{ex}/E_2^{ex} of ReS_2 and $ReSe_2$, respectively. The solid lines are fits to the Bose–Einstein-type expression:

$$\Gamma_i(T) = \frac{\Gamma_i(0) + \Gamma_{i,ep}}{\exp(E_{i,LO}/kT) - 1} \quad [4]$$

where $i = 1$ or 2. The first term of eqn [4] corresponds to broadening mechanisms due to intrinsic lifetime, electron–electron interaction, impurity, dislocation, and alloy scattering effects. The parameter Γ_{ep} is an electron (exciton)–phonon coupling coefficient and E_{LO} is the zone-center LO phonon energy. The obtained values of Γ_{ep} for these materials have been compared with those for the direct gap in GaAs and ZnSe.

Surfaces/Interfaces

Almost since the inception of modulation spectroscopy it has been recognized that EM could be used as an effective probe of surfaces and interfaces. For EM this is true not only because of the sharp, derivative-like spectral features but also due to its sensitivity to electric fields, i.e., FKOs. In addition, PR and CER are of considerable interest and usefulness since they are contactless, employ low light levels, require no special mounting of the sample, and can be performed in any transparent ambient including ultrahigh vacuum (UHV).

Of particular usefulness in the study of Fermi level pinning at surfaces are UN(SIN)/UP(SIP) and δ-doping configurations. The UN(SIN)/UP(SIP) structures are made by fabricating an undoped layer of thickness L (~1000 Å) on a buried doped $n^+(p^+)$ buffer on a doped $n^+(p^+)$ substrate. In the $n^+(p^+)$ buffer/substrate the Fermi level occurs near the conduction (valence) band edge. At the surface the Fermi level is pinned at some value, V_F. Therefore, there exists in the undoped region, which has a small broadening parameter, a large, almost constant electric field, F. A similar constant field can be created by placing a δ-doping layer a distance L from the surface.

The relation between F, as observed from the FKOs, and V_F is given by:

$$FL = V_F - (kT/q) - \mathrm{SCC} - V_p(T) \quad [5]$$

where the second, third, and fourth terms on the right-hand side are the Debye length, space charge layer, and temperature dependent photovoltaic effect corrections, respectively.

The properties of the metal–semiconductor interface have been reported by a number of investigators using PR and/or ER. Jonker *et al.* have demonstrated that an epitaxial Fe film on GaAs (UN structure) provides a ferromagnetic contact, suppresses midgap state formation, and does not pin the Fermi level. Four samples were investigated: a native oxide reference sample, the same sample after sulfur passivation, Al/GaAs-(2 × 4), and Fe/GaAs-(2 × 4). The width of the undoped GaAs layer was 1070 Å for the first three samples and 1500 Å for the last sample. These results are shown in Figure 5a–d, respectively. The fields obtained from the FKOs are 58 kV/cm, 61 kV/cm, 55 kV/cm and 60 kV/cm. For the first three samples the measured barrier height is about 0.64 eV (not corrected for the photovoltaic effect) while for the Fe/GaAs material it is about 40% higher (~0.9 eV). Another significant feature of the Fe/GaAs spectrum is the large number of FKOs (in relation to the other samples) extending to about 2 eV. The significant reduction in damping indicates a substantial increase in carrier lifetimes.

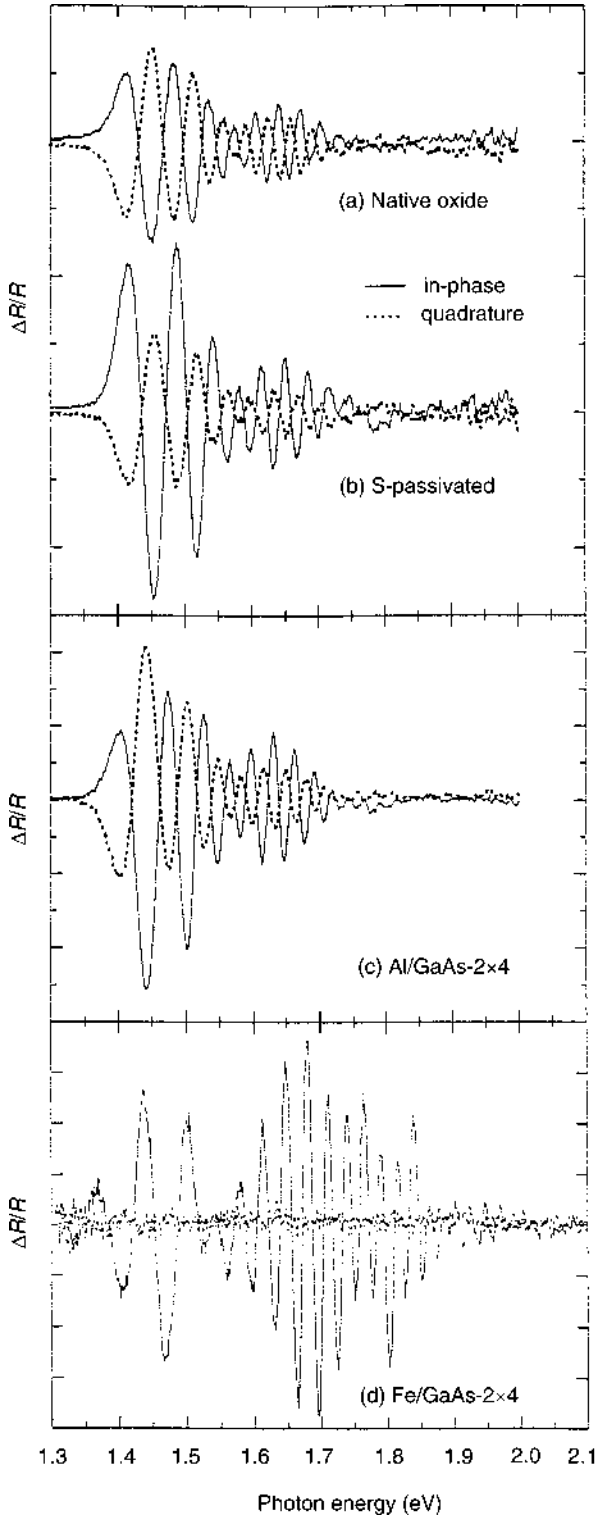

Figure 5 The PR spectra at 300 K for (a) a native oxide reference sample, $F = 58$ kV/cm; (b) the same sample following sulfur passivation, $F = 61$ kV/cm; (c) Al/GaAs-(4 × 2) sample, $F = 55$ kV/cm, and (d) Fe/GaAs-(2 × 4) sample, $F = 60$ kV/cm. The thickness of the undoped spacer region is 1070 Å for (a)–(c) and 1500 Å for (d).

Micro- and Nanostructures

Modulation spectroscopy is a very powerful tool for investigating many of the fundamental aspects of compositional single quantum wells (SQWs)/multiple quantum wells (MQWs), superlattices (SLs), quantum wires, and quantum dots (QDs). Various phenomena can be studied, including band offset, well and barrier widths, excitons, strain, coupling and decoupling between wells, miniband formation, two-dimensional electron gas (2DEG) effects, zone-folding in short-period SLs, built-in electric fields, etc. Work on a large variety of GeSi, III–V and II–VI systems has been reported, including both lattice-matched and strain-layer configurations. In addition the effects of various external perturbations can be evaluated.

The properties of a vertically coupled InAs/GaAs QD-based laser structure fabricated by a self-assembled technique have been investigated by CER (300 K and 20 K). Signals have been observed from all the relevant regions of the sample. The dotted lines are the CER spectra at 300 K and 20 K in Figure 6a,b, respectively. The solid lines are lineshape fits which yield the energies indicated by arrows. The low-energy features, labeled QD_0 (lateral coupling), QD_1 (vertical coupling) and QD_2 (vertical coupling), originate from the QDs while the resonances designated W_1 and W_2, correspond to the first conduction to first heavy- (1C-1H) and light-hole (1C-1L) transitions, respectively, in the QW (approximately one monolayer) formed by the

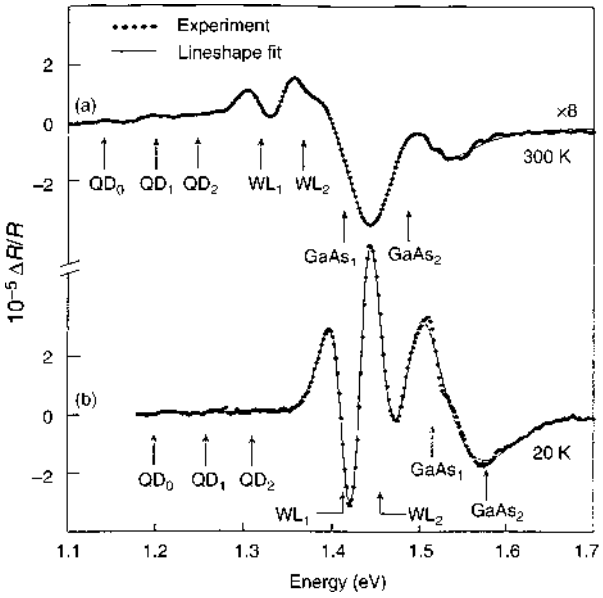

Figure 6 CER spectra (dotted lines) of an InAs/GaAs quantum dot-based laser structure at (a) 300 K and (b) 20 K. The solid lines are a lineshape fit to a first derivative Gaussian.

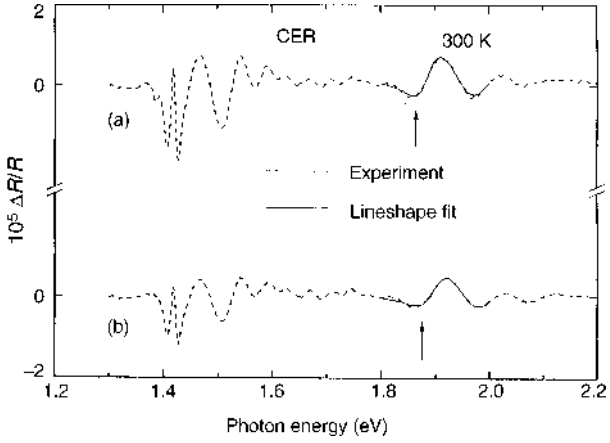

Figure 7 Room-temperature CER spectra (dashed lines) for a GaInP/GaAs HBT sample fabricated by metalorganic chemical vapor deposition for incident light polarized (a) $\boldsymbol{E}\|[110]$ and (b) $\boldsymbol{E}\|[1\bar{1}0]$. The solid lines are a lineshape fit which yields the bandgap energies denoted by arrows.

wetting layer (WL). Although the oscillator strength of the QD features is small compared to the WL resonances, they are clearly visible even at 300 K. The $GaAs_1$ and $GaAs_2$ features are attributed to GaAs-related transitions. The former at both 300 K and 20 K corresponds to the bandgap energy of GaAs and probably comes from the undoped 400 Å-thick GaAs layers which envelop the QD region. The $GaAs_2$ feature is most likely related to the p-doped 6000 Å-thick GaAs cap layer. The heavy doping of this layer induces a Burstein–Moss blue shift of the bandgap.

Device Structures

The methods of PR and CER are valuable tools in the evaluation of important device parameters for structures such as heterojunction bipolar transistors (HBTs), pseudomorphic high electron mobility transistors (PHEMTs), edge-emitting and vertical cavity surface-emitting (VCSEL) quantum well lasers, multiple quantum well infrared detectors, solar cells, superlattice optical mirrors, resonant tunneling structures, metal-oxide-semiconductor (MOS) configurations, among others.

The dashed lines in Figure 7a,b are the 300 K CER spectra of a $GaInP_2$(emitter)/GaAs(collector) HBT (fabricated by metalorganic chemical vapor deposition) for incident light with electric field vector $\boldsymbol{E}\|[110]$ (polarization A) and $\boldsymbol{E}\|[1\bar{1}0]$ (polarization B), respectively. Both collector and emitter signals exhibit well-pronounced FKOs. The solid lines are least-squares fits to a function which contains Lorentzian broadened electro-optic functions, yielding 1.873 ± 0.001 eV and 1.885 ± 0.001 eV for the $GaInP_2$ bandgaps for polarizations A and B, respectively, as indicated by the arrows. The ordering parameter of 0.3 deduced from the differences in the measured bandgaps for the two polarizations was consistent with transmission electron microscope measurements. The electric fields evaluated from the emitter and collector region FKOs were compared with a computer simulation of the field profiles using a comprehensive, self-consistent model, including the photovoltaic effect. The deduced emitter and collector doping densities were in good agreement with the intended growth conditions.

Summary

Modulation spectroscopy is a very powerful tool for investigating many of the fundamental and applied aspects of a wide variety of semiconductors and semiconductor micro- and nanostructures, including bulk/thin films, surfaces/interfaces, single quantum wells, multiple quantum wells, superlattices, quantum wires, quantum dots, etc. In addition the effects of various perturbations such as alloy composition, temperature, electric and magnetic fields, hydrostatic pressure, internal and/or external stain, processing, etc., can be evaluated. Recently these methods have also been applied for the characterization of actual device structures such as HBTs, PHEMTs, quantum well lasers, etc.

See also

Semiconductor Materials: Quantum Dots. **Semiconductor Physics:** Band Structure and Optical Properties; Impurities and Defects; Outline of Basic Electronic Properties; Quantum Wells and GaAs-based Structures; Surface Photovoltage Spectroscopy of Semiconductors.

Further Reading

Aspnes DE (1980) Modulation spectroscopy/electric field effects on the dielectric properties of semiconductors. In: Balkanski M (ed.) *Handbook on Semiconductors*, vol. 2, pp. 109–154. Amsterdam, The Netherlands: North-Holland.

Glembocki OJ and Shanabrook BV (1992) Photoreflectance spectroscopy of microstructures. In: Seiler DG and Littler CL (eds) *Semiconductors and Semimetals*, vol. 36, pp. 222–289. New York: Academic Press.

Pollak FH (1994) Modulation spectroscopy of semiconductors and semiconductor microstructures. In: Balkanski M (ed.) *Handbook on Semiconductors*, vol. 2, pp. 527–635. Amsterdam, The Netherlands: North Holland.

Pollak FH (1995) Electromodulation spectroscopy of semiconductor surfaces and interfaces. In: Halevi P (ed.) *Photonic Probes of Surfaces*, pp. 175–209. Amsterdam, The Netherlands: North-Holland.

Pollak FH (1996) Modulation spectroscopy of quantum wells and superlattices. In: Bhattacharya P (ed.) *Properties of III–V Quantum Wells and Superlattices*: *EMIS Data Reviews Series #15*, pp. 232–240. London: INSPEC.

Pollak FH (1998) Modulation spectroscopy of the group III-nitrides. In: Gil B (ed.) *Group III Nitride Semiconductor Compounds*, pp. 158–181. Oxford, UK: Clarendon.

Pollak FH (2001) Modulation spectroscopy characterization of semiconductors and semiconductor microstructures. In: Buschow KH, Cahn W, Flemmings MC, *et al.* (eds) *Encyclopedia of Materials: Science and Technology*, vol. 6, pp. 5743–5748. Amsterdam, The Netherlands: Elsevier.

Ramdas AK and Rodriguez S (1992) Piezospectroscopy of semiconductors. In: Seiler DG and Littler CL (eds) *Semiconductors and Semimetals*, vol. 36, pp. 137–220. New York: Academic Press.

Quantum Dots

R J Warburton, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Quantum dots are nanometer-sized clusters of semiconductor material. An island of low-bandgap material is embedded in a sea of high-bandgap material so that both electrons and holes are spatially confined in all three directions. In an unstructured semiconductor, the electron energy levels of the individual atoms hybridize to form energy bands. In a quantum dot, however, there are discrete energy levels. This can be understood from the band structure of the unstructured material where there is a so-called dispersion relation connecting the energy, E, and the momentum, k. The confinement potential of a quantum dot quantizes the k-values, giving rise to discrete energy levels. If the energy levels of a quantum dot are separated by more than the typical energy level broadenings and by more than the thermal energy then the quantum dot will behave quite differently to the bulk material. The quantum dot will exhibit atom-like properties – discrete energy levels with small degeneracies – despite the fact that it is embedded in a semiconductor matrix. It is this ability to prepare a variety of atom-like systems in a semiconductor environment which has driven the research into semiconductor quantum dots.

In some sense, nanosized quantum dots have been the subject of research for quite some time. Color filters, for instance, consist of nanosized particles dispersed in a glass matrix, and this is a very mature technology. Also, lithography of layered semiconductor systems already possessing confinement in one dimension has been pursued for at least 15 years. However, developments in the past few years have been rapid, all based on the discovery of self-assembly of quantum dots directly in the growth of semiconductor heterostructures. These quantum dots can be produced with standard materials in standard growth systems, and this allows an easy integration of quantum dots into devices such as diode lasers. Furthermore, the self-assembled quantum dots are of extremely high quality and this is of great value not just for devices but also for fundamental physics experiments, aimed at studying or exploiting the coherence of excitons confined in the quantum dots. For these reasons, this short review concentrates on the properties of self-assembled quantum dots, giving just brief descriptions of other systems. It should be noted that the bibliography is extremely limited and

in no way does justice to the large number of research groups which have made valuable contributions to this field. Instead, the bibliography is simply a list of suggestions for finding out more about this fascinating research area.

Quantization in Quantum Dots

Quantum dots are on the one hand large enough that the bulk band structure of the host material can be used to describe the material's electronic properties. In the effective mass approximation, this amounts to giving both electrons and holes a quadratic dispersion on momentum, just as for a free electron, but with an effective mass, typically about 0.1 for the electrons, and 0.3 for the holes. On the other hand, quantum dots are small enough that confinement effects are important. For instance, the energy of an exciton in a nanometer-sized quantum dot is a strong function of the quantum dot size. As the dots become smaller, confinement effects become larger, and the energy separating electron and hole levels increases. This does not increase indefinitely, however. There is a finite-size energy barrier to the matrix surrounding the dot such that in the limit of small dot sizes, the wavefunction spreads out into the barrier region and confinement effects become very weak. Ideally, an electron in a quantum dot has a confinement energy of at least several meV, and an energy barrier to the surrounding matrix of several 100 meV so that confinement effects are visible even at room temperature.

The motivation for the study of quantum dots is to combine the atom-like properties with the flexibility of a solid-state environment. This is perfectly illustrated by considering a quantum dot as the gain medium for a laser. In a conventional semiconductor diode, a band of levels has to be inverted in order to achieve population inversion. This is costly in terms of the threshold current and the temperature stability. In a quantum dot, inversion can be achieved simply by inverting the population of the two-fold degenerate ground state, which is clearly trivial. Furthermore, a strong confinement energy and a large barrier to the surrounding matrix imply that this population inversion is stable against rises in temperature, implying better room-temperature performance. This example also exemplifies the motivation for investigating quantum dots using well-developed inorganic semiconductor technology as in this case the dots can be integrated into established device geometries. Generally speaking, the important concept in quantum dot physics is spatial confinement of electrons in all three directions, giving rise to discrete energy levels.

Lithography for Quantum Dots

Developments in semiconductor growth technology, largely during the 1970s and 1980s, enabled thin layers of different semiconductors to be grown on top of each other. There are two principal growth techniques, molecular beam epitaxy (MBE), in which a heated substrate is bombarded with pure species in a high vacuum, and metal organic chemical vapor deposition (MOCVD), in which chemical precursors flow over a heated substrate and react. In the simplest structure, termed a quantum well, a thin layer of low-bandgap material is sandwiched between high-bandgap material. The discontinuities in the potentials at the interface give rise to confined electronic states for both carrier types, electrons and holes. The quantum well is an example of a two-dimensional system: there is confinement in one direction, the growth direction, but free dispersion in the other two directions. To make a quantum dot, it is necessary to confine the electron and hole in all three directions. An obvious approach is to take a quantum well, with perfect confinement in one direction, and process it laterally with lithographic techniques to complete the quantization. Note that an analogy can be made to the physics of the quantum Hall effect, where a two-dimensional electron gas is subjected to a very high magnetic field. The magnetic field quantizes the lateral motion in a similar way to the electrostatic potential in a quantum dot.

This approach to the formation of optically active quantum dots obviously has some advantages, notably that the shape and position of the quantum dots can be chosen in the lithography. However, there are some significant problems associated with these techniques. As most of these problems do not plague the properties of self-assembled quantum dots, self-assembled quantum dots are more suitable for advanced photonic devices.

A number of different lithographic techniques has been used to process quantum-well heterostructures into quantum dots. For research applications, electron-beam lithography is the method of choice. The sample is coated with a resist, for instance PMMA, and exposed with electrons in an area with nanometer size. A protective layer, either a metal or insulator, is transferred to the exposed area by developing the resist, by depositing the protective layer material, and by removing the unwanted material with a lift-off step. The wafer is then etched, typically in a dry environment, to produce a free-standing column.

The photoluminescence of these quantum dots has been measured by various research groups, and although the results vary quite widely there are some systematic trends. The photoluminescence

efficiency decreases rapidly for pillars smaller than a few microns. This is somewhat improved at low temperature, where reasonably strong emission can be observed down to 100-nm diameter pillars. For pillars with micron diameters, the lateral confinement effects are weak, and it is debatable if true quantum dots have been achieved. There are a number of effects which degrade the emission. First, pinning of the Fermi energy by surface states generates large electric fields near the surface, which can ionize electron–hole pairs. This is well known for quantum wells: if a GaAs quantum well for example is placed less than about 25 nm from the surface, the photoluminescence is very weak because of the surface electric field. In an etched pillar, these effects are clearly magnified by the new surface area generated. Secondly, the etching process damages the material to a greater or lesser degree, and the damaged region contains centers for nonradiative recombination.

Other lithography techniques have also been used to generate semiconductor quantum dots for optical experiments. An elegant idea is to deposit small regions of a highly strained film onto the sample surface. In this way, a laterally varying strain film in an underlying quantum well can be set up, generating quantum dots through the effect of strain on the band gap. (This idea has also been married with self-assembly, by using Stranski–Krastanow-grown InP stressors on a GaAs surface, generating lateral confinement in an underlying InGaAs quantum well.) A further idea is to generate a quantum dot by deliberately growing a quantum well with rough interfaces. In such a system, there will inevitably be regions which confine an electron–hole pair through localization. Both of these approaches clearly circumvent the problems generated by the exposed surface in an etched quantum dot. However, they generate small lateral confinement potentials, typically a few meV. This is of value in a research environment where fundamental experiments can be carried out at low temperature, but in a possible device at room temperature, where the thermal energy is 25 meV, quantization effects will not be discernible.

Self-Assembled Quantum Dots

A solution to the difficulties inherent in quantum dots made by lithographic techniques began to emerge in the early 1990s. It was discovered that quantum dots can be produced directly in the semiconductor growth by depositing a material onto a substrate with a substantially smaller lattice constant. The most well-known example is the growth of InAs (lattice constant 6.05 Å) on GaAs (lattice constant 5.65 Å) where In-rich quantum dots form. Because billions of dots form spontaneously in this process, the dots are referred to as self-assembled quantum dots. Since the early experiments, considerable progress has been made, such that the state-of-the-art quantum dots possess many of the hoped-for properties: they are coherent (meaning that they contain no dislocations), highly homogeneous (meaning that fluctuations in shape and composition from dot to dot are small), and have large quantum efficiencies for the emission of light.

Growth of Self-Assembled Quantum Dots

When growing one material on top of another, three principal growth mechanisms have been identified: Frank–van der Merwe, Volmer–Weber, and Stranski–Krastanov. In the Frank–van der Merwe mode, material is deposited layer by layer, as in the growth of the lattice-matched pair GaAs and AlAs. In the Volmer–Weber mode, island formation occurs: the material does not wet the surface because it is energetically unfavorable. Another possibility arises if there is a lattice mismatch. In this case, the epilayer is strained, and the strain energy obviously increases with increasing epilayer thickness. In the Stranski–Krastanow mode, the initial growth is layer by layer, as in Frank–van der Merwe, but beyond a certain thickness, islands form, as shown schematically in Figure 1. On the one hand, the islands limit the strain energy, because some strain relaxation is allowed, which is not possible in a thin film. On the other hand, island formation results in an increase of the

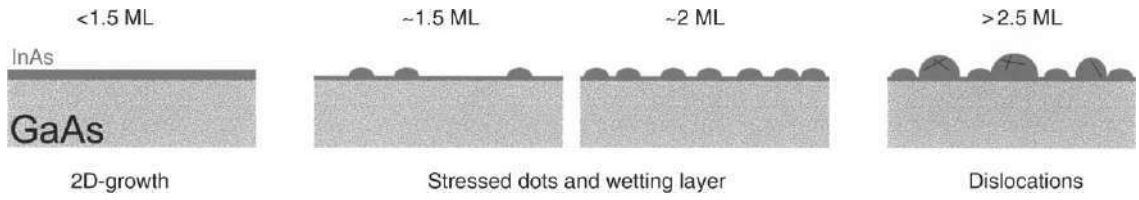

Figure 1 A schematic of the Stranski–Krastanow mode. The material deposited has a substantially larger lattice constant than the substrate, for example InAs on GaAs. Initially, a thin and uniform layer is deposited. At the critical thickness (1.5 monolayers for InAs on GaAs), islands spontaneously form. Further growth results in an increase in the number of islands. At large coverages, large and dislocated islands appear.

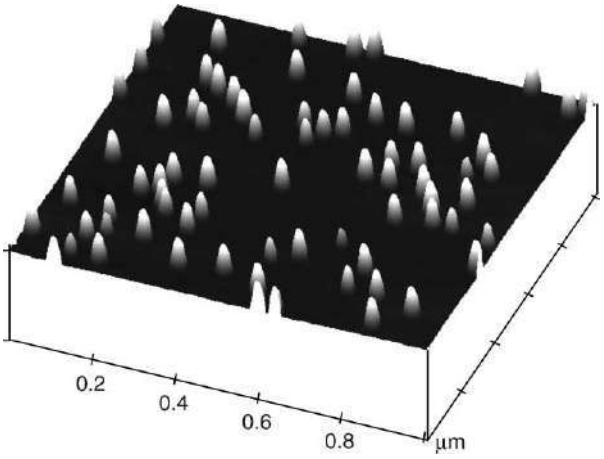

Figure 2 An atomic force micrograph of InAs quantum dots on a GaAs substrate. The InAs dots were grown by MBE at a growth temperature of 530 °C. The image was taken by Axel Lorke.

surface energy. This means that there is a size at which the total energy is minimized.

It is now well established that many combinations of semiconductors grow in the Stranski–Krastanow mode, and this can be utilized for self-assembly of quantum dots. The most studied example is InAs on GaAs where high-quality quantum dots can be grown with both MBE and MOCVD. Dots form at the so-called critical thickness of 1.5 monolayers, corresponding to just 4 Å of material. For thicknesses less than the critical thickness, the covering is not completely uniform; there are island-like structures, elongated along the $[0\bar{1}1]$ direction. At the critical thickness dots form, leaving a thin InAs layer (the wetting layer) between the dots. As more InAs is added, the density of dots increases with only small changes in the dot size. As an example, Figure 2 shows InAs dots on a GaAs substrate.

The dots grown in this Stranski–Krastanow mode are free of dislocations, which is an essential prerequisite in III–V compounds for a high quantum efficiency for optical emission. Furthermore, the dots are remarkably homogeneous. Statistical fluctuations during growth will always give rise to a distribution in dot size and height. Nevertheless, the fluctuations in dot size can be as small as ~10%. The dots are randomly arranged in the lateral plane if the growth proceeds on a flat substrate, and densities are between 10^9 and 10^{11} cm^{-2}. The shape and composition of the dots depend on the growth parameters. For instance, InAs/GaAs dots can have facets along particular crystal directions, or they can be lens-shaped without facets, depending on the growth technique, and in particular the growth temperature. In fact, measuring the shape and composition is a very challenging task, particularly for buried dots.

Stranski–Krastanow growth has turned out to be a robust phenomenon for III–V materials; quantum dots have been produced in this way for a number of material combinations. The most important condition is that the deposited material has a substantially larger lattice constant than the substrate. InAs forms dots not just on GaAs but also on InP. The InAs/InP dots are particularly interesting because they emit at the technologically important 1.5 μm wavelength. InP can also form the dot material by using GaInP substrates where the bandgap is pushed up into the red region of the spectrum. Dots can also be grown with II–VI materials; the growth of CdSe on ZnSe is analogous to the growth of InAs on GaAs. These II–VI dots emit in the green. Finally, dots can also be formed in the nitrides, for instance by depositing InGaN on an AlGaN surface. In fact, the high emission efficiency of GaN, despite the large defect density, has been attributed to the formation of quantum dot-like structures in the active layers.

It is a challenge to arrange the dots in the plane of the substrate. An easy route with self-assembly has not been found to accomplish this. Nevertheless, dots can be encouraged to grow at particular positions by pre-patterning the substrate. This obviously limits the separation of the dots in any such array to the limits set by lithography. However, self-assembly does provide a means to order the dots in the growth direction: if the separation between successive layers is small enough (less than 15 nm for InAs/GaAs) then the dots grow directly above one another, as shown in Figure 3. In a new layer, the residual strain field from the dots in the layer below is sufficient to seed the dots. In this way, vertical stacks of dots can be built up. In fact, it has been argued that the dots tend to become more laterally ordered with each layer. The point is that if two dots are close together in one layer, their strain fields overlap, and provide only one seeding center in the subsequent layer.

Optical Properties

The main interest in quantum dot physics is the quantization in all three spatial directions, giving rise to atomic-like energy levels. For a system of self-assembled quantum dots, it is important to measure the fundamental bandgap, the energetic separation between the confined states both for electrons and for holes, the barriers to the electron continuum and hole continuum, and the oscillator strengths for the various interband transitions. All this information can be gleaned with optical spectroscopy.

Perhaps the simplest experiment to perform, if not to understand, is photoluminescence in which electron–hole pairs are excited with a laser beam.

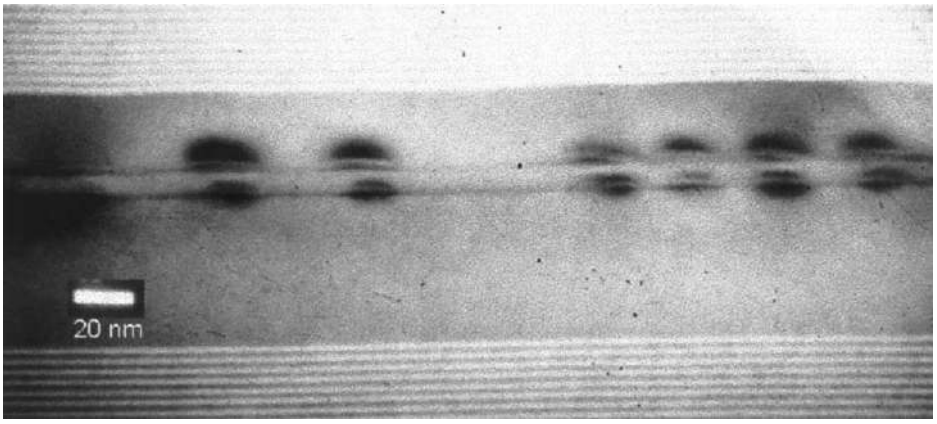

Figure 3 A transmission electron micrograph of two layers of InAs quantum dots in GaAs showing how the dots in the upper layer lie directly above the dots in the first. The image was taken by Gio Medeiros-Ribeiro.

The laser energy can be tuned to lie just above the bandgap of the wetting layer in which case the electron–hole pairs are excited in the wetting layer itself. Alternatively, if the laser energy is larger than the bandgap of the barrier material, electron–hole pairs are excited in the barrier. In both cases, the carriers relax into the dots, and then down the ladder of states in the dot, so that the dot electron and hole ground states are occupied. Emission is detected when an electron and a hole recombine. As such, the experiment measures only the energy of the ground-state exciton. However, if the pump intensity is increased such that electron–hole pairs are generated faster than they can recombine, the excited states of the dots are also occupied and recombination can also occur from excited states. An example of such an experiment is shown in Figure 4. At low pump powers, only the ground state emission is detected; at higher pump powers, the first excited state emerges, and at higher powers still, the second excited state also emerges, and so on. The spectra are broadened by the inhomogeneous broadening, i.e., from fluctuations from dot to dot, but the broadening is small enough that the excited state emission can be easily distinguished from the ground state emission. For these particular quantum dots, the energetic separation between the ground and excited state emission is 15 meV, and this corresponds to the sum of the electron and hole confinement energies. The separation between the other emissions is also about 15 meV, implying that the states are approximately equally spaced, implying in turn that the confining potential is approximately parabolic. Furthermore, in this experiment, emission from the wetting layer can also be made out, and this enables the electron and hole confining potentials to be estimated.

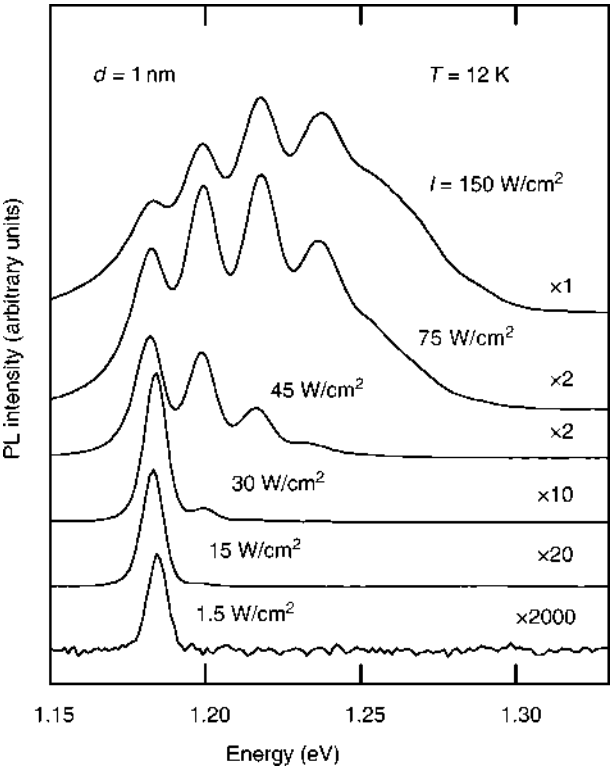

Figure 4 Photoluminescence from an ensemble of quantum dots as a function of laser excitation power. The measurement temperature was 12 K. The quantum dots are induced in an InGaAs/GaAs quantum well by a strain field generated by InP stressors on the sample surface. The stressors are grown in the Stranski–Krastanow mode, and are 1 nm away from the quantum well. At low excitation power, only the ground state emission is observed. As the power increases, emission from excited states is also observed through state filling. Reprinted with permission from Lipsanen H, Sopanen M and Ahopelto J (1995) Luminescence from excited states in strain-induced $In_xGa_{1-x}As$ quantum dots. *Physical Review* 51: 13868. Copyright 1995 American Physical Society.

Photoluminescence invariably yields the sum of electron and hole energies. In order to measure just the electronic properties, it is possible to occupy the dots with, say, two electrons (so that in each dot the two-fold degenerate ground state is fully occupied) and to excite electric dipole transitions between the electron levels. This has been accomplished with so-called charge-tunable heterostructures, with which electrons can be controllably loaded into the dots from an electron reservoir. The interlevel absorption lies in the far-infrared region of the spectrum; for InAs/GaAs quantum dots for instance, the transition is at a wavelength of 25 μm (an energy of 50 meV), as shown in Figure 5. The results of this experiment also point to an approximately parabolic confining potential from the fact that as electrons are added to the dots, the energy of the far-infrared absorption does not change much. (In the limit of a perfect parabolic potential, a famous result known as Kohn's theorem states that the long-wavelength absorption should show no energy dependence on the electron occupation.)

In self-assembled quantum dots, the confining potential can be thought of as very steep in the growth direction, and relatively shallow in the lateral plane. In other words, the dots are essentially two-dimensional disks. The electron ground state is, in analogy with atomic physics, s-like, and the excited state p-like. However, because of the disk-like nature of the dots, the p-state is four-fold degenerate, and not six-fold degenerate as in a conventional atom. One way to demonstrate this is to perform interband absorption experiments on charge-tunable structures, as shown in Figure 6. The first transition corresponds to the transition from the hole s-state to the electron s-state, and disappears when the dots are occupied with two electrons. This is simply because the Pauli principle forbids the transition once the final level is fully occupied. The second transition corresponds to the transition from the hole p-state to the electron p-state, and disappears when the dots are occupied with four electrons, confirming the degeneracies of

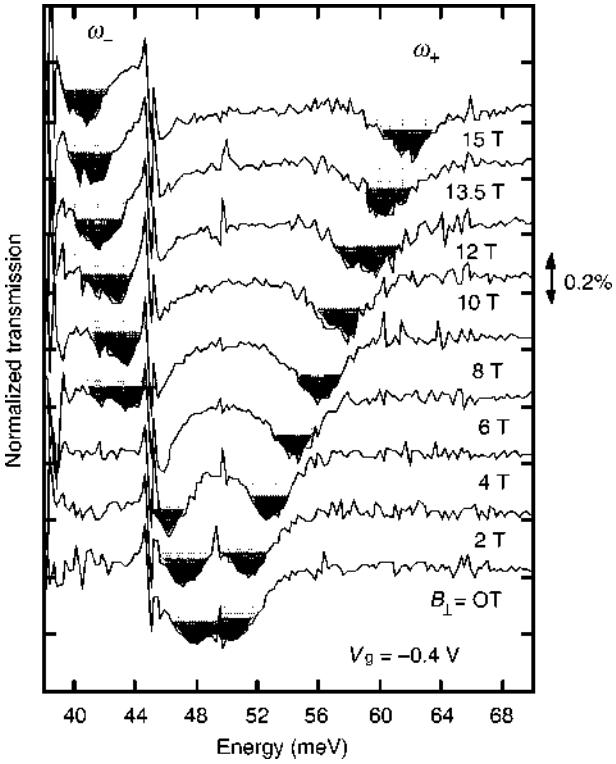

Figure 5 Transmission in the far infrared of an ensemble of InAs/GaAs quantum dots where each quantum dot is occupied with two electrons. The data were taken at a temperature of 4.2 K. The shaded transmission minima correspond to absorption from the quantum dots. The transitions are between the electron ground state and the first excited state. Two peaks are observed even at zero magnetic field, suggesting that the dots are slightly oval in shape. In a magnetic field, the two peaks move apart, which is a consequence of the Zeeman effect. The strong feature at 45 meV arises from an electronic interaction with an interface phonon. Reprinted with permission from Fricke M, Lorke A, Kotthaus JP, Medeiros-Ribeiro G and Petroff PM (1996) Shell structure and electron-electron interaction in self-assembled InAs quantum dots. *Europhysics Letters* 36: 197. Copyright 1996 European Physical Society.

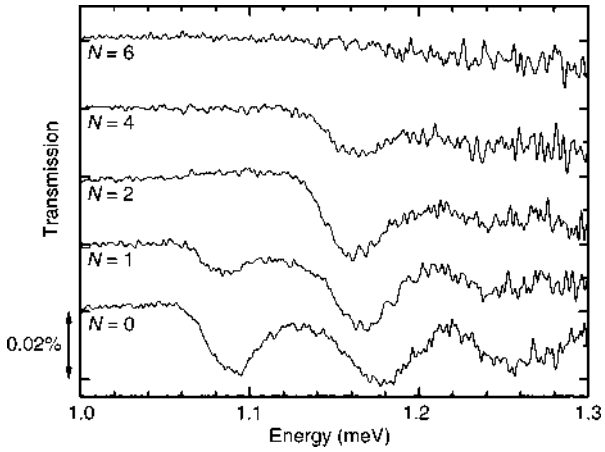

Figure 6 Transmission in the near infrared of an ensemble of InAs/GaAs quantum dots at 4.2 K. Spectra are shown for different values of N, the electron occupation. At $N = 0$, there are three transitions corresponding to transitions from a valence state to an electron state. These are the s–s, p–p and d–d transitions, where the states are labeled in analogy to atomic physics. At $N = 2$, the electron ground state is fully occupied, and the s–s transition disappears because it is blocked, a consequence of the Pauli principle. Similarly, the p–p transition is blocked at $N = 6$ when both the electronic s and p states are fully occupied. Reprinted with permission from Warburton RJ, Dürr CS, Karrai K *et al.* (1997) Charged excitons in self-assembled semiconductor quantum dots. *Physical Review Letters* 79: 5282. Copyright 1997 American Physical Society.

the confined states. This experiment, unlike a photoluminescence experiment, also yields a value for the oscillator strength. The result can be understood simply in terms of the overlap integral between the hole and electron states. In other words, these dots are in the strong confinement regime where the confinement energies are larger than the exciton binding energy. In the other limit, the weak confinement regime, the exciton binding energy dominates over the confinement energies, and it makes sense only in this limit to think of an exciton moving as a coherent entity throughout the dot. In the weak confinement regime, the oscillator strength is actually larger than in the strong confinement regime.

An important issue in interpreting photoluminescence experiments is the rate at which a highly excited electron–hole pair can relax to the ground state of a quantum dot. Relaxation in the continuum is known to occur on a picosecond time-scale: electrons and holes relax by emitting LO-phonons if possible; and if not, acoustic phonons. Capture by a dot also occurs quickly (apart from in low-density samples where capture may take a few hundred picoseconds). It was initially thought that relaxation of carriers once captured by a quantum dot would be hindered by the discrete nature of the density of states. The argument is that the energy separation between the levels is too large for acoustic phonon emission; and there is no reason why the separation between the levels should match the LO-phonon energy (which has a very weak dispersion). Nonetheless, experiments have shown that relaxation from excited dot states to the ground state is rapid, occurring on a picosecond time-scale, pretty much as for quantum wells. Auger processes can be important, whereby an electron in a dot loses energy by promoting an electron in the continuum to a higher energy. But relaxation can also be fast without a large population of mobile carriers in the wetting layer, and this is thought to arise through multiphonon processes. The radiative lifetime is typically 1 ns for self-assembled quantum dots, depending to some extent on the size and material system. As for quantum wells, relaxation occurs much faster than recombination.

Lasers and Self-Assembled Quantum Dots

The motivation for using quantum dots as the gain medium for semiconductor lasers is that they should have advantageous properties over quantum wells. These include a lower threshold current, better temperature stability, and a higher modulation frequency. All of these possible advantages stem from the quantization. In a quantum dot, there are discrete levels each with low degeneracy so that population inversion is easy to achieve. Also, there are large potential barriers to the continuum of states in the surrounding matrix, and an energy separation between levels in the dot at least comparable to the thermal energy at room temperature. This should make a quantum dot device relatively insensitive to temperature. The threshold current of a semiconductor laser usually behaves as $I = I_0 e^{T/T_0}$ where T is the temperature, T_0 is the characteristic temperature, and I_0 the low-temperature threshold current. Generally speaking, lasers are desirable with a small I_0 and a large T_0.

The main obstacle to realizing the benefits of full spatial confinement in self-assembled quantum dots is the inhomogeneity. For a particular lasing wavelength only a fraction of the dots can contribute to the lasing signal. Additionally, the dot densities are typically about 10^{11} cm^{-2} for these applications, implying that in the layer plane there is more unoccupied space than dot material. All in all, the gain at a particular wavelength is limited. Nevertheless, it would appear that even with inhomogeneous dot ensembles the advantages inherent in zero-dimensional states can be exploited.

In addition to the density of states, there is an important difference in the nature of the gain spectrum between quantum well and quantum dot lasers. In a quantum well, gain exists over a range of energies through band filling. If an electron–hole pair is stimulated to emit by the lasing field, then the vacancies will be quickly filled by carrier–carrier scattering in the bands, maintaining a quasi-thermal equilibrium in the conduction and valence bands. In quantum dots at low temperature, capture into a particular dot is a random process, and given the rapid relaxation and the high barriers to the continuum, an electron–hole pair is very likely to emit from the dot where it was created. In other words, the system is not in thermal equilibrium. At higher temperatures, thermal escape out of the dots becomes likely, enabling the system to come into thermal equilibrium. The transition into thermal equilibrium will depend critically on the barrier height, and with appropriate design may take place close to, or even above, room temperature.

Quantum dot lasers have now been realized, and emit continuous wave at room temperature. Progress has been rapid in the past few years. In fact the best quantum dot lasers now have the lowest I_0 and the highest T_0 of any semiconductor laser diode.

Even if the intrinsic advantages of quantum dots are largely offset by the inhomogeneous broadening for laser applications, there are some other advantages. The InAs/GaAs quantum dot emission wavelength is

significantly higher than that of InGaAs/GaAs quantum wells, and with subtle changes in the growth can be increased to 1.3 μm, corresponding to one of the low-loss windows in optical fiber. Also, much higher powers have been realized with quantum dot lasers than with quantum well lasers. In quantum well lasers, device failure occurs at high current, often due to degradation of the end mirrors. It would appear that the carrier localization in quantum dots is sufficiently strong to prevent diffusion to the end facets, allowing the devices to withstand higher currents.

Colloidal Quantum Dots

Progress has been made recently in making inorganic quantum dots with techniques from organic chemistry. Organometallic precursors are employed. The growth temperature is typically below 200 °C which has been found to give material with quite high crystalline quality. A crucial development has been the use of various organic groups to passivate the surface so that the quantum dots have bright photoluminescence. The II–VIs, being more polar than the III–Vs, are easier to produce with these techniques, and most work has concentrated on CdS and CdSe quantum dots which emit in the visible. Nevertheless, it is also possible to produce III–V quantum dots with this technology, and although the bulk GaAs bandgap lies in the near-infrared, the quantum confinement effects can be so high as to push the photoluminescence into the visible. Furthermore, HgCdTe has also been developed as a material system for quantum dots, and the emission can be tuned to the all-important 1.5 μm.

The spectral width of emission from these colloidal quantum dots tends to be very large because of a large spread in particle size. However, techniques are being developed to isolate a specific size. With a selective precipitation technique for instance, the spectral width can be reduced considerably to the extent that a fine structure can be observed in the optical response.

The advantage of these quantum dots lies in the possibility of mass-production for applications such as solar cells. The ease with which one material system can be used to cover a large spectral bandwidth is already attracting a lot of interest, for instance in using passivated and water-soluble quantum dots as biological labels. By using different materials, a huge spectral range can be covered with this technology, as illustrated in Figure 7. Of course, there are still challenges to be met in this field. The quantum dots tend to degrade in the course of time, losing their high quantum efficiency for emission,

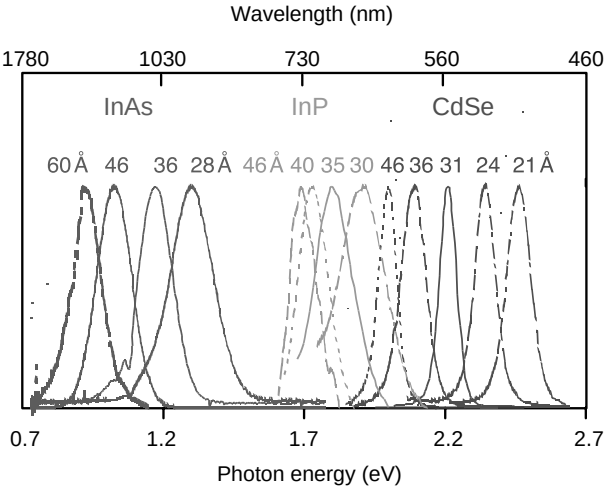

Figure 7 Room temperature emission spectra of surfactant-coated colloidal quantum dots. The quantum dots consist of InAs (left), InP (middle), and CdSe (right). Selection of the size has been used to give a particular wavelength in each band. Reprinted (abstracted/excerpted) with permission from Bruchez M Jr, Moronne M, Gin P, Weiss S and Alivisatos AP (1998) Semiconductor nanocrystals as fluorescent biological labels. *Science* 281: 2013. Copyright 1998 American Association for the Advancement of Science.

presumably as a result of changes to the surface structure. Improvements have been made recently by capping the dots with a high-bandgap semiconductor, such as ZnS, which hinders interactions of the exciton with the surface states. The emission from organic-based dots also tends to show temporal fluctuations in wavelength, revealed experimentally from the emission of a single dot. This is thought to come about through the charging and discharging of traps in the vicinity of the quantum dot. Fundamentally, there are difficulties injecting carriers across an interface into the colloidal dots. However, progress can be expected in all these areas, and these quantum dots may well find applications in novel photonic devices.

Single-Dot Spectroscopy

No matter how quantum dots are prepared, there is always a distribution in emission energy. This constitutes an inhomogeneous broadening mechanism. For self-assembled InAs dots on GaAs, the lowest emission width from an ensemble of dots is about 20 meV. This energy is large compared to several other energies in the system. For example, the binding energy of a charged exciton, or of an exciton–exciton complex (a biexciton), is about 2 meV in this system. Also, the width of single dot emission is two to three orders of magnitude smaller

than the ensemble width at low temperature. The width of single-dot emission is inversely related to the time over which the exciton maintains its coherence, and this should be as long as possible for future applications of quantum dots which manipulate the phase of the excitonic wavefunction. In order to investigate the properties of quantum dots in detail it is clearly necessary to measure the optical response of a single quantum dot rather than the response of a large ensemble of dots. The development of this field has been aided by the techniques developed for single-molecule spectroscopy which was first reported in the late 1980s and has subsequently become almost a standard experimental technique.

Typical densities of self-assembled quantum dots correspond to 100 per μm^2. Given that the emission wavelength is around 1 μm, it is clear that an optical technique with subwavelength spatial resolution is required to excite and detect a signal from just one quantum dot. The techniques employed include cathodoluminescence, where a focused electron beam provides the spatial resolution, and near-field optics where a subwavelength-sized aperture in a metal mask determines the spatial resolution. In the latter case, the aperture is either permanently connected to the sample surface, or it is formed on a tapered optical fiber positioned just a few nanometers away from the sample surface. An example is shown in Figure 8: with a confocal microscope the spatial resolution is not high enough to isolate spectroscopically a single dot. With increased spectral resolution, however, just a small number of peaks remain, each corresponding to the emission from a single dot. A further technique is to etch nanometer-sized mesas in the sample such that each mesa contains on average just one quantum dot. In some cases, single quantum dot emission has been detected with a conventional optical microscope (usually a confocal microscope) by adjusting the growth conditions to give particularly low dot densities. This can be achieved with colloidal nanocrystals by spinning a dilute suspension of the dots onto a substrate, a technique which has been frequently used in single-molecule spectroscopy. In each case, the photoluminescence signal from a single quantum dot is not large, and a detector capable of counting individual photons with a high quantum efficiency is required. A cooled silicon CCD camera is very convenient, giving not only a very low dark count and high quantum efficiency, but also a huge multiplexing advantage. However, the response of a Si detector tails off in the visible, and decreases rapidly at about 1050 nm in the near infrared, so it is by no means a universal solution.

The most striking feature of the emission from a single inorganic quantum dot is its spectral purity. Linewidths as small as a few μeV (a few hundred MHz) have been reported at low temperature for

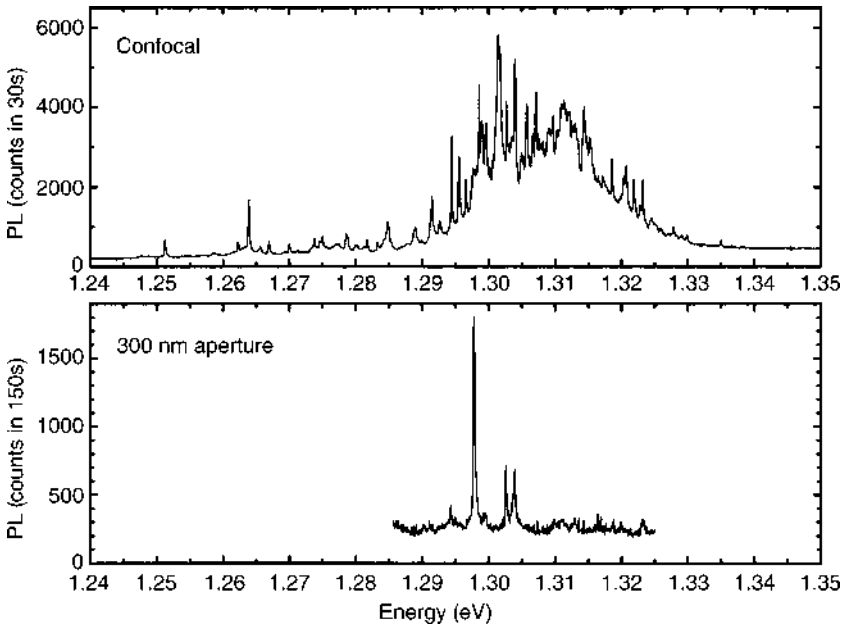

Figure 8 Emission from quantum rings from a 1 μm^2 area of the sample, and from a 0.07 μm^2 area of the sample, both taken at 4.2 K. The quantum rings are generated with a Stranski–Krastanow growth procedure using InAs and GaAs. The first spectrum was taken with a low-temperature confocal microscope; the second with a near-field technique. It can be seen how large statistical fluctuations arise in the first spectrum, because the number of rings probed is not sufficient to give a smooth distribution function. In the second spectrum, just a few peaks can be discerned above the noise level, each arising from the emission from an individual quantum ring.

quantum dots formed in a localizing potential in a quantum well. These linewidths can be interpreted as de-phasing times of the excitons. The maximum possible de-phasing time is obviously the radiative lifetime, in which case the line is homogeneously broadened. This limit has been observed at low temperature on natural quantum dots, but not at elevated temperature: as the temperature increases, scattering of the excitons by phonons decreases the coherence time, observed through a broadening of the emission.

The sharp lines observed in single-dot spectroscopy give access to fine structure on a meV energy scale. Most of this fine structure arises through Coulomb interactions between the electrons and holes. For example, two recent experiments have explored the emission as a function of electron occupation, and exciton population. In both cases, new lines emerge when a single dot is populated by more than one electron–hole pair.

Perspectives

The field of semiconductor quantum dots has made rapid advances in the past decade largely through the discovery of self-assembly in the growth of lattice-mismatched inorganic semiconductors. The most immediate application of these defect-free nanostructures is as the gain medium for semiconductor laser diodes, the aim being to reduce the threshold currents beyond those achievable with quantum wells. This approach is already proving highly successful. However, there are also other possibilities for novel photonic applications of quantum dots. These include memory elements for image processing, detectors and sources for single photons, and perhaps even elements for the coherent manipulation of wavefunctions.

See also

Semiconductor Materials: Band Structure Engineering; GaAs Based Compounds. **Semiconductor Physics:** Band Structure and Optical Properties; Excitons; Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures; Recombination Processes.

Further Reading

Bayer M, Stern O, Hawrylak P, Fafard S and Forchel A (2000) Hidden symmetries in the energy levels of excitonic artificial atoms. *Nature* 405: 923.

Bimberg D, Grundmann M and Ledentsov NN (1998) *Quantum Dot Heterostructures*. Chichester, UK: Wiley.

Bruchez M, Jr, Moronne M, Gin P, Weiss S and Alivisatos AP (1998) Semiconductor nanocrystals as fluorescent biological labels. *Science* 281: 2013.

Empedocles SA and Bawendi MG (1997) Quantum-confined Stark effect in single CdSe nanocrystalline quantum dots. *Science* 278: 2114.

Fricke M, Lorke A, Kotthaus JP, Medeiros-Ribeiro G and Petroff PM (1996) Shell structure and electron–electron interaction in self-assembled InAs quantum dots. *Europhysics Letters* 36: 197.

Gaponenko SV (1998) *Optical Properties of Semiconductor Nanocrystals*. Cambridge, UK: Cambridge University Press.

Leonard D, Pond K and Petroff PM (1994) Critical layer thickness for self-assembled InAs islands on GaAs. *Physical Review B* 50: 11687.

Lipsanen H, Sopanen M and Ahopelto J (1995) Luminescence from excited states in strain-induced $In_xGa_{1-x}As$ quantum dots. *Physical Review B* 51: 13868.

Tersoff J, Teichert C and Lagally MG (1996) Self-organization in growth of quantum dot superlattices. *Physical Review Letters* 76: 1675.

Warburton RJ, Dürr CS, Karrai K, *et al.* (1997) Charged excitons in self-assembled semiconductor quantum dots. *Physical Review Letters* 79: 5282.

Warburton RJ, Schäflein C, Haft D, *et al.* (2000) Optical emission from a charge-tunable quantum ring. *Nature* 405: 926.

Weisbuch C and Vinter B (1991) *Quantum Semiconductor Structures*. San Diego, CA: Academic Press.

Xie Q, Madhukar A, Chen P and Kobayashi NP (1995) Vertically self-organized InAs quantum box islands on GaAs(100). *Physical Review Letters* 75: 2542.

Type-II Quantum Wells and Superlattices

I Vurgaftman and J R Meyer, Naval Research Laboratory, Washington, DC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

When thin layers of two semiconductor materials are combined to form a heterojunction, quantum well (QW), or superlattice, the properties of the resulting 'quantum heterostructure' are to a large extent governed by the alignment of the conduction and valence bands at the interfaces. Usually the overall conduction-band (CB) minimum lies in the same material as the overall valence-band (VB) maximum, as for example in the $GaAs/Al_{0.3}Ga_{0.7}As$ QW illustrated in Figure 1a. In such a type-I structure,

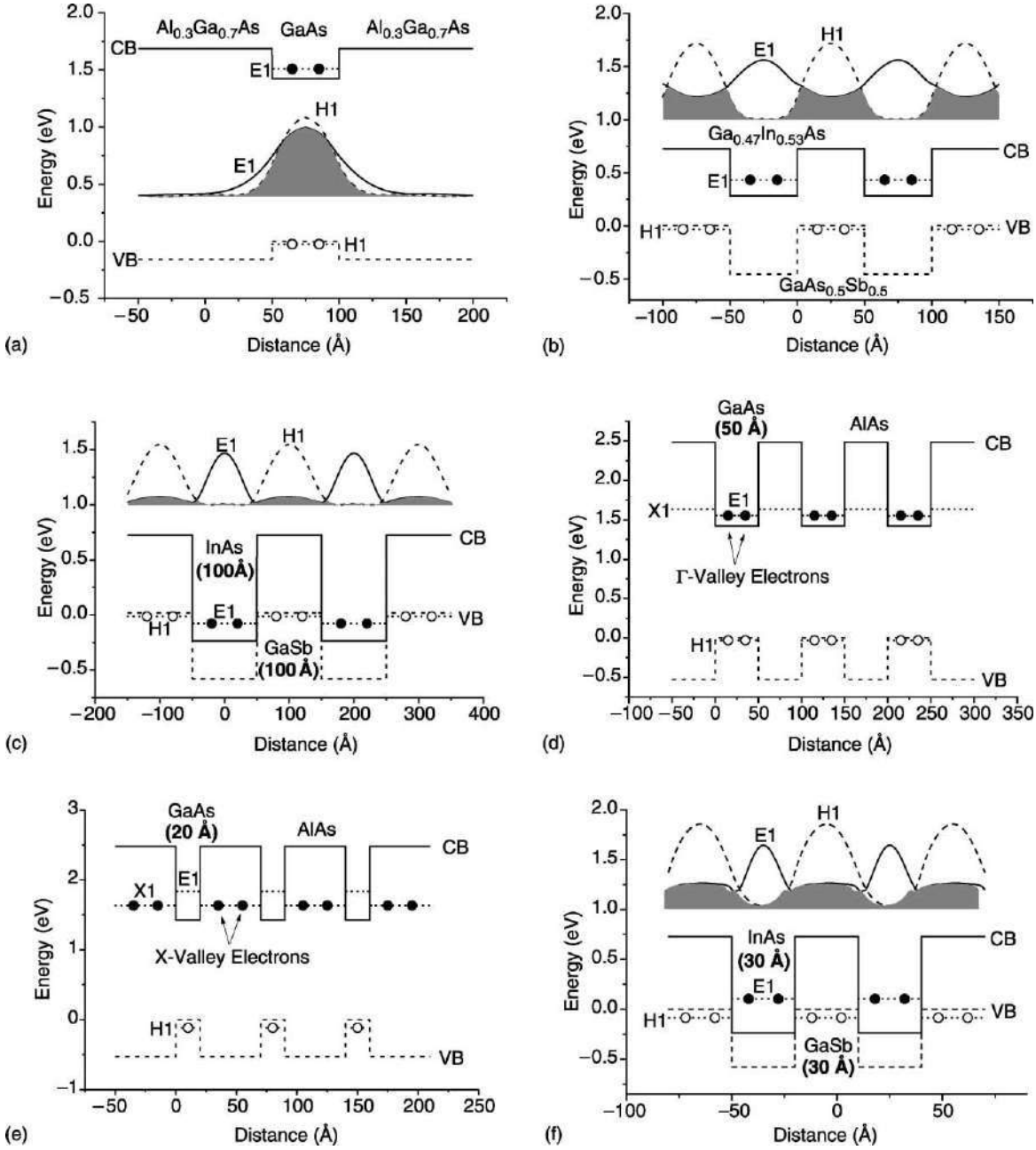

Figure 1 Conduction band (CB) and valence band (VB) profiles along with selected sub-band energies and wavefunctions for: (a) 50 Å GaAs/$Al_{0.3}Ga_{0.7}As$ type-I quantum well (GaAs substrate); (b) 50 Å/50 Å $Ga_{0.47}In_{0.53}As/Ga_{0.5}As_{0.5}Sb$ type-II staggered-gap superlattice (InP substrate); (c) 100 Å/100 Å InAs/GaSb type-II broken-gap semimetallic superlattice (GaSb substrate); (d) 50 Å/50 Å GaAs/AlAs type-I superlattice (GaAs substrate); (e) 20 Å/50 Å GaAs/AlAs type-II superlattice with indirect band alignment both in real and momentum spaces; and (f) 30 Å/30 Å InAs/GaSb type-II semiconducting superlattice. E1, H1, and X1 represent the first sub-band in the Γ-valley conduction band, Γ-valley valence band, and X-valley conduction band, respectively. The overlap between the electron and hole wavefunctions is shown in gray.

the electrons and holes both reside mostly in the GaAs and their wavefunctions overlap almost perfectly as shown in the figure.

There are cases, however, when the CB minimum is in one layer while the VB maximum is in the other. For this so-called type-II alignment, which is exemplified by the $Ga_{0.47}In_{0.53}As/GaAs_{0.5}Sb_{0.5}$ superlattice of Figure 1b, the electrons and holes tend to separate spatially. It should not be surprising that this significantly influences many of the basic optical and electronic properties. Just one example is the band-edge absorption, whose strength is proportional to the square of the wavefunction overlap integral.

In contrast to the staggered type-II heterostructure of Figure 1b, another possibility is the broken-gap type-II alignment of the InAs/GaSb superlattice shown in Figure 1c. Since the CB minimum in InAs actually lies below the VB maximum in GaSb, this combination becomes a semimetal even though the bulk constituents are both semiconductors.

The nature of the band alignment can also be strongly influenced by shifts of the electron and hole energy levels induced by quantum confinement. Figure 1d illustrates a type-I GaAs/AlAs superlattice with relatively thick GaAs layers. Notice that the same superlattice becomes type-II if, as in Figure 1e, the GaAs layers are thin enough that the GaAs Γ-valley electron level (center of the Brillouin zone) is pushed to a higher energy than the AlAs X-valley level (boundary of the Brillouin zone). Similarly, quantum confinement can turn the broken-gap (semimetallic) type-II InAs/GaSb superlattice of Figure 1c into a staggered (semiconducting) type-II superlattice as in Figure 1f.

Although many different type-II heterostructures have been identified over the last three decades, only a few are likely to have long-term technological importance for optics or electronics. The system that has received the most attention in terms of both fundamental scientific studies and opto-electronic device development is InAs/GaSb and related alloy combinations such as InAs/GaInSb. Following a brief overview of wider-gap type-II heterostructures, this article will concentrate primarily on that interesting narrow-gap system.

Wide-Gap Type-II Heterostructures

The best-known wide-gap type-II heterostructure is the GaAs/AlAs superlattice of Figure 1e. The band lineup is type-II provided the GaAs thickness is no greater than ≈10 monolayers (28 Å) and the AlAs layer is no thinner than the GaAs layer. Since the CB minimum is in the AlAs X-valley and the VB maximum is in the GaAs Γ-valley, this structure has an indirect energy gap in both real space and momentum space. In type-II GaAs/AlAs, the X-valley states at the CB minimum form one or more superlattice minibands, which display relatively low resistance at small biases, negative differential velocity, etc. A related structure is the GaAs/AlAs/GaAs/AlAs/GaAs double-barrier intervalley resonant tunneling diode, in which resonant or sequential phonon-assisted tunneling proceeds via intervalley transitions whenever the Fermi level in the GaAs emitter lines up with the X-valley minimum in AlAs. Structures containing AlGaAs do not necessarily need quantum confinement to establish a type-II alignment, since it occurs naturally in $Al_xGa_{1-x}As$/AlAs heterojunctions with $x > 0.3$.

The device potential of wide-gap type-II heterostructures is rather limited, owing to their low luminescence efficiencies (due in large part to the doubly indirect band alignment) and low mobilities (due to the heavy X-valley effective mass). A few proposals have focused on modulation of the band alignment between the type-I and type-II regimes. The $Ga_{0.47}In_{0.53}As/GaAs_{0.5}Sb_{0.5}$ superlattices of Figure 1b and related structures have been investigated as alternative active regions for semiconductor lasers emitting in the important 1.3–2 μm wavelength range.

InAs/GaSb/AlSb-Based Type-II Superlattices and Quantum Wells

Molecular beam epitaxy (MBE) was used to fabricate the first InAs/GaSb-based superlattices in 1977, only a few years after the initial GaAs/AlGaAs superlattice growths. It was quickly demonstrated that for relatively thick layers the band alignment is broken-gap type-II, with 100–150 meV overlap between the conduction band minimum in InAs and the valence band maximum in GaSb. In this semimetallic superlattice, GaSb valence-band electrons spill over into the InAs and leave behind holes (see Figure 1c). It was also observed that at equal layer thicknesses of ≈90 Å quantum confinement induces a semimetal-to-semiconductor transition (Figure 1f), and that the gap becomes increasingly positive when the thickness is reduced further. This is no mere scientific curiosity, since being able to tune in a wide range of energy gaps spanning 0 to >0.5 eV has enormous practical consequences for infrared (IR) emitters and detectors. Such a capability cannot be replicated by any other III–V heterostructure system.

Another happy accident is that the lattice constants of InAs (6.058 Å) and GaSb (6.096 Å) are matched closely enough to allow them to be grown together by MBE or metalorganic chemical vapor deposition (MOCVD). A third binary member of the 'antimonide' or '6.1-Å' family of growth-compatible heterostructures is AlSb (6.136 Å), which is valuable in that it serves as a quantum barrier to both InAs conduction states and GaSb valence states. The extended family includes a variety of ternary alloys (GaInSb, GaAlSb, InAsSb, AlAsSb, etc.), whose energy gaps as a function of lattice constant are given by the curves in Figure 2, and also at least three quaternary alloys: GaInAsSb, AlGaAsSb, and InAlAsSb. Taken together, this menagerie of materials

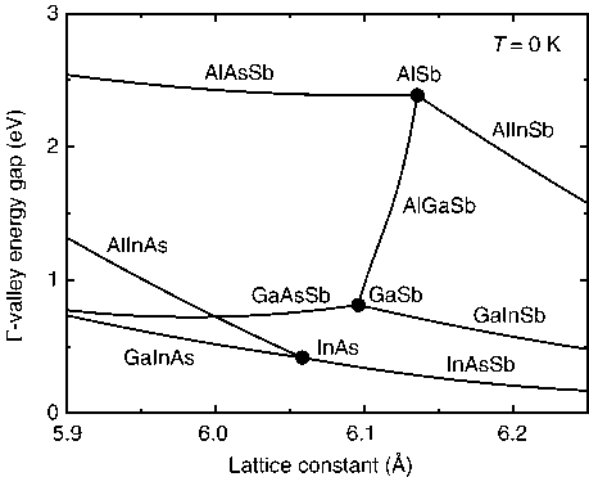

Figure 2 Direct energy gaps at zero temperature for III–V compound semiconductors and their ternary alloys with lattice constants in the vicinity of 6.1 Å.

provides the type-II IR device designer with an exceptional degree of configurational flexibility.

Type-II IR structures are typically grown on either a GaSb or InAs substrate. Since even a small amount of strain build-up can produce dislocation defects that seriously degrade the performance, one usually balances the net tensile strain (associated with layers having a lattice constant slightly larger than that of the substrate) with a compensating amount of net compressive strain (from layers whose lattice constant is slightly smaller than that of the substrate).

Band structures for complicated type-II layering configurations have been calculated by a variety of methods, including $k{\cdot}p$, tight-binding, and empirical pseudopotentials. While all reproduce the main features of the electron and hole sub-bands and dispersion relations near the center of the Brillouin zone, the relative merits of the different approaches remain controversial. Due to the small energy gaps, band-mixing effects are especially important in this system.

One consequence of the small type-II antimonide energy gap is that the electron and hole bands repel each other and both effective masses can be lighter than in any other III–V semiconductor (e.g., smaller than in the bulk constituents). This leads to high mobilities at room temperature, and promising possibilities for high-electron-mobility transistors (HEMTs). Also of interest are resonant tunneling diodes (RTDs) based on InAs/AlSb/GaSb/AlSb/InAs and related structures, in which interband tunneling via the GaSb valence states results in negative differential resistance with a very high peak-to-valley ratio. Since these devices are primarily electronic rather than optical, we will not discuss them further here.

While interfaces play an important role in any semiconductor heterostructure, they are especially crucial in InAs/GaSb and related type-II systems because the two materials on opposite sides of the interface have neither a common cation nor a common anion (unlike GaAs/AlGaAs, for example, in which both wells and barriers have As anions). By timing the shutter sequences in the growth, one can in principle end the InAs on an In layer and begin the GaSb on Sb, to produce InSb-like interface bonds. Alternatively, one can end the InAs on As and begin the GaSb on Ga to form GaAs-like bonds. Detailed characterizations by a variety of methods have shown that the device properties and even the band structure can be relatively sensitive to whether the interface bonds are InSb-like or GaAs-like.

With regard to opto-electronic devices, a potential limitation of InAs/GaSb superlattices is that in order to reach a long wavelength (small energy gap), the layers must be rather thick (e.g., ≈90 Å for zero gap). In such a structure the overlap between the electron and hole wavefunctions becomes quite small (see Figures 1c and 1f), and consequently the interband optical interactions are weak. However, Mailhiot and Smith pointed out that this issue can be circumvented to a large extent by employing $Ga_{1-x}In_xSb$ (with $x \leq 0.4$) rather than GaSb for the hole QWs. Compressive strain in the GaInSb produces a large splitting of the heavy and light hole sub-bands, along with a substantial increase of the type-II offset energy. The result is that much thinner layers (with large wavefunction overlaps) may be employed in conjunction with energy gaps corresponding to long IR wavelengths. In fact, the overlap for a short-period InAs/GaInSb superlattice such as that illustrated schematically in Figure 3 can be large enough to yield interband optical interaction strengths comparable to those in bulk materials and type-I superlattices.

While practical constraints on the minimum reproducible layer thickness set a lower wavelength limit on the order of 2 μm, the upper limit governed by uniformity considerations is quite high, perhaps in the $\lambda = 25-100$ μm range. This makes type-II antimonide quantum heterostructures a natural choice for mid-IR to far-IR photodetectors and lasers.

Type-II IR Photodetectors

Because IR detectors sense the radiant black-body emission from warm objects rather than relying on reflected visible light, they can be used for 'night vision.' Both photoconductive (PC) and

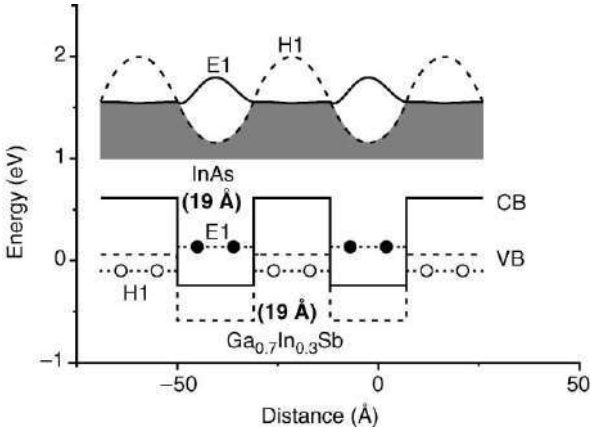

Figure 3 Conduction and valence band profiles along with electron (E1) and hole (H1) sub-band energies and wavefunctions for a 19 Å/19 Å InAs/$Ga_{0.7}In_{0.3}Sb$ type-II superlattice (GaSb substrate) with an energy gap equal to that of the InAs/GaSb superlattice in Figure 1d.

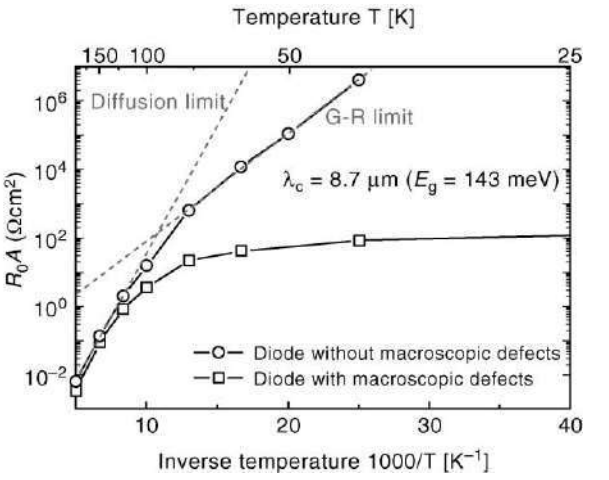

Figure 4 Temperature dependence of the dynamic impedance at zero bias R_0A for two different InAs/GaInSb superlattice diodes with cutoff wavelengths of 8.7 μm. The generation-recombination (G-R) and diffusion-limited R_0A values are indicated by the dashed lines. Reproduced with permission from Bürkle *et al.* (2000) *MRS Symp. Proc.*, 607, pp. 77–82.

photovoltaic (PV) IR detectors are now being developed with type-II active regions, usually consisting of simple InAs/GaInSb superlattices. The PV geometry, in which a reverse bias is applied to a p–n junction, is ultimately preferred since it is easily adaptable to pixelation in a 2D array with many discrete detector elements. That allows an IR camera to construct a detailed thermal picture of any given scene.

An optimized PV photodetector must have high absorption in the active region, unimpeded vertical transport to assure collection of the photogenerated carriers, and a low dark current. The thin layers of a type-II InAs/GaInSb active region help to satisfy the first two criteria, since they assure both a large wavefunction overlap (Figure 3) and rapid electron and hole tunneling through the superlattice minibands.

The dark current in a high-quality diode tends to be dominated by generation-recombination currents at low temperatures and diffusion currents at higher T. However, trap-assisted tunneling via dislocations and other macroscopic defects can significantly increase the dark current. Figure 4 plots the resistance–area product (R_0A) for diodes with and without macroscopic defects. Surface leakage resulting from inadequate sidewall passivation can also lead to a substantial contribution that is proportional to the perimeter of the mesa. The diffusion current scales as $1/\tau^{1/2}$, where τ is the recombination lifetime. To date, the defect-mediated Shockley–Read lifetimes of <100 ns have been significantly shorter than in competing HgCdTe IR materials (>1 μs). However, Auger lifetimes are substantially enhanced, as will be discussed further in the context of type-II IR lasers. In an Auger event, the energy and momentum of the recombining electron–hole pair is transferred to a third carrier, which scatters to a higher-lying state in either the same or a different band. Because three different carriers must interact simultaneously, the recombination rate scales with the cube of the carrier density. Usually Shockley–Read recombination dominates at lower temperatures and Auger processes at higher T.

Typically the n-side of a type-II diode is heavily doped (10^{18} cm^{-3} range) and relatively thin, while most of the absorption occurs on the lightly doped (10^{16} cm^{-3} range) p-side. A standard figure of merit is the detectivity (D^*), which is essentially the signal-to-noise ratio in a device of standard size. For 77 K operation, type-II PV detectors with cutoff wavelengths of 7.5 μm and 12 μm achieved D^* of 1×10^{12} cm Hz$^{1/2}$/W and 5×10^{10} cm Hz$^{1/2}$/W, respectively. Those values are nearly as high as typical results for the much more mature HgCdTe FPA technology.

It is anticipated that type-II detectors will be particularly attractive at very long wave IR (VLWIR), at which it becomes increasingly difficult to maintain adequate control over the HgCdTe compositional uniformity. Recent PC detectors operated out to 22 μm, and PV devices to 16 μm. Due in part to the strong suppression of Auger recombination, type-II detectors are also expected to be advantageous at noncryogenic temperatures where future systems will increasingly operate.

Type-II IR Lasers

At visible and near-IR wavelengths out to $\approx 2\ \mu m$, the QW diode laser is by now a mature technology that dominates the commercial marketplace due to its low cost (<$1 per laser in some cases), compactness, and reliability. However, the technical challenges become much more severe in the mid-IR spectral region (3–5 μm). Before the advent of the type-II laser, no interband III–V semiconductor device emitting at $\lambda \geq 3\ \mu m$ had ever operated at room temperature. Alternatives are the lead salt laser, which is limited to very low-output powers, and the quantum cascade laser (QCL) that is based on optical transitions between two electron QW sub-bands rather than interband electron–hole processes. Mid-IR lasers are needed for high-power military applications such as IR countermeasures against heat-seeking missiles and for the sensitive detection of trace chemicals using the spectroscopic 'fingerprint' that distinguishes each species from any other.

From the 1960s to the late 1990s, conventional III–V double heterostructure (DH) and type-I QW lasers for the mid-IR never achieved the desired levels of performance because they failed to overcome certain fundamental limitations. Those included:

1. inadequate band offsets, which at higher temperatures made it difficult for the QW or DH active regions to retain the injected electrons and holes;
2. rapid Auger nonradiative decay, which led to excessive threshold current densities at higher temperatures; and
3. restricted spectral range, limited to wavelengths shorter than the cutoff for bulk $InAs_{1-x}Sb_x$.

However, it has more recently become apparent that all of these limitations can be either minimized or removed by employing a type-II InAs/GaSb-based active region.

For example, electrical confinement is far more effective because the GaSb conduction band confines InAs electrons and the InAs valence band confines GaSb holes (see Figure 1f). Auger recombination is also significantly suppressed, in part because the holes are nearly as light in the plane as the electrons, making it much more difficult to conserve both momentum and energy in the three-carrier process. Nearly an order-of-magnitude suppression of the room-temperature Auger rate has been reported, which translates directly into a similar reduction of the threshold current density. And as was discussed above in connection with IR photodetectors, very long wavelengths are readily attainable because the energy gap in a type-II active region can be arbitrarily small. Type-II lasers have already operated cw at $\lambda > 7\ \mu m$, and cryogenic operation out to wavelengths as long as 100 μm has been projected theoretically.

In 1986 the first type-II semiconductor laser, which emitted at $\lambda = 1.86\ \mu m$ from single liquid-phase epitaxy (LPE)-grown heterojunctions, was demonstrated. Since the mid-1990s, 3–4 μm GaInAsSb/InAs-based single-interface type-II diodes have been developed that operate up to 205 K in pulsed mode.

In 1995, lasing at $\lambda = 3.2–3.5\ \mu m$ was reported from type-II $InAs/Ga_{0.75}In_{0.25}Sb$ superlattice diodes grown by MBE. Subsequent work on these devices extended the temperature range to 255 K in pulsed mode and 180 K for cw operation, and the emission wavelength to 4.3 μm.

While the wavefunction overlap and hence the gain for a two-constituent type-II superlattice active region can be quite favorable, penetration of the electron wavefunctions into the thin GaInSb layers is in fact so great that a wide miniband forms along the growth axis. The electrons then have strong energy dispersion in all three dimensions, whereas present-generation near-IR and visible semiconductor lasers nearly always have QW active regions with 2D carriers. The stepped quasi-2D density of states yields a much higher differential gain per injected carrier at threshold, and also narrower spectral emission lines than are attainable with a 3D electron population. The most straightforward approach to blocking the formation of a miniband in the superlattice is to add a high electron barrier between the periods, e.g., InAs/GaInSb/AlSb. However, the drawback of that approach is that the wavefunction overlap becomes much smaller than in the simple superlattice, resulting in significantly lower gain.

On the other hand, if a second electron well is added on the other side of the hole well, the resulting structure with four constituents per period (e.g., InAs/GaInSb/InAs/AlSb) combines the large wavefunction overlap of the two-constituent superlattice with 2D electron and hole confinement. Figure 5 illustrates band profiles, energy levels, and wavefunctions for the 'W' laser, which takes its name from the shape of the conduction-band profile.

Optically pumped type-II W lasers display high-power pulsed operation at temperatures well above ambient. Wavelengths have spanned the range 2.6–7.3 μm, and mid-IR vertical-cavity surface-emitting lasers (VCSELs) were also fabricated. Optical pumping is viewed primarily as a means of achieving high cw or quasi-cw output powers in the mid-IR, and in that regard W lasers have exceeded all other semiconductor approaches with 1.5 W peak output power being obtained at 71 K for 100 μs

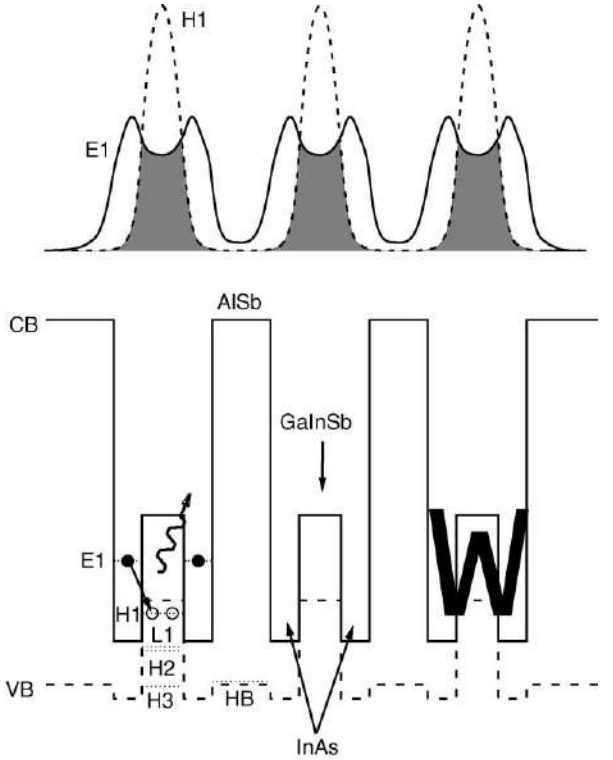

Figure 5 Conduction and valence band profiles along with selected sub-band energies and wavefunctions for InAs/$Ga_{0.7}In_{0.3}Sb$/InAs/AlSb type-II 'W' structure (GaSb substrate). Electron (E1), heavy hole (H1, H2, H3), light hole (L1), and barrier heavy hole (HB) sub-bands are indicated.

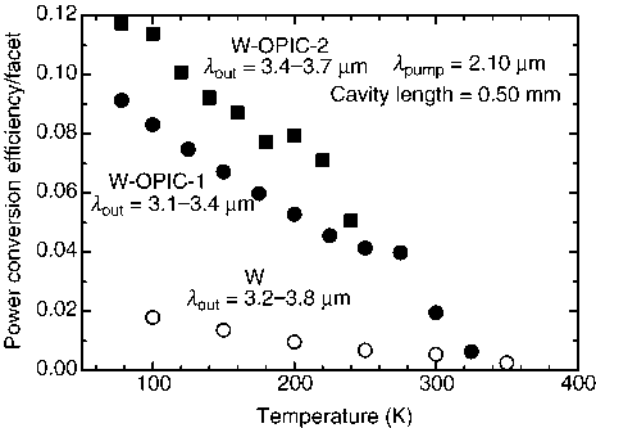

Figure 6 Power conversion efficiency per uncoated facet as a function of temperature for two different W-OPIC lasers emitting in the wavelength ranges of 3.1–3.4 μm (filled circles) and 3.4–3.7 μm (filled squares), respectively. For comparison, conversion efficiencies for a typical 'W' laser without an OPIC emitting in the same spectral range (open circles) are also shown. All of the lasers were optically pumped in pulsed mode by a 2.1 μm Ho:YAG laser, and the cavity length for all three devices was 0.5 mm.

pulses with a 10% duty cycle. The average output at 82 K for that device was up to 360 mW for 20 μs pulses and a 50% duty cycle.

Cw operation at $\lambda = 3.0$ μm was observed nearly to room temperature (290 K), and even a 6.1 μm emitter operated cw to 210 K. The cw output power of 540 mW/facet from a device with 80 W QW periods was the highest ever reported for 78 K operation in the mid-IR. Despite the rather high output power, however, the differential power conversion efficiency was only 2.5% at 78 K and fell to 1% at 140 K. Much higher efficiencies were obtained using the optical pumping injection cavity (OPIC) approach, in which the active region was surrounded by semiconductor quarter-wave mirror stacks that formed an etalon cavity resonant at the 2.1 μm pump wavelength. For pulsed operation, OPIC devices with uncoated facets and only 10 active QWs yielded lower thresholds, lower internal losses, and higher power conversion efficiencies than any previous results at $T \geq 200$ K. Figure 6 illustrates that the efficiencies were 7.1% at 220 K for one device and 4% at 275 K for another.

An alternative approach to maximizing the pump absorbance despite a small number of QWs is the 'integrated absorber' (IA) concept. In this approach, the W QWs are separated by intermediate-gap GaInAsSb absorbing layers which donate optically generated electrons and holes to the active wells. At 80 K with 35 μs pulses at 2.5% duty cycle, an IA laser with only five QWs separated by ≈1600-Å-thick absorber layers had a power conversion efficiency of 9.8% and peak output power of 2.1 W. Thus far the IA has not done as well as the OPIC at high temperatures, due mostly to an inadequate valence band offset between the absorber layers and the active QWs.

The beam quality for optically pumped type-II lasers was substantially improved by using the angled-grating distributed-feedback (α-DFB) approach pioneered at shorter wavelengths. Since both the grating and the facets are needed to produce optical feedback, only a narrow angular cone is fed back into the tilted cavity. For a pump-stripe width of 50 μm, the near-diffraction-limited output beam had 15 times narrower angular divergence (1.4° full width at half-maximum (FWHM)) than the double-lobed beam emitted from a conventional Fabry–Perot laser occupying an unpatterned portion of the same bar. Even more promising results have been reported using the photonic-crystal distributed-feedback (PCDFB) laser, in which the grating is defined on a two-dimensional rectangular lattice that is tilted with respect to the facets as in the α-DFB. For a pump-stripe

width of 200 μm, the beam quality for the direction along the laser stripe was four times the diffraction limit.

Although optical pumping is the simplest way to inject electrons and holes into the active region, type-II diode lasers will ultimately be much more inexpensive, convenient, compact, and energy efficient. However, generalization of the W configuration to electrical pumping is not entirely straightforward. This is because holes injected from the p-type cladding layer may have poor growth-axis transport, and hence hole populations in the various GaInSb QWs could be highly nonuniform. Whereas the holes in a superlattice are required to tunnel through only one InAs electron well at a time (see Figure 3), the barrier for the structure in Figure 5 consists of two InAs layers in addition to the AlSb barrier. To overcome this problem the AlSb was replaced with a specially designed quaternary $Al_{0.15}Ga_{0.85}As_{0.05}Sb_{0.95}$ barrier, which raised the L1 light-hole sub-band to within ≈100 meV of the H1 heavy-hole valence band maximum. That permitted thermally excited light holes to provide the needed tunneling transport to all of the QWs. Another key design element was the insertion of relatively thick (0.6 μm) undoped $Al_{0.35}Ga_{0.65}As_{0.03}Sb_{0.97}$ 'broadened-waveguide' layers on each side of the active region, in order to minimize free-carrier absorption in the doped $Al_{0.9}Ga_{0.1}As_{0.07}Sb_{0.93}$ cladding layers. Some of the devices also contained InAs/AlSb 'hole-blocking' layers, which prevented hole leakage into the n-type broadened waveguide layer.

A pulsed $\lambda = 3.27$ μm W-diode of this type with 10 active QW periods achieved the first room temperature operation (to 310 K) of an electrically pumped III–V interband laser in the mid-IR. Another device with five QWs operated in cw mode to 200 K, the highest T_{max} to date for any interband III–V laser beyond 3 μm. The threshold at 78 K was 63 A/cm^2, and Figure 7 shows that the cw output was 140 mW at 78 K and 20 mW at 180 K. Following optimization of the turn-on voltages, series resistances, and internal losses, cw diodes with one–three QWs are projected to operate in the thermo-electric cooler range and ultimately at room temperature.

A more exotic variation on the type-II diode emitter is the interband cascade laser (ICL) that was first proposed by RQ Yang. Like the intersub-band QCL, the ICL generates multiple photons per injected electron by making an optical transition at each step of a staircase-like QW structure. While the threshold current density for the QCL is high because of the inherently short lifetimes associated with intersub-band phonon scattering, the interband configuration of the ICL eliminates that nonradiative relaxation path. Early projections that the ICL should display low thresholds, high cw operating temperatures, high output powers, and high differential quantum efficiencies (well above the conventional limit of one photon/electron) have all been confirmed experimentally.

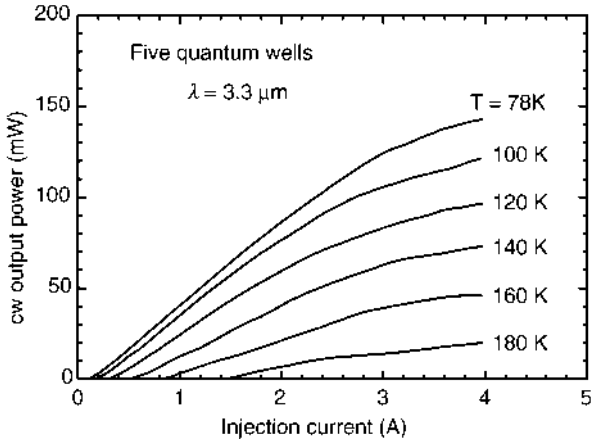

Figure 7 Cw output power as a function of injection current at temperatures between 78 K and 180 K for a 'W' diode laser with a 2-mm-long cavity and 100-μm-wide stripe emitting at a wavelength of 3.3 μm.

Figure 8 illustrates conduction and valence band profiles for a typical single stage of the ICL staircase. While this example employs a W configuration, in which two InAs electron QWs (19 Å and 13 Å) surround the active GaInSb hole QW (32 Å), some designs employ only a single InAs active well. Electrons tunneling into the InAs QWs from the preceding injection region at left emit a photon by making spatially indirect radiative transitions to the GaInSb valence band. They next tunnel into the valence states of the adjacent GaSb and GaInSb hole wells, whose function is to provide a thick barrier to prevent electron leakage from the active InAs QWs directly to the 70 Å InAs well that begins the next injection region. From the final GaInSb hole QW (40 Å), the carriers undergo near-elastic interband scattering into conduction states of the 70 Å InAs QW. Electrons traversing the superlattice miniband of the injection region, which in this case consists of eight digitally graded InAs wells, finally tunnel into the active electron QWs of the next period. This series of events is repeated at each step of the ICL staircase.

Following the first ICL demonstration in 1997, there have been reports of threshold current densities 13 A/cm^2 at 80 K, maximum operating temperatures of up to 300 K for pulsed operation and 150 K for cw (with epitaxial-side-up mounting), pulsed output powers of 6 W/facet and slope quantum efficiencies exceeding 600%. A cw wall plug efficiency of 22%

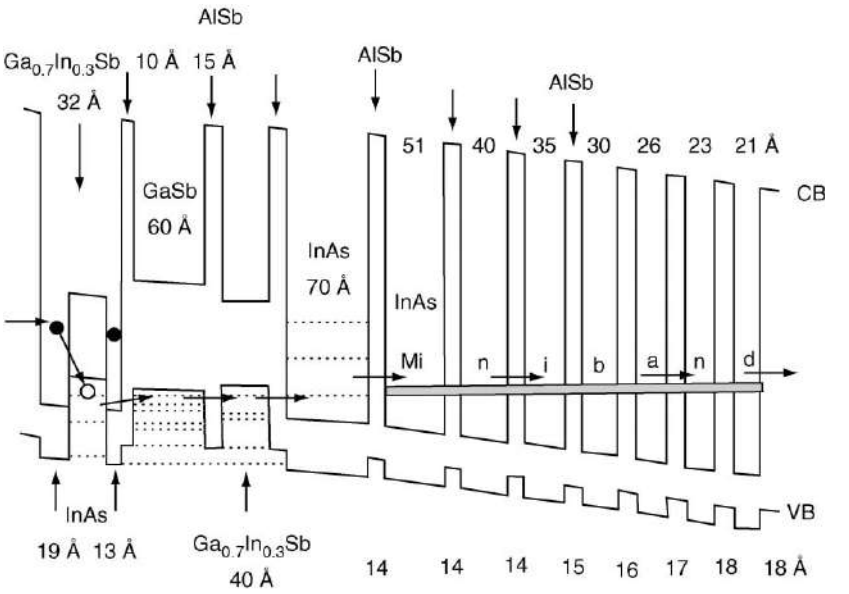

Figure 8 Conduction and valence band profiles for one period of an interband cascade laser with a W active region and three hole wells. Sub-band energy levels are also indicated.

at 80 K was recently reported. While substantial further improvements are expected, the primary disadvantage of the ICL is its complexity.

Despite numerous dramatic advances by type-II mid-IR lasers since their first demonstration in 1995, the performance has not yet progressed to the level that will be required to support a robust commercial technology. One key issue is the cw operating temperature, which must increase to at least the thermo-electric cooler range if the potential market is to extend beyond niche applications. Higher-temperature operation will result automatically once the number of QWs, and hence the threshold current density, can be reduced. Further suppression of the Auger decay rate would also be highly beneficial if it is proven possible.

The laser output power and efficiency are governed by the internal loss, which in type-II QWs tends to be dominated by resonant free-hole intersub-band transitions. The details of the valence sub-band structure are not well known at present, although recent experiments have apparently identified a number of the transition energies. As in the case of the threshold, reducing the number of QWs is highly beneficial since it reduces the number of free carriers that can absorb the lasing photons. The optically pumped OPIC and IA structures discussed above achieved dramatic reductions in the internal losses by following that strategy.

The beam quality attainable from a semiconductor laser is governed by the linewidth enhancement factor (LEF), which is proportional to the variation in the refractive index induced by a small change in the carrier density. A large positive value of the LEF leads to filamentation in broad-area lasers as tiny peaks in the optical intensity cause carrier depletion and a local increase in the refractive index, which in turn stimulates further focusing of the output beam. The output then consists of an incoherent superposition of many rapidly diverging filaments, resulting in a beam quality that is several orders of magnitude worse than the diffraction limit. While theoretical LEFs for optimized type-II W structures can be in the 1.5–2 range even at room temperature, recent Hakki–Paoli measurements have yielded values in the 2.5–5 range. This is possibly due to additional inhomogeneous broadening of the gain spectrum induced by interface roughness.

See also

Semiconductor Materials: Band Structure Engineering; Quantum Dots. **Semiconductor Physics:** Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures; Recombination Processes.

Further Reading

Mailhiot C and Smith DL (1989) Long-wavelength infrared detectors based on strained InAs-GaInSb type-II superlattices. *Journal of Vacuum Science and Technology* A7: 445–449.

Meyer JR, Malin JI, Vurgaftman I, Hoffman CA and Ram-Mohan LR (1997) Antimonide-based quantum heterostructure devices. In: Manasreh MO (ed.) *Strained*

Quantum Wells and Their Applications, pp. 235–272. Amsterdam, The Netherlands: Gordon and Breach.
Olafsen LJ, Vurgaftman I and Meyer JR (in press) Antimonide mid-IR lasers. In: Choi HK (ed.) *Long-Wavelength Semiconductor Lasers*. New York: Wiley.
Ram-Mohan LR, Vurgaftman I, Meyer JR and Dossa D (2000) Wavefunction engineering: a new paradigm in quantum nanostructure modeling. In: Nalwa HS (ed.) *Handbook of Nanostructured Materials and Nanotechnology*, vol. 2, pp. 707–739. New York: Academic Press.
Razeghi M (1998) Current status and future trends of infrared detectors. *Optoelectronics Review* 6: 155–194.
Vurgaftman I, Felix CL, Aifer EH and Meyer JR (2000) Interband III–V mid-IR semiconductor lasers. In: Perera AGU and Liu HU (eds) *Handbook of Thin Film Devices*, vol. 2, pp. 171–193. San Diego, CA: Academic Press.
Vurgaftman I, Felix CL, Bewley WW, *et al.* (2001) Mid-IR "W" lasers. *Philosophical Transactions of the Royal Society A* 359: 489–504.
Vurgaftman I, Meyer JR and Ram-Mohan LR (2001) Band structure parameters for III–V compound semiconductors and their alloys. *Applied Physics Reviews, Journal of Applied Physics* 89: 5815–5875.
Wilson BA (1988) Carrier dynamics and recombination mechanisms in staggered-alignment heterostructures. *IEEE Journal of Quantum Electronics* 24: 1763–1777.
Yang RQ, Bradshaw JL, Bruno JD, Pham JT and Wortman DE (2002) Mid-IR type-II interband cascade lasers. *IEEE Journal of Quantum Electronics* (Feature Section on Mid-IR Quantum Cascade Lasers) 38: 559–568.

SEMICONDUCTOR PHYSICS

Contents

Outline of Basic Electronic Properties
Band Structure and Optical Properties
Excitons
Impurities and Defects
Infrared Lattice Properties
Light Scattering
Polarons
Quantum Wells and GaAs-Based Structures
Recombination Processes
Spin Transport and Relaxation in Semiconductors; Spintronics
Surface Photovoltage Spectroscopy of Semiconductors

Outline of Basic Electronic Properties

C R Pidgeon, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

We briefly introduce the main concepts underlying the detailed treatment of specialist aspects of semiconductor physics in ensuing sections, viz: the free electron model, density of states, Fermi-Dirac statistics, energy bands and energy gaps, effective mass, electrons and holes, intrinsic and extrinsic conductivity, direct and indirect interband optical transitions, excitons and low-dimensional systems.

Introduction

A proper treatment of optoelectronic devices needs to include the physical properties of the materials from which the devices are made, and this subject is covered under the section of the Encyclopedia entitled 'Semiconductor Fundamentals'. This includes sections on the basic physics common to all semiconductors useful in optoelectronic applications, the experimental techniques used to investigate the optical properties of

these semiconductors and the more detailed properties of the most widely used semiconductors, and semiconductor systems. The properties of greatest importance here are those of electrons in crystalline solids, but the interaction between the electrons and the atoms of the crystal lattice is also significant. The combination of these microscopic effects is responsible for determining the macroscopic optical and electrical properties of semiconductors, and in this section we give the necessary background to these topics.

The classical treatment is extremely useful in describing the electrical conductivity of solids; the electrons are regarded as charged particles obeying Newton's laws of motion, and electrical resistance is assumed to arise from collisions with the vibrating lattice and/or impurities. However, the classical theory does not give a method of predicting the natural frequencies, or energies, of these electronic oscillators. To do this we need to use quantum mechanics, and the elements of this are reviewed here. We start with the free electron theory of metals, where the electrons are treated as particles constrained to move in a potential box. For convenience we begin by considering the situation in one dimension (1D).

Free Electron Theory of Metals

Free electrons in a 1D potential box of length L in the x-direction are described by a plane wave function that is normalized within the box, $\psi = (1/L)e^{ikx}$, and satisfies the Schrödinger equation in the x-direction:

$$\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + E\psi = 0 \qquad [1]$$

where E is the electron energy. This gives the familiar parabolic E–k dispersion relation (see Figure 1), $E = \hbar^2k^2/2m$ (with electron momentum $p = \hbar k$). In order to evaluate physical problems it is necessary to define the boundary conditions for the wave function. We assume fixed boundary conditions (i.e., $\psi = 0$ at $x = 0, L$) leading to the quantization condition $k = n\pi/L$ where n is a positive integer.

Density of States

Physical properties of solids depend on the number of allowed k-values per unit range of k (or E-values per unit range of E), a quantity referred to as the density of states, D(k) or D(E) – for example, the number of allowed states between k and $k + dk = D(k)dk$. From the fixed boundary condition ($k = n\pi/L$, the number of allowed values of k per unit range of k in 1D is just

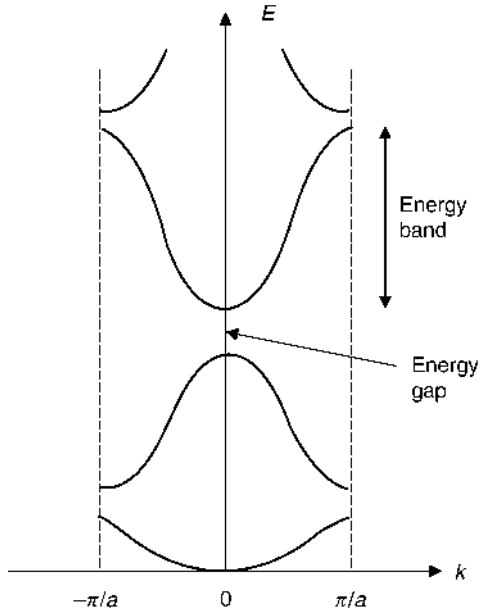

Figure 1 Schematic representation of energy bands in one dimension, characterized by gaps at the high symmetry points of the Brillouin zone.

$D(k) = L/\pi$. In terms of energy, using the E–k dispersion relation, we can write (including the factor 2 for two possible spin states):

$$D(E)dE = D(k)(dk/dE)dk$$
$$= (L/2\pi)(2m/\hbar^2)^{1/2}E^{-1/2}dE \qquad [2]$$

i.e., the density of states depends on $E^{-1/2}$ for the 1D situation.

For a cube of side L, in three dimensions we generalize the identity for fixed boundary conditions:

$$\frac{L^2k^2}{\pi^2} = n_x^2 + n_y^2 + n_z^2 = R^2 \qquad [3]$$

where $k_x = n_x\pi/L$ etc., and the states fill up the positive quadrant of a sphere of radius ($R = Lk/\pi$) in k-space. Thus, the number of states in a given volume of k-space is numerically equal to that volume, so that the number of states between R and $R + dR$ in the positive quadrant (hence $1/8^{th}$) of a spherical shell of radius R in k-space is

$$D(R)dR = 4\pi R^2 dR(1/8) \qquad [4]$$

i.e., $(1/8)^{th}$ of the surface area of the shell multiplied by the thickness of the element. Changing the variable to k, the number of states between k and $k + dk$ is

$$D(k)dk = D(R)\frac{dR}{dk}dk = \frac{V}{2\pi^2}2k^2dk \qquad [5]$$

where the volume $V = L^3$, and a factor of 2 has again been included to account for the two possible spin states of the electron. This in turn, through the relation between E and k, can be expressed in terms of the density of states per unit energy so that the number of states between E and $E + \mathrm{d}E$ is given by:

$$D(E)\mathrm{d}E = D(k)\frac{\mathrm{d}E}{\mathrm{d}k}\mathrm{d}E = \frac{2V}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}}\mathrm{d}E \quad [6]$$

In two dimensions, by the same argument:

$$\frac{L^2k^2}{(\pi)^2} = n_x^2 + n_y^2 = R^2 \quad [7]$$

Therefore $D(R)\mathrm{d}R = (2\pi R/4)\,\mathrm{d}R$, where the factor 1/4 implies the positive quarter of the 2D ring between R and $R + \mathrm{d}R$ in k-space, $A = L^2$ and

$$D(k)\mathrm{d}k = D(R)\frac{\mathrm{d}R}{\mathrm{d}k}\mathrm{d}k = \frac{A}{\pi}k\mathrm{d}k \quad [8]$$

From the E–k relation we have the number of states between E and $E + \mathrm{d}E$ as:

$$D(E)\mathrm{d}E = D(k)\frac{\mathrm{d}E}{\mathrm{d}k}\mathrm{d}E = \frac{A}{\pi}\left(\frac{m}{\hbar^2}\right)\mathrm{d}E \quad [9]$$

Thus, in 2D systems we have the important result that $D(E)$ is a constant.

Fermi-Dirac Distribution of Electrons

At zero temperature an atom exists in what is known as the ground state. All the electrons rest in the lowest energy states allowed to them by the Pauli principle, and fill the allowed states up to an energy known as the Fermi energy, E_F. Above absolute zero of temperature this is no longer the case, and thermal energy allows some electrons to exist in higher energy states. The probability $f(E)$ that a state of energy E will be occupied for an ideal electron gas in thermal equilibrium (thermal energy kT) is governed by Fermi-Dirac statistics:

$$f(E) = \frac{1}{1 + \exp\left(\dfrac{E - E_\mathrm{F}}{kT}\right)} \quad [10]$$

For a metal at room temperature the difference between the Fermi energy and the energy of the conduction band minimum is much greater than the thermal energy, $(E_\mathrm{F} - E_\mathrm{co}) \gg kT$; this is also true for a strongly n-type semiconductor (see below) and the distribution is said to be degenerate.

Energy Bands and Energy Gaps: Semiconductors

In the presence of the (perturbing) periodic crystal potential, $V(r)$, the above simplicity is lost: this is the so-called 'nearly free electron model', and the perturbed Schrödinger equation is now:

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + V(x)\psi = E\psi \quad [11]$$

for the x-direction. This topic is discussed in all undergraduate texts on solid state physics, but we list the main results of the model as follows:

1. The electrons that we are concerned with are confined within a periodic lattice. Consequently the wave function ψ is modulated with the periodicity of the lattice. By symmetry arguments the correct wave function can be shown to be:

$$\psi_k(x) = \frac{1}{\sqrt{L}}\mathrm{e}^{ikx}u_k(x) \quad [12]$$

   where $u_k(x)$ is a function that has the symmetry of the lattice. $\psi_k(x)$ is known as the Bloch function and is a satisfactory solution to the Schrödinger equation. Because of the periodic nature of the Bloch equation, it is always possible to select the wavevector index k of any Bloch function to lie within the range $\pi/\mathrm{a} < k < \pi/a$. This region is referred to as the first Brillouin zone, and the representation where the E–k dispersion is mapped back entirely into the first Brillouin zone is called the 'reduced zone scheme' (Figure 1).
2. Regions of E appear where solutions $\psi_k(x)$ are not allowed, corresponding to Bragg reflections of the electron waves by the lattice; these are the so-called energy gaps (forbidden energy regions) which separate energy bands (allowed energy regions). In 3D these discrete bands of energy form the conduction and valence bands within materials. The forbidden energy gap separates these bands. An additional quantum number n is required to label the individual energy bands corresponding to successive Brillouin zones in the wavevector, $2n\pi/a$ (where $n = 0, 1\ldots$). The position of the Fermi level determines the electron population of the energy bands, and the character of the solid (see below).
3. The parabolic E–k dispersion relation is retained near $k = 0$, and at other symmetry points of the Brillouin zone, but with the substitution of an effective mass, m^* (see below) for the free electron mass, i.e., $E = \hbar^2k^2/2m^*$ for $k \ll \pi/a$. This is referred to as the effective mass approximation.

The nearly free electron model is one limit for the description of the formation of energy bands and energy gaps in a solid. In real crystals the energy bands derive from, and retain the symmetry of, the parent atomic states of the host atoms. In this limit, the model that starts out from the wave functions of the free atoms is known as the tight binding model or the LCAO (linear combination of atomic orbitals) approximation. As neutral separated atoms are brought together to form a crystal, their wave functions overlap and the Coulomb interaction between the atom cores and the electrons splits the atomic energy levels, spreading them into bands. Each state of given quantum number of the free atom is spread in the crystal into a band of energies; a given s, p, d-state in N separate atoms will produce the corresponding N-orbital s, p, d-energy band in the solid. Including the factor of 2 for spin, a full s-band would contain 2N electrons, a full p-band 6N electrons and so on. The width of the band is proportional to the strength of the overlap interaction between neighboring atoms. Each energy band will not have the same energy as any other band over any substantial range of the wavevector.

Metals, Insulators and Semiconductors

The position of the Fermi level distinguishes whether the solid is a metal, insulator or semiconductor. Generally those bands that derive from full atomic orbital states will be fully occupied with electrons, those that derive from empty atomic states lead to empty energy bands, and those from partially full atomic states lead to partially full energy bands. The highest band of those that are filled with electrons is called the valence band; the next highest band is called the conduction band (the fundamental energy gap, E_G, is the valence/conduction band separation, referred to from now on as simply the energy gap). For a metal, E_F is in the conduction band at all temperatures; for an insulator, E_F is close to the middle of the energy gap. Semiconductors are a class of crystals with small energy gaps; they are insulators at $T = 0\,K$, but are conductors at finite temperatures due to either thermal excitation of electrons across the gap or doping, n- or p-type (see below).

Effective Mass

Consider what happens to an electron when an external electric field F is applied. Suppose the electron is initially in a state k. When the field has acted on the electron for a small time dt, it has gained an energy:

$$dE = eFv\,dt = (eF/\hbar)(dE/dk)dt \qquad [13]$$

since the group velocity of the de Broglie wave (energy, $E = \hbar\omega$) is $v = \hbar^{-1}(dE/dk)$. Or simplifying we have: $dk/dt = eF/\hbar$. This gives for the resulting acceleration:

$$a = dv/dt = (1/\hbar)(d^2k/dt^2)(dk/dt)$$
$$= (eF/\hbar^2)(d^2E/dk^2) \qquad [14]$$

Comparing this result for the acceleration of a free electron of mass m under the influence of F, $a = eF/m$, it follows that the electron behaves as if it had an effective mass, m^*, given by

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{d^2E}{dk^2} \qquad [15]$$

Thus, the electron effective mass m^* gives a measure of the conduction band curvature near the minimum at $k = 0$. We note that the same result is achieved by differentiating twice the 'nearly free electron' E–k relation, $E = \hbar^2k^2/2m^*$.

Electrons and Holes

Consider the situation where a single electron is excited either optically or thermally from the top of the filled valence band to the bottom of the conduction band of a 1D lattice, and consider the collective behavior of the valence band when an external electric field is applied. Denoting the charge of an electron by $-e$ and the velocities of the electrons by v_i, we may write for the current associated with all electrons in the completely filled valence band in the absence of an external field:

$$I = -e\sum_i v_i = -e[v_j + \sum_{i\neq j} v_i] = 0 \qquad [16]$$

Thus, if the electron j were missing, we should have current

$$I' = -e\sum_{i\neq j} v_i = ev_j \qquad [17]$$

Applying an external field F, the rate of change of the current due to the field is:

$$dI'/dt = e(dv_j/dt) = -e^2F/m_j^* \qquad [18]$$

Since the holes are at the top of the valence band, m_j^* is negative and the right hand side of the equation is positive – i.e., a positive external field F gives a positive change of current at the top of the valence band. Thus, a band in which an electron is missing behaves as a 'positive hole' with an effective mass $m_h = -m_e$, and positive charge $+e$. Since m_e is negative near the top of the valence band, m_h is positive.

Impurities, Intrinsic and Extrinsic Conductivity

The key point about semiconductors from a device point of view is that one can control their conductivity. The basic method is by doping the host crystal with impurities either of valence greater than the host (donors giving excess electrons, called n-type) or of valence less than the host (acceptors giving excess holes, called p-type). The conductivity is referred to as intrinsic conductivity (from intrinsic carrier concentration, N_i) and extrinsic conductivity (from extrinsic carrier concentration, N_e) for undoped and doped material, respectively.

The impurities are chosen to have very low binding energies, so that the electrons (holes) are substantially ionized at room temperature to provide n-type (or p-type) conductivity. These shallow impurities can be treated using the simple Bohr theory of the hydrogen atom, substituting the electron orbiting around the proton in the hydrogen atom under the Coulomb potential e/r, by the electron in the crystal (electron effective mass m^*) orbiting round the ionized (donor/acceptor) impurity atom under the Coulomb potential $e/\varepsilon r$ in the crystal – where ε is the static dielectric constant of the crystal. As with the hydrogen atom, this gives a set of discrete energy levels:

$$E = \frac{m^*/m_o}{\varepsilon^2}\frac{13.6}{n^2}\text{eV} \qquad [19]$$

where 13.6 eV is the binding energy of the H-atom and n is the principal (Bohr) quantum number. The binding energy is then

$$E_B = \frac{m^*/m_o}{\varepsilon^2} \times 13.6\ \text{eV} \qquad [20]$$

For typical semiconductors (Ge, Si…) we have $m^*/m_o \sim 0.05$, $\varepsilon \sim 16$. Thus, E_B is in the range of a few meV (i.e., much less than E_G, and kT at room temperature). The energy level scheme for a doped semiconductor is shown schematically in real space (at room temperature) in Figure 2.

The total conductivity will be $\sigma = \sigma_i + \sigma_e$, where the intrinsic conductivity $\sigma_i = n_i e^2\tau/m^*$ and the extrinsic conductivity $\sigma_e = n_e e^2\tau/m^*$. τ is the so-called carrier scattering time. The intrinsic carrier concentration is related to temperature by $n_i = n_0 \times \exp(-E_G/2kT)$, where n_0 is a constant. In the extrinsic regime ($n_e \gg n_i$), at temperatures such that $kT \gg E_d$ (so that the impurities are all ionized), the conduction (valence) carrier concentration $n_e \rightarrow n_d$ (or n_a), the donor (or acceptor) concentration.

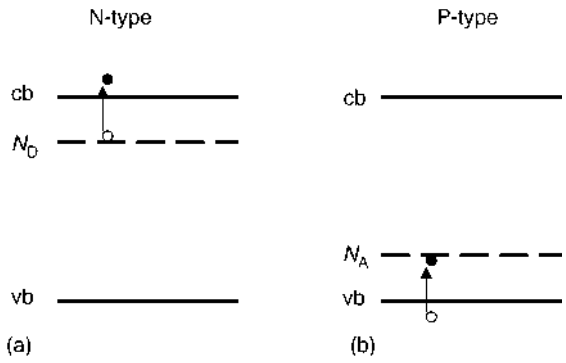

Figure 2 Doping in semiconductors. In n-type material (a), carriers are excited thermally from the donor states to the conduction band where they can contribute to the conductivity. In p-type material (b), electrons are effectively 'captured' by the acceptor states, leaving behind holes in the valence band which can also contribute to conduction. N_D and N_A are the concentration of donor and acceptor impurities, respectively.

Infrared Absorption – Interband Optical Transitions, Excitons

Apart from thermal promotion of electrons and holes via interband (valence to conduction band) transitions responsible for intrinsic conductivity, it is possible to directly excite such transitions optically by subjecting the sample to radiation with $\hbar\omega > E_G$. If the lowest conduction band minimum is situated vertically above the valence band maximum (usually at $k = 0$), then the optical transition process is referred to as a direct transition. If the lowest conduction band minimum is situated at some other point in k-space than that of the valence band maximum, the process is referred to as an indirect transition; this requires the participation of optical phonons (i.e., quantized lattice vibrations) to satisfy momentum conservation, and is therefore a weaker (higher order) process. The semiconductors are respectively referred to as Direct Gap and Indirect Gap Semiconductors. Simplified energy band diagrams for direct and indirect transitions are shown in Figure 3.

The optical spectrum corresponding to the onset of interband electric dipole transitions is referred to as the absorption edge. The most widely used indirect gap materials are the group IV semiconductors Si and Ge. The best known direct gap materials are the III–V's (InSb, GaAs, GaP, etc.), II–VI's (HgTe, CdTe, ZnSe, etc.) and IV–VI's (lead salts PbTe, PnSe, SnTe) and alloy combinations of all these. An important additional effect, which becomes progressively stronger as the energy gap becomes larger in a semiconductor is the Coulomb interaction between the electrons and holes created by interband optical transitions. The electron is raised into an excited state but remains bound in a hydrogen-like orbit

around the positive hole creating a two particle excitation called an exciton. The composite particle can then move throughout the crystal. In an ideal crystal, therefore, the absorption spectrum should consist of a series of discrete lines corresponding to the ground and excited exciton states, followed at somewhat high energies by the continuum associated with interband electronic transitions. The energy levels of the exciton spectrum measured from the beginning of the ionization continuum will be given by the Bohr formula:

$$E_{ex} = \frac{\mu}{\varepsilon^2}\frac{13.6}{n^2}\text{eV} \qquad [21]$$

where μ is the reduced mass given by: $1/\mu = 1/m_e + 1/m_h$ (m_e, m_h are electron and hole effective masses). A schematic diagram of the absorption edge is shown in Figure 4 (dashed line – in the absence of excitons; solid line – in the presence of excitons).

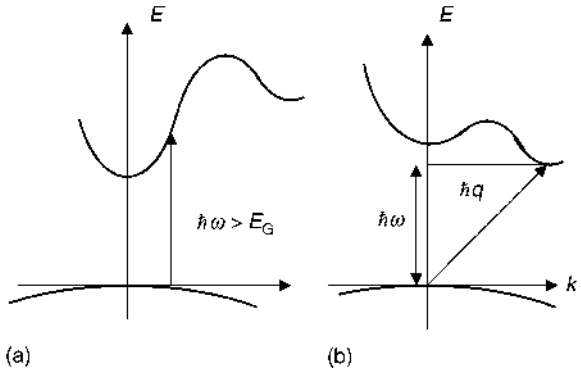

Figure 3 (a) Direct and (b) indirect gap semiconductors. Direct gap transitions involve only the absorption or emission of a single photon and are therefore very nearly 'vertical' transitions in *k*-space. On the other hand, indirect transitions have a 'two-stage' nature, involving the participation of both a photon and an optical phonon that carries sufficient momentum $\hbar q$ to allow the conduction band electron created in the absorption process to have a large non-zero wavevector.

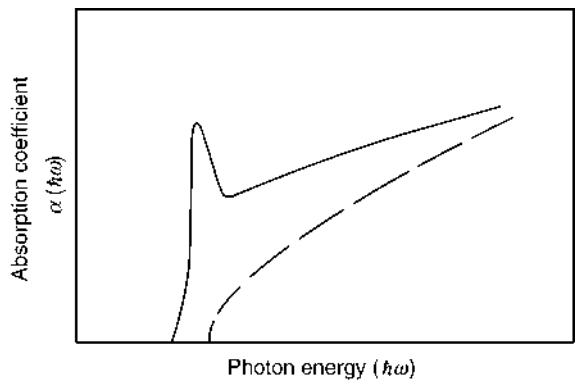

Figure 4 Absorption edge spectra without (dashed curve) and with (solid curve) the excitonic effect.

Conduction and Valence Band Parameters in Typical Semiconductors

As has been described earlier in real semiconductors the conduction and valence bands retain the symmetry of the parent atomic states from which they derive. For example, the conduction and valence bands of silicon can be regarded as deriving from 3s and 3p electrons, and it is generally true that for all group IV, III–V and II–VI semiconductors commonly used in electronic applications the conduction band has s-like symmetry and the valence band has p-like symmetry. This means that the valence band has the additional complication that it is triply degenerate (c.f. the p_x, p_y, p_z components of the hydrogen atom p-orbital); and further, in the presence of spin-orbit coupling (familiar from atomic physics) part of this degeneracy is raised, resulting in doubly degenerate (at $k = 0$) so-called heavy hole (hh) and light hole (lh) bands, and a band split-off by the spin-orbit splitting, Δ:

$$E_{hh} = -\hbar^2k^2/2m_{hh}\ ;\ E_{lh} = -\hbar^2k^2/2m_{lh}\ ;$$
$$E_{so} = -\Delta - \hbar^2k^2/2m_{so} \qquad [22]$$

The resulting conduction and valence bands near $k = 0$ are shown in Figure 5.

Low Dimensional Systems: Quantum Wells

So far we have been concerned almost entirely with the bulk properties of crystalline solids. Such properties can normally be specified in terms of coefficients that are independent of the shape and size of a particular specimen. Thus, for example the electrical

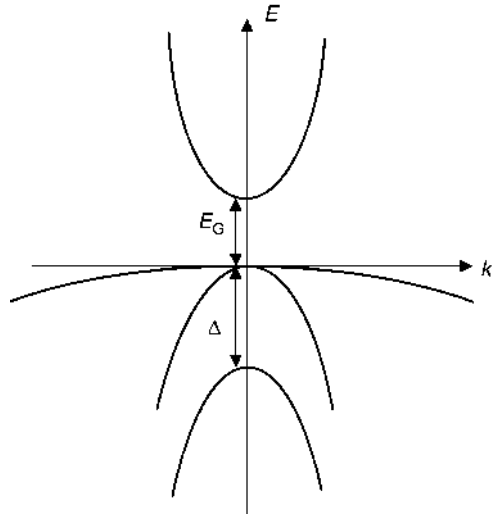

Figure 5 Energy bands (s-like conduction and p-like valence bands) of a direct gap semiconductor near the Brillouin zone center.

conductivity and resistivity, or the optical absorption coefficient and refractive index are properties only of the material and not of the sample dimensions. When one or more dimensions of a solid are reduced sufficiently, the properties are no longer given by these bulk coefficients. The sample is then described as a low-dimensional system (LDS). Such systems form the basis of many optoelectronic devices. LD systems can be classified according to the number of dimensions that are small: thin films (and heterojunctions) are two-dimensional since only the thickness is reduced, and these form so-called quantum wells; fine wires are one-dimensional since two dimensions are reduced, forming quantum wires; dots or specks are zero-dimensional, forming quantum dots.

From an optoelectronic point of view, the most important example is that of electrons confined to a semiconductor film or junction; this system is called the two-dimensional electron gas (2-DEG) and the structure a quantum well. Thus, it is possible to create a narrow electric potential junction region so that the electrons are restricted in 1D (say z) according to the quantum mechanics of a particle in an infinite potential well (described earlier), but 'free' within the xy junction plane.

See also

Semiconductor Materials: Band Structure Engineering; Quantum Dots. **Semiconductor Physics:** Band Structure and Optical Properties; Impurities and Defects; Infrared Lattice Properties; Quantum Wells and GaAs-Based Structures; Recombination Processes.

Further Reading

Hook JR and Hall HE (1991) *Solid State Physics*, 2nd edn. Chichester, UK: John Wiley and Sons, Inc.

Kittel C (1996) *Introduction to Solid State Physics*, 7th edn. Chichester, UK: John Wiley and Sons, Inc.

Seeger K (1989) *Semiconductor Physics: an Introduction*, 4th edn. Berlin: Springer.

Band Structure and Optical Properties

W Zawadzki, Polish Academy of Sciences, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The purpose of this article is to describe the influence of the electronic energy band structure of semiconductors on their optical properties. By the band structure one understands the energy–wavevector dispersion relation $\varepsilon(\mathbf{k})$ for the electron energy bands and the relative positions of band maxima and minima in the Brillouin zone.

The main problem in calculating the band structure of a semiconductor is that of the periodic potential of the lattice in which the electrons move. It is, in fact, a self-consistent problem since the moving electrons partly screen the potential. Different approximations have been developed to deal with the question, and in all of them the symmetry of the lattice is of fundamental importance. Thus, in the so-called empty-lattice approximation one deals exclusively with the symmetry and periodicity of the lattice without specifying the potential. This can give qualitative ordering and symmetry of the bands but no quantitative results. In the opposite limit one uses the tight binding approximation, in which the bands are constructed from the atomic states of separate atoms. This method gives quite a good description of lower (valence) bands but poor results for higher (conduction) bands. A powerful way to describe real energy bands is obtained by various forms of pseudopotential calculation, where one approximates the potential near actual atoms by simple parametrized forms and then adjusts the parameters to fit experimental (mostly optical) data. The pseudopotential methods give a good overall description of various bands in the entire Brillouin zone, but they do not provide sufficiently precise results near the band extrema. A semi-empirical way to serve the latter purpose is called $\mathbf{k}\cdot\mathbf{p}$ theory which we describe below. Various methods of band structure calculation are reviewed in books on solid state physics as outlined in the suggestions for Further Reading at the end of this article.

In the majority of important semiconductor materials (Si, Ge, many III–V and II–VI compounds) the maxima of the valence bands are at the center of the Brillouin zone (Γ point). The minima of the conduction bands in Si and Ge are not at the Γ point. This means that the fundamental optical interband transitions in these materials are indirect (in the ε–$\mathbf{k}$ diagram), i.e., they require phonon assistance. On the other hand, the minima of conduction bands in important III–V and II–VI compounds are at the Γ point, so that the interband optical transitions are direct and they do not require phonon assistance.

Both systems are utilized in opto-electronic devices, particularly for detectors. However, for emitters (light-emitting diodes and lasers) we are mostly concerned with the second case, on which we concentrate here.

Bloch States

The Bloch theorem states that if the potential $V(\mathbf{r})$ in which the electron moves is periodic with the periodicity of the lattice, then the solutions $\Psi(\mathbf{r})$ of the Schrödinger wave equation

$$\left[\frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r})\right]\Psi(\mathbf{r}) = \varepsilon\Psi(\mathbf{r}) \qquad [1]$$

are of the form $\Psi(\mathbf{r}) = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})u_k(\mathbf{r})$, where $u_\mathrm{k}(\mathbf{r})$ is periodic with the periodicity of the direct lattice, and $\mathbf{k}$ is the wavevector ($\hbar\mathbf{k}$ is the pseudomomentum). The proof of this theorem can be found, for example, in undergraduate textbooks on solid state physics.

k·p Theory

The underlying idea of semi-empirical $\mathbf{k}\cdot\mathbf{p}$ theory is to describe the energy bands in the vicinity of a given point of the Brillouin zone (usually near a band extremum). Symmetry properties are used to minimize the number of unknown band parameters (effective mass, energy gap, etc.) which are then determined by experiment. The initial Schrödinger equation for an electron in the periodic potential $V(\mathbf{r})$ of the crystal lattice reads

$$[\mathbf{p}^2/2m_0 + V(r) + H_\mathrm{so}]\Psi = \varepsilon\Psi \qquad [2]$$

where m_0 is the free electron mass and H_so is the spin–orbit interaction. Since $H_\mathrm{so}(\mathbf{r})$ is also periodic with the lattice periodicity, the solutions of eqn [2] are given by the Bloch states above.

We use $\mathbf{k}\cdot\mathbf{p}$ perturbation theory, as follows, to describe the energy bands near $\mathbf{k} = 0$ (i.e., the so-called Γ point of the Brillouin zone). Since we are interested in small values of $\mathbf{k}$, we expand our unknown cell periodic functions, $u_{l\mathbf{k}}(\mathbf{r})$, in terms of the set of Γ-point functions, $u_{lo}(\mathbf{r})$, whose symmetry we know. One defines a representation $\Phi_{l\mathbf{k}} = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})u_{lo}(r)$, where $u_{lo}(\mathbf{r})$ is the periodic $\mathbf{k}$-independent function satisfying the equation

$$[\mathbf{p}^2/2m_0 + V + \mathrm{H}_\mathrm{so}]u_{lo} = \varepsilon_{lo}u_{lo} \qquad [3]$$

whose solution we know in terms of the band-edge energies, ε_{lo}. This representation is sometimes referred to as the Luttinger and Kohn (LK) representation. The u_{lo} functions are in general mixtures of spin up and spin down states because of the spin–orbit interaction. One looks for the solutions of eqn [2] in the form

$$\Phi_{n\mathbf{k}} = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})\sum_l c_l^{(n)}(\mathbf{k})u_{lo}(\mathbf{r}) \qquad [4]$$

where the sum is over all bands and $c_l^{(n)}(\mathbf{k})$ are the coefficients to be determined. Inserting the above form into eqn [2], performing the operation $\mathbf{p}^2$ (i.e., operating twice on the Bloch product function with the operator $\mathrm{p} = -\mathrm{i}\hbar\nabla$) and using the property [3] one eliminates the unknown potential $V(\mathbf{r})$. Multiplying the resulting equation from the left by $u_{l'o}^*$, integrating over the unit cell Ω and using the orthonormality of u_{lo}, one finally obtains

$$\sum_l\left[(\varepsilon_{lo} - \varepsilon')\delta_{l'l} + \frac{\hbar}{m_0}\mathbf{k}\cdot\mathbf{p}_{l'l}\right]c_l^{(n)}(\mathbf{k}) = 0 \qquad [5]$$

for $l' = 1, 2, 3, \ldots$ Here $\delta_{l'l}$ is the Kronecker delta function, $\varepsilon' = \varepsilon - \hbar^2k^2/2m_0$, and $\mathbf{p}_{l'l}$ is the so-called matrix element of the momentum given by $\mathbf{p}_{l'l} = (1/\Omega)\int_\Omega u_{l'o}^*\mathbf{p}u_{lo}\,\mathrm{d}\mathbf{r}$. The index l' runs over all energy bands.

Equation [5] formulates the famous $\mathbf{k}\cdot\mathbf{p}$ theory. The nondiagonal part $\mathbf{k}\cdot\mathbf{p}_{l'l}$ can be eliminated by the perturbation procedure, and the method is referred to as $\mathbf{k}\cdot\mathbf{p}$ perturbation theory. In the second order of perturbation the bands are separated and in each band the carriers are described by a dispersion relation $\varepsilon_l = (\hbar^2/2)\mathbf{k}(\widehat{1/m_{lo}^*})\mathbf{k}$ where $(\widehat{1/m_{lo}^*})$ is an inverse effective mass tensor at the band edge in question. The inverse mass is generally a 3×3 tensor, but for cubic materials in the center of the Brillouin zone it is a scalar, so that $\varepsilon_l(\mathrm{k}) = \hbar^2\mathrm{k}^2/2\mathrm{m}_{lo}^*$ (where $1/\mathrm{m}_{lo}^* \equiv 1/\mathrm{m}_0 + (2/\mathrm{m}_0^2)\Sigma_{l\neq l'}(\mathrm{p}_{ll'}^2)/[\varepsilon_{l0} - \varepsilon_{l'0}]$). We then say that the band is spherical and parabolic. The second order of perturbation is a good approximation if the energy ε (as counted from a nondegenerate band edge) is small compared to forbidden energy gaps. For the triply degenerate p-like valence band one has to do degenerate perturbation theory and treat the bands together as a 3×3 matrix equation.

In the same approximation the wavefunction for a given band is: $\Psi_{l\mathbf{k}}(\mathbf{r}) = \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})\,u_{lo}(\mathbf{r})$. In the absence of spin–orbit coupling, H_so, the u_{lo} function would just have the symmetry of the parent band (i.e., labeled S for the s-like conduction band, and X, Y or Z for the triply degenerate p-like valence band – each with a single spin up or spin down component). In the presence of H_so, the triply degenerate valence band states become mixtures of X, Y and Z with mixed spin character and part of the degeneracy is raised.

In atomic notation the presence of H_{so} has transformed the basis from the (l,s) to the (J,m_J) representation. This results in the well-known light and heavy hole bands (degenerate at $k=0$) and the spin–orbit split-off band.

Under the influence of a radiation field, of vector potential $\mathbf{A}'$ and frequency ω, the optical transition probability for an electron to be raised from state Ψ_i (initial) to Ψ_f (final) is proportional to the square of $M=(e/mc)\langle\Psi_f|\mathbf{A}'\cdot\mathbf{p}|\Psi_i\rangle$ i.e., it is determined by the same momentum matrix element, p_{fi}, which governs the effective mass. For interband transitions one immediately gets the selection rule $\Delta k=0$ (i.e., direct transitions). The polarization selection rules are determined by the $u_{lo}(r)$ components of the initial and final states, through the momentum matrix elements.

Narrow-Gap Semiconductors

Semiconductors such as InSb and $Hg_{1-x}Cd_xTe$ (with $x<0.3$) have small energy gaps and are referred to as narrow gap semiconductors (NGSs). They are commonly used in opto-electronic applications which may be accessing energies in the conduction or valence band which are a significant fraction of the energy gap. Under these circumstances it is not valid to cut off the expansion to order k^2, and one has to deal with so-called nonparabolic bands. Thus, in NGSs (or indeed in any situation where the energy becomes an appreciable fraction of ε_g) the above procedure is not valid and an alternative approximation is in order. Following semidegenerate perturbation theory one includes exactly in eqn [5] a finite number of bands (near each other in energy) and neglects distant bands. This is referred to as the Kane model, and the energy bands near $k=0$ are shown for InSb-type III–V compounds and for HgTe-type II–VI compounds in Figure 1. In this case we include only the conduction and valence band in our set of states (eqn [5]). Including spin and degeneracy of the Γ_8 symmetry the set [5] then has eight equations. They can be solved analytically and the resulting energies are given by

$$\varepsilon'(\varepsilon'+\varepsilon'_g)(\varepsilon'+\varepsilon'_g+\Delta)-\left(\varepsilon'+\varepsilon'_g+\frac{2\Delta}{3}\right)\kappa^2k^2=0 \quad [6]$$

where $\varepsilon'=\varepsilon-\hbar^2k^2/2m_0$. We see that for this restricted set of states the only unknown parameters are the energy gap, ε_g, the spin–orbit splitting energy, Δ, and $\kappa=-i(\hbar/m_0)\langle S|p_z|Z\rangle$ (note that this is the only non-vanishing matrix element of the momentum). The fourth energy is $\varepsilon'=-\varepsilon_g$. All four energies are spin degenerate. In NGSs one may neglect the free electron term, i.e., set $\varepsilon'\approx\varepsilon$. For $\varepsilon\ll\varepsilon_g+(2/3)\Delta$ the resulting quadratic equation for the conduction and the light-hole bands is

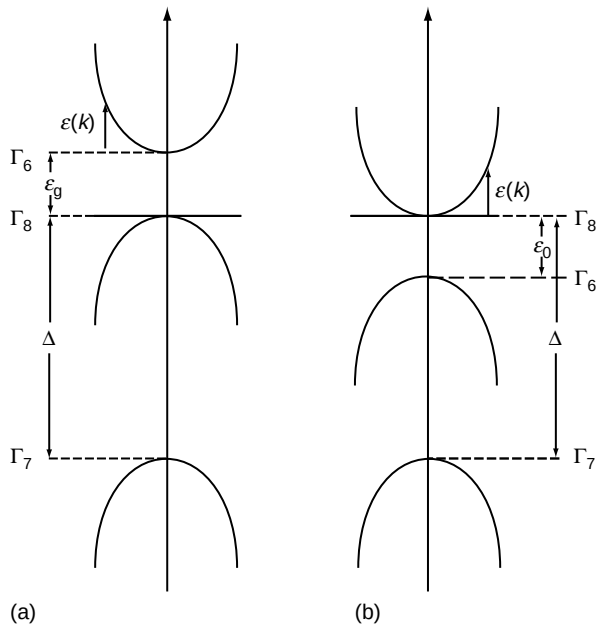

Figure 1 Three-level model of band structure near the Γ point of the Brillouin zone: (a) InSb-type semiconductors; (b) HgTe-type semiconductors.

$$\varepsilon(1+\varepsilon/\varepsilon_g)=\hbar^2k^2/2m_0^* \quad [7]$$

where $1/m_0^*=(4\kappa^2/3\hbar^2\varepsilon_g)(\varepsilon_g+2\Delta/3)/(\varepsilon_g+\Delta)$ defines the effective mass at the band edge. The root for the heavy holes is $\varepsilon=-\varepsilon_g$ i.e., this band is not correctly described within the three-level model. The bands given by eqn [6] are spherical but nonparabolic. For $\varepsilon\ll\varepsilon_g$ one recovers the standard dispersion, $\varepsilon=\hbar^2k^2/2m_0^*$.

For HgTe-type materials, setting the zero of energy at the Γ_8 edge and replacing ε_g by $-\varepsilon_0$ (cf. Figure 1), one obtains

$$\varepsilon'(\varepsilon'+\varepsilon_0)(\varepsilon'+\Delta)-\left(\varepsilon'+\frac{2\Delta}{3}\right)\kappa^2k^2=0 \quad [8]$$

For $\varepsilon'\ll(2/3)\Delta$ the dispersion [7] is valid with $1/m_0^*=4\kappa^2/3\hbar^2\varepsilon_0$. In this case the conduction and the heavy hole bands ($\varepsilon\equiv0$) are degenerate at $k=0$, i.e., the thermal gap is zero.

The important property of the above model is that it holds also in the limit of $\varepsilon_g\to0$ (i.e., for mixed $Hg_{1-x}Cd_xTe$ crystals). The effective mass in eqn [7] tends to zero, but the dispersion described by eqn [6] is valid and it gives $\varepsilon=(2/3)\kappa k$, i.e., a linear band.

The electron and hole wavefunctions in NGSs resulting from semidegenerate perturbation theory are given by eqn [4], in which the sum runs over

all the included states. Thus they involve truly **k**-dependent amplitudes of the Bloch states, cf. eqn [1]. In addition, these functions are mixtures of spin-up and spin-down states. These features have important consequences for optical and dc transport phenomena.

If the conduction band minimum is not at the Γ point (Ge and Si), there are at least two different matrix elements of momentum and the resulting band is ellipsoidal.

Effective Mass

The nonparabolic dispersions $\varepsilon(k)$ in NGSs require generalizations of important relations. The momentum mass $\widehat{m}^*$ is introduced as a connection between the psuedomomentum and the velocity: $\hbar\mathbf{k} = \widehat{m}^*\mathbf{v}$. Since $v_i = \delta\varepsilon/\delta(\hbar k_i)$, one obtains for a spherical band

$$\frac{1}{m^*} = \frac{1}{\hbar^2 k}\frac{\mathrm{d}\varepsilon}{\mathrm{d}k} \qquad [9]$$

The above mass is a scalar, but it depends on the energy (or momentum). This mass is more useful than the one usually defined in textbooks, relating force to acceleration. The latter is not a scalar even for a spherical band. However, for the standard parabolic band, $\varepsilon = \hbar^2 k^2/2m_0^*$, both masses are the same and are equal to m_0^*. The mass [9] is related via velocity to the electric current, so that it enters naturally in transport phenomena (also free carrier optics). In particular, this mass enters the definition of mobility $\mu = q\tau/m^*$, where q is the charge and τ is the relaxation time. It enters the density of states (see below). Finally, it is the momentum mass [9] which is measured in cyclotron resonance.

For the dispersion [7] (the so-called two-level model), the mass [9] in the conduction band is

$$m^*(\varepsilon) = m_0^*\left(1 + \frac{2\varepsilon}{\varepsilon_g}\right) \qquad [10]$$

Again, for $\varepsilon/\varepsilon_g \ll 1$ the energy dependence of the mass is negligible, which represents the standard regime. Figures 2 and 3 illustrate energy dependences of the electron masses in InSb and HgSe, as measured by various optical and transport methods. The increase of the mass with energy is very strong in both cases.

Density of States

The density of states in energy space is calculated beginning with pseudomomentum space. For the spherical band one obtains

$$\rho(\varepsilon) = \frac{k^2}{\pi^2}\frac{\mathrm{d}k}{\mathrm{d}\varepsilon} = \frac{m^* k}{\pi^2\hbar^2} \qquad [11]$$

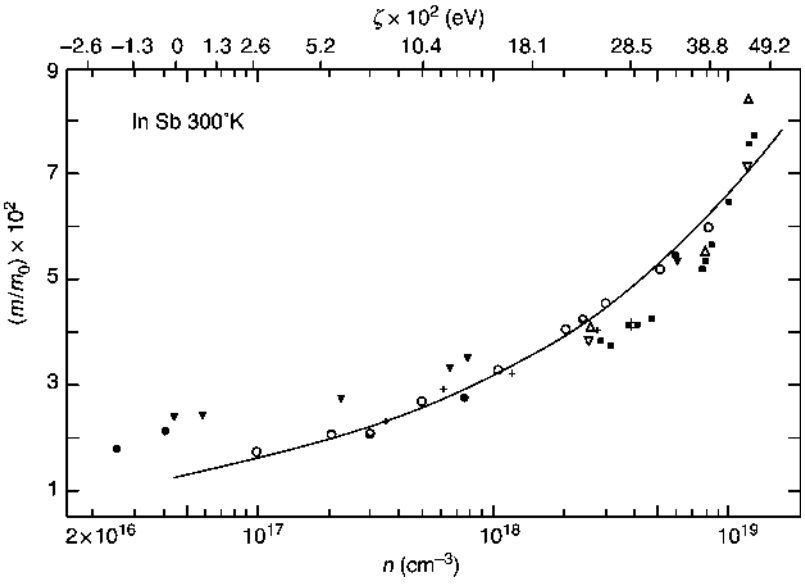

Figure 2 Electron effective mass in InSb at room temperature versus free electron concentration. The solid line, calculated for the two-level model, represents the mass value at the Fermi energy, as indicated on the upper abscissa. The experimental results of various authors are also indicated. Reproduced from Zawadzki W (1974) Electron transport phenomena in Sm all-gap semiconductors. *Advances in Physics* 23: 435–512.

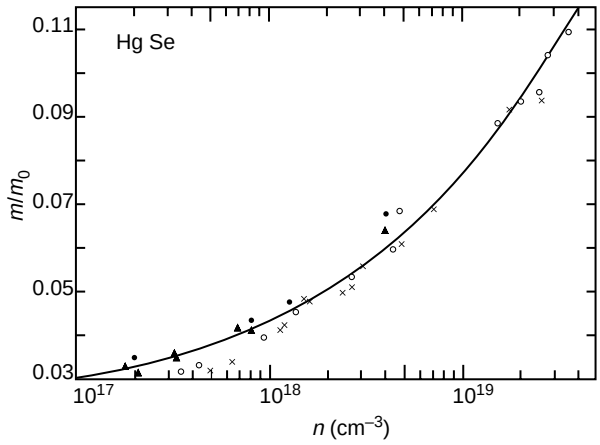

Figure 3 Electron effective mass in zero-gap material HgSe versus free electron concentration. The solid line is calculated for the three-level model. Experimental results of various authors are also indicated. Reproduced from Zawadzki W (1974) Electron transport phenomena in Sm all-gap semiconductors. *Advances in Physics* 23: 435–512.

The momentum mass [9] enters naturally into this important quantity. For the parabolic case $m^* = \text{const.}$ and $k \sim \varepsilon^{1/2}$, so that the standard result is recovered. In the general case one should use eqn [10] for $m^*(\varepsilon)$ and determine k from eqns [6] or [7].

Electron–photon Interaction

The electron–photon interaction can be introduced into the theory on two levels. If one replaces in the initial equation [2] the momentum $\mathbf{p}$ by $\mathbf{p} + (e/c)\mathbf{A}'$, the interaction term is obtained in the scalar form $H_{\text{int}} = (e/m_0c)\mathbf{A}'\cdot\mathbf{p}$, where $\mathbf{A}'$ is the vector potential of radiation. The wavefunctions to be used for the calculation of matrix elements are of the form [4], i.e., they include the periodic LK amplitudes. In fact, the matrix elements of momentum for optical transitions are identical to those of the $\mathbf{k}\cdot\mathbf{p}$ theory, as pointed out earlier. Since (as noted above) the wavefunctions for all bands have the same form [4], differing only by the coefficients $c_l^{(n)}(\mathbf{k})$, the matrix elements for interband and intraband optical transitions have the same form.

The electron–photon interaction can also be introduced directly to the $\mathbf{k}\cdot\mathbf{p}$ theory of eqn [5] replacing $\hbar\mathbf{k}$ by $\hbar\mathbf{k} + (e/c)\mathbf{A}'$. If the free electron term $\hbar^2k^2/2m_0$ is neglected, the interaction matrix involving $\mathbf{A}'$ terms is a number matrix. In this procedure the wavefunctions for initial and final states are simply columns and rows of $c_l^{(n)}$ coefficients and LK amplitudes no longer come into play. Here the $\mathbf{p}_{l'l}$ elements of $\mathbf{k}\cdot\mathbf{p}$ theory determine directly the electron–photon interaction. Both procedures described above give the same results.

Quantum Wells

If charge carriers are placed in a quantum well described by a potential $U(z)$, the motion in the z-direction is quantized while the motion in the x–y plane remains free. One takes the same LK basis (cf. eqn [3]), but the general form of solution is given by

$$\Psi = \sum_l f_l(\boldsymbol{r})u_{l0}(\boldsymbol{r}) \qquad [12]$$

since, in the presence of an external potential, $\mathbf{k}$ is not a good quantum number and the envelope functions $f_l(\mathbf{r})$ are unknown. If $U(z)$ is a slowly varying function within the unit cell the potential appears only on the diagonal of the set [5] and $\hbar\mathbf{k}$ is replaced by $\mathbf{p}$. One obtains

$$\sum_l \left[\left(\varepsilon_{l0} + \frac{\mathbf{p}^2}{2m_0} + U - \varepsilon\right)\delta_{l'l} + \frac{\mathbf{p}_{l'l}\cdot\mathbf{p}}{m_0}\right] f_l(\mathbf{r}) = 0 \qquad [13]$$

One can now eliminate the nondiagonal terms applying the standard Luttinger and Kohn scheme of second-order perturbation theory and arrive at the decoupled band equations with effective masses. This results in the standard quantization due to electric (and magnetic, if present) fields, and wavefunctions are of the form $\Psi_{lm} = f_m u_{l0}$ for the m-th sub-band of the l-th band.

For NGSs this procedure is a poor approximation and one should proceed as before including exactly a finite number of bands (cf. eqn [12]). This results in a set of coupled differential equations for the envelope functions $f_l(\mathbf{r})$. In some important cases one can find the eigenenergies without finding the functions by using the semiclassical approximation (the so-called WKB method).

We shall omit the calculations with coupled differential equations by using a semiclassical procedure also in another sense. Namely, we shall generalize the nonparabolic dispersion relation [7] to include the presence of external fields. To this end we observe that, including the potential $U(\mathbf{r})$ on the diagonal of eqn [13], one replaces $-\varepsilon$ by $-\varepsilon + U$. Further, if a magnetic field is introduced to the $\mathbf{k}\cdot\mathbf{p}$ theory one replaces $\hbar\mathbf{k}$ by $\mathbf{P} = \mathbf{p} + (e/c)\mathbf{A}$, where $\mathbf{A}$ is the vector potential of the magnetic field. Thus the semiclassical equation resulting from eqn [7] is

$$\left[\frac{\mathbf{p}^2}{2m_0^*} - (\varepsilon - U)\left(1 + \frac{\varepsilon - U}{\varepsilon_g}\right)\right]\Psi = 0 \qquad [14]$$

It can be seen that for $\varepsilon - U \ll \varepsilon_g$ one recovers the standard one-band approximation mentioned above. However, below we consider the general situation described by eqn [14].

Let us first consider the case of no magnetic field, i.e., $\mathbf{P} = \mathbf{p}$. For $U(\mathbf{r}) = U(z)$ one can separate the variables by looking for solutions in the form $\Psi = \exp(\mathrm{i}k_x x + \mathrm{i}k_y y)\Phi(z)$. One can now transform eqn [14] into

$$\left[\frac{1}{2m_0^*}p_z^2 - \frac{1}{\varepsilon_g}(\varepsilon - \varepsilon_\perp - U)(\varepsilon_g + \varepsilon + \varepsilon_\perp - U)\right]\Phi(z) = 0 \qquad [15]$$

where

$$\varepsilon_\perp = \frac{\varepsilon_g}{2} + \left[\left(\frac{\varepsilon_g}{2}\right)^2 + \varepsilon_g \frac{\hbar^2 k_\perp^2}{2m_0^*}\right]^{1/2} \qquad [16]$$

in which $k_\perp^2 = k_x^2 + k_y^2$. Equation (15) is suitable for the WKB semiclassical quantization with which one determines the eigenenergies ε by one integration.

For a magnetic field applied along the z-direction (parallel electric and magnetic fields) the effects of both fields are separated and the form of eqn [15] is valid, with $\hbar^2 k_\perp^2/2m_0^*$ in eqn [18] replaced by $\hbar\omega_c(n + 1/2) \pm (1/2)|g_0^*|\mu_B B$. Here $\omega_c = eB/m_0^* c$ is the cyclotron frequency, $n = 0, 1, 2, \ldots$ is the Landau quantum number, g_0^* is the spin splitting factor, and μ_B is the Bohr magneton. This results in nonequal spacing of Landau levels for a fixed value of B and in nonlinear B dependence of the Landau energies as functions of B (similarly but not identically to the bulk case). These features are illustrated by the cyclotron resonance experiment and theory shown in Figure 4.

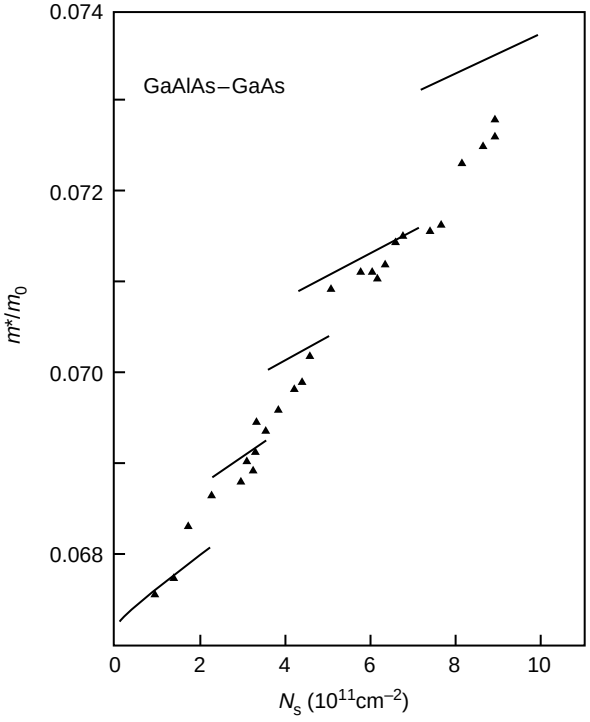

Figure 4 Electron effective mass versus two-dimensional electron density in GaAs/GaAlAs heterostructures, as measured by far infrared cyclotron resonance. The solid lines are calculated using the effective two-level model for GaAs. Reproduced from Zawadzki W, *et al.* (1994) Cyclotron emission study of electron masses in GaAs-GaAlAs heterostructures. *Semiconductor Science and Technology* 9: 320–328.

Selection Rules for Intersub-band Optical Transitions

For the one-band effective mass approximation the electron–photon interaction is $H_{\mathrm{int}} \sim \mathbf{A}'\cdot\mathbf{p}$ (cf. above), and the wavefunctions for the initial and final sub-bands are $\Psi_m = \exp(\mathrm{i}k_x x + \mathrm{i}k_y y)\Phi_m(z)$, in which the $\Phi_m(z)$ must be orthogonal to each other since they describe different energies. It is then clear that the matrix element $\langle fn|\mathbf{A}'\cdot\mathbf{p}|in\rangle$ is finite only if $\mathbf{A}'$ is polarized along the z-direction (also called the growth direction) since for $\mathbf{A}'$ parallel to x or y the integral over z is $\langle\Phi_{m'}|\Phi_m\rangle = 0$ giving a vanishing matrix element. Experimentally this is a serious problem for spectroscopic applications involving intraband (i.e., intersub-band) transitions in n-type semiconductors, since it is not possible to access these using normal incidence radiation.

This selection rule is somewhat relaxed for p-type semiconductors as a result of the mixed spin states of the complex valence band. In addition the narrow-gap band structure introduces interesting new possibilities into the selection rules. The wavefunctions have the form [12] and for the light polarizations A'_x or A'_y the momentum operators p_x or p_y act also on the periodic amplitudes $u_{l0}(\mathbf{r})$, which leads to a nonvanishing transition probability for intersub-band transitions. This has indeed been observed in narrow-gap materials.

See also

Magneto-Optics: Interband Magnetoabsorption, Cyclotron Resonance, Spin Flip Raman Scattering. **Semiconductor Materials:** Mercury Cadmium Telluride; Quantum Dots. **Semiconductor Physics:** Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Ashcroft NW and Mermin ND (1976) *Solid State Physics*. Holt, Rinehart and Winston.

Kane EO (1980) Band structure of narrow gap semiconductors. In: Zawadzki W (ed.) *Narrow Gap Semiconductors. Physics and Applications*, pp. 13–31. Berlin: Springer.

Luttinger JM and Kohn W (1955) Motion of electrons and holes in perturbed periodic fields. *Physical Review* 97: 869–883.
Zawadzki W (1974) Electron transport phenomena in Sm all-gap semiconductors. *Advances in Physics* 23: 435–512.
Zawadzki W (1984) Electric and magnetic quantisation of two-dimensional systems – elementary theory. In: Bauer G, *et al.* (eds) *Two-Dimensional Systems, Heterostructures, and Superlattices*, pp. 2–11. Berlin: Springer.
Zawadzki W (1991) Intraband and interband magneto-optical transitions in semiconductors. In: Landwehr G and Rashba EI (eds) *Landau Level Spectroscopy*, pp. 483–512. Amsterdam: North-Holland.

Excitons

I Galbraith, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

In semiconductors electron–hole pairs can lower their energy by correlating their motion in an exciton state. Such states dominate the bandedge optical properties of direct-gap semiconductors. The resultant absorption peaks are affected by quantum confinement, electric, and magnetic fields and carrier scattering.

Introduction

In the simplest, single-particle picture, the ground state of a semiconductor consists of a full valence band and an empty conduction band. The lowest lying excitation above the ground state in such a scheme is produced by excitation of one electron from the valence band to the conduction band. To achieve this optically requires a photon energy greater than the bandgap, E_g. However in many semiconductor materials this simple picture is unable to explain the observed absorption spectrum in the neighborhood of the fundamental bandgap. The origin of the discrepancy lies in the neglect of the Coulomb interaction between the negatively charged excited electron and the hole which is left in the valence band.

The term exciton refers to a two-particle excitation, consisting of a bound electron and hole. Such excitations dominate the bandedge optical spectra of most direct gap semiconductors, especially at low temperatures. In particular there exists a series of hydrogen-like bound states lying below the bandedge. These states are bound by energies

$$E_b^{3D} = \frac{\mu e^4}{8\mathrm{h}^2\varepsilon_r^2\varepsilon_0^2}\frac{1}{n^2} \qquad [1]$$

where $\mu = m_e m_h/(m_e + m_h)$, is the reduced mass, $m_{e(h)}$ the electron (hole) effective mass, e is the electronic charge, h is Planck's constant, ε_0 is the permittivity of free space, ε_r the relative permittivity of the material and $n = 1, 2, 3, \ldots\infty$ is the principal quantum number.

The binding energy of the lowest lying exciton state varies considerably from one semiconductor to another being, e.g., 0.5 meV in InSb and over 60 meV in Zn0. Similarly the spatial extent of the 1 s electron-hole relative wave function or effective Bohr radius is given by

$$a_0^{3D} = \frac{\hbar}{\sqrt{2\mu E_b^{3D}}} \qquad [2]$$

and ranges from 750 μm in InSb to ≈2 nm in Zn0. Typically, in semiconductors the exciton spatial extent is much larger than the lattice constant of around 0.6 nm, and such excitons are classed Wannier excitons. At the opposite limit, which occurs in many molecular materials, the exciton is localized around a particular atomic site and is termed a Frenkel exciton. Wannier excitons are characterized by the hydrogen-like quantum numbers of their relative motion and an overall center of mass momentum which describes the wave-like motion of the bound entity through the crystal.

Another type of exciton is often referred to as a bound exciton. This consists of an electron–hole pair bound to a neutral impurity center. Such excitons are localized around the impurity and cannot move through the crystal in the same way as a free exciton.

Optical Properties

Excitons manifest themselves primarily as strong modifications to the bandedge optical properties of semiconductors. In particular the bound states give rise to discrete absorption lines at lower energy than the bandgap energy. Several peaks corresponding to the different hydrogen-like bound states

(1s, 2s, etc.) can be seen provided the lines are sufficiently narrow. A theoretical calculation of this is shown in Figure 1.

Associated with these absorption changes are concomitant changes in the refractive index. Contributions to the spectral width of the absorption peaks are characterized as either homogeneous or inhomogeneous. Homogeneous broadening refers to broadening associated with the finite lifetime of a particular state, e.g., due to phonon scattering. Inhomogeneous broadening is due to non-uniformity in the system, e.g., in a sample having a spatially non-uniform strain field.

There is also enhancement of the absorption at photon energies above the bandgap due to the residual influence of the Coulomb attraction on the electron-hole scattering states. Although not bound, electrons and holes in these states still have an enhanced probability of being found close together. This is called the Sommerfeld Enhancement Factor (see Figure 1).

It is important to note that although there may be an excitonic peak in the absorption spectrum this does not imply that excitons form a stable population in such a sample. Often, especially at room temperature, the lifetime of an optically generated exciton is determined by the scattering time with optical phonons which may be very short (<1 ps). In this case the excitons are quickly ionized and the quasi-thermal equilibrium which is formed is dominated by essentially plasma-like behavior. In photoluminescence experiments, where carriers are generated high in the band and the spectrum of the resulting emitted luminescence is measured, peaks associated with the free exciton are often observed and taken as a signature of the presence of excitons. This view has been challenged recently by theoretical calculations which show that a photoluminescence peak at the exciton energy can be also be produced from an uncorrelated plasma if proper account of the Coulomb interaction is taken. This issue remains controversial.

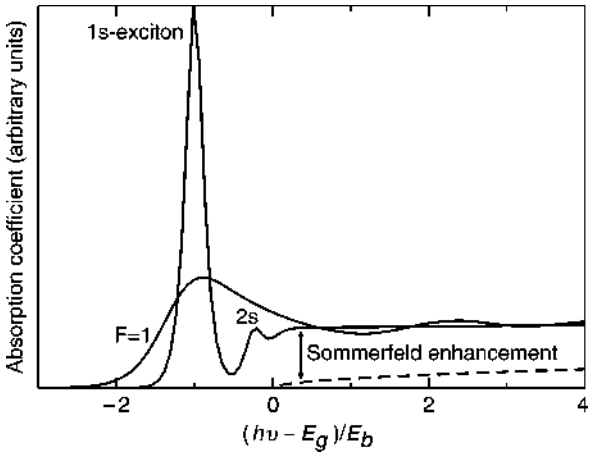

Figure 1 Calculated bulk absorption spectra including (solid-line) and neglecting (dashed line) the Coulomb interaction. The Sommerfeld enhancement is seen as photon energies above the bandgap. Also shown is the spectrum with an applied electric field of one exciton Rydberg per Bohr radius.

In bulk samples, experiments such as absorption and reflectivity are complicated by the existence of polariton effects. An exciton-polariton is a quantum mixture of the propagating photon inside the semiconductor sample and the material excitation. Where the photon dispersion and the exciton energy dispersion meet there is an anti-crossing and this is manifested in, for example, the appearance of two lines in the reflectivity spectrum.

It was hoped during the 1970s that laser action in wide bandgap material such as ZnSe would be possible based on excitonic lasing. In this process a quasi-equilibrium exciton population would form the injected excitation and stimulated recombination would occur with the associated emission of a scattering partner which would take up the necessary momentum to ensure overall momentum conservation. This leads to a light emission wavelength below the absorption edge which is advantageous for minimizing losses. In fact, this process has only been seen at low temperature for scattering with LO-phonons, electrons and other excitons. Each mechanism has its own characteristic lasing threshold and temperature dependence.

Influence of Quantum Confinement

The advent of Molecular Beam Epitaxy enabled the growth of semiconductor layers of thickness similar to the electronic de Broglie wavelength. By sandwiching a low bandgap semiconductor between two high bandgap layers, one can make quantum well structures in which the electronic motion is effectively confined to the plane of the layer. In such structures the excitonic properties are also radically altered by this confinement. An example of the absorption spectrum of a ZnCdSe quantum well embedded in ZnSe is shown in Figure 2. The heavy and light hole excitons correspond to different spin quantum numbers for the electron and hole making up the exciton.

The exciton binding energy is enhanced by a factor of 4 compared to the bulk case and the spatial extent of the 1s wavefunction is also reduced by half:

$$E_b^{2D} = \frac{\mu e^4}{8\mathrm{h}^2\varepsilon_r^2\varepsilon_0^2}\frac{1}{(n-\frac{1}{2})^2} \quad n = 1, 2, 3\ldots \qquad [3]$$

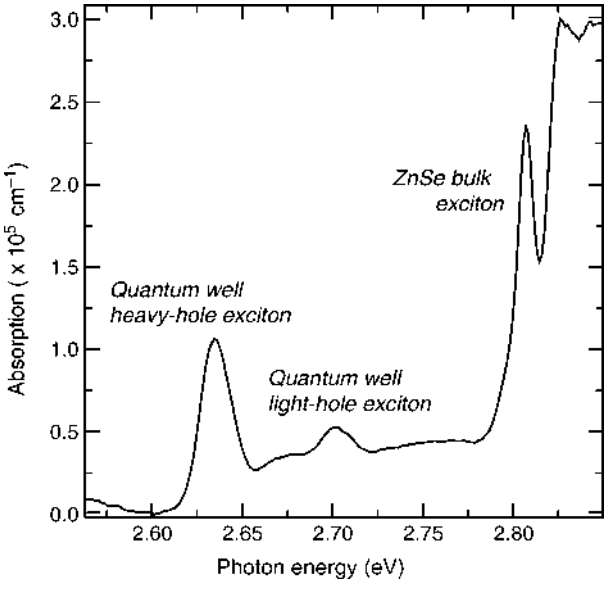

Figure 2 Heavy and light hole exciton absorption spectrum for ZnCdSe quantum wells embedded between ZnSe barriers. Also shown is the bulk exciton of the ZnSe barriers.

This enhancement can be understood since the particles are confined to lie in the plane and are hence closer together on average than in the three-dimensional case. In practice, a four-fold enhancement in the binding energy is never observed since the quantum wells are neither infinitely deep nor infinitely thin. Both these limitations lead to a reduction in the actual enhancement from 4 to about 2, depending on the structure. For example, the binding energy in a realistic quantum well corresponds to the bulk value of the well material for very wide wells and to the bulk value of the barrier material for very narrow wells. In between the value is enhanced by the quantum confinement. The maximum enhancement typically occurs when the well width is around half the bulk Bohr radius for the well material.

Accurate calculation of the exciton binding energy is complicated by the fact that the quantum confinement influences the band structure, making it strongly anisotropic and splitting heavy and light holes. A complete inclusion of these effects is at the limit of current computational power due to the need to calculate the electron-hole interaction matrix elements. Approximate schemes, including variational approaches, have proved successful in fitting a variety of samples, including GaAs and ZnSe based quantum wells. The limitation to such calculations is often the requirement for accurate values of the material parameters such as effective masses and energy band offsets.

The Sommerfeld enhancement of the absorption into the continuum states of quantum wells leads to a doubling of the absorption at the bandedge. This enhancement reduces only slowly as the photon energy is increased.

If multiple quantum wells with very thin barrier layers are used, the electronic states in the neighboring wells couple together to form a superlattice. The superlattice dispersion in the growth direction behaves like a heavy mass and the resulting exciton is three-dimensional but strongly anisotropic.

More extreme quantum confinement is possible making quantum wires, by etching or selective growth of quantum well samples. In these quasi-one dimensional structures, as in quantum wells, the exciton binding energy is again increased further. The Coulomb problem in one-dimension is pathological in that the ground-state binding energy diverges logarithmically, and higher states are pairwise degenerate. However, accounting for the finite cross-section of realistic wires eliminates this problem and recovers a finite binding energy. One consequence of this is that there is no simple enhancement factor which corresponds to the four-fold enhancement in going from bulk to two-dimensional excitons. The density of states in one-dimensional systems diverges at the bandedge, implying the existence of an infinite absorption in a perfect quantum wire. However, the influence of excitonic correlations cancels this divergence, resulting in an enhanced, but not divergent, bandedge absorption feature.

The ultimate limit in quantum confinement can be achieved in quantum dots, where the carriers are confined in all three directions. In this case the term exciton should be used with some care as there are no bandedge states to compare with. The discrete energy levels dictated by the confinement are affected by the Coulomb interaction and the energy is renormalized – in some cases substantially. The absorption spectrum should consist of a series of sharp atomic-like transitions corresponding to s-s, p-p, d-d,... transitions and indeed luminescence spectra obtained from single quantum dots do display such fine structure. Even in the best controlled material systems, such as InAs quantum dots, the spectra are strongly inhomogeneously broadened due to variations in the dot size, shape, and alloy composition.

In large quantum wells, where the width is larger than the exciton Bohr radius the center-of-mass part of the wavefunction can be confined by the barriers. This leads to a quantized motion for the exciton as a whole, rather than for the individual electron and hole separately. One implication of this is that for

such excitons the center-of-mass momentum is no longer a good quantum number.

The above effects are a controlled case of disorder induced effects in semiconductors. This is an area which is only now beginning to be understood. The presence of local potential fluctuations will influence the energy spectrum of excitons in that vicinity. Clearly both the depth and spatial extent of these fluctuations will be important – very short range (less than Bohr radius), shallow potential fluctuations will not have any influence on the exciton. While deep potentials can lead to the center of mass confinement above.

Influence of Electric and Magnetic Fields

Just as an externally applied electric field affects the absorption edge via the Franz–Keldysh Effect, broadening the edge itself and inducing oscillations in the continuum part of the spectrum, so, in the case of an excitonic bandedge, there is substantial modification due to the presence of the field. At low fields, when the Coulomb interaction is strong compared to the potential drop across the exciton diameter, the exciton will be little affected by the applied field. At higher fields the exciton will become gradually more polarized by the field until the exciton is field ionized. This ionization process is manifested in the field broadening of the exciton line (see Figure 1).

Of more technological importance is the influence of an electric field applied perpendicular to the plane of a quantum well. The field forces the electron and hole in opposite directions leading to a reduction in the exciton binding energy. This would lead to a blue shift in the exciton absorption peak were it not for an even larger reduction in the single particle electron and hole energy levels within the quantum wells. Overall a strong red-shift of the excitonic absorption peak is observed with increasing field. This is accompanied by a reduction in the oscillator strength due to the reduced electron-hole overlap. This is termed the Quantum Confined Stark Effect and is the basis for the operation of some semiconductor electro-optic modulators. An electric field applied in the plane of the quantum well gives rise to Franz–Keldysh oscillations analogous to the bulk case.

The influence of magnetic fields on excitons is more subtle and richer than the electric field. This is because the magnetic field seeks to induce a circular cyclotron orbit motion on the carriers, but this is influenced by their mutual Coulomb interaction. Two regimes exist where either the cyclotron energy:

$$\hbar\omega_c = \hbar(eB/\mu) \qquad [4]$$

or the exciton binding energy dominates. In the former strong field case, one can treat the Coulomb interaction as a small perturbation on the electron and hole states in the presence of the magnetic field. In exciton states, which have a finite magnetic moment at zero field, we find a linear Zeeman shift in the magneto-exciton energy. In the other case we must use the magnetic field as the perturbation which produces a mixing of the zero-field exciton states. For example, for the ground state the magnetic field induces an admixture of the p-like excited state with the s-like ground state. The total angular momentum of the mixed state is proportional to B and as the energy of a magnetic dipole in a magnetic field is also proportional to B the shift of the magneto-exciton is proportional to B^2. This distinction between linear and quadratic shifts has been used to identify the nature of carrier populations (excitons or unbound free carriers) in a variety of samples. Clearly there exists an intermediate regime where both the magnetic and Coulomb energies are comparable and in this case the precise energy shifts need to be evaluated numerically.

In quantum well samples the orientation of the magnetic field with respect to the confinement direction is crucial in determining the physics. When the field is perpendicular to the confinement plane the cyclotron orbits exist as before and essentially the same phenomenon as in the bulk emerge. For other orientations, the behavior is much more complex and beyond the scope of this article.

Exciton Scattering

An electron and hole in an excitonic bound state execute a correlated motion which can be disturbed by scattering of either partner. This may lead to the ionization of the exciton and the destruction of the correlation or, alternatively, it may change the center of mass momentum of the exciton as a whole. Almost all of the important scattering mechanisms for excitons arise from the charged nature of the electron and hole. Via the Coulomb interaction, excitons can scatter with lattice phonons, other excitons, free carriers, and impurities. Each of these scattering processes has its own regime of dominance, dependent on, for example, temperature or material quality.

When the scattering rate with other particles is much larger than the recombination rate for electrons

and holes, a quasi-equilibrium state is reached. The detailed nature of such an interacting electron/hole plasma remains a a long-standing open question. The reasons for this can be traced to interplay between the intrinsic Coulomb interactions within the plasma, which give rise to both bound and scattering states, and Pauli exclusion arising from the fermionic nature of the electrons and holes. This leads to a complex phase diagram which encompasses, e.g., electron-hole droplets, the ionized electron-hole plasma, biexciton phase, and the exciton gas which, some suggest, may undergo Bose–Einstein condensation. The parameters of this phase diagram are carrier density, temperature, and the semiconductor material parameters. Interest in this essentially fundamental question has remained high, stimulated by the stream of technological benefits that even partial answers have brought.

In this context it is worth mentioning that there have been two theoretical approaches to exciton physics which each have their advantages and problems. One approach is to treat excitons as bosonic quasi-particles and derive results from Hamiltonians formulated using bosonic operators. This approach has the appeal of simplicity and produces acceptable results in the low-density regime. It does, however, omit a key feature of the constituent particles making up an exciton namely the fermionic nature of electrons and holes. A more rigorous approach based around electron and hole operators has been followed but this is considerably more complex and numerically involved.

Excitonic Molecules

Just as two hydrogen atoms can lower their total energy by forming a bound hydrogen molecule, so two excitons can make a bound complex which has lower energy than the two isolated excitons. Such complexes are called biexcitons and are mostly important in high quality material and at low temperatures. Their binding energy is less than that of the exciton by a factor of about 0.1–0.3, depending on the ratio of electron to hole effective masses.

Three particle complexes have also been observed consisting of two electrons bound to a hole. These are termed trions or charged excitons. Their binding energy lies intermediate between excitons and biexcitons.

See also

Semiconductor Materials: GaAs Based Compounds; Large Gap II–VI Semiconductors; Type-II Quantum Wells and Superlattices. **Semiconductor Physics:** Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Bastard G (1998) Wave Mechanics Applied to Semiconductor Heterostructures. New York: Halstead John Wiley.

Haug H and Koch SW (1993) *Quantum Theory of the Optical and Electronic Properties of Semiconductors*. Singapore: World Scientific.

Klingshirn CF (1997) *Semiconductor Optics*. Berlin: Springer.

Rashba EI and Sturge MD (eds) (1982) *Excitons*. Amsterdam: North Holland.

Schmitt-Rink S, Chemla DS and Miller DAB (1989) Linear and nonlinear optical properties of semiconductor quantum wells. *Advances in Physics* 38: 89–188.

Ueta M, Kanzaki H, Kobayashi K, Toyozawa Y and Hanamura E (1986) *Excitonic Processes in Solids*. Berlin: Springer.

Yu PY and Cardona M (1999) *Fundamentals of Semiconductors*. Berlin: Springer.

Impurities and Defects

K A Prior, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

A simple definition of a defect is any feature that disturbs the periodicity of the perfect lattice. This definition is sufficiently broad to encompass a whole range of different entities, which are normally divided into three categories based on their dimensionality, or extent in real space.

0D defects are those which have an extent of the order of only a few atomic spacings at most in any direction. They include all cases where one, or at most a few, atoms are removed or displaced, or are replaced by atoms of a different type.

1D defects are linear features within the crystal that extend over many atomic dimensions. The most

important type of 1D defect is the dislocation and in a single crystal it is possible for a dislocation to extend from one side of the crystal to the other.

2D defects include a variety of different features that occur primarily in one plane. Any junction between two different types of crystal in an epitaxially grown layer (a heterointerface) fits this description, as of course does the sample surface. There are also crystal defects which exist in two dimensions of which the most important is the stacking fault.

Finally, there are also 3D defects. This category includes voids and inclusions, which are both major perturbations in a crystal structure. In most pure crystals grown by modern techniques these are eliminated and they will not be discussed further.

In the sections below the main types of defect will be discussed according to the scheme described above. Despite the substantial research effort to understand the basic properties of many semiconductors, and of course the huge effort devoted to producing semiconductor devices, it remains true that the defect structure of most semiconductors is still poorly understood, with the exceptions of silicon and gallium arsenide. This is especially true in the case of the nitride semiconductors with the wurtzite structure which have recently been the subject of much commercial interest. In the following sections examples will mainly be taken from gallium arsenide.

Lastly, it is important to note that defects of different dimensionality interact. Several possible interactions are discussed in the last section.

0D Defects

Dopants

There are two basic types of 0D defects to consider: chemical defects, or impurities, and native or point defects caused by missing or misplaced atoms. Table 1 gives the major types of 0D defect found in compound semiconductors together with the symbol used in the common notation for point defects and impurities.

Impurity atoms are those which are introduced either accidentally during the production of the bulk crystal or epitaxial layer, or deliberately by the crystal grower in precisely controlled amounts to modify some property of the material. Usually, this means that the material added is a dopant, and has been added to create a surplus of either free electrons or holes in the semiconductor. Typically, the amounts of dopants added are between 10^{-6} and 10^{-3} mole fraction, giving doping levels between $10^{22}\ \mathrm{m}^{-3}$ and $10^{25}\ \mathrm{m}^{-3}$.

Table 1 Impurities and point defects in a compound *AB* with the zincblende lattice structure

Symbol	*Description*
X_A^n, X_B^n	Impurity atom, *X*, on *A* or *B* site with charge *n*
X_i^n	Interstitial impurity atom with charge *n*
V_A^n, V_B^n	Vacancy on *A* or *B* site with charge *n*
A_i^n, B_i^n	Interstitial *A* or *B* atom with charge *n*
A_B, B_A	*A* atom on *B* site or *B* atom on *A* site (antisite)

In general, any atom added to a material will introduce a localized state into the band structure of the semiconductor. This state may rarely occur in either the conduction or valence band, but is more likely to be located in the bandgap. If the localized energy level is within a few kT of the valance or conduction band edge, then the state is an electron acceptor or donor. If it is much further from the band edge then it is usually known as a deep level. In this case its main effect on the electrical properties will be to act as a trap for electrons or holes, and introduce noise into the current flow. Its effect on the optical properties of the material will be to reduce the radiative recombination rate for band-to-band transitions by acting as an alternate recombination pathway for either radiative or nonradiative transitions.

It is generally assumed that dopant atoms occur only on the site of the appropriate sublattice and are well behaved. While it is true that some elements are well behaved substitutional dopants, such as sulfur in gallium arsenide, which can be denoted S_{As}^{+} when acting as a singly charged donor, the behavior of other species is more complex. In many III–V compounds, silicon is an amphoteric dopant, occupying sites on both the metal and nonmetal sublattices which ionize as shown:

$$\mathrm{Si}_{\mathrm{Ga}}^{0} = \mathrm{Si}_{\mathrm{Ga}}^{+} + \mathrm{e}^{-}$$

$$\mathrm{Si}_{\mathrm{As}}^{0} = \mathrm{Si}_{\mathrm{As}}^{-} + \mathrm{h}^{+}$$

In molecular beam epitaxy (MBE) growth on GaAs(100) the ratio between the donor $\mathrm{Si}_{\mathrm{Ga}}^{+}$ and the acceptor $\mathrm{Si}_{\mathrm{As}}^{-}$ varies as the concentration of doping is increased. At low levels silicon is a donor, while at high concentrations the fraction which is an acceptor increases rapidly. The semiconductor is therefore compensated, effectively limiting the maximum concentration of free carriers.

Another type of behavior occurs with beryllium in III–V semiconductors and lithium in II–VI semiconductors. Here, the dopant atom is so small that it can easily be displaced from the lattice site. Thus, for example, in InGaAs, it is energetically favorable for the acceptor $\mathrm{Be}_{\mathrm{In}}^{-}$ atom to move from the lattice site to

an interstitial position:

$$Be_{In}^{-} = V_{In}^{-} + Be_{i}^{+} + e^{-}$$

The interstitial Be ion is a donor and causes some compensation, but even worse, is highly mobile. It can move a substantial fraction of a micron in an epitaxial device and in so doing completely destroys a carefully engineered abrupt p-n junction. It can be seen therefore that dopant interactions with the host lattice are important in determining device performance.

Native Defects

A large variety of native defect species are possible in semiconductors. Interstitial atoms and vacancies are found in both the diamond structure (silicon) and the zincblende structure (most III–V and II–VI semiconductors). In zincblende structures there is also the possibility of placing an atom on the wrong atomic site, or sublattice; for example in GaAs a Ga atom may be found on an As site. This type of defect is known as an antisite.

Bulk semiconductor crystals are grown under conditions close to equilibrium. In the long periods of time needed to grow the crystals, equilibrium concentrations of point defects are established. Epitaxial growth techniques (in particular molecular beam epitaxy, MBE and metalorganic chemical vapor deposition, MOCVD) operate under conditions far from equilibrium and the point defect concentrations may differ markedly from the equilibrium values. Some semiconductor processing operations can also increase point defect concentrations. Etching semiconductors with ion beams or energetic plasmas is a very efficient way of creating in the subsurface region high concentrations of point defects well above the equilibrium values.

It is possible to anneal crystals in the vapor of one of the constituent elements to alter the point defect concentrations. Interstitial and vacancy concentrations can be changed relatively rapidly over the course of several hours (interstitials) or days (vacancies) as these species are much more mobile than dopant atoms. Usually, annealing is done to reduce point defect concentrations, but on occasions, for example with the alloy HgCdTe, it is done to increase the concentration of a type of defect, in this case mercury vacancies which are used to dope the material.

Like dopant atoms, point defects can also introduce defect levels into the band structure of the semiconductor, and some defects, for example the gallium vacancy V_{Ga}^{n-}, exist in multiple charge states. In some materials, the native defects act either as dopants or as compensating traps, and are present in concentrations that can severely limit the maximum obtainable doping levels. The gallium vacancy, for example, is believed to limit the maximum doping level obtainable in n-type GaAs, and its concentration increases dramatically as the concentration of n-type dopant is increased.

It should be noted that the concentrations of the various types of defect are not independent of one another, and at any given growth (or annealing) temperature and applied partial pressures of the constituent elements, the equilibrium concentrations of the defects are fixed. Examples of the interactions between the types of defect may be seen by considering the displacement of a gallium atom from its lattice site to create a vacancy and an interstitial (the Frenkel reaction):

$$Ga_{Ga}^{0} = Ga_{i}^{0} + V_{Ga}^{0}$$

and the interaction of an arsenic interstitial with a gallium vacancy to create an arsenic antisite:

$$As_{i}^{0} + V_{Ga}^{0} = As_{Ga}^{0}$$

The relative importance of the types of defect varies between semiconductors. In compounds where one of the elements is very volatile, then high concentrations of vacancies on that particular sublattice are to be expected. Antisites occur only at very low concentrations in the ionic II–VI compounds, as the energy involved in placing an atom on the wrong sublattice is very high. However, in the more covalent III–V compounds, there is not such a large energy penalty, and it has been suggested that in GaAs the As_{Ga} antisite may be the most common point defect.

In addition to the simple defect types described above, there are other possibilities. One of these is that complexes of defects may form. The best known are the complexes that occur between dopant atoms and vacancies, for example the $[Te_{As}V_{Ga}]^{-}$ complex found in Te doped GaAs (**Figure 1**). In this material, partial compensation of the n-type dopant occurs with the formation of gallium vacancies, as described above:

$$\frac{1}{2}As_2(\nu) = V_{Ga}^{0} + As_{As}$$

$$V_{Ga}^{0} = V_{Ga}^{2-} + 2h^{+}$$

These vacancies are mobile and diffuse in the material. In the vicinity of Te_{As}^{+} they experience a strong Coulombic attraction and the complex forms when the vacancy and dopant atom are on nearest neighbor sites:

$$V_{Ga}^{2-} + Te_{As}^{+} = [V_{Ga}Te_{As}]^{-}$$

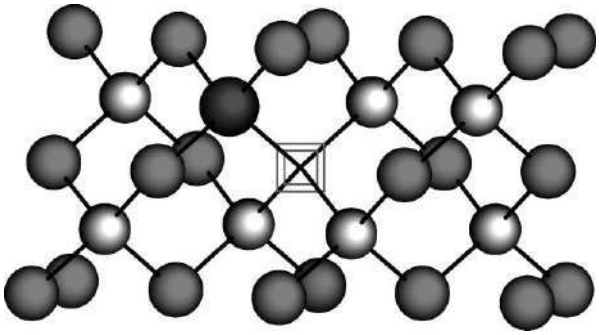

Figure 1 Diagram of the tellurium atom–vacancy complex in GaAs. Gallium and arsenic atoms are shown in mid-gray and white, respectively, while the tellurium atom is shown in dark gray.

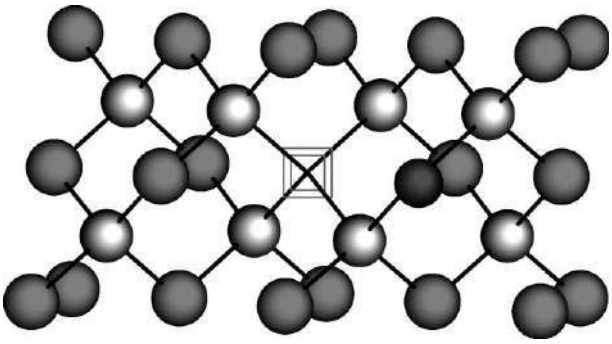

Figure 2 Diagram of the silicon atom–vacancy complex in GaAs. Gallium and arsenic atoms are shown in mid-gray and white, respectively, while the silicon atom is shown in dark gray.

A similar complex also forms in Si doped GaAs: $[Si_{Ga}V_{Ga}]^-$ (Figure 2), although here both species exist on the same sublattice, and consequently occur as next nearest neighbors in the complex.

In the description above it is assumed that all atoms which are located on a particular site are centered on that site. However, it is possible for a dopant atom to replace an atom of the host semiconductor, but not occupy the lattice site. *Ab initio* calculations of the bonding around the dopant atom suggest that it can be displaced away from one of its nearest neighbors towards the other three atoms (Figure 3). The dopant can be considered as having one bond weaker than the other three.

In the case of n- and p-type doped semiconductors the corresponding defects are known as DX and AX centers respectively. DX centers occur in many semiconductors but are particularly important in AlGaAs, where Si (or other n-type dopant) atoms form DX centers that partially compensate the material and reduce the electron mobility. However, irradiation with white light excites the DX center and converts it back to a normal dopant centered on the lattice site. In AlGaAs the DX center is the stable

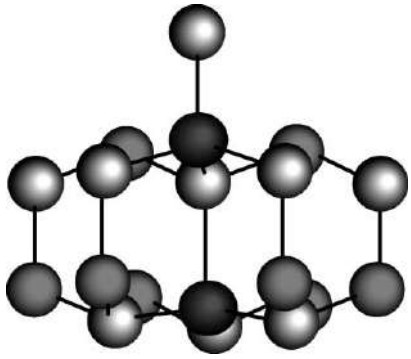

Figure 3 A GaAs crystal shown with [111] oriented vertically. Two Ga atoms have been substituted by silicon atoms which are shown in dark gray. One Si atom is on a substitutional site (top atom) and the other is displaced along [111] away from one of its nearest neighbors (bottom atom). This is one of the structures suggested for the DX center.

species and the dopant atom is a higher energy state. Conversion back to the DX center requires surmounting an energy barrier and thus the normal dopant is metastable. The irradiated semiconductor has a lower concentration of DX and a higher doping concentration than before, and consequently has a higher conductivity than the unirradiated material. This results in persistent photoconductivity in the irradiated material, which is a characteristic feature of semiconductors containing DX centers. At low temperatures, this effect can persist for many hours.

1D Defects

Within any crystal, there must exist a set of planes where the interplanar bonding is weakest. In both the diamond and zincblende crystal structures these are the {111} planes. When a crystal is subjected to a shear stress above its elastic limit, then the material accommodates that stress by changing shape, alternately breaking and reforming bonds across these planes where the bonding is weakest. The resulting movement is easily observable under a microscope in many materials and is known as slip or glide; the planes on which this happens are known as slip or glide planes.

Movement on a slip plane does not happen over the whole crystal at once, and there is always a boundary between the material in which slip has occurred and material where it has not. This boundary is the dislocation. A section through a piece of material where slip has occurred is shown in Figure 4. The curved line DD′ marks the dislocation and slip has occurred in the region marked DQD′. Because it marks a boundary, the dislocation may not end within the crystal and the only permitted ways that a dislocation may terminate are to finish at the surface of

the crystal (as shown) or to meet another dislocation, forming either a closed loop or a junction.

Across the slip plane, the material has moved by a certain amount, which is, typically, a lattice vector within the crystal. The direction of slip is known as the Burgers vector of the dislocation, and it is a constant for any one dislocation. The Burgers vector at any point may be determined by describing a simple loop in the crystal passing through adjacent lattice points lying in one plane. If a dislocation lies within the loop then the ends will not close, and the displacement needed to close the loop is the Burgers vector. This can be seen in Figure 4 on the left-hand face of the crystal around point D, where the Burgers vector is parallel to the line of the dislocation. It can also be seen on the right hand face of the crystal at point D′, where the Burgers vector, which has not changed direction, now lies perpendicular to the dislocation line. Dislocations where the Burgers vector is parallel or perpendicular to the dislocation line are known as screw and edge dislocations, respectively.

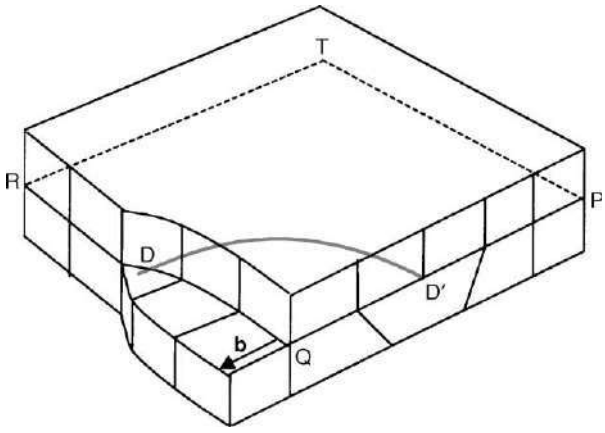

Figure 4 Diagram of a crystal showing a region where slip has occurred. For clarity only the lattice points are shown. The slip plane is PQRT and the slipped region is DQD′. The curve DD′ is the line of the dislocation and has a Burgers vector **b**.

Examples of edge and screw dislocations are shown in Figures 5 and 6, respectively. Edge and screw dislocations are considered to be 'pure' types of dislocations. If the Burgers vector makes any other angle to the dislocation line, the dislocation is referred to as a mixed dislocation. In semiconductors mixed dislocations are the most common type. There are many different ways of describing dislocations; however, the most common one is to give the angle between the dislocation line and the Burgers vector. In this system screw and edge dislocations are referred to as 0° and 90° dislocations.

At the core of the dislocation there is a small region, typically only one or two atoms wide, where the atoms have highly distorted bonding. Around this

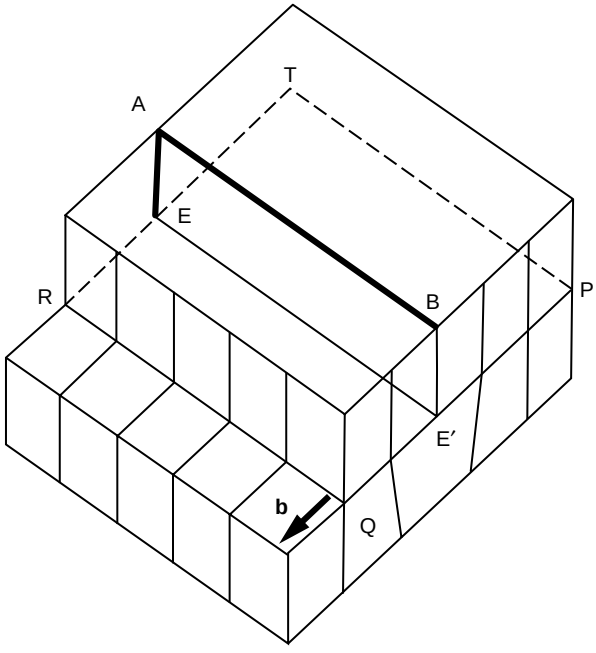

Figure 5 A schematic diagram of an edge dislocation. Only the lattice points are shown for clarity. The slip plane is PQRT and contains the Burgers vector denoted **b**. The line of the edge dislocation is EE′, which can be thought of as terminating a half-plane of atoms denoted ABEE′.

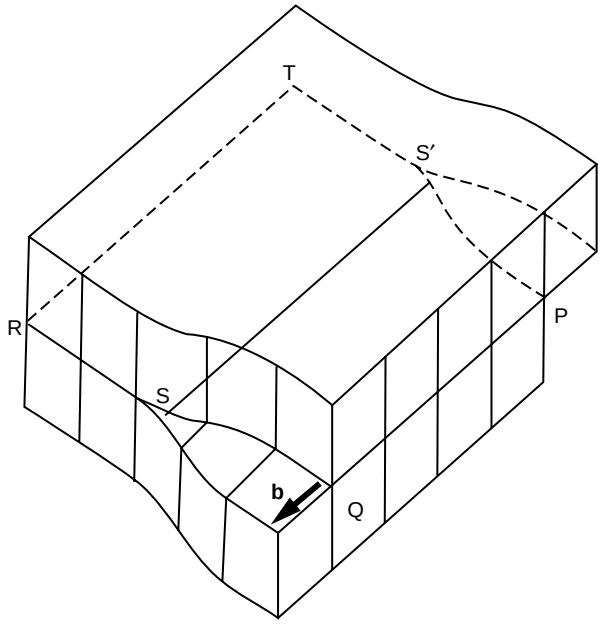

Figure 6 A schematic diagram of a screw dislocation. Only the lattice points are shown for clarity. The slip plane is PQRT and contains the Burgers vector denoted **b**. The line of the screw dislocation is SS′.

core the atoms adopt the normal (tetrahedral) bonding, although they are strained for many nanometers around the dislocation. The strain field around the dislocation means that low dislocation densities are very easy to detect using a technique such as X-ray topography, which images the strain introduced locally in a set of crystal planes by a dislocation. Using this technique, individual dislocations may be determined within crystals. Currently, it is possible to obtain silicon crystals which are essentially dislocation free, while the best GaAs crystals have dislocation densities of $<10^5$–$10^6\,\mathrm{m}^{-2}$. In other materials the densities may be 100–1000 times higher.

The highly distorted bonding at the core of the dislocation inhibits dislocation motion. Under the action of an external stress, dislocations normally move on their glide plane to relieve the stress and deform the material plastically. This is dislocation glide. In covalent semiconductors, such movement means breaking and reforming strong bonds, and does not occur in silicon below the crystal growth temperature. Dislocations are also immobile in GaAs at room temperature, although they can be induced to move in softer, more ionic II–VI semiconductors. Dislocation densities are therefore usually considered 'frozen in' at room temperature.

In addition to macroscopic shear forces, each dislocation experiences the stresses introduced by other dislocations and point defects present in the crystal, and is either attracted or repelled by them. At high densities, therefore, the motion of dislocations is inhibited by the presence of the surrounding defects – the dislocations are said to be pinned.

Generation and movement of dislocations occurs in most semiconductors easily at their growth temperature, and is important in the relief of strain introduced in growing a material on a substrate with a different lattice constant. Edge dislocations always introduce, or remove, extra planes of atoms (Figure 5), and so an array of dislocations can add or remove atoms to accommodate the strain between two semiconductor layers of different lattice constant. These are known as mismatch dislocations and they occur frequently in epitaxial growth when strain relief has occurred.

Typically, growth in MBE or MOCVD occurs on the (001) surface. Mismatch dislocations lie on the {111} planes, which intersect this plane, and at low densities occur in crystals regularly spaced along [110] and [1$\bar{1}$0] directions. The Burgers vectors are also of ⟨110⟩ type and are typically at 60° to the dislocation line. The edge component of this vector, projected on to the (001) plane, gives the amount of strain relief to be expected.

Despite the distorted bonding that exists around the dislocation core, there are thought to be remarkably few broken bonds there. An exception to this is thought to be when the dislocation line is not straight, but kinked, the atoms at the kinks having broken bonds. As a result, nonradiative transitions involving dislocations are not strong. They are important, however, during operation of lasers and LEDs. Dislocations can be seen in electroluminescence of light-emitting devices as dark lines. These so-called dark-line defects (DLDs) are the result of nonradiative recombination occurring at dislocations. Not only do they reduce the efficiency of the device, but the energy released by the recombination of the electron and hole is deposited at the dislocation core as thermal energy. Under these conditions, dislocation movement becomes easy and DLDs propagate, eventually destroying the device.

2D Defects

In a simple description of metallic structures, the hexagonal close packed structure is described as a sequence of hexagonal sheets with the stacking pattern *ABABABAB*..., while that of the face centered cubic structure is *ABCABCA*... In a similar manner, we can describe the wurtzite crystal structure by the stacking sequence *ABABABAB*..., although the layers are no longer close packed and the stacking units comprise one metal and one nonmetal atom. The corresponding stacking sequence for the zincblende structure then becomes *ABCABCA*... These structures are shown in Figures 7a and b, respectively. It can be seen that the two atom units are oriented perpendicular to the stacking sequence, and it is possible to rotate the structure about the vertically oriented bonds between any two given planes. A rotation of 180° would place all the atoms back in registry but on the wrong site. For example, in a zincblende crystal, a rotation after the third layer in the sequence *ABCABCA*... would result in *ABCBCAB*... This means that the top part of the crystal is related to the bottom half by a rotation of 180°, and the two parts are said to be twinned. The twin plane extends throughout the crystal, and is a two-dimensional defect. Both above and below the twin plane, the stacking sequence is the usual *ABC*... found in the zincblende structure; however, at the junction there is a region denoted *BCB*. This region is three planes of the wurtzite stacking sequence, and is an example of what is called a stacking fault. In zincblende crystals, stacking faults introduce small regions of the wurtzite structure, while the reverse happens in wurtzite crystals.

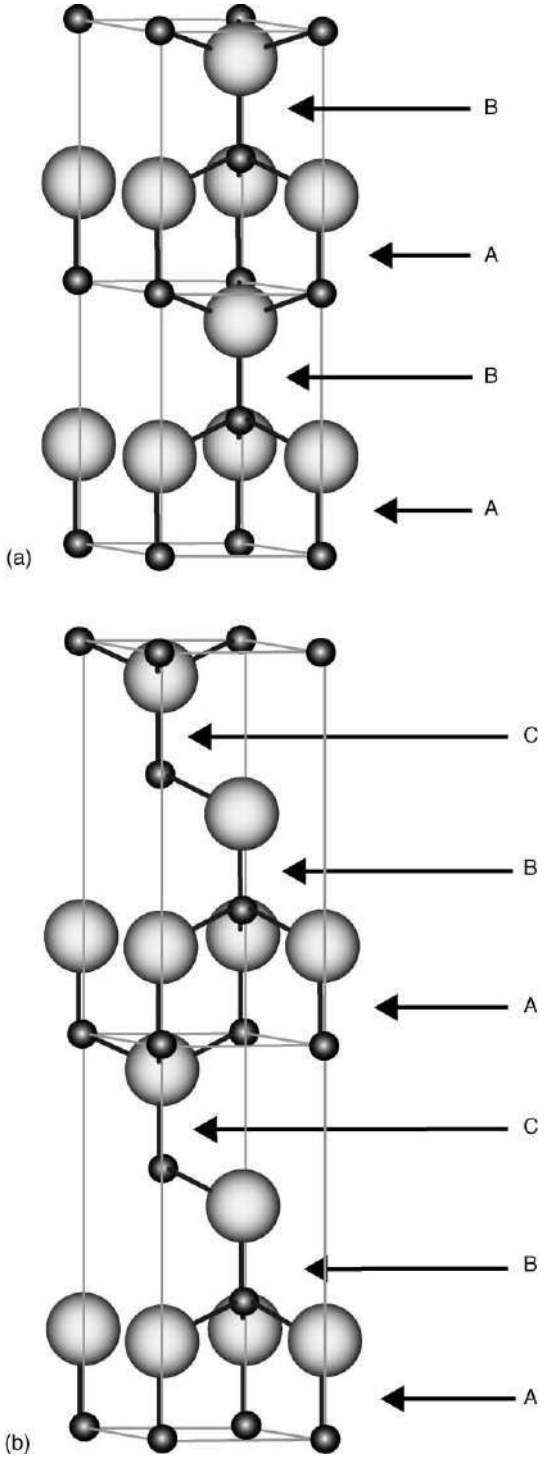

Figure 7 Diagrams of the stacking sequence of layers for two different semiconductor structures. In both cases the metal atoms are small and dark while the larger pale ones are the nonmetal atoms. The structures are: (a). The wurtzite crystal structure. The atoms which occupy the A, B and C sites are marked. Here the vertical direction in which the planes stack is the [0001] direction. (b). The zincblende crystal structure. The atoms which occupy the A and B sites are marked and the vertical direction corresponds to one of the ⟨111⟩ directions in the crystal.

The density of stacking faults in a material is primarily a function of the ease of placing the crystal planes in the wrong order. Under nonequilibrium growth conditions, this can occur easily for many different materials, but for slow growth near equilibrium, the prevalence of stacking faults is governed by the stacking fault energy. This is a measure of the wrong stacking sequence occurring and is also related to the energy of transformation between the zincblende and wurtzite crystal structures. For GaAs, the energy is high and stacking faults are rare, while for II–VI semiconductors the stacking fault energies are low and stacking faults are relatively common.

Typically, stacking faults are observed in TEM planar view micrographs, and on (001) oriented epitaxial samples are clearly visible as triangular regions lying on {111} planes, as described above, nucleated at the interface between the substrate and epitaxial layer.

While stacking faults are accidental introductions in a crystal structure, other 2D defects are introduced deliberately. Heterojunctions between different semiconductor alloys are an obvious example and are responsible for most of the useful properties of advanced semiconductor devices. However, by far the most important 2D crystal defect is the surface of the crystal. This creates a high density of broken bonds, each of which can introduce a defect level in the band structure. Surfaces usually lower their free energy by moving atoms from the bulk equilibrium positions to new sites and forming extra bonds in the plane of the surface. This removes some of the defect levels, but it is still possible to have surface defect densities as large as $10^{18}\ \mathrm{m}^{-2}$ which are associated with enhanced rates of nonradiative recombination of electrons and holes.

Interactions of Defects

Earlier it was assumed that the Burgers vector associated with a dislocation is a lattice vector. The energy of a dislocation per unit length is proportional to the magnitude of the Burgers vector squared:

$$E \propto b^2$$

In some crystals, however, it is possible for the displacement to be composed of two smaller steps, neither of which is a complete lattice vector:

$$\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$$

where $\mathbf{b}$ is a lattice vector and $\mathbf{b}_1$ and $\mathbf{b}_2$ are the smaller displacements. In this case, the total energy

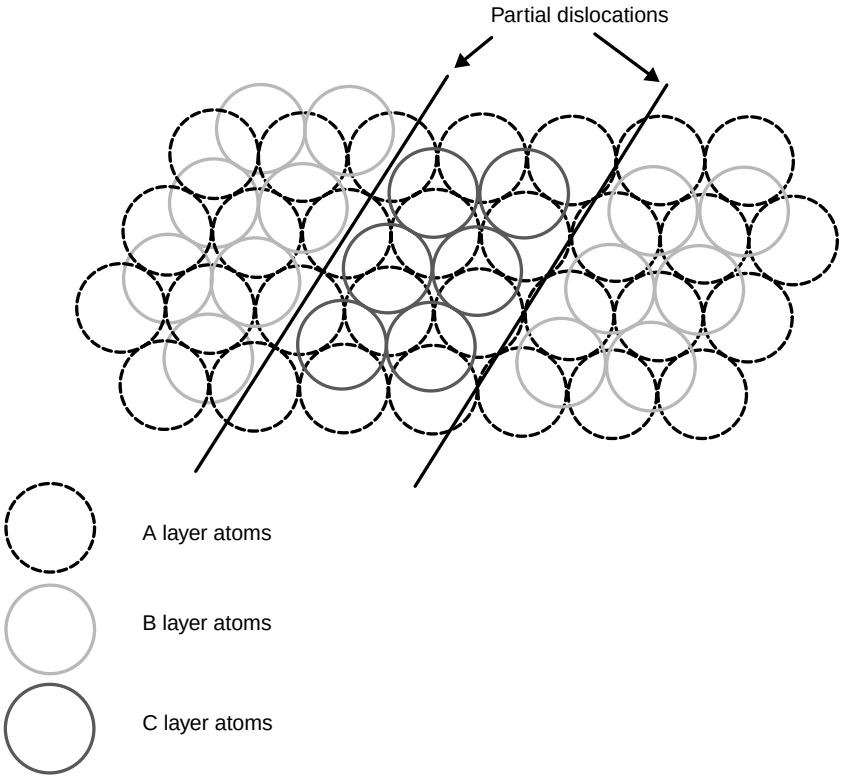

Figure 8 An example of a close packed structure where a region with a stacking fault lies between two partial dislocations.

can be reduced if the dislocation splits into two:

$$E_1 + E_2 \propto b_1^2 + b_2^2 < E$$

These dislocations are known as *partial* dislocations. A normal dislocation splits into two partial dislocations if it is energetically favorable to do so. Figure 8 shows an example in a close packed structure, which illustrates the point without the problem of introducing two different types of atoms. Two partial dislocations run through the top layer. The material lying between them has a different stacking sequence (AC) from the material around it (AB) and must contain a stacking fault. This shows that a 2D defect, such as a stacking fault, is bounded by two 1D defects, in this case partial dislocations.

There also exists an intimate relationship between point defects and dislocations, which may be seen by examining the edge dislocation shown in Figure 5. The edge dislocation may be considered as an extra half-plane of atoms, ABEE′ which terminates along line EE′. The last row of atoms in the plane can be removed from the sheet by one of two processes. Firstly, the atoms at the dislocation core can become interstitials and diffuse away from the dislocation. Alternatively, vacancies can diffuse towards the dislocation. In either case the result is the same; the dislocation line EE′ moves upwards towards AB. The opposite sequence of events (diffusion of interstitials towards the dislocation or vacancies away from it) is also possible and results in a movement of line EE′ downwards. Such processes are known as dislocation climb. Climb is not as common as glide, but it is important as the dislocation acts as a source or sink for point defects. This process contributes to the reduction in radiative efficiency in LEDs and lasers, particularly as the devices age.

See also

Semiconductor Physics: Outline of Basic Electronic Properties; Quantum Wells and GaAs-based Structures.

Further Reading

Beanland R, Kiely CJ and Pond RC (1992) Dislocations in heteroepitaxial films. In: Mahajan S (ed.) *Handbook on Semiconductors. Materials, Properties and Preparation*, vol. 3a, pp. 1149–1225. Amsterdam: North-Holland.

Fukuda M (1991) Reliability and Degradation of Semiconductor Lasers and LEDs. Boston, MA: Artech House.

Hirth JP and Lothe J (1982) Theory of Dislocations. New York: Wiley.

Hull D (1965) Introduction to Dislocations. Oxford, UK: Pergamon Press.

Hurle DTJ (1999) A comprehensive thermodynamic analysis of native point defect and dopant solubilities

in gallium arsenide. *Journal of Applied Physics* 85: 6957–7022.
Kelly A, Groves GW and Kidd P (2000) Crystallography and Crystal Defects, revised edn. Chichester, UK: Wiley.
Kossevich AM (1999) The Crystal Lattice. Berlin: Wiley-VCH.
Malloy KJ and Khachaturyan K (1993) DX and related defects in semiconductors. In: Willardson RK and Beer AC (eds) *Semiconductors and Semimetals*, pp. 235–291. New York: Academic Press.
Schubert EF (1993) Doping in III–V semiconductors. In: Ahmed H (ed.) *Cambridge Studies in Semiconductor Physics and Microelectronic Engineering*, vol. 1. Cambridge, UK: Cambridge University Press.
Skowronski M (1992) Impurity- and defect-centres in compound semiconductors. In: Mahajan S (ed.) *Handbook on Semiconductors. Materials, Properties and Preparation*, vol. 3a, pp. 1343–1398. Amsterdam: North-Holland.

Infrared Lattice Properties

T J Parker and S R P Smith, University of Essex, Colchester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The vibrations of atoms in solids were first studied theoretically at the beginning of the twentieth century to explain the observed temperature dependence of the heat capacity. Following earlier work by Einstein and Debye, Born and von Karman showed in 1912 that the vibrational properties of a linear chain of alternate light and heavy atoms connected by springs account very well for the observed thermal behavior of solids. It was shown many years later that, if the atoms are alternately positively and negatively charged, the model also gives a good description of the optical properties.

In this article the vibrational properties of a linear diatomic chain of atoms are first described. The model gives a good description of the basic vibrational properties of solids with two atoms per unit cell, such as Si, Ge, and the binary III–V (e.g., GaAs) and II–VI (e.g. CdTe) semiconductors, and explains the origin of optic and acoustic branches of the phonon dispersion curves. This leads naturally to an understanding of the basic far-infrared reflectivity of binary semiconductors. The model is then further developed to account for the effects of impurities and the optical properties of alloys.

The far-infrared properties of semiconductors are then reviewed. We begin with the group IV semiconductors Si and Ge in which there is no first-order interaction with infrared radiation, and explain the origins of second-order interactions, and the effects of disorder and impurities. Examples of the first-order reflection spectra of bulk single crystal binary semiconductors and their alloys are then presented. Localized vibrational modes and phonon combination bands are discussed next. We finish with a brief review of more exotic excitations that occur in layered semiconductors: confined optic modes, surface phonon polaritons, guided waves, and interface modes.

Background Theory of Phonons in Semiconductor Materials

Phonons in a Diatomic Linear Chain Lattice

Consider a line of atoms of alternating type A and B, separated by a distance a, as shown in Figure 1. Suppose that the nth A atom is displaced by a small amount $u_{A,n}$ from its equilibrium position. The forces attempting to return it to its equilibrium position will be proportional to the distance it has moved relative to the neighboring atoms. If we assume that the only significant force involves the nearest neighbor B atoms n and $n+1$, the equation of motion for the nth A atom is

$$M_A \frac{d^2 u_{A,n}}{dt^2} = C(u_{B,n} + u_{B,n+1} - 2u_{A,n}) \qquad [1]$$

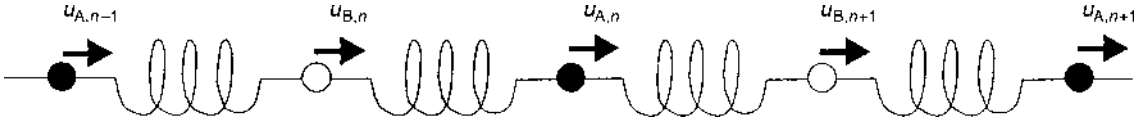

Figure 1 Atomic displacements in a diatomic linear chain.

where M_A is the mass of the atom, and C is the Hooke's law nearest-neighbor force constant. There is a similar equation

$$M_B \frac{d^2 u_{B,n}}{dt^2} = C(u_{A,n-1} + u_{A,n} - 2u_{B,n}) \qquad [2]$$

for the motion of the nth B atom. We assume that the solution of the equations of motion has the wavelike form

$$\begin{aligned} u_{A,n} &= U_A \exp\{i(qan - \omega t\} \\ u_{B,n} &= U_B \exp\{i(qan - \omega t\} \end{aligned} \qquad [3]$$

in which U_A and U_B are amplitude factors, ω is the angular frequency, and q is the wavevector, so that qa is the phase factor by which the phase of the wave changes from one site n to the next site $n+1$. The equations of motion are satisfied so long as the angular frequency ω is given by

$$\omega^2 = \frac{C}{M_A M_B}\Big[M_A + M_B \pm \sqrt{(M_A + M_B)^2 - 4M_A M_B \sin^2(qa/2)}\Big] \qquad [4]$$

There are two possible solutions here because there are two ions in the unit cell. A plot of ω as a function of q is shown in Figure 2a, with q in the range $-\pi/a$ to $+\pi/a$. In the lower branch, the *acoustic* phonon branch, the vibrations at long wavelength (equivalent to small q) are essentially the same as ordinary sound waves. The A and B atoms vibrate in phase, and at long wavelength there is a linear relation between ω and q:

$$\omega \approx v_s q \qquad [5]$$

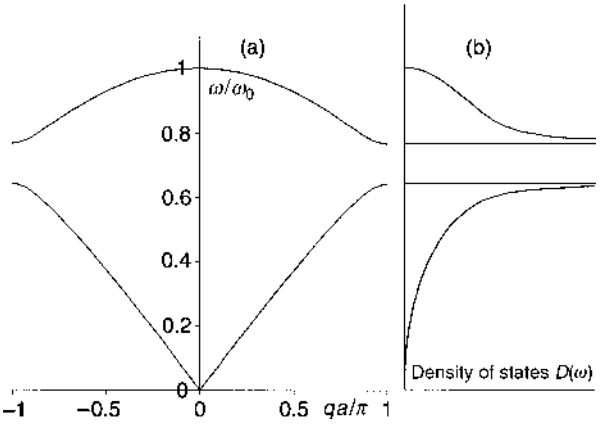

Figure 2 (a) Phonon frequency ω versus wavevector, q, in a one-dimensional diatomic lattice; (b) three-dimensional density of states $D(\omega)$.

where $v_s = \omega_0 a/2$ is the velocity of sound of the acoustic phonons.

The upper branch in Figure 2a, which is of more interest in the present context, is called the *optic* phonon branch. Here the A and B atoms vibrate in antiphase, and because the A and B atoms have equal but opposite charges there is an electric field associated with the vibration. The optic phonon vibrations will consequently affect the propagation of electromagnetic waves through the medium – hence the name – although the vibration frequencies (around 10 THz) are in the far infrared rather than the visible part of the electromagnetic spectrum. At this frequency, the infrared wavelength λ is about 30 μm. The propagation of the infrared will be affected only by phonons whose wavevector q matches the infrared wavevector $k = 2\pi/\lambda$. Since λ is around 10^4 times larger than the typical lattice spacing a in semiconductor materials, it follows that the optical properties of the material may be largely understood merely by considering the long-wavelength phonons with wavevector $q \approx 0$. Thus, in the diatomic case, only the optic phonon of wavevector $q \approx 0$ and frequency

$$\omega_0 = \sqrt{\frac{2C(M_A + M_B)}{M_A M_B}} \qquad [6]$$

is of importance to the propagation of electromagnetic radiation.

Many semiconductor materials are of the diatomic AB type – for example group III–V materials such as GaAs, and II–VI semiconductors such as CdTe. Their phonons are described in detail by three-dimensional versions of the simple linear chain model discussed above. The most basic semiconductor materials, Si and Ge, are clearly monatomic, but their crystal structure is identical to the zincblende structure of GaAs, and they show two phonon bands as in Figure 2a. However, because there is only one type of atom, these materials differ from the diatomic materials in that the optic phonon branch has no electric dipole moment. In Si and Ge, there is no dielectric anomaly around the frequency ω_0 of the type discussed in the following section. In these materials, the only significant optic effects occur when phonons combine in pairs with equal but opposite wavevector q, so that the resultant has very small wavevector that can match the infrared wavevector $k = 2\pi/\lambda$. The significant factor here is the density of states $D(\omega)$, representing the number of phonon states per unit frequency range. In Figure 2b, we show $D(\omega)$ for the diatomic lattice. This is

obtained as

$$D(\omega) = D(q)\frac{\mathrm{d}q}{\mathrm{d}\omega} \qquad [7]$$

where $D(q)$ is the density of states with respect to the wavevector, which is proportional to q^2 for a three-dimensional lattice. Note that $D(\omega)$ has peaks at frequencies corresponding to the zone boundary phonon frequencies because the dispersion curves are flat in these regions. These features are known as critical points, and they have well-defined shapes. Their spectroscopic importance is discussed later.

In bulk three-dimensional crystals, both transverse and longitudinal optic and acoustic phonons occur, each represented by different branches of the phonon dispersion curves, and the transverse branches in cubic materials are doubly degenerate. In three dimensions, more elaborate theoretical models are required, such as the rigid ion model and the shell model, and improved results can sometimes be achieved by including next-nearest-neighbor atoms in the model. As an example, the phonon dispersion curves of GaAs are shown in Figure 3.

It should be noted that the dispersion curves, and hence the associated optical properties, differ for propagation in different crystallographic directions, e.g., [100], [110], and [111].

Born–Huang Optic Phonon Model

The infrared lattice properties of materials at long wavelength are usually described by using the relative permittivity ε (often called the 'dielectric constant'), which is defined through the relation

$$D = \varepsilon\varepsilon_0 E \qquad [8]$$

between the electric field E in a medium and the displacement field D (here, ε_0 is the permittivity of free space). D is the sum of the field E and the polarization P of the medium:

$$D = P + \varepsilon_0 E \qquad [9]$$

P has contributions from lattice displacements (phonons), free electrons (plasmons), and the basic atomic polarizabilities. Since these quantities depend on frequency, ε is also a function $\varepsilon(\omega)$ of angular frequency ω. The charges of the A and B sites are taken as $+Q$ and $-Q$ respectively, and their equations of motion [1] and [2] are extended to include the electric forces $\pm EQ$. As a result of these considerations, the dielectric function can be written in the form

$$\varepsilon(\omega) = \varepsilon_\infty\left[1 + \frac{\omega_\mathrm{L}^2 - \omega_\mathrm{T}^2}{\omega_\mathrm{T}^2 - \omega^2}\right] = \varepsilon_\infty\frac{\omega_\mathrm{L}^2 - \omega^2}{\omega_\mathrm{T}^2 - \omega^2} \qquad [10]$$

Here, ε_∞ is the high-frequency dielectric constant due to the atomic polarizabilities, and ω_L and ω_T are frequencies that are close to the natural vibrational frequency ω_0 that occurs in the absence of electric field effects; the difference $(\omega_\mathrm{L}^2 - \omega_\mathrm{T}^2)$ is of order $\omega_0^2 Q^2/(Ca^3\varepsilon_0)$.

Figure 4 shows a plot of $\varepsilon(\omega)$ as a function of ω. Note that $\varepsilon(\omega)$ diverges at $\omega = \omega_\mathrm{T}$, and that $\omega = 0$ at $\omega = \omega_\mathrm{L}$ ($\omega_\mathrm{L}/\omega_\mathrm{T} = 1.25$ in Figure 4). ω_T is called the transverse optic (TO) phonon frequency: it is the resonance frequency of propagation of transverse

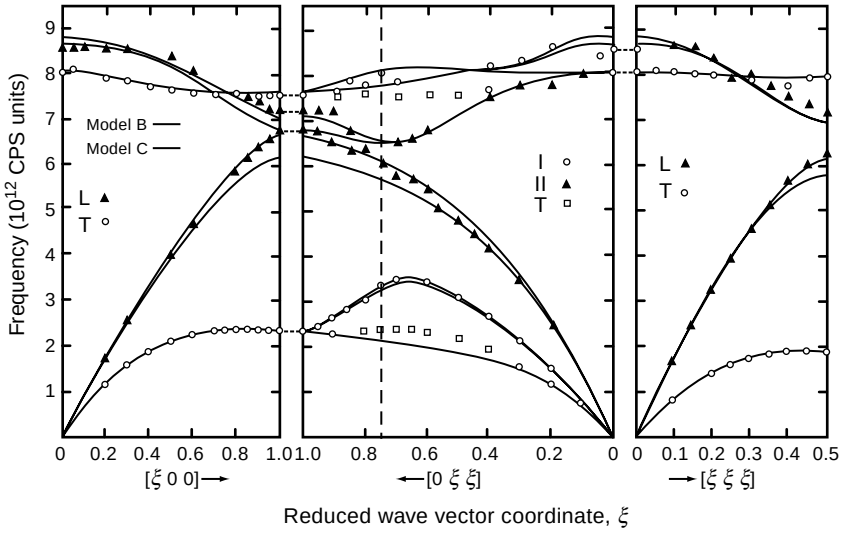

Figure 3 Phonon dispersion curves of GaAs determined by neutron spectrometry (points) compared with calculation (dotted and solid lines). Reproduced with permission from Waugh JLT and Dolling G (1963) Crystal dynamics of gallium arsenide. *Physical Review* 132: 2410.

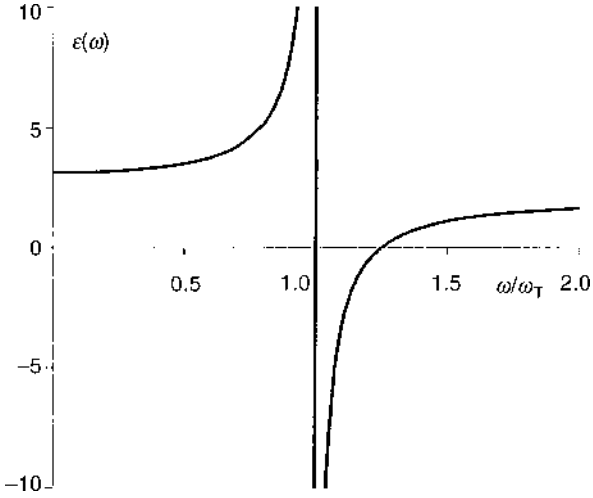

Figure 4 Dielectric function $\varepsilon(\omega)$ as a function of angular frequency, ω.

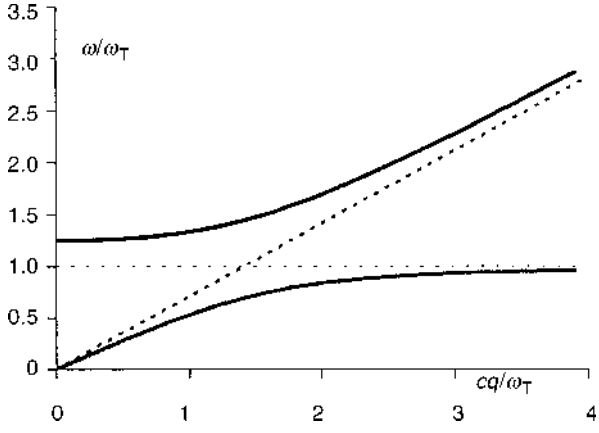

Figure 5 The dispersion curves for phonon polariton modes (solid lines) for a diatomic linear chain of atoms showing the gap in bulk phonon modes between the frequencies ω_T and ω_L.

electromagnetic waves in the material. ω_L is called the longitudinal optic (LO) phonon frequency: it is the frequency of the longitudinal electrostatic wave. Physically, the dielectric constant determines the velocity v of propagation of electromagnetic waves in the material

$$v^2 = \frac{\omega^2}{q^2} = \frac{c^2}{\varepsilon(\omega)} \qquad [11]$$

where c is the velocity of light. Equation [11] has two solutions for ω as a function of wavevector: the two solutions are shown in Figure 5. Note that there is no solution with ω in the range between ω_T and ω_L. The excitations in the region where the graphs are curved are called 'polaritons' – a combination of polar phonon and photon.

The refractive index n of the material is the ratio of v, the velocity of light in the medium, to c, the velocity in free space; thus $n = \sqrt{\varepsilon}$. n is often found from a measurement of the optical reflectivity R. At normal incidence

$$R = \left|\frac{1-n}{1+n}\right|^2 \qquad [12]$$

A graph of R is shown in Figure 6; note that the reflectivity is 100% for ω in the range $\omega_T < \omega < \omega_L$.

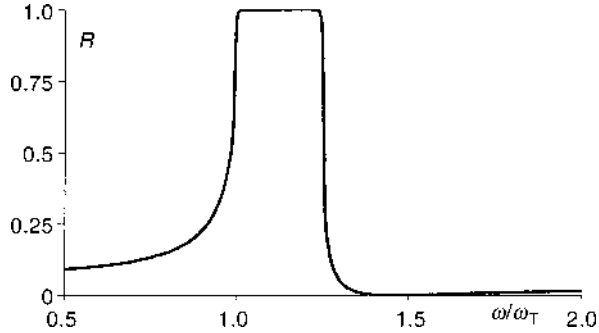

Figure 6 Calculated reflectivity plotted versus angular frequency for a diatomic linear chain of atoms for the case of zero damping.

Impurity Phonon Modes

Impurity atoms in crystal lattices can significantly affect the infrared properties. The simplest case is when an impurity atom has the same binding but a different mass. Consider the monatomic, one-dimensional case, in which the impurity atom has a mass $M + \delta M$, where M is the normal mass. Let the impurity be at the site $n = 0$. The equations of motion are

$$-M\omega^2 u_n = C(u_{n-1} + u_{n+1} - 2u_n) \quad \text{for } n \neq 0$$

$$-(M + \delta M)\omega^2 u_0 = C(u_{-1} + u_1 - 2u_0) \quad \text{for } n = 0 \qquad [13]$$

These equations have a solution provided that ω satisfies the equation

$$N^{-1}\frac{\delta M}{M}\sum_q \frac{\omega^2}{\omega_q^2 - \omega^2} = 1 \qquad [14]$$

where N is the number of atoms in the linear chain, and ω_q the unperturbed normal mode frequencies. This equation has solutions ω that lie close to the bulk phonon frequencies ω_q. However, if $\delta M < 0$ (i.e. the impurity atom is lighter than the bulk atoms), one additional mode is possible at a frequency ω that is higher than the maximum frequency $\omega_0 = \sqrt{C/M}$ of the bulk phonons. This is the so-called impurity or localized vibrational mode (LVM – see Figure 7).

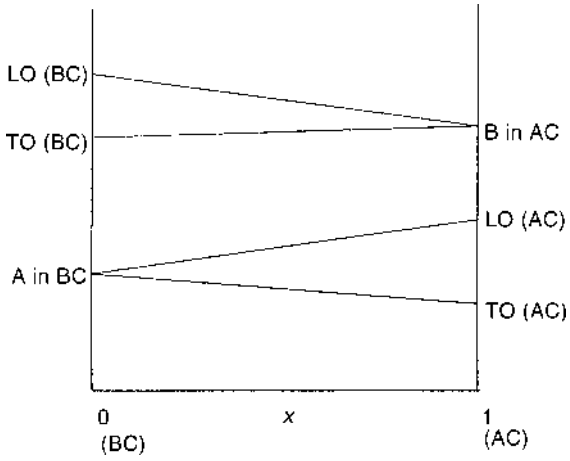

Figure 7 Representation of TO and LO phonon frequencies in a mixed crystal $A_xB_{1-x}C$ plotted versus composition x.

The impurity atom, being lighter than the material atoms, vibrates at a higher frequency than the bulk frequencies; the amplitude of the vibration is greatest at the impurity site, and decreases the further away one goes from the impurity site. If the impurity atom is heavier than the bulk atoms ($\delta M > 0$), there is no distinguishable local mode. In the case of diatomic lattices, an analogous situation pertains: if the impurity is lighter than the bulk atoms, an impurity mode is distinguishable above the top of the optic phonon band; if the impurity is heavier, there is the possibility that a distinct impurity mode may appear in the gap between the acoustic and optic phonon branches.

Mixed Crystals

Mixed compound semiconductor crystals are in widespread use because the bandgap can be adjusted by varying the composition. For example, the ternary material $A_xB_{1-x}C$ is an alloy of a fraction x of the material AC with a fraction $1 - x$ of the material BC. A typical example is $Al_xGa_{1-x}As$. The infrared lattice properties of these materials are classified as 'one-mode' or 'two-mode' depending on how many distinct optic phonon frequencies are observed. The distinction can easily be understood in the context of the discussion above of impurity modes. If an A atom in the bulk material BC has a distinct local impurity optic phonon mode frequency, and similarly for a B impurity in AC, then there will usually be two distinct optic phonon modes throughout the whole composition range $0 < x < 1$: one mode is AC-like, and the other BC-like, as shown in Figure 7. On the other hand, if neither A in BC, nor B in AC, provide distinct optic phonons, there is only one distinguishable optic phonon in the alloy $A_xB_{1-x}C$.

Review of Far-Infrared Properties of Semiconductors

Group IV Semiconductors

We begin with Si and Ge, both of which crystallize in the diamond structure. As mentioned earlier, there is no first-order interaction between far infrared radiation and the TO mode. However, there are several other mechanisms that result in absorption. The dipole moment of the zone center TO phonon in an ionic solid is an example of a symmetry property that, in group theory, is described by a set of selection rules for each different crystal lattice. Other mechanisms that can lead to the creation of a dipole moment are:

Disorder

Disorder destroys the translational symmetry, and hence the selection rules. Consequently, phonon modes that, in a perfect lattice, are not infrared active will become active. The most general case occurs if the crystal becomes completely disordered, in which case all phonon modes become dipole-active and the far-infrared absorption spectrum resembles the phonon density of states. This is illustrated in Figure 8, in

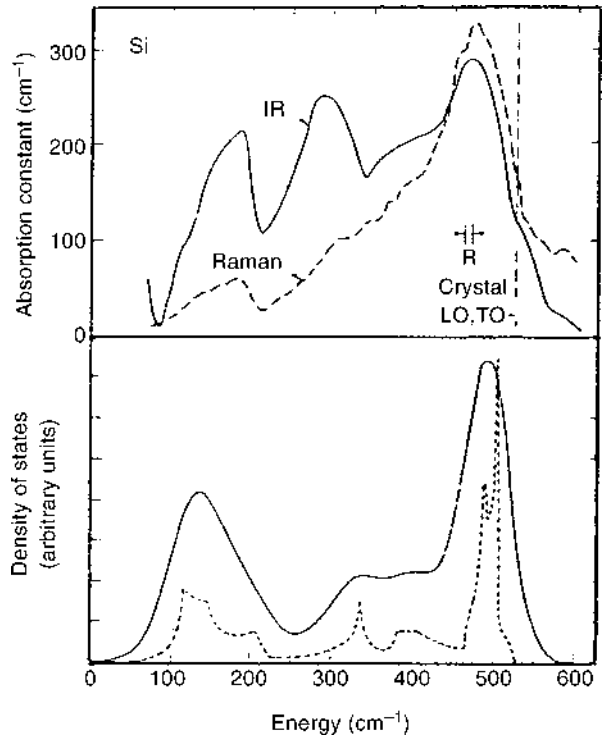

Figure 8 Top: infrared absorption coefficient (solid line) and Raman spectrum (dashed line) of amorphous Si. Bottom: calculated density of states (solid line), and broadened density of states (dashed line). Reproduced with permission from Bradskyn H and Lunio (1974) Infrared vibrational spectra of amorphous Si and Ge. *Physical Review B* 9: 1646, ADS.

which measured infrared and Raman spectra of amorphous Si are compared with the calculated one-phonon density of states.

Phonon combination bands

A combination of the lattice distortions due to two phonons excited simultaneously can produce a resultant displacement of the ions in the lattice with a dipole moment. Such phonon combinations can be excited simultaneously by a far-infrared photon, and they are observed as weak bands in the infrared absorption spectrum. Apart from the dipole moment selection rule, it is also necessary for wavevector conservation to occur. Since $q \sim 0$ for the photon, the two phonons must either be at the zone center, or have equal and opposite wavevectors. An additional consideration is that, if the process is to lead to observable absorption, it should have a high probability as determined by the density of states, so observable absorption features usually arise from critical-point phonon combinations. An example of phonon combination bands in Si is shown in Figure 9. Assignment of the features in the spectrum is usually possible by inspection of the phonon dispersion curves at the zone boundaries, a procedure known as critical-point analysis. Alternatively, the two-phonon density of states can be calculated, taking into account the selection rules, and then compared with the measured spectrum. Similar considerations apply to higher-order phonon combination bands, e.g., three or four phonons, etc., but the strength of the combination bands usually decreases rapidly as the order of interaction increases.

The presence of impurities

As mentioned above, dipole-active localized vibrational modes (LVMs) can occur if the sample is doped with lighter elements as this destroys the local translational symmetry. For instance, an LVM occurs if Si is doped with B, as shown in Figure 10. The LVM frequency is proportional to the ratio $(C/M)^{1/2}$ as seen earlier where C is the force constant and M is the isotopic mass of the substituted atom. As such modes have narrow linewidths, the features in the spectrum due to different isotopes of the same element are easily distinguished, as shown in Figure 10.

Group III–V and II–VI Semiconductors: Bulk Crystals

The group III–V and group II–VI semiconductors crystallize in either the cubic (zincblende) structure, or the hexagonal (wurtzite) structure. Many mixed crystal materials, such as $Al_xGa_{1-x}As$, are grown for

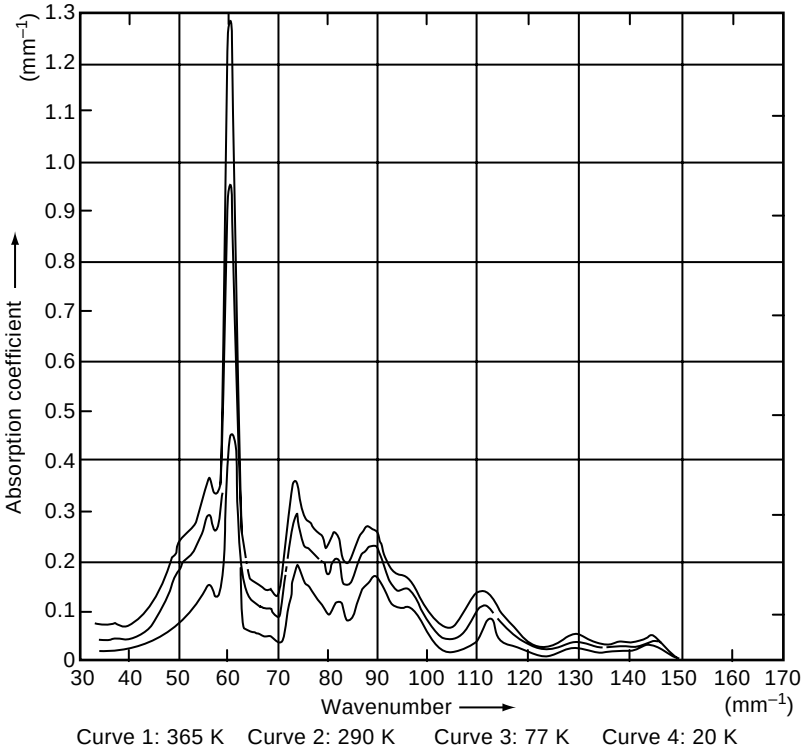

Figure 9 Measured lattice combination bands in a single crystal of silicon at 290 K (top), 77 K (middle), and 20 K (bottom). (Note that 100 mm^{-1} = 1000 cm^{-1}.) Reproduced with permission from Johnson FA (1959) Lattice absorption bands in silicon. *Proceedings of the Physical Society* 73: 265.

electronic applications that require materials with bandgaps not found naturally. Furthermore, a host of low-dimensional structures, such as epilayers, multiple quantum wells (MQWs) and superlattices, are fabricated by molecular beam epitaxy (MBE) or metalorganic vapor phase epitaxy (MOVPE) to create devices that exploit the novel properties which occur when one or more of the physical dimensions are small.

Reflectivity from a bulk single crystal

In the reststrahl frequency range bulk phonon modes cannot propagate and incident radiation is perfectly reflected, i.e., $R = 1$, as shown in the calculated spectrum in Figure 6. Figures 4 and 6 represent the ideal case of a harmonic crystal, in which there is no interaction between the phonon modes, so that all modes have infinite lifetime or, equivalently, spectral features have zero linewidth, corresponding to zero damping. This is close to the behavior of a crystal at absolute zero but, in practice, the phonon modes are coupled, so each mode will decay into combinations of other phonon modes. This leads to finite phonon lifetimes and finite linewidths of all features in the measured spectra. This damping has the effect of

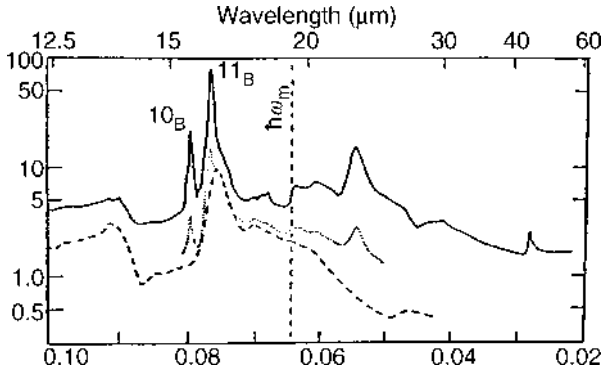

Figure 10 The far-infrared absorption spectrum of Si doped with B ($^{10}B + {}^{11}B$). Reproduced with permission from Angress JF, Goodwin AR and Smith SD (1965) A study of the vibrations of boron and phosphorus in silicon by infrared-red absorption. *Proceedings of the Royal Society A* 287: 64.

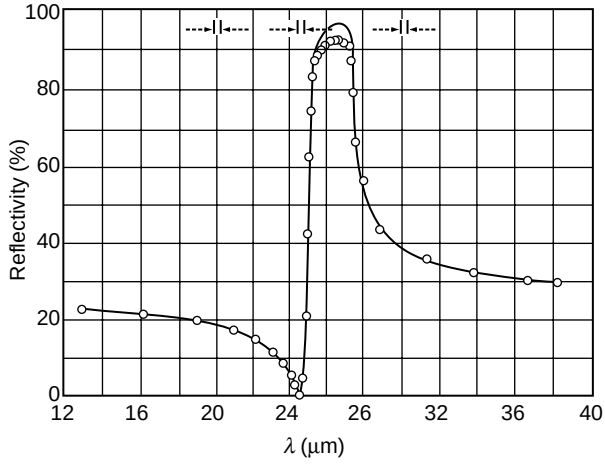

Figure 11 The reststrahlen band reflectivity of GaP, measured (points) and calculated from a damped simple harmonic oscillator (line). Reproduced with permission from Kleinman DA and Spitzer WG (1960) Infrared lattice absorption of GaP. *Physical Review* 118: 110.

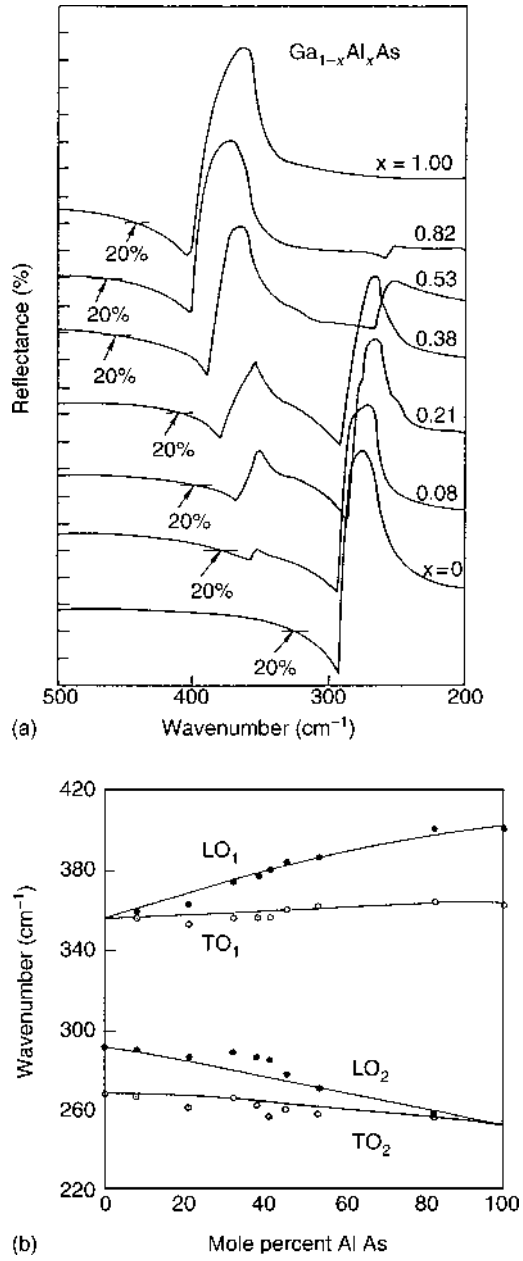

Figure 12 (a) Reflection spectra for GaAs, AlAs, and five $Ga_{1-x}Al_xAs$ mixed crystals. The curves have been displaced for clarity. (b) TO and LO mode frequencies obtained analytically from the spectra, plotted versus alloy composition. Reproduced with permission from Ilegens M and Pearson GL (1970) Infrared reflection spectra of $Ga_{1-x}AL_xAs$ mixed crystals. *Physical Review B* 1: 1576.

rounding off features in the reflectivity compared with Figure 6, and it can be represented analytically by the equation:

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{(\varepsilon(0) - \varepsilon_\infty)\omega_T^2}{\omega_T^2 - \omega^2 - i\gamma\omega} \qquad [15]$$

This differs from eqn [10] only by the addition of the damping coefficient, γ, which takes account in a phenomenological way of the interactions between the phonon modes. Here $\varepsilon(0)$ is the dielectric function at zero frequency, and $\varepsilon(0) - \varepsilon_\infty$ is the oscillator strength (i.e., the dipole strength) of the mode. We shall consider γ again later.

The reflectivity of GaP is shown in Figure 11, and this should be compared with Figure 6. In pure diatomic crystals like the simple ionic solids, e.g., NaCl and KCl, or the binary semiconductors, e.g., GaAs and CdTe, this is the only region of high reflectivity at frequencies below the visible or ultraviolet, and it is often described as the reststrahlen band, from the German term for residual ray.

Reflectivity from mixed crystals of binary semiconductors

The spectra of mixed crystals were discussed above and an example is shown in Figure 12a, which shows the reflection spectra of GaAs, AlAs and five $Ga_{1-x}Al_xAs$ mixed crystals. The TO and LO phonon frequencies were obtained from the spectra using a Kramers–Kronig analysis and classical dispersion theory. Changes in the mode frequencies with composition are interpreted using the so-called random element isodisplacement (REI) model. In Figure 12b, the TO and LO frequencies are plotted versus composition, x, and it can be seen that this system displays typical two-mode behavior, following closely the behavior shown in Figure 7.

Localized vibrational modes due to impurities

LVMs have been widely investigated in binary semiconductors because they offer a useful diagnostic tool for the investigation of impurities. In addition to the identification of isotopic species, high-resolution transmission spectroscopy can be used to determine the sites of the impurity atoms in the lattice. In Figure 13 the LVM absorption band for ^{12}C in GaAs is shown at different resolutions. The ^{12}C atoms are substitutional impurities on the As sites and the different components of the band arise from the different nearest-neighbor isotopic arrangements

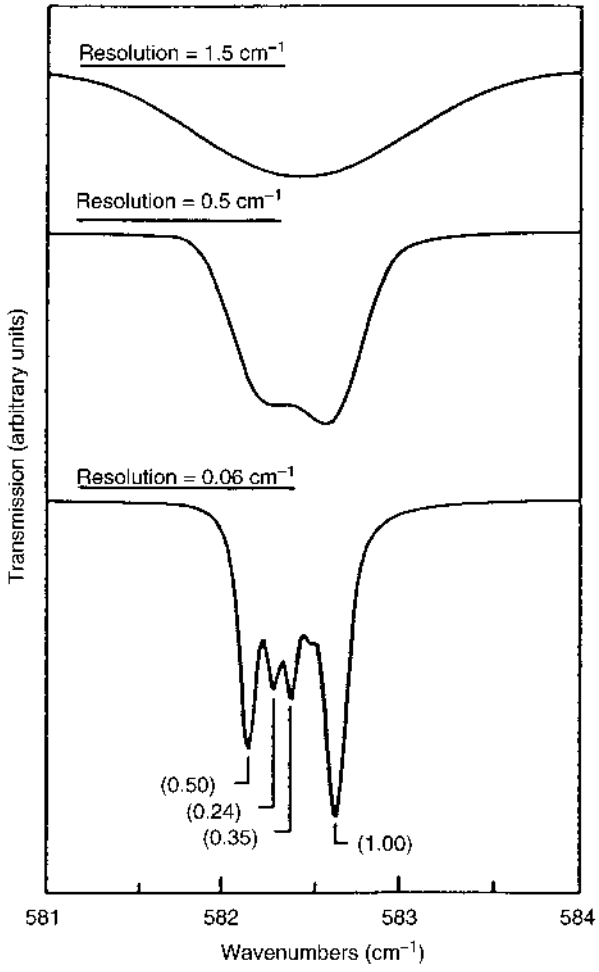

Figure 13 Localized vibrational mode (LVM) absorption band for ^{12}C in GaAs measured at 80 K. The numbers in brackets are estimated band strengths. Reproduced with permission from Theis WM, Bajaj KK, Litton CW and Spitzer WG (1982) Direct evidence for the site of substitutional carbon impurity in GaAs. *Applied Physics Letters* 41: 70.

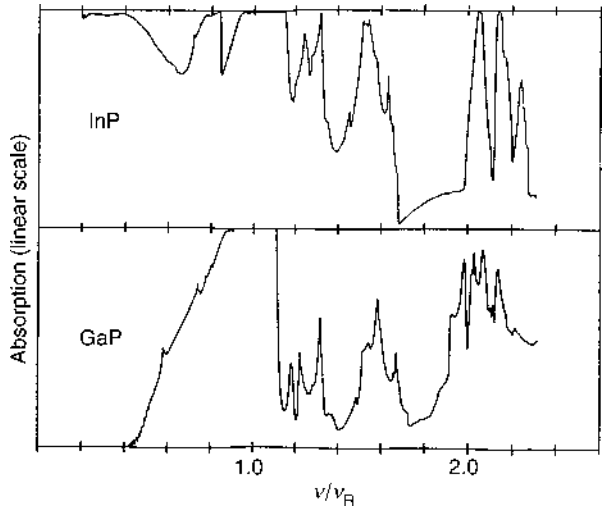

Figure 14 High-resolution far-infrared absorption spectra of GaP and InP with the frequency scale normalized in each case to the reststrahl frequency ($\nu/\nu_R = 1$ corresponds to $\omega = \omega_T$). Reproduced with permission from Koteles ES and Datars WR (1976) Two-phonon absorption in InP and GaP. *Solid State Communications* 19: 221.

surrounding the different ^{12}C atoms in the lattice. Note that ^{12}C is much lighter than either of the host lattice atoms, Ga or As, as required theoretically.

Phonon combination bands

Weak features due to the simultaneous excitation of two or more phonons are observed in the spectra of all semiconductors, and there are two excitation mechanisms. The first is direct excitation of two or more phonons by an incident photon via nonlinear terms in the dipole moment. The second process is indirect: the incident photon first excites a zone center TO phonon and the TO phonon then decays into a combination of phonons elsewhere in the Brillouin zone, a process described as anharmonic decay of the TO phonon. Mathematically, it is described by a complex frequency-dependent damping coefficient, $\gamma(\omega)$, in the oscillator equation [15]. Clearly, the latter mechanism is not available in Si and Ge as the TO phonons are not dipole-active. Anharmonicity is very important in strongly ionic crystals like the alkali halides (i.e., group I–VII compounds like NaCl), leading to prominent phonon combination bands, but it becomes progressively less important in semiconductors as the interatomic bonding becomes less ionic and more covalent, which is the case in the sequence through group II–VI (e.g., semiconductors like CdTe or ZnSe), group III–V (e.g., GaAs or InP), and group IV (Ge and Si), which are entirely covalent. Nearly all two-phonon combinations are dipole-active in zinc-blende structure crystals with the result that the transmission spectrum either side of the reststrahlen band is very rich in weak phonon combination bands, as shown in Figure 14 for GaP and InP.

Layered Semiconductor Structures

Confined optical phonons

Semiconductor superlattices are fabricated for a variety of electronic and optical applications, and the

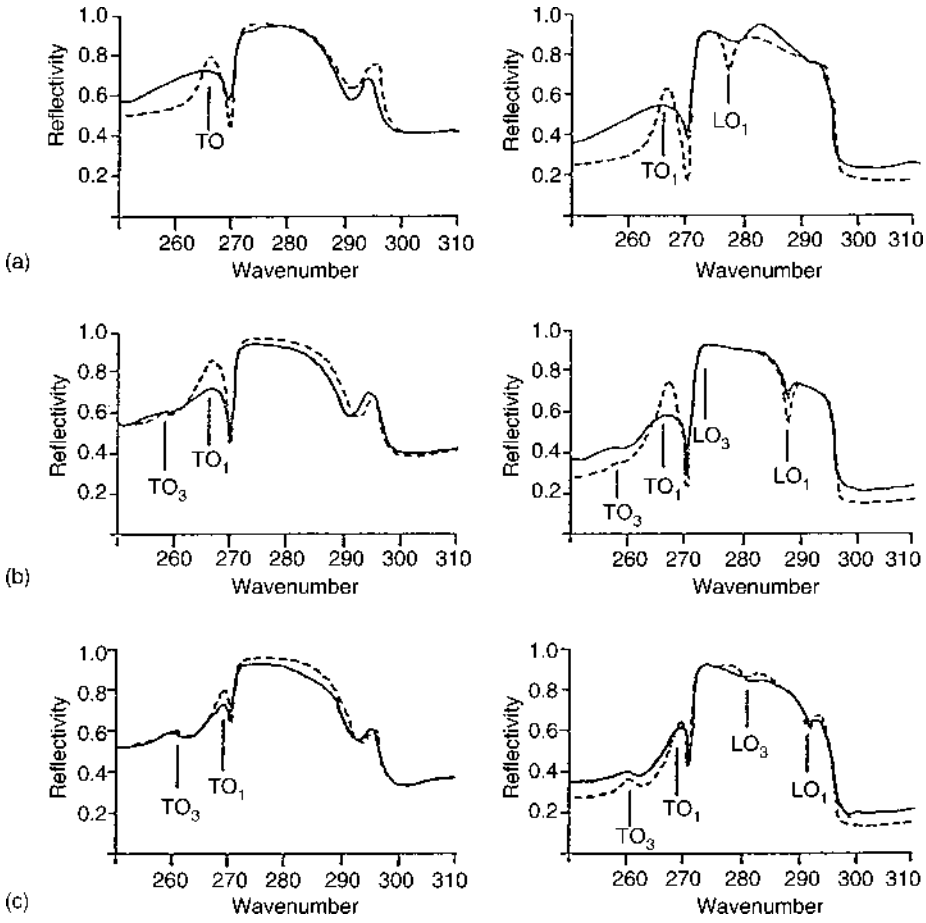

Figure 15 Measured (solid lines) and calculated (dashed lines) oblique incidence (45°) reflectivity spectra at 77 K of three short-period $(GaAs)_n/(AlAs)_n$ superlattices on GaAs substrates: *n* denotes the number of monolayers. Left curves: s-polarization, right curves: p-polarization. Note that the strong reststrahl reflectivity is from the GaAs substrates and the fine structure is due to confined modes. The subscripts 1, 2, and 3 denote the orders of the confined modes. Reproduced with permission from Dumelow T, Hamilton AA, Parker TJ, Tilley DR, Foxon CT, Hilton D and Moore KJ (1990) Far infrared measurements of bulk and surface phonons in GaAs/AlAs superlattices. *International Journal of Infrared and Millimeter Waves* 11: 901.

quality of the interfaces in such structures is important as it determines the amount of electronic scattering at the interfaces. Some structures are fabricated from alternate layers of materials such as GaAs and AlAs, in which there is no overlap between the frequencies of the optical branches of the phonon dispersion curves. This has the effect that optical phonons excited in one type of layer, e.g., in GaAs, cannot propagate into the other type of layer, i.e., AlAs, and vice versa, so that the optical phonon modes are confined in their respective layers. The confined modes behave like standing waves, with frequencies determined by the thicknesses of the layers and the bulk optical phonon dispersion curves. Thus, if the superlattice structure is of poor quality, and the interfaces are broad due to alloying, rather than abrupt, there will be a change in the effective widths of the layers. This leads to small shifts in the frequencies of the confined phonons which can be measured and used to obtain a quantitative estimate of the amount of interface broadening. An example is shown in Figure 15. The degree of interface broadening determined from these spectra was 1.4 atomic spacings, in agreement with estimates using other techniques.

Surface phonon polaritons and guided waves

The reststrahl reflection band (Figure 6) occurs because there is a gap between ω_T and ω_L in the polariton dispersion curves (Figures 4 and 5), with no solutions for bulk phonon modes. Surface phonon polaritons can, however, propagate in this frequency band, but they have wavevectors that are larger than those, ω/c, of far-infrared radiation, so special techniques are required to observe them. The necessary wavevector enhancement can be obtained in two ways. Prism coupling can be used, as shown in Figure 16. When the angle of incidence of the incident beam at the base of the prism exceeds the critical angle for total internal reflection an evanescent wave with advanced wave vector, $q > \omega/c$, is set up at the base of the prism. This wave can couple to a surface polariton, and the energy used to excite the polariton results in a dip in the reflectivity at the frequency of the surface polariton, hence the name of the technique: attenuated total reflection (ATR). An example is shown in Figure 16. Alternatively, wavevector enhancement can be achieved by placing a grating on the surface of the sample. The incident photon wavevector is then enhanced by $+2m\pi/d$, where d is the grating period and m is an integer.

Interface phonons

Under certain conditions phonons can propagate along the interface between two different media that are in intimate contact, as for instance, in the case of thin epilayers on substrates, or in multilayered

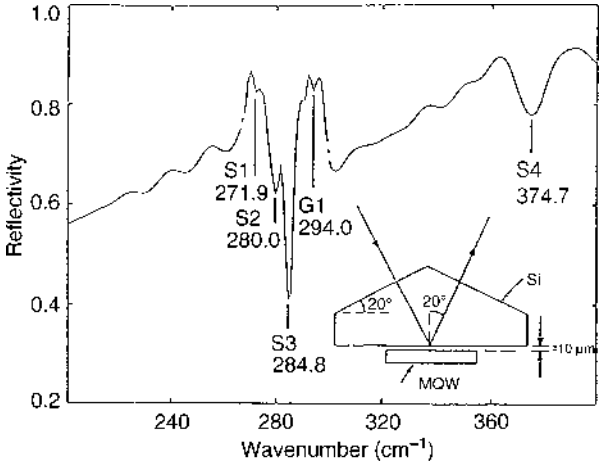

Figure 16 ATR reflection spectrum with dips due to surface phonon polaritons (S) and guided waves (G) in a GaAs/$Al_{0.38}Ga_{0.65}As$ MQW. Inset: schematic diagram of the ATR stage. Reproduced with permission from Dumelow T, Hamilton AA, Parker TJ, Tilley DR, Foxon CT, Hilton D and Moore KJ (1990) Far infrared measurements of bulk and surface phonons in GaAs/AlAs superlattices. *International Journal of Infrared and Millimeter Waves* 11: 901.

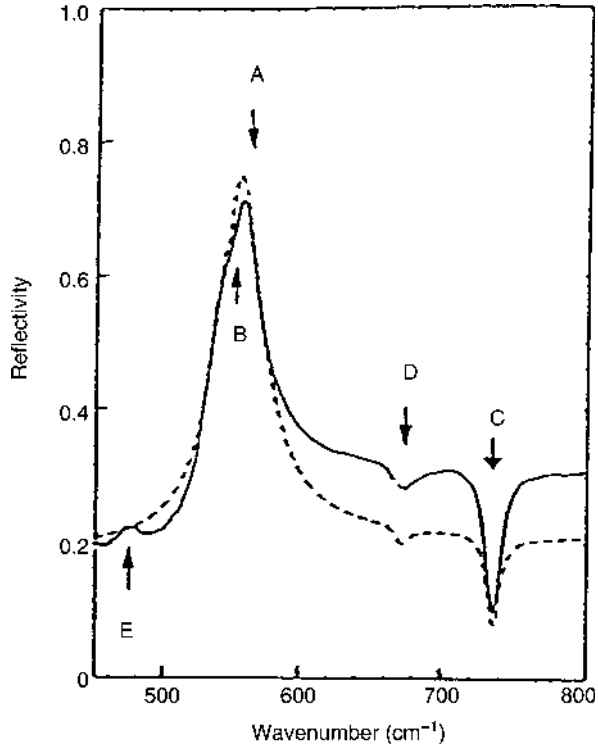

Figure 17 Far-infrared reflectivity of a GaN epilayer on a GaAs substrate at 290 K with dips due to a Brewster mode at C and a Berreman mode at D. Reproduced with permission from Mirjalili G, Parker TJ, Farjami Shayesteh S, Bulbul MM, Smith SRP, Cheng TS and Foxon CT (1998) Far infrared and Raman analysis of phonons and phonon interface modes in GaN epilayers on GaAs and GaP substrates. *Physical Review B* 57: 4656.

structures. There are two common sets of criteria for the existence of such interface modes. The first, leading to so-called Brewster modes, is when the numerator of the Fresnel reflection coefficient passes through zero due to impedance matching in the two media. The second, leading to a Berreman mode, is when the real part of the dielectric constant has a zero crossing in one of the media at a frequency where it is negative in the other medium. Both modes are observed as dips in the reflectivity. A reflection spectrum for a GaN epilayer on a GaAs substrate is shown in Figure 17. The two interface modes occur at frequencies close to the frequencies of GaN LO phonons; the Berreman mode is close to the GaN LO phonon frequency at the base of the GaN epilayer, which is shifted down by about 50 cm^{-1} from the bulk GaN value (735 cm^{-1}) due to disorder arising from the large lattice mismatch between the epilayer and the substrate. Both interface modes couple strongly to free carriers (i.e., plasmons) propagating normal to the interface, and thus act as useful probes of the structural and electronic properties of the epilayer. No other technique is available for probing the electronic properties of the epilayer for propagation normal to the layer.

See also

Semiconductor Materials: GaAs Based Compounds. **Semiconductor Physics:** Band Structure and Optical Properties; Light Scattering; Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Angress JF, Goodwin AR and Smith SD (1965) A study of the vibrations of boron and phosphorus in silicon by infrared-red absorption. *Proceedings of the Royal Society A* 287: 64.

Barker AS Jr and Sievers AJ (1975) Optical studies of the vibrational properties of disordered solids. *Reviews of Modern Physics* 47 (Suppl. 2): S1.

Berreman DW (1963) Infrared absorption at longitudinal optic frequency in cubic crystal films. *Physical Review* 130: 2193.

Born M and Huang K (1954) *Dynamical Theory of Crystal Lattices*. Oxford: Clarendon Press.

Dumelow T and Tilley DR (1993) Optical properties of semiconductor superlattices in the far infrared. *Journal of the Optical Society of America* 10: 633.

Dumelow T, Hamilton AA, Parker TJ, Tilley DR, Foxon CT, Hilton D and Moore KJ (1990) Far infrared measurements of bulk and surface phonons in GaAs/AlAs superlattices. *International Journal of Infrared and Millimeter Waves* 11: 901.

Johnson FA (1959) Lattice absorption bands in silicon. *Proceedings of the Physical Society* 73: 265.

Kittel C (1996) *Introduction to Solid State Physics*, 7th edn. New York: John Wiley.

Kleinman DA and Spitzer WG (1960) Infrared lattice absorption of GaP. *Physical Review* 118: 110.

Koteles ES and Datars WR (1976) Two-phonon absorption in InP and GaP. *Solid State Communications* 19: 221.

Mirjalili G, Parker TJ, Farjami Shayesteh S, Bulbul MM, Smith SRP, Cheng TS and Foxon CT (1998) Far infrared and Raman analysis of phonons and phonon interface modes in GaN epilayers on GaAs and GaP substrates. *Physical Review B* 57: 4656.

Newman RC (1974) *Studies of Crystal Defects*. London: Taylor and Francis.

Theis WM, Bajaj KK, Litton CW and Spitzer WG (1982) Direct evidence for the site of substitutional carbon impurity in GaAs. *Applied Physics Letters* 41: 70.

Waugh JLT and Dolling G (1963) Crystal dynamics of gallium arsenide. *Physical Review* 132: 2410.

Light Scattering

M Balkanski, Université Pierre et Marie Curie, Paris, France

© 2005, Elsevier Ltd. All Rights Reserved.

When a monochromatic light beam passes through a transparent medium, a small part of the light is scattered out of the incident direction with a change in frequency. The scattering processes are usually classified into three groups. The light scattering is named:

Brillouin scattering when the light is scattered with a small frequency shift that varies continuously with scattering angle.

Raman scattering when the light is scattered with a relatively large frequency shift that is independent of scattering angle. The possibility of observing a given transition, however, depends on the orientation of the crystal relative to the polarization of the incident light.

Rayleigh scattering when the light is scattered without frequency shift. In this case the scattering is elastic with the incident and scattered frequencies equal: $\omega_i = \omega_s$. Rayleigh scattering is particularly useful in the study of critical phenomena or aspects related to the size and polarizability of particles.

Brillouin Scattering

Brillouin scattering is caused by the interaction of light with the acoustic modes of vibration of the crystal. Consider an acoustic wave of frequency ω_a propagating with velocity $\pm v$ as shown in Figure 1. Light with incident frequency ω_i interacts strongly with acoustic waves satisfying the Bragg condition $\omega_a = 2\omega_i(v/c)n\sin(\theta/2)$, where n is the refractive index of the material and θ is the angle of deviation of the scattered beam. Since the acoustic waves are moving with velocity $\pm v$ the scattered light suffers a Doppler shift in frequency and exhibits a frequency doublet at the angle θ given by the Brillouin equation

$$\omega_i \pm \omega_a = \omega_i \pm 2\omega_i(v/c)n\sin(\theta/2)$$

In terms of the quantum picture, Brillouin scattering corresponds to scattering of incident photons of frequency ω_i and wavevector $\mathbf{k}_i$ into scattered photons of frequency ω_s and wavevector $\mathbf{k}_s$ with the emission or absorption of an acoustic phonon of frequency ω_a and wavevector $\mathbf{q}$. The predicted Brillouin spectrum is shown in Figure 2. It consists of Stokes ($\omega_s = \omega_i - \omega_a$) and anti-Stokes ($\omega_s = \omega_i + \omega_a$) components corresponding to phonon emission and absorption, respectively.

The frequency shift $|\omega_i - \omega_s|$ for a well-defined scattering angle θ directly gives the acoustic phonon frequency and velocity. The width Γ of the Brillouin component of the spectrum is a measure of the damping or attenuation of the wave.

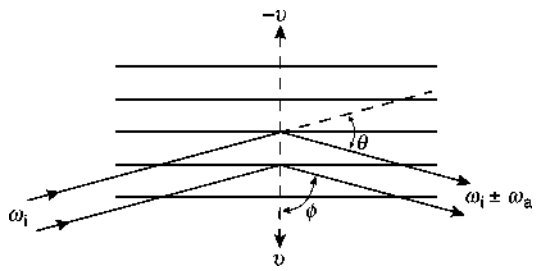

Figure 1 Light scattering at angle ϕ and Bragg reflection at angle θ by moving thermal waves.

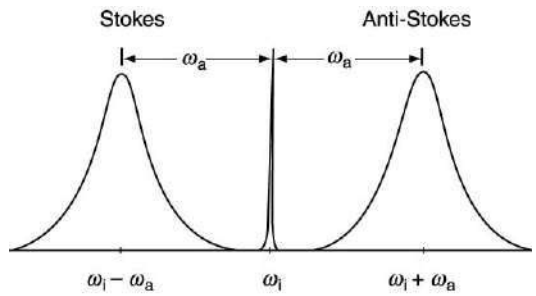

Figure 2 Schematic diagram of a Brillouin spectrum.

Raman Scattering

The essential difference between Brillouin and Raman scattering is that in Raman scattering the incident light beam is scattered with relatively large frequency shift independent of the scattering angle. The same basic considerations apply to Raman scattering that apply to Brillouin scattering. The Raman spectrum has Stokes and anti-Stokes branches corresponding to the emission and absorption, respectively, of an elementary excitation. A variety of elementary excitations are important. They include optical phonons and, in the case of magnetic materials, magnons. Also of interest are electronic excitations such as intraband single-particle excitations, interband excitations, and collective excitations (plasmons).

Raman Scattering by Phonons

An inelastic light scattering event involves the destruction of a photon of frequency ω_i incident from a light source, the creation of a scattered photon of frequency ω_s and the creation or destruction of an optical phonon of frequency ω_o.

Destruction of a photon involves interaction of the radiation field with matter by which a pair of free carriers is created in the semiconductor corresponding to a virtual transition between the valence and conduction bands. The excited state, free electron in the conduction band, interacts with the lattice via the optical deformation potential. This interaction results in the creation or annihilation of an optical phonon of frequency ω_s. When the resulting process corresponds to the creation emission of a phonon, the frequency of the scattered photon is $\omega_s = \omega_i - \omega_a$ and is referred to as the Stokes component of the spectrum. When the process is the annihilation or absorption of a phonon the scattered frequency is $\omega_s = \omega_i + \omega_a$, referred to as the anti-Stokes component.

These two first-order scattering events can be described in terms of Feynman diagrams as shown in Figure 3.

Considering these elementary processes one can calculate, using time-dependent perturbation theory, the Raman amplitude and hence the differential scattering cross-section. The remarkable feature here is that for incident frequencies approaching the frequencies of an electronic excited state of the material the Raman amplitude diverges. This leads to the phenomenon of resonant Raman scattering.

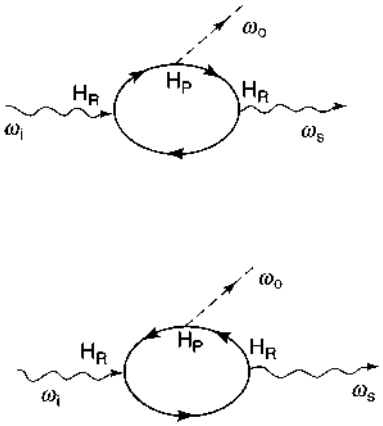

Figure 3 Feynman diagrams for two first-order scattering events.

The Raman scattering efficiency, which is the scattering cross-section per illuminated area can be calculated numerically. Numerical estimates give efficiencies typically of order 10^{-6} or 10^{-7}.

In polar semiconductors the macroscopic electric field leads to an additional contribution to the scattering cross-section for the longitudinal optical (LO) phonon modes. The macroscopic field is a consequence of the Coulomb field of the ionic charges and gives rise to an electron–phonon interaction.

There is no simple relationship between the scattering intensities for transverse optical (TO) and LO phonons in polar semiconductors.

When the probing frequency coincides with the transverse optical mode frequency, i.e., for the regime $cq \cong \omega_{TO}$ (where c is the light velocity, q the photon wavevector and ω_{TO} the transverse optical phonon frequency) in polar semiconductors it is necessary to take into account the polariton nature of the coupled photon–TO-phonon modes and their associated macroscopic electrical field. The dispersion relation of the coupled photon–TO mode, called a polariton, consists of two branches: an upper branch ω_+, in which the electro-optical contribution dominates, and the lower branch ω_-, which contains essentially the mechanical contribution. This leads to a contribution to the Raman amplitude from both mechanical and electro-optical origins. An exception occurs for the lower polariton branch with $\omega = \omega_{TO}$; then the electro-optic contribution is very small compared to the mechanical contribution. If $cq \gg \omega_{TO}$ as in the upper polariton branch, destructive interference between the mechanical and electro-optic contributions can occur.

Selection Rules in Raman Scattering

Not all elementary excitations in semiconductors scatter light. The Raman active modes are determined by selection rules established using group-theoretical methods. The various normal modes in a given crystal correspond to various symmetries of the vibrations of the atoms in the crystal and are characterized by the irreducible representations of the space group of the crystal lattice. One can show that a normal mode can participate in a first-order Raman transition if and only if its irreducible representation is the same as one of the irreducible representations that occur in the reduction of the representation of the Raman tensor.

An important result of group theory is the rule of mutual exclusion which states that, in crystals with a center of inversion, excitations that are active in the first-order infrared spectrum are inactive in the first-order Raman spectrum, and conversely, excitations that are active in the first-order Raman spectrum are inactive in the first-order infrared spectrum. In particular, the first-order spectra due to optical phonons in NaCI are infrared active, but Raman inactive, whereas in Si they are infrared inactive, but Raman active. This difference is related to the fact that each atomic site in NaCI is a center of inversion and the active optical phonons have odd parity whereas each midpoint between two nearest-neighbor atomic sites in Si is a center of inversion and the active optical phonons have even parity.

In a crystal with the zinc blend structure, there is no center of inversion, and the crystal is both infrared and Raman active. The only non-vanishing elements of the Raman tensor $R(\hat{e})$ with $\hat{e}\|z$ are the xy and yx elements. To observe the Raman effect of an LO phonon propagating in the z-direction, one can arrange the polarization of the incident light parallel to the x-axis and observe the scattered light with its polarization parallel to the y-axis or vice versa. An analogous set of constraints applies to TO phonons. A typical spectrum is illustrated in Figure 4.

For higher-order Raman processes in which several phonons participate, it is the product of the irreducible representations of the phonons involved that must be the same as the irreducible representation of the Raman tensor.

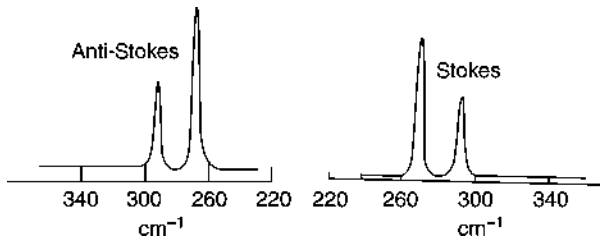

Figure 4 Raman spectrum for optical phonons.

Geometrical Aspects of First-Order Raman Scattering

The observation of light scattering by optical phonons in transparent crystals is usually done in a geometry in which the linearly polarized incident light beam is directed along, say, the x-axis and the scattered beam is observed along the y-axis (see Figure 5). When the crystal is not transparent, as is often the case in semiconductors, the observation is made in the backscattering geometry in which the scattered beam is observed in the opposite direction to the incident beam.

The scattering geometry affects the range of phonon wavevectors that is accessible in first-order Raman scattering. The condition of wavevector conservation is

$$\mathbf{K}_i - \mathbf{K}_s = \pm\mathbf{q}$$

where $\mathbf{q}$ is the phonon wavevector and the plus (minus) sign refers to the Stokes (anti-Stokes) process. The Stokes geometry is shown in Figure 6 and satisfies the following relation:

$$q^2 = k_i^2 + k_s^2 - 2k_i\, k_s \cos\theta$$

Forward scattering is characterized by $\theta = 0$ and a minimum value of q given, for isotropic media, by

$$q_{\min} = \frac{n(\omega_i)\,\omega_i - n(\omega_s)\omega_s}{c}$$

where $n(\omega_i)$ and $n(\omega_s)$ are the refractive indices of the crystal for the incident and scattered light, respectively, and $k_{i(s)} = n(\omega_{i(s)})\omega_{i(s)}/c$.

Back scattering is characterized by the maximum value of q when $\theta = 180°$ and is given by

$$q_{\max} = \frac{n(\omega_i)\omega_i - n(\omega_s)\omega_s}{c}$$

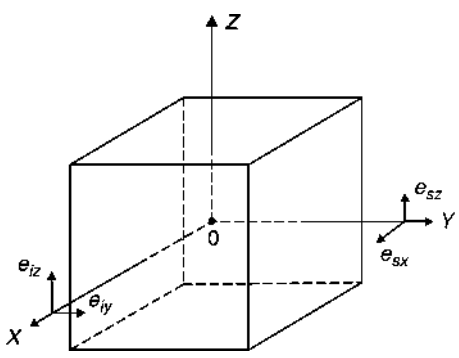

Figure 5 Geometry for light-scattering experiments in transparent crystals.

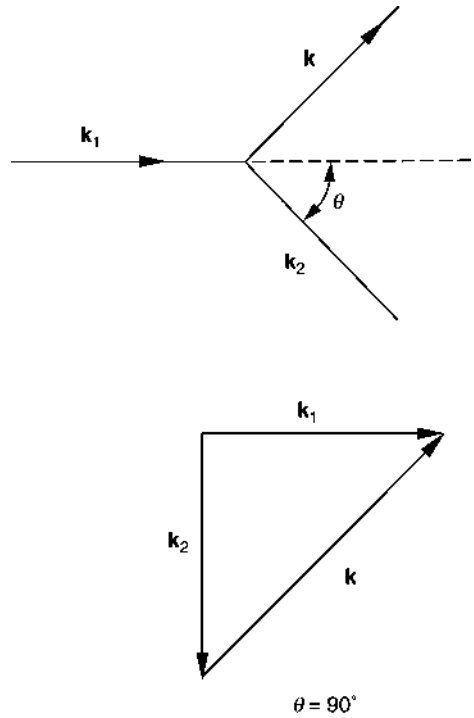

Figure 6 Diagram for Stokes processes.

For typical light-scattering experiments in the visible region the range of the incident wavevector is $0 < k_i < 10^6\ \mathrm{cm}^{-1}$. This implies that for first-order scattering processes the accessible range of $\mathbf{q}$ under conditions of wavevector conservation is small compared to a nonzero reciprocal lattice wavevector. Light-scattering experiments yield the frequencies of optical modes at essentially the center of the Brillouin zone.

The energy and momentum conservation rules have to be modified when the lifetime of the crystal excitations are strongly limited by their decay into other crystal excitations. Momentum conservation breaks down in imperfect crystals, in solids lacking translational symmetry like amorphous materials, and in crystals which are opaque to incident and scattered light.

In those cases where the incident and scattered waves are damped inside the scattering volume, such as occur in small-gap semiconductors that are opaque at the light frequencies involved, $\mathbf{k}_i$ and $\mathbf{k}_s$ are complex. The inelastic scattering is due to excitations having a range of wavevector

$$\Delta q = |\mathrm{Im}\,\mathbf{k}_i| + |\mathrm{Im}\,\mathbf{k}_s|$$

about $\mathbf{q} = \mathrm{Re}(\mathbf{k}_i - \mathbf{k}_s)$. Effects associated with such a wavevector uncertainty have been reported in

Raman scattering spectra of III–V semiconductor compounds.

Second-Order Raman Scattering

Second-order Raman scattering involves two phonons rather than a single phonon as in first-order scattering. The possible second-order processes are the following: two phonons may be created, giving a Stokes component in the scattered light, or one is created and the other destroyed, giving a Stokes or anti-Stokes component, or both may be destroyed giving an anti-Stokes component. For each of these cases there are the possibilities of a pair of first-order electron–phonon interactions or a single second-order electron–phonon interaction.

In the case of second-order Raman scattering the wavevector conservation condition is $\mathbf{q}_1 \pm \mathbf{q}_2 \cong 0$, where $\mathbf{q}_1$ and $\mathbf{q}_2$ are the wavevectors of the phonons involved. This condition places no restriction on the magnitudes of the individual wavevectors, other than $|q_1| \cong |q_2|$, which is in contrast to first-order scattering where $q \cong 0$.

The possible phonon frequency pairings associated with second-order processes are $\omega_1 + \omega_2$ (combination band), $\omega_1 - \omega_2$ (difference band), and $2\omega_1$, (overtone band). The second-order scattering spectrum covers a broad range of frequencies. The overtone spectrum provides a measure of the phonon density of states.

Resonant Light Scattering

General Formulation

Resonant Raman scattering occurs when the incident or scattered photon energy is close to the energy of an intermediate electronic state relative to the ground electronic state. Certain terms in the Raman amplitude then diverge leading to a very large scattering cross-section. Divergences occur in the Stokes spectrum when $h\omega_i = h\omega_a$, and $h\omega_s = h\omega_b$: the photon energies *resonate* with the excitation energies of the intermediate states a or b. If a and b are the same state, resonance occurs nearly simultaneously for both ω_i and ω_s, and gives rise to a particularly strong enhancement of the scattering efficiency. It is thus clear that the resonance phenomenon is specific to the nature of the intermediate state, and its investigation leads to basic information concerning the electronic states of the system. We shall now examine the resonance behavior in several cases where the intermediate states are carrier Bloch states, free excitons, or bound excitons.

Resonance at the Fundamental Bandgap

The divergent behavior of the Raman amplitude is associated with the factor $(h\omega_i - E_g)^{-1}$ where E_g is the lowest direct bandgap of the material.

Resonance at Free Exciton States

The Raman intensity as a function of the incident photon energy has a Lorentzian line shape in the range of a single intermediate state. It is centered at the resonance frequency and has a width determined by the lifetime of the intermediate state. The distinction between scattering at resonance and scattering off resonance in the wings of the Lorentzian is that at resonance the intensity is determined by the exciton lifetime, whereas off resonance it is determined by the frequency separation from resonance. Since in many instances excitons have very long lifetimes, they can lead to a very large Raman cross-section at resonance.

Resonance at Bound Exciton States

Scattering induced by the presence of impurities involving bound exciton states has the following essential features: (a) observation of sharp resonance at the energies of the bound exciton which form discrete levels below the free exciton resonance; (b) dependence upon impurity concentration; (c) involvement of LO phonons of wavevector of the order of the inverse of the impurity state radius. Resonance at bound exciton energies has been observed in impurity-induced resonant Raman scattering by LO phonons in CdS. The resonance is centered at the absorption peak for the impurity state and has approximately Lorentzian shape. If the impurity gives rise to a localized vibrational mode, bound exciton resonance in Raman scattering due to the local mode can occur.

Anharmonic Effects on Raman Spectra

In a perfect crystal whose vibrations are harmonic and which is transparent to the incident and scattered light, the Raman lineshape is a delta function centered on the optical phonon frequency. Experimentally one finds that even in crystals of very high quality, the Raman line is broadened into a roughly Lorentzian shape with a width that increases with increasing temperature. This width can be attributed to anharmonic terms in the vibrational Hamiltonian.

The temperature dependence of the line width is determined by the phonon occupation factors. As the temperature approaches zero, the phonon

occupation factor approaches zero also, and the width approaches the value due to the zero-point motion of the nuclei. At temperatures above the Debye temperature, the width due to cubic anharmonicity becomes proportional to T. Quartic anharmonicity imparts a T^2 dependence to the width which becomes significant at very high temperatures.

In addition to line broadening, anharmonicity causes a shift of the frequency of peak intensity to lower values as the temperature increases. Contributions to the shift proportional to T and T^2 arise in the high-temperature regime from cubic and quartic anharmonicity, respectively.

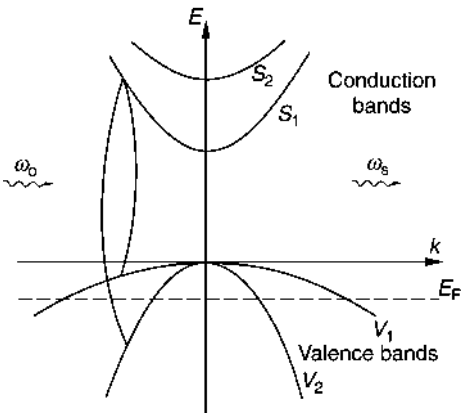

Figure 7 Electronic transitions occurring in intervalence band scattering.

Light Scattering due to Electronic Excitations

Light Scattering by Plasmons

Plasmons are similar to LO phonons in that they have a macroscopic electric field associated with them. The field modulates the electric susceptibility and gives rise to Raman scattering. The difference in frequency of the incident and scattered light is equal to the plasma frequency ω_p specified by $\omega_p = e^2 n/\varepsilon_0 \varepsilon_\infty m^*$, where n is the electron concentration, m^* is their effective mass, and ε^∞ is the high-frequency dielectric constant.

When the plasma frequency is close to the LO phonon frequency, there is a plasmon–phonon interaction via the macroscopic electric fields that leads to coupled modes and forces the frequencies apart.

Light Scattering Associated with Interband Transitions

The plasmons dealt with in the preceding section are an intraband type of excitation. Also of interest are interband transitions in which a carrier undergoes a transition from one band to a different band during the scattering process. An example is provided by p-type silicon in which an electron makes a transition from an occupied state in the light-hole band to an unoccupied state in the heavy-hole band via the conduction band as shown in Figure 7.

For the case of isotropic bands the scattering efficiency spectrum at 0 K has sharp edges. If the anisotropy of the valence bands is taken into account, the sharp edges of the spectrum are smoothed out.

The experimental Raman spectrum for p-type Si exhibits a broad continuum due to the electronic transitions but in the same spectral range there are also Raman active optical phonons which superpose on the electronic continuum and distort it. The resulting distortion of the spectrum is associated with resonant interference that occurs between the discrete phonon line and the electronic continuum.

See also

Scattering: Raman Scattering. **Semiconductor Physics:** Outline of Basic Electronic Properties.

Further Reading

Balkanski M (1972) In: Abeles F (ed.) *Optical Properties of Solids*. Amsterdam: North-Holland.
Balkanski M, Jain KP, Beserman R and Jouanne M (1975) *Phys. Rev.* B12: 4328.
Birman JL (1974) *Encyclopedia of Physics*, vol XXV 12b. Berlin: Springer-Verlag.
Born M and Huang K (1954) *Dynamical Theory of Crystal Lattice*. Oxford, UK: Oxford University Press.
Brillouin L (1922) *Ann. Physique* 17: 88.
Buchler S and Burstein E (1974) *Phys. Rev. Letter* 33: 908.
Daman TC and Shah J (1971) *Phys. Rev. Letter* 27: 1506.
Hayes W and Loudon R (1978) Scattering of Light by Crystals. New York: John Wiley.
Haro E, Balkanski M, Wallis RF and Wanser KH (1986) *Phys. Rev.* B34: 5358.
Kanehisa MA, Wallis RF and Balkanski M (1982) *Phys. Rev.* B25: 7619.
Wallis RF and Balkanski M (1986) Many-Body Aspects of Solid Spectroscopy. Amsterdam: North-Holland.
Yu PY, Pilkhun MH and Evangelisti F (1978) *Solid State Communications*, vol. 25, p. 371.

SEMICONDUCTOR PHYSICS

Polarons

J T Devreese, Universiteit Antwerpen, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

The Polaron Concept

If an electron (or hole) is placed in a polarizable medium (like an ionic crystal or a polar semiconductor), it induces a deformation of polarization field around itself. On the other hand, the electron is attracted to the potential well of this disturbed polarization field. Thus, a feedback loop between the electron and the polarization field arises, which leads to a highly correlated state of the charge carrier and the polarization field. A conduction electron (or hole) together with its self-induced polarization in an ionic crystal or in a polar semiconductor forms a quasiparticle, which is called a polaron. The polaron concept was introduced by Landau in 1933.

The physical properties of the polaron differ from those of the band carrier. In particular, the polaron is characterized by its binding (or self-) energy E_0, effective mass m^* and by its response to external electric and magnetic fields (e.g., dc mobility and optical absorption coefficient).

If the spatial extension of the polaron is large compared to the lattice parameters of the solid, the latter can be treated as a polarizable continuum. This is the case of a 'large polaron', to which most of this article is confined.

The polaron characteristics are determined by the coupling between an electron and the long-wavelength longitudinal optical (LO) phonons. The strength of this coupling is expressed in terms of the constant introduced by Fröhlich:

$$\alpha = \frac{e^2}{\hbar c}\sqrt{\frac{m_b c^2}{2\hbar\omega_{LO}}}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right) \qquad [1]$$

In this definition, ω_{LO} is the frequency of LO phonons, c is the velocity of light, and ε_∞ and ε_0 are, respectively, the electronic and the static dielectric constant of the polar crystal.

In order to find the coupling constant according to eqn [1], it is crucial to know the electron (hole) band mass m_b. This unknown parameter can, in principle, be determined from experiment. However, in ionic crystals and polar semiconductors, experiment measures the polaron effective mass m^* rather than the band mass, because of the polaron effect. The problem of determining the band mass is therefore far from trivial and should be tackled on the basis of the results of polaron theory. In Table 1 the coupling constants are indicated for a number of crystals.

Landau and Pekar were the first to investigate the self-energy and the effective mass of the strong-coupling polaron, when $\alpha \gg 1$. As seen from Table 1, in substances with large values of the electron–phonon coupling constant (like alkali halides, TlBr, CdF_2, $SrTiO_3$) α ranges between 2.5 and 4.5. Though, strictly speaking, polarons in such materials are of intermediate coupling, their properties can be approached within the strong-coupling approximation.

The cases when $\alpha \ll 1$ (like InSb, InAs, GaAs) are referred to as the weak-coupling regime. Polarons in many other substances with $\alpha < 1$ (e.g., GaP, CdTe, ZnSe, CdSe, CdS, CdO) are well described in the weak-coupling approximation.

In the weak-coupling regime, the states of the polaron can be imagined as those of a band carrier perturbed by the electron–phonon interaction. Fröhlich provided in 1954 the first weak-coupling perturbation theory results for the polaron characteristics:

$$E_0 = -\alpha\hbar\omega_{LO} \qquad [2]$$

$$m^* = \frac{m_b}{1 - \alpha/6} \qquad [3]$$

Feynman formulated the polaron problem in the Lagrangian (path integral) form of quantum mechanics and obtained an upper bound for the polaron self-energy at all values of the α, which at weak and strong coupling gave accurate limits. Over the years the Feynman all-coupling model

Table 1 Electron–phonon coupling constants

Material	α	*Material*	α
InSb	0.02	AgCl	1.8
InAs	0.052	KI	2.5
GaAs	0.068	TlBr	2.55
GaP	0.201	KBr	3.05
CdTe	0.286	RbI	3.16
ZnSe	0.43	CdF_2	3.2
CdSe	0.49	KCl	3.5
CdS	0.52	CsI	3.67
CdO	0.74	RbCl	3.81
AgBr	1.6	$SrTiO_3$	4.5

for the polaron has been proven to be in many respects the most successful approach to the polaron problem.

Optical Properties and Internal Structure of Polarons

Optical Absorption of Polarons at Weak Coupling

At zero temperature and in the weak-coupling limit, the optical absorption is due to the following elementary polaron scattering process, schematically shown in Figure 1. An incoming photon is absorbed by a polaron. The polaron emits a phonon during the absorption process and takes recoil energy from the incident light. At zero temperature, the absorption coefficient for absorption of light with frequency Ω can be expressed in terms of elementary functions in two limiting cases: in the region of comparatively low frequencies $(\hbar(\Omega - \omega_{LO})/\zeta \ll 1)$

$$\Gamma(\omega)=\frac{1}{\varepsilon_0 nc}\frac{2^{1/2}N^{2/3}\alpha}{(3\pi^2)^{1/3}}\frac{e^2}{(\hbar m_b\omega_{LO})^{1/2}}\frac{\omega-1}{\omega^3}\Theta(\omega-1) \quad [4]$$

and in the high-frequency region $(\hbar(\Omega-\omega_{LO})/\zeta \gg 1)$

$$\Gamma(\omega)=\frac{1}{\varepsilon_0 nc}\frac{2Ne^2\alpha}{3m_b\omega_{LO}}\frac{(\omega-1)^{1/2}}{\omega^3}\Theta(\omega-1) \quad [5]$$

Here $\omega=\Omega/\omega_{LO}$, ε_0 is the dielectric permittivity of the vacuum, n is the refractive index of the medium, N is the concentration of polarons, and ζ is the Fermi level for electrons. The step function

$$\Theta(\omega-1)=\begin{cases}1 & \text{if } \omega>1\\ 0 & \text{if } \omega<1\end{cases} \quad [6]$$

reflects the fact that at zero temperature absorption of light accompanied by emission of a phonon can occur only if the energy of a photon is larger than that of a phonon ($\omega>1$). In the weak-coupling limit, according to eqn [4], the absorption spectrum consists of a one-phonon line.

Experimentally, the one-phonon line has been clearly seen for free polarons in the infrared absorption spectra of CdO-films (see Figure 2). In CdO, which is a weakly polar material with $\alpha \approx 0.74$, the optical absorption band is observed in the spectral region between 6 and 20 μm (above the LO phonon frequency). As seen from Figure 2, this optical absorption band can be attributed to the weak-coupling polaron absorption as described by eqn [4].

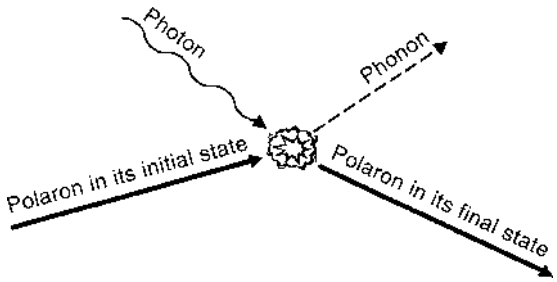

Figure 1 The elementary polaron scattering process leading to absorption of an incoming photon and to generation of an outgoing phonon.

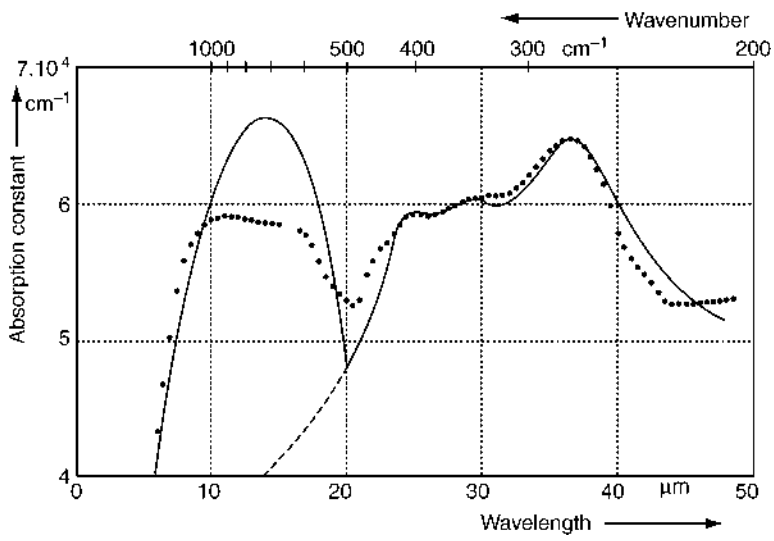

Figure 2 Optical absorption spectrum of a CdO film with the carrier concentration $N = 5.9 \times 10^{19}$ cm^{-3} at $T = 300$ K. The theoretical results are shown with (solid curves) and without (dashed curves) the polaron contribution of eqn [4], and compared to the experimental data (solid dots). (Reproduced with permission from Finkenrath H, Uhle N and Waidelich W (1969) The influence of phonons and polarons on the infrared absorption of cadmium oxide[*1]. *Solid State Communications* 7: 11–14.)

With strengthening coupling, it becomes possible that absorption of a photon is accompanied by emission of two, three, …, K phonons, thus giving rise to two-, three-, …, K-phonon sub-bands in the optical absorption spectra. At high temperatures, absorption of a photon can be accompanied not only by emission, but also by absorption of one or more phonons.

Optical Absorption of Polarons at Arbitrary Coupling

The theoretical absorption spectrum of a single large polaron (at all electron–phonon coupling strengths) is given by the expression

$$\Gamma(\omega) = \frac{1}{n\varepsilon_0 c}\frac{\mathrm{Im}\Sigma(\omega)}{[\omega - \mathrm{Re}\Sigma(\omega)]^2 + [\mathrm{Im}\Sigma(\omega)]^2} \qquad [7]$$

The so-called memory function $\Sigma(\omega)$ contains the dynamics of the polaron and depends on α and ω. This function is related to the impedance function $Z(\omega)$ of the polaron through $-\mathrm{i}Z(\omega) = \omega - \Sigma(\omega)$.

Optical Absorption of Polarons at Strong Coupling

The absorption of light by free large polarons was treated by Kartheuser, Evrard and Devreese using the adiabatic strong-coupling method by Landau and Pekar. The polaron ground state in this scheme has the energy

$$E_0 = -\frac{\alpha^2}{3\pi}\hbar\omega_{\mathrm{LO}} = -0.106\alpha^2\hbar\omega_{\mathrm{LO}} \qquad [8]$$

If the lattice polarization is allowed to relax or adapt to the electronic distribution of the excited electron (which itself then adapts its wavefunction to the new potential, etc., leading to a self-consistent final state), the so-called relaxed excited state (RES) results. Its energy is

$$E_{\mathrm{RES}} = -0.041\alpha^2\hbar\omega_{\mathrm{LO}} \qquad [9]$$

It was argued by Kartheuser, Evrard and Devreese, that for a sufficiently large α ($\alpha > 3$), the (first) RES of a polaron is a stable state, which can participate in optical absorption transitions. This idea was important in order to understand the optical absorption spectrum in the strong-coupling regime. In particular, the following scenario of a transition, which leads to a zero-phonon peak in the absorption by strong-coupling polaron, can be suggested. If the frequency of a photon is equal to

$$\Omega_{\mathrm{RES}} \equiv \frac{E_{\mathrm{RES}} - E_0}{\hbar} = 0.065\alpha^2\omega_{\mathrm{LO}} \qquad [10]$$

then the electron goes from the ground state (following the analogy with the hydrogen atom, let us call it '1s') to an excited state ('2p'), while the lattice polarization in the final state is adapted to the '2p' state of the polaron.

Each RES of the polaron can be accompanied by 1, 2, 3, … free optical phonons in the lattice. Such states are called scattering states (ScS). They are shown in Figure 3. Starting with the photon energy

$$\Omega_{\mathrm{RES}} + \omega_{\mathrm{LO}} \qquad [11]$$

a transition of the polaron towards the first ScS, belonging to the RES, becomes possible. In the course of this transition, while absorbing a photon, the polaron goes from the ground state to the RES with emission of one extra phonon. This process is called one-phonon sideband absorption.

The one-, two-, …, K-phonon sidebands give rise to the broad structure of the absorption coefficient. These transitions occur without lattice relaxation. The (unstable) polaron state, in which the lattice polarization corresponds to the electron ground state, while the electron is excited ('a frozen lattice'), is referred to as the *Franck–Condon* (FC) state of a polaron. Within the adiabatic approach, the energy of the lowest FC state is

$$E_{\mathrm{FC}} = \frac{\alpha^2}{9\pi}\hbar\omega_{\mathrm{LO}} = 0.035\alpha^2\hbar\omega_{\mathrm{LO}} \qquad [12]$$

The superposition of the one-, two-, …, K-phonon sidebands is centered at the FC transition frequency

$$\Omega_{\mathrm{FC}} \equiv \frac{E_{\mathrm{FC}} - E_0}{\hbar} = 0.141\alpha^2\omega_{\mathrm{LO}} \qquad [13]$$

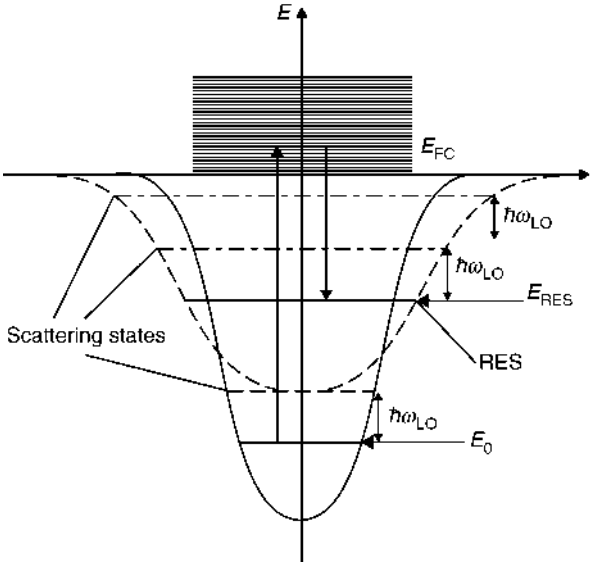

Figure 3 Internal excitations of a polaron at strong coupling: E_0 – the ground state, E_{RES} – the (first) relaxed excited state; the Franck–Condon states (E_{FC}) are shown with a set of horizontal lines above the zero energy. In fact, both the Franck–Condon states and the relaxed excited states lie in the continuum and, strictly speaking, are resonances.

The optical absorption spectrum at strong coupling is, thus, characterized by the following features.

(a) There is an intense absorption peak ('zero-phonon line'), which corresponds to a transition from the ground state to the first RES at Ω_{RES}.
(b) At $\Omega_{RES} + 1$, a phonon sideband structure appears (namely, a K-phonon sideband at $\Omega_{RES} + K$). The maximum of this structure occurs at the FC transition frequency Ω_{FC}.

The qualitative behavior predicted by Kartheuser, Evrard and Devreese, namely, an intense zero-phonon (RES) line with a broader sideband at the high-frequency side, was revealed using a general all-coupling expression for the optical absorption coefficient [7] at $\alpha = 5, 6, 7$.

Internal Structure of Polarons at Arbitrary Coupling

The optical absorption spectra calculated according to the theory of Devreese *et al.* are displayed in Figure 4 for values of α in the range from 1 to 6.

The absorption spectrum for $\alpha = 1$ consists of a one-phonon line, similar to the weak-coupling result. The absorption peak for $\alpha = 3$ might correspond to a one-phonon peak at this intermediate coupling. It was also argued that part of the oscillator strength in the absorption for $\alpha = 3$ is due to lattice relaxation in the final states, be it to a smaller extent than for $\alpha = 5$.

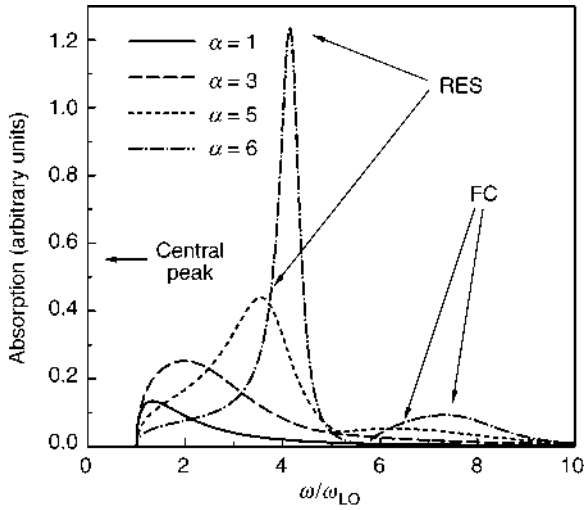

Figure 4 Polaron optical absorption spectra at zero temperature, calculated within the path integral method (according to eqns [11a], [12], [13] of the paper (Devreese JT, De Sitter J and Goovaerts M (1972) Optical absorption of polarons in the Feynman–Hellwarth–Iddings–Platzman approximation. *Physical Review B* 5: 2367–2381)) for different values of the electron–phonon coupling constant α. The peaks are labeled in the picture according to their physical origin as discussed in the text. A δ-like central peak is schematically shown by a vertical line.

It is remarkable that from [7] the three different kinds of polaron excitations appear in the calculated optical absorption spectra for polarons at $\alpha \geq 5$:

- scattering states where, e.g., one real phonon is excited (the structure starting at $\omega = 1$);
- relaxed excited state (RES);
- Franck–Condon (FC) states.

Besides that, at zero temperature, the optical absorption spectrum for one polaron contains a central peak [$\sim\delta(\omega)$] at the origin. For nonzero temperature, this central peak smears out and the optical absorption spectrum consists of both a broad envelope and an anomalous Drude-type low-frequency component.

For example, in Figure 4 the main peak of optical absorption for $\alpha = 5$ at $\omega = 3.51$ is interpreted as due to transitions to RES. The shoulder at the low-frequency side of the main peak is attributed to one-phonon transitions to scattering states. The structure centered at about $\omega = 6.3$ is attributed to a FC band. As seen from Figure 4, when increasing the electron–phonon coupling constant to $\alpha = 6$, the RES peak at $\omega = 4.3$ becomes very intense compared with the FC peak centered at $\omega = 7.5$.

The free-carrier polaron effects were revealed through measurements of optical constants, of cyclotron resonance and of mobility in semiconductors and insulating photoconductors. Experimentally, only the scattering states have been seen for free polarons. However, the full structure of eqn [7] has been revealed through cyclotron resonance measurements.

Magneto-absorption of Polarons

Polaron Cyclotron Resonance

A polaron in a magnetic field is referred to as a *magnetopolaron*. The most powerful technique to study the electron–phonon coupling in semiconductors is cyclotron resonance. Changing the magnetic field B allows one to vary the cyclotron frequency $\omega_c = eB/m_b c$. When it is brought closer to resonance with the phonon frequency ω_{LO}, a strongly renormalized magnetopolaron state is formed. This results in a splitting of the cyclotron resonance, which is a measure of the electron–phonon (polaron) coupling constant α.

A theory of the polaron effects in cyclotron resonance was proposed by Larsen. In particular, the variational approach of Larsen is based on an intermediate-coupling theory to calculate the energy levels (modified Landau levels) of a polaron in a

magnetic field. The polaron mass is then defined from the energy differences between the polaron (Landau-) energy levels.

A rigorous way to find the polaron cyclotron mass is to calculate the magneto-optical absorption spectrum of the polaron (the quantity which is actually measured) and to define the polaron mass, in the same way as the experimentalists do, from the peak positions in the spectrum. The magneto-optical absorption of polarons for all α and ω_c at $T = 0$ was calculated by Peeters and Devreese. They extended the memory-function formalism described above to study the response of a polaron in a magnetic field. The magneto-absorption was then obtained from

$$\lim_{\varepsilon \to 0} \mathrm{Re} \frac{1}{z - \omega_c - \Sigma(z)} \qquad z = \omega + \mathrm{i}\varepsilon \qquad [14]$$

where $\omega_c = eB/m_b$ is the cyclotron frequency for a rigid-lattice band mass m_b, and $\Sigma(z)$ is the memory function, which takes into account all the polaron internal states as well as all the Landau levels and depends on the coupling constant α and on the magnetic field. The explicit results for $\Sigma(z)$ depend on the optimal parameters of the anisotropic Feynman polaron model, which are found by a variational calculation of the polaron ground-state energy.

At $T = 0$ and when $\omega < 1$, it follows that $\mathrm{Im}\Sigma(\omega) = 0$. According to eqn [14], the position of the cyclotron resonance line is determined by the equation $\omega - \omega_c - \Sigma(\omega) = 0$. If solutions of this equation are denoted as $\omega = \omega_c^*$ then the polaron cyclotron mass m_p^* is calculated from the relation $\omega_c^* = eB/m_p^*$.

Cyclotron Resonance of Polarons in Silver Halides

In pioneering experimental studies, Brown and coworkers have combined mobility experiments and cyclotron resonance measurements to clearly demonstrate the polaron effect in AgBr. From a theoretical plot of polaron mobility versus band mass compared to experimental Hall mobility data at a given temperature, they estimate the band mass. This allows them to calculate α and the polaron mass m^*. This value of m^* can then be compared to the measured cyclotron mass in order to test the polaron theory. For AgBr, the experimental value $m^*/m_e = 0.27 \pm 0.01$ obtained from the cyclotron resonance at millimeter wave frequencies and at 18 K is shown to agree well with the value $m^*/m_e = 0.27 \pm 0.05$ determined by comparison of mobility experiment and theory.

Precise cyclotron mass measurements in AgBr and AgCl covered the range from zero magnetic field to 16 T (see Figure 5). Several polaron theories were compared in analyzing these experimental data. It should be pointed out that the weak-coupling theories (Rayleigh–Schrödinger perturbation theory, Wigner–Brillouin and its improvements) fail (and are all off by at least 20% at 16 T) to describe the

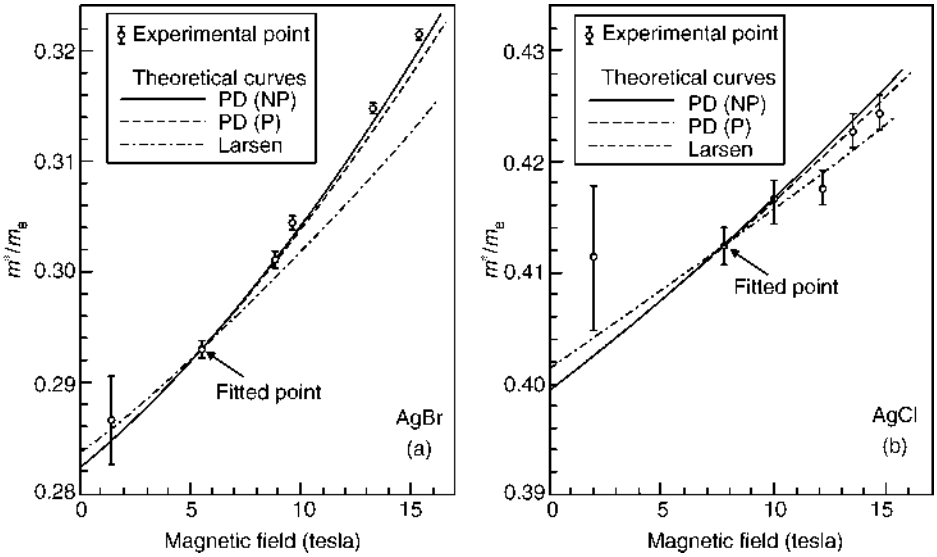

Figure 5 The polaron cyclotron mass in AgBr (a) and in AgCl (b): comparison of experiment and theory. Larsen: (Larsen D (1974) *Journal of Physics C* 7: 2877–2889); PD: (Peeters FM and Devreese JT (1986) *Physical Review B* 34: 7246–7259) (P) – with parabolic band, (NP) – with corrections of the two-band Kane model. In each case the band mass was adjusted to fit the experimental point at 525 GHz. (Reproduced with permission from Hodby JW, Russell G, Peeters F, Devreese JT and Larsen DM (1987) Cyclotron resonance of polarons in the silver halides. *Physical Review Letters* 58: 1471–1474). ©(1987) American Physical Society.

experimental data for the silver halides. The variational approach by Larsen underestimates the polaron cyclotron mass by 2% at 15.3 T. It turns out that the magneto-absorption calculated by Peeters and Devreese leads to the best quantitative agreement between theory and experiment as was analyzed for AgBr and AgCl. This analysis provides a clear confirmation of the Fröhlich description of the polaron in the case where weak-coupling approximations are adequate.

Cyclotron Resonance of Polarons in CdTe

The early infrared transmission study of hydrogen-like shallow donor impurity states in n-CdTe was reported by Cohn, Larsen and Lax (see Figure 6). By studying the Zeeman splitting of the ($1s \rightarrow 2p, m = \pm 1$) transition in the Faraday configuration at magnetic fields up to ~160 kOe, they performed the first quantitative determination of polaron shifts of the energy levels of a bound electron.

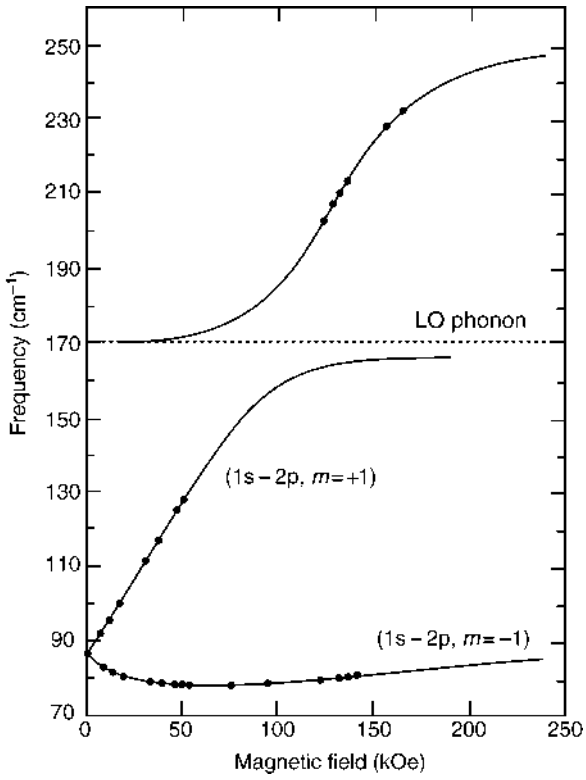

Figure 6 Plot of the experimentally determined magnetic field dependence of the ($1s \rightarrow 2p, m = \pm 1$) transition frequencies in n-CdTe. The solid lines represent the theoretical results for the effects of the electron–LO-phonon interaction upon the shallow donor impurity levels. α treated as a fitting parameter is taken to be 0.4. (Reproduced with permission from Cohn DR, Larsen DM and Lax B (1972) Polaron Zeeman effect in CdTe. *Physical Review B* 6: 1367–1374). ©(1972) American Physical Society.

The experimental data were shown to be in good agreement with the weak-coupling theory of the polaron Zeeman effect. In this comparison, however, the value $\alpha = 0.4$ was used instead of $\alpha = 0.286$, which comes from the definition [1]. Similarly, the value $\alpha \sim 0.4$ was suggested to explain the measured variation of the cyclotron mass with magnetic field in CdTe. This discrepancy gave rise to a discussion where important contributions have been, in particular, due to Harper and Zawadzki.

Grynberg *et al.* applied the far-infrared photoconductivity technique to study the energy spectrum of shallow In donors in CdTe layers and obtained experimental data over the energy range relevant to study the magnetopolaron effect, as shown in Figure 7. An overall good agreement is found between experiment and the theoretical approach, where the electron–phonon interaction is treated within a second-order improved Wigner–Brillouin perturbation theory and a variational calculation is performed for the lowest-lying donor states ($1s$, $2p^{\pm}$, $2s$, $2p_z$, $3d^{-2}$, $4f^{-3}$). It is to be stressed that this agreement is achieved with the coupling constant $\alpha = 0.286$ as follows from eqn [1] with the known material parameters of CdTe.

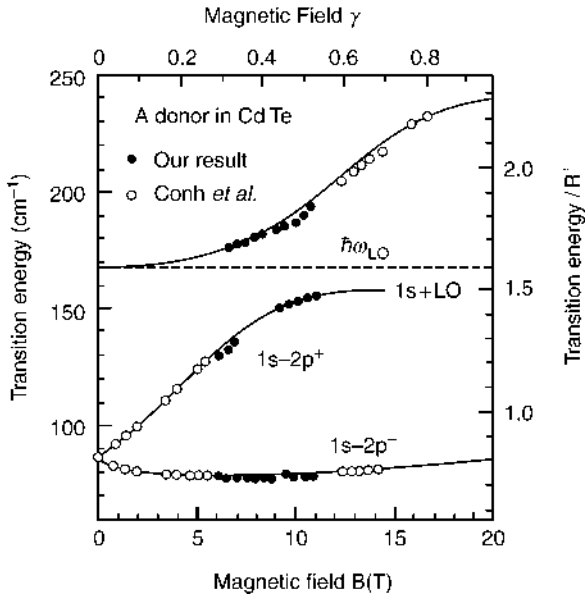

Figure 7 Plot of the experimentally determined magnetic field dependence of the $1s \rightarrow 2p^{\pm 1}$ transition energies in CdTe layers grown by molecular beam epitaxy. The solid lines represent the results of the calculation described in the text without any fitting parameters. The solid dots are the experimental data of the present work and the open circles represent the data of Cohn, Larsen and Lax. (Reproduced with permission from Grynberg M, Huant S, Martinez G, *et al.* (1996) Magnetopolaron effect on shadow indium donors in CdTe. *Physical Review B* 54: 1467–1470). ©(1996) American Physical Society.

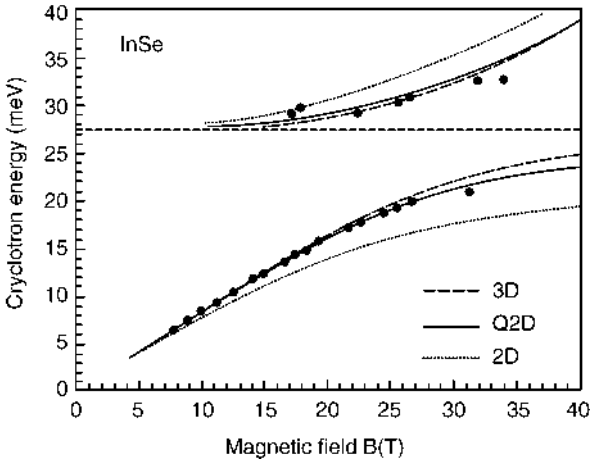

Figure 8 The cyclotron resonance position plotted as a function of magnetic field for InSe. (Reproduced with permission from Nicholas RJ, Watts M, Howell DF, *et al.* (1992) Cyclotron resonance of both magnetopolaron branches for polar and neutral. *Physical Review B* 45: 12144–12147). ©(1992) American Physical Society.

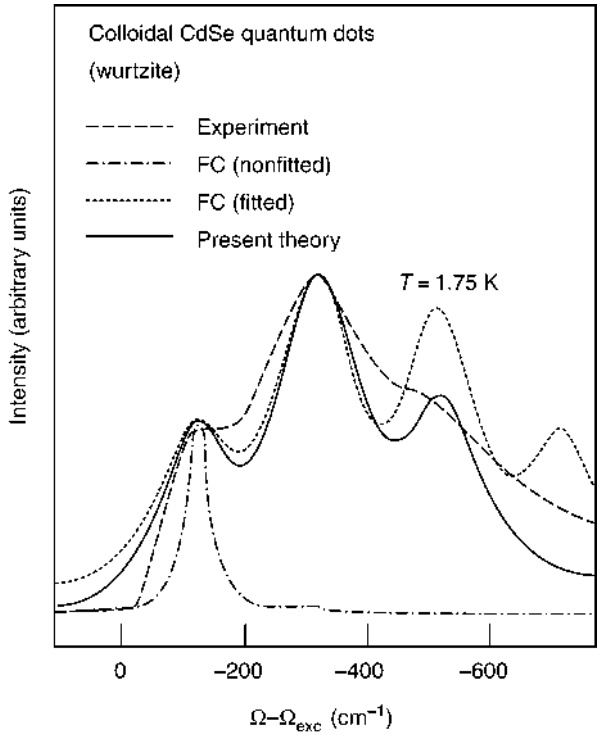

Figure 9 Fluorescence spectra of CdSe quantum dots with wurtzite structure at the average radius $\langle R \rangle = 1.25$ nm. The dashed line represents the experimental data of Nirmal M, Murray CB, Norris DJ and Bawendi MG (1993) *Z. Phys. D* 26: 361–363, the dot-dashed line displays a Franck–Condon progression with the Huang–Rhys parameter $S = 0.06$ calculated using the adiabatic approximation, the dotted line shows another Franck–Condon progression with the Huang–Rhys parameter $S = 1.7$, which is obtained by fitting the ratio of one-phonon and zero-phonon peak heights to the experimental value, and the solid line results from the nonadiabatic theory. (Reproduced with permission from Fomin VM, Gladilin VN, Devreese JT, *et al.* (1988) Photoluminescence of spherical quantum dots. *Physical Reviews B* 57: 2415–2425). ©(1998) American Physical Society.

Polarons in the Layered Compound InSe and in Superlattices

Nicholas *et al.* provided a clear demonstration of the polaron coupling by the cyclotron resonance in a two-dimensional electron gas (2DEG), which naturally occurs in the polar semiconductor InSe. One clearly sees, over a wide range of magnetic fields ($B = 18-34$ T), the two distinct magnetopolaron branches separated by as much as 11 meV ($\sim 0.4\omega_{LO}$) at resonance (Figure 8). The lines show the results of theoretical calculations for coupling to the LO phonon in bulk (3D), sheet (2D) and after correction for the quasi-2D systems at $\alpha = 0.29$ (for motion perpendicular to the c-axis). The agreement is reasonable for the 3D case, and is better for the quasi-2D system, where the finite spatial extent of the 2D electron gas in the symmetric planar layer is taken into account.

For GaAs/$Al_xGa_{1-x}As$ quantum wells and superlattices, the polaron effect is found to decrease the energy of the shallow donor states at low magnetic fields and to lead to a resonant splitting of the energies at high magnetic fields. The results are in very good agreement with available experimental far-infrared optical data in the whole magnetic-field range.

Optical Properties of Quantum Dots: Effects of the Polaron Interaction

In order to interpret the phonon-assisted optical transitions in semiconductor quantum dots, a theory was developed by Fomin, Gladilin, Devreese *et al.* comprising the exciton interaction with both adiabatic and Jahn–Teller phonons, and also the external nonadiabaticity (pseudo-Jahn–Teller effect). The effects of nonadiabaticity of the exciton–phonon system are shown to lead to a significant enhancement of phonon-assisted transition probabilities and to multiphonon optical spectra which are considerably different from the Franck–Condon progression (see Figure 9). The calculated relative intensity of the phonon satellites and its temperature dependence compare well with the experimental data on the photoluminescence of CdSe quantum dots, both colloidal and embedded in glass. These results clearly demonstrate a breakdown of the adiabatic approximation when describing the polaron interaction in quantum dots.

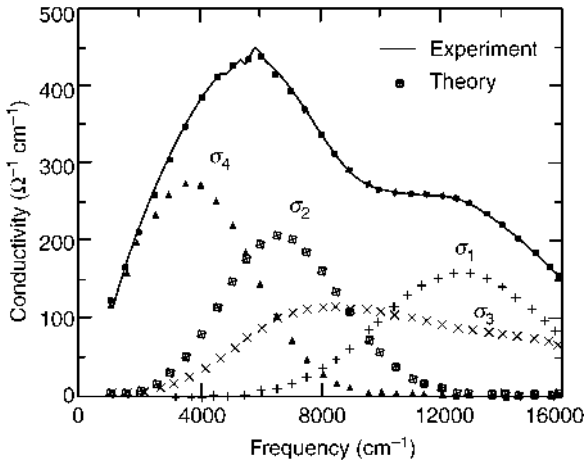

Figure 10 Experimental and theoretical conductivity data at 250 K for $Pr_2NiO_{4.22}$ and four contributions to the theoretical curves resulting from the transitions between the following polaron states: σ_1 – small → small; σ_2 – large → small; σ_3 – small → large; σ_4 – large → large. (Reproduced with permission from Eagles DM, Lobo RPSM and Gervais F (1995) Infrared absorption in oxides in the presence of both large and small polarons. *Physical Review B* 52: 6440–6450). ©(1995) American Physical Society.

Small Polarons

An electron or a hole trapped by its self-induced atomic (ionic) displacement field in a region of linear dimension, which is of the order of the lattice constant, is called a 'small polaron'. As distinct from large polarons, small polarons appear due to short-range forces.

Experimentally small-polaron effects have been analyzed, e.g., in KCl, LiF, NiO, MnO, TiO_2, $BaTiO_3$, $SrTiO_3$, $LaCoO_3$. More recently, Alexandrov and Mott (1996) surveyed both the principles and the main results of the small-polaron theory in the context of the (bi)polaronic approach to describe the physics of high-T_c superconductors.

Spectroscopic manifestations of both large and small polarons have been found by Eagles *et al.* in the infrared optical absorption spectra of $Pr_2NiO_{4.22}$. The infrared reflectivity spectra on a crystal of this material were measured between 30 and 20 000 cm^{-1}. Mid-infrared conductivity data for frequencies between 1000 and 16 000 cm^{-1} were obtained by Kramers–Kronig transformation of those reflectivity spectra. The resulting optical absorption spectrum was quantitatively interpreted (see Figure 10) using the assumption that in $Pr_2NiO_{4.22}$ two types of polarons, small and large, coexist but do not mix with each other. The contribution due to the transitions between the large-polaron states, which dominates the absorption in the low-frequency region, was calculated using the arbitrary coupling theory by Devreese *et al.* for the coupling constant $\alpha = 3$.

In summary, the polaron concept has been proven to be one of the corner-stones of modern solid-state optics.

Acknowledgments

This work has been supported by the BOF NOI (UA-UIA), GOA BOF UA 2000, IUAP, FWO-V projects G.0287.95, 9.0193.97, G.0274.01 and the W.O.G. WO.025.99N (Belgium).

See also

Semiconductor Materials: Quantum Dots.

Further Reading

Alexandrov AS and Mott N (1996) *Polarons and Bipolarons*. Singapore: World Scientific.

Devreese JT (ed.) (1972) *Polarons in Ionic Crystals and Polar Semiconductors*. Amsterdam: The Netherlands North-Holland.

Devreese JT (1996) Polarons. In: Trigg GL (ed.) *Encyclopedia of Applied Physics*, vol. 14, pp. 383–413. Weinheim: VCH.

Devreese JT, De Sitter J and Goovaerts M (1972) Optical absorption of polarons in the Feynman–Hellwarth–Iddings–Platzman approximation. *Physical Review B* 5: 2367–2381.

Feynman RP (1955) Slow electrons in a polar crystal. *Physical Review* 97: 660–665.

Kartheuser E, Evrard R and Devreese J (1969) Mechanism of absorption of light by free continuum polarons. *Physical Review Letters* 22: 94–97.

Kuper CG and Whitfield GD (eds) (1963) *Polarons and Excitons*. Edinburgh, UK: Oliver and Boyd.

Landwehr G and Rashba EI (eds) (1991) *Landau Level Spectroscopy*, 2 vols. Amsterdam, The Netherlands: North-Holland.

Peeters FM and Devreese JT (1986) Magneto-optical absorption of polarons. *Physical Review B* 34: 7246–7259.

Pekar SI (1951) *Issledovanija po Ekektronnoj Teorii Kristallov*. Moscow: Gostekhizdat (in Russian); German translation: (1954) *Untersuchungen über die Elektronentheorie der Kristalle*. Berlin: Akademie Verlag.

Quantum Wells and GaAs-Based Structures

P Blood, Cardiff University, Cardiff, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

A quantum well is a potential minimum within a semiconductor structure which is sufficiently thin to localize charge carriers on a length-scale similar to their de Broglie wavelength which, for an electron in GaAs at room temperature, is about 30 nm. When electrons are localized in this way their electronic and optical properties are determined by quantum mechanical aspects of their behavior which are not apparent in larger-scale structures. The potential well is usually formed by a sandwich of a thin layer of narrow-gap semiconductor between two layers of a wider gap material as depicted in Figure 1. Typical quantum wells have widths of about 5 nm and the key to their routine production has been the development of advanced epitaxial semiconductor crystal growth techniques such as metalorganic vapor phase epitaxy (MOVPE) and molecular beam epitaxy (MBE). The most significant application of these structures has been in opto-electronic devices, especially laser diodes, and in this article I describe how electrons behave in quantum wells with particular reference to optical properties.

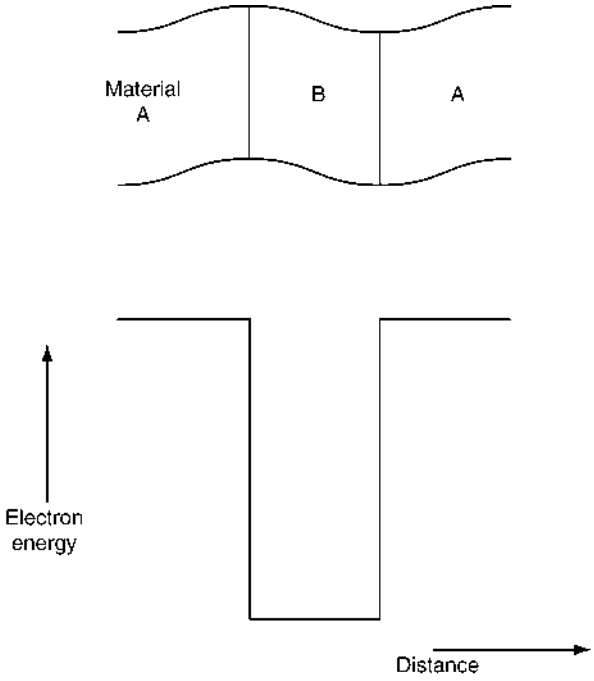

Figure 1 Electron energy diagram illustrating the formation of a potential well by sandwiching a layer of narrow gap semiconductor material (B) between two wider gap layers (A).

The optical properties of semiconductors are usually considered in terms of transitions of electrons between quantum mechanical energy levels, as in the Bohr atom. An alternative description is to consider the spatial displacement of electrons in the oscillating electric field of the electromagnetic radiation. These two descriptions are linked through Schrödinger's equation. For each energy state there is an associated wavefunction which determines the spatial probability distribution of electrons in the state. Consequently, a change in quantum state implies a change in spatial distribution and energy of the electron system. Both approaches are used: here I adopt the viewpoint of transitions between energy states.

We begin with a quantum mechanical description of electrons in a 'bulk' material, as appears in many textbooks. Here electrons are constrained by a three-dimensional potential well of extent given by the sample dimensions, typically millimeters or centimeters and therefore large compared with the de Broglie wavelength. Then we examine what happens when the size of the sample is reduced in one direction: in the limiting case of a quantum well electrons are able to move in only two directions. Quantum confinement is also possible in two or three dimensions to produce a quantum wire or quantum dot, respectively, and it is easy to use the concepts developed in this article to determine the properties of such systems.

Electron States in Bulk Material

We first consider the behavior of an electron in one direction (the x-direction) within a potential well with large dimensions (side length L), then generalize this to three directions (see Figure 2). The wavefunction ψ_n and energy E_n (measured with respect to the bottom of the well) are given by solutions of Schrödinger's equation, which within the well is:

$$\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi_n}{\mathrm{d}x^2} = E_n \qquad [1]$$

It is assumed that the potential is infinitely deep compared with the energy of the electrons.

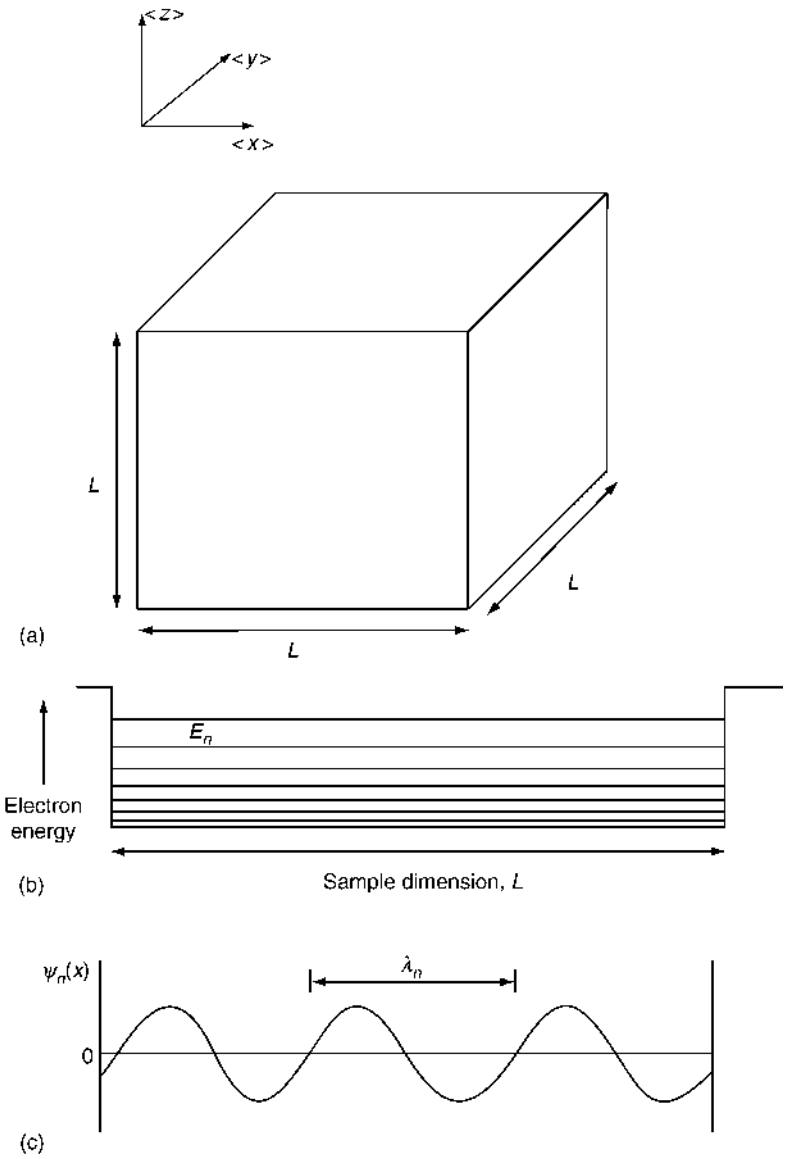

Figure 2 The properties of electrons in a large cubic sample (a) can be obtained by solving Schrödinger's equation for the potential well formed by the sample in each direction (b). This potential well keeps the electrons within the material. The wavefunctions $\psi(x)$ shown in part (c) satisfy cyclic boundary conditions as described in the text, leading to a series of electron energy levels, E_n, given by eqn [3], shown in (b) of the figure.

Within the sample the electrons are described by plane waves of the form $\psi_n = A_n \sin(k_n x)$, where k_n is a wavevector ($= 2\pi/\lambda_n$, where λ_n is the wavelength), and substitution into eqn [1] gives the corresponding energy eigenvalues as

$$E_n = \frac{\hbar^2 k_n^2}{2m} \qquad [2]$$

where m is the mass of the electron. [Substituting $E_n = k_B T$ at room temperature (0.025 eV) and using an effective mass of $0.067m_0$ (for GaAs) in eqn [2] gives $k_n = 2.1 \times 10^8\ \mathrm{m}^{-1}$ and $\lambda = 30$ nm.] Applying cyclic boundary conditions which permit traveling-wave solutions of eqn [1] the wavelength must satisfy the condition $\lambda_n = nL$, i.e., $k_n = 2\pi n/L$, as illustrated in Figure 2, where n is an integer. The dimensions of a typical sample are much greater than the 'size' of an electron so n can be a very large number. The energy eigenvalues are therefore given by

$$E_n = \frac{\hbar^2}{2m}\left(\frac{2\pi n}{L}\right)^2 \qquad [3]$$

This treatment can be extended by solving eqn [1] for motion in three orthogonal directions, x, y, and z. In this case electron motion is described by a wavevector $\mathbf{k}$ comprising components k_x, k_y, k_z, each of which satisfies the cyclic boundary condition in the respective direction. We take the sample to be of dimension L in each direction with no loss of generality. Thus (representing unit vectors by $\langle\mathbf{x}\rangle$, etc.)

$$\mathbf{k} = k_x\langle\mathbf{x}\rangle + k_y\langle\mathbf{y}\rangle + k_z\langle\mathbf{z}\rangle = \frac{2\pi n_x}{L}\langle\mathbf{x}\rangle + \frac{2\pi n_y}{L}\langle\mathbf{y}\rangle + \frac{2\pi n_z}{L}\langle\mathbf{z}\rangle \quad [4]$$

the energy is

$$E = \frac{\hbar^2}{2m}\left\{k_x^2 + k_y^2 + k_z^2\right\} \quad [5]$$

and the wavefunction takes the vector form

$$\psi_k(\mathbf{r}) = A_k \sin(\mathbf{k}\cdot\mathbf{r}) \quad [6]$$

$\psi_n^2(\mathbf{r})$ represents the probability of an electron being at the location $\mathbf{r}$. Each allowed electron state obtained by solution of Schrödinger's equation is specified by a unique combination of the numbers $(n_x\, n_y\, n_z)$, which take both positive and negative values corresponding to plane waves traveling in positive and negative directions. The Pauli exclusion principle states that only one electron of each spin can occupy a given quantum state so the amplitude of each wavefunction is normalized such that

$$\int \psi_k^2(\mathbf{r})\,\mathrm{d}\mathbf{r} = \int [A_k \sin(\mathbf{k}\cdot\mathbf{r})]^2\,\mathrm{d}\mathbf{r} = 1 \quad [7]$$

where the integrals are evaluated over the volume of the sample. Since the amplitude of these solutions is constant throughout the sample electrons may be anywhere in the sample with equal probability. The motion of an *individual* electron of momentum $\hbar\mathbf{k}$ is represented by a wavepacket formed by combination of a number of wavefunctions having similar values of $\mathbf{k}$.

Because the sample dimensions are large the energies of the states allowed by Schrödinger's equation take a series of very closely spaced values for increasing integer values of n as shown in Figure 2b. Substituting a typical sample size of $L = 1$ mm into eqn [3], the $n = 1$ and 2 states are separated by only 7×10^{-11} eV in GaAs and these values are very small compared with thermal energies: at room temperature, $k_\mathrm{B}T = 0.025$ eV. The allowed values of $\mathbf{k}$ and E are discrete and there is a finite but very large number of allowed states in the sample. Since these are very closely spaced we can regard $\mathbf{k}$ and E as continuous variables and the allowed energy states form a continuum. We can calculate the number of allowed energy states $\mathrm{d}N$ in a small energy interval, $\mathrm{d}E$, and hence determine the density of states in energy $g = (\mathrm{d}N/\mathrm{d}E)$ (the number per unit energy interval) at any value of energy.

Electron States in Quantum Wells

Now imagine a sample where the length of one side is reduced in the z-direction (L_z), as illustrated in Figure 3. Gradually the allowed values of k_z become more widely spaced as a consequence of the boundary conditions (eqn [4]). Eventually the width of the potential well becomes similar to the wavelength of the plane-wave state corresponding to the lowest energy level in the well. In these circumstances electron motion is not possible in the z-direction and the wavefunctions become standing waves with boundary conditions $n(\lambda_n/2) = L_z$ in an infinitely deep potential (where n_z takes only positive values) and the amplitude of the wavefunction outside the well must be zero. The lowest energy state, given by eqn [5] with $n_x = n_y = n_z = 1$, no longer lies near the bottom of the potential well. Since the energies associated with motion in the x and y directions remain very small (of order 10^{-10} eV derived earlier) because L_x and L_y are large, the energy of the lowest state is effectively determined by the z-dimension because it is very small, so eqn [5] gives

$$E_{n_z} = \frac{\hbar^2}{2m}\left(\frac{n_z\pi}{L_z}\right)^2 \quad [8]$$

for the lowest energy state in the well. For $L_z = 5$ nm (and $m = 0.067$) E_1 is 0.22 eV above the bottom of the well. Equation [8] shows that this energy can be changed by choice of the well thickness. A simple interpretation of this behavior is as follows. When L_z becomes very small it is no longer possible for the z-component of the wavefunction of the lowest energy state to satisfy the boundary condition of zero amplitude at opposite sides of the sample. The condition can only be satisfied by wavefunctions which have a smaller wavelength, and consequently a higher energy.

Electrons can occupy states defined by all values of $(n_x\, n_y\, n_z)$, thus for $n_z = 1$ there is a continuum of allowed states at increasing energies corresponding to increasing values of $(n_x\, n_y)$ and corresponding to motion in the (x, y) plane (Figure 4a). Since L_x and L_y are both large these states are closely spaced and form a continuum. There is a similar continuum of energy

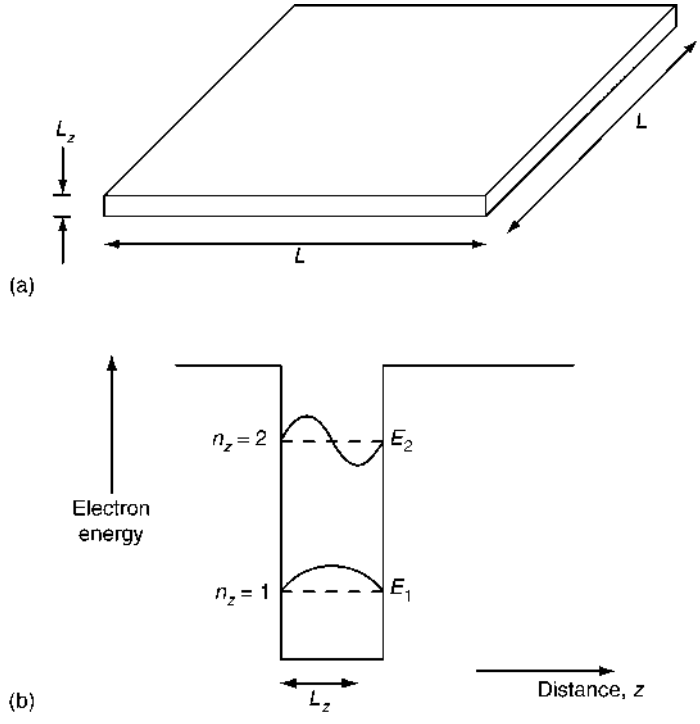

Figure 3 Electron energy diagram (b) for a sample (a) in which one dimension, L_z, is very small. The electron motion in the z-direction is constrained and the associated energy levels become widely spaced, defined by the integers $n_z = 1, 2, \ldots$ in eqn [8]. The wavefunctions in the z-direction, $\psi(z)$, are also illustrated for the first two electron levels.

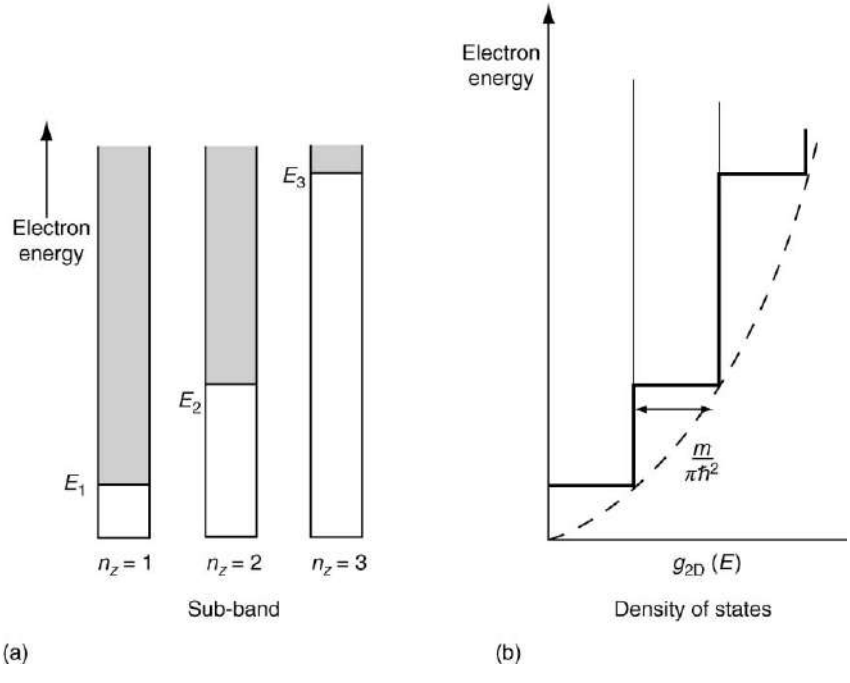

Figure 4 (a) Electron energy diagram showing the energy states associated with unconstrained motion in the (x, y) plane for each sub-band formed by localizing the electrons in the z-direction, defined by $n_z = 1, 2, 3$, etc. When all these allowed states are summed at any energy and expressed as a number of states per unit energy interval we obtain the density of states function $g_{2D}(E)$ shown in (b). This has a series of steps corresponding to each sub-band edge.

states above the energy levels for $n_z = 2, 3$, etc. Thus the energy level diagram is a series of sub-bands, each defined by n_z and with a continuum of states associated with motion in two dimensions in the (x, y) plane, as shown in Figure 4. The number of continuum energy states in a given sub-band in a small energy interval gives the density of states for both spin directions

$$g_{2D}(E) = \frac{m}{\pi\hbar^2} \qquad [9]$$

per unit energy interval per unit sample area.

The density of states is independent of n_z and therefore the same for all sub-bands, and is independent of energy. The density of states function is therefore a series of steps of height given by eqn [9] at each sub-band energy as illustrated in Figure 4b. The density of states per unit area within a given sub-band is independent of L_z as a consequence of dealing with a two-dimensional system.

In real samples the quantum well is not infinitely deep: typically the well depth is in the range 100–400 meV, consequently electrons are not totally confined to the well but are able to penetrate into the barrier material by quantum mechanical tunneling. Solution of Schrödinger's equation for a finite well (treated in many quantum mechanics textbooks) again yields a series of sub-bands, finite in number and with an energy spacing smaller than that given by eqn [8]. In a typical GaAs well the lowest energy state may be about 100 meV from the bottom of the well. The density of states in each sub-band, which is due to the (x, y) motion, remains unchanged. The wavefunction in the z-direction $F(z)$ comprises a sine wave-like standing wave within the well and an exponentially decaying wave in the barrier representing tunneling (Figure 5). The wavefunction is therefore made up of (x, y) plane-wave components and the localized z component and is of the form

$$\Psi_k(\mathbf{r}) = A_k \sin(\mathbf{k}_{xy} \cdot \mathbf{r}_{xy}) F(z) \qquad [10]$$

where $\mathbf{k}_{xy}$ and $\mathbf{r}_{xy}$ are the wavevector and position vector in the (x, y) plane. This wavefunction is again normalized according to eqn [7] such that each state is occupied by one electron of a given spin direction. We see from Figure 5 that we cannot regard the electron as being localized in the well: the electron distribution in the z-direction is specified by the probability distribution $F^2(z)$.

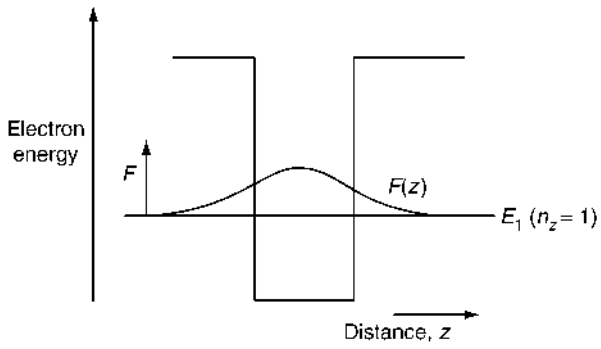

Figure 5 Illustration of the envelope wavefunction $F(z)$ in the z-direction for an electron localized in the $n_z = 1$ sub-band of a well of finite depth. The electron is able to penetrate the barrier by tunneling and this is represented by the decaying part of the wavefunction outside the well.

Occupancy of States in Quantum Wells

The density of states function describes the energy distribution of allowed states in the quantum well. Whether or not particular states are occupied by electrons is determined by the electron concentration and the temperature through the thermodynamic properties of the system. In most cases electron–electron scattering is sufficiently rapid to bring the electron population into an internal equilibrium. Furthermore the interaction between the electrons and the crystal lattice is able to redistribute energy between the two systems so that the lattice and the electron distribution have the same temperature. The electron distribution can therefore be described by Fermi–Dirac statistics for which the probability of a state at energy E being occupied with an electron is

$$f(E) = \frac{1}{1 + \exp\left\{\dfrac{E - E_f}{k_B T}\right\}} \qquad [11]$$

where E_f is the chemical potential (which can be equated with the Fermi level for electrons), T is the lattice temperature and all energies are positive quantities measured with respect to the same arbitrary zero. The energy distribution of electrons in the well, $n(E)$, is then the product of the density of states function, $g_{2D}(E)$, and the occupation probability, $f(E)$, as depicted in Figure 6. The total number of electrons, n, is the integral over energy. Combining eqns [9] and [11], for a single sub-band gives:

$$n_a = \frac{mk_B T}{\pi\hbar^2} \ell n\left\{1 + \exp\left(-\frac{E_n - E_f}{k_B T}\right)\right\} \qquad [12]$$

per unit area, single sub-band.

We cannot specify the carrier density in the well 'per unit volume' since the extent of the distribution in the z-direction is not defined because electrons

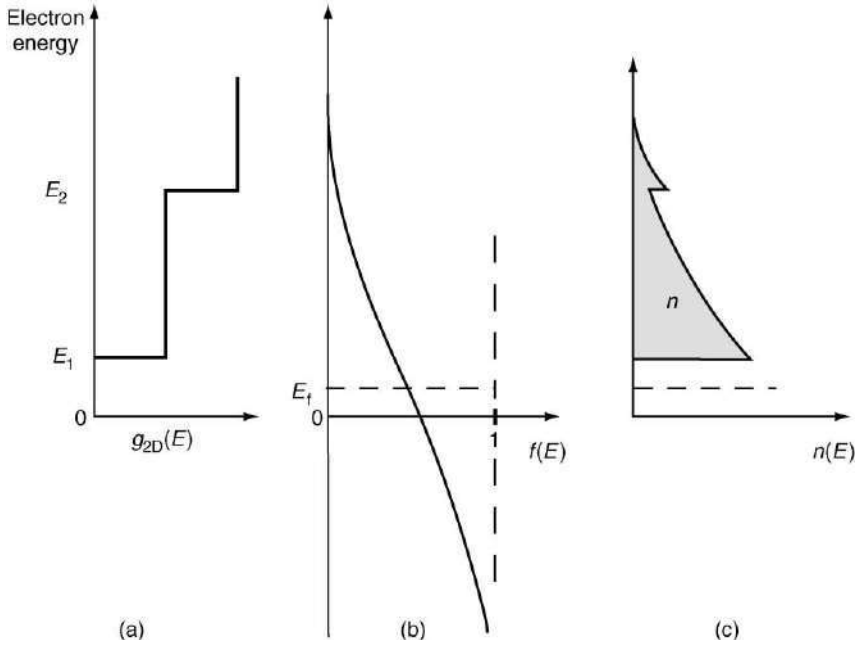

Figure 6 (a) The density of states function for a quantum well, with the value of $g_{2D}(E)$ on the horizontal axis indicating how the density of states varies with increasing electron energy. (b) is an illustration of the probability that a state is occupied by an electron, showing how $f(E)$ varies with increasing electron energy, given by eqn [11]. The actual density of electrons in a given energy interval at any energy $n(E)$ is given by the product of the density of available states and the probability that the state is occupied as shown in (c). The integral of $n(E)$ over energy gives the total density of electrons given for each sub-band by eqn [12].

may tunnel into the barrier material. The electron distribution is properly specified as a number per unit area having a probability distribution in the z-direction given by the function $F(z)$. Where more than one sub-band is populated the total electron density is obtained by adding the contributions from each sub-band each given by eqn [12] with the same Fermi energy and the appropriate sub-band energy. We always use Fermi factors to specify the probability of occupation of a state by an *electron*. The probability that a state is empty is $(1 - f)$.

Formation of Quantum Wells

A quantum well is formed by sandwiching a narrow gap semiconductor layer (E_{g1}) between two layers of wider gap material (E_{g2}). To avoid the formation of defects and dislocations, which have a deleterious effect on many properties of the structure, the constituent materials must have the same lattice parameter. Figure 7a is an energy band diagram for an n-type double heterostructure with layers of micron dimensions. At each interface there are discontinuities in the conduction and valence bands, ΔE_c and ΔE_v, respectively, which account for the difference in the band gaps of the two materials, ΔE_g. While the value of ΔE_g is known for any system, the manner in which this difference is apportioned between the two band edges is not known a priori and the values of ΔE_c and ΔE_v must be obtained from experiment. The discontinuities are often expressed as fractions Q_c, Q_v of the band gap difference:

$$\Delta E_c = Q_c \Delta E_g, \qquad \Delta E_v = Q_v \Delta E_g \qquad [13]$$

In equilibrium the Fermi energy is constant across the diagram and there is a band-bending region on each side of the discontinuity such that the conduction band edges return to their equilibrium energy values at large distances from the interface. The extent of these regions is determined by the discontinuity and the doping level and is typically 200–1000 nm. At each single interface a triangular well is formed which localizes electrons and whose precise shape is determined by the doping density, the carrier distribution and discontinuity through simultaneous solution of Schrödinger's equation and Poisson's equation. Single heterointerfaces only localize the majority carrier. Quantum confinement of both types of carrier can be achieved by a very thin narrow gap semiconductor layer.

When the layer of narrow gap material is made thinner than the band-bending regions, the

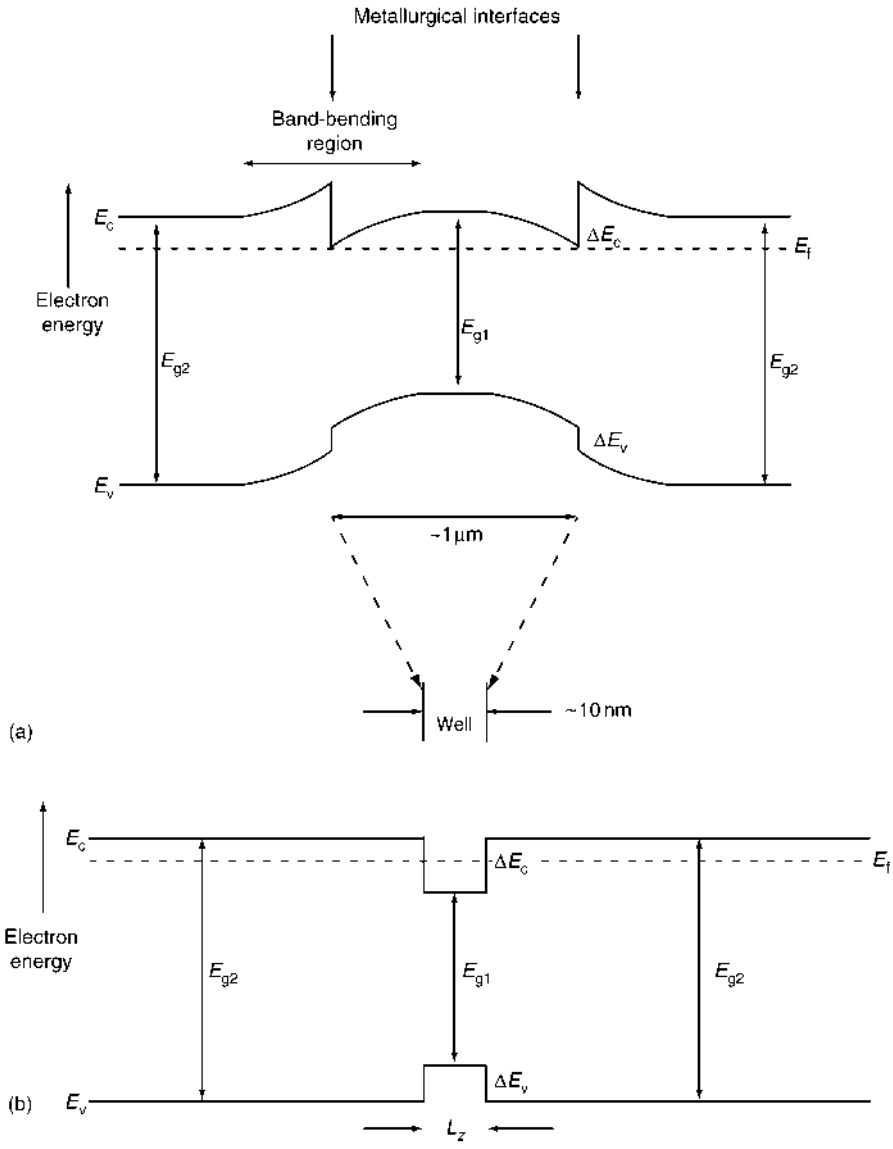

Figure 7 Electron energy diagrams showing the formation of a quantum well as a very thin double heterostructure made up of materials of band gaps E_{g1} and E_{g2}. When the layers are thick the energy diagram is influenced by band bending at the interfaces as shown in (a). When the narrow gap layer is made thinner than the band-bending distances a rectangular well is formed in the conduction and valence bands with depths determined by the respective band offset as shown in (b). On the small distance scale of the well the bands are flat in each region of the structure.

conduction band edge in this layer is not able to regain its bulk equilibrium position and the two triangular wells coalesce as illustrated in Figure 7b. When the well is sufficiently thin, say L_z less than 20 nm, the band-bending across this layer is negligibly small and the well becomes effectively rectangular. This well accumulates electrons from the surrounding barrier material and its potential relative to the Fermi level rises such that there is band bending in the barrier material on each side of the well inhibiting further accumulation. Since the band bending in the barrier occurs on a distance of hundreds of nanometers (being determined by the doping density) it is not apparent over the distance scale of the diagram of the quantum well, which is on the order of 10 nm. Consequently, as shown in Figure 7b, the well can be drawn as a rectangle to a good degree of approximation. This thin, narrow gap layer localizes both electrons and holes. The form of the well is determined by the respective band discontinuities, ΔE_c and ΔE_v, and the thickness of the layer.

Potential wells formed in this way are the basis for most quantum-confined device structures.

Other forms are possible. In certain material combinations ΔE_c or ΔE_v may be negative, meaning that, at the interface, both discontinuities are in the same sense and only one carrier type is confined in the thin narrow gap material. These are known as 'type II' structures. If there is an electric field across the well due to doping or strain (piezoelectric) effects the well becomes triangular. Wells can also be engineered with steps in the potential profile and multiple well systems can be grown in which the states are quantum mechanically coupled. All these variants provide opportunities to engineer the properties of the structure.

One important development of the rectangular quantum well deserves mention. It is possible for the well material to have a different lattice parameter to that of the surrounding barrier material provided that the strain energy can be accommodated elastically within the structure. This means that the strain energy in the layer must be below the energy necessary for formation of dislocations and this translates into an upper limit on the layer thickness (the critical thickness) for a given mismatch. Incorporation of elastic strain is significant because relaxing the lattice match requirement widens the choice of well and barrier materials and because strain modifies the properties of the electronic states in the quantum well and these effects can be used to advantage in the design of devices.

Here we concentrate on the rectangular potential well. Such wells are widely used, particularly in opto-electronic devices, and this structure provides the basic concepts which lie at the heart of all quantum-confined systems. We consider the 'model' lattice-matched quantum well system: a GaAs well bounded by AlGaAs barriers.

GaAs/AlGaAs Quantum Well Structures

As Al is substituted for Ga in the $Al_xGa_{1-x}As$ alloy system the direct band gap increases from 1.424 eV in GaAs ($x = 0$) to 3.018 eV in AlAs ($x = 1$) (Figure 8) while the lattice parameter remains unchanged. A quantum well is formed using a narrow GaAs layer sandwiched between barrier layers of $Al_xGa_{1-x}As$ having a composition chosen to give the desired well depth. The whole structure is grown on a GaAs substrate. The band discontinuity ratio between GaAs and $Al_xGa_{1-x}As$ is independent of alloy composition (ΔE_c and ΔE_v are constant fractions of ΔE_g as x is varied) with $Q_c = 0.66$ and $Q_v = 0.33$. For a typical barrier composition of $x = 0.3, \Delta E_g = 0.374$ eV and ΔE_c and ΔE_v are 0.247 eV and 0.127 eV, respectively.

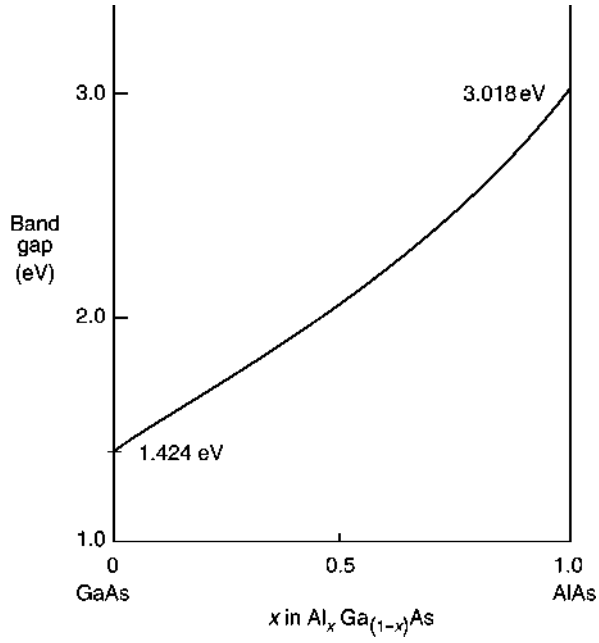

Figure 8 Variation of the direct band gap of $Al_xGa_{(1-x)}As$ with Al content (x).

Optical Properties of AlGaAs/GaAs Quantum Wells

General Principles

Figure 9 illustrates a transition of an electron between an occupied state at E_1 in the $n_z = 1$ sub-band in the conduction band and an empty state (i.e., a hole) at E_2 in the $n_z = 1$ sub-band in the valence band, resulting in emission of a photon of energy $h\nu = E_1 - E_2$ as required by energy conservation. This is the process of luminescence (or spontaneous emission), which requires external excitation such as illumination (photoluminescence) or biasing a p-n junction (electroluminescence, as in a light-emitting diode). Alternatively, incident light may promote an electron from an occupied state in the valence band to an empty state in the conduction band with the consumption of the energy of a photon. This is the process of optical absorption.

At typical temperatures the electrons and holes are at energies near to their respective band edges so the photon energy $h\nu$ is close to the separation of the $n_z = 1$ sub-bands. Since these energies are determined by the well width it is possible to change the photon energy for emission or absorption by changing the well width. As the well width is reduced the energy separation increases and the wavelength of light emitted by the structure is reduced. This ability to engineer the emission wavelength by simply adjusting

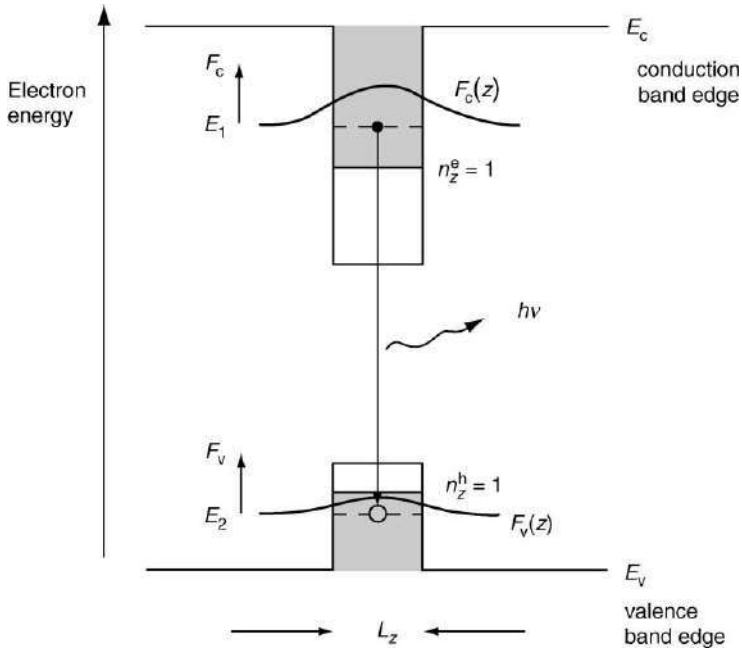

Figure 9 Illustration of a downward optical transition of an electron from a state in the $n_z = 1$ conduction sub-band to an empty state (hole) in the $n_z = 1$ valence sub-band resulting in emission of a photon of energy $h\nu = E_1 - E_2$. The spatial probability distribution of the electron and hole is specified by their respective envelope functions $F_c(z)$ and $F_v(z)$ and the probability of the transition occurring is proportional to their overlap integral, eqn [14].

the well width, without changing the chemical composition of the constituent materials, is one of the major attractions of quantum-confined structures. The shortest wavelength possible for a given material combination is determined by the band gap of the wide gap barrier material alongside the well.

The strength of the luminescence signal or the absorption is determined by the rates at which the electronic transitions occur. In general these rates are determined by the following intrinsic factors.

1. *The quantum mechanical probability of an electron making a transition between a full and an empty state*. This depends upon the 'matrix element' for the transition, which is a fundamental property of the well material and the overlap of the envelope functions. The electron and hole in the initial and final states must be in the same region of space and if the spatial distributions of electrons and holes, determined by F_c and F_v, do not overlap the probability for an electron to make a transition to the hole state is very low. The overlap is almost complete in a rectangular well (see Figure 9), but is only partial in a triangular well. The transition probability is proportional to

$$I_{\text{overlap}} = \left\{ \int F_v(z) F_c(z)\, dz \right\}^2 \qquad [14]$$

2. *The probability of occupation of initial and final states by electrons, specified by the Fermi factors* (eqn [11]). The initial state must be occupied and the Pauli exclusion principle prohibits an electron entering a final state which is already occupied. Thus for a downward transition resulting in emission of light the rate is proportional to $f_c(1 - f_v)$.
3. *The density of initial and final states* (eqn [9]). The more states there are within a given small energy interval, the greater the emission rate per unit energy.
4. *Photon distribution*. For processes such as optical absorption, which are 'induced' by the presence of a photon, the transition rate is also proportional to the photon density in the region of the well. Because the electrons are not wholly localized in the well, and because the spatial extent of the optical field is larger than that of the carriers, the coupling between the photon field (intensity $I_{\text{ph}}(z)$) and the carrier probability distributions is expressed as

$$I_{\text{coupling}} = \frac{\left\{ \int F_v(z) [I_{\text{ph}}(z)]^{1/2} F_c(z)\, dz \right\}^2}{\int I(z)\, dz} \qquad [15]$$

5. *Optical mode density*. The emission rate also depends upon the density of optical modes

available to receive the recombination radiation. In most situations the emission occurs into the cavity of free space which is very large compared with the wavelength of the radiation so the modes are very closely spaced and the cavity can accept radiation at any transition energy (this is the well-known Raleigh–Jeans theory of cavity radiation). However, in some special situations the cavity has dimensions comparable with the wavelength of the radiation (microcavities) and then the modes are widely spaced. In such a microcavity, the optical cavity effectively controls the electronic transitions.

Optical Absorption

Figure 10 is an optical absorption spectrum for a GaAs quantum well measured with light incident normal to the plane of the quantum wells. A series of steps can be seen which correspond to absorption at increasing photon energy between increasing orders of sub-bands, $E_{1e} \rightarrow E_{1h}$, $E_{2e} \rightarrow E_{2h}$, etc. Normally $n_z = 1$ to $n_z = 2$ transitions are forbidden in rectangular wells. At each step there is a pair of peaks (not always resolved) and above each step the absorption is roughly constant, following the step-like density of states function sketched in Figure 4. The peaks are due to the formation of excitons which are weakly bound electron–hole pairs. In bulk materials excitons are only observed at low temperatures whereas in quantum wells the localization increases the binding energy (E_{bX}, typically 6–8 meV) such that they are observed at room temperature. The excitonic peak is at an energy E_{bX} below the sub-band separation. There are in fact two different kinds of electrons in the valence band with different masses so there are pairs of $n_z = 1, 2, 3, \ldots$ sub-bands due to the different confinement energies for the different valence electrons. Spectra of this kind provided the first evidence for the distinctive sub-band structure of quantum well systems produced by carrier confinement. Absorption measurements provide data on the energy spacing of the sub-bands in the well, and it is possible to determine the well width and depths, from which the band offsets can also be determined.

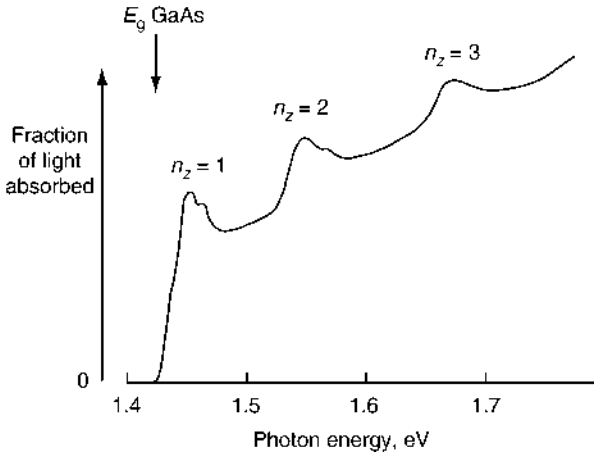

Figure 10 Absorption spectrum of a multiple quantum well structure having 10 nm GaAs wells measured at room temperature. The absorption edges corresponding to transitions between $n_z = 1, 2$, and 3 pairs of sub-bands can be identified. The peaks are due to absorption by formation of excitons.

The strength of the absorption is also of interest because it provides a measure of the optical matrix element if the Fermi factors f_c and f_v are known. If the incident optical beam is very weak such that the absorption process excites very few electrons then the system remains close to thermal equilibrium so the upper states are empty ($f_c = 0$) and the lower states are full ($f_v = 1$).

In *bulk material* the change in photon flux over a small distance Δx is proportional to the flux Φ and the distance traveled

$$\Delta\Phi = -\Phi\alpha\Delta x \qquad [16]$$

where α is the absorption coefficient which has units L^{-1}. This results in an exponential decrease of photon flux with distance traveled through the material

$$\Phi(x) = \Phi_0 \exp(-\alpha x) \qquad [17]$$

For a quantum well structure with light incident in the z-direction perpendicular to the plane of the quantum well, the fraction of light absorbed by a single well by transitions between a single sub-band pair (e.g., transitions between the $n_z = 1$ conduction and valence sub-bands) is independent of the thickness of the well because the density of states for a single sub-band is independent of well thickness. The strength of the absorption cannot be expressed by an absorption coefficient but by the fraction of light absorbed per well:

$$\Delta\Phi = -\gamma_w\Phi \qquad [18]$$

This arises because the quantum well is effectively a sheet in the z-direction: changing its thickness does not change the density of states per unit area (eqn [9]) and we cannot specify the electron concentration along the z-direction. This behavior also occurs because the well is much thinner than the wavelength of the light and it is not possible to

specify the variation of photon flux on a distance scale smaller than the wavelength. The barrier material surrounding the well has a wider band gap and does not absorb at the photon energies absorbed by the well. The well thickness does affect the sub-band spacing and at a fixed photon energy the absorption increases as the well width is decreased as light is absorbed by transitions between more sub-bands. Equation [18] applies individually to each sub-band because the density of states is the same for each sub-band.

In a multiple well system, the fraction of light transmitted through N independent wells is

$$T_N = \{1 - \gamma_w\}^N \qquad [19]$$

so the fraction of light absorbed by such a system is

$$\frac{\Delta\Phi}{\Phi} = -\left[1 - \{1 - \gamma_w\}^N\right] = -N\gamma_w \qquad [20]$$

which is proportional to the number of wells when $\gamma_w << 1$.

The fraction of light absorbed by transitions between a single pair of sub-bands for a rectangular GaAs quantum well is predicted to be about 0.01, and a similar value has been obtained from absorption measurements. In a rectangular well the envelope function overlap is complete; however, in a triangular well of the same material the absorption is reduced by the smaller overlap of the envelope functions.

Photoluminescence and Spontaneous Emission

We turn now to discuss the optical emission from quantum wells as a result of external excitation. An external light source with photon energy above the effective band gap produces excess electron–hole pairs by excitation of electrons from the valence band. These rapidly lose their excess energy and thermalize to the sub-band edges to take up Fermi energy distributions. The electrons subsequently recombine to vacant valence band states by spontaneous emission. The low energy edge of this spectrum corresponds to the sub-band separation and the shape of the spectrum at higher energies corresponds to the thermal distribution of carriers in the bands. At very low excitation intensities where the excess electron population is small the luminescence may be due to recombination of electrons within an exciton and occur at an energy E_{bX} below the sub-band separation. Generally the sub-band separation is several k_BT so only the lowest sub-band is populated and photoluminescence is not seen from higher sub-bands. In this respect absorption measurements provide a more complete characterization of the structure.

While the energy position of features in the photoluminescence spectrum provides useful information, it is difficult to use the strength of the photoluminescence signal to provide quantitative information about the rate of optical transitions within the material because the excitation power and volume within the sample must be known and a known fraction of emitted light must be collected and measured with a calibrated detector. The relative intensity of photoluminescence under standard conditions does provide comparative information and the linewidth of the spectrum at low intensity provides a qualitative indication of 'quality' (e.g., well width variations in the structure). A quantitative determination of the internal recombination rate can be obtained by measuring the decay of the luminescence signal following short-pulse excitation. Fuller discussion can be found in books dealing with the characterization of semiconductor structures.

Concluding Remarks

We have shown that an understanding of the electronic structure of quantum wells follows naturally from quantum mechanical treatments of electrons in bulk materials. When one of the dimensions of the structure is reduced to be similar to the de Broglie wavelength, quantization of the electron motion in this direction becomes apparent. The sub-bands which are formed have a profound effect on the electronic and optical properties when their spacing exceeds thermal energies such that only one band is significantly populated with carriers. The density of states function is a series of steps, each corresponding to a sub-band, and this produces characteristic step features in the optical absorption spectrum. The effective band gap of the structure is controlled by the width of the well. Quantum confinement also increases the binding energy of excitons compared with bulk materials and under low excitation conditions absorption and luminescence spectra are dominated by excitonic features even at room temperature.

One of the most successful application areas of quantum well structures has been in laser diodes. Figure 11 shows experimental data for laser emission wavelengths as a function of well width, showing the dramatic reduction which can be achieved simply by changing the width of the quantum well. Furthermore the step density of

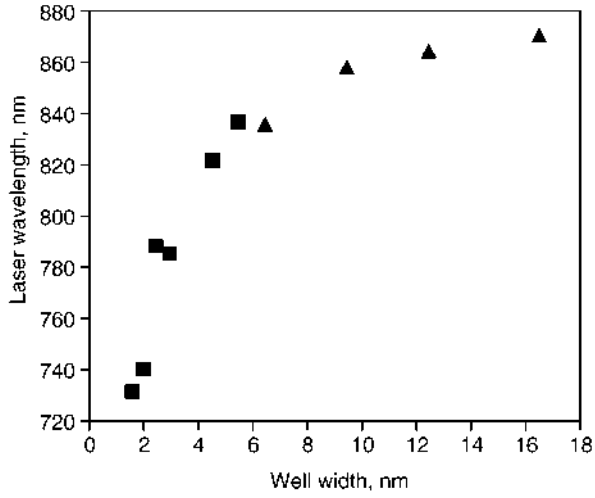

Figure 11 Measured values of the emission wavelength of GaAs/AlGaAs quantum well lasers as a function of well width, showing how the wavelength can be engineered by choice of the well width. Data from: squares: Woodbridge K, Blood P, Fletcher ED and Hulyer PJ (1984) Short wavelength (visible) GaAs quantum well lasers grown by molecular beam epitaxy. *Applied Physics Letters* 45: 16–18; and triangles: Chen HZ, Ghaffari A, Morkoç H and Yariv A (1987) Effect of substrate tilting on molecular beam epitaxial grown AlGaAs/GaAs lasers having very low threshold current densities. *Applied Physics Letters* 51: 2094–2096.

states function, characteristic of quantum wells, is one of the fundamental reasons for the reduction in threshold current of quantum well lasers relative to bulk GaAs devices.

We have considered GaAs because it is a model material system for production of quantum well structures. The concepts developed in this chapter are, however, quite general and can be applied to other material systems, other forms of quantum well and quantum wires and dots.

List of Units and Nomenclature

E	energy
E_f	Fermi energy
E_g	bandgap
F	envelope wavefunction
$f(E)$	Fermi–Dirac distribution function
g_{2D}	two-dimensional density of states $[\text{energy}]^{-1}\,[\text{area}]^{-1}$
$\hbar$	Planck's constant divided by 2π
I_{ph}	optical field intensity
$\mathbf{k}$	wavevector
k_B	Boltzmann's constant
m	electron mass
n	carrier density, expressed per unit area in quantum well structures
Q_c, Q_v	heterostructure band offset ratios
$\mathbf{r}$	position vector
t	time
T_N	fraction of light transmitted at normal incidence through N quantum wells
$\langle\mathbf{x}\rangle$	unit vector along x-direction, similarly for y, z directions
α	optical absorption coefficient in a bulk material, per unit length
γ	fraction of light absorbed by a quantum well at normal incidence to the plane of the well
λ	wavelength
Φ	photon flux (photons per unit time crossing unit area)
Ψ	electronic wavefunction
ν	light frequency

See also

Semiconductor Materials: Band Structure Engineering; Quantum dots; Type-II Quantum Wells and Superlattices. **Semiconductor Physics:** Excitons; Outline of Basic Electronic Properties.

Further Reading

Bastard G (1988) *Wave Mechanics Applied to Semiconductor Heterostructures*. New York: Les Editions de Physique, Halstead Press.

Bimberg D, Grundmann M and Ledentsov NN (1999) *Quantum Dot Heterostructures*. Chichester, UK: Academic Press.

Blood P (1991) Heterostructures in lasers. In: Morgan DV and Williams RH (eds) *Physics and Technology of Heterojunction Devices*. Stevenage, UK: Peter Peregrinus.

Blood P (2000) On the dimensionality of optical absorption, gain and recombination in quantum-confined structures. *Journal of Quantum Electronics* 36: 354–362.

Blood P and Orton JW (1992) *The Electrical Characterisation of Semiconductors: Majority Carriers and Electron States*. London: Academic Press.

Casey HC Jr and Panish MB (1978) *Heterostructure Lasers*. San Diego, CA: Academic Press.

Dingle R, Gossard AC and Weigmann W (1975) Direct observation of superlattice formation in semiconductor heterostructures. *Physical Review Letters* 34: 1327–1330.

Eisberg R and Resnick R (1985) *Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles*, 2nd edn. New York: John Wiley.

Hook JR and Hall HE (1991) *Solid State Physics*. Chichester, UK: John Wiley.

Kelly MJ (1995) *Low-Dimensional Semiconductors*. Oxford, UK: Clarendon Press.
Orton JW and Blood P (1990) *The Electrical Characterisation of Semiconductors: Measurement of Minority Carrier Properties*. London: Academic Press.
Perkowitz S (1993) *Optical Characterisation of Semiconductors*. London: Academic Press.
Wolfe CM, Holonyak N Jr and Stillman GE (1989) *Physical Properties of Semiconductors*. Englewood Cliffs, NJ: Prentice Hall.

Recombination Processes

P T Landsberg, The University of Southampton, Southampton, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In studies of recombination there occur considerable complications; many are associated with interactions between electrons, electrons and phonons, excitons, bi-excitons, impurity centers, etc. Most of these are not required in the present exposition. Modern work normally assumes that the basics of recombination physics are understood and the present exposition offers an appropriate outline.

Basic Assumptions

Electron–Electron Interactions for Electrons in Bands

The recombination problem in semiconductors is greatly complicated by the interaction of the electrons with each other. This allows one to speak only of the quantum states of the semiconductor crystal as a whole. However, as in metals, so also in semiconductors, a simplified picture is successful. In this, the electron interactions, and other interactions, are first neglected, but are later taken into account as a perturbation. Thus, the electrons are first treated as if they can move through each other; the fact that they collide and deflect each other by virtue of their Coulomb interaction is treated as a perturbation. The transitions are still described within the framework of the single-particle states of the unperturbed problem.

The Effect of Electron–Boson Interactions

Our first approximation is to neglect most (but not all) electron interactions. Later we take into account the two-electron transitions that arise. This two-particle recombination process is referred to as Auger recombination and its inverse as the generation of current carriers by impact ionization. In addition, electrons interact with the radiation and lattice fields and emit or absorb photons or phonons. These electron–boson interactions result in transitions of single electrons in an energy band scheme. We shall attach a superfix 'S' (for single-electron) to the recombination coefficients for such processes. These are then denoted by B^{S} or T^{S} depending whether only bands are involved or whether traps are also involved. They are illustrated in Figure 1, where the solid horizontal lines represent the conduction and valence band edges and the dashed line refers to traps. Note that the two-electron transitions do not have the superfix 'S', while n and p refer to the electron and hole concentrations; N_0 and N_1 are the concentration of unoccupied and occupied trap states, respectively.

Electron–Electron Interaction in Traps

Electron–electron interactions can at least formally be taken into account in connection with electrons trapped in a center: the spectrum is a function of the number, $r\,(=1,2,3,\ldots,M)$, of electrons captured. The 'irremovable' electrons can be included with the ion core. Given that the center captured r electrons (say), it can still be in a variety of quantum states, and they will be denoted by the symbol ℓ, yielding a set of quantum states (ℓ, r) for a center. The energy of such a state, divided by kT to make it dimensionless, is denoted by $\eta(\ell, r)$, yielding the canonical partition function

$$Z_r = \sum_{\ell} \exp[-\eta(\ell, r)] \qquad [1]$$

The location of a center in space will here be considered to be of no significance.

Centers can capture several electrons. This brings in a need for the chemical potentials (or Fermi levels) for r-electron centers. They are here denoted by γ (when divided by kT) and the suffix 'eq' denotes an equilibrium value. Thus, the equilibrium probability of finding an r-electron center in state ℓ is given by:

$$P(\ell, r)_{\mathrm{eq}} = \frac{\exp[r\gamma_{\mathrm{eq}} - \eta(\ell, r)]}{\sum_{s=0}^{M} (\exp s\gamma_{\mathrm{eq}}) Z_S} \qquad [2]$$

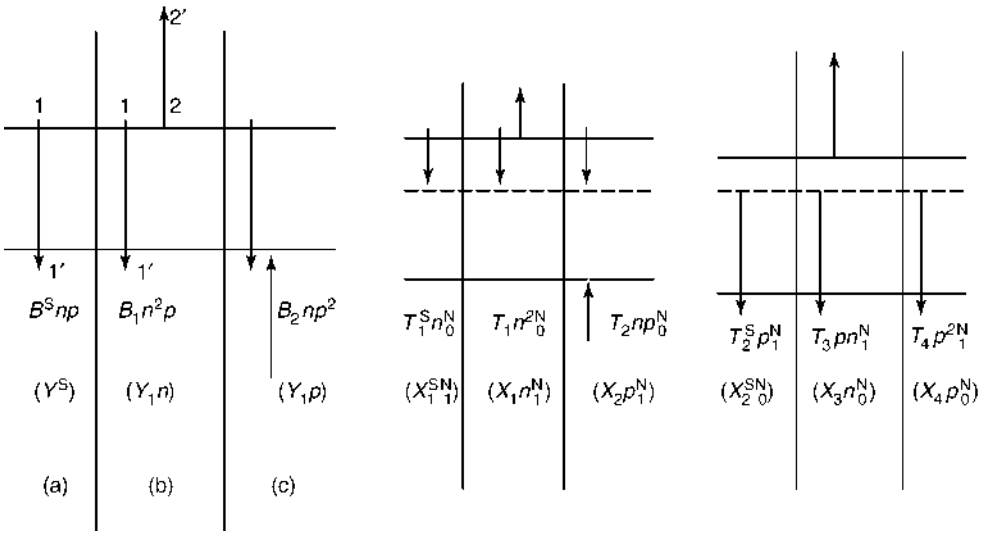

Figure 1 Definition of recombination coefficients. Transition rates per unit volume are stated with each process, and, in brackets, for the reverse process. Thus B_1, B_2, T_1-T_4 refer to Auger processes; B^s, T_1^s, T_2^s refer to single-electron recombination; Y^s, X_1^s, X_2^s refer to carrier generation processes; Y_1, Y_2, X_1-X_4 refer to impact ionization processes. Arrows indicate transitions made by electrons.

One can identify the dominant energies from optical or electrical experiments.

To understand the properties of these centers, suppose we can arrange for the equilibrium Fermi level to rise from the valence band to the conduction band. At first practically no electrons are captured ($r = 0$). As the Fermi level rises the states corresponding to $r = 1$ begin to appear. The states ℓ of the center which correspond to the ground states in equilibrium are always more highly populated than the states ℓ corresponding to excited states, so that we can often confine attention to them. As the Fermi level rises, the ground states for $r = 2$ become more important, and there may now be hardly any centers which have captured fewer electrons. If a larger number of electrons cannot be captured, by the time the Fermi level reaches the conduction band, then states of the center with $r > 2$ can (normally) be neglected as unstable. That this is a satisfactory picture is our second approximation.

Assumptions for Nonequilibrium Statistics

The now much simplified recombination problem is still complex because of the many states available to electrons in bands and centers. A key simplification arises from the fact that it is often possible to talk about a small number of groups of quantum states: let $I\,(= 1, 2, \ldots)$ label quantum states in a group i, let $J\,(= 1, 2, \ldots)$ label quantum states in a group j, etc. Within each group it is supposed that the transitions are much more rapid than they are between groups. In that sense electrons in each group are in equilibrium among themselves. Their quasi-equilibrium is then characterized by what is called a quasi-Fermi level. The dimensionless version, obtained by dividing it by kT, is denoted by γ_i, γ_j, etc., for groups i, j, etc. That this is reasonable is our third assumption. Thus γ_c, γ_v are quasi-Fermi levels which refer to the quasi-equilibrium of the conduction and valence band, respectively, and γ_I to an I-electron center. Groups of states with different quasi-Fermi levels are not in equilibrium with each other.

Recombination problems can now be discussed by neglecting transitions within one (quasi-equilibrium) group, because they proceed at exactly the same rate as their reverse transitions. The number of transition types to be considered is thereby greatly reduced. Away from equilibrium we shall adopt eqn [3] instead of [2] in order to allow for distinct quasi-Fermi levels:

$$P(\ell, r) = \frac{\exp[r\gamma_r - \eta(\ell, r)]}{\sum_{s=0}^{M} (\exp s\gamma_S) Z_S} \qquad [3]$$

Equation [3] is not always correct, but is an approximation arising from the quasi-Fermi level assumption. Among the improved theories are, for example, the 'cascade' theories.

The assumption of a quasi-Fermi level for each state of charge of a center implies that all the excited states of an r-electron center are populated according to eqn [2].

The Main Recombination Rates

General Results and The Two-Band Case

Recombination and its converse, generation, consist of a transition of an electron from one state to another. The observed rate is the net rate of recombination and is the algebraic sum of the recombination and generation processes. During these processes both total energy and total momentum must be conserved. This is achieved by creation or absorption of photons, or phonons, excitation of secondary electrons, etc.

The transition probability per unit time, S_{IJ}, for a single-electron transition from state I in a band to state J requires that state I should be occupied, with probability p_I say, and state J should be vacant with probability q_J say. The general expression for the average rate of the transition from I to J then takes the form $p_I S_{IJ} q_J$. For the reverse process electron state J has to be occupied and state I vacant. Thus the rate of this reverse process is $p_J S_{JI} q_I$. The net recombination rate per unit volume of the process I to J can then be written as:

$$u_{IJ} = (p_I S_{IJ} q_J - p_J S_{JI} q_I) V^{-1} \qquad [4]$$

By the principle of detailed balance, this expression vanishes at equilibrium. If one puts $X_{JI} \equiv S_{IJ} p_I q_J / S_{JI} p_J q_I$, one has $u_{IJ} = p_J S_{JI} q_I (X_{JI} - 1) V^{-1}$. In equilibrium $X_{JI} \rightarrow (X_{JI})_{\mathrm{eq}} = 1$, and the recombination rate is zero.

One can say a little more if one assumes that states I and J are in conduction or valence bands, each with a quasi-Fermi level (divided by kT to make it dimensionless). If these are denoted by γ_{e} and γ_{h}, then $p_I = [1 + \exp(\eta_I - \gamma_{\mathrm{e}})]^{-1}$ for a conduction band. For the total recombination rate per unit volume between the bands i and j one finds:

$$u_{ij} \equiv \left[\sum_{I \in i} \sum_{J \in j} p_J S_{JI} q_I \right] \left[\exp\left(\frac{|e|\phi}{kT} \right) - 1 \right] V^{-1} \qquad [5]$$

where ϕ is essentially the difference between the quasi-Fermi levels: $\gamma_{\mathrm{e}} - \gamma_{\mathrm{h}}$. The first factor depends on the bands involved and it has been assumed that the transition probability ratio S_{IJ}/S_{JI} is independent of the excitation.

A transition rate, when multiplied by the charge of current carriers, is a current, and when divided by the area of the surface involved, is a current density. If i and j denote the states of the conduction and valence bands of a semiconductor, excitation independence may often be assumed, and one then expects a current density proportional to $\exp(|e|\phi/kT) - 1$. This is characteristic of the current through pn junctions, metal semiconductor junctions, etc. In these configurations the Fermi level difference between the ends of the device determines the voltage across it. When radiation is involved, however, excitation dependence of some of the parameters introduced above (e.g. S_{JI}) tends to spoil this simple story.

The Case of Defects

So far we have considered only two bands. When a trap is involved matters are rather different. Because of the interactions among the electrons on a center it is not possible to talk of the same level being occupied or vacant. Consequently, identification of forward and reverse processes in terms of levels becomes impossible. Instead one deals with a center, say an r-electron center, as a whole; we then need the probability that a given center is an r-electron center. For example, the capture of an electron converts an $(r-1)$-electron center into an r-electron center. Thus the p's and q's must be replaced by more complicated expressions. It is convenient to denote u_{ij} by u_{cv} in the simple two-band case and its structure is given (with a sign change) by a recombination coefficient:

$$u_{\mathrm{cv}} = B^{\mathrm{S}} np [1 - \exp(\gamma_{\mathrm{h}} - \gamma_{\mathrm{e}})] \equiv B^{\mathrm{S}} np - Y^{\mathrm{S}} \qquad [6]$$

using Figure 1a. Here, n and p are the electron and hole concentrations. Auger effects in Figures 1b and 1c can also be included, at least formally. Then B^{S} has to be replaced in eqn [6] by $B^{\mathrm{S}} + B_1 n + B_2 p$. Analogous replacements are found for recombination processes involving defects.

In the simplest case of one type of localized defect one finds a steady-state recombination rate per unit volume of the form:

$$u = \frac{np - (np)_0}{\tau_{n0}(p + p_1) + \tau_{p0}(n + n_1)} \qquad [7]$$

Here τ_{n0}, τ_{p0} are parameters with the dimension of time and p_1, n_1 are parameters with the dimension of concentration. The recombination increases with the defect concentration, which is actually in the denominators of τ_{n0} and τ_{p0}. This is an old and much used result associated with the names of W Shockley and W T Read.

Radiative Transitions

Spontaneous and Stimulated Emission

Quantum theory was initiated by Planck's law for black-body radiation at temperature T. This gives the spatial energy density as a function of frequency ν in this system as

$$f(\nu, T) = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{\exp(h\nu/kT) - 1} \qquad [8]$$

For low frequencies one finds the Rayleigh–Jeans law which makes [8] proportional to kT in agreement with the classical equipartition theorem. For high temperatures [8] diverges, as required.

The result [8] may also be obtained by writing

$$N = [A + Bf(\nu, T)]N_u \qquad [9]$$

for the emission rate of photons by atoms, where N_u is the number of atoms in the upper of two states which are separated by the energy $h\nu$. The first term on the right-hand side is due to the normal decay of an excited state ('spontaneous emission'). The second term refers to additional emission ('stimulated emission') induced by the radiation of frequency $h\nu$ itself.

The first factor in [8] is due to the density of states and the second factor is due to the fact that the energy is considered. If one considers the number of photons of energy $h\nu$ at temperature T one comes up with

$$N_\nu = \left[\exp\frac{h\nu - \mu}{kT} - 1\right]^{-1} \qquad [10]$$

(Bose–Einstein distribution)

Here μ is a possible chemical potential of the radiation which is non-zero only in nonequilibrium situations such as in a semiconductor laser.

In a pn junction the two bulk materials several diffusion lengths away from the junction are approximately in equilibrium even if a modest current is flowing. On one side one has then just one quasi-Fermi level, say μ_I, and on the other side one has just one quasi-Fermi level, say μ_J. Then the difference

$$\mu \equiv \mu_I - \mu_J = q\varphi \qquad [11]$$

corresponds to the applied voltage φ. This is in agreement with eqn [10], in that $\varphi = 0$ implies no current and hence thermal equilibrium is possible.

Statistical mechanics teaches one the rule that, in equilibrium, occupation probabilities of individual quantum states are always less for states of higher energies. This ensures that the total energy of a quantum system converges, the occupation probability p being a kind of convergence factor. However, if one is away from equilibrium, the above rule can be suspended and one can have population inversion.

For $h\nu = \mu$ there is trouble with [10] because the steady-state photon occupation diverges. This does not correspond to a 'death ray', but is the result of imperfect modeling; for example, the leakage of photons from the cavity may have been neglected. A formula of type [10] is also needed in connection with solar cells.

Donor–Acceptor Pair Recombination

A striking demonstration of radiative donor–acceptor transitions in GaP at low temperature (1.6 K) was revealed by sharp lines (Figure 2) at photon energies $h\nu_i$ given by

$$h\nu_i = E_G - E_A - E_D + e^2/\varepsilon R_i\,[-E] \qquad [12]$$

where R_i is the distance between the ith-order nearest neighbors.

The discrete nature of the peaks is due to the fact that the impurities involved settle in general on lattice sites so that only definite separations R_i are possible. The energy gap is E_G; the value of $E_A + E_D$ can be inferred from the experimental lines by extrapolation to $R_i \to \infty$. The energy term E is sometimes neglected.

The ZnS phosphors were the first materials in which donor–acceptor radiation was hypothesized. However, it is hard to control its stoichiometry and its impurity content, and it has relatively large carrier mass and hence relatively large impurity activation energies. In such cases spectra such as those shown in Figure 2 are hard or impossible to obtain. This applies in general to many II–VI compounds.

Quantum Efficiency

The quantum efficiency is the radiative recombination as a fraction of the total recombination. It can also be regarded as the average number of electrons produced per incident photon. Consider an intrinsic semiconductor, i.e., one in which the electron and hole concentrations are equal. Then with the notation of Figure 1 the radiative band–band recombination rate is $B^s n^2$. The nonradiative rate is $(B_1 + B_2)n^3$. We shall add another nonradiative rate, An say, proportional to the injected carrier density $n \sim p$. The quantum efficiency is then

$$\eta = \frac{B^s n^2}{An + (B_1 + B_2)n^3 + B^s n^2} \qquad [13]$$

Traps are here neglected, and one finds a maximum

$$\eta_{\max} = \left\{1 + \frac{2}{B^{s}} A^{1/2}(B_1 + B_2)^{1/2}\right\}^{-1} \qquad [14]$$

For a high-quality epitaxial AlGaAs/GaAs double heterostructure of great purity we may take

$$A \sim 0.5 \times 10^{6}\ \mathrm{s}^{-1},\ \ B \sim 10^{-10}\ \mathrm{cm}^{3}\ \mathrm{s}^{-1},$$
$$B_1 + B_2 \sim 10^{-29}\ \mathrm{cm}^{6}\ \mathrm{s}^{-1} \qquad [15]$$

whence $n_1 \sim 2 \times 10^{17}\ \mathrm{cm}^{-3}$, $\eta_{\max} \sim 0.96$.

Detailed Balance

It is clear that the radiative recombination rate from a material should be obtainable in terms of its optical 'constants', the absorption coefficient $\alpha(\lambda)$ and the refractive index $\mu(\lambda)$. Both are functions of the wavelength. This connection can be formalized by comparing the optical absorption process with the emission process and this can be done by using the principle of detailed balance. This equates the rate of disappearance by absorption of photons with the rate of production of photons by radiative recombination. The radiative recombination rate can thus be obtained as an integral over the optical functions. This relationship, pioneered by van Roosbroeck and Shockley in 1954, has been used a great deal because the optical quantities are often known with some accuracy.

This result corresponds to balancing the rate $B^{s}np$ and the rate Y^{s} in Figure 1a. Analogous detailed balance results are obtainable for the other pairs of processes in Figure 1. Thus, the Auger recombination rate can be related to an integral over the impact ionization rate. However, this connection is less useful since impact ionization data are generally less well known than the optical absorption information.

Nonradiative Processes

Auger Effects

The Auger effect was discovered in 1925 in gases by Pierre Auger. An atom is ionized in an inner shell. An electron drops into the vacancy from a higher orbit and a second electron takes up the energy which is used to eject it from the atom. In solids the effect is roughly analogous. One of its characteristics is that it is not radiative. Since radiation is what is seen in many experiments, the Auger effect is suspected if and when there is less radiation than expected in the first place. It is rather harder to investigate than radiative transitions.

The effect has proved to be important as it limits the performance of semiconductor lasers, light-emitting diodes and solar cells, and it can be crucial in transistors and similar devices whose performance is governed by lifetimes. When heavy doping is required, as it is in the drive towards microminiaturization, its importance tends to increase, since the Auger recombination rate behaves roughly as n^2p or p^2n (see Figure 1) compared with the radiative rate which behaves more like np. The inverse process is impact ionization, and is important in the photodiode, the impact avalanche transit time (IMPATT) diode, and hot electron devices.

Impact ionization can be regarded as an autocatalytic reaction of order one:

$$\mathrm{e} \rightarrow 2\mathrm{e} + \mathrm{h} \quad \text{or} \quad \mathrm{h} \rightarrow 2\mathrm{h} + \mathrm{e} \qquad [16]$$

i.e., one extra particle is produced of the type present in the first place. This is a key feature for impact-induced nonequilibrium phase transitions in semiconductors.

In the theory one needs wavefunctions for four states of two electrons which are relevant after the many-electron problem has been reduced to a two-electron problem. In a matrix element calculation the four wavevectors imply a 12-fold integration in $\boldsymbol{k}$-space to cover all possible states of the two electrons. However, momentum and energy conservation usually reduce this to an eight-fold integration. Such a calculation is hard, for it requires (1) good wavefunctions and (2) accurate integrations.

Here we shall merely concentrate on the broad principles. In addition, the many-electron nature of the problem implies that another approximation is often inherent in the treatment apart from the use of perturbation theory. This is clear if one considers that the electron interactions are screened twice: once by an exponential screening factor and a second time by the dielectric constant. So there is some double counting, and the treatment of the effect as uncorrelated electronic transitions mediated by screened Coulomb interactions is another approximation. The collective effects that enter require more sophisticated field theoretic methods which take care of electron–hole correlations, plasmon effects and the effect of free excitons, the so-called excitonic Auger effect. These calculations broadly confirm the results obtained by the simpler methods used for energy gaps large compared with the plasmon energies,

provided the high-frequency dielectric constant is used and doping is not too heavy. Significant corrections are required for the narrow gap lead compounds, for example, and a considerable computational effort is required.

Let us now consider lifetimes for high carrier concentrations, as limited by band–band processes. They are best studied by looking at the emitted radiation. In this connection we recall the striking rise of the threshold current density J in a semiconductor laser as the material is changed from a direct material like GaAs to an indirect material, by mixing it with a compound such as GaP. **Figure 3** shows this spectacular rise for $GaAs_{1-x}P_x$ at 77 K near $x = 0.38$. It is due to the drop in the radiative transition probabilities, as the substance becomes indirect, thus requiring a higher current density for threshold. So we should start with band–band processes in direct materials as most favorable for the experimentally accessible radiative transitions.

The essential point here is that the injected carrier density behaves as:

$$n_{\text{inj}} = \frac{J\tau}{qd} \sim \frac{(5\times10^3\ \text{A cm}^{-2})(10^{-9}\ \text{s})}{(1.6\times10^{-19}\ \text{C})(10^{-4}\ \text{cm})} \sim 10^{17}\ \text{cm}^{-3} \qquad [17]$$

where J/q is the particle current density at threshold, τ is the lifetime of the carriers, and d is the thickness of the active layer. As active layers are made thinner n_{inj} increases to $10^{19}\ \text{cm}^{-3}$ or so, far in excess of the defect concentration in the (undoped) material.

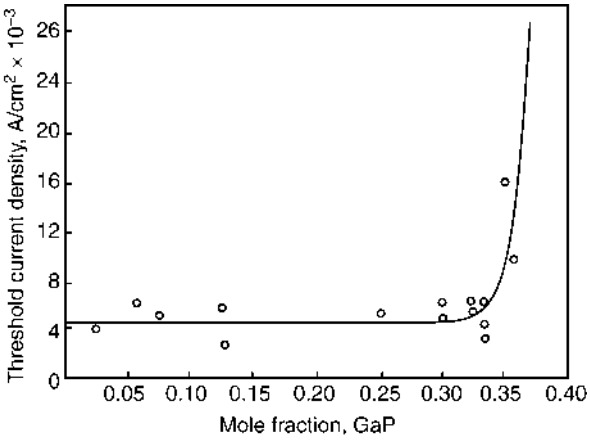

Figure 3 Lowest values of laser threshold current density at 77 K as a function of mole fraction of GaP for $Ga(As_{1-x}P_x)$ junction lasers. Reproduced from Neill CJ, Stillman GE, Sirkis MD, *et al.* (1966) Gallium arsenide-phosphide: Crystal, diffusion and laser properties. *Solid-State Electronics* 9: 735–742, with permission from Elsevier.

This brings in band–band Auger effects. These have also been invoked to explain the undesirable increase in J with temperature. Thus the interest in direct band–band Auger effects in III–V compounds is fuelled by the need for better and smaller opto-electronic devices which work at long wavelengths (1.3–1.5 μm). It matches the interest in indirect band–band Auger and impurity Auger effects in silicon due to the importance of heavy doping in VLSI (very large scale integration) and transistor technology. **Figure 4** shows a typical Auger process, called CHHS, as the conduction band (C) and the split-off band (S) are involved. Two relevant states are in the heavy hole band (H). State 2′ is referred to as the Auger carrier, as it has more kinetic energy than the others.

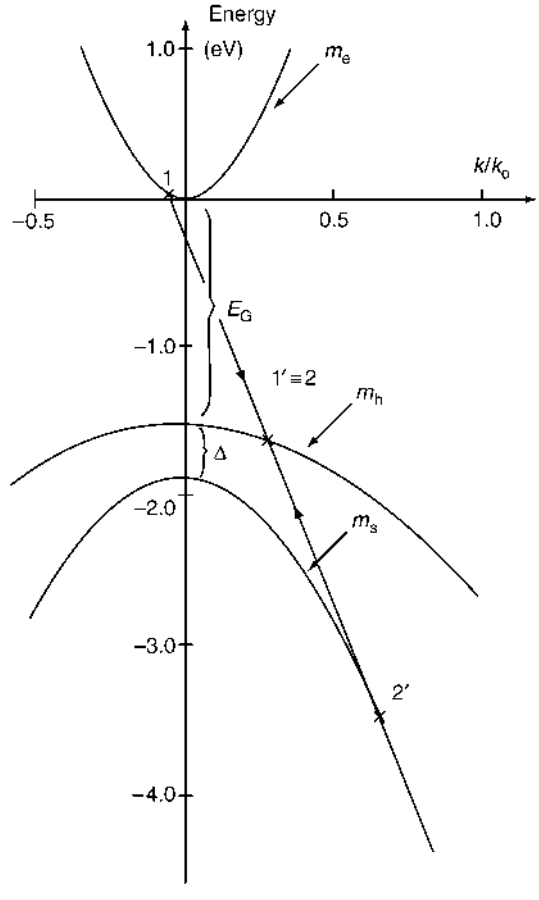

Figure 4 Conduction band, heavy hole and split-off valence bands of GaAs, all treated as parabolic with $\hbar^2k_0^2/2m_h \equiv E_G - \Delta$. x denotes a quadruplet of states for a most probable transition. The two states in the heavy hole band are not shown separately. The arrows indicate electron transitions. Reproduced from Neill CJ, Stillman GE, Sirkis MD, *et al.* (1966) Gallium arsenide-phosphide: Crystal, diffusion and laser properties. Solid-State Electronics 9: 735–742, with permission from Elsevier.

Impact Ionization

The CHHS and CHCC processes and their inverses can be written as

$$e_C + 2h_H \leftrightarrow h_s \quad 2e_C + h_H \leftrightarrow e_C \qquad [18]$$

where the suffix refers to the band. Viewed from left to right these are Auger processes. Viewed from right to left they are autocatalytic and impact ionizations.

Such processes are important for the impact-induced nonequilibrium phase transition in semiconductors. Also reaction rates with autocatalytic elements imply nonlinear equations in the concentrations and this gives rise to much interesting behavior as regards stability and bifurcation phenomena.

Momentum and energy conservation are very restrictive conditions on the four states involved in [18], and it is easily seen that they cannot normally be satisfied if in a direct band semiconductor the recombining electron drops from the band minimum to the valence band maximum. This effect raises all the kinetic energies of possible processes. The Auger electron (on the right-hand sides of [18]) must also have a minimum energy in order that the impact-ionization process can proceed. This leads to an activation energy for the process. For Figure 4, for example, and in the case of nondegeneracy, the energetic hole has a kinetic energy at threshold of

$$E_{th} = \frac{m_e + 2m_h}{m_e + 2m_h - m_s}(E_G - \Delta) \quad \text{(CHSS)} \qquad [19]$$

As the band gap increases we see that E_{th} goes up and so the Auger rate for simple parabolic bands decreases. Values of E_{th} for other transitions can be obtained from equation [19].

The total kinetic or threshold energy E_{th} can be converted to an activation energy by subtracting the basic energy which must under all conditions be part of the Auger particle energy. In the case of eqn [19] one finds for $E_G > \Delta$:

$$\begin{aligned} E_a &= E_{th} - (E_G - \Delta) \\ &= \frac{m_s}{m_e + 2m_h - m_s}(E_G - \Delta) \end{aligned} \qquad [20]$$

This can result in a strong temperature dependence in accordance with an Arrhenius factor $\exp(-E_{th}/kT)$. However, this is again lost, and the direct band–band Auger effect will certainly be important and only weakly temperature dependent, if $E_G \sim \Delta$. This can occur, for example, for InAs, GaSh and their solid solutions.

The connection between Auger processes and impact ionization has also been formalized. Of all Auger quadruplets of states in a set of nondegenerate bands, the most probable Auger transition involves the quadruplet, which yields the threshold for impact ionization. The three crosses in Figure 4 indicate such a quadruplet (the central cross represents two states).

Identification of Auger Effects

How does one know that an Auger effect has occurred? In pure but highly excited materials there is the original solid-state Auger effect which leads to a lifetime broadening of the electronic states at the band edge. For a semiconductor this effect causes fuzziness in the band edge which is a contributory factor to the overall bandgap shrinkage. The mechanism is illustrated in Figure 5, which shows how the lifetime of a vacant electron state near the band edge is shortened by the Auger processes within this band. Under normal conditions this effect is small in a semiconductor. But under degenerate conditions the effect leads by lifetime broadening to a low-energy tail in emission.

The blanket term 'Auger effect' for all Coulombically excited two-particle transitions has been used here. In semiconductor device work the 'Auger effect' has become associated with transitions in which one electron bridges an energy gap. But there is no need to limit the concept in this way. Its competition with radiative effects means that it is often detrimental and a full discussion of means of suppressing it has recently been given by Pidgeon *et al.*

More spectacular and more convincing is the *detection* of the energetic Auger electron or hole. For this purpose one may look for the weak luminescence emitted when this carrier recombines radiatively. This has been done for the band–band process in Si and for the band–impurity process

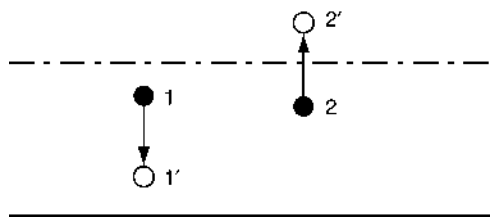

Figure 5 The filling of a state *k* by the Auger effect. The arrows indicate electron transitions.

in GaAs. It leads to so-called $2E_G$-emission. Its rate-limiting step is the rate of the Auger process per unit volume, $B_1 n^2 p$, which populates the high-energy level. The radiative process then proceeds at a rate $B'_1 n^2 p^2$ per unit volume with $B'_1 \sim 10^{-54}$–10^{-51} cm^9 s^{-1}.

More usual is the identification of the band–band Auger recombination mechanism from the minority carrier lifetime τ, which behaves as

$$\frac{1}{\tau_p} = B_1 n^2 \quad \text{or} \quad \frac{1}{\tau_n} = B_2 P^2 \qquad [21]$$

where B_1 and B_2 are Auger coefficients when the Auger particle is an electron or hole, respectively. A more complete expression allows also for trapping processes.

The departure from parabolic bands plays an important part for the energetic Auger particle (in state 2′). Theory shows that when this effect is taken into account, it tends to lower the band–band Auger coefficient. In fact in GaAs the CHCC process is ruled out altogether at 0 K for nonparabolic bands, suggesting that it should be an ideal material for radiative transitions. However, the possibility of phonon participation brings the effect back again, though it is still comparatively weak. The difficulty of conserving electron energy and momentum, which gives rise to the activation energies (eqn [20]) is greatly alleviated if phonons can take up some of the momentum in a direct gap semiconductor.

We have seen that the activation energy barrier against the band–band Auger effect can be overcome by a suitable disposition ($E_G = \Delta$) of the three direct bands and by phonon participation. It can also be overcome if there is an indirect minimum near a Brillouin zone edge at about half the direct energy gap.

Acknowledgments

The author wishes to acknowledge support from NATO under their grant PST.CLG 975758.

See also

Semiconductor Physics: Band Structure and Optical Properties; Excitons; Impurities and Defects; Infrared Lattice Properties; Outline of Basic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Fraser DA (1986) *The Physics of Semiconductor Devices*, 4th edn. Oxford, UK: Clarendon Press.

Hangleiter A (1985) Experimental proof of impurity Auger recombination in silicon. *Physical Review Letter* 55: 2976–2978.

Harrison D, Abram RA and Brand S (1999) Characteristics of impact ionization rates in direct and indirect gap semiconductors. *Journal of Applied Physics* 85: 8186.

Landsberg PT (1991) (paperback edn., 2003) *Recombination in Semiconductors*. Cambridge, UK: Cambridge University Press.

Nimtz G (1980) Recombination in narrow gap semiconductors. *Physics Reports* 63: 265.

Pidgeon CR, Ciesla CM and Murdin BN (1997) Suppression of non-radiative processes in semiconductor mid-infrared emitters and detectors. *Progress in Quantum Electronics* 21: 361–419.

Spin Transport and Relaxation in Semiconductors; Spintronics

M E Flatté, University of Iowa, Iowa City, IA, USA

D D Awschalom, University of California, Santa Barbara, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Control of the generation and transport of coherent electron spin in semiconductors suggests new ways to probe the fundamentals of quantum decoherence in solids and to explore device applications that rely on coherence, including quantum computation. The focus here will be on the physical phenomena that govern spin transport and relaxation, with a particular emphasis on those phenomena that are amenable to direct manipulation. Manipulation of material properties in general has been taken to its most advanced level in the design of semiconductor electronic devices, in which operations are performed by perturbing the motion of electrons through their electric charge; hence, 'spintronics' is a term introduced to describe both the manipulation of electrons by their spin and control of their spin properties, such as g-factors.

The study of the evolution of spin coherence first requires the generation and detection of spin coherence. Here the ultrafast optical technique for generating and detecting spin-polarized populations will be emphasized. The next logical component to treat will be the persistence of spin coherence, followed by the transport of spin coherence from one region of a material to another. Finally the direct ultrafast manipulation of spin-polarized populations will be described. These components are likely to play a prominent role in any attempt to apply the fundamental physical phenomena of spin coherence to specific device technologies.

Optical Generation of Spin-Polarized Distributions

Optical pulses with definite circular polarization can be used to generate spin-polarized distributions of excited electrons in crystals. In order for this process to work one of the energy bands involved in the optical transition must be characterized by a substantial spin–orbit interaction. The selection rules for circularly polarized photons require a change in the orbital angular momentum of the electron projected along the propagation direction of the photon. For the optical transitions of an atom in free space this azimuthal orbital angular momentum l_z must increase or decrease by $\hbar$, permitting the coupling of the p-state with $l_z = 1$ to an s-state for left circular polarization, and of the p-state with $l_z = -1$ to that s-state for right circular polarization. Thus the optical selection rules pick out a particular orbital angular momentum projection.

These selection rules by themselves do not explain how spin-polarized populations are generated, for both spin directions are possible for each value of l_z. In the presence of spin–orbit interaction, however, the energy of a state with a particular l_z is correlated with the spin orientation. The six p-states, which are degenerate in the absence of spin–orbit interaction, split into two groups: one of four degenerate states and the other of two degenerate states. The higher-energy group of four degenerate states consists of two states with **L** and **S** parallel and two states with a one-third probability for **L** and **S** to be antiparallel and a two-thirds probability for **L** and **S** to be perpendicular. Thus for a transition involving the four degenerate states, a circularly polarized photon (that selects a particular initial l_z) is three times more likely to generate a carrier with spin parallel to the initial l_z than antiparallel to it.

In a solid, although full rotational symmetry is lost, the band symmetries preserve some of these selection rules. In crystals of tetrahedral symmetry, such as zincblende semiconductors, the selection rules for transitions from the valence to the conduction band are identical to the selection rules between p and s states for the atom in free space. For example, the valence heavy-hole states are those for which **L** and **S** are parallel. Thus by setting the optical photon energy to the fundamental gap, the resulting excited conduction electrons and valence holes will be spin polarized. In quantum wells, where the energy of the heavy hole is split from that of the light hole by quantum confinement (and, sometimes, strain), fully spin-polarized carrier distributions are possible. Commonly used group IV, III–V, and II–VI semiconductors have a sufficient spin–orbit interaction to permit the straightforward optical generation of spin-polarized distributions. Shown in Figure 1 are the selection rules

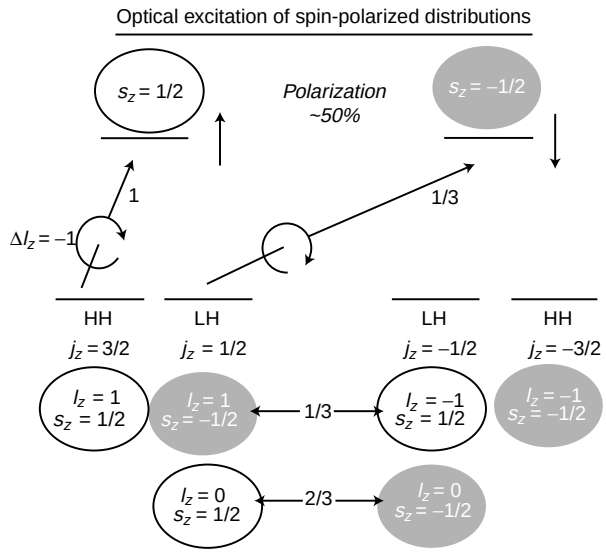

Figure 1 Selection rules for bulk zincblende crystals. Circularly polarized light changes the azimuthal orbital quantum number l_z by $+1$ or -1. Here $\Delta l_z = -1$ is shown. The orbital and spin quantum numbers for the two degenerate conduction band edge states and the four degenerate valence band edge states are shown – unshaded areas correspond to spin up and shaded to spin down. A transition with $\Delta l_z = -1$ couples one heavy-hole state to the conduction band up-spin state, and one light-hole state to the conduction band down-spin state. The orbital angular momentum of the heavy-hole state is entirely parallel to its spin angular momentum. In contrast only one-third of the light-hole state has antiparallel orbital and spin angular momentum. The rest of the light-hole state has perpendicular orbital and spin angular momentum. Thus the oscillator strength of the transition creating the conduction up-spin state is three times larger than that of the transition creating the conduction down-spin state, yielding a 50% spin polarized distribution. In quantum wells for which the heavy–light splitting is larger than the linewidth of the light source driving the transition, only the heavy-hole transition is excited and the spin polarization approaches 100%.

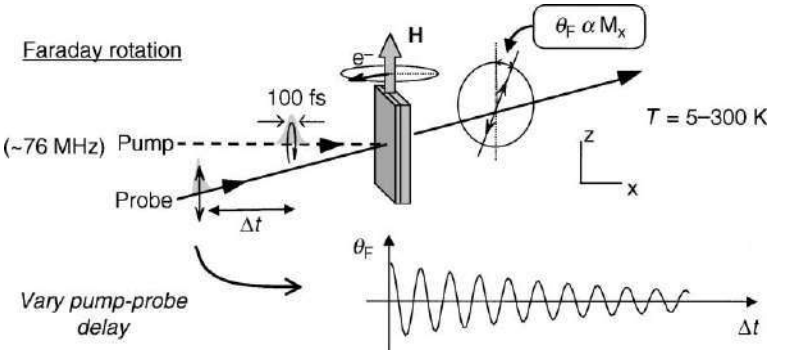

Figure 2 Diagram of the experimental geometry for time-resolved Faraday rotation. A pulse of circularly polarized light illuminates the sample, followed by a probe of linearly polarized light. The linearly polarized light can be considered as a sum of equal parts left and right circularly polarized light. The spin-polarization of the electron population (along the photon propagation direction) induces a difference in index of refraction for left and right circularly polarized light (circular dichroism). This produces a phase change between the left and right polarized components of the probe, which is measured as a rotation (θ_F) of the linear polarization direction. Monitoring the angle of rotation of the probe as a function of time directly indicates the spin polarization along the propagation axis of the photon as a function of time. Reproduced with permission from Awschalom DD, Samarth N and Loss D (eds) (2002) *Semiconductor Spintronics and Quantum Computation*. New York: Springer

which permit circularly polarized optical pulses to generate spin-polarized carriers in the conduction and valence band.

The evolution of a coherent spin-polarized population can be monitored through optical means as well. If the optical transitions used to generate the spin-polarized population are pumped hard enough, the transitions will be bleached. By measuring the polarization dependence of this bleaching the decay of the spin-polarized population can be determined. This can be done either by monitoring the transmission of time-delayed probe pulses of circularly polarized light, or by monitoring the rotation angle (Faraday rotation) of a pulse of linearly polarized light. The first of these depends directly on the imaginary component of the spin-polarized dielectric function, whereas the second depends directly on the real components. Shown in Figure 2 is the geometry for observing Faraday rotation from a spin-polarized distribution in the presence of an applied magnetic field perpendicular to the spin polarization.

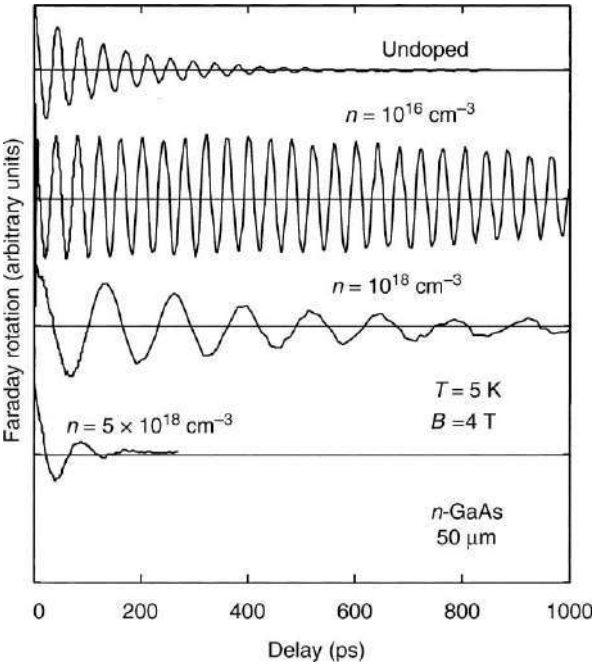

Figure 3 Time-resolved Faraday rotation measurements on bulk GaAs at several doping levels, indicating a sensitive doping dependence of the transverse spin coherence time T_2. T_2 for ideal doping densities can exceed 150 ns at 5 K. The magnetic field is 4 T, and the sample is 50 μm thick. Reproduced with permission from Kikkawa JM and Awschalom DD (1998) Resonant spin amplification in *n*-type GaAs. *Physical Review Letters* 80: 4313–4316.

Mobile Electron Spin Time Evolution: Decoherence and Precession

Two natural phenomena in the time evolution of a coherent spin-polarized population are decay of that population (decoherence) and rotation of the macroscopic magnetization in the presence of an applied magnetic field (precession). The time-scale of decay of the spin coherence can be very much longer than time-scales for the decay of other coherent phenomena in solids. Typical orbital coherence times, for example, are 100 fs, whereas spin coherence times can exceed microseconds. In Figure 3 are shown time-resolved Faraday rotation data for three bulk GaAs

crystals with different doping levels. The oscillations are characteristic of the spin precession, and the oscillation frequency is the Larmor frequency of the electron spin in this crystal. The crystal with the middle doping level, 10^{16} cm^{-3}, has oscillations that persist for the longest time, and therefore this material has the longest spin coherence time.

The dominant mechanism for the decay of these spin polarized distributions works via the same spin–orbit interaction that permits the generation and detection of these spin-polarized distributions. For mobile electrons in homogeneous environments, the mechanisms of spin coherence relate to the crystal momentum-dependence of the eigenstates of the electronic system. Because of time reversal invariance the electronic states with zero momentum (at the Γ point in the Brillouin zone) are doubly degenerate. The electronic states for momentum $\mathbf{k}$ near the Γ point occur in nearly degenerate pairs. Although it is common usage to refer to these eigenstates as degenerate spin eigenstates (correct in the absence of spin–orbit interaction), with spin–orbit interaction the states are not spin eigenstates and are not necessarily degenerate. As a consequence, when a spin-polarized carrier is scattered by ordinary, non-spin-dependent scattering from one momentum state $\mathbf{k}$ to another $\mathbf{k}'$, and then back again to $\mathbf{k}$, the effective orientation of the carrier spin will have rotated by an angle dependent on the intermediate momentum state $\mathbf{k}'$. For a population of carriers in thermal equilibrium this process will lead to misorientation of the individual spin moments from the mean, and thus a decoherence of the spin population. This process is characterized by an increasing decoherence rate with increasing orbital scattering. Thus the dirtier a material is made, the faster the spin orientations decohere.

In crystals with inversion symmetry the pairs of nearly degenerate states are precisely degenerate. Lack of inversion symmetry, and the consequential energy splitting, can be viewed as an effective (crystal) magnetic field, in which the approximate spin eigenstates precess. The magnitude and direction of this effective field depends on the orbital momentum $\mathbf{k}$, and because of time reversal invariance, the effective fields for $\mathbf{k}$ and $-\mathbf{k}$ are opposite. Thus an initially spin-polarized population will dephase due to experiencing different effective magnetic fields. Ordinary spin-independent scattering can now be seen as randomly changing the precession axis for the spin in this effective field. Once the information about the precession axis orientation has been lost the dephasing is not reversible and the spin orientations have decohered. As the precession angle increases linearly between scattering events, the longer the time between such scattering events the faster the spin population decoheres. Thus for this process the cleaner a material is made, the faster the spin orientations decohere. This precessional decoherence process dominates for mobile conduction electrons in bulk and quantum-well III–V semiconductors near room temperature.

The influence of the spin–orbit interaction is far stronger for the valence electrons than the conduction electrons for two reasons. First, as the effective magnetic fields come from relativistic transformations of internal electric fields, and the electric fields are largest near the nucleus, the valence electrons, which are closer to the nucleus than the conduction electrons, experience a stronger spin–orbit interaction. Second, the valence electrons have a dominant p-like character, whereas the conduction electrons have dominant s-like character (and thus lack orbital angular momentum to be coupled to the spin). Thus the spin coherence times of valence electrons (or holes) are commonly much shorter than the coherence times of the conduction electrons. Spin-flip exchange between conduction electrons and valence holes can therefore be the dominant source of decoherence for conduction electrons. This process can dominate at low temperatures in clean III–V semiconductors, and even up to room temperature in III–V quantum wells grown along the unusual (110) growth direction.

As a consequence one would expect the spin coherence times of electrons in n-doped semiconductors to be longer than those of electrons in p-doped or undoped semiconductors (in which excess holes must also be optically generated). This expected trend is present in Figure 3, for the n-doped materials have longer spin coherence times than the semi-insulating material. However, as the effect of the spin–orbit interaction increases with increasing momentum $\mathbf{k}$, the spin coherence times of the more heavily doped material are shorter than those of the lightly doped material. The doping corresponding to the longest spin coherence times may lie near the metal–insulator transition, although the theory of spin coherence times on the insulating side of that transition is not as well developed as on the metallic side.

Transport and Manipulation of Spin Coherence

An essential component of the motion of spin coherence in semiconductors is the response of carriers to electric fields. Outside of very unusual electronic states in semiconductors, such as the quantum Hall state, there is no spin–charge

separation of elementary carriers. The carriers in the semiconductors will move under the influence of an electric field, and if they happen to be spin-polarized, then a spin current will exist. Although there are numerous possible regimes of spin transport (at least as many as there are regimes of charge transport), only two will be presented here, both in the low-field diffusive limit.

Mobilities and diffusion constants are the material properties that describe the low-field charge transport in a homogeneous material. Spin transport can be described similarly, and the spin mobility and diffusion constant can be related to those of charge motion. The relationship between the two emerges from the nature of carrier packets that can be maintained under low-field conditions. Deviations from local charge neutrality cause very large space-charge fields in the semiconductor, and thus under low-field conditions approximate local charge neutrality must be maintained. Shown in Figure 4 are charge and spin packets in an n-doped semiconductor. The charge polarization packet, incorporating extra conduction electrons and valence holes, is shown in Figure 4a. The electron spin polarization packet shown in Figure 4b is qualitatively different from the charge packet as well as the hole spin polarization packet (shown in Figure 4c). Only the electron spin polarization packet does not incorporate both conduction electrons and valence holes.

The difference in packet structure profoundly changes the mobility and diffusion properties of the packet. For packets consisting of both electrons and holes the low-conductivity carrier species dominates the packet motion (in this case, the holes). For the electron spin packet, however, majority carriers dominate the packet motion. Thus motion of this spin packet is very rapid, and also sensitive to the higher Fermi temperature of electrons relative to holes. Experimentally this has been seen by optically creating a local spin-polarized packet of conduction electrons (an electron spin packet) in a thin sample of n-doped GaAs. The application of a lateral electric field drags that spin packet, and an optical probe measures the transport distance (thus the mobility) and the packet spread (yielding the diffusion constant). Such an approach is very similar in concept to the electrical probe of charge packet motion developed by Shockley and Haynes to measure minority hole mobility and diffusion in n-doped Ge.

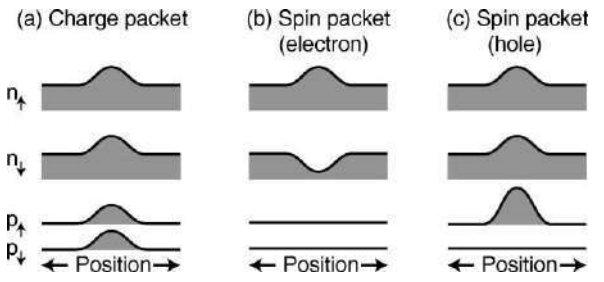

Figure 4 Inhomogeneous carrier packets in n-doped semiconductors in the low-field regime. The charge polarization packet (a) is constructed of approximately equal numbers of electrons and holes. The electron spin polarization packet (b) can be constructed without any excess holes, whereas the hole spin polarization packet (c) requires both excess electrons and excess holes. The presence of minority carriers in a packet significantly alters the mobility and diffusion properties of the packet.

The picture of approximately charge-neutral packets is very helpful for describing spin motion in homogeneous materials, but most semiconductor devices are based on inhomogeneous configurations of semiconducting material. For example, the most ubiquitous charge-based semiconductor device is the p–n diode, whose nonlinear current–voltage characteristics have been used as building blocks for numerous other devices. Such devices rely on the ability of a current to traverse an interface between two dissimilar materials – either different in doping density or different in host structure. Experiments in metallic ferromagnetic systems indicated that, whereas charge current was easily passed from one ferromagnetic metal to another through an interface, the same was not always true for the spin current. Differences in band structure or interface roughness are usually expected in such systems to strongly diminish the efficiency of spin current transfer across the interface.

Semiconductors, however, offer a possible new regime. The bulk band structures of most III–V semiconductors are very compatible – the variations in bandgap may be substantial, but the orbital wavefunctions are very similar. Current growth techniques permit the fabrication of atomically smooth interfaces between two semiconductors, so long as the lattice constants are almost identical. Thus one might expect the spin current to be transferred across such interfaces without tremendous degradation.

The transfer process for spin current across the GaAs/ZnSe interface has been investigated in detail. Here the ZnSe material was n-doped, and GaAs layers that were n-doped, p-doped, and semi-insulating were investigated. Electron spin polarization was generated optically in the GaAs layer. For n-doped GaAs in the absence of an electric field only a small signal for spin-polarized electrons in the ZnSe was detected. An applied electric field, however, which drove the electrons within the GaAs towards the interface, produced a substantial spin-polarized population in the ZnSe. This indicates that spin-polarized packets can be created in one

semiconductor, and then controllably moved into a second semiconductor with an applied electric field. This ability may become a key component of manipulating spin in semiconductors.

One reason the p–n diode is a paradigm of charge-based semiconducting electronics is that it exhibits the crucial phenomenon of minority spin injection. Manipulation of minority spin injection permits great control of charge current, for its amplitude is very sensitive to barriers (such as those introduced by gates). Spin-polarized minority carrier injection has not been shown yet experimentally, however the complementary process of minority carrier extraction has. This process corresponds to optically generating spin-polarized electrons in p-doped GaAs, where the electrons are minority carriers. The built-in electric field of the p–n junction drives these electrons out of the GaAs into the n-doped ZnSe with great efficiency. Thus one may imagine that spin-polarized minority carrier injection is plausible in these systems.

A hybrid of these two situations is the spin-polarized light-emitting diode. Here spin-polarized carriers (either electrons or holes) are electrically injected into the insulating region of p–i–n structure. In the insulating region they recombine with unpolarized carriers of the other species and emit light. Due to the selection rules described before, the emitted light is circularly polarized, and the degree of circular polarization is an indication of the spin-polarization of the recombining carriers. This device demonstrates that large numbers of spin-polarized carriers can be accumulated in the insulating region, although the spin-polarized carriers are not minority carriers.

Methods of manipulating electron spin coherence externally without electric fields are also under investigation. One demonstrated technique relies on the shift of energy levels in a semiconductor under illumination by laser light with frequencies below the fundamental bandgap. When the laser light is circularly polarized this 'ac Stark effect' shifts the energy levels of spin-up states and spin-down states differently, generating an effective magnetic field. Rotations of electron spin by angles greater than $\pi/2$ have been demonstrated in this ultrafast effective magnetic field in semiconductor quantum wells.

Transfer of Spin Coherence from Mobile Electrons to Other Spin Systems

Another fundamental phenomenon in spin coherence is the transfer of spin coherence from one system to another. If one spin system has an extremely long coherence time, this transfer can be viewed as establishing a reservoir of spin coherence that can be drawn on at later times. Explorations of transfer phenomena have focused on establishing spin coherence in a system that is coupled to the conduction electrons, such as core electron polarization or nuclear polarization. The polarization of these other spin systems can be detected because, when polarized, they exert an effective field on the conduction electrons. For example, when the nuclear system is polarized it exerts an effective field on the electrons through the hyperfine coupling.

It is also possible to polarize these other systems indirectly through their interaction with the conduction electron system. When the conduction electron system is driven far from equilibrium by optically pumping the spin polarization, the coupled spin systems also become spin polarized. In the case of the nuclear spin system this effect is the Overhauser effect. While continually generating conduction electron spin polarization with optical pumping the polarization of the nuclear system begins to rise with a time-scale characterized by the coupling between the nuclear spins and the conduction electron spins (hyperfine coupling). Because of the substantial difference between the nuclear spin coherence times parallel to and perpendicular to the applied magnetic field, the effective nuclear polarization is generated parallel to the applied magnetic field independent of the orientation of the conduction electron spin polarization. In bulk GaAs at 5 K the percentage polarization achieved was a few per cent, yielding an effective field through the hyperfine interaction of ~ 0.4 T. These nuclear polarizations and effective fields are increased over an order of magnitude in certain quantum wells, suggesting that nanostructure design can be used to control the effective fields. The induced nuclear polarization persists for time-scales characteristic of the nuclear longitudinal spin coherence time, on the order of minutes. If the optical pumping is turned off, the conduction electron spin polarization decays away rapidly, whereas the nuclear spin polarization persists for time-scales characteristic of the nuclear longitudinal spin coherence time, on the order of minutes. This nuclear hyper-polarization, and its manipulation through an optically pumped time-dependent conduction electron spin polarization, suggests a possible general technique of ultrafast all-optical nuclear magnetic resonance in solids.

A similar spin-coherence storage effect has been obtained for core electron polarization in Mn-doped ZnSe quantum wells. Again, the optically pumped

conduction electron spin polarization generates a nonequilibrium spin polarization of the paramagnetic core electron spins (3d in the case of Mn) through the core-conduction electron coupling. This core electron spin polarization persists long after the conduction electron spin polarization (and, in undoped systems, long after the conduction electron density) has decayed away. The magnetic polarization of the paramagnetic Mn spins dephases over the very long time-scales determined by spin–lattice relaxation, and thus is an interesting avenue towards information storage.

The above systems are not ferromagnetic – the spin systems involved in the process are paramagnetic spin systems coupled to the conduction electrons, which are 'super-polarized' through the conduction electron spin polarization. An alternative approach to manipulating the conduction electrons has recently been demonstrated using a thin layer of a ferromagnetic material placed on top of n-doped GaAs. Here the presence of this ferromagnetic material, even if only a few nanometers thick, leads to a remarkable enhancement of the nuclear spin polarization in the GaAs. For example in the presence of a mere 1000 G applied magnetic field, the ferromagnetic layer generates an effective field on the electrons from the nuclei of up to 9000 G. This effective field also depends on the orientation of the magnetism in the ferromagnetic layer. This intriguing effect has yet to be fully explored, but may yield a new type of magnetic 'gate' on the mobile electrons in a semiconductor without a direct form of electrical spin injection.

The Future of Semiconductor Spintronic Devices

Semiconductor charge-based electronics (e.g., transistors and integrated circuit chips) and magnetoelectronics (e.g., hard disk read heads) are two of the most successful electronic technologies of the twentieth century. A union of elements of these two technologies is anticipated in the area of semiconductor spintronics. Immediate approaches involve placing hybrid magnetic elements within known semiconductor devices. Semiconductor spintronic devices, however, are unlikely to compete with the most successful devices from each of these two areas – the semiconductor transistor is an extremely effective and efficient device, as is the magnetoelectronic hard disk read head. As has occurred with many other new technologies perhaps the most important applications of semiconductor spintronics have not even been envisaged yet.

Despite this, one significant application without a dominant technology suggests itself – that of quantum computation. Recent advances in the theoretical understanding of quantum information technology, including novel algorithms, error-correcting codes, and plausible scalable device architectures, suggest that calculations involving quantum mechanical coherent states are possible. Here the extremely long spin coherence times in semiconductor systems, and the demonstrated ability to manipulate that coherence on ultrafast optical times, indicate that spins in semiconductors (whether electron or nuclear) may constitute a good basis for a physical realization of a quantum computer.

See also

Quantum Optics: Quantum Computing with Atoms.
Semiconductor Physics: Outline of Basic Electronic Properties.

Further Reading

Abragam A (1961) *Principles of Nuclear Magnetism*. Oxford, UK: Oxford University Press.

Awschalom DD and Kikkawa JM (1999) Electron spin and optical coherence in semiconductors. *Physics Today* 52: 33–38.

Awschalom DD, Samarth N and Loss D (eds) (2002) *Semiconductor Spintronics and Quantum Computation*. New York: Springer.

Ivchenko EL and Pikus GE (1997) *Superlattices and Other Heterostructures*, 2nd edn. New York: Springer.

Jeffries CD (1963) *Dynamic Nuclear Orientation*. New York: Wiley.

Meier F and Zachachrenya BP (eds) (1984) *Optical Orientation*. Amsterdam, The Netherlands: North-Holland.

Slichter CP (1992) *Principles of Magnetic Resonance*, 3rd edn. New York: Springer.

Wolf SA, Awschalom DD, Buhrman RA, *et al.* (2001) Spintronics: a spin-based electronics vision for the future. *Science* 294: 1488–1495.

Yafet Y (1963) g factors and spin-lattice relaxation of conduction electrons. In: *Solid State Physics*, vol. 14. New York: Academic Press.

Surface Photovoltage Spectroscopy of Semiconductors

L Kronik, Weizmann Institute of Science, Rehovot, Israel

Y Shapira, Tel-Aviv University, Tel-Aviv, Israel

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Surface photovoltage spectroscopy (SPS) is a well-established contactless technique for semiconductor characterization, which relies on analyzing illumination-induced changes in the surface voltage. SPS traces its origins to the pioneering work of Brattain and Bardeen in the early 1950s, and was extended into a powerful spectroscopic tool by Gatos, Lagowski and Balestra in the early 1970s. It has been used as an extensive source of surface and bulk information on various semiconductors and semiconductor interfaces. In the following, we present the basic theory behind SPS, its experimental setup, and a range of its applications.

Principles of Operation

In general, a surface is defined as a boundary of media with different physical properties. The periodic structure of an ideal crystalline semiconductor results in the appearance of allowed energy bands separated by forbidden energy gaps. In particular, semiconductors are characterized by a nearly full 'valence band', separated by a forbidden bandgap from a higher lying, nearly empty 'conduction band.' Electron transport takes place within the latter and hole transport takes place within the former. The mere termination of this periodic structure at the surface, as well as the chemical changes in the surface associated with this termination, allow for the formation of surface-localized electronic states within the semiconductor bandgap. Electrons or holes in these local states are referred to as 'trapped carriers', as opposed to the 'free carriers' in the conduction and valence bands.

The appearance of surface-localized states induces charge transfer between bulk and surface in order to establish thermal equilibrium between the two. The free carrier density in the vicinity of the surface therefore deviates from its equilibrium value in the bulk. This results in a surface space charge region (SCR), which is electrically non-neutral, implying a nonzero electric field in it and therefore a potential drop across it. Thus, even under equilibrium conditions the surface potential, V_s, is different from the electric potential far away in the bulk. As a specific example which illustrates these concepts, consider an n-type semiconductor with a depleted surface. The majority carriers are electrons and some of them have been trapped in surface states, such that their concentration in the vicinity of the surface is smaller than its equilibrium value but larger than that of the holes (see Figure 1a). The potential drop across the SCR is manifested by the bending of the semiconductor bands, which is such that electrons are repelled from the surface and holes are attracted to it, due to the trapped surface electrons. We note that, by definition, the energy band edge is lower the higher the electrical potential, so that a positive V_s corresponds to downward-bent bands.

Generally, the photovoltaic effect comprises an illumination-induced change in the equilibrium potential distribution and is typically the result of some charge transfer and/or redistribution within the device due to the incident illumination. This is the basic principle behind devices ranging from photodiode detectors to photovoltaic solar panels. A specific variant of the photovoltaic effect is the *surface* photovoltaic effect, which is at the center of this text. It is important to note that the formation of a surface photovoltage (SPV) occurs only if carrier photogeneration *per se* is followed by net charge redistribution. Usually, no significant driving force for such redistribution is found beyond the SCR and the underlying bulk remains quasineutral. Thus, we are interested primarily in the surface SCR.

The SPV mechanism depends strongly on whether the incident photon energy is larger or smaller than the semiconductor bandgap (denoted below as the super-bandgap and sub-bandgap, respectively). Dominant SPV mechanisms are demonstrated below for a depleted n-type surface. The dominant mechanism by which super-bandgap illumination results in surface photovoltage generation is shown in Figure 1b. The electric field in the SCR causes excess electrons to be swept away from the surface and excess holes to be swept towards it. This serves to reduce the density of surface-trapped electrons and decreases the band-bending. In a second mechanism, shown in Figure 1c, either electrons or holes are preferentially trapped at surface defects. This effectively charges the surface and thus modifies the

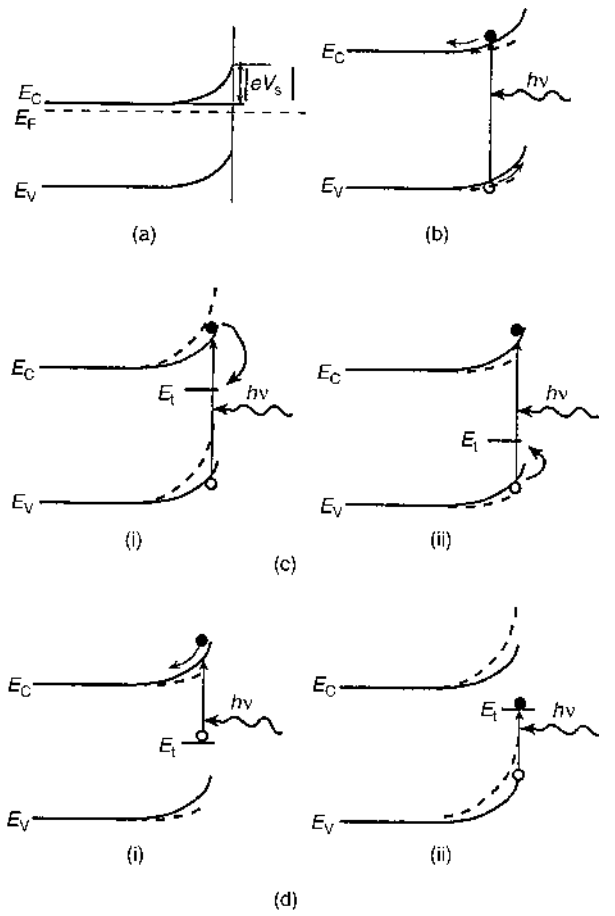

Figure 1 Schematic band diagrams of the surface space charge region at a depleted n-type semiconductor surface under different conditions: (a) thermal equilibrium; (b) super-bandgap illumination with carrier separation under an electric field; (c) super-bandgap illumination with preferential trapping of: (i) electrons, (ii) holes; (d) sub-bandgap illumination with excitation of trapped: (i) electrons, (ii) holes. In all diagrams, solid and dashed lines indicate band positions in the dark and under illumination, respectively. Straight arrows denote carrier generation and curved arrows denote carrier trapping. Adapted from Gal D, Mastai Y, Hodes G and Kronik L (1999) Bandgap determination of semiconductor powders via surface photovoltage spectroscopy. *Journal of Applied Physics* 86: 5573. Reproduced by permission of the American Institute of Physics.

surface potential. This second mechanism is usually apparent in bulk samples only if the equilibrium surface band-bending is fairly small, but increases in importance in polycrystalline materials with decreasing crystallite size, because of the increasing surface to volume ratio.

The most common mechanism for sub-bandgap SPV involves the direct modification of the surface charge, and hence potential, by excitation of trapped carriers, as shown in Figure 1d. Illumination by photons with energy $h\nu > E_c - E_t$ may produce electron transitions from a surface state at an energy E_t into the conduction band, where they are quickly swept to the semiconductor bulk by means of the surface electric field. Hence, the negative surface charge is reduced and the band-bending decreases. Conversely, illumination by photons with energy $h\nu > E_t - E_v$ may produce electron transitions from the valence band into a surface state situated at an energy E_t above the valence band maximum, E_v (which are equivalent to hole transitions from the surface state to the valence band). Such transitions increase the surface negative charge and therefore the surface band-bending. For a SPV to develop in this case, it is necessary to have a significant diffusion of the excess holes into the bulk and/or a significant recombination of electrons and holes inside the SCR, as there is no field-assisted driving force of holes into the bulk.

A typical SPV spectrum shows the SPV as a function of incident photon energy. Its most important features – the ones to watch for when attempting to interpret an SPV spectrum – are the sharp slope changes ('knees') associated with an abrupt onset of an additional carrier excitation mechanism. Consider, as an illustrative example, the top SPV curve in the right-hand part of Figure 2 (the rest of the figure is discussed below). In this curve, two 'knees' are shown. The first 'knee' (marked by an arrow) occurs at sub-bandgap photon energies and corresponds to the onset of carrier excitation from surface states (as in Figure 1d). The second 'knee' occurs in the vicinity of the bandgap energy and corresponds to the onset of band-to-band carrier excitation (as in Figure 1b,c). The proper interpretation of the energy position of these 'knees', their sign (negative or positive SPV change), and the relation of both to the SPV mechanism at work, is elaborated below in the 'Applications' sections.

So far, our discussion has been devoted solely to a study of semi-infinite, uniform bulk samples. However, actual semiconductor samples may be thin, i.e., not much larger than the diffusion length and/or the absorption length, so that excess carriers may be present near their back part as well. Many optically and electronically interesting structures (advanced detectors, transistors, lasers) consist of a multilayer structure, which has several buried interfaces.

Any change in the potential drop across a buried surface will also be measurable at the front surface, because potential drops are additive. For example, the photo-induced potential drop across a photodiode is measured at its contacts, which can be many microns away from its active region. Thus, the sensitivity of SPV measurements to a certain region in the semiconductor is limited only by the

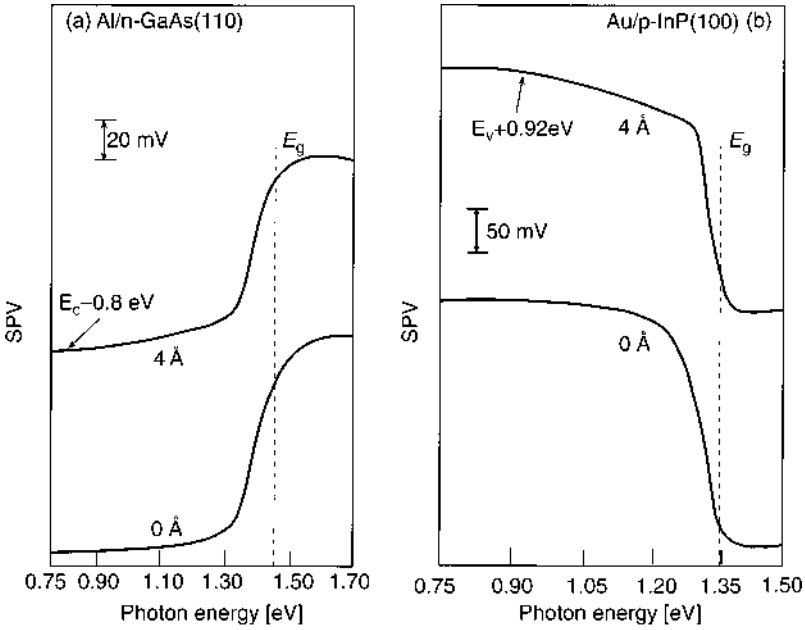

Figure 2 SPV spectra of: (a) an n-type GaAs(110) surface, UHV cleaved (top curve) and with an Al overlayer (bottom curve); (b) a p-type InP(110) surface, UHV cleaved (top curve) and with an Au overlayer (bottom curve). Arrows denote the onset of optically induced transitions and labels denote the corresponding defect level. Adapted from Burstein L, Bregman J and Shapira Y (1991) Characterization of interface states at III–V compound semiconductor–metal interfaces. *Journal of Applied Physics* 69: 2312. Reproduced by permission of the American Institute of Physics.

absorption length of the photons, i.e., by the possibility of introducing an excess of free carriers in the region of interest.

Experimental Details

SPV measurements are nontrivial because the surface potential is a *built-in* potential, rather than an external potential. In other words, the surface potential is not equal to the difference in Fermi levels between the front and back surfaces. Therefore, it cannot be simply measured using some form of voltmeter. Moreover, in the case of a free surface, the application of any contact to indirect electrical measurements of the built-in voltage (e.g., current–voltage or capacitance–voltage analyses) will invariably alter the surface properties and hence the quantity under measurement. Consequently, most of the prevalent techniques for measuring the surface potential (or at least changes of it upon illumination) are based on capacitive coupling schemes that do not require a direct electrical contact.

The two most commonly used methods rely on bringing a metallic probe close to the free surface, but never touching it, thus forming a parallel-plate capacitor. An alternating current, from which surface potential changes can be calculated, is then excited by (a) vibrating the metallic probe, which results in alternating capacitance, or (b) using 'chopped' illumination. The two approaches are known as the Kelvin probe and the ac-SPV schemes, respectively. In principle, they provide similar information, but their somewhat different physical principles result in several relative strengths and weaknesses, which make the method of choice application-dependent. In a nutshell, the Kelvin probe is more suitable for following phenomena with large time constants (in particular surface states with long relaxation times) and does not perturb the true band-bending of the free surface. The ac-SPV technique is more suitable for following fast relaxation phenomena, allows for a systematic modification of the surface band-bending in the dark as an extra degree of freedom, and does not require an ohmic back-contact.

Both Kelvin probe and ac-SPV measurements may also be employed in a scanning mode, which allows for recording lateral variations in the SPV, by scanning a small probe or a well-focused light beam, respectively, across the surface. Both approaches are typically limited to a lateral resolution roughly of the order of ~1 μm. SPV measurements can also be performed by a limited modification of two high-resolution techniques for topographic

mapping of surfaces that have emerged in the 1980s: scanning tunneling microscopy (STM) and atomic force microscopy (AFM). This opens new horizons for SPV mapping with a resolution on the nanometer scale or even the atomic scale.

A schematic view of a complete, generic SPS setup is given in Figure 3. The sample under study is typically placed inside a metallic box, which serves as both a dark box and a Faraday cage. This box must be such that it effectively includes the probe.

Clearly, for spectroscopic purposes one needs to generate the SPV signal using a broadly tunable light source. The default choice would therefore be the use of a 'white-light' source in conjunction with a monochromator. Because the typical super-bandgap absorption coefficient is typically orders of magnitude larger than the sub-bandgap coefficient, SPS is vulnerable to the spurious contribution of high-order diffraction peaks and stray light. Use of a monochromator where these effects are as small as possible is therefore highly recommended. For some applications, use of an auxiliary light source (e.g., white light, laser light, ultraviolet light, etc.) allows for a 'photobias' of the system and/or for additional quantitative analysis.

The actual recording of an SPV spectrum is coordinated by a real-time computer program. This program controls the monochromator motor and thus wavelength changes, operates the control electronics of the probe, reads the measurement results by interfacing with the probe read-out electronics via an analog-to-digital converter, and finally stores the obtained spectrum for subsequent display and analysis.

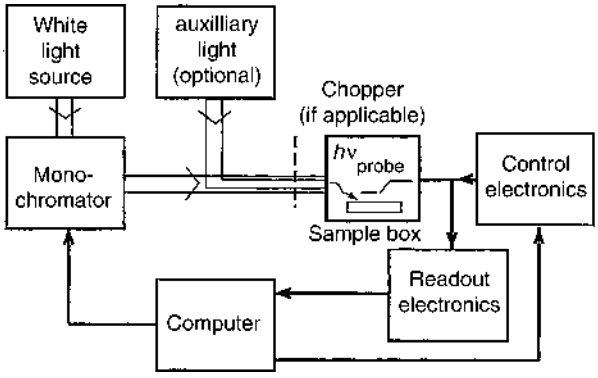

Figure 3 Schematic block diagram of a generic SPS setup. Adapted from Kronik L and Shapira Y (2001) Surface photovoltage spectroscopy of semiconductor structures: at the cross-roads of physics, chemistry and electrical engineering. *Surface and Interface Analysis* 21: 954. New York: John Wiley and Sons, Inc.

Applications I: Uniform Semiconductors

An elementary application of SPS is the approximate determination of the semiconductor bandgap. This determination is based on the large increase in absorption coefficient near the bandgap energy, E_g, found in most semiconductors. This increase brings about a significant change of the SPV signal, which is easily identified as a sharp change in slope of the SPV curve and is often the most significant one in a given spectrum. An example is shown in Figure 2, where the slope changes related to the bandgaps of GaAs and InP are clearly identified. Indeed, the bandgaps of most known semiconductors have been analyzed using this approach.

The use of SPS for extraction of E_g is at heart nothing more than an emulation of an absorption spectrum. However, as opposed to, e.g., transmission spectroscopy, SPS does not require light collection. It can therefore be performed on arbitrarily thick samples (or on layers sandwiched within a heterostructure – see below) and does not require the sample to be removed from the substrate or grown on a transparent one. It is also inherently insensitive to reflection and scattering and is thus eminently useful for micro- and nanocrystallites.

One must bear in mind that the value of E_g obtained from an SPS plot by inspection is only approximate. Indeed, detailed experimental comparisons between SPV and absorption spectra reveal that the two are often similar, but never identical. Thus, the nominal bandgap is nearly always found within the onset of the largest SPV signal, but is usually relatively broad, the exact position of E_g within it is by no means obvious, and the error in E_g is often ~ 0.1 eV. However, it has been shown that this error can be reduced significantly – often by one to two orders of magnitude – by a careful quantitative analysis of the experimental data.

We now turn our attention to determining the type (p or n) of the semiconductor. This may be achieved with the help of the sign of the knee associated with the onset of the super-bandgap SPV. Most semiconductor surfaces are depleted, which means that the bands of p-type semiconductors are bent downwards towards the surface, whereas the bands of n-type semiconductors are bent upwards. Since super-bandgap illumination typically tends to decrease the surface band-bending, this would result in a positive SPV in n-type semiconductors and a negative SPV in p-type semiconductors. For example, the SPV spectra of n-type GaAs and p-type InP, shown in Figure 2, clearly feature opposite onset signs which obey the above rules.

While in many cases the type of the semiconductor is known a priori, this is not always the case and SPS can become very useful in determining the semiconductor type. For example, SPS was used for fast and nondestructive verification of the semiconductor type in GaN films, where the p-doping is known to be nontrivial and subject to compensation by its native n-type doping.

We note that under conditions where the SPV is dominated by trapping of carriers, as shown in Figure 1c, the sign convention described above clearly does not apply. For example, etched CdSe quantum dot films were found to exhibit a p-type response in a humid ambient and an n-type response in a dry ambient, on account of preferential trapping of holes or electrons in a dry or humid ambient, respectively.

A key strength of SPS is in the determination of the energy position and type of surface states. Because photons of sufficient energy may excite charge carriers from a surface state to a band or vice versa, we expect a knee in the SPV spectrum whenever the photon energy exceeds the threshold energy of a certain transition. If the latter involves excitation of electrons from the surface state to the conduction band (Figure 1d(i)),a positive change in the surface charge is induced, and hence a positive SPV is expected. Conversely, excitation of holes to the valence band (Figure 1d(ii)) makes the surface charge more negative and a negative SPV is expected. Thus, the combination of the SPV threshold energy and slope sign makes the determination of the approximate position of the surface state within the bandgap possible.

As an illustrative example, consider how SPS can be used to monitor metal-induced surface states, as shown in Figure 2. After cleaving in ultrahigh vacuum, neither GaAs nor InP surfaces display any knees associated with deep surface states, as expected for these surfaces. However, upon deposition of a very thin metallic overlayer, distinct sub-bandgap knees emerge. These knees, associated with positive and negative SPV changes at the Al/GaAs and Au/InP interfaces, respectively, indicate the formation of surface states situated ~0.8 eV below the conduction band edge and ~0.92 eV above the valence band edge, respectively.

Because SPS is contactless and nondestructive, it can be applied not only to clean surfaces, but also to real ones, in practically any ambient. This makes SPS highly suitable for a direct correlation of surface electronic structure with chemical treatments. One such area of active research is the study of surface passivation treatments. An illustrative example, in which the passivating effects of various liquids on a specific p-InP surface (denoted as (100)) was studied *in situ*, is shown in Figure 4. The negative

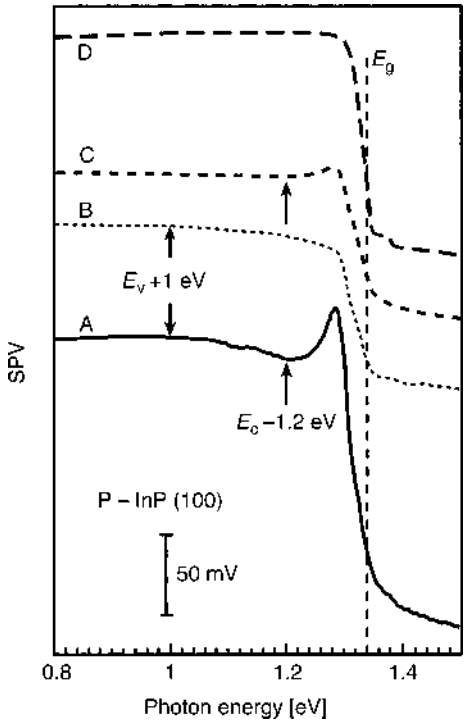

Figure 4 SPV spectra of the p-InP(100) surface in air and after successive immersion in (from top to bottom in the lower energy part): A – 18 MΩ H_2O; B – 0.4% HF; C – 30% H_2O_2; D – a 1.3 M Na_2S solution. Arrows denote the onset of optically induced transitions and labels denote the corresponding defect level. Adapted from Bastide S, Gal D, Cahen D and Kronik L (1999) Surface photovoltage measurements in liquids. *Review of Scientific Instruments* 70: 4032. Reproduced by permission of the American Institute of Physics.

SPV change at ~1.3 eV found in all spectra shown is clearly due to the onset of band-to-band transitions in the InP sample, with its sign indicating a p-type sample, as appropriate. The SPV spectrum of the water-exposed sample (which is similar to that obtained in air (not shown)) features two surface states: the negative SPV slope change, at ~1 eV, is due to a photo-induced population of a surface state situated ~1 eV above the valence band edge. The positive SPV slope change at ~1.2 eV is due to a photo-induced depopulation of a surface state situated ~1.2 eV below the conduction band edge. Interestingly, exposure to HF eliminates the latter state, but not the former. The opposite is true for exposure to H_2O_2. Finally, exposure to a Na_2S solution results in removal of both surface states, in agreement with the known passivating action of this solution. This is just one example of many illustrating the strength of SPS for revealing the link between chemical/physical treatments and device-affecting electronic properties, already at the material level.

While we have so far considered only surface states, SPS is sensitive to bulk states too, as long as charge excited from or trapped in them can be separated (typically under the electric field of a space charge region). Accordingly, SPS has been used for studying, e.g., Cr-related defects in GaAs or Fe-related defects in InP. The simplest distinction between surface and bulk states in SPS is obtained by examining whether the pertinent spectral feature disappears after surface (e.g., etching) or bulk (e.g., doping) treatment, respectively.

We note that in recent years, SPS has also been applied successfully along the lines outlined here for the study of bandgaps, doping type, and surface and bulk defects in the emerging field of organic semiconductors and organic treatments of semiconductors, as well as for nanostructures, such as fullerenes and nanotubes.

Additional, quantitative analyses of the SPV spectra allow for a determination of densities, thermal and optical transition rates, and the spatial site of gap states. Such analyses are usually based on the SPV response to a systematic modification of one or more external parameters, e.g., time (for transient responses), illumination intensity, temperature, etc.

Applications II: Multilayer Structures

Beyond the study of surface properties, the application of SPS to heterojunctions, multilayer structures, and actual device structures has been most rewarding in recent years. The SPS of multilayer structures is illustrated by Figure 5. This figure features the SPV spectra of a ZnO:Al/ZnO/CdS/Cu(In,Ga)Se_2 layered structure, used for the fabrication of thin-film solar cells. Distinct spectral knees at ~1 eV, ~2.4 eV, and ~3.1 eV, corresponding to the bandgaps of Cu(In,Ga)Se_2, CdS, and ZnO, respectively, are clearly observed. The unequivocal identification of all three bandgaps confirms the above-mentioned theoretical prediction of SPS sensitivity to buried interfaces. Interestingly, the magnitude of the SPV signal is seen to increase with increasing annealing time (in air, at 200 °C), in agreement with the increase in overall solar cell efficiency with such annealing.

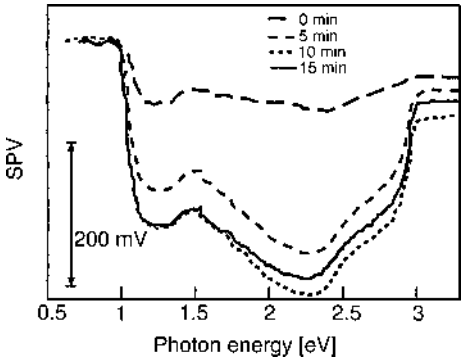

Figure 5 SPV spectra of ZnO:Al/ZnO/CdS/CIGS structures for various annealing times, in air, at 200 °C. Adapted from Gal D, Beier J, Moons E, *et al.* (1995) Band diagram and band line-up of the polycrystalline Cds/Cu(In,Ga)se_2 heterojunction and its response to air annealing. *AIP Conference Proceedings* 353: 453. Reproduced by permission of the American Institute of Physics.

Another feature of growing importance in SPS is that it allows for the characterization of quantum wells and other low-dimensional heterostructures. An illustrative example of the SPS of a multiple-quantum-well (MQW) GaAs/AlGaAs structure is shown in Figure 6a. The spectral knees at ~1.4 eV and 1.78 eV correspond to the onset of band-to-band absorption in the GaAs epilayer and the AlGaAs cap layer, respectively. In between these two knees, the shape of the SPV spectrum resembles typical absorption spectra of MQW structures and the observed MQW-related maxima agree well with calculated energy values of the hole level-electron level transitions denoted as 1HH-1E and 2HH-2E transitions. Note that before etching, a nonnegligible sub-bandgap SPV signal was apparent at photon energies below 1.38 eV, indicating the presence of optically active gap states. Since no such signal was observed after etching, these states were located at the external AlGaAs surface and were not associated with the MQW region. Moreover, the SPV spectrum in the MQW absorption energy range has not changed at all (other than undergoing a uniform shift) after etching, showing that the signal SPV in the MQW absorption range is not interfered by surface effects. Thus, all allowed hole–electron transitions have been resolved, at room temperature.

As SPV analyses of multilayer structures offer the possibility of performing contactless and nondestructive electrical and optical characterization of buried interfaces, they have found significant uses in the field of process monitoring and quality control of actual device structures. This is because the possibility of assessing the quality of a layered structure before any actual device processing takes place presents an opportunity for significant savings in time and money. Recent applications include the identification of defective batches and prediction of ultimate performance of solar cells and x-ray detectors, the prediction of energy levels and lasing wavelength in modern laser heterostructures, the assessment of bandgap narrowing and ultimate gain of heterojunction bipolar transistors, and the

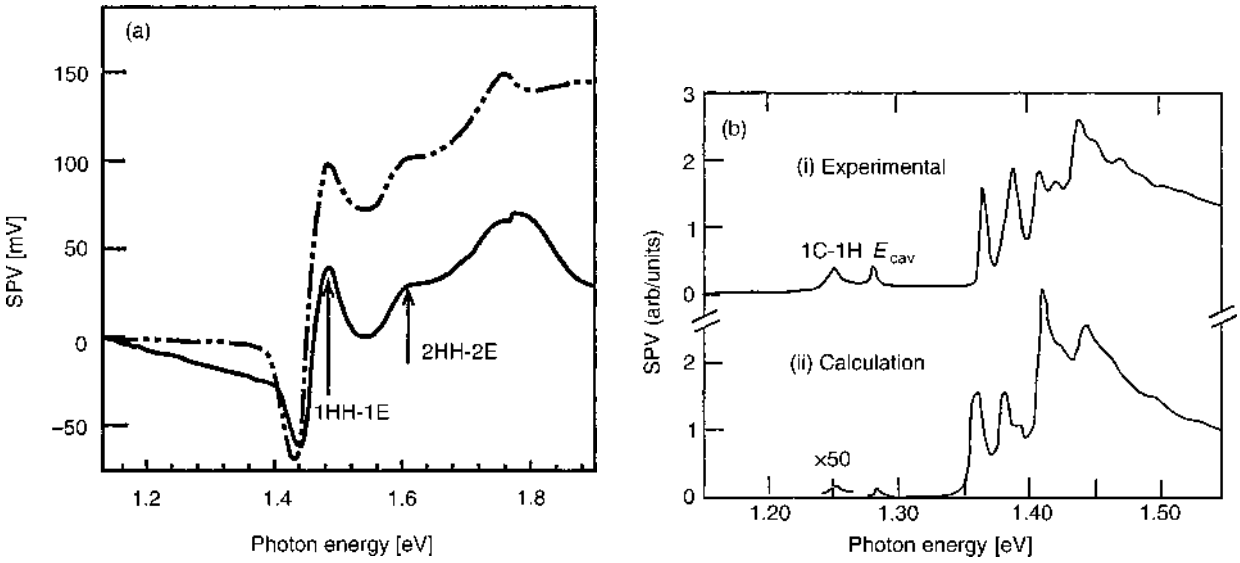

Figure 6 (a) SPV spectra of a GaAs/AlGaAs MQW sample before (solid curve) and after (dashed curve) etching. Adapted from Bachrach-Ashkenasy N, Kronik L, Shapira Y, *et al.* (1996). Surface photovoltage spectroscopy of quantum welts and superlattices. *Applied Physics Letters* 68: 879. Reproduced by permission of the American Institute of Physics. (b) Room temperature experimental SPV spectra (i) and calculated photo-excited carrier spectra (ii) of an InGaAs/GaAs/AlGaAs vertical-cavity surface-emitting laser structure. Adapted from Huang YS, Malikova L, Pollak FH, *et al.* (2000). Surface photovoltage spectroscopy, photoreflectance, and reflectivity characterization of an InGaAs/GaAs/GaAlAs vertical-cavity surface-emitting laser including temperature dependence. *Applied Physics Letters* 77: 37. Reproduced by permission of the American Institute of Physics.

characterization of the electronic structure of high electron mobility transistors and vertical cavity surface emitting lasers.

As an example of such studies, consider the recent characterization of an In-GaAs/GaAs/AlGaAs vertical-cavity surface-emitting laser, shown in Figure 6b. It is readily observed that SPS resolves both the gain providing 1C-1H transition and the lasing-wavelength-related cavity mode, as well as a rich oscillatory structure above ~1.37 eV due to interference effects from the distributed Bragg reflector stacks. The measured SPV spectrum is also in very good agreement with the calculated dependence of the photo-excited carrier density on photon energy, also shown in Figure 6b. In fact, SPS is the first contact-less technique to resolve all of these features in one measurement, and it also allows for a temperature-resolved determination of these quantities.

Summary

SPS features a theory that is relatively straightforward (at least at the qualitative level), and offers typical experimental setups that make the method contactless, nondestructive, ambient insensitive, and capable of any lateral resolution. It can provide a detailed picture of the electronic structure at surfaces, interfaces, heterojunctions, and even complicated multilayer structures. The wide range of SPS applicability also makes it easy to compare and link changes in electronic structure to changes in chemical/physical properties on the one hand, and electrical/optical device properties and performance on the other.

See also

Semiconductor Materials: Modulation Spectroscopy of Semiconductors and Semiconductor Microstructures. **Semiconductor Physics:** Band Structure and Optical Properties; Impurities and Defects; Outline of Basic Electronic Properties; Quantum Wells and GaAs-Based Structures.

Further Reading

Brattain WH and Bardeen J (1953) Surface properties of germanium. *Bell System Technical Journal* 32: 1–41.

Brillson LJ (1982) The structure and properties of metal–semiconductor interfaces. *Surface Science Reports* 2: 123–326.

Gatos HC and Lagowski J (1973) Surface photovoltage spectroscopy – a new approach to the study of high-gap

semiconductor surfaces. *Journal of Vacuum Science and Technology* 10: 130–135.
Kronik L and Shapira Y (1999) Surface photovoltage phenomena: theory, experiment, and applications. *Surface Science Reports* 37: 1–206.
Kronik L and Shapira Y (2001) Surface photovoltage spectroscopy of semiconductor structures: at the crossroads of physics, chemistry and electrical engineering. *Surface and Interface Analysis (special issue on Optical Techniques)* 21: 954–965 (main source for the present text).
Lüth H (1993) *Surfaces and Interfaces of Solids*, 2nd edn. Berlin: Springer-Verlag.
Many A, Goldstein Y and Grover NB (1971) *Semiconductor Surfaces*, 2nd edn. Amsterdam, The Netherlands: North-Holland.
Pollak FH (2000) Modulated electric fields produce semiconductor spectra. *Laser Focus World* 36: S9.

SOLITONS

Contents

Bright Spatial Solitons
Optical Fiber Solitons, Physical Origin and Properties
Soliton Communication Systems
Temporal Solitons

Bright Spatial Solitons

C Conti and G Assanto, University 'Roma Tre', Rome, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical spatial solitons, once considered an exotic and rather abstract concept belonging to the realm of mathematics or mathematical physics, nowadays are pursued both experimentally and theoretically in terms of their rich dynamics and variegated features, with important perspectives in applications to signal processing and switching, routing, filtering and – generally – light-by-light controlling. In an intuitive description of a bright spatial soliton, a self-focusing response balances natural diffraction of a light beam, allowing for its invariant propagation in transverse size and intensity. Spatial solitons can be regarded as the stationary or time-independent self-guided 'eigenwaves' of Maxwell equations in a material system with a nonlinear optical response. Basically, 'bright' spatial solitons (we will not address 'dark' cases here, i.e., a dip on an illuminated background) bear the same relationship to nonlinear optics in space as temporal, or fiber, solitons do in time. While both temporal and spatial solitons are special cases of light localization in the (3 + 1) dimensions (light bullets) encompassing time and two transverse coordinates in space in addition to propagation; the spatial case requires particular attention compared to the fiber inasmuch as, in general, two transverse coordinates (2D) complement the direction of propagation. In other words, the scenario is potentially one-dimension richer than for 'standard' temporal solitons, where light confinement, i.e., the effect of the nonlinearity, takes place along a single coordinate. In this respect, the latter case finds a direct counterpart with spatial solitons in planar waveguides, i.e., (1 + 1)D solitons, where light confinement occurs linearly in one spatial dimension, providing group-velocity dispersion in time is replaced by beam diffraction in space, respectively.

When both transverse dimensions are involved, i.e. when a light beam propagates in bulk, in the simplest case of Kerr self-focusing, the nonlinearity can become overcritical and cause filamentation and catastrophic collapse. Additional mechanisms (such as saturation, nonlocality, or higher-order effects) or inherently different nonlinear responses (parametric or molecular), need therefore to be involved in the description of spatial solitons. In several instances, the mathematics differs from the nonlinear Schrödinger equation originally invoked to define a 1D bright optical soliton, leading to nonintegrable systems and

to a wider class of solutions, called solitary waves. Since solitary waves constitute the vast majority of the observed spatial light-localization phenomena, we will refer to spatial solitons even in those cases when their model is nonintegrable, defining them as non-diffracting beams in a self-focusing material. While retaining the distinction between 1D and 2D cases, in the following sections we will discuss bright spatial solitons and their features, grouping them on the basis of the underlying nonlinear mechanism and corresponding model, trying to provide a direct link with experimental observations. In doing so, we will leave out discussions on dark solitons and modulationally unstable phenomena, as well as more involved subjects ranging from higher-order to vector, vortex, photonic-crystal, and incoherent spatial solitons.

The General Model of Bright Spatial Solitons

Assuming the paraxial approximation to be valid, the general structure of the evolution equations which lead to spatial solitons propagating along z takes the so-called Foch–Leontovich form:

$$2\mathrm{i}k\frac{\partial A}{\partial z} + \nabla_\perp^2 A = -2k^2\frac{\Delta n}{n_0}A \qquad [1]$$

where A is the complex amplitude of the linearly polarized electric field, $k = \omega/c = 2\pi n_0/\lambda$, with λ the wavelength, $\omega = 2\pi\nu$ the pulsation, n_0 the bulk refractive index, and $\Delta n(\mathbf{r})$ the index change induced by the nonlinear effect. In anisotropic materials, both n_0 and Δn may depend on the polarization of the envelope A. The latter is normalized such that the optical intensity is $I = |A|^2$. Hence the whole optical field is

$$E_{\mathrm{opt}} = \mathrm{Re}\left[\sqrt{\frac{2Z_0}{n_0}}A\mathrm{e}^{-\mathrm{i}\omega t+\mathrm{i}kz}\right] \qquad [2]$$

with Z_0 being the vacuum impedance.

Hereafter we will neglect material losses and assume $\Delta n(\mathbf{r})$ to be real-valued. In the simplest case it is a function of the field intensity I calculated in $\mathbf{r}$. Typical models are the Kerr medium with $\Delta n = n_2 I$, or the saturable Kerr with $\Delta n = n_2 I/(1 + I/I_s)$ and I_s a saturation value; they are particular cases of local nonlinear media defined by a relation of the type $F[\Delta n(\mathbf{r}), I(\mathbf{r})] = 0$, with F a 'well-behaved' function.

There is also the possibility that the refractive index modification depends on the electromagnetic perturbation (e.g. the intensity) in a finite region, as happens with nonlocal solitons. In the latter case, F needs to be substituted by an integro-differential system of equations, i.e., it represents a functional between $\Delta n(\mathbf{r})$ and $I(\mathbf{r})$.

Relevant exceptions to eqn [1] are spatial solitons due to a quadratic nonlinearity, where a system of two coupled equations must be taken into consideration, and a description in terms of a refractive index change only holds in an approximate, and by no means physical, sense. Quadratic solitons will be discussed later. Here, instead of embarking into a general classification, we will describe the relationships to some representative cases. In this simple framework, spatial solitons are defined as solutions of the type $A = a\exp(\mathrm{i}\beta z)$, with a real-valued and satisfying the equation:

$$\frac{1}{2k^2}\nabla_\perp^2 a + \frac{\Delta n(I)}{n_0}a = \frac{\beta}{k}a \qquad [3]$$

Equation [3] is the z-independent counterpart of eqn [1]. It has the form of the stationary Schrödinger equation (with a field-dependent Hamiltonian operator), where the kinetic energy and the potential are given by the diffraction term and by the nonlinear correction to the refractive index, respectively. Its solutions correspond to solutions of eqn [1] due to the balance between diffraction and nonlinearity. β is the nonlinear correction to the dispersion relation $k = n_0\omega/c$, i.e., it yields the resulting longitudinal wavevector $K = k + \beta$. Solitary solutions are often classified in terms of their dimensionality. Solutions of eqn [3], which depend on either 1 or 2 transverse coordinates, define 1D or 2D solitons, respectively. 1D solitons refer to situations in which a bulk nonlinear medium is illuminated with a highly elliptic beam (such that the diffraction along a transversal dimension, denoted by y, is negligible), or a planar waveguide is excited with a linear mode fixing the transverse profile in one of the dimensions (denoted by y). The nonlinear dynamics and the corresponding steady-state solutions are then described by eqns [1] and [3], with $\partial_y = 0$.

Spatial Solitons in Kerr Media

1D Kerr Spatial Solitons and the Nonlinear Schrödinger Equation

The simplest example of a spatial soliton is an optical beam diffracting in one transverse dimension in the presence of the so-called Kerr effect, i.e., a refractive index that increases linearly with the local intensity I:

$$n = n_0 + n_2 I \qquad [4]$$

where $n_2(>0)$ is measured in $m^2\,W^{-1}$ and has typical values of 10^{-20} for silica and 10^{-17} for semiconductors at photon energies below half-bandgap, i.e., leaving aside two photon absorption and higher-order effects. For 1D Kerr spatial solitons, eqn [1] becomes

$$2ik\frac{\partial A}{\partial z} + \frac{\partial^2 A}{\partial x^2} = -2k^2\frac{n_2}{n_0}|A|^2 A \qquad [5]$$

Equation [5] is known as the nonlinear Schrödinger equation (NLS), and is the basic model in describing temporal soliton, i.e., nondispersing pulses propagating in optical fibers, provided the variable x is intended as the temporal coordinate. The importance of eqn [5] is due to several remarkable, in some sense unique, mathematical properties, and to the fact that it can be considered a fundamental model for diverse nonlinear phenomena. This circumstance is often referred to as the 'universality of the NLS'.

In the framework of optical spatial solitons, the NLS relevance is moderated by the fact that most experimentally explored environments involve processes significantly more complicated than the Kerr effect. Nevertheless, it settles the basic principles and constitutes a reference point in optical soliton assessment.

Trying to gain some physical insight, the starting point is the beam spreading when $n_2 = 0$ in eqn [5], i.e., the linear regime corresponding to low intensities. The diffraction is inherent to the spectrum of plane waves in a finite beam (along x). To underline the role of the nonlinear response, one can consider the NLS without the diffraction term (the second derivative with respect to x), obtaining:

$$A(x,z) = \sqrt{I(x)}\,\exp\left[ik\frac{n_2}{n_0}I(x)z\right] \qquad [6]$$

Equation [6] shows how the nonlinearity yields a phase-front curvature related to the sign of n_2: the medium is self-focusing (a converging beam) for $n_2 > 0$ or self-defocusing (a diverging beam) for $n_2 < 0$ (see **Figure 1**). In the absence of diffraction, i.e., for a plane wave, the Kerr effect modifies the field phase by increasing ($n_2 > 0$) or decreasing ($n_2 < 0$) the local refractive index (and hence the effective optical path) proportional to the intensity. For a beam of finite transverse extent, nonlinearity and diffraction act together. In a self-focusing medium, the concave phase-front due to diffraction can be balanced by the nonlinearity, and the natural tendency of a beam to spread can therefore be compensated. For this to happen, the length-scales associated to the diffraction, i.e., the Rayleigh range $L_R = kw_0^2$, and the Kerr effect must be comparable. A characteristic length $L_N = n_0/(kn_2I_0)$, over which the nonlinear phase shift equals 1, can be associated

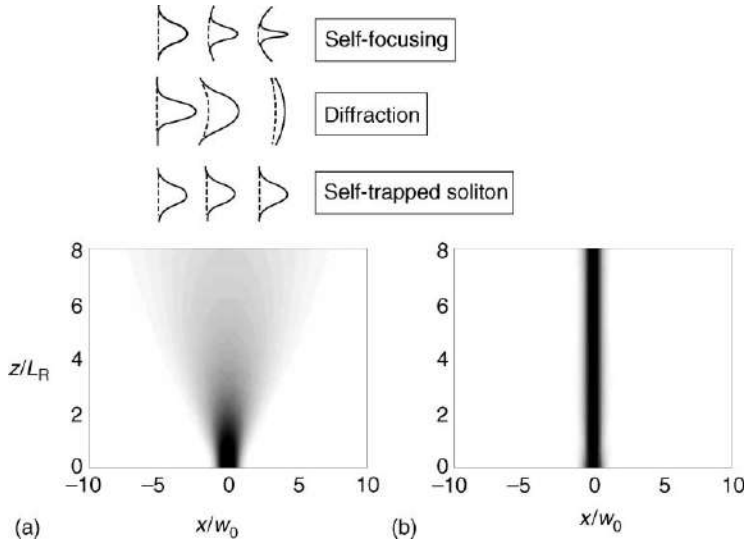

Figure 1 Kerr spatial solitons. Top: sketch of the mechanism leading to Kerr soliton formation. In the absence of diffraction, self-focusing of a Gaussian beam provides a converging phase-front (dashed lines). Conversely, diffraction by itself causes beam divergence. When these two effects act jointly, they can balance each other and soliton generation can be observed, with a flat phase-front. (Adapted with permission from Stegeman GI, Christodoulides DN and Segev M (2000) Optical spatial solitons: historical perspectives. *IEEE Journal of Selected Topics in Quantum Electronics* 6:1419–1426). Bottom: beam propagation method used to simulate the propagation of a one-dimensional Gaussian beam by integrating the nonlinear Schrödinger equation, for (a) low-power input, (b) input high enough to generate a self-trapped solitary wave. The transverse coordinate x is expressed in units of the beam waist, the propagation coordinate z in Rayleigh lengths.

to the nonlinearity, with I_0 the peak intensity across the beam profile. When the two lengths L_R and L_N are nearly equal, i.e., $w_0^2 I_0 \approx n_0/(k^2 n_2)$, the balance between the phase-front curvatures, due to diffraction and nonlinearity, yields a self-localized nondiffracting beam. Such a situation is described by the exact solutions of eqns [1] and [3], which takes the form:

$$A(x,z)=a(x)\mathrm{e}^{\mathrm{i}\beta z} \quad \text{with} \quad a(x)=\sqrt{I_0}\,\mathrm{sech}\left(x\sqrt{\frac{k^2 n_2}{n_0}I_0}\right) \qquad [7]$$

and $\beta = n_2 I_0/2n_0$ is determined by the peak intensity or, equivalently, by the power P_1 carried by the soliton per unit of transverse wavefront along y:

$$P_1=\int_{-\infty}^{\infty}|A|^2\,\mathrm{d}x=\sqrt{\frac{4n_0 I_0}{k^2 n_2}}$$

Note that the overall phase-front is flat, and the spatial extension in width yields $1/(n_2 k^2 I_0)^{1/2}$, as anticipated. The relationship between the soliton width – which determines the strength of diffraction – and the intensity – responsible for the nonlinear phase shift – is known as the existence curve, and is characteristic of the specific nonlinear mechanism.

Equation [7] is the fundamental soliton of the NLS and the paradigmatic 1D spatial soliton. Its sech shape specifically results from the type of nonlinearity; as discussed below, other effects (including parametric generation) can yield different profiles.

Figure 1 shows the simulation of a 1D Kerr soliton generated from a Gaussian input. Self-trapping is attained when the peak intensity or, equivalently, the power, reaches a threshold value depending on experimental details, the size of the nonlinear response, and the beam profile.

Figure 2 shows experimental results on the formation of a 1D Kerr soliton in a semiconductor (AlGaAs, $n_0 \approx 3.5$) waveguide. At low power (center panel) the output beam exhibits diffraction, whereas the soliton (bottom) reproduces the input beam profile (top panel) at the end facet of the sample. These results were obtained at $\lambda = 1.55\ \mu$m, using pulsed laser beams, with temporal pulse durations of about 1 ps and peak powers of about 500 W; the generated 1D solitons had waists of tens of microns.

In a physically intuitive sense, self-trapping is due to the refractive index increase across the beam profile. Such an increase gives rise to a graded index optical waveguide able to confine all the plane wave components with angles of propagation within the numerical aperture of the input beam. Large spatial-frequency components k_x cannot be trapped and are emitted as radiation during soliton formation. It is possible to experimentally verify that this self-induced waveguide is also able to guide a weak probe uncorrelated to the soliton, provided it is above the cutoff for dielectric confinement. The use of spatial solitons for intensity-controlled or all-optical lightwave circuitry can therefore be envisaged. The first experiments on optical 1D Kerr spatial solitons date back to 1985, and were performed in carbon disulfide. Since then, 1D Kerr spatial solitons have been reported in glasses, semiconductors, and organic materials.

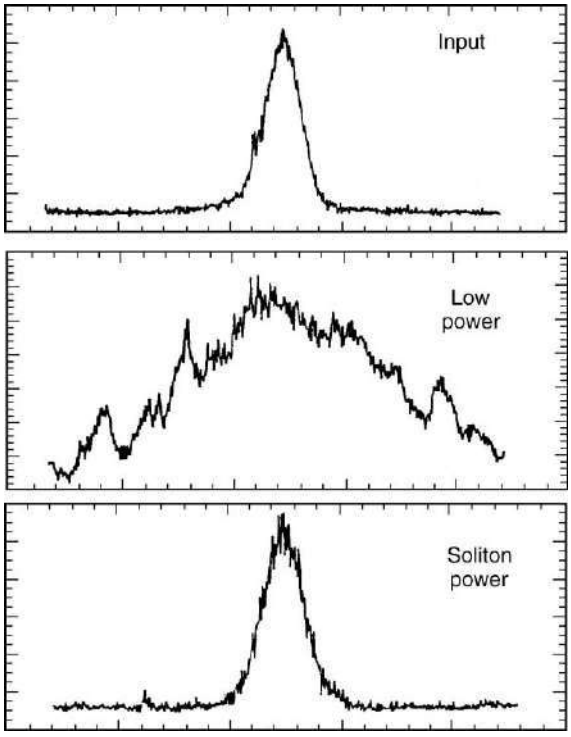

Figure 2 1D Kerr soliton in a planar AlGaAs waveguide. Transverse beam profiles at the input (top) and output facets of a planar waveguide in AlGaAs, at low (center) and high (bottom) power, showing diffraction and soliton formation, respectively. (Adapted with permission from Aitchison JS, Al-hemyari K, Ironside CN, Grant RS and Sibbett W (1992) Observation of spatial solitons in AlGaAs waveguides. *Electronic Letters* 28: 1879.)

Kerr Spatial Solitons in Two Transverse Dimensions: Saturable and Nonlocal Media

When the dynamics of the optical beam encompasses two transverse dimensions, the idealized discussion above no longer applies. As long as the medium is described by the simple Kerr law (eqn [4]), 2D propagation does not admit stable self-trapped

stationary beams. Because of this, two alternatives are available: (i) the beam diffracts; or (ii) the beam progressively self-focuses until it undergoes catastrophic collapse: in the absence of additional limiting phenomena, such as plasma formation or nonparaxial/vectorial coupling, its waist keeps shrinking with a corresponding unbound increase in peak intensity (beam power is conserved). Which regime actually applies, depends on whether the excitation is lower or higher than a self-focusing critical power $P_c = (0.61\lambda)^2/8n_0n_2$. The latter corresponds to a diffraction-driven divergence equal to the critical angle for total internal reflection in the induced waveguide. The self-focusing instability, first predicted by Kelly in 1965, is related to the so-called Derrick's theorem: particularly related to spatial solitons in a positive ($n_2 > 0$) Kerr medium, it states that the equilibrium between diffraction and nonlinearity in 2D is unstable.

The situation is markedly different when saturation is present: self-focusing arrests at the maximum value of the nonlinear refractive index increase, and stable self-trapped beams are allowed. The simplest model for a saturating nonlinearity yields the following stationary-state equation:

$$\frac{1}{2k^2}\nabla_\perp^2 a + \frac{n_2}{n_0}\frac{I}{1+I/I_s}a = \frac{\beta}{k}a \qquad [8]$$

Its solutions, while retaining a bell-shaped profile, depart from the purely Kerr model corresponding to $I_s \to \infty$, and can only be obtained numerically.

Gases, and specifically sodium vapors, were the first media in which spatial solitons of a saturable Kerr-like nonlinearity were observed in 1974. Historically, these were the first 2D optical spatial solitons.

Most recent developments on optical spatial solitons involve more complicated environments and nonlinear responses. In the next sections we will consider those most investigated experimentally: photorefractive solitons, nonlocal solitons in liquid crystals and parametric solitons.

Spatial Solitons in Photorefractive Media

In recent years, research on spatial optical solitons in photorefractives has been among the most active. Self-trapping in photorefractive materials was first predicted in 1992 and observed in 1993. However, a clear assessment of this result came with the demonstration of the so-called steady-state screening photorefractive solitons in 1994.

The principle underlying the photorefractive response can be described as a feedback mechanism relating to the optical induction of a static electric field **E**. In specific materials, such as semiconductors or ferroelectrics, a light beam generates free charges which diffuse and drift under an externally applied voltage. Owing to free carriers and donor/acceptor sites, the resulting net spatial charge produces a static electric field which, in turn, modifies the refractive index n_0 via the electro-optic effect, phenomenologically described by

$$\Delta n_{ij} = -\frac{1}{2}n_0^3 r_{ijk}E_k - \frac{1}{2}n_0^3\varepsilon_0^2(\varepsilon_r - 1)^2 g_{ijkl}E_kE_l \qquad [9]$$

Equation [9] is a series expansion of the index change in powers of the static electric field: all terms are present if the medium is noncentrosymmetric, otherwise only even-powered terms need be included. In essence, eqn [9] entails the feedback mediated by the static electric field on the optical wave.

The feedback leads to stable solutions under appropriate conditions. These are steady-state solitons if the inhomogeneous light-induced index change makes a suitable waveguide for the beam. The three main types of such self-localized beams are screening, photovoltaic, and centrosymmetric photorefractive solitons. Boundary conditions determine the induced static field to be inserted in eqn [9] and in the propagation eqn [1]. In the simplest cases we consider, the light-wave is linearly polarized and the tensorial nature of the index perturbation in eqn [9] can be neglected.

The equations encompassing most of the physics of the photorefractive effect constitute the Kukhtarev model and can be shown as:

$$\begin{aligned} &s(I + I_b + I_{dark})(N_d - N_d^i) - \gamma_r\rho N_d^i = 0 \\ &\nabla\cdot\mathbf{J} = \nabla\cdot[q\mu\rho\mathbf{E} + k_BT\mu\nabla\rho + \beta_{ph}(N_d - N_d^i)I\mathbf{c}_{opt}] = 0 \\ &\nabla\cdot\mathbf{E} + \frac{q}{\varepsilon_S}(\rho + N_A - N_d^i) = 0 \end{aligned} \qquad [10]$$

The relevant quantities in eqns [10] are: the current density **J**; the free charges (electrons) density ρ; the density of ionized (total) donors N_d^i (N_d), the density of negatively charged sites N_A, the photo-ionization cross-section s, the background density I_b (due to background laser light), the dark irradiance I_{dark} (sI_{dark} is the dark generation rate, typically $I_b \gg I_{dark}$), the recombination rate γ_r, the electron mobility μ, the low frequency dielectric constant $\varepsilon_S = \varepsilon_0\varepsilon_r$, the Boltzmann constant k_B, the versor $\mathbf{c}_{opt}$ of the optic axis, and the temperature T. In the second of eqn [10], the pertinent element β_{ph} of the photovoltaic tensor needs be taken into account only for the photovoltaic solitons described below. The above must be completed by boundary conditions for the bias V – the line integral of

E – and the geometry of the metal contacts. Typically, $N_d \approx 10^{18}\ \mathrm{cm}^{-3} \gg N_a$, $I_{dark} \approx 10^{-3}$–$10^{-6}\ \mathrm{W/cm}^2$, $I_b \approx 10\ \mathrm{W/cm}^2$, $\varepsilon_r \approx 10^3$ and solitons are observed in the visible.

Equations [10] determine the field **E** in terms of the optical intensity I in steady-state, and **E** is inserted in eqn [9] to yield the optically induced index change. A major signature of photorefractive solitons is the cumulative process they rely on: the space-charge field **E** builds up in time along with the distribution of carriers, until equilibrium is reached. Hence the phenomenon does not require high optical intensities but, conversely, a sufficiently long time (of the order of minutes or hours). Photorefractive solitons are typically observed at modest excitations, as low as a few microwatts, with typical waists of tens of microns.

Screening and Photovoltaic Solitons

When an externally applied bias lowers the refractive index of a crystal through its electro-optic response, a light beam can screen this effect, defining a self-waveguide supporting a spatial soliton. In a noncentrosymmetric photorefractive, the beam raises the conductivity in the illuminated region, locally lowering the voltage drop compared to dark regions. The induced space-charge screens the applied static field, and it may bring the refractive index back to its unbiased value. This can be seen as a voltage divider formed by a small optically controlled resistance and two large ones corresponding to the dark regions. The resulting solitary waves are called screening photorefractive solitons. For them to be stable (in time), additional illumination of intensity I_b at the same wavelength is required on the crystal in order to increase the conductivity of the (dark) regions surrounding the soliton. The charge accumulation in the beam-tails would otherwise flatten the induced index distribution and destroy the soliton.

The simplest description of screening solitons is provided by the saturating relationship between the electric field E and the light intensity I:

$$E = -\frac{V}{L}\frac{1}{1 + I/I_b} \qquad [11]$$

to be used in eqn [9] to get the index perturbation via the electro-optic effect. L is the width across which the voltage is applied. Since the quadratic term in eqn [9] (weighed by the tensor g_{ijkl}) can be neglected in noncentrosymmetric crystals:

$$\Delta n \approx -\frac{1}{2} n_0^3 r_{eff} E \qquad [12]$$

where r_{eff} is the relevant component of the electro-optics tensor; typically $n_0 \approx 2$, $r_{eff} \approx$ tens or hundreds of pm/V, and the overall index perturbation is of the order of 10^{-4}–10^{-3}, with $V/L \approx 10^3$ V/cm. This description applies to 1D solitons; for 2D solitons, sometimes referred to as 'needle' solitons, the whole set of constitutive equations must be solved numerically. Nevertheless, in most cases the saturating model, even if not rigorous, describes well solitons and related phenomena both theoretically and experimentally, especially when nonlocal effects can be ignored.

Figure 3 shows the propagation of a 2D spatial screening soliton (top) compared to the linear case (bottom). The diffractive regime is realized by artificially 'turning off' the nonlinearity acting on the applied voltage (when $V = 0$ in eqn [11] then $\Delta n = 0$, corresponding to no induced perturbation).

Another class of photorefractive spatial solitary waves is that of photovoltaic solitons. In this case the electric field in the crystal has an additional component through the photovoltaic effect, the latter being relevant in materials such as $LiNbO_3$. To sustain bright solitons the index perturbation due to the photovoltaic current, which induces the space-charge field, must be positive. This is the case for Cu:KNSBN, where 2D photovoltaic solitons were first observed.

The space-charge field, due to the photovoltaic effect, can be derived from the Kukhtarev model:

$$E = -E_{ph}\frac{I/I_b}{1 + I/I_b} \qquad [13]$$

with E_{ph} is a constant depending on the material ($E_{ph} \approx 10^3$ V/cm) and I_b is the dark-irradiance controlled by the background laser light. Typical

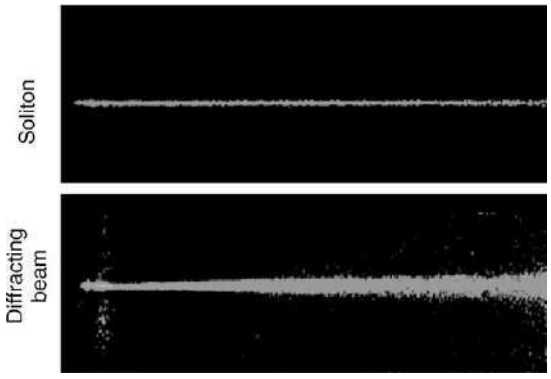

Figure 3 2D photorefractive soliton. Top view photograph of a 10 mm wide spatial soliton propagating in strontium barium niobate (top), and for comparison, the same beam diffracting when the nonlinearity if 'turned off' (bottom). (Adapted with permission from Stegeman GI, Christodoulides DN and Segev M (2000) Optical spatial solitons: historical perspectives. *IEEE Journal of Selected Topics in Quantum Electronics* 6: 1419–1426).

geometries account for the tensorial nature of the effect, and are arranged to make I_b of the same magnitude of the beam intensity I. If $E_{ph} > 0$, as in Cu:KNSBN, eqn [13] in eqn [12] shows that a positive index change of the saturating type can be induced and support bright solitons. This model is successfully used for both 1D and 2D bright photovoltaic spatial solitons.

Photorefractive Solitons in Centrosymmetric Media

Most experiments on screening solitons were performed in ferroelectric noncentrosymmetric crystals. When brought above the Curie temperature T_c, such materials exhibit a phase-transition to a centrosymmetric paraelectric state. Crystals such as Potassium Lithium Tantalate Niobate have room-temperature phase-transitions, conveniently enabling experimental investigations. In paraelectrics, the second-order electro-optic tensor r_{ijk} is zero because of symmetry, but the photorefractive effect is still available through the third-order coefficient g_{ijkl} in eqn [9]. By properly choosing the orientation, eqn [12] can then be replaced by

$$\Delta n = \frac{1}{2} n_0^3 \varepsilon_0^2 (\varepsilon_r - 1)^2 g_{eff} E^2 \qquad [14]$$

Operation near T_c greatly enhances the relevant electro-optic coefficient $g_{eff} > 0$ ($g_{eff} \approx 0.1\ \mathrm{m^4\ C^{-2}}$), making the observation of photorefractive screening solitons possible. Using eqn [11] in eqn [14] the model is saturating and works for both 1D and 2D bright solitons.

Thanks to the low powers involved and a mature crystal technology, the photorefractive effect and related soliton phenomena have contributed several impressive results in optical morphology of nonlinear light propagation, including incoherent and white light excitations.

Spatial Solitons in Nonlocal Media, Liquid Crystalline Media

A link, such as the Kerr law (eqn [4]), between the index perturbation and the optical intensity is 'local', because Δn at a given point (x, y, z) only depends on the intensity at the same location. Otherwise the medium is said to be 'nonlocal', e.g., one of those described by the general Kukhtarev model (eqn [10]) although, sometimes well approximated by eqns [11] or [13].

There are various mechanisms yielding nonlocality, mainly through spatial diffusion of the index perturbation far from the excitation. Remarkable nonlocal effects are expected when the spatial features of the beam, e.g., its waist, are significantly smaller than those of the induced polarization.

One of the most important effects of nonlocality is the stabilization against the catastrophic self-focusing in 2D, as experimentally investigated by Suter and Blasberg in 1992. Evidence of a highly nonlocal behavior has been very recently reported in liquid crystals, where the nonlinearity can be thermal or re-orientational. In the former case the index perturbation is much wider than the beam waist due to thermal diffusion associated to (absorption mediated) heat transfer. In the latter case the elastic properties of the medium produce a widening of the self-waveguide far from the beam axis. In nematic liquid crystal (NLC), the re-orientational response stems from the optically induced tilt of the elongated molecules towards the linearly polarized field versor. The light forced re-orientation increases the refractive index, thus creating the self-guiding conditions for spatial solitons. However, since in the nematic phase the molecules are linked together in a wide-scale order, the region (kernel) affected by this process is determined by the elastic properties of the medium and can extend to tens of microns.

An effective approach to describe spatial solitons in a highly nonlocal medium (where the kernel is much larger than the beam) postulates that the punctual relationship between Δn and I is replaced by a link between Δn and the beam power P. In such a way the transverse integral of the intensity is fully embedded in the index perturbation.

Assuming $n = n(P, r)$ with the beam mainly subject to the refractive perturbation near its axis, one can write:

$$n^2(r, P) = n_0^2(P) - r^2 \gamma(P) \qquad [15]$$

with $n_0^2(P) \cong n_0^2$ the on-axis value. $\gamma(P)$ depends on the specific geometry and determines the wavefront convexity at the origin. The overall profile is basically determined by the constitutive relationships of the medium. The propagation equation can be shown as

$$2ik \frac{\partial A}{\partial z} + \nabla_\perp^2 A = \frac{k^2}{n_0^2} \gamma^2(P) r^2 A \qquad [16]$$

Equations like [16] also describe the transverse profile of a beam linearly propagating in optical fibers, as well as the quantum harmonic oscillator (interpreting z as time). The soliton is its nondiffracting solution, a particular stationary case of a more general Gaussian beam with waist variable in propagation.

For intensities:

$$I(z) = \frac{P}{\pi w^2(z)} e^{-r^2/w^2(z)} \qquad [17]$$

the spot-size obeys:

$$\frac{w^2(z)}{w_0^2} = 1 + \left[\frac{1}{k_0^2 w_0^4 \gamma^2(P)} - 1 \right] \sin^2 \left[\frac{n_0 z}{\gamma(P)} \right] \qquad [18]$$

Therefore, for a Gaussian excitation, the waist oscillates along z between the values w_0 – in general determined by the launch conditions – and $w_m = w_0 \gamma(P_S)/\gamma(P)$. P_s is the 'soliton power' for which $w_m = w_0$ and $k^{02} w_0^4 \gamma^2(P_S) = 1$, hence for $P < P_s$ or $P > P_s$ the beam pulsates cyclically with z (breathes). Such a dynamic balance between diffraction and nonlinearity stems from the spectral broadening associated to an increased beam shrinking: when the spectrum widens sufficiently, diffraction starts to prevail and the beam spreads again. This process repeats itself in propagation.

The stationary solution for $P = P_s$ corresponds to a point in the existence curve of waist versus power, i.e.:

$$k_0 w_0^2 \gamma(P_S) = 1 \qquad [19]$$

Such exact balance is critical, and experimentally observed solitary waves in highly nonlocal media appear as breathers.

Figure 4 displays the intensity profile of a soliton propagating in nematic liquid crystals, as acquired by a digital camera placed on the top of the cell and retrieving the scattered light (see Figure 5). The waist oscillation in the self-trapped beam and the dependence of the period on excitation are apparent.

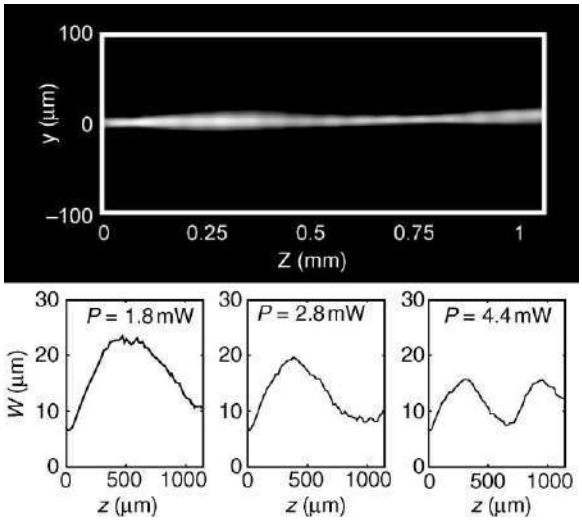

Figure 4 Nonlocal breather. (Top) top view photograph of a spatial soliton in NLC. Beam waist oscillations are apparent in propagation, as graphed (bottom panels) versus z for various excitations.

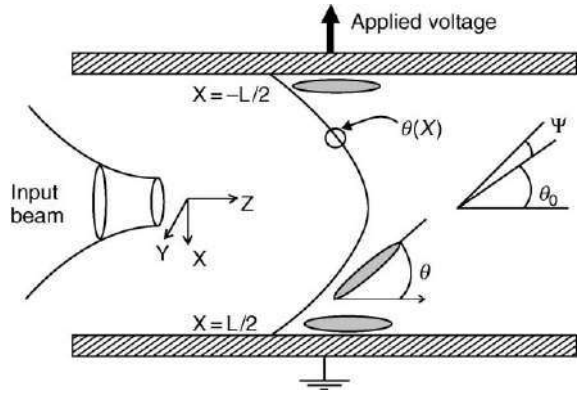

Figure 5 Sketch of a liquid crystal cell for the observation of nonlocal solitons. $\theta(x)$ is the orientation of the main axis of the ellipsoidal molecules, fixed by an external bias. The alignment is planar, with molecules parallel to the cell facets. The optical beam propagates in the center of the cell, where $\theta(x) = \theta_0$, and induces an additional re-orientation ψ. (Adapted from Conti C, Peccianti M and Assanto G (2003) Route to nonlocality and observation of accessible solitons. *Physics Review Letters* 91: 073901).

A specific highly nonlocal nonlinear system is formed by a linearly polarized beam in NLC with molecules pre-tilted (by an external electric or magnetic field) at $\theta_0 = \pi/4$ with respect to the input polarization. As sketched in Figure 5, the light wave induces an additional tilt Ψ. At the lowest order of approximation, the relevant set of equations can be shown as:

$$\begin{aligned} 2ik \frac{\partial A}{\partial z} + \nabla_\perp^2 A &= -k^2 \frac{n_a^2}{n_0^2} \Psi A \\ K \nabla_\perp^2 \Psi - G\Psi + \frac{n_a^2}{2cn_0} |A|^2 &= 0 \end{aligned} \qquad [20]$$

with K the NLC elastic constant ($K \approx 10^{-11}$ N, for NLC of the E7 type), $n_0 \approx 1.5$, n_a^2 the optical anisotropy ($n_a \approx 1$), i.e., the difference between relative permittivities along the two principal axes, G a constant determined by the specific geometry of the NLC cell and the pre-tilt angle θ_0. Since, for $K = 0$, eqn [20] represents a Kerr medium, it can be considered one of the simplest nonlocal generalizations of local models for spatial solitons. The length $R_c = \sqrt{K/G}$ quantifies the diffusion of the index perturbation away from the beam axis, such that Ψ decays according to the modified Bessel function $K_0(r/R_c)$. Typically, $R_c \approx 20\ \mu$m and solitons are observed in the visible and near-infrared.

Solitary wave profiles are found with the ansatz $A = a\exp(\mathrm{i}\beta z)$, obtaining:

$$\frac{\nabla_\perp^2 a}{2k^2} + \frac{n_\mathrm{a}^2}{2n_0^2}\Psi a = \frac{\beta}{k}a$$
$$R_\mathrm{c}^2\nabla_\perp^2\Psi + \frac{R_\mathrm{c}^2 n_\mathrm{a}^2}{2Kcn_0}a^2 = \Psi \qquad [21]$$

No analytical solutions of eqns [21] are known in 2D, hence the profiles have to be derived numerically. Conversely, it can be shown that 1D sech^2 solutions exist (see next section on parametric solitons). From eqn [21], the whole family can be parameterized by a dimensionless $\alpha = 1/(2k\beta R_\mathrm{c}^2)$. Since $(1/2k\beta)^{1/2}$ measures the soliton waist (see eqn [7]), α relates the beam size and the extension of the induced perturbation. Small α solutions access the highly nonlocal regime, as shown in Figure 6, where the ratio R between the index perturbation and the soliton waist is graphed. The insets show sample profiles of a and ψ in two limits, respectively. On the other hand, when $K = 0$, $R_\mathrm{c} = 0$, and $\alpha \to \infty$, the local limit yields an index perturbation comparable in size to the beam waist.

It is worth pointing out that a model such as eqn [16], also derived in the framework of plasma physics, predicts unconditionally stable self-trapped beams. Nonlocal solitons realize an energy minimum, thus preventing beam collapse such as in saturable (Kerr) media. The oscillations observed in experiments can be regarded as fluctuations around the lowest energy state.

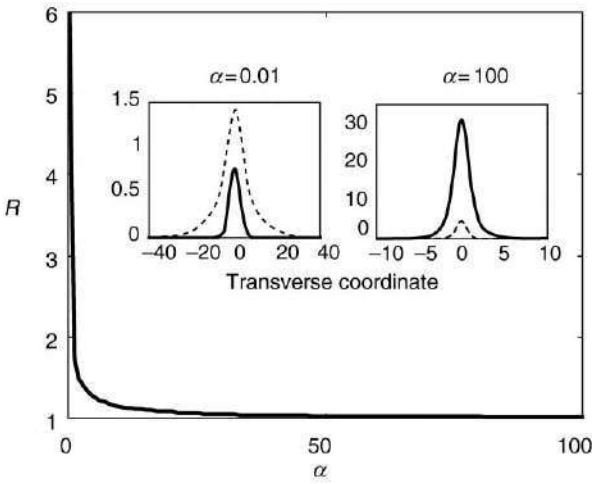

Figure 6 Spatial solitons versus nonlocality. Graph of the ratio R between the waist of the induced index perturbation in a nonlocal medium and the soliton waist versus parameter α. The insets show (in normalized units) soliton (solid line) and perturbation profiles (dashed line) for two different α, i.e., in the local ($\alpha = 100$) and highly nonlocal ($\alpha = 0.01$) regimes. (Adapted from Conti C, Peccianti M and Assanto G (2003) Route to nonlocality and observation of accessible solitons. *Physics Review Letters* 91: 073901).

Parametric Spatial Solitons in Quadratic Media

Solitons due to parametric processes, such as second-harmonic generation, belong to a different class with respect to previous sections. In fact, no refractive index modification can be physically associated to parametric effects. Parametric solitons are characterized by a power spectral density peaked around at least two carrier harmonic frequencies. It is the interaction between these components which provides the diffraction beating mechanism. Their existence was predicted in the early 1970s by Sukhorukov and Karamzin, while the first experimental evidence was reported in 1995.

The simplest process to address is second-harmonic generation (SHG): a fundamental frequency (FF) wave at ω interacts with a component at 2ω. The latter can be spontaneously generated in noncentrosymmetric crystals (e.g., $LiNbO_3$, $KNbO_3$, BBO, KTP, etc.) or externally seeded.

For plane waves, this process is ruled by:

$$\mathrm{i}\frac{\mathrm{d}A_1}{\mathrm{d}z} + \frac{\omega d_\mathrm{eff}}{c}A_1^* A_2\,\mathrm{e}^{-\mathrm{i}\Delta kz} = 0$$
$$\mathrm{i}\frac{\mathrm{d}A_2}{\mathrm{d}z} + \frac{\omega d_\mathrm{eff}}{c}A_1^2\,\mathrm{e}^{\mathrm{i}\Delta kz} = 0 \qquad [22]$$

with $\Delta k = k_1 - 2k_2 = 2\omega/c(n_1 - n_2)$ the phase mismatch, d_eff the nonlinear coefficient, n_2 and n_1 the indices at SH and FF, respectively, and

$$E_\mathrm{opt} = \mathrm{Re}\left[\sqrt{\frac{2Z_0}{n_1}}A_1\mathrm{e}^{-\mathrm{i}\omega t+\mathrm{i}k_1 z} + \sqrt{\frac{2Z_0}{n_2}}A_2\mathrm{e}^{-2\mathrm{i}\omega t+\mathrm{i}k_2 z}\right]$$

the overall electric field. d_eff relates to the quadratic susceptibility $d_\mathrm{eff}^{(2)}$ (measured in mV^{-1}) through to $d_\mathrm{eff} = d_\mathrm{eff}^{(2)}\sqrt{2Z_0}/(n_1\sqrt{n_2})$, and typically $d_\mathrm{eff} \approx 10^{-10}\,\mathrm{m}^2\,\mathrm{W}^{-1/2}$.

Equations [22] show that, in general, the harmonics with amplitudes A_1 and A_2 affect each other's phase and intensity during propagation. The solution with constant moduli, i.e.:

$$A_1 = \sqrt{I_1}\exp\left[\mathrm{i}\frac{\omega}{c}\Delta n_1 z\right]$$
$$A_2 = \sqrt{I_2}\exp\left[\mathrm{i}\frac{2\omega}{c}\Delta n_2 z\right] \qquad [23]$$

with I_1 and I_2 the respective intensities, provides a 'pure' phase variation and 'apparent' index changes Δn_1 and Δn_2 on either wave. The latter satisfy:

$$\Delta n_1 = d_{\text{eff}}^2 \frac{I_2 - I_1/2}{n_1 - n_2}$$
$$\Delta n_2 = d_{\text{eff}}^2 \frac{I_1}{2I_2} \frac{I_2 - I_1/2}{n_1 - n_2} \qquad [24]$$

with the constraint (for plane waves to exist):

$$I_1 = 2I_2 - 2\frac{n_1 - n_2}{d_{\text{eff}}}\sqrt{I_2} \qquad [25]$$

Equations [24] encompass a great deal of the physics in self-trapping via SHG. When eqn [25] is satisfied (at least approximately), the effect of SHG is exclusively on the phase. Assuming eqn [24] is valid for (wide) beams, the propagation in the presence of diffraction can be described by two coupled equations of the type [1] with relevant perturbations (eqn [24]). This shows that the sign of the term $(2I_2 - I_1)/(n_1 - n_2)$ is crucial, because focusing at both harmonics ($\Delta n_{1,2} > 0$) requires $(2I_2 - I_1)/(n_1 - n_2) > 0$.

When just a fundamental frequency (ω) is launched, solitons are generated for $\Delta k > 0$: a small 2ω component is provided via SHG and $I_1 > 2I_2$. In fact, since SHG ($\omega + \omega = 2\omega$) is followed by difference frequency generation ($2\omega - \omega = \omega$, DFG), for any finite mismatch $\Delta k > 0$, a portion of the FF travels for a while at 2ω before it is back-converted: this is commonly named 'cascading' and, as a consequence, part of the FF undergoes a positive phase shift. The converse happens in the DFG + SHG process. Finally, if an SH beam with a small FF seed is input, solitons form for $\Delta k < 0$ being $2I_2 > I_1$. From a physical point of view, these mechanisms create the phase-front distortion required to counteract diffraction. Figure 7 is an artist's sketch on the formation of an SHG soliton.

Equations [24] are approximately valid and rely on a plane-wave insight, but they point out the wealth of possibilities offered by three-wave interactions for light transverse localization via parametric processes. The full set of coupled equations describing parametric solitons from SHG differs from eqn [1]:

$$\mathrm{i}\frac{\mathrm{d}A_1}{\mathrm{d}z} + \frac{1}{2k_0 n_1}\nabla_\perp^2 A_1 + \frac{\omega d_{\text{eff}}}{c} A_1^* A_2 \mathrm{e}^{-\mathrm{i}\Delta kz} = 0$$
$$\mathrm{i}\frac{\mathrm{d}A_2}{\mathrm{d}z} + \frac{1}{4k_0 n_2}\nabla_\perp^2 A_2 + \frac{\omega d_{\text{eff}}}{c} A_1^2 \mathrm{e}^{\mathrm{i}\Delta kz} = 0 \qquad [26]$$

as it includes diffraction terms added to eqn [22].

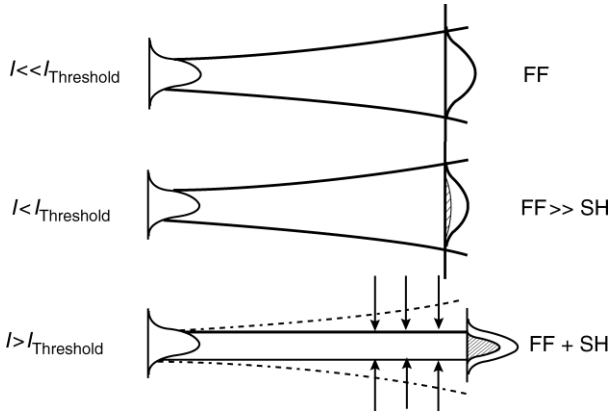

Figure 7 Parametric solitons via SHG. Intuitive description of soliton formation with an FF Gaussian input. At low powers (top) and for a small amount of generated SH (center) linear diffraction dominates. Above threshold parametric gain efficiently counteracts diffraction, originating a spatial soliton via quadratic cascading.

The self-trapped solutions are found as in eqn [23], with

$$A_1 = a_1(x, y)\exp(\mathrm{i}\beta z)$$
$$A_2 = a_2(x, y)\exp(\mathrm{i}\Delta kz)\exp(2\mathrm{i}\beta z), \quad \text{and} \qquad [27]$$
$$a_{1,2}(x, y) = \sqrt{I_{1,2}(x, y)}$$

These show that, even in the presence of diffraction, SHG can give rise to pure phase effect if the profiles satisfy:

$$\frac{1}{2k_0 n_1}\nabla_\perp^2 a_1 + \frac{\omega d_{\text{eff}}}{c} a_1 a_2 = \beta a_1$$
$$\frac{1}{8k_0 n_2}\nabla_\perp^2 a_2 + \frac{\omega d_{\text{eff}}}{2c} a_1^2 = \beta + \frac{\Delta k}{2} a_2 \qquad [28]$$

which are the space-dependent counterpart of eqn [25]. Again, β is a nonlinear contribution to the FF wavevector, and determines the corresponding correction $\beta_2 = 2\beta + \Delta k$ at SH. β is univocally determined by the power of the two harmonics in the solutions of eqn [28].

In the particular case:

$$\Delta k = \beta \frac{n_1 - 4n_2}{2n_2} \qquad [29]$$

1D parametric solitons exhibit profiles

$$a_1 = \sqrt{\frac{n_1}{n_2}} \frac{3c}{4\omega d_{\text{eff}}} \beta \operatorname{sech}\left(\sqrt{\frac{\beta k_0 n_1}{2}} x\right)^2$$
$$a_2 = \frac{3c}{4\omega d_{\text{eff}}} \beta \operatorname{sech}\left(\sqrt{\frac{\beta k_0 n_1}{2}} x\right)^2 \qquad [30]$$

and the powers per unit of transverse wavefront along y at the two components are:

$$P_1 = \int_{-\infty}^{\infty} a_1^2(x)\mathrm{d}x = \frac{n_1}{n_2}\frac{3\sqrt{2}c^2}{4\omega^2 d_{\mathrm{eff}}^2\sqrt{k_0 n_1}}\beta^{3/2}$$
$$P_2 = \int_{-\infty}^{\infty} a_2^2(x)\mathrm{d}x = \frac{3\sqrt{2}c^2}{4\omega^2 d_{\mathrm{eff}}^2\sqrt{k_0 n_1}}\beta^{3/2} \qquad [31]$$

with the total soliton power $P_1 + P_2$, directly related to β.

Equation [29] implies that, for $n_1 > 4n_2$ and therefore $\Delta k > 0$, 1D parametric solitons exhibit the sech^2 profile. Otherwise, no analytical solutions are known even in 1D, and numerical techniques need to be employed. The same applies to 2D soliton profiles in all instances.

Notably, eqns [28] are essentially identical to eqns [21] for solitons in NLC. Therefore, Figure 4 provides the features of 2D parametric solitons as well, with 'a' replaced by 'a_1', and 'Ψ' by 'a_2'. For parametric solitons $\alpha = 4n_2(1 + \Delta k/2\beta)/n_1$, showing that at phase-matching $(n_2 = n_1,\ \Delta k = 0)$ $\alpha = 4$, while large α's correspond to large Δk or small β's, e.g., low power and very wide solitons. The sech^2 profile for 1D solitons corresponds to $\alpha = 1$. 1D and 2D soliton solutions exist for any $\alpha > 0$.

The model for SHG solitons is more complicated than the previously discussed Kerr or saturable cases. The stability question can be formulated as the constrained-minimum problem of a specific functional, the Hamiltonian H. In general, for a given soliton power (the sum of the powers at the two harmonics) two branches exist for H. They correspond to two solitary solutions, differing in the ratio between the two powers. The solution with $P_2 > P_1$ corresponds to higher H, and is unstable via parametric amplification. When a harmonic field propagates in the presence of energy at a subharmonic (e.g., a small amount of FF), it tends to break up and transfer energy to it. Looking at Figure 4, this corresponds to small α's, only attainable when $\Delta k < 0$. As a result, quadratic solitons with a relevant SH content are unstable and thus, more difficult to observe.

The generation of a parametric soliton with a Gaussian input at FF, has a threshold (see Figure 7). This is because the input has to generate sufficient SH in order to produce and 'feel' the nonlinear phase-shift. A numerical simulation of eqn [26] shows the formation of a parametric spatial soliton, as displayed in Figure 8. As shown in (a), if only a small second-harmonic is generated by an FF excitation, diffraction dominates. Above threshold, as in (b), a stable self-trapped two-color beam is generated.

Two significant experimental results are reproduced in Figures 9 and 10. The photograph in Figure 10 shows a soliton propagating in periodically poled KTP (PPKTP) over five Rayleigh ranges, obtained by imaging the out-of-plane scattered light. Parametric solitons are typically observed with the FF in the near-infrared and the SH in the visible.

Figure 10 shows the output spots at FF and SH of a beam propagating in PPKTP for various FF excitations, until a soliton is generated (right end).

We close this section by observing that more general models must be considered to describe the whole physics of solitons in media with a quadratic

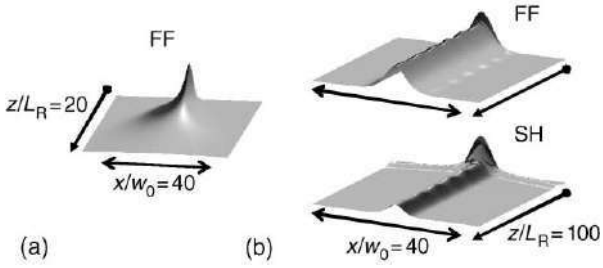

Figure 8 Parametric solitons via SHG. 1D numerical simulations using the beam propagation method. (a) An FF Gaussian excitation diffracts at low powers; (b) at high FF powers, after a transient both harmonics propagate undiffracted as a parametric spatial soliton. (Adapted with permission from Assanto G and Stegeman GI (2002) Simple physics of quadratic spatial solitons. *Optics Express* 10: 388–396).

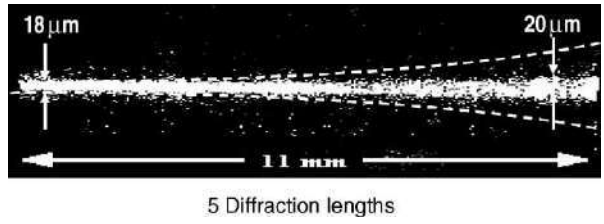

Figure 9 2D parametric solitons via SHG. Top view photograph of a soliton in PPKTP (20 ps pulses at $\lambda \approx 983$ nm and pulse energy of about 0.3 μJ). The dashed lines illustrate the diffractive behavior in the absence of nonlinear effects. (Adapted with permission from Malendevich R, Jankovic L, Polyakov S, Fuerst R, Stegeman G, Bosshard C and Gunter P (2002) Two-dimensional type I quadratic spatial solitons in KNbO3 near non-critical phase-matching. *Optics Letters* 27: 631–673.)

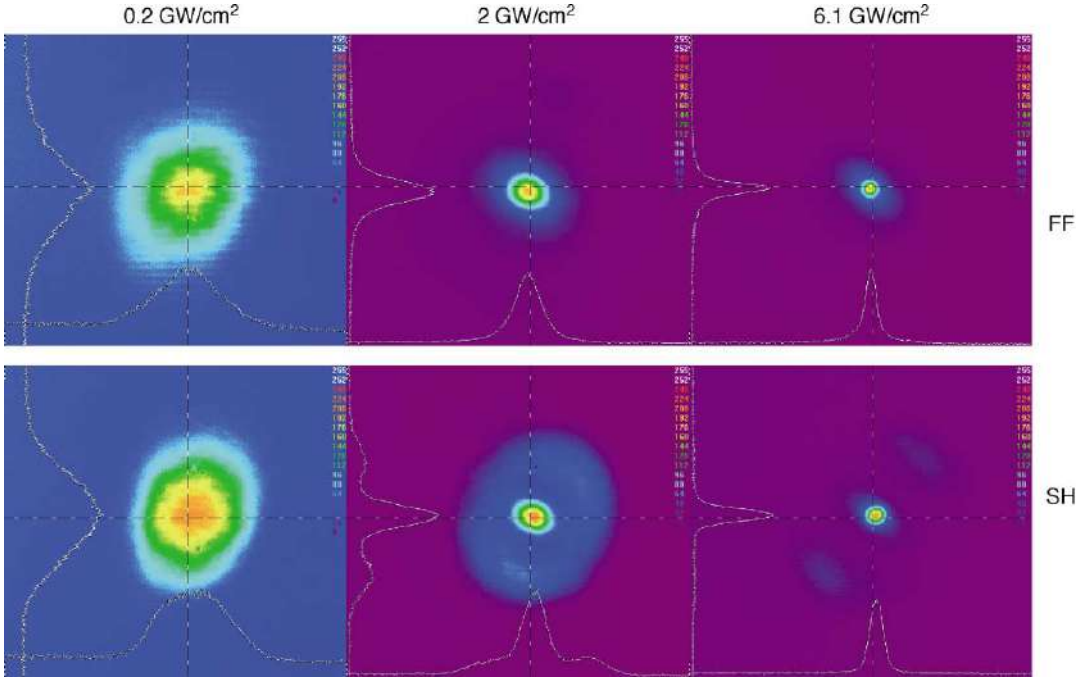

Figure 10 2D parametric solitons via SHG. Beam spots at FF (top row) and SH (bottom) for various input intensities in a PPKTP crystal. (Adapted with permission from Stegeman GI, Jankovic L, Kim H, *et al.* (2003) Generation of, and interactions between quadratic spatial solitons in non-critically-phase-matched crystals. *Journal of Nonlinear Optical Physics and Materials*. 12: 1–20).

response. We mention, for example, the subtleties associated to anisotropy (walk-off), nondegenerate three wave interactions, additional Kerr and higher-order effects.

Conclusions

Bright spatial solitons are widespread, and have been reported in several materials using various wavelengths and geometries. The brief overview presented above was intended to introduce the variety of basic mechanisms underlying their generation and propagation, but could not even touch upon the wealth of additional properties and potential applications of individual and multiple spatial solitons, their interactions and collisional properties, their stability and their control with other linear and nonlinear light waves, as well as with additional nonoptical effects (temperature, voltage, acoustic, and magnetic fields, etc.). Fascinating phenomena involve bright spatial solitons and are the subject of current research. Among them we like to mention all-optical logic and signal readdressing, beam shaping and filtering, limiting, sub-wavelength spatial compression, and digital imaging. To learn more on the subject, we encourage the interested readers to look up the vast literature covering both fundamental and applied aspects of optical spatial solitons.

List of Units and Nomenclature

Angular frequency [s^{-1}]	ω
Beam waist [m]	w
Boltzmann constant [$J\,K^{-1}$]	k_B
Current density vector [$A\,m^{-2}$]	$\mathbf{J}$
Density of ionized donors [m^{-3}]	N_d^i
Dielectric constant of vacuum [$F\,m^{-1}$]	ε_0
Director angle perturbation [rad]	ψ
Director orientation angle [rad]	θ_0
Electric field amplitude, normalized [$W^{1/2}\,m^{-1}$]	a, A
Electric field amplitude, optical frequency [Vm^{-1}]	E_{opt}
Electric field vector, low frequency [$V\,m^{-1}$]	$\mathbf{E}$
Electron charge [C]	q
Electron density [m^{-3}]	ρ
Electron mobility [$m^2\,V^{-1}\,s^{-1}$]	μ
Electro-optic second-order tensor [$m\,V^{-1}$]	r, r_{ijk}

Electro-optic third-order tensor [$m^4\ C^{-2}$]	g_{eff}, g_{ijkl}
Element of the photovoltaic tensor [$C\ m^3\ J^{-1}$]	β_{ph}
Kerr nonlinear coefficient [$m^2 W^{-1}$]	n_2
Low frequency permittivity [$F\ m^{-1}$]	ε_S
Medium elastic constant [N]	K
Medium forcing constant [$N\ m^{-2}$]	G
Non local decay distance [m]	R_c
Nonlinear length [m]	L_N
Nonlinear wavevector correction [m^{-1}]	β, β_2
Optic axis versor	$\mathbf{c}_{opt}$
Optical background intensity [Wm^{-2}]	I_b
Optical dark intensity [Wm^{-2}]	I_{dark}
Optical index anisotropy	n_a
Optical intensity [Wm^{-2}]	I
Optical saturation intensity [Wm^{-2}]	I_S
Photo-ionization cross-section [$m^2\ J^{-1}$]	s
Photovoltaic index perturbation constant [$V\ m^{-1}$]	E_{ph}
Power [W]	P
Power per wavefront unit [$W\ m^{-1}$]	P_1, P_2
Propagation coordinate [m]	z
Rayleigh length [m]	L_R
Recombination rate [$m^3\ s^{-1}$]	γ_r
Refractive index	n
Refractive index perturbation	Δn
Relative dielectric constant	ε_r
Sample width [m]	L
Second order effective susceptibility [$m\ V^{-1}$]	$d^{(2)}_{eff}$
Second order nonlinear coefficient [$m\ W^{-1/2}$]	d_{eff}
Self-focusing power [W]	P_c
Soliton power [W]	P_S
Speed of light in vacuum [ms^{-1}]	c
Static dielectric constant [$F\ m^{-1}$]	ε_s
Temperature [K]	T
Total density of negatively charged sites [m^{-3}]	N_A
Total donor density [m^{-3}]	N_d
Transverse coordinates [m]	x, y, r
Transverse wave-vector [m^{-1}]	k_x
Vacuum impedance [Ω]	Z_0
Wavelength [m]	$\lambda = 2\pi n/k = 2\pi/k_0$
Wavevector in medium [m^{-1}]	$k = n\omega/c$
Wavevector in vacuum [m^{-1}]	$k_0 = \omega/c$

See also

Babinet's Principle. **Fiber and Guided Wave Optics:** Nonlinear Optics. **Nonlinear Optics, Applications:** Self-focusing and Related Effects (Solitons and Multiphoton Absorption). **Nonlinear Optics, Basics:** Cascading; Photorefraction.

Further Reading

Assanto G and Peccianti M (2003) Spatial solitons in nematic liquid crystals. *IEEE Journal of Quantum Electronics* 39: 13–21.

Assanto G and Stegeman GI (2002) Simple physics of quadratic spatial solitons. *Optics Express* 10: 388–396.

Assanto G, Peccianti M and Conti C (2003) Nematicons: optical spatial solitons in nematic liquid crystals. *Optics and Photonics News* 14: 44–48.

Buriak AV, Di Trapani P, Skryabin DV and Trillo S (2002) Optical solitons due to quadratic nonlinearities: from basic physics to futuristic applications. *Physics Reports* 370: 63–235.

Chiao RY, Garmire E and Townes CH (1964) Self-trapping of optical beams. *Physics Review Letters* 13: 479–482.

Conti C, Peccianti M and Assanto G (2003) Route to nonlocality and observation of accessible solitons. *Physics Review Letters* 91: 073901.

Conti C, Peccianti M and Assanto G (2004) Observation of optical spatial solitons in a highly nonlocal medium. *Physics Review Letters* 92: 113902.

Crosignani B and Salamo G (2002) Photorefractive solitons. *Optics and Photonics News* 13: 38–41.

Kivshar Y and Agrawal G (2003) *Optical Solitons: From Fibers to Photonic Crystals*. New York: Academic Press.

Stegeman GI, Christodoulides DN and Segev M (2000) Optical spatial solitons: historical perspectives. *IEEE Journal of Selected Topics in Quantum Electronics* 6: 1419–1426.

Torner L and Sukhorukov AP (2002) Quadratic solitons. *Optics and Photonics News* 13: 42–47.

Various Authors (1998) In: Kajzar F and Reinisch R (eds) *Beam Shaping and Control with Nonlinear Optics*. New York: Plenum Press.

Various Authors (2001) In: Boardman AD and Sukhorukov AP (eds). *Soliton-driven Photonics*. London: Kluwer Academic.

Various Authors (2002) In: Trillo S and Torruellas WE (eds) *Spatial Solitons*. Berlin: Springer-Verlag.

Optical Fiber Solitons, Physical Origin and Properties

G Millot and P Tchofo-Dinda, Université de Bourgogne, Dijon, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In the last few decades there has been considerable effort made in understanding energy-localization phenomena in many areas of physics. Optical fibers constitute excellent media where such phenomena are routinely observed by means of relatively simple experimental setups. A well-known example of those energy-localization phenomena is the optical soliton, which corresponds to a light pulse that can propagate in optical fibers over relatively long distance without experiencing significant degradations. From a fundamental point of view, the high stability of the optical soliton results from a delicate balance between the fiber chromatic dispersion and the nonlinear refractive (or Kerr) index. The idea of using optical solitons as information bits in high-speed communication systems was first proposed in 1973, and then demonstrated experimentally in 1980. After these discoveries, interest in optical soliton transmission started to increase. In an ideal fiber, optical solitons can be modeled approximately by the nonlinear Schrödinger equation, whose solutions are well known. In this article we present the main types of soliton behavior and the fundamental physical processes that are involved in soliton stability.

The Nonlinear Schrödinger Equation

In optical fibers, there are many nonlinear phenomena that are difficult to understand by direct analysis of the nonlinear polarization (e.g., the soliton effect) but that can be readily understood by use of the nonlinear Schrödinger equation (NLSE). Let us consider an optical pulse that propagates in the z-direction, in a dielectric Kerr medium such as an optical fiber, with frequency ω_0, wavenumber k_0, and an envelope $\mathcal{A}$:

$$E = \mathcal{A}\exp[\mathrm{i}(\mathbf{k}_0\cdot\mathbf{r} - \omega_0 t)] \qquad [1]$$

We assume that the wave envelope $\mathcal{A}$ varies slowly in time and space compared with the fast variations of the carrier wave ω_0. In this situation, the Fourier spectrum of the wave is restricted to frequencies and wavenumbers close to $\omega = \omega_0$ and $k = k_0$, respectively. Then, the wavenumber k can be expanded in a Taylor series around k_0, as follows:

$$k(\omega) = k_0 + \left(\frac{\partial k}{\partial \omega}\right)_{\omega_0}(\omega - \omega_0) + \frac{1}{2}\left(\frac{\partial^2 k}{\partial \omega^2}\right)_{\omega_0}(\omega - \omega_0)^2 + \left(\frac{\partial k}{\partial |\mathcal{A}|^2}\right)_{\mathcal{A}_0=0}|\mathcal{A}|^2 + \cdots \qquad [2]$$

The nonlinear term $Q = (\partial k/\partial|\mathcal{A}|^2)_{\mathcal{A}_0=0}$ arises from the optical Kerr effect, as $k = n\omega/c = k(\omega, |\mathcal{A}|^2)$.

Setting $K = k - k_0$, $\Omega = \omega - \omega_0$, $\beta_1 = (\partial k/\partial\omega)_{\omega_0}$ and $\beta_2 = (\partial^2 k/\partial\omega^2)_{\omega_0}$, one can rewrite eqn [2] as

$$K = \beta_1\Omega + \frac{1}{2}\beta_2\Omega^2 + Q|\mathcal{A}|^2 \qquad [3]$$

where β_1 is the inverse of the group velocity and β_2 is the group velocity dispersion (GVD) coefficient. Equation [3] represents the nonlinear dispersion relation for the envelope. Using a Fourier transform method in which K is replaced by a spatial operator $-\mathrm{i}\,\partial/\partial z$, and Ω by a temporal operator $\mathrm{i}\,\partial/\partial T$, one can apply eqn [3] on the envelope $\mathcal{A}$ to obtain the following equation:

$$\mathrm{i}\frac{\partial\mathcal{A}}{\partial z} + \mathrm{i}\beta_1\frac{\partial\mathcal{A}}{\partial T} - \frac{1}{2}\beta_2\frac{\partial^2\mathcal{A}}{\partial T^2} + Q|\mathcal{A}|^2\mathcal{A} = 0 \qquad [4]$$

In practice it is more convenient to make use of the quantity $A = (\alpha A_{\mathrm{eff}})^{1/2}\mathcal{A}$, for which $|A|^2$ represents the wave power. The parameter A_{eff} is the effective core area, which takes into account the variation of the light intensity across the transverse section of the fiber. The value of A_{eff} strongly depends on the mode distribution. Using a system of coordinates that moves at the group velocity v_{g}, i.e., replacing T by $t = T - z/v_{\mathrm{g}}$, eqn [4] reduces to the basic NLSE:

$$\mathrm{i}\frac{\partial A}{\partial z} - \frac{1}{2}\beta_2\frac{\partial^2 A}{\partial t^2} + \gamma|A|^2 A = 0 \qquad [5]$$

where γ is the nonlinear coefficient defined by $\gamma = Q/(\alpha A_{\mathrm{eff}}) = \omega_0 n_2/(cA_{\mathrm{eff}})$ and expressed in $\mathrm{W}^{-1}\mathrm{m}^{-1}$. This equation is formally analogous to the Schrödinger equation in quantum mechanics but with a nonlinear potential. Furthermore, in this basic NLSE, all higher-order terms that describe effects such as the fiber losses, third-order dispersion, or stimulated Raman scattering, are neglected. The NLSE for pulse propagation in optical fibers was

first derived in 1973 by Hasegawa and Tappert. It has also found applications in other areas of physics, such as in plasma physics, fluid dynamics and electrical transmission lines. From a fundamental point of view, the basic NLSE involves the combined effects of the group velocity dispersion (GVD) and self-phase modulation (SPM), which are described by the second and third terms in eqn [5], respectively. Each of these two effects is analyzed below.

Group Velocity Dispersion

The individual effect of the GVD on pulse propagation can be obtained by setting the nonlinear coefficient to zero ($\gamma = 0$). In this case, eqn [5] reduces to

$$\mathrm{i}\frac{\partial A}{\partial z} = \frac{1}{2}\beta_2 \frac{\partial^2 A}{\partial t^2} \qquad [6]$$

In the frequency domain, eqn [6] becomes

$$\mathrm{i}\frac{\partial \tilde{A}}{\partial z} = -\frac{1}{2}\beta_2 \omega^2 \tilde{A} \qquad [7]$$

where $\tilde{A}$ represents the Fourier transform of A. The solution of eqn [7] is given by

$$\tilde{A}(z,\omega) = \tilde{A}(0,\omega)\exp(\mathrm{i}\beta_2\omega^2 z/2) \qquad [8]$$

As eqn [8] shows, the GVD changes the phase of each spectral component of the pulse by an amount that depends on the propagated distance and the frequency of the spectral component. But this effect does not modify the power spectrum: $(|\tilde{A}(z,\omega)|^2 = |\tilde{A}(0,\omega)|^2)$. This phase change can modify the shape of the pulse envelope. For example, a Gaussian pulse with a width of δ_0 at $z = 0$, $A(z = 0, t) = \exp\{(-t^2/(2\delta_0^2)\}$, will transform into

$$A(z,t) = \left[\delta_0^2/(\delta_0^2 - \mathrm{i}\beta_2 z)\right]^{1/2} \exp\left[-t^2/2(\delta_0^2 - \mathrm{i}\beta_2 z)\right] \qquad [9]$$

at a distance z. In other words, the pulse width increases from δ_0 to $\delta_0[1+(z/L_\mathrm{D})^2]^{1/2}$. The parameter $L_\mathrm{D} = \delta_0^2/|\beta_2|$ defines a dispersion length, that is, a length scale over which the dispersive effects become important. At $z = L_\mathrm{D}$ the Gaussian pulse broadens by a factor of $\sqrt{2}$. Thus, the GVD causes the spectral components of the pulse to move at different velocities, which broadens the pulse, as shown by Figure 1.

Another important effect of the GVD comes from the time dependence of the pulse phase, which one can easily obtain by rewriting eqn [9] as $A(z,t) = |A(z,t)|\exp[\mathrm{i}\phi(z,t)]$. It turns out that the GVD induces an instantaneous variation of the frequency across the pulse, called a frequency chirp. For the initially unchirped Gaussian pulse, the GVD induces a chirp given by

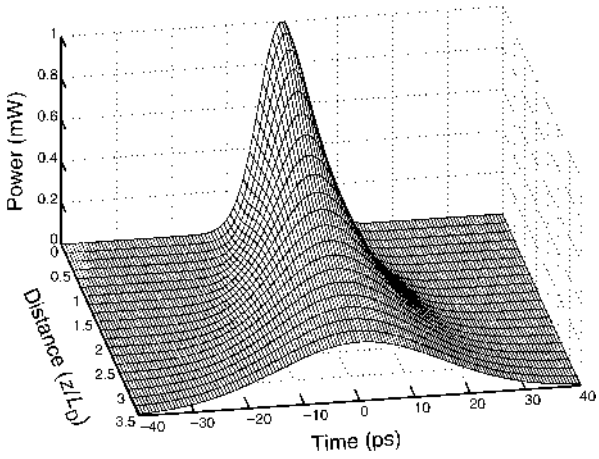

Figure 1 Temporal evolution of the power of a Gaussian pulse, as a function of the propagation distance. The simulation parameters are the following: input peak power $P_0 = 1$ mW, $\beta_2 = -20\ \mathrm{ps}^2\ \mathrm{km}^{-1}$, $\gamma = 3\times10^{-3}\ \mathrm{m}^{-1}\ \mathrm{W}^{-1}$ and $\delta_0 = 10$ ps. The dispersion length is $L_\mathrm{D} = 1.8$ km.

$$\delta\omega = -\frac{\partial\phi}{\partial t} = \frac{t\beta_2(z/L_\mathrm{D})}{|\beta_2|\delta_0^2[1+(z/L_\mathrm{D})^2]} \qquad [10]$$

As eqn [10] illustrates, the GVD-induced frequency chirp depends on the sign of β_2, and varies linearly across the pulse.

Self-Phase Modulation

Similarly to dispersion effects, nonlinear effects in optical fibers require a minimum propagation distance to develop significantly. The distance over which nonlinear effects become important is defined as the nonlinear length L_NL, given by $L_\mathrm{NL} = 1/(\gamma P_0)$, where P_0 is the peak power.

Let us consider the propagation of a light pulse in a fiber of length L, such that $L_\mathrm{NL} < L \ll L_\mathrm{D}$. In this particular case, the effects of dispersion can be neglected, and then the NLS equation becomes

$$\mathrm{i}\frac{\partial A}{\partial z} + \gamma|A|^2 A = 0 \qquad [11]$$

It results from this equation that $|A(z,t)|^2 = |A(z=0,t)|^2$. This means that the pulse shape remains unchanged during propagation. The solution of eqn [11] is given by

$$A(z,t) = A(0,t)\exp[\mathrm{i}\Phi_\mathrm{NL}(z,t)] \qquad [12]$$

where $\Phi_{NL}(z,t) = \gamma|A(0,t)|^2 z$. Thus, the Kerr effect causes the phase of the wave envelope to vary during the propagation. The nonlinear phase shift Φ_{NL} increases with the propagated distance z. The maximum phase shift Φ_{NL}^{max} occurs at the pulse center and is given by $\Phi_{NL}^{max} = \gamma|A(0,0)|^2 z = \gamma P_0 z$. One can see that the nonlinear length represents the propagation distance at which $\Phi_{NL}^{max} = 1$. Since the nonlinear phase shift Φ_{NL} depends on the initial pulse shape, it varies nonlinearly with time t. This implies that the instantaneous optical frequency differs across the pulse from its central value ω_0. The frequency chirp is given by

$$\delta\omega(t) = \omega(t) - \omega_0 = -\frac{\partial \Phi_{NL}}{\partial t} = -\frac{\partial |A(0,t)|^2}{\partial t}\gamma z \qquad [13]$$

The modification of the phase of the wave envelope by itself is called self-phase modulation. The frequency chirp increases with the propagated distance, thus implying that new frequencies are continuously generated as the pulse propagates through the fiber. This spectral broadening depends on the initial pulse shape. Figure 2 shows the variation of Φ_{NL} and the corresponding induced frequency chirp across the pulse at $z = L_{NL}$ for the case of an initial Gaussian pulse.

The SPM-induced chirp exhibits several interesting features. $\delta\omega$ is negative near the leading edge (red shift) and is positive near the trailing edge (blue shift). The chirp is linear (and increases) in the central region of the pulse. In fact, the chirp $\delta\omega$ at a given propagation distance z depends on the initial chirp. As Figure 3 shows, at a sufficiently long propagated distance, the SPM-induced spectrum exhibits several peaks due to interference effects across the pulse. Indeed as Figure 2 shows, the same frequency chirp occurs at two distinct times associated with two distinct nonlinear phase shifts.

The spectrum width can be estimated by calculating the maximum value $\delta\omega_{max}$ of the frequency chirp. For an initial unchirped Gaussian pulse one obtains

$$\delta\omega_{max} \approx \frac{0.86\gamma P_0 z}{\delta_0} \qquad [14]$$

Thus, the spectral width increases linearly with the propagated distance z and the input peak power P_0, and is inversely proportional to the initial pulse width δ_0. The SPM can be detrimental for optical communication systems. The SPM broadens the spectrum of the signal and makes it then more susceptible to GVD effects. Note that in wavelength division multiplexing (WDM) systems, the phase modulation of a given channel can come not only from the power in the considered channel itself (SPM) but also from the power in a different channel of the WDM system (CPM, i.e., cross-phase modulation). By appropriate dispersion compensation, the impact of SPM and CPM can be significantly reduced. On the other hand, the SPM, combined with dispersive effects, leads to very useful effects such as pulse compression, or the generation of optical solitons (described below).

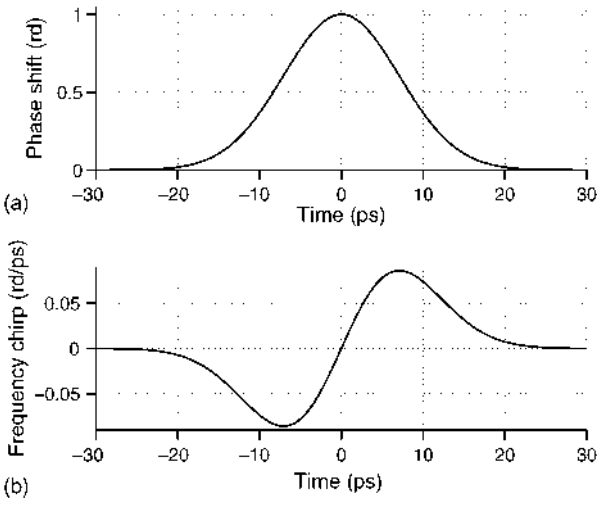

Figure 2 Temporal variation of the nonlinear phase shift (a) and the corresponding frequency chirp induced by SPM (b), for a Gaussian pulse.

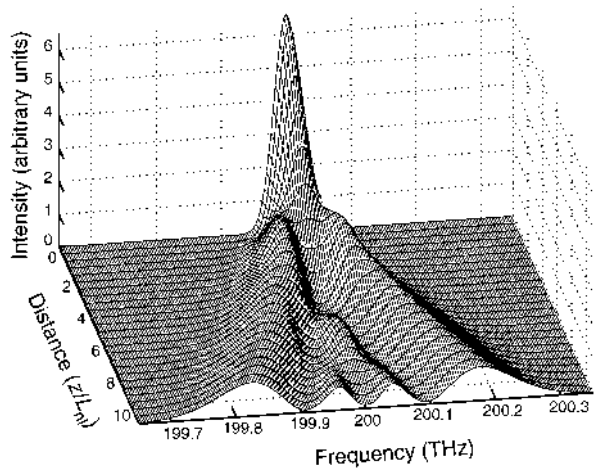

Figure 3 Evolution of the SPM-induced spectral broadening of an initially unchirped Gaussian pulse, as a function of the propagation distance. Here the normalized distance is z/L_{NL}.

Optical Solitons

Bright Solitons

Soliton propagation results fundamentally from a delicate balance between two phenomena: GVD and SPM. As we mentioned above, the GVD causes the spectral components of the pump to move at different velocities, thus leading to a temporal broadening of the pulse as shown in Figure 1 in the case of a Gaussian input pulse. Moreover, eqn [10] shows that the effect of GVD on a propagating Gaussian pulse is associated with a linear frequency chirp $\delta\omega$ and a parabolic temporal phase, proportional to β_2. On the other hand, eqn [12] shows that the SPM induces a frequency chirp across the pulse proportional to the temporal derivative of the pulse intensity. The chirp induced by SPM is then linear only in the central region of the pulse as long as the initial pulse shape is nonparabolic. As a consequence a Gaussian input pulse cannot propagate in a distortion-free manner since the GVD for anomalous dispersion ($\beta_2 < 0$) is able to completely cancel the intensity-dependent phase shift from SPM only over the central region of the pulse. On the other hand we show below that, in the anomalous dispersion regime, the exact cancellation of the nonlinear and dispersive temporal phase components (and corresponding chirps) is obtained in the case of the propagation of a hyperbolic secant pulse, which corresponds to the fundamental bright optical fiber soliton.

We consider the propagation of an unchirped hyperbolic secant pulse at a wavelength of 1.55 μm in a standard telecommunications fiber with $\beta_2 = -21 \times 10^{-3}\ \mathrm{ps}^2\ \mathrm{m}^{-1}$ and $\gamma = 1.2 \times 10^{-3}\ \mathrm{W}^{-1}\ \mathrm{m}^{-1}$. The input pulse at $z = 0$ is given by $A(0,t) = \sqrt{P_0}\ \mathrm{sech}\ (t/\delta_0)$ with $\delta_0 = 10$ ps and $P_0 = 175$ mW. With this choice of parameters the dispersion length is equal to the nonlinear length, i.e. $L_\mathrm{D} = L_\mathrm{NL} = 4.76$ km. In fact, these conditions correspond to those which would support a propagating soliton. However, let us consider first how SPM and GVD modify the incident pulse when acting independently. Figure 4 shows numerically computed results from the solutions of eqns [12] and [8] to illustrate the effects of nonlinear and dispersive evolution over a range of propagation distances. The figure shows the pulse intensity as well as the time-dependent phase and the corresponding frequency chirp. The evolution is clearly very different in the presence of only nonlinear or dispersive effects. Nonlinear SPM alone induces no change in the temporal intensity of the pulse, but the magnitudes of the SPM-induced phase and chirp increase with propagation distance. In contrast, GVD induces changes in both the intensity and the phase of the pulse as it propagates, and the temporal broadening with propagation is associated with a significant change in the functional forms of the phase and frequency chirp. In particular, whilst for large propagation distances it can be seen that GVD induces a parabolic temporal phase and corresponding linear chirp, this is clearly not the case for small propagation distances. This result strongly differs from that obtained for a propagating Gaussian pulse for which the temporal phase is always parabolic. In the context of soliton propagation, the results in Figure 4b are of particular significance, since they clearly show that at $z = L_\mathrm{D}/10$ where neither SPM nor GVD noticeably modify the pulse intensity profile, the SPM and GVD-induced phases and frequency chirps have qualitatively similar forms, yet opposite sign. The split-step Fourier method demonstrates that over a sufficiently small propagation distance (e.g., $\delta z = L_\mathrm{D}/10$) the effects of nonlinearity and dispersion can be considered to act independently on a propagating pulse. This suggests that, over the propagation segment ($0 < z < \Delta z$), the phases are capable of cancelling each other across the entire pulse profile. It is obvious that, if the effects cancel over this distance, then the propagation in the following segment will begin from the same initial condition of an unchirped pulse which was the case at $z = 0$. The nonlinear and dispersive phase components will therefore continue to cancel over successive iterations for an arbitrary length of fiber, and we thus obtain a simple physical interpretation to explain the distortion-free propagation of an optical soliton. Thus, the combined effects of GVD and SPM can prevent the occurrence of a frequency chirp across the pulse. In this situation the pulse propagates with a constant temporal profile. Such light pulses are called optical solitons.

Mathematically, the basic NLSE, eqn [5], belongs to a remarkable class of integrable nonlinear equations, which can be solved exactly for an arbitrary initial condition by means of the inverse scattering method. The solutions depend strongly on the sign of dispersion, and the relative importance of dispersive and nonlinear effects, measured by a dimensionless parameter N:

$$N^2 = \frac{L_\mathrm{D}}{L_\mathrm{NL}} = \frac{\gamma P_0 \delta_0^2}{|\beta_2|} \qquad [15]$$

In the anomalous dispersion regime ($\beta_2 < 0$) the NLS equation admits envelope-soliton solutions. The soliton order is given by the parameter N. Only the fundamental soliton, defined by $N = 1$, can propagate in the fiber with a constant profile, as schematically represented in Figure 5.

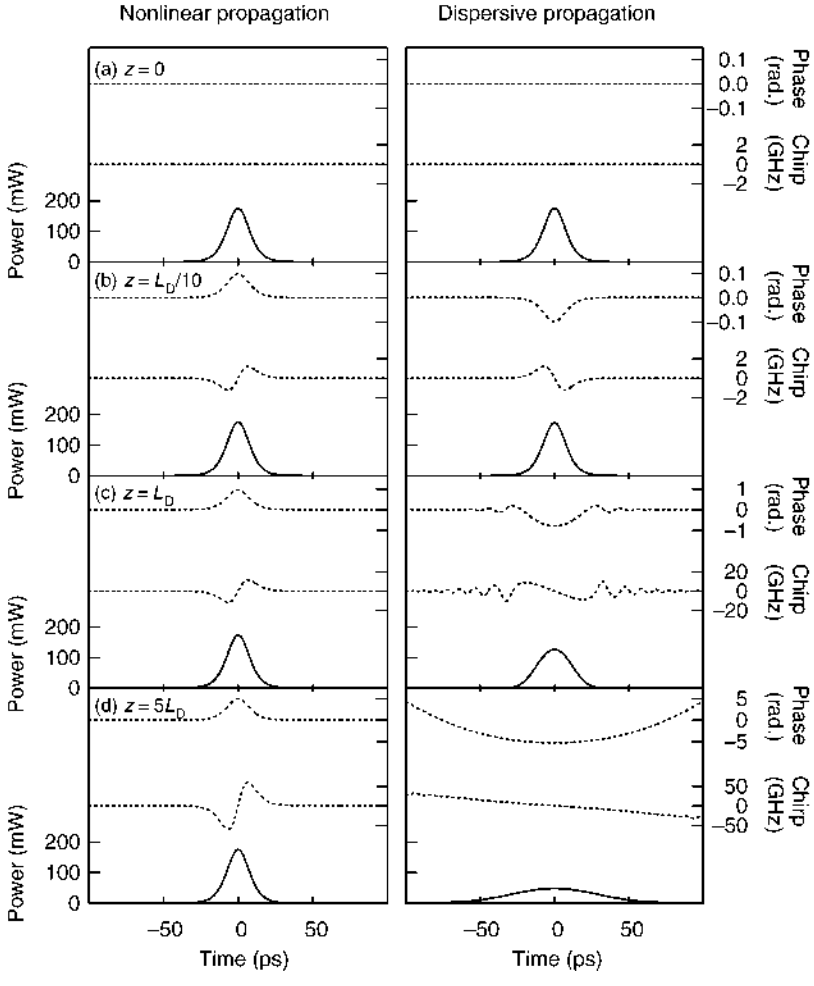

Figure 4 Numerical results showing nonlinear (left column) and dispersive (right column) propagation in an optical fiber, with parameters as described in the text, and for propagation distances as indicated. The solid lines show the pulse intensity (left axis) whilst the dashed lines show the calculated phase (right axis, top) and chirp (right axis, bottom). Constant phase offsets have been removed for clarity.

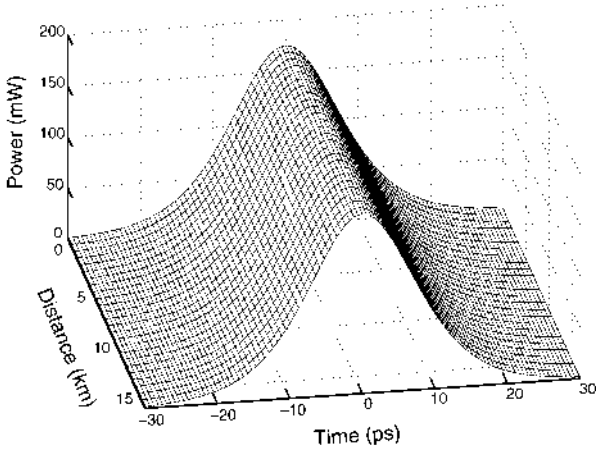

Figure 5 Temporal evolution of the power of a fundamental bright soliton, as a function of the propagation distance. The simulation parameters are the following: input peak power $P_0 = 67$ mW, $\beta_2 = -20\ \mathrm{ps^2\ km^{-1}}$, $\gamma = 3\times10^{-3}\ \mathrm{m^{-1}\ W^{-1}}$ and $\delta_0 = 10$ ps. The nonlinear length is $L_{\mathrm{NL}} = 5$ km.

For the fundamental soliton, dispersion and nonlinearity exactly balance each other in a way such that neither the pulse shape nor the pulse spectrum change during the pulse propagation. The fundamental soliton ($N = 1$) is given by

$$A(z,t) = \sqrt{P_0}\ \mathrm{sech}(t/\delta_0)\ \exp(\mathrm{i}\gamma P_0 z/2) \qquad [16]$$

where the peak power required to generate this soliton is determined by $L_{\mathrm{D}} = L_{\mathrm{NL}}$, and given by

$$P_0 = \frac{|\beta_2|}{\gamma\delta_0^2} \qquad [17]$$

Note that the phase of this bright soliton remains constant across the entire pulse. For a standard single-mode fiber (SMF), at 1.55 μm wavelength, typical values of the dispersion and nonlinear parameters are $\beta_2 = -21\ \mathrm{ps^2\ km^{-1}}$ and $\gamma = 1.2\times10^{-3}\ \mathrm{W^{-1}\ m^{-1}}$, respectively.

So, the peak power required to propagate a fundamental soliton with pulse width $\delta_0 = 10$ ps, is $P_0 = 175$ mW. This soliton is represented in Figure 5. A further reduction of P_0 can be obtained by using dispersion-shifted fibers (DSF), for which the dispersion can be significantly reduced to values of the order of 1 $\text{ps}^2\,\text{km}^{-1}$. In this situation, the power required to create a soliton reduces to $P_0 = 8.3$ mW. Thus, fundamental solitons can be generated at power levels that are easily obtainable from conventional laser diodes. The first experimental observation of soliton propagation in optical fibers was reported in 1980 by Mollenauer, Stolen and Gordon.

Note that all the higher-order solitons ($N > 1$) execute a recurrent deformation during their propagation, as illustrated in Figure 6 in the case of the second-order soliton ($N = 2$).

Dark Solitons

In the normal dispersion regime ($\beta_2 > 0$) the chirps induced by GVD and SPM reinforce each other and thus make the hyperbolic secant pulse more susceptible to temporal broadening. Thus, in the normal dispersion regime bright solitons cannot propagate. However, in this case the NLS equation, eqn [5], admits a hole-soliton solution. In other words, the intensity profile of this soliton exhibits a dip in a uniform background, as represented in Figure 7.

Mathematically, this hole-soliton, commonly called a dark soliton, takes the following form:

$$A(z,t) = \pm\sqrt{P_0}\,\tanh(t/\delta_0)\,\exp(\mathrm{i}\gamma P_0 z) \qquad [18]$$

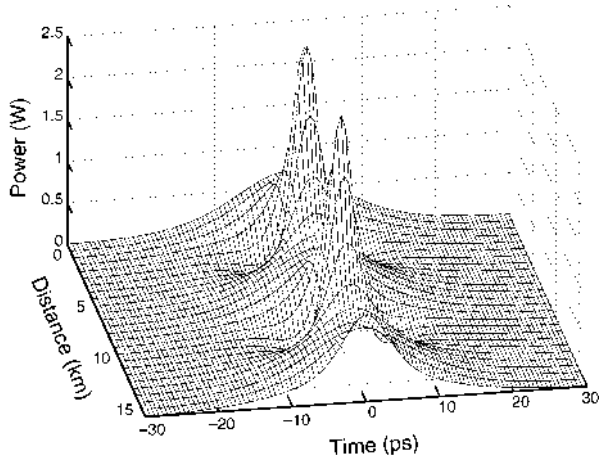

Figure 6 Temporal evolution of the power of a second-order soliton, as a function of the propagation distance. The simulation parameters are the following: input peak power $P_0 = 267$ mW. The fiber parameters are the same as in Figure 5. The nonlinear length is $L_{\text{NL}} = 1.25$ km.

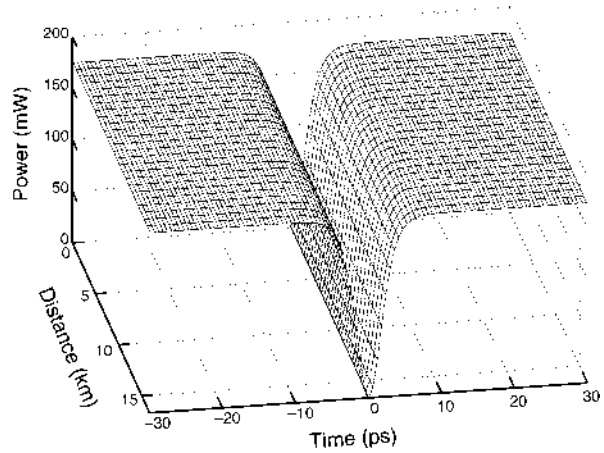

Figure 7 Temporal evolution of the power of a dark soliton, as a function of the propagation distance, for an input peak power $P_0 = 67$ mW.

As eqn [18] shows, the amplitude of the dark soliton is given by an odd function of time. Thus its phase exhibits an abrupt π jump at the dip center, thus implying that dark solitons are chirped. The phase between each jump is constant. Note that there exist other soliton solutions that are similar to the dark soliton, but for which the intensity at the dip center does not fall to zero and the phase variation is more gradual and smaller. These solutions are often called gray solitons.

A common feature of bright and dark solitons is their robustness. This property is very important for ensuring practical applications in optical communications. Moreover, optical solitons emerge unchanged from collision processes. However, dissipative perturbations such as material loss or Raman scattering can destroy such solitons.

Dispersion-Managed Solitons

As we mentioned above, the fundamental soliton is based on a delicate balance between dispersion and nonlinearity, through the relation $L_{\text{D}} = L_{\text{NL}}$. In addition, the existence of this soliton requires that only these two effects are present in the fiber. But in practice, additional linear and nonlinear effects such as fiber losses, Raman scattering, or third-order dispersion, are always present in the fiber. In particular, the fiber losses cause the soliton power to continually decrease during the propagation. The use of lumped amplifiers along the transmission line allows one to compensate for the power loss, but does not prevent the occurrence of chirp. As a consequence, the fundamental soliton does not survive over long distances (a few tens of kilometers), without in-line control. Moreover, this soliton is not stable

against phenomena such as CPM and FWM processes that occur unavoidably in WDM systems. The CPM between the pulses in a given channel and the pulses in other channels induce a frequency chirp (in the considered channel), which is converted into amplitude variations by the fiber dispersion. The FWM induces energy-exchange processes between different channels, which lead to pulse distortions.

A new transmission scheme, called *dispersion management*, was recently proposed to resolve the above-mentioned limitations. Basically, the dispersion-management technique utilizes a transmission line with a periodic dispersion map, such that each period is built up by two types of fibers, generally with different lengths and opposite group-velocity dispersion. Whereas the fundamental soliton requires total cancellation of the chirp at each propagation distance, the dispersion-managed (DM) soliton induces a small variation of the chirp, with a zero mean value over each period of dispersion management.

The DM soliton exhibits major differences with respect to the fundamental bright soliton. First, whereas the fundamental bright soliton possesses a smooth profile (as Figure 5 shows), the profile of the DM soliton exhibits an internal oscillatory structure, with several side-lobes on each side of the soliton's center. A typical representation of a DM-soliton profile is shown in Figure 8, at the free-chirp point of the dispersion map. Another outstanding difference is the following: whereas the fundamental soliton propagates with a constant profile, the DM soliton executes a periodic deformation, with a period that coincides with that of the dispersion map. The minimum value of the soliton's width occurs at the free-chirp point. From this point, the soliton's width can execute several oscillations within one dispersion map, with a peak value that can be quite large compared with the width at the free-chirp point. Figure 9 illustrates the propagation of a DM soliton within one dispersion map of a lossless DM line.

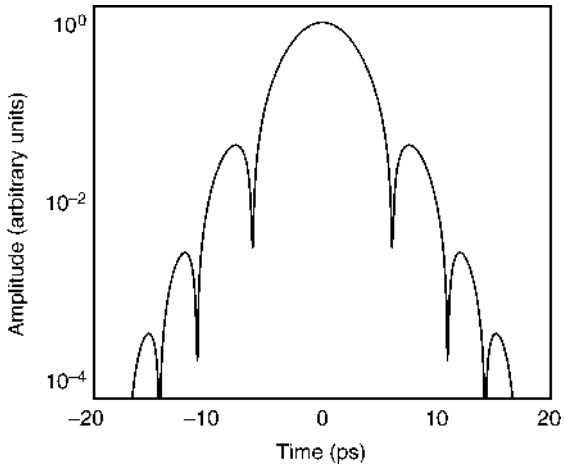

Figure 8 Plot showing the profile of a DM soliton at the free-chirp point of a typical DM line made up of 20.5 km of anomalous-dispersion fiber [$\beta_2 = -1.5 \times 10^{-3}$ ps^2 m^{-1}, $\gamma = 1.4 \times 10^{-3}$ W^{-1} m^{-1}], and 19.5 km of normal-dispersion fiber [$\beta_2 = 1.5 \times 10^{-3}$ ps^2 m^{-1}, $\gamma = 1.4 \times 10^{-3}$ W^{-1} m^{-1}].

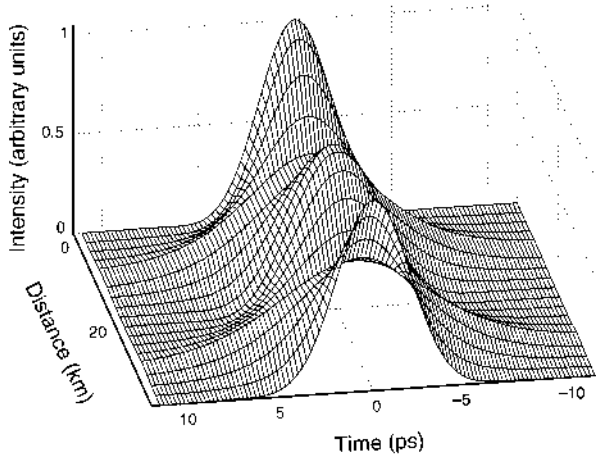

Figure 9 Plot showing the evolution of the DM soliton's profile as a function of the propagation coordinate *z*.

Recent experiments have demonstrated that in this new transmission scheme the pulse propagation becomes much more stable than fundamental solitons against linear and nonlinear effects such that SPM, FWM, CPM, or third-order dispersion. The DM transmission lines are currently subject to intense research with a view both to upgrade the capacity of existing terrestrial networks, and to design submarine fiber systems with ultrahigh capacity.

Modulational Instability

In optical fibers, the interaction between nonlinear and dispersive effects may lead to modulational instability (MI), a phenomenon in which a continuous or quasicontinuous wave undergoes a modulation of its amplitude or phase in the presence of noise or any other small perturbation. The perturbation can originate from quantum noise (spontaneous-MI) or from a frequency shifted signal wave (induced-MI). MI was observed for the first time by Tai, Hasegawa and Tomita in 1986 for a single pump wave propagating in a standard nonbirefringent fiber (scalar MI). Scalar MI occurs only when the group velocity dispersion (GVD) is negative (anomalous dispersion regime). Modulational instability in eqn [5] is examined through a linear stability analysis (LSA)

of its steady state solution

$$A = \sqrt{P_0}\,\exp[\mathrm{i}\gamma P_0 z] \qquad [19]$$

Equation [19] shows that the power remains constant through propagation along the fiber. The linear stability of the steady state solution is examined by looking into the system in the presence of small amplitude and phase perturbations *a;* that is, we consider

$$A = (\sqrt{P_0} + a)\,\exp[\mathrm{i}\gamma P_0 z] \qquad [20]$$

Then, the linearization of the NLSE with respect to a yields

$$\frac{\partial a}{\partial z} + \frac{\mathrm{i}}{2}\beta_2\frac{\partial^2 a}{\partial t^2} = \mathrm{i}\gamma P_0(a + a^*) \qquad [21]$$

Then one can assume for the perturbation a modulational ansatz with wavenumber K and frequency Ω, of the form

$$a(z,t) = u\cos(Kz - \Omega t) + \mathrm{i}v\sin(Kz - \Omega t) \qquad [22]$$

Substitution of eqn [22] into eqn [21] leads to the following dispersion relation for the perturbation

$$K^2 = \frac{\beta_2^2\Omega^2}{4}\left[\Omega^2 + \frac{4\gamma P_0}{\beta_2}\right] \qquad [23]$$

The MI phenomenon manifests itself by an exponential growth of the amplitude of the perturbation. This behavior occurs when the wavenumber K possesses a nonzero imaginary part. The dispersion relation [23] clearly shows that MI strongly depends on the sign of the dispersion. For normal dispersion ($\beta_2 > 0$), K is always real and the steady state is stable, whereas in the anomalous dispersion regime, MI appears for $\Omega < \Omega_c = \sqrt{4\gamma P_0/|\beta_2|}$. The importance of the phenomenon is measured by a power gain defined by

$$g(\Omega) = 2|\mathrm{Im}(K)| = |\beta_2|\Omega\sqrt{\Omega_c^2 - \Omega^2} \qquad [24]$$

Figure 10 shows the MI-gain spectra at three power levels with fiber parameters appropriate for standard silica fibers at $\lambda_0 = 1.319\ \mu\mathrm{m}$.

The peak value of the gain, $g_{\max} = 2\gamma P_0$, occurs at a frequency

$$\Omega_{\mathrm{opt}} = \sqrt{\frac{2\gamma P_0}{|\beta_2|}} \qquad [25]$$

called the optimal modulational frequency. In practice, when the perturbation corresponds to quantum noise, the spectral components that are amplified by the MI-process are close to Ω_{opt}. Here, as shown by Figure 11, the MI phenomenon leads to two sidebands that are situated symmetrically with respect to the pump, at frequency detunings of $\approx \pm\Omega_{\mathrm{opt}}$ (see Figure 10 for a pump peak power $P_0 = 3$ W).

In the time domain, induced-MI leads to the break-up of the quasi-cw pump wave into a train of ultrashort pulses. The repetition rate of the pulses is determined by the frequency detuning between the signal and the pump (i.e. the modulational frequency). The temporal shape of these ultrashort pulses depends not only on the powers

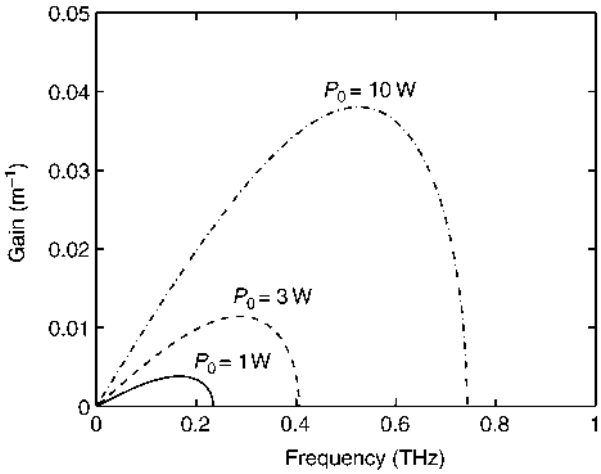

Figure 10 Spectral profiles of MI gain g for a pump peak power $P_0 = 1$ W (solid), 3 W (dashed) and 10 W (dot-dashed). The fiber parameters are $\beta_2 = -3.5\ \mathrm{ps}^2\ \mathrm{km}^{-1}$ and $\gamma = 1.9 \times 10^{-3}\ \mathrm{m}^{-1}\mathrm{W}^{-1}$.

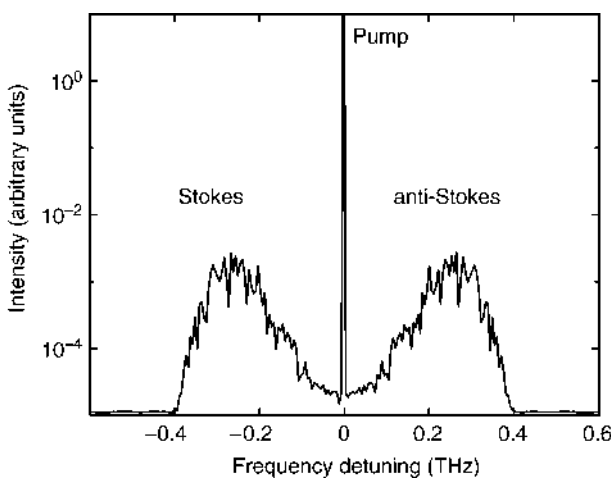

Figure 11 Spontaneous modulational instability spectrum (i.e., parametrically amplified spontaneous noise), as obtained for a peak pump power $P_0 = 3$ W and a fiber length $L = 1$ km. The fiber parameters are the same as in Figure 10.

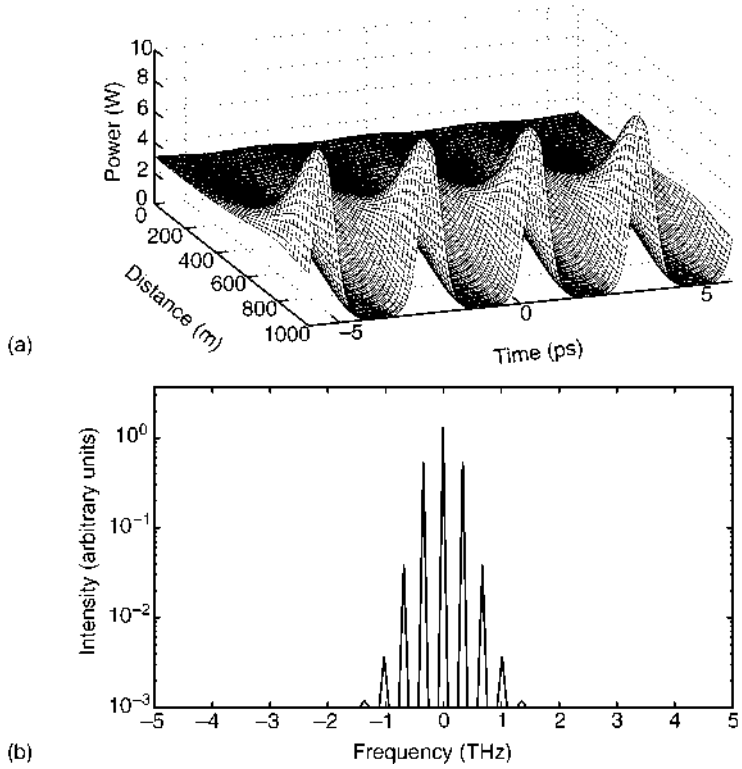

Figure 12 (a) Theoretical temporal evolution of the power for induced-MI, as a function of the propagation distance. (b) Output power spectrum in logarithmic intensity scale. The pump (signal) power is 3 W (0.5 mW), the pump-signal frequency detuning is 0.34 THz, and the fiber length is 1 km. The fiber parameters are the same as in **Figure 10**.

of the input waves but also on the modulational frequency. In particular, trains of bright and dark solitons with a terahertz repetition rate have been generated by MI in optical fibers. **Figure 12a** shows theoretically the evolution of a small initial modulation.

As **Figure 12a** shows, MI causes a small initial perturbation to evolve into a periodic pulse train, whose period is inversely proportional to the frequency spacing Δf between the pump and signal waves. Thus the technique of induced-MI can be exploited for generating high-repetition-rate ultrashort pulse trains. As shown in **Figure 12b**, such pulse train generation manifests itself in the frequency domain by the growth of a cascade of sidebands. The number of harmonics ($n\Delta f$) is determined by the temporal shape of the generated pulses and the initial conditions at the fiber input.

Cross-phase modulation (CPM) between two modes can be used to extend the instability domain to the normal dispersion regime (positive GVD). This is achieved by using orthogonally polarized pump waves. This CPM-induced MI is called vector modulational instability. The optical field evolutions are described by two coupled NLS equations. Several experiments involving two polarization modes, or two spatial modes, have been done in order to observe MI in the normal GVD regime.

Conclusion

The invention of soliton transmission, predicted in 1973 and experimentally demonstrated in 1980, is one of the major discoveries made in the area of optical telecommunication. Although from the period of invention of optical solitons up to very recent years the main line of research activities was based on the concept of the 'classical soliton', which represents an exact balance between the fiber group-velocity dispersion and its intensity-dependent refractive index, most recent experimental and theoretical developments focus on dispersion-managed solitons. It is now widely recognized that DM solitons in fiber-optic links are the most promising way both to upgrade the capacity of existing terrestrial telecommunication networks and to design submarine fiber systems. Most of the current research and development activities in this context aim to develop a non-empirical methodology of designing ultrahigh-speed DM fiber systems.

List of Units and Nomenclature

Dispersion length L_D [m]
Group velocity dispersion coefficient β_2 [$ps^2\ km^{-1} = 10^{-27}\ s^2\ m^{-1}$]
Nonlinear length L_{NL} [m]
Nonlinear parameter γ [$W^{-1}\ m^{-1}$]
Nonlinear refractive index n_2 [$m^2\ W^{-1}$]
Optical intensity I [$GW\ cm^{-2} = 10^9\ W\ cm^{-2}$]

CPM: cross-phase modulation
DM Soliton: dispersion-managed soliton
FWM: four-wave mixing
GVD: group velocity dispersion
LSA: linear stability analysis
MI: modulational instability
NLS (E): nonlinear Schrödinger (equation)
SPM: self-phase modulation
WDM: wavelength division multiplexing

See also

Fiber and Guided Wave Optics: Dispersion; Light Propagation; Nonlinear Effects (Basics). **Lasers:** Optical Fiber Lasers. **Scattering:** Scattering Theory. **Solitons:** Soliton Communication Systems.

Further Reading

Ablowitz MJ and Clarkson PA (1991) *Solitons, Evolution Equations and Inverse Scattering*. Cambridge, UK: Cambridge University Press.
Agrawal GP (1997) *Fiber-Optic Communication Systems*. New-York: Wiley-Interscience.
Agrawal GP (2001) *Nonlinear Fiber Optics*. San Diego, CA: Academic Press.
Agrawal GP (2001) *Applications of Nonlinear Fiber Optics*. San Diego, CA: Academic Press.
Hasegawa A (1989) *Solitons in Optical Fibers*. New-York: Springer Verlag.
Hasegawa A and Kodama Y (1995) *Solitons in Optical Communications*. Oxford, UK: Clarendon Press.
Remoissenet M (1999) *Waves Called Solitons, Concepts and Experiments*, 3rd edn. Berlin: Springer.
Taylor JR (ed.) (1992) *Optical Solitons – Theory and Experiment*. Cambridge, UK: Cambridge University Press.
Zakharov VE and Wabnitz S (eds) (1999) *Optical Solitons: Theoretical Challenges and Industrial Perspectives*. Les Ulis (France) and Berlin (Germany): EDP Sciences and Springer.

Soliton Communication Systems

M Karlsson and P Andrekson, Chalmers University of Technology, Gothenburg, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In 1973 Hasegawa and Tappert proposed that optical soliton pulses in fibers could exist. A soliton is a propagating wave packet that is localized in the sense that it does not spread its energy during propagation, and with the additional property that it is so stable that it can collide with other solitons and emerge unaffected with respect to energy, shape, and momentum after the collision. Solitons are based on some kind of nonlinearity in the system, and for optical fibers the weak Kerr nonlinearity (which makes the refractive index increase in proportion to the optical intensity) can counteract the pulse broadening induced by group-velocity dispersion (GVD). The two effects can form a stable balance in the form of a soliton pulse, which then propagates without changing shape.

Due to the lack of short-pulse laser sources at wavelengths above 1.3 μm, and low-loss silica fibers, it took another seven years for optical soliton pulses to be experimentally verified. In an experiment by Mollenauer, Stolen, and Gordon in 1980, soliton pulse transmission over 700 m fiber was demonstrated. During the 1980s soliton research aimed towards the use of solitons as information carriers in optical communications, and in 1990 the first data transmission experiment using solitons (2.8 Gb/s over 23 km of fiber) were reported by Iwatsuki *et al*.

During the 1990s, soliton-based communication systems have matured, an important reason being the development of the erbium-doped fiber amplifier (EDFA), which made high power levels commercially available. The most recent developments have been towards the use of solitons in alternating dispersion maps (so-called dispersion management), and such systems have reached performance levels near commercialization.

It should be stressed that fiber soliton research has been almost exclusively devoted to single-mode fibers at a wavelength of 1550 nm. In principle, solitons can also be used to counteract mode dispersion in multimode fibres, but that generally requires unrealistically high peak powers, and is therefore regarded to be of less interest. Similarly, solitons with wavelengths around 1300 nm in single-mode fibers have been

considered, but found to be of less interest due to the lack of low-noise optical amplifiers in this spectral range. Another kind of soliton, the so-called dark soliton, can also be generated in single-mode fibers. However, dark solitons have only been used as information carriers in a very limited number of experiments, and will also be omitted from this discussion.

In the present review we will discuss both theoretical and experimental aspects of soliton transmission. We will distinguish between conventional solitons, which have constant dispersion during transmission, and dispersion-managed solitons (although solitary waves would be the proper mathematical name) for which the dispersion varies periodically during transmission.

Conventional Fiber Solitons

Group-velocity dispersion means that the group velocity of the light changes with the optical frequency, and it is usually defined as the second derivative of the propagation constant β with respect to the angular frequency ω, and denoted β_0'' [ps^2 km^{-1}]. Standard single-mode fibers have a $\beta_0'' \approx -20$ [ps^2 km^{-1}] around a wavelength of 1550 nm, whereas dispersion-shifted fibers (DSFs) have β_0'' close to zero in the 1550 nm region. The dispersion is defined with respect to sign, so that the GVD is said to be normal if the group velocity decreases with frequency (i.e., $\beta_0'' > 0$), and anomalous if the group velocity increases with frequency (i.e., $\beta_0'' < 0$). A pulse propagating in a dispersive fiber will broaden with propagation distance at a rate proportional to the product of β_0'' and the spectral width of the pulse. In a communication system such a broadening is devastating, since neighboring pulses will interfere and disperse power into adjacent bit slots. The problem is increased for high data rates (as this requires short pulses with broad spectra) and/or long transmission distances.

The nonlinearity in optical fibers manifests itself as a dependence of the refractive index with power, and is known as the optical Kerr effect. For pulse propagation the Kerr effect will give rise to self-phase modulation (SPM), which means that the pulse will get a phase modulation proportional to its own intensity. Mathematically speaking, the amplitude of an optical pulse $u(0, t)$ [W$^{1/2}$] launched at $z = 0$ will evolve according to $u(z,t) = u(0,t) \exp(-\mathrm{i}\gamma z|u(0,t)|^2)$, where γ [W^{-1} km^{-1}] is the nonlinear coefficient of the fiber, z [km] is the propagation distance along the fiber, t [ps] is the time coordinate, and $|u(z, t)|^2$ [W] is the power level of the light. The value of γ varies between different fiber types due to the dependence on the core area of the fiber, but for standard single-mode fibers, in the 1550 nm region, it is approximately 2.2 km^{-1} W^{-1}.

Taking both self-phase modulation and dispersion into account, one can derive a propagation equation for light, known as the nonlinear Schrödinger (NLS) equation. Of particular importance for the solution to the NLS equation is solitons, which is a solution that propagates without dispersive broadening. The soliton solution arises as a result of a cancellation of the dispersion and the self-phase modulation effects, and it can only be formed when the dispersion is anomalous. For normal dispersion the nonlinearity and the dispersion will instead act together to increase the spectral broadening of the pulse. The soliton solution of duration t_0 and dispersion β_0'' reads

$$u_{\mathrm{sol}}(z,t) = \sqrt{\frac{1}{\gamma L_{\mathrm{D}}}}\,\mathrm{sech}\left(\frac{t}{t_0}\right)\exp\left(-\mathrm{i}\frac{z}{2L_{\mathrm{D}}}\right) \qquad [1]$$

where the dispersive length is $L_{\mathrm{D}} = t_0^2/|\beta_0''|$ [km]. The soliton peak power P_{s} [W] and energy E_{s} [J] can be expressed as $P_{\mathrm{s}} = |u_{\mathrm{sol}}(z,0)|^2 = |\beta_0''|/\gamma t_0^2$ and $E_{\mathrm{s}} = 2P_{\mathrm{s}}t_0 = 2|\beta_0''|/\gamma t_0$.

The fact that solitons are not broadened by dispersion will make them ideally suited as information carriers in a communication system. However, the solitons are affected by various perturbations that will affect the balance by dispersion and nonlinearity. Unless proper care is taken in the design of the system, those perturbations can destroy the performance of the communication system. Next, we will review the most important perturbations and design issues of soliton-based communication systems.

Solitons in the Presence of Third-Order Dispersion

The third-order dispersion (TOD), or the dispersion slope, is modeled with the derivative of the GVD, β_0''' [ps^3 km^{-1}], which is typically 0.1 ps^3 km^{-1} in standard and dispersion-shifted fibers. For solitons in dispersion-shifted fibers the effect of the dispersion slope is increasingly important the closer the solitons are to the zero dispersion wavelength.

When a soliton is placed close to the zero-dispersion wavelength, it will emit radiation in the normal dispersion regime of the spectrum, lose energy, and recoil further into the anomalous dispersion region so that the radiation stops and a stable soliton is formed. As a design criterion $\beta_0'''/(|\beta_0''|t_0) < 0.24$ have been suggested for this phenomenon to provide negligible energy loss.

Solitons in the Presence of Amplification and Loss

The most important property that has been neglected in the derivation of the NLS equation for optical

pulses is the effect of loss. The loss will decrease the power and hence the nonlinearity so that the balance sustaining the soliton is no longer maintained. As a result of the loss (which is of the order of 0.2 dBkm^{-1} in standard transmission fibers) the soliton pulse will be broadened. Partly, this can be circumvented by requiring the launched peak power of the soliton after each amplifier to be such that the path-average power between the amplifiers equals the soliton power, P_s. The peak power P_{peak} falls off as $\exp(-\alpha z)$, where $\alpha = 0.042\ \mathrm{km}^{-1}$ is the fiber loss coefficient, and the average peak power over one amplifier span is then $P_{peak}\ (1 - \exp(-\alpha L_a))/\alpha L_a = P_{peak}(G - 1)/G\ \ln(G) = P_s$, where $G = \exp(\alpha L_a)$ is the gain of each amplifier and L_a [km] is the amplifier spacing.

Nevertheless the periodic fluctuation of power and width over each amplifier length may cause the solitons to become unstable and lose power to dispersive wave radiation. To minimize this loss a useful design criterion is that the amplifier length L_a should be smaller than the dispersive length, L_D.

Sources of Timing Jitter

Another transmission obstacle is the various sources of random movement of the pulse in the bit slot, i.e., timing jitter of the pulses in the data transmission link. There are various sources of timing jitter, e.g., soliton interaction, Gordon–Haus, acoustic, and wavelength division multiplexing collision induced jitter.

Soliton interactions

Solitons, being nonlinear pulses, will interact with adjacent pulses in the pulse train. The interaction between solitons of the same polarization state and wavelength is phase sensitive so that in-phase solitons will attract each other whereas out-of-phase solitons will repel each other. To reduce this effect the solitons must be sufficiently spaced in the data stream. A typical design condition is that the pulse width of the solitons are one fifth of the bit slot, or one fifth of the inverse bit rate. A straightforward way of reducing the interaction is to have alternating amplitudes of the soliton in the pulse train, and the solitons can then be packed more densely.

Orthogonally polarized solitons interact substantially less, since it is the intensity overlap that causes the interaction, rather than the amplitude overlap as for copolarized solitons. In polarization multiplexed systems where adjacent pulse have orthogonal polarization it is usually enough with half the above separation, i.e., a pulse duration of 2.5 less than the bit separation.

Gordon–Haus jitter

Noise from the in-line amplifiers will give rise to a small jitter in the carrier frequency of each soliton, which, via the dispersion, will alter the group velocity and hence affect the arrival time of each pulse. This is known as the Gordon–Haus effect, and it has to be accounted for in long-distance systems. The variance of the timing jitter is proportional to the ratio of the power spectral density of the noise to the energy of the soliton, and also proportional to the transmission distance cubed. The fact that Gordon–Haus jitter grows cubically with distance makes it particularly important at transoceanic lengths, typically exceeding 1 Mm.

Acoustic jitter

The electrostriction nonlinearity in the fiber gives rise to a mechanical pressure proportional to the optical intensity, which in turn modifies the refractive index of the fiber. In particular, an intense optical pulse like a soliton will give rise to a pressure (acoustic) wave moving radially outwards from the fiber center. The pulses in the wake of this wave will experience a randomly changing local refractive index, and hence (just as for Gordon–Haus jitter) a randomly changing carrier frequency that transforms into a timing jitter. The variance of this jitter will grow as the transmission distance to the power of four.

WDM Considerations

Wavelength division multiplexing (WDM) is another way of increasing the bit-rate of soliton systems. That is, several frequency channels are used for solitons transmission. The problem with WDM transmission using solitons stems mainly from collisions between solitons in the different wavelength channels. In a perfectly ideal NLS equation solitons would collide elastically without changing carrier wavelength. However, the presence of losses and amplification may cause the collisions to be asymmetric if they occur around an amplifier, and the result will be a frequency displacement and a concomitant timing jitter.

Soliton Control

Soliton control is the common name for methods to control the soliton parameters such as wavelength and position, mainly in order to reduce the deleterious timing jitter effects. There are two different approaches: passive and active control.

Passive soliton control has been suggested in the form of filters that are inserted along the transmission path. This helps to keep the soliton wavelength fixed. In this way not only Gordon–Haus and acoustic jitter can be remedied, but also interaction jitter and

WDM-collision-induced jitter. A problem with this kind of filtering is that it defines a spectral region with excess gain, in which amplifier noise will grow excessively. A way around that problem is to slightly shift the center wavelength of the filters along the transmission path. In that way the solitons will follow the frequency shift, but the linear noise will not. Such sliding filter experiments have demonstrated 8×10 Gb/s WDM soliton transmission over 10 Mm.

Active control usually acts in the time domain by using phase or amplitude modulators to retime and reshape the solitons. Using this technique, 10 Gbit/s over unlimited distances has been demonstrated. However, this kind of active reshaping of the pulses suffers from the same drawbacks as conventional electronic regeneration, i.e., incompatibility with WDM, complexity, and high cost.

Dispersion-Managed Solitons

Introduction

In the last years of the 1990s, solitons have become substantially more attractive through the rapidly emerging strategy of improving the performance of soliton transmission with dispersion management (DM). While the DM strategy, which involves altering the local dispersion between a large positive and a large negative GVD such that the average GVD is small, has long been used in linear systems, it was only relatively recently appreciated that the same technique, if properly implemented, gives rise to several very striking improvements over conventional soliton transmission systems. While DM-solitons are clearly nonlinear pulses they are by no means classical solitons.

From a commercial viewpoint, however, the most important benefit with using DM-solitons is that they, in principle, can use the already installed conventional fibers (with zero dispersion at 1300 nm) allowing a much more cost-effective upgrade together with dispersion-compensating fibers (DCF) or chirped fiber gratings, which can be located together with in-line amplifiers at certain intervals in the link.

In Figure 1, an example of breathing in spectral and temporal widths of a DM-soliton can be seen, together with its dispersion map at the bottom.

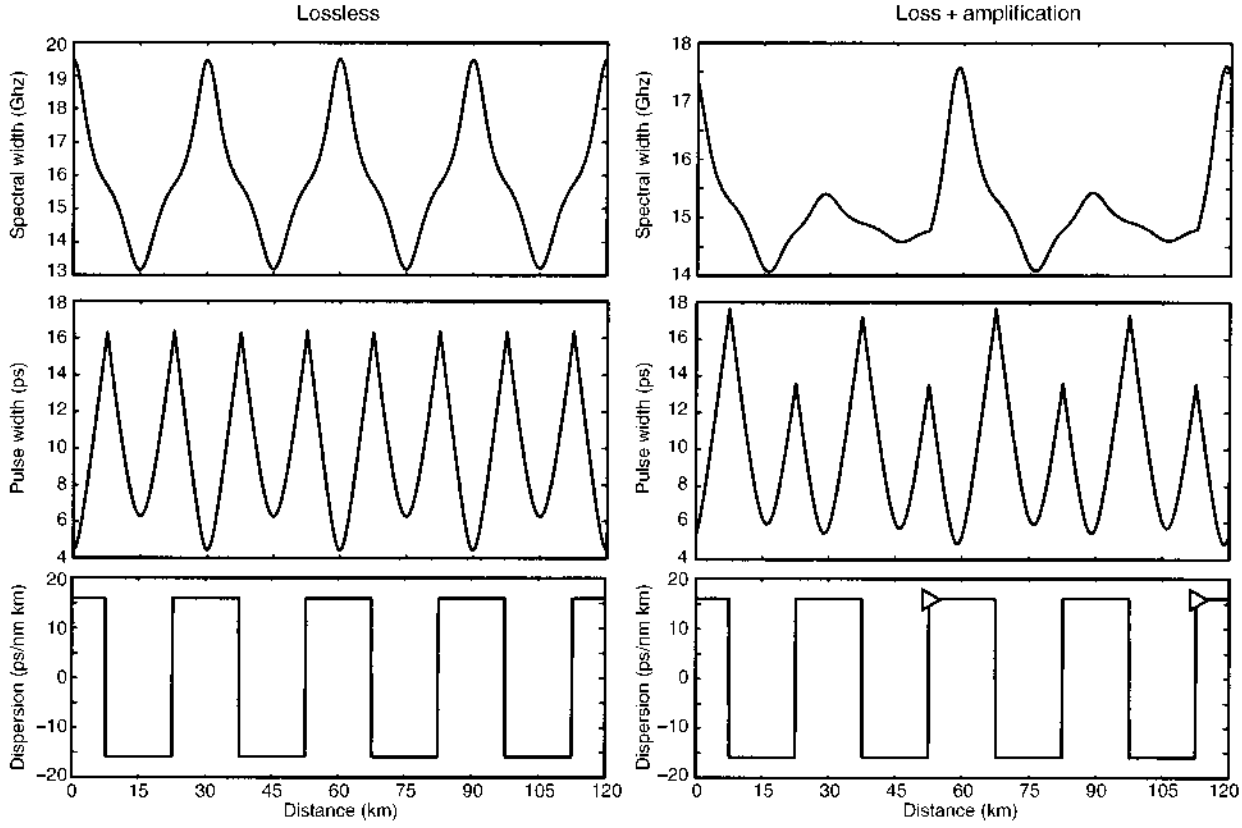

Figure 1 Plot of the spectral width (top), temporal width (middle), dispersion map (bottom) of a dispersion-managed soliton in the lossless case (left column) and the lossy case that is periodically amplified (right column). The positions of the amplifiers are denoted by the small triangles in the plot of the dispersion map.

Two cases are shown; one without fiber attenuation (left column) and one including fiber losses and periodic, lumped amplification every 60 km. One may note the strictly periodic behavior of both the spectral and temporal widths in both cases, and that the evolution of the spectral width is significantly different in the lossy case. This is because a change of the pulse spectrum requires nonlinearities and high power, and the spectral broadening is therefore highest directly after the amplifiers.

Properties

DM-solitons emerged from extensive simulation work, and it was quite surprising to many researchers that the simulations revealed such stable and strictly periodic pulses. The periodicity follows that of the dispersion variation of the system. Usually the system under study has a dispersion map, i.e., a plot of how the GVD varies along the transmission path. The map has a characteristic period (typically 10–100 km) and an average GVD which is significantly lower than the local GVD in the map.

As a result of the massive simulation work done by many groups the following properties of DM-solitons have been found:

- The pulse width, chirp, and spectral width oscillates periodically in the dispersion map. There are two points within the period at which the pulse is chirp-free, and those correspond to local minima of the pulse width. One of those points is the global minimum width, referred to as the 'shortest pulse' below.
- A central parameter that is useful for the characterization of DM-solitons is the map strength, $S = (L_1|\beta_1''| + L_2|\beta_2''|)T_{\mathrm{FWHM}}^2$, where L is the length, β'' is the dispersion, T_{FWHM} is the minimum pulse width in the full width half maximum sense, and the subscripts 1 and 2 refer to the two fibers in the dispersion map. In this example we thus assume the map to consist of two fibers, one with normal and one with anomalous dispersion. Physically the map strength is the number of dispersive lengths the pulse propagates in one period. DM-solitons have been found for map strengths ranging from $S = 0$ (which is the same as conventional solitons) to $S \approx 12$, although this upper limit is a transition regime in which the pulses radiate and perfect periodic evolution never arises.
- DM-solitons have been found for anomalous, normal, and zero average dispersion β'', defined as $\beta_0'' = (L_1\beta_1'' + L_2\beta_2'')/(L_1 + L_2)$. Normal average dispersion is only possible for map strengths above 3.9.
- The shortest pulse shape ranges from hyperbolic secant at $S = 0$ to Gaussian for higher map strengths, and this is also evident from the time-bandwidth product, which increases with S from 0.32 (sech-shape) at $S = 0$ to 0.44 (Gaussian shape) and even higher for large values of S. In addition, the shortest pulse has oscillating tails in the wings.
- The energy of a DM-soliton pulse is enhanced relative to a soliton with the same average dispersion and pulse width.
- The interaction between DM-solitons is less than that of conventional solitons, and an optimum map strength exists that minimizes the interaction.

The fact that DM solitons can work for an average net zero GVD and even for normal dispersion, was a striking and unexpected difference from conventional solitons. This can be understood by the fact that the spectrum is widest in the anomalous dispersion regime (i.e. positive values of the dispersion parameter D, cf. Figure 1), and it has been conjectured that the average dispersion, weighted by the spectral width of the DM-soliton, should always be anomalous. This is then a generalization of the fact that conventional solitons must always have a constant anomalous dispersion.

The technical improvements with DM solitons over conventional ones are numerous. The signal-to-noise ratio is improved since the DM-solitons have a larger peak power than the corresponding conventional solitons. DM-solitons have less Gordon–Haus and acoustic timing jitter, since the system average GVD is much smaller in these systems. A very important added benefit appears in WDM systems. Due to the alteration between large positive and negative GVD along the path, the jitter induced from WDM-soliton collisions is greatly reduced. This, in turn, allows for very dense WDM which will improve the spectral efficiency substantially.

An important practical consequence of using DM-solitons is that they reduce (or eliminate) the need for in-line soliton control such as synchronous modulation or sliding filters. Yet soliton control methods are still applicable and will give improvement in terms of signal-to-noise ratio for DM-solitons as well.

Quite impressive circulating loop experiments including WDM have been reported. For example, 51 densely packed WDM channels each operating at 20 Gb/s were transmitted over 1000 km with 100 km sections of standard fiber, clearly demonstrating the strength of the DM soliton technique.

Quasi-linear Transmission Intrachannel Impairments

As a rule, map strengths in the range 1–8 make best use of the unique features of DM-solitons. This means that the use of installed standard fiber becomes difficult at very high bit-rates (say beyond 30–40 Gb/s) as shorter pulses require a more rapidly (sub-km) varying dispersion map to maintain a proper S-value. If this cannot be maintained, i.e. for S values significantly above 10, one reaches the quasi-linear regime, in which the dispersion significantly dominates the transmission.

In the quasi-linear regime, there are a couple of nonlinear transmission impairments that need to be accounted for and analyzed in more detail. They are the so-called intra-channel effects; intra-channel four-wave mixing (ICFWM), and intra-channel cross-phase modulation (ICXPM), and arise due to the nonlinear interaction between two neighboring pulses. Four-wave mixing (FWM) and cross-phase modulation (XPM) are usually effects associated with WDM transmission. However, the fact that the pulses are chirped and broadened, will cause different frequency components from neighboring pulses within the same wavelength channel to overlap in time, thereby causing FWM and/or XPM within the same channel.

ICFWM arises for two neighboring pulses that, via four-wave mixing (FWM), creates new frequency components that in the time domain will give rise to a new pulse (commonly referred to as a ghost pulse), next to the two. The ghost pulse will then give rise to intersymbol interference and reduction of the eye opening. ICFWM is most prominent for large map strengths and high power.

ICXPM can be viewed as pulse-to-pulse interaction, and physically, it manifests as the frequency shift of one pulse induced by the presence of a neighboring pulse, which, by the dispersion transforms into a timing jitter. The effect can be minimized by selecting proper map strength and pre-chirp of the pulses.

Finally it should be emphasized that these intra-channel impairments will affect also non-soliton systems (e.g., non-return-to-zero (NRZ) modulated systems) if the nonlinearities are significant.

Experiments and Field Trials

Soliton Pulse Sources

When doing soliton experiments, be it conventional or DM-solitons, particular importance is placed on the properties of the pulse source, as it sets the lower limit of the system performance. A high bit-rate soliton pulse source needs to produce low-chirp, low timing-jitter pulses with proper duration (in the ps regime), repetition rate (10–40 GHz), and shape.

One possible choice is gain-switched (GS) laser diodes, possibly with an external cavity for tunability. However, they suffer the drawback of producing pulses that are strongly chirped, asymmetric and often too wide.

For laboratory experiments, fiber ring lasers (FRL) are very attractive, as they provide wavelength and pulse width tunability, besides meeting the above demands. Their drawback is that they are bulky, need active stabilization, and sometimes also temperature control to achieve long-term stability.

Finally it appears quite clear that electro-absorption modulators (EAM), which can be integrated or not with a distributed feedback (DFB) laser, are very useful and simple sources for soliton transmission. While such sources were developed for linear NRZ systems, they have now proven to be near ideal in soliton systems as well. Although EAMs are not commercially available at 40 GHz yet, they are likely to be so soon.

Special considerations need to be taken in DM-soliton systems, however, as the launch condition is different from that for conventional solitons. The pulses should have a linear chirp such that it fits seamlessly in the periodically induced chirp variation along the link. This can be achieved by incorporating a proper length of fiber (or chirped fiber grating) in the transmitter once the overall dispersion map is known.

Loop Experiments

In order to investigate really long distances (megameters) of transmission, loop experiments were developed in the early 1990s. This means that the data pulses are injected in a loop consisting of transmission fiber and amplifier, and then left to propagate a number of laps corresponding to a certain transmission distance. Acousto-optic switches are used to switch the pulse train in and out from the loop at proper time intervals. The drawback of loop experiments is that they may be poor models of reality when it comes to things like dispersion variation along the fiber, polarization mode dispersion (PMD), or various kinds of drift that may arise. In addition, a real system has more options to fine tune, e.g., amplifiers along the transmission line. However, as long as these drawbacks are recognized, loop experiments are very powerful indeed, and invaluable in laboratory evaluations of long-distance transmission.

Field Trials

In the field, there are many transmission link design restrictions and fiber properties which make the systems far from optimal. The actual fiber parameters are nonperiodic with propagation distance as the systems are straight lines rather than relatively short loops. Both loss (in particular when including many contacts and splices along the link) and dispersion might vary significantly along a fiber span. In addition, it may not be possible to tailor the dispersion map and amplifier locations to reach an optimal state. All these examples of nonidealities justify the need for field experiments.

Several soliton field experiments have been conducted over the past few years in Japan by NTT, in the US by MCI/Pirelli, as well as in Europe by ACTS projects, and the results from those experiments are discussed below. It is a good indication that solitons are indeed foreseen as very interesting candidates in commercial systems. Again, it is not very easy to compare the results as the situation in each case differs. All of the systems operated in the 1550 nm range, used optical time division demultiplexing with a 10 Gb/s electronic base rate, and the average loss/km ranged from 0.24 to 0.33 dB/km. Dispersion-shifted fiber was always used, apart from in two cases, where standard fiber was used. One field trial has used DCF for dispersion compensation, which makes this the only DM-soliton field experiment to date.

Polarization multiplexing, for which adjacent pulses have orthogonal polarization, have been used in a few experiments and this serves mainly to allow the use of relatively wide pulses, which in turn allows for larger amplifier spans. Polarization-multiplexing, however, is not as useful if the PMD of the system is high, as then the orthogonally polarized pulses would start to drift statistically in time relative to each other, thereby creating intersymbol interference and increasing the soliton interaction. In the 40 Gb/s cases and above, PMD was found to be the main capacity limiting factor.

Most of the recent experiments used a mode-locked fiber ring laser as a source, mainly because these provide excellent pulse quality as well as tunability in terms of wavelength and pulse width. Other experiments used either gain-switched lasers or electro-absorption modulators.

Future soliton field trials are expected to (1) take advantage of the now well-known strategy of improving soliton transmission performance with dispersion-management, this being very attractive for upgrading existing fiber plants, (2) implement dense WDM (in fiber lines that do not contain dispersion-shifted fibers) to boost aggregate capacity, (3) utilize different forms of in-line control, particularly at high bit rates, and (4) further address the implications of PMD and techniques to combat it. The interesting trade-off between wavelength and time division multiplexing (WDM–TDM) for optimization of overall aggregate capacity will depend on the details of the fiber line parameters.

Future Outlook

To conclude, we note that the motivation for using solitons as information carriers has changed over the years. The property of being resistant to dispersive broadening was originally the main feature, but this was considered less important when dispersion-compensating fibers became commercially available. Instead, this led to the development of the dispersion-managed soliton. The advantages of the DM-soliton over linear transmission are features like the large power which gives a high signal-to-noise ratio. On the other hand, the difference between linear and nonlinear pulses is becoming increasingly fuzzy, and perhaps the distinction should be made between return-to-zero (RZ) and NRZ modulation rather than between linear and nonlinear transmission.

It is nevertheless interesting to note that solitons are now not only considered for oceanic systems but also for shorter terrestrial systems. There are still several challenges and opportunities remaining in order to take full advantage of solitons and to reach a better understanding. Further work is also needed on WDM soliton- and very high-speed TDM soliton-systems. The use of DM-solitons is a very recently established technique and there are thus many issues to consider. These include studies of robustness to deviations of optimum conditions, e.g., improper pulse launch conditions, the impact of nonperiodic dispersion maps and of PMD (both of which are difficult to study in loop experiments), and intrachannel effects. Nevertheless, solitons have now reached a level of maturity such that commercialization seems very near.

List of Units and Nomenclature

Amplifier gain	G	[−]
Amplifier separation	L_a	[km]
Average dispersion	β_0''	[$ps^2\,km^{-1}$]
Dispersive length	L_D	[km]
Fiber loss coefficient	α	[km^{-1}]
Fiber nonlinear coefficient	γ	[$W^{-1}\,km - 1$]
Group velocity dispersion (GVD)	β_0''	[$ps^2\,km^{-1}$]

Map strength	S	[−]
Optical field amplitude	$u(z, t)$	[$W^{1/2}$]
Pulse duration	t_0	[ps]
Soliton energy	E_s	[J]
Soliton peak power	P_s	[W]
Third-order dispersion	β_0''	[$ps^3\ km^{-1}$]
Time	t	[ps]
Transmission distance	z	[km]

See also

Fiber and Guided Wave Optics: Nonlinear Effects (Basics). **Solitons:** Temporal Solitons.

Further Reading

Agarwal GP (1995) *Nonlinear Fiber Optics*, 2nd edn. San Diego: Academic Press.

Agarwal GP (2001) *Nonlinear Fiber Optics*, 3rd edn. San Diego: Academic Press.

Agarwal GP (2001) *Applications of Nonlinear Fiber Optics*. San Diego: Academic Press.

Hasegawa A and Kodama Y (1995) *Solitons in Optical Communications*. Oxford: Clarendon Press.

Iannone E, Matera F, Mecozzi A and Settembre M (1998) *Nonlinear Optical Communication Networks*. New York: Wiley.

Mollenauer LF, Gordon JP and Mamyshev PV (1997) Solitons in high bit-rate long-distance transmission. In: Kaminow IP and Koch TL (eds). *Optical Fiber Telecommunications III A*. San Diego: Academic Press.

Smith NJ and Doran NJ (1995) Evaluating the capacity of phase modular-controlled long-haul soliton transmission. *Optical Fiber Technology* 1: 218–235.

Suzuki K and Nakazawa M (1995) Recent progress in optical soliton communication. *Optical Fiber Technology* 1: 289–308.

Taylor JR (ed) (1992) *Optical Solitons–Theory and Experiment*. Cambridge: Cambridge University Press.

The final report of the ACTS-MIDAS project, Jan 1999, can be downloaded at *www.infowin.org/ACTS/RUS/PROJECTS/FINAL-REPORTS/fr-053.pdf*.

Turitsyn SK, Mezentsev VK and Shapiro EG (1998) Dispersion-managed solitons and optimization of the dispersion management. *Optical Fiber Technology* 4: 384–452.

Temporal Solitons

S Trillo, University of Ferrara, Ferrara, Italy

A Tonello, Université de Bourgogne, Dijon, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The fact that wave propagation phenomena are intrinsically dispersive, that is localized disturbances tend to spread, is part of our experience. Less well known is the fact that the spreading of wavepackets can be counteracted by the nonlinear response of the host medium. Solitons are wavepackets that exploit this principle, realizing a perfect balance of dispersion and nonlinearity that allows them to self-trap and travel undistorded, or at least to recover their shape periodically, for long (virtually unlimited) distances. More specifically, here we deal with optical temporal solitons constituted by envelopes, or pulses with relatively narrow bandwidths around a carrier optical frequency, whose temporal profile remains unchanged (either strictly or on average) upon propagation. In contrast with spatial solitons predicted earlier and such that nonlinearities are exploited to compensate for diffractive beam spreading even in the two transverse dimensions (i.e., in bulk samples, for details the reader is referred to **Solitons:** Bright Spatial Solitons), temporal soliton trapping is intrinsically a one-dimensional phenomenon typically observed in optical fibers where diffraction is compensated by the standard fiber guidance mechanism. Temporal solitons were first predicted and observed in single-mode fibers (SMFs). Their applications in digital optical communications have increased due to their capability to behave as nearly ideal bits. Meanwhile, however, the concept of temporal soliton has been considerably generalized by investigating new settings and effects such as polarization, parametric amplification (*see* **Nonlinear Optics, Basics:** $\chi^{(2)}$-Harmonic Generation), gratings (*see* **Fiber Gratings**), stimulated scattering (*see* **Scattering:** Scattering Phenomena in Optical Fibers), and fiber lasers (*see* **Lasers:** Optical Fiber Lasers), etc. Here we briefly review the essential features of these different families of temporal solitons. In particular, in the section on Fiber Temporal Solitons below, we introduce the basic model for fiber solitons, whereas we discuss their application to optical communications in the section on Solitons in Optical Communications below. In the sections on Multi-component Temporal Solitons, and Temporal Solitons in Bragg Gratings below, we introduce significant generalizations of the concept involving

multicomponent and slow-light grating solitons, respectively. Finally, in the section on Solitons on Dissipative nonlinear optics we discuss issues related to solitons in strongly dissipative systems. Note that, throughout the paper, the word soliton is used in the aforementioned general sense, which does not necessarily require the underlying model to be integrable (*see* the section on Fiber Temporal Solitons below,), as implied by the more restrictive and rigorous mathematical meaning of soliton. We further focus our discussion to bell-shaped solitons (for dark solitons, holes on a continuous-wave background, *see* **Solitons:** Optical Fiber Solitons, Physical Origin and Properties), which can be excited by means of standard laser pulses.

Fiber Temporal Solitons

An optical pulse is described by the complex electric-field envelope $E(z,t)$ that modulates the carrier frequency ω_0. Its evolution along the z axis of a silica SMF is described with good approximation by the following universal model for dispersive and weakly nonlinear systems, known as the nonlinear Schrödinger equation (NLSE),

$$\mathrm{i}\frac{\partial E}{\partial z} - \frac{k''}{2}\frac{\partial^2 E}{\partial t^2} + \gamma|E|^2 E = 0 \qquad [1]$$

In eqn [1] the second term is responsible for dispersive spreading of pulses ruled by the group-velocity dispersion

$$k' = \left.\frac{\mathrm{d}^2 k}{\mathrm{d}\omega}\right|_{\omega_0}$$

(GVD, known also as chromatic or second-order dispersion), while the first-order dispersion

$$k' = \left.\frac{\mathrm{d}k}{\mathrm{d}\omega}\right|_{\omega_0}$$

has been removed in eqn [1] by introducing the retarded time $t = T - k'z$ measured in the pulse frame traveling at the natural group-velocity $V \equiv k'^{-1}$ of light, T being the physical time in the laboratory frame. The cubic term in eqn [1] is responsible for self-phase modulation (SPM) that stems from the instantaneous optical Kerr effect characterized by the effective coefficient $\gamma = k_0 n_{2I}/A_{\mathrm{eff}} = (\omega_0/cA_{\mathrm{eff}})n_{2I}$, where n_{2I} is the nonlinear index of silica, i.e., the nonlinear refractive index change divided by the intensity which has produced it. The real-world total electric field can be recovered as $\mathcal{E}(z,r,t) = \mathrm{Re}[f(r)E(z,t)\exp(\mathrm{i}k_0 z - \mathrm{i}\omega_0 T)]$, where the transverse modal profile $f(r)$ has been averaged out in eqn [1] to yield an effective fiber area A_{eff}.

The mechanism behind soliton formation can be understood as due to the instantaneous frequency changes along the pulse caused by the Kerr SPM, compensating those due to the group velocity dispersion (GVD) in the anomalous case ($k'' < 0$, at carrier wavelengths longer than 1.3 μm in standard SMFs) where the GVD induces a blue-shift on the pulse leading edge and a red-shift on the pulse trailing edge (in the absence of such compensation the pulse broadens because anomalous GVD means that blue-shifted frequencies travel faster than red-shifted frequencies). Such mechanism of compensation becomes ideal in the case of a soliton waveform. This concept finds its mathematical support in the fact that eqn [1] is a conservative (Hamiltonian) model with the remarkable property to possess infinite conserved quantities, a rather exceptional feature which makes it exactly integrable. Among the solutions, the following fundamental (so-called $N = 1$, see below) bright temporal soliton exists whenever $k'' < 0$:

$$E(z,t) = \frac{1}{\sqrt{\gamma L_{\mathrm{d}}}}\,\mathrm{sech}[\eta(t - \mathrm{d}k'z - t_0)] \times \exp\left[-\mathrm{i}\delta\omega t + \frac{\mathrm{i}}{2}\left(\frac{1}{L_{\mathrm{d}}} + k''\delta\omega^2\right)z + \mathrm{i}\phi_0\right] \qquad [2]$$

where t_0 and ϕ_0 are arbitrary initial position and phase, and $L_{\mathrm{d}} = (k''|\eta^2)^{-1}$ is the characteristic distance after which the GVD causes a significant pulse broadening, the so-called dispersion length. Temporal solitons [2] are characterized by two independent parameters, namely the duration $1/\eta$ and the frequency detuning $\delta\omega$, and possess several remarkable features:

(i) they travel with invariant shape by realizing the condition $N \equiv L_{\mathrm{d}}/L_{\mathrm{nl}} = 1$, i.e., exact balance between the two length scales associated with eqn [1], namely L_{d} and the nonlinear length $L_{\mathrm{nl}} = (\gamma P_1)^{-1}$ associated with the peak power $P_1 = (\gamma L_{\mathrm{d}})^{-1} = |k''|\eta^2/\gamma$ needed for the input pulse [2] to be a soliton;

(ii) in any given fiber with fixed GVD k'' and nonlinearity γ, a larger soliton power P_1 is required for a shorter duration, i.e., narrower solitons are taller;

(iii) solitons experience a phase retardation as a whole (particle), which increases with P_1, i.e., taller solitons have larger shifts;

(iv) as a consequence of Galileian invariance, a frequency shift $\delta\omega$ from ω_0 induces the soliton velocity to change from k'^{-1} to $(k' + \mathrm{d}k')^{-1}$, with $\mathrm{d}k' = \mathrm{d}k''\delta\omega$;

(v) unlike other waves in nonlinear media, collisions between solitons with different velocities are strictly elastic, as shown in the example of Figure 1a; and

(vi) solitons are extremely robust: for instance they can be excited with input peak powers differing from P_1, whereas a weak attenuation α ($\alpha L_\mathrm{d} \ll 1$) induces the soliton to reshape adiabatically, i.e., the peak power decreases, yet the pulse remains locally a soliton by adjusting its width, as displayed in Figure 1b.

More general (and cumbersome) solutions of eqn [1] describe periodic soliton evolutions, such as the interaction of in-phase adjacent input pulses at the same frequency (see Figure 1c), or a single input pulse with sech-shape but peak powers corresponding to higher integer values of N, so-called breathers because of their ability to recover periodically the initial shape after 'breathing' (see Figure 1d for $N = 3$) over the distance $\pi L_\mathrm{d}/2$.

Solitons [2] describe stable propagation-invariant pulses under ideal conditions (lossless case, relatively long pulses and short SMFs). In practical settings, the importance of solitons stems from their robustness against real-world perturbing phenomena described by terms that in turn break the integrability of the underlying model [1]. Situations of this kind arise, for instance, when one accounts for stimulated Raman scattering whose effect is to down-shift continuously the central frequency of ultra-short solitons, or in optical communication systems treated below.

Solitons in Optical Communications

Solitons [2] represent ideal bits for RZ (return-to-zero) optical data sequences. However, fibers have unavoidable losses of about 0.2 dB/km at the usual operating wavelength $\lambda_0 = 1550$ nm. To ensure long-haul connections, the optical signal must be periodically amplified, commonly using erbium-doped fiber amplifiers (EDFAs). The pulse peak power hence

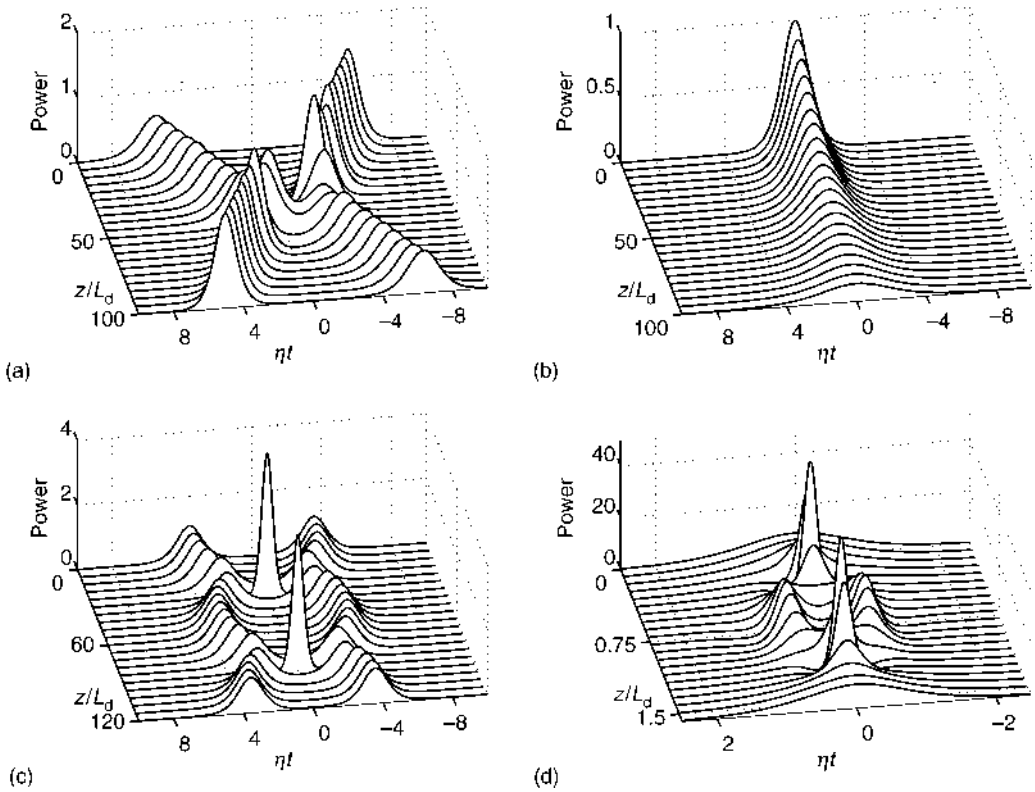

Figure 1 Fundamental features of fiber (NLSE) solitons: (a) elastic collision of two $N = 1$ solitons with different values of parameter pairs (frequency–velocity and amplitude–width); (b) adiabatic soliton damping in the presence of losses; (c) two-soliton bound state (in-phase interacting solitons of equal amplitudes); (d) evolution of a $N = 3$ soliton breather along one period of its evolution. Here the power $|E|^2$ is scaled in units of fundamental soliton power P_1, while distance z and time t are reported in units of L_d and initial pulse-width $1/\eta$, respectively.

undergoes both large exponential decay, because of distributed losses and abrupt lumped amplification under the action of EDFAs. Therefore, the stable local balance between GVD and nonlinearity cannot be rigorously sustained in the fiber. If both losses and amplifiers are accounted for, the mathematical model for pulse envelope propagation along M spans becomes:

$$\mathrm{i}\frac{\partial E}{\partial z} - \frac{k''}{2}\frac{\partial^2 E}{\partial t^2} + \chi|E|^2 E = -\mathrm{i}\alpha E + \mathrm{i}[\exp(\alpha Z_\mathrm{A}) - 1] \times \sum_{n=1}^{M} \delta(z - nZ_\mathrm{A})E(nZ_\mathrm{A}, t) \quad [3]$$

where the RHS contains the perturbing terms arising form distributed losses (α) and lumped amplifications (described by the comb with spacing Z_A).

The analysis of nonintegrable systems, such as eqn [3], can be carried out by appropriate transformations that allow the recovery of a reduced integrable model. The key result is that, under the limitations set by an appropriate length scale, the solitons of the reduced system can be close to those of the original problem. For temporal solitons, such transformation was first called 'the guiding center soliton', envisaging a similarity with the guiding center motion of a charged particle in a magnetic field. The main outcome is that eqn [3] can be reduced to the standard NLSE [1], clearly showing that soliton-like propagation still occurs in the presence of losses (even large, $\alpha L_\mathrm{d} \gg 1$) and rapid periodic amplification. The appropriate scale for these systems is the amplifier spacing Z_A, that must be much shorter than L_d. To reproduce an average behavior similar to the lossless case, it is sufficient to multiply the input condition for an enhancement factor that counteracts the effect of fiber losses. Therefore soliton-like solutions do exist even in fiber systems with loss management, though on average (over several fiber spans) rather than locally.

Another issue that makes solitons suitable as optical bits, lies in their intrinsic aptness to be controlled either through passive or through active systems. Solitons tend to follow their relative center of mass in the temporal or spectral domain. For instance, a powerful technique to overcome one of the major problems, i.e., the growth of noise due to amplified spontaneous emission, may be obtained by deploying periodically along the line optical passband filters. Mollenauer and co-workers have demonstrated that, by sliding the filters progressively (i.e., changing slightly the central peak from filter to filter), one can guide a shape-invariant soliton which adapts its carrier frequency while effectively suppressing the noise. Similar control techniques involve amplitude or phase modulators and a consistent amount of work has focused on the problem of all optical soliton regeneration.

To exploit the idea of soliton in optical communications, however, we should consider the fact that ultra-short optical pulses, suitable for high transmission rates, require extremely low fiber GVD to accomplish the stable balance between dispersion and nonlinearity for a fixed pulse-width and available power. For instance, in order to transmit at 40 Gbit/s, we should achieve a fixed GVD coefficient $D = -(2\pi c/\lambda_0^2)k'' \simeq 0.25$ ps/(km nm) for the whole link.

On the other hand, in many optical systems the fiber nonlinearity is not exploited at all but rather considered as a perturbing effect and a source of penalty that limits the maximum transmission distance. In such an approach the desired cumulated GVD should be as close as possible to zero to avoid errors coming from inter-symbol interference caused by dispersive broadening. Dispersion-shifted fibers may ensure low GVD only in a narrow spectral region, thus limiting the transmission bandwidth. Conversely GVD can be reduced to zero on average, simply combining fiber spans with opposite GVD. The periodic alternance of fiber types may also carry several benefits, e.g., limiting the impact of resonant nonlinear interactions such as four-wave mixing (*see* **Fiber and Guided Wave Optics:** Nonlinear Effects (Basics) for its definition). The different architectures of dispersion management proposed with the twofold scope of limiting nonlinear effects and reducing the cumulated GVD, have immediately stimulated the investigation of their impact on soliton propagation. In the early schemes, the attention was focused on the so-called adiabatic dispersion management, where the GVD was changed proportionally to the soliton power to improve soliton system performances. In contrast with previous works, nonadiabatic map profiles were introduced in the experiments by Suzuki and co-workers to compensate the cumulated dispersion by locating periodically dispersion compensating fibers (DCFs). In net contrast, Doran and co-workers have suggested the use of a combination of fibers with alternate normal and anomalous GVD (i.e., a periodic stepwise variation of GVD), predicting the existence of solitons with a peak power larger than the equivalent classical soliton power, i.e., dispersion-managed solitons (DMSs) The periodic stepwise GVD map is not only conceptually different from earlier designs (with exponentially decaying GVD), but has opened the field to real application due to its simplicity of implementation. The innovative discovery that special pulse waveforms may

propagate in periodically dispersion managed fiber links has triggered the interest of several groups, leading to a complete characterization of the problem, both theoretically and experimentally.

Similar to the case of guiding center solitons, a DMS usually shows a solitonic behavior. However, the DMS has a Gaussian shape surrounded by oscillating tails (see Figure 8 in **Solitons:** Optical Fiber Solitons, Physical Origin and Properties), thus being more suitable, due to its faster decaying tails, to be closely packed in a bit sequence in comparison with a sech-soliton (eqn [2]) of the same FWHM. We show an example of DMS transmission line in the top frame of Figure 2, where we sketched the variation of the GVD of a standard SMF periodically balanced by shorter segments of DCFs with opposite GVD. We report, in the middle frame, the DMS pulsewidth and chirp and in the bottom frame the pulse peak power. As shown, the macroscopic soliton parameters are characterized by large fluctuations: the local minima of pulsewidth correspond to chirp-free points and at such points one observes either a minimum or a maximum of the pulse spectral bandwidth. Nonetheless, all these pulse parameters have periodical evolutions that are fingerprints of an invariant and soliton-like pulse when observed at integer multiples of Z_A. The exact periodicity preserves the pulse waveform and hence the carried information. In DMS transmission we assist at the threefold compensation of energy, GVD and nonlinearity. Similarly to the guiding center solitons, the threefold compensation has an average, rather than pointwise, connotation. However, unlike guiding center solitons, for DMSs the definition of average should take into account the different relevance of pulse peak power in the map. Consequently the nonlinearity is compensated in a weighted manner, and it is not surprising in this context (though astonishing if placed back in the context of the NLSE [1] which requires anomalous GVD), that DMSs have been observed even when the uniform average GVD is weakly normal.

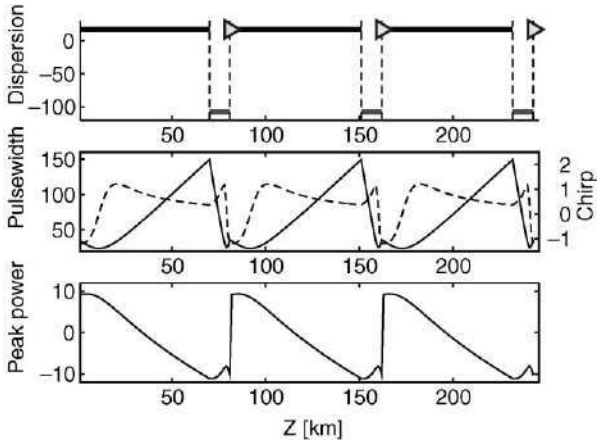

Figure 2 Evolution of soliton parameters along three periods of a fiber link with lumped amplifiers and dispersion management. Top frame: GVD map D(z) [ps/(nm km)] and EDFAs (triangles) location; Middle frame: soliton chirp (in units of 1000 ps^{-2}, dashed line), and pulse-width (in ps, solid line) vs. distance. Bottom frame: DMS peak power (dBm) vs. distance.

The net difference between the local and average GVD allows simultaneous control of different detrimental phenomena: for instance four-wave mixing is reduced by the large local GVD, whereas a small average GVD limits the impact of timing jitter due to amplified spontaneous emission (Gordon–Haus effect). As a result, DMSs perform better than conventional solitons.

Figure 3 shows a typical high-speed (20 Gb/s) single-channel experiment, and the achieved performance in terms of quality factor Q versus propagation distance, with the output after 8100 km. We recall that the Q factor measures the bit-error-rate (an error rate lower than 10^{-9} corresponds to a Q higher than $\sim$ 16 dB). DMSs also allow the control of collisions between solitons that pertain to different channels (we recall that solitons at different carrier frequencies go at different velocities, and hence collide), thus enhancing performances of standard or dense WDM systems.

Multi-Component Temporal Solitons

So far we have discussed temporal solitons described by a single envelope. However, the same mechanism of mutual balance between nonlinearity and dispersion can also be effective for multiple pulses, leading to trap simultaneously multicomponent (or so-called vector) solitons. Solitons of this type are, for instance, those due to the interaction of polarization modes in isotropic or birefringent Kerr media, higher-order modes in large-core (multimode) fibers, or mixing of pulses at different carrier frequencies that interact parametrically (i.e., without net exchange of energy with the medium). The latter case entails a broad class of solitons since virtually any parametric four-photon mixing process in cubic media (e.g., nondegenerate or partially degenerate four-wave mixing, third harmonic generation, …), or three-photon mixing process in quadratic media (sum- and difference-frequency generation, second-harmonic generation (SHG), optical rectification, etc.) is allowed to sustain them. This opens up the field to an entirely new class of materials, namely noncentrosymmetric (quadratic) materials which usually have stronger nonlinear responses if compared with silica fibers. SHG can be taken as a paradigm to

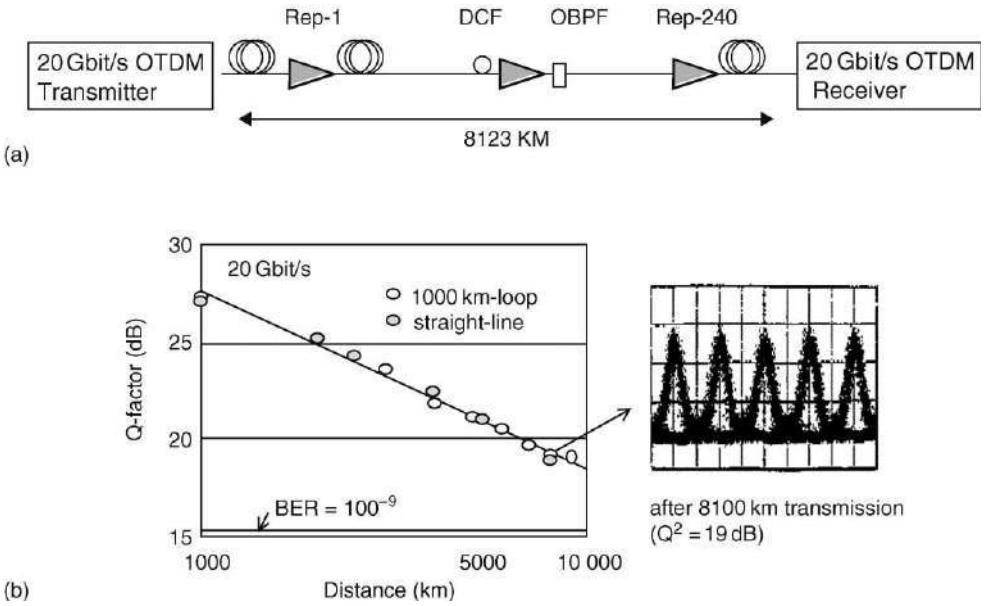

Figure 3 (a) Schematic diagram of the straight-line experiment with optical time-division multiplexing (OTDM) of soliton at 20 Gbit/s; (b) Measured Q factor vs. distance comparing experimental results obtained in the straight-line and loop configuration (loop length of 1000 km), respectively. The waveform received after 8100 km is also shown. With permission from Suzuki M and Edagawa N (2003) Dispersion-managed high-capacity ultra-long-haul transmission. *Journal of Lightwave Technology* 21: 916–929. © 2003 IEEE.

understand the main features of these multicomponent solitons. In this case, the evolution of the interacting envelopes E_j at carrier frequencies $j\omega_0$, $j = 1, 2$, is ruled by the coupled equations:

$$\left(\mathrm{i}\frac{\partial}{\partial z} - \frac{k_1''}{2}\frac{\partial^2}{\partial t^2}\right)E_1 + \chi E_2 E_1^* \exp(\mathrm{i}\Delta k z) = 0,$$
$$\left(\mathrm{i}\frac{\partial}{\partial z} + \mathrm{i}\delta k'\frac{\partial}{\partial t} - \frac{k_2''}{2}\frac{\partial^2}{\partial t^2}\right)E_2 + \chi E_1^2 \exp(-\mathrm{i}\Delta k z) = 0, \quad [4]$$

where the subscript $j = 1, 2$ indicates that the relative quantities evaluated at frequency $j\omega_0$, $t = T - k'_1 z$, $k''_{1,2}$ are GVD coefficients, $\delta k' \equiv k'_2 - k'_1$ is the group-delay, χ is the effective nonlinear coefficient, and $\Delta k = k_2 - 2k_1$ is the wave-number (or phase) mismatch. In spite of the fact that eqns [4] are not integrable, their stable soliton solutions can be found (numerically) in a wide range of parameters as symbiotic phase-locked pulses $E_{1,2}(z,t)$ that travel undistorted. For instance, SHG solitons exist either in phase-matching ($\Delta k = 0$) or far-off phase-matching (with some limitations for $\Delta k > 0$), and, unlike fiber solitons, also in normally dispersive media ($k''_{1,2} > 0$). As for fiber solitons, their width is related to their peak power, since the basic mechanism is the mutual balance between the dispersive broadening (that depend on the pulsewidth) and nonlinear phase-shifts (that depend on power), which occur due to repeated up-conversion and down-conversion processes, often referred to as cascading. A SHG soliton is robust enough to permit the observation of trapping, even from an input that differs considerably from the soliton (e.g., from the fundamental only), or in the presence of a relatively strong group-delay which tends to rip apart the two-component soliton, inducing walk-off between the two envelopes.

In contrast with fibers, in this case, experiments are necessarily made in relatively short bulk samples, thus resulting in a limited number of dispersion lengths. Usually broad (quasi-planewave) beams are conveniently employed in order to make diffraction ineffective.

The main experimental problem is that, in crystals commonly employed for SHG, the dynamics of typical laser pulses (duration above 100 fs) is fully dominated by the group-delay $\delta k'$. In other words, the material GVD becomes relevant only for pulses as short as a few fs. Therefore, the observation of solitons with longer pulses has been required to tailor the dispersion, which can be accomplished by means of pulses with tilted phase front. As displayed in **Figure 4**, successful compression of 200 fs (FWHM) tilted pulses was observed at relatively large negative Δk in a BBO bulk crystal with effective zero group-delay and enhanced (anomalous) GVD.

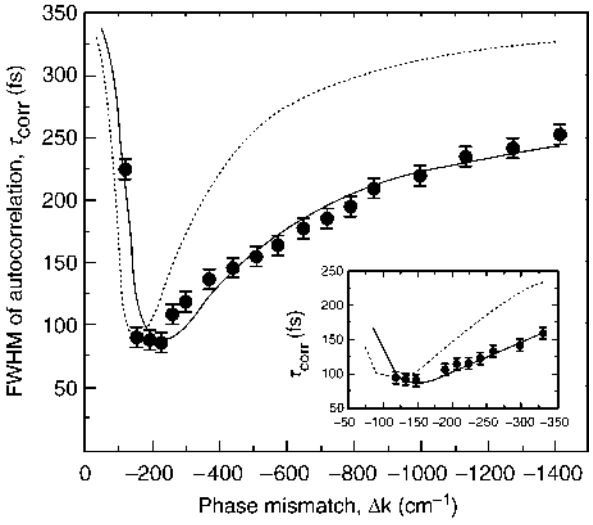

Figure 4 Soliton fomation via SHG: the circles show the output pulse duration (full width at half maximum (FWHM) of autocorrelation τ_{corr}) vs. mismatch Δk, measured after propagation of an input 200 fs tilted pulse (input $\tau_{corr} = 284$ fs) with 13.4 GW/cm^2 peak intensity, in a 7 mm BBO crystal (inset: same with 10.4 GW/cm^2). Superimposed (dotted line) is the corresponding theoretical value obtained from eqn [4]. The solid curve that fits better the data accounts for the additional effect of the material Kerr effect. With permission from Di Trapani P, Caironi D, Valiulis G, Dubietis A, Danielius R and Piskarskas A (1998) Observation of temporal solitons in second-harmonic generation with tilted pulses. *Physical Review Letters* 81: 570–573. © 1998 APS.

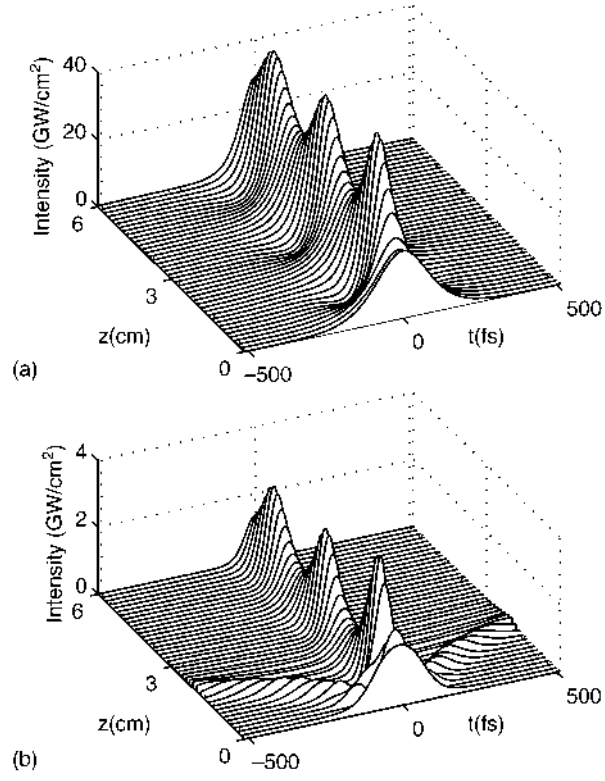

Figure 5 Dynamics of temporal soliton formation ruled by eqn [4]. A compressed soliton is formed via strongly mismatched ($\Delta k = -100$ cm^{-1}) SHG from a 200 fs (FWHM) input pulse at fundamental carrier frequency, in a crystal with $\chi = 1.4 \times 10^{-4}$ W$^{-1/2}$, effective dispersion $\delta k' = 0$ (group–velocity matching) and $k_1'' = -1.8$ ps^2/m, $k_2'' = -0.75$ ps^2/m.

The measured compression is a signature of soliton formation, as clearly indicated by the example of simulated pulse dynamics reported in Figure 5: the pulse at second-harmonic is generated and then trapped with the input one as a two-component bound state. Radiation and oscillations occur, in this case, because the launching conditions (single envelope) are far enough from the two-component soliton.

Temporal Solitons in Bragg Gratings

A linear periodic structure, such as a Bragg grating with pitch Λ, couples counter-propagating waves when their frequency lies close to the Bragg frequency ω_B (vacuum wavelength $\lambda_B = 2n_0\Lambda$) defined by the resonance condition $k(\omega_B) = \pi/\Lambda$. Such a type of structure supports a different type of temporal solitons, so-called gap solitons (GSs) since they originate from a linear dispersion relationship of the type shown in Figure 6a, characterized by a stop-band (gap) centered around ωB. The gap is characterized by a unitary grating reflectivity, as shown in Figure 6b, and the forward linear propagation is therefore forbidden. However, if the incoming wave is intense enough to induce the refractive index to increase through the optical Kerr effect, the Bragg resonance (and the whole gap) is shifted downwards (in frequency) allowing for self-transparency at the operating frequency. GSs can be understood as bell-shaped envelopes which are allowed to travel along the grating due to a high-intensity core that bleaches the reflectivity, while the high reflectivity seen by the tails takes the pulse together. GSs are expected to be slow because the slope of the dispersion curve tends to zero when approaching the gap, while its curvature results in the grating GVD (huge if compared with material GVD in real-world units), which is exactly balanced by the nonlinearity when the GS is formed. This naive portrait is supported by a more thorough analysis starting from the coupled-mode equations obeyed by the forward $E_+(z,T)$ and backward $E_-(z,T)$ envelopes at frequency ω_B, propagating along a (e.g., fiber) Bragg grating with Kerr effect (inducing SPM and

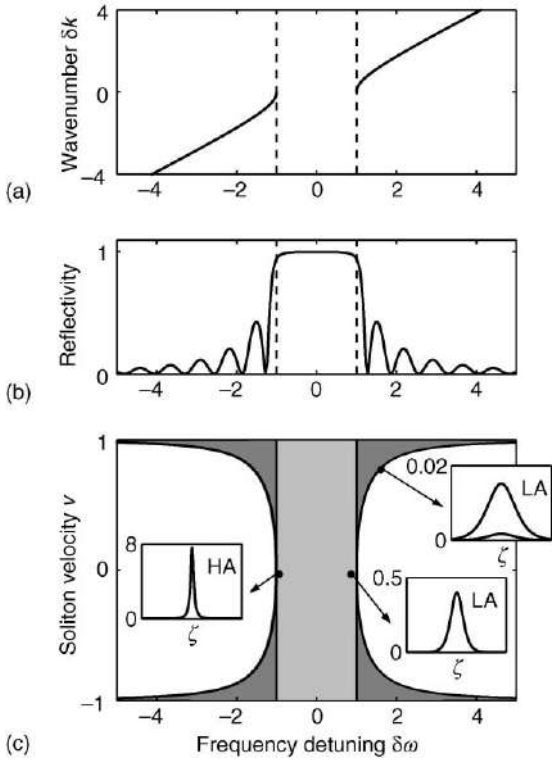

Figure 6 (a) Bragg linear dispersion relationship in normalized (dimensionless) units: wave-number shift $\delta k = (k(\omega) - \pi/\Lambda)/\Gamma$ vs. frequency detuning $\delta\omega = k'(\omega - \omega_B)/\Gamma$; (b) linear grating reflectivity vs. $\delta\omega$ for a grating figure of merit $\Gamma L = 4$; (c) existence domain (whole shaded area) of GSs in the normalized detuning–velocity plane ($\delta\omega, v \equiv k V_{sol}$). The light shaded area corresponds to the stop-band $|\delta\omega| < 1$ (inner region between dashed lines in a–b). The insets show samples of GS intensity profiles vs. the moving frame coordinate $\zeta = (1 - v^2)(z - V_{sol} T)$. Zero-velocity GSs have symmetric envelopes ($|E_+| = |E_-|$) and range from high-amplitude (HA) to low-amplitude (LA) across the gap. The third inset shows a moving GS in the experimentally accessible region, where it retains LA but has a stronger component in the direction of motion.

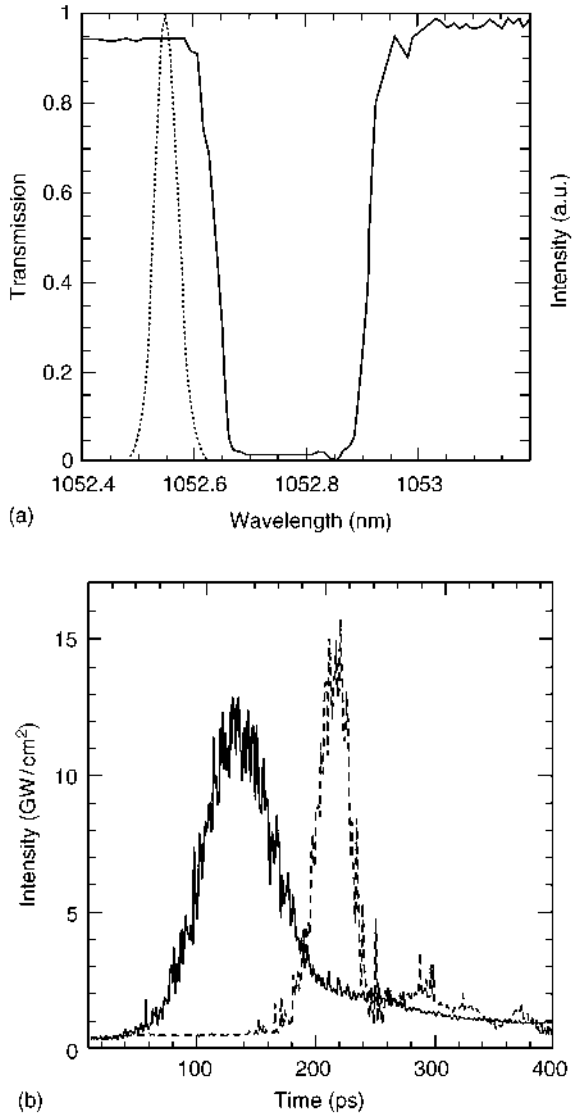

Figure 7 Gap solitons in a fiber Bragg grating: (a) measured pulse spectrum compared with the linear grating transmission, (b) compressed (15 ps) and delayed pulse transmitted at high intensity (dashed line), compared with the pulse transmitted at frequency far from the Bragg resonance (solid line) where it retains its input width of 80 ps. With permission from Eggleton BJ, Slusher RE, de Sterke CM, Krug PA and Sipe JE (1996) Bragg grating solitons. *Physical Review Letters* 76: 1627–1630. ©1996 APS.

cross-phase modulation):

$$i\left(k'\frac{\partial E_\pm}{\partial T} \pm \frac{\partial E_\pm}{\partial z}\right) + \Gamma E_\mp + \gamma(|E_\pm|^2 + 2|E_\mp|^2)E_\pm = 0 \qquad [5]$$

The linear dispersion associated with eqn [5] is as in Figure 6a, with a stop-band $|\omega - \omega_B| \le \Gamma/k'$ proportional to the index corrugation depth through the Bragg coupling coefficient Γ. In the nonlinear regime, GS solutions of eqn [5] differ qualitatively from conventional fiber solitons [2]:

(i) first, GSs are indeed slow since the two envelopes can travel locked with any soliton velocity V_{sol} such that $|v| \equiv |k' V_{sol}| \le 1$, i.e., lower than the natural group velocity k'^{-1};

(ii) GSs exist only in a narrow bandwidth, i.e., the shaded area in Figure 6c, which corresponds to the gap seen in the soliton frame (it reduces to the conventional stop-band for still solitons and gets wider for faster GSs); and

(iii) the gap is strongly asymmetric since lower intensity is required to induce transparency close to the high-frequency band-edge where, as a consequence, GSs have low amplitudes and

are stable (the amplitude grows moving towards the low-frequency bandedge, where GSs become unstable).

Experimentally, GSs have been typically observed with pulses at frequency close to the blue bandedge of fibers (see Figure 7a) and AlGaAs waveguides with built-in Bragg gratings of a few centimeters. The typical signature of GSs, illustrated in Figure 7b, is the enhanced pulse transmission observed under the operating conditions of Figure 7a when the input intensity is sufficiently high. The transmitted pulse experiences pulse compression and a significant group-delay (measured with respect to the transit-time at frequency far off the stop-band), witnessing its nature of slow-light solitons.

Finally, we emphasize that grating solitons can also be supported in the case where the two modal envelopes propagate in the forward direction along a grating whose period Λ is long enough to be resonant with the modal phase slippage, thus inducing efficient linear coupling between them. In this case, the structure has a gap in wave-number, and solitons are slow in the frame traveling at the average modal group-velocity. Soliton formation is expected to lead to grating-induced compensation of the group-delay between the modes. Experiments on such types of structures have been carried out only recently.

Solitons in Dissipative Nonlinear Optics

Temporal solitons can also be observed in nonlinear optical interactions that involve changes of the state (energy level) of matter as in inverted media or scattering phenomena. For instance, early studies in the 1960s have demonstrated that coherent pulses with energy above a critical level can pass through resonant (two-level) atoms without being absorbed, a phenomenon denoted as self-induced transparency. Such pulses are temporal solitons solutions of the Maxwell–Bloch model which describes the atom–light interaction.

Also scattering phenomena occurring in optical fibers, such as Raman and Brilluoin scattering, possess three-wave solitons, in which the pump and scattered (Stokes) light waves are coupled to a phonon field of the acoustic or optical branch (for Brillouin or Raman, respectively).

Perhaps more important from the point of view of the modern applications, are fiber lasers composed by spans of active and/or passive fiber closed to form a recirculating loop. These types of structures can often be described by distributed (averaged over several round-trips) models, which take the form of a dissipative perturbed NLSE, known as the Ginzburg–Landau equation (GLE):

$$\mathrm{i}\frac{\partial E}{\partial z} - \frac{k''}{2}\frac{\partial^2 E}{\partial t^2} + \chi|E|^2 E = \mathrm{i}gE + \mathrm{i}g_{\mathrm{nl}}|E|^2 E + \mathrm{i}q|E|^4 E + \frac{\mathrm{i}}{\Omega_{\mathrm{f}}^2}\frac{\partial^2 E}{\partial t^2} \qquad [6]$$

where g is the excess (net) gain ($g > 0$) or loss ($g < 0$), g_{nl} and q describe the nonlinear gain mechanism and its saturation (usually introduced to control the mode-locking), and Ω_{f} is linked to spectral filtering. Several soliton (pulse) solutions of the GLE [6] can be found even when the parameters (g, g_{nl}, q, Ω_{f}) are such as to constitute a strong perturbation of the NLSE. Unlike NLSE solitons [2], pulse solutions of the GLE [6] are usually chirped (i.e., possess a nonlinear phase profile), and have fixed amplitude for a given value of the parameters. The set of soliton waveforms is generally richer and includes, for instance, flat-top stationary pulses, two-pulse bound-states, solitons moving at fixed velocity, etc. Although instabilities of the background (for $g > 0$) or of the pulse itself can affect their observability, large islands of stability in the parameter space can be found. The relevant role of stable solitons in mode-locking and intra-cavity pulse formation mechanisms has been demonstrated in many different fiber laser geometries, for which we refer the reader to the more specialized papers listed as Further Reading at the end of this article. We point out that some schemes, for instance the stretched pulse-fiber laser developed at MIT, involve a basic configuration (alternating GVD within the cavity length) which is quite similar to that seen in the framework of DMSs, thus exhibiting the same basic physical mechanism (strong compression and stretching of the pulse along each round-trip which play the role of the map period for a DMS).

List of Units and Nomenclature

BBO	β-barium borate
DCF	dispersion compensating fiber
DMS	disperion managed soliton
EDFA	Erbium doped fiber amplifier
GLE	Ginzburg–Landau equation
GS	gap soliton
GVD	group-velocity dispersion
NLSE	nonlinear Schrödinger equation
OTDM	optical time-division multiplexing
SHG	second-harmonic generation

SPM self-phase modulation
WDM wavelength division multiplexing
Amplifier spacing L_A [m]
Bragg frequency ω_B [Hz]
Bragg period Λ [m]
Central wavelength λ_0 [nm]
Coupling coefficient Γ [m^{-1}]
Dispersion coefficient D [ps nm^{-1} km^{-1}]
Dispersion length L_d [m]
Effective area A_{eff} [m^2]
First-order dispersion k' [s/m]
Fundamental soliton power P_1 [W]
Gain (linear) g [m^{-1}]
Gain (nonlinear) g_{nl} [m^{-1}]
Group-delay δ [s/m]
Group-velocity $V = k'^{-1}$ [m/s]
Group velocity dispersion k'' [ps^2 m^{-1}]
Inverse pulse-width η [s^{-1}]
Light velocity in vacuum c [m/s]
Loss coefficient α [m^{-1}]
Nonlinear coefficient (fiber) γ [m^{-1} W^{-1}]
Nonlinear coefficient (quadratic bulk materials) χ [$W^{-1/2}$]
Nonlinear refractive index n_{2I} [m^2 W^{-1}]
Nonlinear length L_{nl} [m]
Refractive index (linear) n_0
Retarded time (in the moving frame) t [s]
Spectral filtering coefficient Ω_f [Hz $m^{1/2}$]
Time in the laboratory frame T [s]
Wave-number mismatch Δk [m^{-1}]

See also

Fiber and Guided Wave Optics: Nonlinear Effects (Basics); Nonlinear Optics. **Fiber Gratings**. **Lasers:** Optical Fiber Lasers. **Nonlinear Optics, Basics:** $\chi^{(2)}$–Harmonic Generation. **Optical Communication Systems:** Wavelength Division Multiplexing. **Scattering:** Scattering Phenomena in Optical Fibers. **Solitons:** Bright Spatial Solitons; Optical Fiber Solitons, Physical Origin and Properties.

Further Reading

Ablowitz MJ, Biondini G and Ostrovsky LA (eds) (2000) Focus issue: optical solitons: perspectives and applications. *Chaos* 10: 471–640.

Agrawal GP (1995) *Nonlinear Fiber Optics*, 2nd edn. London: Academic Press.

Agrawal GP and Kivshar YS (2003) *Optical Solitons*. London: Academic Press.

Akhmediev NN and Ankiewicz A (1997) *Solitons: Nonlinear Pulses and Beams*. London: Chapman and Hall.

Buryak AV, Di Trapani P, Skryabin D and Trillo S (2002) Optical solitons due to quadratic nonlinearities: from basic physics to futuristic applications. *Physics Reports* 370: 63–235.

De Sterke CM, Eggleton BJ and Sipe JE (2001) In: Trillo S and Torruellas W (eds) *Spatial Solitons*. Berlin: Springer-Verlag.

Di Trapani P, Caironi D, Valiulis G, Dubietis A, Danielius R and Piskarskas A (1998) Observation of temporal solitons in second-harmonic generation with tilted pulses. *Physics Review Letters* 81: 570–573.

Eggleton BJ, Slusher RE, de Sterke CM, Krug PA and Sipe JE (1996) Bragg grating solitons. *Physics Review Letters* 76: 1627–1630.

Forysiak W, Knox FM and Doran NJ (1994) Average soliton propagation in periodically amplified systems with stepwise dispersion-profiled fiber. *Optics Letters* 19: 174–176.

Hasegawa A (2000) Soliton-based optical communications: An overview. *IEEE Journal of Selected Topics in Quantum Electronics* 6: 1161–1172.

Hasegawa A and Kodama Y (1995) *Solitons in Optical Communications*. London: Oxford University Press.

Hasegawa A and Matsumoto M (2002) *Optical Solitons in Fibers*. Berlin: Springer-Verlag.

Haus HA (2000) Mode-locking of lasers. *IEEE Journal of Selected Topics in Quantum Electronics* 6: 1173–1185.

Mollenauer LF and Mamyshev PV (1998) Massive wavelength-division multiplexing with solitons. *IEEE Journal of Quantum Electronics* 34: 2102.

Nakazawa N, Kubota H, Sahara A, Suzuki K and Yamada E (2000) Ultrahigh-speed long-distance TDM and WDM soliton transmission technologies. *IEEE Journal of Selected Topics in Quantum Electronics* 6: 363.

Porsezian K and Kuriakose VC (2003) *Optical Solitons: Theoretical and Experimental Challenges*. Berlin: Springer-Verlag.

Slusher RE and Eggleton BJ (eds) (2003) *Nonlinear Photonics Crystals*. Berlin: Springer-Verlag.

Suzuki M and Edagawa N (2003) Dispersion-managed high-capacity ultra-long-haul transmission. *Journal of Lightwave Technology* 21: 916–929.

Suzuki M, Morita I, Edagawa N, Yamamoto S, Toga H and Akiba S (1995) Reduction of Gordon-Haus jitter by periodic dispersion compensation in soliton transmission. *Electronic Letters* 31: 2027–2028.

Turitsyn SK, Aceves AB, Jones CKRT, Zharnitsky V and Mezentsev VK (1999) Hamiltonian averaging in soliton-bearing systems with a periodically varying dispersion. *Physics Review E* 59: 3843–3846.

Wabnitz S, Kodama Y and Aceves AB (1995) Control of optical soliton interactions. *Optics and Fiber Technology* 1: 187–217.

Zakharov VE and Wabnitz S (1999) *Optical Solitons: Theoretical Challenges and Industrial Perspectives*. Berlin: Springer-Verlag.

SPECTROSCOPY

Contents

Absolute Optical Frequency Metrology
Fourier Transform Sepctroscopy
Hadamard Spectroscopy and Imaging
Nonlinear Laser Spectroscopy
Raman Spectroscopy
Second-Harmonic Spectroscopy
Single Molecule Spectroscopy

Absolute Optical Frequency Metrology

S T Cundiff, National Institute of Standards and Technology and University of Colorado, Boulder, CO, USA

L Hollberg, National Institute of Standards and Technology, Boulder, CO, USA

Published by Elsevier Ltd.

Introduction

Visible light occupies a portion of the electromagnetic spectrum between frequencies of approximately 400 terahertz (THz) to 750 THz, where 1 THz is 10^{12} oscillations per second. Optical radiation, which includes adjacent infrared and ultraviolet portions of the spectrum, covers a frequency range of approximately 15–1500 THz. The techniques required for measuring frequencies in this range are qualitatively different from those currently employed in the radio frequency (RF) and microwave portions of the spectrum, where high-speed electronic counters make a direct measurement possible.

The frequency of an oscillatory phenomenon is the number of oscillations that occurs in unit time, or alternatively, the inverse of the period of one oscillation. Thus, to understand frequency, we must understand time, which is intrinsically related to phase. Only a time interval, the time elapsed between two events, has fundamental physical meaning. The 'time' that we are most familiar with, the time of day (including the date), is based on an arbitrarily chosen starting point. The starting point chosen in current international agreements can be traced to noon on December 31st, 1899. Furthermore, the time defined by those agreements, Universal Coordinated Time (UTC), is periodically adjusted relative to atomic time (discussed below) to compensate for irregularities in the rotation of the Earth.

Knowledge of the transition frequencies (energies) in simple atoms gives detailed information on the structure of the atoms. This can be used to test our understanding of the fundamental interactions in nature by comparing the measured frequencies with predictions of fundamental physical theories, in this case quantum electrodynamics (QED). Beautiful results along these lines are exemplified by precision measurements of transition frequencies in hydrogen and helium. In addition, the ability to measure and precisely control laser frequencies promises tremendous advances in the performance of the next generation of atomic clocks and frequency standards. These have potential applications to navigation and communication systems.

The measurement of frequency and time intervals also involves a choice, that of the unit of time, i.e., the definition of the second. The original definition of the second was based on the rotation of the Earth. However, not only is it irregular, but it is slowing down. Both of these facts limit its utility as a frequency standard. Once atomic clocks were highly developed, the second was redefined in 1967 to be the time it takes for the $F = 4$, $m_F = 0 \rightarrow F = 3$, $m_F = 0$ transition in the hyperfine structure of the ground state of ^{133}Cs to undergo 9192 631 770 oscillations. Thus an 'absolute' measurement of time interval, and hence frequency, must be directly connected to this defining frequency. The basic structure of an atomic clock is shown in Figure 1.

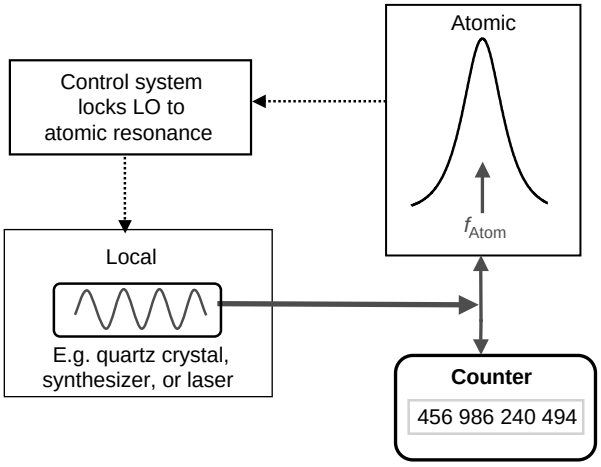

Figure 1 An atomic clock has three basic components: an atomic resonance, an oscillator to probe and lock to the atomic resonance, and a counter that records the number of oscillations and thus displays the time interval since some chosen starting time.

Frequency standards (or clocks) are characterized by three properties: accuracy, stability, and reproducibility. Accuracy describes how well the natural and fundamental atomic frequency is produced by the atomic frequency standard. Stability describes the frequency fluctuations on short timescales (an accurate but unstable standard produces a frequency that fluctuates around a constant mean that is the atomic frequency). Reproducibility between standards addresses whether or not two implementations of the standard produce exactly the same frequency. Atomic clocks are intrinsically accurate because the frequency of the transitions they are based on is determined by identical atoms in nature. The largest inaccuracies are usually due to environmental perturbations such as tiny variations in the local magnetic field.

A careful analysis of all of these issues, together with technical considerations, yielded the choice of cesium for the current definition of the second. Other common frequency standards include quartz crystal oscillators, hydrogen masers, and rubidium vapor cell standards. Quartz oscillators are inexpensive and have good stability, but in comparison to atomic standards they suffer from accuracy and reproducibility issues. Hydrogen masers have the best stability, but also poor accuracy and reproducibility compared to cesium. Vapor cell rubidium standards can be small and inexpensive but have poorer accuracy than cesium beam standards due to cell and buffer gas effects.

There are several motivations for using optical sources rather than RF or microwaves to measure time intervals and frequencies. Obviously, a higher oscillation frequency can divide time into smaller units and this can provide higher measurement precision. It is thus natural to use the highest oscillation frequency that can be precisely counted to measure time and frequency. Not long after lasers were invented, suggestions were made that it would be possible to make optical frequency standards using lasers locked to narrow atomic resonances.

A basic feature of optical transitions is that they have a very high Q factor (the ratio of the center frequency to the linewidth). This allows them to be measured to very high precision (current measurements are good to a few parts in 10^{15}). It may be possible to use this high precision to build improved atomic clocks that use optical transitions instead of the microwave transition currently being used. Candidates for optical standards include both trapped ions (Hg^+, Yb^+, Sr^+, and In^+ are under investigation) and laser-cooled neutral atoms (Ca, Sr, Mg, Ba, and Ag). The high precision is also utilized for fundamental tests of quantum electrodynamics by comparing transition frequencies (usually of hydrogen or helium atoms) to first-principle calculations. For comparison Figure 2 shows the relevant transitions and energy levels in the Cs microwave standard and the Hg^+ optical frequency. It is even possible to use the precision of atomic clocks for high-sensitivity tests of variation in the fundamental constants with time. Finally, because the speed of light is constant length measurements based on the wavelength of light with a known frequency, the precision of length metrology ultimately rests on optical frequency metrology.

An obvious question, which might occur to many in the field of optics, is why a standard spectrometer does not provide absolute frequency measurements. All spectrometers measure wavelength, not frequency, based on interference between light traveling two different path lengths. Thus, knowledge of the path length is needed to determine the wavelength. However, since length is now defined using the speed of light, the definitions become circular unless a source of light with known absolute frequency is available. In addition, frequency measurement is intrinsically more precise and reliable than mechanical length measurement.

Recently, there has been a significant breakthrough in optical-frequency metrology and optical clocks by using mode-locked lasers that generate optical pulses with durations of a few femtoseconds. Before discussing optical frequency metrology based on mode-locked lasers, we will review earlier techniques to provide the necessary background. After describing

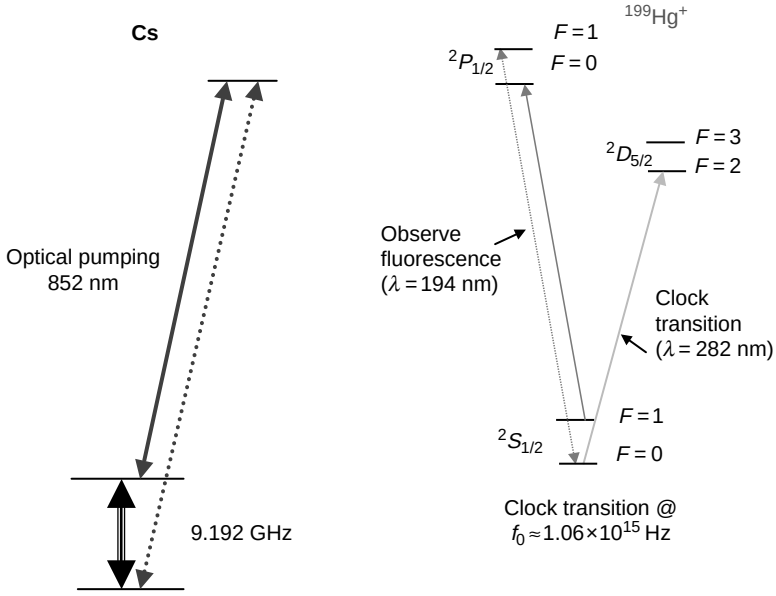

Figure 2 Simplified energy level diagram of Hg^+ and Cs.

methods that use mode-locked lasers, we will then provide a summary of current measurements and standards. Finally we will briefly describe the outlook for the future of absolute optical frequency metrology.

Background

To imagine being able to measure the frequency of light we need to have laser sources with high coherence, like that of radio waves, where the phase is stable enough to be measured. Even though the first beat-note between two lasers was demonstrated in the 1960s, most lasers were not coherent (stable) enough that the optical phase could be tracked for much longer than about 10 to 100 ns. After many years of research and inspiring technological achievements, Hall, Chebotayev, Bergquist and others were able to stabilize the frequency of tunable laser sources to the point that lasers could indeed be considered coherent sources, with the phase continuously measurable and even predictable for times as long as seconds. Once the lasers were stable enough to be counted, there was still the essential problem that no counter was fast enough to actually count optical frequencies.

With spectrally pure laser sources it was possible to conceive of measuring optical frequencies by extending the nonlinear methods that are used to generate and measure RF and microwave frequencies. By using simple nonlinear components such as diodes, it is easy to generate high harmonics of an input frequency in the RF and microwave regions of the spectrum. At least in principle, and by direct analogy, it seemed that it should be possible to multiply coherent sine waves from radio frequencies up through the microwave region to reach the THz frequencies of far-infrared lasers, and then to use nonlinear optical methods to multiply on up to the infrared (IR) and eventually to the visible region of the spectrum. The concept is simple enough, $f \rightarrow 100 \times f \rightarrow 1000 \times f \rightarrow 2000 \times f \rightarrow 4000 \times f \ldots$ until you reach an optical frequency. Multiplication up to the microwave region works quite well, but unfortunately the nonlinear optical processes in the far-infrared, IR, and optical regions are typically quite inefficient; the power at the second harmonic is typically only 10^{-5} times the square of the fundamental power in watts. Just multiplying a single-frequency laser by a factor of two can be a technological challenge. After reaching a microwave frequency of about 50 GHz we still require a multiplication factor of 10^4 to reach the visible at 500 THz. Since we are usually forced to multiply in the IR and optical by factors of 2, we have $2^N = 10^4$, which means we require $N \approx 13$ multiplication stages. This is approximately the number of stabilized laser oscillators that are required to span the frequency gap from the microwave to the visible. Notwithstanding these technological limitations, Evenson and collaborators did a groundbreaking demonstration of a harmonic frequency chain in 1972 that connected a microwave atomic frequency standard and a stabilized HeNe laser in the IR at 88 THz. With a concerted multiperson effort, more lasers, and 10 years of research, Jennings and co-workers finally succeeded in extending the harmonic optical frequency chain to the visible. That original optical frequency chain from the RF to the

visible is diagramed in Figure 3. In the succeeding 18 years, only three other harmonic frequency chains were built world-wide that connected the microwave standards to optical frequencies.

An alternative scheme for measuring optical frequencies is the optical bisection method, a conceptually simple technique proposed by Telle, Meschede and Hänsch, for dividing down from

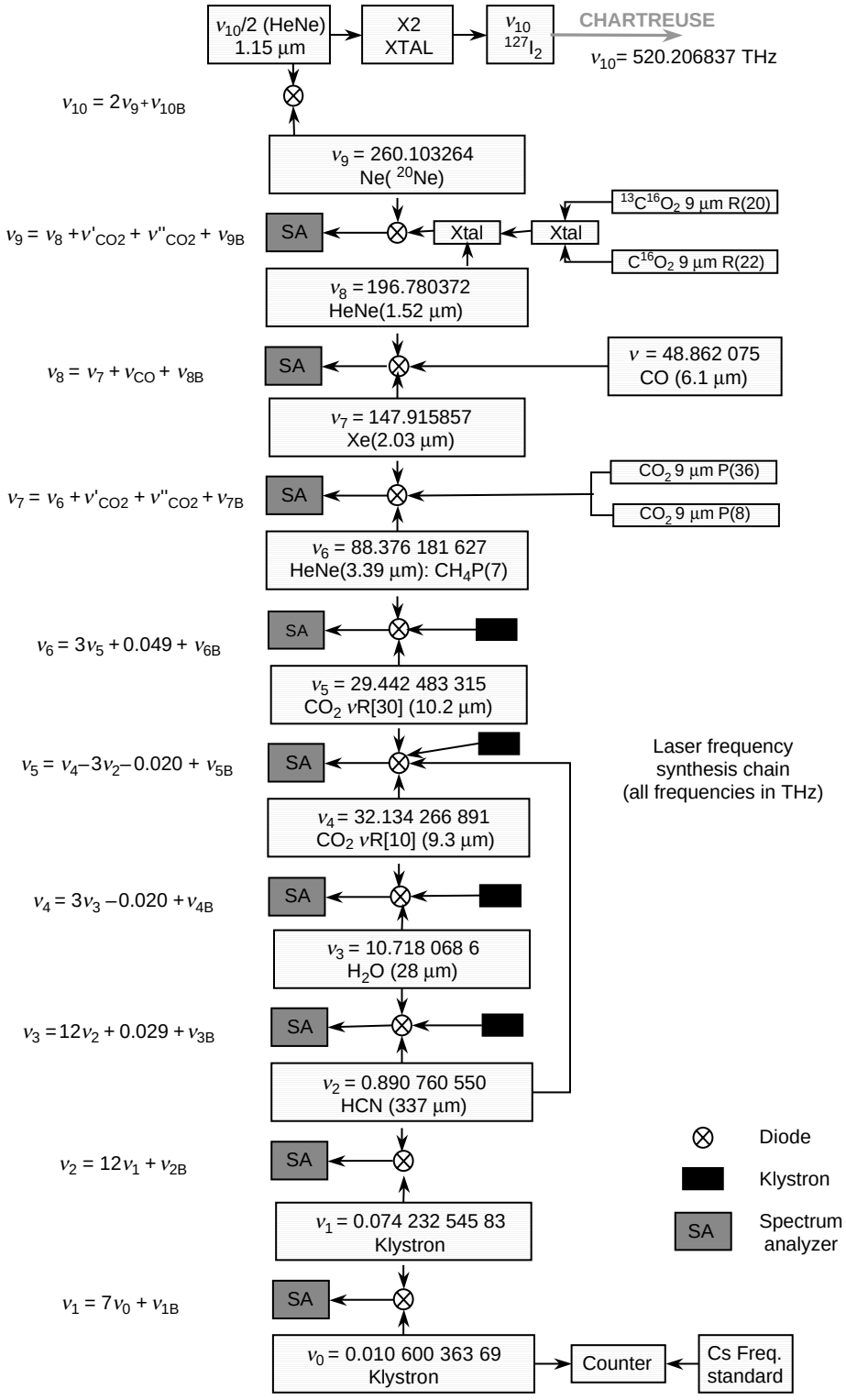

Figure 3 Schematic of the first frequency chain used to measure optical frequencies relative to the Cs standard.

optical frequencies to microwaves. In fact, the bisection method divides an arbitrary optical frequency interval in half by utilizing second-harmonic generation and sum-frequency-mixing, and forcing the condition that $2f_3 = f_1 + f_2$. Thus, the laser at frequency $f_3 = (f_1 + f_2)/2$ bisects the frequency interval between the lasers at f_1 and f_2. The optical bisection method has the following two significant advantages over the multiplication method: the phase noise is decreased in successive divisions (rather than increased by multiplications), and the system could be constructed out of one convenient and reproducible laser technology, say diode lasers in the near-visible. In principle, any optical frequency could be measured by successive operations of the optical bisector. Optical frequency intervals as large as 8 THz have been measured with this method, but even though it is feasible, no bisection system has been built to date that connects the optical region all the way to countable microwave frequencies. The main limitation is the same as that facing the multiplication scheme: it still requires lots of stabilized laser sources to divide by factors of 2 from 10^{15} to 5×10^{10} Hz. The problem can be reduced somewhat by taking advantage of an optical-frequency comb generator based on an electro-optic modulator inside a resonant optical cavity. These systems, developed by Korougi and others, generate many coherent microwave sidebands on a laser so that optical frequency intervals on the order of 5 THz can be measured.

Even combining all of these ideas and techniques, a convenient system to count optical frequencies remained elusive. This has now changed with the revolutionary new optical frequency combs based on mode-locked femtosecond lasers.

Optical Frequency Metrology with Femtosecond Combs

Recently, the introduction of mode-locked lasers into the field of absolute optical frequency metrology has resulted in an important advance. The extremely complex frequency chains described above can be replaced by a single mode-locked laser if it produces pulses with sufficiently large bandwidth. Although the underlying ideas are not new (they were originally discussed by Hänsch and Chebotayaev in the 1970s) it is only recently that mode-locked lasers with the required characteristics were developed and the concept carefully tested.

The dramatic simplification of a frequency chain that uses a mode-locked laser means that absolute optical-frequency measurements can now be made by a single person as compared to the team of approximately 10 highly trained scientists required to run the previous chains. This has resulted in a large number of measurements being reported in the last 2 years. The simplicity also allows longer averaging times and a greater number of measurements to be used in determining the final reported frequency. In addition, since a simpler system has fewer sources of error, it has improved the quality of measurements made in a given amount of time.

Mode-locked lasers produce ultrashort pulses of light. For a typical modern high-quality mode-locked laser, the pulse duration is around 10 fs (1 fs = 10^{-15} second), corresponding to three or four optical cycles in the near-infrared, where these lasers typically operate. The current best mode-locked lasers can produce pulses that are shorter than two optical cycles in duration (5–6 fs). Since the spectrum of these lasers is complicated, the bandwidth is significantly larger than it would be for a smooth pulse such as one with a Gaussian or hyperbolic-secant temporal intensity profile.

At first sight, using mode-locked lasers for optical-frequency metrology seems counter-intuitive, because optical-frequency metrology clearly requires very well-defined frequencies in contrast to the broad frequency spectrum of ultrashort pulses. This apparent conundrum is resolved by the fact that a mode-locked laser actually produces a very regularly spaced train of ultrashort pulses. The ultrashort pulses emitted by a mode-locked laser are replicas of a steady-state pulse that circulates inside the cavity of the laser. Every time the internal pulse impinges on the output coupler, which is a partially reflective mirror, a portion of it is transmitted, producing the output. The timing between these output pulses is determined purely by the time, τ_c, for the intracavity pulse to make one round trip, which is typically between 1 and 20 ns. The spectrum of a train of pulses is a regularly spaced 'comb' of frequencies spaced by the repetition rate of the train, $f_{rep} = 1/\tau_c$. If all of the pulses are identical, then these comb lines are just integer multiples of f_{rep}. However, due to dispersion, the group and phase velocities are different inside the cavity of a mode-locked laser. This leads to a pulse-to-pulse shift in the phase between the carrier and envelope (which we call the carrier–envelope phase; ϕ_{ce}, the pulse-to-pulse shift of ϕ_{ce} is designated by $\Delta\phi_{ce}$). The presence of $\Delta\phi_{ce}$ results in a rigid shift of all the comb lines by an amount $\delta = 2\pi\Delta\phi_{ce}/f_{rep}$. Thus the frequencies of the comb lines are given by $\nu_n = nf_{rep} + \delta$, where n is a large integer of order 10^6. The essential point of this equation is that it gives optical frequencies in terms of RF frequencies (f_{rep} and δ) that can easily be measured with conventional electronics and

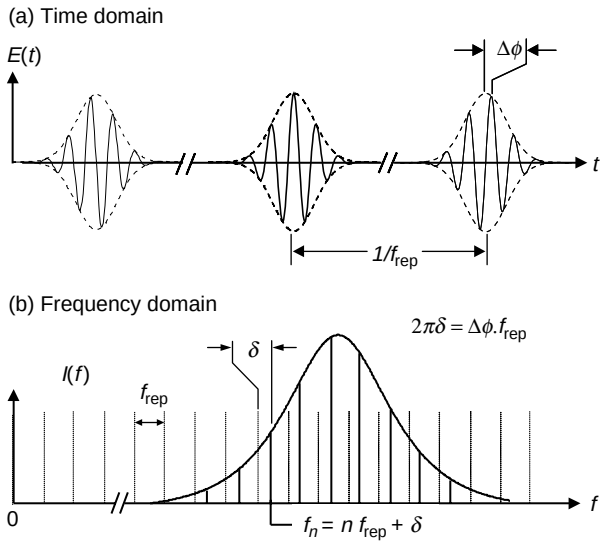

Figure 4 Time–frequency correspondence and relationship between $\Delta\phi$ and δ. (a) In the time domain, the relative phase between the carrier and the envelope evolves from pulse to pulse by the amount $\Delta\phi$. (b) In the frequency domain, the elements of the frequency comb of a mode-locked pulse train are spaced by f_{rep}. The entire comb is offset from integer multiples of f_{rep} by an offset frequency $\delta = \Delta\phi f_{\text{rep}}/2\pi$.

compared to the microwave cesium frequency standard. The relationship between the time and frequency domains is shown in Figure 4.

Given such a comb of frequencies, a heterodyne measurement readily yields the frequency difference between the unknown optical frequency of a narrow-band source (typically a single frequency laser) and the optical frequency of a nearby comb line. Thus, the absolute unknown frequency can be determined if the absolute frequencies of the comb lines are known. This requires measurement of f_{rep} and δ. The repetition rate, f_{rep}, is easily measured with a fast photodiode. Measurement of δ requires somewhat greater effort.

The development of a 'self-referencing' method to easily measure δ is the key enabling breakthrough that has brought about the recent revolution in optical-frequency metrology. It is called self-referencing because it provides a direct measure of δ with no other optical frequencies as input. It works by comparing frequencies that differ by a factor of two (typically these lie in the wings of the spectrum) using second-harmonic generation. Specifically, the frequency difference between the second harmonic of a comb line n and the comb line $2n$ is given by $2f_n - f_{2n} = 2(nf_{\text{rep}} - \delta) - (2nf_{\text{rep}} - \delta) = \delta$. Such a frequency difference is easily measured using a heterodyne beat. This technique is shown schematically in Figure 5. Other variations on this basic idea are possible; for example, rather than frequency doubling

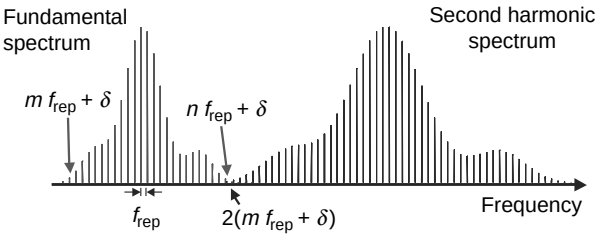

Figure 5 Schematic of the self-referencing technique for determining the offset frequency of the frequency comb produced by a femtosecond mode-locked laser. The frequency difference between the fundamental comb and its second harmonic is the offset frequency. In the region where overlap occurs, this is easily detected as a heterodyne beat.

the comb itself, it is also possible to extract the same information by comparing the comb lines to a single-frequency laser and its second harmonic.

Implementation of the self-referencing technique requires a pulse spectrum that spans a factor of two in frequency (an optical octave). Although a laser has recently been demonstrated that produces such a broad spectrum, they are not commonly available. However, the development of microstructured fiber has made it easy to broaden the spectrum of ordinary mode-locked lasers so that it spans an octave, as was first demonstrated by Ranka and co-workers at Bell Labs. The broadening occurs because optical nonlinearity in the fiber results in self-phase modulation of the laser pulse. Microstructure fiber achieves guiding by surrounding the core region with microscopic air holes, as compared to regular fiber that uses doping to produce a difference in the index of refraction. The much larger difference in the index of refraction in microstructure fiber allows a smaller core region to be used, which increases the effective nonlinearity. It also makes it possible to modify the dispersion such that the group velocity dispersion goes through zero for light with a wavelength close to 800 nm. This is crucial because group-velocity dispersion in ordinary fiber causes ultrashort pulses to temporally spread, which lowers the peak power and hence nonlinearity. It is possible to implement self-referencing techniques using less bandwidth if higher orders of nonlinearity are used; for example, if the difference between the second harmonic of the high-frequency end and third harmonic of the low-frequency end of the comb are used, only a half-octave bandwidth is required.

Given the ability to measure both f_{rep} and δ, and as long as its frequency lies within the comb spectrum, it is possible to measure the optical frequency of any optical source by measuring the heterodyne beat

with a nearby comb line. Thus, an absolute optical frequency measurement comes down to measuring three RF frequencies. It is often advantageous to actively control either one or both of f_{rep} and δ to hold them to a fixed frequency. This requires real-time adjustment of some parameters of the mode-locked laser to affect the necessary feedback. Controlling f_{rep} is straightforward; it is achieved by simply changing the cavity length of the laser, typically by mounting a mirror on a piezoelectric transducer. Controlling δ is more challenging, as it requires a differential change between the cavity group delay and phase delay. This has been achieved by adjusting the pump power, which in turn controls the energy of the intracavity pulse, and/or by tilting the end mirror of the cavity in lasers that employ prism dispersion compensators. Since the optical spectrum is spread out across this mirror, tilting it can be thought of as a frequency-dependent change in the cavity length, with the optical frequency corresponding to the pivot point of the tilt experiencing no length change, while those on opposite sides of the pivot point see changes with opposite signs.

In addition to the three RF frequencies, it is necessary to know the large integer n that describes the comb line that is beat against the unknown source. If a priori knowledge about the unknown source is available with precision somewhat better than $f_{rep}/2$ or better, n can readily be determined. For typical mode-locked lasers, standard commercial wavemeters can provide this level of precision. In the absence of additional outside information, n can also be determined with some effort and precision measurements by changing f_{rep} and δ in a systematic fashion. In the end, the simple formula $f_n = nf_{rep} + \delta$ gives the optical frequency of any arbitrary nth mode of the femtosecond optical comb. With this established, the frequency of any single-mode laser within the octave span of the comb can be determined simply by measuring the frequency of the beat-note generated in a photodiode between the cw laser and the known frequencies of the comb.

The diagram in Figure 6 shows how a femtosecond comb acts as an optical frequency synthesizer, which provides frequencies that are known relative to the input frequency (provided by a hydrogen maser in the figure). A typical series of measurements is shown in Figure 7 for the NIST Hg^+ single ion standard. The inset shows the results of numerous measurements for the frequency of Hg^+ relative to the NIST primary cesium standard.

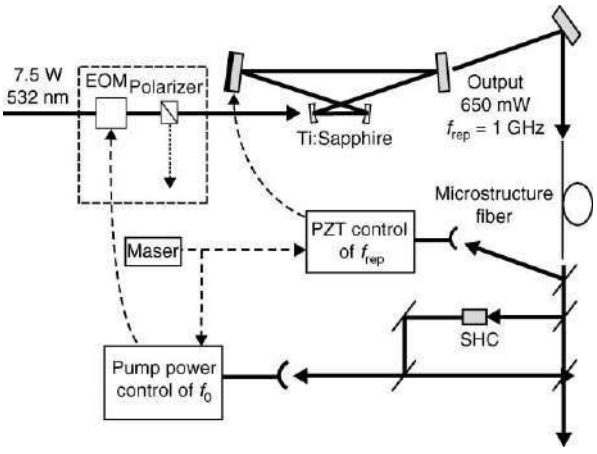

Figure 6 Schematic example of how the femtosecond comb is configured to make optical frequency measurements relative to the reference frequency provided by a hydrogen maser (Maser) that is in turn calibrated to a cesium atomic frequency standard. The repetition rate of the pulses from the Ti:Sapphire laser are detected in a photodiode and are controlled using a piezoelectric transducer (PZT) that changes the length of the laser cavity. The microstructure fiber and second harmonic crystal (SHC), provide the 'self-referencing' beat-note signal that is used to control the offset frequency, δ, by changing the power of the pump laser with an electro-optic modulator (EOM). This 'self-referenced' system then provides a set of modes with known frequencies spanning the visible. An unknown laser frequency can then be measured by making a beat-note with one of the modes of the comb.

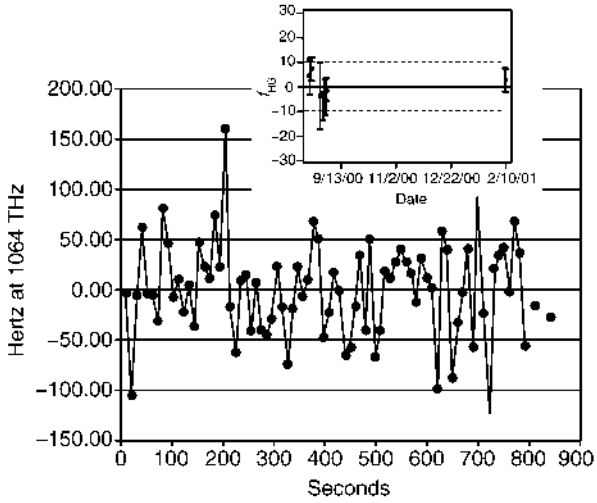

Figure 7 Actual frequency measurements of the beat-note between a hydrogen maser stabilized optical frequency comb (as shown in Figure 6) and the NIST Hg^+ optical frequency standard. The scatter in the data for counter gate time of 10 seconds corresponds to a fractional frequency uncertainty of about 4×10^{-14}, consistent within the instability in the maser. The inset shows the results from numerous datasets of measurement of the frequency of the Hg^+ optical standard relative to the cesium primary standard at NIST. The frequency data are plotted relative to 1064 721 609 899 142.6 Hz. The dotted lines give estimates of systematic uncertainties in lieu of a complete accuracy evaluation.

Table 1 Summary of optical frequency measurements of a selection of molecular, atomic, and ionic transitions. Recent measurements listed for all and additional measurements are given for a few to indicate the level of agreement

Atom/molecule/ion	*Wavelength (nm)*	*Frequency (THz)*	*Uncertainty*		*Year*	*First author and institution*
			± Hz	*Frac.*		
O_sO_4	10318.0	29.054 057 446 660	50	2×10^{-12}	1985	Clarion, LPTF
		29.054 057 446 579	10	3×10^{-13}	1999	Ducos, LPTF
CH_4	3392.2	88.376 181 627 000	50 000	6×10^{-10}	1972	Evenson, NBS
		88.373 149 028 553	200	2×10^{-11}	1998	Ering, PTB
Rb–2 photon	778.11	385.285 142 367 000	8000	2×10^{-11}	1993	Nez, ENS
		385.285 142 374 800	3000	8×10^{-12}	2000	Diddams, JILA
Sr^+	674.03	444.779 904 409 540	200	4×10^{-13}	1999	Bernard, NRC
Ca	657.45	455.986 240 495 150	8	10×10^{-14}	2003	Helmcke, PTB
		455.986 240 494 158	26	6×10^{-14}	2000	Udem, NIST
I_2 a_{15} of R(127) 11-5	632.99	473.612 340 492 000	74 000	2×10^{-10}	1983	Jennings, NBS
I_2 a_{16} of R(127) 11-5	632.99	473.612 353 604 800	1200	3×10^{-12}	2000	Ye, JILA
I_2 a_{10} of R(56) 32-0	532.24	563.260 223 480 000	70 000	1×10^{-10}	1995	Jungner, JILA
		563.260 223 514 000	5000	9×10^{-12}	2000	Diddams, JILA
Yb^+	435.51	688.358 979 230 931	6	9×10^{-15}	2001	Stenger, PTB
H 2*S*-8*S*	389.01	770.649 561 581 100	5900	8×10^{-12}	1997	de Bouroir, ENS
Hg^+	281.57	1064.721 609 899 140	10	9×10^{-15}	2000	Udem, NIST
In^+	236.54	1267.402 452 899 920	230	2×10^{-13}	2000	von Zanthier, MPQ
H 1*S*-2*S*	121.57	2466.061 413 187 100	46	2×10^{-14}	2000	Holzwarth, MPQ

Summary of Measurements/Standards

Starting with the first measurement of the frequency of the methane-stabilized HeNe laser in 1972, a new laser frequency measurement was reported every year or so using harmonic frequency chains. These are summarized in Table 1. An indication of the history of frequency measurements of stable laser references is given in Figure 8. The accuracy, in terms of fractional frequency uncertainty, improved from 10^{-10} in 1972 to 10^{-12} in 1999, at which time the femtosecond optical frequency combs came on line and there was a dramatic improvement in the precision as well as the number of measurements that were completed. It is impossible to predict with certainty but some have projected that optical standards can ultimately reach uncertainties as small as 10^{-18}.

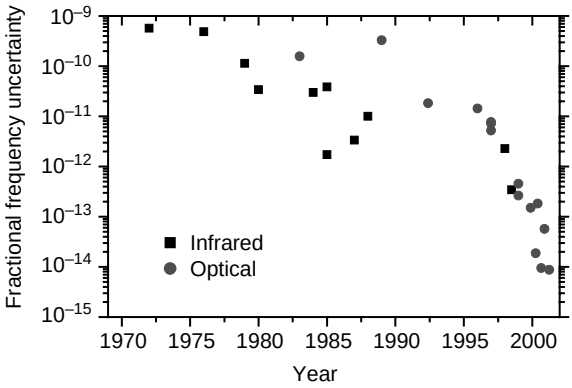

Figure 8 Progress in the accuracy of optical frequency measurements.

Outlook

Three separate technologies have now reached a level of maturity that it is possible to build high-performance optical-frequency standards and clocks. The essential achievements are: laser cooling and trapping of atoms (first proposed by Wineland and Dehmelt, and Hänsch and Schawlow), highly stabilized narrow-linewidth cw lasers, and femtosecond optical frequency combs. Combining these key ingredients, we can construct an optical atomic clock as shown schematically in Figure 9.

Here, for comparison with Figure 1, the cold atoms or single ion provide the narrow atomic resonance, the cw laser serves as the local oscillator to probe the resonance, and the femto-comb serves as the counter. Optical frequency standards of the future are expected to provide orders of magnitude better stability and improved accuracy over the existing atomic frequency standards that now use microwave transitions in atoms. As described in the previous section, frequency combs produced by femtosecond lasers can directly measure the frequency of a stable laser locked to an atomic transition relative to a known microwave frequency standard. This gives fundamental information about the atomic energy levels, and structure, and allows comparisons between different elements. However, to take

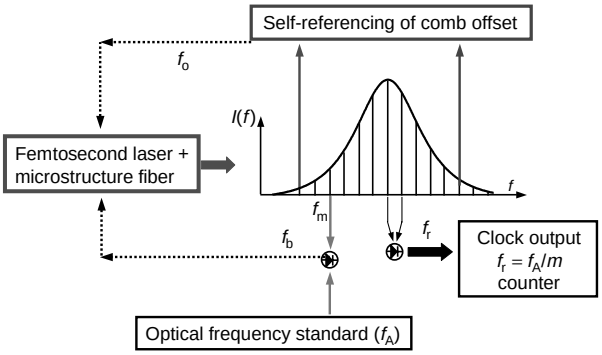

Figure 9 Simplified schematic of an optical clock.

advantage of the high stability of the optical references we run the system as an optical clock (Figure 9) where the stable laser and the femto-comb are locked to the atomic resonance, and the clock output comes as pulses at the repetition rate (e.g., 1 GHz) of the femtosecond mode-locked laser. With a judicious choice of control parameters it is possible to have the pulse repetition frequency, i.e., the clock output, at an exact subharmonic of the optical transition frequency. It is intriguing to note that a portable optical clock could measure time and length at the same time.

See also

Instrumentation: Spectrometers. **Quantum Electrodynamics:** Quantum Theory of the Electromagnetic Field.

Further Reading

Barnett JE (1998) Time's Pendulum. New York: Plenum.

Bergquist JC, Jefferts SR and Wineland DJ (2001) Time measurement at the millennium. *Physics Today* March: 37–42.

Cundiff ST, Ye J and Hall JL (2001) Optical frequency synthesis based on mode-locked lasers. *Reviews of Scientific Instruments* 72: 3749–3771.

Diddams SA, Jones DJ, Ye J, *et al.* (2000) Towards the ultimate control of light. *Optics and Photonics News* October: 16–22.

Diddams SA, Udem T, Bergquist JC, *et al.* (2001) An optical clock based on a single trapped $^{199}Hg^{+}$ ion. *Science* 293: 825–828.

Diels JC and Rudolph W (1996) Ultrashort Laser Pulse Phenomena. San Diego, CA: Academic Press.

Hall JL (2000) Optical frequency measurement: 40 years of technology revolutions. *IEEE Journal of Selected Topics in Quantum Electronics* 6: 1136–1144.

Hall JL and Ye J (2001) A new era of frequency standards and optical frequency measurements. *Optics and Photonics News* February: 45–50.

Hänsch TW and Walther H (1999) Laser spectroscopy and quantum optics. *Reviews of Modern Physics* 71: S242–S252.

Hollberg L, Diddams SA, Curtis EA, Oates CW and Fox RW (2001) Optical frequency standards for clocks of the future. In: The International Society for Optical Engineering. *Harnessing Light: Optical Science and Metrology at NIST.* Bellingham, DC. *Proceedings of SPIE* 4450: 54–59.

Jennings DA, Evenson KE and Knight DJE (1986) Optical frequency measurements. *Proceedings of the IEEE* 74: 168–179.

Jespersen J and Fitz-Randolph J (1977) From Sundials to Atomic Clocks. New York: Dover.

Jones T (2000) Splitting the Second: The Story of Atomic Time. Bristol, UK: Institute of Physics.

Luiten AN (2001) Frequency Measurement and Control: Advanced Techniques and Future Trends. Berlin: Springer-Verlag.

Menoni CS (ed.) (2001) Feature section on optical frequency synthesis: a new tool for precision optical frequency metrology. *IEEE Journal of Quantum Electronics* 37: 1481–1513.

Quinn TJ (1999) Practical realization of the definition of the metre. *Metrologia* 36: 211–244.

Fourier Transform Spectroscopy

T Fromherz, Johannes Kepler Universität, Linz, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Fourier transform spectroscopy (FTS) has emerged as one of the most powerful spectroscopic techniques since the first instruments became commercially available in the late 1960s. With this method, high spectral resolution, high wavenumber accuracy, broad spectral range, high optical throughput, and high signal to noise ratio can be achieved simultaneously. Because of these combined advantages that will be discussed in more detail throughout this article, especially in the far- (10–500 cm^{-1}) and mid-infrared (FIR and MIR) spectral region (500–5000 cm^{-1}), FTS has become the method of choice for the most sensitive spectral investigations. In this article, an overview of the setup, the working principle, and the computation of spectra from the

measured interferograms will be given. Finally, by comparing FTS to dispersive spectroscopic methods, the conditions under which FTS can be exploited to advantage, will be discussed.

Spectrometer Setup

Almost all commercially available Fourier transform (FT) spectrometers are based on a Michelson interferometer. A typical layout of this interferometer is sketched in Figure 1. A collimated beam with spectral density $S(\sigma)$ entering the spectrometer is divided into two beams by a beamsplitter. Here, σ denotes the wavenumber that is related to the wavelength (λ) of the radiation by $\sigma = 1/\lambda$. As shown in Figure 1, one part of the radiation is reflected by the beamsplitter and propagates through the compensator plate to a fixed mirror. The mirror reflects the beam through the compensator plate back onto the beamsplitter. The other part of the radiation beam is transmitted through the beamsplitter and its substrate, and hits a moveable mirror from which it is reflected back onto the beamsplitter. At the beamsplitter, both beams are divided again. As shown in Figure 1, two beams, one propagating to the source, the other propagating to the detector, finally result. Both beams contain a contribution that was reflected by the fixed mirror and a second contribution that was reflected by the moveable mirror. In Figure 1, lines indicate the paths of the beams schematically. In the case of a reflection on the beamsplitter, the line type is changed in Figure 1, whereas for a transmission, the line type remains unchanged. The widths of the lines in Figure 1 indicate the size of the field amplitudes and, therefore, are reduced for each reflection/transmission on the beamsplitter. For clarity, the lines indicating the beams reflected into themselves by the fixed and moveable mirrors in Figure 1 are offset laterally.

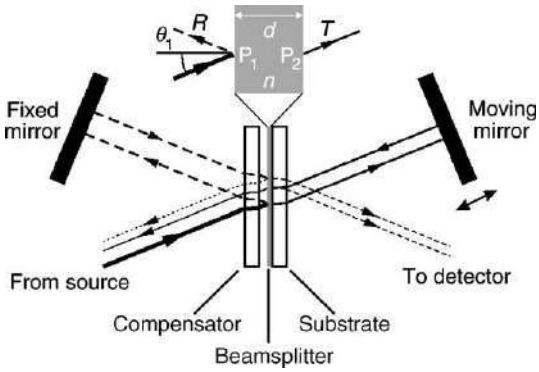

Figure 1 Schematic sketch of the layout of a Michelson interferometer. The paths of the rays are indicated by lines and arrows. At a reflection on the beamsplitter, the line type is changed, whereas for a transmission it remains unchanged. The widths of the lines indicate the size of the field amplitudes. For clarity, the lines indicating the beams reflected into themselves by the fixed and moveable mirrors, are offset laterally. The inset shows a magnified cross-section of the beamsplitter. In the inset, the geometrical details used for calculating the phase difference between R and T are defined.

By changing the position of the moveable mirror along the beam axis, a difference between the optical paths from the beamsplitter to the fixed and moveable mirror is generated. The compensator consists of the same material and has the same thickness as the beamsplitter substrate and, therefore, minimizes the influence of the dispersion of the beamsplitter substrate on the optical path difference.

Fourier Analysis and Interferometry

To calculate the intensities of the combined beams falling on the detector and on the source, we start by considering the phase difference between the reflected and transmitted beams leaving the beamsplitter. For simplicity, we ignore the compensator and beamsplitter substrate, since ideally, they generate the same phases in both arms of the interferometer, and therefore, do not contribute to the phase difference. Thus the beamsplitter is regarded as a freestanding, dielectric slab with thickness d. The complex reflection (R) and transmission (T) coefficients, including the effect of multiple internal reflections within the beamsplitter layer, are given by:

$$\begin{aligned} R &= r(1 - e^{i2\gamma})/(1 - r^2 e^{i2\gamma}) \\ T &= e^{i\gamma}(1 - r^2)/(1 - r^2 e^{i2\gamma}) \end{aligned} \qquad [1]$$

Here, r is the Fresnel reflection coefficient that describes the ratio of the moduli and the phase difference between incoming and reflected electrical field for a electromagnetic wave incident under a angle Θ_1 on the surface of a dielectric with dielectric constant n, and $\gamma = 2\pi\sigma nd\cos(\Theta_2)$ is the phase difference evaluated at P_1 and P_2 in Figure 1 for a single pass of the electromagnetic wave through the dielectric slab under the angle Θ_2 (by Snellius law, $\sin(\Theta_1) = n\sin(\Theta_2)$). Since n is real, so is r and, therefore, the phase difference $\Delta\varphi$ between R and T is calculated from eqn [1] according to:

$$\begin{aligned} \frac{R}{T} &= \frac{|R|}{|T|} e^{i\Delta\varphi_{RT}} = \frac{r}{1 - r^2}(e^{-i\gamma} - e^{i\gamma}) \\ &= -2i\frac{r}{1 - r^2}\sin(\gamma) \end{aligned} \qquad [2]$$

i.e., $\Delta\varphi_{RT} = -\pi/2$. With this result, we are able to calculate the phase difference between the two rays falling on the detector and the two rays falling back

on the source. The geometry considered in the following is sketched in Figure 1. The coefficients R and T connect the phases and magnitudes of reflected and transmitted electric field amplitudes at the points P_1 and P_2. From these points, the radiation beams travel back and forth to the fixed and moveable mirrors, thereby accumulating phases according to: $\varphi_f = 2\pi 2x_f\sigma$ and $\varphi_m = 2\pi 2(x_f + \Delta x)\sigma$, where we have defined $\Delta x = x_m - x_f$ and x_m, x_f denote the distances of moveable and fixed mirrors from the points P_1 and P_2. The two planewaves falling on the detector interact twice with the beamsplitter. Each planewave is once reflected at and once transmitted through the beamsplitter. Therefore, the beamsplitter adds no additional phase difference to the two planewaves and the total phase difference is given by $\Delta\varphi = \varphi_m - \varphi_f$. The situation is different for the two beams falling back on the source. Here, one beam is reflected at both interactions with the beamsplitter, whereas the other is transmitted at both interactions. According to eqn [2], between these two beams an additional phase shift of π results. Quantitatively, the total field amplitude at the detector, neglecting an arbitrary phase, is given by $E_D = E_0RT(e^{i\varphi_m} + e^{i\varphi_f})$, and the corresponding intensity is:

$$\begin{aligned} I_D \propto E_D E_D^* &= 2|E_0|^2|R|^2|T|^2\{1 + \mathrm{Re}[e^{i(\varphi_m - \varphi_f)}]\} \\ &= 2|E_0|^2|R|^2|T|^2\{1 + \cos(4\pi\sigma\Delta x)\} \end{aligned} \quad [3]$$

The sum of the field amplitudes falling back on the source is $E_S = E_0(R^2e^{i\varphi_m} + T^2e^{i\varphi_f})$. For the intensity one gets:

$$I_S \propto E_S E_S^* = |E_0|^2\{|R|^4 + |T|^4 + 2\mathrm{Re}[R^2T^{*2}e^{i(\varphi_m - \varphi_f)}]\} \quad [4]$$

Denoting the complex coefficients R and T by their modulus and their phase $R = |R|e^{i\varphi_R}$, $T = |T|e^{i\varphi_T}$ and using eqn [2], one gets:

$$I_S \propto |E_0|^2\{|R|^4 + |T|^4 + 2|R|^2|T|^2\cos(4\pi\sigma\Delta x + \pi)\} \quad [5]$$

Both intensities, I_D and I_S contain a constant component and a component that varies with $\cos(4\pi\sigma\Delta x)$. The varying parts have the same amplitude $2|E_0|^2|R|^2|T^2|$ but are out of phase by π. Consequently, a maximum in I_D corresponds to a minimum in I_S and the sum of both intensities is independent of the mirror displacement Δx.

In most of the commercially available spectrometers, only I_D is used for spectral measurements. Usually, the detector signal is amplified by an AC coupled electronics. Therefore, in a measurement of $I_D(\Delta x)$, the constant part of the signal is suppressed. Assuming a spectral density $S(\sigma)$ of the source and an ideal detector with a responsivity R_D linear in the intensity and independent of the wavenumber σ, the detector signal $D(\Delta x)$ is given by:

$$D(\Delta x) = R_D\int_0^\infty S(\sigma)\cos(4\pi\sigma\Delta x)\mathrm{d}\sigma \quad [6]$$

The information about the spectrum is contained in $S(\sigma)$ and can be extracted from the interferogram $D(\Delta x)$ by a cosine Fourier transformation. The transformation has to be done numerically and, therefore, it is more convenient to use a complex, fast Fourier transform algorithm. For that purpose, eqn [6] is modified by using the complex expansion of the cosine $\cos(x) = (e^{ix} + e^{-ix})/2$:

$$\begin{aligned} D(\Delta x) &= \frac{R_D}{2}\left(\int_0^\infty S(\sigma)e^{i4\pi\sigma\Delta x}\mathrm{d}\sigma + \int_0^\infty S(\sigma)e^{-i4\pi\sigma\Delta x}\mathrm{d}\sigma\right) \\ &= \frac{R_D}{2}\left(\int_0^\infty S(\sigma)e^{i4\pi\sigma\Delta x}\mathrm{d}\sigma + \int_{-\infty}^0 S(-\sigma)e^{i4\pi\sigma\Delta x}\mathrm{d}\sigma\right) \end{aligned} \quad [7]$$

Extending $S(\sigma)$ to negative wavenumbers according to:

$$S'(\sigma) = \begin{cases} S(\sigma) & \text{for } \sigma \geq 0 \\ S(-\sigma) & \text{for } \sigma < 0 \end{cases} \quad [8]$$

leads to the following expression for the interferogram:

$$D(\Delta x) = \frac{R_D}{2}\int_{-\infty}^\infty S'(\sigma)e^{i4\pi\sigma\Delta x}\mathrm{d}\sigma \quad [9]$$

and the spectral density $S'(s)$ can be obtained by the complex Fourier transform according to:

$$S'(\sigma) = \frac{4}{R_D}\int_{-\infty}^\infty D(\Delta x)e^{-i4\pi\sigma\Delta x}\mathrm{d}(\Delta x) \quad [10]$$

In the following, we will skip the prime and S will refer to the spectral density extended to negative values of σ.

To calculate the spectral density according to eqn [10], one has to use an interferogram extending from $-\infty$ to $+\infty$ measured at infinitely small sampling distances $\mathrm{d}(\Delta x)$. In practise, of course, the interferogram is measured only between finite mirror displacements $-\Delta x_{max}$ and Δx_{max}, at sampling points equally spaced by ξ.

The effect of the finite mirror displacement on the resulting spectral density S can be included into eqn [10] by multiplying the interferogram $D(\Delta x)$ with a function that vanishes for $|\Delta x| > \Delta x_{max}$. As an example, we use the boxcar function

$\theta(\Delta x + \Delta x_{\max})\theta(\Delta x_{\max} - \Delta x)$, where $\theta(x)$ is the Heavyside step function with values 0 and 1 for $x < 0$ and $x > 0$, respectively, and calculate the output spectrum for a spectral density $S_1(\sigma)$ consisting of a single line at wavenumber $\sigma_0 : S_1(\sigma) = (\delta(\sigma - \sigma_0) + \delta(\sigma + \sigma_0))/2$ (note that the spectral density is extended to negative wavenumbers according to the definitions given above). According to eqn [9], the interferogram is $D_1(\Delta x) = (R_D/2)\cos(4\pi\sigma_0\Delta x)$. Multiplying the interferogram with the boxcar function and Fourier transforming the product back using eqn [10], the resulting spectral density S_1^{box} is:

$$S_1^{\text{box}} = 2\Delta x_{\max}\left\{\frac{\sin[4\pi(\sigma_0 - \sigma)\Delta x_{\max}]}{4\pi(\sigma_0 - \sigma)\Delta x_{\max}} + \frac{\sin[4\pi(\sigma_0 + \sigma)\Delta x_{\max}]}{4\pi(\sigma_0 + \sigma)\Delta x_{\max}}\right\} \qquad [11]$$

Therefore, the approximation to a monochromatic spectrum is a sum of $\text{sinc}(z_\pm)$ $(= \sin(z_\pm)/z_\pm)$ functions where $z_\pm = 4\pi(\sigma_0 \pm \sigma)\Delta x_{\max}$. The lineshape function S_1^{box} is plotted around σ_0 in Figure 2 by the full line. Compared to the monochromatic spectrum S_1, the lines of S_1^{box} have a finite linewidth with a fullwidth at half maximum (FWHM) of approximately $0.3/\Delta x_{\max}$. As a general rule in spectroscopy, two spectral lines of equal intensity and linewidth have to be separated at least by their FWHM, in order to be clearly distinguishable. According to this definition, the limit of resolution of a FT spectrometer $\Delta\sigma_{\min}$ is inversely proportional to the maximum mirror displacement and given by $\Delta\sigma_{\min} = 0.3/\Delta x_{\max}$.

A less precise, but more intuitive definition of the resolution refers to the difference of the phases of the interferograms corresponding to two monochromatic lines separated by $\Delta\sigma$. The two lines are defined to be resolvable if the phases of their interferograms differ by more than 2π between the mirror positions $-\Delta x_{\max}$ and $\Delta x_{\max}$, i.e., the beating in the superposition of the interferograms shows at least one period. It is easily seen that, according to this definition $\Delta\sigma_{\min} = 1/(2\Delta x_{\max})$, a result in good agreement but slightly larger than that obtained by the more precise definition given above.

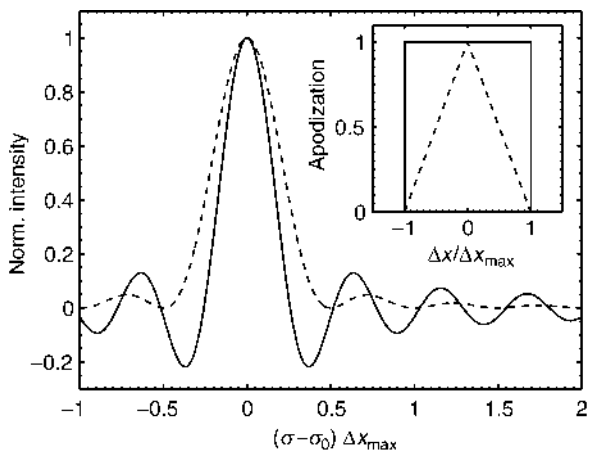

Figure 2 Normalized lineshape functions for boxcar (full line) and triangular apodization (broken line). In the inset, the corresponding apodization functions are shown.

As shown in Figure 2 by the full line, the sinc function has negative side-lobes with magnitudes of up to 22% of the maximum value. For some spectroscopic applications this side lobes are unwanted since they lead to a suppression of smaller peaks separated by approximately the resolution limit of the spectrometer. Thus, in these cases, it appears to be reasonable to multiply the interferogram by a (apodization) function different from the natural boxcar function. Usually, triangular or trapezoidal functions vanishing for $|\Delta x| > \Delta x_{\max}$ are used. By Fourier transformation of these functions, it can be shown that no negative side lobes are generated. However, these functions have a larger (up to ~50%) peak FWHM than the sinc function. For comparison, in Figure 2, also the lineshape of a monochromatic spectrum using a triangular apodization function $A(\Delta x) = \theta(\Delta x + \Delta x_{\max})\theta(\Delta x_{\max} - \Delta x)(1 - |\Delta x/\Delta x_{\max}|)$ is shown by the broken line. In this case, the negative sidelobes are strongly suppressed but the FWHM is increased.

Generally, the output spectrum $S^A(\sigma)$ corresponding to an arbitrary spectral density $S(\sigma)$ and an apodization function $A(\Delta x)$ is given by the convolution $S^A(\sigma) = 2[S \otimes A^{FT}](\sigma)$, where the lineshape function A^{FT} is the Fourier transform of $A(\Delta x)$:

$$A^{FT}(\sigma) = \int_{-\infty}^{\infty} A(\Delta x)e^{-4\pi\sigma\Delta x}d(\Delta x) \qquad [12]$$

and

$$[f \otimes g](t) \equiv \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \qquad [13]$$

Discrete Sampling

In order to discuss the effect of the discrete, equally spaced sampling points with a mutual distance of ξ on the output spectrum, Δx is replaced by $x_n = n\xi$ in eqn [10], where n is an integer and runs from $-N_{\max}$ to $N_{\max}$ and $N_{\max}\xi = \Delta x_{\max}$. Instead of the continuous interferogram $D(\Delta x)$, a discrete interferogram:

$$D_n = \frac{R_D}{2}\int_{-\infty}^{\infty} S(\sigma)e^{i4\pi n\xi\sigma}d\sigma \qquad [14]$$

results. From eqn [14], it is readily seen (by shifting the integration variable) that replacing $S(\sigma)$ with $S[\sigma + m/(2\xi)]$, where m is an integer, does not change the discrete interferogram D_n. Keeping in mind that according to our definition $S(-\sigma) = S(\sigma)$, it is clear that $S[-\sigma + m/(2\xi)]$ also leads to the same interferogram as $S(\sigma)$. Therefore, to be able to transform the interferogram to the spectral density $S(\sigma)$ for a wavenumber band $[0, \sigma_{\max}]$ unambiguously, one has to make sure, by electronically and/or optical filtering, that no wavenumbers outside this range contribute to the interferogram and that the sampling interval ξ is small enough so that $\sigma_{\max} < -\sigma_{\max} + 1/(2\xi)$, i.e., $\sigma_{\max} < 1/(4\xi)$. In Figure 3, a schematic sketch of a spectral density $S(\sigma)$, consisting of two spectral bands shown by the full and dashed lines, is plotted. For the spectral band shown by the full line, the sampling interval ξ is small enough so that in the interval $[0, 1/(4\xi)]$ no ambiguity arises. For the spectral band shown by the dashed line, the interferogram appears to be under-sampled: ξ is too large, and, therefore, the part of the spectral band at negative wavenumbers shifted by $+1/(2\xi)$ falls into the interval $[0, 1/(4\xi)]$ (shown by the dotted line in Figure 3). Thus, for the sampling interval ξ, the interferogram for the sum of spectral bands, shown by the full and the dashed lines in Figure 3, is the same as for the sum of the spectral bands shown by the full and dotted lines. Since the interferograms are the same, the two spectra are indistinguishable at sampling intervals ξ. This ambiguity of spectra for an under-sampled interferogram is called aliasing. Comparing the correct sampling interval $\xi < 1/(4\sigma_{\max})$ with the period of the interferogram D_n corresponding to a spectrum consisting of a single line at $\sigma_{\max}$ ($D_n = R_D \cos(4\pi n\xi\sigma_{\max})$), it follows that the interferogram has to be sampled more than two times per period in order to obtain an unambiguous result (Nyquist Theorem).

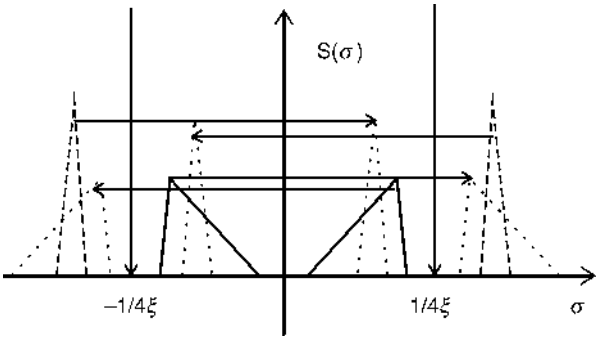

Figure 3 Schematic sketch of aliasing for under-sampled interferograms. For spectral density shown by the full line, the sampling interval ξ is chosen correctly and the $1/(2\xi)$ wavenumber ambiguity indicated by the arrows in the plot does not result in erroneous features in the spectral range $[0, 1/(4\xi)]$. The situation is different for the spectral band sketched by the dashed line: For this band, the sampling interval ξ is too large, and consequently, the $1/(2\xi)$ wavenumber ambiguity maps the line from the negative wavenumber region into the range $[0, 1/(4\xi)]$. The mapped spectral band is indicated by the dotted line.

Mathematically, the spectral density is calculated from the discrete interferogram D_n by a discrete Fourier transformation:

$$S_m = \frac{2}{R_D N} \sum_n^N D_n e^{-2\pi i n m/N} \qquad [15]$$

where N is the number of sampling points ($N = 2N_{\max} + 1$) and the wavenumbers corresponding to S_m are given by $\sigma_m = m/2\xi N$.

In order to obtain a high accuracy in the sampling intervals ξ, usually a reference interferogram of a monochromatic laser line (for example from a HeNe laser) is measured simultaneously with the interferogram of the actual experiment. By counting, for example, the zeros with positive slope in the interferogram of the laser line, uniformly spaced sampling points can be achieved with high accuracy. If, in an experiment many interferograms have to be averaged, the absolute position of the sampling points is important, in order to ensure that only points of equal mirror displacement are averaged. Therefore, in some spectrometers, the interferogram of a broadband source is measured simultaneously with the interferograms of the laser and the experiment and the maximum of the interferogram of the broadband source is taken as the origin for counting the zeros of the laser interferogram.

Phase Correction

Ideally, the interferogram is expected to be symmetrically around the position of equal optical path in the two arms of the interferometer. However, in practise this is seldom the case. Many reasons can lead to an asymmetric interferogram, the most obvious being that none of the discrete sampling points of the interferogram coincides with the mirror position for equal optical path in the two interferometer arms. Let Δx_m be the measured mirror displacement that differs from the real mirror displacement Δx by x_0, i.e., $\Delta x = \Delta x_m + x_0$. In that case the measured interferogram is:

$$D(\Delta x_m) = \frac{R_D}{2} \int_{-\infty}^{\infty} S(\sigma) e^{i[4\pi\sigma(\Delta x_m + x_0)]} d\sigma \qquad [16]$$

and the Fourier transformation, according to eqn [10] with Δx_m as the integration variable, results in

$S(\sigma)e^{i4\pi x_0\sigma}$, that is, in a spectrum with both real and imaginary parts. Also, mirror or compensator misalignment can result in a wavenumber-dependent optical path at a fixed mirror position. Finally, for rapid-scan interferometers, in that the movable mirror is scanned at a constant velocity v, the electronic filtering and amplification of the detector signal will also add different phases, depending on the signal frequency ν that is correlated to the wavenumber σ by $\nu = 2\sigma v$. In general, as long as the phase φ depends only on the wavenumber and not on the mirror displacement, the measured interferogram is:

$$D(\Delta x_m) = \frac{R_D}{2}\int_{-\infty}^{\infty} S(\sigma)e^{i[4\pi\sigma\Delta x_m+\varphi(\sigma)]}d\sigma \qquad [17]$$

and by the Fourier transformation, according to eqn [10], one obtains $S(\sigma)e^{i\varphi(\sigma)}$. There are several methods implemented in the software of commercially available Fourier spectrometers to eliminate the phase from the Fourier transform. In most experimental cases, in that the measured interferograms correspond to spectra with $S(\sigma) \geq 0$ for all $\sigma \in [0, \sigma_{max}]$, these methods usually give the correct results. However, care has to be taken if spectra contain positive and negative components. This can be the case, for example, if the transmission of a sample is periodically modulated by externally applied voltage or a laser light and the detector signal is amplified by a lock-in technique before it is fed into the electronics of the Fourier transform spectrometer. If the modulation increases the sample absorbance in one part of the spectral range and decreases it in a different part, then the measured interferogram corresponds to a spectrum with both negative and positive regions. In the following, first the phase correction methods for spectra with positive values only are discussed before it is argued, why these methods might fail for spectra with positive and negative parts.

For spectra with only positive values, the most simple method to eliminate the unknown phase is to measure the interferogram symmetrically around $\Delta x_m = 0$ and calculate $S(\sigma)e^{i\varphi(\sigma)}$ according to eqn [10]. Including the effect of finite scan length, the output spectrum becomes $S^A = 2[S(\sigma)e^{i\varphi(\sigma)}]\otimes A(\sigma)^{FT}$. Under normal experimental conditions, the phase $\varphi(\sigma)$ does not vary over the peak FWHM of $A(\sigma)^{FT}$ and, therefore, the phase term can be put outside the integral of the convolution and $S^A = 2e^{i\varphi(\sigma)}[S(\sigma)\otimes A(\sigma)^{FT}]$ results. Since $A(\Delta x_m) \geq 0$, and chosen to be symmetric around $\Delta x_m = 0$, the convolution is real and positive and the true spectrum with instrumental resolution $2[S(\sigma)\otimes A(\sigma)^{FT}]$ can be obtained by calculating the modulus of S^A without any knowledge about $\varphi(\sigma)$. However, for several reasons this method for phase correction is not ideal. First of all, since the maximum spectral resolution of a Fourier transform spectrometer is inversely proportional to the maximum mirror displacement, for a given spectrometer layout with a fixed maximal mirror scan length L, the maximal achievable spectral resolution is reduced by a factor of two if scans symmetric around $\Delta x = 0$ have to be performed. Secondly, for signals closer to or smaller than the noise level of the spectrometer, the random noise will be entirely positive by taking the modulus. Hence, for small signals, the noise level will be increased relative to the signal level by taking the modulus.

Therefore, usually the 'Mertz phase correction method' is used. This method is based on the assumption that the phase $\varphi(\sigma)$ is a slowly varying function of σ. In this case, $\varphi(\sigma)$ can be determined with low spectral resolution, i.e., by a short, symmetric scan of length $2L_1$ around $\Delta x = 0$. It can be shown that from this scan the phase can be obtained according to:

$$\varphi(\sigma) = \arctan\left\{\frac{\mathrm{Im}[S(\sigma)e^{i\varphi(\sigma)}\otimes A^{FT}(\sigma)]}{\mathrm{Re}[S(\sigma)e^{i\varphi(\sigma)}\otimes A^{FT}(\sigma)]}\right\} + \pi \text{ (if denominator is negative)} \qquad [18]$$

Here, $S(\sigma)e^{i\varphi(\sigma)}\otimes A(\sigma)^{FT}$ results from the Fourier transformation of the short interferogram measured between $-L_1$ and L_1 and the apodization function $A(\Delta x_m)$ is adjusted to $-L_1$ and L_1. The addition of π for negative denominators is a consequence of the assumption that $S(\sigma)$ is positive everywhere. Under this assumption, a negative denominator in eqn [18] corresponds to a phase angle $|\varphi(\sigma)| > \pi/2$, that is mapped into $[-\pi/2, \pi/2]$ by the arctan-function. To get the correct phase angle, π has to be added to the phase angle, thereby extending the range of $\varphi(\sigma)$ to $[0, 2\pi]$.

With $\varphi(\sigma)$ known, the interferogram has to be recorded only for mirror displacements $\Delta x_m \geq 0$, since the information about the interferogram for $\Delta x_m < 0$ is contained in the phase spectrum and the interferogram for $\Delta x_m \geq 0$. Therefore, with this method, the totally available mirror displacement length can be used for increasing the spectral resolution.

However, for spectra containing positive and negative spectral bands, the Mertz phase correction usually fails, due to the addition of π if the denominator in eqn [18] becomes negative. For spectra with positive and negative bands, the sign of the denominator is determined by both $S(\sigma)$ and $\varphi(\sigma)$. Applying the Mertz algorithm in a negative region of

the spectrum with a vanishing physical phase angle, results in a calculated phase angle of $\varphi = \pi$. As a consequence, the negative spectral band appears as a positive band in the output spectrum calculated according to the Mertz algorithm.

To avoid erroneous phase-flips, the addition of π in eqn [18] for negative denominators can be omitted and the allowed values for $\varphi(\sigma)$ are reduced to $[-\pi/2, \pi/2]$. The corresponding phase correction method is usually referred to as the 'Mertz signed method'. However, it is immediately clear from the discussion above, that in this case any physical phase angle outside the range $[-\pi/2, \pi/2]$ results to an erroneous π-jump of the phase. According to eqn [16], a physical phase shift larger than $\pi/2$ can, for example, be the consequence of a large offset x_0, i.e., of an inaccurate alignment of Δx_m and Δx. At $\Delta x = 0$, all contributions to the interferogram add constructively and the height of the interferogram is proportional to the area under the spectral curve $S(\sigma)$, as can be readily seen by setting $\Delta x = 0$ in eqn [10]. Therefore, for $S(\sigma) > 0$ the position $\Delta x_m = 0$ is usually estimated by the global maximum of the measured interferogram. However, for spectra with both positive and negative bands, in most cases the interferogram does not exhibit a maximum at $\Delta x = 0$ (for positive and negative bands of equal height and width, the interferogram vanishes at $\Delta x = 0$), and estimating $\Delta x_m = 0$ by an extremum of the interferogram results, therefore, in a large offset x_0. Thus, in order to be able to use the 'Mertz signed' algorithm, $\Delta x_m = 0$ has to be determined separately in a calibration experiment in that the condition $S(\sigma) > 0$ is fulfilled. However, even for accurately determined $\Delta x_m = 0$, erroneous phase jumps can appear if lines of different intensity and opposite sign are spaced approximately equal to the resolution used for determining the phase spectrum $\varphi(\sigma)$.

In conclusion, there are no general strategies to completely exclude phase jumps in the Fourier transformation of interferograms corresponding to spectra with positive and negative bands. The experimentalist has to check from case to case if lines with strange shapes appear in the output spectra, and modify the phase correction accordingly. More elaborate strategies for phase correction, that are beyond the scope of this article, are reported in literature.

Example of Experimental Results

To illustrate the features of FTS discussed in the previous paragraphs by a practical example, in Figure 4 the results of a typical measurement in the MIR between 500 cm^{-1} and 5000 cm^{-1} are shown. The experiments were performed with a BRUKER IFS113 spectrometer using a silicon carbide glowbar as a source for the infrared radiation, a potassium bromide beamsplitter, and a liquid-nitrogen cooled mercury cadmium telluride detector. The spectrometer used for the experiments is set up in a vacuum chamber. The two interferograms shown in Figure 4a were measured at two different pressures (broken line: 10 mbar; full line: 300 mbar), i.e., at two different levels of absorption of the atmospheric gas in the interferometer chamber (for clarity, the interferograms are offset vertically). The abscissa axis is broken in Figure 4a and shows portions of the interferograms around zero mirror displacement and around the maximum mirror displacement of 2.5 cm (chosen for 0.2 cm^{-1} spectral resolution). For higher chamber pressures, the increased absorption of the atmospheric gas results in an increased amplitude of the wiggles of the interferogram apparent for large mirror displacements in Figure 4a. The amplitude of these wiggles is approximately three orders of magnitude smaller than the detector signal observed at the center-burst of the interferogram around zero mirror displacement. However, in the corresponding spectra (Figure 4b) calculated from the interferograms after phase correction, the absorption of the atmospheric gas is clearly evident by a strong increase of the dense and narrow absorption lines in the spectral regions 1200 cm^{-1}–2200 cm^{-1}, 2300 cm^{-1}–2400 cm^{-1}, and 3200 cm^{-1}–4000 cm^{-1} (note that in Figure 4b the spectra are offset vertically by 0.1 for clarity). The lines observed between 2300 cm^{-1}–2400 cm^{-1} are due to absorption of the CO_2 molecule and are shown in more detail in Figure 4c from 2335 cm^{-1} to 2345 cm^{-1}. In order to simulate the contribution of the CO_2 lines to the interferogram shown in Figure 4a, two neighboring lines in Figure 4c were fitted to a doublet of Lorentz-absorption lines (indicated by the symbols in Figure 4c). The corresponding interferogram was calculated by Fourier transformation of the Lorentz lines and is shown in Figure 4d. It consists of a rapid oscillating, cosine-like signal (see inset for an expanded plot between 0.6 and 0.602 cm) with an amplitude modulation corresponding to the difference of the resonance wavenumbers of the two Lorentz lines. The ordinate axes of Figures 4a,d have the same arbitrary units and, therefore, the amplitudes of the interferograms shown in these two panels can be directly compared. From Figure 4a, d we conclude that the two CO_2 lines contribute approximately with only one in 10^5 parts to the center-burst of the measured interferogram, i.e., theses lines absorb only a fraction of 10^{-5} of the integrated (over the wavenumbers) intensity of

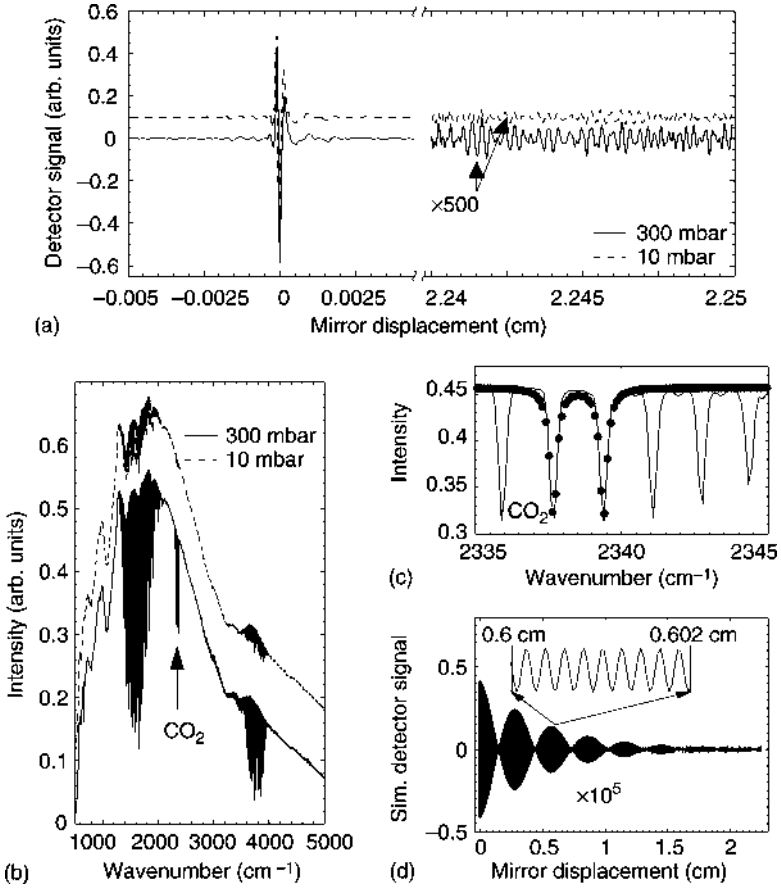

Figure 4 Measurement of the absorption of atmospheric gas for different pressures in the interferometer chamber (broken lines: 10 mbar, full lines: 300 mbar). (a) Interferograms (note the break in the axis of the mirror displacement). For clarity, the interferograms are offset vertically by 0.1 and 0.0002 on the left and right side of the axis break. On the right side of the axis break, the interferograms are amplified by a factor 500. (b) Spectra corresponding to the interferograms shown in (a). The spectra are offset vertically by 0.1 for clarity. (c) Part of the CO_2 absorption lines on an expanded wavenumber axis. The symbols indicate a fit of two Lorentz lines to neighbouring CO_2 absorption lines. The fit is used for calculating the interferogram shown in (d). (d) Simulated interferogram corresponding to the two CO_2 lines shown by the symbols in (c). The arbitrary units of the ordinate axis are the same as those used in (a) but the interferogram is amplified by a factor 10^5. In the inset, the interferogram is shown on an enlarged scale between 0.6 and 0.602 cm mirror displacement.

the incident radiation. Despite this small fraction, they are extremely well resolved with a large signal-to-noise ratio in Figures 4b,c, thus demonstrating the ability of FTS to measure spectra simultaneously with high resolution and over a large spectral range. This ability is one of the advantages (the so-called multiplex advantage) inherent in FTS and will be discussed in more detail in the following paragraph.

Advantage of FTS Over Dispersive Spectrometer

The two main advantages of FTS over dispersive spectroscopic methods, using, for example, a grating for the spectral decomposition of an incoming radiation, are directly connected with the working principle of FT spectrometers; high optical throughput and multiplexed measurement over the whole spectral range of interest. These advantages are discussed in the following, in more detail.

For a fair comparison of the optical throughput, spectrometers with equal spectral resolving power ρ have to be compared. Here, $\rho = \sigma_{max}/\Delta\sigma$, where σ_{max} is the maximum wavenumber analyzed and $\Delta\sigma$ is the spectral resolution. For an ideal Fourier transform spectrometer, the limit for σ_{max} depends on the divergence angle of the beams propagating in the spectrometer. As sketched in Figure 5, the divergence angle α is determined by the radius r_a of the entrance aperture and the focal length f of the collimating lens by $\alpha \approx r_a/f$. This angle causes a phase shift between the rays propagating parallel to and those propagating under the angle of α inclined to the optical axis

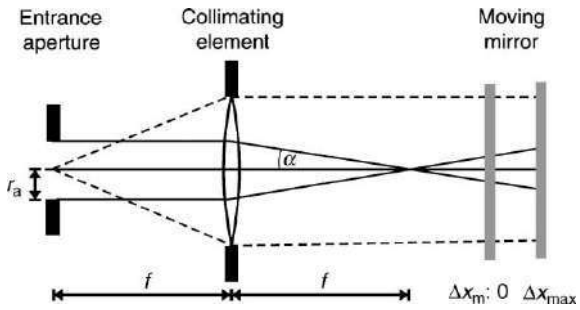

Figure 5 Entrance aperture and collimating element for an interferometer. For simplicity, the beamsplitter and fixed mirror are omitted in the sketch. The extreme rays entering the interferometer are indicated by the full and broken lines. Due to divergence α, a monochromatic light with wavenumber σ_0 produces all periods in the range $[4\pi\sigma_0, 4\pi\sigma_0/\cos(\alpha)]$ in the interferogram.

through the interferometer. The phase difference φ_{div} depends on the mirror displacement according to $\varphi_{div} = 4\pi\sigma\Delta x((1/\cos(\alpha)) - 1)$ and can be approximated for small α by $\varphi_{div} \approx 2\pi\sigma\Delta x\alpha^2$. It can be shown, that as long as φ_{div} does not exceed approximately π, the influence of the divergence on the interferogram can be tolerated, i.e., $\sigma_{max} \approx 1/(2\Delta x_{max}\alpha^2)$. Using $\Delta\sigma \approx 1/(2\Delta x_{max})$ for the resolution, one obtains for the resolving power of a FT spectrometer, $\rho = (f/r_a)^2$.

The optical throughput Θ describes the allowable energy per unit time that the system can let through and is related to area πr_a^2 of the aperture and the solid angle Ω of the collimating (or focusing) optics by $\Theta = \pi r_a^2\Omega$. Using $\Omega = A_{FTS}/f^2$, where A_{FTS} is the clear area of the collimating lens (for a collimating mirror, A_{FTS} is the area of the mirror projected on a plane perpendicular to the beam direction), results in $\Theta_{FTS} = \pi A_{FTS} r_a^2/f^2 = \pi A_{FTS}/\rho$. For a grating spectrometer, $\Theta_{grating} = hA_G/f\rho$. (see respective chapter of this encyclopedia) where h is the height of the slit, f the focal length of the collimating optics, and A_G the projected area of the grating. Generally, h/f is smaller than 1/20 in up-to-date grating spectrometers. Thus, for the same resolving power ρ and similar instrument size, FT spectrometer can offer a more than $\pi f/h \approx 60$ times larger energy gathering capability. Therefore, for measurements with weak signals in that the detector noise is the dominant noise source, the large throughput of a FTS greatly increases the spectral signal to noise ratio that can be achieved for a fixed measuring time.

The other major principal advantage compared to dispersive methods relies on the simultaneous (multiplexed) gathering of information about all spectral bands $\Delta\sigma$ over a broad spectral range from σ_{min} and σ_{max}. For a grating spectrometer using a single detector element, the measurement time T_m available for recording the spectrum in that range has to be distributed between the $M = (\sigma_{max} - \sigma_{min})/\Delta\sigma$ spectral bands so that for a single spectral band the available integration time is T_m/M and the integrated signal increases proportional to T_m/M. If the dominant noise source is the detector noise which is independent of the signal level, then the noise will be proportional to $(T_m/M)^{1/2}$ and, therefore, the signal to noise ratio for the grating spectrometer $(S/N)_G$ is:

$$(S/N)_G \propto (T_m/M)^{1/2} \qquad [19]$$

For a FTS the situation is different, since it detects in the band $[\sigma_{min}, \sigma_{max}]$ all M small bands over the whole measurement time T_m via their contribution to the interferogram. So the integrated signal in a small band $\Delta\sigma$ is proportional to T_m. If the noise is again assumed to be random and independent of the signal level, for the FTS:

$$(S/N)_{FTS} \propto T_m^{1/2} \qquad [20]$$

results. For the same detector, assumed in the grating spectrometer and in the FTS, the proportionality constants are the same in eqns [19] and [20] and, therefore

$$\frac{(S/N)_{FTS}}{(S/N)_G} = M^{1/2} \qquad [21]$$

Equation [21] shows that the multiplex advantage becomes increasingly important for measurements of broad spectral bands with high resolution. In this case, an enhancement of the $(S/N)_{FTS}$ over $(S/N)_G$ by 2–3 orders of magnitude can be achieved. However, it has to be noted that the multiplex advantage of a FTS can be exploited only in cases where the detector noise is the dominant source of noise. In the visible and near infrared spectral region, low noise detectors, that allow single photon detection, are available. With these detectors, the spectral noise is dominated by the photon noise that is proportional to the square root of the intensity. It is evident that in this case the multiplexed detection is not an advantage, since for M spectral bands of width $\Delta\sigma$ contributing to the interferogram at a given mirror position, the noise is enhanced by $M^{1/2}$ compared to the noise that would be measured for a single band (here, it is assumed that all M bands contribute with the same intensity to the interferogram). Consequently, in this case, the noise enhancement by $M^{1/2}$ cancels the beneficial factor $M^{1/2}$ in eqn [21] and with single-detector grating and FT spectrometers, the same S/N ratio is achieved for a given measurement time T_m. In addition, using linear detector arrays with dispersive spectrometers, the multiplex advantage increases the S/N ratio for the dispersive spectrometer. However, detector arrays

and low-noise detectors are mainly available for spectral regions with $\sigma > 6000\ \mathrm{cm}^{-1}$, i.e., in the near-infrared and visible spectral region. In the FIR and MIR region ($\sigma < 6000\ \mathrm{cm}^{-1}$) in that such detectors are absent, the multiplexing inherent in FTS makes it the method of choice if low-noise, broadband, and high-resolution spectroscopic data are required.

List of Units and Nomenclature

Angle of incidence for radiation on beamsplitter	Θ_1
Angle of propagation of radiation in the beamsplitter layer	Θ_2
Beam divergence angle	α
Beam splitter thickness (cm)	d
Complex reflection coefficient	R, r
Complex transmission coefficient	T, t
Detector Signal normalized to detector area (A cm^{-2})	D
Difference between real and measured mirror displacement (cm)	x_0
Dirac delta-function	δ
Discrete interferogram (A cm^{-2})	D_n
Discrete spectral density (W cm^{-1})	S_m
Distance from beamsplitter to fixed mirror (cm)	x_f
Distance from beamsplitter to movable mirror (cm)	x_m
Electric field amplitude of incident radiation (V/m)	E_0
Electric field amplitude of radiation falling back on source (V/m)	E_S
Electric field amplitude of radiation reaching detector (V/m)	E_D
Focal length (cm)	f
Fourier transform of apodization function A (cm)	A^FT
General apodization function	A
Heavyside step function	θ
Intensity falling back on source (W cm^{-2})	I_S
Intensity on detector (W cm^{-2})	I_D
Maximum mirror displacement (cm)	Δx_max
Maximum wavenumber (cm^{-1})	σ_max
Measured mirror displacement (cm)	Δx_m
Measured single line spectral density for a finite scan length (W cm^{-1})	S_1^box
Measured single line spectral density for a finite scan length using the apodization function A (W cm^{-1})	S^A
Measurement time for complete spectrum (s)	T_m
Minimum wavenumber (cm^{-1})	σ_min
Mirror displacement	Δx
Mirror velocity (cm s^{-1})	v
Modulation frequency (s^{-1})	ν
Number of sampling points of the interferogram	N
Number of spectral bands	M
Optical throughput (cm^2)	Θ
Optical throughput for FTS (cm^2)	Θ_FTS
Optical throughput for grating spectrometer (cm^2)	Θ_grating
Phase accumulation due to optical path to fixed mirror	φ_f
Phase accumulation due to optical path to moveable mirror	φ_m
Phase angle	γ
Phase angle	φ
Phase difference	$\Delta\varphi$
Phase difference between beams reflected at and transmitted through beamsplitter	$\Delta\varphi_\mathrm{RT}$
Phase difference due to beam divergence	φ_div
Phase jump for reflection at beamsplitter	φ_R
Phase jump for transmission through beamsplitter	φ_T
Projected area of collimating element of FTS (cm^2)	A_FTS
Projected area of grating (cm^2)	A_G
Radius of entrance aperture (cm)	r_a
Refractive index	n
Responsivity (A W^{-1})	R_D
Sampling interval (cm)	ξ
Signal to noise ratio for FTS	$(\mathrm{S/N})_\mathrm{FTS}$
Signal to noise ratio for grating spectrometer	$(\mathrm{S/N})_\mathrm{G}$
Single line spectral density (W cm^{-1})	S_1
Solid angle	Ω
Spectral bandwidth (cm^{-1})	$\Delta\sigma$
Spectral density (W cm^{-1})	S
Spectral resolving power	ρ
Wave length (μm)	λ
Wave number (cm^{-1})	σ, σ_0
Wave number resolution (cm^{-1})	$\Delta\sigma_\mathrm{min}$

See also

Interferometry: Overview; Phase Measurement Interferometry.

Further Reading

Bell RJ (1972) *Introductory Fourier Transform Spectroscopy*. New York, London: Academic Press.

Born M and Wolf E (1970) *Principles of Optics*. New York: Pergamon Press.

Griffiths PR (1975) *Chemical Infrared Fourier Transform Spectroscopy*. New York, London, Sydney, Toronto: John Wiley & Sons.
Hecht E (1987) *Optics*. Reading, MA: Addison-Wesley.
Mertz L (1965) *Transformations in Optics*. New York, London, Sydney: John Wiley & Sons.
Saptari V (2004) *Fourier-Transform Spectroscopy Instrumentation Engineering*. Bellingham: SPIE Press.

Hadamard Spectroscopy and Imaging

R A DeVerse, Plain Sight Systems, Inc., Kailua-Kona, HI, USA

R M Hammaker, Kansas State University, Manhattan, KS, USA

W G Fateley, Plain Sight Systems, Inc., Manhattan, KS, USA

F B Geshwind and A C Coppi, Plain Sight Systems, Inc., Hamden, CT, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Consider the task of measuring the spectra of a source, the spectra of an array of contiguous sources, or simply the broadband irradiance of an array of sources. These measures are commonly known as spectrometry, spectral imaging, and imaging, respectively. The sensitivity at which spectrometric, spectral imaging, and imaging systems can be operated is often dominated by noise associated with the detection system and the ability of the optics employed to deliver sufficient photonic flux to the detector subsystem to overcome this noise limitation. The ideal case is where the optical system can be fashioned in such a way as to deliver greater signal to the detector in order to maximize the signal-to-noise ratio (SNR) for a given finite source irradiance and detector noise.

Let us consider the number of discrete spectral resolution elements of spectrometers, spatio-spectral resolution elements of spectral imagers, and spatial resolution elements of imaging systems as being the N quantities of interest we would like to measure. We wish to measure these N quantities in time T with the greatest accuracy and precision as determined by the maximum SNR possible given a finite source flux and a fixed detector noise floor. One option might be to use a multi-element detection system (multi-channel detection system) with N detectors, so each detector is measuring one of the N quantities for the full time T in order to maximize the available integration time. However, in practice, these detectors can be problematic or simply not available for many applications when compared to single-element detectors. When a single-element detector is employed in spectrometry, spectral imaging, or imaging, the measurement methodology required becomes N measurements, each for a time T/N, so that each of the N quantities is measured separately or alternatively, one can record N measurements, each for time T/N, for various groups of the N quantities. The measurement procedure for various groups of the N quantities requires an encoding procedure to choose which of the N quantities to include in each of the N measurements and a decoding procedure to recover the values for each of the N individual quantities. When combinations of N quantities are measured, the signal at the single detector element is increased and the procedure is called multiplexing. Multiplexing may be defined as measuring the sum of groups of quantities rather than measuring each quantity separately with the primary goal of increasing SNR. Multiplexing can and does increase SNR when the detection system of the measuring device is operating in the detector noise limit. The detector noise limit is defined as the noise of the detector output that is independent of the amplitude of the signal. Photon-noise-limited detector systems can benefit from multiplexing strategies, however, one must be careful not to realize a multiplex disadvantage when the source of noise is not detector noise limited.

Fourier Transform Multiplexing

In optical spectrometry a common multiplexing method employs the Michelson interferometer in a Fourier transform (FT) spectrometer where, in the best case, as much as half of the spectral energy of the source entering the aperture is incident upon the single detector element for the duration of the measurement. The output of the detector is recorded as an interferogram, which is the detector response as a function of optical retardation (twice the path difference between the arms of the Michelson interferometer). The recorded interferogram is decoded using a Fourier transform algorithm to yield the desired spectrum. FT spectrometers are routinely employed in near-infrared, mid-infrared, and

far-infrared spectrometry. In Raman spectrometry, FT spectrometers are used in the near-infrared spectral region typically using an Nd:YAG laser providing an excitation wavelength of 1064 nm. FT techniques are also employed in nuclear magnetic resonance (NMR) spectrometry and mass spectrometry.

Hadamard Transform Multiplexing

Hadamard transform multiplexing in dispersive optical spectrometry can be accomplished by use of an encoding mask at the focal plane of a dispersive spectrometer, that is operable to select different combinations of spectral resolution elements that are allowed to pass to the single detector element for each of the N required measurements. The specification of which ones of the N quantities are included in each of the N measurements of groups of the N quantities is called a weighing design. The term weighing design originally referred to a procedure for accurately determining the mass of a number of objects by weighing them in groups rather than one at a time. The optimal weighing designs and corresponding optimal encoding masks are specified by Hadamard mathematics. The Hadamard matrix or **H** matrix has all elements of +1 and −1 and provides the optimal weighing design for a two-pan balance or an optical instrument having two detector elements for the +1 and −1 measures. The Hadamard matrices specify which quantities go to each balance pan or detector element.

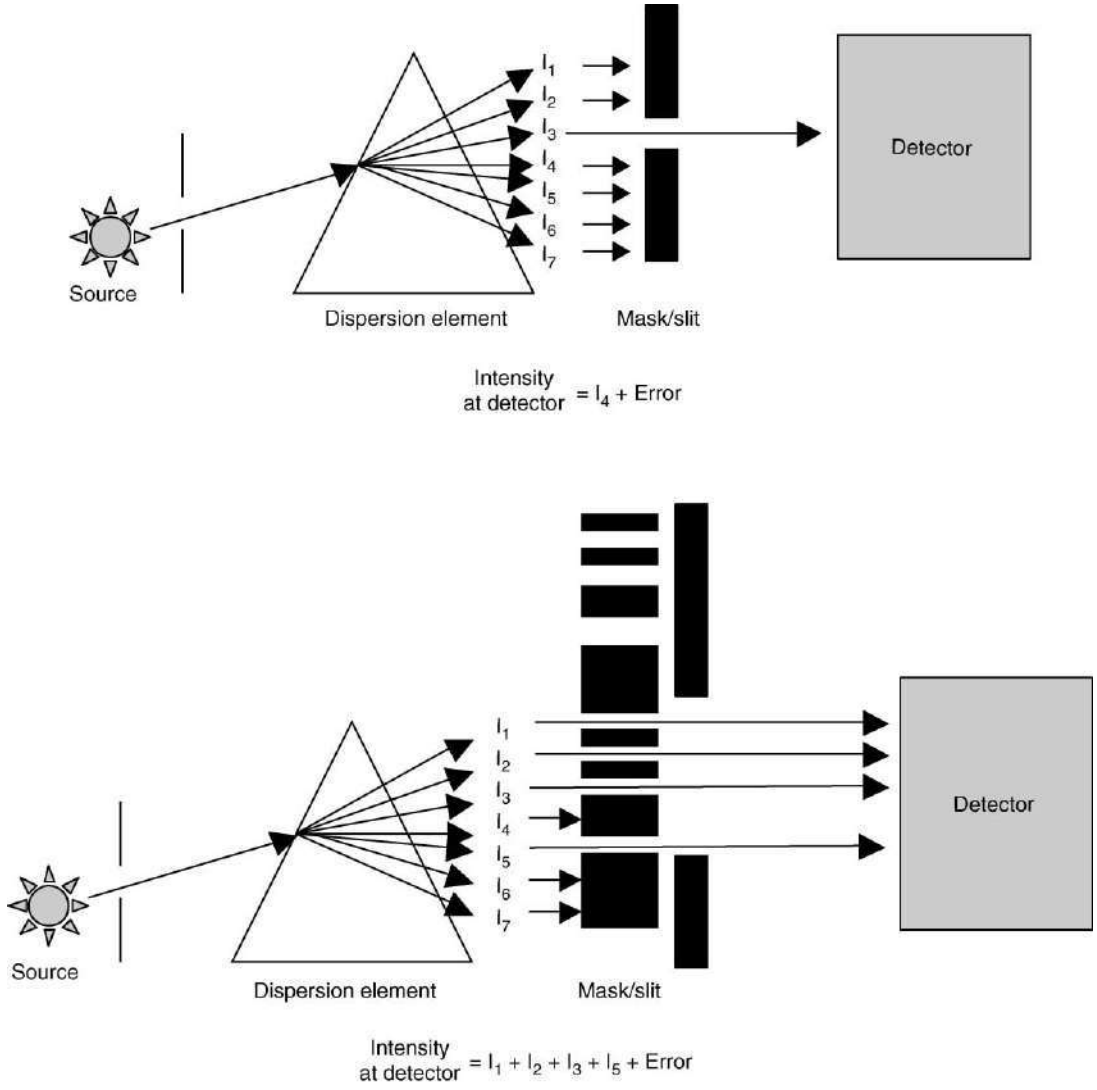

Figure 1 Simple schematic of a conventional and a Hadamard transform spectrometer. The top shows a conventional spectrometer where only one of seven spectral resolution elements ($N = 7$) impinge upon the single element detector for each of N measurements. The bottom shows a Hadamard transform spectrometer where $N = 7$. One more than half of the spectral resolution elements impinge upon the single element detector for each of the N measurements.

One and Two Detector Hadamard Encodement Matrices

In the case of a dispersive spectrometer with a single detector element, the optimal weighing designs and corresponding optimal encoding masks are provided by the simplex matrix or **S** matrix derived from the **H** matrix. The type of **S** matrix that is most convenient for optical applications is the cyclic or left circulant **S** matrix where only the first row needs to be known to generate the complete cyclic **S** matrix of N columns and N rows. Each row is obtained by shifting the previous row by a left-shift registration in which the element in the far left column moves to the far right column of the next row and each of the other $N-1$ elements moves one column to the left. Each of the N rows of the cyclic **S** matrix specifies one of the N encodements and each of the N columns specifies one of the N spectral resolution elements. The elements of the cyclic **S** matrix are 1 and 0 instead of the +1 and −1 found in the **H** matrix implementation. Spectral resolution elements allowed and not allowed to reach the single detector element are given as 1's and 0's, respectively. Each row and each column of the cyclic **S** matrix is a different sequence of $(N+1)/2$ elements of 1's and $(N-1)/2$ elements 0's so that each encodement in the row of the cyclic **S** matrix has a different combination of $(N+1)/2$ spectral resolution elements that are incident upon the single detector element. Just as the positions of the 1's in a row show which spectral resolution elements are included in a particular encodement, the positions of the 1's in a column show which encodements include that particular spectral resolution element. The result is that provided the detector noise is independent of its output signal, the SNR is increased by a factor of $(N+1)/(2\sqrt{N})$, approximately $(\sqrt{N})/2$, when compared to the method of measuring each spectral resolution element separately as in a raster scan performed by a classical scanning dispersive spectrometer.

Multiplexing in dispersive optical spectrometry as introduced in the previous paragraph is called Hadamard transform spectrometry (HTS) (see Figure 1). It is a unique combination of dispersive and multiplexing spectrometries. The choice of transparent or opaque for the elements of a multislit array or Hadamard encoding mask provides an encoding of spectral information that may be decoded into a conventional spectrum by Hadamard mathematics. In a typical dispersive spectrometer, the source energy is collected and separated into its individual spectral resolution elements by a spectral separator and then collected and focused for spatial presentation on a focal plane. Unlike a scanning dispersive spectrometer, which possesses a single exit slit, the Hadamard transform (HT) spectrometer employs a multislit array as a Hadamard encoding mask. This arrangement allows for the simultaneous measurement of a multitude of spectral resolution elements at a single-element detector and produces a multiplexing spectrometer.

Theory of Hadamard Multiplexing

The operation of the HT spectrometer can be described using matrix algebra with column vectors [**D**] and [**I**] and the N-order square matrices $\mathbf{S}_N$, $\mathbf{S}_N^{-1}$, $\mathbf{S}_N^{\mathrm{T}}$, and $\mathbf{T}_N$. [**D**] is a column vector of N rows containing the detector responses for the N encodements. [**I**] is a column vector of N rows containing the N spectral resolution elements to be determined. $\mathbf{S}_N$ is the cyclic **S** matrix of order N that is the basis for construction and design of the Hadamard encoding mask. $\mathbf{S}_N^{-1}$ is the inverse of $\mathbf{S}_N$ and is given by $(2/(N+1))\mathbf{T}_N$. $\mathbf{T}_N$ is a matrix of +1 and −1

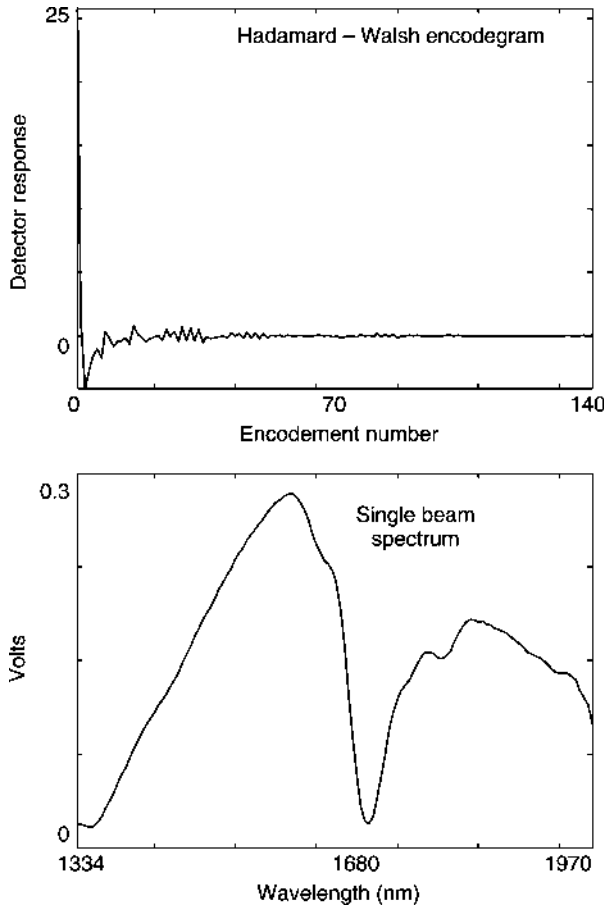

Figure 2 An encodegram (top plot) and the corresponding single-beam spectrum (bottom plot) as a result of decoding the encodegram using Hadamard mathematics.

elements obtained from $\mathbf{S}_N^T$, the transpose of $\mathbf{S}_N$, with $\mathbf{T}_N$ having +1's and −1's where $\mathbf{S}_N^T$ has 1's and 0's, respectively. A cyclic matrix is required to be symmetric so that an interchange of columns and rows can recover the original matrix and $\mathbf{S}_N^T$ then becomes equal to $\mathbf{S}_N$. The encoding of the radiation by the HT spectrometer is given by $[\mathbf{D}] = \mathbf{S}_N[\mathbf{I}]$ and the resulting primary data are a record of detector response versus encodement number. The plot of detector response versus encodement number is called an encodegram (analogous to the interferogram produced by an FT spectrometer). The decoding of the encodegram to recover the conventional spectrum using a Hadamard transform can be described by $[\mathbf{I}] = \mathbf{S}_N^{-1}[\mathbf{D}]$, obtained by left multiplying $[\mathbf{D}] = \mathbf{S}_N[\mathbf{I}]$ by $\mathbf{S}_N^{-1}$. In practice, it is not necessary to invert $\mathbf{S}_N$ because there is a fast Hadamard transform (FHT) algorithm available when N is of the form $N = 2^w - 1$, w being an integer. There are three known methods for generating the first row of cyclic $\mathbf{S}$ matrices of order $N = 2^w - 1$. Consider the example of $N = 7$ for $w = 3$ where the first row of $\mathbf{S}_7$ is 1 1 1 0 1 0 0. Figure 2 shows an encodegram and the result of decoding using Hadamard mathematics. Figure 3 contains the matrices and matrix equations for $N = 7$.

Hadamard Encoding Masks

The key component of any HT spectrometer is the Hadamard encoding mask (Figure 4). Two choices for the Hadamard encoding mask are the moving mechanical mask and the stationary electro-optical mask. The moving mask for N spectral resolution elements requires $2N - 1$ mask elements that are either completely open with a transmittance of 1 or completely closed with a transmittance of 0. The stationary mask for N spectral resolution elements requires only N mask elements that are selected to be either transparent: transmittance $T_t = 1$ or <1 or opaque: transmittance $T_o = 0$ or >0. The resulting SNR is dependent upon $\Delta T = T_t - T_o$, which has its optimal value of unity only for an ideal mask which is $T_t = 1$, $T_o = 0$. The advantage of the moving mask is

$$[\mathbf{D}] = \begin{matrix} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \\ D_6 \\ D_7 \end{matrix} = \begin{matrix} I_1 + I_2 + I_3 + I_5 \\ I_1 + I_2 + I_4 + I_7 \\ I_1 + I_3 + I_6 + I_7 \\ I_2 + I_5 + I_6 + I_7 \\ I_1 + I_4 + I_5 + I_6 \\ I_3 + I_4 + I_5 + I_7 \\ I_2 + I_3 + I_4 + I_6 \end{matrix} \qquad \mathbf{S}_7 = \begin{matrix} 1\,1\,1\,0\,1\,0\,0 \\ 1\,1\,0\,1\,0\,0\,1 \\ 1\,0\,1\,0\,0\,1\,1 \\ 0\,1\,0\,0\,1\,1\,1 \\ 1\,0\,0\,1\,1\,1\,0 \\ 0\,0\,1\,1\,1\,0\,1 \\ 0\,1\,1\,1\,0\,1\,0 \end{matrix} \qquad [\mathbf{I}] = \begin{matrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ I_7 \end{matrix}$$

$$\mathbf{T}_7 = \begin{matrix} 1\ 1\ 1\ -1\ 1\ -1\ -1 \\ 1\ 1\ -1\ 1\ -1\ -1\ 1 \\ 1\ -1\ 1\ -1\ -1\ 1\ 1 \\ -1\ 1\ -1\ -1\ 1\ 1\ 1 \\ 1\ -1\ -1\ 1\ 1\ 1\ -1 \\ -1\ -1\ 1\ 1\ 1\ -1\ 1 \\ -1\ 1\ 1\ 1\ -1\ 1\ -1 \end{matrix} \qquad \mathbf{S}_7^{-1} = (1/4)\mathbf{T}_7 = \begin{matrix} (1/4)(1\ 1\ 1\ -1\ 1\ -1\ -1) \\ (1/4)(1\ 1\ -1\ 1\ -1\ -1\ 1) \\ (1/4)(1\ -1\ 1\ -1\ -1\ 1\ 1) \\ (1/4)(-1\ 1\ -1\ -1\ 1\ 1\ 1) \\ (1/4)(1\ -1\ -1\ 1\ 1\ 1\ -1) \\ (1/4)(-1\ -1\ 1\ 1\ 1\ -1\ 1) \\ (1/4)(-1\ 1\ 1\ 1\ -1\ 1\ -1) \end{matrix}$$

$$[\mathbf{I}] = \begin{matrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ I_7 \end{matrix} = \mathbf{S}_7^{-1}[\mathbf{D}] = \begin{matrix} (1/4)(D_1 + D_2 + D_3 - D_4 + D_5 - D_6 - D_7) \\ (1/4)(D_1 + D_2 - D_3 + D_4 - D_5 - D_6 + D_7) \\ (1/4)(D_1 - D_2 + D_3 - D_4 - D_5 + D_6 + D_7) \\ (1/4)(-D_1 + D_2 - D_3 - D_4 + D_5 + D_6 + D_7) \\ (1/4)(D_1 - D_2 - D_3 + D_4 + D_5 + D_6 - D_7) \\ (1/4)(-D_1 - D_2 + D_3 + D_4 + D_5 - D_6 + D_7) \\ (1/4)(-D_1 + D_2 + D_3 + D_4 - D_5 + D_6 - D_7) \end{matrix}$$

For example: Note $I_1 = (1/4)\,[\,(D_1 + D_2 + D_3 + D_5) - (D_4 + D_6 + D_7)\,]$ so

$$I_1 = (1/4)\,\{\,[\,(I_1 + I_2 + I_3 + I_5) + (I_1 + I_2 + I_4 + I_7) + (I_1 + I_3 + I_6 + I_7) + (I_1 + I_4 + I_5 + I_6)\,] - [\,(I_2 + I_5 + I_6 + I_7) + (I_3 + I_4 + I_5 + I_7) + (I_2 + I_3 + I_4 + I_6)\,]\,\} = (1/4)\,\{\,4I_1\,\} = I_1$$

Figure 3 Matrices and matrix equations (note since $\mathbf{S}_7$ is symmetric $\mathbf{S}_7^T = \mathbf{S}_7$).

Figure 4 A linear encodement of length $N = 43$ generated by using a quadratic residue construction.

no optical transmission problems. However, it suffers the disadvantage of having moving parts and position repeatability problems. For the case of the stationary mask, it has the advantage that it has no moving parts and no positioning repeatability problems. The disadvantage of the stationary mask is the optical transmission problem where $T_t < 1$ and $T_o > 0$. The moving Hadamard encoding mask of $2N - 1$ mask elements is changed to the next encodement by translation of the Hadamard encoding mask to the left past the aperture by a single mask element. For $N = 7$ and $2N - 1 = 13$, the assignment of 1 and 0 to positions 1 through 13 is [1 1 1 0 1 0 0 1 1 1 0 1 0]. For example, the first encodement corresponding to the first row of $\mathbf{S}_7$ in Figure 3 is from positions 1 through 7 as [1 1 1 0 1 0 0] and the fourth encodement corresponding to the fourth row of $\mathbf{S}_7$ in Figure 3 is from positions 4 through 10 as [0 1 0 0 1 1 1].

2D Encoding Masks for Multiplexing in Spectral Imaging and Imaging

If the conventional one-dimensional (1D) Hadamard encoding mask is folded in some manner, the result is the generation of a two-dimensional (2D) Hadamard encoding mask (Figure 5). Similar to the 1D Hadamard encoding masks that are used for spectral mutiplexing, 2D Hadamard encoding masks can be used for the spatial multiplexing done in Hadamard transform spectral imaging (HTSI) and Hadamard transform imaging (HTI). It is important to note that spectrometry as HTS, spectral imaging as HTSI, and imaging as HTI all realize the same theoretical SNR improvement based on the number of spectral, spatio-spectral, or spatial resolution elements (N), respectively. Consider the folding of a 1D Hadamard encoding mask for $N = 15$ into a 2D Hadamard encoding mask of three rows and five columns (Figure 6). A cyclic $\mathbf{S}_{15}$ has a first row of [0 0 0 1 0 0 1 1 0 1 × 0 1 1 1 1] so the 2D Hadamard encoding mask has a first row of [0 0 0 1 0], a second row of [0 1 1 0 1], and a third row of [0 1 1 1 1] from columns 1 through 5, 6 through 10 and 11 through 15, respectively, from $\mathbf{S}_{15}$.

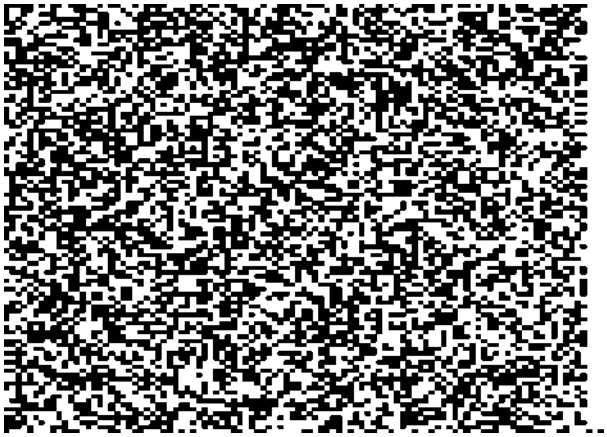

Figure 5 A two dimensional encodement of length $N = 10{,}007$ in a 100×100 format generated by quadratic residue construction. The remaining 7 elements are shown as the last 7 elements in the bottom row on the right. In practice these 'extra' data points can be masked off and used as a measure of system noise characteristics.

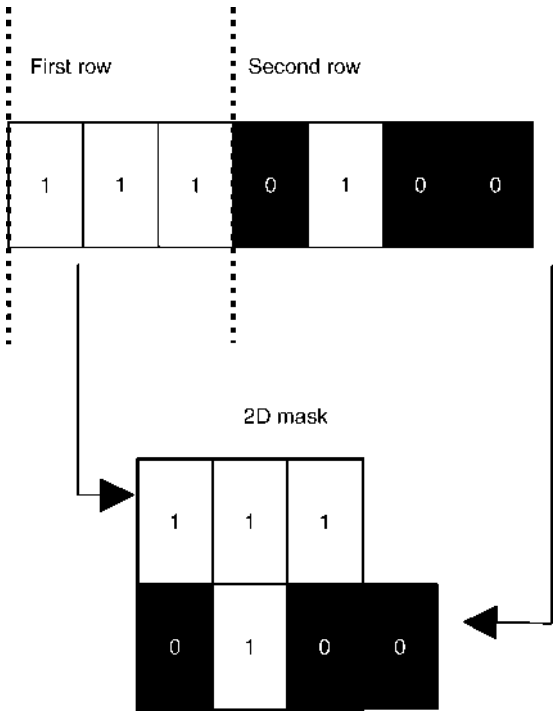

Figure 6 Spectrometry to imagery: The folding of a ID encoding mask that can be used for spectrometry is used to generate a 2D encoding mask that can be used for imaging. Note that using a symmetric aperture window where only 3×2 elements are used, the mask element that is not included in the windowed aperture can be used as a measure of the dark noise of the system, however, a more conventional encodement matrix might be one that can be divided into an integer number of rows and columns such as $N = 15$ where one can construct a 5×3 2D encodement mask for imagery and use a symmetric aperture that uses the entire length of the encodement matrix.

Hadamard Mathematics

Consider a general multiplexing problem where one seeks to measure N quantities, $X = \{x_0, x_1, \ldots, x_{N-1}\}$, in the presence of noise, and wishes to determine a vector function $F(X)$, such that the measurement of $F(X)$ yields the best estimate of X. Conservation of energy and superposition principles dictate a restriction to linear functions. The principle of rank suggests that we restrict our measurements to N-dimensional linear combinations. Determining the best linear combination is the challenge. In the case of constant error for each measurement, an answer is provided by Hotelling's theorem, which proves Hadamard matrices are optimal for such measurements. Hadamard matrices uniquely achieve the optimal bound for root-mean-square measurement error. A Hadamard matrix is a matrix $\mathbf{H}$ for which $\mathbf{H}_{i,j} = \pm 1$, and $\mathbf{H}^{-1} = \frac{1}{N}\mathbf{H}^{\mathrm{T}}$. The definitions of matrix inverse, transpose, and multiplication dictate that the rows of the matrix $\mathbf{H}$ form an orthogonal basis with each basis element having norm $\sqrt{N}$. The norms of the rows of $\mathbf{H}$ follow from the entries of the matrix that are ± 1, and place no additional restriction on $\mathbf{H}$. The columns of $\mathbf{H}$ are also an orthogonal basis with these same norms. The rows (columns) of the matrix $\frac{1}{\sqrt{N}}\mathbf{H}$ form an orthonormal basis. From these few statements and definitions, one can already derive estimates of the multiplex advantage. For example, one seeks to measure X and in one case measures $Y_1 = X + \varepsilon$, where ε is a random noise or error that has a Gaussian distribution with a mean of 0 and variance σ^2. In another case, suppose that one measures $Y_2 = HX + \varepsilon$. In this second case, one would estimate

$$X_2 = \frac{1}{N}\mathbf{H}^{\mathrm{T}}(\mathbf{H}X + \varepsilon) = X + \left(\frac{1}{\sqrt{N}}\right)^2 \mathbf{H}^{\mathrm{T}}\varepsilon$$
$$= X + \frac{1}{\sqrt{N}}\varepsilon_1$$

where ε_1 is random noise with the same characteristics as ε. It follows that the measurement error in the second case has a standard deviation σ that is less than in the former case, with an improvement by the factor $\frac{1}{\sqrt{N}}$.

Hadamard Computations

The ± 1 Hadamard matrices can be applied to two detector optical systems, and 0,1 matrices can be applied to single detector systems. In a two detector system $\mathbf{H}^{\mathrm{T}}$ is proportional to $\mathbf{H}^{-1}$ and one can invert these transforms. Suppose that one has chosen a particular Hadamard matrix $\mathbf{H}$, with which to conduct a multiplexed measurement. One can show that multiplication of any row or any column by -1 does not affect the efficacy of $\mathbf{H}$ in improving SNR. Hence one may assume that the first row and first column of $\mathbf{H}$ are all $+1$. From $\mathbf{H}$, we can form a new $(N-1)\times(N-1)$ matrix $\mathbf{S}$ by deleting the first row and column of $\mathbf{H}$, and then changing each 1 to a 0, and each -1 to a 1. This new matrix $\mathbf{S}$, an $\mathbf{S}$-matrix, is the one corresponding to $\mathbf{H}$ that is to be used for single-element detector optical systems. It is known that $\mathbf{S}$ matrices are optimal for 0,1 matrices just as Hadamard matrices are optimal for ± 1 matrices, with signal to noise improvement by a factor $\frac{\sqrt{N-1}}{2}$, where here N is the dimension of $\mathbf{H}$.

Letting $\mathbf{I}$ be the identity matrix, and $\mathbf{J}$ be the matrix whose entries are all 1, it follows that $\mathbf{S}^{\mathrm{T}}\mathbf{S} = \frac{N}{4}(\mathbf{I} + \mathbf{J})$, where N is the dimension of $\mathbf{H}$. From this we can ascertain that a multiplexed system using the matrix $\mathbf{S}$, multiplied by the encoded parameters of $\mathbf{S}^{\mathrm{T}}$ recovers the original signal to within the addition of an offset. In practice this is what is done to recover the original X. In certain cases, there is also a fast algorithm for computing $\mathbf{S}^{\mathrm{T}}Y$, without ever explicitly computing $\mathbf{S}^{\mathrm{T}}$.

Constructions for Hadamard Matrices

It is known that if $\mathbf{H}$ is an $N \times N$ Hadamard matrix, then N must be 1, 2, or a multiple of 4 and so nontrivial $\mathbf{S}$-matrices must be of size one less than a multiple of 4. It is conjectured that Hadamard matrices exist for each N which is a multiple of 4. Some constructions of Hadamard matrices are enumerated below.

Walsh–Hadamard Construction

A Hadamard matrix of size $N = 2^K$ can be constructed by sampling the first N Walsh functions at $0, 1/N, 2/N, 3/N, \ldots, (N-1)/N$, where the Walsh functions are defined by $W_0(x) = 1$, for $0 \le x < 1$, $W_0(x) = 0$ otherwise, and the recurrence relations, for each non-negative integer n:

$$\begin{aligned} W_{2n}(x) &= W_n(2x) + W_n(2x-1), \text{ and} \\ W_{2n+1}(x) &= W_n(2x) - W_n(2x-1) \end{aligned} \qquad [1]$$

Tensor Product Construction

Walsh is a special case of iterating this construction, starting with the matrix $\mathbf{H}_1 = [1]$. In general, if $\mathbf{H}_N$ is

a Hadamard matrix of size $N \times N$, then:

$$\mathbf{H}_{2N} = \begin{bmatrix} H_N & H_N \\ H_N & -H_N \end{bmatrix} \qquad [2]$$

is a Hadamard matrix of size $2N \times 2N$.

Quadratic Residue Construction

This construction is an example of a cyclic **S**-matrix: one for which each row is a cyclic shift of the previous row. Given a prime number $p = 4n + 3$, define, for $i = 0, \ldots, p - 1$, the values $\mathbf{S}_{1,i} = 1$ if there is some integer k, such that the remainder of dividing k^2 by p is i, and $\mathbf{S}_{1,i} = 0$ otherwise. Taking the cyclic matrix with the first row given by $\mathbf{S}_{1,i}$, yields a Hadamard **S**-matrix.

Other constructions are possible. Maximal length shift register codes, and a construction involving quadratic residues of twin primes are two examples.

Fast Hadamard Transforms

The fast Walsh transform can be deduced from eqn [1]. The basic idea is that, in eqn [1], the function W_n is specified by $\log_2 n$ recurrence relations. For this reason, a Walsh transform (or its transpose) can be computed in $O(n \log_2 n)$ steps. The algorithm goes as follows, for a sequence of length $N = 2^n$:

(i) begin with $Y_{0,0}, Y_{0,1}, \ldots, Y_{0,N-1} = Y_0, Y_1, \ldots, Y_{N-1}$
(ii) for $i = 1, \ldots, n - 1$, let $Y_{i,k} = Y_{i-1,2k} + Y_{i-1,2k+1}$, and $Y_{i,n/2+k} = Y_{i-1,2k} - Y_{i-1,2k+1}$, for $k = 0, \ldots, N/2$
(iii) output $X_i = Y_{n-1,i}$

Then the resulting X_i is the inverse Walsh transform of Y_i.

Note that cyclic **S** matrices that come from maximal length shift registers are known to be permutations of the Wzalsh–Hadamard **S** matrices, and hence there is also a fast algorithm for inverting these, albeit involving some tedious permutations.

Optics for Hadamard Encoded Apertures

Hadamard optical systems utilize spatially encoded apertures that can be employed either at the entrance aperture of the optical system, the exit aperture or both. They have the common attribute that they encode the available aperture spatially where the spatial resolution elements that make up the encodement dictate the spectral, spatio-spectral, or spatial resolution elements that propagate through the optical system, including diffractive optical elements and on to the sensor or exit aperture. These masks have some spatial extent that places special requirements on the optics of the system. As the encodement mask grows either by longer length encodements with fixed subapertures or as the subaperture dimension grows for a fixed encodement length, the spatial resolution elements making up the subapertures in the encodement mask depart from the optical axis. When the resolution elements depart from the optical axis or paraxial condition, it becomes important to employ optics that can image the off-axis resolution elements without inducing excessive aberrations that degrade the performance and cripple the advantages gained by HT multiplexing.

Typically the optical path for conventional monochromators begins with a source that is focused onto an aperture plane that has a large aspect ratio aperture know as a slit. This slit is often very small in extent in the dispersion plane compared to the other extent in the spatial plane. However, it is not required that this aspect ratio is large. If the aspect ratio is close to 1 then simple spherical optics can be employed that perform well as long as the departure from the optical axis is kept to a minimum. However, most monochromators have a large aspect ratio in order to increase the opportunity to maximize throughput, and detectors must be able to 'see' the large extent of the slit aperture. The light entering the slit aperture is then dispersed and focused onto an exit slit aperture. Monochromators are only required to perform well on the optical axis and do not typically employ optics that can manage rays that depart from the optical axis in the plane of dispersion as required by HT multiplexing instruments. To employ encoding techniques the optical system is required to utilize optical performance attributes normally found only in imaging and spectral imaging systems. This requirement is driven by the extent of the encoding mask. The extent of the encoding mask is governed by the diffraction limit of the wavelengths within the bandpass, the encodement length N, and the attributes of the optical system.

In a conventional dispersive spectrometer the radiation from a source is collected and separated into its individual spectral resolution elements by a spectral separator such as a diffraction grating or prism and then is collected and focused for spatial

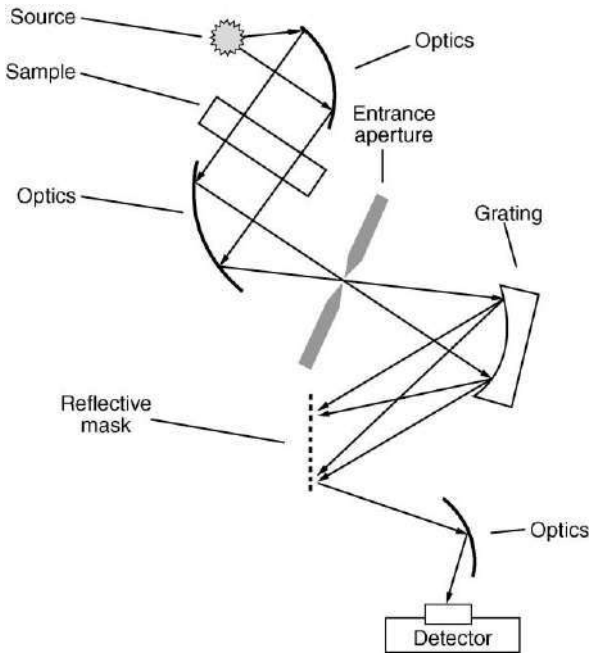

Figure 7 Example implementation of a Hadamard transform spectrometer with reflective optics. Light from a source passes through a sample, is transferred into an entrance aperture, bounced off of a concave grating, and then a reflective mask, and finally transferred to a detector.

presentation on a focal plane. The dispersive spectrometer uses a single exit slit to select one spectral resolution element at any given time, out of N possible spectral resolution elements for measurement by the detector. The Hadamard transform spectrometer (HTS, Figure 7) uses an array of slits (i.e., a mask) at the focal plane to select one more than half, $(N+1)/2$, of the spectral resolution elements at the focal plane for measurement by the detection system. The optical challenge to effect an HT multiplexing spectrometer is to collect all of the spatially distributed individual bandpass images of the entrance slit and transfer them to as small a detector as possible. One desires to keep the area of the detector at a minimum as the noise of many detectors increases with the square of the area. If the optics are able to illuminate a single detector element with all of the available light impinging upon the focal plane containing the spatially distributed images of the slit for each of the N bandpass resolution elements, a multitude of spectral resolution elements can be measured simultaneously using a single detector element. This arrangement results in a multiplexing spectrometer. The recovery of N spectral resolution elements requires measuring the detector response for N different encodements of $(N+1)/2$ open mask elements. The raw data are recorded as the detector response versus encodement number and is called an encodegram. Hadamard transformation of the encodegram yields the spectrum.

The History of Applied Hadamard Multiplexing

The Hadamard transform instruments developed in the 1960s and 1970s employed moving masks. Significant problems such as misalignment and jamming associated with a moving mask led to a reputation of poor reliability and contributed to a dormant period in the development of HTS and HTI. Interest was rekindled in the 1980s using stationary Hadamard encoding masks based on liquid crystal (LC) technology. The first generation 1D stationary Hadamard encoding mask was a cholesteric LC with $N = 127$ mask elements and used polarization as its operating phenomenon. Two parallel polarizers and rotation or lack of rotation of the polarized radiation generated the opaque and transparent states, respectively. The second generation 1D stationary Hadamard encoding mask was fabricated using a polymer dispersed liquid crystal (PDLC) material with $N = 255$ mask elements and used light scattering as its operating phenomenon. The PDLC contained LC droplets dispersed in a polymer matrix whose index of refraction matched the index of refraction in one direction in the birefringent LC droplet. Alignment of the LC droplets optical axis under an applied voltage removed discontinuities in index of refraction at the polymer matrix/LC interface to generate a good transparent state while random orientation of LC droplets in the polymer matrix generated the opaque state from light scattering by the discontinuities in index of refraction at the polymer matrix/LC droplet interface. A 2D stationary Hadamard encoding mask was also based on LC technology. A ferro-electric liquid crystal positioned between a pair of polarizers with perpendicular orientation operated as an electro-optic half-wave plate when a +value of applied voltage rotated the plane of polarization by 90 degrees to produce the transparent state and a −value of applied voltage left the plane of polarization unaltered to produce the opaque state.

Development based on stationary Hadamard encoding masks continued in the 1990s and a 2D moving Hadamard encoding mask was also fabricated and used to perform imaging in the near-infrared and mid-infrared spectral regions. Note that the mid-infrared spectral region is not generally accessible via Hadamard encoding masks based on

LC technology, since any LC material is expected to have strong absorption bands in the mid-infrared spectral region. From the mid-1990s to the present, the stationary Hadamard encoding mask of choice for the visible and near-infrared spectral regions has been the digital micromirror device (DMD), a device based on micro-optoelectromechanical systems (MOEMS) technology and developed by Texas Instruments for projector display applications. One DMD format incorporates 508 800 micromirrors in an 848 column by 600 row array that is 14.4 mm wide by 10.2 mm high. Each individual micromirror is 16 μm square and adjacent micromirrors are separated by a 1 μm gap. The micromirrors are individually addressable and rotate by +10 or −10 degrees about the diagonal axis to produce binary 'on' and 'off' states. The on state has T_t determined by the mirror reflectivity and approaches 1 while the off state approaches $T_o = 0$. The ideal condition of on and off is not realized due to diffraction of the light off of the small and periodic features of the micromirror device. The DMD is an array of spatial resolution elements that may be selected as groups of super-resolution elements or as individual resolution elements consisting of a single micromirror. The DMD resolution elements are realized as spectral resolution elements in the spectrometer with the columns attributed to the frequency or wavelength dimension and the rows attributed to the slit height dimension. The DMD resolution elements are utilized as spatio-spectral resolution elements in the imaging spectrograph with the columns as the frequency or wavelength dimension and the rows as a vertical spatial dimension with the horizontal spatial dimension being accessed, if desired, by translating the sample relative to the imaging spectrograph. The DMD resolution elements are spatial resolution elements in the imager with the columns for the horizontal dimension and the rows for the vertical dimension and the frequency or wavelength dimension provided by other instrumentation. If a photo-acoustic detection system is present then the depth dimension of the sample may also be accessed by changing the modulation frequency used in the photo-acoustic detection system.

Some important features of HTS to keep in mind are:

(i) it is a dispersive technique using a spectral separator;
(ii) it is a multiplexing technique using a single-element detector;
(iii) it uses a Hadamard encoding mask (multislit array) in a focal plane;
(iv) it sends one more than half the resolution elements to the single-element detector in each encodement;
(v) it uses a number of encodements equal to the number of resolution elements desired and the number of mask elements (pixels) in the stationary Hadamard encoding mask (a moving Hadamard encoding mask has $2N - 1$ mask elements);
(vi) it has each encodement containing a different combination of one more than half the resolution elements;
(vii) it has as its primary data the encodegram, a record of detector response versus encodement number; and
(viii) it uses a FHT of the encodegram to decode the encodegram and generate the spectrum or image.

Note that all of these, except for item (i), also apply to HTSI and HTI. The DMD promises to be the best Hadamard encoding mask yet developed for the visible and near-infrared spectral regions. However, its potential applications in spectrometry and imaging are by no means limited to HT techniques since the information corresponding to any micromirror in the DMD may be included in or excluded from any measurement as desired by the investigator. An instrument with no moving parts, other than the micromirrors in the DMD, promises to provide a compact and robust instrument for operation in potentially hostile environments ranging from process control to outer space. It is our belief that the combination of a DMD with a single-element detector may provide an important advance in spectroscopic instrumentation and that instrumentation based on the DMD may lead to a host of environmental, industrial, medical, and military applications.

See also

Fourier Optics. **Imaging:** Multiplex Imaging. **Optical Communication Systems:** Optical Time Division Multiplexing. **Spectroscopy:** Fourier Transform Spectroscopy.

Further Reading

Bellamy MK, Mortensen AN, Hammaker RM and Fateley WG (1997) Chemical mapping in the mid- and near-IR spectral regions by Hadamard transform/FT-IR spectrometry. *Applied Spectroscopy* 51: 477–486.

Fateley WG, Hammaker RM, Paukstelis JV, *et al.* (1993) Application of a two-dimensional Hadamard encoding

mask for the imaging of thin-layer chromatography plates by laser-induced fluorescence or surface-enhanced Raman scattering and for use with a photoacoustic detector to generate three-dimensional photoacoustic images. *Applied Spectroscopy* 47: 1464–1470.

Fateley WG, Hammaker RM, DeVerse RA, Coifman RR and Geshwind FB (2002) The other spectroscopy: demonstration of a new de-dispersion imaging spectrograph. *Vibrational Spectroscopy* 29: 163–170.

Hammaker RM, Bohlke AP, Freeman RD, *et al.* (1996) Hadamard transform Raman spectrometry. In: Laserna JJ (ed.) *Modern Techniques in Raman Spectroscopy*, pp. 143–226. Chichester: Wiley.

Hammaker RM, DeVerse RA, Asunskis DJ and Fateley WG (2002) Hadamard transform near-infrared spectrometers. In: Chalmers JM and Griffiths PR (eds) *Handbook of Vibrational Spectroscopy*, vol. 1, pp. 453–460. Chichester: Wiley.

Harwit M and Sloane NJA (1979) *Hadamard Transform Optics*. New York: Academic Press.

Nonlinear Laser Spectroscopy

P Ewart, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction: Linear and Nonlinear Spectroscopy

Light propagates in transparent or weakly absorbing media by inducing the atoms or molecules in the material to oscillate in response to the electromagnetic field of the light. The induced dipole oscillations set up a polarization field **P** which radiates the wave in the propagation direction. The induced polarization **P** is proportional to the amplitude of the electric field **E** in the incident light:

$$\mathbf{P} = \varepsilon_0 \chi^{(1)}(\omega)\mathbf{E} \qquad [1]$$

where ε_0 is the permittivity of free space and $\chi^{(1)}(\omega)$ is the first-order, frequency-dependent, susceptibility of the material. This interaction of the light with the medium generally results in some absorption of energy from the field and a shift in the phase of the propagating light. Both the phase delay and the absorption depend on the frequency ω (and wavelength λ) of the light. The phase shifting is responsible for dispersion – the variation, as a function of frequency, of the speed (phase velocity) of light in the medium. The dispersion and absorption effects are described by the real and imaginary parts, respectively, of the complex susceptibility:

$$\chi(\omega) = \chi_{\mathrm{Re}}(\omega) + \mathrm{i}\chi_{\mathrm{Im}}(\omega) \qquad [2]$$

This susceptibility parameter describes the optical properties of the medium. Linear spectroscopy uses the frequency-dependent property of the $\chi(\omega)$ to create spectra by absorption or dispersion of specific frequencies that are resonant with transitions between internal energy states of the molecules. The electric field in the incident light is considered to be weak when it leaves the atomic and molecular properties essentially unaffected. Nonlinear optical spectroscopy, however, is based on the ability of intense laser light to modify the optical properties of the medium through which it propagates. These effects are also frequency dependent and so can be used to generate spectra. When the strength of the electric field becomes comparable with the electric fields within the atoms or molecules (typically $\sim 10^{11}\ \mathrm{V\,m^{-1}}$) then the induced polarization becomes nonlinear and is described by a power series expansion:

$$\mathbf{P} = \varepsilon_0[\chi^{(1)}(\omega)\mathbf{E}_1 + \chi^{(2)}(\omega)\mathbf{E}_1\mathbf{E}_2 + \chi^{(3)}(\omega)\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3 + \text{higher-order terms}] \qquad [3]$$

$\chi^{(2)}(\omega)$ and $\chi^{(3)}(\omega)$ are the second- and third-order susceptibilities respectively. The real and imaginary parts of these susceptibilities (and of the higher-order terms) describe observable effects arising from absorption and dispersion. The associated interactions involve two and three photon processes, respectively. In general, nonlinear interactions involve multiphoton processes and will be described by the appropriate term in the expansion of the induced polarization.

Broadly speaking, nonlinear spectroscopy is based on detection of these multiphoton interactions by either incoherent or coherent processes. Incoherent processes include fluorescence and ionization where, following the multiphoton interaction, photons or electrons respectively are emitted spontaneously in all directions. Coherent processes, on the other hand, arise from the coupling of the input driving fields via the nonlinear polarization of the medium.

The waveform of the polarization oscillations is a distorted form of the driving field. Fourier analysis of these nonlinear oscillations would reveal extra frequencies not present in the incident field. When a suitably phased array of dipoles is established, the nonlinear optical process is detected by emission of coherent directional light. The efficiency with which this coherent signal beam is generated is crucially dependent upon the correct phasing of the radiating molecules. This process is known as phase matching and is achieved by geometrical arrangements of the input beams or by modifying the refractive index to counteract dispersion between the generating polarization wave and the radiated signal wave.

Apart from the general division of nonlinear spectroscopy according to whether the detection uses incoherent or coherent emission, the techniques are characterized as frequency or time domain methods, according to how the signal is generated and detected. In most cases the signals recorded in either domain may be related by a Fourier transformation to an equivalent signal in the other.

The advantages of nonlinear spectroscopic techniques lie in their ability to provide high spectral, temporal, and spatial resolution. In particular, it is possible, using a variety of nonlinear optical processes, to eliminate inhomogeneous broadening of the spectral lines recorded. Second, when coherent signal beams are generated they can be readily distinguished from incoherent background emissions, thus permitting high signal-to-noise ratios. Third, the use of short laser pulses, down to the order of 10^{-15} s, allows high temporal resolution. High spatial resolution is achieved by generating signals from interaction regions defined by crossed and focused beams. Although the coefficients of successive terms in the power series expansion of the nonlinear polarization decrease by many orders of magnitude with increasing order, signal strengths may be made comparable to those of linear spectroscopy. Strong signals are obtained by the use of resonant enhancement of the nonlinear susceptibility, high field strengths from laser sources, and the use of sensitive photodetectors.

Second-Order Nonlinear Effects

Weak light fields leave unaffected the optical properties of the medium through which they propagate. When, however, the second-order susceptibility $\chi^{(2)}$ contributes significantly to the induced polarization, the complex refractive index includes a field-dependent term:

$$\mathbf{P} = \varepsilon_0[\chi^{(1)}(\omega) + \chi^{(2)}(\omega)\mathbf{E}_1]\mathbf{E}_2 \qquad [4]$$

Owing to the symmetry properties of $\chi^{(2)}$, second-order nonlinear effects are absent in isotropic media or those possessing a center of inversion symmetry. They may, however, be observed in anisotropic materials such as crystalline solids where they are the basis of frequency mixing schemes leading to second-harmonic generation, sum- and difference-frequency generation, and optical parametric amplification and oscillation. The symmetry of gas phase media, such as atomic vapors, may be broken by application of an external electric or magnetic field. Resonantly enhanced second-harmonic generation has been achieved in alkali metal vapors in the presence of a magnetic field giving potential for spectroscopy. Symmetry breaking is achieved also on solid surfaces where sum-frequency generation has been used to study surface morphology, monatomic layers, and catalytic effects of adsorbed materials.

Third-Order Nonlinear Spectroscopy

In gas phase or isotropic solid and liquid media, the first nonlinear term is of third order and leads to a refractive index that is dependent on the field intensity $|\mathbf{E}|^2$:

$$\mathbf{P} = \varepsilon_0[\chi^{(1)}(\omega) + \chi^{(3)}(\omega)\mathbf{E}^2]\mathbf{E} \qquad [5]$$

Intensity-dependent refractive index effects lead to a wide range of nonlinear optical phenomena. In general, the polarization contains terms that are the product of three input fields generating Fourier components at sum and difference frequencies. Phase matching constraints will generally determine which of the possible interactions will be observed. A wide range of physical effects contribute to $\chi^{(3)}(\omega)$. The simplest of these nonlinear effects is saturation of the absorption.

Linear absorption is defined by the situation where the intensity decays exponentially with propagation distance x, according to Beer's Law:

$$I(x) = I(0)\exp\{-\alpha_0(\omega)x\} \qquad [6]$$

where $I(0)$ and $I(x)$ are the intensity values at $x = 0$ and x, respectively, and $\alpha_0(\omega)$ is the weak field absorption coefficient. The fractional power absorbed in this case is, therefore:

$$[I(0) - I(x)]/I(0) = \alpha_0(\omega)x \qquad [7]$$

and is independent of the intensity. When the incident intensity increases, until the rate of stimulated absorption from the lower level equals the spontaneous emission rate out of the upper level, the medium is said to be saturated. The intensity is

then I_{sat}. A simple rate-equation analysis shows that the absorption coefficient in the presence of a strong light field is:

$$\alpha(\omega) = \frac{\alpha_0(\omega)}{[1 + I/I_{sat}]} \qquad [8]$$

The absorption is then no longer independent of the light intensity.

Saturation Spectroscopy

In the case of a homogeneously broadened transition the effect of a saturating monochromatic light field is to reduce the absorption across the entire line profile. This is shown in Figure 1a where the frequency of the incident light w_L is detuned from line center ω_0. The degree of saturation and the effect on the absorption coefficient depends on the intensity I and the detuning $\Delta\omega = \omega_L - \omega_0$.

For a transition dominated by inhomogeneous broadening a monochromatic light field interacts strongly only with a sub-group of atoms. This sub-group is determined by those having a resonance frequency within the spectral range covered by the homogeneous width and centered on the incident frequency. Saturation in this case results in 'hole burning' as shown in Figure 1b. A dip in the absorption coefficient $\alpha(\omega)$ is produced centered on ω_L of width given by the homogeneous width.

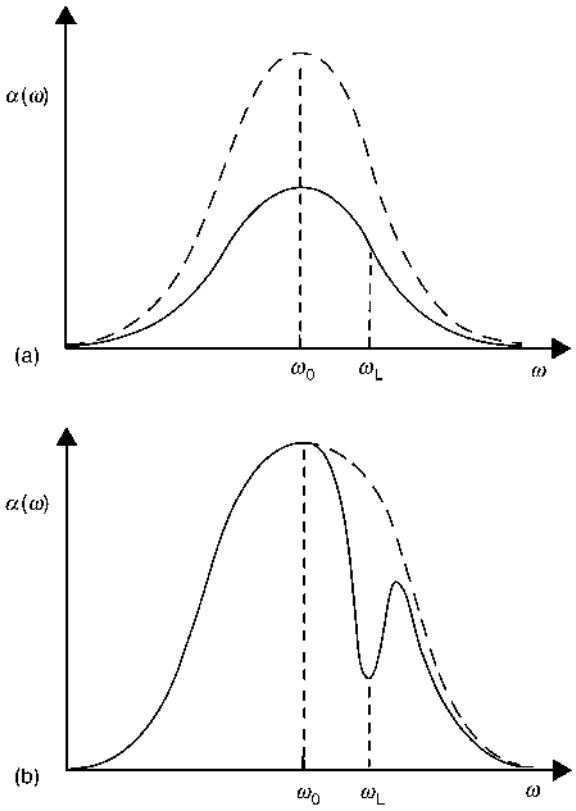

Figure 1 Absorption lineshapes for (a) homogeneously broadened and (b) inhomogeneously broadened lines centered at ω_0. Dashed lines in both figures show lineshape measured by weak probe. Solid lines show lineshape measured by weak probe in presence of a monochromatic, saturating beam at frequency $\omega_L \neq \omega_0$.

Saturation affects the population of both upper and lower levels involved in an optical transition and so affects both absorption and amplification (gain). The phenomenon shows itself as a narrow dip in the spectral profile of lasers operating on an inhomogeneously broadened gain line (e.g., a HeNe gas laser). These 'Lamb dips', as they are called, are the result of saturation by the laser mode which has a very narrow spectral width. Only those atoms in the Maxwell–Boltzmann velocity distribution are saturated, whose velocity results in a Doppler shift of the light that brings it into resonance, i.e., to within the atom's homogeneous linewidth of the resonance frequency. When the laser mode frequency is detuned from atomic line center two velocity subgroups of atoms, within the distribution, are saturated corresponding to positive and negative Doppler shifts for counter-propagating fields of the mode. When the mode frequency is tuned to atomic line center, the counter-propagating fields interact with the same (zero) velocity subgroup. The saturation effect is thereby enhanced, reducing the gain by depletion of the population inversion, leading to a dip in output intensity at line center. The width of the dip is the homogeneous width of the transition (Figure 2).

The variation of output intensity around the Lamb dip is used to provide an error signal in one scheme for laser frequency stabilization. A feedback system is employed to adjust the cavity length, L, in order to maintain the mode frequency, ν, exactly on atomic line-center ($\nu = mc/2L$ where m is an integer). This scheme uses gain saturation of the amplifying medium, whereas an alternative uses saturated absorption. In this scheme, a cell of gas having a narrow absorption line, within the gain profile of the laser, is placed inside the cavity. When the laser mode is tuned to absorption, line-center saturation reduces the absorption loss, resulting in a narrow peak in the laser emission. Drifting off this peak can be used as an error signal, as before, to stabilize the mode frequency.

Saturated absorption spectroscopy uses the same mechanism by placing the absorbing specimen outside the laser cavity. The effect is detected by monitoring the transmission of a weak probe beam in the presence of a strong counter-propagating 'pump' beam. Both probe and pump beams are

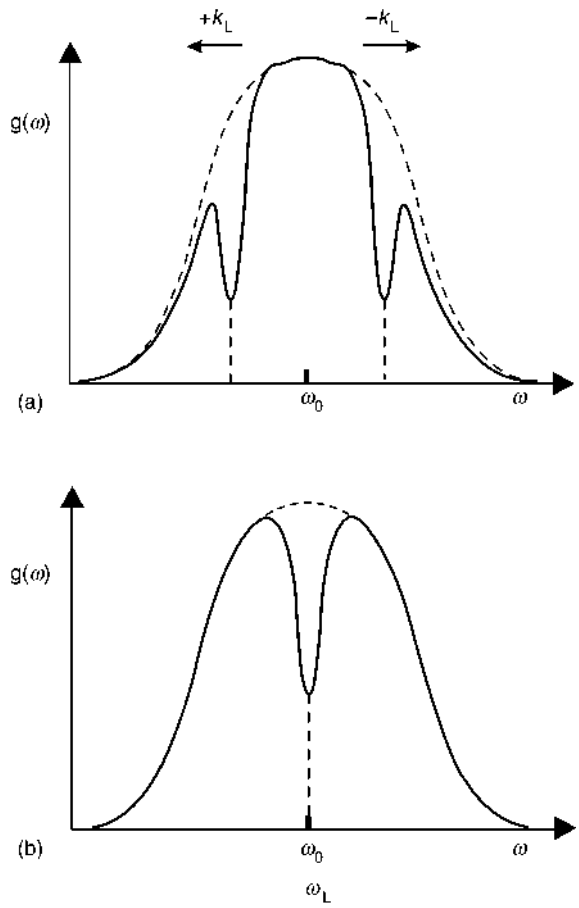

Figure 2 (a) Holes burned in gain profile by two counter-propagating monochromatic waves at $\omega_L \neq \omega_0$. (b) 'Lamb dip' in gain profile $g(\omega)$ (or absorption profile $\alpha(\omega)$) when $\omega_L = \omega_0$.

derived from the same laser using a suitable beamsplitter and, thus, have identical frequency. When tuned to line-center of the absorbing sample, the transmission of the probe increases owing to saturation by the strong pump beam interacting with the same subgroup of the inhomogeneously broadened medium (Figure 3).

The elimination of the first-order Doppler effect, in gaseous media, using this technique results in very high-resolution spectra. The residual linewidth is determined by homogeneous processes such as lifetime broadening, collisional (or pressure) broadening and power (or saturation) broadening. The latter effect is inevitable, as saturation is necessary, but may be controlled so as to contribute very little to the overall broadening. Power broadening becomes significant only when the laser intensity is appreciably in excess of the saturation intensity. Another limiting effect is transit time broadening: moving atoms spend only a limited time interacting with the narrow beams as they pass through. The consequent limit on the

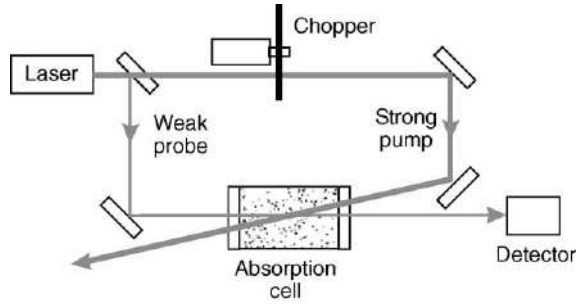

Figure 3 Schematic diagram of saturated absorption spectroscopy experiment. The saturating pump beam is chopped or modulated and the signal is processed using phase-sensitive detection to improve the signal-to-noise ratio.

interaction time leads to conjugate broadening of the frequency of the transition as a result of the uncertainty principle.

The general problem of the spectral lineshape of signals in saturation spectroscopy is complex. Apart from the effects mentioned above, the lineshape will be affected by relativistic effects (second-order Doppler shifts), velocity-changing collisions, recoil effects from photon absorption and emission, velocity-dependent transit time broadening, optical pumping and laser lineshape effects.

Saturation spectroscopy is widely used in high-resolution spectroscopy for measurement of atomic and molecular transition frequencies, time and frequency standards using stabilized lasers, measurement of fundamental constants such as the Rydberg constant and the fine structure constant. Dynamical effects such as atomic collisions are also studied using saturation, whereby the relaxation of hole burning is measured in the time or frequency domain.

Polarization Spectroscopy

An important variation of saturation spectroscopy uses the change induced by a saturating beam in the real part of the complex refractive index, i.e., dispersion rather than absorption. Specifically, the technique of polarization spectroscopy employs a saturating beam that is circularly polarized to effect optical pumping on the degenerate lower level of a transition. The degenerate states may be identified, for example, with magnetic substates based on orientation of the intrinsic atomic or molecular angular momentum and labeled by the quantum numbers m_J. In equilibrium, the populations of all these degenerate substates will be equal and the optical properties of the system will then be isotropic. In particular, the polarization state of a linearly polarized probe beam will be unaffected by passage

through the medium. This probe is totally rejected by an analyzer (a linear polarizer set to transmit only light in an orthogonal plane). The linear polarized beam is a linear combination of two oppositely circular polarization states, σ^+ and σ^-, and both components experience the same complex refractive index. In the presence of a circularly polarized saturating pump beam, the distribution of population amongst the degenerate substates is disturbed from its equilibrium by optical pumping. The result will be an excess of population in higher or lower m_J states (a condition known as orientation) and this leads to optical anisotropy in the medium. The refractive index for σ^+ light is now different from that for σ^- light and a relative phase shift is introduced between these components of the linearly polarized probe resulting in elliptical polarization. A component of this elliptically polarized light will then be transmitted by the analyzer and is detected. When the strong pump beam is closely counter-propagating to the probe, then both beams interact with the same velocity subgroup as in Doppler-free saturation spectroscopy. The inhomogeneous (Doppler) broadening is thus significantly reduced. The technique provides a signal above an essentially 'zero' background and is thus capable of a higher signal-to-noise ratio than saturated absorption spectroscopy. How close to 'zero' the background is depends crucially on the degree of extinction by the analyzer. High-quality polarizer/analyzer devices can achieve extinction ratios of $1{:}10^{-8}$. Stress-induced birefringence in the windows of the cell containing the sample is a common limiting factor (Figure 4).

A probe laser of intensity I_{probe} is transmitted through a sample of length l with an intensity $I_{\text{PS}}(\omega)$ given by the following expression, that takes account of different contributions to the detected signal:

$$I_{\text{PS}}(\omega) = I_{\text{probe}}\left[\xi + \varphi^2 + b^2 + \frac{1}{2}\frac{b\Delta\alpha_{\pm}l}{(1+x^2)} + \frac{1}{2}\varphi\Delta\alpha_{\pm}l\frac{x}{(1+x^2)} + \left(\frac{1}{4}\Delta\alpha_{\pm}l\right)^2\frac{1}{(1+x)^2}\right] \quad [9]$$

where $x = 2(\omega_0 - \omega)/\gamma$ is the detuning from the resonance frequency ω_0 normalized by the halfwidth γ of the collision broadened line. The frequency-dependent terms provide absorptive (Lorentzian) and dispersive contributions to the total lineshape. The first three, frequency independent, terms in eqn [9] are responsible for a constant background signal. They originate from residual transmission of the polarizers (ξ), their accidental uncrossing (φ), and a possible residual birefringence (b) from optical elements between both polarizers in the probe beam path. ($\xi + b^2$), normally is in the range 10^{-6}. The differential absorption coefficient for left and right circularly polarized light, $\Delta\alpha_{\pm}$, at resonance ($x = 0$), is given by:

$$\Delta\alpha_{\pm}(x=0) = \Delta\alpha_{\pm}^0 = \alpha^+ - \alpha^- = \alpha_0 S_0 \Delta C_{JJ_1} \quad [10]$$

where α_0 is the unsaturated absorption coefficient, S_0 is the saturation parameter for the pump wave ($I_{\text{pump}}/I_{\text{sat}}$), and ΔC_{JJ_1} are Clebsch–Gordan coefficients for the respective transition and coupling case.

The dispersion-shaped component may be made to dominate by deliberately uncrossing the polarizer/analyzer combination to a small extent. This technique enhances the resolution of the transition frequency. Linewidths are limited by residual Doppler broadening owing to finite crossing angles of pump and probe beams, collisional broadening, transit time broadening, etc.

Applications of the technique have included Doppler-free spectroscopy of atomic transitions for precision measurements of fundamental constants, molecular spectroscopy and detection of atomic, molecular, and radical species in flames. By employing a different laser to accomplish the induced optical anisotropy from that used to probe the medium, molecular spectra can be simplified by the technique of 'polarization labeling'. Signals are obtained only when the probe and pump excite transitions sharing a common level and the particular transitions observed are sensitive to the selection rules for ΔJ, governing both pump and probe transitions. In this way P- and R-branch transitions can be distinguished.

Advantages of polarization spectroscopy include high sensitivity (10^2–10^3 times that of saturation spectroscopy), high signal-to-noise, simplification of

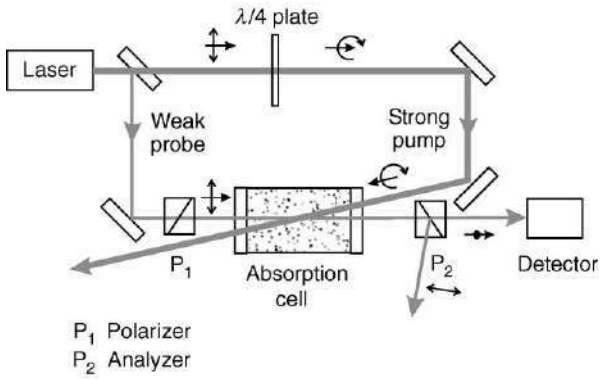

Figure 4 Schematic diagram of polarization spectroscopy experiment. Circularly polarized pump beam may be modulated for phase-sensitive detection of transmitted probe beam.

molecular spectra by polarization labeling, dispersion profiles for enhanced resolution, and for frequency stabilization techniques.

Multiphoton Absorption Spectroscopy

Multiphoton absorption is the term used to describe a process in which an atom or molecule makes a single transition between two of its allowed energy levels by absorbing the energy from more than a single photon. Sequential transitions, by which successive photons are absorbed via single-photon transitions between stationary energy levels, take place provided photons arrive within the lifetime of the intermediate state. Multiphoton transitions take place via 'virtual' energy levels. The lifetime of a virtual level is determined by the uncertainty principle to be the inverse frequency detuning from the nearest stationary energy level. Thus a high photon flux is usually required for multiphoton effects to be observed (Figure 5).

The simplest multiphoton effect, two-photon absorption, may be observed using strong atomic or molecular electronic transitions, even with modest laser power. The absorption is detected by observing the fluorescence from the spontaneous, radiative decay of the excited state. Two-photon absorption spectroscopy may be made Doppler-free by arranging absorption of one photon from each of two counter-propagating beams. The moving atoms experience opposite Doppler shifts $\mathbf{k}\cdot\mathbf{v}$ from each beam. The transition energy is thus:

$$E = (\hbar\omega + \mathbf{k}\cdot\mathbf{v}) + (\hbar\omega - \mathbf{k}\cdot\mathbf{v}) = 2\hbar\omega \qquad [11]$$

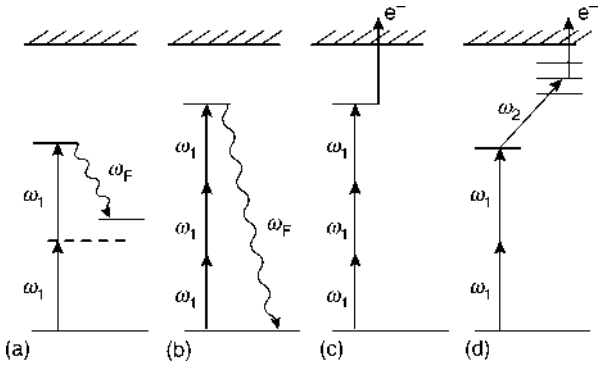

Figure 5 Multiphoton absorption spectroscopy. (a) Two-photon absorption using fluorescence detection of radiative decay at ω_F. (b)Three-photon absorption spectroscopy using fluorescence detection of radiative decay at ω_F. (c) Three-photon absorption spectroscopy using ionization detection following photoionization, collisional or field ionization. (d) Two-photon resonant multiphoton ionization spectroscopy; laser at frequency ω_1 is fixed and laser at ω_2 is tuned over transitions to high-lying levels.

Absorption of two photons from a single beam is prevented by choosing a transition between levels of the same parity and angular momentum quantum number, i.e., $\Delta J = 0$ and using circularly polarized light. Absorption of two σ^+ or σ^- photons gives $\Delta J = \pm 2$, which would be forbidden, e.g., on a n^2S–n'^2S transition in an alkali metal. This method of Doppler-free spectroscopy has the advantage that it involves all the atoms and not just the zero-velocity subgroup, as in saturation or polarization spectroscopy. In principle, it is also possible to use three-photon absorption, with the momentum of the wavevectors arranged to sum to zero, to achieve Doppler-free absorption on a transition between opposite parity states.

Instead of observing fluorescence following multiphoton absorption, the excited atom or molecule may be ionized by atomic collisions or by applying a d.c. electric field. Excitation of Rydberg levels (i.e., having high value of the principal quantum number n) by two- or three-photon absorption is readily accomplished. The sensitivity of this technique is enhanced by using two lasers. One laser at frequency ω_1 is tuned to a two-photon transition (energy $= 2\hbar\omega_1$) and the second laser ω_2 is tunable to excite transitions from the two-photon intermediate state to higher-lying levels (energy $= 2\hbar\omega_1 + \hbar\omega_2$). Thus, energy levels normally requiring excitation in the vacuum UV may be reached using visible frequency lasers. In such cases ionization detection may be more efficient than fluorescence since the nonradiative decay rate of these highly excited levels may exceed the fluorescence decay rate. The ion signal produced may be significantly amplified by the process of space charge amplification. A weak discharge established in the atomic vapor creates a space charge around the cathode. A single ion moving slowly through the space charge cloud acts to cancel the local field allowing many (up to 10^6) electrons to escape to the anode with a resulting amplification of 10^6 in the 'photoionization current'.

Multiphoton ionization spectroscopy is a sensitive method of recording highly excited energy levels. The sensitivity is enhanced in the technique of resonance ionization spectroscopy (RIS): two lasers are used such that the frequency of one laser is fixed close to a two-photon allowed transition and the second is tuned to excite a range of higher excited states. Absorption of an even number of photons allows excitation of levels having the same parity as the ground state and thus not accessible to normal absorption spectroscopy. In combination with a mass spectrometer to detect the ions produced by the multiphoton absorption resonance ionization mass spectrometry (RIMS), affords additional species

selectivity and sensitivity. RIMS finds application in high-resolution spectroscopy of rare isotopic species and is the basis for a scheme for isotope separation.

Coherent Third-Order Processes

An important class of spectroscopic techniques exploits the coherent interaction of three input waves via the nonlinear response of the medium. Specifically, the third-order term has the form:

$$\mathbf{P}_i^{(3)}(\omega_4, \vec{r}) = \chi_{ijkl}^{(3)}(\omega_1, \omega_2, \omega_3, \omega_4)\mathbf{E}_j(\omega_1, \vec{r}) \times \mathbf{E}_k(\omega_2, \vec{r})\mathbf{E}_l(\omega_3, \vec{r}) \quad [12]$$

where $\mathbf{E}_m(\omega_n)$ are complex field amplitudes and the subscript $m = i, j, k, l$ refers to each of the Cartesian coordinates x, y, z of the electric vector. $\chi_{ijkl}^{(3)}$ is a fourth-rank tensor. In condensed phase media, the third-order nonlinearity is readily observed, even with modest laser intensities. In the gas phase, although $\chi_{ijkl}^{(3)}(\omega)$ is small relative to $\chi^{(1)}$, its effect is enhanced when the incident fields are resonant and this nonlinear response may be used for spectroscopy. Equation [12] shows that the induced polarization couples three waves to radiate a fourth and describes a class of interactions known collectively as four-wave mixing.

The physics of four-wave mixing interactions may be visualized as the formation and read-out of a volume hologram. The three input beams are denoted as two pumps and one probe. The process involves the interference of one pump with the probe to form a spatially periodic grating pattern in the electric field distribution. The nonlinear response of the medium via $\chi^{(3)}$ leads to a spatial modulation of the refractive index – a laser-induced grating. The second pump beam generates a signal by scattering off this grating in a manner analogous to Bragg scattering of X-rays from regular crystal planes. The condition for generating a coherent scattered beam is equivalent to matching the momentum (or wavevectors) of the interacting photons, i.e., phase matching.

Degenerate Four-Wave Mixing (DFWM) Spectroscopy

The simplest case of four-wave mixing spectroscopy involves pump and probe beams of identical (i.e., degenerate) frequency to generate a signal at the same frequency. Spectra are generated as the frequency is tuned through atomic or molecular resonances where the interaction is strongly enhanced. In this degenerate case, the pump and probe interference pattern is stationary and the remaining pump beam scatters off this grating with the same frequency. As with all parametric processes, the interaction is phase matched by a suitable geometrical arrangement of the beams. A probe intersecting with counter-propagating pump beams leads to a signal that counter-propagates with the probe. This arrangement provides 'Doppler-free' lineshapes, since only those atoms with essentially zero velocity along the line of intersection of the beams contribute to the signal. Alternatively, pump and probe beams may be arranged to cross at an angle (usually small) in the forward direction – the 'forward geometry', with the resulting signal propagating at the same small angle to the incoming beams. Signals in this case are stronger, since all velocity classes may contribute but the lineshape is Doppler broadened (Figure 6).

An additional feature of the counter-propagating pump geometry is that the phase-matched signal arises from terms that involve products of fields with the complex conjugate of the incident

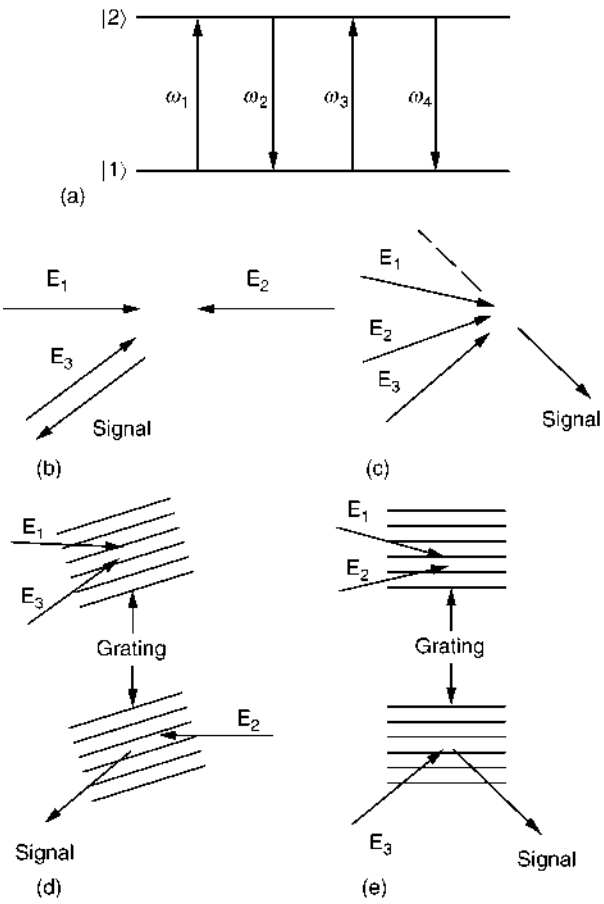

Figure 6 Degenerate four-wave mixing: (a) all input frequencies are degenerate, $\omega_1 = \omega_2 = \omega_3$ and give signal at the same frequency, ω_4. Phase matching geometries are shown in (b) counter-propagating pump (phase conjugate) and (c) forward geometry. (d) and (e) show the grating formation and read-out respectively.

probe field. The induced polarization radiating the signal wave has a component given by:

$$\mathbf{P}^{(3)}_{\text{signal}}(\omega) = \chi^{(3)}(\omega)\mathbf{E}_{\text{pump}}(\omega)\mathbf{E}_{\text{pump}}(\omega)\mathbf{E}^{*}_{\text{probe}}(\omega) \quad [13]$$

The backward propagating signal appears as a phase conjugated reflection of the incident probe wave. A probe wave that has propagated through a phase-distorting medium thus generates a signal with an initial phase distortion that will be undone as it retraces the path through the distorting medium. This phase conjugating property finds application in adaptive optics but also lends an advantage when DFWM is used as a spectroscopic probe in hostile or distorting environments, such as flames or plasmas. The counter-propagating pump geometry is often referred to as the phase conjugating geometry.

The intensity of the DFWM signal beam may be calculated using perturbation theory to derive $\chi^{(3)}(\omega)$ arising from saturated absorption. In the perturbation approximation, the probe beam is assumed to be weak and interacts with stationary atoms in the presence of counter-propagating, saturating pump fields. The signal is found from the phase-matched terms in the wave radiated from the induced polarization. When integrated over the atomic resonance line, the signal intensity, $I^{\text{int}}_{\text{sig}}$, is given by analytic expressions in the two limiting regimes of weak and strong saturation by the pump:

$$I^{\text{int}}_{\text{sig}} = \frac{|\mu_{12}|^8 N^2 k^2 T_1^2 T_2^3}{[1 + 4I_{\text{pump}}/I_{\text{sat}}(0)]^{5/2}} I^2_{\text{pump}} I_{\text{probe}} \quad [14]$$

when $I_{\text{pump}} \ll I_{\text{sat}}(0)$ and

$$I^{\text{int}}_{\text{sig}} \propto |\mu_{12}|^3 N^2 k^2 T_1^{-1/2} T_2^{1/2} I^{-1/2}_{\text{pump}} I_{\text{probe}} \quad [15]$$

when $I_{\text{pump}} \gg I_{\text{sat}}(0)$.

In these relations μ_{12} is the atomic dipole moment, N the atomic number density, $I_S(0) = \hbar^2/T_1T_2|\mu_{12}|^2$ is the line-center saturation intensity, and T_1 and T_2 are the longitudinal and transverse relaxation times, respectively. Extensions of the basic theory accommodate the effects of moving absorbers (Doppler effect) in both the phase-conjugate and forward phase matching geometry.

These results are derived using perturbation theory, by assuming that the probe field is weak. In practice the signal is optimized when pump and probe are of equal intensity and approximately equal to the saturation intensity. This situation violates the perturbation theory assumption but the signal intensity may be derived using nonperturbative analytical methods, although the resulting expression is more complicated. Alternatively, the quantum mechanical equations for the nonlinear susceptibility can be solved numerically to find the signal level for arbitrary intensity of the incident fields. Optimum signal-to-noise ratio in the spectra is achieved using approximately saturating intensities.

The sensitivity of resonant DFWM and the coherent nature of the signal beam suits the process to spectroscopy of nonfluorescing or trace molecular species, such as radicals created in chemical reactions. Equation [15] shows that the DFWM signal becomes relatively insensitive to collisional effects when the incident lasers saturate the transition. This feature of DFWM gives it an additional advantage over laser-induced fluorescence in optical diagnostic applications where quantitative evaluation of fluorescence intensity is hampered by collisional quenching. The state and species selectivity afforded by resonant DFWM is exploited in diagnostics of combustion and plasma processes, where the reduction of Doppler broadening allows high-resolution spectra from high-temperature gases. The relative intensity of spectral lines is used to derive temperature and concentration information with time and space resolution.

Additional selectivity is provided by using two input frequencies in two-color four-wave mixing (TC-FWM), to probe molecular excited state dynamics. The TC-FWM process uses resonant enhancement on two coupled transitions, i.e., the pump beams excite one transition and the probe and signal are generated on another transition having a common level with that excited by the pumps.

Coherent Raman Spectroscopy

The classical theory of spontaneous Raman scattering shows that a nonresonant pump of frequency ω_p wave can excite a resonance ω_m in a molecule with a nonlinear polarizability $\beta(x)$. (x is the spatial coordinate of the internuclear potential.) The wave resulting from this inelastic scattering is a Stokes wave of frequency $\omega_S = \omega_p - \omega_m$. The propagation of the Stokes wave is governed by the complex Raman susceptibility (**Figure 7a**):

$$\chi_{\text{Raman}}(\omega_s) = \chi'_{\text{Raman}}(\omega_s) + i\chi''_{\text{Raman}}(\omega_s) \quad [16]$$

This generated Stokes wave may then couple with the pump to drive the molecular resonance coherently leading to amplification of the Stokes wave at the expense of the pump. This process of stimulated Raman scattering (SRS), provides gain if the amplification exceeds the loss due to absorption, scattering, etc. There is, therefore, a definite threshold intensity that the pump must satisfy to observe SRS. The SRS

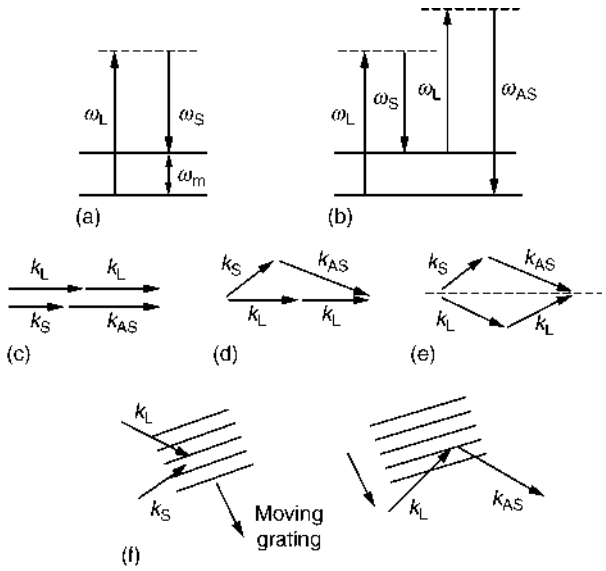

Figure 7 Coherent Raman spectroscopy. (a) Spontaneous Raman scattering of input pump ω_L producing Stokes wave ω_S in inelastic scattering. (b) Coherent anti-Stokes Raman scattering (CARS). Phase matching geometries for CARS is shown for a nondispersive medium in (c) and dispersive media in (d) and (e). The planar boxcars geometry in (e) may be folded along the dotted line (folded boxcars geometry) to separate the signal into an orthogonal plane from the pump beams to reduce scattered pump light reaching the detector.

gain $g(\omega_s)$ at the Stokes frequency ω_s is given by:

$$g(\omega_s) = -\frac{k_s}{2n_s^2}|E(\omega_L)|^2\chi''_{\text{Raman}}(\omega_s) \qquad [17]$$

where k_s and n_s are the wavevector and refractive index for the Stokes wave, $E(\omega_L)$ is the field amplitude at the pump laser frequency ω_L, and $\chi''_{\text{Raman}}(\omega_s)$ is the imaginary part of the complex Raman susceptibility:

$$\chi''_{\text{Raman}}(\omega_s) = \frac{-\varepsilon_0 N\left(\frac{\partial\beta}{\partial x}\right)^2\gamma}{32m_e\omega_v\left\{(\omega_m - (\omega_L - \omega_S))^2 + \frac{\gamma^2}{4}\right\}} \qquad [18]$$

In this expression $\partial\beta/\partial x$ is the variation of polarizability β with internuclear displacement x, N is the molecular population density, and γ is the linewidth due to damping. Spectroscopy based on SRS provides signals that are many orders of magnitude stronger than those from spontaneous Raman scattering, and is widely used to study spectra of complex molecules having no absorption spectra in the visible region.

A four-wave mixing interaction, involving a pump and probe beam of different frequencies, generates a nonstationary grating. The signal generated by scattering of a second pump beam off this moving grating is Doppler-shifted to lower or higher frequency. These signal beams correspond respectively to Stokes or anti-Stokes waves in Raman scattering. A probe beam at the Stokes frequency will thus interact with a pump beam to generate a coherent signal at the anti-Stokes frequency; a process known as coherent anti-Stokes Raman scattering (Figure 7b). A probe at the anti-Stokes frequency correspondingly gives the process of coherent Stokes Raman scattering (CSRS). Phase matching is necessary for both CARS and CSRS and is satisfied by an appropriate arrangement of the incident wavevectors (Figure 7c,d and e).

CARS has proven to be a valuable spectroscopic technique in a wide variety of applications. A CARS spectrum is generated by scanning the frequency of the probe, i.e., the Stokes laser. The signal strength is given by

$$I_4 = I_{\text{CARS}} = \frac{\omega_4^2}{n_1^2 n_2 n_4 c^4 \varepsilon_0^2} I_1^2 I_2 |\chi_{\text{CARS}}|^2 l^2 \times\left(\frac{\sin(\Delta kl/2)}{\Delta kl/2}\right)^2 \qquad [19]$$

where I_1 and I_2 are the intensity of the pump and Stokes beams, respectively, l is the interaction length, and n_i are the refractive indices at frequency ω_i. The maximum signal is generated when the phase mismatch $\Delta k = 0$. The third-order susceptibility, $\chi^{(3)} = \chi_{\text{CARS}}$, in general consists of a nonresonant part $\chi^{(3)}_{\text{nr}}$, which is independent of the frequency of the exciting beams, and a Raman-resonant contribution $\chi^{(3)}_{\text{r}}$:

$$\chi^{(3)} = \chi^{(3)}_{\text{nr}} + \chi^{(3)}_{\text{r}} \qquad [20]$$

The lineshape in coherent Raman spectroscopy is affected by interference between resonant and nonresonant contributions. This effect is exploited to derive the concentration of the resonant (Raman active) species relative to that of the background of known density. A more common application of CARS, however, is for temperature measurement. The dependence of the signal intensity on $|\chi_{\text{CARS}}|^2$, and therefore on the square of the molecular population of the initial state, yields a spectrum from which the temperature may be derived by assuming a Maxwell–Boltzmann distribution over the state populations.

A particularly useful application of CARS employs a narrow linewidth pump laser and a

broad bandwidth Stokes laser. CARS signals are generated simultaneously on all frequencies such that the corresponding Stokes frequency is available from the broadband probe. The spectrum recorded with a single laser pulse can thus provide population information on a wide range of molecular states, giving time- and space-resolved measurement of temperature and concentration.

CARS and CSRS spectroscopy is applied to both gas and condensed phase media. Owing to the higher density in the latter, signals may be strong in spite of the two-photon nature of the resonant interaction.

Time Domain Four-Wave Mixing Spectroscopy

Four-wave mixing spectroscopy may also be executed in the time domain. The three incident waves may arrive as pulses separated in time by variable delays. The interaction is, therefore, mediated by the atom or molecule's induced polarization or coherence. DFWM, CARS, and CSRS processes may thus be used to study ultrafast relaxation since the signal intensity gives a measure of the coherence decay in the intervening delay between the pulses. These processes are intimately related to coherent transient spectroscopy using measurements of free induction decay, photon echoes, and stimulated photon echoes.

Laser-Induced Grating Spectroscopy

The four-wave mixing processes discussed above involve a spatial modulation of the molecular population and a consequent modulation of the refractive index for resonantly interacting waves. The signals are radiated by the coherent oscillations of the molecular populations that decay rapidly. Relaxation processes result in transfer of energy from the excited molecular states, leading to a temperature and density grating in the bulk medium with a period Λ, given by:

$$\Lambda = \lambda_{\text{pump}}/(2 \sin \theta_1) \qquad [21]$$

where θ_1 is the angle between the interfering beams of wavelength λ_{pump}. A nonresonant wave of wavelength λ_{probe} may be scattered by this laser-induced grating provided it is incident on the grating at the angle θ_2 satisfying the Bragg condition:

$$\theta_2 = \sin^{-1}(\lambda_{\text{probe}}/2\Lambda) \qquad [22]$$

The periodic temperature/density grating, when induced by a short duration laser pulse, is accompanied by two oppositely propagating acoustic waves. The reflectivity of Bragg scattering is temporally modulated as the acoustic waves traverse the stationary grating at the local sound speed. Measurement of the period of these oscillations yields the sound speed from which the temperature may be derived. The stationary grating and acoustic waves decay by diffusive and damping processes, dependent upon the bulk medium properties. Measurement of the decay time of the grating induced in gas phase media yields the gas pressure if the gas dynamic parameters are known (Figure 8).

Laser-induced grating spectroscopy (LIGS), has been widely employed for studies of condensed phase relaxation processes. Applications have also been developed for gas phase studies using resonant absorption to enhance the thermal grating component. A grating may, however, also be induced in response to the electrostrictive effect of the high electric field in the interference pattern of two intense pump waves. Such electrostrictive gratings do not require resonant absorption and their spectroscopic utility centers on the study of relaxation effects in the molecular medium.

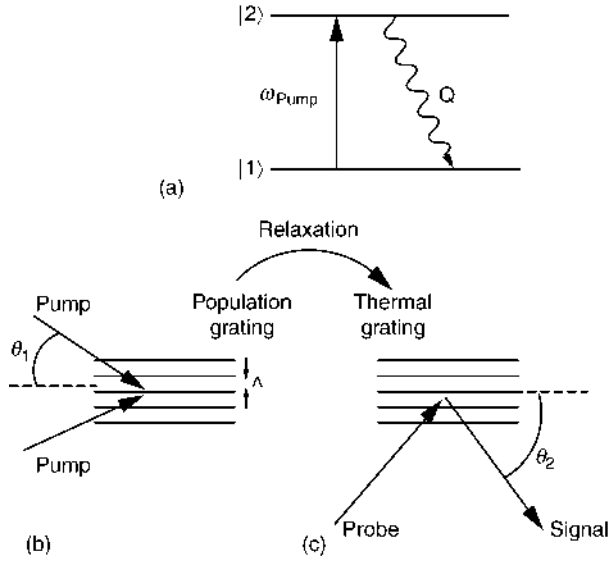

Figure 8 Laser-induced thermal grating spectroscopy. (a) Absorption of energy from pump lasers creates excited molecular population in grating pattern shown in (b). Spacing of laser induced grating Λ is set by angle of intersection of pumps. Quenching, *Q*, of excited population by collisions transfers heat to bulk medium and forms thermal grating. (c) Probe incident at Bragg angle θ_2 is scattered to form signal.

List of Units and Nomenclature

Formulae are expressed in SI units

c	Speed of electromagnetic wave *in vacuo*
$\mathbf{E}$	Electric field amplitude
$g_0(\omega)$, $g(\omega)$	Gain coefficient at line center, at frequency ω
h	Planck's constant
I	Intensity of electromagnetic wave
I_{sat}	Saturation intensity of medium
$\mathbf{k}$, k	Wave vector, scalar value of wave vector ω/c
m_e	Mass of electron
n_m	Refractive index of medium for mth beam
N	Atomic or molecular population density
$\mathbf{P}$	Polarization field amplitude
T_1	Longitudinal relaxation time
T_2	Transverse relaxation time
α_0, $\alpha(\omega)$	Absorption coefficient at line center, at frequency ω
$\beta(x)$	Molecular polarizability as function of internuclear separation x
ε_0	Permittivity of free space
μ_{12}	Dipole moment for transition between energy levels 1 and 2
ν	Frequency of electromagnetic wave
$\chi^{(n)}$	nth order susceptibility
ω	Angular frequency of electromagnetic wave

See also

Coherent Transients: Coherent Transient Spectroscopy in Atomic and Molecular Vapours; Foundations of Coherent Transients in Semiconductors; Ultrafast Studies of Semiconductors. **Fiber and Guided Wave Optics:** Nonlinear Optics. **Imaging:** Adaptive Optics. **Nonlinear Optics, Applications:** Phase Matching. **Nonlinear Optics, Basics:** $\chi^{(2)}$–Harmonic Generation; Four-Wave Mixing; Kramers–Kronig Relations in Nonlinear Optics. **Phase Control:** Phase Conjugation and Image Correction. **Physical Applications of Lasers:** Sum-Frequency Generation at Surfaces. **Scattering:** Raman Scattering; Stimulated Scattering. **Spectroscopy:** Raman Spectroscopy; Second-Harmonic Spectroscopy.

Further Reading

Clark RJH and Hester RE (eds) (1988) Quantitative CARS spectroscopy. In: *Advances in Non-Linear Spectroscopy*, pp. 193–251. New York: John Wiley.

Corney A (1986) *Atomic and Laser Spectroscopy*. Oxford, UK: Oxford University Press.

Demtröder W (1996) *Laser Spectroscopy*, 2nd edn. Berlin: Springer-Verlag.

Dreier T and Ewart P (2002) Coherent techniques for measurements with intermediate concentrations. In: Kohse-Höinghaus K and Jeffries JB (eds) *Applied Combustion Diagnostics*. London: Taylor and Francis.

Eesley GL (1981) *Coherent Raman Spectroscopy*. London: Pergamon.

Eichler HJ, Günter P and Pohl DW (1986) *Laser Induced Dynamic Gratings*. Berlin: Springer-Verlag.

Feld MS and Letokhov VS (eds) (1980) *Coherent Nonlinear Optics, Recent Advances*. Berlin: Springer-Verlag.

Fisher RA (ed.) (1983) *Optical Phase Conjugation*. San Diego, CA: Academic Press.

Levenson MD and Kano SS (1987) *Introduction to Nonlinear Laser Spectroscopy*. Revised edn. San Diego, CA: Academic Press.

Raman Spectroscopy

R Withnall, University of Greenwich, Chatham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Inelastic light scattering, the optical analogue of Compton scattering, had been predicted to occur by Smekal in 1923, but it was Chandresekar Venkataraman Raman who provided the first experimental demonstration of the phenomenon in February 1928. Only a few months later in 1928 the Russian scientists, Landsberg and Mandelstam observed inelastic light scattering from a quartz crystal. In spite of initial claims to the contrary, this was the same effect that had been observed by Raman, although the Russian scientists did not acknowledge this, preferring to call the effect 'combinatorial scattering'. It was not until the 1970s that the effect became known as Raman scattering everywhere, including Russia.

As typically practised, Raman spectroscopy involves laser excitation of a sample and measurement of the wavelength and intensity distribution of the scattered Raman light. When a microscope is used for delivering laser excitation and/or collecting the inelastically scattered light, the technique is referred to as Raman microscopy. However, it is important to

recognize that, apart from the differences in sampling configuration, there are no fundamental differences between Raman microscopy and Raman spectroscopy; the terms merely identify the different sampling techniques.

Since its invention in 1960, major advances in the technology of the laser have occurred, resulting in the wide variety of laser light sources that are currently available for Raman spectroscopy. Characteristic properties of laser systems for Raman microscopy will be described in this chapter, following an introduction to the technique and its applications in the next section below. Then, the commonly used laser sources for Raman microscopy are categorized according to the wavelength regions of their emissions, along with their merits and drawbacks for each application. The chapter will consider mainly continuous wave laser sources, because pulsed laser beams that are tightly focused with microscope optics give very high irradiance that would destroy most samples.

Raman Microscopy

The first experimental Raman microscopes were developed in the mid-1970s by two independent groups, and it was not long before the first generation of Raman microscopes were subsequently commercialized. Many of these early instruments simply consisted of commercial optical microscopes coupled to Raman spectrometers, as shown in Figure 1. It was recognized that the use of a microscope for sampling in this way can facilitate the Raman examination of tiny particles of micrometer dimensions.

The optical lay-out in Figure 1 shows an infinity corrected microscope. The advantage of using this, rather than a standard tube length microscope, is that its length can be extended in order to incorporate additional optical elements that are required for coupling to the Raman excitation and collection optics. In the optical layout, the incident laser beam is spatially filtered by a pinhole, D_1, in order to remove the diffraction rings and give a point source. The laser light is then partially reflected by beamsplitter, B_1, and focused on to the sample by the microscope objective. The back-scattered Raman radiation is collected by the same microscope objective, partially transmitted by the beamsplitter, and directed into the entrance slit of the Raman spectrometer by means of coupling optics. The accurate location of an aperture, D_2, at the focal point improves the spatial resolution and allows depth profiling of transparent samples. The two pinholes, D_1 and D_2, act as spatial filters and are referred to as confocal diaphragms. The Raman microscope is described as being confocal, because out-of-focus light, collected from outside the focal volume, is rejected.

The use of the beamsplitter, in the optical path of early Raman microscope designs, gives rise to significant losses of the precious Raman

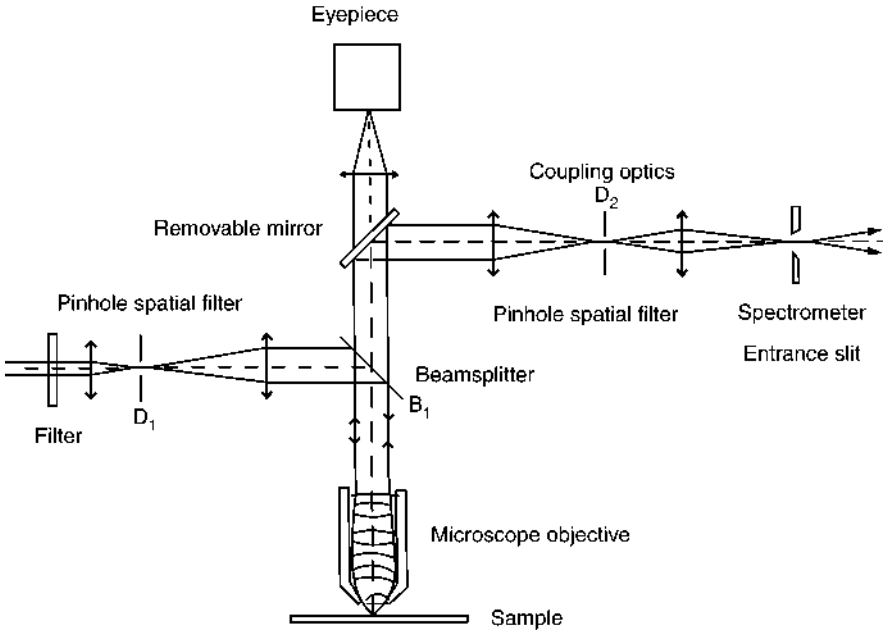

Figure 1 Diagram of the optical layout of a visible Raman microscope (Reproduced from Turrell G, Delhaye M and Dhamelincourt P (1996) Characteristics of Raman microscopy. In: Turrell G and Corset J (eds) *Raman Microscopy. Development and Applications*, 2. London: Academic Press).

scattered light. If a 50:50 beamsplitter is used then half of the incoming laser light is lost by transmission at B_1 and, more significantly, half of the Raman scattered light is lost by reflection at B_1. As long as a higher laser power can be employed, it is better to use a 90:10 beamsplitter that transmits 90% of the light and reflects 10%; then, only 10% of the incoming laser light is directed by B_1 on to the sample, but 90% of the Raman scattered light is transmitted by B_1 towards the detector and only 10% of the Raman light is lost by reflection at B_1. The efficiency of a beamsplitter can be defined as the product of the reflectivity at the laser wavelength and the transmission at the Raman-shifted wavelength. Thus, conventional 50:50 and 90:10 beamsplitters would have efficiencies of 25% and 9%, respectively. Recently, however, holographic notch filters have been developed which can be used as higher efficiency beamsplitters. These typically have a laser rejection contrast ratio of better than 10^4, reflecting ca. 90% of the laser excitation and transmitting ca. 90% of the Raman scattered light, thus giving an efficiency of ca. 81%. Due to these considerations, higher throughput, commercial Raman microscope designs usually incorporate a holographic notch filter.

The importance of Raman microscopy stems from the fact that it is the only microanalytic method available today, by use of which it is possible to identify, or characterize, small particles of micrometer dimensions *in situ*. It is also advantageous that no sample preparation is required when performing Raman microscopy.

The fundamental limit of the lateral spatial resolution (i.e., for a diffraction limited focus) is the separation at which the maximum of one Airy disc function just touches the first minimum of an adjacent Airy disc, given by:

$$\text{Lateral spatial resolution} = \frac{1.22\lambda}{2\text{NA}} \qquad [1]$$

where λ is the wavelength of the light and NA is the numerical aperture of the microscope objective.

The axial Airy disc function determines the resolution in the longitudinal direction. A good estimate of the axial resolution limit for low numerical aperture is given by:

$$\text{Axial spatial resolution} = \frac{2\lambda}{(\text{NA})^2} \qquad [2]$$

When the sample is heterogeneous and exhibits fluorescence that is not evenly distributed within the sample, a region of the sample can be selected with the microscope that shows the minimum amount of fluorescence. If the fluorescence is intrinsic to the sample itself, then it is possible to use the shift subtract technique and/or temporal discrimination between the fluorescence and the Raman scattering. The latter can be achieved by Kerr gate fluorescence rejection, in which a pulse of light is used to close a Kerr gate shutter. The problem with this approach, however, is that it uses a laser pulse, and even pulses of modest energy will have extremely high irradiance when focused to a tiny spot by a microscope objective. Such laser pulses would inevitably destroy the sample under a microscope.

Another approach for reducing fluorescence is to use near infrared excitation, which is low enough in energy so that absorption cannot occur to promote electronic transitions. The technique of Fourier Transform Raman (FTR) spectroscopy often offers the best chance of obtaining Raman spectra from fluorescent samples for this reason. Low energy excitation of wavelength equal to 1064 nm, provided by a Nd:YAG laser, is normally used for FTR. Disadvantages are that the Raman scattering efficiency is low relative to that obtained with visible excitation, due to the ν^4 dependency (ν^4 refers to Rayleigh scattering of the Raman light). This is compensated to some extent by employing a high throughput (Jacquinot advantage) interferometer. However, this works better for macro rather than micro samples. This is because the coupling of a microscope to an FT Raman spectrometer has a fundamental drawback; the microscope is throughput limited, so the high throughput advantage of the FTR spectrometer cannot be realized in FTR microscopy. The image of the Jacquinot stop (the large circular aperture which is the entrance to the interferometer) at the sample plane typically has a diameter of a few hundred μm, which is much larger than particles with diameters of ca. 1 μm having similar dimensions to the waist diameter of the 1064 nm light excitation at its diffraction-limited focus. Consequently the throughput advantage of the interferometer is not fully exploited and there is a trade-off of spatial resolution with signal-to-noise at the detector. For this reason, the typical spatial resolution that is achieved is in the range of 15–100 μm rather than 4 μm, as claimed in the literature. Bruker has commercialized an FTR microscope by coupling an optical microscope to an FTR spectrometer with near infrared optical fibers.

An approach that offers more promise for reducing fluorescence and achieving spatial resolution close to the diffraction limit is to use near infrared excitation in conjunction with dispersive Raman microscopy. For example, semiconductor lasers operating from 785 to 852 nm can be used in conjunction with sensitive multichannel silicon-based CCD arrays. If longer excitation wavelengths are necessary, in order to

overcome the problems with fluorescence, then diode lasers can be used that emit at longer wavelengths in conjunction with dispersive Raman spectrometers equipped with multichannel near infrared (NIR) detectors, e.g., Ge- or InGaAs-array detectors.

The applications of Raman microscopy cover many areas, including material science, the earth sciences, environmental science, biology and medicine, forensic science, and even the analysis of artworks. These areas are too wide ranging to be described in detail here, but the interested reader is referred to the Further Reading section at the end of this article.

Characteristics of Laser Sources

The laser is an excellent light source for Raman spectroscopy to such an extent that the terms 'Raman spectroscopy' and 'laser Raman spectroscopy' are synonymous for all but the most esoteric experiments, for example, with synchotron sources. Indeed, the advent of the laser was the stimulation for the renaissance of Raman spectroscopy in the 1960s, given its special properties, such as monochromaticity, high intensity, beam collimation, and coherence.

The characteristics of laser light and the advantages it offers to Raman microscopy are now considered.

Beam Quality

The light inside a laser tube is formed from a number of standing waves having distinct vibrational modes. There are a limited number of these modes transverse to the beam and these are characterized by the TEM_{pq} number (where p and q can be 0, 1, 2,...) where TEM is an acronym for 'transverse electromagnetic'. When a laser is operating in its fundamental TEM_{00} mode, light rays are reflected on axes between the end mirrors of the laser cavity. The '00' indicates that there are no nodes in the beam profile (Figure 2a), and the laser beam has a Gaussian intensity profile in the radial direction:

$$I(r) = I_0 \exp\left(-\frac{2r^2}{w^2}\right) \qquad [3]$$

where $I(r)$ is the irradiance at a radial distance r from the axis of the beam, I_0 is the axial irradiance, and w is the beam radius.

For higher-order modes, a number of nodes are observed in the beam profile, which arise from off-axis light rays being reflected between the end mirrors, for example, the TEM_{11} mode has two nodes which are mutually perpendicular (Figure 2b).

The propagation characteristics of a Gaussian beam can be fully defined, either by the diameter of the beam waist or by the far-field divergence.

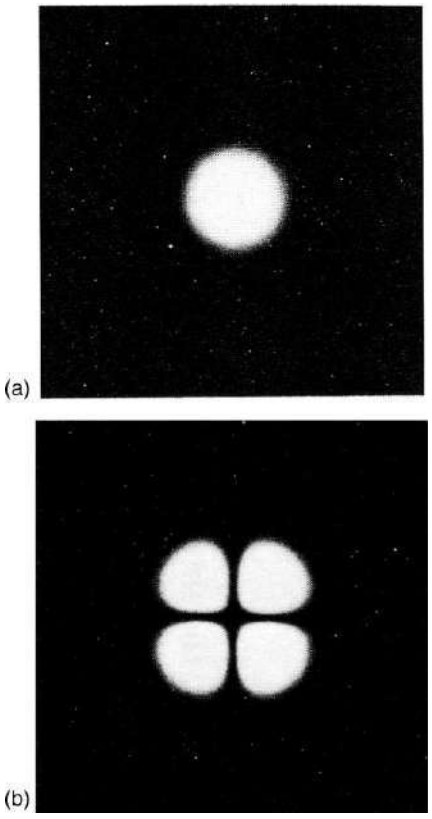

Figure 2 Transverse electromagnetic modes formed with confocal concave mirrors, (a) TEM_{00}, and (b) TEM_{11}. (Reproduced with permission from Young M (1977) *Optics and Lasers: an Engineering Approach.* Berlin, Heidelberg: Springer-Verlag

Consequently, it is only necessary to know the diameter of the beam waist ($2w_0$), or the diameter of the beam ($2w_z$) at a longitudinal distance z from the waist, in order to determine the propagation characteristics of the beam:

$$w(z) = w_0\sqrt{1 + \left(\frac{z}{z_R}\right)^2} \qquad [4]$$

$$R(z) = z\left[1 + \left(\frac{z_R}{z}\right)^2\right] \qquad [5]$$

where the quantity $z_R = \pi w_0^2/\lambda$ is known as the Rayleigh range of the beam, λ is the wavelength of the laser radiation, and $R(z)$ is the radius of curvature of the wavefront at a distance z from the beam waist.

The wavefront is planar at the minimum beam waist and the Rayleigh range is the distance from the beam waist to the location at which the wavefront is most curved (Figure 3), the region from the waist to

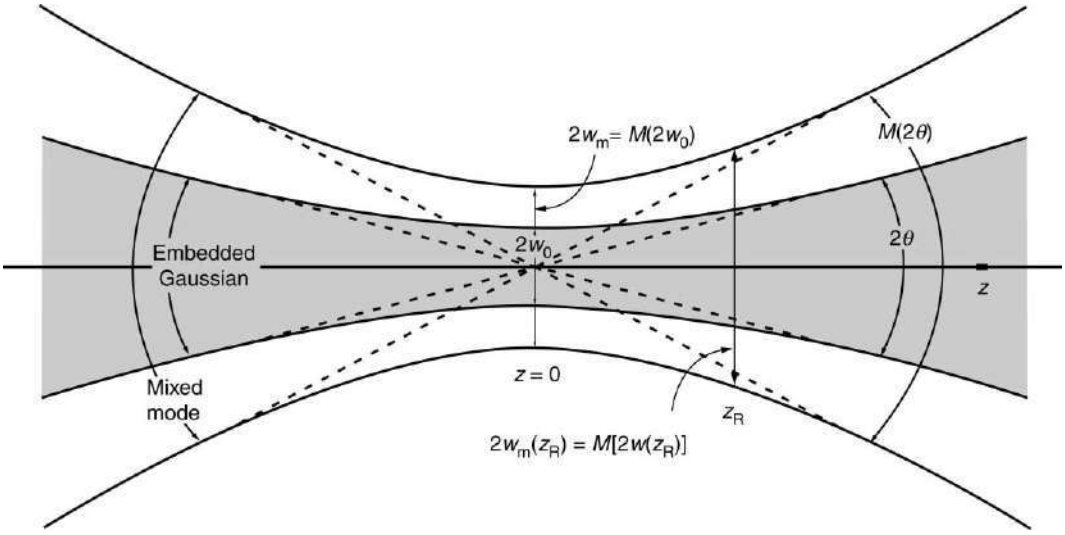

Figure 3 The Rayleigh range and embedded Gaussian (adapted from http://beammeasurement.mellesgriot.com/tut_m2.asp).

the Rayleigh range being the near field. Beyond approximately ten times the Rayleigh range, in the far field, the beam diverges as a cone with approximately straight sides. It can be seen by substituting $z = z_R$ into eqn [4] that the beam diameter at the Rayleigh range is $\sqrt{2}$ times the waist diameter.

Unfortunately laser beams do not conform to pure Gaussian functions and therefore they do not propagate according to the above equations. As a result, a dimensionless beam propagation parameter was devised in the early 1970s. This parameter, known as the M^2 factor or the 'times diffraction limit', is based on the brightness theorem, which states that for any laser beam the product of the beam diameter and the far-field divergence is a constant. Thus, M^2 is defined as the ratio of the laser beam's multimode diameter-divergence product to the ideal diffraction-limited (TEM_{00}) beam diameter-divergence product:

$$M^2 = \left(\frac{2w_m\,\Theta_m}{2w_0\,\theta_0}\right) \qquad [6]$$

where Θ_m is the laser beam's multimode divergence, θ_0 is the theoretical diffraction-limited divergence, $2w_m$ is the laser beam's multimode waist diameter, and $2w_0$ is the ideal diffraction-limited beam waist diameter.

Alternatively, M^2 can be defined as the square of the ratio of the multimode beam diameter ($2w_m$) to the diffraction-limited beam diameter ($2w_0$):

$$M^2 = \left(\frac{2w_m}{2w_0}\right)^2 \qquad [7]$$

The intensity of the beam has a Gaussian distribution in the radial direction, but the accepted definition of beam diameter is the distance at which the intensity has fallen to $1/e^2$ (i.e., ca. 13.5%) of its axial intensity, as can be seen by setting $r = w$ in eqn [3] above. This definition only applies properly to Gaussian beams; for other beam profiles the diameter can be calculated using a second moment measurement of the entire beam.

The real beam, then, can be treated as Gaussian by substituting an artificial wavelength $M^2\lambda$ into the equations that apply to an ideal diffraction-limited TEM_{00} beam. The ideal TEM_{00} Gaussian can be superimposed as an 'embedded Gaussian' on the optical diagram of the real multimode beam (Figure 3). It can be seen that the multimode beam has a waist diameter and a divergence that are both M times larger than those of the embedded Gaussian.

The M^2 factor has a value of unity for an ideal diffraction-limited TEM_{00} beam, and values of greater than unity for all other beams; it can be as high as several hundred for a distorted beam of poor quality. This is an important parameter where a tightly focused laser spot (e.g., confocal Raman microscopy) or low divergence (e.g., remote sensing) are required.

Polarization

Although polarization is not an inherent property of laser light, lasers provide plane polarized light when the end windows of the laser tube are mounted at Brewster's angle (θ_B):

$$\theta_B = \tan^{-1} n \qquad [8]$$

where n is the refractive index of the window material for the appropriate wavelength of laser light. For this angle, reflected light is completely plane polarized and consequently the laser light that is generated is also plane polarized. Brewster windows are often used inside the laser cavity in order to reduce reflection losses. This is important in low-gain lasers such as HeNe lasers where laser action could be prevented by reflection losses. Air-cooled ionic gas lasers typically deliver highly linearly polarised output with ratios exceeding 1000:1.

The polarization property of the exciting light is used a great deal in Raman spectroscopy. In combination with the polarization properties of the Raman scattered light, it can be used to determine the depolarization ratios, ρ, of Raman bands in the solution state. In solid state studies it can be used to determine the degree of orientation, for example, in polymeric fibers or films, and in single crystal studies it can be used to determine the symmetry classes of the Raman active vibrations. In such studies a linear polarizer is used to improve the degree of plane polarization of the laser light, and a half-wave plate is used to rotate the plane of polarization by 90°.

Plane polarized laser light can be converted to circularly polarized laser light which is required for Raman optical activity (ROA) studies. For example, plane polarized 514.5 nm laser light provided by an argon ion laser can be converted to circularly polarized light by an electro-optic modulator. As CCD detectors are usually employed in present-day ROA experiments and they have relatively long sampling times, slow polarization modulation is required which can be achieved by periodically rotating a quarter-wave plate or by applying a square quarter-wave voltage to a Pockels cell.

Whereas optical isomers or enantiomers give identical Raman spectra, they are distinguished in the ROA experiment by the sign of the ROA signal. Indeed, the sign can be used to determine the absolute configuration of a chiral molecule. Also the enantiomeric excess of mixtures of enantiomers can be determined in such experiments, with obvious applications in the pharmaceutical industry, and the conformations of biological molecules, for example, proteins, nucleic acids, sugars, and viruses, can be determined.

Up until now, the ROA technique has been practised by no more than a handful of research groups world-wide, but it is expected that this situation will change in the near future, since the first commercial ROA spectrometer (the ChiralRAMAN spectrometer) has recently been launched by Biotools Inc. This spectrometer makes use of a solid state laser as the source of the polarized laser light.

Beam Divergence

Referring to Figure 3, the full divergence angle, Θ, for the fundamental TEM_{00} Gaussian beam is given by:

$$\Theta = \lim_{z\to\infty} \frac{2w(z)}{z} = \frac{2\lambda}{\pi w_0} \qquad [9]$$

From eqn [4] it can be seen that the spot size increases linearly with the longitudinal distance z and, from eqn [9], that it diverges at a constant cone angle, Θ. It can also be seen from eqn [9] that the smaller the spot size, w_0, the greater the beam divergence angle, Θ.

Indeed, a highly divergent beam is a problem with edge-emitting laser diodes due to the small dimensions of the light source. Furthermore, it is much smaller in the vertical direction than in the horizontal, giving rise to an elliptical output beam that diverges much more rapidly in the vertical direction than in the horizontal. The highly divergent, elliptical beam can be corrected, to an extent, with a cylindrical lens, but the inherent problem of a small, elliptical source can never be completely corrected.

In contrast, vertical-cavity surface-emitting semiconductor lasers (VCSELs) do not have the same limitations on beam divergence because the cavity is in the vertical direction and the light emission occurs from the surface, which has a much larger area than the light source of an edge emitter. Indeed, frequency-doubled green and blue semiconductor lasers are now available, having beams with M^2 values of less than 1.1 and beam divergences of a few milliradians. In terms of these beam characteristics, it appears that these lasers are beginning to challenge the argon ion laser.

Emission Linewidth

The laser linewidth limit, $\Delta\nu_L$, is given by the Schawlow–Townes expression:

$$\Delta\nu_L = \frac{2\pi h\nu(\Delta\nu_c)^2}{P_{out}} \qquad [10]$$

where $\Delta\nu_c$ is the linewidth of the passive resonator, $h\nu$ is the photon energy, and P_{out} is the laser power output.

The power output, P_{out}, is equal to the number of photons, N_p, in the resonator times the energy per photon, $h\nu$, divided by the photon lifetime, τ_c:

$$P_{out} = \frac{N_p h\nu}{\tau_c} \qquad [11]$$

Furthermore, the width $\Delta\nu$ of the resonance curve, at which the intensity has fallen off to half the maximum

value, is given by:

$$\Delta\nu_c = (2\pi\tau_c)^{-1} \quad [12]$$

Substituting for P_{out} and $\Delta\nu_c$ from eqns [8] and [9] into eqn [10] gives the limit for the linewidth, $\Delta\nu_L$:

$$\Delta\nu_L = \frac{\Delta\nu_c}{N_p} \quad [13]$$

The linewidth of a laser is characterized by its Q factor:

$$Q = \frac{\nu}{\Delta\nu} \quad [14]$$

where ν is the frequency of the laser line and $\Delta\nu$ is the linewidth.

Other useful relationships involving linewidth parameters are:

$$Q = \frac{\lambda}{\Delta\lambda} \quad [15]$$

and

$$\Delta\tilde{\nu} = -\frac{\Delta\lambda}{\lambda^2} \quad [16]$$

where λ and $\Delta\lambda$ are the wavelength and linewidth (in units of metres), respectively, and $\Delta\tilde{\nu}$ is the linewidth expressed as a wavenumber (in units of cm^{-1}).

The Q factor can be as high as 10^8 which is of great use in high resolution spectroscopy.

Visible Lasers

Helium-Neon Laser

The helium-neon laser is continuously pumped electrically using a high dc voltage of up to 1 kV. The gain medium consists of a gas mixture of about 10 parts helium to each part of neon at a pressure of about 3 Torr. The gain of this medium is extremely low, hence the requirement for Brewster-angle windows to eliminate reflection loss of the light polarized in the plane that includes the axis of the laser and the normal to the Brewster window. Due to this requirement, the light output is necessarily plane polarized.

The pumping excites helium atoms by electron impact, and resonant energy transfer to neon atoms then occurs via collisions of the gaseous atoms (Figure 4). This creates a population inversion in the neon atoms, enabling the laser transition to occur at 632.8 nm. Following emission, the neon atoms decay nonradiatively down to their metastable $2p^5 3s^1$ level from which they decay back down to the ground state via collisional de-excitation with the walls of the tube. This mechanism restricts the maximum achievable power output to ca. 50 mW, because of the need to depopulate the metastable neon atoms by wall collisions.

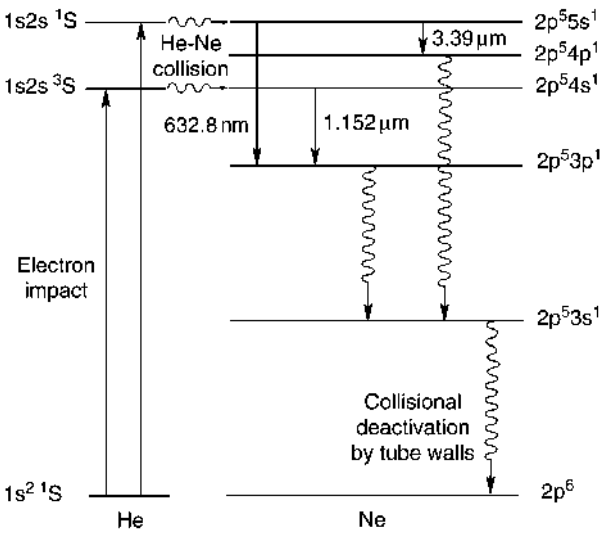

Figure 4 Energy levels of the He-Ne laser.

An advantage of using helium-neon lasers for Raman spectroscopy is that they generally operate in the fundamental TEM_{00} mode, which is critically dependent on the ratio of the length of the tube to its diameter. For this reason they are designed to have small tube diameters of around a few millimetres and lengths of 0.15–0.50 m, the small diameter also aiding collisional de-activation with the walls due to the relatively large surface to volume ratio of the tube. Another advantage of this laser for Raman spectroscopy is that the linewidth of the 632.8 nm emission line is ca. 1.5 GHz.

A disadvantage of the helium-neon laser is that it emits a large number of spontaneous emission lines originating from the excited neon atoms. These plasma emissions are observed in the Raman spectrum as sharp lines unless they are filtered out, for example, by a pre-monochromator or an interference filter. It should be mentioned, in passing, that these plasma lines are not always unwanted, because they can be very useful for wavenumber calibration of the Raman spectrometer.

The low output power of <50 mW is not generally a disadvantage for Raman microscopy since higher laser powers can often cause photolytic or thermal degradation of the sample. This is due to the high irradiance at the sample when the laser light is focused to a tight spot by the microscope objective. The excitation wavelength of 632.8 nm is suitable for combined use of the laser with a silicon-based CCD

detector in Raman spectroscopy, because large Stokes shifts in the 3000–4000 cm^{-1} region lie well within the quantum efficiency curve of this detector. It is also possible, though less common, to operate the helium-neon laser on weaker transitions that include the green 543.5 nm line.

Argon Ion Laser

The argon ion laser was one of the first lasers to be discovered following the invention of the laser and up until now it has been used extensively for Raman spectroscopy, among many other applications. Excitation is provided by a continuous electrical discharge, and because of the high energy required to ionize the argon atoms and then promote the ions to an excited state, the efficiency of the laser is very small (ca. 0.1%). In spite of this, once population inversion has been achieved, the gain of the laser is very high and output powers of up to 25 W can be obtained for the strong lines at 488.0 and 514.5 nm and up to 50 W for multiline operation.

An advantage of the argon ion laser is that it can provide emission at more than 35 discrete wavelengths, the strongest of which are listed in Table 1. These lie in the green, blue, and near ultraviolet regions of the spectrum, a number of the ultraviolet lines only being obtained on the larger frame argon ion lasers. Discrete laser emission lines are selected by tuning a prism or grating inside the cavity. The argon ion laser can also be operated in multiline mode, for example, for pumping dye or Ti:sapphire lasers.

Table 1 Continuous wave argon ion laser wavelengths and output powers for the Coherent Inc. Innova 300 argon ion laser system

Wavelength (nm)	*Power (mW)*
228.9	30
238.3	100
244.0	400
248.3	180
257.3	750
275.4	5
300–305.5	20
333.4	40
333.8	30
335.8	20
351.1	200
351.4	60
363.8	240
454.5	140
457.9	420
465.8	180
472.7	240
476.5	720
488.0	1800
496.5	720
501.7	480
514.5	2400
528.7	420

The five lines below 260 nm are frequency doubled.

The linewidths of the argon ion emission lines at 488.0 and 514.5 nm are around 4.0 GHz. The high temperature laser tube has a diameter of approximately 12 mm and a length in the range of 0.1 to 1.8 m. A 240 V three phase power supply and ca. 35 A are required for pumping a medium power 5 W argon ion laser. The tube also requires water cooling at a flow rate of ca. 10 L/min and a pressure of 25 psi because of the large amount of heat dissipation.

For Raman microscopy smaller, air-cooled argon ion lasers, which only require a 240 V single phase power supply, can be used to provide output of a few hundred milliwatts on the 488.0 and 514.5 nm lines. For this application, the excitation is provided by argon ion lasers, which are designed to operate in the TEM_{00} mode.

Until recently, the argon ion laser has been a workhorse as the most common excitation source for Raman spectroscopy, but it is now losing ground to solid state lasers. The principle reasons for this are that the latter are much more efficient and consequently do not in general have three phase power and water-cooling requirements. Nevertheless, argon ion lasers still have a niche when high excitation powers on the order of watts are required in conjunction with a good beam quality ($M^2 < 1.1$), for example, for Raman spectroscopy of gaseous samples. Another advantage of argon ion lasers over solid state lasers is their multiline capability.

Krypton Ion Laser

The krypton ion laser has useful discrete emission lines in the near ultraviolet, blue, yellow, red, and near infrared regions of the electromagnetic spectrum; the strongest lines of a Coherent Inc. Innova 400 krypton ion laser are listed in Table 2.

The argon ion and krypton ion lasers are close relatives, thus the large frame krypton ion laser has similar power and water-cooling requirements to those mentioned above for the argon ion laser. Also, like the argon ion laser, air-cooled models with lower power output are available. The characteristics of the two lasers are also similar; for example, the linewidths of the krypton ion transitions at 530.9, 568.2, and 647.1 nm are about 4.0 GHz, the laser tube has similar dimensions (0.1–2.0 m in length) and the laser can be operated in either TEM_{00} or multimode.

The krypton ion laser is even less efficient than the argon ion laser, consequently lower-power outputs of

Table 2 Continuous wave krypton ion laser wavelengths and output powers for the Coherent Inc. Innova 400 krypton ion laser system

Wavelength (nm)	*Power (mW)*
206.5	4
234	8
337.5–356.4	2000
406.7	900
413.1	1800
415.4	280
468.0	500
476.2	400
482.5	400
520.8	700
530.9	1500
568.2	1100
647.1	3500
676.4	900
720.8	45
752.5	1200
793.1–799.3	300

The 206.5 and 234 nm lines are frequency doubled.

up to about 20 W can be achieved when operated in multiline mode.

Mixed argon and krypton ion laser tubes are also commonly used that provide laser lines originating from both argon ion and krypton ion transitions. As with the individual argon ion and krypton ion lasers, sharp plasma lines, due to spontaneous emission from the rare gas ions, are observed in the Raman spectrum unless they are filtered out. If interference filters are used for this, each laser line requires its own filter that is tailored to the specific emission wavelength.

HeCd Laser

Laser lines at 325 and 442 nm, both capable of providing milliwatts of output power, can be obtained from the helium-cadmium laser. These emission lines result from electronic transitions in free cadmium atoms.

It is obviously advantageous to use blue, rather than longer wavelength, excitation (e.g., provided by the 442 nm line of the HeCd laser) for off-resonance Raman spectroscopy, due to the ν^4 dependence of the Raman light scattering efficiency, provided that the efficiencies of the illumination/collection optics and Raman spectrometer throughput are optimized in the blue region of the spectrum. Unfortunately, far more samples fluoresce when excited with blue light than with red or near infrared radiation, which explains why red or near infrared lasers (e.g., HeNe, semiconductor lasers) are more commonly used for general Raman applications.

Near Infrared Lasers

Although the traditional laser systems for Raman spectroscopy have been the argon ion, krypton ion, and helium-neon lasers for discrete excitation wavelengths, the diode laser has gained popularity over recent years.

The principal advantage of near infrared semiconductor lasers for Raman microscopy is that they generally excite less fluorescence in the Raman spectrum than visible lasers due to their longer wavelength. Commonly available wavelengths are 670, 785, 830, and 852 nm and, of these, the first two can be used in conjunction with silicon-based CCD detectors to give Stokes Raman shifts over the whole range of fundamental vibrations, up to 4000 cm^{-1}. However, for excitation wavelengths of 830 and 852 nm, only Stokes Raman bands in the fingerprint region can be detected with a silicon-based CCD detector because higher wavenumber Raman shifts approach the bandgap of the silicon semiconductor ($\lambda >$ ca. 1050 nm).

The disadvantages of diode lasers are that they cannot supply high power of narrow linewidth in single-mode and they are susceptible to mode hopping which gives rise to a shift in the excitation wavelength. This latter drawback is particularly disadvantageous for Raman spectroscopy because it results in a shift in the wavenumber positions of the Raman bands. When good beam quality is not required, broad stripe, high power laser diodes can be used; these can have output powers that are greater than 1 W but their emission linewidths are equal to ca. 2 nm, which is far too wide for Raman spectroscopic applications. These laser diodes find applications as pumps of solid state lasers, for example, Nd:YAG, Nd:YVO_4 lasers, however. Additionally, amplified stimulated emission (ASE) gives an unwanted background signal that can be very broad (ca. 20–30 nm) and 0.1–1% of the intensity of the laser line. This necessitates the use of bandpass filters in order to reduce this unwanted background.

Although laser diodes having an output of less than 200 mW can operate in TEM_{00} and in single longitudinal modes, mode hopping can occur due to optical feedback or fluctuations in environmental factors and this causes severe broadening of the linewidth. The sensitivity to optical feedback can be greatly reduced, hence the laser linewidth can be narrowed appreciably, by confining the frequency of the photons either in an internal cavity containing a small diffraction grating or in an external cavity. The former types of laser are sometimes called 'distributed feedback' (DFB) lasers because the feedback is distributed over the length of the grating, rather

than occurring all at once at a mirror. The wavelength that is fed back is determined by the period of the grating. Usually, a DFB laser has a grating fabricated into the entire length of the laser. A variation referred to as a distributed Bragg reflector (DBR) laser has a distinct grating fabricated into the substrate on each side of the active area. The external cavity semiconductor laser (ECSL) is fabricated by placing the laser diode in a separate resonator like a conventional gas or solid-state laser. The DBR and ECSL lasers are now discussed.

Distributed Bragg Reflector Lasers

The DBR laser consists of a grating on each side of the active region (Figure 5); these gratings act as mirrors having a reflectivity that is optimized at one particular wavelength, in addition to narrowing the laser linewidth. At present, DBR lasers are only commercially available having an excitation wavelength of 852 nm.

The emission linewidth is less than 4 MHz, which is suitable for the vast majority of Raman spectroscopic applications and the laser operates in single TEM_{00} mode. The disadvantages of DBR lasers are that the maximum output power is restricted to around 150 mW, so an optical isolator is usually necessary in order to prevent external facet damage caused by external optical feedback, and only an excitation wavelength of 852 nm is currently available.

External-Cavity Semiconductor Lasers (ECSL)

ECSLs provide higher output power (up to ca. 1 W) and a wider range of excitation wavelengths (630 nm to around 850 nm), as well as being less expensive than DBR lasers. They use the semiconductor chip only as the gain medium and employ an external grating, both as the frequency selector and the reflective mirror. Specific excitation wavelengths are becoming widely adopted in Raman microscopy employing single grating spectrograph in conjunction with CCD detection, such as 670, 785, 830, and 852 nm. The ECSLs have emission linewidths as low as a few MHz, but in general they span the range 2 MHz–30 GHz, and can operate in a nearly diffraction-limited transverse mode.

Three different cavity designs employing a diffraction grating have been discussed in the literature: Littrow type-I, Littrow type-II, and Littman configurations. In the Littrow type-I design, the laser diode light is incident on the grating at an angle of incidence equal to $\theta_{Littrow}$ (Figure 6a). The diffracted light in first order is re-directed back to the laser diode to provide feedback and the output is the zeroth order (i.e., specularly reflected) light. A disadvantage of this design is that the bandwidth of the grating is relatively large because a single pass geometry is used. In the Littrow type-II design, light from the rear facet of the laser diode is collimated by a lens and directed on to a grating at an angle of incidence equal to $\theta_{Littrow}$ (Figure 6b). The diffracted light in first order is re-directed back to the laser diode, in a similar fashion to the Littrow type-I design, in order to provide optical feedback, and the output is the laser light emitted from the front facet. Disadvantages of the Littrow type-II design are the requirement for

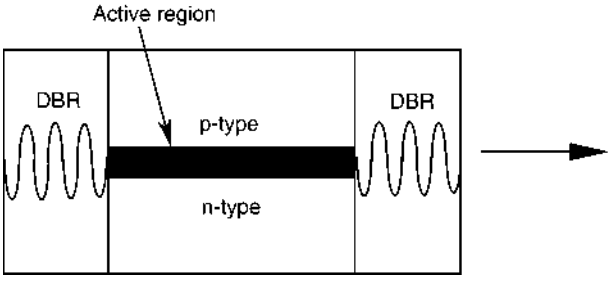

Figure 5 Diagram of the cavity of the DBR laser (adapted from Pan M-W, Benner RE and Smith LM (2002) Continuous lasers for Raman spectrometry. In: Chalmers JM and Griffiths PR (eds) *Handbook of Vibrational Spectroscopy*, 1. Chichester, UK: Wiley).

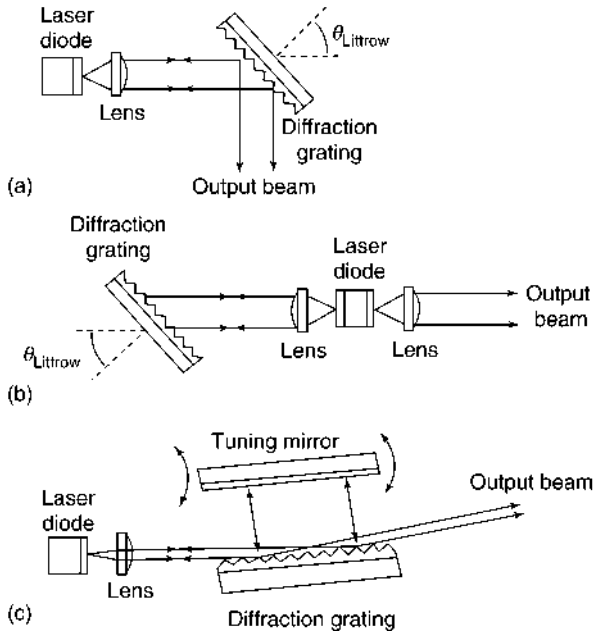

Figure 6 Diagrams of the cavities of (a) Littrow type-I, (b) Littrow type-II, and (c) Littman external cavity diode lasers (adapted from Pan M-W, Benner RE and Smith LM (2002) Continuous lasers for Raman spectrometry. In: Chalmers JM and Griffiths PR (eds) *Handbook of Vibrational Spectroscopy*, 1. Chichester, UK: Wiley).

custom fabrication of antireflection coatings on both facets and the sensitivity to external optical feedback. In the Littman design, the laser diode light is collimated by a lens and directed on to a grating at a grazing angle of incidence. The diffracted light is reflected by a mirror and diffracted back to the laser diode by the grating in order to provide optical feedback (Figure 6c). An advantage of this design is that the grating bandwidth is less than half that of the Littrow designs due to the double pass geometry, but a disadvantage is that the external cavity length is longer, resulting in a narrower spacing of the cavity modes. Another advantage of this design is that the tuning is accomplished by rotating the mirror instead of the grating, thus the alignment of the output beam is not altered when tuning. For Littrow type I and II and Littman ECSL designs, filtering of the ASE is required, as it can have an intensity of ca. 0.1 to 1% of the intensity of the laser line. Thus, these designs necessitate the incorporation of a bandpass filter (having a rejection of better than 10^{-4} at the ASE).

Nd:YAG Laser

Under normal conditions, the Nd:YAG laser oscillates on a transition in the near infrared at 1064 nm, and this is the excitation line that is used in most commercial Fourier Transform Raman spectrometers. The gain medium is a crystal of yttrium aluminium garnet ($Y_3Al_5O_{12}$, YAG) doped with ca. 1.0 mol% Nd^{3+} cations that substitute Y^{3+} ions in the cubic YAG lattice.

The Nd:YAG laser is a four-level system (Figure 7), which has high gain and low threshold due to the narrow fluorescent linewidth of the laser transition. Absorption bands of the Nd^{3+}ions around 808 nm conveniently match the energy of commercially available high-power multimode diode lasers that serve as the pump. The Nd^{3+} ions decay nonradiatively to the upper laser level, the $^4F_{3/2}$ state, thereby creating a population inversion. This is because the lower laser level, the $^4I_{11/2}$ state, has no appreciable thermal population at room temperature, since it is $>2000\ cm^{-1}$ above the $^4I_{9/2}$ ground state, and the $^4F_{3/2}$ excited state has a relatively long lifetime of 230 μs.

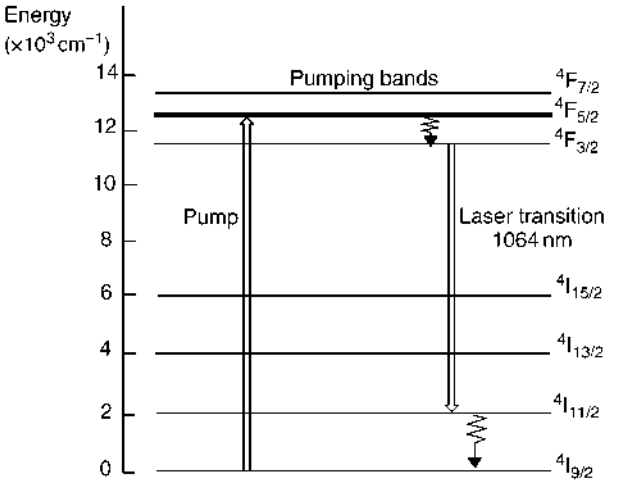

Figure 7 Energy levels of the Nd:YAG laser.

The high thermal conductivity of Nd:YAG enables the laser to operate at high power in either continuous wave (CW) or *Q*-switched modes, and the diode pumped solid state (DPSS) variety of the cw Nd:YAG laser can currently achieve an output power in excess of 20 W on the 1064 nm line. The laser can be designed to operate in the fundamental TEM_{00} mode and the full width at half maximum (FWHM) linewidth of the spontaneous emission of the 1064 nm laser transition is 120 GHz (ca. 0.45 nm). Advantages of the diode pumped Nd:YAG laser are that it is air-cooled and can be operated at 240 V single phase. Also, being all solid state and having a small footprint, it is robust and portable. The lifetime of the laser is dependent on the laser diodes used for pumping, but some Nd:YAG laser designs enable these to be replaced in the field.

The frequency-doubled Nd:YAG laser emitting green light, having a wavelength of 532 nm, is also commonly used nowadays for dispersive Raman spectroscopy, and an output power of 10 W on the 532 nm laser line is provided by commercially available frequency-doubled diode-pumped Nd:YAG lasers.

It is worth mentioning that Nd:YAG lasers have also been fabricated that emit either at 946 or 1330 nm, by suppressing the strong emission at 1064 nm and optimising the optics for the desired wavelength.

Nd:YVO_4 and Nd:YLF Lasers

An intracavity, frequency-doubled DPSS Nd:YVO_4 laser has recently become commercially available, and it has some advantages over the Nd:YAG laser including a larger stimulated emission cross-section and a higher absorption coefficient (along the extraordinary direction of the birefringent crystal). It has the same emission wavelength as the frequency-doubled Nd:YAG laser, i.e., 532 nm, and Nd:YVO_4 is the material of choice for cw end-pumped lasers having around 5 W output power. This is because the diode laser pump beam is tightly focused in the end-pumped system, but a small waist diameter cannot be retained over a distance of more than a

few millimeters, consequently a high absorption coefficient and gain are very beneficial.

The gain medium of the Nd:YLF laser consists of a crystal of yttrium lithium fluoride (YLF) that is doped with Nd^{3+} ions on the Y^{3+} cation sites. Unlike the Nd:YAG and Nd:YVO_4 lasers, the emission does not occur at 1064 nm; instead the $^4F_{3/2}$–$^4I_{11/2}$ emission occurs at wavelengths of either 1047 or 1053 nm, depending on the polarization that is selected. The former is due to extraordinary polarized light, whereas the latter is due to ordinary polarization, and either of these emission wavelengths can be selected using an intracavity polarizer. The Nd:YLF laser can offer benefits in *Q*-switched operation when the longer fluorescence lifetime (480 μs) of Nd^{3+} ions in the $^4F_{3/2}$ state enables a higher energy to be stored for the same number of pump laser diodes.

Ti:Sapphire Laser

The Ti:sapphire laser is tunable over the approximate range of 670–1070 nm with the peak of the gain curve at ca. 800 nm. In the gain medium, ca. 0.1% by weight Ti^{3+} is doped into a crystal of sapphire grown by the Czochralski method. The Ti^{3+} ions substitute for Al^{3+} ions in the Al_2O_3 of the sapphire and the laser emission is due to the ^2E–$^2T_2$ transition of the Ti^{3+} cation, which has a $3d^1$ valence electronic configuration (Figure 8). The laser has a large stimulated emission cross-section but the fluorescence lifetime of the upper laser level (^2E state) is quite short (3.2 μs), thus the laser is usually laser (e.g., by argon ion or frequency-doubled Nd:YAG or Nd:YVO_4 lasers) rather than flashlamp pumped because a very high pump flux is required.

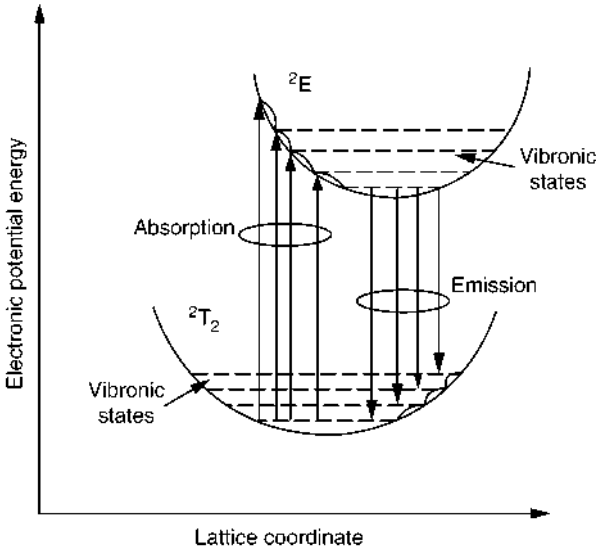

Figure 8 Energy levels of the Ti:sapphire laser (adapted from Pan M-W, Benner RE and Smith LM (2002) Continuous lasers for Raman Spectrometry. In: Chalmers JM and Griffiths PR (eds) *Handbook of Vibrational Spectroscopy*, 1. Chichester, UK: Wiley).

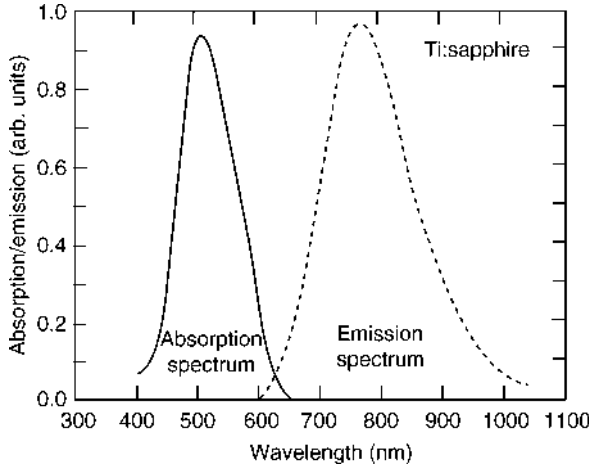

Figure 9 Absorption and emission spectra of the Ti^{3+} ion in sapphire. Al_2O_3 (adapted from Pan M-W, Benner RE and Smith LM (2002) Continuous lasers for Raman Spectrometry. In: Chalmers JM and Griffiths PR (eds) *Handbook of Vibrational Spectroscopy*, 1. Chichester, UK: Wiley).

The absorption and emission bands are broad and widely separated, due to the vibronic coupling between the Ti^{3+} host and the Al_2O_3 lattice (Figure 9). The lower laser level is any one of the vibronic levels of the $^2T_2$ state. Following the laser transition, the Ti^{3+} ions decay from the upper vibronic levels of the $^2T_2$ state down to the lower vibronic levels. Hence, this is a four-level laser.

The broad spontaneous emission linewidth of the ^2E–$^2T_2$ laser transition is around 100 THz, the ouput power can be as high as 50 W, and the laser can be operated in either TEM_{00} or multimodes.

UV Lasers

UV laser sources offer numerous advantages over visible laser sources for Raman spectroscopy. A major consideration is that many analytes have absorptions in the near UV, making them amenable to resonance Raman spectroscopic studies. The signal enhancement (up to 10^6), that can be achieved in resonance Raman spectra, results in a large increase in detection sensitivity. Furthermore, some vibrational modes, which normally give rise to weak bands in the off resonance Raman spectrum, can show strong enhancement in the UV excited resonance Raman spectrum. A good example of this is the amide II band of peptides and proteins, which is due to a combination of N–H in plane bending and C–N stretching modes of the peptide linkage. This band is normally weak in the Raman spectrum but strong in the infrared spectrum; however, the amide II band

can show strong enhancement in the UV excited resonance Raman spectrum and this can be useful for determining secondary structure in proteins. A further benefit of UV excited Raman spectroscopy is that fluorescence can often be avoided as it tends to occur at a lower energy, outside the Stokes Raman spectral window.

It can be advantageous to use cw rather than pulsed excitation for UV resonance Raman spectroscopy, if the analyte has a long excited state lifetime relative to the pulsewidth of the pulsed laser. This is because the concentration of analyte molecules in the ground state is significantly depleted during pulsed excitation, due to Raman saturation giving a lower signal to noise ratio in the Raman spectrum than for the case of cw excitation. It has been found that pulse energy flux densities should be less than 1 mJ/cm^2 and flowing or spinning samples should be used, in order to ensure that nonlinear phenomena and saturation effects do not occur.

For UV Raman microscopy, where the laser beam is focused into a small spot of micrometer dimensions on the sample, it is essential to avoid using pulsed lasers delivering high peak powers. This is because such pulsed lasers can cause dielectric breakdown, and even at lower peak powers they can cause nonlinear effects and Raman saturation phenomena. Consequently, with UV Raman microscopy, one is restricted to the use of cw or quasicontinuous laser excitation.

It is only recently that high thoughput UV Raman microspectrometers have been developed. It has been demonstrated for visible Raman spectroscopy that a significantly higher throughput can be achieved by employing a single stage spectrograph with a notch filter instead of a double or triple monochromator. In the UV region, blocking the Rayleigh scattering is a problem because notch filters are not currently available. However, it has been discovered that this can be overcome by introducing two modifications to the design of the optical layout of a visible Raman microscope. First, two novel dielectric longpass filters were used instead of a notch filter in order to reject the elastically scattered light. Second, an all-reflecting Cassegrain microscope objective was used, instead of a lens, in order to block the specular reflection, and thereby further reduce the Rayleigh scattering background. These design modifications have enabled a single stage spectrograph with a holographic grating to be used to disperse the Stokes Raman radiation in a UV Raman microspectrometer (Figure 10).

In the optical layout of Figure 10 it can be seen that the laser excitation is not focused by the Cassegrain microscope objective because an epi-illumination configuration is not employed. Instead the laser is

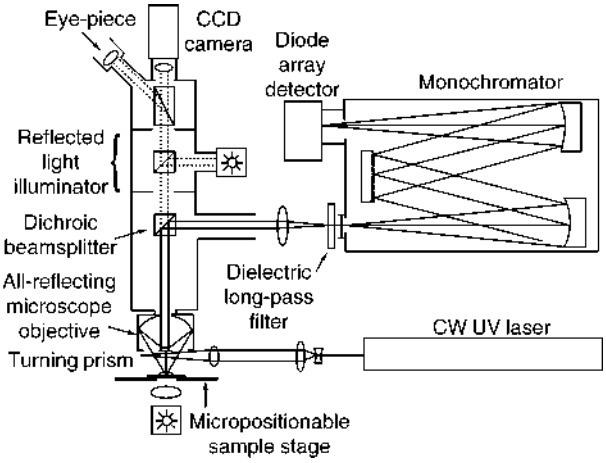

Figure 10 Diagram of the optical layout of a UV Raman microscope (adapted from Pajcini V, Munro CH, Bormett RW, Witkowski RE and Asher SA (1997) UV Raman microspectroscopy: Spectral and spatial selectivity and simplicity. *Applied Spectroscopy* 51: 81–86).

focused by a separate lens and directed on to the sample with a turning prism located directly underneath the Cassegrain objective. This minimizes loss of laser beam throughput or scattered light throughput to the spectrometer at the expense of spatial resolution, since the laser light is focused by a longer focal length lens than the microscope objective. An additional difference between the optical layouts of the visible and UV Raman microscopes shown in Figures 1 and 10, respectively, is that pinhole spatial filters are not shown in the latter. However, better axial spatial resolution could be achieved if this UV Raman microscope was made confocal by introducing a pinhole and lens.

Frequency-Doubled Argon Ion Laser

The frequency doubled argon ion laser is a popular choice among UV laser sources. The cw UV laser contains a nonlinear optical beta-barium borate (BBO) crystal, which is located within the laser cavity; this crystal frequency doubles the strong Ar^+ lines to give five UV lines below 260 nm which are listed in Table 1 for the Coherent Inc. Innova 300 Ar^+ ion laser system. Of these UV lines, the 228.9 nm line is almost ideal for resonance Raman enhancement of tyrosine and trpytophan residues in proteins and the 244.0 nm line is well suited for studying tyrosinate groups. The 244.0 nm line has also been used to excite selectively UV resonance Raman spectra from spatially resolved areas of biological samples within the nucleus of a single cell, from DNA in particular, using low laser powers and short acquisition times to keep sample damage to a minimum under the microscope.

Frequency-Doubled Krypton Ion Laser

Like the cw frequency-doubled argon ion laser mentioned above, the cw frequency-doubled krypton ion laser contains a nonlinear optical BBO crystal, which is located within the laser cavity; this crystal frequency doubles the strong Kr^+ lines to give 206.5 and 234 nm UV lines which are listed in Table 2 for the Coherent Inc. Innova 400 Kr^+ ion laser system. The 206.5 nm laser line is useful for exciting resonance Raman spectra within the $\pi-\pi^*$ amide transition of the peptide backbone of proteins, and the enhanced protein amide vibrational bands can be used to determine the protein secondary structure.

HeCd Laser

The 325 nm HeCd excitation line, in the near ultraviolet, can be used for combined micro-Raman/photoluminescence studies of, for example, semiconductors. These lasers are suitable for low-power applications, typically having output powers in the 1–100 mW range, and they can be designed to operate in TEM_{00} single-mode or multimode.

Quasi-CW Mode-Locked Ti:Sapphire Laser

The second harmonic of the mode-locked Ti:sapphire laser (ca. 350–500 nm), third harmonic (ca. 233–333 nm), and fourth harmonic (ca. 200–250 nm), are all available with conventional nonlinear crystals.

Typically, the Ti:sapphire crystal is pumped by a cw argon ion laser, for example, as the excitation source for a UV Raman microscope. This quasicontinuous laser system produces 2–3 ps pulses at a 76 MHz repetition rate and is continuously tunable over the 200–300 nm range. The typical power output is ca. 50 mW at 250 nm, ca. 10 mW at 240 nm, and 1 mW at 230 nm. In future, it is anticipated that a frequency-doubled Nd:YAG laser will be increasingly favored as the pump, creating an all-solid-state laser source.

Conclusions

Laser systems for Raman microscopy have been described in this article. Although the differences in the terms 'Raman spectroscopy' and 'Raman microscopy' only refer to whether a macro or micro sampling configuration is used, this does have a bearing on the types of laser systems that are employed. This is because the latter imposes a restriction on using low duty pulsed lasers with high pulse peak power, as both spatial and temporal confinement of the laser excitation can lead to dielectric breakdown of the sample, nonlinear phenomena, or Raman saturation. Thus, for this reason, one is limited to cw lasers or quasi-cw lasers for Raman microscopy whereas there is no such restriction for Raman spectroscopy.

The quality of the laser beam, measured by the M^2 value or 'times diffraction limit' influences the ultimate spatial resolution that can be achieved in Raman microscopy.

In the visible region, argon ion, krypton ion, and frequency-doubled Nd:YAG lasers have good beam quality and high power, with helium-neon and helium-cadmium lasers also having good beam quality but lower power; this makes these lasers good light sources for visible Raman microscopy.

For FT Raman microscopy using Nd:YAG excitation, there is a lack of fluorescence interference from a wide variety of samples, but sensitivity and spatial resolution are compromised in comparison to visible dispersive Raman microscopy.

Red and near infrared semiconductor lasers operating in TEM_{00} single mode can have good beam quality, approaching that of gas lasers, but they are power limited. Obviously, when good beam quality is not necessary, for example, for optically pumping other laser sources, high-power diode lasers can be used in multimode operation. Solid state red and near infrared laser sources are becoming very popular for routine Raman microscopic analysis due to their compactness, robustness, and economical power consumption.

Another reason for the popularity of the red and near infrared diode laser sources for Raman microscopy is that fluorescence is less of a problem due to the lower energy of the light, compared to conventional green 514.5 nm excitation. Although there is a penalty, due to the dependency of the scattering efficiency on the ν^4 term, present day silicon-based CCD arrays can exhibit excellent detection quantum efficiency of Stokes shifted Raman radiation, across the whole vibrational spectrum, when using standard 670 and 785 nm diode lasers and, at least across the fingerprint region, when using 830 and 852 nm diode lasers.

For ultimate spatial resolution in the Raman microscopic experiment, short wavelength excitation is advantageous because it is fundamentally possible to focus to a smaller diffraction-limited laser spot. In this regard, it would be preferable to use UV Raman microscopy where fluorescence is also not as problematic as it is in the visible region. However, in spite of recent improvements in UV laser sources such as the intracavity frequency-doubled argon and krypton ion cw lasers, there are current instrumental limitations on the throughput efficiency of the UV Raman microscope imposed by the lack of availability of

suitable notch filters and low stray light and high throughput optics.

It is anticipated that in the future all-solid-state, quasi-cw, diode pumped, mode-locked Ti:sapphire lasers operating on their third or fourth harmonics will become increasingly popular for providing tunable UV excitation affording good beam quality. It is now also possible to fabricate hollow cathode metal ion deep UV lasers, 10–15 cm in length, 2–4 cm in diameter, weighing 50–100 g and requiring only 2–3 W of electrical power. These lasers have low beam quality (M^2 equal to ca. 18), but they make possible the development of portable UV fluorescence imaging and Raman microprobes for geobiological exploration of terrestrial and extraterrestrial environments.

See also

Fourier Optics. **Imaging:** Infrared Imaging. **Introductory Article:** Early Development of the He-Ne Laser (Title TBC). **Lasers:** Noble Gas Ion Lasers; Semiconductor Lasers. **Nonlinear Optics, Applications:** Raman Lasers. **Scattering:** Raman Scattering.

Further Reading

Andrews DL (1990) *Lasers in Chemistry* (2nd edition). Berlin, Heidelberg: Springer-Verlag.

Asher SA, Bormett RW, Chen XG, *et al.* (1993) UV resonance Raman spectroscopy using a new cw laser source: Convenience and experimental simplicity. *Applied Spectroscopy* 47: 628–633.

Bell SEJ, Bourguignon ESO, O'Grady A, Villaumie J and Dennis AC (2002) Extracting Raman spectra from highly fluorescent samples with "Scissors" (SSRS, Shifted-Subtracted Raman Spectroscopy). *Spectroscopy Europe* 14: 17.

Best SP, Clark RJH and Withnall R (1992) Non-destructive pigment analysis of artefacts by Raman microscopy. *Endeavour* 16: 66.

Boustany NN, Manoharan R, Dasari RR and Feld MS (2000) Ultraviolet resonance Raman spectroscopy of bulk and microscopic human colon tissue. *Applied Spectroscopy* 54: 24–30.

Delhaye M, Barbillat J, Aubard J, Bridoux M and Da Silva E (1996) Instrumentation. In: Turrell G and Corset J (eds) *Raman Microscopy: Developments and Applications*, chap. 3. London: Academic Press.

Delhaye M and Dhamelincourt P (1975) Raman microprobe and microscope with laser excitation. *Journal of Raman Spectroscopy* 3: 33.

Derbyshire A and Withnall R (1999) Pigment analysis of portrait miniatures using Raman microscopy. *Journal of Raman Sprectroscopy* 30: 185.

Gordeyev SA, Nikolaeva GYu, Prokhorov KA, Withnall R, Dunkin IR and Shilton SJ (2001) Super-selective polysulfone hollow fiber membranes for gas separation: assessment of molecular orientation by Raman spectroscopy. *Laser Physics* 11: 82–85.

Hayward CL, Best SP, Clark RJH, Ross NL and Withnall R (1994) Polarised single crystal Raman spectroscopy of sinhalite, $MgAlBO_4$. *Spectrochimica Acta* 50A: 1287–1294.

Hendra P, Jones C and Warnes G (1991) *Fourier Transform Raman Spectroscopy*. London: Ellis Horwood.

Hitz B (2003) Solid state lasers are gunning for argon ion's place. *Photonics Spectra* 54–58.

Holtz JSW, Bormett RW, Chi Z, *et al.* (1996) Applications of a new 206.5 nm continuous wave laser source: UV Raman determination of protein secondary structure and CVD diamond material properties. *Applied Spectroscopy* 50: 1459–1468.

http://www.btools.com

Koechner W and Bass M (2003) *Solid-State Lasers*. New York: Springer-Verlag.

Landsberg G and Mandelstam L (1928) Eine neue Erschienung bei der Lichtzerstreaung in Krystallen. *Naturwiss* 16: 557.

Matousek P, Towrie M, Stanley A and Parker AW (1999) Efficient rejection of fluorescence from Raman spectra using picosecond Kerr gating. *Applied Spectroscopy* 53: 1485.

Nakashima S, Okumura H, Yamamoto T and Shimidzu R (2004) Deep-ultraviolet Raman microspectroscopy: Characterization of wide-gap semiconductors. *Applied Spectroscopy* 58: 224–229.

Pajcini V, Munro CH, Bormett RW, Witkowski RE and Asher SA (1997) UV Raman microspectroscopy: Spectral and spatial selectivity and simplicity. *Applied Spectroscopy* 51: 81–86.

Pallister DM and Morris MD (1993) In: Morris MD (ed.) *Microscopic and Spectroscopic Imaging of the Chemical State*, chap. 1. New York: Marcel Dekker.

Pan M-W, Benner RE and Smith LM (2002) Continuous lasers for Raman spectrometry. In: Chalmers JM and Griffiths PR (eds) *Handbook of Vibrational Spectroscopy,* 1. Chichester, UK: Wiley.

Pan M-W, Benner RE, Johnson CW and Smith LM (2000) Near-IR lasers rejuvenate Raman spectroscopy. *Optoelectronics World* S5–S10.

Raman CV and Krishnan KS (1928) A new type of secondary radiation. *Nature* 121: 501.

Rosasco GJ (1980) Raman microprobe spectroscopy. In: Clark RJH and Hester RE (eds) *Advances in Infrared and Raman Spectroscopy*, vol. 7. Chichester: Wiley.

Rosasco J, Etz ES and Cassatt WA (1975) The analysis of discrete fine particles by Raman spectroscopy. *Applied Spectroscopy* 29: 396.

Schawlow AL and Townes CH (1958) Infrared and optical masers. *Physics Review* 112: 1940.

Smekal A (1923) Sur Quantentheorie der Dispersion. *Naturwiss* 11: 873.

Storrie-Lombardi MC, Hug WF, McDonald GD, Tsapin AI and Nealson KH (2001) Hollow cathode ion lasers for deep ultraviolet Raman spectroscopy and fluorescence

imaging. *Review of Scientific Instruments* 72: 4452–4459.
Turrell G, Delhaye M and Dhamelincourt P (1996) Characteristics of Raman microscopy. In: Turrell G and Corset J (eds) *Raman Microscopy: Development and Applications*, 2. London: Academic Press.
Young M (1977) *Optics and Laser: An Engineering Approach*. Berlin, Heidelberg: Springer-Verlag.
Yu Y-M, Nam S, Byungsung O, *et al.* (2002) Resonant Raman scattering in ZnS epilayers. *Materials Chemistry and Physics* 78: 149–153.

Second-Harmonic Spectroscopy

J I Dadap and T F Heinz, Columbia University, New York, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The demonstration of second-harmonic generation (SHG), one of the early experiments with pulsed lasers, is considered to have marked the birth of the field of nonlinear optics. The importance of second-harmonic generation, and the related phenomena of sum- and difference-frequency generation, as methods for producing new frequencies of coherent radiation was recognized immediately and has in no way diminished over the years. Associated with these applications came an interest in the fundamental nature of the second-order nonlinear response, and some spectroscopic studies were undertaken with this motivation. These investigations were of rather limited scope and were not developed as a general probe of materials. In contrast, the third-order nonlinear optical interactions have given rise to a panoply of significant spectroscopic measurement techniques, including, to name a few, coherent Raman spectroscopy, pump-probe and other four-wave mixing measurements, two-photon absorption and hole burning measurements, photon echoes, and Doppler-free spectroscopy.

A central factor in the lack of spectroscopic applications of the second-order nonlinear response lies in a fundamental symmetry constraint. Unlike the third-order nonlinearity, which is present in all materials, only materials lacking a center of inversion give rise to an (electric-dipole) allowed second-order nonlinear response. As we demonstrate in this article, however, it is this property that makes SHG, as well as the sum-frequency generation process discussed elsewhere in this encyclopedia, such a remarkable probe of surfaces and interfaces. For centrosymmetric materials, the second-order nonlinear response is strong only at interfaces, where only a monolayer or so of material contributes to the response. This yields a purely optical probe (with incident and emitted photons) that gives an inherent surface sensitivity comparable to the best electron spectroscopies, such as electron diffraction, Auger-electron spectroscopy, photoemission spectroscopy, and scanning tunneling microscopy, the workhorses of surface science. Neither linear nor third-order nonlinear optical measurements provide this inherent interface sensitivity. The possibility of probing interfaces by purely optical means permits one to take advantage of several important features of optical radiation. Perhaps most noteworthy is the possibility of probing buried interfaces of every sort, from solid/solid to liquid/liquid, as well as the more conventional cases of solid/vapor, liquid/vapor, and liquid/solid. In addition, laser-based techniques offer unequalled capabilities for spectral or temporal resolution.

In this article, we present a brief overview of SH spectroscopy for the study of surfaces and interfaces. The article is divided into two principal sections: an introductory section that presents some of the fundamentals of SHG and the behavior at interfaces; and a discussion of the different types of information that the method can yield, as illustrated by some representative examples. A brief section on the experimental technique links these two parts.

Fundamentals of Surface SHG

Basic Notions

Within the domain of classical optics, the response of a material to electromagnetic radiation is described by an induced polarization **P** that varies linearly with the electric field strength, **E**. This situation reflects the fact that the strength of electric fields for light encountered under conventional conditions is minute compared to that of the electric fields binding atoms and solids together. The electric field binding an atom may be estimated as $E_a \sim 1\ \text{V/Å} = 10^{10}\ \text{V/m}$. To reproduce this field strength by an electromagnetic wave requires an irradiance of $I_a \sim 10^{17}\ \text{W/m}^2$, so the validity of this linear approximation in classical optics is clear. The advent of the laser, with its capability for producing light beams with very high optical power and irradiance, has permitted us to

probe and exploit the nonlinear response of materials, giving rise to the field of nonlinear optics.

The next-order material response beyond the linear approximation consists of an induced nonlinear source polarization $\mathbf{P}^{(2)} = \varepsilon_0 \boldsymbol{\chi}^{(2)} : \mathbf{EE}$ that depends quadratically on the driving electric field $\mathbf{E}$, described by a second-order nonlinear susceptibility $\boldsymbol{\chi}^{(2)}$. For excitation by a single laser at a frequency ω, this response gives rise to SHG at a frequency of 2ω, as well as optical rectification with a dc induced-polarization. For beams at two distinct frequencies of ω_1 and ω_2, the processes of sum- or difference-frequency generation (SFG or DFG) become operative with the production of the new frequencies of $|\omega_1 \pm \omega_2|$. For a relationship of the given form, which assumes only a spatially local quadratic dependence of the polarization on the driving electric field, one can show that $\boldsymbol{\chi}^{(2)}$ must vanish within any centrosymmetric medium. This is a direct consequence of the fact that $\boldsymbol{\chi}^{(2)}$ as a quantity describing a centrosymmetric medium should not change under an inversion operation, while the quantities $\mathbf{P}^{(2)}$ and $\mathbf{E}$ as polar vectors will change sign. The interface specificity of the SHG technique for centrosymmetric media follows from this general observation: only at interfaces is the bulk symmetry broken and the SHG process allowed.

Anharmonic Oscillator Model

Before we embark on our discussion of surface SHG, we first briefly consider a simple classical model of SHG based on an anharmonic oscillator. Although this is not a realistic treatment of the response of materials, it helps to illustrate the characteristic features of the surface SHG process and its spectroscopic attributes. In the anharmonic oscillator model, the medium is treated as a collection of electrons bound to fixed ion cores. While the usual Lorentz model only includes the effect of a linear restoring force, to describe second-order response, we generalize this description to include a leading-order anharmonic restoring force. For simplicity, we consider motion of the anharmonic oscillator in one spatial dimension, in which case the classical equation of motion for the displacement, x, of the electron can then be written as

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\gamma\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2 x - bx^2 = -\frac{e}{m}E(t) \qquad [1]$$

Here $-e$ and m denote, respectively, the electron charge and mass, and $E(t)$ is the optical driving field. The (angular) frequency ω_0 defines the resonance of the harmonic component of the response, and γ represents a phenomenological damping rate for the oscillator. The nonlinear restoring force corresponds to the term containing the parameter b, a material-dependent constant. We are interested in the solution of this equation for a monochromatic driving field at frequency ω with amplitude $E(\omega)$ which we write as

$$E(t) = E(\omega)\exp(-\mathrm{i}\omega t) + \text{c.c.} \qquad [2]$$

An exact solution to eqn [1] in the presence of this electric field is not possible. Let us, however, consider a regime in which we do not excite the oscillator too strongly, and the effect of the anharmonic term remains relatively small compared to the harmonic one. In this case, we may solve the problem perturbatively in powers of the driving field. We find a linear response at the fundamental frequency and a nonlinear response at the SH frequency, which we can write in terms of the corresponding induced dipole moments $p = -ex$:

$$p(t) = p^{(1)}(\omega)\exp(-\mathrm{i}\omega t) + p^{(2)}(2\omega)\exp(-\mathrm{i}2\omega t) + \text{c.c.} \qquad [3]$$

We can then write the amplitudes for the harmonically varying dipole moments as

$$p^{(1)}(\omega) = \alpha^{(1)}(\omega)E(\omega) \qquad [4]$$

$$p^{(2)}(2\omega) = \alpha^{(2)}(2\omega = \omega + \omega)E(\omega)E(\omega) \qquad [5]$$

where $\alpha^{(1)}$, the linear polarizability, and $\alpha^{(2)}$, the second-order polarizability (or first hyperpolarizability) for SHG are given, respectively, by

$$\alpha^{(1)}(\omega) = \frac{e^2}{m}\frac{1}{D(\omega)} \qquad [6]$$

$$\alpha^{(2)}(2\omega = \omega + \omega) = \frac{(e^3/m^2)b}{D(2\omega)D^2(\omega)} \qquad [7]$$

Here we have introduced the resonant response at the fundamental frequency by $D(\omega) \equiv \omega_0^2 - \omega^2 - 2\mathrm{i}\gamma\omega$, with a corresponding relation for the response $D(2\omega)$ at the SH frequency. A similar derivation can, of course, be applied to obtain the corresponding material response for the SFG or DFG processes.

The spectral dependence of the polarizabilities $\alpha^{(1)}$ and $\alpha^{(2)}$ is illustrated in Figure 1 for the anharmonic oscillator model with a resonance frequency of ω_0. In the linear case, one sees the expected behavior of the loss, associated with the $\mathrm{Im}[\alpha^{(1)}]$, and dispersion associated with $\mathrm{Re}[\alpha^{(1)}]$. A single resonance in the linear response is seen for the driving frequency ω near the oscillator frequency ω_0. For the case of SHG, on the other hand, we observe both resonances when the fundamental frequency $\omega \approx \omega_0$ and when the SH

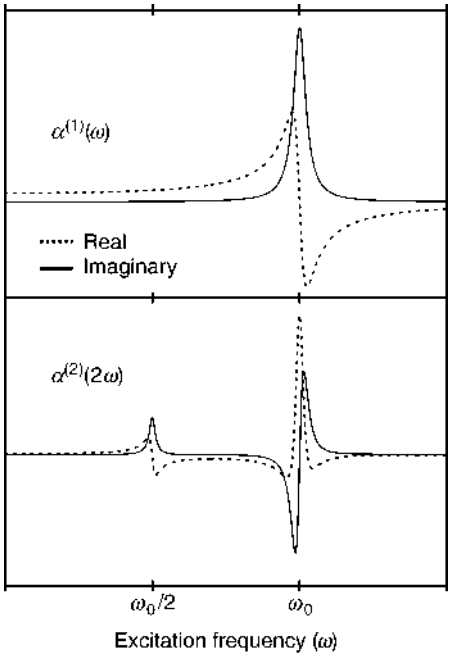

Figure 1 Real and imaginary components of the linear polarizability $\alpha^{(1)}(\omega)$ (top panel) and the second-order nonlinear polarizability $\alpha^{(2)}(2\omega = \omega + \omega)$ (bottom panel) plotted versus frequency according for the anharmonic oscillator model. For the SH response, the resonance features at $\omega = \omega_0$ and $\omega = \omega_0/2$ correspond, respectively, to single-photon and two-photon resonances.

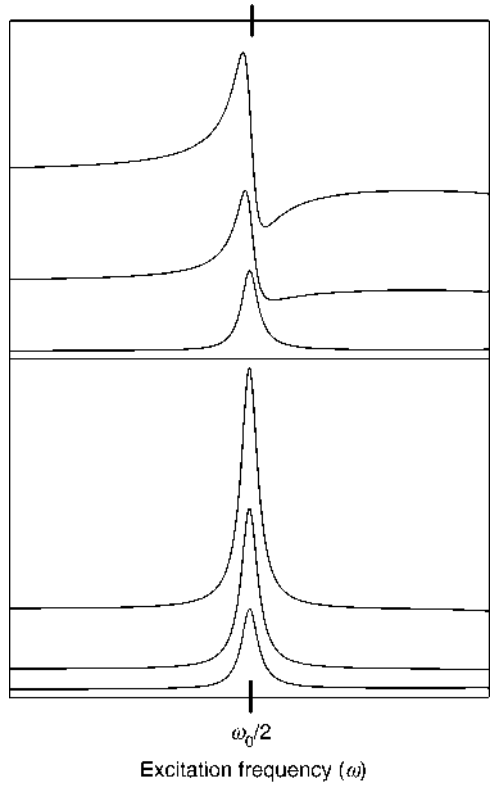

Figure 2 SH spectra $I(2\omega)$ versus pump laser frequency ω corresponding to Figure 1 in the vicinity of the two-photon resonance, with and without the presence of a nonresonant (frequency-independent) background: $I(2\omega) \sim |\alpha^{(2)}(2\omega = \omega + \omega) + k|^2$, where k represents a constant background and parametrized through $k = m|\alpha^{(2)}(2\omega = \omega_0)|$ and m is constant factor. Top panel: $m = 0$, 0.5, 1 (bottom to top curves). Bottom panel: $m = 0$, $0.5i$, i (bottom to top curves). In the top panel, the upper two lineshapes were displaced vertically to distinguish the three curves.

frequency $2\omega \approx \omega_0$. In quantum mechanical language introduced below, these peaks can be considered as arising, respectively, from matching of the fundamental and second-harmonic photon energies with the transition energy in the material.

In SHG spectroscopy, one normally measures the SH power as a function of the laser frequency. The SH field will scale with the nonlinear material response, i.e., with $\alpha^{(2)}(2\omega = \omega + \omega)$, so that the SH power will be proportional to $|\alpha^{(2)}(2\omega = \omega + \omega)|^2$. Figure 2 shows the corresponding spectra in the vicinity of the second-harmonic resonance. An important aspect of this type of spectroscopy, which is also present in other nonlinear spectroscopies, such as coherent Raman measurements, is the role that may be played by a nonresonant background. This may arise either from the off-resonant response of the system under study or from a coherent background from another material. In either case, the effect can be described by adding a spectral flat background response to the nonlinearity. Now when we detect the optical power, we then measure a quantity proportional to $|\alpha^2(2\omega = \omega + \omega) + k|^2$, where k represents the nonresonant background. The presence of k obviously elevates the baseline response near a resonance. In addition, however, depending on the relative phase of k, this extra term can introduce significant changes in the observed lineshapes. Figure 2 illustrates this effect, which must always be borne in mind in the interpretation of experimental spectra.

For the purposes of studies of surfaces and interfaces, the key property of second-order nonlinear processes is the fact that they exhibit, for centrosymmetric bulk media, an inherent interface specificity. This feature can be readily understood in the context of the nonlinear oscillator model in which the SH polarization scales with the material parameter b. Since this term is associated with a potential that varies as x^3, it is clear that the value of the b parameter can be taken as a measure of the material's departure from inversion symmetry. At the interface, such a term can be present, reflecting the inherent

asymmetry at a boundary, while it must necessarily be absent in the bulk of a centrosymmetric material.

Quantum-Mechanical Description of SHG

A correct description of SHG naturally requires a quantum mechanical treatment of the material response to the optical field. For our present purposes, we consider a localized entity with a nonlinear response, such as a non-centrosymmetric molecule at a surface. We consider below how these individual units can be added together to yield the surface nonlinear response of the material. Within this picture, we write the induced nonlinear dipole moment $\mathbf{p}^{(2)}$ of each molecule in terms of the driving electric field $\mathbf{E}$ as

$$\mathbf{p}^{(2)}(2\omega) = \boldsymbol{\alpha}^{(2)}(2\omega = \omega + \omega) : \mathbf{E}(\omega)\mathbf{E}(\omega) \qquad [8]$$

This relation is simply the generalization of eqn [5] to include a full 3-dimensional description, where $\mathbf{p}^{(2)}$ and $\mathbf{E}$ are vectors and the second-order nonlinear polarizability $\boldsymbol{\alpha}^{(2)}$ is a third-rank tensor. The quantum mechanical description enters in how we relate the response function $\boldsymbol{\alpha}^{(2)}(2\omega = \omega + \omega)$ to the underlying properties of the material.

We can obtain an expression for the second-order nonlinear polarizability by application of second-order perturbation theory with the light–matter interaction treated as the perturbation. Within the electric-dipole approximation, the interaction Hamiltonian can be taken as $H_{\text{int}} = -\boldsymbol{\mu}\cdot\mathbf{E}(t)$, where $\boldsymbol{\mu} = -e\mathbf{r}$ denotes the electric-dipole operator and $\mathbf{E}(t)$ the electric field of the driving laser beam. Using the standard density-matrix formalism for second-order perturbation theory, one calculates for the Cartesian components $\alpha^{(2)}_{ijk}$ of the tensor $\boldsymbol{\alpha}^{(2)}$ as

$$\alpha^{(2)}_{ijk}(2\omega = \omega + \omega) = -\frac{1}{\hbar^2}\sum_{g,n,n'}\left[\frac{(\mu_i)_{gn}(\mu_j)_{nn'}(\mu_k)_{n'g}}{(2\omega - \omega_{ng} + \mathrm{i}\Gamma_{ng})(\omega - \omega_{n'g} + \mathrm{i}\Gamma_{n'g})} + \cdots\right]\rho^{(0)}_g \qquad [9]$$

In this expression, the letters g, n', and n represent energy eigenstates $|g\rangle$, $|n'\rangle$, and $|n\rangle$ of the system, with $\rho^{(0)}_g$ corresponding to the thermal population for differing available ground states g. As illustrated in Figure 3, the SHG process can be regarded as involving a series of three transitions: two transitions associated with the absorption of two pump photons, each of energy $\hbar\omega$, and a transition associated with emission of a second-harmonic photon of energy $2\hbar\omega$. These transitions occur through the electric-dipole operator $\boldsymbol{\mu}$ and are characterized by the matrix

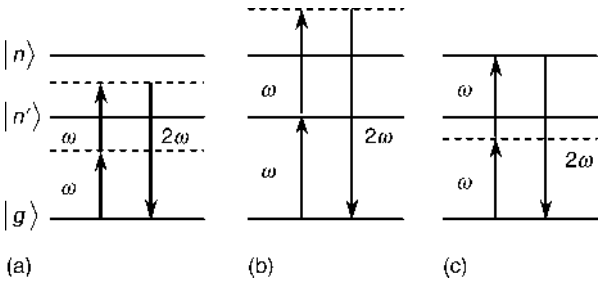

Figure 3 Quantum mechanical description of SHG process for the situations of (a) nonresonant, (b) single-photon resonant, and (c) two-photon resonant interactions. SHG occurs through transition from the initial state electronic state $|g\rangle$ and a first intermediate state $|n'\rangle$, then to a second intermediate state $|n\rangle$, followed by a return to the initial state. The process is just a parametric conversion of two photons at frequency ω into one photon at frequency 2ω.

elements $(\mu_i)_{gn}$. In eqn [9], the energy denominators involve the energy differences $\hbar\omega_{ng} \equiv E_n - E_g$ and widths $\hbar\Gamma_{ng}$ for transitions between eigenstates $|n\rangle$ and $|g\rangle$, and similarly for other combinations of states. In addition to the term indicated explicitly in eqn [9], there are seven additional terms with different Cartesian coordinates in the matrix elements and/or frequency denominators.

The frequency denominators in the eight terms of eqn [9] introduce a resonant enhancement in the nonlinearity when either the fundamental frequency ω or the SH frequency 2ω coincides with a transition from a ground state $|g\rangle$ to one of the intermediate states $|n'\rangle$ or $|n\rangle$. Figure 3a shows the situation for a nonresonant nonlinear response, while Figures 3b,c illustrate, respectively, single- and two-photon resonances. The numerator in the perturbation theory expression consists of products of the three dipole matrix elements of the form $(\mu_i)_{gn}(\mu_j)_{nn'}(\mu_k)_{n'g}$. These terms reflect the structure and symmetry of the material that is built into the third-rank tensor $\alpha^{(2)}_{ijk}(2\omega = \omega + \omega)$.

The Surface Nonlinear Response

As a model of the surface nonlinearity, let us consider a monolayer of oriented molecules. The surface nonlinear response accessible to a macroscopic measurement is given, under the neglect of local-field effects, simply by summing the response of the individual molecules. The resulting surface nonlinear susceptibility tensor $\chi^{(2)}_{\text{s},ijk}$ connecting the induced nonlinear sheet polarization to the driving electric field can then be expressed as

$$\chi^{(2)}_{\text{s},ijk} = \frac{N_\text{s}}{\varepsilon_0}\langle T_{i\lambda}T_{j\mu}T_{k\nu}\rangle\alpha^{(2)}_{\lambda\mu\nu} \qquad [10]$$

Here N_s denotes the adsorbate surface density, $\alpha^{(2)}_{\lambda\mu\nu}$ the nonlinear polarizability of the molecule expressed in its own coordinate system, $T_{i\lambda}$ the transformation tensor from the molecule's coordinate system to the laboratory frame, $\langle\,\rangle$ an ensemble average over the orientation of the different molecules in the monolayer, and $\varepsilon_0 = 8.85 \times 10^{-12}$ F/m the permittivity of free space. This expression for the surface nonlinear susceptibility has been successful in describing properties of adsorbed molecules, including their coverage, orientation, and spectroscopic features.

For the surfaces of materials such as semiconductors or metals with delocalized electronic states, a somewhat different formulation of the nonlinear response is appropriate. A modified version of eqn [9] can be obtained using the relevant band states. The underlying concepts involved in the surface nonlinear response are, however, similar to those presented for the molecular case.

Regardless of the details of the model describing the surface nonlinear susceptibility, its tensor properties must reflect the symmetry of the interface. In that respect $\chi^{(2)}_{s,ijk}$ is quite analogous to the bulk nonlinear response, $\chi^{(2)}_{ijk}$, in a non-centrosymmetric medium. In the absence of any symmetry constraints, $\boldsymbol{\chi}^{(2)}_s$ will exhibit $3 \times 3 \times 3 = 27$ independent elements for SFG and DFG. For SHG, the order of the last two indices has no significance, and the number of independent elements is reduced to 18. If the surface exhibits a certain in-plane symmetry, then the form of $\boldsymbol{\chi}^{(2)}_s$ will be simplified correspondingly. For the common situation of an isotropic surface, for example, SHG possesses three allowed elements of $\boldsymbol{\chi}^{(2)}_s$ and may be denoted as $\chi^{(2)}_{s,\perp\perp\perp}$, $\chi^{(2)}_{s,\perp\|\|}$, $\chi^{(2)}_{s,\|\perp\|} = \chi^{(2)}_{s,\|\|\perp}$, where $\perp$ corresponds to the direction of the surface normal and $\|$ to an in-plane direction. The nonvanishing elements of $\boldsymbol{\chi}^{(2)}_s$ for other commonly encountered surface symmetries are summarized in several textbooks and reviews listed below.

Radiation Properties for Surface SHG

In order to probe and extract information from interfaces through surface SHG, it is necessary to understand how radiation interacts with the relevant media and gives rise to the experimentally observable signals. Here we present the principal results for the usual case of a spatially homogeneous planar interface. Nonplanar and inhomogeneous geometries, which are also of considerable interest, will be considered briefly under the heading of applications of the SHG technique below.

The SHG process is coherent in nature. Thus, for the planar geometry, a pump beam impinging on the interface will give rise to a well-collimated SH radiation emerging in distinct reflected and transmitted directions. To describe this situation in a more detailed fashion and provide formulas for the radiation efficiency, we introduce a general description of SHG by a planar interface excited by a plane wave, as shown in Figure 4. The nonlinear response of the interface described by the surface nonlinear susceptibility tensor $\boldsymbol{\chi}^{(2)}_s$ is incorporated through a nonlinear source polarization $\mathbf{P}^{(2)}$ of

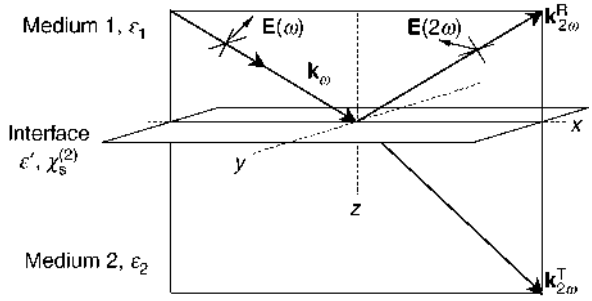

Figure 4 Schematic representation of SHG at the interface between two centrosymmetric media. The pump field $\mathbf{E}(\omega)$ at frequency ω with corresponding wavevector $\mathbf{k}_\omega$ is incident on the interface from medium 1. The SH field $\mathbf{E}(2\omega)$ at frequency 2ω is emitted in directions described by the wavevectors $\mathbf{k}^{R}_{2\omega}$ (reflected in medium 1) and $\mathbf{k}^{T}_{2\omega}$ (transmitted in medium 2). The linear dielectric constants of media 1, 2, and the interface are denoted, respectively, by ε_1, ε_2, and ε'. The nonlinear response of the interface ($z = 0$) is given by the surface nonlinear susceptibility tensor $\boldsymbol{\chi}^{(2)}_s$.

$$\begin{aligned}\mathbf{P}^{(2)}(2\omega) &= \mathbf{P}^{(2)}_s(2\omega)\delta(z) \\ &= \varepsilon_0 \boldsymbol{\chi}^{(2)}_s(2\omega = \omega + \omega) : \mathbf{E}(\omega)\mathbf{E}(\omega)\delta(z) \qquad [11]\end{aligned}$$

localized at the interface ($z = 0$). In addition to the strong nonlinear response at the interface, one must also generally consider the nonlocal nonlinear response of the bulk media. This response is much weaker than the symmetry-allowed interfacial response, but is permitted even in centrosymmetric materials. Since it is present in a much larger volume, however, its cumulative effect, while generally weaker than the surface response, may be of comparable magnitude. For simplicity, however, we neglect this constant background response in our discussion in this article, but a complete discussion can be found in the referenced works. The overall SH response is naturally also influenced by the linear optical properties of the surrounding media. The two bulk media and the interfacial region are characterized, respectively, by frequency-dependent dielectric functions ε_1, ε_2, and ε', as shown in Figure 4. Again for simplicity, and in accordance with most applications, we take the linear response to be isotropic.

Figure 4 shows the incoming pump radiation and the reflected and transmitted SH beams. The directions of these beams are determined by conservation of the component of the momentum parallel to the interface in the SHG process, i.e., $2\mathbf{k}_{\omega,\|} = \mathbf{k}^{\mathrm{R}}_{2\omega,\|} = \mathbf{k}^{\mathrm{T}}_{2\omega,\|}$. Thus the reflected and transmitted beams both remain in the plane of incidence, with directions governed by the so-called nonlinear Snell's law. For bulk media without dispersion between the frequencies of ω and 2ω, the reflected and transmitted beams simply maintain the same angle of incidence as the pump beam. However, for dispersive bulk media, the directions of the beams are, however, altered, as can be understood immediately from the expression $2k_\omega/k_{2\omega} = n_\omega/n_{2\omega}$, where n denotes the refractive index of the medium. The surface SHG process does, nonetheless, differ in an important respect. In contrast to the situation for bulk media, the interfacial region is of negligible thickness, so phase shifts can obviously be disregarded. The issue of phase matching, which is of great significance for SHG from bulk media, consequently does not enter into the surface SHG problem.

By application of the Maxwell equations with the indicated nonlinear source polarization, one can derive explicit expressions for the SH radiation in terms of the linear and nonlinear response of the materials and the excitation conditions. For the case of the nonlinear reflection, the irradiance for the SH radiation can be written as

$$I(2\omega) = \frac{\omega^2 \sec^2\theta |\mathbf{e}'(2\omega)\cdot\boldsymbol{\chi}^{(2)}_{\mathrm{s}} : \mathbf{e}'(\omega)\mathbf{e}'(\omega)|^2 [I(\omega)]^2}{2\varepsilon_0 c^3 [\varepsilon_1(2\omega)]^{1/2}\varepsilon_1(\omega)} \qquad [12]$$

where $I(\omega)$ denotes the irradiance of the pump beam at the fundamental frequency; c is the speed of light in vacuum; θ is the angle of emission of the SH beam with respect to the surface normal. The vectors $\mathbf{e}'(\omega)$ and $\mathbf{e}'(2\omega)$ represent the polarization vectors $\hat{\mathbf{e}}_1(\omega)$, and $\hat{\mathbf{e}}_1(2\omega)$, respectively, after they have been adjusted to account for the linear propagation of the waves to the interface. More specifically, we may write $\mathbf{e}'(\omega) = \mathbf{F}_{1\to2}\hat{\mathbf{e}}_1(\omega)$. Here the $\mathbf{F}_{1\to2}$ describes the relationship between the electric field $E\hat{\mathbf{e}}_1$ in medium 1 (propagating towards medium 2), which yields a field $E\mathbf{e}'$ at the interface. For light incident in the x–z plane as shown in Figure 4, $\mathbf{F}_{1\to2}$ is a diagonal matrix whose elements are $F^{xx}_{1\to2} = 2\varepsilon_1 k_{2,z}/(\varepsilon_2 k_{1,z} + \varepsilon_1 k_{2,z})$, $F^{yy}_{1\to2} = 2k_{1,z}/(k_{1,z} + k_{2,z})$, and $F^{zz}_{1\to2} = 2(\varepsilon_1\varepsilon_2/\varepsilon')k_{1,z}/(\varepsilon_2 k_{1,z} + \varepsilon_1 k_{2,z})$, where the quantity $k_{i,z}$ denotes the magnitude of the z-component of the wavevector in medium i at the relevant wavelength (ω or 2ω) (Figure 4).

A few observations may be made about eqn [12], which describes the relationship between the typical experimental observables of the pump irradiance, $I(\omega)$, and the irradiance of the emitted SH reflection, $I(2\omega)$, from the interface. First, one can immediately recognize the quadratic relationship between the pump irradiance and the SH irradiance that is expected for a second-order nonlinear optical effect. Second, the SH radiation is determined not only by the nonlinear optical response, but by the linear optical properties of the relevant media. Consequently, obtaining the surface nonlinear susceptibility tensor $\boldsymbol{\chi}^{(2)}_{\mathrm{s}}$ from experiment requires knowledge or measurement of the linear response. (It should be noted, however, that the interfacial linear properties always enter into the response in a fixed manner and can be incorporated into an effective surface nonlinear response.) The linear response can significantly enhance or reduce the efficiency of the SHG process. The use of a total-internal reflection geometry, for example, provides a significant enhancement in the SH radiation through the linear optical response of the media. Third, while full knowledge of the linear and nonlinear response will obviously allow one to predict the strength of the SH radiation for any set of conditions, a simple measurement of the SH response does not permit one to determine $\boldsymbol{\chi}^{(2)}_{\mathrm{s}}$, even if the linear optical properties are known. The determination of the tensor $\boldsymbol{\chi}^{(2)}_{\mathrm{s}}$ requires probing the material with different polarization states and/or angles of incidence. Protocols for making these measurements have been discussed extensively in the literature. For the general case of a material with in-plane anisotropy, such as the surface of a crystal, measurements of the response as the crystal is rotated about its surface normal are also very useful in a complete determination of $\boldsymbol{\chi}^{(2)}_{\mathrm{s}}$. Fourth, the simple measurement of the SH intensity as assumed here obviously eliminates information about the overall phase of the $\boldsymbol{\chi}^{(2)}_{\mathrm{s}}$. This phase information is often of considerable interest and importance. For example, if the response of the interface is dominated by an oriented molecular monolayer, the sign of $\boldsymbol{\chi}^{(2)}_{\mathrm{s}}$ will be inverted if the monolayer is reflected through the plane of the surface, i.e., the phase of $\boldsymbol{\chi}^{(2)}_{\mathrm{s}}$ yields information on the polar ordering of the layer. To obtain such phase information, one must employ a scheme using interference with a known SH reference. This approach, which is sketched briefly in the next section, is implemented quite commonly in experiment.

Experimental Considerations

The merits of the SHG technique for selective probing of interfaces are clear. On the other hand, one needs

to be able to detect the optical signal associated with a second-order nonlinear response of a material that arises from an effective thickness of just a monolayer or so. Consequently, it is important to consider various factors that aid in optimizing the strength and detectability of the SH radiation. To discuss how best to achieve this end, let us recast eqn [12] in terms of more convenient experimental parameters. We express the SH signal, S, in units of photons/s and describe the laser excitation by its pulse energy E_p, pulse width τ, and repetition rate R. For irradiation of an area A of the sample, we then obtain

$$S = \frac{\omega \sec^2\theta E_p^2 R |\chi_{\text{eff}}^{(2)}|^2}{4\hbar\varepsilon_0 c^3 \tau A} \qquad [13]$$

where $|\chi_{\text{eff}}^{(2)}| \equiv |\mathbf{e}'(2\omega)\cdot\boldsymbol{\chi}_s^{(2)} : \mathbf{e}'(\omega)\mathbf{e}'(\omega)|/[\varepsilon_1^{1/2}(2\omega)\,\varepsilon_1(\omega)]^{\frac{1}{2}}$, whose typical value is of the order of 10^{-21} m^2/V. From this relation, we see that increasing either the pulse energy or repetition rate of the laser will enhance our SH count rate. These parameters are, however, often limited by the characteristics of the available laser, as well as possible damage of the sample. For a given pulse energy and repetition rate, the SH signal is also seen to increase as the laser pulse duration or sample area is decreased. In practice, adequate SH signals can generally be obtained from high-power Q-switched lasers (with nanosecond pulses) with a fairly large focal spot. A preferred solution in most cases is, however, provided by modelocked lasers with high repetition rate and pulses of sub-picosecond duration. Surface SH signals obtained, for example, using a femtosecond modelocked Ti:sapphire oscillator, are typically in the range of 10^2–10^6 photons/s. Although these signals are relatively weak, they are easily detectable with photomultiplier tubes (PMT) and appropriate electronics.

The basic arrangement for surface SHG measurements is shown in Figure 5. It makes use of the directionality and frequency selectivity of the SHG process not only to detect the desired signal with high efficiency, but to discriminate strongly against background signals, including the reflected laser beam. In the setup, irradiation of the sample involves a pulsed laser source, LS; a halfwave plate, λ, and polarizer, P, for controlling the input polarization. Spectral control is achieved with a filter F1 that passes the fundamental radiation in the pump beam, but rejects other wavelengths, particularly spurious SH radiation arising from other optical components; and a color filter F2 that passes SH, but rejects the reflected fundamental radiation. In addition, a spectrometer is commonly employed to limit the detection bandwidth further. An analyzer A serves to select the desired SH polarization. The SH signal is then measured by a high-quantum efficiency detector and the signal is monitored by processing electronics, typically a photon counter, a gated integrator, or a lock-in amplifier, depending on the strength of the signal and the repetition rate of the laser. The laser source may be either of fixed wavelength or tunable for spectroscopic measurements.

There are other important experimental configurations for surface SHG that extend beyond the basic SH reflection geometry, shown schematically in Figure 6a. A common improvement (Figure 6b) is the implementation of a SH reference in which a small fraction of the pump radiation is directed into a crystal with a bulk second-order nonlinear response. This approach provides a reference against which to compare the measured signal, thus compensating appropriately for pulse-to-pulse variation in duration,

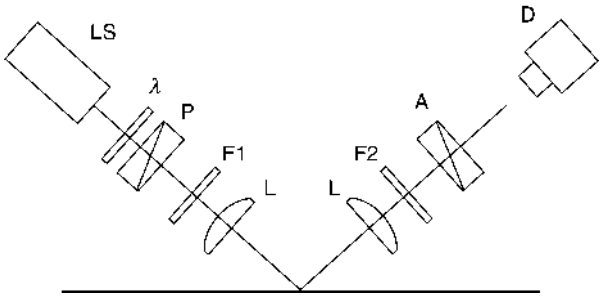

Figure 5 Schematic representation of a basic surface experimental setup for surface SHG. It is comprised of laser source (LS), halfwave plate (λ), polarizer (P), lenses (L), spectral filters (F1, F2), analyzer (A), and detector (D).

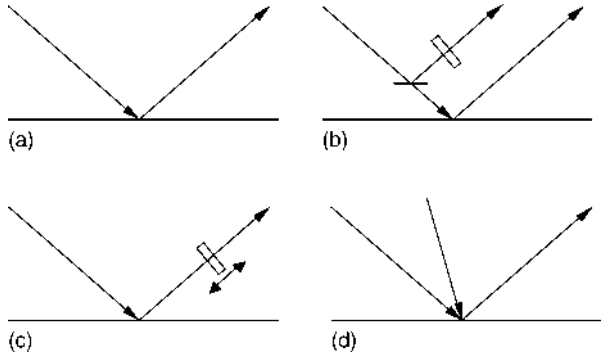

Figure 6 Various schematic representations of different experimental schemes for surface SH measurements. Panel (a) represents the basic scheme as in Figure 5. Panel (b) describes use of a separate SH source as a reference for normalizing the surface SH signals against laser fluctuation and drift. Panel (c) represents the arrangement for a measurement of the phase of the surface SH signal by mixing the fields from the surface and an external source SH radiation. Panel (d) describes the configuration for a time-resolved measurement using the pump-probe scheme.

energy, and spatial profile, as well as for long-term drifts. This type of normalization (against a material with a suitably flat spectral response) is essential for spectroscopy measurements in which the laser frequency is scanned. In addition, use of a calibrated reference is the best way of determining the surface nonlinear response in absolute terms.

It may also be desirable to measure the phase of the signal. For this type of an experiment, a reference field, $\mathbf{E}_{\mathrm{ref}}(2\omega)$, of similar amplitude and known phase response is generated along the path of the surface SH signal $\mathbf{E}(2\omega)$ (Figure 6c). The two collinear SH beams of the same polarization interfere at the detector. The SH power scales as $|\mathbf{E}_{\mathrm{ref}}(2\omega)+\mathbf{E}(2\omega)|^2 = |\mathbf{E}_{\mathrm{ref}}|^2 + 2E_{\mathrm{ref}}E\cos\varphi + |\mathbf{E}|^2$, where φ is the relative phase of the two fields. The relative phase φ may be varied by translating a reference plate along the path of the beam and making use of the dispersion of the ambient air or by inserting another control phase-shifting element. A variant of this phase measurement method is the homodyne scheme in which the reference electric field amplitude $\mathbf{E}_{\mathrm{ref}}(2\omega)$ is significantly stronger than that of the signal amplitude $\mathbf{E}(2\omega)$. The measured SH power then scales as $|\mathbf{E}_{\mathrm{ref}}|^2 + 2E_{\mathrm{ref}}E\cos\varphi$, which yields a signal that varies linearly with the desired SH field strength from the sample.

In the SHG technique, a short-pulse laser is used to interrogate the sample and thus the measurement provides high time-resolution. To make use of this capability, we can add a pump pulse to excite the sample away from equilibrium. This provides the opportunity to examine ultrafast surface dynamics induced by excitation of the sample with a suitable pump laser pulse. For such measurements (Figure 6d), the pump pulse is generally derived from the same laser source as the probe for the SHG measurement, although one or both of these pulses may undergo frequency conversion before being applied to the surface. The critical element is that time synchronism is maintained. The dynamics can then be followed by repeating the SHG measurement at various time delays relative to pump excitation, as controlled by an optical delay line. The ultimate time resolution of this technique is limited only by the duration of the laser pulses and may be as short as a few femtoseconds with state-of-the-art modelocked lasers.

Applications of Surface SHG

The technique of surface SHG is extremely flexible in application, since it relies only on basic symmetry properties for its remarkable surface or interface specificity. SHG measurements can provide quite varied information about the interface, depending on the mode of application of the technique and the material system under study. In this section, we attempt to illustrate the principal types of measurements that can be performed with SHG and the information that can be obtained thereby. For systems with molecular adsorbates, SHG studies can yield information on the density and orientation of molecular species, and can typically do so in real time. For crystalline surfaces, the symmetry and order of the surface can be examined through studies of tensor properties of the surface nonlinear susceptibility. Spectroscopic studies of electronic transitions are also of great value. These can be accomplished for both solids and molecular systems through the same experimental approach of tuning the frequency of the laser source. In all of these investigations, one can achieve very high time resolution, down into the femtosecond range, by the use of pump-probe techniques. In addition, as we discuss below, SHG provides interesting additional capabilities for systems with applied electric and magnetic fields. Finally, the method permits one to probe spatially inhomogeneous systems.

In addition to this diverse range of information that can be provided by surface SHG measurements, the technique is unique in the extremely wide range of material and chemical systems to which it can be – and, indeed, has been – applied. These span the gamut from solids under ultrahigh vacuum to solid/solid, solid/liquid, liquid/vapor, and even liquid/liquid interfaces. Since our examples were drawn to illustrate the type of information obtainable, rather than the range of material systems investigated, the reader is unfortunately obliged to refer to other more detailed treatments of surface SHG measurements to appreciate fully this important facet of the technique.

Adsorbate Density

The remarkably high sensitivity of the SHG technique to the nature of the surface is illustrated by the data in Figure 7, which shows the dramatic change in the SH response from a clean Si(111)-7 × 7 surface upon adsorption of atomic hydrogen. Indeed, it is possible to detect well below 1% of a monolayer (ML) of adsorbed hydrogen. A useful point to note is that the change in $\chi_{\mathrm{s}}^{(2)}$ is linear with adsorbate density for modest values of the coverage, θ. Thus, it is quite easy to obtain a relative gauge of the adsorbate coverage in this regime.

The behavior of the SH response to adsorbates at low coverages can often be described by a simple linearized model of the surface nonlinear susceptibility of $\chi_{\mathrm{s}}^{(2)} = (1-a\theta)\chi_{\mathrm{s},1}^{(2)} + \theta\chi_{\mathrm{s},2}^{(2)}$, where

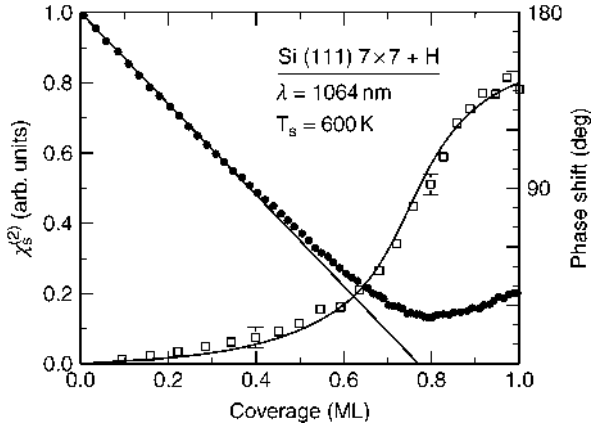

Figure 7 Dependence of the magnitude (full circles) and the phase (open squares) of the nonlinear susceptibility $\chi_s^{(2)}$ for a Si(111)-7 × 7 surface as a function of the coverage of adsorbed atomic hydrogen using a fundamental photon energy 1.17 eV. Reproduced with permission from Höfer U (1996) Nonlinear optical investigations of the dynamics of hydrogen interaction with silicon surfaces. *Applied Physics A* 63: 533.

$\chi_{s,i}^{(2)}$ $(i = 1, 2)$ are complex quantities and a is a constant. This relation describes the surface nonlinear susceptibility as arising from a contribution $\chi_{s,1}^{(2)}$ from the clean surface and a contribution $\chi_{s,2}^{(2)}$ from a saturated monolayer $(\theta = 1)$ of the adsorbate. The term $-a\theta\chi_{s,1}^{(2)}$ then represents the modification of the surface nonlinearity arising from interaction with the adsorbate. For the case of hydrogen adsorption of Figure 7, this is the dominant term, since hydrogen atoms would have no nonlinearity on their own. In other systems, however, the response can be considered as arising essentially from the nonlinear response of aligned adsorbed molecules. The simple linearity of this relation of $\chi_s^{(2)}$ with adsorbate coverage obviously neglects important effects such as adsorbate–adsorbate chemical interactions, as well as local-field effects and the possible influence of differing adsorption sites. All of these factors have been considered in the literature. Nonetheless, the linearity of response with adsorbate coverage is often found to be a good approximation, even for fairly high adsorbate coverages. From the point of view of simply following adsorption dynamics with SHG, it is not, we note, necessary to have any a priori knowledge of the relation between the adsorbate coverage and the SH response. This relation can be established empirically in static measurements and then used advantageously to follow the dynamics of surface processes.

Surface Symmetry and Molecular Orientation

The surface SHG process is described by the third-rank tensor $\boldsymbol{\chi}_s^{(2)}$. The tensorial properties of this response reflect the symmetry of the surface. Other than for a few cases of high symmetry, such a third-rank tensor gives a distinct signature of surfaces with crystalline order compared to surfaces, such as disordered ones, exhibiting effective in-plane isotropy. The form of the $\boldsymbol{\chi}_s^{(2)}$ tensor can be probed both by measurements that vary the polarization of the fundamental and second-harmonic beams and by measurements in which the plane of incidence is changed with respect to the orientation of the surface, generally by rotating the sample about its normal with a fixed experimental geometry.

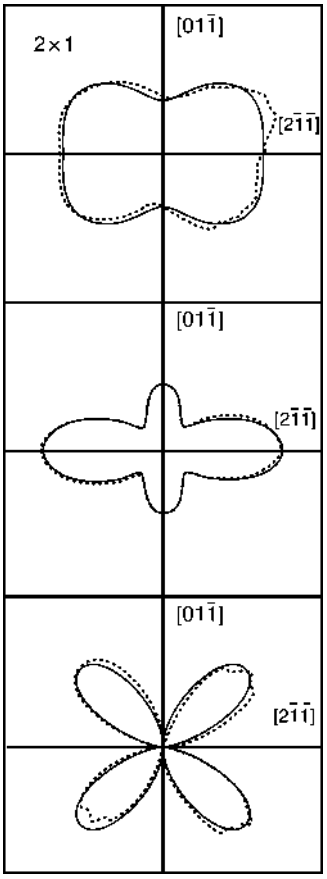

Figure 8 Polarization dependence of SHG for normal incidence excitation of the cleaved Si(111)-2 × 1 surface. The polar plot of the SH signal versus the pump polarization orientation is shown for the case of no polarization analysis (top panel), for the SH analyzer along the [2$\bar{1}\bar{1}$] crystallographic direction of the surface (middle panel), and along the [01$\bar{1}$] direction (bottom panel). The solid curves are fits to theory assuming that the surface has *m* symmetry. Reproduced with permission from Heinz TF, Log MMT and Thomson WA (1985) Study of Si(111) surfaces by optical second-harmonic generation: reconstruction and surface phase-transformation. *Physical Review Letters* 54: 63.

Figure 8 illustrates the capability of surface SHG measurements to probe surface symmetry. The figure shows SHG polarization dependence for a Si(111) surface cleaved in ultrahigh vacuum. While the

underlying bulk symmetry suggests that the surface might exhibit 3m symmetry, the analysis of data indicates that only a single mirror m plane is present. This lowered symmetry corresponds to the metastable 2×1 reconstruction induced by the crystal cleavage. By heating the sample, one can observe a phase transformation to the equilibrium 7×7 reconstruction and the emergence of the corresponding 3m surface symmetry.

The analysis of surface or interface symmetry by SHG relies only on the form of the nonlinear susceptibility tensor, i.e., which elements are independent and nonvanishing. Within the context of a microscopic model of the surface response, however, the relative values of the tensor elements of $\boldsymbol{\chi}_s^{(2)}$ can also yield valuable information about the nature of the surface. This approach has been developed extensively in the context of the analysis of molecular orientation at surfaces. The basic notion behind this treatment relies on the knowledge of the microscopic molecular nonlinear polarizability $\alpha_{\lambda\mu\nu}^{(2)}$ and the measured surface nonlinear susceptibility $\chi_{s,ijk}^{(2)}$ to infer information about the orientational average that relates these quantities. As discussed above in conjunction with eqn [10], under the neglect of local-field corrections, the relationship between these two quantities involves the ensemble average of the coordinate transformation tensor from the molecular to the laboratory coordinate systems $\langle T_{i\lambda} T_{j\mu} T_{k\nu} \rangle$. Hence a knowledge of the quantities $\alpha_{\lambda\mu\nu}^{(2)}$, N_s, and $\chi_{s,ijk}^{(2)}$ allows us to infer information about the orientational distribution of the molecules as described by the quantity $\langle T_{i\lambda} T_{j\mu} T_{k\nu} \rangle$. Even in the absence of knowledge of the magnitudes of the nonlinear polarizabilities and the adsorbate density, one may still obtain useful information for molecules with a simple form of $\alpha_{\lambda\mu\nu}^{(2)}$.

An important example of this analysis of molecular orientation concerns the case of molecules with a dominant nonlinear polarizability along a single axis, i.e., $\boldsymbol{\alpha}^{(2)} = \alpha_{z'z'z'}^{(2)} \hat{\mathbf{z}}' \hat{\mathbf{z}}' \hat{\mathbf{z}}'$. For the usual situation of in-plane isotropy, one can then show that the nonlinear susceptibility elements are given by $\chi_{s,\perp\perp\perp}^{(2)} = N_s \langle \cos^3\theta \rangle \alpha_{z'z'z'}^{(2)}/\varepsilon_0$ and $\chi_{s,\parallel\perp\parallel}^{(2)} = \chi_{s,\parallel\parallel\perp}^{(2)} = \frac{1}{2} N_s \langle \cos\theta \times \sin^2\theta \rangle\ \alpha_{z'z'z'}^{(2)}/\varepsilon_0$, where θ denotes the angle between the surface normal and the z' molecular axis. It should be noted that these orientational moments, reflecting their origin in a third-rank tensor, give a signature of the polar ordering, since a sign change in the elements of $\boldsymbol{\chi}_s^{(2)}$ follows if the molecular orientation is inverted. Also, when only one element of $\boldsymbol{\alpha}^{(2)}$ is present, we can form ratios of the experimentally measurable ratio of tensor elements that depend only on the molecular orientational factors, such as $R \equiv (2\chi_{s,\parallel\perp\parallel}^{(2)} + \chi_{s,\perp\perp\perp}^{(2)})/\chi_{s,\perp\perp\perp}^{(2)} = \langle \cos\theta \rangle / \langle \cos^3\theta \rangle$. For a narrow distribution of the orientations centered at θ_0, this ratio yields directly the molecular orientation angle through $R \approx \sec^2\theta_0$.

This type of molecular orientation analysis has been applied to a wide range of material systems. The method, while powerful, relies on several assumptions, including an adequate knowledge of the molecular nonlinear polarizability, the lack of significant local field corrections, the availability of information about the effective linear dielectric constant in the interfacial region, and a suitable method of eliminating any substrate nonlinear response in the measurement. The method is, consequently, best suited for adsorbed layer of molecules with strong and well-defined nonlinear response, present at relatively modest coverages. The use of resonant excitation is a valuable adjunct in these studies, since it generally simplifies the form and strengthens the molecular nonlinear response. An analogous approach to the analysis of molecular orientation can be accomplished with infrared–visible sum-frequency generation. These measurements often provide a superior method of analysis, since the molecular response and its symmetry are better defined by resonant excitation of a molecular vibration.

Surface and Interface Spectroscopy

Surface SHG can be readily adapted to yield spectroscopic information on surface transitions. As indicated above, both resonances at the fundamental and second-harmonic frequencies are manifested in the SHG response. The examination of resonances at the SH frequency is often convenient, since it permits the application of a relatively intense pump radiation at a frequency well separated from that of the resonance. Such SHG spectroscopy measurements have been widely applied to examine electronic transitions at surfaces and interfaces of solids, as well as in adsorbed molecules under a wide variety of conditions.

An example of SHG spectroscopy of electronic transitions at a buried solid/solid interface is presented in Figure 9. For the epitaxial CaF_2/Si(111) interface, distinctive new electronic transitions, associated with interface states, are observed that are absent from the response of either of the bulk materials. These excitations occur in a spectral range where the bulk silicon sample is highly absorbing, which would render the identification of the new interface transitions very difficult using conventional linear optics. The buried nature of the interface naturally also constrains application of conventional surface probing techniques involving electron

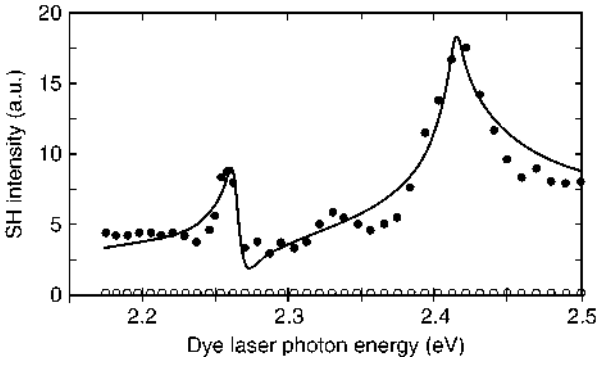

Figure 9 SHG spectrum (solid dots) for the CaF_2/Si(111) interface. For comparison, the open circles represent the corresponding signals for the SiO_2/Si(111) interface, which is seen to give a very weak signal in this frequency range. The solid line is a theoretical fit based on a two-dimensional direct transition of the interface states, together with an excitonic feature slightly below the band gap. Reproduced with permission from Heinz TF, Himpsel FJ, Palange E and Burstein E (1989) Electronic transitions at the CaF_2/Si(111) interface probed by resonant three-wave-mixing spectroscopy. *Physical Review Letters* 63: 644.

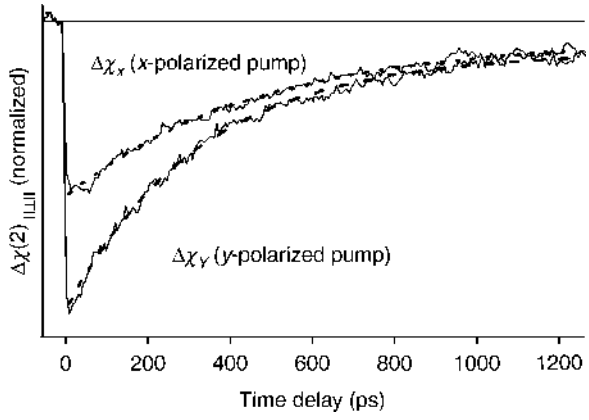

Figure 10 Rotational relaxation of Coumarin 314 molecules at the air–water interface subsequent to photoexcitation. The initial anisotropy in the molecular orientational distribution is created by linearly polarized pump radiation in two orthogonal directions in the surface. Reproduced with permission from Zimdars D, Dadap JI, Eilsenthal KB and Heinz TF (1999) Anisotropic orientational motion of molecular adsorbates at the air–water interface. *Journal of Physical Chemistry B* 103: 3425–3433.

irradiation or emission. Two additional general observations are appropriate. First, the lineshape analysis for such measurements must take into account the fact that the SHG response derives from both the real and imaginary parts of $\chi_s^{(2)}$, as we discussed above. Second, in the absence of any other knowledge, SHG spectroscopic data contain an ambiguity about whether the resonances correspond to the fundamental or SH photon energy. This ambiguity, however, can be resolved in a straightforward fashion by complementing the SHG measurement with a sum-frequency measurement in which one frequency is scanned and one is held fixed, as was demonstrated in the study associated with Figure 9.

Time-Resolved Measurements

One of the fascinating frontiers that can be addressed by SHG measurements is probing the ultrafast dynamics of surfaces and interfaces by the pump-probe method. This possibility is exemplified by a variety of studies ranging from the dynamics of laser-driven phase transitions, charge-transfer, solvation, phonon dynamics, and orientational and torsional dynamics.

An example of applying time-resolved SHG to examine ultrafast molecular dynamics at the liquid/vapor interface of water is presented in Figure 10. In these studies, the rotational motion of dye molecules at the interface, which probes the local environment, is followed in real time. This is accomplished by photo-exciting the dye molecules at the interface. Because polarized pump radiation is used, an anisotropic orientational distribution of excited-state molecules is established. Since the excited molecules have a different SHG response, a change in the signal is observed immediately. Subsequently, the anisotropic orientational distribution relaxes to the equilibrium configuration. The time evolution of the rotational anisotropy is followed by detecting the SH of a probe laser pulse as a function of the delay time. Through a comparison of the results for different initial anisotropic distributions (produced by pump beams with differing polarizations), one may deduce rates for both in-plane and out-of-plane orientational relaxation. This example illustrates the possibilities of time-resolved SHG measurements for probing the detailed dynamics of molecules at an interface on the ultrafast timescale.

Spatially Resolved Measurements

Up to this point, our discussions have centered on spatially homogenous surfaces and interfaces. Surface SHG has also been exploited to probe the lateral variation of inhomogeneous surfaces with the micron to sub-micron resolution characteristic of optical microscopy techniques. Spatial resolution may be achieved simply by detecting the nonlinear response with a focused laser beam that is scanned across the surface. A large area of the surface may also be illuminated and the emitted nonlinear radiation imaged. A wide range of applications of this imaging capability has emerged, including the probing of biological materials, such as cells and tissues, to probing spatially varying electric fields

and magnetic domains. Within the biological context, SH microscopy has recently become an active research direction. In these applications, while probing with SHG typically senses non-centrosymmetric structures in bulk samples as opposed to interfaces, most of the aspects of SHG presented in this article still apply. To further improve the spatial resolution of SHG microscopy, researchers have also employed near-field techniques for spatial resolutions below the diffraction limit.

Complementary approaches to the real-space imaging just described have also been developed based on SH diffraction. In this scheme, periodically modulated surfaces are established holographically on a surface by utilizing two interfering laser beams. In addition to the reflected and transmitted SH beams, additional beams are then generated at well-defined angles as dictated by the grating period. This approach has been used to monitor the relaxation of a monolayer grating of adsorbates in time, which yields precise information on the rate of adsorbate surface diffusion.

Probing Electric Fields at Interfaces

For centrosymmetric media, the SHG process shows a high degree of sensitivity to the presence of electric fields. This property follows from the fact that an applied electric field can break the inversion symmetry, just as an interface does. The resulting process, termed electric-field induced SHG (or EFISH), has been known for many years. From a phenomenological standpoint, it can usually be described as a nonlinear response of the form $\mathbf{P}^{(3)}(2\omega) = \varepsilon_0 \boldsymbol{\chi}^{(3)} : \mathbf{E}(\omega)\mathbf{E}(\omega)\mathbf{E}$, where $\mathbf{E}$ is the electric field being probed and $\mathbf{E}(\omega)$ is the laser pump field. Within the context of surfaces and interfaces, there are many circumstances where probing such electric fields is of interest and importance. The SHG technique provides high sensitivity, good spatial resolution, accessibility to buried interfaces, and, in the pump-probe implementation, extremely high time resolution. In addition, with appropriate measurements of the polarization dependence and homodyne detection, the full vector character of the electric field can be determined.

The behavior of electric-field induced SHG is similar to that arising from the asymmetry of an interface. Indeed, it is formally equivalent to the response of an interface provided that the electric field is present in (or radiates from) a region of less than a wavelength's thickness. The applications of this measurement capability are quite diverse, ranging from studies of charging and acid/base reactions at liquid/solid interfaces to probing fields in metal-oxide-semiconductor structures and Schottky barriers to the measurement of freely-propagating terahertz radiation. In Figure 11,

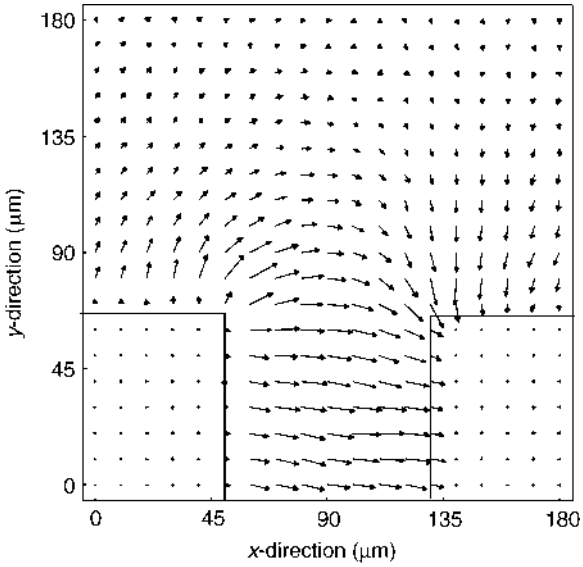

Figure 11 Vector map of the electric field in Si-on-sapphire substrate as deduced by SHG measurements. A voltage is imposed on the rectangular metallic electrodes, which are separated by 80 μm. Reproduced with permission from Dadap JI, Shan J, Weling AS, Misewich JA, Nahata A and Heinz TF (1999) Measurement of the vector character of electric fields by optical second-harmonic generation. *Optics Letters* 24: 1059–1061.

SHG with appropriate polarization control and homodyne detection scheme has been applied to the measurement of the vector character of electric fields present in a dc-biased dipole structure on silicon. By spatially scanning the laser beam, a two-dimensional electric field map can be produced. In a pump/probe scheme, the same measurements can be made with femtosecond time resolution.

Probing Magnetization at Interfaces

In contrast to the application of an external electric field (a polar vector), the presence of an external magnetic field (an axial vector) does not break the inversion symmetry of a material. Thus, SHG retains its interface selectivity for magnetic effects. SHG measurements, as an interface-specific probe, thus provide an excellent complement to the linear magneto-optical Kerr effect (MOKE), which is sensitive to magnetic effects within the penetration depth of the light.

The ability of the SHG technique to sense the magnetization at a surface is illustrated in Figure 12 for a film of epitaxially-grown magnetic Co on nonmagnetic substrate Cu. A hysteresis loop of the magnetization versus the strength of an applied magnetic field is shown, where the magnetization is gauged from the SHG response. Through measurements of the response as a function of the thickness of

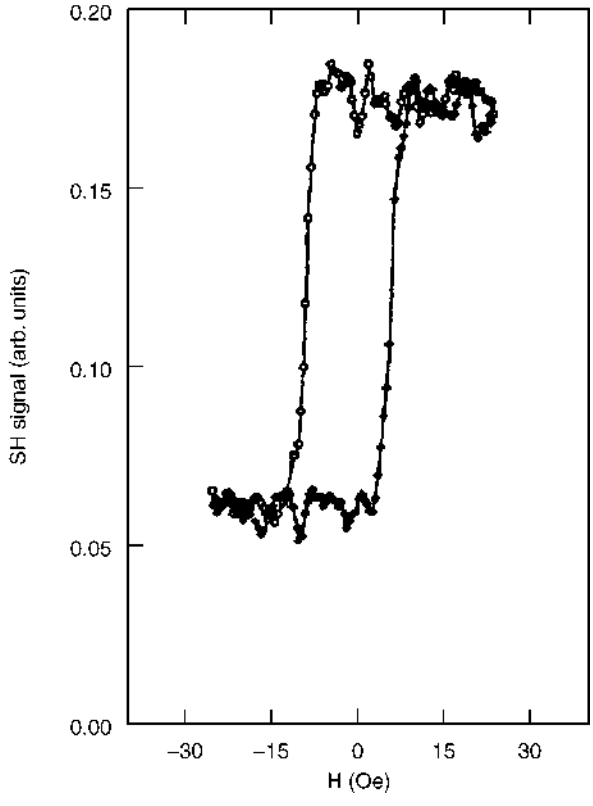

Figure 12 SH response of a Co film on a Cu(001) substrate as a function of the applied magnetic field. A hysteresis loop corresponding to the surface response is observed. Reproduced with permission from Wierenga HA, de Jong W, Prins MWJ *et al.* (1995) Interface magnetism and possible quantum-well oscillations in ultrathin Co/Cu films observed by magnetization induced second harmonic generation. *Physical Review Letters* 74: 1462–1465.

the magnetic film, the researchers demonstrated that the SH response reflected the interface properties, rather than those of the bulk film.

Probing Micro- and Nanoscale Structures

In addition to the microscopy studies described above that probe spatially inhomogeneous surfaces, there has been considerable interest in SHG from a variety of strongly textured surfaces and small particles. From the standpoint of the optical response of materials, particular interest has been attached to studies of metallic surfaces with surface roughness. For appropriate roughness of low-loss materials, such as silver, very significant local-field enhancements can be achieved. Since these field enhancements enter into the strength of the SH emission nonlinearly, overall enhancements in the SH power by several orders of magnitude have been demonstrated. This provides both a useful method of study of such surface enhancement effects and a scheme for amplifying the relatively weak surface nonlinear response.

Beyond these problems associated with roughened surfaces, considerable attention has recently been directed towards the application of the SHG technique for probing micro- and nano-scale particles. The fundamental principle of the interface selectivity of the SHG process for such systems comprised of centrosymmetric materials remains unaltered. The interaction of these particles with the radiation fields is, however, significantly modified. For the case of particles with dimensions of several optical wavelengths, the overall SHG signal strength may be comparable to that from a corresponding area of a planar surface. The radiation pattern will, of course, be strongly modified and will generally be diffuse in character. While this makes detection (and background discrimination) somewhat more difficult than for the planar geometry, many SHG measurements of surface and interface properties have been carried out in this regime. As the particle size becomes progressively smaller, particles with overall inversion symmetry begin to exhibit appreciable cancellation of the SH emission. Indeed, in the limit of vanishing size, full cancellation is expected. This tendency is, however, offset by the increasing surface-to-bulk ratio and the possibility of increasing the number of particles probed. It has been demonstrated experimentally that the surfaces of particles in the nanometer length scale can indeed be probed by SHG.

From the point of view of applications, this extension of the SHG technique to probe micro- and nano-structured materials opens up a rich variety of new possibilities. Results have been reported on metal and semiconductor nanoparticles in suspension, as well as on colloids of minerals. In addition, liquid droplets have been investigated. Self-assembling structures such as liposomes and vesicles have been examined. For these structures, many of the same capabilities of SHG as demonstrated for the planar geometry, such as probing of adsorption and charging effects, have been demonstrated. More uniquely, molecular transport *through* the liposome membrane could also be observed and characterized. The possibility of extension of these measurements to biological systems is of obvious interest and promise.

Concluding Remarks

In this article, we provided a brief review of surface SHG as a technique for interface-specific probing of materials. In addition to discussing some of the fundamental ideas underlying the method, we outlined the wide variety of information that may be obtained from such measurements, ranging from

surface symmetry and molecular ordering to interface spectroscopy and ultrafast dynamics. While several aspects of the method have now become quite well established, developments in widely available mode-locked lasers with excellent performance characteristics have greatly enhanced the ease of making surface SHG measurements in a wide class of material systems. In addition, new developments and applications of the method have continued to emerge apace. Of particular note in this regard are applications to probe electric fields and magnetization at interfaces, as well as advances in the use of the method for interface-specific probing of micro- and nanoscale structures. The sharp increase in applications of SHG for probing systems of biological importance constitutes another important development, and one with much promise for the future.

See also

Materials for Nonlinear Optics: Organic Nonlinear Materials. **Nonlinear Optics, Basics:** $\chi^{(2)}$–Harmonic Generation. **Physical Applications of Lasers:** Sum-Frequency Generation at Surfaces.

Further Reading

Bennemann KH (ed.) (1998) *Nonlinear Optics in Metals*. Oxford, UK: Clarendon.

Bloembergen N (1965) *Nonlinear Optics*. New York: WA Benjamin.

Boyd RW (1992) *Nonlinear Optics*. New York: Academic.

Brevet PF (1997) *Surface Second Harmonic Generation*. Lausanne, Switzerland: Presses Polytechniques et Universitaires Romandes.

Butcher PN and Cotter D (1990) *The Elements of Nonlinear Optics*. Cambridge, UK: Cambridge.

Corn RM and Higgins DA (1994) Optical second-harmonic generation as a probe of surface-chemistry. *Chemical Review* 94: 107–125.

Dadap JI and Heinz TF (2001) Nonlinear optical spectroscopy of surfaces and interfaces. In: Moore JH and Spencer ND (eds) *Encyclopedia of Chemical Physics and Physical Chemistry*, pp. 1089–1125. Bristol, UK: IOP.

Eisenthal KB (1996) Liquid interfaces probed by second-harmonic and sum-frequency spectroscopy. *Chemical Review* 96: 1343–1360.

Heinz TF (1991) Second-order nonlinear optical effects at surfaces and interfaces. In: Ponath H-E and Stegeman GI (eds) *Nonlinear Surface Electromagnetic Phenomena*, pp. 353–416. Amsterdam: North Holland.

Liebsch A (1995) Electronic excitations at metal surfaces. In: Halevi P (ed.) *Photonic Probes of Surfaces*, pp. 479–532. Amsterdam: Elsevier.

Lüpke G (1999) Characterization of semiconductor interfaces by second-harmonic generation. *Surface Science Report* 35: 75–161.

McGilp JF (1995) Optical characterisation of semiconductor surfaces and interfaces. *Progress in Surface Science* 49: 1–106.

McGilp JF (1999) Second-harmonic generation at semiconductor and metal surfaces. *Surface Review Letters* 6: 529–558.

McGilp JF, Patterson CH and Weaire DL (eds) (1995) *Epioptics: Linear and Nonlinear Optical Spectroscopy of Surfaces and Interfaces*. Berlin, New York: Springer.

Prasad PN and Williams DJ (1991) *Introduction to Nonlinear Optical Effects in Molecules and Polymers*. New York: Wiley.

Reider GA and Heinz TF (1995) Second-order nonlinear optical effects at surfaces and interfaces: recent advances. In: Halevi P (ed.) *Photonic Probes of Surfaces*, pp. 413–478. Amsterdam: Elsevier.

Richmond GL (1997) Vibrational spectroscopy of molecules at liquid/liquid interfaces. *Analytical Chemistry* 69: A536–A543.

Richmond GL, Robinson JM and Shannon VL (1988) Second harmonic generation studies of interfacial structure and dynamics. *Progress in Surface Science* 28: 1–70.

Shen YR (1984) *The Principles of Nonlinear Optics*. New York: Wiley.

Shen YR (1989) Optical second harmonic-generation at interfaces. *Annual Review of Physical Chemistry* 40: 327–350.

Single Molecule Spectroscopy

X Michalet and S Weiss, University of California, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Textbooks in physics, chemistry, and biology abound in diagrams depicting single-particle, single-molecule or single-enzyme properties, reactions or interactions with their environment. Such pictures and the numbers associated with them, however, were, until recently, extrapolated from measurements performed on large ensembles of molecules, which only yield quantities averaged out over time, space, conformations and/or the local environment. Advances in detector sensitivity and improvements in instrument design do now provide scientists

with tools to probe single molecules with light, and monitor their photophysical properties with exquisite sensitivity and time resolution. This spectroscopic information is used to explore the molecular structure, conformational dynamics, local environment, and intermolecular interactions. The collected data generate a wealth of new information on single-molecule physical chemistry or biomechanics, but also in polymer gel physics or glass dynamics and other domains of solid state physics.

This article reviews the field of single-molecule spectroscopy. First, the principles underlying single-molecule spectroscopy approaches are introduced, and this is followed by a discussion of the different spectroscopic techniques available to the experimentalist and what kind of information can be extracted with each of them. It also discusses important photophysical properties of molecules that influence the outcome of single-molecule spectroscopic measurements. The next section presents some applications of these techniques, focusing on physical properties of materials and biological applications of single-molecule spectroscopy. The last part briefly reviews recent developments and outlines the future prospects of single-molecule spectroscopy.

Principles of Single-Molecule Spectroscopy

Motivation

Measurements performed on ensembles of molecules only yield averaged-out information. This averaging process is of several kinds: average over molecular heterogeneity (e.g., chemical composition, oxidation state, or spatial conformation) and average over different or changing local environments. Moreover, dynamic processes studied by ensemble measurements cannot sort out molecules exhibiting fast or slow reaction times, although such differences are expected from the stochastic nature of any chemical kinetics. The interest in single-molecule analysis is for these reasons manifold. At low temperature, being able to study a single molecule and its evolution under the influence of various external parameters allows for precise tests of molecular quantum electrodynamics. At higher temperature, it gives access to the temporal evolution of different nanodomains of a bulk material in which single molecules are embedded as reporters of their environment. In biology and biochemistry, the building blocks of any process are individual molecules (genes, enzymes, motors, filaments, etc.); the ability to fully understand their kinetic behavior at the single-molecule level is thus of utmost importance. Finally, reliable observation of single molecules is a prerequisite of the full development of nanotechnology and quantum computing.

Historical Outlook

Single-molecule experiments were first performed on isolated ion channels by the patch clamp technique, which allows the recording of single-ion translocation through a pore-like protein embedded in a fluid membrane. Scanning tunneling and atomic force microscope observations later opened the way to investigations of nanometer-scale objects on surfaces. Only recently have optical methods reached the sensitivity required to detect, image, manipulate, and follow the spectroscopic evolution of single molecules on surfaces, in solids, liquids, and biological membranes as well as inside live cells.

Single-molecule spectroscopy (SMS) was initially demonstrated in a system of pentacene molecules embedded in a *p*-terphenyl host crystal at liquid helium temperature. Absorption (leading to fluorescence excitation) spectroscopy in this system takes advantage of the very narrow zero-phonon absorption line (ZPL) of the rigid pentacene molecule (Figure 1). The position in laser-frequency space of each molecule's ZPL indeed depends on the local fields, allowing for precise molecular selectivity. However this approach is limited to a few molecule–host pairs and requires cryogenic techniques, due to broadening of the ZPL above 10 K.

Room temperature studies started with total internal reflection (TIR) microscopy and standard flow-cytometry methods in the mid-1980s. In the latter case, single molecules randomly passing through the excitation volume gave rise to sudden bursts of photons emitted during their brief transit. Initially developed for rapid DNA sequencing, this simple approach is now widely used for more sophisticated studies in its confocal version (Figure 2), which will be detailed in the following.

The first image of a single molecule at room temperature was later obtained using scanning near-field optical microscopy (Figure 3), rapidly followed by similar achievements using far-field confocal microscopy, at the expense of a lower spatial resolution, but with much more ease of use. TIR and wide-field fluorescence imaging are now used to image and study the two-dimensional and three-dimensional diffusion trajectories of single fluorescent molecules (Figure 4), making a come-back after the pioneering studies of the early 1980s.

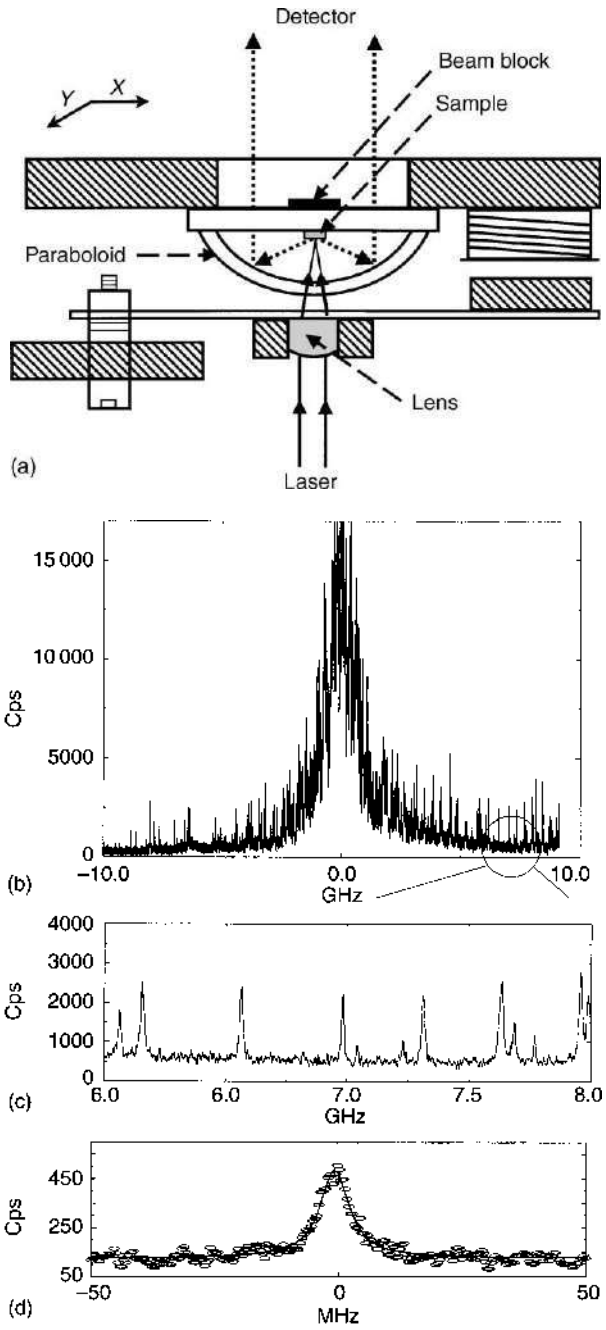

Figure 1 Fluorescence excitation spectra of pentacene in *p*-terphenyl at 1.5 K. (a) Schematic of the cryogenic setup used. A laser beam is focused onto the sample through a lens into the cryogenic chamber. The sample can be moved in the *XY* plane, with a magnet and coil pair. A beam block rejects the incident laser light (black arrow). A paraboloidal mirror collects the emitted fluorescence (dashed arrows) and redirects it towards the detector (photomultiplier tube). The laser frequency (linewidth 3 MHz) is tuned around the absorption line center at 592.321 nm. (b) Broad scan showing the inhomogeneously broadened line shape of pentacene molecules embedded in a *p*-terphenyl crystal at 1.5 K. The sharp lines correspond to spectra of individual molecules (see panel c) (cps: counts per second). (c) Detail of the previous spectrum showing spectra of several individual molecules. The variability of intensities is due to the different positions of the molecule in the excitation beam. (d) Individual molecule spectrum obtained at lower excitation power exhibiting a Lorentzian shape (solid line) centered at 592.407 nm with a lifetime-limited width of 7.8 MHz (intensity is measured in counts-per-second for all three spectra). (a): adapted from Ambrose WP, Basché T and Moerner WE (1991) Detection and spectroscopy of single pentacene molecules in a *p*-terphenyl crystal by means of fluorescence excitation. *Journal of Chemical Physics* 95: 7150–7163. (b)–(d): after Moerner WE (1994) Examined nanoenvironments in solids on the scale of a single, isolated impurity molecule. *Science* 265: 46–53. Copyright 1994, The American Association for the Advancement of Science.

Requirements for Single-Molecule Sensitivity

Observing single molecules may be as easy as finding a needle in a haystack if background, signal and noise are not carefully optimized. Simple considerations illustrated in Figure 5a help measure the technical challenge bearing on single-molecule optical detection. Whatever the type of detected signal (fluorescence, Raman scattering, or others), it is characterized by an absorption cross-section σ (probability of photon absorption) and a quantum yield Q (number of emitted photons per absorbed photon). The optical setup transmits an excitation power P at a frequency ν, resulting in a signal rate $s = EQ\sigma P/(Ah\nu)$ expressed in counts per second (or Hz). This rate is proportional to P/A where A is the cross-section of the excitation beam in the plane of the molecule. E is the collection efficiency of the setup. The environment adds a background rate b per unit volume and unit power, and the detector contributes a dark count rate D. Neglecting noise from the readout electronics, a measurement of duration τ will yield the following signal-to-noise ratio (SNR):

$$\mathrm{SNR} = \frac{s\tau}{\sqrt{(s + bVP + D)\tau}} \qquad [1]$$

where V is the excited volume. The denominator is the root mean square of all contributions to the measured count rate, considered shot-noise limited. The SNR obviously increases at larger integration time τ. Another important quantity is the signal-to-background ratio (SBR):

$$\mathrm{SBR} = \frac{EQ\sigma}{bVAh\nu} \qquad [2]$$

These ratios can be increased by improving the collection efficiency E or decreasing the excitation volume V. A larger excitation power or a longer integration time will improve the SNR without affecting the SBR. Typical values for the above

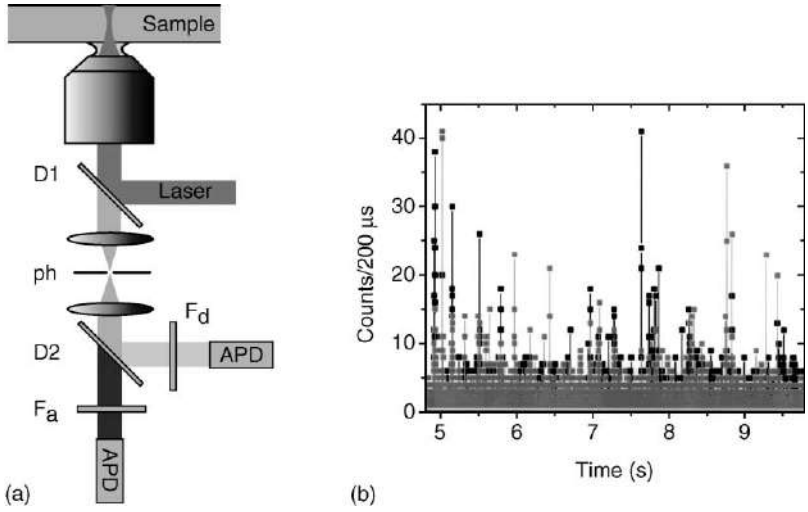

Figure 2 Single-molecule fluorescence burst detection. (a) Schematic of the confocal setup used for single-molecule fluorescence detection. A collimated laser beam is focused through a high numerical aperture (NA) objective in a liquid cell. Molecules diffusing through the diffraction-limited excitation volume (~1 fl) emit photons during their passage, which are collected by the same objective. A dichroic mirror (D1) separates excitation and emission wavelengths. The emitted light is focused on a pinhole (ph) in order to reject out-of-focus background light, and finally recollimated onto two different avalanche photodiodes (APDs) after spectral separation by a dichroic mirror (D2). Filters specific for each channel (F_a: acceptor channel, F_d: donor channel) may be used to improve spectral separation. (b) 5 s time trace of recorded bursts for a 30 pM solution of doubly labeled DNA in 20 mM sodium phosphate buffer. The DNA constructs are used for distance measurement (detailed in **Figure 9**) and have a TMR fluorophore attached to one end of the DNA molecule and a CY5 fluorophore attached to the *n*th base from the end. The time traces correspond to $n = 12$. Gray: TMR fluorescence (donor); black: CY5 fluorescence (acceptor). The laser excitation (0.6 mW, 514 nm) is focused 10 μm within the solution through an oil immersion objective (numerical aperture: 1.4). For each detected photon, the APD generates a pulse, which is recorded by a computer-embedded acquisition board. Counting these events into 200 μs bins, the resulting time trace exhibits distinct bursts distributed randomly in time, with duration and amplitude corresponding to a Brownian diffusion of individual molecules through the excitation volume. (b): is adapted from Deniz AA, Dahan M, Grunwell JR, *et al.* (1999) Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. *Proceedings of the National Academy Sciences USA* 96: 3670–3675. Copyright 1999, The National Academy of Sciences.

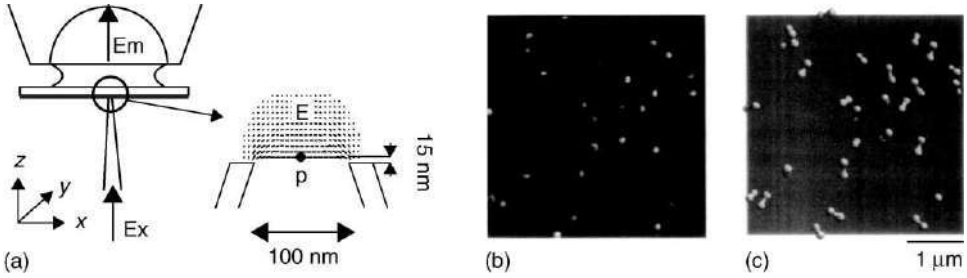

Figure 3 Scanning near-field imaging of single molecules at room temperature. (a) Schematic of a scanning near-field optical microscope setup used for single-molecule imaging. An aluminum-coated tapered optical fiber (raster-scanned at nanometer distance from the sample) serves as a waveguide for laser excitation. Shear-force feedback keeps the tip at a constant distance from the sample, resulting in a signal used for nanometer-resolution topographic reconstruction of the scanned area. The excitation volume and the corresponding local evanescent electric field are detailed in the expanded view. Fluorescence light emitted by individual molecules is collected by an oil immersion, high NA objective and recorded by an avalanche photodiode. (b) $4 \times 4\ \mu m^2$ area of a polymethylmethacrylate (PMMA) coated coverslip on which a diluted methanol solution of a lipophilic carbocyanine dye, $diIC_{12}$, has been deposited. Scanning was performed with randomly polarized light. (c) Orientation of the corresponding individual emitting dipole determined by modeling the field distribution (expanded view in a), simulation of the absorption intensity with polarized excitation and comparison with data obtained at two different excitation polarizations. (a): Adapted and (b), (c) reprinted with permission from Betzig E and Chichester RJ (1993) Single molecules observed by near-field scanning optical microscopy. *Science* 262: 1422–1425. Copyright the American Association for the Advancement of Science.

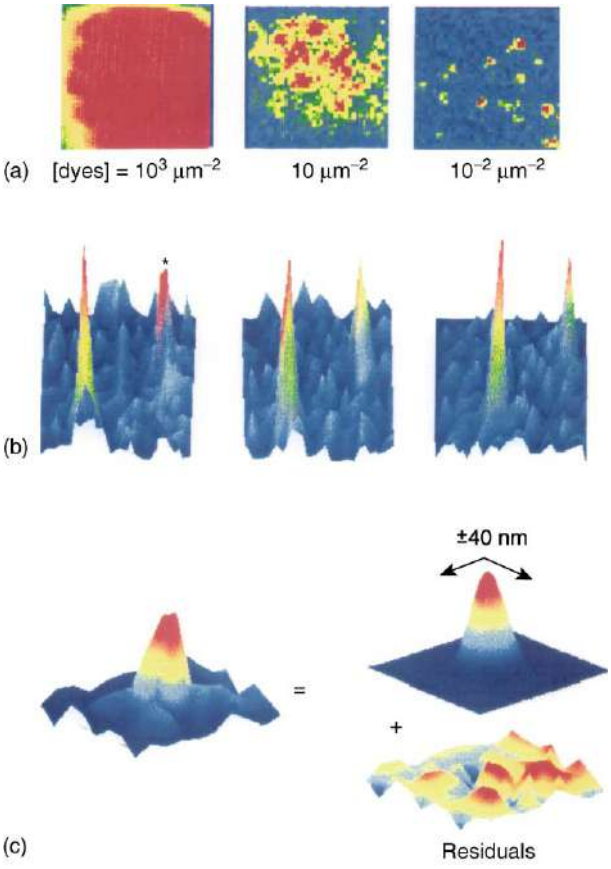

Figure 4 Single molecule diffusion in two dimensions. (a) $6.8 \times 6.8\ \mu m^2$ images of fluorescently labeled lipids (tetramethylrhodamineinthiocarbamoyl, linked to 1,2-dihexadecanoyl-*sn*-glycero-3-phospho-ethanolamine, DHPE) in a lipid double layer membrane (palmitoyloleoylphosphocholine, POPC) at different concentration. From left to right: 10^3 dyes/μm^2, 10 dyes/μm^2, 10^{-2} dyes/μm^2. Each image was obtained with an epifluorescence microscope, illuminated by a defocused 514 nm Ar laser line ($\sim$60 kW/cm^2), and 5 ms integration on a nitrogen cooled CCD camera. Color scale: blue = 0, $d = 60$ counts). Image intensity at 10^3 dyes/μm^2 was divided by 60. (b) Sequence of three $5.4 \times 5.4\ \mu m^2$ images of two labeled lipids observed at 105 ms intervals (actual acquisition rate was one frame every 35 ms, integration time: 5 ms). (c) Subregion of a $2.4 \times 2.4\ \mu m^2$ membrane area showing the peak marked by a star in B. The nonlinear least-square fit of a two-dimensional Gaussian profile to the peak yields values of the fluorescence intensity $I = 168 \pm 25$ counts, width $r = 480 \pm 80$ nm and two-dimensional position with 40 nm resolution. Adapted from Schmidt T, Schutz GJ, Baumgartner W, *et al.* (1996) Imaging of single molecule diffusion. *Proceedings of the National Academy of Sciences USA* 93: 2926–2929. Copyright 1996, The Academy of Sciences.

parameters in the case of a confocal microscopy study of freely diffusing fluorescein isothiocyanate (FITC) in water, using avalanche photodiode detectors (APD) are: $E = 10^{-2}$, $Q = 0.85$, $\sigma = 2.8 \times 10^{-16}\,cm^2$, $A = 0.1\,\mu m^2$, $\lambda = c/\nu = 525\,nm$, $bV = 10\,Hz/\mu W$, $D = 100$ Hz. With an excitation power of 100 μW and a 1 ms integration time, a count rate of 629 counts/ms can be obtained, of which 1 comes from the background sources, leading to a SNR ~25 and SBR ~629. Although these are acceptable figures, they have to be balanced by the fact that single molecules have a finite lifespan. Single molecules in an oxygen-rich environment typically emit on the order of 10^6 photons before irreversible degradation (photobleaching). With the above parameters, this amounts to a total emission of ~3.2 ms. This is enough to observe freely diffusing molecules during their transit time of a few hundred microseconds. For a fixed molecule, however, it sets a stringent limit on the total duration of a single-molecule observation, and all efforts are put on reducing the excitation volume and intensity, as well as background sources. Similar considerations hold on the detector side, where a trade-off has to be found between time-resolution and quantum efficiency.

In addition to SNR and SBR issues, care has to be taken to ensure that the collected signal originates from a single molecule. Two different strategies can be envisioned:

(i) work at low concentration, such that at most one molecule is present in the excitation volume (Figure 5b), or equivalently, minimize the excitation volume;
(ii) use a selective excitation or emission detection protocol, such that only one molecule is excited or detected within the sampled volume (Figure 5c).

The first case has already been illustrated in Figure 2, in which a solution of fluorophores at low concentration yields separate bursts corresponding most of the time to transits of individual molecules. The second strategy can be adopted for the example illustrated in Figure 1, in which the ZPL of individual molecules lying in the wings of the ensemble spectrum are well separated. Tuning the laser to the peak absorption frequency of a molecule will excite this molecule only, allowing for a separate study of this molecule.

In summary, a set of six criteria can be defined for SMS:

1. The observed density of emitter is compatible with the known concentration of individual molecules and scales with the original concentration.
2. The observed intensity level is consistent with that of a single emitting molecule.
3. Each immobilized emitter has a well-defined absorption or emission dipole (for organic dyes usually linear, but for other emitters such as a fluorescent semiconductor nanocrystal, possibly circular).

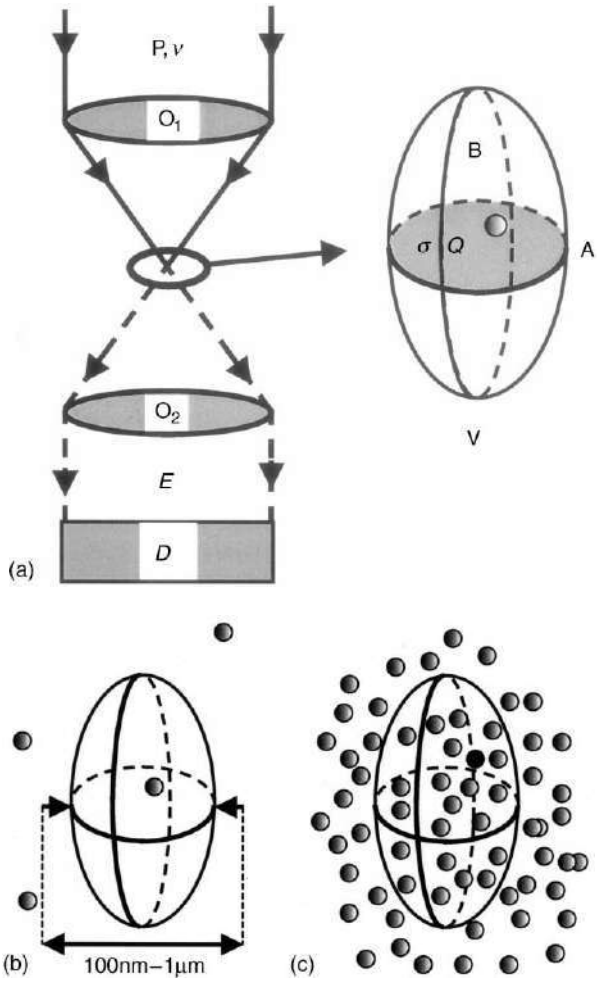

Figure 5 Signal, background and noise issues in single-molecule detection. (a) In a single-molecule spectroscopic experiment, excitation light (power *P*, frequency ν) is focused on the sample through an optical element O_1 (objective lens, fiber tip), exciting a diffraction-limited volume of section *A*, represented by an ellipsoid in the expanded view. The absorption process is characterized by an absorption cross-section σ and the emission by a quantum yield *Q*. The background contribution per watt of incident power is *B*, whereas the detector will have a dark count rate *D*. The signal is collected by an optical element O_2 (possibly confounded with O_1), and further guided to the detector, with an overall efficiency *E*. (b) The signal of a single molecule can be separated from that of other molecules if the probability of finding two or more molecules in the probed volume is negligible. This can be obtained either by a low concentration, or by reducing the excitation volume. (c) If each molecule in the excited volume can be characterized by a different spectral property (which may or may not depend on its exact environment), it is possible to use very selective excitation or detection techniques to probe just one molecule in the sampled volume (the black sphere in the ellipsoid).

4. Fluorescence emission exhibits only two levels (on/off) exemplified in blinking or photobleaching.
5. If there are more emission levels due to spectral jumps (see the section on fluorescence), blinking and spectral jumps are correlated.
6. Antibunching of the emitted photons permits us to test whether the detected signal comes from a single quantum emitter or more (see the section on biological studies below).

Types of Spectroscopy

Fluorophores

A variety of fluorescence emitters (or fluorophores) have been investigated with single-molecule spectroscopy techniques. Fluorescent organic molecules, green fluorescent proteins (GFP) and other biological fluorescent proteins, conjugated polymers (J-aggregates), light-harvesting complexes (e.g., B-phycoerythrin) comprising several fluorophores acting effectively as a single quantum emitter, or semiconductor nanocrystals, are but a few examples of systems which have been extensively studied. Each of these systems exhibits fluorescence based on specific processes, which cannot be described in detail here. A simple picture valid for single fluorescent dyes will be presented instead. It captures, however, the essential ingredients needed to describe the different parameters accessible with single-molecule spectroscopy.

Fluorescence Emission

Most single-molecule experiments use one-photon fluorescence excitation because of its relatively large cross-section ($\sim 10^{-16}\,\mathrm{cm}^2$). It also allows for a simple rejection of the excitation light: the emitted photons have a lower energy because of losses in intramolecular relaxation processes (the so-called Stokes shift). Excitation and emission processes are conveniently represented by a Jablonski diagram (Figure 6a) depicting the initial, final, and intermediate electronic and vibrational states of the molecule. Upon excitation to the first excited state, the molecule has a certain probability to be shelved in a nonemitting triplet state for up to milliseconds, resulting in 'dark' or off-states (Figure 6b,c). This results in a saturation of the emitted count rate *R* (Figure 6d) according to:

$$R = R_s \frac{I/I_s}{1 + I/I_s}, \qquad [3]$$

where the saturation intensity $I_s \sim 10\text{–}100\,\mathrm{kW/cm^2}$ and $R_s \sim 1\text{–}20\,\mathrm{MHz}$ for typical fluorophores. Incidentally, this puts a limit in SNR enhancement due to saturation of the emitted signal (triplet bottleneck). Other fluorescence emitting 'single molecules' like semiconductor nanocrystals

exhibit similar dark state intervals, although for different reasons (Auger ionization or surface trapping of carriers). Finally, spectral jumps are also observed at room temperature (Figure 6e). This phenomenon results in a shift of absorption and emission maxima due to the changing environment of the molecule, or a sudden conformational change. The corresponding changes in excitation and emission spectra are noticed by sudden fluctuations in the emitted intensity

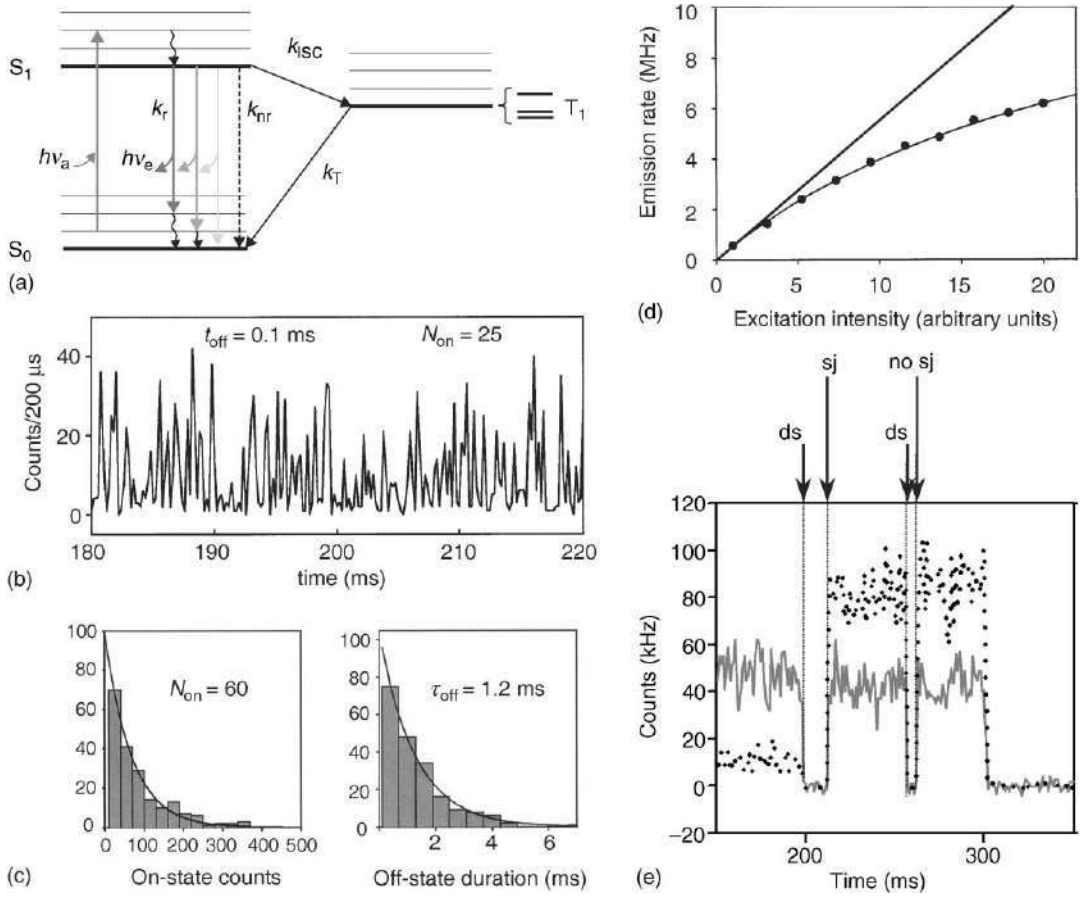

Figure 6 Single-molecule fluorescence. (a) Jablonski diagram for fluorescence (at room temperature). Upon absorption of a photon of energy $h\nu_a$ close to the resonance energy $E_{S_1} - E_{S_0}$, a molecule in a vibronic sublevel of the ground singlet state S_0 is promoted to a vibronic sublevel of the lowest excited singlet state S_1. Nonradiative, fast relaxation brings the molecule down to the lowest sublevel in picoseconds. Emission of a photon of energy $h\nu_e < h\nu_a$ (radiative rate k_r) can take place within nanoseconds and bring the molecule back to one of the vibronic sublevels of the ground state. Alternatively, collisional quenching may bring the molecule back to its ground state without photon emission (nonradiative rate k_{nr}). A third type of process present in organic dye molecules is intersystem crossing to the first excited triplet state T_1 (rate k_{ISC}). Relaxation from this excited state back to the ground state is spin-forbidden and thus the lifetime of this state is in the order of microseconds to milliseconds. Relaxation to the ground state takes place either by photon emission (phosphorescence) or nonradiative relaxation. The fluorescence lifetime is defined by $\tau = 1/\Gamma = (k_r + k_{nr})^{-1}$. (b) Fluorescence time trace (40 ms excerpt) of a Texas Red (TR) molecule attached to a 20-nucleotides-long single-stranded DNA molecule immobilized on a silanized glass coverslip. The confocal setup is similar to that depicted in Figure 2. Excitation: 514 nm, circularly polarized, 6 kW cm^{-2}, binning: 200 μs. On- and off-states (the latter corresponding to the molecule being in the triplet state) alternate stochastically. (c) Histograms of on-state count (left) and off-state duration (right) for the time trace partly shown in (b). The decay parameters of the exponential fits (solid line) are indicated. $t_{off} = 1.2$ ms is the triplet state lifetime. $N_{on}^{-1} = k_{ISC} = 1.7\%$. (d) Intensity dependence of the emission rate of a single Texas Red molecule. The data departs from a linear dependence because of saturation due to intersystem crossing. The fitted curve yields a saturation rate of $R_S = 6.5$ MHz and a triplet lifetime $t_{off} = (k_{ISC}\ R_S)^{-1} = 10$ μs. The difference between this value and the value obtained in (c) illustrates the heterogeneity uncovered by single-molecule measurements. (e) Spectral jumps of a single Texas Red molecule attached to a single-stranded DNA molecule. ds: dark state, sj: spectral jump, no sj: dark state not followed by a spectral jump. Cases where spectral jumps are observed without noticeable dark states may still correspond to dark states with durations shorter than the time resolution of the experiment. (b)–(d) Adapted from Ha T, Enderle T, Chemla DS, *et al.* (1997) Quantum jumps of single molecules at room temperature. *Chemical Physics Letters* 271: 1–5. Copyright 1997, Elsevier Science B. V. (e) Adapted from Ha T, Enderle T, Chemla DS and Weiss S (1996) Dual-molecule spectroscopy: molecular rules for the study of biological macromolecules. *IEEE Journal of Selected Topics in Quantum Electronics* 2: 1115–1127. Copyright 1996 IEEE.

(due to wavelength dependence of the absorption cross-section).

Fluorescence Lifetime

A fluorescence photon is emitted within nanoseconds after absorption of the excitation photon. The decay law is usually mono-exponential, characterized by a lifetime $\tau = \Gamma^{-1}$ (Figure 6a). However, the lifetime depends on the emission spectrum (which can change due to spectral shifts) and the local environment. For instance, the proximity to a dielectric or metallic surface markedly changes it via perturbations of the intramolecular transition matrix elements. Thus, fluorescence lifetime is a sensitive probe of the environment of the molecule (Figure 7). Lifetime measurements require a pulsed laser source and time-correlated single-photon detection and analysis techniques.

Fluorescence Polarization

The efficiency of photon absorption by a fluorescent molecule is proportional to $(\boldsymbol{E}\cdot\boldsymbol{\mu})^2$ where $\boldsymbol{E}$ represents the local electric field and $\boldsymbol{\mu}$ the molecule's absorption dipole moment. For a fixed fluorophore, modulation of the excitation polarization thus allows the determination of the absorption dipole orientation (the emission dipole being generally nearly parallel to it). This information in turn allows the determination of the spatial orientation of the molecule. For a mobile molecule, however, more information is needed, since the emission dipole has time to tumble significantly during the few nanoseconds between absorption and emission. The emission polarization is needed to fully recover the relevant information, as illustrated in Figure 8. Different experimental schemes can be used, besides the near-field optical one described in Figure 3 and the confocal one described in Figure 8. In particular, it is important to recover the projection of the polarization on more than two orthogonal axes. Approaches taking advantage of aberrations induced by index mismatch in high numerical aperture (NA) objectives, or using TIR excitation with alternating polarizations, or using amplitude or phase masks, allow a complete determination of this orientation.

Fluorescence Resonance Energy Transfer

A special case of the local influence of the environment on fluorescence is encountered in fluorescence resonance energy transfer (FRET, see Figure 9a), first theoretically explained by Förster in the late 1940s. In case of the presence of a nearby molecule (acceptor) having an absorption spectrum overlapping that of the donor emission, there is a possibility that the energy absorbed by the donor is transferred nonradiatively to the acceptor. The efficiency E of this transfer is given by:

$$E = \left(1 + \left(\frac{r}{R_0}\right)^6\right)^{-1} \qquad [4]$$

where r is the distance between the two emitting centers and R_0 is the Förster radius (in Å):

$$R_0 = a(\kappa^2 n^{-4} Q_D J(\lambda))^{1/6} \qquad [5]$$

In this complex expression, $a = 9.78 \times 10^3$, n is the medium refractive index, Q_D the donor fluorescence quantum yield in the absence of the acceptor and $J(\lambda)$ is a spectral overlap integral (in M^{-1} cm^3). κ^2 is a geometric factor, which averages out to 2/3 in the case of freely rotating dyes. A typical order of magnitude for R_0 is a few nanometers, which gives the range of the energy transfer. FRET is thus very sensitive to the donor–acceptor distance, which makes it a useful molecular ruler. Experimentally, the efficiency is measured as a function of the recorded donor (F_D) and acceptor (F_A) emissions (dual-channel measurement) via:

$$E = \left(1 + \alpha \frac{F_D}{F_A} \frac{Q_A}{Q_D}\right)^{-1} \qquad [6]$$

where Q_A and Q_D are the fluorescence quantum efficiencies of the donor and acceptor, respectively, and α is the ratio of the acceptor and donor channel detection efficiencies. An illustration of this utilization is given in Figure 9b–d, where different predefined distances between donors and acceptors are set by rigid double-stranded DNA molecules. The corresponding donor–acceptor distances are readily measured on diffusing molecules, with subnanometer precision.

Alternatively, the transfer efficiency can be measured via the fluorescence lifetimes as:

$$E = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \qquad [7]$$

where $\tau_{D(A)}$ and $\tau_{D(0)}$ are the donor lifetimes in the presence or absence of an acceptor, respectively.

Other Spectroscopies

Other types of spectroscopic techniques can be used at the single-molecule level. Fluorescence can for instance be excited via a two-photon absorption process, whereby a laser excitation with half the photon energy needed to attain the excited state

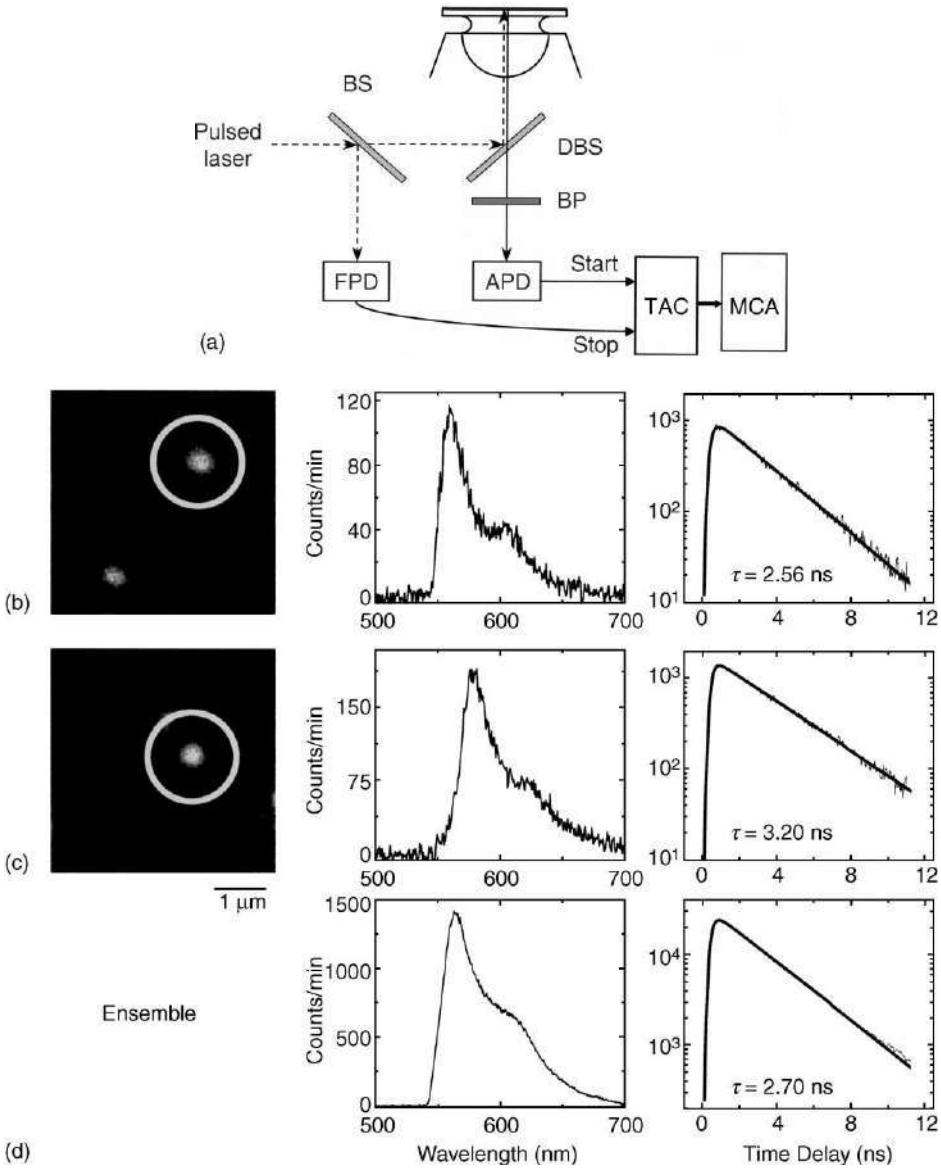

Figure 7 Lifetime measurements. (a) Schematic of a lifetime measurement setup. BS: beamsplitter used to deflect part of the incident pulse towards a fast photodiode (FPD); DBS: dichroic beam splitter used to spectrally separate incident from emitted light; BP: bandpass filter further rejecting stray light; APD: avalanche photodiode. For each detected photon, the APD emits a pulse, which is fed to the start input of a time-to-amplitude converter (TAC). A stop pulse from the laser immediately follows it (regularly emitted at a rate of τ_L^{-1}), allowing the measurement of the separation between both pulses. A multichannel analyzer (MCA) histograms the arrival times, resulting in the curves shown in the right part of Figure 2b–d. (b) and (c) Fluorescence images (left panels), corresponding spectra (middle), and fluorescence decays (right) for two molecules located at a PMMA–air interface. The peak emission wavelength λ_p was 560 nm in (b) and 578 nm in (c). Lifetimes were fitted to a single exponential (dotted curves), with decay times of $\tau = 2.56$ ns (b) and 3.20 ns (c). A ensemble measurement of spectrum and fluorescence decay averaged over several hundred of molecules is shown in (d). The ensemble values are $\lambda_p = 565$ nm and $\tau = 2.70$ ns. (b)–(d): Adapted from Macklin JJ, Trautman JK, Harris TD and Brus LE (1996) Imaging and time-resolved spectroscopy of single molecules at an interface. *Science* 272: 255–258. Copyright 1996, The American Association for the Advancement of Science.

is used. However, due to the very low cross-section and quadratic dependence in the incident energy (two simultaneous photons are needed for excitation), an excitation power several orders of magnitude larger than that for one-photon excitation is needed. In compensation, excitation takes place in a much more reduced volume of the sample, reducing the out-of-focus background contribution. The higher power required increases the probability of photobleaching due to the increased probability of

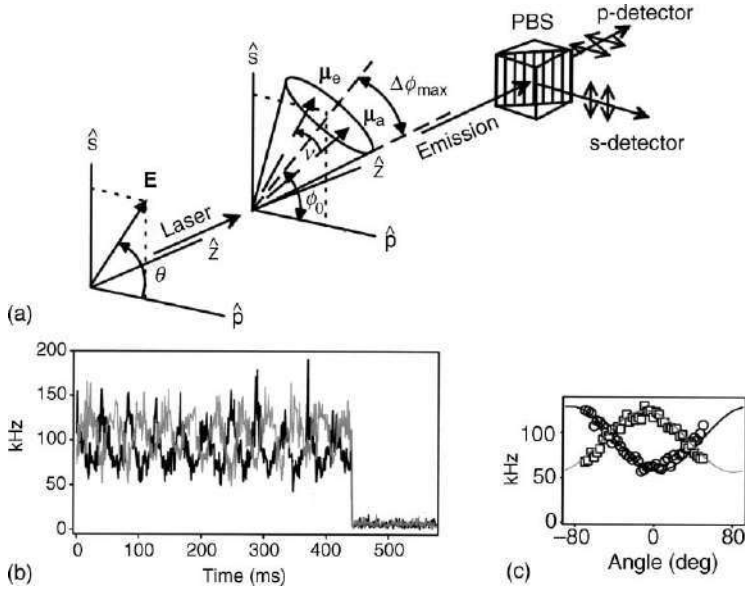

Figure 8 Polarization Measurements. (a) Experiment schematic. **E** is the electric field, making an angle θ with the *p* polarization axis. The excitation propagates along the *z* axis, which is also the collection axis. $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_e$ are the absorption and emission dipole moments, initially aligned. ν represents the rotational diffusion of the emission dipole during the excited lifetime. The dipole is supposed to be confined in a cone positioned at an angle ϕ_0 projected on the (*s*, *p*) plane, and having a half-angle $\Delta\phi_{max}$. A polarizing beam-splitter (PBS) splits the collected emission into two signals I_s and I_p, which are simultaneously recorded by APDs. (b) Simultaneously recorded I_s (black) and I_p (gray) of a molecule rapidly rotating in a liquid. (c) Same data as in (b), but averaged over 11 on-periods. The fit corresponds to a $\Delta\phi_{max}$ close to 90° (freely rotating molecule) and permits the determination of the constrained rotational diffusion parameters. Adapted from Ha T, Glass J, Enderle T, *et al.* (1998) Hindered rotational diffusion and rotational jumps of single molecules. *Physical Review Letters* 80: 2093–2096. Copyright 1998, The American Physical Society.

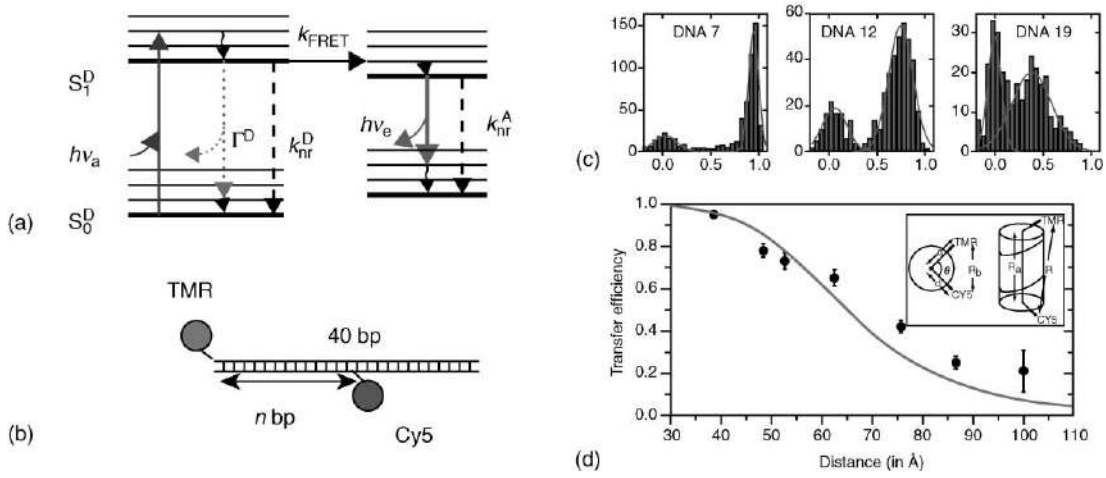

Figure 9 Single pair fluorescence resonance energy transfer. (a) Jablonski diagram for FRET. Fluorescence energy transfer involves two molecules: a donor D and an acceptor A whose absorption spectrum overlaps the emission spectrum of the donor. Excitation of the acceptor to the lowest singlet excited state is a process identical to that previously described for single-molecule fluorescence (**Figure 6a**). In the presence (within a few nm) of a nearby acceptor molecule, donor fluorescence emission is largely prevented by energy transfer to the acceptor by dipole–dipole interaction. The donor exhibits fluorescent emission following the rules outlined in **Figure 6a**. (b) DNA *n* constructs used for the FRET distance study. Tetramethylrhodamine (TMR) is attached to one end of the DNA and CY5 is attached to the *n*th base from the end ($n = 7$, 12, 14, 16, 19, 24 and 27). (c) FRET histograms extracted from time traces for DNAs 7, 12, and 19, as shown in **Figure 2b**. Double Gaussian fits extract numbers for the mean (width) of the higher efficiency peak of 0.95 ± 0.05, 0.75 ± 0.13, and 0.38 ± 0.21, respectively. The peak around 0 efficiency corresponds to bleached or inactive acceptor molecules. (d) Mean FRET efficiencies extracted from FRET histograms plotted as a function of distance for DNA 7, 12, 14, 16, 19, 24, and 27. Distances are calculated using the known B-DNA double helix structure. The solid line corresponds to the expected Förster transfer curve with $R_0 = 65$ Å. (b)–(c) Adapted from Deniz AA, Dahan M, Grunwell JR, *et al.* (1999) Single-pair fluorescence resonance energy transfer on freely diffusing, molecules: Observation of Förster distance dependence and subpopulations. *Proceedings of the National Academy of Sciences USA* 96: 3670–3675. Copyright 1999, The National Academy of Sciences.

photochemical degradation in the long-lived triplet state and the interplay of multiphoton ionization processes.

Although Raman scattering cross-sections ($\sim 10^{-30}$ cm^2) are orders of magnitude lower than the fluorescence cross-section, they can be enormously amplified if the single molecule is trapped in a 'hot spot' of a metallic nanoparticle, giving rise to surface-enhanced Raman scattering (SERS). Highly resolved vibrational information can be recovered, exhibiting much more stable emission than fluorescence. Moreover, this technique can be applied to nonfluorescent molecules. Similarly to fluorescence, the technique can be extended to nonlinear (two-photon) regimes, as in surface-enhanced hyper-Raman scattering and surface-enhanced anti-Stokes Raman scattering. However, since they require the single molecule to be located on rare sites of a metallic particle, these methods have only been applied to fundamental chemical studies of molecular species.

Finally, extending the range of the FRET 'molecular ruler', electron transfer allows measuring distances between an electron donor and an acceptor within a few ångströms (Å), via opening of a nonradiative channel for the relaxation of the electron in the excited state. Lifetime measurements ($\tau = 1/\Gamma = (k_r + k_{nr})^{-1}$ as defined in Figure 6) yield distance information via the electron transfer rate E_{ET}:

$$E_{ET} = k_0 e^{-\beta r}, \qquad [8]$$

where β^{-1} is of the order of one Å and k_0 of the order of 10^{15} s^{-1} are constants characterizing the donor–acceptor pair.

Applications of Single-Molecule Spectroscopy

Single-molecule spectroscopy has already proven to be a useful tool in various domains. The following examples are but a few of an expanding number of applications of SMS.

Molecular Physics

The realization that the fluorescence of a single molecule could be detected with a good SNR over extended periods of time, especially embedded in a crystalline matrix at cryogenic temperature, has led to high-resolution experiments aimed at determining molecular energy levels and transitions. In particular, studies of transitions between sublevels of the triplet state by electron spin resonance (ESR) using superimposed radio-frequency pulses have permitted a precise test of the quantum description of such systems. The dc and ac Stark effect (the shift of a transition energy in the presence of a fixed or alternating electric field) has also been studied in several systems, allowing the symmetry and deformations of the molecule to be determined. The method has been extended to study the photophysics of luminescent polymers (J-aggregates), light-harvesting complexes, green fluorescent proteins (GFP) or semiconductor nanocrystals, and other fundamental systems.

Material Science Studies

Single-molecule photophysical properties are extremely sensitive to their local (nanometer-sized) environment (Figure 1). A precise knowledge of these properties allows using single molecules as sensitive probes of their environment, as illustrated by experiments designed to study the local dynamics of a polymer matrix at the onset of the glass transition (Figure 10).

Molecules of the organic fluorophore Rhodamine 6G (R6G) dispersed within a thin (250 nm) poly(methacrylate) film were observed at temperatures slightly above the melting temperature of the polymer, using a confocal microscope detecting the *s* and *p* polarization of the emitted fluorescence (Figure 8a). Observations are performed under nitrogen atmosphere and at low excitation power (20 nW) in order to prevent oxidation, thus allowing long (several hours) observation time. Each observed R6G molecule exhibits a slow rotational diffusion, sometimes accompanied by lateral or transverse diffusion on a much longer time-scale (Figure 10a). Computation of the autocorrelation function of the emission polarization shows a multi-exponential decay, pointing to a distribution of time-scales of rotational diffusion. A closer look at subwindows of the time trace reveals, however, that the decay comes closer to mono-exponential behavior at shorter time-scales (Figure 10b,c). This study demonstrates that an individual molecule probes an increasing number of different environments over time. At long time-scales, the observable characteristics (such as the autocorrelation function) are similar to those measured on an ensemble of molecules, as expected from the ergodicity hypothesis. At short time-scale, the molecule reveals the local homogeneous and stable characteristic of its nanoenvironment.

Biophysical and Biological Studies

SMS has allowed a reassessment of long-standing questions in biophysics, biochemistry, and biology, by giving scientists the possibility to look at individual building blocks of biological processes: DNA, RNA, proteins and other biomolecules or biomolecular

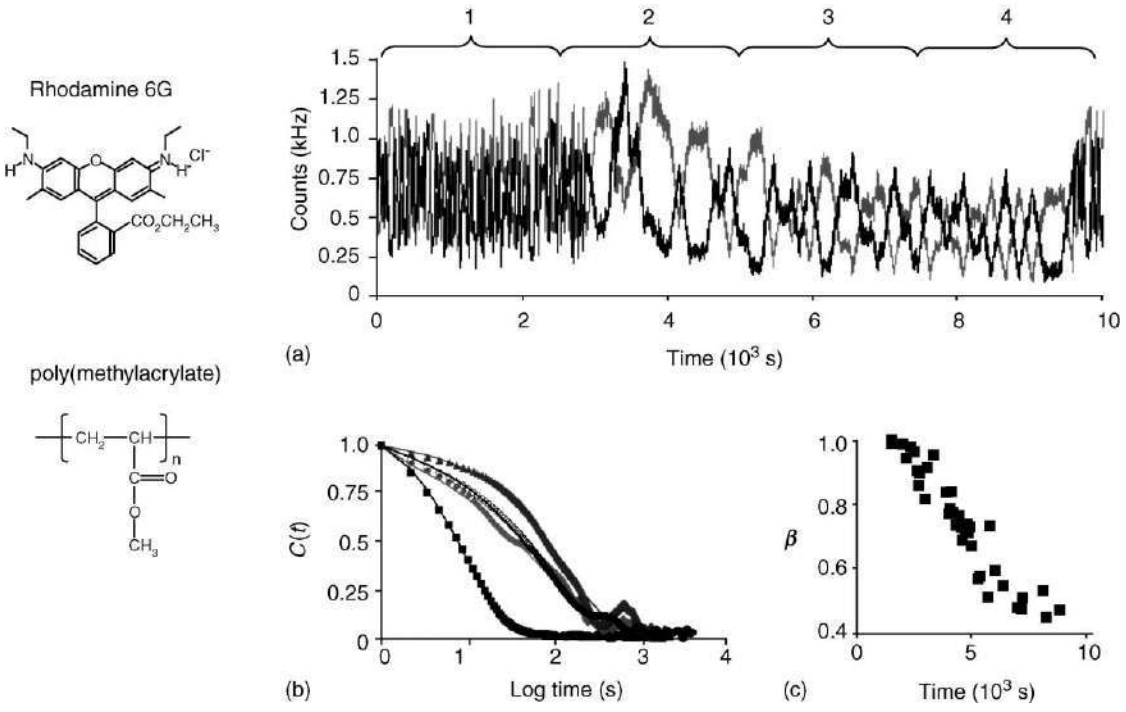

Figure 10 Polymer dynamics near the glass transition. (a) Fluorescence intensity of a single long-lived Rhodamine 6G molecule in poly(methylacrylate) (PMA) at 291 K detected at orthogonal polarizations (gray: *s* and black: *p* polarizations). The melting temperature of PMA is $T_g = 281$ K. The time trace integration is 1 s. This time trace shows the anticorrelation of both polarizations, characteristic of rotational diffusion of the molecule. Different regimes are observed along the time trace (1–4), which could be due to either a modification of the local 'cage' in which the molecule rotates, or escape of the molecule to another nearby nano-environment. The autocorrelation function of polarization computed over the whole trace (not shown) is well fitted by a model adapted to the nonexponential relaxation of polymers, the Kohlrausch–Williams–Watt (KWW) model:

$$\mathrm{KWW}(t) \propto e^{-(t/\tau)^{\beta}}.$$

with $\tau = 88$ s, $\beta = 0.46$. (b) Correlation functions of subsets 1–4 of the full trace. 1 (squares): $\tau = 15$ s, $\beta = 1.0$. 2 (triangles): $\tau = 112$ s, $\beta = 0.77$. 3 (open circles): $\tau = 90$ s, $\beta = 0.59$. 4 (diamonds): $\tau = 64$ s, $\beta = 0.59$. The β exponent is closer to 1, showing that the rotational diffusion is an unhindered one at short time-scales, converging to the bulk exponent at larger time-scale due to ergodicity. This experiment illustrates changes of environment as the observed molecule diffuses and rotates through the polymer mesh, changes that are hidden in ensemble measurements. (c) Values of β as a function of photochemical survival time in PMA at $T_g + 5$ K. Adapted from Deschenes L and Vanden Bout DA (2001) Single-molecule studies of heterogeneous dynamics in polymer melts near the glass transition. *Science* 292: 255–258. Copyright 2001, The American Association for the Advancement of Science.

complexes formed by association of a few of these molecules.

To simplify the task, some biological molecules are intrinsically fluorescent in their active state, as the cholesterol oxidase (COx) enzyme of Brevibacterium sp., an enzyme that catalyzes the oxidation of cholesterol by oxygen. As a result, it is relatively easy to study the dynamics of this enzyme at the single-molecule level using a confocal setup with enzymes immobilized in an agarose gel (Figure 11). The intervals during which the enzyme is active ('on' state) are followed by intervals during which it is inactive ('off' states). The fluorescence recording of a single enzyme thus appears as a 'blinking' time trace, as shown in Figure 11b. Relevant kinetic parameters can be extracted from the distribution of 'on' and 'off' times. Apart from the usual heterogeneity between enzymes of the same batch characterized by different kinetic parameters for the catalysis reaction, this experiment reveals a dynamic heterogeneity, which shows up for each individually studied enzyme as a tendency for short 'on' times to follow short 'on' times, and for long 'on' times to follow long 'on' times (Figure 11c). This variable behavior of a single enzyme is reminiscent of the variable rotational dynamics of Rhodamine 6G molecules in a polymer matrix (see the previous section), and can be studied only thanks to single-molecule techniques.

Not all biological molecules can be studied in such a convenient way using their intrinsic fluorescence. Several labeling schemes are currently available, which allow quantitative, site-specific labeling of proteins, nucleic acid or lipid molecules with few biochemical side effects.

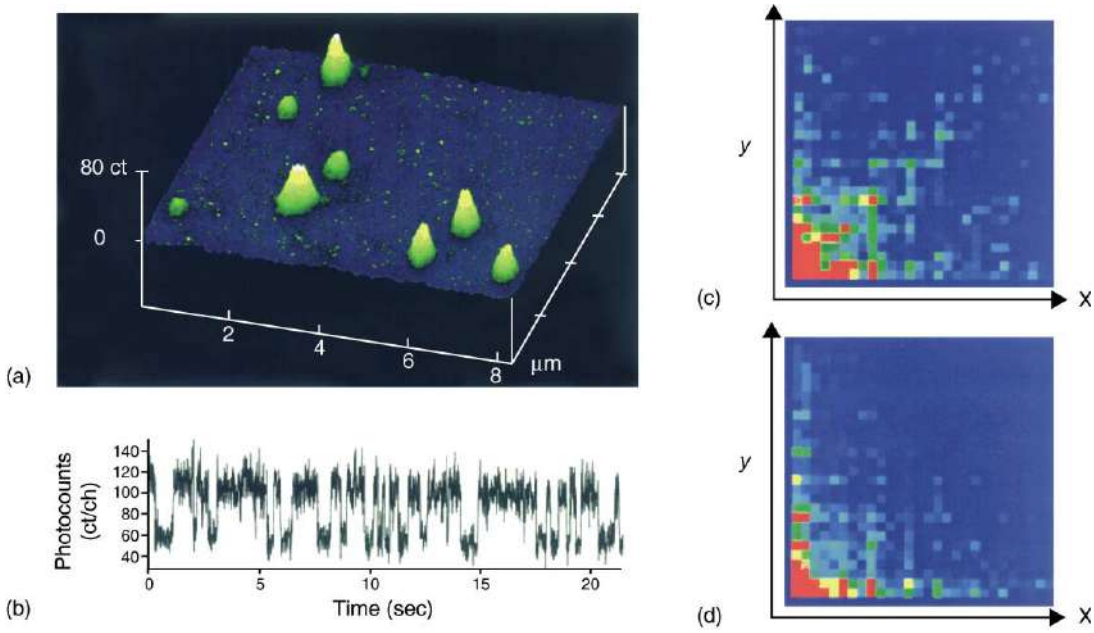

Figure 11 Single-enzyme dynamics. (a) Fluorescence image of single CO_x molecules immobilized in a 10-µm-thick film of agarose gel of 99% pH 7.4 buffer solution (8 µm × 8 µm, 4 min scan, 500 nW excitation at 442 nm). The emission is from the fluorescent active site, FAD, which is tightly bound to the center of CO_x. Each individual peak is attributed to a single CO_x molecule. The intensity variation between the molecules is due to different vertical positions in the excitation volume. (b) Real-time observation of enzymatic turnovers of a single CO_x molecule catalyzing oxidation of cholesterol molecules. This panel shows a portion of an emission intensity trajectory recorded in 13.1 ms per channel. The trajectory was recorded with a cholesterol concentration of 0.2 mM and saturated oxygen concentration of 0.25 mM. The emission exhibits stochastic blinking behavior as the flavin adenine dinucleotide (FAD), the CO_x fluorescent active site toggles between oxidized (fluorescent) and reduced (nonfluorescent) states, each on–off cycle corresponding to an enzymatic turnover. More than 500 on–off cycles are recorded in this trajectory. (c) The 2D conditional probability distribution for on-times (*x* and *y*) of two adjacent turnovers as derived from the trajectories of 33 CO_x molecules with 2 mM 5-pregene-3β-20α diol substrate, a derivative of cholesterol. The scales of the *x*- and *y*-axes are from 0 to 1 s. A subtle diagonal feature is present, indicative of a memory effect. (d) The 2D conditional histogram for two on-times separated by 10 turnovers for the CO_x molecules in (c). The diagonal feature vanishes because the two on-times become independent of each other at the 10-turnover separation. The color code in (c) and (d) represents the occurrence (*z*-axis from 350 (red) to 0 (purple). Adapted from Lu HP, Xun L and Xie XS (1998) Single-molecule enzymatic dynamics. *Science* 282: 1877–1882. Copyright 1998, The American Association for the Advancement of Science.

An example of this versatility has been provided by the labeling of the central part of a rotary protein, F_1-ATPase. In this experiment (not shown), the molecular rotor was labeled with a single fluorophore whose orientation, detected by emission polarization measurements, directly reported the angle of the rotor with respect to the shaft. The small size of the fluorophore guaranteed that the protein motion would not be hindered, contrary to previous experiments, which used much larger reporters (micron-size latex beads or fluorescent actin filaments). This experiment reproduced the previous results, showing that the rotor performed 120° steps (recent work hints at substeps of 90° and 30°).

One of the most promising domains where SMS can be of great importance is the study of protein folding. Doubly labeled proteins can be monitored as they undergo conformational changes, taking advantage of the distance dependence of fluorescence energy transfer (demonstrated for a model DNA system in the section on FRET, Figure 9).

Figure 12 shows one of many examples of such a study, performed on diffusing molecules in solution, a configuration that suppresses uncontrolled effects of nearby surfaces. The enzyme chymotrypsin inhibitor 2 (CI2) is a model system for protein folding, believed to have two clearly distinct folding states, which can be controlled by the concentration of denaturant guanidinium chloride. Using single-pair FRET (spFRET) techniques, it is possible to sort out the proportion of folded and unfolded molecules present at different denaturant concentration (Figure 12b). These measurements yield the same titration curve as obtained by ensemble measurements (Figure 12b) and can give access, under some simplifying assumptions, to the energy landscape of the folding reaction. The additional information provided by spFRET is

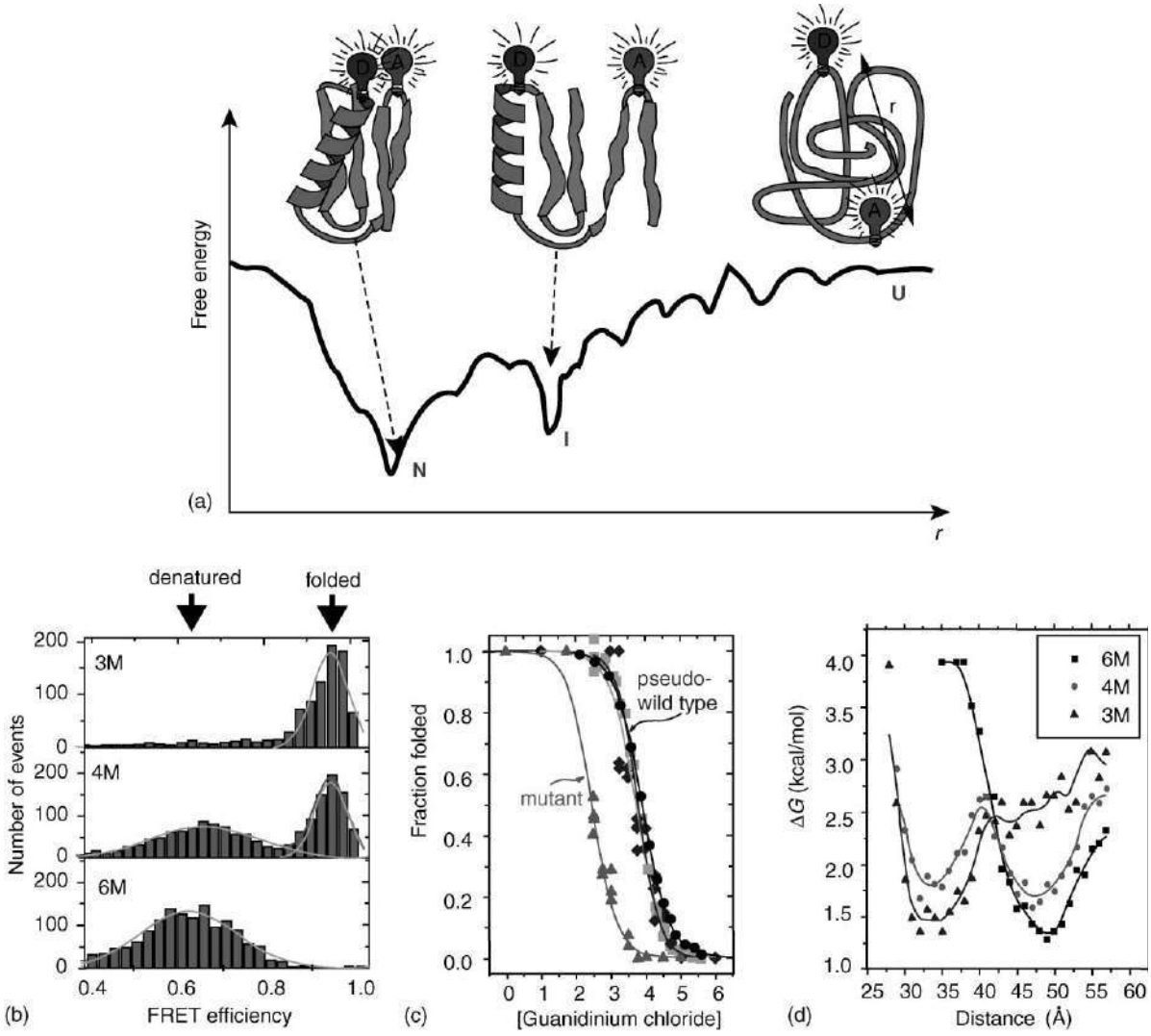

Figure 12 Protein folding. (a) Energy landscape for a protein folding reaction represented as a cartoon of the free energy as a function of distance *r* between donor and acceptor. The rugged energy landscape has a funnel shape with three extrema: U (unfolded state), I (intermediate state) and N (folded or native state). In this labeling scheme, donor and acceptor are close to one another in the native, folded state (N), resulting in a high FRET efficiency. There is reduced energy transfer (larger distance) in the denatured, unfolded state (U). (b) Single-molecule protein folding data. Single-pair FRET (spFRET) histograms of the pseudo-wild type enzyme chymotrypsin inhibitor 2 (PWT CI2) at 3, 4, and 6M denaturant (guanidinium chloride). The FRET efficiency exhibits two peaks, one for folded molecules (high FRET) and one for unfolded molecules (low FRET). Population is transferred from folded to unfolded at higher denaturant concentration, as expected from the protein labeling schematically depicted in (a). (c) Ensemble and single molecule denaturation curves for PWT and K17G mutant CI2. Symbols correspond to data; lines showing sigmoidal fits. PWT CI2: ensemble tryptophan fluorescence experiment (circles), ensemble FRET (squares), spFRET (diamonds); K17G mutant CI2 spFRET (triangles). (d) Free energy functions for PWT CI2 at 3, 4, and 6M denaturant. The solid lines represent a smoothing of the data and are only meant to guide the eye. (a) Adapted from Weiss S (2000) Measuring conformational dynamics of biomolecules by single-molecule fluorescence spectroscopy. *Nature Structural Biology* 7: 724–729. Copyright 2000, Nature Structural Biology. (b)–(d) Adapted from Deniz AA, Laurence TA, Beligere GS, *et al.* (2000) Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proceedings of the National Academy of Sciences USA* 97: 5179–5184. Copyright 2000, The National Academy of Sciences.

the number of molecules in the two different states, providing direct evidence of the two-states model hypothesized from ensemble measurements.

Other biological systems have been addressed using the spFRET approach. Intramolecular spFRET (used in the CI2 study) allows monitoring conformational changes, whereas by labeling two different molecules, intermolecular spFRET permits the detection of association and dissociation events. For instance, spFRET has revealed transient intermediate states in

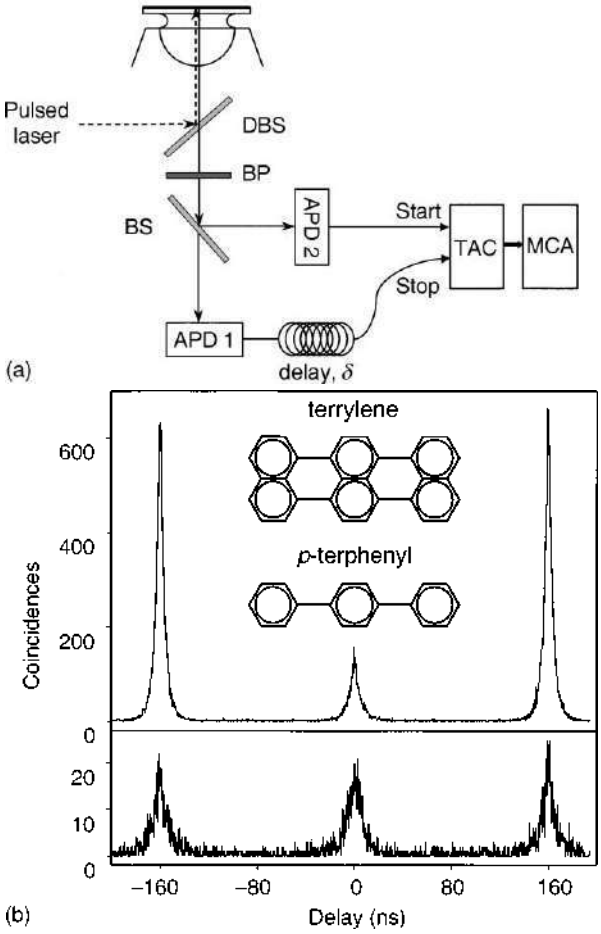

Figure 13 Single-photon source at room temperature. (a) Antibunching measurement setup. The basic elements are similar to those of a lifetime measurement setup (**Figure 7a**) except that the emitted filtered signal (BP: bandpass filter) is split in two with a 50/50 beamsplitter (BS). Two avalanche photodiodes (APD) detect the arrival of photons. One APD pulse is used as the start signal for a time-to-amplitude converter (TAC), which is stopped by the pulse coming from the second APD. Since the TAC response time is finite, it cannot detect simultaneously arriving pulses. A delay time δ is thus added to the stop signal, such that two photons impinging simultaneously on the two APDs will generate pulses with a separation of δ. If the start APD detects a photon from the previous excitation, the measured delay will be 160 ns $+ \delta$, whereas if the opposite happens, the delay will be $\delta - 160$ ns. (b) Results of measurements on a terrylene molecule (upper chemical structure) embedded in a p-terphenyl (lower chemical structure) at room temperature. Upper panel: histogram of delay measurements shifted by an amount δ. The histogram corresponds to 120 s of recording on an isolated emitting spot. The residual central peak corresponding to coincidence of photon arrival at APD 1 and 2 is the result of background contamination. Lower panel: histogram acquired over 300 s on a background spot. In both measurements, the laser pulse width was 35 ps, each pulse being separated by 160 ns. The lifetime of terrylene is measured to be 3.8 ns, which sets the width of the observed peaks. Excitation is performed at 532 nm. (b): Adapted with permission from Lounis B and Moerner WE (2000) Single photons on demand from a single molecule at room temperature. *Nature* 407: 491–493. Copyright 2000, Nature Publishing Group.

the Tetrahymena ribozyme, which had remained unnoticed in ensemble studies.

Quantum Optics

Low-temperature spectroscopic studies of single molecules have turned into a benchmark for testing predictions of quantum mechanics. In particular, one effect expected for a single quantum emitter is the absence of simultaneous emission of two photons: a fluorescent system which is excited at time t_0 has to emit a photon before it can return to the ground state, and be re-excited at a later time t_1. In conditions where the excitation takes place immediately as the electron returns to the ground state, the time difference $t_1 - t_0$ is of the order of the excited state lifetime, of the order of a few nanoseconds. This antibunching phenomenon was first observed in atoms and ions, and rapidly confirmed by low-temperature SMS. This observation is possible either using continuous wave or using pulsed excitation. In fact, nothing prevents this effect from occurring at room temperature, as demonstrated in **Figure 13** for pulsed excitation in which excitation photons are provided as well-separated bunches of a few hundred femtoseconds duration.

In this experiment, a system originally used at low temperature (terrylene in a crystal of p-terphenyl) is excited with a repetition rate of 6.25 MHz. The fluorescence emitted by a single molecule is detected in a confocal setup, split in two and directed to two different detectors (**Figure 13a**). In this way, the time interval between each pair of photons detected by the two detectors can be precisely measured and histogrammed. The result of this time-correlated single-photon counting approach is the absence of coincidence of two photons (**Figure 13b**), besides the residual counts due to background contamination.

Since the quantum efficiency of this system is high, the experimental setup can be seen as a periodic single-photon source, with an extremely low probability for the simultaneous emission of two photons. This is precisely the kind of device that is needed for quantum cryptography, in which information encoded in the polarization of a beam should not be sensitive to a 'beamsplitter' attack. Future efforts will undoubtedly build on such 'turnstile' photon sources to move forward into optical quantum computing applications.

Perspectives

SMS is a mature discipline, which will certainly prove to be most useful in domains beyond those previously mentioned. Its relative ease of implementation will

certainly attract more scientists of other disciplines to this methodology, each of them having specific goals in mind that are difficult to foresee. The principal advantages of SMS outlined at the beginning of this article will remain, but improvements, especially towards simultaneous acquisition of all the fluorescence parameters described earlier are promising. New detectors will permit the combination of the high time resolution of single-photon counting devices with the large field of view and spectral resolution allowed by two-dimensional detectors. SMS will be associated with single molecule manipulation techniques, to correlate applied forces and molecular conformations. New ways of controlling local fields (electric, magnetic, or others) would take advantage of the non-invasiveness, high-temporal and spatial resolution of SMS to get a direct feedback of events at the nanometer scale. Undoubtedly, SMS will be a major tool in the future of nanotechnology.

List of Units and Nomenclature

Å	Ångström
cm^2	square centimeter
fl	femtoliter
K	kelvin
ms	millisecond
MHz	megahertz
μW	microwatt
nm	nanometer

APD	Avalanche photodiode
DNA	Deoxyribonucleic acid
FRET	Fluorescence resonance energy transfer
NA	Numerical aperture
PMMA	Polymethylmethacrylate
SERS	Surface-enhanced Raman scattering
SMS	Single-molecule spectroscopy
SNR	Signal-to-noise ratio
SNE	Signal-to-background ratio
TIR	Total internal reflection
ZPL	Zero phonon line

See also

Scattering: Raman Scattering.

Further Reading

Basché T, Moerner WE, Orrit M and Wild UP (1997) Single-Molecule Optical Detection, Imaging and Spectroscopy. Weinheim: VCH.

Craig DP and Thirunamachandran T (1998) Molecular Quantum Electrondynamics. An Introduction to Radiation Molecule Interactions. Mineola: Dovers.

Forkey JN, Quinlan ME and Goldman YE (2000) Protein structural dynamics by single-molecule fluorescence polarization. *Progress in Biophysics and Molecular Biology* 74: 1–35.

Kneipp K, Katrun H, Kneipp I, *et al.* (1999) Surface-enhanced non-linear Raman scattering at the single-molecule level. *Chemical Physics* 247: 155–162.

Marcus RA and Sutin N (1985) Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta* 811: 265–322.

Mehta AD, Rief M, Spudich JA, Smith DA and Simmons RM (1999) Single-molecule biomechanics with optical method. *Science* 283: 1689–1695.

Moerner WE and Orrit M (1999) Illuminating single molecules in condensed matter. *Science* 283: 1670–1676.

Nie S and Zare RN (1997) Optical detection of single molecules. *Annual Review of Biophysics and Biomolecular Structure* 26: 567–596.

Plakhotnik T, Donley EA and Wild UP (1997) Single-molecule spectroscopy. *Annual Review of Physical Chemistry* 48: 181–212.

Tamarat P, Maali A, Lounis B and Orrit M (2000) Ten years of single-molecule spectroscopy. *Journal of Physical Chemistry A* 104: 1–16.

Weiss S (1999) Fluorescence spectroscopy of single biomolecules. *Science* 283: 1676–1683.

Xie XS and Trautman JK (1988) Optical studies of single molecules at room temperature. *Annual Review of Physical Chemistry* 49: 441–480.

T

TERAHERTZ TECHNOLOGY

Contents

Coherent Terahertz Sources
Terahertz Physics of Semiconductor Heterostructures

Coherent Terahertz Sources

L Wang, Chinese Academy of Sciences, Beijing, China

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Terahertz (THz) radiation is usually referred to as an electromagnetic wave with frequencies ranging from 1 to a few terahertz ($1\,\text{THz} = 10^{12}\,\text{Hz}$). In the electromagnetic spectrum, terahertz radiation lies between the infrared and microwave, as shown in Figure 1. The techniques for generation and detection of THz radiation bridge photonics and electronics in the sense of the concepts and the techniques discussed below.

Physical quantities corresponding to 1 THz are listed as follows;

- Frequency, 1 THz
- Wavelength, 300 μm
- Wavenumber, 33 cm^{-1}
- Energy, 4.1 meV
- Temperature, 48 K

In this spectral region, there exist many rich physical, chemical, and biological phenomena. For example, many phonon resonance and other elementary excitations in condensed matter fall in this region; many molecular vibrations and rotations occur at THz frequency; conformation-related collective vibrational modes in macromolecules, especially in bio-molecules such as proteins and DNA, also lie in THz frequency. However, the development of THz technique lags far behind that of photonics and electronics, due to lack of feasible, reliable, and economic coherent THz sources. The situation has been rapidly changing since the 1990s. Benefiting from advanced material science and the ultrafast laser techniques, various coherent THz sources are being developed and commercialization is soon to follow. We will introduce these coherent THz sources and their main features, and discuss their generation mechanisms. More details can be found in the listed books and articles in the Further Reading section at the end of this article.

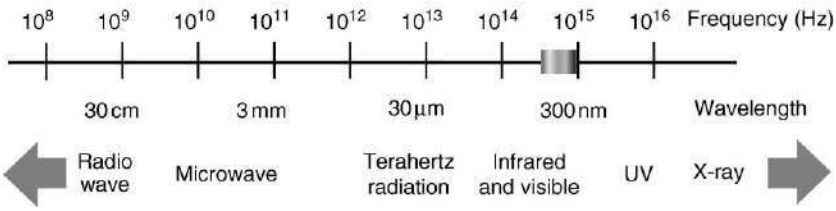

Figure 1 The electromagnetic spectrum with THz radiation sitting between the microwave and the infrared.

Coherent THz Radiation

When a time varying polarization, $\mathbf{P}(\mathbf{r},t)$, is generated in a medium and oscillates at THz frequency, it will radiate a THz wave, $\mathbf{E}(\mathbf{r},t)$, according to Maxwell's equation:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r},t) + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{E}(\mathbf{r},t) = -\mu_0 \frac{\partial^2}{\partial t^2}\mathbf{P}(\mathbf{r},t) \qquad [1]$$

or

$$\nabla \times \nabla \times \mathbf{E}(\omega) - \frac{\omega^2}{c^2}\mathbf{E}(\omega) = \omega^2 \mu_0 \mathbf{P}(\omega) \qquad [2]$$

in frequency domain, for a monochromatic wave or a Fourier component of a wave with a finite bandwidth. The properties of THz radiation are determined by the polarization source $\mathbf{P}(\mathbf{r},t)$. Incoherent THz sources exist in blackbody radiation, and are also detected in astronomy observations. If the phase evolution of the THz wave keeps in step, both spatially and temporally, it is called coherent radiation. A coherent THz source is more powerful and useful in spectroscopy, material characterization, imaging, and many other applications. It is clear that a coherent polarization source is first required for generation of a coherent THz wave. In practice, the working media for THz emitting can be free electrons, quasi-free electrons in solids, or bound electrons and other charge oscillations, such as plasmon oscillation, lattice vibration, etc. Classified by the excitation properties, the emitting process can be of resonance or nonresonance. Coherent polarization sources with different features can be created using various excitation techniques under different principles and device configurations, which lead to different coherent THz sources. Here we focus on coherent THz sources, primarily based on photonic techniques, and working in the spectral range from 1 THz to a few THz.

Gas THz Laser

Molecules have many rotational and vibrational energy levels spaced at THz frequency. For polar molecules, the radiative transition between these levels can radiate THz wave. Following the standard laser technology, molecules pumped either optically or electronically can be used as laser gain media for coherent THz wave generation. For THz gas lasers, the very low THz photon energy poses a problem for lasing, because the coherent radiative transition can be disrupted by thermal depopulation and dephasing. An effective population inversion is difficult to set up and maintain. Therefore, it is important to carefully choose working medium and energy levels, as well as the way to excite the molecules. A number of media has been used for THz laser, such as methanol and HCN vapor, working in both continuous wave (CW) and pulsed modes. For example, methanol lasers operate under excitation of the rocking and asymmetric deformation modes of methanol excited by a CO_2 infrared laser. HCN laser is pumped electrically, which needs a large voltage and current to excite the molecules, a long cavity to obtain sufficient gain, and dedicated control of the temperature in the cavity for stable operation.

In general, THz gas lasers can be tuned discretely in thousands of lasing lines, ranging from a few tens to a few hundreds of micrometers, with power output from μW to mW. THz gas lasers have existed since the early 1970s, and are the only commercialized THz lasers to date. They are still subject to development, for less bulky volume, enhanced tunability and efficiency, and reduced costs.

Free Electron THz Laser

A free electron laser (FEL) uses free electrons as working medium, rather than bound atomic or molecular states in a conventional laser. A typical FEL consists of three parts: (i) an electron source that generates an electron beam with high current; (ii) an accelerator to raise the electron energy; and (iii) a lasing cavity. After the electron beam is generated, electrons are first accelerated to a relativistic velocity, typically with energy of hundreds of MeV, and then enter the cavity consisting of end mirrors and a spatially periodic magnets array called a wiggler (Figure 2). Due to the Lorentzean force imposed by the periodic magnetic field, the high-energy electrons move in the cavity along a sinusoidal path, and emit coherent radiation with a wavelength determined by the spacing between magnets and the electron velocity, as well as the magnetic induction. With the feedback provided by the cavity mirrors, electrons are accelerated or decelerated continuously by the optical field, and are bunched via the resonant interaction. The collective motion of the electron bunches radiates powerful coherent synchrotron radiation. In a standard configuration, the wavelength of the radiation can be expressed as:

$$\lambda = \frac{\Lambda}{2\gamma^2}[1 + (k\Lambda B_0)^2/2] \qquad [3]$$

where Λ is the spatial period of the wiggler magnets, γ the Lorentz factor of the electron beam, B_0 the peak magnetic induction, and k a constant. With a suitable

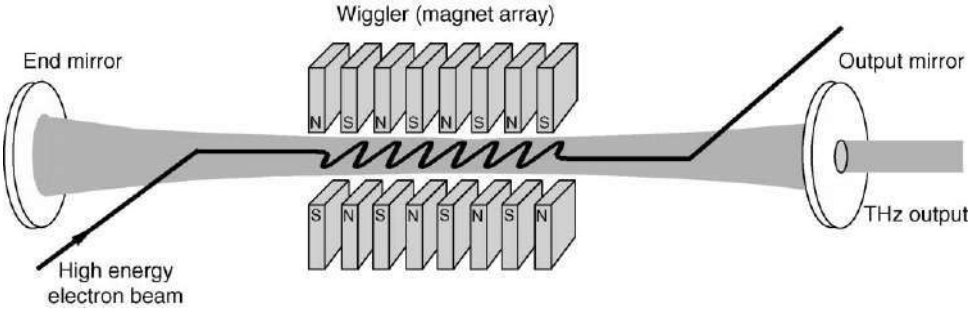

Figure 2 A schematic of THz generation in a free electron laser.

arrangement of the magnet array, as well as other related factors, the facility can work in THz spectral region.

An FEL generates tunable, coherent, high-power THz radiation. It is the most powerful coherent source of THz radiation available. Up to 50 W of average THz power has been generated in the Jefferson Lab recently. The free electron THz laser is becoming an important tools for studying THz radiation and its interactions with condensed matter and biological materials. On the other hand, the FEL is a complicated and expensive facility, and needs dedicated maintenance.

Semiconductor THz Laser

In the twentieth century, semiconductor devices have achieved tremendous success, both in electronic and photonic regime, such as transistors and diode lasers. The semiconductor industry has resulted in revolutionary changes to our world and everyday life. To fill the THz gap between electronics and photonics, great efforts have been made to meet the demand in compact, economic, tunable, and highly efficient THz sources. Encouraging advances have been made since we stepped into the new millennium.

In a semiconductor quantum well, the energy difference between a pair of sublevels can be artificially designed at THz frequency. As early as 1971, the concept of far infrared lasers based on intersub-band transition in semiconductor superlattices was proposed. Although the sublevels can be populated by electrical or optical pumping, the low photon energy at THz frequency causes serious difficulties for lasing, because the two levels sit so close to each other that many thermal processes could destroy the population inversion. Various designs and configurations have been proposed to tackle the problem. An attractive idea is so-called quantum cascade heterostructure. In this kind of device, the electrons are injected into a series of coupled quantum wells electrically. When the electrons are driven by the biased voltage, they can make radiative transition between a pair of sub-bands in a well. Subsequently, electrons enter the next well through resonant tunneling, and so on. However, this kind of device needs the materials of very high quality and suffers severely from fast depopulation of the excited states, even at very low temperatures.

As the advance in material science and unremitting efforts in pursuing the dedicated design of the device structures continued, the real breakthrough came in 2001, when Köhler *et al.* at Pisa demonstrated THz lasing in a quantum cascade heterostructure. Many improvements have been made since then; for example, the idea of using phonon resonance to selectively deplete the population of the lower sub-band has been successful, so that much stabler and robuster population inversion can be set up. The operation temperature has been raised to 136 K, well above the liquid nitrogen temperature, which opens the way for semiconductor THz lasers stepping into many more daily applications. These semiconductor devices are pumped with low voltages and currents at mA levels, and generate narrow bandwidth radiation with mW output at a frequency of a few THz. It is also probable that a four-level lasing system could even working at room temperature.

Another promising device is the p-germanium (Ge) laser developed recently. The Ge laser operates through the electrical excitation of hot holes in p-doped Ge. The laser cavity is formed by polishing the surfaces of the Ge crystal. These lasers could run at 5% duty cycles, with several mW output power. The output wavelength can be continuously tuned from 1 to 4 THz. At the moment, the devices have to operate at 20–30 K and consume about 10–20 W for refrigeration.

It is safe to predict that the feasibility, low cost, and compact semiconductor THz laser will emerge in the near future.

Beside the THz lasers, coherent THz radiation can also be generated by coherently exciting suitable media in other ways. Since the mid-1990s, the

development of ultrafast laser techniques has led to two important pulsed coherent THz sources, i.e., photoconductive antenna and optical rectification THz sources. A unique feature of the pulsed coherent sources is broad bandwidth, which is useful in spectroscopic applications. Figure 3 is a typical setup for pulsed THz generation and spectroscopic study, using photoconductive antenna or optical rectification as THz sources.

THz Photoconductive Antenna (PCA)

Different from a conventional antenna working in radio and microwave frequency that are driven directly by electrical current, a photoconductive antenna for THz wave generation works with ultrafast pulsed lasers. When the photoconductive gap is irradiated by an ultrafast laser pulse, photocarriers are generated in the semiconductor. These photocarriers are driven by a DC voltage bias and form a current transient. The antenna couples the electromagnetic field associated with the time varying current into free space, and produces a THz pulse. The THz pulse usually has a duration of a few picoseconds, with a broad frequency bandwidth centered at THz frequency. The output THz power scales with both the optical pulse power and the DC bias field:

$$E_{\mathrm{THz}} \propto \frac{\partial J}{\partial t} \propto \frac{\partial^2 n_{\mathrm{e}}}{\partial t^2} E_{\mathrm{bias}} \qquad [4]$$

The output power, frequency bandwidth, and polarization of the THz radiation depend on the photoconductive features of the semiconductor layer and the geometric structure of the antenna. For an optimal performance, the substrate should have a suitable bandgap for optical interband excitation, a high electron mobility and short carrier lifetime to support an ultrafast current transient, and high dielectric breakdown threshold and to sustain high bias voltage. High-power PCAs have been demonstrated with semi-insulating GaAs, ion-implanted GaAs, and In-GaAs. In the aspects of antenna structures, coplanar strip lines and large aperture emitters are the most effective. Average output power from a good system can reach 30–40 μW.

A broad bandwidth is a special characteristic of a THz source, which is particularly desirable in the spectroscopic study as a coherent probe source. From the property of Fourier transform, we know that the shorter the THz pulse duration, the broader the spectral band. PCAs made of GaAs typically have a useful bandwidth extending from ~100 GHz to 3 THz, which can be extended to 4 THz by tuning the pumping laser wavelength close to the bandedge. Up to 6 THz bandwidths have been reported for PCAs. The pulse duration and, therefore the achievable bandwidth, are ultimately limited by carrier mobility and TO phonon resonant absorption which is around 8.3 THz in GaAs.

In many applications, THz radiation is coupled to free space from the antenna using a closely attached silicon hemispherical lens. This practice increases the system's output and also provides control of the radiation pattern. The radiation pattern for the common dipole antenna is essentially dipolar, with a weak quadrupole component perpendicular to the bias field. In the far-field, the THz beam has a Gaussian cross-section with high-frequency components concentrated in the center.

As the first practical pulsed coherent THz source, PCA has been widely used for many applications in

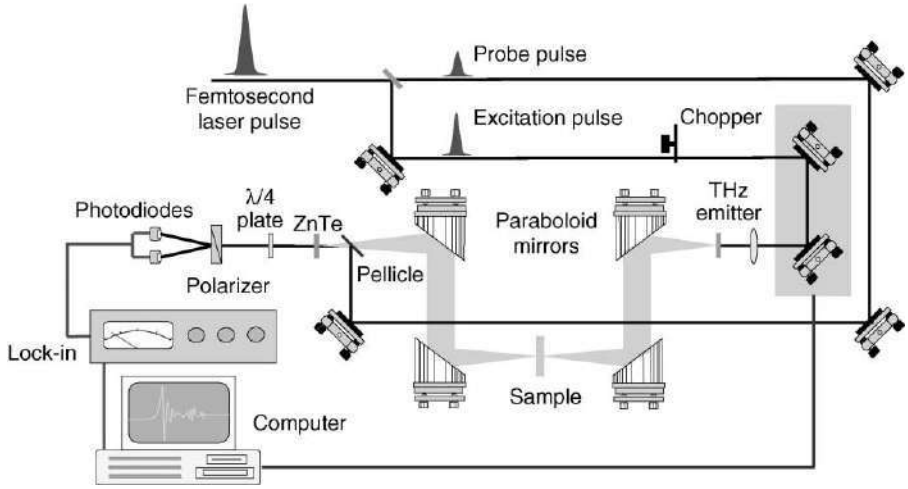

Figure 3 A typical setup for broadband pulsed THz generation and spectroscopic study with a femtosecond laser as excitation source.

scientific research and material characterization. Combined with commercial ultrafast optical fiber lasers, full functional and self-contained THz spectroscopy systems with PCAs as emitters, have been demonstrated and are under commercialization.

Optical Rectification

Optical rectification is a second-order nonlinear optical effect originating from susceptibility $\chi^{(2)}$, and was first observed in 1962 using two intense monochromatic nanosecond laser pulses. In this process, the interaction between laser and nonlinear optical crystal produces a frequency difference polarization:

$$P(0) = \chi^{(2)} E(\omega_1) E(-\omega_1) \qquad [5]$$

When a nonlinear optical crystal is irradiated by an ultrafast laser pulse, the different frequency components in the laser pulse will be coupled with each other by $\chi^{(2)}$, to induce frequency difference polarizations as:

$$P(\Omega) = \chi^{(2)}(\omega_1 - \omega_2 = \Omega) E(\omega_1) E(-\omega_2) \qquad [6]$$

A femtosecond laser pulse usually has a spectral linewidth of at least a few nm, the induced polarization $P(\Omega)$ has a finite frequency distribution centered at THz frequency in the far field, and results in THz radiation $E_{\mathrm{THz}} \propto \partial^2 P(t)/\partial t^2$.

THz radiation in free space by optical rectification was obtained in the early 1990s. One of the differences from a PCA device is that the THz output by optical rectification scales with both the optical pulse power and the DC bias field, while the THz radiation generated by optical rectification solely results from the incident laser. The optimal performance depends on several factors. The generation efficiency first depends on the magnitude of the $\chi^{(2)}$ and phase-matching condition between the optical and THz pulses. The effective magnitude of the $\chi^{(2)}$ coefficient varies with the crystal cut and orientation. Up to now, the most popular material is ZnTe, for its physical durability and excellent phase matching, when a Ti:sapphire femtosecond laser is used as the excitation source. With a moderately focused optical pump beam and a ZnTe crystal, nW T-ray average power can be generated by optical rectification. In general, the generation efficiency may be increased by increasing the laser power, but there are two factors limiting the effect. First, the incident laser power cannot exceed the damage threshold. Second, other second-order nonlinear processes may compete with the optical rectification, such as second-harmonic generation of the pump beam at high optical flux.

For THz radiation generated by optical rectification, the ultimate bandwidth is primarily determined only by the linewidth of the pump laser pulse. Because it is a nonresonance process in most cases, optical rectification can reach broader bandwidths than PCA. Using ultrashort optical pulses and thin nonlinear crystal, THz pulses have been generated with spectra extending to the mid-infrared, well beyond the phonon band of the materials. Like the photoconductive antenna, optical rectification has become a popular technique to generate coherent broadband THz radiation in various applications. Beside the conventional bulk inorganic crystals, many different media have been used for THz generation, such as organic crystal DAST, biased quantum wells, periodically poled LiNbO3 (PPLN), poled polymers, super-conducting thin films, etc.

Other Coherent THz Sources

When two light sources with slightly different wavelengths interact together with a nonlinear optical crystal, a beating polarization will be generated and will radiate a THz wave. It is called difference frequency generation, or three-wave mixing. Optical parametric processing is another way to generate a THz wave. In this process, a near-infrared pump beam generates a second NIR idler beam in a nonlinear crystal, and THz radiation can be generated from the beating of the pump and the idler. The merit of this technique is the continuous frequency tunability of the THz output, which is valuable in spectroscopic applications, and relatively cheaper nanosecond lasers can be used. If the nonlinear crystal is further placed in a cavity and the idler beam is amplified by feedback from end mirrors, a THz optical parametric oscillator is formed. A number of improvements have been made since the mid-1990s with this kind of device. More recently, THz parametric generation with an injection seeded idler beam has shown a reduced linewidth ($\Delta\nu/\nu \approx 10^{-4}$) and a peak THz power of over 100 mW for 3.4 ns pulses. A narrow linewidth combined with reasonable tunability make this kind of THz source attractive in spectroscopic studies.

Besides using photonic techniques, coherent THz radiation can also be generated by electronic techniques through increasing the output frequency of the microwave devices. For example, an electrically driven microwave generator, backward wave oscillator, can generate CW THz radiation up to 2 THz. A backward wave oscillator, running under optimal conditions, can provide up to 300 mW of polarized, narrowband radiation with a limited tunability about 30% of its central frequency.

See also

Imaging: Infrared Imaging. **Lasers:** Free Electron Lasers. **Optical Parametric Devices:** Overview. **Semiconductor Physics:** Infrared Lattice Properties. **Terahertz Technology:** Terahertz Physics of Semiconductor Heterostructures.

Further Reading

Grüner G (1998) *Millimeter and Submillimeter Wave Spectroscopy of Solids*, *Vol. 74 of Topics in Applied Physics*. Berlin: Springer-Verlag.

Kaiser W (1988) *Ultrashort Laser Pulses and Applications*, *Vol. 60 of Topics in Applied Physics*. Berlin: Springer-Verlag.

Khurgin JB (1994) Optical rectification and terahertz emission in semiconductors excited above the bandgap. *Journal of the Optical Society of America B* 11: 2492–2501.

Rullière C (1998) *Femtosecond Laser Pulses*. Berlin: Springer-Verlag.

Saldin EL, Schneidmiller EA and Yurkov MV (2000) *The Physics of Free Electron Lasers*. *Advanced Texts in Physics*. Berlin: Springer-Verlag.

Siegel PH (2002) Terahertz technology. *IEEE Transactions on Microwave Theory and Techniques* 50: 910–928.

Tredicucci A, Gmachl C, Capasso F, *et al.* (2001) Novel quantum cascade devices for long wavelength IR emission. *Optics Materials* 17: 211–217.

Tsen KT (2001) *Ultrafast Physical Processes in Semiconductors*, *Vol. 67 of Semiconductors and Semimetals*. New York: Academic Press.

Terahertz Physics of Semiconductor Heterostructures

R Bratschitsch and K Unterrainer, Technische Universität Wien, Vienna, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The terahertz (THz) regime of the electromagnetic spectrum holds promise for a large number of important applications in areas like environmental sensing, quality control, medicine, military applications, astronomy, etc. The development of THz devices has been mainly hindered by the lack of compact sources. In this article we overview the types of tabletop semiconductor sources that are available right now. We focus on the use of pulsed THz sources that are optically excited by femtosecond laser pulses. The generated bursts of THz radiation are extremely useful for time-resolved THz spectroscopy of semiconductor heterostructures.

THz Generation from Semiconductors

The first freely propagating THz pulses using femtosecond lasers were generated in 1984 when Auston *et al.* fabricated a Hertzian dipole antenna. It mainly consists of an electrically biased transmission line structure mounted on a semiconductor, i.e. a photoconductive switch that is triggered by a femtosecond laser pulse (Figure 1). The generated transient photocurrent $J(t)$ emits radiation according to $E(t) \propto \partial J(t)/\partial t$. Since 1989, sapphire half-spheres or refractive silicon lenses have been attached to these dipole antennas, which made it possible to produce collimated THz beams for use in everyday spectroscopic experiments. Today generation and detection of THz radiation by semiconductor photoconductive antennas is widely used. However, most of the antennas are limited to the low-frequency regime of the THz spectrum since they usually have a high-frequency cutoff of about 1–2 THz.

THz generation from semiconductor surfaces was introduced by Zhang in 1990. There are three different physical mechanisms that are responsible for the surface emission of THz radiation. One is the transient current, i.e., the acceleration of photogenerated carriers in the semiconductor surface electric field. This process is accompanied by field induced optical rectification which is often called instantaneous polarization, where the electron–hole pairs are polarized in the dc electric field. Finally, bulk optical rectification as a source of THz radiation can be observed in noncentrosymmetric crystals.

The surface emission intensity can be enhanced by using an externally applied electric field instead of the intrinsic electric field. Until recently, the use of laser amplifiers that optically pump these so-called large-aperture antennas has led to the brightest ultrafast tabletop THz emitters.

Recently, Sarukura and co-workers found that the application of a magnetic field dramatically increases the THz emission from bulk InAs. This led to a re-examination of Zhang's work, in which he reported the magnetic field enhancement in bulk GaAs,

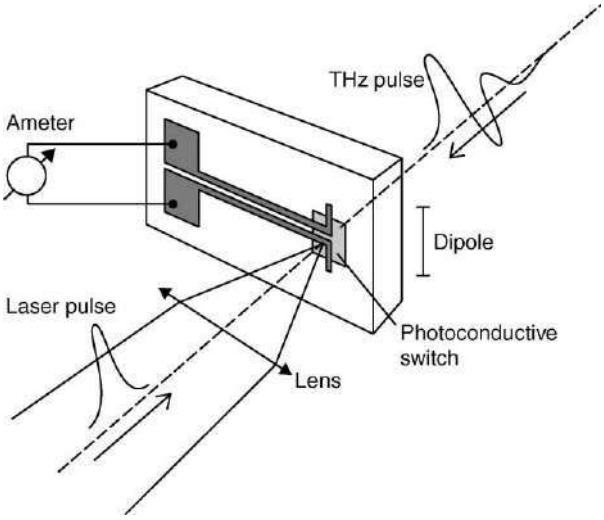

Figure 1 Photoconducting dipole antenna. Reproduced with permission from Gornik E and Kersting R (2001) Coherent THz emission in semiconductors. In: Tsen KT (ed.) *Ultrafast Physical Processes in Semiconductors*, vol. 25, pp. 396. New York: Academic Press.

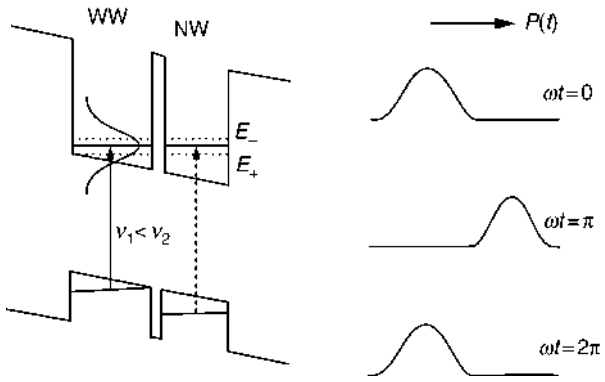

Figure 2 Schematic band diagram of an asymmetric double quantum well structure. A short laser pulse with a spectral width larger than the intersub-band splitting can excite an electronic wavepacket which tunnels back and forth between both wells. Reproduced with permission from Roskos H, Nuss M, Shah J, Leo K, Miller D, Fox A, Schmitt-Rink S and Köhler K (1992) Coherent submillimeter-wave emission from charge oscillations in a double-well potential. *Physical Review Letters* 68: 2216. Copyright (1992) by the American Physical Society.

and to a large number of studies of different types of other semiconductors in various magnetic field geometries. Currently, there is no unified model of the enhancement mechanism(s).

The use of a semiconductor heterostructure by Roskos *et al.* in 1992 (a coupled double quantum well) marked the advent of a new class of semiconductor THz emitters that can be deliberately designed using bandgap engineering. In these devices the quantum nature of electrons in coupled potential wells is used to generate electron wavepackets that tunnel back and forth between different potential wells (Figure 2). The coherent charge oscillation is triggered by a femtosecond laser pulse that has a spectral width that is larger than the separation between the quantized states in the well. In this way a coherent superposition of these levels is created. THz emission from these quantum beats has also been observed from charge oscillations due to light and heavy hole excitons in a quantum well by Planken and co-workers. The concept of quantum beats can even be transferred to the high-THz regime (tens of THz) which was successfully demonstrated by Bonvalet and co-workers. By using an InAs step quantum well they generated THz radiation at a frequency of 30 THz.

Bloch oscillations, one of the fundamental concepts of solid state physics, were predicted in 1934 by Zener. He stated that electrons in a periodic potential under an applied electric field should perform oscillations. Using a quantum mechanical picture, Bloch oscillations can be described as a quantum beat of Wannier–Stark states and in 1993 Waschke *et al.* detected THz emission from these oscillations.

These heterostructure THz sources are inherently frequency tunable since the energy level spacing can be altered by an applied electric field that tilts the quantum wells. The Bloch frequency is given by:

$$\omega_{\text{Bloch}} = \frac{eFd}{\hbar} \qquad [1]$$

where F is the applied electric field and d is the superlattice period. Unfortunately, these devices suffer from their low emission power.

An alternative concept is the use of collective excitations in semiconductor bulk material or heterostructures. Dekorsy and co-workers showed in 1995 that coherent phonons emit THz radiation when excited by femtosecond laser pulses. In the same year Sha and co-workers studied the absorption changes within an electrically biased GaAs p-i-n structure and found distinct oscillations which they ascribed to another collective phenomenon, namely coherent plasma oscillations of the photogenerated carriers (Figure 3). This explanation was supported by the square-root dependence of the observed frequencies on the photogenerated carrier density which is predicted by a simple oscillator model:

$$\omega_{\text{p}} = \sqrt{\frac{ne^2}{\varepsilon\varepsilon_0 m^*}} \qquad [2]$$

where n is the carrier density, ε the dielectric constant, and m^* the effective mass of the carriers. Fischler and co-workers confirmed this finding and in particular investigated the damping behavior of the plasmons. In 1997, Kersting *et al.* detected THz radiation from

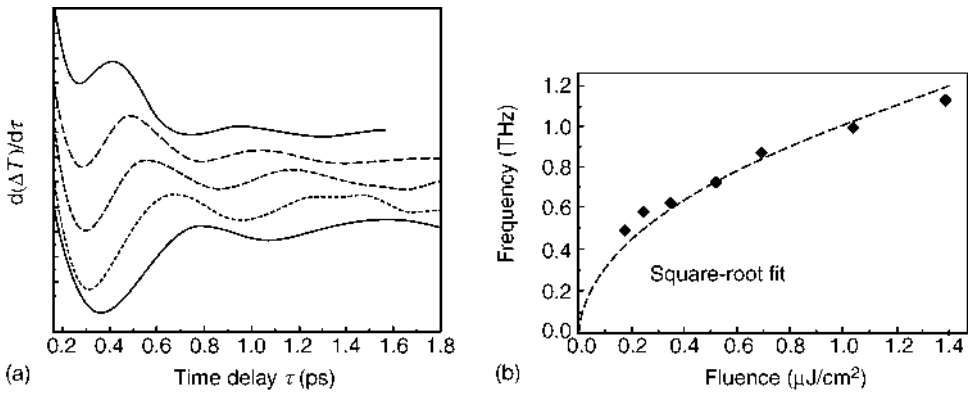

Figure 3 (a) Time-derivative of the charge in electro-absorption versus pump-probe time delay for decreasing fluences (top to bottom respectively); (b) the frequency of the oscillations as a function of fluence. The dashed curve is a square-root function fit. Reproduced with permission from Sha W, Smirl A and Tseng W (1995) Coherent plasma oscillations in bulk semiconductors. *Physical Review Letters* 74: 4273. Copyright (1995) by the American Physical Society.

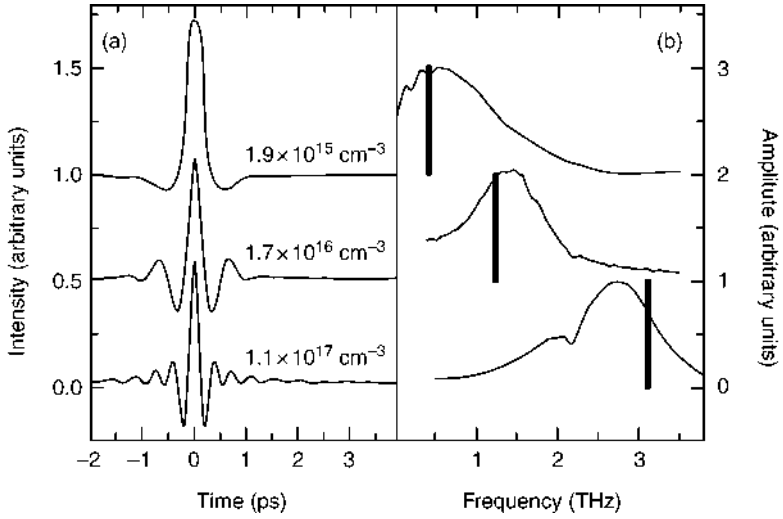

Figure 4 (a) THz autocorrelation traces for three different doping densities; (b) THz spectra; lines denote the frequencies calculated by formula [2]. Reproduced with permission from Kersting R, Heyman J, Strasser G and Unterrainer K (1998) Coherent plasmas in n-doped GaAs. *Physical Review B* 58: 4553. Copyright (1998) by the American Physical Society.

these p-i-n structures and identified heating of the photogenerated carriers as the origin for the strong damping of the plasma oscillations resulting in a low emission intensity.

This problem can be circumvented by using coherent plasma oscillations of extrinsic charge carriers. These electrons are cold compared to the photo-excited carriers and should undergo less scattering. Kersting *et al.* showed that THz pulses are indeed emitted from coherent three-dimensional (3D) plasma oscillations in n-doped GaAs epilayers. The THz pulses are temporally and spatially coherent with intensities up to 100 nW.

Figure 4 shows the THz emission from three GaAs epilayer samples which have been doped with densities 1.9×10^{15}, 1.7×10^{16}, and $1.1 \times 10^{17}\,\mathrm{cm}^{-3}$. The curves are recorded using an autocorrelation technique: 13 nJ, 100 fs laser pulses at a repetition rate of 80 MHz and a central wavelength of 800 nm are split into two pulses with a Michelson interferometer. Every visible pulse generates a THz pulse in the transmission or pseudo-reflection direction. After collimating and focusing of the THz radiation by parabolic mirrors, the interference of the two THz pulses can be recorded by changing the time delay between the visible pulses (Figure 5). Assuming negligible nonlinear interactions between the two excitation processes, the linear autocorrelation function can be recorded. The spectrum is calculated by simply Fourier transforming (Figure 4b). The experiments are performed at room temperature, and the whole setup is purged with dry nitrogen gas to prevent

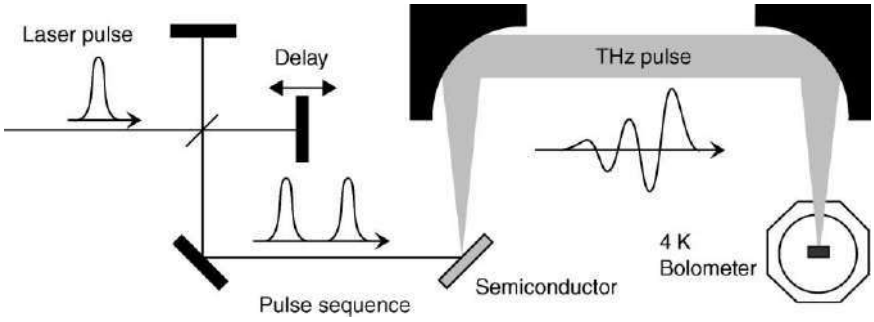

Figure 5 THz autocorrelation setup. Reproduced with permission from Kersting R, Heyman J, Strasser G and Unterrainer K (1998) Coherent plasmons in n-doped GaAs. *Physical Reviews B* 58: 4553. Copyright (1998) by the American Physical Society.

water vapor absorption. Although the autocorrelation traces are recorded at the same optical excitation density of $n_{ex} = 2 \times 10^{16}\ \mathrm{cm}^{-3}$, the emission frequency changes significantly. In addition, the autocorrelation trace of the highly doped structure ($1.1 \times 10^{17}\ \mathrm{cm}^{-3}$) shows multiple oscillations while in the case of the sample with the lowest doping ($1.9 \times 10^{15}\ \mathrm{cm}^{-3}$) the slow oscillation is damped out rather quickly. The frequency behavior of the three different samples can be analyzed from the spectra shown in Figure 4b. The bars indicate the frequency calculated with [2] and setting n equal to the doping density determined by standard Hall measurements. Obviously, the calculated frequencies match very well the experimentally determined ones. This indicates that the THz emission results from a collective oscillation of the extrinsic electrons within the epilayer. This is also supported by the observation that a variation of the excitation density of the photogenerated carriers over two orders of magnitude does not change the THz emission frequency. This is in clear contrast to the experiments with undoped semiconductors where the emission frequency shows a square-root dependence on the optically excited carrier density.

Excitation of the coherent plasma oscillation results from the ultrafast screening of the surface electric field by the femtosecond laser pulse. Before excitation, the doping electrons are confined between the surface electric field and the undoped GaAs substrate (Figure 6). A short laser pulse with an energy above the bandgap of the semiconductor generates electron–hole pairs that screen this surface electric field for a short time. The doping electrons inside the material respond to the change in potential at the surface of the sample leading to the oscillation behavior. A detailed model calculation of the excitation mechanism can be found in Kersting *et al.* (1998). The oscillation frequency – the eigenfrequency of the electron system – is simply the plasma frequency given by the simple formula [2].

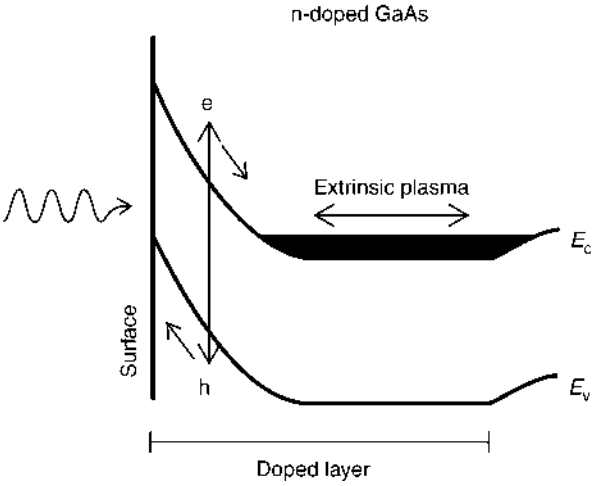

Figure 6 Schematic drawing of the 3D plasmon excitation. After Kersting R, Bratschitsch R, Heyman JN and Unterrainer K (1999) Unpublished.

There are two mechanisms that are responsible for the damping of the plasma oscillations. If the doping density is lower than $10^{16}\ \mathrm{cm}^{-3}$ the coherence time of the 3D plasmon is mainly limited by optical phonon scattering. This can be verified by measuring the plasmon damping times at different temperatures and comparing the results with mobility data which show the same tendency. Since optical phonon scattering is suppressed at low temperatures, the THz emission intensity can be substantially increased by cooling the sample. In the high-doping-density regime $>10^{16}\ \mathrm{cm}^{-3}$ the observed damping times are shorter than those expected due to the optical phonon scattering mechanism alone. This is because in this regime impurity scattering is the dominant scattering mechanism.

One way around this problem is the use of a remotely doped structure. In this case the scatterers are not in the same region as the electrons that perform the plasma oscillation (Figure 7). The higher mobility should lead to a longer decoherence time.

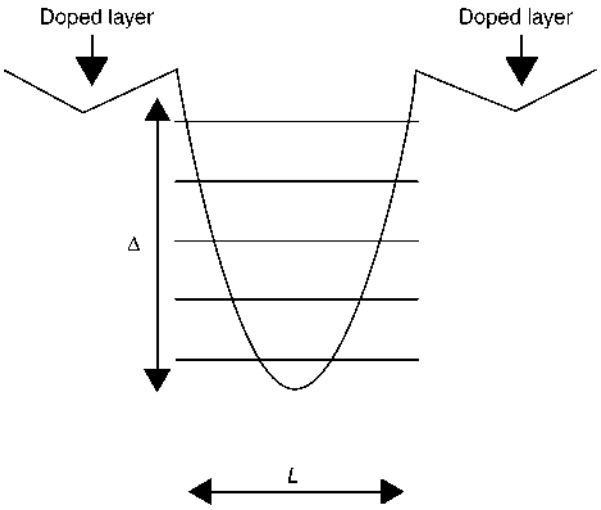

Figure 7 Schematic drawing of a modulation-doped PQW. The doped layers are outside the quantum well.

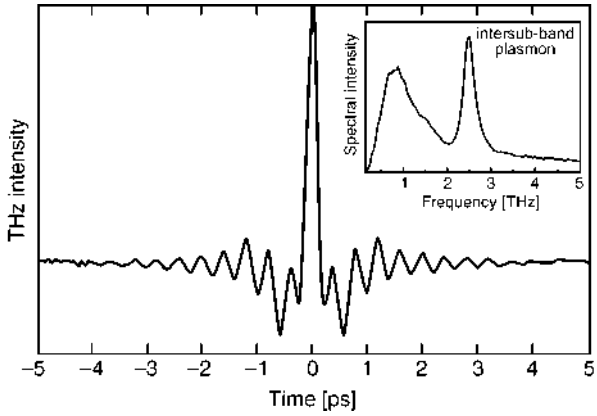

Figure 8 THz autocorrelation signal of a 1400 Å PQW excited by 780 nm laser pulses ($T = 5$ K). Inset: Fourier transform of the recorded autocorrelation. Reproduced with permission from Bratschitsch R, Müller T, Kersting R, Strasser G and Unterrainer K (2000) Coherent THz emission from optically pumped intersubband plasmons in parabolic quantum wells. In: *Ultrafast phenomena XII*. Springer series in Chemical Physics, vol. 66, pp. 204. New York: Springer-Verlag.

For this purpose Bratschitsch *et al.* used modulation doped GaAs/AlGaAs parabolic quantum wells (PQWs), with widths L in the range of 1200–2000 Å and carrier sheet densities n_s of 1.7×10^{11}–5×10^{11} cm^{-2}. The electrons confined within the parabolic potential form a so-called quasi-3D electron system. Figure 8 shows the THz autocorrelation trace of a modulation doped PQW ($L = 1400$ Å, $n_s = 5 \times 10^{11}$ cm^{-2}) excited by 780 nm laser pulses. The density of the optically generated carriers is kept well below the carrier density due to the modulation doping inside the PQW.

The spectrum of the emitted THz radiation (inset of Figure 8) consists of two components: a broad background and a narrow line (FWHM: 0.3 THz) with a center frequency of 2.55 THz. The low-frequency broadband component is found for all the different PQW samples and is independent of the PQW sample structure. It can also be observed at room temperature while the narrowband component completely vanishes. The origin of the broadband component is due to THz generation at the surface of the sample.

The THz emission at 2.55 THz can be observed within a wide range of excitation wavelengths (815–760 nm) above the 819 nm GaAs bandgap ($T = 5$ K). When varying the excitation density over one order of magnitude ($5.6 \times 10^{10} - 6.2 \times 10^{11}$ cm^{-2}), the emission frequency does not change. The narrowband emission results from the oscillation of the carriers inside the PQW. The observed frequency corresponds to the intersub-band plasmon of the PQW. The expected frequency of the intersub-band plasmon is given by

$$\omega_0 = \sqrt{\frac{8\Delta}{L^2 m^*}} \qquad [3]$$

where Δ is the energetic depth of the well, L is the well width, and m^* the effective mass. For the 1400 Å wide PQW the formula yields 2.2 THz for the intersubband plasmon frequency.

The excitation mechanism for the intersub-band plasmon is due to screening of the surface depletion field by the electron–hole pairs injected by the ultrafast laser pulse. In this way the electrons inside the quantum well experience a kick and begin to oscillate with their eigenfrequency (Figure 9). This is supported by the fact that the oscillation can be excited over a large wavelength range of the femtosecond pulses. The onset of intersub-band plasmon emission appears at 815 nm near the 819 nm GaAs bandgap ($T = 5$ K) and can be observed to up to 760 nm. This large range implies that the excitation mechanism is clearly a nonresonant phenomenon, in contrast to the THz quantum beat experiments. Due to ultrafast field screening, a symmetric nanostructure can emit optically driven THz radiation.

The width of the intersub-band plasmon emission line is 0.3 THz (full width at half maximum; FWHM) for the 1400 Å PQW and even 0.2 THz for the 2000 Å PQW sample. The linewidth is a factor of 5 smaller than the smallest linewidth measured for a 3D plasmon emitter. The width of the intersub-band emission is nearly unaffected when increasing the excitation wavelength from near bandgap to approximately 60 nm above bandgap. However, it shows a linear dependence ($5.6 \times 10^{10} - 6.2 \times 10^{11}$ cm^{-2}) on

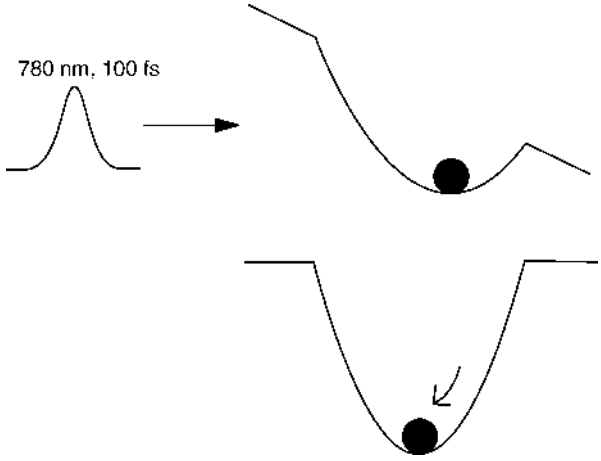

Figure 9 Symbolic drawing of the excitation mechanism: an oscillation of the carriers in the parabolic potential is initiated by ultrafast field screening. Reproduced from Bratschitsch R, Müller T, Kersting R, Strasser G and Unterrainer K (2000) Coherent terahertz emission from optically pumped intersubband plasmons in parabolic quantum wells. *Applied Physics Letters* 76: 3501, with permission from the American Institute of Physics.

the number of optically excited carriers. These optically injected carriers are added to those from the modulation doping and cause a faster dephasing of the intersubband plasmon oscillation. This linear dependence due to carrier–carrier scattering has also been observed in Wolter's experiments dealing with the dephasing of Bloch oscillations in semiconductor superlattices. The linewidth stays constant up to about 50 K (0.33 THz = 1.36 meV) and then starts to broaden until the sample temperature reaches 150 K where the PQW emission is buried by the surface-generated THz emission. The broadening of the line between $T = 50$ K and $T = 150$ K can be explained by the onset of optical phonon scattering in addition to interface roughness scattering which is the dominant scattering mechanism at low temperatures. The onset and linear increase of the line broadening has also been observed with Fourier Transform Infrared (FTIR) absorption and THz time-domain cross-correlation spectroscopy measurements by Kersting and co-workers.

The combination of the designability of the transition frequency, the narrowband emission, and the absence of any processing of the sample make modulation-doped PQWs attractive and easy-to-use THz emitters.

Other THz emission experiments of heterostructures include the optically pumped emission of THz radiation from grating-coupled intrasub-band (2D) plasmons in a doped single quantum well observed by Sekine and co-workers or optically excited THz emission from coherent plasmons in doped superlattices detected by Bratschitsch and co-workers.

Few-Cycle THz Spectroscopy of Semiconductor Quantum Structures

There are a large number of studies that use few-cycle THz radiation for the investigation of semiconductor bulk materials. The complex index of refraction in the THz regime gives insight into the conductivity (i.e., carrier mobility) of the material. The free carriers that change the index of refraction by Drude absorption can be provided either by doping (n-doped silicon or GaAs) or by photo-excitation. An optical pump-THz probe experiment with undoped GaAs reveals the different intervalley scattering times.

There are compelling reasons to use THz radiation to excite and probe carriers in semiconductor quantum structures. The photon energies are comparable to the sub-band spacings, carrier kinetic energies, and phonon energies. In addition, THz radiation does not generate minority carriers so that the experiments directly probe free carriers rather than excitons.

Heyman *et al.* investigated a modulation-doped GaAs/AlGaAs multiple quantum well ($d = 510$ Å, 10×) with a transition energy between the two lowest sub-bands of 1.5 THz and a carrier density of $n_s = 2.75 \times 10^{10}\ \mathrm{cm}^{-2}$. On top of the sample an aluminum Schottky gate is evaporated and the well is contacted with AuGe alloyed ohmic contacts. The wells can be depleted of carriers by applying a negative gate voltage to the Schottky gate (−10 V gate bias). By modulating the voltage between 0 V and −10 V the response of the 2D electron system can be measured. Otherwise, the recorded signal is dominated by the contribution of the bulk GaAs substrate that dramatically chirps the THz pulse due to its frequency-dependent index of refraction. The THz pulses are coupled into the cleaved edge of the quantum well sample so that the electric field is parallel to the growth direction in order to excite the intersub-band transition. The electrons' response is measured with a cross-correlation technique. Here the THz radiation transmitted through the quantum well is mixed with a short THz pulse emitted from low-temperature grown GaAs to get the desired time resolution in the experiment. This allows the detection of the transmitted signal in amplitude and phase. Figure 10 shows the cross-correlation signal of the QW. The signal from the carriers rises during the first 2 ps in response to the THz field and then continues to oscillate at a constant frequency of 1.5 THz. Due to the femtosecond time resolution in this experiment it is possible to observe the phase of the electrons even during the process of excitation. Initially, the incident THz pulse generates a coherent superposition of states in the first and second sub-band. After the THz driving pulse is over

(>2 ps) the carriers continue to radiate at the intersub-band transition frequency, and the signal is damped out by the free induction decay. Since the amplitude and phase of the radiated electric field is recorded, not only the absorption due to the QW electrons but also the phase shift associated with the absorption can be extracted (Figure 11).

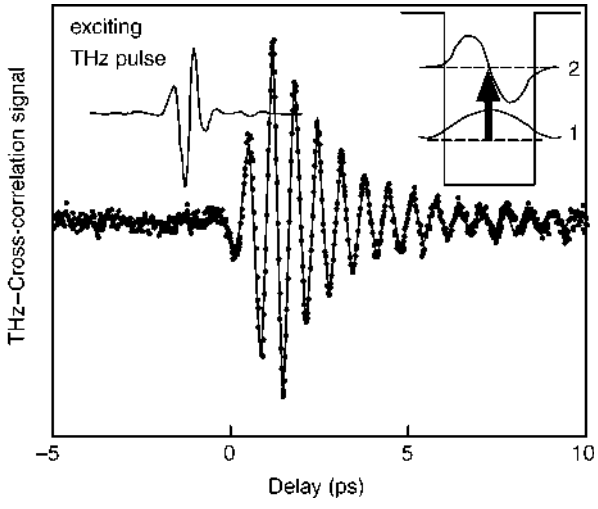

Figure 10 Cross-correlation signal obtained by electrically modulating the charge density in a 510 Å wide quantum well. The THz pulse generates a coherent superposition of the $n = 1$ and $n = 2$ sub-band states. The resulting oscillating polarization radiates and is damped out by the free induction decay. The upper trace shows the incident THz pulse. Reprinted from Unterrainer K, Kersting R, Bratschitsch R, *et al.* (2001) Few-cycle THz spectroscopy of semiconductor quantum structures. *Physica E* 9: 76, with permission from Elsevier.

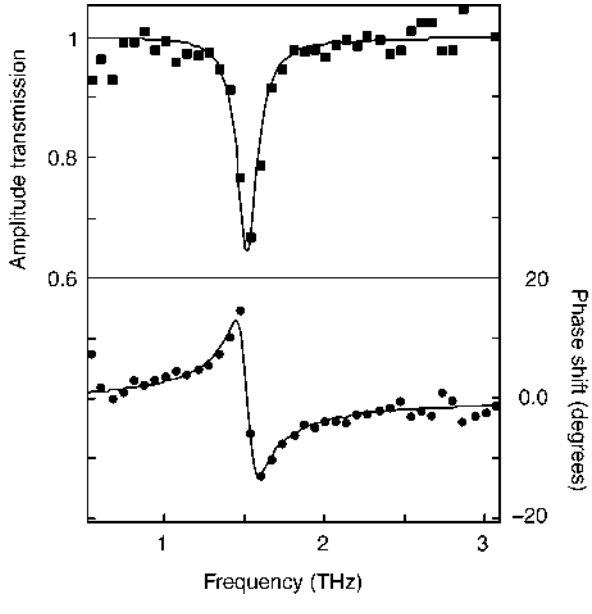

Figure 11 Change in the absorption coefficient and index of refraction in the quantum well due to the quantum well electrons. Solid points are calculated from the time-domain data. Reproduced from Heyman J, Kersting R and Unterrainer K (1998) Time-domain measurement of intersubband oscillations in a quantum well. *Applied Physics Letters* 72: 644, with permission from the American Institute of Physics.

This experiment showed the response of the QW electrons to radiation which is resonant to the intersub-band transition. Using the same experimental technique Kersting *et al.* investigated the response of electrons to a non-resonant excitation of the intersub-band transition. Instead of a square QW, a modulation-doped parabolic quantum well is used in this experiment which has the advantage that the transition frequency is independent of the carrier concentration (Kohn's theorem) and the applied electric field. Furthermore the THz radiation is coupled to the intersub-band transition via a metallic Schottky grating. Figure 12 shows the response of the QW electrons to resonant THz radiation. The electrons follow the driving pulse and continue to radiate at the intersub-band transition frequency. The response to a nonresonant THz pulse is shown in Figure 13. Initially, the electrons inside the QW follow the driving pulse but when the driving pulse has ended a phase jump occurs and the electrons lock to their eigenfrequency and slowly lose their coherence. This behavior can be understood by using time-dependent perturbation theory to solve Schrödinger's

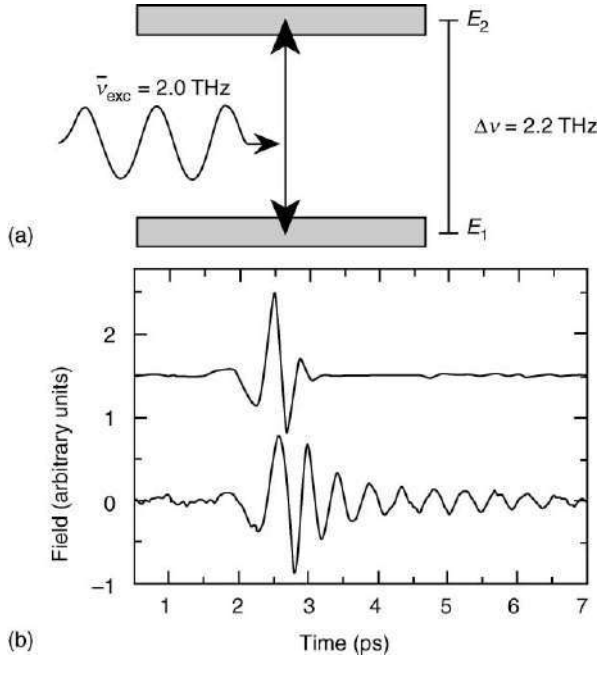

Figure 12 Resonant excitation: (a) band diagram of an intersub-band transition of 2.2 THz and driving THz field at 2.0 THz; (b) exciting THz pulse (upper curve) and cross-correlation signal (lower curve). Reproduced from Kersting R, Bratschitsch R, Strasser G, Unterrainer K and Heyman J (2000) Sampling a THz dipole transition with sub-cycle time-resolution. *Optics Letters* 25: 272, with permission from the Optical Society of America.

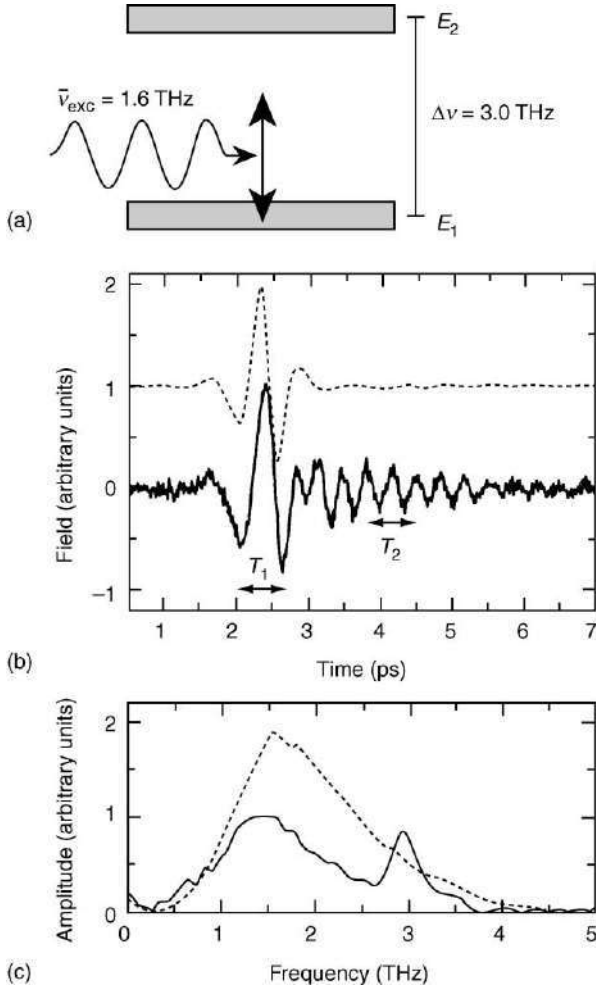

Figure 13 Nonresonant excitation: (a) band diagram of an intersub-band transition of 3.0 THz and driving THz field at 1.6 THz; the oscillation frequency changes from T_1 to T_2; (b) exciting THz pulse (upper curve) and cross-correlation signal (lower curve); (c) spectra of the exciting THz pulse (dotted curve) and cross-correlation signal (solid curve). Reproduced from Kersting R, Bratschitsch R, Strasser G, Unterrainer K and Heyman J (2000) Sampling a THz dipole transition with sub-cycle time-resolution. *Optics Letters* 25: 272, with permission from the Optical Society of America.

equation for a two-level system with an electromagnetic driving field. While the two-level system is treated quantummechanically, the dipole Hamiltonian is used for the incoming THz radiation: $H = H^0 + H'(t)$ and $H'(t) = \mu E(t)$. Making the ansatz $\Psi(t) = a_1(t)\Psi_1 + a_2(t)\Psi_2$ one gets a system of two coupled equations for $a_1(t)$ and $a_2(t)$. They can be decoupled by the condition that the population in the lower level is close to one ($\left\|a_1^2\right\| \approx 1$) since the THz radiation field is not very strong. Figure 14 shows the result of the model calculation that reproduces the change in frequency as well as the observed phase jump after the driving pulse is over.

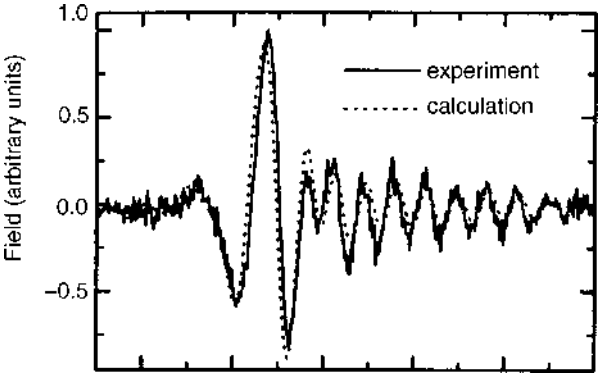

Figure 14 Model calculation of the polarization (dotted curve) and experimental data (solid line). Reproduced from Kersting R, Bratschitsch R, Strasser G, Unterrainer K and Heyman J (2000) Sampling a THz dipole transition with sub-cycle time-resolution. *Optics Letters* 25: 272, with permission from the Optical Society of America.

These experiments are all performed in the linear regime but with nonadiabatic excitation conditions (half-cycle THz pulses). The excellent time resolution together with sufficiently high THz intensities can open up a previously unexplored field of quantum optics beyond the slowly varying envelope approximation.

Conclusions

The recent progress in the development of semiconductor sources together with room temperature THz detection techniques (free-space electro-optic sampling) paves the way for the commercialization of THz technology. In particular, spectrally sensitive THz imaging seems to have great potential in medicine and quality control of products.

See also

Semiconductor Materials: GaAs Based Compounds. **Semiconductor Physics:** Outline of Basic Electronic Properties; Quantum Wells and GaAs-based Structures.

Further Reading

Auston DH, Cheung KP and Smith PR (1984) Picosecond photoconducting Hertzian dipoles. *Applied Physics Letters* 45: 284.

Bonvalet A, Nagle J, Berger V, Migus A, Martin J-L and Joffre M (1996) Femtosecond infrared emission resulting from coherent charge oscillations in quantum wells. *Physical Review Letters* 76: 4392.

Bratschitsch R, Kersting R, Müller T, Strasser G, Unterrainer K, Fischler W and Höpfel RA (1999) Coherent THz plasmons in GaAs/AlGaAs superlattices. *Physica B* 272/1-4: 375.

Bratschitsch R, Müller T, Kersting R, Strasser G and Unterrainer K (2000) Coherent terahertz emission from optically pumped intersubband plasmons in parabolic quantum wells. *Applied Physics Letters* 76: 3501.

Chamberlain M (ed.) (1999) *Terahertz Spectroscopy and Applications*, SPIE Proceedings (Europto Series), vol. 3828, pp. 1–454.

Dekorsy T, Auer H, Waschke C, Bakker H, Roskos H, Kurz H, Wagner V and Grosse P (1995) Emission of submillimeter electromagnetic waves by coherent phonons. *Physical Review Letters* 74: 738.

Gornik E and Kersting R (2001) Coherent THz emission in semiconductors. In: Tsen KT (ed.) *Ultrafast Physical Processes in Semiconductors*, vol. 25, pp. 396. New York: Academic Press.

Heyman J, Kersting R and Unterrainer K (1998) Time-domain measurement of intersubband oscillations in a quantum well. *Applied Physics Letters* 72: 644.

Kersting R, Unterrainer K, Strasser G, Kauffmann H and Gornik E (1997) Few-cycle thz emission from cold plasma oscillations. *Physical Review Letters* 79: 3038.

Kersting R, Heyman J, Strasser G and Unterrainer K (1998) Coherent plasmons in n-doped GaAs. *Physical Reviews B* 58: 4553.

Kersting R, Bratschitsch R, Strasser G, Unterrainer K and Heyman J (2000) Sampling a THz dipole transition with sub-cycle time-resolution. *Optics Letters* 25: 272.

Nun MC and Orenstein J (1997) Terahertz time-domain spectroscopy. In: Gruener G (ed.) *Millimeter Wave Spectroscopy of Solids*. Berlin: Springer.

Roskos H, Nuss M, Shah J, Leo K, Miller D, Fox A, Schmitt-Rink S and Köhler K (1992) Coherent submillimeter-wave emission from charge oscillations in a double-well potential. *Physical Review Letters* 68: 2216.

Sarukura N, Ohtake H, Izumida S and Liu Z (1998) High average power radiation from femtosecond laser-irradiated InAs in a magnetic field and its elliptical polarization characteristics. *Journal of Applied Physics* 84: 654.

Sekine N, Yamanaka K, Hirakawa K, Voßebürger M, Haring-Bolivar P and Kurz H (1999) Observation of terahertz radiation from higher-order two-dimensional plasmon models in GaAs/AlGaAs single quantum wells. *Applied Physics Letters* 74: 1006.

Sha W, Smirl A and Tseng W (1995) Coherent plasma oscillations in bulk semiconductors. *Physical Review Letters* 74: 4273.

Shah J (1999) *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*. Springer Series in Solid-State Sciences, vol. 115. Berlin: Springer.

Waschke C, Roskos H, Schwedler R, Leo K, Kurz H and Köhler K (1993) Coherent submillimeter-wave emission from bloch oscillations in a semiconductor superlattice. *Physical Review Letters* 70: 3319.

Wolter F, Martini R, Tolk S, *et al.* (1999) Influence of carrier–carrier scattering on intraband dephasing. *Superlattice and Microstructure* 26: 93.

Zhang X-C, Hu B, Darrow J and Auston D (1990) Generation of femtosecond electromagnetic pulses from semiconductor surfaces. *Applied Physics Letters* 56: 1011.

TIME-RESOLVED FLUORESCENCE

Contents

Laser Applications
Measurements in Polymer Science

Laser Applications

B Valeur, Conservatoire National des Arts et Métiers, Paris, France

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Fluorescence is extensively used in physical, chemical, material, biological, and medical sciences, for investigating the structure and dynamics of matter or living systems. After excitation of a fluorescent sample by a short pulse of light, the fluorescence decay is, in the simplest case, a single exponential whose time constant is the lifetime of the excited state of the fluorescent species (i.e., the average time during which these species stay in the excited state). However, interactions of a fluorescent species with its micro-environment often perturb the fluorescence decay so that information on this micro-environment can be obtained. More specifically, when intermolecular photophysical processes

(such as electron transfer, proton transfer, energy transfer, excimer, or exciplex formation) occur during the lifetime of the excited state, competition with normal de-excitation results in a fluorescence decay that contains the rate constants for those processes. Consequently, the fluorescence decay is generally a sum of discrete exponentials, or a more complicated function; for example, sometimes the system is characterized by a distribution of decay times. Thus it is of major interest to recover the parameters characterizing the δ-pulse response of a fluorescent sample (i.e., the response to an iznfinitely short pulse of light expressed as the Dirac function δ). With this aim, two time-resolved techniques, pulse fluorometry and phase-modulation fluorometry, are usually employed.

In pulse fluorometry, the sample is excited by a short pulse of light, and the fluorescence response is the δ-pulse response of the sample convoluted by the instrument response. In phase-modulation fluorometry, the sample is excited by modulated light at variable frequency, and the harmonic response of the sample obtained is the Fourier transform of the δ-pulse response. The first technique works in the time domain, and the second technique in the frequency domain. Pulse fluorometry and phase-modulation fluorometry are theoretically equivalent, but the principles of instrumentation are different. In both techniques, lasers are currently used as sources of pulsed or modulated light.

It is the aim of this article to explain the principles of pulse and phase fluorometries using lasers as light sources, and to compare these techniques.

General Principles of Time-Resolved Fluorometry

The principles of time-resolved fluorometry are illustrated in Figure 1. For any excitation function $E(t)$, the fluorescence response $R(t)$ of the sample is the convolution product of this function by the δ-pulse response $I(t)$:

$$R(t) = E(t) \otimes I(t) = \int_{-\infty}^{t} E(t')I(t-t')\,\mathrm{d}t' \qquad [1]$$

In pulse fluorometry, the sample is excited by a short pulse of light and the fluorescence response is recorded as a function of time. If the duration of the pulse is not short with respect to the time constants of the fluorescence decay, the fluorescence response is the convolution product given by eqn [1]: the fluorescence intensity increases, goes through a maximum, and becomes identical to the true δ-pulse response $i(t)$ as soon as the intensity of the light pulse becomes negligible. In this case, data analysis for the determination of the parameters characterizing the δ-pulse response requires deconvolution of the fluorescence response.

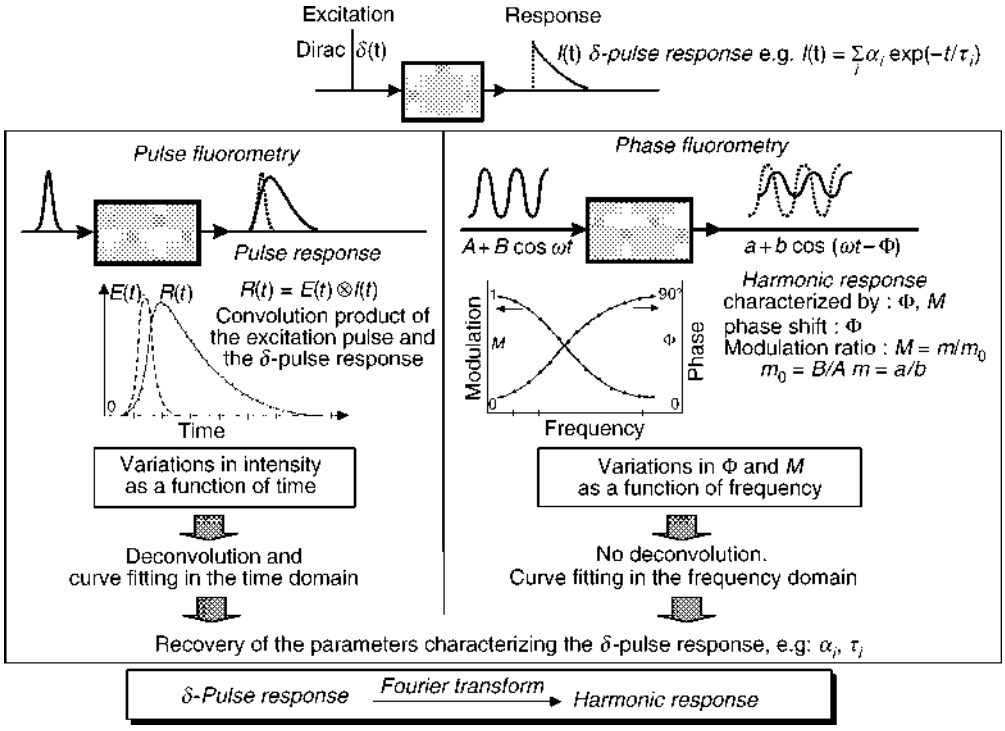

Figure 1 Principles of time-resolved fluorometry.

In phase fluorometry, the sample is excited by a sinusoidally modulated light at high frequency. The fluorescence response, which is the convolution product (eqn [1]) of the δ-pulse response by the excitation function, is sinusoidally modulated at the same frequency, but delayed in phase and partially demodulated with respect to the excitation. The modulation ratio M is defined as the ratio of the modulation depth m (AC/DC ratio) of the fluorescence and the modulation depth of the excitation m_0 (see Figure 1). The phase shift Φ and the modulation ratio M characterize the harmonic response of the system. These parameters are measured as a function of the modulation frequency. No deconvolution is necessary because the data are directly analyzed in the frequency domain.

The relationship between the harmonic response and the δ-pulse response can be obtained as follows. When the excitation function is sinusoidal, i.e., of the following form:

$$E(t) = E_0[1 + m_0 \exp(\mathrm{i}\omega t)] \qquad [2]$$

where ω is the angular frequency $(= 2\pi f)$, the response of the system can be calculated using eqn [1]. It can then be shown that

$$M \exp(-\mathrm{j}\Phi) = \int_0^\infty i(t) \exp(-\mathrm{i}\omega t)\, \mathrm{d}t \qquad [3]$$

where $i(t)$ is the normalized δ-pulse response according to

$$\int_0^\infty i(t)\, \mathrm{d}t = 1 \qquad [4]$$

This important expression shows that the harmonic response expressed as $M \exp(-\mathrm{j}\Phi)$ is the Fourier transform of the δ-pulse response.

It is convenient to introduce the sine and cosine transforms P and Q of the δ-pulse response:

$$P = \int_0^\infty i(t) \sin(\omega t)\, \mathrm{d}t \qquad [5]$$

$$Q = \int_0^\infty i(t) \cos(\omega t)\, \mathrm{d}t \qquad [6]$$

If the δ-pulse response is not normalized according to eqn [4], then eqns [5] and [6] should be replaced by

$$P = \frac{\int_0^\infty I(t) \sin(\omega t)\, \mathrm{d}t}{\int_0^\infty I(t)\, \mathrm{d}t} \qquad [7]$$

$$Q = \frac{\int_0^\infty I(t) \cos(\omega t)\, \mathrm{d}t}{\int_0^\infty I(t)\, \mathrm{d}t} \qquad [8]$$

Since eqn [3] can be rewritten as $M \cos \Phi - iM \times \sin \Phi = Q - iP$, it is easy to show that the phase shift and the modulation ratio are given by

$$\Phi = \tan^{-1}\left(\frac{P}{Q}\right) \qquad [9]$$

$$M = [P^2 + Q^2]^{1/2} \qquad [10]$$

For a single exponential decay, the δ-pulse response is

$$I(t) = \alpha \exp(-t/\tau) \qquad [11]$$

where τ is the decay time and α is the pre-exponential factor or amplitude. The phase shift and relative modulation are related to the decay time by

$$\tan \Phi = \omega\tau \qquad [12]$$

$$M = \frac{1}{(1 + \omega^2\tau^2)^{1/2}} \qquad [13]$$

For a multi-exponential decay with n components, the δ-pulse response is

$$I(t) = \sum_{i=1}^{n} \alpha_i \exp(-t/\tau_i) \qquad [14]$$

Note that the fractional intensity of component i, i.e., the fractional contribution of component i to the total steady-state intensity, is

$$f_i = \frac{\int_0^\infty I_i(t)\, \mathrm{d}t}{\int_0^\infty I(t)\, \mathrm{d}t} = \frac{\alpha_i\tau_i}{\sum_{i=1}^{n} \alpha_i\tau_i} \qquad [15]$$

with, of course, $\sum_{i=1}^{n} f_i = 1$.

Using eqns [7] and [8], the sine and cosine Fourier transforms, P and Q, are given by

$$P = \frac{\omega \sum_{i=1}^{n} \dfrac{\alpha_i\tau_i^2}{1 + \omega^2\tau_i^2}}{\sum_{i=1}^{n} \alpha_i\tau_i} = \omega \sum_{i=1}^{n} \frac{f_i\tau_i}{1 + \omega^2\tau_i^2} \qquad [16]$$

$$Q = \frac{\sum_{i=1}^{n} \dfrac{\alpha_i\tau_i}{1 + \omega^2\tau_i^2}}{\sum_{i=1}^{n} \alpha_i\tau_i} = \sum_{i=1}^{n} \frac{f_i}{1 + \omega^2\tau_i^2} \qquad [17]$$

It should be noted that the optimum frequency for decay time measurements using either the phase shift or the modulation ratio is, according to eqns [12] and [13], such that $\omega\tau$ is close to 1, i.e., $f \approx 1/(2\pi\tau)$. Therefore, for decay times of 10 ps, 1 ns and 100 ns, the optimum frequencies are about 16 GHz, 160 MHz and 1.6 MHz, respectively.

In the case of a single exponential decay, eqns [12] and [13] provide two independent ways of measuring the decay time:

– by phase measurements:

$$\tau_{\Phi} = \frac{1}{\omega}\tan\Phi \qquad [18]$$

– by modulation measurements:

$$\tau_{M} = \frac{1}{\omega}\left(\frac{1}{M^2} - 1\right)^{1/2} \qquad [19]$$

The values measured in these two ways should, of course, be identical and independent of the modulation frequency. This provides two criteria to check whether an instrument is correctly calibrated, by using a lifetime standard whose fluorescence decay is known to be a single exponential.

It should be noted that the measurement of a decay time is fast (a fraction of a second) for a single exponential decay since a single frequency suffices. It is also of interest that a significant difference between values obtained by means of eqns [18] and [19] is compelling evidence for nonexponentiality of the fluorescence decay.

Pulse Fluorometers

The most widely used technique in time domain is the time-correlated single-photon counting technique, preferably called single-photon timing technique (SPT). Time-gated systems are less popular. Streak cameras offer a very good time resolution (a few picoseconds or less) but the dynamic range is smaller than that of the single photon-timing technique. The instruments that provide the best time resolution (a few tens of femtoseconds) are based on fluorescence up-conversion, but they are very expensive and not commercially available. Because of space limitations, only the single-photon timing technique will be presented below.

Principle of the Single-Photon Timing Technique

The basic principle of this technique relies on the fact that the probability to detect a single photon at time t, after an exciting pulse, is proportional to the fluorescence intensity at that time. After timing and recording the single photons following a large number of exciting pulses, the decay of the fluorescence intensity is reconstructed.

Figure 2 shows a schematic diagram of a conventional single-photon counting instrument using a pulsed laser as an excitation source. An electrical pulse associated to the optical pulse is generated (e.g. by a photodiode or the electronics associated to the excitation source) and routed – through a constant-fraction discriminator – to the start input of the time-to-amplitude converter (TAC). Meanwhile, the sample is excited by the optical pulse and emits fluorescence. The optics is tuned (e.g., by means of a neutral density filter) so that the photomultiplier detects no more than one fluorescence photon for each exciting pulse. The corresponding electrical pulse is routed – through a constant-fraction discriminator – to the stop input of the TAC. The latter generates an output pulse whose amplitude is directly proportional to the delay time between the start and the stop pulses. The height analysis of this pulse is achieved by an analog-to-digital converter and a multichannel

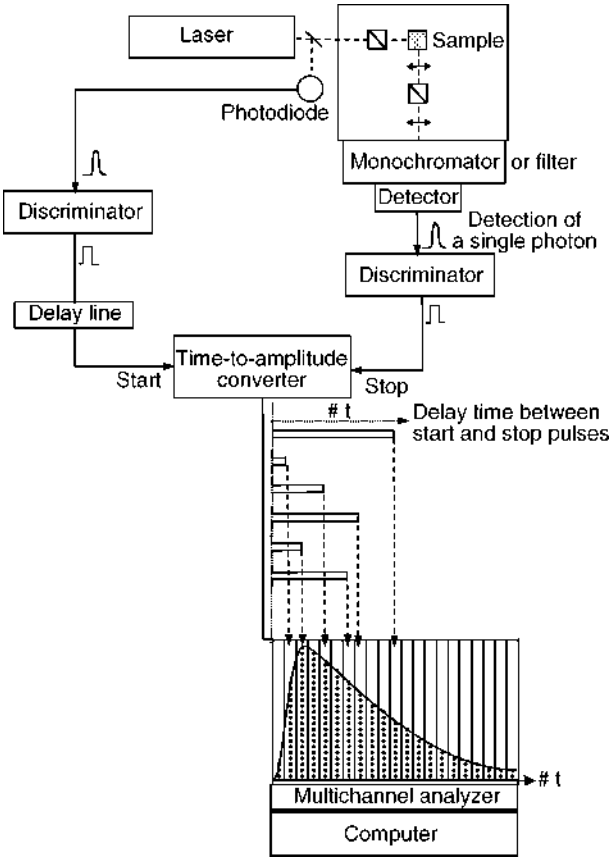

Figure 2 Schematic diagram of a single-photon timing fluorometer.

analyzer (MCA), which increases by one the contents of the memory channel corresponding to the digital value of the pulse. After a large number of excitation and detection events, the histogram of pulse heights represents the fluorescence decay curve.

Obviously, the larger the number of events, the better the accuracy of the decay curve. The required accuracy depends on the complexity of the δ-pulse response of the system; for instance, a high accuracy is, of course, necessary for recovering a distribution of decay times.

When deconvolution is required, the time profile of the exciting pulse is recorded under the same conditions by replacing the sample with a scattering solution (e.g., a suspension of colloidal silica (Ludox)).

It is important to note that the number of detected fluorescence photons must be kept much smaller than the number of exciting pulses (<0.01–0.05 stops per pulse), so that the probability of detecting two fluorescence photons per exciting photon is negligible. Otherwise, the TAC will take into account only the first fluorescence pulse and the counting statistics will be distorted: the decay will appear shorter than it is in reality. This effect is called the 'pile-up effect'.

Laser Sources for the Single-Photon Timing Technique

Mode-locked dye lasers and titanium-doped sapphire lasers are well suited to the single-photon timing technique: in fact, they can generate pulses over broad wavelength ranges, and the pulse widths are in the femtosecond and picosecond ranges with a high repetition rate. Flash lamps running in air, or filled with gas, can also be used, but the range of excitation wavelengths is limited; they deliver nanosecond pulses, and the repetition rate is not high (10^4–10^5 Hz). Dye lasers require the handling of large volumes of dye solutions which must be replaced when the dye is bleached. In contrast, the Ti:sapphire laser is a solid-state laser that does not have this drawback and delivers shorter light pulses. Moreover, long operational lifetime, operational simplicity, reliability, and broad tunability are distinct advantages. For these reasons, Ti:sapphire lasers have become more popular, and are described below.

Ti:sapphire lasers

Titanium-doped sapphire crystals are obtained by replacing a small fraction of the Al atoms of the host (Al_2O_3) with Ti atoms. When the concentration of Ti is low, the optical quality and the mechanical properties of Ti:sapphire crystals are comparable to those of ruby crystals in ruby lasers. The absorption spectrum of Ti:sapphire crystals exhibits a maximum at 490 nm and a shoulder at 540 nm. Therefore, pumping can be achieved by an Argon ion laser or a green solid-state laser. The emission spectrum ranges from 600 to 1,100 nm. Such a broad emission spectrum results from electronic coupling between the electronic levels of the 3d electrons with the lattice vibrations. Since excited-state absorption is not possible in the spectral region of the pump and the fluorescence spectrum, the tuning range is broad: it extends from 680 to 1,100 nm, i.e., almost over the entire fluorescence spectrum.

The width of the pulses that can be generated by Ti:sapphire lasers ranges from a few tens of femtoseconds to about 100 picoseconds. Commercially available Ti:sapphire lasers can operate either in the femtosecond configuration or in the picosecond configuration. In time-resolved fluorescence experiments using the single-photon timing technique, the instrument response is limited by the time response of the photomultiplier; the shortest width of this response that can be measured with the faster detectors (i.e., microchannel plate photomultipliers) is about 30–40 ps (see below). Therefore, there is no reason to operate the Ti:sapphire laser in the femtosecond configuration. In the picosecond configuration, the pulse width is <2–100 ps.

Figure 3 shows the six-mirror folded cavity design for the picosecond configuration in a Tsunami

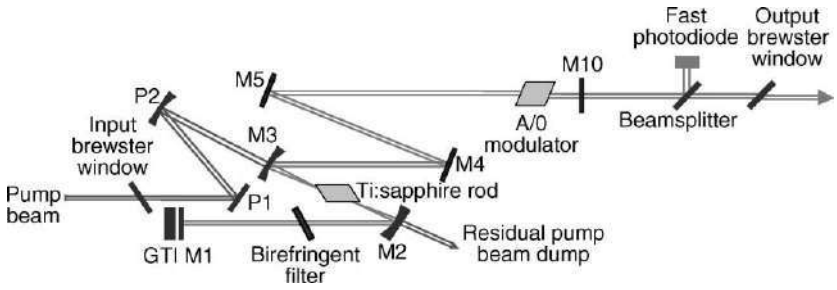

Figure 3 Beam path of the laser cavity of the Tsunami Ti:sapphire laser (Spectra Physics) operating in the picosecond configuration (Adapted from the Tsunami brochure with permission of Spectra Physics).

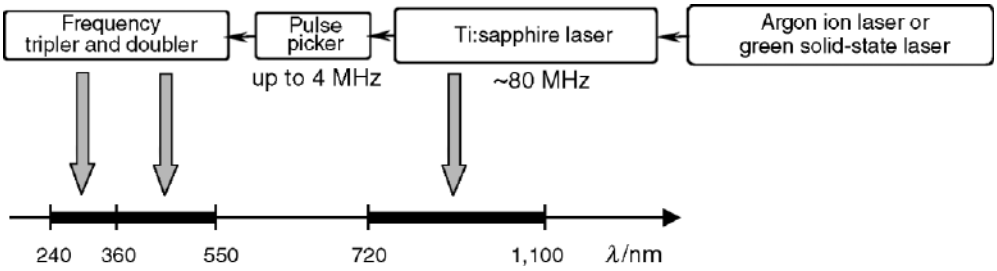

Figure 4 Schematics of the excitation system in time-resolved fluorescence experiments.

Ti:sapphire laser (Spectra Physics). A Brewster's angle acousto-optic modulator, driven by a low-power rf source, insures an 80 MHz repetition rate. Consequently, in contrast to standard passive mode-locking systems, the laser can operate for long periods without dropouts or shut-downs. Gires–Tournois Interferometer (GTI) mirror spacing is used for tuning the laser cavity. Wavelength tuning is achieved by means of a two-plate birefringent filter placed within the laser cavity at Brewster's angle. The tuning range depends on the mirror sets that are used. With broadband cavity mirrors, the entire tuning range of Ti:sapphire is covered, i.e., from below 700 nm to greater than 1,000 nm, and changes in wavelength can be made in a matter of minutes.

When pumping with a 10 W argon ion laser, the average output power is about 1.5 W at 800 nm and the pulse energy is about 15 nJ. Argon ion lasers can now be advantageously replaced by cw green solid-state lasers (532 nm) in which the gain medium ($Nd:YVO_4$) is pumped by diodes, and a LBO crystal placed inside the cavity ensures frequency doubling. These lasers offer several advantages: the head is compact and does not require external cooling water; it is insensitive to temperature, and thermal equilibration is very rapid.

Since the excited-state lifetime of fluorescent compounds can be as long as several hundreds of nanoseconds, the repetition rate of 80 MHz is too high to let the fluorescence of long lifetime samples vanish before a new exciting pulse is generated. Therefore, it must be limited to a few MHz, thanks to a pulse picker using a Bragg-angle acousto-optic modulator synchronized to the mode-locked 80 MHz pulse train. Selection of pulses from the train can thus be achieved at a maximum pulse selection rate of 8 MHz.

For excitation of fluorescent samples, wavelengths in the visible and UV regions are desirable. Frequency doublers and triplers using LBO and BBO crystals, respectively, have been designed with this aim. The possible tuning ranges are shown in Figure 4. When pumping with a 10 W argon ion laser, the average output powers are about 300 mW and 120 mW for frequency doubling and tripling, respectively.

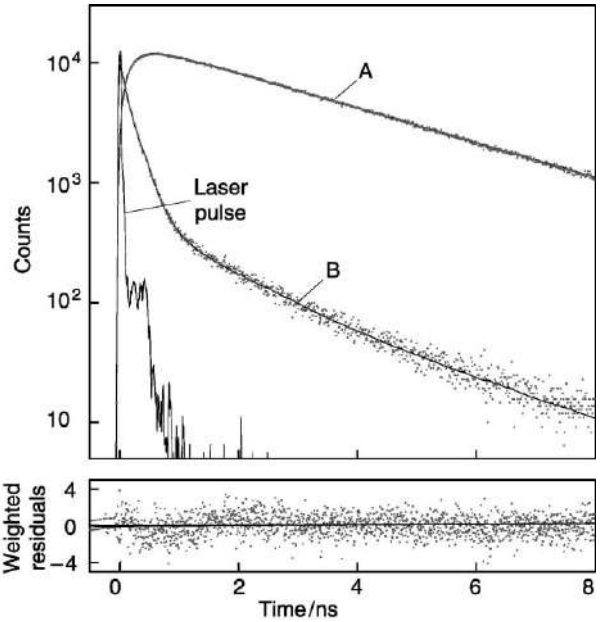

Figure 5 Example of data obtained by the single-photon timing technique using a Spectra-Physics setup composed of a titanium-sapphire Tsunami laser pumped by an argon ion laser, a pulse selector, and doubling (LBO) and tripling (BBO) crystals. The detector is a Hamamatsu MCP photomultiplier R3809U. The sample is a solution of DANS1 (a derivative of dansylamide) in a mixture CH_3CN/H_2O (60:40 v/v) at pH 0.8. Excitation wavelength: 286 nm. Channel width: 9.8 ps. A: Observation through a bandpass filter (300–370 nm) of the protonated form that undergoes photoinduced deprotonation. B: Observation through a cut-off filter ($\lambda > 400$ nm) of the neutral form; the observed risetime corresponds to the formation of the neutral form from the protonated form. Global analysis of the two curves leads to two time constants: 183 ± 9 ps and 3.01 ± 0.03 ns (Global $\chi^2_R = 1.26$) (Métivier R, Leray I and Valeur B, unpublished results. Métivier R, doctoral thesis, Ecole Normale Supérieure de Cachan, 2003).

With Ti-sapphire lasers and microchannel plate photomultipliers, decay times as short as 10 ps can be measured.

The results of an experiment are shown in Figure 5.

Pulsed diode lasers

Picosecond diode laser heads offer an interesting alternative to mode-locked lasers and are much less expensive. They can produce light pulses as short as 50–90 ps with repetition rates from single shot to 40–80 MHz. Peak powers up to 500 mW can be obtained. However, the main disadvantage is the absence of tunability, and the number of wavelengths is limited: 375, 400, 440, 635 to 1,550 nm. No wavelength below 375 nm is presently available.

Phase-Modulation Fluorometers

In the frequency domain, the phase and modulation measurements can be taken by using either a cw laser and an optical modulator, or the harmonic content of a pulsed laser.

Phase Fluorometers Using a cw Laser and an Optical Modulator

The optical modulator is generally a Pockels cell. Low-power cw lasers are sufficient and not as expensive as pulsed lasers, so that one can afford several lasers of this type in order to get various excitation wavelengths. Available wavelengths from cw lasers are given in Table 1. A Xenon lamp can also be used; however, the optical modulator works better with a cw laser.

Figure 6 shows the schematic diagram of a multifrequency phase-modulation fluorometer using a cw laser and a Pockels cell. A beamsplitter reflects a few percent of the incident light towards a reference photomultiplier (via, or not, a cuvette containing a reference scattering solution). The fluorescent sample and a reference solution (containing either a scatter or a reference fluorescent compound) are placed in a rotating turret. The emitted fluorescence or scattered light is detected by a photomultiplier through a monochromator or an optical filter. The Pockels cell is driven by a frequency synthesizer and the photomultiplier response is modulated by varying the voltage at the second dynode by means of another frequency synthesizer locked in phase with the first one. The two synthesizers provide modulated signals that differ in frequency by a few tens of Hz in order to achieve cross-correlation (heterodyne detection). This procedure offers excellent accuracy because the phase and modulation information contained in the signal is transposed to the low-frequency domain, where phase shifts and modulation depths can be measured with a much better accuracy than in the high-frequency domain.

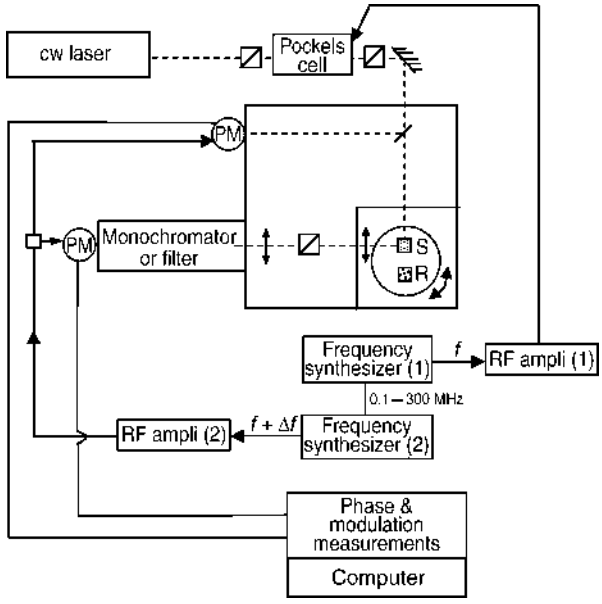

Figure 6 Schematic diagram of a multifrequency phase-modulation fluorometer using a cw laser and a Pockels cell (S: sample; R: reference; PM: photomultiplier).

Practically, the phase delay ϕ_R and the modulation ratio m_R of the light emitted by the scattering solution are measured with respect to the signal detected by the reference photomultiplier. Then, after rotation of the turret, the phase delay ϕ_F and the modulation ratio m_F for the sample fluorescence are measured with respect to the signal detected by the reference photomultiplier. The absolute phase shift and modulation ratio of the sample are then $\Phi = \phi_F - \phi_R$ and $M = m_F/m_R$, respectively.

Figure 7 shows an example of an experiment.

Table 1 Examples of available wavelengths from low-power cw lasers

Laser	*Wavelength (nm)*
He–Cd	325; 442.5
He–Ne	543.3; 594; 611.9; 632.8
Ar^{+a}	457.9; 476.5; 488.0; 514.5

[a]High-power argon ion lasers can also provide wavelengths of 351.1 and 363.8 nm.

Phase Fluorometers Using the Harmonic Content of a Pulsed Laser

The type of laser source that can be used is exactly the same as that for single-photon timing pulse fluorometry (see above). Such a laser system that delivers pulses in the picosecond range, with a repetition rate of a few MHz, can be considered as an intrinsically modulated source. The harmonic content of the pulse train – which depends on the

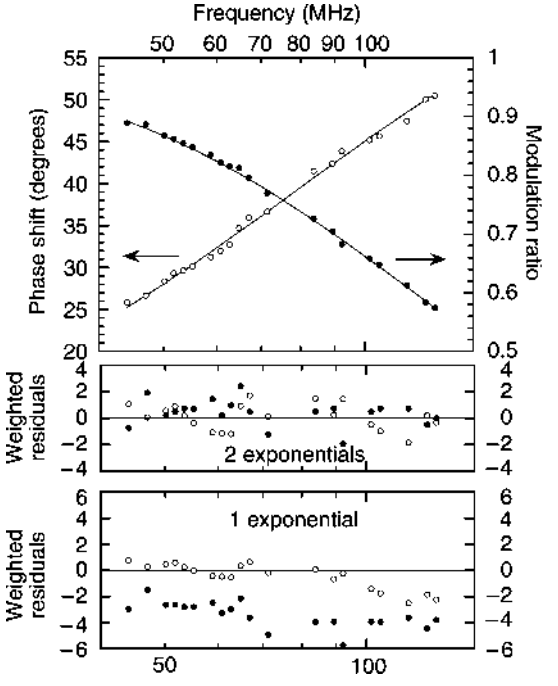

Figure 7 Example of data obtained by the phase and modulation technique using a fluorometer is equipped with a cw Omnichrome He–Cd laser (442 nm) and a Pockels cell operating at frequencies ranging from 0.1 to 200 MHz. The sample is a solution of a calcium complex of a crowned merocyanine (DCM-crown) in acetonitrile (observation through a Balzers bandpass filter centered at 581 nm). The solid line corresponds to the best fit with a biexponential decay. Weighted residuals are shown for the best fit with one and two exponentials. The time constants are 1.87 ± 0.04 and 0.35 ± 0.09 ns with fractional intensities of 0.85 and 0.15, respectively. The value of χ_r^2 is 1.01. (Redrawn from Martin MM, Plaza P, Meyer YH, *et al.* (1996) Steady-state and picosecond spectroscopy of Li^+ or Ca^{2+} complexes with a crowned merocyanine. Reversible photorelease of cations. *Journal of Physical Chemistry* 100: 6879–6888, with permission from the American Chemical Society).

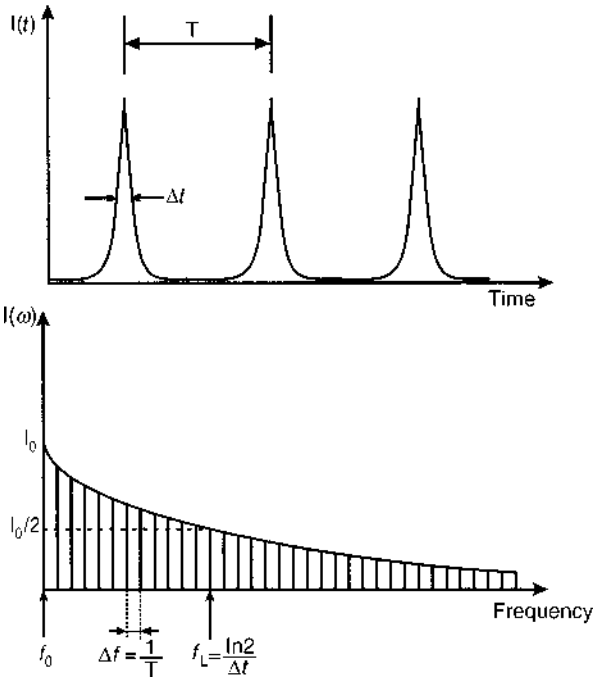

Figure 8 Harmonic content of a pulse train. Example of numerical values: $\Delta f = 4$ MHz; $\Delta t = 4$ Ps; $f_L = 44$ GHz.

pulse width (as illustrated in Figure 8) – extends to several gigahertz.

For high-frequency measurements, normal photomultipliers are too slow, and microchannel plate photomultipliers are required. However, internal cross-correlation is not possible with the latter and an external mixing circuit must be used.

The time resolution of a phase fluorometer, using the harmonic content of a pulsed laser and a microchannel plate photomultiplier, is comparable to that of a single-photon timing instrument using the same kind of laser and detector.

Data Analysis

In both pulse and phase fluorometries, the most widely used method of data analysis is based on a nonlinear least-squares method. The basic principle of this method is to minimize a quantity which expresses the mismatch between data and fitted function. This quantity is the reduced chi-square χ_r^2, defined as the weighted sum of the squares of the deviations of the experimental response $R(t_i)$ from the calculated responses $R_c(t_i)$:

$$\chi_r^2 = \frac{1}{\nu}\sum_{i=1}^{N}\left[\frac{R(t_i) - R_c(t_i)}{\sigma(i)}\right]^2 \qquad [20]$$

where N is the total number of data points and $\sigma(i)$ is the standard deviation of the ith data point, i.e., the uncertainty expected from statistical considerations (noise). ν is the number of degrees of freedom ($\nu = N - p$, where p is the number of fitted parameters). The value of χ_r^2 should be close to 1 for a good fit.

In the single-photon timing technique, the statistics obey the Poisson distribution and the expected deviation $\sigma(i)$ is approximated to $[R(t_i)]^{1/2}$ so that eqn [20] becomes

$$\chi_r^2 = \frac{1}{\nu}\sum_{i=1}^{N}\frac{[R(t_i) - R_c(t_i)]^2}{R(t_i)} \qquad [21]$$

In phase fluorometry, no deconvolution is required: curve fitting is performed in the frequency domain, directly using the variations of the phase shift Φ and the modulation ratio M as functions of the modulation frequency. Phase data and

modulation data can be analyzed separately, or simultaneously. In the latter case the reduced chi-squared is given by

$$\chi_r^2 = \frac{1}{\nu}\left[\sum_{i=1}^{N}\left[\frac{\Phi(\omega_i) - \Phi_c(\omega_i)}{\sigma_\Phi(\omega_i)}\right]^2 + \sum_{i=1}^{N}\left[\frac{M(\omega_i) - M_c(\omega_i)}{\sigma_M(\omega_i)}\right]^2\right] \quad [22]$$

where N is the total number of frequencies. In this case, the number of data points is twice the number of frequencies, so that the number of degrees of freedom is $\nu = 2N - p$.

In addition to the value of χ_r^2, it is useful to display graphical tests. The most important of them is the plot of the weighted residuals defined as

$$W(t_i) = \frac{R(t_i) - R_c(t_i)}{\sigma(i)} \quad [23]$$

where $\sigma(i) = [R(t_i)]^{1/2}$ for single-photon counting data. The fit is satisfactory when the weighted residuals are randomly distributed around zero.

Concluding Remarks

From the instrumental point of view, the latest generations of pulse and phase fluorometers both use pulsed lasers (usually Ti:sapphire lasers) and microchannel plate detectors; only the electronics is different. Since the instrument response is mainly limited by the time response of the detector, the time resolution is the same for both techniques. Moreover, the optical module is identical so that the total cost of instrumentation is similar.

There are, however, significant differences between pulse and phase fluorometers from the methodological point of view. In pulse fluorometry, the single-photon timing technique offers various advantages: direct visualization of the fluorescence decay; high sensitivity; outstanding dynamic range and linearity; and well-defined statistics (Poisson distribution) allowing the accurate weighing of each point in data analysis. In phase-modulation fluorometry, the main advantage is that no deconvolution is necessary, while this operation is often necessary in pulse fluorometry and requires great care in recording the instrument response, especially for very short decay times. Moreover, the short acquisition time for phase shift and modulation ratio measurements at a given frequency is a distinct advantage.

See also

Ultrafast Laser Techniques: Pulse Characterization Techniques.

Further Reading

Alcala JR, Gratton E and Jameson DM (1985) A multifrequency phase fluorometer using the harmonic content of a mode-locked laser. *Analytical Instrumentation* 14: 225–250.

Demas JN (1983) *Excited-State Lifetime Measurements*. New York: Academic Press.

Jameson DM, Gratton E and Hall RD (1984) The measurement and analysis of heterogeneous emissions by multifrequency phase and modulation fluorometry. *Applied Spectroscopy Reviews* 20: 55–106.

Lakowicz JR (ed.) (1991) *Topics in Fluorescence Spectroscopy, Vol. 1: Techniques*. New York: Plenum Press.

Lakowicz JR, Laczko G and Gryczynski I (1986) 2 GHz frequency-domain fluorometer. *Review of Scientific Instruments* 57: 2499–2506.

O'Connor DV and Phillips D (1984) *Time-Correlated Single Photon Counting*. London: Academic Press.

Valeur B (2002) *Molecular Fluorescence. Principles and Applications*, chap. 6. Weinheim, Germany: Wiley-VCH.

Measurements in Polymer Science

T A Smith, University of Melbourne, Melbourne, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In this article we explore the use of ultrashort pulsed lasers in time-resolved fluorescence measurements for the study of the dynamics of excimer formation and energy transfer/migration in synthetic aromatic polymers. The temporal profile of monomer and excimer emission, and the use of time-resolved fluorescence polarization approaches are discussed.

The use of lasers in chemistry has most often exploited the high degree of spectral purity achievable with these devices, in both photochemical and

high-resolution spectroscopic applications. Through the use of lasers, we also have the ability to generate pulses of light of ultrashort duration (a few femtoseconds) and of incredibly high peak powers (petawatts), characteristics significantly beyond those achievable with conventional (incoherent) light sources. This enables the kinetics of the many photophysical and photochemical processes that occur on ultrafast time-scales with high efficiency within polymers, to be determined through various time-resolved measurements.

These processes are dealt with in detail in specialist photochemistry textbooks and so knowledge of the terms associated with photochemistry and photophysics is assumed (e.g., electronic configurations, orbitals, spin, multiplicity, etc.).

Following the absorption of light, an isolated chromophore will be in one of its excited electronic and vibrational states. Once formed, the excited state of a molecule will only survive for a limited time, with the molecule losing its excess energy to return to a preferred lower energy state (electronic configuration). It can do this through a number of spontaneous, radiative, and nonradiative deactivation processes, each with its own characteristic rate constant k_{ij}. Nonradiative pathways include internal conversion (IC), such as intramolecular vibrational redistribution, and vibrational coupling to the bath, and inter-system crossing (ISC) between singlet and triplet manifolds. Radiative pathways involve the emission of a photon of light, in the form of fluorescence or phosphorescence.

The time the molecule spends in the excited state is determined by the rates of all the pathways that contribute to the overall deactivation of that state. The lifetime, τ, of a singlet state is therefore defined as:

$$\tau = \frac{1}{k_{\mathrm{ISC}} + k_{\mathrm{IC}} + k_{\mathrm{Rad}} + \cdots} = \frac{1}{\sum k_{ij}} \qquad [1]$$

In systems of higher chromophore concentration, such as aromatic-containing polymers, additional, bimolecular pathways are available through which the excess excited state energy may be dissipated, making it rare that a single fluorescence 'lifetime' is observed from polymer systems. The observation of complex (multi- or nonexponential) fluorescence decay profiles sometimes invokes the use of the terms 'average lifetimes' or 'lifetime distributions'.

The complex nonexponential fluorescence decay behavior observed, particularly from polymer systems, has been variously attributed to one or more of the following factors: molecular aggregation (e.g., excimer/exciplex equilibria); acid–base equilibria (e.g., proton transfer); one-, two-, or three-dimensional energy transfer and energy migration (and variations in the singlet energy migration length); electron transfer; molecular motion (translational diffusion, rotational relaxation); vibrational or solvent relaxation; concentration quenching; fluorescent impurities, etc. The kinetics of excimer formation are further complicated by heterogeneity of kinetically distinct monomer and/or excimer species, configurational and conformational influences upon excimer formation, and the time dependence of energy trapping (excimer formation) through nonequilibrium diffusion-controlled processes. Here, we will concentrate on excimer formation and energy migration in synthetic aromatic polymers, as shown schematically in Figure 1.

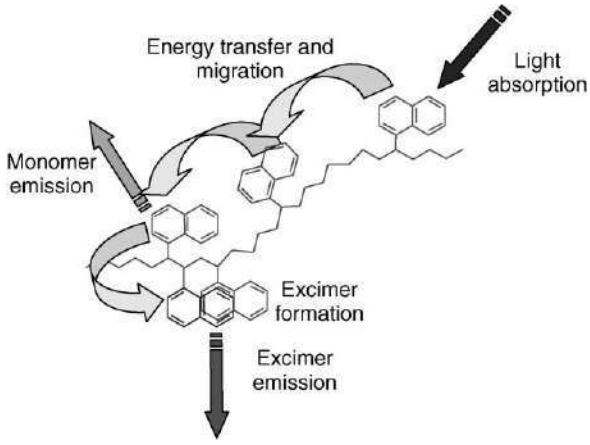

Figure 1 Schematic representation of some of the photophysical processes in polymers.

Ultrafast Laser Spectroscopic Techniques

A wide range of experimental techniques and instrumentation is available for the study of time-resolved processes in polymers. The most widely used methods rely on the detection of emission from the polymer. Fluorescence-based techniques are popular for investigating time-dependent processes in polymers, since the time-scales involved in many of the photochemical and photophysical processes discussed correspond with the time-scales on which fluorescence occurs (normally on the subpicosecond to nanosecond time-scales). Phosphorescence usually occurs on significantly longer time-scales and therefore is less likely to be competitive with fluorescence or other radiative or nonradiative deactivation pathways under normal conditions. Fluorescence-based techniques are attractive for reporting on a number of time-dependent processes since fluorescence is inherently multidimensional, it can be readily resolved temporally, and fluorescence detection can

be extremely sensitive, achieving single photon and single molecule detection levels.

Amongst the most commonly used fluorescence-based techniques are: time-correlated single photon counting (TCSPC, or single photon timing), fluorescence transient digitization, and ultrafast streak cameras. Methods such as transient absorption (flash photolysis) can also provide some details regarding the identity of short-lived species formed following photo-excitation. Other ultrafast laser techniques have been applied to the study of time-dependent processes in polymers in order to achieve higher temporal resolution. So-called pump–probe methods provide the ultimate in time-resolution and can be combined with fluorescence detection through the use of Kerr shutters or ultrafast fluorescence upconversion techniques. Other, more complex pump–probe methods, such as transient grating and multiple pulse photon echo measurements, are potentially very powerful tools in the study of time-resolved processes in polymers.

The generation of short pulses of light can be achieved through numerous mechanisms, including active or passive mode-locking and Q-switching techniques, and superradiance processes (e.g., in nitrogen lasers). Lasers capable of producing short (~ns) pulses of light by Q-switching are usually characterized by high energies per pulse (from μJ up to many hundreds of mJ/pulse) at low (~Hz–kHz) pulse repetition rates. Lasers producing pulses of ultrashort duration (often in the picosecond or femtosecond region) through mode-locking techniques, characteristically operate at high pulse repetition rates (~MHz) and with lower energies per pulse (~nJ/pulse). It should be noted that even pulses of a few nJ energy delivered in ~100 fs, still have peak powers of tens of kW. These high peak powers enable the generation of a range of excitation wavelengths suitable for the study of polymer systems through nonlinear optical wavelength conversion processes, such as second-harmonic generation (SHG) and optical parametric oscillation (OPO) devices.

Excimer Formation

Excimers (or *exci*ted-state di*mers*) are formed when an aromatic molecule, while in its excited electronic state, $^1M^*$, interacts with a second molecule of the same chemical type in its ground electronic state, ^1M. Excimer formation is often observed in concentrated molecular systems such as concentrated solutions and molecular crystals; however, it can also occur in dilute solutions of polymers containing aromatic chromophores, or in the confined environment of solid polymer systems such as homopolymer films.

The structure of excimers in solution is generally thought to be a face-to-face, coplanar, sandwiched, or in some cases, partially eclipsed arrangement of the two constituent aromatic moieties, which attempt to adopt the lowest possible energy configuration, with a separation of 3–4 Å, as depicted in Figure 2. A range of possible excimer configurations may exist rather than just the two extremes.

A feature of excimers is that they have a dissociative ground state (Figure 3), i.e., they do not exist prior to the absorption of light by one monomer chromophore and thus cannot be formed directly without the coming together of the two constituent chromophores within the lifetime of the $^1M^*$. Excimers, being of lower energy than the corresponding excited monomer species and with this dissociative ground state, usually exhibit characteristically

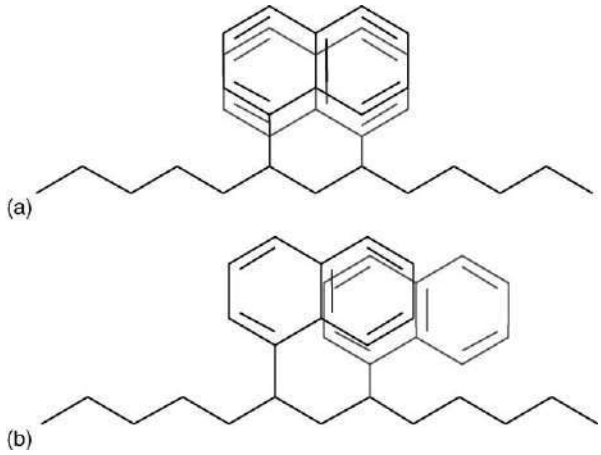

Figure 2 (a) Sandwiched and (b) partially eclipsed excimer configurations.

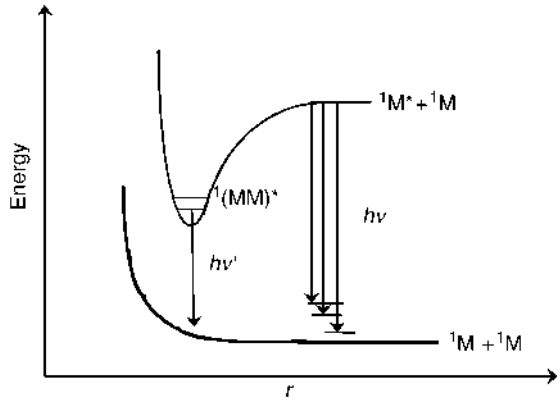

Figure 3 Schematic potential energy diagram for excimer formation and decay. *r* represents the monomer–monomer separation.

broad, structureless emission to the red of the monomer emission (Figure 4). Upon the emission of fluorescence from an excimer, the complex loses its excess energy and dissociates back to two ground state monomer units (i.e., 2 ^{1}M). Excimers may also dissociate back to a noninteracting pair of ^{1}M* and ^{1}M.

Excimer formation of simple aromatic molecules in solution is often modeled on the kinetic scheme proposed by Birks and co-workers (Figure 5). From Figure 5, if a δ-pulse excitation, represented by $h\nu_{abs}$, is used to produce an initial concentration of excited singlet monomer molecules, $[^1M^*]_o$, then such a scheme results in the following rate equations:

$$\frac{d[^1M^*]}{dt} = k_{MD}[^1(MM)^*] - X[^1M^*] \qquad [2]$$

$$\frac{d[^1(MM)^*]}{dt} = k_{DM}[^1M][^1M^*] - Y[^1(MM)^*] \qquad [3]$$

where $X = k_{FM} + k_{NM} + k_{DM}[^1M^*] = k_M + k_{DM}[^1M]$, $Y = k_{FD} + k_{ND} + k_{MD} = k_D + k_{MD}$.

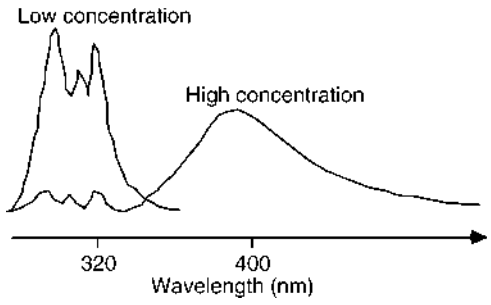

Figure 4 Schematic representation of monomer and excimer emission.

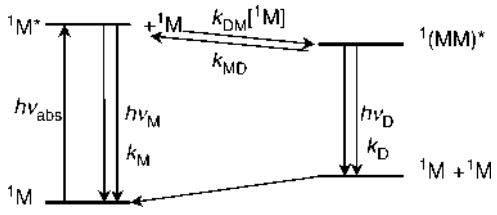

Figure 5 Kinetic scheme for excimer formation where $h\nu_M$ and $h\nu_D$ represent the emission of photons by the monomer and excimer species respectively, k_M and k_D comprise the sum of the various rate constants for all deactivation pathways for the monomer and excimer respectively (excluding k_{MD} and k_{DM}) (e.g., $k_M = k_{FM} + k_{NM}$ where k_{FM} and k_{NM} are the radiative (fluorescence) and nonradiative deactivation rate constants respectively of the monomer).

Solving these rate equations and applying the initial conditions $[^1M^*]_t = [^1M^*]_o$ and $[^1(MM)^*]_o = 0$ at $t = 0$, yields expressions given below for the decay of monomer and excimer fluorescence intensity with time. The emission intensity (monitored in a spectral region in which only the monomer $[^1M^*]$ species emits), will have the temporal form given by eqn [4]:

$$i_M(t) = k_{FM}\frac{[^1M^*]}{[^1M^*]_o} = \left[k_{FM}\frac{(\lambda_2 - X)}{(\lambda_2 - \lambda_1)}\right]\left[\exp(-\lambda_1 t) + A\exp(-\lambda_2 t)\right] \qquad [4]$$

where

$$A = \frac{(X - \lambda_1)}{(\lambda_2 - X)},$$

$$\lambda_{1,2} = \frac{1}{2}\left[X + Y \mp \left\{(Y - X)^2 + 4k_{MD}k_{DM}[^1M^*]\right\}^{1/2}\right]$$

The emission intensity collected from a spectral region in which the excimer ($^1[MM]^*$) species only emits, will have the temporal form given by eqn [5]:

$$i_D(t) = k_{FD}k_{DM}\frac{[^1D^*]}{[^1M^*]_o} = \left[\frac{k_{FD}k_{DM}[^1M]}{(\lambda_2 - \lambda_1)}\right]\left[\exp(-\lambda_1 t) - \exp(-\lambda_2 t)\right] \qquad [5]$$

If excimer formation follows Birks' scheme, then the decay profile of the monomer fluorescence should therefore be described by the sum of two exponential decay terms, and the excimer emission by an exponential growth term and a decay term (Figure 6). Only two decay constants, λ_1 and λ_2, should be required to fit both the monomer and excimer fluorescence profiles successfully. This is often not observed, which has led to the introduction of more complex interpretations.

In synthetic aromatic polymers, the effective local concentration of the aromatic molecules can be high while the concentration of polymer chains in solution may be very low. This means that excimer formation can be significant in such polymers, even in dilute solution. In fact, in some polymers the emission is dominated by excimers over monomers. The fact that the excimers are of lower energy than the corresponding excited monomer species also means that they can act as effective traps of excitation energy that might be migrating/relocating within a polymer chain as discussed below.

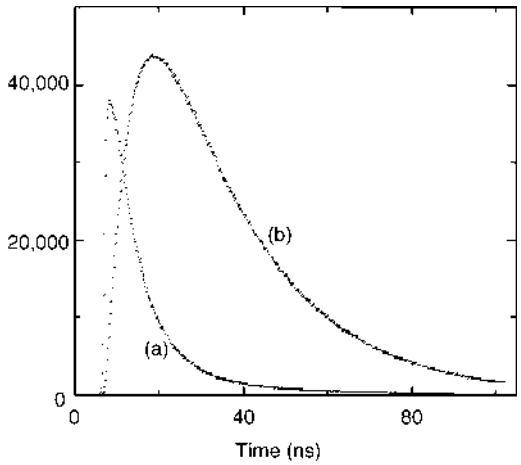

Figure 6 Examples of (a) monomer and (b) excimer fluorescence decays.

Excimer formation in solutions of aromatic polymers is generally thought to proceed via a complicated route that may involve a combination of electronic energy transfer and migration, segmental rotation, and chromophore diffusion. It is believed that energy migrates along a polymer chain until sites suitable for excimer formation are encountered, at which point the energy is trapped. A number of mechanisms for the formation of suitable excimer sites in polymers are then conceivable. These include:

(i) interactions between chromophores belonging to different chains (this *inter*molecular excimer formation is highly concentration dependent);
(ii) interactions between two chromophores far removed from each other, but on the same chain. This necessitates mainchain conformational changes and side group motions during the excited state lifetime, so as to bring both molecules together;
(iii) near neighbor interactions (c.f. Hirayama's '$n = 3$' rule). The configuration required for excimer formation need not exist prior to excitation provided that the two constituent groups can attain the required geometry within the lifetime of the excited chromophore; or
(iv) preformed sites, suitable for excimer formation, present in the polymer before the initial absorption of light occurs. These may be formed by interactions between nearest neighbor chromophores or between two chromophores on loops of the same, or different chains in cases where chain motion is restricted.

The first two of these mechanisms depend upon segmental motion of the polymer chain and the third mechanism may depend upon rotation of the chromophore/s to adopt the close proximity and optimum orientational configuration required for excimer formation. In solution, this may be, at least partially, diffusion-controlled, causing a dominant complicating factor in the kinetics of excimer formation in polymeric systems. One approach is therefore to substitute a time-dependent expression for the rate coefficient of excimer formation, $k_{DM}(t) = A + B/\sqrt{t}$, (based on the Einstein–Smoluchowski diffusion theory), for k_{DM}, into the rate equations. Such an approach produces the following expression for the time dependence of the monomer fluorescence, $I_M(t)$ in the presence of diffusion-controlled excimer formation and the absence of excimer back dissociation:

$$I_M(t) = A \exp\left\{ -at - b\sqrt{t} \right\} \qquad [6]$$

In some cases, such as when excimer back dissociation cannot be neglected, more complex functional forms have been derived, such as:

$$I_M(t) = A \exp\left\{ -at - b\sqrt{t} \right\} + B \exp\left\{ -t/\tau \right\} \qquad [7]$$

Similar expressions can be derived for the time dependence of the excimer fluorescence, $I_E(t)$. More detailed expressions for the term $k(t)_{DM}$, involving the complete form of the diffusion rate coefficient, have also been used in such approaches.

A related approach is the concept of 'convolution kinetics', in which the intimate dependence of excimer formation upon the monomer decay is emphasized. Once formed, the excimer is assumed to decay exponentially. The excimer fluorescence temporal profile is predicted to take the form:

$$I_E(t) = k_1(t) I_M(t) \otimes \exp\left(\frac{-t}{\tau} \right) \qquad [8]$$

where $\otimes$ represents the convolution integral and $I_E(t)$ and $I_M(t)$ are the measured decay profiles in the excimer and monomer emission regions respectively.

Electronic Energy Transfer, Energy Migration and Trapping

In polymers the high local concentration of chromophores makes additional deactivation pathways competitive with the other photophysical processes.The excess energy may transfer from the excited fluorophore of interest to a lower energy chromophore. Three of the more common possible mechanisms, through which energy transfer can occur, are:

(i) *Radiative* ('*trivial*') – involves the reabsorption of donor emission by an acceptor chromophore.

(ii) *Resonance* (*Dipole–Dipole*, *Förster*) – long-range (15–100 Å), nonradiative, single-step energy transfer process arising from Coulombic (dipole–dipole) interactions between an excited-state donor chromophore, D^*, and an initially unexcited acceptor, *A*. Förster derived a statistical expression for the (first-order) rate constant, for such a process, considering three-dimensional, dipole–dipole interactions between stationary molecules:

$$k(R)_{D^*\to A} = \frac{9000\kappa^2\phi_o \ln 10}{128\pi^5\eta^4 N\tau_D R^6}\int_0^\infty \frac{f_D(\bar{\nu})\varepsilon_A(\bar{\nu})}{\bar{\nu}}\mathrm{d}\bar{\nu} = \frac{1}{\tau_D}\left(\frac{R_o}{r_D}\right)^6 \qquad [9]$$

where R is the donor–acceptor distance, ϕ_o is the quantum yield for emission from the donor in the absence of acceptor molecules, η is the refractive index of the solvent, τ_D is the lifetime of the donor excited state (reduced by the onset of electronic energy transfer), N is Avogadro's number, κ is a constant dependent on the mutual orientation of the donor and acceptor molecules, and the integral term is a measure of the spectral overlap between the donor emission profile, $f_D(\bar{\nu})$ (in quanta cm^{-1} and normalized to unit area) and the acceptor absorption profile, $\varepsilon_A(\bar{\nu})$ (converted to molar extinction coefficients and measured on a wavenumber scale). R_o is the critical transfer distance, at which the probability of resonance transfer, $k(R)_{D^*\to A}$, is equal to the probability of spontaneous decay, $1/\tau_D$.

(iii) *Exchange mechanism* – also a nonradiative, single-step process but involves the exchange of electronic excitation when the donor and acceptor are close enough so that their electron clouds overlap. Consequently it only operates over small distances. Dexter has derived a rate expression for exchange energy transfer, again assuming that the donor and acceptor molecules are stationary:

$$k_{D^*\to A} = \left(\frac{2\pi}{\hbar}\right)Z^2\int f_{D^*}(\bar{\nu})\varepsilon_A(\bar{\nu})\mathrm{d}\bar{\nu} \qquad [10]$$

where $Z^2 = K^2\exp(-2R/L)$ and K and L are simply constants.

It has been clearly demonstrated that energy transfer, from pendant aromatic monomer units to chemically bound acceptor chromophores, can occur with high efficiency and is influenced significantly by polymer structure and conformation.

In addition to energy transfer between D^* and A (being different chemical species), more than one exchange of excitation energy between chemically and spectroscopically identical chromophores, attached by covalent bonds to a polymer chain, can occur. A single step between identical units is still termed energy transfer, while more than one such step in a sequence would constitute intramolecular energy migration.

Energy migration is a long-range process, thought to occur via a series of energy transfer steps occurring between chromophores, either: (i) along the polymer chain (often the adjacent groups); (ii) across loops in a single polymer chain; or (iii) across different polymer chains. The migration continues until either a trap site is encountered or one of the chromophores excited in the migration process deactivates spontaneously. Each step is believed to occur with a probability defined by the exchange or Förster mechanisms, but since the energy transfer occurs between donor and acceptor groups, which are of the same type, the spectral overlap integral involved in both mechanisms will usually be small. The distances over which such energy transfer can occur therefore cannot be large and typically takes values of <10–20 Å. It is worth noting that many authors consider the energy to migrate as an incoherent exciton (where the term exciton is used in a much looser sense than is customary among solid-state physicists).

Long-range energy migration can only occur by energy transfer from the initially excited chromophore to ground state chromophores of the same, or lower, energy. If energy transfer is favorable between one chromophore and another of the same energy, then back transfer to the initially excited chromophore is also highly probable. When there is a change in the emission and absorption dipole orientations, caused by rotation of the chromophores relative to one another on the time-scale of the energy transfer step, or the transfer is to a chromophore of significantly lower energy, then the probability of energy transfer back to the original chromophore might be reduced. The lower energy chromophores on the polymer chain may therefore act as 'traps' of the excitation originally absorbed by a higher energy chromophore, even in the absence of excimer energy trap sites.

Various approaches have been proposed for modeling the kinetics of energy migration. One can consider the time-dependent nature of each energy transfer step of the migration process, and introduce a time-dependent rate coefficient for each step of the energy migration process, $k_{DD}(t)$. The form for $k_{DD}(t)$ by Sienicki *et al.*, is given by eqn [11], which leads to nonexponential functions predicted for the ensemble averaged probability that an originally excited donor

molecule is still excited at time t:

$$k_{DD}(t) = \left(\frac{\Delta}{n}\right)\gamma_\Delta c_D \Gamma(1-\Delta/n)\left(\frac{t}{\tau_{OD}}\right)^{\Delta/n-1} \quad [11]$$

where $\Delta(=1, 2, 3)$ is the dimensionality of the medium, $n(=6, 8, 10,\ldots)$ is the order of the interaction, τ_{OD} is the lifetime of the donor fluorescence, $\Gamma(x)$ is the gamma function, γ_Δ is a constant dependent upon Δ and c_D is the donor concentration.

Energy migration has also been considered as a diffusion-like process, and the Smoluchowski diffusion theory adopted. The rate coefficient for energy migration, $k(t)_{EM}$, is then expressed in a form similar to the time-dependent rate coefficient for partially diffusion-controlled reactions (eqn [12]). In this type of treatment, the rate of change of the concentration of excited monomer chromophores, $^1M^*$, is assumed to be given by eqn [13]:

$$k(t)_{EM} = 4\pi N'_A R D \rho\left[1 + R\rho(\pi D t)^{-1/2}\right] \quad [12]$$

$$\frac{d[^1M^*]}{dt} = I_o - \left(\frac{1}{\tau_o} + k(t)_{EM}\right)[^1M^*] \quad [13]$$

In these expressions, τ_o is the reciprocal of the sum of the rate constants for radiative and nonradiative deactivation, ρ is the reaction probability per collision, R is the interaction distance, and N'_A is the number density of the energy acceptors or traps ($N'_A = NC_A$, with the concentration of acceptors, C_A, in mol cm^{-3}). D is the mutual diffusion coefficient which can be replaced by a migration coefficient, Λ, R can be replaced by the Förster critical transfer distance, R_o, and the reaction probability, ρ, can be set to 0.5 (since R_o is the distance at which the probability of energy transfer is 0.5). This leads to a simplified expression for $k(t)_{EM}$, which can be substituted into eqn [13], to produce an expression (eqn [14]) for the monomer decay:

$$I(t) = A\exp\left(-\left(\frac{1}{\tau_o} + 2\pi N'_A R_o \Lambda[^1M^*]\right)t - 2N'_A R_o^2(\pi\Lambda)^{1/2}[^1M^*]\sqrt{t}\right) \quad [14]$$

Yokota and Tanimoto have also developed a statistical representation of Förster's theory and derived an expression for dipole–dipole transfer in a fluid medium in which the distribution of excited-state donor and acceptor molecules is determined, both by diffusion and by the decay and transfer of the donor excited state. According to the Yokota and Tanimoto approach, the decay law is approximated by the expression:

$$I(t) = A\exp\left\{\frac{-t}{\tau_o} - 2B\delta\left(\frac{-t}{\tau_o}\right)^{1/2}\right\} \quad [15]$$

where $\delta = [C]/[C]_o$ (C_o is the critical molar concentration of acceptor molecules given by: $[C_o] = 3000/2\pi^{3/2}NR_o^3$, and B is a complex, time-dependent function given by: $B = (1 + 10.87x + 15.5x^2/1 + 8.743x)^{3/4}$ with $x = D\alpha^{-1/3}t^{2/3}$, $\alpha = R_o^6/\tau_o$ and D is the 'diffusion coefficient' representing energy migration.

Emission depolarization is often exploited in the study of energy migration in polymers. Only molecules with some vector component of their absorption transition dipole lying parallel to the plane of polarization of the incident radiation, will absorb incoming radiation. In the absence of angular motion of the absorbing chromophore within the lifetime of the excited state, emission from the same chromophore will also be highly polarized; however, upon each step in the energy migration process, some loss of polarization will be incurred (**Figure 7**). The extent of depolarization of emission is usually referred to by the term fluorescence anisotropy, $r(t)$ (eqn [16]):

$$r(t) = \frac{I_{\parallel}(t) - I_{\perp}(t)}{I_{\parallel}(t) + 2I_{\perp}(t)} \quad [16]$$

The rate of energy migration can be estimated directly through observation of the time-dependent decay of $r(t)$ in the absence of other fluorescence depolarizing processes, such as molecule rotation (**Figure 8**).

One of the important properties which characterizes excitation energy transport in a system of n identical molecules, which are distributed randomly in space, is the Green function, $G(r,t)$. The Green function can be thought of as the time-dependent probability of finding an excitation at position r on the polymer chain at time t. It has been shown that the overwhelming contribution to fluorescence polarization is due to fluorescence from the sites that were initially excited. Thus, $G^s(r,t)$ is directly related to the fluorescence anisotropy if the polarization memory of secondary excited molecules can be neglected. The $r(t)$ for donor–donor energy transfer is then given by:

$$r(t) = \alpha r_{rot}(t)G^s(t) \quad [17]$$

in which r_{rot} describes the anisotropy in the absence of energy transfer, i.e., in the presence of chromophore

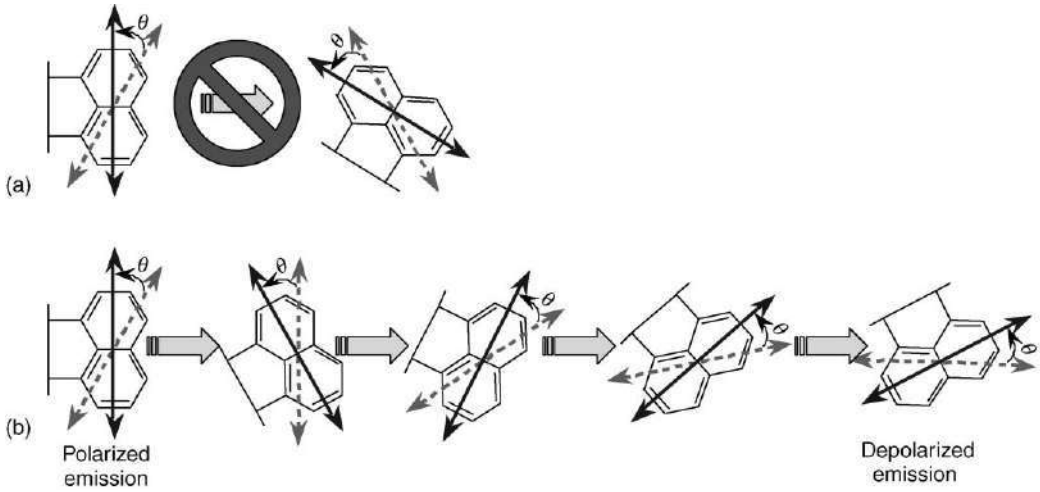

Figure 7 Schematic representation of energy migration. θ is the inherent angle between the donor emission (dotted arrow) and acceptor absorption (solid arrow) transition dipoles of the chromophore (a) orthogonal donor emission and acceptor absorption dipoles – no energy transfer/migration, emission depolarized only due to θ. (b) some vector component of the absorption dipole of the acceptor is parallel with that of the donor emission dipole – this and subsequent energy transfer steps can occur resulting in loss of emission polarization.

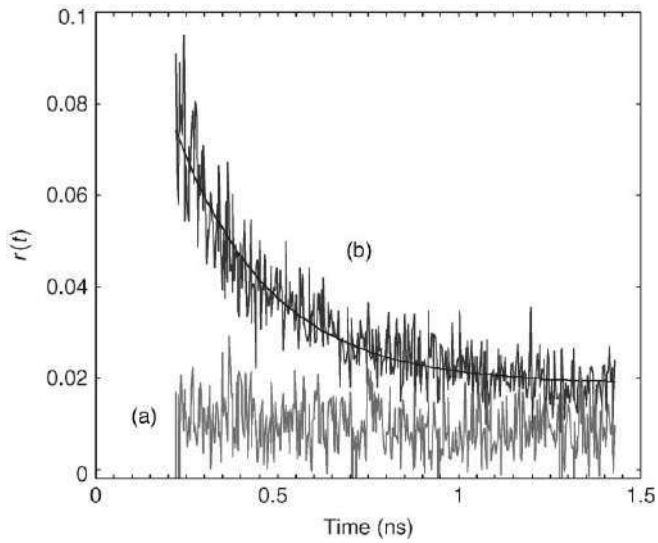

Figure 8 Temporal decays of $r(t)$ from (a) a polymer in which emission depolarization due to energy migration is complete on ultrafast time-scales, and (b) a polymer in which the rate of energy migration has been inhibited. Emission collected from polymers in frozen glass to exclude polymer motion contributions to emission depolarization.

rotation only, and α is the anisotropy from time-independent sources of depolarization.

The determination of the functional form of $G^s(r, t)$, resulting from excitation energy migration in polymer systems, is a challenging theoretical problem to which a number of approaches have been applied. The Gochanour, Anderson and Fayer approach has had some success in arriving at an expression for $G^s(r, t)$, represented as the time decay of emission polarization curves.

Polymeric systems that undergo energy migration, yet contain no excimer forming or other nonmonomeric, energy-trapping sites, are rare. In the presence of singlet energy traps, the problem can be treated as being analogous to energy transfer (via the Förster mechanism), following a diffusive process in a similar manner to that described above.

Fredrickson and Frank have derived expressions for the temporal decays of the monomer and excimer fluorescence when one-dimensional electronic singlet

energy migration is the major excimer-forming path, and segmental rotation and chromophore diffusion are considered. The Fredrickson and Frank model assumes inherently that excimers are formed predominantly via a nearest-neighbor mechanism, where rapid energy migration is at least an order of magnitude faster than any competing rotationally sampled pathways. The expressions for the fluorescence decay of both the monomer, $I_M(t)$, and excimer species, $I_E(t)$, are highly nonexponential, with the expression for $I_M(t)$ being given by eqn [18]:

$$I_M(t) = Q_{fm} k_m (1-q)^2 \times \exp[(4q^2 W - k_m - k_{rot})t] \mathrm{erfc}[2q(Wt)^{1/2}] \quad [18]$$

In eqn [18], Q_{fm} is the quantum efficiency for monomer fluorescence and k_m is a rate constant, which encompasses the total rate of monomer decay by radiative and nonradiative pathways in the absence of energy migration and rotational sampling. Similarly, k_{rot} represents the rate at which an excited monomer is transformed into an excimer by segmental rotation. The parameter W represents the rate constant for nearest-neighbor electronic energy transport between the donor (monomer) chromophores, averaged over both the transition moment orientation and the interchromophore separation associated with each dyad.

Energy transfer from pendant aromatic monomer units to a specific chemically bound acceptor chromophore can occur in some polymers following energy migration. This phenomenon is known as the 'antenna effect', which has commercial applications such as in the sensitization of photo-oxidation–reduction, i.e. reactions, in the stabilization of polymers against photodegradation or in the modeling of photosynthesis by light harvesting. The kinetic behavior of such processes is yet to be modeled definitively, but the approaches discussed above (and others) have been used with varying degrees of success to date.

List of Units and Nomenclature

Ångstrom (Å)	10^{-10} m
chromophore	An aromatic molecule that absorbs (and may emit) light
exciton	An excited state localized on a particular chromophore at any given time
femtoseconds	10^{-15} s
intermolecular	Occurring between molecules or polymer chains
intramolecular	Occurring within a single molecule or polymer chain
petawatts	10^{15} W
δ-pulse excitation	Excitation by an infinitely short laser pulse

See also

Chemical applications of lasers: Non-linear spectroscopies. **Lasers:** Excimer Laser. **Optical Parametric Devices:** Optical Parametric Oscillators (Pulsed). **Spectroscopy:** Second Harmonic Spectroscopy.

Further Reading

Andrews DL and Demidov AA (1999) *Resonance Energy Transfer*. New York: Wiley.

Birks JB (1970) *Photophysics of Aromatic Molecules*. London: Wiley-Interscience.

Fleming GR (1986). *Chemical Applications of Ultrafast Spectroscopy*, vol. 13. New York: Oxford University Press.

Fredrickson GH and Frank CW (1983) Nonexponential transient behavior in fluorescent polymeric systems. *Macromolecules* 16: 572–577.

Ghiggino KP and Smith TA (1993) Dynamics of energy migration and trapping in photoirradiated polymers. *Progress in Reaction Kinetics* 18: 375–436.

Gochanour CR, Andersen HC and Fayer MD (1979) Electronic excited state transport in solution. *Journal of Chemical Physics* 70: 4254–4271.

Guillet JE (1985) *Polymer Photophysics and Photochemistry: An Introduction to the Study of Photoprocesses in Macromolecules*. New York: Cambridge University Press.

Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. New York: Plenum.

O'Connor DV and Phillips D (1984) *Time-Correlated Single Photon Counting*. London: Academic Press.

Sienicki K, Itagaki H and Mattice WL (1989) On the theory of concentration depolarization of fluorescence in one- and two-component systems for multipole interactions in one, two, and three dimensional medium. *Journal of Chemical Physics* 91: 4515–4521.

Van der Meer BW, Coker G and Chen S-YS (1994) *Resonance Energy Transfer: Theory and Data*. New York: VCH.

Van Hecke GR and Karukstis KK (1998) *A Guide to Lasers in Chemistry*. Boston, MA: Jones and Bartlett.

Yokota M and Tanimoto O (1967) Effects of diffusion on energy transfer by resonance. *Journal of the Physics Society of Japan* 22: 779–784.

TOMOGRAPHY

Contents

Optical Coherence Tomography
Tomography and Optical Imaging

Optical Coherence Tomography

S A Boppart, University of Illinois, Urbana, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical coherence tomography (OCT) is an emerging imaging technique for a wide range of biological, medical, and material investigations. OCT was initially developed for imaging biological tissue because it permits the imaging of tissue microstructure *in situ*, yielding micron-scale image resolution without the need for excision of a specimen and tissue processing. OCT is analogous to ultrasound B-mode imaging except that it uses low-coherence light rather than sound and performs cross-sectional imaging by measuring the backscattered intensity of light from structures in tissue. The OCT image is a gray-scale or false-color two-dimensional representation of backscattered light intensity in a cross-sectional plane. The OCT image represents the differential backscattering contrast between different tissue types on a micron scale. Because OCT performs imaging using light, it has a one to two order-of-magnitude higher spatial resolution than ultrasound and does not require contact with the specimen or sample.

OCT was originally developed and demonstrated in ophthalmology for high-resolution tomographic imaging of the retina and anterior eye. Because the eye is transparent and is optically accessible, it is well suited for diagnostic OCT imaging. OCT is promising for the diagnosis of retinal disease because it can provide images of retinal pathology with 10 micron resolution, almost one order-of-magnitude higher than previously possible using ultrasound. Recently, OCT has been applied for imaging in a wide range of nontransparent tissues. In tissues other than the eye, the imaging depth is limited by optical attenuation due to scattering and absorption. Ophthalmic imaging is typically performed at 800 nm wavelengths. However, because optical scattering decreases with increasing wavelength, OCT imaging in nontransparent tissues is possible using longer near-infrared wavelengths. In most tissues, imaging depths of 2–3 mm can be achieved using a system detection sensitivity of 100 to 110 dB. Imaging studies have been performed in a wide range of biological, medical, and surgical specialties including developmental biology, cardiology, gastroenterology, urology, and neurosurgery. High-resolution OCT using short coherence length, short pulsed light sources has also been demonstrated and axial resolutions less than 2 μm have been achieved. High-speed real-time OCT at image acquisition rates of 4 to 8 frames per second have also been demonstrated. OCT has been extended to perform Doppler imaging of blood flow and birefringence imaging to investigate tissue injury. Different imaging delivery systems including transverse imaging catheters and forward imaging devices have been developed to enable internal body OCT imaging. Most recently, an OCT catheter has been combined with endoscope-based delivery to perform *in vivo* imaging in animal models and human patients.

This chapter will provide an overview of the OCT technology, beginning with the basic principles of operation. Technological advancements over the last decade have enabled OCT to transition from laboratory- to clinical-based imaging. These enabling advancements will be discussed, followed by representative applications in biology, medicine, surgery, and materials.

Principles of Operation

OCT is based on optical ranging, the high-resolution, high dynamic range detection of backscattered light. In contrast to ultrasound, the velocity of light is extremely high. Therefore, the echo time delay of reflected light cannot be measured directly and interferometric detection techniques must be used. One method for measuring echo time delay is to use low coherence interferometry or optical coherence

domain reflectometry. Low-coherence interferometry was first developed for measuring reflections in fiber optics and optoelectronic devices.

The echo time delay of reflected light is measured by using a Michelson-type interferometer (Figure 1).

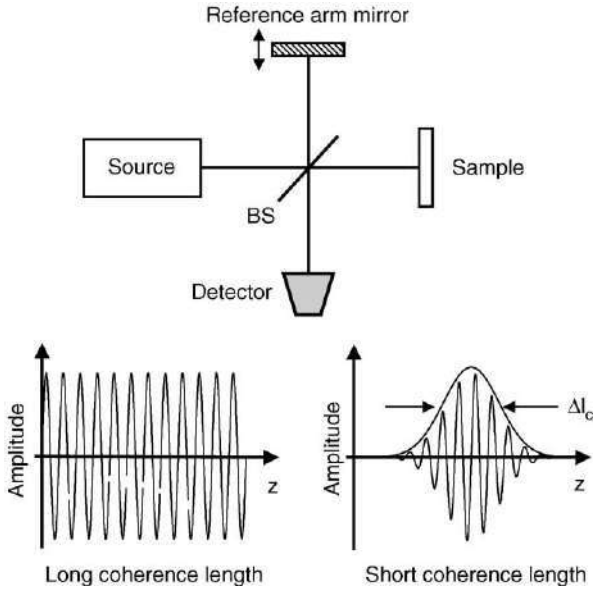

Figure 1 Schematic illustrating the concept of low-coherence interferometry. Using a short coherence-length light source and a Michelson-type interferometer, interference fringes are observed only when the path lengths of the two interferometer arms are matched to within the coherence length of the optical source. Abbreviations: BS, beam splitter; Δl_c, coherence length.

The light reflected from the specimen or sample is interfered with light that is reflected from a reference path of known path length. Interference of the light reflected from the sample arm and reference arm of the interferometer can occur only when the optical path lengths of the two arms match to within the coherence length of the optical source. As the reference arm optical path length is scanned, different echo delays of backscattered light from within the sample are measured. The interference signal is detected at the output port of the interferometer, electronically band-pass filtered, demodulated, digitized, and stored on a computer. The position of the incident beam on the specimen is typically scanned in the transverse direction and multiple axial measurements are performed. This generates a two-dimensional data array that represents the optical backscattering through a cross-sectional plane in the specimen (Figure 2). The logarithm of the backscatter intensity is then mapped to false-color or gray-scale and displayed as an OCT image. The interferometer in an OCT instrument can be implemented using a fiber optic coupler and beam-scanning can be performed with small mechanical galvanometers in order to yield a compact and robust system (Figure 3).

In contrast to conventional microscopy, the axial resolution in OCT images is determined by the coherence length of the light source. The axial point spread function of the OCT measurement as defined by the signal detected at the output of the

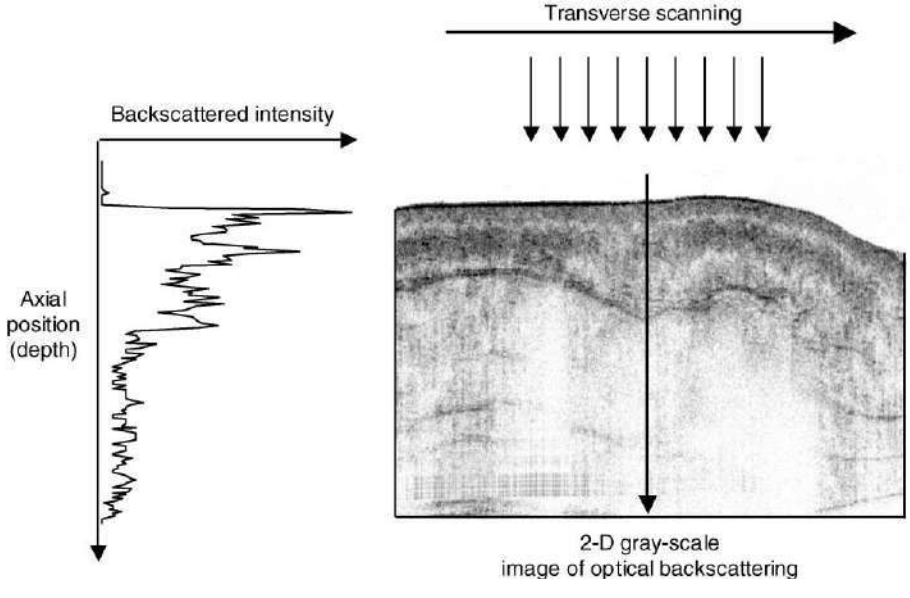

Figure 2 OCT image generation. The OCT image is typically acquired by performing axial measurements of optical backscatter at different transverse positions on the specimen and displaying the resulting two-dimensional data set as a gray-scale (or false-color) image.

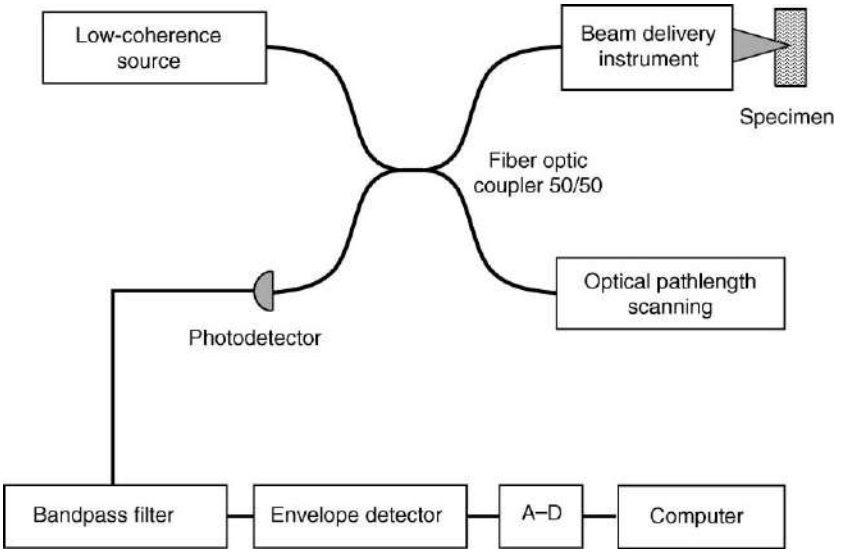

Figure 3 Schematic representation of an OCT system implemented using fiber optics. The Michelson interferometer is implemented using a fiber-optic coupler (beam splitter). The sample arm contains a beam-delivery instrument and the reference arm contains a mechanism for rapidly scanning the optical pathlength. The output of the interferometer is detected with a photodiode, filtered, demodulated, analog-to-digital (A–D) converted, and processed by a computer.

interferometer is the electric-field autocorrelation of the source. The coherence length of the light is the spatial width of the field autocorrelation and the envelope of the field autocorrelation is equivalent to the Fourier transform of its power spectrum. Thus, the width of the autocorrelation function, or the axial resolution, is inversely proportional to the width of the power spectrum. For a source with a Gaussian spectral distribution, the axial resolution Δz is given:

$$\Delta z = \frac{2\ln 2}{\pi} \cdot \frac{\lambda^2}{\Delta\lambda}$$

where Δz and $\Delta\lambda$ are the full-widths-at-half-maximum of the autocorrelation function and power spectrum respectively and λ is the source central wavelength. Figure 4 illustrates the dependence of the coherence length (axial resolution) on the bandwidth of the optical source. To achieve high axial resolution requires broad bandwidth optical sources. Resolution is also improved by using shorter wavelengths, however, shorter wavelengths are scattered and absorbed more in biological tissue.

The transverse resolution in an OCT imaging system is determined by the focused spot size in analogy with conventional microscopy and is given by:

$$\Delta x = \frac{4\lambda}{\pi} \cdot \frac{f}{d}$$

Where d is the spot size on the objective lens and f is its focal length. High transverse resolution can be obtained by using a large numerical aperture and

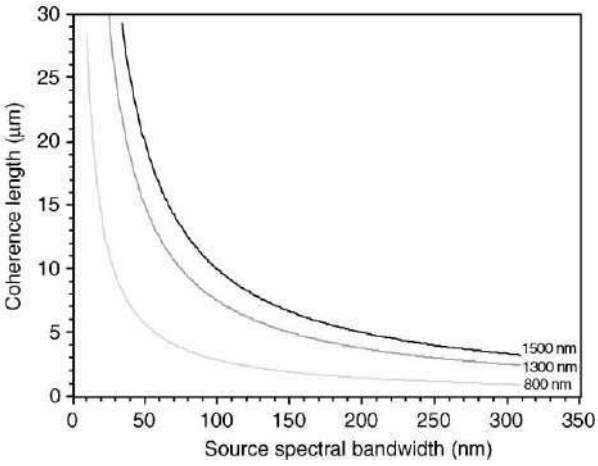

Figure 4 Dependence of coherence length (axial resolution) on the bandwidth of the optical source. Curves are shown for imaging wavelengths of 800 nm, 1300 nm, and 1500 nm. To achieve micron-scale resolution imaging, broad spectral bandwidths are necessary.

focusing the beam to a small spot size. The transverse resolution is also related to the depth of focus or the confocal parameter $2z_R$ (two times the Raleigh range).

$$2z_R = \frac{\pi \Delta x^2}{2\lambda}$$

Thus, increasing the transverse resolution results in a reduced depth of field. Typically, the confocal parameter or depth of focus is chosen to match the desired depth of imaging. Increased resolution may

also be obtained by using a higher numerical aperture lens or objective and spatially tracking the focus through the specimen.

Finally, the detection signal-to-noise is given by the optical power backscattered from the sample divided by the noise equivalent bandwidth:

$$\mathrm{SNR} = 10\log\left(\frac{\eta}{\hbar\omega}\frac{P_{\mathrm{SAM}}}{\mathrm{NEB}}\right)$$

Depending upon the desired signal to noise performance, incident powers of 5–10 mW are typically required for OCT imaging of 250–500 square pixel images at several frames per second. If lower data acquisition speeds or signal-to-noise can be tolerated, power requirements can be reduced accordingly.

Technological Developments

Since the inception of OCT in the early 1990s, there has been rapid technological developments aimed at improving the imaging resolution, acquisition rate, and methods for beam delivery to the tissue or sample. Investigators have also explored other imaging methods using the principles of OCT to extract information from the tissue or sample. Some of these methods have included acquiring optical Doppler signals from moving scatterers or structures, obtaining images based on the polarization state of the returned light, and extracting spectroscopic information based on the local absorption or scattering properties of the tissue.

Optical Sources / High-Resolution Imaging

The majority of OCT imaging systems to date have used superluminescent diodes (SLDs) as low coherence light sources. SLDs are manufactured with a similar structure to that of laser diodes, but commonly end facets are angle-cleaved to suppress lasing. These device structures do not support oscillation modes and generate output based on amplified spontaneous emission, resulting in emission spectra that are broader than laser diodes. SLDs are commercially available at a range of wavelengths including 800 nm, 1.3 μm, and 1.5 μm and are attractive because they are compact, have high efficiency, and low noise. However, output powers are typically limited to only a few milliwatts which limits fast real-time acquisition rates, and the available bandwidths are relatively narrow, permitting imaging with 10–15 micron resolution. Recent advances in short-pulse solid-state laser technology make these sources attractive for OCT imaging in research applications. Femtosecond solid-state lasers can generate tunable, low-coherence light at powers sufficient to permit high-speed OCT imaging. Short pulse generation has been achieved across the full wavelength range in titanium:sapphire ($Ti:Al_2O_3$) from 0.7 μm to 1.1 μm and over more limited tuning ranges near 1.3 μm and 1.5 μm in chromium: forsterite ($Cr^{4+}:Mg_2SiO_4$) and chromium:yttrium-aluminum-garnet (Cr^{4+}:YAG) lasers, respectively. OCT imaging with resolutions of 1 μm and 5 μm has been demonstrated at 800 nm and 1.3 μm respectively using $Ti:Al_2O_3$ and $Cr^{4+}:Mg_2SiO_4$ sources. More compact and convenient sources such as superluminescent fiber sources, are currently under investigation. The titanium:sapphire laser technology is routinely used in multi-photon microscopy applications for its high peak intensities to enable multi-photon absorption and subsequent emission of fluorescence from exogenous fluorescent contrast agents. Combined OCT and multi-photon microscopy has been used to provide complementary image data using a single optical source.

Fast Scanning

The short-pulse solid-state laser technology not only provides broad spectral bandwidths for high-resolution OCT imaging, but also higher output powers to enable fast real-time OCT imaging. Higher incident powers are required to maintain equivalent signal-to-noise ratios when scanning at a faster rate. Linearly translating a reference arm mirror is problematic at high rates and provides axial scan frequencies of approximately 100 Hz, depending on the mirror size and the translating galvanometer. Several investigators have utilized rotating glass cubes, piezoelectric modulators, and multi-pass optical cavities to increase axial scan rates while maintaining scan ranges of 1–2 mm. An optical delay based on the principles used in femtosecond pulse shaping has been demonstrated for OCT. This delay line spectrally disperses the reference arm beam with a grating. The dispersed beam is then focused by a lens on to a rotating mirror mounted on a galvanometer. The mirror, located in the Fourier-transform plane of the lens, imparts a wavelength-dependent phase shift on the light. Subsequently, when re-coupled back into the interferometer, this phase-shift is equivalent to a time-delay in the time domain. The use of high-speed resonant galvanometers has permitted axial scan rates as high as 8 kHz over scan ranges of several millimeters. Depending on the image pixel size (number of axial scans within each image), this scan rate can provide video-rate OCT imaging (30 frames per second). An example of high-speed *functional* OCT imaging is shown in Figure 5. Images of a

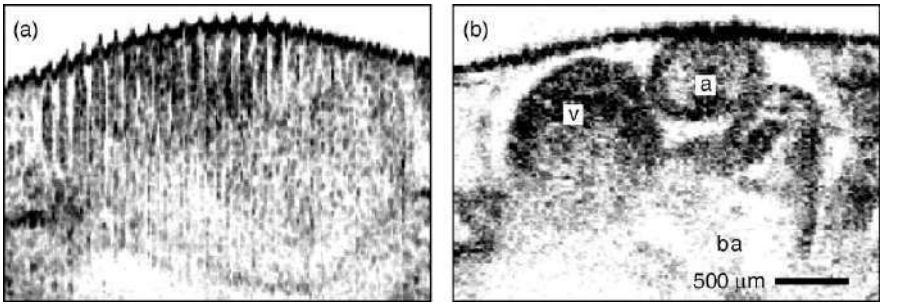

Figure 5 Comparison of slow and fast OCT imaging of a beating *Xenopus* (African frog) tadpole heart. The OCT image in (a) was acquired in 30 seconds and shows oscillations and blurring from the beating heart. The image on the right (b) was acquired in 250 ms (4 frames per second), fast enough to capture the anatomy of the heart with minimal motion artifacts. Abbreviations: a, atrium; ba, bulbous arteriosus; v, ventricle. (Images reprinted with permission from Boppart SA, Tearney GJ, Bouma BE, *et al.* (1997) Noninvasive assessment of the developing *Xenopus* cardiovascular system using optical coherence tomography. *Proceedings of the National Academy of Sciences* 94: 4256–4261.)

beating *Xenopus laevis* (African frog) tadpole were compared at slow (30 s per image) and at fast (250 ms per image) acquisition rates. Artifacts due to cardiac motion were present at slow acquisition rates, but were minimized with acquisition rates of 4 frames per second. Functional cardiac parameters could be measured and high-speed processes such as chamber and valve function could be visualized in real-time.

Depth-priority scanning, as described above, is performed by rapidly varying the optical delay in the reference arm and collecting a single axial scan before translating the beam laterally and repeating this depth-scanning. An alternative method for generating OCT images is with transverse-priority scanning. This is equivalent to optical sectioning in confocal and multi-photon microscopy. Three-dimensional OCT imaging using transverse-priority scanning can be obtained by stepping the position of the reference arm mirror after each *en face* image is acquired. The OCT images produced from transverse-priority scanning can be correlated with confocal or multi-photon microscopy images, which are acquired with the same scanning method. This method can also utilize higher numerical aperture objective lenses to provide high transverse resolutions since a large depth of focus is not needed as in depth-priority OCT scanning. The combination of OCT with high-numerical aperture objectives has been termed optical coherence microscopy.

Doppler Optical Coherence Tomography

Conventional OCT detects the amplitude of the optical backscatter versus position within the sample. Doppler OCT is a technique which has been applied to measuring fluid flow within small capillary tubes and *in vivo* vessels in the skin, retina, and heart of small animal models and humans. This technique is based on digitally sampling the interference fringes that are produced from a moving scatterer within the sample or specimen and determining the Doppler frequency shift in the signal due to the moving scatterer. Figure 6 illustrates two- and three-dimensional optical Doppler data for fluid flow through a cylindrical tube. A two-dimensional cross-sectional profile of fluid flow through a silicon tube with a 600 μm inner-diameter is plotted in Figure 6. The acquired data (points) closely match the predicted profile (line) for these tube dimensions. Below the plot, one method of obtaining a three-dimensional fluid flow profile is illustrated. While imaging the tube in cross-section, the fluid flow velocity was altered with a perfusion pump. The detection filter bandwidth of the OCT electronics served a windowing function, detecting only flow velocities with frequencies within the detection bandwidth. For an unknown flow profile, the center frequency of the detection filter can be shifted to map the flow velocities present within the sample.

Beam Delivery

The OCT imaging technology is modular in design and a variety of optical instruments can be used to deliver the OCT beam to the tissue or sample. Because OCT is fiber-optic based, single optical fibers can be used to deliver the OCT beam and collect the reflected light. The OCT technology can readily be integrated into existing optical instruments such as research and surgical microscopes, ophthalmic slit-lamp biomicroscopes, and hand-held imaging probes.

Imaging penetration is determined by the optical absorption and scattering properties of the tissue or specimen. The imaging penetration for OCT ranges from tens of millimeters for transparent tissues such as the eye to less than 3 mm in highly-scattering tissues such as skin. To image highly-scattering tissues deep within the body, novel beam-delivery

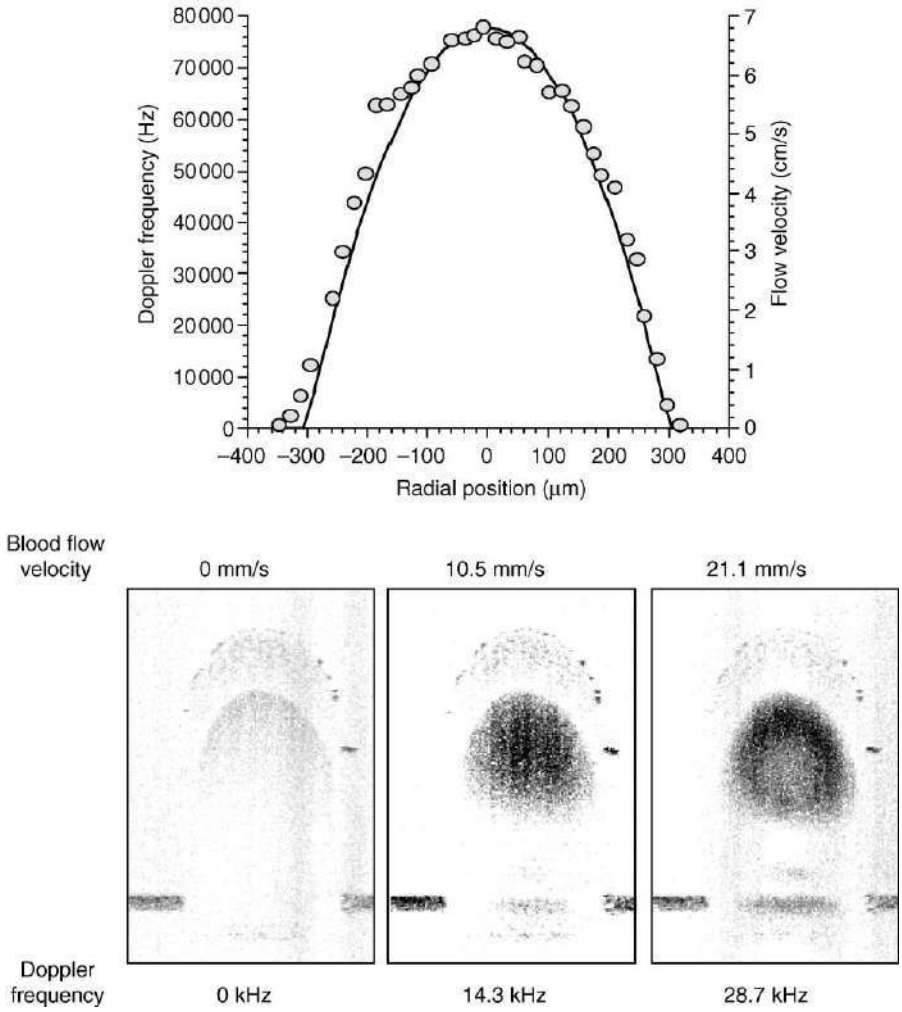

Figure 6 Doppler OCT. A two-dimensional plot of the fluid-flow velocity profile of flowing blood through a 600 μm inner-diameter tube is shown. Data points acquired using OCT closely follow the theoretical laminar flow profile (line). Two-dimensional Doppler OCT images of blood flow through the tube are shown below the plot. As fluid flow increases, only fluid-flow velocities within the bandwidth of the detection electronics are detected. This represents one method by which 3-D fluid flow can be imaged.

instruments have been developed to relay the OCT beam to the site of the tissue to be imaged. An OCT catheter has been developed for insertion into biological lumens such as the gastrointestinal tract. Used in conjunction with endoscopy, the 1 mm diameter catheter can be inserted through the working channel of the endoscope for simultaneous OCT and video imaging. Minimally-invasive surgical procedures utilize laparoscopes, which are long, thin, rigid optical instruments that permit video-based imaging within the abdominal cavity. Laparoscopic OCT imaging has been demonstrated by passing the OCT beam through the optical elements of a laparoscope. Deep solid-tissue imaging is possible with the use of fiber-needle probes. Small (400 μm diameter) needles housing a single optical fiber and micro-optic elements can be inserted into solid tissues and rotated to acquire OCT images. Recently, microfabricated micro-electro-optical-mechanical systems (MEOMS) technology has been used to miniaturize the OCT beam scan mechanism.

Applications

Developmental Biology

OCT has been demonstrated in the field of developmental biology as a method to perform high-resolution, high-speed imaging of developing morphology and function. Cellular-level imaging is possible, providing a non-invasive technique for visualizing cellular processes such as mitosis and migration. Imaging studies have been performed on several standard biological animal models commonly employed in developmental biology investigations including *Rana pipiens* (Leopard frog), *Xenopus laevis*

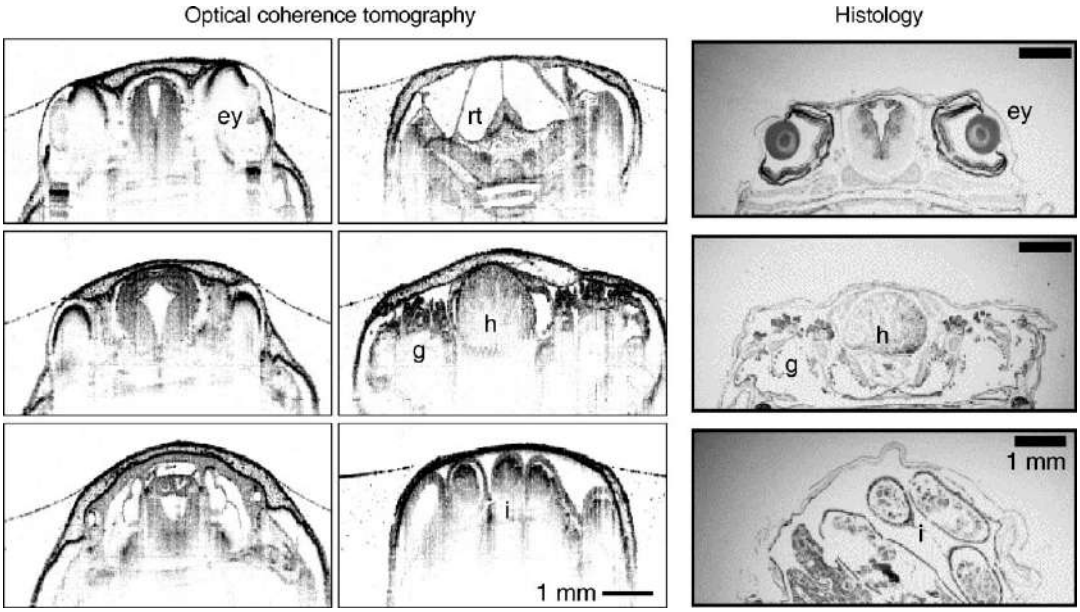

Figure 7 OCT images and corresponding histology of a Stage 49 (12 day) *Rana pipiens* (Leopard frog) tadpole. OCT images in the left and middle columns were acquired with the OCT beam incident from the dorsal and ventral sides of the specimen, respectively. The corresponding histology in the right column illustrates strong correlations between OCT images and the actual tissue morphology. Abbreviations: ey, eye; g, gills; h, heart; i, intestines; rt, respiratory tract. (Images reprinted with permission from Boppart SA, Brezinski ME, Bouma BE, *et al.* (1996) Investigation of developing embryonic morphology using optical coherence tomography. *Dev. Biol.* 177: 54–64.)

(African frog), and *Brachydanio rerio* (zebrafish) embryos and eggs, and the murine (mouse) model.

A series of cross-sectional images acquired *in vitro* from the dorsal and ventral sides of a Stage 49 (12 day) *Rana pipiens* (Leopard frog) tadpole is shown in Figure 7. Features of internal architectural morphology are clearly visible in the images. The image of the eye differentiates structures corresponding to the cornea, lens, and iris. Internal morphology not accessible in one orientation due to the specimen size or shadowing effects can be imaged by reorienting the specimen and scanning in the same cross sectional image plane. With the OCT beam incident on the ventral side, images of the respiratory tract, ventricle of the heart, internal gills, and gastrointestinal tract were acquired.

These images can be compared with corresponding histology (Figure 7). Histological images are acquired by euthanizing the specimen, immersing the specimen in a chemical fixative, and physically sectioning thin (2–5 micron-thick) slices using a microtome. The slices are placed on a microscope slide, selectively stained to highlight particular features, and viewed with light microscopy. The correlations between OCT and histology images are strong, suggesting that OCT images can accurately represent the *in vivo* specimen morphology. The potential exists to repeatedly image specimens to quantify organo- and morphogenesis throughout development. Technologies such as OCT are likely to become increasingly important in functional genomics, relating genetic features to the morphology and function in living specimens.

OCT images represent the optical backscatter intensity from regions within the tissue or sample. Because OCT relies on the inherent optical scattering changes to produce imaging contrast, no exogenous contrast agents or fluorophores are necessary. This permits long-term sequential imaging of development *in vivo* without loss of specimen viability. Repeated images of a developing zebrafish embryo within its egg beginning immediately after fertilization and up until hatching have been demonstrated without loss of specimen viability or without developmental abnormalities. In this example, the zebrafish egg and embryo were semi-transparent and the use of OCT significantly complemented observations made using light microscopy. By imaging subtle differences in backscattering intensity, interfacial structural layers millimeters deep within specimens can be clearly delineated.

Previous OCT images have characterized morphological features within biological specimens. These structures are static even though they may have been acquired from *in vivo* specimens. *In vivo* imaging in living specimens, particularly in larger organisms and

for medical diagnostic applications, must be performed at high speeds to eliminate motion artifacts within the images. Functional OCT imaging is the quantification of *in vivo* images which yield information characterizing the functional properties of the organ system or organism. High speed OCT permits both the positioning and manipulation of specimens as well as imaging in real time and is a powerful technology for functional imaging in developmental biology animal models.

Studies investigating normal and abnormal cardiac development have been frequently limited by an inability to access cardiovascular function within the intact organism. OCT has been demonstrated for the high-resolution assessment of structure and function in the developing *Xenopus laevis* (African frog) cardiovascular system (Figure 5). The morphology of the *in vivo* cardiac chambers is clearly delineated. Image acquisition rates are fast enough to capture the cardiac chambers in mid-cycle. With this capability, images can be acquired at various times during the cardiac cycle. These frames can be displayed in real-time to produce a movie illustrating the dynamic, functional behavior of the developing heart. OCT, unlike technologies such as computed tomography and magnetic resonance imaging, provides high-speed *in vivo* imaging, allowing quantitative dynamic activity, such as ventricular ejection fraction, to be assessed.

Cellular Imaging

Although previous studies have demonstrated *in vivo* OCT imaging of tissue morphology, most have imaged tissue at $\sim 10-15\ \mu m$ resolutions, which does not allow differentiation of cellular structure. The ability of OCT to identify the mitotic activity, the nuclear-to-cytoplasmic ratio, and the migration of cells has the potential to not only impact the fields of cell and developmental biology, but also impact medical and surgical disciplines for the early diagnostics of disease such as cancer.

The *Xenopus laevis* (African frog) tadpole has been used to demonstrate the feasibility of OCT for high-resolution *in vivo* cellular and subcellular imaging. Many of the cells in this common developmental biology animal model are rapidly dividing and migrating during the early growth stages of the tadpole, providing an opportunity to image dynamic cellular processes. Three-dimensional volumes of high-resolution OCT data have been acquired from these specimens throughout development. From this 3-D data set, cells undergoing division were identified and tracked in three dimensions. In a similar manner, 3-D data sets were acquired to track single melanocytes (neural crest cells) as they migrated through the living specimens. The ability of OCT to characterize cellular processes such as mitosis and migration not only are of interest in cell and developmental biology, but also have relevance for cancer diagnostics and tumor metastasis.

An example of cellular-level OCT imaging in these specimens is shown in Figure 8. This composite image (0.83×1 mm, 1800×1000 pixels) was acquired using a titanium:sapphire laser with a broad bandwidth (~ 260 nm). The axial and transverse resolution in this image is 1 μm and 5 μm, respectively (see Figure 4). Because the high transverse resolution reduced the depth-of-focus to 49 μm, separate OCT images were first acquired with the focus at different depths within the specimen. These images were then assembled to produce the composite image shown in Figure 8. This type of image construction is similar to C-mode ultrasound. Cellular features including cell membranes, nuclei, and nuclear morphology are clearly observed.

Medicine – Imaging Barrett's Esophagus

OCT performs *in situ* imaging and has the potential to be used as a screening technique of early pathological changes in patients. OCT imaging has the potential to be a general diagnostic in many organ systems,

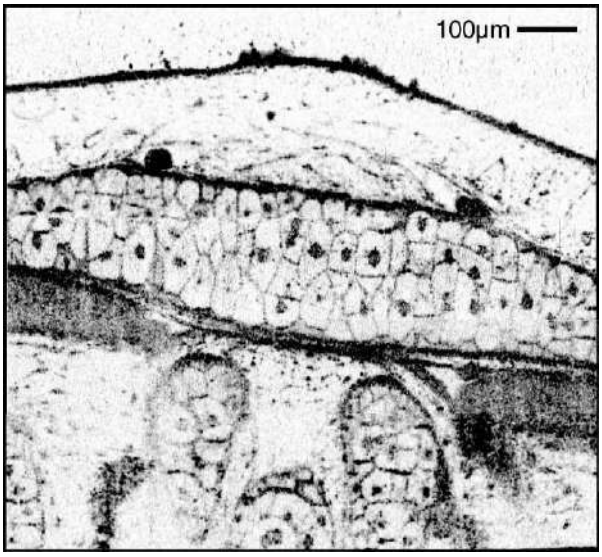

Figure 8 Cellular-level OCT imaging using a broad bandwidth titanium:sapphire laser. Axial and transverse image resolutions are 1 μm and 5 μm, respectively, enabling the visualization of cell membranes, nuclear morphology, and sub-cellular organelles. Cellular processes such as mitosis and migration can be visualized in real-time in living specimens. (Image reprinted with permission from Drexler W, Morgner U, Kartner FX, *et al.* (1999) *In vivo* ultrahigh resolution optical coherence tomography. *Optics Letters* 24: 1221–1223.)

particularly where current methods of screening by excisional biopsy are limited. Barrett's esophagus is a condition where the cells of the distal esophagus undergo a metaplastic change, resembling cells of the lower gastrointestinal tract. This condition is believed to be caused by chronic gastroesophageal reflux. Several studies have demonstrated that Barrett's esophagus is associated with a 30–125 times increased risk of developing adenocarcinoma. For this reason, endoscopic surveillance of Barrett's epithelium every 12–18 months is recommended. Endoscopic screening currently involves random four-quadrant biopsies every 1–2 cm along the length of suspect mucosa. However, excisional biopsy is prone to sampling errors and small foci of carcinoma or dysplasia may be missed. Because of the imprecision and high cost associated with screening, new methods are being developed to assess patients at increased risk. Endoscopic ultrasound catheters have been used for imaging the gastrointestinal tract with 50–100 μm resolution. However, these resolutions are insufficient to resolve early epithelial changes that occur in Barrett's esophagus and the pre-malignant changes that lead to adenocarcinoma of the esophagus.

The ability of OCT to differentiate normal and pathologic tissue is a central question being addressed by many research groups. Comparisons with histology show good correlations between OCT images and histological findings. The image resolution of conventional OCT (10–15 μm) is sufficient to differentiate architectural but not cellular morphology. Still, endoscopic OCT resolution can differentiate normal from Barrett's epithelium in real-time based on differences in epithelial architecture. Crypt- and gland-like structures that disrupt the relatively uniform layers of squamous epithelium can be readily identified, enabling differentiation between normal and Barrett's epithelium (Figure 9). The ability to differentiate normal from Barrett's epithelium suggests that the OCT technology could be used for screening applications.

Oncology – Identifying Tumors and Tumor Margins

OCT has been used to differentiate between the morphological structure of normal and cancerous tissue for a wide-range of tumors. The use of OCT to identify tumors and tumor margins *in situ* will represent a significant advancement for medical or image-guided surgical applications. OCT has been demonstrated for the detection of brain tumors and their margins with normal brain parenchyma, suggesting a role for guiding surgical resection. A hand-held surgical imaging probe was constructed for this application. The compact and portable probe permits OCT imaging within the surgical field while

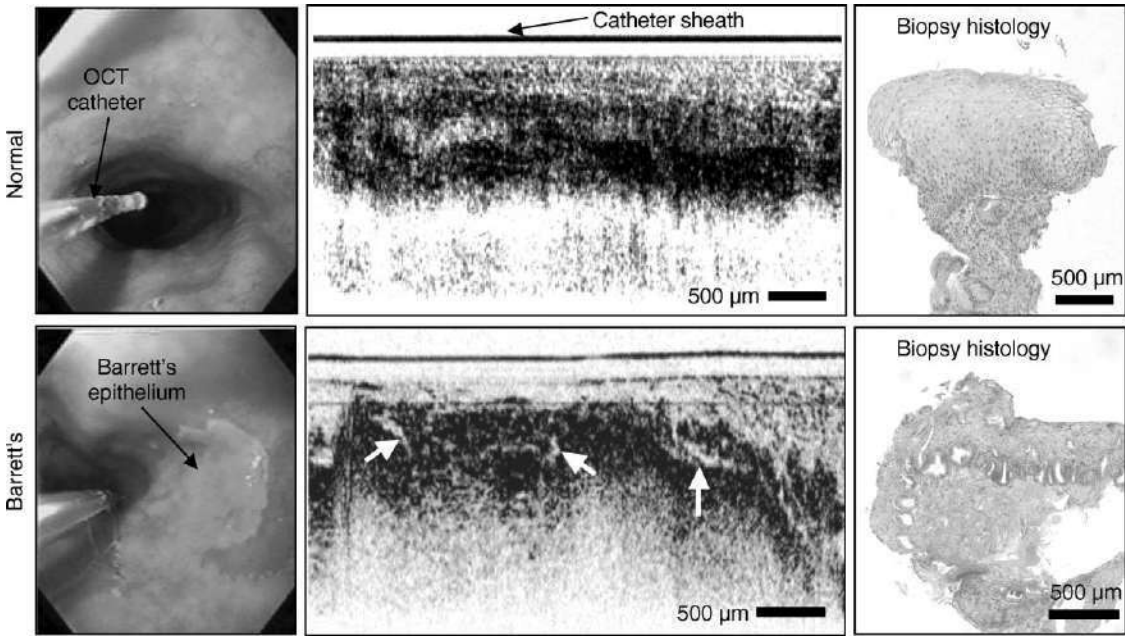

Figure 9 Endoscopic OCT imaging in humans. An OCT catheter was inserted down the working channel of an endoscope. Guided by video imaging, regions of normal and Barrett's esophagus were imaged. Corresponding biopsy histology is shown. OCT can identify Barrett's esophagus by the presence of gland- and crypt-like structures (arrows). (Images reprinted with permission from Li XD, Boppart SA, Van Dam J, *et al.* (2000) Optical coherence tomography: advanced technology for the endoscopic imaging of Barrett's esophagus. *Endoscopy* 32: 921–930.)

the OCT instrument can be remotely located in the surgical suite.

Figure 10 shows a specimen of outer human cerebral cortex with metastatic melanoma. The OCT images in Figure 10a,b were acquired through the tumor. These original images were threshold segmented to identify regions of high backscatter within the tumor. The original images were then overlaid with the segmented data and are shown in Figure 10c,d. The OCT images show increased optical backscattering in the region of the larger tumor (arrows). Smaller tumor lesions also appear within the image (arrows). A shadowing effect is observed below each tumor site due to the increased optical backscatter and the subsequent loss of optical power penetrating beneath the tumor. In Figure 10a,c, the boundary of the tumor can be identified. In Figure 10b,d, the tumor is identified below the surface of normal cortex. The histology in Figure 10e,f confirms the presence and relative size of the tumor.

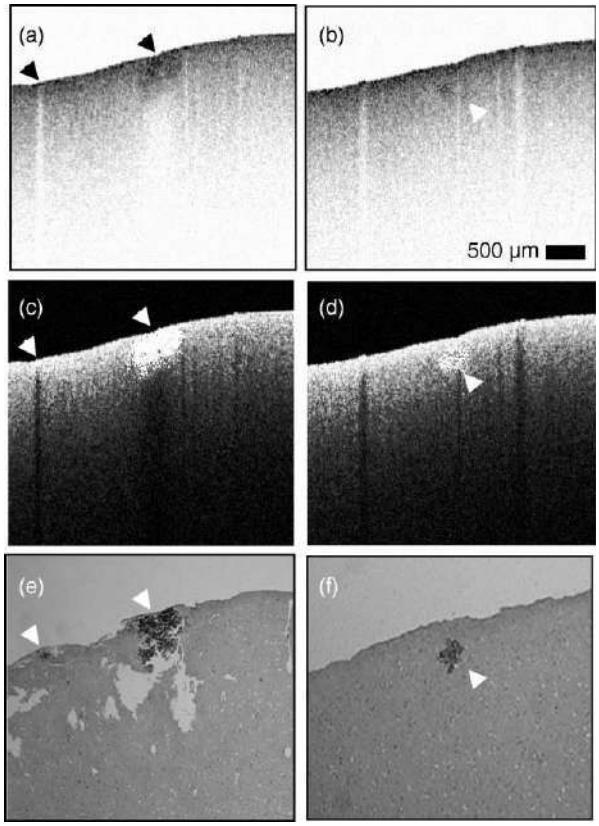

Figure 10 Cancer diagnostics. OCT imaging was used to image metastatic melanoma in an *ex vivo* specimen of human cortex (brain). Raw OCT images (a,b) were threshold-segmented (c,d) to highlight regions of increased optical backscatter and to identify tumor margins. OCT images correlate well with histological findings (e,f). (Images reprinted from Boppart SA, Brezinski ME, Pitris C and Fujimoto JG (1998) Optical coherence tomography for neurosurgical imaging of human intracortical melanoma. *Neurosurgery* 43: 834–841, with permission from Lippincott Williams and Wilkins.)

The image resolutions used to acquire the images in Figure 10 were as high as 16 μm, higher than any current ultrasound, CT, or MRI intraoperative imaging technique. This allowed the tumor-cortex interface and the extent of tumor below the surface to be defined with high resolution. At higher imaging resolutions, it may be possible to image individual tumor cells which have migrated away from the central tumor. OCT represents a new high-resolution optical imaging technology that has the potential for identifying tumors and tumor margins on the micron scale and in real-time. OCT offers imaging performance not achievable with current imaging modalities and may contribute significantly toward the surgical resection of neoplasms.

Image-Guided Surgery

The repair of vessels and nerves is necessary to restore function following traumatic injury. Although the repair of these sensitive structures is performed with the aid of surgical microscopes and loupes to magnify the surgical field, surgeons are limited to the *en face* view that they provide. A technique capable of subsurface, three-dimensional, micron-scale imaging in real-time would permit the intraoperative monitoring of microsurgical procedures. The capabilities of OCT for the intraoperative assessment of microsurgical procedures have been demonstrated. High-speed OCT imaging was integrated with a surgical microscope to performed micron-scale three-dimensional imaging on microsurgical specimens.

OCT has been used to image *in vitro* peripheral nerves and identify individual fascicles. Longitudinal tracking of the spatial orientation of rabbit peripheral nerve fascicles is demonstrated in Figure 11. Representative cross-sectional images of the peripheral nerve are shown in Figure 11a–d. For each slice, one fascicle was manually segmented, colored white, and tracked through the acquired volume of data. Forty images at 100 μm spacing were assembled for the 3-D projections shown in Figure 11e,f. The horizontally- and vertically-rotated projections of the peripheral nerve dramatically reveal the twisted path of the segmented fascicle along the longitudinal axis of the nerve. In addition, a branch in an unsegmented fascicle is observed in Figure 11f. The use of OCT to acquire multiple cross-sectional images and three-dimensionally reconstruct the peripheral nerve offers the opportunity to determine the relative diameters of individual fascicles and to longitudinally track their spatial orientation. OCT may also be useful at identifying and grading the degree of injury in nerves during surgical repair. These results have shown how 2-D OCT images and 3-D OCT projections can

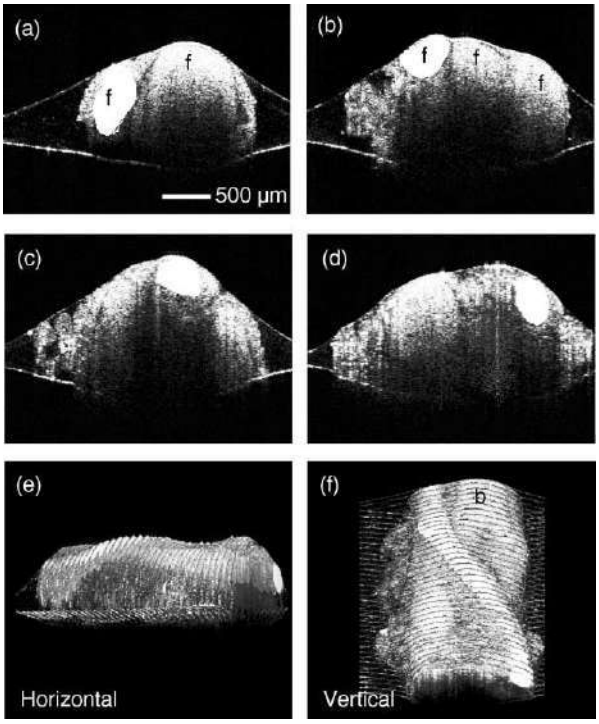

Figure 11 OCT image-guided microsurgery. Multiple cross-sectional OCT images were acquired through a resected rabbit peripheral nerve (a–d). A single nerve fascicle (f) was identified and segmented longitudinally along the length of the nerve. Three-dimensional reconstruction projections of the nerve in (e,f) reveal the tortuous path of the segmented fascicle as well as reveal a bifurcation (b). Intraoperative OCT can guide surgical repair and potentially identify regions of damaged tissue. (Images reprinted from Boppart SA, Bouma BE, Pitris C, *et al.* (1998) Intraoperative assessment of microsurgery with three-dimensional optical coherence tomography. *Radiology* 208: 81–86, with permission from the Radiology Society of North America.)

provide diagnostic feedback to assess microsurgical anastomoses. This previously unavailable diagnostic ability offers the potential to directly impact and improve patient outcome by incorporating high-speed, high-resolution intraoperative image-guidance during microsurgical procedures.

Surgical intervention requires visualization to identify tissue morphology, precision to avoid sensitive tissue structures, and continuous feedback to monitor the extent of the intervention. OCT may provide the technological advancements to improve the operative procedure. The feasibility of OCT to perform image-guided surgical intervention has been investigated. OCT has been used to monitor laser ablation therapy in real-time and may enable more precise control of laser delivery. An argon-laser ablation sequence of fresh *ex vivo* rat rectus abdominis muscle is shown in Figure 12. A pair of blood-filled vessels were located 1.5 mm below the tissue surface and centered within the OCT image. Because blood has a higher absorption coefficient than muscle at the 514 nm argon laser wavelength, coagulation and thermal heating of the blood occurred prior to thermal damage to the overlying tissue. The accompanying histology reveals both the thermally-damaged overlying tissue and coagulated blood present within the vessel pair.

These examples demonstrate the use of OCT for guiding and monitoring surgical intervention. The laser is only a representative interventional surgical technique for a wide range of instruments and techniques including scalpels, electrosurgery, radio-frequency, microwaves, and ultrasound ablation. OCT imaging was performed at 8 frames per second, fast enough to capture dynamic changes in the optical properties of the tissue during thermal ablation. These image sequences provided interesting insight into ablation mechanisms for a variety of tissue types. OCT can monitor the extent of thermal injury below the surface of the tissue by imaging the changes in optical backscatter. OCT imaging can therefore provide empiric information for dosimetry to minimize the extent of collateral injury. The use of OCT for guiding surgical interventions has the potential to improve intraoperative monitoring and more effectively control interventional procedures.

Materials

While the majority of OCT applications have been in the fields of biology and medicine. OCT has also been demonstrated in the non-biological areas of materials investigation, optical data storage, and microfluidic devices. The highly-scattering or reflecting optical properties of many materials prohibits deep imaging penetration using OCT. Many material defects, however, originate near the surface or at interfacial boundaries, making the use of OCT a possibility for inspection and quality control. OCT has been used to identify subsurface defects in ceramics and polymer composites. The optical ranging capabilities of OCT through scattering materials has been utilized for increasing the data storage capacity by assembling multiple layers of optically-accessible data.

The advancement of microfabrication techniques has led to increasingly complex microfluidic and bioMEM (biological micro-electro-mechanical) systems. Microstructures within microfluidic systems range from 10–1000 μm, within both the imaging depth and resolution of OCT. In addition, microfluidic systems are typically fabricated from transparent or semi-transparent substrates, facilitating imaging penetration to deeper three-dimensional features.

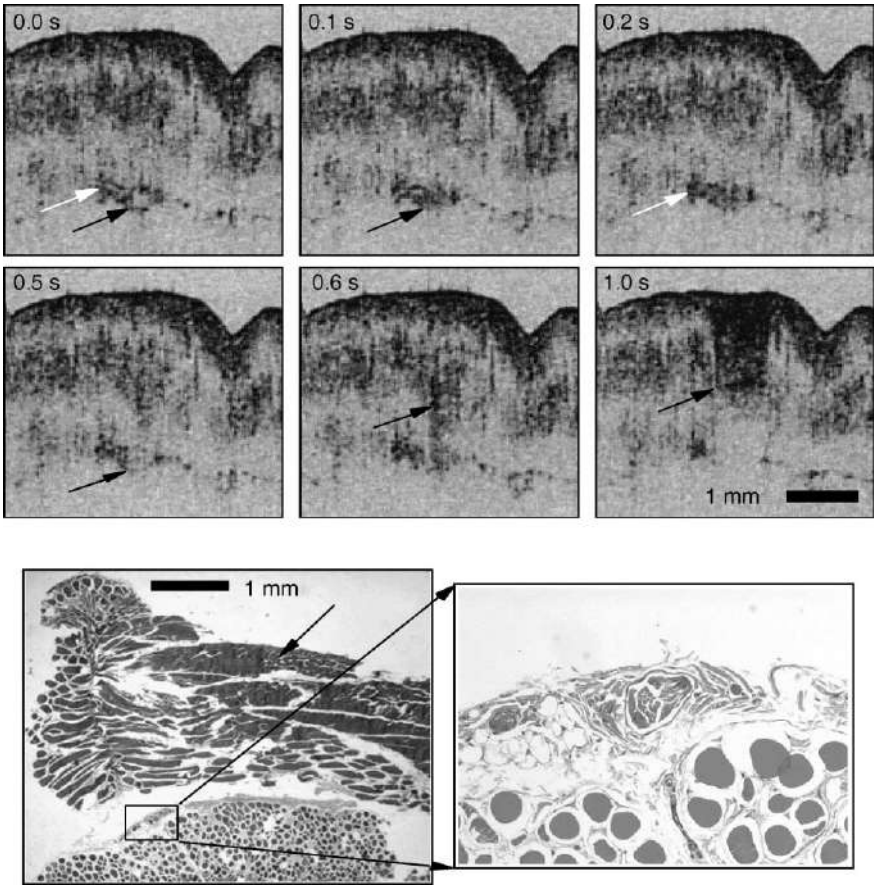

Figure 12 OCT image-guided laser ablation. High-speed OCT image sequence of argon-laser ablation of blood vessels in *ex vivo* rat muscle tissue. Thermal heating within the blood vessels (arrows, 0.0–0.5 s) induces coagulation prior to injury to superficial muscle tissue (0.6–1.0 s). Accompanying histology shows coagulated blood within the vessels.

OCT is not only capable of imaging three-dimensional microstructures at micron-scale resolutions, but is also capable of obtaining dynamic functional data from microfluidic systems. Microfluidic devices made from polymeric materials are attractive since they provide flexibility in the design of the channels, can be manufactured in large volumes, and are mostly bio-compatible. A three-dimensional imaging technique, such as OCT, will be instrumental in the analysis of these devices and the monitoring of device performance.

Conclusions

The capabilities of OCT offer a unique and informative means of imaging biological specimens and non-biological samples. The non-contact nature of OCT and the use of low-power near-infrared radiation for imaging causes few harmful effects on living cells or damage to materials. OCT imaging does not require the addition of fluorophores, dyes, or stains in order to improve contrast in images. Instead, OCT relies on the inherent optical contrast generated from variations in optical scattering and index of refraction. These factors permit the use of OCT for extended imaging over the course of hours, days, or weeks. OCT permits the cross-sectional imaging of tissue and samples and enables *in vivo* structure to be visualized in opaque specimens or in specimens too large for high-resolution confocal or light microscopy.

Imaging at cellular and subcellular resolutions with OCT is an important area of ongoing research. The *Xenopus* (African frog) developmental animal model has been commonly used because its care and handling are relatively simple while allowing cells with a high mitotic index to be assessed. Many of the cells observed were as large as 100 μm in diameter, but ranged in size down to a few microns, below the resolution of most OCT systems. Imaging human cells *in vivo* remains a challenge since differentiated human cells are 10 to 20 microns in size, too small for most OCT systems. With further advances in OCT technology, improved discrimination and

imaging of more detailed structures should be possible. New laser sources at other wavelengths in the near-infrared can enhance tissue contrast as well as potentially provide functional information since tissue scattering and absorbance properties in specimens are wavelength dependent. Short coherence length short pulse laser sources have been used to achieve higher axial resolutions on the order of 1 to 3 microns. Unfortunately, unlike superluminescent diode source, these high-speed and high-resolution systems utilize femtosecond lasers that are relatively complex and costly. Developing compact and portable optical sources at near-infrared wavelengths, broad spectral bandwidths, and high output powers is an area of active research.

Optical coherence tomography provides high-resolution morphological, functional, and cellular information of biological, medical, and material specimens and samples. OCT represents a multifunctional investigative tool that not only complements many of the existing imaging technologies available today, but also is poised to become a major optical imaging modality.

List of Units and Nomenclature

Cr^{4+}:Mg_2SiO_4	chromium:forsterite
Cr^{4+}:YAG	chromium:yttrium-aluminum-garnet
CT	computed tomography
d	beam diameter
f	focal length
MRI	magnetic resonance imaging
NEB	noise equivalent bandwidth
OCT	optical coherence tomography
P_{SAM}	power from sample
SNR	signal-to-noise ratio
SLD	superluminescent diode
Ti:Al_2O_3	titanium:sapphire
z_R	Rayleigh range
3-D	three-dimensional
Δx	transverse resolution
$\Delta\lambda$	spectral bandwidth
Δz	axial resolution
η	quantum efficiency
λ	center wavelength

See also

Coherence: Coherence and Imaging; Overview. **Coherent Lightwave Systems**. **Interferometry:** Overview. **Fiber and Guided Wave Optics:** Optical Fiber Cables. **Imaging:** Imaging Through Scattering Media; Interferometric Imaging. **Terahertz Technology:** Coherent Terahertz Sources. **Tomography:** Tomography and Optical Imaging. **Microscopy:** Confocal Microscopy; Imaging multiple Photon Fluorescence Microscopy; Overview.

Further Reading

Boppart SA, Tearney GJ, Bouma BE, *et al.* (1997) Noninvasive assessment of the developing *Xenopus* cardiovascular system using optical coherence tomography. *Proceedings of the National Academy of Sciences* 94: 4256–4261.

Boppart SA, Bouma BE, Pitris C, *et al.* (1998) *In vivo* cellular optical coherence tomography imaging. *Nature Medicine* 4: 861–864.

Boppart SA, Deutsch TF and Rattner DW (1999) Optical imaging technology in minimally invasive surgery. *Surgical Endoscopy* 13: 718–722.

Boppart SA, Brezinski ME and Fujimoto JG (2000) Optical Coherence Tomography Imaging. In: Tuan RS and Lo CW (eds) *Methods in Molecular Biology, Vol. 135: Developmental Biology Protocols*, vol. 1, pp. 217–233. Totowa, NJ: Humana Press, Inc.

Bouma BE and Tearney GJ (eds) (2001) *The Optical Coherence Tomography Handbook*. New York: Marcel Dekker Inc.

Chen Z, Zhao Y, Srinivas SM, *et al.* (1999) Optical Doppler tomography. *IEEE Journal of Selected Topics in Quantum Electronics* 5(4): 1134–1142.

Fujimoto JG, Brezinski ME, Tearney GJ, *et al.* (1995) Biomedical imaging and optical biopsy using optical coherence tomography. *Nature Medicine* 1: 970–972.

Fujimoto JG, Pitris C, Boppart SA and Brezinski ME (2000) Optical coherence tomography: An emerging technology for biomedical imaging and optical biopsy. *Neoplasia* 2(1–2): 9–25.

Huang D, Swanson EA, Lin CP, *et al.* (1991) Optical coherence tomography. *Science* 254: 1178–1181.

Masters BR (ed.) (2001) *Selected Papers on Optical Low-Coherence Reflectometry & Tomography.* SPIE Milestone Series, Vol. MS 165, Thompson BJ (general ed.) Bellingham, WA: SPIE Optical Engineering Press.

Puliafito CA, Hee MR, Schuman JS and Fujimoto JG (1995) Optical Coherence Tomography of Ocular Diseases. Thorofare, NJ: Slack, Inc.

Schmitt M (1999) Optical coherence tomography (OCT): A review. *IEEE Journal of Selected Topics in Quantum Electronics* 5(4): 1205–1215.

Tadrous PJ (2000) Methods for imaging the structure and function of living tissues and cells: 1. Optical coherence tomography. *Journal of Pathology* 191(2): 115–119.

Tearney GJ, Brezinski ME, Bouma BE, *et al.* (1997) *In vivo* endoscopic optical biopsy with optical coherence tomography. *Science* 276: 2037–2039.

Welzel J (2001) Optical coherence tomography in dermatology: a review. *Skin Research and Technology* 7(1): 1–9.

Tomography and Optical Imaging

Z Chen, University of California, Irvine, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

'Tomos' means section, in Greek. A tomogram is a sectional image of the internal structure of a three-dimensional object. One of the challenges for tomographic imaging of biological tissue using visible or near-infrared light is tissue optical scattering. When photons propagate inside a tissue, they undergo multiple scattering. The mean free path of photons in the skin, for example, is 5 μm. If one focuses the light 1 mm below the skin surface, most photons undergo multiple scattering and never reach the target point. Less than one out of a million photons will be able to reach the target, undergo a single backscattering, and return to the detector (Figure 1). Most photons that are backscattered and reach the detector are multiple scattered photons that carry little information on the target point. The challenge for tomographic imaging of tissues is to separate the single backscattered photons, that carry information about the target point, from multiple scattered photons.

Several optical tomography techniques have been developed. Each uses different techniques to discriminate the multiple scattering photons from the minimally scattered photons for imaging constructions. The confocal microscope, for example, uses a spatial filtering technique to reject multiple scattering photons because multiple scattered photons are more spread out in their angular distribution than the minimally scattered photons. Time-of-flight imaging uses a time gating technique to discriminate diffuse photons because single backscattered photons arrive earlier than diffuse photons. Optical coherence tomography (OCT) uses coherent gating to discriminate single scattered photons from multiple scattered photons. These gating techniques provide good discrimination of minimally scattered photons from multiple scattered photons and can form high-resolution sectional images, albeit with a limited imaging depth. Diffuse photons can propagate deep into the tissue and these photons can also be used for optical tomography. If a modulated photon density wave is propagated inside the tissue, the scattering and absorption properties of the tissue will modulate the phase and amplitude of the photon density wave. However, the resolution of optical tomography, based on diffuse photons, is usually much lower than that of the tomography constructed from single or minimally scattered photons. Three of the most important parameters for optical tomography are contrast, spatial resolution, and image depth.

Most optical tomography uses absorption, scattering, and fluorescent properties of materials as contrast. OCT, for example, uses variations in tissue scattering as contrast for tomographic imaging. However, in many instances, and especially during the early stages of disease, the change in tissue scattering properties between normal and diseased tissues is small and difficult to measure. One of the great challenges for extending applications of OCT is to find more contrast mechanisms that can provide physiological information in addition to morphological structure. This article will focus on optical tomography technologies that use nonconventional contrast mechanisms to obtain tomographic images. These include optical Doppler tomography (ODT), optical polarization tomography, and optical second-harmonic tomography. These optical tomography technologies are extensions of OCT with contrast mechanisms that can provide physiological information, in addition to morphological structure. ODT, also known as Doppler OCT, combines the Doppler principle with OCT to obtain high-resolution tomographic images of tissue structure and blood flow simultaneously. Optical polarization tomography, also known as polarization sensitive OCT (PS-OCT), combines polarization sensitive

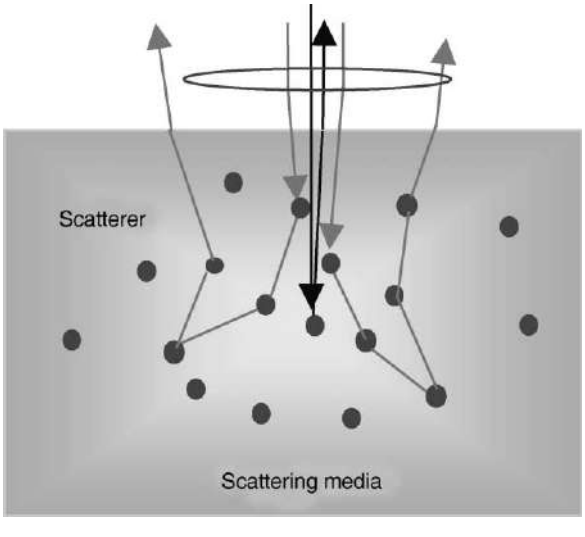

Figure 1 Optical imaging in scattering media.

detection with OCT, to determine tissue birefringence. Optical second-harmonic tomography, also known as second-harmonic optical coherence tomography (SH-OCT), combines second-harmonic generation (SHG) with coherence gating to obtain images with molecular contrast. These functional extensions of OCT provide clinically important information on tissue physiology, such as tissue blood perfusion, oxygen saturation, hemodynamics, and structural remodeling. Each provides several potential clinical applications, such as vasoactive drug screening, tissue viability and burn depth determination, tumor angiogenesis studies, tumor diagnosis, bleeding ulcer management, and ocular pathology evaluation.

Optical Doppler Tomography

ODT combines the Doppler principle with OCT to obtain high-resolution tomographic images of static and moving constituents in high scattering media. When light backscattered from a moving particle interferes with the reference beam, a Doppler frequency shift (f_D) occurs in the interference fringe:

$$f_D = \frac{1}{2\pi}(\mathbf{k}_s - \mathbf{k}_i)\cdot\mathbf{v} \qquad [1]$$

where $\mathbf{k}_i$ and $\mathbf{k}_s$ are wavevectors of incoming and scattered light, respectively, and $\mathbf{v}$ is the velocity vector of the moving particle (Figure 2). Since ODT measures the backscattered light, assuming the angle between flow and sampling beam is θ, the Doppler shift equation is simplified to:

$$f_D = \frac{2V\cos\theta}{\lambda_0} \qquad [2]$$

where λ_0 is the vacuum center wavelength of the light source.

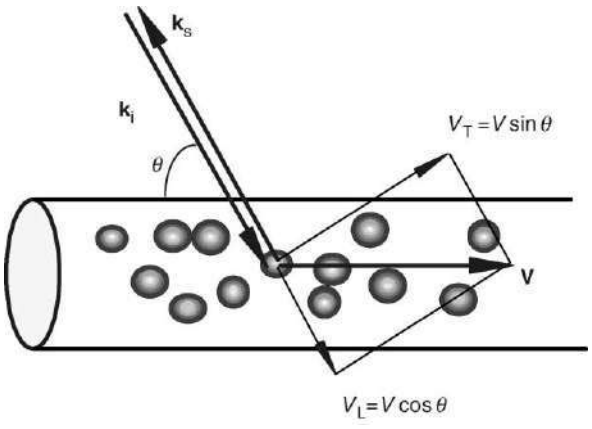

Figure 2 Schematic of flow direction and probe beam angle.

The optical system of ODT is similar to that of OCT. The primary difference is in signal processing. Figure 3 illustrates an ODT instrument that uses a fiber optic Michelson interferometer with a broadband light source. Light from a broadband partial coherence source is coupled into a fiber interferometer by a 2 × 2 fiber coupler and then split equally into reference and target arms of the interferometer. Light backscattered from the turbid sample is coupled back into the fiber and forms interference fringes with the light reflected from the reference arm. A rapid-scanning optical delay line is used for group delay and axial scanning. Because this delay line can decouple the group delay from the phase delay, an electro-optical phase modulator is introduced to produce a stable carrier frequency. The interference fringe intensity signal is amplified, bandpass filtered, and digitized with a high-speed analog-to-digital converter. The signal processing is carried out at the same time as data are transferred to the computer, and real-time display can be accomplished with the use of a digital signal processing board.

To understand the signal processing of ODT, let us look at the fringe signal due to the moving particles.

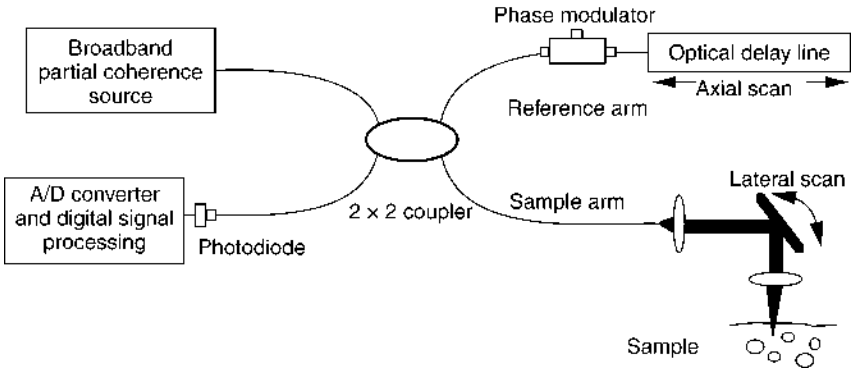

Figure 3 Schematic of ODT system consists of a fiber-based Michelson interferometer with a partially coherent light source.

If we denote $U(t)$ as a complex-valued analytic signal of a stochastic process representing the field amplitude emitted by a low coherent light source and $\bar{U}(\nu)$ as the corresponding spectral amplitude at optical frequency ν, the amplitude of a partially coherent source light coupled into the interferometer at time t, is written as a harmonic superposition:

$$U(t) = \int_0^\infty \bar{U}(\nu) e^{2\pi i \nu t} \, d\nu \qquad [3]$$

Because the stochastic process of a partially coherent light source is stationary, the cross spectral density of $\bar{U}(\nu)$ satisfies:

$$\langle \bar{U}^*(\nu)\bar{U}(\nu')\rangle = S_0(\nu)\delta(\nu - \nu') \qquad [4]$$

where $S_0(\nu)$ is the source power spectral density, and $\delta(\nu - \nu')$ is the Dirac delta function. Assuming that light couples equally into the reference arm and sample arm with spectral amplitude of $\bar{U}_0(\nu)$, the light coupled back to the detector from the reference, $\bar{U}_r(\nu)$, and sample, $\bar{U}_s(\nu)$, are:

$$\bar{U}_r(\nu) = e^{2\pi i \nu (2L_r + L_d)/c} K_r(\nu) e^{i\alpha_r} \bar{U}_0(\nu) \qquad [5]$$

$$\bar{U}_s(\nu) = e^{2\pi i \nu (2L_s + L_d)/c} K_s(\nu) e^{i\alpha_s} \bar{U}_0(\nu) \qquad [6]$$

where L_r and L_s are the optical pathlengths from the beamsplitter to the reference mirror and sample, respectively; L_d is the optical pathlength from the beamsplitter to the detector; and $K_r(\nu)e^{i\alpha_r(\nu)}$ and $K_s(\nu)e^{i\alpha_s(\nu)}$ are the amplitude reflection coefficients of light backscattered from the reference mirror and turbid sample, respectively.

The total power detected at the interferometer output is given by a time-average of the squared light amplitude:

$$P_d(\tau) = \langle |U_s(t) + U_r(t+\tau)|^2 \rangle \qquad [7]$$

where τ is the depth scan time delay. Combining harmonic expansions for $U_s(t)$ and $U_r(t)$ and applying eqn [4] when calculating the time-average, the total power detected is a sum of three terms representing reference (I_r), sample (I_s), and the interference fringe intensity ($\Gamma_{ODT}(\tau)$):

$$\begin{aligned} P_d(\tau) &= \int_0^\infty \left(P_r(\nu) + P_s(\nu) + P_{ODT}(\nu)\right) d\nu \\ &= I_r + I_s + \Gamma_{ODT}(\tau) \end{aligned} \qquad [8]$$

with

$$P_r(\nu) = S_0(\nu)|K_r(\nu)|^2 \qquad [9]$$

$$P_s(\nu) = S_0(\nu)|K_s(\nu)|^2 \qquad [10]$$

$$\begin{aligned} P_{ODT}(\nu) = {} & 2S_0(\nu)K_r(\nu)K_s(\nu) \\ & \times \cos\left[2\pi\nu(\Delta_d/c + \tau) + \alpha_s(\nu) - \alpha_r(\nu)\right] \end{aligned} \qquad [11]$$

and

$$I_r = \int_0^\infty P_r(\nu) d\nu \qquad [12]$$

$$I_s = \int_0^\infty P_s(\nu) d\nu \qquad [13]$$

$$\Gamma_{ODT}(\tau) = \int_0^\infty P_{ODT}(\nu) d\nu \qquad [14]$$

When there is a moving particle, light scattered from this particle is equivalent to a moving phase front; therefore, Δ_d in eqn [11] can be written as:

$$\Delta_d = \Delta + 2\bar{n}V_z t \qquad [15]$$

where Δ is the optical pathlength difference between light in the sampling and reference arms; V_z is the velocity of a moving particle parallel to the probe beam; and $\bar{n}$ is the refractive index of flow media.

To simplify the computation, we assume α_s and α_r are constants over the source spectrum and can be neglected. The spectral domain fringe signal, $P_{ODT}(\nu)$, is simplified to:

$$\begin{aligned} P_{ODT}(\nu) = {} & 2S_0(\nu)K_r(\nu)K_s(\nu) \\ & \times \cos\left[2\pi\nu\left((\Delta + 2\bar{n}V_z t)/c + \tau\right)\right] \end{aligned} \qquad [16]$$

The corresponding time domain signal, $\Gamma_{ODT}(\tau)$, is given by:

$$\begin{aligned} \Gamma_{ODT}(\tau) = {} & 2\int_0^\infty S_0(\nu)K_r(\nu)K_s(\nu) \\ & \times \cos\left[2\pi\nu\left((\Delta + 2\bar{n}V_z t)/c + \tau\right)\right] d\nu \end{aligned} \qquad [17]$$

A comparison of eqns [16] and [17] shows that there is a Fourier transformation relation between spectral domain and time domain signals. Consequently, there are two methods to acquire the ODT signal: the time domain method and the spectral domain method.

In the time domain method, a delay line is incorporated in the reference arm to generate a delay. A spectrogram analysis or phase-resolved algorithm is then used to determine the Doppler frequency shift. In the spectral domain method, the reference mirror is fixed, and there is no depth scan ($\tau = 0$). The spectral domain fringe signal, $P_{ODT}(\nu)$, is obtained either by a spectrometer at the detection

arm, or by a frequency sweeping light source. The time domain signal, $\Gamma_{\mathrm{ODT}}(\tau)$, is determined from the spectral domain signal by a Fourier transformation.

Spectrogram Method

The spectrogram method uses time–frequency analysis to calculate the Doppler shift. Signal processing algorithms to obtain structural and velocity images from recorded temporal interference fringe intensity using the spectrogram method, are illustrated in Figure 4.

The spectrogram is an estimate of the power spectrum of the temporal interference fringe intensity ($\Gamma_{\mathrm{ODT}}(\tau_i, f_j)$) in the ith time delay window. The power spectrum of the temporal interference fringe at the ith pixel, corresponding to time delay τ_{i} in the structural and velocity images, is calculated by a short-time fast Fourier transformation (STFFT) or a wavelet transformation:

$$\hat{\Gamma}_{\mathrm{ODT}}(\tau_i, f_m) = |\mathrm{STFFT}(f_m; \Gamma_{\mathrm{ODT}}(\tau))|^2 \qquad [18]$$

where f_m is the discrete frequency value. A tomographic structural image is obtained by calculating the value of the power spectrum at the phase modulation frequency (f_0). Fluid flow velocity is determined from the Doppler frequency shift (f_{D}), which is the difference between the carrier frequency established by the optical phase modulation (f_0) and the centroid (f_{c}) of the measured power spectrum at the ith pixel:

$$V_{\mathrm{ODT}}(i) = \frac{\lambda_0 f_{\mathrm{D}}}{2\bar{n}\cos\theta} = \frac{\lambda_0(f_{\mathrm{c}} - f_0)}{2\bar{n}\cos\theta} \qquad [19]$$

where we have assumed, $\mathbf{k}_{\mathrm{s}} = -\mathbf{k}_{\mathrm{i}}$ and θ is the angle between $\mathbf{k}_{\mathrm{i}}$ and $\mathbf{v}$. The centroid of the measured power spectrum is determined by:

$$f_{\mathrm{c}} = \sum_m f_m \hat{\Gamma}_{\mathrm{ODT}}(\tau_i, f_m) \Big/ \sum_m \hat{\Gamma}_{\mathrm{ODT}}(\tau_i, f_m) \qquad [20]$$

Figure 5 shows the *in vivo* structural and blood flow images from a chick chorioallantoic membrane. The lumen wall, chorion membrane, and yolk sac membrane are observed in structural image (a). In the velocity image (b), static regions ($V = 0$) appear dark, while blood flowing at different velocities appears as different brightnesses on the gray scale. The velocity profile taken from a horizontal cross-section passing through the center of the vessel is shown in (c).

Phase-Resolved ODT Method

Although spectrogram methods allow simultaneous imaging of *in vivo* tissue structure and flow velocity, the velocity sensitivity is limited for high-speed imaging. When STFFT or wavelet transformation is used to calculate flow velocity, the resolution is determined by the window size of the Fourier transformation for each pixel. The minimum detectable Doppler frequency shift, f_{D}, varies inversely with the STFFT window size. Because pixel acquisition time is proportional to the STFFT window size, the image frame rate is limited by velocity resolution. Furthermore, spatial resolution is also proportional to the STFFT window size. Therefore, a large STFFT window size increases velocity resolution while decreasing imaging speed and spatial resolution. This coupling between velocity sensitivity, spatial resolution, and imaging speed prevents the spectrogram method from achieving simultaneously both high imaging speed and high velocity sensitivity, which are essential for measuring flow in small blood vessels where flow velocity is low.

Phase-resolved ODT overcomes the compromise between velocity sensitivity and imaging speed by using the phase change between sequential scans to construct flow velocity images. The phase information of the fringe signal can be determined from the complex analytical signal $\tilde{\Gamma}(t)$, which is determined through analytic continuation of the measured interference fringe function, $\Gamma(t)$, using a Hilbert transformation:

$$\tilde{\Gamma}(t) = \Gamma(t) + \frac{i}{\pi} P \int_{-\infty}^{\infty} \frac{\Gamma(\tau)}{\tau - t}\, \mathrm{d}\tau = A(t)\mathrm{e}^{i\phi(t)} \qquad [21]$$

where P denotes the Cauchy principle value, i is the complex number, and $A(t)$ and $\phi(t)$ are amplitude and phase of $\tilde{\Gamma}(t)$, respectively. Because the interference signal $\Gamma(t)$ is quasi-monochromatic, the complex

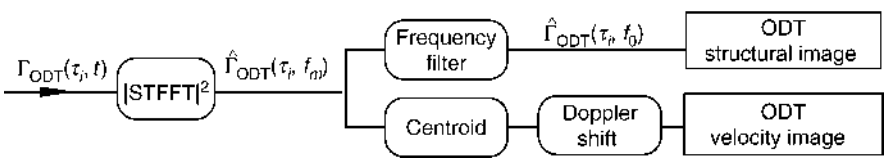

Figure 4 Signal processing algorithms for ODT structural and velocity images.

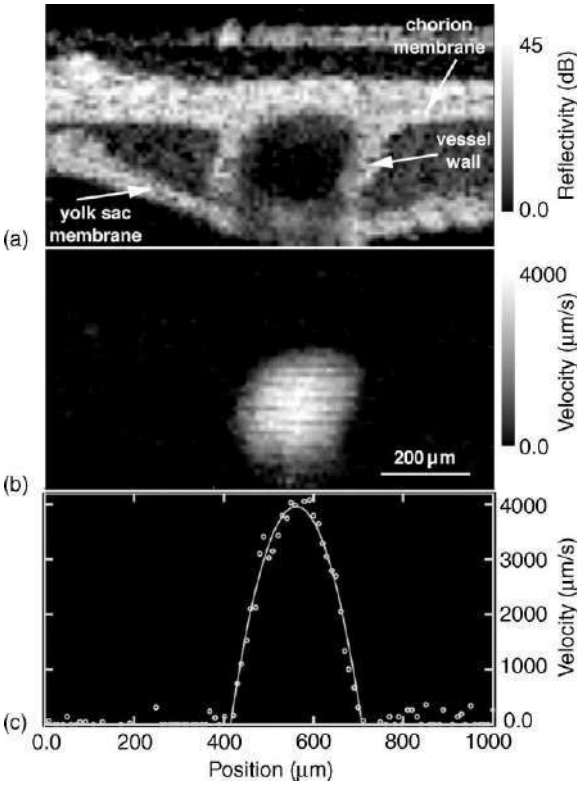

Figure 5 ODT images of *in vivo* blood flow in a CAM vein. (a) Structural image, (b) velocity image, and (c) velocity profile taken from a horizontal cross-section passing through the center of the vein, where the open circles are experimental data and the solid line is a parabolic fit. Reproduced with permission from Chen Z, MilnerTE, Srinivas S, Wang XJ, Malekafzali A, van Gemert MJC and Nelson JS (1997) Noninvasive imaging of *in vivo* blood flow velocity using optical doppler tomography. *Optics Letters* 22: 1119–1121.

analytical signal can be determined by:

$$\tilde{\Gamma}(t) = 2\int_0^{\infty}\int_0^{\tau}\Gamma(t')\exp(-2\pi i\nu t')\mathrm{d}t' \times \exp(2\pi i\nu t)\mathrm{d}\nu \qquad [22]$$

where τ is the time duration of the fringe signal in each axial scan.

The Doppler frequency shift (f_n) at nth pixel in the axial direction is determined from the average phase shift between sequential A-scans. This can be accomplished by calculating the phase change of sequential scans from the individual analytical fringe signal:

$$f_n = \frac{\Delta\phi}{2\pi T} = \frac{1}{2\pi T}\sum_{m=(n-1)M}^{nM}\sum_{j=1}^{N}\left[\tan^{-1}\left(\frac{\mathrm{Im}\tilde{\Gamma}_{j+1}(t_m)}{\mathrm{Re}\tilde{\Gamma}_{j+1}(t_m)}\right) - \tan^{-1}\left(\frac{\mathrm{Im}\tilde{\Gamma}_{j}(t_m)}{\mathrm{Re}\tilde{\Gamma}_{j}(t_m)}\right)\right] \qquad [23]$$

Alternatively, the phase change can also be calculated by the cross-correlation method:

$$f_n = \frac{1}{2\pi T}\tan^{-1}\left(\frac{\mathrm{Im}\left[\sum_{m=(n-1)M}^{nM}\sum_{j=1}^{N}\tilde{\Gamma}_j(t_m)\cdot\tilde{\Gamma}^*_{j+1}(t_m)\right]}{\mathrm{Re}\left[\sum_{m=(n-1)M}^{nM}\sum_{j=1}^{N}\tilde{\Gamma}_j(t_m)\cdot\tilde{\Gamma}^*_{j+1}(t_m)\right]}\right) \qquad [24]$$

where $\tilde{\Gamma}_j(t_m)$ and $\tilde{\Gamma}^*_j(t_m)$ are the complex signals at axial time t_m corresponding to the jth A-scan and its respective conjugate; $\tilde{\Gamma}_{j+1}(t_m)$ and $\tilde{\Gamma}^*_{j+1}(t_m)$ are the complex signals at axial time t_m corresponding to the next A-scan and its respective conjugate; M is an even number that denotes the window size in the axial direction for each pixel; N is the number of sequential scans used to calculate the cross correlation; and T is the time duration between A-scans. Because T is much longer than the pixel time window within each scan used in the spectrogram method, high-velocity sensitivity can be achieved.

Phase-resolved ODT decouples spatial resolution and velocity sensitivity in flow images and increases imaging speed by more than two orders of magnitude, without compromising spatial resolution and velocity sensitivity. In addition, because two sequential A-line scans are compared at the same location, speckle modulations in the fringe signals cancel each other and, therefore, will not affect the phase difference calculation. Consequently, the phase-resolved method reduces the speckle noise in the velocity image. Furthermore, if the phase difference between sequential frames is used, the velocity sensitivity can be further increased. Real-time imaging with velocity sensitivity on the order of 10 μm/s has been demonstrated.

In addition to digital processing of the fringe signal using Hilbert transformation, the complex analytical signal can also be achieved through hardware implementation. Optical Hilbert transformation using polarization optics has been implemented for real-time phase-resolved ODT imaging. Real-time ODT imaging using hardware demodulation of the ODT signal has also been demonstrated.

Spectral Domain Phase-Resolved ODT Method

A schematic diagram of a spectral domain ODT system is shown in Figure 6. The signal from the Michelson interferometer is directly coupled to a

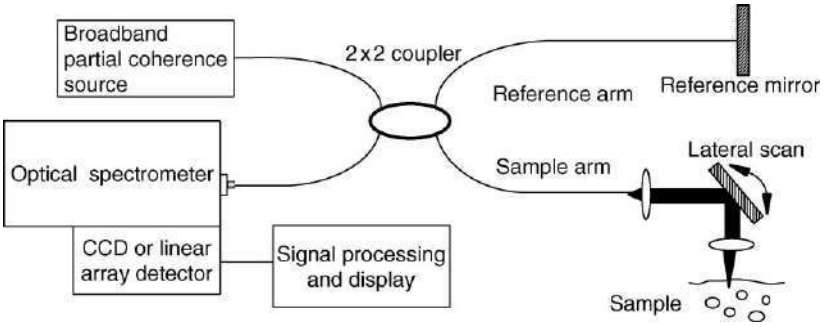

Figure 6 Schematic diagram of a spectral domain ODT instrument.

spectrometer that records the spectral fringe pattern, $P_{\mathrm{ODT}}(\nu)$. The temporal interference fringe can be calculated by a Fourier transformation of the spectral fringe pattern (eqn [17]). The Doppler shift can then be determined from the phase shift between sequential scans using the phase-resolved ODT algorithm (eqns [23] and [24]).

Spectral domain phase-resolved ODT has the advantage that no optical delay line is required. In addition, the spectral domain method can achieve high signal to noise ratio over time domain ODT. Since parallel detection can be implemented with a high-speed spectrometer, high-speed imaging acquisition is possible. Finally, because the dynamic range of the phase-resolved ODT depends on the speed of the line scans, spectral domain ODT has the advantage in terms of imaging speed and velocity dynamic range.

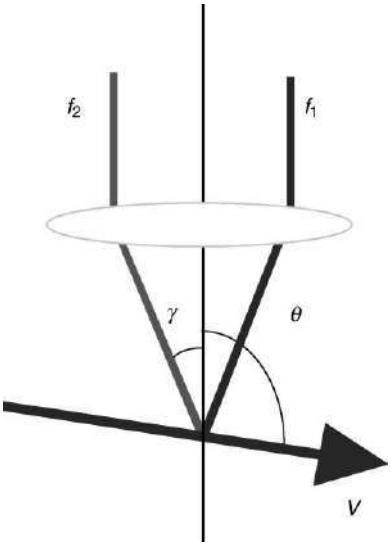

Figure 7 Effect of numerical aperture and transverse flow velocity on Doppler bandwidth.

Transverse Flow Velocity and Doppler Angle Determination

One of the limitations of using the Doppler shift to determine the flow is that the technique is only sensitive to longitudinal flow velocity (Figure 2). If one knows the flow direction, Doppler shift measurement can fully quantify the flow. However, in many biological cases, where flow direction is not known, Doppler shift measurement alone is not enough to fully quantify the flow. Furthermore, there are many clinical cases, such as ocular bloodflow, where vessels are in the plane perpendicular to the probing beam. When flow direction is perpendicular to the probing beam, the Doppler shift is not sensitive to transverse bloodflow (V_{T}). Therefore, a method to measure transverse flow velocity is essential.

Standard deviation of the Doppler spectra can be used to determine the transverse flow. The technique is based on the fact that ODT imaging uses a relatively large numeric aperture lens in the sampling arm. The beam from different sides of the edges will produce different Doppler shifts, f_1 and f_2, as indicated in Figure 7. Consequently, the Doppler spectra will be broadened by the transverse flow. If the incident beam has a Gaussian spectral profile and contributions from Brownian motion and other sources that are independent of the macroscopic flow velocity are included, a linear relation between standard deviation of the Doppler spectra and transverse flow velocity, $V_{\mathrm{T}} = V \sin\theta$, can be derived:

$$\sigma = \frac{\pi V \sin\theta NA_{\mathrm{eff}}}{8\lambda} + b \qquad [25]$$

where b is a constant, and NA_{eff} is the effective numeric aperture. The standard deviation can be determined from the measured analytical fringe

signal:

$$\sigma^2 = \frac{1}{2\pi^2 T^2} \times \left(1 - \frac{\left| \sum_{m=(n-1)M}^{nM} \sum_{j=1}^{N} \tilde{\Gamma}_j(t_m) \cdot \tilde{\Gamma}^*_{j+1}(t_m) \right|}{\frac{1}{2} \sum_{m=(n-1)M}^{nM} \sum_{j=1}^{N} [\tilde{\Gamma}_j(t_m) \cdot \tilde{\Gamma}^*_j(t_m) + \tilde{\Gamma}_{j+1}(t_m) \cdot \tilde{\Gamma}^*_{j+1}(t_m)]} \right) \quad [26]$$

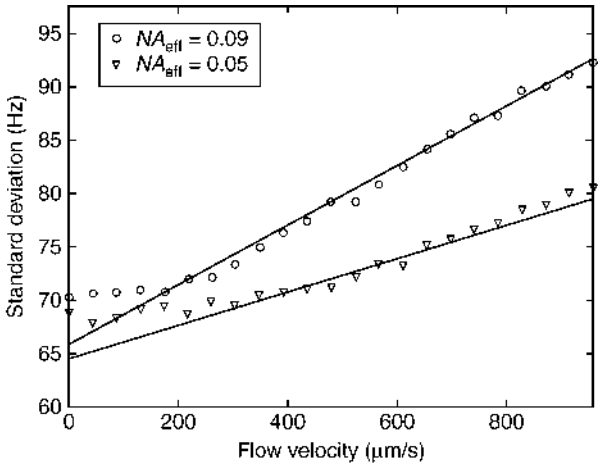

Figure 8 Standard deviation as a function of flow velocity for two different numeric apertures. Reproduced with permission from Ren *et al.* (2002) *Optics Letters* 27: 409–411.

The measured standard deviation as a function of transverse flow velocity is shown in Figure 8. As predicted, the Doppler bandwidth is a linear function of transverse flow velocity above a certain threshold level. The effective numerical aperture of the optical objective in the sample arm determines the slope of this dependence. This result indicates that standard deviation can be used to determine the transverse flow velocity. Since both longitudinal and transverse flow velocity (V_L and V_T) can be measured by the Doppler shift and standard deviation, respectively, flow direction can be determined from a single measurement of the Doppler fringe signal.

Applications

Due to its exceptionally high spatial resolution and velocity sensitivity, several clinical applications of ODT have been demonstrated, including screening vasoactive drugs, monitoring changes in image tissue morphology and hemodynamics following pharmacological intervention and photodynamic therapy, evaluating the efficacy of laser treatment in port wine stain (PWS) patients, assessing the depth of burn wounds, mapping cortical hemodynamics for brain research, imaging ocular bloodflow, and

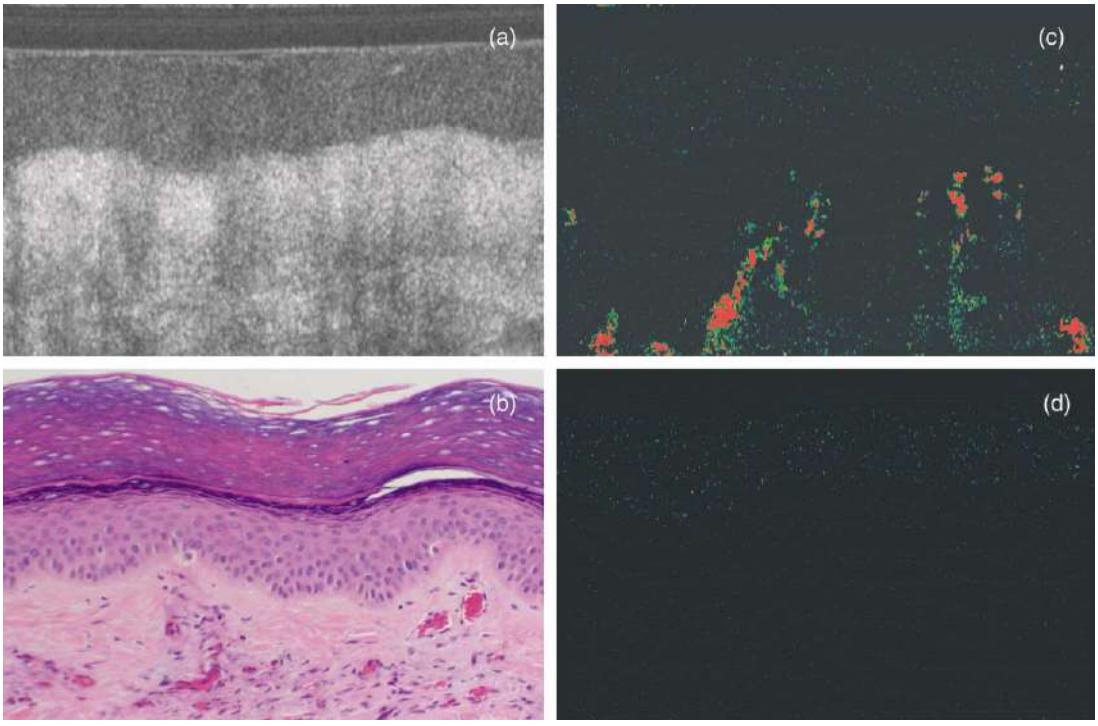

Figure 9 ODT images taken *in situ* from PWS human skin. (a) Structural image, (b) histological section, (c) image before laser treatment, and (d) image after laser treatment (scale bar 100 μm). Reproduced with permission from Nelson *et al.* (2001) *Archives of Dermatology* 137: 741–744. Copyright © 2001, American Medical Association. All rights reserved.

mapping bloodflow in gastrointestinal tracts. Figure 9 shows ODT structural and flow velocity images of a patient with PWS before and after laser treatment, respectively. For comparison, a histology picture taken at the same site is included. The vessel location from the ODT measurement and histology agree very well. Furthermore, the destruction of the vessel by laser can be identified, since no flow appears on the Doppler flow image after laser treatment.

Three-dimensional images of a microvascular network can be obtained by stacking 2D scans together. It is known that the microvasculature of mammary tumors has several distinct differences from normal tissues. Mapping of a 3D cancer microvascular network could provide additional information for cancer diagnosis. Figure 10 shows multiple blood vessels imaged from a patient with an abnormal microvascular network. Different colors represent different signs of the Doppler shift, which depends on the angle between the direction of flow and probing beam.

ODT is also a powerful tool to study flow dynamics in microfluidic channels. In addition to imaging flow velocity profile, ODT can also be used to measure osmotic mobility, quantify size of the scattering particle, and study flow dynamics of microfluids in microchannels of different materials, geometry, and surface treatment.

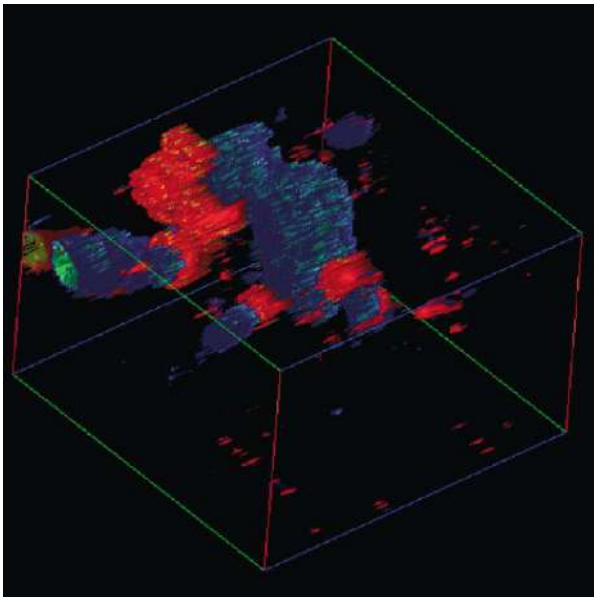

Figure 10 Three-dimensional ODT images of multiple blood vessels in human skin from a patient with a PWS birthmark. Reproduced with permission from Zhao *et al.* (2001) *IEEE J. Select Topics Quantum Electro.* 7: 931–935. © 2001, IEEE.

Polarization Sensitive OCT

PS-OCT combines polarization sensitive detection with OCT to determine tissue birefringence. PS-OCT can obtain enhanced image contrast and additional physiological information by studying polarization properties of biological tissues. Many biological tissues, such as tendon, muscle, nerve, bone, cartilage, and skin, exhibit birefringence properties. Consequently, PS-OCT can provide contrast mechanisms that are directly related to the physiological conditions of these tissues.

The polarization state of a light beam can be described by a Stokes vector **S** as:

$$\mathbf{S} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \begin{bmatrix} a_x^2 + a_y^2 \\ a_x^2 - a_y^2 \\ 2a_x a_y \cos\varphi \\ 2a_x a_y \sin\varphi \end{bmatrix} \qquad [27]$$

where I, Q, U, V are Stokes parameters, a_x and a_y are amplitudes of two orthogonal components of the electric vector, and φ represents the phase difference between the two components.

The effect of an optical device on the polarization of light can be characterized by a 4×4 Mueller matrix. The matrix acts on the input state, $\mathbf{S}_1$, to give the output state, $\mathbf{S}_2$:

$$\mathbf{S}_2 = \begin{bmatrix} I_2 \\ Q_2 \\ U_2 \\ V_2 \end{bmatrix} = \mathbf{M}\mathbf{S}_1$$

$$= \begin{bmatrix} M_{00} & M_{01} & M_{02} & M_{03} \\ M_{10} & M_{11} & M_{12} & M_{13} \\ M_{20} & M_{21} & M_{22} & M_{23} \\ M_{30} & M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} I_1 \\ Q_1 \\ U_1 \\ V_1 \end{bmatrix} \qquad [28]$$

In the Poincaré sphere representation, a polarization state, **S**, can be represented by a point (Q, U, V). The Poincaré sphere representation provides a convenient method to evaluate changes of Stokes vectors. Determination of either the Stokes vector or the Mueller matrix allows full quantification of birefringence properties of the biological sample.

A fiber-based, high-speed PS-OCT system is shown in Figure 11. The system is similar to the ODT system, except a polarization modulator and two orthogonal detection channels are included. Due to the coherence detection intrinsic to OCT, backscattered light that

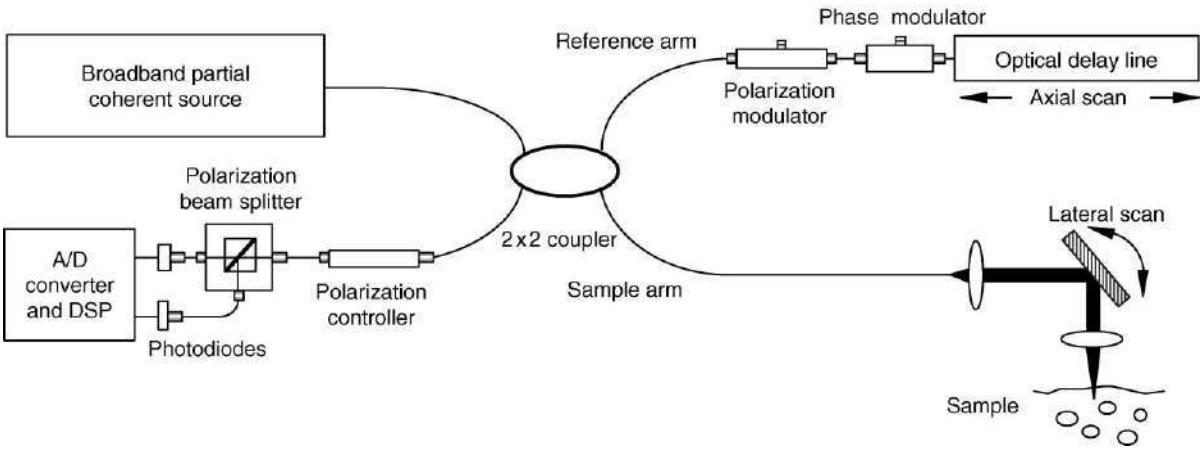

Figure 11 Fiber-based PS-OCT system. DSP: digital signal processing.

originates from incident light in different polarization states can be determined by controlling polarization states in the reference arm. Modulating the reference light in four orthogonal polarization states allows coherent detection of backscattered light from the sample under equivalent illumination in four different polarization states. The choice of orthogonal polarization states in the Poincaré sphere is important because it ensures that the birefringence measurements will be independent of the orientation of the optical axis in the sample.

For every polarization state controlled by the polarization modulator, the A-scan signals corresponding to the two orthogonal polarization diversity channels are digitized. The coherence matrix can be calculated from the complex electrical field vector from these two channels:

$$J = \begin{bmatrix} \langle E_{\mathrm{H}}^{*}(t)E_{\mathrm{H}}(t)\rangle & \langle E_{\mathrm{H}}^{*}(t)E_{\mathrm{V}}(t)\rangle \\ \langle E_{\mathrm{V}}^{*}(t)E_{\mathrm{H}}(t)\rangle & \langle E_{\mathrm{V}}^{*}(t)E_{\mathrm{V}}(t)\rangle \end{bmatrix} = \begin{bmatrix} J_{\mathrm{HH}} & J_{\mathrm{HV}} \\ J_{\mathrm{VH}} & J_{\mathrm{VV}} \end{bmatrix} \qquad [29]$$

where E_{H}, E_{V} are the components of the complex electric field vector corresponding to horizontal and vertical polarization channels, respectively, and E_{H}^{*}, E_{V}^{*} are their conjugates, respectively. The Stokes vector can be derived from the coherence matrix:

$$S_0 = J_{\mathrm{HH}} + J_{\mathrm{VV}} \qquad S_2 = S_0^{-1}(J_{\mathrm{HV}} + J_{\mathrm{VH}})$$
$$S_1 = S_0^{-1}(J_{\mathrm{HH}} - J_{\mathrm{VV}}) \qquad S_3 = iS_0^{-1}(J_{\mathrm{VH}} - J_{\mathrm{HV}}) \qquad [30]$$

where S_1, S_2, and S_3 are the normalized coordinates of the Stokes vector in the Poincaré sphere, characterizing the polarization state of the backscattered light, and S_0 is the module of the Stokes vector characterizing light intensity. The Stokes vector, corresponding to each of the input polarization states for the nth pixel in the jth A-scan, can be calculated as:

$$S_{0,n} = \sum_{m=(n-1)M}^{nM} [\tilde{\Gamma}_j^{\mathrm{H}}(t_m)\tilde{\Gamma}_j^{\mathrm{H}^*}(t_m) + \tilde{\Gamma}_j^{\mathrm{V}}(t_m)\tilde{\Gamma}_j^{\mathrm{V}^*}(t_m)]$$

$$S_{1,n} = S_{0,n}^{-1} \sum_{m=(n-1)M}^{nM} [\tilde{\Gamma}_j^{\mathrm{H}}(t_m)\tilde{\Gamma}_j^{\mathrm{H}^*}(t_m) - \tilde{\Gamma}_j^{\mathrm{V}}(t_m)\tilde{\Gamma}_j^{\mathrm{V}^*}(t_m)]$$

$$S_{2,n} = S_{0,n}^{-1} \sum_{m=(n-1)M}^{nM} 2\,\mathrm{Re}(\tilde{\Gamma}_j^{\mathrm{H}^*}(t_m)\tilde{\Gamma}_j^{\mathrm{V}}(t_m))$$

$$S_{3,n} = S_{0,n}^{-1} \sum_{m=(n-1)M}^{nM} 2\,\mathrm{Im}(\tilde{\Gamma}_j^{\mathrm{H}^*}(t_m)\tilde{\Gamma}_j^{\mathrm{V}}(t_m)) \qquad [31]$$

where M is an integer number that determines the size of the average in the axial direction for each pixel; $\tilde{\Gamma}_j^{\mathrm{H}}(t_m)$ and $\tilde{\Gamma}_j^{\mathrm{V}}(t_m)$ are complex signals detected from the two orthogonal polarization channels at axial time t_m, for the jth A-scan; and $\tilde{\Gamma}_j^{\mathrm{H}^*}(t_m)$ and $\tilde{\Gamma}_j^{\mathrm{V}^*}(t_m)$ are their conjugates, respectively.

In order to measure the birefringence properties of the sample accurately, four states of light polarization are generated for each lateral pixel. For each polarization state, one A-line scan is performed. Therefore, a total of four A-line scans are used to calculate the Stokes vector and phase retardation images simultaneously.

For a sample with an assumed linear birefringence, there exists two eigenwaves that are polarized along the projected fast and slow axes of the sample normal to the propagation direction of incident light. The Stokes vectors of these eigenwaves determine a rotation axis in the equator plane of a Poincaré sphere.

The effect of birefringence is to rotate the Stokes vector about this axis through an angle that is equal to the phase retardation of the sample. Conversely, the rotation axis can be determined from the known polarization states of incident and backscattered light at the sample location. The phase retardation image, which characterizes the accumulated birefringence distribution in the sample, is calculated by the rotation of the Stokes vectors in the Poincaré sphere.

The simultaneous Stokes vectors and phase retardation images of a rat muscle are shown in Figure 12.

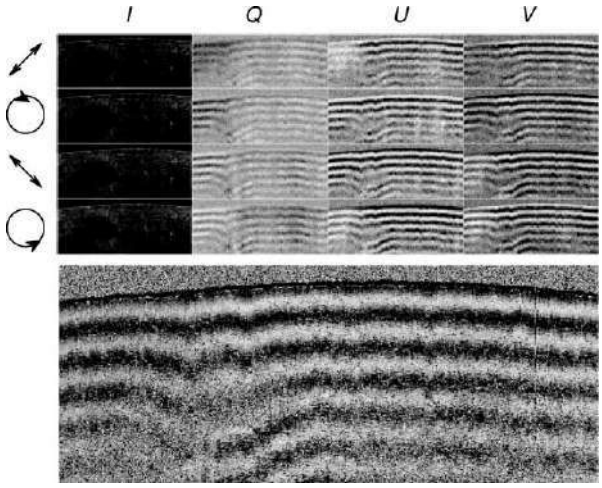

Figure 12 Stokes vectors images (top) and phase retardation image (bottom) in fresh rat muscle. The image area is 2 mm × 5 mm. Reproduced with permission from Ren *et al.* (2003) *SPIE Proceedings* 4956: 320–327.

Stokes vectors, corresponding to four different polarization states, are shown in the top figure. From Stokes vector images, the phase retardation image can be calculated as shown in the bottom figure. The banded structure in the phase retardation image indicates the accumulated tissue birefringence distribution in the rat muscle. In addition, differential optical axis and birefringence images can also be calculated.

There are a number of clinical applications for PS-OCT, including burn depth determination and retinal fiber nerve layer thickness evaluation. Figure 13 shows *in vivo* structural and phase retardation images from rat skin burned at 100°C for 10 seconds. Although the structure image shows very little contrast, the phase retardation image clearly shows the contrast between the normal and burned regions. As light propagates into the tissue, the retardation increases due to tissue birefringence. For normal tissue, it only takes less than 20 μm to cause a phase retardation of 90°. However, for the burned tissue, it takes much thicker tissue to cause the same phase retardation. The decrease of birefringence in thermally damaged tissue can clearly be seen in PS-OCT scans in comparison to normal rat skin.

Second Harmonic OCT

SH-OCT combines SHG with coherence gating for high-resolution tomographic imaging of tissue structures with molecular contrast. SHG is the lowest-order nonlinear optical process where the second-order nonlinear optical susceptibility is

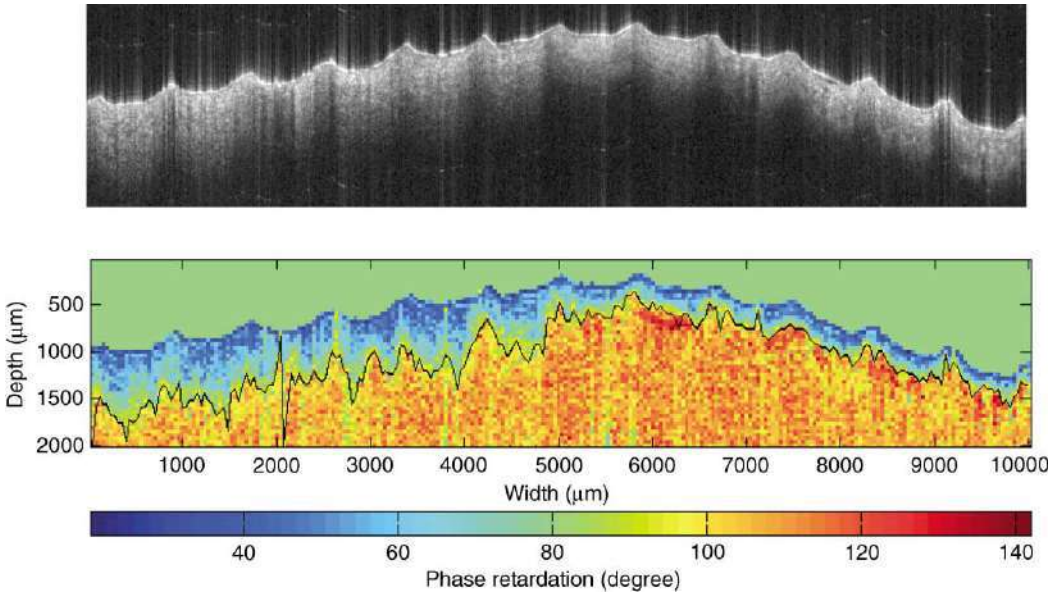

Figure 13 *In vivo* structure (top) and phase retardation image (bottom) from rat skin burned at 100°C for 10 seconds. The black contour line in the phase retardation image demarcates the depth at which 90° phase retardation has been reached with respect to the incident polarization (image size 2 mm × 10 mm).

responsible for the generation of light at second-harmonic frequency. Because second-order nonlinear optical susceptibility is very sensitive to electronic configurations, molecular structure and symmetry, local morphology, and ultrastructures, SHG provides molecular contrast for the coherence image.

The schematic diagram for SH-OCT is shown in Figure 14. A broadband short pulse laser is used as a light source in a Michelson interferometer. The light is split into the reference arm and the sample arm. In the reference arm, a thin nonlinear crystal is used to convert the input radiation to SH photons. Both SH and fundamental waves are then reflected by a metal mirror (M1) mounted on a motorized translation stage, which acts as the delay line in this SH-OCT system. A backscattered SH wave from the sample recombines with the SH wave from the reference arms to form interference fringe. The SH interference fringe signal is detected by a photomultiplier tube (PMT) after passing through a short-pass filter (F1). The fundamental interference fringe signal is detected by a photodiode (PD) after passing through a long-pass filter (F2).

A SH-OCT image of a rat tendon is shown in Figure 15. Compared with conventional OCT performed at fundamental wavelengths, SH-OCT offers enhanced molecular contrast and spatial resolution. It is also an improvement over existing SHG scanning microscopy technology as the intrinsic coherence gating mechanism enables the detection and discrimination of SH signals generated at deeper locations. The enhanced molecular contrast of SH-OCT extends conventional OCT's capability for detecting small changes in molecular structure. SH-OCT is promising for the diagnosis of cancers and other diseases at an early stage when changes in tissue and molecular structure are small.

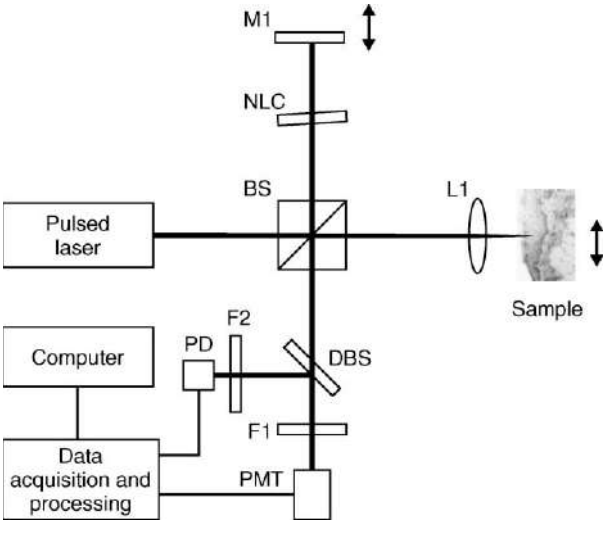

Figure 14 Schematic of experiment set-up for SH-OCT. M1, mirror; NLC, nonlinear crystal; BS, broadband nonpolarization beamsplitter; DBS, dichroic beamsplitter; L1, lens; F1–F2, filters; PD, photo diode; PMT, photomultiplier tube.

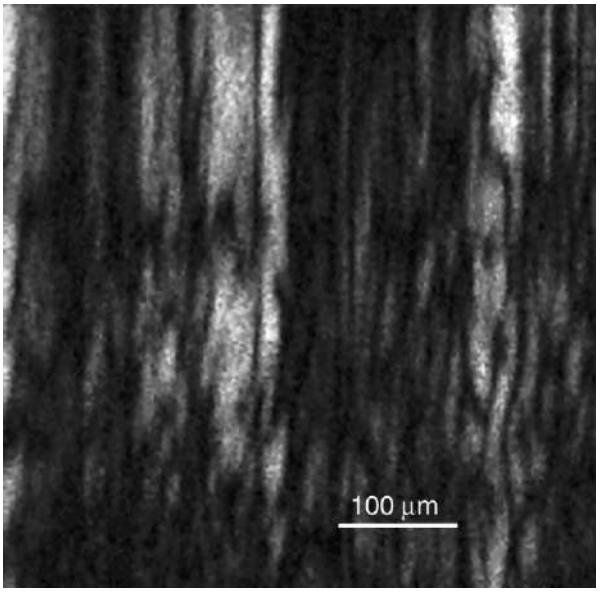

Figure 15 SH-OCT image of a rat tendon; image size 250 μm × 250 μm. Scale bar 100 μm. Image reprinted with permission from Jiang Y *et al.* (2004) Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Conclusions

Optical tomography is a rapidly developing imaging technology with many potential applications. Through different contrast enhancement mechanisms, ODT, PS-OCT, and SH-OCT provide clinically important physiological information that is not available in the structure image. Given the noninvasive nature and exceptionally high spatial resolution, optical tomography can simultaneously provide tissue structure, blood perfusion, birefringence, and other physiological information and have great potential for basic biomedical research and clinical medicine.

Acknowledgments

I would like to thank many of my colleagues who have contributed to the functional OCT project at the Beckman Laser Institute and Center for Biomedical Engineering at UCI, particularly the students and postdoctoral fellows. I also want to acknowledge the research grants awarded from the National Institutes of Health (EB-00293, NCI-91717, RR-01192 and EB-00255), the National Science Foundation (BES-86924), the Whitaker

Foundation (WF-23281) and the Defense Advanced Research Program Agency (Bioflip program). Institutional support from the Air Force Office of Scientific Research (F49620-00-1-0371), and the Beckman Laser Institute Endowment are also gratefully acknowledged. Dr Chen's e-mail address is zchen@bli.uci.edu.

List of Units and Nomenclature

OCT	Optical coherence tomography
ODT	Optical Doppler tomography
PS-OCT	Polarization sensitive OCT
SH-OCT	Second harmonic OCT
STFFT	Short time fast Fourier transformation

See also

Coherence: Coherence and Imaging; Overview. **Fiber and Guided Wave Optics:** Nonlinear Optics; Optical Fiber Cables. **Imaging:** Imaging Through Scattering Media; Interferometric Imaging; Photon Density Wave Imaging. **Interferometry:** Overview. **Microscopy:** Confocal Microscopy; Imaging Multiple Photon Fluorescence Microscopy. **Polarization:** Introduction; Matrix Analysis. **Tomography:** Optical Coherence Tomography; Tomography and Optical Imaging.

Further Reading

Bouma BE and Tearney GJ (eds) (2001) *The Optical Coherence Tomography Handbook*. New York: Marcel Dekker, Inc.

Chen Z (2003) Functional optical coherence tomograpy. In: Hwang NHC and Woo LY (eds) *Frontiers in Biomedical Engineering*, pp. 345–364. New York: Kluwer Academic Publishers.

Chen Z, Milner TE, Srinivas S, Wang XJ, Malekafzali A, van Gemert MJC and Nelson JS (1997) Noninvasive imaging of *in vivo* blood flow velocity using optical doppler tomography. *Optics Letters* 22: 1119–1121.

Chen Z, Zhao Y, Srinivas SM, Nelson S, Prakash N and Frostig RD (1999) Optical Doppler tomography. *IEEE Journal of Selected Topics in Quantum Electronics* 5(4): 1134–1142.

de Boer JF and Milner TE (2002) Review of polarization sensitive optical coherence tomography and Stokes vector determination. *Journal of Biomedical Optics* 7: 359–371.

Fercher AF and Hizenberger CK (2002) Optical coherence tomography. *Progress in Optics* 44: 215–296.

Huang D, Swanson EA, Lin CP, *et al.* (1991) Optical coherence tomography. *Science* 254: 1178–1181.

Izatt JA, Kulkarni MD, Yazdanfar S, Barton JK and Welch AJ (1997) *In vivo* bidirectional color Doppler flow imaging of picoliter blood volumes using optical coherence tomography. *Optics Letters* 22: 1439–1441.

Jiang Y, Tomov I, Wang Y and Chen Z (2004) Second harmonic optical coherence tomography. *Optics Letters* 29: 1090–1092.

Jiao S and Wang LV (2002) Jones-matrix imaging of biological tissues with quadruple-channel optical coherence tomography. *Journal of Biomedical Optics* 7: 350–358.

Masters BR and Thompson BJ (eds) (2001) *Selected Papers on Optical Low-Coherence Reflectometry & Tomography*. SPIE Milestone Series, vol. MS 165. Bellingham, WA: SPIE Optical Engineering Press.

Tuchin VV (ed.) (2004) *Coherent-Domain Optical Methods for Biomedical Diagnostics, Environmental and Material Science*, vol. 2. Boston, MA: Kluwer Academic Publishers.

Vo-Dinh T (ed.) (2003) *Biomedical Photonics Handbook*. New York: CRC Press.

U

ULTRAFAST LASER TECHNIQUES

Contents

Generation of Femtosecond Pulses
Pulse Characterization Techniques

Generation of Femtosecond Pulses

D T Reid, Heriot-Watt University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Optical pulses from femtosecond lasers represent some of the shortest physical events ever observed and their high instantaneous powers allow them to be focused to intensities that can heat and compress matter to resemble the extreme conditions found in a nuclear explosion. The extreme units of time and power needed to describe this area of physics can easily distract from the simple and elegant techniques by which femtosecond pulses can be generated. The review presented here is divided into two main sections. The first explains the theory of pulse generation and shaping and the second concentrates on actual femtosecond laser sources in common use today.

Theory of Femtosecond Pulse Generation

Active Mode-locking

In an inhomogeneously broadened laser system, optical frequencies within the emission bandwidth of the material can experience gain simultaneously at discrete frequencies corresponding to the longitudinal modes of the laser resonator. In the absence of any mechanism to coherently couple energy between these modes, the laser will operate with a continuous-wave (cw) output containing a narrow band of frequencies whose roundtrip cavity gain is highest. An alternative operating regime exists in which each mode is related to its neighbor by a fixed phase relationship and it is this mode-locked condition that is used to generate a periodic sequence of ultrashort optical pulses.

To understand how mode-locking works, consider the frequency separation between adjacent longitudinal resonator modes which can be expressed as

$$\Delta\omega = 2\pi c/l \qquad [1]$$

where l is the optical roundtrip length of the resonator and c is the speed of light. If the nth longitudinal mode has an amplitude E_n then the total optical field can be denoted in complex notation as

$$E(t) = \sum_n E_n \exp \mathrm{i}[(\omega_0 + n\Delta\omega)t + n\Delta\varphi] \qquad [2]$$

with ω_0 being the center frequency of the output and $\Delta\varphi$ being the phase difference between adjacent modes. Figure 1 illustrates how a simple pulse can be produced by coherently adding together adjacent modes. In Figure 1a, three modes with equal amplitudes and the same initial phase ($\Delta\varphi = 0$) are interfered to give a pulse. As further modes are added (Figure 1b) the pulse duration decreases.

Coupling of energy between adjacent cavity modes can be achieved in practice by modulating either the

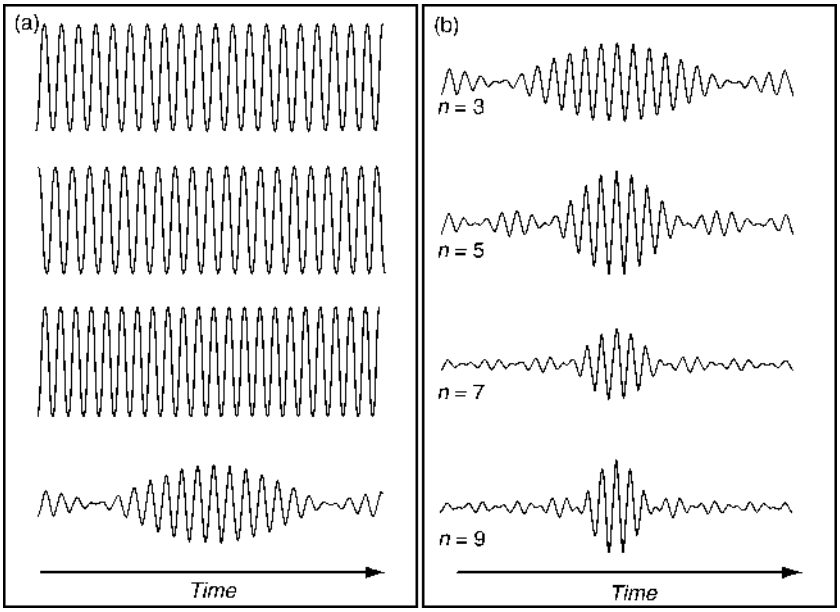

Figure 1 The principles of mode-locking: (a) three adjacent modes add coherently to produce a pulse; (b) shorter pulses require a larger number of *n* coupled modes.

intracavity loss (amplitude modulation (AM) mode-locking) or the intracavity phase (frequency modulation (FM) mode-locking). A common practical amplitude modulator is an acousto-optic modulator and for operation in the visible and near-infrared, fused silica is a common acousto-optic modulator material. Lithium-niobate devices based on the Pockels effect are the most commonly used phase modulators and are attractive for mode-locked fiber-laser applications because fiber-pigtailed waveguide devices are readily available.

Active mode-locking has been employed in a wide variety of laser systems including argon ion, dye, erbium-doped fiber, semiconductor, and solid-state lasers. Unlike passive mode-locking, no optical nonlinearity is required which means that actively mode-locked lasers can be operated with low intracavity powers. The technique is particularly suitable for mode-locked sources which require active synchronization to an external clock source.

A form of 'optical' active mode-locking exists in which gain modulation resulting from optical pumping using another mode-locked laser is the mode-locking mechanism. Known as synchronously pumped mode-locking, this technique has been used to obtain ultrashort pulses from oscillators incorporating gain media of laser or nonlinear optical materials. The technique is most appropriate to laser gain materials which have a short-lived fluorescence lifetime and whose gain therefore exhibits a significant change when the pump light is modulated at high frequencies. Synchronous pumping has been used to generate mode-locked pulses from color-center lasers, dye lasers, and optical parametric oscillators.

Passive Mode-locking

To produce the shortest optical pulses, passive mode-locking, a technique based on exploiting intracavity optical nonlinearities, must be employed. Unlike the active technique, the strength of the mode-locking action increases as the propagating pulses shorten and their peak intensities rise. One consequence of this nonlinear behavior is that the passive technique commonly generates and mode-locks entirely new cavity frequencies which were not originally present during cw operation.

Passive mode-locking has been implemented using a wide range of different nonlinear effects, but in every technique the mode-locking process can be described in terms of dynamic loss saturation combined, in certain systems, with dynamic gain saturation. The principal mode-locking element is a saturable absorber whose loss decreases as the incident pulse intensity increases. A saturable absorber can take the form either of a resonant component such as a solid-state semiconductor device or a dye jet with a fixed absorption wavelength, or a nonresonant component that replicates a saturable-absorber-like action using self-focusing or self-phase-modulation effects. Saturable absorbers can be categorized as either slow saturable absorbers or fast saturable

absorbers, depending on whether their recovery time is longer or shorter than the pulse duration, respectively. The pulse-shaping mechanisms are different in each case and will be described separately.

Slow saturable absorbers are physical devices, commonly dye molecules in solution or semiconductor-doped glass, which rely on resonant excitation to produce an excited electronic state whose transmission is greater than that of the ground state. When an optical pulse is incident on a slow absorber, its leading edge is absorbed and creates an excited state which is relatively transparent to the trailing edge. In a laser medium whose gain is readily saturated, a pulse propagating first through the absorber and then through the gain medium experiences a 'gain window' with a finite duration which strongly shapes the pulse (see Figure 2a,b).

A fast saturable absorber is required to realize mode-locking in laser media which are not easily saturated. The time-dependent gain and loss profiles of the absorber and the gain medium are again depicted in Figure 2c,d and refer to an absorber with both a rapid response and recovery time. In this system, the dynamics of the absorber and gain medium no longer represent the limiting factor influencing the pulse duration, and other effects such as linear and nonlinear dispersion become the dominant pulse-shaping mechanisms.

Pulse Shaping by Material Dispersion

Intense pulses propagating in a mode-locked laser experience strong linear and nonlinear shaping effects associated, respectively, with group-velocity dispersion and self-phase-modulation. Contributions to the net intracavity dispersion arise from the gain medium, the cavity mirrors, other elements such as intracavity prisms, even the air in the beam path. Dispersive effects in materials originate as a result of a frequency-dependent dielectric susceptibility:

$$P(\omega) = \varepsilon_0\chi^{(1)}(\omega)E(\omega) + \varepsilon_0\chi^{(2)}(\omega)E(\omega)^2 + \varepsilon_0\chi^{(3)}(\omega)E(\omega)^3 + \cdots \quad [3]$$

which is due to physical resonances (poles) in the electronic response to an applied E-field.

Linear dispersion

Linear dispersion refers to the variation of the refractive index associated with the real part of the $\chi^{(1)}$ term, namely,

$$n_0 = \sqrt{1 + \mathrm{Re}\{\chi^{(1)}\}} \quad [4]$$

and leads to a wavelength-dependent group velocity which is responsible for pulse broadening and break-up in optical systems. Dispersive effects are best analyzed in terms of optical phase which is related to the refractive index by

$$\varphi(\omega) = \frac{n_0(\omega)\omega L}{c} \quad [5]$$

where L is the medium length and c is the vacuum speed of light. The local variation of the spectral

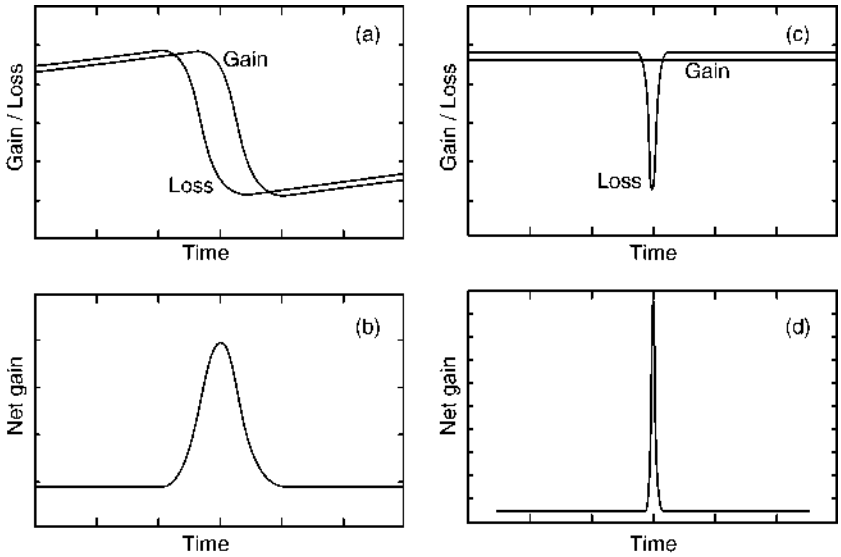

Figure 2 Saturable absorber action: (a) intracavity gain and loss profiles for a laser comprising a slow saturable absorber and a rapid relaxation time gain medium, and (b) the net gain (gain − loss) for this system; (c) gain and loss profiles for a laser comprising a near-instantaneous saturable absorber and a slow relaxation time gain medium, and (d) the net gain for this system.

phase due to an optical medium can be represented as the Taylor series

$$\varphi(\omega) = \varphi(\omega_0) + (\omega - \omega_0)\left.\frac{d\varphi}{d\omega}\right|_{\omega_0} + \frac{1}{2}(\omega - \omega_0)^2\left.\frac{d^2\varphi}{d\omega^2}\right|_{\omega_0} + \left|\frac{1}{6}(\omega - \omega_0)^3\frac{d^3\varphi}{d\omega^3}\right|_{\omega_0} + \cdots \qquad [6]$$

where ω_0 is the center frequency of the incident light. When an ultrashort pulse propagates through the medium, only the quadratic and higher-order terms in [6] have a direct influence on the shape of the pulse in time. The $(\omega - \omega_0)^2$ term is responsible for adding linear chirp which refers to a time-varying linear increase or decrease in the instantaneous optical frequency across the pulse. Obtaining the shortest pulses from a mode-locked laser requires the quadratic and higher-order terms in the roundtrip intracavity spectral phase to be zero. In practice this means minimizing the material group-velocity dispersion of the gain medium by using a short but highly doped laser crystal and then identifying additional optical elements whose group-velocity dispersion is equal in magnitude but opposite in sign to that of the crystal. Optical components for compensating group-velocity dispersion fall into two categories: those which use bulk optics and geometry to achieve a wavelength-dependent path length, and others based on interference effects within a dielectric optical coating.

Geometrical methods of dispersion compensation are commonly based on pairs of diffraction gratings or prisms and in both cases the geometry of the systems means that the group delay is wavelength dependent and can be made greater for longer wavelengths than for shorter ones. Analytic expressions describing the quadratic and cubic spectral phase of prism and diffraction grating pairs are summarized in Table 1. Dispersion compensation using prisms has been extremely successful and 11-fs pulses were reported using this approach. The simultaneous compensation of arbitrary second- and third-order spectral phase is, however, difficult and involves identifying a suitable prism material, apex spacing, and tip insertion. The expressions listed in Table 1 indicate that the form of the spectral phase depends on the refractive index of the prism material and its wavelength derivatives, while the apex separation, l, also contributes to a large extent. A further contribution to the spectral phase comes from the glass encountered by the beam passing through the prism tips which contributes a positive dispersion described by the equations for a bulk material given in Table 1. In general, an iterative solution of the equations given in Table 1 is needed to determine the geometry of a prism pair suitable for producing a desired spectral phase profile.

Laser cavity mirrors can be designed with chirped multilayer dielectric coatings which possess a chosen group-velocity dispersion profile. A Bragg mirror with smaller layer periods at the surface and larger periods deeper in the coating provides negative group-velocity dispersion because the longer wavelengths experience larger group delays within the coating than the shorter ones. These chirped mirrors provide constant reflectivity across their stopband and therefore only modify the spectral phase of a pulse, leaving the intensity unchanged. The design of chirped mirrors has been refined in recent years to enhance their operating bandwidth and to reduce undesirable modulations in the group-delay response. The technique has been used to obtain 6.5-fs pulses directly from a Ti:sapphire oscillator and to achieve compression of amplified pulses to below 5 fs.

Nonlinear dispersion

Nonlinear effects arise in all materials due to the $\chi^{(3)}$ term in [3]. In a transparent medium the refractive index due to this net susceptibility is approximately

$$n \approx n_0 + \frac{\chi^{(3)}E^2}{2n_0} = n_0 + n_2 I \qquad [7]$$

where I is the optical intensity. The importance of the $\chi^{(3)}$ susceptibility is therefore that it causes the refractive index to become intensity-dependent, and significant changes in the refractive index can be induced by the high peak intensities of ultrashort pulses.

Self-phase-modulation (SPM) resulting from the nonlinear refractive index is essential in generating the shortest pulses from modern ultrafast lasers and is caused by the time-varying intensity profile of the pulse. The higher-intensity components induce a larger nonlinear refractive index change and therefore experience a greater phase shift than the weaker components. The resulting frequency shift across the pulse can be shown to be

$$\Omega(t) = -\frac{d\phi(t)}{dt} = -\frac{\omega n_2 L}{c}\frac{dI(t)}{dt} \qquad [8]$$

which describes a redshift of the pulse leading edge and a blueshift of the trailing edge. Only the pulse center frequency remains unchanged and SPM therefore broadens the pulse spectrum by redistributing energy from the center to the wings. Across the center of the pulse, SPM leads to an approximately linear positive chirp that, with appropriate optics, can be

Table 1 Expressions for the second and third-order dispersion of common optical systems

Bulk material, thickness t

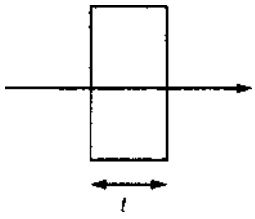

$$\frac{d^2\varphi}{d\omega^2} = t\frac{\lambda^3}{2\pi c^2}\frac{d^2n}{d\lambda^2}$$

$$\frac{d^3\varphi}{d\omega^3} = -t\frac{\lambda^4}{4\pi^2c^3}\left(3\frac{d^2n}{d\lambda^2} + \lambda\frac{d^3n}{d\lambda^3}\right)$$

Double-pass prism-pair, apex separation l, angular deviation within ray of β

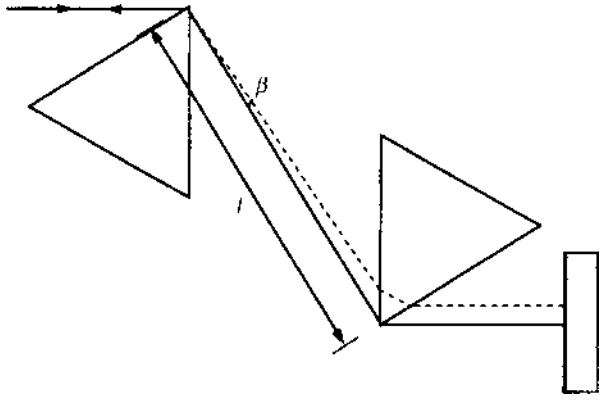

$$\frac{d^2\varphi}{d\omega^2} = \frac{\lambda^3}{2\pi c^2}\frac{d^2P}{d\lambda^2}$$

$$\frac{d^3\varphi}{d\omega^3} = -\frac{\lambda^4}{4\pi^2c^3}\left(3\frac{d^2P}{d\lambda^2} + \lambda\frac{d^3P}{d\lambda^3}\right)$$

$$\frac{d^2P}{d\lambda^2} = 4l\sin\beta\left[\frac{d^2n}{d\lambda^2} + (2n - 1/n^3)\left(\frac{dn}{d\lambda}\right)^2\right] - 8l\cos\beta\left(\frac{dn}{d\lambda}\right)^2$$

$$\frac{d^2P}{d\lambda^3} = l\cos\beta\left[\left(24/n^3 - 48n\right)\left(\frac{dn}{d\lambda}\right)^3 - 24\frac{dn}{d\lambda}\frac{d^2n}{d\lambda^2}\right] + l\sin\beta\left[\begin{array}{l}\left(\frac{dn}{d\lambda}\right)^3\left(12/n^6 + 12/n^4 + 8/n^3 - 16/n^2 + 32n\right)\\ +\left(24n - 12/n^3\right)\frac{dn}{d\lambda}\frac{d^2n}{d\lambda^2} + 4\frac{d^3n}{d\lambda^3}\end{array}\right]$$

Double-pass grating-pair, separation l, angle of incidence γ, line spacing d

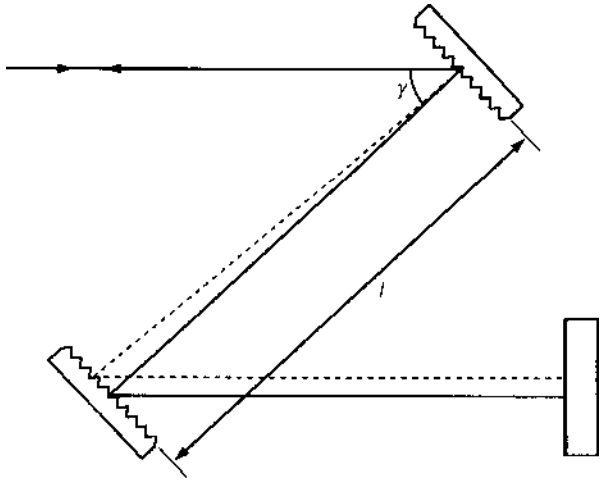

$$\frac{d^2\varphi}{d\omega^2} = l\frac{\lambda^3}{\pi c^2 d^2}\left[1 - \left(\frac{\lambda}{d} - \sin\gamma\right)^2\right]^{-3/2}$$

$$\frac{d^3\varphi}{d\omega^3} = -6l\frac{\lambda^4}{c^3d^2}\left(1 + \frac{\lambda}{d}\sin\gamma - \sin^2\gamma\right)\left[1 - \left(\frac{\lambda}{d} - \sin\gamma\right)^2\right]^{-5/2}$$

corrected using an equal but opposite amount of negative second-order dispersion. The combination of spectral broadening using SPM and subsequent chirp-removal using linear dispersion is the basis for optical pulse compression techniques that have been applied effectively to create pulses which exist for only a few carrier-wave periods.

When the pulse propagates in an environment where SPM and linear dispersion are well balanced and evenly distributed, a soliton will evolve which, in the absence of loss, can propagate indefinitely without changing shape. The nonlinear Schrödinger equation predicts that lasers in which the dominant pulse-shaping mechanisms are negative

group-velocity dispersion and self-phase-modulation will produce soliton-like pulses with a $\text{sech}^2(t)$ intensity profile. In experimental measurements of ultrafast lasers the pulse shape is commonly assumed to be $\text{sech}^2(t)$ and the chirp of the pulses is then estimated by measuring their duration–bandwidth product, $\Delta\nu\Delta\tau$, which is a dimensionless quantity that can be treated as a figure-of-merit for ultrashort pulses. For unchirped pulses the duration–bandwidth product is independent of their duration or spectral bandwidth but is sensitive to their shape: chirp-free Gaussian pulses have $\Delta\nu\Delta\tau = 0.441$ while $\text{sech}^2(t)$ pulses have $\Delta\nu\Delta\tau = 0.315$.

Sources of Femtosecond Pulses

Dye Lasers

The first practical passively mode-locked ultrafast lasers were those based on a gain medium of a thin (~10 μm) organic dye jet and incorporating a slow saturable absorber jet at a separate cavity focus. The most common combination was Rhodamine 6G in ethylene glycol (gain jet) and 3,3-diethyloxadicarbocyanine iodide in ethylene glycol (absorber jet) which produced pulses at a wavelength of 620 nm. The most successful approach was colliding-pulse mode-locking in which two counter-propagating pulses within a ring cavity are synchronized to arrive simultaneously at the absorber jet, so achieving greater saturation of the absorber and allowing the pulses to be transmitted with a loss significantly smaller than in the single-pulse case. The first sub-100-fs optical pulses were reported in 1981 from a colliding-pulse mode-locking dye laser and later work reported pulse durations as low as 27 fs but the configuration was inefficient and produced low average output powers. Other approaches such as synchronous pumping and hybrid mode-locking produced higher output powers but the maintenance-intensive nature of dye lasers and the hazardous nature of many of the chemicals used led to their gradual replacement by more convenient solid-state alternatives, based principally on Ti:sapphire.

Color-Center Lasers

A new source of femtosecond pulses in the near-infrared was reported in 1984. Coined the soliton laser the system was a synchronously pumped KCl:Tl color-center laser tunable from 1.4 to 1.6 μm and modified to allow direct generation of pulses as short as 50 fs, substantially shorter than the 8 ps achieved in synchronously pumped operation alone. Femtosecond operation was achieved by coupling a length of polarization-preserving anomalously dispersive fiber to the main laser cavity using a beamsplitter and retroreflector arrangement. This coupled-cavity arrangement later became known as additive pulse mode-locking (APM) and produced short pulses when the optical length of the fiber was adjusted to be an integral multiple of the main cavity length. The long upper-state lifetime of the laser-active centers in the KCl:Tl gain medium allowed particularly stable APM operation, but mode-locking of this kind was also successfully employed in an actively stabilized $NaCl:OH^-$ color-center laser to produce 110-fs pulses. The need to maintain color-center crystals at liquid-nitrogen temperatures limited the convenience of these lasers but until the advent of Ti:sapphire they remained the only solid-state sources capable of femtosecond operation.

Ti:sapphire Lasers

The creation in 1986 of a new broadband laser material – titanium-doped sapphire or Ti:sapphire – began a revolution in ultrafast laser sources that still continues today. Unlike other contemporary solid-state materials for mode-locked lasers such as Nd:YAG, Ti:sapphire is a vibronic gain medium which exhibits a strong electron-phonon coupling between the titanium ion transition and the host lattice. This vibronic coupling leads to a broad continuum of possible transition energies and results in a fluorescence bandwidth covering 670–1050 nm, making Ti:sapphire capable of supporting extremely short femtosecond pulses. The long fluorescence lifetime of Ti:sapphire means that, unlike dye lasers, femtosecond mode-locking requires a passive method using a fast saturable absorber and two methods have prevailed and are discussed below.

The first passively mode-locked Ti:sapphire lasers were based on the APM technique in which mode-locking is achieved by a fast-saturable absorber action due to interference between the intracavity pulse and a self-phase-modulated replica propagating in an auxiliary cavity. An important breakthrough was made in 1990 when it was reported that an APM Ti:sapphire laser continued to operate even when the auxiliary cavity was blocked. This discovery was attributed to an entirely new mode-locking effect known as Kerr-lens mode-locking (KLM) or self-mode-locking in which the presence of a Kerr lens within the gain medium during mode-locked operation changes the mode focusing within the cavity so that, compared to cw operation, mode-locked pulses experience higher gain. Two configurations of KLM exist and are known as soft aperture mode-locking, where the presence of the Kerr lens increases the gain by improving the overlap between the pump and laser

modes, and hard-aperture mode-locking, in which a physical aperture such as a slit or the edge of a prism is adjusted to introduce greater loss for cw operation. Because KLM is based on a nonresonant nonlinearity, Ti:sapphire lasers can be mode-locked at any wavelength in their gain bandwidth using this technique. Recent advances in controlling the intracavity group-velocity dispersion characteristics have led to the generation of sub-5-fs pulses directly from a Ti:sapphire oscillator. The principal drawback of the KLM method is that it is not self-starting because mode-locked operation must be seeded by an intense noise spike or other short fluctuation. Different starting methods have been applied successfully including mirror tapping, acousto-optic modulation, and mode-dragging, and once mode-locking has been initiated it is generally stable until some external perturbation disturbs the intracavity beam.

An alternative method for obtaining femtosecond pulses has been pioneered largely in parallel with the KLM efforts and is based on using resonant nonlinearities in semiconductors as fast saturable absorbers. The approach is to fabricate a semiconductor multiple-quantum-well (MQW) device whose bandgap has been chosen to match the laser wavelength and whose absorption can be saturated at high fluences. Initially these devices were utilized in an APM geometry in which the MQW device was situated in an auxiliary cavity but it was later realized that an equivalent monolithic configuration could be obtained by sandwiching the MQW layer between two high-reflectivity mirrors (see Figure 3, inset) to form a Fabry–Perot structure and the resulting element was known as an antiresonant Fabry–Perot saturable absorber. The first application of a such a semiconductor saturable absorber mirror (SESAM) for mode-locking Ti:sapphire was reported in 1995 where self-starting operation was demonstrated. A typical modern Ti:sapphire laser design is shown in Figure 3 and illustrates how a saturable absorber mirror can be integrated into the cavity. Further design enhancements such as reducing the finesse of the Fabry–Perot cavity resulted in SESAMs with a broader wavelength response which enabled sub-10-fs self-starting operation.

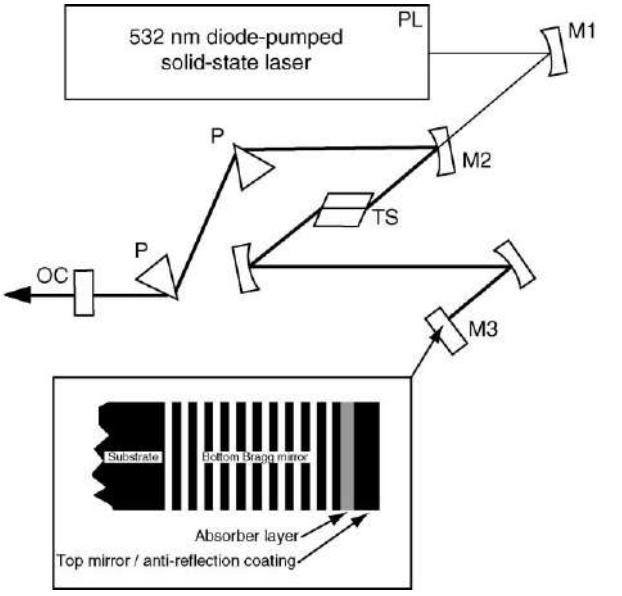

Figure 3 Typical configuration of a modern femtosecond Ti:sapphire laser. PL, pump laser; M1, pump focusing mirror; M2, high-reflectivity cavity mirror (800 nm); TS, Ti:sapphire crystal, typically, 2–10 mm long; M3, SESAM mirror; P, dispersion-compensating prisms; OC, output coupler. Inset: a simplified schematic of a saturable-absorber Bragg mirror.

Fiber Lasers

Ultrafast fiber lasers, particularly those based on silica doped with rare earth Er^{3+}, Nd^{3+}, or Yb^{3+}ions, have enjoyed renewed attention in recent years as efforts to scale their average output powers and pulse energies to practical levels have proved successful. Four main mode-locking strategies have been applied to these systems: active mode-locking which is generally restricted to picosecond generation; intensity-dependent feedback using a nonlinear amplifying loop mirror; polarization APM; and semiconductor saturable absorber mode-locking. The upper-state lifetime of rare-earth-doped fibers is long and therefore a fast saturable absorber mechanism is necessary for passive mode-locking, limiting the choice of mode-locking elements to those based on Kerr or semiconductor nonlinearities.

The development of mode-locked fiber sources has concentrated on Er-doped lasers because of the compatibility of their output wavelengths with the 1.5 μm optical communication window in standard silica fiber. Passively mode-locked designs based on the nonlinear loop mirror work by exploiting the Kerr effect in an intracavity fiber Sagnac interferometer containing a gain section. With suitable control of the intracavity polarization, such a nonlinear amplifying loop mirror (NALM) can be set up to transmit light of high intensities but reflect low intensities because of the differential phase shift induced between the two interferometer arms. By adding the loop mirror to a unidirectional ring cavity a figure-of-eight laser is formed (Figure 4) which is capable of producing femtosecond soliton pulses with average powers of around 1 mW. Drawbacks of the figure-of-eight laser include the presence of spectral sidebands at the shortest pulse durations and a tendency for multiple pulse operation. Mode-locking can also be achieved

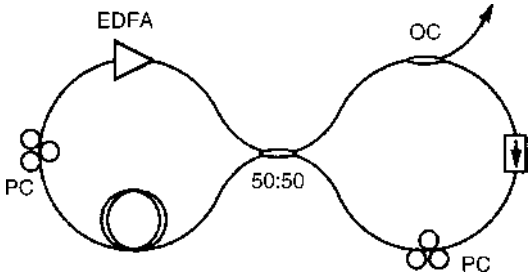

Figure 4 Schematic of a figure-of-eight femtosecond fiber laser. EDFA, erbium-doped fiber amplifier gain section; OC, output coupler; PC, polarization controller.

using a variation on the APM technique and significant average output powers have been demonstrated from stretched-pulse-APM lasers, with up to 90 mW being reported at pulse energies of 2.25 nJ.

Semiconductor saturable absorber mode-locking has been demonstrated in fiber lasers using a variety of configurations including bulk InGaAsP on InP and saturable Bragg reflectors based on InGaAs/InP MQWs on a AlAs/GaAs mirror structure. Fiber lasers mode-locked using semiconductor saturable absorbers are self-starting and exhibit stable mode-locking, normally at the fundamental cavity frequency. The narrow bandwidth of saturable absorber devices at near-infrared wavelengths makes the generation of sub-100-fs pulses difficult.

Semiconductor Sources

Diode lasers and external-cavity lasers based on semiconductor optical amplifiers (SOAs) are of importance because of their ability to produce ultrashort pulses from a compact system at wavelengths of relevance to optical fiber communications. The two principal methods used to generate ultrashort pulses from such devices are mode-locking with an external cavity or gain-switching (equivalently known as Q-switching) in a bow-tie waveguide structure. Subpicosecond pulse durations are now routinely produced using dispersion compensation and/or nonlinear pulse compression techniques and, by combining these strategies with Er:fiber amplification, 20-fs pulses have been produced. The mode-locking and gain-switching techniques each have their individual advantages. Gain-switching provides a simple and compact approach offering high average output power, but significant timing jitter and poor spectral quality can accompany pulses produced in this way. Mode-locked systems are physically larger than gain-switched devices because they require an external cavity and additional optical components, but shorter and higher quality pulses are achievable.

Other Common Solid-State Laser Sources

Certain solid-state gain media offer the potential for direct diode pumping, miniaturization and operation at wavelengths or average powers not accessible with other sources. Systems based on Nd or Cr ion transitions have been shown to be capable of ultrashort picosecond or femtosecond operation. Femtosecond mode-locking (175 fs) has been reported in Nd:glass which has one of the broadest fluorescence spectra of the Nd materials. Crystals with active ions of Cr^{2+}, Cr^{3+}, and Cr^{4+} have broad transitions covering wavelengths in the 2.0 μm, 800 nm and 1.4 μm regions and femtosecond operation has been demonstrated in Cr^{4+}:YAG at 1.52 μm, Cr^{4+}:Mg_2SiO_4 (Cr:forsterite) at 1.25 μm, Cr^{3+}:$LiSrAlF_4$ (Cr:LiSAF) at 850 nm, and Cr^{3+}:$LiSrGaF_6$ (Cr:LiSGAF). Direct diode pumping of Cr-ion-based lasers is attractive because of the opportunities for producing low-noise, compact, and efficient femtosecond sources, but the limited power available from high-quality pump diodes results in low intracavity powers which make KLM operation difficult. As a result, stable operation is better achieved using a physical saturable absorber, and SESAM mode-locking has been applied successfully to many of these lasers.

Sources Based on Nonlinear Frequency Conversion

Laser sources alone are unable to provide ultrashort pulses continuously tunable from the ultraviolet to the mid-infrared, and large gaps exist at wavelengths where no broadband laser materials are available. Harmonic generation using Ti:sapphire can be used to access some parts of the visible and ultraviolet regions, but infrared and certain visible wavelengths cannot be generated in this way and coverage is limited by the ability to tune the source itself. The nonlinear production of infrared wavelengths requires parametric down-conversion using a synchronously pumped optical parametric oscillator. Ti:sapphire pumped synchronously pumped optical parametric oscillators have produced, either directly or using harmonic generation, outputs from the green to beyond 7 μm.

Sources of Amplified Ultrashort Pulses

Modern ultrafast oscillators can produce pulses with peak powers of >1 MW, average powers of several watts, and cavity frequencies from 4 MHz to 2 GHz, but many applications require pulses with higher intensities or lower pulse repetition frequencies. Additional amplification stages can be used to satisfy these requirements and these are commonly

implemented in Nd:glass for picosecond pulses and Ti:sapphire for femtosecond pulses. In order to avoid damage and unwanted nonlinear effects caused by high pulse intensities within the amplifier system, most practical sources now apply the chirped-pulse amplification (CPA) approach in which a low-energy femtosecond seed pulse is stretched to sub-nanosecond durations to reduce its peak power then amplified and finally compressed to its original duration. Using the chirped-pulse amplification technique, pulses with peak powers as high as 1.5 petawatt have been generated.

Two possible amplifier geometries are used which are categorized either as multipass or regenerative amplifiers. In the multipass configuration the injected pulse is refocused many times through the same gain crystal using a system of mirrors. The gain crystal is pumped above the saturation fluence in order to extract the maximum stored energy from the medium per pass. The multipass geometry is more commonly used for high-average-power operation and critical alignment is required to maximize the overlap between the pump mode profile and the multiple beams intersecting the gain medium. The alternative regenerative amplifier is essentially a stable laser resonator which includes a Pockel cell and a polarizing beamsplitter to enable pulses to be switched into and out of the cavity. The gain per pass is considerably lower than in the multipass configuration but regenerative systems produce superior beam quality and their performance can be optimized without affecting the beam pointing of the output. Commercially available Ti:sapphire amplifiers commonly use the regenerative method and are available with pulse energies of up to 1 mJ at repetition frequencies in the 1–300 kHz range.

Further Reading

Agrawal GP (1995) *Nonlinear Fiber Optics*. San Diego, CA: Academic Press.

Duling IN (1995) *Compact Sources of Ultrashort Pulses*. Cambridge University Press.

Fork RL, Martinez OE and Gordon JP (1984) Negative dispersion using pairs of prisms. *Optics Letters* 9: 150–152.

Keller U, Weingarten KJ, Kartner FX, *et al*. (1996) Semiconductor saturable absorber mirrors for femtosecond to nanosecond pulse generation. *Solid-State Lasers, IEEE Selected Topics in Quantum Electronics* 2: 435–453.

Siegman AE (1986) *Lasers*. University Science Books.

Spence DE, Kean PN and Sibbett W (1991) 60-fsec pulse generation from a self-mode-locked Ti-sapphire laser. *Optics Letters* 16: 42–44.

Szipocs R, Ferencz K, Spielmann C and Krausz F (1994) Chirped multilayer coatings for broad-band dispersion control in femtosecond lasers. *Optics Letters* 9: 201–203.

Yariv A (1989) *Quantum Electronics*. New York: Wiley.

Pulse Characterization Techniques

D J Kane, Southwest Sciences, Inc., Santa Fe, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Ultrashort optical pulses have very short time durations, typically less than a few tens of picoseconds. As a result, these pulses are spectrally broad. Because the index of refraction of materials is a function of wavelength, different wavelengths of light travel at different speeds in optical materials, causing the properties of ultrashort optical pulses to change as they propagate. However, the shape of the pulse can influence how the pulse itself interacts with materials. Thus, the goal of ultrafast laser pulse measurement is to obtain not only the intensity profile of the pulse, but also the actual variation of the frequencies that make up the pulse. This is called measuring the intensity and phase of the pulse, respectively.

Even though the development of techniques to characterize ultrashort optical pulses has not been easy, a myriad of pulse measurement techniques have been developed. In this chapter, we will confine our discussion to the most well-known pulse and generally accepted pulse characterization methods in an effort to provide a basic working knowledge of pulse measurement techniques and a fundamental understanding of the principles behind pulse measurement. After a brief discussion of the mathematical representation of ultrashort optical pulses, we will discuss pulse measurement in (roughly) chronological order, starting with autocorrelation methods. The next section introduces a useful method for the measurement of the relative phase between two pulses called spectral interferometry (SI). Following SI, we introduce the notion of the time-frequency

representation of ultrashort optical pulses where techniques such as frequency-resolved optical gating (FROG) are discussed. Following the time-frequency section, a relatively new technique known as spectral phase interferometry for direct electric field reconstruction (SPIDER) is discussed.

Mathematical Representation of an Optical Pulse

The time-dependent variations of the pulse are embodied in the pulse electric field, $A(t)$, which can be written:

$$A(t) = \mathrm{Re}[E(t)\mathrm{e}^{\mathrm{i}\omega_0 t}] \qquad [1]$$

where ω_0 is the carrier frequency and Re refers to the real part. While we can use $A(t)$ as it stands for calculations, it is much easier to remove the rapidly varying ω_0 part, $\mathrm{e}^{\mathrm{i}\omega_0 t}$, and use a slowly varying envelope together with a phase term that contains only the frequency variations, not the rapidly varying carrier frequency:

$$E(t) = [I(t)]^{1/2}\mathrm{e}^{\mathrm{i}\varphi(t)} \qquad [2]$$

where $I(t)$ and $\varphi(t)$ are the time-dependent intensity and phase of the pulse. (Note that $E(t)$ is complex.) The frequency variation, $\Omega(t)$, is the derivative of $\varphi(t)$ with respect to time:

$$\Omega(t) = -\mathrm{d}\varphi(t)/\mathrm{d}t \qquad [3]$$

The pulse field can be written equally well in the frequency domain by taking the Fourier transform of eqn [2]:

$$\tilde{E}(\omega) = [\tilde{I}(\omega)]^{1/2}\mathrm{e}^{-\mathrm{i}\phi(\omega)} \qquad [4]$$

where $\tilde{I}(\omega)$ is the spectrum of the pulse, and $\phi(\omega)$ is its phase in the frequency domain. The spectral phase contains time versus frequency information; that is, the derivative of the spectral phase with respect to frequency yields the time arrival of the frequency – the group delay.

Obtaining the intensity and phase, $I(t)$ and $\varphi(t)$ (or $\tilde{I}(\omega)$ and $\phi(\omega)$ is called full characterization of the pulse. Common phase distortions include linear chirp, where the phase (either the time domain or frequency domain) is parabolic. When the frequency is increasing in time, the pulse is said to have positive linear chirp; negative linear chirp is when the high frequencies lead the lower frequencies. Higher-order chirps are common, but for these, differentiation between spectral and temporal chirp is required because spectral phase and temporal phase are not interchangeable.

Autocorrelation

Traditionally, we measure events using shorter events. Unfortunately, for the ultrafast researcher, shorter events do not exist and modern electronics are not fast enough. Therefore, because the shortest event we have is the event we wish to measure, traditional pulse measurement methods use the pulse itself to determine the approximate duration of the pulse in question giving birth to the ubiquitous intensity autocorrelation (see Figure 1). While autocorrelations cannot be used to fully characterize ultrashort optical pulses, the methods used in autocorrelations are fundamental to all pulse measurement schemes.

The intensity autocorrelation is measured by combining a pulse and a delayed replica of a pulse in a nonlinear medium such as a second harmonic generation (SHG) crystal. A pulse is sent onto a beamsplitter to produce the two replicas. One pulse is delayed relative to the other and both are focused together into a SHG crystal. As the delay of one pulse relative to the other is varied, the intensity of the second-harmonic signal is recorded. The intensity autocorrelation is not limited to using SHG; any nonlinearity may be utilized. Indeed, two-photon absorption in a semiconductor LED or photodiode often acts as a convenient nonlinearity. Sometimes a third-order nonlinearity is used to provide some direction of time information about the pulse. When the generation of the signal involves phase matching, such as second harmonic generation, care must be taken to use a thin crystal. Typically, 10 μm to 1 mm thick SHG crystals are used, depending on the bandwidth of the pulse to be measured. The exact thickness depends on the SHG crystal, the pulse width, and the requirements of the measurement.

Regardless of the nonlinearity used, an autocorrelation yields only the approximate duration and

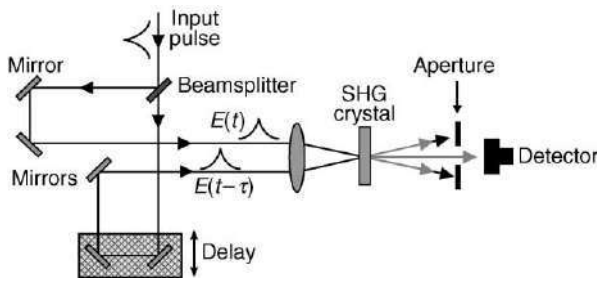

Figure 1 An intensity autocorrelator used to determine the approximate duration of a pulse. The pulse is split into two replicas that are sent into delay lines. The outputs from both delay lines are focused into a doubling crystal. The second harmonic output is monitored as a function of delay between the two pulses. A plot of the signal versus time is the intensity autocorrelation of the pulse.

shape of the pulse. The structure of the pulse is smeared, producing a smooth, featureless profile for even a complex pulse. Thus, the intensity autocorrelation alone does not determine the intensity profile of the pulse, $I(t)$. Only the interaction of the intensity of the pulse is recorded, producing no phase information for the pulse. Some qualitative information can be gleaned from the spectrum and the autocorrelation, but usually only that the pulse is chirped, and higher-order chirps and complex pulse structures elude such an analysis.

Another useful form of intensity autocorrelation is the single-shot autocorrelator. In this case, fairly large beams are used, and these beams are set to intersect at an angle of 2Φ, tilting the pulse fronts, so that delay is mapped onto a spatial coordinate (see Figure 2). The beams are then focused into a nonlinear medium using a cylindrical lens. The interaction region in the SHG crystal is imaged onto a linear array or CCD camera. Typically, when second-harmonic generation is used, the SHG crystal is oriented for Type I phase matching. The time window is proportional to $\Delta w(\sin\Phi)/\Delta v_g$, where Δw is the beam waist at the SHG crystal, and v_g is the group velocity of the pulse in the crystal. Like all SHG interactions, care must be exercised to insure that the phase matching is sufficient to mix the entire bandwidth of the pulse.

Jean-Claude Diels improved the intensity autocorrelation by the development of the interferometric autocorrelation. In this configuration, a Michelson interferometer is typically used so that the beams are collinear when arriving at the SHG crystal, which allows the beams to interfere. As a result, fringes appear on the autocorrelation. By examining the shape and extent of the fringes, some information about the pulse chirp can be obtained. While this does add to the information provided by an intensity autocorrelation, the added information is only the spectrum of the second harmonic. Unfortunately the addition of the second harmonic spectrum does not provide enough information for full retrieval of the intensity and phase.

Even though autocorrelation alone does not completely determine pulse intensity and phase, it is a very simple and useful technique to determine approximate pulse duration. Sometimes, cross-correlation can be used to determine the duration of a long pulse, if a shorter one is available, or a pulse too weak (or at an inappropriate wavelength) to measure with autocorrelation. The addition of the spectrum and/or the second harmonic spectrum can provide more,

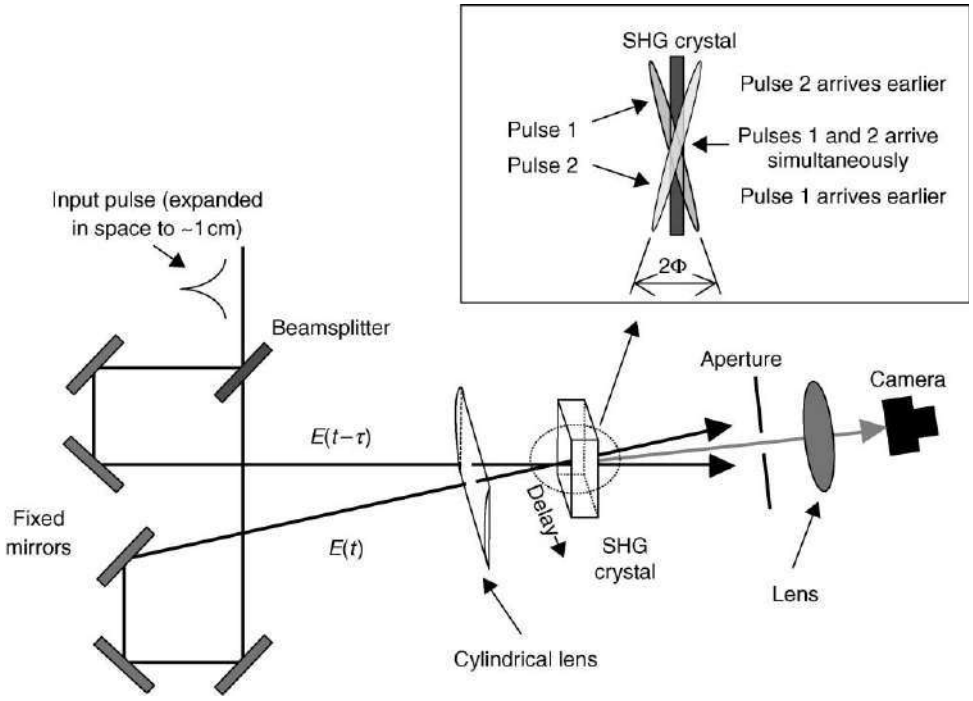

Figure 2 A single shot intensity autocorrelator. For this type of autocorrelator, the beam size is large, on the order of 1 cm in diameter, and the pulse fronts are tilted with respect to each other in order to map relative delay to position. The two beams are focused using a cylindrical lens and the output is recorded using an array detector. The inset shows how delay is mapped into position. Inset: the tilted pulse fronts cause time delay to be mapped spatially.

albeit, incomplete information about pulse chirp. More importantly, the experimental techniques used by correlation methods are fundamental to all pulse measurement methods.

Spectral Interferometry – Relative Phase Measurements

While determination of the absolute intensity and phase of an ultrashort laser pulse is difficult, determination of the relative phase is not. Invented in the late 1800s, spectral interferometry (SI), measures the relative phase between two pulses. SI, a linear technique, can also be modified (albeit, a some-what complex modification) to be self-referencing for full characterization of the ultrashort optical pulse (See the section on SPIDER below). In addition, when combined with a full pulse characterization technique, SI, because it is a linear technique, provides a method by which very weak optical pulses can be characterized (see below).

Typically, an SI apparatus uses a Mach–Zender interferometer (see Figure 3). A pulse is split into two replicas; one is sent through a region or medium to measure and the other through a known path. The two pulses remain separated by some known time and are sent, collinearly, into a spectrometer; no time scanning is required. An analysis of the fringe pattern yields the relative phase between the two pulses. In other words, SI provides $\phi_{\text{unk}}(\omega) - \phi_{\text{ref}}(\omega)$, where $\phi_{\text{unk}}(\omega)$ is the unknown pulse phase versus frequency. The spectrum of the two pulses in the frequency domain is

$$\tilde{I}_{\text{SI}}(\omega) = |\tilde{E}_0(\omega) + \tilde{E}_{\text{unk}}(\omega)|^2$$

$$\tilde{I}_{\text{SI}}(\omega) = \tilde{I}_0(\omega) + \tilde{I}_{\text{unk}}(\omega) + 2\sqrt{\tilde{I}_0(\omega)}\sqrt{\tilde{I}_{\text{unk}}(\omega)} \times \cos(\phi_{\text{unk}}(\omega) - \phi_0(\omega) - \omega\tau) \quad [5]$$

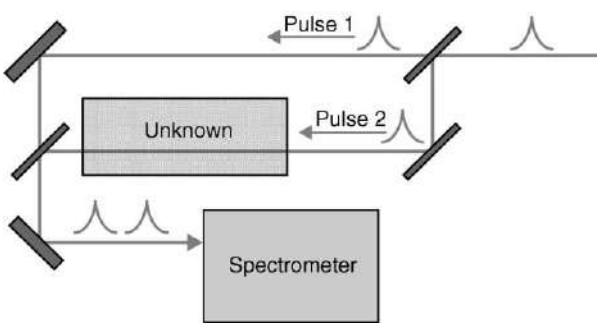

Figure 3 A Mach–Zender interferometer is the typical arrangement for spectral interferometry. One path acts as the reference arm and the sample to be measured is placed in the other arm. Unlike standard interferometry, the delay is fixed in a spectral interferometry experiment.

where $\tilde{I}_{\text{SI}}(\omega)$ is the output spectrum, $\tilde{E}_0(\omega)$ is the reference pulse electric field, $\tilde{E}_{\text{unk}}(\omega)$ is the unknown pulse electric field, $\tilde{I}_0(\omega)$ is the reference pulse spectrum, $\tilde{I}_{\text{unk}}(\omega)$ is the unknown pulse spectrum, $\phi_{\text{unk}}(\omega)$ is the unknown pulse phase, $\phi_0(\omega)$ is the reference pulse phase, and τ is the delay between the two pulses. To facilitate extracting the phase difference from the SI output, τ is chosen to yield fringes in the sum spectrum. The spectral fringes, which have a period inversely proportional to the optical path difference between the two beams (see Figure 4), contain all of the phase difference information.

Figure 4 shows the steps required to obtain the phase difference between the two pulses. The first step is to subtract out the spectra of the individual pulses in order to isolate the spectral interferogram, $S(\omega)$ (Figure 4a), where

$$S(\omega) = 2\sqrt{\tilde{I}_0(\omega)}\sqrt{\tilde{I}_{\text{unk}}(\omega)}\cos(\phi_{\text{unk}}(\omega) - \phi_0(\omega) - \omega\tau) \quad [6]$$

By Fourier transforming $S(\omega)$, we obtain

$$\Im^{-1}[S(\omega)] = f(t-\tau) + f(-t-\tau) \quad [7]$$

where $f(t)$ is the correlation product between the reference and the unknown electric field (see Figure 4b). The Fourier transform of eqn [7] multiplied by a Heaviside function, $\Theta(t)$ (to remove $f(-t-\tau)$), recovers the amplitude and phase of $f(\omega)$, the Fourier transform of $f(t-\tau)$, which contains the relative spectral phase between the unknown pulse and the unmodified pulse. Some care, however, needs to be taken to correctly remove the linear phase due to the time delay between the two pulses, $\omega\tau$. Also, for best results, the spectrometer should be well calibrated and care should be taken to properly window the spectrum before taking the Fourier transform.

Spectral interferometry is a straightforward and simple technique that always provides a relative phase. For this very reason, care must be taken to make sure the beams are mode-matched as well as collinear; also, interferometric stability must be maintained over the course of the measurement. Furthermore, SI is nongating; that is, CW background in the laser can add to the interferogram, masking transient effects. Nevertheless, spectral interferometry is a useful technique that has been applied to the measurement of the linear and nonlinear spectral phase introduced by optical fibers. More recently, SI has been applied to phase-locking and to phase-resolved pump probe experiments.

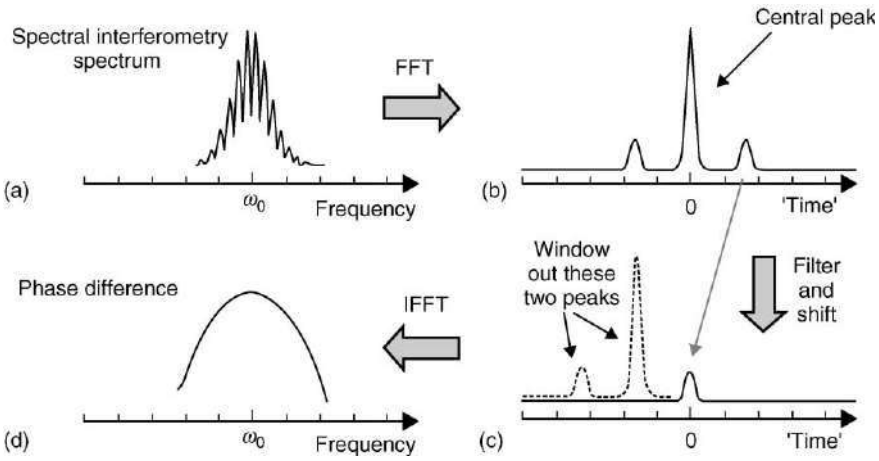

Figure 4 The steps required for the analysis of spectral interferograms. Part (a) shows a sample SI interferogram. The dominant frequency of the fringes is the delay multiplied by the speed of light. The actual phase differences of interest are the perturbations on this frequency. Part (b) is the Fourier transform of the SI interferogram in Part (a). (Note that this is not the true time domain.) The central peak contains only spectral information. All of the phase information is contained in the satellite peaks. The next step is to mask out the central peak and one of the satellite peaks (Part (c)). The phase of the inverse Fourier transform of Part (c) yields the relative phase between the two pulses (Part (d)).

Time-Frequency Representation

In 1971, E.B. Treacy laid the foundations for a revolution in pulse measurement by introducing the idea of measuring the intensity versus time for different spectral splices of a pulse. Because his measurements provided both time and frequency information simultaneously, these measurements could be thought of as applying in a hybrid time-frequency domain. At first, this may seem confusing, but similar methods have been used to visualize sounds patterns (a musical score, for example) and speech. A time-frequency plot of this type, where spectral slices of a pulse are plotted versus time, is called a sonogram. The mathematical formalism of a sonogram is:

$$S_{\mathrm{E}}(\Omega, T) = \left| \int_{-\infty}^{\infty} \tilde{E}(\omega) h(\omega - \Omega) \mathrm{e}^{-\mathrm{i}\omega T} \mathrm{d}\omega \right|^2 \qquad [8]$$

where $\tilde{E}(\omega)$ is the electric field of the pulse to be measured in the frequency domain and $h(\omega - \Omega)$ is a frequency gate that varies with frequency. The magnitude squared of the inverse Fourier transform of spectral slices yields the sonogram.

In 1991, Chilla and Martinez showed that the sonogram could be used to reconstruct the full intensity and phase of the pulse. That is, under certain conditions, the approximate group delay could be determined as a function of frequency from the sonogram by finding the peak time arrival of each spectral slice; integration of the group delay yields the spectral phase, which together, with the pulse spectrum, provides the intensity and phase of the pulse in the frequency domain. They labelled this technique as frequency domain phase measurement (FDPM).

An experimental diagram of an FDPM apparatus is shown in Figure 5. The pulse is split into two replicas, and one of the pulse replicas is spectrally filtered. The spectrally filtered pulse is cross-correlated with the original pulse to find the 'time arrival' of the spectrally filtered pulse – defined as the peak of the cross-correlated pulse. Because the spectrally filtered pulse has a much longer time duration than the original pulse, the small perturbation caused by the finite length of the original pulse is neglected.

The main difficulty of the Chilla–Martinez method is that the frequency gate must be very narrow, reducing the filtered pulse energy significantly, which greatly reduces the measured signal strength. If the spectral phase is not well behaved, the filtered pulse may not be much longer than the original pulse. In addition, for a given frequency, the sonogram may have two or more peaks in time, that is, the group delay may not be a function. However, a method called 2D phase retrieval, discussed in the next section, can alleviate many of these issues.

A spectrogram is a relative of the sonogram. Rather than determining the time arrival of spectral slices of a pulse, a spectrogram is obtained when the spectrum of time slices of the pulse are measured. The spectrogram is experimentally much easier to obtain than the sonogram; however, the measured quantity is not

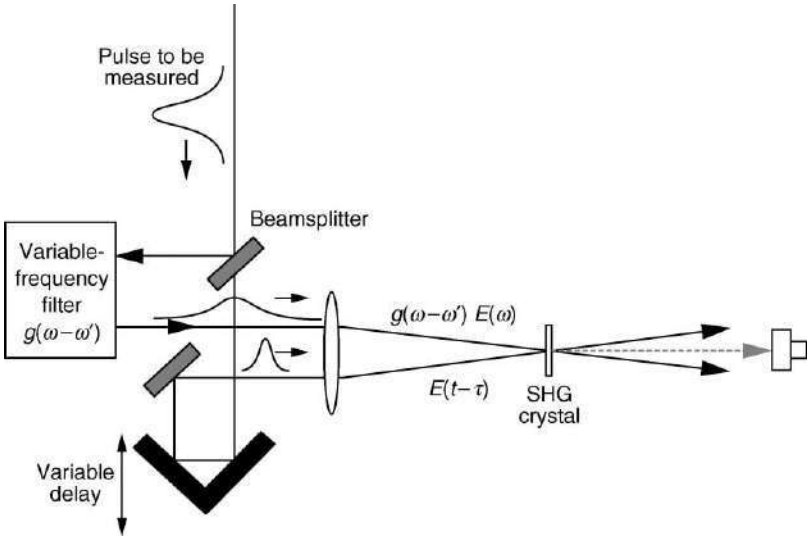

Figure 5 Measuring a sonogram requires determining the time arrival of spectral slices of the pulse. The pulse is split into two replicas using a beamsplitter. One replica is spectrally filtered using a tunable filter. The other pulse is cross-correlated with the spectrally filtered pulse to determine the time arrival.

quite as mathematically useful as the sonogram because the time slicing of the pulse occurs in the time domain. Thus, if a narrow time gate is used, the time domain phase is obtained. The time domain phase together with the intensity of the pulse, $I(t)$, provides the full intensity and phase of the pulse. Unfortunately, the intensity profile is not as readily measurable as the pulse spectrum. Consequently, to obtain the full intensity and phase of a pulse from its spectrogram requires an iterative 2D phase retrieval algorithm. This is the basis of a pulse measurement technique called frequency-resolved optical gating (FROG).

Frequency-Resolved Optical Gating

Frequency-resolved optical gating (FROG), developed by Kane and Trebino, measures the spectrum of a particular temporal component of the pulse (see Figures 6 and 7) by spectrally resolving the signal pulse in an autocorrelation-type experiment using an instantaneously responding nonlinear medium. As shown in Figure 6, FROG involves splitting a pulse and then overlapping the two resulting pulses in an instantaneously responding $\chi^{(3)}$ or $\chi^{(2)}$ medium. Even though any instantaneous nonlinear interaction may be used to implement FROG, perhaps the most intuitive is the polarization-gating configuration. In this case, induced birefringence, due to the electronic Kerr effect, is used as the nonlinear-optical process. In other words, the 'gate' pulse causes the $\chi^{(3)}$ medium, which is placed between two crossed polarizers, to become slightly birefringent. The polarization of the 'gated' probe pulse (which is cleaned up by the first polarizer) is rotated slightly by the induced birefringence allowing some of the 'gated' pulse to leak through the second polarizer. This is referred to as the signal. Because most of the signal emanates from the region of temporal overlap between the gate pulse and the probe pulse, the signal pulse contains the frequencies of the 'gated' probe pulse within this overlap region. The signal is then spectrally resolved, and the signal intensity is measured as a function of wavelength and delay time τ. The resulting trace of intensity versus delay and frequency is a spectrogram, a time- and frequency-resolved transform that intuitively displays time-dependent spectral information of a waveform.

The spectrogram can be expressed as:

$$S_{\mathrm{E}}(\omega,\tau) = \left| \int_{-\infty}^{\infty} E(t) g(t-\tau) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t \right|^2 \qquad [9]$$

where $E(t)$ is the measured pulse's electric field, $g(t-\tau)$ is the variable-delay gate pulse, and the subscript E on S_{E} indicates the spectrogram's dependence on $E(t)$. The gate pulse $g(t)$ is usually somewhat shorter in length than the pulse to be measured, but not infinitely short. This is an important point: an infinitely short gate pulse yields only the intensity $I(t)$ and conversely, a CW gate yields only the spectrum $I(\omega)$. On the other hand,

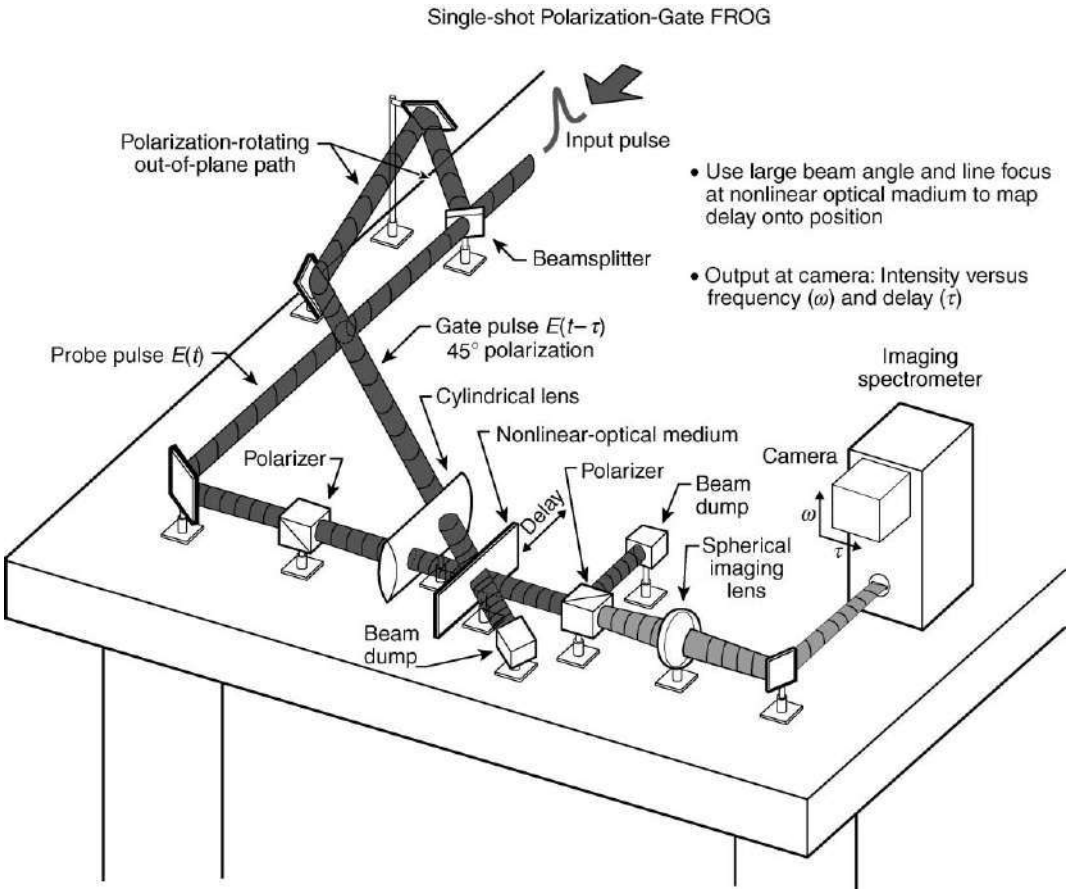

Figure 6 Measuring the spectrogram of a pulse is easier than measuring its sonogram – a spectrogram is a spectrally resolved autocorrelation. In this figure, the optical Kerr–effect is used (polarization-gate) as the nonlinearity. (Adapted with permission from Kane DJ and Trebino R (1993) Single shot measurement of the intensity and phase of an arbitrary ultrashort pulse by using frequency-resolved optical gating. *Optics Letters* 18(10): 823–825.)

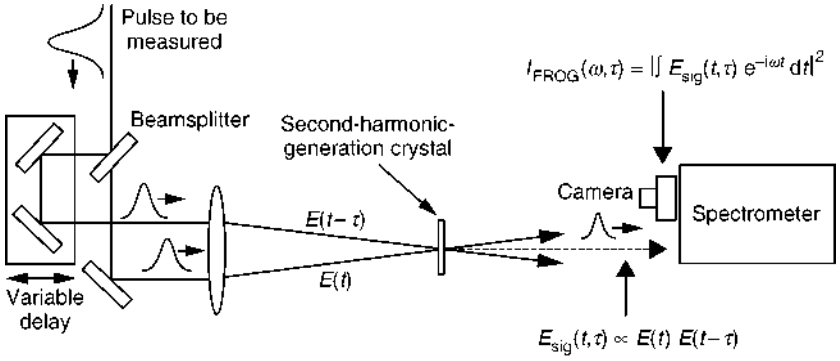

Figure 7 This figure shows a schematic of an SHG FROG device. The SHG signal from the autocorrelation is spectrally resolved. SHG FROG is very simple and sensitive, but it has a direction-of-time ambiguity. For example, if the pulse has chirp, only the magnitude of the chirp is determined – the sign of the chirp remains unknown.

a finite-length gate pulse yields the spectrum of all of the finite pulse segments with duration equal to that of the gate. While the phase information remains lacking in each of these short-time spectra, having spectra of an infinitely large set of pulse segments compensates for this loss. The spectrogram has been shown to nearly uniquely determine both the intensity $I(t)$ and phase $\varphi(t)$ of the pulse, even if the gate pulse is longer

than the pulse to be measured (although if the gate is too long, sensitivity to noise and other practical problems arise).

In FROG, when using optically induced birefringence as the nonlinear effect, the signal pulse is given by:

$$E_{\mathrm{sig}}(t,\tau) \propto E(t)|E(t-\tau)|^2 \qquad [10]$$

So the measured signal intensity $I_{\mathrm{FROG}}(\omega,\tau)$, after the spectrometer is:

$$I_{\mathrm{FROG}}(\omega,\tau) = \left|\int_{-\infty}^{\infty} E(t)|E(t-\tau)|^2 \mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t\right|^2 \qquad [11]$$

We see that the FROG trace is a spectrogram of the pulse $E(t)$ although the gate, $|E(t-\tau)|^2$, is a function of the pulse itself.

To see that the FROG trace essentially uniquely determines $E(t)$ for an arbitrary pulse, it is first necessary to observe that $E(t)$ is easily obtained from $E_{\mathrm{sig}}(t,\tau)$. Then it is simply necessary to write eqn [11] in terms of $E_{\mathrm{sig}}(t,\Omega)$, the Fourier transform of the signal field $E_{\mathrm{sig}}(t,\tau)$, with respect to delay variable τ. We then have what appears to be a more complex expression, but one that will give us better insight into the problem:

$$I_{\mathrm{FROG}}(\omega,\tau) = \left|\int_{-\infty}^{\infty} E_{\mathrm{sig}}(t,\Omega)\mathrm{e}^{-\mathrm{i}\omega t-\mathrm{i}\Omega\tau)}\mathrm{d}t\,\mathrm{d}\Omega\right|^2 \qquad [12]$$

Equation [12] indicates that the problem of inverting the FROG trace $I_{\mathrm{FROG}}(\omega,\tau)$, to find the desired quantity $E_{\mathrm{sig}}(t,\Omega)$, is that of inverting the squared magnitude of the two-dimensional (2D) Fourier transform of $E_{\mathrm{sig}}(t,\Omega)$. This problem, which is called the 2D phase-retrieval problem, is well known in many fields, especially in astronomy, where the squared magnitude of the Fourier transform of a 2D image is often measured. At first glance, this problem appears unsolvable; after all, much information is lost when the magnitude is taken. It is well known that the one-dimensional (1D) phase retrieval problem is unsolvable (for example, infinitely many pulse fields give rise to the same spectrum). Intuition fails in this case, however; two- and higher-dimensional phase retrieval essentially always yields unique results.

The Simplified FROG – GRENOUILLE

A simplified version of the FROG device, known as grating eliminated no-nonsense observation of ultrafast incident laser light e-fields (GRENOUILLE), can be constructed using the doubling crystal as both the gating medium and the spectrometer (see Figure 8). In this case, the phase matching condition in certain thick doubling crystals causes a wavelength dependent output from the doubling crystal. A cylindrical lens is used in a Fourier transform configuration to image the wavelength variation onto a CCD camera. A cylindrical lens set at 90 degrees to the Fourier transform lens images the time axis onto the CCD camera. A Fresnel biprism replaces the traditionally used Mach–Zender interferometer to produce two pulses propagating at an angle with respect to each other. Presently, the GRENOUILLE technique works with BBO, in the wavelength region around 500 nm–1200 nm, and with proustite in the wavelength region around 1200 nm–2000 nm.

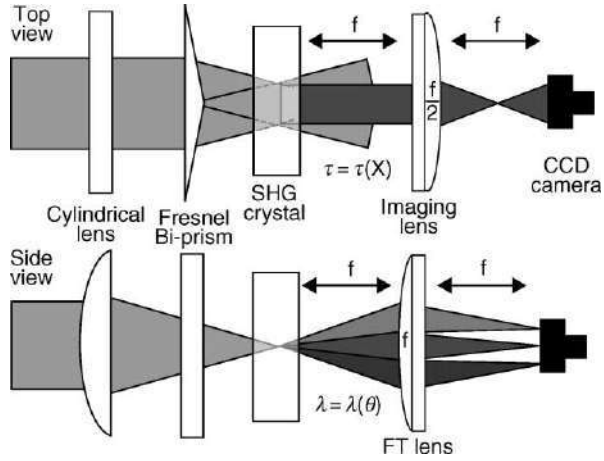

Figure 8 Figure showing a schematic of a GRENOUILLE FROG device which uses a thick doubling crystal as both the gating nonlinearity and the spectrometer. Like single-shot autocorrelators, a cylindrical lens focuses a spatially large input beam into an SHG crystal in only one dimension. Unlike most autocorrelators, the Mach–Zender interferometer is replaced with a Fresnel biprism that forces each half of the input beam to propagate, at an angle, toward the SHG crystal. An imaging cylindrical lens images the spatial dependence of the delay onto a CCD camera. A Fourier transform cylindrical lens maps the angular dependence of wavelength from the SHG crystal to a spatial coordinate on the CCD camera. (Adapted with permission from O'Shea P, Kimmel M, Xun G and Trebino R (2001) Highly simplified device for ultrashort-pulse measurement. *Optics Letters* 26(12): 932–934.)

FROG Inversion Algorithms

An iterative 2D phase retrieval algorithm is required to extract the pulse information from the measured FROG trace (see Figures 9 and 10). This algorithm converges to a pulse that minimizes the difference

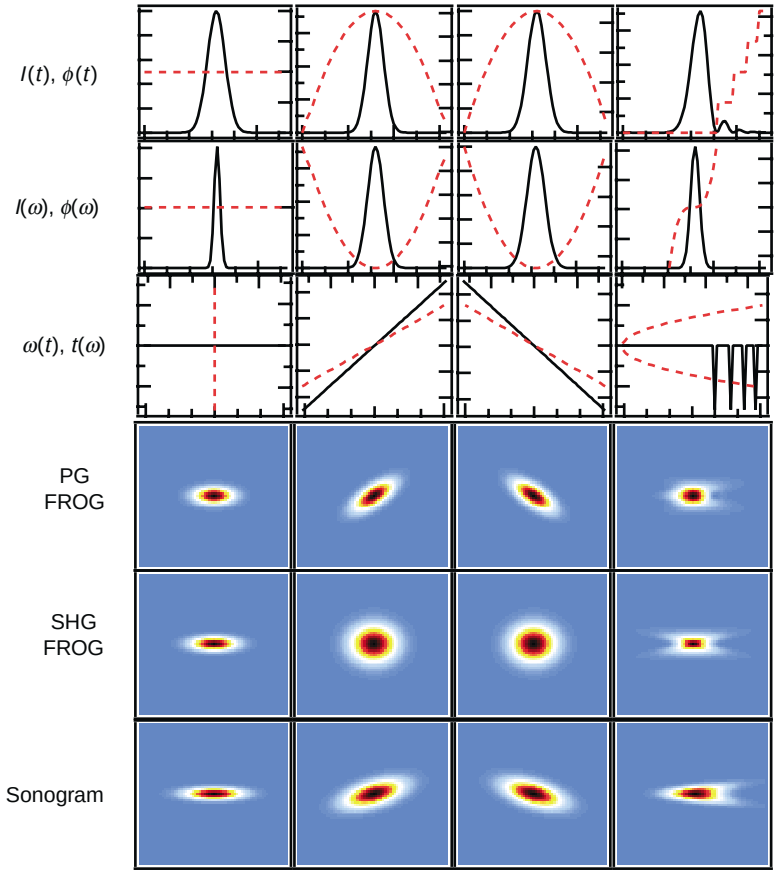

Figure 9 Example FROG traces and sonograms are shown for four different pulses: a transform limited pulse, a positively chirped pulse, and a negatively chirped pulse. The top series of plots are the time domain representation of the pulse intensity (solid line) and phase (dashed line). The second row of four plots is the frequency domain representation of the sample pulse's intensity and phase. The next four plots show the instantaneous frequency and the group delay. The vertical axis is frequency and the horizontal axis is time. The remaining rows show the PG FROG traces, the SHG FROG traces and sonograms of the pulse shown in the first two rows.

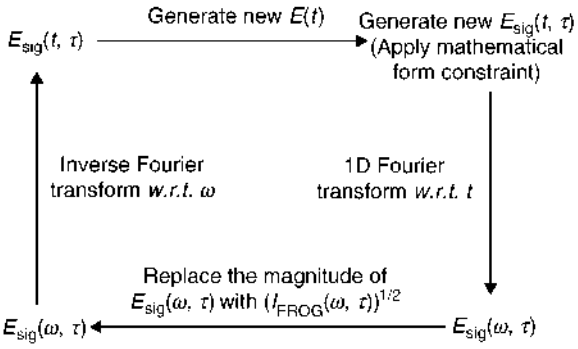

Figure 10 Phase retrieval algorithm for the inversion of FROG spectrograms. Start with an initial guess for the pulse to generate an initial $E_{sig}(t,\tau)$. A 1D Fourier transform generates the FROG trace. The next step is to replace the magnitude of the calculated FROG trace with the square root of the measured FROG trace. Inverse Fourier transform with respect to ω to produce the new signal field and generate a new guess for $E(t)$. Interestingly, it is only the step that produces $E(t)$ from $E_{sig}(t,\tau)$ that differentiates all the FROG algorithms.

between the measured and the calculated FROG trace. While this aspect of FROG has been its Achilles heel in the past, in reality, new generalized projections algorithms (together with faster computers) converge quickly and can track pulse changes at rates up to 30 Hz and beyond, making FROG a true, real-time pulse measurement technique. Indeed, excellent algorithms for the analysis of FROG traces in real-time are commercially available.

The original FROG inversion algorithm, commonly referred to as the vanilla algorithm, is simple and iterates quickly, but tends to stagnate, giving erroneous results, especially for geometries that use a complex gate function such as SHG or self-diffraction. An improved algorithm, incorporating several different algorithms including brute force minimization, was developed to alleviate stagnation programs at the expense of both speed and convergence time. By the mid-1990s, a significant

advance in both speed and stability was made by the addition of a numerical method called generalized projections. This algorithm determines the next guess by constructing a projection – minimizing the error (distance) between the FROG electric field, $E_{\text{sig}}(t, \tau)$, obtained immediately after the application of the intensity constraint, and the FROG electric field calculated from the mathematical form constraint.

The first generalized projections algorithm used a standard minimization procedure to find the electric field for the next iteration, which can still be slow. For the most common FROG geometries, PG and SHG, a different algorithm that determines the next iteration directly has been developed. This algorithm, called the principal components generalized projections (PCGP) algorithm, converts the generalized projections algorithm into an eigenvector problem. Using this algorithm, pulse measurement rates of at least 30 Hz have been achieved.

The goal of the phase retrieval algorithm is to find the $E(t)$ that satisfies two constraints. The first is the FROG trace itself which is the magnitude squared of the 1D Fourier transform of $E_{\text{sig}}(t, \tau)$:

$$I_{\text{FROG}}(\omega, \tau) = \left| \int_{-\infty}^{\infty} E_{\text{sig}}(t, \tau) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t \right|^2 \qquad [13]$$

The other constraint is the mathematical form of the signal field, $E_{\text{sig}}(t, \tau)$, for the nonlinear interaction used. The signal forms for a variety of FROG beam geometries are:

$$E_{\text{sig}}(t, \tau) \propto \left\{ \begin{array}{ll} E(t)|E(t-\tau)|^2 & \text{PG FROG} \\ E(t)^2 E^*(t-\tau) & \text{SD FROG} \\ E(t)E(t-\tau) & \text{SHG FROG} \\ E(t)^2 E(t-\tau) & \text{THG FROG} \end{array} \right\} \qquad [14]$$

where PG is polarization gate, SD is self-diffraction, SHG is second harmonic generation, and THG is third harmonic generation FROG.

All FROG algorithms work by iterating between two different data sets: the set of all signal fields that satisfy the data constraint, $I_{\text{FROG}}(\omega, \tau)$, and the set of all signal fields that satisfy eqn [14]. The difference between the FROG algorithms is how the iteration between the two sets is completed. In the case of generalized projections, the $E(t)'$s are chosen such that the distance between the $E(t)$ on the magnitude set and the $E(t)$ on the mathematical form set is minimized. This is accomplished by minimizing the following equation:

$$Z = \sum_{i,j=1}^{N} \left| E^{(k)}_{\text{sig(DC)}}(t_i, \tau_j) - E^{(k+1)}_{\text{sig(MF)}}(t_i, \tau_j) \right|^2 \qquad [15]$$

where $E^{(k)}_{\text{sig(DC)}}(t_i, \tau_j)$ is the signal field generated by the data constraint, and $E^{(k+1)}_{\text{sig(MF)}}(t_i, \tau_j)$ is the signal field produced from one of the beam geometry equations in eqn [14]. For the normal generalized projection, the minimization of Z is completed using a standard steepest descent algorithm; the derivative of Z with respect to the signal field is computed to determine the direction of the minimum. The computation of the derivatives are tedious; they are tabulated elsewhere.

An alternative to an algorithm that minimizes eqn [15] is the PCGP algorithm. This algorithm converts the computation of the next guess into an eigenvector problem, reducing the computation of the next guess to simple matrix–vector multiplications. This algorithm works for both the PG and SHG beam geometries, it is robust and the fastest FROG algorithm. Indeed, the PCGP algorithm was used in the original real-time FROG work by D.J. Kane.

Self Checks in FROG Measurements

Unlike any other pulse measurement techniques, FROG can provide a great deal of feedback about both the quality of the measurement (systematic errors) and the quality of the algorithm's performance. The most common check for convergence is the FROG trace error together with a visual comparison between the retrieved FROG trace and the measured FROG trace. The FROG trace error is given by:

$$G = \sqrt{\frac{1}{N^2} \sum_{i,j=1}^{N} \left| I_{\text{FROG}}(\omega_i, \tau_j) - \alpha I^{(k)}_{\text{FROG}}(\omega_i, \tau_j) \right|^2} \qquad [16]$$

where α is a renormalization constant, I_{FROG} is the measured trace, and $I^{(k)}_{\text{FROG}}(\omega, \tau)$ is computed from the retrieved electric field. Typically, the FROG trace error of a PG measurement should be less than 2% for a 64×64 pixel trace, while the FROG trace error of a 64×64 pixel SHG FROG trace should be about 1% or less. Acceptable FROG trace errors decrease as FROG trace size increase for smaller FROG traces. These values are only rules of thumb only;

for example, acceptable retrievals of large and very complicated FROG traces can produce larger FROG trace errors.

In a good FROG measurement, the spectrum of the retrieved pulse should faithfully reproduce the salient features of the pulse's measured spectrum. SHG FROG even provides an additional check called the frequency marginal. The sum of an SHG FROG trace along the time axis yields the autoconvolution of the pulse's spectrum, providing two ways the FROG measurement can be checked. First, the autoconvolution of an independently measured spectrum can be compared to the sum of the FROG trace along the time axis, providing an indication of how well the measurement was made. For example, if the doubling crystal was too thick in the pulse measurement, the FROG trace's frequency marginal will be narrower than the autoconvolution of the measured spectrum. Second, comparing the autoconvolution of the retrieval pulse spectrum with the FROG trace marginal can provide a test of algorithm convergence in addition to a test of the measurement. Phase matching problems appear as a mismatch between the FROG trace frequency marginal and the autoconvolution of the retrieved spectrum.

Because FROG is a spectrally resolved autocorrelation, summing any FROG trace along the frequency axis yields the autocorrelation of the measured pulse. This autocorrelation can be compared to an independently measured autocorrelation, or a comparison can be made between the frequency sum of the FROG trace and the autocorrelation calculated from the retrieved pulse to determine algorithm convergence and the quality of the measurement.

Measuring Pulses Directly – SPIDER

SPIDER is a novel technique that measures the pulse directly – no iterative phase retrieval method is required. To accomplish this feat, spectral interferometry is conducted on two pulse replicas that are shifted in frequency with respect to one another. The original pulse is split into two replicas. The two pulse replicas are sent into a phase modulator that shifts the center frequency of each pulse slightly. The frequency-shifted pulses are sent into a spectrometer and a standard spectral interferometry analysis yields the derivative of the phase. Integration of the phase derivative, together with the spectrum of pulse, provides the full intensity and phase.

Thus, analysis of the spectral interferogram gives:

$$\Delta\phi = (\phi(\omega_1) - \phi(\omega_2))\Delta\omega \qquad [17]$$

where $\Delta\phi$ is the measured phase difference from the spectral interferogram, $\Delta\omega$ is the difference in the center frequency between the two pulses, and $\phi(\omega)$ is the pulse phase as a function of frequency. Consequently, the phase difference must be divided by the frequency difference between the two pulses to determine the true derivative. It is also important to subtract out the linear phase difference resulting from the delay between the two pulses (the delay produces the fringes in the resulting interferogram) as this integrates to a linear chirp.

Because phase modulators are too slow to be used in the femtosecond region, a different method is used to shift the frequency of the two pulses – they are mixed with a highly chirped pulse (see Figure 11).

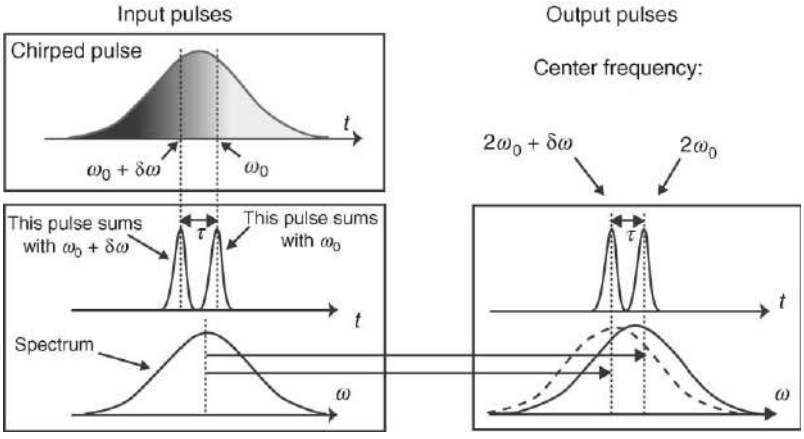

Figure 11 Figure showing how a SPIDER device works. The highly chirped pulse has a frequency that varies with time. Mixing the chirped pulse with two time delayed replicas produces two pulses, separated in time, that have slightly different frequencies, but are otherwise identical. By interfering the two pulses in a spectrometer, the relative phase between $\omega_0 + \delta\omega$ portions of the pulse and ω_0 portions of the pulse are determined.

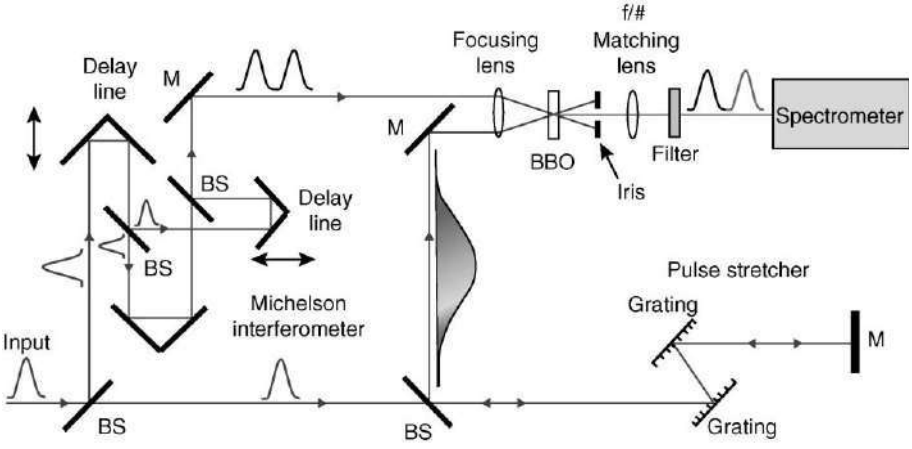

Figure 12 SPIDER apparatus. The input pulse is first split into two pulses. One pulse is stretched in a pulse stretcher. The other pulse is sent into a Michelson interferometer to produce two time-delayed pulses. The two time-delayed pulses are mixed with the stretched pulse to produce two pulses at a slightly different center frequency. The two pulses are sent into a spectrometer to produce a spectral interferogram. Analysis of the spectral interferogram yields the derivative of the pulse phase.

Because the pulses are slightly time delayed, each one aligns with a slightly different frequency in the chirped pulse. The chirped pulse must be so highly chirped that the pulse frequency does not change significantly over the duration of the pulse to be measured.

A schematic of a SPIDER device for the measurement of femtosecond laser pulses is shown in Figure 12. The input pulse is first split into two replicas. The first replica is sent to a pulse stretcher to produce to highly chirped pulse. The other replica is sent into a Michelson interferometer (or a simple étalon may sometimes suffice) to create two time-delayed replicas. The time-delayed replicas are mixed with the chirped pulse. Because each replica is mixed with a slightly different frequency in the chirped pulse, each replica produces a pulse with a slightly different center frequency. When the time-delayed replicas are spectrally resolved, each frequency of the pulse interferes with a frequency-shifted portion of the same pulse. Thus, the measured phase difference is proportional to the derivative of the phase of the original pulse.

Measuring Weak Pulses

Because virtually all the pulse measurement techniques, that do not require a reference pulse, use a nonlinearity, measuring weak ultrafast pulses directly is difficult. Fortunately, because most weak pulses are generated from a stronger pulse, spectral interferometry can be used to determine the relative phase between the weaker pulse and the stronger pulse (for regions where there is spectral overlap between the two pulses). If the stronger pulse is characterized using another pulse measurement technique, then the weaker pulse phase can be determined from the relative phase measurement and the known phase of the stronger pulse. For example, the Trebino group termed the combination of FROG and SI as TADPOLE (Temporal Analysis by Dispersing a Pair Of Light E-fields).

Which Pulse Measurement Method Should I Use?

Choosing a pulse measurement method depends on the required measurement as well as the precision and accuracy required. All the pulse measurement techniques have advantages and disadvantages, but excellent measurements have been made with all of them. Autocorrelation is adequate when all that is required is the approximate pulse duration without any phase information. Spectral interferometry is quite useful for measuring transient changes in a sample in pump-probe experiments. FROG is perhaps the most commonly used and general-purpose pulse measurement technique, providing a great deal of feedback about the quality of the pulse measurement. Inexpensive and convenient FROG devices, together with software, are available commercially, adding to their acceptance. While SPIDER is experimentally complex and does not have the cross checks that FROG has, it provides the quickest answer, requiring no iterative

phase retrieval algorithm to obtain pulse intensity and phase.

Acknowledgments

This material is based in part upon work supported by the National Science Foundation under Grant numbers DMI-9801116 and DMI-0091454. The author would also like to acknowledge Rick Trebino for the use of figures from his lectures.

List of Units and Nomenclature

$A(t)$	Full electric field
$E(t)$	Complex electric field without carrier frequency
$E_{\mathrm{sig}}(t,\tau)$	Complex electric field of the FROG trace versus time and time delay
$E_{\mathrm{sig}}(t,\Omega)$	Fourier transform of $E_{\mathrm{sig}}(t,\tau)$ with respect to τ
$\tilde{E}(\omega)$	Complex electric field in the frequency domain
$f(t)$	Correlation product
$f(\omega)$	Fourier transform of the correlation product
$g(t-\tau)$	time gate or filter
G	FROG trace error
$h(\omega-\Omega)$	frequency gate or filter
$I(t)$	Intensity profile of the electric field in the time domain
$\tilde{I}(\omega)$	spectral intensity
$I_{\mathrm{FROG}}(\omega,\tau)$	Intensity of the FROG trace
$\Theta(t)$	Heaviside function $= 0\ t<0,\ = 1\ t \geqq 0$
$\phi(\omega)$	Frequency domain phase
$S_{\mathrm{E}}(\omega,\tau)$	Spectrogram
$S_{\mathrm{E}}(\Omega,T)$	Sonogram
$S(\omega)$	Spectral interferogram
t, T	variables for time
τ	Fourier transform
v_{g}	group velocity
$\varphi(t)$	Time domain phase
Δw	Beam waist
Z	Distance metric between the FROG signal field generated by the data constraint and the FROG signal field generated by the mathematical form constraint
Φ	Geometric angle between two beams
$\Delta\phi$	Phase difference
$\Pi^{(n)}$	nth order nonlinear susceptibility
ω	Primary variable for frequency
$\Delta\omega$	Difference in center frequency between two pulses
Ω	Secondary variable for frequency
$\Omega(t)$	Instantaneous frequency

See also

Coherent Transients: Ultrafast Studies of Semiconductors. **Nonlinear Optics, Basics:** Ultrafast and Intense-Field Nonlinear Optics. **Ultrafast Laser Techniques:** Generation of Femtosecond Pulses. **Ultrafast Technology:** Femtosecond Chemical Dynamics: Gas-Phase; Femtosecond Condensed Phase Spectroscopy: Structural Dynamics; Ultrafast Illumination and Processing.

Further Reading

Diels JC and Rudolph W (1996) *Ultrashort Laser Pulse Phenomena*. San Diego, CA: Academic Press Inc.

Dorrer C, Belabas N, Likforman J-P and Joffre M (2000) Spectral resolution and sampling issues in Fourier-transform spectral interferometry. *Journal of the Optical Society B* 17: 1795.

Kane DJ, Weston J and Chu K-CJ (2003) Real-time inversion of polarization gate frequency-resolved optical gating spectrograms. *Applied Optics* 42: 1140–1144.

Iaconis C and Walmsley IA (1999) Self-referencing spectral interferometry for measuring ultrashort optical pulses. *IEEE Journal of Quantum Electronics* 35: 501–509.

Lepetit L, Cheriaux G and Joffre M (1995) Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy. *Journal of the Optical Society of America B* 12: 2467–2474.

Rulliere C (ed.) (2000) *Femtosecond Laser Pulses*. New York: Springer Verlag.

Shuman TM, Anderson ME, Bromage J, Iaconis C, Waxer L and Walmsley IA (1999) Real-time SPIDER: ultrashort pulse characterization at 20 Hz. *Optics Express* 5: 134–143.

Treacy EB (1971) Measurement and interpretation of dynamic spectrograms of picosecond light pulses. *Journal of Applied Physics* 42: 3848–3858.

Trebino R (2000) *Frequency Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses*. Boston, MA: Kluwer Academic Publisher.

Trebino R, DeLong KW, Fittingholf DN, Sweetser JN, Krunbugel MA, Richman BA and Kane DJ (1997) Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating. *Review of Scientific Instruments* 68: 3277–3295.

Various authors (1999) Feature issue on ultrashort-laser-pulse measurement and applications. *IEEE Journal of Quantum Electronics* 35: 4.

ULTRAFAST TECHNOLOGY

Contents

Femtosecond Chemical Dynamics: Gas-Phase
Femtosecond Condensed Phase Spectroscopy: Structural Dynamics
Ultrafast Illumination and Processing

Femtosecond Chemical Dynamics: Gas-Phase

M Dantus, Michigan State University, East Lansing, MI, USA

E J Brown, McDaniel College, Westminster, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Chemical reactions are at the heart of chemistry; the quest to understand them has sparked research and study along many different avenues. The fundamental dynamics involved in bond breakage, bond formation, and transition states take place on a very fast time-scale, usually in the tens to hundreds of femtoseconds, where a femtosecond (fs) is equal to 10^{-15} seconds. Therefore, ultrashort laser pulses are required to measure these fundamental processes as well as the rotational and vibrational motion before and after the reaction. The whole set of measurements can then be used for mapping a potential energy surface for the reaction. As a result of these studies, one learns the key details that are required to develop a picture of how chemical reactions occur. The measurement of these fundamental dynamics, using ultrashort laser pulses, is the subject of this article.

To illustrate how long it takes to form or break a chemical bond, one needs to consider the following analogy. The time it takes to break a chemical bond is approximately a tenth of a trillionth of a second (10^{-13} s). During such a short time, light can travel only 0.03 mm in vacuum. This time is so short that it cannot be measured electronically. Its measurement requires the use of ultrashort laser pulses, with pulse durations of 50 femtoseconds or shorter. The development of methods aimed at measuring the fundamental steps involved in chemical reactions culminated in the development of femtosecond transition state spectroscopy (FTS), a method developed in the 1980s by Zewail and co-workers at Caltech. This work allows researchers to observe bond breakage and formation, vibrational and rotational dynamics, and transition state configurations, as the reaction occurs. The Nobel Prize in Chemistry was awarded in 1999 to Ahmed Zewail for this development. Since the 1980s, femtosecond lasers have been used to study molecular dynamics and chemical dynamics on very fast time-scales. The field, known as femtochemistry, has grown quickly and diversified to involve hundreds of scientists around the world. This article discusses the measurement of ultrafast dynamics measured in gas-phase samples. The presentation encompasses some of the measurement techniques and the fundamental values to be measured and concludes with an outlook towards the future.

From Nanosecond to Picosecond to Femtosecond Reaction Kinetics

The quest for shorter light pulses to study chemical reactions predates the invention of the laser. In the early days, electric discharges were used to measure fast dynamics in the millisecond (10^{-3} s) and then the microsecond (10^{-6} s) time-scales. With the invention of the laser, measurements improved drastically with temporal resolution going from the nanosecond (10^{-9} s) to the picosecond (10^{-12} s) time-scales. Since the early days, it was recognized that the best time resolution was obtained when pairs of light pulses were used – one to initiate the reaction and the second to probe it. These pump-probe measurements have progressed with the laser technology almost to the single femtosecond level.

The measurements being considered here go beyond the measurement of fast reaction kinetics. When a chemical reaction occurs in a beaker, one usually measures a statistical rate of reaction that involves diffusion of reagents in a solvent. These kinetic measurements, which involve an ensemble of molecules, are much slower and do not reveal the

fundamental steps of the reaction such as bond formation. The introduction of ultrafast pulses allowed, for the first time, the measurement of the motion of the atoms in real time, as if the motion was frozen by ultrafast flash photography. Capturing this motion helps to reveal the internal forces, described by a potential energy surface, that act on the atoms in the molecule during the chemical reaction. This concept is illustrated in Figure 1, where we see the reagents and the products for the reaction. Typically, the starting and ending point of a chemical reaction are very well known and the compounds are very well characterized. However, how the reagents become the products is not always known. Chemists typically deduce a reaction mechanism based on a large body of experiments where different parameters including the structure of the reagents are modified to evaluate their effect on the outcome of the reaction. Unfortunately in most cases, reaction mechanisms, even when consistent with all available experimental evidence, may not accurately reflect how a particular chemical reaction takes place.

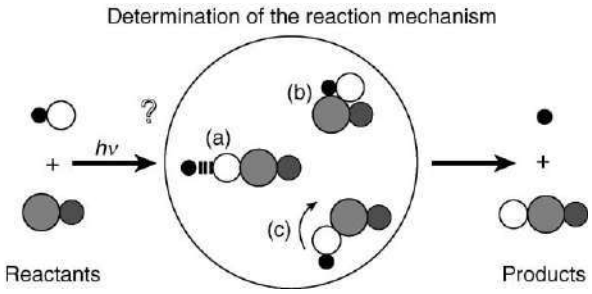

Figure 1 A sketch of a chemical reaction, AB + CD → A + BCD. The reactants and products are well known, but it is not known how they change from reactants to products. This mechanism may involve a transition state which is (a) a linear complex with vibrational motion, (b) a nonlinear complex, (c) a bent complex with rotational motion, or many other possibilities and combinations. The elucidation of the mechanism and transition state dynamics are the goals of femtosecond time-resolved studies.

The ideal method to make the determination involves direct observation. As mentioned earlier, direct observation requires femtosecond laser pulses, just like a fast camera shutter is required to take pictures of fast-moving objects. The concept of a femtosecond pump-probe measurement is illustrated in Figure 2. The chemical reaction is initiated at time $t = 0$ fs by the pump laser. The pump laser provides the energy required for initiating the chemical transformation. The probe laser, delayed in time, probes the formation of the product. Figure 2 illustrates a number of pump-probe measurements

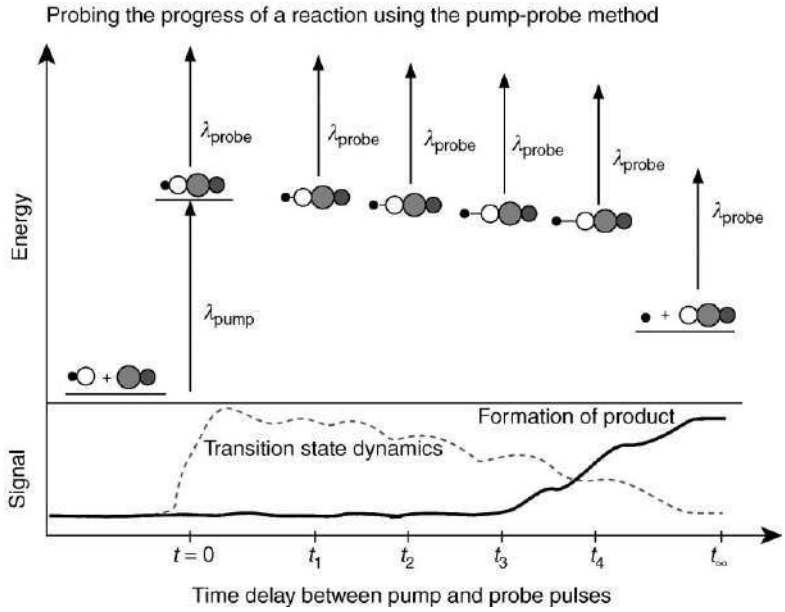

Figure 2 Energy schematic of the reaction of AB + CD → A + BCD assuming vibrational motion in a linear transition state. The pump pulse provides enough energy to excite the reactants to the transition state at $t = 0$. This transition state is probed by the second pulse. This pulse arrives at different time delays, from no delay ($t = 0$), to increasing delays (t_1, t_2,…), to infinite delay (t_∞). The signal varies depending on the time delay between the two pulses and these changes in signal intensity correspond to vibrational and rotational information about the transition state. To obtain a transient as shown at the bottom, the time delay between the pump and probe pulses is scanned. The dashed line shows the transient that would be obtained when the transition state is monitored; the solid line shows the transient that would be obtained when the product is monitored.

each obtained at a different time delay between the pump and probe laser pulses. When a series of pump-probe measurements are gathered as a function of time delay, one obtains a transient that contains a record for the evolution of the reagents to product as a function of time.

The Formation of Coherent Wave Packets and their Motion

The illustration shown in Figure 2 shows atoms in a molecule moving as classical particles. Because of their small size, atoms behave quantum mechanically and this has some implications on the observed motion. A prototypical pump-probe measurement is illustrated for this purpose in Figure 3. For resonant excitation, a photon is absorbed by a molecular system, originally in the ground electronic state (V_0). This absorption causes a vertical transition to an excited electronic state (V_1). Consistent with Heisenberg's uncertainty principle, a pulse that has a short temporal duration has a large spectral bandwidth; for most molecular systems and ultrashort pulses, the frequency spread of the pulse is wider than the vibrational energy spacing in the excited state. Thus, a number of quantum vibrational states are populated on V_1 by the absorption of a short laser pulse. When a short pulse excites a number of states, a coherent superposition of states is formed. One can understand the formation of a wave packet by simply realizing that in the short time the laser pulse excites the system there is no opportunity for the atoms in the molecule to move. For this reason the initial wave packet is very similar to the ground state atomic arrangement.

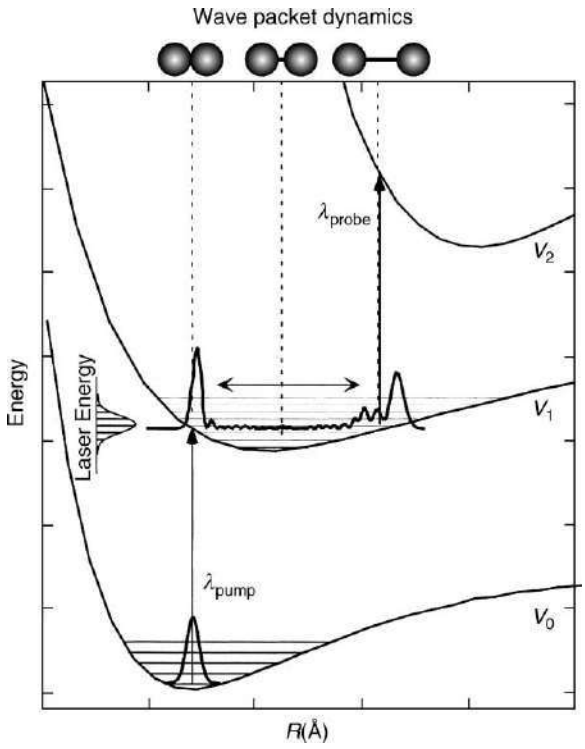

Figure 3 Pump-probe experiment. The pump pulses excite the molecular system from the ground state V_0 to excited state V_1. The large spectral bandwidth of femtosecond pulses allows for a number of vibrational levels in V_1 to be populated. A wave packet is formed that vibrates on V_1; the classical motion of the particle is shown as well at the top. In this case the energy of the probe pulse will excite the molecules to V_2 when the bond is stretched (right side of oscillation); however, when the bond is compressed, the energy is not resonant with any excitation and the probe pulse will not be absorbed.

When a short laser pulse excites a molecule to a bound electronic state, the resulting wave packet, $\Psi(t)$, is a sum of all the possible states φ_n in V_1 with amplitudes, a_n, $\Psi = \sum_n a_n \varphi_n$. The amplitudes, a_n, depend on the overlap between the initial (ground state) and final (excited state) spatial overlap. This overlap indicates that the excitation process is faster than the motion of atoms in a molecule. Based on this approximation for the overlap integral, the amplitudes are replaced by the Franck–Condon overlaps to obtain $\Psi = C\sum_n \langle \varphi_n | \varphi_0^{(0)} \rangle \varphi_n$. The wave packet on V_1 evolves in time according to $\Psi(t) = \sum_n a_n \exp[-iE_n t/\hbar]\varphi_n$, where E_n is the energy that corresponds to the state φ_n and $\hbar$ is Planck's constant divided by 2π. When a pulse is very short in time, the coherent superposition of states closely resembles the initial distribution of atoms in the molecule. The motion from inner to outer turning point of the quantum mechanical wave packet is illustrated in Figure 3. One can imagine this wave packet as a classical particle oscillating on the excited potential energy curve with the periodic nature of the molecular vibrations. When the pulse is very long and only one state is populated, the dynamics can no longer be observed, as is the case with conventional frequency-resolved spectroscopy.

Molecular dynamics can be inferred from high-resolution frequency-resolved spectroscopy; however, the spectra can become congested and the data analysis becomes more complicated for large molecules, for high temperatures, or in solutions. Molecular dynamics are obtained from frequency-resolved spectra by taking the Fourier transform of the data. It seems more intuitive to consider these dynamics in the time domain where the dynamics can be observed directly as 'snapshot' taken by the femtosecond pulses. Ultrafast time-resolved spectroscopy has other advantages – high peak intensities allow for multiphoton excitation and other nonlinear processes, the selection of the detection wavelength discriminates between different species and results in

an increased signal-to-noise ratio for intermediate short-lived species, collisions take place on a longer time-scale and do not affect these measurements, and the characteristics of the ultrafast laser pulse (such as phase, frequency components, and chirp) can be manipulated to control the dynamics and the yield of chemical reactions.

Methods for Measuring Femtosecond Dynamics

Pump-Probe Method

Pump-probe techniques use a pair of pulses separated in time, one to initiate the reaction and the second to determine its progress. The first femtosecond pulse with wavelength λ_{pump} excites a molecule from its (bound) ground electronic state (V_0) to an excited (V_1) state. After a variable time delay, a second femtosecond pulse with wavelength λ_{probe} causes a transition to another excited state (V_2) as shown in Figures 3 and 4.

The progress of the reaction dynamics is followed through the measurement of signal resulting from the absorption of the probe pulse by the system. There are different detection modes depending on the potential energy surfaces reached by the pump and probe pulses. In one case (Figure 4a), the first excited state (V_1) is nonfluorescent but the second one (V_2) fluoresces and this fluorescence is measured. In another case (Figure 4b), state V_1 is fluorescent but state V_2 is not; in this case the depletion of the fluorescence by the probe is measured. Finally (Figure 4c), excited states V_1 and V_2 may be dissociative; therefore, the fluorescence detected is emitted by the products. In any of these cases, as the wave packet on the intermediate state propagates (e.g., vibrates, dissociates), the transition probability varies (i.e., the ability of the probe pulse to be absorbed by the molecule changes). Therefore, the amount of laser-induced fluorescence (LIF) emitted varies with the pump-probe time delay. In some cases, the probe pulse produces ions and there are a number of methods to detect the electrons or the ions with great sensitivity.

Measurement of rotational dynamics

Measurement of rotational dynamics requires a vectorial frame of reference and this is achieved with linearly polarized laser beams. Excitation by the pump pulse selects an initial population distribution described by $\cos^2\theta$ where θ is the angle between the transition dipole and the pump laser polarization vector. As the rotational wave packet dephases due to the distribution of the population in different rotational levels, the probe pulse probes the transient alignment of the molecules in space. As the molecules rotate away from the initial alignment, the signal decreases. In Figure 5 rotational dynamics are depicted schematically. First, a single (classical) molecule is considered (top panel). As the molecule rotates, its dipole changes from being aligned with the polarization vector of the laser field

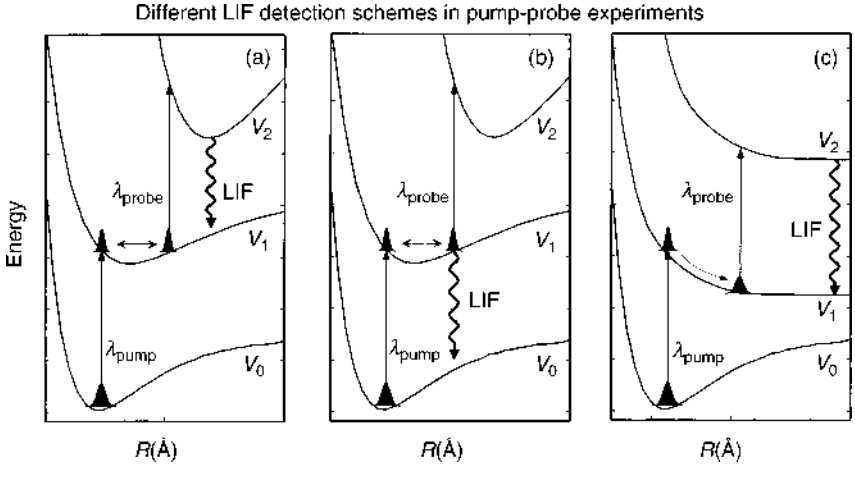

Figure 4 Different pump-probe experiments.(a) Bound-to-bound transitions from $V_0 \rightarrow V_1$ and from $V_1 \rightarrow V_2$. The V_2 state fluoresces and this increase in LIF from V_2 is measured as signal. (b) Bound-to-bound transition from $V_0 \rightarrow V_1$ where V_1 fluoresces. When the probe pulse causes an excitation from $V_1 \rightarrow V_2$, a loss of LIF signal is observed. The depletion of fluorescence is measured in this case. (c) Bound-to-repulsive surface transition from $V_0 \rightarrow V_1$. The energy of the probe causes an excitation to V_2 at a particular molecular distance. Fluorescence of a product fragment is measured.

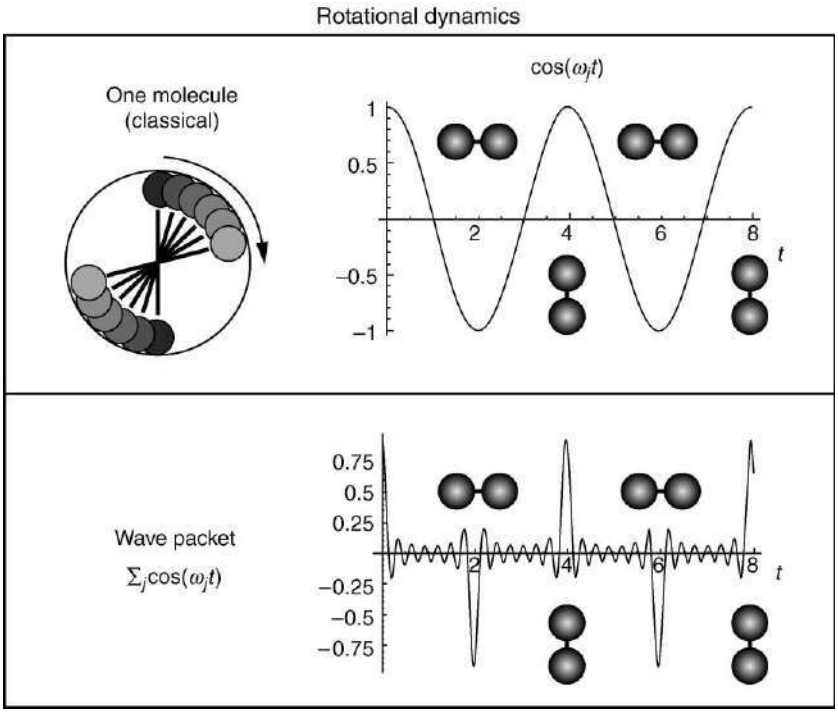

Figure 5 Rotational motion of molecules. Excitation by a linearly polarized laser pulse initially selects molecules that are well aligned ($\cos^2\theta$) with its polarization. Considering only one molecule evolving in time (top), the rotational motion in the molecule causes it to move away from the initial alignments and then eventually return to it. The signal will decrease when the molecule is not aligned with the polarization of the probe laser and will increase when aligned well. The signal looks like a cosine function that depends on the rotational frequency of the molecule (ω_j). For a rotational wave packet, the rotational dynamics are similar except that a number of rotational levels are populated; thus, the signal depends on the summation of many cosine functions and the features in the data become sharper.

(assumed vertical) to being perpendicular. The rotational motion can be described by a cosine function. The rotational frequency ω_j is proportional to the rotational level and the rotational constant of the molecule. Because the rotational constant is inversely proportional to the angular momentum, larger molecules rotate slower than smaller molecules. When several rotational levels are contained in a rotational wave packet the motion produces sharper features (bottom panel). The sharp features are called rotational recurrences or revivals. Note that the opposite trend is observed if the probe pulse is polarized perpendicular to the pump pulse; the signal increases as the molecules rotate away from the initial alignment and decreases as the alignment recurs. The time-resolved data for each polarization arrangement (parallel and perpendicular) contain both an isotropic component (vibrational motion) and an anisotropic component (rotational motion). In rotational anisotropy measurements, data from these two polarization arrangements are manipulated mathematically to yield just the rotational motion; rotational dynamics and rotational information such as the spectroscopic rotational constants and the rotational energy partitioning in the reactants and products can be extracted from these measurements.

Three-Pulse Four-Wave Mixing Method

Some methods are more complicated than pump-probe technique but they are more powerful. Three-pulse four-wave mixing (FWM) is one such method and is briefly presented here. Time-resolved four-wave mixing techniques can be understood based on the formation of transient gratings by two of the incident lasers. When two laser beams cross, the spatial modulation of their electric fields varies due to constructive and destructive interference as shown in Figure 6. The molecules in the interaction region experience varying electric field intensities according to their position and this leads to the formation of a transient grating of polarized molecules in space. The transient grating formation can be probed easily by the detection of Bragg scattering of a third laser beam, also known as homodyne detection; here, the signal beam is scattered in the phase-matching direction $\mathbf{k}_{\text{signal}}$ as indicated in Figure 6a. Alternatively, probing can be achieved by detection of changes in the intensity of a fourth beam that propagates in the

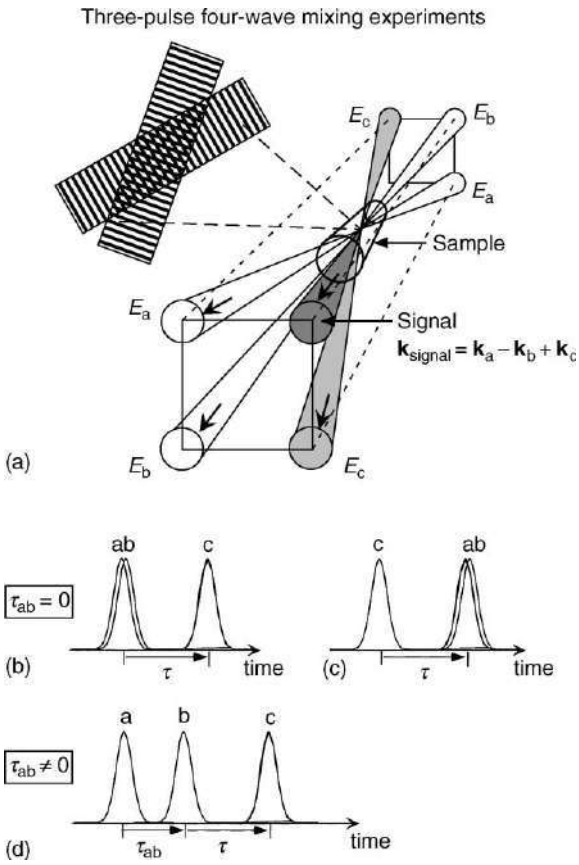

Figure 6 (a) FWM experimental configuration. Two of the beams cross which forms a grating and the third beam scatters off the grating to generate a fourth beam (the signal). The three laser pulses are noted as electric fields E_a, E_b, and E_c. These three laser pulses can be delayed or advanced in time relative to each other giving rise to a number of different pulse sequences. (b) Pulse sequence with fields E_a and E_b overlapped in time ($\tau_{ab} = 0$). Field E_c follows at a time delay τ. (c) Pulse sequence in which E_c precedes E_a and E_b by time delay τ. For both of these cases, τ is scanned in time to obtain the transients. (d) Pulse sequence in which no pulses are overlapped in time and are applied in the order E_a, E_b, and E_c. Time delay τ_{ab} can be set to specific values while τ is scanned in time.

same direction as the signal; this is called heterodyne detection. The medium that constitutes the gratings is composed of the sample molecules, which in the gas phase move freely. Therefore as time evolves, the molecular dynamics (rotation and vibration) modulate the transient grating and hence the signal.

Three-pulse four-wave mixing is a nonlinear spectroscopic method that combines the interaction of three laser pulses in a phase-matched geometry with a well-defined time sequence of the pulses. The signal from these measurements arises from a third-order polarization resulting from the interaction of the three electric fields. Three-pulse FWM is similar to the pump-probe technique in that a preparation step is followed, after some variable time delay, by a probing step. However, three-pulse FWM allows for a greater degree of control over the preparation and probing processes.

When the lasers are in resonance with an electronic state, FWM measurements can provide dynamic information from the ground and excited electronic states. When two pulses are incident on the sample at $t = 0$ and a third pulse probes the dynamics at a later time τ (as in Figure 6b), the measurements provide information that is very similar to that from a pump-probe measurement. However, having three pulses allows for other different setups. For example, when one pulse arrives at the sample first and the other two pulses follow after a time delay (as in Figure 6c), the resulting signal provides dynamic information from the excited state. This type of measurement also provides a measurement of the electronic coherence of the sample, i.e., for how long the molecules remember the phase of the incident laser pulses. This information is important if one wants to take advantage of the phase of laser pulses to cause quantum mechanical interference in the molecular sample. These measurements are known as photon echo and are much more common in condensed-phase samples. There are numerous pulse sequences possible; only three examples are shown in Figure 6.

FWM measurements do not depend on resonance excitation to an electronic excited state. In the absence of such resonance, the grating is formed in the ground state and ground state vibrational and rotational dynamics are observed. These measurements, also known as transient grating (TG), depend on the impulsive Raman scattering process. By changing the time delay (in Figure 6b) between the electric fields that generate the grating and the third (probe) beam, the signal is modulated by the molecular dynamics. Analysis of the time-dependent signal yields the rotational and vibrational spectroscopic values from the sample being studied.

Studies of Reaction Dynamics

Dissociation on a Repulsive Surface

In 1987, Zewail and co-workers studied the photodissociation of cyanogen iodide (ICN) using femtosecond lasers; the first direct observation of a chemical reaction from reactants to products was conducted. The excitation and decay of the reactants (ICN), the formation and decay of the intermediates (I–CN), and the formation of the products (I + CN) were observed in this experiment.

The pump pulse (306 nm) excited ground state cyanogen iodide to a repulsive excited state of ICN which dissociated into I and CN fragments (both in the ground state). The probe pulse when set to 388.5 nm was resonant with an excitation of the CN fragment; therefore, the reaction was monitored by detecting the laser-induced fluorescence (LIF) of this CN^* fragment using a probe pulse at different time delays (as shown in Figure 4c). The wavelength of the probe pulse could be detuned to different wavelengths for making observations at different lengths of the I–C bond. The results of these experimental studies showed that ICN dissociated in about 200 fs and that the transition state lived for only about 50 fs. The data from the wavelength detuning studies were used to map the excited state potential of ICN both in time and in space; this data analysis is called inversion to the potential.

A schematic representation of the ICN experiment is shown in Figure 7. The potential energy curves are shown in Figure 7a. The pump laser initiates the chemical reaction. Three different probe wavelengths are depicted, the first probing early in the reaction, the second at an intermediate distance, and the final one probing the free products. Notice that the wavelength of the probe determines the location in the potential energy curve that is probed. This concept allows one to study reagents as they are transformed into products. The signals obtained from the three different probe wavelengths are shown in Figure 7b.

Reactions Involving Crossing of Potential Energy Surfaces

For many chemical reactions, the progress of the reaction requires changes of the electronic state. This can take place where potential energy surfaces cross. The best-known example of reaction dynamics involves the experiment on the photodissociation of NaI by Zewail and co-workers. This experiment is depicted in Figure 8. The pump pulse excites the wave packet from ground state (ionic) to an excited state (covalent) where it begins to oscillate. The ground and excited states overlap with an avoided crossing in this molecule (ionic versus covalent nature, as shown in Figure 8a); when the wave packet approaches this avoided crossing, it splits into two partial waves. One wave stays on the excited state while the other 'escapes' onto the dissociative surface and separates into Na and I. The partial wave on the excited state again oscillates and then splits at the avoided crossing and the pattern continues. The time delay of the probe pulse shows the oscillations of the wave packet on the excited surface when detecting the activated complex $[NaI]^{*\ddagger}$ (thin line), while the data of the product fragments (free Na atoms detected by using the D-line absorption) (thick line) show a step-wise accumulation in the signal corresponding to the sequential formation of products during each oscillation near the avoided crossing region of the surfaces (see Figure 8b).

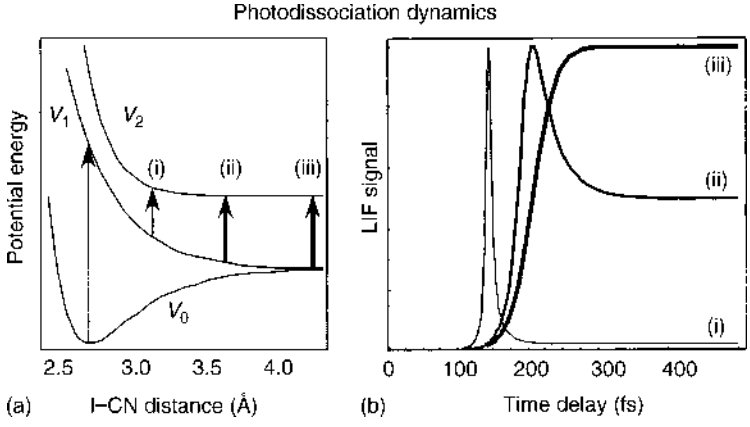

Figure 7 Photodissociation of ICN. (a) Potential energy surface schematic showing the initial excitation by the pump laser to a dissociative surface. The reaction could be monitored at different I–CN distances by using different probe wavelengths as shown by three examples here, (i), (ii) and (iii), from shorter to longer distances. (b) At short distances (long probe wavelengths), the signal is observed at shorter time delays only. The signal decreases at later time delays because the internuclear distance continues to lengthen as the dissociation occurs over time. When the distance increases as monitored by shorter probe wavelengths, the signal rises to a maximum at longer time delays reflecting the dissociation of the reaction in progress. At 'infinite' I–CN distance (i.e., the molecule is dissociated and only free I atoms and CN fragments remain), the signal reaches a maximum and remains there because the reaction is complete.

Chemical Dynamics with Multiple Pathways

The pump-probe concept can be used to probe more complex reactions. To date, a great number of different chemical reactions have been studied (see Further Reading). Among the most important reactions we list ring opening, isomerization, proton transfer, and concerted chemical reactions. We illustrate these measurements with experiments on the photodissociation of dihalogenated alkanes. Upon high-energy excitation (12 eV), the methylene iodide (CH_2I_2) molecule dissociates and forms molecular iodine in excited states D′ and f as well as producing I atoms (main channel). The detection wavelength used in the experiment can distinguish between these products; D′ fluorescence can be collected at 342 nm and f fluorescence is collected at 272 nm. The Dantus group has studied the dynamics of this type of concerted molecular elimination reaction extensively. The photodissociation reaction occurs by a multiphoton excitation (three-photons from the pump beam at 310 nm) reaching a dissociative state where the CH_2 fragment and excited molecular iodine are formed. The 620 nm probe pulse depletes the fluorescence from the excited I_2 molecule (either D′ or f state) as shown in Figure 4b; this fluorescence depletion is measured and analyzed to determine the dynamics of the reaction. The object of this experimental work was to determine if the photodissociation mechanism for forming molecular iodine followed a step-wise, synchronous concerted, or asynchronous concerted mechanism as shown in Figure 9.

The time transients reveal coherent vibrational motion of the I_2 product. Analysis of the transition state dynamics shows that the breaking of the carbon–halogen bond and the formation of the halogen–halogen bond occurs within 50 fs. Further, the iodine molecule is formed with a very hot rotational distribution indicating that the symmetry of the molecule is broken during the process; the torque imparts a large amount of rotational motion. Overall, the results of these studies show that the photodissociation process of CH_2I_2 follows an asynchronous concerted reaction mechanism and corresponds to a symmetry barrier that exists in C_{2v} molecules preventing the formation of the interhalogen bond. The existence of this symmetry barrier was explored by using CH_2ICl, a molecule that does not have C_{2v} symmetry; as expected, the halogen molecule (ICl) was formed with a cold rotational distribution indicating that it was able to follow a synchronous concerted mechanism. The dissociation times were also studied on the family of compounds CX_2Y_2 and $R{-}CHI_2$ to investigate the influence of thermodynamics and reduced mass changes on the reaction.

The Dynamics of Bond Formation

Most chemical reactions that are studied with femtosecond spectroscopy are unimolecular reactions (photodissociation); however, bimolecular reactions (photo-association) are interesting and chemically relevant in trying to understand how chemical reactions occur. In the photo-association process, a pair of atoms or molecules cooperatively absorbs the photon undergoing a free-to-bound transition and forming a bond. When the ultrashort pulse initiates the reaction in femtosecond photo-association spectroscopy, a clearly defined initial time for the reaction is given

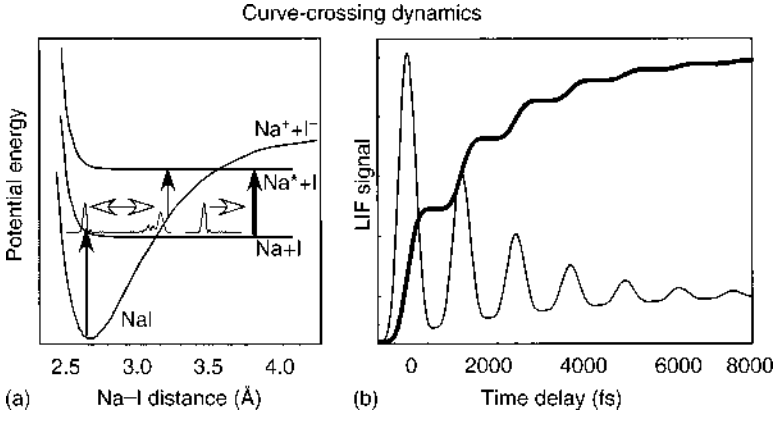

Figure 8 Photodissociation of sodium iodide. (a) Potential energy curve schematic showing the curve-crossing between the ground (ionic) and excited (covalent) states. As the wave packet oscillates on the excited curve, it splits into two wave packets at the curve crossing. (b) The thin line shows the vibrations of the excited complex as it oscillates on the excited surface. The signal intensity decreases because some of the molecules have dissociated to form Na and I atoms when the wave packet splits. The thick line shows the stepwise increase in free Na atom signal each time the wave packet oscillates and splits.

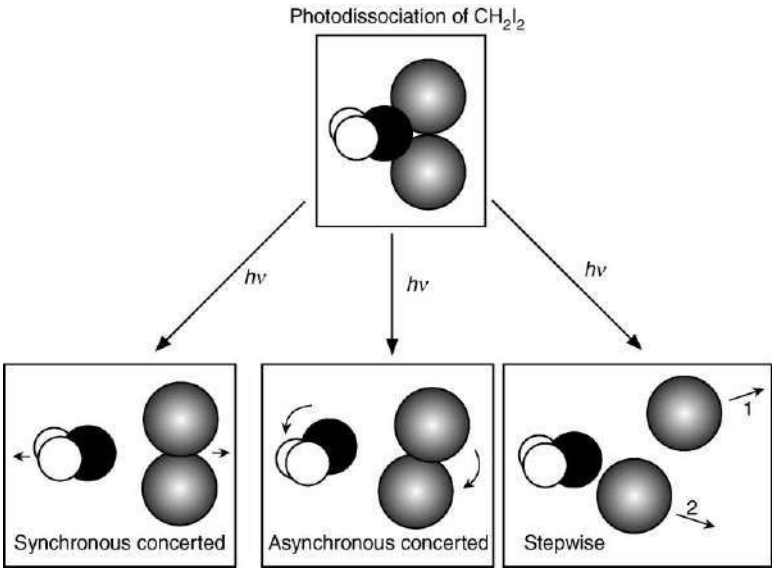

Figure 9 Possible mechanisms for the photodissociation of CH_2I_2 following multiphoton excitation. The molecule could break both C–I bonds and form an I–I bond all at the same time (synchronous concerted mechanism); this would result in very little rotational motion in the products. The molecule could begin to break one C–I bond slightly before breaking the second one and forming the I–I bond, but all three events still occur within the duration of the laser pulse (asynchronous concerted mechanism). This would result in significant rotational motion in the products. The molecule could also lose one I atom and at some later time lose the other I atom; these I atoms could collide later to form I_2 (stepwise mechanism).

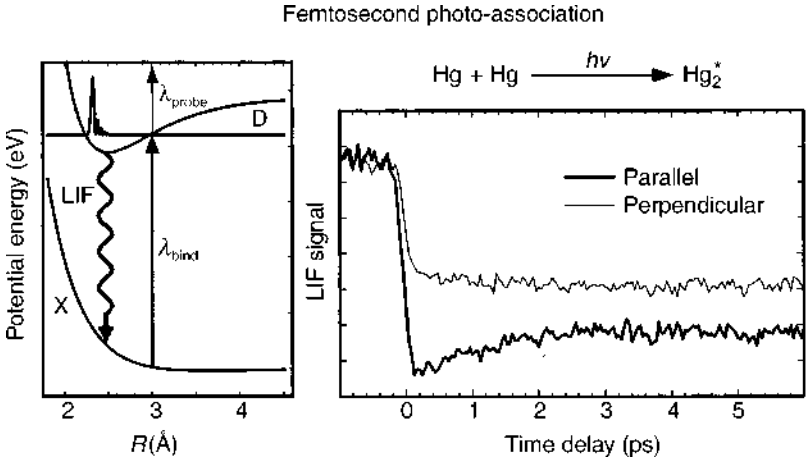

Figure 10 Photo-association of mercury dimers. The pump (bind) pulse is absorbed by two ground-state Hg atoms to form Hg_2^* on the excited state which is fluorescent. The probe pulse depletes this fluorescence causing the LIF signal to decrease at $t = 0$. Analysis of the transient reveals that the time of association is <60 fs. Parallel and perpendicular pump-probe data were examined to obtain the rotational distribution of this reaction.

and the alignment conditions of the reaction can be examined by taking advantage of the polarization of the lasers. The femtosecond photo-association spectroscopy technique was used by the Dantus group to study the formation of an excited mercury dimer. Their experiment is depicted in Figure 10. Two ground-state mercury atoms absorbed the pump (bind) pulse (at 312 nm) forming a wave packet in the excited D state. The probe pulse depleted the fluorescence from the newly formed molecules. The time of association was determined to be within the laser pulse duration (<60 fs). The dimers were found to have a narrow rotational distribution which together with the wavelength of the binding pulse reflects the impact parameter (geometry) of the collision. Calculations reveal that the 'optimal'

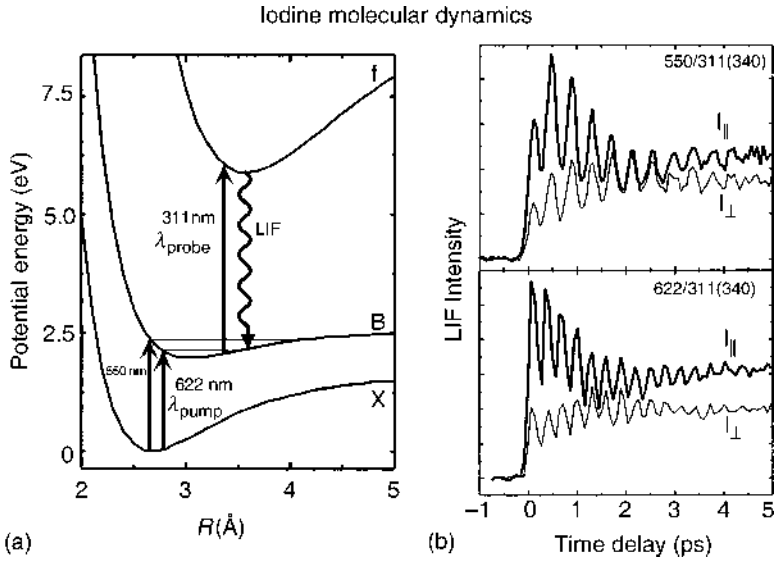

Figure 11 Molecular dynamics of iodine using two different pump laser wavelengths. (a) The pump laser, either at 550 or 622 nm, causes an excitation from the ground state to the B excited state. The choice of wavelength determines the vibrational levels which are excited on the B state. In either case, the probe at 311 nm causes an excitation to the f excited state which fluoresces around 340 nm. (b) Parallel and perpendicular pump-probe transients were obtained using each pump wavelength. At higher pump energies (550 nm), the molecule vibrates more slowly (higher vibrational level) and is reflected by the 410 fs vibrational period observed in the data. At lower pump energies (622 nm), the molecule vibrates faster and corresponds to the 300 fs vibrational period observed in the data.

binding pulse for this association reaction is 10 fs; such short pulses would overlap the entire absorption band from the repulsive ground state to the bound D state.

Bimolecular reactions can also be studied by preparing a low-temperature cluster that contains the reagents of interest. For example, a cluster containing one HI and one CO_2 molecule can be used to study the reaction $H + CO_2 \rightarrow OH + CO$. Initiation of the reaction takes place when the pump laser cleaves the H–I bond and the H atom attacks the CO_2 molecule present in the cluster. The reaction dynamics are probed as a function of time delay between pump and probe pulses as for other pump-probe experiments. The femtosecond dynamics of this reaction have been studied independently by Wittig's and Zewail's research groups.

Studies of Molecular Dynamics

I_2-Pump-Probe, Rotation and Vibration (Bound Surface)

Aside from observing chemical reactions, using ultrafast spectroscopy allows for studying the motion of isolated molecules to elucidate the internal motions of molecules themselves and to extract spectroscopic parameters. Iodine is one of the most frequently used molecules in femtosecond spectroscopy. Here we highlight only a small subset of the femtosecond studies on I_2 that have been conducted by Zewail's and other research groups. Using a pump beam at 550 nm, ground state I_2 is excited to the B state; probing with 310 nm excites the molecule to an even higher bound potential (f state) from which fluorescence at 340 nm can be obtained (as shown in Figure 11). By changing the pump-probe time delay, oscillations on the B excited state are clearly visible and show a vibrational period of approximately 410 fs. Excitation at 622 nm produces a wave packet lower in the potential well, with a vibrational period of approximately 300 fs. The differences observed for parallel and perpendicular probing correspond to the early rotational motion of the molecules. Fourier transfer analysis of the data yields the energy of each vibrational state that forms part of the vibrational wave packet. Different pump wavelengths can be used to access different vibrational levels such that vibrational frequencies and rotational constants can be obtained for a wide range of vibrational levels of the potential. This information can then be used to refine the B excited state potential energy curve of iodine.

Different Measurements Possible with Three-Pulse FWM

Ground and excited state molecular dynamics

Molecular dynamics can also be examined with four-wave mixing techniques. Once again we use iodine as a model system to illustrate that unlike the previous experimental transients of iodine obtained with pump-probe measurements, both excited and

ground state dynamics of I_2 are revealed with FWM. Materny, Kiefer, and co-workers have carried out coherent anti-Stokes Raman spectroscopy and degenerate four-wave mixing measurements on this molecule using two pulses overlapped in time followed by a third pulse after a time delay; both ground and excited state dynamic are observed in the Fourier transform of data. Dantus and co-workers explored all the pulse sequences possible with three-pulse FWM using molecular iodine as a model system (see below).

Double-sided Feynman diagrams and third-order response functions can be used to describe the interaction of the three pulses with the molecular system in three-pulse four-wave mixing. These mathematical tools are described in several books on nonlinear optical spectroscopy (see Further Reading). Figure 12 illustrates how different pulse sequences reveal different molecular dynamics. The transient grating method (Figure 12a) shows both ground (160 fs oscillations) and excited state (310 fs oscillations) wave packet dynamics. When data collection occurs over long time delays, the transient grating signal persists with a similar time constant to that of the lifetime of the excited state. The photon echo pulse sequence (shown in Figure 12b) reveals only excited state dynamics (310 fs oscillations). Note that at longer time delays, the signal decays exponentially with a time constant that reflects the rate of electronic coherence dephasing. This fundamental decay parameter indicates how long molecules remember the phase of the electric field that acted on them.

Off-resonance transient grating molecular dynamics

Off-resonance transient grating can be used to study the molecular motion of molecules in the ground state. The pulse sequence used is shown in Figure 6b. With this method it is possible to examine the rotational and vibrational dynamics of diatomic and polyatomic molecules. The vibrational dynamics belong to the Raman-active modes. From the rotational recurrences, one can obtain rotational constants with great accuracy. Dantus and co-workers have studied carbon dioxide at room temperature and in flames. At high temperatures, rotational recurrences narrow because more rotational levels are populated and the rotational recurrence moves towards longer times because of centrifugal distortion. Further studies on carbon dioxide and carbon disulfide have shown that strong fields can impart an increase in initial alignment (due to torque along the polarization vector of the laser) as well as molecular deformation such as bending (due to electronic state mixing). Heterodyne detected off-resonance transient grating data for CO_2 at room temperature are shown in Figure 13. The data closely resemble the schematic illustration in Figure 5 where rotational dynamics were introduced. The data in Figure 13 have been analyzed to yield the B and D rotational constants which give the structural parameters of the molecules and the centrifugal distortions, respectively. The type of analysis given here for CO_2 can be done on molecules

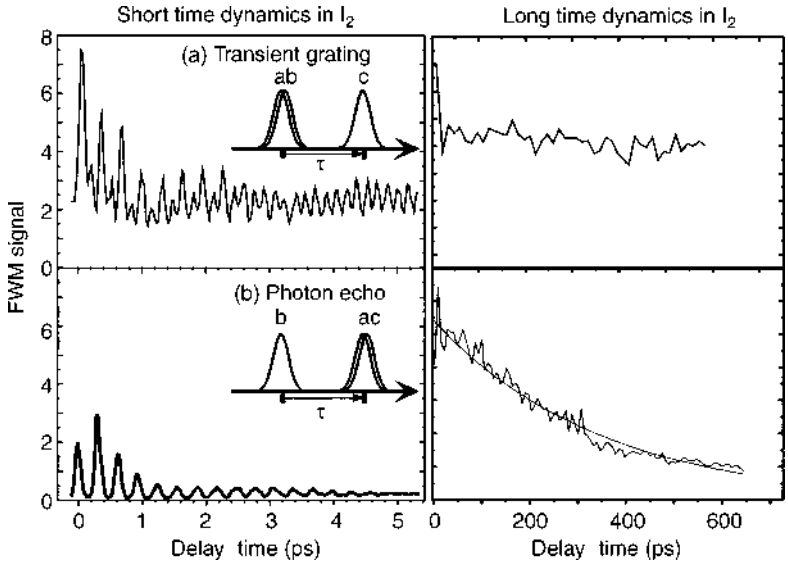

Figure 12 FWM signal in molecular iodine using two pulse sequences at short and long delay times. (a) Transient grating – E_a and E_b overlapped in time followed by E_c. Similar amounts of ground (208 cm^{-1}) and excited (105 cm^{-1}) state dynamics are observed with short delay times. In the transients, these correspond to oscillations of 160 fs and 307 fs respectively. (b) Photon echo – E_a and E_c overlapped in time but preceded by E_b. Excited state dynamics are observed in the transient with short delay times. At long time delays, the decrease in the signal reflects (a) the lifetime of the excited state or (b) the rate of the electronic coherence dephasing.

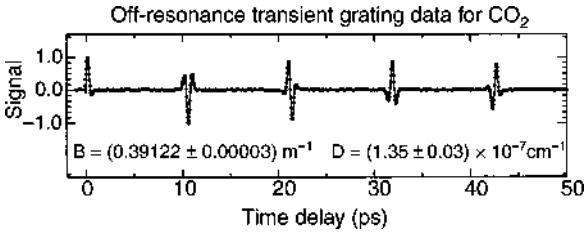

Figure 13 Heterodyne detected off-resonance transient grating on carbon dioxide at room temperature. The experimental data (dots) reflect the rotational motion of the CO_2 molecule. The data have been analyzed (the fit is shown as a line that goes through the experimental data points) to obtain the rotational constant (B) and the centrifugal distortion constant (D).

and molecular fragments resulting from chemical reactions. This method can also be used for molecular identification.

Perspectives for the Future

Most of this article has focused on the observation of ultrafast dynamics involved in chemical reactions. Presently, it is possible to generate pulses with durations as short as 5 fs. Such short pulses are capable of observing motion involving frequencies near 3000 cm^{-1}. This implies that one can observe the dynamics of most chemical bonds. The availability of shorter pulses will allow one to observe electronic dynamics. Advances in laser technology have also produced ultrashort pulses in the infrared region as well as in the vacuum ultraviolet and X-ray region. The wide spectrum available permits experiments that follow specific chemical bonds or specific transitions in elements. This in turn gives a wide range of possibilities for probing ultrafast chemical reaction dynamics.

Having pulses that are shorter than the duration of the dynamics they probe allows one to consider influencing the evolution of a chemical reaction as it is taking place. Controlling chemical reactions with lasers is not a new theme; however, the availability of ultrashort laser pulses has given new power to these efforts. Here we present a brief sample of some of the current efforts.

Control Using Three-Pulse Four-Wave Mixing

Coherent control of chemical reactions depends on the relative phase of two different laser pulses that interact with the sample in order to optimize the transfer of population between electronic states. The three-pulse four-wave mixi\ng technique can be utilized to attain coherent control; the specific timing between the pulses can be used to achieve near-unity or near-zero values of the transfer of population for molecules that

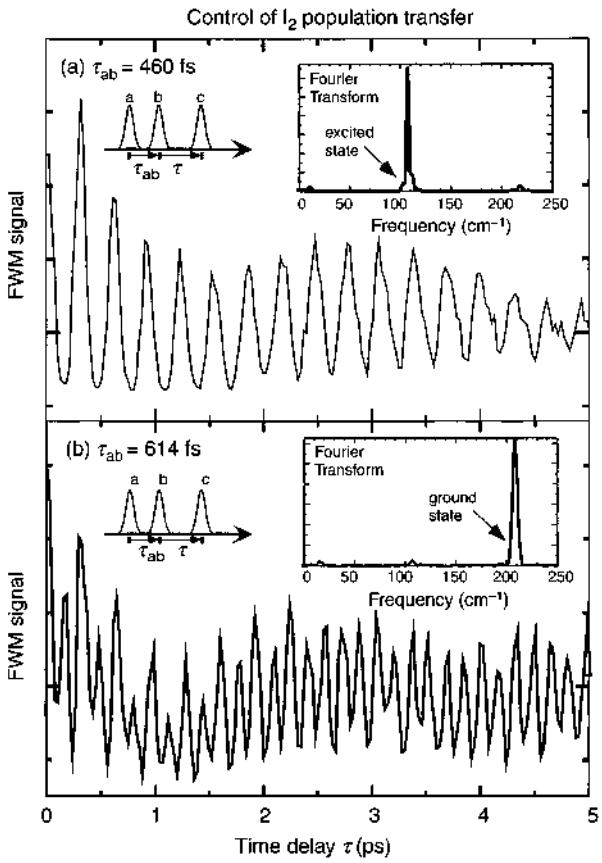

Figure 14 FWM signal in molecular iodine demonstrating control over observation of molecular dynamics. By using the pulse sequence shown, τ_{ab} was defined to be either (a) 460 fs or (b) 614 fs while τ was scanned in time. As is clearly evident in the data and the FFT, excited-state dynamics alone (307 fs oscillations) are observed with $\tau_{ab} = 460$ fs and ground-state dynamics (160 fs oscillations) are observed with $\tau_{ab} = 614$ fs.

interacted with the two electric fields. The Dantus group has also used three-pulse FWM to demonstrate control over optimizing the population transfer between two electronic states in iodine. The populations of the ground and excited states are controlled based on the time delay between the first two pulses τ_{ab} in the pulse sequence shown in Figure 6d. The transient is obtained as a function of the time delay between the second and third pulses, which probes the vibrational dynamics of the population; this experiment is illustrated in Figure 14. For Figure 14a using $\tau_{ab} = 460$ fs, only excited-state dynamics are observed; while for Figure 14b using $\tau_{ab} = 614$ fs, primarily ground-state dynamics are observed. This observation is confirmed by the Fourier transform of the time-resolved transient. The top transient has a vibrational frequency of 108 cm^{-1}, corresponding to the vibrational frequency of the excited B state. The bottom transient has a vibrational

frequency of 208 cm^{-1}, corresponding to the vibrational frequency of the ground state of iodine molecules.

Control Using Feedback and Evolutionary Algorithm

For complex molecules, one seldom knows the precise sequence of pulses or phase modulation of a given pulse that can optimize the yield of a chemical reaction. In these cases, a pulse-shaping device which manipulates the phase and amplitude of frequencies within an ultrashort pulse can be used. The device is part of a learning loop that includes a computer and an experimental setup. The learning algorithm running in the computer generates different pulses to be evaluated experimentally. Signal from the experiments is sent back to the computer for further refinement. The learning algorithm iteratively arrives at the optimum pulse. This method, first introduced by Rabitz, was first tested by Wilson. Gerber and co-workers have demonstrated automated control over the photodissociation reaction of organometallic compounds using an evolutionary algorithm with feedback. The organometallic compounds chosen were iron pentacarbonyl ($Fe(CO)_5$) and dicarbonylchloro(η^5-cyclopentadienyl)iron ($CpFe(CO)_2Cl$). For the first molecule, multiphoton absorption of femtosecond pulses at 800 nm caused ionization and fragmentation resulting in numerous ionic species – Fe^+, $Fe(CO)^+$, $Fe(CO)_2^+$, $Fe(CO)_3^+$, $Fe(CO)_4^+$, and $Fe(CO)_5^+$. These species were detected by mass spectrometry; the measured signal was used as feedback for the evolutionary algorithm that adjusts the phase of the femtosecond laser pulse. The $Fe^+$and $Fe(CO)_5^+$ exit channels were used to demonstrate the control available over branching ratios with this automated control method. Maximization (5.0) of the $Fe(CO)_5^+/Fe^+$ ratio was achieved with very short laser pulses while minimization (0.057) of this ratio was achieved with long (ps) pulses. These results were found to correspond well to changing the pulse duration of bandwidth-limited pulses. The optimization procedure did not attempt to control the other ionic species. This same automated optimization procedure was used to control the branching ratio of $CpFeCOCl^+/FeCl^+$ for the second organometallic compound. The maximum branching ratio was 4.9; the minimum was 1.2; and bandwidth-limited pulses yielded 2.4. Unlike the simpler carbonyl compound, simple changes to the pulse duration of a laser pulse could not optimize this branching ratio. The evolutionary algorithm with the feedback from the mass spectrometer allowed for these tailored pulses to be obtained and optimized the desired exit channel without any knowledge of the specific molecular system. The full potential of shaped femtosecond pulses has yet to be realized. It is possible that shaped femtosecond pulses will play an important role in future time-resolved experiments where they will be used to improve time resolution or to access potential energies that could not be reached with simpler pulses. Pulse shapers will also play a role in the preparation of ultrashort pulses with single femtosecond or attosecond duration.

List of Units and Nomenclature

Å	angstrom (10^{10} Å = 1 meter)
Attosecond	10^{18} as = 1 second
B	rotational constant
cm^{-1}	wavenumbers
D	centrifugal distortion constant
electron volts	1 eV = 8065 cm^{-1}
fs	femtosecond (10^{15} fs = 1 second)
FTS	femtosecond transition state spectroscopy
FWM	four-wave mixing
k	wavevector
LIF	laser-induced fluorescence
nm	nanometer (10^9 nm = 1 meter)
ps	picosecond (10^{12} ps = 1 second)
R(Å)	internuclear distance (given in angstroms)
$t = 0$	initiation of reaction (pump and probe are overlapped in time)
τ	time delay between pump and probe pulses being applied
TG	transient grating
ω_j	rotational frequency

See also

Chemical Applications of Lasers: Non-Linear Spectroscopies; Pump and Probe Studies of Femtosecond Kinetics. **Coherent Control:** Experimental; Theory. **Nonlinear Optics, Applications:** Phase Matching. **Nonlinear Optics, Basics:** Four-Wave Mixing; Ultrafast and Intense-Field Nonlinear Optics; $\chi^{(3)}$ – Third-Harmonic Generation. **Spectroscopy:** Nonlinear Laser Spectroscopy; Raman Spectroscopy. **Ultrafast Laser Techniques:** Generation of Femtosecond Pulses. **Ultrafast Technology**: Femtosecond Condensed Phase Spectroscopy: Structural Dynamics.

Further Reading

Bandrauk AD, Fujimura Y and Gordon RJ (eds) (2002) *Laser Control and Manipulation of Molecules*, vol. 821 Washington, DC: ACS Press.

Chergui M (ed.) (1996) *Femtochemistry: Ultrafast Chemical and Physical Processes in Molecular Systems*. Singapore: World Scientific.

Dantus M (2001) Coherent nonlinear spectroscopy: from femtosecond dynamics to control. *Annual Reviews of Physical Chemistry* 52: 639–679.
De Schriber FC, De Feyter S and Schweitzer G (eds) (2001) *Femtochemistry: With the Nobel Lecture of A. Zewail*. Weinheim, Germany: Wiley-VCH.
Douhal A and Santamaria J (eds) (2002) *Femtochemistry and Femtobiology*. Singapore: World Scientific.
Goswami D (2003) Optical pulse shaping approaches to coherent control. *Physics Reports* 374: 385–481.
Grimberg I, Lozovoy VV, Dantus M and Mukamel S (2002) Femtosecond three pulse spectrocopies in the gas phase: density matrix representation. *Journal of Physical Chemistry* 106: 697–718.
Manz J and Woste L (eds) (1995) *Femtosecond Chemistry*, vols. I and II. Heidelberg, Germany: Verlag Chemie.
Mukamel S (1995) Principles of Nonlinear Optical Spectroscopy. New York: Oxford University Press.
Sundstrom V (ed.) (1997) *Femtochemistry and Femtobiology*. Singapore: World Scientific.
Zewail AH (ed.) (1994) *Femtochemistry – Ultrafast Dynamics of the Chemical Bond*, vols. I and II. Singapore: World Scientific.

Femtosecond Condensed Phase Spectroscopy: Structural Dynamics

E T J Nibbering, Max Born Institut, Berlin, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Ultrafast Chemistry

Chemistry means rearrangement of atoms in/between molecules. Chemical bonds are broken and/or formed, leading to a change in chemical composition, from reactants to products, during a reaction. Following the temporal behavior of the reaction all the way from reactant – via possible intermediates – towards the product state(s), amounts to deciphering the molecular processes that lead to these chemical reaction dynamics. Molecular rearrangements from reactant to product may encompass many orders of magnitude in time (see Figure 1). Typically the larger and more involved the rearrangements are, the more time it takes for a molecular species to reach the final state. Rearrangements of biomolecular structure occur in micro- to milliseconds (spontaneous folding of small proteins consisting of several hundreds of amino acid units), or longer (DNA multiplication, protein synthesis in ribosomes). In these conformational dynamics, hundreds to thousands of chemical bonds change their nature. Bimolecular reaction dynamics in liquid solution occur on time-scales of nanoseconds or longer, being controlled by the relatively slow diffusional motions of the reaction partners to each other. Many elementary processes in chemistry, however, occur on much faster time-scales. In fact, when one considers the dynamical event of a single-bond rearrangement, the relocation of the (relatively few) atoms appears to take place as an ultrafast event. Bond fission, hydrogen and proton transfer, electron transfer, and cis/trans-isomerizations have been found to occur on femto- to picosecond time-scales ('femtochemistry'). The dynamics of these elementary processes are not only determined by the energy landscapes of the reacting partners. The energy landscapes (potential energy surfaces) are determined by molecular parameters such as relative orientation (distances, angles) of the reaction partners, the energy levels of the reactant, intermediate and product species, and – very important – the energy barriers represented by transition states. When considering condensed phase reaction dynamics, the important role of the surrounding solvent has been recognized. The solvent shells modulate energy levels of the reactant, intermediate and product states, through electrostatic interactions. In the liquid phase the solvent fluctuations lead through these interactions to fluctuating energy levels of the reaction partners. In addition, the solvent may facilitate chemical reactions through energy exchange with the reaction partners, leading to efficient dissipation of excess energy, making chemical reactions often irreversible. The processes that are extremely dominated by these solvent interactions (such as molecular collisions, electronic and vibrational dephasing, vibrational relaxation) are also often found to occur on femto- to picosecond time-scales. The outcome of a bond fission may be a null effect, when the dissociating fragments are forced by the surrounding solvent to recombine and relax ('cage effect').

Spectroscopy has been a prime tool to obtain insight into these key molecular processes. Reactant, intermediate, and products contribute to the spectra with their respective molecular resonances. With spectral domain spectroscopy one may, in principle, obtain a precise determination of structural information. Fluctuations of energy levels of these species by the solvent interactions, and finite lifetimes, contribute to a broadening of these molecular resonances. Using steady-state spectroscopy one

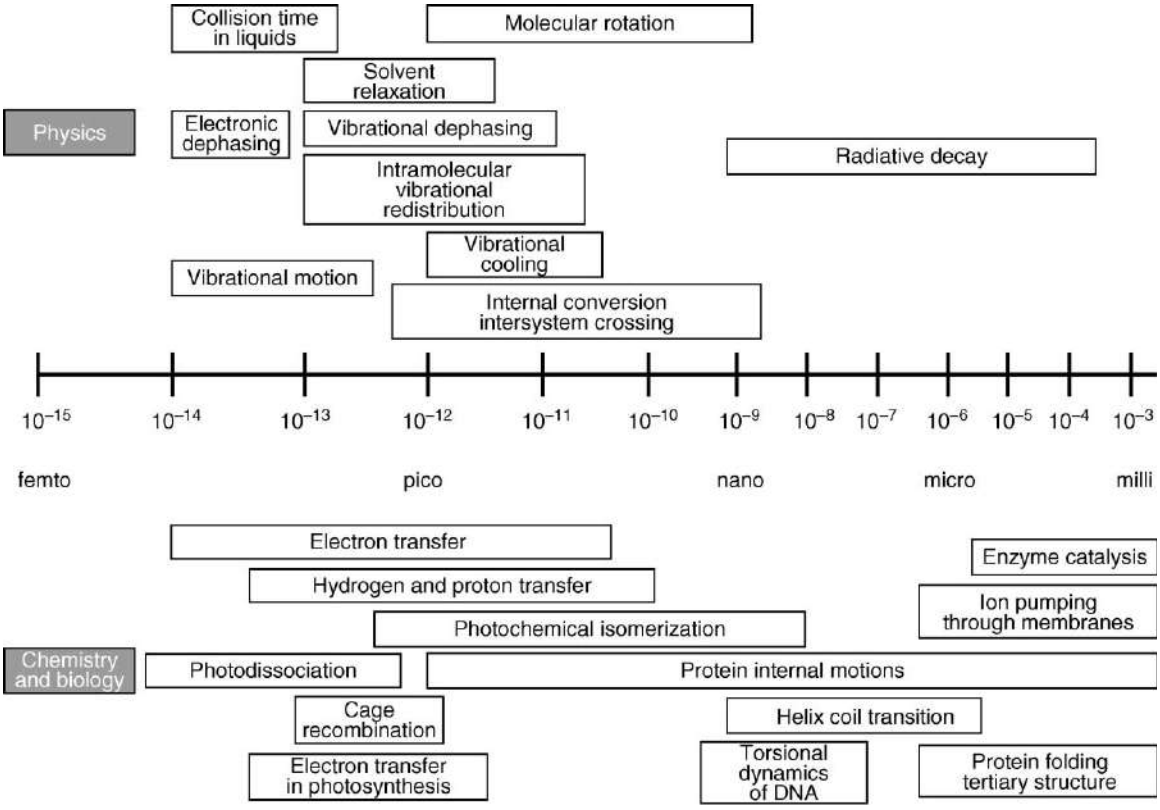

Figure 1 Chart showing the ultrafast time-scales of fundamental physical and chemical processes (inspired by charts of Fleming GR (1986) *Chemical Applications of Ultrafast Spectroscopy*. Oxford, UK: Clarendon Press and Zewail AH (2000) Femtochemistry: atomic-scale dynamics of the chemical bond. Nobel lecture as appeared in *Journal of Physical Chemistry A* 104(24): 5660–5694).

could try to determine these effects through a lineshape analysis. Using this approach, the results have to be analyzed with an a priori defined model representing the dynamics. The disadvantage lies in the averaging over all time-scales inherent in spectral domain spectroscopy, where details of transient states may be lost in this averaging procedure. In contrast, in experiments designed to grasp the molecular dynamics in real time, the temporal resolution is the prerequisite quantity in observation of these processes. In time-resolved experiments one can trigger a chemical reaction at a well-defined point in time, after which one can follow the conformational changes from reactant – via possible intermediates – to the product, and identify any possible transition states along the reaction pathways.

During the last century the technological advances in time-resolved spectroscopy have prompted significant breakthroughs in the study of chemical reaction dynamics. Initially the temporal resolution was given by the duration of flashes from light bulbs ('flash' spectroscopy) or by rapid mixing of reaction partners (stopped-flow technique), as developed by Nobel Laureates Eigen, Norrish, and Porter, and many others. These techniques allow a temporal resolution of at best milliseconds. Already, up to the early 1960s, it was understood that elementary reaction dynamics occur on much faster time-scales. Fourier spectroscopy may improve time resolution down to microseconds or slightly less. With the advent of pulsed laser sources with ever-increasing temporal resolution, a wide arena of time-resolved spectroscopic techniques has emerged. Nanosecond laser systems are known since the early 1960s, picosecond pulses can be generated since the 1970s, the first femtosecond laser system was reported in 1981, and currently laser engineering groups are breaking into the attosecond domain. The majority of pulsed laser systems operate at visible and near-infrared wavelengths. This means that usually chemical events in ultrafast time-resolved spectroscopy are initiated by electronic excitation ('photochemistry'). Subsequent molecular rearrangements have then to be followed by probing the electronic states through electronic resonances (UV/VIS pump–UV/VIS probe or UV/VIS four-wave mixing

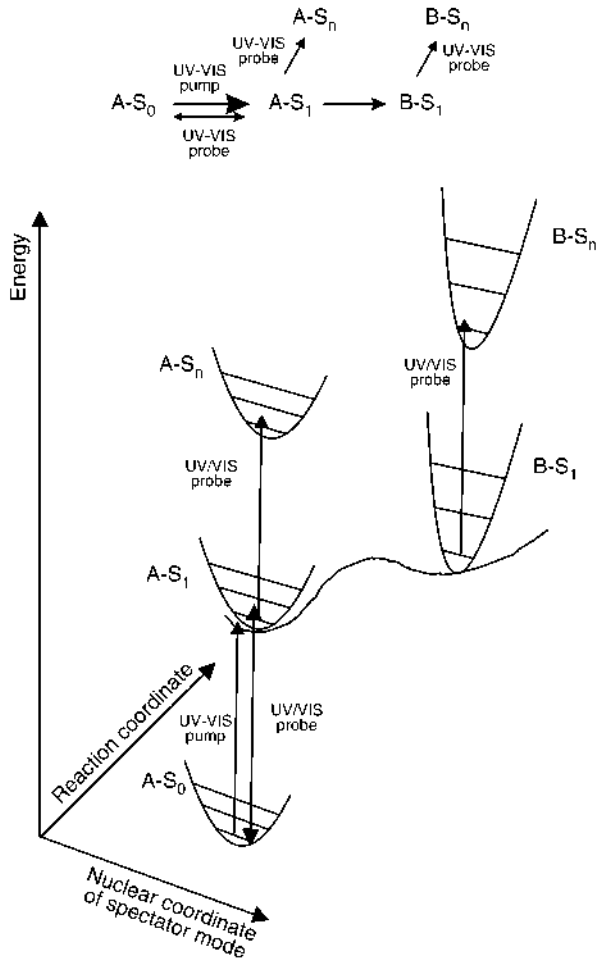

Figure 2 Ultrafast time-resolved electronic spectroscopy. A UV/VIS pump pulse excited an electronic resonance, and a probe pulse determines, depending on the tuning in the UV/VIS population changes in the reactant ground or excited states or in the product state.

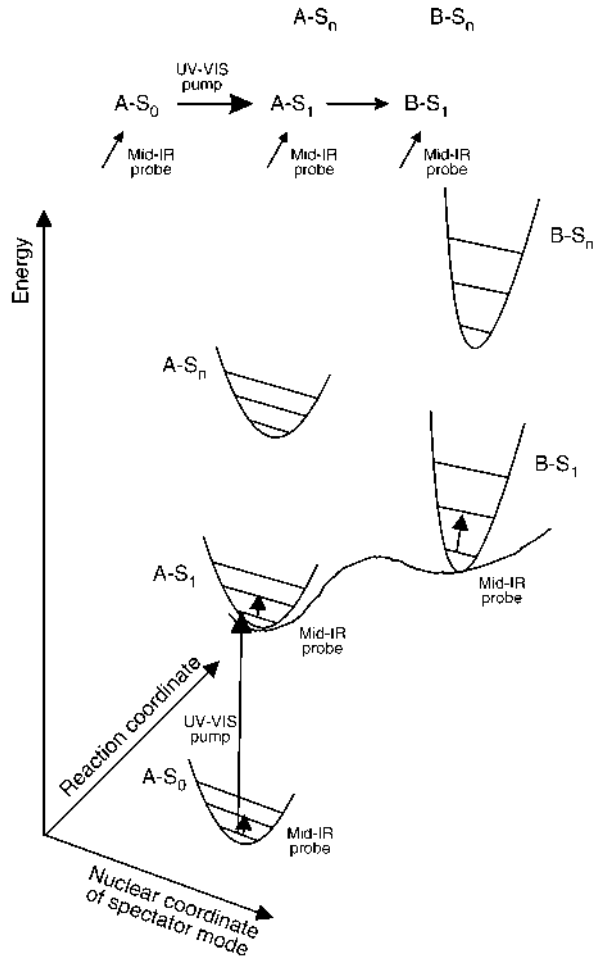

Figure 3 Ultrafast time-resolved infrared spectroscopy. A UV/VIS pump pulse excites an electronic resonance, after which the populations are followed with an infrared probe pulse by inspection of marker modes in the different electronic states.

spectroscopy) (Figure 2), or via vibrational resonances (UV/VIS pump–IR/Raman probe spectroscopy) (see Figures 3 and 4).

Laser Technology

Since the beginning of the 1990s, a major advancement in femtosecond laser technology has emerged by the discovery of Kerr lens mode locking in laser oscillators with Ti:sapphire as lasing material. Ultrastable and ultrashort laser pulses tunable between 700 and 1,000 nm can be generated with durations as short as 6 fs. The second advancement is due to the ability of Ti:sapphire as laser amplifying material, using the method of chirped pulse amplification. Pulses with output powers up to several W, with repetition rates in the kHz-regime, are now routinely generated. These developments facilitate the efficient generation of ultrashort pulses tunable from the UV to the mid-infrared, and even beyond to the THz-regime using schemes based on nonlinear optics. Frequency conversion of the fundamental output of an amplified Ti:sapphire laser system is then achieved with nonlinear processes such as self-phase modulation and supercontinuum generation, n^{th}-order harmonic generation ($n = 2, 3, 4, \ldots$), sum and difference frequency generation, or parametric generation and amplification. As a result, one can now almost arbitrarily tune the excitation and probe wavelengths of the applied laser pulses. Additional parameters, that one may alter, are pulse duration (from less than 10 fs up to several ps) and pulse energy (in the mJ range or less; typically one uses for

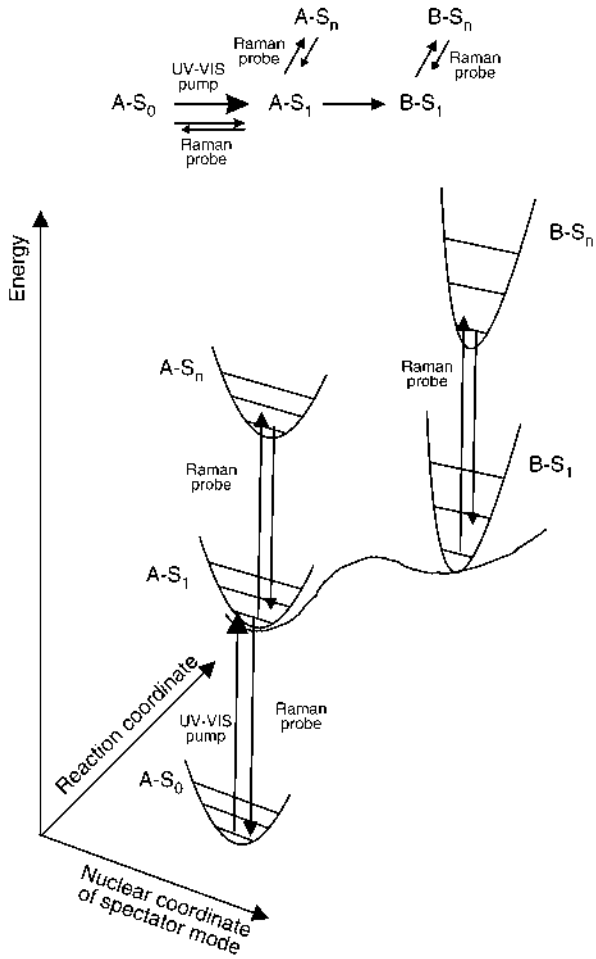

Figure 4 Ultrafast time-resolved Raman spectroscopy. A UV/VIS pump pulse excites an electronic resonance, after which the populations are followed with (resonance) Raman probe pulse by inspection of marker modes in the different electronic states.

condensed phase spectroscopy pulse energies in the range of nJ to μJ). More advanced pulse manipulation schemes use amplitude and phase masks. The latter aspect holds the promise of amplitude and phase control of molecular excitation, feeding the ultrafast chemist's dream of laser control of molecular reaction dynamics.

Ultrafast Electronic Spectroscopic Techniques

Pulsed laser sources enable a variety of experimental approaches to determine transient states. Already, since the development of nanosecond laser sources, time-resolved pump–probe and four-wave mixing spectroscopy have been widely used. In a first approach, often one uses these forms of spectroscopy to probe in real-time populations of reactant,

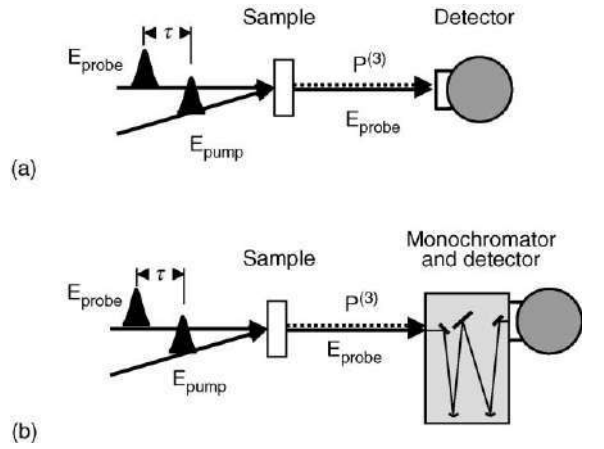

Figure 5 Pump–probe technique: A pump pulse induces an absorbance change in a sample, that is subsequently measured by a time delayed probe pulse, either spectrally integrated (a) or spectrally resolved (b).

transient, and product states as function of pulse delay. With the advent of pico- and femtosecond laser sources, coherence properties of the material response have been investigated as well.

The following techniques have been widely used in ultrafast electronic condensed phase spectroscopy.

UV/VIS Pump–UV/VIS Probe Spectroscopy

In this technique a 'pump' pulse resonant to a transition promotes the molecules to an electronic excited state, and the time-dependent populations are determined by measurement of the transmission of a time-delayed 'probe' pulse (Figure 5). The absorbance change $\Delta A(\tau) = -{}^{10}\log[T(\tau)/T(\tau = 0)]$ (with T for transmission change and τ the pulse delay) is a direct indication of populations of states. Here one can tune the probe pulse to the same transition as the pump pulse, and one measures an absorbance decrease due to a bleach of ground state population (i.e., fraction of molecules being excited), and due to molecules in the excited state stimulated back to the ground state. One could also tune the probe pulse to transitions between the transient state and higher lying states. For electronic spectroscopy, often transitions between the ground and first excited state ($S_0 \rightarrow S_1$) overlap with excited state absorptions ($S_1 \rightarrow S_n$). One has to rely then on numerical analysis procedures such as singular value decomposition and decay associated spectra, that correlate spectral features with temporal behavior.

The time resolution of the experiment is given by the cross-correlation between pump and probe pulses. In the case of extremely short pulses one has

to consider the effects of group velocity dispersion between different frequency components due to different travel times through the samples, leading to pulse temporal broadening. For the case of pump and probe tuned to different wavelengths, the time-resolution is usually dominated by group velocity mismatch between pump and probe pulses.

When ultrashort laser pulses are used, often the applied spectral bandwidth enables simultaneous excitation of several vibrational sublevels in vibronic transitions. Due to the phase-relationship of the applied laser pulses, these vibrational sublevels are prepared in states that are well-defined in temporal phase with respect to each other. Such a coherent superposition of vibrational eigenstates is nothing less than a vibrational wavepacket, that will evolve in time. The occurrence of vibrational wavepacket motions leads to oscillatory modulations of the pump–probe signals.

Time-Resolved Fluorescence Spectroscopy

Here again a 'pump' pulse excites an electronic resonance. The transient excited states are now probed through their spontaneous fluorescence emission, either by use of time-correlated single photon counting or by fluorescence upconversion with a gating pulse in a nonlinear medium (Figure 6). Single-photon counting has a time resolution of at best a few picoseconds. Fluorescence upconversion can be performed with a time resolution of 100–200 fs. The advantage lies in the fact that the measured signals are more likely due to the specific excited states only, and not corrupted by other contributions.

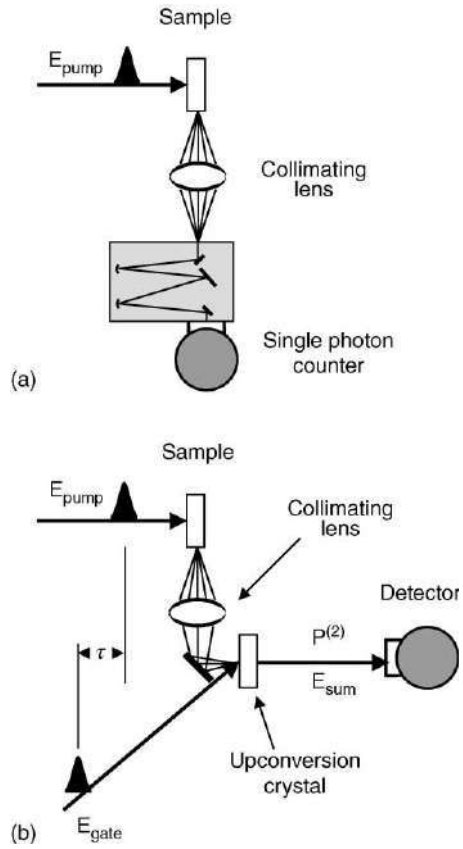

Figure 6 Time-resolved emission detection: The fluorescence emission is either temporally resolved by time-correlated single photon counting (a) or by up-conversion with a gating pulse (b).

Electronic Four-Wave Mixing Spectroscopy

In time-resolved degenerate electronic four-wave mixing spectroscopy, where all the three applied laser pulses (with wavevectors $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$) are tuned to the same electronic resonance, one can detect a nonlinear signal in the phase-matched direction $\mathbf{k}_S = \mathbf{k}_3 + \mathbf{k}_2 - \mathbf{k}_1$ (Figure 7). The first laser pulse generates a coherent superposition between the electronic ground and excited states. This electronic coherence evolves in time during the period (denoted as coherence time τ), until the second pulse converts this coherence in a population frequency grating in the electronic ground and excited states. After a second time period of free evolution (denoted as population or waiting time T), this frequency grating is again transformed into an electronic coherence, which after some time evolution generates a macroscopic polarization in the phase-matched directions. This macroscopic polarization is either directly measured by a time-integrating detector (in which case one measures the intensity signal proportional to the absolute square of the nonlinear polarization), or one uses another light pulse that mixes as local oscillator with the nonlinear polarization (and the signal is proportional to the amplitude of the polarization). The time resolution is determined by the third-order autocorrelation of the applied pulses.

Depending on the delay between the three input pulses, and which delay one varies, different dynamics can be measured. When one performs an experiment with τ scanned and T set at zero, one measures the electronic coherence decay (due to electronic dephasing) time in a two-pulse photon echo experiment. When, on the other hand, τ is set to zero, and T is scanned, one measures the evolution of a population grating, and electronic lifetimes can be deduced. In a three-pulse photon echo experiment both pulse delays are scanned, and the signals give insight into both electronic coherence and frequency grating decay (due to electronic dephasing and spectral diffusion, respectively). A novel extension of four-wave mixing technique is the three-pulse echo peak shift (3PEPS). In a 3PEPS measurement, the delay time τ_{max} is

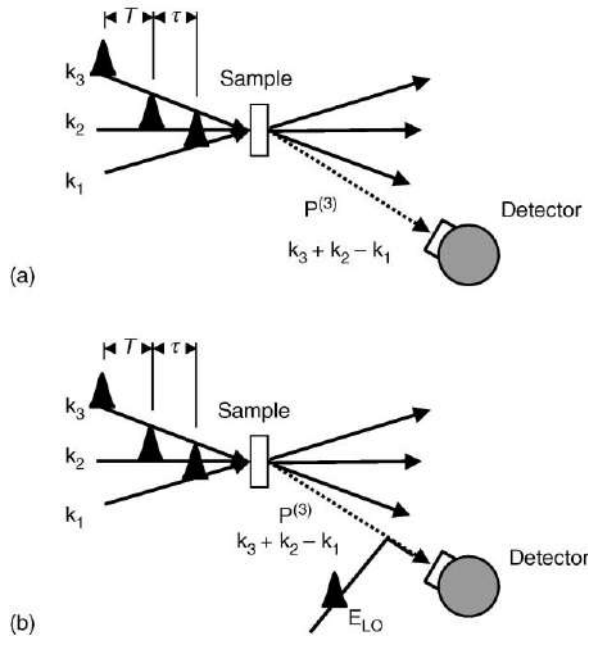

Figure 7 Beam geometry for time-resolved four-wave mixing (grating/photon echo) spectroscopy: The nonlinear signal is transmitted by the sample in a phase-matched direction and its intensity signal (a) or its amplitude by interference with a local oscillator field (b) is detected by a time-integrating detector.

determined, where the echo signal has a maximum along the coherence delay τ, as a function of the population delay T. This method provides direct insight into the frequency fluctuation correlation function that governs the linear and nonlinear signals. Extension of the method to multi-color four-wave mixing spectroscopy has been demonstrated.

Electronic Condensed Phase Spectroscopy: Femtochemistry and Solvation Dynamics

In the narrow sense of meaning, 'femtochemistry' deals with the real-time observation of nuclear motions during chemical reactions. Quantum mechanics teaches us that this is only possible when molecular systems are prepared in vibrational superpositions. Only in the case of quantum states describing vibrational wavepackets evolving in time the narrow-defined positions of nuclei allow for a real-time description of molecular rearrangements, or in other words: dynamics. This has been demonstrated in the gas-phase landmark experiments by Nobel Laureate Zewail (Figure 8).

In the condensed phase at room temperature, however, often such well-defined quantum states are hard to prepare, since the ensembles of molecular systems exist in mixed states. Coherent laser excitation on an ensemble will then prepare mixed states as well, and then information that one obtains from the experiments is less explicit. In addition, room temperature liquid solutions induce extremely fast fluctuations in the states of the molecules under study and as a consequence the electronic and vibrational coherences show ultrafast decaying behavior. This limits the temporal window with which one can follow the evolution of vibrational wavepackets.

Most of the ultrafast spectroscopic experimental work performed on photochemistry in the condensed phase, being described as 'femtochemistry', only deals with the time-scales of changes in electronic states (usually indicated with reaction rates typified with exponentially decaying or rising functions). In this regard it is better to speak of 'kinetics' rather than 'dynamics', although both terms are used in condensed phase spectroscopy. Due to the relative broad electronic spectra of condensed phase molecules masking any structural detail, it is hard to derive nuclear motions in real time, and as such it is difficult to make any statements about dynamics of molecular structures during chemical reactions.

Electronic resonances have also been used to probe the dynamical interactions between a nonreactive solute and the surrounding fluctuating solvent. The idea behind these experiments is to achieve information about the time-scales of solvent fluctuations on (and solvent motions due to a change in) the electronic charge distribution in a molecule (as induced by an electronic excitation). Usually the solvent motions are characterized with a frequency fluctuation correlation function (Figure 9). This information is highly relevant for the case of reacting molecules, where electronic motions accompany the relocations of nuclei, even in electron transfer reactions, where the nucleic rearrangements are modest. Solvent motions have been shown to often control reaction rates. Solvent rearrangement, as a response to a change in electronic charge distribution, is known as solvation dynamics. Experimental probes for solvent motional fluctuations are electronic coherence decay (electronic dephasing) and population frequency grating decay (spectral diffusion) that can be measured in two- and three-pulse photon echoes. Solvent adaptation to new charge distributions can be followed with the time-dependent Stokes-shift in fluorescence emission, as well as in three-pulse photon echoes (in so-called three-pulse echo peak shift measurements described above). Ultimately these experiments resolve the temporal characteristics of the frequency fluctuation correlation function.

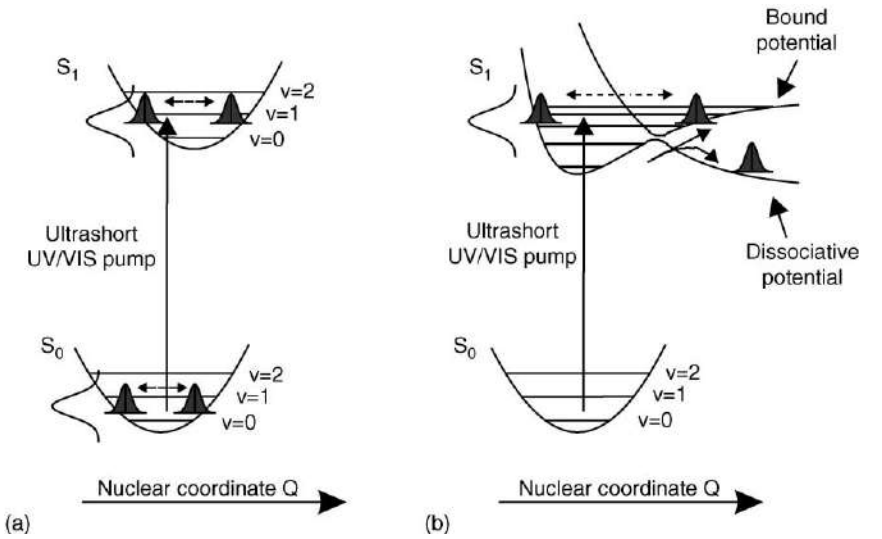

Figure 8 Ultrafast excitation with broadband pulses creates coherent superpositions of vibrational eigenstates, i.e., vibrational wavepackets. These wavepackets are generated in electronic excited and ground states (a), and their time-dependence can be determined in pump–probe and four-wave mixing experiments. When level crossing occurs to another electronic state, the vibrational wavepacket may be observed to evolve along the reaction coordinate (b).

Ultrafast Vibrational Spectroscopy in Photochemistry: Structural Dynamics

Vibrational spectroscopy has, despite a smaller signal sensitivity due to smaller transition cross-sections, several advantages over electronic spectroscopy. Vibrational transitions have typically smaller bandwidths than electronic transitions, due to longer dephasing times (the exception to the rule: O–H/O–D stretching bands of hydrogen bonded hydroxyl groups). Specific signal contributions are thus easier tractable in vibrational spectroscopy than in electronic spectroscopy. Vibrational bands can often be correlated to specific vibrational motions by inspection of the transition frequencies, (e.g., O–H, N–H, and C–H stretching bands can be found in the 3000 cm^{-1} range, C=O and C=N stretching modes are located near 1,600–1,750 cm^{-1}, O–H, N–H, and C–H bending vibrations between 1,500–1,650 cm^{-1}, C–O stretching mode around 1,200–1,300 cm^{-1} etc.), and, in particular, the frequency range between 1,000 and 1,800 cm^{-1} is called the fingerprint region because of this reason. Measurement of vibrational bands thus leads to identification of particular vibrational motions, and conclusions can be drawn on specific structural motifs in the molecules. Vibrational bands can be infrared (IR) or Raman active, sometimes both, and both IR and Raman techniques have been widely applied in structural determination studies ranging from small molecules to larger biomolecular systems. In ultrafast photochemistry one thus excites the molecules with a UV/VIS pump pulse, and an IR/Raman probe pulse follows the outcome of the chemical reaction by inspection of vibrational bands of reactant, transients, and products.

Vibrational spectroscopy has the potential of revealing site-specific information if the marker modes are due to nuclear motions of specific molecular side-groups. For instance, hydrogen bonding induces marked shifts of O–H, N–H, C=O, and C=N bands. Observation of changes in spectral shifts reveals important information on hydrogen bond interactions (weakening/strengthening or even hydrogen bond cleavage). Infrared spectroscopy is able to probe small molecular species in solution that typically have their electronic resonances in the far-UV (making electronic spectroscopy impossible, since normally the solvent would absorb this radiation) (Figure 10). For instance, in acid–base neutralization reactions, where the acidity of a so-called photoacid is switched on by use of a UV pulse, a VIS probe pulse would only be able to probe the photoacid (or its conjugate photobase), thus revealing only when a proton leaves the photoacid. An infrared pulse can, besides probing vibrational resonances of the photoacid, also probe bands of a small base (or its conjugated acid) indicating when a proton arrives at the base.

In the case where the vibrational normal modes do not allow such a structural insight into site-specific groups, one can make a comparison of the

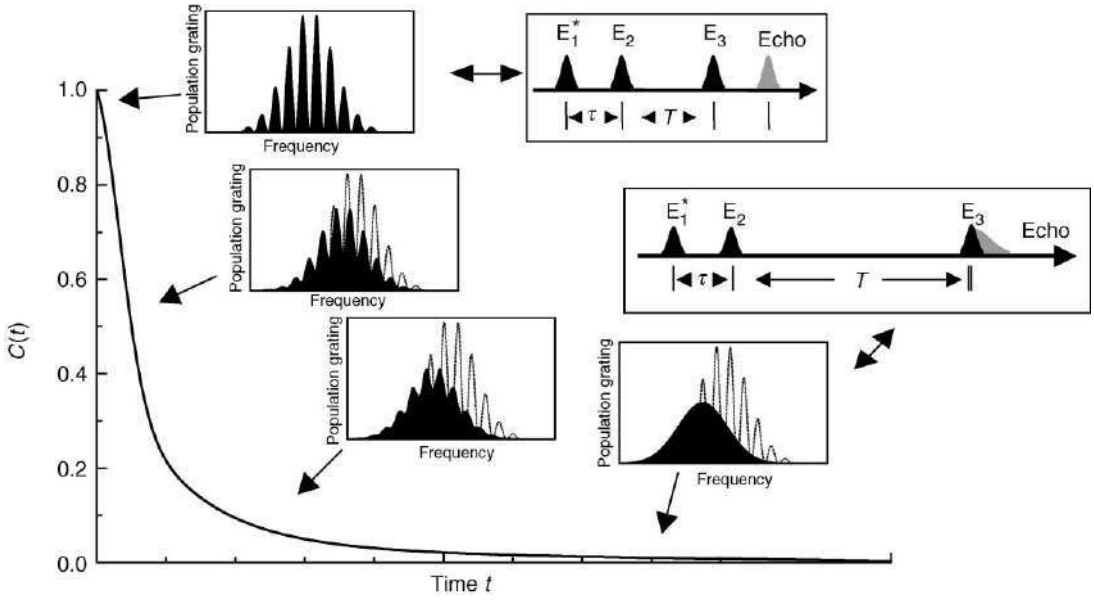

Figure 9 Solute–solvent interactions as characterized by the transition frequency fluctuation correlation function $C(t)$. This quantity can be measured in photon echo and time-resolved fluorescence experiments. In a three-pulse photon echo experiment phase information as given by the excited state frequency grating created with the first two interactions, is washed out due to solvent fluctuations and Stokes shifts to lower frequencies due to solvent rearrangement. A similar plot can be drawn for the ground state frequency grating.

experimentally observed vibrational bands with predictions made by quantum chemical calculations. When a full correspondence between experiment and theory is possible, one can make statements about the three-dimensional structure. With current quantum chemical calculational routines, such as density functional theory, medium-sized molecules are routinely calculated for electronic ground-state conformations. Reliable results of transient and product states in electronic excited states can be obtained with the routine *ab initio* complete active space self-consistent field (CASSCF), albeit for mid-size molecules of at most on the order of about 20 atoms. New developments in numerical procedures, such as time-dependent density functional theory (TD-DFT) may prove fruitful in the calculation of larger molecular systems in electronic excited states.

During recent years, femtosecond IR spectroscopy has been used in photo-induced chemical reactions ranging from excited state hydrogen and proton transfer, transformations of photochromic switches, excited state charge transfer, and cis–trans isomerizations. Here an ultrashort UV/VIS 'pump' pulse promotes the molecule to an electronic excited state, and the reaction is followed by a measurement of the absorbance change of a 'probe' pulse tuned in the mid-IR region where the vibrational marker modes sensitive to structural changes have their resonances. Experimentally, one performs spectrally resolved transient IR spectroscopy. Femtosecond IR parametric devices deliver pulses with bandwidths of 150 cm^{-1} or more. In order to be able to observe shifts as small as the linewidths of IR-active vibrations, one usually measures the IR absorbance change with a detector after spectral dispersion with a monochromator. As a side-effect of this spectral dispersion ground-state bleach signals often appear to grow in at negative pulse delay with the dephasing time of the transition. This effect, known as 'perturbed free induction decay' is a common feature of spectrally resolved nonlinear pump–probe spectroscopy of bleached transitions with dephasing times much longer than the time resolution of the experiment. The time resolution of the experiment is given by the cross-correlation between the UV/VIS-pump and IR-probe pulse (about 100–200 fs), and is typically dominated by group velocity mismatch in samples with thicknesses of about 100 μm.

In principle, the same approach can be followed by probing Raman-active vibrations. In this case the spectral resolution is not only determined by the monochromator through which the spontaneous Raman emission dispersed, but also by the bandwidth of the gating pulse by which the Raman effect is induced. As a result, UV/VIS pump–Raman probe

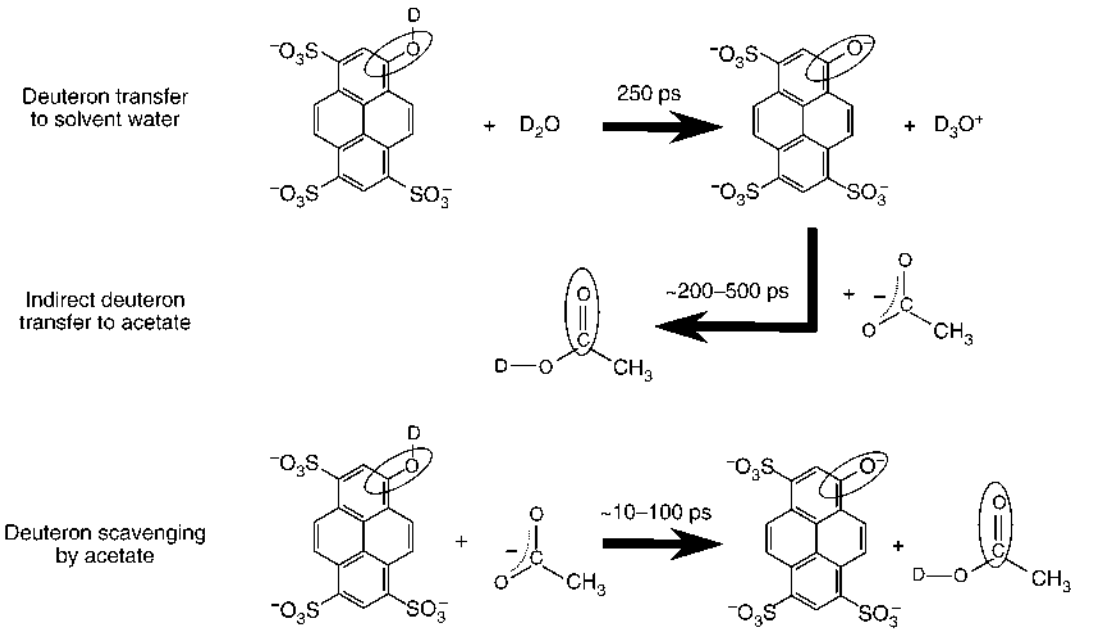

Figure 10 In excited state acid–base neutralization experiments, the proton dissociation can be followed on two sides with ultrafast infrared spectroscopy. Vibrational marker modes of photoacid and conjugated photobase indicate when the proton leaves the acid. The marker mode of acid formed by proton pick-up by the accepting base is a direct measure of the arrival time of the proton at the base. Depending on relative concentrations, either the indirect proton transfer via the solvent or direct scavenging of the proton by the base dominates the dynamics.

spectroscopy has a temporal resolution of around 1 ps (a compromise between spectral and time resolution). Due to the even weaker cross-sections of Raman transitions, one often uses resonance enhancement by tuning the gating pulse close to or resonant with an electronic transition of the state that is probed. The advantage is that one usually only observes Raman bands of the state under inspection. The drawback is that fluorescence resulting from the resonant electronic excitation often inhibits detection of Raman bands in extended spectral ranges.

Raman spectroscopy offers an additional insight into how chemical reactions evolve. By comparison of the intensities of anti-Stokes and Stokes lines of a particular vibration, it is possible to derive time-dependent changes in the excitation level of this vibration. One can draw conclusions on whether particular modes initially drive a chemical reaction ('state transition promoting modes'), or only get excited after the transition is made, by taking up the excess energy released by the reaction ('accepting modes'). In such a way, insight is obtained on the energy flow inside a molecular system (intramolecular vibrational redistribution, abbreviated as IVR) and vibrational energy dissipation to the surrounding solvent (vibrational cooling). Infrared spectroscopy is less powerful in revealing this, since red-shifted transient absorption of vibrational bands may either mean that the particular mode is highly excited (the red-shift is a consequence of the diagonal anharmonicity of the vibration), or it may mean that other modes are highly excited and these cause a red-shift of the marker mode under inspection (the red-shift is then due to off-diagonal anharmonic coupling with other modes).

Equilibrium Structural Dynamics in the Electronic Ground State

Existence of anharmonic coupling between vibrational modes means that the vibrational motions are not decoupled from each other. These couplings always exist, since otherwise no IVR and vibrational cooling (no relaxation) would occur. When the couplings are significantly large, excitation of one specific mode will induce a significant instantaneous shift of other vibrations anharmonically coupled to this particular mode. Estimation of the magnitude of these couplings should lead to a determination of the curvatures of potential energy surfaces along the respective coordinates. With current state-of-the-art femtosecond infrared technology (with time resolution of about 150 fs or less) it is possible to excite

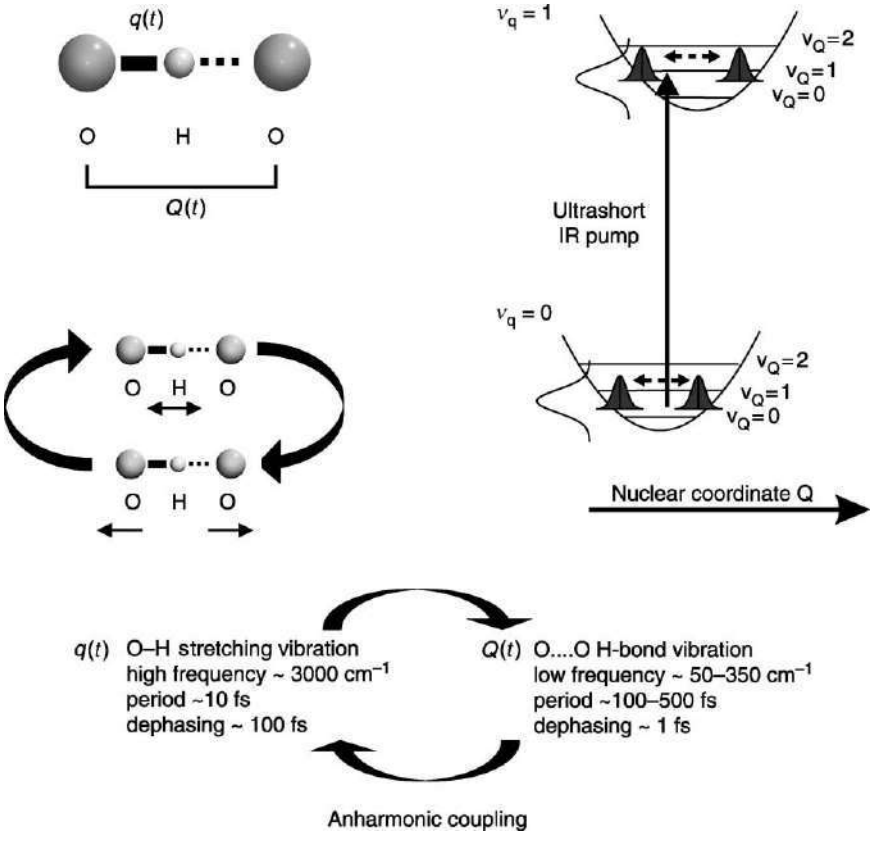

Figure 11 In hydrogen-bonded hydroxyl groups a strong anharmonic coupling exists between the hydroxyl stretching and modes modulating the hydrogen-bond distance. A Born–Oppenheimer-like separation of time-scales of vibrational motion between the high-frequency O–H/O–D stretching and the low-frequency hydrogen bond modes enables a description with potential energy surfaces for the low-frequency mode as a function of the quantum state of the hydroxyl stretching vibration. The formation of low-frequency vibrational wavepackets in ultrafast IR spectroscopy of O–H/O–D stretching modes is then fully analogous to wavepacket formation in ultrafast electronic spectroscopy.

infrared-active vibrations and to probe the same or other vibrations (either in the IR or by the Raman process), enabling insight into the anharmonic couplings between these vibrational modes.

Direct anharmonic couplings are probed in IR-pump–IR-probe and IR photon echo spectroscopy (Figure 11). In the case of peptides, the excitonic (Davydov) couplings between amide I vibrations of the amino acid units have been the subject of extensive study since these lead to spatial information of the relative orientations of the different amide I modes, and thus, of the spatial orientation of the amino acid units inside the peptide. Exploring the effects of the delay between the IR pulses on the observed signals, give insight into the fluctuations of these orientational features, or in other words, into structural dynamics of peptides. In the case of hydrogen bonded O–H/O–D stretching vibrations, the anharmonic couplings with under-damped low-frequency modes modulating the hydrogen bond distance, lead to marked modulations of observed pump–probe and echo signals as function of pulse delay. These experiments reveal which modes couple strongly to the hydrogen stretching oscillator.

Fluctuations in couplings between vibrational modes (in particular with those of the solvent) induce dephasing of the vibrational coherences. The dephasing and spectral diffusion dynamics of vibrational transitions, e.g., the O–H stretch vibration of HOD dissolved in D_2O, have been determined by IR photon echo spectroscopy. From these studies the time-scales of structural memory decay of hydrogen bond networks can be derived. Couplings between vibrational modes ultimately lead to population relaxation through intramolecular vibrational redistribution effects and vibrational cooling. Ultrafast two-color IR-pump–IR-probe reveals whether excitation of one IR-active vibration is followed by population transfer into another IR-active mode.

With IR-pump–Raman-probe, one can estimate whether Raman-active vibrations are transiently excited. In the case of hydrogen-bonded O–H/O–D stretching bands, an extremely short population relaxation time T_1 (below 1 ps) is observed, as well as an ultrarapid dephasing time T_2 (200 fs or less). These values are definitely longer for other vibrations, from which it has been concluded that the strong anharmonicities in hydrogen-bonded vibrational systems enable dephasing and relaxation processes to have such a rapid impact.

Future Developments

Future activities in the determination of dynamically evolving molecular structures in the condensed phase, may involve the aforementioned application of amplitude and phase controlled excitation pulses in optimal control of chemical reactions. Vibrational spectroscopy may be used as a tool in determination of evolving structures after electronic excitation by a shaped pulse tuned in the UV/VIS. Infrared pulses could also be shaped in amplitude and phase, and may be used in the exploration of steering of chemical reactions in the electronic ground state. Multidimensional vibrational spectroscopy is an object of extensive research. Extension of the method to determination of structure of transient states will lead to new results on molecular rearrangements, e.g., polypeptides reaching a new equilibrium after inducing a UV-excitation induced geometric change of a photochromic switch incorporated in the peptide structure. Much activity is currently also being put in the development of ultrafast structure resolving techniques such as electron diffraction, X-ray diffraction, and X-ray spectroscopy. While still extremely demanding at a technological level, final implementation of these latter techniques will open up new areas of ultrafast chemical dynamics.

See also

Chemical Applications of Lasers: Pump and Probe Studies of Femtosecond Kinetics.

Further Reading

Cerullo G and De Silvestri S (2003) Ultrafast optical parametric amplifiers. *Review of Scientific Instruments* 74(1): 1–18.

de Boeij WP, Pshenichnikov MS and Wiersma DA (1998) Ultrafast solvation dynamics explored by femtosecond photon echo spectroscopies. *Annual Review of Physical Chemistry* 49: 99–123.

Elsaesser T and Bakker HJ (eds) (2002) *Ultrafast Hydrogen Bonding Dynamics and Proton Transfer in the Condensed Phase*. Dordrecht, Germany: Kluwer.

Elsaesser T and Kaiser W (1991) Vibrational and vibronic relaxation of large polyatomic molecules. *Annual Review of Physical Chemistry* 42: 83–107.

Fayer MD (ed.) (2001) *Ultrafast Infrared and Raman Spectroscopy*. New York: Marcel Dekker.

Fleming GR (1986) *Chemical Applications of Ultrafast Spectroscopy*. Oxford, UK: Clarendon Press.

Fleming GR and Cho M (1996) Chromophore-solvent dynamics. *Annual Review of Physical Chemistry* 47: 109–134.

Fleming GR and Hänggi P (eds) (1993) *Activated Barrier Crossing: Applications in Physics, Chemistry and Biology*. Singapore: World Scientific.

Kaiser W (ed.) (1988) *Ultrashort Laser Pulses and Applications*. Berlin: Springer.

Mukamel S (1995) *Principles of Nonlinear Optical Spectroscopy*. Oxford, UK: Oxford University Press.

Nibbering ETJ and Elsaesser T (2004) Ultrafast vibrational dynamics of hydrogen bonds in the condensed phase. *Chemical Reviews* 104: 1887–1914.

Owrutsky JC, Raftery D and Hochstrasser RM (1994) Vibrational relaxation dynamics in solutions. *Annual Review of Physical Chemistry* 45: 519–555.

Sundström V (ed.) (1996) *Nobel Symposium: Femtochemistry and Femtobiology: Ultrafast Reaction Dynamics at Atomic-Scale Resolution*. Singapore: World Scientific.

Wiersma DA (ed.) (1994) *Femtosecond Reaction Dynamics*. Amsterdam: North-Holland.

Zewail AH (2000) Femtochemistry: atomic-scale dynamics of the chemical bond. Nobel lecture as appeared in *Journal of Physical Chemistry A* 104(24): 5660–5694.

Web-information: http://www.nobel.se
Nobel web-site showing information on Nobel Laureates M Eigen, RGW Norrish and G Porter (The Nobel Prize in Chemistry 1967: 'for their studies of extremely fast chemical reactions, effected by disturbing the equilibrium by means of very short pulses of energy') and AH Zewail (The Nobel Prize in Chemistry 1999: 'for his studies of the transition states of chemical reactions using femtosecond spectroscopy').

Ultrafast Illumination and Processing

Y Fainman, University of California, San Diego, CA, USA

D M Marom, Bell Laboratories, Holmdel, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Mode-locked lasers emitting ultrashort pulses have expanded the range of available instruments and techniques to the scientific community. The unique properties of these ultrashort pulses – their high peak powers and short duration – enable such diverse applications as multiphoton microscopy, pump-probe experiments, and coherent control of molecular and atomic wavepackets, to name a few. Ultrashort pulses can also be used as a basis for ultrahigh data rate optical communications using time division multiplexing (TDM) in a fiber communication application. These ultrafast applications often require complementary real-time optical processors based on instantaneous nonlinear phenomena, since conventional electronic means are limited by the response time of electronic devices.

During the last few years, we have developed real-time optical processors utilizing nonlinear three- and four-wave mixing of signals originating from spatial or temporal channels for synthesizing, processing, and detecting ultrafast waveforms. We perform real-time optical signal processing that can be applied to data conversion between slow parallel channels in space and an ultrafast serial signal. A time-to-space mapping processor has been built for detection of ultrafast waveforms. It mixes two spatially inverted temporal frequency information waves in a three-wave mixing arrangement, generating a quasi-monochromatic spatial signal carrying the temporal image for detection by slower electronic means. A space-to-time mapping processor was constructed for generation of arbitrary ultrafast waveforms from an input ultrashort pulse. The waveform is a time-scaled replica of the employed input spatial image, converted by a four-wave mixing arrangement. By introducing only temporal information channels to our four-wave mixing processors, real-time optical processing of temporal waveforms was enabled. We have demonstrated time reversal experiments based on performing spectral phase conjugation and spectral inversion operations, achieving time reversal of the electrical field and of the complex amplitude waveform, respectively. The three- and four-wave mixing arrangements are performed within a spectral processing device – a free-space optical setup consisting of diffraction gratings and lenses – utilizing a $\chi^{(2)}$ nonlinear crystal. A cascaded second-order nonlinearity technique is used for four-wave mixing within the $\chi^{(2)}$ crystal (a frequency-sum generation process followed by a frequency-difference generation process satisfying the type-II noncollinear phase matching condition). However, a $\chi^{(3)}$ nonlinear crystal may be readily substituted without loss of generality (albeit with a likely smaller conversion efficiency).

This article reviews the spatial/temporal processing techniques for synthesizing, processing, and detecting ultrafast waveforms. The generation of spectrally decomposed waves within a spectral processing device is described in the following section. The subsequent section articulates on the employed nonlinear wave mixing processes, and the ensuing sections express the synthesis, processing, and detection of ultrafast waveforms.

Generation and Characteristics of Spectrally Decomposed Waves

Our wave mixing experiments are performed with spectrally decomposed waves, in which the spectral frequency content is spatially dispersed. When the pulse is decomposed with high resolution, the signal at every spatial location can be treated as quasi-monochromatic. This allows for increased interaction length within a nonlinear crystal, allowing us to use long crystals, whereas typically the interaction length is limited by temporal walkoff. In this section we analyze the creation of spectrally decomposed waves.

We introduce an ultrashort optical waveform, with a temporal envelope of $s(t/\tau)$, into the spectral processing device, where τ parameterized the waveform's duration. The waveform is propagating in free space towards a diffraction grating, at an incidence angle θ (see Figure 1). The input ultrashort waveform is characterized in its coordinate system (x_1, z_1) as:

$$E_{\text{pulse}}(x_1, z_1; t) = w\left(\frac{x_1}{L_1}\right) s\left(\frac{t - t_0 - z_1/c}{\tau}\right) \times \exp\left(\mathrm{i}\left(\omega_0 t - \frac{\omega_0}{c} z_1\right)\right) \quad [1]$$

where $w(\cdot)$ defines the transversal field distribution with a scale size parameter L_1, t_0 is a time delay parameter, c is the speed of light in vacuum, and ω_0 is the center optical frequency. The waveform is

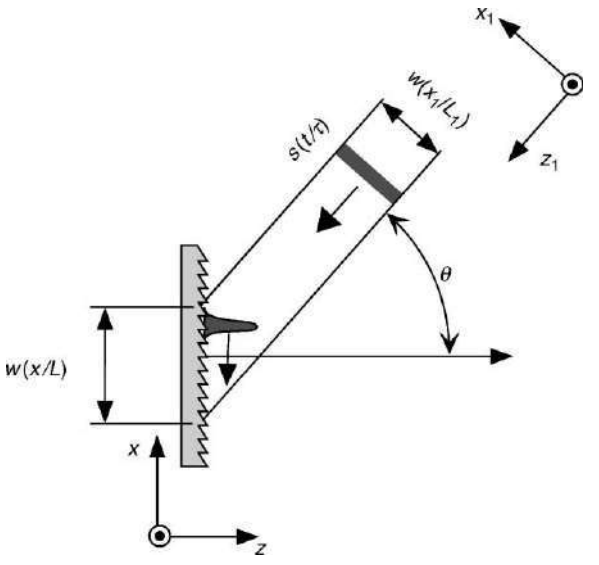

Figure 1 An ultrashort optical waveform, $s(t/\tau)$, propagating in free-space is incident upon a diffraction grating at angle θ. The transverse spatial mode of the pulse is stretched due to the projection onto the grating.

propagating in the z_1 direction at a group velocity and phase velocity of the speed of light in vacuum. The spatial confinement in the y_1 direction is omitted for brevity. We perform a time-domain Fourier transform (TFT) on eqn [1] to perform a frequency decomposition of the ultrafast waveform, yielding:

$$\tilde{E}_{\text{pulse}}(x_1, z_1; \omega) = \tau w\left(\frac{x_1}{L_1}\right) \times \exp\left(-\mathrm{i}\frac{\omega}{c}z_1\right)\tilde{s}(\tau(\omega-\omega_0)) \times \exp(-\mathrm{i}(\omega-\omega_0)t_0) \quad [2]$$

where the tilde over the script denotes a FT relationship, defined for the TFT as $\tilde{s}(\omega)A\int s(t)\exp(-\mathrm{i}\omega t)\mathrm{d}t$. Next we perform a rotation of the coordinate system from (x_1, z_1) to (x, z), for compatibility with the coordinate system of the spectral processing device, yielding:

$$\tilde{E}_{\text{pulse}}(x, z; \omega) = \tau w\left(\frac{x\cos\theta - z\sin\theta}{L_1}\right) \times \exp\left(-\mathrm{i}\frac{\omega}{c}(x\sin\theta + z\cos\theta)\right) \times \tilde{s}(\tau(\omega-\omega_0))\exp(-\mathrm{i}(\omega-\omega_0)t_0) \quad [3]$$

To find the incident field on the input grating of the spectral processing device, the field of eqn [3] is evaluated at $z = 0$. The effect of the grating diffraction can be modeled by adding the grating momentum, k_g, in the x direction to the k-vector that characterizes the propagation direction of the wave, resulting in:

$$\tilde{E}_{\text{input}}(x, 0; \omega) = \tau w\left(\frac{x\cos\theta}{L_1}\right) \times \exp\left(-\mathrm{i}\left(\frac{\omega}{c}\sin\theta - k_g\right)x\right) \times \tilde{s}(\tau(\omega-\omega_0))\exp(-\mathrm{i}(\omega-\omega_0)t_0) \quad [4]$$

The grating momentum, k_g, and the incidence angle, θ, are chosen such that the center frequency ω_0 will diffract in the direction of the optical axis of the system. Setting $\omega_0/c\cdot\sin\theta = k_g$ and substituting the grating's k-vector, $k_g = 2\pi/\Lambda$, where Λ is the grating period, yields $\sin\theta = \lambda_0/\Lambda \equiv \alpha$, where λ_0 is the optical wavelength corresponding to the center carrier frequency. To further simplify the notation, we define a new spatial width scaling parameter $L = L_1/\cos(\theta)$, accounting for the aperture size increase due to the projection onto the diffraction grating. With these substitutions, the expression takes on the more familiar form, given by

$$\tilde{E}_{\text{input}}(x; \omega) = \tau w\left(\frac{x}{L}\right)\exp\left(-\mathrm{i}\frac{(\omega-\omega_0)\alpha x}{c}\right) \times \tilde{s}(\tau(\omega-\omega_0))\exp(-\mathrm{i}(\omega-\omega_0)t_0) \quad [5]$$

We may perform an inverse TFT on eqn [5] to characterize the input signal in the time domain, yielding:

$$E_{\text{input}}(x; t) = w\left(\frac{x}{L}\right)s\left(\frac{t - t_0 - \alpha x/c}{\tau}\right)\exp(\mathrm{i}\omega_0 t) \quad [6]$$

Equation [6] describes the input waveform scanning across the fixed aperture at velocity of c/α in the x-direction.

The spectrally decomposed wave is generated by performing a spatial Fourier transform (SFT) employing a lens of focal length f on the input wave (either frequency domain representation, eqn [4], or time domain representation, eqn [6]). Using the frequency domain representation, the spectrally decomposed wave U_{SDW} of the input waveform is given by

$$U_{\text{SDW}}(x'; \omega) = \tau\tilde{s}(\tau(\omega-\omega_0)) \times \exp(-\mathrm{i}(\omega-\omega_0)t_0)\int_{-\infty}^{\infty} w\left(\frac{x}{L}\right) \times \exp\left(-\mathrm{i}(\omega-\omega_0)\frac{\alpha x}{c}\right) \times \exp\left(-\mathrm{i}2\pi\frac{xx'}{\lambda_0 f}\right)\mathrm{d}x \quad [7]$$

where we use the central wavelength in the Fourier transform kernel. Evaluating the integration of eqn [7], yields:

$$U_{\mathrm{SDW}}(x';\omega) = L\tau\tilde{s}(\tau(\omega-\omega_0))\exp(-\mathrm{i}(\omega-\omega_0)t_0) \times \tilde{w}\left(\frac{\alpha L}{2\pi c}\left[\frac{\omega_0 x'}{\alpha f} + (\omega-\omega_0)\right]\right) \quad [8]$$

where the SFT of the input pupil function, $\tilde{w}(f_x)A\int w(x)\exp(-\mathrm{i}2\pi x f_x)\mathrm{d}x$, has mixed space-domain and time-domain frequency variables in its argument, characterizing the dispersion. A spectral component shifted away from the center frequency, $\Delta\omega = \omega - \omega_0$, will be centered at a displaced location $\Delta x' = -\Delta\omega\cdot\alpha f/\omega_0$, defining the spatial dispersion term $\partial x'/\partial\omega = -\alpha f/\omega_0$. One can see that in order to get fine localization of each spectral component, the spatial extent of the function $\tilde{w}(\cdot)$ must be minimized. This can be readily achieved by increasing the size of the input aperture, or the parameter L, as well as increasing α (grating's angular dispersion). Let us make the following observation on system resolution: The ultrafast pulse's spectral width is proportional to $1/\tau$, and its spatial extent is therefore proportional to $\alpha f/\omega_0\tau$ (multiplying by the spatial dispersion term). The spatial width of each spectral component is proportional to $cf/\omega_0\tau$. The dimensionless ratio of these two spatial widths, which is indicative of the number of resolvable spectral components, is $N \sim \alpha L/c\tau$, where N is defined as the system spectral resolution. When the spectral processing device is designed to satisfy the requirement of $N \gg 1$, then the spectrally decomposed wave can be considered quasi-monochromatic at every spatial position. In our experiments, we use $\lambda_0 = 800$ nm, gratings of 600 lines/mm frequency ($\Lambda = 1666$ nm), $L \approx 1$ cm, and $\tau \approx 100$ femtoseconds, yielding $\alpha \cong 0.5$ and $N \sim 160$, easily satisfying the high spectral resolution requirement. Let us make a second observation on system resolution using the time domain representation of eqn [6]: The spatial extent of the input aperture is proportional to L, whereas the instantaneous spatial extent of the scanning waveform is proportional to $\tau c/\alpha$ (see Figure 2). The dimensionless ratio of these two spatial widths is again $N \sim \alpha L/c\tau$, exemplifying the concept of resolvable positions and demonstrating the conservation of the space-bandwidth product.

To further facilitate the ensuing analysis, let us assume that we operate in the high-resolution regime (satisfied with the values given above), and we seek to simplify the expressions defining the spectrally decomposed wave. The spectrally decomposed wave

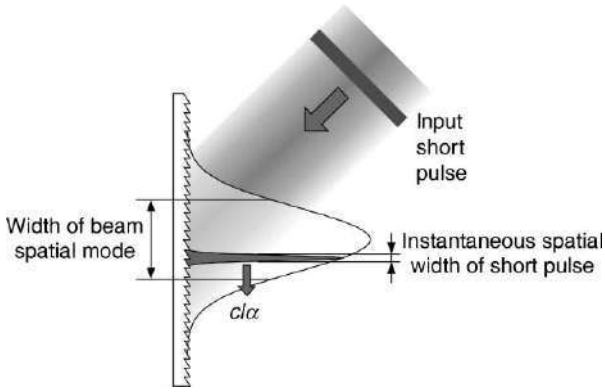

Figure 2 The input short pulse scans across the diffraction grating at velocity of c/α. The ratio of the width of the spatial mode of the beam to the instantaneous spatial width of the scanning waveform determines the system resolution. © 2001 Optical Society of America.

can be defined in the time domain by performing a SFT on eqn [6], yielding:

$$\begin{aligned} U_{\mathrm{SDW}}(x';t) &= \int w\left(\frac{x}{L}\right)s\left(\frac{t-t_0-\alpha x/c}{\tau}\right)\exp(\mathrm{i}\omega_0 t) \times \exp\left(-\mathrm{i}\frac{2\pi}{\lambda_0 f}xx'\right)\mathrm{d}x \\ &\approx w\left(\frac{c(t-t_0)}{\alpha L}\right)\exp\left(-\mathrm{i}\frac{\omega_0(t-t_0)x'}{\alpha f}\right) \times \tilde{s}\left(-\frac{\omega_0\tau x'}{\alpha f}\right)\exp(\mathrm{i}\omega_0 t) \end{aligned} \quad [9]$$

where by account of the high-resolution approximation, the instantaneous location of the traveling signal waveform samples the value of the aperture function and removes the aperture dependence from the Fourier integral. This approximation decouples the aperture function from the temporal waveform, and illustrates that the duration of the spectrally decomposed wave is defined by the aperture function (time of flight through aperture of size L), and that the spatial dependence is a mapping of the incident waveform spectrum. A rotating wavevector in time is also present, on account of the traveling (moving) waveform on the input plane. Again, we may perform a TFT on eqn [9], defining the high-resolution limit of the spectrally decomposed wave, yielding:

$$U_{\mathrm{SDW}}(x';\omega) = \tilde{s}\left(-\frac{\omega_0\tau x'}{\alpha f}\right)\exp(-\mathrm{i}(\omega-\omega_0)t_0) \times \tilde{w}\left(\frac{\alpha L}{2\pi c}\left[\frac{\omega_0 x'}{\alpha f} + (\omega-\omega_0)\right]\right) \quad [10]$$

One can compare eqns [8] and [10], as both characterize the spectrally decomposed wave in the

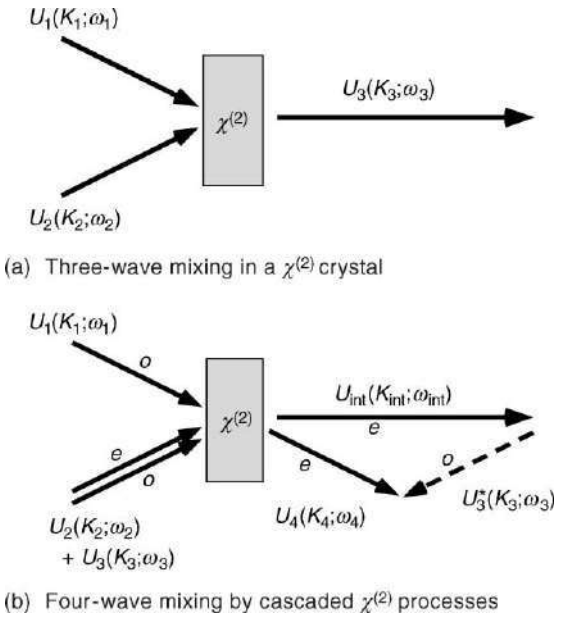

Figure 3 (a) Three-wave mixing interaction: two input waves U_1 and U_2, generate a third wave U_3. (b) Four-wave mixing by cascaded second order nonlinearities: three input waves, U_1, U_2, and U_3, generate a fourth wave U_4. The first two waves generate an intermediate wave U_{int}, by a frequency-sum process, followed by a frequency-difference process between the intermediate wave and the third input wave.

temporal frequency domain, and observe that the difference lies with the argument of the spectral term. If we apply the high-resolution approximation to eqn [8], then each spectral component would occupy a small region in space, as defined by the extent of the aperture function's SFT. Substituting for $\omega - \omega_0$ by $-\omega_0 x'/\alpha f$ for the spectral argument of eqn [8], corresponding to the high-resolution limit, equates the two expressions (essentially the same approximation as was performed in deriving eqn [9]).

Nonlinear Wave Mixing With Nondepleting Pumps

A nonlinear crystal exhibiting a large nonlinear susceptibility $\chi^{(2)}$ is placed at the Fourier plane of the spectral processing device. We employ noncollinear phase matching in our experiments. Either type-I or type-II phase matching conditions can be used in three-wave mixing experiments, with the resultant frequency-sum wave propagating along the bi-sector direction (see Figure 3a). Since both input waveforms are at ω_0 frequency, the generated field will be at a doubled carrier frequency, $2\omega_0$, due to the frequency-sum process. Four-wave mixing experiments use type-II phase matching in the $\chi^{(2)}$ crystal, involving a frequency-sum process generating an intermediate wave followed by a frequency-difference process between the intermediate wave and the third input wave. The resultant wave will co-propagate with one of the input waves and at the same temporal carrier frequency (propagation direction and frequency ω_0), albeit at an orthogonal polarization state (see Figure 3b). The output signal can be extracted using a polarizer or a polarizing beam splitter. The interacting input waves propagate noncollinearly at the Fourier transform plane by offsetting the signals from each other at the input plane of the spectral processing device.

While it is possible, though often difficult, to obtain a precise analytic solution to the coupled mode equations governing the wave-mixing processes for the generated waves, it is beyond the scope of this article whose focus is on the signal processing functionality. We therefore limit the analysis to weak interactions with nondepleting input waves. Using the nondepleted input waves approximation, the nonlinear polarization term, which drives the generation of new waves, is constant and an ordinary differential equation characterizes the generation process.

The nonlinear polarization for the frequency-sum process is proportional to the product of the input waves, when expressed in the time domain: $P_{NL}^{sum}(x';t) = \chi_{eff}^{(2)} U_1(x';t) U_2(x';t)$. The nonlinear polarization for the frequency-difference process is proportional to the product of an input waves and a conjugated input wave: $P_{NL}^{diff}(x';t) = \chi_{eff}^{(2)} U_1(x';t) U_2^*(x';t)$.

The solution of the ordinary differential equation driving the generation of the new wave, takes into account the possible phase mismatch between the propagation of the nonlinear polarization and the generated wave within the nonlinear crystal medium. Since we are interested in the signal processing functionality only, we approximate the generated waves as being directly proportional to the driving nonlinear polarization. This approach is valid for obtaining the generated signal forms in the case of short nonlinear crystals (limited interaction length) satisfying the phase matching condition. In practice, the temporal and spatial bandwidths, as well as the crystal length, will introduce filtering effects which will slightly modify the output signals. However, it can be shown that when wave mixing with spectrally decomposed waves, especially mutually inverted waves, the phase matching requirement is well satisfied and the filtering effects are negligible for the employed crystal lengths. In our experiments, we typically used 2 mm long β-barium borate (BBO) crystals. Using the simplified method for obtaining

the form of the generated waves, arrangements for synthesizing, processing, and detecting ultrafast waveforms can be explored.

Detection of Ultrafast Waveforms

Ultrafast optical waveforms are too rapid for acquisition by opto-electronic means alone, due to the latter's finite response time. An optical processor is usually required to convert a finite time window (typically in the tens to hundreds of ps range) to a stationary signal in space for detection by slower means. We mix two mutually inverted spectrally decomposed waves in the Fourier plane of a spectral processing device, and spatially Fourier transform the resultant wave to the output plane. As will be shown below, when one of the spectrally decomposed waves is generated by a transform limited short pulse, the output spatial signal carries the temporal information of the input waveform. We call this process a time-to-space converter.

Let two ultrafast waveforms, $s((t-t_0)/\tau)$ and $r(t/\tau)$, enter the spectral processing device from opposite directions (one strikes the grating at angle θ and the other at $-\theta$, see **Figure 4**). Each waveform will generate a spectrally decomposed wave in the spatial Fourier transform plane. The spectrally decomposed wave of the signal $s(t)$, as defined by eqn [9], yields the first input field:

$$U_1(x';t) \approx w\left(\frac{c(t-t_0)}{\alpha L}\right)\exp\left(-\mathrm{i}\frac{\omega_0(t-t_0)x'}{\alpha f}\right) \times \tilde{s}\left(-\frac{\omega_0 \tau x'}{\alpha f}\right)\exp(\mathrm{i}\omega_0 t) \qquad [11a]$$

whereas the inverted spectrally decomposed wave for $r(t/\tau)$ is given by

$$U_2(x';t) \approx w\left(-\frac{ct}{\alpha L}\right)\exp\left(\mathrm{i}\frac{\omega_0 t x'}{\alpha f}\right)\tilde{r}\left(\frac{\omega_0 \tau x'}{\alpha f}\right)\exp(\mathrm{i}\omega_0 t) \qquad [11b]$$

Without loss of generality, we set the time delay parameter in eqn [11b] to zero, such that the parameter t_0 in the spectrally decomposed wave of the signal waveform is a measure of the relative time delay of the two waveforms. With full knowledge of the two input waves, U_1 and U_2, we can express the resultant waveform of the three-wave mixing process as:

$$U_3(x';t) \propto P_{\mathrm{NL}}^{\mathrm{sum}}(x';t) = w\left(\frac{c(t-t_0)}{\alpha L}\right)w\left(-\frac{ct}{\alpha L}\right) \times \exp\left(\mathrm{i}\frac{\omega_0 t_0 x'}{\alpha f}\right)\tilde{s}\left(-\frac{\omega_0 \tau x'}{\alpha f}\right)\tilde{r}\left(\frac{\omega_0 \tau x'}{\alpha f}\right) \times \exp(\mathrm{i}2\omega_0 t) \qquad [12]$$

Next we apply a SFT to propagate the wave U_3 to the output plane, yielding:

$$U_{\mathrm{out}}(x'';t) = \int U_3(x';t)\exp\left(-\mathrm{i}\frac{2\pi}{\lambda_0 f/2}x'x''\right)\mathrm{d}x' = y\left(\frac{2\alpha x''}{c} - t_0\right)w\left(\frac{c(t-t_0)}{\alpha L}\right)w\left(-\frac{ct}{\alpha L}\right) \times \exp(\mathrm{i}2\omega_0 t_0) \qquad [13]$$

where $y(t) = s(t/\tau) \otimes r(-t/\tau)$, as the Fourier transform of a product is the convolution of their Fourier transforms (the $\otimes$-symbol denotes the convolution operation), and we use the spatial Fourier transform kernel at the doubled carrier frequency. The stationary spatial signal, $y(\cdot)$, is proportional to the convolution of the two input waveforms, where

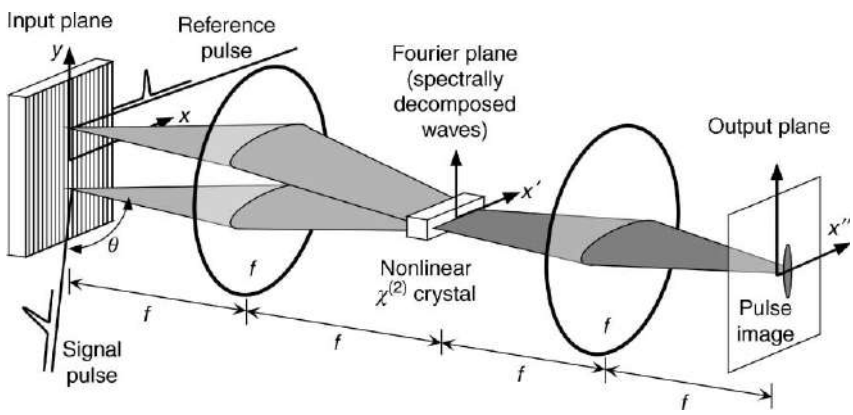

Figure 4 Arrangement for imaging ultrafast waveforms using a time-to-space conversion. The mutually inverted, spectrally decomposed waves of a signal waveform and that of a reference pulse are mixed in a nonlinear crystal placed at the Fourier plane. The generated wave is spatially Fourier transformed to the output plane, yielding a stationary image that carries the signal waveform information. © 2001 Optical Society of America.

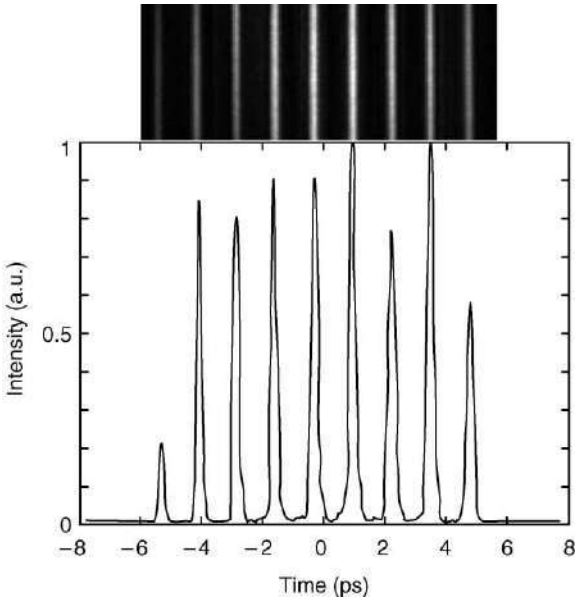

Figure 5 Output of a time-to-space converter. Top: image acquired by a CCD camera of the stationary image of a pulse sequence. Botttom: extracted time domain data from the image, showing inter-pulse time separation of 1.25 ps.

one of them is reversed in time. This mixing process was developed for time-to-space conversion by using a reference transform limited pulse for the reversed waveform. The resulting spatial image will be that of the signal waveform, albeit smoothed out due to the convolution operation with the reference pulse. The spatial signal is centered at $x'' = ct_0/2\alpha$, illustrating that the relative time delay between the two waveforms shifts the location of the output signal. The temporal characteristic is determined by the product of the time-scaled input apertures. The relative time delay, t_0, serves as an attenuation mechanism when the time windows are not perfectly aligned. When the time delay between the signal pulse and the reference pulse is too large, no output signal will emerge as the two signals do not overlap for the wave mixing process. Therefore, the time window of the spectral processing device is determined by the input aperture size, not the physical dimension of the nonlinear crystal. Consequently, we have been able to demonstrate high-resolution temporal imaging with time windows in the range of tens of picoseconds (see Figure 5). Additionally, as mentioned earlier, one of the significant advantages of this technique is the favorable phase matching condition with the spectrally decomposed wave and its inverted counterpart.

Processing of Ultrafast Waveforms

The simplest form of ultrafast waveform processing uses linear filtering, or placement of an element with transmittance characteristics that are frequency selective. However, a linear filter is a time invariant process and cannot perform more advanced processing operations of ultrafast waveforms such as interaction between ultrafast waveforms. For such processing applications, we mix three spectrally decomposed waves in the Fourier plane of a spectral processing device to generate a resultant output ultrafast waveform in real time. We illustrate advanced processing by performing a time reversal experiment using the spectral phase conjugation and spectral inversion techniques in four-wave mixing experiments.

Consider three ultrafast waveforms, $s_1(t - t_0/\tau)$, $s_2((t - t_1)/\tau)$ and $s_3(t/\tau)$, which enter a spectral processing device for generation of their respective spectrally decomposed waves. Let all three waves enter the spectral processing device from the same direction, such that the spatial dispersion is in the same direction for each signal (as defined by eqn [11a]), and generate the spectrally decomposed waves $U_1(x';t)$, $U_2(x';t)$, and $U_3(x';t)$, respectively (see Figure 6a). Since we utilize a cascaded second-order nonlinearity process, we let the first two waves interact first in a frequency-sum process, generating the intermediate field:

$$U_{\text{int}}(x';t) \propto P_{\text{NL}}^{\text{sum}}(x';t) = w\left(\frac{c(t-t_0)}{\alpha L}\right) w\left(\frac{c(t-t_1)}{\alpha L}\right) \times \exp\left(-\mathrm{i}\frac{\omega_0(2t-t_0-t_1)x'}{\alpha f}\right)\tilde{s}_1\left(-\frac{\omega_0 \tau x'}{\alpha f}\right) \times \tilde{s}_2\left(-\frac{\omega_0 \tau x'}{\alpha f}\right)\exp(\mathrm{i}2\omega_0 t) \quad [14]$$

It should be noted that upon a more rigorous analysis of the generated frequency-sum wave, phase matching is not well met across the temporal bandwidth of the waveform, as is the case when wave mixing mutually inverted spectrally decomposed waves. This will introduce time domain spectra filtering effects, but for simplicity these effects are ignored below. The generated frequency-sum wave interacts with the third input wave in a frequency-difference process, giving rise to the fourth wave:

$$U_4(x';t) \propto P_{\text{NL}}^{\text{diff}}(x';t) = w\left(\frac{c(t-t_0)}{\alpha L}\right) w\left(\frac{c(t-t_1)}{\alpha L}\right) \times w\left(\frac{ct}{\alpha L}\right)\exp\left(-\mathrm{i}\frac{\omega_0(t-t_0-t_1)x'}{\alpha f}\right) \times \tilde{s}_1\left(-\frac{\omega_0 \tau x'}{\alpha f}\right)\tilde{s}_2\left(-\frac{\omega_0 \tau x'}{\alpha f}\right)\tilde{s}_3^*\left(-\frac{\omega_0 \tau x'}{\alpha f}\right)\exp(\mathrm{i}\omega_0 t) \quad [15]$$

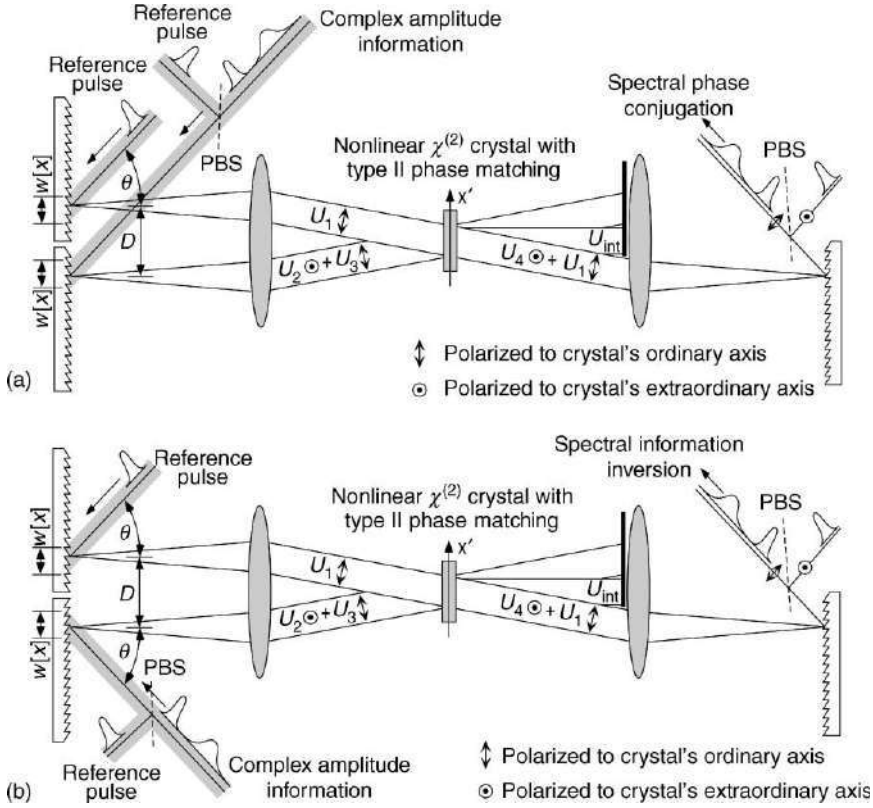

Figure 6 Arrangements of processing ultrafast waveforms within a spectral processing device. (a) Three input waveforms are introduced from the same direction, and their spectrally decomposed waves interact via four-wave mixing. (b) The three input waveforms are introduced from opposite directions, giving rise to mutually inverted spectrally decomposed waves. These two arrangements can perform time reversal, the first by spectral phase conjugation and the second by spectral information inversion. © 2001 IEEE.

where we assume that the window function $w(\cdot)$ is real. Next we apply a SFT to propagate the wave U_4 to the output plane, yielding:

$$\begin{aligned} U_{\text{out}}(x'';t) &= \int U_4(x';t)\exp\left(-\mathrm{i}\frac{2\pi}{\lambda_0 f}x'x''\right)\mathrm{d}x' \\ &= y\left(\frac{\alpha x''}{c}+t-t_0-t_1\right)w\left(\frac{c(t-t_0)}{\alpha L}\right) \\ &\quad\times w\left(\frac{c(t-t_1)}{\alpha L}\right)w\left(\frac{ct}{\alpha L}\right)\exp(\mathrm{i}\omega_0 t_0) \end{aligned} \quad [16]$$

where $y(t)=s_1(t/\tau)\otimes s_2(t/\tau)\otimes s_3^*(-t/\tau)$. The new ultrafast waveform $y(\cdot)$ scans along the output plane at a velocity of $-c/\alpha$ in the x-direction (compare to the input signals scanning in the opposite direction along the input plane, eqn [6]). By placing another diffraction grating at this plane, the new ultrafast waveform can be recomposed and made to propagate once more in free space. The waveform information consists of a convolution between the first and second signals, followed by a convolution with the third signal, which is both conjugated and reversed. If we set the first and second signals to be transform limited pulses, then the resultant waveform will carry the reversed and conjugated information of the third signal. Such capability can be very important for dispersion compensation in optical fiber communication links.

Let us now consider an alternate arrangement, where the second and third input signals are introduced to the spectral processing device from the opposite direction, and consequently have reversed spatial dispersion direction (see Figure 6b). Therefore, the first signal will generate a spectral decomposed wave $U_1(x';t)$, as defined by eqn [11a], whereas the second and third signals will give rise to the spectrally decomposed waves $U_2(x';t)$ and $U_3(x';t)$, with inverted spectra relative to that of U_1, as defined by eqn [11b]. The interaction of the first two waves gives rise to:

$$\begin{aligned} U_{\text{int}}(x';t) &\propto P_{\text{NL}}^{\text{sum}}(x';t) = w\left(\frac{c(t-t_0)}{\alpha L}\right) \\ &\quad\times w\left(-\frac{c(t-t_1)}{\alpha L}\right)\exp\left(\mathrm{i}\frac{\omega_0(t_0+t_1)x'}{\alpha f}\right) \\ &\quad\times \tilde{s}_1\left(-\frac{\omega_0\tau x'}{\alpha f}\right)\tilde{s}_2\left(\frac{\omega_0\tau x'}{\alpha f}\right)\exp(\mathrm{i}2\omega_0 t) \end{aligned} \quad [17]$$

which is identical to the wave mixing process of the time-to-space converter (see eqn [12]). However, in this case, we follow the first wave mixing process by a frequency-difference process, generating:

$$U_4(x';t) \propto P_{\mathrm{NL}}^{\mathrm{diff}}(x';t) = w\left(\frac{c(t-t_0)}{\alpha L}\right) w\left(-\frac{c(t-t_1)}{\alpha L}\right) \times w\left(-\frac{ct}{\alpha L}\right) \exp\left(-\mathrm{i}\frac{\omega_0(t-t_0-t_1)x'}{\alpha f}\right) \tilde{s}_1\left(-\frac{\omega_0 \tau x'}{\alpha f}\right) \times \tilde{s}_2\left(\frac{\omega_0 \tau x'}{\alpha f}\right) \tilde{s}_3^*\left(\frac{\omega_0 \tau x'}{\alpha f}\right) \exp(\mathrm{i}\omega_0 t) \quad [18]$$

Next we apply a SFT to propagate the wave U_4 to the output plane, yielding the output waveform:

$$U_{\mathrm{out}}(x'';t) = \int U_4(x';t) \exp\left(-\mathrm{i}\frac{2\pi}{\lambda_0 f} x' x''\right) \mathrm{d}x' = y\left(\frac{\alpha x''}{c} + t - t_0 - t_1\right) w\left(\frac{c(t-t_0)}{\alpha L}\right) \times w\left(-\frac{c(t-t_1)}{\alpha L}\right) w\left(-\frac{ct}{\alpha L}\right) \exp(\mathrm{i}\omega_0 t_0) \quad [19]$$

where $y(t) = s_1(t/\tau) \otimes s_2(-t/\tau) \otimes s_3^*(t/\tau)$. As before, the new ultrafast waveform $y(\cdot)$ scans along the output plane at a velocity of $-c/\alpha$ in the x-direction, and will propagate in free-space upon another grating diffraction. This time, however, the waveform information consists of a convolution between the first signal and a time-reversed second signal, followed by a convolution with the conjugated third signal. If we set the first and second signals to be transform limited pulses, then the resultant waveform will carry the conjugated information of the third signal. If we set the first and third signals to be transform limited pulses, then the resultant waveform will carry the reversed information of the second signal.

We have experimentally demonstrated the two time-reversal techniques: the first technique whereby the resultant signal is both reversed and conjugated is labeled 'spectral phase conjugation', as the signal's spectrum is conjugated by the frequency difference process, and the second technique, where the signal is purely reversed in time, is named 'spectral information inversion', as the spectrum is flipped about the carrier frequency (see Figure 7). By proper selection of the input waveforms, one can also obtain higher-level operations and functionality such as convolutions and correlations among the different signals. All these operations are time-variant, and cannot be performed by passive filtering alone.

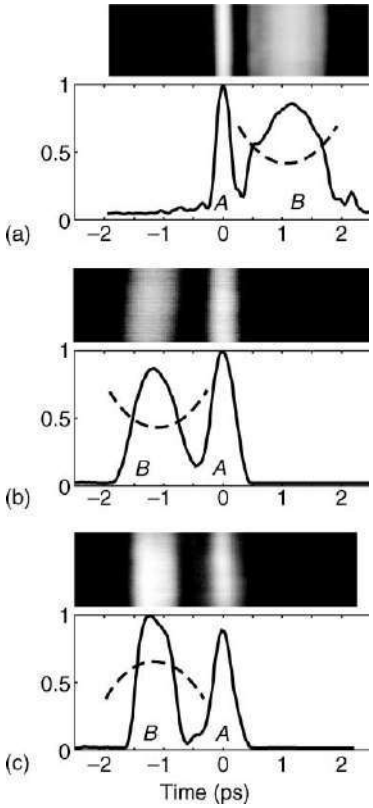

Figure 7 Time reversal experimental results. (a) Input pulse pair consisting of a transform limited pulse followed by a chirped pulse. (b) Time reversal by spectral information inversion technique. Chirped pulse leads transform limited pulse and maintains the sign of the chirp. (c) Time reversal by spectral phase conjugation technique. Chirped pulse leads transform limited pulse and flips the sign of the chirp. Data acquired with time-to-space imaging technique and sign of quadratic term (dashed line) determined by translating output plane.

Synthesis of Ultrafast Waveforms

Shaping of ultrafast waveforms is usually performed by filtering an input transform limited pulse by a fixed filter. Adaptive filtering can be performed by various means incorporating opto-electronic modulators such as liquid-crystal-based or acousto-optic-based spatial light moduloators, however the response time required for changing the waveform is typically limited to the ms and μs regime, respectively. For ultrafast reconfigurability, we utilize our spectral processing device and wave mixing, to achieve pulse shaping controlled by light, or a second optical signal.

Let the three input signals entering the spectral processing device be an ultrashort waveform, $s(t/\tau)$, and two additional spatial domain signals; one carrying spatial information $m(x/\chi)$, where χ is a scaling factor, and the second a point source (see Figure 8). Both spatial domain signals are illuminated

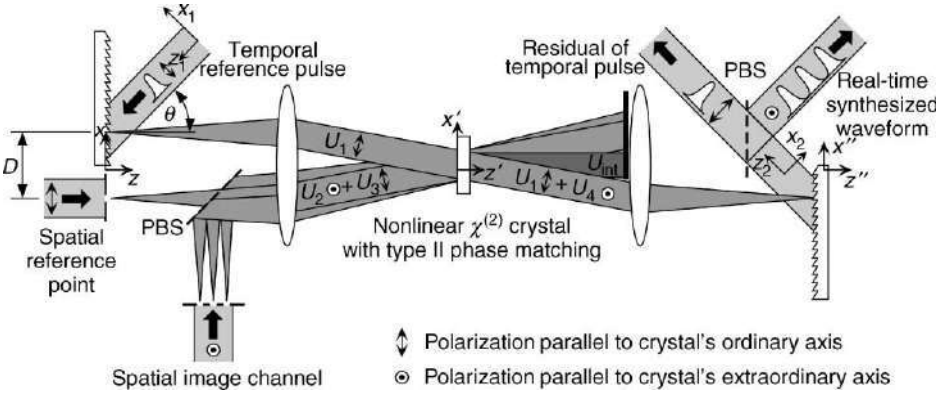

Figure 8 Arrangement for ultrafast waveform synthesis by wave mixing temporal and spatial information channels. Three input channels consist of a temporal channel, a spatial channel carrying information, and a second spatial channel consisting of a point source. The information from the spatial channel is transferred to the temporal channel via the wave mixing process. © 2001 IEEE.

by a monochromatic light source at frequency ω_0. The input ultrashort waveform gives rise to a spectrally decomposed wave $U_1(x';t)$, as described by eqn [11a]. The spatial information channel is SFT by a lens, generating the wave:

$$U_2(x';t) \propto \tilde{m}\left(\frac{\omega_0 \chi x'}{2\pi c f}\right)\exp(\mathrm{i}\omega_0 t) \qquad [20]$$

The signal from the point source (or $\delta(x)$) gives rise to a wave $U_3(x';t)$ with no spatial information (plane wave). As before, we utilize cascaded second-order nonlinearities for the wave mixing process. The interaction of the first two waves gives rise to the intermediate frequency-sum wave:

$$U_{\mathrm{int}}(x';t) \propto P_{\mathrm{NL}}^{\mathrm{sum}}(x';t) = w\left(\frac{ct}{\alpha L}\right)\tilde{s}\left(-\frac{\omega_0 \tau x'}{\alpha f}\right)\tilde{m}\left(\frac{\omega_0 \chi x'}{cf}\right)\exp(\mathrm{i}2\omega_0 t) \qquad [21]$$

The frequency-difference process with an informationless waveform serves to reduce the carrier frequency of resultant waveform back to ω_0. The down-converted wave U_4 is SFT to the output plane, yielding:

$$U_{\mathrm{out}}(x'';t) = y\left(\frac{\alpha x''}{c} + t\right)w\left(\frac{ct}{\alpha L}\right)\exp(\mathrm{i}\omega_0 t_0) \qquad [22]$$

where $y(t) = s(t/\tau) \otimes m(-ct/\alpha\chi)$. As in the processing of ultrafast waveforms section above, the new ultrafast waveform $y(\cdot)$ scans along the output plane at a velocity of $-c/\alpha$ in the x-direction. The synthesized ultrafast waveform can propagate in freespace after an additional grating diffraction. The waveform information, in this case, consists of a convolution between the input temporal signal and a time-scaled and reversed version of the spatial information channel. Since the input temporal channel is typically a transform limited pulse, the output waveform carries the spatial information image, mapped to the time domain. In analogy to the imaging of ultrafast waveforms section, we call this technique space-to-time conversion. We have generated experimentally numerous waveforms that are controlled by a space domain image, or mask (flat-top waveforms, Figure 9, and pulse sequences of Figure 5 were generated by a cylindrical lens array in the spatial information channel).

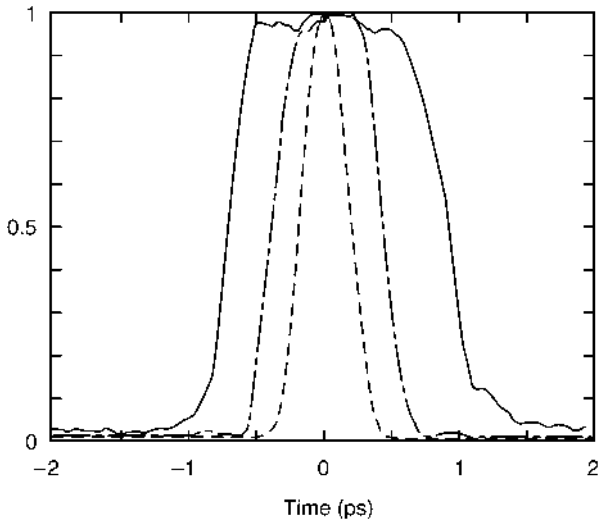

Figure 9 Synthesized flat-top waveforms via space-to-time conversion utilizing a variable slit in the spatial information channel. Increasing the width of the slit results in broader square-shaped pulses in real-time. Solid line: 2 ps pulse, dot-dash line: ~1 ps pulse, dashed line: ~300 fs pulse.

Conclusions

To properly utilize ultrafast waveforms in scientific and advanced technology applications requires instantaneous techniques from generation to processing and imaging of the waveforms. We have studied the nonlinear wave-mixing approach within a spectral processing device, and demonstrated the ability to perform all of these tasks. The wave mixing was performed in a second-order nonlinear crystal, using three-wave mixing for detection and four-wave mixing for synthesis and processing of ultrafast waveforms. The three-wave mixing process results in a second-harmonic output image, suitable for acquisition by silicon-based CCD imagers. The four-wave mixing processes result in an output waveform of an identical carrier frequency, as typically required in processing and synthesis applications. The high nonlinear coefficient of $\chi^{(2)}$ crystals enabled us to demonstrate these techniques with ultrashort pulse energies ranging from nJ to mJ levels. Further advantages afforded by the techniques are expanded time windows and favorable phase matching.

This review of the instantaneous optical signal processing capabilities with optical nonlinearities and interaction with ultrafast waveforms introduced and demonstrated the basic signal processing capabilities to manipulate optical signals in amplitude and phase. The same techniques can be used and expanded upon to achieve different signal processing abilities, which are beyond the scope of this article.

See also

Coherent Transients: Ultrafast Studies of Semiconductors. **Nonlinear Optics, Basics:** Ultrafast and Intense-Field Nonlinear Optics. **Ultrafast Laser Techniques:** Pulse Characterization Techniques. **Ultrafast Technology:** Femtosecond Chemical Dynamics: Gas-Phase; Femtosecond Condensed Phase Spectroscopy: Structural Dynamics; Ultrafast Illumination and Processing.

Further Reading

Diels J-C and Rudolph W (1996) *Ultrashort Laser Pulse Phenomena: Fundamentals, Techniques, and Applications on a Femtosecond Time Scale*. San Diego: Academic Press.

Marom DM, Panasenko D, Sun P-C and Fainman Y (2000) Femtosecond-rate space-to-time conversion. *Journal of the Optical Society of America B* 17: 1759–1773.

Marom DM, Panasenko D, Sun P-C, Mazurenko YY and Fainman Y (2001) Real-time spatial-temporal signal processing with optical nonlinearities. *IEEE Journal of Selected Topics in Quantum Electronics* 7: 683–693.

Miller DR, Murnane MM, Scherer NF, Weiner AM (eds) (2003) *Ultrafast Phenomena XIII, Proceeding of the 13th International Conference*. Berlin: Springer-Verlag.

Weiner AM (1995) Femtosecond optical pulse shaping and processing. *Progress in Quantum Electronics* 19: 161–238.

Weiner AM and Kanan M (1998) Femtosecond pulse shaping for synthesis, processing, and time-to-space conversion of ultrafast optical waveforms. *IEEE Journal of Selected Topics in Quantum Electronics* 4: 317–331.

EARLY HISTORY OF QUANTUM ELECTRONICS

C H Townes, University of California, Berkeley, CA, USA

© 2005, C H Townes. Published by Elsevier Ltd. All rights reserved.

All the ideas essential to making a laser were known before 1930, but there was no operating laser before 1960. So why didn't the laser come sooner? There are several reasons. One important impediment to the laser invention was that a combination of ideas from quantum mechanics and from electrical engineering was needed, and these two fields were not well mixed in the early days. Another is that, while some physicists recognized that amplification could occur if there was a population inversion of states, they did not consider the coherence of such amplifications and did not recognize its importance or usefulness. It was just something that in principle could happen, but was not very interesting. What was required was a combination of physics and electrical engineering thinking and recognition of the coherence of such amplification. In addition the importance of such a development had to be visualized and recognized in a way that it led to very devoted work towards its achievement.

It is striking that lasers grew out of the study of microwave spectra of molecules, for which engineering and quantum mechanics were both important and which helped orient thinking in the appropriate direction. That this origin was not accidental is convincingly demonstrated by the fact that three independent ideas for amplification by stimulated emission were generated about 1950. They were from Joe Weber, at the University of Maryland, from Nikolai Basov and Alexander Prokhovov at the Soviet Academy, and myself at Columbia University. Weber primarily wanted to point out the possibility, but didn't try to do it. In addition, his numbers were a bit off and no practical system was suggested. Basov and Prokhovov actually worked towards microwave amplification using a beam of molecules, as did I.

That stimulated amplification was recognized early, but not thought through, is illustrated by the fact that Prof. Richard Tolman, a theoretical physicist, wrote a discussion in 1924 of the net absorption of light by molecules, pointing out in particular that induced emission counteracted absorption and noting that if there were more molecules in the upper than in the lower state there could be 'negative absorption'. But, he wrote, 'This would usually be very small'. The Russian physicist Vitaly Ginsburg wrote me, after the maser and laser had appeared, that his professor, S.M. Levi, had been well aware of such effects back in the 1930s and had told him 'create an overpopulation at higher atomic levels and you will obtain an amplifier; the whole trouble is that it is difficult to create a substantial overpopulation of levels'.

The German physicist F.G. Houtermans said to me that in 1932, when told by a colleague of an unusual light intensity in a gaseous discharge, he had thought it might be a 'photon avalanche', i.e. multiplication of photons by stimulated emission.

Another Russian physicist, V.A. Fabricant, wrote a thesis in 1939 in which he discussed absorption and emission of light radiation in a gas and looked for 'negative absorption', or amplification. He did not discuss coherence or a resonant cavity, and was not able to achieve any amplification so his work was quickly forgotten. None of these early mentions of stimulated emission proposed how to actually get amplification, that it would be useful, nor clearly noted its coherence. Tolman did, however, write in 1927 that, 'We should expect radiation induced by an external field to be coherent with the radiation associated with that field'. I know of no proposal to actually make use of such amplification, until those made by microwave spectroscopists in the early 1950s.

Another clear indicator that even in the 1950s physicists and engineers did not think amplification by stimulated emission was particularly interesting nor useful is that during the $2\frac{1}{2}$ years when Jim Gordon, Herb Zeiger, and I were working on trying to obtain microwave amplification (the maser), a large number of scientists visited my laboratory, saw what we were doing, but no one bothered to also try to obtain such amplification. After the maser worked, it hit the newspapers and then very quickly became a popular and intense field of interest for a number of physicists.

My own drive to produce oscillators by stimulated emission came from my strong interest in obtaining sources of waves shorter than the few millimeters wavelength which could be produced by electronic devices, in order to extend the high resolution study of molecular spectra to wavelengths shorter than microwaves, down into the far infrared. My students and I worked on several possible schemes for producing waves shorter than those produced by klystrons or magnetrons – frequency multiplication by nonlinearities, electronic beams passing over surfaces of solid materials with resonances, and anything else I could think of. None worked very well.

In early 1950, I was asked by the Office of Naval Research to form and chair a committee which would examine possible research towards obtaining wavelengths down to one millimeter and shorter. I chose outstanding scientists and engineers in a variety of fields which might touch on this problem. We met together, and visited many pertinent laboratories and individuals interested in such research. Nothing very promising seemed to turn up. But we wanted, of course, to at least provide a report summarizing the situation as we saw it. Our last meeting was April 26, 1951, in Washington, D.C. Worrying over our lack of success, I woke up early before the meeting. Breakfast was not ready, so I walked over to nearby Franklin Park, sat on a bench in front of beautiful blooming azaleas, and puzzled over why neither I nor the Committee had found any promising solutions.

I went over all the ideas I had had previously. Molecules, of course, can produce high frequencies. But I had previously concluded, I thought wisely and rigorously, that one could not obtain intense radiation from them because radiation intensity was a function of temperature, and the temperature could not be very high without destroying the molecules. Suddenly I realized that they did not need to have a temperature in the usual sense – they need not be in temperature equilibrium. There could be more molecules in an upper than a lower excitation state, which could in principle produce indefinitely intense radiation. I pulled a piece of paper out of my pocket and wrote down the equations and numbers for such a case, using a molecule beam sent into a resonant cavity and ammonia molecules with which I was very familiar. My equations said one could get enough excited molecules, low enough loss in a resonant cavity, and it would work! Why hadn't I thought of it before?!

Back at Columbia, where I worked, and about 4 months later, the graduate student Jim Gordon agreed to work on trying to obtain such an oscillator using a beam of ammonia. I assured him that if it didn't work he could modify the experiment to do interesting spectroscopy and thus complete a thesis. But we both thought he had a good chance of making it work. And a young post doc working with me, Herb Zeiger, joined the effort.

Actually, I had first thought of obtaining stimulated emission from a molecular beam back in 1948, but simply as a demonstration of physical principles rather than as a useful amplification. Also, about a year later a young post doc, J.W. Trischka, working on molecular beams with Professors Rabi and Kusch, had also thought of demonstrating stimulated emission. We talked about it together, and he decided it wasn't worthwhile just demonstrating the effect because it wouldn't really prove any new physics. Neither of us at that time had recognized the real point and the possibility of useful amplification, which is one of the reasons mentioned above that the idea was delayed as long as it was.

It is perhaps important to emphasize again and to illustrate how out-of-the-way the use of stimulated emission was at that time for physicists, and how distant stimulated emission was from engineers. While we were working on the ammonia beam maser, Prof. L.H. Thomas, an outstanding physicist known for the 'Thomas Effect', frequently would run into me in the hallway at Columbia University and say that I didn't understand, and that my proposed ammonia oscillator could not work. However, I never got a clear explanation from him of what it was I didn't understand. And after we had been working on the ammonia maser system for about 2 years (a more-or-less normal time for a student thesis project), the Physics Department Chairman, Prof. Polycarp Kusch, and the previous chairman, Prof. I.I. Rabi, came into my office to object. They were excellent physicists, both received Nobel Prizes, and both were experts on molecular beams. They sat down in my office and said 'Look, Charlie, that is not going to work. We know it won't work and you know it won't work. You are wasting departmental money, and must stop!' Other people had also questioned what I was doing, frequently in particular whether the stimulated radiation would be coherent. So I had already thought over the situation many times. I had full notes from the quantum mechanics course I took as a student back in 1938, and could derive from them a proof of coherence. I felt also that I knew the quantitative numbers, such as the molecular beam intensity and the possible 'Q', or loss, in the cavity resonator, very well. I was also by then an Associate Professor, and the department chairman could not fire me simply because of stupidity. I replied to Rabi and Kusch 'No, I believe it has a reasonable chance of working and I'm going to continue!' Annoyed, they stomped out of the room.

About 2 months after the Rabi/Kusch incident, Jim Gordon dashed into the classroom where I was lecturing and said loudly 'It's working!' Most of the class then went up to the lab to see the new device. Rabi and Kusch were not against me; even though they were outstanding physicists, they probably just didn't quite comprehend the device. A couple of months after its successful operation, Kusch more or less apologized by saying 'Well, I should have realized that you probably know more about what you are doing than I do.'

After the maser successfully operated, there were other incidences showing the lack of appropriate

focus of physicists on stimulated emission. I was a friend of Aage Bohr, the son of Niels Bohr, and in that connection was visiting him in Denmark. Niels Bohr asked me what research I was presently doing, so I told him about our new oscillator, the maser, and its remarkably pure frequency. He looked at me and said 'Oh no, that's not possible. You must be misunderstanding something.' I emphasized again what it was really doing, but he still seemed not to believe it could function that way. I presumed he was thinking in terms of the uncertainty principle and the finite time of passage of a molecule through the cavity, though I never was quite sure just why he felt it impossible. A similar thing happened shortly after that at a cocktail party in Princeton, where John Von Neumann asked the same question – what was my research at the moment? After telling him about the maser oscillator and the purity of frequency he reacted very similarly 'Oh no,' he said, 'That can't be right. You must be misunderstanding something.' After arguing a little more, he left to get another drink. Fifteen minutes later he came back, saying 'Hey, you are right!' He had understood, and wanted to talk much more about the maser and of possibly using excited semiconductors. Only after his death did I learn from his notebooks that he had considered exciting electrons in semiconductors with neutrons from a reactor, and had written Edward Teller about whether some experiments might be done with this to obtain intense light. However, he had not considered coherence, and Teller was apparently not interested enough to respond, so the matter was dropped.

The above account reminds me a bit of the amusing comments of Arthur Clark on 'Change'. He writes:

'People go through four states before any revolutionary development:

1. It's nonsense, don't waste my time
2. It's interesting, but not important
3. I always said it was a good idea
4. I thought of it first'.

It's clear that the world of physics was thinking very little in the direction of masers or lasers, and that many preconceptions as well as lack of interest stood in the way. My own experience with engineering at Bell Telephone Labs during World War II, in designing radar and electronic systems, plus my intense interest in obtaining short-wave oscillators, were clearly important in bringing me to the right ideas.

As masers became very interesting to the physics community and the field grew rapidly, my engineering experiences continued to be of help. I was well acquainted with the theoretical examination of noise in vacuum tube amplifiers by various individuals at Bell Labs, and recognized that the maser could provide much more sensitive amplification than could common electronic amplifiers, where discrete electronic charges produce the basic noise. On sabbatical in France, I worked on electron spin masers with Jean Combrisson and Arnold Honig who had the appropriate equipment. And then in Japan, Koichi Shimoda, Hidetoshi Takahashi, and I wrote a theoretical paper on the basic quantum noise of maser (or laser) amplification. The theory showed that maser amplifiers of microwaves should be about 100 times more sensitive than the existing electronic ones.

After 2–3 years of maser experiments and development I felt I wanted to move on to the shorter wavelengths for which I had generated the maser idea. Although I had first tried the idea at microwave frequencies because that seemed the easiest way to test out the general idea and the result had been exciting, I still wanted those shorter wavelengths. I had not come up with any great ideas of just how to get to much shorter wavelengths, which is why I waited several years after the maser worked before moving on. However, in 1957, 3 years after the first successful operation of the maser, I decided it was high time to simply figure out what was the best way I could imagine to move on into the infrared region and do it. A number of physicists had concluded that of course one couldn't make masers work at much shorter wavelengths, certainly not in the visible region, because spontaneous emission becomes so much faster as the wavelength is shortened and adequate inversion of population was not practical. But that was intuition, not quantitative science. As I wrote down equations for what might be done to move towards shorter wavelengths using a model of atomic or molecular excitation by radiation and a reasonably high Q resonator, it became clear that it was quite practical to move on even into the visible region. That was exciting! Why hadn't I, or someone else, looked at it carefully and quantitatively before that time?

I was at the time consulting at Bell Labs, with the assignment to spend a day there every 2 weeks and just talk with the Lab's scientific personnel. Since my former post doc and now brother-in-law Arthur Schawlow was then at Bell Labs, I of course talked with him. On telling him of my ideas for an 'optical maser', using a resonant cavity and excitation of atoms with optical radiation, he said he too was interested in that, and we decided to work together to optimize a system. It was Art who then suggested use of a Fabry–Perot resonator rather than the cavity with large holes I had used for a model, and that was

an excellent addition. Why I did not think of that is a mystery, but Art had done his thesis at the University of Toronto in Fabry–Perot spectroscopy, and that might have been why the thought came to him.

Since Art Schawlow was participating, I decided the patent for the new 'optical maser' should belong to the Bell Labs (I already claimed ownership of the basic maser patent, which covered all wavelengths). Hence we recognized that the new idea must be kept confidential until Bell Labs lawyers had worked out and applied for an appropriate patent. This delay in public information sheds some additional light on how the scientific and technical world was thinking at the time. I had written down my original ideas for an 'optical maser' in my notebook and had it witnessed by my student Joe Giordmaine at Columbia University. I had also talked with a Columbia student Gordon Gould because he had been doing research with an intense light source and I wanted to know how much intensity he had in order to be sure I could get enough excited atoms and provide an oscillator at these short wavelengths. Except for these two persons, neither Art Schawlow nor I told anyone outside of Bell Labs about the 'optical maser', or laser idea until after Bell Labs had properly prepared its patent, which was about 11 months after my first notebook entry. For that entire time, no one has produced a record of any thoughts about extending the maser to optical wavelengths except Gordon Gould, who entered some ideas in his notebook about 1 month after I talked with him about the possibility of an optical maser, and 2 months after my original notebook entry. His notes were later to be the source of a long patent case.

The striking observation is that no one outside of Bell Labs except Gordon Gould, to whom I had explained my ideas, seems to have written or noted down anything about extending masers to these shorter wavelengths during the 11 months from my recognition that it could well be done until after the Schawlow and Townes paper on how to do it became available. After that there was considerable excitement and a number of ideas.

It is also noteworthy that, because of low general interest and competition, I did not publish a paper on how a maser might be made, but waited for publication until we demonstrated its operation, about 3 years after the idea had arrived. But after the maser worked the field became exciting and competitive. Hence, Schawlow and I thought we surely should publish a theoretical paper establishing the idea before taking the time to make a laser work. And indeed, there was much competition to build the first laser.

I myself helped a couple of my graduate students start work on trying to build an 'optical maser' or laser. But at about that time (1958), I was urged to undertake a job in Washington as Vice President and Director of Research of the Institute for Defense Analysis, an organization put together largely by the presidents of several universities to try to help advise the government on matters of science and technology. Sputnik had gone up the year before, and everyone was worried about the position of the U.S. with respect to Communist Russia, which seemed to be ahead particularly in some areas important to defense. I decided I should try to help, and accepted a two-year appointment in Washington. I recognized that this seriously distracted my attention from developing a laser quickly, but knew there were many others working towards lasers and so such devices would certainly be developed and our doing it just at Columbia University was neither critically important nor highly probable.

The field of masers and 'optical masers' or lasers was becoming so exciting and a bit scrambled that The Office of Naval Research asked me if I would organize a meeting on the subject. I did, with the help of a committee of many distinguished people in the field or closely related science and technology. And it was in a meeting of the Committee that we christened the field with the name 'Quantum Electronics'. This first international meeting on the subject was at Shawanga Lodge in New York State in September 1959. It made an occasion for the Russians Basov and Prokhorov to visit the United States (and my lab and home), and was about 6 months before Ted Maiman made the first working laser. There were many interesting discussions of masers and their coming operation at optical wavelengths.

It is significant to note that, while the maser grew out of basic research in universities (with Russian work at the Russian Academy) and industry had little to do with early masers, all the first lasers were made in industrial laboratories. This illustrates the sociology and the strengths and weaknesses of industrial and of academic laboratories. While masers (and from them lasers) originated from research on microwave spectroscopy of molecules, industrial laboratories believed the field of microwave spectroscopy had little to offer in the way of commercial results. Because of equipment available and the interest of physicists in industry, the field was initiated and pursued immediately after World War II in commercial laboratories – by myself at Bell Labs, my friends at the RCA Labs, at Westinghouse, and at General Electric. For lack of interest in industry, such work was soon shut down,

except at Bell Labs, and it moved to universities. Bell Labs generously allowed me to continue such work, although they wanted me to do some 'more useful' engineering. Once the field had obvious commercial possibilities, commercial laboratories began to support it well and the clear importance of masers and lasers made industry very interested. Really interested industry can more easily put strong new resources and push harder on a field than can academic laboratories, where money has to be granted and professors have a variety of other assignments. The first laser was made to work, of course, by Ted Maiman at Hughes Research Laboratories. Ted had been a student of Willis Lamb at Stanford, and there worked on radio spectroscopy. The second type of laser, rather similar to Maiman's but using a different material, was made to work at the General Electric Laboratories by Peter Sorokin and Mirek Stevenson. Sorokin had been a student of Bloembergen at Harvard in microwave or radio spectroscopy and Stevenson was one of my students in microwave spectroscopy at Columbia University. The next type of laser, and one I particularly admire, was made by Ali Javan, William Bennett, and Don Herriott. Javan had been a student with me in microwave spectroscopy at Columbia University, Bill Bennett a student in the molecular beam group at Columbia working on radio spectroscopy, and Don Harriott was an optics specialist. All of these originators, with the exception of Herriott, had been working at universities in the field which originated the idea and were recently hired by industry. The next important laser, involving semiconductors, was invented by Robert Hall and collaborators at General Electric Labs. Note that every one of these early lasers was created in industry.

After the first laser was operated, my own students at Columbia quickly turned from trying to make a laser to using lasers to explore more physics, the normal university function. And I am delighted that masers and lasers have provided such excellent tools for research, as well as for commercial and medical applications.

After a few years of exploring new physics with lasers, I decided that since there were many excellent scientists doing such work, I should move into fields which it seemed to me were being relatively neglected. I moved to the University of California at Berkeley to look for molecules in interstellar space by microwave spectroscopy, and to do infrared astronomy. Very soon after initiating work at Berkeley, one of my students, Albert Cheung, not only discovered the first polyatomic molecules in space, he found powerful water masers. A while before our discovery of water and identification of its radiation as due to maser action, it had been deduced that some microwave radiation of OH must be due to maser action. And since then, many, many masers due to a wide variety of molecules have been found in astronomical sources as well as a few powerful lasers. Since deviations from thermal equilibrium are common in the very thin gases excited by powerful sources in space, we should have expected this, but didn't. And these masers in space could have been easily detected with radio technology available back in the 1930s if anyone had searched the microwave spectrum carefully.

Clearly, masers and lasers could have been discovered and used much sooner than they actually were. We simply were neither thinking nor looking in the right directions. And this raises the natural question – what important and more-or-less obvious ideas are we missing now because of our lack of imaginative exploration?

Further Reading

Basov N and Prokhorov A (1954) Application of molecular beams to radio spectroscopic studies of rotation spectra of molecules. *J. Experimental. Theoretical Physics U.S.S.R.* 27: 431–438.

Bloembergen N (1956) Proposal for a new type of solid state maser. *Physical Review* 104: 324–327.

Cheung A, Rank D, Townes C, Thornton D and Welch J (1969) Detection of water in interstellar regions by its microwave radiation. *Nature* 121: 626–628.

Combrisson J, Honig A and Townes C (1956) Utilization de la resonance de spins electroniques pour realiser un oscillateur ou un amplificateur en hyperfrequencies. *Comptes Rendues* 242: 2451–2453.

Gordon J, Zeiger H and Townes C (1954) Molecular microwave oscillator and new hyperfine spectrum of NH_3. *The Physical Review* 95: 282–284.

Hall R, Fenner G, Kingsley G, Soltys J and Carlson R (1962) Coherent light emission from GaAs junction. *Physical Review Letters* 9: 366–368.

Javan A, Bennett W and Herriot D (1961) Population inversion and continuous optical maser oscillation in a gas discharge containing a He–Ne mixture. *Physical Review Letters* 6: 106–110.

Knowles S, Mayer C, Cheung A, Rank D and Townes C (1969) Spectra, variability, size, and polarization of H_2O microwave emission sources in the galaxy. *Science* 163: 1055–1057.

Maiman T (1960) Stimulated optical radiation in ruby. *Nature* 187: 493–494.

Perkins F, Gold T and Salpeter E (1966) Maser action in interstellar OH. *Ap. J.* 145: 361–366.

Schawlow A and Townes C (1958) Infrared and optical masers. *Physical Review* 112: 1940–1949.

Shimoda K, Takahashi H and Townes C (1957) Fluctuations in amplification of quanta with

application to maser amplifiers. *Journal of the Physical Society of Japan* 12: 686–700.

Sorokin P and Stevenson M (1960) Stimulated infrared emission from trivalent uranium. *Physical Review Letters* 5: 557–559.

Tolman R (1924) Weak quantization. *The Physical Review* 24: 287–295.

Tolman R (1927) Statistical mechanics with applications to physics and chemistry. The Chemical Catalog Co., New York, pp. 169.

Weber J (1953) Amplification of microwave radiation by substances not in thermal equilibrium. *Trans. Inst. Radio Engineers Professional Group on Electron Devices* 3: 1–4.

LET THERE BE COHERENT LIGHT: THE DISCOVERY OF THE LASER

T H Maiman, Vancouver, BC, Canada

© 2005, T H Maiman. Published by Elsevier Ltd. All rights reserved.

Editor's Note

The discovery of the first laser was a remarkable culmination of human creativity and ingenuity. However the achievement was finally made by one individual young scientist, Theodore Maiman, working independently of the mainstream and without the benefit of strong government or private financial support. With persuasive arguments, he convinced his management at the Hughes Research Laboratories to allow him to continue his work, albeit with limited resources. And then, working alone with his assistant he saw the light that now illuminates the world of laser science and engineering. This is the story of how thinking outside the box created a new field of science, whose developing technology has advanced the state of medicine and engineering and the quality of life in general.

Introduction

I am frequently asked how it was that I discovered the laser. Many assume that the concept evolved from some sudden, inspirational thought. It didn't happen that way.

It is dramatic and exciting to have a scientific discovery emerge from a dream or a vision that comes out of nowhere, but it seldom does. In reality, almost all scientific discoveries come from building on other, prior, scientific developments.

So it was with the laser. In 1916, Albert Einstein laid the foundation and conceived the basic underlying principles on which lasers are based. He formulated and explained the relations that govern the way that atoms and molecules absorb and emit radiation. He introduced the key concept of stimulated emission. Then, in the 1920s, physicists C.H. Füchtbauer and Rudolph Ladenburg added formulations that used the Einstein theory to go further, and tie the absorption of light in a material medium to the fundamental properties of its constituent atoms. But it was the Russian physicist A.V. Fabricant who first had the vision to propose the concept of a laser in 1940.

Unlike the Sun and other incoherent light sources, the laser produces light of one precise color. Its waves are emitted in phase with each other, and its radiation can be efficiently focused to a precise spot. In short, the laser produces coherent light, with the properties we were already used to with radio, television and microwave sources.

Fabricant specified the conditions needed for amplification of light via stimulated emission. He appreciated the concept of an inverted population and the concept of coupling such an inverted population medium to a resonant structure (resonator). He proposed using a gaseous electrical discharge to achieve laser action as one possibility. Later, he proposed the use of a helium discharge lamp to optically pump the gaseous form of cesium.

Scientists Purcell and Pound were the first to report evidence of net stimulated emission in 1951. Their observation, in the radio part of the spectrum, was a by-product of their pioneering developments in the techniques of nuclear induction, the foundation for magnetic resonance imaging.

The first achievement of a working device that utilized stimulated emission as its operating principle was a microwave oscillator/amplifier. That device used a beam of ammonia molecules as the working medium. The design details were worked out independently by groups at the Lebedev Institute in the Soviet Union and Columbia University in the USA. The device was dubbed a *maser*, which is an acronym for *microwave amplification by stimulated emission of radiation*.

In 1955, the year after I was awarded my Ph.D., I joined the Hughes Research Laboratory in Culver City, California. One of my major assignments was to head a project developing a miniature, liquid-nitrogen ruby maser for the US Army Signal Corps.

Race to the Light

Before starting work on the maser I had begun to germinate some ideas about the possibility of a laser. My concept so far was to try to use a solid material for the lasing medium, and to fabricate the potential laser material into a rod shape. In microwave thinking this configuration would be called a *dielectric wave-guide*. In optics it is known as a *light pipe*.

Abridged from "The Laser Odyssey" by Theodore H. Maiman. Copyright © 2000 Theodore H. Maiman. Reproduced with the permission of the author and Laser Press, Blaine, WA, USA.

I planned to put mirrors on each end of the rod to form a resonator.

Despite the advances in physics it represents, the maser does not in any way extend the coherent electromagnetic spectrum. Furthermore, its use as an amplifier turned out to be impractical because cryogenic temperatures were required for proper operation. There was strong motivation to push the frontiers of coherent electromagnetic radiation to higher frequencies, principally the promise of higher energies and greater concentrations. The reduction in wavelength from microwave to visible is a factor of about 10^4. This would lead to a consequent 10^4-fold reduction in the minimum achievable focal size, and a 10^8-fold increase in energy concentration delivered by the laser.

Work on the maser project therefore proved something of a distraction, and it was not until August 1959 that I was able to devote full effort to the consideration and analysis of my concepts for a laser.

Why would I be willing to enter such a race? There was rather formidable global competition already in play, well-funded and very competently manned. The answer lies in my knowledge of the proposals that had been floated about. Generally, the scientists who pursued these notions were not particularly secretive about what they were working on. They published and gave talks at conferences. On the whole, however, I found that the authors and conference presenters were only offering vague proposals. To be sure, this served as an exchange of information and stimulated ideas. But it didn't appear to me that anyone was close to the answer. In short, no viable laser concept was yet in existence.

As I look back, I was a little brash. I would be thrusting myself, in a sense, into a technological Olympics. The competition was of the best quality and of international scope. But my competitive spirit won out. The challenge of working in the top league of such an exciting project, that had so many questions and problems to resolve, was very compelling to me.

We know *now* that many kinds of lasers can be made. But back then, in 1959, we didn't know. We didn't even know with any confidence if it was really possible to make a laser at all. My strategy was to limit myself to potential solutions to the making of a laser that did not have appreciable distractions in the design. That way I could focus strictly on just the laser problem itself.

For various reasons I was reluctant to work on alkali-vapor or gas discharge systems. I chose to work with solid-state crystals. The main appeal that solids held for me was simplicity. By that, I mean simple in analysis and understanding, and simple in device conception.

In contrast to a gas discharge, the energy level diagram in an appropriate crystal is very limited. There are relatively few possibilities for the energy states, and by and large, the pertinent parameters for a potential laser candidate are amenable to a combination of calculations and relatively straightforward direct measurements.

Another advantage, in principle, to a solid crystal is its relatively high gain coefficient. By that, I mean the amplification in a given length of material is of reasonable proportions. This meant that the laser medium could be relatively small in size and short in extent, and I would not have the problem of developing or depending on the use of special mirrors. Indeed, my first laser used a crystal that was only 2 cm long.

I was also intrigued with the concept of a solid medium since I would not have to deal with vacuum pumps, impurity problems and gas handling apparatus, or complex mirror mechanisms. I could put simple silver mirror coatings directly on the crystal as I had done with my small ruby maser. In principle, a solid crystal laser could be designed to be very simple, compact and rugged.

My first choice to study and contemplate was that of a ruby crystal. Ruby is the result when a water-clear cast of aluminum oxide is doped with a small amount of chromium oxide impurity. It is the chromium that is responsible for the red color. The rubies used for devices are usually not natural gemstones, which have a chromium impurity level of about 0.5%. The chromium concentration in device rubies is around ten times less than this, and hence they display a lighter red color than gemstone ruby and are referred to as *pink ruby*.

Why choose ruby as a potential laser candidate? Most importantly, I was quite familiar with and fascinated by the interesting optical properties of the crystal. Ruby is a fluorescent mineral; if ultraviolet light is shone on a ruby, it will glow with deep red fluorescence. Furthermore, ruby also gives off a red glow when either blue or green light is shone upon it. It is these blue and green absorption bands that give ruby its red color. When a green photon impinges on and is absorbed by the ruby, a chromium impurity ion is raised from its ground state into a broad, excited band. Although the chromium ion has the possibility of radiating by spontaneous emission from that excited level, another process comes into play. The competing process uses the thermal vibrations of the crystal lattice to interact with the excited ion and deposit most of the excitation energy to another slightly lower excited level of the ruby chromium

where it stays for a while. This latter process is much more probable and dominates.

The energy from this third level is radiated as spontaneous incoherent emission. This spontaneous emission is made up of red photons and is the observed fluorescence. The level where the fluorescence emanates is sometimes called a metastable level, since the chromium ions linger in that energy state for a comparatively long time before they radiate red photons.

I developed a model that could be mathematically analyzed and I set up kinetic equations to take into account the various mechanisms taking place in the fluorescent process. I also set up simple intuitive criteria for establishing the condition for laser action. This model and these equations have subsequently become a standard way for others to analyze crystal lasers.

I was able to determine which material parameters were important and relevant to a laser by getting the solution to the equations describing the model. Using known and estimated values for the pertinent parameters in ruby, I found that ruby would require a very bright pump lamp to excite the crystal sufficiently to allow laser action to take place. The brightness of a lamp is an important parameter for lasers. It is a measure not of the total power radiated by the lamp, but rather the power per unit area.

Obstacles and Solutions

Ruby has many desirable qualities. It is a very stable, very rugged crystal to work with. It has some broad pumping bands that make it helpful to get a reasonable energy transfer from an incoherent pumping lamp, and compared to some other crystals I considered (e.g., gadolinium-doped crystals), a working ruby laser would produce visible light.

However, there was some uncertainty over the crystal's measured quantum efficiency. The quantum efficiency for a fluorescent crystal is the number of fluorescent photons emitted, divided by the number of pumping photons. In the case of ruby, this is the number of red photons radiated out compared to the number of green photons absorbed. A report in November 1959 had suggested that ruby's quantum efficiency was only 1%. This would effectively mean that it would be impossible for ruby to achieve laser action. However, I was able to make my own measurements of ruby's fluorescent efficiency, and these showed that it was nearer 75%.

I now had enough information that I could proceed to work out an actual laser design. But how would I pump the ruby? What would be the ruby's shape and size?

I knew that I needed a very bright lamp. Among the brightest around was a high-pressure mercury arc lamp manufactured by General Electric. In addition to being one of the brightest available laboratory lamps, it had the advantage of radiating most of its energy in the green and blue-violet parts of the spectrum. This was a good match to what I needed for the ruby.

I drew up a paper design with the mercury arc lamp at the focus of an elliptic cylinder. A small ruby rod was to be mounted on the other focus of the elliptic cylinder. It is a property of an elliptically shaped reflector that a point of light placed at one focus of the ellipse will be imaged at the other focus. The elliptic cylinder that I had in mind would be highly polished on its inner surface to a mirror finish.

But I had a problem. My paper analysis of the design showed that although it should work, it would only do so by a slim margin. I studied the design in more detail, looking for ways to optimize and improve it, but I couldn't convince myself that it was anything but marginal.

I stewed over the problem at hand. I started to think about other ways to look at the problem. I was frustrated: I felt that, on the one hand I was very close to an answer, but on the other hand it was eluding me.

I went back to my analytical model. I pondered my options and decided to put the pumping lamp requirement for the ruby in a different form. I calculated the equivalent black-body temperature of a suitable pumping lamp. The temperature I calculated to have enough brightness capable of driving a ruby into laser action was close to 5000 K. By contrast an ordinary tungsten light bulb has a brightness temperature of about 2800 K.

Once I put the pumping requirement in terms of brightness temperature, I began to think in a different way. I remembered reading an article about photographic strobe lamps, a camera's flash mechanism. The article said that strobe lamps could reach brightness temperatures of 8000 K or more.

I now had my 'aha'!

Most scientists had been thinking in terms of a continuous laser, and that was certainly my thinking to start with. But why should I place such restrictions on myself? At this juncture I was only trying to demonstrate that coherent light could be made at all. Besides, a pulsed source is not only acceptable in many applications, it may even be preferable.

I went back to my analytical model, modified the equations to account for a pulsed light source, and then analyzed the results to guide me in the actual laser design. I had already determined that the most important parameter of the light source would be its

brightness, that is, a lamp that maximized the power per unit area into the useful pump bands of the ruby.

My calculations also instructed me that this lamp brightness requirement is largely independent of the ruby crystal dimensions and its chromium concentration over a reasonably wide range. The model assumption was that of a three-level system (appropriate to ruby) where the lower laser-energy level is the ground state.

Next, I scoured every flashlamp catalogue that I could find. I calculated the luminous intensity for promising candidates and found that the highest intensity came from three General Electric spiral-shaped flashlamps. My calculations showed that I would have an adequate margin of safety, in terms of excess brightness intensity for the laser, of some two to three times, even with the smallest lamp.

Let there be 'Coherent' Light

Based on my most current calculations and measurements, I was becoming optimistic about the possibility of creating a laser. The tension was building and I started to have dreams that I actually did it. It was a wildly exciting thought!

Still, I was worried that I might be missing something. Instead of proceeding directly to the actual design, I decided to do an experiment that would check the validity of my model.

I made up a cube of pink ruby with each dimension equal to one centimeter in length, about the size and shape of a sugar cube. The crystal axis was perpendicular to one set of faces of the cube. I then placed the cube between two parallel plates, thus forming a microwave cavity. The cavity resonance was designed to be equal to the natural ground-state splitting of the ruby. These are the ground-state sublevels used in a ruby maser.

The purpose of this arrangement and the subsequent experiment was to monitor the ground state population of the ruby. My calculations had shown that it would be possible to substantially reduce the ruby ground state ion population necessary to make a laser, but I wanted tangible evidence to confirm and justify my optimism.

I was still harassed by the thought that no one had yet ever made a laser, although by now several teams of scientists had been diligently working away for nearly two years. Was there a fundamental insurmountable problem that I had missed?

I proceeded with the experiment using the ruby cube. As described above, one set of parallel faces of the cube served as the resonant microwave cavity. I connected one of the open (second set of) faces of the cube to one end of a polished quartz rod called a light pipe. The other end of the light pipe was connected to the flashlamp. The third set of faces of the cube was used to probe with selected wavelengths of light. I used the 'loading' on the parallel-plate microwave cavity to monitor the ground state population of the ruby, making use of the microwave properties of ruby that I had learned from my work on the ruby maser.

When I flashed the lamp, the cavity loading did indeed change. It decreased because, as expected, there were fewer chromium ions in the ground state. More important, the *magnitude* of the population change, 3%, was very close to the value I had predicted from my analysis of the experimental parameters. This was an extremely gratifying result since it gave further confirmation to my model and its analysis.

Now I was getting more excited. Since I couldn't think of anything else to check, it was time to proceed with the definitive design of a laser.

The Laser Design

My first thought was to use some modification of the elliptic-cylinder configuration that I considered for the mercury arc lamp as discussed in the previous section. The problem was, I would need a straight, rod-shaped flashlamp, but the straight lamps in the catalogues that I checked didn't have high enough intensity. Therefore, I resorted to a different design and stuck with what was available – the spiral-shaped lamps. I didn't want to get sidetracked at this time into the development of a special lamp.

It dawned on me that I would not have to devise a focusing arrangement, since the brightness of the radiation that is attainable at the focus of a mirror or lens can only approach, but not exceed, the brightness of the source. So, why not just place the crystal in close proximity to the lamp? That is, why not place the crystal *inside* the lamp helix?

To help gather the light, I placed a polished aluminum cylinder around the outside of the spiral. To fit the lamp, I used a pink ruby crystal in the shape of a right circular cylinder rod, about 1×2 cm. The ends of the ruby cylinder were polished flat, parallel to each other and perpendicular to the axis of the cylinder. For mirrors, I used evaporated silver.

I chose silver for the mirrors because it has the highest reflectivity of any metal at the deep red ruby laser wavelength. However, silver tarnishes quickly, and so with time, the thin output layer will change its transmission characteristics; it is not stable. To solve this problem, I had a thick layer of silver evaporated on both ends of the crystal and I scraped a tiny hole in the coating at one end. The laser beam would emanate from that coupling hole.

The Crucial Test

As I neared completion of my laser design, Bob Hellwarth, one of my colleagues in the Atomic Physics Department, asked, 'How will you know if it's working?'

At first, I was concerned that if I were stuck with a ruby crystal that deviated too much from optical perfection, these imperfections could prevent measurable laser behavior.

To get a better understanding of the processes, I elaborated on my existing analysis to account for an imperfect crystal. I found that if I were able to drive the ruby crystal reasonably far above the point of inverted population, or laser threshold, that I would see very substantial evidence of an inverted population and impressive laser behavior even with a poor crystal.

Just exactly what would I see?

I planned to monitor the red light emitted from the ruby through the hole in the output silver coating and expected to observe three kinds of laser behavior evidence: a reduction in the decay time of the fluorescent level; the bunching of the radiation into a beam; and a significant reduction of the spectral width of the red light.

The Laser Process

Laser action can be understood by following through the operational details in the above-mentioned ruby laser design. The process starts when the ruby crystal is excited by the flashlamp and chromium ions are excited into the metastable fluorescent level. The ions lose energy by radiating randomly red photons. This is the familiar red fluorescence (spontaneous radiation).

When the excitation level is great enough, that revered 'inverted population' condition is reached. In that case, more chromium ions are in the ruby's metastable upper fluorescent level than are in the ground state. Therefore, chromium ions can radiate by *stimulated* emission as well as by the normal spontaneous emission process.

Because the metastable level now empties much faster than it normally would, the fluorescent lifetime is reduced. That decreased lifetime can be observed by monitoring the red ruby glow with a photoelectric cell connected to an oscilloscope when the ruby is flashlamp-excited.

When the ruby is in the inverted population condition, as explained previously, it becomes an amplifier. The red photons are amplified as they progress through the crystal.

An important selection process starts to take place. The red photons are initially emitted in random directions. But the fluorescent photons that happen to be radiated at large angles to the mechanical axis of the ruby cylinder are lost through the sides of the crystal. On the other hand, photons radiated along the crystal axis, or at small angles with respect to the crystal axis, are in effect, trapped. They are reflected when they strike either one of the end mirrors and move back and forth through the crystal. As these axial photons move through the ruby they are amplified and consequently generate more photons in the same direction. They quickly become the dominant stimulated radiation from the excited metastable level.

Keep in mind that photons moving through the crystal are responsible for the stimulated emission. The axial photons pick up a following and march down that crystal axis. Consequently, the photons that emerge from the coupling hole in the output silver mirror are concentrated in a direction along, or nearly along, the crystal axis.

The red ruby fluorescent (spontaneous) emission extends over a distribution of frequencies in a curve that resembles a bell shape. The fluorescent photons are most concentrated at the center of that curve. When the inverted population condition exists, the top of the curve has the highest amplification. The consequence is (as with the discussion of beam angle) that photons near the center of the distribution are favored, since it is these photons that get amplified most in the back and forth transits, through the crystal from the multiple mirror reflections. It is this last process that explains why the frequency distribution, the linewidth of a laser, is so small.

I planned to vary the excitation to the flashlamp. In doing so, as explained above, I would expect to see a reduction of the fluorescent lifetime giving evidence of stimulated emission. As I varied the ruby excitation level, the fluorescent intensity should increase proportionately, as long as I was below threshold. But, when above threshold, small increases in excitation should make much bigger changes in the output since the detector and monochromator combination are more responsive to the smaller beam and narrower linewidth.

Do it!

It was the afternoon of May 16, 1960; it was time to confirm or deny all the fears of why 'ruby can't work' or why 'lasers can't be made to work.' No more new calculations, no more diversionary experiments. This was the moment of truth!

The laser head was mounted on a workbench. The flashlamp was connected to the power supply. The trigger electrode was connected to the spark

coil (the mechanism that initiates the flash from the strobe lamp). The light output from the coupling hole in the end of the ruby was directed through the monochromator to a photomultiplier tube. The electrical signal from the photomultiplier was connected to an oscilloscope.

Irnee D'Haenens, my laboratory assistant, and I were the only ones performing and observing the experiment. We first took a test shot so that we could adjust the monitoring equipment. We turned the power supply up to about 500 volts. We fired the flashtube. Indeed, we observed a trace on the oscilloscope!

The trace was a recording of the red ruby fluorescence. The decay in the trace was about three milliseconds, the lifetime of the upper possible laser level. We made the appropriate adjustments to optimize the monitor display.

We continued. We progressively increased the supply voltage, each time monitoring and recording the light output trace. As we did so, the peak output increased proportionately to the energy input, the decay time remained the same. So far, so good.

But, when we got past 950 volts on the power supply, everything changed! The output trace started to shoot up in peak intensity and the initial decay time rapidly decreased.

Voilà. This was it! The laser was born!

The Light Fantastic

When Irnee and I observed the first laser go into action, Irnee was smiling and jumping up and down with glee. I was numb and emotionally drained from all the tension and excitement.

Imagine: in the 10 years prior to the laser, the coherent electromagnetic spectrum had been extended by perhaps a factor of five. Now, with the advent of the laser, there was a quantum jump in that spectrum of ten thousand!

The significance of my historic accomplishment didn't sink in right away. I'm not sure that it has yet.

THE HELIUM–NEON LASER

A Javan, Massachusetts Institute of Technology, Cambridge, MA, USA

© 2005, A Javan. Published by Elsevier Ltd. All rights reserved.

The date, twelve of December nineteen sixty, and the time, 4:20 PM, are the date and the time of an event that triggered such an imortant segment of what has now evolved in modern optics. Here is how:

It had snowed heavily on that day. The laboratory at Murray Hill, New Jersey, the Bell Telephone Laboratory, BTL, had been closed at 4. What occurred was a triumph of our depth of understanding of the atomic spectra, the physics of the inter-atom collisions and electron impacts and exchange effects, the fascinating sequence of what takes place in a *gas discharge plasma* when the discharge is switched on. The triumph was also that of the art in spectroscopy, this time that of the colorful glow of a gas discharge plasma in a controlled mixture of helium and neon. Performing the spectroscopy, one could *effectively* watch and follow a set of fascinating physics effects in the discharge plasma, that would lead to an *inversion* to set-in in a specific transition between a high-lying pair of energy levels in neon, by the now well-known transition at 1.15 μm wavelength in neon. Helium atoms in a metastable state in the discharge plasma played the key role.

On that day: The laser had already been de-gassed for 5 days and nights at 600 °C under the high vacuum, to drive off the impurities prior to introducing the gas mixture in the laser-plasma quartz tubing and sealing it off. We had a way to optimize the mixture, at the exact mixture ratio for the optimum inversion and the laser gain at the 1.15 μm neon transition wavelength. The laser design had a *daring* interferometer length *one meter long*. Why *daring* at *one meter long*, I will have a chance to explain why below. The laser-mirrors at each end of the one-meter-long interferometer were situated internally in the laser, exposed to the gas mixture. Two fine micrometers at each end of the laser provided the fine control for the laser–mirror alignment control externally.

To convey the scene, we had known already, that the inversion and the laser gain at the 1.15 μm neon transition will be *higher* in the discharge '*afterglow*', the afterglow following *rapidly switching-off the discharge*. The inversion and the laser gain at the 1.15 μm wavelength, we knew, would last for as long as a millisecond in the afterglow. Certain that the inversion and the gain will be higher in the afterglow, we had set-up the electronics to switch on the discharge for about 10 milliseconds, and switching it off fast at a one microsecond switch-off time, and leaving it off for the same 10 milliseconds time interval, and repeating it at a low rep-rate, about 20 times a second on that day. Expecting that while searching for the laser signal, the laser will first break into oscillation in the afterglow where the laser-gain will be higher, knowing that pulling it into the glow afterward will be easy.

With me on that day I had two of the three good people who worked with me on the experiment. One of the two, Donald Herriott, was about to miss his car pool. He had his hand at the time on the laser–mirror tuning-alignment, the fine micrometer at one end of the laser interferometer on his side. I had my back towards the oscilloscope that would display the laser signal, aligning the detection system. My journey had begun over two years previously. I was *certain* on that day, and at that time, that possibly right then, or no later than hours later or by midnight, the moment would come for me and others that I had with me, to witness the onset of the event sought after so hard. And then … .

I had my back towards Don Herriott, urging him to stay a bit longer. I heard Don's voice, saying '*What is that*?'. I turned and looked. The laser signals in the discharge afterglow on the oscilloscope screen, synchronized to the gas discharge rep rate. I reached the laser interferometer fine-tuning micrometer on my end … a gentle tuning, in a mere few seconds time, optimized the afterglow signal, and then pulled the laser signal right into the glow. I loudly called on 'Ed' to switch on to CW. He, Edward Ballick, jumped over the equipment and the electronics on the floor in the lab, reached the switching electronics, switched it on to CW:

> The laser signal on the oscilloscope … switched to CW … a coherent light beam … CW … continuous … and at a degree of coherence … that as it proved later and I know then … to the limits that the laws of nature will permit. I looked at my watch, 4:20 PM.

The year 1960 was the beginning. Some months earlier, in the summer of that year, Theodore Maiman at Hughes Laboratories, in Malibu, CA, succeeded to operate an optically pumped laser, operating in short-duration laser pulses, the Ruby laser at 0.694 μm. That approach to the laser differed fundamentally

from the discharge plasma approach. The two were different inventions and served different purposes. One, the *optically pumped*, converts the *optical energy* in the pump, to the laser light output. The other, *electrically pumped,* converts the *electrical energy* in the *electrical current* that excites the laser plasma, to the laser light output. The rest on how things evolved is history.

The Scene Before Lasers

Before lasers, optics had long passed its peaks. The last words had been written by Born and Wolf in 1957, in an impressive volume on Physical Optics. When I went through my graduate studies at Columbia University in the early fifties, there were no graduate courses offered in optics. The subject was dealt with incidentally in courses in electricity and magnetism. Action was all on the microwaves at the time. Optical spectroscopy had reached its peaks much earlier yet, before World War II. Condon and Shortly had written the last words on Theory of Atomic Spectra in 1935, amazingly, as it seems, in *hardly a decade* after the discoveries in quantum mechanics. As events had it, optics gave way to the microwaves after the war, and optical spectroscopy to the RF and microwave spectroscopy – and all this, a result of the discoveries in Radar at the microwave frequencies in World War II.

We owed everything we knew of the atomic spectra to the *gas discharges* used as spectroscopic light sources at the time, from the high-resolution spectra of the hydrogen atoms at one end of the periodic table, to that of the trans-uranium elements on the other. We need not be reminded, everyone knows, how much we owed then and do now, to the high-resolution spectra of the hydrogen atoms, all because of the early hydrogen discharge light sources and then at the beginning, what led to the discovery of the quantum mechanics, and the wave nature of the atoms. The excitement that followed has remained unparalleled in physics since. Within hardly a decade after as a result, the cleverly designed spectroscopic gas-discharge sources, of different makes and designs, exhaustively revealed every feature predicted by the quantum mechanics in the high-resolution spectra of every element we know in nature. Gas discharge light sources set the course at the time, as did the lasers after the events in 1960.

Optics had given way to the microwaves in the fifties after the war. The physics of the gas discharges, however, continued as an active research field after the war, although of short duration. Bell Telephone Laboratory had a most distinguished research department in gas discharges in the early fifties. Towards the late fifties, as it seemed, the gas discharges as a field in research, also seemed to have been exhausted. The payoff at the time had evolved in the colorful neon light sources and fluorescence lamps, a thriving industry all over the country.

In the summer of 1958, Bell Telephone Laboratory had dissolved their research department in gas discharges. They had transferred every one of their distinguished research people to the other research areas in the laboratory. Some good people had left earlier, already.

As the fate in gas discharge physics had it, I joined the Laboratory in September 1958 – with ZERO backgrounds in gas discharge physics. I knew physics.

How Things Evolved

It has been said that when time is ripe for a new idea, if one person misses it, the next one will not. Time was ripe in the late fifties for the optically pumped idea for the laser, but far from it for the *electrically pumped*. Optical pumping was an active field in research at the time. The French had introduced the 'pompage optic', the optical pumping. With the French pompage optic, one would by optical pumping, resonantly excite an atomic species from its ground into its excited electronic state, and perform a double-resonance at an RF frequency on a fine or hyperfine structure of the excited electronic state. There were other versions of the French pompage optic. Each version used the optical pumping to induce resonantly, an atomic species into an excited electronic state.

Important, and more so, in the fifties, in the second half, the discoveries of the three-level masers at the microwave frequencies had attracted much attention. The three-level masers are the counterparts at the microwaves of our now optically pumped lasers. In three-level masers, one drives to saturation a pair of energy levels at microwave energy-spacing in a paramagnetic solid, and allows the inversion and the gain to set-in in a transition to a low-lying third energy level at the microwave frequencies. Nicholas Blumbergen at Harvard had proposed the three-level maser idea in an important publication in 1956. I had my own original three-level maser ideas, and had them published in 1957. In my work I had discovered a fascinating new effect, and for the first time, that in the transition to a third level in a three-level maser, it will be possible to drive the maser into its self-oscillation, *without* requiring the inversion. The effect had led me to the idea of the Raman masers, requiring a '*pumping*' microwave field, followed by a Raman transition across a pair of energy levels at microwave energy-spacing. In my publication of the

effect, I had also considered the pumping field at an optical frequency, instead of at the microwave.

The ideas were 'in the air' at the time, to extend the three-level maser approach into the optical frequencies. Yet, in the summer of 1958, on an occasion in June of that year, I think on the first Monday in June, *flashed* through my mind a far different approach, all because of the *helium metastables*. Ideas always come to mind in a *flash*, the rest follows.

Summer of 1958: A Bit of History

I think it was on the first Monday in June 1958. I had been on an occasion visiting the Bell Telephone Laboratory (BTL) on that day. They had approached me that I join their research staff at Bell. My work on masers – I had done as a part of my post-doctoral research position in physics at Columbia University in New York City – had in part overlapped the then ongoing work at Bell. They knew me, and I knew Bell.

At the visit meeting with Arthur Schawlow, whom I knew well, he gave me his news on that day, that he and Charles Townes, had submitted to the *Physical Review* a paper for publication proposing an optically pumped counterpart of a three-level maser at the optical frequencies – by-now, the well-known Schawlow–Townes optically pumped laser publication. Specially, I was attracted to Schawlow's idea on a two-mirror Perot–Fabry-type interferometer as a resonator at the optical frequencies for the laser. The use of it in that publication was Schawlow's idea.

My thoughts, it may seem paradoxical, right there and then moved far away from an optically pumped laser as a possibility. The helium atoms in a *metastable state* flashed through my mind. I will avoid explaining why or how, other than, of a three-level atom, two of the levels were now to belong to a *different* atomic species, the helium atom, and that there will be no optical pumping … and so much '*electrical energy*' could be converted to and stored in the energetic metastable helium. A new possibility presented itself, with no resemblance to the optical pumping – an energy-exchange effect taking place resonantly, between the highly energetic helium metastables, and – as it evolved – another inert gas, the neon atoms in a helium–neon gas mixture. The resonant nature of the effect embodied everything we know of the wave nature of the atoms. I will have a chance to explain the effect shortly below.

Driving home at the end of the day on that day, a two hour drive from Murray Hill, New Jersey to uptown New York City, the thought of the metastable helium had crystallized before me the task to examine the effect in a helium-plasma in a gas discharge, in a gas mixture with the other *four* well-behaved *inert gases*, from neon through xenon. There are four.

Within weeks in June, searching through the literature I learned, that over the years, and mostly in the early fifties, the helium plasmas in a gas discharge in mixtures with *every other* inert gases had been the subject of most extensive research, *except in a mixture with neon*. The mixture with neon, for the reason that the reader will find below, had been thought of as *uninteresting*. The matter was left for me.

I had accepted in June already, to join the Bell research staff at Murray-Hill, and I did on September. From my work that entire summer, unfolded the entire theory for the helium–neon gas laser. There was also another far-reaching offshoot.

The Physics Effect

Next to the hydrogen atom, the helium is the most fundamental of all atoms in physics. Hydrogen has one electron. Its spectrum, detached from the complexities of the multi-electron atoms, carries the entire signature of the wave nature of the atom, in ways that are subject to accurate predictions and the measurements. The helium atom has two electrons. Its spectrum carries the entire signature of the physics effect of the entire multi-electron atoms, also in ways that are subject to accurate predictions and the measurements. In its metastable states, it carries special *dynamic features*, deeply fundamental, and special *only* to the helium metastables.

In a gas discharge in pure helium at a moderate gas pressure, the discharge color takes on a grayish-white with a tint of blue. Over 500 excited electronic states in helium have been identified – their excitation by the energetic electrons in the discharge, followed by their spontaneous radiative decays, the emissions of photons, accounts for the high-resolution spectra of the discharge grayish-white with a tint of blue in color. Of these 500, all of them, high lying in energy, are short-lived radiatively, with the exception, however, of two amongst them. The two have the lowest excitation energy. Their radiative decays to the ground electronic state are deeply forbidden; they cannot decay radiatively. The two form the well-known helium metastables, identified by their spectroscopic designations: $^3S_{1/2}$, the triplet S one half, and $^1S_0$, the singlet S zero – generally referred to in the literature as, the triplet S and the singlet S metastables. Of the two, the triplet S metastable has the lowest excitation energy. It lies at 19.8 electron volts (eV), above the helium ground state. The singlet S metastable lies higher, at 20.2 eV.

Helium atoms in their ground electronic state are chemically inert. A helium atom's two electrons form a closed shell. It cannot enter in a chemical reaction, neither with its own species, nor with any other, as we all know. The situation is, however, far different with every one of the 500 excited electronic states in a helium gas discharge. They are energetically high-reactive, except, and with the exception of the two metastables, they are all *short-lived*, and cannot play an important role.

The effect manifests itself differently in the long-lived metastables, and accounts for their *special dynamic features,* special only to the helium metastables. What follows outlines the effect.

Dynamic Features: Why so Special

By far the majority of the atoms in the helium gas discharge are in their inert, ground electronic states. Once in its energetic metastable state, a helium metastable in the gas discharge encounters in an impact – in a collision – with a helium atom in its ground electronic state. The encounter in the impact is of a short duration in our scale of time, but long in the atomic time-scale, of about 2×10^{-13} second, a fifth of a picosecond. What occurs is strictly wave-mechanical, dictated by the wave nature of the electrons in the atoms. During the encounter, the electron wave functions in the encountering atoms overlap momentarily. A momentary binding occurs. In no ways can one tell during the encounter which electron is which: Electrons are the properties of their wave functions – once overlapped, individual electrons lose their identity, in no way can one tell which is which. As the atoms fly apart, caused by the kinematics of the impact, an *electron exchange* occurs, *wave-mechanically.* The helium atom initially in its ground electronic state, will fly away from the encounter with one of its two electrons now in the *highly excited* energetic *metastable* state. The other, the one initially in the excited metastable state, emerges from the encounter in the *unexcited* ground electronic state.

At a gas pressure in the discharge of about one torr, a millimeter, the mean free-time between two impacts is about one microsecond, a very long time in the scale of the atoms. The helium atom that flew away from the encounter in the metastable state remains in its now metastable state, during the entire mean-free collision time. Once it encounters the next helium atom in its ground electronic state, the act repeats again, a momentary binding occurs, an electron exchange takes place, and a new metastable flying away. The effect repeats over and over again, at about every microsecond. A *diffusion* of the metastability through the helium in the discharge sets in. Depending on the diameter of the discharge tubing, it takes about a millisecond until the helium that emerges from the encounter in the metastable state reaches the walls of the discharge tubing. The act will then *terminate*, in a *most violent* collision effect with the wall. The energetic metastable helium *knocks out* an electron off the wall, as itself emerges from the collision in its ground electronic state – all intact.

With one exception, the encounter between a helium metastable with any species other than of its own kind, a helium, is energetically *highly violent*. The exception is the neon atoms. In encounters with all atomic species other than neon, the metastable helium *knocks out* an electron off the species, ionizes the species, and itself emerges from the encounter in its ground electronic state, intact. The ionizing impacts had attracted so much attentions in the early fifties, the effect in the helium gas discharge in mixtures specially with every other inert gases, had been exhaustively investigated, except with neon. Neon has its ionization potential higher than the two metastable excitation energies. It can not undergo an ionizing impact with either of the two metastables. The mixture with neon had thus been considered uninteresting, and was discarded.

What emerged by the end of the summer of 1958 was the entire theory for the helium neon laser – and the recognition of a deeply wave-mechanical effect, the resonant energy exchange taking place between the energetic helium metastable and an inert neon atom in the discharge gas mixture. In the mixture, a momentary binding occurs between the helium metastable encountering an inert neon atom in the gas discharge. Wave-function overlaps occur for the electrons in the encountering atoms. A *resonant* energy exchange takes place during the encounter. Helium metastable will fly away from the impact in its now ground electronic state, while neon emerges in one of its high-lying excited energy states, at the near-resonant with the helium metastable.

A special effect sets in in the energy exchange, as well. The kinetics of the impact in the encounter enters in satisfying the exact resonance in the energy exchange. The singlet S metastable flying away from the encounter after the energy exchange, will be slightly 'cooler' kinetically, and the triplet S flying away from the encounter after the energy-exchange, will be slightly 'hotter'. The kinetic energy of the impact enters in the energy exchange.

Optical Gain and Amplification Effect

Within weeks, early in the summer of 1958, investigating the energy transfer effect led to the

prediction that there can be inversion and optical gain in two sets of transitions in neon, one as a result of the excitation transfer from the triplet S metastables, and the other a result of the transfer from the singlet S. By far the more challenging task that summer, however, was to estimate theoretically the *values* for the *gain and amplification factors* obtainable for the predicted inverted transitions, and more specifically, for the inverted transitions that will have the highest gain and amplification factors. There followed the theoretical estimates on which relied the entire experimental phase of the work, the planning of it commenced within weeks after my joining the Bell Telephone Laboratory on September of that year.

The set of transitions originating from the triplet S energy transfer to neon, occur in the near-red region of the spectrum, the 1 μm wavelength region. The strongest inverted transition at the highest predicted gain value, lies at 1.15 μm transition wavelength, the spectral laser line that first broke into self-oscillation (see the opening page in this article). The set of transitions originating from the singlet S energy transfer to neon occur in the red-yellow region of the spectrum. The strongest, although at less gain-value than 1.15 μm, occurs at the 0.633 μm in the red.

In the early phase of the investigation, it came to light also that the achievement of the inverted transitions with optical gains, was, interestingly, a *commune occurrence* in gaseous discharges in a variety of the atomic species, including in pure neon. However, the complexities in the excitations by the hot electrons of the high-lying energy levels, and in particular in the presence of the secondary excitation mechanism, and a list of other similar effects, in no way would allow a quantitative estimate of the gain and amplification factors in the inverted transitions obtainable in the gas discharge.

The reliance on the excitations of the high-lying energy levels in neon via the resonant energy transfer from the helium metastables, as per the physics effect I have outlined above, placed at my disposal the one gas discharge system that lent itself to the theoretical estimates of the predicted optical gains and the amplification factor for the inverted transitions.

Without a knowledge of the predicted values for the optical gain and the amplification factor, it would have been in no way possible to perform the experiment.

Experimental Phase

The Bell Telephone Laboratory, considered as a foremost research institution in the country at the time, operated at a $250 million yearly budget – in 1960s' dollars. Transistors had been invented at Bell in the forties. With the awareness at the time at Bell, that a light beam at the degree of frequency-purity and the coherence expected of the laser, will have far-reaching consequences on the telecommunication technology, without hesitation I received from Bell an open ticket to do what it takes – one key person in the administration understood in-depth my approach to the laser – via the gas discharge and the physics effects I had invoked. Within weeks following joining Bell in September, I had my planning underway for what came to be known in a short while later, as a most aggressive research in gas discharges – and at a time when Bell had already, in the summer of 1958, dismantled the entire research activities in gas discharges at Bell. That speaks a great deal for Bell at the time … at the time before the AT&T break up.

In lasers, the knowledge of the full length of the amplifying medium across which the optical gain obtainable at the laser transition will suffice to drive the laser into its self-oscillation, will dictate the full length of the laser interferometer resonator. The spacing between the two high-reflectance interferometer end-mirrors will have to accommodate the full length of the amplifying medium, thus dictating the interferometer full length. With respect to the mirror reflectivity obtainable at the time, one had to be content with the values at the highest close to 98%. The highest gain transition from my theoretical estimates, that of the neon transition at the 1.15 μm wavelength in the near-red, indicated an interferometer length of no less than 50 centimeters, or possibly as long as one meter – and there lay the challenge, achieving the required degree of parallelism needed for the interferometer end-mirrors at such spacing. One could show, from simple inspection, that the parallelism needed for the mirrors had to be at values better than one part in several-hundredths of a wavelength. This value applied to the use of the flat interferometer mirrors known to me at that time.

An alignment to such a degree of accuracy could well be obtainable at the time for Perot–Fabry-type interferometer mirrors at the mirror-spacing not exceeding about eight or ten centimeters at the most – where one would rely on the observation of the interference fringes using the available spectroscopic light sources to observe the fringes, a principle well known in physical optics. At longer mirror spacing, the spectroscopic light sources lacked the degree of coherence, the color purity, needed to observe fringes for the mirror alignment. The fringes will get wiped out.

These days, using a helium–neon laser light source available at every laboratory in optics, one can align a two-mirror interferometer at tens of meters mirror-spacing, or hundreds, or even as far apart as

several miles in some experiments. Ideas are floating around now of the possibility, using gas lasers, to align mirrors placed on satellites in orbit, with the mirror on two different satellites at controlled spacing as far apart as thousands of miles – the experiment thought of is that of the detection of the gravity waves. Without a laser operating and available at the time, in no ways would it have been possible to align an interferometer at a *daring* mirror-spacing as far apart as one meter, or even half as long. One had to rely on an auto-collimator, an optical instrument that at the best would provide an alignment to within one or two wavelengths, nowhere sufficient to drive the laser into its self oscillation.

The challenge mapped the entire course for the experimental phase, foremost aimed at showing every feature of the predicted effects experimentally, from the energy transfer effect, to observing and measuring the gain and amplification effect in the predicted inverted transitions in neon. The feat was that of the art in spectroscopy. The observation of gain and amplification effect was a highlight in the course of the experiment. Up to that point, in optical spectroscopy, one always measured the spectral lines in either emission or in absorption. The feat was the first in spectroscopy, observing the spectral line of an atomic species not in absorption, but in amplification – the amplification of an external light source transmitted through the gas discharge in the experiment. The feat, yet, lay in the art in spectroscopy, throughout the experiments.

The event at 4:20 PM in December two years later, speaks for itself.

> It may seem now, as if one needed an operating helium–neon laser at the time, to align the laser mirrors for the first operating helium–neon laser. It seemed evident from the beginning, that the presence in the very first helium–neon laser, of the optical-gain at the predicted value in the predicted inverted transition, will guide achieving the alignment of its own mirrors, to the exact optical degree of the alignment accuracy needed for the self-oscillation. Driving the laser into its self-oscillation signaled the feat. The feat, however, was that of the power in spectroscopy, making it possible to predetermine the presence in the first helium–neon laser, of the optical gain, at the predicted optimum value needed for self oscillation.
>
> There were doubters and doubters, in the technical staff and the administration at Bell. Donald Herriott, William Bennett and Edward Ballick collaborated with me in the experiment.

The Impact

Within months of the publication of the work in January 1961 in the *Physical Review Letters*, walking through the long hallways at BTL where the laboratories were situated, one could see through the open doors in many laboratories down the long halways, the pink colors of long helium–neon gas discharges plasma tubing, with people in the labs searching for new laser transitions. BTL had revived the gas discharge research activities they had dismantled in the summer of 1958. Not surprisingly, the second set of new laser transitions driven into self-oscillation took place at Bell in 1962, that of the set of transitions in neon originating from the energy transfer from the helium singlet S metastable – with the most intense in the series, that of the by now well-known 0.6328 μm neon transition in the red. There followed the discoveries of the argon–ion gas laser and shortly thereafter the CO_2 gas laser, the two amongst the foremost in the many new gas lasers driven to self-oscillation at Bell.

Every gas laser discovered since the December 12 event, has enjoyed the use of a '*no-longer*' a '*daring*' '*long*' interferometer resonator, merely because one would use a He–Ne laser to *pre-align* the laser mirrors to within a small fraction of the wavelength needed, prior to searching for a new laser line in the gas discharge chosen for the search. The already aligned resonator removed the need to *predetermine* the presence of the gain and the amplification effect at the values needed for self-oscillation. Within a span of about a decade, virtually hundreds of different gases were driven to self-oscillation with this method, operating literally at thousands of different wavelengths, ranging from the far infrared, the 100 100ηn region, to the near-UV.

The helium–neon laser triggered so many of the key discoveries that followed. Within months, in the spring of 1961, one of the distinguished engineers at BTL, Rudolf Komphner, the discoverer of the backward tube oscillators at the microwave frequencies, began experimenting with the transmission through the improvised thin fiberoptic, of the output beam of a helium–neon laser at the 1.15 μm transition in the near-red, the near-IR, having his eyes on the potential someday in communication on the use of a continuously operating laser light beam at the near-IR frequencies. I recollect the very first ad that appeared on lasers nationally was a BTL ad in the spring of 1961, showing a cross-section of a beam of light, underneath it saying: '*Someday a Beam of Light This Size Will Carry Millions of Telephone Conversations Simultaneously*'. The ad speaks for itself.

In every gas laser, the active laser medium – being a gas – is highly homogeneous. As a result, with a reasonably good optics, one can obtain a nearly perfect Gaussian output beam. For this reason, the helium–neon lasers available commercially since the

mid-sixties, have served as the indispensable tools for the alignment of the complex optics in every research laboratory in lasers, or in every manufacturer of the lasers in the country, or the manufacturers of the precise optical components, or the medical uses and so on.

In the applied areas, the helium–neon laser triggered so many of the key experiments in the early days. Philips at Eindhoven can be considered as the first discoverer of a video disc recorder. The model used a helium–neon red laser, at a most impressive fidelity and color resolution, in recording and in reading for the display. I recollect, in an exhibit shown to me at a Philips Lab, one groove on the recorder was that of a horse in a prairie, and the next groove, one '*micrometer away*' on the recorder, was that of a crowd in a park – so impressive at the time, with no cross talks. They had it packaged in a compact unit already. In other areas, all bar-code readings were done using a helium–neon laser, and some still do.

In the scientific area, from high-resolution spectroscopy at unprecedented accuracy, to the accurate clocks operating at optical frequencies, or in the measurements of the absolute frequencies of the accurate laser clocks, or the early experiments in relativity and so on, all are owed to the precision gas lasers. Helium–neon laser has played the key role, in the early days throughout, and continues today.

A note to the reader

In the early publications on laser, the device was generally known and referred to at the time as the 'Optical Maser'. The word laser became an accepted designation for the device in the late 1960s and the early 1970s.

Suggested Readings

Javan A (1961) In: Singer J (ed.) *Advances in Quantum Electronics*, pp. 18. Columbia University Press.

Javan A, Bennett WR, Jr. and Herriott DR (1961) Publications reporting operation of the first helium–neon laser. *Physical Review Letters* 6: 106.

Field Guide to

Microscopy

Tomasz S. Tkaczyk

Field Guide to

Microscopy

Tomasz S. Tkaczyk

SPIE Field Guides
Volume FG13

John E. Greivenkamp, Series Editor

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Tkaczyk, Tomasz S.
Field guide to microscopy / Tomasz S. Tkaczyk.
p. cm. -- (The field guide series)
Includes bibliographical references and index.
ISBN 978-0-8194-7246-5
1. Microscopy--Handbooks, manuals, etc. I. Title.
QH205.2.T53 2009
502.8'2--dc22

2009049648

Published by

SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: spie@spie.org
Web: http://spie.org

Copyright © 2010 The Society of Photo-Optical Instrumentation Engineers (SPIE)

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means without written permission of the publisher.

The content of this book reflects the work and thought of the author. Every effort has been made to publish reliable and accurate information herein, but the publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Printed in the United States of America.

Introduction to the Series

Welcome to the *SPIE Field Guides*—a series of publications written directly for the practicing engineer or scientist. Many textbooks and professional reference books cover optical principles and techniques in depth. The aim of the *SPIE Field Guides* is to distill this information, providing readers with a handy desk or briefcase reference that provides basic, essential information about optical principles, techniques, or phenomena, including definitions and descriptions, key equations, illustrations, application examples, design considerations, and additional resources. A significant effort will be made to provide a consistent notation and style between volumes in the series.

Each *SPIE Field Guide* addresses a major field of optical science and technology. The concept of these *Field Guides* is a format-intensive presentation based on figures and equations supplemented by concise explanations. In most cases, this modular approach places a single topic on a page, and provides full coverage of that topic on that page. Highlights, insights, and rules of thumb are displayed in sidebars to the main text. The appendices at the end of each *Field Guide* provide additional information such as related material outside the main scope of the volume, key mathematical relationships, and alternative methods. While complete in their coverage, the concise presentation may not be appropriate for those new to the field.

The *SPIE Field Guides* are intended to be living documents. The modular page-based presentation format allows them to be easily updated and expanded. We are interested in your suggestions for new *Field Guide* topics as well as what material should be added to an individual volume to make these *Field Guides* more useful to you. Please contact us at **fieldguides@SPIE.org.**

John E. Greivenkamp, *Series Editor*
College of Optical Sciences
The University of Arizona

The Field Guide Series

Keep information at your fingertips with all of the titles in the *Field Guide* series:

Field Guide to Geometrical Optics, John E. Greivenkamp (FG01)

Field Guide to Atmospheric Optics, Larry C. Andrews (FG02)

Field Guide to Adaptive Optics, Robert K. Tyson & Benjamin W. Frazier (FG03)

Field Guide to Visual and Ophthalmic Optics, Jim Schwiegerling (FG04)

Field Guide to Polarization, Edward Collett (FG05)

Field Guide to Optical Lithography, Chris A. Mack (FG06)

Field Guide to Optical Thin Films, Ronald R. Willey (FG07)

Field Guide to Spectroscopy, David W. Ball (FG08)

Field Guide to Infrared Systems, Arnold Daniels (FG09)

Field Guide to Interferometric Optical Testing, Eric P. Goodwin & James C. Wyant (FG10)

Field Guide to Illumination, Angelo V. Arecchi; Tahar Messadi; R. John Koshel (FG11)

Field Guide to Lasers, Rüdiger Paschotta (FG12)

Field Guide to Microscopy, Tomasz Tkaczyk (FG13)

Field Guide to Laser Pulse Generation, Rüdiger Paschotta (FG14)

Field Guide to Infrared Systems, Detectors, and FPAs, Second Edition, Arnold Daniels (FG15)

Field Guide to Optical Fiber Technology, Rüdiger Paschotta (FG16)

Preface to the Field Guide to Microscopy

In the 17th century Robert Hooke developed a compound microscope, launching a wonderful journey. The impact of his invention was immediate; in the same century microscopy gave name to "cells" and imaged living bacteria. Since then microscopy has been the witness and subject of numerous scientific discoveries, serving as a constant companion in humans' quest to understand life and the world at the small end of the universe's scale.

Microscopy is one of the most exciting fields in optics, as its variety applies principles of interference, diffraction, and polarization. It persists in pushing the boundaries of imaging limits. For example, life sciences in need of nanometer resolution recently broke the diffraction limit. These new super-resolution techniques helped name microscopy the method of the year by *Nature Methods* in 2008.

Microscopy will critically change over the next few decades. Historically, microscopy was designed for visual imaging; however, enormous recent progress (in detectors, light sources, actuators, etc.) allows the easing of visual constrains, providing new opportunities. I am excited to witness microscopy's path toward both integrated, digital systems and nanoscopy.

This Field Guide has three major aims: (1) to give a brief overview of concepts used in microscopy; (2) to present major microscopy principles and implementations; and (3) to point to some recent microscopy trends. While many presented topics deserve a much broader description, the hope is that this Field Guide will be a useful reference in everyday microscopy work and a starting point for further study.

I would like to express my special thanks to my colleague here at Rice University, Mark Pierce, for his crucial advice throughout the writing process and his tremendous help in acquiring microscopy images.

This Field Guide is dedicated to my family: my wife, Dorota, and my daughters, Antonina and Karolina.

Tomasz Tkaczyk
Rice University

Table of Contents

Glossary of Symbols **xi**

Basics Concepts **1**
- Nature of Light 1
- The Spectrum of Microscopy 2
- Wave Equations 3
- Wavefront Propagation 4
- Optical Path Length (OPL) 5
- Laws of Reflection and Refraction 6
- Total Internal Reflection 7
- Evanescent Wave in Total Internal Reflection 8
- Propagation of Light in Anisotropic Media 9
- Polarization of Light and Polarization States 10
- Coherence and Monochromatic Light 11
- Interference 12
- Contrast vs Spatial and Temporal Coherence 13
- Contrast of Fringes (Polarization and Amplitude Ratio) 15
- Multiple Wave Interference 16
- Interferometers 17
- Diffraction 18
- Diffraction Grating 19
- Useful Definitions from Geometrical Optics 21
- Image Formation 22
- Magnification 23
- Stops and Rays in an Optical System 24
- Aberrations 25
- Chromatic Aberrations 26
- Spherical Aberration and Coma 27
- Astigmatism, Field Curvature, and Distortion 28
- Performance Metrics 29

Microscope Construction **31**
- The Compound Microscope 31
- The Eye 32
- Upright and Inverted Microscopes 33
- The Finite Tube Length Microscope 34

Infinity-Corrected Systems 35
Telecentricity of a Microscope 36
Magnification of a Microscope 37
Numerical Aperture 38
Resolution Limit 39
Useful Magnification 40
Depth of Field and Depth of Focus 41
Magnification and Frequency vs Depth of Field 42
Köhler Illumination 43
Alignment of Köhler Illumination 45
Critical Illumination 46
Stereo Microscopes 47
Eyepieces 48
Nomenclature and Marking of Objectives 50
Objective Designs 51
Special Objectives and Features 53
Special Lens Components 55
Cover Glass and Immersion 56
Common Light Sources for Microscopy 58
LED Light Sources 59
Filters 60
Polarizers and Polarization Prisms 61

Specialized Techniques 63
Amplitude and Phase Objects 63
The Selection of a Microscopy Technique 64
Image Comparison 65
Phase Contrast 66
Visibility in Phase Contrast 69
The Phase Contrast Microscope 70
Characteristic Features of Phase Contrast 71
Amplitude Contrast 72
Oblique Illumination 73
Modulation Contrast 74
Hoffman Contrast 75
Dark Field Microscopy 76
Optical Staining: Rheinberg Illumination 77

Optical Staining: Dispersion Staining 78
Shearing Interferometry: The Basis for DIC 79
DIC Microscope Design 80
Appearance of DIC Images 81
Reflectance DIC 82
Polarization Microscopy 83
Images Obtained with Polarization Microscopes 84
Compensators 85
Confocal Microscopy 86
Scanning Approaches 87
Images from a Confocal Microscope 89
Fluorescence 90
Configuration of a Fluorescence Microscope 91
Images from Fluorescence Microscopy 93
Properties of Fluorophores 94
Single vs Multi-Photon Excitation 95
Light Sources for Scanning Microscopy 96
Practical Considerations in LSM 97
Interference Microscopy 98
Optical Coherence Tomography/Microscopy 99
Optical Profiling Techniques 100
Optical Profilometry: System Design 101
Phase-Shifting Algorithms 102

Resolution Enhancement Techniques 103
Structured Illumination: Axial Sectioning 103
Structured Illumination: Resolution Enhancement 104
TIRF Microscopy 105
Solid Immersion 106
Stimulated Emission Depletion 107
STORM 108
4Pi Microscopy 109
The Limits of Light Microscopy 110

Other Special Techniques 111
Raman and CARS Microscopy 111
SPIM 112

Array Microscopy 113

Digital Microscopy and CCD Detectors 114
Digital Microscopy 114
Principles of CCD Operation 115
CCD Architectures 116
CCD Noise 118
Signal-to-Noise Ratio and the Digitization of CCD 119
CCD Sampling 120

Equation Summary 122

Bibliography 128

Index 133

Glossary of Symbols

$a(x, y)$, $b(x, y)$	Background and fringe amplitude
AO	Vector of light passing through amplitude Object
A_R, A_S	Amplitudes of reference and sample beams
b	Fringe period
c	Velocity of light
C	Contrast, visibility
C_{ac}	Visibility of amplitude contrast
C_{ph}	Visibility of phase contrast
$C_{ph\text{-}min}$	Minimum detectable visibility of phase contrast
d	Airy disk dimension
d	Decay distance of evanescent wave
d	Diameter of the diffractive object
d	Grating constant
d	Resolution limit
D	Diopters
D	Number of pixels in the x and y directions (N_x and N_y, respectively)
DA	Vector of light diffracted at the amplitude object
DOF	Depth of focus
DP	Vector of light diffracted at phase object
$D_{pinhole}$	Pinhole diameter
d_{xy}, d_z	Spatial and axial resolution of confocal microscope
E	Electric field
E	Energy gap
E_x and E_y	Components of electric field
F	Coefficient of finesse
F	Fluorescent emission
f	Focal length
f_C, f_F	Focal length for lines C and F
f_e	Effective focal length
FOV	Field of view
FWHM	Full width at half maximum
h	Planck's constant
h, h'	Object and image height
H	Magnetic field
I	Intensity of light
i, j	Pixel coordinates
I_o	Intensity of incident light

Glossary of Symbols (cont.)

I_{max}, I_{min}	Maximum and minimum intensity in the image
I, I_t, I_r	Irradiances of incident, transmitted, and reflected light
I_1, I_2, I_3 …	Intensity of successive images
I_0, $I_{2\pi/3}$, $I_{4\pi/3}$	Intensities for the image point and three consecutive grid positions
k	Number of events
k	Wave number
L	Distance
l_c	Coherence length
m	Diffraction order
M	Magnification
M_{min}	Minimum microscope magnification
MP	Magnifying power
MTF	Modulation transfer function
M_u	Angular magnification
n	Refractive index of the dielectric media
n	Step number
N	Expected value
N	Intensity decrease coefficient
N	Total number of grating lines
n_a	Probability of two-photon excitation
NA	Numerical aperture
n_e	Refractive index for propagation velocity of extraordinary wave
n_m, n_r	Refractive indices of the media surrounding the phase ring and the ring itself
n_o	Refractive index for propagation velocity of ordinary wave
n_1	Refractive index of media 1
n_2	Refractive index of media 2
o, e	Ordinary and extraordinary beams
OD	Optical density
OPD	Optical path difference
OPL	Optical path length
OTF	Optical transfer function
OTL	Optical tube length
P	Probability
P_{avg}	Average power
PO	Vector of light passing through phase object

Glossary of Symbols (cont.)

PSF	Point spread function
Q	Fluorophore quantum yield
r	Radius
r	Reflection coefficients
r, m, and o	Relative, media, or vacuum
r_{AS}	Radius of aperture stop
r_{PR}	Radius of phase ring
r_λ	Radius of filter for wavelength λ
s, s'	Shear between wavefront object and image space
S	Pinhole/slit separation
$s_\perp$	Lateral shift in TIR perpendicular component of electromagnetic vector
$s_\parallel$	Lateral shift in TIR for parallel component of electromagnetic vector
S_{FD}	Factor depending on specific Fraunhofer approximation
SM	Vector of light passing through surrounding media
SNR	Signal-to-noise ratio
SR	Strehl ratio
t	Lens separation
t	Thickness
t	Time
t	Transmission coefficient
T	Time required to pass the distance between wave oscillations
T	Throughput of a confocal microscope
t_c	Coherence time
u, u'	Object, image aperture angle
V_d, V_e	Abbe number as defined for lines d, F, C or e, F′, C′
V_m	Velocity of light in media
w	Width of the slit
x, y	Coordinates
z	Distance along the direction of propagation
z	Fraunhofer diffraction distance
z, z'	Object and image distances
z_m, z_o	Imaged sample depth, length of reference path

Glossary of Symbols (cont.)

α	Angle between vectors of interfering waves
α	Birefringent prism angle
α	Grating incidence angle in plane perpendicular to grating plane
α_s	Visual stereo resolving power
β	Grating diffraction angle in plane perpendicular to grating plane
γ	Angle of fringe localization plane
γ	Convergence angle of a stereo microscope
γ	Incidence / diffraction angle from the plane perpendicular to grating plane
Γ	Retardation
δ	Birefringence
δ	Excitation cross section of dye
δz	Depth perception
Δ_b	Axial phase delay
Δf	Variation in focal length
$\Delta \mathbf{k}$	Phase mismatch
Δz, $\Delta z'$	Object and image separations
$\Delta\lambda$	Wavelength bandwidth
$\Delta\nu$	Frequency bandwidth
$\Delta\varphi$	Phase delay
$\Delta\varphi$	Phase difference
$\Delta\varphi$	Phase shift
$\Delta\varphi_{min}$	Minimum perceived phase difference
ε	Angle between interfering beams
ε	Dielectric constant, i.e., medium permittivity
η	Quantum efficiency
θ'	Refraction angle
θ_{cr}	Critical angle
θ_i	Incidence angle
θ_r	Reflection angle
λ	Wavelength of light
λ_p	Peak wavelength for the mth interference order
μ	Magnetic permeability
ν	Frequency of light
ν	Repetition frequency
ξ	Propagation direction
ρ	Spatial frequency
σ	Molecular absorption cross section

Glossary of Symbols (cont.)

σ_{dark}	Dark noise
σ_{photon}	Photon noise
σ_{read}	Read noise
τ	Integration time
τ	Length of a pulse
τ	Transmittance
φ	Incident photon flux
φ	Optical power
φ	Phase difference generated by a thin film
φ_o	Initial phase
φ_o	Phase delay through the object
φ_p	Phase delay in a phase plate
φ_{TF}	Phase difference generated by a thin film
ω	Angular frequency
ω	Bandgap frequency
$\perp$ and $\parallel$	Perpendicular and parallel components of the light vector

Nature of Light

Optics uses two approaches, termed the **particle** and **wave models,** to describe the nature of light. The former arises from atomic structure, where the transitions between energy levels are quantized. Electrons can be excited into higher energy levels via external processes with the release of a discrete quantum of energy (a photon) upon their decay to a lower level.

The wave model complements this corpuscular theory and explains optical effects involving diffraction and interference. The wave and particle models can be related through the frequency of oscillations, with the relationship between quanta of energy and frequency given by

$$E = h\nu \text{ [in eV or J]},$$

where $h = 4.135667\times10^{-15}$ [eV·s] = 6.626068×10^{-34} [J·s] is **Planck's constant**, and ν is the **frequency of light** [Hz]. The frequency ν of a wave is the reciprocal of its period T [s]:

$$\nu = \frac{1}{T}.$$

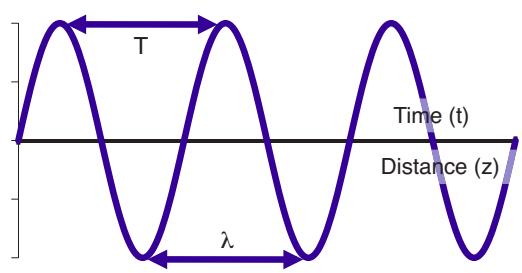

The period T [s] corresponds to one cycle of a wave or, if defined in terms of the distance required to perform one full oscillation, describes the **wavelength of light** λ [m]. The **velocity of light** in free space is 2.99792×10^8 m/s and is defined as the distance traveled in one period (λ) divided by the time taken (T):

$$c = \frac{\lambda}{T}.$$

Note that wavelength is often measured indirectly as time T required to pass the distance between wave oscillations.

The relationship for the velocity of light c can be rewritten in terms of wavelength and frequency as

$$c = \lambda\nu.$$

The Spectrum of Microscopy

The range of electromagnetic radiation is called the **electromagnetic spectrum**. Various microscopy techniques employ wavelengths spanning from **x-ray radiation** and **ultraviolet radiation** (UV) through the **visible spectrum** (VIS) to **infrared radiation** (IR). The wavelength in combination with the microscope parameters determines the resolution limit of the microscope (0.61λ/NA). The smallest feature resolved using light microscopy and determined by diffraction is approximately 200 nm for UV light with a high numerical aperture (for more details, see *Resolution Limit* on page 39). However, recently emerging super-resolution techniques overcome this barrier, and features are observable in the 20–50 nm range.

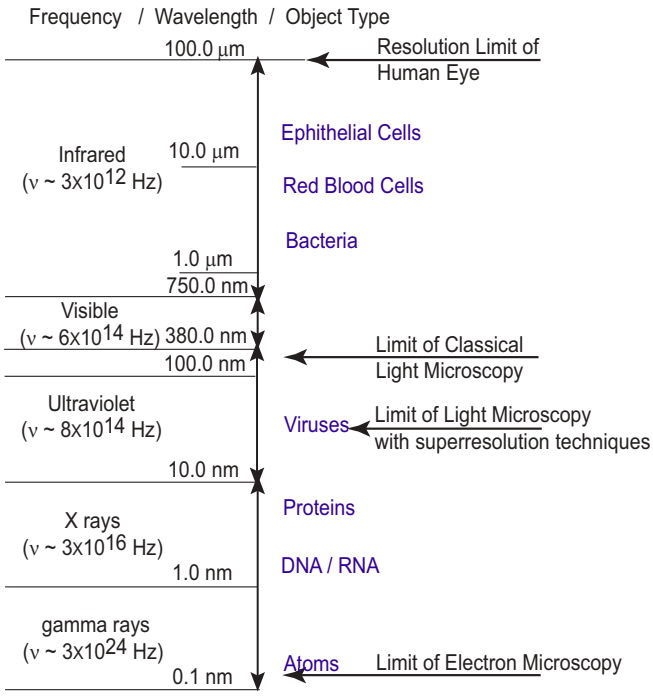

Wave Equations

Maxwell's equations describe the propagation of an electromagnetic wave. For homogenous and isotropic media, magnetic (*H*) and electric (*E*) components of an electromagnetic field can be described by the wave equations derived from Maxwell's equations:

$$\nabla^2 \vec{E} - \varepsilon_m \mu_m \frac{\partial^2 \vec{E}}{\partial t^2} = 0$$

$$\nabla^2 \vec{H} - \varepsilon_m \mu_m \frac{\partial^2 \vec{H}}{\partial t^2} = 0 \,,$$

where ε is a **dielectric constant,** i.e., **medium permittivity** while μ is a **magnetic permeability**:

$$\varepsilon_m = \varepsilon_o \varepsilon_r \qquad \mu_m = \mu_o \mu_r$$

Indices r, m, and o stand for relative, media, or vacuum, respectively.

The above equations indicate that the time variation of the electric field and the current density creates a time-varying magnetic field. Variations in the magnetic field induce a time-varying electric field. Both fields depend on each other and compose an electromagnetic wave. Magnetic and electric components can be separated.

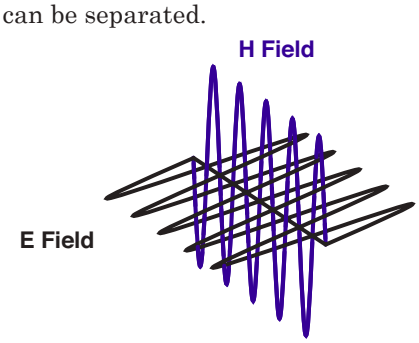

The electric component of an electromagnetic wave has primary significance for the phenomenon of light. This is because the magnetic component cannot be measured with optical detectors. Therefore, the magnetic component is most often neglected, and the electric vector $\vec{E}$ is called a light vector.

Wavefront Propagation

Light propagating through isotropic media conforms to the equation

$$E = A\sin\left(\omega t - kz + \varphi_o\right),$$

where t is time, z is distance along the direction of propagation, and ω is an **angular frequency** given by

$$\omega = \frac{2\pi}{T} = \frac{2\pi V_m}{\lambda}.$$

The term kz is called the **phase of light**, while φ_o is an initial phase. In addition, k represents the **wave number** (equal to $2\pi/\lambda$) and V_m is the velocity of light in media:

$$kz = \frac{2\pi}{\lambda} nz$$

The above propagation equation pertains to the special case when the electric field vector varies only in one plane.

The **refractive index** of the dielectric media n describes the relationship between the speed of light in a vacuum and in media. It is

$$n = \frac{c}{V_m},$$

where c is the speed of light in vacuum.

An alternative way to describe the propagation of an electric field is with an exponential (complex) representation:

$$E = A\exp\left[-i\left(\omega t - kz + \varphi_o\right)\right].$$

This form allows the easy separation of phase components of an electromagnetic wave.

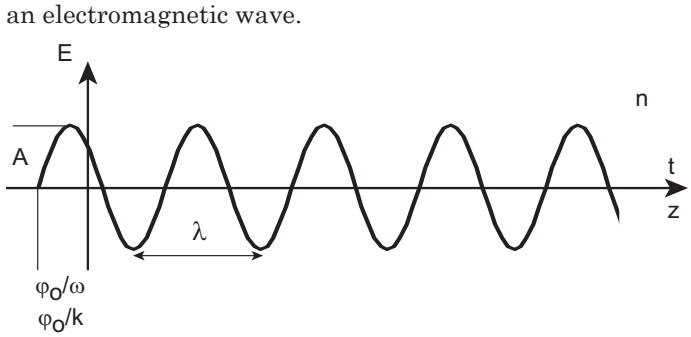

Optical Path Length (OPL)

Fermat's principle states that "The path traveled by a light wave from a point to another point is stationary with respect to variations of this path." In practical terms, this means that light travels along the minimum time path.

Optical path length (OPL) is related to the time required for light to travel from point P1 to point P2. It accounts for the media density through the refractive index:

$$t = \frac{1}{c}\int_{P1}^{P2} n\,ds$$

or

$$OPL = \int_{P1}^{P2} n\,ds\ ,$$

where

$$ds^2 = dx^2 + dy^2 + dz^2\ .$$

Optical path length can also be discussed in the context of the number of periods required to propagate a certain distance L. In a medium with refractive index n, light slows down and more wave cycles are needed. Therefore, OPL is an equivalent path for light traveling with the same number of periods in a vacuum:

$$OPL = nL\ .$$

Optical path difference (OPD) is the difference between optical path lengths traversed by two light waves:

$$OPD = n_1 L_1 - n_2 L_2\ .$$

OPD can also be expressed as a phase difference:

$$\Delta\varphi = 2\pi\frac{OPD}{\lambda}\ .$$

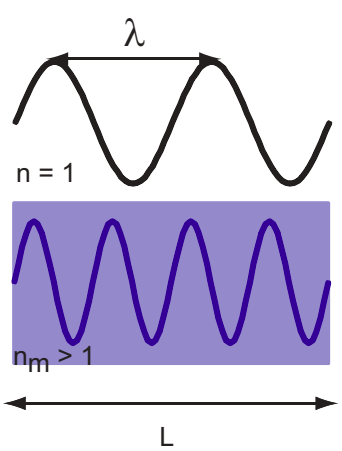

Laws of Reflection and Refraction

Light rays incident on an interface between different dielectric media experience reflection and refraction as shown.

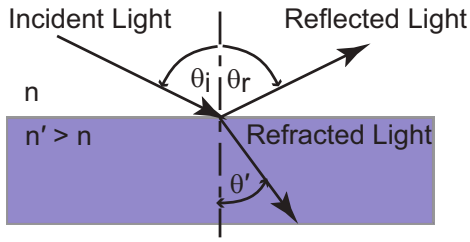

Reflection Law: Angles of **incidence** and **reflection** are related by

$$\theta_i = -\theta_r .$$

Refraction Law (**Snell's Law**): Incident and refracted angles are related to each other and to the refractive indices of the two media by Snell's Law:

$$n \sin\theta_i = n' \sin\theta' .$$

Fresnel reflection: The division of light at a dielectric boundary into transmitted and reflected rays is described for nonlossy, nonmagnetic media by the Fresnel equations:

Reflectance coefficients	Transmission Coefficients
$r_\perp = \frac{I_{r\perp}}{I_\perp} = \frac{\sin^2(\theta_i - \theta')}{\sin^2(\theta_i + \theta')}$	$t_\perp = \frac{I_{t\perp}}{I_\perp} = \frac{4\sin^2\theta'\cos^2\theta_i}{\sin^2(\theta_i + \theta')}$
$r_\parallel = \frac{I_{r\parallel}}{I_\parallel} = \frac{\tan^2(\theta_i - \theta')}{\tan^2(\theta_i + \theta')}$	$t_\parallel = \frac{I_{t\parallel}}{I_\parallel} = \frac{4\sin^2\theta'\cos^2\theta_i}{\sin^2(\theta_i + \theta')\cos^2(\theta_i - \theta')}$

t and r are transmission and reflection coefficients, respectively. I, I_t, and I_r are the irradiances of incident, transmitted, and reflected light, respectively. $\perp$ and $\parallel$ denote perpendicular and parallel components of the light vector with respect to the plane of incidence. θ_i and θ' in the table are the angles of incidence and reflection/refraction, respectively. At a normal incidence ($\theta' = \theta_i = 0$ deg), the Fresnel equations reduce to

$$r = r_\perp = r_\parallel = \frac{(n' - n)^2}{(n' + n)^2} \quad \text{and} \quad t = t_\perp = t_\parallel = \frac{4nn'}{(n' + n)^2} .$$

Total Internal Reflection

When light passes from one medium to a medium with a higher refractive index, the angle of the light in the medium bends toward the normal, according to Snell's Law. Conversely, when light passes from one medium to a medium with a lower refractive index, the angle of the light in the second medium bends away from the normal. If the angle of refraction is greater than 90 deg, the light cannot escape the denser medium, and it reflects instead of refracting. This effect is known as **total internal reflection** (TIR). TIR occurs only for illumination with an angle larger than the **critical angle**, defined as

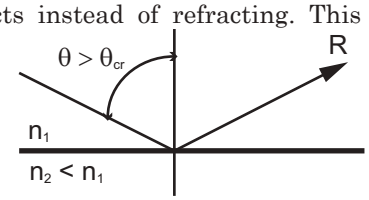

$$\theta_{cr} = \arcsin\left(\frac{n_1}{n_2}\right).$$

It appears, however, that light can propagate through (at a limited range) to the lower-density material as an **evanescent wave** (a nearly standing wave occurring at the material boundary), even if illuminated under an angle larger than θ_{cr}. Such a phenomenon is called **frustrated total internal reflection**. Frustrated TIR is used, for example, with thin films (TF) to build beam splitters. The proper selection of the film thickness provides the desired beam ratio (see the figure below for various split ratios). Note that the maximum film thickness must be approximately a single wavelength or less, otherwise light decays entirely. The effect of an evanescent wave propagating in the lower-density material under TIR conditions is also used in total internal reflection fluorescence (**TIRF**) microscopy.

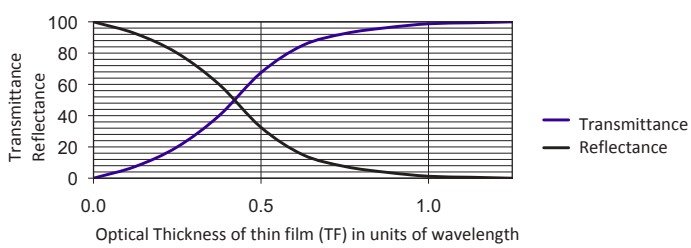

Evanescent Wave in Total Internal Reflection

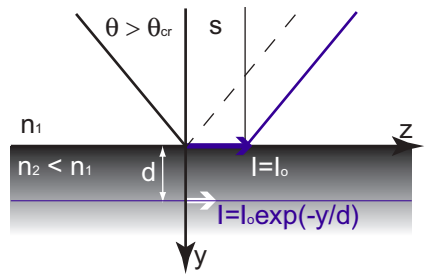

A beam reflected at the dielectric interface during **total internal reflection** is subject to a small lateral shift, also known as the **Goos-Hänchen effect**. The shift is different for the parallel and perpendicular components of the electromagnetic vector.

	Lateral shift in TIR
Parallel* component of e/m vector	$s_{\parallel} = \frac{\lambda n_1}{\pi n_2^2} \frac{\tan\theta}{\sqrt{\sin^2\theta - \sin^2\theta_c}}$
Perpendicular* component of e/m vector	$s_{\perp} = \frac{\lambda}{\pi n_1} \frac{\tan\theta}{\sqrt{\sin^2\theta - \sin^2\theta_c}}$
* to the plane of incidence	

The intensity of the illuminating beam decays exponentially with distance y in the direction perpendicular to the material boundary:

$$I = I_o \exp(-y/d).$$

Note that d denotes the distance at which the intensity of the illuminating light I_o drops by e. The decay distance is smaller than a wavelength. It is also inversely proportional to the illumination angle.

Decay distance (d) of evanescent wave	
Decay distance as a function of incidence and critical angles	$d = \frac{\lambda}{4\pi n_1} \frac{1}{\sqrt{\sin^2\theta - \sin^2\theta_c}}$
Decay distance as a function of incidence angle and refractive indices of media	$d = \frac{\lambda}{4\pi} \frac{1}{\sqrt{n_1^2 \sin^2\theta - n_2^2}}$

Propagation of Light in Anisotropic Media

In anisotropic media the velocity of light depends on the direction of propagation. Common anisotropic and optically transparent materials include **uniaxial crystals**. Such crystals exhibit one direction of travel with a single propagation velocity. The single velocity direction is called the **optic axis** of the crystal. For any other direction, there are two velocities of propagation.

The wave in birefringent crystals can be divided into two components: ordinary and extraordinary. An **ordinary wave** has a uniform velocity in all directions, while the velocity of an **extraordinary wave** varies with orientation. The extreme value of an extraordinary wave's velocity is perpendicular to the optic axis. Both waves are linearly polarized, which means that the electric vector oscillates in one plane. The vibration planes of extraordinary and ordinary vectors are perpendicular. Refractive indices corresponding to the direction of the optic axis and perpendicular to the optic axis are $n_o = c/V_o$ and $n_e = c/V_e$ respectively.

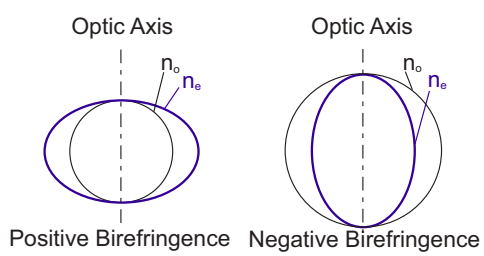

Uniaxial crystals can be positive or negative (see table). The **refractive index** $n(\xi)$ for velocities between the two extreme (n_o and n_e) values is

$$\frac{1}{n^2(\xi)} = \frac{\cos^2\xi}{n_o^2} + \frac{\sin^2\xi}{n_e^2}.$$

Positive Birefringence	$V_e \le V_o$
Negative Birefringence	$V_e \ge V_o$

Note that the propagation direction ξ (with regard to the optic axis) is slightly off from the ray direction, which is defined by the Poynting (energy) vector.

Uniaxial Crystal	**Refractive index**	**Abbe Number**	**Wavelength Range** [μm]
Quartz	$n_o = 1.54424$ $n_e = 1.55335$	70 69	0.18–4.0
Calcite	$n_o = 1.65835$ $n_e = 1.48640$	50 68	0.2–2.0
The refractive index is given for the *D* spectral line (589.2 nm).			

(Adapted from Pluta, 1988)

Polarization of Light and Polarization States

The orientation characteristic of a light vector vibrating in time and space is called the **polarization of light**. For example, if the vector of an electric wave vibrates in one plane, the state of polarization is linear. A vector vibrating with a random orientation represents unpolarized light.

The wave vector E consists of two components E_x and E_y.

$$E(z,t) = E_x + E_y$$

$$E_x = A_x \exp\left[-i\left(\omega t - kz + \varphi_x\right)\right]$$

$$E_y = A_y \exp\left[-i\left(\omega t - kz + \varphi_y\right)\right].$$

The **electric vector** rotates periodically (the periodicity corresponds to wavelength λ) in the plane perpendicular to the propagation axis (z) and generally forms an elliptical shape. The specific shape depends on the A_x and A_y ratios and the phase delay between the E_x and E_y components, defined as $\Delta\varphi = \varphi_x - \varphi_y$.

Linearly polarized light is obtained when one of the components E_x or E_y is zero, or when $\Delta\varphi$ is zero or π. **Circularly polarized light** is obtained when $E_x = E_y$ and $\Delta\varphi = \pm\pi/2$. The light is called right circularly polarized (RCP) if it rotates in a clockwise direction or left circularly polarized (LCP) if it rotates counterclockwise when looking at the oncoming beam.

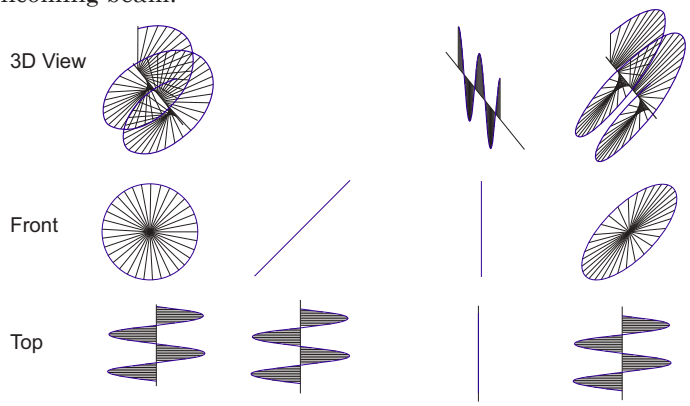

Polarization	Circular	Linear	Linear	Elliptical
Amplitudes	1 1	1 1	0 1	1 1
$\Delta\varphi$	$\pi/2$	0	0	$\pi/4$

Coherence and Monochromatic Light

An ideal light wave that extends in space at any instance to $\pm\infty$ and has only one wavelength λ (or frequency ν) is said to be **monochromatic**. If the range of wavelengths λ or frequencies ν is very small (around λ_o or ν_o, respectively), the wave is **quasi-monochromatic**. Such a wave packet is usually called a **wave group**.

Note that for monochromatic and quasi-monochromatic waves, no phase relationship is required, and the intensity of light can be calculated as a simple summation of intensities from different waves; phase changes are very fast and random, so only the average intensity can be recorded.

If multiple waves have a common phase relation dependence, they are **coherent** or **partially coherent**. These cases correspond to full- and partial-phase correlation respectively. A common source of a coherent wave is a laser where waves must be in resonance and therefore in phase. The average length of the wave packet (group) is called the **coherence length**, while the time required to pass this length is called **coherence time**. Both values are linked by the equations

$$\Delta\nu = \frac{1}{t_c} = \frac{V}{l_c}$$

where coherence length is

$$l_c = \frac{\lambda^2}{\Delta\lambda}.$$

The coherence length l_c and **temporal coherence** t_c are inversely proportional to the bandwidth $\Delta\lambda$ of the light source.

Spatial coherence is a term related to the coherence of light with regard to the extent of the light source. The fringe contrast varies for interference of any two spatially different source points. Light is **partially coherent** if its coherence is limited by the source bandwidth, dimension, temperature, or other effects.

Interference

Interference is a process of superposition of two coherent (correlated) or partially coherent waves. Waves interact with each other, and the resulting intensity is described by summing the **complex amplitudes** (**electric fields**) E_1 and E_2 of both wavefronts:

$$E_1 = A_1 \exp\left[-i\left(\omega t - kz + \varphi_1\right)\right]$$

and

$$E_2 = A_2 \exp\left[-i\left(\omega t - kz + \varphi_2\right)\right].$$

The resultant field is

$$E = E_1 + E_2 .$$

Therefore, the **interference** of the two beams can be written as

$$I = EE^*,$$

$$I = A_1^2 + A_2^2 + 2A_1A_2\cos\left(\Delta\varphi\right),$$

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\left(\Delta\varphi\right),$$

$$I_1 = E_1E_1^*, \quad I_2 = E_2E_2^*, \text{ and}$$

$$\Delta\varphi = \varphi_2 - \varphi_1,$$

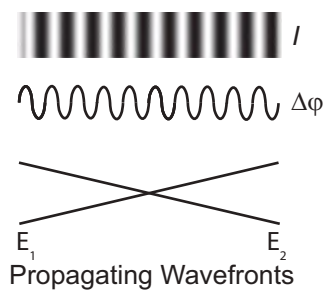

where * denotes a conjugate function, I is the **intensity of light**, A is an amplitude of an electric field, and $\Delta\varphi$ is the phase difference between the two interfering beams. **Contrast** C (called also **visibility**) of the interference fringes can be expressed as

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}.$$

The fringe existence and visibility depends on several conditions. To obtain the interference effect,

- Interfering beams must originate from the same light source and be temporally and spatially coherent;
- The polarization of interfering beams must be aligned;
- To maximize the contrast, interfering beams should have equal or similar amplitudes.

Conversely, if two noncorrelated, random waves are in the same region of space, the sum of intensities (irradiances) of these waves gives a total intensity in that region: $I = I_1 + I_2$.

Contrast vs Spatial and Temporal Coherence

The result of interference is a periodic intensity change in space, which creates fringes when incident on a screen or detector. **Spatial coherence** relates to the contrast of these fringes depending on the extent of the source and is not a function of the phase difference (or OPD) between beams. The intensity of interfering fringes is given by

$$I = I_1 + I_2 + 2C\left(\text{Source Extent}\right)\sqrt{I_1 I_2}\cos\left(\Delta\varphi\right),$$

where C is a constant, depending on the extent of the source.

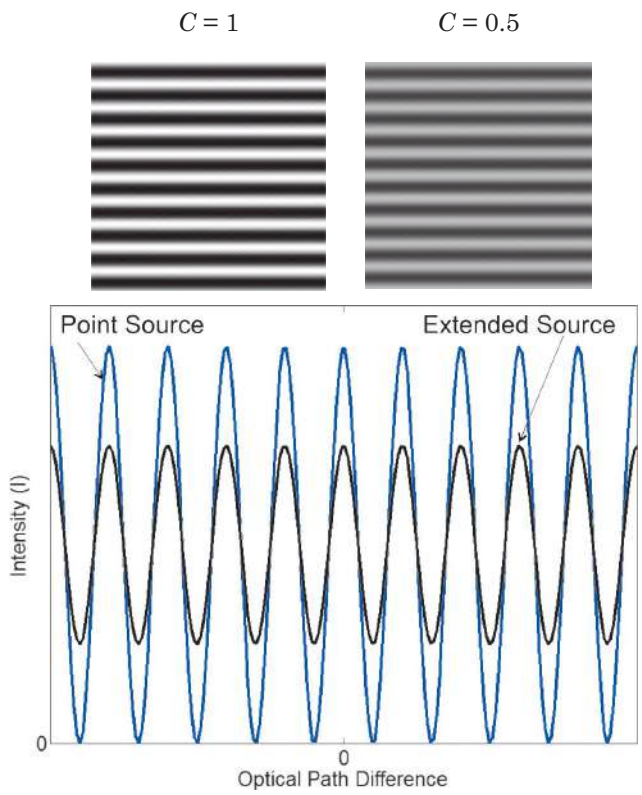

The spatial coherence can be improved through spatial filtering. For example, light can be focused on the pinhole (or coupled into the fiber) by using a microscope objective. In microscopy, spatial coherence can be adjusted by changing the diaphragm size in the conjugate plane of the light source.

Contrast vs Spatial and Temporal Coherence (cont.)

The intensity of the fringes depends on the OPD and the **temporal coherence** of the source. The fringe contrast trends toward zero as the OPD increases beyond the coherence length:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\, C(OPD)\cos(\Delta\varphi).$$

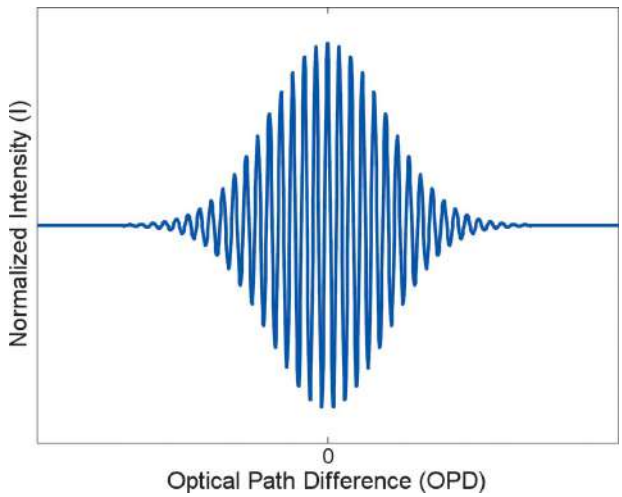

The spatial width of a fringe pattern envelope decreases with shorter temporal coherence. The location of the envelope's peak (with respect to the reference) and narrow width can be efficiently used as a gating mechanism in both imaging and metrology. For example, it is used in interference microscopy for optical profiling (white light interferometry) and for imaging using optical coherence tomography. Examples of short coherence sources include white light sources, luminescent diodes, and broadband lasers.

Long coherence length is beneficial in applications that do not naturally provide near-zero OPD, e.g., in surface measurements with a Fizeau interferometer.

Contrast of Fringes (Polarization and Amplitude Ratio)

The contrast of fringes depends on the polarization orientation of the electric vectors. For linearly polarized beams, it can be described as $C = |\cos\alpha|$, where α represents the angle between polarization states.

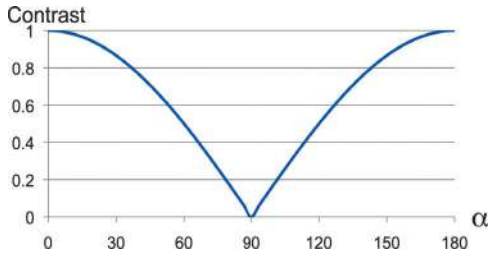

Angle between Electric Vectors of Interfering Beams			
0	π/6	π/3	π/2
Interference Fringes			
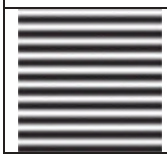	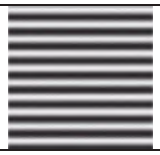	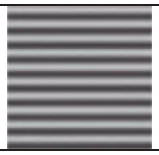	

Fringe contrast also depends on the interfering beams' amplitude ratio. The intensity of the fringes is simply calculated by using the main interference equation. The contrast is maximum for equal beam intensities and for the interference pattern, defined as

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\Delta\varphi),$$

it is

$$C = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}.$$

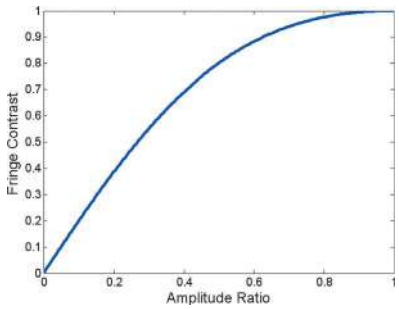

Ratio between Amplitudes of Interfering Beams			
1	0.5	0.25	0.1
Interference Fringes			
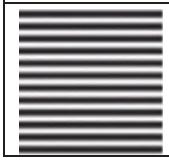	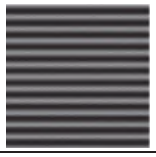		

Multiple Wave Interference

If light reflects inside a thin film, its intensity gradually decreases and **multiple beam interference** occurs.

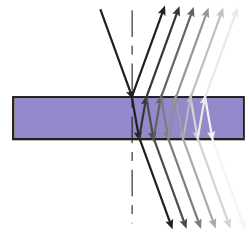

The intensity of reflected light is

$$I_r = I_i \frac{F \sin^2(\varphi/2)}{1 + F \sin^2(\varphi/2)} ,$$

and for transmitted light is

$$I_t = I_i \frac{1}{1 + F \sin^2(\varphi/2)} .$$

The coefficient of finesse F of such a resonator is

$$F = \frac{4r}{(1-r)^2} .$$

Because phase depends on the wavelength, it is possible to design selective **interference filters**. In this case, a dielectric thin film is coated with metallic layers. The peak wavelength for the mth interference order of the filter can be defined as

$$\lambda_p = \frac{2nt\cos\theta'}{m - \left(\frac{\varphi_{TF}}{2\pi}\right)} ,$$

where φ_{TF} is the phase difference generated by a thin film of thickness t and a specific incidence angle. The **interference order** relates to the phase difference in multiples of 2π.

Interference filters usually operate for illumination with flat wavefronts at a normal incidence angle. Therefore, the equation simplifies as

$$\lambda_p = \frac{2nt}{m} .$$

The **half bandwidth** (HBW) of the filter for normal incidence is

$$HBW = \frac{1-r}{m\pi}\lambda_p \text{ [m]} .$$

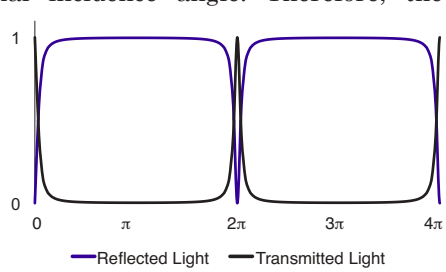

The peak intensity transmission is usually at 20% to 50%, or up to 90% of the incident light for metal-dielectric or multi-dielectric filters respectively.

Interferometers

Due to the high frequency of light, it is not possible to detect the phase of the light wave directly. To acquire the phase, interferometry techniques may be used. There are two major classes of interferometers, based on **amplitude splitting** and **wavefront splitting**.

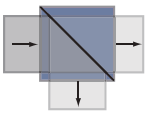

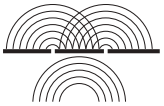

From the point of view of fringe pattern interpretation, interferometers can also be divided into those that deliver fringes, which directly correspond to the phase distribution, and those providing a derivative of the phase map (also called shearing interferometers). The first type is realized by obtaining the interference between the test beam and a reference beam. An example of such a system is the **Michelson interferometer**. **Shearing interferometers** provide the intereference of two shifted object wavefronts.

Michelson Interferometer

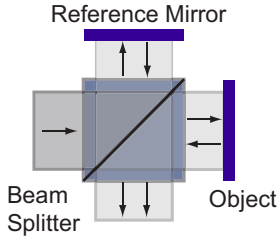

There are numerous ways of introducing shear (linear, radial, etc.). Examples of shearing interferometers include the parallel or wedge plate, the grating interferometer, the **Sagnac**, and polarization interferometers (commonly used in microscopy). The **Mach-Zehnder interferometer** can be configured for direct and differential fringes, depending on the position of the sample.

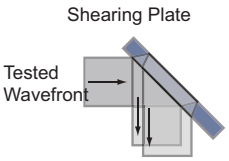

Mach Zehnder Interferometer (Direct Fringes)

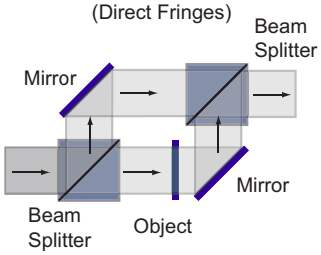

Mach Zehnder Interferometer (Differential Fringes)

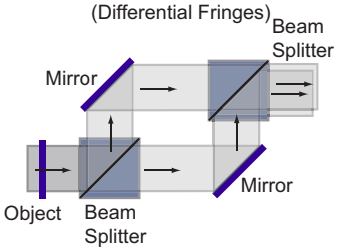

Diffraction

The bending of waves by apertures and objects is called **diffraction of light**. Waves diffracted inside the optical system interfere (for path differences within the coherence length of the source) with each other and create diffraction patterns in the image of an object.

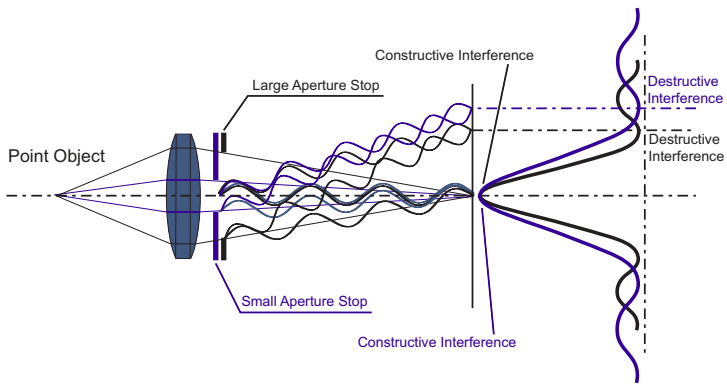

There are two common approximations of diffraction phenomena: **Fresnel diffraction** (near-field) and **Fraunhofer diffraction** (far-field). Both diffraction types complement each other but are not sharply divided due to various definitions of their regions. Fraunhofer diffraction occurs when one can assume that propagating wavefronts are flat (collimated), while the Fresnel diffraction is the near-field case. Thus, Fraunhofer diffraction (distance z) for a free-space case is infinity, but in practice it can be defined for a region

$$z > S_{\mathrm{FD}} \frac{d^2}{\lambda},$$

where d is the diameter of the diffractive object, and S_{FD} is a factor depending on approximation. A conservative definition of the Fraunhofer region is $S_{\mathrm{FD}} = 1$, while for most practical cases it can be assumed to be 10 times smaller ($S_{\mathrm{FD}} = 0.1$).

Diffraction effects influence the resolution of an imaging system and are a reason for fringes (ring patterns) in the image of a point. Specifically, Fraunhofer fringes appear in the conjugate image plane. This is due to the fact that the image is located in the Fraunhofer diffracting region of an optical system's aperture stop.

Diffraction Grating

Diffraction gratings are optical components consisting of periodic linear structures that provide "organized" diffraction of light. In practical terms, this means that for specific angles (depending on illumination and grating parameters), it is possible to obtain constructive (2π phase difference) and destructive (π phase difference) interference at specific directions. The angles of constructive interference are defined by the diffraction grating equation and called **diffraction orders** (m).

Gratings can be classified as transmission or reflection (mode of operation), amplitude (periodic amplitude changes) or phase (periodic phase changes), or ruled or holographic (method of fabrication). Ruled gratings are mechanically cut and usually have a triangular profile (each facet can thereby promote select diffraction angles). Holographic gratings are made using interference (sinusoidal profiles).

Diffraction angles depend on the ratio between the grating constant and wavelength so various wavelengths can be separated. This makes them applicable for spectroscopic detection or spectral imaging. The **grating equation** is

$$m\lambda = d \cos\gamma \, (\sin\beta \pm \sin\alpha)$$

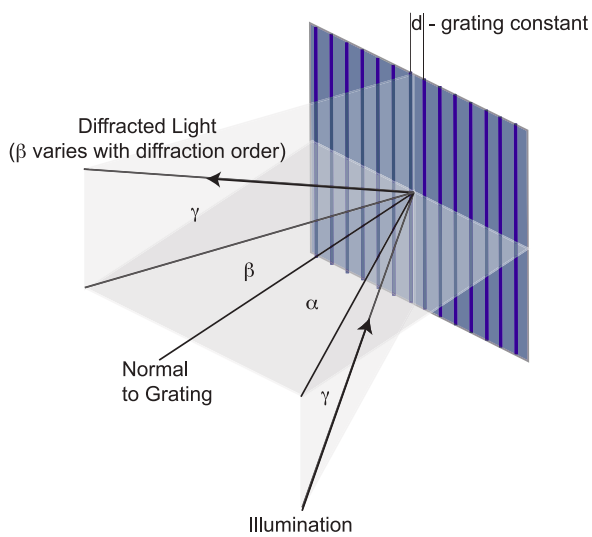

Diffraction Grating (cont.)

The sign in the diffraction grating equation defines the type of grating: A **transmission grating** is identified with a minus sign (–), and a **reflective grating** is identified with a plus sign (+).

For a grating illuminated in the plane perpendicular to the grating, its equation simplifies and is

$$m\lambda = d\left(\sin\beta \pm \sin\alpha\right).$$

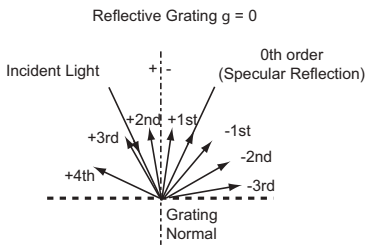

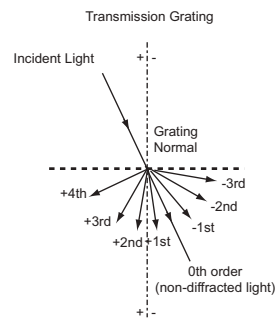

For normal illumination, the grating equation becomes

$$m\lambda = d\sin\beta.$$

The **chromatic resolving power** of a grating depends on its size (total number of lines N). If the grating has more lines, the diffraction orders are narrower, and the resolving power increases:

$$\frac{\lambda}{\Delta\lambda} = mN.$$

The **free spectral range** $\Delta\lambda$ of the grating is the bandwidth in the mth order without overlapping other orders. It defines useful bandwidth for spectroscopic detection as

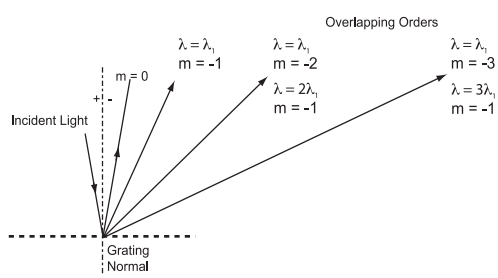

$$\Delta\lambda = \frac{m+1}{m}\lambda_1 - \lambda_1 = \frac{\lambda_1}{m}.$$

Useful Definitions from Geometrical Optics

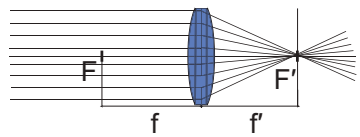

All optical systems consist of refractive or reflective interfaces (surfaces) changing the propagation angles of **optical rays**. Rays define the propagation trajectory and always travel perpendicular to the wavefronts. They are used to describe imaging in the regime of geometrical optics and to perform optical design.

There are two **optical spaces** corresponding to each optical system: **object space** and **image space**. Both spaces extend from $-\infty$ to $+\infty$ and are divided into real and virtual parts. Ability to access the image or object defines the real part of the optical space.

Paraxial optics is an approximation assuming small ray angles and is used to determine first-order parameters of the optical system (image location, magnification etc.).

Thin lenses are optical components reduced to zero thickness and are used to perform first-order optical designs (see next page).

The **focal point** of an optical system is a location that collimated beams converge to or diverge from. Planes perpendicular to the optical axis at the focal points are called **focal planes**. **Focal length** is the distance between the lens (specifically, its principal plane) and the focal plane. For thin lenses, principal planes overlap with the lens.

Sign Convention: The common sign convention assigns positive values to distances traced from left to right (the direction of propagation) and to the top. Angles are positive if they are measured counterclockwise from normal to the surface or optical axis. If light travels from right to left, the refractive index is negative. The surface radius is measured from its vertex to its center of curvature.

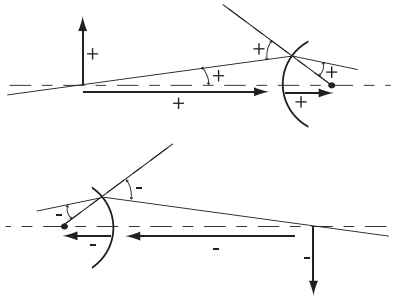

Image Formation

A simple model describing imaging through an optical system is based on thin lens relationships. A **real image** is formed at the point where rays converge.

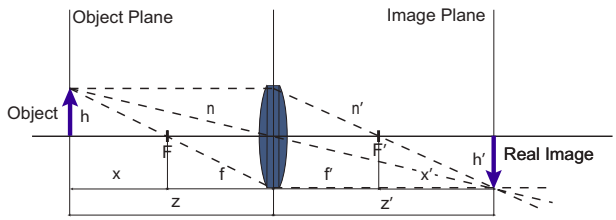

A **virtual image** is formed at the point from which rays appear to diverge.

For a lens made of glass, surrounded on both sides by air ($n = n' = 1$), the imaging relationship is described by the **Newtonian equation**

$$xx' = ff' \qquad \text{or} \qquad xx' = -f'^2 .$$

Note that Newtonian equations refer to the distance from the focal planes so, in practice, they are used with thin lenses.

Imaging can also be described by the **Gaussian imaging equation**

$$\frac{f}{z} + \frac{f'}{z'} = 1 .$$

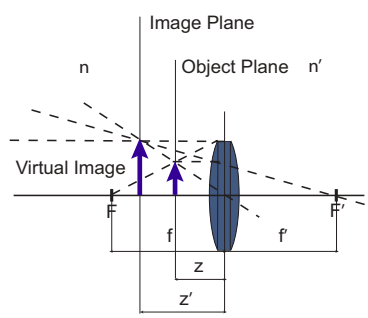

The **effective focal length** of the system is

$$f_e = \frac{1}{\varphi} = -\frac{f}{n} = \frac{f'}{n'} ,$$

where φ is the **optical power** expressed in diopters D [m^{-1}].

Therefore,

$$\frac{n'}{z'} = \frac{n}{z} + \frac{1}{f_e} .$$

For air, when n and n' are equal to 1, the imaging relation is

$$\frac{1}{z'} = \frac{1}{z} + \frac{1}{f_e} .$$

Magnification

Transverse magnification (also called **lateral magnification**) of an optical system is defined as the ratio of the image and object size measured in the direction perpendicular to the optical axis:

$$M = \frac{h'}{h} = -\frac{x'}{f'} = -\frac{f}{x} = \frac{z'}{z} = \frac{-f}{z-f} = \frac{-(z'-f')}{f'}.$$

Longitudinal magnification defines the ratio of distances for a pair of conjugate planes:

$$\frac{\Delta z'}{\Delta z} = \left(-\frac{f'}{f}\right) M_1 M_2,$$

where

$$\Delta z = z_2 - z_1$$
$$\Delta z' = z_2' - z_1'$$

and

$$M_1 = \frac{h_1'}{h_1}$$
$$M_2 = \frac{h_2'}{h_2}.$$

Angular magnification is the ratio of angular image size to angular object size and can be calculated with

$$M_u = \frac{u'}{u} = \frac{z}{z'}.$$

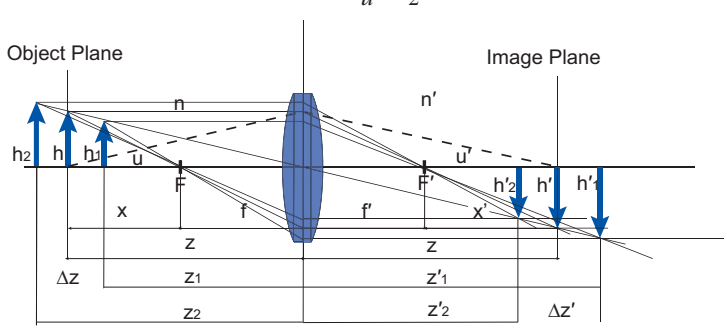

Stops and Rays in an Optical System

The primary stops in any optical system are the **aperture stop** (which limits light) and **field stop** (which limits the extent of the imaged object or the field of view). The aperture stop also defines the resolution of the optical system. To determine the aperture stop, all system diaphragms including the lens mounts should be imaged to either the image or the object space of the system. The aperture stop is determined as the smallest diaphragm or diaphragm image (in terms of angle) as seen from the on-axis object/image point in the same optical space.

Note that there are two important conjugates of the aperture stop in object and image space. They are called the **entrance pupil** and **exit pupil**, respectively.

The physical stop limiting the extent of the field is called the **field stop**. To find a field stop, all of the diaphragms should be imaged to the object or image space, with the smallest diaphragm defining the actual field stop, as seen from the entrance/exit pupil. Conjugates of the field stop in the object and image space are called the **entrance window** and **exit window**, respectively.

Two major rays that pass through the system are the **marginal ray** and the **chief ray**. The marginal ray crosses the optical axis at the conjugate of the field stop (and object) and the edge of the aperture stop. The chief ray goes through the edge of the field and crosses the optical axis at the aperture stop plane.

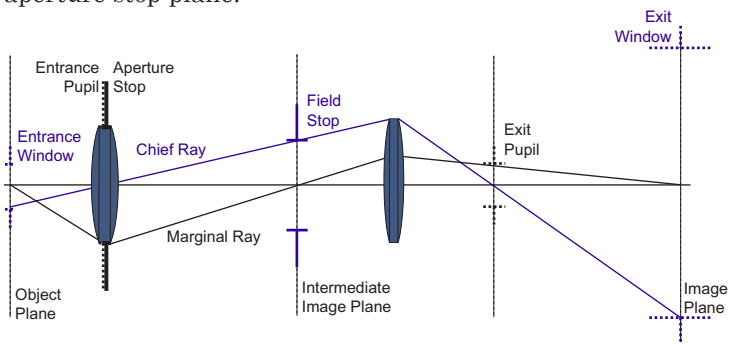

Aberrations

Individual spherical lenses cannot deliver perfect imaging because they exhibit errors called aberrations. All aberrations can be considered either **chromatic** or **monochromatic**. To correct for aberrations, optical systems use multiple elements, aspherical surfaces, and a variety of optical materials.

Chromatic aberrations are a consequence of the **dispersion** of optical materials, which is a change in the refractive index as a function of wavelength. The parameter-characterizing dispersion of any material is called the **Abbe number** and is defined as

$$V_{\mathrm{d}} = \frac{n_{\mathrm{d}} - 1}{n_{\mathrm{F}} - n_{\mathrm{C}}} .$$

Alternatively, the following equation might be used for other wavelengths:

$$V_{\mathrm{e}} = \frac{n_{\mathrm{e}} - 1}{n_{\mathrm{F'}} - n_{\mathrm{C'}}} .$$

λ [nm]	Symbol	Spectral Line
656	C	red hydrogen
644	C′	red cadmium
588	d	yellow helium
546	e	green mercury
486	F	blue hydrogen
480	F′	blue cadmium

In general, V can be defined by using refractive indices at any three wavelengths, which should be specified for material characteristics. Indices in the equations denote spectral lines. If V does not have an index, V_{d} is assumed.

Geometrical aberrations occur when optical rays do not meet at a single point. There are longitudinal and transverse ray aberrations describing the axial and lateral deviations from the paraxial image of a point (along the axis and perpendicular to the axis in the image plane), respectively.

Wave aberrations describe a deviation of the wavefront from a perfect sphere. They are defined as a distance (the optical path difference) between the wavefront and the reference sphere along the optical ray.

Chromatic Aberrations

Chromatic aberrations occur due to the dispersion of optical materials used for lens fabrication. This means that the refractive index is different for different wavelengths; consequently, various wavelengths are refracted differently.

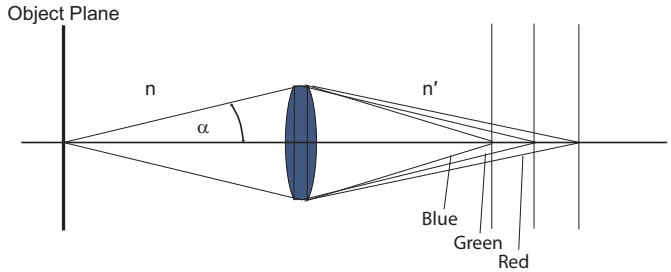

Chromatic aberrations include axial (longitudinal) or transverse (lateral) aberrations. **Axial chromatic aberration** arises from the fact that various wavelengths are focused at different distances behind the optical system. It is described as a variation in focal length:

$$\frac{\Delta f}{f} = \frac{f_F - f_C}{f} = \frac{1}{V}.$$

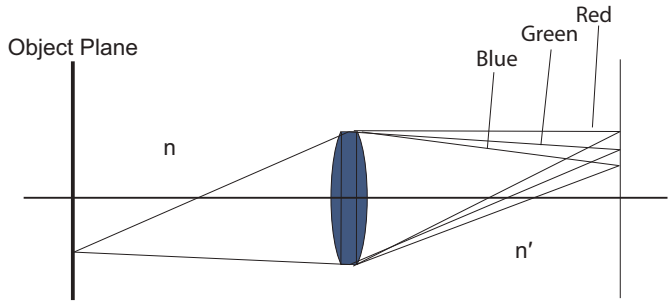

Transverse chromatic aberration is an off-axis imaging of colors at different locations on the image plane.

To compensate for chromatic aberrations, materials with low and high Abbe numbers are used (such as flint and crown glass). Correcting chromatic aberrations is crucial for most microscopy applications, but it is especially important for multi-photon microscopy. Obtaining multi-photon excitation requires high laser power and is most effective using short pulse lasers. Such a light source has a broad spectrum, and chromatic aberrations may cause pulse broadening.

Spherical Aberration and Coma

The most important wave aberrations are spherical, coma, astigmatism, field curvature, and distortion.

Spherical aberration (on-axis) is a consequence of building an optical system with components with spherical surfaces. It occurs when rays from different heights in the pupil are focused at different planes along the optical axis. This results in an axial blur. The most common approach for correcting spherical aberration uses a combination of negative and positive lenses. Systems that correct spherical aberration heavily depend on imaging conditions. For example, in microscopy a cover glass must be of an appropriate thickness and refractive index in order to work with an objective. Also, the media between the objective and the sample (such as air, oil, or water) must be taken into account.

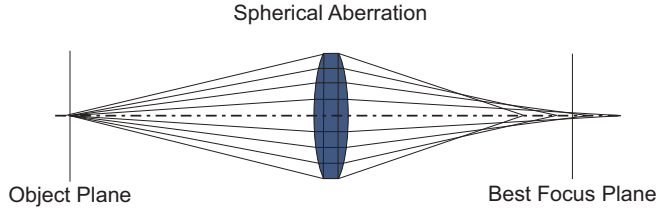

Coma (off-axis) can be defined as a variation of magnification with aperture location. This means that rays passing through a different azimuth of the lens are magnified differently. The name "coma" was inspired by the aberration's appearance, because it resembles a comet's tail as it emanates from the focus spot. It is usually stronger for lenses with a larger field, and its correction requires accommodation of the field diameter.

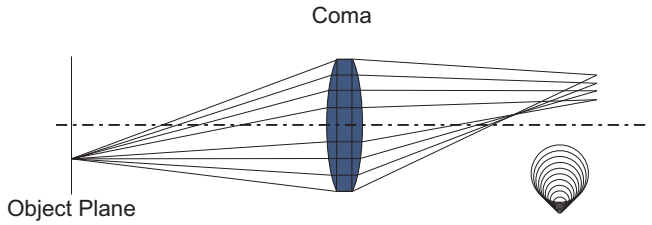

Astigmatism, Field Curvature, and Distortion

Astigmatism (off-axis) is responsible for different magnifications along orthogonal meridians in an optical system. It manifests as elliptical, elongated spots for the horizontal and vertical directions on opposite sides of the best focal plane. It is more pronounced for an object farther from the axis and is a direct consequence of improper lens mounting or an asymmetric fabrication process.

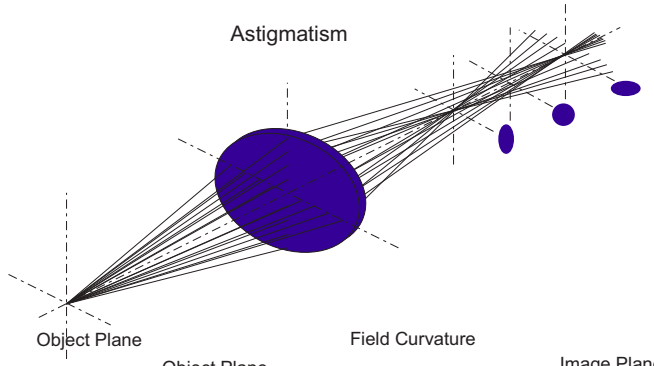

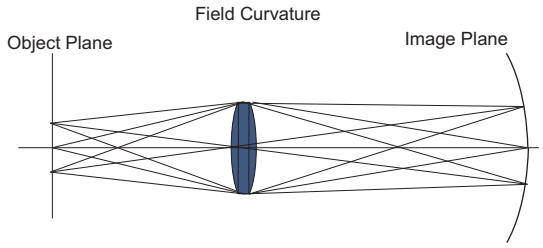

Field curvature (off-axis) results in a non-flat image plane. The image plane created is a concave surface as seen from the objective; therefore, various zones of the image can be seen in focus after moving the object along the optical axis. This aberration is corrected by an objective design combined with a tube lens or eyepiece.

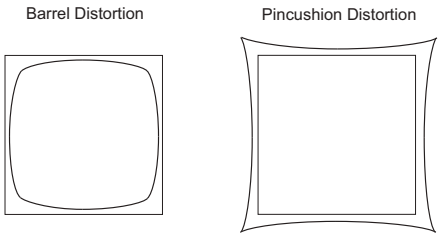

Distortion is a radial variation of magnification that will image a square as a pincushion or barrel. It is corrected in the same manner as field curvature. If preceded with system calibration, it can also be corrected numerically after image acquisition.

Performance Metrics

The major metrics describing the performance of an optical system are the **modulation transfer function** (MTF), the **point spread function** (PSF), and the **Strehl ratio** (SR).

The MTF is the modulus of the **optical transfer function** described by

$$OTF(\rho) = MTF(\rho)\exp\left[i\varphi(\rho)\right],$$

where the complex term in the equation relates to the phase transfer function. The MTF is a contrast distribution in the image in relation to contrast in the object as a function of spatial frequency ρ (for sinusoidal object harmonics) and can be defined as

$$MTF(\rho) = \frac{C(\rho)_{\text{image}}}{C(\rho)_{\text{object}}}.$$

The PSF is the intensity distribution at the image of a point object. This means that the PSF is a metric directly related to the image, while the MTF corresponds to spatial frequency distributions in the pupil. The MTF and PSF are closely related and comprehensively describe the quality of the optical system. In fact, the amplitude of the Fourier transform of the PSF results in the MTF.

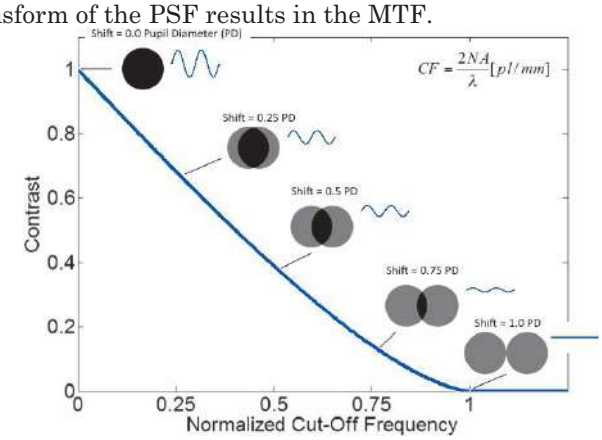

The MTF can be calculated as the autocorrelation of the **pupil function**. The pupil function describes the field distribution of an optical wave in the pupil plane of the optical system. In the case of a uniform pupil's transmission, it directly relates to the field overlap of two mutually shifted pupils where the shift corresponds to spatial frequency.

Performance Metrics (cont.)

The **modulation transfer function** has different results for coherent and incoherent illumination. For incoherent illumination, the phase component of the field is neglected since it is an average of random fields propagating under random angles.

For coherent illumination, the contrast of transferring harmonics of the field is constant and equal to 1 until the location of the spatial frequency in the pupil reaches its edge. For higher frequencies, the contrast sharply drops to zero since they cannot pass the optical system. Note that contrast for the coherent case is equal to 1 for the entire MTF range.

The **cutoff frequency** for an incoherent system is two times the cutoff frequency of the equivalent aperture coherent system and defines the **Sparrow resolution limit**.

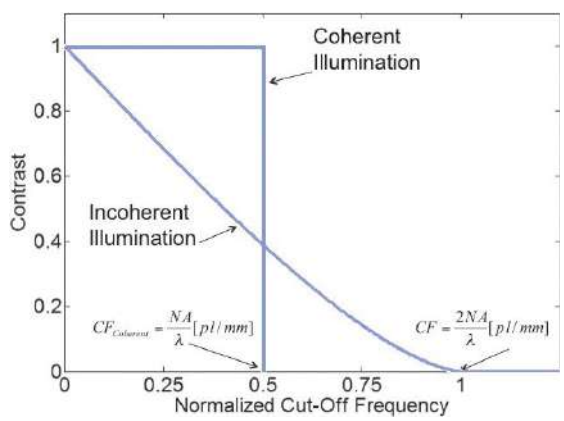

The **Strehl ratio** is a parametric measurement that defines the quality of the optical system in a single number. It is defined as the ratio of irradiance within the theoretical dimension of a diffraction-limited spot to the entire irradiance in the image of the point. One simple method to estimate the Strehl ratio is to divide the field below the MTF curve of a tested system by the field of the diffraction-limited system of the same numerical aperture. For practical optical design consideration it is usually assumed that the system is diffraction limited if the Strehl ratio is equal to or larger than 0.8.

The Compound Microscope

The primary goal of microscopy is to provide the ability to resolve the small details of an object. Historically, microscopy was developed for visual observation and, therefore, was set up to work in conjunction with the human eye. An effective high-resolution optical system must have resolving ability and be able to deliver proper magnification for the detector. In the case of visual observations, the detectors are the cones and rods of the retina.

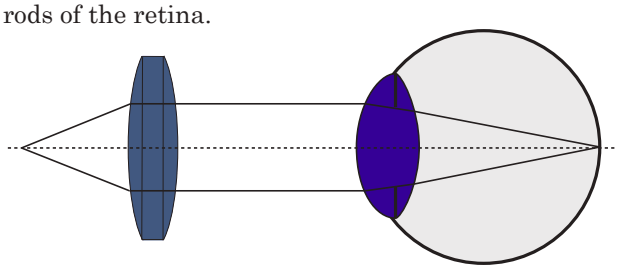

A basic microscope can be built with a single-element, short-focal-length lens (**magnifier**). The object is located in the focal plane of the lens and is imaged to infinity. The eye creates a final image of the object on the retina. The system stop is the eye's pupil.

To obtain higher resolution for visual observation, the **compound microscope** was first built in the 17th century by Robert Hooke. It consists of two major components: the **objective** and the **eyepiece**. The short-focal-length lens (objective) is placed close to the object under examination and creates a real image of an object at the focal plane of the second lens. The eyepiece (similar to the magnifier) throws an image to infinity, and the human eye creates the final image. An important function of the eyepiece is to match the eye's pupil with the system stop, which is located in the back focal plane of the microscope objective.

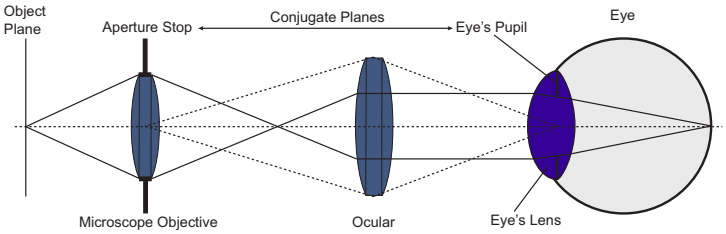

The Eye

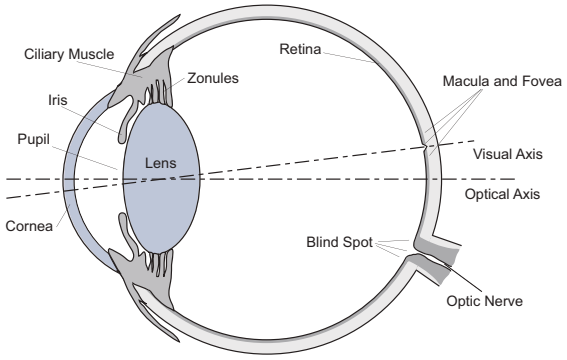

The **eye** was the first—and for a long time, the only—real-time detector used in microscopy. Therefore, the design of the microscope had to incorporate parameters responding to the needs of visual observation.

Cornea—the transparent portion of the sclera (the white portion of the human eye), which is a rigid tissue that gives the eyeball its shape. The cornea is responsible for two thirds of the eye's refractive power.

Lens—the lens is responsible for one third of the eye's power. Ciliary muscles can change the lens's power (accommodation) within the range of 15–30 diopters.

Iris—controls the diameter of the pupil (1.5–8 mm).

Retina—a layer with two types of photoreceptors: rods and cones. **Cones** (about 7 million) are in the area of the **macula** (~3 mm in diameter) and **fovea** (~1.5 mm in diameter, with the highest cone density), and they are designed for bright vision and color detection. There are three types of cones (red, green, and blue sensitive), and the spectral range of the eye is approximately 400–750 nm. **Rods** are responsible for night/low-light vision, and there are about 130 million located outside the fovea region.

It is arbitrarily assumed that the eye can provide sharp images for objects between 250 mm and infinity. A 250-mm distance is called the **minimum focus distance** or **near point**. The maximum eye resolution for bright illumination is 1 arc minute.

Upright and Inverted Microscopes

The two major microscope geometries are **upright** and **inverted.** Both systems can operate in reflectance and transmittance modes.

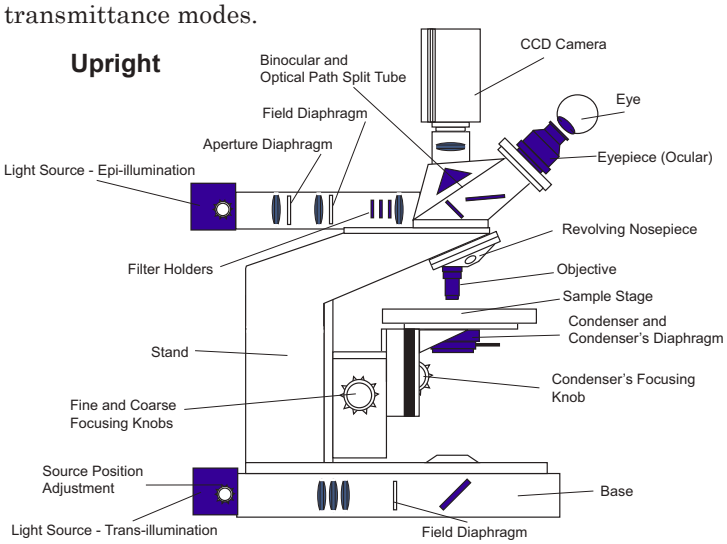

The inverted microscope is primarily designed to work with samples in cell culture dishes and to provide space (with a long-working distance condenser) for sample manipulation (for example, with patch pipettes in electrophysiology).

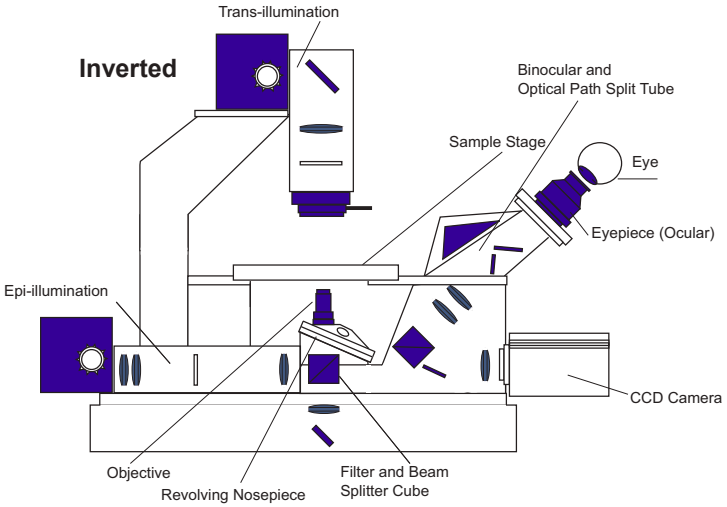

The Finite Tube Length Microscope

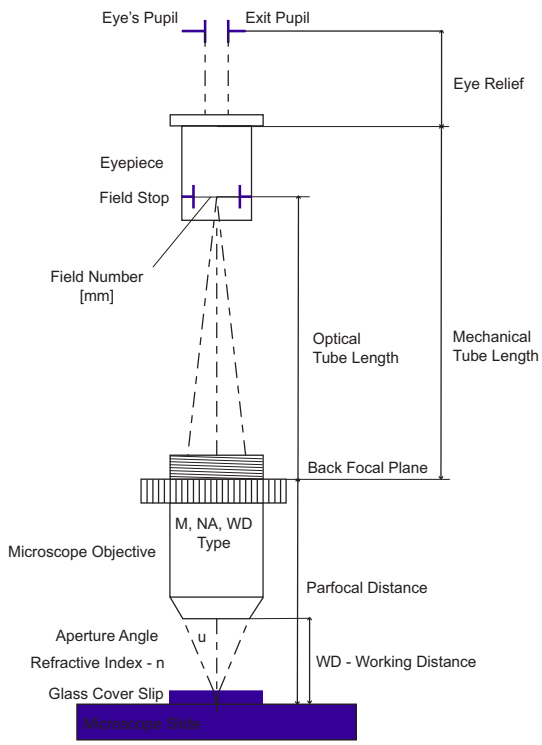

Historically, microscopes were built with a **finite tube length**. With this geometry, the microscope objective images the object into the tube end. This intermediate image is then relayed to the observer by an eyepiece. Depending on the manufacturer, different **optical tube lengths** are possible (for example, the standard tube length for Zeiss is 160 mm). The use of a standard tube length by each manufacturer unifies the optomechanical design of the microscope.

A constant **parfocal distance** for microscope objectives enables switching between different magnifications without defocusing. The **field number**, corresponding to the physical dimension (in millimeters) of the **field stop** inside the eyepiece, allows the microscope's field of view (FOV) to be determined according to

$$FOV = \frac{FieldNumber}{M_{\text{objective}}} \ [\text{mm}].$$

Infinity-Corrected Systems

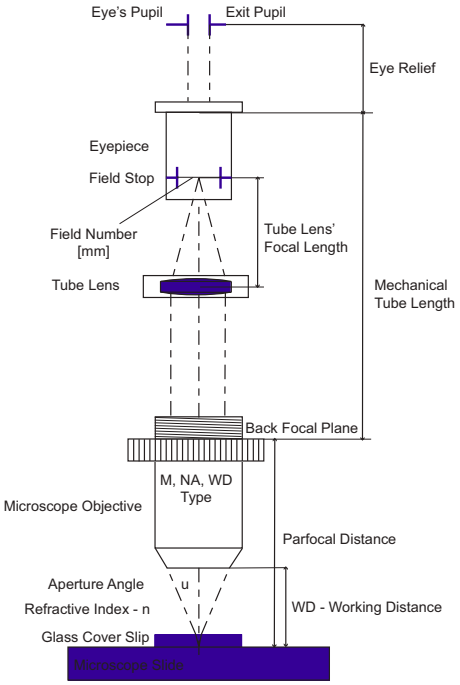

Microscope objectives can be corrected for a conjugate located at infinity, which means that the microscope objective images an object into infinity. An infinite conjugate is useful for introducing beam splitters or filters that should work with small incidence angles. Also, an infinity-corrected system accommodates additional components like DIC prisms, polarizers. etc. The collimated beam is focused to create an intermediate image with additional optics, called a **tube lens**. The tube lens either creates an image directly onto a CCD chip or an intermediate image, which is further reimaged with an eyepiece. Additionally, the tube lens might be used for system correction. For example, Zeiss corrects aberrations in its microscopes with a combination objective-tube lens.

In the case of an **infinity-corrected objective**, the transverse magnification can only be defined in the presence of a tube lens that will form a real image. It is given by the ratio between the tube lens's focal length and the focal length of the microscope objective.

Manufacturer	Focal Length of Tube Lens
Zeiss	164.5 mm
Olympus	180.0 mm
Nikon	200.0 mm
Leica	200.0 mm

Telecentricity of a Microscope

Telecentricity is a feature of an optical system where the principal ray in object, image, or both spaces is parallel to the optical axis. This means that the object or image does not shift laterally, even with defocus; the distance between two object or image points is constant along the optical axis.

An optical system can be **telecentric** in

- Object space, where the entrance pupil is at infinity, and the aperture stop is in the back focal plane;
- Image space, where the exit pupil is at infinity, and the aperture stop is in the front focal plane; or
- Both (doubly telecentric), where the entrance and exit pupils are at infinity, and the aperture stop is at the center of the system, in the back focal plane of the element before the stop and front focal length of the element after the stop (afocal system).

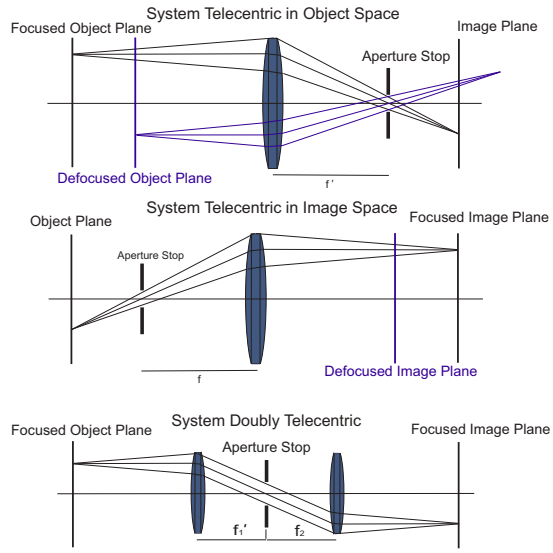

The aperture stop in a microscope is located at the back focal plane of the microscope objective. This makes the microscope objective telecentric in object space. Therefore, in microscopy, the object is observed with constant magnification, even for defocused object planes. This feature of microscopy systems significantly simplifies their operation and increases the reliability of image analysis.

Magnification of a Microscope

Magnifying power (MP) is defined as the ratio between the angles subtended by an object with and without magnification. The magnifying power (as defined for a single lens) creates an enlarged virtual image of an object. The angle of an object observed with magnification is

$$u' = \frac{h'}{z'-l} = \frac{h(f-z')}{f(z'-l)}.$$

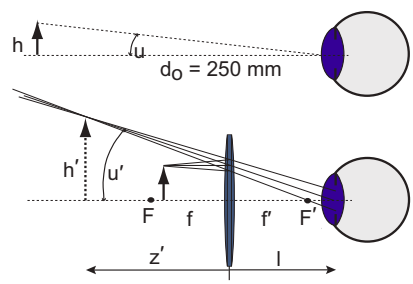

Therefore,

$$MP = \frac{u'}{u} = \frac{d_o(f-z')}{f(z'-l)}.$$

The angle for an unaided eye is defined for the **minimum focus distance** (d_o) of 10 inches or 250 mm, which is the distance that the object (real or virtual) may be examined without discomfort for the average population. A distance l between the lens and the eye is often small and can be assumed to equal zero:

$$MP \approx \frac{250\text{mm}}{f} - \frac{250\text{mm}}{z'}.$$

If the virtual image is at infinity (observed with a relaxed eye), $z' = -\infty$, and

$$MP \approx \frac{250\text{mm}}{f}.$$

The total magnifying power of the microscope results from the magnification of the microscope objective and the magnifying power of the eyepiece (usually 10×):

$$M_{\text{objective}} = \frac{OTL}{f_{\text{objective}}}$$

$$MP_{\text{microscope}} = M_{\text{objective}} MP_{\text{eyepiece}} = -\frac{OTL}{f_{\text{objective}}} \frac{250\text{mm}}{f_{\text{eyepiece}}}.$$

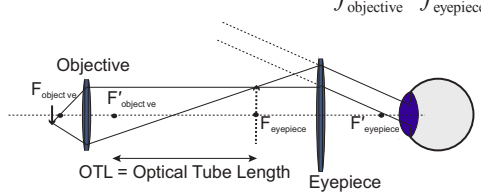

Numerical Aperture

The aperture diaphragm of the optical system determines the angle at which rays emerge from the axial object point and, after refraction, pass through the optical system. This acceptance angle is called the object space aperture angle. The parameter describing system throughput and this acceptance angle is called the **numerical aperture** (NA):

$$NA = n \sin u \,.$$

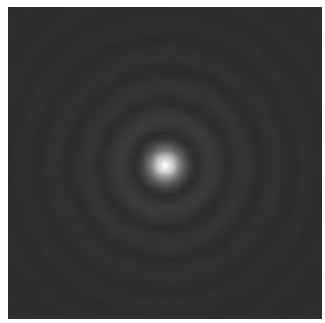

As seen from the equation, throughput of the optical system may be increased by using media with a high refractive index n, e.g., oil or water. This effectively decreases the refraction angles at the interfaces.

The dependence between the numerical aperture in the object space NA and the numerical aperture in the image space between the objective and the eyepiece NA′ is calculated using the objective magnification:

$$NA = NA' M_{\text{objective}} \,.$$

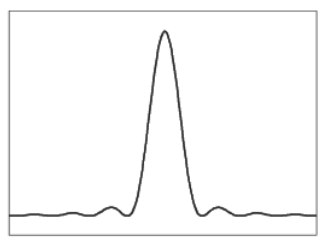

As a result of diffraction at the aperture of the optical system, self-luminous points of the object are not imaged as points but as so-called Airy disks. An **Airy disk** is a bright disk surrounded by concentric rings with gradually decreasing intensities. The disk diameter (where the intensity reaches the first zero) is

$$d = \frac{1.22\lambda}{n \sin u} = \frac{1.22\lambda}{NA} \,.$$

Media	Refractive Index
Air	1
Water	1.33
Oil	1.45–1.6 (1.515 is typical)

Note that the refractive index in the equation is for media between the object and the optical system.

Resolution Limit

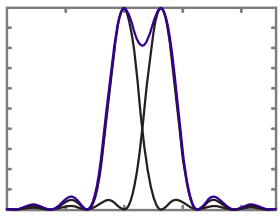

The lateral resolution of an optical system can be defined in terms of its ability to resolve images of two adjacent, self-luminous points. When two **Airy disks** are too close, they form a continuous intensity distribution and cannot be distinguished. The **Rayleigh resolution limit** is defined as occurring when the peak of the Airy pattern arising from one point coincides with the first minimum of the Airy pattern arising from a second point object. Such a distribution gives an intensity dip of 26%. The distance between the two points in this case is

$$d = \frac{0.61\lambda}{n \sin u} = \frac{0.61\lambda}{NA}.$$

The equation indicates that the resolution of an optical system improves with an increase in *NA* and decreases with increasing wavelength λ. For example, for λ = 450 nm (blue) and oil immersion *NA* = 1.4, the microscope objective can optically resolve points separated by less than 200 nm.

The situation in which the intensity dip between two adjacent self-luminous points becomes zero defines the **Sparrow resolution limit**. In this case $d = 0.5\lambda/NA$.

The **Abbe resolution limit** considers both the diffraction caused by the object and the *NA* of the optical system. It assumes that if at least two adjacent diffraction orders for points with spacing *d* are accepted by the objective, these two points can be resolved. Therefore, the resolution depends on both imaging and illumination apertures and is

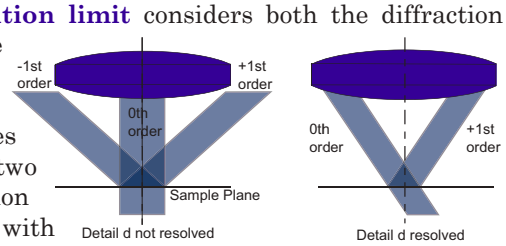

$$d = \frac{\lambda}{NA_{\text{objective}} + NA_{\text{condenser}}}.$$

Useful Magnification

For visual observation, the **angular resolving power** can be defined as 1.5 arc minutes. For unmagnified objects at a distance of 250 mm (the eye's minimum focus distance), 1.5 arc minutes converts to d_{eye} = 0.1 mm. Since microscope objectives by themselves do not usually provide sufficient magnification, they are combined with oculars or eyepieces. The resolving power of the microscope is then

$$d_{\text{mic}} = \frac{d_{\text{eye}}}{M_{\text{min}}} = \frac{d_{\text{eye}}}{M_{\text{obj}} M_{\text{eyepiece}}} .$$

In the Sparrow resolution limit, the minimum microscope magnification is

$$M_{\text{min}} = 2 d_{\text{eye}} NA / \lambda .$$

Therefore, a total minimum magnification M_{min} can be defined as approximately 250–500 NA (depending on wavelength). For lower magnification the image will appear brighter, but imaging is performed below the overall resolution limit of the microscope. For larger magnification, the contrast decreases and resolution does not improve. While a theoretically higher magnification should not provide additional information, it is useful to increase it to approximately 1000 NA to provide comfortable sampling of the object. Therefore, it is assumed that the **useful magnification** of a microscope is between 500 NA and 1000 NA. Usually, any magnification above 1000 NA is called **empty magnification**. The image size in such cases is enlarged, but no additional useful information is provided. The highest useful magnification of a microscope is approximately 1500 for an oil-immersion microscope objective with NA = 1.5.

Similar analysis can be performed for digital microscopy, which uses CCD or CMOS cameras as image sensors. Camera pixels are usually small (between 2 and 30 microns), and useful magnification must be estimated for a particular image sensor rather than the eye. Therefore, digital microscopy can work at lower magnification, and magnification of the microscope objective alone is usually sufficient.

Depth of Field and Depth of Focus

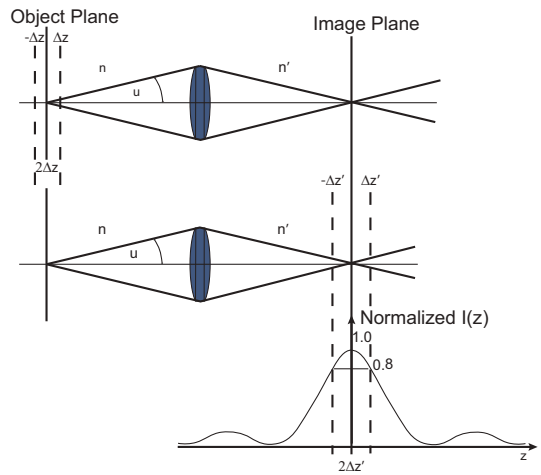

Two important terms are used to define axial resolution: **depth of field** and **depth of focus**. Depth of field refers to the object thickness for which an image is in focus, while depth of focus is the corresponding thickness of the image plane. In the case of a diffraction-limited optical system, the depth of focus is determined for an intensity drop along the optical axis to 80%, defined as

$$DOF = 2\Delta z' = \frac{n'\lambda}{NA'^2}.$$

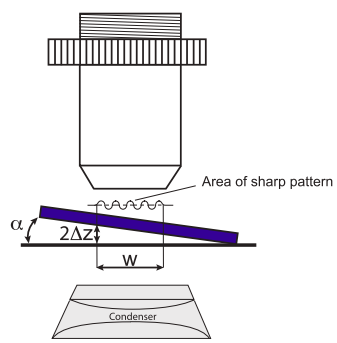

The relation between depth of field ($2\Delta z$) and depth of focus ($2\Delta z'$) incorporates the objective magnification:

$$2\Delta z' = M^2_{\text{objective}}\, 2\Delta z \frac{n'}{n},$$

where n' and n are medium refractive indexes in the object and image space, respectively.

To quickly measure the depth of field for a particular microscope objective, the grid amplitude structure can be placed on the microscope stage and tilted with the known angle α. The depth of field is determined after measuring the grid zone w while in focus:

$$2\Delta z = nw \tan\alpha\,.$$

Magnification and Frequency vs Depth of Field

Depth of field for visual observation is a function of the total microscope magnification and the NA of the objective, approximated as

$$2\Delta z = n\frac{0.5\lambda}{NA^2} + n\frac{340}{M_{\text{microscope}}NA}.$$

Note that estimated values do not include eye accommodation. The graph below presents depth of field for visual observation. Refractive index n of the object space was assumed to equal 1. For other media, values from the graph must be multiplied by an appropriate n.

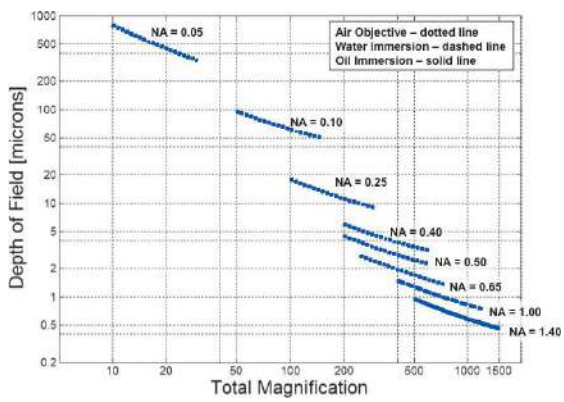

Depth of field can also be defined for a specific frequency present in the object because imaging contrast changes for a particular object frequency, as described by the modulation transfer function. The approximated equation is

$$2\Delta z = \frac{0.4}{\rho NA},$$

where ρ is the frequency in cycles per millimeter.

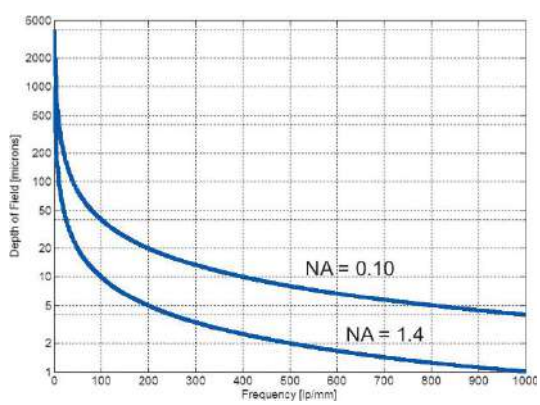

Köhler Illumination

One of the most critical elements in efficient microscope operation is proper set up of the illumination system. To assist with this, August Köhler introduced a solution that provides uniform and bright illumination over the field of view, even for light sources that are not uniform (e.g., a lamp filament). This system consists of a light source, a field-stop diaphragm, a condenser aperture, and collective and condenser lenses. This solution is now called **Köhler illumination** and is commonly used for a variety of imaging modes.

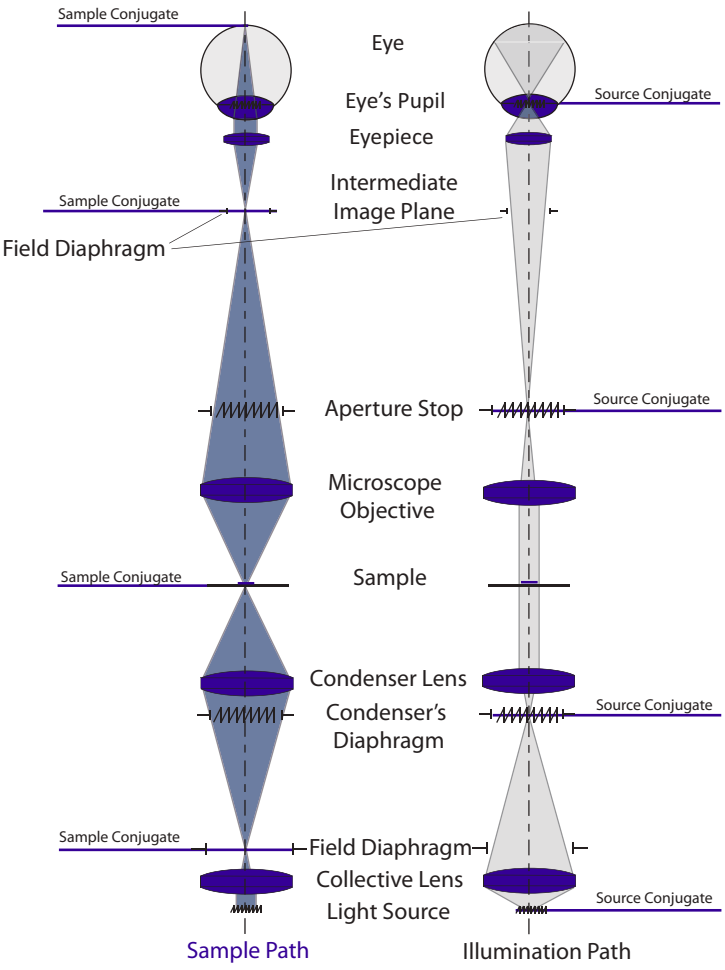

Köhler Illumination (cont.)

The illumination system is configured such that an image of the light source completely fills the condenser aperture.

Köhler illumination requires setting up the microscope system so that the field diaphragm, object plane, and intermediate image in the eyepiece's field stop, retina, or CCD are in conjugate planes. Similarly, the lamp filament, front aperture of the condenser, microscope objective's back focal plane (aperture stop), exit pupil of the eyepiece, and the eye's pupil are also at conjugate planes. Köhler illumination can be set up for both the transmission mode (also called **DIA**) and the reflectance mode (also called **EPI**).

The uniformity of the illumination is the result of having the light source at infinity with respect to the illuminated sample.

EPI illumination is especially useful for the biological and metallurgical imaging of thick or opaque samples.

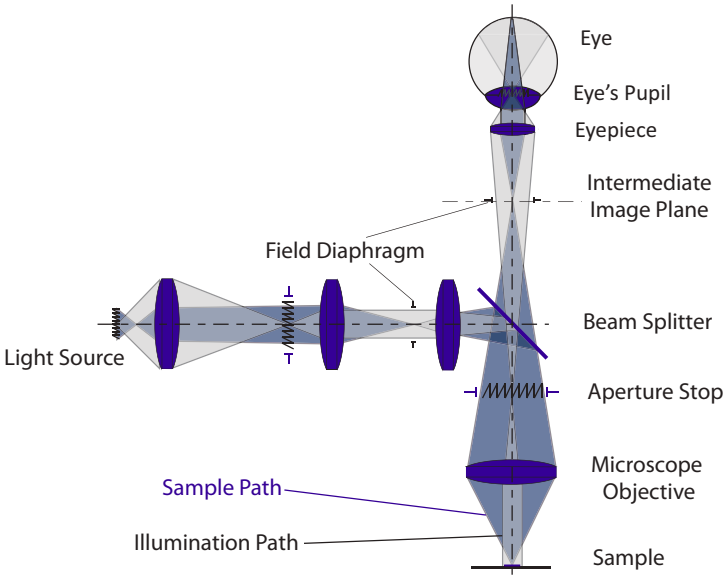

Alignment of Köhler Illumination

The procedure for **Köhler illumination** alignment consists of the following steps:

1. Place the sample on the microscope stage and move it so that the front surface of the condenser lens is 1–2 mm from the microscope slide. The condenser and collective diaphragms should be wide open.
2. Adjust the location of the source so illumination fills the condenser's diaphragm. The sharp image of the lamp filament should be visible at the condenser's diaphragm plane and at the back focal plane of the microscope objective. To see the filament at the back focal plane of the objective, a **Bertrand lens** can be applied (also see *Special Lens Components* on page 55). The filament should be centered in the aperture stop using the bulb's adjustment knobs on the illuminator.
3. For a low-power objective, bring the sample into focus. Since a microscope objective works at a parfocal distance, switching to a higher magnification later is easy and requires few adjustments.
4. Focus and center the condenser lens. Close down the field diaphragm, focus down the outline of the diaphragm, and adjust the condenser's position. After adjusting the x-, y-, and z-axis, open the field diaphragm so it accommodates the entire field of view.
5. Adjust the position of the condenser's diaphragm by observing the back focal objective plane through the Bertrand lens. When the edges of the aperture are sharply seen, the condenser's diaphragm should be closed to approximately three quarters of the objective's aperture.
6. Tune the brightness of the lamp. This adjustment should be performed by regulating the voltage of the lamp's power supply or, preferably, through the neutral density filters in the beam path. Do not adjust the brightness by closing the condenser's diaphragm because it affects the illumination setup and the overall microscope resolution.

Note that while three quarters of the aperture stop is recommended for initial illumination, adjusting the aperture of the illumination system affects the resolution of the microscope. Therefore, the final setting should be adjusted after examining the images.

Critical Illumination

An alternative to Köhler illumination is **critical illumination**, which is based on imaging the light source directly onto the sample. This type of illumination requires a highly uniform light source. Any source non-uniformities will result in intensity variations across the image. Its major advantage is high efficiency, since it can collect a larger solid angle than Köhler illumination and therefore provides a higher energy density at the sample. For parabolic or elliptical reflective collectors, critical illumination can utilize up to 85% of the light emitted by the source.

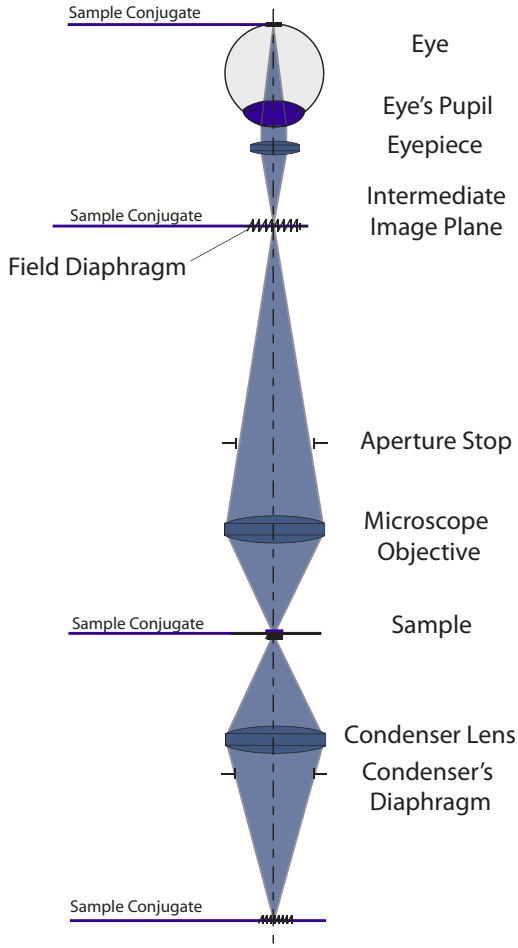

Stereo Microscopes

Stereo microscopes are built to provide depth perception, which is important for applications like micro-assembly and biological and surgical imaging. Two primary stereo-microscope approaches involve building two separate tilted systems, or using a common objective combined with a binocular system.

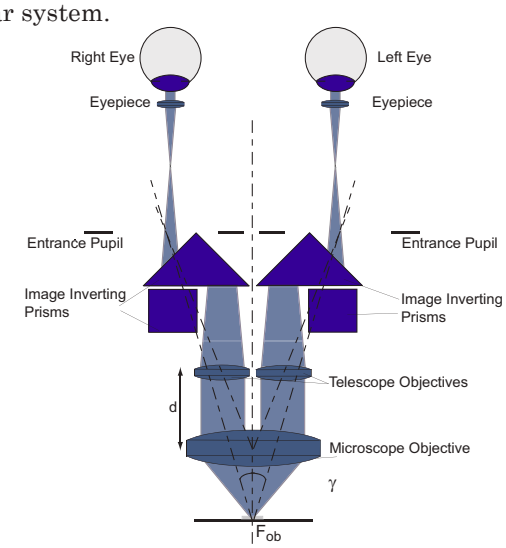

In the latter, the angle of convergence γ of a stereo microscope depends on the focal length of a microscope objective and a distance d between the microscope objective and telescope objectives. The entrance pupils of the microscope are images of the stops (located at the plane of the telescope objectives) through the microscope objective.

Depth perception δz can be defined as

$$\delta z = \frac{250\alpha_s}{M_{\text{microscope}} \tan\gamma} [\text{mm}],$$

where α_s is a **visual stereo resolving power**, which for daylight vision is approximately 5–10 arc seconds.

Stereo microscopes have a convergence angle γ in the range of 10–15 deg. Note that γ is 15 deg for visual observation and 0 for a standard microscope. For example, depth perception for the human eye is approximately 0.05 mm (at 250 mm), while for a stereo microscope of $M_{\text{microscope}} = 100$ and $\gamma = 15$ deg, it is $\delta z = 0.5$ μm.

Eyepieces

The eyepiece relays an intermediate image into the eye, corrects for some remaining objective aberrations, and provides measurement capabilities. It also needs to relay an exit pupil of a microscope objective onto the entrance pupil of the eye. The location of the relayed exit pupil is called the **eye point** and needs to be in a comfortable observation distance behind the eyepiece. The clearance between the mechanical mounting of the eyepiece and the eye point is called **eye relief**. Typical eye relief is 7–12 mm.

An eyepiece usually consists of two lenses: one (closer to the eye) which magnifies the image, and a second working as a collective lens and also responsible for the location of the exit pupil of the microscope. An eyepiece contains a field stop that provides a sharp image edge.

Parameters like **magnifying power** of an eyepiece and **field number** (FN), i.e., field of view of an eyepiece, are engraved on an eyepiece's barrel: M × / FN. Field number and magnification of a microscope objective allow quick calculation of the imaged area of a sample (see also *The Finite Tube Length Microscope* on p. 34). Field number varies with microscopy vendors and eyepiece magnification. For 10× or lower magnification eyepieces, it is usually 20–28 mm while for higher magnification, wide-angle oculars can get down to approximately 5 mm.

The majority of eyepieces are **Huygens**, **Ramsden**, or derivations of them. The Huygens eyepiece consists of two plano-convex lenses with convex surfaces facing the microscope objective:

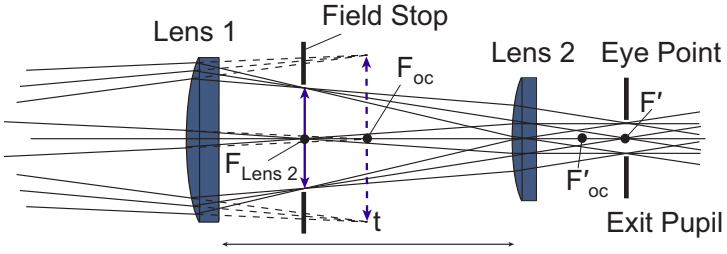

Eyepieces (cont.)

Both lenses are usually made with crown glass and allow for some correction of lateral color, coma, and astigmatism. To correct for color, the following conditions should be met:

$$f_1 \approx f_2 \quad \text{and} \quad t \approx 1.5 f_2 .$$

For higher magnifications, the exit pupil moves toward the ocular, making observation less convenient, and so this eyepiece is used only for low magnifications (≤10×). In addition, Huygens oculars effectively correct for chromatic aberrations, and therefore can be more effectively used with lower-end microscope objectives (e.g., achromatic).

The Ramsden eyepiece consists of two plano-convex lenses with convex surfaces facing each other. Both focal lengths are very similar, and the distance between lenses is smaller than f_2:

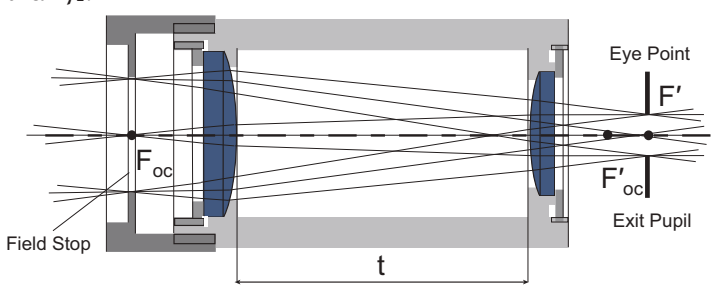

The field stop is placed in the front focal plane of the eyepiece. A collective lens does not participate in creating an intermediate image, so the Ramsden eyepiece works as a simple magnifier:

$$f_1 = f_2 \quad \text{and} \quad t < f_2 .$$

Compensating eyepieces work in conjunction with microscope objectives to correct for lateral color (apochromatic).

High-eye point oculars provide an extended distance between the last mechanical surface and the exit pupil of the eyepiece. They allow users with glasses to comfortably use a microscope. The convenient high eye-point location should be 20–25 mm behind the eyepiece.

Nomenclature and Marking of Objectives

Objective parameters include:

Objective correction—such as Achromat, Plan Achromat, Fluor, Plan Fluor, Apochromat (Apo), and Plan Apochromat.

Magnification—the lens magnification for a finite tube length, or for a microscope objective in combination with a tube lens. In infinity-corrected systems, magnification depends on the ratio between the focal lengths of a tube lens and a microscope objective. A particular objective type should be used only in combination with the proper tube lens.

Magnification	Zeiss Code
1×, 1.25×	Black
2.5×	Khaki
4×, 5×	Red
6.3×	Orange
10×	Yellow
16×, 20×, 25×, 32×	Green
25×, 32×	Green
40×, 50×	Light Blue
63×	Dark Blue
> 100×	White

Application—specialized use or design of an objective, e.g., H (bright field), D (dark field), DIC (differential interference contrast), RL (reflected light), PH (phase contrast), or P (polarization).

Tube length—an infinity corrected (∞) or finite tube length in mm.

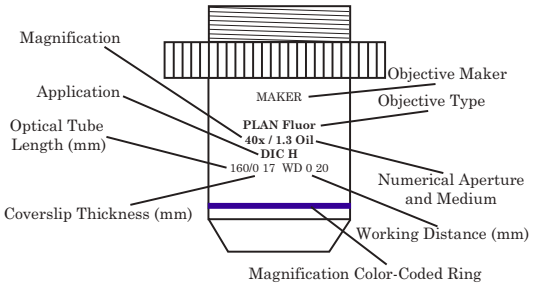

Cover slip—the thickness of the cover slip used (in mm). "0" or "–" means no cover glass or the cover slip is optional, respectively.

Numerical aperture (NA) and medium—defines system throughput and resolution. It depends on the media between the sample and objective. The most common media are air (no marking), oil (Oil), water (W), or Glycerine.

Working distance (WD)—the distance in millimeters between the surface of the front objective lens and the object.

Objective Designs

Achromatic objectives (also called **achromats)** are corrected at two wavelengths: 656 nm and 486 nm. Also, spherical aberration and coma are corrected for green (546 nm), while astigmatism is very low. Their major problems are a secondary spectrum and field curvature.

Achromatic objectives are usually built with a lower NA (below 0.5) and magnification (below 40×). They work well for white light illumination or single wavelength use. When corrected for field curvature, they are called plan-achromats.

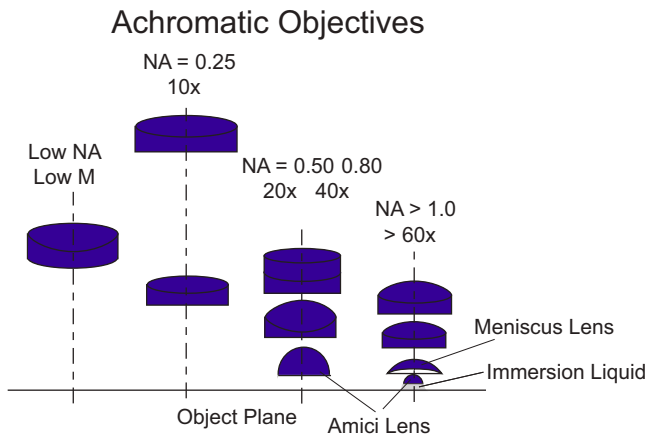

Fluorites or **semi-apochromats** have similar color correction as **achromats**; however, they correct for spherical aberration for two or three colors. The name "fluorites" was assigned to this type of objective due to the materials originally used to build this type of objective. They can be applied for higher NA (e.g., 1.3) and magnifications, and used for applications like immunofluorescence, polarization, or differential interference contrast microscopy.

The most advanced microscope objectives are **apochromats**, which are usually chromatically corrected for three colors, with spherical aberration correction for at least two colors. They are similar in construction to fluorites but with different thicknesses and surface figures. With the correction of field curvature, they are called plan-apochromats. They are used in fluorescence microscopy with multiple dyes and can provide very high NA (1.4). Therefore, they are suitable for low-light applications.

Objective Designs (cont.)

Apochromatic Objectives

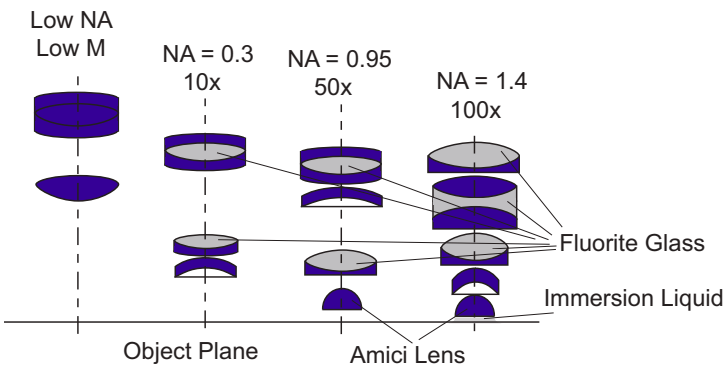

Type	Number of wavelengths for spherical correction	Number of colors for chromatic correction
Achromat	1	2
Fluorite	2–3	2–3
Plan-Fluorite	2–4	2–4
Plan-Apochromat	2–4	3–5

(Adapted from Murphy, 2001 and http://www.microscopyu.com/)

Examples of typical objective parameters are shown in the table below:

M	Type	Medium	*WD*	NA	*d*	DOF
10	Achromat	Air	4.4	0.25	1.34	8.80
20	Achromat	Air	0.53	0.45	0.75	2.72
40	Fluorite	Air	0.50	0.75	0.45	0.98
40	Fluorite	Oil	0.20	1.30	0.26	0.49
60	Apochromat	Air	0.15	0.95	0.35	0.61
60	Apochromat	Oil	0.09	1.40	0.24	0.43
100	Apochromat	Oil	0.09	1.40	0.24	0.43
Refractive index of oil is $n = 1.515$						

(Adapted from Murphy, 2001)

Special Objectives and Features

Special types of objectives include long working distance objectives, ultra-low-magnification objectives, water-immersion objectives, and UV lenses.

Long working distance objectives allow imaging through thick substrates like culture dishes. They are also developed for interferometric applications in Michelson and Mirau configurations. Alternatively, they can allow for the placement of instrumentation (e.g., micropipettes) between a sample and an objective. To provide a long working distance and high NA, a common solution is to use reflective objectives, or extend the working distance of a standard microscope objective by using a reflective attachment. While reflective objectives have the advantage of being free of chromatic aberrations, their serious drawback is a smaller field of view relative to refractive objectives. The central obscuration also decreases throughput by 15%–20%.

LWD Reflective Objective

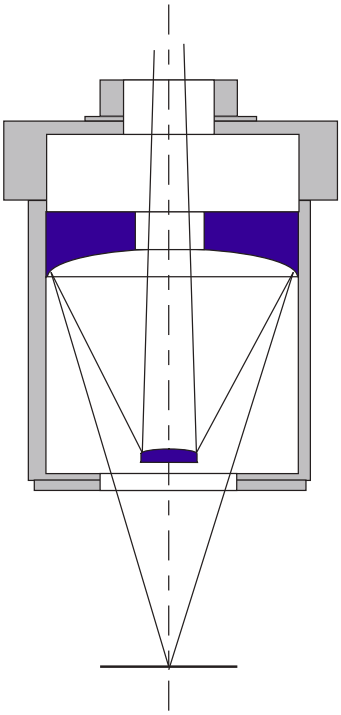

Special Objectives and Features (cont.)

Low magnification objectives can achieve magnifications as low as 0.5. However, in some cases they are not fully compatible with microscopy systems, and may not be telecentric in the image space. They also often require special tube lenses and special condensers to provide Köhler illumination.

Water immersion objectives are increasingly common, especially for biological imaging, because they provide a high NA and avoid toxic immersion oils. They usually work without a cover slip.

UV objectives are made using UV transparent, low-dispersion materials such as quartz. These objectives enable imaging at wavelengths from the near-UV through the visible spectrum, e.g., 240–700 nm. Reflective objectives can also be used for the IR bands.

Reflective Adapter to extend working distance WD

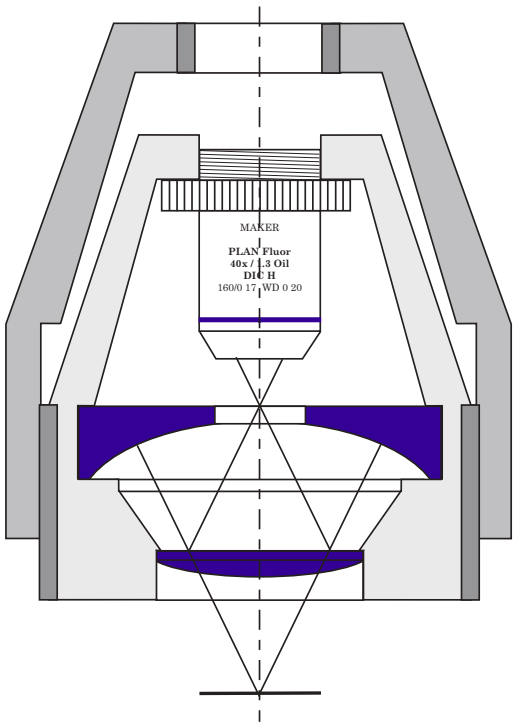

Special Lens Components

The **Bertrand lens** is a focusable eyepiece telescope that can be easily placed in the light path of the microscope. This lens is used to view the back aperture of the objective, which simplifies microscope alignment, specifically when setting up Köhler illumination.

To construct a high-numerical-aperture microscope objective with good correction, numerous optical designs implement the **Amici lens** as a first component of the objective. It is a thick lens placed in close proximity to the sample. An Amici lens usually has a plane (or nearly plane) first surface and a large-curvature spherical second surface. To ensure good chromatic aberration correction, an achromatic lens is located closely behind the Amici lens. In such configurations, microscope objectives usually reach an NA of 0.5–0.7.

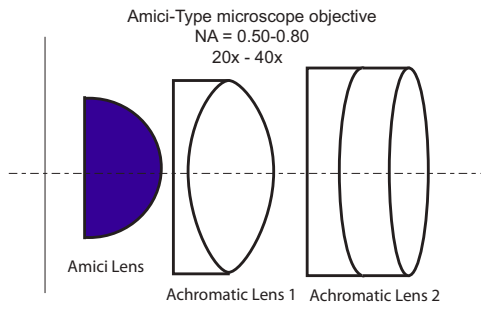

To further increase the NA, an Amici lens is improved by either cementing a high-refractive-index meniscus lens to it or placing a meniscus lens closely behind the Amici lens. This makes it possible to construct well-corrected high magnification (100×) and high numerical aperture (NA > 1.0) oil-immersion objective lenses.

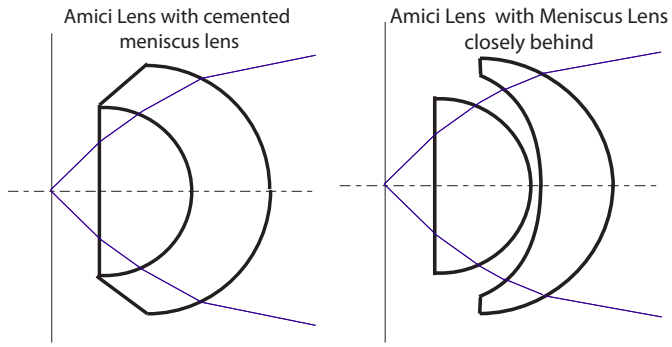

Cover Glass and Immersion

The **cover slip** is located between the object and the microscope objective. It protects the imaged sample and is an important element in the optical path of the microscope. Cover glass can reduce imaging performance and cause spherical aberration, while different imaging angles experience movement of the object point along the optical axis toward the microscope objective. The object point moves closer to the objective with an increase in angle.

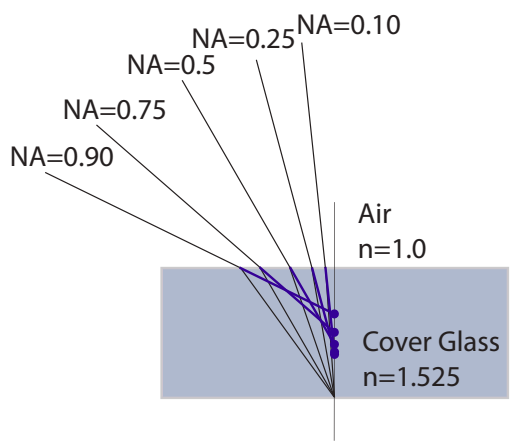

The importance of the cover glass increases in proportion to the objective's NA, especially in high-NA dry lenses. For lenses with an NA smaller than 0.5, the type of cover glass may not be a critical parameter, but should always be properly used to optimize imaging quality. Cover glasses are likely to be encountered when imaging biological samples. Note that the presence of a standard-thickness cover glass is accounted for in the design of high-NA objectives. The important parameters that define a cover glass are its thickness, index of refraction, and Abbe number. The ISO standards for refractive index and Abbe number are $n = 1.5255 \pm 0.0015$ and $V = 56 \pm 2$, respectively.

Microscope objectives are available that can accommodate a range of cover-slip thicknesses. An adjustable collar allows the user to adjust for cover slip thickness in the range from 100 microns to over 200 microns.

Cover Glass Grade Number	Thickness (in microns)
0	83–130
1	130–160
1.5	160–190
2	190–250

Cover Glass and Immersion (cont.)

The table below presents a summary of acceptable cover glass thickness deviations from 0.17 mm and the allowed thicknesses for Zeiss air objectives with different NAs.

NA of the Objective	Allowed Thickness Deviations (from 0.17 mm)	Allowed Thickness Range
<0.3	–	0.000–0.300
0.30–0.45	0.07	0.100–0.240
0.45–0.55	0.05	0.120–0.220
0.55–0.65	0.03	0.140–0.200
0.65–0.75	0.02	0.150–0.190
0.75–0.85	0.01	0.160–0.180
0.85–0.95	0.005	0.165–0.175

(Adapted from Pluta, 1988)

To increase both the microscope's resolution and system throughput, immersion liquids are used between the cover-slip glass and the microscope objective, or directly between the sample and objective. An **immersion liquid** effectively increases the NA and decreases the angle of refraction at the interface between the medium and the microscope objective. Common immersion liquids include oil (n = 1.515), water (n = 1.34), and glycerin (n = 1.48).

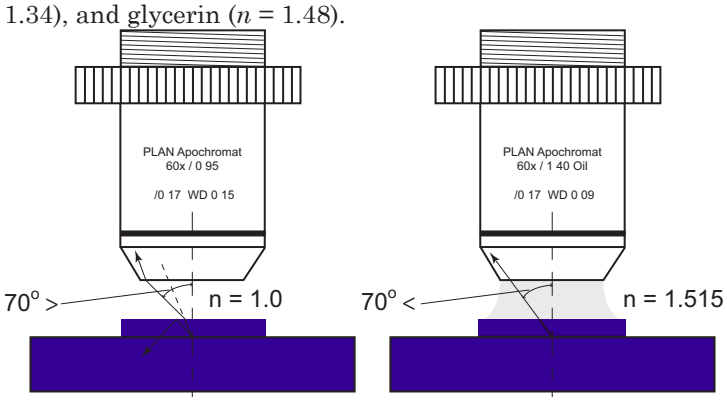

Water, which is more common, or glycerin immersion objectives are mainly used for biological samples, such as living cells or a tissue culture. They provide a slightly lower NA than oil objectives but are free of the toxicity associated with immersion oils. Water immersion is usually applied without a cover slip. For fluorescent applications it is also critical to use non-fluorescent oil.

Common Light Sources for Microscopy

Common light sources for microscopy include **incandescent lamps**, such as tungsten-argon and tungsten (e.g., quartz halogen) lamps. A **tungsten-argon lamp** is primarily used for bright-field, phase-contrast, and some polarization imaging. **Halogen lamps** are inexpensive and a convenient choice for a variety of applications that require a continuous and bright spectrum.

Arc lamps are usually brighter than incandescent lamps. The most common examples include xenon (XBO) and mercury (HBO) arc lamps. Arc lamps provide high-quality monochromatic illumination if combined with the appropriate filter. Arc lamps are more difficult to align, are more expensive, and have a shorter lifetime. Their spectral range starts in the UV range and continuously extends through visible to the infrared. About 20% of the output is in the visible spectrum, while the majority is in the UV and IR. The usual lifetime is 100–200 hours.

Common Mercury Lines	
Wavelength [nm]	**Color**
254	Far UV
366	Near UV
405	Violet
435	Deep Blue
546	Yellow-green
578	Yellow doublet

Another light source is the **gas-arc discharge lamp**, which includes mercury, xenon, and halide lamps. The **mercury lamp** has several strong lines, which might be 100 times brighter than the average output. About 50% is located in the UV range, and should be used with protective glasses. For imaging biological samples, the proper selection of filters is necessary to protect living cell samples and micro-organisms (e.g., UV-blocking filters/cold mirrors). The xenon arc lamp has a uniform spectrum, can provide an output power greater than 100 W, and is often used for fluorescence imaging. However, over 50% of its power falls into the IR; therefore, IR-blocking filters (hot mirrors) are necessary to prevent the overheating of samples.

Metal halide lamps were recently introduced as high-power sources (over 150 W) with lifetimes several times longer than arc lamps. While in general the metal-halide lamp has a spectral output similar to that of a mercury arc lamp, it extends further into longer wavelengths.

LED Light Sources

Light-emitting diodes (LEDs) are a new, alternative light source for microscopy applications. The LED is a semiconductor diode that emits photons when in forward-biased mode. Electrons pass through the depletion region of a p-n junction and lose an amount of energy equivalent to the bandgap of the semiconductor. The characteristic features of LEDs include a long lifetime, a compact design, and high efficiency. They also emit narrowband light with relatively high energy.

Wavelengths [nm] of High-power LEDs Commonly Used in Microscopy		Total Beam Power [mW] (approximate)
455	Royal Blue	225–450
470	Blue	200–400
505	Cyan	150–250
530	Green	100–175
590	Amber	15–25
633	Red	25–50
435–675	White Light	200–300

An important feature of LEDs is the ability to combine them into arrays and custom geometries. Also, LEDs operate at lower temperatures than arc lamps, and due to their compact design they can be cooled easily with simple heat sinks and fans.

The radiance of currently available LEDs is still significantly lower than that possible with arc lamps; however, LEDs can produce an acceptable fluorescent signal in bright microscopy applications. Also, the pulsed mode can be used to increase the radiance by 20 times or more.

LED Spectral Range [nm]	Semiconductor
350–400	GaN
400–550	$In_{1-x}Ga_xN$
550–650	$Al_{1-x-y}In_yGa_xP$
650–750	$Al_{1-x}Ga_xAs$
750–1000	$GaAs_{1-x}P_x$

Filters

Neutral density (ND) filters are neutral (wavelength wise) gray filters scaled in units of **optical density** (OD):

$$OD = \log_{10}\left(\frac{1}{\tau}\right),$$

where τ is the transmittance. ND filters can be combined to provide OD as a sum of the ODs of the individual filters. ND filters are used to change light intensity without tuning the light source, which can result in a spectral shift.

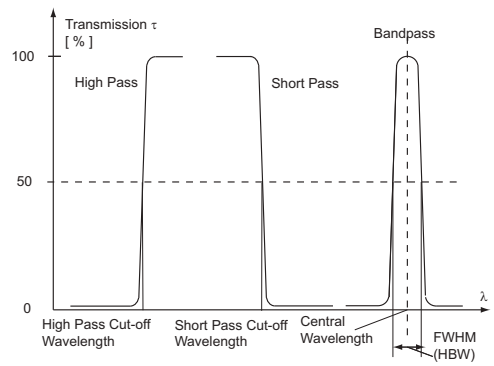

Color **absorption filters** and **interference filters** (also see *Multiple Wave Interference* on page 16) isolate the desired range of wavelengths: e.g. bandpass and edge filters.

Edge filters include short-pass and long-pass filters. Short-pass filters allow short wavelengths to pass and stop long wavelengths, and long-pass filters allow long wavelengths to pass while stopping short wavelengths. Edge filters are defined for wavelengths with a 50% drop in transmission.

Bandpass filters allow only a selected spectral bandwidth to pass and are characterized with a central wavelength and full-width-half-maximum (FWHM) defining the spectral range for transmission of at least 50%.

Color absorption glass filters usually serve as broad bandpass filters. They are less costly and less susceptible to damage than interference filters.

Interference filters are based on multiple beam interference in thin films. They combine between three to over 20 dielectric layers of $\lambda/2$ and $\lambda/4$, separated by metallic coatings. They can provide a sharp bandpass transmission down to the sub-nm range and full-width-half-maximum of 10–20 nm.

Polarizers and Polarization Prisms

Polarizers are built with birefringent crystals using polarization at multiple reflections, selective absorption (dichroism), form birefringance, or scattering.

For example, polarizers built with birefringent crystals (Glan-Thompson prisms) use the principle of total internal reflection to eliminate ordinary or extraordinary components (for positive or negative crystals).

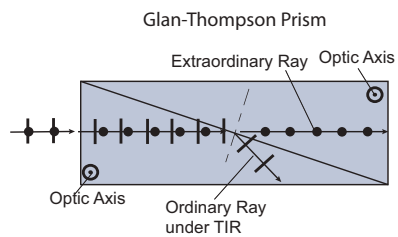

Birefringent prisms are crucial components for numerous microscopy techniques, e.g., differential interference contrast. The most common birefringent prism is a **Wollaston prism**. It splits light into two beams with orthogonal polarization and propagates under different angles.

The angle between propagating beams is

$$\varepsilon = 2\left|n_{\mathrm{e}} - n_{\mathrm{o}}\right| \tan\alpha \,.$$

Both beams produce interference with fringe period b:

$$b = \frac{\lambda}{\varepsilon} = \frac{\lambda}{2\left|n_{\mathrm{e}} - n_{\mathrm{o}}\right| \tan\alpha} \,.$$

The localization plane of fringes is

$$\gamma = \frac{1}{2}\left(\frac{1}{n_{\mathrm{e}}} + \frac{1}{n_{\mathrm{o}}}\right)\alpha \,.$$

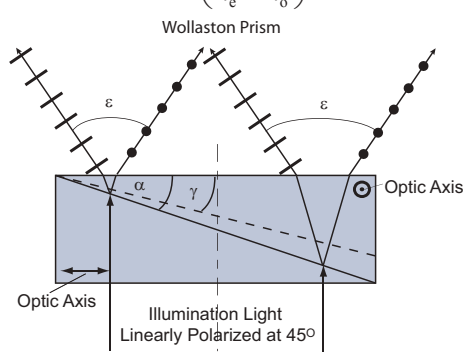

Polarizers and Polarization Prisms (cont.)

The fringe localization plane tilt can be compensated by using two symmetrical Wollaston prisms.

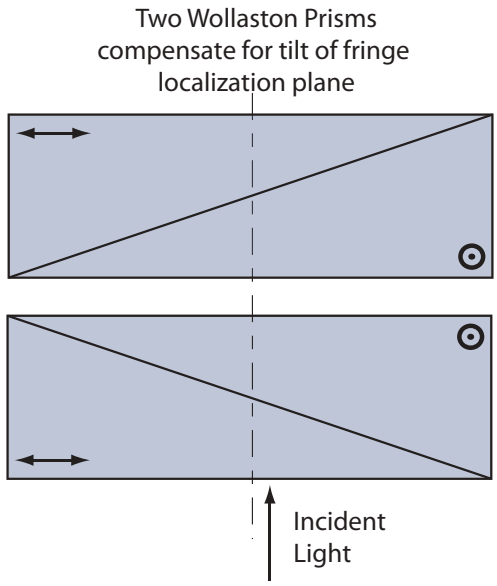

Wollaston prisms have a fringe localization plane inside the prism. One example of a modified Wollaston prism is the **Nomarski prism**, which simplifies the DIC microscope setup by shifting the plane outside the prism, i.e., the prism does not need to be physically located at the condenser front focal plane or the objective's back focal plane.

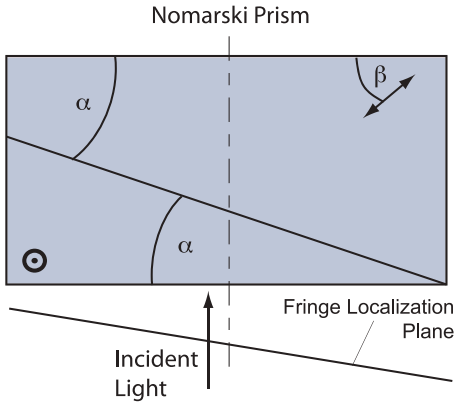

Amplitude and Phase Objects

The major object types encountered on the microscope are **amplitude** and **phase** objects. The type of object often determines the microscopy technique selected for imaging.

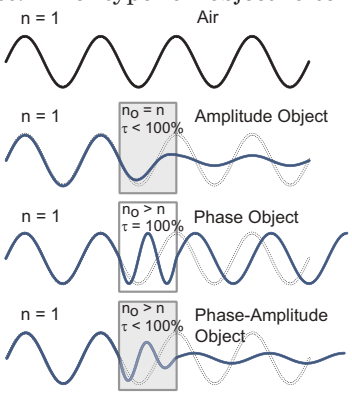

The amplitude object is defined as one that changes the amplitude and therefore the intensity of transmitted or reflected light. Such objects are usually imaged with bright-field microscopes. A stained tissue slice is a common amplitude object.

Phase objects do not affect the optical intensity; instead, they generate a phase shift in the transmitted or reflected light. This phase shift usually arises from an inhomogeneous refractive index distribution throughout the sample, causing differences in optical path length (OPL). Mixed amplitude-phase objects are also possible (e.g., biological media), which affect the amplitude and phase of illumination in different proportions.

Classifying objects as **self-luminous** and **non-self-luminous** is another way to categorize samples. Self-luminous objects generally do not directly relate their amplitude and phase to the illuminating light. This means that while the observed intensity may be proportional to the illumination intensity, its amplitude and phase are described by a statistical distribution (e.g., fluorescent samples). In such cases, one can treat discrete object points as secondary light sources, each with their own amplitude, phase, coherence, and wavelength properties.

Non-self-luminous objects are those that affect the illuminated beam in such a manner that discrete object points cannot be considered as entirely independent. In such cases, the wavelength and temporal coherence of the illuminating source needs to be considered in imaging. A diffusive or absorptive sample is an example of such an object.

The Selection of a Microscopy Technique

Microscopy provides several imaging principles. Below is a list of the most common techniques and object types:

Technique	Type of sample
Bright-field	Amplitude specimens, reflecting specimens, diffuse objects
Dark-field	Light-scattering objects
Phase contrast	Phase objects, light-scattering objects, light-refracting objects, reflective specimens
Differential interference contrast (DIC)	Phase objects, light-scattering objects, light-refracting objects, reflective specimens
Polarization microscopy	Birefringent specimens
Fluorescence microscopy	Fluorescent specimens
Laser scanning, Confocal microscopy, and Multi-photon microscopy	3D samples requiring optical sectioning, fluorescent and scattering samples
Super-resolution microscopy (**RESOLFT**, **4Pi**, **I^{5}M**, **SI**, **STORM**, **PALM**, and others)	Imaging at the molecular level; imaging primarily focuses on fluorescent samples where the sample is a part of an imaging system
Raman microscopy, **CARS**	Contrast-free chemical imaging
Array microscopy	Imaging of large FOVs
SPIM	Imaging of large 3D samples
Interference Microscopy	Topography, refractive index measurements, 3D coherence imaging

All of these techniques can be considered for transmitted or reflected light. Below are examples of different sample types:

Sample type	Sample Example
Amplitude specimens	Naturally colored specimens, stained tissue
Specular specimens	Mirrors, thin films, metallurgical samples, integrated circuits
Diffuse objects	Diatoms, fibers, hairs, micro-organisms, minerals, insects
Phase objects	Bacteria, cells, fibers, mites, protozoa
Light-refracting samples	Colloidal suspensions, minerals, powders
Birefringent specimens	Mineral sections, crystallized chemicals, liquid crystals, fibers, single crystals
Fluorescent specimens	Cells in tissue culture, fluorochrome-stained sections, smears and spreads

(Adapted from www.microscopyu.com)

Image Comparison

The images below display four important microscope modalities: **bright field**, **dark field**, **phase contrast**, and **differential interference contrast**. They also demonstrate the characteristic features of these methods. The images are of a blood specimen viewed with the Zeiss upright Axiovert Observer Z1 microscope. The microscope was set with an objective at 40×, NA = 0.6, Ph2, LD Plan Neofluor, and a cover slip glass of 0.17 mm. The pictures were taken with a monochromatic CCD camera.

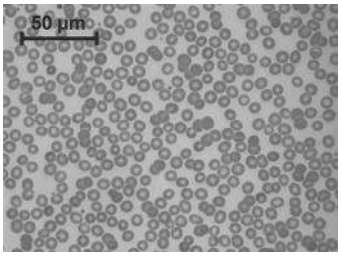

Bright Field

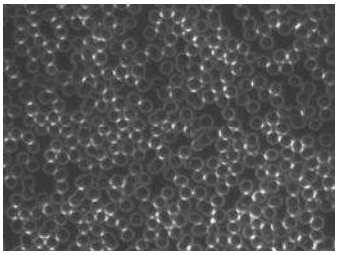

Dark Field

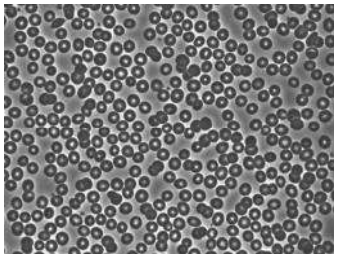

Phase Contrast

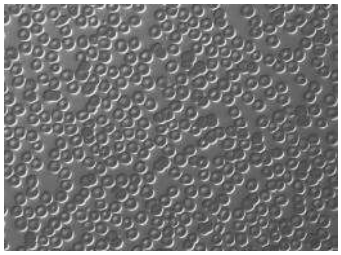

Differential Interference Contrast (DIC)

The bright-field image relies on absorption and shows the sample features with decreasing amounts of passing light. The dark-field image only shows the scattering sample components. Both the phase contrast and the differential interference contrast demonstrate the optical thickness of the sample. The characteristic halo effect is visible in the phase-contrast image. The 3D effect of the DIC image arises from the differential character of images; they are formed as a derivative of the phase changes in the beam as it passes through the sample.

Phase Contrast

Phase contrast is a technique used to visualize phase objects by phase and amplitude modifications between the direct beam propagating through the microscope and the beam diffracted at the phase object. An object is illuminated with monochromatic light, and a phase-delaying (or advancing) element in the aperture stop of the objective introduces a phase shift, which provides interference contrast. Changing the amplitude ratio of the diffracted and non-diffracted light can also increase the contrast of the object features.

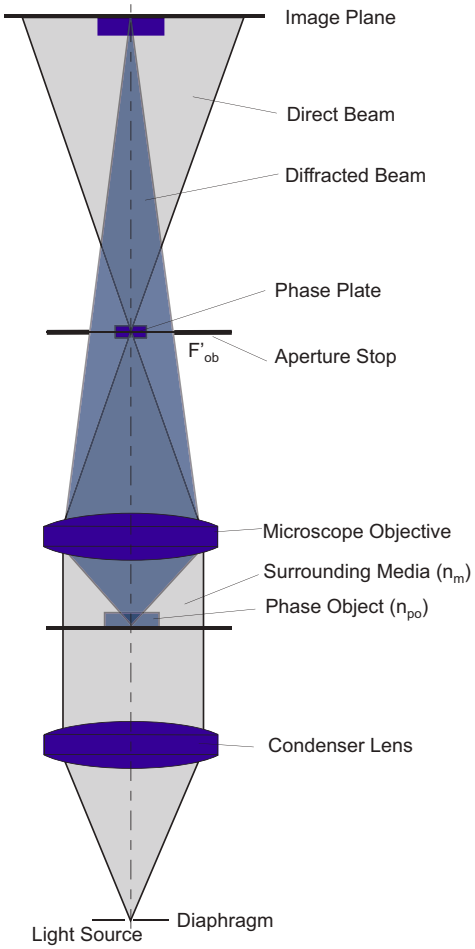

Phase Contrast (cont.)

Phase contrast is primarily used for imaging living biological samples (immersed in a solution), unstained samples (e.g., cells), thin film samples, and mild phase changes from mineral objects. In that regard, it provides qualitative information about the optical thickness of the sample. Phase contrast can also be used for some quantitative measurements, like an estimation of the refractive index or the thickness of thin layers. Its phase detection limit is similar to the **differential interference contrast** and is approximately $\lambda/100$. On the other hand, phase contrast (contrary to DIC) is limited to thin specimens, and phase delay should not exceed the depth of field of an objective. Other drawbacks compared to DIC involve its limited optical sectioning ability and undesired effects like halos or shading-off. (See also *Characteristic Features of Phase Contrast Microscopy* on page 71).

The most important advantages of phase contrast (over DIC) include its ability to image birefringent samples and its simple and inexpensive implementation into standard microscopy systems.

Presented below is a mathematical description of the phase contrast technique based on a vector approach. The phase shift in the figure (See page 68) is represented by the orientation of the vector. The length of the vector is proportional to amplitude of the beam. When using standard imaging on a transparent sample, the length of the light vectors passing through sample *PO* and surrounding media *SM* is the same, which makes the sample invisible. Additionally, vector *PO* can be considered as a sum of the vectors passing through surrounding media *SM* and diffracted at the object *DP*.

$$PO = SM + DP$$

If the wavefront propagating through the surrounding media can be the subject of an exclusive phase change (diffracted light *DP* is not affected), the vector *SM* is rotated by an angle corresponding to the phase change. This exclusive phase shift is obtained with a small circular or ring phase plate located in the plane of the aperture stop of a microscope.

Phase Contrast (cont.)

Consequently, vector *PO*, which represents light passing through a phase sample, will change its value to *PO′* and provide contrast in the image:

$$PO' = SM' + DP,$$

where *SM′* represents the rotated vector *SM*.

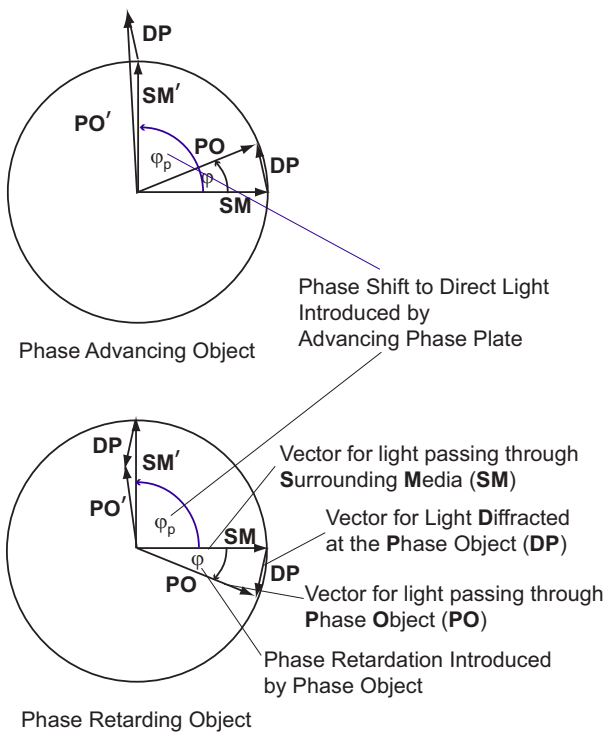

Phase samples are considered to be **phase retarding objects** or **phase advancing objects** when their refractive index is greater or less than the refractive index of the surrounding media, respectively.

Phase plate	Object Type	Object Appearance
$\varphi_p = +\pi/2$ (+90°)	phase-retarding	brighter
$\varphi_p = +\pi/2$ (+90°)	phase-advancing	darker
$\varphi_p = -\pi/2$ (–90°)	phase-retarding	darker
$\varphi_p = -\pi/2$ (–90°)	phase-advancing	brighter

Visibility in Phase Contrast

Visibility of features in **phase contrast** can be expressed as

$$C_{\text{ph}} = \frac{I_{\text{media}} - I_{\text{object}}}{I_{\text{media}}} = \frac{|SM'|^2 - |PO'|^2}{|SM'|^2}.$$

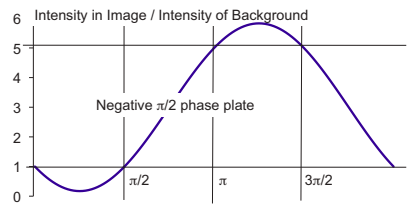

This equation defines the phase object's visibility as a ratio between the intensity change due to phase features and the intensity of surrounding media $|SM'|^2$. It defines the negative and positive contrast for an object appearing brighter or darker than the media, respectively. Depending on the introduced phase shift (positive or negative), the same object feature may appear with a **negative** or **positive contrast**. Note that C_{ph} relates to classical contrast C as

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{|I_1 - I_2|}{I_1 + I_2} = |C_{\text{ph}}| \frac{|SM'|^2}{|SM'|^2 + |PO'|^2}.$$

Contrast	Phase Plate
-2φ	$\varphi_p = +\pi/2$ (+90°)
$+2\varphi$	$\varphi_p = -\pi/2$ (–90°)

For phase changes in the 0–2π range, intensity in the image can be found using vector relations; small phase changes in the object $\varphi \ll 90$ deg can be approximated as $\pm 2\varphi$.

To increase the contrast of images, the intensity in the direct beam is additionally changed with beam attenuation τ in the phase ring. τ is defined as a transmittance, $\tau = 1/N$, where N is a dividing coefficient of intensity in a direct beam (intensity is decreased N times). The contrast in this case is

$$C_{\text{ph}} = -2\varphi / \sqrt{\tau} = -2\varphi\sqrt{N}$$

for a $+\pi/2$ phase plate, and

$$C_{\text{ph}} = +2\varphi / \sqrt{\tau} = +2\varphi\sqrt{N}$$

for $-\pi/2$ phase plate. The minimum perceived phase difference with phase contrast is

$$\Delta\varphi_{\min} = \frac{C_{\text{ph-min}}\lambda}{4\pi\sqrt{N}}.$$

$C_{\text{ph-min}}$ is usually accepted at the contrast value of 0.02.

The Phase Contrast Microscope

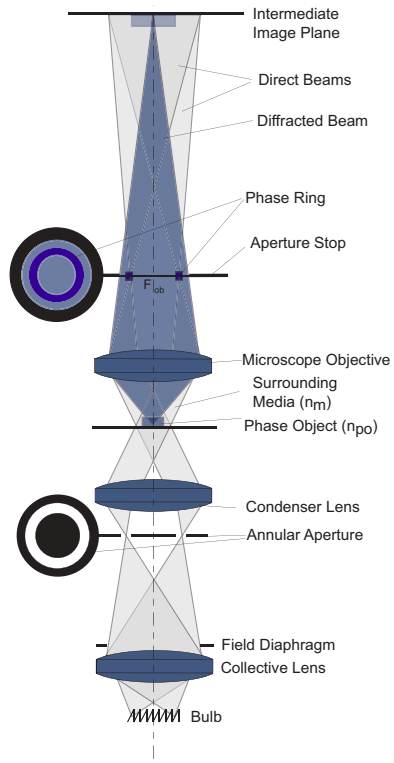

The common phase-contrast system is similar to the bright-field microscope but with two modifications:

1. The condenser diaphragm is replaced with an annular aperture diaphragm.
2. A ring-shaped phase plate is introduced at the back focal plane of the microscope objective. The ring may include metallic coatings, typically providing 75% to 90% attenuattion of the direct beam.

Both components are in conjugate planes and are aligned such that the image of the annular condenser diaphragm overlaps with the objective phase ring. While there is no direct access to the objective phase ring, the anular aperture in front of the condenser is easily accessible.

The phase shift introduced by the ring is given by

$$\varphi_p = 2\pi \frac{OPD}{\lambda} = \frac{2\pi}{\lambda}\left(n_m - n_r\right)t\,,$$

where n_m and n_r are refractive indices of the media surrounding the phase ring and the ring itself, respectively. t is the physical thickness of the phase ring.

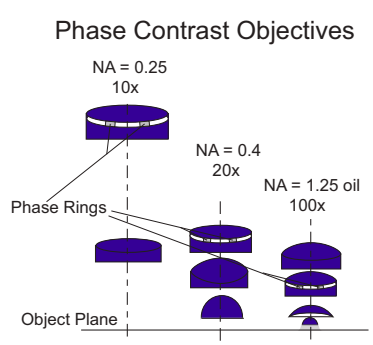

Characteristic Features of Phase Contrast

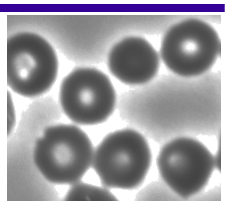

Images in phase contrast are dark or bright features on a background (positive and negative contrast, respectively). They contain undesired image effects called halo and shading-off, which are a result of the incomplete separation of direct and diffracted light. The **halo effect** is a phase contrast feature that increases light intensity around sharp changes in the phase gradient.

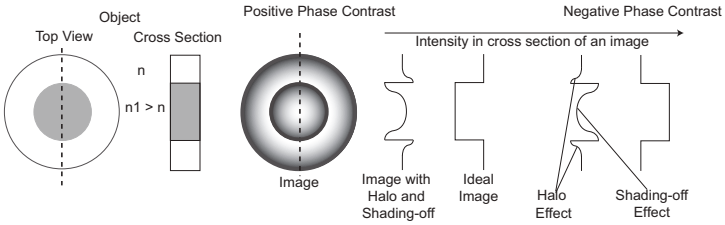

The **shading-off effect** is an increase or decrease (for dark or bright images, respectively) of the intensity of the phase sample feature.

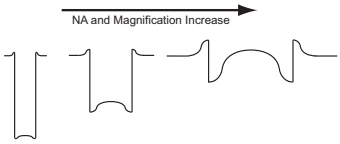

Both effects strongly increase with an increase in numerical aperture and magnification. They can be reduced by surrounding the sides of a phase ring with ND filters.

Lateral resolution of the phase contrast technique is affected by the annular aperture (the radius of phase ring r_{PR}) and the aperture stop of the objective (the radius of aperture stop is r_{AS}). It is

$$d = f'_{\text{objective}} \frac{\lambda}{(r_{\text{AS}} + r_{\text{PR}})},$$

compared to the resolution limit for a standard microscope:

$$d = f'_{\text{objective}} \frac{\lambda}{r_{\text{AS}}}.$$

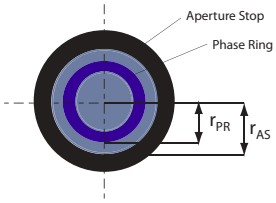

Amplitude Contrast

Amplitude contrast changes the contrast in the images of absorbing samples. It has a layout similar to phase contrast; however, there is no phase change introduced by the object. In fact, in many cases a phase-contrast microscope can increase contrast based on the principle of amplitude contrast. The visibility of images is modified by changing the intensity ratio between the direct and diffracted beams.

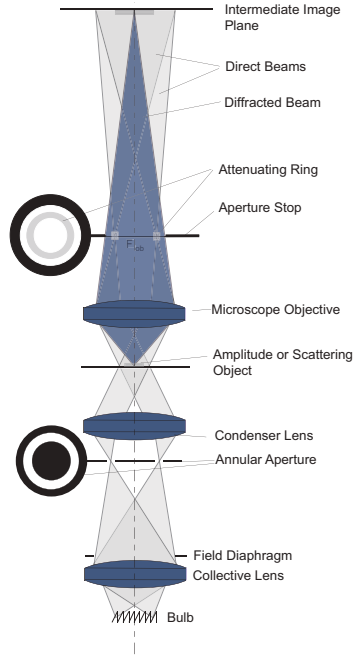

A vector schematic of the technique is as follows:

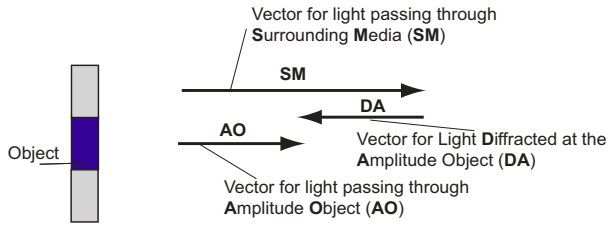

Similar to visibility in phase contrast, image contrast can be described as a ratio of the intensity change due to amplitude features surrounding the media's intensity:

$$C_{\mathrm{ac}} = \frac{I_{\mathrm{media}} - I_{\mathrm{object}}}{I_{\mathrm{media}}} = \frac{2|SM||DA| - |DA|^2}{|SM|^2}.$$

Since the intensity of diffraction for amplitude objects is usually small, the contrast equation can be approximated with $C_{\mathrm{ac}} = 2|DA|/|SM|$. If a direct beam is further attenuated, contrast C_{ac} will increase by factor of $N^{-1} = 1/\tau^{-1}$.

Oblique Illumination

Oblique illumination (also called **anaxial illumination**) is used to visualize phase objects. It creates pseudo-profile images of transparent samples. These reliefs, however, do not directly correspond to the actual surface profile.

The principle of oblique illumination can be explained by using either refraction or diffraction and is based on the fact that sample features of various spatial frequencies will have different intensities in the image due to a nonsymmetrical system layout and the filtration of spatial frequencies.

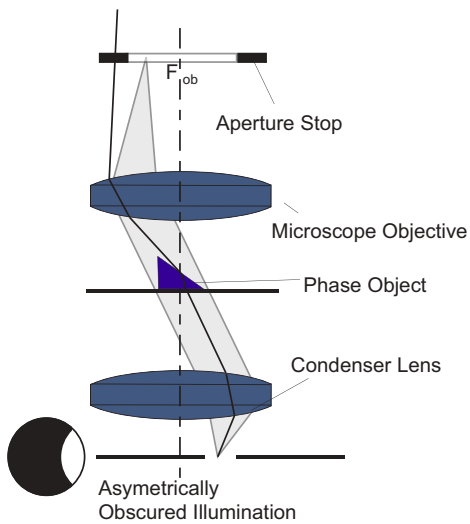

In practice, oblique illumination can be achieved by obscuring light exiting a condenser; translating the sub-stage diaphragm of the condenser lens does this easily.

The advantage of oblique illumination is that it improves spatial resolution over Köhler illumination (up to two times). This is based on the fact that there is a larger angle possible between the 0th and 1st orders diffracted at the sample for oblique illumination. In Köhler illumination, due to symmetry, the 0th and -1st orders at one edge of the stop will overlap with the 0th and $+1^{st}$ order on the other side. However, the gain in resolution is only for one sample direction; to examine the object features in two directions, a sample stage should be rotated.

Modulation Contrast

Modulation contrast microscopy (MCM) is based on the fact that phase changes in the object can be visualized with the application of multilevel gray filters.

The intensities of light refracted at the object are displayed with different values since they pass through different zones of the filter located in the stop of the microscope. MCM is often configured for oblique illumination since it already provides some intensity variations for phase objects. Therefore, the resolution of the MCM changes between normal and oblique illumination.

MCM can be obtained through a simple modification of a bright-field microscope by adding a slit diaphragm in front of the condenser and amplitude filter (also called the modulator) in the aperture stop of the microscope objective. The modulator filter consists of three areas: 100% (bright), 15% (gray), and 1% (dark). This filter acts in a similar way to the knife-edge in the Schlieren imaging techniques.

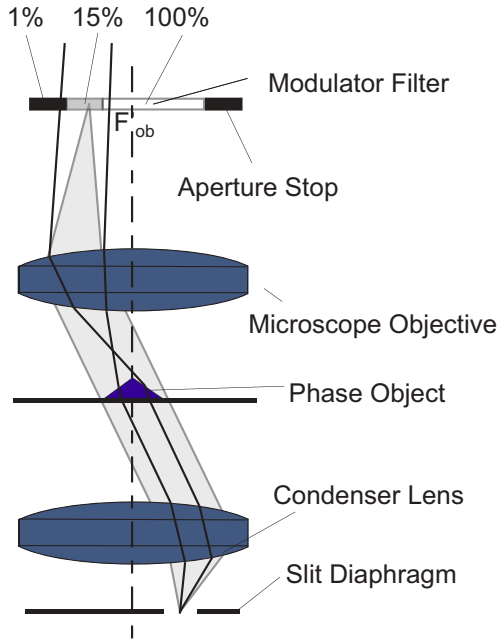

Hoffman Contrast

A specific implementation of MCM is **Hoffman modulation contrast**, which uses two additional polarizers located in front of the slit aperture. One polarizer is attached to the slit and obscures 50% of its width. A second can be rotated, which provides two intensity zones in the slit (for crossed polarizers, half of the slit is dark and half is bright; the entire slit is bright for the parallel position).

The major advantages of this technique include imaging at a full spatial resolution and a minimum depth of field, which allows optical sectioning. The method is inexpensive and easy to implement. The main drawback is that the images deliver a pseudo-relief image, which cannot be directly correlated with the object's form.

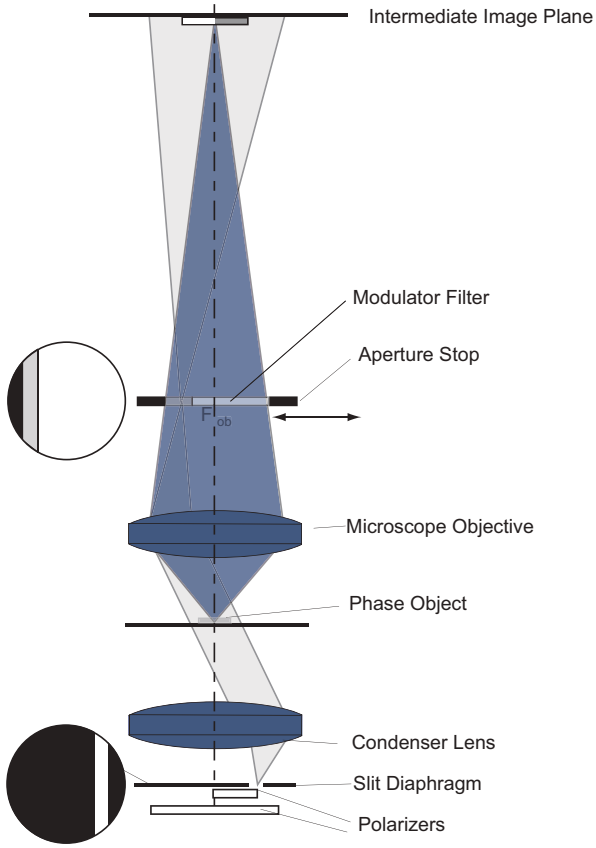

Dark Field Microscopy

In **dark-field microscopy**, the specimen is illuminated at such angles that direct light is not collected by the microscope objective. Only diffracted and scattered light are used to form the image.

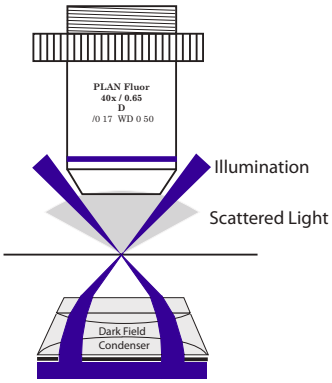

The key component of the dark-field system is the condenser. The simplest approach uses an annular condenser stop and a microscope objective with an NA below that of the illumination cone. This approach can be used for objectives with NA < 0.65. Higher-NA objectives may incorporate an adjustable iris to reduce their NA below 0.65 for dark-field imaging. To image at a higher NA, specialized reflective dark-field condensers are necessary. Examples of such condensers include the **cardioid** and **paraboloid** designs, which enable dark-field imaging at an NA up to 0.9–1.0 (with oil immersion).

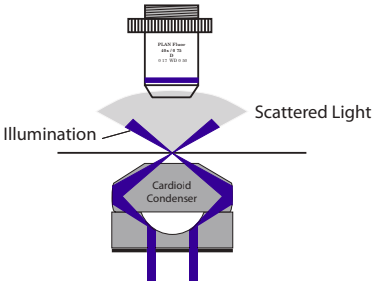

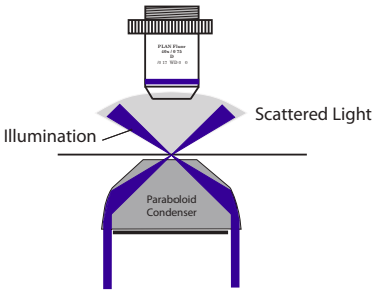

For dark-field imaging in epi-illumination, objective designs consist of external illumination, and internal imaging paths are used.

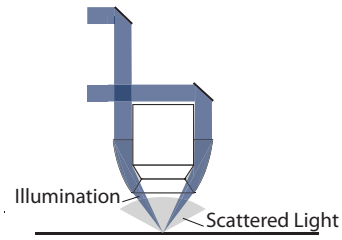

Optical Staining: Rheinberg Illumination

Optical staining is a class of techniques that uses optical methods to provide color differentiation between different areas of a sample.

Common Color Combinations	
Inner Filter	Annular Filter
Violet	Red
Blue	Yellow-Orange
Green	Red

Rheinberg illumination is a modification of dark-field illumination. Instead of an annular aperture, it uses two-zone color filters located in the front focal plane of the condenser. The overall diameter of the outer annular color filter is slightly greater than that corresponding to the NA of the system:

$$2r_{\lambda 2} > 2NAf'_{\text{condenser}} .$$

To provide good contrast between scattered and direct light, the inner filter is darker. Rheinberg illumination provides images that are a combination of two colors. Scattering features in one color are visible on the background of the other color.

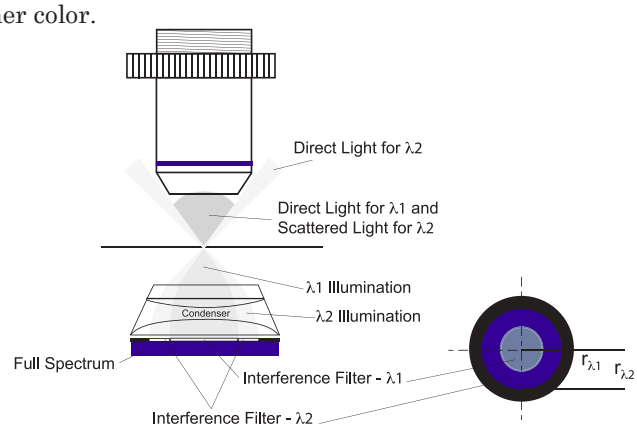

Color-stained images can also be obtained in a simplified setup where only one (inner) filter is used. For such an arrangement, images will present white-light scattering features on the colored background. Other deviations from the above technique include double illumination where transmittance and reflectance are separated into two different colors.

Optical Staining: Dispersion Staining

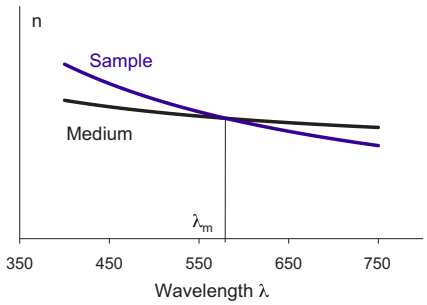

Dispersion staining is based on using highly dispersive media that matches the refractive index of the phase sample for a single wavelength λ_m (the matching wavelength). The system is configured in a setup similar to dark-field or oblique illumination. Wavelengths close to the matching wavelength will miss the microscope objective, while the other will refract at the object and create a colored dark-field image. Various configurations include illumination with a dark-field condenser and an annular aperture or central stop at the back focal plane of the microscope objective.

Color	λ Range [nm]	λ Mean [nm]
Violet	400–450	440
Blue	450–480	470
Blue-Green	480–510	495
Green	510–550	530
Green-Yellow	550–570	560
Yellow	570–590	580
Orange	590–630	605
Red	630–700	650

(Adapted from Pluta, 1989)

The annular-stop technique uses low-magnification (10×) objectives with a stop built as an opaque screen with a central opening, and is used for bright-field dispersion staining. The annular stop absorbs red and blue wavelengths and allows yellow through the system. A sample appears as white with yellow borders. The image background is also white.

The central-stop system absorbs unrefracted yellow light and direct white light. Sample features show up as purple (a mixture of red and blue) borders on a dark background. Different media and sample types will modify the color's appearance.

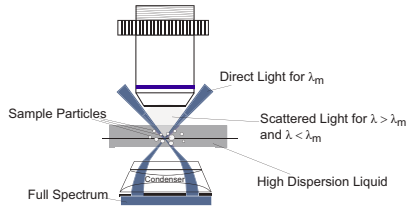

Shearing Interferometry: The Basis for DIC

Differential interference contrast (DIC) microscopy is based on the principles of shearing interferometry. Two wavefronts propagating through an optical system containing identical phase distributions (introduced by the sample) are slightly shifted to create an interference pattern. The phase of the resulting fringes is directly proportional to the derivative of the phase distribution in the object, hence the name "differential." Specifically, the intensity of the fringes relates to the derivative of the phase delay through the object (φ_o) and the local delay between wavefronts (Δ_b):

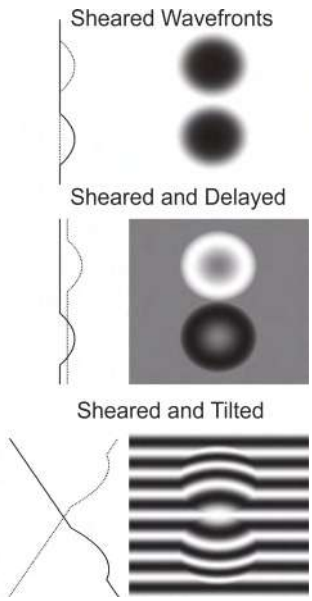

$$I \propto I_{max} \cos^2\left(\Delta_b \pm \Delta\varphi_o\right)$$

or

$$I \propto I_{max} \cos^2\left(\Delta_b \pm s\frac{d\varphi_o}{dx}\right),$$

where s denotes the shear between wavefronts, and Δ_b is the axial delay.

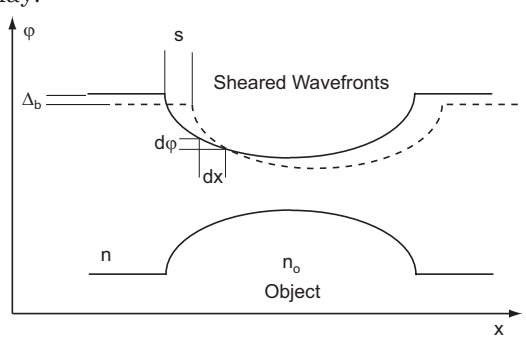

Wavefront shear is commonly introduced in DIC microscopy by incorporating a Mach-Zehnder interferometer into the system (e.g., Zeiss) or by the use of birefringent prisms. The appearance of DIC images depends on the sample orientation with respect to the shear direction.

DIC Microscope Design

The most common **differential interference contrast** (**Nomarski DIC**) design uses two **Wollaston** or **Nomarski birefringent prisms**. The first prism splits the wavefront into ordinary and extraordinary components to produce interference between these beams. Fringe localization planes for both prisms are in conjugate planes. Additionally, the system uses a crossed polarizer and analyzer. The polarizer is located in front of prism I, and the analyzer is behind prism II. The polarizer is rotated by 45 deg with regard to the shear axes of the prisms.

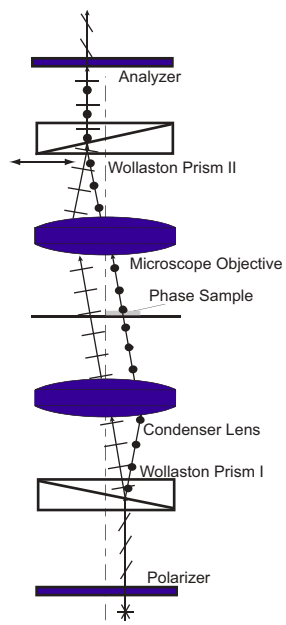

If prism II is centrally located, the intensity in the image is

$$I \propto \sin^2\left(\frac{\pi s \frac{d\varphi_o}{dx}}{\lambda}\right).$$

For a translated prism, phase **bias** Δ_b is introduced, and the intensity is proportional to

$$I \propto \sin^2 \frac{\pi\left(\Delta_b \pm s\frac{d\varphi_o}{dx}\right)}{\lambda}.$$

The sign in the equation depends on the direction of the shift. Shear s is

$$s = \frac{s'}{M_{\text{objective}}} = \varepsilon \frac{OTL}{M_{\text{objective}}},$$

where ε is the angular shear provided by the birefringent prisms, s' is the shear in the image space, and OTL is the optical tube length. The high spatial coherence of the source is obtained by the slit located in front of the condenser and oriented perpendicularly to the shear direction. To assure good interference contrast, the width of the slit w should be

$$w = \frac{\lambda f'_{\text{condenser}}}{4s}.$$

Low-strain (low-birefringence) objectives are crucial for high-quality DIC.

Appearance of DIC Images

In practice, shear between wavefronts in **differential interference contrast** is small and accounts for a fraction of a fringe period. This provides the differential character of the phase difference between interfering beams introduced to the interference equation.

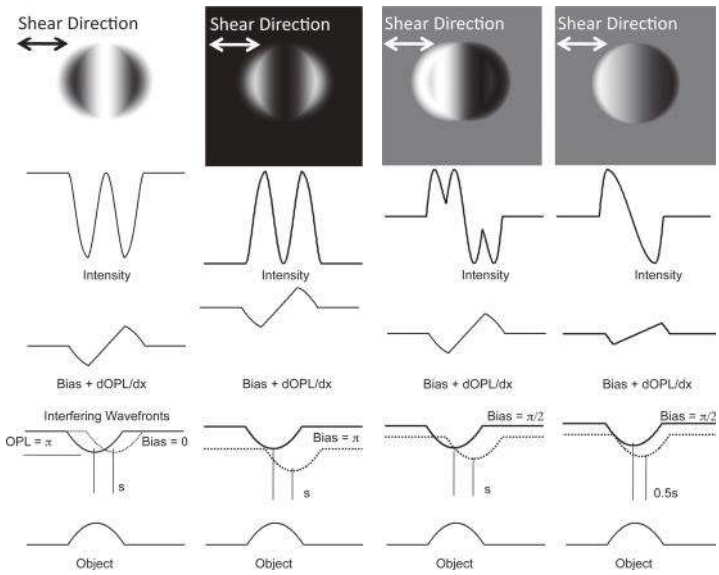

Increased shear makes the edges of the object less sharp, while increased bias introduces some background intensity. Shear direction determines the appearance of the object.

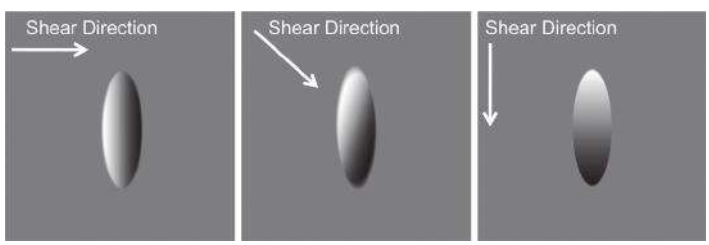

Compared to phase contrast, DIC allows for larger phase differences throughout the object and operates in full resolution of the microscope (i.e., it uses the entire aperture). The depth of field is minimized, so DIC allows optical sectioning.

Reflectance DIC

A **Nomarski interference microscope** (also called a polarization interference contrast microscope) is a reflectance DIC system developed to evaluate roughness and the surface profile of specular samples. Its applications include metallography, microelectronics, biology, and medical imaging. It uses one Wollaston or Nomarski prism, a polarizer, and an analyzer. The information about the sample is obtained for one direction parallel to the gradient in the object. To acquire information for all directions, the sample should be rotated.

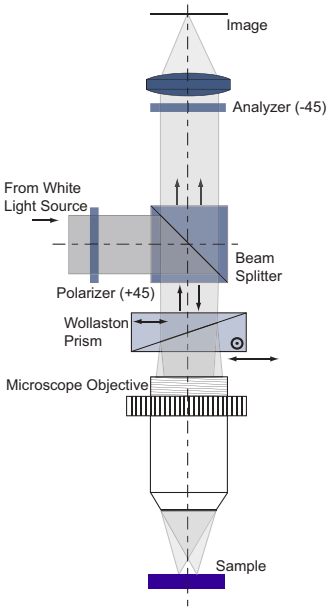

If white light is used, different colors in the image correspond to different sample slopes. It is possible to adjust the color by rotating the polarizer.

Phase delay (**bias**) can be introduced by translating the birefringent prism. For imaging under monochromatic illumination, the brightness in the image corresponds to the slope in the sample. Images with no bias will appear as bright features on a dark background; with bias they will have an average background level, and slopes will vary between dark and bright. This way, it is possible to interpret the direction of the slope. Similar analysis can be performed for the colors from white-light illumination.

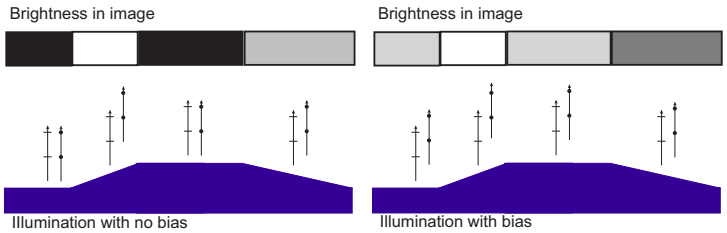

Polarization Microscopy

Polarization microscopy provides images containing information about the properties of anisotropic samples. A polarization microscope is a modified compound microscope with three unique components: the **polarizer**, **analyzer**, and **compensator**. A linear polarizer is located between the light source and the specimen, close to the condenser's diaphragm. The analyzer is a linear polarizer with its transmission axis perpendicular to the polarizer. It is placed between the sample and the eye/camera, at or near the microscope objective's aperture stop. If no sample is present, the image should appear uniformly dark. The compensator (a type of retarder) is a birefringent plate of known parameters used for quantitative sample analysis and contrast adjustment.

Quantitative information can be obtained by introducing known retardation. Rotation of the analyzer correlates the analyzer's angular position with the retardation required to minimize the light intensity for object features at selected wavelengths. **Retardation** introduced by the compensator plate can be described as

$$\Gamma = (n_e - n_o)t ,$$

where t is the sample thickness, and subscripts e and o denote the extraordinary and ordinary beams. Retardation in concept is similar to the optical path difference for beams propagating through two different materials:

$$OPD = (n_1 - n_2)t = (n_e - n_o)t .$$

The phase delay caused by sample birefringence is therefore

$$\delta = \frac{2\pi OPD}{\lambda} = \frac{2\pi\Gamma}{\lambda} .$$

A polarization microscope can also be used to determine the orientation of the optic axis.

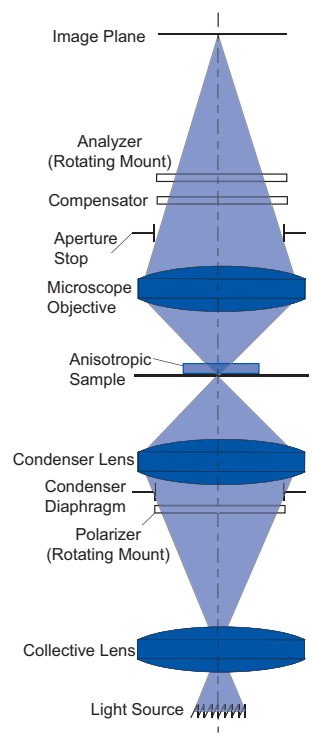

Images Obtained with Polarization Microscopes

Anisotropic samples observed under a **polarization microscope** contain dark and bright features and will strongly depend on the geometry of the sample. Objects can have characteristic elongated, linear, or circular structures:

- linear structures appear as dark or bright, depending on the orientation of the analyzer.
- circular structures have a "Maltese Cross" pattern with four quadrants of different intensities.

While polarization microscopy uses monochromatic light, it can also use white-light illumination. For broadband light, different wavelengths will be subject to different retardation; as a result, samples can provide different output colors. In practical terms, this means that color encodes information about retardation introduced by the object. The specific retardation is related to the image color. Therefore, the color allows for determination of sample thickness (for known retardation) or its retardation (for known sample thickness). If the polarizer and the analyzer are crossed, the color visible through the microscope is a complementary color to that having full wavelength retardation (a multiple of 2π, and the intensity is minimized for this color). Note that the two complementary colors combine into white light. If the analyzer could be rotated to be parallel to the analyzer, the two complementary colors will be switched (the one previously displayed will be minimized, while the other color will be maximized).

Polarization microscopy can provide quantitative information about the sample with the application of compensators (with known retardation). Estimating retardation might be performed visually or by integrating with an image acquisition system and CCD. This latter method requires one to obtain a series of images for different retarder orientations and calculate sample parameters based on saved images.

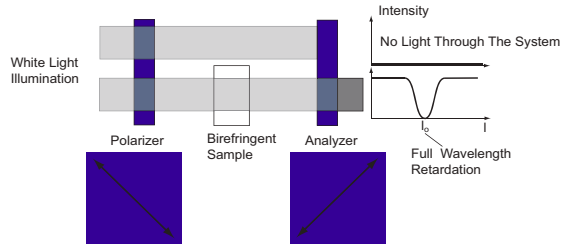

Compensators

Compensators are components that can be used to provide quantitative data about a sample's retardation. They introduce known retardation for a specific wavelength and can act as nulling devices to eliminate any phase delay introduced by the specimen. Compensators can also be used to control the background intensity level.

A **wave plate compensator** (also called a first-order red or red-I plate) is oriented at a 45-deg angle with respect to the analyzer and polarizer. It provides retardation equal to an even number of waves for 551 nm (green), which is the wavelength eliminated after passing the analyzer. All other wavelengths are retarded with a fraction of the wavelength and can partially pass the analyzer and appear as a bright red magenta. The sample provides additional retardation and shifts the colors towards blue and yellow. Color tables determine the retardation introduced by the object.

The **de Sénarmont compensator** is a quarter-wave plate (for 546 nm or 589 nm). It is used with monochromatic light for samples with retardation in the range of $\lambda/20$ to λ. A quarter-wave plate is placed between the analyzer and the polarizer with its slow ellipsoid axis parallel to the transmission axis of the analyzer. Retardation of the sample is measured in a two-step process. First, the sample is rotated until the maximum brightness is obtained. Next, the analyzer is rotated until the intensity maximally drops (the extinction position). The analyzer's rotation angle θ finds the phase delay related to retardation: $\Gamma_{\text{sample}} = 2\theta$.

A **Brace Köhler** rotating **compensator** is used for small retardations and monochromatic illumination. It is a thin birefringent plate with an optic axis in its plane. The analyzer and polarizer remain fixed while the compensator is rotated between the maximum and minimum intensities in the sample's image. Zero position (the maximum intensity) is determined for a slow axis being parallel to the transmission axis of the analyzer, and retardation is

$$\Gamma_{\text{sample}} = \Gamma_{\text{compensator}} \sin 2\theta .$$

Confocal Microscopy

Confocal microscopy is a scanning technique that employs pinholes at the illumination and detection planes so that only in-focus light reaches the detector. This means that the intensity for each sample point must be obtained in sequence through scanning in the x and y directions. An illumination pinhole is used in conjunction with the imaged sample plane and only illuminates the point of interest. The detection pinhole rejects the majority of out-of-focus light.

Confocal microscopy has the advantage of lowering the background light from out-of-focus layers, increasing spatial resolution, and providing the capability of imaging thick 3D samples, if combined with z scanning. Due to detection of only the in-focus light, confocal microscopy can provide images of thin sample sections. The system usually employs a photo-multiplier tube (PMT), avalanche photodiodes (APD), or a charge-coupled device (CCD) camera as a detector. For point detectors, recorded data is processed to assemble x-y images. This makes it capable of quantitative studies of an imaged sample's properties. Systems can be built for both reflectance and fluorescence imaging.

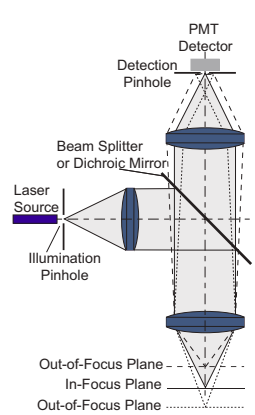

Spatial resolution of a confocal microscope can be defined as

$$d_{xy} \approx \frac{0.4\lambda}{NA},$$

and is slightly better than wide-field (bright-field) microscopy resolution. For pinholes larger than an Airy disk, the spatial resolution is the same as in a wide-field microscope. The **axial resolution** of a confocal system is

$$d_z \approx \frac{1.4n\lambda}{NA^2}.$$

The optimum pinhole value is at the full width at half maximum (FWHM) of the Airy disk intensity. This corresponds to ~75% of its intensity passing through the system, minimally smaller than the Airy disk's first ring:

$$D_{\text{pinhole}} = \frac{0.5\lambda M}{NA}.$$

Scanning Approaches

A scanning approach is directly connected with the temporal resolution of confocal microscopes. The number of points in the image, scanning technique, and the frame rate are related to the signal-to-noise ratio SNR (through time dedicated to detection of a single point). To balance these parameters, three major approaches were developed: point scanning, line scanning, and disk scanning.

A **point-scanning confocal microscope** is based on point-by-point scanning using, for example, two mirror galvanometers or resonant scanners. Scanning mirrors should be located in pupil conjugates (or close to them) to avoid light fluctuations. Maximum spatial resolution and maximum background rejection are achieved with this technique.

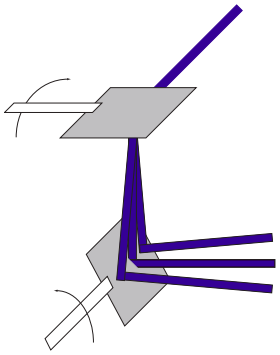

A **line-scanning confocal microscope** uses a slit aperture that scans in a direction perpendicular to the slit. It uses a cylindrical lens to focus light onto the slit to maximize throughput. Scanning in one direction makes this technique significantly faster than a point approach. However, the drawbacks are a loss of resolution and sectioning performance for the direction parallel to the slit.

Feature	Point Scanning	Slit Slit Scanning	Disk Spinning
z resolution	High	Depends on slit spacing	Depends on pinhole distribution
x,*y* resolution	High	Lower for one direction	Depends on pinhole spacing
Speed	Low to moderate	High	High
Light sources	Lasers	Lasers	Laser and other
Photobleaching	High	High	Low
QE of detectors	Low (PMT) Good (APD)	Good (CCD)	Good (CCD)
Cost	High	High	Moderate

Scanning Approaches (cont.)

Spinning-disk confocal imaging is a parallel-imaging method that maximizes the scanning rate and can achieve a 100–1000 speed gain over point scanning. It uses an array of pinholes/slits (e.g., Nipkow disk, Yokogawa, Olympus DSU approach). To minimize light loss, it can be combined (e.g., with the Yokogawa approach) with an array of lenses, so each pinhole has a dedicated focusing component. Pinhole disks contain several thousand pinholes, but only a portion is illuminated at one time.

	Throughput of a confocal microscope T [%]
Array of pinholes	$T_{\text{pinhole}} = 100 \times (D/S)^2$
Multiple slits	$T_{\text{slit}} = 100 \times (D/S)$
Equations are for a uniformly illuminated mask of pinholes/slits (array of microlenses not considered). D is the pinhole's diameter or slit's width respectively, while S is the pinhole/slit separation.	

Crosstalk between pinholes and slits will reduce the background rejection. Therefore a common pinhole/slit separation S is 5–10 times larger than the pinhole's diameter or the slit's width.

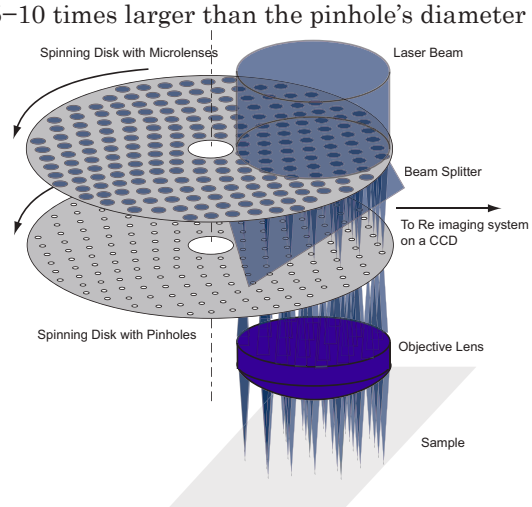

Spinning-disk and slit-scanning confocal microscopes require high-sensitivity array image sensors (CCD cameras) instead of point detectors. They can use lower excitation intensity for fluorescence confocal systems; therefore, they are less susceptible to photobleaching.

Images from a Confocal Microscope

Laser-scanning confocal microscopy rejects out-of-focus light and enables optical sectioning. It can be assembled in reflectance and fluorescent modes. A 1483 cell line, stained with membrane anti-EGFR antibodies labeled with fluorecent Alexa488 dye, is presented here. The top image shows the bright-field illumination mode. The middle image is taken by a confocal microscope with a pinhole corresponding to approximately 2.5 Airy disks. In the bottom image, the pinhole was entirely open. Clear cell boundaries are visible in the confocal image, while all out-of-focus features were removed.

Images were taken with a Zeiss LSM 510 microscope using a 63× Zeiss oil-immersion objective. Samples were excited with a 488-nm laser source.

Another example of 1483 cells labeled with EGF-Alexa647 and proflavine obtained on a Zeiss LSM 510 confocal at 63× / 1.4 oil is shown below. The proflavine (staining nuclei) was excited at 488 nm, with emission collected after a 505-nm long-pass filter. The EGF-Alexa647 (cell membrane) was excited at 633 nm, with emission collected after a 650–710 nm bandpass filter. The sectioning ability of a confocal system is nicely demonstrated by the channel with the membrane labeling.

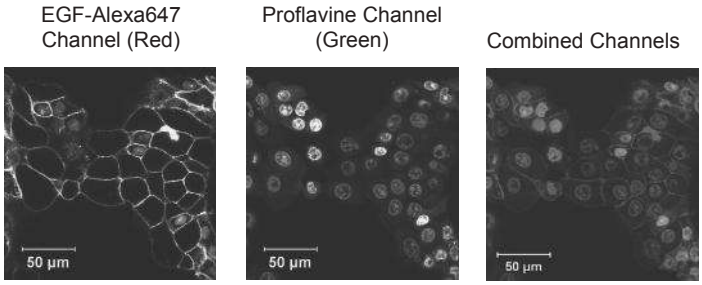

Fluorescence

Specimens can absorb and re-emit light through **fluorescence**. The specific wavelength of light absorbed or emitted depends on the energy level structure of each molecule. When a molecule absorbs the energy of light, it briefly enters an excited state before releasing part of the energy as fluorescence. Since the emitted energy must be lower than the absorbed energy, fluorescent light is always at longer wavelengths than the excitation light. Absorption and emission take place between multiple sub-levels within the ground and excited states, resulting in absorption and emission spectra covering a range of wavelengths. The loss of energy during the fluorescence process causes the **Stokes shift** (a shift of wavelength peak from that of excitation to that of emission). Larger Stokes shifts make it easier to separate excitation and fluorescent light in the microscope. Note that energy quanta and wavelength are related by $E = hc/\lambda$.

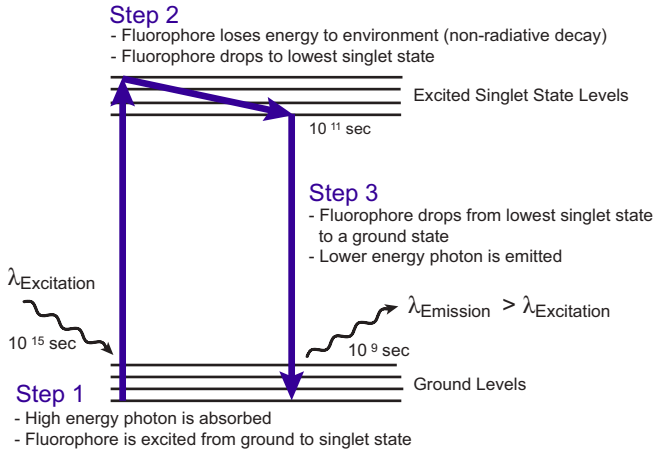

The fluorescence principle is widely used in fluorescence microscopy for both organic and inorganic substances. While it is most common for biological imaging, it is also possible to examine samples like drugs and vitamins.

Due to the application of filter sets, the fluorescence technique has a characteristically low background and provides high-quality images. It is used in combination with techniques like wide-field microscopy and scanning confocal-imaging approaches.

Configuration of a Fluorescence Microscope

A fluorescence microscope includes a set of three **filters**: an **excitation filter**, **emission filter**, and a **dichroic mirror** (also called a **dichroic beam splitter**). These filters separate weak emission signals from strong excitation illumination. The most common fluorescence microscopes are configured in epi-illumination mode. The dichroic mirror reflects incoming light from the lamp (at short wavelengths) onto the specimen. Fluorescent light (at longer wavelengths) collected by the objective lens is transmitted through the dichroic mirror to the eyepieces or camera. The transmission and reflection properties of the dichroic mirror must be matched to the excitation and emission spectra of the fluorophore being used.

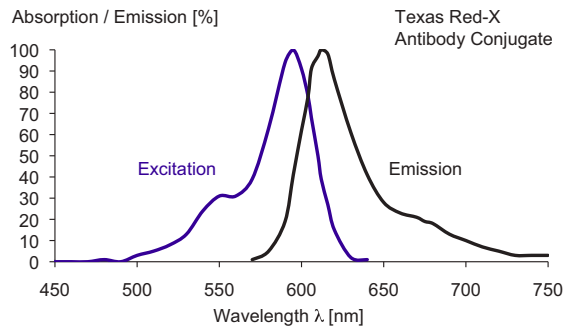

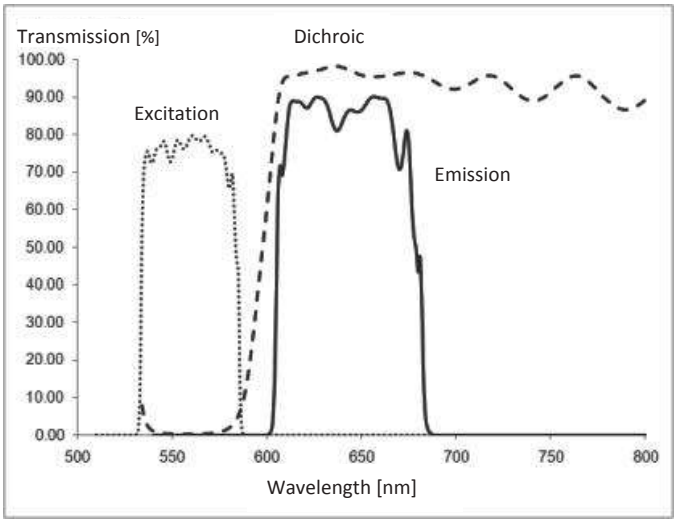

Configuration of a Fluorescence Microscope (cont.)

A sample must be illuminated with light at wavelengths within the excitation spectrum of the fluorescent dye. Fluorescence-emitted light must be collected at the longer wavelengths within the emission spectrum. Multiple fluorescent dyes can be used simultaneously, with each designed to localize or target a particular component in the specimen.

The **filter turret** of the microscope can hold several cubes that contain filters and dichroic mirrors for various dye types. Images are then reconstructed from CCD frames captured with each filter cube in place. Simultanous acquisition of emission for multiple dyes requires application of multiband dichroic filters.

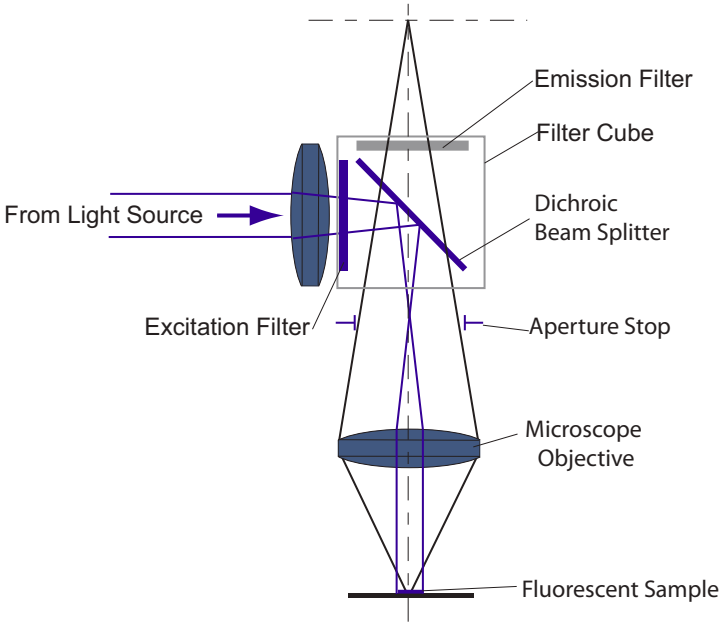

Images from Fluorescence Microscopy

Fluorescent images of cells labeled with three different fluorescent dyes, each targeting a different cellular component, demonstrate fluorescence microscopy's ability to target sample components and specific functions of biological systems. Such images can be obtained with individual filter sets for each fluorescent dye, or with a multiband (a triple-band in this example) filter set, which matches all dyes.

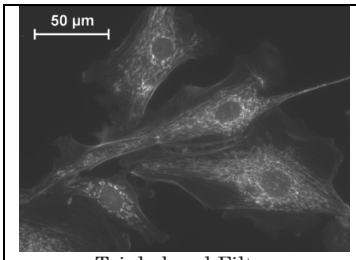

Triple-band Filter

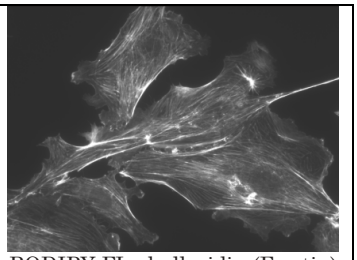

BODIPY FL phallacidin (F-actin)

MitoTracker Red CMXRos (mitochondria)

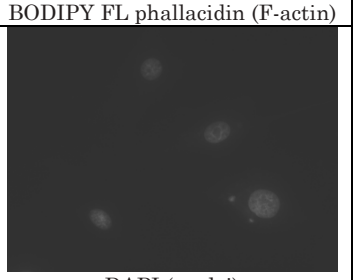

DAPI (nuclei)

Sample: Bovine pulmonary artery epithelial cells (Invitrogen FluoCells No. 1)

Components: Plan-apochromat 40×/0.95, MRm Zeiss CCD (1388×1040 mono)

Fluorescent labels:	Peak Excitation	Peak Emission
DAPI	358 nm	461 nm
BODIPY FL	505 nm	512 nm
MitoTracker Red CMXRos	579 nm	599 nm

Filter:	Excitation	Dichroic	Emission
Triple-band	395–415	435	448–472
	480–510	510	510–550
	560–590	600	600–650
DAPI	325–375	395	420–470
GFP	450–490	495	500–550
Texas Red	530–585	600	615LP

Properties of Fluorophores

Fluorescent emission F [photons/s] depends on the intensity of excitation light I [photons/s] and fluorophore parameters like the **fluorophore quantum yield** Q and **molecular absorption cross section** σ:

$$F = \sigma Q I \ ,$$

where the molecular absorption cross section is the probability that illumination photons will be absorbed by fluorescent dye. The intensity of excitation light depends on the power of the light source and throughput of the optical system. The detected fluorescent emission is further reduced by the quantum efficiency of the detector.

Quenching is a partially reversible process that decreases fluorescent emission. It is a non-emissive process of electrons moving from an excited state to a ground state.

Fluorescence emission and excitation parameters are bound through a photobleaching effect that reduces the ability of fluorescent dye to fluoresce. **Photobleaching** is an irreversible phenomenon that is a result of damage caused by illuminating photons. It is most likely caused by the generation of long-living (in triplet states) molecules of singlet oxygen and oxygen free radicals generated during its decay process. There is usually a finite number of photons that can be generated for a fluorescent molecule. This number can be defined as the ratio of emission quantum efficiency to bleaching quantum efficiency, and it limits the time a sample can be imaged before entirely bleaching. Photobleaching causes problems in many imaging techniques, but it can be especially critical in time-lapse imaging. To slow down this effect, optimizing collection efficiency (together with working at lower power at the steady-state condition) is crucial.

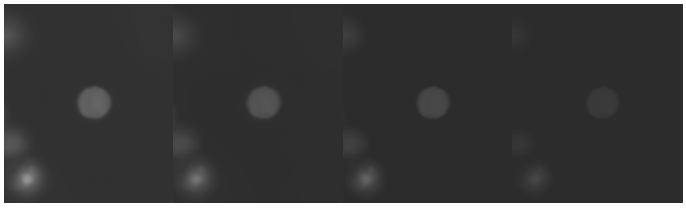

Photobleaching effect as seen in consecutive images

Single vs Multi-Photon Excitation

Multi-photon fluorescence is based on the fact that two or more low-energy photons match the energy gap of the fluorophore to excite and generate a single high-energy photon. For example, the energy of two 700-nm photons can excite an energy transition approximately equal to that of a 350-nm photon (it is not an exact value due to thermal losses). This is contrary to traditional fluorescence, where a high-energy photon (e.g., 400 nm) generates a slightly lower-energy (longer wavelength) photon. Therefore, one big advantage of multi-photon fluorescence is that it can obtain a significant difference between excitation and emission.

Fluorescence is based on stochastic behavior, and for single-photon excitation fluorescence, it is obtained with a high probability. However, multi-photon excitation requires at least two photons delivered in a very short time, and the probability is quite low:

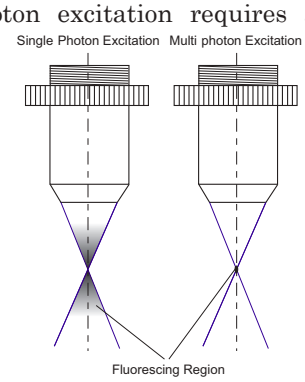

$$n_a \propto \delta\left(\frac{P_{avg}^2}{\tau\nu^2}\right)\left(\pi\frac{NA^2}{hc\lambda}\right)^2.$$

δ is the excitation cross section of dye, P_{avg} is the average power, τ is the length of a pulse, and ν is the repetition frequency. Therefore, multi-photon fluorescence must be used with high-power laser sources to obtain favorable probability. Short-pulse operation will have a similar effect, as τ will be minimized.

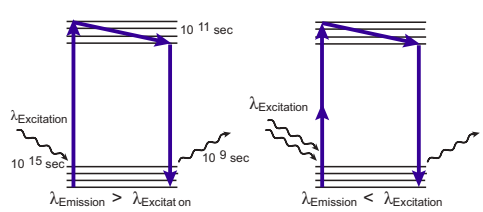

Due to low probability, fluorescence occurs only in the focal point. Multi-photon imaging does not require scanning through a pinhole and therefore has a high signal-to-noise ratio and can provide deeper imaging. This is also because near-IR excitation light penetrates more deeply due to lower scattering, and near-IR light avoids the excitation of several autofluorescent tissue components.

Light Sources for Scanning Microscopy

Lasers are an important light source for scanning microscopy systems due to their high energy density, which can increase the detection of both reflectance and fluorescence light. For laser-scanning confocal systems, a general requirement is a single-mode TEM_{00} laser with a short coherence length. Lasers are used primarily for point-and-slit scanning modalities. There are a great variety of laser sources, but certain features are useful depending on their specific application:

- Short-excitation wavelengths are preferred for fluorescence confocal systems because many fluorescent probes must be excited in the blue or ultraviolet range.
- Near-IR wavelengths are useful for reflectance confocal microscopy and multi-photon excitation. Longer wavelengths are less suseptible to scattering in diffused objects (like tissue) and can provide a deeper penetration depth.
- Broadband short-pulse lasers (like a Ti-Sapphire mode-locked laser) are particularily important for multi-photon generation, while shorter pulses increase the probability of excitation.

Laser Type	Wavelength [nm]
Argon-Ion	351, 364, 458, 488, 514
HeCd	325, 442
HeNe	543, 594, 633, 1152
Diode lasers	405, 408, 440, 488, 630, 635, 750, 810
DPSS (diode pumped, solid state)	430, 532, 561
Krypton-Argon	488, 568, 647
Dye	630
Ti-Sapphire*	710–920, 720–930, 750–850, 690–100, 680–1050

*High-power (1000mW or less), pulses between 1 ps and 100 fs

Non-laser sources are particularly useful for disk-scanning systems. They include arc lamps, metal-halide lamps, and high-power photo-luminescent diodes (See also pages 58 and 59).

Practical Considerations in LSM

A light source in laser-scanning microscopy must provide enough energy to obtain a sufficient number of photons to perform successful detection and a statistically significant signal. This is especially critical for non-laser sources (e.g., arc lamps) used in disk-scanning systems. While laser sources can provide enough power, they can cause fast photobleaching or photo-damage to the biological sample.

Detection conditions change with the type of sample. For example, fluorescent objects are subject to photobleaching and saturation. On the other hand, back-scattered light can be easily rejected with filter sets. In reflectance mode, out-of-focus light can cause background comparable or stronger than the signal itself. This background depends on the size of the pinhole, scattering in the sample, and overall system reflections.

Light losses in the optical system are critical for the signal level. Collection efficiency is limited by the NA of the objective; for example, a NA of 1.4 gathers approximately 30% of the fluorescent/scattered light. Other system losses will occur at all optical interfaces, mirrors, filters, and at the pinhole. For example, a pinhole matching the Airy disk allows approximately 75% of the light from the focused image point.

Quantum efficiency (QE) of a detector is another important system parameter. The most common photodetector used in scanning microscopy is the photomultiplier tube (PMT). Its quantum efficiency is often limited to the range of 10–15% (550–580 nm). In practical terms, this means that only 15% of the photons are used to produce a signal. The advantages of PMTs for laser-scanning microscopy are high speed and high signal gain. Better QE can be obtained for CCD cameras, which is 40–60% for standard designs and 90–95% for thinned, back-illuminated cameras. However, CCD image sensors operate at lower rates.

Spatial and temporal sampling of laser-scanning microscopy images imposes very important limitations on image quality. The image size and frame rate are often determined by the number of photons sufficient to form high-quality images.

Interference Microscopy

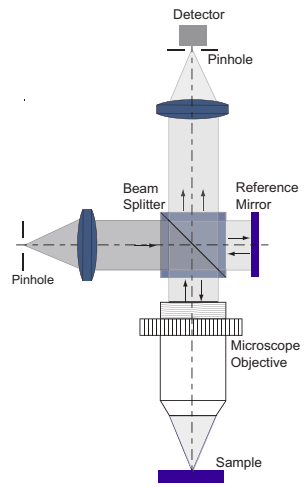

Interference microscopy includes a broad family of techniques for quantitative sample characterization (thickness, refractive index, etc.). Systems are based on microscopic implementation of interferometers (like the Michelson and Mach-Zehnder geometries), or polarization interferometers.

In interference microscopy, short coherence systems are particularly interesting and can be divided into two groups: **optical profilometry** and **optical coherence tomography** (OCT) [including **optical coherence microscopy** (OCM)]. The primary goal of these techniques is to add a third (z) dimension to the acquired data.
Optical profilers use interference fringes as a primary source of object height. Two important measurement approaches include **vertical scanning interferometry** (VSI) (also called **scanning white light interferometry** or **coherence scanning interferometry**) and **phase-shifting interferometry**. Profilometry techniques are capable of achieving nanometer-level resolution in the z direction while x and y are defined by standard microscope limitations.

Profilometry is used for characterizing the sample surface (roughness, structure, microtopography, and in some cases film thickness) while OCT/OCM is dedicated to 3D imaging, primarily of biological samples.
Both types of systems are usually configured using the geometry of the Michelson or its derivatives (e.g., Mirau).

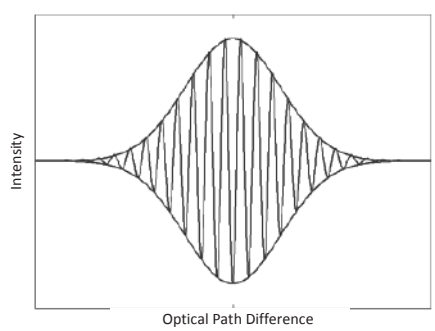

Optical Coherence Tomography/Microscopy

In early 3D coherence imaging, information about the sample was gated by the coherence length of the light source (**time-domain OCT**). This means that the maximum fringe contrast is obtained at a zero optical path difference, while the entire fringe envelope has a width related to the coherence length. In fact, this width defines the axial resolution (usually a few microns) and images are created from the magnitude of the fringe pattern envelope. **Optical coherence microscopy** is a combination of OCT and confocal microscopy. It merges the advantages of out-of-focus background rejection provided by the confocal principle and applies the coherence gating of OCT to improve optical sectioning over confocal reflectance systems.

In the mid-2000s, time-domain OCT transitioned to **Fourier-domain OCT** (FDOCT), which can acquire an entire depth profile in a single capture event. Similar to Fourier Transform Spectroscopy, the image is acquired by calculating the Fourier transform of the obtained interference pattern. The measured signal can be described as

$$I\left(k, z_{\mathrm{o}}\right) = \int A_{\mathrm{R}} A_{\mathrm{S}}\left(z\right) \cos\left[k\left(z_{\mathrm{m}} - z_{\mathrm{o}}\right)\right] dz\,,$$

where A_{R} and A_{S} are amplitudes of the reference and sample light, respectively, and z_{m} is the imaged sample depth (z_{o} is the reference length delay, and k is the wave number).

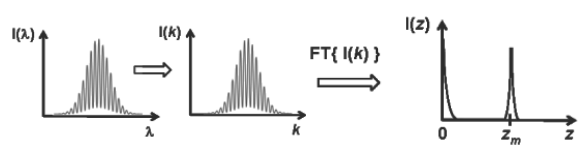

The fringe frequency is a function of wave number k and relates to the OPD in the sample. Within the Fourier-domain techniques, two primary methods exist: **spectral-domain OCT** (SDOCT) and **swept-source OCT** (SSOCT). Both can reach detection sensitivity and signal-to-noise ratios over 150 times higher than time-domain OCT approaches. SDOCT achieves this through the use of a broadband source and spectrometer for detection. SSOCT uses a single photodetector but rapidly tunes a narrow linewidth over a broad spectral profile. Fourier-domain systems can image at hundreds of frames per second with sensitivities over 100 dB.

Optical Profiling Techniques

There are two primary techniques used to obtain a surface profile using white-light profilometry: **vertical scanning interferomtry** (VSI) and **phase-shifting interferometry** (PSI).

VSI is based on the fact that a reference mirror is moved by a piezoelectric or other actuator, and a sequence of several images (in some cases hundreds or thousands) is acquired along the scan. During post-processing, algorithms look for a maximum fringe modulation and correlate it with the actuator's position. This position is usually determined with capacitance gauges; however, more advanced systems use optical encoders or an additional Michelson interferometer. The important advantage of VSI is its ability to ambiguously measure heights that are greater than a fringe.

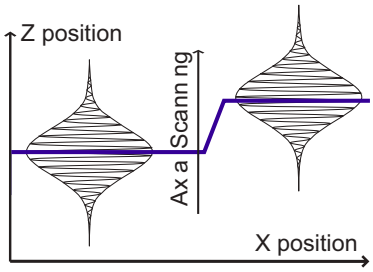

The PSI technique requires the object of interest to be within the coherence length and provides one order of magnitude or greater *z* resolution compared to VSI. To extend coherence length, an optical filter is introduced to limit the light source's bandwidth. With PSI, a phase-shifting component is located in one arm of the interferometer. The reference mirror can also provide the role of the phase shifter. A series of images (usually three to five) is acquired with appropriate phase differences to be used later in phase reconstruction. Introduced phase shifts change the location of the fringes. An important PSI drawback is the fact that phase maps are subject to 2π discontinuities. Phase maps are first obtained in the form of so-called phase fringes and must be unwrapped (a procedure of removing discontinuities) to obtain a continuous phase map.

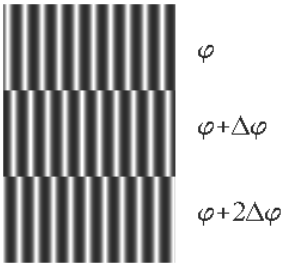

Note that modern short-coherence interference microscopes combine phase information from PSI with a long range of VSI methods.

Optical Profilometry: System Design

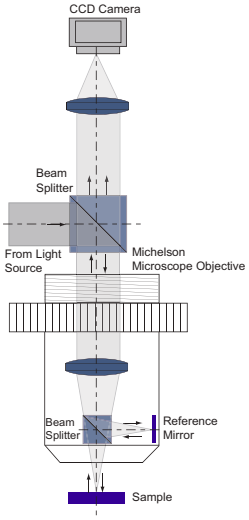

Optical profilometry systems are based on implementing a Michelson interferometer or its derivatives into their design. There are three primary implementations: classical **Michelson** geometry, **Mirau**, and **Linnik**. The first approach incorporates a Michelson inside a microscope objective using a classical interferometer layout. Consequently, it can work only for low magnifications and short working distances. The Mireau objective uses two plates perpendicular to the optical axis inside the objective. One is a beam-splitting plate, while the other acts as a reference mirror. Such a solution extends the working distance and increases the possible magnifications.

Design	Magnification
Michelson	1× to 5×
Mireau	10× to 100×
Linnik	50× to 100×

The Linnik design utilizes two matching objectives. It does not suffer from NA limitation, but it is quite expensive and susceptible to vibrations.

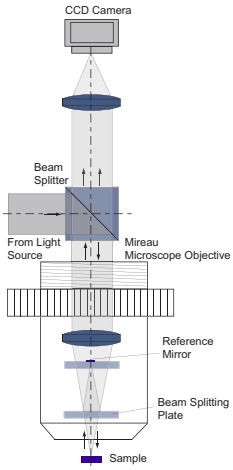

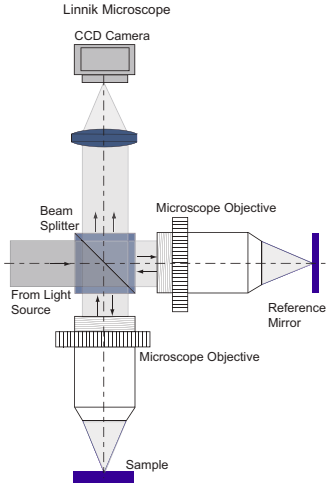

Phase-Shifting Algorithms

Phase-shifting techniques find the phase distribution from interference images given by

$$I(x,y) = a(x,y) + b(x,y)\cos(\varphi + n\Delta\varphi),$$

where $a(x, y)$ and $b(x, y)$ correspond to background and fringe amplitude, respectively. Since this equation has three unknowns, at least three measurements (images) are required. For image acquisition with a discrete CCD camera, spatial (x, y) coordinates can be replaced by (i, j) pixel coordinates.

The most basic PSF algorithms include the **three-**, **four-**, and **five-image** techniques.

I denotes the intensity for a specific image (1st, 2nd, 3rd, etc.) in the selected i, j pixel of a CCD camera. The phase shift for a three-image algorithm is $\pi/2$. Four- and five-image algorithms also acquire images with $\pi/2$ phase shifts.

$$\varphi = \arctan\left(\frac{I_3 - I_2}{I_1 - I_2}\right)$$

$$\varphi = \arctan\left(\frac{I_4 - I_2}{I_1 - I_3}\right)$$

$$\varphi = \arctan\left(\frac{2[I_2 - I_4]}{I_1 - I_3 + I_5}\right)$$

A reconstructed phase depends on the accuracy of the phase shifts. $\pi/2$ steps were selected to minimize the systematic errors of the above techniques. Accuracy progressively improves between the three- and five-point techniques. Note that modern PSI algorithms often use seven images or more, in some cases reaching thirteen.

After the calculations, the wrapped phase maps (modulo 2π) are obtained (arctan function). Therefore, unwrapping procedures have to be applied to provide continuous phase maps.

Common unwrapping methods include the line by line algorithm, least square integration, regularized phase tracking, minimum spanning tree, temporal unwrapping, frequency multiplexing, and others.

Structured Illumination: Axial Sectioning

A simple and inexpensive alternative to confocal microscopy is the **structured-illumination technique**. This method uses the principle that all but the zero (DC) spatial frequencies are attenuated with defocus. This observation provides the basis for obtaining the **optical sectioning** of images from a conventional wide-field microscope. A modified illumination system of the microscope projects a single spatial-frequency grid pattern onto the object. The microscope then faithfully images only that portion of the object where the grid pattern is in focus. The structured-illumination technique requires the acquisition of at least three images to remove the illumination structure and reconstruct an image of the layer. The reconstruction relation is described by

$$I = \sqrt{\left(I_0 - I_{2\pi/3}\right)^2 + \left(I_0 - I_{4\pi/3}\right)^2 + \left(I_{2\pi/3} - I_{4\pi/3}\right)^2} \ ,$$

where I denotes the intensity in the reconstructed image point, while I_0, $I_{2\pi/3}$, and $I_{4\pi/3}$ are intensities for the image point and the three consecutive positions of the sinusoidal illuminating grid (zero, 1/3 and 2/3 of the structure period).

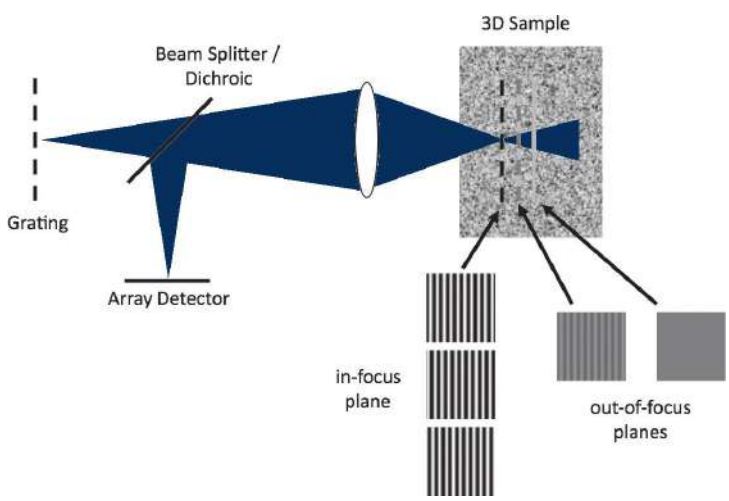

Note that the maximum sectioning is obtained for 0.5 of the objective's cut-off frequency. This sectioning ability is comparable to that obtained in confocal microscopy.

Structured Illumination: Resolution Enhancement

Structured illumination can be used to enhance the spatial resolution of a microscope. It is based on the fact that if a sample is illuminated with a periodic structure, it will create a low-frequency beating effect (Moiré fringes) with the features of the object. This new low-frequency structure will be capable of aliasing high spatial frequencies through the optical system. Therefore, the sample can be illuminated with a pattern of several orientations to accommodate the different directions of the object features. In practical terms, this means that system apertures will be doubled (in diameter), creating a new synthetic aperture.

To produce a high-frequency structure, the interference of two beams with large illumination angles can be used. Note that the blue dots in the figure represent aliased spatial frequencies.

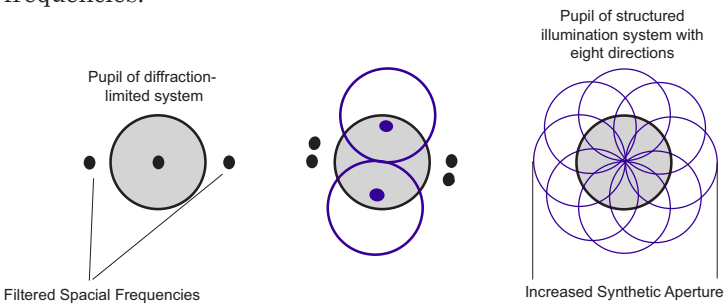

To reconstruct distribution, at least three images are required for one grid direction. In practice, obtaining a high-resolution image is possible with seven to nine images and a grid rotation, while a larger number can provide a more efficient reconstruction.

The linear-structured illumination approach is capable of obtaining two-fold resolution improvement over the diffraction limit. The application of nonlinear gain in fluorescence imaging improves resolution several times while working with higher harmonics. Sample features of 50 nm and smaller can be successfully resolved.

TIRF Microscopy

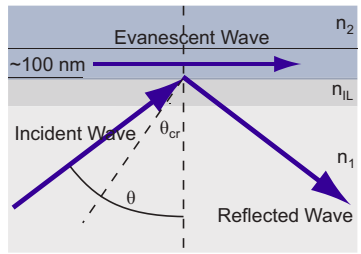

Total internal reflection fluorescence (TIRF) **microscopy** (also called evanescent wave microscopy) excites a limited width of the sample close to a solid interface. In TIRF, a thin layer of the sample is excited by a wavefront undergoing total internal reflection in the dielectric substrate, producing an **evanescent wave** propagating along the interface between the substrate and object.

The thickness of the excited section is limited to less than 100–200 nm. Very low background fluorescence is an important feature of this technique, since only a very limited volume of the sample in contact with the evanescent wave is excited.

This technique can be used for imaging cells, cytoplasmic filament structures, single molecules, proteins at cell membranes, micro-morphological structures in living cells, the adsorption of liquids at interfaces, or Brownian motion at the surfaces. It is also a suitable technique for recording long-term fluorescence movies.

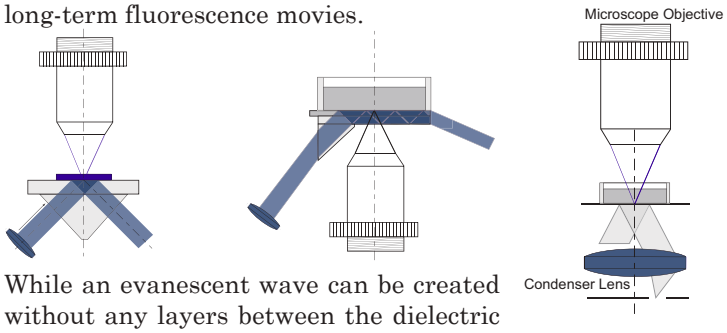

While an evanescent wave can be created without any layers between the dielectric substrate and sample, it appears that a thin layer (e.g., metal) improves image quality. For example, it can quench fluorescence in the first 10 nm and therefore provide selective fluorescence for the 10–200 nm region. TIRF can be easily combined with other modalities like optical trapping, multi-photon excitation, or interference. The most common configurations include oblique illumination through a high-NA objective or condenser and TIRF with a prism.

Solid Immersion

Optical resolution can be enhanced using **solid immersion** imaging. In the classical definition, the diffraction limit depends on the NA of the optical system and the wavelength of light. It can be improved by increasing the refractive index of the media between the sample and optical system or by decreasing the light's wavelength. The solid immersion principle is based on improving the first parameter by applying **solid immersion lenses** (SILs). To maximize performance, SILs are made with a high refractive index glass (usually greater than 2). The most basic setup for a SIL application involves a hemispherical lens. The sample is placed at the center of the hemisphere so the rays propagating through the system are not refracted (they intersect the SIL surface direction) and enter the microscope objective. The SIL is practically in contact with the object, but has a small sub-wavelength gap (<100 nm) between the sample and optical system; therefore, an object is always in an evanescent field and can be imaged with high resolution. Therefore, the technique is confined to a very thin sample volume (up to 100 nm) and can provide optical sectioning. Systems based on solid immersion can reach sub-100-nm lateral resolutions.

The most common solid-immersion applications are in microscopy, including fluorescence, optical data storage, and lithography. Compared to classical oil-immersion techniques, this technique is dedicated to imaging thin samples only (sub-100 nm), but provides much better resolution depending on the configuration and refractive index of the SIL.

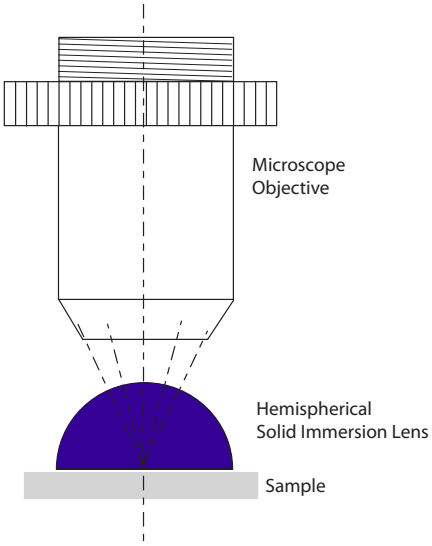

Stimulated Emission Depletion

Stimulated emission depletion (STED) microscopy is a fluorescence super-resolution technique that allows imaging with a spatial resolution at the molecular level. The technique has shown an improvement 5–10 times beyond the possible diffraction-resolution limit.

The STED technique is based on the fact that the area around the excited object point can be treated with a pulse (depletion pulse or STED pulse) that depletes high-energy states and brings fluorescent dye to the ground state. Consequently, an actual excitation pulse excites only the small sub-diffraction resolved area. The depletion pulse must have high intensity and be significantly shorter than the lifetime of the excited states. The pulse must be shaped, for example, with a half-wave phase plate and imaging optics to create a 3D doughnut-like structure around the point of interest. The STED pulse is shifted toward red with regard to the fluorescence excitation pulse and follows it by a few picoseconds. To obtain the complete image, the system scans in the x, y, and z directions.

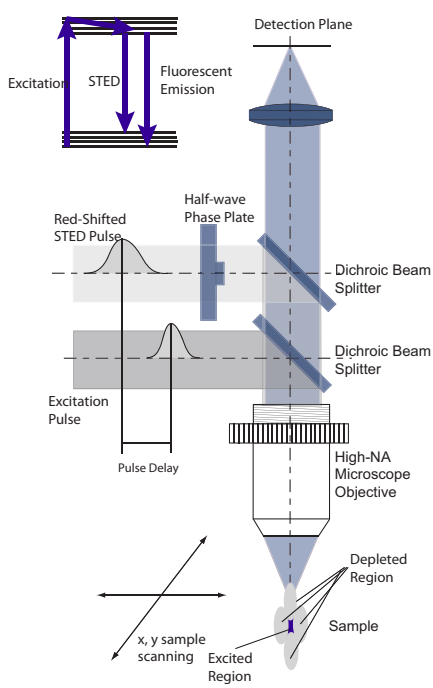

STORM

Stochastic optical reconstruction microscopy (STORM) is based on the application of multiple photo-switchable fluorescent probes that can be easily distinguished in the spectral domain. The system is configured in **TIRF** geometry. Multiple fluorophores label a sample, and each is built as a dye pair where one component acts as a fluorescence activator. The activator can be switched on or deactivated using a different illumination laser (a longer wavelength is used for deactivation—usually red). In the case of applying several dual-pair dyes, different lasers can be used for activation and only one for deactivation.

A sample can be sequentially illuminated with activation lasers, which generate fluorescent light at different wavelengths for slightly different locations on the object (each dye is sparsely distributed over the sample and activation is stochastic in nature). Activation for a different color can also be shifted in time. After performing several sequences of activation/deactivation, it is possible to acquire data for the centroid localization of various diffraction-limited spots. This localization can be performed with precision to about 1 nm. The spectral separation of dyes detects closely located object points encoded with a different color.

The STORM process requires 1000+ activation/deactivation cycles to produce high-resolution images. Therefore, final images combine the spots corresponding to individual molecules of DNA or an antibody used for labeling. STORM images can reach a lateral resolution of about 25 nm and an axial resolution of 50 nm.

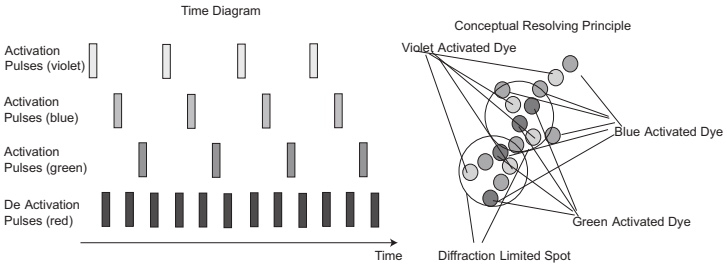

4Pi Microscopy

4Pi microscopy is a fluorescence technique that significantly reduces the thickness of an imaged section. The 4Pi principle is based on coherent illumination of a sample from two directions. Two constructively interfering beams improve the axial resolution by three to seven times. The possible values of z resolution using 4Pi microscopy are 50–100 nm and 30–50 nm when combined with STED.

The undesired effect of the technique is the appearance of side lobes located above and below the plane of the imaged section. They are located about $\lambda/2$ from the object. To eliminate side lobes, three primary techniques can be used:

- Combine 4Pi with confocal detection. A pinhole rejects some of the out-of-focus light.
- Detect in multi-photon mode, which quickly diminishes the excitation of fluorescence.
- Apply a modified 4Pi system, which creates interference at both the object and detection planes (two imaging systems are required).

It is also possible to remove side lobes through numerical processing, if they are smaller than 50% of the maximum intensity. However, side lobes increase with the NA of the objective; for an NA of 1.4, they are about 60–70% of the maximum intensity. Numerical correction is not effective in this case.

4Pi microscopy improves resolution only in the z direction; however, the perception of details in the thin layers is a useful benefit of the technique.

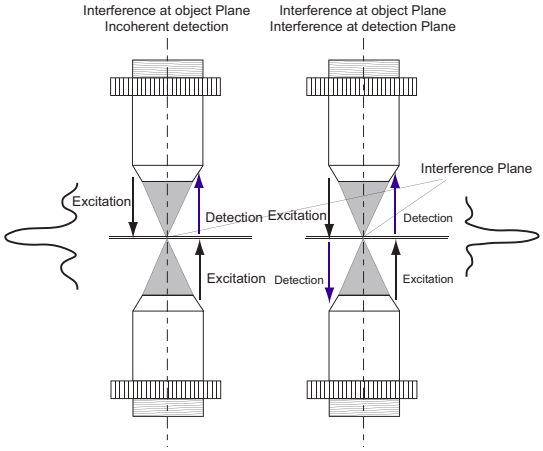

The Limits of Light Microscopy

Classical light microscopy is limited by diffraction and depends on the wavelength of light and the parameters of an optical system (NA). However, recent developments in super-resolution methods and enhancement techniques overcome these barriers and provide resolution at a molecular level. They often use a sample as a component of the optical train [resolvable saturable optical fluorescence transitions (**RESOLFT**), saturated structured-illumination microscopy (**SSIM**), photoactivated localization microscopy (**PALM**) and **STORM**] or combine near-field effects (**4Pi, TIRF, solid immersion**) with far-field imaging. The table below summarizes these methods.

		Demonstrated Values	
Method	**Principles**	**Lateral**	**Axial**
Bright field	Diffraction	200 nm	500 nm
Confocal	Diffraction (slightly better than bright field)	200 nm	500 nm
Solid immersion	Diffraction / evanescent field decay	< 100 nm	< 100 nm
TIRF	Diffraction / evanescent field decay	200 nm	< 100 nm
4Pi / I5	Diffraction/ interference	200 nm	50 nm
RESOLFT (e.g.,STED)	Depletion, molecular structure of sample–fluorescent probes	20 nm	20 nm
Structured illumination (SSIM)	Aliasing, nonlinear gain in fluorescent probes (molecular structure)	25–50 nm	50–100 nm
Stochastic techniques (PALM, STORM)	Fluorescent probes (molecular structure), centroid localization, time-dependant fluorophore activation	25 nm	50 nm

Raman and CARS Microscopy

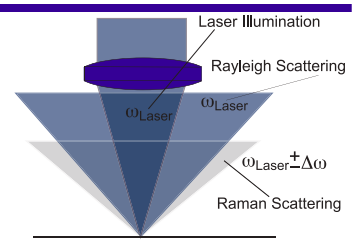

Raman microscopy (also called chemical imaging) is a technique based on inelastic **Raman scattering** and evaluates the vibrational properties of samples (minerals, polymers, and biological objects).

Raman microscopy uses a laser beam focused on a solid object. Most of the illuminating light is scattered, reflected, and transmitted, and preserves the parameters of the illumination beam (the frequency is the same as the illumination). However, a small portion of the light is subject to a shift in frequency: $\omega_{Raman} = \omega_{laser} \pm \Delta\omega$. This frequency shift between illumination and scattering bears information about the molecular composition of the sample. Raman scattering is weak and requires high-power sources and sensitive detectors. It is examined with spectroscopy techniques.

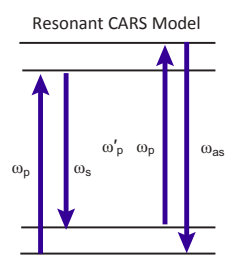

Coherent Anti-Stokes Raman Scattering (**CARS**) overcomes the problem of weak signals associated with Raman imaging. It uses a pump laser and a tunable Stokes laser to stimulate a sample with a four-wave mixing process. The fields of both lasers [pump field $E(\omega_p)$, Stokes field $E(\omega_s)$, and probe field $E'(\omega_p')$] interact with the sample and produce an anti-Stokes field [$E(\omega_{as})$] with frequency ω_{as}, so $\omega_{as} = 2\omega_p - \omega_s$.

CARS works as a resonant process, only providing a signal if the vibrational structure of the sample specifically matches $\omega_p - \omega_s$. It also must assure phase matching so that l_c (the coherence length) is greater than $\pi/\Delta\mathbf{k} = \pi/\left|\mathbf{k}_{as} - (2\mathbf{k}_p - \mathbf{k}_s)\right|$, where $\Delta\mathbf{k}$ is a phase mismatch. CARS is orders of magnitude stronger than Raman scattering, does not require any exogenous contrast, provides good sectioning ability, and can be set up in both transmission and reflectance modes.

SPIM

Selective plane illumination microscopy (SPIM) is dedicated to the high-resolution imaging of large 3D samples. It is based on three principles:

- The sample is illuminated with a **light sheet**, which is obtained with cylindrical optics. The light sheet is a beam focused in one direction and collimated in another. This way, the thin and wide light sheet can pass through the object of interest (see figure).
- The sample is imaged in the direction perpendicular to the illumination.
- The sample is rotated around its axis of gravity and linearly translated into axial and lateral directions.

Data is recorded by a CCD camera for multiple object orientations and can be reconstructed into a single 3D distribution. Both scattered and fluorescent light can be used for imaging.

The maximum resolution of the technique is limited by the longitudinal resolution of the optical system (for a high numerical aperture, objectives can obtain micron-level values). The maximum volume imaged is limited by the working distance of a microscope and can be as small as tens of microns or exceed several millimeters. The object is placed in an air, oil, or water chamber.

The technique is primarily used for bio-imaging, ranging from small organisms to individual cells.

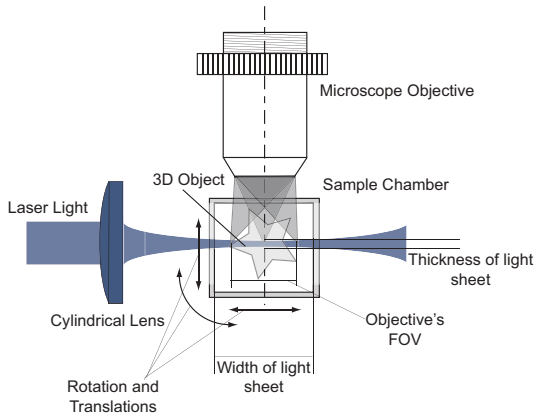

Array Microscopy

An **array microscope** is a solution to the trade-off between field of view and lateral resolution. In the array microscope, a miniature microscope objective is replicated tens of times. The result is an imaging system with a field of view that can be increased in steps of an individual objective's field of view, independent of numerical aperture and resolution. An array microscope is useful for applications that require fast imaging of large areas at a high level of detail. Compared to conventional microscope optics for the same purpose, an array microscope can complete the same task several times faster due to its parallel imaging format.

An example implementation of the array-microscope optics is shown. This array microscope is used for imaging glass slides containing tissue (histology) or cells (cytology). For this case, there are a total of 80 identical miniature 7×/0.6 microscope objectives. The summed field of view measures 18 mm. Each objective consists of three lens elements. The lens surfaces are patterned on three plates that are stacked to form the final array of microscopes. The plates measure 25 mm in diameter. Plate 1 is near the object at a working distance of 400 μm. Between plates 2 and 3 is a baffle plate. A second baffle is located between the third lens and the image plane. The purpose of the baffles is to eliminate cross talk and image overlap between objectives in the array.

The small size of each microscope objective and the need to avoid image overlap jointly dictate a low magnification. Therefore, the array microscope works best in combination with an image sensor divided into small pixels (e.g., 3.3 μm).

Focusing the array microscope is achieved by an up/down translation and two rotations: a pitch and a roll.

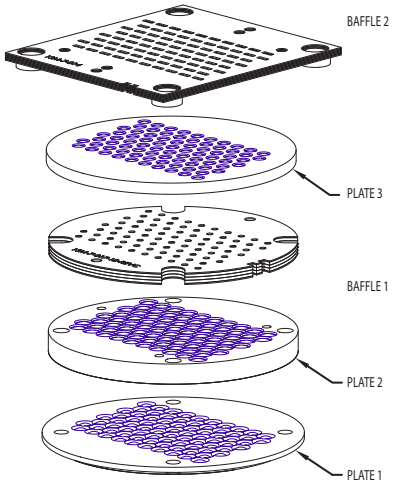

Digital Microscopy

Digital microscopy emerged with the development of electronic array image sensors and replaced the microphotography techniques previously used in microscopy. It can also work with point detectors when images are recombined in post or real-time processing. Digital microscopy is based on acquiring, storing, and processing images taken with various microscopy techniques. It supports applications that require:

- Combining data sets acquired for one object with different microscopy modalities or different imaging conditions. Data can be used for further analysis and processing in techniques like hyperspectral and polarization imaging.
- Image correction (e.g., distortion, white balance correction, or background subtraction). Digital microscopy is used in deconvolution methods that perform numerical rejection of out-of-focus light and iteratively reconstruct an object's estimate.
- Image acquisition with a high temporal resolution. This includes short integration times or high frame rates.
- Long time experiments and remote image recording.
- Scanning techniques where the image is reconstructed after point-by-point scanning and displayed on the monitor in real time.
- Low light, especially fluorescence. Using high-sensitivity detectors reduces both the excitation intensity and excitation time, which mitigates photobleaching effects.
- Contrast enhancement techniques and an improvement in spatial resolution. Digital microscopy can detect signal changes smaller than possible with visual observation.
- Super-resolution techniques that may require the acquisition of many images under different conditions.
- High throughput scanning techniques (e.g., imaging large sample areas).
- UV and IR applications not possible with visual observation.

The primary detector used for digital microscopy is a CCD camera. For scanning techniques, a photomultiplier or photodiodes are used.

Principles of CCD Operation

Charge-coupled device (CCD) sensors are a collection of individual photodiodes arranged in a 2D grid pattern. Incident photons produce electron-hole pairs in each pixel in proportion to the amount of light on each pixel. By collecting the signal from each pixel, an image corresponding to the incident light intensity can be reconstructed.

Here are the step-by-step processes in a CCD:

1. The CCD array is illuminated for integration time τ.
2. Incident photons produce mobile charge carriers.
3. Charge carriers are collected within the p-n structure.
4. The illumination is blocked after time τ (full-frame only).
5. The collected charge is transferred from each pixel to an amplifier at the edge of the array.
6. The amplifier converts the charge signal into voltage.
7. The analog voltage is converted to a digital level for computer processing and image display.

Each pixel (**photodiode**) is reverse biased, causing photoelectrons to move towards the positively charged electrode. Voltages applied to the electrodes produce a potential well within the semiconductor structure. During the integration time, electrons accumulate in the potential well, up to the full-well capacity. The full-well capacity is reached when the repulsive force between electrons exceeds the attractive force of the well. At the end of the exposure time, each pixel has stored a number of electrons in proportion to the amount of light received. These charge packets must be transferred from the sensor, from each pixel to a single amplifier without loss. This is accomplished by a series of parallel and serial shift registers. The charge stored by each pixel is transferred across the array by sequences of gating voltages applied to the electrodes. The packet of electrons follows the positive clocking waveform voltage from pixel-to-pixel or row-to-row. A potential barrier is always maintained between adjacent pixel charge packets.

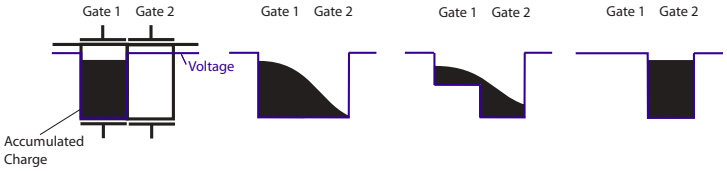

CCD Architectures

In a **full-frame architecture**, individual pixel charge packets are transferred by a parallel row shift to the serial register, then one-by-one to the amplifier. The advantage of this approach is a 100% photosensitive area, while the disadvantage is that a shutter must block the sensor area during readout, and consequently limits the frame rate.

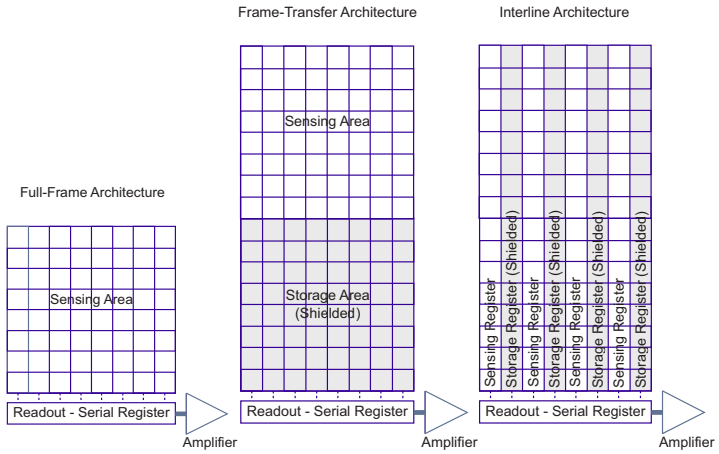

The **frame-transfer architecture** has the advantage of not needing to block an imaging area during readout. It is faster than full-frame, since readout can take place during the integration time. The significant drawback is that only 50% of the sensor area is used for imaging.

The **interline transfer architecture** uses columns with exposed imaging pixels interleaved with columns of masked storage pixels. A charge is transferred into an adjacent storage column for readout. The advantage of such an approach is a high frame rate due to a rapid transfer time (< 1 ms), while again the disadvantage is that 50% of the sensor area is used for light collection. However, recent interleaved CCDs have an increased fill factor due to the application of microlens arrays.

CCD Architectures (cont.)

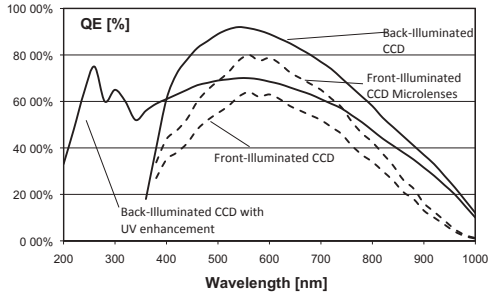

Quantum efficiency (QE) is the percentage of incident photons at each wavelength that produce an electron in a photodiode. CCDs have a lower QE than individual silicon photodiodes due to the charge-transfer channels on the sensor surface, which reduces the effective light-collection area. A typical CCD QE is 40–60%. This can be improved for back-thinned CCDs, which are illuminated from behind; this avoids the surface electrode channels and improves QE to approximately 90–95%.

Recently, **electron-multiplying CCDs** (EM-CCD), primarily used for biological imaging, have become more common. They are equipped with a gain register placed between the shift register and output amplifier. A multi-stage gain register allows high gains. As a result, single electrons can generate thousands of output electrons. While the read noise is usually low, it is therefore negligible in the context of thousands of signal electrons.

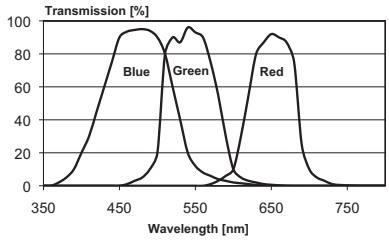

CCD pixels are not color sensitive, but use color filters to separately measure red, green, and blue light intensity. The most common method is to cover the sensor with an array of **RGB** (red, green, blue) filters, which combine four individual pixels to generate a single color pixel. The **Bayer mask** uses two green pixels for every red and blue pixel, which simulates human visual sensitivity.

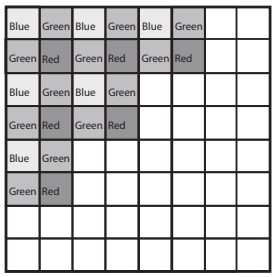

CCD Noise

The three main types of noise that affect CCD imaging are **dark noise**, **read noise**, and **photon noise**.

Dark noise arises from random variations in the number of thermally generated electrons in each photodiode. More electrons can reach the conduction band as the temperature increases, but this number follows a statistical distribution. These random fluctuations in the number of conduction band electrons results in a **dark current** that is not due to light. To reduce the influence of dark current for long time exposures, CCD cameras are cooled (which is crucial for the exposure of weak signals, like fluorescence, with times of a few seconds and more).

Read noise describes the random fluctuation in electrons contributing to measurement, due to electronic processes on the CCD sensor. This noise arises during the charge transfer, the charge-to-voltage conversion, and the analog-to-digital conversion. Every pixel on the sensor is subject to the same level of read noise, most of which is added by the amplifier.

Dark noise and read noise are due to the properties of the CCD sensor itself.

Photon noise (or **shot noise**) is inherent in any measurement of light, due to the fact that photons arrive at the detector randomly in time. This process is described by **Poisson statistics**. The probability P of measuring k events, given an expected value N is

$$P(k,N) = \frac{N^k e^{-N}}{k!}.$$

The standard deviation of the Poisson distribution is $N^{1/2}$. Therefore, if the average number of photons arriving at the detector is N, then the noise is $N^{1/2}$. Since the average number of photons is proportional to incident power, shot noise increases with $P^{1/2}$.

Signal-to-Noise Ratio and the Digitization of CCD

Total **noise** as a function of the number of electrons from all three contributing noises is given by

$$Noise\left[N_{\text{electrons}}\right] = \sqrt{\sigma_{\text{Photon}}^2 + \sigma_{\text{Dark}}^2 + \sigma_{\text{Read}}^2}\ ,$$

where

$\sigma_{\text{Photon}} = \sqrt{\varphi\eta\tau}$, $\sigma_{\text{Dark}} = \sqrt{I_{\text{Dark}}\tau}$, and $\sigma_{\text{Read}} = N_{\text{R}}$, I_{Dark} is dark current (electrons per second), and N_{R} is read noise (electrons, rms).

The quality of an image is determined by the **signal-to-noise ratio** (SNR). The signal can be defined as

$$Signal\left[N_{\text{electrons}}\right] = \varphi\eta\tau\ ,$$

where φ is the incident photon flux at the CCD (photons per second), η is the quantum efficiency of the CCD (electrons per photon), and τ is the integration time (in seconds). Therefore, SNR can be defined as:

$$SNR = \frac{\varphi\eta\tau}{\sqrt{\varphi\eta\tau + I_{\text{dark}}\tau + N_{\text{R}}^2}}\ .$$

It is best to use a CCD under photon-noise-limited conditions. If possible, it would be optimal to increase the integration time to increase the SNR and work in photon-noise-limited conditions. However, an increase in integration time is possible until reaching full-well capacity (saturation level):

$$SNR = \frac{\varphi\eta\tau}{\sqrt{\varphi\eta\tau}} = \sqrt{\varphi\eta\tau}.$$

The **dynamic range** can be derived as a ratio of full-well capacity and read noise. Digitization of the CCD output should be performed to maintain the dynamic range of the camera. Therefore, an analog-to-digital converter should support (at least) the same number of gray levels calculated by the CCD's dynamic range. Note that a high bit depth extends readout time, which is especially critical for large-format cameras.

CCD Sampling

The maximum spatial frequency passed by the CCD is one half of the sampling frequency—the **Nyquist frequency**. Any frequency higher than Nyquist will be aliased at lower frequencies.

Undersampling refers to a frequency where the sampling rate is not sufficient for the application. To assure maximum resolution of the microscope, the system should be optics limited, and the CCD should provide sampling that reaches the diffraction resolution limit. In practice, this means that at least two pixels should be dedicated to a distance of the resolution. Therefore, the maximum pixel spacing that provides the diffraction limit can be estimated as

$$d_{\text{pix}} = M\frac{0.61\lambda}{2NA},$$

where M is the magnification between the object and the CCD's plane. A graph representing how sampling influences the MTF of the system, including the CCD, is presented below.

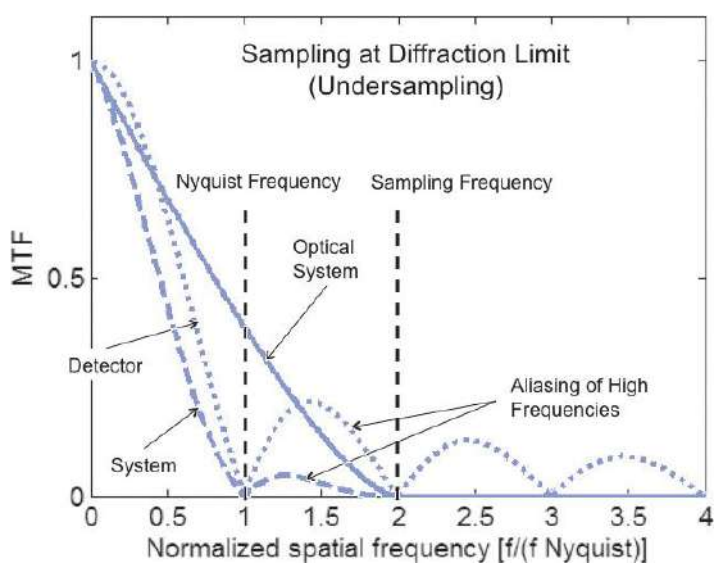

CCD Sampling (cont.)

Oversampling means that more than the minimum number of pixels, according to the Nyquist criteria, are available for detection, which does not imply excessive sampling of the image. However, excessive sampling decreases the available field of view of a microscope. The relation between the extent of field D, the number of pixels in the x and y directions (N_x and N_y, respectively), and the pixel spacing d_{pix} can be calculated from

$$D_x = \frac{N_x d_{\text{pix}-x}}{M}$$

and

$$D_y = \frac{N_y d_{\text{pix}-y}}{M} .$$

This assumes that the object area imaged by the microscope is equal to or larger than D. Therefore, the size of the microscope's field of view and the size of the CCD chip should be matched.

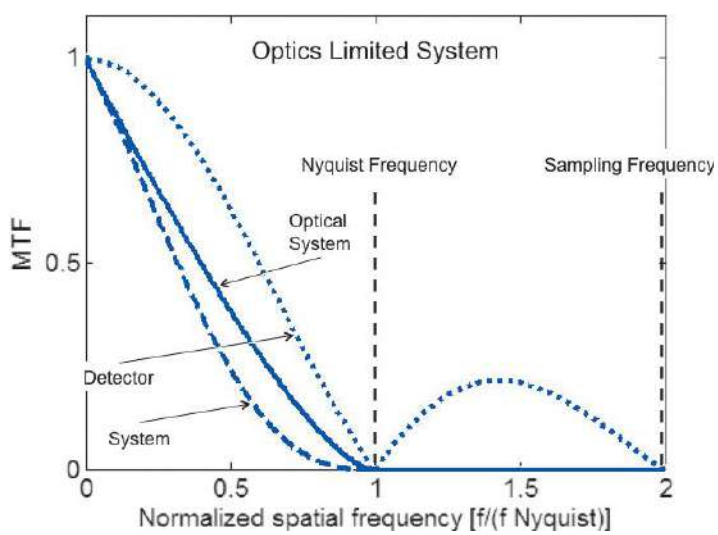

Equation Summary

Quantized Energy:

$$E = h\nu = hc/\lambda$$

Propagation of an electric field and wave vector:

$$E = A\sin\left(\omega t - kz + \varphi_o\right) = A\exp\left[-i\left(\omega t - kz + \varphi_o\right)\right]$$

$$\omega = \frac{2\pi}{T} = \frac{2\pi V_m}{\lambda} \qquad E\left(z,t\right) = E_x + E_y$$

$$E_x = A_x \exp\left[-i\left(\omega t - kz + \varphi_x\right)\right] \qquad E_y = A_y \exp\left[-i\left(\omega t - kz + \varphi_y\right)\right]$$

Refractive index:

$$n = \frac{c}{V_m}$$

Optical path length:

$$\mathrm{OPL} = nL$$

$$\mathrm{OPL} = \int_{P1}^{P2} n\,ds$$

$$ds^2 = dx^2 + dy^2 + dz^2$$

Optical path difference and phase difference:

$$\mathrm{OPD} = n_1 L_1 - n_2 L_2$$

$$\Delta\varphi = 2\pi\frac{\mathrm{OPD}}{\lambda}$$

TIR:

$$\theta_{cr} = \arcsin\left(\frac{n_1}{n_2}\right)$$

$$I = I_o \exp\left(-y/d\right)$$

$$d = \frac{\lambda}{4\pi n_1}\frac{1}{\sqrt{\sin^2\theta - \sin^2\theta_c}}$$

Coherence length:

$$l_c = \frac{\lambda^2}{\Delta\lambda}$$

Equation Summary (cont'd)

Two-beam interference:

$$I = EE^*$$

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\Delta\varphi)$$

$$\Delta\varphi = \varphi_2 - \varphi_1.$$

Contrast:

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

Diffraction grating equation:

$$m\lambda = d(\sin\beta \pm \sin\alpha)$$

Resolving power of diffraction grating:

$$\frac{\lambda}{\Delta\lambda} = mN$$

Free spectral range:

$$\Delta\lambda = \frac{m+1}{m}\lambda_1 - \lambda_1 = \frac{\lambda_1}{m}.$$

Newtonian equation:

$$xx' = ff'$$

$$xx' = -f'^2$$

Gaussian imaging equation:

$$\frac{n'}{z'} = \frac{n}{z} + \frac{1}{f_e}$$

$$\frac{1}{z'} = \frac{1}{z} + \frac{1}{f_e}$$

$$f_e = \frac{1}{\varphi} = -\frac{f}{n} = \frac{f'}{n'}$$

Transverse magnification:

$$M = \frac{h'}{h} = -\frac{x'}{f'} = -\frac{f}{x} = \frac{z'}{z} = \frac{-f}{z-f} = \frac{-(z'-f')}{f'}$$

Longitudinal magnification:

$$\frac{\Delta z'}{\Delta z} = \left(-\frac{f'}{f}\right)M_1 M_2$$

Equation Summary (cont'd)

Optical transfer function:

$$\mathrm{OTF}(\rho) = \mathrm{MTF}(\rho)\exp\left[i\varphi(\rho)\right]$$

Modulation transfer function:

$$\mathrm{MTF}(\rho) = \frac{C(\rho)_{\mathrm{image}}}{C(\rho)_{\mathrm{object}}}$$

Field of view of microscope:

$$\mathrm{FOV} = \frac{FieldNumber}{M_{\mathrm{objective}}}[\mathrm{mm}]$$

Magnifying power:

$$\mathrm{MP} = \frac{u'}{u} = \frac{d_o(f - z')}{f(z' - l)}$$

$$MP \approx \frac{250\ \mathrm{mm}}{f}$$

Magnification of the microscope objective:

$$M_{\mathrm{objective}} = \frac{OTL}{f_{\mathrm{objective}}}$$

Magnifying power of the microscope:

$$\mathrm{MP}_{\mathrm{microscope}} = M_{\mathrm{objective}}\mathrm{MP}_{\mathrm{eyepiece}} = -\frac{\mathrm{OTL}}{f_{\mathrm{objective}}}\frac{250\,\mathrm{mm}}{f_{\mathrm{eyepiece}}}$$

Numerical aperture:

$$NA = n\sin u$$

$$NA = NA'M_{\mathrm{objective}}$$

Airy disk:

$$d = \frac{1.22\lambda}{n\sin u} = \frac{1.22\lambda}{NA}$$

Rayleigh resolution limit:

$$d = \frac{0.61\lambda}{n\sin u} = \frac{0.61\lambda}{NA}$$

Sparrow resolution limit:

$$d = 0.5\lambda/NA$$

Equation Summary (cont'd)

Abbe resolution limit:

$$d = \frac{\lambda}{NA_{\text{objective}} + NA_{\text{condenser}}}$$

Resolving power of the microscope:

$$d_{\text{mic}} = \frac{d_{\text{eye}}}{M_{\text{min}}} = \frac{d_{\text{eye}}}{M_{\text{obj}} M_{\text{eyepiece}}}$$

Depth of focus and depth of field:

$$DOF = 2\Delta z' = \frac{n'\lambda}{\text{NA}'^2}$$

$$2\Delta z' = M^2_{\text{objective}}\, 2\Delta z \frac{n'}{n}$$

Depth perception of stereoscopic microscope:

$$\delta z = \frac{250\alpha_s}{M_{\text{microscope}} \tan\gamma} [\text{mm}]$$

Minimum perceived phase in phase contrast:

$$\Delta\varphi_{\text{min}} = \frac{C_{\text{ph-min}}\lambda}{4\pi\sqrt{N}}$$

Lateral resolution of the phase contrast:

$$d = f'_{\text{objective}} \frac{\lambda}{\left(r_{\text{AS}} + r_{\text{PR}}\right)}$$

Intensity in DIC:

$$I \propto \sin^2\left(\frac{\pi s \frac{d\varphi_o}{dx}}{\lambda}\right)$$

$$I \propto \sin^2 \frac{\pi\left(\Delta_b \pm s\frac{d\varphi_o}{dx}\right)}{\lambda}$$

Retardation:

$$\Gamma = \left(n_e - n_o\right)t$$

Equation Summary (cont'd)

Birefringence:

$$\delta = \frac{2\pi OPD}{\lambda} = \frac{2\pi\Gamma}{\lambda}$$

Resolution of a confocal microscope:

$$d_{xy} \approx \frac{0.4\lambda}{NA},$$

$$d_z \approx \frac{1.4n\lambda}{NA^2}$$

Confocal pinhole width:

$$D_{\text{pinhole}} = \frac{0.5\lambda M}{NA}$$

Fluorescent emission:

$$F = \sigma QI$$

Probability of two-photon excitation:

$$n_{\text{a}} \propto \delta\left(\frac{P_{\text{avg}}^2}{\tau\nu^2}\right)\left(\pi\frac{NA^2}{hc\lambda}\right)^2$$

Intensity in FD-OCT:

$$I(k, z_{\text{o}}) = \int A_{\text{R}} A_{\text{S}}(z)\cos\left[k(z_{\text{m}} - z_{\text{o}})\right]dz$$

Phase-shifting equation:

$$I(x, y) = a(x, y) + b(x, y)\cos(\varphi + n\Delta\varphi)$$

Three-image algorithm (π/2 shifts):

$$\varphi = \arctan\left(\frac{I_3 - I_2}{I_1 - I_2}\right)$$

Four-image algorithm:

$$\varphi = \arctan\left(\frac{I_4 - I_2}{I_1 - I_3}\right)$$

Five-image algorithm:

$$\varphi = \arctan\left(\frac{2[I_2 - I_4]}{I_1 - I_3 + I_5}\right)$$

Equation Summary (cont'd)

Image reconstruction structure illumination (sectioning):

$$I = \sqrt{\left(I_0 - I_{2\pi/3}\right)^2 + \left(I_0 - I_{4\pi/3}\right)^2 + \left(I_{2\pi/3} - I_{4\pi/3}\right)^2}$$

Poisson statistics:

$$P\left(k, N\right) = \frac{N^k e^{-N}}{k!}$$

Noise:

$$Noise\left[N_{\text{electrons}}\right] = \sqrt{\sigma_{\text{Photon}}^2 + \sigma_{\text{Dark}}^2 + \sigma_{\text{Read}}^2}$$

$$\sigma_{\text{Photon}} = \sqrt{\varphi\eta\tau}\ ,\ \sigma_{\text{Dark}} = \sqrt{I_{\text{Dark}}\tau}\ ,\ \sigma_{\text{Read}} = N_{\text{R}}\ ,\ I_{\text{Dark}}$$

Signal-to-noise ratio (SNR):

$$\text{Signal}\left[N_{\text{electrons}}\right] = \varphi\eta\tau$$

$$\text{SNR} = \frac{\varphi\eta\tau}{\sqrt{\varphi\eta\tau + I_{\text{dark}}\tau + N_{\text{R}}^2}}$$

$$\text{SNR} = \frac{\varphi\eta\tau}{\sqrt{\varphi\eta\tau}} = \sqrt{\varphi\eta\tau}$$

Bibliography

M. Bates, B. Huang, G.T. Dempsey, and X. Zhuang, "Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes." *Science* **317**, 1749 (2007).

J. R. Benford, Microscope objectives, Chapter 4 (p. 178), in *Applied Optics and Optical Engineering, Vol. III,* R. Kingslake, ed., Academic Press, New York, NY (1965).

M. Born and E. Wolf, *Principles of Optics*, Sixth Edition, Cambridge University Press, Cambridge, UK (1997).

S. Bradbury and P. J. Evennett, *Contrast Techniques in Light Microscopy*, BIOS Scientific Publishers, Oxford, UK (1996).

T. Chen, T. Milster, S. K. Park, B. McCarthy, D. Sarid, C. Poweleit, and J. Menendez, "Near-field solid immersion lens microscope with advanced compact mechanical design." *Optical Engineering* **45**(10), 103002 (2006).

T. Chen, T. D. Milster, S. H. Yang, and D. Hansen, "Evanescent imaging with induced polarization by using a solid immersion lens." *Optics Letters* **32**(2), 124–126 (2007).

J.-X. Cheng and X. S. Xie, "Coherent anti-stokes raman scattering microscopy: instrumentation, theory, and applications." *J. Phys. Chem. B* **108***,* 827-840 (2004).

M. A. Choma, M. V. Sarunic, C. Yang, and J. A. Izatt, "Sensitivity advantage of swept source and Fourier domain optical coherence tomography." *Opt. Express* **11**, 2183–2189 (2003).

J. F. de Boer, B. Cense, B. H. Park, M.C. Pierce, G. J. Tearney, and B. E. Bouma, "Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography." *Opt Lett* **28**, 2067–2069 (2003).

E. Dereniak, materials for SPIE Short Course on Imaging Spectrometers, SPIE, Bellingham, WA (2005).

E. Dereniak, *Geometrical Optics*, Cambridge University Press, Cambridge, UK (2008).

M. Descour, materials for OPTI 412, "Optical Instrumentation," University of Arizona (2000).

Bibliography

D. Goldstein, *Polarized Light*, Second Edition, Marcel Dekker, New York, NY (1993).

D. S. Goodman, Basic optical instruments, Chapter 4, in Geometrical and Instrumental Optics, D. Malacara, ed., Academic Press, New York, NY (1988).

J. Goodman, *Introduction to Fourier Optics*, 3rd Edition, Roberts and Company Publishers, Greenwood Village, CO (2004).

E. P. Goodwin and J. C. Wyant, *Field Guide to Interferometric Optical Testing*, SPIE Press, Bellingham, WA (2006).

J. E. Greivenkamp, *Field Guide to Geometrical Optics*, SPIE Press, Bellingham, WA (2004).

H. Gross, F. Blechinger, and B. Achtner, *Handbook of Optical Systems*, *Vol. 4*: *Survey of Optical Instruments*, Wiley-VCH, Germany (2008).

M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution." *PNAS*, **102**(37) 13081–13086 (2005).

M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy." *Journal of Microscopy*, **198**(2), 82-87 (2000).

Gerd Häusler and Michael Walter Lindner, ""Coherence Radar" and "Spectral Radar"—New tools for dermatological diagnosis." *Journal of Biomedical Optics* **3**(1), 21–31 (1998).

E. Hecht, *Optics*, Fourth Edition, Addison-Wesley, Upper Saddle River, New Jersey (2002).

S. W. Hell, "Far-field optical nanoscopy." *Science* **316**, 1153 (2007).

B. Herman and J. Lemasters, *Optical Microscopy: Emerging Methods and Applications*, Academic Press, New York, NY (1993).

P. Hobbs, *Building Electro-Optical Systems: Making It All Work*, Wiley and Sons, New York, NY (2000).

Bibliography

G. Holst and T. Lomheim, *CMOS/CCD Sensors and Camera Systems*, JCD Publishing, Winter Park, FL (2007).

B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy." *Science* **319**, 810 (2008).

R. Huber, M. Wojtkowski, and J. G. Fujimoto, "Fourier domain mode locking (FDML): A new laser operating regime and applications for optical coherence tomography." *Opt. Express* **14**, 3225–3237 (2006).

Invitrogen, http://www.invitrogen.com

R. Jozwicki, *Teoria Odwzorowania Optycznego* (in Polish), PWN (1988).

R. Jozwicki, *Optyka Instrumentalna* (in Polish), WNT (1970).

R. Leitgeb, C. K. Hitzenberger, and A. F. Fercher, "Performance of Fourier-domain versus time-domain optical coherence tomography." *Opt. Express* **11**, 889–894 (2003).

D. Malacara and B. Thompson, Eds., *Handbook of Optical Engineering*, Marcel Dekker, New York, NY (2001).

D. Malacara and Z. Malacara, *Handbook of Optical Design*, Marcel Dekker, New York, NY (1994).

D. Malacara, M. Servin, and Z. Malacara, *Interferogram Analysis for Optical Testing*, Marcel Dekker, New York, NY (1998).

D. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss, Wilmington, DE (2001).

P. Mouroulis and J. Macdonald, *Geometrical Optics and Optical Design*, Oxford University Press, New York, NY (1997).

M. A. A. Neil, R. Juškaitis, and T. Wilson, "Method of obtaining optical sectioning by using structured light in a conventional microscope." *Optics Letters*, **22**(24), 1905–1907 (1997).

Nikon Microscopy U, http://www.microscopyu.com/

Bibliography

C. Palmer (Erwin Loewen, First Edition), *Diffraction Grating Handbook*, Newport Corp. (2005).

K. Patorski, *Handbook of the Moiré Fringe Technique*, Elsevier, Oxford, UK(1993).

J. Pawley, Ed., *Biological Confocal Microscopy*, Third Edition, Springer, New York, NY (2006).

M. C. Pierce, D. J. Javier and R. Richards-Kortum, "Optical contrast agents and imaging systems for detection and diagnosis of cancer." *Int. J. Cancer* **123**, 1979–1990 (2008).

M. Pluta, *Advanced Light Microscopy, Volume One: Principle and Basic Properties*, PWN and Elsevier, New York, NY (1988).

M. Pluta, *Advanced Light Microscopy Volume Two: Specialized Methods*, PWN and Elsevier, New York, NY (1989).

M. Pluta, *Advanced Light Microscopy, Volume Three: Measuring Techniques*, PWN, Warsaw, Poland; and North Holland, Amsterdam, Holland (1993).

E. O. Potma, C. L. Evans, and X. S. Xie, "Heterodyne coherent anti-Stokes Raman scattering (CARS) imaging." *Optics Letters*, **31**(2), 241–243 (2006).

D. W. Robinson, G. T.Reed,, Eds., *Interferogram Analysis*, IOP Publishing, Bristol, UK (1993).

M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)." *Nature Methods*, **3**, 793–796 (2006).

B. Saleh and M. C. Teich, *Fundamentals of Photonics*, SecondEdition, Wiley, New York, NY (2007).

J. Schwiegerling, *Field Guide to Visual and Ophthalmic Optics*, SPIE Press, Bellingham, WA (2004).

W. Smith, *Modern Optical Engineering,* Third Edition, McGraw-Hill, New York, NY (2000).

D. Spector and R. Goldman, Eds., *Basic Methods in Microscopy*, Cold Spring Harbor Laboratory Press, Woodbury, NY (2006).

Bibliography

Thorlabs website resources: http://www.thorlabs.com/

P. Török and F. J. Kao, Eds., *Optical Imaging and Microscopy*, Springer, New York, NY (2007).

Veeco Optical Library entry, http://www.veeco.com/pdf/Optical_Library/Challenges_in_White_Light.pdf

R. Wayne, *Light and Video Microscopy*, Elsevier (reprinted by Academic Press), New York, NY (2009).

H. Yu, P. C. Cheng, P. C. Li, F. J. Kao, Eds., *Multi Modality Microscopy*, World Scientific, Hackensack, NJ (2006).

S. H. Yun, G. J. Tearney, B. J. Vakoc, M. Shishkov, W. Y. Oh, A. E. Desjardins, M. J. Suter, R. C. Chan, J. A. Evans, I. K. Jang, N. S. Nishioka, J. F. de Boer, and B. E. Bouma, "Comprehensive volumetric optical microscopy *in vivo*." *Nature Med* **12**, 1429–1433 (2006).

Zeiss Corporation, http://www.zeiss.com

Index

4Pi, 64, 109, 110
4Pi microscopy, 109

Abbe number, 25
Abbe resolution limit, 39
absorption filter, 60
achromatic objectives, 51
achromats, 51
Airy disk, 38
Amici lens, 55
amplitude contrast, 72
amplitude object, 63
amplitude ratio, 15
amplitude splitting, 17
analyzer, 83
anaxial illumination, 73
angular frequency, 4
angular magnification, 23
angular resolving power, 40
anisotropic media, propagation of light in, 9
aperture stop, 24
apochromats, 51
arc lamp, 58
array microscope, 113
array microscopy, 64, 113
astigmatism, 28
axial chromatic aberration, 26
axial resolution, 86

bandpass filter, 60
Bayer mask, 117
Bertrand lens, 45, 55
bias, 80
Brace Köhler compensator, 85
bright field, 64, 65, 76

charge-coupled device (CCD), 115
chief ray, 24
chromatic aberration, 26
chromatic resolving power, 20
circularly polarized light, 10
coherence length, 11
coherence scanning interferometry, 98
coherence time, 11
coherent, 11
coherent anti-Stokes Raman scattering (CARS), 64, 111
coma, 27
compensators, 83, 85
complex amplitudes, 12
compound microscope, 31
condenser's diaphragm, 46
cones, 32
confocal microscopy, 64, 86
contrast, 12
contrast of fringes, 15
cornea, 32
cover glass, 56
cover slip, 56
critical angle, 7
critical illumination, 46
cutoff frequency, 30

dark current, 118
dark field, 64, 65
dark noise, 118
de Sénarmont compensator, 85
depth of field, 41
depth of focus, 41
depth perception, 47
destructive interference, 19

DIA, 44
DIC microscope design, 80
dichroic beam splitter, 91
dichroic mirror, 91
dielectric constant, 3
differential interference contrast (DIC), 64–65, 67, 79
diffraction, 18
diffraction grating, 19
diffraction orders, 19
digital microscopy, 114
digitiziation of CCD, 119
dispersion, 25
dispersion staining, 78
distortion, 28
dynamic range, 119

edge filter, 60
effective focal length, 22
electric fields, 12
electric vector, 10
electromagnetic spectrum, 2
electron-multiplying CCDs, 117
emission filter, 91
empty magnification, 40
entrance pupil, 24
entrance window, 24
epi-illumination, 44
epithelial cells, 2
evanescent wave microscopy, 105
excitation filter, 91
exit pupil, 24
exit window, 24
extraordinary wave, 9
eye, 32, 46
eye point, 48
eye relief, 48
eye's pupil, 46
eyepiece, 48

Fermat's principle, 5
field curvature, 28
field diaphragm, 46
field number (FN), 34, 48
field stop, 24, 34
filter turret, 92
filters, 91
finite tube length, 34
five-image technique, 102
fluorescence, 90
fluorescence microscopy, 64
fluorescent emission, 94
fluorites, 51
fluorophore quantum yield, 94
focal length, 21
focal plane, 21
focal point, 21
Fourier-domain OCT (FDOCT), 99
four-image technique, 102
fovea, 32
frame-transfer architecture, 116
Fraunhofer diffraction, 18
frequency of light, 1
Fresnel diffraction, 18
Fresnel reflection, 6
frustrated total internal reflection, 7
full-frame architecture, 116

gas-arc discharge lamp, 58
Gaussian imaging equation, 22
geometrical aberrations, 25

Index

geometrical optics, 21
Goos-Hänchen effect, 8
grating equation, 19–20

half bandwidth (HBW), 16
halo effect, 65, 71
halogen lamp, 58
high-eye point, 49
Hoffman modulation contrast, 75
Huygens, 48–49

I5M, 64
image comparison, 65
image formation, 22
image space, 21
immersion liquid, 57
incandescent lamp, 58
incidence, 6
infinity-corrected objective, 35
infinity-corrected systems, 35
infrared radiation (IR), 2
intensity of light, 12
interference, 12
interference filter, 16, 60
interference microscopy, 64, 98
interference order, 16
interfering beams, 15
interferometers, 17
interline transfer architecture, 116
intermediate image plane, 46
inverted microscope, 33
iris, 32

Köhler illumination, 43–45

laser scanning, 64
lasers, 96
laser-scanning microscopy (LSM), 97
lateral magnification, 23
lateral resolution, 71
laws of reflection and refraction, 6
lens, 32
light sheet, 112
light sources for microscopy, common, 58
light-emitting diodes (LEDs), 59
limits of light microscopy, 110
linearly polarized light, 10
line-scanning confocal microscope, 87
Linnik, 101
long working distance objectives, 53
longitudinal magnification, 23
low magnification objectives, 54

Mach-Zehnder interferometer, 17
macula, 32
magnetic permeability, 3
magnification, 21, 23
magnifier, 31
magnifying power (MP), 37
marginal ray, 24
Maxwell's equations, 3
medium permittivity, 3
mercury lamp, 58
metal halide lamp, 58
Michelson, 17, 53, 98, 100, 101

minimum focus distance, 32
Mirau, 53, 98, 101
modulation contrast microscopy (MCM), 74
modulation transfer function (MTF), 29
molecular absorption cross section, 94
monochromatic light, 11
multi-photon fluorescence, 95
multi-photon microscopy, 64
multiple beam interference, 16

nature of light, 1
near point, 32
negative birefringence, 9
negative contrast, 69
neutral density (ND), 60
Newtonian equation, 22
Nomarski interference microscope, 82
Nomarski prism, 62, 80
non-self-luminous, 63
numerical aperture (NA), 38, 50
Nyquist frequency, 120

object space, 21
objective, 31
objective correction, 50
oblique illumination, 73
optic axis, 9
optical coherence microscopy (OCM), 98
optical coherence tomography (OCT), 98
optical density (OD), 60
optical path difference (OPD), 5
optical path length (OPL), 5
optical power, 22
optical profilometry, 98, 101
optical rays, 21
optical sectioning, 103
optical spaces, 21
optical staining, 77
optical transfer function, 29
optical tube length, 34
ordinary wave, 9
oversampling, 121

PALM, 64, 110
paraxial optics, 21
parfocal distance, 34
partially coherent, 11
particle model, 1
performance metrics, 29
phase-advancing objects, 68
phase contrast microscope, 70
phase contrast, 64, 65, 66
phase object, 63
phase of light, 4
phase retarding objects, 68
phase-shifting algorithms, 102
phase-shifting interferometry (PSI), 98, 100
photobleaching, 94
photodiode, 115
photon noise, 118
Planck's constant, 1

point spread function (PSF), 29
point-scanning confocal microscope, 87
Poisson statistics, 118
polarization microscope, 84
polarization microscopy, 64, 83
polarization of light, 10
polarization prisms, 61
polarization states, 10
polarizers, 61–62
positive birefringence, 9
positive contrast, 69
pupil function, 29

quantum efficiency (QE), 97, 117
quasi-monochromatic, 11
quenching, 94

Raman microscopy, 64, 111
Raman scattering, 111
Ramsden eyepiece, 48
Rayleigh resolution limit, 39
read noise, 118
real image, 22
red blood cells, 2
reflectance coefficients, 6
reflected light, 6, 16
reflection law, 6
reflective grating, 20
refracted light, 6
refraction law, 6
refractive index, 4
RESOLFT, 64, 110

resolution limit, 39
retardation, 83
retina, 32
RGB filters, 117
Rheinberg illumination, 77
rods, 32

Sagnac interferometer, 17
sample conjugate, 46
sample path, 44
scanning approaches, 87
scanning white light interferometry, 98
Selective plane illumination microscopy (SPIM), 64, 112
self-luminous, 63
semi-apochromats, 51
shading-off effect, 71
shearing interferometer, 17
shearing interferometry, 79
shot noise, 118
SI, 64
sign convention, 21
signal-to-noise ratio (SNR), 119
Snell's law, 6
solid immersion, 106
solid immersion lenses (SILs), 106
Sparrow resolution limit, 30
spatial coherence, 11
spatial resolution, 86
spectral-domain OCT (SDOCT), 99
spectrum of microscopy, 2
spherical aberration, 27
spinning-disc confocal imaging, 88
stereo microscopes, 47

stimulated emission depletion (STED), 107
stochastic optical reconstruction microscopy (STORM), 64, 108, 110
Stokes shift, 90
Strehl ratio, 29, 30
structured illumination, 103
super-resolution microscopy, 64
swept-source OCT (SSOCT), 99

telecentricity, 36
temporal coherence, 11, 13–14
thin lenses, 21
three-image technique, 102
total internal reflection (TIR), 7
total internal reflection fluorescence (TIRF) microscopy, 105
transmission coefficients, 6
transmission grating, 20
transmitted light, 16
transverse chromatic aberration, 26
transverse magnification, 23
tube length, 34
tube lens, 35
tungsten-argon lamp, 58

ultraviolet radiation (UV), 2
undersampling, 120
uniaxial crystals, 9
upright microscope, 33
useful magnification, 40
UV objectives, 54

velocity of light, 1
vertical scanning interferometry (VSI), 98, 100
virtual image, 22
visibility, 12
visibility in phase contrast, 69
visible spectrum (VIS), 2
visual stereo resolving power, 47

water immersion objectives, 54
wave aberrations, 25
wave equations, 3
wave group, 11
wave model, 1
wave number, 4
wave plate compensator, 85
wavefront propagation, 4
wavefront splitting, 17
wavelength of light, 1
Wollaston prism, 61, 80
working distance (WD), 50

x-ray radiation, 2

Tomasz S. Tkaczyk is an Assistant Professor of Bioengineering and Electrical and Computer Engineering at Rice University, Houston, Texas, where he develops modern optical instrumentation for biological and medical applications. His primary research is in microscopy, including endo-microscopy, cost-effective high-performance optics for diagnostics, and multi-dimensional imaging (snapshot hyperspectral microscopy and spectro-polarimetry).

Professor Tkaczyk received his M.S. and Ph.D. from the Institute of Micromechanics and Photonics, Department of Mechatronics, Warsaw University of Technology, Poland. Beginning in 2003, after his postdoctoral training, he worked as a research professor at the College of Optical Sciences, University of Arizona. He joined Rice University in the summer of 2007.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.1

Fundamental Optics 1

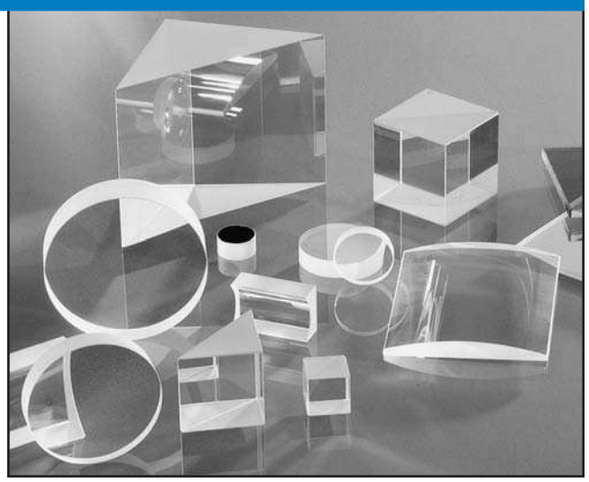

Introduction	**1.2**
Paraxial Formulas	**1.3**
Imaging Properties of Lens Systems	**1.6**
Lens Combination Formulas	**1.8**
Performance Factors	**1.11**
Lens Shape	**1.17**
Lens Combinations	**1.18**
Diffraction Effects	**1.20**
Lens Selection	**1.23**
Spot Size	**1.26**
Aberration Balancing	**1.27**
Definition of Terms	**1.29**
Paraxial Lens Formulas	**1.32**
Principal-Point Locations	**1.36**
Prisms	**1.37**
Polarization	**1.41**
Waveplates	**1.46**
Etalons	**1.49**
Ultrafast Theory	**1.52**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.2

Introduction

The process of solving virtually any optical engineering problem can be broken down into two main steps. First, paraxial calculations (first order) are made to determine critical parameters such as magnification, focal length(s), clear aperture (diameter), and object and image position. These paraxial calculations are covered in the next section of this chapter.

Second, actual components are chosen based on these paraxial values, and their actual performance is evaluated with special attention paid to the effects of aberrations. A truly rigorous performance analysis for all but the simplest optical systems generally requires computer ray tracing, but simple generalizations can be used, especially when the lens selection process is confined to a limited range of component shapes.

In practice, the second step may reveal conflicts with design constraints, such as component size, cost, or product availability. System parameters may therefore require modification.

Because some of the terms used in this chapter may not be familiar to all readers, a glossary of terms is provided in *Definition of Terms*.

Finally, it should be noted that the discussion in this chapter relates only to systems with uniform illumination; optical systems for Gaussian beams are covered in *Gaussian Beam Optics*.

Engineering Support

CVI Melles Griot maintains a staff of knowledgeable, experienced applications engineers at each of our facilities worldwide. The information given in this chapter is sufficient to enable the user to select the most appropriate catalog lenses for the most commonly encountered applications. However, when additional optical engineering support is required, our applications engineers are available to provide assistance. Do not hesitate to contact us for help in product selection or to obtain more detailed specifications on CVI Melles Griot products.

THE OPTICAL ENGINEERING PROCESS

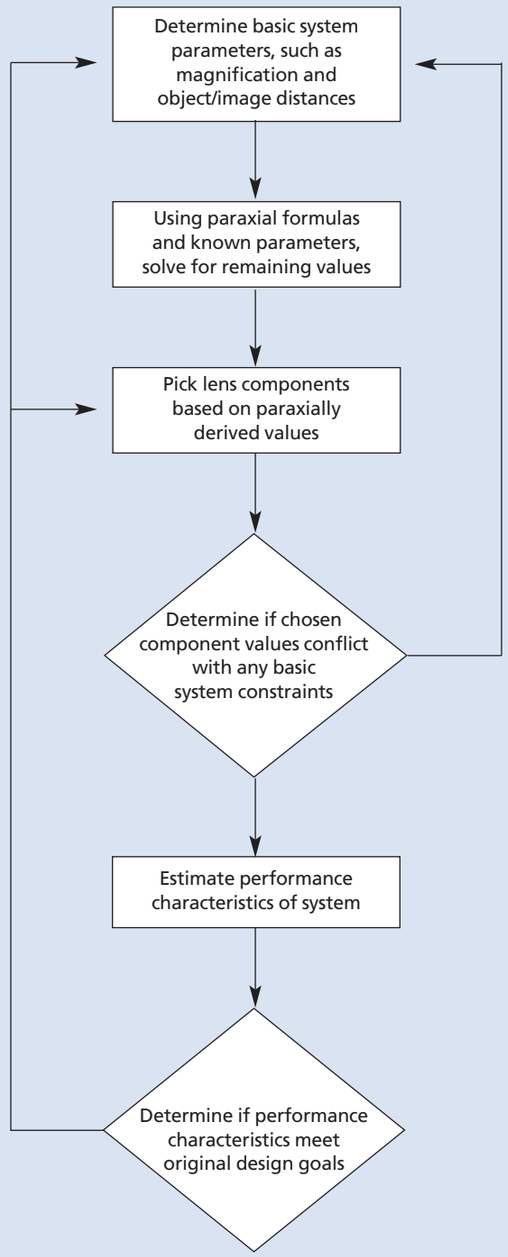

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.3

Paraxial Formulas

Sign Conventions

The validity of the paraxial lens formulas is dependent on adherence to the following sign conventions:

For lenses: (refer to figure 1.1)

s is + for object to left of H (the first principal point)

s is − for object to right of H

s'' is + for image to right of H″ (the second principal point)

s'' is − for image to left of H″

m is + for an inverted image

m is − for an upright image

For mirrors:

f is + for convex (diverging) mirrors

f is − for concave (converging) mirrors

s is + for object to left of H

s is − for object to right of H

s'' is − for image to right of H″

s'' is + for image to left of H″

m is + for an inverted image

m is − for an upright image

When using the thin-lens approximation, simply refer to the left and right of the lens.

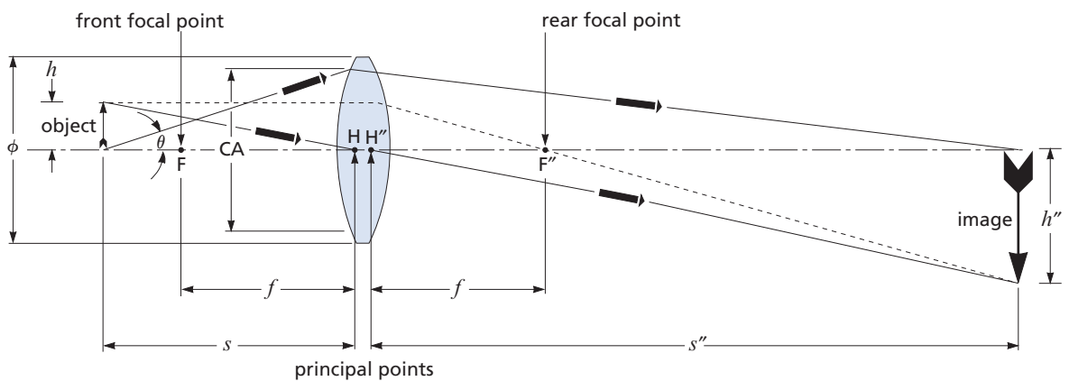

Note location of object and image relative to front and rear focal points.

ϕ = lens diameter

CA = clear aperture (typically 90% of ϕ)

f = effective focal length (EFL) which may be positive (as shown) or negative. f represents both FH and H″F″, assuming lens is surrounded by medium of index 1.0

m = $s''/s = h''/h$ = magnification or conjugate ratio, said to be infinite if either s'' or s is infinite

θ = arcsin (CA/2s)

s = object distance, positive for object (whether real or virtual) to the left of principal point H

s'' = image distance (s and s'' are collectively called conjugate distances, with object and image in conjugate planes), positive for image (whether real or virtual) to the right of principal point H″

h = object height

h'' = image height

Figure 1.1 **Sign conventions**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.4

Typically, the first step in optical problem solving is to select a system focal length based on constraints such as magnification or conjugate distances (object and image distance). The relationship among focal length, object position, and image position is given by

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{s''} \tag{1.1}$$

This formula is referenced to figure 1.1 and the sign conventions given in *Sign Conventions*.

By definition, magnification is the ratio of image size to object size or

$$m = \frac{s''}{s} = \frac{h''}{h}. \tag{1.2}$$

This relationship can be used to recast the first formula into the following forms:

$$f = m\frac{(s+s'')}{(m+1)^2} \tag{1.3}$$

$$f = \frac{sm}{m+1} \tag{1.4}$$

$$f = \frac{s+s''}{m+2+\frac{1}{m}} \tag{1.5}$$

$$s(m+1) = s+s'' \tag{1.6}$$

where $(s+s'')$ is the approximate object-to-image distance.

With a real lens of finite thickness, the image distance, object distance, and focal length are all referenced to the principal points, not to the physical center of the lens. By neglecting the distance between the lens' principal points, known as the hiatus, $s+s''$ becomes the object-to-image distance. This simplification, called the thin-lens approximation, can speed up calculation when dealing with simple optical systems.

Example 1: Object outside Focal Point

A 1-mm-high object is placed on the optical axis, 200 mm left of the left principal point of a LDX-25.0-51.0-C ($f = 50$ mm). Where is the image formed, and what is the magnification? (See figure 1.2.)

$$\frac{1}{s''} = \frac{1}{f} - \frac{1}{s}$$

$$\frac{1}{s''} = \frac{1}{50} - \frac{1}{200}$$

$$s'' = 66.7 \text{ mm}$$

$$m = \frac{s''}{s} = \frac{66.7}{200} = 0.33$$

or real image is 0.33 mm high and inverted.

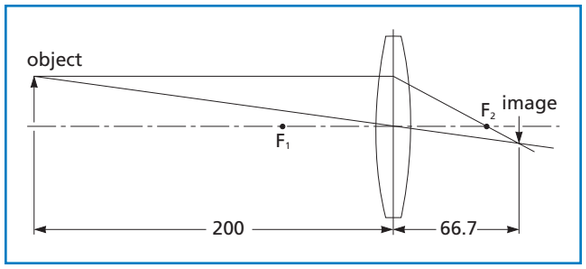

Figure 1.2 **Example 1** ($f = 50$ mm, $s = 200$ mm, $s'' = 66.7$ mm)

Example 2: Object inside Focal Point

The same object is placed 30 mm left of the left principal point of the same lens. Where is the image formed, and what is the magnification? (See figure 1.3.)

$$\frac{1}{s''} = \frac{1}{50} - \frac{1}{30}$$

$$s'' = -75 \text{ mm}$$

$$m = \frac{s''}{s} = \frac{-75}{30} = -2.5$$

or virtual image is 2.5 mm high and upright.

In this case, the lens is being used as a magnifier, and the image can be viewed only back through the lens.

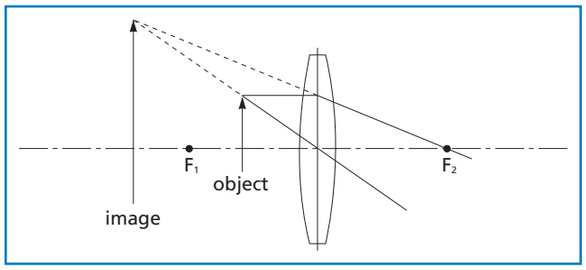

Figure 1.3 **Example 2** ($f = 50$ mm, $s = 30$ mm, $s'' = -75$ mm)

Example 3: Object at Focal Point

A 1-mm-high object is placed on the optical axis, 50 mm left of the first principal point of an LDK-50.0-52.2-C ($f = -50$ mm). Where is the image formed, and what is the magnification? (See figure 1.4.)

$$\frac{1}{s''} = \frac{1}{-50} - \frac{1}{50}$$

$$s'' = -25 \text{ mm}$$

$$m = \frac{s''}{s} = \frac{-25}{50} = -0.5$$

or virtual image is 0.5 mm high and upright.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.5

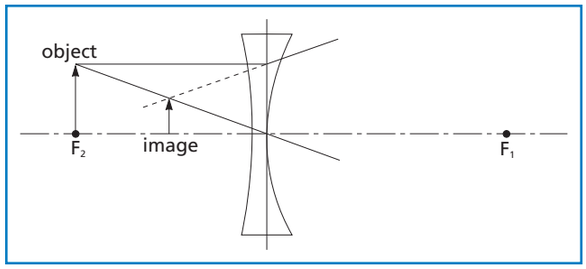

Figure 1.4 **Example 3** ($f = -50$ mm, $s = 50$ mm, $s'' = -25$ mm)

A simple graphical method can also be used to determine paraxial image location and magnification. This graphical approach relies on two simple properties of an optical system. First, a ray that enters the system parallel to the optical axis crosses the optical axis at the focal point. Second, a ray that enters the first principal point of the system exits the system from the second principal point parallel to its original direction (i.e., its exit angle with the optical axis is the same as its entrance angle). This method has been applied to the three previous examples illustrated in figures 1.2 through 1.4. Note that by using the thin-lens approximation, this second property reduces to the statement that a ray passing through the center of the lens is undeviated.

F-NUMBER AND NUMERICAL APERTURE

The paraxial calculations used to determine the necessary element diameter are based on the concepts of focal ratio (f-number or f/#) and numerical aperture (NA). The f-number is the ratio of the focal length of the lens to its "effective" diameter, the clear aperture (CA).

$$\text{f-number} = \frac{f}{\text{CA}} \cdot \tag{1.7}$$

To visualize the f-number, consider a lens with a positive focal length illuminated uniformly with collimated light. The f-number defines the angle of the cone of light leaving the lens which ultimately forms the image. This is an important concept when the throughput or light-gathering power of an optical system is critical, such as when focusing light into a monochromator or projecting a high-power image.

The other term used commonly in defining this cone angle is numerical aperture. The NA is the sine of the angle made by the marginal ray with the optical axis. By referring to figure 1.5 and using simple trigonometry, it can be seen that

$$\text{NA} = \sin\theta = \frac{\text{CA}}{2f} \tag{1.8}$$

and

$$\text{NA} = \frac{1}{2(\text{f-number})}. \tag{1.9}$$

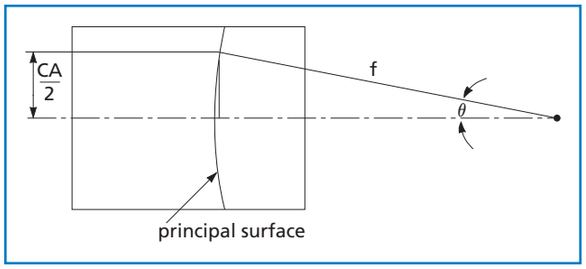

Figure 1.5 **F-number and numerical aperture**

Ray f-numbers can also be defined for any arbitrary ray if its conjugate distance and the diameter at which it intersects the principal surface of the optical system are known.

NOTE

Because the sign convention given previously is not used universally in all optics texts, the reader may notice differences in the paraxial formulas. However, results will be correct as long as a consistent set of formulas and sign conventions is used.

1ch_FundamentalOptics_Final_a.qxd 7/6/2009 1:42 PM Page 1.6

Imaging Properties of Lens Systems

THE OPTICAL INVARIANT

To understand the importance of the NA, consider its relation to magnification. Referring to figure 1.6,

$$\text{NA (object side)} = \sin\theta = \frac{\text{CA}}{2s} \tag{1.10}$$

$$\text{NA}'' \text{ (image side)} = \sin\theta'' = \frac{\text{CA}}{2s''} \tag{1.11}$$

which can be rearranged to show

$$\text{CA} = 2s\ \sin\theta \tag{1.12}$$

and

$$\text{CA} = 2s''\sin\theta'' \tag{1.13}$$

leading to

$$\frac{s''}{s} = \frac{\sin\theta}{\sin\theta''} = \frac{\text{NA}}{\text{NA}''}. \tag{1.14}$$

Since $\frac{s''}{s}$ is simply the magnification of the system,

we arrive at

$$m = \frac{\text{NA}}{\text{NA}''}. \tag{1.15}$$

The magnification of the system is therefore equal to the ratio of the NAs on the object and image sides of the system. This powerful and useful result is completely independent of the specifics of the optical system, and it can often be used to determine the optimum lens diameter in situations involving aperture constraints.

When a lens or optical system is used to create an image of a source, it is natural to assume that, by increasing the diameter (ϕ) of the lens, thereby increasing its CA, we will be able to collect more light and thereby produce a brighter image. However, because of the relationship between magnification and NA, there can be a theoretical limit beyond which increasing the diameter has no effect on light-collection efficiency or image brightness.

Since the NA of a ray is given by CA/2s, once a focal length and magnification have been selected, the value of NA sets the value of CA. Thus, if one is dealing with a system in which the NA is constrained on either the object or image side, increasing the lens diameter beyond this value will increase system size and cost but will not improve performance (i.e., throughput or image brightness). This concept is sometimes referred to as the optical invariant.

SAMPLE CALCULATION

To understand how to use this relationship between magnification and NA, consider the following example.

Example: System with Fixed Input NA

Two very common applications of simple optics involve coupling light into an optical fiber or into the entrance slit of a monochromator. Although these problems appear to be quite different, they both have the same limitation — they have a fixed NA. For monochromators, this limit is usually expressed in terms of the f-number. In addition to the fixed NA, they both have a fixed entrance pupil (image) size.

Suppose it is necessary, using a singlet lens, to couple the output of an incandescent bulb with a filament 1 mm in diameter into an optical fiber as shown in figure 1.7. Assume that the fiber has a core diameter of 100 μm and an NA of 0.25, and that the design requires that the total distance from the source to the fiber be 110 mm. Which lenses are appropriate?

By definition, the magnification must be 0.1. Letting $s + s''$ total 110 mm (using the thin-lens approximation), we can use equation 1.3,

$$f = m\frac{(s+s'')}{(m+1)^2}, \qquad \text{(see eq. 1.3)}$$

to determine that the focal length is 9.1 mm. To determine the conjugate distances, s and s'', we utilize equation 1.6,

$$s(m+1) = s + s'', \qquad \text{(see eq. 1.6)}$$

and find that s = 100 mm and s'' = 10 mm.

We can now use the relationship NA = CA/2s or NA″ = CA/2s'' to derive CA, the optimum clear aperture (effective diameter) of the lens.

With an image NA of 0.25 and an image distance (s'') of 10 mm,

$$0.25 = \frac{\text{CA}}{20}$$

and

$$\text{CA} = 5 \text{ mm}.$$

Accomplishing this imaging task with a single lens therefore requires an optic with a 9.1-mm focal length and a 5-mm diameter. Using a larger diameter lens will not result in any greater system throughput because of the limited input NA of the optical fiber. The singlet lenses in this catalog that meet these criteria are LPX-5.0-5.2-C, which is plano-convex, and LDX-6.0-7.7-C and LDX-5.0-9.9-C, which are biconvex.

Making some simple calculations has reduced our choice of lenses to just three. The following chapter, *Gaussian Beam Optics*, discusses how to make a final choice of lenses based on various performance criteria.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.7

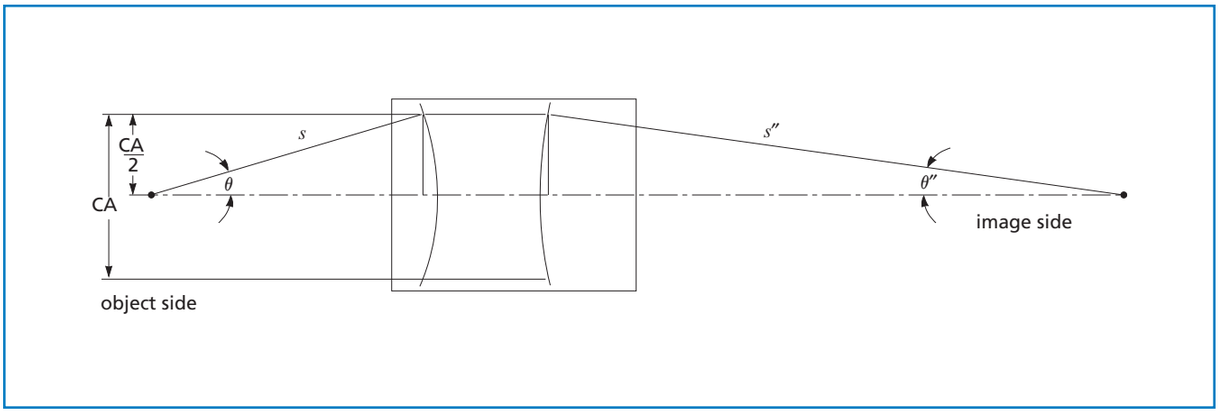

Figure 1.6 **Numerical aperture and magnification**

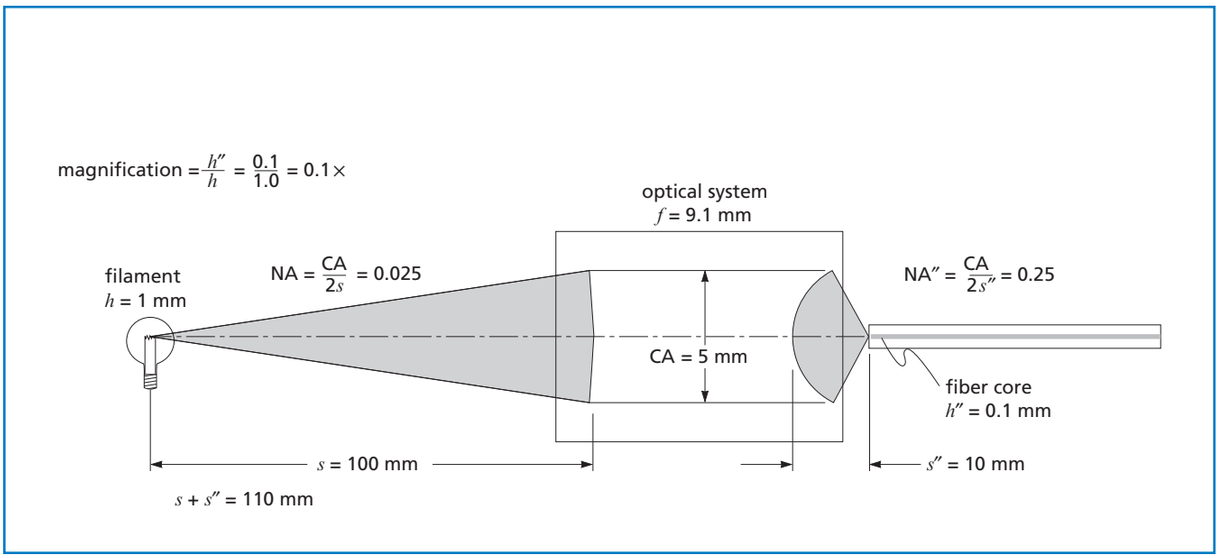

Figure 1.7 **Optical system geometry for focusing the output of an incandescent bulb into an optical fiber**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.8

Lens Combination Formulas

Many optical tasks require several lenses in order to achieve an acceptable level of performance. One possible approach to lens combinations is to consider each image formed by each lens as the object for the next lens and so on. This is a valid approach, but it is time consuming and unnecessary.

It is much simpler to calculate the effective (combined) focal length and principal-point locations and then use these results in any subsequent paraxial calculations (see figure 1.8). They can even be used in the optical invariant calculations described in the preceding section.

EFFECTIVE FOCAL LENGTH

The following formulas show how to calculate the effective focal length and principal-point locations for a combination of any two arbitrary components. The approach for more than two lenses is very simple: Calculate the values for the first two elements, then perform the same calculation for this combination with the next lens. This is continued until all lenses in the system are accounted for.

The expression for the combination focal length is the same whether lens separation distances are large or small and whether f_1 and f_2 are positive or negative:

$$f = \frac{f_1 f_2}{f_1 + f_2 - d}. \qquad (1.16)$$

This may be more familiar in the form

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}. \qquad (1.17)$$

Notice that the formula is symmetric with respect to the interchange of the lenses (end-for-end rotation of the combination) at constant *d*. The next two formulas are not.

COMBINATION FOCAL-POINT LOCATION

For all values of f_1, f_2, and d, the location of the focal point of the combined system (s_2''), measured from the secondary principal point of the second lens (H_2''), is given by

$$s_2'' = \frac{f_2(f_1 - d)}{f_1 + f_2 - d}. \qquad (1.18)$$

This can be shown by setting $s_1 = d - f_1$ (see figure 1.8a), and solving

$$\frac{1}{f_2} = \frac{1}{s_1} + \frac{1}{s_2''}$$

for s_2''.

Symbols

f_c = combination focal length (EFL), positive if combination final focal point falls to the right of the combination secondary principal point, negative otherwise (see figure 1.8c).

f_1 = focal length of the first element (see figure 1.8a).

f_2 = focal length of the second element.

d = distance from the secondary principal point of the first element to the primary principal point of the second element, positive if the primary principal point is to the right of the secondary principal point, negative otherwise (see figure 1.8b).

s_1'' = distance from the primary principal point of the first element to the final combination focal point (location of the final image for an object at infinity to the right of both lenses), positive if the focal point is to left of the first element's primary principal point (see figure 1.8d).

s_2'' = distance from the secondary principal point of the second element to the final combination focal point (location of the final image for an object at infinity to the left of both lenses), positive if the focal point is to the right of the second element's secondary principal point (see figure 1.8b).

z_H = distance to the combination primary principal point measured from the primary principal point of the first element, positive if the combination secondary principal point is to the right of secondary principal point of second element (see figure 1.8d).

z_H'' = distance to the combination secondary principal point measured from the secondary principal point of the second element, positive if the combination secondary principal point is to the right of the secondary principal point of the second element (see figure 1.8c).

Note: *These paraxial formulas apply to* ***coaxial*** *combinations of both thick and thin lenses immersed in air or any other fluid with refractive index independent of position. They assume that light propagates from left to right through an optical system.*

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.9

COMBINATION SECONDARY PRINCIPAL-POINT LOCATION

Because the thin-lens approximation is obviously highly invalid for most combinations, the ability to determine the location of the secondary principal point is vital for accurate determination of d when another element is added. The simplest formula for this calculates the distance from the secondary principal point of the final (second) element to the secondary principal point of the combination (see figure 1.8b):

$$z = s_2'' - f \,. \qquad (1.19)$$

COMBINATION EXAMPLES

It is possible for a lens combination or system to exhibit principal planes that are far removed from the system. When such systems are themselves combined, negative values of d may occur. Probably the simplest example of a negative d-value situation is shown in figure 1.9. Meniscus lenses with steep surfaces have external principal planes. When two of these lenses are brought into contact, a negative value of d can occur. Other combined-lens examples are shown in figures 1.10 through 1.13.

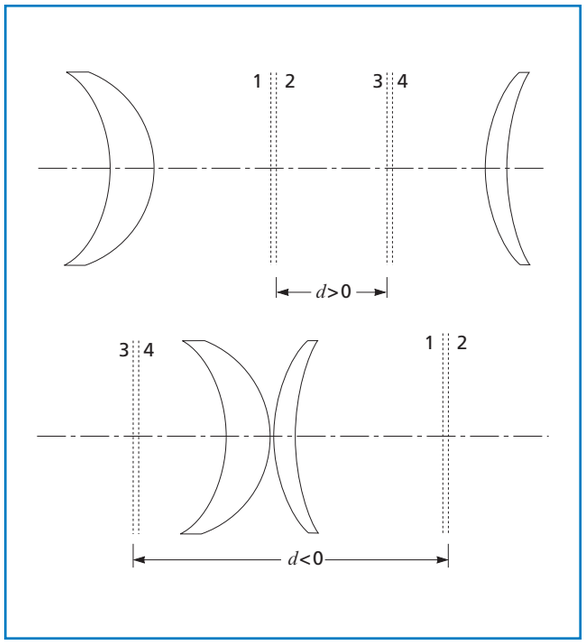

Figure 1.9 **"Extreme" meniscus-form lenses with external principal planes** (drawing not to scale)

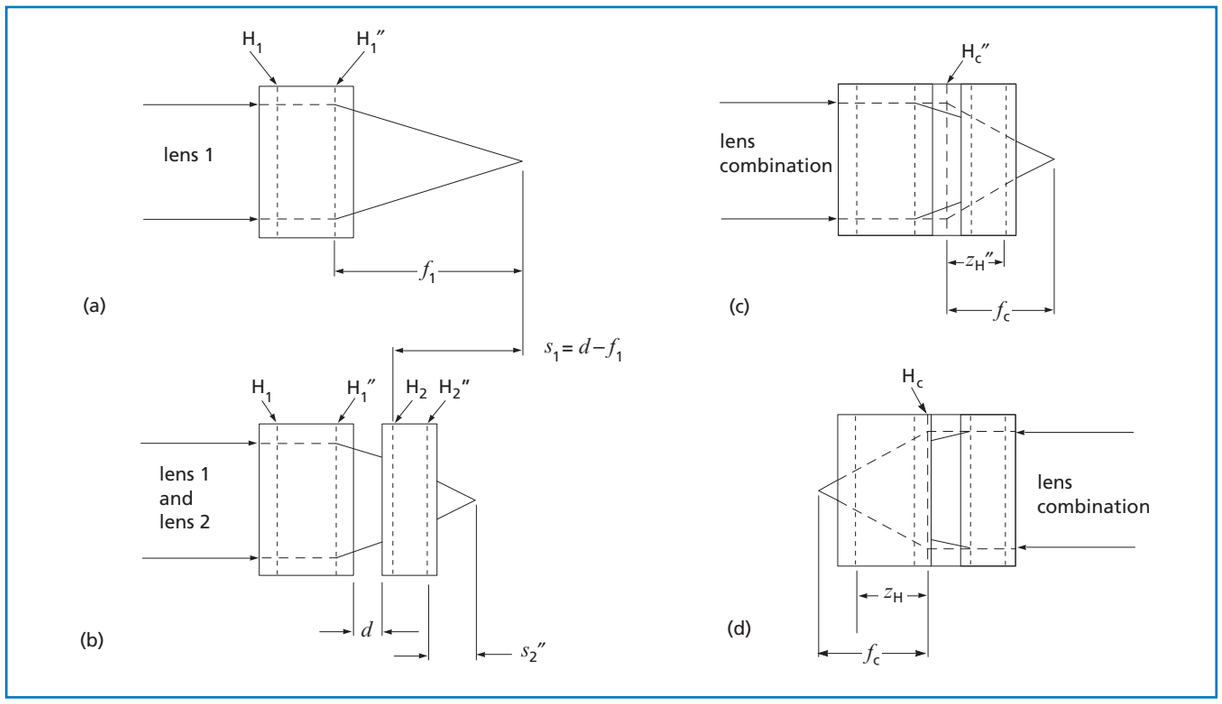

Figure 1.8 **Lens combination focal length and principal planes**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 4:09 PM Page 1.10

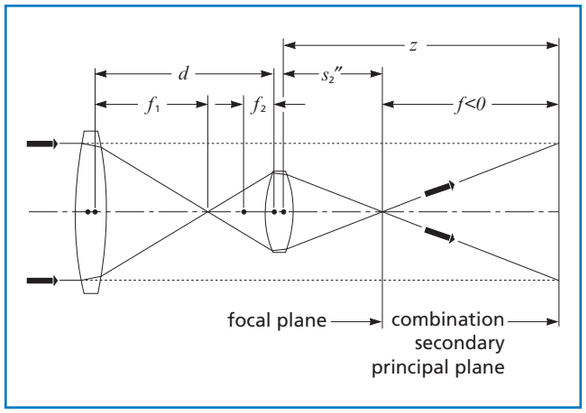

Figure 1.10 **Positive lenses separated by distance greater than $f_1 + f_2$:** f is negative and both s_2'' and z are positive. Lens symmetry is not required.

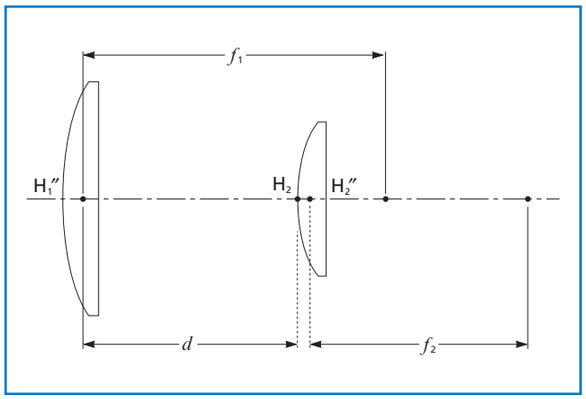

Figure 1.11 **Achromatic combinations:** Air-spaced lens combinations can be made nearly achromatic, even though both elements are made from the same material. Achieving achromatism requires that, in the thin-lens approximation,

$$d = \frac{(f_1 + f_2)}{2}.$$

This is the basis for Huygens and Ramsden eyepieces.

This approximation is adequate for most thick-lens situations. The signs of f_1, f_2, and d are unrestricted, but d must have a value that guarantees the existence of an air space. Element shapes are unrestricted and can be chosen to compensate for other aberrations.

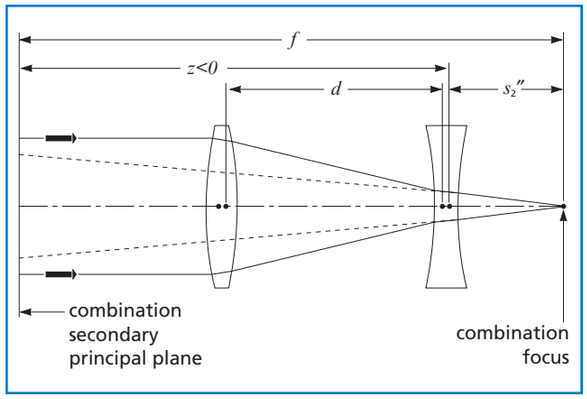

Figure 1.12 **Telephoto combination:** The most important characteristic of the telephoto lens is that the EFL, and hence the image size, can be made much larger than the distance from the first lens surface to the image would suggest by using a positive lens followed by a negative lens (but not necessarily the lens shapes shown in the figure). For example, f_1 is positive and $f_2 = -f_1/2$. Then f is negative for d less than $f_1/2$, infinite for $d = f_1/2$ (Galilean telescope or beam expander), and positive for d larger than $f_1/2$. To make the example even more specific, catalog lenses LDX-50.8-130.4-C and LDK-42.0-52.2-C, with $d = 78.2$ mm, will yield $s_2'' = 2.0$ m, $f = 5.2$ m, and $z = 43.2$ m.

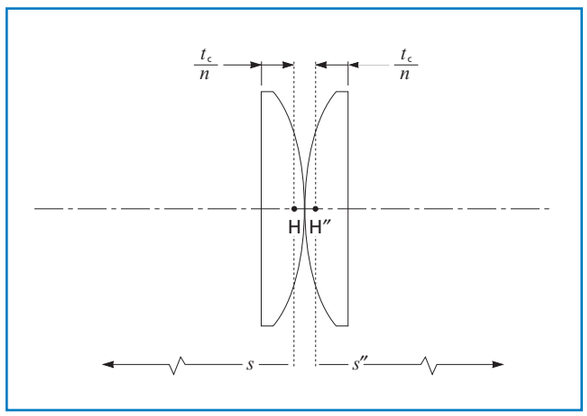

Figure 1.13 **Condenser configuration:** The convex vertices of a pair of identical plano-convex lenses are on contact. (The lenses could also be plano aspheres.) Because $d = 0$, $f = f_1/2 = f_2/2$, $f_1/2 = s_2''$, and $z = 0$. The secondary principal point of the second element and the secondary principal point of the combination coincide at H″, at depth t_c/n beneath the vertex of the plano surface of the second element, where t_c is the element center thickness and n is the refractive index of the element. By symmetry, the primary principal point of the combination is similarly located in the first element. Combination conjugate distances must be measured from these points.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.11

Performance Factors

After paraxial formulas have been used to select values for component focal length(s) and diameter(s), the final step is to select actual lenses. As in any engineering problem, this selection process involves a number of tradeoffs, including performance, cost, weight, and environmental factors.

The performance of real optical systems is limited by several factors, including lens aberrations and light diffraction. The magnitude of these effects can be calculated with relative ease.

Numerous other factors, such as lens manufacturing tolerances and component alignment, impact the performance of an optical system. Although these are not considered explicitly in the following discussion, it should be kept in mind that if calculations indicate that a lens system only just meets the desired performance criteria, in practice it may fall short of this performance as a result of other factors. In critical applications, it is generally better to select a lens whose calculated performance is significantly better than needed.

DIFFRACTION

Diffraction, a natural property of light arising from its wave nature, poses a fundamental limitation on any optical system. Diffraction is always present, although its effects may be masked if the system has significant aberrations. When an optical system is essentially free from aberrations, its performance is limited solely by diffraction, and it is referred to as diffraction limited.

In calculating diffraction, we simply need to know the focal length(s) and aperture diameter(s); we do not consider other lens-related factors such as shape or index of refraction.

Since diffraction increases with increasing f-number, and aberrations decrease with increasing f-number, determining optimum system performance often involves finding a point where the combination of these factors has a minimum effect.

ABERRATIONS

To determine the precise performance of a lens system, we can trace the path of light rays through it, using Snell's law at each optical interface to determine the subsequent ray direction. This process, called ray tracing, is usually accomplished on a computer. When this process is completed, it is typically found that not all the rays pass through the points or positions predicted by paraxial theory. These deviations from ideal imaging are called lens aberrations.

The direction of a light ray after refraction at the interface between two homogeneous, isotropic media of differing index of refraction is given by Snell's law:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \qquad (1.20)$$

where θ_1 is the angle of incidence, θ_2 is the angle of refraction, and both angles are measured from the surface normal as shown in figure 1.14.

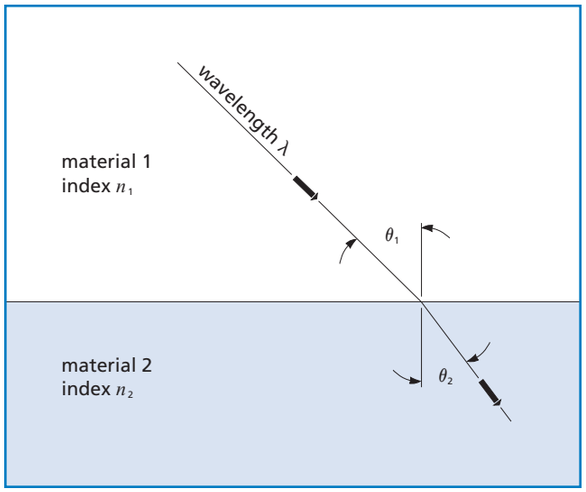

Figure 1.14 **Refraction of light at a dielectric boundary**

APPLICATION NOTE

Technical Assistance

Detailed performance analysis of an optical system is accomplished by using computerized ray-tracing software. CVI Melles Griot applications engineers are able to provide a ray-tracing analysis of simple catalog-component systems. If you need assistance in determining the performance of your optical system, or in selecting optimum components for your particular application, please contact your nearest CVI Melles Griot office.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.12

Even though tools for the precise analysis of an optical system are becoming easier to use and are readily available, it is still quite useful to have a method for quickly estimating lens performance. This not only saves time in the initial stages of system specification, but can also help achieve a better starting point for any further computer optimization.

The first step in developing these rough guidelines is to realize that the sine functions in Snell's law can be expanded in an infinite Taylor series:

$$\sin\theta_1 = \theta_1 - \theta_1^3/3! + \theta_1^5/5! - \theta_1^7/7! + \theta_1^9/9! - \ldots \qquad (1.21)$$

The first approximation we can make is to replace all the sine functions with their arguments (i.e., replace $\sin\theta_1$ with θ_1 itself and so on). This is called first-order or paraxial theory because only the first terms of the sine expansions are used. Design of any optical system generally starts with this approximation using the paraxial formulas.

The assumption that $\sin\theta = \theta$ is reasonably valid for θ close to zero (i.e., high f-number lenses). With more highly curved surfaces (and particularly marginal rays), paraxial theory yields increasingly large deviations from real performance because $\sin\theta \neq \theta$. These deviations are known as aberrations. Because a perfect optical system (one without any aberrations) would form its image at the point and to the size indicated by paraxial theory, aberrations are really a measure of how the image differs from the paraxial prediction.

As already stated, exact ray tracing is the only rigorous way to analyze real lens surfaces. Before the advent of electronic computers, this was excessively tedious and time consuming. Seidel* addressed this issue by developing a method of calculating aberrations resulting from the $\theta_1^3/3!$ term. The resultant third-order lens aberrations are therefore called Seidel aberrations.

To simplify these calculations, Seidel put the aberrations of an optical system into several different classifications. In monochromatic light they are spherical aberration, astigmatism, field curvature, coma, and distortion. In polychromatic light there are also chromatic aberration and lateral color. Seidel developed methods to approximate each of these aberrations without actually tracing large numbers of rays using all the terms in the sine expansions.

In actual practice, aberrations occur in combinations rather than alone. This system of classifying them, which makes analysis much simpler, gives a good description of optical system image quality. In fact, even in the era of powerful ray-tracing software, Seidel's formula for spherical aberration is still widely used.

** Ludwig von Seidel, 1857.*

SPHERICAL ABERRATION

Figure 1.15 illustrates how an aberration-free lens focuses incoming collimated light. All rays pass through the focal point F″. The lower figure shows the situation more typically encountered in single lenses. The farther from the optical axis the ray enters the lens, the nearer to the lens it focuses (crosses the optical axis). The distance along the optical axis between the intercept of the rays that are nearly on the optical axis (paraxial rays) and the rays that go through the edge of the lens (marginal rays) is called longitudinal spherical aberration (LSA). The height at which these rays intercept the paraxial focal plane is called transverse spherical aberration (TSA). These quantities are related by

$$\mathrm{TSA} = \mathrm{LSA} \times \tan(u'') \qquad (1.22)$$

Spherical aberration is dependent on lens shape, orientation, and conjugate ratio, as well as on the index of refraction of the materials present. Parameters for choosing the best lens shape and orientation for a given task are presented later in this chapter. However, the third-

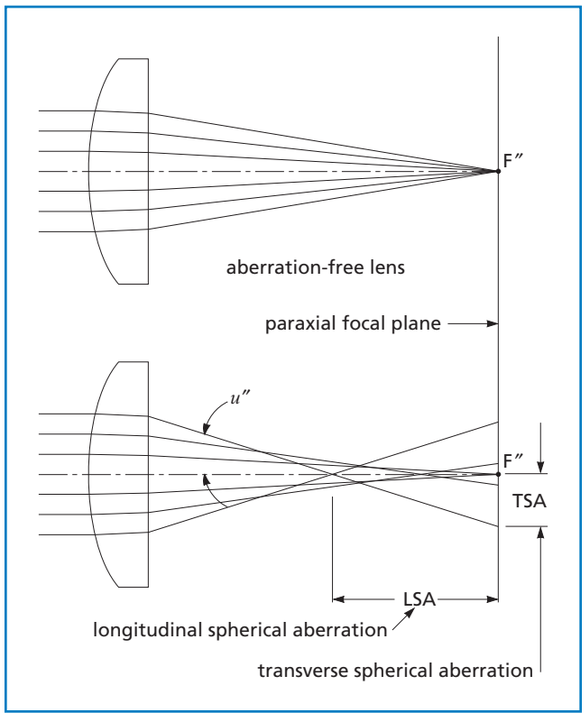

Figure 1.15 **Spherical aberration of a plano-convex lens**

order, monochromatic, spherical aberration of a plano-convex lens used at infinite conjugate ratio can be estimated by

$$\text{spot size due to spherical aberration} = \frac{0.067\,f}{\text{f/\#}^3}. \qquad (1.23)$$

Theoretically, the simplest way to eliminate or reduce spherical aberration is to make the lens surface(s) with a varying radius of curvature (i.e., an aspheric surface) designed to exactly compensate for the fact that $\sin\theta \neq \theta$ at larger angles. In practice, however, most lenses with high surface accuracy are manufactured by grinding and polishing techniques that naturally produce spherical or cylindrical surfaces. The manufacture of aspheric surfaces is more complex, and it is difficult to produce a lens of sufficient surface accuracy to eliminate spherical aberration completely. Fortunately, these aberrations can be virtually eliminated, for a chosen set of conditions, by combining the effects of two or more spherical (or cylindrical) surfaces.

In general, simple positive lenses have undercorrected spherical aberration, and negative lenses usually have overcorrected spherical aberration. By combining a positive lens made from low-index glass with a negative lens made from high-index glass, it is possible to produce a combination in which the spherical aberrations cancel but the focusing powers do not. The simplest examples of this are cemented doublets, such as the LAO series which produce minimal spherical aberration when properly used.

ASTIGMATISM

When an off-axis object is focused by a spherical lens, the natural asymmetry leads to astigmatism. The system appears to have two different focal lengths.

As shown in figure 1.16, the plane containing both optical axis and object point is called the tangential plane. Rays that lie in this plane are called tangential, or meridional, rays. Rays not in this plane are referred to as skew rays. The chief, or principal, ray goes from the object point through the center of the aperture of the lens system. The plane perpendicular to the tangential plane that contains the principal ray is called the sagittal or radial plane.

The figure illustrates that tangential rays from the object come to a focus closer to the lens than do rays in the sagittal plane. When the image is evaluated at the tangential conjugate, we see a line in the sagittal direction. A line in the tangential direction is formed at the sagittal conjugate. Between these conjugates, the image is either an elliptical or a circular blur. Astigmatism is defined as the separation of these conjugates.

The amount of astigmatism in a lens depends on lens shape only when there is an aperture in the system that is not in contact with the lens itself. (In all optical systems there is an aperture or stop, although in many cases it is simply the clear aperture of the lens element itself.) Astigmatism strongly depends on the conjugate ratio.

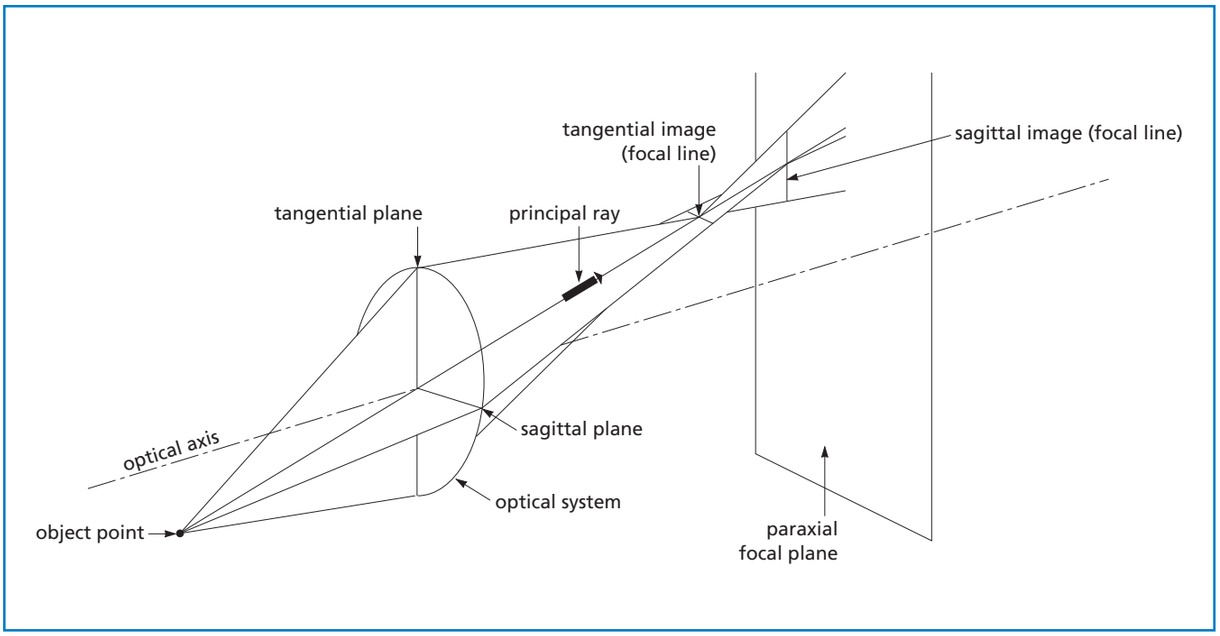

Figure 1.16 **Astigmatism represented by sectional views**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:28 PM Page 1.14

COMA

In spherical lenses, different parts of the lens surface exhibit different degrees of magnification. This gives rise to an aberration known as coma. As shown in figure 1.17, each concentric zone of a lens forms a ring-shaped image called a comatic circle. This causes blurring in the image plane (surface) of off-axis object points. An off-axis object point is not a sharp image point, but it appears as a characteristic comet-like flare. Even if spherical aberration is corrected and the lens brings all rays to a sharp focus on axis, a lens may still exhibit coma off axis. See figure 1.18.

As with spherical aberration, correction can be achieved by using multiple surfaces. Alternatively, a sharper image may be produced by judiciously placing an aperture, or stop, in an optical system to eliminate the more marginal rays.

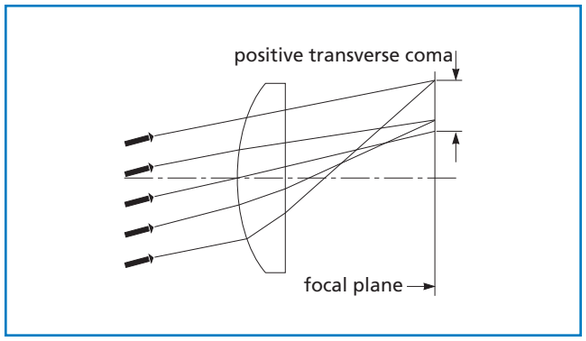

Figure 1.18 **Positive transverse coma**

FIELD CURVATURE

Even in the absence of astigmatism, there is a tendency of optical systems to image better on curved surfaces than on flat planes. This effect is called field curvature (see figure 1.19). In the presence of astigmatism, this problem is compounded because two separate astigmatic focal surfaces correspond to the tangential and sagittal conjugates.

Field curvature varies with the square of field angle or the square of image height. Therefore, by reducing the field angle by one-half, it is possible to reduce the blur from field curvature to a value of 0.25 of its original size.

Positive lens elements usually have inward curving fields, and negative lenses have outward curving fields. Field curvature can thus be corrected to some extent by combining positive and negative lens elements.

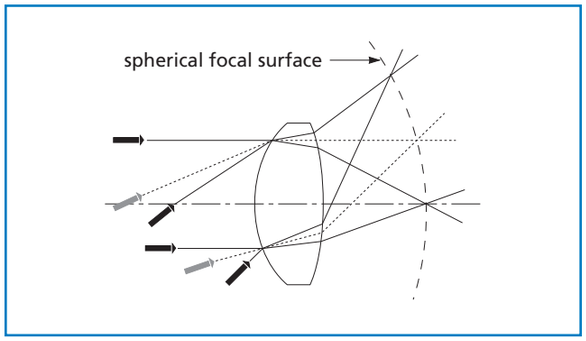

Figure 1.19 **Field curvature**

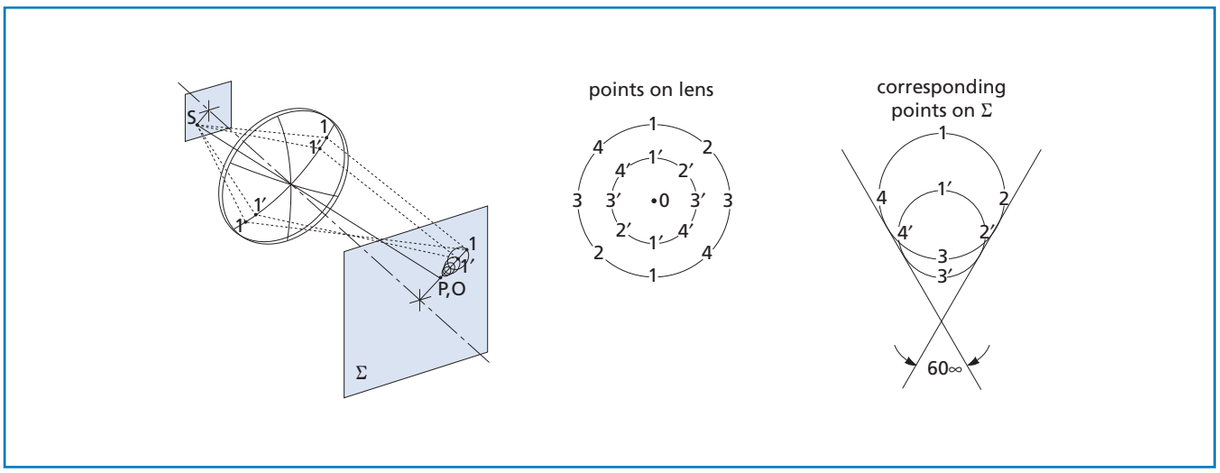

Figure 1.17 **Imaging an off-axis point source by a lens with positive transverse coma**

DISTORTION

The image field not only may have curvature but may also be distorted. The image of an off-axis point may be formed at a location on this surface other than that predicted by the simple paraxial equations. This distortion is different from coma (where rays from an off-axis point fail to meet perfectly in the image plane). Distortion means that even if a perfect off-axis point image is formed, its location on the image plane is not correct. Furthermore, the amount of distortion usually increases with increasing image height. The effect of this can be seen as two different kinds of distortion: pincushion and barrel (see figure 1.20). Distortion does not lower system resolution; it simply means that the image shape does not correspond exactly to the shape of the object. Distortion is a separation of the actual image point from the paraxially predicted location on the image plane and can be expressed either as an absolute value or as a percentage of the paraxial image height.

It should be apparent that a lens or lens system has opposite types of distortion depending on whether it is used forward or backward. This means that if a lens were used to make a photograph, and then used in reverse to project it, there would be no distortion in the final screen image. Also, perfectly symmetrical optical systems at 1:1 magnification have no distortion or coma.

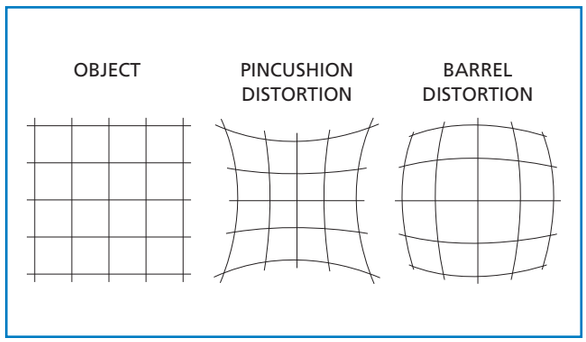

Figure 1.20 **Pincushion and barrel distortion**

CHROMATIC ABERRATION

The aberrations previously described are purely a function of the shape of the lens surfaces, and they can be observed with monochromatic light. Other aberrations, however, arise when these optics are used to transform light containing multiple wavelengths. The index of refraction of a material is a function of wavelength. Known as dispersion, this is discussed in *Material Properties*. From Snell's law (see equation 1.20), it can be seen that light rays of different wavelengths or colors will be refracted at different angles since the index is not a constant. Figure 1.21 shows the result when polychromatic collimated light is incident on a positive lens element. Because the index of refraction is higher for shorter wavelengths, these are focused closer to the lens than the longer wavelengths. Longitudinal chromatic aberration is defined as the axial distance from the nearest to the farthest focal point. As in the case of spherical aberration, positive and negative elements have opposite signs of chromatic aberration. Once again, by combining elements of nearly opposite aberration to form a doublet, chromatic aberration can be partially corrected. It is necessary to use two glasses with different dispersion characteristics, so that the weaker negative element can balance the aberration of the stronger, positive element.

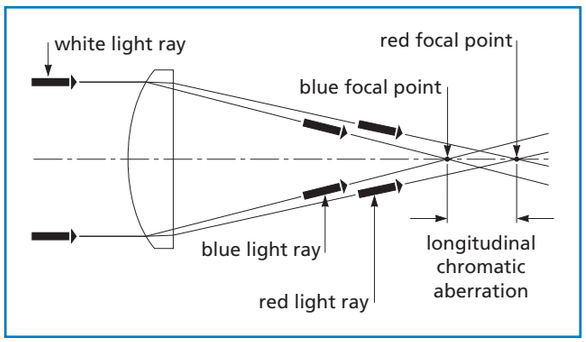

Figure 1.21 **Longitudinal chromatic aberration**

Variations of Aberrations with Aperture, Field Angle, and Image Height

Aberration	Aperture (ϕ)	Field Angle (θ)	Image Height (y)
Lateral Spherical	ϕ^3	—	—
Longitudinal Spherical	ϕ^2	—	—
Coma	ϕ^2	θ	y
Astigmatism	ϕ	θ^2	y^2
Field Curvature	ϕ	θ^2	y^2
Distortion	—	θ^3	y^3
Chromatic	—	—	—

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.16

LATERAL COLOR

Lateral color is the difference in image height between blue and red rays. Figure 1.22 shows the chief ray of an optical system consisting of a simple positive lens and a separate aperture. Because of the change in index with wavelength, blue light is refracted more strongly than red light, which is why rays intercept the image plane at different heights. Stated simply, magnification depends on color. Lateral color is very dependent on system stop location.

For many optical systems, the third-order term is all that may be needed to quantify aberrations. However, in highly corrected systems or in those having large apertures or a large angular field of view, third-order theory is inadequate. In these cases, exact ray tracing is absolutely essential.

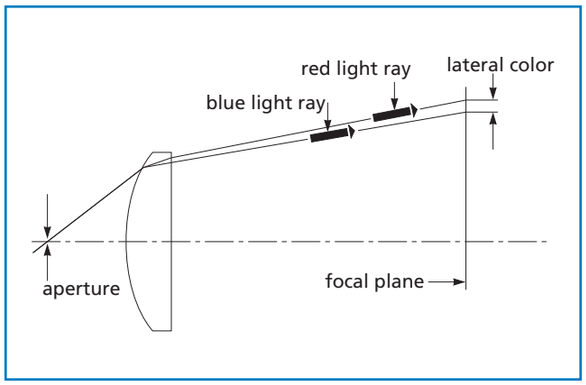

Figure 1.22 **Lateral Color**

APPLICATION NOTE

Achromatic Doublets Are Superior to Simple Lenses

Because achromatic doublets correct for spherical as well as chromatic aberration, they are often superior to simple lenses for focusing collimated light or collimating point sources, even in purely monochromatic light.

Although there is no simple formula that can be used to estimate the spot size of a doublet, the tables in *Spot Size* give sample values that can be used to estimate the performance of catalog achromatic doublets.

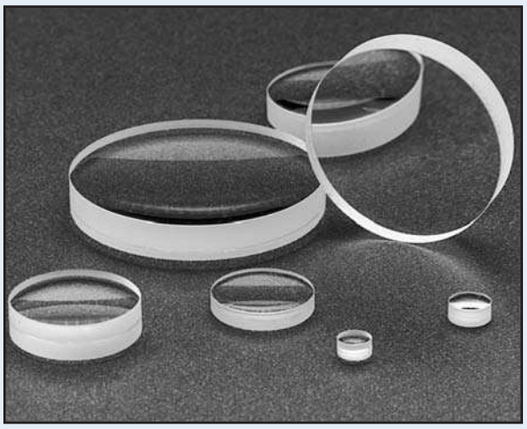

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.17

Lens Shape

Aberrations described in the preceding section are highly dependent on application, lens shape, and material of the lens (or, more exactly, its index of refraction). The singlet shape that minimizes spherical aberration at a given conjugate ratio is called best-form. The criterion for best-form at any conjugate ratio is that the marginal rays are equally refracted at each of the lens/air interfaces. This minimizes the effect of $\sin\theta \neq \theta$. It is also the criterion for minimum surface-reflectance loss. Another benefit is that absolute coma is nearly minimized for best-form shape, at both infinite and unit conjugate ratios.

To further explore the dependence of aberrations on lens shape, it is helpful to make use of the Coddington shape factor, q, defined as

$$q = \frac{(r_2 + r_1)}{(r_2 - r_1)}. \qquad (1.24)$$

Figure 1.23 shows the transverse and longitudinal spherical aberrations of a singlet lens as a function of the shape factor, q. In this particular instance, the lens has a focal length of 100 mm, operates at f/5, has an index of refraction of 1.518722 (BK7 at the mercury green line, 546.1 nm), and is being operated at the infinite conjugate ratio. It is also assumed that the lens itself is the aperture stop. An asymmetric shape that corresponds to a q-value of about 0.7426 for this material and wavelength is the best singlet shape for on-axis imaging. It is important to note that the best-form shape is dependent on refractive index. For example, with a high-index material, such as silicon, the best-form lens for the infinite conjugate ratio is a meniscus shape.

At infinite conjugate with a typical glass singlet, the plano-convex shape ($q = 1$), with convex side toward the infinite conjugate, performs nearly as well as the best-form lens. Because a plano-convex lens costs much less to manufacture than an asymmetric biconvex singlet, these lenses are quite popular. Furthermore, this lens shape exhibits near-minimum total transverse aberration and near-zero coma when used off axis, thus enhancing its utility.

For imaging at unit magnification ($s = s'' = 2f$), a similar analysis would show that a symmetric biconvex lens is the best shape. Not only is spherical aberration minimized, but coma, distortion, and lateral chromatic aberration exactly cancel each other out. These results are true regardless of material index or wavelength, which explains the utility of symmetric convex lenses, as well as symmetrical optical systems in general. However, if a remote stop is present, these aberrations may not cancel each other quite as well.

For wide-field applications, the best-form shape is definitely not the optimum singlet shape, especially at the infinite conjugate ratio, since it yields maximum field curvature. The ideal shape is determined by the situation and may require rigorous ray-tracing analysis. It is possible to achieve much better correction in an optical system by using more than one element. The cases of an infinite conjugate ratio system and a unit conjugate ratio system are discussed in the following section.

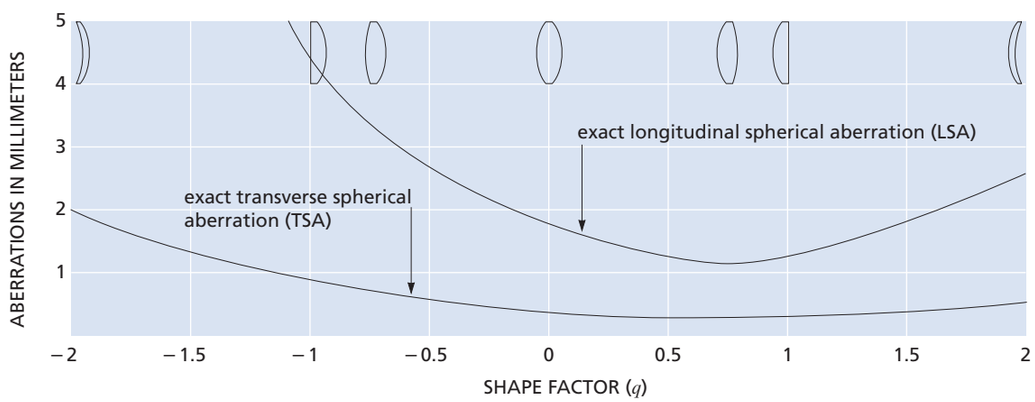

Figure 1.23 **Aberrations of positive singlets at infinite conjugate ratio as a function of shape**

Lens Combinations

INFINITE CONJUGATE RATIO

As shown in the previous discussion, the best-form singlet lens for use at infinite conjugate ratios is generally nearly plano-convex. Figure 1.24 shows a plano-convex lens (LPX-15.0-10.9-C) with incoming collimated light at a wavelength of 546.1 nm. This drawing, including the rays traced through it, is shown to exact scale. The marginal ray (ray f-number 1.5) strikes the paraxial focal plane significantly off the optical axis.

This situation can be improved by using a two-element system. The second part of the figure shows a precision achromat (LAO-21.0-14.0), which consists of a positive low-index (crown glass) element cemented to a negative meniscus high-index (flint glass) element. This is drawn to the same scale as the plano-convex lens. No spherical aberration can be discerned in the lens. Of course, not all of the rays pass exactly through the paraxial focal point; however, in this case, the departure is measured in micrometers, rather than in millimeters, as in the case of the plano-convex lens. Additionally, chromatic aberration (not shown) is much better corrected in the doublet. Even though these lenses are known as achromatic doublets, it is important to remember that even with monochromatic light the doublet's performance is superior.

Figure 1.24 also shows the f-number at which singlet performance becomes unacceptable. The ray with f-number 7.5 practically intercepts the paraxial focal point, and the f/3.8 ray is fairly close. This useful drawing, which can be scaled to fit a plano-convex lens of any focal length, can be used to estimate the magnitude of its spherical aberration, although lens thickness affects results slightly.

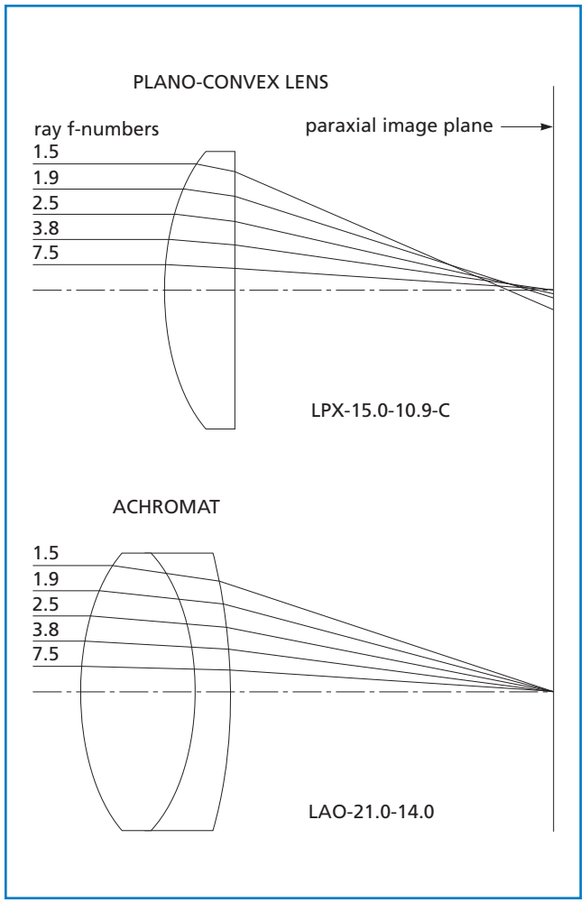

Figure 1.24 **Single-element plano-convex lens compared with a two-element achromat**

UNIT CONJUGATE RATIO

Figure 1.25 shows three possible systems for use at the unit conjugate ratio. All are shown to the same scale and using the same ray f-numbers with a light wavelength of 546.1 nm. The first system is a symmetric biconvex lens (LDX-21.0-19.2-C), the best-form singlet in this application. Clearly, significant spherical aberration is present in this lens at f/2.7. Not until f/13.3 does the ray closely approach the paraxial focus.

A dramatic improvement in performance is gained by using two identical plano-convex lenses with convex surfaces facing and nearly in contact. Those shown in figure 1.25 are both LPX-20.0-20.7-C. The combination of these two lenses yields almost exactly the same focal length as the biconvex lens. To understand why this configuration improves performance so dramatically, consider that if the biconvex lens were split down the middle, we would have two identical plano-convex lenses, each working at an infinite conjugate ratio, but with the convex surface toward the focus. This orientation is opposite to that shown to be optimum for this shape lens. On the other hand, if these lenses are reversed, we have the system just described but with a better correction of the spherical aberration.

Previous examples indicate that an achromat is superior in performance to a singlet when used at the infinite conjugate ratio and at low f-numbers. Since the unit conjugate case can be thought of as two lenses, each working at the infinite conjugate ratio, the next step is to replace the plano-convex singlets with achromats, yielding a four-element system. The third part of figure 1.25 shows a system composed of two LAO-40.0-18.0 lenses. Once again, spherical aberration is not evident, even in the f/2.7 ray.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.19

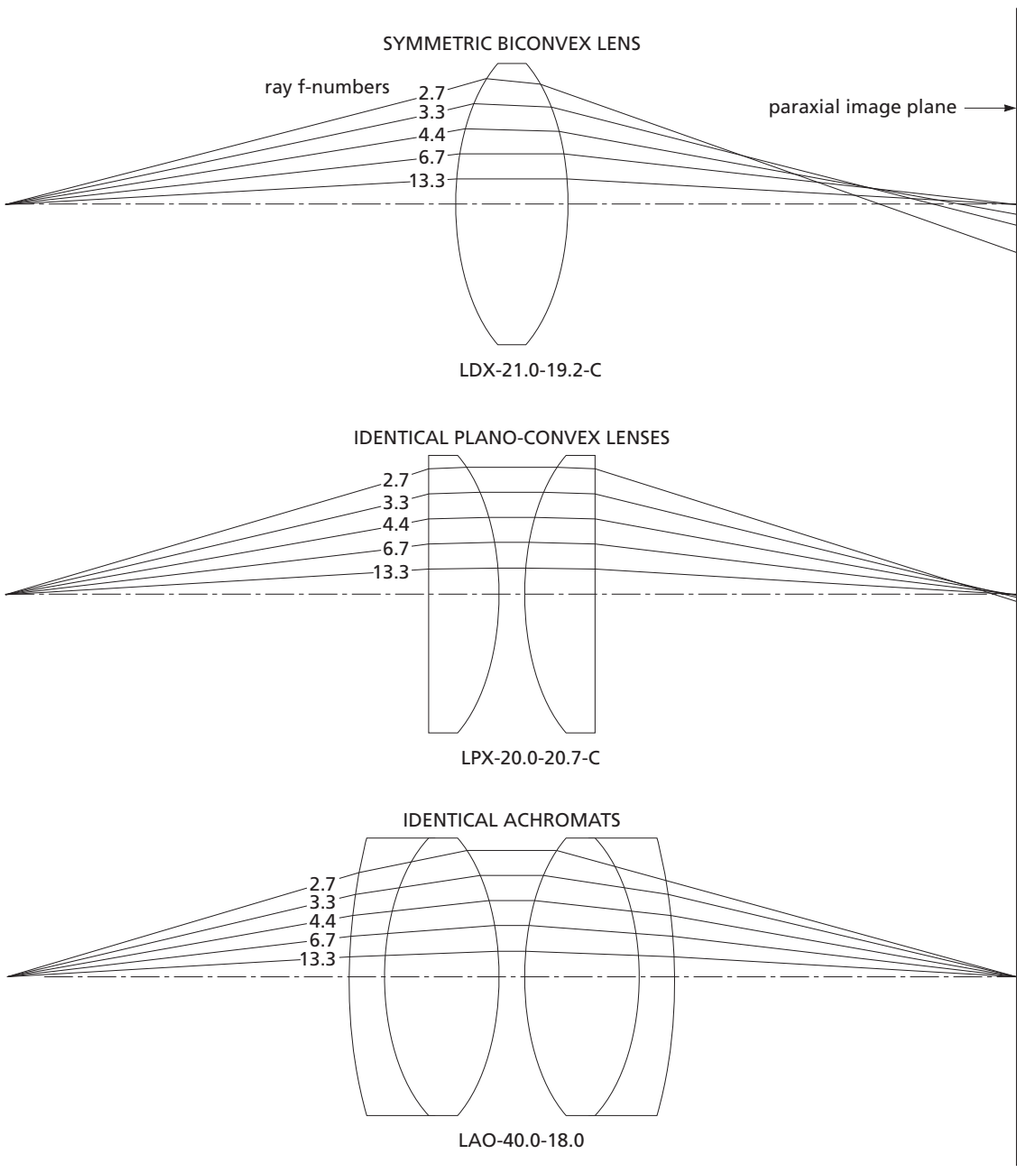

Figure 1.25 **Three possible systems for use at the unit conjugate ratio**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.20

Diffraction Effects

In all light beams, some energy is spread outside the region predicted by geometric propagation. This effect, known as diffraction, is a fundamental and inescapable physical phenomenon. Diffraction can be understood by considering the wave nature of light. Huygens' principle (figure 1.26) states that each point on a propagating wavefront is an emitter of secondary wavelets. The propagating wave is then the envelope of these expanding wavelets. Interference between the secondary wavelets gives rise to a fringe pattern that rapidly decreases in intensity with increasing angle from the initial direction of propagation. Huygens' principle nicely describes diffraction, but rigorous explanation demands a detailed study of wave theory.

Diffraction effects are traditionally classified into either Fresnel or Fraunhofer types. Fresnel diffraction is primarily concerned with what happens to light in the immediate neighborhood of a diffracting object or aperture. It is thus only of concern when the illumination source is close to this aperture or object, known as the near field. Consequently, Fresnel diffraction is rarely important in most classical optical setups, but it becomes very important in such applications as digital optics, fiber optics, and near-field microscopy.

Fraunhofer diffraction, however, is often important even in simple optical systems. This is the light-spreading effect of an aperture when the aperture (or object) is illuminated with an infinite source (plane-wave illumination) and the light is sensed at an infinite distance (far-field) from this aperture.

From these overly simple definitions, one might assume that Fraunhofer diffraction is important only in optical systems with infinite conjugate, whereas Fresnel diffraction equations should be considered at finite conjugate ratios. Not so. A lens or lens system of finite positive focal length with plane-wave input maps the far-field diffraction pattern of its aperture onto the focal plane; therefore, it is Fraunhofer diffraction that determines the limiting performance of optical systems. More generally, at any conjugate ratio, far-field angles are transformed into spatial displacements in the image plane.

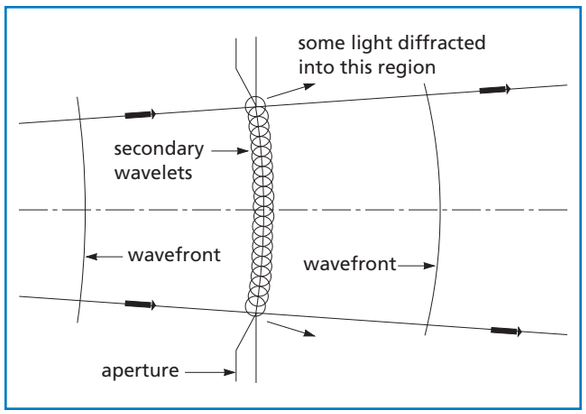

Figure 1.26 **Huygens' principle**

CIRCULAR APERTURE

Fraunhofer diffraction at a circular aperture dictates the fundamental limits of performance for circular lenses. It is important to remember that the spot size, caused by diffraction, of a circular lens is

$$d = 2.44\lambda(\text{f/\#}) \qquad (1.25)$$

where d is the diameter of the focused spot produced from plane-wave illumination and λ is the wavelength of light being focused. Notice that it is the f-number of the lens, not its absolute diameter, that determines this limiting spot size.

The diffraction pattern resulting from a uniformly illuminated circular aperture actually consists of a central bright region, known as the Airy disc (see figure 1.27), which is surrounded by a number of much fainter rings. Each ring is separated by a circle of zero intensity. The irradiance distribution in this pattern can be described by

$$I_x = I_0 \left[\frac{2J_1(x)}{x}\right]^2 \qquad (1.26)$$

where

I_0 = peak irradiance in the image

$J_1(x)$ = Bessel function of the first kind of order unity

$$= x\sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^{2n-2}}{(n-1)!\,n!\,2^{2n-1}}$$

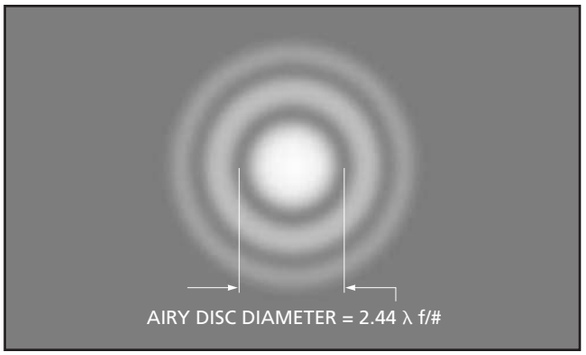

Figure 1.27 **Center of a typical diffraction pattern for a circular aperture**

and

$$x = \frac{\pi D}{\lambda}\sin\theta$$

where

λ = wavelength

D = aperture diameter

θ = angular radius from the pattern maximum.

This useful formula shows the far-field irradiance distribution from a uniformly illuminated circular aperture of diameter D.

SLIT APERTURE

A slit aperture, which is mathematically simpler, is useful in relation to cylindrical optical elements. The irradiance distribution in the diffraction pattern of a uniformly illuminated slit aperture is described by

$$I_x = I_0\left[\frac{\sin x}{x}\right]^2 \qquad (1.27)$$

where I_0 = peak irradiance in image

$$x = \frac{\pi w \sin\theta}{\lambda}$$

where λ = wavelength

w = slit width

θ = angular deviation from pattern maximum.

APPLICATION NOTE

Rayleigh Criterion

In imaging applications, spatial resolution is ultimately limited by diffraction. Calculating the maximum possible spatial resolution of an optical system requires an arbitrary definition of what is meant by resolving two features. In the Rayleigh criterion, it is assumed that two separate point sources can be resolved when the center of the Airy disc from one overlaps the first dark ring in the diffraction pattern of the second. In this case, the smallest resolvable distance, d, is

$$d = \frac{0.61\lambda}{NA} = 1.22\lambda(f/\#)$$

ENERGY DISTRIBUTION TABLE

The accompanying table shows the major features of pure (unaberrated) Fraunhofer diffraction patterns of circular and slit apertures. The table shows the position, relative intensity, and percentage of total pattern energy corresponding to each ring or band. It is especially convenient to characterize positions in either pattern with the same variable x. This variable is related to field angle in the circular aperture case by

$$\sin\theta = \frac{\lambda x}{\pi D} \qquad (1.28)$$

where D is the aperture diameter. For a slit aperture, this relationship is given by

$$\sin\theta = \frac{\lambda x}{\pi w} \qquad (1.29)$$

where w is the slit width, π has its usual meaning, and D, w, and λ are all in the same units (preferably millimeters). Linear instead of angular field positions are simply found from $r = s''\tan\theta$ where s'' is the secondary conjugate distance. This last result is often seen in a different form, namely the diffraction-limited spot-size equation, which, for a circular lens is

$$d = 2.44\,\lambda(f/\#) \qquad \text{(see eq. 1.25)}$$

This value represents the smallest spot size that can be achieved by an optical system with a circular aperture of a given f-number, and it is the diameter of the first dark ring, where the intensity has dropped to zero.

The graph in figure 1.28 shows the form of both circular and slit aperture diffraction patterns when plotted on the same normalized scale. Aperture diameter is equal to slit width so that patterns between x-values and angular deviations in the far-field are the same.

GAUSSIAN BEAMS

Apodization, or nonuniformity of aperture irradiance, alters diffraction patterns. If pupil irradiance is nonuniform, the formulas and results given previously do not apply. This is important to remember because most laser-based optical systems do not have uniform pupil irradiance. The output beam of a laser operating in the TEM_{00} mode has a smooth Gaussian irradiance profile. Formulas used to determine the focused spot size from such a beam are discussed in *Gaussian Beam Optics*. Furthermore, when dealing with Gaussian beams, the location of the focused spot also departs from that predicted by the paraxial equations given in this chapter. This is also detailed in *Gaussian Beam Optics*.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.22

Energy Distribution in the Diffraction Pattern of a Circular or Slit Aperture

	Circular Aperture			Slit Aperture		
Ring or Band	Position (x)	Relative Intensity (I_x/I_0)	Energy in Ring (%)	Position (x)	Relative Intensity (I_x/I_0)	Energy in Band (%)
Central Maximum	0.0	1.0	83.8	0.0	1.0	90.3
First Dark	1.22π	0.0		1.00π	0.0	
First Bright	1.64π	0.0175	7.2	1.43π	0.0472	4.7
Second Dark	2.23π	0.0		2.00π	0.0	
Second Bright	2.68π	0.0042	2.8	2.46π	0.0165	1.7
Third Dark	3.24π	0.0		3.00π	0.0	
Third Bright	3.70π	0.0016	1.5	3.47π	0.0083	0.8
Fourth Dark	4.24π	0.0		4.00π	0.0	
Fourth Bright	4.71π	0.0008	1.0	4.48π	0.0050	0.5
Fifth Dark	5.24π	0.0		5.00π	0.0	

Note: Position variable (x) is defined in the text.

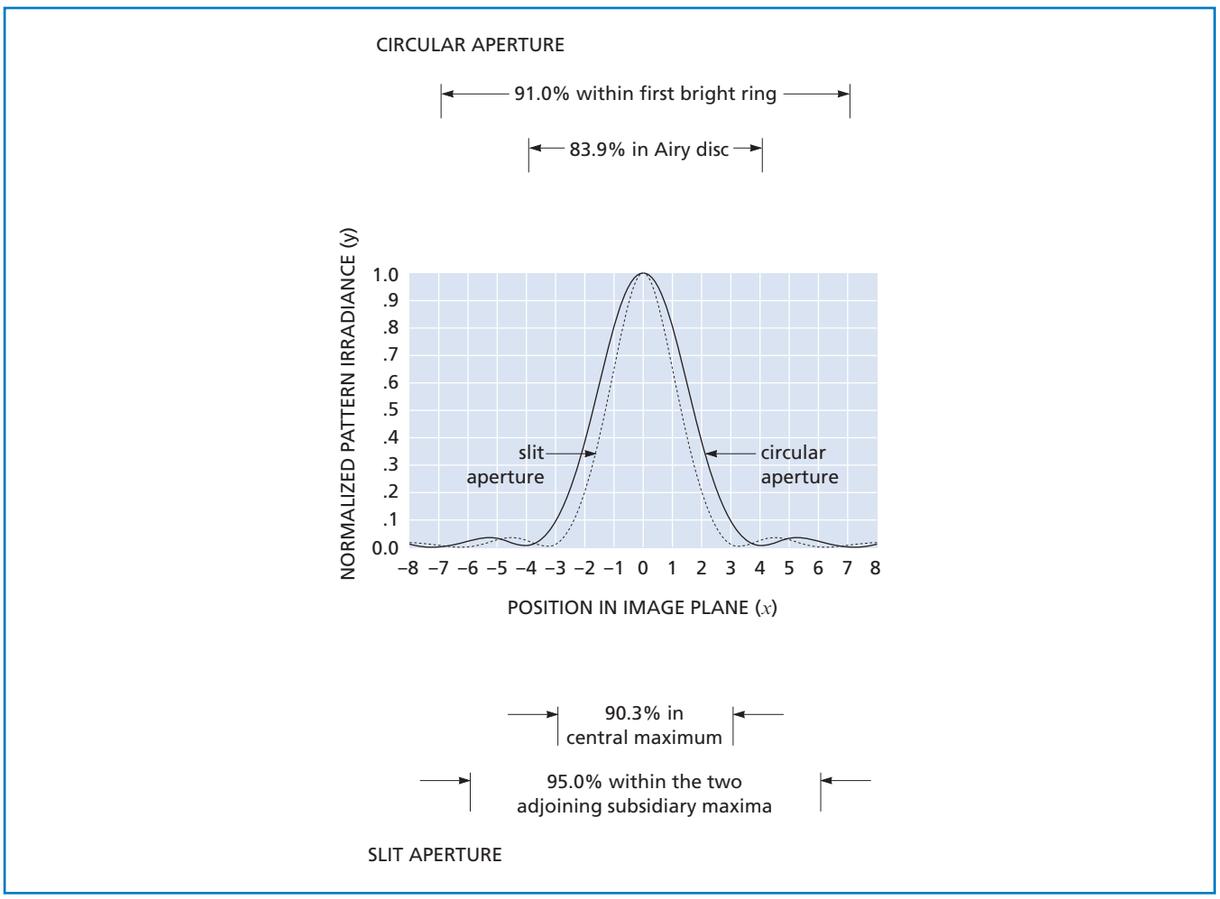

Figure 1.28 **Fraunhofer diffraction pattern of a singlet slit superimposed on the Fraunhofer diffraction pattern of a circular aperture**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.23

Lens Selection

Having discussed the most important factors that affect the performance of a lens or a lens system, we will now address the practical matter of selecting the optimum catalog components for a particular task. The following useful relationships are important to keep in mind throughout the selection process:

- Diffraction-limited spot size = 2.44 λ f/#
- Approximate on-axis spot size of a plano-convex lens at the infinite conjugate resulting from spherical aberration = $\frac{0.067\,f}{\mathrm{f}/\#^3}$
- Optical invariant = $m = \frac{\mathrm{NA}}{\mathrm{NA}''}$.

Example 1: Collimating an Incandescent Source

Produce a collimated beam from a quartz halogen bulb having a 1-mm-square filament. Collect the maximum amount of light possible and produce a beam with the lowest possible divergence angle.

This problem, illustrated in figure 1.29, involves the typical tradeoff between light-collection efficiency and resolution (where a beam is being collimated rather than focused, resolution is defined by beam divergence). To collect more light, it is necessary to work at a low f-number, but because of aberrations, higher resolution (lower divergence angle) will be achieved by working at a higher f-number.

In terms of resolution, the first thing to realize is that the minimum divergence angle (in radians) that can be achieved using any lens system is the source size divided by system focal length. An off-axis ray (from the edge of the source) entering the first principal point of the system exits the second principal point at the same angle. Therefore, increasing the system focal length improves this limiting divergence because the source appears smaller.

An optic that can produce a spot size of 1 mm when focusing a perfectly collimated beam is therefore required. Since source size is inherently limited, it is pointless to strive for better resolution. This level of resolution can be achieved easily with a plano-convex lens.

While angular divergence decreases with increasing focal length, spherical aberration of a plano-convex lens increases with increasing focal length. To determine the appropriate focal length, set the spherical aberration formula for a plano-convex lens equal to the source (spot) size:

$$\frac{0.067\,f}{\mathrm{f}/\#^3} = 1\ \mathrm{mm}.$$

This ensures a lens that meets the minimum performance needed. To select a focal length, make an arbitrary f-number choice. As can be seen from the relationship, as we lower the f-number (increase collection efficiency), we decrease the focal length, which will worsen the resultant divergence angle (minimum divergence = 1 mm/*f*).

In this example, we will accept f/2 collection efficiency, which gives us a focal length of about 120 mm. For f/2 operation we would need a minimum diameter of 60 mm. The LPX-60.0-62.2-C fits this specification exactly. Beam divergence would be about 8 mrad.

Finally, we need to verify that we are not operating below the theoretical diffraction limit. In this example, the numbers (1-mm spot size) indicate that we are not, since

$$\text{diffraction-limited spot size} = 2.44 \times 0.5\ \mu\mathrm{m} \times 2 = 2.44\ \mu\mathrm{m}.$$

Example 2: Coupling an Incandescent Source into a Fiber

In *Imaging Properties of Lens Systems* we considered a system in which the output of an incandescent bulb with a filament of 1 mm in diameter was to be coupled into an optical fiber with a core diameter of 100 µm and a numerical aperture of 0.25. From the optical invariant and other constraints given in the problem, we determined that f = 9.1 mm, CA = 5 mm, s = 100 mm, s'' = 10 mm, NA″ = 0.25, and NA = 0.025 (or f/2 and f/20). The singlet lenses that match these specifications are the plano-convex LPX-5.0-5.2-C or biconvex lenses LDX-6.0-7.7-C and LDX-5.0-9.9-C. The closest achromat would be the LAO-10.0-6.0.

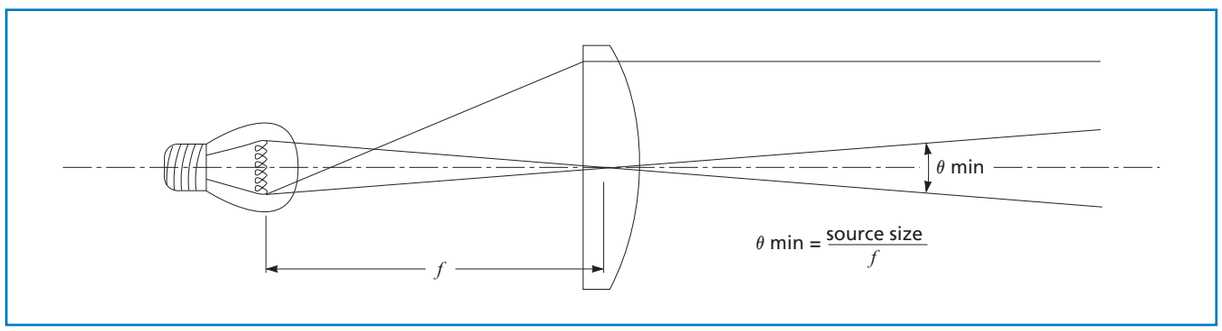

Figure 1.29 **Collimating an incandescent source**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.24

We can immediately reject the biconvex lenses because of spherical aberration. We can estimate the performance of the LPX-5.0-5.2-C on the focusing side by using our spherical aberration formula:

$$\text{spot size} = \frac{0.067\ (10)}{2^3} = 84\ \mu\text{m}.$$

We will ignore, for the moment, that we are not working at the infinite conjugate.

This is slightly smaller than the 100-µm spot size we are trying to achieve. However, since we are not working at infinite conjugate, the spot size will be larger than that given by our simple calculation. This lens is therefore likely to be marginal in this situation, especially if we consider chromatic aberration. A better choice is the achromat. Although a computer ray trace would be required to determine its exact performance, it is virtually certain to provide adequate performance.

Example 3: Symmetric Fiber-to-Fiber Coupling

Couple an optical fiber with an 8-µm core and a 0.15 numerical aperture into another fiber with the same characteristics. Assume a wavelength of 0.5 µm.

This problem, illustrated in figure 1.30, is essentially a 1:1 imaging situation. We want to collect and focus at a numerical aperture of 0.15 or f/3.3, and we need a lens with an 8-µm spot size at this f-number. Based on the lens combination discussion in *Lens Combination Formulas*, our most likely setup is either a pair of identical plano-convex lenses or achromats, faced front to front. To determine the necessary focal length for a plano-convex lens, we again use the spherical aberration estimate formula:

$$\frac{0.067\ f}{3.3^3} = 0.008\ \text{mm}.$$

This formula yields a focal length of 4.3 mm and a minimum diameter of 1.3 mm. The LPX-4.2-2.3-BAK1 meets these criteria. The biggest problem with utilizing these tiny, short focal length lenses is the practical considerations of handling, mounting, and positioning them. Because using a pair of longer focal length singlets would result in unacceptable performance, the next step might be to use a pair of the slightly longer focal length, larger achromats, such as the LAO-10.0-6.0. The performance data, given in *Spot Size*, show that this combination does provide the required 8-μm spot diameter.

Because fairly small spot sizes are being considered here, it is important to make sure that the system is not being asked to work below the diffraction limit:

$$2.44 \times 0.5\ \mu\text{m} \times 3.3 = 4\ \mu\text{m}.$$

Since this is half the spot size caused by aberrations, it can be safely assumed that diffraction will not play a significant role here.

An entirely different approach to a fiber-coupling task such as this would be to use a pair of spherical ball lenses (LMS-LSFN series) or one of the gradient-index lenses (LGT series).

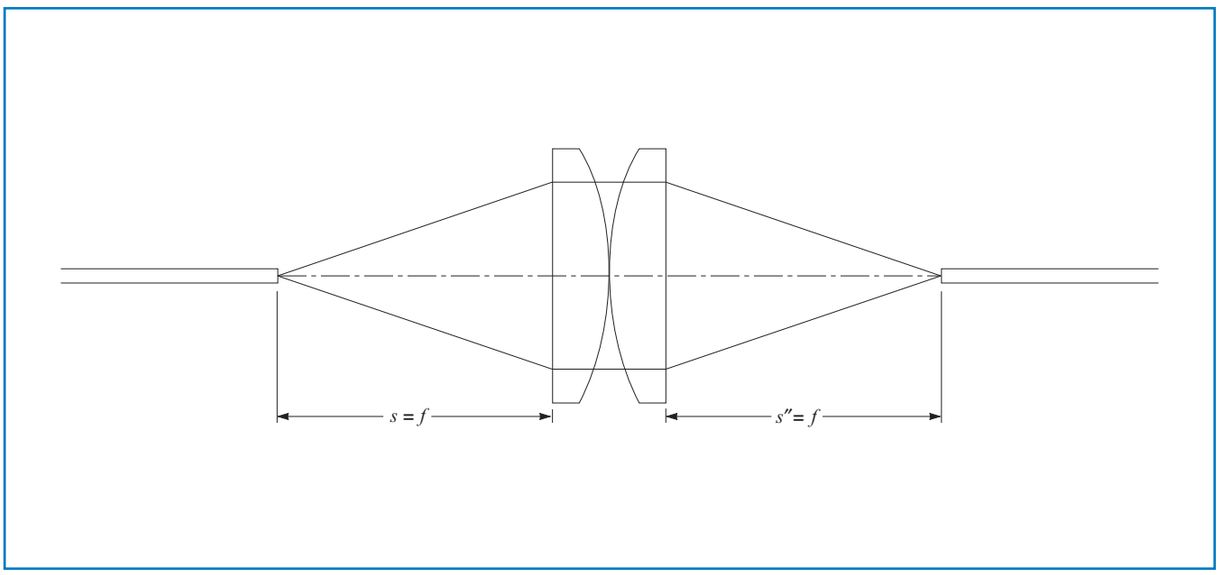

Figure 1.30 **Symmetric fiber-to-fiber coupling**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.25

Example 4: Diffraction-Limited Performance

Determine at what f-number a plano-convex lens being used at an infinite conjugate ratio with 0.5-μm wavelength light becomes diffraction limited (i.e., the effects of diffraction exceed those caused by aberration).

To solve this problem, set the equations for diffraction-limited spot size and third-order spherical aberration equal to each other. The result depends upon focal length, since aberrations scale with focal length, while diffraction is solely dependent upon f-number. By substituting some common focal lengths into this formula, we get f/8.6 at $f = 100$ mm, f/7.2 at $f = 50$ mm, and f/4.8 at $f = 10$ mm.

$$2.44 \times 0.5\ \mu\text{m} \times \text{f/\#} = \frac{0.067 \times f}{\text{f/\#}^3}$$

or

$$\text{f/\#} = (54.9 \times f)^{1/4}.$$

When working with these focal lengths (and under the conditions previously stated), we can assume essentially diffraction-limited performance above these f-numbers. Keep in mind, however, that this treatment does not take into account manufacturing tolerances or chromatic aberration, which will be present in polychromatic applications.

APPLICATION NOTE

Spherical Ball Lenses for Fiber Coupling

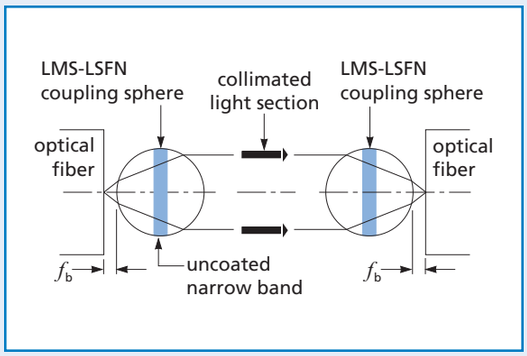

Spheres are arranged so that the fiber end is located at the focal point. The output from the first sphere is then collimated. If two spheres are aligned axially to each other, the beam will be transferred from one focal point to the other. Translational alignment sensitivity can be reduced by enlarging the beam. Slight negative defocusing of the ball can reduce the spherical aberration third-order contribution common to all coupling systems. Additional information can be found in "Lens Coupling in Fiber Optic Devices: Efficiency Limits," by A. Nicia, *Applied Optics*, vol. 20, no. 18, pp 3136–45, 1981. Off-axis aberrations are absent since the fiber diameters are so much smaller than the coupler focal length.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.26

Spot Size

In general, the performance of a lens or lens system in a specific circumstance should be determined by an exact trigonometric ray trace. CVI Melles Griot applications engineers can supply ray-tracing data for particular lenses and systems of catalog components on request. In certain situations, however, some simple guidelines can be used for lens selection. The optimum working conditions for some of the lenses in this catalog have already been presented. The following tables give some quantitative results for a variety of simple and compound lens systems, which can be constructed from standard catalog optics.

In interpreting these tables, remember that these theoretical values obtained from computer ray tracing consider only the effects of ideal geometric optics. Effects of manufacturing tolerances have not been considered. Furthermore, remember that using more than one element provides a higher degree of correction but makes alignment more difficult. When actually choosing a lens or a lens system, it is important to note the tolerances and specifications clearly described for each CVI Melles Griot lens in the product listings.

The tables give the diameter of the spot for a variety of lenses used at several different f-numbers. All the tables are for on-axis, uniformly illuminated, collimated input light at 546.1 nm. They assume that the lens is facing in the direction that produces a minimum spot size. When the spot size caused by aberrations is smaller or equal to the diffraction-limited spot size, the notation "DL" appears next to the entry. The shorter focal length lenses produce smaller spot sizes because aberrations increase linearly as a lens is scaled up.

The effect on spot size caused by spherical aberration is strongly dependent on f-number. For a plano-convex singlet, spherical aberration is inversely dependent on the cube of the f-number. For doublets, this relationship can be even higher. On the other hand, the spot size caused by diffraction increases linearly with f-number. Thus, for some lens types, spot size at first decreases and then increases with f-number, meaning that there is some optimum performance point at which both aberrations and diffraction combine to form a minimum.

Unfortunately, these results cannot be generalized to situations in which the lenses are used off axis. This is particularly true of the achromat/aplanatic meniscus lens combinations because their performance degrades rapidly off axis.

Focal Length = 10 mm

	Spot Size (μm)*		
f/#	LDX-5.0-9.9-C	LPX-8.0-5.2-C	LAO-10.0-6.0
f/2	—	94	11
f/3	36	25	7
f/5	8	6.7 (DL)	6.7 (DL)
f/10	13.3 (DL)	13.3 (DL)	13.3 (DL)

**Diffraction-limited performance is indicated by DL.*

Focal Length = 30 mm

	Spot Size (μm)*		
f/#	LPX-18.5-15.6-C	LAO-30.0-12.5	LAO-50.0-18.0 & MENP-18.0-4.0-73.5-NSF8
f/2	295	—	3
f/3	79	7	4 (DL)
f/5	17	6.7 (DL)	6.9 (DL)
f/10	13.3 (DL)	13.3 (DL)	13.8 (DL)

**Diffraction-limited performance is indicated by DL.*

Focal Length = 60 mm

	Spot Size (μm)*			
f/#	LDX-50.0-60.0-C	LPX-30.0-31.1-C	LAO-60.0-30.0	LAO-100.0-31.5 & MENP-31.5-6.0-146.4-NSF8
f/2	816	600	—	—
f/3	217	160	34	4 (DL)
f/5	45	33	10	6.7 (DL)
f/10	13.3 (DL)	13.3 (DL)	13.3 (DL)	13.3 (DL)

**Diffraction-limited performance is indicated by DL.*

Aberration Balancing

To improve system performance, optical designers make sure that the total aberration contribution from all surfaces taken together sums to nearly zero. Normally, such a process requires computerized analysis and optimization. However, there are some simple guidelines that can be used to achieve this with lenses available in this catalog. This approach can yield systems that operate at a much lower f-number than can usually be achieved with simple lenses.

Specifically, we will examine how to null the spherical aberration from two or more lenses in collimated, monochromatic light. This technique will thus be most useful for laser beam focusing and expanding.

Figure 1.31 shows the third-order longitudinal spherical aberration coefficients for six of the most common positive and negative lens shapes when used with parallel, monochromatic incident light. The plano-convex and plano-concave lenses both show minimum spherical aberration when oriented with their curved surface facing the incident parallel beam. All other configurations exhibit larger amounts of spherical aberration. With these lens types, it is now possible to show how various systems can be corrected for spherical aberration.

A two-element laser beam expander is a good starting example. In this case, two lenses are separated by a distance that is the sum of their focal lengths, so that the overall system focal length is infinite. This system will not focus incoming collimated light, but it will change the beam diameter. By definition, each of the lenses is operating at the same f-number.

The equation for longitudinal spherical aberration shows that, for two lenses with the same f-number, aberration varies directly with the focal lengths of the lenses. The sign of the aberration is the same as focal length. Thus, it should be possible to correct the spherical aberration of this Galilean-type beam expander, which consists of a positive focal length objective and a negative diverging lens.

If a plano-convex lens of focal length f_1 oriented in the normal direction is combined with a plano-concave lens of focal length f_2 oriented in its reverse direction, the total spherical aberration of the system is

$$\mathrm{LSA} = \frac{0.272\, f_1}{\mathrm{f/\#}^2} + \frac{1.069\, f_2}{\mathrm{f/\#}^2}.$$

After setting this equation to zero, we obtain

$$\frac{f_1}{f_2} = -\frac{1.069}{0.272} = -3.93.$$

To make the magnitude of aberration contributions of the two elements equal so they will cancel out, and thus correct the system, select the focal length of the positive element to be 3.93 times that of the negative element.

Figure 1.32(a) shows a beam-expander system made up of catalog elements, in which the focal length ratio is 4:1. This simple system is corrected to about 1/6 wavelength at 632.8 nm, even though the objective is operating at f/4 with a 20-mm aperture diameter. This is remarkably good wavefront correction for such a simple system; one would normally assume that a doublet objective would be needed and a complex diverging lens as well. This analysis does not take into account manufacturing tolerances.

A beam expander of lower magnification can also be derived from this information. If a symmetric-convex objective is used together] with a reversed plano-concave diverging lens, the aberration coefficients are in the ratio of 1.069/0.403 = 2.65. Figure 1.32(b) shows a system of catalog lenses that provides a magnification of 2.7 (the closest possible given the available focal lengths). The maximum wavefront error in this case is only a quarter-wave, even though the objective is working at f/3.3.

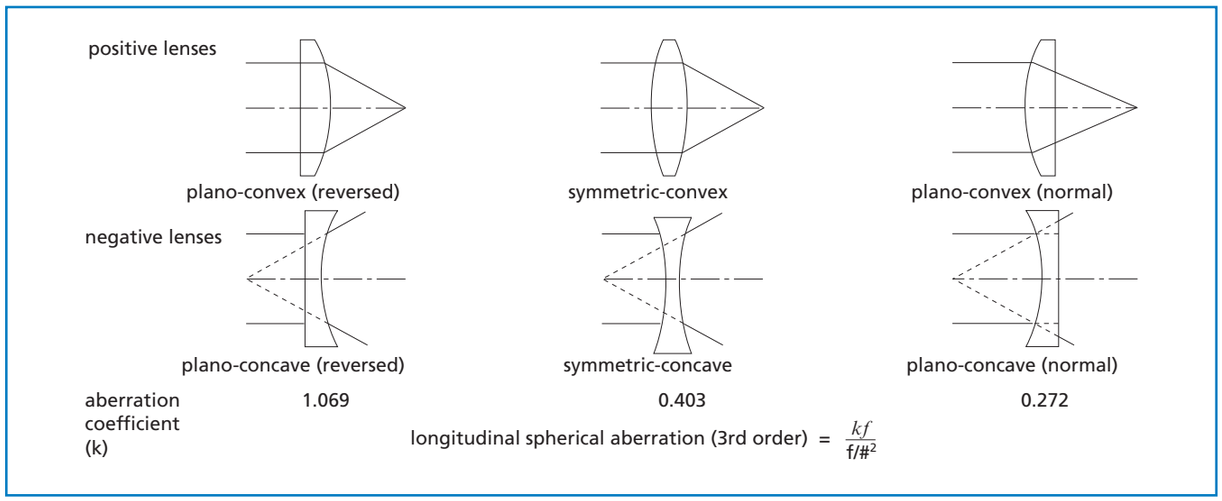

Figure 1.31 **Third-order longitudinal spherical aberration of typical lens shapes**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.28

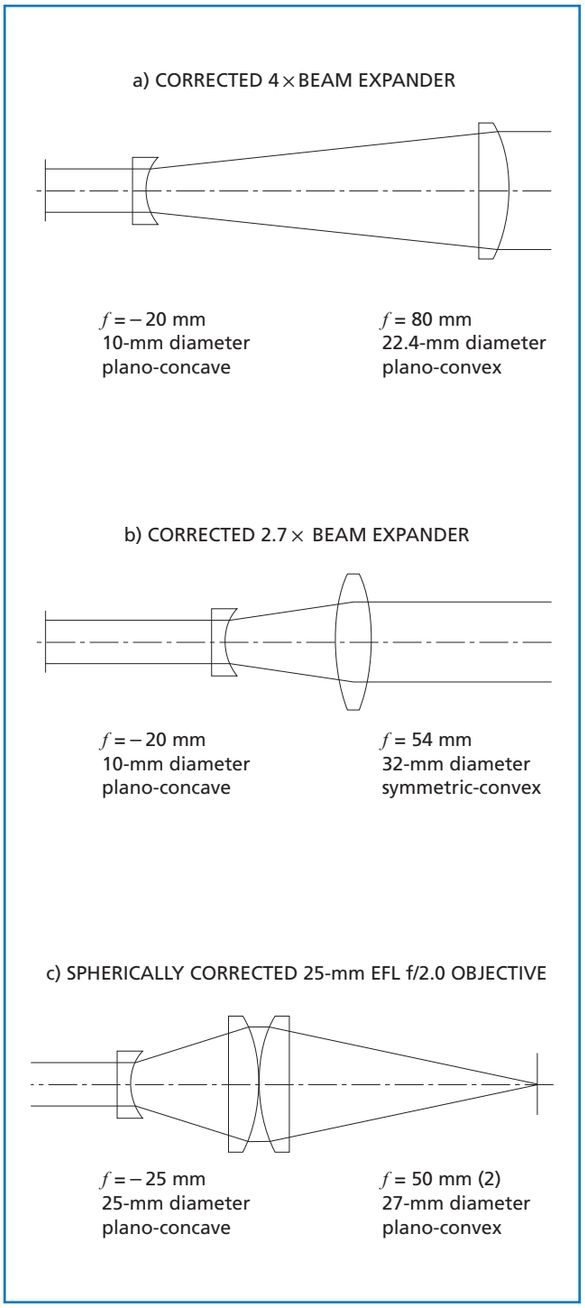

Figure 1.32 **Combining catalog lenses for aberration balancing**

The relatively low *f* numbers of these objectives is a great advantage in minimizing the length of these beam expanders. They would be particularly useful with Nd:YAG and argon-ion lasers, which tend to have large output beam diameters.

These same principles can be utilized to create high numerical aperture objectives that might be used as laser focusing lenses. Figure 1.32(c) shows an objective consisting of an initial negative element, followed by two identical plano-convex positive elements. Again, all of the elements operate at the same f-number, so that their aberration contributions are proportional to their focal lengths. To obtain zero total spherical aberration from this configuration, we must satisfy

$$1.069\, f_1 + 0.272\, f_2 + 0.272\, f_2 = 0$$

or

$$\frac{f_1}{f_2} = -0.51.$$

Therefore, a corrected system should result if the focal length of the negative element is just about half that of each of the positive lenses. In this case, $f_1 = 425$ mm and $f_2 = 50$ mm yield a total system focal length of about 25 mm and an f-number of approximately f/2. This objective, corrected to 1/6 wave, has the additional advantage of a very long working distance.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.29

Definition of Terms

FOCAL LENGTH (f)

Two distinct terms describe the focal lengths associated with every lens or lens system. The effective focal length (EFL) or equivalent focal length (denoted f in figure 1.33) determines magnification and hence the image size. The term f appears frequently in lens formulas and in the tables of standard lenses. Unfortunately, because f is measured with reference to principal points which are usually inside the lens, the meaning of f is not immediately apparent when a lens is visually inspected.

The second type of focal length relates the focal plane positions directly to landmarks on the lens surfaces (namely the vertices) which are immediately recognizable. It is not simply related to image size but is especially convenient for use when one is concerned about correct lens positioning or mechanical clearances. Examples of this second type of focal length are the front focal length (FFL, denoted f_f in figure 1.33) and the back focal length (BFL, denoted f_b).

The convention in all of the figures (with the exception of a single deliberately reversed ray) is that light travels from left to right.

FOCAL POINT (F OR F″)

Rays that pass through or originate at either focal point must be, on the opposite side of the lens, parallel to the optical axis. This fact is the basis for locating both focal points.

PRIMARY PRINCIPAL SURFACE

Let us imagine that rays originating at the front focal point F (and therefore parallel to the optical axis after emergence from the opposite side of the lens) are singly refracted at some imaginary surface, instead of twice refracted (once at each lens surface) as actually happens. There is a unique imaginary surface, called the principal surface, at which this can happen.

To locate this unique surface, consider a single ray traced from the air on one side of the lens, through the lens and into the air on the other side. The ray is broken into three segments by the lens. Two of these are external (in the air), and the third is internal (in the glass). The external segments can be extended to a common point of intersection (certainly near, and usually within, the lens). The principal surface is the locus of all such

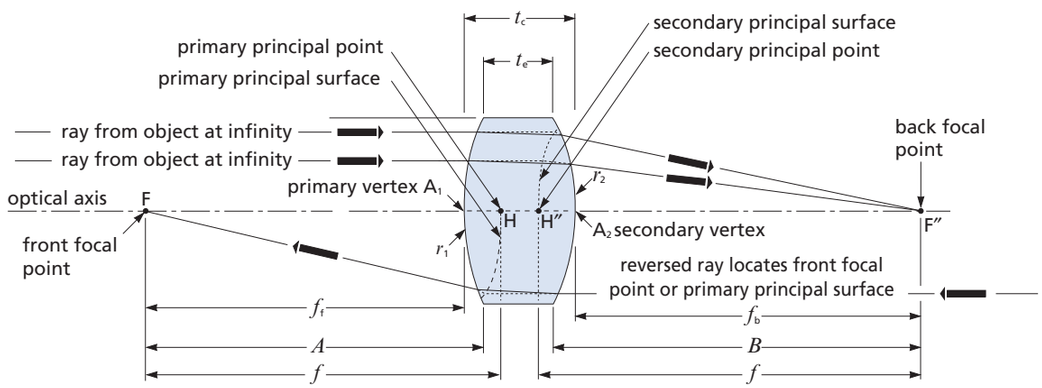

A = front focus to front edge distance

B = rear edge to rear focus distance

f = effective focal length; may be positive (as shown) or negative

f_f = front focal length

f_b = back focal length

t_e = edge thickness

t_c = center thickness

r_1 = radius of curvature of first surface (positive if center of curvature is to right)

r_2 = radius of curvature of second surface (negative if center of curvature is to left)

Figure 1.33 **Focal length and focal points**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:29 PM Page 1.30

points of intersection of extended external ray segments. The principal surface of a perfectly corrected optical system is a sphere centered on the focal point.

Near the optical axis, the principal surface is nearly flat, and for this reason, it is sometimes referred to as the principal plane.

SECONDARY PRINCIPAL SURFACE

This term is defined analogously to the primary principal surface, but it is used for a collimated beam incident from the left and focused to the back focal point F″ on the right. Rays in that part of the beam nearest the axis can be thought of as once refracted at the secondary principal surface, instead of being refracted by both lens surfaces.

PRIMARY PRINCIPAL POINT (H) OR FIRST NODAL POINT

This point is the intersection of the primary principal surface with the optical axis.

SECONDARY PRINCIPAL POINT (H″) OR SECONDARY NODAL POINT

This point is the intersection of the secondary principal surface with the optical axis.

CONJUGATE DISTANCES (s AND s'')

The conjugate distances are the object distance, s, and image distance, s''. Specifically, s is the distance from the object to H, and s'' is the distance from H″ to the image location. The term infinite conjugate ratio refers to the situation in which a lens is either focusing incoming collimated light or is being used to collimate a source (therefore, either s or s'' is infinity).

PRIMARY VERTEX (A_1)

The primary vertex is the intersection of the primary lens surface with the optical axis.

SECONDARY VERTEX (A_2)

The secondary vertex is the intersection of the secondary lens surface with the optical axis.

EFFECTIVE FOCAL LENGTH (EFL, f)

Assuming that the lens is surrounded by air or vacuum (refractive index 1.0), this is both the distance from the front focal point (F) to the primary principal point (H) and the distance from the secondary principal point (H″) to the rear focal point (F″). Later we use f to designate the paraxial EFL for the design wavelength (λ_0).

FRONT FOCAL LENGTH (f_f)

This length is the distance from the front focal point (F) to the primary vertex (A_1).

BACK FOCAL LENGTH (f_b)

This length is the distance from the secondary vertex (A_2) to the rear focal point (F″).

EDGE-TO-FOCUS DISTANCES (A AND B)

A is the distance from the front focal point to the primary vertex of the lens. B is the distance from the secondary vertex of the lens to the rear focal point. Both distances are presumed always to be positive.

REAL IMAGE

A real image is one in which the light rays actually converge; if a screen were placed at the point of focus, an image would be formed on it.

VIRTUAL IMAGE

A virtual image does not represent an actual convergence of light rays. A virtual image can be viewed only by looking back through the optical system, such as in the case of a magnifying glass.

F-NUMBER (f/#)

The f-number (also known as the focal ratio, relative aperture, or speed) of a lens system is defined to be the effective focal length divided by system clear aperture. Ray f-number is the conjugate distance for that ray divided by the height at which it intercepts the principal surface.

$$\mathrm{f/\#} = \frac{f}{\mathrm{CA}}. \qquad \textit{(see eq. 1.7)}$$

NUMERICAL APERTURE (NA)

The NA of a lens system is defined to be the sine of the angle, θ_1, that the marginal ray (the ray that exits the lens system at its outer edge) makes with the optical axis multiplied by the index of refraction (n) of the medium. The NA can be defined for any ray as the sine of the angle made by that ray with the optical axis multiplied by the index of refraction:

$$\mathrm{NA} = n\sin\theta. \qquad (1.30)$$

MAGNIFICATION POWER

Often, positive lenses intended for use as simple magnifiers are rated with a single magnification, such as 4 × . To create a virtual image for viewing with the human eye, in principle, any positive lens can be used at an infinite number of possible magnifications. However, there is usually a narrow range of magnifications that will be comfortable for the viewer. Typically, when the viewer adjusts the object distance so that the image appears to be essentially at infinity (which is a comfortable viewing distance for most individuals), magnification is given by the relationship

$$\text{magnification} = \frac{254\ \text{mm}}{f} \qquad (f \text{ in mm}). \tag{1.31}$$

Thus, a 25.4-mm focal length positive lens would be a 10 × magnifier.

DIOPTERS

The term diopter is used to define the reciprocal of the focal length, which is commonly used for ophthalmic lenses. The inverse focal length of a lens expressed in diopters is

$$\text{diopters} = \frac{1000}{f} \qquad (f \text{ in mm}). \tag{1.32}$$

Thus, the smaller the focal length is, the larger the power in diopters will be.

DEPTH OF FIELD AND DEPTH OF FOCUS

In an imaging system, depth of field refers to the distance in object space over which the system delivers an acceptably sharp image. The criteria for what is acceptably sharp is arbitrarily chosen by the user; depth of field increases with increasing f-number.

For an imaging system, depth of focus is the range in image space over which the system delivers an acceptably sharp image. In other words, this is the amount that the image surface (such as a screen or piece of photographic film) could be moved while maintaining acceptable focus. Again, criteria for acceptability are defined arbitrarily.

In nonimaging applications, such as laser focusing, depth of focus refers to the range in image space over which the focused spot diameter remains below an arbitrary limit.

APPLICATION NOTE

Technical Reference

For further reading about the definitions and formulas presented here, refer to the following publications:

Rudolph Kingslake, *Lens Design Fundamentals* (Academic Press)

Rudolph Kingslake, System Design (Academic Press)

Warren Smith, *Modern Optical Engineering* (McGraw Hill).

Donald C. O'Shea, *Elements of Modern Optical Design* (John Wiley & Sons)

Eugene Hecht, *Optics* (Addison Wesley)

Max Born, Emil Wolf, *Principles of Optics* (Cambridge University Press)

If you need help with the use of definitions and formulas presented in this guide, our applications engineers will be pleased to assist you.

Paraxial Lens Formulas

PARAXIAL FORMULAS FOR LENSES IN AIR

The following formulas are based on the behavior of paraxial rays, which are always very close and nearly parallel to the optical axis. In this region, lens surfaces are always very nearly normal to the optical axis, and hence all angles of incidence and refraction are small. As a result, the sines of the angles of incidence and refraction are small (as used in Snell's law) and can be approximated by the angles themselves (measured in radians).

The paraxial formulas do not include effects of spherical aberration experienced by a marginal ray — a ray passing through the lens near its edge or margin. All EFL values (f) tabulated in this catalog are paraxial values which correspond to the paraxial formulas. The following paraxial formulas are valid for both thick and thin lenses unless otherwise noted. The refractive index of the lens glass, n, is the ratio of the speed of light in vacuum to the speed of light in the lens glass. All other variables are defined in figure 1.33.

Focal Length

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) + \frac{(n-1)^2}{n}\,\frac{t_c}{r_1 r_2} \qquad (1.33)$$

where n is the refractive index, t_c is the center thickness, and the sign convention previously given for the radii r_1 and r_2 applies. For thin lenses, $t_c \cong 0$, and for plano lenses either r_1 or r_2 is infinite. In either case the second term of the above equation vanishes, and we are left with the familiar lens maker's formula:

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right). \qquad (1.34)$$

Surface Sagitta and Radius of Curvature

(refer to figure 1.34)

$$r^2 = (r-s)^2 + \left(\frac{d}{2}\right)^2 \qquad (1.35)$$

$$s = r - \sqrt{r^2 - \left(\frac{d}{2}\right)^2} > 0 \qquad (1.36)$$

$$r = \frac{s}{2} + \frac{d^2}{8s}. \qquad (1.37)$$

An often useful approximation is to neglect $s/2$.

Symmetric Lens Radii ($r_2 = -r_1$)

With center thickness constrained,

$$r_1 = (n-1)\left[f \pm \sqrt{f^2 - \left(\frac{f t_c}{n}\right)}\right] \qquad (1.38)$$

$$= (n-1) f \left[1 + \sqrt{1 - \left(\frac{t_c}{nf}\right)}\right]$$

where, in the first form, the + sign is chosen for the square root if f is positive, but the − sign must be used if f is negative. In the second form, the + sign must be used regardless of the sign of f.

Plano Lens Radius

Since r_2 is infinite,

$$r_1 = (n-1) f. \qquad (1.39)$$

Principal-Point Locations (signed distances from vertices)

$$A_2H'' = \frac{-r_2 t_c}{n(r_2 - r_1) + t_c(n-1)} \qquad (1.40)$$

$$A_1H = \frac{-r_1 t_c}{n(r_2 - r_1) + t_c(n-1)} \qquad (1.41)$$

where the above sign convention applies.

For symmetric lenses ($r_2 = -r_1$),

$$A_1H = -A_2H''$$

$$= \frac{r_1 t_c}{2nr_1 - t_c(n-1)}. \qquad (1.42)$$

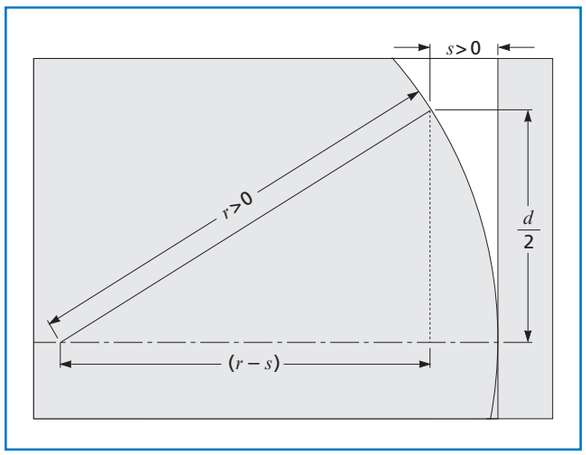

Figure 1.34 **Surface sagitta and radius of curvature**

If either r_1 or r_2 is infinite, l'Hôpital's rule from calculus must be used. Thus, referring to *Aberration Balancing*, for plano-convex lenses in the correct orientation,

$$A_1H = 0$$

and

$$A_2H'' = -\frac{t_c}{n}. \tag{1.43}$$

For flat plates, by letting $r_1 \to \infty$ in a symmetric lens, we obtain $A_1H = A_2H'' = t_c/2_n$. These results are useful in connection with the following paraxial lens combination formulas.

Hiatus or Interstitium (principal-point separation)

$$HH'' = t_c\left\{1 - \frac{f}{n}\left[\frac{1}{f} - \frac{(n-1)^2}{n}\frac{t_c}{r_1 r_2}\right]\right\} \tag{1.44}$$

which, in the thin-lens approximation (exact for plano lenses), becomes

$$HH'' = t_c\left(1 - \frac{1}{n}\right). \tag{1.45}$$

Solid Angle

The solid angle subtended by a lens, for an observer situated at an on-axis image point, is

$$\begin{aligned}\Omega &= 2\pi(1-\cos\theta)\\ &= 4\pi\sin^2\left(\frac{\theta}{2}\right)\end{aligned} \tag{1.46}$$

where this result is in steradians, and where

$$\theta = \arctan\left(\frac{CA}{2s}\right) \tag{1.47}$$

is the apparent angular radius of the lens clear aperture. For an observer at an on-axis object point, use s instead of s''. To convert from steradians to the more intuitive sphere units, simply divide Ω by 4π. If the Abbé sine condition is known to apply, θ may be calculated using the arc sine function instead of the arctangent.

Back Focal Length

$$\begin{aligned}f_b &= f'' + A_2H''\\ &= f'' - \frac{r_2 t_c}{n(r_2 - r_1) + t_c(n-1)}\end{aligned} \tag{1.48}$$

where the sign convention presented above applies to A_2H'' and to the radii. If r_2 is infinite, l'Hôpital's rule from calculus must be used, whereby

$$f_b = f'' - \frac{t_c}{n}. \tag{1.49}$$

Front Focal Length

$$\begin{aligned}f_f &= f - A_1H\\ &= f + \frac{r_1 t_c}{n(r_2 - r_1) + t_c(n-1)}\end{aligned} \tag{1.50}$$

where the sign convention presented above applies to A_1H and to the radii. If r_1 is infinite, l'Hôpital's rule from calculus must be used, whereby

$$f_f = f - \frac{t_c}{n}. \tag{1.51}$$

Edge-to-Focus Distances

For positive lenses,

$$A = f_f + s_1 \tag{1.52}$$

and

$$B = f_b + s_2 \tag{1.53}$$

where s_1 and s_2 are the sagittas of the first and second surfaces. Bevel is neglected.

Magnification or Conjugate Ratio

$$\begin{aligned}m &= \frac{s''}{s}\\ &= \frac{f}{s-f}\\ &= \frac{s''-f}{f}.\end{aligned} \tag{1.54}$$

PARAXIAL FORMULAS FOR LENSES IN ARBITRARY MEDIA

These formulas allow for the possibility of distinct and completely arbitrary refractive indices for the object space medium (refractive index n'), lens (refractive index n''), and image space medium (refractive index n). In this situation, the EFL assumes two distinct values, namely f in object space and f'' in image space. It is also necessary to distinguish the principal points from the nodal points. The lens serves both as a lens and as a window separating the object space and image space media.

The situation of a lens immersed in a homogenous fluid (figure 1.35) is included as a special case ($n = n''$). This case is of considerable practical importance. The two values f and f'' are again equal, so that the lens combination formulas are applicable to systems immersed in a common fluid. The general case (two different fluids) is more difficult, and it must be approached by ray tracing on a surface-by-surface basis.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:30 PM Page 1.34

Lens Constant (*k*)

This number appears frequently in the following formulas. It is an explicit function of the complete lens prescription (both radii, t_c and n') and both media indices (n and n''). This dependence is implicit anywhere that k appears.

$$k = \frac{n'-n}{r_1} + \frac{n''-n'}{r_2} - \frac{t_c(n'-n)(n''-n')}{n'r_1r_2}. \qquad (1.55)$$

Effective Focal Lengths

$$f = \frac{n}{k} \qquad f'' = \frac{n''}{k}. \qquad (1.56)$$

Lens Formula (Gaussian form)

$$\frac{n}{s} + \frac{n''}{s''} = k. \qquad (1.57)$$

Lens Formula (Newtonian form)

$$xx'' = f''f = \frac{nn''}{k^2}. \qquad (1.58)$$

Principal-Point Locations

$$\mathrm{A_1H} = \frac{nt_c}{k}\left(\frac{n''-n'}{n'r_2}\right) \qquad (1.59)$$

$$\mathrm{A_2H''} = \frac{-n''t_c}{k}\left(\frac{n'-n}{n'r_1}\right). \qquad (1.60)$$

Object-to-First-Principal-Point Distance

$$s = \frac{ns''}{ks''-n''}. \qquad (1.61)$$

Second Principal-Point-to-Image Distance

$$s'' = \frac{n''s}{ks-n}. \qquad (1.62)$$

Magnification

$$m = \frac{ns''}{n''s}. \qquad (1.63)$$

Lens Maker's Formula

$$\frac{n}{f} = \frac{n''}{f''} = k. \qquad (1.64)$$

Nodal-Point Locations

$$\mathrm{A_1N} = \mathrm{A_1H} + \mathrm{HN} \qquad (1.65)$$

$$\mathrm{A_2N''} = \mathrm{A_2H''} + \mathrm{H''N''}. \qquad (1.66)$$

Separation of Nodal Point from Corresponding Principal Point

HN = H″N″ = $(n''-n)/k$, positive for N to right of H and N″ to right of H″.

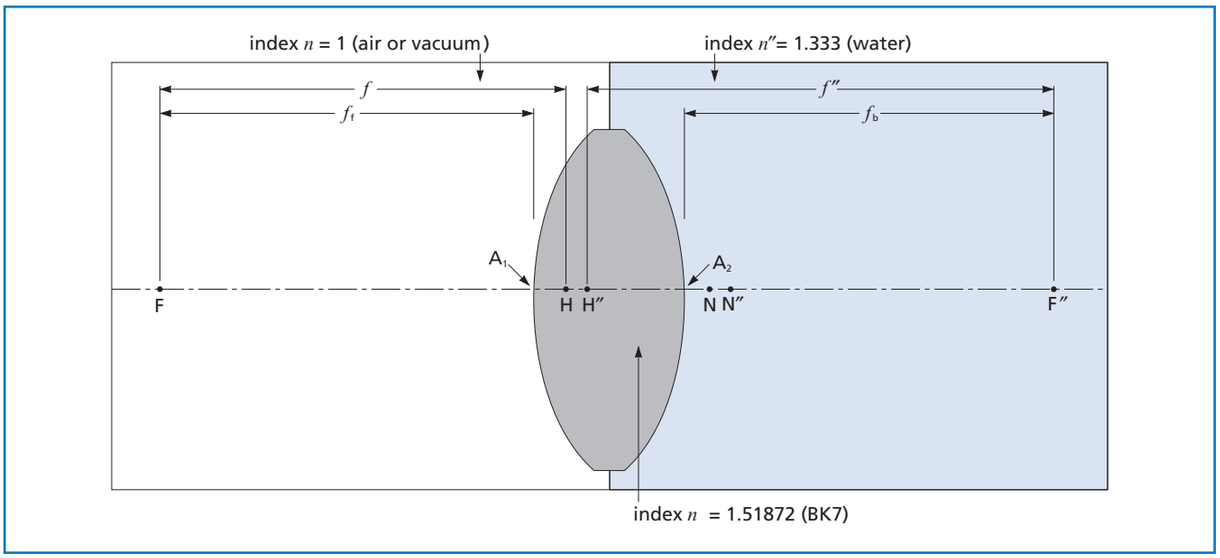

Figure 1.35 **Symmetric lens with disparate object and image space indexes**

Back Focal Length

$$f_b = f'' + A_2H''. \qquad \text{(see eq. 1.48)}$$

Front Focal Length

$$f_f = f - A_1H. \qquad \text{(see eq. 1.50)}$$

Focal Ratios

The focal ratios are f/CA and f''/CA, where CA is the diameter of the clear aperture of the lens.

Numerical Apertures

$$n \sin\theta$$

$$\text{where } \theta = \arcsin\left(\frac{CA}{2s}\right)$$

and

$$n'' \sin\theta''$$

$$\text{where } \theta'' = \arcsin\left(\frac{CA}{2s''}\right).$$

Solid Angles (in steradians)

$$\Omega = 2\pi(1 - \cos\theta) \qquad \text{(see eq. 1.46)}$$

$$= 4\pi \sin^2\left(\frac{\theta}{2}\right)$$

$$\text{where } \theta = \arctan\left(\frac{CA}{2s}\right)$$

$$\Omega = 2\pi(1 - \cos\theta'')$$

$$= 4\pi \sin^2\left(\frac{\theta}{2}\right)$$

$$\text{where } \theta'' = \arctan\left(\frac{CA}{2s''}\right).$$

To convert from steradians to spheres, simply divide by 4π.

APPLICATION NOTE

For Quick Approximations

Much time and effort can be saved by ignoring the differences among f, f_b, and f_f in these formulas by assuming that $f = f_b = f_f$. Then s becomes the lens-to-object distance; s'' becomes the lens-to-image distance; and the sum of conjugate distances $s + s''$ becomes the object-to-image distance. This is known as the thin-lens approximation.

APPLICATION NOTE

Physical Significance of the Nodal Points

A ray directed at the primary nodal point N of a lens appears to emerge from the secondary nodal point N″ without change of direction. Conversely, a ray directed at N″ appears to emerge from N without change of direction. At the infinite conjugate ratio, if a lens is rotated about a rotational axis orthogonal to the optical axis at the secondary nodal point (i.e., if N″ is the center of rotation), the image remains stationary during the rotation. This fact is the basis for the nodal slide method for measuring nodal-point location and the EFL of a lens. The nodal points coincide with their corresponding principal points when the image space and object space refractive indices are equal ($n = n''$). This makes the nodal slide method the most precise method of principal-point location.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:30 PM Page 1.36

Principal-Point Locations

Figure 1.36 indicates approximately where the principal points fall in relation to the lens surfaces for various standard lens shapes. The exact positions depend on the index of refraction of the lens material, and on the lens radii, and can be found by formula. in extreme meniscus lens shapes (short radii or steep curves), it is possible that both principal points will fall outside the lens boundaries. For symmetric lenses, the principal points divide that part of the optical axis between the vertices into three approximately equal segments. For plano lenses, one principal point is at the curved vertex, and the other is approximately one-third of the way to the plane vertex.

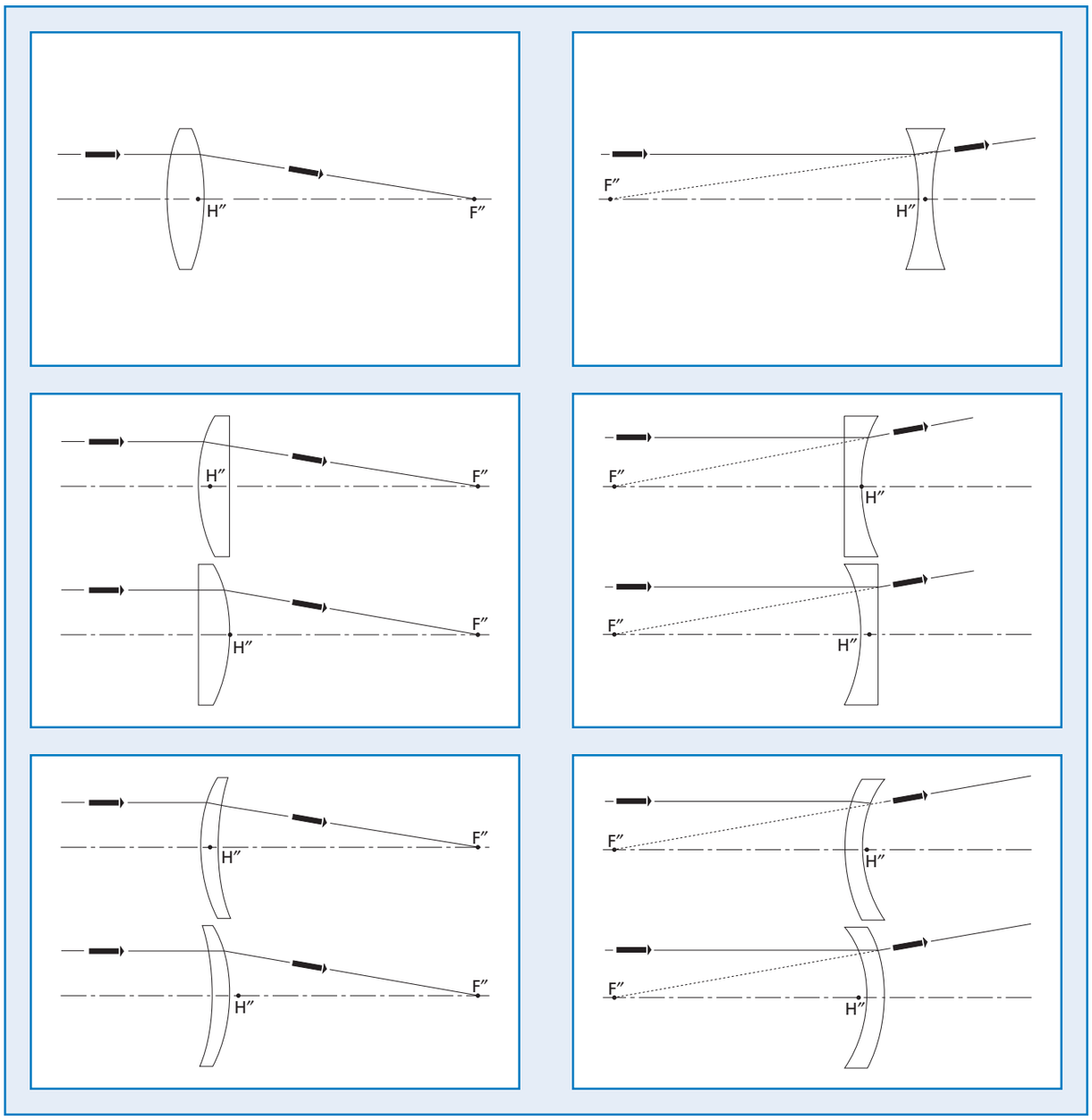

Figure 1.36 **Principal points of common lenses**

1ch_FundamentalOptics_Final_a.qxd 7/6/2009 1:54 PM Page 1.37

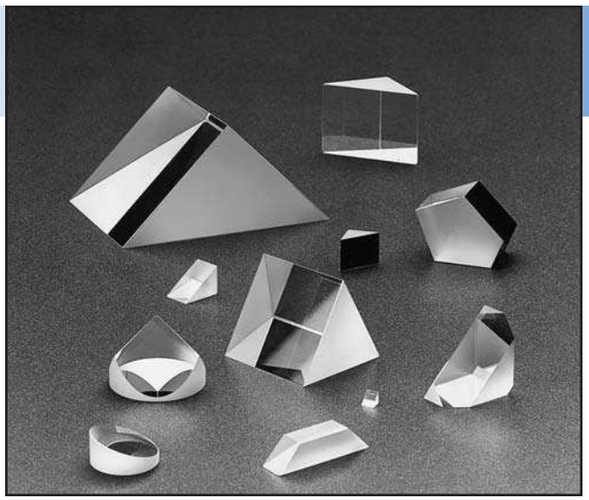

Prisms

Prisms are blocks of optical material whose flat, polished sides are arranged at precisely controlled angles to each other. Prisms may be used in an optical system to deflect or deviate a beam of light. They can invert or rotate an image, disperse light into its component wavelengths, and be used to separate states of polarization.

PRISM ORIENTATION

The orientation of a prism determines its effect on a beam of light or an image.

A viewer looks through a prism at an object and sees a virtual image (see Figure 1.37). This image may be displaced from the original object, or, if a dove prism is used, it may coincide with the object. Furthermore, image orientation may differ from the object; in the case of a right-angle prism, the image is reversed.

A real image (see Figure 1.38) can be formed only if imaging optics are present in the system. Without imaging optics, the image is virtual. A virtual image has the same orientation as the real image shown, but it can be viewed by the observer only by looking back through the prism system.

TOTAL INTERNAL REFLECTION

Rays incident upon a glass/air boundary (i.e., an internal reflection) at angles that exceed the critical angle are reflected with 100-percent efficiency regardless of their initial polarization state. The critical angle is given by

$$\theta_c(\lambda) = \arcsin\frac{1}{n_\lambda} \qquad (1.67)$$

and depends on the refractive index, which is a function of wavelength. If, at some wavelength, the refractive index should fall to less than $\sqrt{2} = 1.414$, the critical angle will exceed 45 degrees, and total internal reflection (TIR) will fail for a collimated beam internally incident at 45 degrees on the hypotenuse face of a right-angle prism. Reflectance decreases rapidly at angles of incidence smaller than the critical angle.

The index of BK7 is sufficiently high to guarantee the TIR of a collimated beam at 45 degrees internal incidence over the visible and near-infrared region. The possibility of significant TIR failure with convergent or divergent beams should be kept in mind if polarization is important. TIR can also fail if the hypotenuse face is not kept extremely clean. Even an almost invisible fingerprint can lead to TIR failure. An aluminum- or silver-coated hypotenuse is recommended for applications where the right-angle prism is frequently handled, or where convergent or divergent beams are used. There is a slight loss of reflectance at all internal angles with the coating, and no critical angle exists.

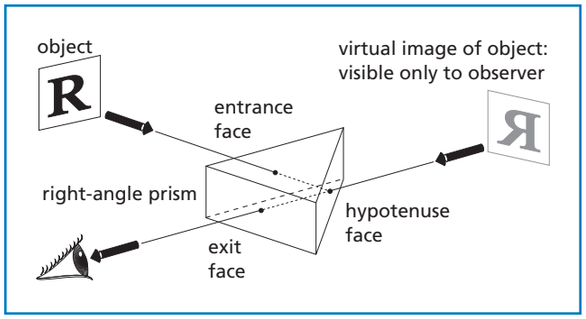

Figure 1.37 **Virtual imaging using a prism**

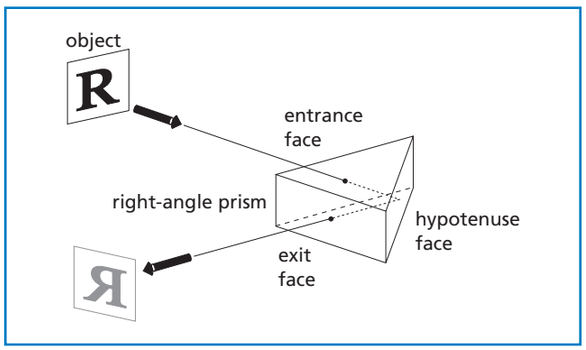

Figure 1.38 **Real imaging using a prism**

ABERRATIONS FOR PRISMS

Prisms will introduce aberrations when they are used with convergent or divergent beams of light. Using prisms with collimated or nearly collimated light will help minimize aberrations. Conjugate distances that include prisms should be long.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:30 PM Page 1.38

DISPERSING PRISMS

Dispersing prisms are used to separate a beam of white light into its component colors. Generally, the light is first collimated and then dispersed by the prism. A spectrum is then formed at the focal plane of a lens or curved mirror. In laser work, dispersing prisms are used to separate two wavelengths following the same beam path. Typically, the dispersed beams are permitted to travel far enough so the beams separate spatially.

A prism exhibits magnification in the plane of dispersion if the entrance and exit angles for a beam differ. This is useful in anamorphic (one-dimensional) beam expansion or compression, and may be used to correct or create asymmetric beam profiles.

As shown in Figure 1.39, a beam of width W_1 is incident at an angle α on the surface of a dispersing prism of apex angle A. The angle of refraction at the first surface, β, the angle of incidence at the second surface, γ, and the angle of refraction exiting the prism, δ, are easily calculated:

$$\beta = \sin^{-1}((\sin\alpha) / \eta)$$
$$\gamma = A - \beta$$
$$\delta = \sin^{-1}(\eta \sin\gamma)$$

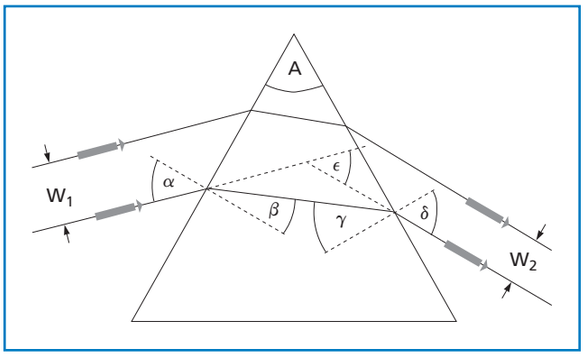

Figure 1.39 **Diagram of dispersing prism**

The beam deviation, ϵ, is of greatest importance. It is the angle the exit beam makes with its original direction.

$$\epsilon = \alpha + \delta - A \qquad (1.68)$$

The magnification W_2/W_1 is given by:

$$M = \frac{\cos\delta\cos\beta}{\cos\alpha\cos\gamma} \qquad (1.69)$$

The resolving power of a prism spectrometer angle α, the angular dispersion of the prism is given by:

$$\frac{d\delta}{d\lambda} = \left(\frac{\sin A}{\cos\delta\cos\beta}\right)\left(\frac{d\eta}{d\lambda}\right) \qquad (1.70)$$

If the spectrum is formed by a diffraction limited focal system of focal length f, the minimum spot size is $dx \sim f\lambda / W$. This corresponds to a minimum angular resolution $d\delta \sim \lambda/w$ for a beam of diameter w. The diffraction limited angular resolution at a given beam diameter sets the limit on the spectral resolving power of a prism. Setting the expression for $d\delta$ equal to the minimum angular resolution, we obtain:

$$RP = \frac{\lambda}{d\lambda} = \left(\frac{w_2 \sin A}{\cos\delta\cos\beta}\right)\left(\frac{d\eta}{d\lambda}\right) \qquad (1.71)$$

where RP is the resolving power of the prism.

At a given wavelength, the beam deviation ϵ is a minimum at an angle of incidence:

$$\alpha_{\text{min dev}} = \sin^{-1}[\eta \sin(A/2)] \qquad (1.72)$$

where η is the prism index of refraction at that wavelength. At this angle, the incident and exit angles are equal, the prism magnification is one, and the internal rays are perpendicular to the bisector of the apex angle.

By measuring the angle of incidence for minimum deviation, the index of refraction of a prism can be determined. Also, by proper choice of apex angle, the equal incident and exit angles may be made Brewster's angle, eliminating losses for p-polarized beams. The apex angle to choose is:

$$A = \pi - 2\theta_B \qquad (1.73)$$

If, in addition, the base angles of the prism are chosen as Brewster's angle, an isosceles Brewster prism results.

Another use is illustrated next.

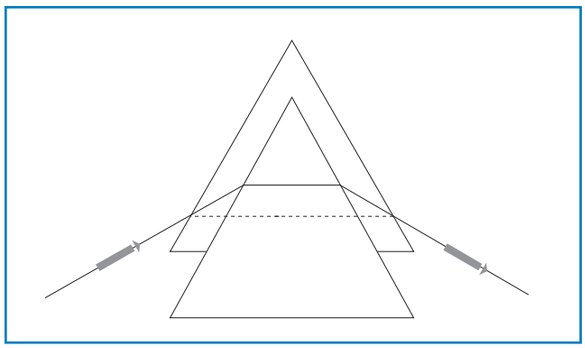

Figure 1.40 **Translation of a prism at minimum deviation**

At minimum deviation, translating a prism along the bisector of the apex angle does not disturb the direction of the output rays. See Figure 1.40. This is important in femtosecond laser design where intracavity prisms are used to compensate for group velocity dispersion. By aligning a prism for minimum deviation and translating it along its apex bisector, the optical path length in material may be varied with no misalignment, thus varying the contribution of the material to overall group velocity dispersion. Finally, it is possible to show that at minimum deviation

$$RP = (b_2 - b_1)\frac{d\eta}{d\lambda} \qquad (1.74)$$

where the relevant quantities are defined in Figure 1.41.

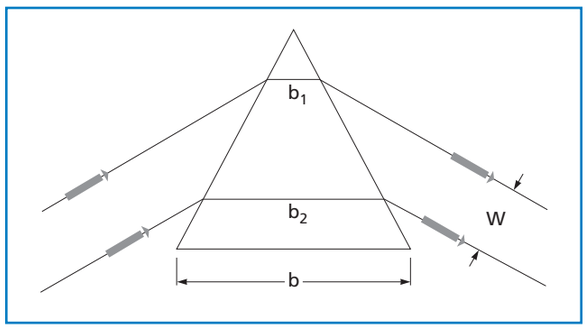

Figure 1.41 **Ray path lengths of a prism at minimum deviation**

If the beam is made to fill the prism completely, $b_1 = 0$, and $b_2 = b$, the base of the prism. So, we have the classical result that the resolving power of a prism spectrometer is equal to the base of the prism times the dispersion of the prism material.

As an example, consider CVI Melles Griot EDP-25-F2 prism, operating in minimum deviation at 590 nm. The angle of incidence and emergence are both then 54.09° and $d\eta/d\lambda$ is $-0.0854\ \mu m^{-1}$ for F2 glass at 590 nm. If the 25-mm prism is completely filled, the resolving power, $\lambda/d\lambda$, is 2135. This is sufficient to resolve the Sodium D lines.

PELLIN BROCA PRISMS

In a Pellin Broca prism, an ordinary dispersing prism is split in half along the bisector of the apex angle. Using a right angle prism, the two halves are joined to create a dispersing prism with an internal right angle bend obtained by total internal reflection, as shown in Figure 1.42.

In principle, one can split any type of dispersing prism to create a Pellin Broca prism. Typically the Pellin Broca prism is based on an Isosceles Brewster prism. Provided the light is *p*-polarized, the prism will be essentially lossless.

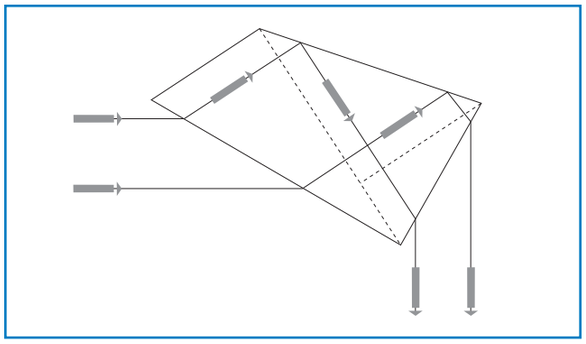

Figure 1.42 **One of the wavelengths deviates at exactly 90° to its initial direction**

Suppose wavelengths λ_1 and λ_2 are superimposed in a collimated beam, as at the output of a harmonic generating crystal, the diagram in Figure 1.42 suggests that it is always possible to find a rotation of the prism in its plane that ensures that one of the two wavelengths will operate at minimum deviation when refracting at the input face of the first of the half-dispersing prisms. This means that it will enter the right angle prism normal to one of its faces, be turned exactly 90°, be presented to the second half-dispersing prism in minimum deviation, and hence exit the Pellin Broca prism deviated at exactly 90° to its initial direction.

A simple dispersing prism always deviates the longer wavelength less than the shorter wavelength. In a Pellin Broca prism, whether the longer wavelength is deviated more or less depends on the orientation of the prism. This is an important consideration when designing a high power Pellin Broca beam separator, as shown in Figures 1.43 and 1.44.

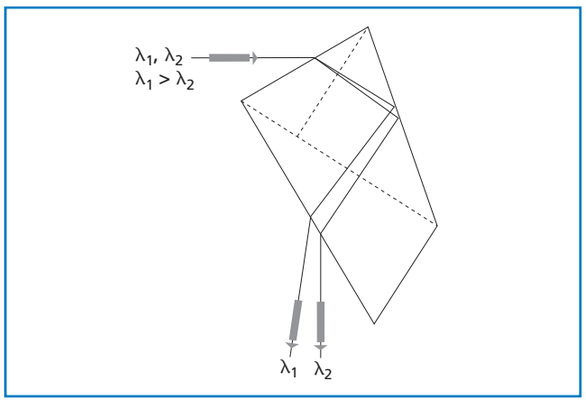

Figure 1.43 **Longer wavelength is deviated more than the shorter wavelength**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:31 PM Page 1.40

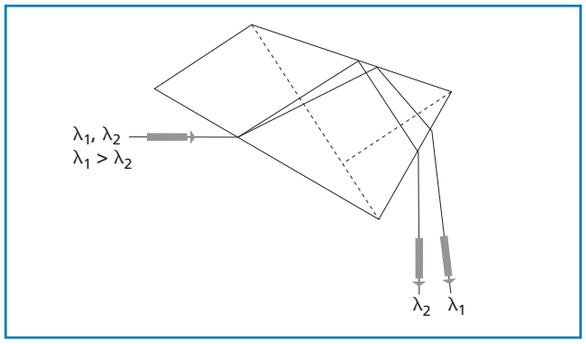

Figure 1.44 **Longer wavelength is deviated less than the shorter wavelength**

CVI Melles Griot offers Brewster angle Pellin Brocca prisms in a number of sizes and materials. BK7 prisms are used in the visible and near IR, and is the least expensive. UV-grade fused silica Pellin Broca prisms are used from 240 nm to 2000 nm. Excimer-grade prisms are used in the 180-nm to 240-nm region. Crystal-quartz Pellin Broca prisms are specifically designed for high-power Q-switched 266-nm laser pulses at fluence levels of 50 mJ/cm^2. Fused silica prisms track (i.e., suffer internal catastrophic damage) above this fluence, probably due to self-focusing.

PORRO PRISMS

A Porro prism, named for its inventor Ignazio Porro, is a type of reflection prism used to alter the orientation of an image. In operation, light enters the large face of the prism, undergoes total internal reflection twice from the 45° sloped faces, and exits again through the large face. An image traveling through a Porro prism is rotated by 180° and exits in the opposite direction offset from its entrance point, as shown in Figure 1.45. Since the image is reflected twice, the handedness of the image is unchanged. Porro prisms have rounded edges to minimize breakage and facilitate assembly.

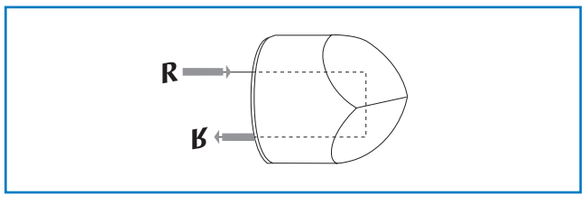

Figure 1.45 **Porro prisms retroreflect and invert the image**

Porro prisms are most often used in pairs, forming a double Porro prism, as shown in Figure 1.46. A second prism, rotated 90° with respect to the first, is placed such that the beam will traverse both prisms. The net effect of the prism system is a beam parallel to but displaced from its original direction, with the image rotated 180°. As before, the handedness of the image is unchanged.

Double Porro prism systems are used in small optical telescopes to reorient an inverted image and in many binoculars to both re-orient the image and provide a longer, folded distance between the objective lenses and the eyepieces.

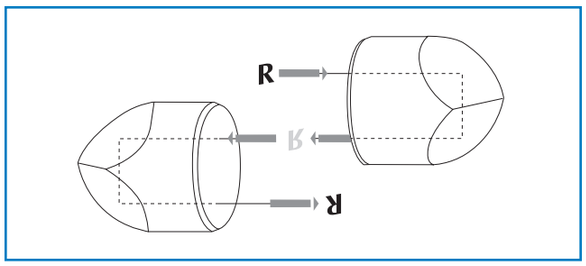

Figure 1.46 **Double porro prism results in a beam parallel to but displaced from its original direction, with the image rotated 180°**

1ch_FundamentalOptics_Final_a.qxd 7/21/2009 9:41 AM Page 1.41

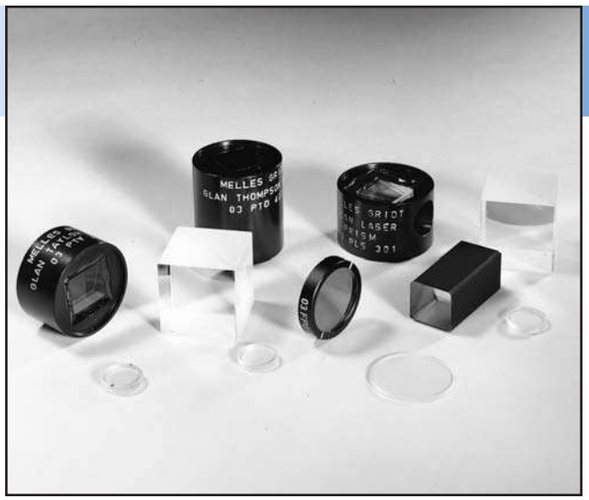

Polarization

Polarization States

Four numbers are required to describe a single plane wave Fourier component traveling in the $+z$ direction. These can be thought of as the amplitude and phase shift of the field along two orthogonal directions.

1. CARTESIAN REPRESENTATION

In Cartesian coordinates, the propagation equation for an electric field is given by the formula

$$E(x,y,z) = (xE_x e^{i\phi_x} + yE_y e^{i\phi_y})e^{i(kz-\omega t)} \qquad (1.75)$$

where E_x, E_y, ϕ_x, and ϕ_y are real numbers defining the magnitude and the phase of the field components in the orthogonal unit vectors x and y. If the origin of time is irrelevant, only the relative phase shift

$$\phi = \phi_x - \phi_y \qquad (1.76)$$

need be specified.

2. CIRCULAR REPRESENTATION

In the circular representation, we resolve the field into circularly polarized components. The basic states are represented by the complex unit vectors

$$e_+ = \left(1/\sqrt{2}\right)(x+iy) \text{ and} \qquad (1.77)$$

$$e_- = \left(1/\sqrt{2}\right)(x-iy) \qquad (1.78)$$

where e_+ is the unit vector for left circularly polarized light; for positive helicity light; for light that rotates counterclockwise in a fixed plane as viewed facing into the light wave; and for light whose electric field rotation obeys the right hand rule with thumb pointing in the direction of propagation.

APPLICATION NOTE

Polarization Convention

Historically, the orientation of a polarized electromagnetic wave has been defined in the optical regime by the orientation of the electric vector. This is the convention used by CVI Melles Griot.

e_- is the unit vector for right circularly polarized light; for negative helicity light; for light that rotates clockwise in a fixed plane as viewed facing into the light wave; and for light whose electric field rotation disobeys the right hand rule with thumb pointing in the direction of propagation.

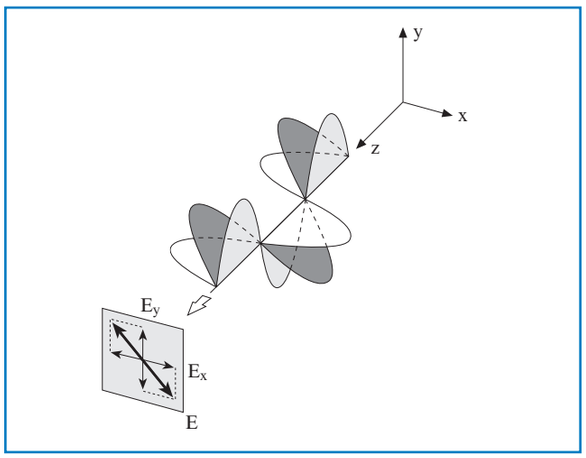

Figure 1.47 **Linearly polarized light. E_x and E_y are in phase**

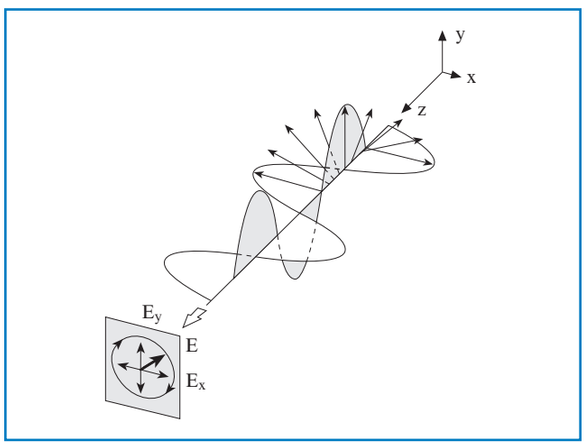

Figure 1.48 **Circularly polarized light. E_x and E_y are out of phase by angular frequency ω**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:31 PM Page 1.42

As in the case of the Cartesian representation, we write

$$E = (e_+ E_+ e^{i\phi +} + e_- E_- e^{i\phi -})\, e^{i(kz-\omega t)} \tag{1.79}$$

where E_+, E_-, ϕ_+, and ϕ_- are four real numbers describing the magnitudes and phases of the field components of the left and right circularly polarized components. Note that

$$E_+ = e_- \bullet E \tag{1.80}$$

$$E_- = e_+ \bullet E \tag{1.81}$$

3. ELLIPTICAL REPRESENTATION

An arbitrary polarization state is generally elliptically polarized. This means that the tip of the electric field vector will describe an ellipse, rotating once per optical cycle.

Let a be the semimajor and b be the semiminor axis of the polarization ellipse. Let ψ be the angle that the semimajor axis makes with the x axis.

Let ξ and η be the axes of a right-handed coordinate system rotated by an angle $+\psi$ with respect to the x axis and aligned with the polarization ellipse as shown in Figure 1.49.

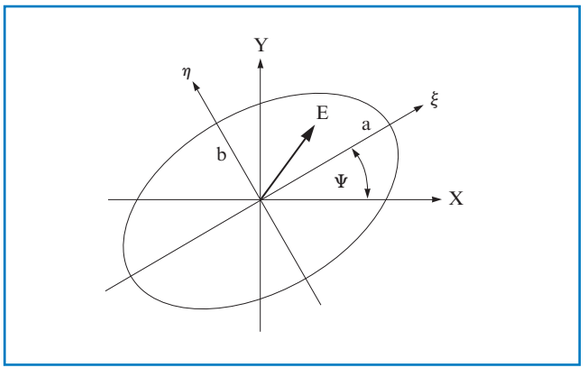

Figure 1.49 **The polarization ellipse**

The elliptical representation is:

$$E = \left(a\hat{\xi} + b\hat{\eta}\right) e^{i\delta_0} e^{i(kz-\omega t)} \tag{1.82}$$

Note that the phase shift δ_0 above is required to adjust the time origin, and the parameter ψ is implicit in the rotation of the ξ and η axes with respect to the x and y axes.

CONVERSIONS BETWEEN REPRESENTATIONS

For brevity, we will provide only the Cartesian to circular and Cartesian to elliptical transformations. The inverse transformations are straightforward. We define the following quantities:

$$g_1 = E_x \cos\phi_x - E_y \sin\phi_y \tag{1.83}$$

$$g_2 = E_x \sin\phi_x + E_y \cos\phi_y \tag{1.84}$$

$$g_3 = E_x \cos\phi_x + E_y \sin\phi_y \tag{1.85}$$

$$g_4 = E_x \sin\phi_x - E_y \cos\phi_y \tag{1.86}$$

$$u = \left(g_1^2 + g_2^2\right)^{1/2} \tag{1.87}$$

$$v = \left(g_3^2 + g_4^2\right)^{1/2} \tag{1.88}$$

$$\phi_{12} = \text{atan}\,(g_1, g_2) \tag{1.89}$$

$$\phi_{34} = \text{atan}\,(g_3, g_4) \tag{1.90}$$

In the above, atan(x,y) is the four quadrant arc tangent function. This means that atan(x,y) = atan(y/x) with the provision that the quadrant of the angle returned by the function is controlled by the signs of both x and y, not just the sign of their quotient; for example, if $g_2 = g_1 = -1$, then ϕ_{12} above is $5\pi/4$ or $-3\pi/4$, not $\pi/4$.

A. CARTESIAN TO CIRCULAR TRANSFORMATION

$$E_+ = v/\sqrt{2} \tag{1.91}$$

$$E_- = u/\sqrt{2} \tag{1.92}$$

$$\phi_+ = \phi_{34} \tag{1.93}$$

$$\phi_- = \phi_{12} \tag{1.94}$$

B. CARTESIAN TO ELLIPTICAL TRANSFORMATION

$$a = (v+u)/2 \tag{1.95}$$

$$b = (v-u)/2 \tag{1.96}$$

$$\psi = (\phi_{12} - \phi_{34})/2 \tag{1.97}$$

$$\delta_0 = (\phi_{12} + \phi_{34})/2 \tag{1.98}$$

LINEAR POLARIZERS

A linear polarizer is a device that creates a linear polarization state from an arbitrary input. It does this by removing the component orthogonal to the selected state. Unlike plastic sheet polarizers which absorb the rejected beam (which turns into heat), cube polarizers and thin-film plate polarizers reflect the rejected beam, creating two usable beams. Still others may refract the two polarized beams at different angles, thereby separating them. Examples are Wollaston and Rochon prism polarizers.

Suppose the pass direction of the polarizer is determined by unit vector p. Then the transmitted field E_2, in terms of the incident field E_1, is given by

$$E_2 = p\left(p \bullet E_1\right) \tag{1.99}$$

where the phase shift of the transmitted field has been ignored.

A real polarizer has a pass transmission, $T_{\|}$, less than 1. The transmission of the rejected beam, $T_{\perp}$, may not be 0. If r is a unit vector along the rejected direction, then

$$E_2 = \left(T_{\|}\right)^{1/2} p\left(p \bullet E_1\right)e^{i\phi_{\|}} + \left(T_{\perp}\right)^{1/2} r\left(r \bullet E_1\right)e^{i\phi_{\perp}} \tag{1.100}$$

In the above, the phase shifts along the two directions must be retained. Similar expressions could be arrived at for the rejected beam. If θ is the angle between the field E_1 and the polarizer pass direction p, the above equation predicts that

$$T = T_{\|}\cos^2\theta + T_{\perp}\sin^2\theta \tag{1.101}$$

The above equation shows that, when the polarizer is aligned so that $\theta = 0$, $T = T_{\|}$. When it is "crossed", $\theta = \pi/2$, and $T = T_{\perp}$. The extinction ratio is $\epsilon = T_{\|} / T_{\perp}$. A polarizer with perfect extinction has $T_{\perp} = 0$, and thus $T = T_{\|}\cos^2\theta$ is a familiar result. Because $\cos^2\theta$ has a broad maximum as a function of orientation angle, setting a polarizer at a maximum of transmission is generally not very accurate. One has to either map the $\cos^2\theta$ with sufficient accuracy to find the $\theta = 0$ point, or do a null measurement at $\theta = \pm\pi/2$.

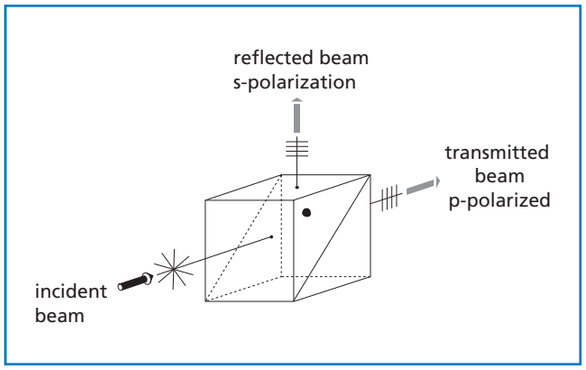

Figure 1.50 **At CVI Melles Griot, the DOT marks preferred input face. This is the tested direction for transmitted wavefront. Damage threshold is also higher for this orientation as well.**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:31 PM Page 1.44

Polarization Definitions

BIREFRINGENCE

A birefringent crystal, such as calcite, will divide an entering beam of monochromatic light into two beams having opposite polarization. The beams usually propagate in different directions and will have different speeds. There will be only one or two optical axis directions within the crystal in which the beam will remain collinear and continue at the same speed, depending on whether the birefringent crystal is uniaxial or biaxial.

If the crystal is a plane-parallel plate, and the optical axis directions are not collinear with the beam, radiation will emerge as two separate, orthogonally polarized beams (see Figure 1.51). The beam will be unpolarized where the beams overlap upon emergence. The two new beams within the material are distinguished from each other by more than just polarization and velocity. The rays are referred to as extraordinary (E) and ordinary (O). These rays need not be confined to the plane of incidence. Furthermore, the velocity of these rays changes with direction. Thus, the index of refraction for extraordinary rays is also a continuous function of direction. The index of refraction for the ordinary ray is constant and is independent of direction.

The two indexes of refraction are equal only in the direction of an optical axis within the crystal. The dispersion curve for ordinary rays is a single, unique curve when the index of refraction is plotted against wavelength. The dispersion curve for the extraordinary ray is a family of curves with different curves for different directions. Unless it is in a particular polarization state, or the crystalline surface is perpendicular to an optical axis, a ray normally incident on a birefringent surface will be divided in two at the boundary. The extraordinary ray will be deviated; the ordinary ray will not. The ordinary ray index n, and the most extreme (whether greater or smaller) extraordinary ray index n_e, are together known as the principal indices of refraction of the material.

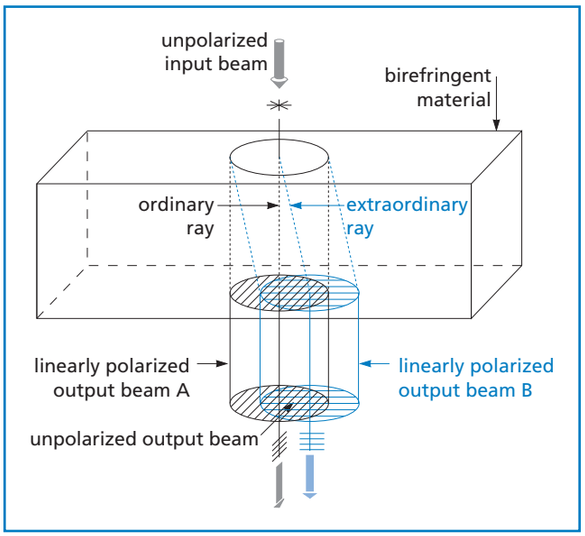

Figure 1.51 **Double refraction in a birefringent crystal**

If a beam of linearly polarized monochromatic light enters a birefringent crystal along a direction not parallel to the optical axis of the crystal, the beam will be divided into two separate beams. Each will be polarized at right angles to the other and will travel in different directions. The original beam energy, which will be divided between the new beams, depends on the original orientation of the vector to the crystal.

The energy ratio between the two orthogonally polarized beams can be any value. It is also possible that all energy will go into one of the new beams. If the crystal is cut as a plane-parallel plate, these beams will recombine upon emergence to form an elliptically polarized beam.

The difference between the ordinary and extraordinary ray may be used to create birefringent crystal polarization devices. In some cases, the difference in refractive index is used primarily to separate rays and eliminate one of the polarization planes, for example, in Glan-type polarizers. In other cases, such as Wollaston and Thompson beamsplitting prisms, changes in propagation direction are optimized to separate an incoming beam into two orthogonally polarized beams.

DICHROISM

Dichroism is selective absorption of one polarization plane over the other during transmission through a material. Sheet-type polarizers are manufactured with organic materials imbedded into a plastic sheet. The sheet is stretched, aligning molecules and causing them to be birefringent, and then dyed. The dye molecules selectively attach themselves to aligned polymer molecules, so that absorption is high in one plane and weak in the other. The transmitted beam is linearly polarized. Polarizers made of such material are very useful for low-power and visual applications. The usable field of view is large (up to grazing incidence), and diameters in excess of 100 mm are available.

POLARIZATION BY REFLECTION

When a beam of ordinary light is incident at the polarizing angle on a transmissive dielectric such as glass, the emerging refracted ray is partially linearly polarized. For a single surface (with n=1.50) at Brewster's angle, 100 percent of the light whose electric vector oscillates parallel to the plane of incidence is transmitted. Only 85 percent of the perpendicular light is transmitted (the other 15 percent is reflected). The degree of polarization from a single-surface reflection is small.

If a number of plates are stacked parallel and oriented at the polarizing angle, some vibrations perpendicular to the plane of incidence will be reflected at each surface, and all those parallel to it will be refracted. By making the number of plates within the stack large (more than 25), high degrees of linear polarization may be achieved. This polarization method

is utilized in CVI Melles Griot polarizing beamsplitter cubes which are coated with many layers of quarter-wave dielectric thin films on the interior prism angle. This beamsplitter separates an incident laser beam into two perpendicular and orthogonally polarized beams.

THIN METAL FILM POLARIZERS

Optical radiation incident on small, elongated metal particles will be preferentially absorbed when the polarization vector is aligned with the long axis of the particle. CVI Melles Griot infrared polarizers utilize this effect to make polarizers for the near-infrared. These polarizers are considerably more effective than dichroic polarizers.

Polarizing thin films are formed by using the patented Slocum process to deposit multiple layers of microscopic silver prolate spheroids onto a polished glass substrate. The exact dimensions of these spheroids determine the optical properties of the film. Peak absorption can be selected for any wavelength from 400 to 3000 nm by controlling the deposition process. Contrast ratios up to 10,000:1 can be achieved with this method. Other CVI Melles Griot high-contrast polarizers exhibit contrasts as high as 100,000:1.

CALCITE

Calcite, a rhombohedral crystalline form of calcium carbonate, is found in various forms such as limestone and marble. Since calcite is a naturally occurring material, imperfections are not unusual. The highest quality materials, those that exhibit no optical defects, are difficult to find and are more expensive than those with some defects. Applications for calcite components typically fall into laser applications or optical research. CVI Melles Griot offers calcite components in two quality grades to meet those various needs.

There are three main areas of importance in defining calcite quality.

Spectral Properties

Trace amounts of chemical impurities, as well as lattice defects, can cause calcite to be colored, which changes absorption. For visible light applications, it is essential to use colorless calcite. For near-infrared applications, material with a trace of yellow is acceptable. This yellow coloration results in a 15-percent to 20-percent decrease in transmission below 420 nm.

Wavefront Distortion (Striae)

Striae, or streaked fluctuations in the refractive index of calcite, are caused by dislocations in the crystal lattice. They can cause distortion of a light wavefront passing through the crystal. This is particularly troublesome for interferometric applications.

Scatter

Small inclusions within the calcite crystal account for the main source of scatter. They may appear as small cracks or bubbles. In general, scatter presents a significant problem only when the polarizer is being used with a laser. The amount of scatter centers that can be tolerated is partially determined by beam size and power.

CVI MELLES GRIOT CALCITE GRADES

CVI Melles Griot has selected the most applicable calcite qualities, grouped into two grades:

Laser Grade

Calcite with a wavefront deformation of $\lambda/4$ at 633 nm or better due to striae only.

Optical Grade

Calcite with a wavefront deformation of 1λ to $\lambda/4$ at 633 nm due to striae only.

Waveplates

Waveplates use birefringence to impart unequal phase shifts to the orthogonally polarized field components of an incident wave, causing the conversion of one polarization state into another.

There are two types of birefringence. With linear birefringence, the index of refraction (and hence the phase shift) differs for two orthogonally polarized linear polarization states. This is the operation mode of standard waveplates. With circular birefringence, the index of refraction and hence phase shift differs for left and right circularly polarized components. This is the operation mode of polarization rotators.

STANDARD WAVEPLATES: LINEAR BIREFRINGENCE

Suppose a waveplate made from a uniaxial material has light propagating perpendicular to the optic axis. This makes the field component parallel to the optic axis an extraordinary wave and the component perpendicular to the optic axis an ordinary wave. If the crystal is positive uniaxial, $\eta_e > \eta_o$, then the optic axis is called the slow axis, which is the case for crystal quartz. For negative uniaxial crystals $\eta_e < \eta_o$, the optic axis is called the fast axis.

The equation for the transmitted field E_2, in terms of the incident field E_1 is:

$$E_2 = \mathrm{s}(\mathrm{s} \bullet E_1)e^{i\phi_s} + \mathrm{f}(\mathrm{f} \bullet E_1)e^{i\phi_f} \qquad (1.102)$$

where s and f are unit vectors along the slow and fast axes. This equation shows explicity how the waveplate acts on the field. Reading from left to right, the waveplate takes the component of the input field along its slow axis and appends the slow axis phase shift to it. It does a similar operation to the fast component.

The slow and fast axis phase shifts are given by:

$$\phi_s = \eta_s(\omega)\omega t / c = 2\pi\eta_s(\lambda)t/\lambda \qquad (1.103)$$

$$\phi_f = \eta_f(\omega)\omega t / c = 2\pi\eta_f(\lambda)t/\lambda \qquad (1.104)$$

where η_s and η_f are, respectively, the indices of refraction along the slow and fast axes, and t is the thickness of the waveplate.

To further analyze the effect of a waveplate, we throw away a phase factor lost in measuring intensity, and assign the entire phase delay to the slow axis:

$$E_2 = \mathrm{s}(\mathrm{s} \bullet E_1)e^{i\phi} + \mathrm{f}(\mathrm{f} \bullet E1) \qquad (1.105)$$

$$\phi = \phi_s - \phi_f = 2\pi\left[\eta_s(\lambda) - \eta_f(\lambda)\right]t/\lambda \qquad (1.106)$$

$$\phi = 2\,\pi\,\Delta\eta(\lambda)t/\lambda \qquad (1.107)$$

In the above, $\Delta\eta(\lambda)$ is the birefringence $\eta_s(\lambda)$ - $\eta_f(\lambda)$. The dispersion of the birefringence is very important in waveplate design; a quarter waveplate at a given wavelength is never exactly a half waveplate at half that wavelength.

Let E_1 be initially polarized along x, and let the waveplate slow axis make an angle θ with the x axis. This orientation is shown in Figure 1.52.

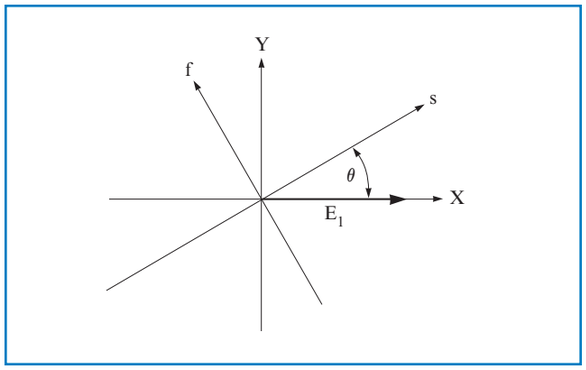

Figure 1.52. Orientation of the slow and fast axes of a waveplate with respect to an x-polarized input field

When the waveplate is placed between parallel and perpendicular polarizers the transmissions are given by:

$$T_{\parallel} \propto |E_2 x|^2 = 1 - \sin^2\ 2\theta \sin^2\phi/2 \qquad (1.108)$$

$$T_{\perp} \propto |E_2 y|^2 = \sin^2\ 2\theta \sin^2\phi/2 \qquad (1.109)$$

Note that θ is only a function of the waveplate orientation, and ϕ is only a function of the wavelength, the birefringence is a function of wavelength and the plate thickness.

For a full-wave waveplate:

$\phi = 2m\pi$, $T_{\parallel} = 1$, and $T_{\perp} = 0$, regardless of waveplate orientation.

For a half-wave waveplate:

$\phi = (2m + 1)\pi$, $T_{\parallel} = \cos^2 2\theta$, and $T_{\perp} = \sin^2 2\theta$. (1.110)

This transmission result is the same as if an initial linearly polarized wave were rotated through an angle of 2θ. Thus, a half-wave waveplate finds use as a polarization rotator.

For a quarter waveplate,

$\phi = (2m + 1)\pi/2$ (i.e., an odd multiple of $\pi/2$). (1.111)

To analyze this, we have to go back to the field equation. Assume that the slow and fast axis unit vectors s and f form a right handed coordinate system such that s x f = +z, the direction of propagation. To obtain circularly polarized light, linearly polarized light must be aligned midway between the slow and fast axes. There are four possibilities listed in the table below.

Phase Shift	Input Field Along $(\mathbf{s} + \mathbf{f})/\sqrt{2}$	Input Field Along $(\mathbf{s} - \mathbf{f})/\sqrt{2}$
$\phi = \pi/2 + 2m\pi$	RCP	LCP
$\phi = 3\pi/2 + 2m\pi$	LCP	RCP

Sometimes, waveplates described by the second line above are called 3/4 waveplates. For multiple order waveplates, CVI Melles Griot permits the use of either of the above classes of waveplates to satisfy the requirements of a quarter-wave waveplate.

MULTIPLE-ORDER WAVEPLATES

For the full-, half-, and quarter-wave waveplate examples given in standard waveplates, the order of the waveplate is given by the integer m. For m > 0, the waveplate is termed a multiple-order waveplate. For m = 0, we have a zero order waveplate.

The birefringence of crystal quartz near 500 nm is approximately 0.00925. Consider a 0.5-mm-thick crystal quartz waveplate. A simple calculation shows that this is useful as a quarter waveplate for 500 nm; in fact, it is a 37 $\lambda/4$ waveplate at 500 nm with m = 18. Multiple-order waveplates are inexpensive, high-damage-threshold retarders. Further analysis shows that this same 0.5mm plate is a 19 $\lambda/2$ half waveplate at 488.2 nm and a 10λ full-wave waveplate at 466.5 nm. The transmission of this plate between parallel polarizers is shown in Figure 1.53 as a function of wavelength. The retardance of the plate at various key points is also shown. Note how quickly the retardance changes with wavelength. Because of this, multiple-order waveplates are generally useful only at their design wavelength.

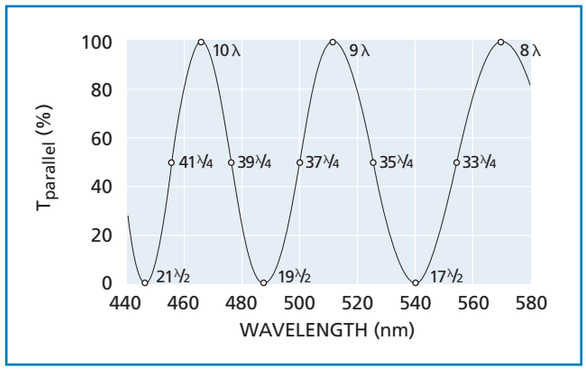

Figure 1.53 **Transmission of a 0.5-mm-thick crystal quartz waveplate between parallel polarizers**

ZERO-ORDER WAVEPLATES

As discussed above, multiple-order waveplates are not useful with tunable or broad bandwidth sources (e.g., femtosecond lasers). A zero-order waveplate can greatly improve the useful bandwidth in a compact, high-damage-threshold device.

As an example, consider the design of a broadband half-wave waveplate centered at 800 nm. Maximum tuning range is obtained if the plate has a single π phase shift at 800 nm. If made from a single plate of crystal quartz, the waveplate would be about 45 μm thick, which is too thin for easy fabrication and handling. The solution is to take two crystal quartz plates differing in thickness by 45 μm and align them with the slow axis of one against the fast axis of the other. The net phase shift of this zero-order waveplate is π. The two plates may be either air-spaced or optically contacted. The transmission of an 800-nm zero-order half-wave waveplate between parallel polarizers is shown in Figure 1.54 using a 0-10% scale. Its extinction is better than 100:1 over a bandwidth of about 95 nm centered at 800 nm.

CVI Melles Griot produces multiple order and zero order crystal quartz waveplates at any wavelength between 193 nm and 2100 nm. Virtually all popular laser wavelengths are kept in stock, and custom wavelength parts are available with short delivery time.

Mica waveplates are an inexpensive zero-order waveplate solution for low-power applications and in detection schemes.

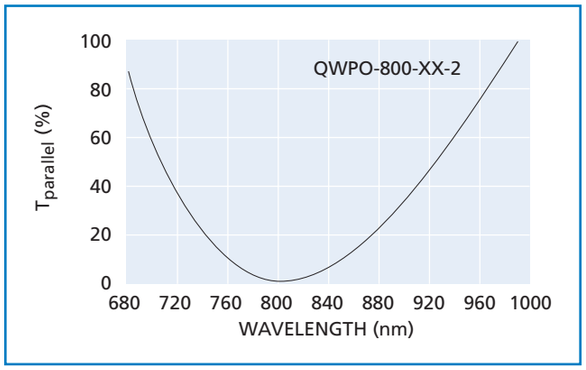

Figure 1.54 **Zero-order crystal quartz half-wave waveplate for 800 nm**

POLYMER WAVEPLATES

Polymer waveplates offer excellent angular field of view since they are true zero-order waveplates. Figure 1.55 compares the change in retardance as function of incidence angle for polymer and quartz waveplates. A polymer waveplate changes by less than 1% over a $\pm 10°$ incidence angle.

Retardance accuracy with wavelength change is often of key concern. For example, an off-the-shelf diode laser has a center wavelength tolerance of ± 10 nm. Changes with temperature and drive conditions cause wavelength shifts which may alter performance. These polymer waveplates maintain excellent waveplate performance even with minor shifts in the source wavelength. The temperature sensitivity of laminated polymer waveplates is about 0.15 nm/°C, allowing operation over moderate temperature ranges without significantly degrading retardance accuracy. A comparison of different waveplate types and their dependence on wavelength is shown in figure 1.56.

1ch_FundamentalOptics_Final_a.qxd 7/6/2009 2:02 PM Page 1.48

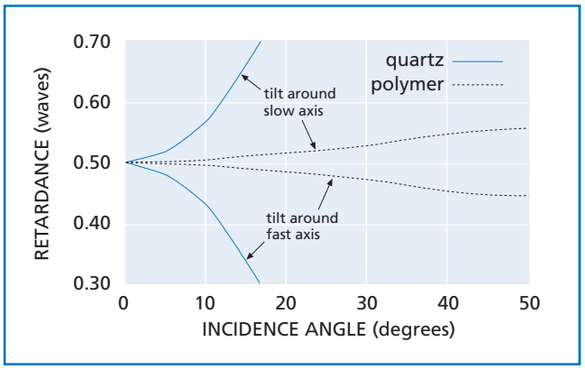

Figure 1.55 **Retardance vs incidence angle for quartz and polymer waveplates**

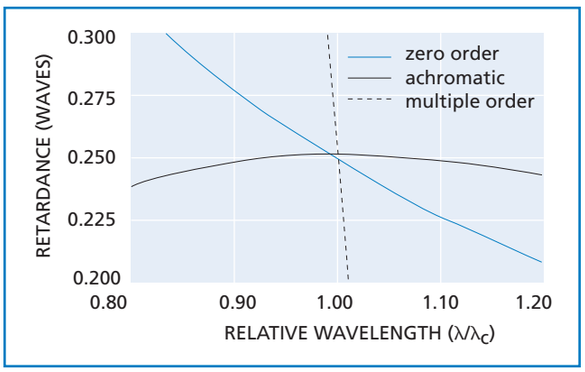

Figure 1.56 **Wavelength performance of common quarter wave retarders.**

ACHROMATIC WAVEPLATES

At 500 nm, a crystal quartz zero-order half-wave waveplate has a retardation tolerance of λ/50 over a bandwidth of about 50 nm. This increases to about 100 nm at a center wavelength of 800 nm. A different design which corrects for dispersion differences over the wavelength range is required for bandwidths up to 300 nm.

If two different materials are used to create a zero-order or low-order waveplate, cancellation can occur between the dispersions of the two materials. Thus, the net birefringent phase shift can be held constant over a much wider range than in waveplates made from one material. ACWP-series achromatic waveplates from CVI Melles Griot (see Figure 1.57) are comprised of crystal quartz and magnesium fluoride to achieve achromatic performance.

Three wavelength ranges are available in both quarter and half wave retardances. Retardation tolerance is better than λ/100 over the entire wavelength range.

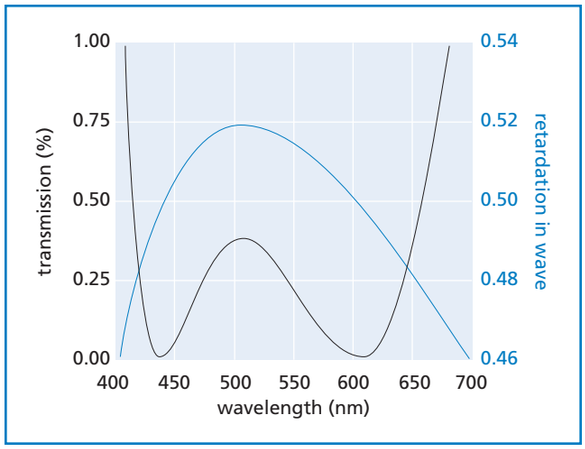

Figure 1.57 **ACWP-400-700-10-2**

For quarter-wave waveplates, perfect retardance is a multiple of 0.25 waves, and transmission through a linear polarizer must be between 33% and 67%. (In all but the shortest wavelength design, quarter-wave retardation tolerance is better than λ/100.) For half-wave waveplates, perfect retardance is 0.5 waves, while perfect transmission through a linear polarizer parallel to the initial polarization state should be zero. A high degree of achromatization is achievable by the dual material design. In addition, we manage low group velocity dispersion for ultrashort pulse applications through the use of thin plates.

DUAL-WAVELENGTH WAVEPLATES

Dual-wavelength waveplates are used in a number of applications. One common application is separation of different wavelengths with a polarizing beamsplitter by rotating the polarization of one wavelength by 90°, and leaving the other unchanged. This frequently occurs in nonlinear doubling or tripling laser sources such as Nd:YAG (1064/532/355/266).

One way to achieve the multiple retardation specifications is through careful selection of multiple-order waveplates which meet both wavelength and retardation conditions. This often results in the selection of a relatively high order waveplate. Therefore, these dual-wavelength waveplates operate best over a narrow bandwidth and temperature range.

Another approach is to combine two quartz waveplates with their optical axes orthogonal to one another, effectively creating a zero-order waveplate. In this configuration, the temperature dependence is a function of the thickness difference between the waveplates, resulting in excellent temperature stability. The retardation of the compound waveplate is also a function of the thickness difference enabling wide bandwidth performance.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:31 PM Page 1.49

Etalons

Etalons are most commonly used as line-narrowing elements in narrow-band laser cavities or as bandwidth-limiting and coarse-tuning elements in broadband and picosecond lasers. Further applications are laser line profile monitoring, diagnosis.

The etalons described in this section are all of the planar Fabry-Perot type. Typical transmission characteristics for this type etalon are shown in Figure 1.58.

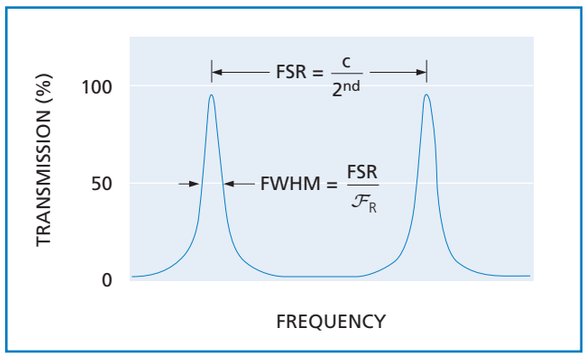

Figure 1.58 **Transmission characteristics of a Fabry-Perot etalon**

Air-Spaced Etalons consist of pairs of very flat plano-plano plates separated by optically contacted spacers. The inner surfaces of the plates are coated with partially reflecting coatings, the outer surfaces are coated with antireflection coatings.

Solid Etalons are made from a single plate with parallel sides. Partially reflecting coatings are then deposited on both sides. The cavity is formed by the plate thickness between the coatings.

Deposited Solid Etalons are a special type of solid etalon in which the cavity is formed by a deposited layer of coating material. The thickness of this deposited layer depends on the free spectral range required and can range from a few nanometers up to 15 micrometers. The cavity is sandwiched between the etalon reflector coatings and the whole assembly is supported on a fused-silica base plate.

Etalon plates need excellent surface flatness and plate parallelism. To avoid peak transmission losses due to scatter or absorption, the optical coatings also have to meet the highest standards.

For a plane wave incident on the etalon, the transmission of the etalon is given by:

$$T = \frac{I_{\text{trans}}}{I_{\text{inc}}} = \frac{1}{1 + \frac{4R}{(1-R)^2}\sin^2(\delta/2)} \qquad (1.112)$$

Here, R is the reflectance of each surface; δ is the phase shift

$$\delta = \frac{2\pi}{\lambda}\eta d\cos\theta \qquad (1.113)$$

where,

η is the refractive index (e.g., 1 for air-spaced etalons)

d is the etalon spacing or thickness

θ is the angle of incidence

The free spectral range (FSR) of the etalon is given by

$$\text{FSR} = \frac{c}{2\eta d} \text{ in Hz} \qquad (1.114)$$

$$= \frac{1}{2\eta d} \text{ in cm}^{-1}$$

$$= \frac{\lambda^2}{2\eta d} \text{ in nm}$$

The reflectivity finesse, $\mathcal{F}_R$ is given by

$$\mathcal{F}_R = \frac{\pi\sqrt{R}}{1-R} \qquad (1.115)$$

Figure 1.59 shows the reflectivity finesse as a function of the coating reflectivity.

The bandwidth (FWHM) is given by

$$\text{FWHM} = \frac{\text{FSR}}{\mathcal{F}_R} \qquad (1.116)$$

However, the above applies to theoretical etalons which are assumed to be perfect. In reality, even the best etalon will show defects that limit theoretically expected performance. Therefore, in a real etalon, the actual finesse will usually be lower than the reflectivity finesse.

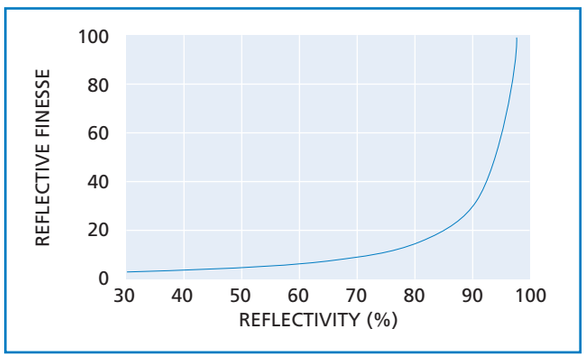

Figure 1.59 **Reflectivity finesse vs. coating reflectance of each surface**

$$\mathcal{F}_R = \frac{\pi\sqrt{R}}{1-R} \tag{1.117}$$

$$\mathcal{F}_S = \frac{M}{2}\frac{\lambda}{633\ \text{nm}} \tag{1.118}$$

$$\mathcal{F}_\theta = \frac{\lambda}{d\tan^2\theta} \tag{1.119}$$

$$\mathcal{F}_D = \frac{CA^2}{2d\lambda} \tag{1.120}$$

where $\mathcal{F}_R$ is the reflectivity finesse, $\mathcal{F}_s$ is the plate spherical deviation finesse coefficient, $\mathcal{F}_\theta$ is the incident beam divergence finesse coefficient, and $\mathcal{F}_d$ is the diffraction-limited finesse coefficient.

The defects that contribute to this reduction are as shown in Figure 1.60 (graphical representations are exagerated for clarification).

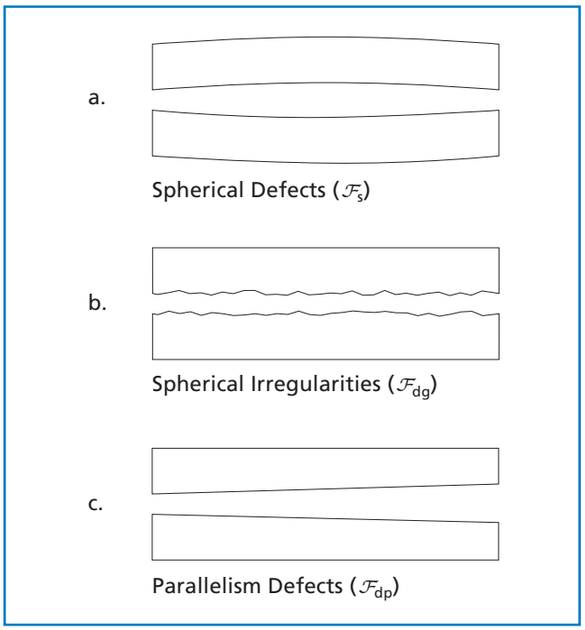

Figure 1.60 **Three types of defects contributing to the total defect finesse**

All three types of defects contribute to the total defect finesse $\mathcal{F}_d$:

$$\frac{1}{\mathcal{F}_d^2} = \frac{1}{\mathcal{F}_S^2} + \frac{1}{\mathcal{F}_{dg}^2} + \frac{1}{\mathcal{F}_{dp}^2} \tag{1.121}$$

The beam divergence also influences the actual finesse of an etalon.

Taking into account all these contributions, the effective finesse ($\mathcal{F}_e$) of an etalon (with $\mathcal{F}_R$ being the reflectivity finesse and $\mathcal{F}_\theta$ the divergence finesse) is:

$$\frac{1}{\mathcal{F}_e} = \sqrt{\frac{1}{\mathcal{F}_R^2} + \frac{1}{\mathcal{F}_D^2} + \frac{1}{\mathcal{F}_\theta^2} + \frac{1}{\mathcal{F}_S^2}} \tag{1.122}$$

The effective finesse a user sees when using the etalon depends not only on the absolute clear aperture, but also on the used aperture of the etalon, especially when a high finesse is required.

The examples below show how the effective finesse varies with plate flatness and clear aperture.

Example 1: Air-spaced etalon,

$R = 95\%\ (\pm 1\%)$ at 633 nm

CA = 25 mm

Used aperture = 20 mm

Spacer (air gap) = 1 mm

Spherical / parallelism defects = $<\lambda/20$

Plate rms = 0.80 nm

Beam divergence = 0.1 mRad

$\mathcal{F}_R = 61$, $\mathcal{F}_e = 10$

Example 2: Air-spaced etalon,

Same parameters as example 1 except:

Used aperture = 5 mm

$\mathcal{F}_R = 61$, $\mathcal{F}_e = 40\ (\pm 4)$

Example 3: Air-spaced etalon,

Same parameters as example 1 except:

Spherical / parallelism defects = $\lambda/100$

Plate rms = 0.40 nm

Beam divergence: 0.1mRad

$\mathcal{F}_R = 61$, $\mathcal{F}_e = 40\ (\pm 8)$

These examples illustrate that, for large-aperture applications, it is important to use very high-quality plates to ensure a high finesse and good transmission values.

TUNING AN ETALON

Etalons can be tuned over a limited range to alter their peak transmission wavelengths. These techniques are:

Angle tuning or tilting the etalon: As the angle of incidence is increased, the center wavelength of the etalon can be tuned down the spectrum.

Temperature tuning: Primarily used for solid etalons, temperature-tuning changes both the actual spacing of the reflective surfaces via expansion and the index of refraction of the material, which changes the optical spacing. The tuning result can be given by

$$\frac{\partial(\mathrm{FSR})}{\partial T} = -(\mathrm{FSR})\left[\frac{1}{\mathrm{n}}\frac{\partial \eta}{\partial T} + \frac{1}{d}\frac{\partial d}{\partial T}\right] \qquad (1.123)$$

Pressure tuning: Air-spaced etalons can be tuned by increasing the pressure in the cavity between the optics, thereby increasing the effective index of refraction, and thus the effective spacing.

The above examples illustrate how critical the optical surface flatness, plate parallelism and surface quality are to the overall performance of an etalon. At CVI Melles Griot we have developed sophisticated software that allows us to simulate all effects that influence the performance of an etalon. To order an etalon, FSR, finesse and used aperture are required.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:31 PM Page 1.52

Ultrafast Theory

The distinguishing aspect of femtosecond laser optics design is the need to control the phase characteristic of the optical system over the requisite wide pulse bandwidth. CVI Melles Griot has made an intensive theoretical study of these effects. Certain coating designs have been modified with control of the phase characteristics in mind. New proprietary designs have been created with desirable characteristics for femtosecond researchers. All optics in this section have been tested by researchers in the field and we are constantly fielding new requests.

Assume that the power, reflectivity, and polarization characteristics of a laser mirror are acceptable over the bandwidth of a femtosecond pulse. This means that, over the entire pulse bandwidth, a cavity mirror may have a reflectivity greater than 99.8%; a 50% beamsplitter may have a fairly constant reflection; a polarizer may maintain its rejection of one polarization with an acceptable transmission of the other. It is not enough, however, to simply preserve the power spectrum $S(\omega) = |E(\omega)|^2$ when dealing with femtosecond pulses. The phase relationship among the Fourier components of the pulse must also be preserved in order that the pulse not be broadened or distorted. What constraint on the performance of a mirror or transmissive optic does this imply?

Consider a general initial pulse shape $E_0(t)$. As a function of its Fourier components, it may be expressed as:

$$E_0(t) = \int E(\omega)^{-i\omega t}\, d\omega \qquad (1.124)$$

Suppose this pulse reflects off of a mirror. For this example, we assume the mirror is "ideal", and use the Fourier transform of its complex amplitude reflectance:

$$r(\omega) = re^{+i\Phi(\omega)} = re^{+i\omega t_d} \qquad (1.125)$$

For this "ideal" mirror, r is a real constant equal to the amplitude reflectivity that is assumed constant over the pulse bandwidth. All phase effects have been assumed to be describable by a single phase shift $\phi(\omega)$ that is linearly proportional to frequency with proportionality constant t_d. The reflected pulse is then:

$$\begin{aligned} E_r(t) &= r\int E(\omega)^{-i\omega(t-t_d)}\, d\omega \\ &= rE(t-t_d) \end{aligned} \qquad (1.126)$$

Thus, provided the phase shift is linear in frequency over the pulse bandwidth, the reflected pulse is scaled by the amplitude reflectance r, and delayed in time by the constant group delay t_d. It is, otherwise, an undistorted replica of the original pulse.

Examined over a large enough bandwidth, no optical system will exhibit the constant group delay over frequency needed for perfect fidelity. In general, the phase shift near some center frequency ω_0 may be expanded in a Taylor series for frequencies near ω_0:

$$\begin{aligned} \Phi(\omega) = \Phi(\omega_0) + \Phi'(\omega_0)(\omega-\omega_0) + \\ \Phi''(\omega_0)(\omega-\omega_0)^2/2! + \\ \Phi'''(\omega_0)(\omega-\omega_0)^3/3! \ldots \end{aligned} \qquad (1.127)$$

These derivatives are, respectively, the group delay $\Phi'(\omega_0)$, the group velocity dispersion $\Phi''(\omega_0)$, and the cubic term $\Phi'''(\omega_0)$, evaluated at a center frequency ω_0. This expansion is heuristically useful, in an exactly soluble model, for the propagation of a transform-limited Gaussian pulse. Note, however, that for extremely short pulses the expansion above may be insufficient. A full numerical calculation may have to be performed using the actual phase shift function $\Phi(\omega)$. CVI Melles Griot will be happy to assist those interested in the modeling of real optical elements.

To illustrate pulse distortion due to the dependence of the group delay on frequency, consider what happens when an unchirped, transform-limited Gaussian pulse passes through a medium, or is incident on a mirror whose dominating contribution to phase distortion is non-zero group velocity dispersion. The field envelope of the pulse is assumed to be of the form:

$$E(t) = \exp\left[-\left(2\ \ln 2 t^2/\tau_0^2\right)\right] \qquad (1.128)$$

where τ_0 is the initial pulse duration (FWHM of the pulse intensity). Let the pulse enter a medium or reflect off of a mirror with non-zero $\Phi''(\omega)$, measured in fsec^2 radians. (For a continuous medium-like glass, $\Phi''(\omega) = \beta''(\omega) \times z$ where $\beta''(\omega)$ is the group velocity dispersion (GVD) per centimeter of material, and z is the physical path length, in centimeters, traveled through the material.) The Gaussian pulse will be both chirped and temporally broadened by its encounter with group velocity dispersion. The power envelope will remain Gaussian; the result for the broadened FWHM is:

$$\tau_1 = \tau_0\left[1+\left(4\ln 2\,\Phi''(\omega)/\tau_0^2\right)^2\right]^{1/2} \qquad (1.129)$$

This result, valid only for initially unchirped, transform-limited Gaussian pulses, is nevertheless an excellent model to study the effects of dispersion on pulse propagation. The graphs shown in Figure 1.61 represent the theoretical broadening from dispersion for initial pulse widths ranging from 10 to 100 femtoseconds.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:31 PM Page 1.53

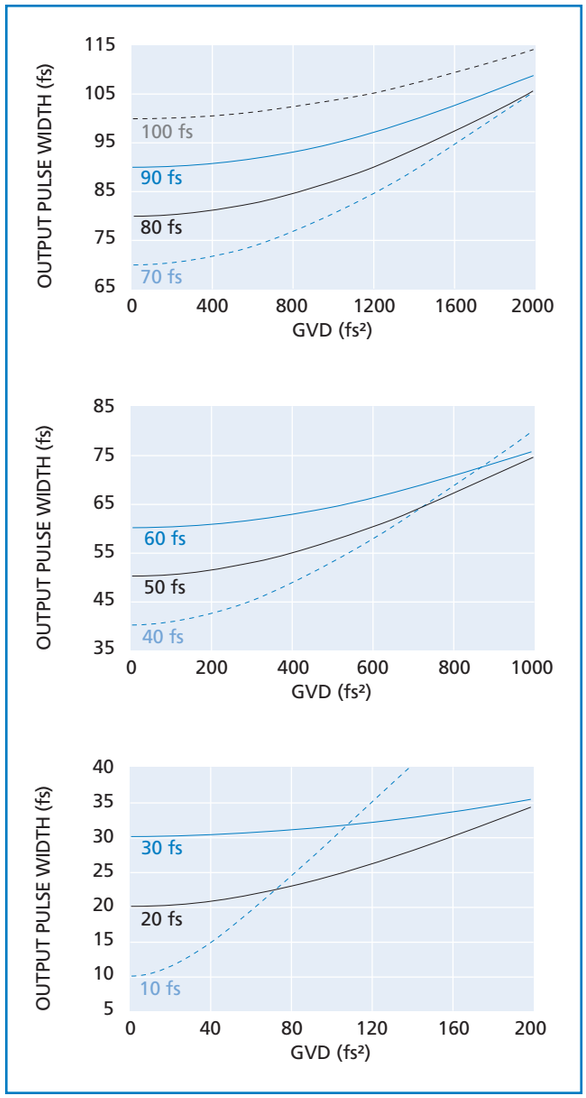

Figure 1.61 **Output pulse width vs. GVD**

GROUP-VELOCITY & CUBIC DISPERSION FOR VARIOUS OPTICAL MATERIALS

Figures 1.62 and 1.63 show the GVD and cubic dispersion respectively for some common used glasses. Some of the glasses can be used in the UV region. They should be useful in estimating material dispersion and pulse distortion effects. Please check these calculations independently before using them in a final design.

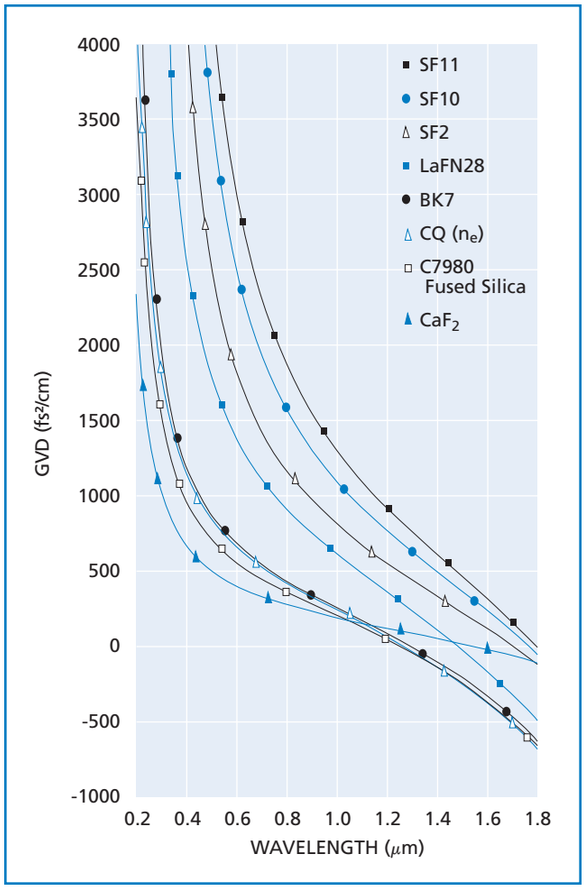

Figure 1.62 **GVD for common glasses**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:31 PM Page 1.54

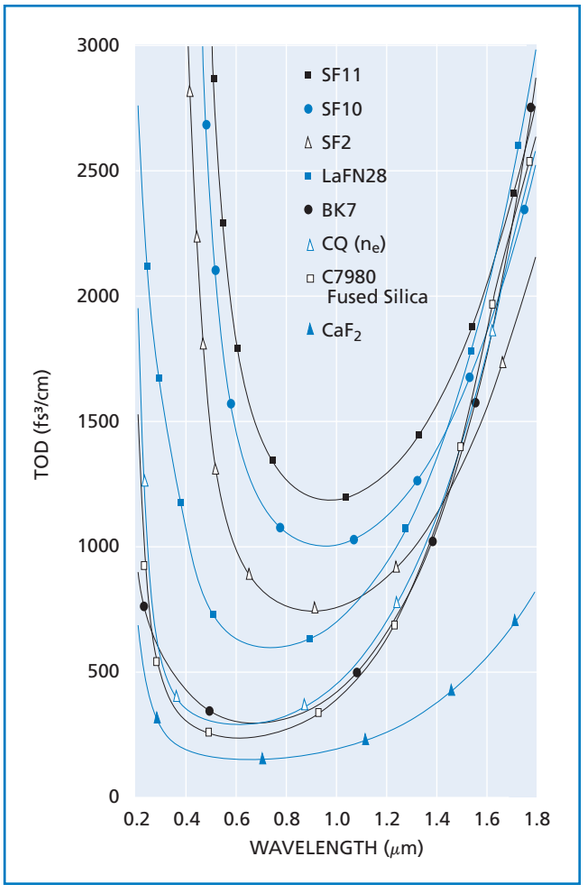

Figure 1.63 **Cubic dispersion for common glasses**

DISPERSIVE PROPERTIES OF MIRRORS

CVI Melles Griot uses three basic designs; TLM1 mirrors for energy fluence greater than 100 mJ/cm^2, TLM2 mirrors for cw oscillators and low-fluence pulses, and TLMB mirrors which are a hybrid of the two. The reflectivity, GVD parameter, and cubic dispersion parameter for TLM2 high reflectors are shown in Figure 1.64. In these examples, the mirrors are centered at 800 nm and are designed for use at normal incidence and at 45 degrees. Note that, at the design wavelength, (a) GVD is zero, (b) the cubic term is minimized, and (c) at 45° incidence, the GVD of the p-polarization component is very sensitive to wavelength, while the GVD for s-polarization component is nearly zero over a broad wavelength range. Thus one should avoid using mirrors at 45° incidence with the p-polarization. On the other hand, at 45° incidence, s-polarization provides very broad bandwidth and minimizes pulse distortion problems and should be used whenever possible.

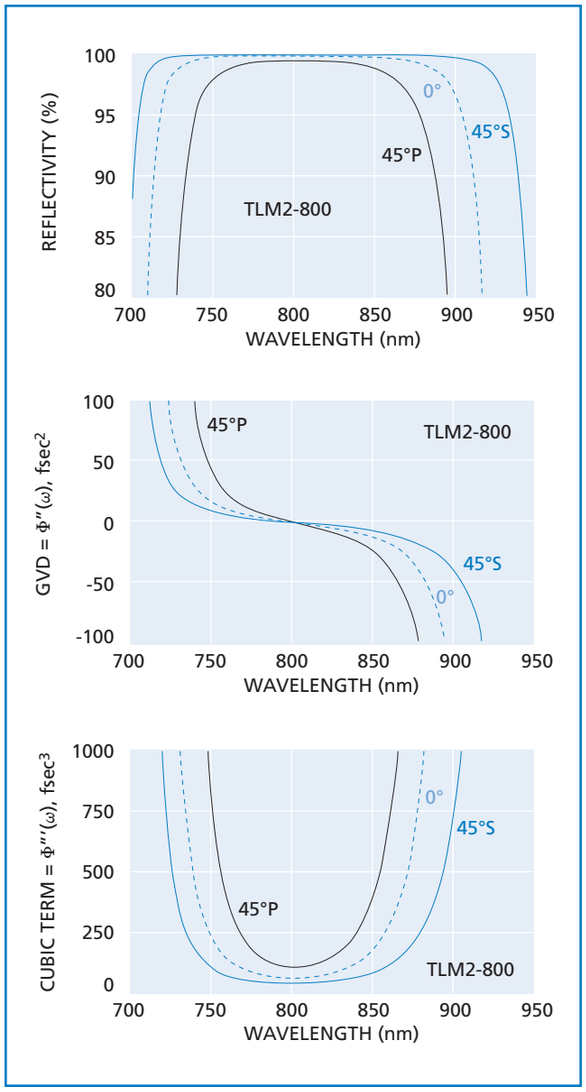

Figure 1.64 **Dispersion and reflectivity for mirrors TLM2-800-0 and TLM2-800-45**

Ti:Sapphire and other femtosecond laser systems need prismless compensation of the built-in positive chirp encountered in the laser optical circuit. This becomes mandatory in industrial and biomedical applications where the laser must provide a compact, stable, and reliable solution.

Negative Group Velocity Dispersion Mirrors (TNM2) meet these needs with off-the-shelf availability, and can be employed both intracavity and extracavity to satisfy chirp control requirements.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:32 PM Page 1.55

In experiments using CVI Melles Griot TNM2 negative group velocity dispersion mirrors, 200-mW, 80-fsec pulses centered at 785 nm were achieved in a simple, prismless, Ti:Sapphire oscillator. The configuration is shown in Figure 1.65.

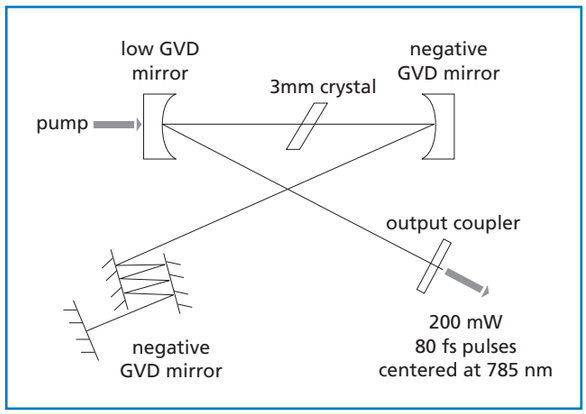

Figure 1.65 **Typical optical setup incorporating low GVD and Negative GVD mirrors in an ultrafast application**

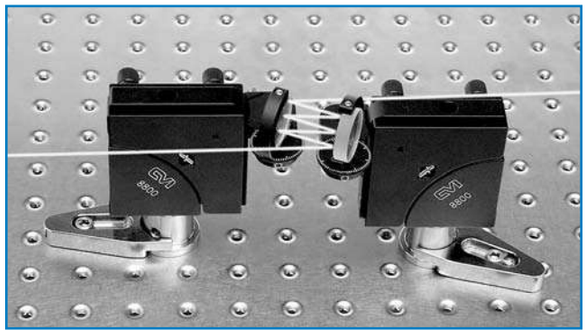

Figure 1.66 **Typical optical set-up of negative GVD mirrors**

OUTPUT COUPLERS AND BEAMSPLITTERS

Output-coupler partial reflectors and beamsplitters behave similarly; however, here is an additional consideration in their analysis. The behavior of the transmitted phase of the coating and the effect of material dispersion within the substrate on the transmitted beam have to be taken into account in a detailed analysis. In general, the coating transmitted phase has similar properties and magnitudes of GVD and cubic to the reflected phase. As usual, centering is important. As a beamsplitter, we recommend the 1.5 mm thick fused silica substrate PW-1006-UV. As an output coupler substrate, we recommend the 3.0 mm thick, 30 minute wedge fused silica substrate IF-1012-UV.

CVI Melles Griot has developed the TFPK Series Broadband Low Dispersion Polarizing Beamsplitters to satisfy requirements for very-high-power, short-pulse lasers. These optics are ideal for intracavity use in femtosecond regenerative amplifiers. The main emphasis is on linear phase characteristics. See Chapter 9 of Lasers, A. E. Siegman (University Science Books, Mill Valley, California, 1986), for a good discussion of linear pulse propagation.

In chirped pulse regenerative amplification, the pulse may have to pass through one or two polarizers twice per round trip. There can be 10 to 20 round trips before the gain is saturated and the pulse is ejected. At this stage the pulse is long (100–1000 psec); however the phase shift at each frequency must still be maintained to minimize the recompressed pulse width. The many round trips of the pulse in the regenerative amplifier put stringent requirements on the phase characteristics of the coatings.

Figure 1.67 shows the power transmission curves for both *s*- and *p*-polarization and the transmitted phase characteristics of the *p* component for a TFPK optimized at 800 nm. (Users may specify any wavelength

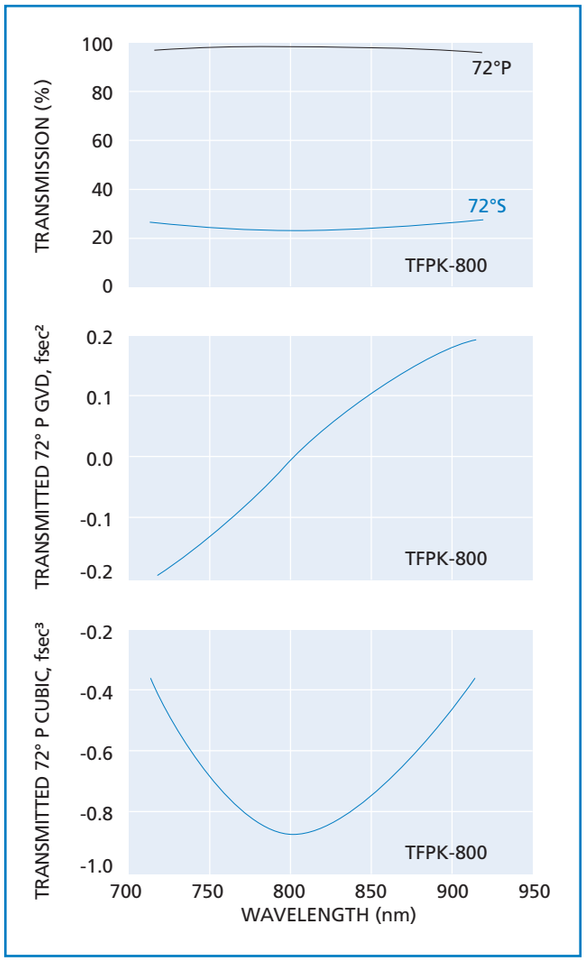

Figure 1.67 **Properties for one coated side of a TFPK polarizing beamsplitter optimized for 800 nm. Both sides are coated for these properties.**

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:32 PM Page 1.56

from 250 nm to 1550 nm.) The phase characteristics shown are the GVD and the cubic phase term. Not shown are the reflected phase characteristics for *s*-polarization; they are similar to the *p*-polarization transmission curves, and have the same low nonlinearity and broad bandwidth. Note that both sides of the optic have the coating whose properties are described in Figure 1.67. Therefore, the *s*- and *p*-polarization transmissions per surface should be squared in determining the specifications. The phase characteristics show that in all modes of operation, the TFPK polarizer performance is dominated by the substrate.

There are some subtleties associated with the TFPK. The near 72° angle has to be set properly and optimized. Some thought has to be given to mechanical clearances of the laser beam at such a steep incidence angle. The reflectivity for *s*-polarization is limited to 75%. Variant designs can increase this at a slight loss in bandwidth, increase in incidence angle, and increase in insertion loss for the transmitted *p*-polarized component.

The FABS autocorrelator beamsplitters from CVI Melles Griot are broadband, 50% all-dielectric beamsplitters. They are useful in many types of pump-probe experiments and in the construction of antiresonant ring configurations. They are essentially lossless and extremely durable. Both have advantages over partially reflecting metal coatings.

Power transmission curves for the *s*- and *p*-polarized versions of the FABS, along with the corresponding reflected phase characteristics for beamsplitters optimized at 800 nm, are shown in Figures 1.68 and 1.69. The linear pulse propagation properties of these beamsplitters are dominated by the substrate material dispersion. As with virtually all dielectric coated optics, the *s*-polarized version is broader than *p*-polarized version. CVI Melles Griot can produce FABS in other than 50:50 with excellent phase characteristics.

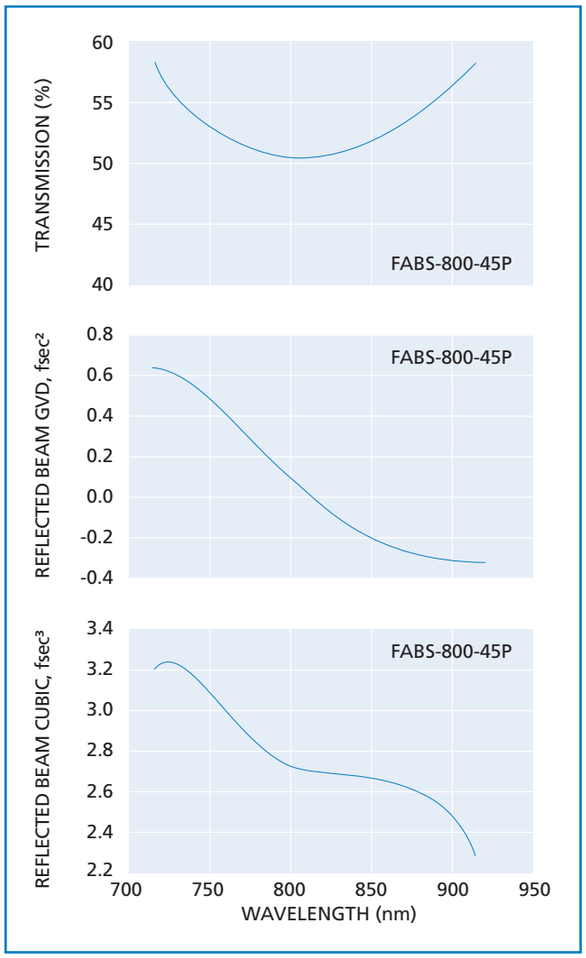

Figure 1.68 **Transmission characteristics for FABS series polarizers with *p*-polarized light**

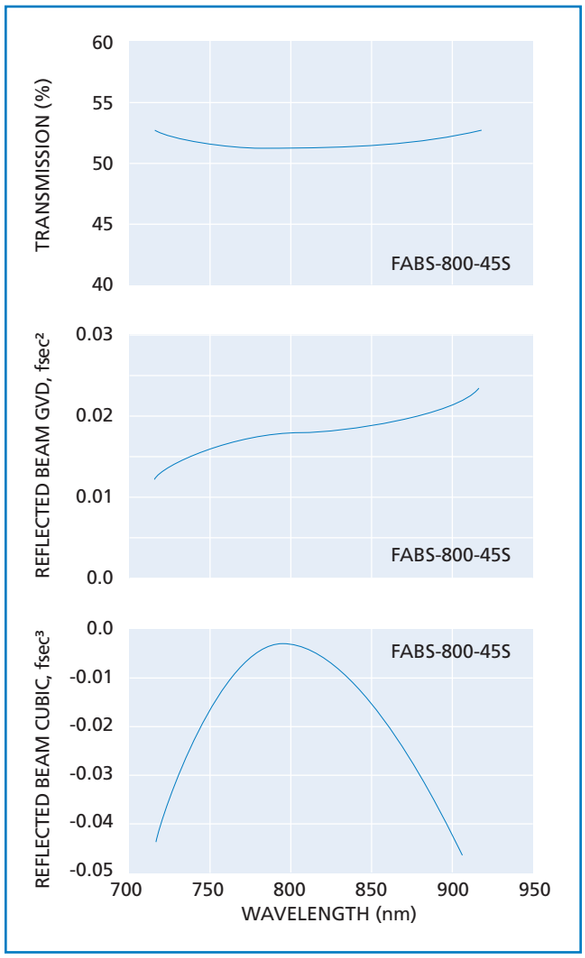

Figure 1.69 **Transmission characteristics for FABS series polarizers with *s*-polarized light**

ANTIREFLECTION COATINGS

All CVI Melles Griot antireflection coating designs work well in femtosecond operation as the forward-going phasor is the dominant contribution to the phase shift; the AR coating is very thin and simply "fixes" the small Fresnel reflection of the substrate.

PRISMS

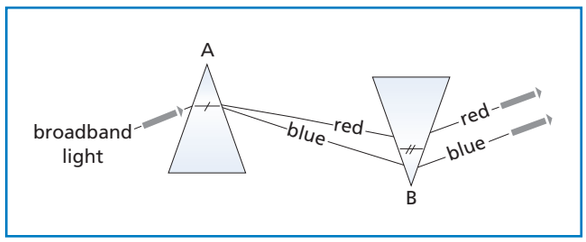

Figure 1.70 **Brewster prism**

Very-high-quality isosceles Brewster's angle prisms for intra and extracavity use are available from CVI Melles Griot. The design of these prisms satisfies the condition of minimum loss due to entrance and exit at Brewster's angle. To calculate GVD at Brewsters angle, refer to Figure 1.70 and use the following equation:

$$\mathrm{GVD} = \frac{d^2}{d\omega^2}\psi\bigg|_{\omega-\omega_l} \approx \frac{\lambda_l^3}{2\pi c^2}\left[\mathrm{L}\,\frac{d^2\eta}{d\lambda^2}\bigg|_{\lambda_l} - 4l\left(\frac{d\eta}{d\lambda}\bigg|_{\lambda_l}\right)^2\right] \qquad (1.130)$$

where

η = refractive index of the prisms (assuming the same material)
l = tip to tip distance ($\overline{AB}$)
L = total avg. glass path
ψ = spectral phase of the electric field
$\omega_l \lambda_l = 2\pi_c$ (assumes Brewster prism at minimum deviation).

For more on the Ultrafast phenomena, see J.C. Diels and W. Rudolph, Ultrashort Laser Pulse Phenomena, Academic Press, 1996.

1ch_FundamentalOptics_Final_a.qxd 6/15/2009 2:32 PM Page 1.58

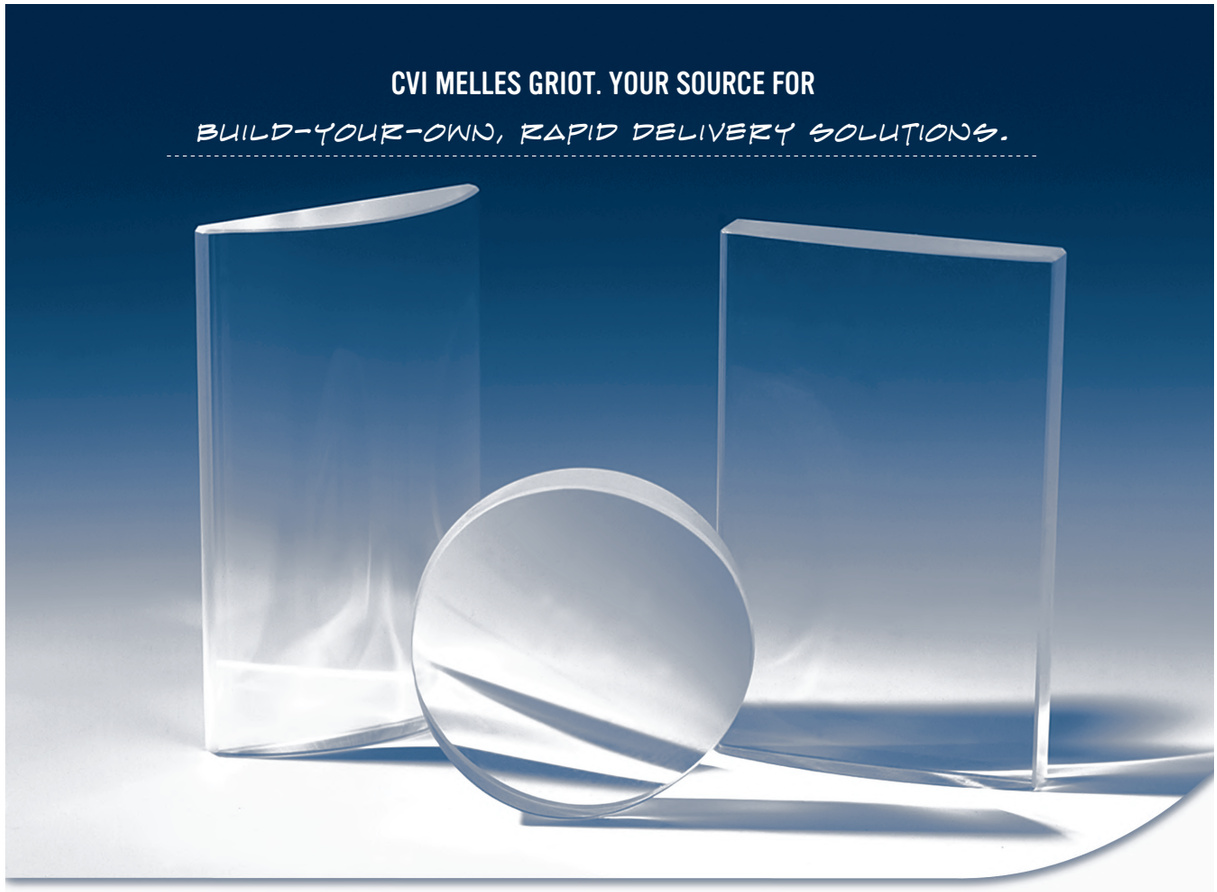

CVI Melles Griot offers two-week delivery of optics at catalog prices.

- 4000 uncoated substrates
- 2500 coating options
- A wide variety of in-stock substrates available (curves, flats, prisms, waveplates)
- Coatings from 193 nm to 2300 nm
- Your order ships in two weeks or less

How can we help you make your project a success?

Lasers | Lenses | Mirrors | Assemblies | Windows | Shutters | Waveplates | Mounts

WELDING SECRETS

SECOND EDITION

by Hal Wilson

WELDING SECRETS

Second Edition

A welding guide for the self taught welder,
as well as the more experienced welder.

by

HAL WILSON

© Copyright, 1990 by the Flyco Machine Company

All rights reserved. No part of this book may be reproduced by any means without the written permission of the Flyco Machine Company.

Published by

FLYCO MACHINE COMPANY
3712 Ringgold Road #155
Chattanooga, TN 37412-1638

Web Site:
http://www.entermall.com/weldingsecrets

Printed in the United States of America

Table of Contents

How To Remove A Bad Bearing Race Or Cup From A Hole 1
How To Build Up A Worn Shaft 1
How To Burn A Nut Off Of A Bolt And Save The Threads On The Bolt 3
How To Expand A Nut 3
How To Shrink A Nut 3
Camouflage 4
How To Remove A Broken-Off Bolt 4
Case Hardening 5
How To Square A Frame By Welding Or Peening 5
How To Make A Warped Steel Frame Lie Flat 6
How To Weld A Screen Wire To A Steel Frame 8
How To Make A Smooth Cut With A Cutting Torch 9
How To Do Overhead Welding 9
How To Weld Thick To Thin 9
How To Make A Smooth Bead 10
Vertical Weld 10
How To Flux Lead 10
Bulge 10
How To Burn A Weld And Save Both Pieces 11
Big John 12
Notch Effect 13
How To Remove A Broken-Off Tap 13
Built In Stresses 14
Crystals 14
How To Burn Through A Large Steel Shaft 15
My First Commercial Weld 15
Cast Iron Welded With Nickel Electrodes 16
Straighten 6 Inch Strip 16
Clevis On Channel Iron 17
How To Temper A Chisel 18
How To Make A Circle Lie Flat 18
How To Remove A Stuck Sleeve From A Hole 19
How To Shrink A Steel Pulley 19
Reinforce Inside Of Tubing Or Pipe 19
How To Repair A Crack In A Piece Of Tubing 20
Weld Inserted Pipe 21
Another Method For Welding Pipe 22
Build Up Inside Of Tubing 23

Test Welds 24
Flying Coaster 25
Roller Coaster 27
Cracked Center 28
Peening Test 28
Cross Brace 29
Test Welds On Low Carbon Steel 30
"Direction Of Travel" and "Back-Stepping" 33
Why A Weld Bead Shrinks More Where It Stops Than It Does Where It Starts 34
Confined Expansion 34
The Exception To The Rule 35
Magnaflux Testing For Cracks 36
Water Tank 36
Back Stepping 37
Broken Trailer Frame 37
Air Lifts 37
Subjects Of Non-Welding 39
Cover A Padded Bar With Upholstery 40
Hooks Or Eyes On Spring 41
Make A Coil Spring 42
Cooling Fin 42
Shaft Will Break At Shoulder 43
How To Temper Small Parts 43
Rod Ends 44
Find Diameter Of Numbered Machine Screws 44
Which Way To Turn A Turnbuckle 45
Dish In Wheel 45
Blowing Grip On Bar 45
How To Find A Tap Drill Size Without A Chart 46
Finding Half Of A Mixed Common Fraction 47
Simple Hardness Tester 49
Fusibility: The Ease Of Melting 50
Heat Of Fusion Of Metals 51
Thermal Conductivity Of Metals 52
Thermal Expansion 53
Typical Ferrous Materials 54
Uses For Steel By Carbon Content 55
AWS Electrode Classification System 57
Temper Colors 58
Melting Points Of Some Elements 58

Foreword

I have seen a lot of welds in my life. I've seen them on new products and on repair jobs. The variety of welds I've seen lately is enough to convince me that some people **still** don't know where **not** to weld.

In this book I've tried to help those of you who will go on welding a long time after I am gone. I've tried to illustrate as clearly as I know how — in words and in pictures — some of the places that should **not** be welded.

I've also included some tips that will help you do a better, safer job in places that should be welded.

I hope this book will help you be a better welder. The information in it — gathered the hard way over 49 years — has certainly helped me.

Hal Wilson

NOTE

This publication is designed to provide the author's opinion in regard to the subject matter covered.

It is sold with the understanding that the Publisher or Author does not assume responsibility or liability for any applications or installations produced from the design, products, processes, techniques or data set forth in this booklet.

How To Remove A Bad Bearing Race Or Cup From A Hole

If the race or cup is bad, it will be discarded as scrap. Take your arc welder and carefully run a bead around the center of the race, keeping it in the center to prevent drifting off to the other surface. After cooling, it will have loosened enough to fall out.

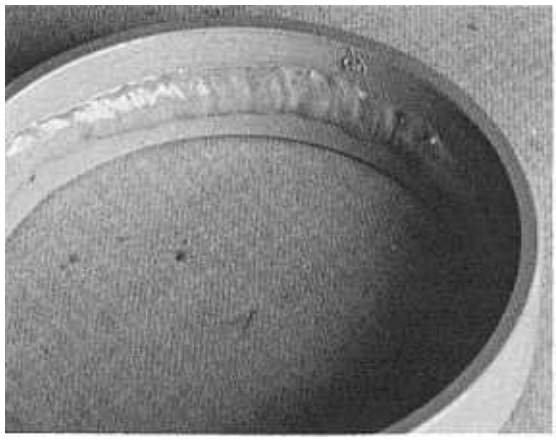

Fig. 1

The above race was 4 7/16 inches in diameter before it was welded. After it cooled it was 7 thousandths of an inch smaller in diameter.

Replace New Race

An easy way to replace the bad race with a new one is to place the new race in the freezing compartment of a refrigerator and freeze it before inserting it. Handle it with gloves.

Occasionally, a race will turn in the hole and wear the hole out too big.

If an internal knurling tool is unavailable, take a center punch and a hammer and make center punch marks all the way around the entire surface that is loose. The amount of wear or looseness determines how hard you must hit the center punch. When a piece of mild steel is hit hard enough to make an indentation, the surrounding metal swells outwardly.

How To Build Up A Worn Shaft

First, turn the worn part off in a lathe. Then, preheat it with a torch to about 250°. Have the shaft in a set of rollers so you can turn it from side to side. Weld a bead on the top side. Turn the shaft a half-turn so the new bead is on the bottom. Then weld a bead on the new top side exactly opposite the first bead. Turn the top of the shaft slightly toward you to make the side of your second bead level. (See Figure 2)

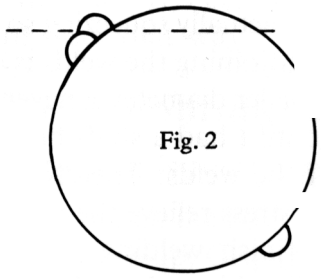

Fig. 2

Then run another bead overlapping to the center of the second bead. Turn the shaft back over to the first bead and run

another bead overlapping the center of the first bead.

You don't need to knock the flux off. After the second and third beads, turn the shaft after each bead to the other side. On the last two beads, weave the rod from side to side to overlap the first and last beads on that side.

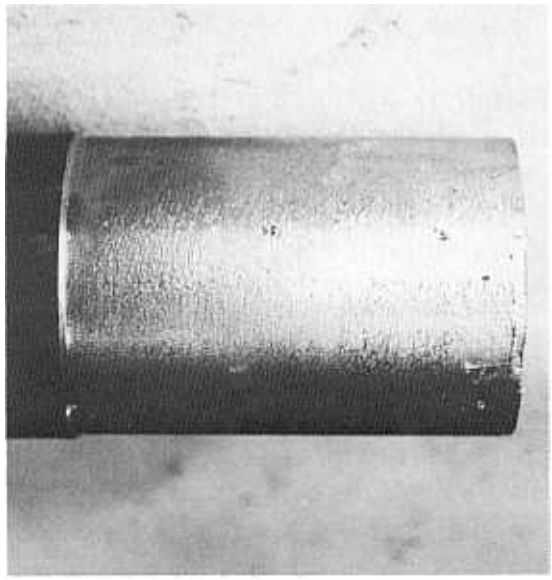

Fig. 3
Shaft turned down ready to be built up.

Fig. 4 shows a built-up shaft. It is hard to believe how much tension is on these welds. They are actually squeezing so hard that the shaft adjoining the welds is compressed to a smaller diameter. I never gave it a thought until I had a shaft break off at the end of the welds. Therefore, you should always stress relieve the welds. As soon as you finish welding, take your acetylene torch and heat the welds to a red heat. Turn the shaft and heat evenly. Hold the heat on it and give it time to get red to the center of the shaft.

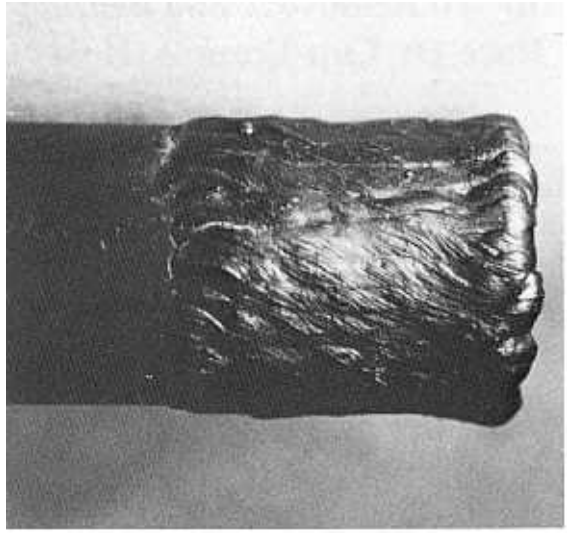

Fig. 4

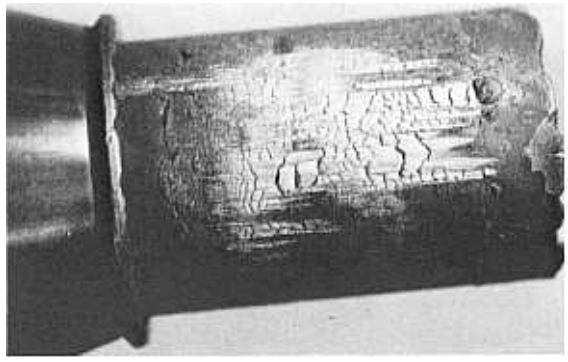

Fig. 5

To prove a point, I turned the welds down on a lathe, to the original size, freely slipped the bearing race on to the other end of the shaft, stress relieved the end that was welded, and let it cool. When it was cold, it was 3 thousandths larger in diameter than it was before I stress relieved it. The bearing race would not come off. I pressed it part of the way off. You can see where I stopped. Then pressed it back on.

How To Burn A Nut Off Of A Bolt And Save The Threads On The Bolt

It's not as hard as it sounds.

You can't burn the threads on the bolt unless you get them red hot. Place the bolt and nut in a horizontal position. Using a small cutting tip, hold the tip parallel to the bolt. (See Below) Start on one of the points and burn through the nut, cutting parallel to the bolt until you get near the bottom. Then, tilt the end of the cutting tip downward and move rapidly through the slot with your stream of oxygen, washing the threads of the nut out of the threads of the bolt.

You can use the same procedure to remove a piece of broken off pipe from a fitting (steel pipe only). Concentrate the heat on the broken off pipe.

Fig. 6

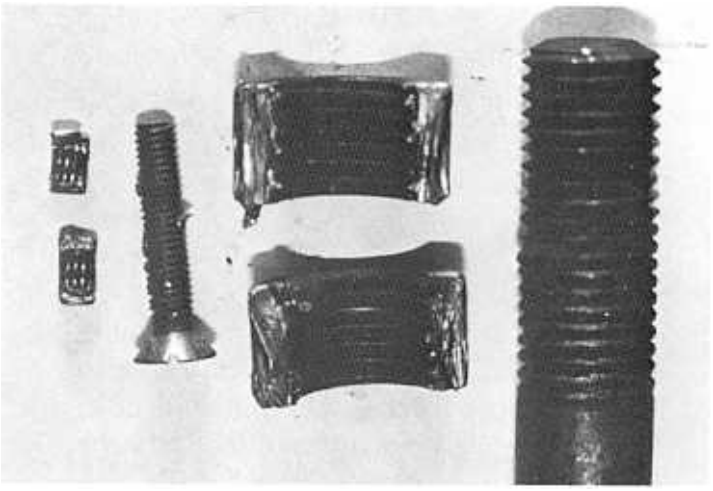

Fig. 7

How To Expand A Nut

To make a nut that is too tight, loose enough to turn with your fingers, do this:

Screw the nut on a bolt. Lay the nut on an anvil and, with a very small ball-peen hammer, peen the flat sides very, very lightly with the ball end. Be extremely careful not to hit it hard enough to make a dent in the flat side of the nut. You can hit it fast, but not hard. Raise the hammer not more than two inches above the nut. If necessary, you may peen all sides that way.

How To Shrink A Nut

When the nut is too loose, do this:

With the nut on the bolt, place the nut in a vise flat side up. Grip the two points of the nut in the jaws of the vise very lightly. Then heat the flat side on top red hot. When it turns black, chill it in water and check it for fit. If necessary, you can heat all six sides this way, one at a time.

Camouflage

The nut below appears to be welded onto the bolt. As you can see, it is not. The nut was welded onto another bolt, cut off, and then drilled to the original "tap drill" size. Then the tap was screwed thru from the back side to tap the weld. This is for larger bolts, the smaller bolts twist off too easily. You can also drill a hole through the nut and bolt, drive a rolled pin through it, putty the ends and paint over it. Occasionally, this procedure may deter a thief.

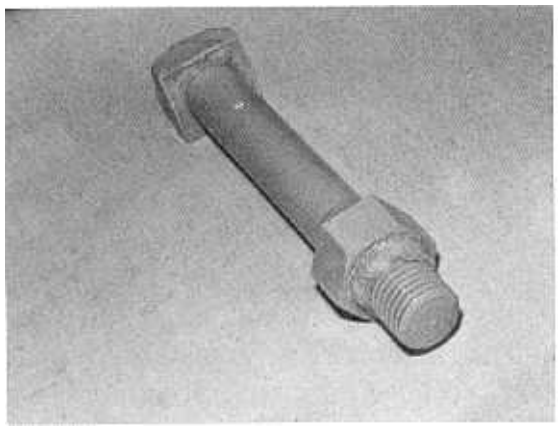

Fig. 8

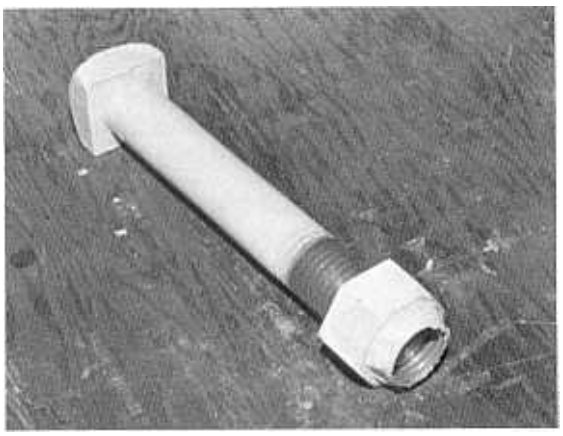

Fig. 9

How To Remove A Broken-Off Bolt

There are several ways to remove a broken-off bolt. "Ridgid" has a very good screw extractor set #10.

Another way is to grind and break off the bent end of an Allen wrench, then grind across the end until it is square. The points should be sharp. Drill a hole in the broken-off bolt the size of the diameter of the flats on the Allen wrench. Then drive the Allen wrench into the hole. The points will cut the metal out and make a hexagonal hole.

These two methods work well — but if the bolt is steel, you can't beat welding a nut on it (See "How To Remove A Broken-Off Tap," page 13).

For example, our shop was once faced with 16 flat head screws, 1/4 by 3/8 — all stuck. The head had a hole for an Allen wrench. We bent an Allen wrench and stripped the hole in another screw. Then, we tried penetrating oil. We hit them with a punch.

Nothing worked — and only three threads holding. Then we welded a nut on it, using the vise grips as shown in "How To Weld A Screen Wire To A Steel Frame." See page 8) It's the only thing that worked.

The reason this method works so well is that when the bolt is welded, the heat radiates through it. Filling the hole in the nut adds still more heat. When the bolt

expands and meets resistance from the hole, something has to give. Either the bolt is compressed or the hole is stretched. Either way, the bolt comes out when it cools.

Shown is one of the 1/4 by 3/8 inch flat head screws with a nut welded to it.

Fig. 10

Case Hardening

Caution: When cutting with a cutting torch, be sure that you do not run out of oxygen. If your oxygen pressure gets too low, a feather will appear on your flame, showing an excess of acetylene. The melted metal on each side of your cut will absorb carbon from the excess acetylene. As you know, when cutting steel with a cutting torch, a very small area gets red hot. The surrounding cold steel absorbs the heat so fast, it is almost like dunking it in water. The result is, the steel that was melted, and absorbed the carbon is case hardened.

I remember 20 years ago, I was cutting some flanges for hydraulic cylinders out of 1/2" plate with a circle burner. My teen-aged son said, "Let me do that dad." I handed him the torch. Later he came to me and said there was something wrong. He was running out of oxygen. We changed oxygen tanks. When I started to machine the 3 pieces that were case hardened, instead of the lathe tool cutting the flange, the hard place cut a groove in the lathe tool. I heated them red hot and covered them with a bucket to retard the cooling. When they were cold, they were soft enough to machine.

Fig. 11

How To Square A Frame By Welding Or Peening

Every weld in a frame pulls the frame toward being square or out of square. The direction of travel causes it. Every weld pulls the metal closer together at the point where you stop welding than where you started. The first welds on the four corners of a frame should be made in the same direction — from inside to outside or vice versa. After you weld the corners, check the frame for square by measuring diagonally across the corners.

If two corners are farther apart, weld a bead on the other side of the frame from the inside to the outside on both corners. If the frame does not square with that, weld another bead on the other two corners, traveling from outside to inside. In other words, **open** an angle by traveling from inside to outside. **Close** an angle by traveling outside to inside.

You can also open an angle by peening on the inside edge. Close it by peening on the outside edge.

How To Flatten A Warped Steel Frame

Flattening a warped steel frame is simpler than it sounds.

When an angle iron frame is warped — if two diagonal corners are raised when the frame is placed on a flat surface, causing the frame to rock — here's how to solve the problem:

Lay the frame on a flat surface with one leg of the angles pointing down and the other ones pointing toward the inside of the frame. (See Figure 12)

Place an adjustable wrench on the end piece of angle on the leg pointing to the inside of the frame. Place it about one inch from the corner that is higher.

At this point, you have to do just the opposite of what it looks like you should do. You push **down**, not up, on the wrench handle until you can feel it give. (See Figure 13)

Then, do the same thing on the opposite high corner. If the frame still refuses to lie flat, place the tool on the angles on each side as shown in Fig. 15 and 16, and push down. The frame is turned over in Fig. 17. If necessary you may pull inward on the high corners and outward on the low corners on the legs that point upward.

Some angle irons are too large to be bent with an adjustable wrench. For these, cut a slot in a piece of heavy steel and weld a long handle on it. Make the handle strong enough to bend the angle iron leg. This procedure works on many different shapes.

When you bend a piece of cold mild steel, it always springs back a little. If you bend it more, it doesn't spring back any more than it did the first time.

When you bend the leg on an angle to flatten a frame, the leg springs back a little. That puts a torsion stress on the angle iron causing it to push down on the high corner. It isn't necessary to bend the leg too much. You have 16 points on an angle iron frame, (four on each corner) to flatten it with.

The iron-worker who taught me how to do this also told me a story:

A new man came to work at the shop where my friend worked one day. The new man built a gate out of very heavy, closely woven wire. When he finished it, it was warped. He worked all morning, hammering and twisting. He even blocked up the low corners and pressed down on the high corners. It flopped — and warped the other way. What had been the high corners were now the low corners.

My friend watched out of the corner of his eye until the new man took off for lunch.

"Then I went over there and straightened it out with a pair of pliers," said my friend.

When the new man came back from lunch, he checked all four corners. They were all flat on the floor. He turned the gate over and checked it. He stared at it, scratched his head, looked all around the shop and demanded, "Who straightened this for me?"

He never did find out.

My friend's secret? He put stress on the wires near the frame with his pliers. Just the opposite from the way it looks.

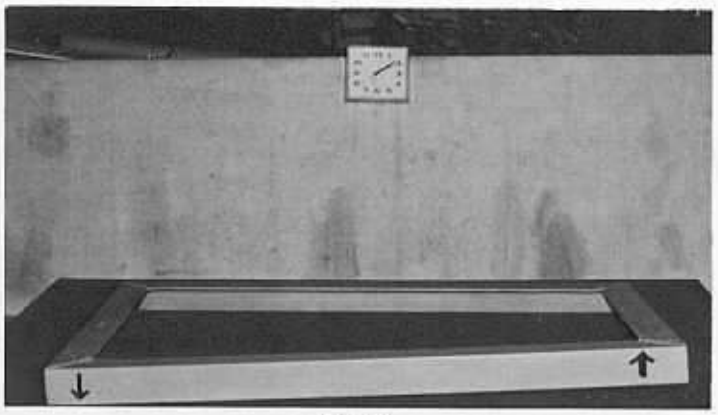

Fig. 12

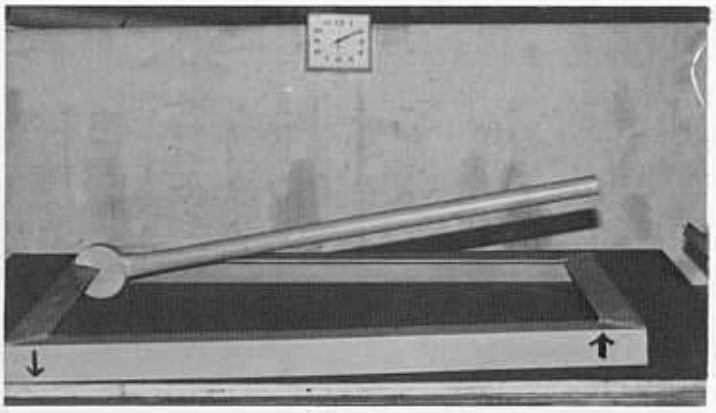

Fig. 13

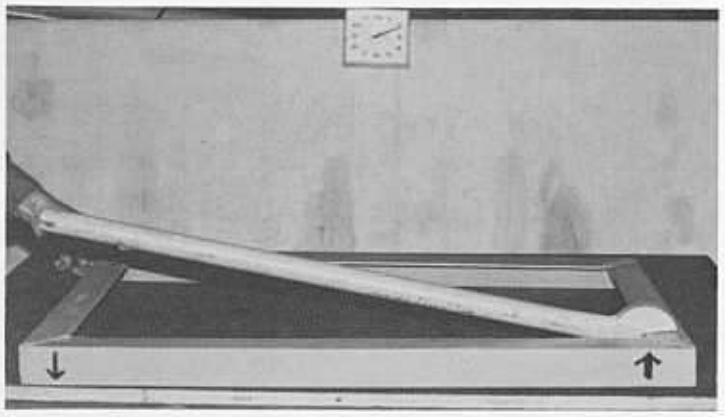

Fig. 14

Notice the clock. I placed the tool, pushed down on the handle, left the tool in place and then made the picture, on each one of these. I straightened one without making pictures in less than one minute.

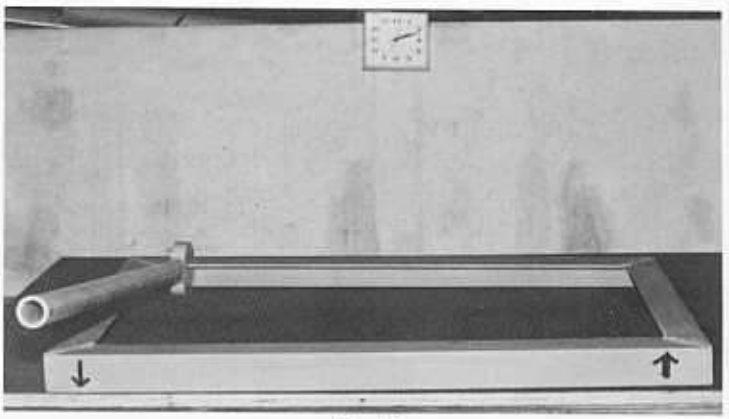

Fig. 15

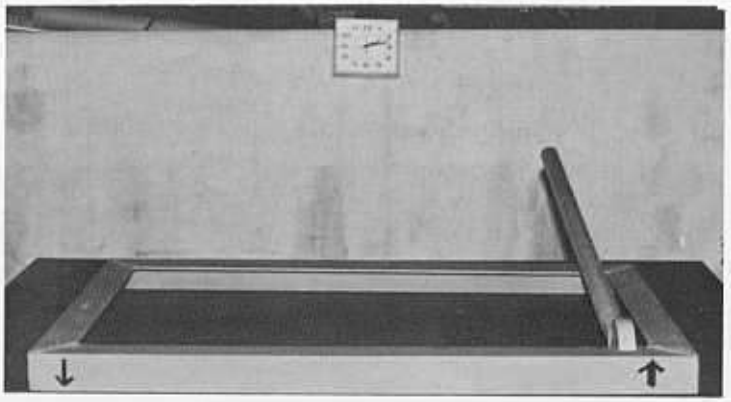

Fig. 16

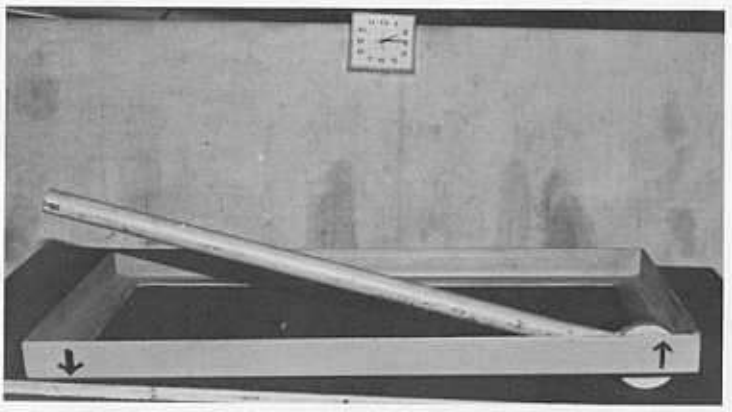

Fig. 17

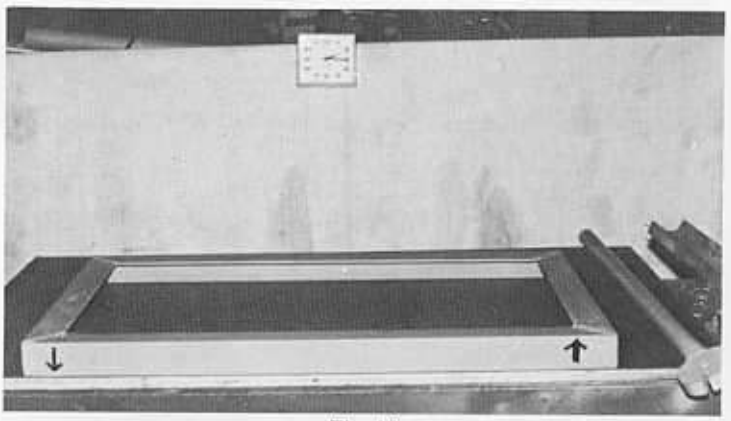

Fig. 18

How To Weld A Screen Wire To A Steel Frame

If you're having trouble welding a screen wire or light expanded metal to a frame, try using washers.

Place a flat washer on the wire where you want to weld it to the frame. Clamp it down with a pair of vise-grips. (See Figure 19) The big, partly cut-off washer is added on, lightly tacked.

Weld through the hole in the washer and into the frame. Fill the holes in the washers with weld metal. (See Figure 20) If you want a really neat job, cut a strip of 1/8 x 3/4 inch to the correct length and drill 5/16 inch holes in the center of it. Clamp it in place and weld through the holes. (See Figures 21 and 22)

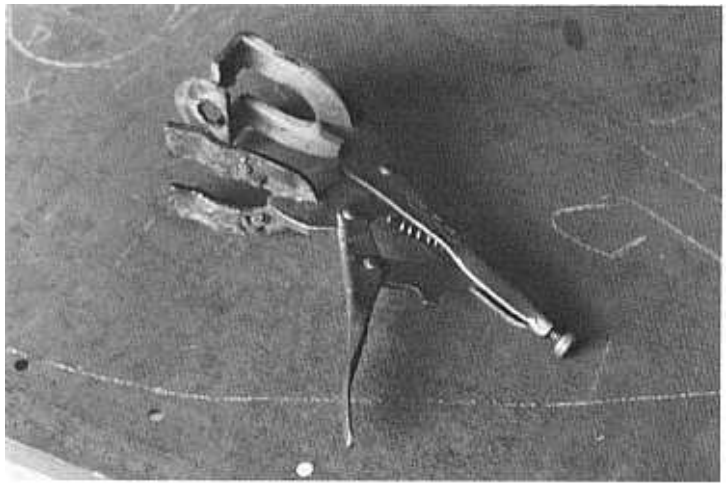

Fig. 19

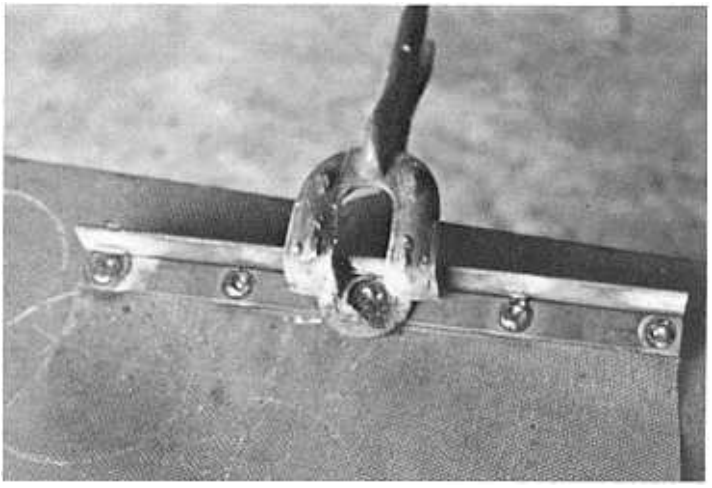

Fig. 21

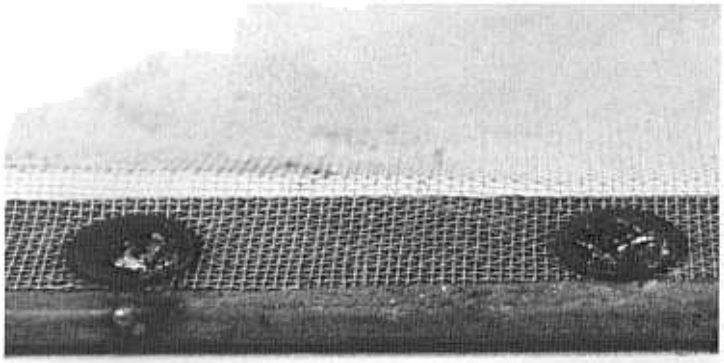

Fig. 20

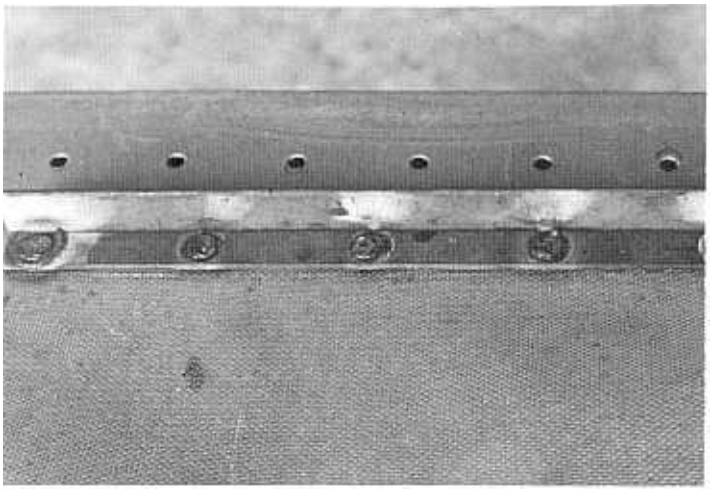

Fig. 22

How To Make A Smooth Cut With A Cutting Torch

For a machine-like cut, use a straight edge. Take a 1/2 by 2-inch piece about 12 to 14 inches long and a 1/8 by 3-inch piece the same length. Weld or tack the 1/8-inch piece on top of the 1/2-inch piece, letting the smaller one overhang a half-inch on each side. File the 1/8-inch edges smooth. Place the file flat on the 1/8-inch edge lengthwise to keep it straight. Keep the edge filed and sanded to keep it smooth.

To cut, place straight edge far enough away from the cutting line to make the center of the cutting tip directly over the cutting line when the tip is rested vertically and lightly against the straight edge. Drag the torch toward you. You will not be able to see the flow of metal. If all the sparks and fire are landing on your feet, you're traveling at the right speed.

If the slag on the bottom of the cut is too hard to knock off, you are using too much heat. Cut down on the flame, use a smaller tip, or travel a little faster.

Incidentally, the acetylene gas is not what does the cutting. When you pull the trigger on red hot steel, the oxygen attacks the carbon in the steel. The carbon burns and supplies most of the heat. You can prove it by burning the end off a piece of 1/4 by 2-inches. Quickly start another cut on the end. Have your hand on the acetylene valve. Get the cut started and turn off the valve. You can finish the cut without a flame.

That's why you can't cut other metals or even stainless steel with a cutting torch.

How To Do Overhead Welding

Never try an overhead weld with a cold machine. Have the machine hot enough to melt the base metal almost instantly. If your machine isn't hot enough, you'll have to stay in one place too long to melt the base metal and your weld will start bulging downward — it may even drop off. Lay in a little rod and move it away for an instant, before it starts to bulge downward. This lets the crater chill. Then go back, but not as far as you were before.

You can also run light stringer beads. Hold a short arc and move along fast enough to keep it from bulging, but slow enough to fuse.

How To Weld Thick To Thin

To weld a thick piece of steel to a thin one without burning a hole in the thin one, try concentrating the arc entirely on the thick piece, moving slowly toward the thin piece with the crater edge until it fuses to the thin piece. Then back away before it burns a hole.

How To Make A Smooth Bead

Every welder is judged by the looks of his work.

To run a smooth bead, position is very important. Get comfortable. Use a rest for your left hand. With the electrode holder in your right hand, rest your right hand on your left hand so that you can burn a complete rod in a smooth, steady movement. Practice the movement without the rod.

Vertical Weld

If a vertical weld bulges in the center or drops off, you are probably coming back down too low. When you raise your rod to give the crater a chance to chill (freeze or solidify), you are coming back down too far. Try this method:

Bring the end of your rod back down and across the top edge of the crater that just chilled. You can also run a vertical bead downward. Use a fairly hot machine, tilt your rod upward, and run a light bead with a close (or short) arc. Travel fast enough to keep the flux and melted weld metal from trying to get ahead of or below the end of your welding rod, but slowly enough to get fusion.

How To Flux Lead

If you have a pot of melted lead and it has a lot of skimmings on top of the lead, you can skim it back and it will be a different color. The color changes back rapidly to the original color. This is because it oxidizes when it comes in contact with the oxygen in the air. Don't skim the skimmings off and throw them away. Sprinkle a little pine rosin on the top of the skimmings and stir it up. The pine rosin is a flux for the lead. It breaks the oxide and it all melts. This also works on babbitt and solder - you have heard of rosin core solder.

Bulge

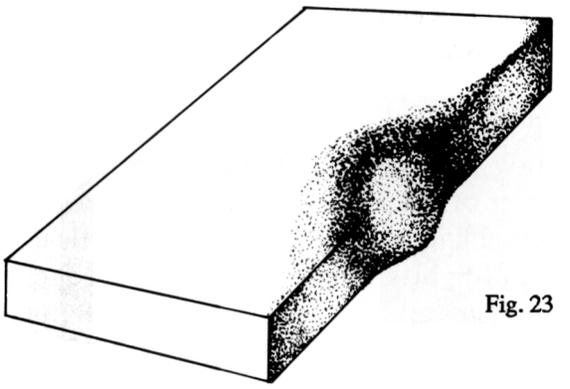

Fig. 23

The bulge in this sketch is highly exaggerated.

If you can imagine that this really happens when a piece of steel is heated red hot in one spot, it will help you figure out a lot of problems that arise in expansion and contraction.

When you start heating one spot, the steel begins to expand on that side. The center of the strip will be under tension and the other side will be under compression. But when the spot gets red hot it has very little strength. The red hot part is pushed together causing it to swell. When it starts to cool and begins to turn black it has more strength. As it contracts, one side of the steel becomes shorter than the other. And the strip bends edgewise.

How To Burn A Weld And Save Both Pieces

If, for instance, you have two quarter-inch pieces lap welded to each other, here's how to handle it:

Light your cutting torch, then cut the oxygen pressure down at the regulator — to the point where you can barely hear it when you pull the trigger.

Position is crucial. Position yourself slightly to the left of the work. Hold the torch with your left hand near the tip. (Have some sort of support handy to rest your left hand on.) Hold the torch with your right hand ready to pull the trigger. Start at the left and travel right. Place your head where you can see the flow of metal. The tip of the torch should be at about a 35-degree angle. Heat the corner red hot and pull the trigger.

When the flow of metal goes through the weld, you will see a black spot at the bottom of the flow. It's okay. Keep on going. If the flow of metal flies back in your face, the oxygen pressure is too high. If you don't see that black spot at the bottom of the flow, you are cutting through the weld and into the other piece of quarter-inch plate. Tilt the tip down more and get below the weld with the flow of metal.

WRONG

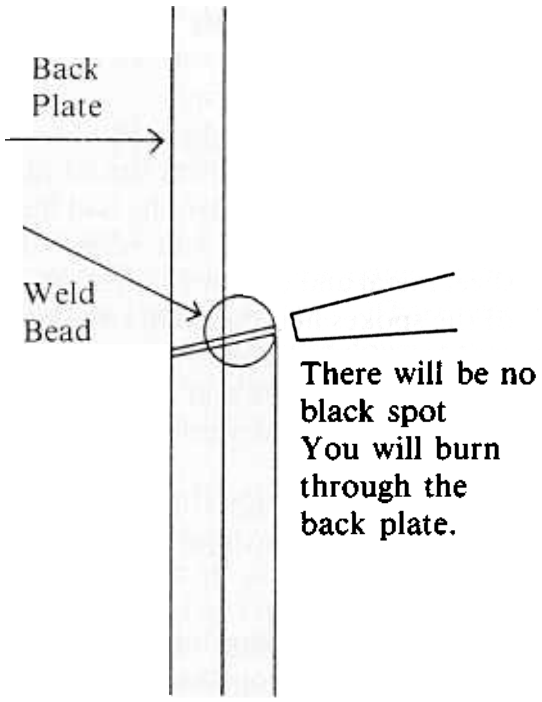

Fig. 24

RIGHT

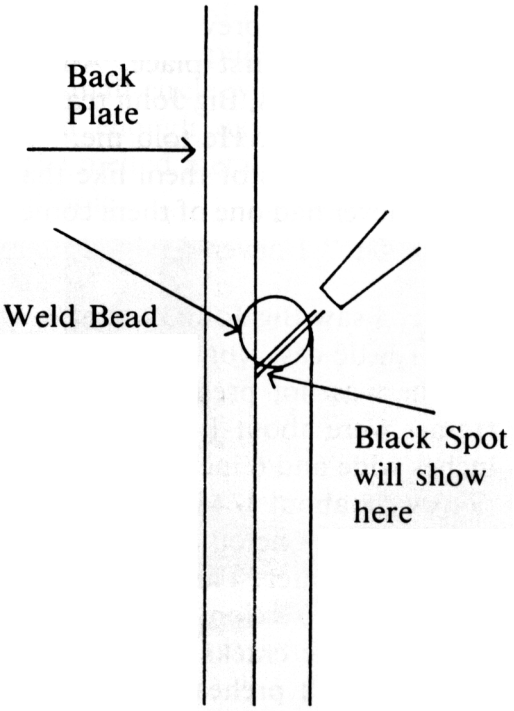

Fig. 25

Big John

Big John had been working at the shop for 16 years when I started my apprenticeship. He was a top notch welder. He welded all of the broken truck frames — and there were a lot of them back then! Big John had his own way of reinforcing them. He didn't put a fishplate on the side of the channel as you might expect. After he welded the crack up, he would weld a flat bar (about 1/4 x 2-inches or thereabouts) on the top or bottom leg of the channel, extending six or eight inches on each side of the weld. Some of the frames started cracking at the top if the load was too far to the rear of the rear axle.

Big John argued that the weld would hold as much weight as any other part of the channel, and if the reinforcement was put on the bottom (or top) leg of the channel it would prevent the crack from starting in the first place. And if the crack never starts, Big John reasoned, it will never break. He told me, "I have welded hundreds of them like that and I have never had one of them come back on me yet." I never doubted him.

Once I saw Big John brazing a whole pile of little cast iron frames. I noticed that he was not preheating them. The frames were about 1/8 inch thick by 4 inches wide and 6 inches long. The sections were about 3/4 inches wide — one on each side, one on each end, and one across the center. They were all broken on the center section. He had previously ground the cracks out. I asked him why he didn't preheat them.

He told me, but only after twisted his arm. Here's his trick:

Big John held the torch at an angle so that, while he as heating the crack on the center section, the other part of the flame extended over to the end section and heated it. That let the two sections expand together and then shrink together. If I had been doing it, I would have heated all three cross sections.

learned a lot from Big John

Another time, I was helping him with a big steel flywheel, about six or eight feet across with a heavy rim and spokes. We blocked it up on bricks to keep it off the dirt floor. Then, we placed several natural gas torches around it and started heating it with soft flames (no air pressure), on the rim only and covered it with some sheet metal. The heat had not been on it for more than five or ten minutes when there was an explosion that shook the building.

The boss came running out of his office to see what catastrophe had hit. By that time, Big John had removed the sheet metal and cut the torches off. One of the spokes had pulled in two, leaving a wide gap (the spokes were oval, three or four inches thick and ten or twelve inches wide — solid steel).

"John," the boss growled, "you heated the rim too hot."

"Look boss," Big John said, "if that crack goes back together when it cools,

you can have my job. But if it doesn't, don't open your mouth."

Here's what happened.

The guys who removed the flywheel had trouble getting it off the shaft. They heated the hub of the flywheel red hot in order to remove it. In doing so, the hub tried to expand outwardly, but it couldn't. The rim and spokes were cold. They removed the flywheel and the hub began to shrink. The shrinking of the hub put tons and tons of tension stress on all of the spokes. When just a little heat was added to the rim, which was under compression, it expanded enough to pull the weakest spoke in two. The gap did not close up when it cooled.

Notch Effect

Cut a notch in a shaft, and you know where it will break - in the notch. This is called a "notch effect." There are many others. A shaft turned down in one place to a smaller diameter will break right next to the larger diameter.

A built-up shaft will break at the end of the welds. A poor weld is a notch effect. A shoulder of almost any kind is a notch effect. The end of a reinforcement is a notch effect which can be minimized by tapering the ends.

How To Remove A Broken Tap

Place a nut over the tap and weld the tap and the nut together. Fill the hole in the nut with weld metal. Use a reverse rod with a light flux. Let it cool.

Before you start, the nut needs to be held in place. One way to do this is to use a washer with a hole bigger than the hole in the nut. Weld it to the end of a flat strip of metal. Place the washer over the nut. Block up the other end of the flat strip to make it level and place a weight on the flat strip. (See Below)

With a little skill, a broken off tap can be removed with a cutting torch. Concentrate the heat on the tap only. Hold the tip in a straight line with the tap. Let about half of the flames go down the side of one of the flutes, and the other half on the center part of the tap. When the tap gets red hot, pull the trigger and make a small circle with your tip. You will need enough oxygen pressure to blow the melted metal back out if it is a blind hole.

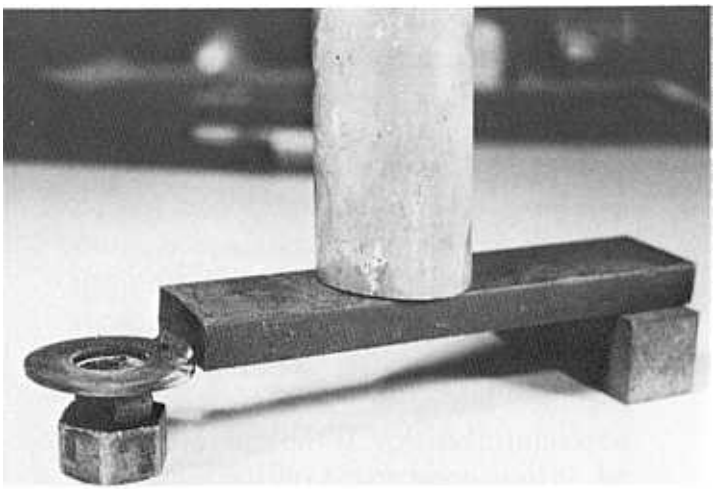

Fig. 26

Built In Stresses

Every weld contracts or shrinks in all directions. Steel is made up of round crystals and expands when it is hot. When a piece of steel is heated uniformly to a red heat, it is free to expand. But if it is heated in one spot, it is not free to expand. When it tries to expand outwardly the cold metal around it prevents it. The crystals are rearranged to make the metal thicker at that point. Then, when the metal cools, it contracts. This puts the crystals under tension. This will cause a thin sheet of metal to buckle. All welds are under stress unless the stress has been relieved. A light bead on a very heavy piece of steel usually cracks the weld wide open. If the hub of a cast iron wheel is heated red hot, it usually pulls one of the spokes in two.

However steel sometimes has stresses already built up in it. When you weld, you may relieve more stresses than you create. Anytime a piece doesn't move the way you expect it to, blame it on the built-in stresses. When a machinist takes a cut on a straight shaft, the shaft sometimes bends a little due to the stresses being cut off. That is probably why the ball-peen hammer was invented. You can peen the inside of a curve on a shaft to straighten it.

If you repair a broken piece of steel you should reinforce it with another piece of steel, after welding it back together. But **never, never** weld the ends of a reinforcement. If the ends are welded, it puts great stress on the metal adjoining the weld. A little more stress and vibration will cause a crack right next to the weld. Cut the ends of the reinforcement on a 45° angle instead of a 90° angle. Always skip weld the reinforcement — weld an inch and skip an inch. Never weld all the way to the ends of a reinforcement. Leave about a quarter-inch, or the thickness of the reinforcement.

Crystals

When a machine part is subjected to severe stresses over a period of years, sometimes the crystals become elongated. If a part breaks, they say that the steel crystalized. I was told that if you heat a piece of steel to a red heat, the crystals would return to their original shape.

To prove this to myself, I cut some 1/2" round, mild steel bars. I hammered one of them cold, on the end, to bring it to a point. It didn't taper much until it cracked lengthways. The crystals in the steel elongate when the steel is stretched, but they will only elongate so far and then they separate, and that is a crack. Then I hammered another one, but I stopped before it cracked, then heated it to a red heat and waited until it was cold before hammering again.

I continued this process a number of times. The center of the bar did not stretch much, if any. A hole was left in the center. It seems to be quite evident that the crystals return to their original shape when heated red hot.

How To Burn Through A Large Steel Shaft

You will need the biggest cutting tip you have and all the oxygen pressure your regulator will give you. If your tip is big enough and your oxygen strong enough, you can cut right through. But if the tip is just a little too small for the job, try this:

You must be able to see the flow of metal all the way through the shaft at all times. With the shaft in a horizontal position, place yourself on one side of it. The cutting tip should be on about a 45° angle from vertical. You should be able to see over the torch and watch the flow. Before you light the torch, practice moving it in the path that it will travel. Then light it and heat a starting spot. Don't be in a hurry to start cutting; get it red hot. When you do start cutting, weave right and left and make a gap about three-quarters of an inch wide (for a four-inch shaft, wider for a larger shaft). If the gap is not wide enough and you can't burn all the way through, go back and start over again on one side. Try burning off another half-inch. If that doesn't work, your tip may be too small.

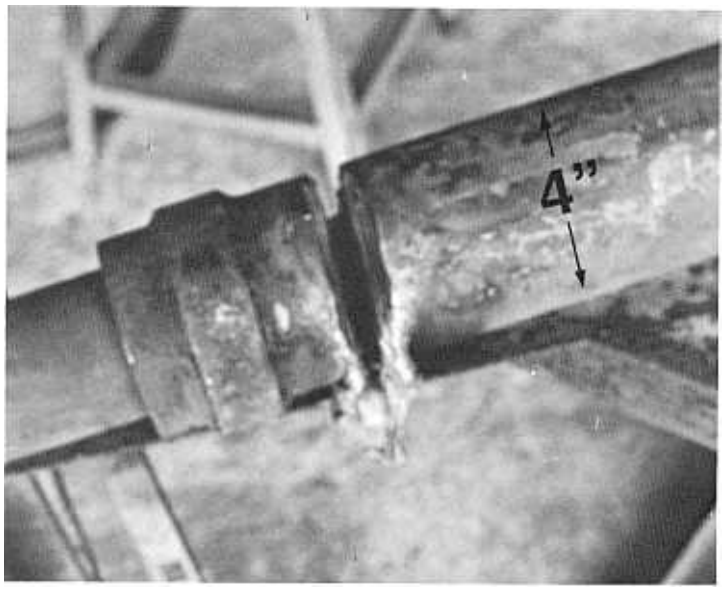

Fig. 27

My First Commercial Weld

In 1936 I was a welder's helper. I had been practicing running a bead for some time. One day someone brought in a broken bumper support from a Model T Ford into the shop. The welder welded it back together and tacked a reinforcement over the weld and looked at me and said: "You can weld this, son. Go ahead."

I welded the top and bottom and both ends. Later, the vice president of the company came through and stopped to talk with the welder. In a few minutes he noticed my weld. He jumped straight up and yelled, "Who did that?"

The welder looked at it and realized what I had done. He said, "It's all my fault, boss. I didn't tell him not to weld the ends. Don't worry about it. I will cut it off and put a longer piece on it."

The boss cooled down. And that was the only time I **ever** welded the ends of a reinforcement.

Cast Iron Welded W ı Nickel Electrodes

If you have a cast iron ornament that was broken accidentally, you can successfully weld it with a nickel electrode. The cast iron should be preheated. However if a cast iron part is subjected to stress and you weld it with a nickel electrode, it usually breaks right next to the weld. I am not a metallurgist and I don't know why, but I think it has something to do with getting the cast iron too hot. Cast iron can be successfully brazed with a bronze rod and an acetylene torch. Chamfer the break and grind 3/4 inches clean on all sides. The bronze will have to be built up a little bigger to equal the strength of the cast iron. Heat the cast iron to a dull red, heat the rod and dip it into "Cast Iron Brazing Flux" and melt it into the Vee. It has to tin the surface of the cast iron (spread out) if it balls up, it will not stick. Fill the Vee and then spread out over the edges that you ground clean. Dip your rod in the flux whenever it is necessary to keep the bronze tinning and flowing into the pores of the cast iron.

Now this procedure is for a single piece of cast iron that is free to expand and contract without putting a stress on another place. On a wheel, pulley, or gear, they can be brazed if you heat them slowly and uniformly to a dull red heat. Some welders can heat the rim of a wheel at each spoke (that allows the heat to radiate each way on the rim and inward on the spokes). As the rim expands and gets larger in diameter, the spokes expand to keep out stresses.

At some places, they weld cast iron blocks and heads for Diesel engines. They grind the cracks out then place the block in a portable furnace. Heat it slowly to a red heat, open a door on top leaving the heaters on, they reach down in the furnace and acetylene weld it with a cast iron rod. Then they close the door and leave the heat on for awhile. When they cut the heat off they leave the block in the furnace. The block is surrounded by red hot fire bricks. It takes some time for it to cool. Then they machine it. It costs about half as much as a new block.

A foundry made me a cast iron sheave with an extra large hub in order to bolt a brake drum to it. When I picked it up, it had a crack in one of the spokes. I showed it to the foreman. He said the mold should have been stripped. He explained that when it took the hub longer to cool than the rim and spokes, they were supposed to remove the sand off of the hub to make it cool as fast as the metal on the rim and spokes. The rim cooled first and shrank, forcing the spokes to squeeze the hub together, making it thicker. Then when the hub cooled and shrank, the rim would not "give" and the hub pulled one of the spokes in two.

Straighten 6 Inch Strip

If you cut a 6" strip off of a 1/2" plate. 8 feet long, with a cutting torch, it will be curved. The side that you cut contracted and now it is shorter than the other side. It can be straightened by stretching the side that you cut, by peening or by shrinking the other side. Stand on edge and with an acetylene torch heat the edge red hot. As it gets red move along. Have someone come behind you cooling it with water.

Clevis On Channel Iron

Figure 28 shows how **NOT** to weld a clevis on to a stress member.

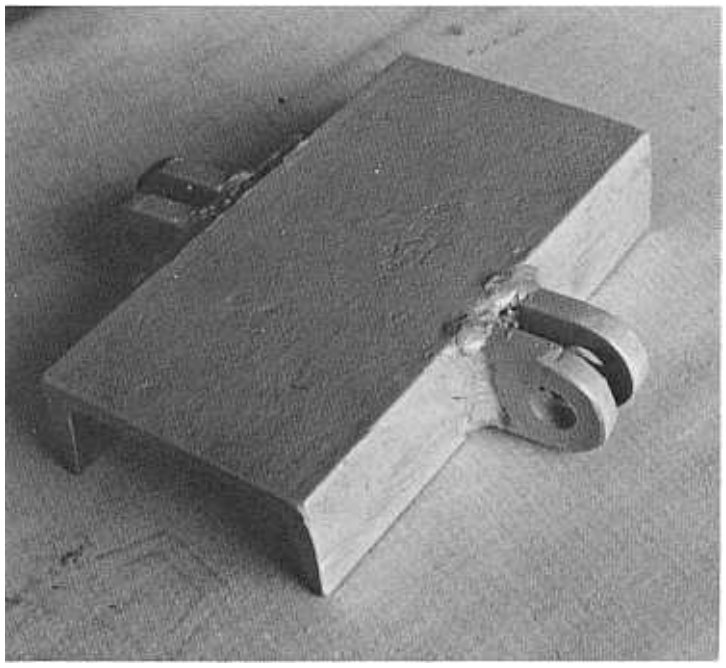

Fig. 28

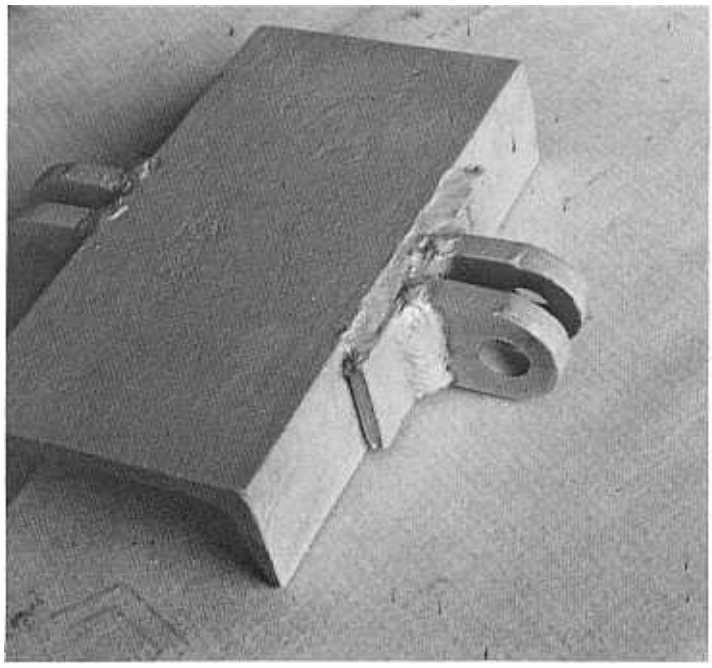

Fig. 29

Figure 29 shows a better way to weld it. The two pieces with the holes are welded on to the patch, then the patch is welded on to the channel iron lengthways with the channel. No doubt you are tired of reading "Do not weld the ends."

How To Temper A Chisel

The cutting edge of a chisel should be soft enough to be filed with a file, but no softer. If it is too hard it will break.

You will need a tub of water, a file, a pair of tongs or vice-grips, and a torch.

Slowly heat the chisel from the cutting edge back about three or four inches to a red heat. With the tongs holding the chisel on the back end in a straight line with the chisel, place the tip of the chisel into the water about a half-inch or more. Hold it there until it cools. Then rapidly plunge the entire chisel into the water and back out. Rest the chisel on the edge of the tub and, with the file, rub the end of the chisel. It should be too hard for the file to cut. Keep stroking the tip until the file starts cutting. Then, instantly shove the whole chisel into the water and back out again. The tip should again be too hard for the file to cut. Continue filing and dunking until it is soft enough to file when you take it out of the water. Then place the chisel in the water and move it around until it is cold.

When I was serving my apprenticeship, a contractor would send in an armful of drills every few weeks to be drawn out and tempered. The drills were about one and a quarter inches in diameter and about 18 inches long. They were used in an air hammer to break rock and concrete.

A friend had told me how to temper a chisel. I decided to try it out on these pointed drills. Without permission.

A few weeks later they brought in another batch with special instructions that the same man who tempered the last ones to temper these.

It worked.

How To Make A Circle Lie Flat

An 18-inch circle that has been burned out of a 1/8-inch plate will bulge in the center. It does so because the outside edge that was heated to a red heat is under tension. To flatten it, remove the slag, place the edge of the circle on an anvil, and hammer it with the flat face of the hammer centered on the edge of the circle. Hold the other side of the circle up so the edge will be in flat contact with the anvil.

Turn the circle and hit the edge all the way around, hard enough to stretch the edge, but not hard enough to distort the metal too much. Try leaving a little space between each stroke and, if that isn't enough, go around again, hitting between the last strokes. If the edge is stretched too much it will buckle. The buckle may be corrected by heating red hot on the edge in very small spots. Cool each spot with water before heating another one. Or you can skip from one side to the other, letting someone else cool the spot you just heated.

On the deck of a ship which has many buckles, one man heats spots and another one cools them with air and water.

How To Remove A Stuck Sleeve From A Hole

Screw a tap that is slightly larger than the hole in the sleeve into the sleeve — it needs only to scratch the inside of the sleeve. Drive it out from the other end with a mild steel rod (never hit a hard surface with another hard surface).

Another way to remove a stuck sleeve is to heat the sleeve red hot and let it cool. You can heat a large sleeve red hot in a straight line from one end to the other — it does not have to be red hot all at the same time. Heat red hot in one spot and, as it reaches red hot stage, move forward. The black part behind you has already shrunk.

This works on steel. Don't heat cast iron unless you are sure you know what you're doing.

How To Shrink A Steel Pulley

We once had a steel fabricated pulley that had a tapered hole in the hub to fit a split Q.D. hub. The Q.D. hub had a flange which would pull up against the hub on the pulley without being tight on the shaft.

To fix it we heated the rim of the pulley red hot with a torch and slowly moved toward the center, heating it red hot as we went — heating the hub red hot in one place also. It helped, but it wasn't enough.

When we heated it in four places on quarters it worked fine. There was a quarter-inch of space between the flange and the pulley and the hub was tight on the shaft.

CAUTION: do not try this method on a cast iron pulley.

Reinforce Inside Of Tubing Or Pipe

If you reinforce a broken-off piece of tubing with a smaller piece of tubing or solid bar that fits on the inside of the broken off tubing, do not weld through the broken piece and into the smaller reinforcement on the inside. If you do, it puts it under stress and may cause a break. You can see that if you run a light bead around a solid shaft, the light bead has to shrink, but it will have a hard time compressing all of that cold steel on the inside of the small piece.

Cut slots lengthways on the broken tubing to weld through and into the reinforcement before you put the reinforcement inside. The size of the slots depend on the size of the tubing. There should be enough welds on each side of your weld (that welds the outside piece back together) to equal the strength of the tubing. Stagger the slots. With the reinforcement in place and the broken tubing lined up, weld the broken tubing back together. Try not to weld through to the inside piece of tubing. Let it cool, then weld up the slots.

How To Repair A Crack In A Piece Of Tubing

Cut out piece as shown in Fig. 30 - shaded area.

Then cut out patch to fit hole, out of same size tubing. Fit patch in hole smoothly and tack at each end, A and B. Then weld one side only — from D to B — C to D, then A to C. Let cool, then weld other side — same way.

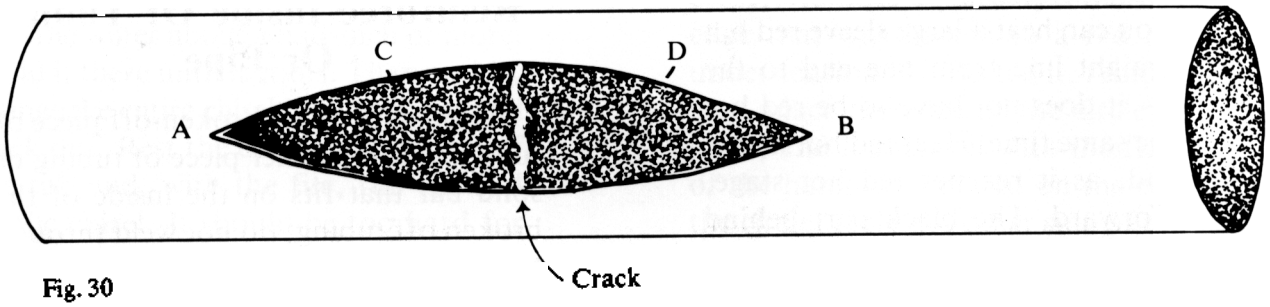

Fig. 30

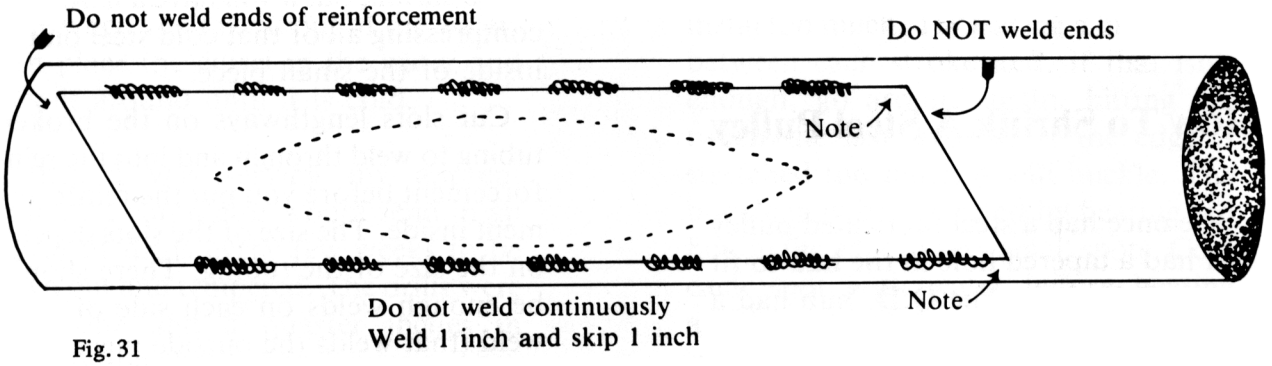

Fig. 31

Grind welds on patch and place reinforcement over patch as shown in Fig. 31. It is best to always grind long ways with a stress member. Cross scratches can sometimes cause a crack. Do not weld all the way to the ends of the reinforcement — leave 1/4 inch or the thickness of the reinforcement.

Weld Inserted Pipe

If you weld around the pipe as shown in Fig. 32, you will create a "notch effect" which usually causes a break adjoining the weld.

Slots may be cut in the larger pipe to weld through as shown in Fig. 33 or holes may be drilled to weld through. Slots or holes should be staggered, with enough weld metal to equal the strength of the pipe.

If you don't have room enough for the slots or holes, you may cut slots on the end of the larger pipe as shown in Fig. 34. The direction of travel should be away from the end of the larger pipe. Start your weld slightly to the inside of the slot.

If the joint must be air-tight, the larger pipe can be cut as shown in Fig. 35. When two beads meet at a point or crotch, extend one bead a little further and try to taper it out to nothing. This prevents stopping the shrinkage abruptly.

Some welders may not agree with me on the "notch effect." You think about it. If every weld shrinks, the last part of your bead will shrink more than the first part. The first part of your bead is already black. Even if the weld didn't shrink, there would still be a shoulder.

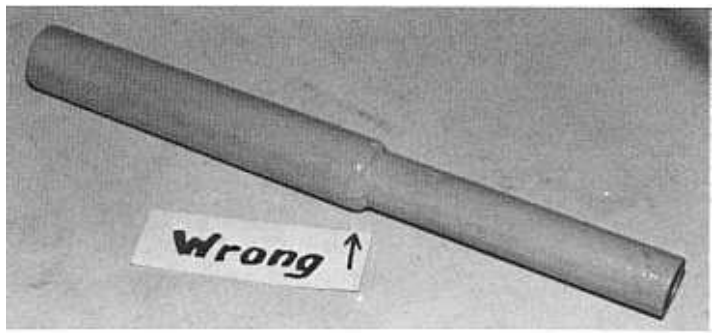

Fig. 32

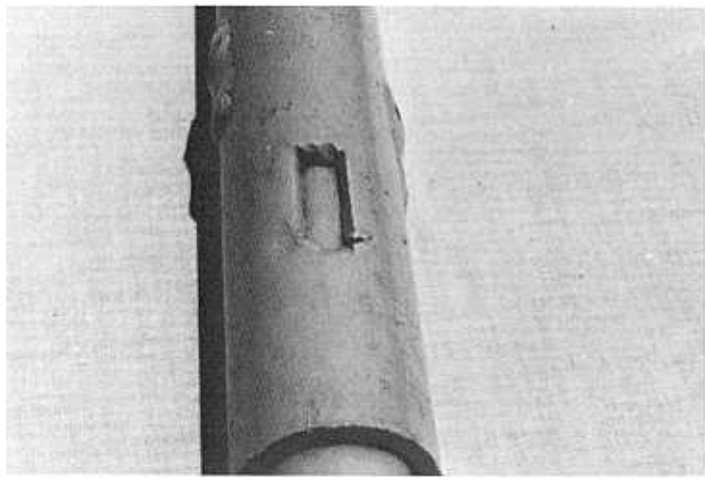

Fig. 33

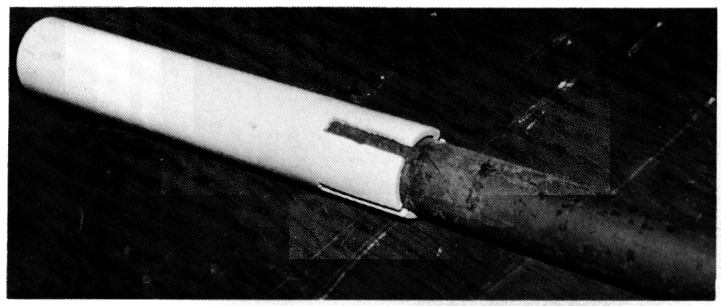

Fig. 34

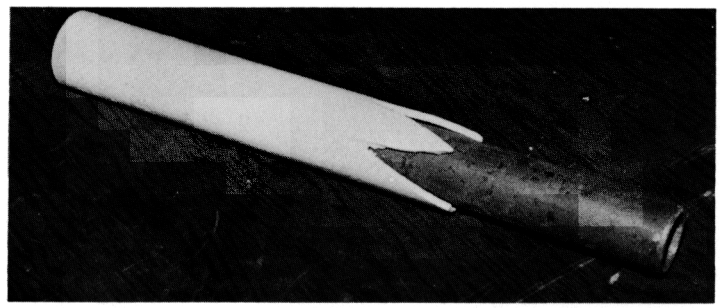

Fig. 35

Another Method For Welding Pipe

It is general practice to bevel two pieces of pipe and butt weld them together. However, if you are using pipe or tubing for structural purposes, and the welds are subjected to tension, torsion, impact or shear stresses, and you have trouble with cracks, you may want to consider the method shown below. The method shown below has no cross welds at all. The welds are back-stepped, and at the ends where the two welds meet, they are reduced to one weld which tapers out to nothing to avoid stopping the shrinkage abruptly. This method is more expensive than a butt weld, but you have a lot more weld metal securing the joint.

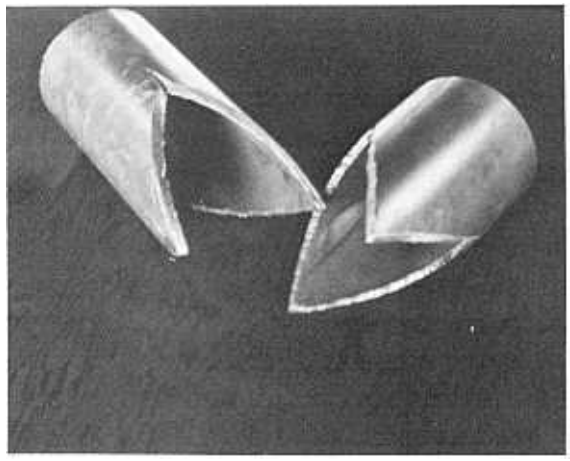

Fig. 36

Make a symmetrical pattern to cut the pipe out with, as shown above.

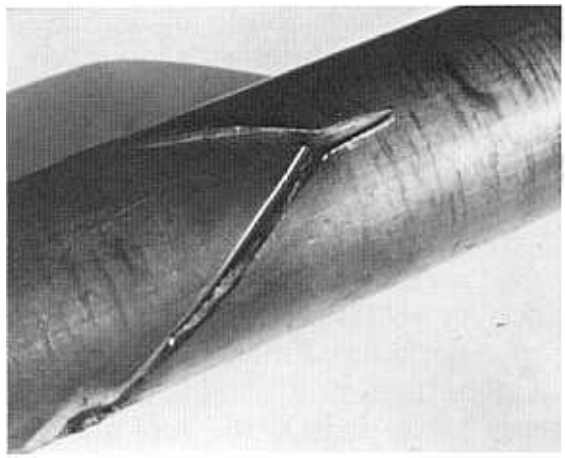

Fig. 37

Bevel edges and grind slot at end.

Another Method For Welding Pipe

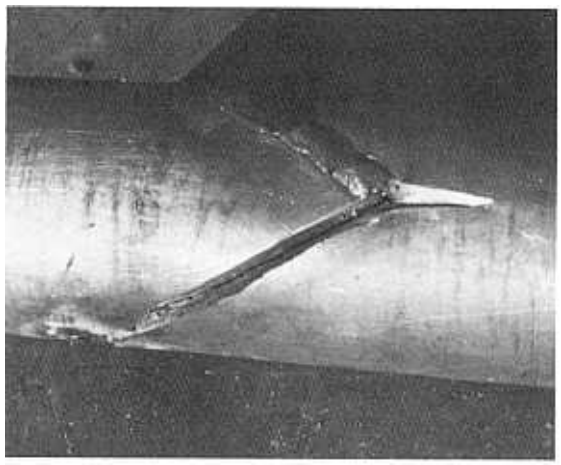

Fig. 38

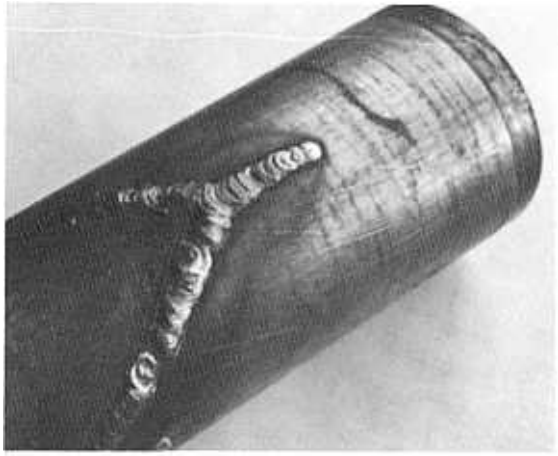

Fig. 39

Weld up two grooves that are exactly opposite of each other and let them cool. The shrinkage will open the other two grooves slightly. Then weld the other grooves. This way, you will have less shrinkage on the joint.

Build Up Inside Of Tubing

The pictures below show that if a tubing is built up on the inside, it will cause the tubing to shrink. This 4 1/4" O.D. piece shrunk .045" in diameter.

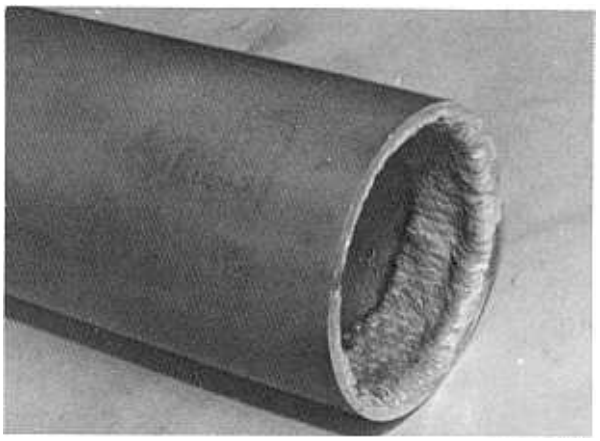

Fig. 40

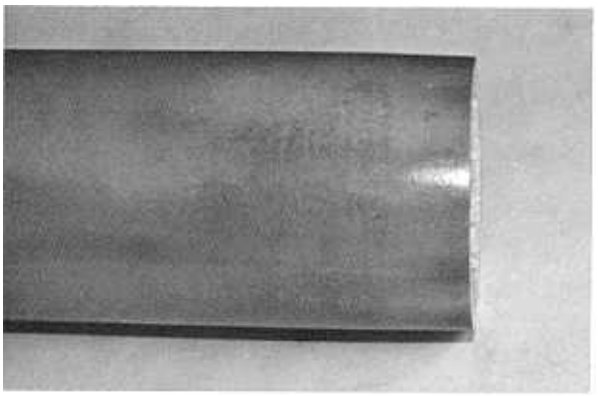

Fig. 41

Test Welds

Figures 42 and 43 are test welds that show how much the tubing shrank from end to end. Figure 42, welded lengthwise, shrank **four** thousandths of an inch in length. Figure 42 probably shrank more crosswise, but that does not cause it to crack.

Figure 43, welded crosswise, shrank **twelve** thousandths of an inch in length. The cold metal surrounding the welds resisted. The cold metal is now under compression stress and the welds and the adjoining metal that got red hot is under tension stress almost to the yield point.

Think about this and **please** don't weld crosswise on a stress member.

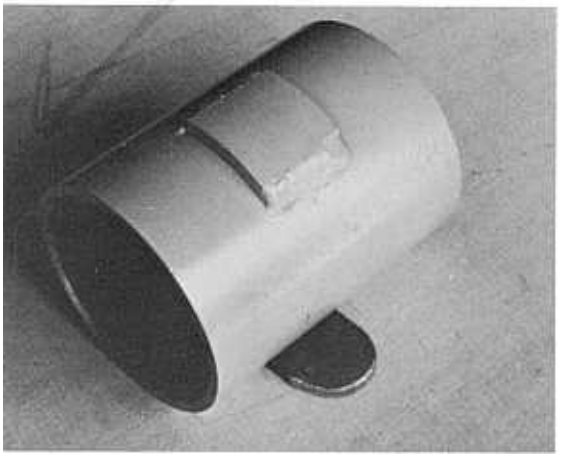

Fig. 42

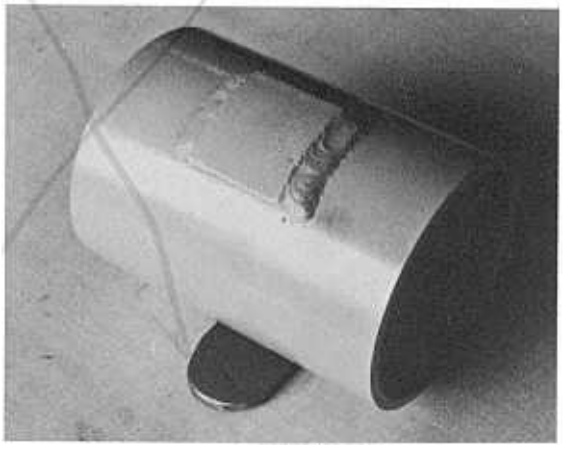

Fig. 43

Flying Coaster

No. 1 shows large gap to prevent someone from welding crossways on the clevis. On the old ride they were welded crossways and they broke off.

No. 2 shows end of patch is not welded.

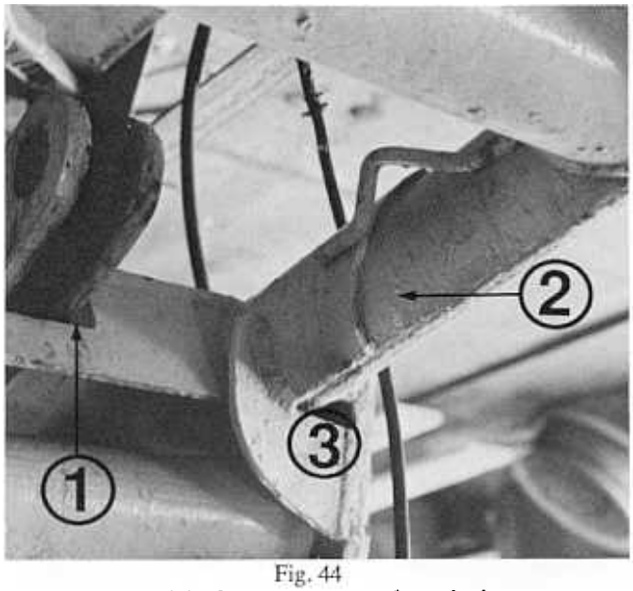

Fig. 44

No. 4 welded crossways, but it is on the inside. I have never seen a failure here.

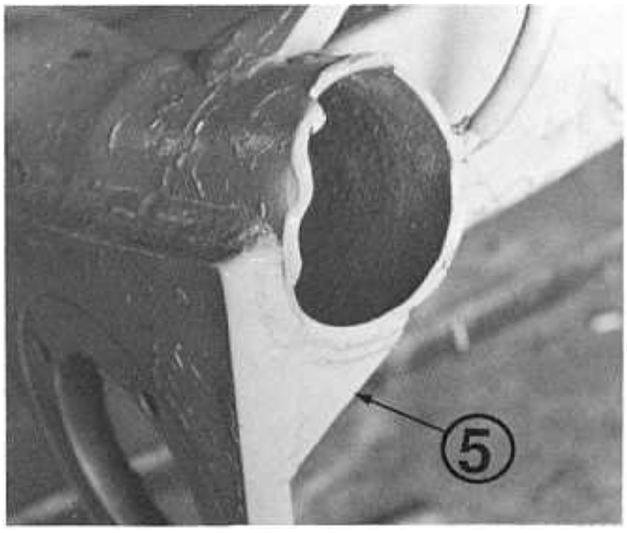

Fig. 45

No. 3 shows that we did not weld a gusset on to the tubing as shown in No. 5.

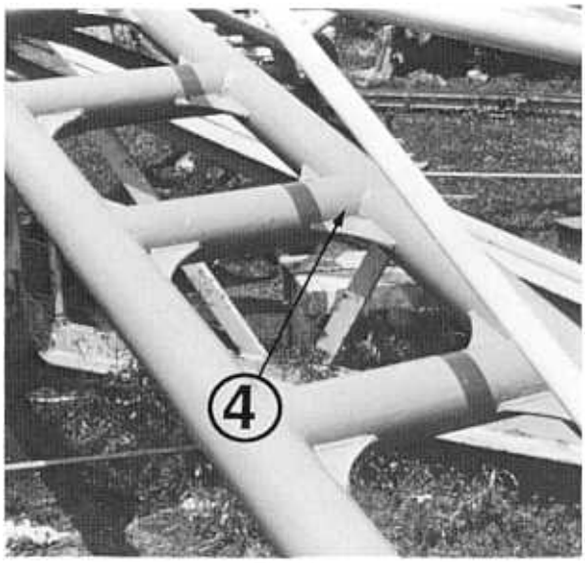

Fig.46

No. 5 weld on gusset and tubing caused the break on both pieces of tubing.

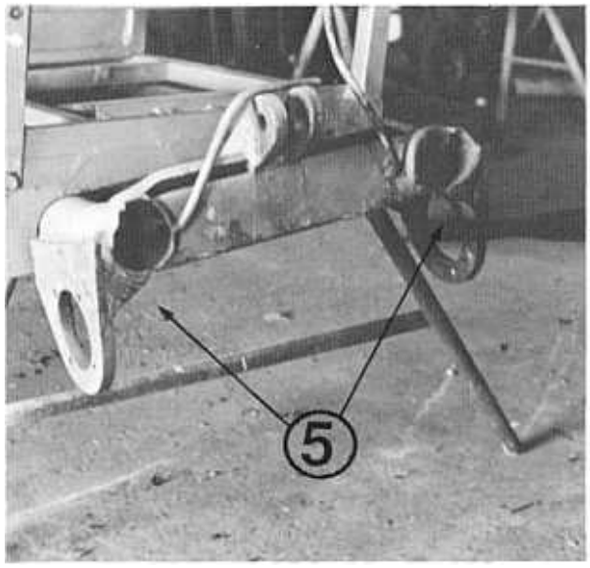

Fig. 47

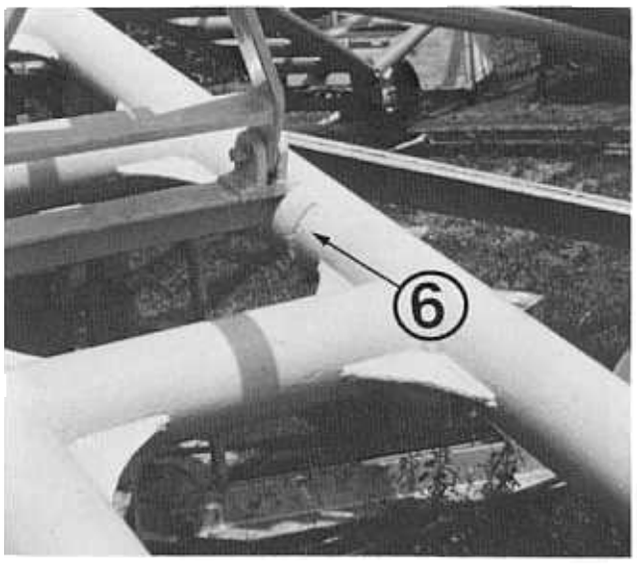

Fig. 48

No. 6 shows T bar welded to patch that is welded lengthways on tubing not welded on ends. On old ride, T bar was welded to tubing and it cracked many times.

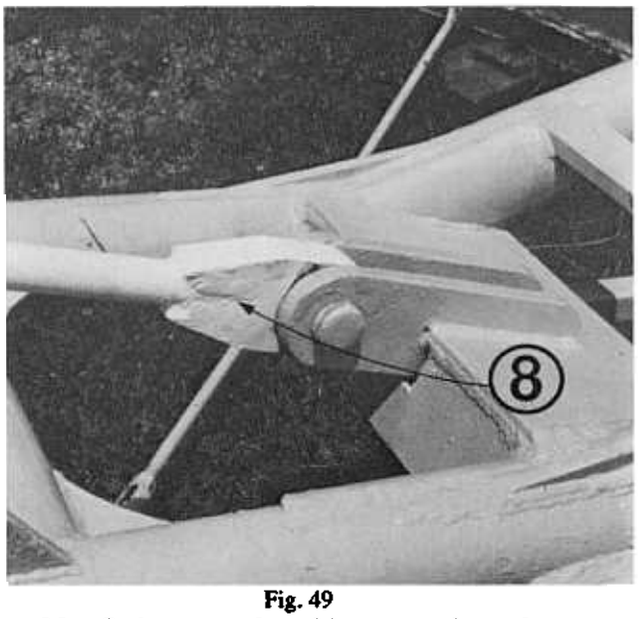

Fig. 49

No. 8 shows rod end is tapered on the end, not welded across the end and the welds are tapered. Even a square shoulder here could cause a break.

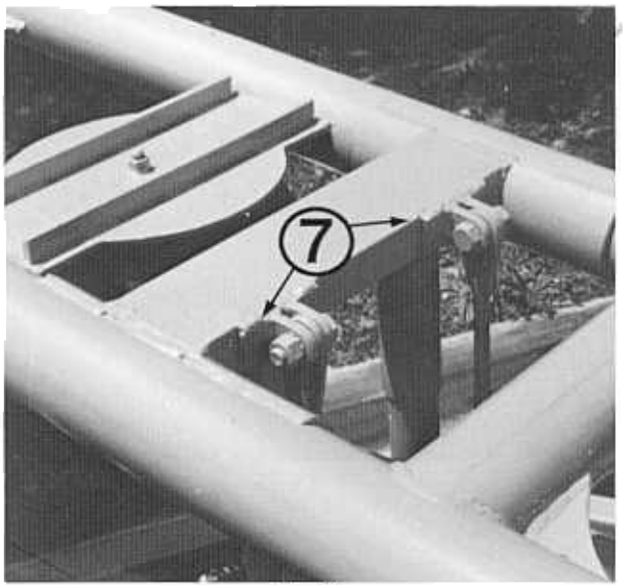

Fig. 50

No. 7 shows short pieces of angle irons welded lengthways with channel iron.

Roller Coaster

Almost every time I saw a certain type of portable steel roller coaster there was a welder on top of it welding something. With a little investigation, I found out why.

The track pieces were angle irons. The cross ties were channel irons. They were fastened together with very short pieces of angle irons. One leg of the angle iron welded flat on top of the channel iron cross tie and the other welded to the side of the angle iron track. The angle irons were not welded lengthways with the angle iron track, but **crosswise**. (See Figure 52) This causes a notch effect. The angle iron track always cracked right next to the weld. They'd weld it. And it would crack again, right next to the new weld. I actually saw one place with 20 or more welds, side by side, overlapping each other.

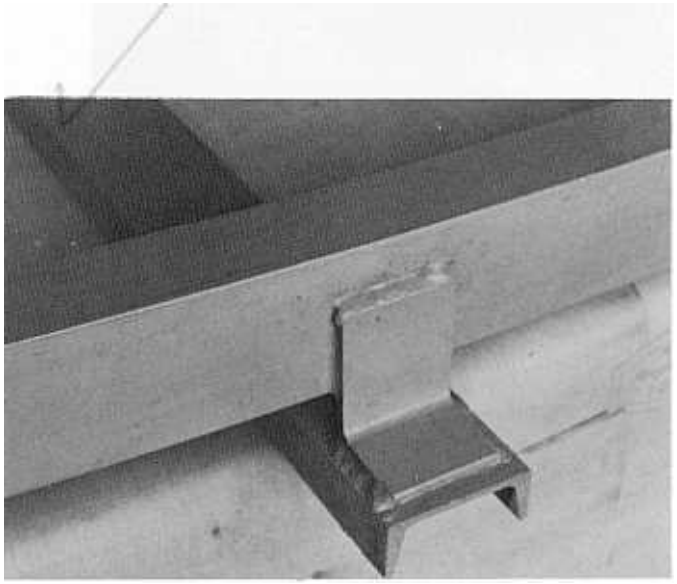

Fig. 51

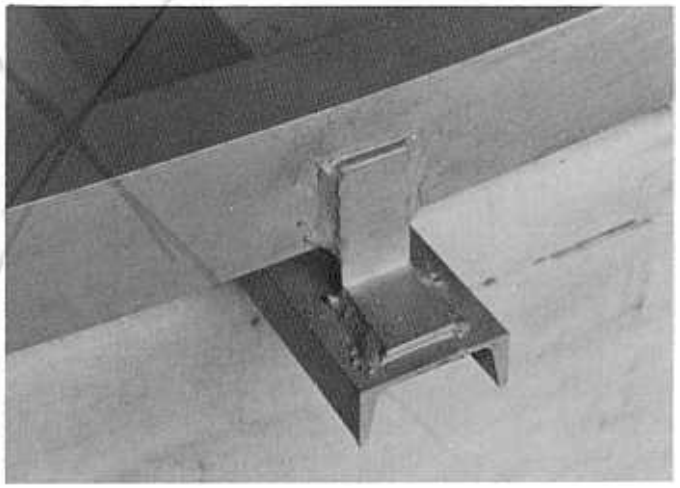

Fig. 52

Cracked Center

This old center for an amusement ride shown below has been abandoned since 1977.

Fig. 53

I remember back in the 1960's, my foreman called me from 600 miles away. He told me the 12" tubing on the center was cracked. I knew that, with all of the gussets and stiffeners that were welded to it, there could be only one reason for it to crack and that is the shrinkage of the welds. I told him to go find a welding shop that could weld it and call me and let me talk to the welder. I asked the welder if he had an air-hammer. He said "I think I can get one." I told him to vee out the crack and then run light stringer beads and peen each bead with the air hammer before running another bead over it. He said "OK." When the ride was brought back to the shop a few weeks later, it had a big beautiful bead, but it was cracked wide open. The welder couldn't get an air hammer. I had to weld it myself. I peened each bead and also peened the top and bottom beads that were also shrinking. It didn't crack again.

Peening Test

If you don't know how much peening should be done to relieve the stress, there is a simple test you can make. Stand a plate straight up on another plate one inch from the edge. Tack it on and square it to a 90 degree. Weld a bead 2 or 3 inches long on the 1 inch side. Place your square on it to see how much the weld pulled it. Then use an air hammer with a blunt chisel and peen the weld and the adjoining metal which got red hot until the vertical plate is back to a 90 degree.

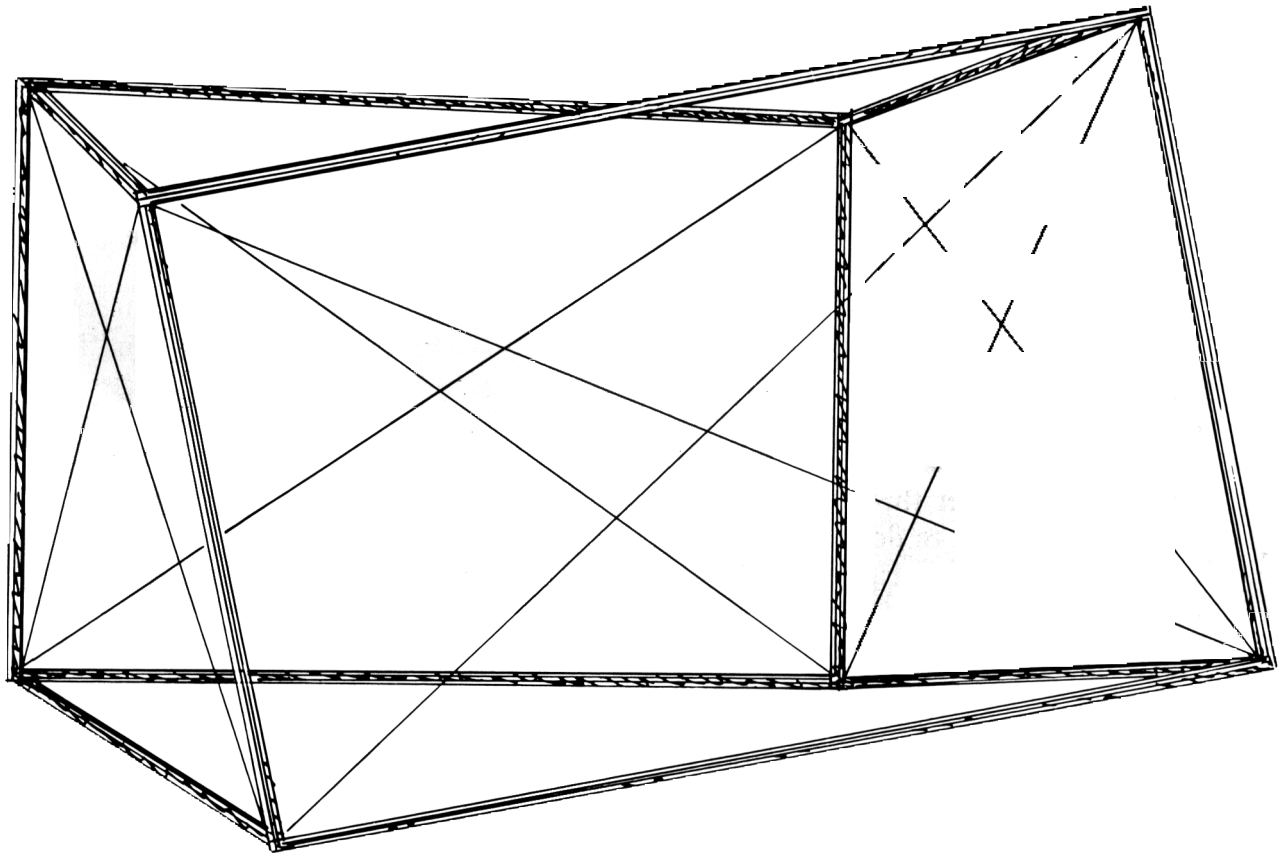

Fig. 54

Cross Brace

A box built of steel, and cross-braced on all four sides will still be flexible. It can be twisted as shown above. To make it rigid, cross-brace the top and bottom also.

For example, twist an open top cardboard box. Two diagonal corners will be farther apart and the other two will be closer together.

Therefore, never cross-brace the top of a truck body. If you do, the cross-brace will tend to lift one wheel off of the ground when it comes to a low place in the road. It could stretch a brace, break something loose or hold the wheel up off of the low place.

Test Welds On Low Carbon Steel

The following subject has been a controversial topic for many years. Whether or not to weld the ends of a reinforcement (or patch).

One writer stated: "Low carbon steel has such a high ductility that it is *not harmed* by being stressed above its yield point. Hence, it is *not weakened* by any stresses that are left in it after welding.

The distortion and the stresses that are caused by welding have been overemphasized. Perhaps this is due to our zeal to make our welds stronger than the parent metal rather than to make them just strong enough to do the job. So frequently after a weld fails, we ask, "Where did it break?" when the question should be, "At what load did it fail?"

We admit that low carbon steel is very ductile, but almost everything has a limit. The above quotation, in my opinion, has helped cause many failures. I have seen hundreds of reinforcements welded on the ends, with a crack adjoining the weld. I would guess that only about 10% of the welders do *not* weld the ends of a reinforcement.

You can see that if you heat a piece of ½" round steel in the center, to a red heat, it will expand in length. If it is placed across the jaws of a vice while heating, the ends of the specimen will be free to move outwardly as it is heated and free to move inwardly as it cools. That is "free expansion". After it cools there will be little if any difference in the length of the specimen. Now open the jaws of the vice and place the ends of the specimen between the jaws and tighten them very lightly, to restrict the expansion.

As the specimen is heated, it expands and puts compressive stress on the specimen. The specimen will compress a little. But when the spot that you are heating gets red hot, it has very little strength. It will push together and cause a bulge. It will be larger in diameter at the place where it was red hot. As it cools, it will fall out of the vice. It is shorter now than it was before.

This is "confined expansion". There is a great difference between "free expansion" and "confined expansion". Some tests were made to show what happens. A 3/16 inch plate was sawed to about a 6"x 6" square. Three overlaping beads, 2 inches long were made in the center of the plate. After they cooled, the welds were ground down flush with the plate, to avoid making a "shoulder effect" which will cause a crack. There were 2 inches on each end of the welds that were not welded.

The cold metal on each end of the welds resisted the expansion and caused the red hot place to be pushed together (up-set). Now, the places which were red hot are thicker than they would have been if the expansion had been "free". As the welds cool, they shrink from all directions, something is forced to stretch. If a weld is not ground down flush, it is thicker than the adjoining metal, a crack usually starts adjoining the weld. If the weld is ground down flush, it usually cracks in the weld.

The 3/16"x 6"x 6" specimen was bent double on about a 3/4" radius. It did *not* crack. See Fig. 55.

Then we sawed a specimen from a 3/8" plate (twice as thick) same size, same welds, same type electrodes and same setting on the machine. We ground it, bent

it to an 81 degree and the welds cracked, as shown in Fig. 56. Of course the metal had to stretch much further on the thicker specimen.

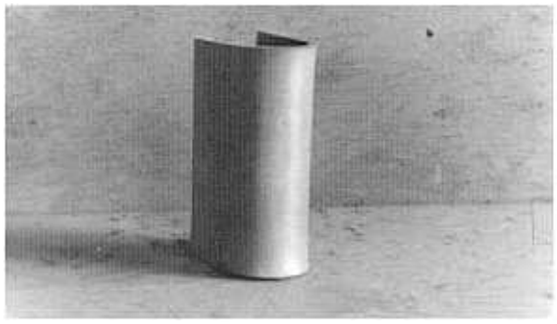

Fig. 55

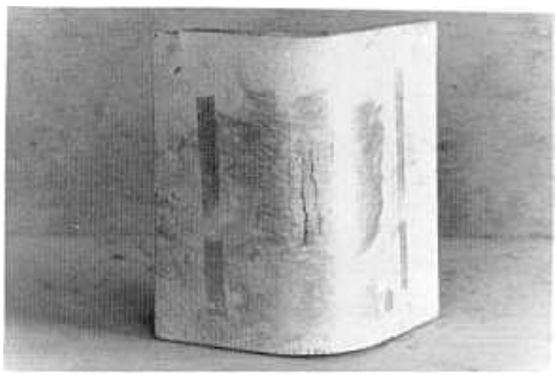

Fig. 56

Then we sawed a specimen 2 inches wide from the 3/8 inch plate. The welds traveled from edge to edge. There was no cold metal on each end of the welds to resist the expansion. Of course, there would be some stress because the welds were not all red hot at the same time. The specimen was ground down and bent double. It did not crack. See Fig. 57.

The specimens in Fig. 56, Fig. 57 and Fig. 58 were all sawed from the same 3/8 inch plate. In Fig. 58 one was sawed across the grain and the other one was

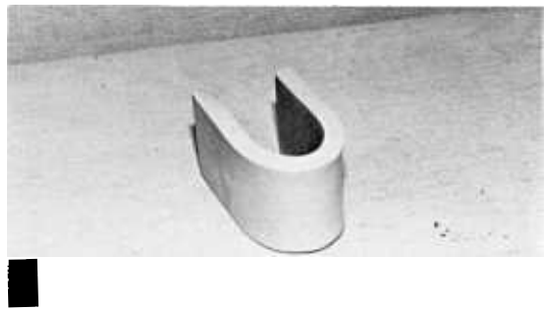

Fig. 57

sawed with the grain, 1/2 inch wide, no welding was done. They were bent double and then pressed flat. Neither of them cracked. That is ductile.

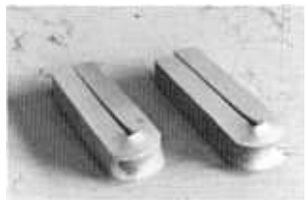

Fig. 58

The specimen in Fig. 59 shows 2 Grade #2 bolts. The one on the left was welded and it cracked. The one on the right was not welded and it did not crack.

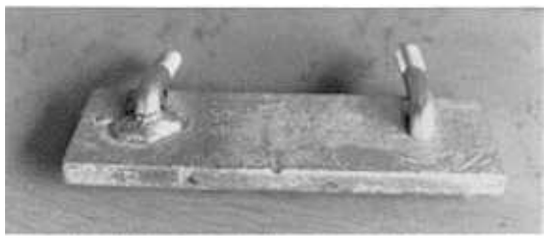

Fig. 59

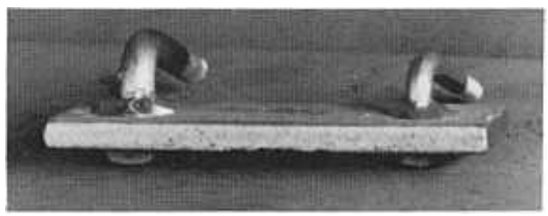

Fig. 60

The bolts in Fig. 60 are ungraded (the square head type). They were both welded. One was a 3/8", the other a 1/2" diameter. They did not crack.

Fig. 61

The 1 inch square tubing in Fig. 61 shows that if you weld across the tubing, it will crack. If you don't weld it crossways, it doesn't crack.

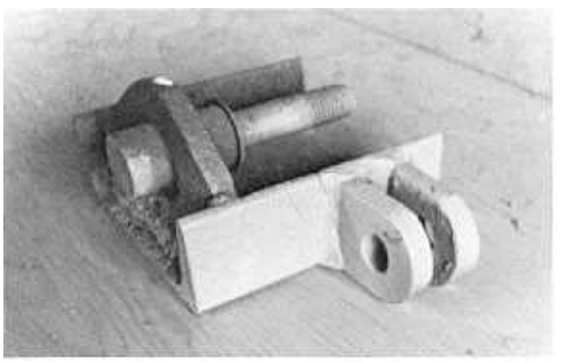

Fig. 62

Fig. 62 shows a crack adjoining a crossweld on a stress member. See a better way on page 26- Fig. 50.

Fig. 63 shows that if you weld around a bolt that comes through a hole, it may break off. Better grades of bolts have markings on top of the head. Grades above one are more likely to break. If you weld the head on the backside, it probably will not affect the bend.

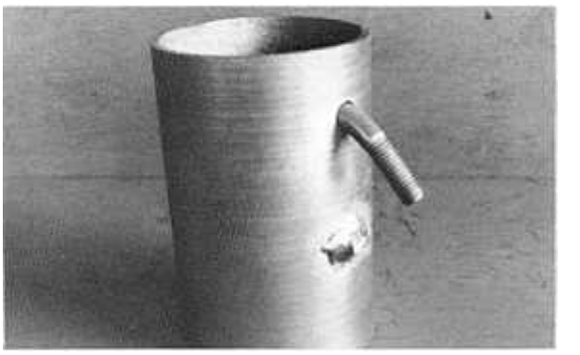

Fig. 63

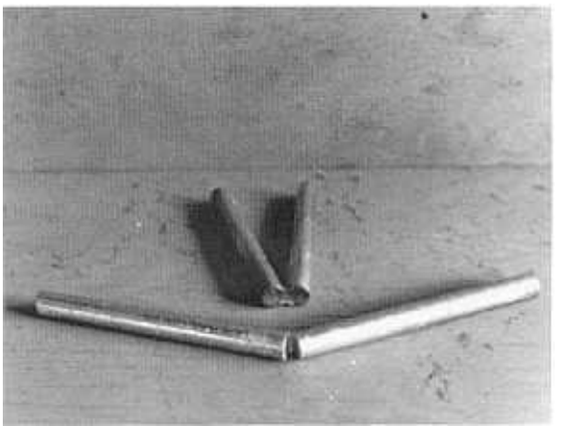

Fig. 64

The above specimens in Fig. 64 were cut from the same 1/2" bar. A stop on the saw was used to cut them the same depth.

The one on the bottom was bent cold to a 38 degree. It cracked because the only place that the metal could stretch was the width of the saw cut at the bottom of the cut.

The specimen on top was bent *cold* also. It was bent cold less than 38 degrees. It had not cracked. Then it was heated red hot and left to cool. When it was cooled to room temperature, it was slightly bent a little more. This procedure was continued until the specimen was bent on a 160 degree. It did not crack. When steel is bent cold, the crystals on the outside stretch. When it is heated to a red heat (non-magnetic), the crystals return to their original shape.

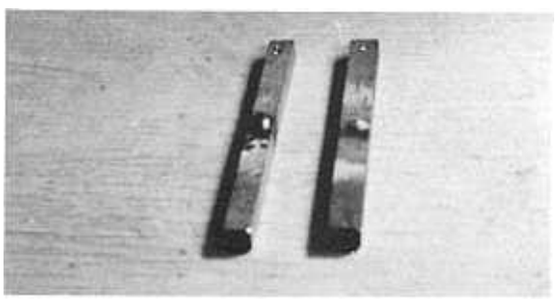

Fig. 65

A very small spot was welded on each of the 3/8" x about 1/2" specimens. The weld area did not get red hot all the way through. The one on the right was heated red hot in the weld area. The one on the left was not. See Fig. 65.

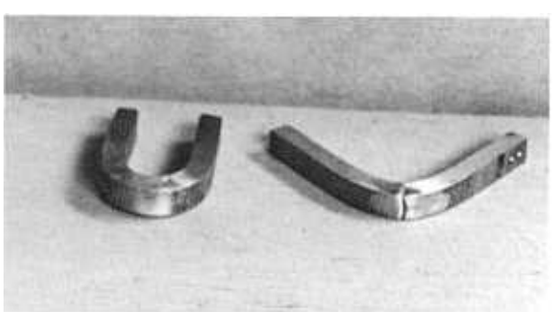

Fig. 66

They were ground and bent. The one that was reheated did *not* crack. The one that was *not reheated* cracked. See Fig. 66.

"Direction of Travel" and "Back Stepping"

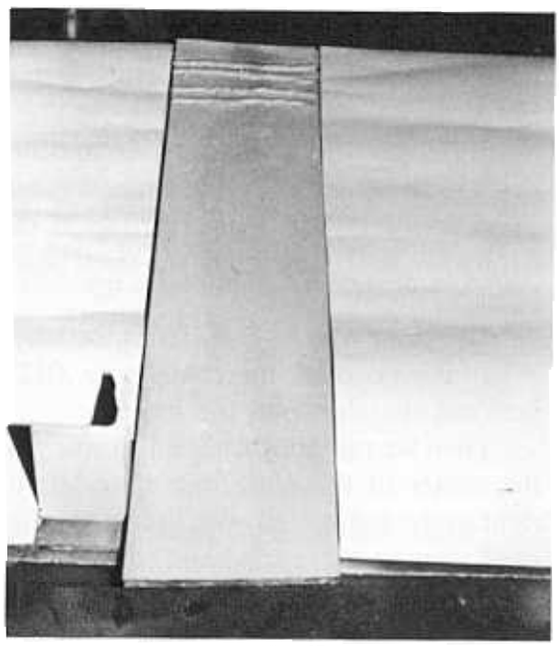

Fig. 67

Fig. 67 shows a piece of scrap steel 1/4"x 4 5/8"x 23" long was tacked (welded) on one end to a steel table. A stop was tacked against one side on the other end to show how much the "Direction of Travel" while welding, effects the other end.

One continuous bead was welded left to right across the end that was welded to the table, 1 inch from the welds that held it to the table. The stop was 19 5/16" from the cross weld. When the weld cooled, there was a space .100" or 1/10 of an inch wide between the stop and the 1/4" plate. The 1/4" plate was not cut in two. The welds were made on a solid 1/4" plate.

The stop was knocked off and again placed against the side of the 1/4" plate and tacked.

To show the results of "Back-Stepping", 4 short beads were "back-stepped" across the plate. That is, start about 1 5/32" from the right hand edge of the plate. Travel left to right to the right hand edge. Then start at about the center of the plate. Travel left to right to the first bead, fusing them together. Then another short bead running into the second bead. Then another short bead starting at the left hand edge running into the third bead. (4 back steps).

The bead was 18 1/4" from the stop. When it all cooled, there was only .012" between the stop and the plate.

Then we ran another bead, starting in the center of the plate, traveling left to right to the right hand edge. Then another bead starting at the left hand edge, traveling left to right to the center, meeting the first bead. (2 back-steps) This bead was 17 1/2" from the stop. It moved away from the stop .021". This is only 9 thousandths of an inch more than the 4 back-steps.

On the next bead, we traveled in both directions, starting at the right hand edge traveling right to left, stopping a little short of the center. (The first bead usually pulls more than the second bead.) Then starting at the left hand edge, traveling left to right, meeting the first bead in the center. This bead was 19 1/2" from the stop. It moved away from the stop .002" (2 thousandths of an inch).

I would not recommend starting in the center and traveling to the outside edges. The center would be under compressive stress, and the edges would be under tensional stress. I think it would be more likely that a crack would start on an edge that was under tension stress.

SUMMARY

	Moved	Distance from Stop
Continuous bead L to R	.100"	19 5/16"
Back-stepped 4 times	.012"	18 1/4"
Back-stepped 2 times	.021"	17 1/2"
2 beads R to L to center and L to R to center	.002"	19 1/2"

Why a Weld Bead Shrinks More Where You Stop Welding Than It Does Where You Start

The weld bead and surrounding metal behind the weld crater are hot. The metal in front of the weld crater is cold. This creates a bending force on the workpiece. The cold metal in front of the weld bead is under a compressive stress, (pushing together). As the weld crater moves forward and melts the compressed metal, the weld crater and the red hot metal around it are relaxed. The crystals are rearranged which makes the weld thicker at that point. This is "confined expansion". As soon as the crater solidifies, it starts shrinking which puts **more** compression on the cold metal in front.

Confined Expansion

A condition which resists the free expansion of a weld bead while welding. Now think back to the time that you made your test welds, were the ends of the specimen free to move in and out? Did you weld from edge to edge on the specimen? Did the speicmen get red hot all the way through? If the answers are "yes" that is "free expansion". The specimen can be bent without cracking.

If you weld a crack on an I beam, angle iron, channel iron, etc. the cold metal partly surrounding the weld will resist the expansion. This is "confined expansion".

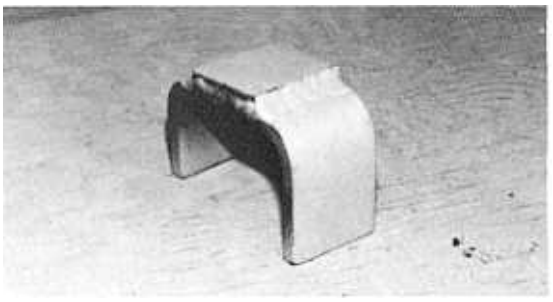

Fig. 68

The specimen in Fig. 68 shows that if it is not too thick, and welded from edge to edge, it can be bent without cracking. It is free to expand in all directions.

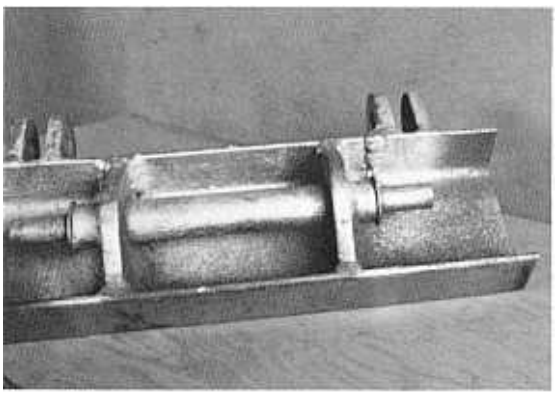

Fig. 69

The specimen in Fig. 69 shows 2 cracks on rewelds, that were welded crossways on a stress member.

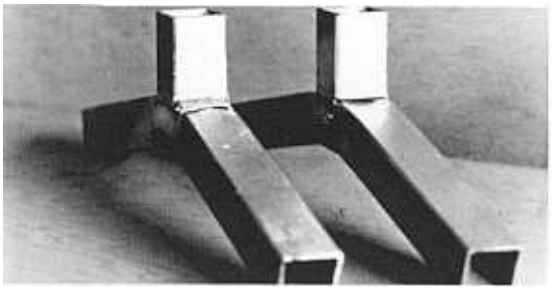

Fig. 70

Fig. 70 shows the same thing. If a crossweld is made on a 1 inch square tubing, it will crack.

The Exception to the Rule

Fig. 71

The test weld in Fig. 71 above was made to show the exception to the rule. This test weld was made with no restraints. The ends were free to move outwardly as it was heated, and free to move inwardly as it cooled. The welds are not under as much tensional stress as they would have been if the ends had been welded to a heavy piece or if the patch had been welded in the center of a much larger plate. This piece can be bent without causing a crack adjoining the weld. However, I can't say what would happen under a work-load.

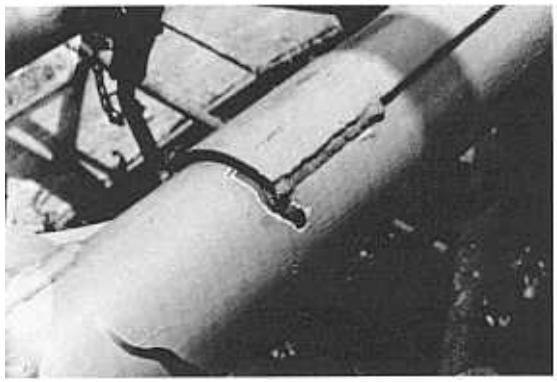

Fig. 72

The above photograph shows a crack (with magnetic particles, see Fig. 72) at the end of a 3 inch weld. The reinforcement is cut on a 90 degree angle. The reinforcement is too wide. It should have been tapered on both sides to about 1 1/2 or 2 inches wide. The 1 1/2 inch width should have been about 3 or 4 inches long and then cut on a 45 degree angle. When a reinforcement wraps half way around a piece of pipe, it makes it rigid. It should not stop abruptly. It causes a shoulder effect which can cause a crack. The shoulder effect and the shrinkage of the weld caused this crack. Both sides of the reinforcement should have been "skip-welded". This means to weld one inch and skip one inch, weld one inch and skip one inch. Repeat all the way on both sides, not the ends.

Magnaflux Testing For Cracks

Magnaflux is a brand name for an Electro magnetic device that assists in finding a crack that cannot be seen with the naked eye. It works like this — place the magnetic poles on each side of the suspected crack. Turn the electric on and spray powdered iron over the surface. The iron dust will stick to the crack. Gently blow the excess away. The magnetic particles will show the crack. **CAUTION:** Do not breathe the iron particles! If you don't have the magnetic device, a large horseshoe magnet will do. If you don't have iron powder, the sawdust from under a metal (steel) saw will do. You may have to sift it to get the smaller pieces. This method also works on a weld that does not go all the way through the work. That is, if you weld over the top of a crack and leave part of the crack below the weld, the dust will stick to the weld bead.

Another way to test for a crack is to clean the surface, pour kerosene over the surface, wipe dry with a clean rag, and then cover with blackboard chalk. The crack will fill with kerosene and when the chalk is applied, the kerosene will wet the chalk to show the crack. You can also buy a kit that works the same way. The kit contains a red liquid dye, and the chalk is brushed on in liquid form.

Water Tank

I remember, back in about 1948, a friend of mine insisted that I help him straighten out a tank that had been screwed up by two beginners. The tank was about 18 or 20 feet in diameter and about 10 or 12 feet high. The 1/4 inch plates had already been rolled to the correct radius. (As you know, the ends of a rolled plate are straight for the distance between the centers on the rolls. They should have been sheared off.) But the beginners had already welded together the bottom of the tank and the first layer of the vertical plates. When they started with the second

layer of the plates, (the seams were staggered) the flat places on the ends of the bottom plates took a short cut across the radius of the top plate and left a wide gap horizontally between the top and bottom plates. With a sledge hammer, we drove the seams out to meet the radius on the top plate.

The beginners had already welded some of the second layer plates solidly, except the gap. I remember one place where they had tacked a plate in place and then welded the end vertically from bottom to top, to the end of the other plate. The vertical weld pulled together so much at the top that it broke the tack on the other end and lifted the plate up on the other end, making a gap too large to be welded. (See "Direction Of Travel", pg. 33, fig. 67.) I started welding horizontally from the bottom of their vertical weld toward the loose end. I would weld a couple of inches, then wait for it to cool. (The plate was 10 or 12 feet long.) Weld a couple more inches and wait for it to cool. This method was continued until the short welds pulled the other end back down to the proper place. The end was tacked again and the rest of the weld was "back-stepped" (See pg. 33, fig. 67.) to reduce some of the stresses.

If you have a long bead to run, it is best to back-step it. Back-stepping will not relieve all of the stress, but it will help.

Back-Stepping

Start a bead 6" from the edge of a plate. Travel to the edge of the plate. Start another bead 12" from the edge and weld toward your first bead and into it. Then back up 6" from where you started the last time, each time.

Broken Trailer Frame

I had a truck driver who had a trailer frame welded in Canfield, Ohio. He drove it to Knoxville, Tennessee, and it was broken when he arrived there. When he arrived in Chattanooga, it was broken again. Both welders had welded the ends of the reinforcements. Both breaks were adjoining the end welds. The reinforcements were removed and replaced with a longer one. The ends were *not* welded. It did *not* break again.

Most people agree that: As it cools, every weld shrinks from all directions. When a weld shrinks, the adjoining metal is forced to stretch or crack. If the metal is stretched, it is thinner. Is thin metal as strong as thicker metal?

Some welding instructors teach the beginners to weld crossways on a stress member. They bend a test weld and it doesn't break. This is explained in "The Exception To The Rule" on page 35, fig. 71.

Air Lifts

Years ago, I had some air-lifts installed on my car. They were rubber bags that fit on the inside of the coil springs on the rear of the car. The rear end could be overloaded and then add air pressure to the bags to raise the car up. They worked really well. There was a rubber hose from the bags to the trunk of the car. The air pressure could be adjusted without going under the car. Then one day, one of the bags went flat. It was taken off and cut open to see what caused the crack. There was a reinforcement molded to fit the inside and probably vulcanized to it. The reinforcement was about 1/8 of an inch thick and where it stopped on the sidewall, it made a sharp shoulder. That is where it cracked. If a shoulder will cause a crack on rubber, what will it do to steel?

Subjects of Non-Welding

The following subjects do not pertain to welding. However, a good welder is expected to perform many tasks other than welding. Welding is not a trade, but a sideline of many trades, such as: machinist, iron worker, carpenter, electrician, pipe fitter, etc.

We hope some of you will benefit from this effort. At least, some of them are interesting.

Cover A Padded Bar With Upholstery

This method is for problem applications, where you have a thousand wrestlers a day trying to twist them off. Use a **sponge** rubber pad that is tough (foam will disintegrate). The holes in foam rubber are connected. You can blow smoke through it, but not in sponge.

Glue the pad on to the bar solidly. Leave a 1/4" gap where the sides of pad come together. Try to place the gap out of sight. The gap will contain the loops and wires, as you will see later.

The padding should be wrapped with strips of cloth to hold it in place until the glue dries. When the glue dries, remove the cloth and determine how wide the cover should be to fit around the padding correctly. Cut a strip of your material 1 inch wide and 2 inches or more longer than the circumference around the padding. Fold 1 inch back and staple it with a paper stapler. Wrap it around the padding with the other end folded back loosely. Bring them together at the gap in the padding. Slip the loose end until the loops or folds touch. This measurement is important. If you like, you may try a piece of scrap material first.

With the material cut to exact size, and sewed lapped over 1 inch on the sides and 1/2 inch on the ends, as shown below fold together, place one end of both sides in the vice and cut both pieces 1/4 inch deep every 1/2 inch, as shown above.

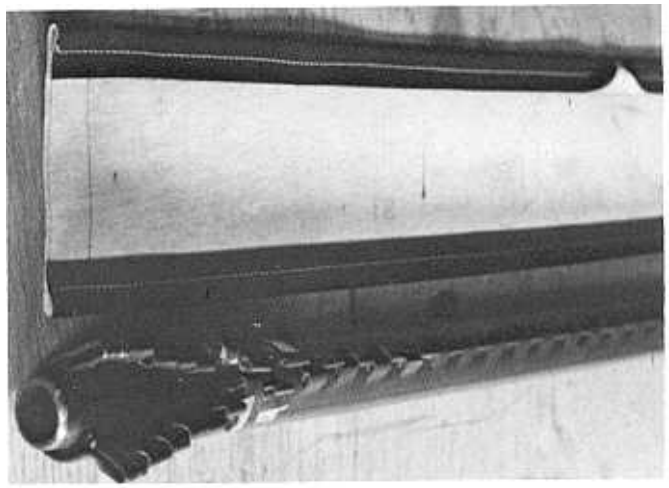

Fig. 73

Apply glue to padded bar, all except 1/2 inch on each side of the gap. If you get glue on the loops, you may not be able to push the wires through. In applying the glue to the cover, raise the flap on the seam up and brush the glue all the way back to the threads, then when your threads rot out, the glue will still be holding.

In placing the cover on the padded bar, line the loops up so that the wire which will be on the inside of the loops will be centered over the gap in the padding.

With the cover glued on, take a piece of 3/32 acetylene welding rod and file one end smooth and bend it 1/2 inch from the end about 1/8 inch off center. Then bend about 2 inches on the other end on a 90 degree for a handle. Push the welding rod through the first loop on the left hand side, then through the second loop on the right hand side, third loop on left, etc.

When you push the wire through one loop, turn the wire about a half turn so that the bent end points to the next loop on the other side. It will start through the loop much easier, and it will pull the cover tight. Push the wire all the way through and out the other end. Then push another welding rod through the remaining loops the same way.

On the ends, punch a hole in the cover so you can run a wire around the ends on the inside of the hem. Now go back to the wires that go through the loops. Cut the wires off and file the sharp edges off and with a pair of needle nose pliers, bend the wires and tuck them under the padding. Then twist the wires together that go around the ends, cut them off, file the sharp edges off and tuck them under the padding. Don't twist them too tight. Leave room to tuck them under.

It is important that you cut the material lengthways with the roll instead of crossways — it will stretch more crossways.

When you cut the material, cut it 1 inch longer and 2 inches wider to allow for the hems. Make a straight line 2 inches in from the sides and 1 inch in from the ends. All lines and cuts should be straight. This gives the one who does the sewing something to lap the edge of the material to, which makes the width correct.

Hooks Or Eyes On Spring

To make the hooks or eyes on an expansion spring, place spring in vise vertically, with one round of the wire above the jaws. Use a pair of duck bill pliers to twist the top round straight up. You may have to hold the second round down with a screwdriver. Now the hook is standing straight up and you have not bent the wire. You have twisted it a quarter turn as shown on one end below and your hook is on one side of the

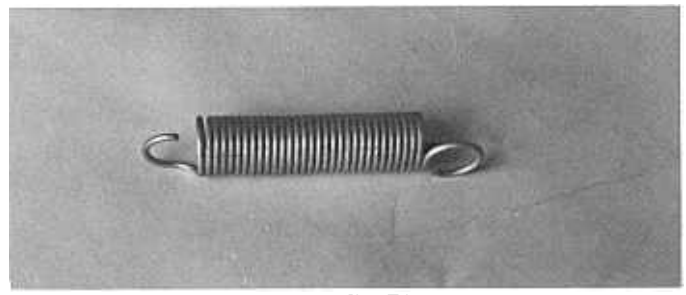

Fig. 74

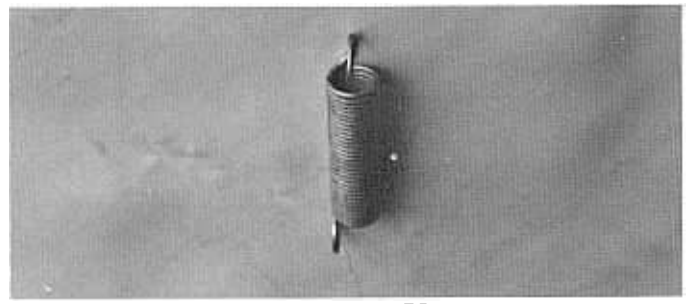

Fig. 75

spring. To center the hook, twist it one quarter turn to the center of the spring as shown above. This method does not bend the spring steel wire, it only twists, it. You can also use this method to replace a broken off hook or eye.

Make A Coil Spring

In an emergency, you can make a coil spring. You will need some spring steel wire of the correct size. You will need a mandrel the right size.

Drill a hole slightly larger than your wire, 1/4 inch from the end. Then saw a slot with a hacksaw, slightly larger than your wire, a few degrees clockwise of your hole as shown below.

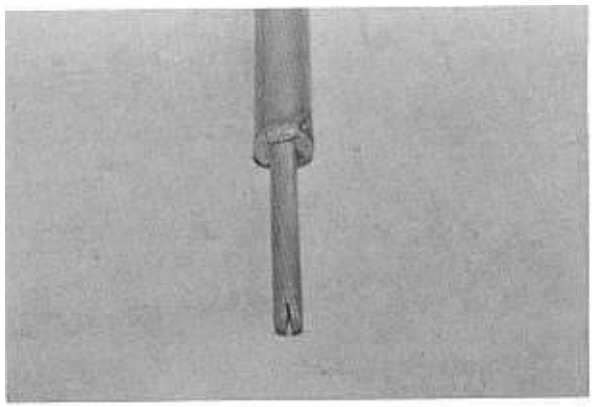

Fig. 76

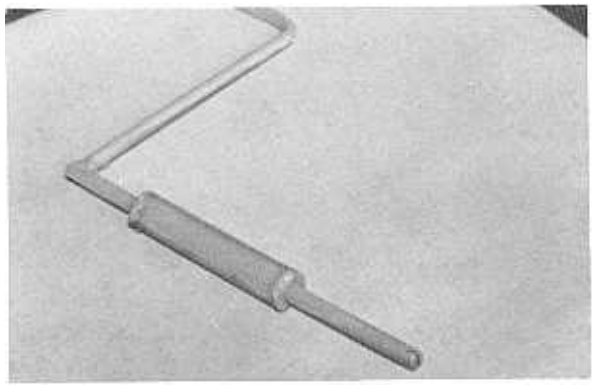

Fig. 77

If you don't have a lathe, cut a piece of pipe 2 inches longer than the jaws of your vise. Weld a washer on each end of the pipe with holes drilled in the washers to fit the mandrel. Weld a handle on the other end of your mandrel to turn it with as shown above. You will need someone else to turn the crank with the pipe held in the vise. Place the end of your spring wire in the slot and bend it. With the mandrel turning, pull tightly on the wire. For an expansion spring, wind the wire close together. For a compression spring, lead the wire away from the last round so the spring will have room to compress. When you finish, turn the crank the other way to relieve the tension. The spring will be much larger in diameter than the mandrel.

Cooling Fin

I was once told by an engineer that a 7" aluminum cooling fin placed on a 4" steel hydraulic cylinder, would contract when heated, which would make it tighter on the cylinder. I didn't believe it. Aluminum expands more than steel. Then I asked another engineer. He agreed with the first one. He explained that it would expand both ways from a point slightly to the inside of the center between the two circumferences, and it would tighten on the cylinder. I still didn't believe it, so I pressed one fin onto a cylinder. With the cylinder standing vertically, heat was added to the inside of the cylinder. When the fin got hot, it dropped. I think the aluminum expanded in **all** directions, causing the hole to be larger as well as the outside circumference. After all, the sides of the fin are only 1 1/2 inches and the circumference is over 21 inches.

Shaft Will Break At Shoulder

Figure 78 shows a broken shaft at the end of a rod end. It broke for two reasons: partly because the threads are the weakest place and partly because the rod end was not tapered enough to give. It broke off right at the end of the lock nut. That is a notch effect.

The same thing happens when you saw a half-inch rod with a hacksaw, leaving only 1/8 inch holding. Bend it and it breaks off. That is because the metal can only stretch in one place, the width of your saw blade. Take a 1/8 inch welding rod and try to break it. It bends in a much larger radius.

Years ago electric irons came with a drop cord the same size all the way. With

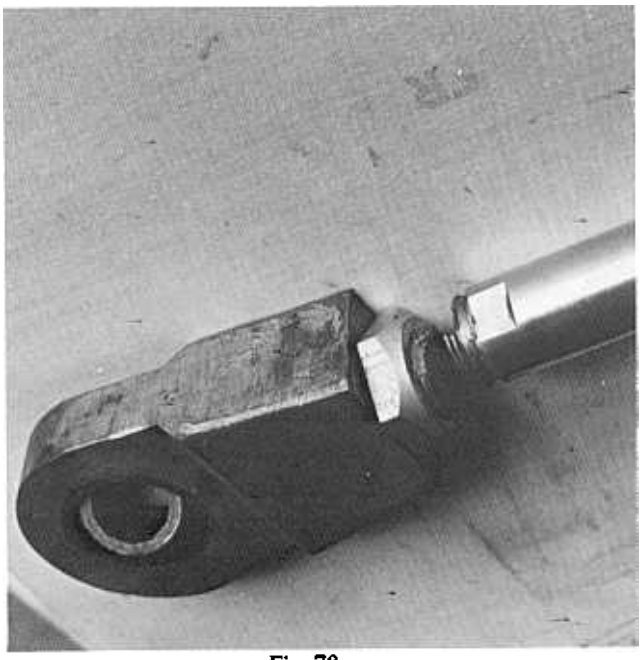

Fig. 78

all the flex required for ironing clothes, the cords inevitably broke off right at the plug end that attaches to the iron. A heavy rubber sheath, tapered out six or eight inches caused the cord to bend in a much larger radius and the breaking stopped.

The same principle applies in both cases.

In 20 years, I had 3 rods to break as shown in Fig. 78 (Shaft will break at shoulder).

Then I made new ones, 1/8 of an inch in diameter larger. To avoid cutting as deep into the shaft, I used 18 threads per inch instead of 12. At the end of the threads, the shaft was turned down about 1/2 inch in length to the size of the bottom of the threads to let it bend in a larger radius. Then it was filed, sanded and polished to remove any cross scratches that might start to crack. They have been in service about 6 years with no failures.

How To Temper Small Parts

Using water hardening drill rod, cut to desired lengths. Heat them to a red heat and plunge them into water. This makes them hard and brittle. Take them out of the water and sand the oxide off so you may be able to see the heat colors. Heat a piece of heavier steel to a red heat, and place the parts on top of it. Watch the heat colors. If you quench them when they first turn dark blue, they will be hard enough to break. If you want them to be soft enough to bend, wait a few seconds until they turn light blue then quench them.

Rod Ends

If you want to pin a threaded rod end to a rod to prevent it from turning, and the rod-end and pin still be interchangeable with the other rods, here's how. Weld a piece of steel onto a nut to hold a hardened drill bushing. The hole for the pin should be near the end of the rod. Screw the nut onto the rod until it is flush with the end of the rod. Then place the rod-end against the rod and start turning the rod-end, nut and drill bushing all together. There may be a space between the nut and the rod-end. If the space is wider than the distance between two threads, take it off. You held the rod-end crooked and didn't catch the first thread. Screw the rod-end on to the desired depth and drill hole through drill bushing into and out the other side of the rod-end and rod.

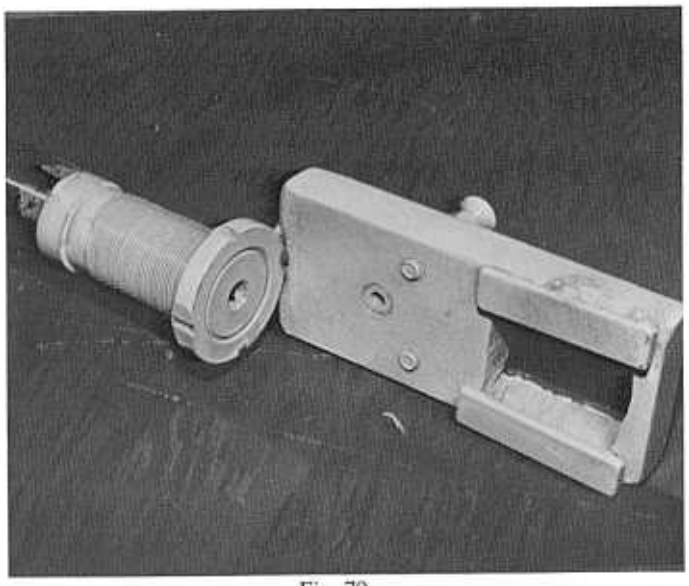

Fig. 79

It is important to have a lock-nut **tightly** against the rod end. I had a nut get loose and wore the threads on the rod-end too much. I shrank the rod-end at the threads (as explained in "How To Shrink A Steel Pulley") a little too much. I had to tap it out to get it on.

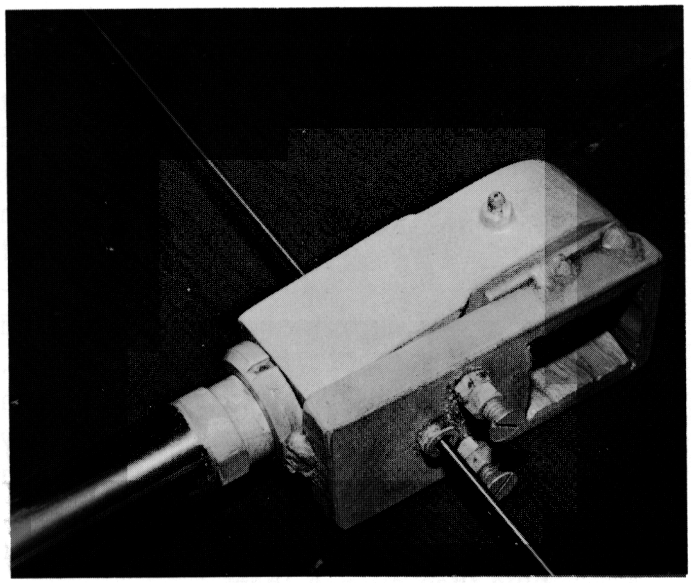

Fig. 80

Find Diameter of Numbered Machine Screws

Multiply number of screw by .013 then add .060.

Example: The screws that hold the cover plates on your electric wall switches at home are 6-32.

$$\begin{array}{r} .013 \\ \times 6 \\ \hline .078 \\ +\,.060 \\ \hline .138 \end{array}$$

Diameter of screw is 138 thousandths of an inch

Which Way To Turn A Turnbuckle

Look at the two threads on this turnbuckle. If the threads on one of them look like they are slanted more than the other ones, they are the left-handed threads. Facing either end of the bolt, turn a left-handed nut counter clockwise to move it away from you. Turn a right-handed nut clockwise to move it away from you.

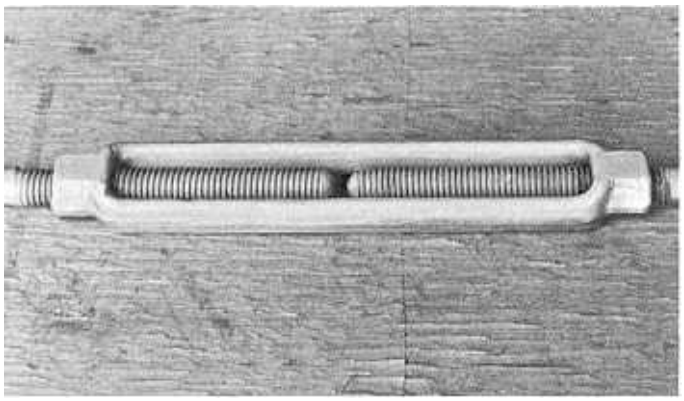

Fig. 81

Dish In Wheel

Almost every wheel has a dish in it. If the hub was not off-center to the rim, it would be like a square tank to hold pressure. A flat surface bends more than a curved surface when pressure is applied. A welder once asked me if I could build a square hot water tank. I said, "yes, but it would take a lot of stay-bolts to prevent the sides from bulging." He said he built one and the next morning it was almost round. Same thing goes for a retaining wall. A straight wall sometimes falls over or breaks. A curved or zig-zag wall is much stronger.

Try to stand a flat card on edge. Then bend it and see the difference.

To make a flat frame rigid, cross-brace from corner to corner. Where the braces cross in the center, they must be bent to make that point higher than the frame. The cross must be bolted or welded to prevent sliding.

Blowing Grip On Bar

The picture below shows a steel bar 3/4" in diameter. The plastic grip has a 5/8" hole in it. The grip was forced on with air pressure. Notice that the grip is not on all the way. The end of the air nozzle hit the end of the bar and stopped. This method may be used to cover a pipe with rubber or plastic tubing. The end of the pipe must be sealed and one end of the rubber tubing fastened to the air nozzle air tight. The rubber tubing will slide over the pipe on a layer of escaping air.

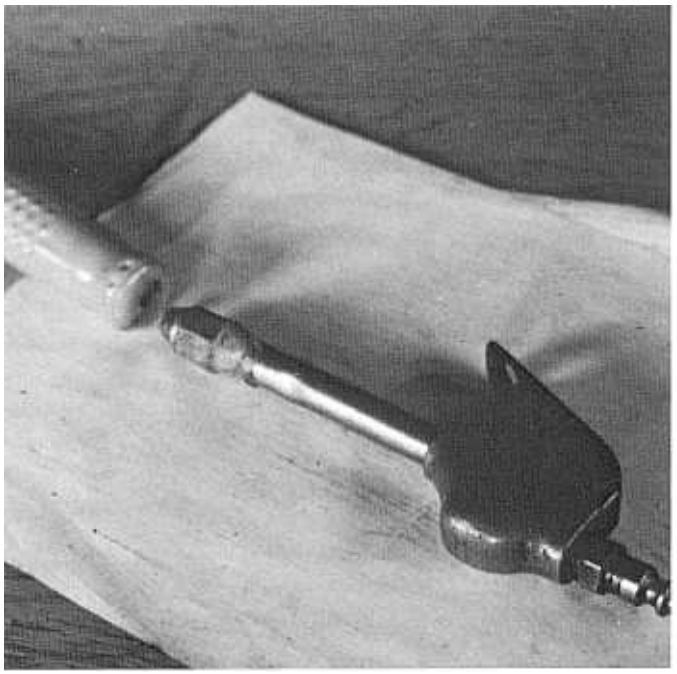

Fig. 82

How To Find A Tap Drill Size Without A Chart

Example: Find a drill size for 1/2 inch 13 thread tap. Divide number of threads per inch into one inch.

```
        .0769
13 ) 1.000
      91
      90
      78
      120
      117
```

Then subtract .0769 from tap size — $\frac{1}{2''}$ = .500

```
.500
.0769
.4231 = drill size
```

$\frac{1}{64''}$ = .0156 — then divide .0156 into .4231

```
          27
.0156 ) .4231
        312
        1111
        1092
          19
```

$\frac{27}{64}$

OR

```
.4231
   64
16924
25386
27.0784
```

$\frac{27}{64''}$

In other words: divide the number of threads per inch into one inch and subtract the answer from the tap size. Then find the nearest standard drill size. A few thousandths does not matter.

Finding Half Of A Mixed Common Fraction

Sometimes it is necessary to find the center of something. That is — find half of the total length.

Example: Whole number **ODD**

$$5\,\frac{1}{2} \div 2 = 2\,\frac{3}{4}$$

DISCARD the odd whole number making the 5 a 4. Divide 4 by 2 and place here.

Multiply the denominator by 2 and place here.

Add the numerator and the denominator together and place here.

Example: Whole number **EVEN**

Divide the whole number by 2 and place here

$$4\,\frac{1}{2} \div 2 = 2\,\frac{1}{4}$$

Multiply denominator by 2 and place here.

MOVE numerator to here.

In other words: Take the odd whole number and throw it away — forget about it. That changes the 5 into a 4. Take half of the 4 and write it down, 2 - with a dash to the right of it. Then add the top number and the bottom number of the fraction together and place it above the dash $2\,\frac{3}{}$, then double the bottom number of the fraction and place it below the dash this way: $2\,\frac{3}{4}$.

$$37\,\frac{21}{32} \div 2 = 18\,\frac{53}{64}$$

If someone tells you something that you don't already know, don't even think about making a defensive statement. You are not supposed to know everything. Nobody else does. Check it out and see Who Is Right.

However, you can't believe everthing you see, hear, or read.

Facts

Nickel is magnetic.

U.S. Paper Currency is slightly magnetic in places. When you turn the steering wheel on your car to the left, the left front wheel turns more than the right one does, and vice versa.

The file and scratch test is a quick procedure for checking hardness. Shopmen frequently use this method to predict the machinability of a metal. It consists of simply trying to scratch or cut the surface of a metal with a file or pointed object of known hardness. All scratch tests reveal only a superficial or outer skin hardness. They tell us nothing about the hardness ⅛ in. below the surface.

Table 3-2 gives a relation between Brinell hardness and the hardness as estimated with a machinist's new hand file.

Table 3-2. File Hardness Test Data

BRINELL HARDNESS	FILE ACTION
100	File bites into surface very easily
200	File removes metal with slightly more pressure
300	Metal exhibits its first real resistance to the file
400	File removes metal with difficulty
500	File just barely removes metal
600	File slides over surface without removing metal. File teeth are dulled.

Simple Hardness Tester

Welders sometimes make their own set of hardness testers, by heat treating ¼-in. high carbon steel electrodes to give a graduated series of hardnesses.The rods are labeled with their respective Rockwell C hardnesses. A sharp point is then ground on the end of each rod, taking care not to soften the steel by overheating.

By drawing these scratch testers, one at a time, over a smooth metal surface, the welder can determine which rod has a hardness about the same as the metal being scratched. If the sharp point of the scratch tester digs in and leaves a scratch, its hardness is greater than that of the metal being tested. If the point only slides over the surface without scratching, the scratch tester is softer than the metal tested.

The hardness set is inexpensive and is convenient to use in hard-to-reach places, where a conventional test would be impossible.

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Fusibility: The Ease of Melting

Fusibility is a measure of the ease of melting. Mercury (the metal with the lowest melting point) melts at -38 degrees Fahrenheit, while tungsten, which has the highest melting point, melts at 6,100 F.

A pure metal has a definite melting point, which is the same temperature as its freezing point. Alloys and mixtures of metals, however, have a temperature at which melting starts and a higher temperature at which the melting is complete.

Figure 3-9 gives the melting point of a few metals and other temperatures of interest.

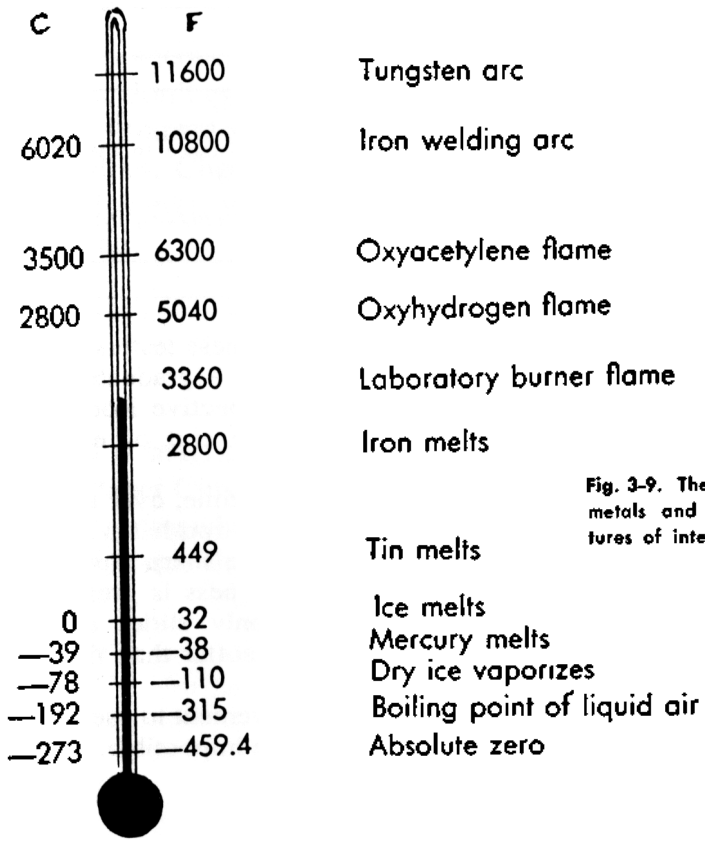

Fig. 3-9. The melting points of some metals and a few other temperatures of interest for comparison.

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Heat of Fusion

The *heat of fusion* is the quantity of heat necessary to change one pound of a solid material to a liquid without temperature change.

The British thermal unit (Btu) is used to measure the quantity of heat; for all practical purposes, it is the amount of heat required to raise the temperature of 1 pound of water 1 degree Fahrenheit.

The heat of fusion of ice is 144 Btu per lb. In comparison, here are the heats of fusion of a few metals:

Table 3-3. Heat of Fusion of Metals

Aluminum	170 Btu per lb	Tungsten	79 Btu per lb
Magnesium	160 " " "	Silver	45 " " "
Chromium	136 " " "	Zinc	43 " " "
Nickel	133 " " "	Gold	29 " " "
Molybdenum	126 " " "	Tin	26 " " "
Iron	117 " " "	Lead	11 " " "
Manganese	115 " " "	Mercury	5 " " "
Copper	91 " " "		

More heat is required to melt a pound of ice than is needed to melt a pound of iron (provided of course that both materials are at their melting point). The figures are 144 Btu for water and 117 Btu for iron. It takes more heat, however, to melt a pound of aluminum than a pound of ice—170 Btu as against 144 Btu.

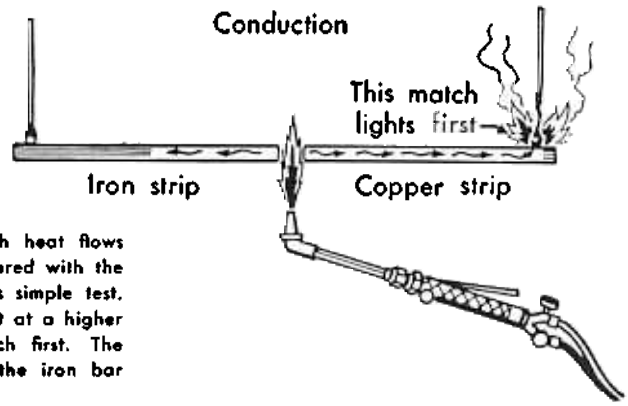

Fig. 3-10. The rate at which heat flows through copper can be compared with the same property in iron by this simple test. The metal which conducts heat at a higher rate (copper) lights the match first. The match at the outer end of the iron bar will burst into flame later.

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Thermal Conductivity

Thermal conductivity is a measure of the rate at which heat will flow through a material. The difference in thermal conductivity between iron and copper is easily demonstrated, Fig. 3-10. The copper conducts heat much faster than does the iron.

If one end of a copper bar is kept in boiling water (212 F) and the other end in chipped ice, heat will flow into the bar from the water, then through the bar and to the ice, causing it to melt. The rate at which the ice melts indicates the rate heat is flowing through the bar.

The amount of ice that melts depends upon:

1. Time	The longer the time, the more ice will melt.
2. Size of the bar	The larger the cross-sectional area of the bar, the more heat will flow.
3. Length of the bar	The shorter the bar, the faster the ice will melt.
4. The temperature to which the bar is heated	The higher the temperature of the hot end of the bar, the faster the ice will melt.
5. Thermal conductivity	The higher the thermal conductivity of the bar, the more heat will flow.

The amount of heat flow, therefore, depends upon time, area, length, temperature difference and thermal conductivity.

Table 3-4. Thermal Conductivity of Metals

Metal	Chemical Symbol	Btu/sq ft/in. hr/°F	Relative Conductivity Based on Silver as 100%
Aluminum	Al	1428	49.7%
Copper	Cu	2664	92.7
Gold	Au	2037	70.9
Iron-Pure	Fe	467	16
Iron-Steel		313	10.9
Iron-Cast		316	11
Lead	Pb	241	8
Mercury	Hg	476	17
Molybdenum	Mo	1004	34.9
Nickel	Ni	412	14.3
Platinum	Pt	483	16.8
Silver	Ag	2873	100
Tin	Sn	450	15.6
Tungsten	W	1381	48
Zinc	Zn	770	27

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Thermal Conductivity

Thermal conductivity is a measure of the rate at which heat will flow through a material. The difference in thermal conductivity between iron and copper is easily demonstrated, Fig. 3-10. The copper conducts heat much faster than does the iron.

If one end of a copper bar is kept in boiling water (212 F) and the other end in chipped ice, heat will flow into the bar from the water, then through the bar and to the ice, causing it to melt. The rate at which the ice melts indicates the rate heat is flowing through the bar.

The amount of ice that melts depends upon:

1. Time	The longer the time, the more ice will melt.
2. Size of the bar	The larger the cross-sectional area of the bar, the more heat will flow.
3. Length of the bar	The shorter the bar, the faster the ice will melt.
4. The temperature to which the bar is heated	The higher the temperature of the hot end of the bar, the faster the ice will melt.
5. Thermal conductivity	The higher the thermal conductivity of the bar, the more heat will flow.

The amount of heat flow, therefore, depends upon time, area, length, temperature difference and thermal conductivity.

Table 3-4. Thermal Conductivity of Metals

Metal	Chemical Symbol	Btu/sq ft/in. hr/°F	Relative Conductivity Based on Silver as 100%
Aluminum	Al	1428	49.7%
Copper	Cu	2664	92.7
Gold	Au	2037	70.9
Iron-Pure	Fe	467	16
Iron-Steel		313	10.9
Iron-Cast		316	11
Lead	Pb	241	8
Mercury	Hg	476	17
Molybdenum	Mo	1004	34.9
Nickel	Ni	412	14.3
Platinum	Pt	483	16.8
Silver	Ag	2873	100
Tin	Sn	450	15.6
Tungsten	W	1381	48
Zinc	Zn	770	27

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Thermal Expansion

Thermal expansion is the increase in dimensions of a body due to a change in its temperature. See Fig. 3-11 for thermal expansion of selected materials.

The coefficient of linear expansion is the ratio of the change in length of a material, caused by heating it one degree, divided by the original length.

The coefficient of linear expansion of iron at room temperature is 0.0000065 per degree F. ($6.5 \times 10^{-6}/°F$).

The total increase in length of an iron bar 100 feet long which is heated from 10 F to 110 F will be:

$0.0000065 \times (110-10) \times 100 = 0.065$ ft or 0.78 in.

The coefficient of cubical expansion equals approximately three times the coefficient of linear expansion.

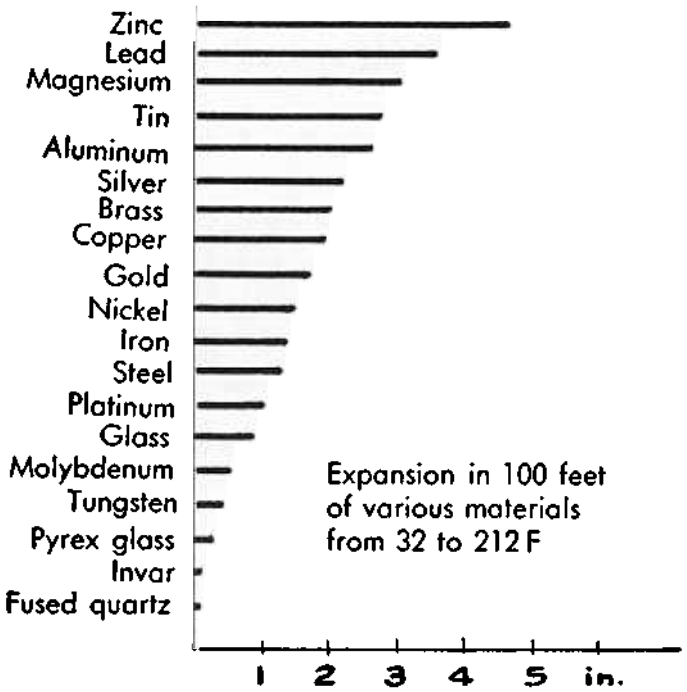

Fig. 3-11. Linear thermal expansion (in in.) of 19 materials when heated from 32 F to 212 F.

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Table 4-1. Typical Ferrous Materials

MATERIAL	SPEC.	CHEMICAL ANALYSIS			MECHANICAL PROPERTIES			
		CARBON	OTHERS		YIELD STRENGTH	TENSILE STRENGTH		
Cast Iron, Gray, Grade 20	ASTM A48-56	3.00-4.00				20,000		163
Gray Grade 30	ASTM A48-56	3.00-4.00				30,000		180
Nickel		2.00-3.50	Ni 0.25-0.50			40,000		310
Chrome-Nickel		2.00-3.50	Ni 1.00-3.00	Cr 0.50-1.00		53,000		510
White		2.00-4.00	Si 0.80-1.50			46,000		420
Malleable	ASTM A47-52	1.75-2.30	Si 0.85-1.20		35,000	53,000	18	140
Iron, Wrought, Plates	ASTM A42-55	0.08	Si 0.15	Slag 1.20	26,000	46,000	35	105
" " Forgings	ASTM A75-55	0.01-0.05	Fe 99.45-99.80		25,000	44,000	30	100
Steel, Cast, Low Carbon		0.11	Mn 0.60	Si 0.40	35,000	60,000	22	120
" " Medium Carbon		0.25	Mn 0.68	Si 0.32	44,000	72,000	18	140
" " High Carbon		0.50			40,000	80,000	17	182
Steel, Rolled, Carbon	SAE 1010	0.05-0.15			28,000	56,000	35	110
" " Carbon	SAE 1015	0.15-0.25			30,000	60,000	26	120
" " Carbon	SAE 1025	0.20-0.30			33,000	67,000	25	135
" " Carbon	SAE 1035	0.30-0.40			52,000	87,000	24	175
" " Carbon	SAE 1045	0.40-0.50			58,000	97,000	22	200
" " Carbon	SAE 1050	0.45-0.55			60,000	102,000	20	207
" " Carbon	SAE 1095	0.90-1.05			100,000	150,000	15	300
" " Nickel	SAE 2315	0.10-0.20	Ni 3.25-3.75		90,000	125,000	21	230
" " Ni.-Cr.	SAE 3240	0.35-0.45	Ni 1.50-2.00	Cr 0.90-1.25	113,000	136,000	21	280
" " Moly.	SAE 4130	0.25-0.35	Cr 0.50-0.80	Mo 0.15-0.25	115,000	139,000	18	280
" " Cr.	SAE 5140	0.35-0.45	Cr 0.80-1.10		128,000	150,000	19	300
" " Cr.-Va.	SAE 6130	0.25-0.35	Cr 0.80-1.10	Va 0.15-0.18	125,000	150,000	18	310
" " Si.-Mn.	SAE 9260	0.55-0.65	Mn 0.60-0.90	Si 1.80-2.20	180,000	200,000	12	390

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Steels in the low-carbon group are generally tough, ductile and easily formed, machined and welded. Although low-carbon steels are not especially hard, some grades respond to heat treatment and are readily case-hardened by carburizing, cyaniding, flame-hardening, etc.

Table 4-2. Uses for Steel by Carbon Content

CARBON CLASS	CARBON RANGE %	TYPICAL USES
Low		Chain, nails, pipe, rivets, screws, sheets for pressing and stamping, wire.
		Bars, plates, structural shapes.
Medium		Axles, connecting rods, shafting.
High	0.45 - 0.60	
	0.60 - 0.75	
Very High	0.75 - 0.90	Chisels, punches, sand tools.
	0.90 - 1.00	Knives, shear blades, springs.
	1.00 - 1.10	Milling cutters, dies, taps.
	1.10 - 1.20	Lathe tools, woodworking tools.
	1.20 - 1.30	Files, reamers.
	1.30 - 1.40	Dies for wire drawing.
	1.40 - 1.50	Metal cutting saws.

Medium-carbon steels—have a carbon range of 0.30 to 0.45%. They are strong, hard and not so easily forged or welded as low-carbon steels. If extensive welding is to be done on medium-carbon steels, it is desirable to use those steels at the lower end of the carbon range—between 0.30 and 0.35% carbon. As the carbon content increases above 0.35%, the steel becomes increasingly difficult to weld since there is a greater tendency toward brittleness of the weld. Special electrodes and procedures are often necessary to prevent weld cracking.

High-carbon (0.45 to 0.75%) and **very-high-carbon** (0.75 to 1.5%) **steels**—are very strong and hard. Both properties increase with an increased carbon content.

While carbon has an important influence on the characteristics of

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

steel, the degree to which impurities are not removed by refining is also important. A slight increase in the percentage of phosphorus or sulphur will materially lower the ductility, malleability, fatigue and shock resistance and welding qualities of a steel.

High- and very-high-carbon steels respond well to heat treating. Nearly any degree of hardness, temper or strength may be obtained. In the annealed state, most of these materials may be readily machined. They also may be hot-worked for forming.

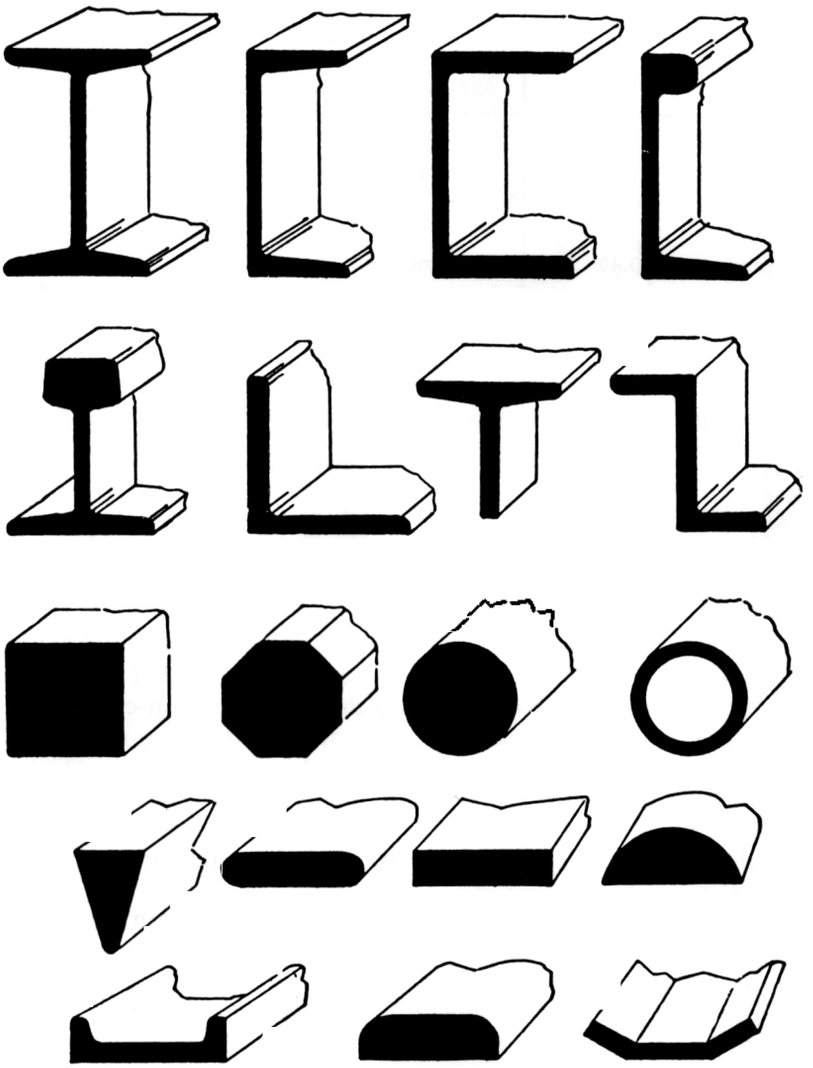

Fig. 4-1. A great many different shapes of rolled carbon and low-alloy steels are available from steel warehouses.

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Table 6-1. AWS Electrode Classification System

DIGIT	SIGNIFICANCE	EXAMPLE
1st two or 1st three	Min. tensile strength (stress relieved)	E-60xx = 60,000 psi (min) E-110xx = 110,000 psi (min)
2nd last	Welding position	E-xx1x = all positions E-xx2x = horizontal and flat E-xx3x = flat
Last	Power supply, type of slag, type of arc, amount of penetration, presence of iron powder in coating	See Table 6-2

Note: Prefix "E" (to left of a 4 or 5-digit number) signifies arc welding electrode

Measuring Preheat and Interpass Temperatures

Instruments are available for measuring surface temperatures. One such instrument has a wire coil, which is placed on the work; within a few seconds the temperature may be read on the dial at the handle end of the pyrometer. Preheat temperatures, however, can be satisfactorily estimated without expensive instruments.

Several suppliers* offer materials that melt over comparatively narrow ranges of temperatures. These indicators are furnished in sticks, liquids or pellets. If a preheat temperature of, say, 400 F, is wanted, the 400 F stick is scratched over the surface of the metal; if it leaves a white line, the metal is not at 400 F. If a transparent liquid line is formed, the metal is at or above the temperature printed on the stick.

Some common items used as rough temperature indicators are listed here:

1. Blue chalk: A mark made with carpenter's blue chalk on hot metal will turn whitish gray when the metal is above 625 F.
2. Red lumber-marking crayon: Its mark melts at about 615 F. The red mark turns pink at about 635 F and turns a dirty gray at about 650 F.
3. Solder: 50-50 solder (50% lead and 50% tin) starts melting at 360 F and is completely molten at 420 F.
4. Pine stick: White pine stick chars at about 635 F (used to determine the proper pouring temperature for some babbitts).

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

Another method for checking preheat temperatures is by temper colors: The temperature and accompanying color that appears on a freshly filed surface of carbon steel are given in Table 13-1.

Table 13-1. Temper Colors

TEMPERATURE, F	COLOR
400	Faint straw
440	Straw
475	Deep straw
520	Bronze
540	Peacock
590	Full blue
640	Light blue

It should be noted that the presence of alloying elements in the steel affects the colors formed, and some of the higher alloy steels have no temper colors.

*The Tempil Corporation; Markal Company; and Curtiss-Wright Corporation.

Appendix 2—MELTING POINTS OF SOME ELEMENTS

ELEMENT	CENT.	FAHR.	ELEMENT	CENT.	FAHR.
Hydrogen	—259	—434	Uranium		
Oxygen	—218	—360	Manganese		
Nitrogen	—210	—346	Beryllium		
Phosphorus	44	111	Silicon		
Sodium	97	207	Nickel		
Sulphur	113	236	Cobalt		
Selenium	218	424	Chromium		
Tin	232	450	Iron		
Bismuth	271	520	Titanium		
Cadmium	321	610	Thorium		
Lead	327	621	Zirconium		
Zinc	419	737	Vanadium		
Antimony	630	1,166	Hafnium		
Magnesium	651	1,204	Boron		
Aluminum	659	1,218	Columbium (Niobium)		
Calcium	810	1,490	Molybdenum		
Barium	850	1,562	Tantalum		
Silver	960	1,761	Tungsten		
Copper	1,083	1,981	Carbon	[illegible]	[illegible]

Reprinted With Permission From: The James F. Lincoln Arc Welding Foundation

WELDING

Skills, Processes and Practices for Entry-Level Welders

BOOK 2

- Gas Metal Arc Welding
- Flux Cored Arc Welding

LARRY JEFFUS • LAWRENCE BOWER

Welding

Skills, Processes and Practices for Entry-Level Welders

Book 2

- Gas Metal Arc Welding
- Flux Cored Arc Welding

First Edition

Larry Jeffus
Lawrence Bower

DELMAR
CENGAGE Learning™

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

DELMAR
CENGAGE Learning™

Welding: Skills, Processes and Practices for Entry-Level Welders: Book Two
Larry Jeffus/Lawrence Bower

Vice President, Editorial: Dave Garza
Director of Learning Solutions: Sandy Clark
Executive Editor: David Boelio
Managing Editor: Larry Main
Senior Product Manager: Sharon Chambliss
Editorial Assistant: Lauren Stone
Vice President, Marketing: Jennifer McAvey
Executive Marketing Manger: Deborah S. Yarnell
Senior Marketing Manager: Jimmy Stephens
Marketing Specialist: Mark Pierro
Production Director: Wendy Troeger
Production Manager: Mark Bernard
Content Project Manager: Cheri Plasse
Art Director: Benj Gleeksman
Technology Project Manager: Christopher Catalina
Production Technology Analyst: Thomas Stover

© 2010 Delmar, Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at
Professional & Career Group Customer Support,
1-800-648-7450

For permission to use material from this text or product, submit a request online at **cengage.com/permissions.**
Further permissions questions can be e-mailed to
permissionrequest@cengage.com.

Library of Congress Control Number: 2008911231

ISBN-13: 978-1-4354-2790-7
ISBN-10: 1-4354-2790-4

Delmar
5 Maxwell Drive
Clifton Park, NY, 12065-2919
USA

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For your lifelong learning solutions, visit **delmar.cengage.com**
Visit our corporate website at **cengage.com.**

Notice to the Reader
Publisher does not warrant or guarantee any of the products described herein or perform any independent analysis in connection with any of the product information contained herein. Publisher does not assume, and expressly disclaims, any obligation to obtain and include information other than that provided to it by the manufacturer. The reader is expressly warned to consider and adopt all safety precautions that might be indicated by the activities described herein and to avoid all potential hazards. By following the instructions contained herein, the reader willingly assumes all risks in connection with such instructions. The publisher makes no representations or warranties of any kind, including but not limited to, the warranties of fitness for particular purpose or merchantability, nor are any such representations implied with respect to the material set forth herein, and the publisher takes no responsibility with respect to such material. The publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or part, from the readers' use of, or reliance upon, this material.

Printed in the United States of America
1 2 3 4 5 XX 11 10 09

Brief Contents

Preface viii

AWS Acknowledgment xi

Acknowledgments xii

About the Authors xiii

Chapter 1 Gas Metal Arc Welding Equipment, Setup, and Operation 1

Chapter 2 Gas Metal Arc Welding 55

Chapter 3 Flux Cored Arc Welding Equipment, Setup, and Operation 113

Chapter 4 Flux Cored Arc Welding 143

Appendix 197

Glossary 199

Index 203

This page intentionally left blank

Contents

Preface viii

AWS Acknowledgment xi

Acknowledgments xii

About the Authors xiii

Chapter 1 Gas Metal Arc Welding Equipment, Setup, and Operation 1

Introduction 2
Metal Transfer 4
Axial Spray Metal Transfer 5
Globular Transfer 7
Pulsed-Spray Transfer: GMAW-P 7
Shielding Gases for Spray or Pulsed-spray Transfer 9
Buried-arc Transfer 9
Short-circuiting Transfer: GMAW-S 11
Filler Metal Specifications 13
Filler Metal Selection 14
Stainless Steel Electrodes 15
Aluminum and Aluminum Alloys 15
Aluminum Bare Welding Rods and Electrodes 20
Wire Melting and Deposition Rates 20
Welding Power Supplies 21
Speed of the Wire Electrode 23
Power Supplies for Short-circuiting Transfer 23
Inductance 25
Molten Weld Pool Control 25
Shielding Gas 26
Power Settings 29
Gun Manipulation 30
Travel Speed 30
Electrode Extension 30
Gun Angle 31
Equipment 33
Power Source 34
Electrode (Wire) Feed Unit 37
Push-type Feed System 37

Pull-type Feed System 38
Push-pull-type Feed System 39
Linear Electrode Feed System 39
Spool Gun 40
Electrode Conduit 41
Welding Gun 41
Spot Welding 42
Face, Eye, and Ear Protection 44
Face and Eye Protection. 44
Ear Protection. 48
General Work Clothing. 49
Special Protective Clothing 51
Hand Protection 51
Body Protection 52
Waist and Lap Protection 52
Arm Protection 52
Leg and Foot Protection. 52

Chapter 2 Gas Metal Arc Welding 55

Introduction 56
Setup. 57
Gas Density and Flow Rates 65
Arc-voltage and Amperage Characteristics 68
Electrode Extension 70
Welding Gun Angle 73
Effect of Shielding Gas on Welding 74
Practices. 77
Metal Preparation 78
Flat Position, 1G and 1F Positions 79
Vertical Up 3G and 3F Positions 84
Vertical Down 3G and 3F Positions 86
Horizontal 2G and 2F Positions 89
Overhead 4G and 4F Positions. 91
Axial Spray Transfer 94

Chapter 3 Flux Cored Arc Welding Equipment, Setup, and Operation 113

Introduction 114
Principles of Operation. 114
Equipment 117
Power Supply 117
Guns 117
Smoke Extraction Nozzles. 119
Electrode Feed 119

Advantages 119
High Deposition Rate. 119
Minimum Electrode Waste 119
Narrow Groove Angle 120
Minimum Precleaning 120
All-position Welding 120
Flexibility 120
High Quality. 120
Excellent Control. 121
Limitations 121
Electrodes 121
Methods of Manufacturing 121
Electrode Cast and Helix 124
Flux. 125
Flux Cored Steel Electrode Identification 128
Metal Cored Steel Electrode Identification 130
Care of Flux Core Electrodes. 130
Shielding Gas 131
Welding Techniques 132
Gun Angle 132
Forehand/Perpendicular/Backhand Techniques. . 134
Travel Speed. 136
Mode of Metal Transfer 137
Electrode Extension 138
Porosity 138

Chapter 4 Flux Cored Arc Welding. 143

Introduction 144
Practices. 145
Flat-position Welds 151
Square-groove Welds 152
V-Groove and Bevel-Groove Welds 154
Root Pass 155
Filler Pass 156
Cover Pass 157
Fillet Welds. 160
Vertical Welds. 164
Horizontal Welds. 170
Overhead-position Welds 175
Thin-gauge Welding. 180

Appendix 197

Glossary 199

Index. 203

Preface

ABOUT THE SERIES

Welding: Skills, Processes and Practices for Entry-Level Welders is an exciting new series that has been designed specifically to support the American Welding Society's (AWS) SENSE EG2.0 training guidelines. Offered in three volumes, these books are carefully crafted learning tools consisting of theory-based texts that are accompanied by companion lab manuals, and extensive instructor support materials. With a logical organization that closely follows the modular structure of the AWS guidelines, the series will guide readers through the process of acquiring and practicing welding knowledge and skills. For schools already in the SENSE program, for those planning to join, or for schools interested in obtaining certifiable outcomes based on nationally recognized industry standards order to comply with the latest Carl D. Perkins Career and Technical Education in requirements, *Welding: Skills, Processes and Practices for Entry-Level Welders* offers a turnkey solution of high quality teaching and learning aids.

Career and technical education instructors at the high school level are often called upon to be multi-disciplinary educators, teaching welding as only one of as many as five technical disciplines in any given semester. The *Welding: Skills, Processes and Practices for Entry-Level Welders* package provides these educators with a process-based, structured approach and the tools they need to be prepared to deliver high level training on processes and materials with which they may have limited familiarity or experience. Student learning, satisfaction and retention are the target of the logically planned practices, supplements and full color textbook illustrations. While the AWS standards for entry level welders are covered, students are also introduced to the latest in high technology welding equipment such as pulsed gas metal arc welding (GMAW-P). Career pathways and career clusters may be enhanced by the relevant mathematics applied to real world activities as well as oral and written communication skills linked to student interaction and reporting.

Book 1, the core volume, introduces students to the welding concepts covered in AWS SENSE Modules 1, 2, 3, 8 and 9 (Occupational Orientation, Safety and Health of Welders, Drawing and Welding Symbol Interpretation, Thermal Cutting, and Weld Inspection Testing and Codes). Book 1 contains all the material needed for a SENSE program that prepares students for qualification in Thermal Cutting processes. The optional Books 2 and 3 cover other important welding processes and are grouped in logical combinations. Book 2 corresponds to AWS SENSE Modules 5 and 6 (GMAW, FCAW), and Book 3 corresponds to AWS SENSE Modules 4 and 7 (SMAW, GTAW).

The texts feature hundreds of four-color figures, diagrams and tight shots of actual welds to speed beginners to an understanding of the most widely used welding processes.

FEATURES

- Produced in close collaboration with experienced instructors from established SENSE programs to maximize the alignment of the content with SENSE guidelines and to ensure 100% coverage of Level I-Entry Welder Key Indicators.
- Chapter introductions contain general performance objectives, key terms used, and the AWS SENSE EG2.0 Key Indicators addressed in the chapter.
- Coverage of Key Indicators is indicated in the margin by a torch symbol and a numerical reference.
- Contains scores of fully illustrated Practices, which are guided exercises designed to help students master processes and materials. Where applicable, the Practices reproduce and reference actual AWS technical drawings in order to help students create acceptable workmanship samples.
- Each section introduces students to the materials, equipment, setup procedures and critical safety information they need in order to weld successfully.
- Hundreds of four-color figures, diagrams and tight shots of actual welds to speed beginners to an understanding of the most widely used welding processes.
- End of chapter review questions develop critical thinking skills and help students to understand "why" as well as "how."

SUPPLEMENTS

Each book in the Welding Skills series is accompanied by a ***Lab Manual*** that has been designed to provide hands-on practice and reinforce the student's understanding of the concepts presented in the text. Each chapter contains practice exercises to reinforce the primary objectives of the lesson, including creation of workmanship samples (where applicable), and a quiz to test knowledge of the material. Artwork and safety precautions are included throughout the manuals.

Instructor Resources (on CD-ROM), designed to support Books 1–3 and the accompanying Lab Manuals, provide a wealth of time-saving tools, including:

- An Instructor's Guide with answers to end-of-chapter Review Questions in the texts and Lab Manual quizzes.
- Modifiable model Lesson Plans that aid in the design of a course of study that meets local or state standards and also maps to the SENSE guidelines.
- An extensive ExamView Computerized Test Bank that offers assessments in true/false, multiple choice, sentence completion and short answer formats. Test questions have been designed to expose students to the types of questions they'll encounter on the SENSE Level 1 Exams.
- PowerPoint Presentations with selected illustrations that provide a springboard for lectures and reinforce skills and processes covered in the texts. The PowerPoint Presentations can be modified or expanded as instructors desire, and can be augmented with additional illustrations from the Image Library.
- The Image Library contains nearly all (well over 1000!) photographs and line art from the texts, most in four-color.
- A SENSE Correlation Chart that shows the close alignment of the *Welding* series to the SENSE Entry Level 1 training guidelines. Each Key Indicator within each SENSE Module is mapped to the relevant text and lab manual page or pages.

TITLES IN THE SERIES

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 1, Occupational Orientation, Safety and Health of Welders, Drawing and Welding Symbol Interpretation, Thermal Cutting, Weld Inspection Testing and Codes
(Order #: 1-4354-2788-2)
Lab Manual, Book One (Order #: 1-4354-2789-0)

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 2, Gas Metal Arc Welding, Flux Cored Arc Welding (Order #:1-4354-2790-4)
Lab Manual, Book Two (Order #: 1-4354-2795-5)

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 3, Shielded Metal Arc Welding, Gas Tungsten Arc Welding (Order #:1-4354-2796-3)
Lab Manual, Book Three (Order #: 1-4354-2797-1)

AWS Acknowledgment

The Authors and Publisher gratefully acknowledge the support provided by the American Welding Society in the development and publication of this textbook series. "American Welding Society," the AWS logo and the SENSE logo are the trade and service marks of the American Welding Society and are used with permission.

For more information on the American Welding Society and the SENSE program, visit **http://www.aws.org/education/sense/** or contact AWS at (800) 443-9353 ext. 455 or by email: **education@aws.org**.

Acknowledgments

The authors and publisher would like to thank the following individuals for their contributions to this series:

Garey Bish, *Gwinnett Technical College, Lawrenceville, GA*
Julius Blair, *Greenup County Area Technology Center, Greenup, KY*
Rick Brandon, *Pemiscot County Career & Technical Center, Hayti, MO*
Stephen Brandow, *University of Alaska Southeast, Ketchikan, Ketchikan, AK*
Francis X Brieden, *Career Technology Center of Lackawanna County, Scranton, PA*
John Cavenaugh, *Community College of Southern Nevada, Las Vegas, NV*
Clay Corey, *Washington-Saratoga BOCES, Fort Edward, NY*
Keith Cusey, *Institute for Construction Education, Decatur, IL*
Craig Donnell, *Whitmer Career Technology Center, Toledo, OH*
Steve Farnsworth, *Iowa Lakes Community College, Emmetsburg, IA*
Ed Harrell, *Traviss Career Center, Lakeland, FL*
Robert Hoting, *Northeast Iowa Community College, Sheldon, IA*
Steve Kistler, *Moberly Area Technical Center, Moberly, MO*
David Lynn, *Lebanon Technology & Career Center, Lebanon, MO*
Frank Miller, *Gadsden State Community College, Gadsden, AL*
Chris Overfelt, *Arnold R Burton Tech Center, Salem, VA*
Kenric Sorenson, *Western Technical College, LaCrosse, WI*
Pete Stracener, *South Plains College, Levelland, TX*
Bill Troutman, *Akron Public Schools, Akron, OH*
Norman Verbeck, *Columbia/Montour AVTS, Bloomsburg, PA*

About The Authors

Larry Jeffus is a dedicated teacher and author with over twenty years experience in the classroom and several Delmar Cengage Learning welding publications to his credit. He has been nominated by several colleges for the Innovator of the Year award for setting up nontraditional technical training programs. He was also selected as the Outstanding Post-Secondary Technical Educator in the State of Texas by the Texas Technical Society. Now retired from teaching, he remains very active in the welding community, especially in the field of education.

Lawrence Bower is a welding instructor at Blackhawk Technical College, an AWS SENSE School, in Janesville, Wisconsin. Mr. Bower is an AWS-certified Welding Inspector and Welding Educator. In helping to create *Welding: Skills, Processes and Practices for Entry-Level Welders*, he has brought to bear an excellent mix of training experience and manufacturing know-how from his work in industry, including fourteen years at United Airlines, and six years in the US Navy as an aerospace welder.

This page intentionally left blank

CHAPTER

1 Gas Metal Arc Welding Equipment, Setup, and Operation

OBJECTIVES

After completing this chapter, the student should be able to

- demonstrate the proper use of personal protective equipment (PPE) for gas metal arc welding (GMAW)
- list the various terms used to describe gas metal arc welding
- describe methods of metal transfer including the axial spray metal transfer process, globular transfer, pulsed-arc metal transfer (GMAW-P), and short-circuiting transfer (GMAW-S)
- list four common shielding gases or gas mixtures used for short-circuiting, spray, and pulsed-spray transfer on plain carbon steel
- locate the GMA welding filler metal on a welding procedure specification (WPS)
- define *deposition efficiency,* and tell how a welder can control the deposition rate
- define *voltage, electrical potential, amperage,* and *electrical current* as related to GMA welding
- tell how wire-feed speed is determined and demonstrate its relationship to welding current
- list five ways the GMAW molten weld pool can be controlled by varying the shielding gas, power settings, travel speed, electrode extension, and gun angle
- describe and demonstrate the backhand and forehand welding techniques and their relationship to weld bead profile and penetration in the short-circuiting transfer mode
- list and describe the basic GMAW equipment
- use a chart to select the correct eye and face protective devices for working and welding in a shop
- describe what type of general work clothing should be worn in a welding shop
- describe special protective clothing worn by welders to protect hands, arms, body, waist, legs, and feet

KEY TERMS

axial spray metal transfer
electrode extension (stickout)
globular transfer
pinch effect
pulsed-arc metal transfer
short-circuiting transfer
slope
synergic systems
transition current
welding helmet

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter:

Module 2: Safety and Health of Welders

Key Indicator 1: Demonstrates proper use and inspection of personal protection equipment (PPE)

Module 5: Gas Metal Arc Welding (GMAW-S, GMAW spray transfer)

Key Indicator 1: Performs safety inspections of GMAW equipment and accessories

Key Indicator 3: Sets up for GMAW-S operations on carbon steel

Key Indicator 8: Sets up for GMAW (spray) operations on carbon steel

INTRODUCTION

In the 1920s, a metal arc welding process using an unshielded wire was being used to assemble the rear axle housings for automobiles. The introduction of the shielded metal arc welding electrode rapidly replaced the bare wire. The shielded metal arc welding electrode made a much higher-quality weld. In 1948, the first inert gas metal arc welding (GMAW) process, as it is known today, was developed and became commercially available, Figure 1.1. In the beginning, the GMAW process was used to weld aluminum using argon (Ar) gas for shielding. As a result, the process was known as MIG, which stands for metal inert gas welding. The later introduction of CO_2 and O_2 to the shielding gas has resulted in the American Welding Society's preferred term *gas metal*

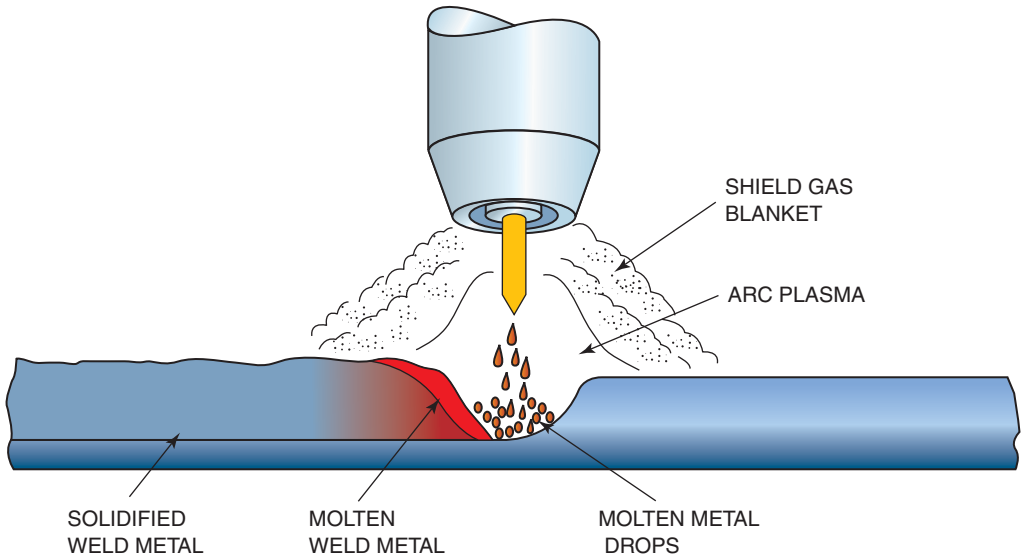

Figure 1.1 Gas-shielded metal arc welding (GMAW)

Table 1.1 Methods of Performing Welding Processes

Function	Manual (MA) (Example: SMAW)	Semiautomatic (SA) (Example: GMAW)	Machine (ME) (Example: GMAW)	Automatic (AU) (Example: GMAW)
Maintain the arc	Welder	Machine	Machine	Machine
Feed the filler metal	Welder	Machine	Machine	Machine
Provide the joint travel	Welder	Welder	Machine	Machine
Provide the joint guidance	Welder	Welder	Welder	Machine

arc welding (GMAW). Although the American Welding Society uses the term *gas metal arc welding* to describe this process, it is known in the field by several other terms, such as

- MIG, which is short for metal inert gas welding
- MAG, which is short for metal active gas welding
- *wire welding*, which describes the electrode used

The GMAW process may be performed as semiautomatic (SA), machine (ME), or automatic (AU) welding, Table 1.1. The GMA welding process is commonly performed as a semiautomatic process and is often mistakenly referred to as "semiautomatic welding." Equipment is available to perform most of the wire-feed processes semiautomatically, and the GMAW process can be fully automated. Robotic arc welding often uses GMAW because of the adaptability of the process in any position, Figure 1.2.

The rising use of all the various types of consumable wire welding processes has resulted in the increased sales of wire. At one time, wire made up less than 1% of the total market of filler metal. The total tonnage of filler metals used has grown and so has the percentage of wire. Today, wire exceeds 50% of the total tonnage of filler metals produced and used.

Much of the increase in the use of the wire welding processes is due to the increases in the quality of the welds produced. This improvement is due to an increased reliability of the wire-feed systems, improvements in the filler metal, smaller wire sizes, faster welding speed, higher weld deposition rates, less expensive shielding gases, and improved welding techniques. Table 1.2 shows the typical weld deposition rates using the GMA welding process. The increased usage has led to a reduction in the cost of equipment. GMA welding equipment is now found even in small shops.

In this chapter, the semiautomatic GMA welding process will be covered. The skill required to set up and operate this process is basic to the understanding and operation of other wire-feed processes. The reaction of the weld to changes in voltage, amperage, feed speed, stickout, and gas is similar to that of most wire-feed processes.

Figure 1.2 GMAW equipment
(A) Semiautomatic GMA welding setup. (B) Machine GMA welding gun with a friction drive, which provides both uniform nozzle-to-work distance and travel speed. (C) Automatic GMA welding.
Source: Courtesy of Lincoln Electric Company

Table 1.2 GMA Weld Deposition Rates

Electrode Diameter	Pounds per Hour		
Amperage	0.35	0.45	0.63
50	2.0	–	–
100	4.8	4.2	–
150	7.5	6.7	5.1
200	–	8.7	7.8
250	–	12.7	11.1
300	–	–	14.4

METAL TRANSFER

When first introduced, the GMA welding process was used with argon as a shielding gas to weld aluminum. Even though argon (Ar) was then expensive, the process was accepted immediately because it was much more productive than the gas tungsten arc (GTA) process and because it produced higher-quality welds than the shielded metal arc (SMA) process. This new arc welding process required very little postweld cleanup because it was slag and spatter free.

Axial Spray Metal Transfer

The freedom from spatter associated with the argon-shielded GMAW process results from a unique mode of metal transfer called **axial spray metal transfer**, Figure 1.3. This process is identified by the pointing of the wire tip from which droplets smaller than the diameter of the electrode wire are projected axially across the arc gap to the molten weld pool. There are hundreds of drops per second crossing from the wire to the base metal. These drops are propelled by arc forces at high velocity in the direction the wire is pointing. In the case of plain carbon steel and stainless steels, the molten weld pool may be too large and too hot to be controlled in vertical or overhead positions. Because the drops are very small and directed at the molten weld pool, the process is spatter free.

The spray transfer mode for carbon steels requires three conditions: argon-rich shielding gas mixtures, direct current electrode positive (DCEP) polarity (also called direct current reverse polarity, DCRP), and a current level above a critical amount called the **transition current**. The shielding gas is usually a mixture of 95% to 98% argon and 5% to 2% oxygen, or 80% to 90% argon and 20% to 10% carbon dioxide. The added percentage of active gases allows greater weld penetration. Figure 1.4 illustrates how the rate of drops transferred changes in relationship to the welding current. At low currents, the drops are large and are transferred at rates below 10 per second. These drops move slowly, falling from the electrode tip as gravity pulls them down. They tend to bridge the gap between the electrode

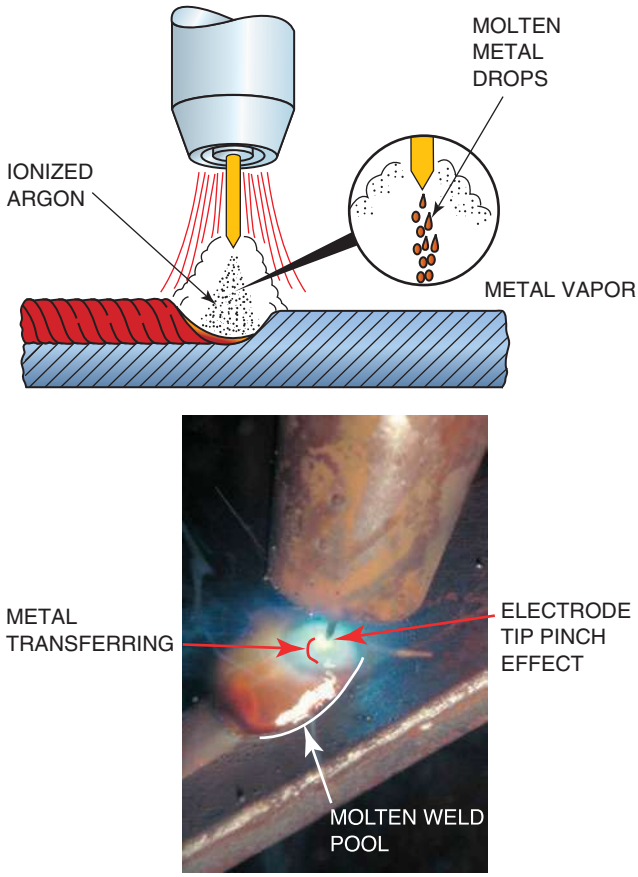

Figure 1.3 Axial spray metal transfer
Note the pinch effect of filler wire and the symmetrical metal transfer column.
Source: Courtesy of Larry Jeffus

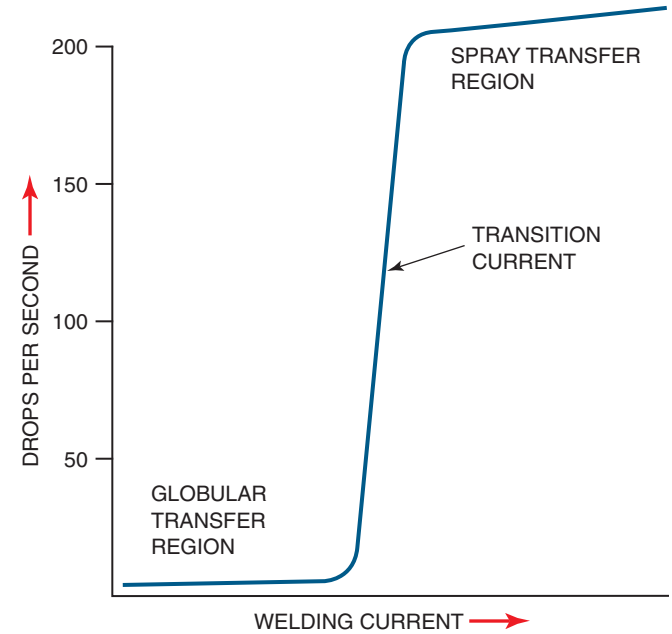

Figure 1.4 Desirable spray transfer shown schematically

tip end and the molten weld pool. This produces a momentary short circuit that throws off spatter. However, the mode of transfer changes abruptly above the critical current, producing the desirable spray.

The transition current depends on the alloy being welded. It is also proportional to the wire diameter, meaning that higher currents are needed with larger diameter wires. The need for high current density imposes some restrictions on the process. The high current hinders welding of sheet metal because the high heat cuts through sheet metal. High current also limits its use to the flat and horizontal welding positions. Weld control in the vertical or overhead position is very difficult to impossible to achieve. Table 1.3 lists the welding parameters for a variety of gases, wire sizes, and metal thicknesses for GMA welding of mild steel.

Table 1.3 GMA Welding Parameters for Mild Steel

Mild Steel	Wire-feed Speed, in./min		Voltage, V				
Base-material Thickness, in.	0.035-in.	0.045-in.	CO_2	75 Ar-25 CO_2	Ar	98 Ar-2O_2	Current A
0.036	105–115	–	18	16	–	–	50–60
0.048	140–160	70	19	17	–	–	70–80
0.060	180–220	90–110	20	17.7	–	–	90–110
0.075	240–260	120–130	20.7	18	20	–	120–130
1/8	280–300	140–150	21.5	18.5	20.5	–	140–150
3/16	320–340	160–175	22	19	21.5	23.5	160–170
1/4	360–380	185–195	22.7	19.5	22.5	24.5	180–190
5/16	400–420	210–220	23.5	20.5	23.5	25	200–210
3/8	420–520	220–270	25	22	25	26.5	220–250
1/2 and up	–	375	28	26	29	31	300

Globular Transfer

The **globular transfer process** is rarely used by itself because it transfers the molten metal across the arc in much larger droplets. It is used in combination with pulsed-spray transfer.

Globular transfer can be used on thin materials and at a very low current range. It can be used with higher current but is not as effective as other welding modes of metal transfer.

Pulsed-Spray Transfer: GMAW-P

Because the change from spray arc to globular transfer occurs within a very narrow current range and the globular transfer occurs at the rate of only a few drops per second, a controlled spray transfer at significantly lower average currents is achievable. Gas metal arc welding with pulsed-spray transfer (GMAW-P), or **pulsed-arc metal transfer** involves pulsing the current from levels below the transition current to those above it. The time interval below the transition current is short enough to prevent a drop from developing. About 0.1 second is needed to form a globule, so no globule can form at the electrode tip if the time interval at the low base current is about 0.01 second. Actually, the energy produced during this time is very low—just enough to keep the arc alive.

The arc's real work occurs during those intervals when the current pulses to levels above the transition current. The time of that pulse is controlled to allow a single drop of metal to transfer. This is typical of the drops normally associated with spray transfer. In fact, with many power supplies, a few small drops could transfer during the pulse interval. As with conventional spray arc, the drops are propelled across the arc gap, allowing metal transfer in all positions.

The average current can be reduced sufficiently to reduce penetration enough to weld sheet metal or reduce deposition rates enough to control the molten weld pool in all positions. This level controlling the weld heat input and rate of weld metal deposit is achieved by changing the following variables, graphed in Figure 1.5:

- Frequency—The number of times the current is raised and lowered to form a single pulse; frequency is measured in pulses per second.
- Amplitude—The amperage or current level of the power at the peak or maximum, expressed in amperage.
- Width of the pulses—The amount of time the peak amperage is allowed to stay on.

Figure 1.6 shows a typical pulsed-arc welding system. Although developed in the mid-1960s, this technology did not receive much attention until solid state electronics were developed to handle the high power required of welding power supplies. Solid state electronics provided a better, simpler, and more economical way to control the pulsing process. The newest generation of pulsed-arc systems interlocks the power supply and wire feeder so that the proper settings of the wire-feed and power supply are obtained for any given job by adjusting a single knob. Such systems have been termed **synergic systems**. The greatest benefit to synergic GMAW-P is that the power supply reacts to changes in the

Caution

The heat produced during pulsed-spray transfer welding using large-diameter wire or high current may be intense enough to cause the filter lens in a welding helmet to shatter. Be sure the helmet is equipped with a clear plastic lens on the inside of the filter lens. Avoid getting your face too close to the intense heat.

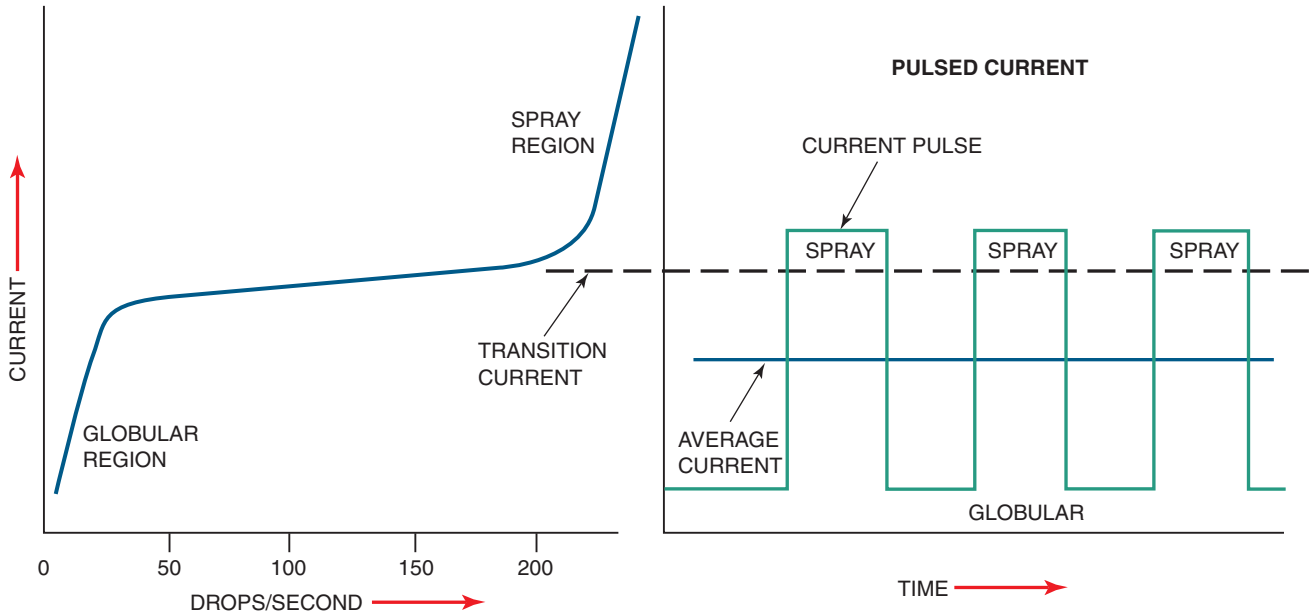

Figure 1.5 Mechanism of pulsed-arc spray transfer at a low average current

Figure 1.6 GMA pulsed-arc welding system controller
Source: Courtesy of Thermal Arc, a Thermadyne Company

work conditions such as inconsistent fitups or operator changes in stickout. For example, with a traditional GMA weld, amperage is reduced as the operator increases the stickout; but with synergic GMAW-P, the power supply senses the change in stickout and adjusts the current to maintain a more consistent weld.

In some respects, synergic GMAW-P systems are more complex to use because the correct relationships between the wire-feed speeds and power supply settings must be programmed into the equipment, and each wire composition, wire size, and shielding gas requires a special program. The manufacturer generally programs the most common combinations, allowing space in the computer for additional user input. In addition, welding power supplies that produce synergic pulsation must be capable of both constant-current (CC) and constant-voltage (CV) power output.

Shielding Gases for Spray or Pulsed-spray Transfer

Axial spray transfer is impossible without shielding gases rich in argon. Pure argon is used with all metals except steels. As much as 80% helium can be added to the argon to increase the heat in the arc without affecting the desirable qualities of the spray mode. With more helium, the transfer becomes progressively more globular, forcing the use of a different welding mode, to be described later. Since these gases are inert, they do not react chemically with any metals. This factor makes the GMAW process the only productive manual or semiautomatic method for welding metals sensitive to oxygen (essentially all metals except iron or nickel). The cathodic cleaning action which helps to remove the thin layer oxides that form on metals is associated with argon at DCEP (DCRP) and is also very important for fabricating metals such as aluminum, which quickly develop these undesirable surface oxides when exposed to air.

This same cleaning action causes problems with steels. Iron oxide in and on the steel surface is a good emitter of electrons that attracts the arc. But these oxides are not uniformly distributed, resulting in very irregular arc movement and in turn irregular weld deposits. This problem was solved by adding small amounts of an active gas such as oxygen or carbon dioxide to the argon. The reaction produces a uniform film of iron oxide on the weld pool and provides a stable site for the arc. This discovery enabled uniform welds in ferrous alloys and expanded the use of GMAW to welding those materials.

The amount of oxygen needed to stabilize arcs in steel varies with the alloy. Generally, 2% is sufficient for carbon and low-alloy steels. In the case of stainless steels, about 0.5% should prevent a refractory scale of chromium oxide, which can be a starting point for corrosion in stainless steels. Carbon dioxide can substitute for oxygen. Between 8% and 10% is optimum for low-alloy steels. In many applications, carbon dioxide is the preferred addition because the weld bead has a better contour and the arc appears to be more stable. Gas mixes of 98% argon–2% oxygen as well as 98% argon–2% carbon dioxide are commonly used for spray transfer GMA welding of stainless steels.

Buried-arc Transfer

Carbon dioxide was one of the first gases studied during the development of the GMAW process. It was abandoned temporarily because of excessive spatter and porosity in the weld. After argon was accepted for shielding, further work with carbon dioxide demonstrated that the spatter was

associated with globular metal transfer. The large drops are partially supported by arc forces, Figure 1.7. As they become heavy enough to overcome those forces and drop into the pool, they bridge the gap between the wire and the weld pool, producing explosive short circuits and spatter.

Additional work showed that the arc in carbon dioxide was very forceful. Because of this, the wire tip could be driven below the surface of the molten weld pool. With the shorter arcs, the drop size is reduced, and any spatter produced as the result of short circuits is trapped in the cavity produced by the arc—hence the name **buried-arc transfer**, Figure 1.8. The resultant welds tend to be more highly crowned than those produced with open arcs, but they are relatively free of spatter and offer a decided advantage of welding speed. These characteristics make the buried-arc process useful for high-speed mechanized welding of thin sections, such as that found in compressor domes for hermetic air-conditioning and refrigeration equipment or for automotive components.

Because carbon dioxide is an oxidizing gas, its applications to welding carbon steels are restricted. It cannot be used to fabricate most nonferrous materials. Neither should it be used to weld stainless steels, because carbon corrodes the weld metal.

Carbon dioxide and helium are similar in that metal transfer in both gases is globular. Helium, too, can be used with the buried-arc technique. It has the advantage of being inert, potentially making it useful for the same types of applications as carbon dioxide but in nonferrous alloys.

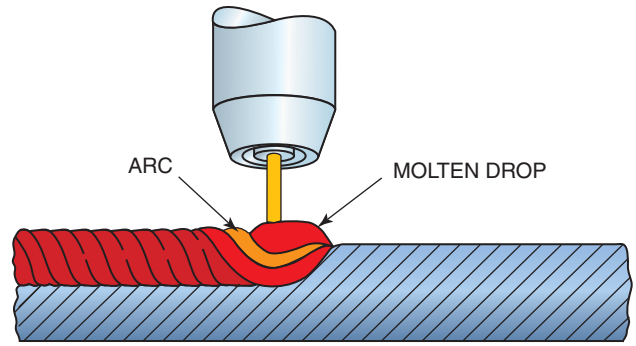

Figure 1.7 Globular metal transfer
Large drop is supported by arc forces.

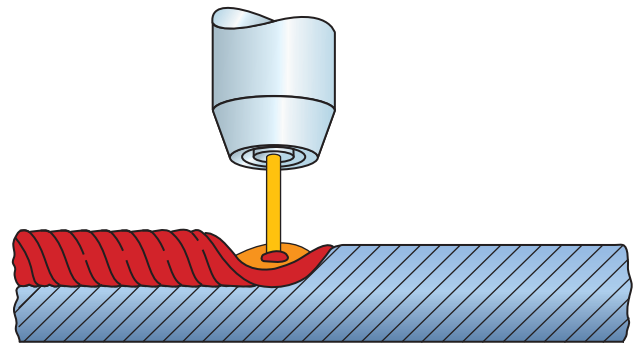

Figure 1.8 Buried-arc transfer
Wire tip is within the weld crater, so spatter is trapped.

Short-circuiting Transfer: GMAW-S

Low currents allow the liquid metal at the electrode tip to be transferred by direct contact with the molten weld pool. This process requires close interaction between the wire feeder and the power supply. This technique is called **short-circuiting transfer**, or gas metal arc welding with short-circuiting transfer (GMAW-S).

The short-circuiting mode of transfer is the most common process used with GMA welding

- on thin or properly prepared thick sections of material
- on a combination of thick to thin materials
- with a wide range of electrode diameters
- with a wide range of shielding gases
- in all positions

The 0.023, 0.030, 0.035, and 0.045 wire electrodes are the most common diameters for the short-circuiting mode. The most popular shielding gases used on carbon steel are 100% carbon dioxide (CO_2) or a combination of 75% argon (Ar) and 25% CO_2. The amperage range may be as low as 35 for materials of 24 gauge or as high as 225 for materials up to 1/8 inch in thickness on square groove weld joints. Thicker base metals can be welded if the edges are beveled to accept complete joint weld penetration.

The transfer mechanisms in this process are quite simple and straightforward, as shown schematically in Figure 1.9. To start, the wire is in direct contact with the molten weld pool, Figure 1.9A. Once the electrode touches the molten weld pool, the arc and its resistance are removed. Without the arc resistance, the welding amperage quickly rises as it begins to flow freely through the tip of the wire into the molten weld pool. The resistance to current flow is highest at the point where the electrode touches the molten weld pool. The resistance is high because both the electrode tip and weld pool are very hot. The higher the temperature,

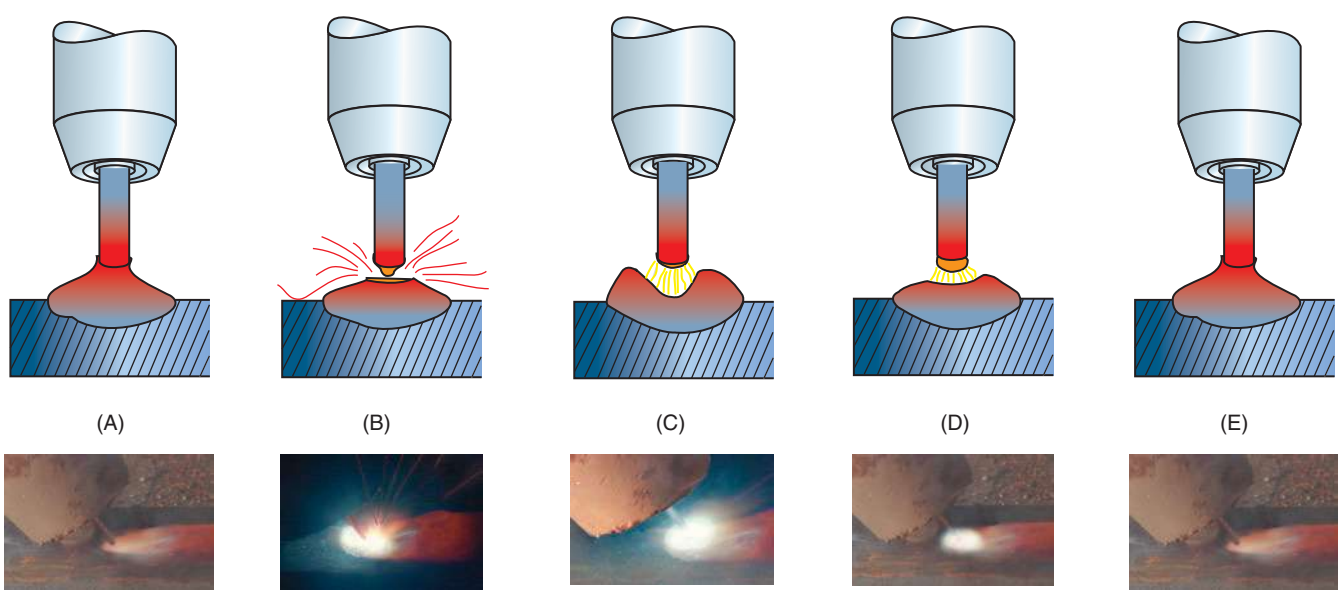

Figure 1.9 Schematic of short-circuiting transfer
Source: Courtesy of Larry Jeffus

the higher the resistance to current flow. A combination of high current flow and high resistance causes a rapid rise in the temperature of the electrode tip.

As the current flow increases, the interface between the wire and molten weld pool is heated until it explodes into a vapor (Figure 1.9B), establishing an arc. This small explosion produces sufficient force to depress the molten weld pool. A gap between the electrode tip and the molten weld pool (Figure 1.9C) immediately opens. With the resistance of the arc reestablished, the voltage increases as the current decreases.

The low current flow is insufficient to continue melting the electrode tip off as fast as it is being fed into the arc. As a result, the arc length rapidly decreases (Figure 1.9D) until the electrode tip contacts the molten weld pool (Figure 1.9E). The liquid formed at the wire tip during the arc-on interval is transferred by surface tension to the molten weld pool, and the cycle begins again with another short circuit.

If the system is properly tuned, the rate of short circuiting can be repeated from approximately 20 to 200 times per second, causing a characteristic buzzing sound. The spatter is low and the process easy to use. The low heat produced by GMAW-S makes the system easy to use in all positions on sheet metal, low-carbon steel, low-alloy steel, and stainless steel ranging in thickness from 25 gauge (0.02 in.; 0.5 mm) to 12 gauge (0.1 in.; 2.6 mm). The short-circuiting process does not produce enough heat to make quality welds in sections much thicker than 1/4 in. (6 mm) unless it is used for the root pass on a grooved weld or to fill gaps in joints. Although this technique is highly effective, lack-of-fusion defects can occur unless the process is perfectly tuned and the welder is highly skilled, especially on thicker metal. For this reason the American Welding Society D1.1 code for structural steel does not list short-circuiting GMAW as a prequalified process. In this code all short-circuiting procedures must be qualified by extensive testing.

Carbon dioxide works well with this short-circuiting process because it produces the forceful, high-energy arc needed during the arc-on interval to displace the weld pool. Helium can be used as well. Pure argon is not as effective because its arc tends to be sluggish and not very fluid. However, a mixture of 25% carbon dioxide and 75% argon produces a less harsh arc and a flatter, more fluid, and desirable weld profile. Although more costly, this gas mixture is preferred. A gas mixture of 98% argon and 2% oxygen may also be used on thinner carbon steels and sheet metal. This mixture produces lower-energy short-circuiting transfer and can be an advantage on thin-gauge metals, producing minimal burn-through at voltages as low as 13 volts.

New technology in wire manufacturing has allowed smaller wire diameters to be produced. These smaller diameters have become the preferred size even though they are more expensive due to the cost of drawing wires down to these desirable sizes. The short-circuiting process works better with a short electrode stickout.

The power supply is most critical. It must have a constant voltage, otherwise known as constant-potential output, and sufficient inductance to slow the time rate of current increase during the short-circuit interval. Too little inductance causes spatter due to high-current surges. Too much inductance causes the system to become sluggish. The short-circuiting rate decreases enough to make the process difficult to use. Also, the

power supply must sustain an arc long enough to premelt the electrode tip in anticipation of the transfer at recontact with the weld pool.

FILLER METAL SPECIFICATIONS

GMA welding filler metals are available for a variety of base metals, Table 1.4. The most frequently used filler metals are AWS specification A5.18 for carbon steel and AWS specification A5.9 for stainless steel. Wire electrodes are produced in diameters of 0.023, 0.030, 0.035, 0.045, and 0.062 inch. Other, larger diameters are available for production work and can include wire diameter sizes such as 1/16, 5/64, and 7/64 inch. Table 1.5 lists the most common sizes and the amperage ranges for these electrodes. The amperage will vary depending on the method of metal transfer, type of shielding gas, and base metal thickness. Some steel wire electrodes have a thin copper coating. This coating provides some protection to the electrode from rusting and improves the electrical contact between the wire electrode and the contact tube. It also acts as a lubricant to help the wire move more smoothly through the liner and contact tube. These electrodes may look like copper wire because of the very thin copper cladding. The amount of copper is so small that it

Table 1.4 AWS Filler Metal Specifications for Different Base Metals

Base Metal Type	AWS Filler Metal Specification
Aluminum and aluminum alloys	A5.10
Copper and copper alloys	A5.6
Magnesium alloys	A5.19
Nickel and nickel alloys	A5.14
Stainless steel (austenitic)	A5.9
Steel (carbon)	A5.18
Titanium and titanium alloys	A5.16

Table 1.5 Filler Metal Diameters and Amperage Ranges

	Electrode Diameter		Amperage
Base Metal	Inch	Millimeter	Range
Carbon Steel	0.023	0.6	35–190
	0.030	0.8	40–220
	0.035	0.9	60–280
	0.045	1.2	125–380
	1/16	1.6	275–450
Stainless Steel	0.023	0.6	40–150
	0.030	0.8	60–160
	0.035	0.9	70–120
	0.045	1.2	140–310
	1/16	1.6	280–450

either burns off or is diluted into the weld pool with no significant effect on the weld deposit.

Filler Metal Selection

Solid Wire

The AWS specification for carbon steel filler metals for gas-shielded welding wire is A5.18. Filler metal classified within this specification can be used for GMAW, gas tungsten arc welding (GTAW), and plasma arc welding (PAW) processes. Because in GTAW and PAW the wire does not carry the welding current, the letters *ER* are used as a prefix. The *ER* is followed by two numbers to indicate the minimum tensile strength of a good weld. The actual strength is obtained by adding three zeroes to the right of the number given. For example, ER70S-x is 70,000 psi.

The *S* located to the right of the tensile strength indicates that this is a solid wire. The last number—2, 3, 4, 5, 6, or 7—or the letter *G* is used to indicate the filler metal composition and the weld's mechanical properties, Figure 1.10.

ER70S-2

This is a deoxidized mild steel filler wire. The deoxidizers allow this wire to be used on metal that has light coverings of rust or oxides. There may be a slight reduction in the weld's physical properties if the weld is made on rust or oxides, but this reduction is only slight, and the weld will usually still pass the classification test standards. This is a general-purpose filler that can be used on killed, semikilled, and rimmed steels. Argon-oxygen, argon-CO_2, and CO_2 can be used as shielding gases. Welds can be made in all positions.

ER70S-3

This is a popular filler wire. It can be used in single or multiple-pass welds in all positions. ER70S-3 does not have the deoxidizers required to weld over rust, over oxides, or on rimmed steels. It produces high-quality welds on killed and semikilled steels. Argon-oxygen, argon-CO_2, and CO_2 can be used as shielding gases. This is the low-carbon steel filler most commonly used to weld galvanized steel.

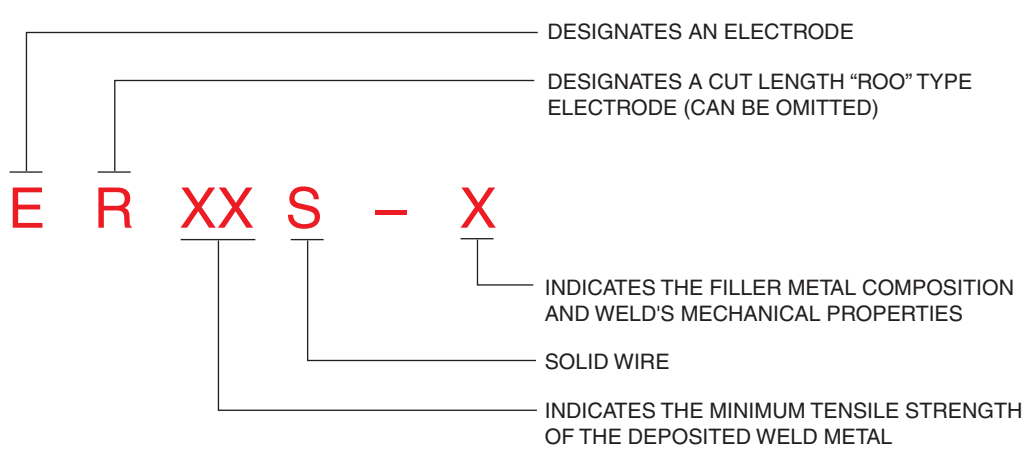

Figure 1.10 AWS numbering system for carbon steel filler metal for GMAW

ER70S-6

This is a good general-purpose filler wire. It has the highest levels of manganese and silicon. The wire can be used to make smooth welds on sheet metal or thicker sections. Welds over rust, oxides, and other surface impurities will lower the mechanical properties, but not normally below the specifications of this classification. Argon-oxygen, argon-CO_2, and CO_2 can be used as shielding gases. Welds can be made in all positions. This filler wire is not suggested for galvanized steel as its higher silicon content may induce cracking.

Stainless Steel Electrodes

The AWS specification for stainless steel–covered arc electrodes is A5.4 and for stainless steel bare, cored, and stranded electrodes and welding rods is A5.9. Filler metal classified within A5.4 uses the letter E as its prefix, and the filler metal within A5.9 uses the letters ER as its prefix.

Following the prefix, the three-digit stainless steel number from the American Iron and Steel Institute (AISI) is used. This number indicates the type of stainless steel in the filler metal, Table 1.6.

ALUMINUM AND ALUMINUM ALLOYS

The International Alloy Designation System is the most widely accepted naming scheme for aluminum alloys. Each alloy has a four-digit number, in which the first digit indicates the major alloying elements. Heat-treatable alloys can be strengthened by furnace operations after they are produced, while non-heat-treatable alloys are work hardened by plastic deformation such as bending and rolling operations. Heat treatment of aluminum alloys is sometimes referred to as precipitation hardening, and work hardening is sometimes referred to as strain hardening.

- 1000 series are essentially pure aluminum with a minimum 99% aluminum content by weight, and can be work hardened.
- 2000 series are alloyed with copper, and are heat treatable.
- 3000 series are alloyed with manganese, and can be work hardened.
- 4000 series are alloyed with silicon and can be work hardened.
- 5000 series are alloyed with magnesium, derive most of their strength from heat treatment, but can also be work hardened.
- 6000 series are alloyed with magnesium and silicon, are easy to machine, and are heat treatable, but not to the high strengths that 2000, 5000, and 7000 can reach.
- 7000 series are alloyed with zinc, and can be heat treated to the highest strengths of any aluminum alloy.
- 8000 series are a miscellaneous category where other elements not listed above may be used.

The AWS specifications for aluminum and aluminum alloy filler metals are A5.3 for covered arc welding electrodes and A5.10 for bare welding rods and electrodes. Filler metal classified within A5.3 uses the atomic symbol Al, and in A5.10 the prefix ER is used with the Aluminum Association number for the alloy, Table 1.7.

Aluminum Bare Welding Rods and Electrodes

ER1100

1100 aluminum has the lowest percentage of alloy agents of all the aluminum alloys, and it melts at 1215°F. The filler wire is also relatively pure. ER1100 produces welds that have good corrosion resistance and high ductility, with tensile strengths ranging from 11,000 to 17,000 psi. The weld deposit has a high resistance to cracking during welding. This wire can be used with oxyfuel gas welding (OFW), GTAW, and GMAW. Preheating to 300°F to 350°F is required for GTA welding on plate or pipe 3/8 in. and thicker to ensure good fusion. Flux is required for OFW. 1100 aluminum is commonly used for items such as food containers, food-processing equipment, storage tanks, and heat exchangers. ER1100 can be used to weld 1100 and 3003 grade aluminum.

ER4043

ER4043 is a general-purpose welding filler metal. It has 4.5% to 6.0% silicon added, which lowers its melting temperature to 1155°F. The lower melting temperature helps promote a free-flowing molten weld pool. The welds have high ductility and a high resistance to cracking during welding. This wire can be used with OFW, GTAW, and GMAW. Preheating to 300°F to 350°F is required for GTA welding on plate or pipe 3/8 in. and thicker to ensure good fusion. Flux is required for OFW. ER4043 can be used to weld on 2014, 3003, 3004, 4043, 5052, 6061, 6062, and 6063 and cast alloys 43, 355, 356, and 214.

ER5356

ER5356 has 4.5% to 5.5% magnesium added to improve the tensile strength. The weld has high ductility but only an average resistance to cracking during welding. This wire can be used for GTAW and GMAW. Preheating to 300°F to 350°F is required for GTA welding on plate or pipe 3/8 in. and thicker to ensure good fusion. ER5356 can be used to weld on 5050, 5052, 5056, 5083, 5086, 5154, 5356, 5454, and 5456.

ER5556

ER5556 has 4.7% to 5.5% magnesium and 0.5% to 1.0% manganese added to produce a weld with high strength. The weld has high ductility and only average resistance to cracking during welding. This wire can be used for GTAW and GMAW. Preheating to 300°F to 350°F is required for GTA welding on plate or pipe 3/8 in. and thicker to ensure good fusion. ER5556 can be used to weld on 5052, 5083, 5356, 5454, and 5456.

WIRE MELTING AND DEPOSITION RATES

The wire melting rates, deposition rates, and wire-feed speeds of the consumable wire welding processes are affected by the same variables. Before discussing them, however, these terms need to be defined. The wire melting rate, measured in inches per minute (in/min) or pounds per hour (lb/hr), is the rate at which the arc consumes the wire. The deposition rate, the measure of weld metal deposited, is nearly always less than the melting rate because not all of the wire is converted to

weld metal. Some is lost as slag, spatter, or fume. The amount of weld metal deposited in ratio to the wire used is called the deposition efficiency.

Deposition efficiencies depend on the process, on the gas used, and even on how the welder sets welding conditions. With efficiencies of approximately 98%, solid wires with argon-rich shield gas mixes are best. Some of the self-shielded cored wires are poorest, with efficiencies as low as 80%.

Welders can control the deposition rate by changing the current, electrode extension, and diameter of the wire. To obtain higher melting rates, they can increase the current or wire extension or decrease the wire diameter. Knowing the precise constants is unimportant. However, it is important to know that current greatly affects melting rate and that the electrode extension must be controlled if results are to be reproducible.

WELDING POWER SUPPLIES

To better understand the terms used to describe the different welding power supplies, you need to know the following electrical terms:

- Voltage, or volts (V), is a measurement of electrical pressure and is the force that causes the current (amperage) to flow, in the same way that pounds per square inch is a measurement of water pressure.
- Electrical potential means the same thing as voltage and is usually expressed by using the term *potential (P)*. The terms *voltage*, *volts*, and *potential* can all be interchanged when referring to electrical pressure.
- Amperage, or amps (A), is the measurement of the total number of electrons flowing, in the same way that gallons are a measurement of the amount of water flowing.
- Electrical current means the same thing as amperage and is usually expressed by using the term *current (C)*. The terms *amperage*, *amps*, and *current* can all be interchanged when referring to electrical flow.

GMAW power supplies are constant-voltage, constant-potential (CV, CP) machines, unlike shielded metal arc welding (SMAW) power supplies, which are constant-current (CC) machines and are sometimes called drooping arc voltage (DAV). It is impossible to make acceptable welds using the wrong type of power supply. GMAW power supplies are available as transformer rectifiers, motor generators, or inverters, Figure 1.11. Some newer machines use electronics, enabling them to supply both types of power at the flip of a switch; they may be referred to as CC/CV power supplies.

The relationships between current and voltage with different combinations of arc length or wire-feed speeds are called volt-ampere characteristics. The volt-ampere characteristics of arcs in argon with constant arc lengths or constant wire-feed speeds are shown in Figure 1.12. To maintain a constant arc length while increasing current, it is necessary to increase voltage. For example, with a 1/8-in. (3-mm) arc length, increasing current from 150 to 300 amperes requires increasing the

Figure 1.11 Transformer rectifier welding power supply
Source: Courtesy of ESAB Welding & Cutting Products

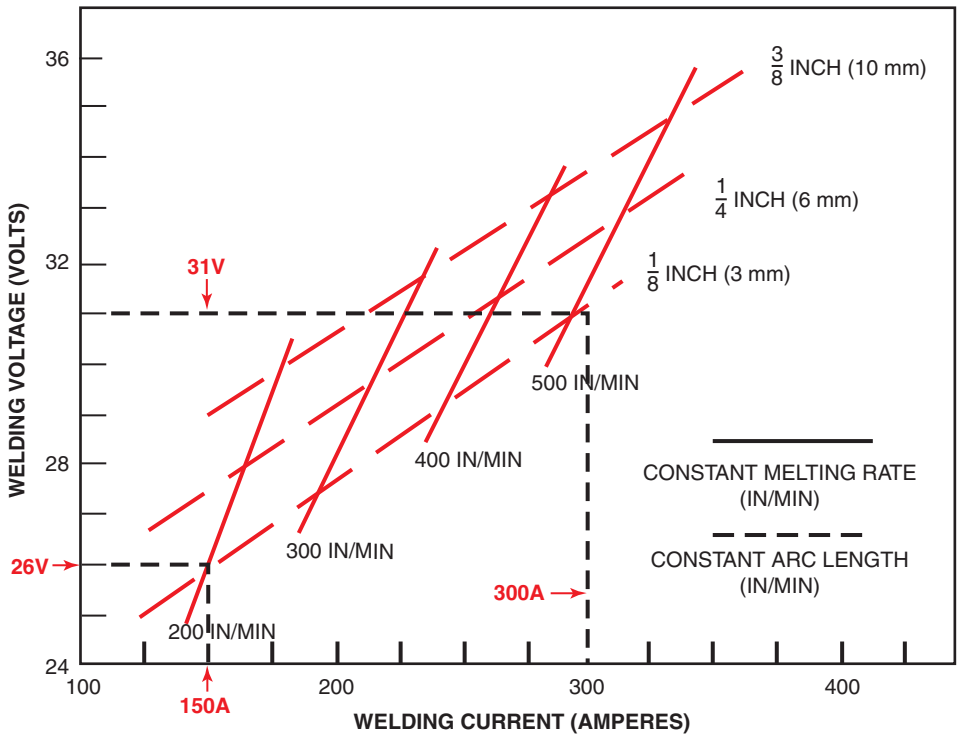

Figure 1.12 Volt-ampere characteristics of arcs with argon
The arc length and arc voltage are affected by the welding current and wire-feed speed (0.045-in. [1.43-mm] wire; 1-in. [25-mm] electrode extension).

voltage from about 26 to 31 volts. The current increase illustrated here results from increasing the wire-feed speed from 200 to 500 inches per minute.

Speed of the Wire Electrode

The wire-feed speed is generally recommended by the electrode manufacturer and is selected in inches per minute (ipm), or how fast the wire exits the contact tube. The welder uses a wire-speed control dial on the wire-feed unit to control ipm. It can be advanced or slowed to control the burn-off rate, or how fast the electrode transfers into the weld pool, to meet the welder's skill in controlling the weld pool. Note the direct relationship between current (amps) and wire-feed speed (wfs); as wire-feed speed increases, the amperage increases. If wire-feed speed is reduced, amperage will decrease, Table 1.8.

To accurately measure wire-feed ipm, snip off the wire at the contact tube. Wearing safety glasses and pointing the contact tube away from your face, squeeze the trigger for ten seconds; release and snip off the wire electrode. Measure the number of inches of wire that was fed out in the ten seconds. Now using basic shop math, multiply its total length in inches by six. The result is how many inches of wire were fed per minute.

Power Supplies for Short-circuiting Transfer

Although the GMA power source is said to have a constant voltage (CV) or constant potential (CP), it is not perfectly constant. The graph in Figure 1.13 shows that there is a slight decrease in voltage as the amperage increases within the working range. The rate of decrease is known as **slope**. It is expressed as the voltage decrease per 100-ampere increase—for example, 10 V/100 A. For short-circuiting welding, some welding power supplies are equipped to allow changes in the slope by steps or continuous adjustment.

The slope, which is called the *volt-ampere curve*, is often drawn as a straight line because it is fairly straight within the working range of the machine. Whether it is drawn as a curve or a straight line, the slope can be found by finding two points. The first point is the set voltage as read from the voltmeter when the gun switch is activated but no welding is

Table 1.8 Typical Amperages for Carbon Steel

Wire-feed Speed* in./min (m/min)	Wire Diameter Amperages			
	.030 in. (0.8 mm)	.035 in. (0.9 mm)	.045 in. (1.2 mm)	.062 in. (1.6 mm)
100 (2.5)	40	65	120	190
200 (5.0)	80	120	200	330
300 (7.6)	130	170	260	425
400 (10.2)	160	210	320	490
500 (12.7)	180	245	365	–
600 (15.2)	200	265	400	–
700 (17.8)	215	280	430	–

*To check feed speed, run out wire for one minute and then measure its length.

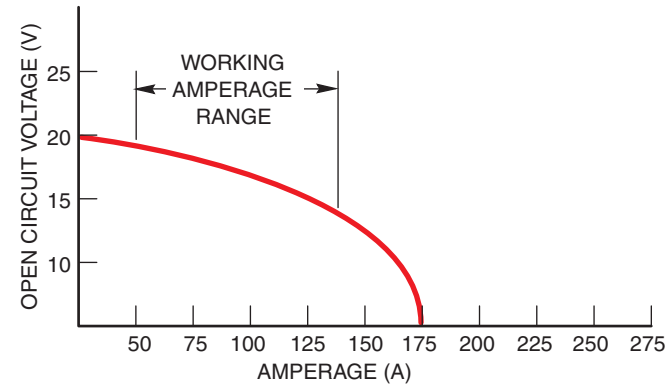

Figure 1.13 Constant potential welder slope

being done. This is referred to as the open circuit voltage. The second point is the voltage and amperage as read during a weld. The voltage control is not adjusted during the test but the amperage can vary. The slope is the voltage difference between the first and second readings. The difference can be found by subtracting the second voltage from the first voltage. Therefore, for settings over 100 amperes, it is easier to calculate the slope by adjusting the wire feed so that you are welding with 100 amperes, 200 amperes, 300 amperes, and so on. In other words, the voltage difference can be simply divided by 1 for 100 amperes, 2 for 200 amperes, and so forth.

The machine slope is affected by circuit resistance. Circuit resistance may result from a number of factors, including poor connections, long leads, or a dirty contact tube. A higher resistance means a steeper slope. In short-circuiting machines, increasing the inductance increases the slope. This increase slows the current's rate of change during short circuiting and the arcing intervals, Figure 1.14. Therefore, slope and inductance become synonymous in this discussion. As the slope increases, both the short-circuit current and the **pinch effect** are reduced. A flat slope has both an increased short-circuit current and a greater pinch effect.

The machine slope affects the short-circuiting metal transfer mode more than it does the other modes. Too much current and pinch effect from a flat slope cause a violent short and arc restart cycle, which results in increased spatter. Too little current and pinch effect from a steep slope result in the short circuit not being cleared as the wire freezes in the molten pool and piles up on the work, Figure 1.15.

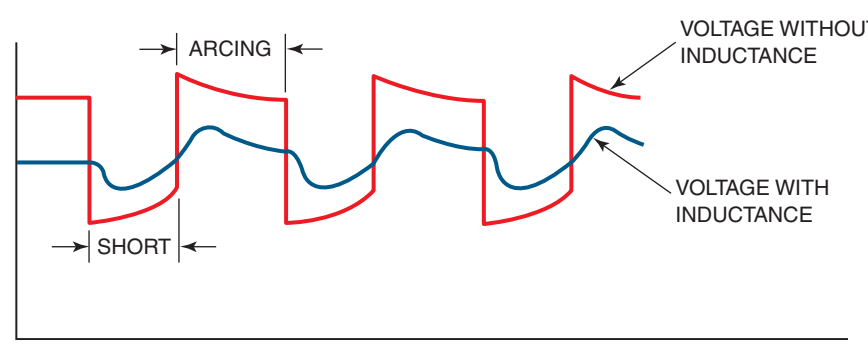

Figure 1.14 Voltage pattern with and without inductance

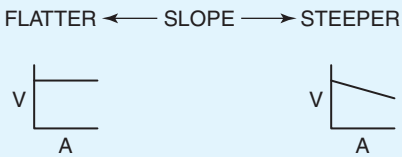

Buildup	Decreases	Increases
Depth of fusion	Increases	Decreases
Spatter	Increases	Decreases
Shorting	Violently cleared	May not clear

Figure 1.15 Effect of slope

The slope should be adjusted so that a proper spatter-free metal transfer occurs. On machines that have adjustable slopes, this is easily set. Experiment 2-2 describes a method of adjusting the circuit resistance to change the slope on machines that have a fixed slope. This is done by varying the contact tube-to-work distance. The GMA filler wire is much too small to carry the welding current and heats up due to its resistance to the current flow. The greater the tube-to-work distance, the greater the circuit resistance and the steeper the slope. By increasing or decreasing this distance, a proper slope can be obtained so that the short circuiting is smoother with less spatter.

Inductance

Many newer GMA welding power supplies are supplied with an inductance control. Inductance controls are used primarily in the GMA short-circuiting transfer mode, especially when open root joints are involved. Inductance controls the rise up to peak current during GMAW short-circuiting transfer. A low inductance control setting will provide a greater short-circuiting frequency, which may be beneficial in welding thin material where burn-through is an issue. A higher setting will reduce the frequency of short circuits, creating slightly longer arc periods, which allows greater current to go to the work. The greater current can be used to increase penetration in thicker sections or where complete penetration is required on open root joints welded from one side.

In the case of the GMA welding arc, the rate of change in the amperage relative to the arc voltage determines how the metal droplet detaches from the end of the electrode. If the rate of change is too rapid (inductance too low), the droplet detaches violently and produces excessive spatter. If the rate of change is too slow (inductance too high) the metal droplet doesn't detach cleanly and the arc is unstable. A midrange setting on the inductance control can be used for general-purpose short-circuit GMA welding and adjustments made for thick, thin, or open root conditions.

MOLTEN WELD POOL CONTROL

The GMAW molten weld pool can be controlled by varying the following factors: shielding gas, power settings, gun manipulation, travel speed, electrode extension, and gun angle.

Shielding Gas

The shielding gas selected for a weld has a definite effect on the weld produced. The properties that can be affected include the method of metal transfer, welding speed, weld contour, arc cleaning effect, and fluidity of the molten weld pool.

In addition to the effects on the weld itself, the metal to be welded must be considered in selecting a shielding gas. Some metals must be welded with an inert gas such as argon or helium or mixtures of argon and helium. Other metals weld more favorably with reactive gases such as carbon dioxide or with mixtures of inert gases and reactive gases such as argon and oxygen or argon and carbon dioxide, Table 1.9. The most commonly used shielding gases are 75% argon + 25% CO_2, argon + 1% to 5% oxygen, and carbon dioxide, Figure 1.16.

- **Argon:** The atomic symbol for argon is *Ar*, and it is an inert gas. *Inert gases* do not react with any other substance and are insoluble in molten metal. One hundred percent argon is used on nonferrous metals such as aluminum, copper, magnesium, nickel, and their alloys; but 100% argon is not normally used for making welds on ferrous metals.

  Because argon is denser than air, it effectively shields welds by pushing the lighter air away. Argon is relatively easy to ionize. Easily ionized gases can carry long arcs at lower voltages. This makes it less sensitive to changes in arc length.

  Argon gas is naturally found in all air and is collected in air separation plants. There are two methods of separating air to extract nitrogen, oxygen, and argon. In the cryogenic process, air is super-cooled to temperatures that cause it to liquefy and the gases are separated. In the noncryogenic process, molecular sieves (strainers with very small holes) separate the various gases, much like using a screen to separate sand from gravel.
- **Argon gas blends:** Oxygen, carbon dioxide, helium, and nitrogen can be blended with argon to change argon's welding characteristics. Adding reactive gases (oxidizing), such as oxygen or carbon dioxide, to argon tends to stabilize the arc, promote favorable metal transfer, and minimize spatter. As a result, the penetration pattern is improved, and undercutting is reduced or eliminated. Adding helium or nitrogen gases (nonreactive or inert) increases the arc heat for deeper penetration.

  The amount of the reactive gases, oxygen or carbon dioxide, required to produce the desired effects is quite small. As little as a half percent change in the amount of oxygen will produce a noticeable effect on the weld. Most of the time, blends containing 1% to 5% of oxygen are used. Carbon dioxide may be added to argon in the range of 10% to 30%. Blends of argon with less than 10% carbon dioxide may not have enough arc voltage to give the desired results. The most commonly used blend for short-circuiting transfer is 75% argon and 25% CO_2.

  When using oxidizing shielding gases with oxygen or carbon dioxide added, a suitable filler wire containing deoxidizers should be used to prevent porosity in the weld. The presence of oxygen in the shielding gas can also cause some loss of certain alloying

Table 1.9 (A) Metals Matched with GMAW Shielding Gases and Gas Blends; (B) GMAW Shielding Gases and Gas Blends Matched with Metals

GMAW Shielding Gas, Gas Blends, Metals, and Welding Process							
Gases/ Blend	**Gas Reaction**	**Application**	**Remarks**	**Gas/ Blend**	**Gas Reaction**	**Application**	**Remarks**
Ar + 5% CO_2	Oxidizing	Low-alloy steel	Pulse spray and short-circuit transfer in out-of-position welds	Ar + CO_2 + O_2	Oxidizing	Low-alloy steel and some stainless steels	All metal transfer for automatic and robotic applications
Ar + 10% CO_2	Oxidizing	Low-alloy steel	Same as above with a wider, more fluid weld pool	Ar + CO_2 + N	Almost inert	Stainless steel	All metal transfer, excellent for thin gauge material
Ar + 25% CO_2	Oxidizing	Mild, low-alloy steels and stainless steel	Smooth weld surface, reduces penetration with short-circuiting transfer	He + 7.5% Ar + 2.5% CO_2	Almost inert	Stainless steel and some low-alloy steels	Excellent toughness, arc stability, wetting characteristics, and bead contour, little spatter with short-circuiting transfer

(B)

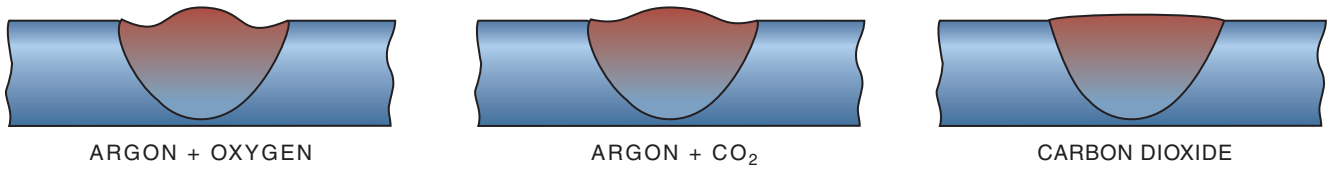

Figure 1.16 **Effect of shielding gas on weld bead shape**

elements, such as chromium, vanadium, aluminum, titanium, manganese, and silicon.

- **Helium:** The atomic symbol for helium is *He*, and it is an inert gas that is a product of the natural gas industry. It is removed from natural gas as the gas undergoes separation (fractionation) for purification or refinement.

  Helium is lighter than air; thus its flow rates must be about twice as high as argon's for acceptable stiffness in the gas stream to be able to push air away from the weld. Proper protection is difficult in drafts unless high flow rates are used. It requires a higher voltage to ionize, which produces a much hotter arc. There is a noticeable increase in both the heat and temperature of a helium arc. This hotter arc makes it easier to make welds on thick sections of aluminum and magnesium.

  Small quantities of helium are blended with heavier gases. These blends take advantage of the heat produced by the lightweight helium and weld coverage by the other heavier gas. Thus, each gas is contributing its primary advantage to the blended gas.
- **Carbon dioxide:** Carbon dioxide is a compound made up of one carbon atom (C) and two oxygen atoms (O_2), and its molecular formula is CO_2. One hundred percent carbon dioxide is widely used as a shielding gas for GMA welding of steels. In the short-circuiting transfer mode it allows higher welding speed, better penetration, good mechanical properties, and costs less than the inert gas mixes. The chief drawback in the use of carbon dioxide is the less-steady arc characteristics and a considerable increase in weld spatter. The spatter can be kept to a minimum by maintaining a very short, uniform arc length and strict attention to amperage and voltage parameters. CO_2 can produce sound, spatter-free welds of the highest quality, provided established procedures are followed and a filler wire having the proper deoxidizing additives is selected.
- **Nitrogen:** The atomic symbol for nitrogen is *N*. It is not an inert gas but is relatively nonreactive to the molten weld pool. It is often used in blended gases to increase the arc's heat and temperature. One hundred percent nitrogen can be used to weld copper and copper alloys and is an economical choice for gas purging of some austenitic stainless steel pipe welds.

Power Settings

As the power settings, voltage, and amperage are adjusted, the weld bead is affected. Making an acceptable weld requires a balancing of the voltage and amperage. If either or both are set too high or too low, the

Module 5
Key Indicator 3, 8

weld penetration can decrease. A GMA welding machine has no direct amperage settings. Instead, the amperage at the arc is adjusted by changing the wire-feed speed. As a result of the welding machine's maintaining a constant voltage when the wire-feed speed increases, more amperage flows across the arc. This higher amperage is required to melt the wire so that the same arc voltage can be maintained. The higher amperage is used to melt the filler wire and does not increase the penetration. In fact, the weld penetration may decrease significantly.

Increasing and decreasing the voltage changes the arc length; however, it may not put more heat into the weld. Like changes in the amperage, these voltage changes may decrease weld penetration.

Gun Manipulation

The GMA welding process is greatly affected by the location of the electrode tip and molten weld pool. During the short-circuiting process if the arc is directed to the base metal and outside the molten weld pool, the welding process may stop. Without the resistance of the hot molten metal, high-amperage surges occur each time the electrode tip touches the base metal, resulting in a loud pop and a shower of sparks. It is something that occurs each time a new weld is started. So when making a weave pattern, which is a gun manipulation technique of moving side to side in order to produce a wider weld bead, you must keep the arc and electrode tip directed into the molten weld pool. Other than the sensitivity to arc location, most of the SMAW weave patterns that keep the electrode wire at or near the leading edge of the weld pool can be used for short-circuiting GMA welds.

Travel Speed

Because the location of the arc inside the molten weld pool is important, the welding travel speed cannot exceed the ability of the arc to melt the base metal. Too high a travel speed can result in overrunning of the weld pool and an uncontrollable arc. Fusion between the base metal and filler metal can completely stop if the travel rate is too fast. If the travel rate is too slow and the weld pool size increases excessively, it can also restrict fusion to the base plate.

Electrode Extension

Module 5
Key Indicator 3, 8

The **electrode extension (stickout)** is the distance from the contact tube to the arc measured along the wire. Adjustments in this distance cause a change in the wire resistance and the resulting weld bead, Figure 1.17.

GMA welding currents are relatively high for the wire sizes, even for the low current values used in short-circuiting arc metal transfer, Figure 1.18. As the length of wire extending from the contact tube to the work increases, the voltage, too, should increase. Since this change is impossible with a constant-voltage power supply, the system compensates by reducing the current. In other words, by increasing the electrode extension and maintaining the same wire-feed speed, the current has to change to provide the same resistance drop. This situation leads to a reduction in weld heat, penetration, and fusion, and an increase in buildup. On the

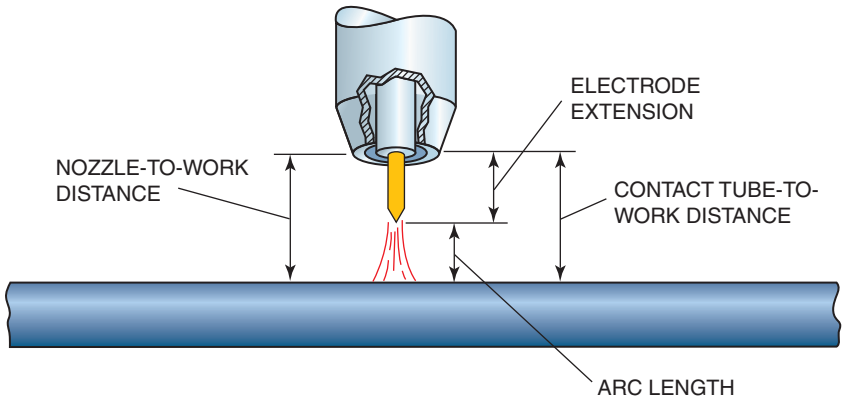

Figure 1.17 Electrode-to-work distances

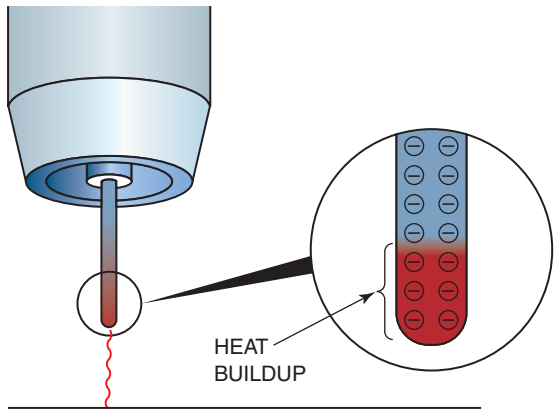

Figure 1.18 Heat buildup due to the extremely high current for the small conductor (electrode)

other hand, as the electrode extension distance is shortened, the weld heats up, penetrates more, and builds up less, Figure 1.19.

Experiment 2-3 explains the technique of using varying extension lengths to change the weld characteristics. Using this technique, a welder can make acceptable welds on metal ranging in thickness from 16 gauge to 1/4 in. (6 mm) or more without changing the machine settings. When using this technique, the nozzle-to-work distance should be kept the same so that enough shielding gas coverage is provided. Some nozzles can be extended to provide coverage. Others must be exchanged with the correct-length nozzle, Figure 1.20.

Gun Angle

Module 5
Key Indicator 3, 8

The GMA welding gun may be held so that the relative angle between the gun, work, and welding bead being made is either vertical or has a drag angle or a push angle. Changes in this angle will affect the weld bead. The effect is most noticeable during the short-circuiting arc and globular transfer modes.

Backhand welding is the welding technique that uses a drag angle, Figure 1.21. The welding technique that uses a push angle is known as forehand welding, Figure 1.22.

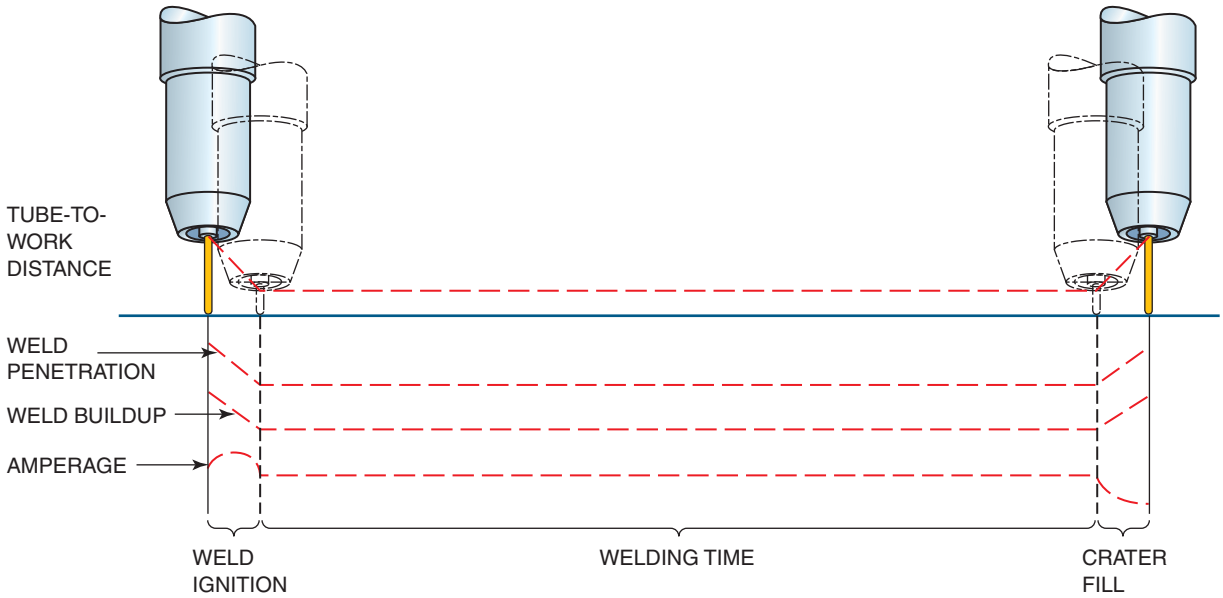

Figure 1.19 Using the changing tube-to-work distance to improve both the starting and stopping points of a weld

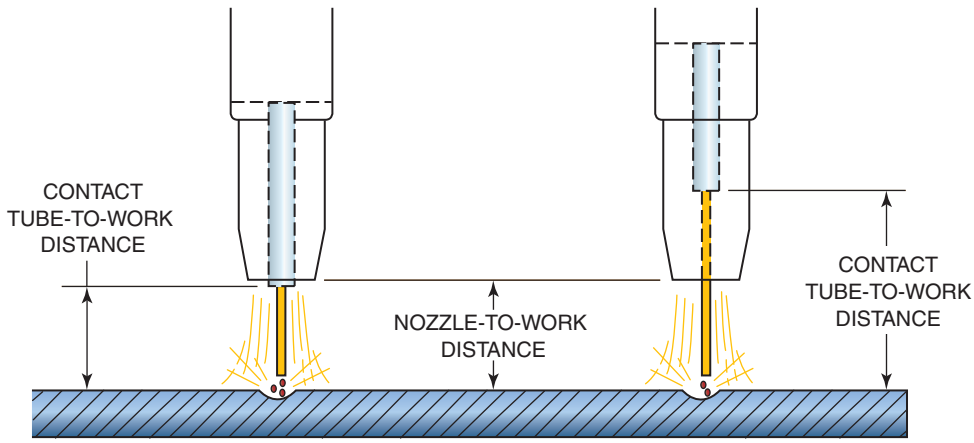

Figure 1.20 Nozzle-to-work distance can differ from the contact tube-to-work distance

Figure 1.21 Backhand welding, or drag angle
Source: Courtesy of Larry Jeffus

Figure 1.22 Forehand welding, or push angle
Source: Courtesy of Larry Jeffus

Backhand Welding

Module 5
Key Indicator 3, 8

A dragging angle, or backhand, welding technique directs the arc force into the molten weld pool of metal. This action, in turn, forces the molten metal back onto the trailing edge of the molten weld pool and exposes more of the unmelted base metal, Figure 1.23. The digging action pushes the penetration deeper into the base metal while building up the weld head. If the weld is sectioned, the profile of the bead is narrow and deeply penetrated, with high buildup.

Forehand Welding

Module 5
Key Indicator 3, 8

In a push angle, or forehand, welding technique, the arc force pushes the weld metal forward and out of the molten weld pool onto the cooler metal ahead of the weld, Figure 1.24. The heat and metal are spread out over a wider area. The sectional profile of the bead is wide, showing shallow penetration with little buildup.

The greater the angle, the more defined is the effect on the weld. As the angle approaches vertical, the effect is reduced. This allows the welder to change the weld bead as effectively as the changes resulting from adjusting the machine current settings.

EQUIPMENT

Module 5
Key Indicator 3, 8

The basic GMAW equipment consists of the gun, electrode (wire) feed unit, electrode (wire) supply, power source, shielding gas supply with flowmeter/regulator, control circuit, and related hoses, liners, and cables, Figure 1.25 and Figure 1.26. Larger, more complex systems may have water for cooling, solenoids for controlling gas flow, and carriages for moving the work or the gun or both, Figure 1.27. The system may be

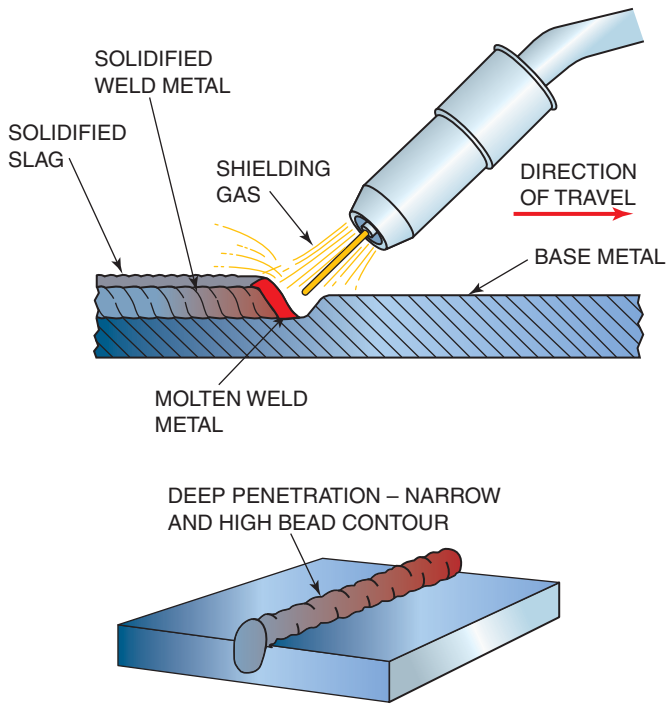

Figure 1.23 Backhand welding, or dragging angle

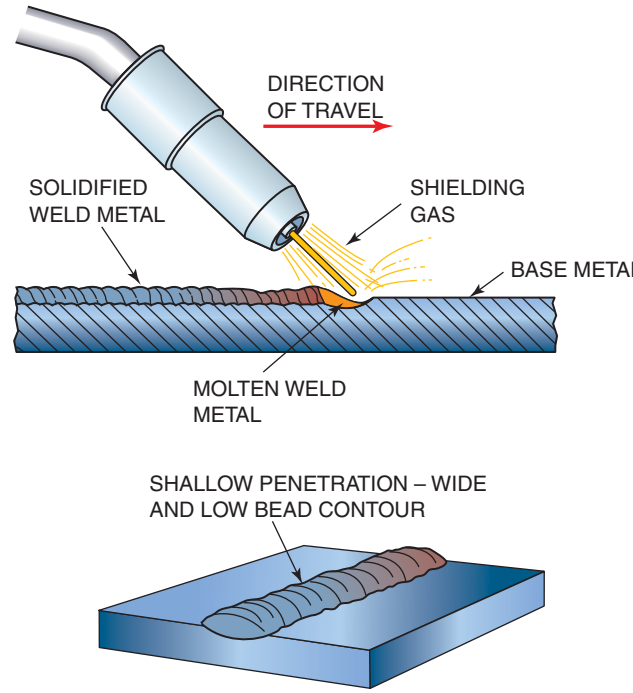

Figure 1.24 Forehand welding, or pushing angle

stationary or portable, Figure 1.28. In most cases, the system is meant to be used for only one process. Some manufacturers, however, do make power sources that can be switched over for other uses.

Power Source

The power source may be either a transformer rectifier, inverter, or generator type. The transformers are stationary and commonly require a three-phase power source. The inverter power sources are smaller, lighter, and may be designed to accept a variety of different electrical inputs, from 208 volts to 440 volts, single or three-phase. Engine generators are ideal for portable use or where sufficient power is not available.

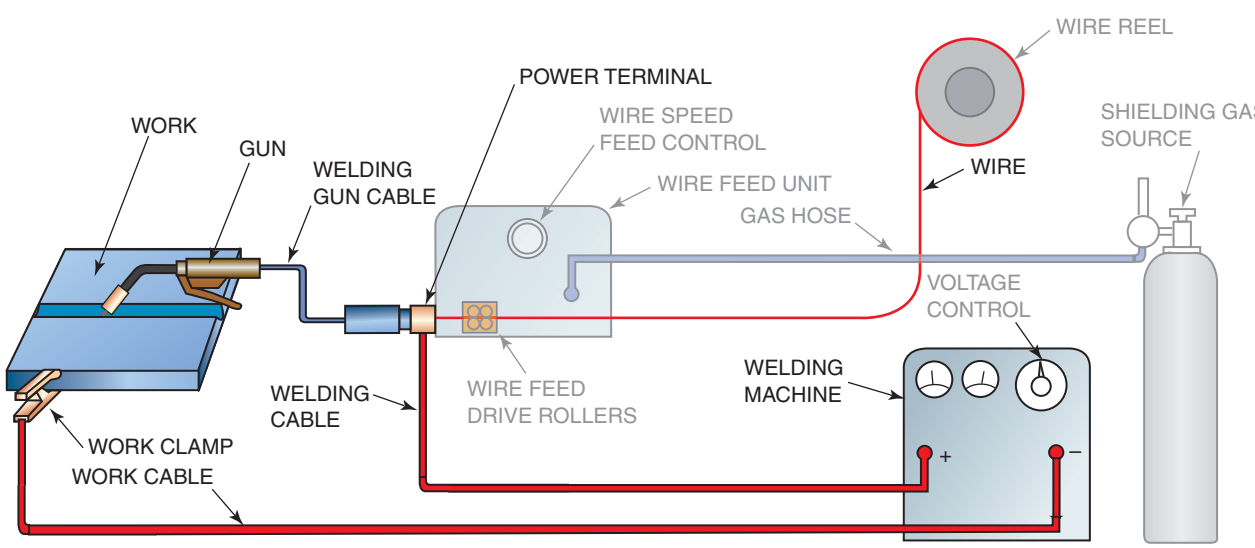

Figure 1.25 Schematic of equipment setup for GMA welding
Source: Courtesy of Hobart Brothers Company

Figure 1.26 Small 110-V GMA welder
Source: Courtesy of Thermal Arc, a Thermadyne Company

Figure 1.27 Robot welding
Source: Courtesy of ESAB Welding & Cutting Products

Figure 1.28 Portable water cooler for GMA welding equipment
Source: Courtesy of Lincoln Electric Company

Typical GMA welding machines produce a DC welding current ranging from 40 amperes to 600 amperes with 10 volts to 40 volts, depending upon the machine. In the past, some GMA processes used AC welding current, but DCRP is used almost exclusively now. Typical power supplies are shown in Figure 1.29.

Because many GMAW power supplies are used in automation and require long periods of continuous use, it is not unusual for GMA welding machines to have a 100% duty cycle. This allows the machine to be run continuously at its highest-rated output without damage.

Figure 1.29 Two power supplies for multipurpose GMAW applications
(A) An expensive 200-ampere constant-voltage power supply, and (B) a 650-ampere constant-voltage and constant-current power supply.
Source: (A) Courtesy of ESAB Welding & Cutting Products (B) Courtesy of Lincoln Electric Company

Electrode (Wire) Feed Unit

The purpose of the electrode feeder is to provide a steady and reliable supply of wire to the weld. Slight changes in the rate at which the wire is fed have distinct effects on the weld.

The motor used in a feed unit is usually a DC type that can be continuously adjusted over the desired range. Figure 1.30 and Figure 1.31 show typical wire-feed units and accessories.

Push-type Feed System

The wire rollers are clamped securely against the wire to provide the necessary friction to push the wire through the conduit to the gun. The pressure applied on the wire can be adjusted. A groove is provided in the roller to aid in alignment and to lessen the chance of slippage. Most manufacturers provide rollers with smooth or knurled U-shaped or V-shaped grooves, Figure 1.32. Knurling (a series of ridges cut into the groove) helps grip larger-diameter wires so that they can be pushed along more easily. Soft wires, such as aluminum, are easy to damage if knurled rollers are used. Soft wires are best used with U-grooved rollers. Even V-grooved rollers can distort the surface of the wire, causing problems. V-grooved rollers are best suited for hard wires, such as mild steel and stainless steel. It is also important to use the correct-size grooves in the rollers.

Variations of the push-type electrode wire feeder include the pull type and push-pull type. The difference is in the size and location of the drive rollers. In the push-type system, the electrode must have enough strength to be pushed through the conduit without kinking. Mild steel and stainless steel can be readily pushed 15 to 20 ft (4 to 6 m), but aluminum is much harder to push more than 10 ft (3 m).

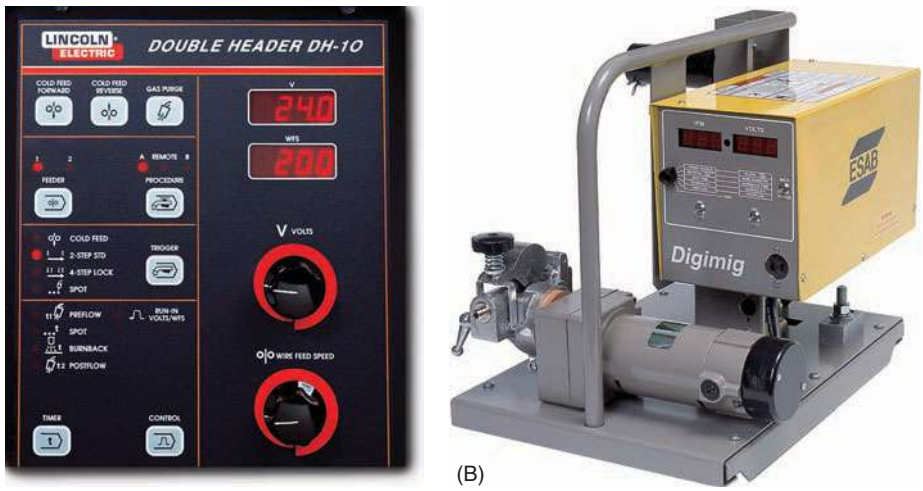

(A) (B)

Figure 1.30 Examples of wire feeders
(A) A 90-ampere power supply and wire feeder for welding sheet steel with carbon dioxide shielding. (B) Modern wire feeder with digital preset and readout of wire-feed speed and closed-loop control.
Source: (A) Courtesy of Lincoln Electric Company (B) Courtesy of ESAB Welding & Cutting Products

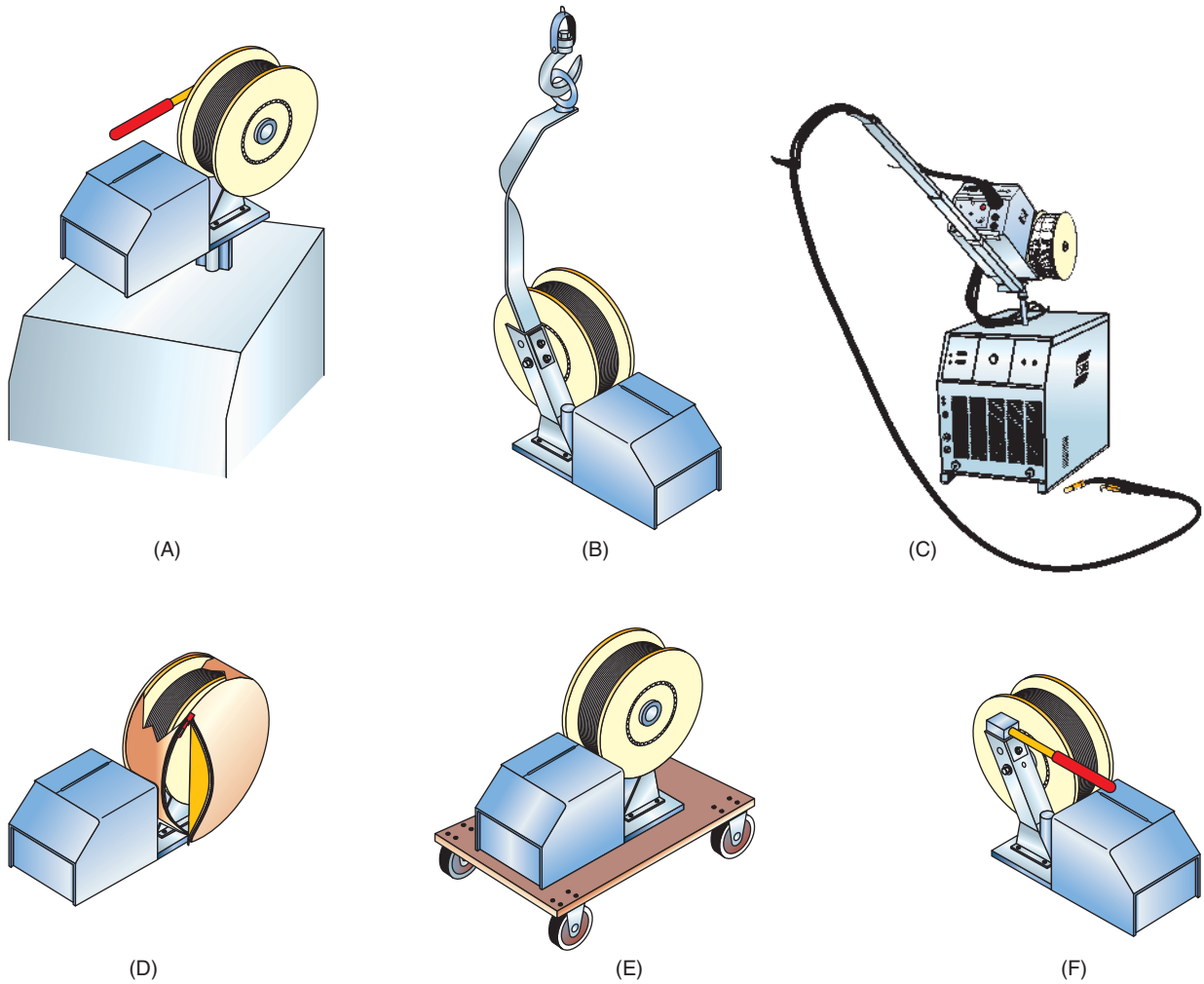

Figure 1.31 A variety of accessories are available for most electrode feed systems.
(A) Swivel post, (B) boom hanging bracket, (C) counterbalance mini-boom, (D) spool cover, (E) wire feeder wheel cart, and (F) carrying handle.

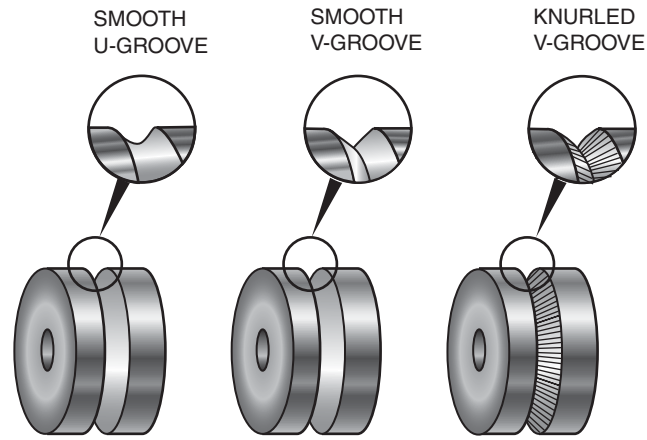

Figure 1.32 Feed rollers

Pull-type Feed System

In pull-type systems, a smaller but higher-speed motor is located in the gun to pull the wire through the conduit. Using this system, it is possible to move even soft wire over great distances. The disadvantages are that

the gun is heavier and more difficult to use, rethreading the wire takes more time, and the operating life of the motor is shorter.

Push-pull-type Feed System

Push-pull-type feed systems use a synchronized system with feed motors located at both ends of the electrode conduit, Figure 1.33. This system can be used to move any type of wire over long distances by periodically installing a feed roller into the electrode conduit. Compared to the pull-type system, the advantages of this system include the ability to move wire over longer distances, faster rethreading, and increased motor life due to the reduced load. A disadvantage is that the system is more expensive.

Linear Electrode Feed System

Linear electrode feed systems use a different method to move the wire and change the feed speed. Standard systems use rollers that pinch the wire between the rollers. A system of gears is used between the motor and rollers to provide roller speed within the desired range. The linear feed system does not have gears or conventional-type rollers.

The linear feed system uses a small motor with a hollow armature shaft through which the wire is fed. The rollers are attached so that they move around the wire. Changing the roller pitch (angle) changes the speed at which the wire is moved without changing the motor speed.

Figure 1.33 Wire-feed system that enables the wire to be moved through a longer cable
Source: Courtesy of Lincoln Electric Company

This system works in the same way that changing the pitch on a screw, either coarse threads or fine threads, affects the rate that the screw will move through a spinning nut.

The advantage of a linear system is that the bulky system of gears is eliminated, thus reducing weight, size, and wasted power. The motor operates at a constant high speed where it is more efficient. The reduced size allows the system to be housed in the gun or within an enclosure in the cable. Several linear wire feeders can be synchronized to provide an extended operating range. The disadvantage of a linear system is that the wire may become twisted as it is moved through the feeder.

Spool Gun

A spool gun is a compact, self-contained system consisting of a small drive system and a wire supply, Figure 1.34A. This system allows the welder to move freely around a job with only a power lead and shielding gas hose to manage. The major control system is usually mounted on the welder. The feed rollers and motor are found in the gun just behind the

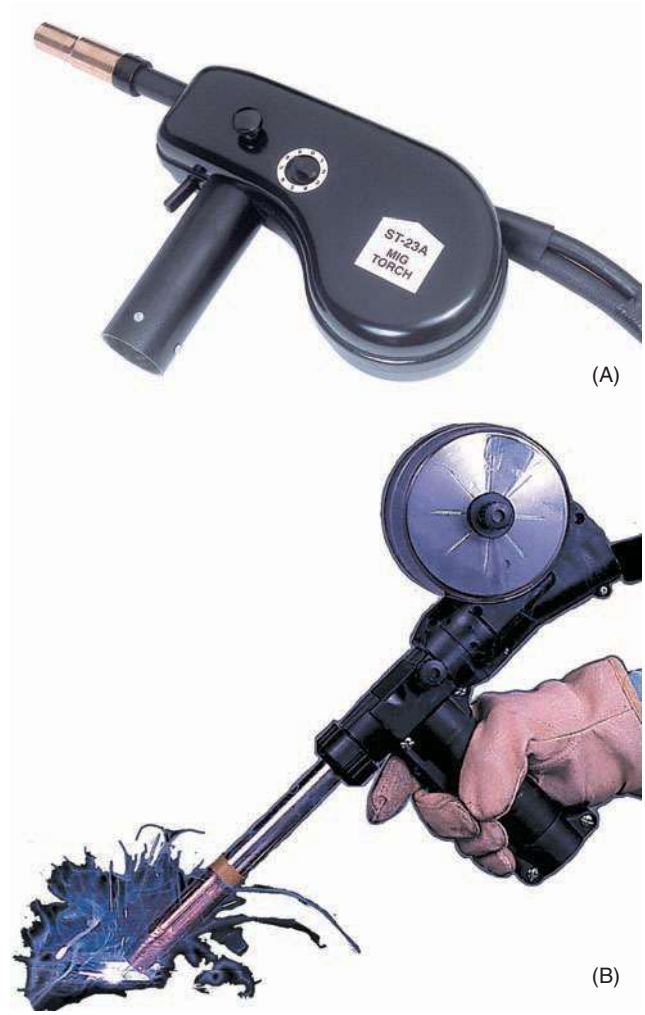

Figure 1.34 Feeder/guns for GMA welding
Source: Courtesy of ESAB Welding & Cutting Products

nozzle and contact tube, Figure 1.34B. Because of the short distance the wire must be moved, very soft wires (aluminum) can be used. A small spool of welding wire is located just behind the feed rollers. The small spools of wire required in these guns are often very expensive. Although the guns are small, they feel heavy when being used.

Electrode Conduit

The electrode conduit or liner guides the welding wire from the feed rollers to the gun. It may be encased in a lead that contains the shielding gas.

Power cable and gun switch circuit wires are contained in a conduit that is made of a tightly wound coil having the needed flexibility and strength. The steel conduit may have a nylon or Teflon liner to protect soft, easily scratched metals, such as aluminum, as they are fed.

If the conduit is not an integral part of the lead, it must be firmly attached to both ends of the lead. Failure to attach the conduit can result in misalignment, which causes additional drag or makes the wire jam completely. If the conduit does not extend through the lead casing to make a connection, it can be drawn out by tightly coiling the lead, Figure 1.35. Coiling will force the conduit out so that it can be connected. If the conduit is too long for the lead, it should be cut off and filed smooth. Too long a lead will bend and twist inside the conduit, which may cause feed problems.

Welding Gun

The welding gun attaches to the end of the power cable, electrode conduit, and shielding gas hose, Figure 1.36. It is used by the welder to produce the weld. A trigger switch is used to start and stop the weld cycle. The gun also has a contact tube, which is used to transfer welding current to the electrode moving through the gun, and a gas nozzle, which directs the shielding gas onto the weld, Figure 1.37.

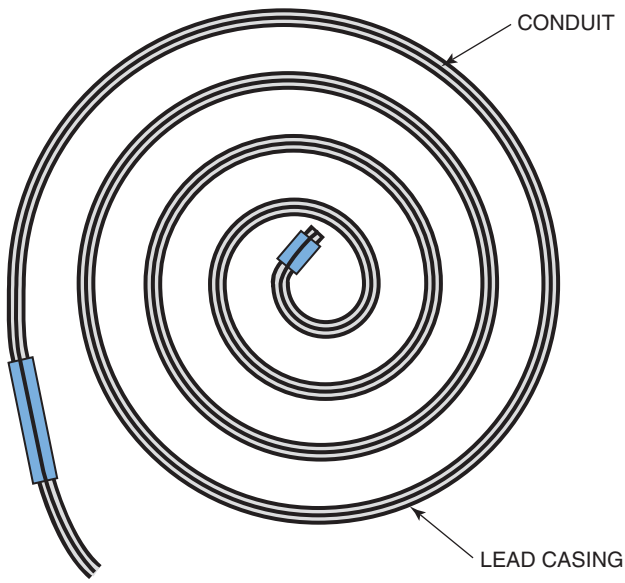

Figure 1.35 Tightly coiled lead casing will force the liner out of the gun

Figure 1.36 A typical GMA welding gun
Guns like this are used for most welding processes with a heat shield attached to protect the welder's gloved hand from intense heat generated when welding with high amperages.
Source: Courtesy of Tweco, a Thermadyne Company

SPOT WELDING

GMA can be used to make high-quality arc spot welds. Welds can be made using standard or specialized equipment, Figure 1.38. The arc spot weld produced by GMAW differs from electric resistance spot welding. The GMAW spot weld starts on one surface of one member and burns through to the other member, Figure 1.39. Fusion between the members occurs, and a small nugget is left on the metal surface.

GMA spot welding has some advantages such as the following: (1) welds can be made in thin-to-thick materials; (2) the weld can be made when only one side of the materials to be welded is accessible; and (3) the weld can be made when there is paint on the interfacing surfaces. The arc spot weld can also be used to assemble parts for welding to be done at a later time.

Thin metal can be attached to thicker sections using an arc spot weld. If a thin-to-thick butt, lap, or tee joint is to be welded with complete joint penetration, often the thin material will burn back, leaving a hole, or there will not be enough heat to melt the thick section. With an arc spot weld, the burning back of the thin material allows the thicker metal to be melted. As more metal is added to the weld, the burn-through is filled, Figure 1.39.

Caution

Safety glasses and/or flash glasses must be worn to protect the eyes from flying sparks.

The GMA spot weld is produced from only one side. Therefore, it can be used on awkward shapes and in cases where the other side of the surface being welded should not be damaged. This makes it an excellent process for auto body repair. In addition, because the metals are melted

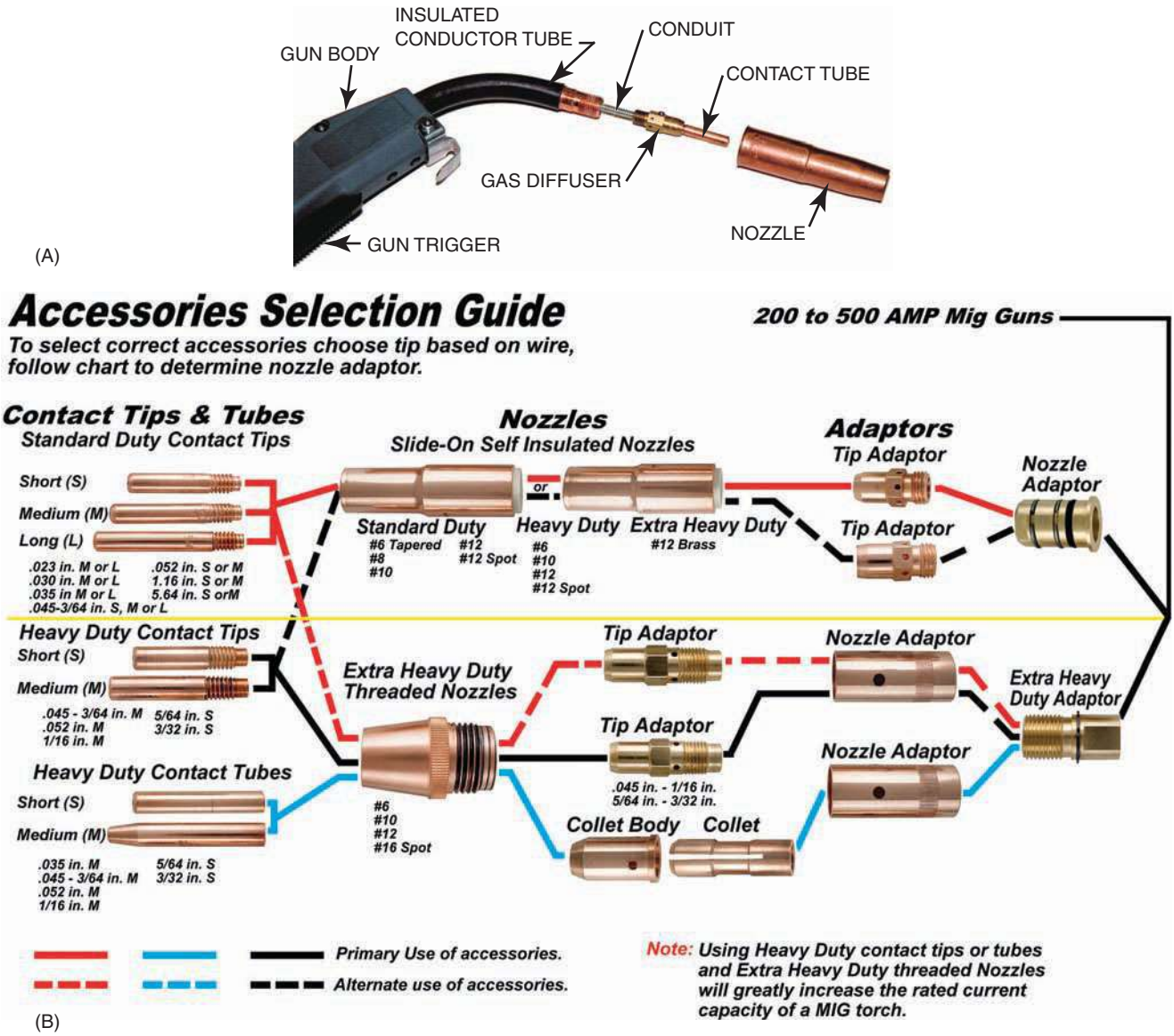

Figure 1.37 GMA welding gun parts
(A) Typical replaceable parts of a GMA welding gun. (B) Accessories and parts selection guide for a GMA welding gun.
Source: Courtesy of ESAB Welding & Cutting Products

and the molten weld pool is agitated, thin films of paint between the members being joined need not be removed. This is an added benefit for auto body repair work.

Specially designed nozzles provide flash protection, part alignment, and arc alignment, Figure 1.40. As a result, for some small jobs it is possible to perform the weld with only safety glasses. The optional control timer provides weld time and burn-back time. To make a weld, the amperage, voltage, and length of welding time must be set correctly. The burn-back time is a short period at the end of the weld when the wire feed stops but the current does not. This allows the wire to be burned back so it does not stick in the weld, Figure 1.39D.

Caution

It is not advisable for any spot welding work requiring more than just a few spot welds to be done without full welder's safety gear. Prolonged exposure to the reflected ultraviolet light will cause skin burns.

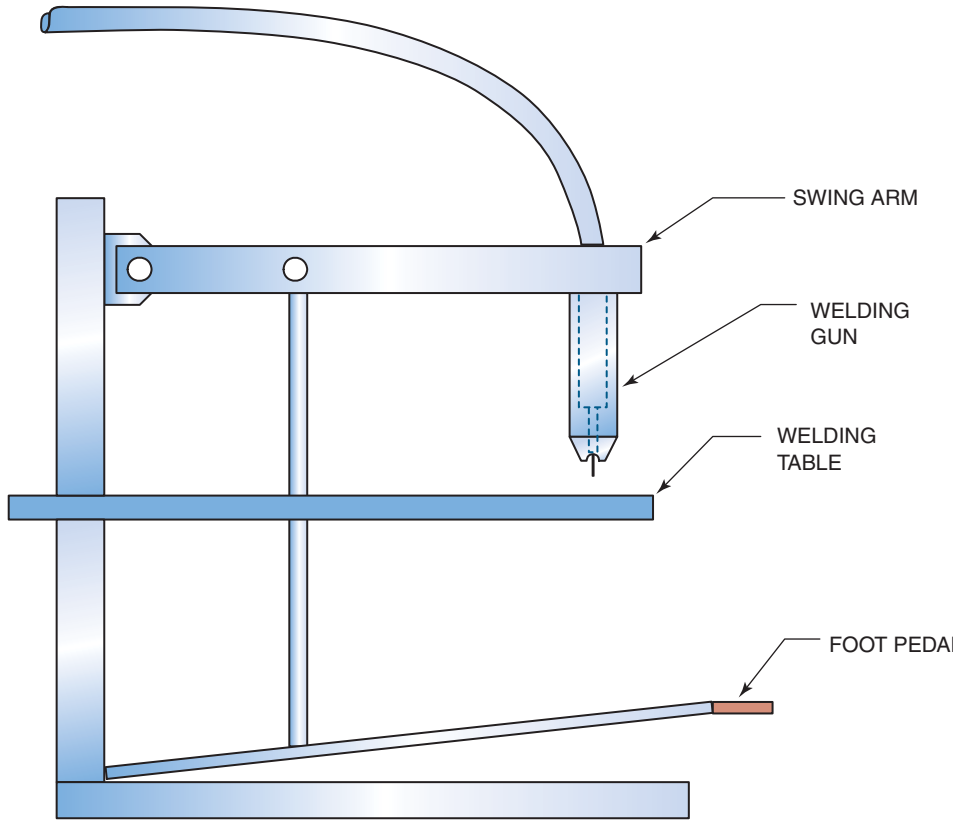

Figure 1.38 GMA spot welding machine

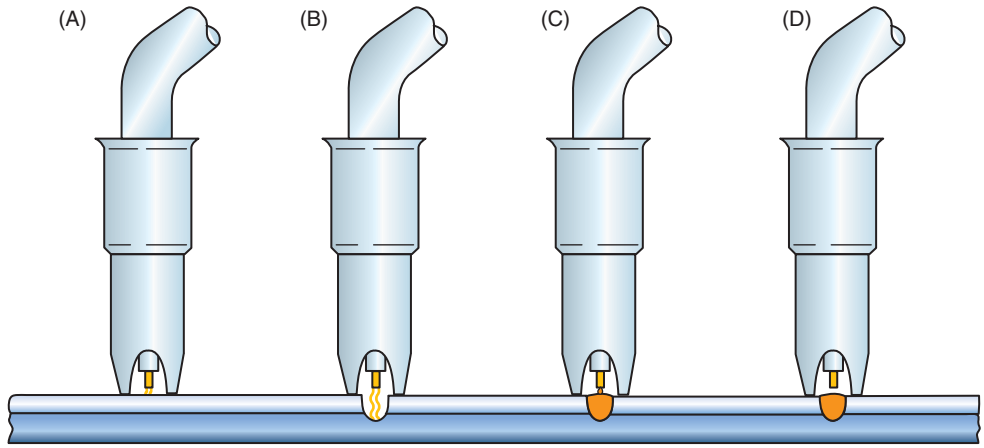

Figure 1.39 A GMA spot weld
(A) The arc starts, (B) a hole is burned through the first plate, (C) the hole is filled with weld metal, and (D) the wire feed stops and the arc burns the electrode back.

FACE, EYE, AND EAR PROTECTION

Module 2
Key Indicator 1

Face and Eye Protection

Module 5
Key Indicator 1

Eye protection must be worn in the shop at all times. Eye protection can be safety glasses with side shields, Figure 1.41, goggles, or a full face shield. To give better protection when working in brightly lit areas or outdoors, some welders wear flash glasses, which are special, lightly tinted, safety glasses. These safety glasses provide protection from both flying debris and reflected light.

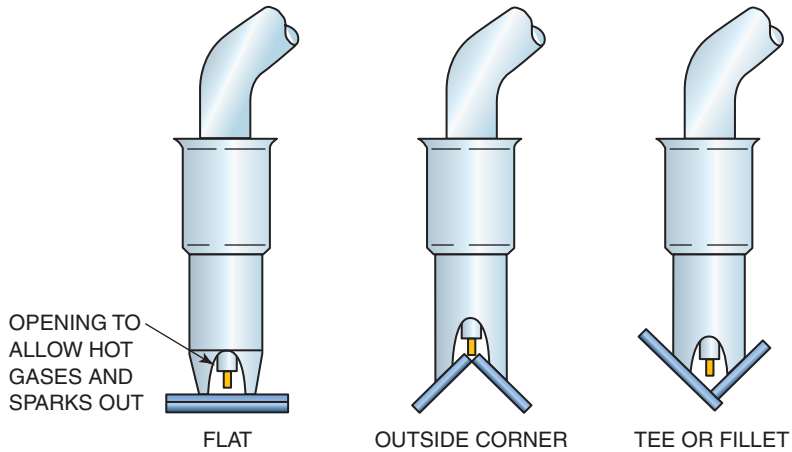

Figure 1.40 Specialized nozzles for GMA spot welding

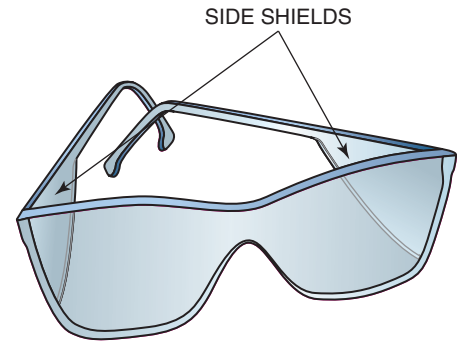

Figure 1.41 Safety glasses with side shields

Suitable eye protection is important because eye damage caused by excessive exposure to arc light is not noticed. Welding light damage often occurs without warning, like a sunburn's effect that is felt the following day. Therefore, welders must take appropriate precautions in selecting filters or goggles that are suitable for the process being used, Figure 1.42. Selecting the correct shade lens is also important, because both extremes of too light or too dark can cause eye strain. New welders often select too dark a lens, assuming it will give them better protection, but this results in eye strain in the same manner as if they were trying to read in a poorly lit room. In reality, any approved arc welding lenses will filter out the harmful ultraviolet light. Select a lens that lets you see comfortably. At the very least, the welder's eyes must not be strained by excessive glare from the arc.

Ultraviolet light can burn the eye in two ways. This light can injure either the white of the eye or the retina, which is the back of the eye, Figure 1.43. Burns on the retina are not painful but may cause some loss of eyesight. The whites of the eyes are very sensitive, and burns are very painful. The eyes are easily infected because, as with any burn, many cells are killed. These dead cells in the moist environment of the eyes will promote the growth of bacteria that cause infection. When the eye is burned, it feels as though there is something in the eye. Without a professional examination, however, it is impossible to tell if there is something in the eye. Because there may actually be something in the eye and because of the high risk of infection, home remedies or other medicines should never be used for eye burns. Anytime you receive an eye injury you should see a doctor.

Welding Helmets

Even with quality **welding helmets**, like those shown in Figure 1.44, the welder must check for potential problems that may occur from accidents or daily use. Small, undetectable leaks of ultraviolet light in an arc welding helmet can cause a welder's eyes to itch or feel sore after a day of welding. To prevent these leaks, make sure the lens gasket is installed correctly, Figure 1.45. The outer and inner clear lenses must be plastic. As shown in Figure 1.46, a lens can be checked for cracks by twisting it between your fingers. Worn or cracked spots on a helmet must be repaired.

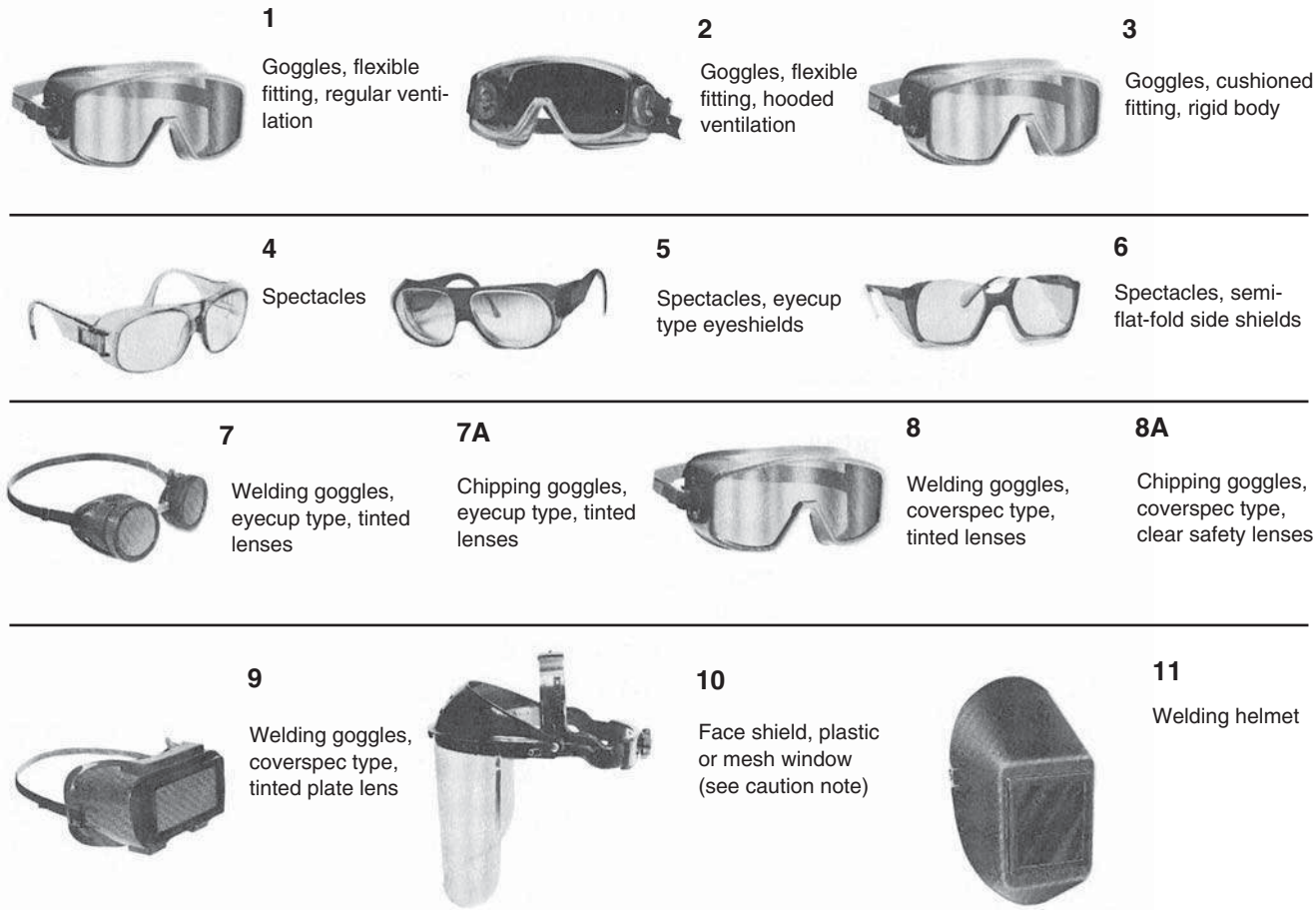

Non-side-shield spectacles are available for limited hazard use requiring only frontal protection.

Applications

Operation	Hazards	Protectors
Acetylene-burning Acetylene-cutting Acetylene-welding	Sparks, harmful rays, molten metal, flying particles	7,8,9
Chemical handling	Splash, acid burns, fumes	2 (for severe exposure add 10)
Chipping	Flying particles	1,2,4,5,6,7A,8A
Electric (arc) welding	Sparks, intense rays, molten metal	11 (in combination with 4,5,6 in tinted lenses advisable)
Furnace operations	Glare, heat, molten metal	7,8,9 (for severe exposure add 10)
Grinding—light	Flying particles	1,3,5,6 (for severe exposure add 10)
Grinding—heavy	Flying particles	1,3,7A,8A (for severe exposure add 10)
Laboratory	Chemical splash, glass breakage	2 (10 when in combination with 5,6)
Machining	Flying particles	1,3,5,6 (for severe exposure add 10)
Molten metals	Heat, glare, sparks, splash	7,8 (10 in combination with 5,6 in tinted lenses)
Spot welding	Flying particles, sparks	1,3,4,5,6 (tinted lenses advisable, for severe exposure add 10)

CAUTION:
Face shields alone do not provide adequate protection. Plastic lenses are advised for protection against molten metal splash.
Contact lenses, of themselves, do not provide eye protection in the industrial sense and shall not be worn in a hazardous environment without appropriate covering safety eyewear.

Figure 1.42 Huntsman selector chart
Source: Courtesy of Kedman Co., Huntsman Product Division

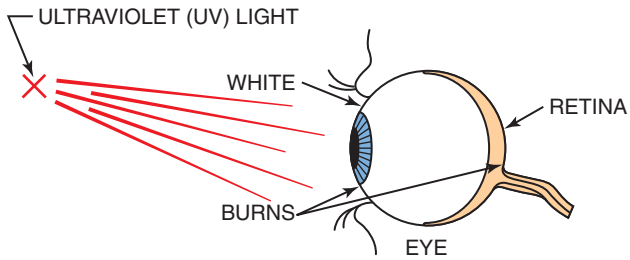

Figure 1.43 The eye can be burned on the white or on the retina by ultraviolet light

Figure 1.44 Typical arc welding helmets for eye and face protection during welding
Source: Courtesy of Hornell, Inc.

Safety Glasses

Safety glasses with side shields are adequate for general use, but if heavy grinding, chipping, or overhead work is being done, goggles or a full face shield should be worn in addition to safety glasses, Figure 1.47. Safety glasses are best for general protection. They must be worn under an arc welding helmet at all times.

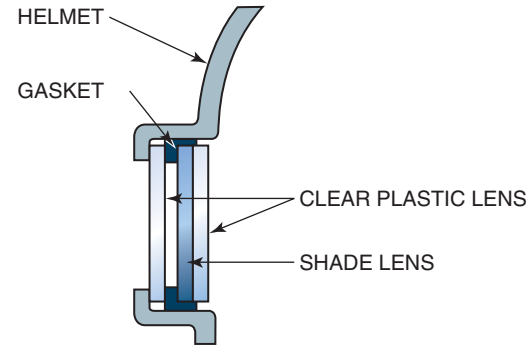

Figure 1.45 Correct placement of the gasket around the shade lens
Correct gasket placement around the shade lens of a welding helmet is important because it can stop ultraviolet light from bouncing around the lens assembly.

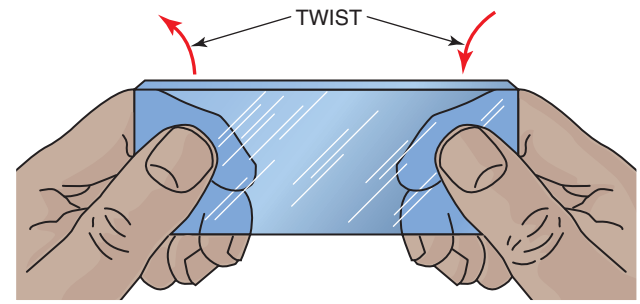

Figure 1.46 To check the shade lens for possible cracks, gently twist it

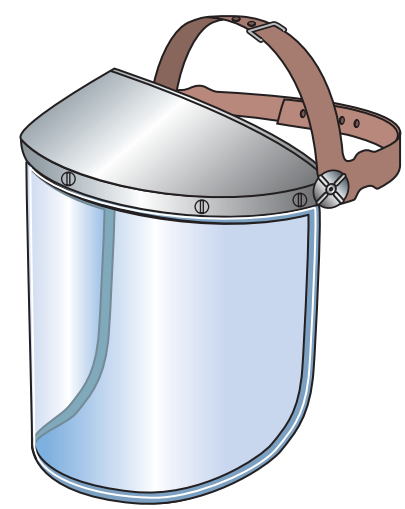

Figure 1.47 Full face shield

Ear Protection

The welding environment can be very noisy. The sound level is at times high enough to cause pain and some loss of hearing if the welder's ears are unprotected. Hot sparks can also drop into an open ear, causing severe burns.

Ear protection is available in several forms. One form of protection is earmuffs that cover the outer ear completely, Figure 1.48. Another form

Figure 1.48 Earmuffs
Earmuffs provide complete ear protection and can be worn under a welding helmet.
Source: Courtesy of Mine Safety Appliances Company

Figure 1.49 Earplugs
Earplugs are used as protection from noise only.
Source: Courtesy of Mine Safety Appliances Company

of protection is earplugs that fit into the ear canal, Figure 1.49. Both of these protect a person's hearing, but only the earmuffs protect the outer ear from burns.

GENERAL WORK CLOTHING

Special protective clothing cannot be worn at all times. It is, therefore, important to choose general work clothing that will minimize the possibility of getting burned because of the high temperature and amount of hot sparks, metal, and slag produced during welding, cutting, or brazing.

Caution

Damage to your hearing caused by high sound levels may not be detected until later in life, and the resulting loss in hearing is permanent. Your hearing will not improve with time, and each exposure to high levels of sound will further damage your hearing.

Work clothing must also stop ultraviolet light from passing through it. This is accomplished if the material chosen is a dark color, thick, and tightly woven. The best choice is 100% wool but it is difficult to find. Another good choice is 100% cotton clothing, the most popular fabric used.

You must avoid wearing synthetic materials including nylon, rayon, and polyester. They can easily melt or catch fire. Some synthetics produce a hot, sticky residue that can make burns more severe. Others may produce poisonous gases.

The following are some guidelines for selecting work clothing:

- Shirts must be long-sleeved to protect the arms, have a high-buttoned collar to protect the neck, Figure 1.50, be long enough to tuck into the pants to protect the waist, and have flaps on the pockets to keep sparks out (or have no pockets).
- Pants must have legs long enough to cover the tops of the boots and must be without cuffs that would catch sparks.
- Boots must have high tops to keep out sparks, have steel toes to prevent crushed toes, Figure 1.51, and have smooth tops to prevent sparks from being trapped in seams.
- Caps should be thick enough to prevent sparks from burning the top of a welder's head.

Figure 1.50 Neck protection
The top button of the shirt worn by the welder should always be buttoned in order to avoid severe neck burns.
Source: Courtesy of Larry Jeffus

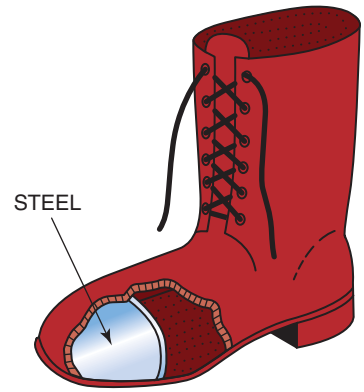

Figure 1.51 Safety boots with steel toes are required by many welding shops

All clothing must be free of frayed edges and holes. The clothing must be relatively tight-fitting in order to prevent excessive folds or wrinkles that might trap sparks.

Some welding clothes have pockets on the inside to prevent the pockets from collecting sparks. However, it is not safe to carry butane lighters or matches in these or any pockets while welding. Lighters and matches can easily catch on fire or explode if they are subjected to the heat and sparks of welding.

Caution

There is no safe place to carry butane lighters or matches while welding or cutting. They can catch fire or explode if subjected to welding heat or sparks. Butane lighters may explode with the force of a quarter of a stick of dynamite. Matches can erupt into a ball of fire. Both butane lighters and matches must always be removed from the welder's pockets and placed a safe distance away before any work is started.

SPECIAL PROTECTIVE CLOTHING

General work clothing is worn by each person in the shop. In addition to this clothing, extra protection is needed for each person who is in direct contact with hot materials. Leather is often the best material to use, as it is lightweight, is flexible, resists burning, and is readily available. Synthetic insulating materials are also available. Ready-to-wear leather protection includes capes, jackets, aprons, sleeves, gloves, caps, pants, knee pads, and spats, among other items.

Hand Protection

All-leather, gauntlet-type gloves should be worn when doing any welding, Figure 1.52. Gauntlet gloves that have a cloth liner for insulation are best for hot work. Noninsulated gloves will give greater flexibility for fine work. Some leather gloves are available with a canvas gauntlet top, which should be used for light work only.

When a great deal of manual dexterity is required for gas tungsten arc welding, brazing, soldering, oxyfuel gas welding, and other delicate processes, soft leather gloves may be used, Figure 1.53. All-cotton gloves are sometimes used when doing very light welding.

Figure 1.52 All-leather, gauntlet-type welding gloves
Source: Courtesy of Larry Jeffus

Figure 1.53 Soft leather gloves
For welding that requires a great deal of manual dexterity, soft leather gloves can be worn.
Source: Courtesy of Larry Jeffus

Body Protection

Full leather jackets and capes will protect a welder's shoulders, arms, and chest, Figure 1.54. A jacket, unlike the cape, protects a welder's back and complete chest. A cape is open and much cooler but offers less protection. The cape can be used with a bib apron to provide some additional protection while leaving the back cooler. Either the full jacket or the cape with a bib apron should be worn for any out-of-position work.

Waist and Lap Protection

Bib aprons or full aprons will protect a welder's lap. Welders will especially need to protect their laps if they squat or sit while working and when they bend over or lean against a table.

Arm Protection

For some vertical welding, a full or half sleeve can protect a person's arm, Figure 1.55. The sleeves work best if the work level is not above the welder's chest. Work levels higher than this usually require a jacket or cape to keep sparks off the welder's shoulders.

Leg and Foot Protection

When heavy cutting or welding is being done and a large number of sparks are falling, leather pants and spats should be used to protect the welder's legs and feet. If the weather is hot and full leather pants are uncomfortable, leather aprons with leggings are available. Leggings can be strapped to the legs, leaving the back open. Spats will prevent sparks from burning through the front of lace-up boots.

Figure 1.54 Full leather jacket
Source: Courtesy of Larry Jeffus

Figure 1.55 Full leather sleeve
Source: Courtesy of Larry Jeffus

SUMMARY

The keys to producing quality GMA welds are equipment, setup, and adjustments. Once you have mastered these skills the only remaining obstacle to your producing consistent, uniform, high-quality welds is your ability to follow, or track, the joint consistently. Some welders find that lightly dragging their glove along the metal surface or edge of the fabrication can aid them in controlling the weld consistency. One of the advantages of the GMA welding process is its ability to produce long, uninterrupted welds. However, this often leads to welder fatigue. Finding a comfortable welding position that you can maintain for several minutes at a time will both improve your weld quality and reduce your fatigue.

Selecting the proper method of metal transfer—short arc, globular, or spray transfer—is normally done by the welding shop foreman or supervisor. He or she makes these selections based on the material being welded, the welding position, and other factors, including welding procedure specifications and applicable codes. A welder must be proficient with each of the various methods of metal transfer. It is therefore important that you spend time practicing and developing your skills with each of these processes.

REVIEW

1. Why is usage of the term *GMAW* preferable to *MIG* for gas metal arc welding?
2. Using Table 1.1, answer the following:
   a. What maintains the arc in machine welding?
   b. What feeds the filler metal in manual welding?
   c. What provides the joint travel in automatic welding?
   d. What provides the joint guidance in semiautomatic welding?
3. What factors have led to increased usage of the GMAW process?
4. In what form is metal transferred across the arc in the axial spray metal transfer method of GMA welding?
5. What three conditions are required for the spray transfer process to occur?
6. Using Table 1.3, answer the following:
   a. What should the wire-feed speed and voltage ranges be to weld 1/8-in. metal with 0.035-in. wire using argon shielding gas?
   b. What should the amperage and voltage ranges be when using 98% Ar + 2% O_2 to weld 1/4-in. metal with 0.045-in. wire?
7. What ranges does the pulsed-arc metal transfer shift between?
8. How do frequency, amplitude, and width of the pulses affect the GMA pulse welding process?
9. How have electronics helped the pulsed-arc process?
10. Why is helium added to argon when making some spray or pulsed-spray transfer welds?
11. Why does DCEP help with welds on metals such as aluminum?
12. Why is CO_2 added to argon when making GMA spray transfer welds?
13. Why should CO_2 not be used to weld stainless steel?
14. How is the metal transferred from the electrode to the plate during the GMAW-S process?
15. Using Figure 1.12, what should be the approximate voltage at 175 amps at 200 in./min when using 0.035-in. ER70S-6 electrode wire?
16. Using Table 1.8, what would the amperage be for 0.035-in. (0.9-mm) wire at 200 in./min (5 m/min)?
17. Using Table 1.9, what shielding gas should be used for welding on copper?
18. What may happen if the GMA welding electrode is allowed to strike the base metal outside the molten weld pool?
19. What effect does shortening the electrode extension have on weld penetration?
20. Describe the weld produced by a backhand welding angle.
21. Describe the weld produced by a forehand welding angle.
22. What components make up a GMA welding system?
23. Why must GMA welders have a 100% duty cycle?
24. What can happen if rollers of the wrong shape are used on aluminum wire?
25. Where is the drive motor located in a pull-type wire-feed system?
26. How is the wire-feed speed changed with a linear feed system?
27. What type of liner should be used for aluminum wire?
28. What parts of a typical GMA welding gun can be replaced?
29. Describe the spot welding process using a GMA welder.

CHAPTER

2 Gas Metal Arc Welding

OBJECTIVES

After completing this chapter, the student should be able to

- set up a GMA welding work station
- thread the electrode wire on a GMAW machine
- set the shielding gas flow rate on a GMAW machine
- use various settings on a GMA welding machine, and compare the effects on a weld
- demonstrate the effect of changing the electrode extension on a weld
- describe the effects of changing the welding gun angle on the weld bead
- list six variables to consider when selecting the shielding gas for a particular application
- evaluate weld beads made with various shielding gas mixtures
- explain why hot-rolled steel should be cleaned to bright metal before welding
- make GMA welds in butt joints, lap joints, and tee joints in all positions that will pass a specified standard's visual or destructive examination criteria

KEY TERMS

bird-nesting
cast
conduit liner
contact tube
feed rollers
flow rate
spool drag
wire-feed speed

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter:

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports or records
Key Indicator 2: Performs housekeeping duties
Key Indicator 3: Follows verbal instructions to complete work assignments
Key Indicator 4: Follows written details to complete work assignments

Module 2: Safety and Health of Welders

Key Indicator 1: Demonstrates proper use and inspection of personal protection equipment (PPE)
Key Indicator 2: Demonstrates proper work area operation

Key Indicator 3: Demonstrates proper use and inspection of ventilation equipment
Key Indicator 4: Demonstrates proper hot zone operation
Key Indicator 7: Demonstrates proper inspection and operation of equipment for each welding or thermal cutting process used (This is best done as part of the process module/unit for each of the required welding and thermal cutting processes.)

Module 3: Drawing and Welding Symbol Interpretation

Key Indicator 1: Interprets basic elements of a drawing or sketch
Key Indicator 2: Interprets welding symbol information
Key Indicator 3: Fabricates parts from a drawing or sketch

Module 5: Gas Metal Arc Welding [GMAW-S, GMAW (spray)]

Key Indicator 1: Performs safety inspections of GMAW equipment and accessories
Key Indicator 2: Makes minor external repairs to GMAW equipment and accessories

Short-circuit Transfer

Key Indicator 3: Sets up for GMAW-S operations on carbon steel
Key Indicator 4: Operates GMAW-S equipment on carbon steel
Key Indicator 5: Makes GMAW-S fillet welds, in all positions, on carbon steel
Key Indicator 6: Makes GMAW-S groove welds, in all positions, on carbon steel
Key Indicator 7: Passes GMAW-S welder performance qualification test (workmanship sample) on carbon steel

Spray Transfer

Key Indicator 8: Sets up for GMAW (spray) operations on carbon steel
Key Indicator 9: Operates GMAW (spray) equipment on carbon steel
Key Indicator 10: Makes GMAW (spray) fillet welds, in the 1F and 2F positions, on carbon steel
Key Indicator 11: Makes GMAW (spray) groove welds, in the 1G position, on carbon steel
Key Indicator 12: Passes GMAW (spray) welder performance qualification test (workmanship sample) on carbon steel

Module 9: Welding Inspection and Testing Principles

Key Indicator 1: Examines cut surfaces and edges of prepared base metal parts.
Key Indicator 2: Examines tacks, root passes, intermediate layers, and completed welds

INTRODUCTION

Performing a satisfactory GMA weld requires more than just manipulative skill. The setup, voltage, amperage, electrode extension, and welding angle, as well as other factors, can dramatically affect the weld produced. The very best welding conditions are those that will allow a welder to produce the largest quantity of successful welds in the shortest period of time with the highest productivity. Because these are semiautomatic or automatic processes, increased productivity may require only that the welder increase the travel

speed and current. This does not mean that the welder will work harder but, rather, that the welder will work more productively, resulting in a greater cost efficiency.

The more cost efficient welders can be, the more competitive they and their companies become. This can make the difference between being awarded a job or losing work.

SETUP

The same equipment may be used for semiautomatic GMAW, flux cored arc welding (FCAW), and submerged arc welding (SAW). Often, FCAW and SAW equipment have a higher amperage range. In addition, equipment for FCAW and SAW is more likely to be automated than that for GMAW. However, GMA welding equipment can easily be automated.

The basic GMAW installation consists of the following: welding gun, gun switch circuit, electrode conduit or liner, welding contractor control, electrode feed unit, electrode supply, power source, shielding gas supply, shielding gas flowmeter regulator, shielding gas hoses, and both power and work cables. Typical water-cooled and air-cooled guns are shown in Figure 2.1. The equipment setup in this chapter is similar to equipment built by other manufacturers, which means that any skills developed can be transferred easily to other equipment.

Figure 2.1 **GMA welding guns are available in a variety of sizes and shapes**
Source: Courtesy of ESAB Welding & Cutting Products

PRACTICE 2-1

GMAW Equipment Setup

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 3
Spray Transfer
Key Indicator 8

For this practice, you will need a GMAW power source, a welding gun, an electrode feed unit, an electrode supply, a shielding gas supply, a shielding gas flowmeter regulator, electrode conduit, power and work leads, shielding gas hoses, assorted hand tools, spare parts, and any other required materials. In this practice, you will properly set up a GMA welding installation.

If the shielding gas supply is a cylinder, it must be chained securely in place before the valve protection cap is removed, Figure 2.2. Standing to one side of the cylinder and making sure no bystanders are in line with the valve, quickly crack the valve to blow out any dirt in the valve before the flowmeter regulator is attached, Figure 2.3. With the flowmeter regu-

Figure 2.2 **Make sure the gas cylinder is chained securely in place before removing the safety cap**
Source: Courtesy of Larry Jeffus

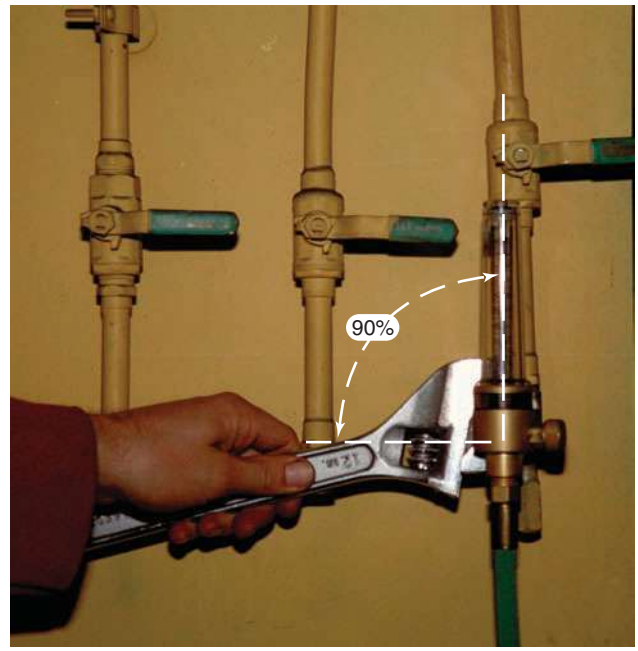

Figure 2.3 **Attaching the flowmeter regulator**
Be sure the tube is vertical.
Source: Courtesy of Larry Jeffus

lator attached securely to the cylinder valve, attach the correct hose from the flowmeter to the "gas-in" connection on the electrode feed unit or machine.

Install the reel of electrode (welding wire) on the holder and secure it, Figure 2.4. Check the feed roller size to ensure that it matches the wire size, Figure 2.5. The conduit liner size should be checked to be sure that it is compatible with the wire size. Connect the conduit to the feed unit. The conduit or an extension should be aligned with the groove in the roller and set as close to the roller as possible without touching, Figure 2.6. Misalignment at this point can contribute to a bird's nest, Figure 2.7. **Bird-nesting** of the electrode wire results when the feed roller pushes the wire into a tangled ball, like a bird's nest, because the wire would not go through the outfeed side conduit.

Figure 2.4 Wire label
When installing the spool of wire, check the label to be sure that the wire is the correct type and size.
Source: Courtesy of Larry Jeffus

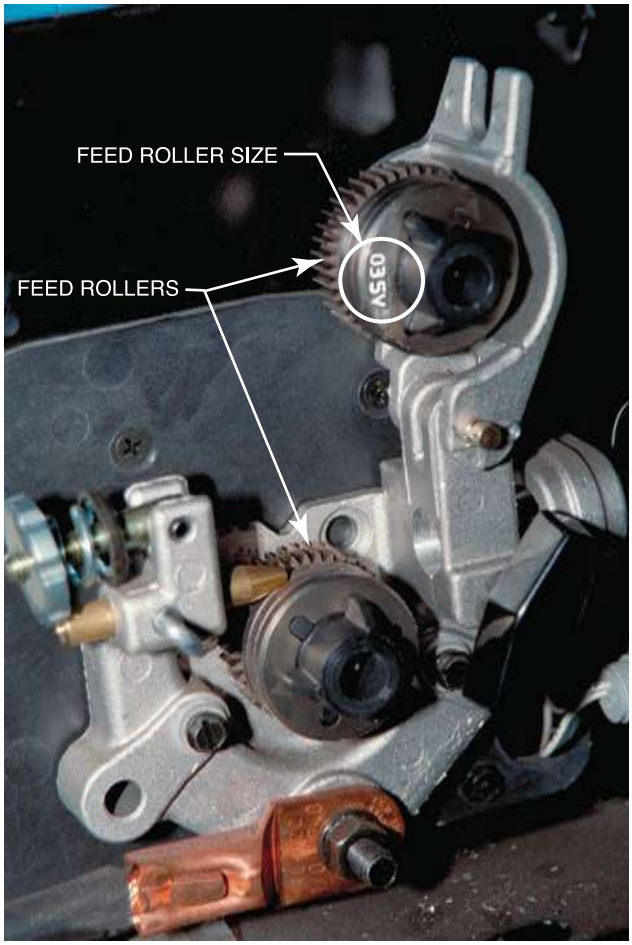

A

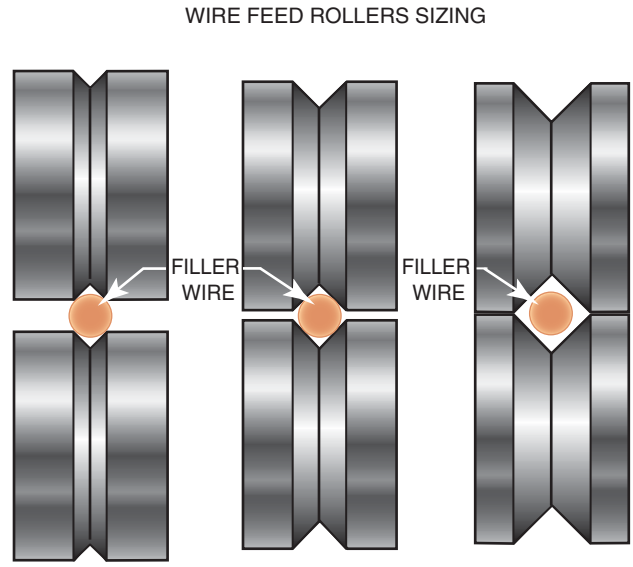

B

Figure 2.5 Checking feed roller size
(A) Check to be certain that the feed rollers are the correct size for the wire being used. (B) If the wirefeed rollers are too small, the welding wire could be damaged. If the wirefeed rollers are too large, the rollers will not grip the wire.
Source: (A) Courtesy of Larry Jeffus

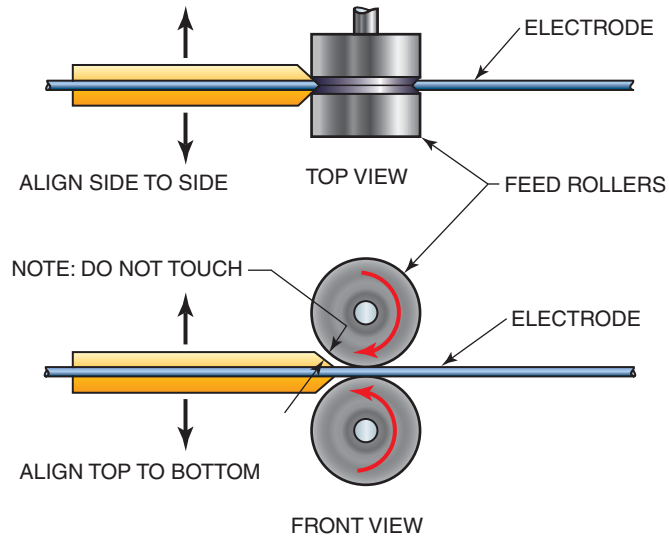

Figure 2.6 Feed

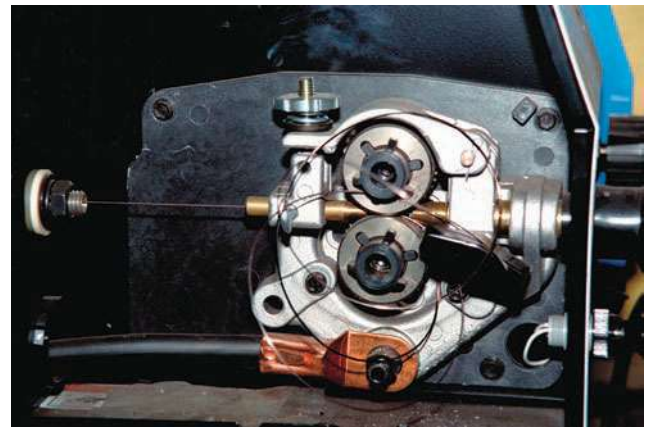

Figure 2.7 "Bird's nest" in the filler wire at the feed rollers
Source: Courtesy of Larry Jeffus

Be sure the power is off before attaching the welding cables. The electrode and work leads should be attached to the proper terminals. The electrode lead should be attached to the terminal marked electrode or positive (+). If necessary, it is also attached to the power cable part of the gun lead. The work lead should be attached to the terminal marked work or negative (–).

The shielding "gas-out" side of the solenoid is then also attached to the gun lead. If a separate splice is required from the gun switch circuit to the feed unit, it should be connected at this time, Figure 2.8. Check to

see that the welding contractor circuit is connected from the feed unit to the power source.

The welding gun should be securely attached to the main lead cable and conduit, Figure 2.9. There should be a gas diffuser attached to the end of the conduit liner to ensure proper alignment. A **contact tube** (tip) of the correct size to match the electrode wire size being used should be installed, Figure 2.10. A shielding gas nozzle is attached to complete the assembly.

Recheck all fittings and connections for tightness. Loose fittings can leak; loose connections can cause added resistance, reducing the welding

(A) (B)

Figure 2.8 **GMAW station setup**
(A) Typical GMA welding machine. (B) Typical interconnecting cables and wires for a semiautomatic GMA welding station.
Source: (A) Courtesy of Lincoln Electric Company (B) Courtesy of Dynatorque

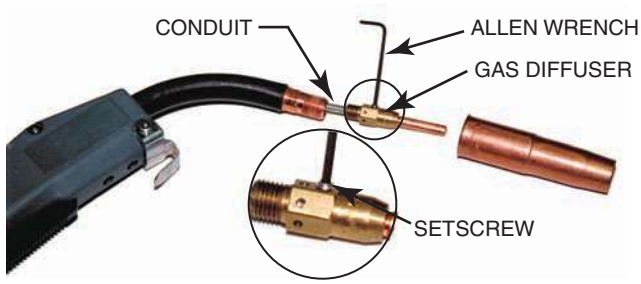

Figure 2.9 **GMA welding gun assembly**
Source: Courtesy of Larry Jeffus

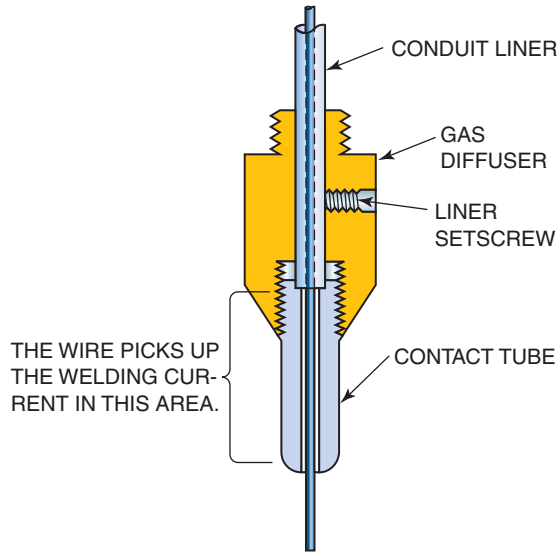

(A)

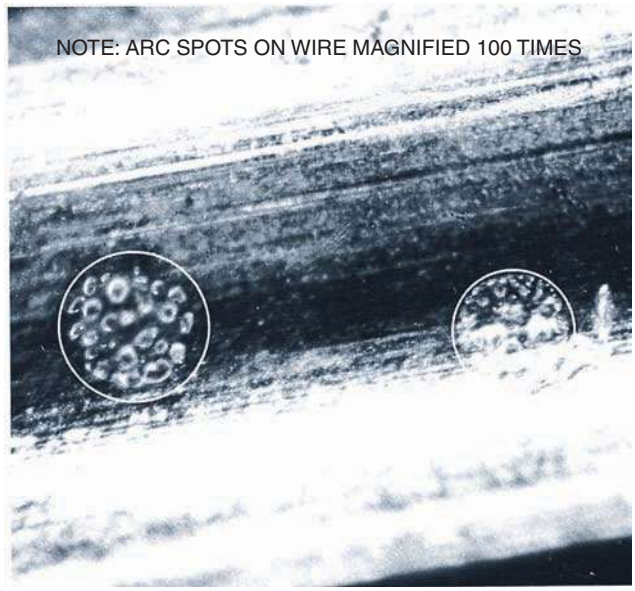

(B)

(C)

Figure 2.10 The contact tube must be the correct size
(A) Too small a contact tube will cause the wire to stick. (B) Too large a contact tube can cause arcing to occur between the wire and tube. (C) Heat from the arcing can damage the tube.
Source: (B) Courtesy of Brett V. Hahn (C) Courtesy of Larry Jeffus

efficiency. Some manufacturers include detailed setup instructions with their equipment, Figure 2.11.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 3
Spray Transfer
Key Indicator 8

PRACTICE 2-2

Threading GMAW Wire

Using the GMAW machine that was properly assembled in Practice 2-1, you will turn the machine on and thread the electrode wire through the system.

Open the side cover.

Remove the empty wire spool.

Release upper feed roller.

Reload the wire spool with the free end unreeling from the bottom.

Thread wire through guide between rollers and into wire cable.

Set the polarity as DCEP from GMA welding.

Turn the input switch on.

With the gun trigger pressed, adjust the feed roller tension.

Check the setting guide inside the machine door.

Set the voltage and wire feed for the metal you are going to be welding.

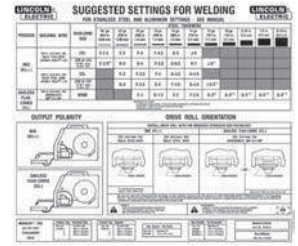

Attach work cable clamp to work to be welded.

Connect gas to coupling at rear of case and turn on shielding gas.

ALWAYS WEAR PROPER SAFETY EQUIPMENT. Pull trigger and weld.

Figure 2.11 Example of manufacturer's setup instructions
Source: Courtesy of Lincoln Electric Company

Check to see that the unit is assembled correctly according to the manufacturer's specifications. Switch on the power and check the gun switch circuit by depressing the switch. The power source relays, feed relays, gas solenoid, and feed motor should all activate.

Cut the end of the electrode wire free. Hold it tightly so that it does not unwind. The wire has a natural curve that is known as its **cast**. The cast is measured by the diameter of the circle that the wire would make if it were loosely laid on a flat surface, Figure 2.12. The cast helps the wire make a good electrical contact as it passes through the contact tube. However, the cast can be a problem when threading the system. To make threading easier, straighten about 12 in. (305 mm) of the end of the wire and cut any kinks off.

Caution

If the wire stops feeding before it reaches the end of the contact tube, stop and check the system. If no obvious problem can be found, mark the wire with tape and remove it from the gun. It then can be held next to the system to determine the location of the problem.

Separate the wire **feed rollers** and push the wire first through the guides, Figure 2.13, then between the rollers, and finally into the **conduit liner**. Reset the rollers so there is a slight amount of compression on the wire, Figure 2.14. Set the **wire-feed speed** control to a slow speed. Hold the welding gun so that the electrode conduit and cable are as straight as possible.

With safety glasses on and the gun pointed away from the welder's face, press the gun switch, or the cold feed switch if your wire feeder is equipped with one. The cold feed switch feeds wire without delivering current to the gun. The wire should start feeding into the liner. Watch to make certain that the wire feeds smoothly and release the gun switch as soon as the end comes through the contact tube.

With the wire feed running, adjust the feed roller compression so that the wire reel can be stopped easily by a slight pressure. Too light a roller pressure will cause the wire to feed erratically. Too high a pressure can turn a minor problem into a major disaster. If the wire jams at a high roller pressure, the feed rollers keep feeding the wire, causing it to bird-nest and possibly short out. With a light pressure, the wire can stop, preventing bird-nesting. This is very important with soft wires. The other

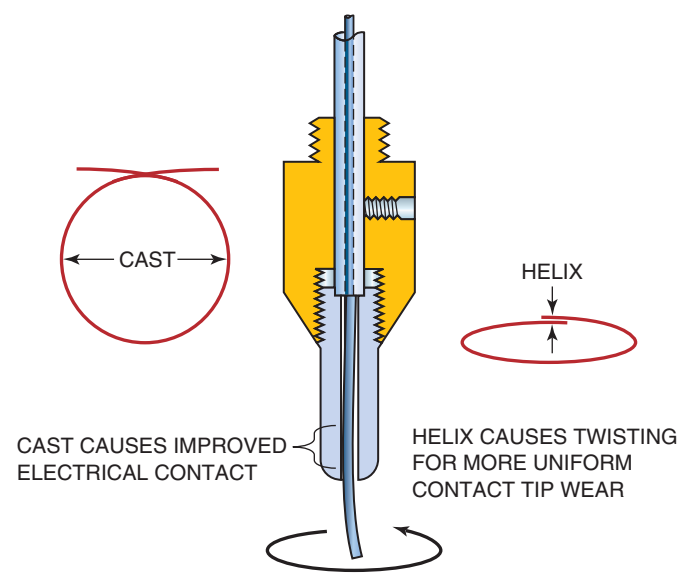

Figure 2.12 Cast of a welding wire
The cast of the welding wire causes it to rub firmly inside the contact tube for good electrical contact. The helix causes the electrode to twist inside the contact tube so that the tube is worn uniformly.

Figure 2.13 **Push the wire through the guides by hand**
Source: Courtesy of Larry Jeffus

Figure 2.14 **Adjust the wire-feed tensioner**
Source: Courtesy of Larry Jeffus

advantage of a light pressure is that the feed will stop if something like clothing or a gas hose is caught in the reel.

With the feed running, adjust the **spool drag** so that the reel stops when the feed stops. The reel should not coast to a stop because the wire can be snagged easily. Also, when the feed restarts, a jolt occurs when the slack in the wire is taken up. This jolt can be enough to momentarily stop the wire, possibly causing a discontinuity in the weld.

When the test runs are completed, the wire can either be rewound or cut off. Some wire-feed units have a retract button. This allows the feed driver to reverse and retract the wire automatically. To rewind the wire on units without this retraction feature, release the rollers and turn them backward by hand. If the machine will not allow the feed rollers to be released without upsetting the tension, you must cut the wire.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Caution

Do not discard pieces of wire on the floor. They present a hazard to safe movement around the machine. In addition, a small piece of wire can work its way into a filter screen on the welding power source. If the piece of wire shorts out inside the machine, it could become charged with high voltage, which could cause injury or death. Always wind the wire tightly into a ball or cut it into short lengths before discarding it in the proper waste container.

GAS DENSITY AND FLOW RATES

Density is the chief determinant of how effective a gas is for arc shielding. The lower the density of a gas, the higher will be the flow rate required for equal arc protection. Flow rates, however, are not in proportion to the

NOTE

If you need more shielding gas coverage in a windy or drafty area, use both a larger-diameter gas nozzle and a higher gas flow rate. The larger the nozzle size, the higher the permissible flow rate without causing turbulence. Larger nozzle sizes may restrict your view of the weld. You might also consider setting up a wind barrier to protect your welding from the wind, Figure 2.16.

densities. Helium, with about one-tenth the density of argon, requires about twice the flow for equal protection.

The correct flow rate can be set by checking welding guides that are available from the welding equipment and filler metal manufacturers. These welding guides list the gas flow required for various nozzle sizes and welding amperage settings. Some welders feel that a higher gas flow will provide better weld coverage, but that is not always the case. High gas flow rates waste shielding gases and may lead to contamination. The contamination comes from turbulence in the gas at high flow rates. Air is drawn into the gas envelope by the venturi effect around the edge of the nozzle. Also, the air can be drawn in under the nozzle if the torch is held at too sharp an angle to the metal, Figure 2.15.

EXPERIMENT 2-1

Setting Gas Flow Rate

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 3
Spray Transfer
Key Indicator 8

Using the equipment setup as described in Practice 2-1, and the threaded machine as described in Practice 2-2, you will set the shielding gas flow rate.

The exact **flow rate** required for a certain job will vary depending upon welding conditions. This experiment will help you determine how those conditions affect the flow rate. You will start by setting the shielding gas flow rate at 35 cubic feet per hour (cfh) (16 L/min).

Standing to one side, turn on the shielding gas supply valve. If the supply is a cylinder, the valve is opened all the way. With the machine power on and the welding gun switch depressed, you are ready to set

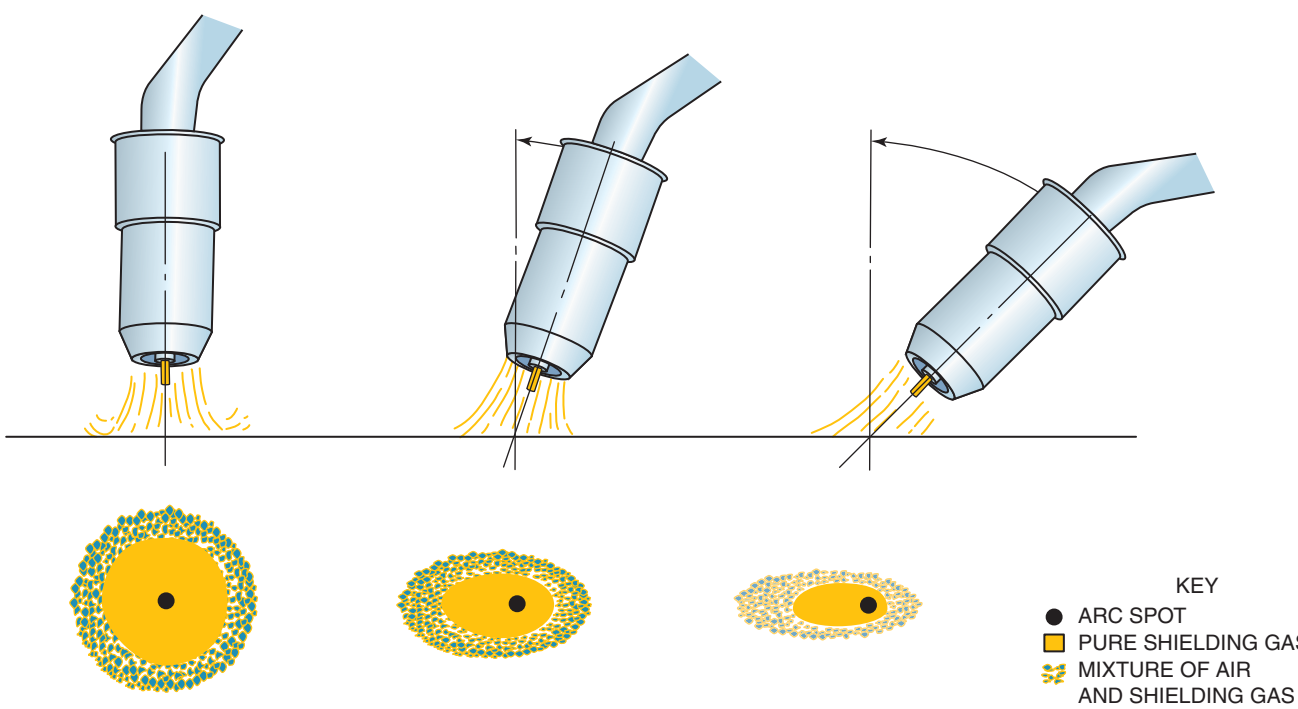

Figure 2.15 The welding gun angle affects the shielding gas coverage for the molten weld pool.

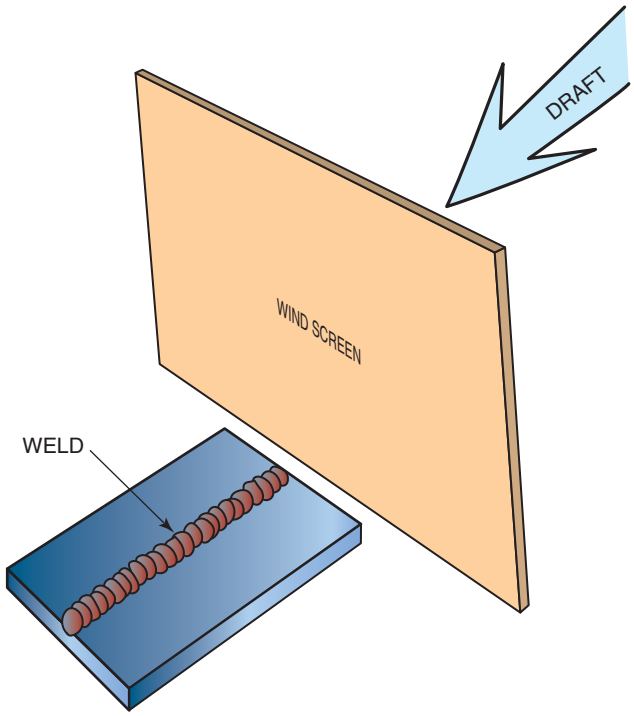

Figure 2.16 **A wind screen can help prevent the shielding gas from being blown away.**

the flow rate. Slowly turn in the adjusting screw and watch the float ball as it rises in a tube on a column of gas. The faster the gas flows, the higher the ball will float. A scale on the tube allows you to read the flow rate. Different scales are used with each type of gas being used. Since various gases have different densities (weights), the ball will float at varying levels even though the flow rates are the same, Figure 2.17. The line corresponding to the flow rate may be read as it compares to the top, center, or bottom of the ball, depending upon the manufacturer's instructions. There should be some marking or instruction on the tube or regulator to tell how it should be read, Figure 2.18.

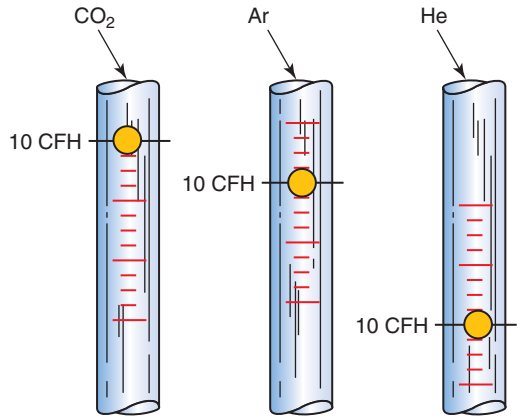

Figure 2.17 **Reading gas flow rates**
Each of these gases is flowing at the same cfh (L/min) rate. Because helium (He) is less dense, its indicator ball is the lowest. Be sure that you are reading the correct scale for the gas being used.

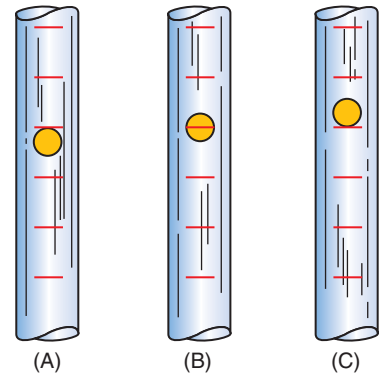

Figure 2.18 **Three methods of reading a flowmeter**
(A) Top of ball, (B) center of ball, and (C) bottom of ball.

Release the welding gun switch, and the gas flow should stop. Turn off the power and spray the hose fittings with a leak-detecting solution.

When you stop for more than a short period, close the shielding gas supply valve and release the hose pressure.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

ARC-VOLTAGE AND AMPERAGE CHARACTERISTICS

The arc-voltage and amperage characteristics of GMA welding are different from those for most other welding processes. The voltage is set on the welder, and the amperage is set by changing the wire-feed speed. At any one voltage setting the amperage required to melt the wire must change as it is fed into the weld. More amperage is required to melt the wire the faster it is fed, and less the slower it is fed.

Because changes in the wire-feed speed directly change the amperage, it is possible to set the amperage by using a chart and measuring the length of wire fed per minute, Table 2.1. The voltage and amperage required for a specific metal transfer method differ for various wire sizes, shielding gases, and metals.

The voltage and amperage setting will be specified for all welding done according to a welding procedure specification (WPS) or other codes and standards. However, most welding—like that done in small production shops, as maintenance welding, for repair work, in farm shops, and the like—is not done to a specific code or standard and therefore no specific setting exists. For that reason, it is important to learn to make the adjustments necessary to allow you to produce quality welds.

EXPERIMENT 2-2

Setting the Current

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 4
Spray Transfer
Key Indicator 9

Using a properly assembled GMA welding machine, proper safety protection, and one piece of mild steel plate approximately 12 in. (305 mm) long × 1/4 in. (6 mm) thick, you will change the current settings and observe the effect on GMAW. On a scale of 0 to 10, set the wire-feed speed control dial at 5, or halfway between the low and high settings of

Table 2.1 Typical Amperages for Carbon Steel

Wire-feed Speed* (in./min)	Wire Diameter .030 in. (0.8 mm)	.035 in. (0.9 mm)	.045 in. (1.2 mm)	.062 in. (1.6 mm)
100 (2.5)	40	65	120	190
200 (5.0)	80	120	200	330
300 (7.6)	130	170	260	425
400 (10.2)	160	210	320	490
500 (12.7)	180	245	365	–
600 (15.2)	200	265	400	–
700 (17.8)	215	280	430	–

*To check feed speed, run out wire for one minute and then measure its length.

the unit. The voltage is also set at a point halfway between the low and high settings. The shielding gas can be CO_2, argon, or a mixture. The gas flow should be adjusted to a rate of 35 cfh (16 L/min).

Hold the welding gun at a comfortable angle, lower your welding hood, and pull the trigger. As the wire feeds and contacts the plate, the weld will begin. Move the gun slowly along the plate. Note the following welding conditions as the weld progresses: voltage, amperage, weld direction, metal transfer, spatter, molten weld pool size, and penetration. Stop and record your observations in Table 2.2. Evaluate the quality of the weld as acceptable or unacceptable.

Reduce the voltage somewhat and make another weld, keeping all other weld variables (travel speed, stickout, direction, amperage) the same. Observe the weld and upon stopping record the results. Repeat this procedure until the voltage has been lowered to the minimum value indicated on the machine. Near the lower end the wire may stick, jump, or simply no longer weld.

Return the voltage indicator to the original starting position and make a short test weld. Stop and compare the results to those first observed. Then slightly increase the voltage setting and make another weld. Repeat the procedure of observing and recording the results as the voltage is increased in steps until the maximum machine capability is obtained. Near the maximum setting the spatter may become excessive if CO_2 shielding gas is used. Care must be taken to prevent the wire from fusing to the contact tube.

Return the voltage indicator to the original starting position and make a short test weld. Compare the results observed with those previously obtained.

Lower the wire-feed speed setting slightly and use the same procedure as before. First lower and then raise the voltage through a complete range and record your observations. After a complete set of test results are obtained from this amperage setting, again lower the wire-feed speed for a new series of tests. Repeat this procedure until the amperage is at the minimum setting shown on the machine. At low amperages and high voltage settings, the wire may tend to pop violently as a result of the uncontrolled arc.

Return the wire-feed speed and voltages to the original settings. Make a test weld and compare the results with the original tests. Slightly raise the wire speed and again run a set of tests as the voltage is changed in small steps. After each series, return the voltage setting to the starting point and increase the wire-feed speed. Make a new set of tests.

All of the test data can be gathered into an operational graph for the machine, wire type, size, and shielding gas. Set up a graph like that in Figure 2.19 to plot the data. The acceptable welds should be marked on

Table 2.2 Setting the Current

Weld Acceptability	Voltage	Amperage	Spatter	Molten Pool Size	Penetration
Good	20	75	Light	Small	Little

Electrode diameter 0.035 in. (0.9 mm)
Shielding gas CO_2
Welding direction Backhand

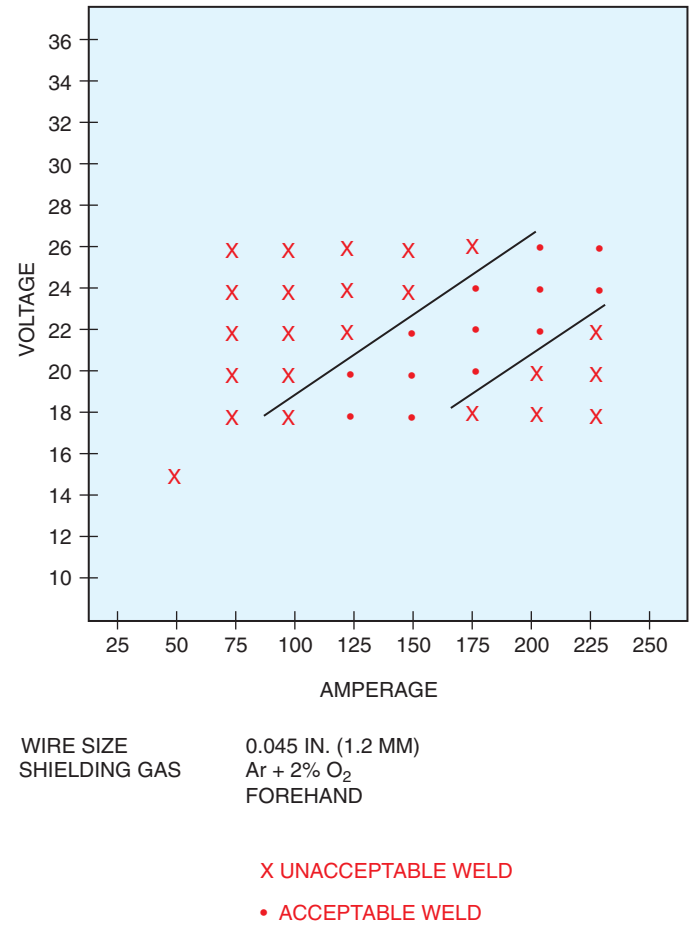

Figure 2.19 Graph for GMAW machine settings

the lines that extend from the appropriate voltages and amperages. Upon completion, the graph will give you the optimum settings for the operation of this particular GMAW setup. The optimum settings are along a line in the center of the acceptable welds.

Experienced welders will follow a much shorter version of this type of procedure anytime they are starting to work on a new machine or testing for a new job. This experiment can be repeated using different types of wire, wire sizes, shielding gases, and weld directions. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

ELECTRODE EXTENSION

Because of the constant-potential (CP) power supply, the welding current will change as the distance between the contact tube and the work changes. Although this change is slight, it is enough to affect the weld being produced. The longer the electrode extension, the greater the

resistance will be to the welding current flowing through the small welding wire. This results in some of the welding current being changed to heat at the tip of the electrode, Figure 2.20. With a standard constant-current (CC) power supply for SMA welding, the heat buildup would also reduce the arc voltage, but with a CP power supply the voltage remains constant and the amperage increases. If the electrode extension is shortened, the welding current decreases.

The increase in current does not result in an increase in penetration, because the current is being used to heat the electrode tip and not being transferred to the weld metal. Penetration is reduced and buildup is increased as the electrode extension is lengthened. Penetration is increased and buildup decreased as the electrode extension is shortened. Controlling the weld penetration and buildup by changing the electrode will help maintain weld bead shape during welding. It will also help you better understand what may be happening if a weld starts out correctly but begins to change as it progresses along the joint. You may be changing the electrode extension without noticing the change. Short electrode stickout gives a hotter weld, and long stickout results in a cooler weld.

EXPERIMENT 2-3

Electrode Extension

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 4
Spray Transfer
Key Indicator 9

Using a properly assembled GMA welding machine, proper safety protection, and a few pieces of mild steel, each about 12 in. (305 mm) long and ranging in thickness from 16 gauge to 1/2 in. (13 mm), you will observe the effect of changing electrode extension on the weld.

Start at a low current setting. Using the graph developed in Experiment 2.2, set both the voltage and amperage. The settings should be equal to those on the optimum line established for the wire type and size being used with the same shielding gas.

Holding the welding gun at a comfortable angle and height, lower your helmet and start to weld. Make a weld approximately 2 in. (51 mm) long. Then reduce the distance from the gun to the work while continuing to weld. After a few inches, again shorten the electrode extension even

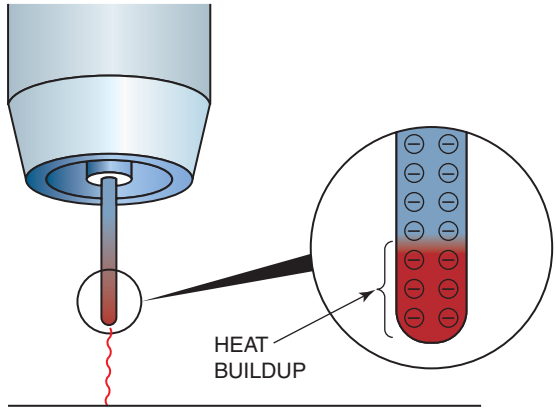

Figure 2.20 Heat buildup due to the extremely high current for the small conductor (electrode)

more. Keep doing this in steps until the nozzle is as close as possible to the work. Stop and return the gun to the original starting distance.

Repeat the process just described but now increase the electrode extension and make welds of a few inches at each extension. Keep increasing the electrode extension until the weld will no longer fuse or the wire becomes impossible to control.

Change the plate thickness and repeat the procedure. When the series has been completed with each plate thickness, raise the voltage and amperage to a medium setting and repeat the process. Upon completing this series of tests, adjust the voltage and amperage upward to a high setting. Make a full series of tests using the same procedures as before.

Record the results in Table 2.3 after each series of tests. The final results can be plotted on a graph, as was done in Figure 2.21, to establish the optimum electrode extension for each thickness, voltage, and amperage. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Table 2.3 Electrode Extension

Weld Acceptability	Voltage	Amperage	Electrode Extension	Contact Tube to Work Distance	Bead Shape
Poor	20	100	1 in. (25 mm)	1-1/4 in. (31 mm)	Narrow, high, with little penetration

Electrode diameter: 0.035 in. (0.9 mm)
Shielding gas: CO_2
Welding direction: Forehand

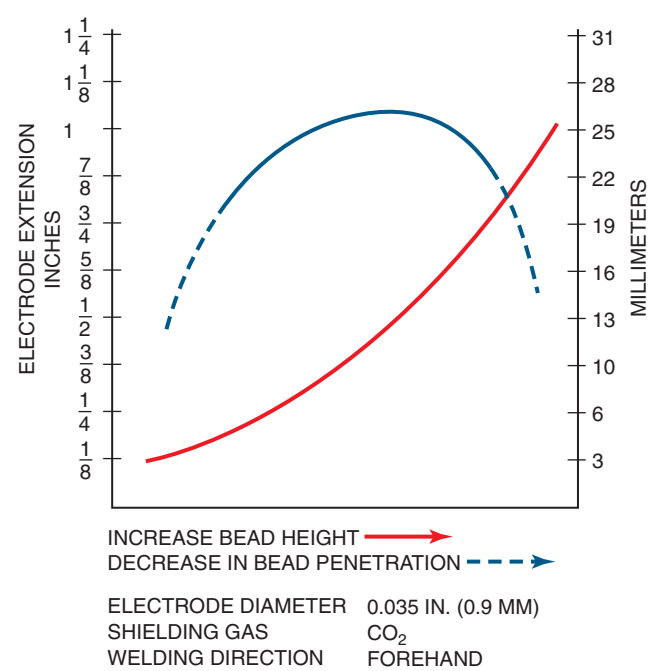

Figure 2.21 Plot of Experiment 2-3 results

WELDING GUN ANGLE

The term *welding gun angle* refers to the angle between the GMA welding gun and the work as it relates to the direction of travel. Backhand welding, or dragging angle, Figure 2.22, produces a weld with deep penetration and higher buildup. Forehand welding, or pushing angle, Figure 2.23, produces a weld with shallow penetration and little buildup.

Slight changes in the welding gun angle can be used to control the weld as the groove spacing changes. A narrow gap may require more penetration, but as the gap spacing increases a weld with less penetration may be required. Changing the electrode extension and welding gun angle at the same time can result in a quality weld being made under less than ideal conditions.

EXPERIMENT 2-4

Welding Gun Angle

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 4
Spray Transfer
Key Indicator 9

Using a properly assembled GMA welding machine, proper safety protection, and some pieces of mild steel, each approximately 12 in. (305 mm) long and ranging in thickness from 16 gauge to 1/2 in. (13 mm), you will observe the effect of changing the welding gun angle on the weld bead.

Starting with a medium current setting and a plate that is 1/4 in. (6 mm) thick, hold the welding gun at a 30° angle to the plate in the direction of the weld, Figure 2.24. Lower your welding hood and depress the trigger. When the weld starts, move in a straight line and slowly pivot the gun angle as the weld progresses. Keep the travel speed, electrode

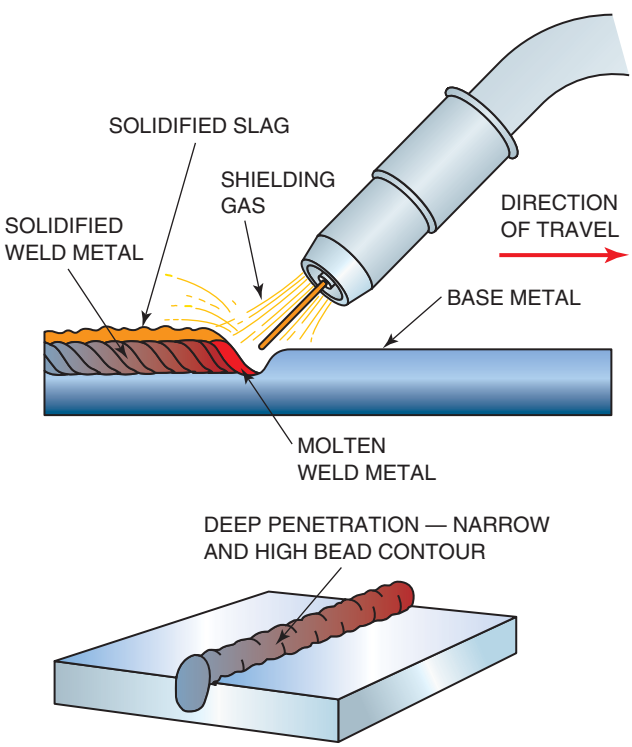

Figure 2.22 Backhand welding, or dragging angle

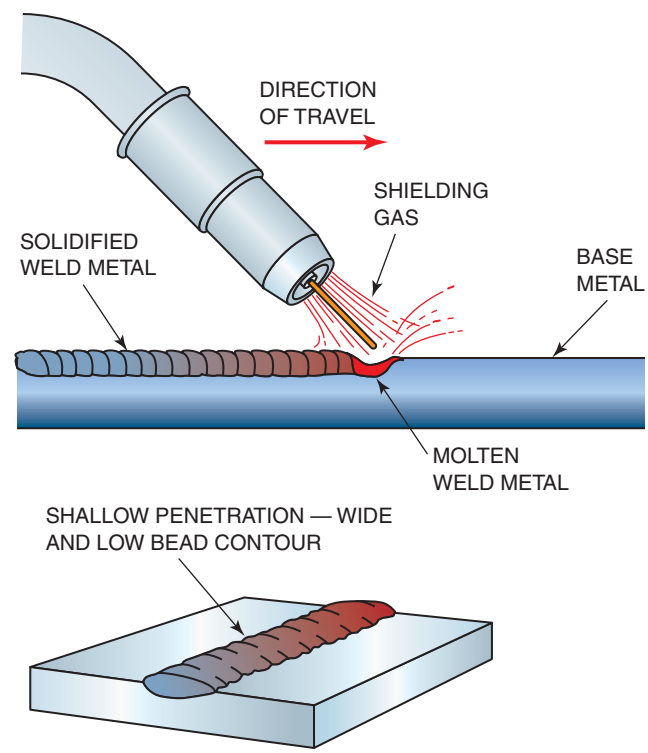

Figure 2.23 Forehand welding, or pushing angle

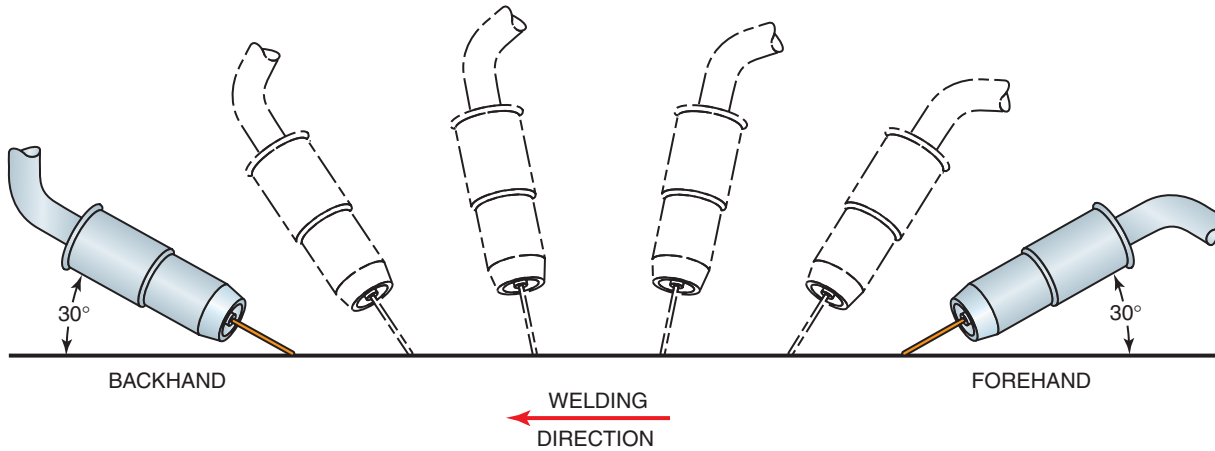

Figure 2.24 Welding gun angle

extension, and weave pattern (the side-to-side motion, if used) constant so that any change in the weld bead is caused by the angle change.

The pivot should be completed in the 12 in. (305 mm) of the weld. You will proceed from a 30° pushing angle to a 30° dragging angle. Repeat this procedure using different welding currents and plate thicknesses.

After the welds are complete, note the differences in width and reinforcement along the welds. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

EFFECT OF SHIELDING GAS ON WELDING

Shielding gases in the gas metal arc process are used primarily to protect the molten metal from oxidation and contamination. Other factors must be considered, however, in selecting the right gas for a particular application. Shielding gas can influence arc and metal transfer characteristics, weld penetration, width of fusion zone, surface shape patterns, welding speed, and undercut tendency. Inert gases such as argon and helium provide the necessary shielding because they do not form compounds with any other substance and are insoluble in molten metal. When used as pure gases for welding ferrous metals, argon and helium may produce an erratic arc action, promote undercutting, and result in other flaws.

It is therefore usually necessary to add controlled quantities of reactive gases to achieve good arc action and metal transfer with these materials. Adding oxygen or carbon dioxide to the inert gas tends to stabilize the arc, promote favorable metal transfer, and minimize spatter. As a result, the penetration pattern is improved and undercutting is reduced or eliminated.

Oxygen or carbon dioxide is often added to argon. The amount of reactive gas required to produce the desired effects is quite small. As little as 0.5% of oxygen will produce a noticeable change; 1% to 5% of oxygen is more common. Carbon dioxide may be added to argon in the 10% to 30% range. Mixtures of argon with less than 10% carbon dioxide may not have enough arc voltage to give the desired results.

Adding oxygen or carbon dioxide to an inert gas causes the shielding gas to become oxidizing. This in turn may cause porosity in some ferrous metals. In this case, a filler wire containing suitable deoxidizers should be used. The presence of oxygen in the shielding gas can also cause some loss of certain alloying elements, such as chromium, vanadium, aluminum, titanium, manganese, and silicon. Again, the addition of a deoxidizer to the filler wire is necessary.

Pure carbon dioxide has become widely used as a shielding gas for GMA welding of steels. It allows higher welding speed, better penetration in the short-circuiting transfer mode, and good mechanical properties, and it costs less than the inert gases. The chief drawback in the use of carbon dioxide is the less-steady-arc characteristics and considerable weld-metal-spatter losses. The spatter can be kept to a minimum by maintaining a very short, uniform arc length. Consistently sound welds can be produced using carbon dioxide shielding, provided that a filler wire having the proper deoxidizing additives is used.

EXPERIMENT 2-5

Effect of Shielding Gas Changes

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 4
Spray Transfer
Key Indicator 9

Using a properly assembled GMA welding machine; proper safety protection; a source of CO_2, argon, and oxygen gases or a variety of premixed shielding gases; two flowmeters (or one two-gas mixing regulator); and some pieces of mild steel plate, each about 12 in. (305 mm) long and ranging in thickness from 16 gauge to 1/2 in. (13 mm), you will observe the effect of various shielding gas mixtures on the weld.

Using a mixing flowmeter regulator will allow the gases to be mixed in any desired mixture. A mixing ratio chart to be used to arrive at the approximate gas percentages appears in Figure 2.25. The exact ratios are not so important to you, as a student, as they are on code work.

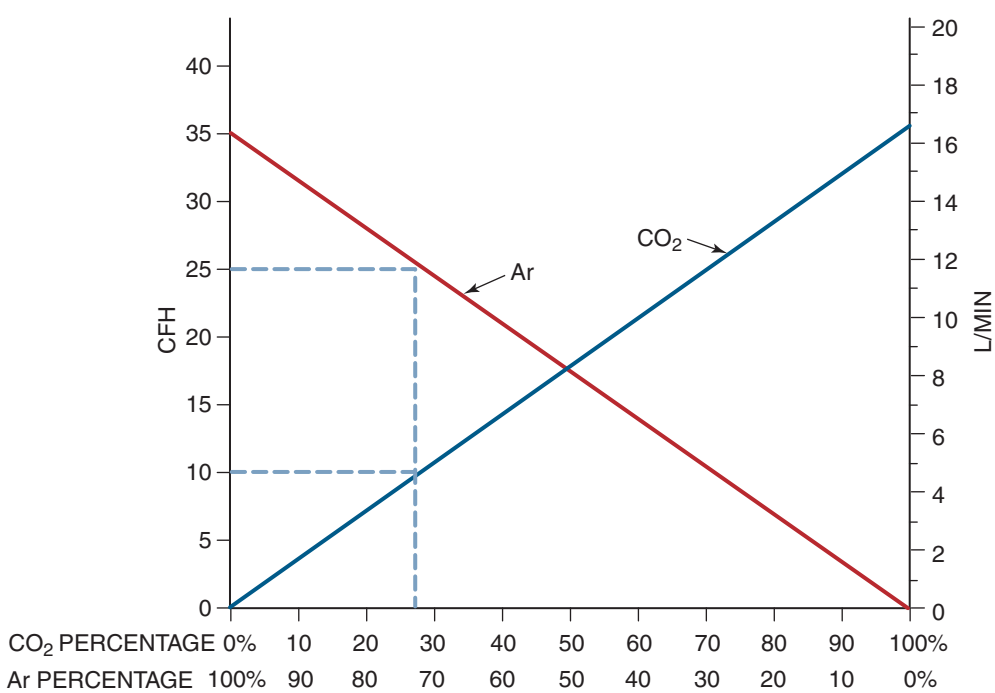

EXAMPLE:

10 CFH	CO_2	= 28%
25 CFH	Ar	= 72%
35 CFH	TOTAL = 100%	
4.75 L/MIN	CO_2	= 28%
11.75 L/MIN	Ar	= 72%
16.5 L/MIN	TOTAL = 100%	

Figure 2.25 Gas mixing percentages

With a medium voltage and amperage setting and using a 100% carbon dioxide (CO_2) shielding gas, start making a weld. Either change the mixture after each weld or have another person change the shielding gas during the weld. Keep the total flow rate the same by adding argon (Ar) while reducing the CO_2 to preserve the same flow rate. During the experiment, change over the shielding gas to 100% argon (Ar).

After the weld is complete, evaluate it for spatter, penetration, undercut, buildup, width, or other noticeable changes along its length. Using Table 2.4, record the results of your evaluation.

Repeat the procedure just explained two times more with both low and high power settings. Again, record your observations.

Starting with 100% argon (Ar), add oxygen (O_2) to the shielding gas. The oxygen percentage will range from 0% to 10%, Figure 2.26. Very slight changes in the percentage will have dramatic effects on the weld. You will make three welds using low, medium, and high power settings. For each weld, you will record your observations.

Table 2.4 Shielding Gas Mixtures

Weld Acceptability	Voltage	Spatter	Penetration	Puddle Size	Bead Appearance
Good	75 Ar 25 CO_2	Very little	Deep	Large	Wide with little buildup

Electrode diameter: 0.035 in. (0.9 mm)
Welding direction: Backhand
Voltage: 25
Amperage: 150

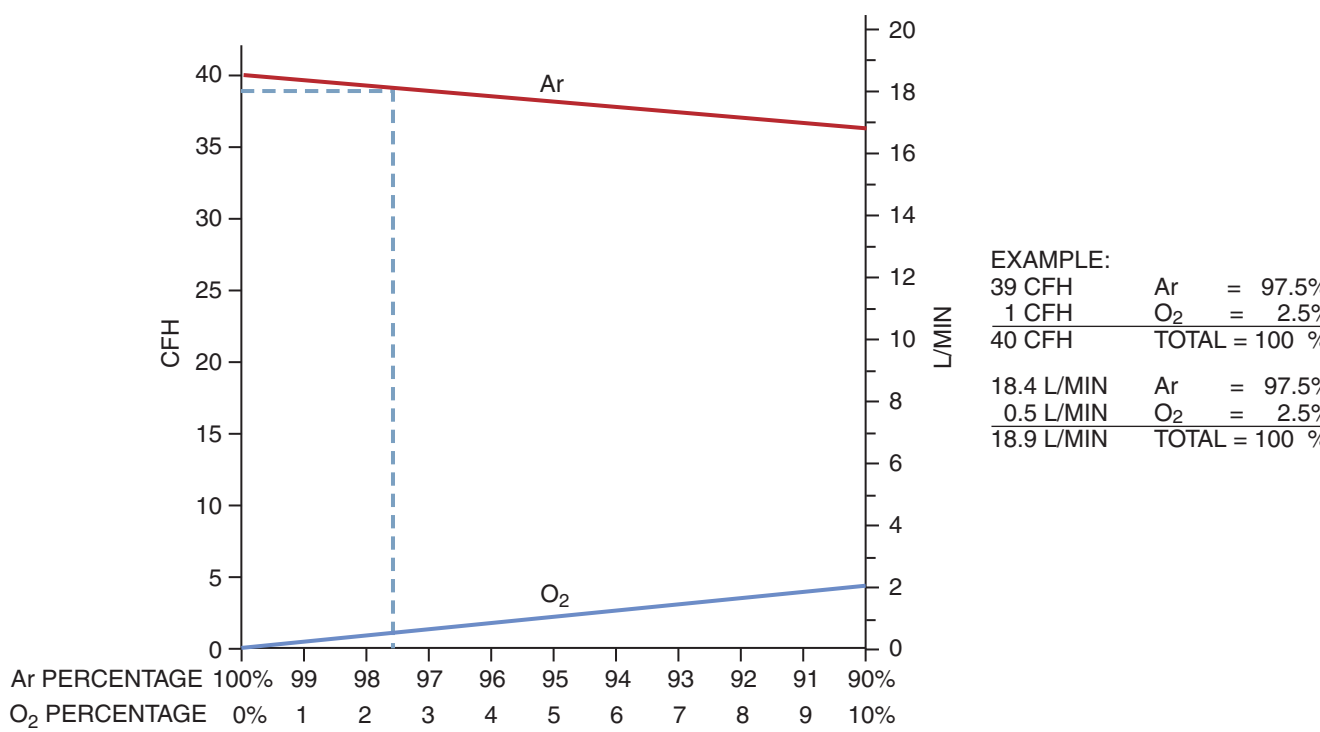

Figure 2.26 Ar and O_2 Mixture Percentages

During some of the welding tests, you will notice a change in the method of metal transfer, weld heat, and general weld performance without a change in the current settings. The shielding gas mixture can have major effects on the rate of metal transfer and the welding speed, as well as other welding variables. Higher speeds and greater production can be obtained by using some gas mixtures. However, the savings can be completely offset by the higher gas cost. Before making a final decision about the gas to be used, all the variables must be compared. Table 2.5 lists premixed shielding gases and their uses.

Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICES

The practices in this chapter are grouped according to those requiring similar techniques and setups. To make acceptable GMA welds consistently, the major skill required is the ability to set up the equipment and weldment. Changes such as variations in material thickness, position, and type of joint require changes in both technique and setup. A correctly set up GMA welding station can, in many cases, be operated with minimum skill. Often the only difference between a welder earning a minimum wage and one earning the maximum wage is the ability to make correct machine setups.

Ideally, only a few tests would be needed for the welder to make the necessary adjustments in setup and manipulation techniques to achieve

Table 2.5 Shielding Gases and Gas Mixtures Used for Gas Metal Arc Welding

Shielding Gas	Chemical Behavior	Uses and Usage Notes
1. Argon	Inert	Welding virtually all metals except steel
2. Helium	Inert	Al and Cu alloys for greater heat and to minimize porosity
3. Ar and He (20% to 80% to 50% to 50%)	Inert	Al and Cu alloys for greater heat and to minimize porosity but with quieter, more readily controlled arc action
4. N_2	Reducing	On Cu, very powerful arc
5. Ar + 25% to 30% N_2	Reducing	On Cu, powerful but smoother operating, more readily controlled arc than with N_2
6. Ar + 1% to 2% O_2	Oxidizing	Stainless and alloy steels, also for some deoxidized copper alloys
7. Ar + 3% to 5% O_2	Oxidizing	Plain carbon, alloy, and stainless steels (generally requires highly deoxidized wire)
8. Ar + 3% to 5% O_2	Oxidizing	Various steels using deoxidized wire
9. Ar + 20% to 30% O_2	Oxidizing	Various steels, chiefly with short-circuiting arc
10. Ar + 5% O_2 + 15% CO_2	Oxidizing	Various steels using deoxidized wire
11. CO_2	Oxidizing	Plain-carbon and low-alloy steels, deoxidized wire essential
12. CO_2 + 3% to 10% O_2	Oxidizing	Various steels using deoxidized wire
13. CO_2 + 20% O_2	Oxidizing	Steels

a good weld. The previous welding experiments should have given the welder a graphic set of comparisons to help that welder make the correct changes. In addition to keeping the test data, you may want to keep the test plates for a more accurate comparison.

The grouping of practices in this chapter will keep the number of variables in the setup to a minimum. Often, the only change required before going on to the next weld is to adjust the power settings.

Figures that are given in some of the practices will give the welder general operating conditions, such as voltage, amperage, and shielding gas or gas mixture. These are general values, so the welder will have to make some fine adjustments. Differences in the type of machine being used and the material surface condition will affect the settings. For this reason, it is preferable to use the settings developed during the experiments.

METAL PREPARATION

All hot-rolled steel has an oxide layer, which is formed during the rolling process, called mill scale. *Mill scale* is a thin layer of dark gray or black iron oxide. Some hot-rolled steels that have had this layer removed either mechanically or chemically can be purchased. However, almost all of the hot-rolled steel used today still has this layer because it offers some protection from rusting.

Mill scale is not removed for noncode welding, because it does not prevent most welds from being suitable for service. For practice welds that will be visually inspected, mill scale can usually be left on the plate. Filler metals and fluxes usually have deoxidizers added to them so that the adverse effects of the mill scale are reduced or eliminated, Table 2.6. But with GMA welding wire it is difficult to add enough deoxidizers to remove all effects of mill scale. The porosity that mill scale causes is most often confined to the interior of the weld and is not visible on the surface, Figure 2.27. Because it is not visible on the surface, it usually goes unnoticed and the weld passes visual inspection.

If the practice results are going to be destructively tested or if the work is of a critical nature, then all welding surfaces within the weld groove and the surrounding surfaces within 1 in. (25 mm) must be cleaned to bright metal, Figure 2.28. Cleaning may be either grinding, filing, sanding, or blasting.

Table 2.6 Deoxidizing Elements in Filler Wire

Deoxidizing Element	Strength
Aluminum (Al)	Very strong
Manganese (Mn)	Weak
Silicon (Si)	Weak
Titanium (Ti)	Very strong
Zirconium (Zr)	Very strong

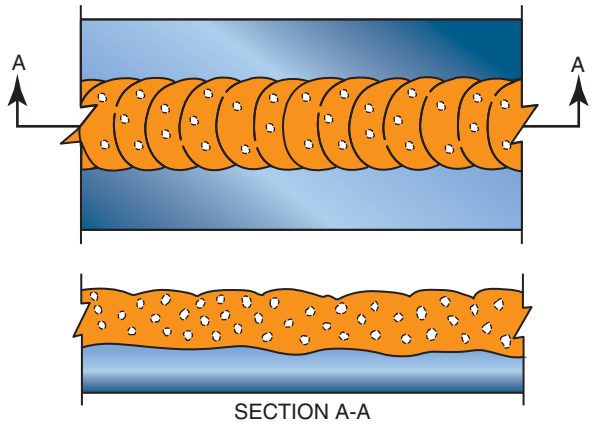

Figure 2.27 Uniformly scattered porosities

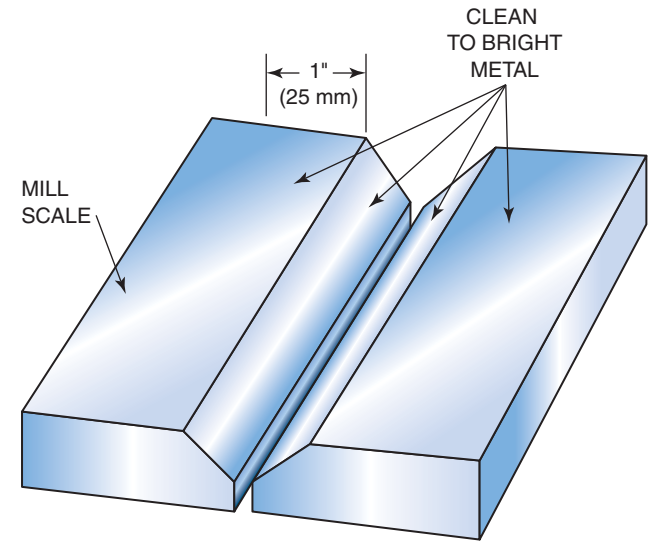

Figure 2.28 Clean all surfaces to bright metal before welding

FLAT POSITION, 1G AND 1F POSITIONS

PRACTICE 2-3

Stringer Beads Using the Short-circuiting Metal Transfer Method in the Flat Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 4

Using a properly set up and adjusted GMA welding machine, Table 2.7, proper safety protection, 0.035-in. and/or 0.045-in. (0.9-mm and/or 1.2-mm) diameter wire, and two or more pieces of mild steel sheet 12 in. (305 mm) long and 16 gauge and 1/8 in. (3 mm) thick, you will make a stringer bead weld in the flat position, Figure 2.29.

Starting at one end of the plate and using either a pushing or dragging technique, make a weld bead along the entire 12-in. (305-mm) length of the metal. After the weld is complete, check its appearance. Make any needed changes in voltage, wire feed speed, or electrode extension to correct the weld (refer to Figure 2.19 and Figure 2.21). Repeat the weld and make additional adjustments. After the machine is set, start to work on improving the straightness and uniformity of the weld.

Keeping the bead straight and uniform can be hard because of the limited visibility due to the small amount of light and the size of the molten weld pool. The welder's view is further restricted by the shielding gas nozzle, Figure 2.30. Even with limited visibility, it is possible to make a

Table 2.7 Typical Welding Current Settings for Short-circuiting Metal Transfer for Mild Steel

Process	Wire Diameter	Amperage Range (Optimum)	Voltage Range (Optimum)	Shielding Gas
Short-circuiting	0.030	60 (100) 140	14 (15) 16	100% CO_2
	0.035	90 (130) 150	16 (17) 20	75% Ar + 25% CO_2 98% Ar + 2% O

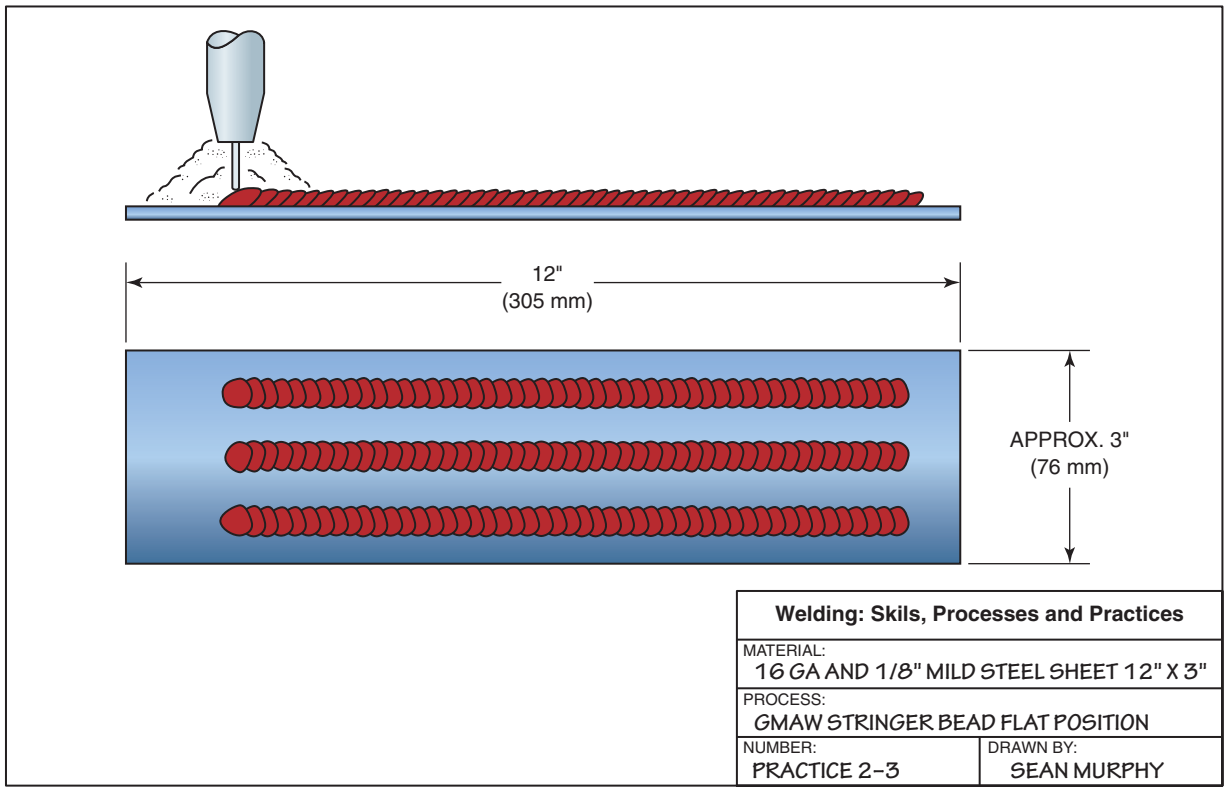

Figure 2.29 **Stringer beads in the flat position**

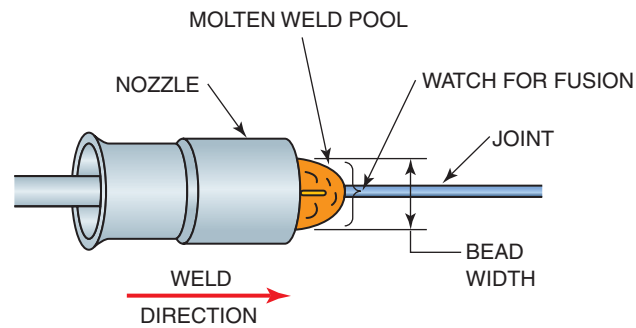

Figure 2.30 **Watching the weld in forehand welding**
The shielding gas nozzle restricts the welder's view of the weld bead when pushing.

satisfactory weld by watching the edge of the molten weld pool, the sparks, and the weld bead produced. Watching the leading edge of the molten weld pool (forehand welding, pushing technique) will show you the molten weld pool fusion and width. Watching the trailing edge of the molten weld pool (backhand welding, dragging technique) will show you the amount of buildup and the relative heat input, Figure 2.31. The quantity and size of sparks produced can indicate the relative location of the filler wire in the molten weld pool. The number of sparks will increase as the wire strikes the solid metal ahead of the molten weld pool. The gun itself will begin to vibrate or bump as the wire momentarily pushes against the cooler, unmelted base metal before it melts. Changes in weld width, buildup, and proper joint tracking can be seen by watching the bead as it appears from behind the shielding gas nozzle.

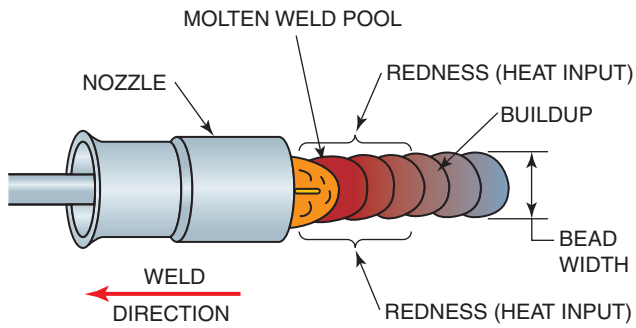

Figure 2.31 Watching the weld in backhand welding
Watch the trailing edge of the molten weld pool.

Repeat each type of bead as needed until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-4

Flat Position Butt Joint, Lap Joint, and Tee Joint

Using the same equipment, materials, and procedures listed in Practice 2-3, make welded butt joints, lap joints, and tee joints in the flat position, Figure 2.32A, 2.32B and 2.32C.

- Tack weld the sheets together and place them flat on the welding table, Figure 2.33.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Flat" position portion of the all-position requirement.

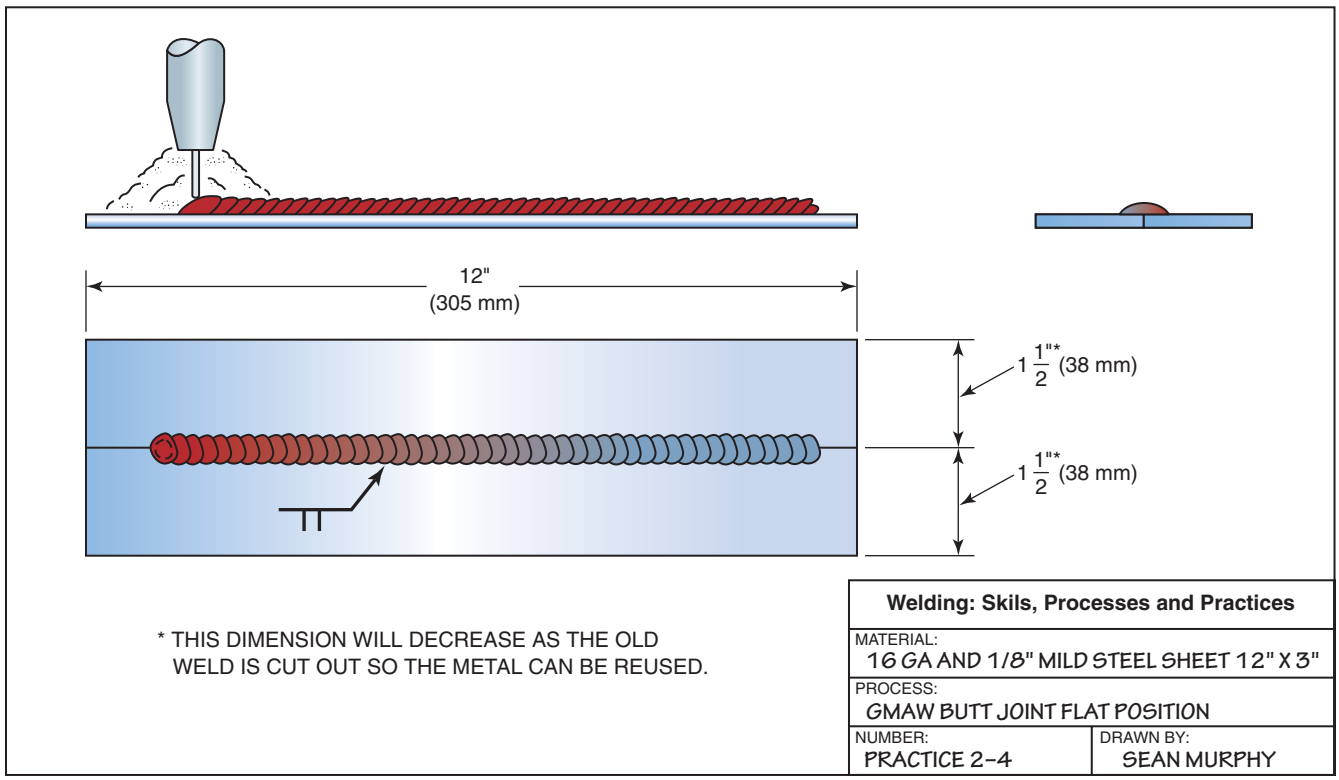

Figure 2.32 (A) Butt joint in the flat position

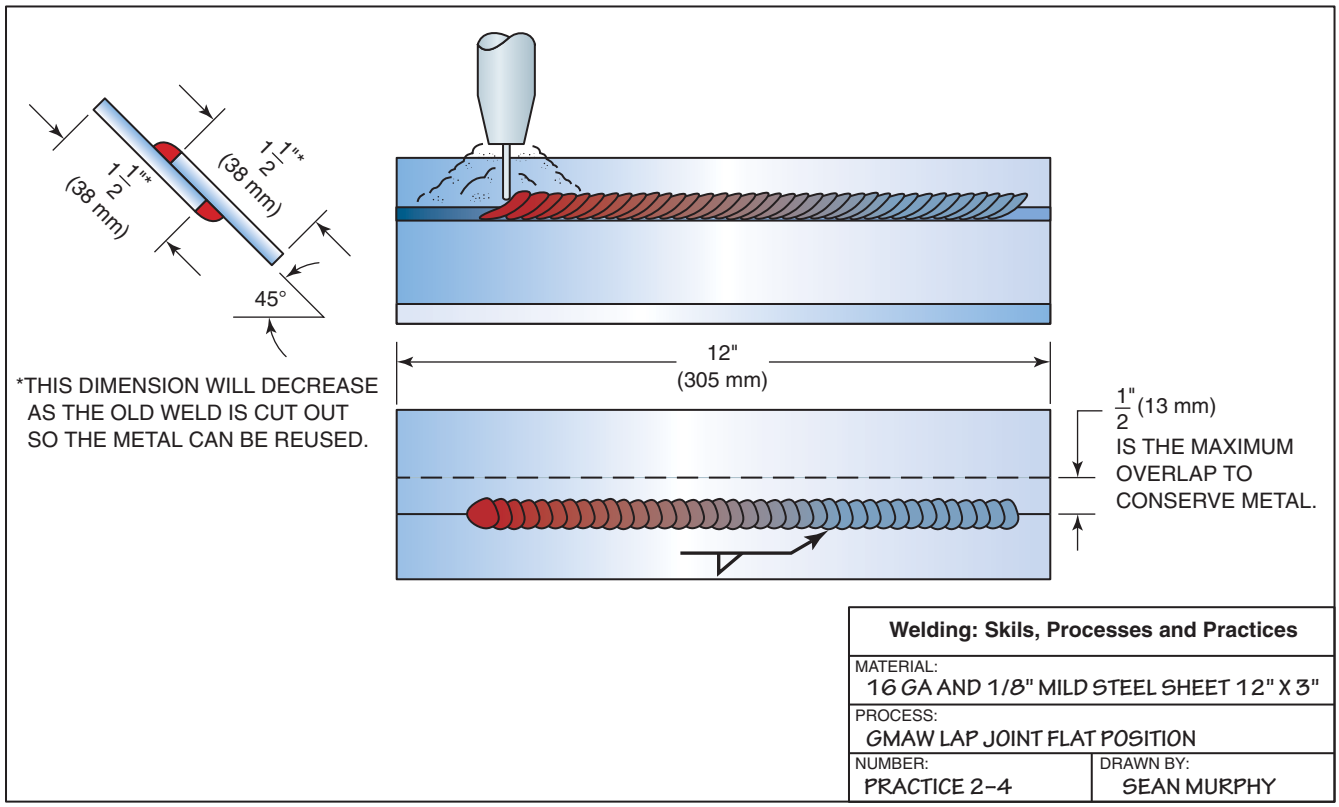

Figure 2.32 (B) Lap joint in the flat position

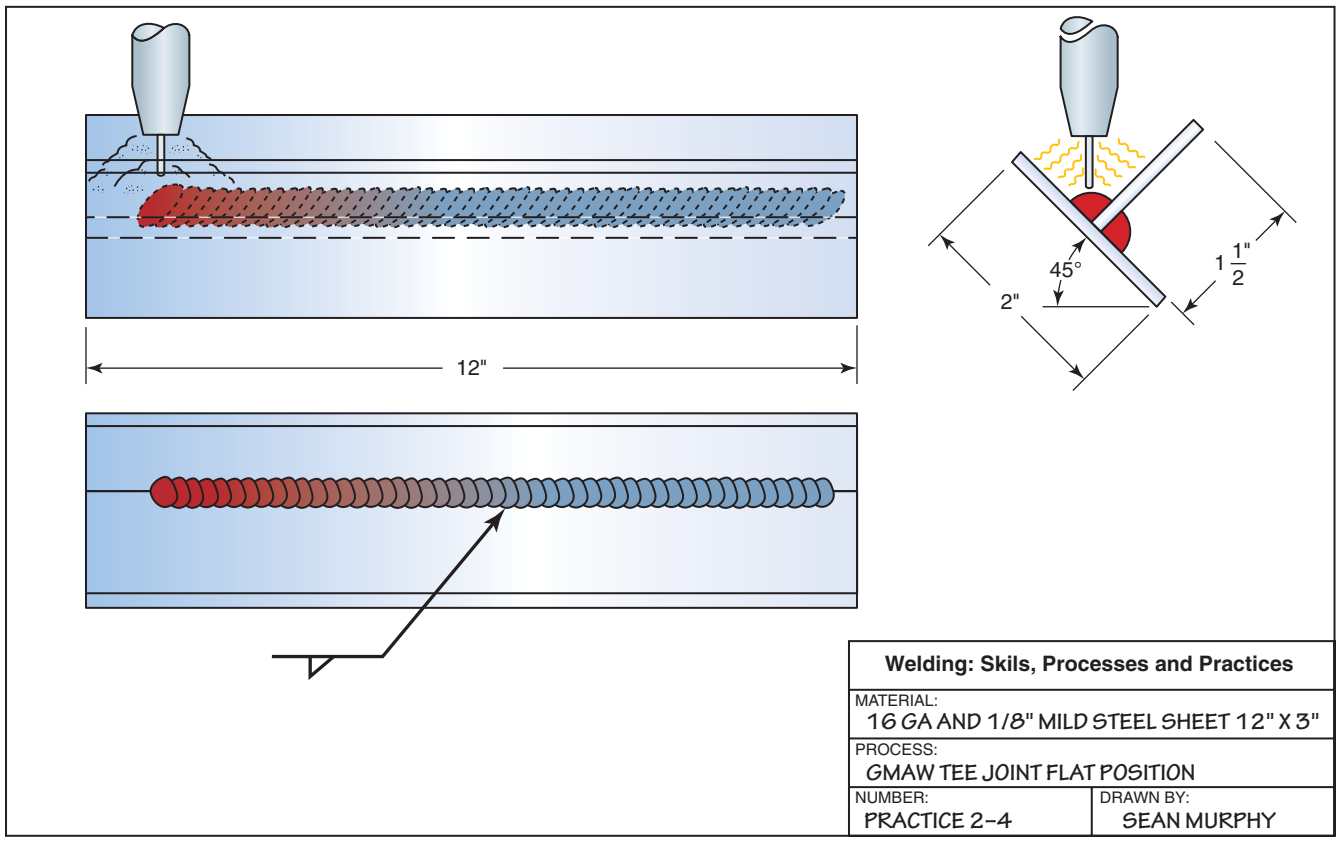

Figure 2.32 (C) Tee joint in the flat position

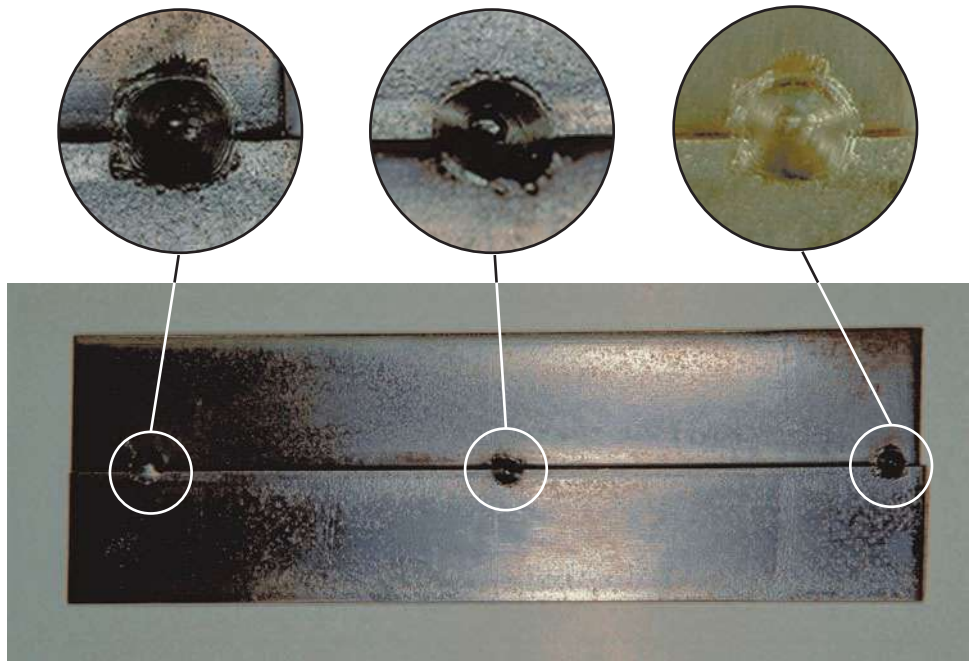

Figure 2.33 Tack welds
Use enough tack welds to keep the joint in alignment during welding. Small tack welds are easier to weld over without adversely affecting the weld.
Source: Courtesy of Larry Jeffus

- Starting at one end, run a bead along the joint. Watch the molten weld pool and bead for signs that a change in technique may be required.
- Make any needed changes as the weld progresses. By the time the weld is complete, you should be making the weld nearly perfectly.
- Using the same technique that was established in the last weld, make another weld. This time, the entire 12 in. (305 mm) of weld should be flawless.

Repeat each type of joint with both thicknesses of metal until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-5

Flat Position Butt Joint, Lap Joint, and Tee Joint, All with 100% Penetration

Using the same equipment, materials, and setup listed in Practice 2-3, make a welded joint in the flat position with 100% penetration along the entire 12-in. (305-mm) length of the welded joint. Repeat each type of joint until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Flat" position portion of the all-position requirement.

VERTICAL UP 3G AND 3F POSITIONS

PRACTICE 2-6

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

Stringer Bead at a 45° Vertical Up Angle

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make a vertical up stringer bead on a plate at a 45° inclined angle.

Start at the bottom of the plate and hold the welding gun at a slight pushing or upward angle to the plate, Figure 2.34. Brace yourself, lower your hood, and begin to weld. Depending upon the machine settings and type of shielding gas used, you will make a weave pattern.

If the molten weld pool is large and fluid (hot), use a "C" or "J" weave pattern to allow a longer time for the molten weld pool to cool, Figure 2.35. Do not make the weave so long or fast that the wire is allowed to strike the metal ahead of the molten weld pool. If this happens, spatter increases and a spot or zone of incomplete fusion may occur, Figure 2.36.

If the molten weld pool is small and controllable, use a small "C," zig-zag, or "J" weave pattern to control the width and buildup of the weld. A slower speed can also be used. Watch for complete fusion along the leading edge of the molten weld pool. Figure 2.37 shows a weld that did not fuse with the plate.

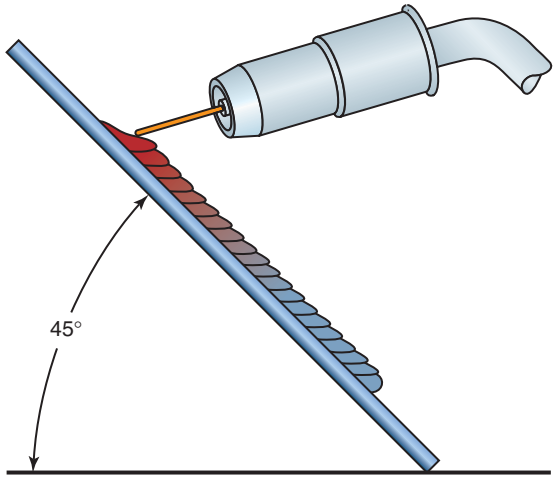

Figure 2.34 Vertical up position
Source: Photo courtesy of Larry Jeffus

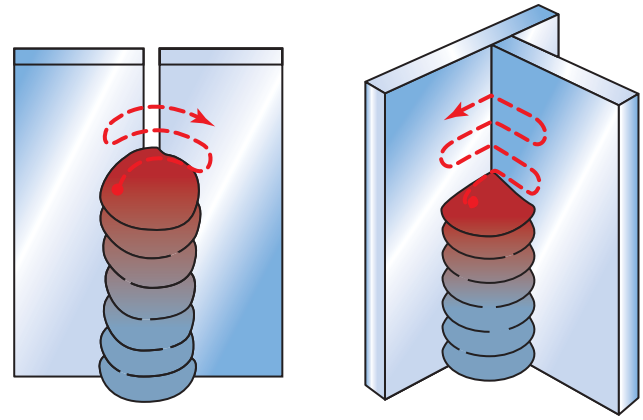

Figure 2.35 Vertical up weld patterns
Left, "C" pattern; right, "J" pattern.

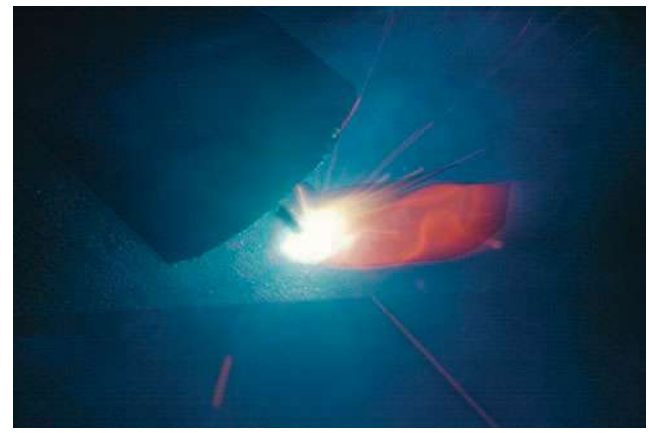

Figure 2.36 Burst of spatter caused by incorrect electrode contact with base metal
Source: Courtesy of Larry Jeffus

Figure 2.37 Weld separated from the plate
There is no fusion between the weld and plate.
Source: Courtesy of Larry Jeffus

A weld that is high and has little or no fusion is too "cold." Changing the welding technique will not correct this problem. The welder must stop welding and make the needed adjustments.

As the weld progresses up the plate, the back or trailing edge of the molten weld pool will cool, forming a shelf to support the molten metal. Watch the shelf to be sure that molten metal does not run over, forming a drip. When it appears that the metal may flow over the shelf, either increase the weave lengths, lengthen the electrode extension, or stop and start the current for brief moments to allow the weld to cool. Stopping for brief moments will not allow the shielding gas to be lost.

Continue to weld along the entire 12-in. (305-mm) length of the plate. Repeat this weld until a straight and uniform weld bead is produced. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-7

Stringer Bead in the Vertical Up Position

Repeat Practice 2-6 and increase the angle of the plate until you have mastered a straight and uniform weld bead in the vertical up position. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 4

PRACTICE 2-8

Butt Joint, Lap Joint, and Tee Joint in the Vertical Up Position at a 45° Angle

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make vertical up welded joints on a plate at a 45° inclined angle.

Tack weld the metal pieces together and brace them in position. Check to see that you have free movement along the entire joint to

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

prevent stopping and restarting during the weld. Avoiding stops and starts both speeds up the welding time and eliminates discontinuities.

The weave pattern should allow for adequate fusion on both edges of the joint. Watch the edges to be sure that they are being melted so that adequate fusion and penetration occur.

Repeat each type of joint as needed until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-9

Butt Joint, Lap Joint, and Tee Joint in the Vertical Up Position with 100% Penetration

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

Using the same equipment, materials, and setup as listed in Practice 2-3, you will increase the plate angle gradually as you develop skill until you are making satisfactory welds in the vertical up position.

Repeat each type of joint as needed until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

VERTICAL DOWN 3G AND 3F POSITIONS

The vertical down welding technique can be useful when making some types of welds. The major advantages of this technique are the following:

- Speed—Very high rates of travel are possible.
- Shallow penetration—Thin sections or root openings can be welded with little burn-through.
- Good bead appearance—The weld has a nice width-to-height ratio and is uniform.

Vertical down welds are often used on thin sheet metals or in the root pass in grooved joints. The combination of controlled penetration and higher welding speeds makes vertical down the best choice for such welds. The ease with which welds having a good appearance can be made is deceiving. Generally, more skill is required to make sound welds with this technique than in the vertical up position. The most common problem with these welds is lack of fusion or overlap. To prevent these problems, the arc must be kept at or near the leading edge of the molten weld pool.

PRACTICE 2-10

Stringer Bead at a 45° Vertical Down Angle

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-Circuit Transfer
Key Indicator 4

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make a vertical down stringer bead on a plate at a 45° inclined angle.

Holding the welding gun at the top of the plate with a slight dragging angle, Figure 2.38, will help to increase penetration, hold back the molten weld pool, and improve visibility of the weld. Be sure that your movements along the 12-in. (305-mm) length of plate are unrestricted.

Lower your hood and start the weld. Watch both the leading edge and the sides of the molten weld pool for fusion. The leading edge should flow into the base metal, not curl over it. The sides of the molten weld pool should also show fusion into the base metal and not be flashed (ragged) along the edges.

The weld may be made with or without a weave pattern. If a weave pattern is used, it should be a "C" pattern. The "C" should follow the leading edge of the weld. Some changes on the gun angle may help to increase penetration. Experiment with the gun angle as the weld progresses.

Repeat these welds until you have established a rhythm and technique that work well for you. The welds must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Figure 2.38 **Vertical down position**
Source: Courtesy of Larry Jeffus

PRACTICE 2-11

Stringer Bead in the Vertical Down Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-Circuit Transfer
Key Indicator 4

Repeat Practice 2-10 and increase the angle of the plate until you have developed the skill to repeatedly make good welds in the vertical down position. The weld bead must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-12

Butt Joint, Lap Joint, and Tee Joint in the Vertical Down Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

Using a properly set up and adjusted GMA welding machine, Table 2.7, proper safety protection, 0.035-in. and/or 0.045-in. (0.9-mm and/or 1.2-mm) diameter wire, and two or more pieces of mild steel sheet 12 in. (305 mm) long and 16 gauge and 1/8 in. (3 mm) thick, you will make vertical down welded joints.

Holding the welding gun at the top of the plate with a slight dragging angle, Figure 2.38, will help to increase penetration, hold back the molten weld pool, and improve visibility of the weld. Be sure that your movements along the 12-in. (305-mm) length of plate are unrestricted.

Lower your hood and start the weld. Watch both the leading edge and sides of the molten weld pool for fusion. The leading edge should flow into the base metal, not curl over it. The sides of the molten weld pool should also show fusion into the base metal and not be flashed (ragged) along the edges.

The weld may be made with or without a weave pattern. If a weave pattern is used, it should be a "C" pattern. The "C" should follow the leading edge of the weld. Some changes on the gun angle may help to increase penetration. Experiment with the gun angle as the weld progresses.

Repeat these welds until you have established a rhythm and technique that work well for you. The welds must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-13

Butt Joint and Tee Joint in the Vertical Down Position with 100% Penetration

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make welded joints with 100% weld penetration.

It may be necessary to adjust the root opening to meet the penetration requirements. The lap joint was omitted from this practice because little additional skill can be developed with it that is not already acquired

with the tee joint. Repeat each type of joint until consistently good welds are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

HORIZONTAL 2G AND 2F POSITIONS

PRACTICE 2-14

Horizontal Stringer Bead at a 45° Angle

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 4

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make a horizontal stringer bead on a plate at a 45° reclined angle.

Start at one end with the gun pointed in a slightly upward direction, Figure 2.39. You may use a pushing or a dragging (a leading or a trailing) gun angle, depending upon the current setting and penetration desired. Undercutting along the top edge and overlap along the bottom edge are problems with both gun angles. Careful attention must be paid to the manipulation "weave" technique used to overcome these problems.

The most successful weave patterns are the "C" and "J" patterns. The "J" pattern is the most frequently used. It allows weld metal to be deposited along a shelf created by the previous weave, Figure 2.40. The length of the "J" can be changed to control the weld bead size. Smaller weld beads are easier to control than large ones.

Repeat these welds until you have established the rhythm and technique that work well for you. The weld must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

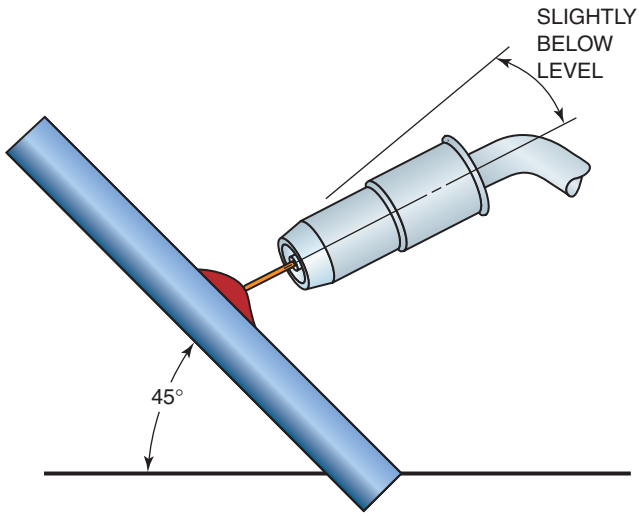

Figure 2.39 45° horizontal position

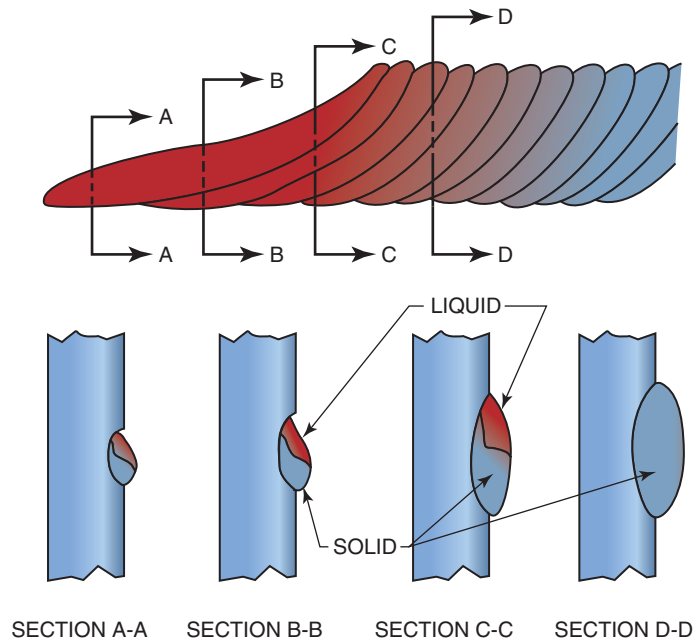

Figure 2.40 Sections of a "J" weave pattern
The actual size of the molten weld pool remains small along the weld.

PRACTICE 2-15

Stringer Bead in the Horizontal Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Repeat Practice 2-14 and increase the angle of the plate until you have developed the skill to repeatedly make good horizontal welds on a vertical surface. The weld bead must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-16

Butt Joint, Lap Joint, and Tee Joint in the Horizontal Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Horizontal" position portion of the all-position requirement.

Using the same equipment, materials, and setup listed in Practice 2-3, you will make horizontal welded joints.

Tack weld the pieces of metal together and brace them in position using the same skills developed in Practice 2-14. Starting at one end, make a weld along the entire length of the joint. When making the butt or lap joints, it may help to recline the plates at a 45° angle until you have developed the technique required. Repeat each type of joint as needed until consistently good welds are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-17

Butt Joint and Tee Joint in the Horizontal Position with 100% Penetration

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make overhead joints having 100% penetration in the horizontal position.

It may be necessary to adjust the root opening to meet the penetration requirements. Repeat each type of joint as needed until consistently good welds are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Horizontal" position portion of the all-position requirement.

OVERHEAD 4G AND 4F POSITIONS

There are several advantages to the use of short-circuiting arc metal transfer in the overhead position, including the following:

- Small molten weld pool size—The smaller size of the molten weld pool allows surface tension to hold it in place. Less molten weld pool sag results in improved bead contour with less undercut and fewer icicles, Figure 2.41.
- Direct metal transfer—The direct metal transfer method does not rely on other forces to get the filler metal into the molten weld pool. This results in efficient metal transfer and less spatter and loss of filler metal.

PRACTICE 2-18

Stringer Bead in the Overhead Position

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make a welded stringer bead in the overhead position.

The molten weld pool should be kept as small as possible for easier control. A small molten weld pool can be achieved by using lower current settings, by using a longer wire stickout, by traveling faster, or by pushing the molten weld pool. The technique used is the welder's choice. Often a combination of techniques can be used with excellent results.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 4

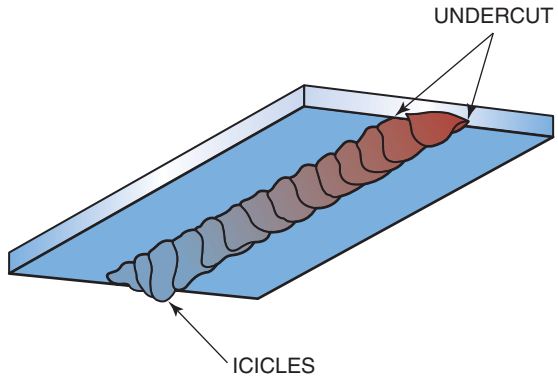

Figure 2.41 Overhead weld

Lower current settings require closer control of gun manipulation to ensure that the wire is fed into the molten weld pool just behind the leading edge. The low power will cause overlap and more spatter if this contact position of wire to molten weld pool is not closely maintained.

Faster travel speeds allow the welder to maintain a high production rate even if multiple passes are required to complete the weld. Weld penetration into the base metal at the start of the bead can be obtained by using a slow start or quickly reversing the weld direction. Both the slow start and reversal of weld direction put more heat into the start to increase penetration, Figure 2.42. The higher speed also reduces the amount of weld distortion by reducing the amount of time that heat is applied to a joint.

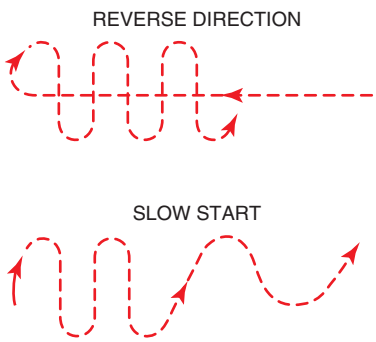

Figure 2.42 Two methods of concentrating heat at the beginning of a weld bead to aid in penetration depth

The pushing, or trailing, gun angle forces the bead to be flatter by spreading it out over a wider area as compared to the bead resulting from a dragging, or backhand, gun angle. The wider, shallow molten weld pool cools faster, resulting in less time for sagging and the formation of icicles.

When welding overhead, extra personal protection is required to reduce the danger of burns. Wear leather sleeves or a leather jacket, and a cap.

Much of the spatter created during overhead welding falls into the shielding gas nozzle. The effectiveness of the shielding gas is reduced, Figure 2.43, and the contact tube may short out to the gas nozzle, Figure 2.44. Turbulence caused by the spatter obstructing the gas may lead to weld contamination. The shorted gas nozzle may arc to the work, causing damage both to the nozzle and to the plate. To control the amount of spatter, a longer stickout and/or a sharper gun-to-plate angle is required to allow most of the spatter to fall clear of the gas nozzle. The nozzle can be dipped, sprayed, or injected automatically, Figure 2.45, with antispatter to help prevent the spatter from sticking. Applying antispatter will not stop the spatter from building up, but it does make its removal much easier. Ensure that the antispatter gel does not restrict the flow of shield gas from the nozzle.

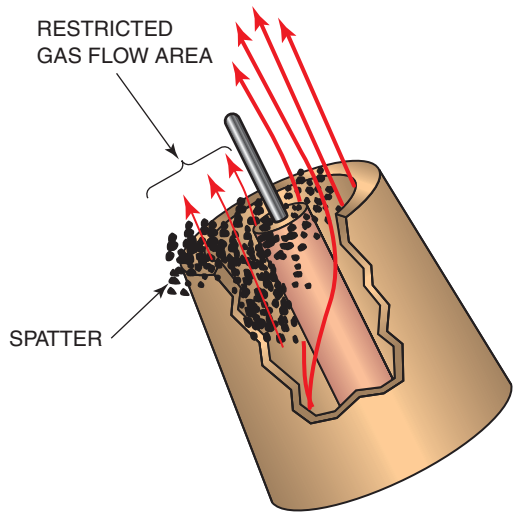

Figure 2.43 Shielding gas flow affected by excessive weld spatter in nozzle

Figure 2.44 Gas nozzle damaged after shorting out against the work
Source: Courtesy of Larry Jeffus

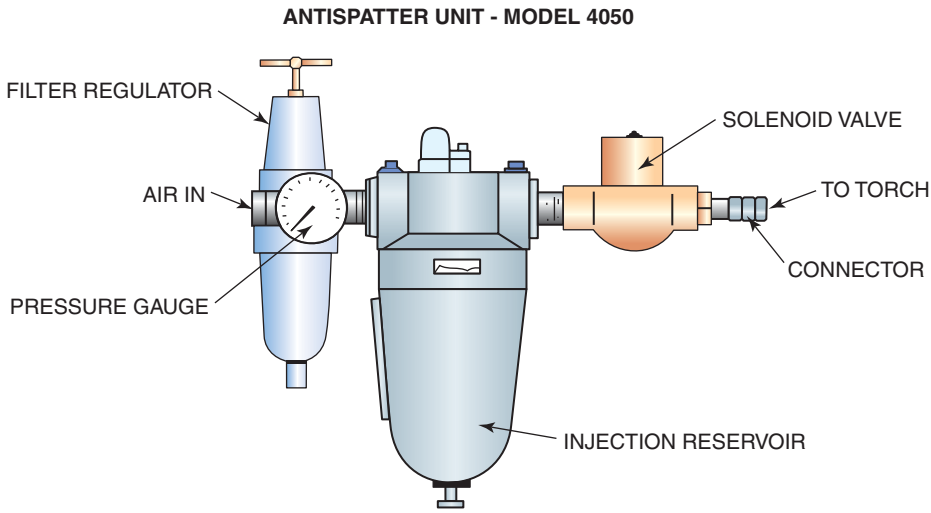

Figure 2.45 **Automatic antispatter system that can be added to a GMA welding gun**

Make several short weld beads using various techniques to establish the method that is most successful and most comfortable for you. After each weld, stop and evaluate it before making a change. When you have decided on the technique to be used, make a welded stringer bead that is 12 in. (305 mm) long.

Repeat the weld until it can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-19

Butt Joint, Lap Joint, and Tee Joint in the Overhead Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Overhead" position portion of the all-position requirement.

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make an overhead welded joint.

Tack weld the pieces of metal together and secure them in the overhead position. Be sure you have an unrestricted view and freedom of movement along the joint. Start at one end and make a weld along the joint. Use the same technique developed in Practice 2-18.

Repeat the weld until it can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-20

Butt Joint and Tee Joint in the Overhead Position with 100% Penetration

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Overhead" position portion of the all-position requirement.

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make overhead welded joints having 100% penetration.

Tack weld the metal together. It may be necessary to adjust the root opening to allow 100% weld metal penetration. During these welds, it may be necessary to use a dragging, or backhand, torch angle. When

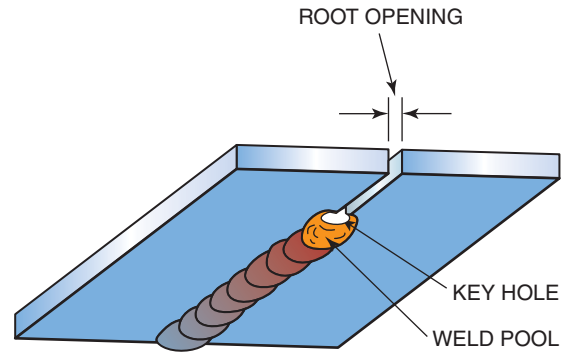

Figure 2.46 Overhead welding

used with a "C" or "J" weave pattern, this torch angle helps to achieve the desired depth of penetration. A key hole just ahead of the molten weld pool is a good sign that the metal is being penetrated, Figure 2.46.

Repeat the weld until it can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

AXIAL SPRAY TRANSFER

PRACTICE 2-21

Stringer Bead, 1G Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Spray Transfer
Key Indicator 9

Using a properly set up and adjusted GMA welding machine (see Table 2.8), proper safety protection, 0.035-in. and/or 0.045-in. (0.9-mm and/or 1.2-mm)-diameter wire, and two or more pieces of mild steel plate 12 in. (305 mm) long × 1/4 in. (6 mm) thick, you will make a welded stringer bead in the flat position.

Start at one end of the plate and use either a push or drag technique to make a weld bead along the entire 12-in. (305-mm) length of the metal using spray or pulsed-arc metal transfer. After the weld is complete, check its appearance and make any changes needed to correct the weld, Figure 2.47. Repeat the weld and make any additional adjustments required. After the machine is set, start working on improving the straightness and uniformity of the weld. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Table 2.8 Typical Welding Current Settings for Axial Spray Metal Transfer for Mild Steel

Process	Wire Diameter	Amperage Range (Optimum)	Voltage Range (Optimum)	Shielding Gas
Axial spray	0.030	115–200	15–27	98% Ar + 2% O_2
	0.035	165–300	18–32	
	0.045	200–450	20–34	

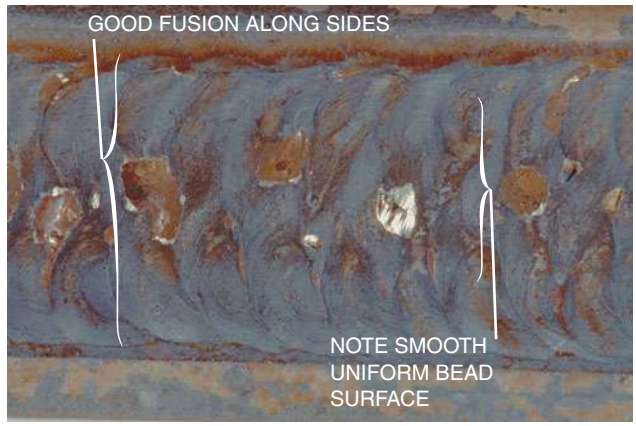

Figure 2.47 Weld bead made with GMAW axial spray metal transfer
Source: Courtesy of Larry Jeffus

PRACTICE 2-22

Butt Joint, Lap Joint, and Tee Joint Using the Spray Transfer Method

Using the same equipment, materials, and setup as listed in Practice 2-21, you will make a flat and horizontal weld using spray transfer or pulsed-spray metal transfer, Figure 2.48.

Tack weld the metal together and place the assembly in the flat position on the welding table. Start at one end and make a uniform weld along the entire 12-in. (305-mm) length of the joint. Watch the sides of the fillet weld for signs of undercutting.

Repeat the weld until it can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Spray Transfer
Key Indicator 10, 11

PRACTICE 2-23

Butt Joint and Tee Joint

Using the same equipment, materials, and setup as listed in Practice 2-21, you will make a flat weld using spray transfer. Each weld must pass the

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Spray Transfer
Key Indicator 10, 11

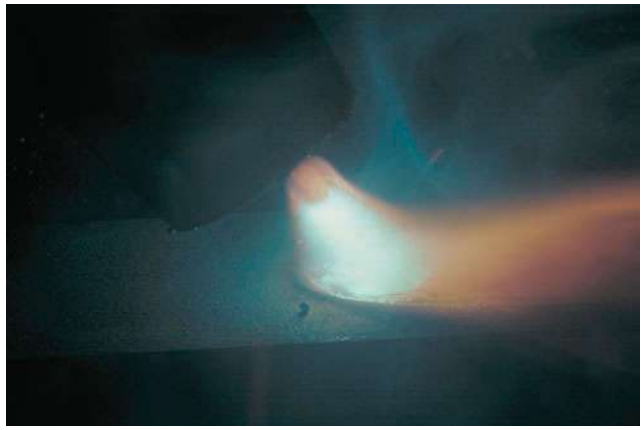

Figure 2.48 GMAW axial spray metal transfer
Source: Courtesy of Larry Jeffus

guided bend test. Repeat each type of weld joint as needed until the bend test can be passed. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-24

Gas Metal Arc Welding—Short-Circuit Metal Transfer (GMAW-S) Workmanship Sample

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 3
Key Indicator 1, 2, 3

Module 5
Short-circuit Transfer
Key Indicator 3, 4, 5, 6, 7

Module 9
Key Indicator 1, 2

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 2-24 Date:

Title:
Welding GMAW-S of plate to plate.

Scope:
This procedure is applicable for square groove and fillet welds within the range of 10 ga. (3.4 mm) through 14 ga. (1.9 mm).

Welding may be performed in the following positions: all.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1 Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. E70S-X from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity:

CO_2 at 30 to 50 cfh or 75% Ar/25% CO_2 at 30 to 50 cfh.

Joint Design and Tolerances:

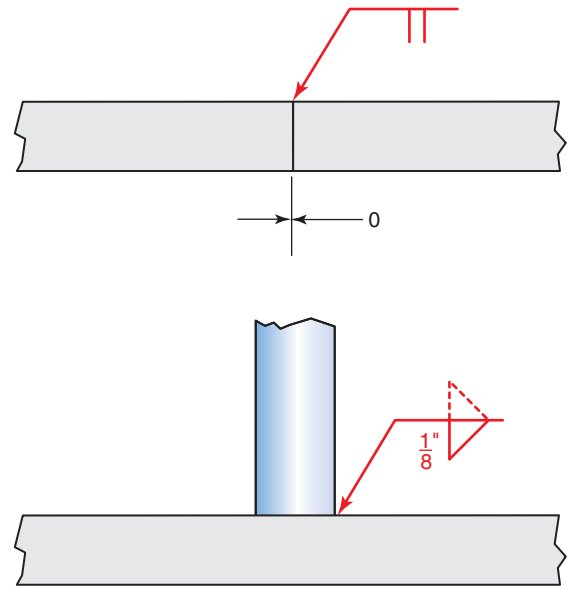

Preparation of Base Metal:
All parts may be mechanically cut or machine PAC unless specified as manual PAC.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct-current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70S-X	0.035 in. (0.9 mm)	90 to 120	180 to 300 (457 to 762)	15 to 19	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70S-X	0.045 in. (1.2 mm)	130 to 200	125 to 200 (318 to 508)	17 to 20	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 1/2-in. (13-mm) or larger gas nozzle for all welding, first tack weld the plates together according to the drawing. Use the E70S-X filler metal to fuse the plates together. Clean any silicon slag, being sure to remove any trapped silicon slag along the sides of the weld.

Using the E70S-X arc welding electrodes, make a series of stringer beads, no thicker than 3/16 in. (4.7 mm). The 1/8-in. (3.1-mm) fillet welds are to be made with one pass. All welds must be placed in the orientation shown in the drawing.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
Any slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position.

Visual Inspection:
Visually inspect the weld for uniformity and discontinuities.

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm).
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.
6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm)
7. Welds shall be free from overlap.

Sketches:
See Figure 2.49.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-25

This practice parallels an AWS D1.1 structural steel limited thickness welder performance qualification test.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 3
Key Indicator 1, 2, 3

Module 5
Short-circuit Transfer
Key Indicator 3, 4, 5, 6

Module 9
Key Indicator 1, 2

Gas Metal Arc Welding—Short-Circuit Metal Transfer (GMAW-S) Limited Thickness Welder Performance Qualification Test Plate for 2G, 3G, and 4G Positions without Backing

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 2-25 Date:

Title:
Welding GMAW-S of plate to plate.

Scope:
This procedure is applicable for V-groove, bevel, or single-bevel welds within the range of 1/8 in. (3.2 mm) through 3/4 in. (19 mm).

Welding may be performed in the following positions: all.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1 Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. E70S-X from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

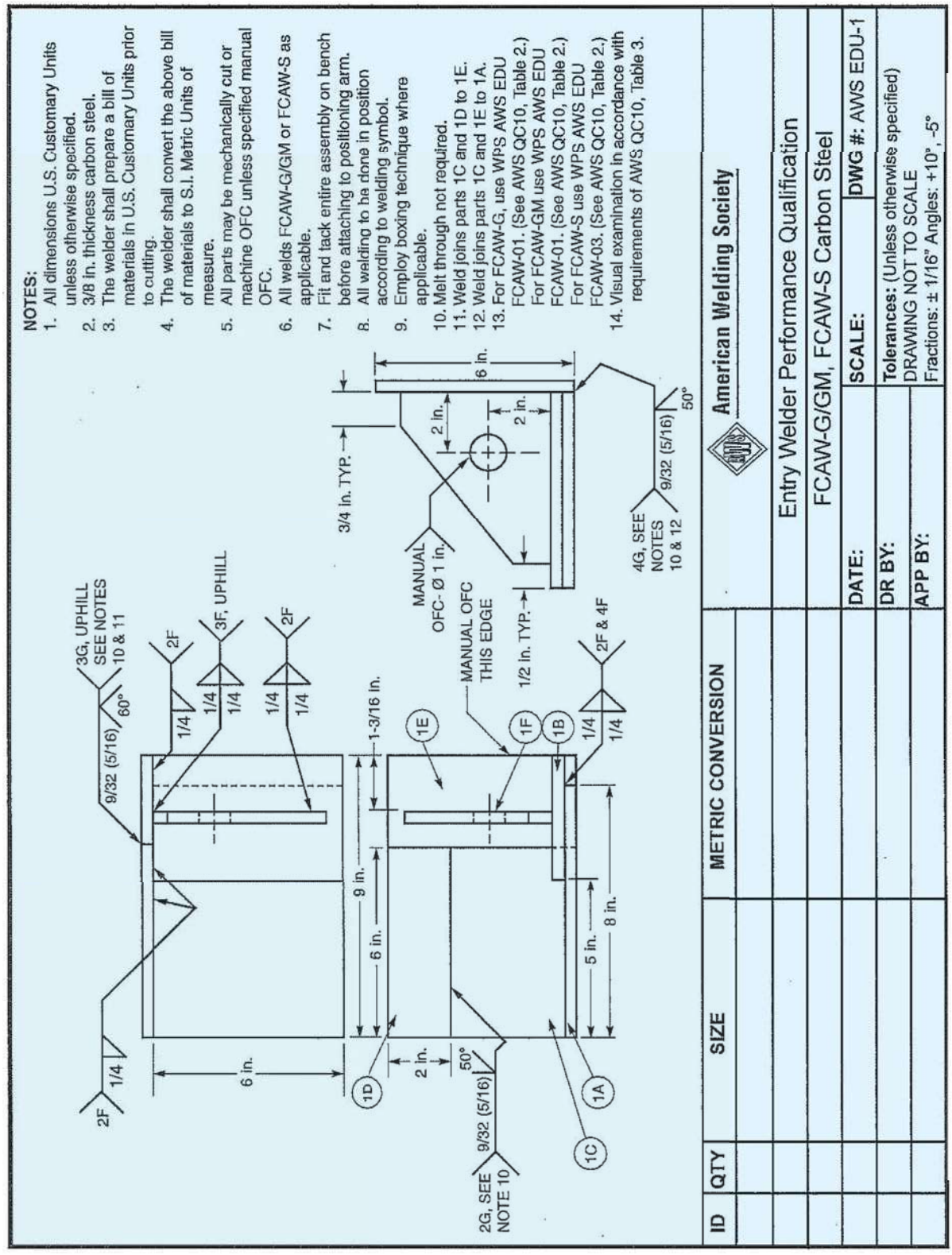

Figure 2.49 GMAW-S workmanship qualification test
Source: Courtesy of the American Welding Society

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity: CO_2 at 30 to 50 cfh or 75% Ar/25% CO_2 at 30 to 50 cfh.

Joint Design and Tolerances:

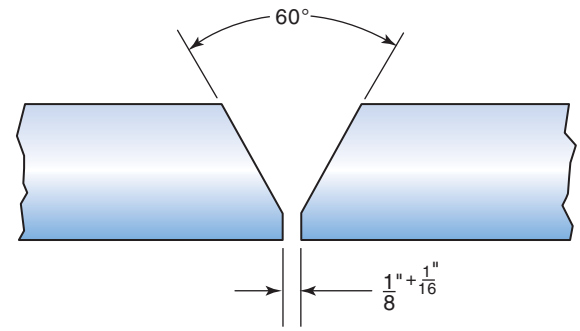

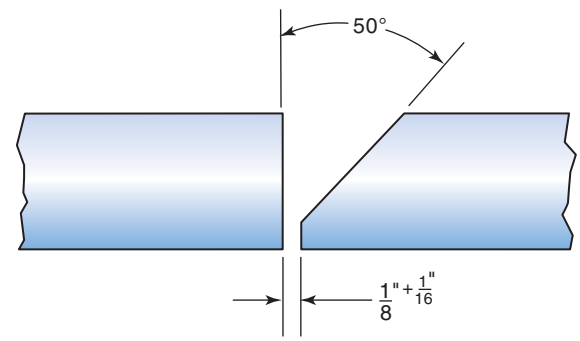

Preparation of Base Metal:
The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct-current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70S-X	0.035 in. (0.9 mm)	90 to 120	180 to 300 (457 to 762)	15 to 19	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70S-X	0.045 in. (1.2 mm)	130 to 200	125 to 200 (318 to 508)	17 to 20	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 1/2-in. (13-mm) or larger gas nozzle for all welding, first tack weld the plates together according to the drawing. There should be about a 1/8-in. (1.6-mm) root gap between the plates with V-grooved or beveled edges and 1/16-in. root faces. Use the E70S-X filler wire to make a root pass to fuse the plates together. Clean any silicon slag from the root pass, being sure to remove any trapped silicon slag along the sides of the weld.

Using the E70S-X filler wire, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled. Note: The horizontal (2G) weldment should be made with stringer beads only.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
Any slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position.

Visual Inspection:*
Visually inspect the weld for uniformity and discontinuities.

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm)
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.

*From Table 3, AWS SENSE QC-10 2004 (Courtesy of the American Welding Society)

6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

Bend Test:
The weld is to be mechanically tested only after it has passed the visual inspection. Be sure that the test specimens are properly marked to identify the welder, the position, and the process.

Specimen Preparation:
For 3/8-in. test plates, two specimens are to be located in accordance with the requirements of the figure below left. One is to be prepared for a "transverse face bend," and the other is to be prepared for a "transverse root bend."
Transverse face bend. The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld face becomes the tension surface of the specimen. Transverse face-bend specimens shall comply with the requirements of the figure below bottom.
Transverse root bend. The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld root becomes the tension surface of the specimen. Transverse face-bend specimens shall comply with the requirements of the figure below bottom.

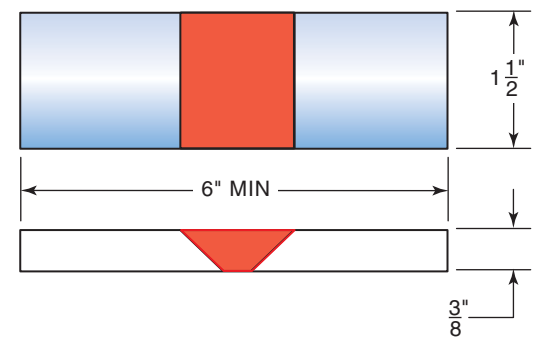

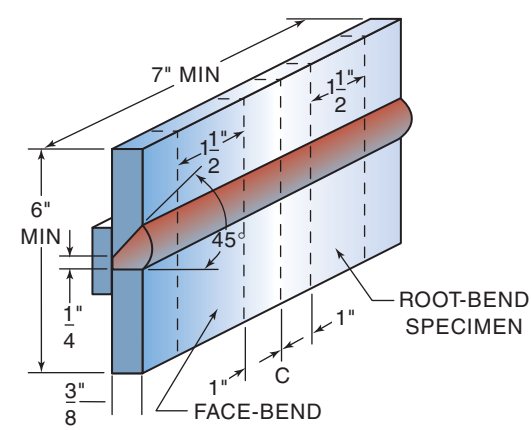

Acceptance Criteria for Face and Root Bends:*
For acceptance, the convex surface of the face- and root-bend specimens shall meet both of the following requirements:

*From Table 4, AWS-SENSE QC-10:2004 (Courtesy of the American Welding Society)

1. No single indication shall exceed 1/8 in. (3.2 mm), measured in any direction on the surface.
2. The sum of the greatest dimensions of all indications on the surface, which exceed 1/32 in. (0.8 mm) but are less than or equal to 1/8 in. (3.2 mm), shall not exceed 3/8 in. (9.6 mm).

Cracks occurring at the corner of the specimens shall not be considered unless there is definite evidence that they result from slag inclusion or other internal discontinuities.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-26

Gas Metal Arc Welding Spray Transfer (GMAW) Workmanship Sample

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 3
Key Indicator 1, 2, 3

Module 5
Spray Transfer
Key Indicator 8, 9, 10, 11, 12

Module 9
Key Indicator 1, 2

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 2-26. Date:

Title:
Welding GMAW of plate to plate.

Scope:
This procedure is applicable for V-groove and fillet welds within the range of 1/8 in. (3.2 mm) through 1-1/2 in. (38 mm).

Welding may be performed in the following positions: 1G, 1F, 2F.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1, Group 1 or 2.

Backing material specification: none.

Filler Metal:
The 0.035 to 0.045 dia. filler metal shall conform to AWS specification no. E70S-X from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions: 98% Ar/2% O_2 or 90% Ar/10% CO_2.

Preparation of Base Metal:
The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct-current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70S-X	0.035 in. (0.9 mm)	180 to 230	400 to 550 (1016 to 1397)	25 to 27	Ar plus 2% O_2 or 90% Ar/ 10% CO_2	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70S-X	0.045 in. (1.2 mm)	260 to 340	300 to 500 (762 to 1270)	25 to 30	Ar plus 2 O_2 or 90% Ar/ 10% CO_2	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 3/4-in. (19-mm) or larger gas nozzle for all welding, first tack weld the plates together according to the drawing. There should be about a 1/16-in. (1.6-mm) root gap between the plates with V-grooved or beveled edges. Use the E70S-X arc welding electrodes to make a root pass to fuse the plates together. Clean any silicon slag from the root pass, being sure to remove any trapped silicon slag along the sides of the weld.

Using the E70S-X arc welding electrodes, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled. The 1/4-in. (6.4-mm) fillet welds are to be made with one pass.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
Any slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position.

Inspection:*
Visually inspect the weld for uniformity and discontinuities.

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.

*From Table 3, AWS SENSE QC-10:2004 (Courtesy of the American Welding Society)

4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm)
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.
6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

Sketches:
See Figure 2.50.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-27

Gas Metal Arc Welding (GMAW) Spray Transfer Limited Thickness Welder Performance Qualification Test Plate for 1-G, Position, with Backing

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 3
Key Indicator 1, 2, 3

Module 5
Spray Transfer
Key Indicator 8, 9, 11

Module 9
Key Indicator 1, 2

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice (2-27). Date:

Title:
Welding GMAW of plate to plate.

Scope:
This procedure is applicable for V-groove and fillet welds within the range of 1/8 in. (3.2 mm) through 1-1/2 in. (38 mm).

Welding may be performed in the following positions: 1G, 2F.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1, Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. 0.035 to 0.045 dia. E70S-X from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity: 98% Ar/2% O_2 or 90% Ar/10% CO_2

Joint Design and Tolerances:

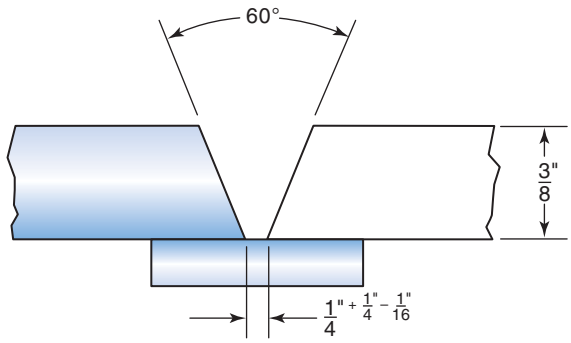

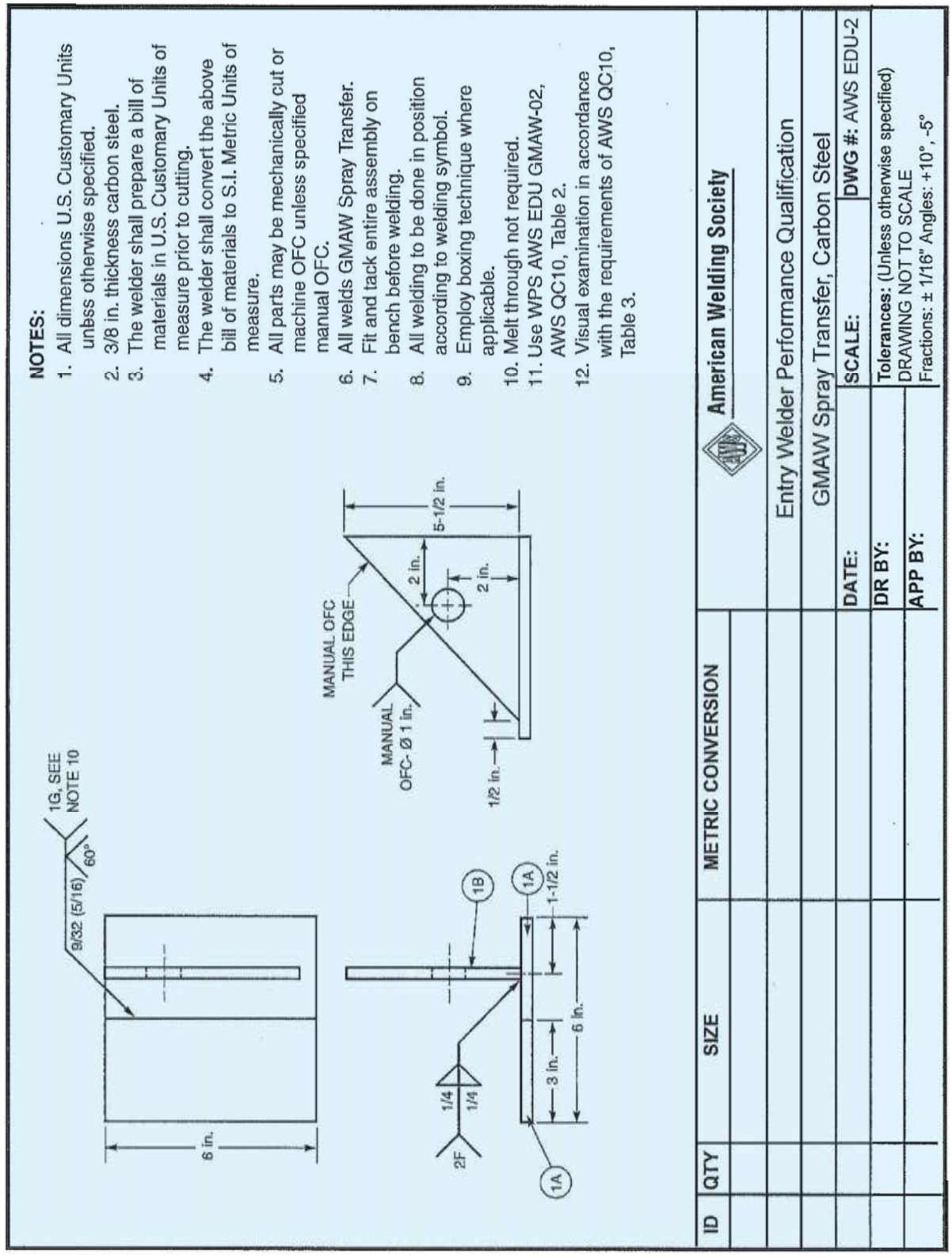

Figure 2.50 GMAW (spray transfer) workmanship qualification test
Source: Courtesy of the American Welding Society

Preparation of Base Metal:
The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct-current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70S-X	0.035 in. (0.9 mm)	180 to 230	400 to 550 (1016 to 1397)	25 to 28	Ar plus 2% O_2 or 90% Ar/10% CO_2	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70S-X	0.045 in. (1.2 mm)	260 to 340	300 to 500 (762 to 1270)	25 to 30	Ar plus 2 O_2 or 90% Ar/10% CO_2	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 3/4-in. (19-mm) or larger gas nozzle for all welding, first tack weld the plates and backing strip together according to the drawing below. There should be about a 1/4-in. (1.6-mm) root gap between the plates with V-grooved or beveled edges. Use the E70S-X arc welding electrodes to make a root pass to fuse the plates together. Clean any silicon slag from the root pass, being sure to remove any trapped silicon slag along the sides of the weld.

Using the E70S-X arc welding electrodes, make a series of stringer filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to

cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
Any slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position.

Visual Inspection:*
Visually inspect the weld for uniformity and discontinuities.

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm).
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.
6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

Sketches:

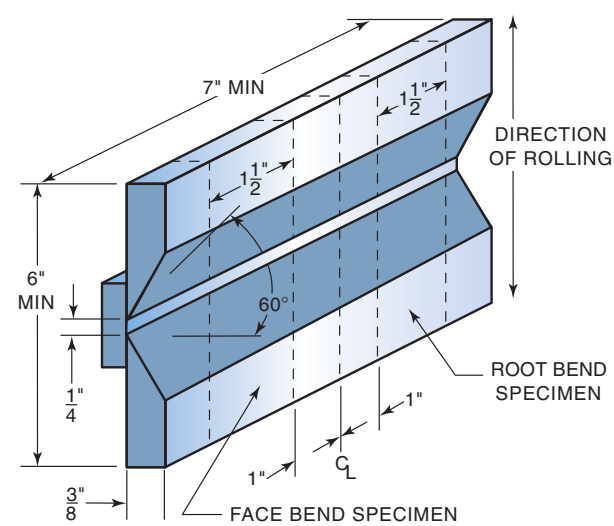

Bend-Test:
The weld is to be mechanically tested only after it has passed the visual inspection. Be sure that the test specimens are properly marked to identify the welder, the position, and the process.

Specimen Preparation
For 3/8-in. test plates, two specimens are to be located in accordance with the requirements below. One is to be prepared for a transverse face bend, and the other is to be prepared for a transverse root bend.

*From Table 3, AWS SENSE QC-10:2004 (Courtesy of the American Welding Society)

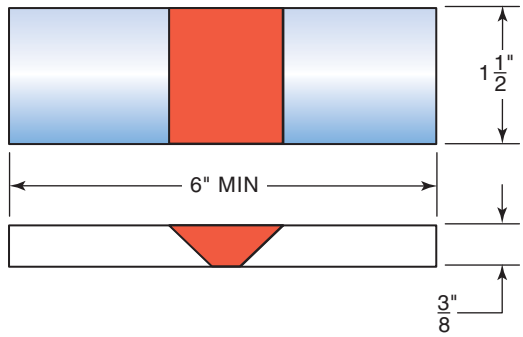

Transverse face bend. The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld face becomes the tension surface of the specimen. Transverse face bend specimens shall comply with the requirements below.

Transverse root bend. The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld root becomes the tension surface of the specimen. Transverse root bend specimens shall comply with the requirements below.

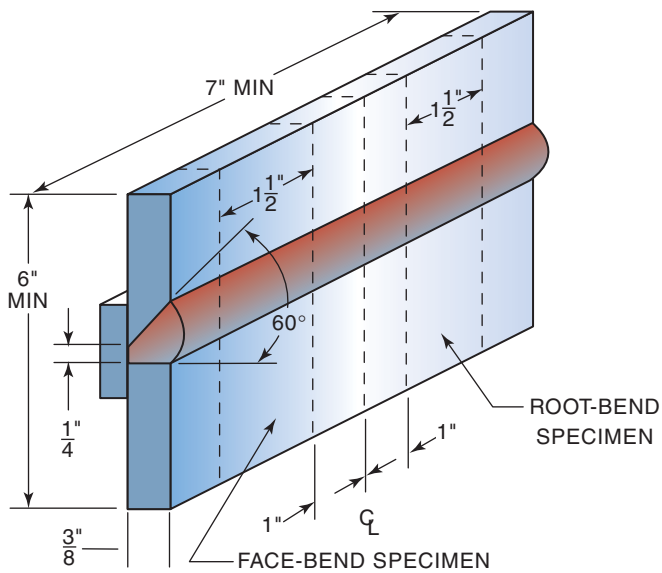

Acceptance Criteria for Face and Root Bends*

For acceptance, the convex surface of the face and root bend specimens shall meet both of the following requirements:

1. No single indication shall exceed 1/8 in. (3.2 mm), measured in any direction on the surface.
2. The sum of the greatest dimensions of all indications on the surface, which exceed 1/32 in. (0.8 mm), but are less than or equal to 1/8 in. (3.2 mm), shall not exceed 3/8 in. (9.6 mm).

Cracks occurring at the corner of the specimens shall not be considered unless there is definite evidence that they result from slag or inclusions or other internal discontinuities.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

*From Table 4, AWS SENSE QC-10:2004 (Courtesy of the American Welding Society)

SUMMARY

Slight changes in welding gun angle and electrode extension can make significant differences in the quality of the weld produced. As a new welder you might find it difficult to tell the effect of these changes if they are slight. Therefore, as you start to learn this process it is a good idea to make more radical changes so it is easier for you to see their effects on the weld. Later as you develop your skills you can use these slight changes to aid in controlling the weld's quality and appearance as it progresses along the joint. Small adjustments in your welding technique are required to compensate for slight changes that occur along a welding joint, such as joint gap and the increasing temperature of the base metal.

Variations in conditions can significantly affect welding setup for the GMA process. Before starting an actual weld in the field you should practice to test your setup. Practice on scrap metal of a similar thickness and type of metal to be welded. A practice weld before you begin welding can significantly increase the chances that your weld will meet standards or specifications. Making these sample or test welds is more important when you are welding in the field, since welds outside the shop are more difficult to control and anticipate. Think of this much as an athlete warms up before competing.

You will find it beneficial when you are initially setting up your welder to have someone assist you, so that he or she can make changes in the welding machine's settings as you are welding. This teamwork can significantly increase your setup accuracy and reduce setup time. Later on in the field, having developed a keen eye for watching the weld, you can then make these adjustments for yourself more rapidly and accurately. Working with another student in a group effort like this will also give you a better understanding of how other individuals' setup preferences affect their welds. Welding is an art, and therefore each welder may have slight differences in preference for voltage, amperage, gas flow, and other setup variables. This gives you an opportunity to learn more from others.

REVIEW

1. What items make up a basic semiautomatic welding system?
2. What must be done to the shielding gas cylinder before the valve protection cap is removed?
3. Why is the shielding gas valve "cracked" before the flowmeter regulator is attached?
4. What causes the electrode to bird-nest?
5. Why must all fittings and connections be tight?
6. What parts should be activated by depressing the gun switch?
7. What benefit does a welding wire's cast provide?
8. What can be done to determine the location of a problem that stops the wire from being successfully fed through the conduit?
9. What are the advantages of using a feed roller pressure that is as light as possible?

10. Why should the feed roller drag prevent the spool from coasting to a stop when the feed stops?
11. Why must you always wind the wire tightly into a ball or cut it into short lengths before discarding it in the proper waste container?
12. Why would the flowmeter ball float at different heights with different shielding gases if the shielding gases are flowing at the same rate?
13. Using Table 2.1, determine the amperage if 400 in. (10.2 m) of 0.45-in. (1.2-mm) steel wire is fed in one minute.
14. How is the amperage adjusted on a GMA welder?
15. What happens to the weld as the electrode extension is lengthened?
16. What is the effect on the weld of changing the welding angle from a dragging to a pushing angle?
17. What are the advantages of adding oxygen or CO_2 to argon for welds on steel?
18. What are the advantages of using CO_2 for making GMA welds on steel?
19. What is mill scale?
20. What type of porosity is most often caused by mill scale?
21. What should the welder watch if the view of the weld is obstructed by the shielding gas nozzle?
22. When you are making a vertical weld and it appears that the weld metal is going to drip over the shelf, what should you do?
23. What are the advantages of making vertical down welds?
24. How can small weld beads be maintained during overhead welds?
25. How can spatter be controlled on the nozzle when making overhead welds?
26. How should the electrode be manipulated for the deepest penetration when using the pulsed-arc metal transfer process?

This page intentionally left blank

CHAPTER

3 Flux Cored Arc Welding Equipment, Setup, and Operation

OBJECTIVES

After completing this chapter, the student should be able to

- describe the flux cored arc (FCA) welding process
- list the equipment required for an FCA welding workstation
- list five advantages of FCA welding, and explain four of its limitations
- tell how electrodes are manufactured and explain the purpose of the electrode cast and helix
- list four things flux can provide to the weld and how fluxes are classified
- explain what each of the digits in a standard FCAW electrode identification number mean
- describe the proper care and handling of FCAW electrodes
- list two common shielding gases used in FCAW, and contrast their benefits related to cost, productivity, and quality
- list three differences in an FCA weld when the gun angle is changed
- identify the two modes of metal transfer and contrast them in regard to application and quality
- list four effects that electrode extension has on FCA welding
- list three things that can cause weld porosity and how it can be prevented

KEY TERMS

air-cooled
coils
deoxidizers
dual shield
flux cored arc welding (FCAW)
lime-based flux
rutile-based flux
self-shielding
slag
smoke extraction nozzles
spools
water-cooled

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter:

Module 6: Flux Cored Arc Welding (FCAW-G, FCAW-S)

Key Indicator 1: Performs safety inspections of FCAW-G/GM, FCAW-S equipment and accessories

Key Indicator 2: Makes minor external repairs to FCAW-G/GM, FCAW-S equipment and accessories

Key Indicator 3: Sets up for FCAW-G/GM, FCAW-S operations on carbon steel

INTRODUCTION

Flux cored arc welding (FCAW) is a fusion welding process in which weld heating is produced from an arc between the work and a continuously fed filler metal electrode. Atmospheric shielding is provided completely or in part by the flux sealed within the tubular electrode, Figure 3.1. Extra shielding may or may not be supplied through a nozzle in the same way as in GMAW.

Although the process was introduced in the early 1950s, it represented less than 5% of the total amount of welding done in 1965. In 2005, it passed the 50% mark and is still rising. The rapid rise in the use of FCAW has been due to a number of factors. Improvements in the fluxes, smaller electrode diameters, increased reliability of the equipment, better electrode feed systems, and improved guns have all led to the increased usage. Guns equipped with **smoke extraction nozzles** and electronic controls are the latest in a long line of improvements to this process, Figure 3.2.

PRINCIPLES OF OPERATION

FCA welding is similar in a number of ways to the operation of GMA welding, Figure 3.3. Both processes use a constant-potential (CP) or constant-voltage (CV) power supply. Constant potential and voltage are terms that have the same meaning. CP power supplies provide a controlled voltage (potential) to the welding electrode. The amperage (current) varies with the speed that the electrode is being fed into the molten weld pool. Just as in GMA welding, higher electrode feed speeds produce higher currents and slower feed speeds result in lower currents, assuming all other conditions remain constant.

The effects on the weld of electrode extension, gun angle, welding direction, travel speed, and other welder manipulations are similar to those experienced in GMA welding. As in GMA welding, having a correctly set welder does not ensure a good weld. The skill of the welder is an important factor in producing high-quality welds.

The flux inside the electrode protects the molten weld pool from the atmosphere, improves strength through chemical reactions and alloys, and improves the weld shape.

Atmospheric contamination of molten weld metal occurs as it travels across the arc gap and within the pool before it solidifies. The major atmospheric contaminations come from oxygen and nitrogen, the major elements in air. The addition of fluxing and gas-forming elements to the core electrode reduces or eliminates their effects.

Improved strength and other physical or corrosion-resistant properties of the finished weld are improved by the flux. Small additions of alloying elements, deoxidizers, and gas-forming and slag agents all can improve the desired weld properties. Carbon, chromium, and vanadium can be added to improve hardness, strength, creep resistance, and corrosion resistance. Aluminum, silicon, and titanium all help remove oxides and/or nitrides in the weld. Potassium, sodium, and zirconium are added to the flux and form a slag.

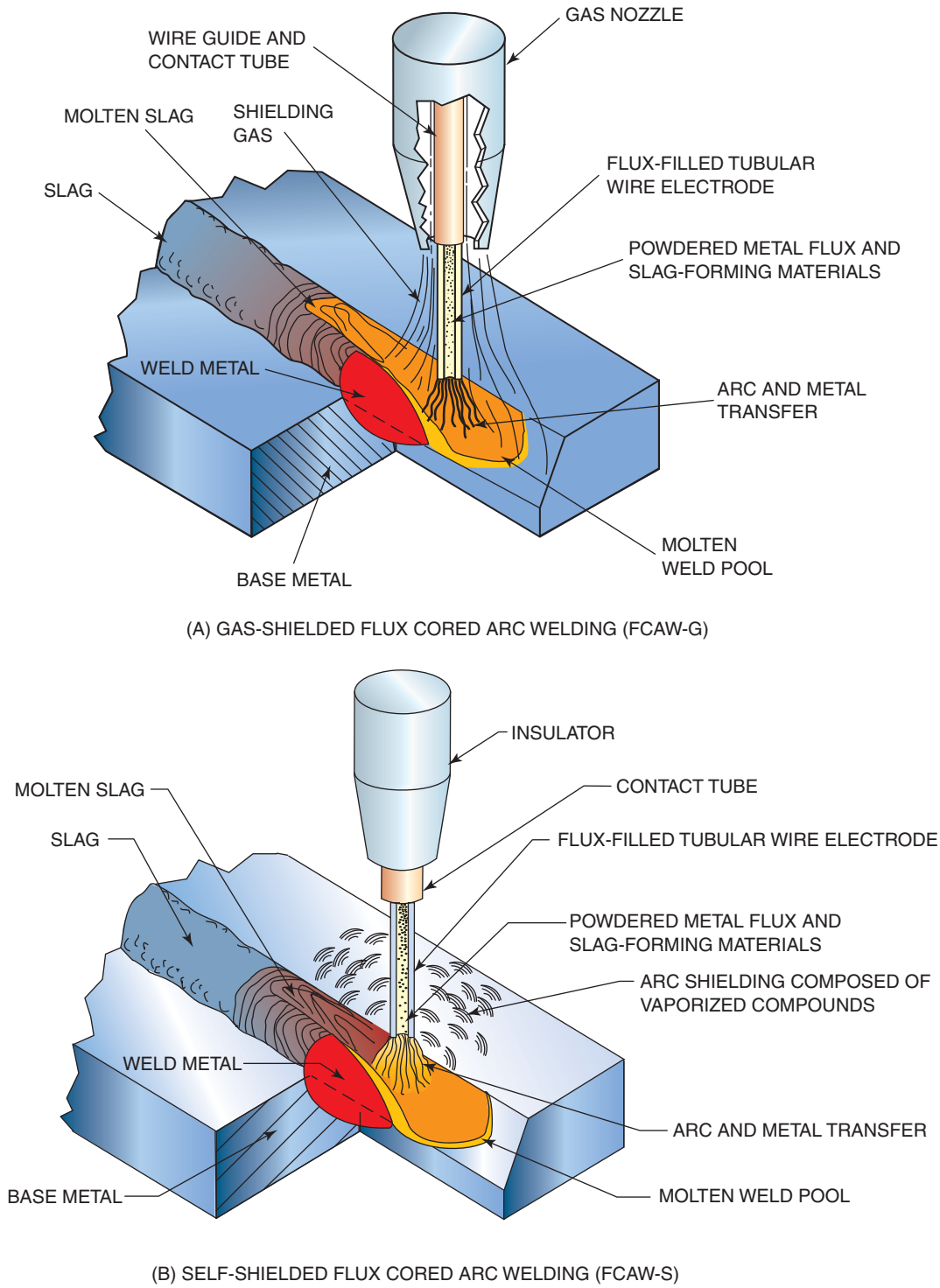

Figure 3.1 Two types of flux cored arc welding
FCA welding may have extra shielding provided by a gas nozzle (A), or be self-shielding only (B).
Source: Courtesy of the American Welding Society

A discussion of weld metal additives and flux elements and their effects on the weld can be found later in this chapter.

The flux core additives that serve as deoxidizers, gas formers, and slag formers either protect the molten weld pool or help to remove impurities from the base metal. Deoxidizers may convert small amounts of surface

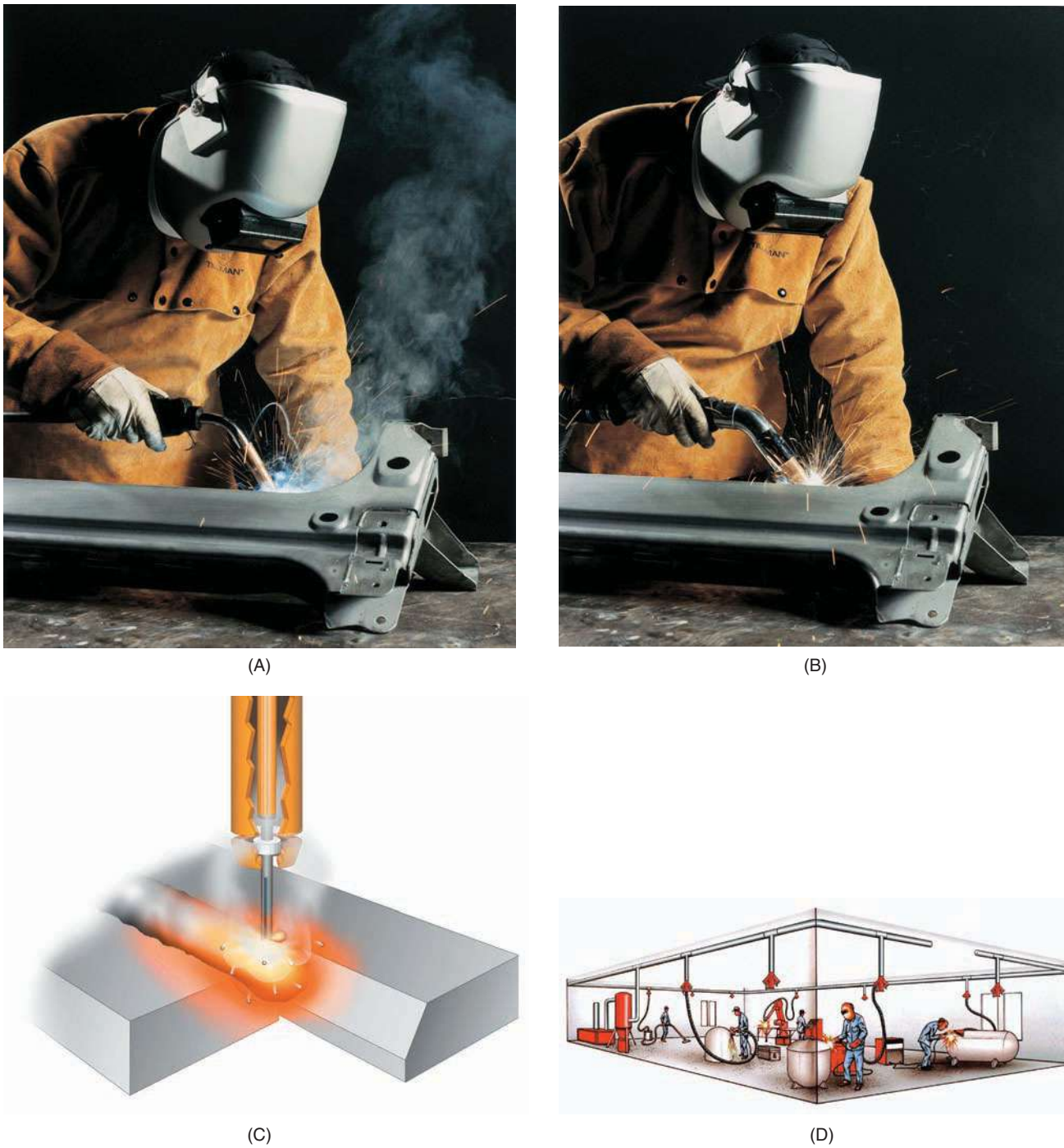

Figure 3.2 Smoke extraction
(A) FCA welding without smoke extraction and (B) with smoke extraction. (C) Typical FCAW smoke extraction gun. (D) Typical smoke exhaust system.
Source: Courtesy of Lincoln Electric Company

oxides like mill scale back into pure metal. They work much like the elements used to refine iron ore into steel.

Gas formers rapidly expand and push the surrounding air away from the molten weld pool. If oxygen in the air were to come in contact with the molten weld metal, the weld metal would quickly oxidize. Sometimes this can be seen at the end of a weld when the molten weld metal erupts in a shower of tiny sparks.

Figure 3.3 Large-capacity wire-feed unit used with FCAW or GMAW
Source: Courtesy of Lincoln Electric Company

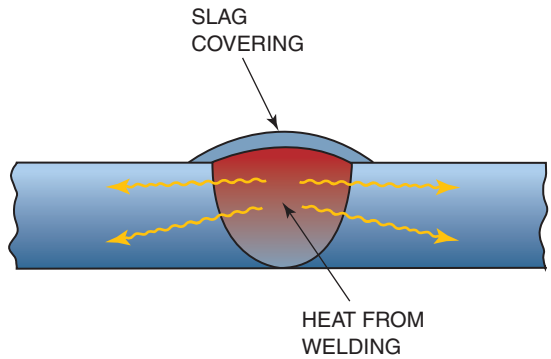

Figure 3.4 Slag blanketing the weld
The slag covering keeps the welding heat from escaping quickly, thus slowing the cooling rate.

The slag covering of the weld is useful for several reasons. **Slag** is a nonmetallic product resulting from the mutual dissolution of the flux and nonmetallic impurities in the base metal. Slag helps the weld by protecting the hot metal from the effects of the atmosphere, controlling the bead shape by serving as a dam or mold, and serving as a blanket to slow the weld's cooling rate, which improves its physical properties, Figure 3.4.

EQUIPMENT

Module 6
Key Indicator 1, 2, 3

Power Supply

The FCA welding power supply is the same type that is required for GMAW, called constant-potential, constant-voltage (CP, CV). The words *potential* and *voltage* have the same electrical meaning and are used interchangeably. FCAW machines can be much more powerful than GMAW machines and are available with up to 1500 amperes of welding power.

Guns

FCA welding guns are available as **water-cooled** or **air-cooled**, Figure 3.5. Although most of the FCA welding guns that you will find in schools are air-cooled, our industry often needs water-cooled guns because of the

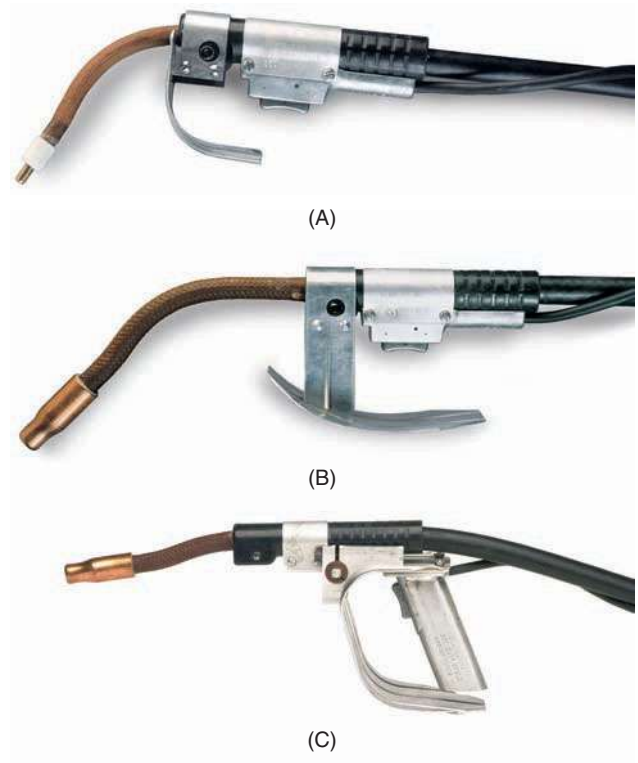

Figure 3.5 Typical FCA welding guns
(A) 350 ampere rating self-shielding, (B) 450 ampere rating gas-shielding, and (C) 600 ampere rating gas-shielding.
Source: Courtesy of Lincoln Electric Company

higher heat caused by longer welds made at higher currents. The water-cooled FCA welding gun is more efficient than an air-cooled gun at removing waste heat. The air-cooled gun is more portable because it has fewer hoses, and it may be made lighter so it is easier to manipulate than the water-cooled gun.

Also, the water-cooled gun requires a water reservoir or another system to give the needed cooling. There are two major ways that water can be supplied to the gun for cooling. Cooling water can be supplied directly from the building's water system, or it can be supplied from a recirculation system.

Cooling water supplied directly from the building's water system is usually dumped into a wastewater drain once it has passed through the gun. When this type of system is used, a pressure regulator must be installed to prevent pressures that are too high from damaging the hoses. Water pressures higher than 35 psi (241 kg/mm^2) may cause the water hoses to burst. Check valves must also be installed in the supply line to prevent contaminated water from being drawn back into the water supply. Some cities and states have laws that restrict the use of open systems because of the need for water conservation. Check with your city or state for any restrictions before installing an open water-cooling system.

Recirculating cooling water systems eliminate any of the problems associated with open systems. Chemicals may be added to the water in recirculating systems to prevent freezing, to aid in pump lubrication, and

to prevent algae growth. Only manufacturer-approved additives should be used in a recirculation system. Read all of the manufacturer's safety and data sheets before using these chemicals.

Smoke Extraction Nozzles

Because of the large quantity of smoke that can be generated during FCA welding, systems for smoke extraction that fit on the gun have been designed, Figure 3.2B. These systems use a vacuum to pull the smoke back into a specially designed smoke extraction nozzle on the welding gun. The disadvantage of this slightly heavier gun is offset by the system's advantages. The advantages of the system are as follows:

- Cleaner air for the welder to breathe because the smoke is removed before it rises to the welder's face.
- Reduced heating and cooling cost because the smoke is concentrated, so less shop air must be removed with the smoke.

Electrode Feed

Electrode feed systems are similar to those used for GMAW; in fact many feed systems are designed with dual feeders so that solid wire and flux core may be run in sequence. The major difference is that the more robust FCAW feeders are designed to use large-diameter wire and most often have two sets of feed rollers. The two sets of rollers help reduce the drive pressure on the electrode. Excessive pressure can distort the electrode wire diameter, which can allow some flux to be dropped inside the electrode guide tube.

ADVANTAGES

FCA welding offers the welding industry a number of important advantages.

High Deposition Rate

High rates of depositing weld metal are possible. FCA welding deposition rates of more than 25 lb/hr (11 kg/hr) of weld metal are possible. This compares to about 10 lb/hr (5 kg/hr) for shielded metal arc (SMA) welding using a very large-diameter electrode of 1/4 in. (6 mm).

Minimum Electrode Waste

The FCA method makes efficient use of filler metal; from 75% to 90% of the weight of the FCAW electrode is metal, the remainder being flux. SMAW electrodes have a maximum of 75% filler metal; some SMAW electrodes have much less. Also, a stub must be left at the end of each SMA welding electrode. The stub will average 2 in. (51 mm) in length, resulting in a loss of 11% or more of the SMAW filler electrode purchased. FCA welding has no stub loss, so nearly 100% of the FCAW electrode purchased is used.

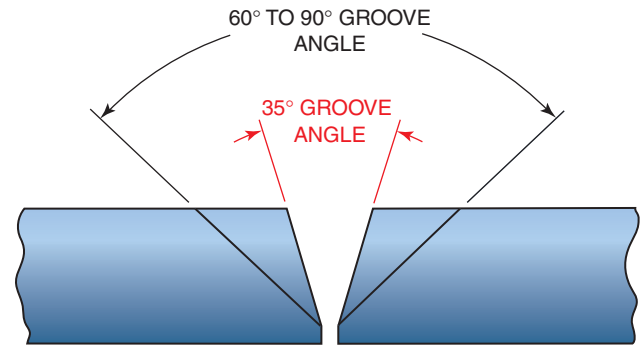

Figure 3.6 Narrower groove angle for FCAW
The narrower groove angle for FCAW compared to other welding processes saves on filler metal, welding time, and heat input into the part.

Narrow Groove Angle

Because of the deep penetration characteristic of FCAW, no edge-beveling preparation is required on some joints in metal up to 1/2 in. (13 mm) in thickness. When bevels are cut, the joint-included angle can be reduced to as small as 35°, Figure 3.6. The reduced groove angle results in a smaller-sized weld. This can save 50% of filler metal with about the same savings in time and weld power used.

Minimum Precleaning

The addition of **deoxidizers**, which combine with and remove harmful oxides on the base metal or its surface, and other fluxing agents permits high-quality welds to be made on plates with light surface oxides and mill scale. This eliminates most of the precleaning required before GMA welding can be performed. Often it is possible to make excellent welds on plates in the "as cut" condition; no cleanup is needed.

All-position Welding

Small-diameter electrode sizes in combination with special fluxes allow excellent welds in all positions. The slags produced assist in supporting the weld metal. This process is easy to use, and, when properly adjusted, it is much easier to use than other all-position arc welding processes.

Flexibility

Changes in power settings can permit welding to be done on thin-gauge sheet metals or thicker plates using the same electrode size. Multipass welds allow joining metals with no limit on thickness. This, too, is attainable with one size of electrode.

High Quality

Many codes permit welds to be made using FCAW. The addition of the flux gives the process the high level of reliability needed for welding on boilers, pressure vessels, and structural steel.

Excellent Control

The molten weld pool is more easily controlled with FCAW than with GMAW. The surface appearance is smooth and uniform even with less operator skill. Visibility is improved by removing the nozzle when using self-shielded electrodes.

LIMITATIONS

The main limitation of flux cored arc welding is that it is confined to ferrous metals and nickel-based alloys. Generally, all low- and medium-carbon steels and some low-alloy steels, cast irons, and a limited number of stainless steels are presently weldable using FCAW.

The equipment and electrodes used for the FCAW process are more expensive. However, the cost is quickly recoverable through higher productivity.

The removal of postweld slag requires another production step. The flux must be removed before the weldment is finished (painted) to prevent crevice corrosion.

With the increased welding output comes an increase in smoke and fume generation. The existing ventilation system in a shop might need to be increased to handle the added volume.

ELECTRODES

Methods of Manufacturing

The electrodes have flux tightly packed inside. One method used to make them is to first form a thin sheet of metal into a U-shape, Figure 3.7. A measured quantity of flux is poured into the U-shape before it is squeezed shut. The wire is then passed through a series of dies to size it and further compact the flux.

A second method of manufacturing the electrode is to start with a seamless tube. The tube is usually about 1 in. in diameter. One end of the tube is sealed, and the flux powder is poured into the open end. The tube is vibrated during the filling process to ensure that it fills completely. Once the tube is full, the open end is sealed. The tube is now sized using a series of dies, Figure 3.8.

In both these methods of manufacturing the electrode, the sheet and tube are made up of the desired alloy. Also in both cases, the flux is compacted inside the metal skin. This compacting helps make the electrode operate more smoothly and consistently.

Electrodes are available in sizes from 0.030 in. to 5/32 in. (0.8 mm to 3.9 mm) in diameter. Smaller-diameter electrodes are much more expensive per pound than the same type in a larger diameter due to the high cost of drawing and filling cored wires to small sizes. Larger-diameter electrodes produce such large welds they cannot be controlled in all positions. The most popular diameters range from 0.035 in. to 3/32 in. (0.9 mm to 2.3 mm).

The finished FCA filler metal is packaged in a number of forms for purchase by the end user, Figure 3.9. The AWS has a standard for the size of

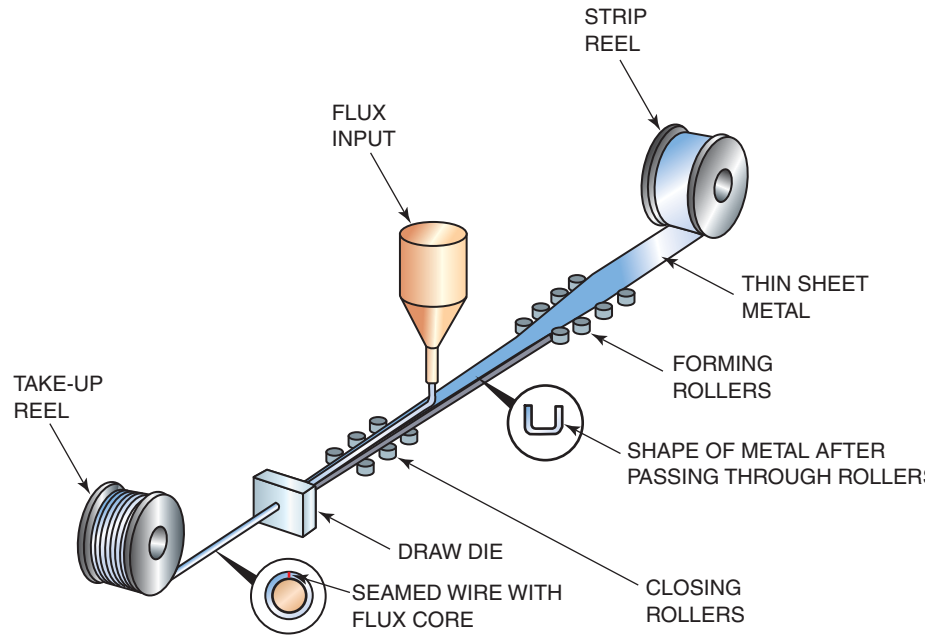

Figure 3.7 **Putting the flux in the flux cored wire**

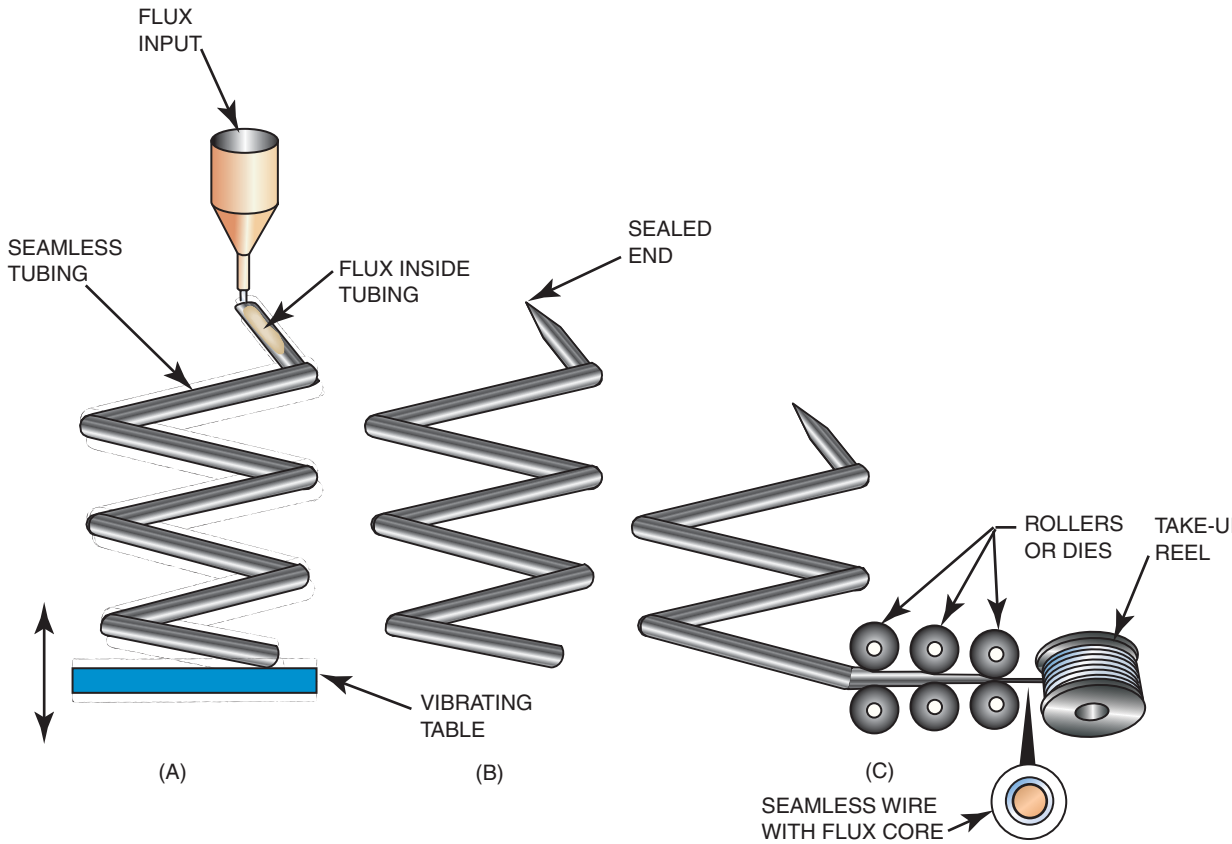

Figure 3.8 **One method of filling seamless FCA welding filler metal with flux**
The vibration helps compact the granular flux inside the tube.

Figure 3.9 A few packaged forms of FCA filler metal
FCAW filler metal weights are approximate. They will vary by alloy and manufacturer.
Source: Courtesy of Lincoln Electric Company

each of the package units. Although the dimensions of the packages are standard, the weight of filler wire is not standard. More of the smaller-diameter wire can fit into the same space compared with a larger-diameter wire, so a package of 0.030-in. (0.8-mm) wire weighs more than the same-sized package of 3/32-in. (2.3-mm) wire. The standard packing units for FCAW wires are **spools**, **coils**, reels, and drums, Table 3.1.

Spools are made of plastic or fiberboard and are disposable. They are completely self-contained and are available in approximate weights from

Table 3.1 Packaging Size Specification for Commonly Used FCA Filler Wire

Packaging	Outside Diameter	Width	Arbor (Hole) Diameter
Spools	4 in. (102 mm)	1-3/4 in. (44.5 mm)	5/8 in. (16 mm)
	8 in. (203 mm)	2-1/4 in. (57 mm)	2-1/16 in. (52.3 mm)
	12 in. (305 mm)	4 in. (102 mm)	2-1/16 in. (52.3 mm)
	14 in. (356 mm)	4 in. (102 mm)	2-1/16 in. (52.3 mm)
Reels	22 in. (559 mm)	12-1/2 in. (318 mm)	1-5/16 in. (33.3 mm)
	30 in. (762 mm)	16 in. (406 mm)	1-5/16 in. (33.3 mm)
Coils	16-1/4 in. (413 mm)	4 in. (102 mm)	12 in. (305 mm)

	Outside Diameter	Inside Diameter	Height
Drums	23 in. (584 mm)	16 in. (406 mm)	34 in. (864 mm)

1 lb up to around 50 lb (0.5 kg to 25 kg). The smaller spools, 4 in. and 8 in. (102-mm and 203 mm), weighing from 1 lb to 7 lb, are most often used for smaller production runs or for home/hobby use; 12-in. and 14-in. (305-mm and 356-mm) spools are often used in schools and welding fabrication shops.

Coils come wrapped and/or wire tied together. They are unmounted, so they must be supported on a frame on the wire feeder in order to be used. Coils are available in weights around 60 lb (27 kg). Because FCAW wires on coils do not have the expense of a disposable core, these wires cost a little less per pound, so they are more desirable for higher-production shops.

Reels are large wooden spools, and drums are shaped like barrels. Both reels and drums are used for high-production jobs. Both can contain approximately 300 lb to 1000 lb (136 kg to 454 kg) of FCAW wire. Because of their size, they are used primarily at fixed welding stations. Such stations are often associated with some form of automation, such as turntables or robotics.

Electrode Cast and Helix

To see the cast and helix of a wire, feed out 10 ft of wire electrode and cut it off. Lay it on the floor and observe that it forms a circle. The diameter of the circle is known as the cast of the wire, Figure 3.10.

Note that the wire electrode does not lay flat. One end is slightly higher than the other. This height is the helix of the wire.

The AWS has specifications for both cast and helix for all FCA welding wires.

The cast and helix cause the wire to rub on the inside of the contact tube, Figure 3.11. The slight bend in the electrode wire ensures a positive electrical contact between the contact tube and filler wire.

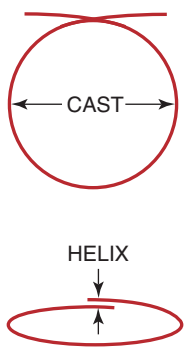

Figure 3.10 Method of measuring cast and helix of FCAW filler wire

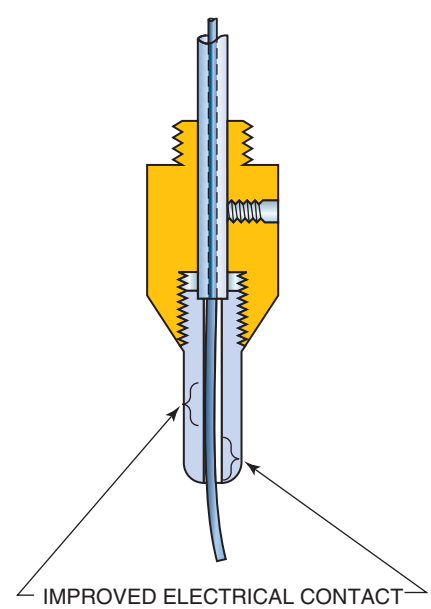

Figure 3.11 Cast forces the wire to make better electrical contact with the tube

FLUX

The fluxes used are mainly based on lime or rutile (a mineral compound consisting of titanium dioxide, usually with a little iron). The purpose of the fluxes is the same as in the shielded metal arc welding (SMAW) process. That is, they can provide all or part of the following to the weld:

- *Deoxidizers:* Oxygen that is present in the welding zone has two forms. It can exist as free oxygen from the atmosphere surrounding the weld. Oxygen can also exist as part of a compound such as an iron oxide or carbon dioxide (CO_2). In either case it can cause porosity in the weld if it is not removed or controlled. Chemicals are added that react to the presence of oxygen in either form and combine to form a harmless compound, Table 3.2. The new compound can become part of the slag that solidifies on top of the weld, or some of it may stay in the weld as very small inclusions. Both methods result in a weld with better mechanical properties because of lower porosity.
- *Slag formers:* Slag serves several vital functions for the weld. It can react with the molten weld metal chemically, and it can affect the weld bead physically. In the molten state it moves through the molten weld pool and acts as a magnet or sponge to chemically combine with impurities in the metal and remove them, Figure 3.12. Slags can be refractory, so that they become solid at a high temperature. As they solidify over the weld, they help it hold its shape and they slow its cooling rate.
- *Fluxing agents:* Molten weld metal tends to have a high surface tension, which prevents it from flowing outward toward the edges of the weld. This causes undercutting along the junction of the

Table 3.2 Deoxidizing Elements Added to Filler Wire (to Minimize Porosity in the Molten Weld Pool)

Deoxidizing Element	Strength
Aluminum (Al)	Very strong
Manganese (Mn)	Weak
Silicon (Si)	Weak
Titanium (Ti)	Very strong
Zirconium (Zr)	Very strong

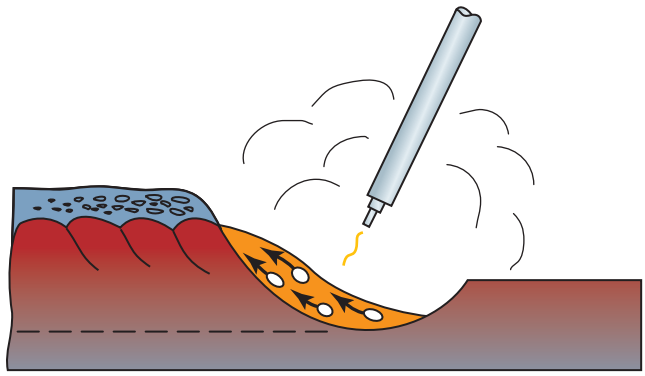

Figure 3.12 Impurities being floated to the surface by slag

weld and the base metal. Fluxing agents make the weld more fluid and allow it to flow outward, filling the undercut.

- *Arc stabilizers:* Chemicals in the flux affect the arc resistance. As the resistance is lowered, the arc voltage drops and penetration is reduced. When the arc resistance is increased, the arc voltage increases and weld penetration is increased. Although the resistance within the ionized arc stream may change, the arc is more stable and easier to control. It also improves the metal transfer by reducing spatter caused by an erratic arc.
- *Alloying elements:* Because of the difference in the mechanical properties of metal that is formed by rolling or forging and metal that is melted to form a weld bead, the metallurgical requirements of the two also differ. Some elements change the weld's strength, ductility, hardness, brittleness, toughness, and corrosion resistance. Other alloying elements in the form of powder metal can be added to increase deposition rates.
- *Shielding gas:* As elements in the flux are heated by the arc, some of them vaporize and form voluminous gaseous clouds hundreds of times larger than their original volume. This rapidly expanding cloud forces the air around the weld zone away from the molten weld metal, Figure 3.13. Without the protection this process affords the molten metal, it would rapidly oxidize. Such oxidization would severely affect the weld's mechanical properties, rendering it unfit for service.

All FCAW fluxes are divided into two groups based on the acid or basic chemical reactivity of the slag. The AWS classifies T-1 as acid and T-5 as basic.

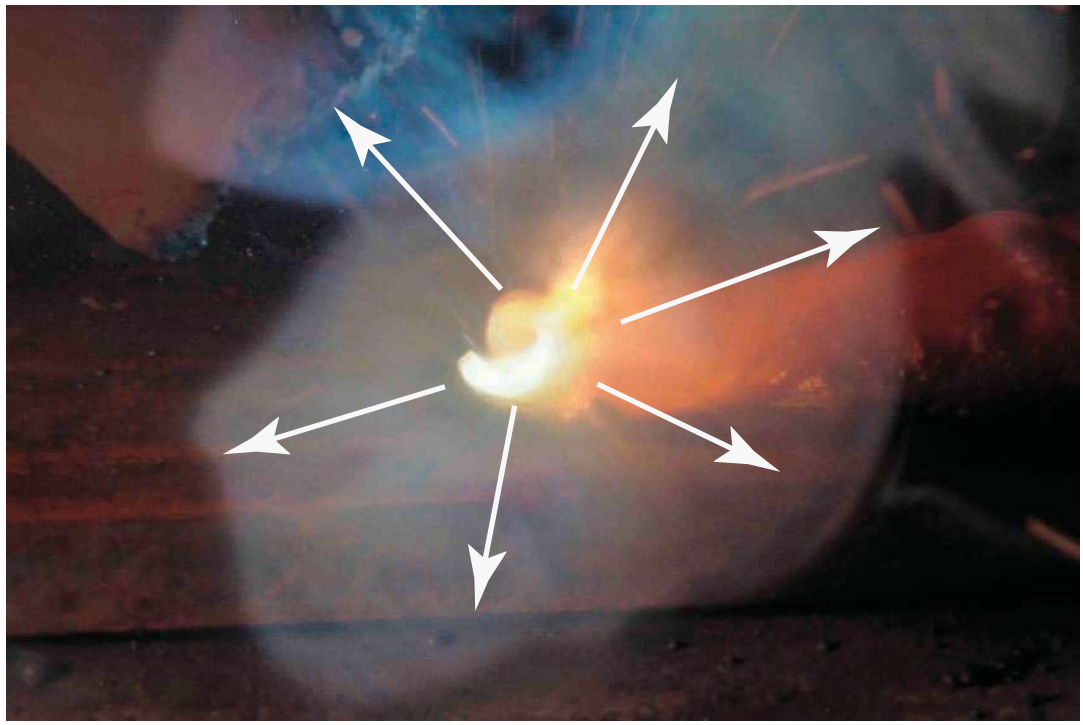

Figure 3.13 Rapidly expanding gas cloud
Source: Courtesy of Larry Jeffus

The **rutile-based flux** is acidic, T-1. It produces a smooth, stable arc and a refractory high-temperature slag for out-of-position welding. These electrodes produce a fine drop transfer, a relatively low fume, and an easily removed slag. The main limitation of the rutile fluxes is that their fluxing elements do not produce as high a quality deposit as do the T-5 systems.

The **lime-based flux** is basic, T-5. It is very good at removing certain impurities from the weld metal, but its low-melting-temperature slag is fluid, which makes it generally unsuitable for out-of-position welding. These electrodes produce a more globular transfer, more spatter, more fume, and a more adherent slag than do the T-1 systems. These characteristics are tolerated when it is necessary to deposit very tough weld metal and for welding materials having a low tolerance for hydrogen.

Some rutile-based electrodes allow the addition of a shielding gas. With the weld partially protected by the shielding gas, more elements can be added to the flux, which produces welds with the best of both flux systems, high-quality welds in all positions.

Some fluxes can be used on both single- and multiple-pass welds, and others are limited to single-pass welds only. Using a single-pass welding electrode for multipass welds may result in an excessive amount of manganese. The manganese is necessary to retain strength when making large, single-pass welds. However, with the lower dilution associated with multipass techniques, it can strengthen the weld metal too much and reduce its ductility. In some cases, small welds that deeply penetrate the base metal can help control this problem.

Table 3.3 lists the shielding and polarity for the flux classifications of mild steel FCAW electrodes. The letter *G* is used to indicate an unspecified classification. The *G* means that the electrode has not been classified by the American Welding Society. Often the exact composition of fluxes is kept as a manufacturer's trade secret. Therefore, only limited information about the electrode's composition will be given. The only information often supplied is current, type of shielding required, and some strength characteristics.

Table 3.3 Welding Characteristics of Seven Flux Classifications

Classification	Comments	Shielding Gas
T-1	Requires clean surfaces and produces little spatter. It can be used for single- and multiple-pass welds in all positions.	Carbon dioxide (CO_2) or argon/carbon dioxide mixes
T-2	Requires clean surfaces and produces little spatter. It can be used for single-pass welds in the flat (1G and 1F) and horizontal (2F) positions only.	Carbon dioxide (CO_2)
T-3	Used on thin-gauge steel for single-pass welds in the flat (1G and 1F) and horizontal (2F) positions only.	None
T-4	Low penetration and moderate tendency to crack for single- and multiple-pass welds in the flat (1G and 1F) and horizontal (2F) positions.	None
T-5	Low penetration and a thin, easily removed slag, used for single- and multiple-pass welds in the flat (1G and 1F) position only.	With or without carbon dioxide (CO_2)
T-6	Similar to T-5 without externally applied shielding gas.	None
T-G	The composition and classification of this electrode are not given in the preceding classes. It may be used for single- or multiple-pass welds.	With or without shielding

Table 3.4 Ferrite-forming Elements Used in FCA Welding Fluxes

Element	Reaction in Weld
Silicon (Si)	Ferrite former and deoxidizer
Chromium (Cr)	Ferrite and carbide former
Molybdenum (Mo)	Ferrite and carbide former
Columbium (Cb)	Strong ferrite former
Aluminum (Al)	Ferrite former and deoxidizer

As a result of the relatively rapid cooling of the weld metal, the weld may tend to become hard and brittle. This factor can be controlled by adding elements to the flux that affect the content of both the weld and the slag, Table 3.4. Ferrite is the softer, more ductile form of iron. The addition of ferrite-forming elements can control the hardness and brittleness of a weld. Refractory fluxes are sometimes called "fast-freeze" because they solidify at a higher temperature than the weld metal. By becoming solid first, this slag can cradle the molten weld pool and control its shape. This property is very important for out-of-position welds.

The impurities in the weld pool can be metallic or nonmetallic compounds. Metallic elements that are added to the metal during the manufacturing process in small quantities may be concentrated in the weld. These elements improve the grain structure, strength, hardness, resistance to corrosion, or other mechanical properties in the metal's as-rolled or formed state. But the deposited weld metal, or weld nugget, is like a small casting because the liquid weld metal freezes in a controlled shape, and some alloys adversely affect the properties of this casting (weld metal). Nonmetallic compounds are primarily slag inclusions left in the metal from the fluxes used during manufacturing. The welding fluxes form slags that are less dense than the weld metal so that they will float to the surface before the weld solidifies.

Flux Cored Steel Electrode Identification

Module 6
Key Indicator 3

The American Welding Society revised its A5.20 *Specification for Carbon Steel Electrodes for Flux Cored Arc Welding* in 1995 to reflect changes in the composition of the FCA filler metals. Table 3.5 lists the AWS specifications for flux cored filler metals.

Mild Steel

The electrode number *E70T-10* is used as an example to explain the classification system for mild steel FCAW electrodes (Figure 3.14):

- *E*—Electrode.
- *7*—Tensile strength in units of 10,000 psi for a good weld. This value is usually either *6* for 60,000 or *7* for 70,000 psi minimum

Table 3.5 Filler Metal Classification Numbers

Metal	AWS Filler Metal Classification
Mild steel	A5.20
Stainless steel	A5.22
Chromium–molybdenum	A5.29

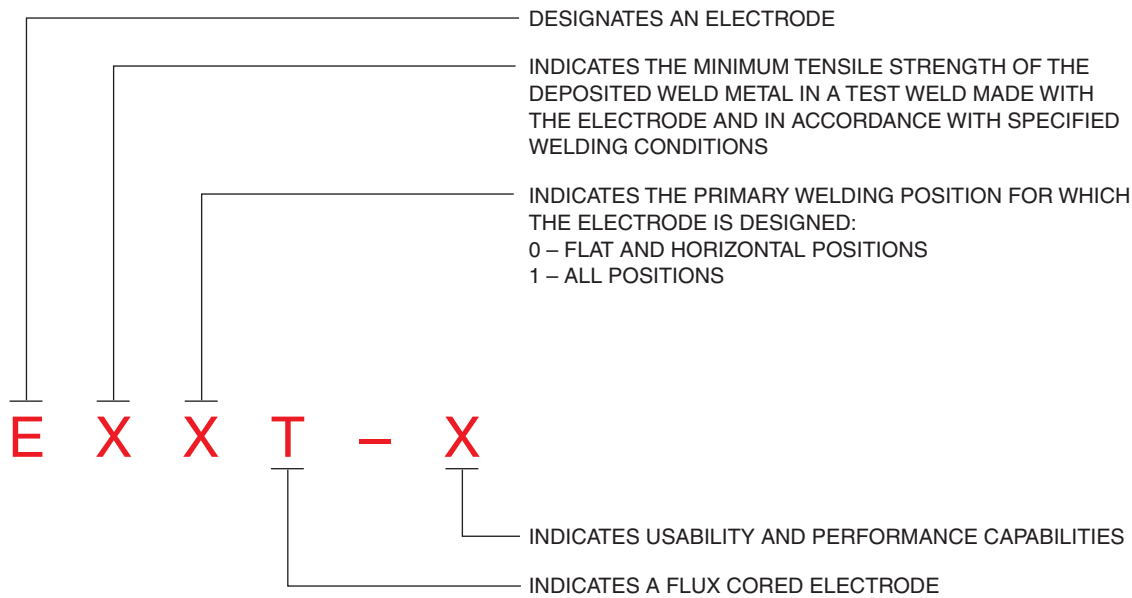

Figure 3.14 Identification system for mild steel FCAW electrodes
Source: Courtesy of the American Welding Society

weld strength. An exception is for the number *12*, which is used to denote filler metals having a range from 70,000 to 90,000 psi.

- *0—0* is used for flat and horizontal fillets only, and *1* is used for all-position electrodes.
- *T*—Tubular (flux cored) electrode.
- *10*—The number in this position can range from *1* to *14* and is used to indicate the electrode's shielding gas if any, number of passes that may be applied one on top of the other, and other welding characteristics of the electrode. The letter *G* is used to indicate that the shielding gas, polarity, and impact properties are not specified. The letter *G* may or may not be followed by the letter *S*. *S* indicates an electrode suitable only for single-pass welding.

The electrode classification *E70T-10* can have some optional identifiers added to the end of the number, as in *E70T-10MJH8*. These additions are used to add qualifiers to the general classification so that specific codes or standards can be met. These additions have the following meanings:

- *M*—Mixed gas of 75% to 80% Ar and CO_2 for the balance. If there is no *M*, either the shielding gas is CO_2 or the electrode is self-shielded.
- *J*—Describes the Charpy V-notch impact test value of 20 ft-lb at 40°F.
- *H8*—Describes the residual hydrogen levels in the weld: *H4* equals less than 4 ml/100 g; *H8*, less than 8 ml/100 g; *H16*, less than 16 ml/100 g.

Stainless Steel Electrodes

The AWS classification for stainless steel for FCAW electrodes starts with the letter *E* as its prefix. Following the *E* prefix, the American Iron and Steel Institute's (AISI) three-digit stainless steel number is used. This number indicates the type of stainless steel in the filler metal.

To the right of the AISI number, the AWS adds a dash followed by a suffix number. The number 1 is used to indicate an all-position filler metal, and the number 3 is used to indicate an electrode to be used in the flat and horizontal positions only.

Metal Cored Steel Electrode Identification

The addition of metal powders to the flux core of FCA welding electrodes has produced a new classification of filler metals. The new filler metals evolved over time, and a new identification system was established by the AWS to identify these filler metals. Some of the earlier flux cored filler metals that already had powder metals in their core had their numbers changed to reflect the new designation. The designation was changed from the letter *T* for *tubular* to the letter *C* for *core*. For example, E70T-1 became E70C-3C. The complete explanation of the cored electrode *E70C-3C* follows:

- *E*—Electrode.
- 7—Tensile strength in units of 10,000 psi for a good weld. This value is usually either *6* for 60,000 or *7* for 70,000 psi minimum weld strength. An exception is for the number *12,* which is used to denote filler metals having a range from 70,000 to 90,000 psi.
- *0—0* is used for flat and horizontal fillets only, and *1* is used for all-position electrodes.
- *C*—Metal-cored (tubular) electrode.
- *3—3* is used for a Charpy impact of 20 ft-lb at 0°F, and *6* represents a Charpy impact of 20 ft-lb at 20°F.
- *C*—The second letter *C* indicates CO_2. The letter *M* in this position would indicate a mixed gas, 75% to 80% Ar, with the balance being CO_2. If there is no *M* or *C,* then the shielding gas is CO_2. The letter *G* is used to indicate that the shielding gas, polarity, and impact properties are not specified. The letter *G* may or may not be followed by the letter *S*. *S* indicates an electrode suitable only for single-pass welding.

NOTE

The powdered metal added to the core flux can provide additional filler metal and/or alloys. This is one way the micro-alloys can be added in very small and controlled amounts, as low as 0.0005% to 0.005%. These are very powerful alloys that dramatically improve the metal's mechanical properties.

Care of Flux Core Electrodes

Wire electrodes may be wrapped in sealed plastic bags for protection from the elements. Others may be wrapped in a special paper, and some are shipped in cans or cardboard boxes.

A small paper bag of a moisture-absorbing material, crystal desiccant, is sometimes placed in the shipping containers to protect wire electrodes from moisture. Some wire electrodes require storage in an electric rod oven to prevent contamination from excessive moisture. Read the manufacturer's recommendations located in or on the electrode shipping container for information on use and storage.

Weather conditions affect your ability to make high-quality welds. Humidity increases the chance of moisture entering the weld zone. Water (H_2O), which consists of two parts hydrogen and one part oxygen, separates in the weld pool. When only one part of hydrogen is expelled, hydrogen entrapment occurs. Hydrogen entrapment can cause weld beads to crack or become brittle. The evaporating moisture will also cause porosity.

To prevent hydrogen entrapment, porosity, and atmospheric contamination, it may be necessary to preheat the base metal to drive out moisture. Storing the wire electrode in a dry location is recommended. The electrode may develop restrictions due to the tangling of the wire or become oxidized with excessive rusting if the wire electrode package is mishandled, thrown, dropped, or stored in a damp location.

Caution

Always keep the wire electrode dry and handle it as you would any important tool or piece of equipment.

SHIELDING GAS

FCA welding wire can be manufactured so that all of the required shielding of the molten weld pool is provided by the vaporization of some of the flux within the tubular electrode. When the electrode provides all of the shielding, it is called **self-shielding** and the welding process is abbreviated FCAW-S (S for self-shielding). Other FCA welding wire must use an externally supplied shielding gas to provide the needed protection of the molten weld pool. When a shielding gas is added, the combined shielding is called **dual shield** and the process is abbreviated FCAW-G (G for gas).

Note: Sometimes the shielding gas(es) are referred to as the shielding *medium*. For example, the shielding gas, or medium, for E71T-5 is either 75% argon with 25% CO_2 or 100% CO_2.

Care must be taken to use the cored electrodes with the recommended gases, and not to use gas at all with the self-shielded electrodes. Using a self-shielding flux cored electrode with a shielding gas may produce a defective weld. The shielding gas will prevent the proper disintegration of much of the deoxidizers. This results in the transfer of these materials across the arc to the weld. In high concentrations, the deoxidizers can produce slags that become trapped in the welds, causing undesirable defects. Lower concentrations may cause brittleness only. In either case, the chance of weld failure is increased. If these electrodes are used correctly, there is no problem.

The selection of a shielding gas will affect the arc and weld properties. The weld bead width, buildup, penetration, spatter, chemical composition, and mechanical properties are all affected as a result of the shielding gas selection.

Shielding gas comes in high-pressure cylinders. These cylinders are supplied with 2000 psi of pressure. Because of this high pressure, it is important that the cylinders be handled and stored safely. For specific cylinder safety instructions see Chapter 2 in Introduction to Welding, the first book in this series.

Gases used for FCA welding include CO_2 and mixtures of argon and CO_2. Argon gas is easily ionized by the arc. Ionization results in a highly concentrated path from the electrode to the weld. This concentration results in a smaller droplet size that is associated with the axial spray mode of metal transfer, Figure 3.15. A smooth, stable arc results and there is a minimum of spatter. This transfer mode continues as CO_2 is added to the argon until the mixture contains more than 25% of CO_2.

As the percentage of CO_2 increases in the argon mixture, weld penetration increases. This increase in penetration continues until a 100% CO_2 shielding gas is reached. But as the percentage of CO_2 is increased the arc stability decreases. The less stable arc causes an increase in spatter.

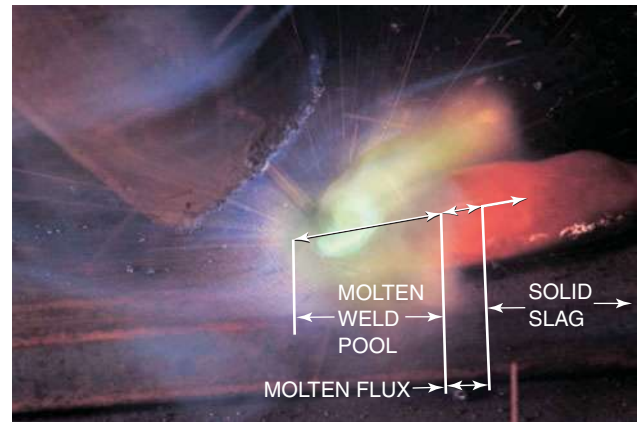

Figure 3.15 **Axial spray transfer mode**
Source: Courtesy of Larry Jeffus

A mixture of 75% argon and 25% CO_2 works best for jobs requiring a mixed gas. This mixture is sometimes called C-25.

Straight CO_2 is used for some welding. But the CO_2 gas molecule is easily broken down in the welding arc. It forms carbon monoxide (CO) and free oxygen (O). Both gases are reactive to some alloys in the electrode. As these alloys travel from the electrode to the molten weld pool, some of them form oxides. Silicon and manganese are the primary alloys that become oxidized and lost from the weld metal.

Most FCA welding electrodes are specifically designed to be used with or without shielding gas and for a specific shielding gas or percentage mixture. For example, an electrode designed specifically for use with 100% CO_2 will have higher levels of silicon and manganese to compensate for the losses to oxidization. But if 100% argon or a mixture of argon and CO_2 is used, the weld will have an excessive amount of silicon and manganese. The weld will not have the desired mechanical or metallurgical properties. Although the weld may look satisfactory, it will probably fail prematurely.

Caution

Never use an FCA welding electrode with a shielding gas it is not designated to be used with. The weld it produces may be unsafe.

WELDING TECHNIQUES

A welder can control weld beads made by FCA welding by making changes in the techniques used. The following explains how changing specific welding techniques will affect the weld produced.

Gun Angle

Gun angle, work angle, and *travel angle* are terms used to refer to the relation of the gun to the work surface, Figure 3.16. The gun angle can be used to control the weld pool. The electric arc produces an electrical force known as the arc force. The arc force can be used to counteract the gravitational pull that tends to make the liquid weld pool sag or run ahead of the arc. By manipulating the electrode travel angle for the flat and horizontal position of welding to a 20° to 45° angle from the vertical, the weld pool can be controlled. A 40° to 50° angle from the vertical plate is recommended for fillet welds.

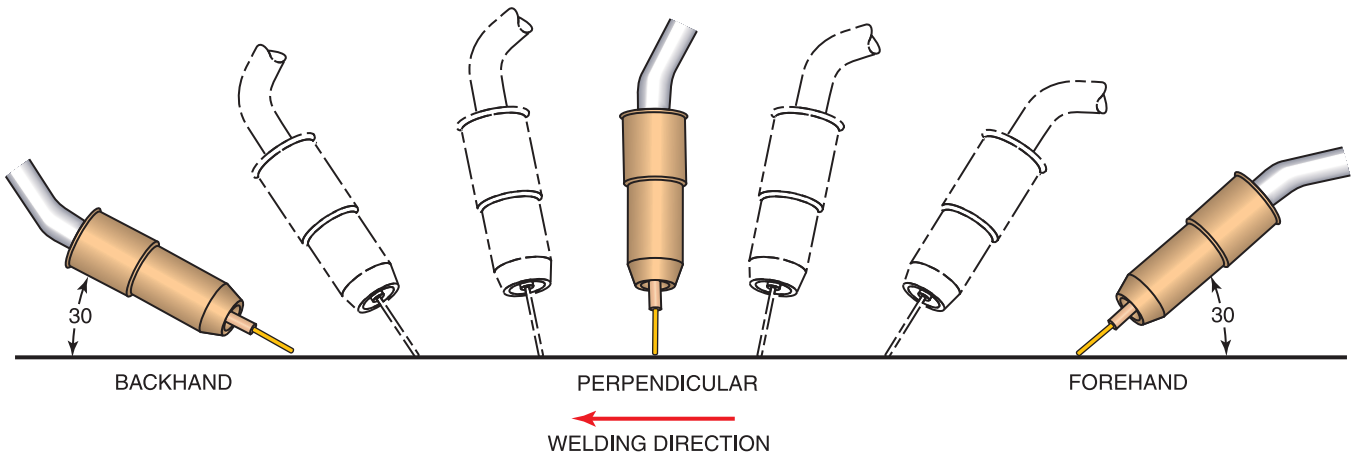

Figure 3.16 Welding gun angles

Changes in this angle will affect the weld bead shape and penetration. Shallower angles are needed when welding thinner materials to prevent burn-through. Steeper, perpendicular angles are used for thicker materials.

FCAW electrodes have a flux that is mineral based, often called low-hydrogen. These fluxes are refractory and become solid at a high temperature. If too steep a forehand, or pushing, angle is used, slag from the electrode can be pushed ahead of the weld bead and solidify quickly on the cooler plate, Figure 3.17. Because the slag remains solid at higher temperatures than the temperature of the molten weld pool, it can be trapped under the edges of the weld by the molten weld metal. To avoid this problem, most flat and horizontal welds should be performed with a backhand angle.

Vertical up welds require a forehand gun angle. The forehand angle is needed to direct the arc deep into the groove or joint for better control of the weld pool and deeper penetration, Figure 3.18. Slag entrapment associated with most forehand welding is not a problem for vertical welds.

A gun angle around 90° to the metal surface either slightly forehand or backhand works best for overhead welds, Figure 3.19. The slight angle aids with visibility of the weld, and it helps control spatter buildup in the gas nozzle.

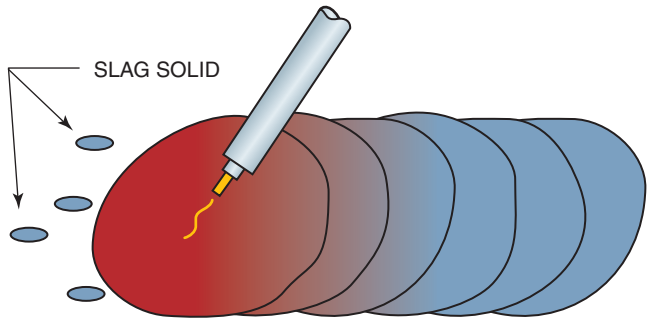

Figure 3.17 Problem of trapped slag
Large quantities of solid slag in front of a weld can cause slag to be trapped under the weld bead.

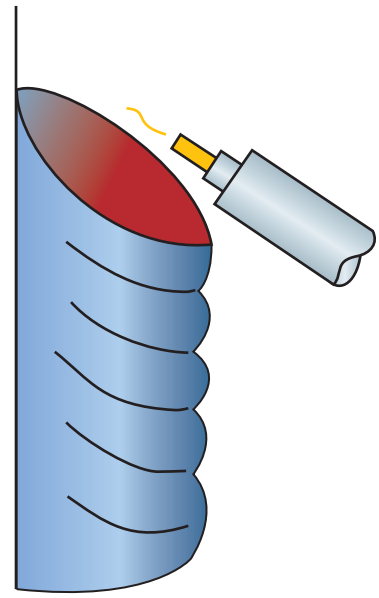

Figure 3.18 Vertical up gun angle

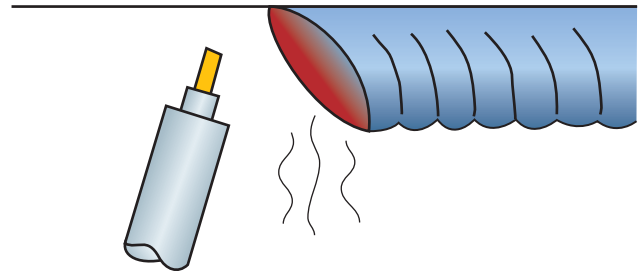

Figure 3.19 Weld gun position to control spatter buildup on an overhead weld

Forehand/Perpendicular/Backhand Techniques

Forehand, perpendicular, and *backhand* are the terms most often used to describe the gun angle as it relates to the work and the direction of travel. The forehand technique is sometimes referred to as *pushing* the weld bead, and backhand may be referred to as *pulling* or *dragging* the weld bead. The term *perpendicular* is used when the gun angle is at approximately 90° to the work surface, Figure 3.20.

Advantages of the Forehand Technique

The forehand welding technique has several advantages:

- Joint visibility—You can easily see the joint where the bead will be deposited, Figure 3.21.
- Electrode extension—The contact tube tip is easier to see, making it easier to maintain a constant extension length.
- Less weld penetration—It is easier to weld on thin sheet metal without melting through.
- Out-of-position welds—This technique works well on vertical up and overhead joints for better control of the weld pool.

Disadvantages of the Forehand Technique

The disadvantages of using the forehand welding technique are the following:

- Weld thickness—Thinner welds may occur because less weld reinforcement is applied to the weld joint.

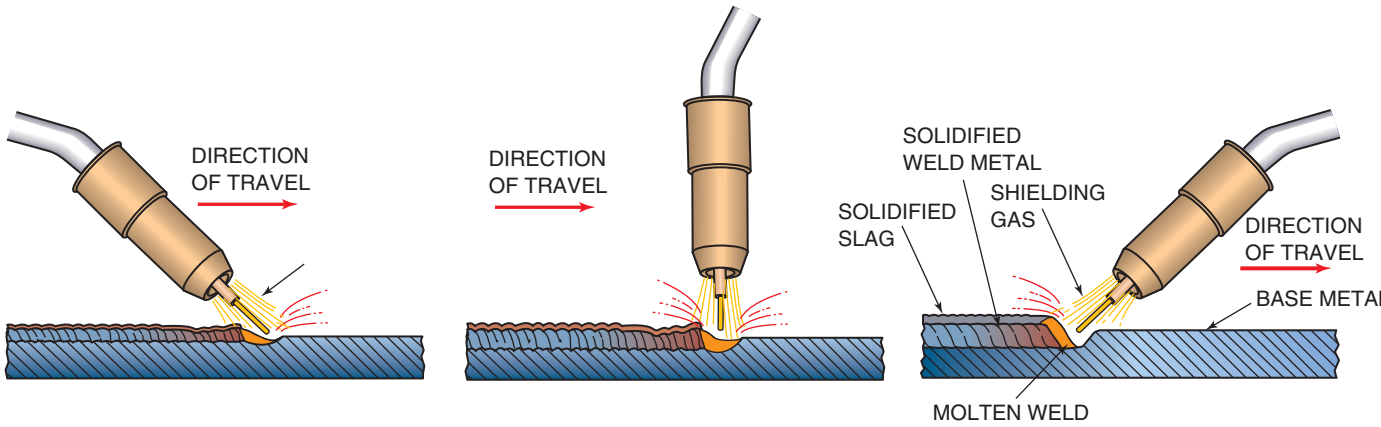

Figure 3.20 Gun angles
Changing the welding gun angle between forehand, perpendicular, and backhand angles will change the shape of the weld bead produced.

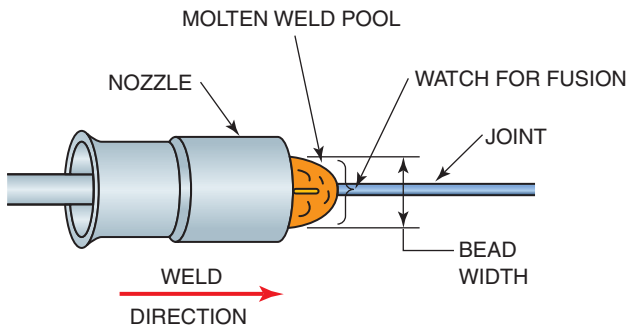

Figure 3.21 Welder's view with a forehand angle
This angle keeps the shielding gas nozzle from restricting the welder's view.

- Welding speed—Because less weld metal is being applied, the rate of travel along the joint can be faster, which may make it harder to create a uniform weld.
- Slag inclusions—Some spattered slag can be thrown in front of the weld bead and be trapped or included in the weld, resulting in a weld defect.
- Spatter—Depending on the electrode, the amount of spatter may be slightly increased with the forehand technique.

Advantages of the Perpendicular Technique

The perpendicular welding technique has the following advantages:

- Machine and robotic welding—The perpendicular gun angle is used on automated welding because there is no need to change the gun angle when the weld changes direction.
- Uniform bead shape—The weld's penetration and reinforcement are balanced between those of forehand and backhand techniques.

Disadvantages of the Perpendicular Technique

The disadvantages of using the perpendicular welding technique are the following:

- Limited visibility—Because the welding gun is directly over the weld, there is limited visibility of the weld unless you lean your head way over to the side.
- Weld spatter—Because the weld nozzle is directly under the weld in the overhead position, more weld spatter can collect in the nozzle, causing gas flow problems or even shorting the tip to the nozzle.

Advantages of the Backhand Technique

The backhand welding technique has the following advantages:

- Weld bead visibility—It is easy to see the back of the molten weld pool as you are welding, which makes it easier to control the bead shape, Figure 3.22.
- Travel speed—Because of the larger amount of weld metal being applied, the rate of travel may be slower, making it easier to create a uniform weld.
- Depth of fusion—The arc force and the greater heat from the slower travel rate both increase the depth of weld joint penetration.

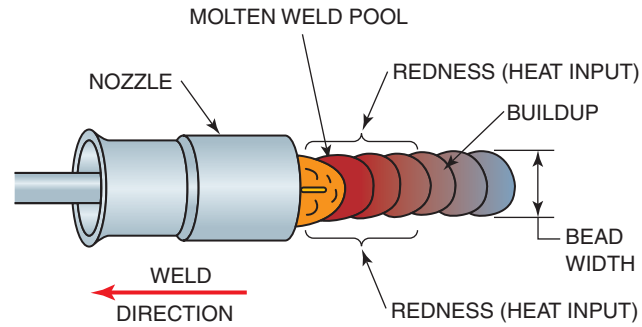

Figure 3.22 Welder's view with a backhand angle
Watch the trailing edge of the molten weld pool.

Disadvantages of the Backhand Technique

The disadvantages of the backhand welding technique are the following:

- Weld buildup—The weld bead may have a convex (raised or rounded) weld face when you use the backhand technique.
- Postweld finishing—Because of the weld bead shape, more work may be required if the product has to be finished by grinding smooth.
- Joint following—It is harder to follow the joint because your hand and the FCAW gun are positioned over the joint, and you may wander from the seam.
- Loss of penetration—An inexperienced welder sometimes directs the wire too far back into the weld pool causing the wire to build up in the face of the weld pool reducing joint penetration.

Travel Speed

The American Welding Society defines travel speed as the linear rate at which the arc is moved along the weld joint. Fast travel speeds deposit less filler metal. If the rate of travel increases, the filler metal cannot be deposited fast enough to adequately fill the path melted by the arc. This causes the weld bead to have a groove melted into the base metal next to the weld and left unfilled by the weld. This condition is known as undercut.

Undercut occurs along the edges or toes of the weld bead. Slower travel speeds will, at first, increase penetration and increase the filler weld metal deposited. As the filler metal increases, the weld bead will build up in the weld pool. Because of the deep penetration of flux cored wire, the angle at which you hold the gun is very important for a successful weld.

If all welding conditions are correct and remain constant, the preferred rate of travel for maximum weld penetration is a travel speed that allows you to stay within the selected welding variables and still control the fluidity of the weld pool. This is an intermediate travel speed, or progression, which is not too fast or too slow.

Another way to figure out correct travel speed is to consult the manufacturer's recommendations chart for the inches per minute (ipm) burn-off rate for the selected electrode.

Mode of Metal Transfer

The mode of metal transfer is used to describe how the molten weld metal is transferred across the arc to the base metal. The mode of metal transfer that is selected, the shape of the completed weld bead, and the depth of weld penetration depend on the welding power source, wire electrode size, type and thickness of material, type of shielding gas used, and best welding position for the task.

Spray Transfer with FCAW-G

The spray transfer mode is the most common process used with gas-shielded FCAW (FCAW-G), Figure 3.15.

As the gun trigger is depressed, the shielding gas automatically flows and the electrode bridges the distance from the contact tube to the base metal, making contact with the base metal to complete a circuit. The electrode shorts and becomes so hot that the base metal melts and forms a weld pool. The electrode melts into the weld pool and burns back toward the contact tube. A combination of high amperage and the shielding gas along with the electrode size produces a pinching effect on the molten electrode wire, causing the end of the electrode wire to spray across the arc.

The characteristic of spray-type transfer is a smooth arc, through which hundreds of small droplets per second are transferred through the arc from the electrode to the weld pool. At that moment a transfer of metal is taking place. Spray transfer can produce a high quantity of metal droplets, up to approximately 250 per second above the transition current, or critical current. This means the current required for a spray transfer to take place is dependent on the electrode size, composition of the electrode, and shielding gas. Below the transition current (critical current), globular transfer takes place.

In order to achieve a spray transfer, high current and larger-diameter electrode wire are needed. A shielding gas of carbon dioxide (CO_2), a mixture of carbon dioxide (CO_2) and argon (Ar), or an argon (Ar) and oxygen (O_2) mixture is needed. FCAW-G is a welding process that, with the correct variables, can be used

- on thin or properly prepared thick sections of material
- on a combination of thick to thin materials
- with small or large electrode diameters
- with a combination of shielding gases

Globular Transfer with FCAW-G

Globular transfer occurs when the welding current is below the transition current, Figure 3.23. The electrode forms a molten ball at its end that grows in size to approximately two to three times the original electrode diameter. These large molten balls are then transferred across the arc at the rate of several drops per second.

The arc becomes unstable because of the gravitational pull from the weight of these large drops. A spinning effect caused by a natural phenomenon takes place when argon gas is introduced to a large ball of molten metal on the electrode. The molten ball spins as it transfers across the arc to the base metal. This unstable globular transfer can produce excessive spatter.

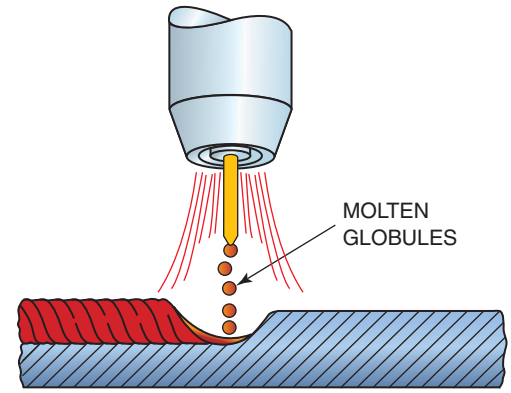

Figure 3.23 Globular transfer method

Both FCAW-S and FCAW-G use direct current electrode negative (DCEN) when welding on thin-gauge materials to keep the heat in the base metal and the small-diameter electrode at a controllable burn-off rate. The electrode can then be stabilized, and it is easier to manipulate and control the weld pool in all weld positions. Larger-diameter electrodes are welded with direct current electrode positive (DCEP) because the larger diameters can keep up with the burn-off rates.

The recommended weld position means the position in which the workpiece is placed for welding. All welding positions use either spray or globular transfer, but for now we will concentrate on the flat and horizontal welding positions.

In the flat welding position the workpiece is placed flat on the work surface. In the horizontal welding position the workpiece is positioned perpendicular to the workbench surface.

The amperage range may be from 30 to 400 amperes or more for welding materials from gauge thickness up to 1-1/2 inches. On square groove weld joints, thicker base metals can be welded with little or no edge preparation. This is one of the great advantages of FCAW. If edges are prepared and cut at an angle (beveled) to accept a complete joint weld penetration, the depth of penetration will be greatly increased. FCAW is commonly used for general repairs to mild steel in the horizontal, vertical, and overhead welding positions, sometimes referred to as out-of-position welding.

Electrode Extension

The electrode extension is measured from the end of the electrode contact tube to the point the arc begins at the end of the electrode, Figure 3.24. Compared to GMA welding, the electrode extension required for FCAW is much greater. The longer extension is required for several reasons. The electrical resistance of the wire causes the wire to heat up, which can drive out moisture from the flux. This preheating of the wire also results in a smoother arc with less spatter.

Porosity

FCA welding can produce high-quality welds in all positions, although porosity in the weld can be a persistent problem. Porosity can be caused by moisture in the flux, improper gun manipulation, or surface contamination.

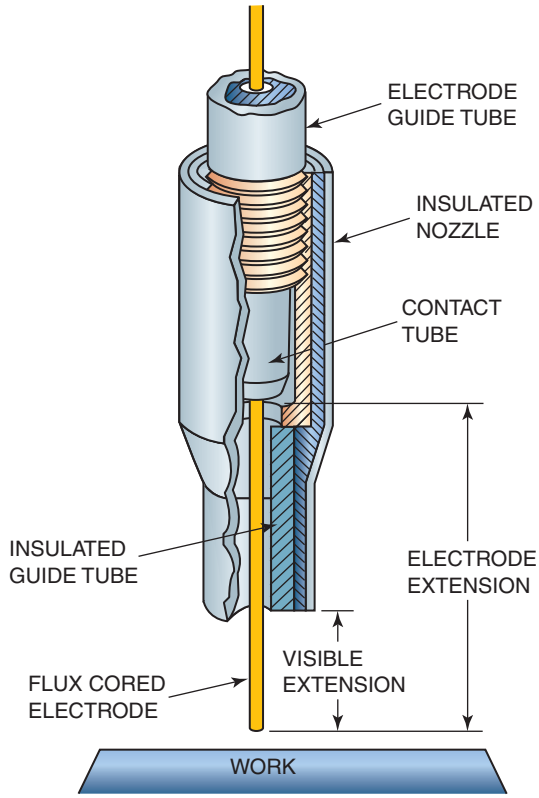

Figure 3.24 Self-shielded electrode nozzle
Source: Courtesy of the American Welding Society

The flux used in the FCA welding electrode is subject to picking up moisture from the surrounding atmosphere, so the electrodes must be stored in a dry area. Once the flux becomes contaminated with moisture, it is very difficult to remove. Water (H_2O) breaks down into free hydrogen and oxygen in the presence of an arc, Figure 3.25. The hydrogen can be absorbed into the molten weld metal, where it can cause postweld cracking. The oxygen is absorbed into the weld metal also, but it forms oxides in the metal.

If a shielding gas is used, the FCA welding gun gas nozzle must be close enough to the weld to provide adequate shielding gas coverage. If there is a wind or if the nozzle-to-work distance is excessive, the shielding will be inadequate and allow weld porosity. If welding is to be done outside or in an area subject to drafts, the gas flow rate must be increased or a wind shield must be placed to protect the weld, Figure 3.26.

A common misconception is that the flux within the electrode will either remove or control weld quality problems caused by surface contaminations. That is not true. The addition of flux makes FCA welding more tolerant to surface conditions than GMA welding, although it still is adversely affected by such contaminations.

New hot-rolled steel has a layer of dark gray or black iron oxide called mill scale. Although this layer is very thin, it may provide a source of enough oxygen to cause porosity in the weld. If mill scale causes porosity, it is usually uniformly scattered through the weld, Figure 3.27. Unless it is severe, uniformly scattered porosity is usually not visible in the finished weld. It is trapped under the surface as the weld cools.

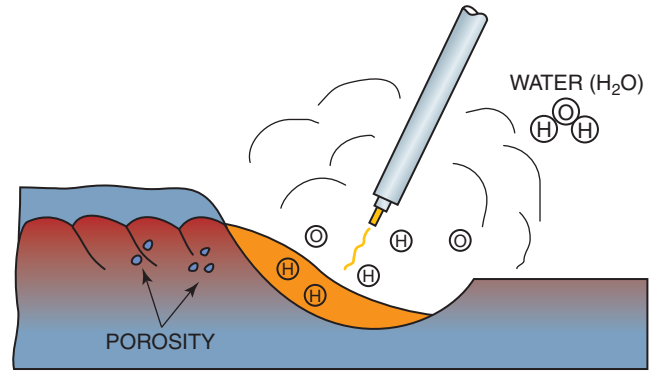

Figure 3.25 Water and porosity
Water (H_2O) breaks down in the presence of the arc and the hydrogen (H) is dissolved in the molten weld metal.

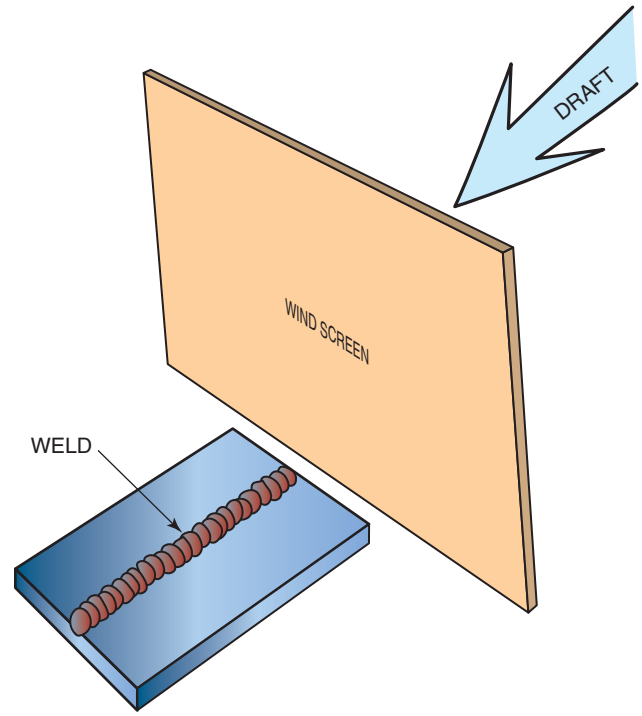

Figure 3.26 Wind and draft protection
A wind screen can keep the welding shielding from being blown away.

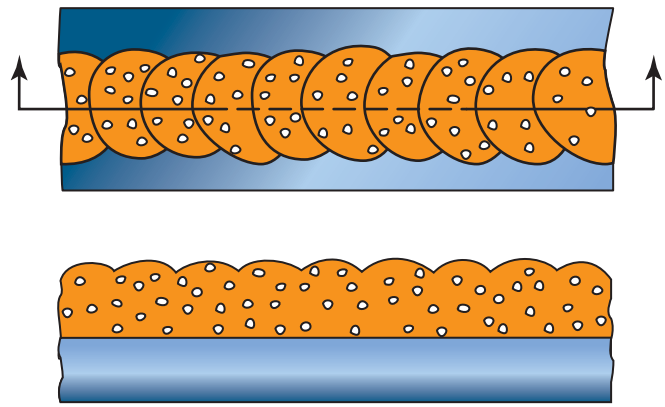

Figure 3.27 Uniformly scattered porosity

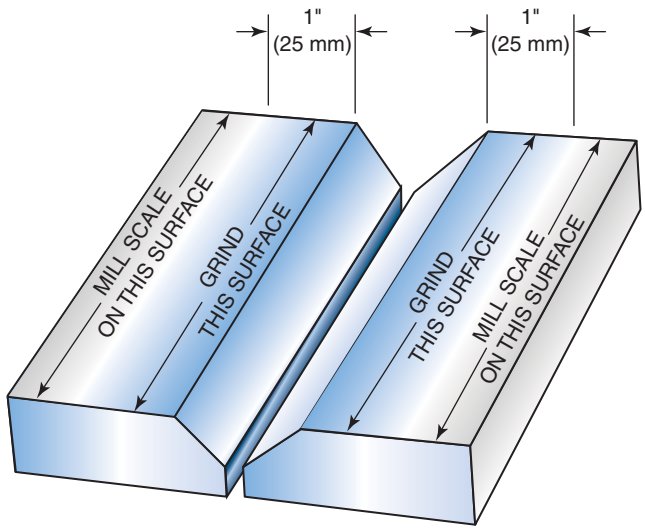

Figure 3.28 Grinding requirements
Grind mill scale off plates within 1 in. (25 mm) of the groove.

Because porosity is under the weld surface, nondestructive testing methods, including X ray, magnetic particle, and ultrasound, must be used to locate it in a weld. It can be detected by mechanical testing such as guided bend, free bend, and nick-break testing for establishing weld parameters. Often it is better to remove the mill scale before welding rather than risking the production of porosity.

All welding surfaces within the weld groove and the surrounding surfaces within 1 in. (25 mm) must be cleaned to bright metal, Figure 3.28. Cleaning may be either grinding, filing, sanding, or blasting.

Any time FCA welds are to be made on metals that are dirty, oily, rusty, or wet or that have been painted, the surface must be precleaned. Cleaning can be done chemically or mechanically.

One advantage of chemically cleaning oil and paint is that it is easier to clean larger areas. Both oil and paint smoke easily when heated, and such smoke can cause weld defects. They must be removed far enough from the weld so that weld heat does not cause them to smoke. In the case of small parts the entire part may need to be cleaned.

Caution

Chemically cleaning oil and paint off metal must be done according to the cleaner manufacturer's directions. The work must be done in an appropriate, approved area. The metal must be dry, and all residues of the cleaner must be removed before welding begins.

SUMMARY

Flux cored arc welding is used to produce more tons of welded fabrications than any other process. The ability to produce high-quality welds on a wide variety of material thicknesses and joint configurations has led to its popularity. As you learn and develop these skills, you will therefore be significantly increasing your employability and productivity in the welding industry.

A wide variety of filler metals and shielding gas combinations for flux cored arc welding are available to you in industry. These various materials aid in producing welds of high quality under various welding conditions. Although the selection of the proper filler metal and gas coverage, if used, will significantly affect the finished weld's quality in the field, there are very few differences in manipulation and setup among these

filler metals. Therefore, as you practice welding in a school or training program and learn to use a specific wire and shielding gas mixture, these skills are easily transferable to the next group of materials you will encounter on the job.

REVIEW

1. List some factors that have led to the increased use of FCA welding.
2. How is FCAW similar to GMAW?
3. What does the FCA flux provide to the weld?
4. What are the major atmospheric contaminations of the molten weld metal?
5. How does slag help an FCA weld?
6. What is the electrical difference between a constant-potential and a constant-current power supply?
7. How can FCA welding guns be cooled?
8. What problems does excessive drive roller pressure cause?
9. List the advantages that FCA welding offers the welding industry.
10. Describe the two methods of manufacturing FCA electrode wire.
11. Why are large-diameter electrodes not used for all-position welding?
12. How do deoxidizers remove oxygen from the weld zone?
13. What do fluxing agents do for a weld?
14. Why are alloying elements added to the flux?
15. How does the flux form a shielding gas to protect the weld?
16. What are the main limitations of the rutile fluxes?
17. Why is it more difficult to use lime-based fluxed electrodes on out-of-position welds?
18. What benefit does adding an externally supplied shielding gas have on some rutile-based electrodes?
19. How do excessive amounts of manganese affect a weld?
20. Why are elements added that cause ferrite to form in the weld?
21. Why are some slags called refractory?
22. Why must a flux form a less dense slag?
23. Referring to Table 3.5, what is the AWS classification for FCA welding electrodes for stainless steel?
24. Describe the meaning of each part of the following FCA welding electrode identification: E81T-5.
25. What does the number 316 in E316T-1 mean?
26. What is the advantage of using an argon-CO_2 mixed shielding gas?
27. What are the primary alloying elements lost if 100% CO_2 shielding gas is used?
28. What can cause porosity in an FCA weld?
29. What happens to water in the welding arc?
30. What is the thin, dark gray or black layer on new hot-rolled steel? How can it affect the weld?
31. Why is uniformly scattered porosity hard to detect in a weld?
32. What cautions must be taken when chemically cleaning oil or paint from a piece of metal?
33. What can happen to slag that solidifies on the plate ahead of the weld?
34. How is the electrode extension measured?

CHAPTER

4 Flux Cored Arc Welding

OBJECTIVES

After completing this chapter, the student should be able to

- set up the FCA weld station
- thread the electrode wire through the system
- list three disadvantages of having to bevel a plate before welding
- make root, filler, and cover passes with the FCAW process
- make butt welds in all positions that can pass a specified standard's visual or destructive examination criteria
- make fillet welds in tee joints and lap joints in all positions that can pass a specified standard's visual or destructive examination criteria

KEY TERMS

amperage range	feed rollers	tee joint
conduit liner	lap joint	voltage range
contact tube	root face	weave bead
critical weld	stringer bead	wire-feed speed

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter:

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports, or records
Key Indicator 2: Performs housekeeping duties
Key Indicator 3: Follows verbal instructions to complete work assignments
Key Indicator 4: Follows written instructions to complete work assignments

Module 6: Flux Cored Arc Welding (FCAW-G/GM, FCAW-S)

Key Indicator 1: Performs safety inspections of FCAW equipment and accessories
Key Indicator 2: Makes minor external repairs to FCAW equipment and accessories

Gas Shielded

Key Indicator 3: Sets up for FCAW-G/GM operations on carbon steel
Key Indicator 4: Operates FCAW-G/GM equipment on carbon steel
Key Indicator 5: Makes FCAW-G/GM fillet welds, in all positions, on carbon steel

Key Indicator 6: Makes FCAW-G/GM groove welds, in all positions, on carbon steel
Key Indicator 7: Passes FCAW-G/GM welder performance qualification testing (workmanship sample) on carbon steel
Self Shielded
Key Indicator 8: Sets up for FCAW-S operations on carbon steel
Key Indicator 9: Operates FCAW-S equipment on carbon steel
Key Indicator 10: Makes FCAW-S fillet welds, in all positions, on carbon steel
Key Indicator 11: Makes FCAW-S groove welds, in all positions, on carbon steel
Key Indicator 12: Passes FCAW-S welder performance qualification test (workmanship sample) on carbon steel

Module 9: Welding Inspection and Testing Principles

Key Indicator 1: Examines cut surfaces and edges of prepared base metal parts
Key Indicator 2: Examines tacks, root passes, intermediate layers, and completed welds

INTRODUCTION

Setup of the flux cored arc welding (FCAW) work station is the key to making quality welds. It may be possible, using a poorly set up FCA welder, to make an acceptable weld in the flat position. The FCA welding process is often forgiving; thus welds can often be made even when the welder is not set correctly. However, such welds will have major defects such as excessive spatter, undercut, overlap, porosity, slag inclusions, and poor weld bead contours. Setup becomes even more important for out-of-position welds. Making vertical and overhead welds can be difficult for a student welder with a properly set up system, but it becomes impossible with a system that is out of adjustment.

Learning to set up and properly adjust the FCA welding system will allow you to produce high-quality welds at a high level of productivity.

FCAW is set up and manipulated in a manner similar to that of GMAW. The results of changes in electrode extension, voltage, amperage, and torch angle are essentially the same.

Although every manufacturer's FCA welding equipment is designed differently, all equipment is set up in a similar manner. It is always best to follow the specific manufacturer's recommendations regarding setup as provided in its equipment literature. You will find, however, that, in the field, manufacturers' literature is not always available for the equipment you are asked to use. It is therefore important to have a good general knowledge and understanding of the setup procedure for FCA welding equipment. Figure 4.1 shows all of the various components that make up an FCA welding station.

Caution

FCA welding produces a lot of ultraviolet light, heat, sparks, slag, and welding fumes. Proper personal, protective clothing and special protective clothing must be worn to prevent burns from the ultraviolet light and hot weld metal. Eye protection must be worn to prevent injury from flying sparks and slag. Forced ventilation and possibly a respirator must be used to prevent fume-related injuries. Refer to the safety precautions provided by the equipment and electrode manufacturers and to Chapter 2 in *Welding Skills, Processes and Practices for Entry-Level Welders: Book One* for additional safety help.

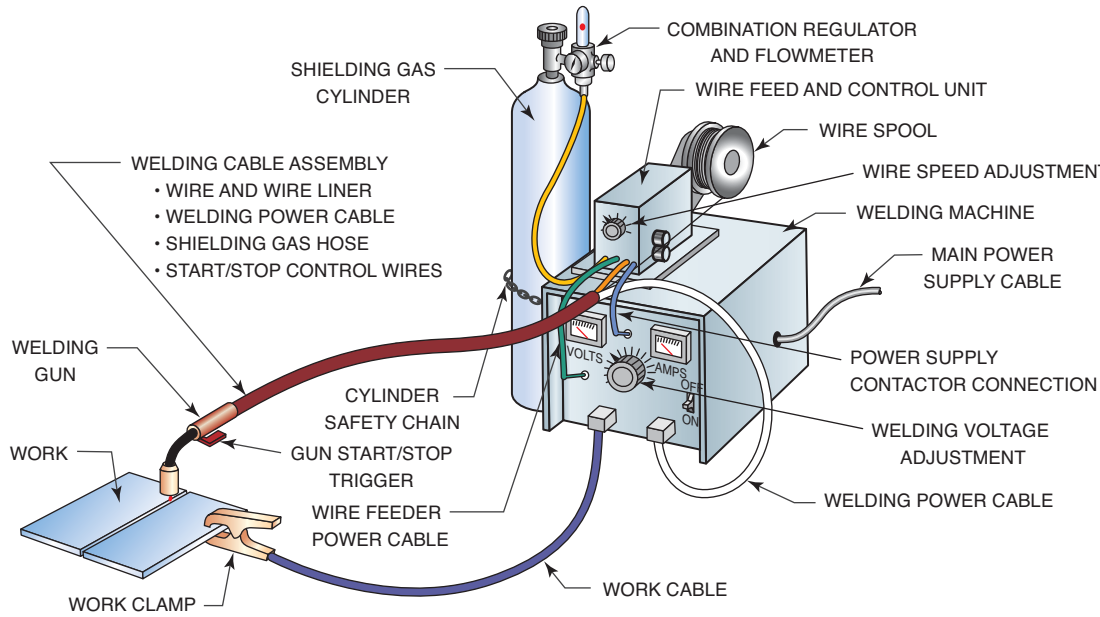

Figure 4.1 **Basic FCA welding equipment identification**

PRACTICES

The practices in this chapter are grouped according to those requiring similar techniques and setups. Plate welds are covered first, then sheet metal. The practices start with 1/4-in. (6-mm) mild steel plates; they are used because they require the least preparation times. The thicker 3/8-in. (9.5-mm) plates provide the basics of practicing groove welding. The 3/4-in. (19-mm) and thicker plates are used to develop the skills required to pass the unlimited thickness test often given to FCA welders. Sheet metal is grouped together because it presents a unique set of learning skills.

The major skill required for making consistently acceptable FCA welds is the ability to set up the welding system. Changes such as variations in material thickness, position, and type of joint require changes both in technique and setup. A correctly set up FCA welding station can, in many cases, be operated by a less-skilled welder. Often the only difference between a welder earning a minimum wage and one earning the maximum wage is the ability to correct machine setups.

For several reasons the FCA welding practice plates will be larger than most other practice plates. Welding heat and welding speed are the major factors that necessitate this increased size. FCA welding is both high energy and fast, and the welding energy (heat) input is so great that small practice plates may glow red by the end of a single weld pass. This would seriously affect the weld quality. To prevent this from happening, wider plates are used. Because of the higher welding speeds, longer plates are usually used.

Plates less than 1/2 in. (13 mm) will be 12 in. (305 mm) long for most practices. In addition to controlling the heat buildup, the longer plates are needed to give the welder enough time to practice welding. Learning to make longer welds is a skill that must also be practiced, because the FCA welding process is used in industry to make long production welds.

Plates thicker than 1/2 in. (13 mm) can be shorter than 12 in. (305 mm). Most codes allow test plates of "unlimited thickness" to be as short as 7 in. (178 mm).

PRACTICE 4-1

FCAW Equipment Setup

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 2
Gas Shielded
Key Indicator 3
Self Shielded
Key Indicator 8

For this practice, you will need a semiautomatic welding power source approved for FCA welding, welding gun, electrode feed unit, electrode supply, shielding gas supply, shielding gas flowmeter regulator, electrode conduit, power and work leads, shielding gas hoses (if required), assorted hand tools, spare parts, and any other required materials. In this practice, you will demonstrate to a group of students and your instructor how to properly set up an FCA welding station. Some manufacturers include detailed setup instructions with their equipment. If such instructions are available for your equipment, follow them. Otherwise, use the following instructions.

If the shielding gas is to be used and it comes from a cylinder, the cylinder must be chained securely in place before the valve protection cap is removed, Figure 4.2. Standing to one side of the cylinder, make sure nobody is in line with the valve and quickly crack the valve to blow out any dirt in the valve before the flowmeter regulator is attached, Figure 4.3. Attach the correct hose from the regulator to the "gas-in" connection on the electrode feed unit or machine.

Install the reel of electrode (welding wire) on the holder and secure it, Figure 4.4. Check the feed roller size to ensure that it matches the wire size, Figure 4.5. Also check the **conduit liner** size for compatibility with the wire size. Connect the conduit to the feed unit. The conduit or an extension should be aligned with the groove in the roller and set as close to the roller as possible without touching, Figure 4.6. Misalignment at this point can contribute to a bird's nest, Figure 4.7. Bird-nesting of the electrode wire, so called because it looks like a bird's nest, results when the feed roller pushes the wire into a tangled ball because the wire would not go through the outfeed side conduit.

Figure 4.2 Make sure the gas cylinder is chained securely in place before removing the safety cap
Source: Courtesy of Larry Jeffus

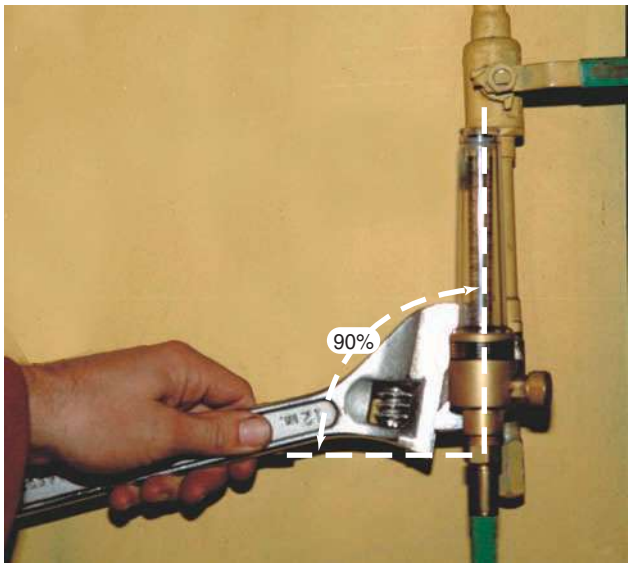

Figure 4.3 Attaching the flowmeter regulator
Be sure the tube is vertical.
Source: Courtesy of Larry Jeffus

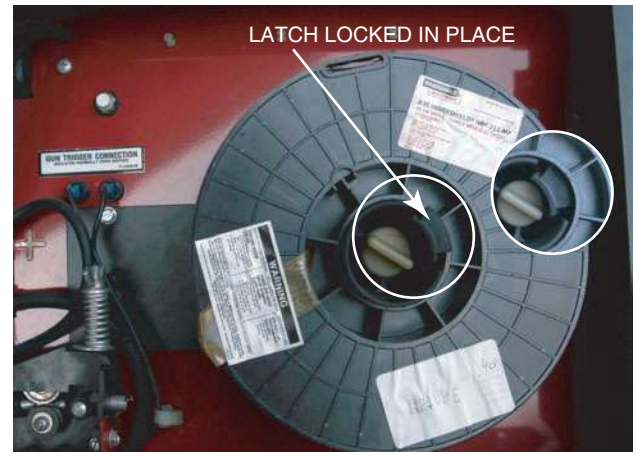

Figure 4.4 Wire reel may be secured by a center nut or locking lever
Source: Courtesy of Larry Jeffus

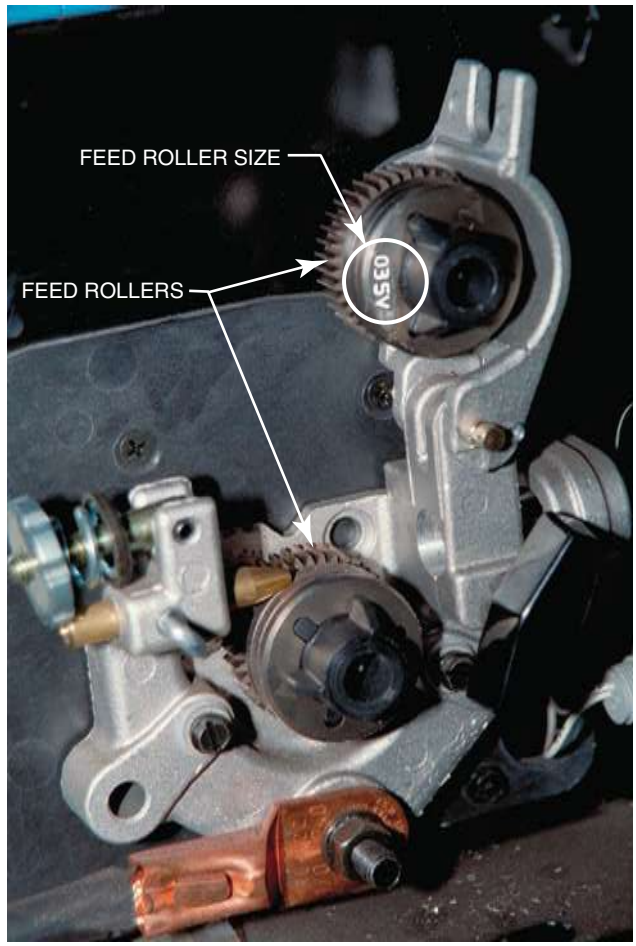

Figure 4.5 Checking feed roller size
Check to be certain that the feed rollers are the correct size for the wire being used.
Source: Courtesy of Larry Jeffus

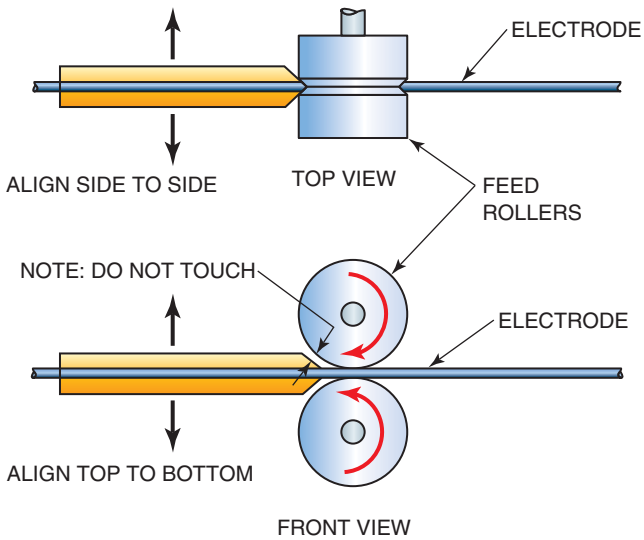

Figure 4.6 Feed roller and conduit alignment

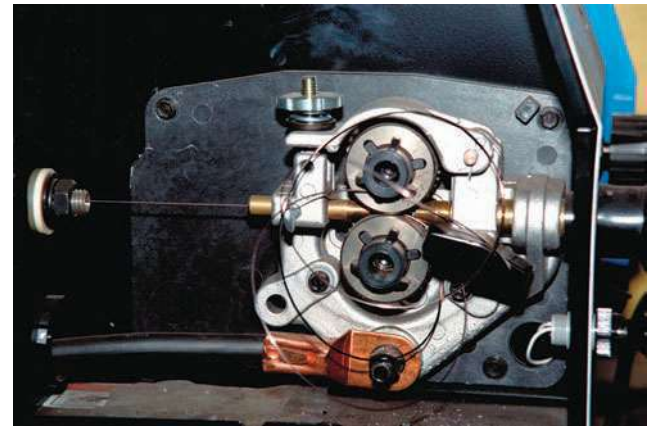

Figure 4.7 "Bird's nest" in the filler wire at the feed rollers
Source: Courtesy of Larry Jeffus

Be sure the power is off before attaching the welding cables. The electrode and work leads should be attached to the proper terminals. The electrode lead should be attached to the electrode or positive (+) terminal. If necessary, it is also attached to the power cable part of the gun lead. The work lead should be attached to the work or negative (–) terminal.

The shielding "gas-out" side of the solenoid is then also attached to the gun lead. If a separate splice is required from the gun switch circuit to the feed unit, it should be connected at this time. Check that the welding contactor circuit is connected from the feed unit to the power source.

The welding cable liner or wire conduit must be securely attached to the gas diffuser and contact tube, Figure 4.8. The **contact tube** (tip), Figure 3.1, must be the correct size to match the electrode wire size being used. If a shielding gas is to be used, a gas nozzle would be attached to complete the assembly. If a gas nozzle is not needed for a shielding gas, it may still be installed. Because it is easy for a student to touch the work with the contact tube during welding, an electrical short may occur. This short-out of the contact tube will immediately destroy the tube. Although the gas nozzle may interfere with some visibility, it may be worth the trouble for a new welder. FCA welding is more sensitive to changes in arc voltage than is SMA (stick) welding. Such variations in FCA welding voltage can dramatically and adversely affect your ability to maintain weld bead control.

A loose or poor connection will result in increased circuit resistance and a loss of welding voltage. To be sure that you have a good work

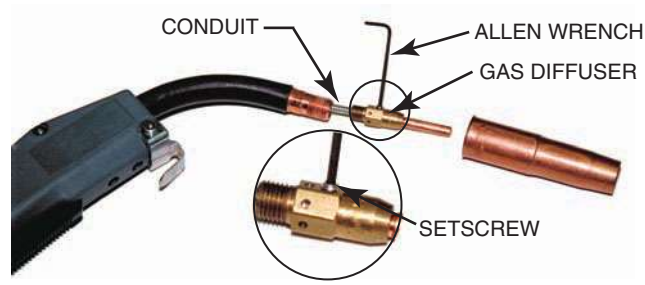

Figure 4.8 Securely attach conduit to gas diffuser and contact tube to prevent wire jams caused by misalignment
Source: Courtesy of Larry Jeffus

connection, remove any dirt, rust, oil, or other surface contamination at the point where the work clamp is connected to the weldment.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-2

Threading FCAW Wire

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 2
Gas Shielded
Key Indicator 3
Self Shielded
Key Indicator 8

Using the FCAW machine that was properly assembled in Practice 4-1, you will turn the machine on and thread the electrode wire through the system.

Check that the unit is assembled correctly according to the manufacturer's specifications. Switch on the power and check the gun switch circuit by depressing the switch. The power source relays, feed relays, gas solenoid, and feed motor should all activate.

Cut off the end of the electrode wire if it is bent. When working with the wire, be sure to hold it tightly. The wire will become tangled if it is released. The wire has a natural curl known as *cast.* Straighten out about 12 in. (300 mm) of the curl to make threading easier. Separate the wire **feed rollers** and push the wire first through the guides, then between the rollers, and finally into the conduit liner, Figure 4.9. Reset the rollers so there is a slight amount of compression on the wire, Figure 4.10. Set the **wire-feed speed** control to a slow speed. Hold the welding gun so that the electrode conduit and cable are as straight as possible.

Wearing safety glasses and pointing the gun away from the welder's face, press the gun switch or the cold feed switch on the feeder. Pressing the gun switch to start the wire feeder is called triggering the gun. The cold feed switch on the feeder is a safety option built into some equipment that advances the wire without current being sent to the gun. The wire should start feeding into the liner. Watch to make certain that the wire feeds smoothly and release the gun switch as soon as the end comes through the gun.

If the wire stops feeding before it reaches the end of the gun, stop and check the system. If no obvious problem can be found, mark the wire with tape and remove it from the gun. It can then be held next to the system to determine the location of the problem.

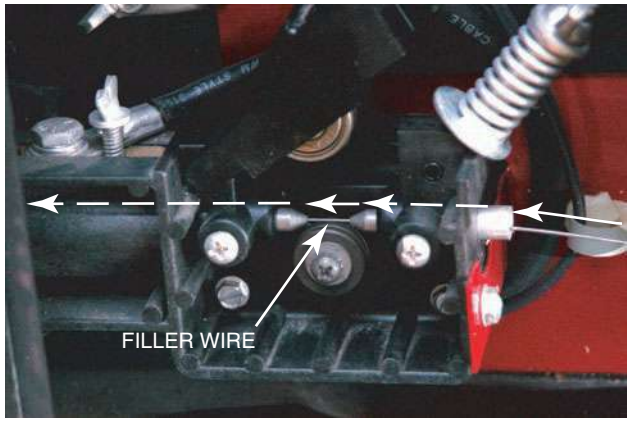

Figure 4.9 Push the wire through the guides by hand
Source: Courtesy of Larry Jeffus

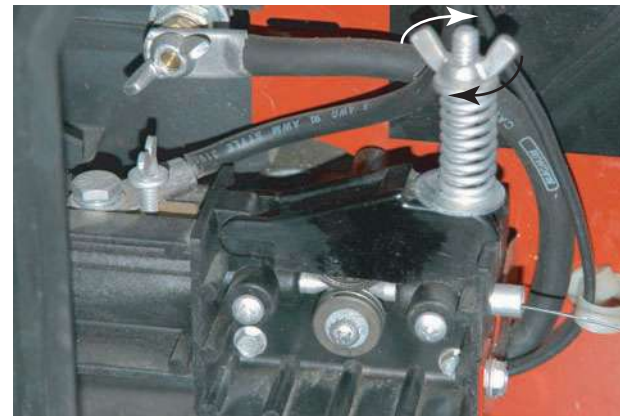

Figure 4.10 Adjust the wire-feed tensioner
Source: Courtesy of Larry Jeffus

With the wire feed running, adjust the feed roller compression so that the wire reel can be stopped easily by a slight pressure. Too light a roller pressure will cause the wire to feed erratically. Too high a pressure can crush some wires, causing some flux to be dropped inside the wire liner. If this happens, you will have a continual problem with the wire not feeding smoothly or jamming.

With the feed running, adjust the spool drag so that the reel stops when the feed stops. The reel should not coast to a stop, because it allows slack in the wire that can easily be snagged. Also, when the feed restarts, a jolt occurs when the slack in the wire is taken up. This jolt can be enough to momentarily stop the wire, possibly causing a discontinuity in the weld.

When the test runs are completed, you can either rewind or cut off the wire. Some wire-feed units have a retract button. This allows the feed driver to reverse and retract the wire automatically. To rewind the wire on units without this retraction feature, release the rollers and turn them backward by hand. If the machine will not allow the feed rollers to be released without upsetting the tension, you must cut the wire. Some wire reels have covers to prevent the collection of dust, dirt, and metal filings on the wire, Figure 4.11.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

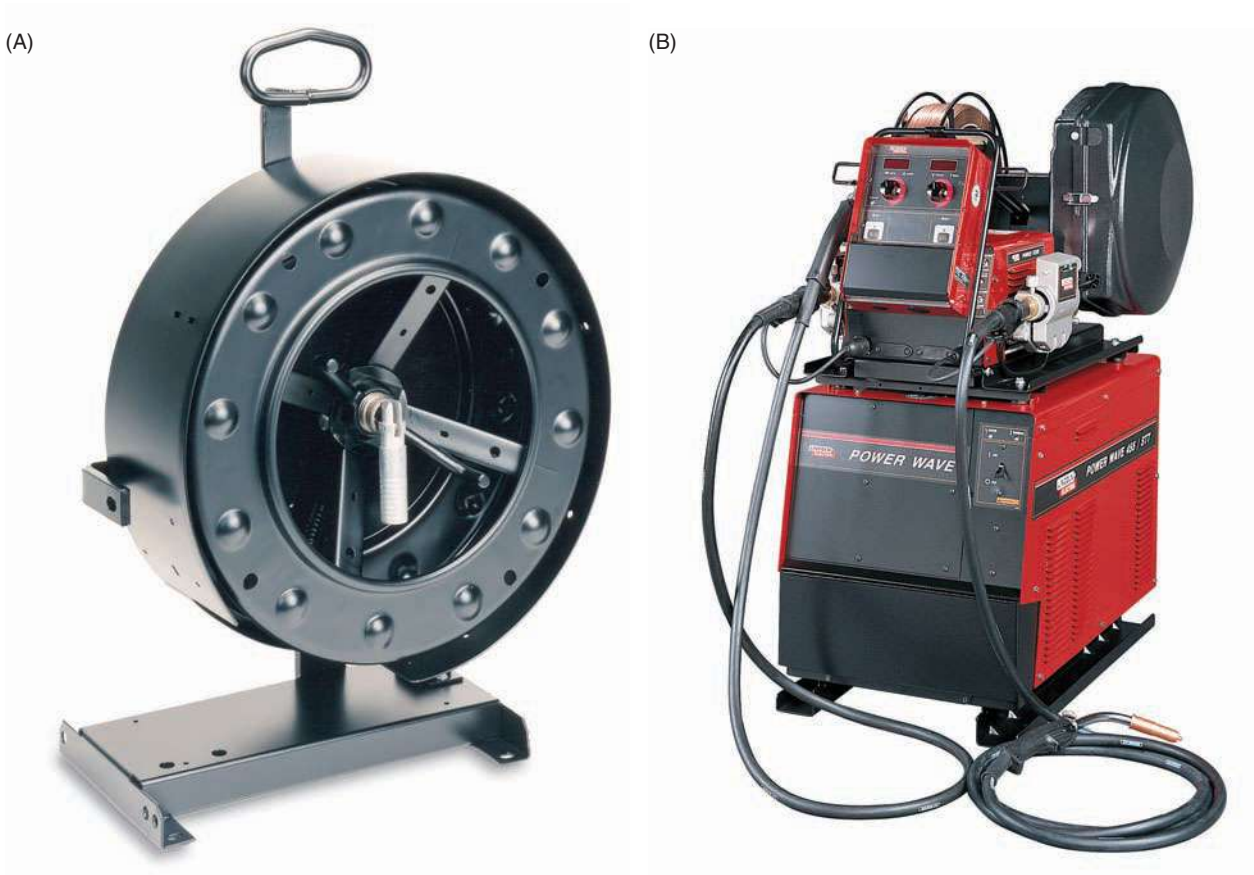

Figure 4.11 Wire protection
(A) Covered wire reel. (B) Wire cover on a dual wire-feed system.
Source: Courtesy of Lincoln Electric Company

FLAT-POSITION WELDS

PRACTICE 4-3

Stringer Beads, Flat Position

Using a properly set up and adjusted FCA welding machine, Table 4.1, proper safety protection, E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) or thicker, you will make a stringer bead weld in the flat position, Figure 4.12.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 1
Gas Shielded
Key Indicator 4
Self Shielded
Key Indicator 9

Table 4.1 FCA Welding Parameters for Use if Specific Settings Are Unavailable from Electrode Manufacturer (base metal thickness 1/4 to 1/2 inch)

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min.)	Volts	Type	Flow	Type	Thickness
E70T-1 E71T-1	0.035 in. (0.9 mm)	130 to 150	288 to 380 (732 to 975)	22 to 25	None	n/a	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70T-1 E71T-1	0.045 in. (1.2 mm)	150 to 210	200 to 300 (508 to 762)	28 to 29	None	n/a	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70T-5 E71T-11	0.035 in. (0.9 mm)	130 to 200	288 to 576 (732 to 1463)	20 to 28	75% argon 25% CO_2	30 cfh	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70T-5 E71T-11	0.045 in. (1.2 mm)	150 to 250	200 to 400 (508 to 1016)	23 to 29	75% argon 25% CO_2	35 cfh	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

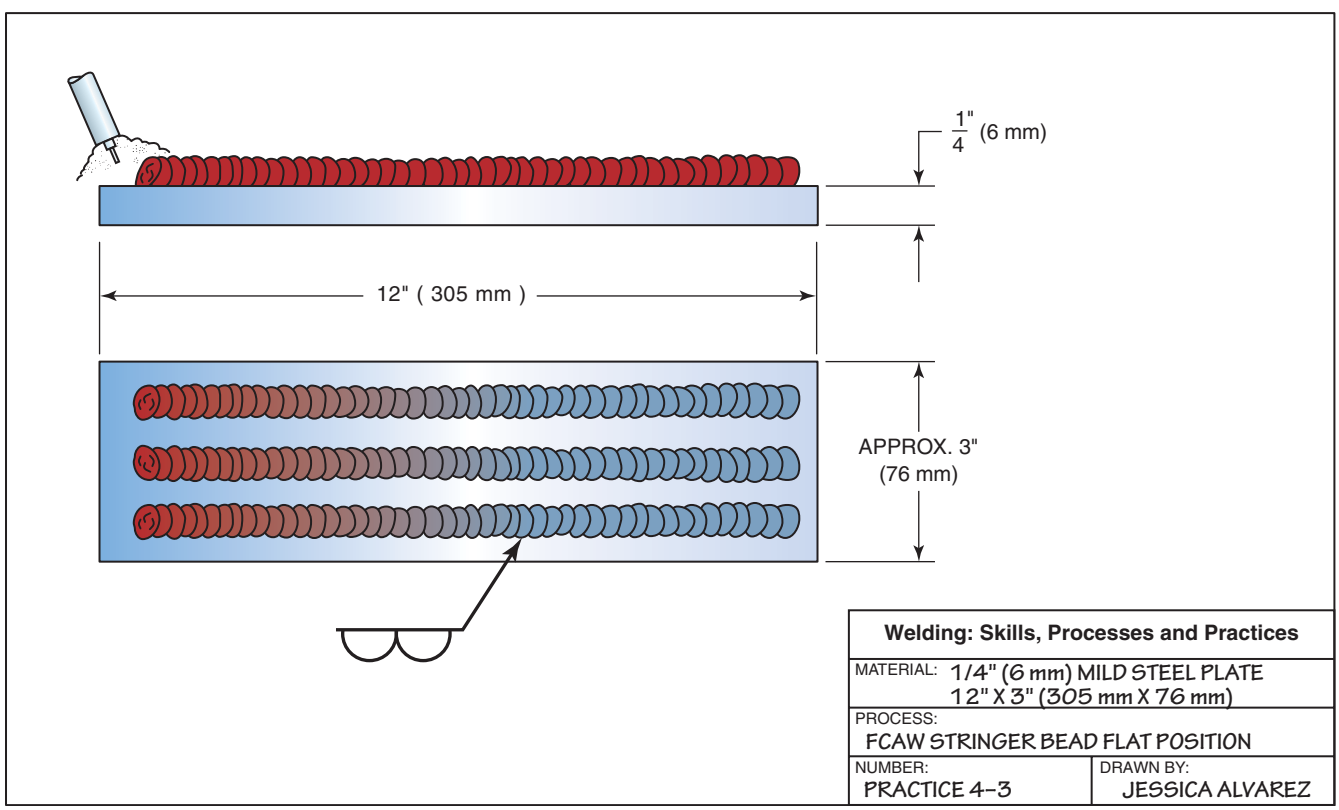

Figure 4.12 FCAW stringer bead, 1/4 in. mild steel, flat position

Starting at one end of the plate and using a dragging technique, make a weld bead along the entire 12-in. (305-mm) length of the metal. After the weld is complete, check its appearance. Make any needed changes to correct the weld. Repeat the weld and make additional adjustments. After the machine is set, start to work on improving the straightness and uniformity of the weld. Use weave patterns of different widths and straight stringers without weaving.

Repeat with both classifications of electrodes until beads can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

SQUARE-GROOVE WELDS

One advantage of FCA welding is the ability to make 100%-joint-penetrating welds without beveling the edges of the plates. These full-joint-penetrating welds can be made in plates that are 1/4 in. (6 mm) or less in thickness. Welding on thicker plates risks the possibility of a lack of fusion on both sides of the **root face**, Figure 4.13.

There are several disadvantages of having to bevel a plate before welding:

- Beveling the edge of a plate adds an operation to the fabrication process.
- Both more filler metal and welding time are required to fill a beveled joint than are required to make a square jointed weld.
- Beveled joints have more heat from the thermal beveling and additional welding required to fill the groove. The lower heat input to the square joint means less distortion.

The major disadvantage of making square jointed welds is that as the plate thickness approaches 1/4 in. (6 mm) or if the weld is out of

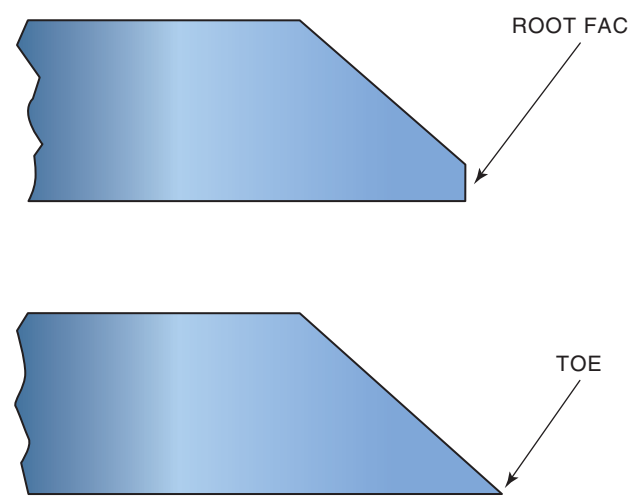

Figure 4.13 A beveled joint may or may not have a flat surface, called a root face

position, a much higher level of skill is required. The skill required to make quality square welds can be acquired by practicing on thinner metal. It is much easier to make this type of weld in metal 1/8 in. (3 mm) thick and then move up in thickness as your skills improve.

PRACTICE 4-4

Butt Joint 1G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) or less in thickness, you will make a groove weld in the flat position, Figure 4.14.

- Tack weld the plates together and place them in position to be welded.
- Starting at one end, run a bead along the joint. Watch the molten weld pool and bead for signs that a change in technique may be required.
- Make any needed changes as the weld progresses in order to produce a uniform weld.

Repeat with both classifications of electrodes until defect-free welds can consistently be made in the 1/4-in.-thick (6-mm-thick) plate. Turn

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Flat" position portion of the all-position requirement of 6 and 11.

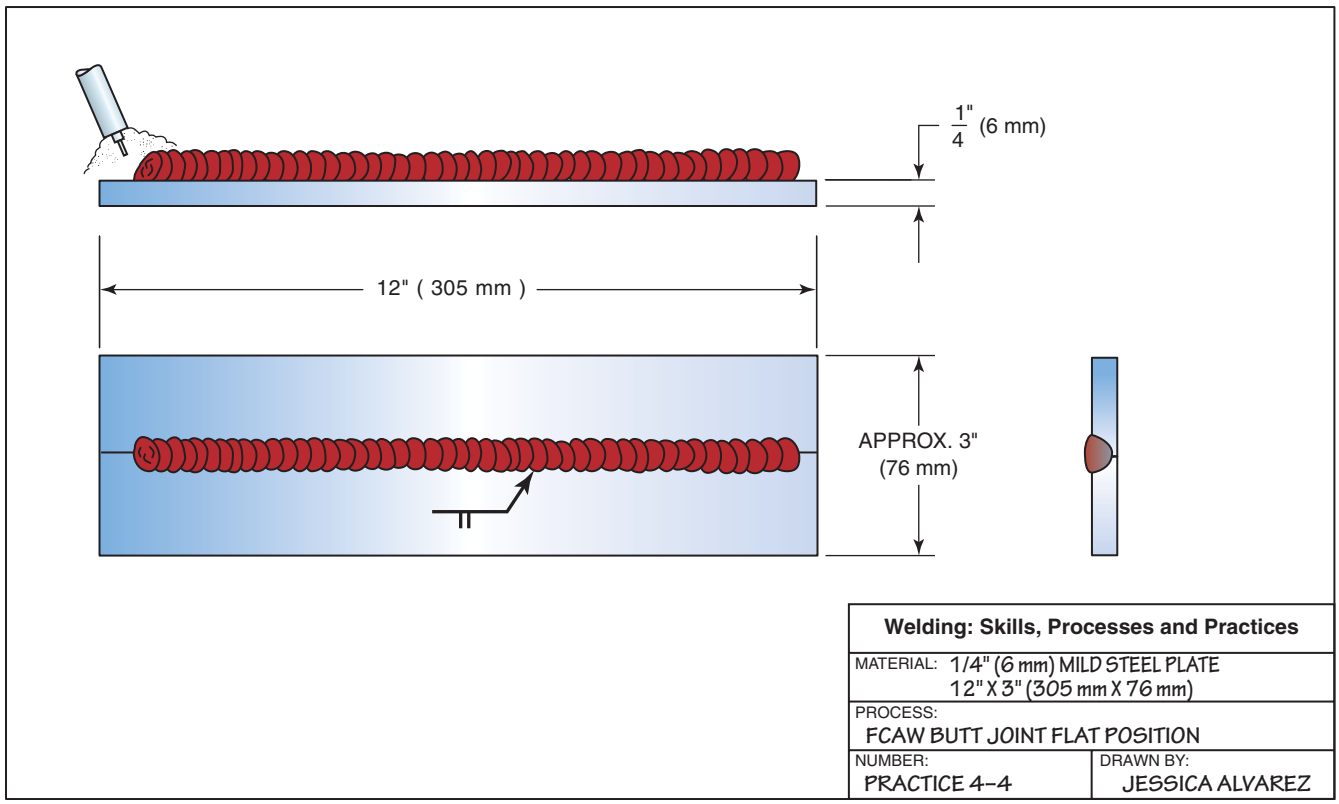

Figure 4.14 FCAW butt joint, 1/4 in. mild steel, flat position

off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

V-GROOVE AND BEVEL-GROOVE WELDS

Although for speed and economy engineers try to avoid specifying welds that require beveling the edges of plates, it is not always possible. Anytime the metal being welded is thicker than 1/4 in. (6 mm) and a 100% joint penetration weld is required, the edges of the plate must be prepared with a bevel. Fortunately, FCA welding allows a narrower groove to be made and still achieve a thorough thickness weld, due to the deeper penetration characteristics of the FCAW process, Figure 4.15.

All FCA groove welds are made using three different types of weld passes, Figure 4.16.

- *Root pass:* The first weld bead of a multiple-pass weld. The root pass fuses the two parts together and establishes the depth of weld metal penetration.
- *Filler pass:* Made after the root pass is completed and used to fill the groove with weld metal. More than one pass is often required.
- *Cover pass:* The last weld pass on a multipass weld. The cover pass may be made with one or more welds. It must be uniform in width, reinforcement, and appearance.

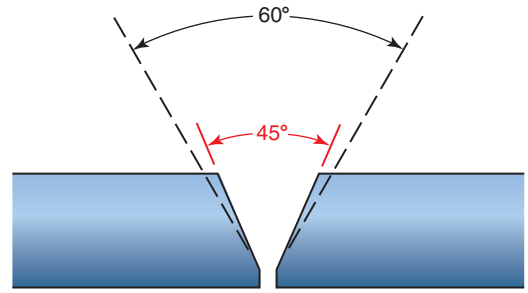

Figure 4.15 Reduced groove angle for FCAW
A smaller groove angle reduces both weld time and filler metal required to make the weld.

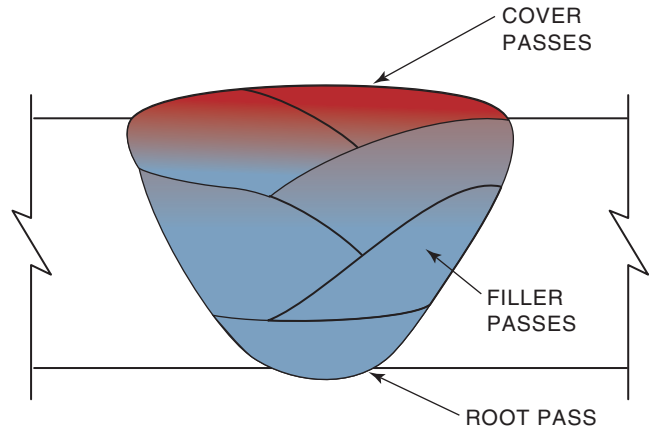

Figure 4.16 Three types of weld passes make up a weld

Root Pass

A good root pass is needed in order to obtain a sound weld. The root may be either open or closed using a backing strip, Figure 4.17.

The backing strips are usually made from a piece of metal 1/4 in. (6 mm) thick and 1 in. (25 mm) wide, and should be 2 in. (50 mm) longer than the base plates. The strip is attached to the plate by tack welds made on the sides of the strip, Figure 4.18.

Most production welds do not use backing strips, so they are made as open root welds. Because of the difficulty in controlling root weld face contours in FCAW, however, open-root joints are often avoided on **critical welds**. If an open-root weld is needed because of weldment design, the root pass may be put in with an SMAW electrode or the root face of the FCA weld can be retouched by grinding and/or back welding.

Care must be taken with any root pass not to have the weld face too convex, Figure 4.19. Convex weld faces tend to trap slag along the toe of

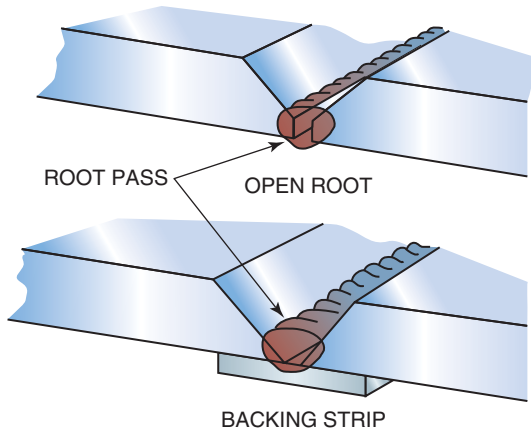

Figure 4.17 Root pass
The maximum deposit for a root pass is 1/4 in. (6 mm) thick.

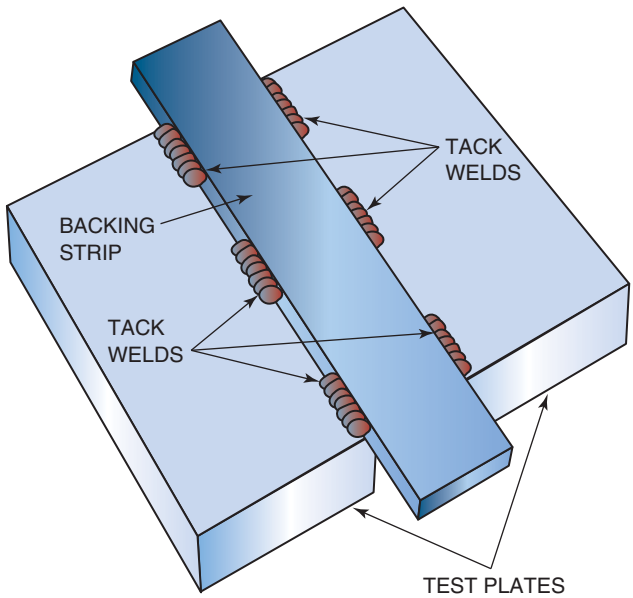

Figure 4.18 Backing strip
Securely tack weld the backing strip to the test plates.

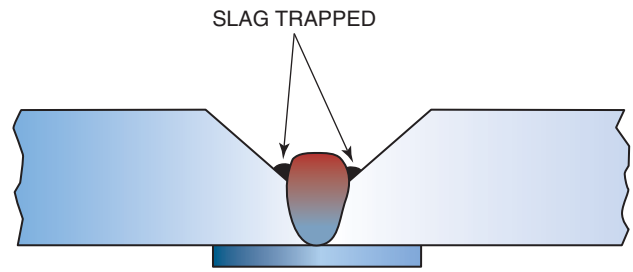

Figure 4.19 Slag trapped beside weld bead is hard to remove

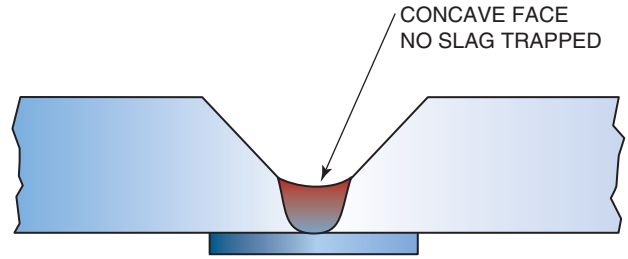

Figure 4.20 Flat or concave weld faces are easier to clean off

the weld. FCA weld slag can be extremely difficult to remove in this area, especially if there is any undercutting. To avoid this, adjust the welding power settings, speed, and weave pattern so that a flat or slightly concave weld face is produced, Figure 4.20.

Filler Pass

Filler passes are made with either **stringer beads**, which are made with a straight progression and very little gun manipulation, or **weave beads**, in which the operator oscillates the gun from side to side in order to widen the weld profile. Either bead type works well for flat or vertically positioned welds, but stringer beads work best for horizontal and overhead-positioned welds. When multiple-pass filler welds are required, each weld bead must overlap the others along the edges. Edges should overlap smoothly enough so that the finished bead is uniform, Figure 4.21. Stringer beads usually overlap about 25% to 50%, and weave beads overlap approximately 10% to 25%.

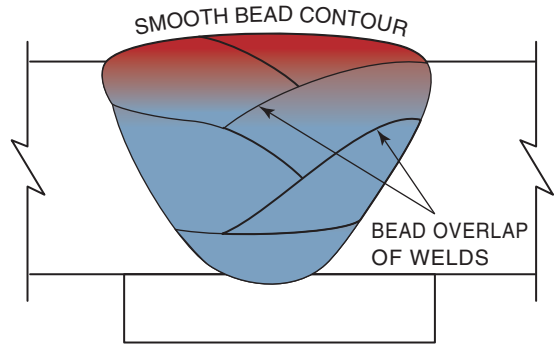

Figure 4.21 The surface of a multipass weld should be as smooth as if it were made by one weld

Each weld bead must be cleaned before the next bead is started. The filler pass ends when the groove has been filled to a level just below the plate surface.

Cover Pass

The cover pass may or may not simply be a continuation of the weld beads used to make the filler pass(es). The major difference between the filler pass and the cover pass is the weld face importance. Keeping the face and toe of the cover pass uniform in width, reinforcement, and appearance and free of defects is essential. Most welds are not tested beyond a visual inspection. For that reason the appearance might be the only factor used for accepting or rejecting welds.

The cover pass must meet a strict visual inspection standard. The visual inspection looks to see that the weld is uniform in width and reinforcement. There should be no arc strikes or hammer marks from chipping or slag removal operations on the plate other than those on the weld itself. The weld must be free of both incomplete fusion and cracks. The weld must be free of overlap, and undercut must not exceed either 10% of the base metal or 1/32 in. (0.8 mm), whichever is less. Reinforcement must have a smooth transition with the base plate and be no higher than 1/8 in. (3 mm), Figure 4.22.

PRACTICE 4-5

Butt Joint 1G

Use a properly set up and adjusted FCA welding machine, Table 4.1; proper safety protection; E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the flat position, Figure 4.23.

Tack weld the backing strip to the plates. There should be a root gap of approximately 1/8 in. (3 mm) between the plates. The beveled surface can be made with or without a root face, Figure 4.24.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Gas Shielded
Key Indicator 11
This practice addresses the "Flat" position portion of the all-position requirement of 6 and 11.

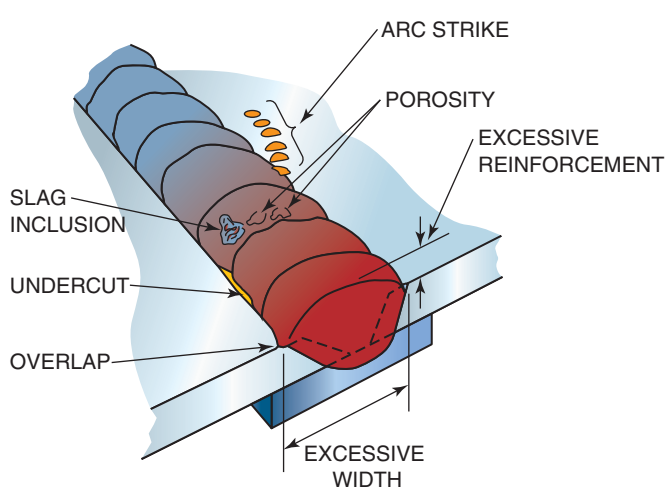

Figure 4.22 Common discontinuities found during a visual examination

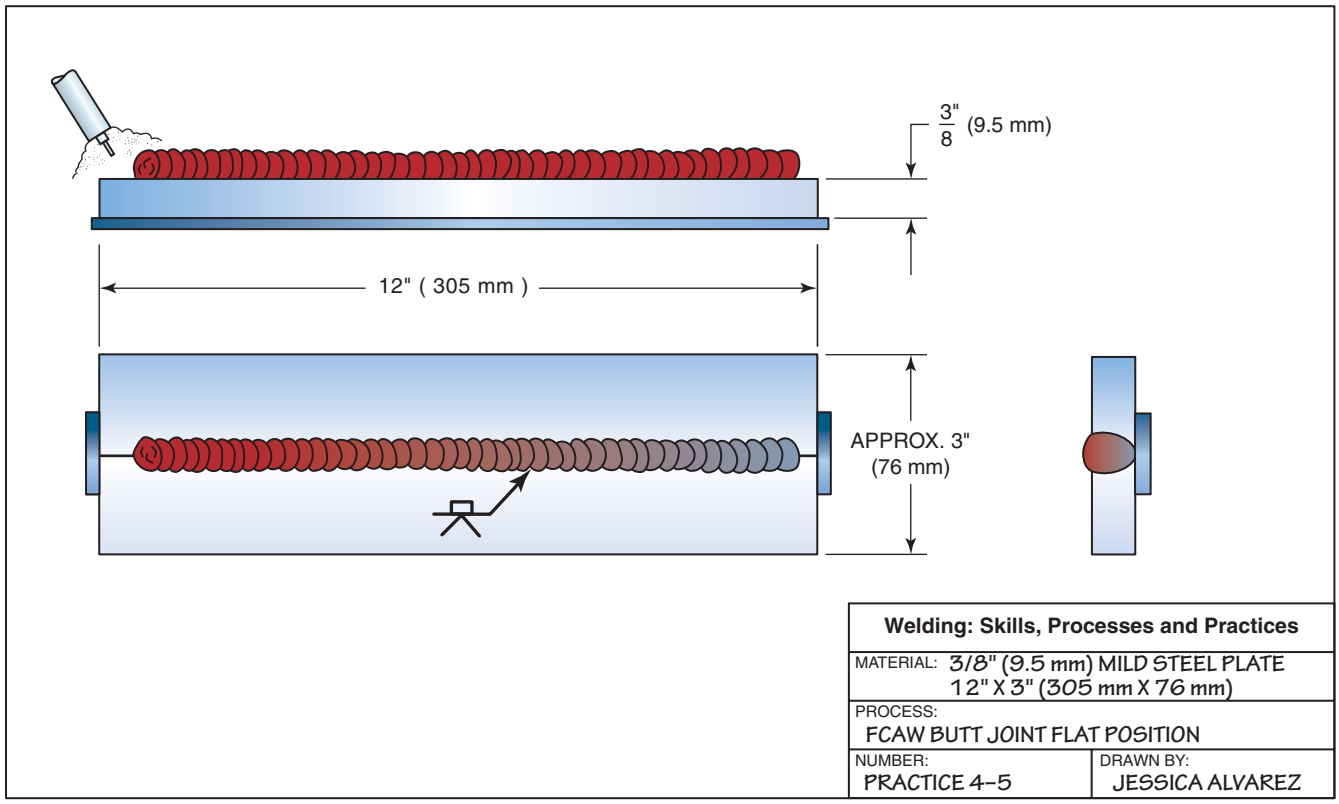

Figure 4.23 FCAW butt joint, 3/8 in. mild steel, flat position

Place the test plates in position at a comfortable height and location. Be sure that you have complete and free movement along the full length of the weld joint. It is often a good idea to make a practice pass along the joint with the welding gun without power to make sure nothing will interfere with your making the weld. Be sure the welding cable is free and will not get caught on anything during the weld.

Start the weld outside the groove on the backing strip tab, Figure 4.25. This is done so that the arc is smooth and the molten weld pool size is

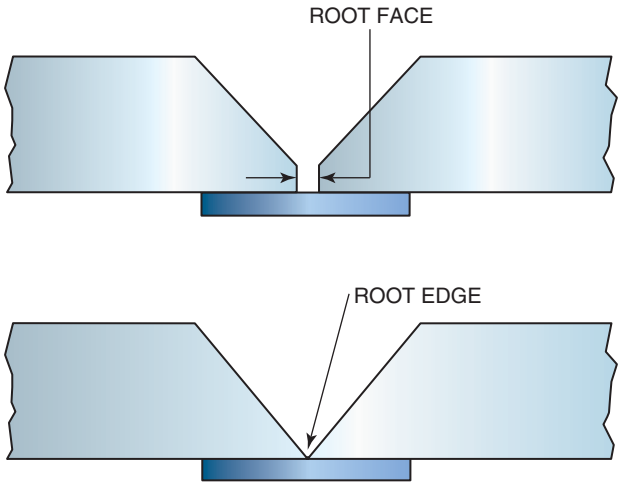

Figure 4.24 Groove layout with and without a root face

Figure 4.25 Using the ends of a backing strip
Run-off tabs, in which the weld starts and stops on the ends of the backing strip, help control possible underfill or burn-back at the starting and stopping points of a groove weld.
Source: Courtesy of Larry Jeffus

established at the beginning of the groove. Continue the weld out onto the tab at the outer end of the groove. This process ensures that the end of the groove is completely filled with weld.

Repeat with both classifications of electrodes until consistently defect-free welds can be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-6

Butt Joint 1G 100% to Be Tested

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Flat" position portion of the all-position requirement of 6 and 11.

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the flat position, Figure 4.26.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-5, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

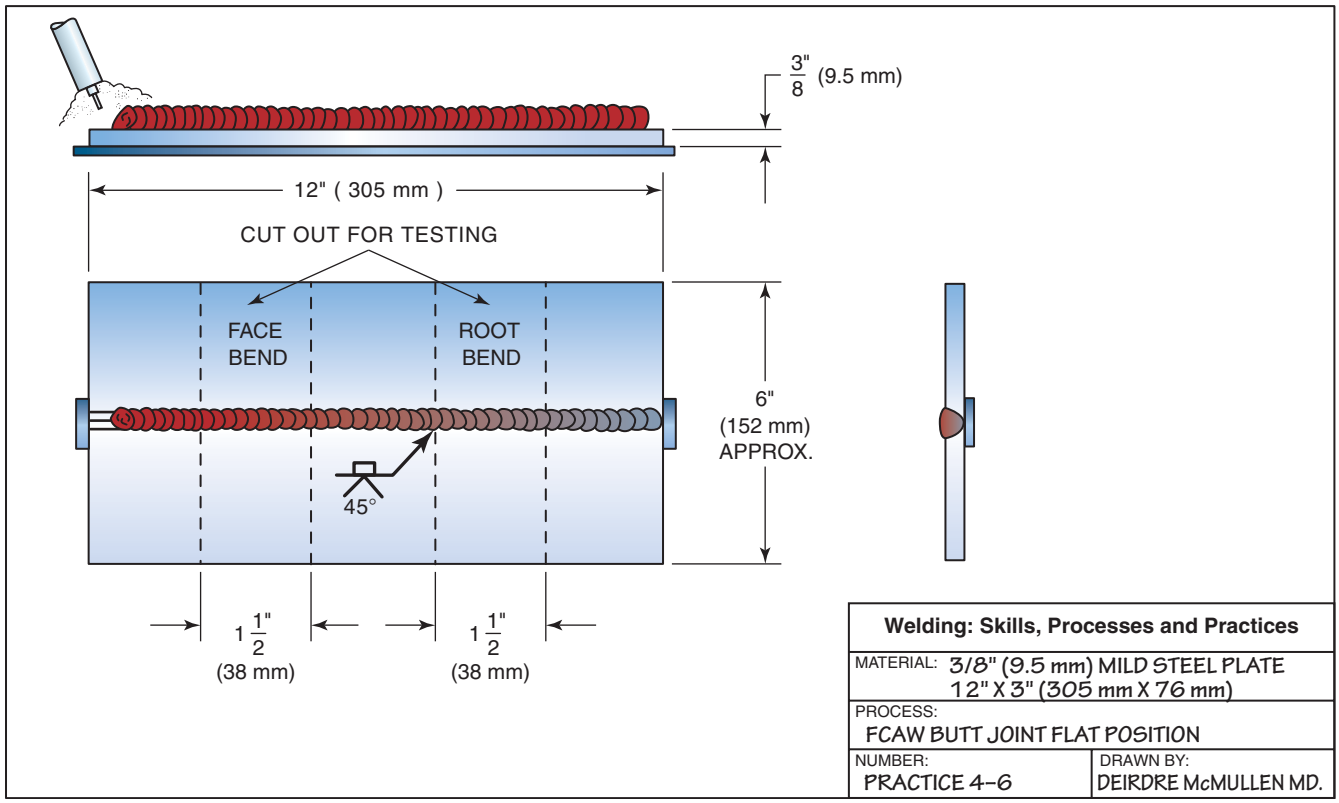

Figure 4.26 FCAW butt joint, 3/8 in. mild steel, flat position

FILLET WELDS

A fillet weld is the type of weld made on a **lap joint** and a **tee joint**. It should be built up equal to the thickness of the plate, Figure 4.27. On thick plates the fillet must be made up of several passes as with a groove weld. The difference with a fillet weld is that a smooth transition from the plate surface to the weld is required. If this transition is abrupt, it can cause stresses that will weaken the joint.

The lap joint is made by overlapping the edges of the plates. They should be held together tightly before tack welding them together. A small tack weld may be added in the center to prevent distortion during welding, Figure 4.28. Chip and wire brush the tacks before you start to weld.

The tee joint is made by tack welding one piece of metal on another piece of metal at a right angle, Figure 4.29. After the joint is tack welded together, the slag is chipped from the tack welds. If the slag is not removed, it will cause a slag inclusion in the final weld.

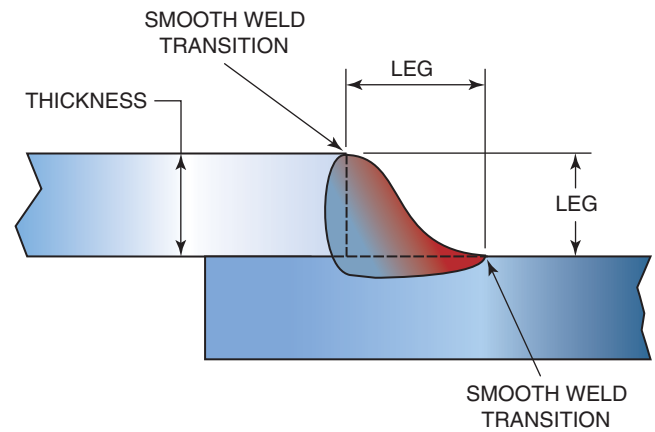

Figure 4.27 The legs of a fillet weld should generally be equal to the thickness of the base metal

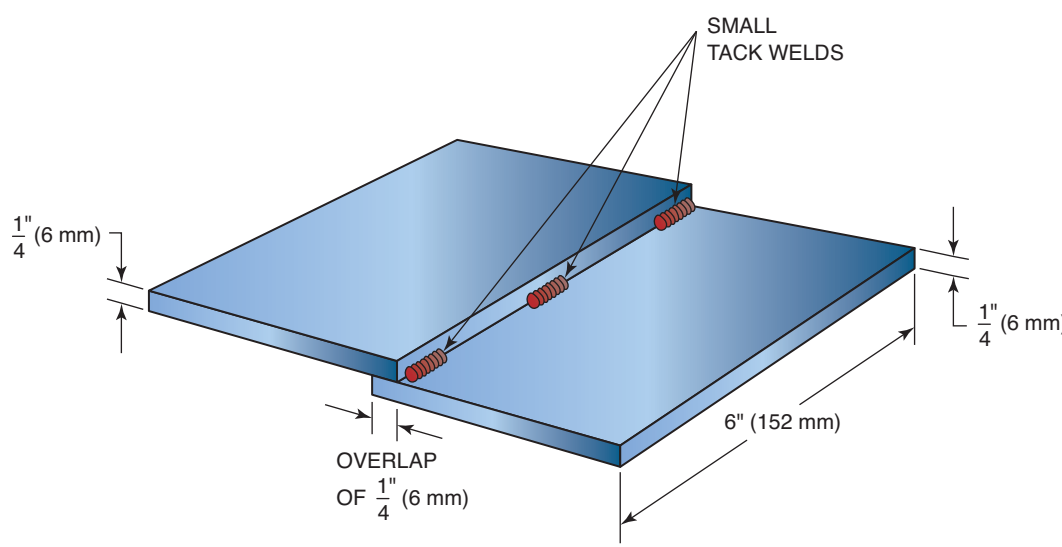

Figure 4.28 Tack welding the plates of a lap joint together

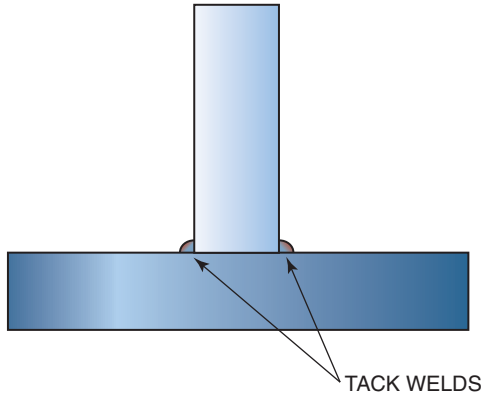

Figure 4.29 Tack welding both sides of a tee joint
Tack welding will help keep the tee square for welding.

Holding thick plates tightly together on tee joints may cause underbead cracking, or lamellar tearing, Figure 4.30. On thick plates the weld shrinkage can be great enough to pull the metal apart well below the bead or its heat-affected zone. In production welds, cracking can be controlled by not assembling the plates tightly together. The space between the two plates can be set by placing a small wire spacer between them, Figure 4.31.

A fillet welded lap or tee joint can be strong if it is welded on both sides, even without having deep penetration, Figure 4.32. Some tee joints may be prepared for welding by cutting either a bevel or a J-groove in the vertical plate. This cut is not required for strength but may be necessary because of design limitations. Unless otherwise specified, most fillet welds will be equal in size to the plates welded. A fillet weld will be as strong as the base plate if the size of the two welds equals the total thickness of the base plate. The weld bead should have a flat or slightly concave appearance to ensure the greatest strength and efficiency, Figure 4.33.

The root of fillet welds must be melted to ensure a completely fused joint. A notch along the root of the weld pool is an indication that the root is not being fused together, Figure 4.34. To achieve complete root fusion, move the arc to a point as close as possible to the leading edge of the weld pool, Figure 4.35. If the arc strikes the unmelted plate ahead of the molten weld pool, it may become erratic, which will increase weld spatter.

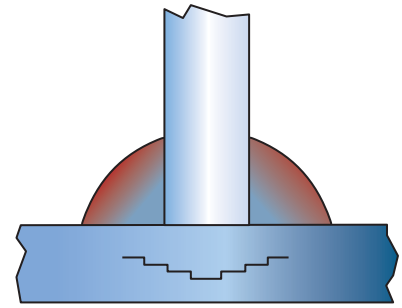

Figure 4.30 Underbead cracking, or lamellar tearing, of the base plate

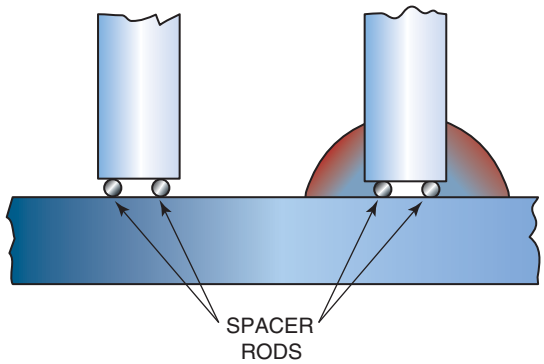

Figure 4.31 Spacers in tee joints
Base plate cracking can be controlled by placing spacers in the joint before welding.

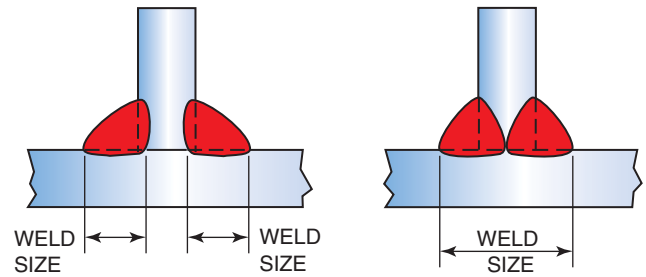

Figure 4.32 Fillet weld size
If the total weld sizes are equal, then both tee joints would have equal strength.

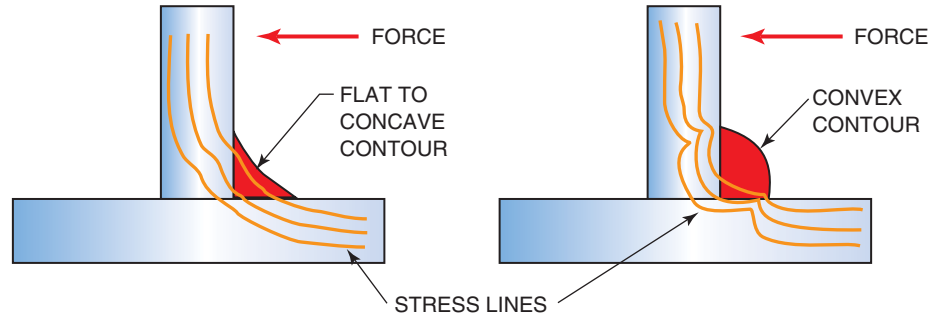

Figure 4.33 Fillet weld shape
The stresses are distributed more uniformly through a flat or concave fillet weld.

Figure 4.34 Watch the root of the weld bead to be sure there is complete fusion
Source: Courtesy of Larry Jeffus

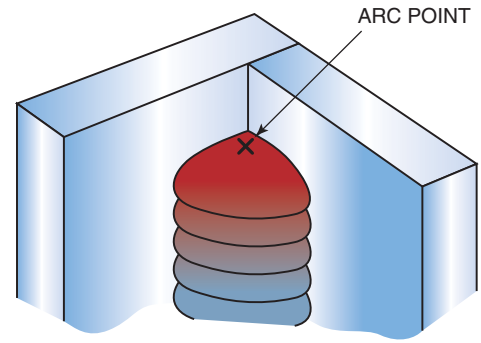

Figure 4.35 Moving the arc as close as possible to the leading edge of the weld will provide good root fusion.

PRACTICE 4-7

Lap Joint and Tee Joint 1F

Module 1
Key Indicator 1, 2, 3, 4
Module 2
Key Indicator 1, 2, 3, 4, 7
Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Flat" position portion of the all-position requirement for 5 and 10.

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); and one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick. You will make a fillet weld in the flat position.

Tack weld the pieces of metal together and brace them in position. When making the lap or tee joints in the flat position, the plates must be at a 45° angle so that the surface of the weld will be flat, Figure 4.36A and Figure 4.36B. Starting at one end, make a weld along the entire length of the joint.

Repeat each type of joint with both classifications of electrodes until consistently defect-free welds can be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-8

Tee Joint 1F

Module 1
Key Indicator 1, 2, 3, 4
Module 2
Key Indicator 1, 2, 3, 4, 7
Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Use a properly set up and adjusted FCA welding machine, Table 4.2; proper safety protection; E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one

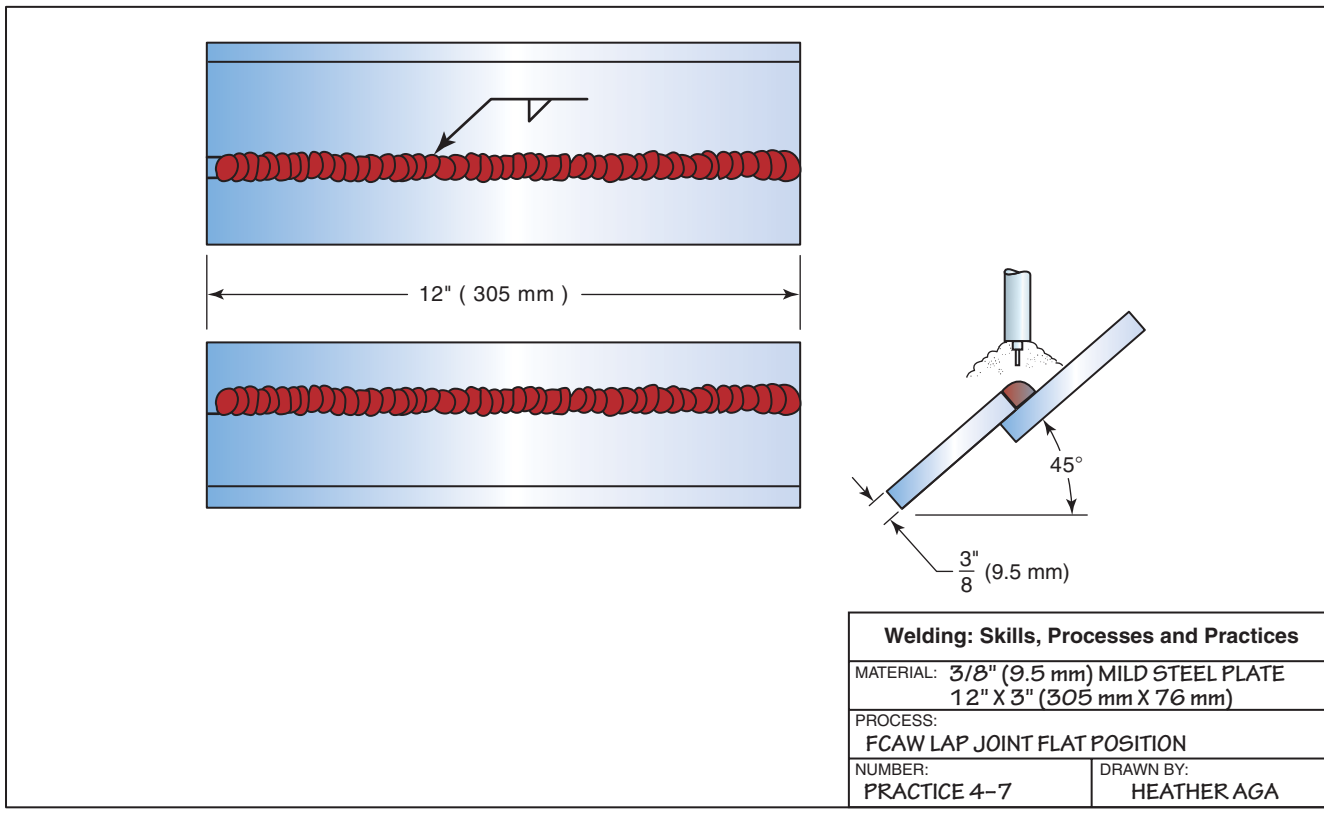

Figure 4.36 (A) FCAW lap joint, 3/8 in. mild steel, flat position

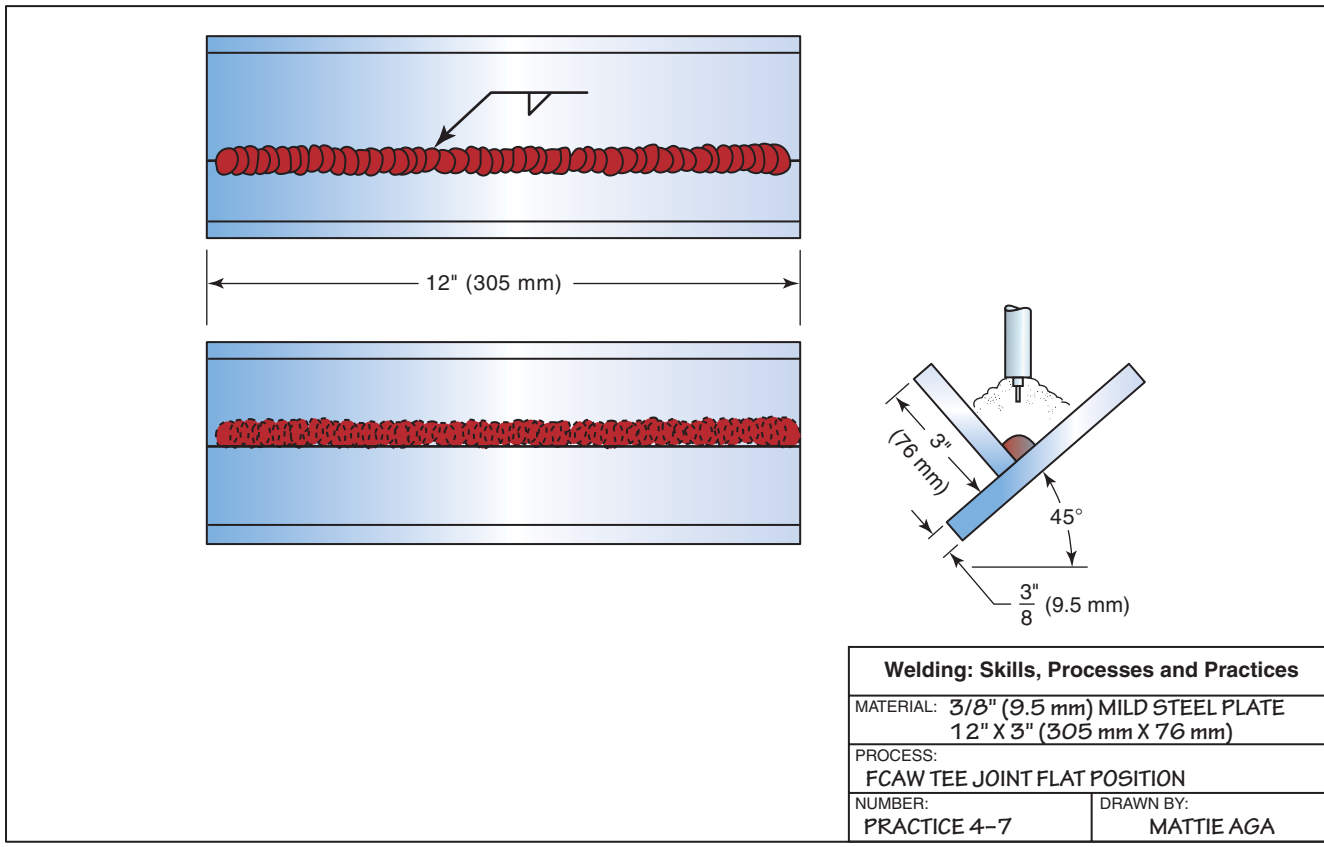

Figure 4.36 (B) FCAW tee joint, 3/8 in. mild steel, flat position

Table 4.2 FCA Welding Parameters for Use if Specific Settings Are Unavailable from Electrode Manufacturer (base metal thickness 1/2 to 3/4 inch)

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70T-1 E71T-1	0.035 in. (0.9 mm)	130 to 150	288 to 380 (732 to 975)	22 to 25	None	n/a	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-1 E71T-1	0.045 in. (1.2 mm)	150 to 210	200 to 300 (508 to 762)	28 to 29	None	n/a	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-1 E71T-1	.052 in. (1.4 mm)	150 to 300	150 to 350 (381 to 889)	25 to 33	None	n/a	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-1 E71T-1	1/16 in. (1.6 mm)	200 to 400	150 to 300 (381 to 762)	27 to 33	None	n/a	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-5 E71T-11	0.035 in. (0.9 mm)	130 to 200	288 to 576 (732 to 1463)	20 to 28	75% argon 25% CO_2	30 cfh	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-5 E71T-11	0.045 in. (1.2 mm)	150 to 250	200 to 400 (508 to 1016)	23 to 29	75% argon 25% CO_2	35 cfh	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-5 E71T-11	0.052 in. (1.4 mm)	150 to 300	150 to 350 (381 to 889)	21 to 32	75% argon 25% CO_2	35 cfh	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-5 E71T-11	1/16 in. (1.6 mm)	180 to 400	145 to 350 (368 to 889)	21 to 34	75% argon 25% CO_2	40 cfh	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)

or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker. You will make a fillet weld in the flat position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-7, repeat each type of joint with both classifications of electrodes until consistently defect-free welds can be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

VERTICAL WELDS

PRACTICE 4-9

Butt Joint at a 45° Vertical Up Angle

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 1, 2
Gas Shielded
Key Indicator 3
Gas Shielded
Key Indicator 4
Self Shielded
Key Indicator 8
Self Shielded
Key Indicator 9

Using a properly set up and adjusted FCA welding machine, proper safety protection, 0.035-in. and/or 0.045-in. (0.9-mm and/or 1.2-mm)-diameter E71T-1 and/or E71T-11 electrodes, and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick or thinner, you will increase the plate angle gradually as you develop skill until you are making satisfactory welds in the vertical up position, Figure 4.37.

- Start practicing this weld with the plate at a 45° angle.
- Gradually increase the angle of the plate to vertical as skill is gained in welding this joint. A straight stringer bead or slight zigzag will work well on this joint.
- Establish a molten weld pool in the root of the joint.
- Cool, chip, and inspect the weld for uniformity and defects.

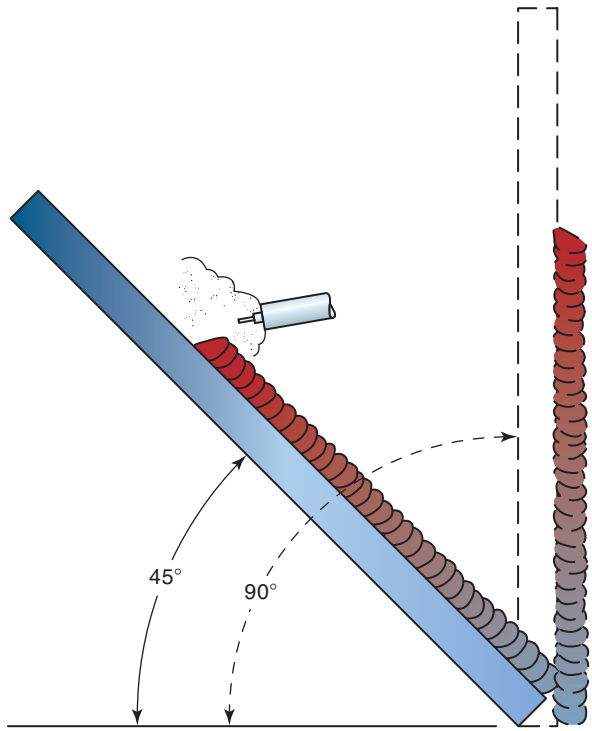

Figure 4.37 Start making welds with the plate at a 45° angle
As your skill develops, increase the angle until the plate is vertical.

It is easier to make a quality weld in the vertical up position if lower settings are used in both the **amperage range** and **voltage range**. This will make the molten weld pool smaller, less fluid, and easier to control. A problem with lower power settings is that the weld bead often can be very convex, Figure 4.38. Faster travel speed and/or slightly wider weave patterns can be used to control the bead shape.

Start at the bottom of the plate and hold the welding gun at a slight upward angle to the plate, Figure 4.39. Brace yourself, lower your hood, and begin to weld. Depending on the machine settings and type of electrode used, you will make a weave pattern.

If the molten weld pool is large and fluid (hot), use a C or J weave pattern to allow a longer time for the molten weld pool to cool, Figure 4.40. Do not make the weave so long or fast that the electrode is allowed to strike the metal ahead of the molten weld pool. If this happens, spatter increases and a spot or zone of incomplete fusion may occur.

A weld that is high and has little or no fusion is too "cold." Changing the welding technique will not correct this problem. The welder must stop welding and make the needed adjustments to the power supply or electrode feeder. Continue to weld along the entire 12-in. (305-mm) length of plate.

Repeat welds with both electrodes until defect-free welds can be consistently made vertically in the 1/4-in. (6-mm)-thick plate. Turn off the

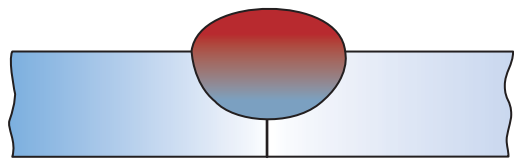

Figure 4.38 Low amperage causes too much buildup and not enough penetration

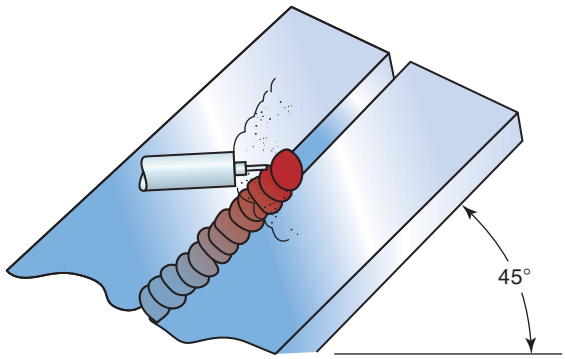

Figure 4.39 45° vertical up

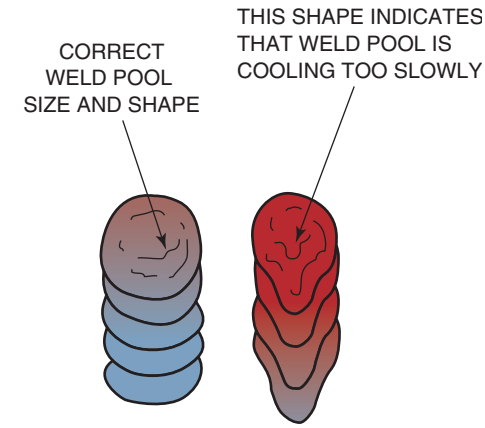

Figure 4.40 The shape of the weld pool can indicate the temperature of the surrounding base metal

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Vertical" portion of the all-position requirement.

welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-10

Butt Joint 3G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild

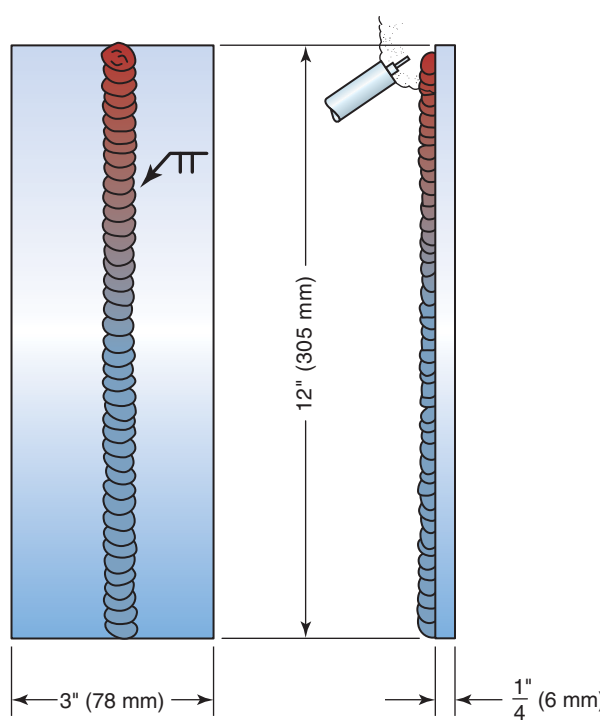

Welding: Skills, Processes and Practices	
MATERIAL: 1/4" (6 mm) MILD STEEL PLATE 12" X 3" (305 mm X 76 mm)	
PROCESS: FCAW 3G BUTT JOINT	
NUMBER: PRACTICE 4-10	DRAWN BY: AMY JEFFUS

Figure 4.41 FCAW 3G butt joint, 1/4 in. mild steel

steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick or thinner, you will make a groove weld in the vertical position, Figure 4.41.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-9, repeat with both classifications of electrodes until defect-free welds can be consistently made in the 1/4-in. (6-mm)-thick plate. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-11

Butt Joint 3G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the vertical position.

Following the same instructions for assembly and welding procedure as outlined in Practice 4-9, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

PRACTICE 4-12

Butt Joint 3G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the vertical position, Figure 4.42.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-9, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Vertical" component of the all-position requirement of 6 and 11.

PRACTICE 4-13

Butt Joint 3G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Vertical" portion of the all-position requirement of 6 and 11.

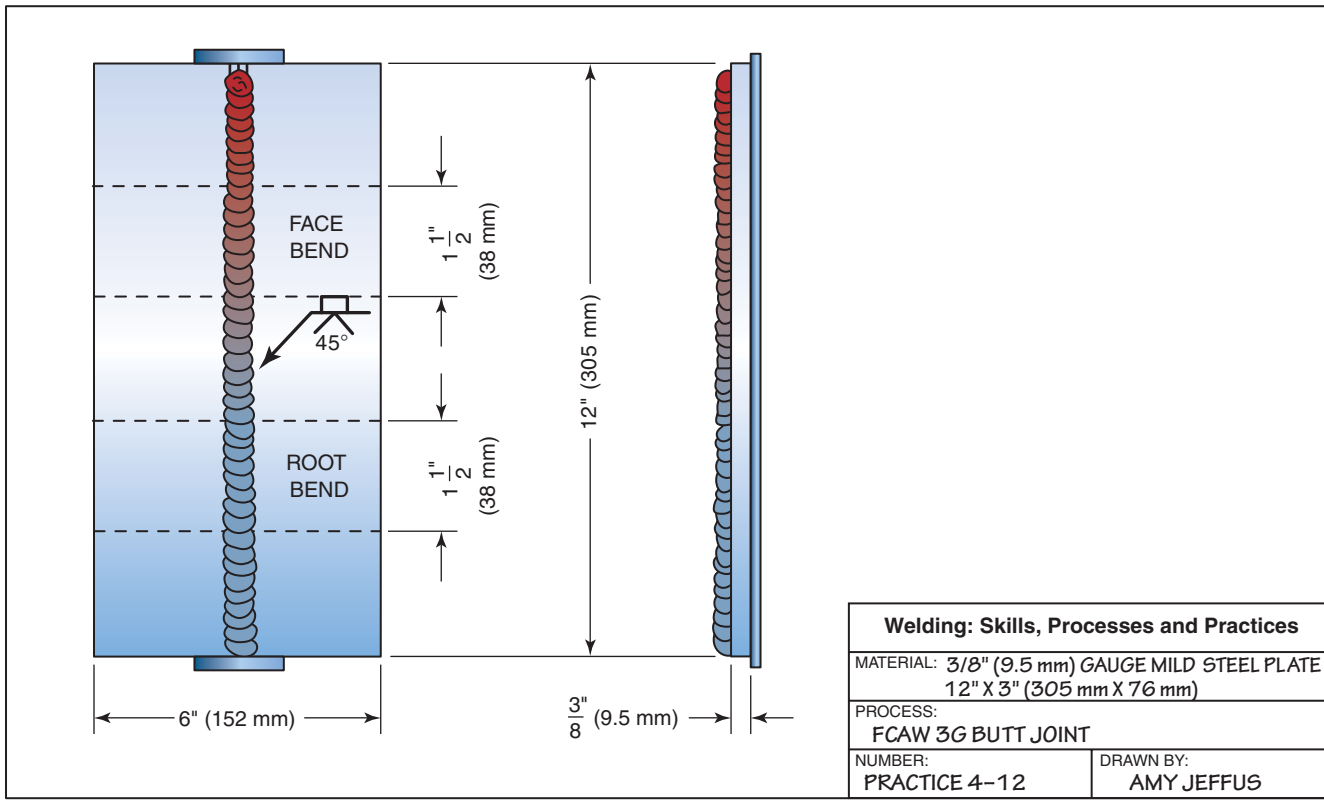

Figure 4.42 FCAW 3G butt joint, 3/8 in. mild steel

3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the vertical position, Figure 4.43.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Following the same instructions for the assembly and welding procedure outlined in Practice 4-9, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-14

Fillet Weld Joint at a 45° Vertical Up Angle

Module 6
Key Indicator 1, 2
Gas Shielded
Key Indicator 3
Gas Shielded
Key Indicator 4
Self Shielded
Key Indicator 8
Self Shielded
Key Indicator 9
Self Shielded
Key Indicator 10
Self Shielded
Key Indicator 11
Self Shielded
Key Indicator 12

Module 9
Key Indicator 1, 2

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick, you will increase the plate angle gradually as you develop skill until you are making satisfactory welds in the vertical up position, Figure 4.44.

Tack weld the metal pieces together and brace them in position. Check to see that you have free movement for your gun along the entire joint to prevent stopping and restarting during the weld. Avoiding stops and starts both speeds up the welding time and eliminates discontinuities.

It is easier to make a quality weld in the vertical up position if both the amperage and voltage are set at the lower end of their ranges. This

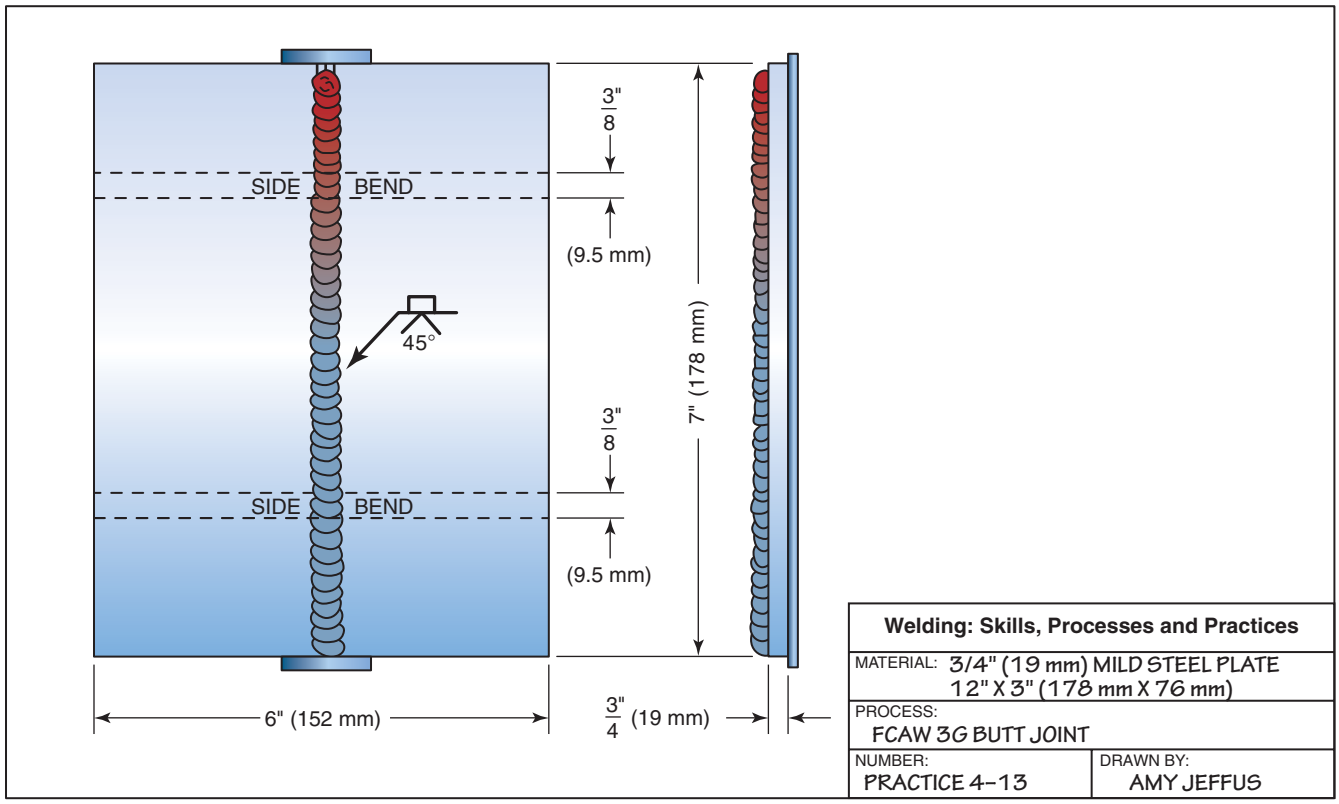

Figure 4.43 FCAW 3G butt joint, 3/4 in. mild steel

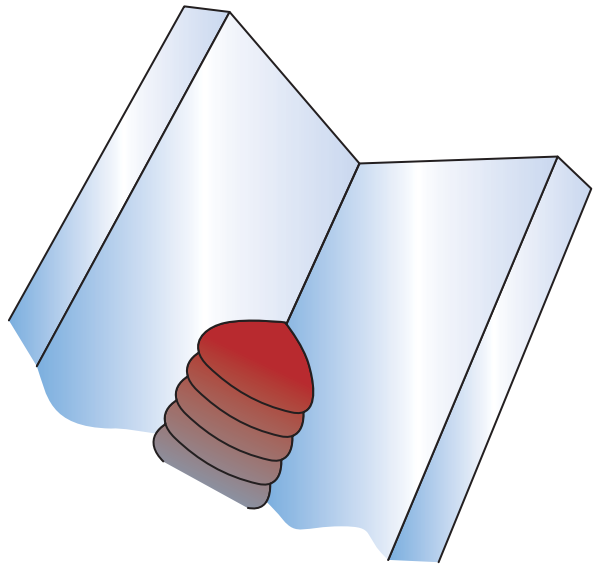

Figure 4.44 45° vertical up fillet weld

will make the molten weld pool smaller, less fluid, and easier to control. A problem with the lower power settings is that the weld bead often is very convex. A convex face on a weld bead often makes it more difficult to remove the slag along the toe of the weld.

The weave pattern should allow for adequate fusion on both edges of the joint. Watch the edges to be sure that they are being melted so that adequate fusion and penetration occur.

Repeat the weld with each electrode type until defect-free welds can consistently be made vertically. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2
This practice addresses the "Vertical" position component of the all-position requirement.

PRACTICE 4-15

Lap Joint and Tee Joint 3F 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); and one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick. You will make a fillet weld in the vertical position.

Following the same instructions for assembly and welding procedure as outlined in Practice 4-14, repeat each type of joint with both classifications of electrodes until you can make welds with 100% penetration that will pass the test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Vertical" component of the all-position requirement for 5 and 10.

PRACTICE 4-16

Tee Joint 3F

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm), and one or more pieces of mild steel plate, 7 in. (178 mm) long and 3/4 in. (19 mm) thick, you will make a fillet weld in the vertical position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-14, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Horizontal" component of the all-position requirement for 5 and 10.

HORIZONTAL WELDS

PRACTICE 4-17

Lap Joint and Tee Joint 2F

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 (or E71T-1 and/or E70T-5) electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); and one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick. You will make a fillet weld in the horizontal position, Figure 4.45A and Figure 4.45B.

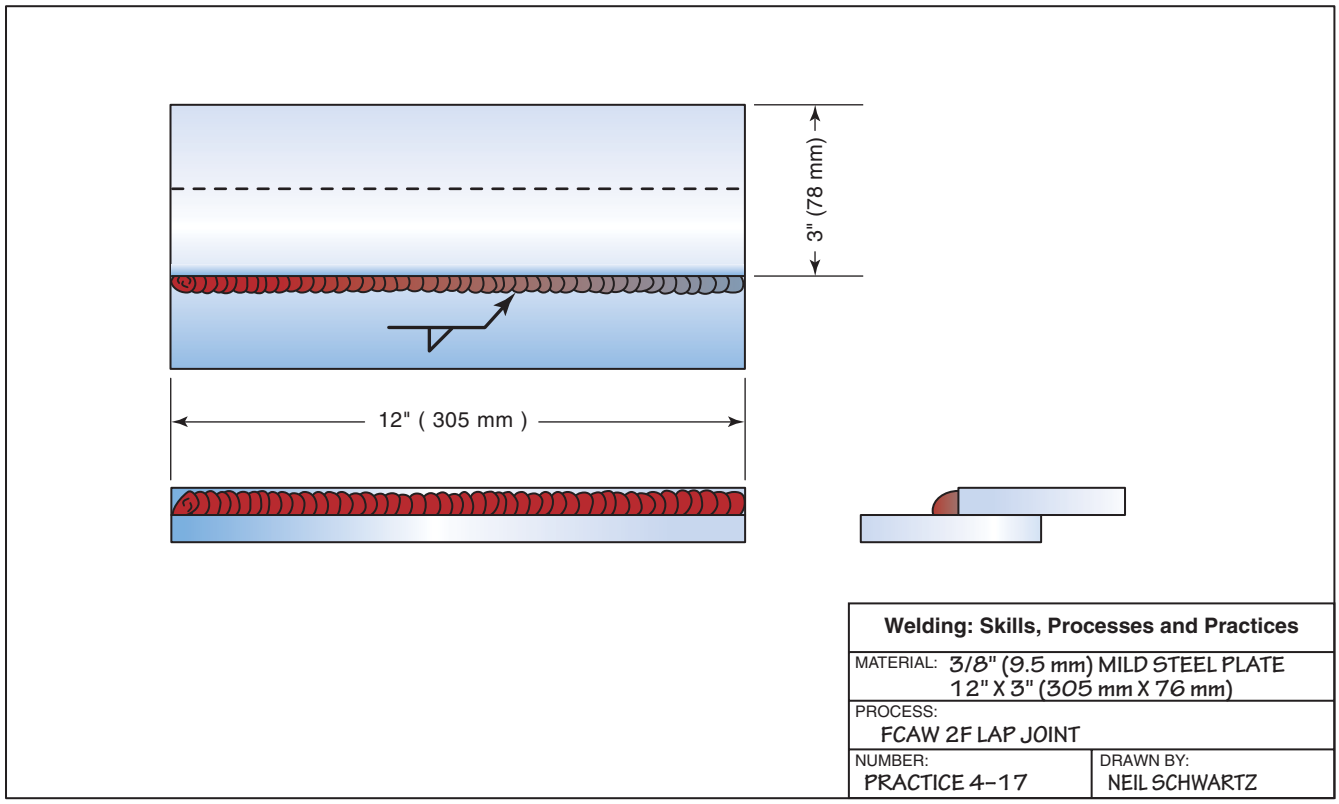

Figure 4.45 (A) FCAW 2F lap joint, 3/8 in. mild steel

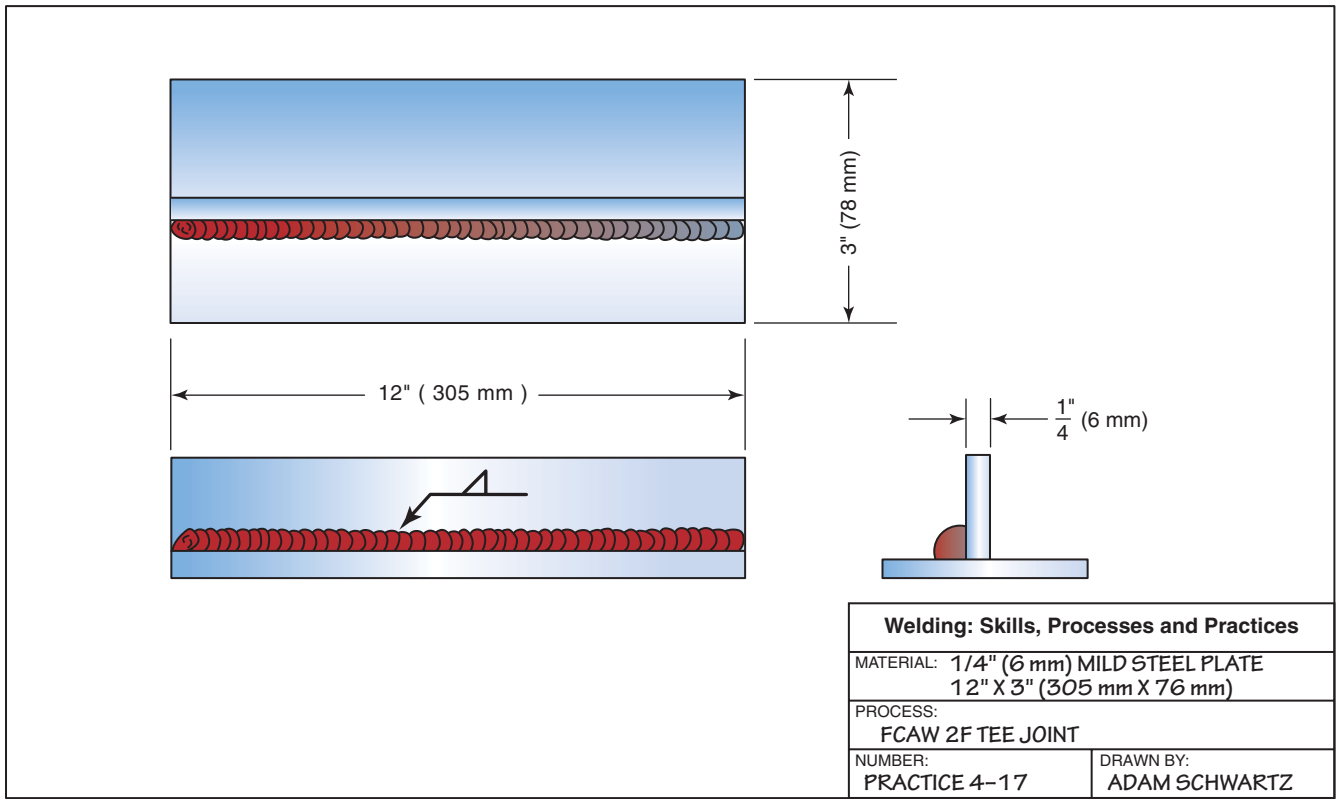

Figure 4.45 (B) FCAW 2F tee joint, 1/4 in. mild steel

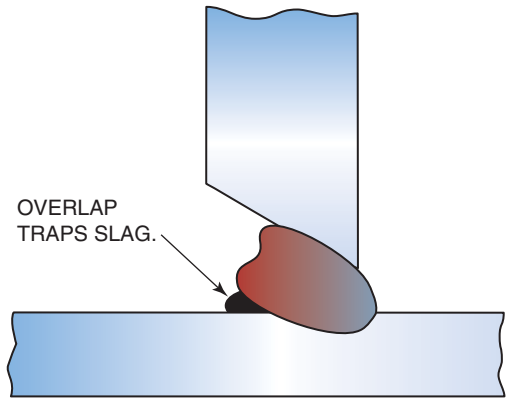

Figure 4.46 Slag can be trapped along the side of the root pass

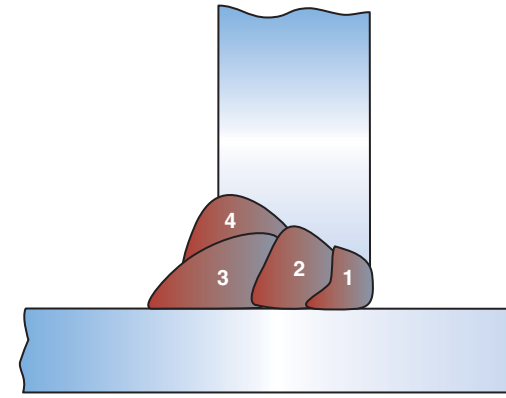

Figure 4.47 FCAW weld bead positions for a 100% penetration grooved tee joint

The root weld must be kept small so that its contour can be controlled. Too large a root pass can trap slag under overlap along the lower edge of the weld, Figure 4.46. Clean each pass thoroughly before the weld bead is started. Follow the weld bead sequence shown in Figure 4.47. Use stringer beads rather than weave beads in order to eliminate overlap. Keeping all of the weld beads small will help control their contour.

Repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Horizontal" component of the all-position requirement for 5 and 10.

PRACTICE 4-18

Tee Joint 2F

Use a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 (or E71T-1 and/or E70T-5) electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm), and one or more pieces of mild steel, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker. You will make a fillet weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-17, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-19

Butt Joint 2G

Use a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 (or E71T-1 and/or E70T-5 electrodes) of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm),

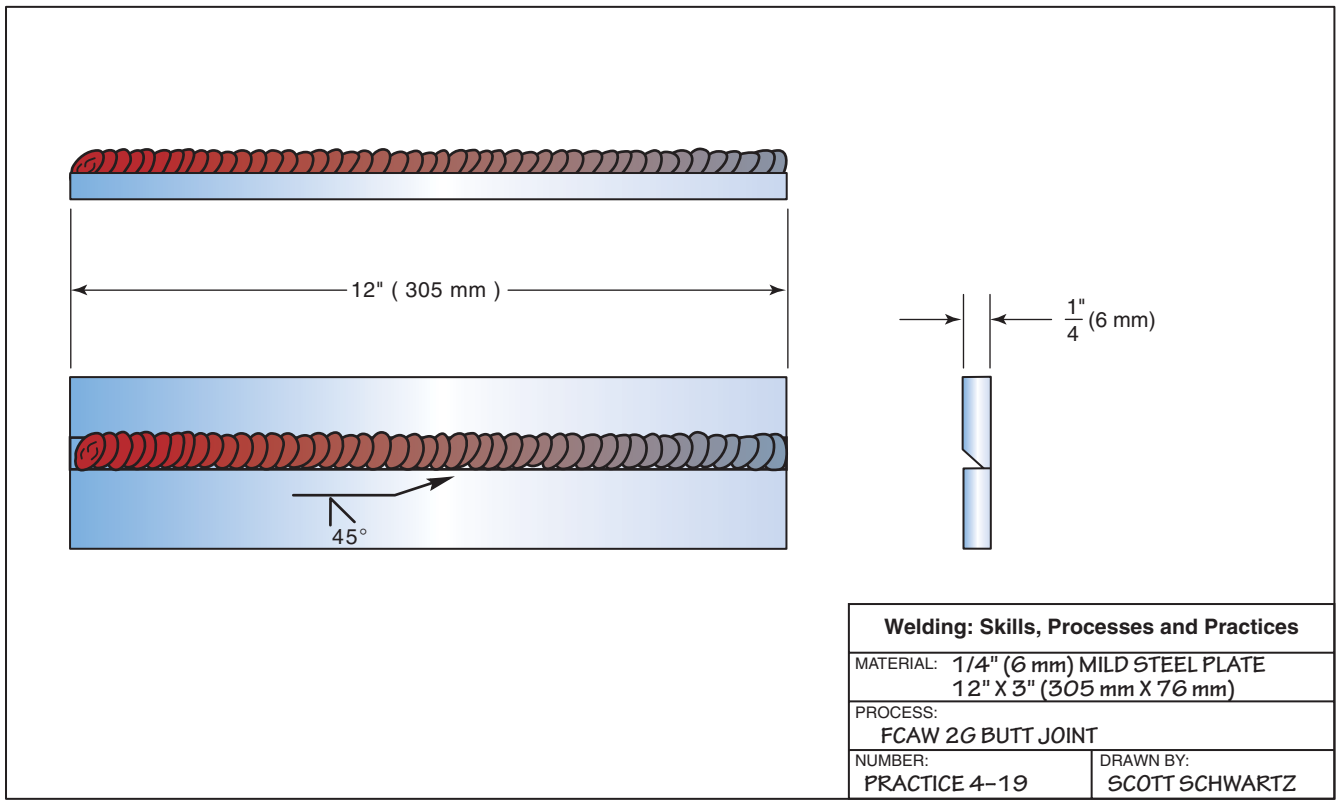

Figure 4.48 FCAW 2G butt joint, 1/4 in. mild steel

and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick or thinner. You will make a groove weld in the horizontal position, Figure 4.48.

Repeat with both classifications of electrodes until defect-free welds can consistently be made in the 1/4-in. (6-mm)-thick plate. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-20

Butt Joint 2G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 or 5 (or E71T-1 and/or E70T-5) electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the horizontal position.

Repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-21

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

Butt Joint 2G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 or 5 (or E71T-1 and/or E70T-5) electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide and 1/4 in. (6 mm) thick. You will make a groove weld in the horizontal position.

Repeat the weld using each electrode classification until you can make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-22

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Self Shielded
Key Indicator 6
Gas Shielded
Key Indicator 11
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

Butt Joint 2G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 electrodes (or E71T-1 and/or E70T-5) of diameter 0.035 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the horizontal position, Figure 4.49.

Repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-23

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

Butt Joint 2G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 electrodes (or E71T-1 and/or E70T-5) of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the horizontal position.

Repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

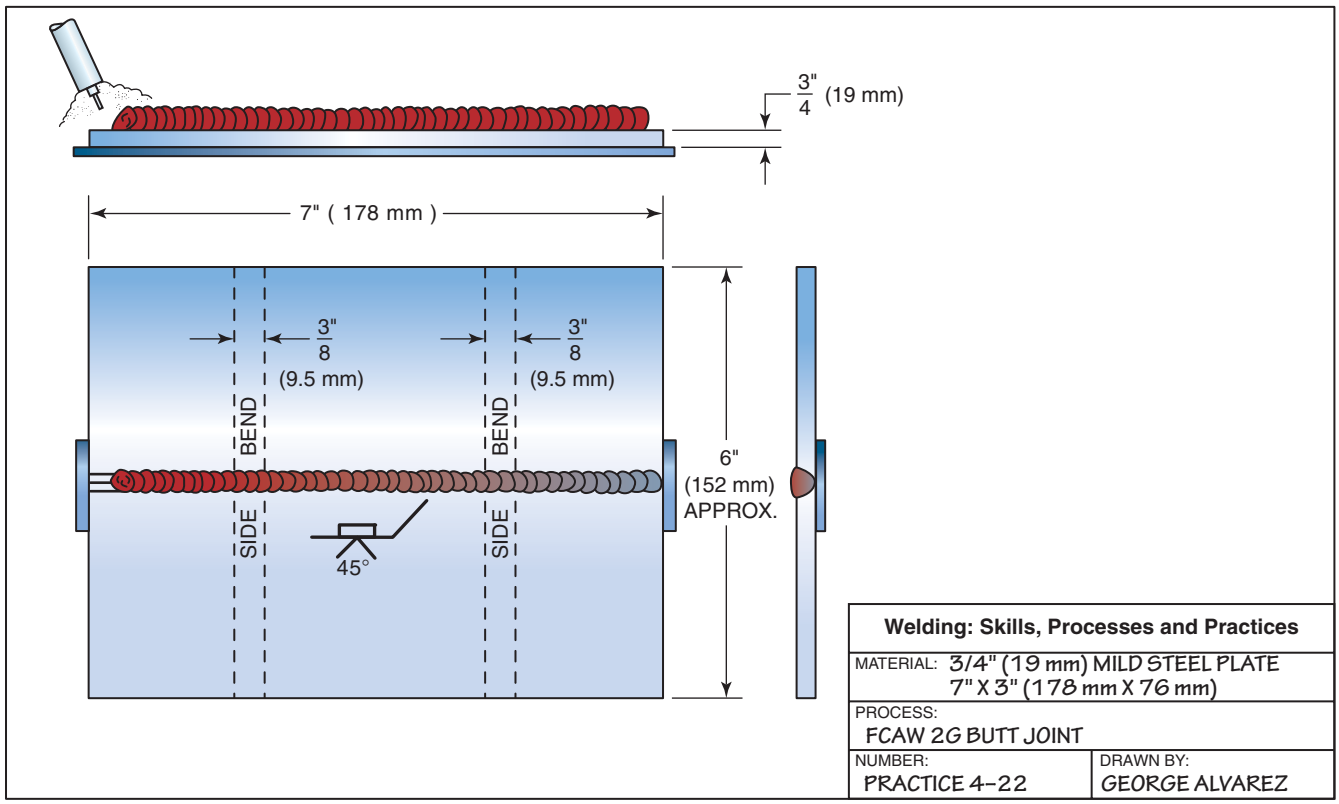

Figure 4.49 FCAW 2G butt joint, 3/4 in. mild steel

OVERHEAD-POSITION WELDS

PRACTICE 4-24

Butt Joint 4G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick or thinner, you will make a groove weld in the overhead position.

The molten weld pool should be kept as small as possible for easier control. A small molten weld pool can be achieved by using lower current, faster traveling speeds, and settings.

Lower current settings require closer control of gun manipulation to ensure that the electrode is fed into the molten weld pool just behind the leading edge. The low power will cause overlap and more spatter if this electrode-to-molten weld pool contact position is not closely maintained.

Faster travel speeds allow the welder to maintain a high production rate even if multiple passes are required to complete the weld. Weld penetration into the base metal at the start of the bead can be obtained by using a slow start or quickly reversing the weld direction. Both the slow start and reversal of weld direction put more heat into the weld start to increase penetration. The higher speed also reduces the amount of weld distortion by reducing the amount of time that heat is applied to a joint.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

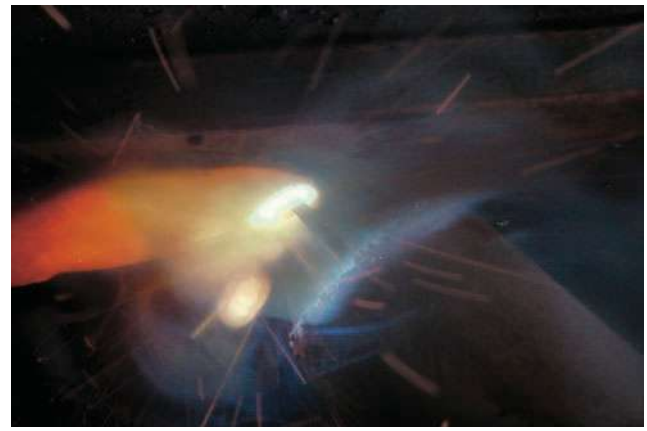

Figure 4.50 **Hold the gun so that weld spatter will not fall onto the gun**
Source: Courtesy of Larry Jeffus

For overhead welding, extra personal protection is required to reduce the danger of burns. Leather sleeves or leather jackets should be worn.

Much of the spatter created during overhead welding falls into or on the nozzle and contact tube. The contact tube may short out to the gas nozzle. The shorted gas nozzle may arc to the work, causing damage both to the nozzle and to the plate. To control the amount of spatter, a longer stickout and/or a sharper gun-to-plate angle is required to allow most of the spatter to fall clear of the gun or nozzle, Figure 4.50.

Make several short weld beads using various techniques to establish the method that is most successful and most comfortable for you. After each weld, stop and evaluate it before making a change. When you have decided on the technique to be used, make a welded stringer bead that is 12 in. (305 mm) long.

Repeat with both classifications of electrodes until defect-free welds can consistently be made in the 1/4-in. (6-mm)-thick plate. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-25

Butt Joint 4G 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick, you will make a groove weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-26

Butt Joint 4G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the overhead position, Figure 4.51.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-27

Butt Joint 4G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the overhead position.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

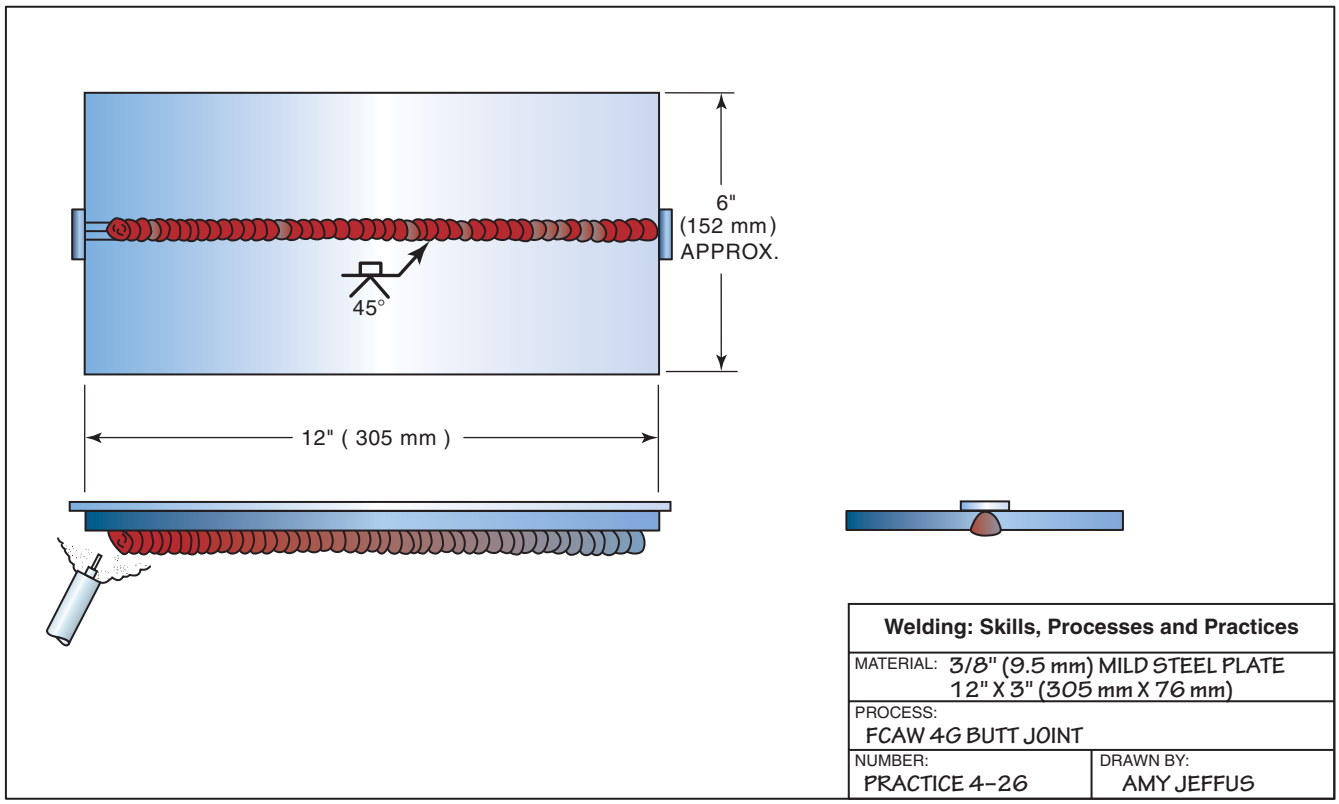

Figure 4.51 FCAW 4G butt joint, 3/8 in. mild steel

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-28

Butt Joint 4G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-29

Butt Joint 4G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Overhead" component of the all-position requirement for 5 and 10.

PRACTICE 4-30

Lap Joint and Tee Joint 4F

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); and one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/4 in. (19 mm) thick. You will make a fillet weld in the overhead position, Figure 4.52A and Figure 4.52B.

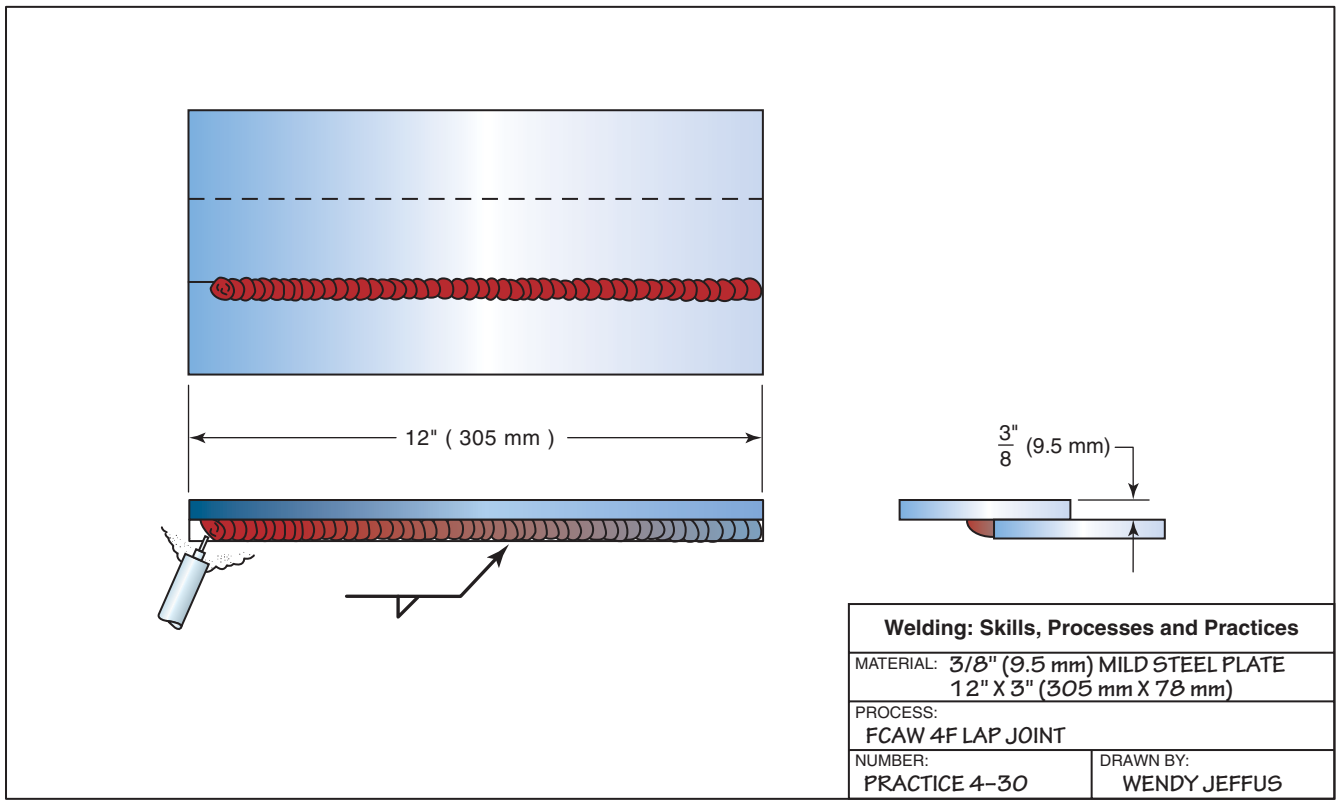

Figure 4.52 (A) FCAW 4F lap joint, 3/8 in. mild steel

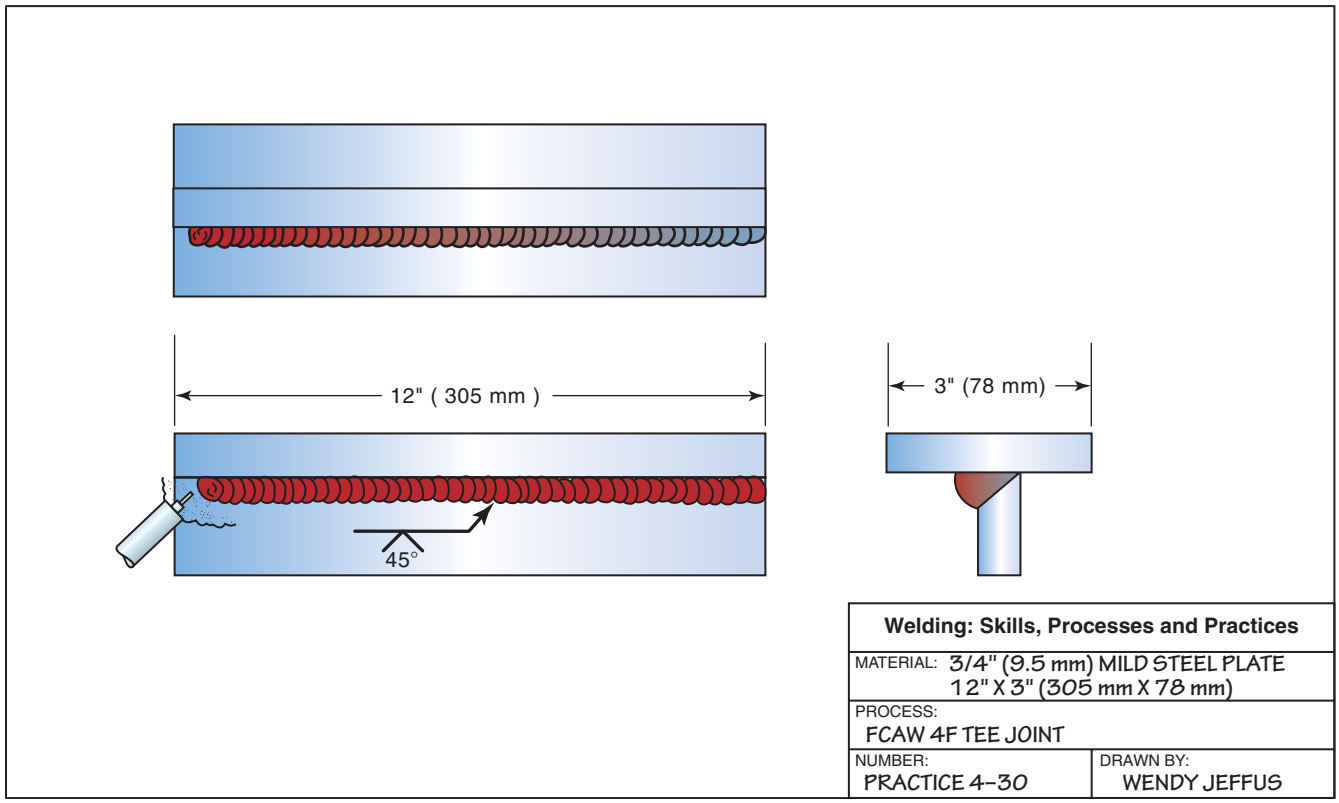

Figure 4.52 (B) FCAW 4F tee joint, 3/4 in. mild steel

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-31

Tee Joint 4F 100% to Be Tested

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" component of the all-position requirement for 5 and 10.

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9-mm and/or through 1.6-mm); and one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker. You will make a fillet weld in the overhead position.

Following the same instructions for the assembly and welding procedure as outlined in Practice 4-24, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass the bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

THIN-GAUGE WELDING

The introduction of small electrode diameters has allowed FCA welding to be used on thin sheet metal. Usually these welds will be a fillet type, the easiest weld to make on thin stock. An effort should be made when possible to design the weld so it is not a butt-type joint. A common use for FCA welding on thin stock is to join it to a thicker member, Figure 4.53. This type of weld is used to put panels in frames.

The following practices include some butt-type joints. You will find that the vertical down welds are the easiest ones to make. If it is possible to position the weldment for a vertical down weld, production speeds for butt joints can be increased.

PRACTICE 4-32

Butt Joint 1G

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Use a properly set up and adjusted FCA welding machine, Table 4.3; proper safety protection; E71T-1 and/or E70T-5 or E70T-1 and/or E71T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm); and one or more pieces of mild steel sheet, 12 in. (305 mm) long

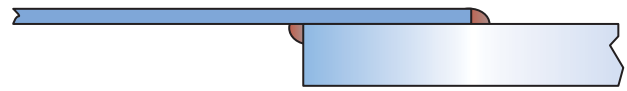

Figure 4.53 FCA welding of thin to thick metal

Table 4.3 FCA Welding Parameters for Use if Specific Settings Are Unavailable from Electrode Manufacturer

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thick
E70T-1 E71T-1	0.030 in. (0.8 mm)	40 to 145	90 to 340 (228 to 864)	20 to 27	None	n/a	Low-carbon steel	16 gauge to 18 gauge
E70T-1 E71T-1	0.035 in. (0.9 mm)	130 to 200	288 to 576 (732 to 1463)	20 to 28	None	n/a	Low-carbon steel	16 gauge to 18 gauge
E70T-5 E71T-11	0.035 in. (0.9 mm)	90 to 200	190 to 576 (483 to 1463)	16 to 29	57% argon 25% CO_2	35 cfh	Low-carbon steel	16 gauge 18 gauge

and 16-gauge to 18-gauge thick. You will make a butt weld in the flat position, Figure 4.54.

Do not leave a root opening for these welds. Even the slightest opening will result in a burn-through. If a burn-through occurs, the welder can be pulsed off and on so that the hole can be filled. This process will leave a larger than usual buildup. Excessive buildup could be ground off if necessary as part of the postweld cleanup.

Repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I provided by your instructor.

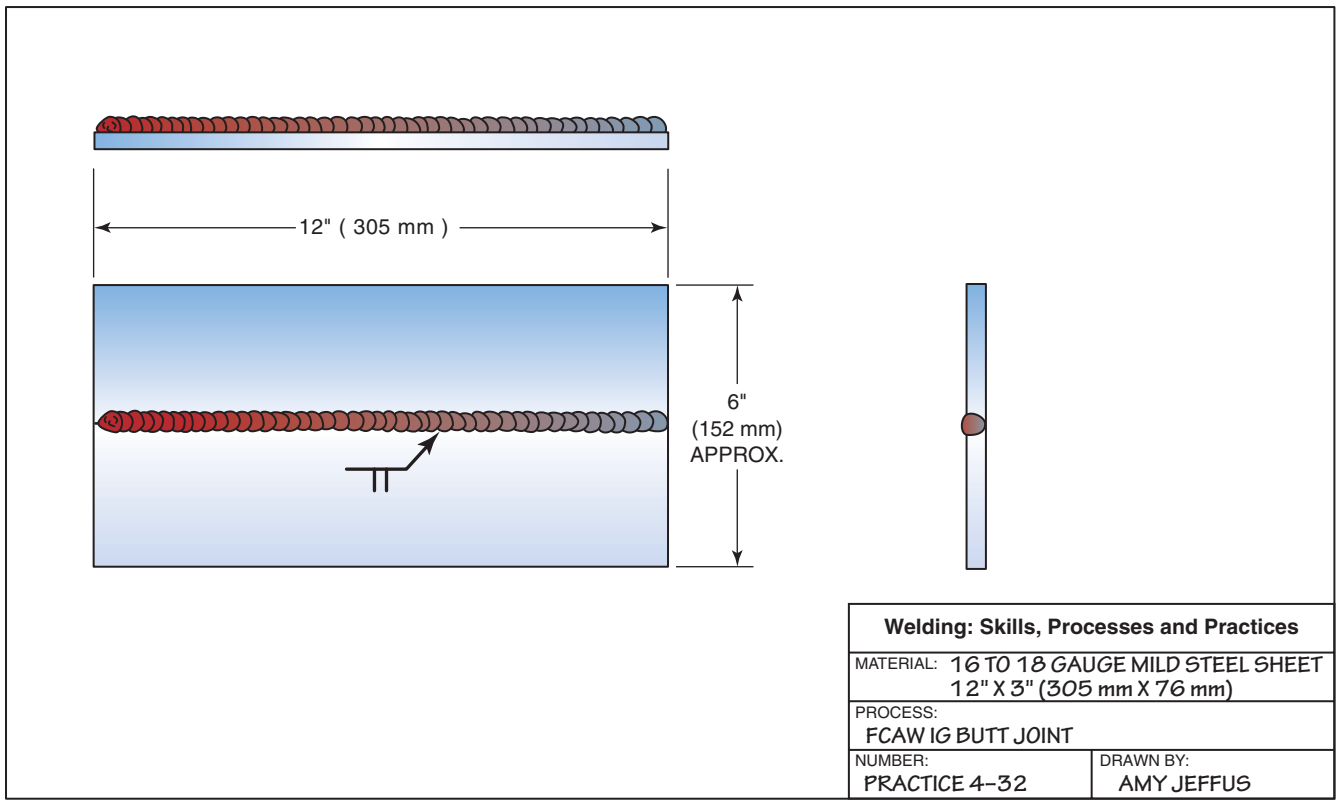

Figure 4.54 FCAW 1G butt joint, 16- to 18-gauge mild steel

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2

PRACTICE 4-33

Butt Joint 1G 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E70T-5 or E70T-1 and/or E70T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the flat position, Figure 4.55.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Gas Shielded
Key Indicator 10
This practice addresses the "Flat" component of the all-position requirement for 5 and 10.

PRACTICE 4-34

Lap Joint and Tee Joint 1F

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030-in. and/or 0.035-in. (0.8-mm and/or 0.9-mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a fillet weld in the flat position.

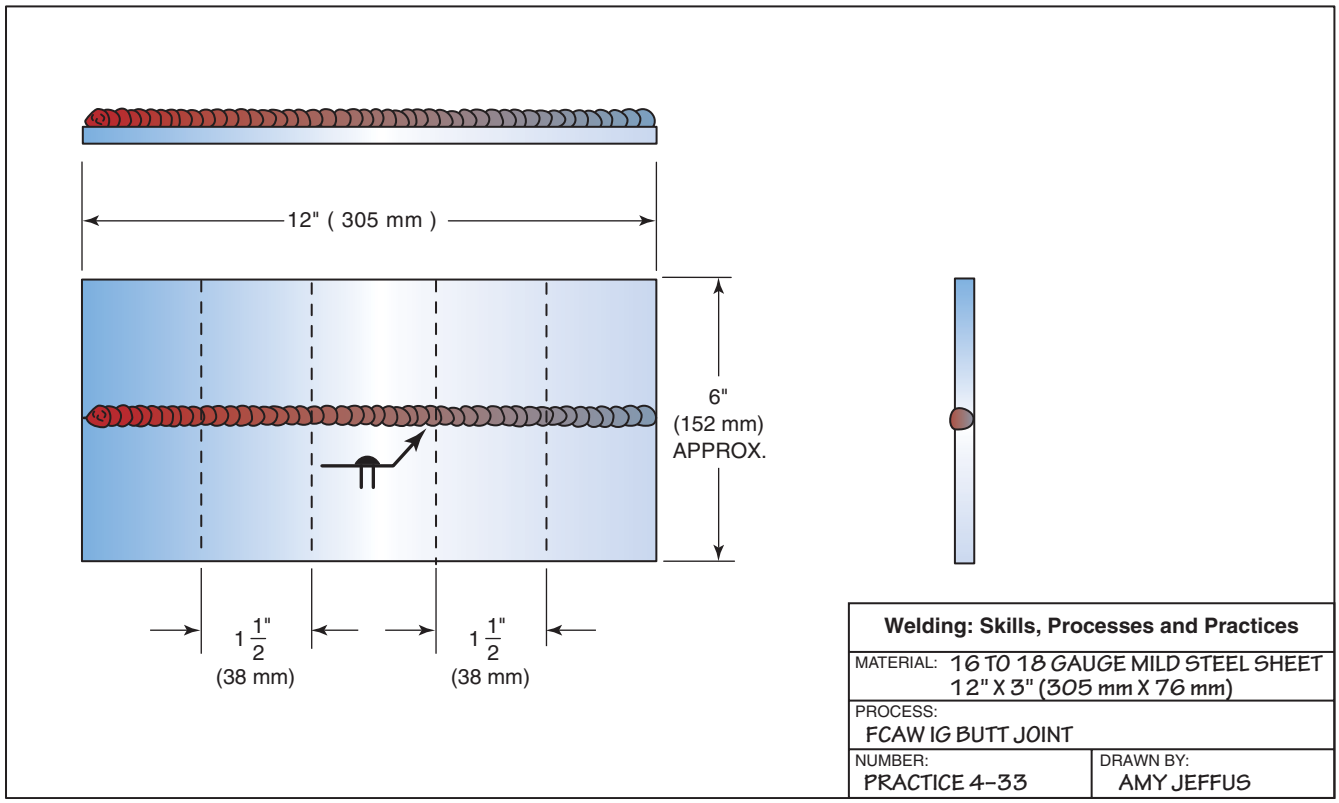

Figure 4.55 FCAW 1G butt joint, 16- to 18-gauge mild steel

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2

PRACTICE 4-35

Lap Joint and Tee Joint 1F 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a fillet weld in the flat position, Figure 4.56A and Figure 4.56B.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until you can make welds with 100% penetration that will pass the bend test, Figure 4.57A and Figure 4.57B. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Vertical" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-36

Butt Joint 3G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild

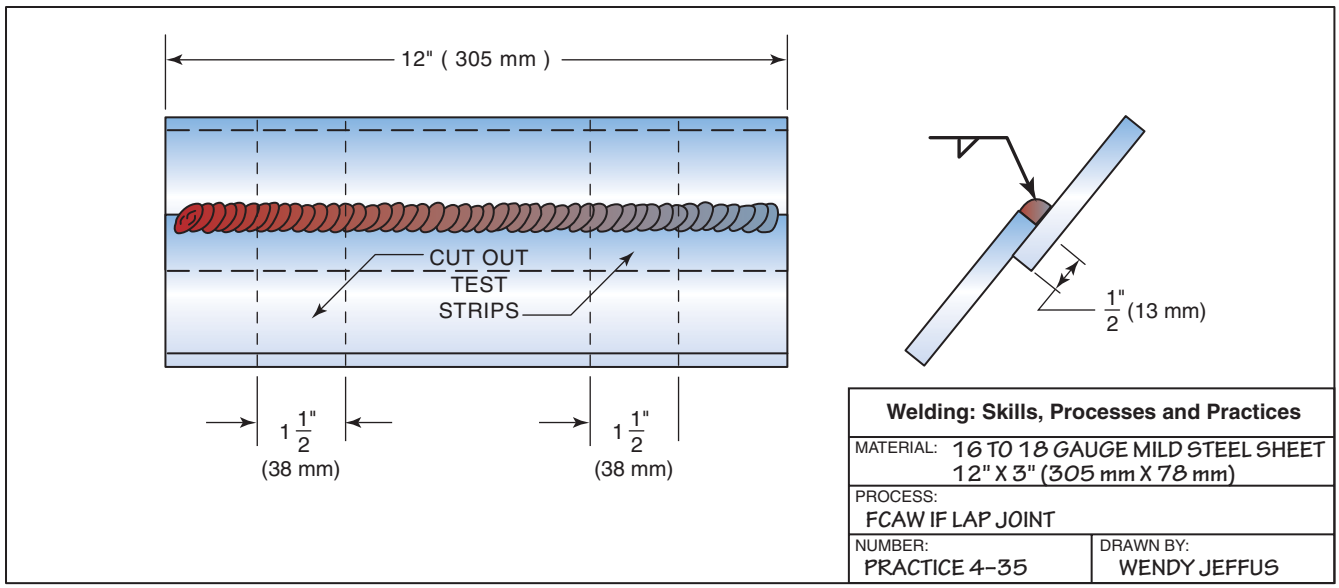

Figure 4.56 (A) FCAW 1F lap joint, 16- to 18-gauge mild steel

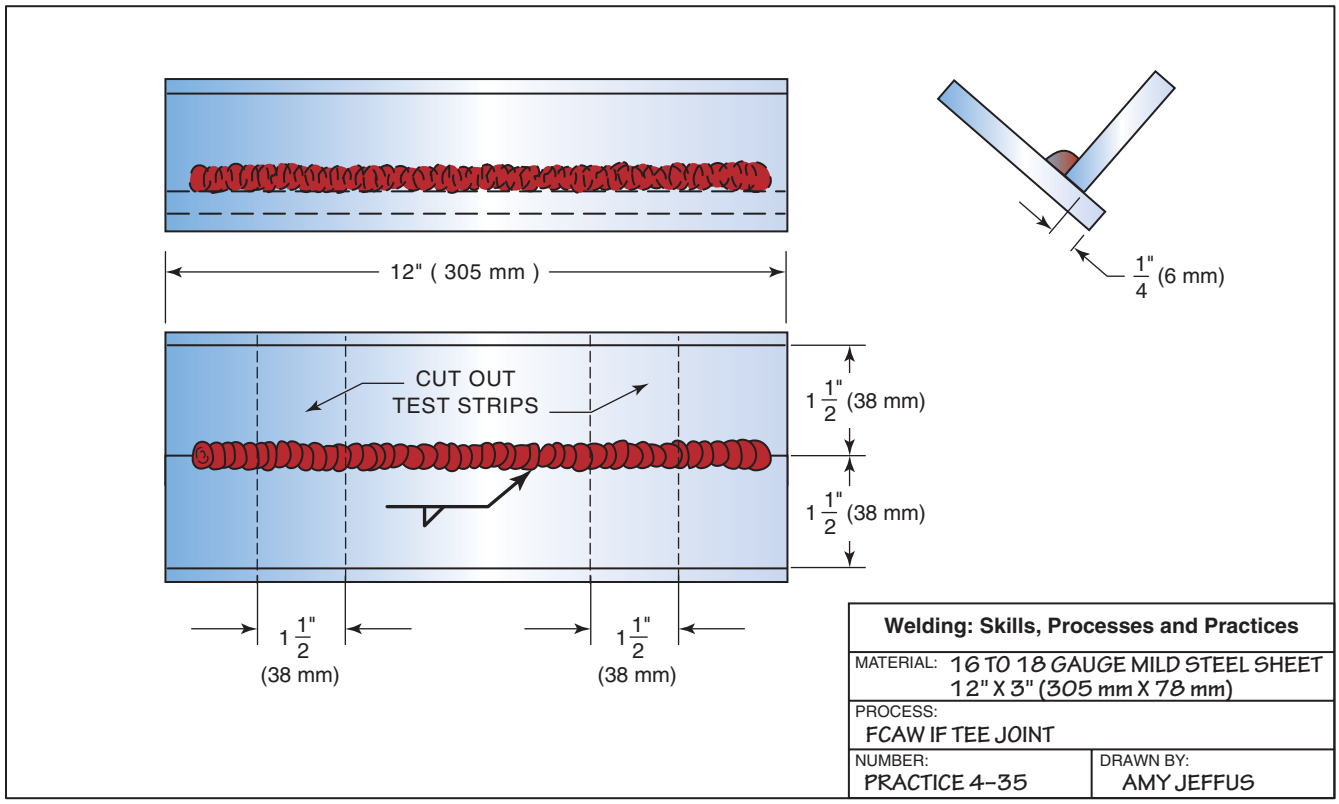

Figure 4.56 (B) FCAW 1F tee joint, 16- to 18-gauge mild steel

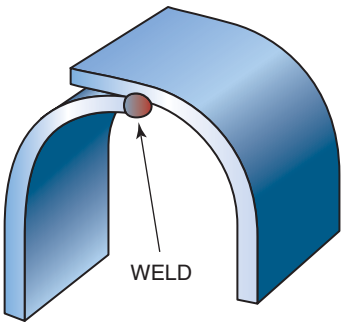

Figure 4.57 (A) 180° bend to test lap weld quality

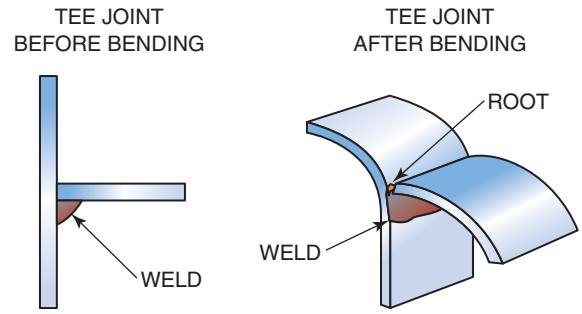

Figure 4.57 (B) Bend the test strip to be sure the weld had good root fusion

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Vertical" component of the all-position requirement for 5 and 10.

steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the vertical up or down position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-37

Lap Joint and Tee Joint 3F

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.030 in.

and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18- gauge thick, you will make a fillet weld in the vertical up or down position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2
This practice addresses the "Vertical" component of the all-position requirement for 5 and 10.

PRACTICE 4-38

Lap Joint and Tee Joint 3F 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a fillet weld in the vertical up or down position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until you can make welds with 100% penetration that will pass the test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Horizontal" component of the all-position requirement for 5 and 10.

PRACTICE 4-39

Lap Joint and Tee Joint 2F

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a fillet weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2
This practice addresses the "Horizontal" component of the all-position requirement for 5 and 10.

PRACTICE 4-40

Lap Joint and Tee Joint 2F 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and

16-gauge to 18-gauge thick, you will make a fillet weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until you can make welds with 100% penetration that will pass the bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-41

Butt Joint 2G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-42

Butt Joint 2G 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-43

Butt Joint 4G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of

mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-44

Butt Joint 4G 100% to Be Tested

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-45

AWS SENSE Entry-Level Welder Workmanship Sample for Flux Cored Arc Welding, Gas-Shielded (FCAW)

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 7

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 4-45. Date:

Title:
Welding FCAW of plate to plate.

Scope:
This procedure is applicable for V-groove, bevel, and fillet welds within the range of 1/8 in. (3.2 mm) through 1-1/2 in. (38 mm).

Welding may be performed in the following positions: all.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1, Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. E71T-1 from AWS specification A5.20. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity:

CO_2 at 30 to 50 cfh or 75% Ar/25% CO_2 at 30 to 50 cfh.

Joint Design and Tolerances:

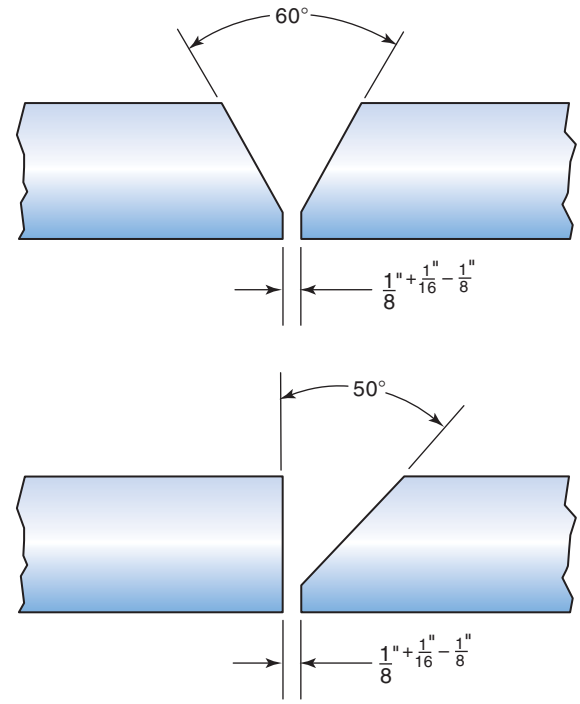

Preparation of Base Metal:
The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E71T-1	0.035 in. (0.9 mm)	130 to 150	288 to 380 (732 to 975)	22 to 25	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E71T-1	0.045 in. (1.2 mm)	150 to 210	200 to 300 (508 to 762)	28 to 29	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 1/2-in. (13-mm) or larger gas nozzle and a distance from contact tube to work of approximately 3/4 in. (19 mm) for all welding, first tack weld the plates together according to Figure 4.58. There should be a root gap of about 1/8 in. (3.2 mm) between the plates with V-grooved or beveled edges. Use an E71T-1 arc welding electrode to make a root pass to fuse the plates together. Clean the slag from the root pass, being sure to remove any trapped slag along the sides of the weld.

Using an E71T-1 arc welding electrode, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled. The 1/4-in. (6.4-mm) fillet welds are to be made with one pass.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
The slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position. A grinder may not be used to remove weld control problems such as undercut, overlap, or trapped slag.

Inspection:
Visually inspect the weld for uniformity and discontinuities. There shall be no cracks, no incomplete fusion, and no overlap. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm). The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length, and the maximum diameter shall not exceed 3/32 in. (2.4 mm).

Sketches:
See Figure 4.58.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-46

AWS SENSE Entry-Level Welder Workmanship Sample for Flux Cored Arc Welding Self-Shielded (FCAW)

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 12

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 4-46 Date:

Title:
Welding FCAW of plate to plate.

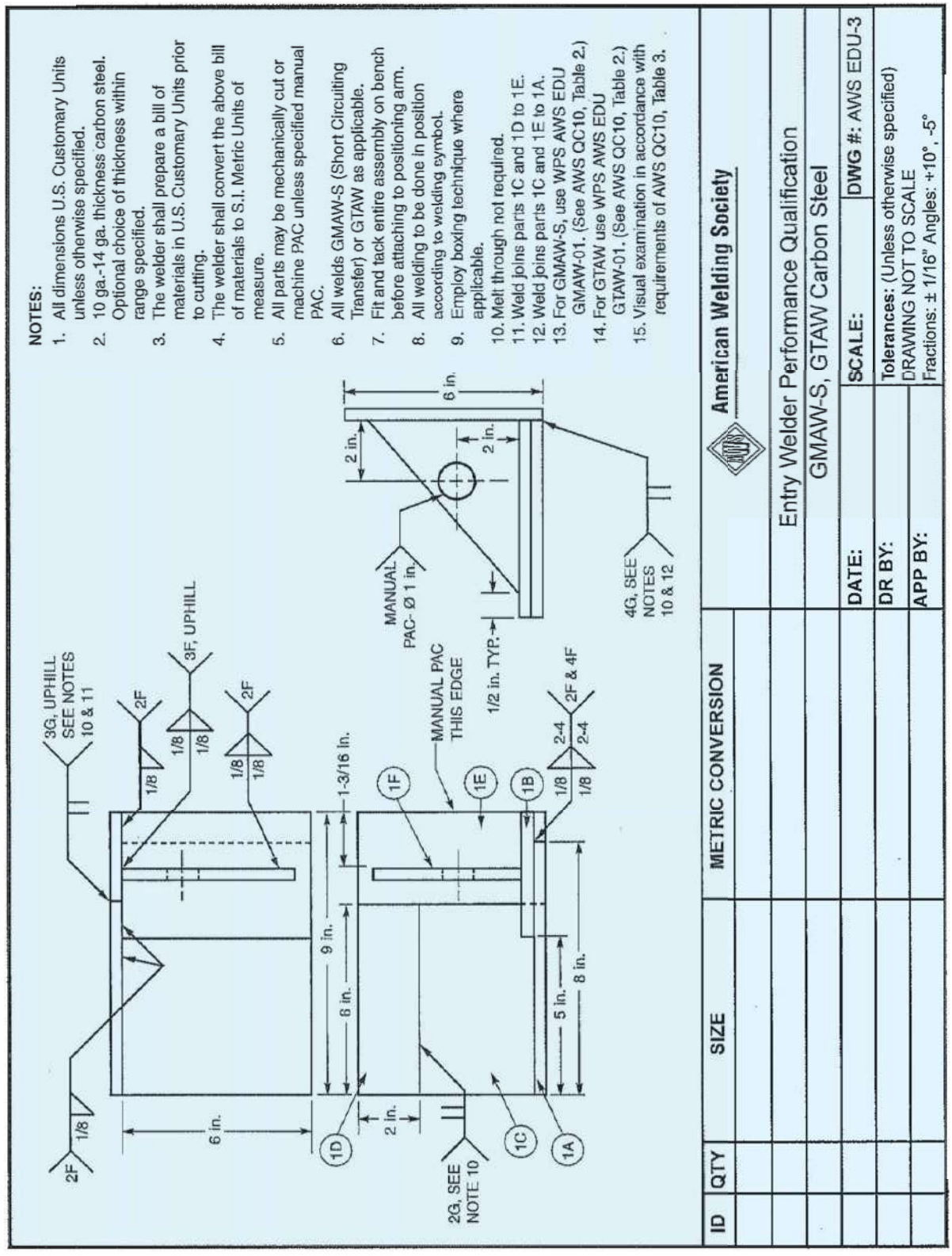

Figure 4.58 FCAW-G workmanship qualification test
Source: Courtesy of the American Welding Society

Scope:
This procedure is applicable for V-groove, bevel, and fillet welds within the range of 1/8 in. (3.2 mm) through 1 1/2 in. (38 mm).

Welding may be performed in the following positions: all.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1, Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. 0.035 to 0.0415 dia. E71T-11 from AWS specification A5.20. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity: none.

Joint Design and Tolerances:

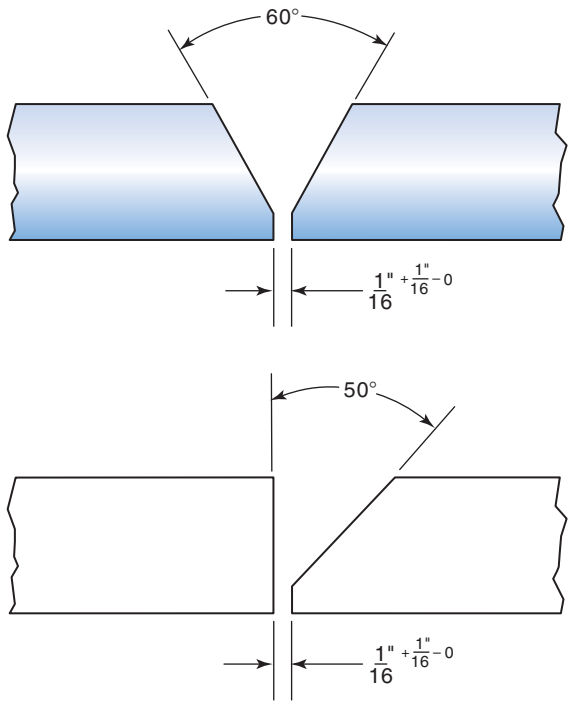

Preparation of Base Metal:
The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct current electrode negative (DCEN). The base metal shall be on the positive side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E71T-11	0.035 in. (0.9 mm)	130 to 150	150 to 225 (381 to 571)	22 to 25	None	—	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E71T-11	0.045 in. (1.2 mm)	150 to 210	105 to 195 (266 to 495)	15 to 18	None	—	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 1/2-in. (13-mm) or larger gas nozzle and a distance from contact tube to work of approximately 3/4 in. (19 mm) for all welding, first tack weld the plates together according to Figure 4.59. There should be a root gap of about 1/8 in. (3.2 mm) between the plates with V-grooved or beveled edges. Use an E71T-11 arc welding electrode to make a root pass to fuse the plates together. Clean the slag from the root pass, being sure to remove any trapped slag along the sides of the weld.

Using an E71T-11 arc welding electrode, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled. The 1/4-in. (6.4-mm) fillet welds are to be made with one pass.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
The slag must cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position. A grinder may not be used to remove weld control problems such as undercut, overlap, or trapped slag.

Inspection:
Visually inspect the weld for uniformity and discontinuities. There shall be no cracks, no incomplete fusion, and no overlap. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm). The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length, and the maximum diameter shall not exceed 3/32 in. (2.4 mm).

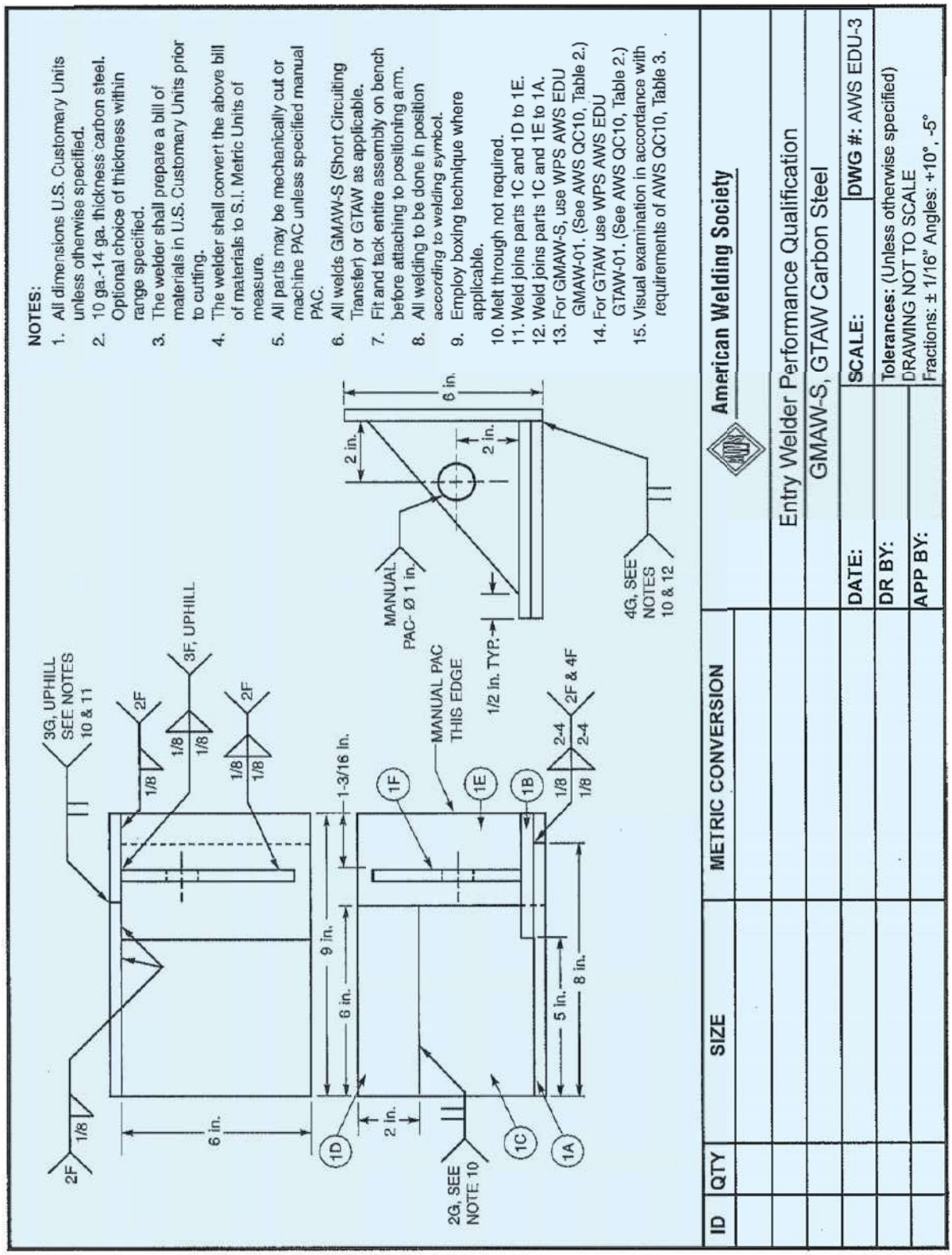

Figure 4.59 FCAW-G workmanship qualification test
Source: Courtesy of the American Welding Society

We Support AWS

TSENSE

WELDING

Skills, Processes and Practices for Entry-Level Welders

BOOK 3

- Shielded Metal Arc Welding
- Gas Tungsten Arc Welding

LARRY JEFFUS • LAWRENCE BOWER

Welding

Skills, Processes and Practices for Entry-Level Welders

Book 3

- Shielded Metal Arc Welding
- Gas Tungsten Arc Welding

Larry Jeffus
Lawrence Bower

DELMAR
CENGAGE Learning™

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

DELMAR
CENGAGE Learning™

Welding Skills, Processes and Practices for Entry-Level Welders: Book Three
Larry Jeffus/Larry Bower

Vice President, Editorial: Dave Garza
Director of Learning Solutions: Sandy Clark
Executive Editor: David Boelio
Managing Editor: Larry Main
Senior Product Manager: Sharon Chambliss
Editorial Assistant: Lauren Stone
Vice President, Marketing: Jennifer McAvey
Executive Marketing Manager: Deborah S. Yarnell
Senior Marketing Manager: Jimmy Stephens
Marketing Specialist: Mark Pierro
Production Director: Wendy Troeger
Production Manager: Mark Bernard
Content Project Manager: Cheri Plasse
Art Director: Benj Gleeksman
Technology Project Manager: Christopher Catalina
Production Technology Analyst: Thomas Stover

© 2010 Delmar, Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at
Professional & Career Group Customer Support, 1-800-648-7450

For permission to use material from this text or product, submit all requests online at **cengage.com/permissions.**
Further permissions questions can be e-mailed to
permissionrequest@cengage.com.

Library of Congress Control Number: 2009920893

ISBN-13: 978-1-4354-2796-9
ISBN-10: 1-4354-2796-3

Delmar
5 Maxwell Drive
Clifton Park, NY 12065-2919
USA

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For your lifelong learning solutions, visit **delmar.cengage.com**
Visit our corporate website at **cengage.com.**

Notice to the Reader

Publisher does not warrant or guarantee any of the products described herein or perform any independent analysis in connection with any of the product information contained herein. Publisher does not assume, and expressly disclaims, any obligation to obtain and include information other than that provided to it by the manufacturer. The reader is expressly warned to consider and adopt all safety precautions that might be indicated by the activities described herein and to avoid all potential hazards. By following the instructions contained herein, the reader willingly assumes all risks in connection with such instructions. The publisher makes no representations or warranties of any kind, including but not limited to, the warranties of fitness for particular purpose or merchantability, nor are any such representations implied with respect to the material set forth herein, and the publisher takes no responsibility with respect to such material. The publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or part, from the readers' use of, or reliance upon, this material.

Printed in Canada
1 2 3 4 5 XX 11 10 09

Brief Contents

Preface ix

AWS Acknowledgment xii

Acknowledgments xiii

About the Authors xiv

Chapter 1 Shielded Metal Arc Equipment, Setup, and Operation 1

Chapter 2 Shielded Metal Arc Welding of Plate 25

Chapter 3 Advanced Shielded Metal Arc Welding 69

Chapter 4 Gas Tungsten Arc Welding Equipment, Setup, Operation, and Filler Metals 105

Chapter 5 Gas Tungsten Arc Welding of Plate 141

Appendix 193

Glossary 195

Index 201

This page intentionally left blank

Contents

Preface ix

AWS Acknowledgment xii

Acknowledgments xiii

About the Authors xiv

Chapter 1 Shielded Metal Arc Equipment, Setup, and Operation 1

Introduction 2
Welding Current 3
Measurement 4
Temperature 4
Currents 5
Types of Welding Power 6
Open Circuit Voltage 7
Operating Voltage 7
Arc Blow 7
Types of Power Sources 9
Transformers 9
Multiple-coil Machines 10
Movable-coil or Movable-core Machines 12
Inverter Machines 13
Generators and Alternators 13
Rectifiers 15
Duty Cycle 17
Welding Cables and Leads 18
Electrode Holders 20
Work Clamps 21
Setup 22
Summary 23
Review 23

Chapter 2 Shielded Metal Arc Welding of Plate 25

Introduction 26
Effect of Too-High or Too-Low Current Settings 29
Electrode Size and Heat 31
Arc Length 32
Electrode Angle 34

Leading Angle 34
Trailing Angle 35
Electrode Manipulation 37
Positioning of the Welder and the Plate 40
Practice Welds 40
Electrodes 40
Stringer Beads 41
Square Butt Joint 44
Edge Weld 49
Outside Corner Joint 54
Lap Joint 58
Tee Joint 63
Summary 66
Review 67

Chapter 3 Advanced Shielded Metal Arc Welding 69

Introduction 70
Root Pass 71
Hot Pass 79
Filler Pass 81
Cover Pass 84
Plate Preparation 85
Preparing Specimens for Testing 88
Acceptance Criteria for Face Bends and Root Bends 88
Restarting a Weld Bead 89
Preheating and Postheating 92
Poor Fitup 101
Summary 103
Review 103

Chapter 4 Gas Tungsten Arc Welding Equipment, Setup, Operation, and Filler Metals 105

Introduction 106
Tungsten 107
Types of Tungsten Electrodes 110
Pure Tungsten Electrodes, EWP 111
Thoriated Tungsten Electrodes, EWTh-1 and EWTh-2 111
Zirconium Tungsten Electrodes, EWZr 112
Cerium Tungsten Electrodes, EWCe-2 112
Lanthanum Tungsten Electrodes, EWLa-1, EWLa-1.5, EWLa-2 112
Alloy Not Specified, EWG 113
Shaping the Tungsten Electrode 113

Grinding 113
Breaking and Remelting 114
Chemical Cleaning and Pointing 116
Pointing and Remelting 116
GTA Welding Equipment 117
Torches 117
Hoses 118
Nozzles 121
Flowmeter 122
Types of Welding Current 123
AC Balance Control 126
Shielding Gases 129
Argon 129
Helium 130
Hydrogen 130
Nitrogen 130
Hot Start 131
Preflow Time and Postflow Time 131
Gas Flow Rate 132
Remote Controls 133
Summary 138
Review 138

Chapter 5 Gas Tungsten Arc Welding of Plate 141

Introduction 143
Torch Angle 143
Filler Rod Manipulation 144
Tungsten Contamination 146
Current Setting 146
Experiments 147
Gas Flow 149
Practice Welds 151
Low-Carbon and Mild Steels 151
Stainless Steel 152
Aluminum 153
Metal Preparation 154
Summary 191
Review 191

Appendix 193

Glossary 195

Index 201

This page intentionally left blank

Preface

ABOUT THE SERIES

Welding: Skills, Processes and Practices for Entry-Level Welders is an exciting new series that has been designed specifically to support the American Welding Society's (AWS) SENSE EG2.0 training guidelines. Offered in three volumes, these books are carefully crafted learning tools consisting of theory-based texts that are accompanied by companion lab manuals, and extensive instructor support materials. With a logical organization that closely follows the modular structure of the AWS guidelines, the series will guide readers through the process of acquiring and practicing welding knowledge and skills. For schools already in the SENSE program, for those planning to join, or for schools interested in obtaining certifiable outcomes based on nationally recognized industry standards in order to comply with the latest Carl D. Perkins Career and Technical Education Requirements. *Welding: Skills, Processes and Practices for Entry-Level Welders* offers a turnkey solution of high quality teaching and learning aids.

Career and technical education instructors at the high school level are often called upon to be multi-disciplinary educators, teaching welding as only one of as many as five technical disciplines in any given semester. The *Welding: Skills, Processes and Practices for Entry-Level Welders* package provides these educators with a process-based, structured approach and the tools they need to be prepared to deliver high level training on processes and materials with which they may have limited familiarity or experience. Student learning, satisfaction and retention are the target of the logically planned practices, supplements and full color textbook illustrations. While the AWS standards for entry level welders are covered, students are also introduced to the latest in high technology welding equipment such as pulsed gas metal arc welding (GMAW-P). Career pathways and career clusters may be enhanced by the relevant mathematics applied to real world activities as well as oral and written communication skills linked to student interaction and reporting.

Book 1, the core volume, introduces students to the welding concepts covered in AWS SENSE Modules 1, 2, 3, 8 and 9 (Occupational Orientation, Safety and Health of Welders, Drawing and Welding Symbol Interpretation, Thermal Cutting, and Weld Inspection Testing and Codes). Book 1 contains all the material needed for a SENSE program that prepares students for qualification in Thermal Cutting processes. The optional Books 2 and 3 cover other important welding processes and are grouped in logical combinations. Book 2 corresponds to AWS SENSE Modules 5 and 6 (GMAW, FCAW), and Book 3 corresponds to AWS SENSE Modules 4 and 7 (SMAW, GTAW).

The texts feature hundreds of four-color figures, diagrams and tight shots of actual welds to speed beginners to an understanding of the most widely used welding processes.

FEATURES

- Produced in close collaboration with experienced instructors from established SENSE programs to maximize the alignment of the content with SENSE guidelines and to ensure 100% coverage of Level I-Entry Welder Key Indicators.
- Chapter introductions contain general performance objectives, key terms used, and the AWS SENSE EG2.0 Key Indicators addressed in the chapter.
- Coverage of Key Indicators is indicated in the margin by a torch symbol and a numerical reference.
- Contains scores of fully illustrated Practices, which are guided exercises designed to help students master processes and materials. Where applicable, the Practices reproduce and reference actual AWS technical drawings in order to help students create acceptable workmanship samples.
- Each section introduces students to the materials, equipment, setup procedures and critical safety information they need in order to weld successfully.
- Hundreds of four-color figures, diagrams and tight shots of actual welds to speed beginners to an understanding of the most widely used welding processes.
- End of chapter review questions develop critical thinking skills and help students to understand "why" as well as "how."

SUPPLEMENTS

Each book in the Welding Skills series is accompanied by a ***Lab Manual*** that has been designed to provide hands-on practice and reinforce the student's understanding of the concepts presented in the text. Each chapter contains practice exercises to reinforce the primary objectives of the lesson, including creation of workmanship samples (where applicable), and a quiz to test knowledge of the material. Artwork and safety precautions are included throughout the manuals.

Instructor Resources (on CD-ROM), designed to support Books 1–3 and the accompanying Lab Manuals, provide a wealth of time-saving tools, including:

- An Instructor's Guide with answers to end-of-chapter Review Questions in the texts and Lab Manual quizzes.
- Modifiable model Lesson Plans that aid in the design of a course of study that meets local or state standards and also maps to the SENSE guidelines.
- An extensive ExamView Computerized Test Bank that offers assessments in true/false, multiple choice, sentence completion and short answer formats. Test questions have been designed to expose students to the types of questions they'll encounter on the SENSE Level 1 Exams.
- PowerPoint Presentations with selected illustrations that provide a springboard for lectures and reinforce skills and processes covered in the texts. The PowerPoint Presentations can be modified or expanded as instructors desire, and can be augmented with additional illustrations from the Image Library.
- The Image Library contains nearly all (well over 1000!) photographs and line art from the texts, most in four-color.
- A SENSE Correlation Chart that shows the close alignment of the *Welding* series to the SENSE Entry Level 1 training guidelines. Each Key Indicator within each SENSE Module is mapped to the relevant text and lab manual page or pages.

TITLES IN THE SERIES

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 1, Occupational Orientation, Safety and Health of Welders, Drawing and Welding Symbol Interpretation, Thermal Cutting, Weld Inspection Testing and Codes
(Order #: 1-4354-2788-2)
Lab Manual, Book One (Order #: 1-4354-2789-0)

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 2, Gas Metal Arc Welding, Flux Cored Arc Welding (Order #: 1-4354-2790-4)
Lab Manual, Book Two (Order #: 1-4354-2795-5)

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 3, Shielded Metal Arc Welding, Gas Tungsten Arc Welding (Order #: 1-4354-2796-3)
Lab Manual, Book Three (Order #: 1-4354-2797-1)

AWS Acknowledgment

The Authors and Publisher gratefully acknowledge the support provided by the American Welding Society in the development and publication of this textbook series. "American Welding Society," the AWS logo and the SENSE logo are the trade and service marks of the American Welding Society and are used with permission.

For more information on the American Welding Society and the SENSE program, visit **http://www.aws.org/education/sense/** or contact AWS at (800) 443-9353 ext. 455 or by email: **education@aws.org**.

Acknowledgments

The authors and publisher would like to thank the following individuals for their contributions:

Garey Bish, *Gwinnett Technical College, Lawrenceville, GA*
Julius Blair, *Greenup County Area Technology Center, Greenup, KY*
Rick Brandon, *Pemiscot County Career & Technical Center, Hayti, MO*
Stephen Brandow, *University of Alaska Southeast, Ketchikan, Ketchikan, AK*
Francis X Brieden, *Career Technology Center of Lackawanna County, Scranton, PA*
John Cavenaugh, *Community College of Southern Nevada, Las Vegas, NV*
Clay Corey, *Washington-Saratoga BOCES, Fort Edward, NY*
Keith Cusey, *Institute for Construction Education, Decatur, IL*
Craig Donnell, *Whitmer Career Technology Center, Toledo, OH*
Steve Farnsworth, *Iowa Lakes Community College, Emmetsburg, IA*
Ed Harrell, *Traviss Career Center, Lakeland, FL*
Robert Hoting, *Northeast Iowa Community College , Sheldon, IA*
Steve Kistler, *Moberly Area Technical Center, Moberly, MO*
David Lynn, *Lebanon Technology & Career Center, Lebanon, MO*
Frank Miller, *Gadsden State Community College , Gadsden, AL*
Chris Overfelt, *Arnold R Burton Tech Center, Salem, VA*
Kenric Sorenson, *Western Technical College, LaCrosse, WI*
Pete Stracener, *South Plains College, Levelland, TX*
Bill Troutman, *Akron Public Schools, Akron, OH*
Norman Verbeck, *Columbia/Montour AVTS, Bloomsburg, PA*

About The Authors

Larry Jeffus is a dedicated teacher and author with over twenty years experience in the classroom and several Delmar Cengage Learning welding publications to his credit. He has been nominated by several colleges for the Innovator of the Year award for setting up nontraditional technical training programs. He was also selected as the Outstanding Post-Secondary Technical Educator in the State of Texas by the Texas Technical Society. Now retired from teaching, he remains very active in the welding community, especially in the field of education.

Lawrence Bower is a welding instructor at Blackhawk Technical College, an AWS SENSE School, in Janesville, Wisconsin. Mr. Bower is an AWS-certified Welding Inspector and Welding Educator. In helping to create *Welding: Skills, Processes and Practices for Entry-Level Welders,* he has brought to bear an excellent mix of training experience and manufacturing know-how from his work in industry, including fourteen years at United Airlines, and six years in the US Navy as an aerospace welder.

CHAPTER

1 Shielded Metal Arc Equipment, Setup, and Operation

OBJECTIVES

After completing this chapter, the student should be able to

- describe the process of shielded metal arc welding (SMAW)
- list and define the three units used to measure a welding current
- contrast how adding various chemicals to the coverings of the electrodes affects the welding arc
- contrast the three different types of current used for welding and their effects on welds
- contrast constant current (CC) and constant voltage (CV) welding power supplies and which type the shielded metal arc welding process requires
- define open circuit voltage and operating voltage
- identify arc blow and apply three different techniques to control it
- tell what the purpose of a welding transformer is and what kind of change occurs to the voltage and amperage with a step-down transformer
- tell the purpose of a rectifier
- read a welding machine duty cycle chart and explain its significance
- determine the proper welding cable size
- service and repair SMAW electrode holders
- identify three problems that can occur as a result of poor work lead clamping
- describe the factors that should be considered when placing an arc welding machine in a welding area

KEY TERMS

amperage
anode
cathode
duty cycle
electrons
inverter
magnetic flux lines
open circuit voltage
operating voltage
output
rectifier
step-down transformer
voltage
wattage
welding cables
welding leads

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports or records
Key Indicator 3: Follows verbal instructions to complete work assignments
Key Indicator 4: Follows written instructions to complete work assignments

Module 4: Shielded Metal Arc Welding (SMAW)

Key indicator 1: Performs safety inspections of SMAW equipment and accessories
Key Indicator 2: Makes minor external repairs to SMAW equipment and accessories
Key Indicator 3: Sets up SMAW operations on carbon steel

INTRODUCTION

Shielded metal arc welding (SMAW) is a welding process that uses a flux-covered metal electrode to carry an electrical current, **Figure 1.1.** The current forms an arc across the gap between the end of the electrode and the work. The electric arc creates sufficient heat to melt both the electrode and the work. Molten metal from the electrode travels across the arc to the molten pool on the base metal, where they mix together. The end of the electrode and molten pool of metal is surrounded, purified, and protected by a gaseous cloud and a covering of molten flux is produced as the flux coating of the electrode burns or vaporizes. As the arc moves away, the mixture of molten electrode and base metal solidifies and becomes one piece. At the same time, the molten flux solidifies forming a solid slag. Some electrode types produce heavier slag coverings than others.

SMAW is a widely used welding process because of its low cost, flexibility, portability, and versatility. The machine and the electrodes are low cost. The machine itself can be as simple as a 110-V step-down transformer. The electrodes are available from a large number of manufacturers in packages from 1 lb (0.5 kg) to 50 lb (22 kg).

The SMAW process is very flexible in terms of the metal thicknesses that can be welded and the variety of positions it can be used in. Metal as thin as 1/16 in. (2 mm) thick, or approximately 16 gauge, to several feet thick can be welded using the same machine with different settings and sizes of electrodes. The flexibility of the process also allows metal in this thickness range to be welded in any position.

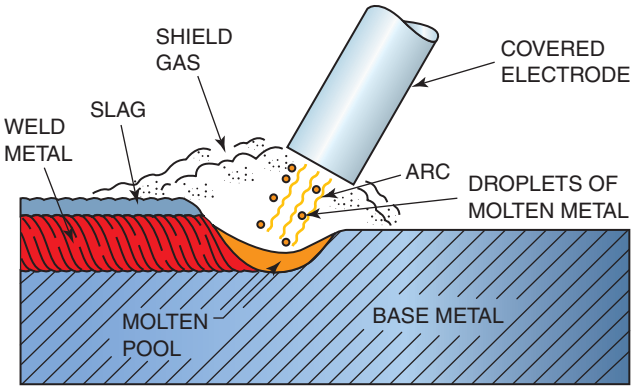

Figure 1.1
Shielded metal arc welding

SMAW is a very portable process because it is easy to move the equipment, and engine-driven generator-type welders are available. Also, the limited amount of equipment required for the process makes moving easy.

The process is versatile, and it is used to weld almost any metal or alloy, including cast iron, aluminum, stainless steel, and nickel.

WELDING CURRENT

The source of heat for arc welding is an electric current. An electric current is the flow of **electrons.** Electrons flow through a conductor from negative (−) to positive (+), **Figure 1.2.** Resistance to the flow of electrons (electricity) produces heat. The greater the resistance, the greater the heat. Air has a high resistance to current flow. As the electrons jump the air gap between the end of the electrode and the work, a great deal of heat is produced. Electrons flowing across an air gap produce an arc.

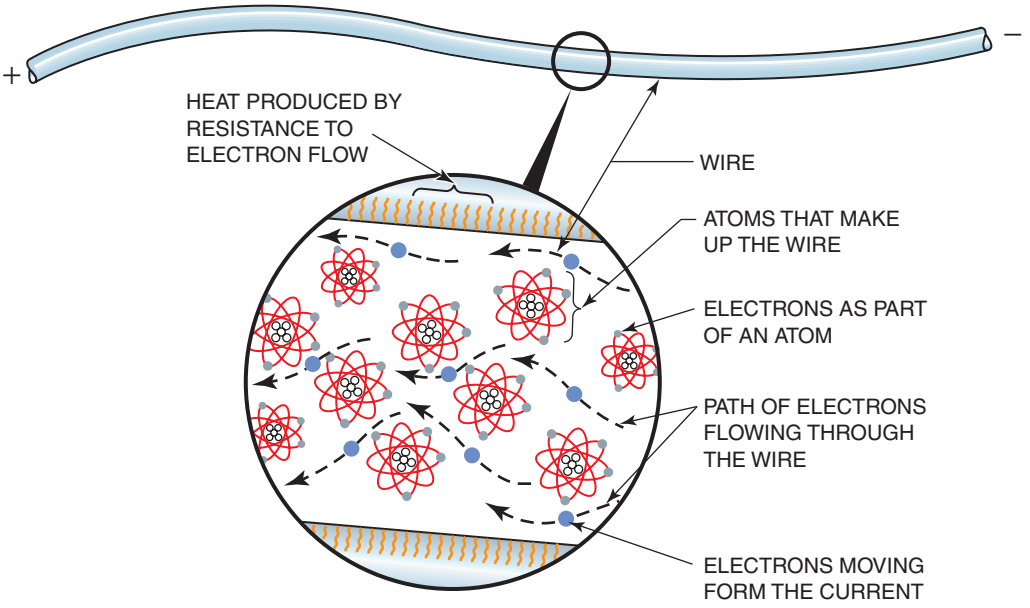

Figure 1.2
Electrons traveling along a conductor

Measurement

Three units are used to measure a welding current. The three units are voltage (V), amperage (A), and wattage (W). **Voltage,** or volts (V), is the measurement of electrical pressure, in the same way that pounds per square inch is a measurement of water pressure. Voltage controls the maximum gap the electrons can jump to form the arc. A higher voltage can jump a larger gap. **Amperage,** or amperes (A), is the measurement of the total number of electrons flowing, in the same way that gallons is a measurement of the amount of water flowing. Amperage controls the size of the arc. **Wattage,** or watts (W), is calculated by multiplying volts (V) times amperes (A), **Figure 1.3.** Wattage is a measurement of the amount of electrical energy or power in the arc. The amount of watts being put into a weld per inch (cm) controls the width and depth of the weld bead, **Figure 1.4.**

Temperature

The temperature of a welding arc exceeds 11,000°F (6000°C). The exact temperature depends on the resistance to the current flow. The resistance is affected by the arc length and the chemical composition of the gases formed as the electrode covering burns and vaporizes. As the arc lengthens, the resistance increases, thus causing a rise in the arc voltage and temperature. The shorter the arc, the lower the arc temperature produced.

Most shielded metal arc welding electrodes have chemicals added to their coverings to stabilize the arc. These arc stabilizers reduce the arc resistance, making it easier to maintain an arc. By lowering the resistance, the arc stabilizers also lower the arc temperature. Other chemicals within the gaseous cloud around the arc may raise or lower the resistance.

The amount of heat produced is determined by the size of the electrode and the amperage setting. Not all of the heat produced by an arc reaches the weld. Some of the heat is radiated away in the form of light and heat waves, **Figure 1.5.** Additional heat is carried away with the hot gases formed by the electrode covering. Heat also is lost through conduction in the work. In total, about 50% of all heat produced by an arc is missing from the weld.

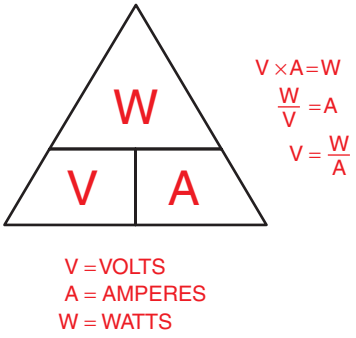

Figure 1.3
Ohm's law

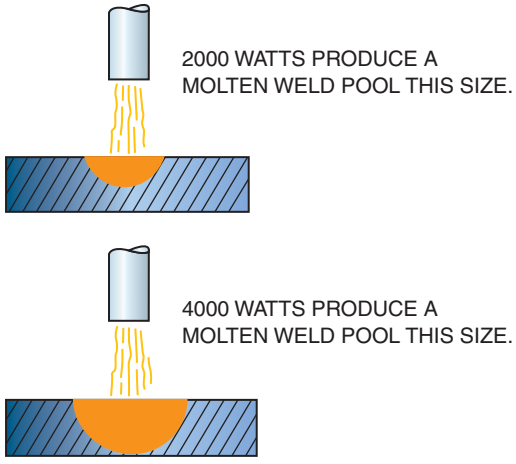

Figure 1.4
The molten weld pool size depends upon the energy (watts), the metal mass, and thermal conductivity

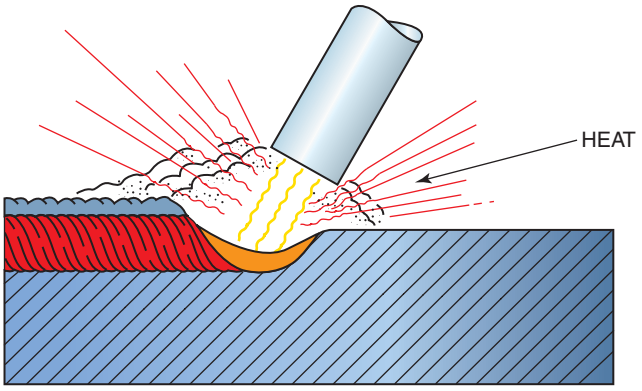

Figure 1.5
Energy is lost from the weld in the forms of radiation and convection

The 50% of the remaining heat the arc produced is not distributed evenly between both ends of the arc. This distribution depends on the composition of the electrode's coating, the type of welding current, and the polarity of the electrode's coating.

Currents

The three different types of current used for welding are alternating current (AC), direct current electrode negative (DCEN), and direct current electrode positive (DCEP). The terms DCEN and DCEP have replaced the former terms *direct current straight polarity (DCSP)* and *direct current reverse polarity (DCRP).* DCEN and DCSP are the same currents, and DCEP and DCRP are the same currents. Some electrodes can be used with only one type of current. Others can be used with two or more types of current. Each welding current has a different effect on the weld.

DCEN

In direct current electrode negative, the electrode is negative, and the work is positive, **Figure 1.6.** DCEN welding current produces a high electrode melting rate.

DCEP

In direct current electrode positive, the electrode is positive, and the work is negative, **Figure 1.7.** DCEP current produces the deepest penetrating welding arc characteristics.

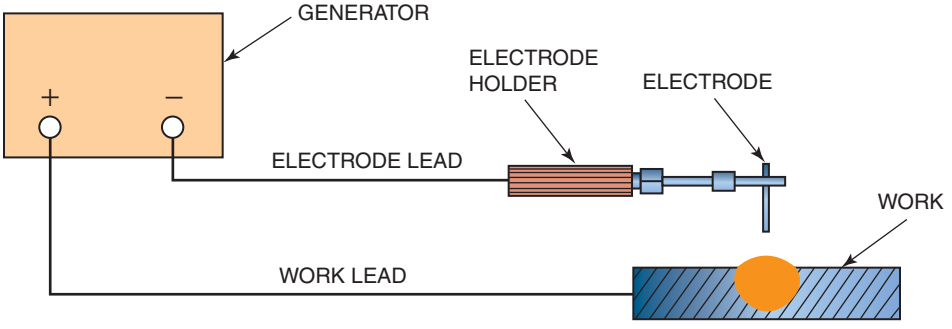

Figure 1.6
Electrode negative (DCEN), straight polarity (DCSP)

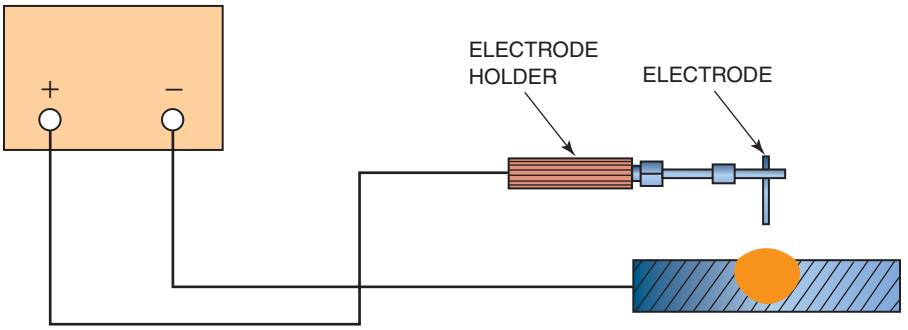

Figure 1.7
Electrode positive (DCEP), reverse polarity (DCRP)

AC

In alternating current, the electrons change direction every 1/120 of a second so that the electrode and work alternate from **anode** to **cathode, Figure 1.8.** The positive side of an electrode arc is called the anode, and the negative side is called the cathode. The rapid reversal of the current flow causes the welding heat to be evenly distributed on both the work and the electrode—that is, half on the work and half on the electrode. The even heating gives the weld bead a balance between penetration and buildup.

TYPES OF WELDING POWER

Welding power can be supplied as

- Constant voltage (CV)—The arc voltage remains constant at the selected setting even if the arc length and amperage increase or decrease.
- Rising-arc voltage (RAV)—The arc voltage increases as the amperage increases.
- Constant current (CC)—The total welding current (watts) remains the same. This type of power is also called drooping-arc voltage (DAV), because the arc voltage decreases as the amperage increases.

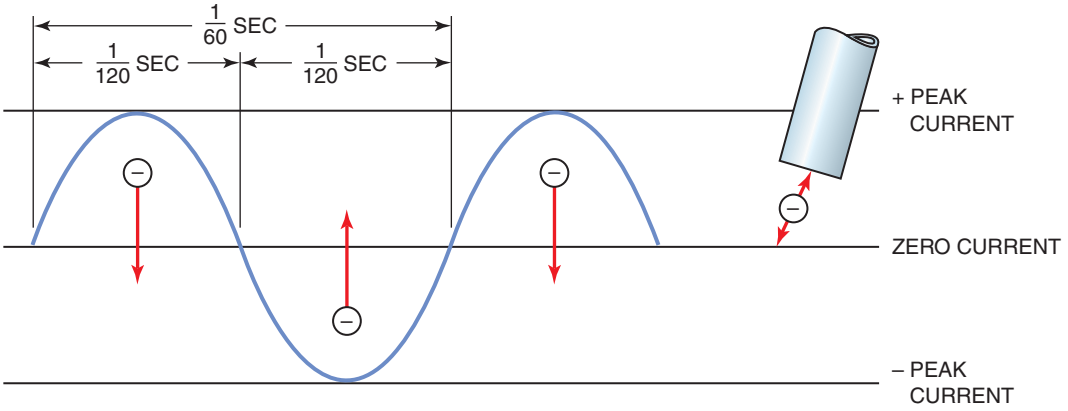

Figure 1.8
Alternating current sine wave (AC)

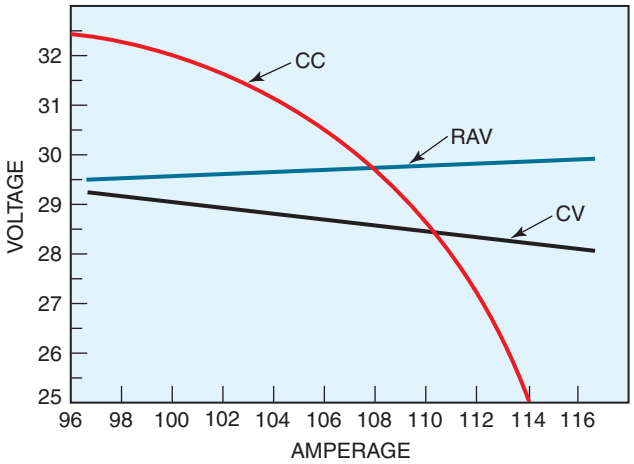

Figure 1.9
Constant voltage (CV), rising arc voltage (RAV), and constant current (CC)

The shielded metal arc welding (SMAW) process requires a constant current arc voltage characteristic, illustrated by the CC line in **Figure 1.9.** The shielded metal arc welding machine's voltage **output** decreases as current increases. This output power supply provides a reasonably high open circuit voltage before the arc is struck. The high open circuit voltage quickly stabilizes the arc. The arc voltage rapidly drops to the lower closed circuit level after the arc is struck. Following this short starting surge, the power (watts) remains almost constant despite the changes in arc length. With a constant voltage output, small changes in arc length would cause the power (watts) to make large swings. The welder would lose control of the weld.

OPEN CIRCUIT VOLTAGE

Open circuit voltage is the voltage at the electrode before striking an arc (with no current being drawn). This voltage is usually between 50 V and 80 V. The higher the open circuit voltage, the easier it is to strike an arc. The higher voltage also increases the chance of electrical shock.

CAUTION

The maximum safe open circuit voltage for welders is 80 V.

OPERATING VOLTAGE

Operating voltage, or closed circuit voltage, is the voltage at the arc during welding. This voltage will vary with arc length, type of electrode being used, type of current, and polarity. The operating voltage will be between 17 V and 40 V.

ARC BLOW

When electrons flow, they create lines of magnetic force that circle around the line of flow, **Figure 1.10.** Lines of magnetic force are referred to as **magnetic flux lines.** These lines space themselves evenly along a current-carrying wire. If the wire is bent, the flux lines on one side are

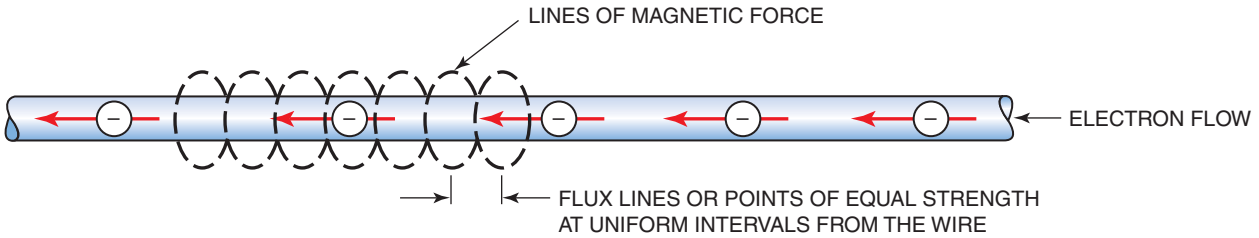

Figure 1.10
Magnetic force around a wire

compressed together, and those on the other side are stretched out, **Figure 1.11.** The unevenly spaced flux lines try to straighten the wire so that the lines can be evenly spaced once again. The force that they place on the wire is usually small. However, when welding with very high amperages, 600 A or more, the force may cause the wire to move.

The welding current flowing through a plate or any residual magnetic fields in the plate will result in uneven flux lines. These uneven flux lines can, in turn, cause an arc to move during a weld. This movement of the arc is called *arc blow.* Arc blow makes the arc drift as a string would drift in the wind. Arc blow is more noticeable in corners, at the ends of plates, and when the work lead is connected to only one side of a plate, **Figure 1.12.** If arc blow is a problem, it can be controlled by connecting the work lead to the end of the weld joint and making the weld in the direction toward the work lead, **Figure 1.13.** Another way of controlling arc blow is to use two work leads, one on each side of the weld. The best way to eliminate arc blow is to use alternating current. AC usually does not allow the flux lines to build long enough to bend the arc before the current changes direction. If it is impossible to move the work connection or to change to AC, a very short arc length can help control arc blow. A small tack weld or a change in the electrode angle can also help control arc blow.

Arc blow may not be a problem as you are learning to weld in the shop, because most welding tables are all steel. If you are using a pipe stand to hold your welding practice plates, arc blow can become a problem. In this case, try reclamping your practice plates or switch to alternating current.

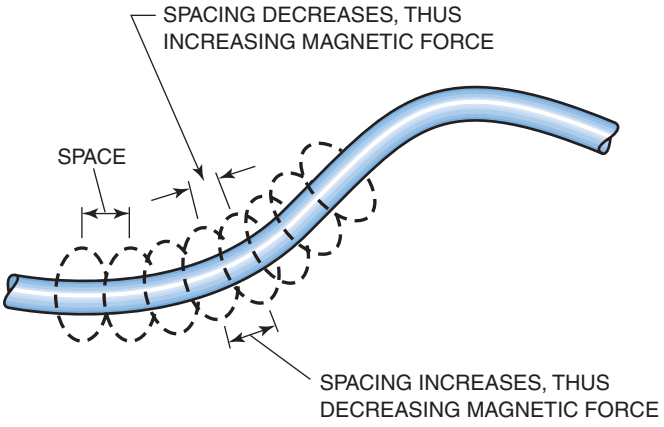

Figure 1.11
Magnetic forces concentrate around bends in wires

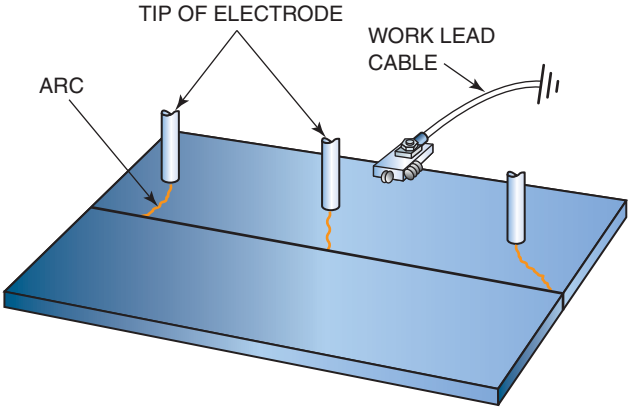

Figure 1.12
Arc blow

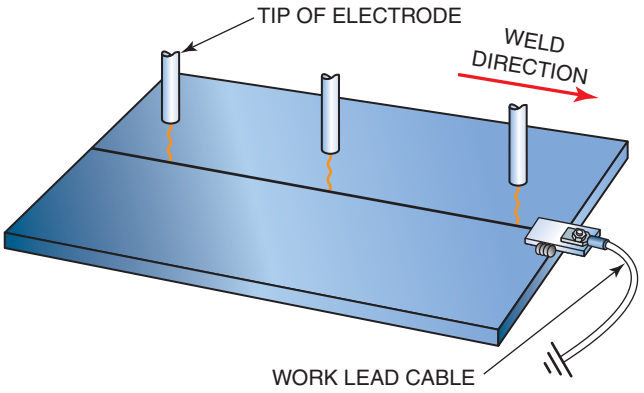

Figure 1.13
Correct current connections to control arc blow

TYPES OF POWER SOURCES

Two types of electrical devices can be used to produce the low-voltage, high-amperage current combination that arc welding requires. One type uses electric motors or internal combustion engines to drive alternators or generators. The other type uses **step-down transformers.** Because transformer-type welding machines are quieter, are more energy efficient, require less maintenance, and are less expensive, they are now the industry standard. However, engine-powered generators are still widely used for portable welding.

Transformers

A welding transformer uses the alternating current (AC) supplied to the welding shop at a high voltage to produce the low-voltage welding power. As electrons flow through a wire they produce a magnetic field around the wire. If the wire is wound into a coil the weak magnetic field of each wire is concentrated to produce a much stronger central magnetic force. Because the current being used is alternating or reversing each 1/120 of a second, the magnetic field is constantly being built and allowed to collapse. By placing a second, or secondary, winding of wire in the magnetic field produced by the first, or primary, winding, a current will be induced in the secondary winding. The placing of an iron core in the center of these coils will increase the concentration of the magnetic field, **Figure 1.14.**

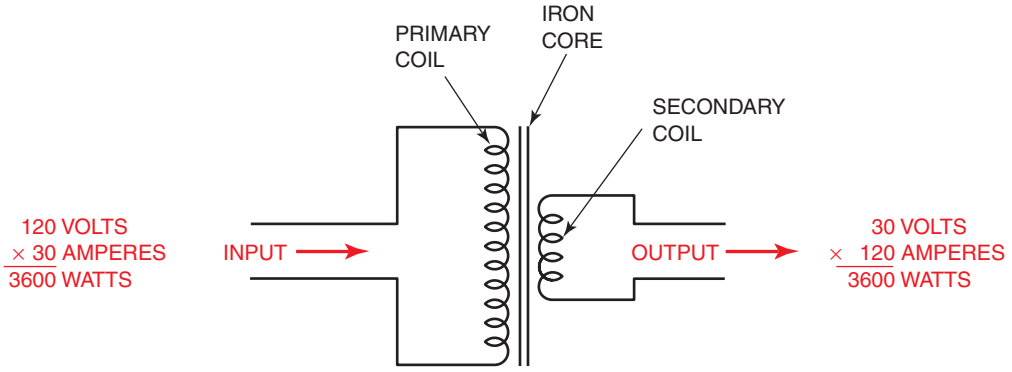

Figure 1.14
Diagram of a step-down transformer

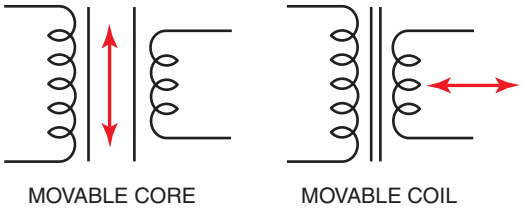

Figure 1.15
Major types of adjustable welding transformers

A transformer with more turns of wire in the primary winding than in the secondary winding is known as a step-down transformer. A step-down transformer takes a high-voltage, low-amperage current and changes it into a low-voltage, high-amperage current. Except for some power lost by heat within a transformer, the power (watts) into a transformer equals the power (watts) out because the volts and amperes are mutually increased and decreased.

A transformer welder is a step-down transformer. It takes the high line voltage (110 V, 220 V, 440 V, etc.) and low-amperage current (30 A, 50 A, 60 A, etc.) and changes it into 17 V to 45 V at 190 A to 590 A.

Welding machines can be classified by the method through which they control or adjust the welding current. The major classifications are multiple-coil, called tap-type, movable-coil or movable-core, **Figure 1.15,** and inverter-type machines.

Multiple-coil Machines

The multiple-coil machine, or tap-type machine, allows the selection of different current settings by tapping into the secondary coil at a different turn value. The greater the number of turns, the higher the amperage is

Figure 1.16
Tap-type transformer welding machine
Courtesy of Lincoln Electric Company

induced in the turns. These machines may have a large number of fixed amperes, **Figure 1.16,** or they may have two or more amperages that can be adjusted further with a fine adjusting knob. The fine adjusting knob may be marked in amperes, or it may be marked in tenths, hundredths, or in any other unit.

EXPERIMENT 1-1

Estimating Amperages

Module 1
Key Indicator 1, 3, 4

Module 4
Key Indicator 3

Using a pencil and paper, you will prepare a rough estimate of the amperage setting of a welding machine. **Figure 1.17** shows a welding machine with low, medium, and high tap amperage ranges. A fine adjusting knob is marked with ten equal divisions, and each division is again divided by ten smaller lines.

The machine is set on the medium range, 50 A to 250 A, and the fine adjusting knob is turned until it points to the line marked 5 (halfway between 0 and 10). This means that the amperage is halfway from 50 to 250, or 150 A. If the fine adjusting knob points between 2 and 3, the resulting amperage is one-quarter of the way from 50 to 250, or about 100 A. If the knob points between 7 and 8, the amperage is three-quarters of the way from 50 to 250, or about 200 A. If the knob points at 4, the amperage is more than 100 but a little less than 150, or about 130 A to 140 A. What is the amperage if the knob points at 6?

Since this is a method of estimating only, the amperage value obtained is close enough to allow an arc to be struck. The welder can then finish the fine adjusting to obtain a good weld.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

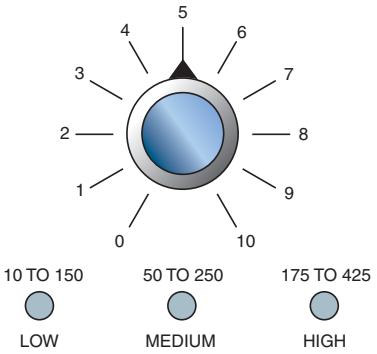

Figure 1.17
Tap-type welder knobs

EXPERIMENT 1-2

Calculating the Amperage Setting

Module 1
Key Indicator 1, 3, 4

Module 4
Key Indicator 3

Using a pencil and paper or calculator, you will calculate the exact value for each space on the fine adjusting knob of a welding machine.

With the machine set on the medium range, from 50 A to 250 A, first subtract the low amperage from the high amperage to get the amperage spread (250 − 50 = 200). Now divide the amperage spread by the number of units shown on the fine adjusting knob (200 ÷ 10 = 20). Each unit is equal to a 20-A increase, **Table 1.1.** When the knob points to 0, the amperage is 50; when the knob points to 1, the amperage is 70; and at 2, the amperage is 90, **Figure 1.18.** There are 100 small units on the fine adjusting knob. Dividing the amperage spread by the number of small units gives the amperage value for each unit (200 ÷ 100 = 2). Therefore, if the knob points to 6.1, the amperage is set at a value of 50 + 120 + 2 = 172 A. This method provides a good starting place for the current setting, but if the welding is to be made in accordance with a welding procedure's specific amperage setting it will be necessary to use a calibrated meter to make the correct setting.

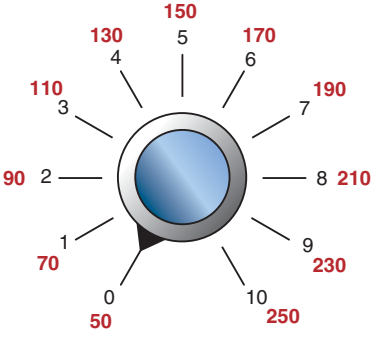

Figure 1.18
Fine adjusting knob

Table 1.1 Example of a Table Used to Calculate the Amperage Setting

Setting	Value in Amperes
0 = 50 + 0,	or 50 A
1 = 50 + 20,	or 70 A
2 = 50 + 40,	or 90 A
3 = 50 + 60,	or 110 A
4 = 50 + 80,	or 130 A
5 = 50 + 100,	or 150 A
6 = 50 + 120,	or 170 A
7 = 50 + 140,	or 190 A
8 = 50 + 160,	or 210 A
9 = 50 + 180,	or 230 A
10 = 50 + 200,	or 250 A

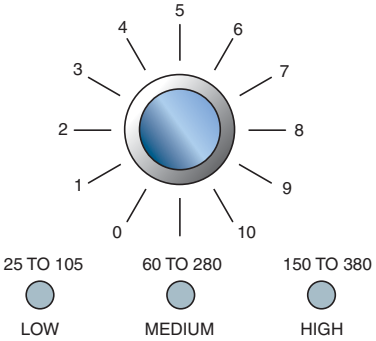

Figure 1.19 Practice 1.1

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 1-1

Estimating Amperages

Module 1
Key Indicator 1, 3, 4

Module 4
Key Indicator 3

Using a pencil and paper and the amperage ranges given in this practice (or from machines in the shop), you will estimate the amperage when the knob is at the 1/4, 1/2, and 3/4 settings, **Figure 1.19.**

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 1-2

Calculating Amperages

Module 1
Key Indicator 1, 3, 4

Module 4
Key Indicator 3

Using a pencil and paper or a calculator, and the amperage ranges given in this practice (or from machines in the shop), you will calculate the amperages for each of the following knob settings: 1, 4, 7, 9, 2.3, 5.7, and 8.5.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Movable-coil or Movable-core Machines

Movable-coil or movable-core machines are adjusted by turning a handwheel that moves the internal parts closer together or farther apart. The adjustment may also be made by moving a lever, **Figure 1.20.** These machines may have a high and low range, but they do not have a fine adjusting knob. The closer the primary and secondary coils are, the greater is the induced current; the greater the distance between the coils, the smaller the induced current, **Figure 1.21.** Moving the core in concentrates more of the magnetic force on the secondary coil, thus increasing the current. Moving the core out allows the field to disperse, and the current is reduced, **Figure 1.22.**

Figure 1.20
A movable core-type welding machine
Courtesy of Lincoln Electric Company

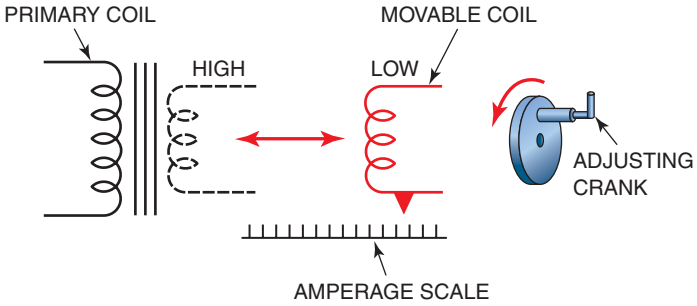

Figure 1.21
Movable coil

Inverter Machines

Inverter welding machines are much smaller than other types of machines of the same amperage range. This smaller size makes the welder much more portable and increases the energy efficiency. In a standard welding transformer, the iron core used to concentrate the magnetic field in the coils must be a specific size. The size of the iron core is determined by the length of time it takes for the magnetic field to build and collapse. By using solid-state electronic parts, the incoming power in an inverter welder is changed from 60 cycles a second to several thousand cycles a second. This higher frequency allows the use of a transformer that may be as light as 7 lb and still do the work of a standard transformer weighing 100 lb. Additional electronic parts remove the high frequency for the output welding power.

The use of electronics in the inverter-type welder allows it to produce any desired type of welding power. Before the invention of this machine, each type of welding required a separate machine. Now a single welding machine can produce the specific type of current needed for shielded metal arc welding, gas tungsten arc welding, gas metal arc welding, and plasma arc cutting. Because the machine can be light enough to be carried closer to work, shorter welding cables can be used. The welder does not have to walk as far to adjust the machine. Welding machine power wire is cheaper than welding cables. Some manufacturers produce machines that can be stacked so that when you need a larger machine all you have to do is add another unit to your existing welder, **Figure 1.23.**

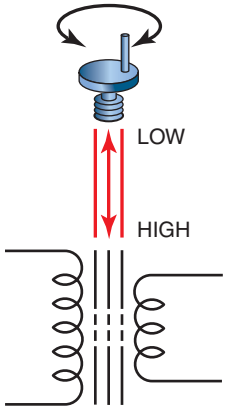

Figure 1.22
Movable core

GENERATORS AND ALTERNATORS

Generators and alternators both produce welding electricity from a mechanical power source. Both devices have an armature that rotates and a stator that is stationary. As a wire moves through a magnetic force field, electrons in the wire are made to move, producing electricity.

Figure 1.23
Typical 300-A inverter-type power supply weighing only 70 lb
Courtesy of Arcon Welding, L.L.C.

In an alternator, magnetic lines of force rotate inside a coil of wire, **Figure 1.24.** An alternator can produce AC only. In a generator, a coil of wire rotates inside a magnetic field. A generator produces DC. It is possible for alternators to use diodes to change the AC to DC for welding. In generators, the welding current is produced on the armature and is picked up with brushes, **Figure 1.25.** In alternators, the welding current is produced on the stator, and only the small current for the electromagnetic force field goes across the brushes. Therefore, the brushes in an alter-

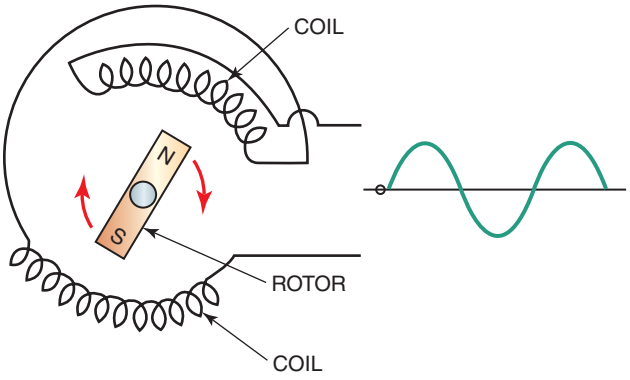

Figure 1.24
Schematic diagram of an alternator

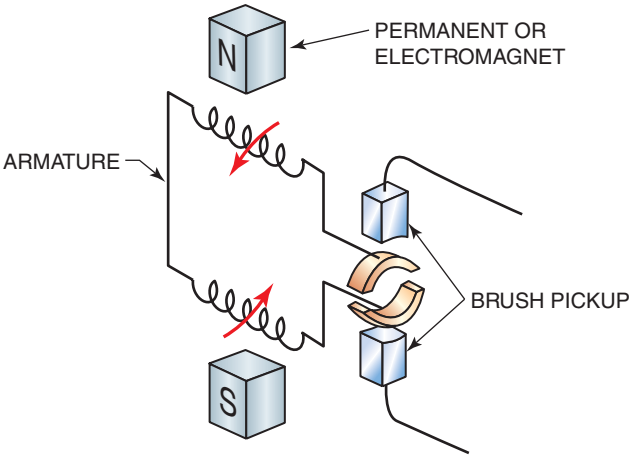

Figure 1.25
Diagram of a generator

(A) (B)

Figure 1.26
Portable engine generator welder
Courtesy of Lincoln Electric Company

nator are smaller and last longer. Alternators can be smaller in size and lighter in weight than generators and still produce the same amount of power.

Engine-driven generators and alternators may run at the welding speed all the time, or they may have an option that reduces their speed to an idle when welding stops. This option saves fuel and reduces wear on the welding machine. To strike an arc when using this type of welder, stick the electrode to the work for a second. When you hear the welding machine (welder) pick up speed, remove the electrode from the work and strike an arc. In general, the voltage and amperage are too low to start a weld, so shorting the electrode to the work should not cause the electrode to stick. A timer can be set to control the length of time that the welder maintains speed after the arc is broken. The time should be set long enough to change electrodes without losing speed.

Portable welders often have 110-V or 220-V plug outlets, which can be used to run grinders, drills, lights, and other equipment. The power provided may be AC or DC. If DC is provided, only equipment with brush-type motors or tungsten light bulbs can be used. If the plug is not specifically labeled 110 V AC, check the owner's manual before using it for such devices as radios or other electronic equipment. A typical portable welder is shown in **Figure 1.26.**

It is recommended that a routine maintenance schedule for portable welders be set up and followed. By checking the oil, coolant, battery, filters, fuel, and other parts, the life of the equipment can be extended. A checklist can be posted on the welder, **Table 1.2.**

RECTIFIERS

Alternating-welding current can be converted to direct current by using a series of rectifiers. A **rectifier** allows current to flow in one direction only, **Figure 1.27.**

Table 1.2 Portable Welder Checklist. The owner's manual should be checked for any additional items that might need attention.

Check Each Day before Starting
Oil level
Water level
Fuel level
Check Each Monday
Battery level
Cables
Fuel line filter
Check at Beginning of Month
Air filter
Belts and hoses
Change oil and filter
Check Each Fall
Antifreeze
Test battery
Pack wheel bearings
Change gas filter

If one rectifier is added, the welding power appears as shown in **Figure 1.28.** It would be difficult to weld with pulsating power such as this. A series of rectifiers, known as a bridge rectifier, can modify the alternating current so that it appears as shown in **Figure 1.29.**

Rectifiers become hot as they change AC to DC. They must be attached to a heat sink and cooled by having air blown over them. The heat produced by a rectifier reduces the power efficiency of the welding machine. **Figure 1.30** shows the amperage dial of a typical machine. Notice that at the same dial settings for AC and DC, the DC is at a lower amperage. The difference in amperage (power) is due to heat lost in the rectifiers. The loss in power makes operation with AC more efficient and less expensive compared to DC.

A DC adapter for small AC machines is available from manufacturers. For some types of welding, AC does not work properly.

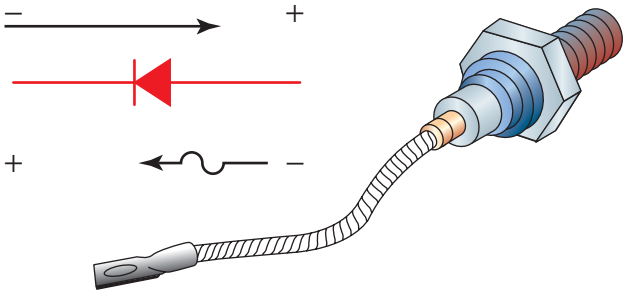

Figure 1.27
Rectifier

Figure 1.28
One rectifier in a welding power supply results in pulsating power

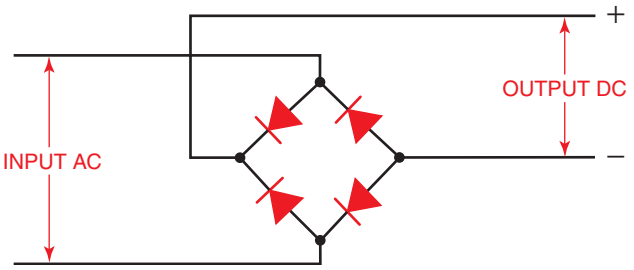

Figure 1.29
Bridge rectifier

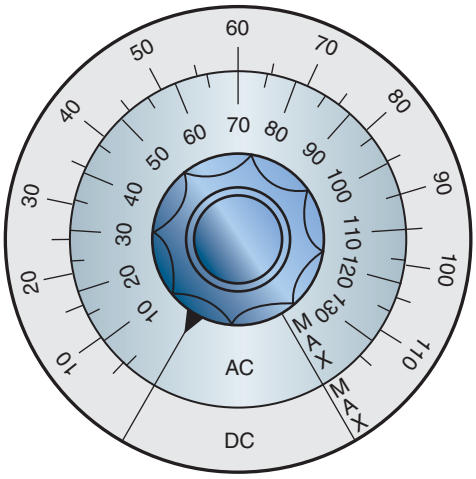

Figure 1.30
Typical dial on an AC-DC transformer rectifier welder

DUTY CYCLE

Welding machines produce internal heat at the same time that they produce the welding current. Except for automatic welding machines, welders are rarely used every minute for long periods of time. The welder must take time to change electrodes, change positions, or change parts. Shielded metal arc welding never continues for long periods of time.

The **duty cycle** is the percentage of time a welding machine can be used continuously. A 60% duty cycle means that out of any 10 minutes, the machine can be used for a total of six minutes at the maximum rated current. When providing power at this level, it must be cooled off for four minutes out of every 10 minutes. The duty cycle increases as the amperage is lowered and decreases for higher amperages, **Figure 1.31.** Most welding machines weld at a 60% rate or less. Therefore, most manufacturers list

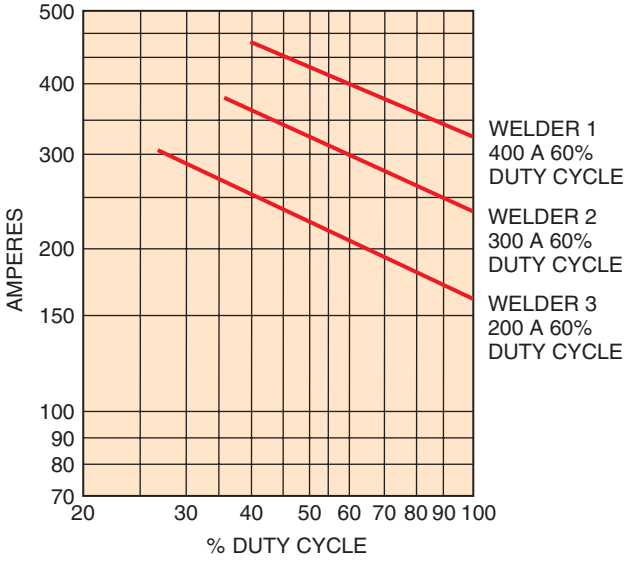

Figure 1.31
Duty cycle of a typical shielded metal arc welding machine

the amperage rating for a 60% duty cycle on the nameplate that is attached to the machine. Other duty cycles are given on a graph in the owner's manual.

The manufacturing cost of power supplies increases in proportion to their rated output and duty cycle. To reduce their price, it is necessary to reduce either their rating or their duty cycle. For this reason, some home-hobby welding machines may have duty cycles as low as 20% even at a low welding setting of 90 A to 100 A. The duty cycle on these machines should never be exceeded because a buildup of the internal temperature can cause the transformer insulation to break down, damaging the power supply.

PRACTICE 1-3

Reading Duty Cycle Chart

Module 1
Key Indicator 1, 3, 4

Module 4
Key Indicator 3

Using a pencil and paper and the duty cycle chart in Figure 1.31 (or one from machines in the shop), you will determine the following:

Welder 1: maximum welding amperage percent duty cycle at maximum amperage
Welder 2: maximum welding amperage percent duty cycle at maximum amperage
Welder 3: maximum welding amperage percent duty cycle at maximum amperage
Welder 1: maximum welding amperage at 100% duty cycle
Welder 2: maximum welding amperage at 100% duty cycle
Welder 3: maximum welding amperage at 100% duty cycle

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

WELDING CABLES AND LEADS

The terms **welding cables** and **welding leads** mean the same thing. Cables to be used for welding must be flexible, well insulated, and the correct size for the job. Most welding cables are made from stranded copper wire. Some manufacturers sell a newer type of cable made from aluminum wires. The aluminum wires are lighter and less expensive than copper. Because aluminum as a conductor is not as good as copper for a given wire size, the aluminum wire should be one size larger than would be required for copper.

The insulation on welding cables is exposed to hot sparks, flames, grease, oils, sharp edges, impact, and other types of wear. To withstand such wear, only specially manufactured insulation should be used for welding cable. Several new types of insulation are available that give longer service against these adverse conditions.

As electricity flows through a cable, the resistance to the flow causes the cable to heat up and increase the voltage drop. To minimize the loss of power and prevent overheating, the electrode cable and work cable must be the correct size. **Table 1.3** lists the minimum size cable that is required for each amperage and length. Large welding lead sizes make electrode

Table 1.3 Copper and Aluminum Welding Lead Sizes

	Amperes		Copper Welding Lead Sizes								
	ft	m	100	150	200	250	300	350	400	450	500
Length of Cable	50	15	2	2	2	2	1	1/0	1/0	2/0	2/0
	75	23	2	2	1	1/0	2/0	2/0	3/0	3/0	4/0
	100	30	2	1	1/0	2/0	3/0	4/0	4/0		
	125	38	2	1/0	2/0	3/0	4/0				
	150	46	1	2/0	3/0	4/0					
	175	53	1/0	3/0	4/0						
	200	61	1/0	3/0	4/0						
	250	76	2/0	4/0							
	300	91	3/0								
	350	107	3/0								
	400	122	4/0								

	Amperes		Aluminum Welding Lead Sizes								
	ft	m	100	150	200	250	300	350	400	450	500
Length of Cable	50	15	2	2	1/0	2/0	2/0	3/0	4/0		
	75	23	2	1/0	2/0	3/0	4/0				
	100	30	1/0	2/0	4/0						
	125	38	2/0	3/0							
	150	46	2/0	3/0							
	175	53	3/0								
	200	61	4/0								
	225	69	4/0								

manipulation difficult. Smaller cable can be spliced to the electrode end of a large cable to make it more flexible. This whip-end cable must not be more than 10 ft (3 m) long.

CAUTION

A splice in a cable should not be within 10 ft (3 m) of the electrode because of the possibility of electrical shock.

PRACTICE 1-4

Determining Welding Lead Sizes

Using a pencil and paper and Table 1.3, Copper and Aluminum Welding Lead Sizes, you will determine the following:

1. the minimum copper welding lead size for a 200-A welder with 100-ft (30-m) leads
2. the minimum copper welding lead size for a 125-A welder with 225-ft (69-m) leads
3. the maximum length aluminum welding lead that can carry 300 A

Module 1
Key Indicator 1, 3, 4

Module 4
Key Indicator 3

Splices and end lugs are available from suppliers. Be sure that a good electrical connection is made whenever splices or lugs are used. A poor electrical connection will result in heat buildup, voltage drop, and poor service from the cable. Splices and end lugs must be well insulated against possible electrical shorting, **Figure 1.32.**

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Figure 1.32
Power lug protection is provided by insulators
Courtesy of ESAB Welding & Cutting Products

CAUTION

Never dip a hot electrode holder in water to cool it off. The problem causing the holder to overheat should be repaired.

Figure 1.33
The amperage capacity of an electrode holder is often marked on its side
Courtesy of Tweco®, a Thermadyne® Company

Module 1
Key Indicator 1, 3, 4

Module 4
Key Indicator 1, 2

ELECTRODE HOLDERS

The electrode holder should be of the proper amperage rating and in good repair for safe welding. Electrode holders are designed to be used at their maximum amperage rating or less. Higher amperage values will cause the holder to overheat and burn up. If the holder is too large for the amperage range being used, manipulation is hard, and operator fatigue increases. Make sure that the correct amperage holder is chosen, **Figure 1.33.**

A properly sized electrode holder can overheat if the jaws are dirty or loose, or if the cable is loose. If the holder heats up, welding power is being lost. In addition, a hot electrode holder is uncomfortable to work with.

Replacement springs, jaws, insulators, handles, screws, and other parts are available to keep the holder in good working order, **Figure 1.34.** To prevent excessive damage to the holder, welding electrodes should not be burned too short. A 2-in. (51-mm) electrode stub is short enough to minimize electrode waste and save the holder.

PRACTICE 1-5

Repairing Electrode Holders

Using the manufacturer's instructions for your type of electrode holder, required hand tools, and replacement parts, you will do the following:

1. Remove the electrode holder from the welding cable.
2. Remove the jaw insulating covers.
3. Replace the jaw insulating covers.

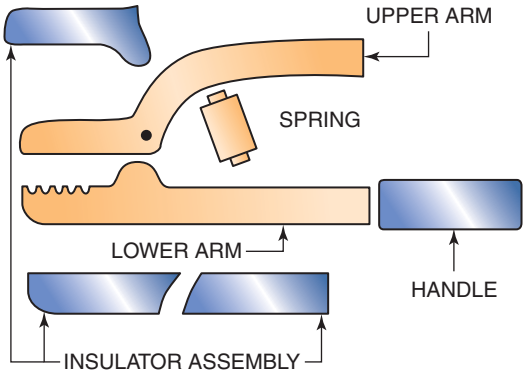

Figure 1.34
Replaceable parts of an electrode holder

4. Reconnect the electrode holder to the welding cable.
5. Turn on the welding power or reconnect the welding cable to the welder.
6. Make a weld to ensure that the repair was made correctly.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

CAUTION

Before starting any work, make sure that the power to the welder is off and locked off or the welding lead has been removed from the machine.

WORK CLAMPS

The work clamp must be the correct size for the current being used, and it must clamp tightly to the material. Heat can build up in the work clamp, reducing welding efficiency, just as was previously described for the electrode holder. Power losses in the work clamp are often overlooked. The clamp should be carefully touched occasionally to find out if it is getting hot.

In addition to power losses due to poor work lead clamping, a loose clamp may cause arcing that can damage a part. Improper work clamp placement may also contribute to arc blow. To avoid arc blow, keep the work clamp situated as close to the weld as practical, and weld away from it. If the part is to be moved during welding, a swivel-type work clamp may be needed, **Figure 1.35.** It may be necessary to weld a tab to thick parts so that the work lead can be clamped to the tab, **Figure 1.36.**

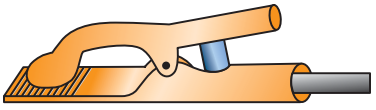

Figure 1.35
A work clamp may be attached to the workpiece

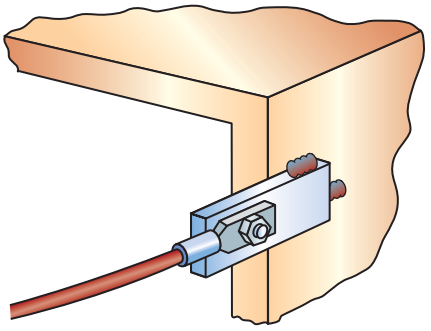

Figure 1.36
Tack welded ground to part

SETUP

Module 1
Key Indicator 1

Module 4
Key Indicator 3

Arc welding machines should be located near the welding site, but far enough away so that they are not covered with spark showers. The machines may be stacked to save space, but there must be enough room between the machines to ensure the air can circulate so as to keep the machines from overheating. The air that is circulated through the machine should be as free as possible of dust, oil, and metal filings. Even in a good location, the power should be turned off periodically and the machine blown out with compressed air, **Figure 1.37.**

The welding machine should be located away from cleaning tanks and any other sources of corrosive fumes that could be blown through it. Water leaks must be fixed and puddles cleaned up before a machine is used.

Power to the machine must be fused, and a power shut-off switch provided. The switch must be located so that it can be reached in an emergency without touching either the machine or the welding station. The machine case or frame must be grounded.

The welding cables should be sufficiently long to reach the work station but not so long that they must always be coiled. Cables should not be placed on the floor in aisles or walkways. If cables must cross a walkway, the cable must be installed overhead, or it must be protected by a ramp, **Figure 1.38.** The welding machine and its main power switch should be off while a person is installing or working on the cables.

The work station must be free of combustible materials. Screens or curtains should be provided to protect other workers from the arc light.

The welding cable should never be wrapped around arms, shoulders, waist, or any other part of the body. If the cable was caught by any moving equipment, such as a forklift, crane, or dolly, a welder could be pulled off balance or more seriously injured. If it is necessary to hold the weight off the cable so that the welding can more easily be done, a free hand can be used. The cable should be held so that if it is pulled it can be easily released.

Check the surroundings before starting to weld. If heavy materials are being moved in the area around you, there should be a safety watch. A safety watch can warn a person of danger while that person is welding.

CAUTION

The cable should never be tied to scaffolding or ladders. If the cable is caught by moving equipment, the scaffolding or ladder may be upset, causing serious personal injury.

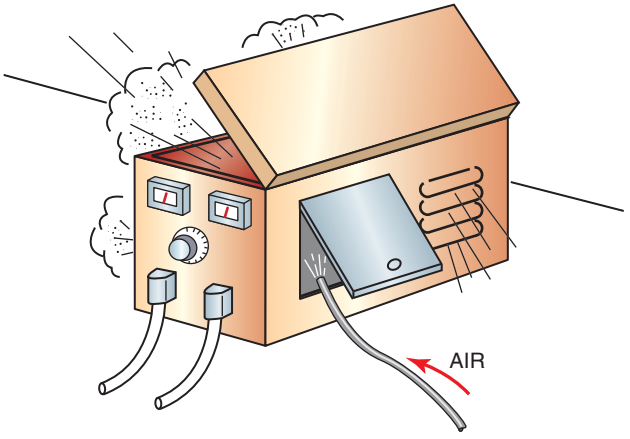

Figure 1.37
Slag, chips from grinding, and dust must be blown out occasionally so that they will not start a fire or cause a short-out or other types of machine failure

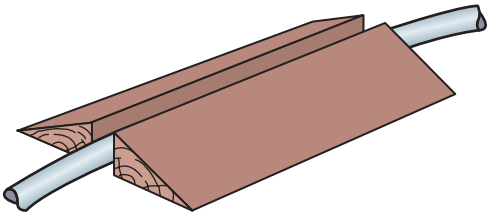

Figure 1.38
To prevent people from tripping, when cables must be placed in walkways, lay two blocks of wood beside the cables

SUMMARY

Understanding the scientific theory of electricity and magnetism will aid you in understanding how the welding currents are produced and their reactions to changes in their physical surroundings. Understanding electromagnetic phenomena will aid you in controlling arc blow. Failure to control arc blow can result in weld failures. In addition, understanding electricity will help you interpret information given on manufacturers' tables, charts, and equipment specifications.

Before starting any new job or welding operation, be sure to check with the equipment manufacturer's safety guidelines for proper operation and maintenance. Follow all recommended guidelines.

Keeping your work area clean and orderly will help prevent accidents.

REVIEW

1. Describe the welding current.
2. What produces the heat during a shielded metal arc weld?
3. Voltage can be described as _____.
4. Amperage can be described as _____.
5. Wattage can be described as _____.
6. What determines the exact temperature of the shielded metal welding arc?
7. Does all of the heat produced by an SMA weld stay in the weld? Why or why not?
8. What do the following abbreviations mean: AC, DCEN, DCEP, DCSP, and DCRP?
9. Sketch a welding machine, an electrode lead, an electrode holder, an electrode, a work lead, and work connected for DCEN welding.
10. Sketch a welding machine, an electrode lead, an electrode holder, an electrode, a work lead, and work connected for DCEP welding.
11. Why is SMA welding current referred to as *constant current?*
12. What is the higher voltage at the electrode before the arc is struck called? What is its advantage to welding?
13. Referring to the graph in Figure 1.9, what would the voltage be for the CC power supply at 110 A? What would the watts be?
14. How does arc blow affect welding?
15. How can arc blow be controlled?
16. How does a welding transformer work?
17. What are *taps* on a welding transformer?
18. What would the approximate amperage setting be if a welder were set to the high range (150 A to 350 A) and the fine adjustment knob were pointing at 5 on a 10-point scale?
19. What are the advantages of the inverter-type welding power supply?
20. What is the difference between the welding current produced by alternators and the one produced by generators?
21. What are the advantages of alternators over generators?
22. What must be checked before using the 110-V power plug on a portable welder?
23. What is meant by a *welder's duty cycle?*

24. Why must a welding machine's duty cycle never be exceeded?
25. Why is copper better than aluminum for welding cables?
26. A splice in a welding cable should never be any closer than _____ to the electrode holder.
27. Why must the electrode holder be correctly sized?
28. What can cause a properly sized electrode holder to overheat?
29. What problem can occur if welding machines are stacked or placed too closely together?
30. Why must welding cables never be tied to scaffolding or ladders?

CHAPTER

2 Shielded Metal Arc Welding of Plate

OBJECTIVES

After completing this chapter, the student should be able to

- demonstrate safe SMAW work practices
- strike an arc at a specific point
- contrast the problems that can result if the welding current is set too low or too high
- select the correct diameter of welding electrode for a weld
- compare a weld bead's shape for width, reinforcement, and appearance with differing heat inputs
- define arc length, and describe the effects of using too short or too long an arc length
- compare a leading electrode angle to a trailing electrode angle
- list three characteristics of the weld bead that can be controlled by the movement or weaving of the welding electrode
- demonstrate ten weave patterns for weld beads
- match various SMAW electrodes to the four filler metal groups
- define stringer beads and tell how they are used
- make a vertical up stringer bead and a horizontal stringer bead
- make a welded square butt joint in the flat, vertical up, and the horizontal positions
- on an edge joint, make a flat weld, a vertical down weld, a vertical up weld, a horizontal weld, and an overhead weld
- on an outside corner joint, make a flat weld, a vertical down weld, a vertical up weld, a horizontal weld, and an overhead weld
- make fillet welds in lap joints in all positions
- make fillet welds in tee joints in all positions
- make groove welds in plate in all positions without backing

KEY TERMS

amperage range
arc length
cellulose-based fluxes
electrode angle
lap joint
mineral-based fluxes
rutile-based fluxes
square butt joint
stringer bead
tee joint
weave pattern

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports or records
Key Indicator 2: Performs housekeeping duties
Key Indicator 3: Follows verbal instructions to complete work assignments
Key Indicator 4: Follows written instructions to complete work assignments

Module 2: Safety and Health of Welders

Key indicator 1: Demonstrates proper use and inspection of personal protective equipment (PPE)
Key Indicator 2: Demonstrates safe operation practices in the work area
Key Indicator 3: Demonstrates proper use and inspection of ventilation equipment
Key Indicator 4: Demonstrates proper hot zone operation
Key Indicator 5: Demonstrates proper work actions for work in confined spaces
Key Indicator 6: Understands proper use of precautionary labeling and MSDS information

Module 4: Shielded Metal Arc Welding (SMAW)

Key Indicator 1: Performs safety inspections of SMAW equipment and accessories
Key Indicator 3: Sets up SMAW operations on carbon steel
Key Indicator 4: Operates SMAW equipment on carbon steel
Key Indicator 5: Makes SMAW fillet welds, in all positions, on carbon steel
Key Indicator 6: Makes SMAW groove welds, in all positions, on carbon steel

Module 9: Welding Inspection and Testing Principles

Key Indicator 2: Examines tacks, root passes, intermediate layers, and completed welds

INTRODUCTION

Shielded metal arc welding (SMAW), or stick welding, is a common method used to join plate. This method provides a high temperature and concentration of heat, which allows a small molten weld pool to be built up quickly. The addition of filler metal from the electrode adds reinforcement and increases

the strength of the weld. SMAW can be performed on almost any type of metal 1/8 in. (3 mm) thick or thicker. A minimum amount of equipment is required, and it can be portable.

High-quality welds can be consistently produced on almost any type of metal and in any position. The quality of the welds produced depends largely upon the skill of the welder. Developing the necessary skill level requires practice. However, practicing the welds repeatedly without changing techniques will not aid in developing the required skills. Each time a weld is completed it should be evaluated, and then a change should be made in the technique to improve the next weld.

PRACTICE 2-1

Shielded Metal Arc Welding Safety

Using a welding workstation, welding machine, welding electrodes, welding helmet, eye and ear protection, welding gloves, proper work clothing, and any special protective clothing that may be required, demonstrate, to your instructor and other students, the safe way to prepare yourself and the welding workstation for welding. Include in your demonstration appropriate references to burn protection, eye and ear protection, material safety data sheets, ventilation, electrical safety, general work clothing, special protective clothing, and area cleanup. More information on welding safety can be found in Chapter 2 in *Welding Skills, Processes and Practices for Entry-Level Welders: Book One* and ANSI Z49.1, *Safety in Welding, Cutting, and Allied Processes.*

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 5, 6

Module 4
Key Indicator 1

Module 9
Key Indicator 2

EXPERIMENT 2-1

Striking the Arc

Using a properly set up and adjusted arc welding machine, the proper safety protection as demonstrated in Practice 2-1, E6011 or E6103 welding electrodes having a 1/8-in. (3-mm) core wire diameter measured at the bare end of the electrode, and one piece of mild steel plate, 1/4 in. (6 mm) thick, you will practice striking an arc, **Figure 2.1.**

With the electrode held over the plate, lower your helmet. Scratch the electrode across the plate (like striking a large match), **Figure 2.2.** As the arc is established, slightly raise the electrode to the desired arc length. Hold the arc in one place until the molten weld pool builds to the desired size. Slowly lower the electrode as it burns off and move it forward to start the bead.

If the electrode sticks to the plate, quickly squeeze the electrode holder lever to release the electrode. Break the electrode free by bending it back and forth a few times. Do not touch the electrode without gloves because it will still be hot. If the flux breaks away very far from the end of the electrode, throw out the electrode because restarting the arc will be very difficult, **Figure 2.3.**

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 4

Module 9
Key Indicator 2

Once you are able to easily strike an arc and make a weld, try to strike the arc where it will be re-melted by the weld you are making. Arc strikes on the metal's surface that are not covered up by the weld are considered to be weld defects by most codes.

Break the arc by rapidly raising the electrode after completing a 1-in. (25-mm) weld bead. Restart the arc as you did before, and make another short weld. Repeat this process until you can easily start the arc each time.

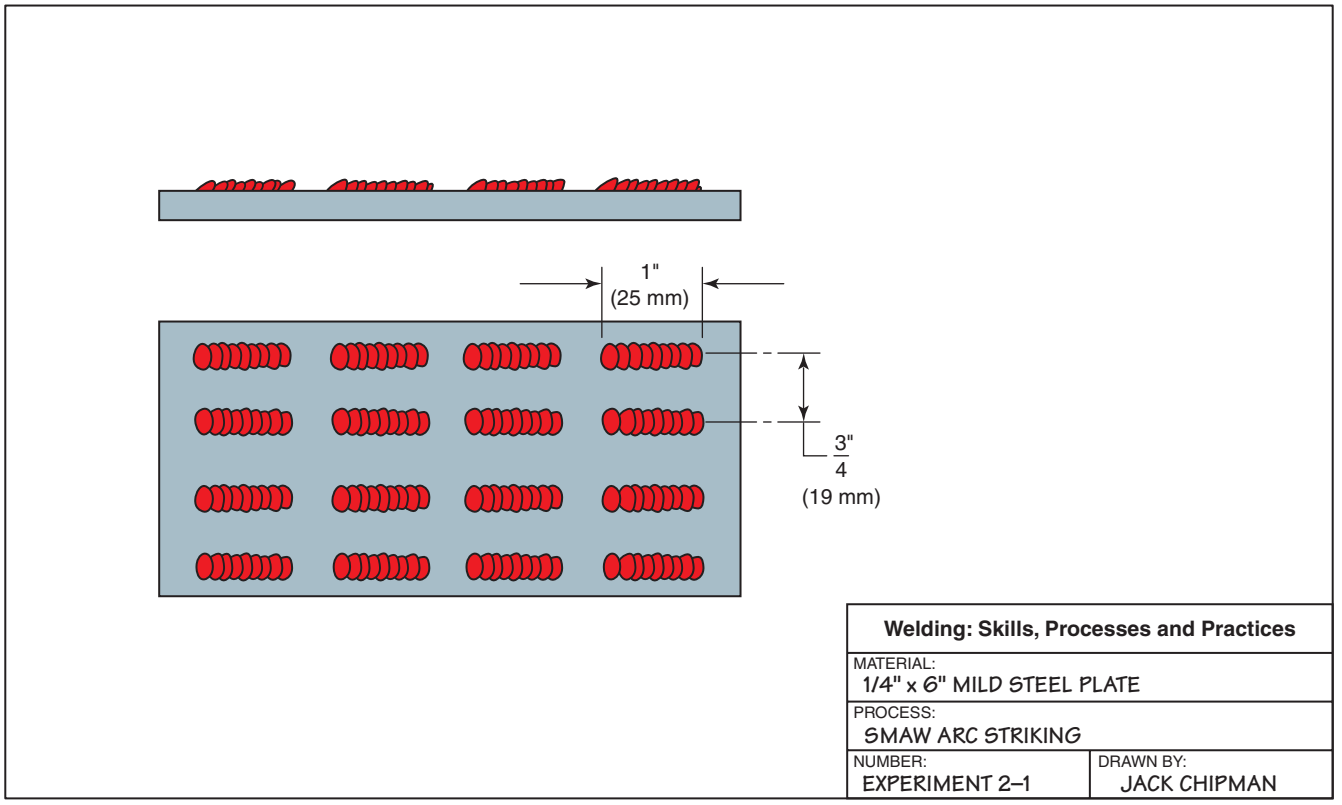

Figure 2.1
Striking an arc and running short beads

Figure 2.2
Striking the arc
Courtesy of Larry Jeffus

Figure 2.3
If the flux is broken off the end completely or on one side, the arc can be erratic or forced to the side
Courtesy of Larry Jeffus

Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

EXPERIMENT 2-2

Striking the Arc Accurately

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 4

Module 9
Key Indicator 2

Using the same materials and setup as described in Experiment 2-1, you will start the arc at a specific spot in order to prevent damage to the surrounding plate.

Hold the electrode over the desired starting point. After lowering your helmet, swiftly bounce the electrode against the plate, **Figure 2.4.** A lot of practice is required to develop the speed and skill needed to prevent the electrode from sticking to the plate.

A more accurate method of starting the arc involves holding the electrode steady by resting it on your free hand like a pool cue. The electrode is rapidly pushed forward so that it strikes the metal exactly where it should. This is an excellent method of striking an arc. Striking an arc in an incorrect spot may cause damage to the base metal.

Practice starting the arc until you can start it within 1/4 in. (6 mm) of the desired location. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

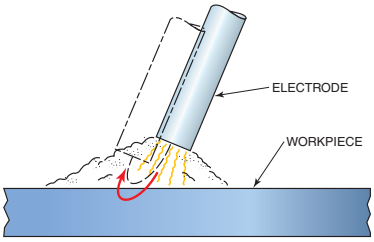

Figure 2.4
Striking the arc on a spot

EFFECT OF TOO-HIGH OR TOO-LOW CURRENT SETTINGS

Each welding electrode must be operated in a specified current (amperage) range, **Table 2.1.** Welding with the current set too low results in poor fusion and poor arc stability, **Figure 2.5.** The weld may have slag or gas inclusions because the molten weld pool was not fluid long enough for the flux to react. Little or no penetration of the weld into the base plate may also be evident. With the current set too low, the arc length is very short. A very short arc length results in frequent shorting and sticking of the electrode.

The core wire of the welding electrode is limited in the amount of current it can carry. As the current is increased, the wire heats up because of electrical resistance. This preheating of the wire causes some of the chemicals in the covering to be burned out too early, **Figure 2.6.** The loss of the proper balance of elements causes poor arc stability. This condition leads to spatter, porosity, and slag inclusions.

Table 2.1 Welding Amperage Range for Common Electrode Types and Sizes

Electrode	Classification					
Size	E6010	E6011	E6012	E6013	E7016	E7018
3/32 in. (2.4 mm)	40–80	50–70	40–90	40–85	75–105	70–110
1/8 in. (3.2 mm)	70–130	85–125	75–130	70–120	100–150	90–165
5/32 in. (4 mm)	110–165	130–160	120–200	130–160	140–190	125–220

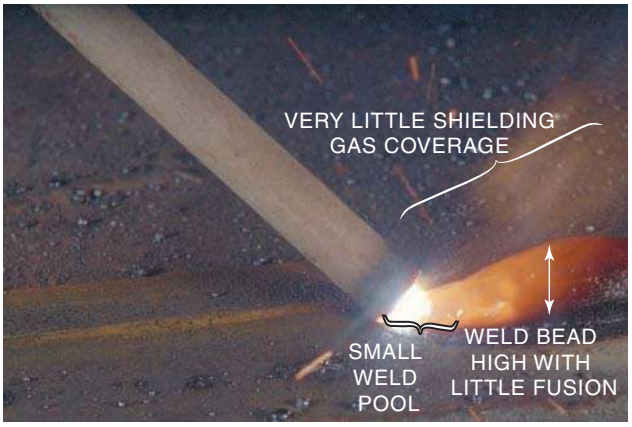

Figure 2.5
Welding with the amperage set too low
Courtesy of Larry Jeffus

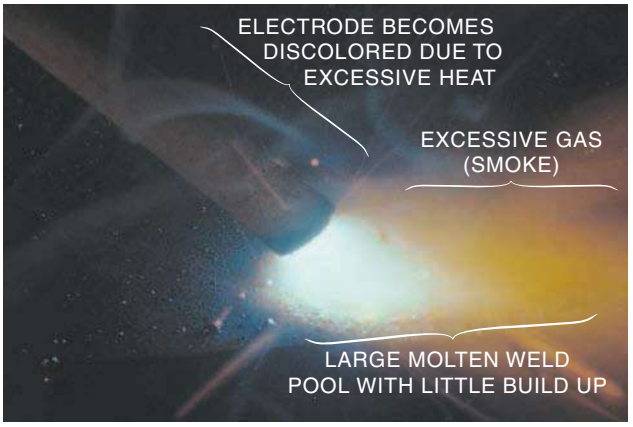

Figure 2.6
Welding with too high an amperage
Courtesy of Larry Jeffus

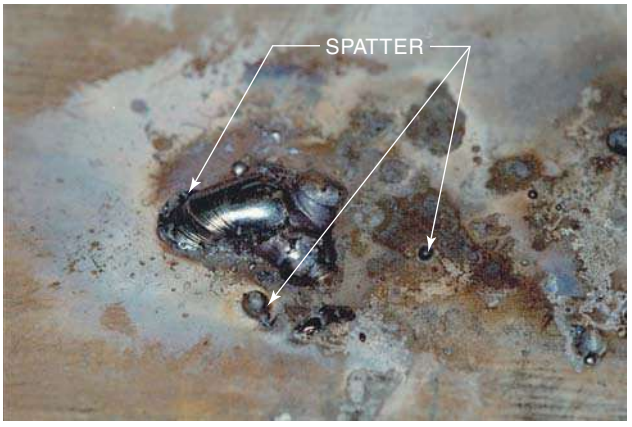

Figure 2.7
Hard weld spatter is fused to base metal and is difficult to remove
Courtesy of Larry Jeffus

An increase in the amount of spatter is also caused by longer arc lengths. The weld bead made at a high amperage setting is wide and flat with deep penetration. The spatter is excessive and is mostly hard. The spatter is called hard because it fuses to the base plate and is difficult to remove, **Figure 2.7.** The electrode covering is discolored more than 1/8 in. (3 mm) to 1/4 in. (6 mm) from the end of the electrode. Extremely high settings may also cause the electrode to discolor, crack, glow red, or burn.

EXPERIMENT 2-3

Effect of Amperage Changes on a Weld Bead

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 3, 4

Module 9
Key Indicator 2

For this experiment, you will need an arc welding machine, welding gloves, safety glasses, welding helmet, appropriate clothing, E6011 or E6013 welding electrodes having a 1/8-in. (3-mm) diameter, and one piece of mild steel plate, 1/4 in. (6 mm) to 1/2 in. (13 mm) thick. You will observe what happens to the weld bead when the amperage settings are raised and lowered.

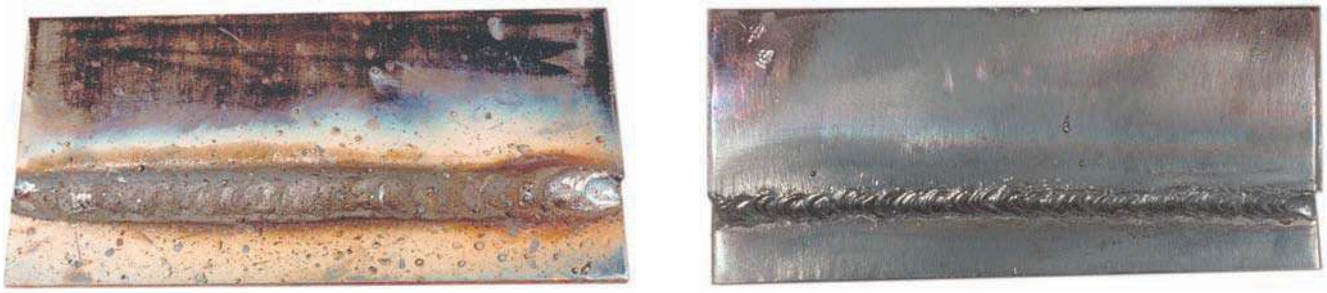

Figure 2.8
Weld before cleaning and after cleaning
Courtesy of Larry Jeffus

Starting with the machine set at approximately 90 A AC or DCRP, strike an arc and make a weld 1 in. (25 mm) long. Break the arc. Raise the current setting by 10 A, strike an arc, and make another weld 1 in. (25 mm) long. Repeat this procedure until the machine amperage is set at the maximum value.

Replace the electrode and reset the machine to 90 A. Make a weld 1 in. (25 mm) long. Stop and lower the current setting by 10 A. Repeat this procedure until the machine amperage is set at a minimum value.

Cool and chip the plate, comparing the different welds for width, buildup, molten weld pool size, spatter, slag removal, and penetration, **Figure 2.8.** In addition, compare the electrode stubs. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

CAUTION

Do not change the current settings during welding. A change in the setting may cause arcing inside the machine, resulting in damage to the machine.

ELECTRODE SIZE AND HEAT

The selection of the correct size of welding electrode for a weld is determined by some or all of the following: the skill of the welder, the thickness of the metal to be welded, the size of the metal, and welding codes or standards. Using small diameter electrodes requires less skill than using large diameter electrodes. The deposition rate, or the rate at which the weld metal is added to the weld, is slower when small diameter electrodes are used. Small diameter electrodes will make acceptable welds on thick plate, but more time is required to make the weld.

Large diameter electrodes may overheat the metal if they are used with thin or small pieces of metal. To determine if a weld is too hot, watch the shape of the trailing edge of the molten weld pool, **Figure 2.9.** Rounded ripples indicate the weld is cooling uniformly and that the heat is not excessive. If the ripples are pointed, the weld is cooling too slowly because of excessive heat. Extreme overheating can cause a burn-through. Once a burn-through occurs, it is hard to repair.

To correct an overheating problem, a welder can turn down the amperage, use a shorter arc, travel at a faster rate, use a chill plate (a large piece of metal used to absorb excessive heat), or use a smaller electrode at a lower current setting.

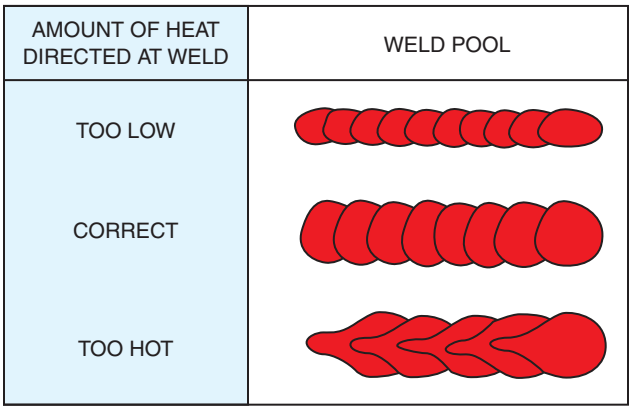

Figure 2.9
The effect on the shape of the molten weld pool caused by the heat input

EXPERIMENT 2-4

Excessive Heat

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 3, 4

Module 9
Key Indicator 2

Using a properly set up and adjusted arc welding machine, the proper safety protection, E6011 or E6013 welding electrodes having a 1/8-in. (3-mm) diameter, and three pieces of mild steel plate, 1/8 in. (3 mm), 3/16 in. (5 mm), and 1/4 in. (6 mm) thick, you will observe the effects of overheating on the weld. Make a stringer weld on each of the three plates using the same amperage setting, travel rate, and arc length for each weld. Cool and chip the welds. Then compare the weld beads for width, reinforcement, and appearance.

Using the same amperage settings, make additional welds on the 1/8-in. (3-mm) and 3/16-in. (5-mm) plates. Vary the arc lengths and travel speeds for these welds. Cool and chip each weld and compare the beads for width, reinforcement, and appearance. Make additional welds on the 1/8-in. (3-mm) and 3/16-in. (5-mm) plates, using the same arc length and travel speed as in the earlier part of this experiment, but at a lower amperage setting. Cool and chip the welds and compare the beads for width, reinforcement, and appearance.

Welders often use the terms *heat* and *amperage* as interchangeable when they are speaking about making changes to the welding current. For example, a welder may say, "Turn up the heat a little," or "Turn up the amperage a little." In both cases what's being asked is that the welding amperage be increased a little.

The plates should be cooled between each weld so that the heat from the previous weld does not affect the test results. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

ARC LENGTH

The **arc length** is the distance the arc must jump from the end of the electrode to the plate or weld pool surface. As the weld progresses, the electrode becomes shorter as it is consumed. To maintain a constant arc length, the electrode must be lowered continuously. Maintaining a con-

Figure 2.10
Welding with too short an arc length
Courtesy of Larry Jeffus

Figure 2.11
Welding with too long an arc length
Courtesy of Larry Jeffus

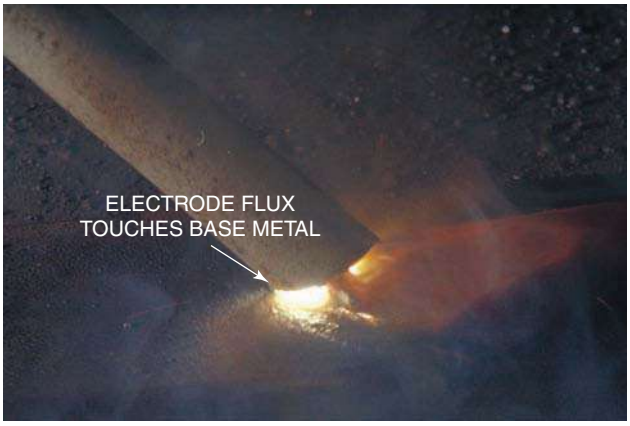

Figure 2.12
Welding with a drag technique
Courtesy of Larry Jeffus

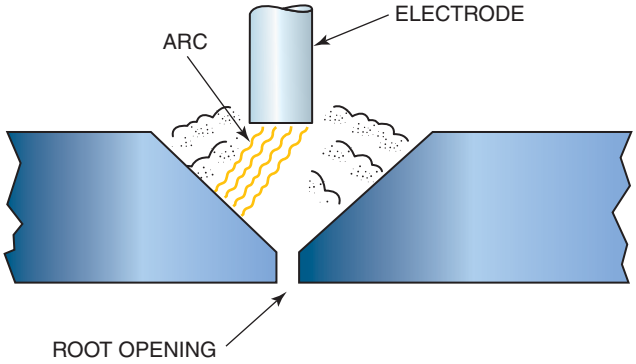

Figure 2.13
The arc may jump to the closest metal, reducing root penetration

stant arc length is important, as too great a change in the arc length will adversely affect the weld.

As the arc length is shortened, metal transferring across the gap may short out the electrode, causing it to stick to the plate. The weld that results from a short arc is narrow and has a high buildup, **Figure 2.10.**

Long arc lengths produce more spatter because the metal being transferred may drop outside of the molten weld pool. The weld is wider and has little buildup, **Figure 2.11.**

There is a narrow range for the arc length in which stability is maintained, metal transfer is smooth, spatter is minimized, and the bead shape is controlled. Factors affecting the length are the type of electrode, joint design, metal thickness, and current setting.

Some welding electrodes, such as E7024, have a thick flux covering. The rate at which the covering melts is slow enough to permit the electrode coating to be rested against the plate. The arc burns back inside the covering as the electrode is dragged along touching the joint, **Figure 2.12.** For that reason, electrodes like E7024 are sometimes referred to as drag electrodes, or drag rods. For this type of welding electrode, the arc length is maintained by the electrode covering.

An arc will jump to the closest metal conductor. On joints that are deep or narrow, the arc is pulled to one side and not to the root, **Figure 2.13.** As

a result, the root fusion is reduced or may be nonexistent, thus causing a poor weld. If a very short arc is used, the arc is forced into the root for better fusion.

Because shorter arcs produce less heat and penetration, they are best suited for use on thin metal or thin-to-thick metal joints. Using this technique, metal as thin as 16 gauge can be arc welded easily. Higher amperage settings are required to maintain a short arc that gives good fusion with a minimum of slag inclusions. The higher settings, however, must be within the **amperage range** for the specific electrode.

Finding the correct arc length often requires some trial and adjustment. Most welding jobs require an arc length of 1/8 in. (3 mm) to 3/8 in. (10 mm) when using a 1/8-in. electrode, but this distance varies. It may be necessary to change the arc length when welding to adjust for varying welding conditions.

EXPERIMENT 2-5

Effect of Changing the Arc Length on a Weld

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 3, 4

Module 9
Key Indicator 2

Using an arc welding machine, welding gloves, safety glasses, welding helmet, appropriate clothing, E6011 or E6013 welding electrodes having a 1/8-in. (3-mm) diameter, and one piece of mild steel plate, 1/4 in. (6 mm) to 1/2 in. (13 mm) thick, you will observe the effect of changing the arc length on a weld.

Starting with the welding machine set at approximately 90 A AC or DCRP, strike an arc and make a weld 1 in. (25 mm) long. Continue welding while slowly increasing the arc length until the arc is broken. Restart the arc and make another weld 1 in. (25 mm) long. Welding should again be continued while slowly shortening the arc length until the arc stops. Quickly break the electrode free from the plate, or release the electrode by squeezing the lever on the electrode holder.

Cool and chip both welds. Compare both welding beads for width, reinforcement, uniformity, spatter, and appearance. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

ELECTRODE ANGLE

The **electrode angle** is measured from the electrode to the surface of the metal. The term used to identify the electrode angle is affected by the direction of travel, generally leading or trailing, **Figure 2.14.** The relative angle is important because there is a jetting force blowing the metal and flux from the end of the electrode to the plate.

Leading Angle

A leading electrode angle pushes molten metal and slag ahead of the weld, **Figure 2.15.** When welding in the flat position, caution must be taken to prevent overlap and slag inclusions. The solid metal ahead of the weld cools and solidifies the molten filler metal and slag before they can melt

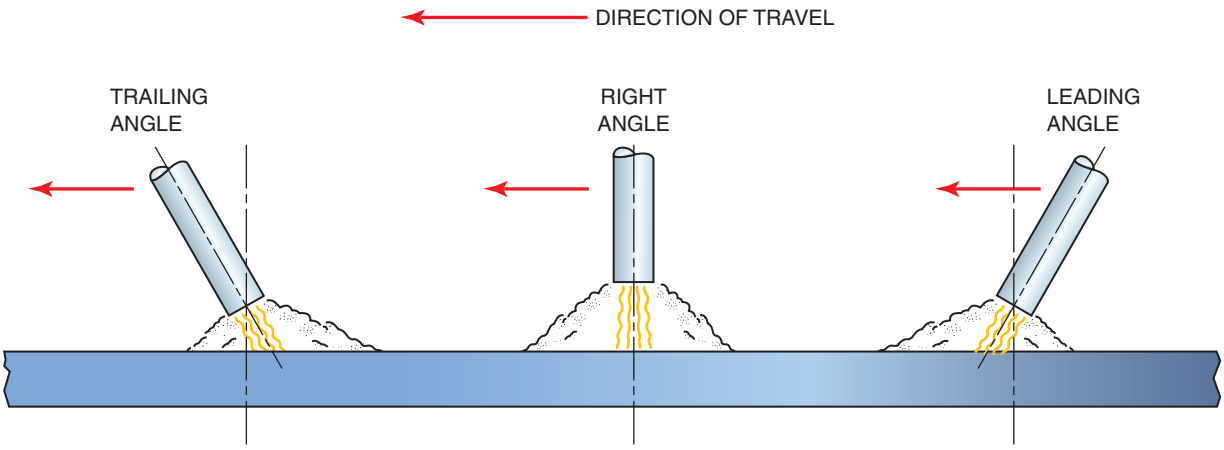

Figure 2.14
Direction of travel and electrode angle

the solid metal. This rapid cooling prevents the metals from fusing together, **Figure 2.16.** As the weld passes over this area, heat from the arc may not melt it. As a result, some overlap and slag inclusions are left.

The following are suggestions for preventing cold lap and slag inclusions:

- Use as little leading angle as possible.
- Ensure that the arc melts the base metal completely, **Figure 2.17.**
- Use a penetrating-type electrode that causes little buildup.
- Move the arc back and forth across the molten weld pool to fuse both edges.

A leading angle can be used to minimize penetration or to help hold molten metal in place for vertical welds, **Figure 2.18.**

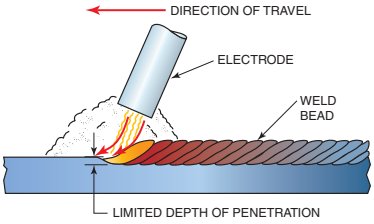

Figure 2.15
Leading, lag, or pushing electrode angle

Trailing Angle

A trailing electrode angle pushes the molten metal away from the leading edge of the molten weld pool toward the back where it solidifies, **Figure 2.19.** As the molten metal is forced away from the bottom of the

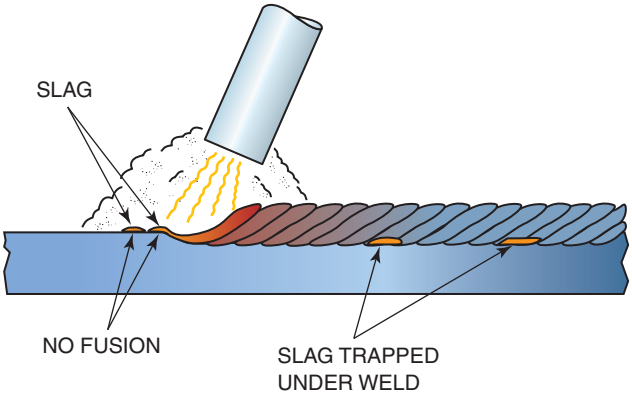

Figure 2.16
Some electrodes, such as E7018, may not remove the deposits ahead of the molten weld pool, resulting in discontinuities within the weld

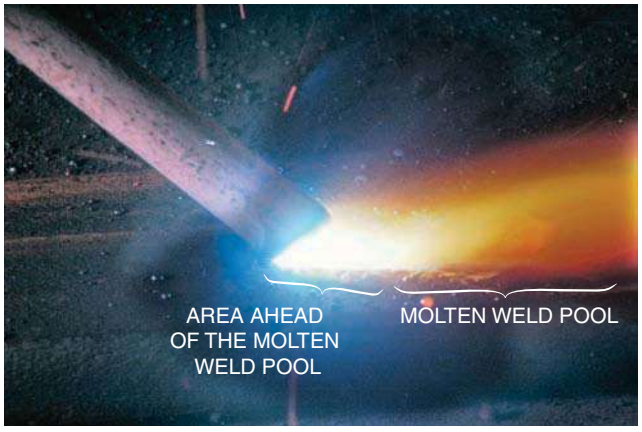

Figure 2.17
Metal being melted ahead of the molten weld pool helps to ensure good weld fusion
Courtesy of Larry Jeffus

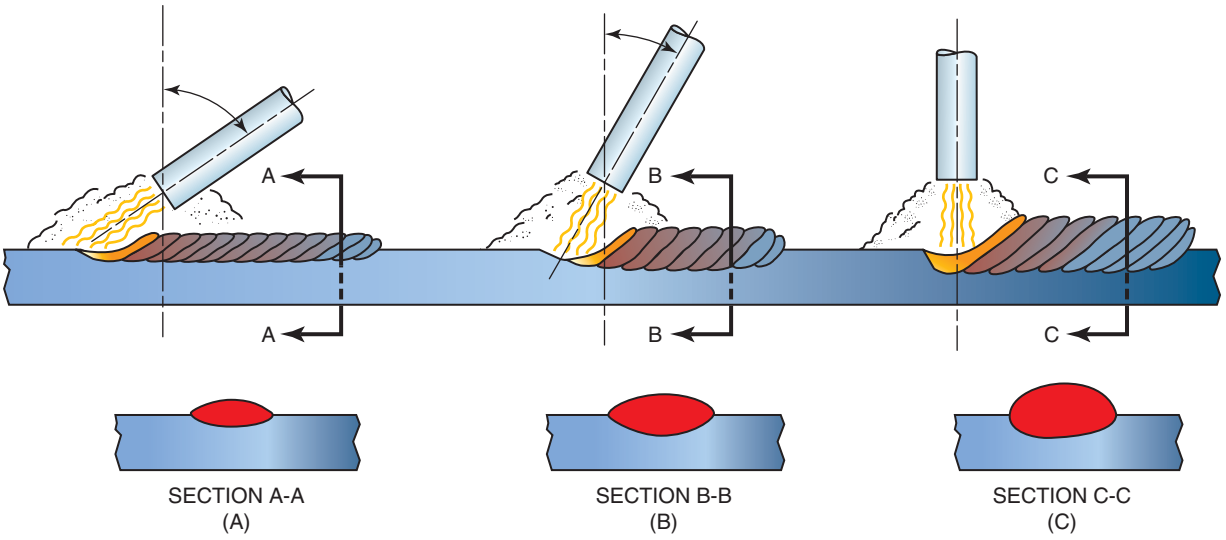

Figure 2.18
Effect of a leading angle on weld bead buildup, width, and penetration. As the angle increases toward the vertical position (C), penetration increases

weld, the arc melts more of the base metal, which results in deeper penetration. The molten metal pushed to the back of the weld solidifies and forms reinforcement for the weld, **Figure 2.20.**

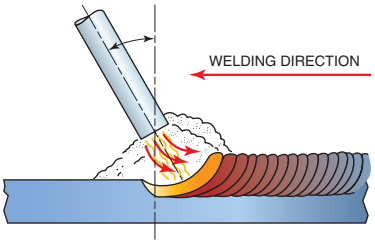

Figure 2.19
Trailing electrode angle

EXPERIMENT 2-6

Effect of Changing the Electrode Angle on a Weld

Using a properly set up and adjusted arc welding machine, the proper safety protection, E6011 welding electrodes having a 1/8-in. (3-mm) diameter, and one piece of mild steel plate, 1/4 in. (6 mm) to 1/2 in. (13 mm) thick, you will observe the effect of changes in the electrode angle on a weld.

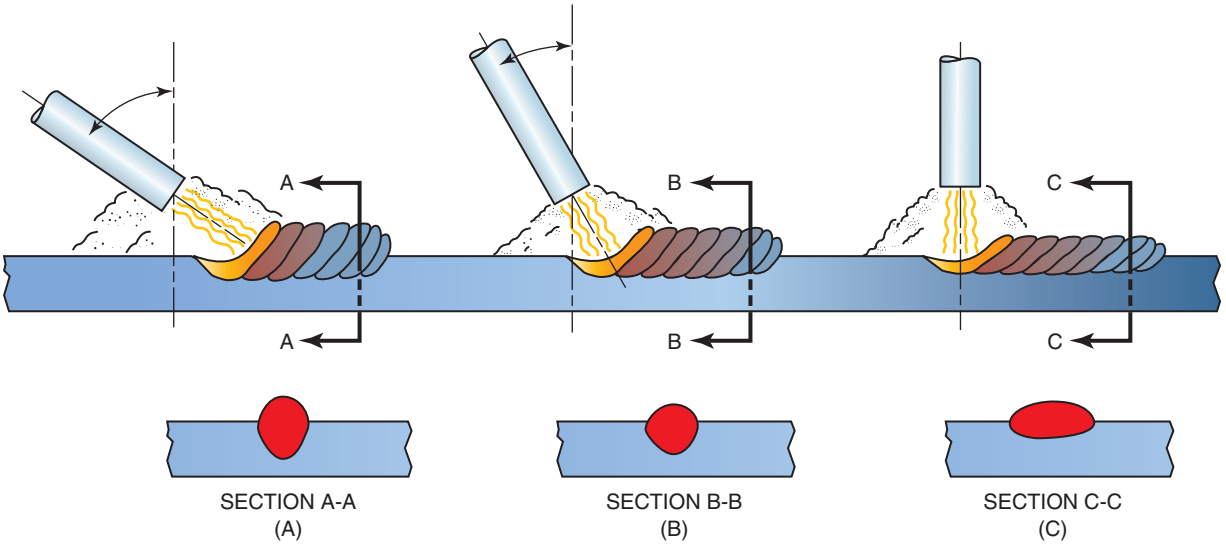

Figure 2.20
Effect of a trailing angle on weld bead buildup, width, and penetration. Section A-A shows more weld buildup due to a greater angle of the electrode

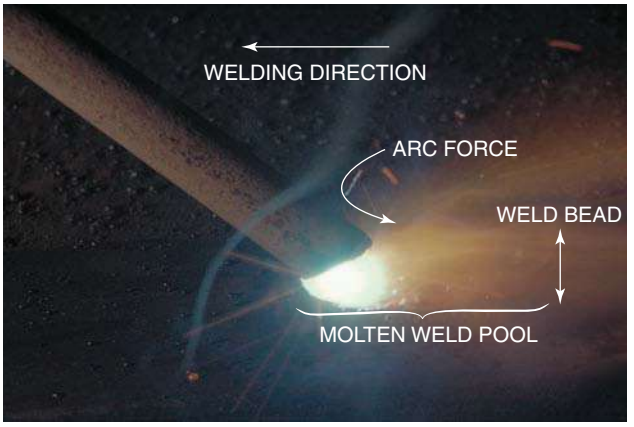

Figure 2.21
Welding with a trailing angle
Courtesy of Larry Jeffus

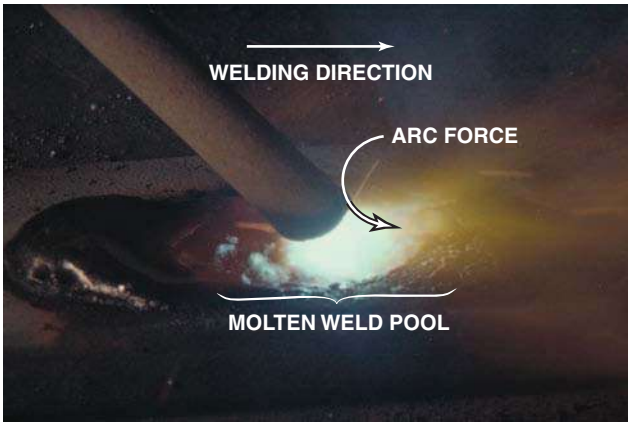

Figure 2.22
Welding with a leading angle
Courtesy of Larry Jeffus

Start welding with a sharp trailing angle. Make a weld about 1 in. (25 mm) long. Closely observe the molten weld pool at the points shown in **Figure 2.21.** Slowly increase the electrode angle and continue to observe the weld.

When you reach a 90° electrode angle, make a weld about 1 in. (25 mm) long. Observe the parts of the molten weld pool as shown in Figure 2.21.

Continue welding and change the electrode angle to a sharp leading angle. Observe the molten weld pool at the points shown in **Figure 2.22.**

During this experiment, you must maintain a constant arc length, travel speed, and weave pattern if the observations and results are to be accurate.

Cool and chip the weld. Compare the weld bead for uniformity in width, reinforcement, and appearance. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 3, 4

Module 9
Key Indicator 2

ELECTRODE MANIPULATION

The movement or weaving of the welding electrode can control the following characteristics of the weld bead: penetration, buildup, width, porosity, undercut, overlap, and slag inclusions. The exact **weave pattern** for each weld is often the personal choice of the welder. However, some patterns are especially helpful for specific welding situations. The pattern selected for a flat (1G) butt joint is not as critical as is the pattern selection for other joints and other positions.

Many weave patterns are available for the welder to use. **Figure 2.23** shows ten different patterns that can be used for most welding conditions.

The circular pattern is often used for flat position welds on butt, tee, outside corner joints, and for buildup or surfacing applications. The circle can be made wider or longer to change the bead width or penetration, **Figure 2.24.**

The "C" and square patterns are both good for most 1G (flat) welds, but can also be used for vertical (3G) positions. These patterns can also be

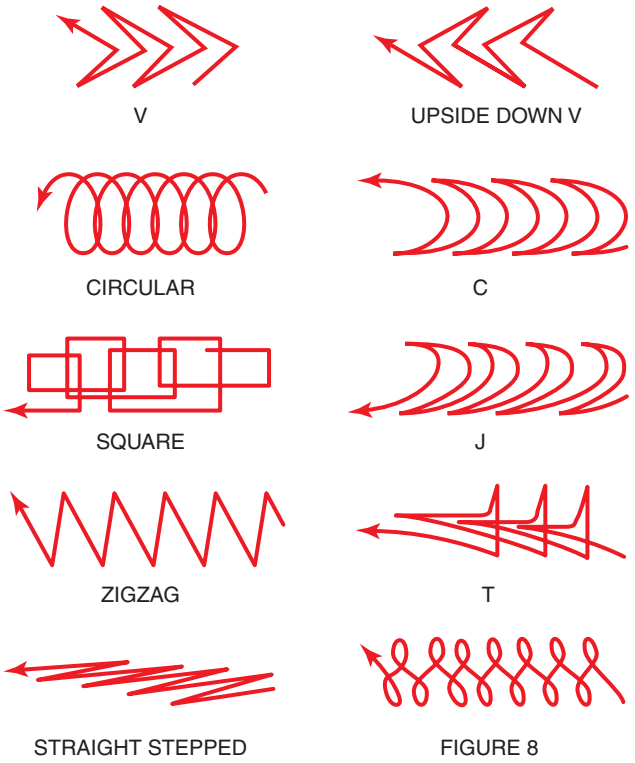

Figure 2.23
Weave patterns

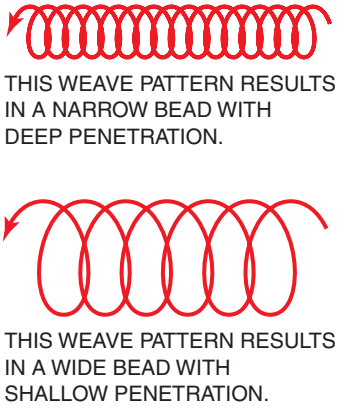

Figure 2.24
Changing the weave pattern width to change the weld bead characteristics

used if there is a large gap to be filled when both pieces of metal are nearly the same size and thickness.

The "J" pattern works well on flat (1F) lap joints, all vertical (3G) joints, and horizontal (2G) butt and lap (2F) welds. This pattern allows the heat to be concentrated on the thicker plate, **Figure 2.25.** It also allows the reinforcement to be built up on the metal deposited during the first part of the pattern. As a result, a uniform bead contour is maintained during out-of-position welds.

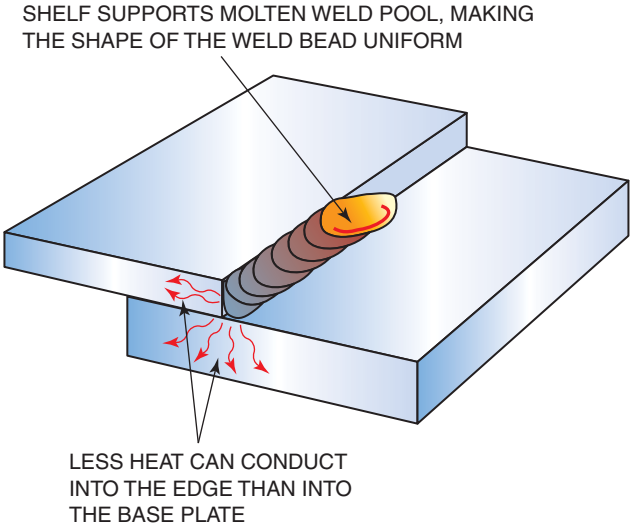

Figure 2.25
The "J" pattern allows the heat to be concentrated on the thicker plate

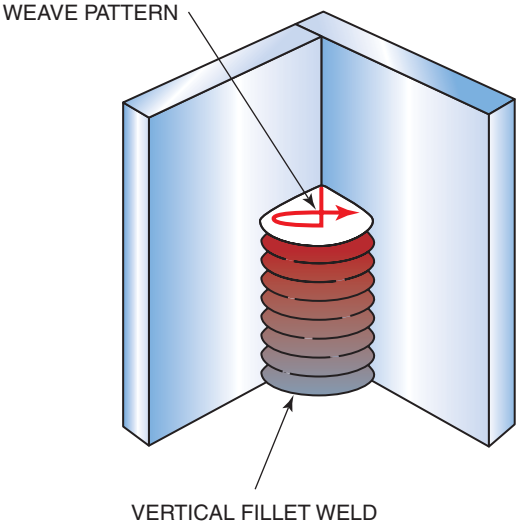

Figure 2.26
"Inverted T" pattern

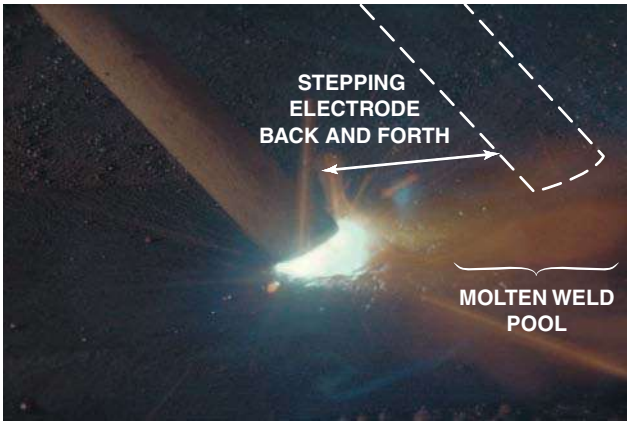

Figure 2.27
The electrode is moved slightly forward and then returned to the weld pool
Courtesy of Larry Jeffus

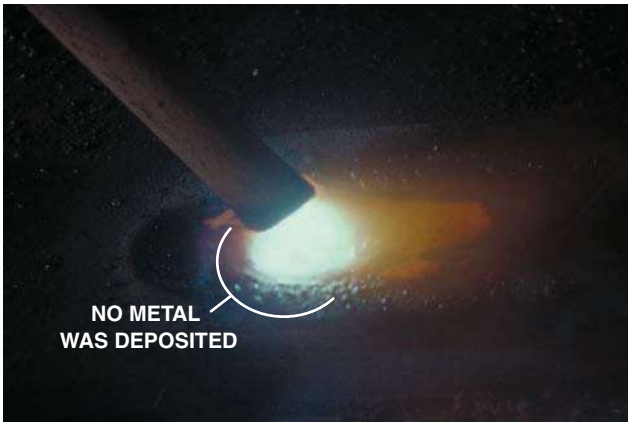

Figure 2.28
The electrode does not deposit metal or melt the base metal
Courtesy of Larry Jeffus

The inverted "T" pattern works well with fillet welds in the vertical (3F) and overhead (4F) positions, **Figure 2.26.** It also can be used for deep groove welds for the hot pass. The top of the "T" can be used to fill in the toe of the weld to prevent undercutting.

The straight step pattern can be used for stringer beads, root pass welds, and multiple pass welds in all positions. For this pattern, the smallest quantity of metal is molten at one time as compared to other patterns. Therefore, the weld is more easily controlled. At the same time that the electrode is stepped forward, the arc length is increased so that no metal is deposited ahead of the molten weld pool, **Figure 2.27** and **Figure 2.28.** This action allows the molten weld pool to cool to a controllable size. In addition, the arc burns off any paint, oil, or dirt from the metal before it can contaminate the weld.

The figure-eight pattern and the zigzag pattern are used as cover passes in the flat and vertical positions. Do not weave more than 2-1/2 times the width of the electrode. These patterns deposit a large quantity of metal at one time. A shelf can be used to support the molten weld pool when making vertical welds using either of these patterns, **Figure 2.29.**

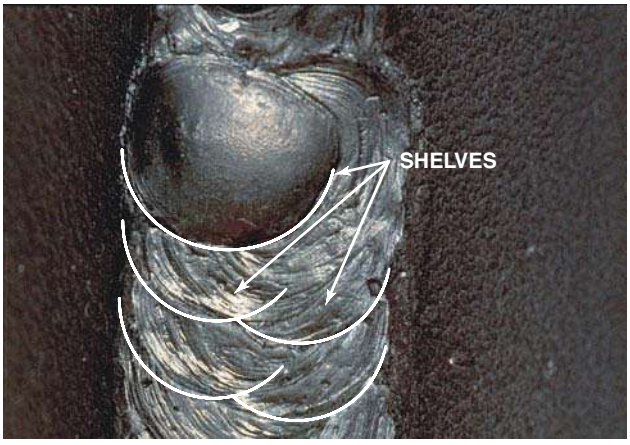

Figure 2.29
Using the shelf to support the molten pool for vertical welds
Courtesy of Larry Jeffus

POSITIONING OF THE WELDER AND THE PLATE

The welder should be in a relaxed, comfortable position before starting to weld. A good position is important for both the comfort of the welder and the quality of the welds. Welding in an awkward position can cause welder fatigue, which leads to poor welder coordination and poor-quality welds. Welders must have enough freedom of movement so that they do not need to change position during a weld. Body position changes should be made only during electrode changes.

When the welding helmet is down, the welder is blind to the surroundings. Due to the arc, the field of vision of the welder is also very limited. These factors often cause the welder to sway. To stop this swaying, the welder should lean against or hold on to a stable object. When welding, even if a welder is seated, touching a stable object will make that welder more stable and will make welding more relaxing.

Welding is easier if the welder can find the most comfortable angle. The welder should be in either a seated or a standing position in front of the welding table. The welding machine should be turned off. With an electrode in place in the electrode holder, the welder can draw a straight line along the plate to be welded. By turning the plate to several different angles, the welder should be able to determine which angle is most comfortable for welding, **Figure 2.30.**

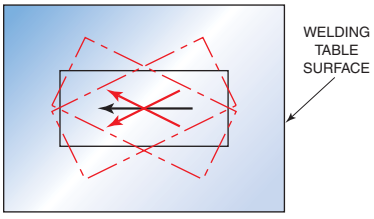

Figure 2.30
Change the plate angle to find the most comfortable welding position

PRACTICE WELDS

Practice welds are grouped according to the type of joint and the type of welding electrode. The welder or instructor should select the order in which the welds are made. The stringer beads should be practiced first in each position before the welder tries the different joints in each position. Some time can be saved by starting with the stringer beads. If this is done, it is not necessary to cut or tack the plate together, and a number of beads can be made on the same plate.

Students will find it easier to start with butt joints. The lap, tee, and outside corner joints are all about the same level of difficulty. Starting with the flat position allows the welder to build skills slowly so that out-of-position welds become easier to do. The horizontal tee and lap welds are almost as easy to make as the flat welds. Overhead welds are as simple to make as vertical welds, but they are harder to position. Horizontal butt welds are more difficult to perform than most other welds.

Electrodes

Arc welding electrodes used for practice welds in the following chapters are grouped into three filler metal (F-number) classes according to their major welding characteristics. E6012 and E6013 fall into the F2 group, E6010 and E6011 fall into the F3 group, and E7016 and E7018 fall into the F4 group. The F1 electrode group (not used in this chapter) is limited flat and horizontal fillet welds, E7024 and E7028 are examples of F1 electrodes.

F1 E7024 and E7028 Electrodes

These electrodes are primarily produced with iron-powder-based fluxes that are substantially thicker than the other groups. As a result, they have

very high deposition rates and produce welds with medium to low penetration, but are restricted to groove welds in the flat position and fillet welds in the flat or horizontal positions.

F2 E6012 and E6013 Electrodes

These electrodes have rutile-based fluxes, giving a smooth, easy arc with a thick slag left on the weld bead. They may be used in all positions and do not require special rod ovens. These electrodes are popular because slag is easily removed, and they may be used with all polarities to produce welds with medium to low penetration.

F3 E6010 and E6011 Electrodes

Both of these electrodes have cellulose-based fluxes. As a result, these electrodes have a forceful deep penetrating arc with little slag left on the weld bead. They do not require a rod oven, and they are often the electrode of choice when surface conditions on the base metal are less than optimal. F3 electrodes are also commonly used for open root welds on plate and pipe.

F4 E7016 and E7018 Electrodes

Both of these electrodes have a mineral-based flux. The resulting medium penetration arc is smooth and easy, with a very heavy slag left on the weld bead. F4 electrodes are also referred to as low hydrogen or low-hi electrodes. They require special handling and storage in a rod oven after being removed from their factory packaging. Refer to manufacturer's requirements or the applicable welding code for specific handling directions.

The cellulose- and rutile-based groups of electrodes have characteristics that make them the best electrodes for starting specific welds. The electrodes with the **cellulose-based fluxes** do not have heavy slags that may interfere with the welder's view of the weld. This feature is an advantage for flat tee and lap joints. Electrodes with the **rutile-based fluxes** (giving an easy arc with low spatter) are easier to control and are used for flat stringer beads and butt joints.

Unless a specific electrode has been required by a welding procedure specification (WPS), welders can select what they consider to be the best electrode for a specific weld. Without a WPS, the welder has the final choice. An accomplished welder can make defect-free welds on all types of joints using all types of electrodes in any weld position.

Electrodes with **mineral-based fluxes** should be the last choice. Welds with a good appearance are more easily made with these electrodes, but strong welds are hard to obtain. Without special care being taken during the start of the weld, porosity will be formed in the weld. **Figure 2.31** shows a starting tab used to prevent this porosity from becoming part of the finished weld.

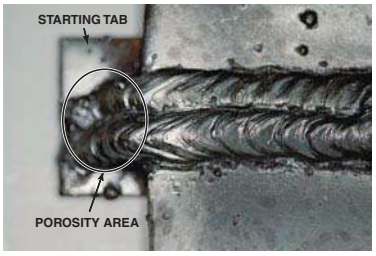

Figure 2.31
Porosity is found on the starting tab where it will not affect the weld
Courtesy of Larry Jeffus

STRINGER BEADS

A straight weld bead on the surface of a plate, with little or no side-to-side electrode movement, is known as a **stringer bead.** Stringer beads are used by students to practice maintaining arc length and electrode angle so that their welds will be straight, uniform, and free from defects. Stringer beads,

Figure 2.32
Stringer bead
Courtesy of Larry Jeffus

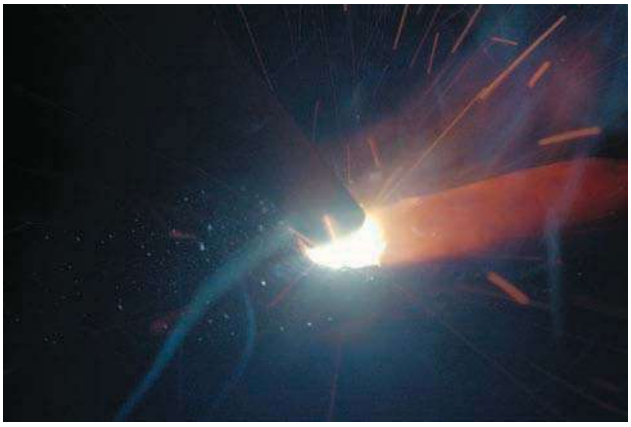

Figure 2.33
New welders frequently see only the arc and sparks from the electrode
Courtesy of Larry Jeffus

Figure 2.32, are also used to set the machine amperage and for buildup or surfacing applications. Stringer beads are the most commonly used type of bead for vertical, horizontal, and overhead welds.

The stringer bead should be straight. A beginning welder needs time to develop the skill of viewing the entire welding area. At first, the welder sees only the arc, **Figure 2.33.** With practice, the welder begins to see parts of the molten weld pool. After much practice, the welder will see the molten weld pool (front, back, and both sides), slag, buildup, and the surrounding plate, **Figure 2.34.** Often, at this skill level, the welder may not even notice the arc.

A straight weld is easily made once the welder develops the ability to view the entire welding zone. The welder will occasionally glance around to ensure that the weld is straight. In addition, it can be noted if the weld is uniform and free from defects. The ability of the welder to view the entire weld area is demonstrated by making consistently straight and uniform stringer beads.

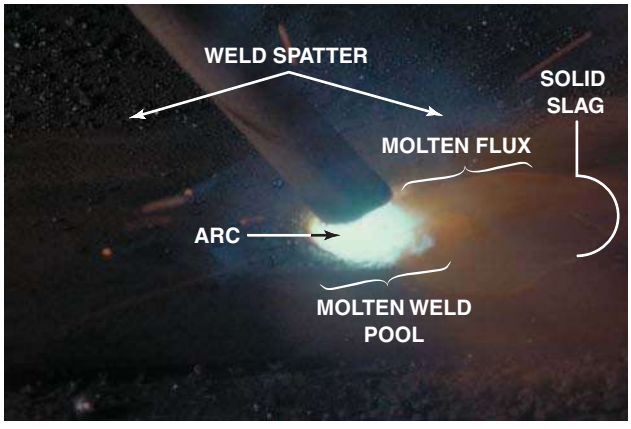

Figure 2.34
More experienced welders can see the molten pool, metal being transferred across the arc, and penetration into the base metal
Courtesy of Larry Jeffus

After making practice stringer beads, a variety of weave bead patterns should be practiced to gain the ability to control the molten weld pool when welding out of position.

PRACTICE 2-2

Straight Stringer Beads in the Flat Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using a properly set up and adjusted arc welding machine, proper safety protection as demonstrated in Practice 2-1, arc welding electrodes with a 1/8-in. (3-mm) diameter, and one piece of mild steel plate, 6 in. (152 mm) long × 1/4 in. (6 mm) thick, you will make straight stringer beads.

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

- Starting at one end of the plate, make a straight weld the full length of the plate.
- Watch the molten weld pool at this point, not the end of the electrode. As you become more skillful, it is easier to watch the molten weld pool.
- Repeat the beads with all three (F) groups of electrodes until you have consistently good beads.
- Cool, chip, and inspect the bead for defects after completing it. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-3

Stringer Beads in the Vertical Up Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using the same setup, materials, and electrodes as listed in Practice 2-2, you will make vertical up stringer beads. Start with the plate at a 45° angle.

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

This technique is the same as that used to make a vertical weld. However, a lower level of skill is required at 45°, and it is easier to develop your skill. After the welder masters the 45° angle, the angle is increased successively until a vertical position is reached, **Figure 2.35.**

Before the molten metal drips down the bead, the back of the molten weld pool will start to bulge, **Figure 2.36.** When this happens, increase the speed of travel and the weave pattern.

Cool, chip, and inspect each completed weld for defects. Repeat the beads as necessary with all three (F) groups of electrodes until consistently good beads are obtained in this position. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

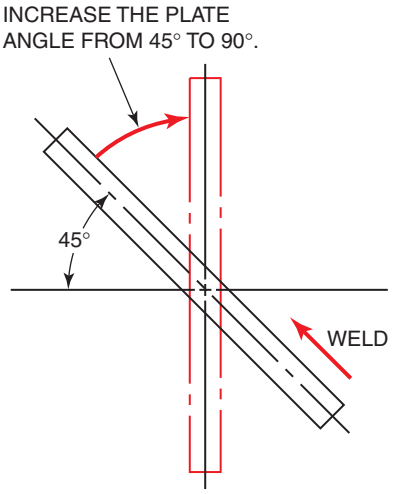

Figure 2.35
Once the 45° angle is mastered, the plate angle is increased successively until a vertical position (90°) is reached

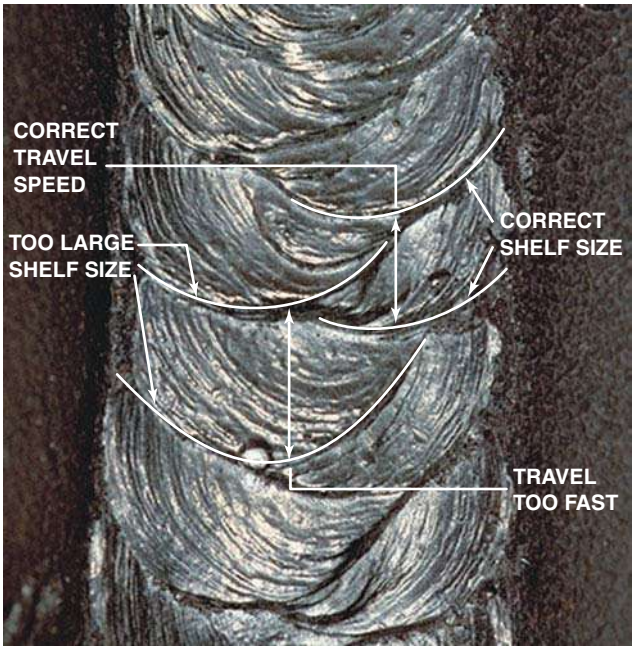

Figure 2.36
E7018 vertical up weld
Courtesy of Larry Jeffus

PRACTICE 2-4

Horizontal Stringer Beads Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-2, you will make horizontal stringer beads on a plate.

When the welder begins to practice the horizontal stringer bead, the plate may be reclined slightly, **Figure 2.37.** This placement allows the welder to build the required skill by practicing the correct techniques successfully. The "J" weave pattern is suggested for this practice. As the electrode is drawn along the straight back of the "J," metal is deposited. This metal supports the molten weld pool, resulting in a bead with a uniform contour, **Figure 2.38.**

Angling the electrode up and back toward the weld causes more metal to be deposited along the top edge of the weld. Keeping the bead small allows the surface tension to hold the molten weld pool in place.

Gradually increase the angle of the plate until it is vertical and the stringer bead is horizontal. Repeat the beads as needed with all three (F) groups of electrodes until consistently good beads are obtained in this position. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

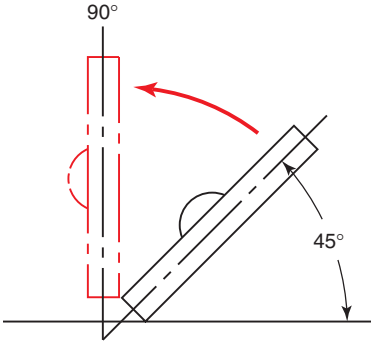

Figure 2.37
Change the plate angle as welding skill improves

SQUARE BUTT JOINT

The **square butt joint** is made by tack welding two flat pieces of plate together, **Figure 2.39.** The space between the plates is called the root opening or root gap. Changes in the root opening will affect penetration. As the

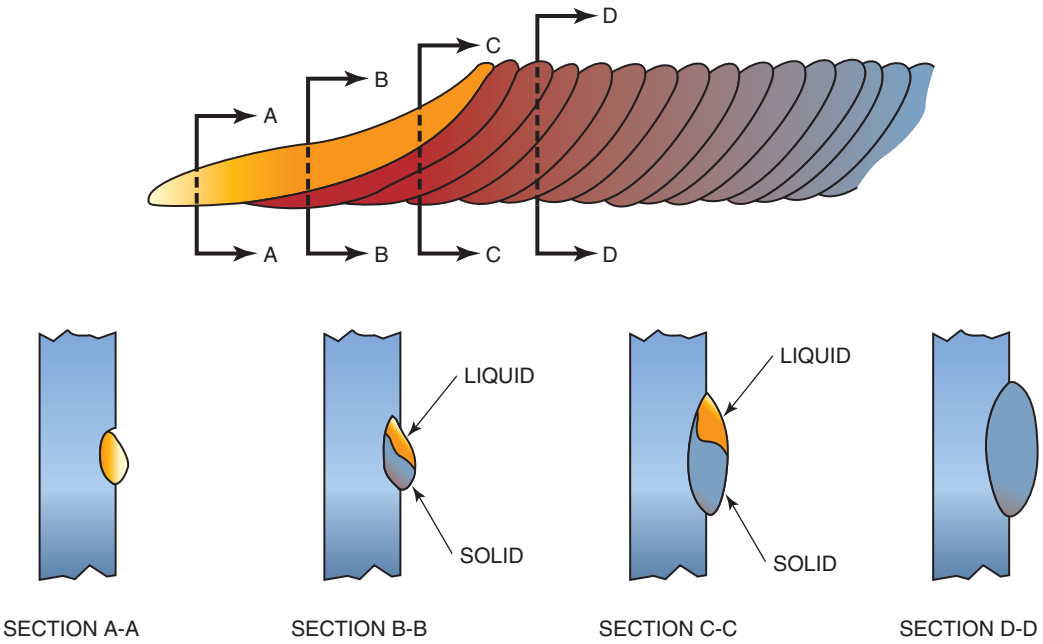

Figure 2.38
The progression of a horizontal bead

space increases, the weld penetration also increases. The root opening for most butt welds will vary from 0 in. (0 mm) to 1/8 in. (3 mm). Excessively large openings can cause burn-through or a cold lap at the weld root, **Figure 2.40.**

After a butt weld is completed, the plate can be cut apart so it can be used for rewelding. The strips for butt welding should be no smaller than 1 in. (25 mm) wide. If they are too narrow, there will be a problem with heat buildup.

If the plate strips are no longer flat after the weld has been cut out, they can be tack welded together and flattened with a hammer, **Figure 2.41.**

Figure 2.39
The tack weld should be small and uniform to minimize its effect on the final weld
Courtesy of Larry Jeffus

PRACTICE 2-5

Welded Square Butt Joint in the Flat Position (1G) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using a properly set up and adjusted arc welding machine, proper safety protection, arc welding electrodes having a 1/8-in. (3-mm) diameter, and two or more pieces of mild steel plate, 6 in. (152 mm) long × 1/4 in.

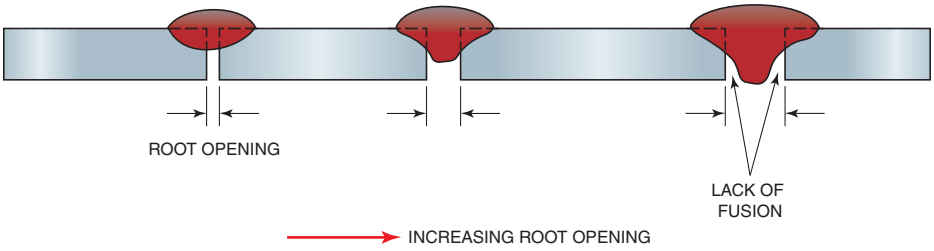

Figure 2.40
Effect of root opening on weld penetration

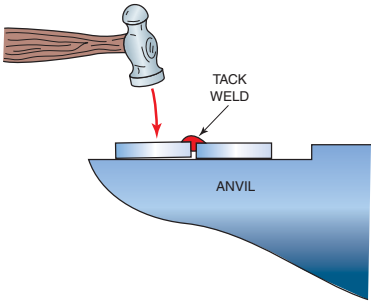

Figure 2.41
After the plates are tack welded together, they can be forced into alignment by striking them with a hammer

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 6
This practice addresses the "Flat" position portion of the all-position requirement of 6.

Module 9
Key Indicator 2

(6 mm) thick, you will make a welded square butt joint in the flat position, **Figure 2.42**.

Tack weld the plates together and place them flat on the welding table. Starting at one end, establish a molten weld pool on both plates. Hold the electrode in the molten weld pool until it flows together, **Figure 2.43.** After the gap is bridged by the molten weld pool, start weaving the electrode slowly back and forth across the joint. Moving the electrode too quickly from side to side may result in slag being trapped in the joint, **Figure 2.44.**

Continue the weld along the 6-in. (152-mm) length of the joint. Normally, deep penetration is not required for this type of weld. If full plate penetration is required, the edges of the butt joint should be beveled or a larger than normal root gap should be used. Cool, chip, and inspect the weld for uniformity and soundness. Repeat the welds as needed to master all three (F) groups of electrodes in this position. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

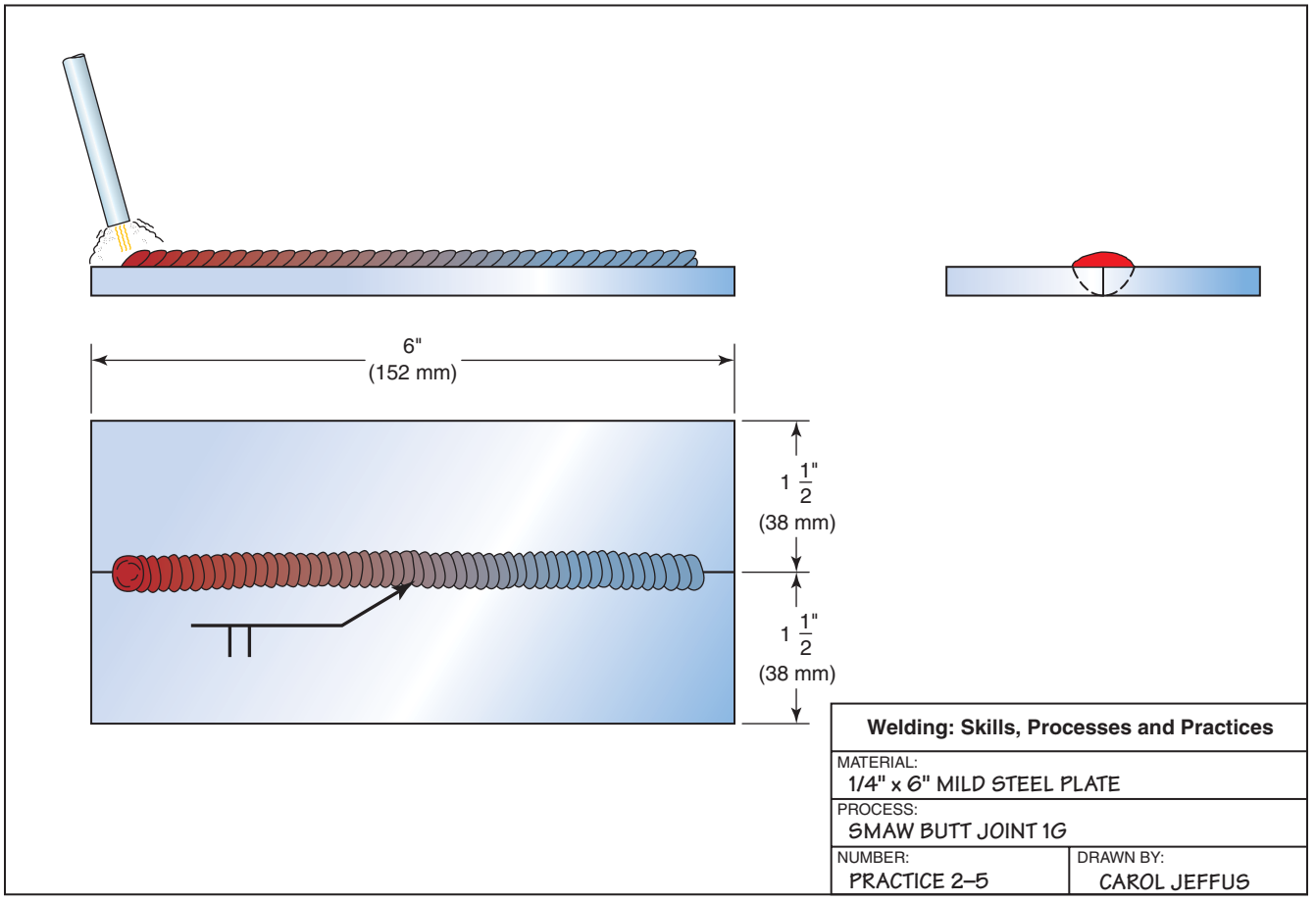

Figure 2.42
Square butt joint in the flat position

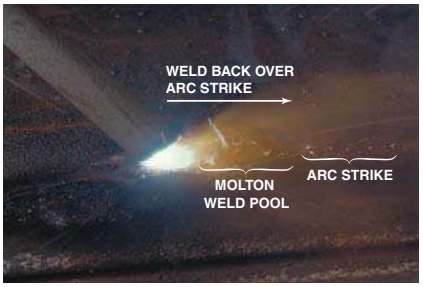

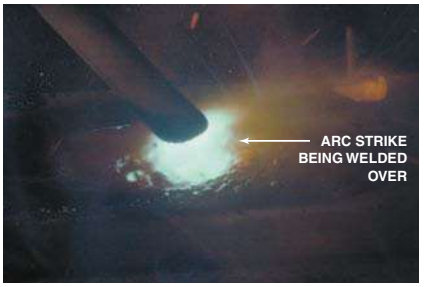

Figure 2.43
(A) After the arc is established, hold it in one area long enough to establish the size of molten weld pool desired. (B) Weld back over the arc strike to melt into the weld.
Courtesy of Larry Jeffus

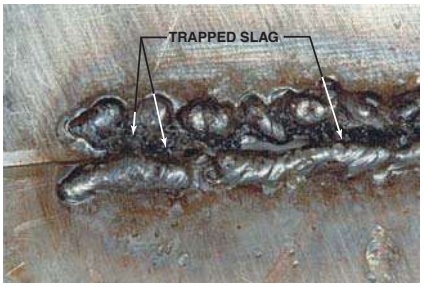

Figure 2.44
Moving the electrode from side to side too quickly can result in slag being trapped between the plates
Courtesy of Larry Jeffus

PRACTICE 2-6

Welded Vertical Up (3G) Square Butt Joint Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using the same setup, materials, and electrodes as listed in Practice 2-5, you will make welded vertical up square butt joints.

With the plates at a 45° angle, start at the bottom and make the molten weld pool bridge the gap between the plates, **Figure 2.45.** Build the bead size slowly so that the molten weld pool has a shelf for support. The "C," "J," or square weave pattern works well for this joint.

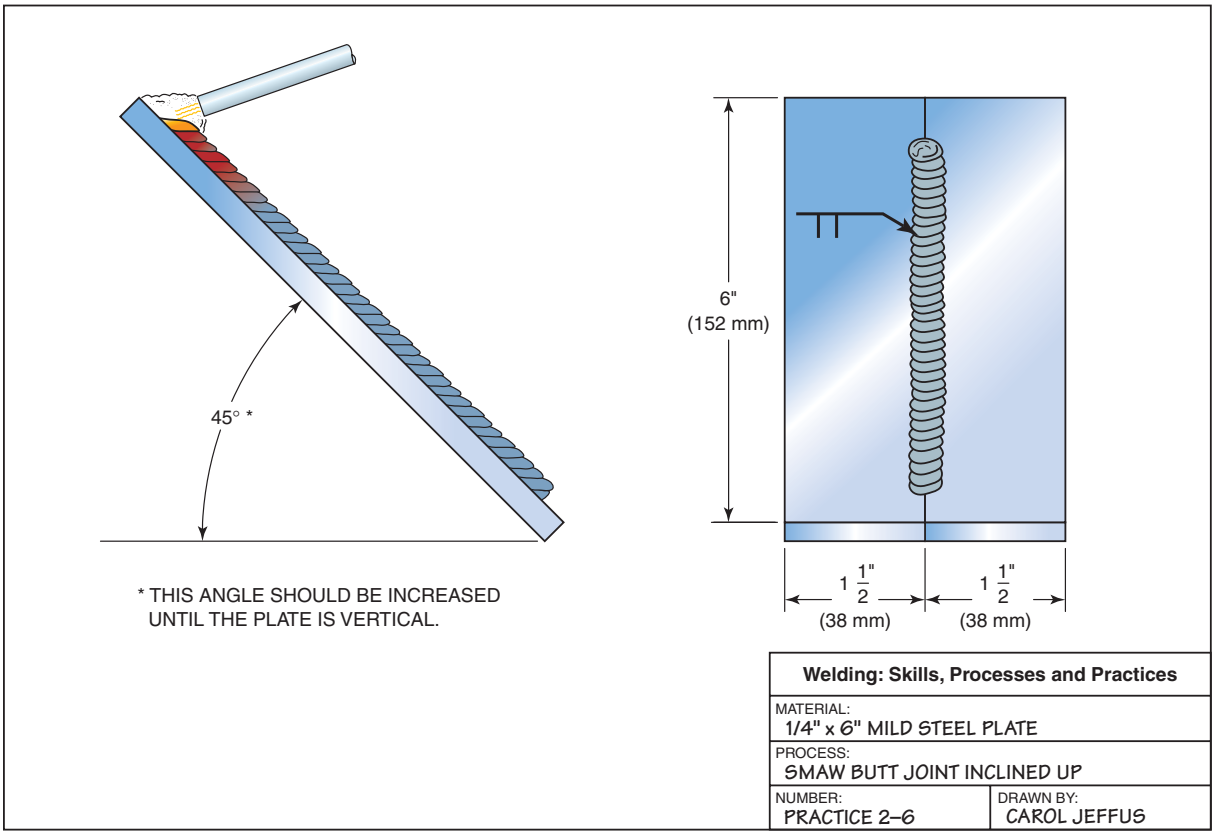

Figure 2.45
Square butt joint in the vertical up position

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 6
This practice addresses the "Vertical" position portion of the all-position requirement of 6.

Module 9
Key Indicator 2

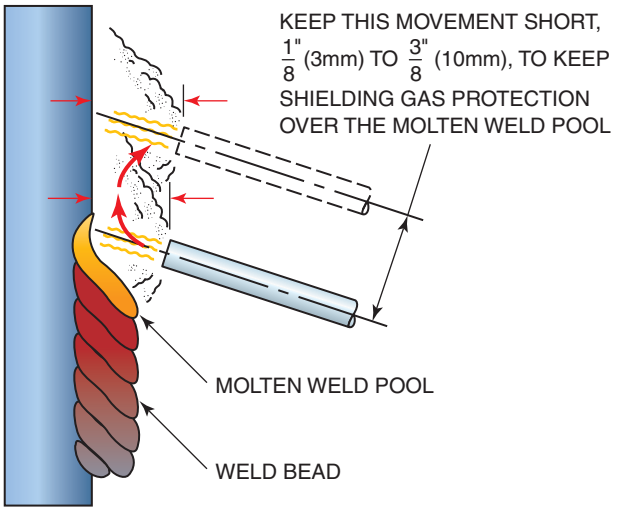

Figure 2.46
Electrode movement for vertical up welds when using F3 electrodes

As the electrode is moved up the weld, the arc is lengthened slightly so that little or no metal is deposited ahead of the molten weld pool. When the electrode is brought back into the molten weld pool, it should be lowered to deposit metal, **Figure 2.46.**

As skill is developed, increase the plate angle until it is vertical. Cool, chip, and inspect the weld for uniformity and defects. Repeat the welds with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-7

Welded Horizontal (2G) Square Butt Joint Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 6
This practice addresses the "Horizontal" position portion of the all-position requirement of 6.

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as described in Practice 2-5, you will make a welded horizontal square butt joint.

- Start practicing these welds with the plate at a slight angle.
- Strike the arc on the bottom plate and build the molten weld pool until it bridges the gap.

If the weld is started on the top plate, slag will be trapped in the root at the beginning of the weld because of poor initial penetration. The slag may cause the weld to crack when it is placed in service.

The "J" weave pattern is recommended in order to deposit metal on the lower plate so that it can support the bead. By pushing the electrode inward as you cross the gap between the plates, deeper penetration is achieved.

As you acquire more skill, gradually increase the plate angle until it is vertical and the weld is horizontal.

- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding

machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

EDGE WELD

An edge weld joint is made by placing the edges of the plate evenly, **Figure 2.47.** When assembling the edge joint the plates should be clamped tightly together; there should not be any gap between the plates. Both edges of the plate assembly can be welded. Make the tack welds to hold the plates together along the ends of the joint, **Figure 2.48.**

The size of the weld should equal the thickness of the plate being joined. A good indication the weld is being made large enough is when the weld bead width is equal to the width of the joint, **Figure 2.49.** The weld bead should also have a slight buildup.

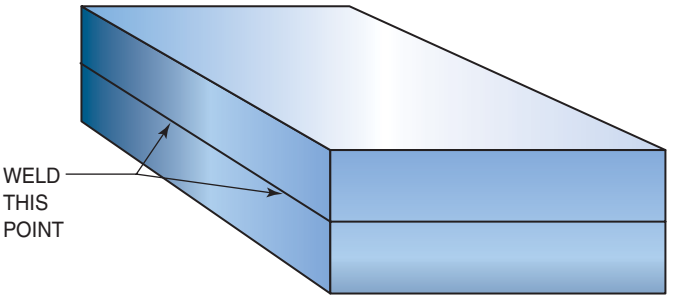

Figure 2.47
Edge joint

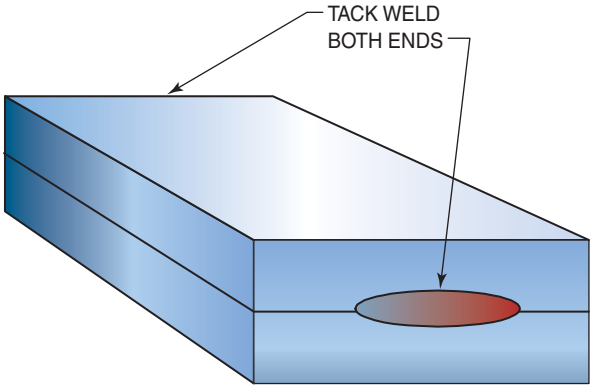

Figure 2.48
Make tack welds at the ends of the joint

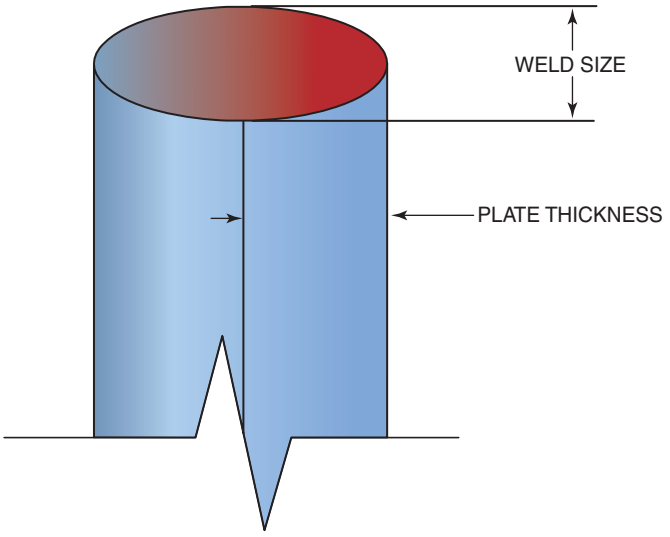

Figure 2.49
Edge weld size

PRACTICE 2-8

Edge Weld in the Flat Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

Using a properly set up and adjusted arc welding machine, proper safety protection as demonstrated in Practice 2-1, arc welding electrodes with a 1/8-in. (3-mm) diameter, and two pieces of mild steel plate, 6 in. (152 mm) long × 1/4 in. (6 mm) thick, you will make a weld on an edge joint, **Figure 2.50.**

- Clamp the plates flat together and make a tack weld along each end of the plates.
- Starting at one end of the plate, make a straight weld the full length of the plate. Make the weld bead as wide as the width of the edge joint.
- Watch the molten weld pool, not the end of the electrode.
- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects.
- Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

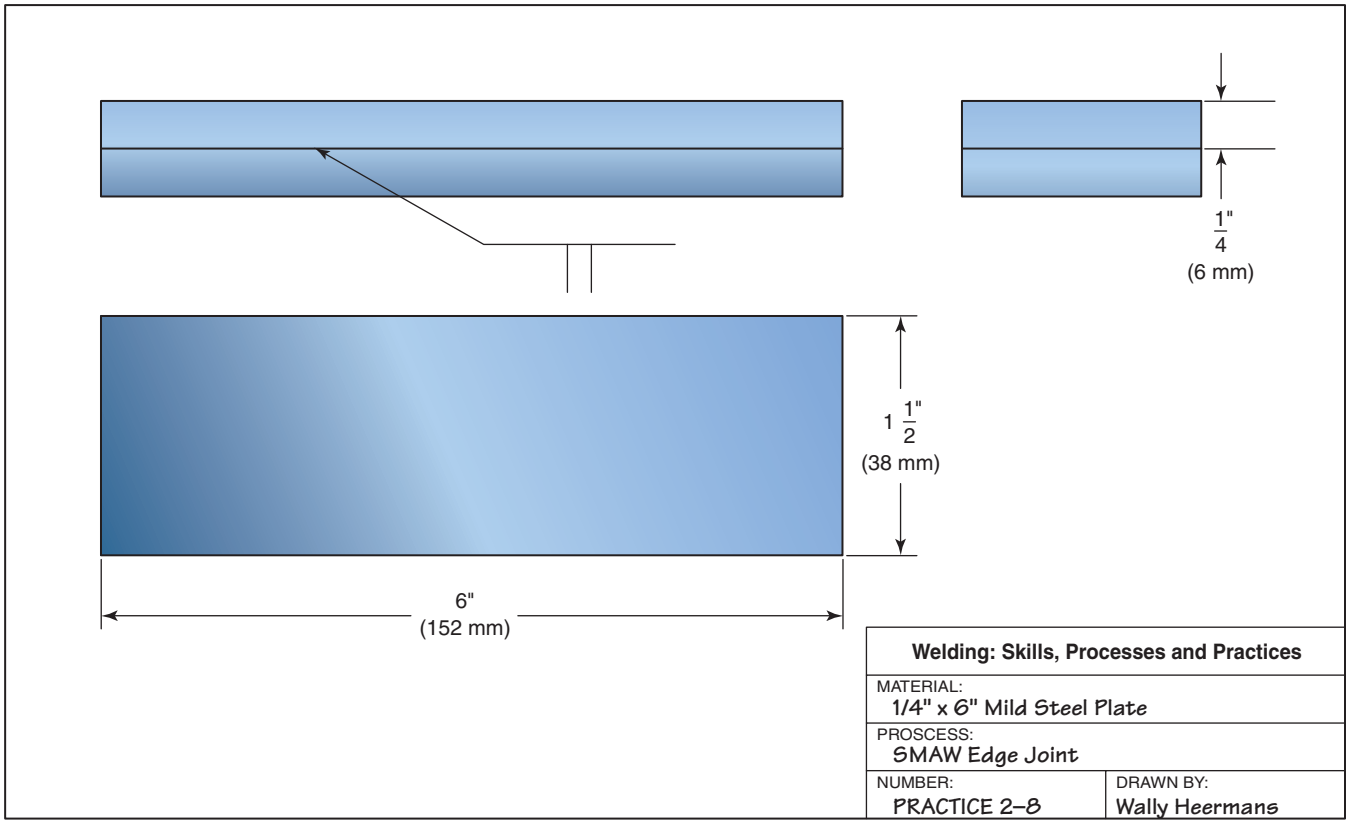

Figure 2.50
Practice 2.8 edge joint

PRACTICE 2-9

Edge Joint in the Vertical Down Position Using E6010 or E6011 Electrodes and E6012 or E6013 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-8, you will make a vertical down weld on an edge joint. Start with the plates at a 45° angle.

This technique is the same as that used to make vertical down welds. However, a lower level of skill is required at 45°, and it is easier to develop your skill. After you master the 45° angle, the angle is increased successively until a vertical position is reached, **Figure 2.51.**

- Make the weld bead as wide as the joint. Controlling a weld bead this size is more difficult, but you must develop the skill required to control this larger molten weld pool.
- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

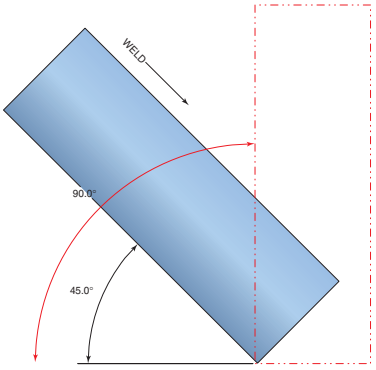

Figure 2.51
Vertical down

PRACTICE 2-10

Edge Joint in the Vertical Up Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-8, you will make a vertical up weld on an edge joint. Start with the plates at a 45° angle.

This technique is the same as that used to make vertical up welds. However, a lower level of skill is required at 45°, and it is easier to develop your skill. After you master the 45° angle, the angle is increased successively until a vertical position is reached, **Figure 2.52.**

Before the molten metal drips down the bead, the back of the molten weld pool will start to bulge, **Figure 2.53.** When this happens, increase the speed of travel and the weave pattern.

- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-11

Edge Joint in the Horizontal Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-8, you will make a horizontal weld on an edge joint. When you begin

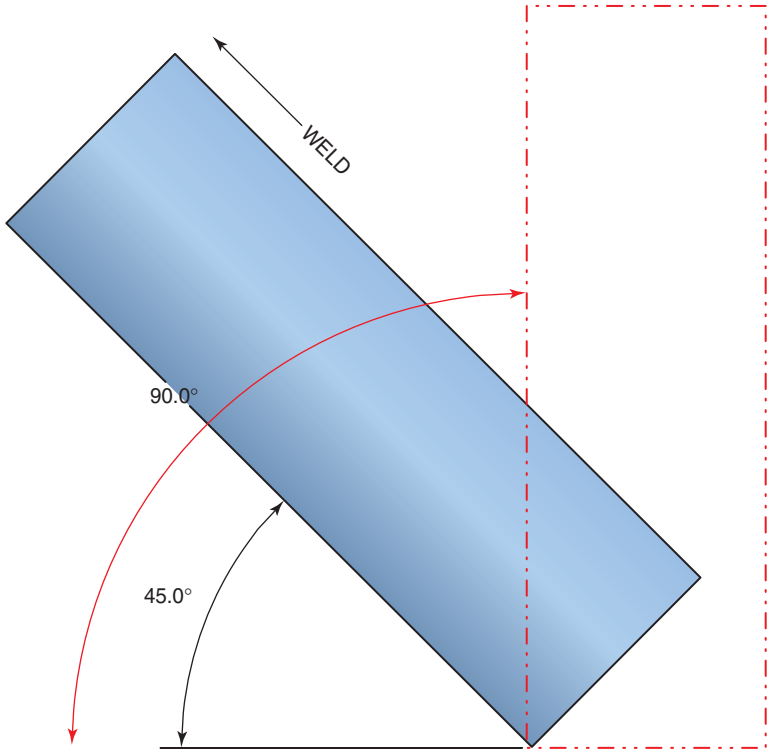

Figure 2.52
Vertical up

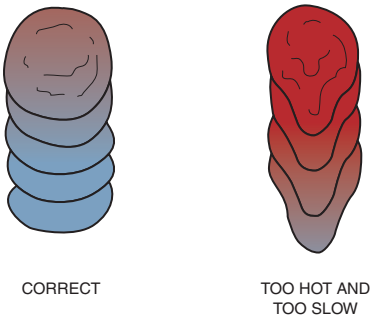

Figure 2.53
Watch the trailing edge of the weld pool to judge the correct travel speed

to practice the horizontal weld, the plate may be reclined slightly, **Figure 2.54.** This placement allows the welder to build the required skill by practicing the correct techniques successfully. The "J" weave or stepped pattern is suggested for this practice. As the electrode is drawn back to the back edge of the weld pool, metal is deposited. Use the metal that's being deposited to support the molten weld pool.

Angling the electrode up and back toward the weld causes more metal to be deposited along the top edge of the weld. Keeping the bead small allows the surface tension to hold the molten weld pool in place.

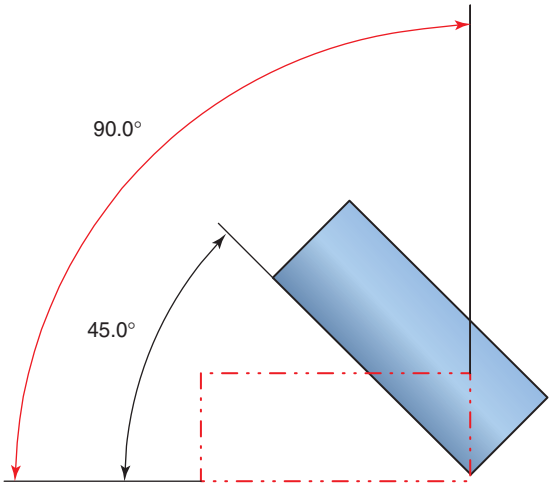

Figure 2.54
Incline angle

Gradually increase the angle of the plate until it and the weld bead are horizontal.

- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-12

Edge Joint in the Overhead Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 1, 2

Using the same setup, materials, and electrodes as listed in Practice 2-8, you will make an overhead weld on an edge joint.

- With the electrode pointed in a slightly trailing angle, **Figure 2.55,** strike the arc in the joint.
- Keep a very short arc length.
- Use the stepped pattern and move the electrode forward slightly when the molten weld pool grows to the correct size, **Figure 2.56.**

As the molten weld pool gets larger, it has a tendency to quickly become convex. If you keep the arc in the molten weld pool once the joint is filled and the weld face is flat, it will quickly overfill and become convex. This can result in the weld face forming drips of metal that hang from the weld like icicles, **Figure 2.57.**

- When the molten weld pool cools and begins to shrink, move the arc back near the center of the weld.

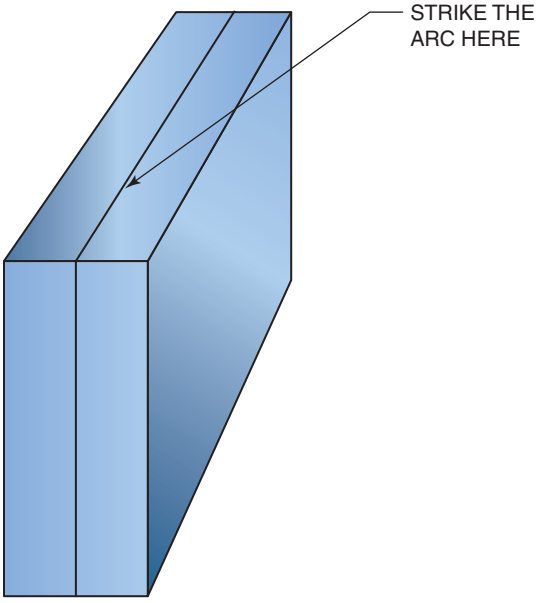

Figure 2.55
Strike the arc in the joint

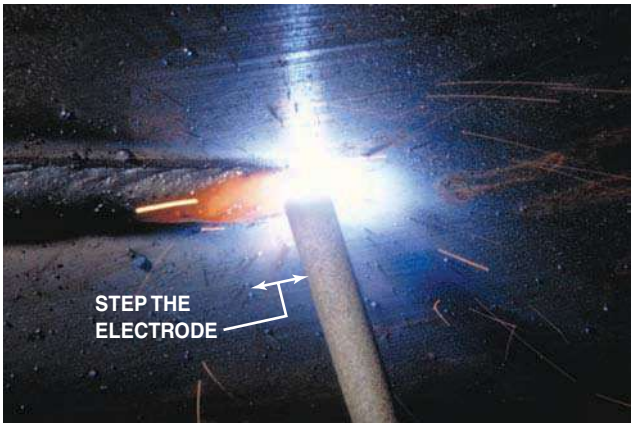

Figure 2.56
Step the electrode
Courtesy of Larry Jeffus

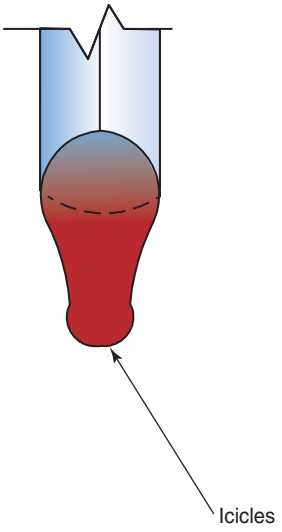

Figure 2.57
Welding too slow or with too high of an amperage setting will result in the weld metal dripping down like icicles

- Hold the arc in this new location until the molten weld pool again grows to the correct size.
- Step the electrode forward again and keep repeating this pattern until the weld progresses along the entire weld joint length.
- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

OUTSIDE CORNER JOINT

An outside corner joint is made by placing the plates at a 90° angle to each other, with the edges forming a V-groove, **Figure 2.58.** There may or may not be a slight root opening left between the plate edges. Small tack welds should be made approximately 1/2 in. (13 mm) from both ends of the joint.

The weld bead should completely fill the V-groove formed by the plates and may have a slightly convex surface buildup. The back side of an outside corner joint can be used to practice fillet welds, or four plates can be made into a box tube shape, **Figure 2.59.**

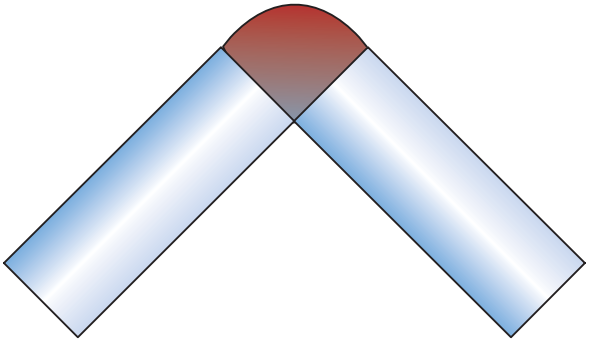

Figure 2.58
V formed by an outside corner joint

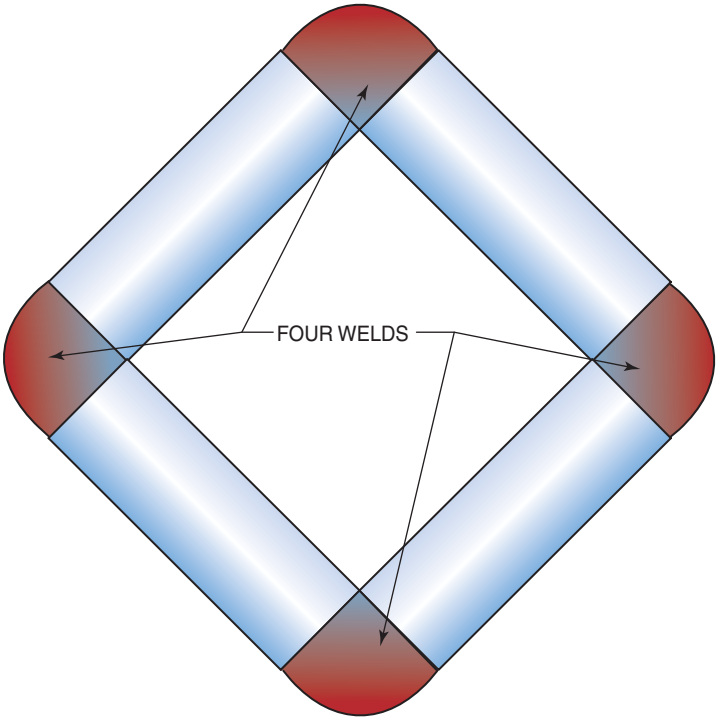

Figure 2.59
Box tube made from four outside corner joint welds

PRACTICE 2-13

Outside Corner Joint in the Flat Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 1, 2

Using a properly set up and adjusted arc welding machine, proper safety protection as demonstrated in Practice 2-1, arc welding electrodes with a 1/8-in. (3-mm) diameter, and two pieces of mild steel plate, 6 in. (152 mm) long × 1/4 in. (6 mm) thick, you will make a weld on an outside corner joint.

- Starting at one end of the plate, make a straight weld the full length of the plate.
- Watch the molten weld pool at this point, not the end of the electrode. As you become more skillful, it is easier to watch the molten weld pool.
- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-14

Outside Corner Joint in the Vertical Down Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes.

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 1, 2

Using the same setup, materials, and electrodes as listed in Practice 2-13, you will make a vertical down weld on an outside corner joint. Start with the plate at a 45° angle.

This technique is the same as that used to make vertical down welds. However, a lower level of skill is required at 45°, and it is easier to develop your skill. After you master the 45° angle, the angle is increased successively until a vertical position is reached, **Figure 2.60.**

- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

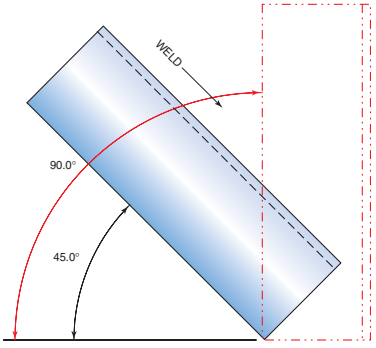

Figure 2.60
Start a 45° angle with and increase it to 90°

PRACTICE 2-15

Outside Corner Joint in the Vertical Up Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using the same setup, materials, and electrodes as listed in Practice 2-13, you will make a vertical up weld on an outside corner joint. Start with the plate at a 45° angle.

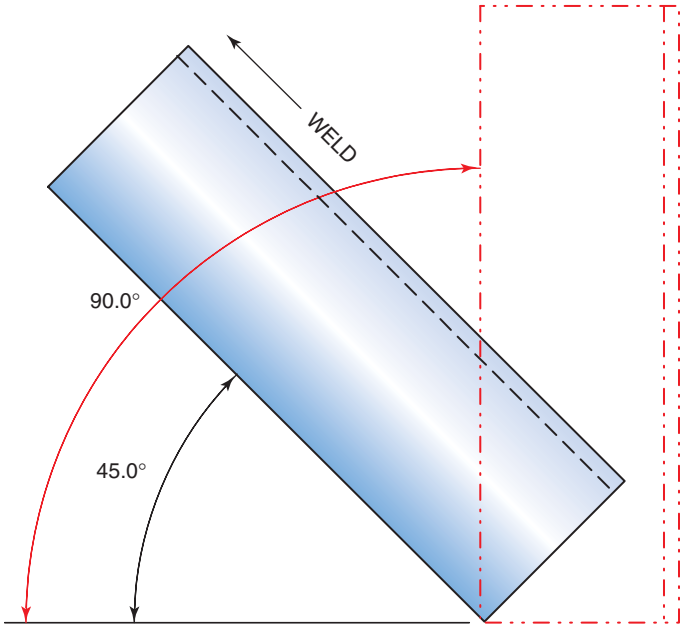

Figure 2.61
Vertical up

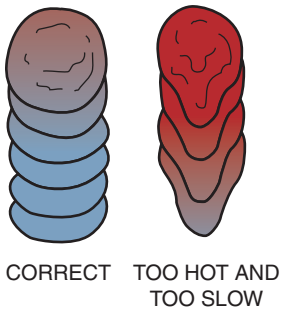

Figure 2.62
Watch the trailing edge of the weld pool to judge the correct travel speed

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

This technique is the same as that used to make vertical up welds. However, a lower level of skill is required at 45°, and it is easier to develop your skill. After the welder masters the 45° angle, the angle is increased successively until a vertical position is reached, **Figure 2.61.**

Before the molten metal drips down the bead, the back of the molten weld pool will start to bulge, **Figure 2.62.** When this happens, increase the speed of travel and the weave pattern.

- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-16

Outside Corner Joint in the Horizontal Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-13, you will make a horizontal weld on an outside corner joint. When the welder begins to practice the horizontal weld, the joint may be reclined slightly, **Figure 2.63.** This placement allows the welder to build the required skill by practicing the correct techniques successfully. The "J" weave or stepped pattern is suggested for this practice. As the electrode is drawn back into the weld pool, metal is deposited. This metal supports the molten weld pool, resulting in a bead with a uniform contour, **Figure 2.64.**

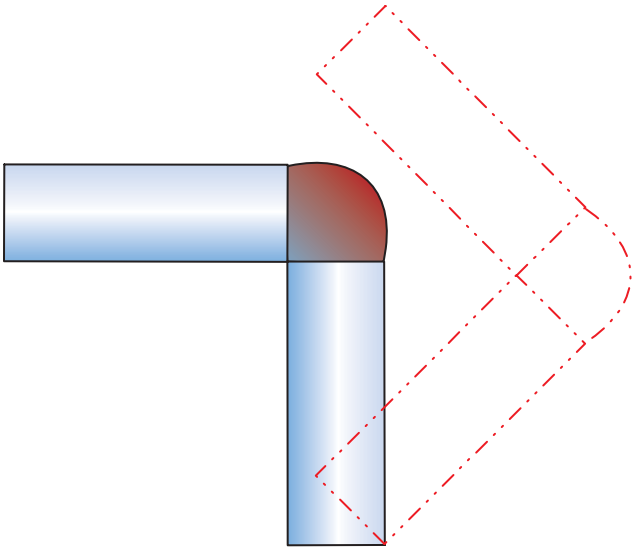

Figure 2.63
Incline angle

Figure 2.64
"J" weave or stepping

Angling the electrode up and back toward the weld causes more metal to be deposited along the top edge of the weld. Keeping the bead small allows the surface tension to hold the molten weld pool in place.

Gradually increase the angle of the plate until it is vertical and the weld bead is horizontal.

- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-17

Outside Corner Joint in the Overhead Position Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 3, 4

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-13, you will make an overhead weld on an outside corner joint.

- With the electrode pointed slightly into the joint, **Figure 2.65,** strike the arc in the joint.
- Keep a very short arc length.
- Use the stepped pattern and move the electrode forward slightly when the molten weld pool grows to the correct size, **Figure 2.66.**

As the molten weld pool gets larger, it has a tendency to quickly become convex. If you keep the arc in the molten weld pool once the joint is filled and the weld face is flat, it will quickly overfill and become convex. This can result in the weld face forming drips of metal that hang from the weld like icicles, **Figure 2.67.**

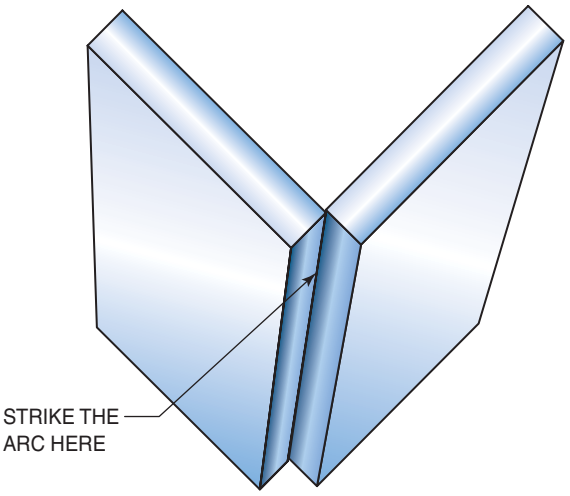

Figure 2.65
Strike arc in the joint

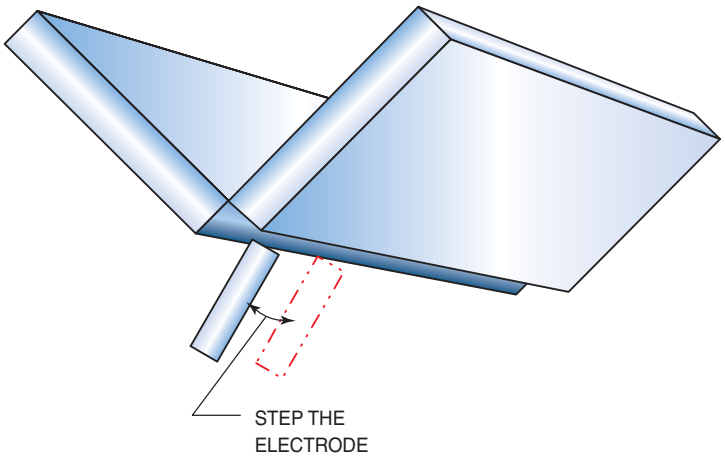

Figure 2.66
Stepping the electrode to control weld size

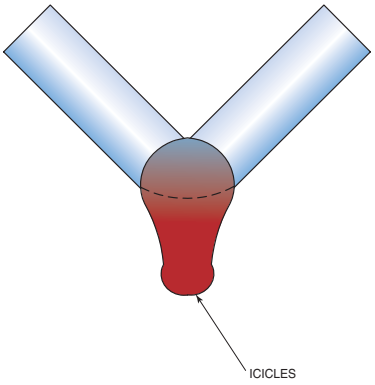

Figure 2.67
Welding too slowly or with too high of an amperage setting will result in the weld metal dripping down like icicles

- When the molten weld pool cools and begins to shrink, move the arc back near the center of the weld.
- Hold the arc in this new location until the molten weld pool again grows to the correct size.
- Step the electrode forward again and keep repeating this pattern until the weld progresses along the entire weld joint length.
- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

LAP JOINT

A **lap joint** is made by overlapping the edges of the two plates, **Figure 2.68.** The joint can be welded on one side or both sides with a fillet weld. In Practice 2-7, both sides should be welded unless otherwise noted.

As the fillet weld is made on the lap joint, the buildup should equal the thickness of the plate, **Figure 2.69.** A good weld will have a smooth transition from the plate surface to the weld. If this transition is abrupt, it can cause stresses that will weaken the joint.

Penetration for lap joints does not improve their strength; complete fusion is required. The root of fillet welds must be melted to ensure a completely fused joint. If the molten weld pool shows a notch during the weld, **Figure 2.70,** this is an indication that the root is not being fused together. The weave pattern will help prevent this problem, **Figure 2.71.**

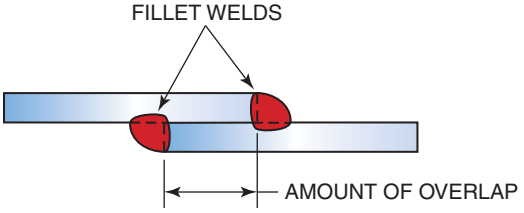

Figure 2.68
Lap joint

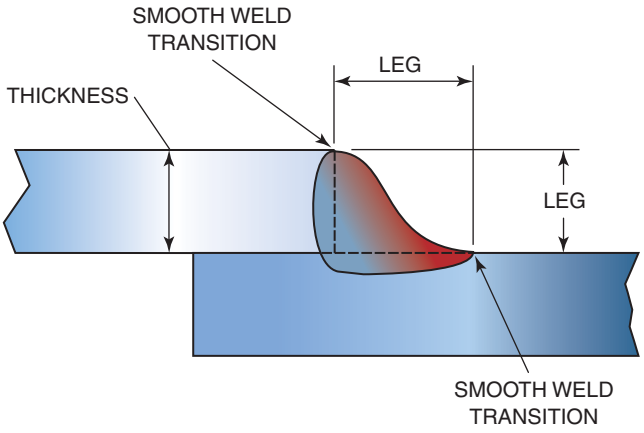

Figure 2.69
The legs of a fillet weld generally should be equal to the thickness of the base metal

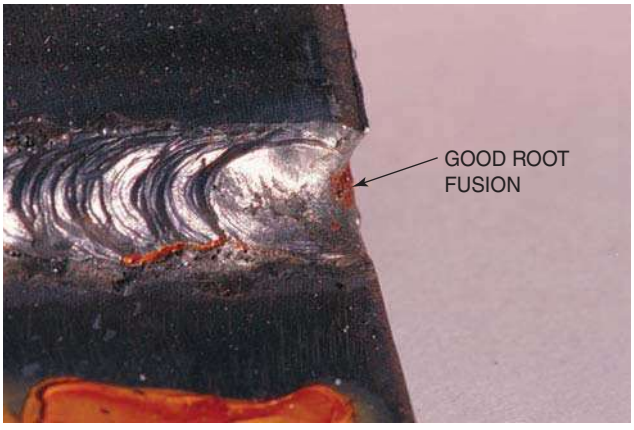

Figure 2.70
Watch the root of the weld bead to be sure there is a complete fusion
Courtesy of Larry Jeffus

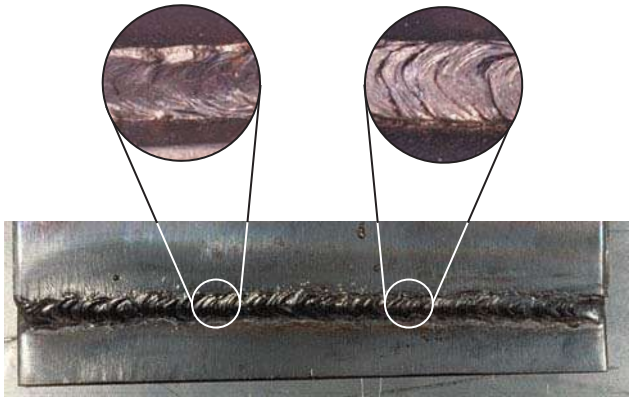

Figure 2.71
Lap joint
Courtesy of Larry Jeffus

PRACTICE 2-18

Welded Lap Joint in the Flat Position (1F) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using a properly set up and adjusted arc welding machine, proper safety protection, arc welding electrodes having a 1/8-in. (3-mm) diameter, and two or more pieces of mild steel plate, 6 in. (152 mm) long × 1/4 in. (6 mm) thick, you will make a welded lap joint in the flat position, **Figure 2.72.**

Hold the plates together tightly with an overlap of no more than 1/4 in. (6 mm). Tack weld the plates together. A small tack weld may be added in the center to prevent distortion during welding, **Figure 2.73.** Chip the tacks before you start to weld.

The "J," "C," or zigzag weave pattern works well on this joint. Strike the arc and establish a molten pool directly in the joint. Move the electrode out on the bottom plate and then onto the weld to the top edge of the top plate, **Figure 2.74.** Follow the surface of the plates with the arc. Do not follow the trailing edge of the weld bead. Following the molten weld pool will not allow for good root fusion and will cause slag to collect in the root. If

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 5

Module 9
Key Indicator 1, 2
This practice addresses the "Flat" position portion of the all-position requirement of 5.

Module 9
Key Indicator 2

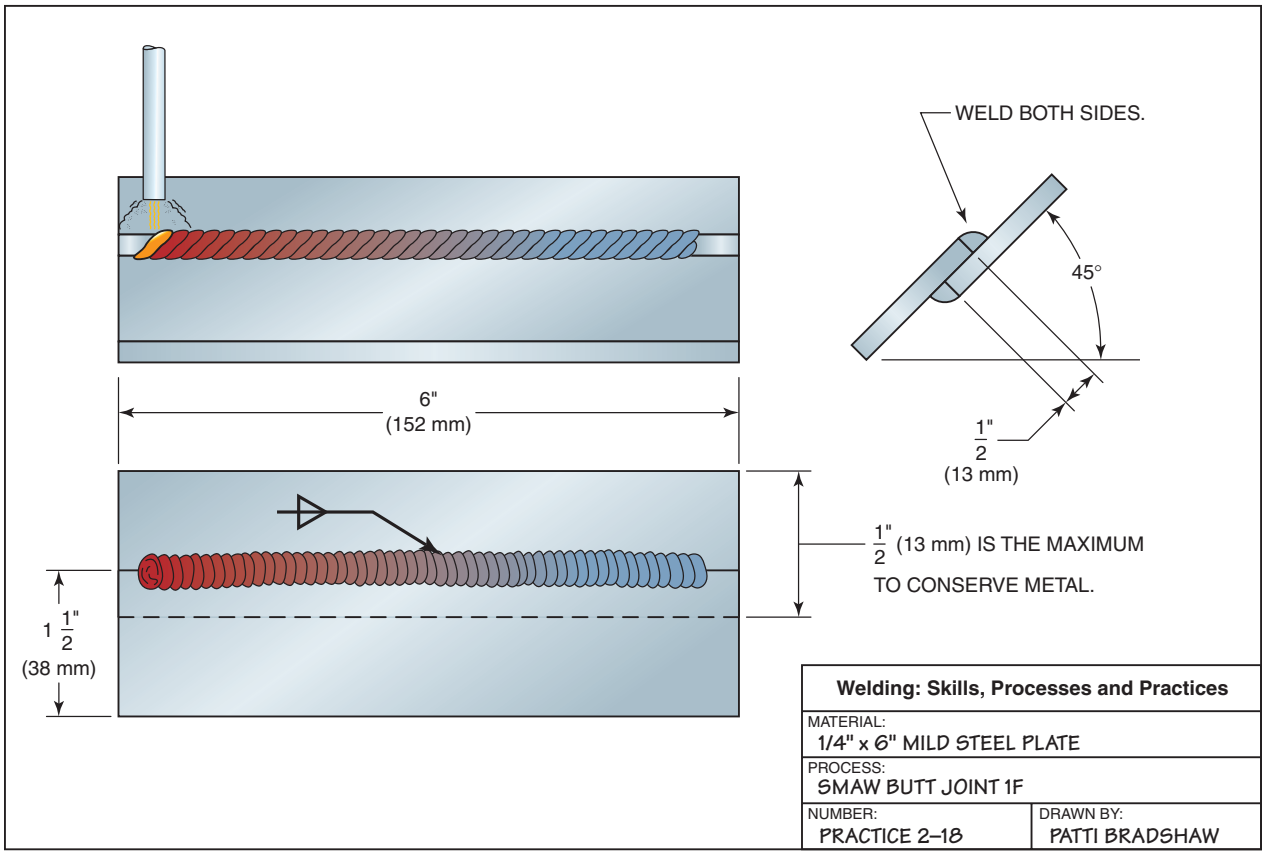

Figure 2.72
Lap joint in the flat position

slag does collect, a good weld is not possible. Stop the weld and chip the slag to remove it before the weld is completed. Cool, chip, and inspect the weld for uniformity and defects. Repeat the welds with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

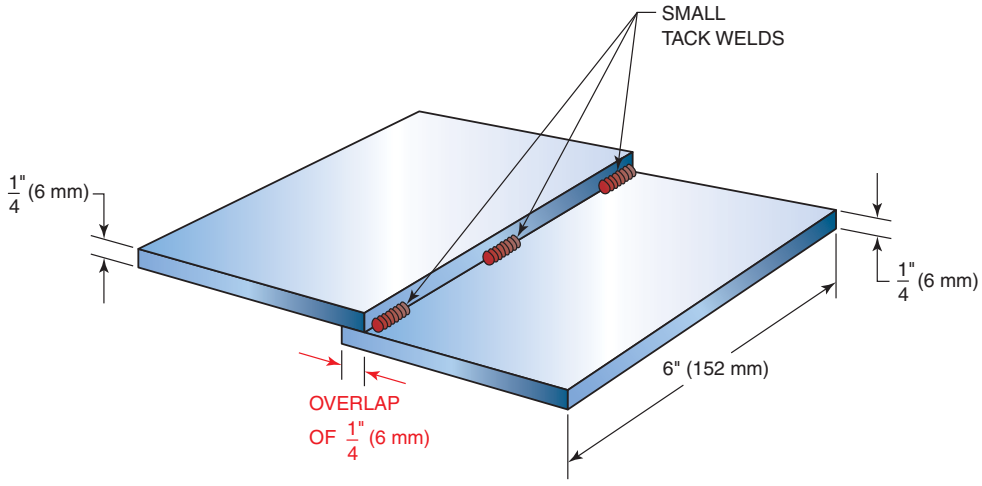

Figure 2.73
Tack welding the plates together

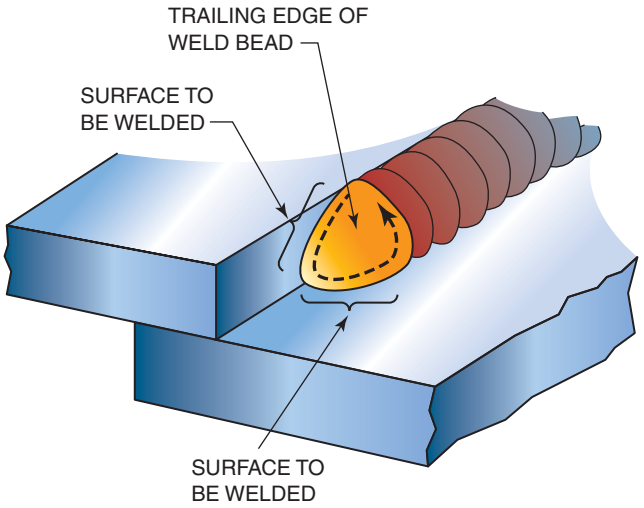

Figure 2.74
Follow the surface of the plate to ensure good fusion

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 5

Module 9
Key Indicator 1, 2
This practice addresses the "Horizontal" position portion of the all-position requirement of 5.

Module 9
Key Indicator 2

PRACTICE 2-19

Welded Lap Joint in the Horizontal Position (2F) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using the same setup, materials, and electrodes as listed in Practice 2-18, you will make a welded horizontal lap joint.

The horizontal lap joint and the flat lap joint require nearly the same technique and skill to achieve a proper weld, **Figure 2.75.** Use the "J," "C," or zigzag weave pattern to make the weld. Do not allow slag to collect in the root. The fillet must be equally divided between both plates for good strength. After completing the weld, cool, chip, and inspect the weld for uniformity and defects. Repeat the welds using all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

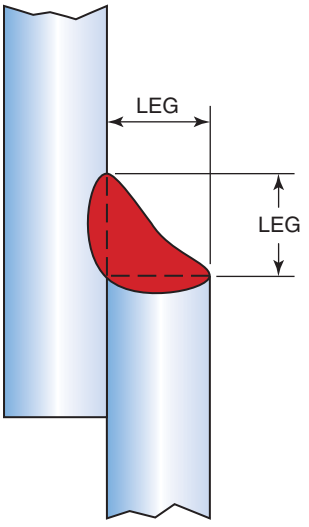

Figure 2.75
The horizontal lap joint should have a fillet weld that is equal on both plates

PROACTICE 2-20

Lap Joint in the Vertical Up Position (3F) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using the same setup, materials, and electrodes as listed in Practice 2-18, you will make a vertical up weld in a lap joint.

- Start practicing this weld with the plate at a 45° angle.
- Gradually increase the angle of the plate to vertical as skill is gained in welding this joint. The "J" or "T" weave pattern works well on this joint.
- Establish a molten weld pool in the root of the joint.

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 5
This practice addresses the "Vertical" position portion of the all-position requirement of 5.

Module 9
Key Indicator 2

- Use the "T" pattern to step ahead of the molten weld pool, allowing it to cool slightly. Do not deposit metal ahead of the molten weld pool.
- As the molten weld pool size starts to decrease, move the electrode back down into the molten weld pool.
- Quickly move the electrode from side to side in the molten weld pool, filling up the joint.
- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as necessary with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-21

Lap Joint in the Overhead Position (4F) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 5
This practice addresses the "Overhead" position portion of the all-position requirement of 5.

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-18, you will make an overhead weld in a lap joint.

- With the electrode pointed slightly into the joint, **Figure 2.76,** strike the arc in the inside corner of the lap joint.
- Keep a very short arc length.
- Use the stepped pattern and move the electrode forward slightly when the molten weld pool grows to the correct size, **Figure 2.77.**

As the molten weld pool gets larger, it has a tendency to quickly become convex. If you keep the arc in the molten weld pool once the joint is filled and the weld face is flat, it will quickly overfill and become convex. This can result in the weld face forming drips of metal that hang from the weld like icicles, **Figure 2.78.**

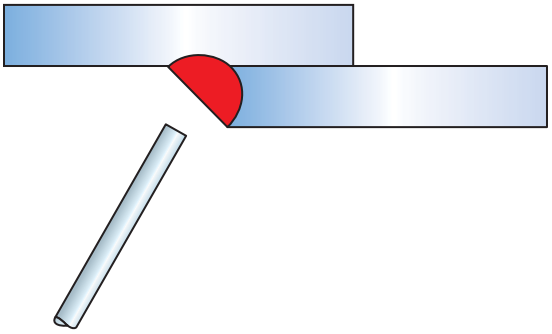

Figure 2.76
Point the electrode slightly toward the root of the joint

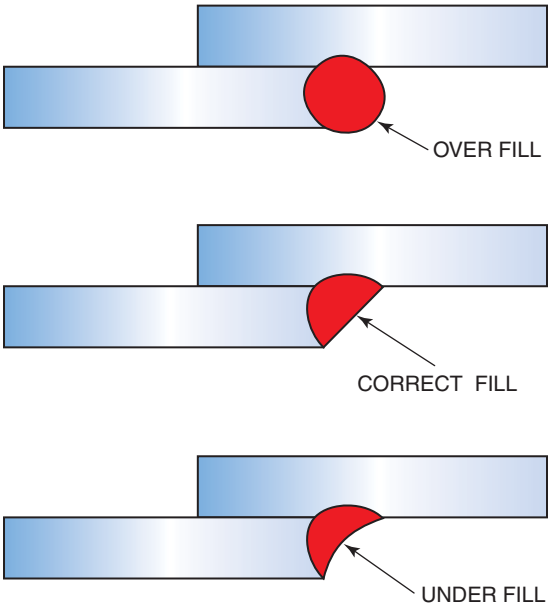

Figure 2.77
Correct fillet weld size for overhead welds

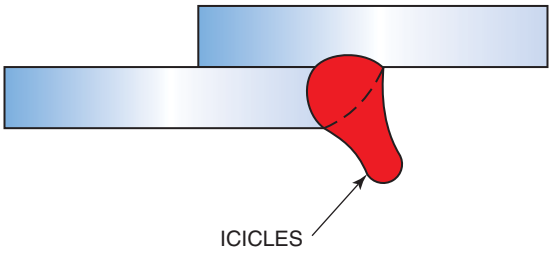

Figure 2.78
Overfilling the molten weld pool will result in drips of metal called icicles

- When the molten weld pool cools and begins to shrink, move the arc back near the center of the weld.
- Hold the arc in this new location until the molten weld pool again grows to the correct size.
- Step the electrode forward again and keep repeating this pattern until the weld progresses along the entire weld joint length.
- Cool, chip, and inspect the weld for uniformity and defects.
- Repeat the welds as needed with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

TEE JOINT

The **tee joint** is made by tack welding one piece of metal on another piece of metal at a right angle, **Figure 2.79.** After the joint is tack welded together, the slag is chipped from the tack welds. If the slag is not removed, it will cause a slag inclusion in the final weld.

The heat is not distributed uniformly between both plates during a tee weld. Because the plate that forms the stem of the tee can conduct heat away from the arc in only one direction, it will heat up faster than the base plate. Heat escapes into the base plate in two directions. When using a weave pattern, most of the heat should be directed to the base plate to keep the weld size more uniform and to help prevent undercut.

A welded tee joint can be strong if it is welded on both sides, even without having deep penetration, **Figure 2.80.** The weld will be as strong as the

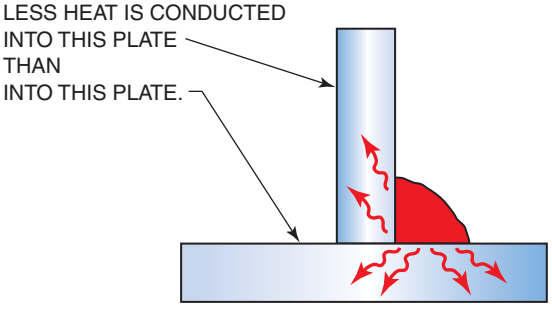

Figure 2.79
Tee joint

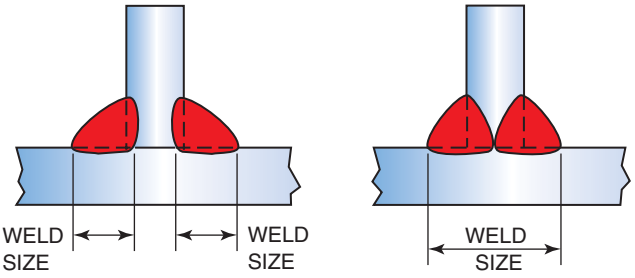

Figure 2.80
If the total weld sizes are equal, then both tee joints would have equal strength

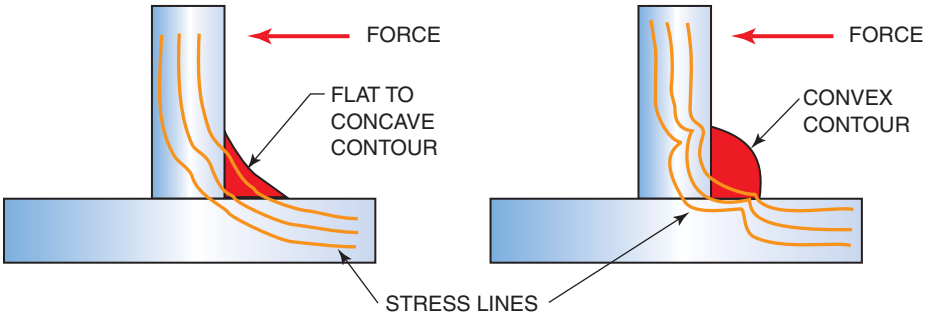

Figure 2.81
The stresses are distributed more uniformly through a flat or concave fillet weld

base plate if the size of the two welds equals the total thickness of the base plate. The weld bead should have a flat or slightly concave appearance to ensure the greatest strength and efficiency, **Figure 2.81.**

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 5
This practice addresses the "Flat" position portion of the all-position requirement of 5.

Module 9
Key Indicator 2

PRACTICE 2-22

Tee Joint in the Flat Position (1F) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Using a properly set up and adjusted arc welding machine, proper safety protection, arc welding electrodes having a 1/8-in. (3-mm) diameter, and two or more pieces of mild steel plate, 6 in. (152 mm) long × 1/4 in. (6 mm) thick, you will make a welded tee joint in the flat position, **Figure 2.82.**

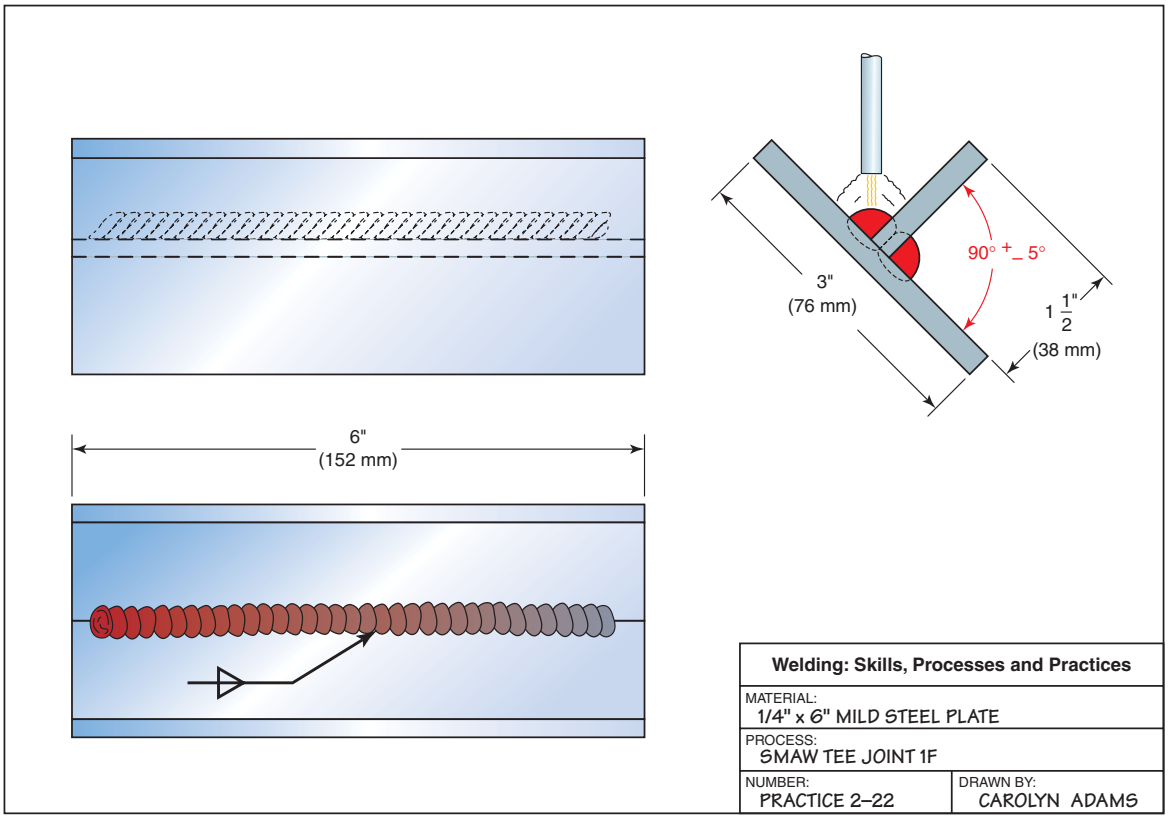

Figure 2.82
Tee joint in the flat position

After the plates are tack welded together, place them on the welding table so the weld will be flat. Start at one end and establish a molten weld pool on both plates. Allow the molten weld pool to flow together before starting the bead. Any of the weave patterns will work well on this joint. To prevent slag inclusions, use a slightly higher than normal amperage setting.

When the 6-in. (152-mm)-long weld is completed, cool, chip, and inspect it for uniformity and soundness. Repeat the welds as needed for all these groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-23

Tee Joint in the Horizontal Position (2F) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 5
This practice addresses the "Horizontal" position portion of the all-position requirement of 5.

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-22, you will make a welded tee joint in the horizontal position.

Place the tack welded tee plates flat on the welding table so that the weld is horizontal and the plates are flat and vertical, **Figure 2.83.** Start the arc on the flat plate and establish a molten weld pool in the root on both plates. Using the "J" or "C" weave pattern, push the arc into the root and slightly up the vertical plate. You must keep the root of the joint fusing together with the weld metal. If the metal does not fuse, a notch will appear on the leading edge of the weld bead. Poor or incomplete root fusion will cause the weld to be weak and easily cracked under a load.

When the weld is completed, cool, chip, and inspect it for uniformity and defects. Undercut on the vertical plate is the most common defect. Repeat the welds with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

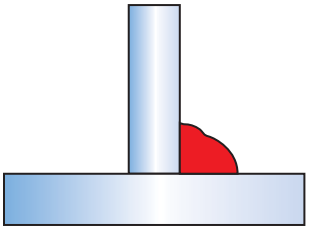

Figure 2.83
Horizontal tee

PRACTICE 2-24

Tee Joint in the Vertical Up Position (3F) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 5
This practice addresses the "Vertical" position portion of the all-position requirement of 5.

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-22, you will make a welded tee joint in the vertical up position.

Practice this weld with the plate at a 45° angle. This position will allow you to develop your skill for the vertical up position. Start the arc and molten weld pool deep in the root of the joint. Build a shelf large enough to support the bead as it progresses up the joint. The square, "J," or "C" pattern can be used, but the "T" or stepped pattern will allow deeper root penetration.

For this weld, undercut is a problem on both sides of the weld. It can be controlled by holding the arc on the side long enough for filler metal

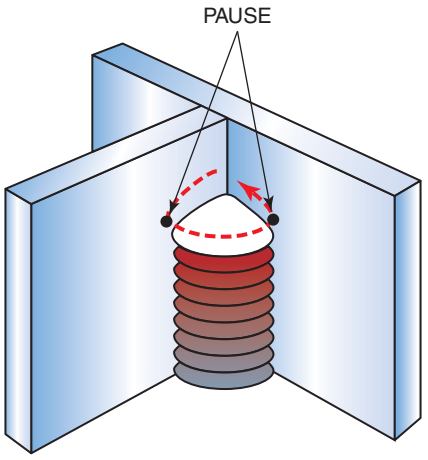

Figure 2.84
Pausing just above the undercut will fill it. This action also causes undercut, but that will be filled on the next cycle.

to flow down and fill it, **Figure 2.84.** Cool, chip, and inspect the weld for uniformity and defects. Repeat the welds as necessary with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-25

Tee Joint in the Overhead Position (4F) Using E6010 or E6011 Electrodes, E6012 or E6013 Electrodes, and E7016 or E7018 Electrodes

Module 1
Key Indicator 1, 2, 3, 4

Module 4
Key Indicator 1, 5
This practice addresses the "Overhead" position portion of the all-position requirement of 5.

Module 9
Key Indicator 2

Using the same setup, materials, and electrodes as listed in Practice 2-12, you will make a welded tee joint in the overhead position.

Start the arc and molten weld pool deep in the root of the joint. Keep a very short arc length. The stepped pattern will allow deeper root penetration.

For this weld, undercut is a problem on both sides of the weld with a high buildup in the center. It can be controlled by holding the arc on the side long enough for filler metal to flow in and fill it. Cool, chip, and inspect the weld for uniformity and defects. Repeat the welds as necessary with all three (F) groups of electrodes until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

SUMMARY

The shielded metal arc welding process is most often referred to in welding shops as stick welding. Some people say that it gets this name for one of two reasons. The first is most obviously a result of the stick shape

of the electrode. The second reason is experienced by all new welders; it is the tendency for the electrode to stick to the workpiece. All new welders experience this, and your ability to control the sticking of the electrode can be improved as you develop the proper arc-striking techniques.

For a new welder, it is often difficult to concentrate on anything other than the bright sparks and glow at the end of the electrode. But, with time, as you develop your skills, your visual field will increase, allowing you to see a much larger welding zone. This skill comes with time and practice. Developing this skill is essential for you to become a highly proficient welder. Nothing enhances your welding skills more than time under the hood, actually welding, cleaning off the weld, inspecting it, determining the necessary corrections to be made, and immediately trying to produce the next weld with a higher level of quality.

REVIEW

1. Describe two methods of striking an arc with an electrode.
2. Why is it important to strike the arc only in the weld joint?
3. What problems may result by using an electrode at too low a current setting?
4. What problems may result by using an electrode at too high a current setting?
5. According to **Table 2.1,** what would the amperage range be for the following electrodes?
   a. 1/8 in. (3 mm), E6010
   b. 5/32 in. (4 mm), E7018
   c. 3/32 in. (2.4 mm), E7016
   d. 1/8 in. (3 mm), E6011
6. What makes some spatter "hard"?
7. Why should you never change the current setting during a weld?
8. What factors should be considered when selecting an electrode size?
9. What can a welder do to control overheating of the metal pieces being welded?
10. What effect does changing the arc length have on the weld?
11. What arc problems can occur in deep or narrow weld joints?
12. Describe the difference between using a leading and a trailing electrode angle.
13. Can all electrodes be used with a leading angle? Why or why not?
14. What characteristics of the weld bead does the weaving of the electrode cause?
15. What are some of the applications for the circular pattern in the flat position?
16. Using a pencil and paper, draw two complete lines of the weave patterns you are most comfortable making.
17. Why is it important to find a good welding position?
18. Which electrodes would be grouped in the following F-numbers: F2, F3, F4?
19. Give one advantage of using electrodes with cellulose-based fluxes.
20. What are stringer beads?
21. Describe an ideal tack weld.

22. What effect does the root opening or root gap have on a butt joint?
23. What can happen if the fillet weld on a lap joint does not have a smooth transition?
24. Which plate heats up faster on a tee joint? Why?
25. Can a tee weld be strong if the welds on both sides do not have deep penetration? Why or why not?

CHAPTER

3 Advanced Shielded Metal Arc Welding

OBJECTIVES

After completing this chapter, the student should be able to

- prepare metal before welding
- make the root pass, filler weld, and cover pass in all positions and techniques
- explain the purpose of a hot pass
- make a visual inspection, and describe the appearance of an acceptable weld
- make a root pass on plate in all positions
- make a root pass on plate with an open root in all positions
- make an open root weld on plate using the step technique in all positions
- make a multiple pass filler weld on a V-joint in all positions using E7018 electrodes
- make groove welds on a limited thickness plate with a backing strip using E6010 root and hot passes and E7018 fill passes in all positions
- make an SMAW workmanship sample and welder qualification test plate for limited thickness horizontal 2G and 3G positions with E7018 electrodes
- make a cover bead in all positions
- make limited thickness welder performance qualification test plate without backing in all positions
- make a single V-groove open root butt joint with an increasing root opening
- make a single V-groove open root butt joint with a decreasing root opening
- make SMAW welds of plate to plate

KEY TERMS

back gouging
burn-through
cover pass
filler pass
guided bend specimen
hot pass
interpass temperature
key hole
molten weld pool
multiple pass weld
postheating
preheating
root pass
wagon tracks
weld groove
weld specimen

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports or records
Key Indicator 2: Performs housekeeping duties
Key Indicator 4: Follows written instructions to complete work assignments

Module 4: Shielded Metal Arc Welding (SMAW)

Key Indicator 1: Performs safety inspections of SMAW equipment and accessories
Key Indicator 2: Makes minor external repairs to SMAW equipment and accessories
Key Indicator 3: Sets up SMAW operations on carbon steel
Key Indicator 4: Operates SMAW equipment on carbon steel
Key Indicator 5: Makes SMAW fillet welds, in all positions, on carbon steel
Key Indicator 6: Makes SMAW groove welds, in all positions, on carbon steel
Key Indicator 7: Passes SMAW welder performance qualification test (2G and 3G, uphill, limited thickness test plates) on carbon steel

Module 9: Welding Inspection and Testing

Key Indicator 1: Examines cut surfaces and edges of prepared base metal parts
Key Indicator 2: Examines tacks, root passes, intermediate layers, and completed welds

INTRODUCTION

The SMAW process can be used to consistently produce high-quality welds. Sometimes it is necessary to make welds in less than ideal conditions. Knowing how to produce a weld of high strength in an out-of-position, difficult situation or on an unusual metal takes both practice and knowledge. A welder is frequently required to make these types of welds to a code or standard. This chapter covers the high-quality welding of plate. The practices are designed to give you the experience of taking code-type tests in a variety of materials and positions as well as to develop good workmanship.

Any time a code-quality weld requiring 100% joint penetration is to be made on metal thicker than 1/4 in. (6 mm), the metal edges must be prepared before welding. A joint is prepared for welding by cutting a groove in the metal along the edge. The preparation is done to allow deeper penetration into the joint of the weld for improved strength. Prepared joints often require more than one weld pass to complete them. By preparing the joint, metal

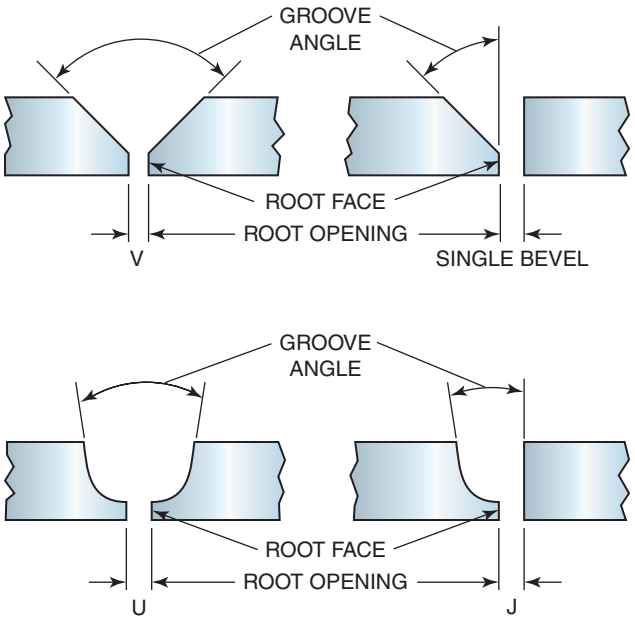

Figure 3.1
Standard grooves used to ensure satisfactory joint penetration

several feet thick can be welded with great success, **Figure 3.1.** The same welding techniques are used when making prepared welds of any thickness.

The root pass is used to fuse the parts together and seal off possible atmospheric contamination from the filler weld. Once the root pass is completed and cleaned, a hot pass may be used to improve the weld contour and burn out small spots of trapped slag. For high-quality welds, a grinder should be used on the root pass to clean it.

Filler welds and cover welds are often made with low hydrogen electrodes such as E7018. These passes are used to fill and cap the weld groove.

Welders are often qualified by passing a required qualification test using a groove weld. The type of joint, thickness of metal, type and size of electrode, and position are all specified by agencies issuing codes and standards. Except for the American Welding Society's Certified Welder program, taking a test according to one company or agency's specifications may not qualify a welder for another company or agency's testing procedures. But being able to pass one type of test will usually help the welder to pass other tests for the same type of joint, thickness of metal, type and size of electrode, and position. Information about the AWS Certified Welder program is available from the AWS's main office in Miami, Florida.

ROOT PASS

The **root pass** is the first weld bead of a **multiple pass weld.** The root pass fuses the two parts together and establishes the depth of weld metal penetration. A good root pass is needed in order to obtain a sound weld. The

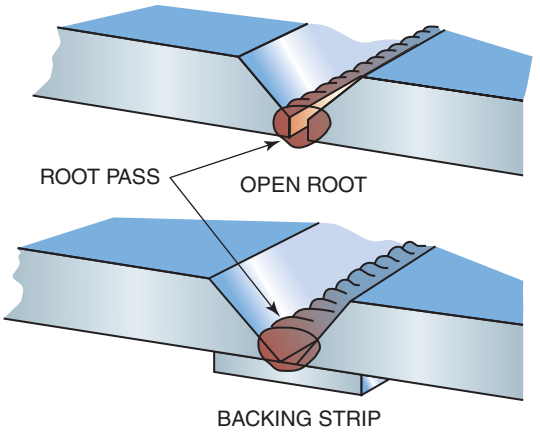

Figure 3.2
Root pass maximum deposit 1/4 in. (6 mm) thick

root may be either open or closed, using a backing strip or backing ring, **Figure 3.2.**

The backing strip used in a closed root may remain as part of the weld, or it may be removed. Because leaving the backing strip on a weld may cause it to fail due to concentrations of stresses along the backing strip, removable backup tapes have been developed. Backup tapes are made of high-temperature ceramics, **Figure 3.3,** that can be used to increase penetration and prevent **burn-through.** The tape can be peeled off after the weld is completed. Most welds do not use backing strips.

On plates that have the joints prepared on both sides, the root face may be ground or gouged clean before another pass is applied to both sides, **Figure 3.4.** This practice has been applied to some large diameter pipes. Welds, however, that can be reached from only one side must be produced adequately, without the benefit of being able to clean and repair the back side.

The open root weld is widely used in plate and pipe designs. The face side of an open root weld is not so important as the root surface on the back or inside, **Figure 3.5.** The face of a root weld may have some areas

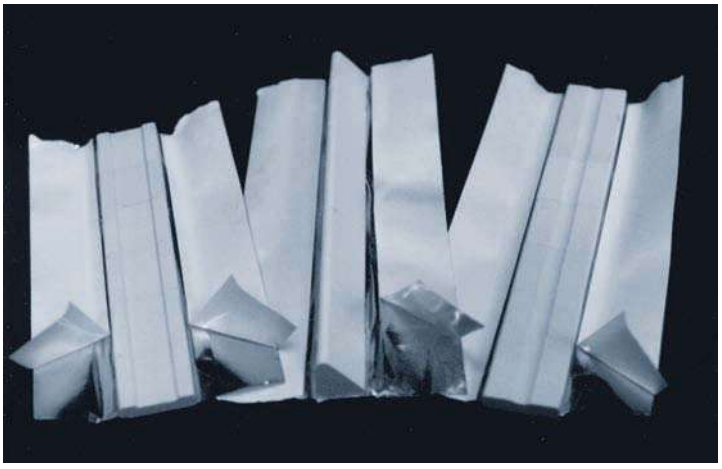

(A) FIBERGLASS

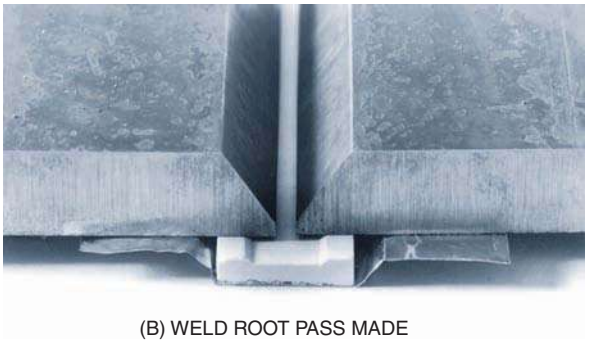

(B) WELD ROOT PASS MADE
USING CERAMIC BACKING TAPE

Figure 3.3
Welding backing tapes are available in different materials and shapes
Courtesy of ESAB Welding & Cutting Products

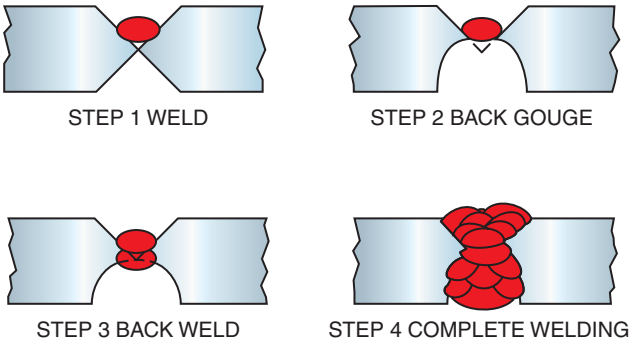

Figure 3.4
Using back gouging to ensure a sound weld root

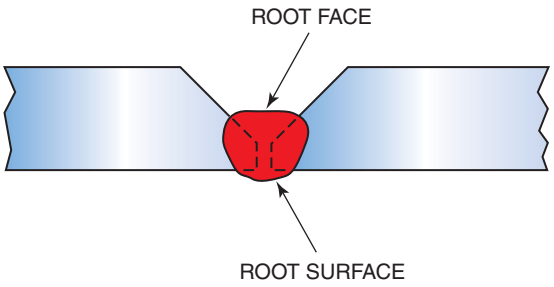

Figure 3.5
Ideal bead shape for the root pass

of poor uniformity in width, reinforcement, and buildup, or it may have other defects, such as undercut or overlap. As long as the root surface is correct, the front side can be ground, gouged, or burned out to produce a sound weld, **Figure 3.6.** For this reason, during the root pass practices, the weld will be evaluated from the root side only, as long as there are not too many defects on the face. To practice the open root welds, the welder will be using mild steel plate that is 1/8 in. (3 mm) thick. The root face for most grooved joints will be about the same size. This thin plate will help the welder build skill without taking too much time beveling the plate just to practice the root pass. Two different methods are used to make a root pass. One method is used only on joints with little or no root gap. This method requires a high amperage and short arc length. The arc length is so short that the electrode flux may drag along on the edges of the joint. The setup for this method must be correct in order for it to work.

The other method can be used on joints with wide, narrow, or varying root gaps. A stepping electrode manipulation and key hole control the penetration. The electrode is moved in and out of the **molten weld pool** as the weld progresses along the joint. The edge of the metal is burned back slightly by the electrode just ahead of the molten weld pool, **Figure 3.7.**

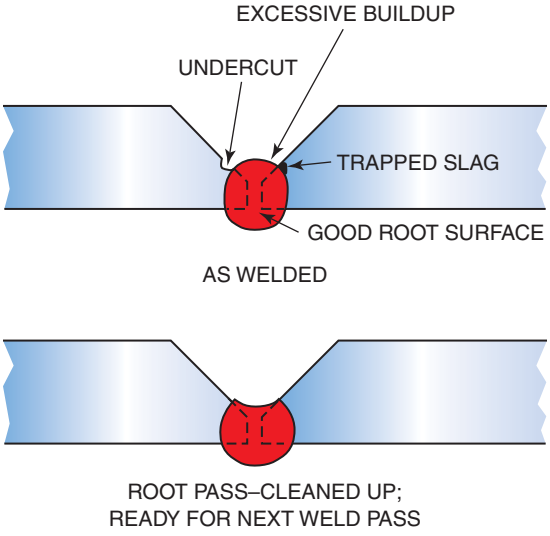

Figure 3.6
Grinding back the root pass to ensure a sound second pass

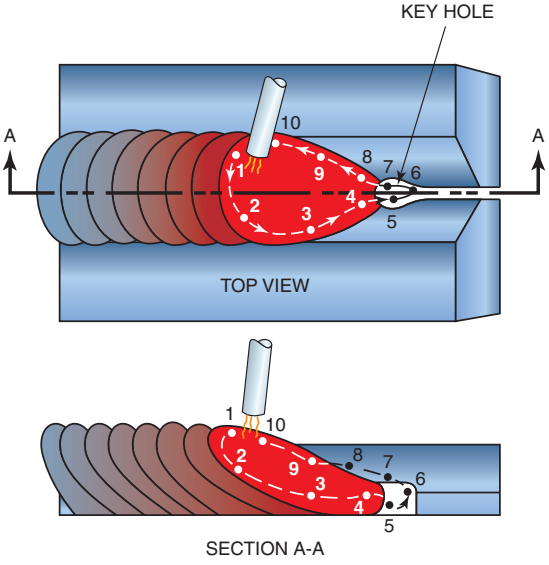

Figure 3.7
Electrode movement to open and use a keyhole

This is referred to as a **key hole,** and metal flows through the key hole to the root surface. The key hole must be maintained to ensure 100% penetration. This method requires more welder skill and can be used on a wide variety of joint conditions. The face of the bead resulting from this technique often is defect free.

PRACTICE 3-1

Root Pass on Plate with a Backing Strip in All Positions

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 6

Module 9
Key Indicator 1, 2

Using a properly set up and adjusted welding machine, proper safety protection, E6010 or E6011 arc welding electrodes having a 1/8-in. (3-mm) diameter, and one or more pieces of mild steel plate, 1/8 in. (3 mm) thick × 6 in. (152 mm) long, and one strip of mild steel, 1/8 in. (3 mm) thick × 1 in. (25 mm) wide × 6 in. (152 mm) long, you will make a root weld in all positions, **Figure 3.8.** Tack weld the plates together with a 1/16-in. (2-mm) to 1/8-in. (3-mm) root opening. Be sure there are no gaps between the backing strip and plates when the pieces are tacked together, **Figure 3.9.** If there is a small gap between the backing strip and the plates, it can be removed by placing the assembled test plates on an anvil and striking the tack weld with a hammer. This will close up the gap by compressing the tack welds, **Figure 3.10.**

Use a straight step or "T" pattern for this root weld. Push the electrode into the root opening so that there is good fusion with the backing strip and bottom edge of the plates. Failure to push the penetration deep into the joint will result in a cold lap at the root, **Figure 3.11.**

Watch the molten weld pool and keep its size as uniform as possible. As the molten weld pool increases in size, move the electrode out of the weld pool. When the weld pool begins to cool, bring the electrode back into the

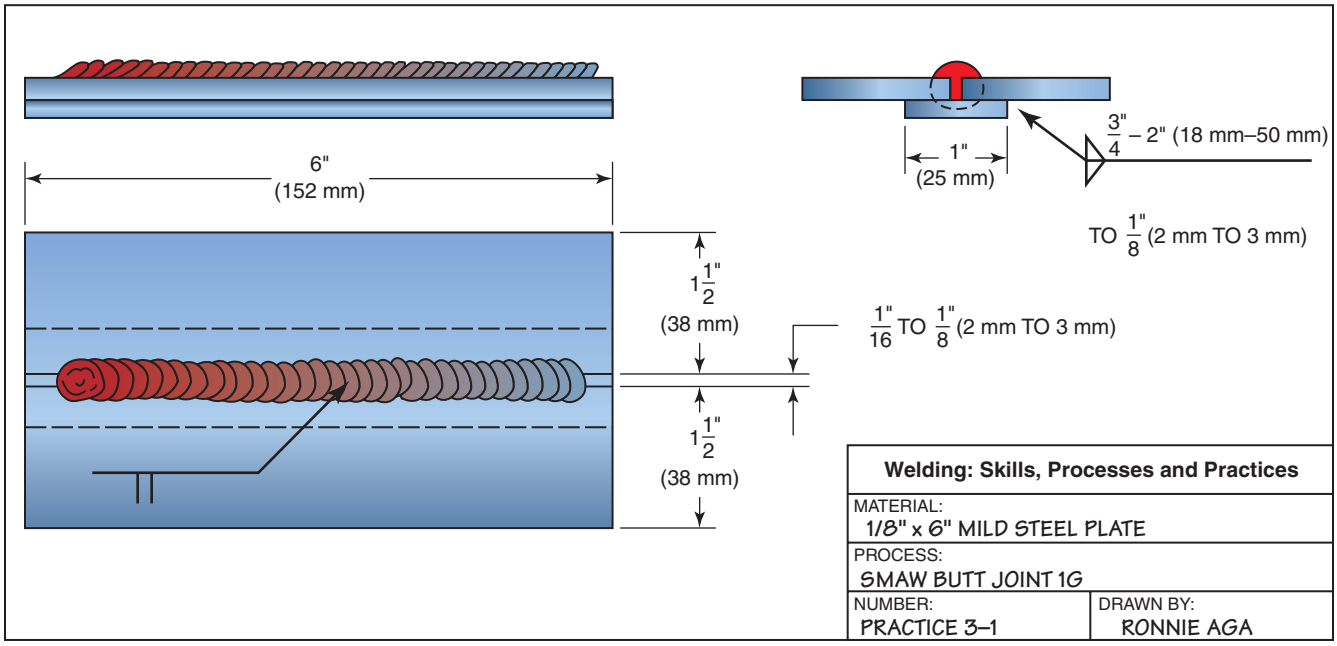

Figure 3.8
Square butt joint with a backing strip

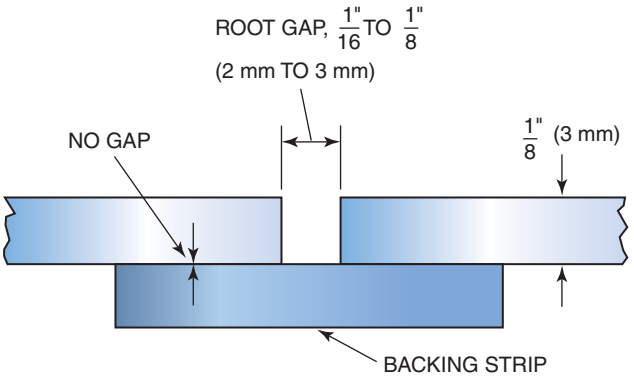

Figure 3.9
Backing strip

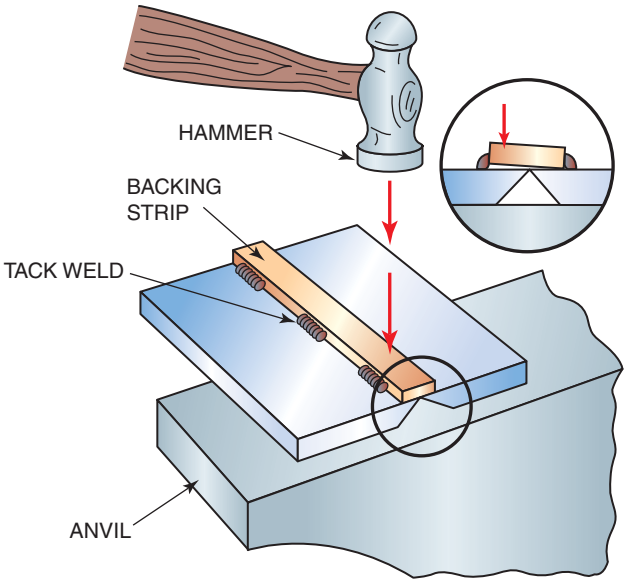

Figure 3.10
Using a hammer to align the backing strip and weld plates

molten weld pool. Use these weld pool indications to determine how far to move the electrode and when to return to the molten weld pool. After completing the weld, cut the plate and inspect the cross section of the weld for good fusion at the edges. Repeat the welds as necessary until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

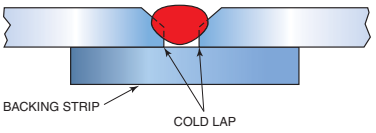

Figure 3.11
Incomplete root fusion

PRACTICE 3-2

Root Pass on Plate with an Open Root in All Positions

Using a properly set up and adjusted arc welding machine, proper safety protection, E6010 or E6011 arc welding electrodes with a 1/8-in. (3-mm) diameter, and two or more pieces of mild steel plate, 6 in. (152 mm) long × 1/8 in (3 mm) or 3/16 in. (4.7 mm.) thick, you will make a welded butt joint in all positions with 100% root penetration.

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 6

Module 9
Key Indicator 1, 2

- Tack weld the plates together with a root opening of 0 in. (0 mm) to 1/16 in. (2 mm).
- Using a short arc length and high amperage setting, make a weld along the joint.

You can change the electrode angle to control penetration and burn-through. As the trailing angle is decreased, making the electrode flatter to the plate, penetration, depth, and burn-through decrease, **Figure 3.12,** because both the arc force and heat are directed away from the bottom of the joint back toward the weld. Surface tension holds the metal in place, and the mass of the bead quickly cools the molten weld pool holding it in place. Increasing the electrode angle toward the perpendicular will increase penetration depth and possibly cause more burn-through. The arc

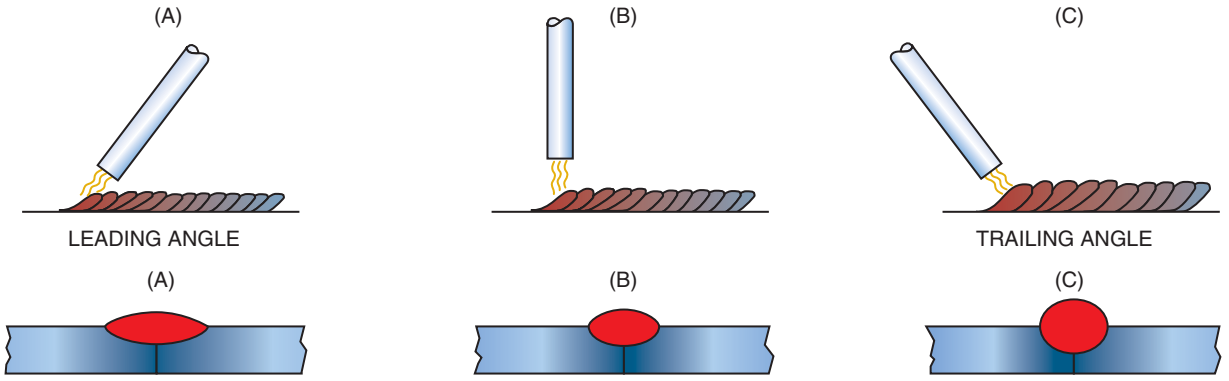

Figure 3.12
Effect of rod angle on weld bead shape

force and heat focused on the gap between the plates will push the molten metal through the joint.

The electrode holder can be slowly rocked from side to side while keeping the end of the electrode in the same spot on the joint, **Figure 3.13.** This will allow the arc force to better tie in the sides of the root to the base metal.

- When a burn-through occurs, rapidly move the electrode back to a point on the weld surface just before the burn-through.
- Lower the electrode angle and continue welding. If the burn-through does not close, stop the weld, chip, and wire brush the weld.
- Check the size of the burn-through. If it is larger than the diameter of the electrode, the root pass must be continued with the step method described in Practice 3-3. If the burn-through is not too large, lower the amperage slightly and continue welding.
- Watch the color of the slag behind the weld. If the weld metal is not fusing to one side, the slag will be brighter in color on one side. The

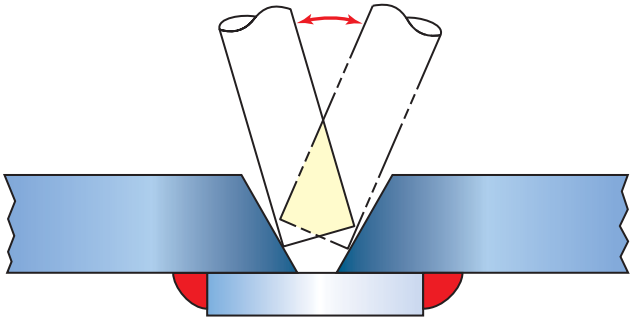

Figure 3.13
Rocking the top of the electrode while keeping the end in the same place helps control the bead shape

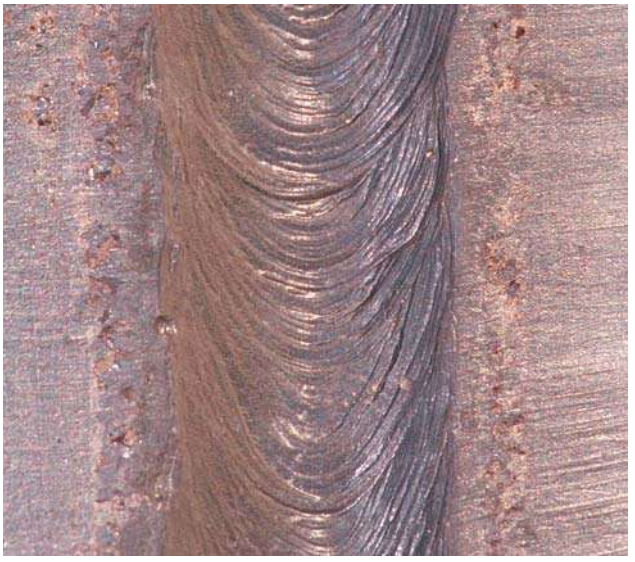

Figure 3.14
The weld toes appear uniform
Courtesy of Larry Jeffus

brighter color is caused by the slower cooling of the slag because there is less fused metal to conduct the heat away quickly.

- After the weld is completed, cooled, and chipped, check the back side of the plate for good root penetration. The root should have a small bead that will look as though it was welded from the back side, **Figure 3.14.** The penetration must be completely free of any drips of metal from the root face, called icicles.
- Repeat the welds as necessary until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 3-3

Open Root Weld on Plate Using the Step Technique in All Positions

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 6

Module 9
Key Indicator 1, 2

Using the same setup, materials, and electrodes as described in Practice 3-2, you will make a welded butt joint in all positions with 100% root penetration.

Tack weld the plates together with a root opening from 0 in. (0 mm) to 1/8 in. (3 mm). Using a medium amperage setting and a short stepping electrode motion, make a weld along the joint.

The electrode should be pushed deeply into the root to establish a key hole that will be used to ensure 100% root penetration. Once the key hole is established, the electrode is moved out and back in the molten weld pool at a steady, rhythmic rate. Watch the molten weld pool and key hole size to determine the rhythm and distance of electrode movement.

If the molten weld pool size decreases, the key hole will become smaller and may close completely. To increase the molten weld pool size and maintain the key hole, slow the rate of electrode movement and shorten the distance the electrode is moved away from the molten weld pool. This will increase the molten weld pool size and penetration because of increased localized heating.

If the molten weld pool becomes too large, metal may drip through the key hole, forming an icicle on the back side of the plate. Extremely large molten weld pool sizes can cause a large hole to be formed or cause burn-through. Repairing large holes can require much time and skill. To keep the molten weld pool from becoming too large, increase the travel speed, decrease the angle, shorten the arc length, or lower the amperage, **Table 3.1.**

The distance the electrode is moved from the molten weld pool and the length of time in the molten weld pool are found by watching the molten

Table 3.1 Changes Affecting Molten Weld Bead Size

	Amperage	Travel Speed	Electrode Size	Electrode Angle
To decrease puddle size	Decrease	Increase	Decrease	Leading
To increase puddle size	Increase	Decrease	Increase	Trailing

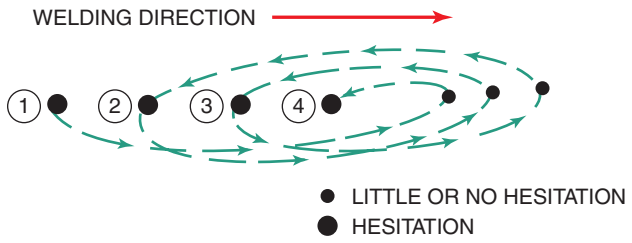

Figure 3.15
Weave pattern used to control the molten weld bead size

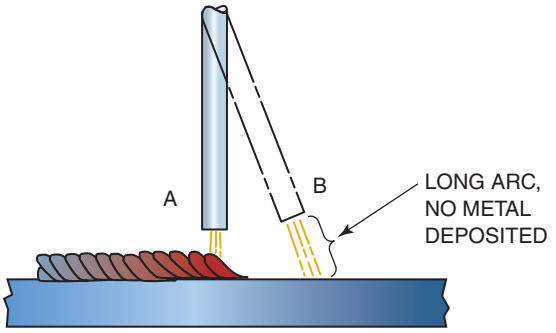

Figure 3.16
A long arc prevents metal or slag from being deposited ahead of the weld bead. A relatively quick movement from position "A" to position "B" and back to "A" is required.

weld pool. The molten weld pool size increases as you hold the arc in the molten weld pool until it reaches the desired size, about twice the electrode diameter, **Figure 3.15.** Move the electrode ahead of the molten weld pool, keeping the arc in the joint but being careful not to deposit any slag or metal ahead of the weld. To prevent metal and/or slag from transferring, raise the electrode to increase the arc length, **Figure 3.16.** Keep moving the electrode slowly forward as you watch the molten weld pool. The molten weld pool will suddenly start to solidify. At that time, move the electrode quickly back to the molten weld pool before it totally solidifies. Moving the electrode in a slight arc will raise the electrode ahead of the molten weld pool and automatically lower the electrode when it returns to the molten weld pool. Metal or slag deposited ahead of the molten weld pool may close the key hole, reduce penetration, and cause slag inclusions. Raising the end of the electrode too high or moving it too far ahead of the molten weld pool can cause all of the shielding gas to be blown away from the molten weld pool. If this happens, oxides can cause porosity. Keeping the electrode movement in balance takes concentration and practice.

Changing from one welding position to another requires an adjustment in timing, amperage, and electrode angle. The flat, horizontal, and overhead positions use about the same rhythm, but the vertical position may require a shorter time cycle for electrode movement. The amperage for the vertical position can be lower than that for the flat or horizontal, but the overhead position uses nearly the same amperage as flat and horizontal. The electrode angle for the flat and horizontal positions is about the same. For the vertical position, the electrode uses a sharper leading angle than does overhead, which is nearly perpendicular and may even be somewhat trailing.

Cool, chip, wire brush, and inspect both sides of the weld. The root surface should be slightly built up and look as though it was welded from that side (refer to Figure 3.14). Repeat the welds as necessary until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

HOT PASS

The surface of a root pass may be irregular, have undercut, overlap, slag inclusions, or other defects, depending upon the type of weld, the code or standards, and the condition of the root pass. The surface of a root pass can be cleaned by grinding or by using a **hot pass.**

On critical, high-strength code welds, it is usually required that the root pass as well as each filler pass be ground (refer to Figure 3.6). This grinding eliminates weld discontinuities caused by slag entrapments. It also can be used to remove most of the E60 series weld metal so that the stronger weld metal can make up most of the weld. When high-strength, low-alloy welding electrodes are used, this grinding is important to remove most of the low-strength weld deposit. This will leave the weld made up of nearly 100% of the high-strength weld metal, **Figure 3.17.**

The fastest way to clean out trapped slag and make the root pass more uniform is to use a hot pass. The hot pass uses a higher than normal amperage setting and a fast travel rate to reshape the bead and burn out trapped slag. After chipping and wire brushing the root pass to remove all the slag possible, a welder is ready to make the hot pass. The ideal way to apply a hot pass is to rapidly melt a large surface area, **Figure 3.18,** so that the trapped slag can float to the surface. The slag, mostly silicon dioxide (SiO_2), may not melt itself, so the surrounding steel must be melted to enable it to float free. The silicon dioxide may not melt because it melts at about 3100°F (1705°C), which is more than 500°F (270°C) hotter than the temperature at which the surrounding steel melts, around 2600°F (1440°C).

A very small amount of metal should be deposited during the hot pass so that the resulting weld is concave. A concave weld, compared to a

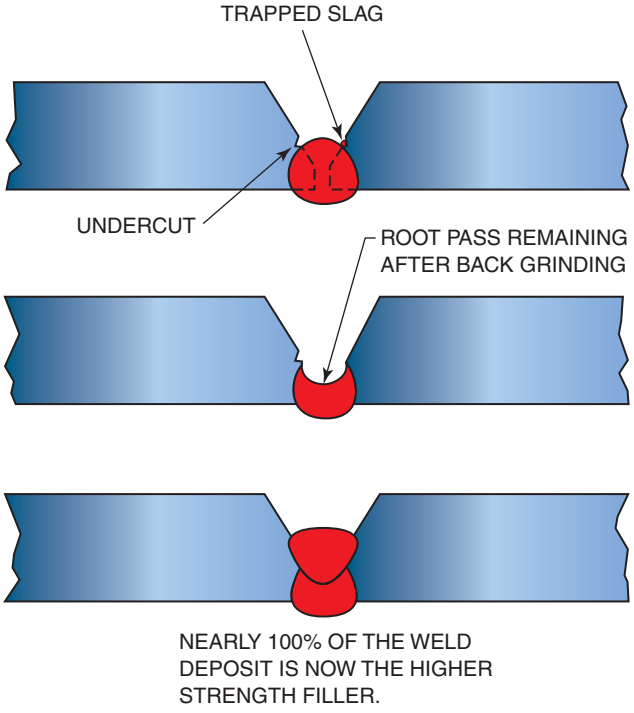

Figure 3.17
Back grinding to remove both discontinuities; filler metal used for the root pass

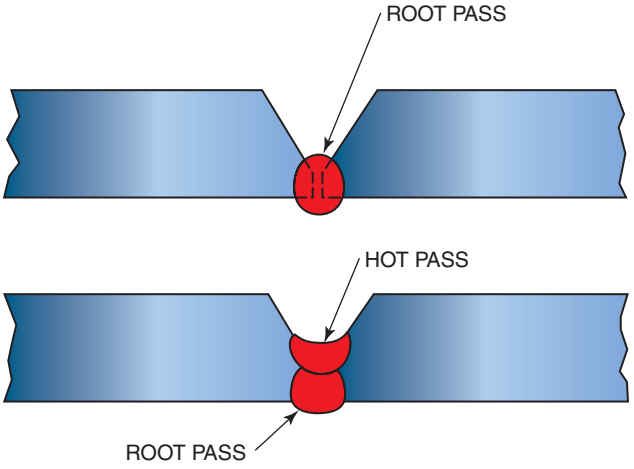

Figure 3.18
Using the hot pass to clean up the face of the root pass

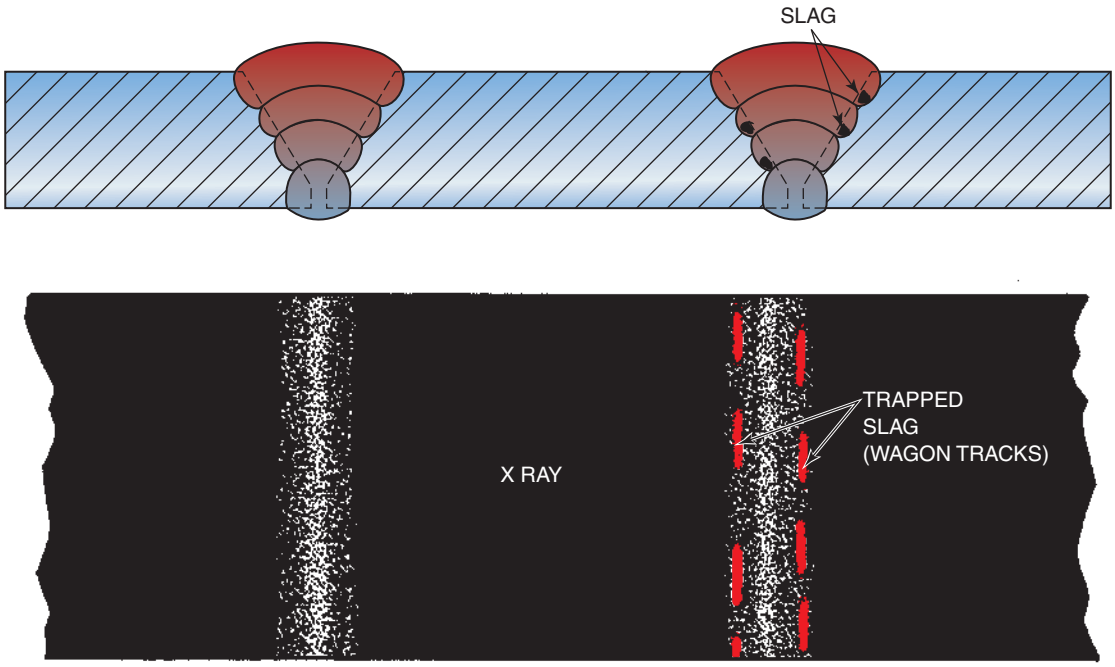

Figure 3.19
Slag trapped between passes will show on an X ray

convex weld, is more easily cleaned by chipping, wire brushing, or grinding. Failure to clean the convex root weld will result in a discontinuity showing up on an X ray. Such discontinuities are called **wagon tracks, Figure 3.19.**

The hot pass can also be used to repair or fill small spots of incomplete fusion or pinholes left in the root pass.

The normal weave pattern for a hot pass is the straight step or "T" pattern. The "T" can be used to wash out stubborn trapped slag better than the straight step pattern. The frequency of electrode movement is dependent upon the time required for the molten weld pool to start cooling. As with the root pass, metal or slag should not be deposited ahead of the bead. Do not allow the molten weld pool to cool completely or let the shielding gas covering to be blown away from the molten weld pool.

The hot pass technique can also be used to clean some welds that may first require grinding or gouging for a repair. The penetration of the molten weld pool must be deep enough to free all trapped slag and burn out all porosity.

EXPERIMENT 3-1

Hot Pass to Repair a Poor Weld Bead

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 4

Module 9
Key Indicator 1, 2

Using a properly set up and adjusted arc welding machine, proper safety protection, E6010 or E6011 arc welding electrodes having a 1/8-in. (3-mm) diameter, and two or more plates that have welds containing slag inclusions, lack of fusion, porosity, or other defects, you will make a hot pass to burn out the defects.

Chip and wire brush the weld bead. If necessary, use a punch to break apart large trapped slag deposits. The poorer the condition of the weld,

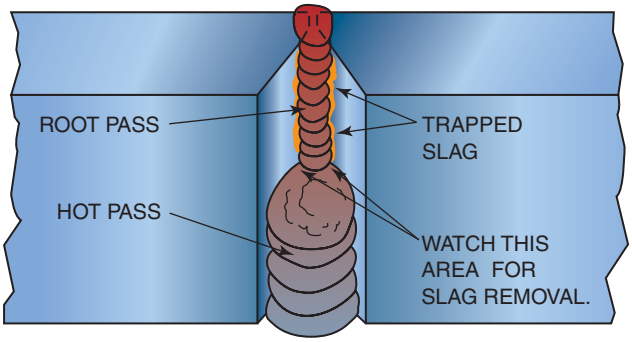

Figure 3.20
Burning out trapped slag by using a hot pass

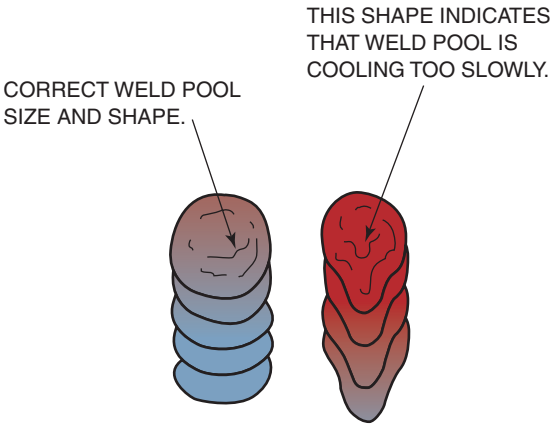

Figure 3.21
The shape of the weld pool can indicate the temperature of the surrounding base metal

the more vertical the joint should be for the hot pass. Large slag deposits tend to float around the molten weld pool and stay trapped in deep pockets in the flat position. With the weld in the vertical position, the slag can run out of the joint and down the face of the weld. Set the amperage as high as possible without overheating and burning up the electrode. Start at the bottom and weld upward using a combination of straight-step and "T" patterns to keep the weld deposit uniform. Watch the back edge of the molten weld pool for size and the weld crater for the complete burning out of impurities, **Figure 3.20.**

The plate may start to become overheated because of the high heat input. If you notice that the weld bead is starting to cool too slowly and is growing in length, you should stop welding, **Figure 3.21.** Allow the plate to cool before continuing the weld.

After the weld is completed, cool, chip, and inspect it for uniformity. The plate can be cut at places where you know large discontinuities existed before to see if they were repaired or only covered up. If you wish, this experiment can be repeated on other defects and joints.

Welds that have large defects in addition to excessive buildup may require some grinding to remove the buildup. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

CAUTION

This hot pass technique is designed to be used on noncritical, noncode welds only. It should not be used to cover bad welds or as a means of repairing the work of a welder who is less skilled.

FILLER PASS

After the root pass is completed and it has been cleaned, the groove is filled with weld metal. These weld beads make up the **filler passes.** More than one pass is often required.

Filler passes are made with stringer beads or weave beads. For multiple pass welds, the weld beads must overlap along the edges. They should overlap enough so that the finished bead is smooth, **Figure 3.22.** Stringer beads usually overlap about 50%, and weave beads overlap approximately 25%.

Each weld bead must be cleaned before the next bead is started. Slag left on the plate between welds cannot be completely burned out because

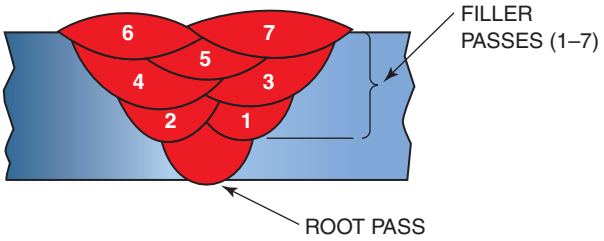

Figure 3.22
Filler passes—maximum thickness 1/8 in. (3 mm) each pass

filler welds should be made with a low amperage setting. Deep penetration will slow the rate of buildup in the joint. Deeply remelting the previous weld metal may weaken the joint. All that is required of a filler weld is that it be completely fused to the base metal.

Chipping, wire brushing, and grinding are the best ways to remove slag between filler weld passes. After the weld is completed, it can be checked by ultrasonic or radiographic nondestructive testing. Most schools are not equipped to do this testing. Therefore, a quick check for soundness can be made by destructive testing. One method of testing the deposited weld metal is by cutting and cross-sectioning the weld with an abrasive wheel and inspecting the weld. Another fast way to inspect filler passes is to cut a groove through the weld with a gouging tip. Watch the hot metal as it is washed away. The black spots that appear in the cut are slag inclusions. If only a few small spots appear, the weld probably will pass most tests. But, if a long string or large pieces of inclusions appear, the weld will most likely fail.

PRACTICE 3-4

Multiple Pass Filler Weld on a V-joint in All Positions

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 5

Module 9
Key Indicator 1, 2

Using a properly set up and adjusted arc welding machine, proper safety protection, E6010 or E6011 arc welding electrodes having a 1/8-in. (3-mm) diameter, and two or more pieces of mild steel plate, 6 in. (152 mm) long × 3 in. (76 mm) wide × 3/8 in. (10 mm) thick, you will make a multiple pass filler weld on a V-joint.

Tack weld the plates together at the corner so that they form a V, **Figure 3.23.** Starting at one end, make a stringer bead along the entire length using the straight step or "T" weave pattern. Thoroughly clean off the slag from the weld before making the next bead. **Figure 3.24** shows the suggested sequence for locating the beads. Continue making welds and cleaning them until the weld is 1 in. (25 mm) or more thick. Both ends of the weld may taper down. If it is important that the ends be square, metal tabs are welded on the ends of the plate for starting and stopping, **Figure 3.25.** The tabs are removed after the weld is completed.

After the weld is completed, it can be visually inspected for uniformity. If nondestructive testing is available, it may be checked for discontinuities. The weld also may be inspected by sectioning it with an abrasive wheel or gouging out with a torch. Repeat these welds until they are mastered. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

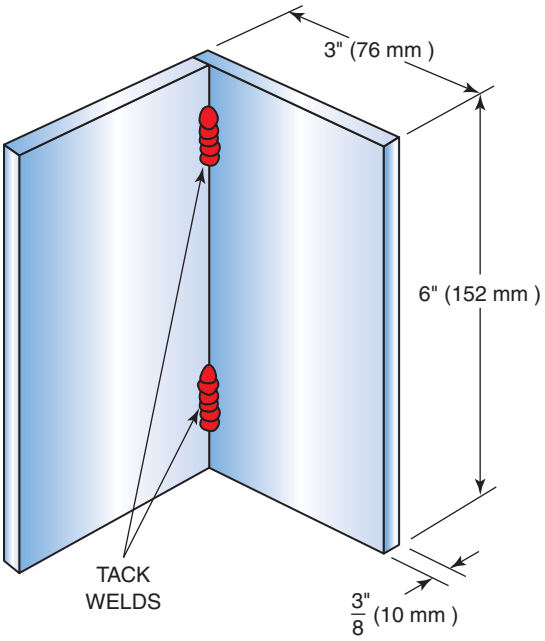

Figure 3.23
Tack weld the plates for the filler weld practice

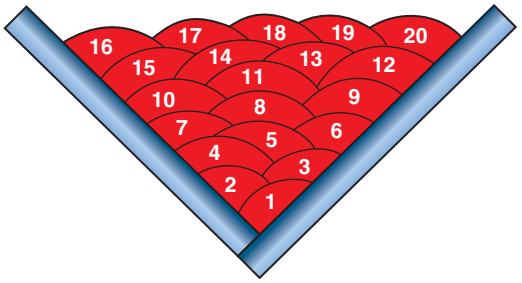

Figure 3.24
Filler pass buildup sequence

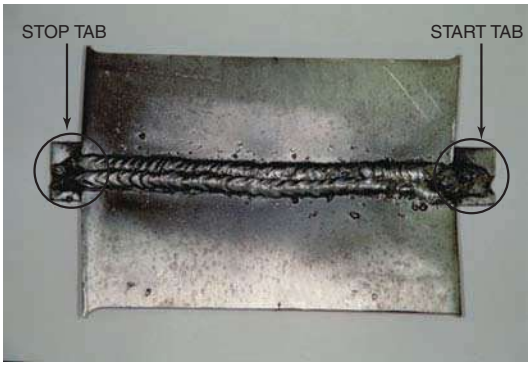

Figure 3.25
Run-off tabs help control possible under-fill or burn back at the starting and stopping points of a groove weld. The tabs can be removed.
Courtesy of Larry Jeffus

PRACTICE 3-5

Multiple Pass Filler Weld on a V-joint in All Positions Using E7018 Electrodes

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 5

Module 9
Key Indicator 1, 2

Using the same setup and materials as in Practice 3-4, and E7018 arc welding electrodes in place of E6010 or E6011 electrodes, you will make a multiple pass filler weld on a V-joint.

- Tack weld the two plates together at the corners so that they form a V.
- Using a slow, straightforward motion, with little or no stepping, "T" or inverted "V" motion, make a stringer bead along the root of the joint.
- Chip the slag and repeat the weld until there is a buildup of 1 in. (25 mm) or more.

If during the weld the buildup should become uneven or large slag entrapments occur, they should be ground out. In industry, groove welds in plate 1 in. (25 mm) or more are normally repaired. Making these repairs now is good experience for the welder. All welders will at some time make a weld that may need repairing.

- After the weld is completed, visually test and nondestructive test the weld for external and internal discontinuities.
- Repeat the welds as necessary until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

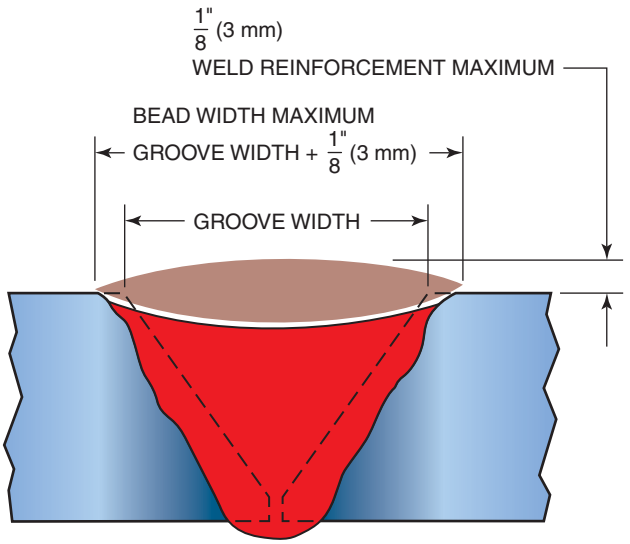

Figure 3.26
The cover pass should not be excessively large

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

COVER PASS

The last weld bead on a multiple pass weld is known as the **cover pass.** The cover pass may use a different electrode weave, or it may be the same as the filler beads. Keeping the cover pass uniform and neat looking is important. Most welds are not tested, and often the inspection program is only visual. Thus, the appearance might be the only factor used for accepting or rejecting welds.

The cover pass should be free of any visual defects such as undercut, overlap, porosity, or slag inclusions. It should be uniform in width and reinforcement. A cover pass should not be more than 1/8 in. (3 mm) wider than the groove opening, **Figure 3.26.** Cover passes that are too wide do not add to the weld strength.

PRACTICE 3-6

Cover Bead in All Positions

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 4

Module 9
Key Indicator 1, 2

Using a properly set up and adjusted arc welding machine, proper safety protection, E7018 arc welding electrodes having a 1/8-in. (3-mm) diameter, and one or more pieces of mild steel plate, 6 in. (152 mm) long × 1/4 in. (6 mm) thick, you will make a cover bead in each position.

Remember, any time an E7018 low-hydrogen-type electrode is to be used, the weave pattern, if used, must not be any larger than 2-1/2 times the diameter of the electrode. This weave cannot be any larger than 5/16 in. (7 mm) wide. Start welding at one end of the plate and weld to the other end. The weld bead should be about 5/16 in. (7 mm) wide having no more than 1/8 in. (3 mm) of uniform buildup, **Figure 3.27.** The weld buildup should have a smooth transition at the toe, with the plate and the face somewhat

Figure 3.27
Practice cover pass
Courtesy of Larry Jeffus

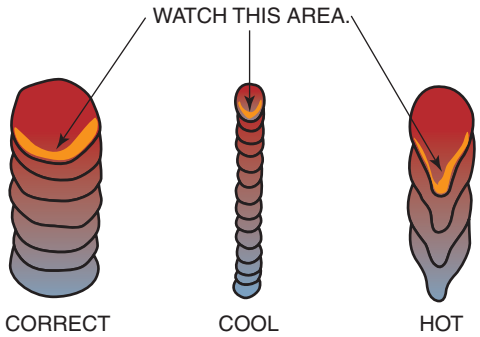

Figure 3.28
Watch the back edge of the weld pool to determine the correct current

convex. Undercut at the toe and a concave or excessively built-up face are the most common problems. Watch the sides of the bead for undercut. When undercut occurs, keep the electrode just ahead of the spot until it is filled in. There should be a smooth transition between the weld and the plate (refer to Figure 3.26). The shape of the bead face can be controlled by watching the trailing edge of the molten weld pool. That trailing edge is the same as the finished bead, **Figure 3.28.**

Deep penetration is not required with this weld and may even result in some weakening. After the weld is completely cooled, chip and inspect it for uniformity and defects. Repeat the welds as necessary until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PLATE PREPARATION

When welding on thick plate, it is impossible or impractical for the welder to try to get 100% penetration without preparing the plate for welding. The preparation of the plate is usually in the form of a **weld groove.** The groove can be cut into one side or both sides of the plate, and it may be cut into either just one plate or both plates of the joint, **Figure 3.29.** The type, depth, angle, and location of the groove are usually determined by a code standard that has been qualified for the specific job.

For SMA welds on plate 1/4 in. (6 mm) or thicker that need to have a weld with 100% joint penetration, the plate must be grooved. The groove may be ground, flame-cut, gouged, or machined on the edge of the plate before or after the assembly. Bevels and V-grooves are best if they are cut before the parts are assembled. J-grooves and U-grooves can be cut either before or after assembly, **Figure 3.30.** The lap joint is seldom prepared with a groove because little or no strength can be gained by grooving this joint. The only advantage to grooving the lap joint design is to give additional clearance.

Plates that are thicker than 3/8 in. (10 mm) can be grooved on both sides but may be prepared on only one side. The choice to groove one

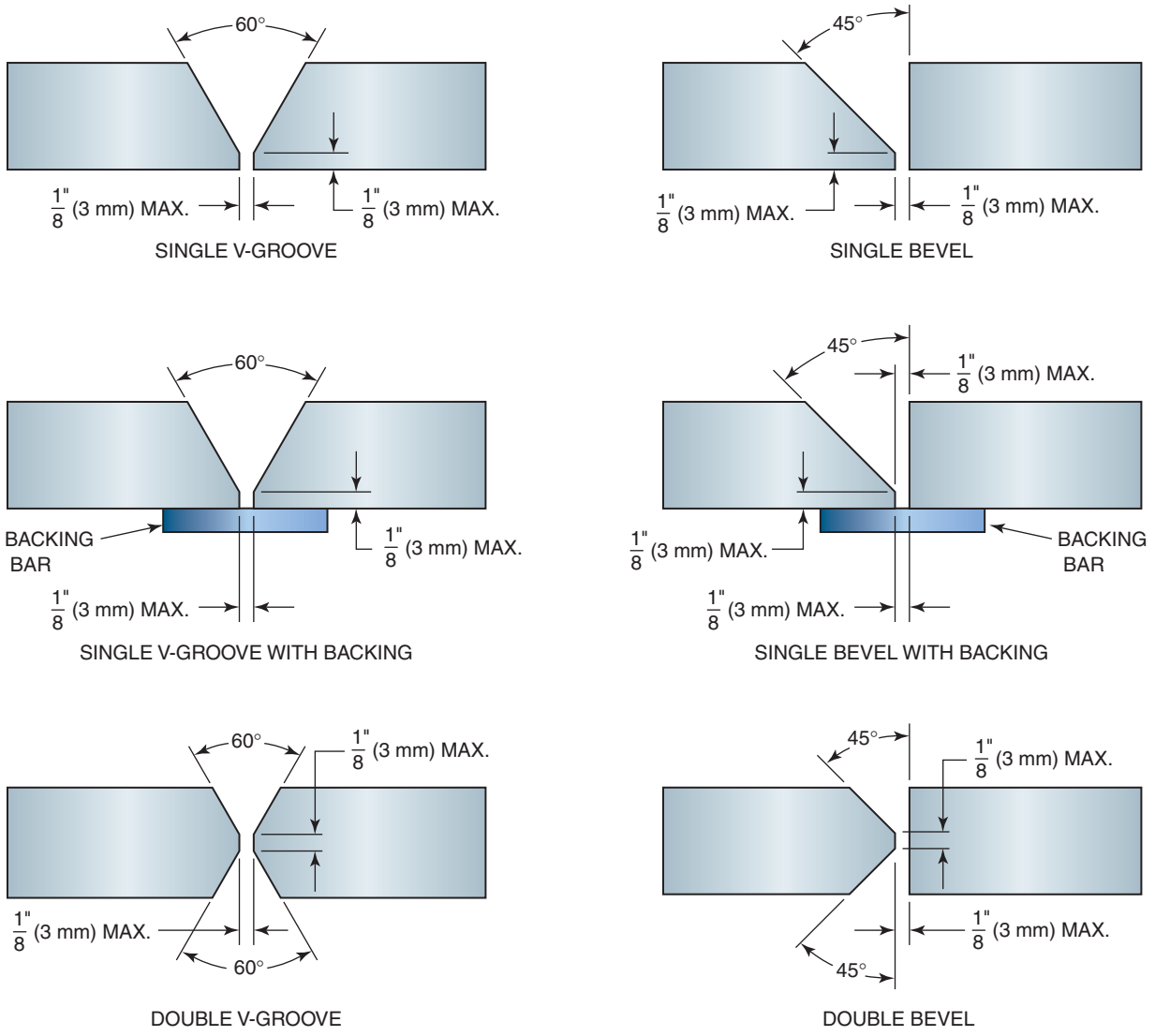

Figure 3.29
Typical butt joint preparations

or both sides is most times determined by joint design, position, and application. A tee joint in thick plate is easier to weld and will have less distortion if it is grooved on both sides. Plate in the flat position is usually grooved on only one side unless it can be repositioned. Welds that must have little distortion or that are going to be loaded equally from both sides are usually grooved on both sides. Sometimes plates are either grooved and welded or just welded on one side, and then back gouged and welded, **Figure 3.31. Back gouging** is a process of cutting a groove in the back side of a joint that has been welded. Back gouging can ensure 100% fusion at the root and remove discontinuities of the root pass. This process can also remove the root pass metal if the properties of the metal are not desirable to the finished weld, **Figure 3.32.** After back gouging, the groove is then welded. See Chapters 6, 7, and 8 in *Welding: Skills, Processes and Practices for Entry-Level Welders: Book One* for more information on the various methods of gouging.

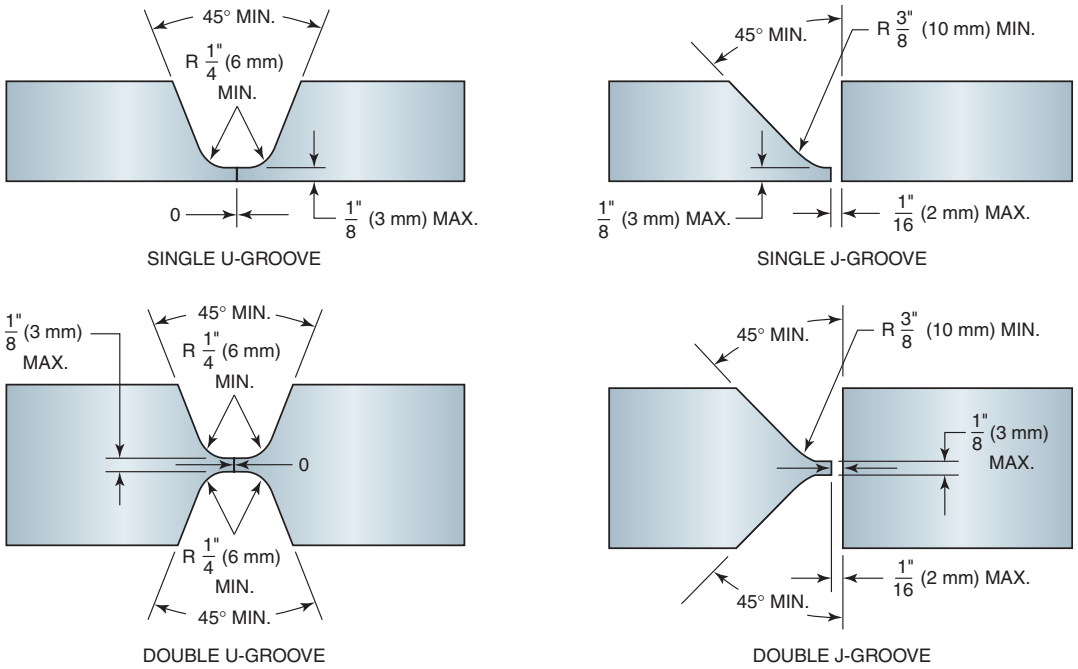

Figure 3.30
Typical butt joint preparations

Heavy plate and pipe sections requiring preparations are often used in products manufactured under a code or standard. The American Welding Society, American Society of Mechanical Engineers, and American Bureau of Ships are a few of the agencies that issue codes and specifications. The AWS D1.1 and the ASME Boiler and Pressure Vessel (BPV) Section IX standards will be used in this chapter as the standards for multiple pass groove welds that will be tested. The groove depth and angle are determined by the plate or pipe thickness and process.

After the plate is beveled, a grinder can be used to clean off oxides and improve the fitup.

Note

Root faces on bevels and grooves with backing bars show a 1/8-in. root face. The 1/8-in. root face is a maximum allowable dimension and many welders prefer root faces prepared thinner, or all the way to a "knife edge" (see Figure 3.2).

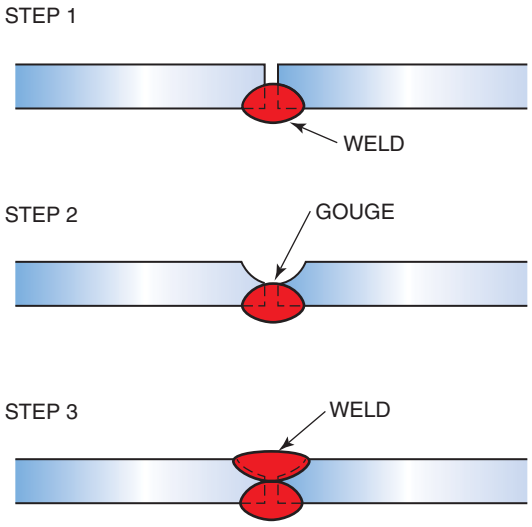

Figure 3.31
Back gouging sequence for a weld to ensure 100% joint penetration

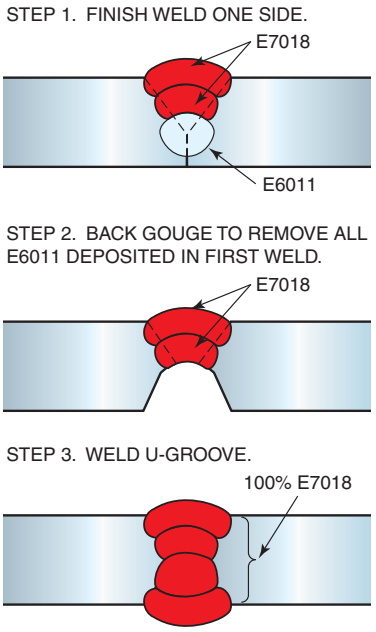

Figure 3.32
Back gouging to remove all weld metal used for the root pass or tacking

PREPARING SPECIMENS FOR TESTING

The detailed preparation of specimens for testing in this chapter is based on the structural welding code AWS D1.1 and the ASME BPV Code, Section IX. The maximum allowable size of fissures (crack or opening) in a **guided bend specimen** is given in codes for specific applications. Some of the standards are listed in ASTM E190 or AWS B4.0, AWS QC10, AWS QC11, and others. Copies of these publications are available from the appropriate organizations.

Acceptance Criteria for Face Bends and Root Bends

The weld specimen must first pass visual inspection before it can be prepared for bend testing. Visual inspection looks to see that the weld is uniform in width and reinforcement. There should be no arc strikes on the plate other than those on the weld itself. The weld must be free of both incomplete fusion and cracks. The joint penetration must be either 100% or as specified by the specifications. The weld must be free of overlap and undercut must not exceed either 10% of the base metal or 1/32 in. (0.8 mm), whichever is less.

Correct **weld specimen** preparation is essential for reliable results. The weld must be uniform in width and reinforcement and have no undercut or overlap. The weld reinforcement and backing strip, if used, must be removed flush to the surface, **Figure 3.33.** They can be machined or ground off. The plate thickness after removal must be a minimum of 3/8 in. (10 mm), and the pipe thickness must be equal to the pipe's original wall thickness. The specimens may be cut out of the test weldment by using an abrasive disc, by sawing, or by cutting with a torch. Flame-cut specimens must have the edges ground or machined smooth after cutting.

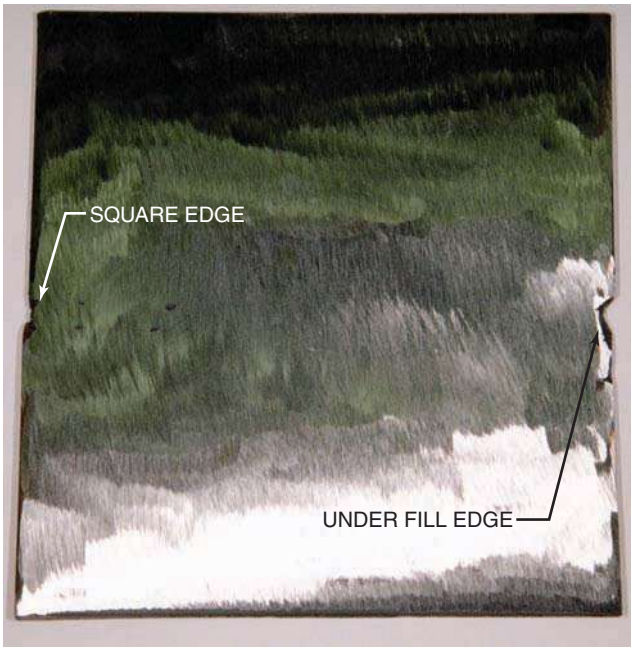

Figure 3.33
Plate ground in preparation for removing test specimens
Courtesy of Larry Jeffus

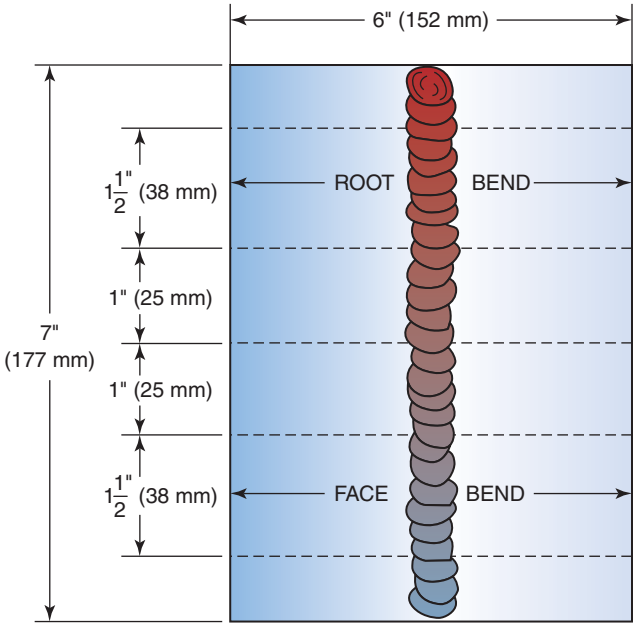

Figure 3.34
Sequence for removing guided bend specimens from the plate once welding is complete

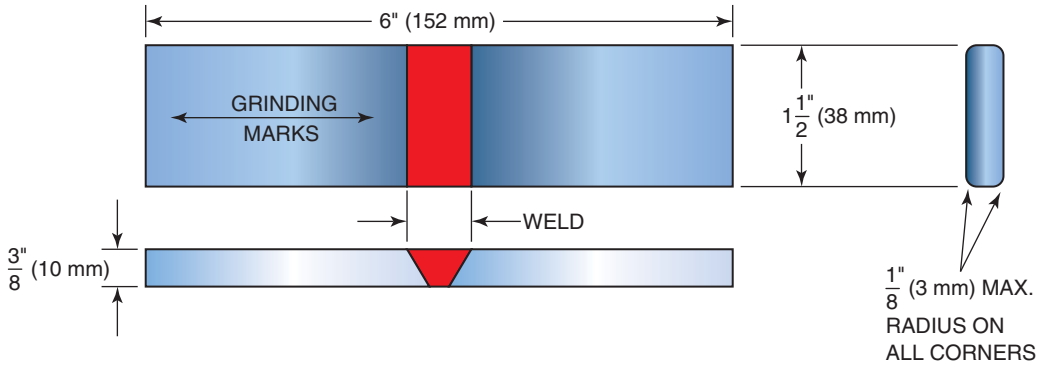

Figure 3.35
Guided bend specimen

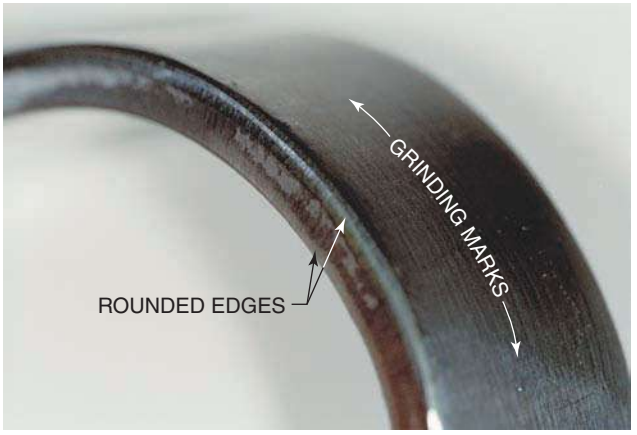

Figure 3.36
Guided bend test specimen
Courtesy of Larry Jeffus

This procedure is done to remove the heat-affected zone caused by the cut, **Figure 3.34.**

All corners must be rounded to a radius of 1/8 in. (3 mm) maximum, and all grinding or machining marks must run lengthwise on the specimen, **Figure 3.35** and **Figure 3.36.** Rounding the corners and keeping all marks running lengthwise reduce the chance of good weld specimen failure due to poor surface preparation.

The weld must pass both the face and root bends in order to be acceptable. After bending there can be no single defects larger than 1/8 in. (3 mm), and the sum of all defects larger than 1/32 in. (0.8 mm) but less than 1/8 in. (3 mm) must not exceed a total of 3/8 in. (10 mm) for each bend specimen. An exception is made for cracks that start at the edge of the specimen and do not start at a defect in the specimen.

RESTARTING A WELD BEAD

On all but short welds, the welding bead will need to be restarted after a welder stops to change electrodes. Because the metal cools as a welder changes electrodes and chips slag when restarting, the penetration and buildup may be adversely affected.

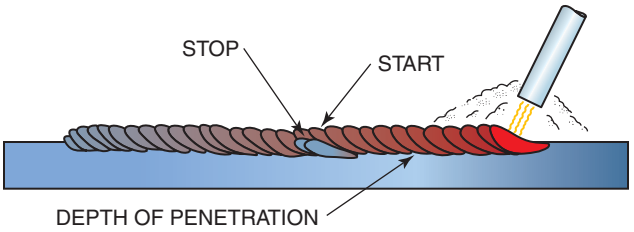

Figure 3.37
Tapering the size of the weld bead helps keep the depth of penetration uniform

When a weld bead is nearing completion, it should be tapered so that when it is restarted the buildup will be more uniform. To taper a weld bead, the travel rate should be increased just before welding stops. A 1/4-in. (6-mm) taper is all that is required. The taper allows the new weld to be started and the depth of penetration reestablished without having excessive buildup, **Figure 3.37.**

The slag should always be chipped and the weld crater should be cleaned each time before restarting the weld. This is important to prevent slag inclusions at the start of the weld.

The arc should be restarted in the joint ahead of the weld. The electrodes must be allowed to heat up so that the arc is stabilized and a shielding gas cloud is reestablished to protect the weld. Hold a long arc as the electrode heats up so that metal is not deposited. Slowly bring the electrode downward and toward the weld bead until the arc is directly on the deepest part of the crater where the crater meets the plate in the joint, **Figure 3.38.** The electrode should be low enough to start transferring metal. Next, move the electrode in a semicircle near the back edge of the weld crater. Watch

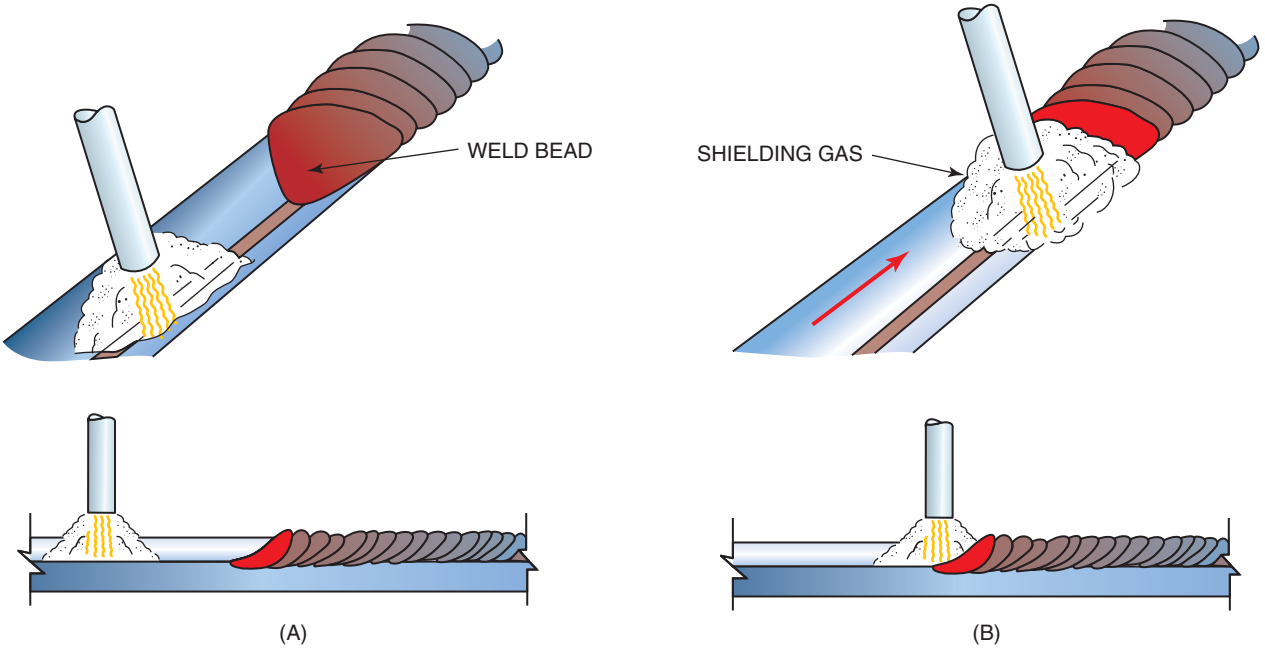

Figure 3.38
When restarting the arc, strike the arc ahead of the weld in the joint (A). Hold a long arc and allow time for the electrode to heat up, forming the protective gas envelope. Move the electrode so that the arc is focused directly on the leading edge (root) of the previous weld crater (B).

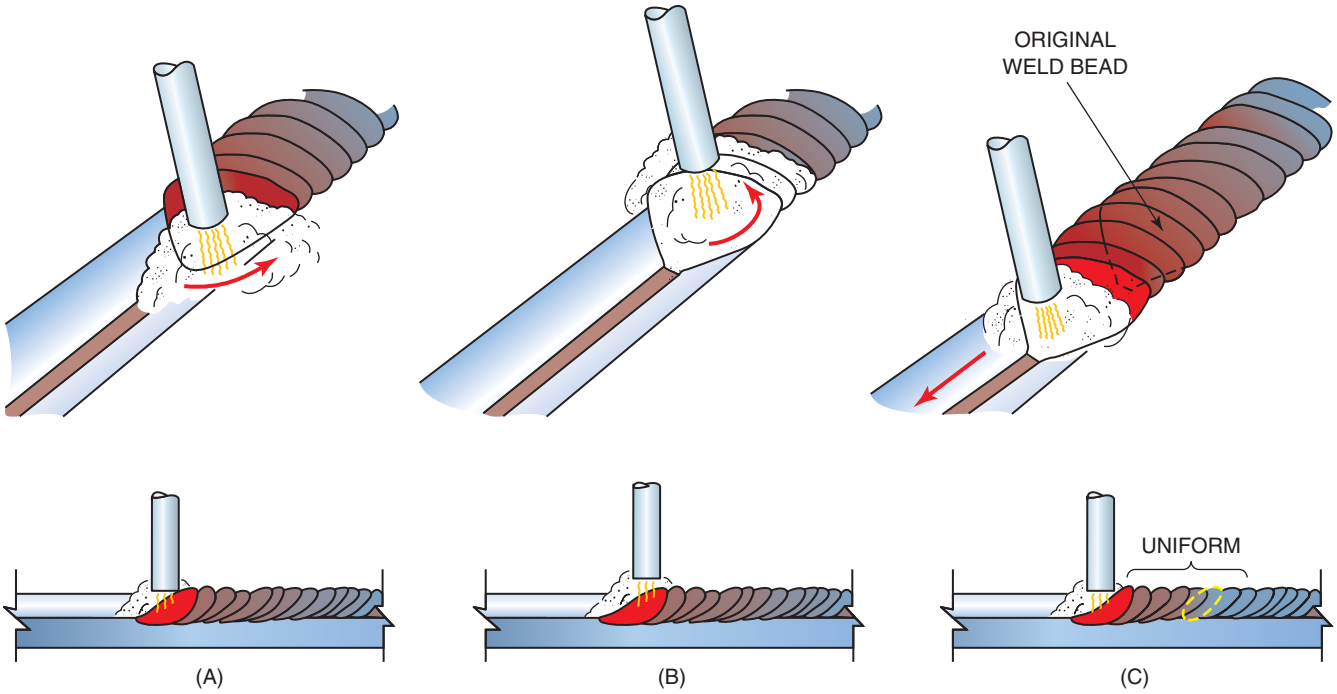

Figure 3.39
When restarting the weld pool after the root has been heated to the melting temperature, move the electrode upward along one side of the crater (A). Move the electrode along the top edge, depositing new weld metal (B). When the weld is built up uniformly with the previous weld, continue along the joint (C).

the buildup and match your speed in the semicircle to the deposit rate so that the weld is built up evenly, **Figure 3.39.** Move the electrode ahead and continue with the same weave pattern that was being used previously.

The movement to the root of the weld and back up on the bead serves both to build up the weld and reheat the metal so that the depth of penetration will remain the same. If the weld bead is started too quickly, penetration is reduced and buildup is high and narrow.

Starting and stopping weld beads in corners should be avoided. Tapering and restarting are especially difficult in corners, and this often results in defects, **Figure 3.40.**

Figure 3.40
Correct method of welding through a corner. Stopping on a corner may cause leaks.
Courtesy of Larry Jeffus

Table 3.2 Preheat Temperatures for Arc Welding on Low-carbon Steels*

Plate Thickness in. (mm)	Minimum Temperature °F	°C
Up to 1/2 in. (13 mm)	70	21
1/2 in. (13 mm) to 1 in. (25 mm)	100	38
1 in. (25 mm) to 2 in. (51 mm)	200	95
Over 2 in. (51 mm)	300	150

*Metal should be above the dew point.
Allow 1 hour for each inch in order to provide uniform heating or for localized preheating. Check to ensure that preheat temperatures are sufficient to melt thermal crayons a minimum of 3 inches (76 mm) in all directions from the area to be welded prior to welding arc application.

PREHEATING AND POSTHEATING

Preheating is the application of heat to the metal before it is welded. This process helps to reduce cracking, hardness, distortion, and stresses by reducing the thermal shock from the weld and slowing the cool-down rate. Preheating is most often required on large, thick plates, when the plate is very cold, on days when the temperature is very cold, when small-diameter electrodes are used, on high-carbon or manganese steels, on complex shapes, or with fast welding speeds.

With the practices that are to be tested in this chapter, preheating should be used if the base metal to be welded is very cold. It may also be used to reduce distortion on thick sections and to reduce hardness caused by the rapid cooling of the weld, which may result in weld failure. Preheating the metal will slow the weld cooling rate, which results in a more ductile weld. **Table 3.2** lists the recommended preheat temperatures for plain carbon steels.

PRACTICE 3-7

Limited Thickness Welder Performance Qualification Test Plate with Backing and F3 and F4 Electrodes

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 1, 2, 3, 4, 6

Module 9
Key Indicator 1, 2

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 3-7. Date:

Title:
Welding SMAW of plate to plate.

Scope:
This procedure is applicable for V-groove plate with a backing strip within the range of 3/8 in. (10 mm) through 3/4 in. (20 mm). Welding may be performed in the following positions: 1G, 2G, 3G, and 4G.

Base Metal:
The base metal shall conform to M1020 or A36.
Backing material specification M1020 or A36.

Filler Metal:
The filler metal shall conform to AWS specification No. E6010 or E6011 root pass and E7018 for the cover pass from AWS specification A5.1. This filler metal falls into F-number F3 and F4 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity: N/A.

Joint Design and Tolerances:

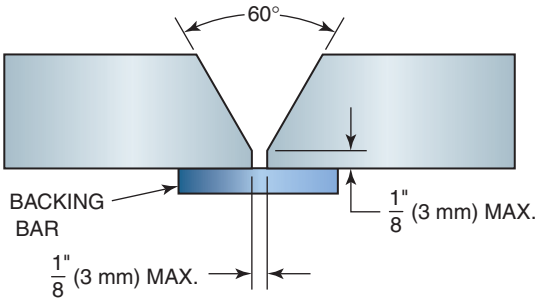

SINGLE V-GROOVE WITH BACKING

Preparation of Base Metal:
The V-groove is to be ground, flame-cut, or machined on the edge of the plate before the parts are assembled. Root faces may be prepared from a knife edge to the maximum allowable thickness of 1/8 in. All parts must be cleaned prior to welding of all contaminants, such as paints, oils, grease, or primers. Both inside and outside surfaces within 1 in. (25 mm) of the joint must be mechanically cleaned using a wire brush or grinder.

Electrical Characteristics:
The current shall be AC or DCRP.

The base metal shall be on the work lead or negative side of the line.

Preheat:
The parts must be heated to a temperature higher than 70°F (21°C) before any welding is started.

Backing Gas:
N/A

Welding Technique:
Tack weld the plates together with the backing strip. There should be about a 1/8-in. (3-mm) root gap between the plates. Use the E6010 or E6011 arc welding electrodes to make a root pass to fuse the plates and backing strip together. Clean the slag from the root pass and use either a hot pass or grinder to remove any trapped slag.

Using the E7018 arc welding electrodes, make a series of filler welds in the groove until the joint is filled.

Interpass Temperature:
The plate should not be heated to a temperature higher than 400°F (205°C) during the welding process. After each weld pass is completed, allow it to cool; the weldment must not be quenched in water.

Cleaning:
The slag can be chipped and/or ground off between passes but can only be chipped off of the cover pass.

Inspection:

Visually inspect the weld for uniformity and other discontinuities with criteria found in practice 3-8. If the weld passes the visual inspection, then it is to be prepared and guided bend tested according to criteria found in practice 3-8. Repeat each of the welds as needed until you can pass this test.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

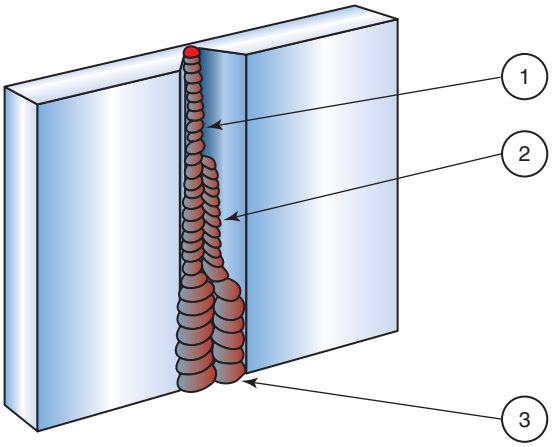

Note the sequence in depositing the beads

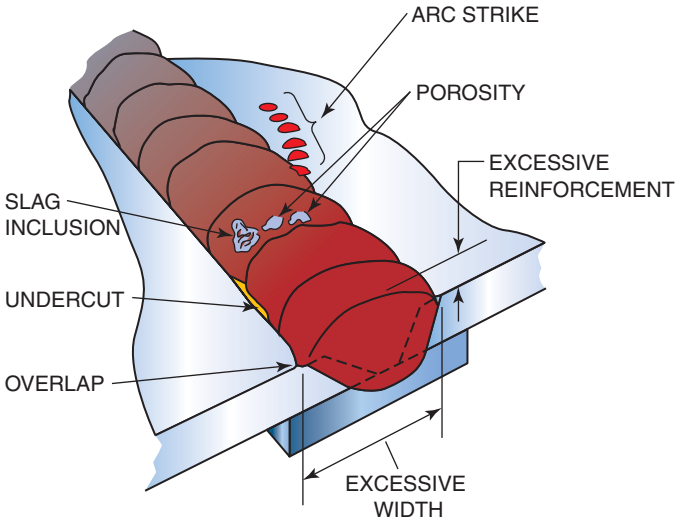

Common discontinuities found during a visual examination

Sketches:

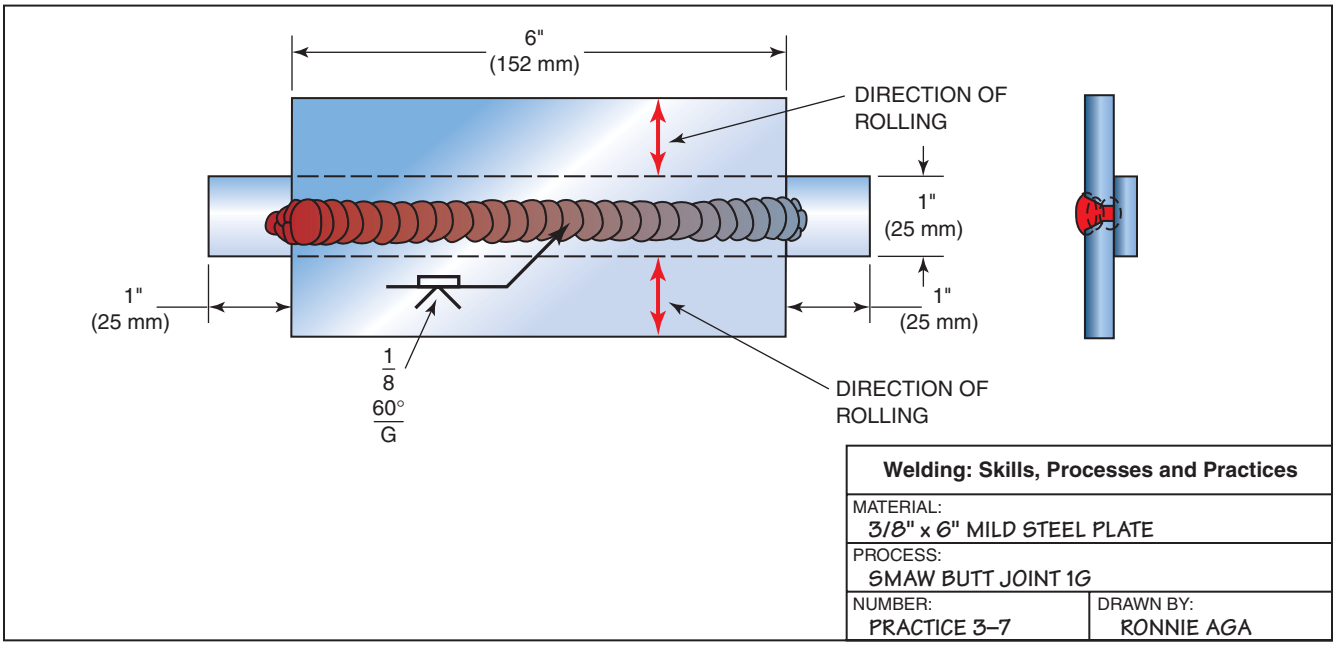

PRACTICE 3-8

SMAW Workmanship Sample and Welder Qualification Test Plate for Limited Thickness Horizontal 2G and 3G Positions with E7018 Electrodes

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 1, 2, 3, 4, 6

Module 9
Key Indicator 1, 2

Dimensional Tolerances:

2G Test plates: two (2); each 3/8 in. (10 mm) thick, 3 in. (75 mm) wide, and 7 in. (175 mm) long, one having a 45° bevel along one edge.
3G Test plates: two (2); each 3/8 in. (10 mm) thick, 3 in. (75 mm) wide, and 7 in. (175 mm) long, one having a 45° included bevel.
Backing strip: one (1); each either 1/4 in. (6 mm) or 3/8 in. (10 mm) thick, 1 in. (25 mm) wide, and 9 in. (228 mm) long.

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 3-8. Date:

Title:

Welding SMAW of plate to plate.

Scope:

This procedure is applicable for single bevel or V-groove plate with a backing strip within the range of 3/16 in. (5 mm) through 3/4 in. (20 mm).
Welding may be performed in the following positions: 2G and 3G.

Base Metal:

The base metal shall conform to carbon steel M-1 or P-1, Group 1 or 2.
Backing material specification carbon steel M-1 or P-1, Group 1, 2, or 3.

Filler Metal:

The filler metal shall conform to AWS specification no. E7018 from AWS specification A5.1. This filler metal falls into F-number F-4 and A-number A-1.

Shielding Gas:

The shielding gas, or gases, shall conform to the following compositions and purity: N/A.

Joint Design and Tolerances:

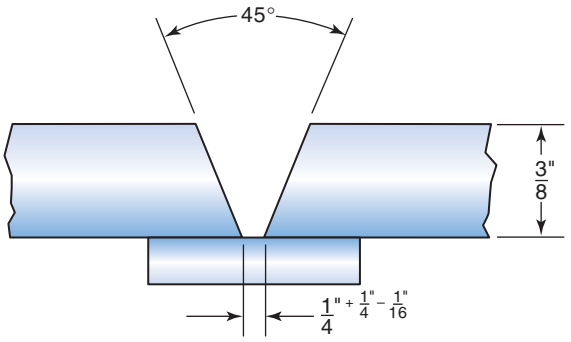

Preparation of Base Metal:

The bevel is to be flame-cut on the edge of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches that are deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The backing strip, groove face, and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Welds	Filler Metal Diameter	Current	Amperage Range
Tack	3/32 in. (2.4 mm)	DCEP	70 to 115
Root and Filler	1/8 in. (3 mm)	DCEP	115 to 165
Filler	1/8 in. (3 mm)	DCEP	115 to 140

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Tack weld the plates together with the backing strip. There should be about a 1/4-in. (6-mm) root gap between the plates. Use the E7018 arc welding electrodes to make a root pass to fuse the plates and backing strip together. Clean the slag from the root pass, being sure to remove any trapped slag along the sides of the weld.

Using the E7018 arc welding electrodes, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6 mm), in the groove until the joint is filled.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
The slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand-chipping hammer, a punch and hammer, or a needle scaler. All weld cleaning must be performed with the test plate in the welding position.

Visual Inspection Criteria for Entry Welders*:
There shall be no cracks, no incomplete fusion. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.

*Courtesy of American Welding Society

The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.

Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm).

Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.

The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).

Welds shall be free from overlap.

Sketches:

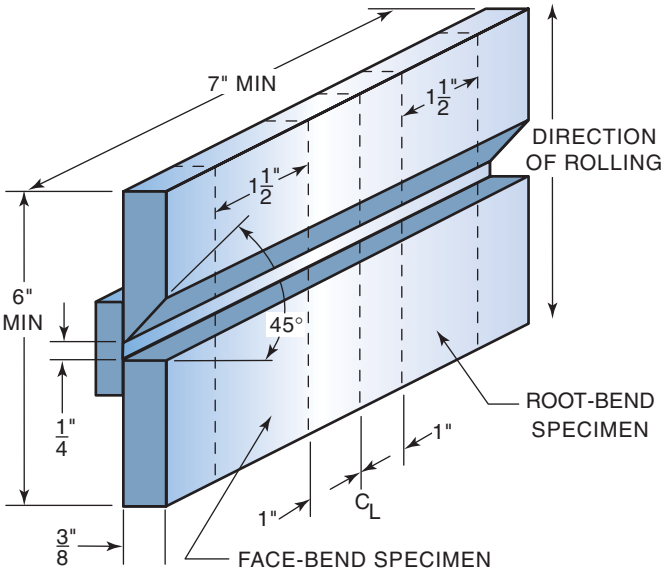

Bend Test:

The weld is to be mechanically tested only after it has passed the visual inspection. Be sure that the test specimens are properly marked to identify the welder, the position, and the process.

Specimen Preparation

For 3/8-in. (10-mm) test plates, two specimens are to be located in accordance with the requirements below. One is to be prepared for a transverse face bend, and the other is to be prepared for a transverse root bend.

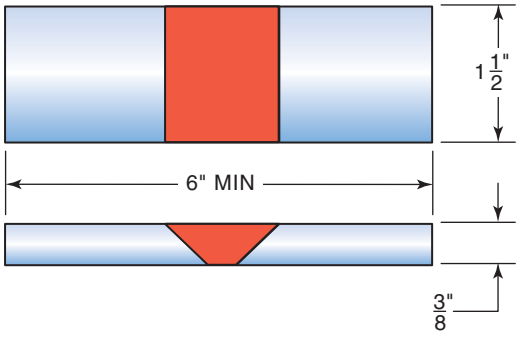

- *Transverse face bend.* The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld face becomes the tension surface of the specimen. Transverse face bend specimens shall comply with the requirements of below.
- *Transverse root bend.* The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld root becomes the tension surface of the specimen. Transverse root bend specimens shall comply with the requirements of above.

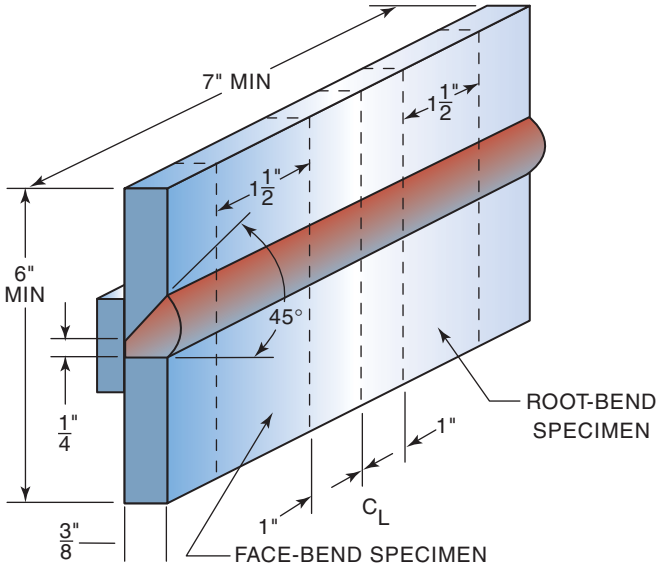

Acceptance Criteria for Bend Test*:

For acceptance, the convex surface of the face and root bend specimens shall meet both of the following requirements:

1. No single indication shall exceed 1/8 in. (3 mm), measured in any direction on the surface.
2. The sum of the greatest dimensions of all indications on the surface, which exceed 1/32 in. (0.8 mm), but are less than or equal to 1/8 in. (3 mm), shall not exceed 3/8 in. (10 mm).

*Courtesy of American Welding Society

Cracks occurring at the corner of the specimens shall not be considered unless there is definite evidence that they result from slag or inclusions or other internal discontinuities.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 3-9

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 1, 2, 3, 6

Module 9
Key Indicator 1, 2

Limited Thickness Welder Performance Qualification Test Plate Without Backing

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 3-9. Date:

Title:

Welding SMAW of plate to plate.

Scope:
This procedure is applicable for V-groove plate without a backing strip within the range of 3/8 in. (10 mm) through 3/4 in. (20 mm).

Welding may be performed in the following positions: 1G, 2G, 3G, and 4G.

Base Metal:
The base metal shall conform to M1020 or A36.
Backing material specification: M1020 or A36.

Filler Metal:
The filler metal shall conform to AWS specification no. E6010 or E6011 root pass and E7018 for the cover pass from AWS specification A5.1. This filler metal falls into F-number F3 and F4 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity: N/A.

Joint Design and Tolerances:

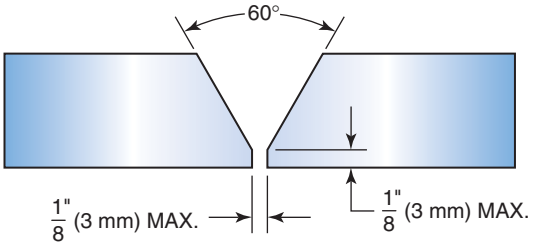

SINGLE V-GROOVE

Preparation of Base Metal:
The V-groove is to be ground, flame-cut, or machined on the edge of the plate before the parts are assembled. All parts must be cleaned prior to welding of all contaminants, such as paints, oils, grease, or primers. Both inside and outside surfaces within 1 in. (25 mm) of the joint must be mechanically cleaned using a wire brush or grinder.

Electrical Characteristics:
The current shall be DCRP. The base metal shall be on the work lead or negative side of the line.

Preheat:
The parts must be heated to a temperature higher than 70°F (21°C) before any welding is started.

Backing Gas:
N/A

Welding Technique:
Tack weld the plates together; there should be about a 1/8-in. (3-mm) root gap between the plates. Use the E6010 or E6011 arc welding electrodes to make a root pass to fuse the plates together. Clean the slag from the root pass and use either a hot pass or grinder to remove any trapped slag.

Using the E7018 arc welding electrodes, make a series of filler welds in the groove until the joint is filled.

Interpass Temperature:
The plate, outside of the heat affected zone, should not be heated to a temperature higher than 400°F (205°C) during the welding process. After each weld pass is completed, allow it to cool; the weldment must not be quenched in water.

Cleaning:
The slag can be chipped and/or ground off between passes but can only be chipped off of the cover pass.

Inspection:
Visually inspect the weld for uniformity and other discontinuities by using the criteria found in Practice 3-8. If the weld passes the visual inspection, then it is to be prepared and guided bend tested according to the "guided bend test" criteria found in Practice 3-8. Repeat each of the welds as needed until you can pass this test.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Sketches:

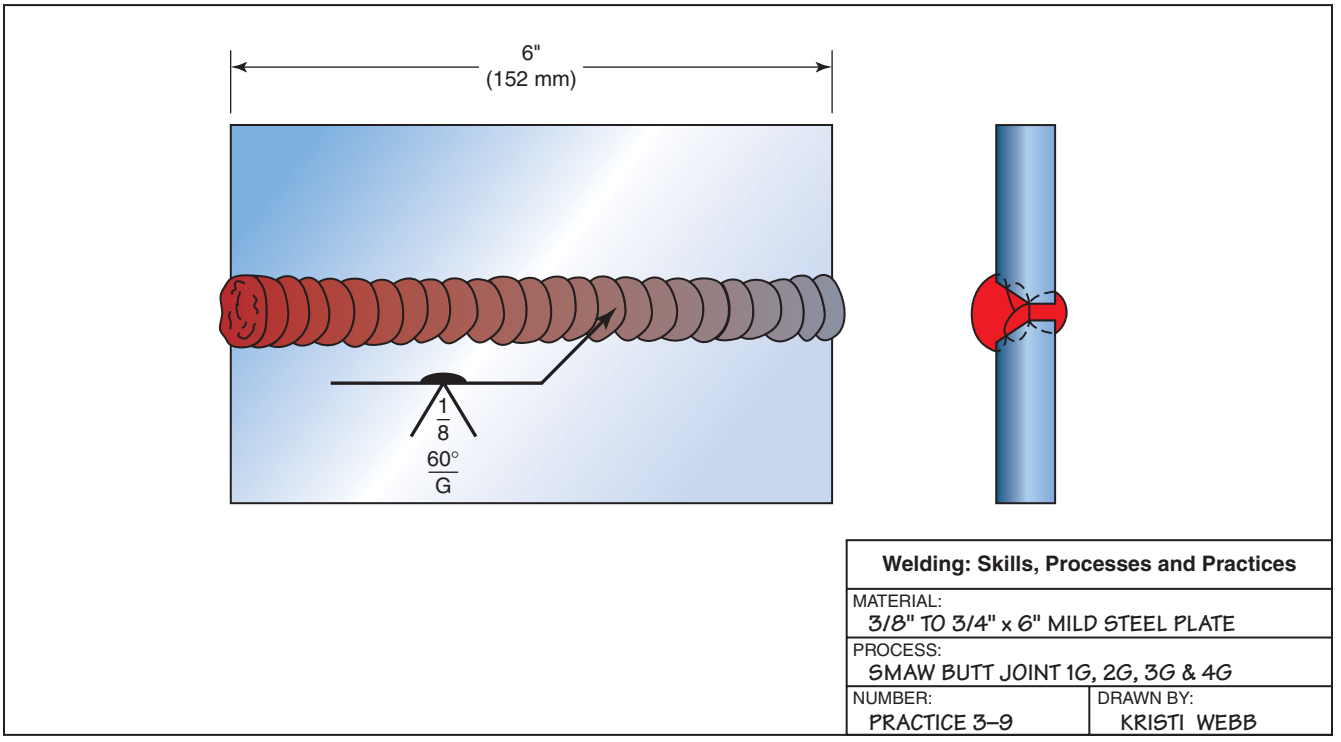

Postheating is the application of heat to the metal after welding. Postheating is used to slow the cooling rate and reduce hardening.

Interpass temperature is the temperature of the metal during welding. The interpass temperature is given as a minimum and maximum. The minimum temperature is usually the same as the preheat temperature. If the plate cools below this temperature during welding, it should be reheated. The maximum temperature may be specified to keep the plate below a certain phase change temperature for the mild steel used in these

practices. The maximum interpass temperature occurs when the weld bead cannot be controlled because of a slow cooling rate. When this happens, the plate should be allowed to cool down, but not below the minimum interpass temperature.

If, during the welding process, a welder must allow a practice weldment to cool so that the weld can be completed later, the weldment should be cooled slowly and then reheated before starting to weld again. A weld that is to be tested or that is done on any parts other than scrap should not be quenched in water.

POOR FITUP

Ideally, all welding will be performed on joints that are properly fitted. Most welds produced to a code or standard are properly fitted. Repair, prototype, and job shop welding, however, may not be cut and fitted properly. These welds must be performed under less than ideal conditions, but they still must be strong and have a good appearance.

To make a good weld on a poorly fitted joint requires some special skills. These welds also require a good welder, one whose skill is developed. A skilled welder can watch the molten weld pool and knows how to correct for problems before they develop into disasters. The welder must be able to read the molten weld pool correctly to make needed changes in amperage, current, electrode movement, electrode angle, and timing.

The amperage setting may have to be adjusted up or down by only a few amperes to make the necessary changes in molten weld pool size. Adjusting the machine is often preferable to lengthening the weave pattern excessively. The current may be changed from AC to DCSP or DCRP to vary the amount of heat input to the molten weld pool. Some electrodes can operate better than other electrodes with lower amperages on some currents. The current also will alter the forcefulness of some electrodes.

The "U," "J," and straight step patterns are usually the best to use, but they should not be moved more than required to close the gap or opening. On some poor-fitting joints, it is necessary to break and restart the arc in order to keep the molten weld pool under control. This will result in a weld with porosity, slag inclusions, and other defects. But it is often better to have a poor weld than to have no weld. A poor root weld can be capped with a sound weld to improve joint strength.

Changing the electrode angle from leading to trailing improves poor fit. Sometimes a very flat angle will also help. The time interval that the electrode is moved into and out of the molten weld pool is critical in maintaining weld control. Returning to a molten weld pool too often or too soon can cause the molten weld pool to drop out of the joint. In most cases, a welder should return to the molten weld pool only after it has started to cool.

On some joints, it is possible to make stringer beads on both sides of the joint until the gap is closed, **Figure 3.41.** Note that the beads are made alternately from the edges of the joint to the center. Welds made in this manner can have good weld soundness and strength, but they require more time to complete.

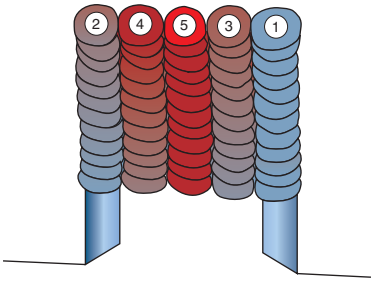

Figure 3.41
Multiple stringer beads used to close a large gap

PRACTICE 3-10

Single V-groove Open Root Butt Joint with an Increasing Root Opening

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 6

Module 9
Key Indicator 1, 2

For this practice, you will need a properly set up and adjusted arc welding machine, proper safety protection, E6010 or E6011 and E7018 arc welding electrodes having a 1/8-in. (3-mm) diameter, and two or more pieces of mild steel plate, 3/8 in. (10 mm) thick × 4 in. (102 mm) wide × 12 in. (305 mm) long. You will weld a single V-groove open root butt joint that has a poor fitup, starting from the close end.

Tack weld the plates together with a root opening of 1/8 in. (3 mm) at one end and 1/2 in. (13 mm) at the other end, **Figure 3.42.** Using the E6010 or the E6011 electrode, start the root pass at the narrow end and weld to the other end. As the root pass progresses along the widening root gap, care must be taken to maintain molten weld pool control. The "J" or "U" weave pattern works best with low current settings. Long time intervals for electrode movements will give the best weld control.

When the root pass is completed, clean the weld and make a hot pass to burn out any trapped slag. Finish the weld with filler passes using the E7018 electrode. Cool, chip, and inspect the weld for uniformity and defects. Repeat these welds until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 4

Module 4
Key Indicator 6

Module 9
Key Indicator 1, 2

PRACTICE 3-11

Single V-groove Open Root Butt Joint with a Decreasing Root Opening

Using the same setup, equipment, and materials as described in Practice 3-10, you will weld a single V-groove open root butt joint that has a poor fitup, starting from the wide end.

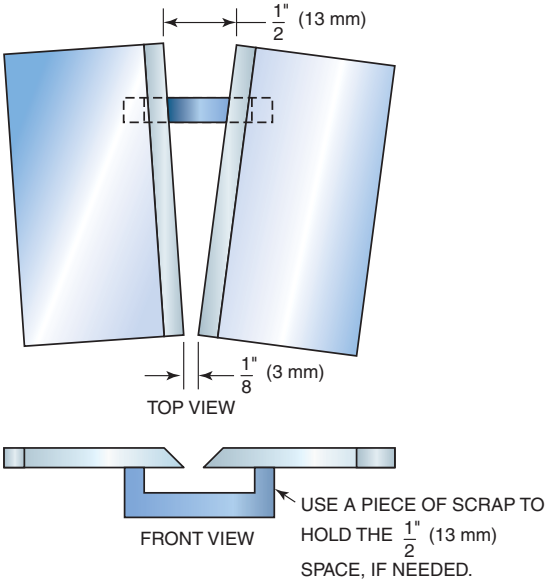

Figure 3.42
Welding specimen with poor fitup

As in Practice 3-10, tack weld the plates together with a root opening of 1/8 in. (3 mm) at one end and 1/2 in. (13 mm) at the other end. Using the E6010 or E6011 electrode, start welding the root pass at the wide end. Both sides of the joint must be built up until it is possible to get the metal to flow together. The "J" or "U" weave pattern works best to control the bead.

When the root pass is completed, make a hot pass to clean out any trapped slag before making the filler passes with the E7018 electrode. After the weld is completed, cool, chip, and inspect it for uniformity and defects. Repeat these welds until you can consistently make welds free of defects. Turn off the welding machine and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

SUMMARY

Grooved welds on approximately 1/2-in. (13-mm)-thick plate are the most common test plates given to new welding applicants. Groove welds are used by many companies as the base welding skills performance test requirement for employment. The vertical and overhead positions are the most commonly used for the test. It is often assumed in the welding industry that a uniform, visually defect-free weld will successfully pass destructive testing. This assumption has great basis in fact because, in most cases, such a weld reflects the welder's skills required to produce quality welds, so, in many cases, applicant test welds are evaluated only by the weld shop foreman or supervisor for visual defects. For this reason, you should always attempt to make your welds as uniform in appearance as possible. Learning how to make a "pretty" groove weld can often mean the difference between successfully earning the job and losing out to another applicant.

REVIEW

1. Why are some weld joints grooved?
2. Sketch four of the standard grooves used for welded joints.
3. Why are some backing strips removed from the finished weld?
4. Why are backing tapes used on some joints?
5. Why is it very important to make a weld with a good root surface?
6. What are the two common methods of making a root pass on an open root joint?
7. How can small gaps between the weld plate and backing strip be closed?
8. What effect does changing from a trailing angle to a leading angle have on a weld?
9. What benefit would there be to the root pass if the electrode holder were rocked back and forth while keeping the electrode tip in the joint?
10. What might cause the bright color on the flux as a weld cools?

11. What can happen if the molten weld pool becomes too large on a root weld?
12. What can be done to increase the amount of high-strength welding electrode in the final weld if the root weld was made with a low-alloy electrode?
13. What is the purpose of the hot pass?
14. Why should a filler weld pass not have deep penetration?
15. Why is it important to have a good cover pass?
16. What can watching the back edge of a weld pool help you determine?
17. Other than penetration, why would thick butt joints be grooved on both sides?
18. List the things that a weld must be inspected for before it is ground for bend testing.
19. What determines the acceptance or rejection of a bend specimen?
20. What technique can be used to make restarting weld beads easier and more uniform?
21. Why should some weldments be preheated before welding starts?
22. What is postheating used for?
23. How can a wide gap in a joint be closed by welding?

CHAPTER

4 Gas Tungsten Arc Welding Equipment, Setup, Operation, and Filler Metals

OBJECTIVES

After completing this chapter, the student should be able to

- describe the gas tungsten arc welding process and list the other terms used to identify the process
- list four attributes of a quality tungsten electrode
- list four precautions taken to limit tungsten erosion
- contrast the various types of tungsten electrodes and how they are used
- shape the end of the tungsten electrode and clean it
- grind a point on a tungsten electrode using an electric grinder
- remove a contaminated tungsten end
- melt the end of the tungsten electrode into the desired shape
- list two advantages and two disadvantages of water-cooled GTA welding torches and air-cooled torches
- describe the purposes of the three hoses connecting a water-cooled torch to the welding machine
- choose an appropriate nozzle for the job
- accurately read a flowmeter
- compare and contrast the three types of welding current used for GTA welding
- list three advantages and three disadvantages of the shielding gases used in the GTA welding process
- set preflow time and postflow time on a GTA welding workstation
- list three problems that can occur as a result of an incorrect gas flow rate
- set up a GTAW workstation for carbon steel, stainless steel, and aluminum operations
- strike a GTA welding arc on carbon steel, stainless steel, and aluminum

KEY TERMS

cleaning action
collet
flowmeter
frequency
inert gas
postflow time
preflow time
rectification
spark gap oscillator
tungsten

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports or records
Key Indicator 4: Follows written instructions to complete work assignments

Module 7: Gas Tungsten Arc Welding (GTAW)

Key Indicator 1: Performs Safety inspections of GTAW equipment and accessories
Key Indicator 2: Makes minor external repairs to GTAW equipment and accessories

Carbon Steel

Key Indicator 3: Sets up for GTAW operations on carbon steel
Key Indicator 4: Operates GTAW equipment on carbon steel

Austenitic Stainless Steel

Key Indicator 8: Sets up for GTAW operations on austenitic stainless steel
Key Indicator 9: Operates GTAW equipment on austenitic stainless steel

Aluminum

Key Indicator 13: Sets up for GTAW operations on aluminum
Key Indicator 14: Operates GTAW equipment on aluminum

INTRODUCTION

The gas tungsten arc welding (GTAW) process is sometimes referred to as TIG, or heliarc. The term *TIG* is short for tungsten inert gas welding, and the much older term *heliarc* was used because helium was the first gas used for the process. The aircraft industry developed the GTAW process for welding magnesium during the late 1930s and the early 1940s. During that time, helium was the primary shielding gas used, along with DCEP welding current. These caused many of the problems that limited application of the GTA welding process. But improvements in gas composition and a better understanding of the importance of polarity improved the process's effectiveness and reduced its cost.

To use this process, an arc is established between a nonconsumable tungsten electrode and the base metal, which is called the work. Under the correct welding conditions, the electrode does not melt, although the work does at the spot where the arc impacts its surface and produces a molten weld pool. The filler metal is thin wire that is fed directly into the molten weld pool, where it melts. Since hot tungsten is sensitive to oxygen contamination, a good inert shielding gas is required to keep the air away from the hot tungsten and mol-

ten weld metal. The inert gas provides the needed arc characteristics and protects the molten weld pool. Because fluxes are not used, the welds produced are sound, free of slags, and as corrosion-resistant as the parent metal.

Before development of the GTAW process, welding aluminum and magnesium was difficult. The welds produced were porous and prone to corrosion.

When argon became plentiful and DCEN was recognized as more suitable than DCEP, the GTA process became more common. Until the development of Gas Metal Arc Welding (GMAW) in the late 1940s, GTAW was the only acceptable process for welding such reactive materials as aluminum, magnesium, titanium, and some grades of stainless steel, regardless of thickness. Reactive metals are ones that are easily contaminated when heated to their molten welding temperatures. Contamination can come from the air or can be picked up from surfaces containing oxides, oils, paints, or similar materials. Although economical for welding sheet metal, the process proved tedious and expensive for joining sections much thicker than 1/4 in. (6 mm). The eye-hand coordination required to make GTA welds is very similar to the coordination required for oxyfuel gas welding. GTA welding is often easier to learn when a person can oxyfuel gas weld.

TUNGSTEN

Tungsten, atomic symbol W, has the following properties:

- high tensile strength: 500,000 lb/in.^{2} (3447 kg/mm^2)
- hardness: Rockwell C45
- high melting temperature: 6170°F (3410°C)
- high boiling temperature: 10,700°F (5630°C)
- good electrical conductivity

Tungsten is produced mainly by reduction of its trioxide with hydrogen. Powdered tungsten is then purified to 99.95+%, compressed, and sintered (heated to a temperature below melting where grain growth can occur) to make an ingot. The ingot is heated to increase ductility and then is swaged and drawn through dies to produce electrodes. These electrodes are available in sizes varying from 0.01 in. to 0.25 in. (0.25 mm to 6 mm) in diameter. The tungsten electrode, after drawing, has a heavy black oxide that is later chemically cleaned or ground off.

The high melting temperature and good electrical conductivity make tungsten the best choice for a nonconsumable electrode. Its arc temperature, around 11,000°F (6000°C), is much higher than its melting temperature but not much higher than its boiling temperature of 10,700°F (5630°C).

As the tungsten electrode becomes hot, the arc between the electrode and the work will stabilize. Because electrons are more freely emitted from hot tungsten, the very highest temperature possible at the tungsten electrode tip is desired. Maintaining a balance between the heat necessary to have a stable arc and heat that is high enough to melt the tungsten requires an understanding of the GTA torch and electrode.

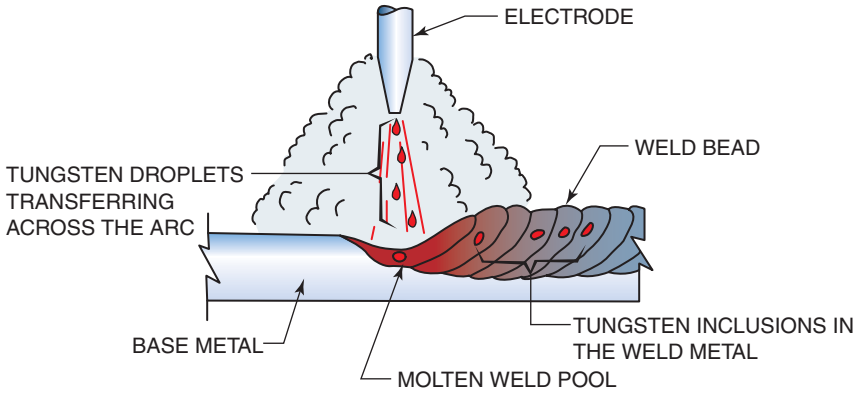

Figure 4.1
Some tungsten will erode from the electrode, be transferred across the arc, and become trapped in the weld deposit

The thermal conductivity of tungsten and the heat input are prime factors in the use of tungsten as an electrode. In general, tungsten is a good conductor of heat. This conductive property is what allows the tungsten electrode to withstand the arc temperature well above its melting temperature. The heat of the arc is conducted away from the electrode's end so fast that it does not reach its melting temperature. For example, a wooden match burns at approximately 3000°F (1647°C). Because aluminum melts at 1220°F (971°C), a match should easily melt an aluminum wire. However, a match will not even melt a 1/16-in. (2-mm) aluminum wire. The aluminum, like a tungsten electrode, conducts the heat away so quickly that it will not melt.

Because of the intense heat of the arc, some erosion of the electrode will occur. This eroded metal is transferred across the arc, **Figure 4.1.** Slow erosion of the electrode results in limited tungsten inclusions in the weld, which are acceptable. Standard codes give the size and amount of tungsten inclusions that are allowable in various types of welds. The tungsten inclusions are hard spots that cause stresses to concentrate, possibly resulting in weld failure. Although tungsten erosion cannot be completely eliminated, it can be controlled. A few ways of limiting erosion include:

- having good mechanical and electrical contact between the electrode and the collet
- using as low a current as possible
- using a water-cooled torch
- using as large a size of tungsten electrode as possible
- using DCEN current
- using as short an electrode extension from the collet as possible
- using the proper electrode end shape
- using an alloyed tungsten electrode

The torch end of the electrode is tightly clamped in a collet. The collet inside the torch is cooled by air or water. The **collet** is the cone-shaped sleeve that holds the electrode in the torch. Heat from both the arc and the tungsten electrode's resistance to the flow of current must be absorbed by the collet and torch. To ensure that the electrode is being cooled properly, be sure the collet connection is clean and tight. And for water-cooled torches, make sure water flow is adequate.

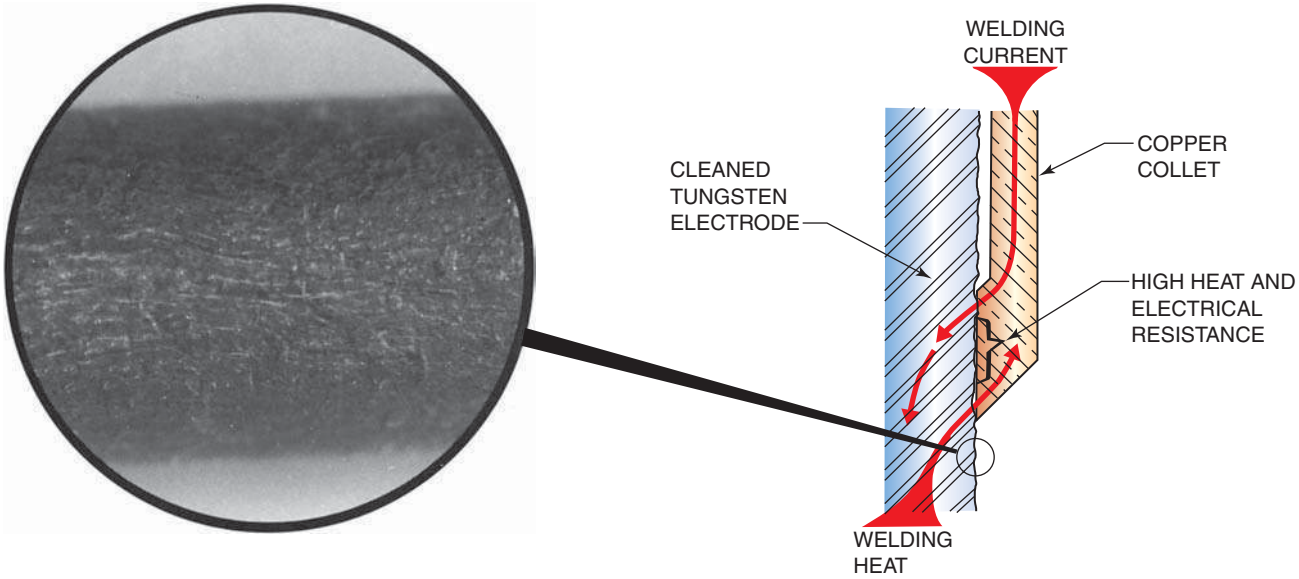

Figure 4.2
Irregular surface of a cleaned tungsten electrode (poor heat transfer to collet)
Photo Courtesy of Larry Jeffus

Collet-tungsten connection efficiency is shown in **Figure 4.2** and **Figure 4.3.**

Large-diameter electrodes conduct more current because the resistance heating effects are reduced. However, excessively large sizes may result in too low a temperature for a stable arc.

In general, the current-carrying capacity at DCEN is about ten times greater than that at DCEP.

The particular electrode tip shape impacts the temperature and erosion of the tungsten. With DCEN, a pointed tip concentrates the arc as much as possible and improves arc, starting with either a high-voltage electrical

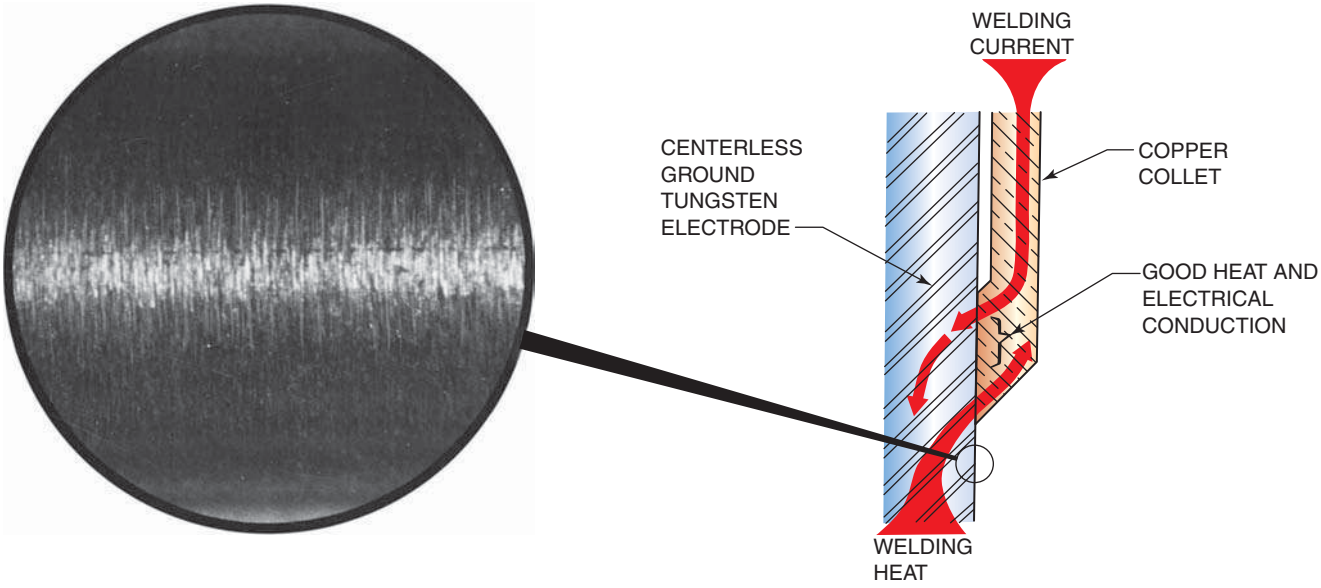

Figure 4.3
Smooth surface of a centerless ground tungsten electrode (good heat transfer to collet)
Photo Courtesy of Larry Jeffus

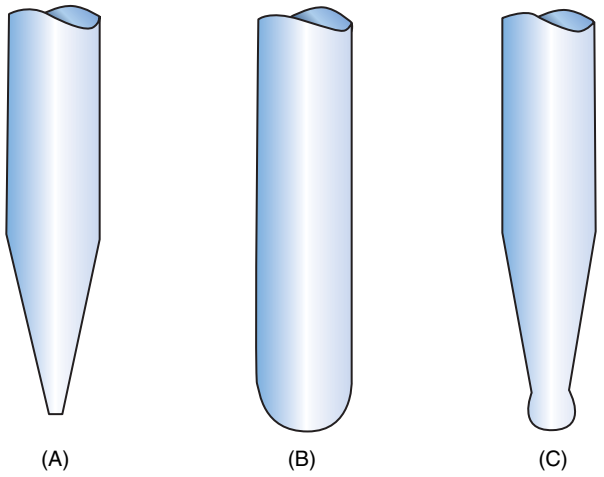

Figure 4.4
Basic tungsten electrode and shapes: pointed (A), rounded (B), and tapered with a balled end (C)

discharge or a touch start. Because DCEN does not put much heat on the tip, it is relatively cool, the point is stable, and it can survive extensive use without damage, **Figure 4.4A.**

With alternating current, the tip is subjected to more heat than with DCEN. To allow a larger mass at the tip to withstand the higher heat, the tip is rounded. The melted end must be small to ensure the best arc stability, **Figure 4.4B.**

DCEP has the highest heat concentration on the electrode tip. For this reason, a slight ball of molten tungsten is suspended at the end of a tapered electrode tip. The liquid tungsten surface tension, with the larger mass above the molten ball, holds it in place like a drop of water on your fingertip, **Figure 4.4C.**

TYPES OF TUNGSTEN ELECTRODES

Pure **tungsten** has a number of properties that make it an excellent nonconsumable electrode for the GTA welding process. These properties can be improved by adding cerium, lanthanum, thorium, or zirconium to the tungsten.

For GTA welding, tungsten electrodes are classified as the following:

- pure tungsten, EWP
- 1% thorium tungsten, EWTh-1
- 2% thorium tungsten, EWTh-2
- 1/4% to 1/2% zirconium tungsten, EWZr
- 2% cerium tungsten, EWCe-2
- 1% lanthanum tungsten, EWLa-1
- alloy not specified, EWG

See **Table 4.1.**

The type of finish on the tungsten must be specified as cleaned or ground. More information on composition and other requirements for tungsten welding electrodes is available in the AWS publication A5.12,

Table 4.1 Tungsten Electrode Types and Identification

AWS Classification	Tungsten Composition	Tip Color
EWP	Pure tungsten	Green
EWTh-1	1% thorium added	Yellow
EWTh-2	2% thorium added	Red
EWZr	1/4% to 1/2% zirconium added	Brown
EWCe-2	2% cerium added	Orange
EWLa-1	1% lanthanum added	Black
EWLa-1.5	1.5% lanthanum added	Gold
EWLa-2	2% lanthanum added	Blue
EWG	Alloy not specified	Grey

Specifications for Tungsten and Tungsten Alloy Electrodes for Arc Welding and Cutting.

Pure Tungsten Electrodes, EWP

Pure tungsten has the poorest heat resistance and electron emission characteristic of all the tungsten electrodes. It has a limited use with AC welding of metals, such as aluminum and magnesium, and is not recommended for use with inverter type welding power supplies. New alloying elements are now available with higher current carrying capacity; better emission characteristics and greater durability have replaced pure tungsten electrodes for AC welding in many shops.

Thoriated Tungsten Electrodes, EWTh-1 and EWTh-2

Thorium oxide (ThO_2), when added to tungsten in percentages of up to 0.6%, improves tungsten's current-carrying capacity. The addition of 1% to 2% of thorium oxide does not further improve current-carrying capacity. It does, however, help with electron emission. This can be observed by a reduction in the electron force (voltage) required to maintain an arc of a specific length. Thorium also increases the serviceable life of the tungsten. The improved electron emission of the thoriated tungsten allows it to carry approximately 20% more current. This also results in a corresponding reduction in electrode tip temperature, resulting in less tungsten erosion and subsequent weld contamination.

Thoriated tungsten also provides a much easier arc starting characteristic than pure or zirconiated electrodes. Thoriated tungsten works well with direct current electrode negative (DCEN). It can maintain a sharpened point well. It is very well suited for making welds on steel, steel alloys (including stainless), nickel alloys, and most other metals, other than aluminum or magnesium.

Thoriated electrodes do not work well with alternating current (AC). It is difficult to maintain a balled end, which used to be required and is still commonly used for AC welding. A thorium spike, **Figure 4.5,** may also develop on the end of a thorium electrode when it is overheated, disrupting a smooth arc. Furthermore, whether balled or sharpened to a point, when a thoriated tungsten electrode is overheated by exceeding its maximum amperage rating, small vaporized droplets of tungsten will transfer across the arc and into the work, possibly making the weld rejectable (see Figure 4.1).

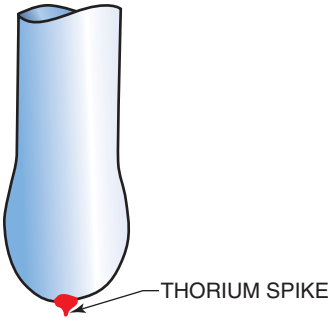

Figure 4.5
Thorium spike on a balled end tungsten electrode

CAUTION

Thorium is a very low-level radioactive oxide, but the level of radioactive contamination from a thorium electrode has not been found to be a health hazard during welding. It is, however, recommended that grinding dust be contained. Because of concern in other countries regarding radioactive contamination to the welder and welding environment, thoriated tungsten has been replaced with other alloys.

When a thorium spike becomes visible on an electrode tip, it should be assumed that the electrode was overheated and that the weld may be contaminated by vaporized droplets.

Zirconium Tungsten Electrodes, EWZr

Zirconium oxide (ZrO2) also helps tungsten emit electrons freely. The addition of zirconium to the tungsten has the same effect on the electrode characteristic as thorium, but to a lesser degree. Because zirconium tungsten is more easily melted than thorium tungsten, ZrO2 electrodes are recommended to be used with AC current. Because of the ease in forming the desired balled end on thorium versus zirconium tungsten, zirconium tungsten is normally the electrode chosen for AC welding of aluminum and magnesium alloys. Zirconiated tungsten is more resistant to weld pool contamination than pure tungsten, thus providing excellent weld qualities with minimal contamination.

Zirconiated tungsten also has the advantage over thoriated tungsten in that it is not radioactive.

Cerium Tungsten Electrodes, EWCe-2

Cerium oxide (CeO2) is added to tungsten to improve the current-carrying capacity in the same manner as does thorium. These electrodes were developed to replace thoriated tungsten electrodes because they are not made of a radioactive material. Cerium oxide electrodes have a current-carrying capacity similar to that of thoriated tungsten; however, they have an improved arc starting and arc stability characteristic, similar to that of thoriated tungsten. They can also provide a longer life than do most other electrodes, including thorium.

These electrodes have strong welder appeal for DC welding at low current settings; however, they are also used more and more with AC processes. In fact, ceriated electrodes along with lanthanum oxide electrodes are being recommended as a multipurpose electrode by the manufacturers of inverter based welding power supplies. With excellent arc starts at low amperages, ceriated tungsten has become more and more common in orbital tube and pipe manufacturing operations as well as with thin sheet and delicate work. Cerium electrodes contain approximately 2% cerium oxide.

Lanthanum Tungsten Electrodes, EWLa-1, EWLa-1.5, EWLa-2

Lanthanated tungsten is available as 1%, 1.5%, and 2%. Lanthanum trioxide has the lowest work function of any of the materials, thus it usually starts easiest and has the lowest temperature at the tip, which resists grain growth and promotes longer service life. Tests have proven that lanthanum electrodes last longer than do thorium electrodes when they are not overheated, and lanthanum electrodes are also resistant to thermal shock, which makes them a good choice for pulsed GTAW operations. Lanthanum electrodes require about 15% less power to initiate and maintain low current arcs. Lanthanum tungsten is a “rare earth” material and is not radioactive. Whereas lanthanated electrodes have been used successfully in Europe and Japan since 1993, lanthanum tungsten is a relative newcomer

to the United States. This electrode has been primarily used for DC welding, but it is now being recommended, along with cerium electrodes, for AC welding, given its high current-carrying capacity and resistance to spitting during AC operations. The higher current-carrying capacity and resistance to spitting and spikes allow lanthanum and cerium electrodes to be used with a pointed tip prep on AC operations with aluminum alloys. The pointed tip allows for a more controllable arc and the ability to make smaller weld profiles in fillet joints. While slightly more expensive than thoriated electrodes, lanthanum can be used for just about any shop operation, thus eliminating confusion about electrode selection when multiple alloys are welded.

Alloy Not Specified, EWG

The EWG classification is for electrodes whose alloys have been modified by manufacturers. Such alloys have been developed and tested by manufacturers to meet specific welding criteria. The blend of alloying oxides are proprietary, meaning that specific alloy compositions are not normally available from manufacturers; however, they do provide welding characteristics for these electrodes.

SHAPING THE TUNGSTEN ELECTRODE

The desired end shape of tungsten can be obtained by grinding, breaking, remelting the end, or using chemical compounds. Tungsten is brittle and easily broken. Welders must be sure to make a smooth, square break where they want it to be located.

Grinding

A grinder or belt sander is often used to clean a contaminated tungsten electrode or to put a point at the end of the tungsten. The grinder or sander used to sharpen tungsten should have a fine, hard stone or a fine grit media for the belt (80-120 grit). It should be used for grinding tungsten only. Because of the hardness of the tungsten and its brittleness, the grinding stone chips off small particles of the electrode. A coarse grinding stone will result in more tungsten breakage and a poorer finish. If the grinder is used for metals other than tungsten, particles of these metals may become trapped on the tungsten as it is ground. The metal particles will quickly break free when the arc is started, resulting in contamination.

EXPERIMENT 4-1

Module 1
Key Indicator 1, 4

Module 7
Key Indicator 2
Carbon Steel
Key Indicator 3
Austenitic Stainless Steel
Key Indicator 8
Aluminum
Key Indicator 13

Grinding the Tungsten to the Desired Shape

Using an electric grinder with a fine grinding stone or a belt sander with 80-120 grit abrasive media, one piece of tungsten 2 in. (51 mm) long or longer, and safety glasses, you will grind a point on tungsten electrodes.

Because of the hardness of the tungsten, it will become hot. Its high thermal conductivity means that the heat will be transmitted quickly to your fingers. To prevent overheating, only light pressure should be applied

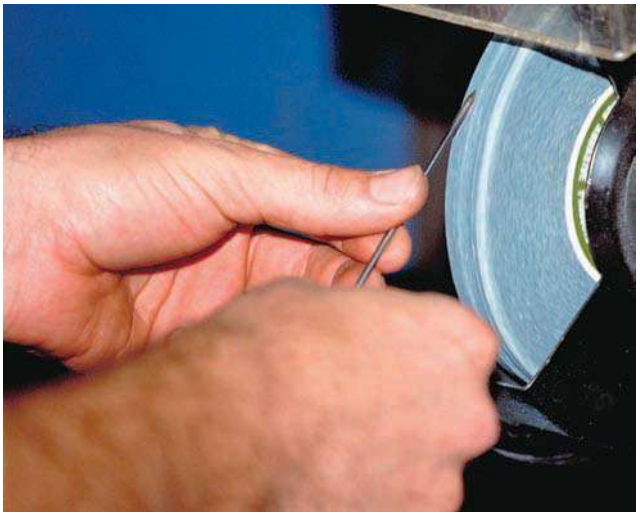

Figure 4.6
Correct method of grinding a tungsten electrode
Courtesy of Larry Jeffus

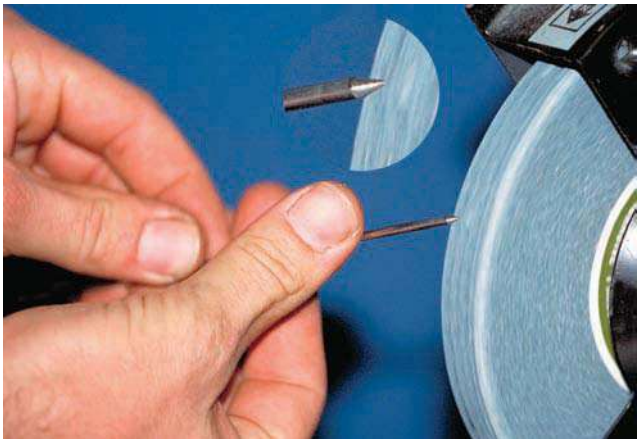

Figure 4.7
Incorrect method of grinding a tungsten electrode
Courtesy of Larry Jeffus

Note

If an electrode has a large aluminum deposit at the end from contact with the work or filler metal, the deposit should be broken off before grinding on a stone wheel because nonferrous materials like aluminum, brass, and copper have a tendency to smear onto the wheel, which may cause a dangerous out-of-balance condition in the wheel.

CAUTION

When holding one end of the tungsten against the grinding wheel, the other end of the tungsten must not be directed toward the palm of your hand, **Figure 4.8**. This will prevent the tungsten from being stuck into your hand if the grinding wheel catches it and suddenly pushes it downward.

against the grinding wheel or belt. This will also reduce the possibility of accidentally breaking the tungsten.

Grind the tungsten so that the grinding marks run lengthwise, **Figure 4.6** and **Figure 4.7.** Lengthwise grinding reduces the amount of small particles of tungsten contaminating the weld. Move the tungsten up and down as it is twisted during grinding. This will prevent the tungsten from becoming hollow-ground.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Breaking and Remelting

Tungsten is hard but brittle, resulting in a low impact strength. If tungsten is struck sharply, it will break without bending. When it is held against a sharp corner and hit, a fairly square break will result. **Figure 4.9, Figure 4.10,**

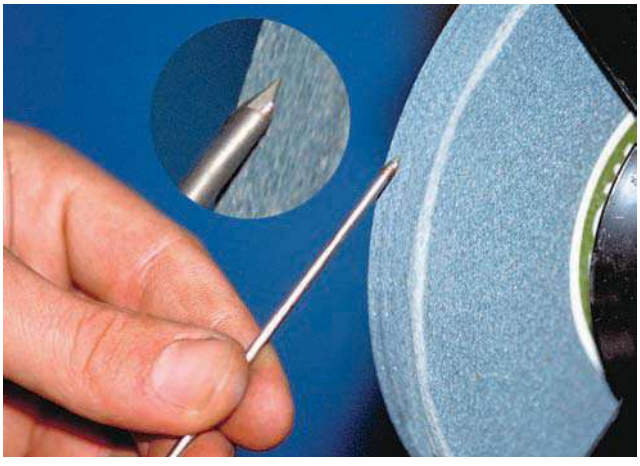

Figure 4.8
Correct way of holding a tungsten when grinding
Courtesy of Larry Jeffus

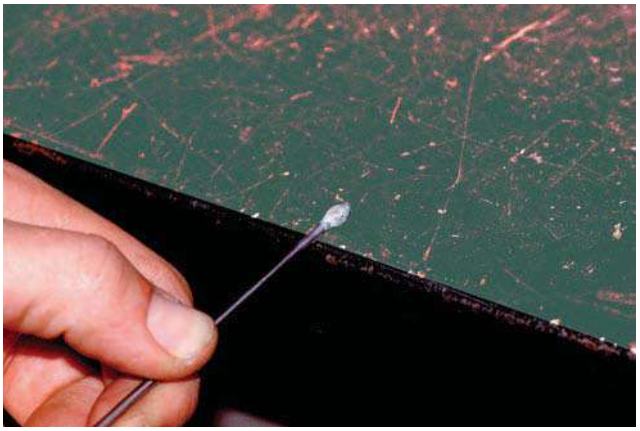

Figure 4.9
Breaking the contaminated end from a tungsten by striking it with a hammer
Courtesy of Larry Jeffus

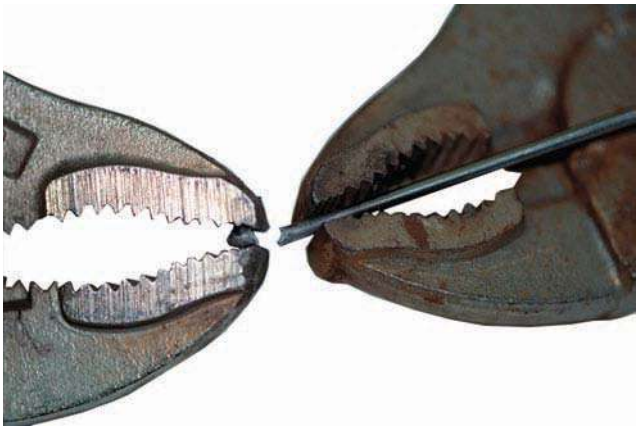

Figure 4.10
Correctly breaking the tungsten using two pairs of pliers
Courtesy of Larry Jeffus

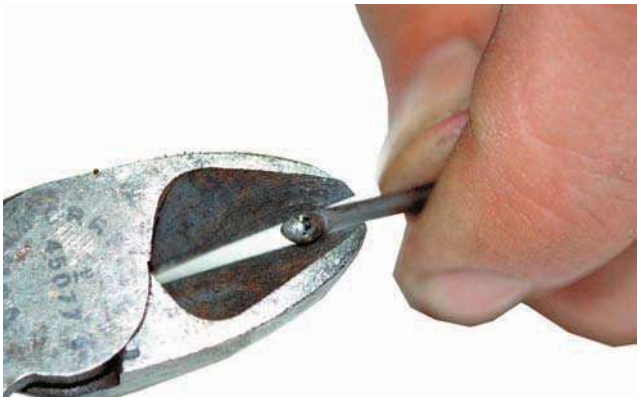

Figure 4.11
Using wire cutters to correctly break the tungsten
Courtesy of Larry Jeffus

and **Figure 4.11** show ways to break the tungsten correctly on a sharp corner using a hammer, with two pliers, and wire cutters.

Once the tungsten has been broken squarely, the end may be melted back so that it becomes somewhat rounded. The breaking and remelting technique is an older technique appropriate for pure tungsten and zirconiated tungsten electrodes due to their lower current-carrying capacity and inability to hold a pointed end. Cerium and lanthanum electrodes will maintain a pointed end much longer; therefore, this technique is rarely used with them unless the wider bead profile produced by a hemispherical tip prep is desired. This is accomplished by switching the welding current to DCEP and striking an arc under argon shielding on a piece of copper. If copper is not available, another piece of clean metal can be used. Do not use carbon, as it will contaminate the tungsten.

Note

Always wear eye protection as the end of the electrode can fly a good distance when struck.

EXPERIMENT 4-2

Removing a Contaminated Tungsten End by Breaking

Using short scrap pieces of tungsten, pliers or wire cutters, and a light machinist's hammer, you will break the end from the tungsten.

Module 1
Key Indicator 1, 4

Module 7
Key Indicator 2
Carbon Steel
Key Indicator 3
Austenitic Stainless Steel
Key Indicator 8
Aluminum
Key Indicator 13

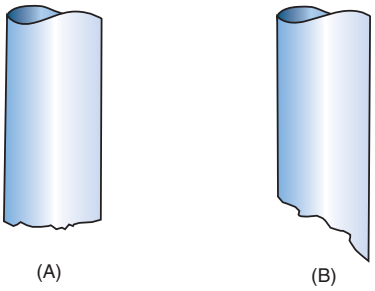

Figure 4.12
(A) Correctly broken tungsten electrode; (B) incorrectly broken tungsten electrode

Break about 1/4 in. (6 mm) from the end of the tungsten using the appropriate method, depending upon the diameter of the tungsten. Observe the break; it should be square and relatively smooth, **Figure 4.12.** The end of the tungsten should be broken only if the tungsten is badly contaminated.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Chemical Cleaning and Pointing

The tungsten can be cleaned and pointed using one of several compounds. The tungsten is heated by shorting it against the work. The tungsten is then dipped in the compound, a strong alkaline, which rapidly dissolves the hot tungsten. The chemical reaction occurs so quickly that enough additional heat is produced to keep the tungsten hot, **Figure 4.13.** When the tungsten is removed, cooled, and cleaned, the end will be tapered to a fine point. If the electrode is contaminated, the chemical compound will dissolve the tungsten, allowing the contamination to fall free.

Pointing and Remelting

The tapered tungsten with a balled end, a shape sometimes used for AC or DCEP welding, is made by first grinding or chemically pointing the electrode. Using DCEP, as in the procedure for the remelted broken end, strike an arc on some copper under argon shielding and slowly increase the current until a ball starts to form on the tungsten. The ball should be made large enough so that the color of the end stays between dull red and bright red. If the color turns white, the ball is too small and should be made larger. To increase the size of the ball, simply apply more current until the end begins to melt. Surface tension will pull the molten tungsten up onto the tapered end. Lower the current and continue welding. DCEP is seldom used for welding. If the tip is still too hot, it may be necessary to increase the size of the tungsten.

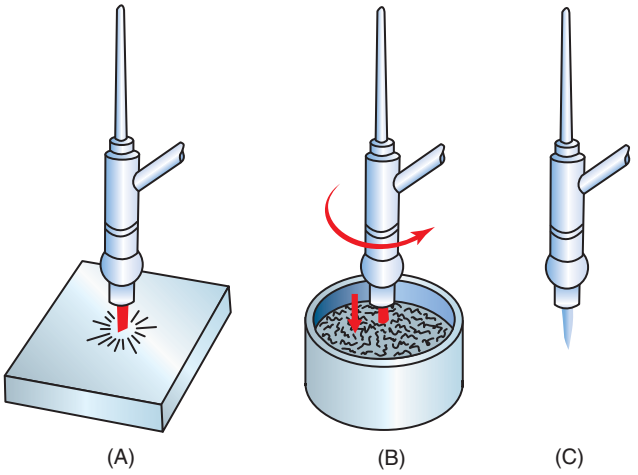

Figure 4.13
Chemically cleaning and pointing tungsten: (A) shorting the tungsten against the work to heat it to red hot, (B) inserting the tungsten into the compound and moving it around, and (C) cleaned and pointed tungsten ready for use

EXPERIMENT 4-3

Melting the Tungsten End Shape

Module 1
Key Indicator 1, 4

Module 7
Key Indicator 2
Carbon Steel
Key Indicator 3
Austenitic Stainless Steel
Key Indicator 8
Aluminum
Key Indicator 13

Using a properly set up GTA welding machine, proper safety protection, one piece of copper or other clean piece of metal, and the tungsten that was sharpened and broken in Experiments 4-1 and 4-2, you will melt the end of the tungsten into the desired shape.

Properly install the tungsten, set the argon gas flow, switch the current to DCEP, and turn on the machine. Strike an arc on the copper and slowly increase the amperage. Watch the tungsten as it begins to melt and stop the current when the desired shape has been obtained, **Figure 4.14.**

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

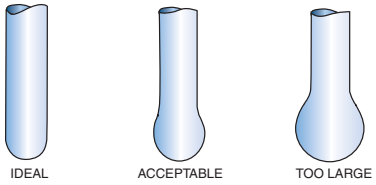

Figure 4.14
Melting the tungsten end shape

GTA WELDING EQUIPMENT

Figure 4.15 and **Figure 4.16** show two industrial applications of gas tungsten arc welding.

Torches

GTA welding torches are available water-cooled or air-cooled. The heat transfer efficiency for GTA welding may be as low as 20%. This means that 80% of the heat generated does not enter the weld. Much of this

Figure 4.15
Semiautomatic operation allows a stainless steel part to be GTA welded as it is turned past the torch
Courtesy of Lincoln Electric Company

Figure 4.16
An operator GTA welds a cap ring on a pneumatic tank
Courtesy of Lincoln Electric Company

Figure 4.17
Power cable safety fuse
Courtesy of ESAB Welding & Cutting Products

Note

With the introduction of multipurpose electrode types, such as cerium and lanthanum, it is often no longer necessary to produce a balled end on an electrode when welding aluminum with alternating current. However, when making surfacing welds or buildups on aluminum castings, a hemispherical end preparation on an electrode will produce a wider arc and weld pool, which is desirable in these types of operations.

heat stays in the torch and must be removed by some type of cooling method.

The water-cooled GTA welding torch is more efficient at removing waste heat than is an air-cooled torch. The water-cooled torch, as compared to the air-cooled torch, operates at a lower temperature, resulting in a lower tungsten temperature and less erosion.

The air-cooled torch is more portable because it has fewer hoses, and it may be easier to manipulate than the water-cooled torch. Also, the water-cooled torch requires a water reservoir or other system to give the needed cooling. The cooling water system should contain some type of safety device, **Figure 4.17,** to make it possible to shut off the power if the water flow is interrupted. The power cable is surrounded by the return water to keep it cool so that a smaller-size cable can be used. Without the cooling water, the cable quickly overheats and melts through the hose.

The water can become stopped or restricted for a number of reasons, such as a kink in the hose, a heavy object set on the hose, or failure to turn on the system. Water pressures higher than 35 psi (241 kg/mm^2) may cause the water hoses to burst. When an open system is used, a pressure regulator must be installed to prevent pressures that are too high from damaging the hoses.

GTA welding torch heads are available in a variety of amperage ranges and designs, **Figure 4.18.** The amperage listed on a torch is the maximum rating and cannot be exceeded without possible damage to the torch. The various head angles allow better access in tight places. Some of the heads can be swiveled easily to new angles. The back cap that both protects and tightens the tungsten can be short or long. Short caps are beneficial for getting the torch into restricted areas. Long caps can accommodate a full-length electrode, which will carry more heat away from the tip and allow slightly more current to be used before the electrode overheats. **Figure 4.19** and **Figure 4.20.**

Hoses

A water-cooled torch has three hoses connecting it to the welding machine. The hoses are for shielding gas to the torch, cooling water to the

Figure 4.18
GTA welding torches
Courtesy of American Torch Tip

torch and cooling water return, and housing the power cables to the torch, **Figure 4.21.** Air-cooled torches may have one hose for shielding gas attached to the power cable, **Figure 4.22.**

The shielding gas hose must be plastic to prevent the gas from being contaminated. Rubber hoses contain oils that can be picked up by the gas, resulting in weld contamination.

The water-in hose may be made of any sturdy material. Water hose fittings have left-hand threads, and gas hose fittings have right-hand threads.

Figure 4.19
Short back caps are available for torches when space is a problem

Figure 4.20
Long back caps allow tungstens that are a full 7 in. (177 mm) long to be used
Courtesy of ESAB Welding & Cutting Products

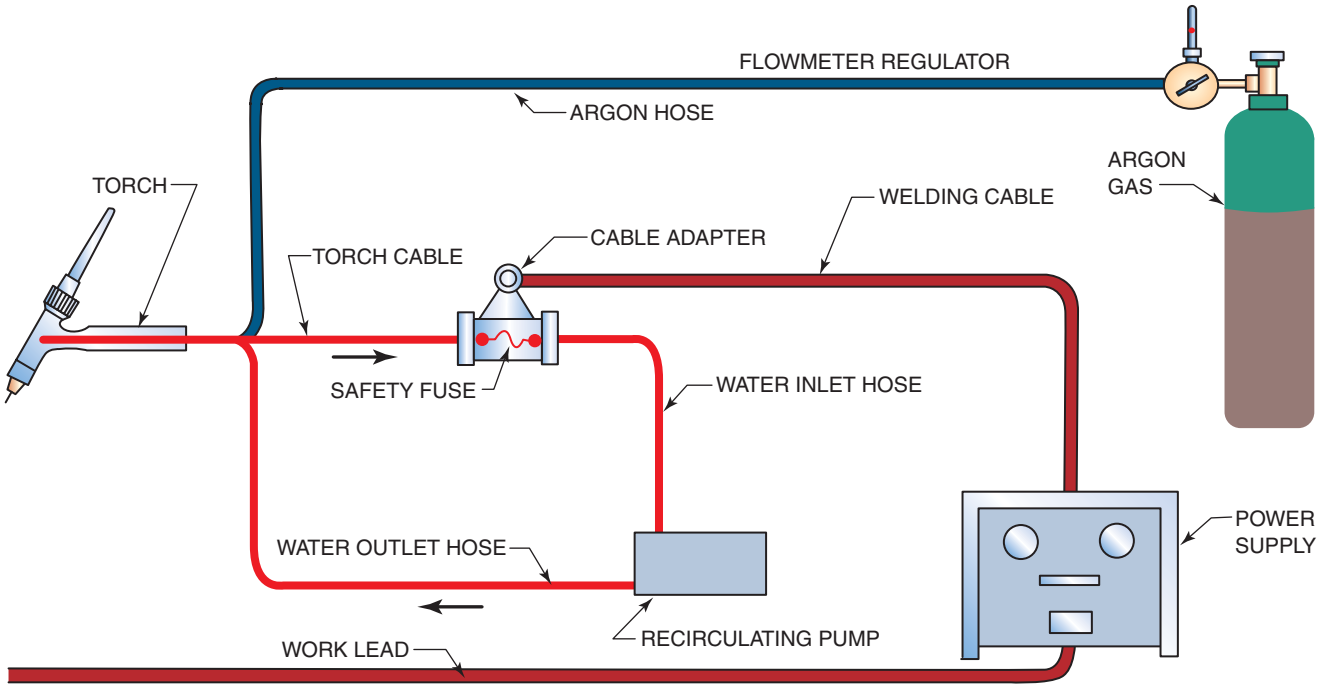

Figure 4.21
Schematic of a GTA welding setup with a water-cooled torch

This prevents the water and gas hoses from accidentally being reversed when they are attached to the welder. The return water hose also contains the welding power cable. This permits a much smaller-size cable to be used because the water keeps it cool.

The water must be supplied to the torch head and return around the cable. This allows the head to receive the maximum cooling from the water before the power cable warms it. Running the water through the torch first has another advantage. That is, when the water solenoid is closed, there is no water pressure in the hoses, which is particularly important. This fea-

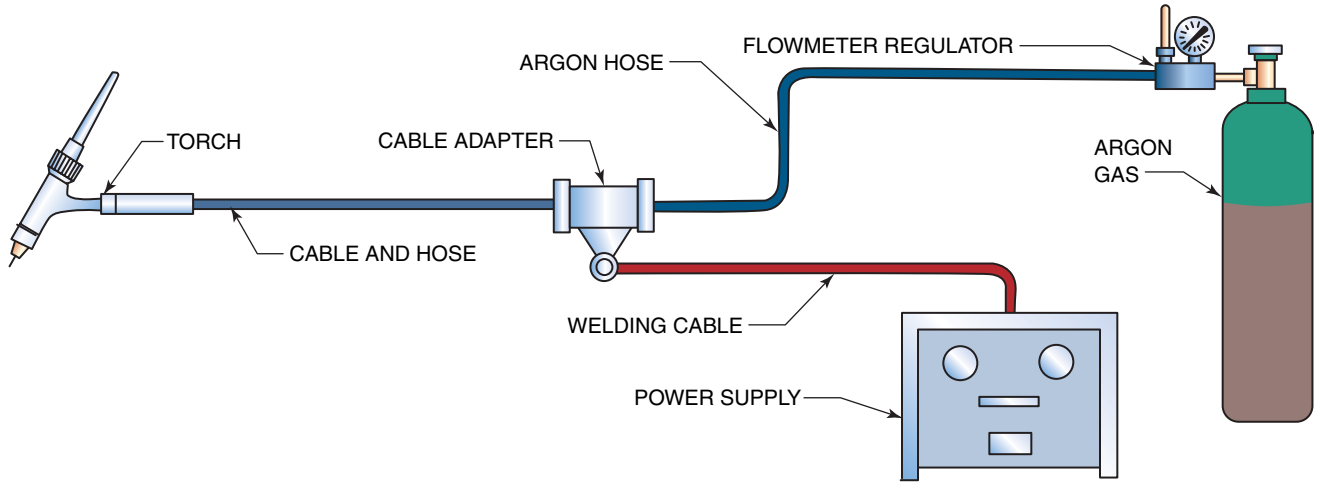

Figure 4.22
Schematic of a GTA welding setup with an air-cooled torch

Figure 4.23
Zip-on protective covering also helps keep the hoses neat
Courtesy of ESAB Welding & Cutting Products

Figure 4.24
A bracket holds the leads off the floor
Courtesy of Larry Jeffus

ture also prevents condensation in the torch. If a water leak should occur during welding, the welding power is stopped, closing the water solenoid and thus stopping the leak.

A protective covering can be used to prevent the hoses from becoming damaged by hot metal, **Figure 4.23.** Even with this protection, the hoses should be supported, **Figure 4.24,** so that they are not underfoot on the floor. By supporting the hoses, the chance of their being damaged by hot sparks is reduced.

Nozzles

The nozzle, or cup, is used to direct the shielding gas directly on the welding zone. The nozzle size is determined by the diameter of the opening and its length, **Table 4.2.** Nozzles may be made from ceramic, such as alumina or silicon nitride (opaque), or from fused quartz (clear). The nozzle may also have a gas lens to improve the gas flow pattern.

The nozzle size, both length and diameter, is often the welder's personal preference. Occasionally, a specific choice must be made based upon joint design or location. Small nozzle diameters allow the welder to better see the molten weld pool and can be operated with lower gas flow rates. Larger nozzle diameters can give better gas coverage, even in drafty places.

Ceramic nozzles are heat resistant and offer a relatively long life. The useful life of a ceramic nozzle is affected by the current level and prox-

Table 4.2 Recommended Cup Sizes

Tungsten Electrode Diameter		Nozzle Orifice Diameter	
in.	(mm)	in.	(mm)
1/16	(2)	1/4 to 3/8	(6 to 10)
3/32	(2.4)	3/8 to 7/16	(10 to 11)
1/8	(3)	7/16 to 1/2	(11 to 13)
3/16	(5)	1/2 to 3/4	(13 to 20)

imity to the work. Silicon nitride nozzles will withstand much more heat, resulting in a longer useful life.

The fused quartz (glass) used in a nozzle is a special type that can withstand the welding heat. These nozzles are no more easily broken than ceramic ones but are more expensive. The added visibility with glass nozzles in tight, hard-to-reach places is often worth the added expense.

The longer a nozzle, the longer the tungsten must be extended from the collet. This can cause higher tungsten temperatures, resulting in greater tungsten erosion. When using long nozzles, it is better to use low amperages or a larger-size tungsten.

Flowmeter

The **flowmeter** may be merely a flow regulator used on a manifold system or it may be a combination flow and pressure regulator used on an individual cylinder, **Figure 4.25** and **Figure 4.26.**

The flow is metered or controlled by opening a small valve at the base of the flowmeter. The rate of flow is then read in units of cfh (cubic feet per hour), or L/min (liters per minute). The reading is taken from a fixed scale that is compared to a small ball floating on the stream of gas. Meters from various manufacturers may be read differently. For example, they may read from the top, center, or bottom of the ball, **Figure 4.27.** The ball floats on top of the stream of gas inside a tube that gradually increases in diameter in the upward direction. The increased size allows

Figure 4.25
Flowmeter
Courtesy of Controls Corporation of America

Figure 4.26
Flowmeter regulator
Courtesy of Controls Corporation of America

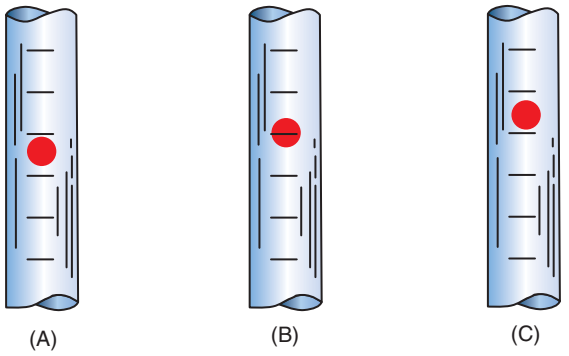

Figure 4.27
Three methods of reading a flowmeter (A) top of ball, (B) center of ball, and (C) bottom of ball

more room for the gas flow to pass by the ball. If the tube is not vertical, the reading is not accurate, but the flow is unchanged. Also, when using a line flowmeter, it is important to have the correct pressure. Changes in pressure will affect the accuracy of the flowmeter reading. In order to get accurate readings, be sure the gas being used is read on the proper flow scale. Less dense gases, such as helium and hydrogen, will not support the ball on as high a column with the same flow rate as a denser gas, such as argon.

TYPES OF WELDING CURRENT

All three types of welding current, or polarities, can be used for GTA welding. Each current has individual features that make it more desirable for specific conditions or with certain types of metals.

The major differences among the currents are in their heat distributions and the presence or degree of arc cleaning. **Figure 4.28** shows the heat distribution for each of the three types of currents.

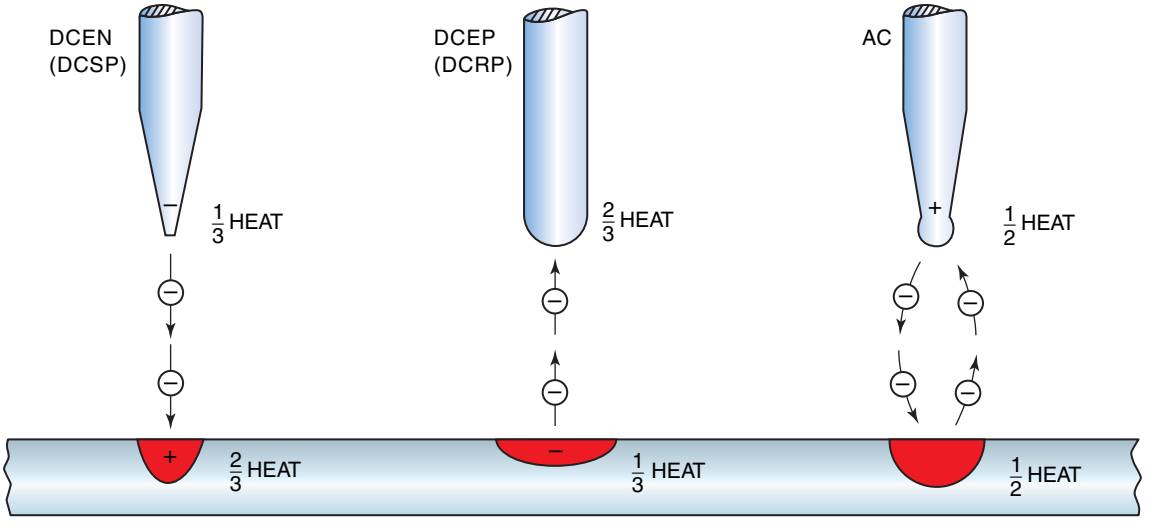

Figure 4.28
Heat distribution between the tungsten electrode and the work with each type of welding current

Direct current electrode negative (DCEN), which used to be called direct current straight polarity (DCSP), concentrates about two-thirds of its welding heat on the work and the remaining one-third on the tungsten. The higher heat input to the weld results in deep penetration. The low heat input into the tungsten means that a smaller-size tungsten can be used without erosion problems.

Direct current electrode positive (DCEP), which used to be called direct current reverse polarity (DCRP), concentrates only one-third of the arc heat on the plate and two-thirds of the heat on the electrode. This type of current produces wide welds with shallow penetration, but it has a strong cleaning action upon the base metal. The high heat input to the tungsten indicates that a large-size tungsten is required, and the end shape with a ball must be used. The low heat input to the metal and the strong cleaning action on the metal make this a good current choice for thin, heavily oxidized metals and magnesium. The metal being welded will not emit electrons as freely as does tungsten, so the arc may wander or be more erratic than DCEN. Because of the near molten state of the electrode tip during DCEP operations, it is almost always done in the flat position.

There are many theories as to why DCEP has a **cleaning action.** The most probable explanation is that the electrons accelerated from the cathode surface lift the oxides that interfere with their movement. The positive ions accelerated to the metal's surface provide additional energy. In combination, the electrons and ions cause the surface erosion needed to produce the cleaning. Although this theory is disputed, it is important to note that cleaning does occur, that it requires argon-rich shield gases and DCEP polarity, and that it can be used to advantage, **Figure 4.29.**

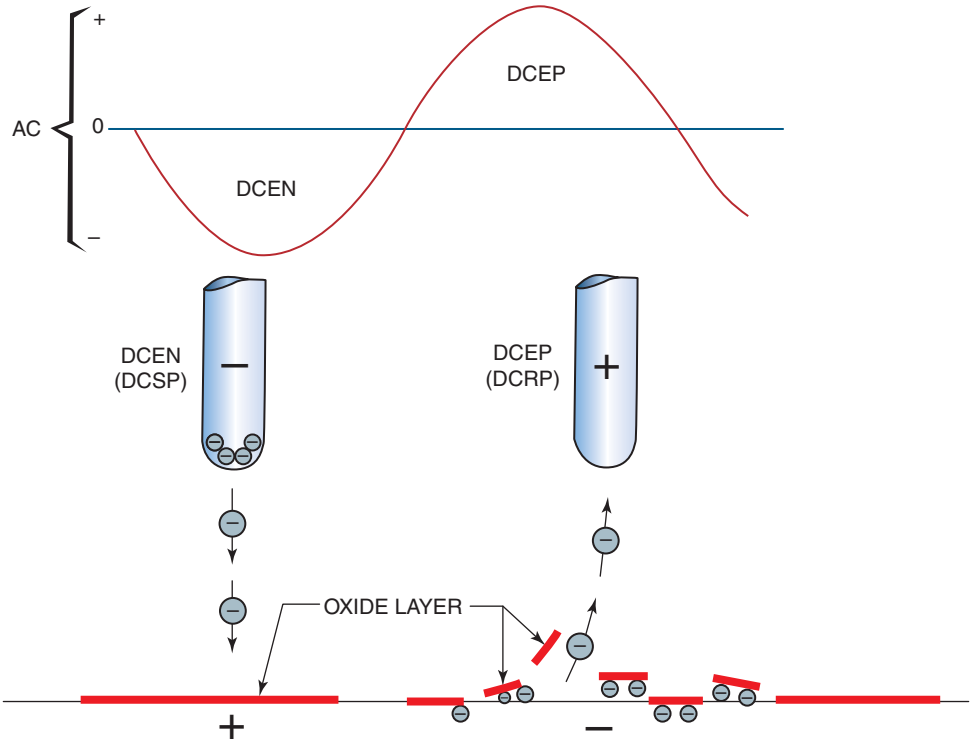

Figure 4.29
Electrons collect under the oxide layer during the DCEP portion of the cycle and lift the oxides from the surface

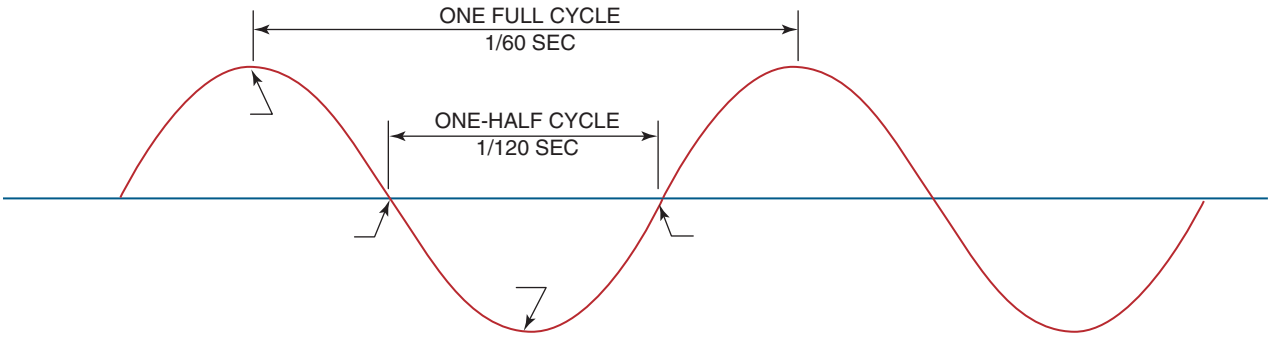

Figure 4.30
Sine wave of alternating current at 60 cycle

Alternating current (AC) concentrates about half of its heat on the work and the other half on the tungsten. Alternating current is DCEN half of the time and DCEP the other half of the time. The **frequency** at which the current cycles is the rate at which it makes a full change in direction, **Figure 4.30.** In the United States, the electric current cycles at the rate of 60 times per second, or 60 hertz (60 Hz). Referring again to Figure 4.30, the current is at its maximum peak at points A and B. The rate gradually decreases until it stops at points C and D. The arc at these points is extinguished and, as the current reversal begins, must be reestablished. This event requires the emission of electrons from the cathode to ionize the shielding gas. When the hot, emissive electrode becomes the cathode, reestablishing the arc is easy. However, it is often quite difficult to reestablish the arc when the colder and less emissive workpiece becomes the cathode. Because voltage from the power supply is designed to support a relatively low voltage arc, it may be insufficient to initiate electron flow. When the arc does not reignite consistently, it becomes destabilized and can cause poor welding performance. This phenomenon is called **rectification.** Thus, a voltage assist from another source is needed. A high-voltage but low-current **spark gap oscillator** commonly provides the assist at a relatively low cost. The high frequency ensures that a voltage peak will occur reasonably close to the current reversal in the welding arc, creating a low-resistance ionized path for the welding current to follow, **Figure 4.31A** and **Figure 4.31B.** This same device is often used to initiate direct current arcs, a particularly useful technique for mechanized welding.

The high-frequency current is established by capacitors discharging across a gap set on points inside the machine. Changing the point gap setting will change the frequency of the current. The closer the points are, the higher the frequency; the wider the spacing between the points, the lower the frequency. The voltage is stepped up with a transformer from the primary voltage supplied to the machine. The available amperage to the high-frequency circuit is very low. Thus, when the circuit is complete, the voltage quickly drops to a safe level. The high frequency is induced on the primary welding current in a coil.

The high frequency may be set so that it automatically cuts off after the arc is established, when welding with DC. It is kept on continuously with AC and transformer rectifier power supplies. When used in this manner, it is referred to as alternating current, high-frequency stabilized, or ACHF. Most of the newer inverter power supplies have advanced circuitry that

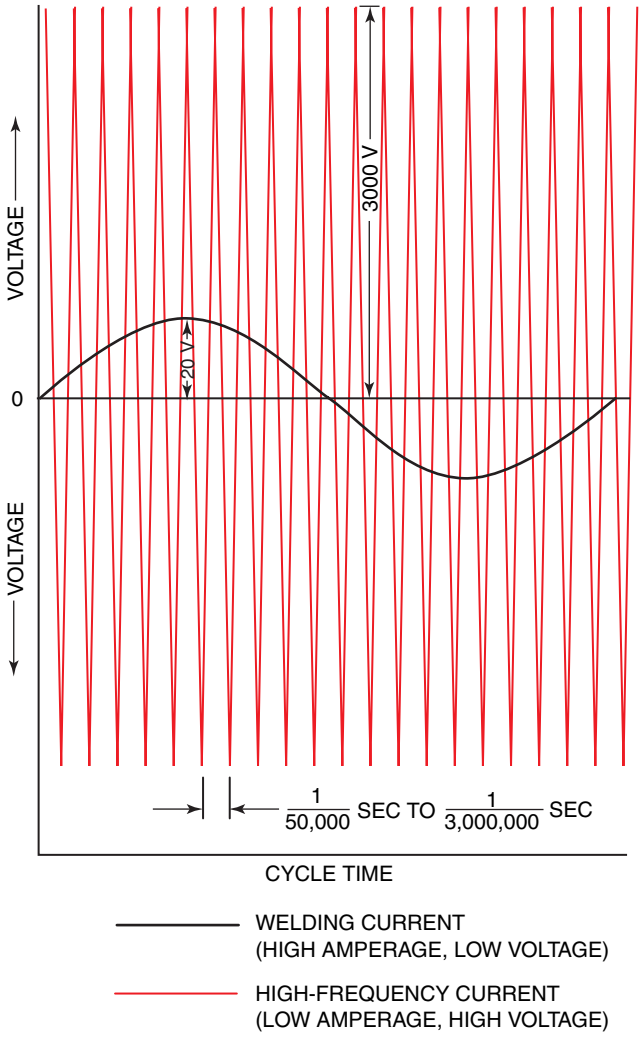

Figure 4.31A
High-frequency arc starting current shown over the low-frequency welding current

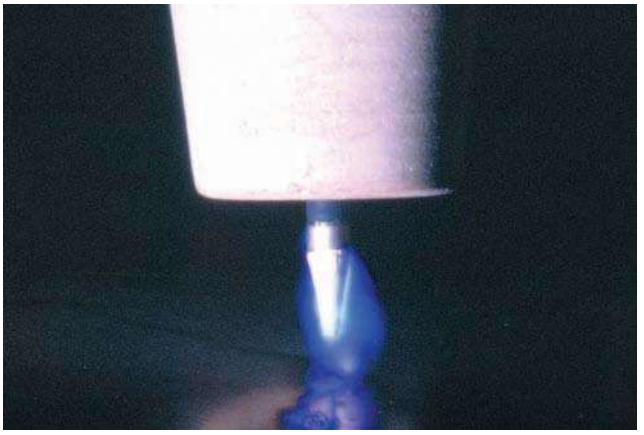

Figure 4.31B
The high frequency first appears as a blue glow around the tungsten before the welding current starts its arc
Courtesy of Larry Jeffus

switches between DCEN and DCEP so quickly that rectification cannot occur and high frequency is not used.

AC BALANCE CONTROL

While older transformer rectifier welding power supplies can only produce AC power in the form of a sine wave (Figure 4.30), newer transformer rectifiers and inverter power supplies can provide enhanced AC output by controlling the dwell time spent on each side of the AC cycle.

The two additional AC waveforms are called square wave and advanced square wave. **Figure 4.32** shows a comparison of the three waveforms. The waveforms represent how the current is measured in amplitude and time as the current moves to a high point, changes direction and polarity, and crosses over the zero point while moving to the maximum amperage on the other side. When the current has moved up and down, hitting both maximum amperages, this represents one cycle. AC electrical power in the United States is delivered at 60 cycles per second.

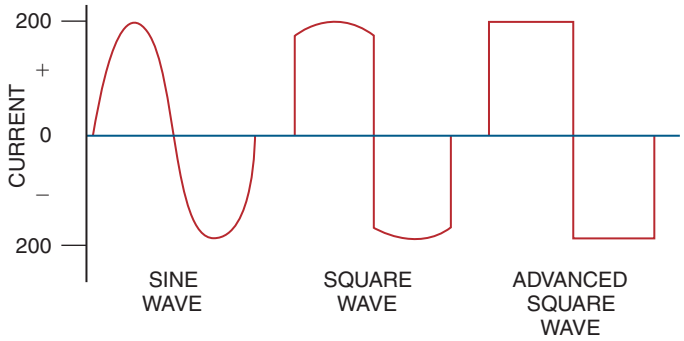

Figure 4.32
Comparison of the three different AC waveforms all representing a time balanced condition and operating at 200 amperes

Welding machines with square wave capabilities have a knob or digital readout that controls the amount of time the current spends on the EN and EP half of the AC cycle. If the machine is adjusted to have more dwell time spent on the EN side, the result will be deeper penetration and reduced cleaning action. The dial usually says "max penetration," **Figure 4.33.** An additional benefit of an AC arc balanced to favor EN is that the extra current delivered to the work will reduce the heat at the tip of the tungsten, which may allow a smaller diameter electrode, or even a pointed end, to be used.

If the machine is adjusted to have more dwell time spent on the EP side of the half cycle, the result will be a wider, shallower bead profile with greater cleaning action. A machine setting favoring more dwell time on EP (called "max cleaning") will be a benefit when it is impractical or impossible to completely remove small amounts of surface contaminants or when welding on aluminum castings. The extra EP will help to etch away more or thicker oxides. When balance controls are set toward "max cleaning," a balled end will be required on the electrode, and if the ball becomes larger than 1-1/2 times the electrode diameter, the next larger size should be selected (Figure 4.14).

Note

Increased EP balance is not a substitute for excellent surface prep, **Figure 4.34.**

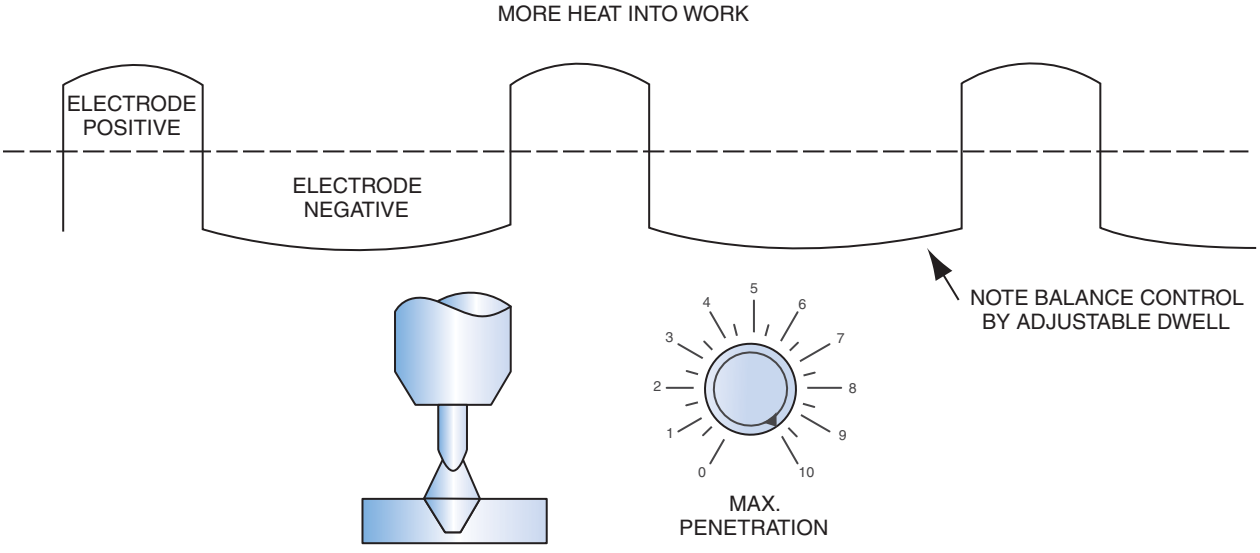

Figure 4.33
Maximum penetration balance control setting

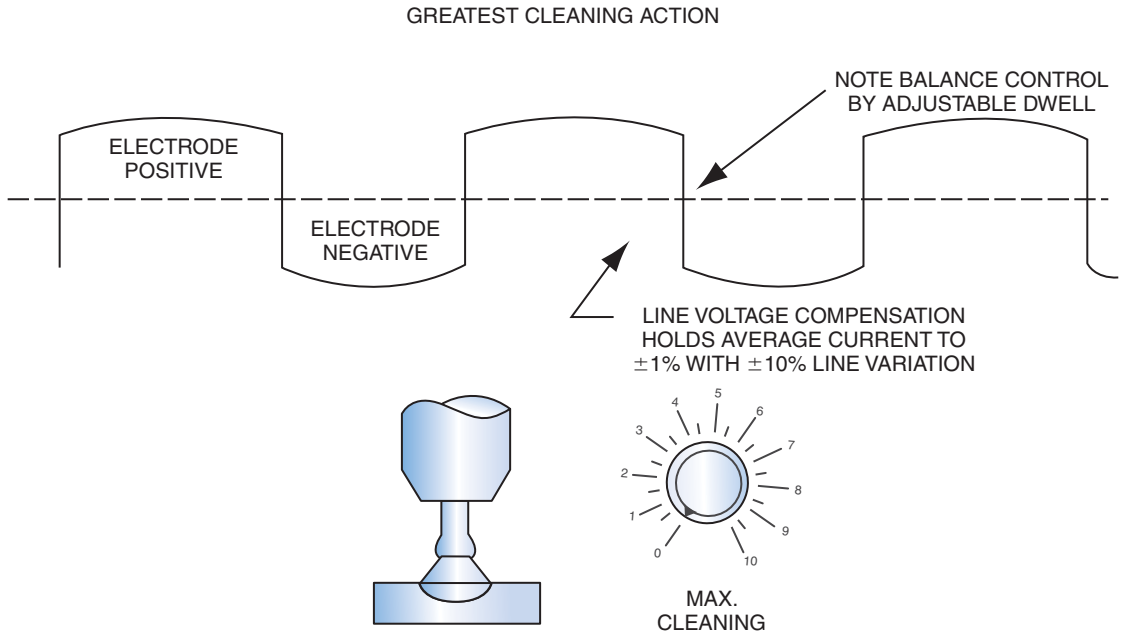

Figure 4.34
Maximum cleaning control setting

When the control knob is set in-between max cleaning and max penetration, it is called a balanced waveform. This is a good place to start for general AC applications, **Figure 4.35.** Balance control can be increased toward max EN if greater penetration or a pointed electrode is required; however, operators must be careful not to turn the balance control too far toward EN. When this happens, the oxides on top of the aluminum will not be completely removed and poor fusion or porosity may result. It is also important to note that transformer rectifiers, when set at max penetration, have 68% dwell time on EN, while advanced square wave inverter power supplies may produce as much as 99% dwell time on EN. In essence, this

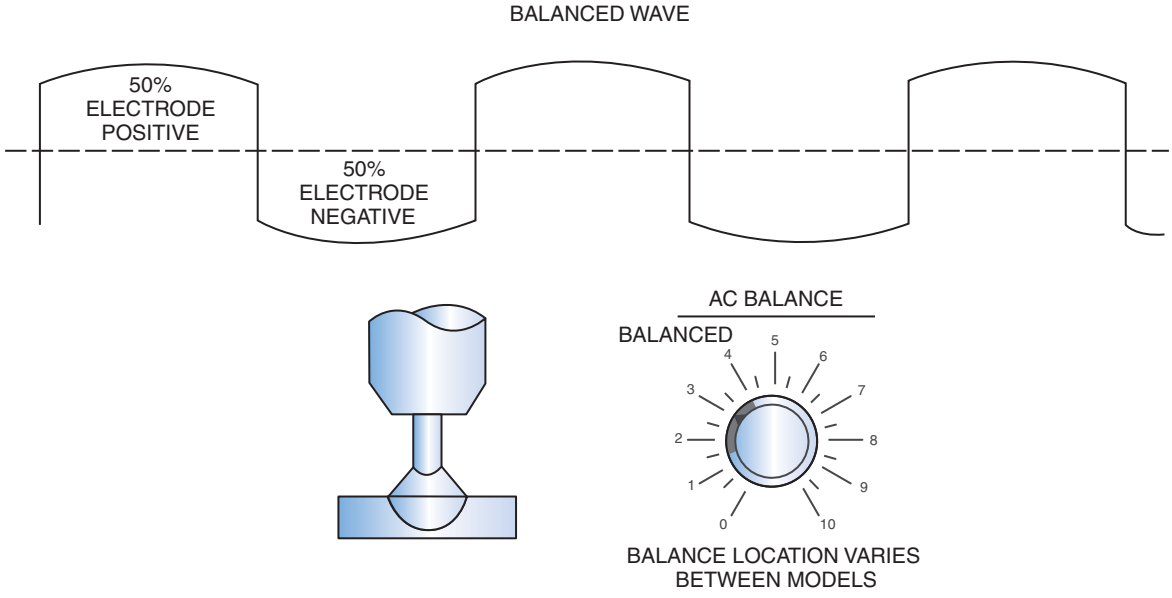

Figure 4.35
Balanced control setting

means that when a transformer square wave is set at 10 on its balance control dial, it delivers about the same amount of EN as when an advanced square wave inverter is set to 7.

If you have a square wave power supply, try adjusting the balance control toward max cleaning and max penetration when working on aluminum practices, and note the results in etching at the weld toes and the overall bead profile.

CAUTION

Never allow noninert gases, such as O_2, CO_2, or N, to come in contact with your inert gas system. Very small amounts can contaminate the inert gas, which may result in the weld failing.

SHIELDING GASES

The shielding gases used for the GTA welding process are argon (Ar), helium (He), hydrogen (H), nitrogen (N), or a mixture of two or more of these gases. The purpose of the shielding gas is to protect the molten weld pool and the tungsten electrode from the harmful effects of air. The shielding gas also affects the amount of heat produced by the arc and the resulting weld bead appearance.

Argon and helium are **inert gases.** This means that they will not combine chemically with any other material. Argon and helium may be found in mixtures but never as compounds. Because they are inert, they will not affect the molten weld pool in any way.

Argon

Argon is a by-product in air separation plants. Air is cooled to temperatures that cause it to liquefy; then its constituents are fractionally distilled. The primary products are oxygen and nitrogen. Before these gases were produced on a tonnage scale, argon was a rare gas. Now it is distributed in cylinders as gas or in bulk as a liquid.

Because argon is denser than air, it effectively shields welds in deep grooves in the flat position. However, this higher density can be a hindrance when welding overhead because higher flow rates are necessary. The argon is relatively easy to ionize and thus suitable for alternating current applications and easier starts. This property also permits fairly long arcs at lower voltages, making it virtually insensitive to changes in arc length. Argon is also the only commercial gas that produces the cleaning discussed earlier. These characteristics are most useful for manual welding, especially with filler metals added, as shown in **Figure 4.36.**

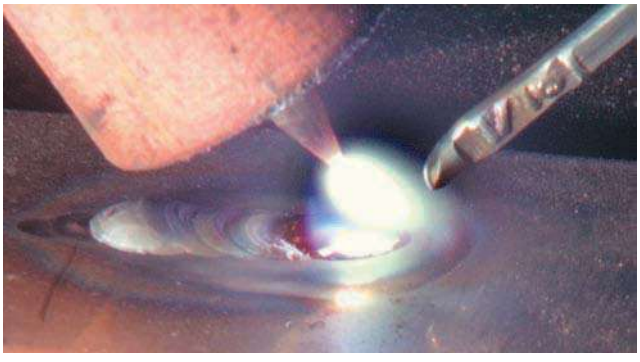

Figure 4.36
Highly concentrated ionized argon gas column
Courtesy of Larry Jeffus

Helium

Helium is a by-product of the natural gas industry. It is removed from natural gas as the gas undergoes separation (fractionation) for purification or refinement.

Helium offers the advantage of deeper penetration. The arc force with helium is sufficient to displace the molten weld pool with very short arcs. In some mechanized applications, the tip of the tungsten electrode is positioned below the workpiece surface to obtain very deep and narrow penetration. This technique is especially effective for welding aged aluminum alloys prone to overaging. It is also very effective at high welding speeds, as for tube mills. However, helium is less forgiving for manual welding. With helium, penetration and bead profile are sensitive to arc length, and the long arcs that are needed for feeding filler wires are more difficult to control.

Helium has been mixed with argon to gain the combined benefits of cathode cleaning and deeper penetration, particularly for manual welding. The most common of these mixtures is 75% helium and 25% argon.

Although the GTA process was developed with helium as the shielding gas, argon is now used whenever possible because it is much cheaper. Helium also has some disadvantages. Because it is lighter than air, helium doesn't allow for good shielding. Its flow rates must be about twice as high as argon's for acceptable stiffness in the gas stream, and proper protection is difficult in drafts unless high flow rates are used. It is difficult to ionize, necessitating higher voltages to support the arc and making the arc more difficult to ignite. Alternating current arcs are very unstable. However, helium is not used with alternating current because the cleaning action does not occur.

Hydrogen

Hydrogen is not an inert gas and is not used as a primary shielding gas. However, it can be added to argon in small amounts (1–3%) when deep penetration and high welding speeds are needed. It also improves the weld surface cleanliness and bead profile on some grades of stainless steel that are very sensitive to oxygen. Hydrogen additions are restricted to stainless steel because hydrogen is the primary cause of porosity in aluminum welds. It can cause porosity in carbon steel and, in highly restrained welds, underbead cracking in carbon and low alloy steel.

Nitrogen

Nitrogen is not an inert gas. Like hydrogen, nitrogen has been used as an additive to argon. But it cannot be used with some materials, such as ferritic steel, because it produces porosity. In other cases, such as with austenitic stainless steel, nitrogen is useful as an austenite stabilizer in the alloy. It is used to increase penetration when welding copper. Nitrogen is also sometimes used as a backing gas on some austenitic stainless steel pipe and tubing because of its low relative cost. When using nitrogen to protect the back side of a weld, care must be taken to be sure that all fitups are tight and that no nitrogen mixes with the shield gas at the top of the weld. Unfortunately, because of the general success with inert gas mix-

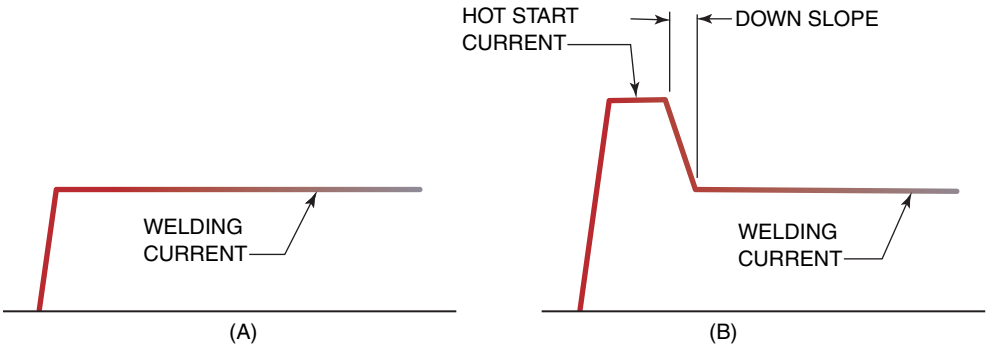

Figure 4.37
Standard method of starting welding current (A); hot start method of starting welding current (B)

tures and because of potential metallurgical problems, nitrogen has not received much attention as an additive for GTA welding.

Hot Start

The hot start allows a controlled surge of welding current as the arc is started to quickly establish a molten weld pool. Rapidly establishing a molten weld pool on metals with a high thermal conductivity is often hard without this higher than normal current. Adjustments can be made in the length of time and the percentage above the normal current, **Figure 4.37.**

Preflow Time and Postflow Time

Preflow time is the time during which gas flows to clear out any air in the nozzle or surrounding the weld zone. The operator sets the length of time that the gas flows before the welding current is started, **Figure 4.38.** Because some machines do not have preflow, many welders find it hard to hold a position while waiting for the current to start. One solution to this problem is to use the postflow for preflow. Switch on the current to engage the postflow. Now, with the current off, the gas is flowing, and the GTA torch can be lowered to the welding position. The welder's helmet should be lowered and the current restarted before the postflow stops. This allows welders to have preflow of shield gas and to start the arc when they are ready.

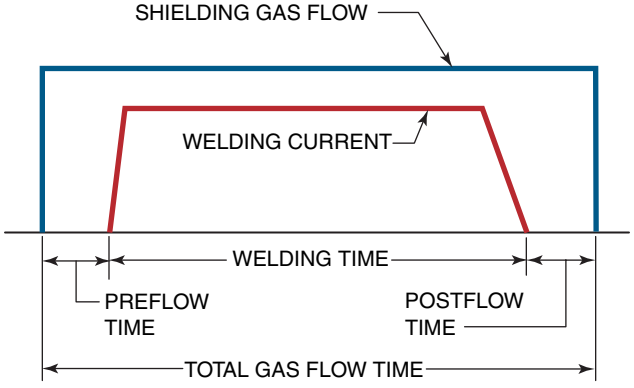

Figure 4.38
Welding time compared to shielding gas flow time

Table 4.3 Postwelding Gas Flow Times

Electrode Diameter		Postwelding Gas Flow Time*
in.	(mm)	
.01	(0.25)	5 sec
0.02	(0.5)	5 sec
0.04	(1)	5 sec
1/16	(2)	8 sec
3/32	(2.4)	10 sec
1/8	(3)	15 sec
5/32	(4)	20 sec
3/16	(5)	25 sec
1/4	(6)	30 sec

*The time may be longer if either the base metal or the tungsten electrode does not cool below the rapid oxidation temperatures within the postflow times shown.

The **postflow time** is the time during which the gas continues flowing after the welding current has stopped. This period serves to protect the molten weld pool, the filler rod, and the tungsten electrode as they cool to a temperature at which they will not oxidize rapidly. The time of the flow is determined by the welding current and the tungsten size, **Table 4.3.**

Gas Flow Rate

The shielding gas flow rate is measured in cubic feet per hour (cfh) or in metric measure as liters per minute (L/min). The rate of flow should be as low as possible and still give adequate coverage. High gas flow rates waste shielding gases and may lead to contamination. The contamination comes from turbulence in the gas at high flow rates. Air is drawn into the gas envelope by a venturi effect around the edge of the nozzle. Also, the air can be drawn in under the nozzle if the torch is held at too sharp an angle to the metal, **Figure 4.39.**

The larger the nozzle size, the higher is the flow rate permissible without causing turbulence. **Table 4.4** shows the average and maximum flow rates for most nozzle sizes. A gas lens can be used in combination with the nozzle to stabilize the gas flow, thus eliminating some turbulence. A gas lens will add to the turbulence problem if there is any spatter or contamination on its surface.

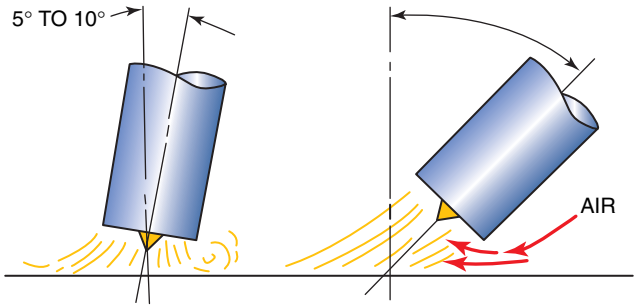

Figure 4.39
Too steep an angle between the torch and work may draw in air

Table 4.4 Suggested Argon Gas Flow Rate for Given Cup Sizes

Nozzle Inside Diameter		Gas Flow*	
in.	(mm)	cfh	(L/min)
1/4	(6)	10–14	(4.7–6.6)
5/16	(8)	11–15	(5.2–7.0)
3/8	(10)	12–16	(5.6–7.5)
7/16	(11)	13–17	(6.1–8.0)
1/2	(13)	17–20	(8.0–9.4)
5/8	(16)	17–20	(8.0–9.4)

*The flow rates may need to be increased or decreased depending upon the conditions under which the weld is to be performed.

REMOTE CONTROLS

A remote control can be used to start the weld, increase or decrease the current, and stop the weld. The remote control can be either a foot-operated or hand-operated device. The foot control works adequately if the welder can be seated. Welds that must be performed away from a welding station may use a hand or thumb control, or may not have any remote welding controls.

Most remote controls have an on-off switch that is activated at the first or last part of the control movement. A variable resistor increases the current as the control is pressed more. A variable resistor works in a manner similar to the accelerator pedal on a car to increase the power (current), **Figure 4.40.** The operating amperage range is determined by the value that has been set on the main controls of the machine.

Figure 4.40
A foot-operated device can be used to increase the current
Courtesy of Larry Jeffus

EXPERIMENT 4-4

Setting Up a GTA Welder

Using a GTA welding machine; remote control welding torch; gas flowmeter; gas source (cylinder or manifold); tungsten; nozzle; collet; collet body; cap; and any other hoses, special tools, and required equipment, you will set up the machine for GTA welding, **Figure 4.41.**

Module 1
Key Indicator 1, 4

Module 7
Carbon Steel
Key Indicator 3
Austenitic Stainless Steel
Key Indicator 8
Aluminum
Key Indicator 13

Figure 4.41
GTA welding unit that can be added to a standard power supply so that it can be used for GTA welding
Courtesy of Lincoln Electric Company

Figure 4.42
Always be sure the power is off when making machine connections
Courtesy of Larry Jeffus

Figure 4.43
Tighten each fitting as it is connected to avoid missing a connection
Courtesy of Larry Jeffus

1. Start with the power switch off, **Figure 4.42.** Use a wrench to attach the torch hose to the machine. The water hoses should have left-hand threads to prevent incorrectly connecting them. Tighten the fittings only as tightly as needed to prevent leaks, **Figure 4.43.** Attach the cooling water "in" to the machine solenoid and the water "out" to the power block.
2. The flowmeter or flowmeter regulator should be attached next. If a gas cylinder is used, secure it in place with a safety chain. Then remove the valve protection cap and crack the valve to blow out any dirt, **Figure 4.44.** Attach the flowmeter so that the tube is vertical.
3. Connect the gas hose from the meter to the gas "in" connection on the machine.
4. With both the machine and main power switched off, turn on the water and gas so that the connection to the machine can be checked for leaks. Tighten any leaking fittings to stop the leakages.

Figure 4.44
During transportation or storage, dirt may collect in the valve. Cracking the valve is the best way to remove any dirt.

Figure 4.45
Setting the current
Courtesy of Larry Jeffus

Figure 4.46
Setting the amperage range
Courtesy of Larry Jeffus

Figure 4.47
The high-frequency switch should be placed in the appropriate position
Courtesy of Larry Jeffus

5. Turn on both the machine and main power switches and watch for leaks in the torch hoses and fittings.
6. With the power off, switch the machine to the GTA welding mode.
7. Select the desired type of current and amperage range, **Figure 4.45** and **Figure 4.46.**
8. Set the fine current adjustment to the proper range, depending upon the size of tungsten used, **Table 4.5.**
9. Place the high-frequency switch in the appropriate position, auto (HF start) for DC or continuous for AC, **Figure 4.47.**
10. The remote control can be plugged in and the selector switch set, **Figure 4.48.**
11. The collet and collet body should be installed on the torch first, **Figure 4.49.**
12. On the Linde brand or copies of Linde torches, installing the back cap first will stop the collet body from being screwed into the torch fully. A poor connection will result in excessive electrical and thermal resistance, causing a heat buildup in the head.
13. The tungsten can be installed and the end cap tightened to hold the tungsten in place. Select and install the desired nozzle size. Adjust the tungsten length so that it does not stick out more than the diameter of the nozzle, **Figure 4.50.**
14. Check the manufacturer's operating manual for the machine to ensure that all connections and settings are correct.

Table 4.5 Amperage Range of Tungsten Electrodes

Electrode Diameter		DCEN	DCEP	AC
in.	(mm)			
0.04	(1)	15–60	Not recommended	10–50
1/16	(2)	70–100	10–20	50–90
3/32	(2.4)	90–200	15–30	80–130
1/8	(3)	150–350	25–40	100–200
5/32	(4)	300–450	40–55	160–300

Figure 4.48
Setting the remote control switch
Courtesy of Larry Jeffus

Figure 4.49
Inserting collet and collet body
Courtesy of Larry Jeffus

Figure 4.50
Install the nozzle (cup) to the torch body
Courtesy of Larry Jeffus

CAUTION

Turn off all power before attempting to stop any leaks in the water system.

The GTA welding system is now ready to be used.

15. Turn on the power, depress the remote control, and again check for leaks.
16. While the postflow is still engaged, set the gas flow by adjusting the valve on the flowmeter.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

EXPERIMENT 4-5

Striking an Arc

Module 1
Key Indicator 1, 4

Module 7
Key Indicator 1
Carbon Steel
Key Indicator 3, 4
Austenitic Stainless Steel
Key Indicator 8, 9
Aluminum
Key Indicator 13, 14

Using a properly set up GTA welding machine, proper safety gear, and clean scrap metal, you will strike a GTA welding arc.

1. Position yourself so that you are comfortable and can see the torch, tungsten, and plate while the tungsten tip is held about 1/4 in. (6 mm) above the metal. Try to hold the torch at a vertical angle ranging from 0° to 15°. Too steep an angle will not give adequate gas coverage, **Figure 4.51.**

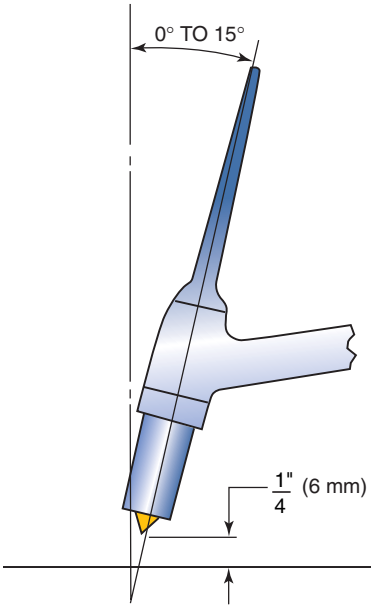

Figure 4.51
GTA torch position

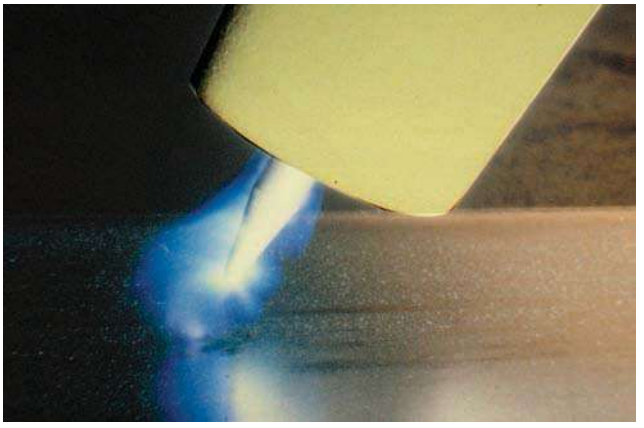

Figure 4.52
High frequency starting before arc starts
Courtesy of Larry Jeffus

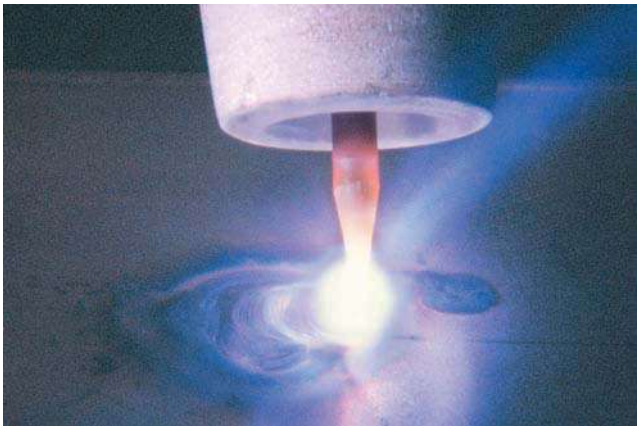

Figure 4.53
Stable gas tungsten arc
Courtesy of Larry Jeffus

2. Lower your arc welding helmet and depress the remote control. A high-pitched, erratic arc should be immediately jumping across the gap between the tungsten and the plate. If the high-frequency arc is not established, lower the torch until it appears, **Figure 4.52.**
3. Slowly increase the current until the main welding arc appears, **Figure 4.53.**
4. Observe the color change of the tungsten as the arc appears.
5. Move the tungsten around in a small circle until a molten weld pool appears on the metal.
6. Slowly decrease the current and observe the change in the molten weld pool.
7. Reduce the current until the arc is extinguished.
8. Hold the torch in place over the weld until the postflow stops.
9. Raise your hood and inspect the weld.

CAUTION

Avoid touching the metal table with any unprotected skin or jewelry. The high frequency can cause an uncomfortable shock.

Repeat this procedure until you can easily start the arc and establish a molten weld pool using both AC and DCEN currents. Turn off the welding machine, water, and shielding gas when you are finished; then clean up your work area.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

SUMMARY

One of the prime considerations for gas tungsten arc welding equipment setup is the cleanliness of the equipment, supplies, base material or material, and the welders themselves. When everything is clean, you will find that the welding process will proceed more easily and more successfully.

Another major factor affecting your ability to produce quality welds is the tungsten end or tip shape. As you practice making the various welds, you will find that keeping the tungsten electrode tip shaped appropriately will assist you in producing uniform welds.

Often, new welders feel that there is some sort of attraction between the tungsten electrode, filler metal, and base metal during the welding process because it seems to continually become contaminated. This almost continuous contamination can be very frustrating. At times it may seem overwhelming; however, with continued practice and diligence, you will be able to control this problem. Even experienced welders in the field can be plagued from time to time with tungsten contamination. At other times they can weld an entire day without contaminating the tungsten. It is often beneficial for students to realize that tungsten contamination is just part of the process, and they must therefore try to ignore the possibility of it happening and concentrate on producing the welds.

REVIEW

1. What early advancements made the GTA welding process more effective and reduced its cost?
2. What metals were weldable only by the GTAW process before GMAW was developed?
3. Which two of tungsten's properties make it the most versatile choice for GTA welding?
4. Why must the tip of the tungsten be hot?
5. Why does some tungsten erosion occur?
6. What function regarding tungsten heat do the collet and torch play?
7. What problem can an excessively large tungsten cause?
8. What holds the molten ball of tungsten in place at the tip of the electrode during DCEP welding?
9. Using **Table 4.1,** answer the following:
   a. What color identifies EWTh-2?
   b. What is the composition of EWCe-2?
   c. What color identifies EWLa-1.5
10. What does adding thorium oxide do for the tungsten electrode?

11. How can the end of a tungsten electrode be shaped?
12. Why should a grinding stone that is used for sharpening tungsten not be used for other metals?
13. Why should the grinding marks run lengthwise on the tungsten electrode end?
14. What are three ways of breaking off the contaminated end of a tungsten electrode?
15. Why should the torch be as cool as possible?
16. What will happen to a water-cooled torch cable if the flow of cooling water stops?
17. Why must shielding gas hoses not be made from rubber?
18. Why should the water solenoid be on the supply side of the water system?
19. What problem can a long nozzle cause to the tungsten?
20. Why must the tube of a flowmeter be vertical?
21. What is the heat distribution with DCEN welding current?
22. What is the heat distribution with DCEP welding current?
23. What is the heat distribution with AC welding current?
24. Why must AC welding power provided by transformer power supplies use high frequencies in order to work?
25. Why are argon and helium known as inert gases?
26. Why is argon's ease of ionization a benefit?
27. What makes helium difficult to use for manual welding?
28. What are the benefits of adding hydrogen to argon for welding?
29. What is the purpose of a hot start?
30. Using **Table 4.3,** determine the gas postflow time for a 3/32-in. (2.4-mm) tungsten.
31. How can air be drawn into the shielding gas?
32. Using **Table 4.4,** determine the minimum gas flow rate for a 1/2-in. (13-mm) nozzle.
33. What functions can a remote control provide the welder?

This page intentionally left blank

CHAPTER

5 Gas Tungsten Arc Welding of Plate

OBJECTIVES

After completing this chapter, the student should be able to

- list four applications for which the gas tungsten arc welding process is more commonly used
- describe four effects on the weld of varying torch angles
- explain why the filler rod end must be kept inside the protective zone of the shielding gas and how to accomplish this
- list four disadvantages to welding with a contaminated tungsten electrode
- list four techniques to reduce electrode contamination
- determine the correct machine settings for the minimum and maximum welding current for the machine used, the types and sizes of tungsten, and the metal types and thicknesses
- list four factors that affect the gas preflow and postflow times required to protect the electrode and the weld
- determine the minimum and maximum gas flow settings for each nozzle size, tungsten size, and amperage setting
- compare the characteristics of low-carbon and mild steels, stainless steel, and aluminum with respect to GTA welding
- prepare carbon steel, stainless steel, and aluminum for GTA welding

KEY TERMS

chill plate	gas coverage	protective zone
contamination	oxide layer	surface tension

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports or records
Key Indicator 3: Follows verbal instructions to complete work assignments
Key Indicator 4: Follows written instructions to complete work assignments

Module 2: Safety and Health of Welders

Key Indicator 1: Demonstrates proper use and inspection of personal protection equipment (PPE)
Key Indicator 2: Demonstrates safe operation practices in the work area
Key Indicator 3: Demonstrates proper use and inspection of ventilation equipment
Key Indicator 4: Demonstrates proper hot zone operation
Key Indicator 6: Understands proper use of precautionary labeling and MSDS information.
Key Indicator 7: Demonstrates proper inspection and operation of equipment for each welding or thermal cutting process used

Module 7: Gas Tungsten Arc Welding (GTAW)

Key Indicator 1: Performs Safety inspections of GTAW equipment and accessories
Key Indicator 2: Makes Minor external repairs to GTAW equipment and accessories

Carbon Steel

Key Indicator 3: Sets up for GTAW operations on carbon steel
Key Indicator 4: Operates GTAW equipment on carbon steel
Key Indicator 5: Makes GTAW fillet welds, in all positions, on carbon steel
Key Indicator 6: Makes GTAW groove welds, in all positions, on carbon steel
Key Indicator 7: Passes GTAW welder performance qualification test (workmanship sample) on carbon steel

Austenitic Stainless Steel

Key Indicator 8: Sets up for GTAW operations on austenitic stainless steel
Key Indicator 9: Operates GTAW equipment on austenitic stainless steel
Key Indicator 10: Makes GTAW fillet welds, in 1F, 2F, and 3F positions, on austenitic stainless steel
Key Indicator 11: Makes GTAW groove welds, in 1G and 2G positions, on austenitic stainless steel
Key Indicator 12: Passes GTAW welder performance qualification test (workmanship sample) on austenitic stainless steel

Aluminum

Key Indicator 13: Sets up for GTAW operations on aluminum
Key Indicator 14: Operates GTAW equipment on aluminum
Key Indicator 15: Makes GTAW fillet welds, in the 1F and 2F positions, on aluminum
Key Indicator 16: Makes GTAW groove welds, in the 1G position, on aluminum
Key Indicator 17: Passes GTAW welder performance qualification test (workmanship sample) on aluminum

Module 9: Welding Inspection and Testing

Key Indicator 1: Examines cut surfaces and edges of prepared base metal parts
Key Indicator 2: Examines tacks, root passes, intermediate layers, and completed welds.

INTRODUCTION

The gas tungsten arc welding process can be used to join nearly all types and thicknesses of metal. Welders can have a clear, unobstructed view of the molten weld pool because GTA welding is fluxless, slagless, and smokeless. The clear view of the weld allows welders to make changes in their welding technique, current, travel speed, and rate at which the filler metal is added to the weld as the weld progresses, ensuring that a quality weld is being made. This gives the welder very fine control of the welding process.

The fine control of the weld that is possible with GTA welding makes it an ideal process for very close-tolerance, high-quality welds. GTA welding is used to make critical welds, such as those on aircraft structures. If these welds fail, serious injury, death, and/or significant loss of property can result. Sometimes GTA welding is used to make the critical root pass of a weld that will be completed using another faster process. It is also used when weld appearance is important to the look of the finished part, as in some furniture, decorations, and/or sculptures.

The proper setup of GTA equipment can often affect the quality of the weld performed. Charts and graphs are available that give the correct amperage, gas flow rate, and time for various types of welds and metals. These charts are designed for optimum laboratory or classroom conditions. Actual conditions in the field will have an effect on these values. The experiments in this chapter are designed to help the welder understand the harmful effects on welding of less than ideal conditions. This will allow the welder to evaluate the appearance of a weld and make the necessary changes in technique or setup to improve the weld.

After a person has learned to weld in the lab, troubleshooting field welding problems will become much easier. The weld should be watched carefully to pick up any changes that could indicate a needed adjustment. When welders can do this, they have mastered the GTA process and have made themselves better potential employees. To make a weld is good; to solve a welding problem is better.

TORCH ANGLE

The torch should be held as close to perpendicular as possible in relation to the plate surface. The torch may be angled from 0° to 15° from perpendicular for better visibility and still have the proper shielding **gas coverage.** As the gas flows out it must form a **protective zone** around the weld. Tilting the torch changes the shape of this protective zone, **Figure 5.1.** Too much tilting of the torch will cause the protective shielding gas zone to become so distorted that the weld may not be protected from contamination from the air. The closer the torch is held to perpendicular, the better the weld is shielded.

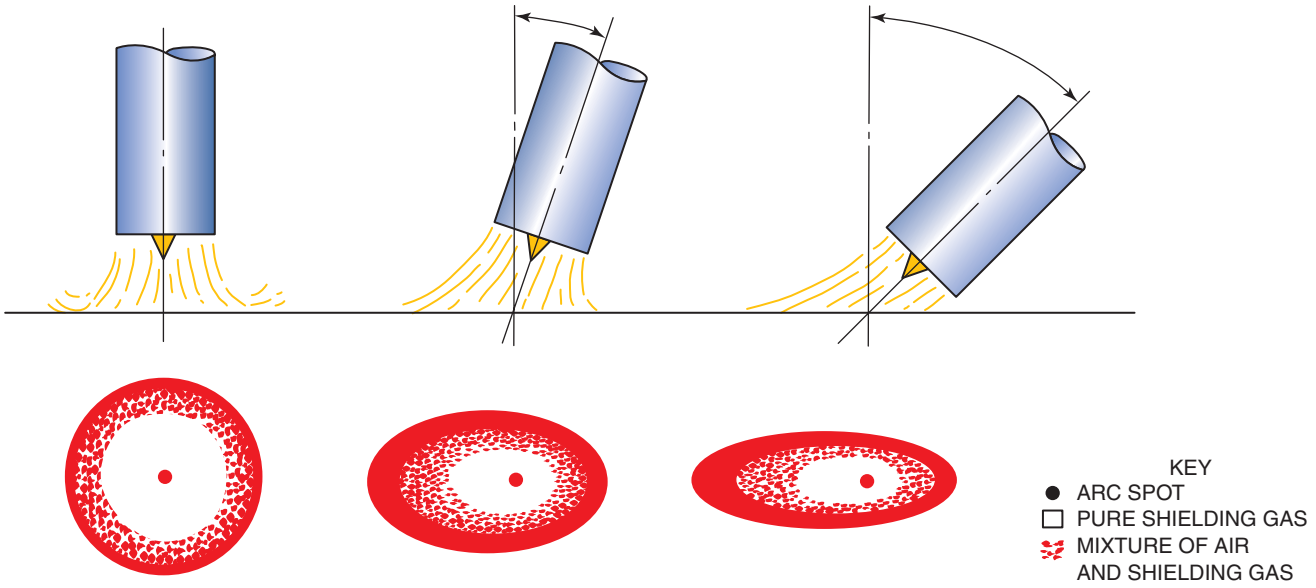

Figure 5.1
Gas coverage patterns for different GTA torch angles. Note how the area covered by the shielding gas becomes narrower and elongates as the angle of the torch increases from the perpendicular.

The velocity of the shielding gas also affects the protective zone as the torch angle changes. As the velocity increases, a low-pressure area develops behind the cup. When the low-pressure area becomes strong enough, air is pulled into the shielding gas. The sharper the angle and the higher the flow rate, the greater the possibility contamination will occur from the onset of turbulence in the gas stream. This causes air to become mixed with the shielding gas. Turbulence caused by the shielding gas striking the work will also cause air to mix with the shielding gas at high velocities.

FILLER ROD MANIPULATION

The filler rod end must be kept inside the protective zone of the shielding gas, **Figure 5.2.** The end of the filler rod is hot, and if it is removed from the gas protection, it will oxidize rapidly. The oxide will then be added to the

Figure 5.2
The hot filler rod end is well within the protective gas envelope
Courtesy of Larry Jeffus

Figure 5.3
(A) Filler properly protected, (B) some oxides on filler, and (C) excessive oxides caused by improper filler rod manipulation
Courtesy of Larry Jeffus

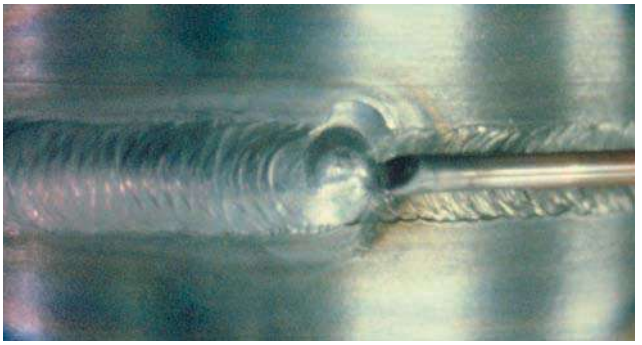

Figure 5.4
Filler being left in the molten weld pool as the arc is extinguished
Courtesy of Larry Jeffus

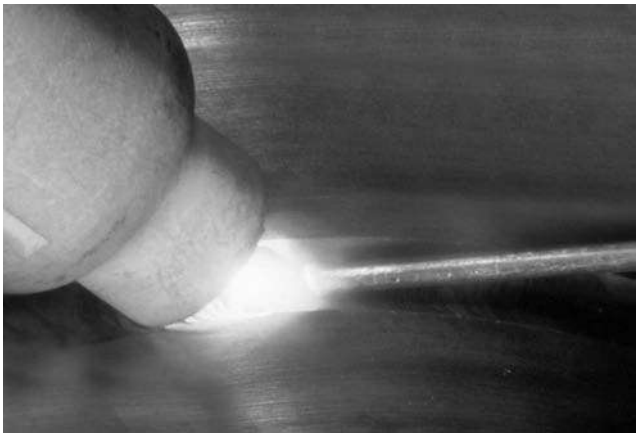

Figure 5.5
Filler being remelted as the weld is continued
Courtesy of Larry Jeffus

molten weld pool, **Figure 5.3.** When a weld is stopped so that the welder can change position, the shielding gas must be kept flowing around the rod end to protect it until it is cool. If the end of the rod becomes oxidized, it should be cut off before restarting. The following method can be used both to protect the rod end and reduce the possibility of crater cracking—that is, breaking the arc but keeping the torch over the crater while, at the same time, sticking the rod in the molten weld pool before it cools, **Figure 5.4.** When the weld is restarted, the rod is simply melted loose again, **Figure 5.5.**

The rod should enter the shielding gas as close to the base metal as possible, **Figure 5.6.** A 15° angle or less to the plate surface prevents air from being pulled into the welding zone behind the rod, **Figure 5.7.** As an example, if a rod is held in a stream of running water, air can be pulled in. The faster the water flows or the steeper the angle at which the rod is held, the more air is pulled in. The same action occurs with the shielding gas as its flow increases or as the rod angle increases.

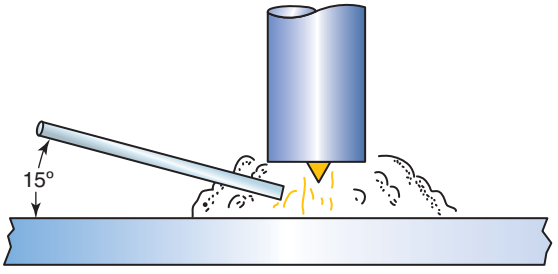

Figure 5.6
Keep the filled metal at approximately a 15° angle

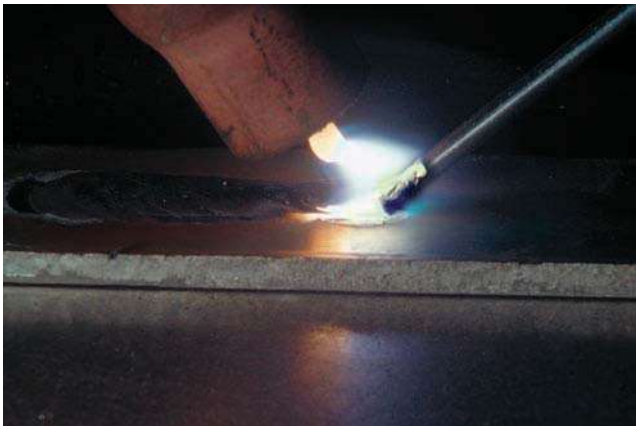

Figure 5.7
Too much filler rod angle has caused oxides to be formed on the filler rod end
Courtesy of Larry Jeffus

TUNGSTEN CONTAMINATION

For new welding students, the most frequently occurring and most time-consuming problem is tungsten **contamination.** The tungsten becomes contaminated when it touches the molten weld pool or when it is touched by the filler metal. When this happens, especially with aluminum, **surface tension** pulls the contamination up onto the hot tungsten, **Figure 5.8.** The extreme heat causes some of the metal to vaporize and form a large, widely scattered **oxide layer.** On aluminum, this layer is black. On iron (steel and stainless steel), this layer is a reddish color.

The contamination caused by the tungsten touching the molten weld pool or filler metal forms a weak weld. On a welding job, both the weld and the tungsten must be cleaned before any more welding can be done. The weld crater must be ground or chiseled to remove the tungsten contamination, and the tungsten end must be reshaped. Extremely tiny tungsten particles will show up if the weld is X-rayed. Failure to remove the contamination properly will result in the failure of the weld.

When starting to weld, the beginning student may save weld practice time by burning off the contamination. On a scrap, usually copper plate, strike an arc using a higher than normal amperage setting. The arc will be erratic and discolored at first, but, as the contamination vaporizes, the arc will stabilize. Contamination can also be knocked off by quickly flipping the torch head.

CAUTION

This procedure should never be used with heavy contaminations or when a welder is on the job in the field. It is designed only to help the new student in the first few days of training to save time and increase weld production.

CURRENT SETTING

The amperage set on a machine and the actual welding current are often not the same. The amperage indicated on the machine's control is the same as that at the arc only for the following conditions:

- The power to the machine is exactly correct.
- The lead length is very short.
- All cable connections are perfect with zero resistance.
- The arc length is exactly the right length.
- The remote current control is in the full on position.

Figure 5.8
Contaminated tungsten
Courtesy of Larry Jeffus

Table 5.1 **Sample Chart Used to Record GTA Welding Machine Settings**

Current and Tungsten Electrode Size	Amperage/Machine Setting				
	Too Low	Low	Good	High	Too High

If any one of these factors changes, the actual welding amperage will change.

In addition to the difference between indicated and actual welding amperage, there is a more significant difference between amperage and welding power. The welding power, in watts, is based on the formula $W = E \times I$, or volts (E) multiplied by amperes (I) equals watts (W). Thus, the indicated power to a weld from two different types of welding machines set at 100 amperes will vary depending upon the voltage of the machine.

The welding machine setting will vary within a range from low to high (cool to hot). The range for one machine may be different from that of another machine. The setting will also be different for various types and sizes of tungstens, polarities, types and thicknesses of metal, joint position or design, and shielding gas used.

A chart, such as the one in **Table 5.1,** and a series of tests can be used to set the lower and upper limits for the amperage settings. As students' welding skills improve with practice, they will become familiar with the machine settings so that a table for these settings is no longer needed. In the welding industry, some welders will mark a line on the dial of the machine to help in resetting the machine. If a welder is required to make a number of different machine setups, a list or chart can be made and taped to the machine. This practice is more professional than marking the machine dials.

EXPERIMENTS

Experiments are designed to help new welders learn some basic skills that will help them troubleshoot welding problems. If you do the experiments listed in this chapter, you will be better able to determine what is causing a problem with your weld. As you learn more about welding, subtle changes will become more noticeable. Even experienced welders make changes in the setup, current, or welding technique as they try to resolve a problem.

Experiment 5-1 will help the welder determine the correct machine settings for the minimum and maximum welding current for the machine used, the types and sizes of tungstens, and the metal types and thicknesses. Most welding will be performed with a medium-range or mid-range machine setting. The exact setting is more important for machines without remote controls. The remote control allows changes in welding current to be made during the welding without having to stop.

EXPERIMENT 5-1

Setting the Welding Current

Module 1
Key Indicator 1, 4

Module 7
Key Indicator 1, 3, 4

Module 9
Key Indicator 1, 2

Using a properly set up GTA welding machine and torch, proper safety protection, one of each available tungsten size and type, and 16-gauge mild steel, 1/8 in. (3 mm) and 1/4 in. (6 mm) thick, you will work with a small group of students to develop a chart of the correct machine current setting for each type and size of tungsten.

Set the machine welding power switch for DCEN (DCSP) and the amperage control to its lowest setting, **Figure 5.9.** Sharpen a point on each tungsten and install one of the smaller diameter tungstens in the GTA torch. Select a nozzle with a 1/2-in. (13-mm)-diameter hole and attach it to the torch head. Set the preflow time to 0 and postflow to 10 to 15 seconds. Connect the remote control if it is available. Turn on the main power and hold the torch so that it cannot short out. Depress the remote controls to start the shielding gas so the flow rate can be set at 20 cfh (8 L/min). Switch the high frequency to start. All other functions, such as pulse, hot start, slope, and so on, should be in the off position.

Place the piece of 16-gauge sheet metal flat on the welding table. Hold the torch vertically with the tungsten about 1/4 in. (6 mm) above the metal. Lower your welding hood and fully depress the remote control. Watch the arc to see if it stabilizes and melts the metal. After a short period of time (15 to 30 seconds), stop, raise your hood, and check the plate for a melted spot. If melting occurred, note the size of the spot and depth of penetration, **Figure 5.10.** Increase the amperage setting by 5 or 10 A, note the setting on the chart, and repeat the process.

After each test, observe and record the results. The important settings to note are:

1. when the tungsten first heats up and the arc stabilizes
2. when the metal first melts
3. when 100% penetration of the metal first occurs
4. when burn-through first occurs
5. when the tungsten starts glowing white hot and/or melts

Figure 5.9
Lower the welding current to zero or as low as possible
Courtesy of Larry Jeffus

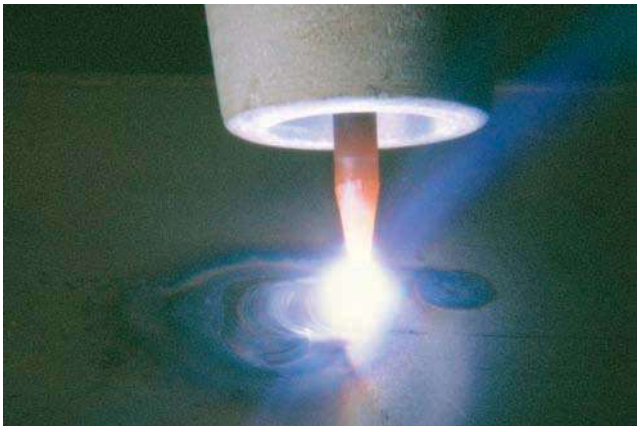

Figure 5.10
Melting first occurring
Courtesy of Larry Jeffus

The lowest (minimum) acceptable amperage setting is when the molten weld pool first appears on the base metal and the arc is stable. The highest (maximum) amperage setting is when the base metal burn-through or melting of the tungsten occurs. Any current setting between the high and low points is within the amperage range for that specific setup.

To establish the range for the next tungsten type or size, repeat the test. After each test, the metal should be cooled to prevent overheating. After each type and size of tungsten has been tested and an operating range established, repeat the procedure using the next thicker metal. Repeat this procedure until you have set up the operating ranges for all of the metals and tungstens you will be using. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

GAS FLOW

The gas preflow and postflow times required to protect both the tungsten and the weld depend upon the following factors:

- wind or draft speed
- nozzle size used
- tungsten size used
- amperage
- joint design
- welding position
- type of metal welded

The weld quality can be adversely affected by improper gas flow settings. The lowest possible gas flow rates and the shortest preflow or postflow time can help reduce the cost of welding by saving the expensive shielding gas.

In Experiment 5-2, the minimum and maximum gas flow settings for each nozzle size, tungsten size, and amperage setting will be determined. The chart a welder prepares based on experiments is to improve that welder's skill and welding technique. Charts may differ slightly from one welder to another. As a welder's skill improves, the chart may change. As experience is gained, a welder will learn how to set the gas flow effectively without the need for this chart.

The minimum flow rates and times must be increased when welding in drafty areas or for out-of-position welds. The rates and times can be somewhat lower for tee joints or welds made in tight areas. The maximum flow rates must never be exceeded. Exceeding these flow rates causes weld contamination and increases the rejection rate.

EXPERIMENT 5-2

Setting Gas Flow

Using a properly set up GTA welding machine and torch, proper safety protection, one of each available tungsten size, metal that is 16 gauge to 1/4 in. (6 mm) thick, and the welding current chart developed in

Module 1
Key Indicator 1, 4

Module 7
Key Indicator 1
Carbon Steel
Key Indicator 3, 4

Module 9
Key Indicator 1, 2

Figure 5.11
Setting the postflow timer
Courtesy of Larry Jeffus

Figure 5.12
Oxides forming due to inadequate gas shielding
Courtesy of Larry Jeffus

Experiment 5-1, you will work with a small group of students to make a chart of the minimum and maximum flow rates and times for each nozzle size, tungsten size, and amperage setting. An assistant will also be needed to change and record the flow rate while you work.

Set the machine welding power switch for DCEN (DCSP). Set the amperage to the lowest setting for the size of tungsten used. Set the preflow time to 0 and postflow to 20 seconds, **Figure 5.11.** Turn on the main power. With the torch held so that it cannot short out, depress the remote control to start the shielding gas flow and set the flow at 20 cfh (9 L/min). Switch the high frequency to start. All other functions, such as pulse, hot start, slope, and so on, should be in the off position.

Starting with the smallest nozzle and tungsten size, strike an arc and establish a molten pool on a piece of metal in the flat position. Watch the molten weld pool and tungsten for signs of oxide formation as another person slowly lowers the gas flow rate. Have that person note this setting (where oxide formation begins), **Figure 5.12,** as the minimum flow rate on the chart next to the nozzle size and current setting. Now slowly increase the flow rate until the molten pool starts to be blown back or oxides start forming. This setting should be noted on the chart as the maximum flow rate for this current and nozzle size, **Table 5.2.** Lower the flow to a rate of 2 cfh or 3 cfh (1 L/min or 2 L/min) above the minimum value noted on the chart, and then stop the arc. Record the length of time from the point when the arc stops and the tungsten stops glowing as the postflow time. Repeat this test at a medium and then high current setting for this nozzle

Table 5.2 Sample Chart for Setting Shielding Gas Flow Rate and Time

Electrode and Nozzle Size	Flow Rate					Postflow Time		
	Too Low	Low	Good	High	Too High	Too Short	OK	Too Long

and tungsten size. When using high current settings, it may be necessary to move the torch or use thicker plate to prevent burn-through.

Repeat this test procedure with each available nozzle and tungsten size. Stainless steel or aluminum is preferred for this experiment because the oxides are more quickly noticeable than when mild steel is used. If aluminum is used, the welding current must be AC, and the high-frequency switch should be set on continuous.

To establish the minimum preflow time for each nozzle and tungsten size, set the amperage to a medium-high setting. Hold the torch above the metal so that an arc will be instantly started. Set the preflow timer to 0 and the gas flow to just above the minimum value noted on the chart. Quickly strike an arc on metal thin enough to cause a weld pool to form instantly at that power setting. Stop the arc and examine the weld pool and tungsten for oxides. Repeat this procedure, increasing the preflow time until no oxides are formed on either the plate or tungsten. Record this time on the chart as the minimum preflow time. Repeat this test with each available nozzle and tungsten size. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE WELDS

The practice welds are grouped according to the weld position and type of joint, and not by the type of metal. The order in which a person decides to do the welds is that person's choice. It is suggested that the stringer beads be done in each metal and position before the different joints are tried. Each metal has its own characteristics that may make one metal easier for a person to work on than another metal.

Mild steel is inexpensive and requires the least amount of cleaning. Slight changes in the metal have little effect on the welding skill required. Stainless steel is somewhat affected by cleanliness, requiring little preweld cleaning. However, the weld pool shows overheating or poor gas coverage. With aluminum, cleanliness is a critical factor. Oxides on aluminum may prevent the molten weld pool from flowing together. The surface tension helps hold the metal in place, giving excellent bead contour and appearance.

The degree of difficulty a welder encounters with each of these metals depends upon the individual's experience. Try each weld with each metal to determine which metal will be easiest to master first. The type of welding machine and materials used will also affect a welder's progress. Practice will help welders overcome any obstacle to their progress.

Low-Carbon and Mild Steels

Low-carbon and mild steel are two basic steel classifications. These steels are the most common type of steels a new GTA welding student will experience welding. Carbon is the primary alloy in these classifications of steel, and it ranges from 0.15% or less for low carbon and 0.15% to 0.30% for mild steel. The GTA welding techniques required for welding steels in both

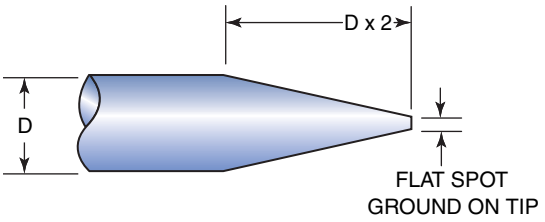

Figure 5.13
Tungsten tip shape for mild steel or stainless steel

classifications are the same. You start with EWTh-1, EWTh-2, EWCe-2, or one of the EWLa pointed tungsten electrodes, **Figure 5.13,** with the welding machine set for DCEN (DCSP) welding current. **Table 5.3** lists the types of filler metal used for both low-carbon and mild steels.

During the manufacturing process, small pockets of primarily carbon dioxide gas sometimes become trapped inside low-carbon and mild steels. There are only a few molecules of gas trapped inside the microscopic pockets within the steel, so they do not affect the steel strength. During most other types of welding, fluxes on the filler metal capture these gas pockets and they are removed. However, because GTA filler metals do not have fluxes when these gas pockets become hot during welding, they expand and can sometimes cause weld porosity. You are most likely to see porosity caused by these gases when you are not using a filler metal. Most GTA filler metals have some alloys, called deoxidizers, that can help prevent porosity caused by gases trapped in the base metal. RG45 gas welding rods do not have these deoxidizers and are not recommended for GTA welding.

Stainless Steel

The setup and manipulation techniques required for stainless steel are nearly the same as those for low-carbon and mild steels, and skills transfer is easy. The major difference is that most welds on steels do not show the effects of contamination as easily as do welds on stainless steels. To make a weld on stainless steel you must do a better job of precleaning the base metal and filler metal; make sure you have adequate shielding gas coverage and do not overheat the weld.

Table 5.3 **Filler Metals for Low-Carbon and Mild Steels**

SAE No.	Carbon %	AWS Filler Metal No.
	Low Carbon	
1006	0.08 max	RG60 or ER70S-3
1008	0.10 max	RG60 or ER70S-3
1010	0.08 to 0.15	RG60 or ER70S-3
	Mild Steel	
1015	0.11 to 0.16	RG60 or ER70S-3
1016	0.13 to 0.18	RG60 or ER70S-3
1018	0.15 to 0.20	RG60 or ER70S-3
1020	0.18 to 0.23	RG60 or ER70S-3
1025	0.22 to 0.29	RG60 or ER70S-3

Table 5.4 **Temperatures at Which Various Colored Oxide Layers Form on Steel**

Surface Color	Approximate Temperature at Which Color is Formed	
	°F	(°C)
Light Straw	400	(200)
Tan	450	(230)
Brown	525	(275)
Purple	575	(300)
Dark Blue	600	(315)
Black	800	(425)

Table 5.5 **Filler Metals for Stainless Steels**

AISI No.	AWS Filler No.	AISI No.	AWS Filler No.
303	ER308	310	ER310
304	ER308	316	ER316L
304L	ER308L	316L	ER316L
309	ER309	410	ER410

The most common sign that there is a problem with a stainless steel weld is the bead color after the weld. The greater the contamination, the darker the color. The exposure of the weld bead to the atmosphere before it has cooled will also change the bead color. It is impossible, however, to determine the extent of contamination of a weld with only visual inspection. Both light-colored and dark-colored welds may not be free from oxides. Thus, it is desirable to take the time and necessary precautions to make welds that are no darker than dark blue, **Table 5.4.** Welds with only slight oxide layers are better for multiple passes.

Using a low arc current setting with faster travel speeds is important when welding stainless steel, because some stainless steels are subject to carbide precipitation. Carbide precipitation, the combining of carbon with chromium, occurs in some stainless steels when they are kept at a temperature between 800°F and 1500°F (625°C and 815°C) for a long time. There are a number of ways of controlling carbide precipitation, including the adding of alloys to the stainless steel and the lowering of the percentage of carbon. Also, during welding, special alloy filler metals can be used to control the problem, but the most important thing a welder can do is travel fast and use as little welding heat as possible.

Black crusty spots may appear on weld beads. These spots are often caused by improper cleaning of the filler rod or failure to keep the end of the rod inside the shielding gas.

Table 5.5 lists some common types of stainless steels and the recommended filler metals.

Aluminum

Aluminum is GTA welded using EWP, EWZr, EWCe-2, or EWLa rounded tip tungsten electrodes on transformer rectifier welding power supplies, **Figure 5.14,** with the welding machine set for ACHF welding current. EWZr, EWCe-2, or EWLa electrodes with a pointed tip may be used with some inverter welding power supplies due to the heat control afforded by

Figure 5.14
Tungsten tip shape for aluminum

Table 5.6 Filler Metals for Aluminum Alloys

AISI No.	AWS Filler No.	AISI No.	AWS Filler No.
1100	ER1100	3004	ER4043
3003	ER4043	6061	ER4043

their advanced circuitry. The alternating current provides good arc cleaning, and the continuous high frequency restarts the arc as the current changes direction.

The molten aluminum weld pool has high surface tension, which allows large weld beads to be controlled easily. **Table 5.6** lists some basic types of filler metals used for aluminum welding.

The high thermal conductivity of the metal may make starting a weld on thick sections difficult without first preheating the base metal. In most cases the preheat temperature is around 300°F (150°C) but will vary depending on metal thickness and alloy type. Specific preheat temperatures are available from the metal supplier.

The processes of cleaning and keeping the metal clean take a lot of time. Removal of the oxide layer is easy using a chemical or mechanical method. Ten minutes after cleaning, however, the oxide layer may again be thick enough to require recleaning. The oxide that forms reduces the ability of the weld pool to flow together. Keep your hands and gloves clean and oil free so the base metal or filler rods do not become recontaminated.

Although aluminum resists oxidation at room temperature, it rapidly oxidizes at welding temperatures. If the filler rod is not kept inside the shielding gas, it will quickly oxidize, but, because of the low melting temperature of the filler rod, the end will melt before it is added to the weld pool if it is held too closely to the arc, **Figure 5.15** and **Figure 5.16.**

CAUTION

The manufacturer's recommendations for using these products must be followed. Failure to do so may result in chemical burns, fires, fumes, or other safety hazards that could lead to serious injury. If anyone should come in contact with any chemicals, immediately refer to the material safety data sheet (MSDS) for the proper corrective action.

Metal Preparation

Both the base metal and the filler metal used in the GTAW process must be thoroughly cleaned before welding. Contamination left on the metal will be deposited in the weld because there is no flux to remove it. Oxides, oils, and dirt are the most common types of contaminants. They can be

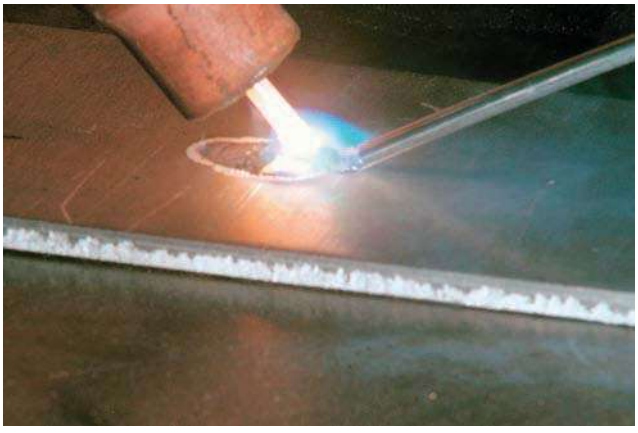

Figure 5.15
Aluminum filler being correctly added to the molten weld pool
Courtesy of Larry Jeffus

Figure 5.16
Filler rod being melted before it is added to the molten pool
Courtesy of Larry Jeffus

removed mechanically or chemically. Mechanical metal cleaning may be done by grinding, wire brushing, scraping, machining, or filing. Chemical cleaning may be done by using acids, alkalies, solvents, or detergents.

PRACTICE 5-1

Stringer Beads, Flat Position, on Mild Steel

Using a properly set up and adjusted GTA welding machine on DCEN, proper safety protection, and one or more pieces of mild steel, 6 in. (152 mm) long and 16 gauge and 1/8 in. (3 mm) thick, you will push a weld pool in a straight line down the plate, **Figure 5.17.** Maintain uniform weld pool size and penetration.

- Starting at one end of the piece of metal that is 1/8 in. (3 mm) thick, hold the torch as close as possible to a 90° angle.
- Lower your hood, strike an arc, and establish a weld pool.
- Move the torch in a stepping or circular oscillation pattern down the plate toward the other end, **Figure 5.18.**
- If the size of the weld pool changes, speed up or slow down the travel rate to keep the weld pool the same size for the entire length of the plate.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 3, 4

Module 9
Key Indicator 1,2

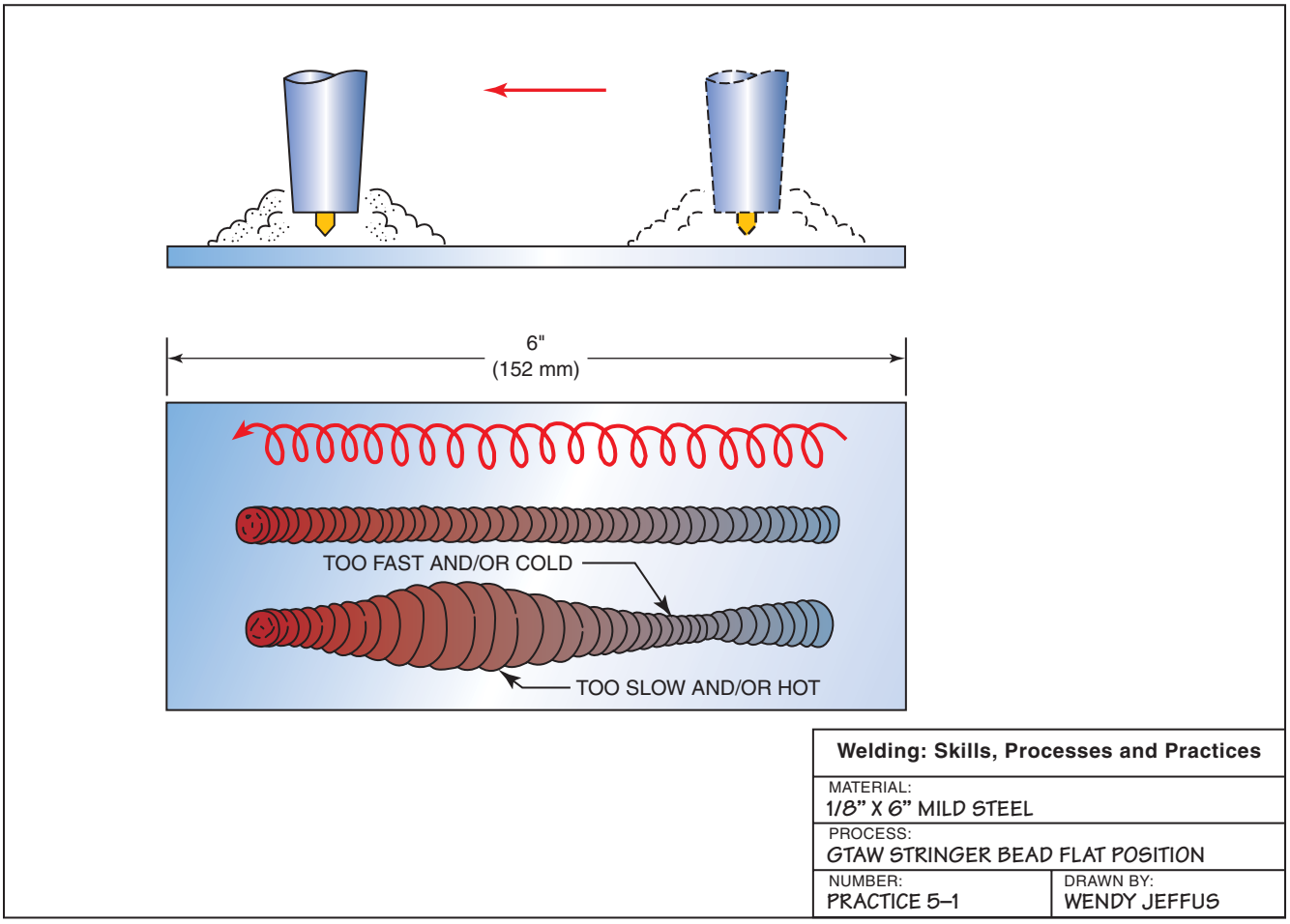

Figure 5.17
Surfacing weld in the flat position

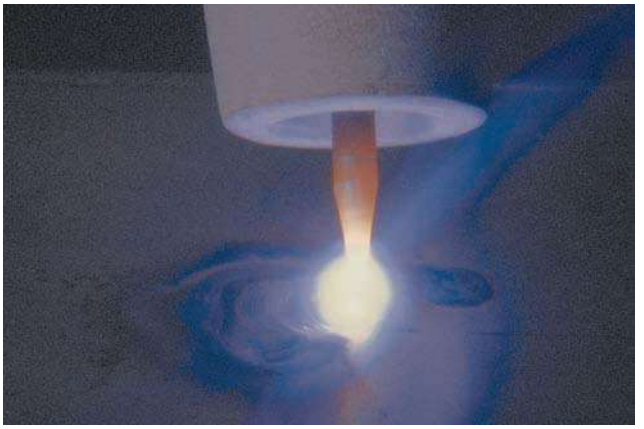

Figure 5.18
Surfacing weld
Courtesy of Larry Jeffus

The ability to maintain uniformity in width and keep a straight line increases as you are able to see more than just the weld pool. As your skill improves, you will relax, and your field of vision will increase.

Repeat the process using both thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Austenitic Stainless Steel
Key Indicator 8, 9

Module 9
Key Indicator 1, 2

PRACTICE 5-2

Stringer Beads, Flat Position, on Stainless Steel

Using the same equipment and material thicknesses as listed in Practice 5-1 and one or more pieces of stainless steel, 6 in. (152 mm) long and 1/4 in. (6 mm) thick, you will push a molten weld pool in a straight line down the plate, keeping the width and penetration uniform.

To keep the formation of oxides on the bead to a minimum, a **chill plate** (a thick piece of metal used to absorb heat) may be required. Another method is to make the bead using as low a heat input as possible. When the weld is finished, the weld bead should be no darker than dark blue.

Repeat the process using both thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Aluminum
Key Indicator 13, 14

Module 9
Key Indicator 1, 2

PRACTICE 5-3

Stringer Beads, Flat Position, on Aluminum

Using the same equipment, setup for AC, and procedure as listed in Practice 5-1 and pieces of aluminum, 6 in. (152 mm) long and 1/16 in. (2 mm), 1/8 in. (3 mm), and 1/4 in. (6 mm) thick, you will push a weld pool

in a straight line, maintaining uniform width and penetration for the length of the plate.

A high current setting will allow faster travel speeds. The faster speed helps control excessive penetration. Hot cracking may occur on some types of aluminum after a surfacing weld. This is not normally a problem when filler metal is added. If hot cracking should occur during this practice, do not be concerned.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 3, 4
Austenitic Stainless Steel
Key Indicator 8, 9
Aluminum
Key Indicator 13, 14

Module 9
Key Indicator 1, 2

PRACTICE 5-4

Flat Position, Using Mild Steel, Stainless Steel, Aluminum

For this practice, you will need a properly set up and adjusted GTA welding machine, proper safety protection, and filler rods, 36 in. (0.9 m) long × 1/16 in. (2 mm), 3/32 in. (2.4 mm), and 1/8 in. (3 mm) in diameter, one or more pieces of mild steel, stainless steel, and aluminum, 6 in. (152 mm) long × 1/16 in. (2 mm) and 1/8 in. (3 mm) thick, and aluminum plate 1/4 in. (6 mm) thick, **Table 5.7, Table 5.8,** and **Table 5.9.** In this practice,

Table 5.7 Suggested Setting for GTA Welding of Mild Steel

Tungsten			Welding			Shielding			Filler	
Type	Size	Tip	Amp	Current	HF	Type	Flow	Size	Type	Size
EWTh-1 or EWTh-2 or EWCe-2 or EWLa	1/16 in. (2 mm)	Point	50 to 100	DCEN DCSP	Start or auto	Argon	16 cfh 7 L/min auto	3/8 in. (10 mm)	RG60 or ER70S-3	1/16–3/32 in. (2–2.4 mm)
	3/32 in. (2.4 mm)	Point	70 to 150	DCEN DCSP	Start or auto	Argon	16 cfh 7 L/min auto	3/8 in. (10 mm)	RG60 or ER70S-3	1/16–3/32 in. (2–2.4 mm)
	1/8 in. (3 mm)	Point	90 to 250	DCEN DCSP	Start or auto	Argon	20 cfh 9 L/min	1/2 in. (13 mm)	RG60 or ER70S-3	3/32–1/8 in. (2.4–3 mm)

Table 5.8 Suggested Setting for GTA Welding of Stainless Steel

Tungsten			Welding			Shielding			Filler	
Type	Size	Tip	Amp	Current	HF	Type	Flow	Size	Type	Size
EWTh-1 or EWTh-2	1/16 in. (2 mm)	Point	70 to 100	DCEN DCSP	Start or auto	Argon	16 cfh 7 L/min	3/8 in. (10 mm)	ER308 or ER316	1/16–3/32 in. (2–2.4 mm)
EWTh-1 or EWCe-2 or EWLa	3/32 in. (2.4 mm)	Point	70 to 150	DCEN DCSP	Start or auto	Argon	16 cfh 7 L/min	3/8 in. (10 mm)	ER308 or ER316	1/16–3/32 in. (2–2.4 mm)
	1/8 in. (3 mm)	Point	90 to 250	DCEN DCSP	Start or auto	Argon	20 cfh 9 L/min	1/2 in. (13 mm)	ER308 or ER316	3/32–1/8 in. (2.4–3 mm)

Table 5.9 Suggested Setting for GTA Welding of Aluminum

Tungsten			Welding			Shielding			Filler	
Type	Size	Tip	Amp	Current	HF	Type	Flow	Size	Type	Size
EWZr or EWLa or EWCe-2 or EWP	1/16 in.	Round 2 mm	50 to 90	AC	Continues or on	Argon	17 cfh 8 L/min	7/16 in. (11 mm)	ER1100 or ER4043	1/16–3/32 in. (2–2.4 mm)
	3/32 in.	Round 2.4 mm	80 to 130	AC	Continues or on	Argon	20 cfh 9 L/min	1/2 in. (13 mm)	ER1100 or ER4043	1/16–3/32 in. (2–2.4 mm)
	1/8 in.	Round 3 mm	100 to 200	AC	Continues or on	Argon	20 cfh 9 L/min	5/8 in. (16 mm)	ER1100 or ER4043	3/32–1/8 in. (2.4–3 mm)

Note: A pointed electrode may be preferred with AC and inverter GTAW power supplies with enhanced balance control capabilities.

you will make a straight stringer bead, 6 in. (152 mm) long, that is uniform in width, reinforcement, and penetration, **Figure 5.19.** Use DCEN current on the steel and AC current on the aluminum.

Starting with the metal that is 1/8 in. (3 mm) thick and the filler rod having a 3/32-in. (2.4-mm) diameter, strike an arc and establish a weld pool, **Figure 5.20.** Move the torch in a circle as in the practice beading. When the torch is on one side, add filler rod to the other side of the molten weld pool, **Figure 5.21.** The end of the rod can be held lightly in the leading edge of the molten weld pool, or it can be dipped into the molten weld pool. If you are using the dipping method, be sure not to allow the tip to melt and drip into the weld pool, **Figure 5.22.** Change to another size filler rod and determine its effect on the weld pool.

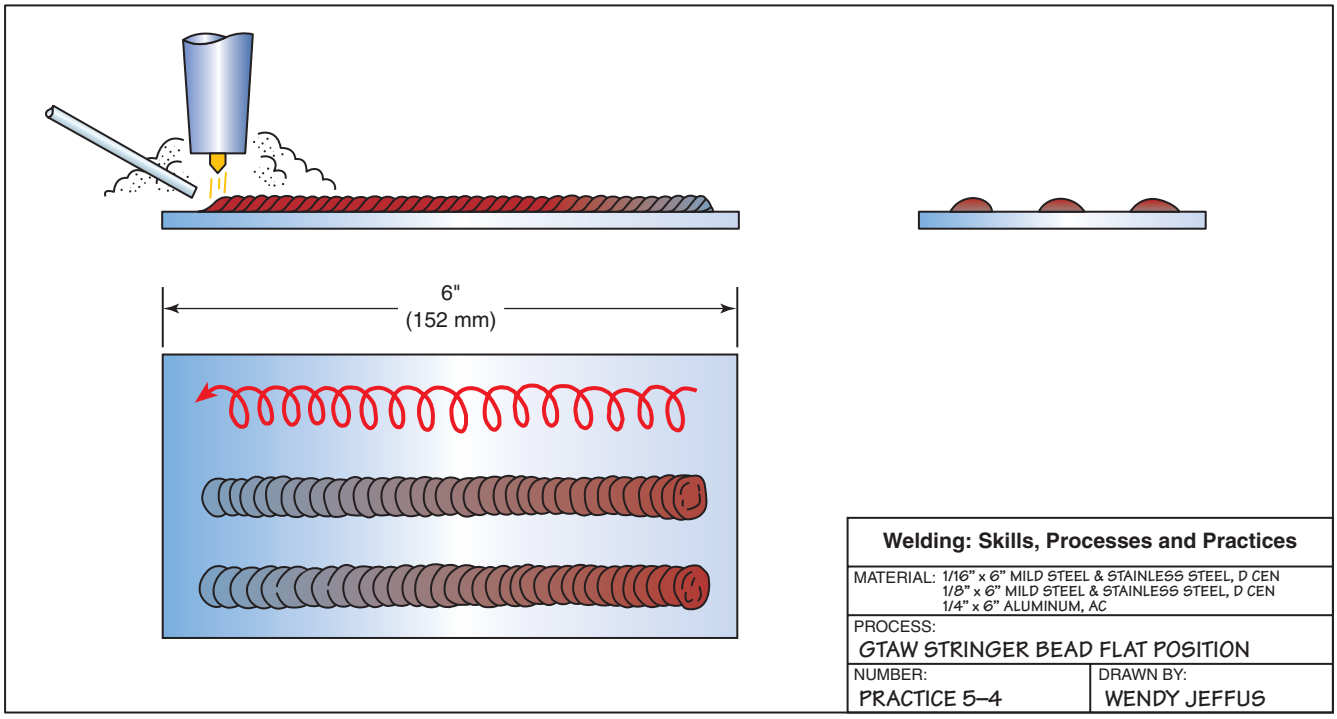

Figure 5.19
Stringer beads in the flat position

Figure 5.20
Establish a molten weld pool and dip the filler rod into it
Courtesy of Larry Jeffu

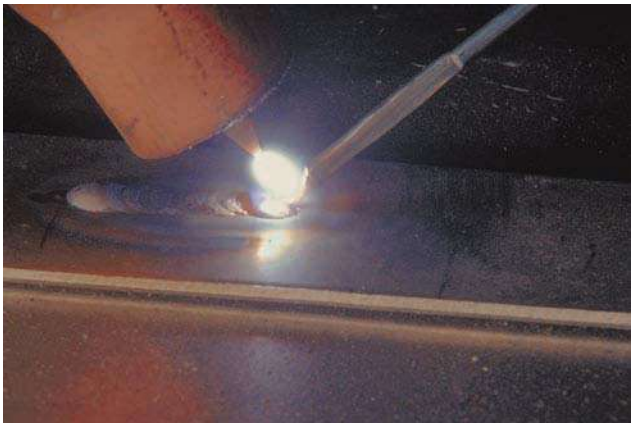

Figure 5.21
Note the difference in the weld produced when different size filler rods are used
Courtesy of Larry Jeffus

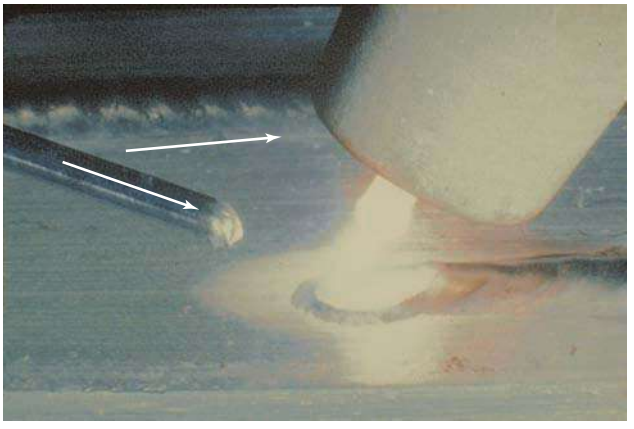

Figure 5.22
Move the electrode back as the filler rod is added
Courtesy of Larry Jeffus

Maintain a smooth and uniform rhythm as filler metal is added. This will help to keep the bead uniform. Vary the rhythms to determine which one is easiest for you. If the rod sticks, move the torch toward the rod until it melts free.

When the full 6-in. (152-mm)-long weld bead is completed, cool and inspect it for uniformity and defects. Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Key Indicator 1, 2
Carbon Steel
Key Indicator 3, 4
Austenitic Stainless Steel
Key Indicator 8, 9
Aluminum
Key Indicator 13, 14

Module 9
Key Indicator 1, 2

PRACTICE 5-5

Outside Corner Joint, 1G Position, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials listed in Practice 5-4, weld an outside corner joint in the flat position, **Figure 5.23.**

- Place one of the pieces of metal flat on the table and hold or brace the other piece of metal horizontally on it.
- Tack weld both ends of the plates together, **Figure 5.24.**
- Set the plates up and add two or three more tack welds on the joint as required, **Figure 5.25.**

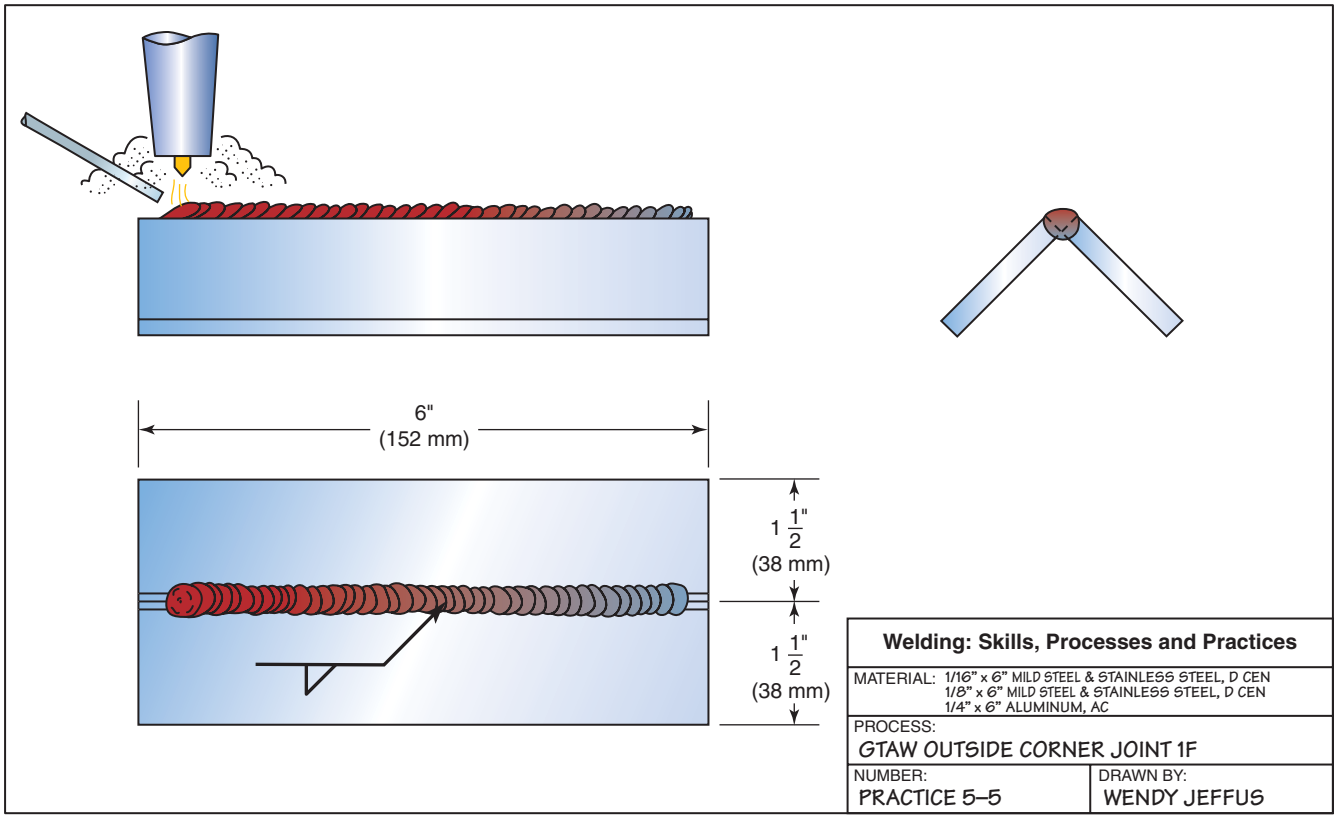

Figure 5.23
Outside corner joint in the flat position

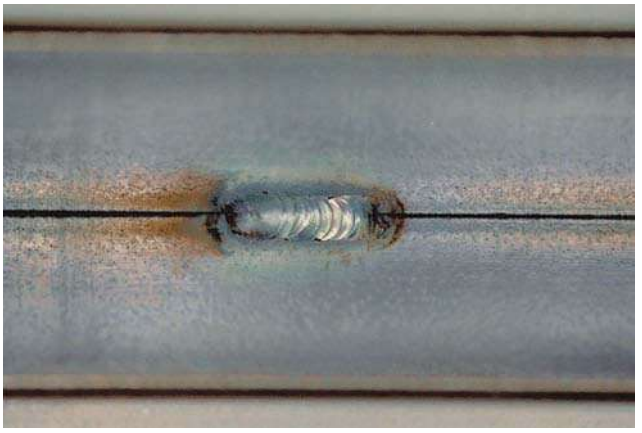

Figure 5.24
Tack weld. Note the good fusion at the start and crater fill at the end.
Courtesy of Larry Jeffus

Figure 5.25
Outside corner tack welded together
Courtesy of Larry Jeffus

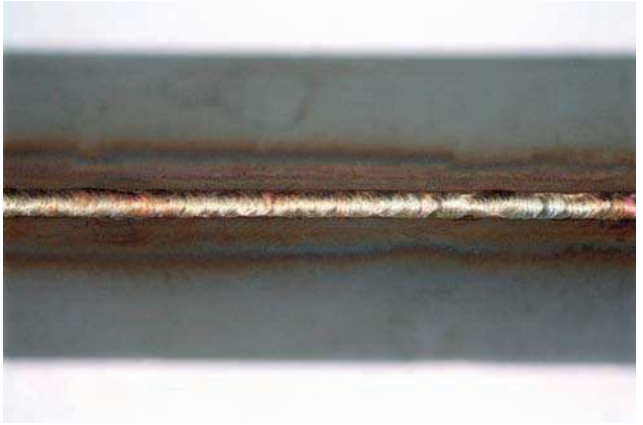

Figure 5.26
Outside corner joint. Note precleaning along weld.
Courtesy of Larry Jeffus

- Starting at one end, make a uniform weld, adding filler metal as needed. In **Figure 5.26,** note the metal areas that are precleaned before the weld is made.

Repeat each weld as needed until all are mastered. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 16

Module 9
Key Indicator 1, 2

PRACTICE 5-6

Butt Joint, 1G Position, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will weld a butt joint in the flat position, **Figure 5.27.**

Place the metal flat on the table and tack weld both ends together, **Figure 5.28.** Two or three additional tack welds can be made along the

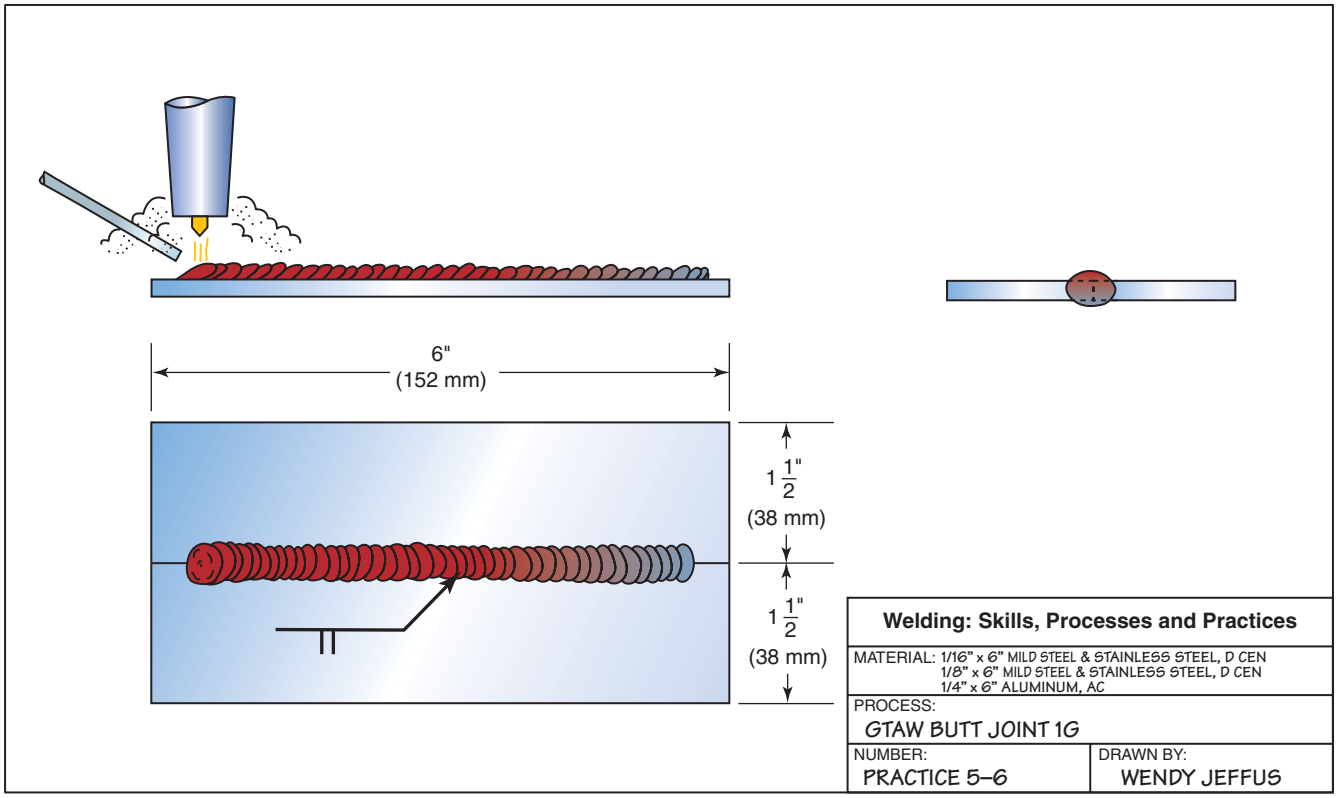

Figure 5.27
Square butt joint in the flat position

joint as needed. Starting at one end, make a uniform weld along the joint. Add filler metal as required to make a uniform weld.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Figure 5.28
Tack weld on butt joint
Courtesy of Larry Jeffus

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-7

Butt Joint, 1G Position, with 100% Penetration, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 16

Module 9
Key Indicator 1, 2

Using the same equipment and materials as listed in Practice 5-4, you will weld a butt joint with 100% penetration, **Figure 5.29,** along the entire 6-in. (152-mm) length of the joint. After the weld is completed, shear out strips 1 in. (25 mm) wide and bend-test them as shown in **Figure 5.30.**

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-8

Butt Joint, 1G Position, with Minimum Distortion, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 16

Module 9
Key Indicator 1, 2

Using the same equipment and materials as listed in Practice 5-4, you will weld a flat butt joint, while controlling both distortion and penetration, **Figure 5.31.**

Tack weld the plates together as shown in **Figure 5.32.** Using a back-stepping weld sequence, make a series of welds approximately 1 in. (25 mm) long along the joint. Be sure to fill each weld crater adequately to reduce crater cracking.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

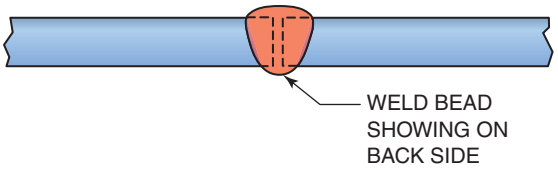

Figure 5.29
100% weld penetration

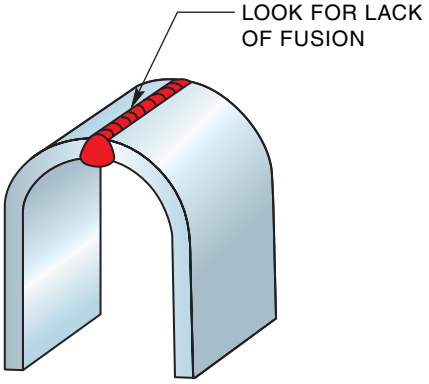

Figure 5.30
Bend the 1-in. (25-mm) strip of butt joint backward and look at the root for 100% penetration

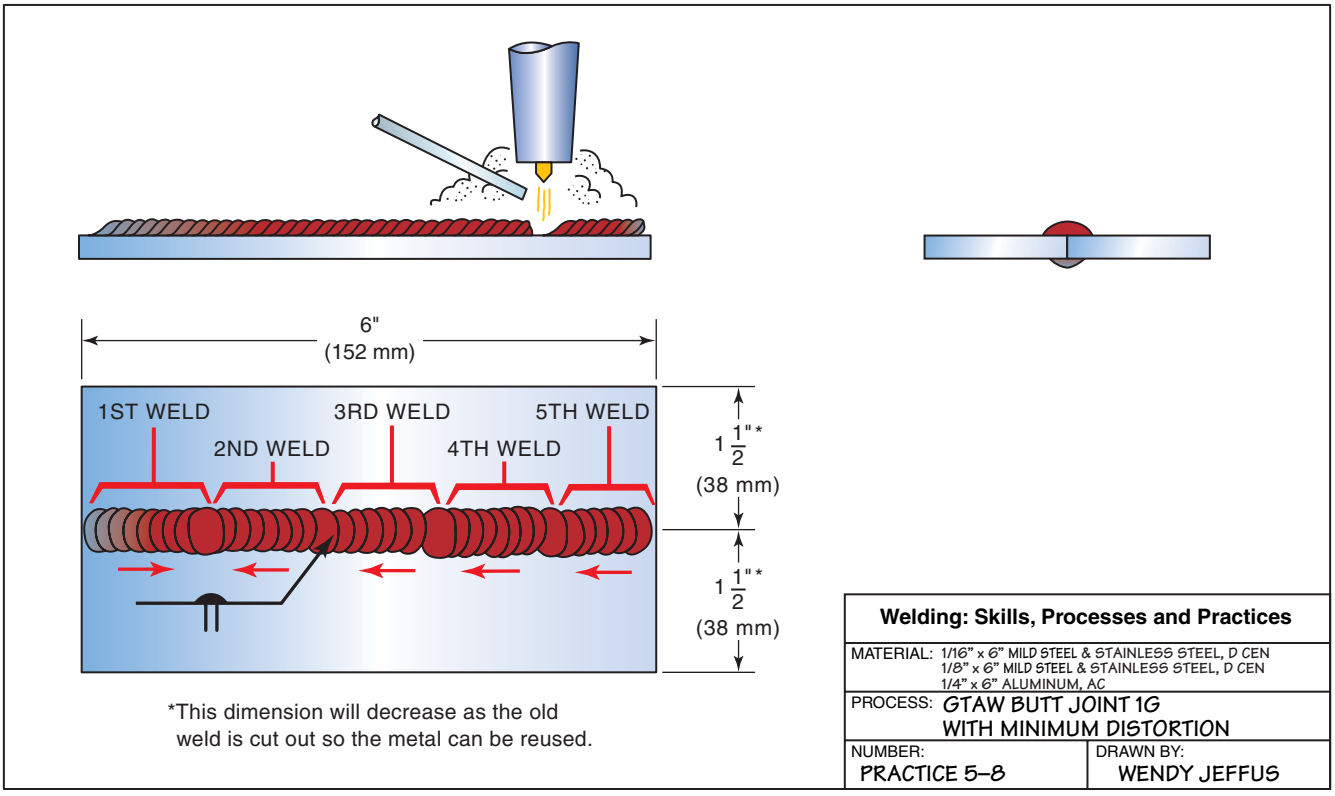

Figure 5.31
Square butt joint in the flat position with minimum distortion

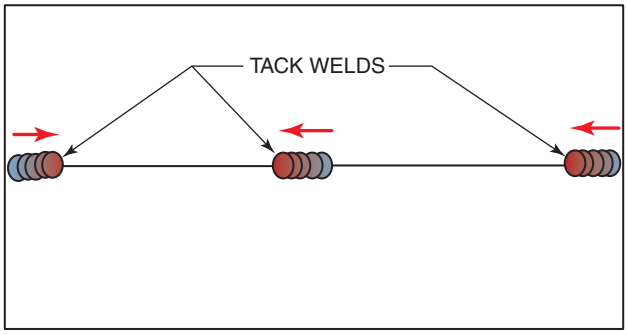

Figure 5.32
Tack welds on a butt joint

PRACTICE 5-9

Lap Joint, 1F Position, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

Module 9
Key Indicator 1, 2

Using the same equipment and materials as listed in Practice 5-4, you will weld a lap joint in the flat position, **Figure 5.33.**

Place the two pieces of metal flat on the table with an overlap of 1/4 in. (6 mm) to 3/8 in. (10 mm). Hold the pieces of metal tightly together and tack weld them as shown in **Figure 5.34** and **Figure 5.35.** Starting at one end, make a uniform fillet weld along the joint. Both sides of the joint can be welded.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

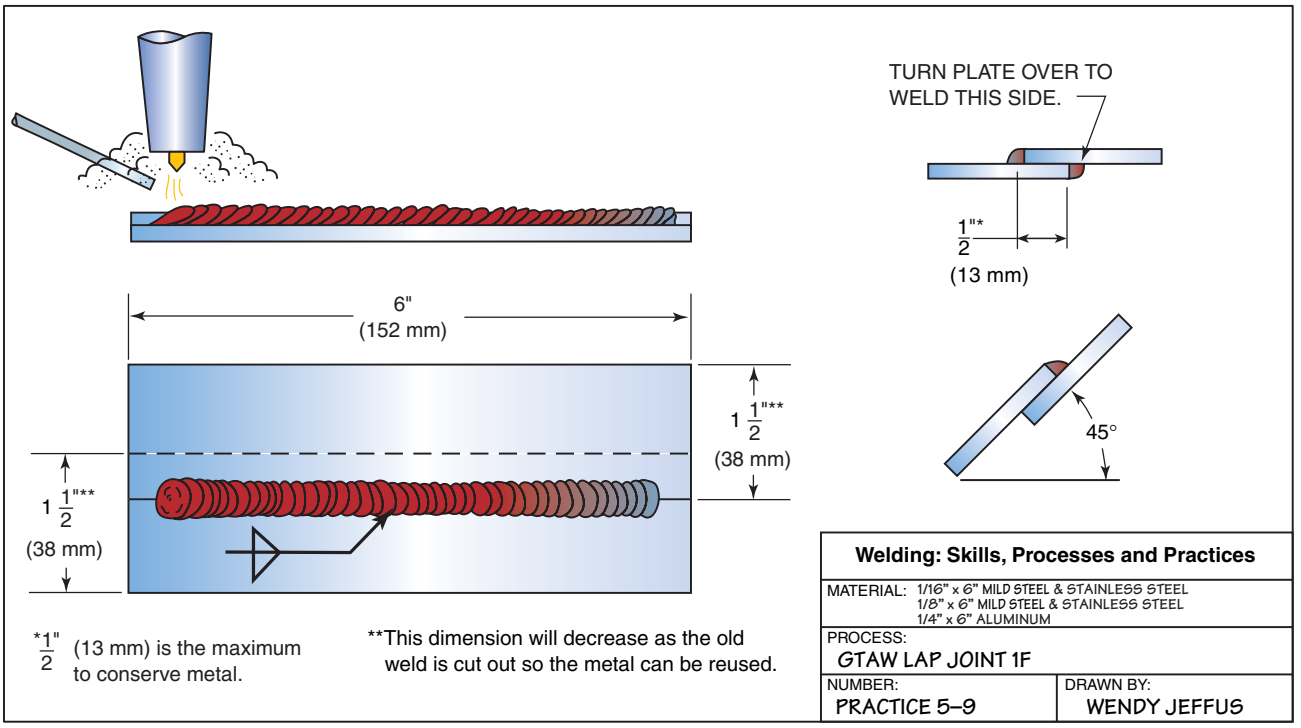

Figure 5.33
Lap joint in the flat position

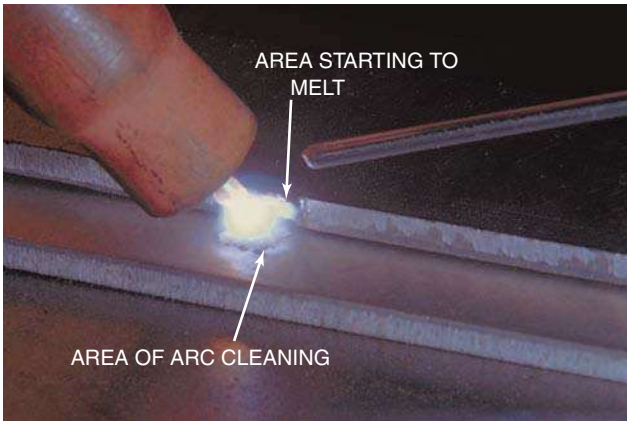

Figure 5.34
Be sure both the top and bottom pieces are melted before adding filler metal
Courtesy of Larry Jeffus

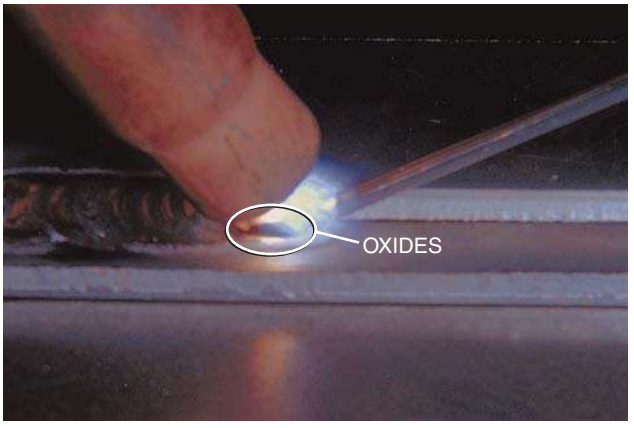

Figure 5.35
Oxides form during tack welding. Do not complete the tack welds. These oxides will become part of the finished weld if the tack is completed.
Courtesy of Larry Jeffus

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-10

Lap Joint, 1F Position, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will make a fillet weld on one side of a lap joint and test it for 100% root penetration, **Figure 5.36, Figure 5.37,** and **Figure 5.38.** After the weld is completed, shear out strips 1 in. (25 mm) wide on one side of the joint and bend-test them as shown in **Figure 5.39.**

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

Module 9
Key Indicator 1, 2

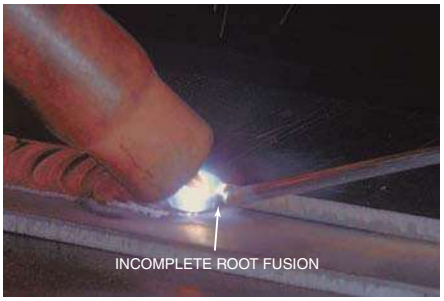

Figure 5.36
A notch indicates that the root was not properly melted and fused
Courtesy of Larry Jeffus

Figure 5.37
Watch the leading edge of the molten weld pool to ensure that there is complete root fusion
Courtesy of Larry Jeffus

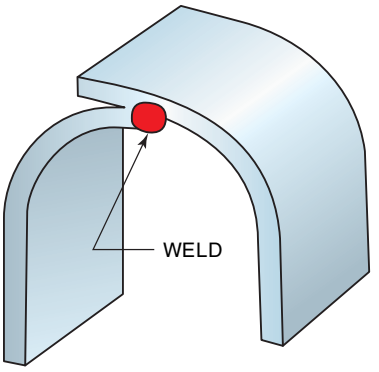

Figure 5.38
Bend the 1-in (25-mm) strip of lap joint backward and look at the root for 100% penetration

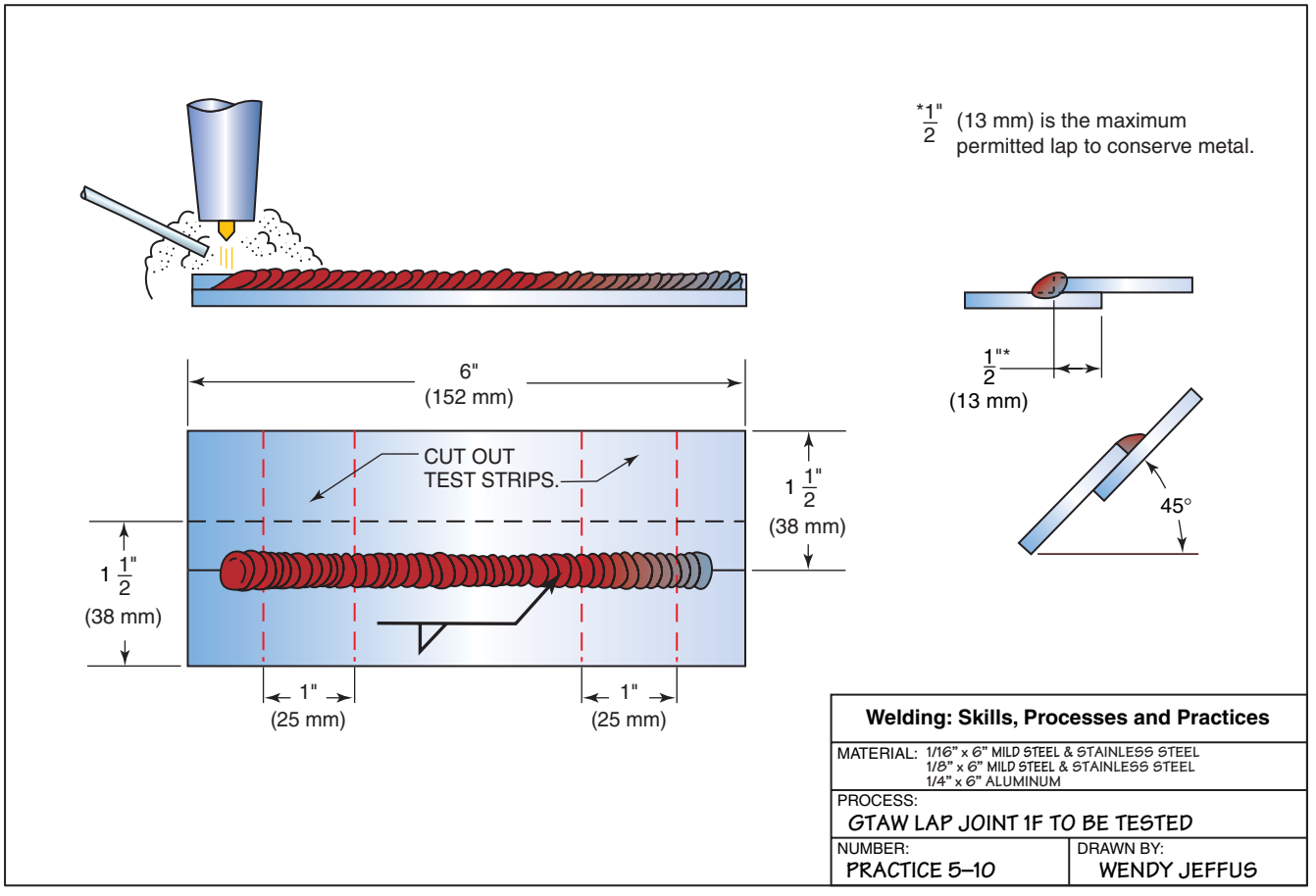

Figure 5.39
Remove strips and test for root fusion

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-11

Tee Joint, 1F Position, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will weld a tee joint in the flat position, **Figure 5.40.**

Place one of the pieces of metal flat on the table and hold or brace the other piece of metal horizontally on it. Tack weld both ends of the plates together, **Figure 5.41.** Set up the plates in the flat position and add two or three more tack welds to the joint as required, **Figure 5.42.**

On the metal that is 1/16 in. (1.5 mm) thick, it may not be possible to weld both sides, but on thicker material a fillet weld can usually be made on both sides. The exception to this is if carbide precipitation occurs on the stainless steel during welding.

Starting at one end, make a uniform weld, adding filler metal as needed.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

Module 9
Key Indicator 1, 2

PRACTICE 5-12

Tee Joint, 1F Position, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will weld a tee joint and test it for 100% root penetration, **Figure 5.43.** After

Module 1
Key Indicator 1, 3, 4

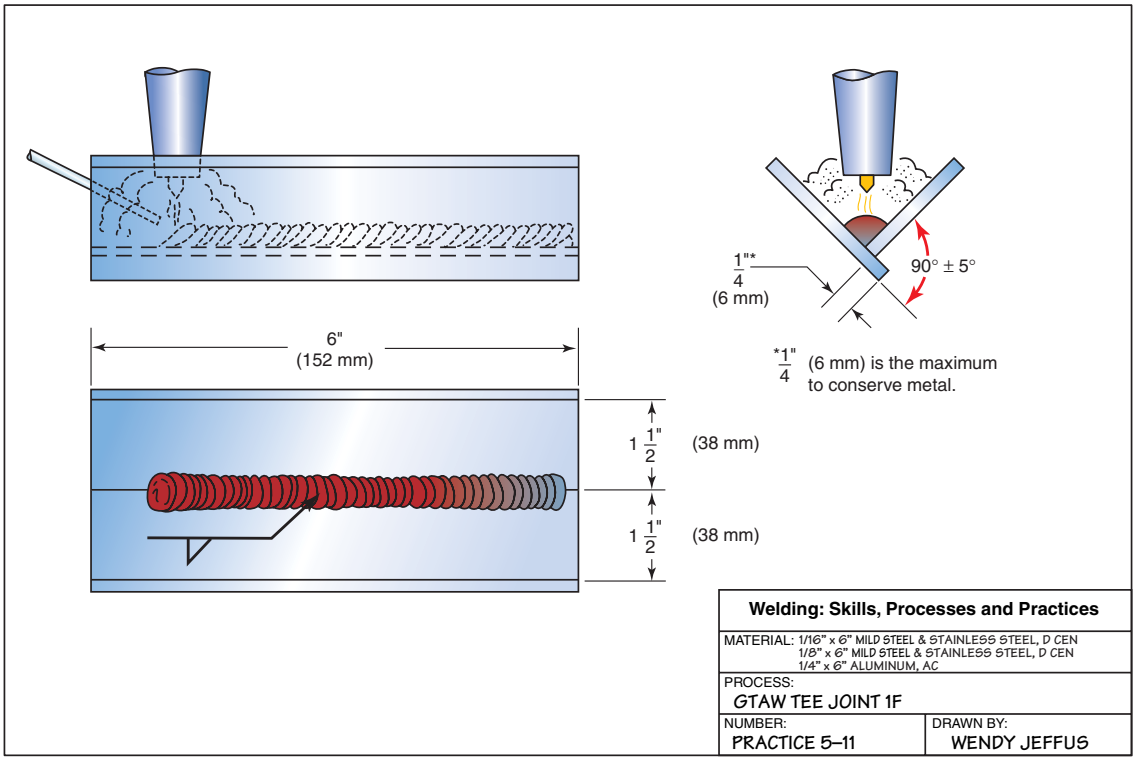

Figure 5.40
Tee joint in the flat position

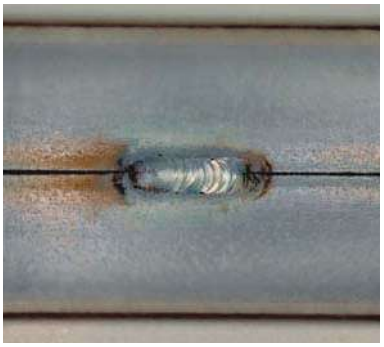

Figure 5.41
Tack weld on a tee joint
Courtesy of Larry Jeffus

Figure 5.42
Keep the tack welds small so that they will not affect the weld
Courtesy of Larry Jeffus

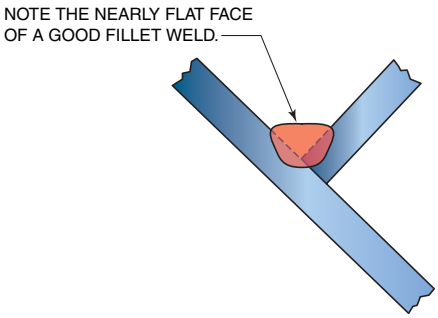

Figure 5.43
100% weld penetration

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

Module 9
Key Indicator 1, 2

the weld is completed, cut or shear out strips 1 in. (25 mm) wide and bend-test them as shown in **Figure 5.44** and **Figure 5.45.**

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

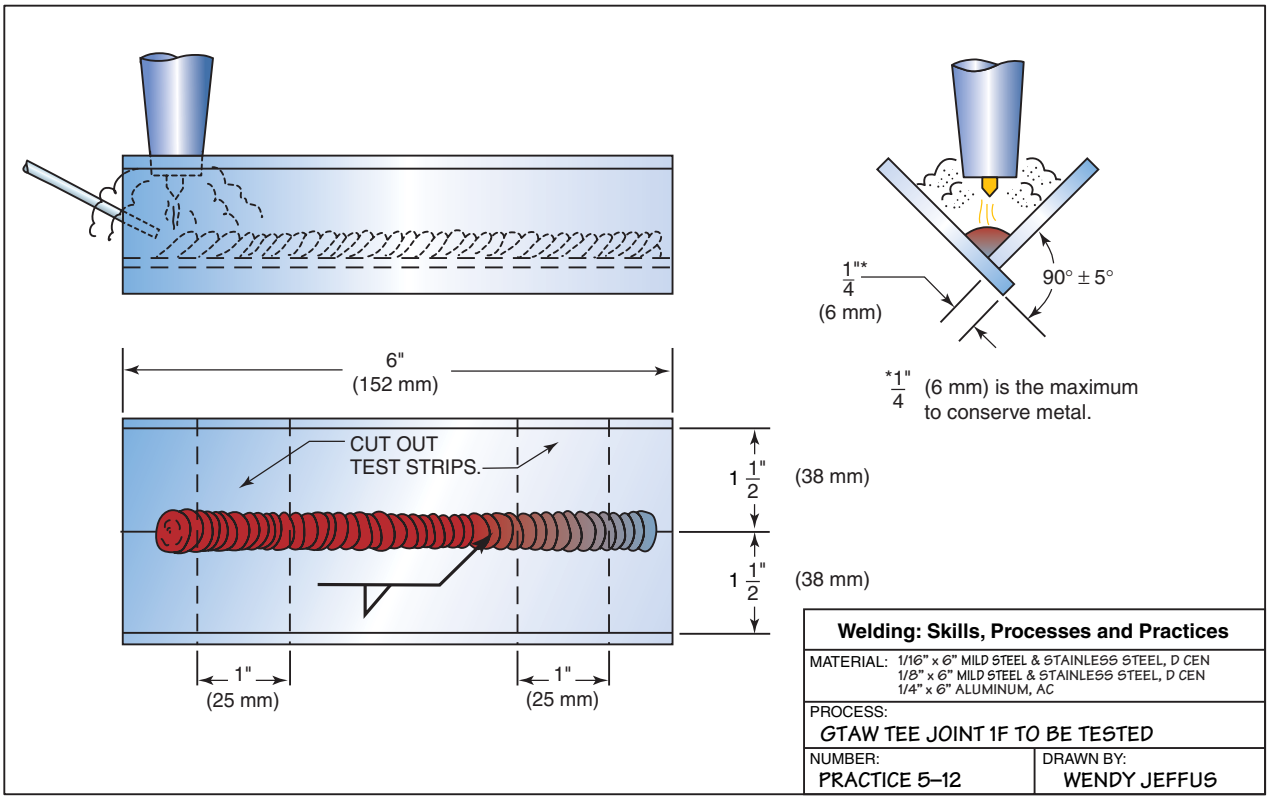

Figure 5.44
Tee joint in the flat position to be tested

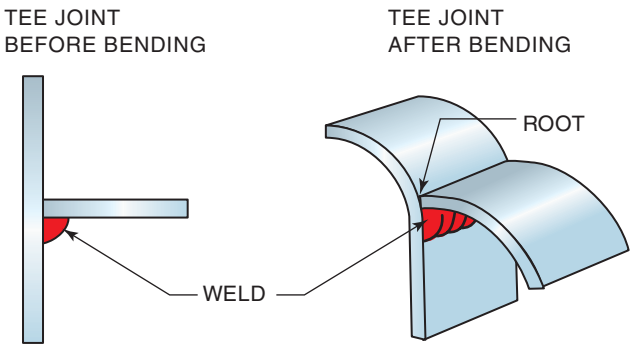

Figure 5.45
Bend the 1-in. (25-mm) strip of the tee joint backward and look at the root for 100% penetration

PRACTICE 5-13

Stringer Bead at a 45° Vertical Angle, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials listed in Practice 5-4, you will make a stringer bead in the vertical up position.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

- Starting at the bottom and welding in an upward direction, add the filler metal to the top edge of the weld pool and move the torch in a circle or "C" pattern, **Figure 5.46.** If the weld pool size starts to in-

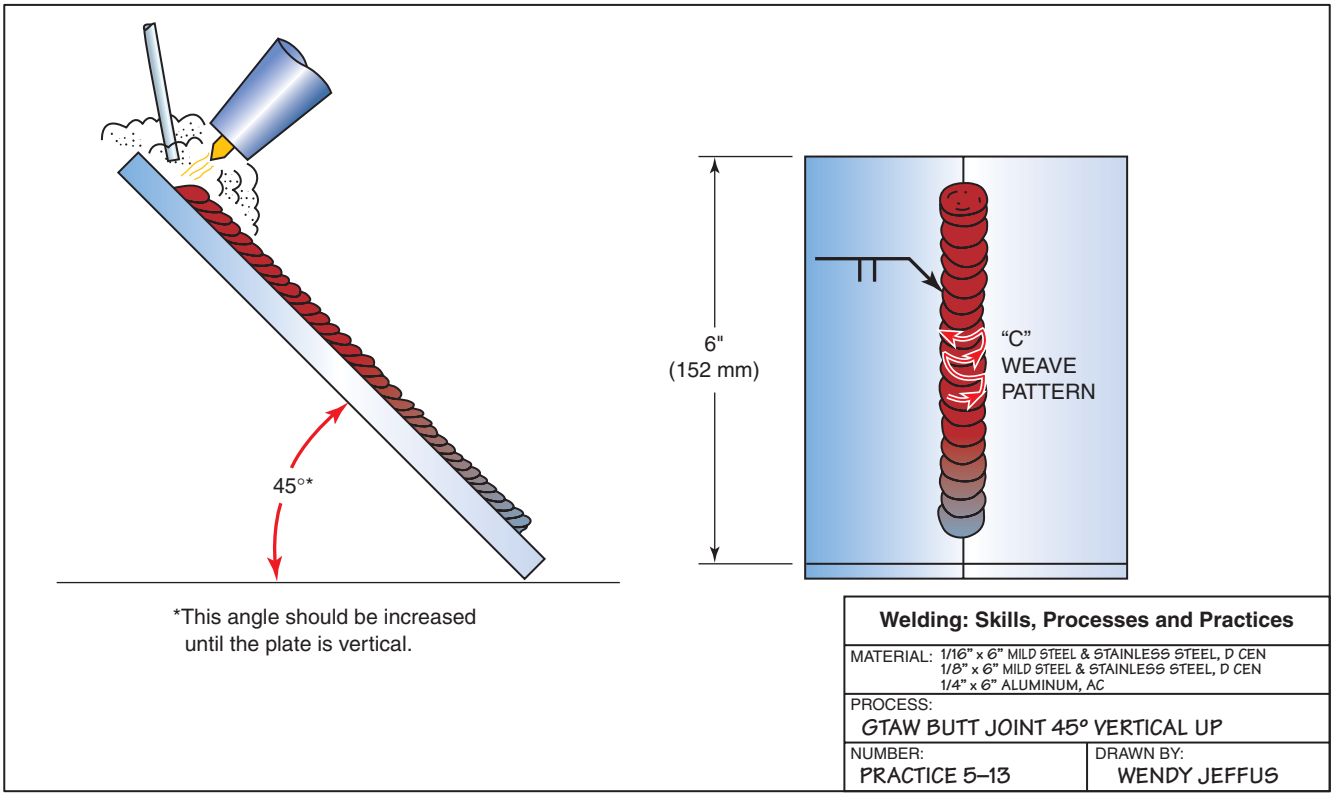

Figure 5.46
45° vertical up

Module 7
Carbon Steel
Key Indicator 4
Austenitic Stainless Steel
Key Indicator 9
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

crease, the "C" pattern can be increased in length or the power can be decreased.

- Watch the weld pool and establish a rhythm of torch movement and addition of rod to keep the weld uniform.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 4
Austenitic Stainless Steel
Key Indicator 9
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

PRACTICE 5-14

Stringer Bead, 3G Position, Using Mild Steel, Stainless Steel, Aluminum

Repeat Practice 5-13. Gradually increase the angle as you develop skill until the weld is being made in the vertical up position, **Figure 5.47.** Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

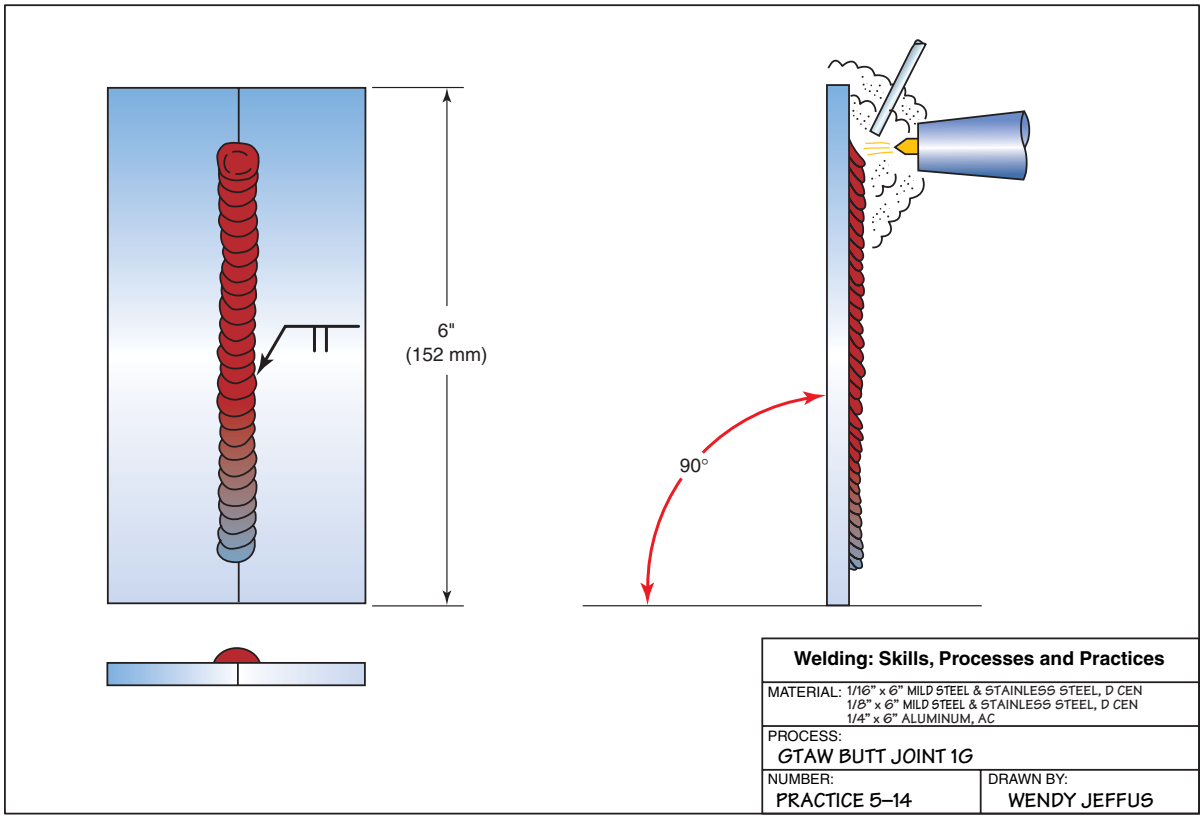

Figure 5.47
Vertical up position

PRACTICE 5-15

Butt Joint at a 45° Vertical Angle, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will weld a butt joint in the vertical up position.

After tack welding the plates together, start the weld at the bottom and weld in an upward direction. The same rhythmic torch and rod movement practiced for the 45° stringer bead should be used to control the weld.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 16

Module 9
Key Indicator 1, 2

PRACTICE 5-16

Butt Joint, 3G Position, Using Mild Steel, Stainless Steel, Aluminum

Repeat Practice 5-15. Gradually increase the plate angle after each weld. As you develop skill, continue increasing the angle until the weld is being made in the vertical up position. Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 16

Module 9
Key Indicator 1, 2

PRACTICE 5-17

Butt Joint, 3G Position, with 100% Penetration, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Repeat Practice 5-16. Make the needed changes in the root opening to allow 100% penetration, **Figure 5.48.** It may be necessary to provide a backing gas to protect the root from atmospheric contamination. After the weld is completed, visually inspect it for uniformity and defects. Then shear out strips 1 in. (25 mm) wide and bend-test them.

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 16

Module 9
Key Indicator 1, 2

PRACTICE 5-18

Lap Joint at a 45° Vertical Angle, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will make a vertical up fillet weld on a lap joint.

Module 1
Key Indicator 1, 3, 4

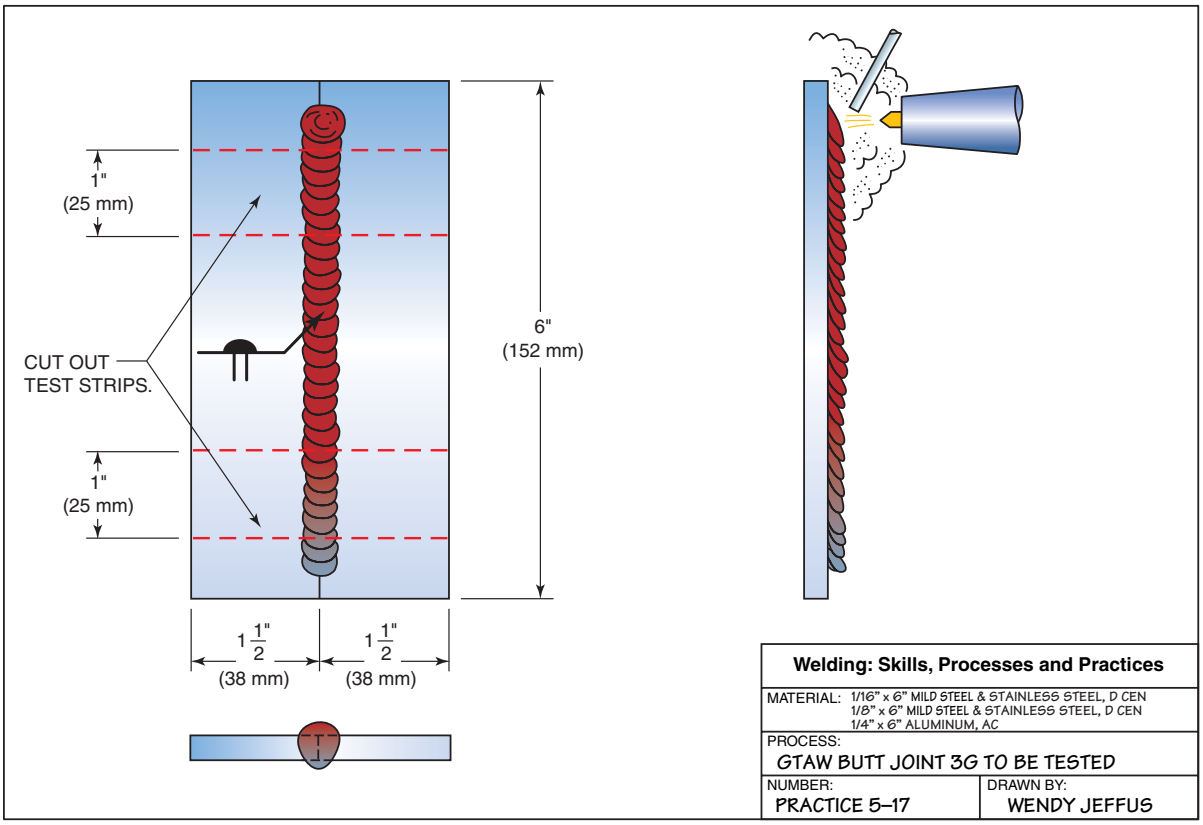

Figure 5.48
Vertical up butt joint with 100% penetration. Strips are to be cut for testing.

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

Module 9
Key Indicator 1, 2

After tack welding the plates together, start the weld at the bottom and weld in an upward direction. It is important to maintain a uniform weld rhythm so that a nice-looking weld bead is formed. It may be necessary to move the torch in and around the base of the weld pool to ensure adequate root fusion, **Figure 5.49.** The filler metal should be added along the top edge of the weld pool near the top plate.

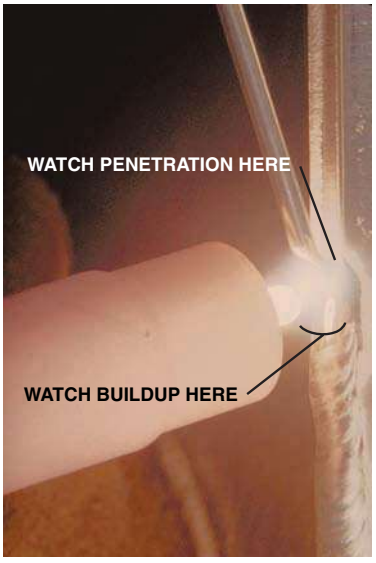

Figure 5.49
Lapjoint
Courtesy of Larry Jeffus

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-19

Lap Joint, 3F Position, Using Mild Steel, Stainless Steel, Aluminum

Repeat Practice 5-18. Gradually increase the plate angle after each weld as you develop your skill. Increase the angle until the weld is being made in the vertical up position, **Figure 5.50.**

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

Module 9
Key Indicator 1, 2

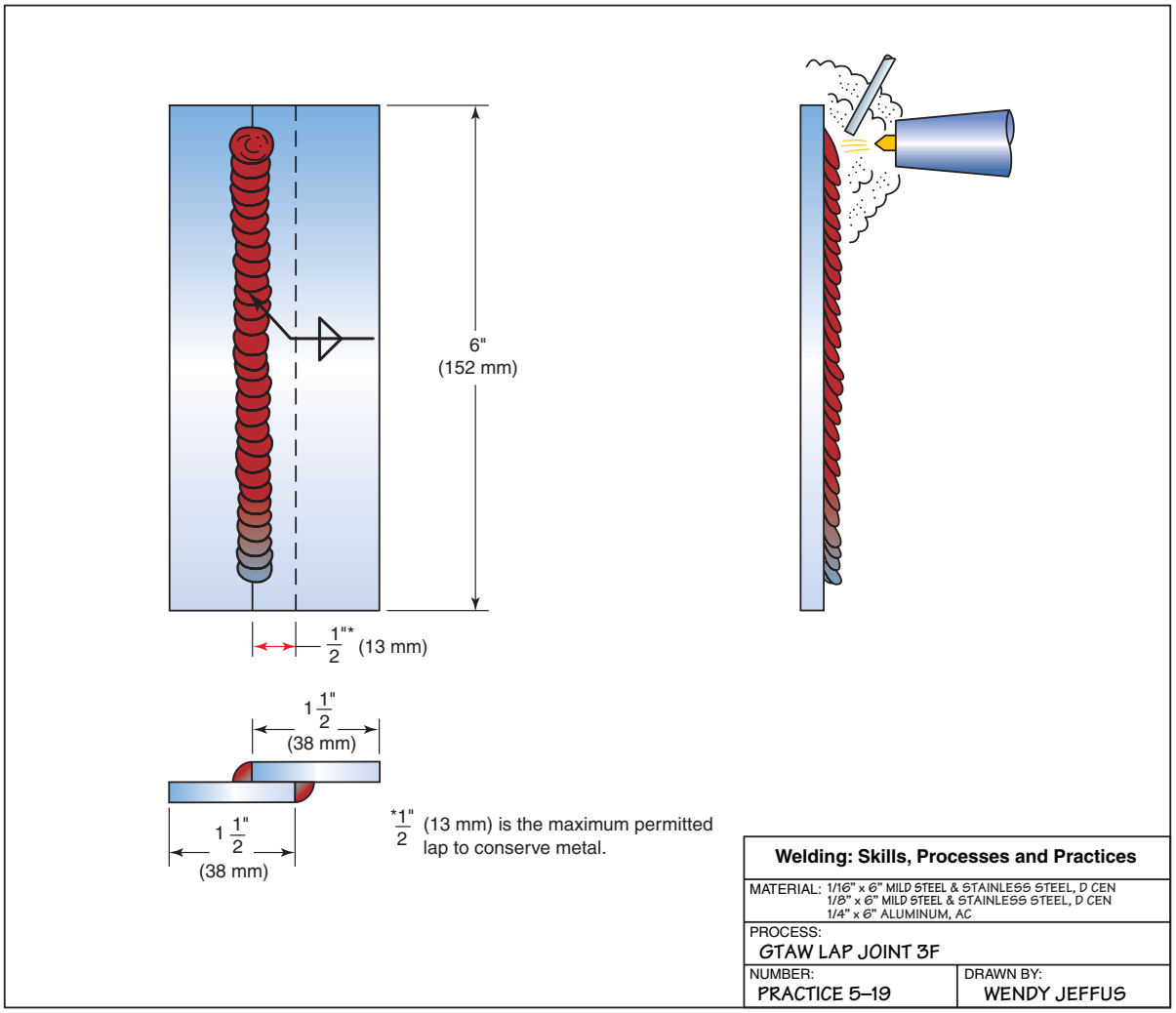

Figure 5.50
Vertical up lap joint

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

PRACTICE 5-20

Lap Joint, 3F Position, with 100% Root Penetration, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Repeat Practice 5-19 and make the fillet weld in a lap joint with 100% root penetration. After the weld is completed, visually inspect it for uniformity and defects before shearing out strips 1 in. (25 mm) wide and bend-testing them.

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

Module 9
Key Indicator 1, 2

PRACTICE 5-21

Tee Joint at a 45° Vertical Angle, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials listed in Practice 5-4, you will make a vertical up fillet weld on a tee joint.

- After tack welding the plates together, start the weld at the bottom and weld in an upward direction. The edge of the side plate, **Figure 5.51,** will heat up more quickly than the back plate. This rapid heating often leads to undercutting along this edge of the weld.
- To control undercutting, keep the arc on the back plate and add the filler metal to the weld pool near the side plate.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

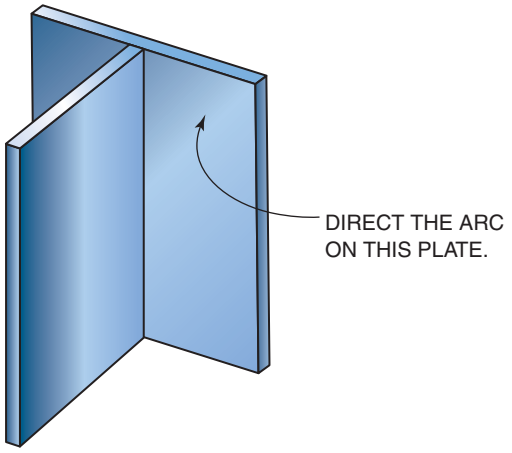

Figure 5.51
The edge of the intersecting plate will heat up faster than the base plate if the heat is not directed away from it

PRACTICE 5-22

Tee Joint, 3F Position, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

Repeat Practice 5-21. Gradually increase the plate angle after each weld as you develop your skill. Increase the angle until the weld is being made in the vertical up position.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-23

Tee Joint, 3F Position, with 100% Root Penetration, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

Repeat Practice 5-21 and make the fillet weld in a lap joint with 100% root penetration. After the weld is completed, visually inspect it for uniformity and defects before shearing out strips 1 in. (25 mm) wide and bend-testing them.

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-24

Stringer Bead at a 45° Reclining Angle, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 16

Module 9
Key Indicator 1, 2

Using the same equipment and materials as listed in Practice 5-4, you will make a weld bead on a plate at a 45° reclining angle, **Figure 5.52.** Add the filler metal along the top leading edge of the weld pool. Surface tension will help hold the weld pool on the top if the bead is not too large. The weld should be uniform in width and reinforcement.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-25

Stringer Bead, 2G Position, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Repeat Practice 5-24. Gradually increase the plate angle as you develop your skill until the weld is being made in the horizontal position on a vertical plate.

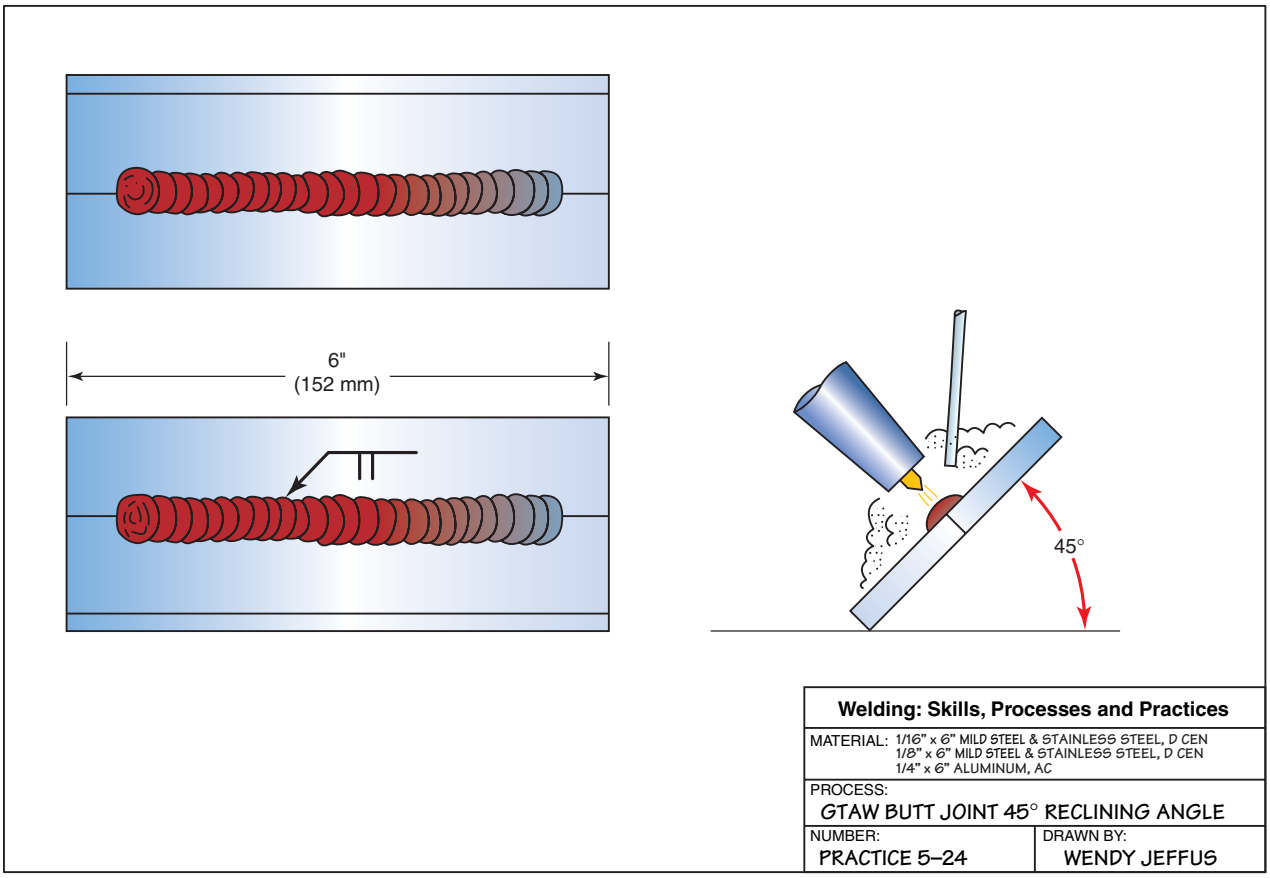

Figure 5.52
45° reclining angle

Module 7
Carbon Steel
Key Indicator 3, 4
Austenitic Stainless Steel
Key Indicator 8, 9
Aluminum
Key Indicator 13, 14

Module 9
Key Indicator 1, 2

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-26

Butt Joint, 2G Position, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 16

Module 9
Key Indicator 1, 2

Using the same equipment and materials as listed in Practice 5-4, you will weld a butt joint in the horizontal position.

The welding techniques are the same as those used in Practice 5-25. Add the filler metal to the top plate, and keep the bead size small so it will be uniform.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-27

Butt Joint, 2G Position, with 100% Penetration, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Repeat Practice 5-26. It may be necessary to increase the root opening to ensure 100% penetration. A backing gas may be required to prevent atmospheric contamination. Using a "J" weave pattern will help to maintain a uniform weld bead. After the weld is completed, visually inspect it for uniformity and defects. Then shear out strips 1 in. (25 mm) wide and bend-test them.

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 11
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

PRACTICE 5-28

Lap Joint, 2F Position, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, make a horizontal fillet weld on a lap joint.

- After tack welding the plates together, start the weld at one end. The bottom plate will act as a shelf to support the molten weld pool, **Figure 5.53.**
- Add the filler metal along the top edge of the weld pool to help control undercutting.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 14, 15

Module 9
Key Indicator 1, 2

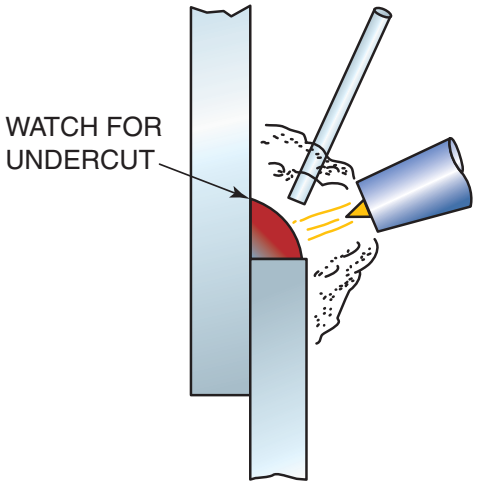

Figure 5.53
Horizontal lap joint

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

Module 9
Key Indicator 1, 2

PRACTICE 5-29

Lap Joint, 2F Position, with 100% Root Penetration, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Repeat Practice 5-28. Be sure the weld is penetrating the root 100%. After the weld is completed, visually inspect it for uniformity and defects before shearing out strips 1 in. (25 mm) wide and bend-testing them.

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 10
Aluminum
Key Indicator 15

PRACTICE 5-30

Tee Joint, 2F Position, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will make a horizontal fillet weld on a tee joint.

After tack welding the plates together, start the weld at one end. The bottom plate will act as a shelf to support the molten weld pool, **Figure 5.54.** As with the horizontal lap joint, add the filler metal along the top leading edge of the weld pool. This will help control undercut.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

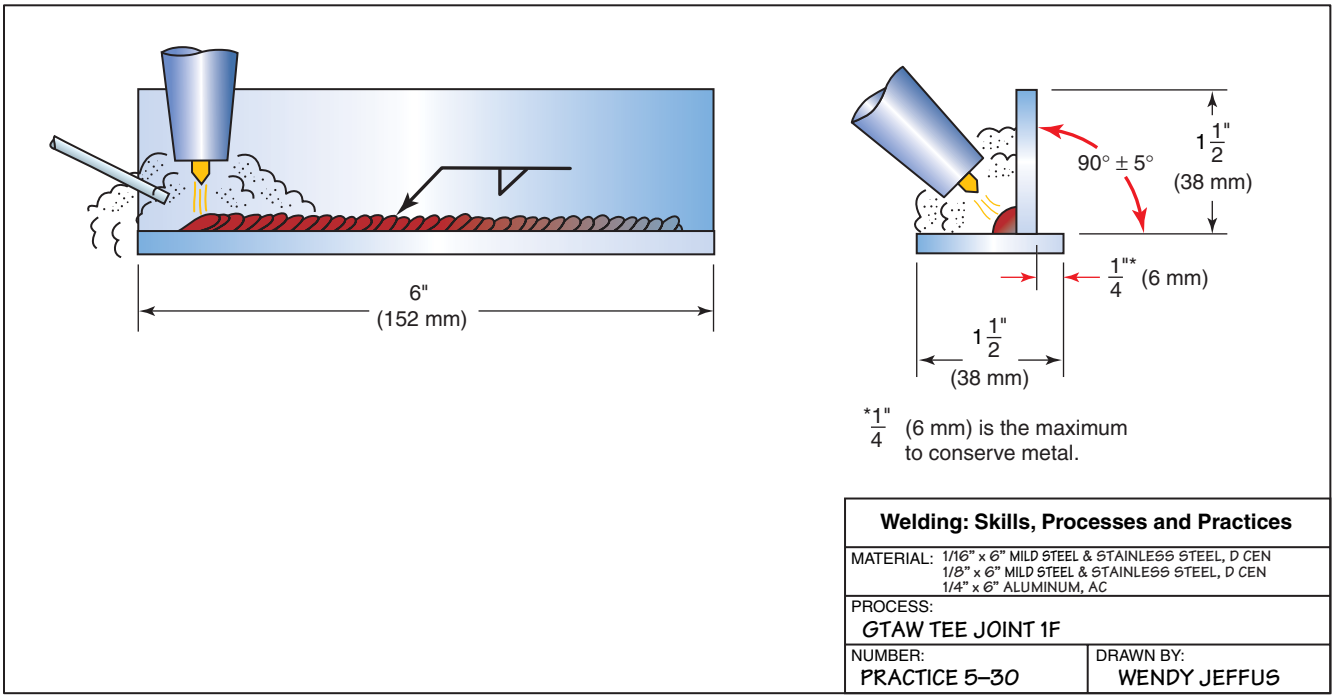

Figure 5.54
Tee joint in the horizontal position

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 9
Key Indicator 1, 2

PRACTICE 5-31

Tee Joint, 2F Position, with 100% Root Penetration, to Be Tested, Using Mild Steel, Stainless Steel, Aluminum

Repeat Practice 5-30. Be sure the weld is penetrating the root 100%. After the weld is completed, visually inspect it for uniformity and defects before shearing out strips 1 in. (25 mm) wide and bend-testing them.

Repeat each weld until all have 100% root penetration. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 4
Austenitic Stainless Steel
Key Indicator 9
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

PRACTICE 5-32

Stringer Bead, 4G Position, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will make a weld bead on a plate in the overhead position.

The surface tension on the molten metal will hold the welding bead on the plate providing that it is not too large. A wide weld with little buildup will be easier to control and less likely to undercut along the edge.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 6
Austenitic Stainless Steel
Key Indicator 9, 10
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

PRACTICE 5-33

Butt Joint, 4G Position, Using Mild Steel, Stainless Steel, Aluminum

Using the same equipment and materials as listed in Practice 5-4, you will weld a butt joint in the overhead position.

The same techniques used to make the stringer beads in Practice 5-32 are also used with the butt joint. The size of the bead should be kept small enough so that you can control the weld. Add the filler metal along the leading edge of the weld pool, **Figure 5.55.** The completed weld should be uniform and free from defects.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 9
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

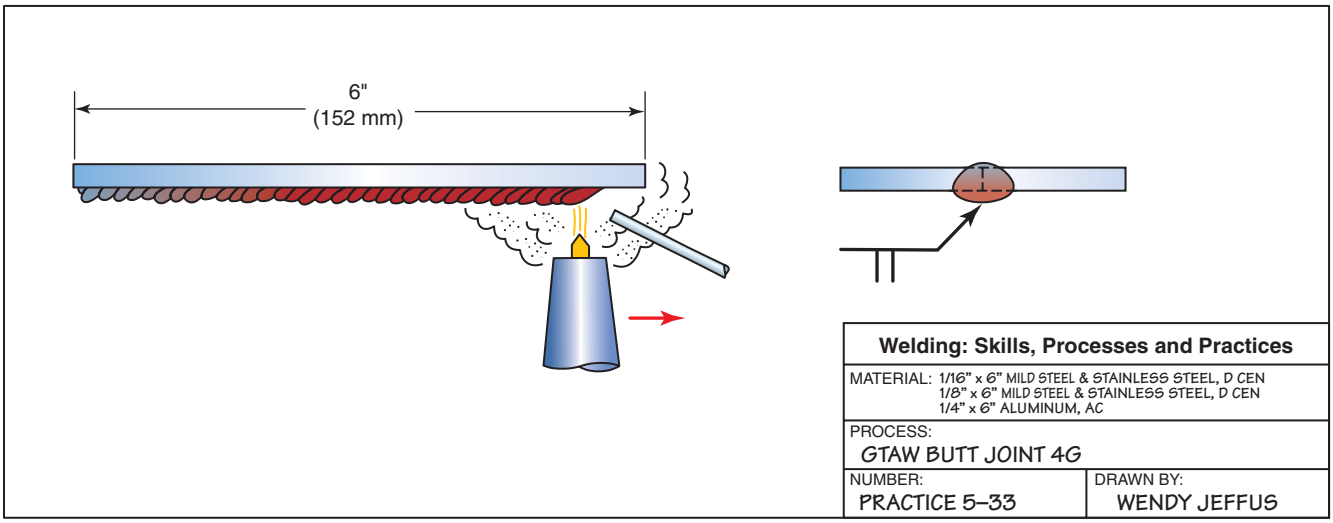

Figure 5.55
Overhead butt joint

PRACTICE 5-34

Lap Joint, 4F Position, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Carbon Steel
Key Indicator 5
Austenitic Stainless Steel
Key Indicator 9
Aluminum
Key Indicator 14

Module 9
Key Indicator 1, 2

Using the same equipment and materials as listed in Practice 5-4, you will make a fillet weld on a lap joint in the overhead position.

The major concentration of heat and filler metal should be on the top plate. Gravity and an occasional sweep of the torch along the bottom plate will pull the weld pool down. Undercutting along the top edge of the weld can be controlled by putting most of the filler metal along the top edge. The completed weld should be uniform and free from defects.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-35

Tee Joint, 4F Position, Using Mild Steel, Stainless Steel, Aluminum

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Key Indicator 1, 2
Carbon Steel
Key Indicator 3, 4, 5, 6, 7

Module 9
Key Indicator 1, 2

Using the same equipment and materials as listed in Practice 5-4, you will make a fillet weld on a tee joint in the overhead position.

The same techniques used to make the overhead lap weld in Practice 5-34 are used with the tee joint. As with the lap joint, most of the heat and filler metal should be concentrated on the top plate. A "J" weave pattern will help pull down any needed metal to the side plate. The completed weld should be uniform and free from defects.

Repeat the process using all thicknesses of metal until you can consistently make the weld visually defect free. Turn off the welding machine, shielding gas, and cooling water, and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-36

Gas Tungsten Arc Welding (GTAW) on Plain Carbon Steel Workmanship Sample

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 5-36. Date:

Title:

Welding GTAW of sheet to sheet.

Scope:

This procedure is applicable for square groove and fillet welds within the range of 18 gauge through 10 gauge.

Welding may be performed in the following positions: 1G and 2F.

Base Metal:

The base metal shall conform to carbon steel M-1, Group 1.
Backing material specification: none .

Filler Metal:

The filler metal shall conform to AWS specification no. 1/16 in. (2 mm) to 3/32 in. (2.4 mm) diameter. E70S-3 from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

Electrode:

The tungsten electrode shall conform to AWS specification no. EWTh-2, EWCe-2, or EWLa from AWS specification A5.12. The tungsten diameter shall be 1/8 in. (3.2 mm) maximum.

The tungsten end shape shall be tapered at two to three times its length to its diameter.

Shielding Gas:

The shielding gas, or gases, shall conform to the following compositions and purity: welding grade argon.

Joint Design and Tolerances:

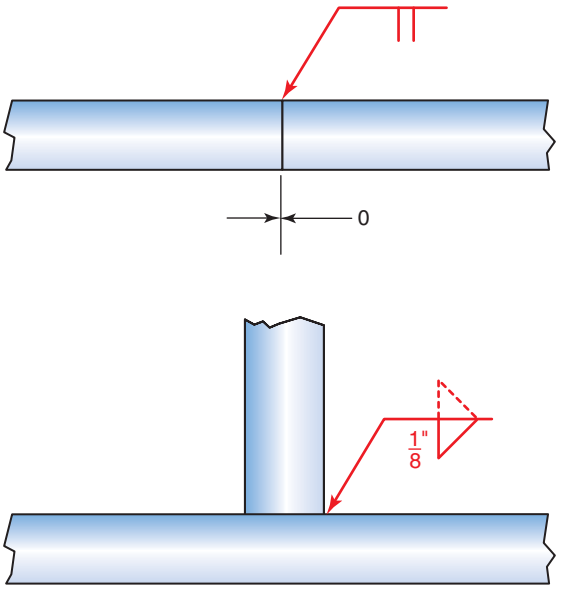

Preparation of Base Metal:

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The joint face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:

The current shall be direct current electrode negative (DCEN). The base metal shall be on the positive side of the line.

Metal Specifications		Gas Flow			Nozzle Size in. (mm)	Amperage Min. Max.
Thickness	Diameter of E70S-3*	Rates cfm) (L/min)	Flow Times Preflow	Postflow		
18 ga	1/16 in. (2 mm)	15 to 20 (7 to 9)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	45 to 65
17 ga	1/16 in. (2 mm)	15 to 20 (7 to 9)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	45 to 70
16 ga	1/16 in. (2 mm)	15 to 20 (7 to 9)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	50 to 75
15 ga	1/16 in. (2 mm)	15 to 20 (7 to 9)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	55 to 80
14 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	60 to 90
13 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	60 to 100
12 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	60 to 110
11 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	65 to 120
10 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	70 to 130

*Other E70S-X filler metal may be used.

Preheat:

The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:

None

Safety:

Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:

TACK WELDS: With the parts securely clamped in place with the correct root gap, the tack welds are to be performed. Holding the electrode so that it is very close to the root face but not touching, slowly increase the current until the arc starts and a molten weld pool is formed. Add filler metal as required to maintain a slightly convex weld face and a flat or slightly concave root face. When it is time to end the tack weld, lower the current slowly so that the molten weld pool can be tapered down in size. When all tack welds are complete, allow the parts to cool as needed before assembling the remaining parts. Repeat the tack welding procedure until the entire part is assembled.

SQUARE GROOVE AND FILLET WELDS: Holding the electrode so that it is very close to the metal surface but not touching, slowly increase the

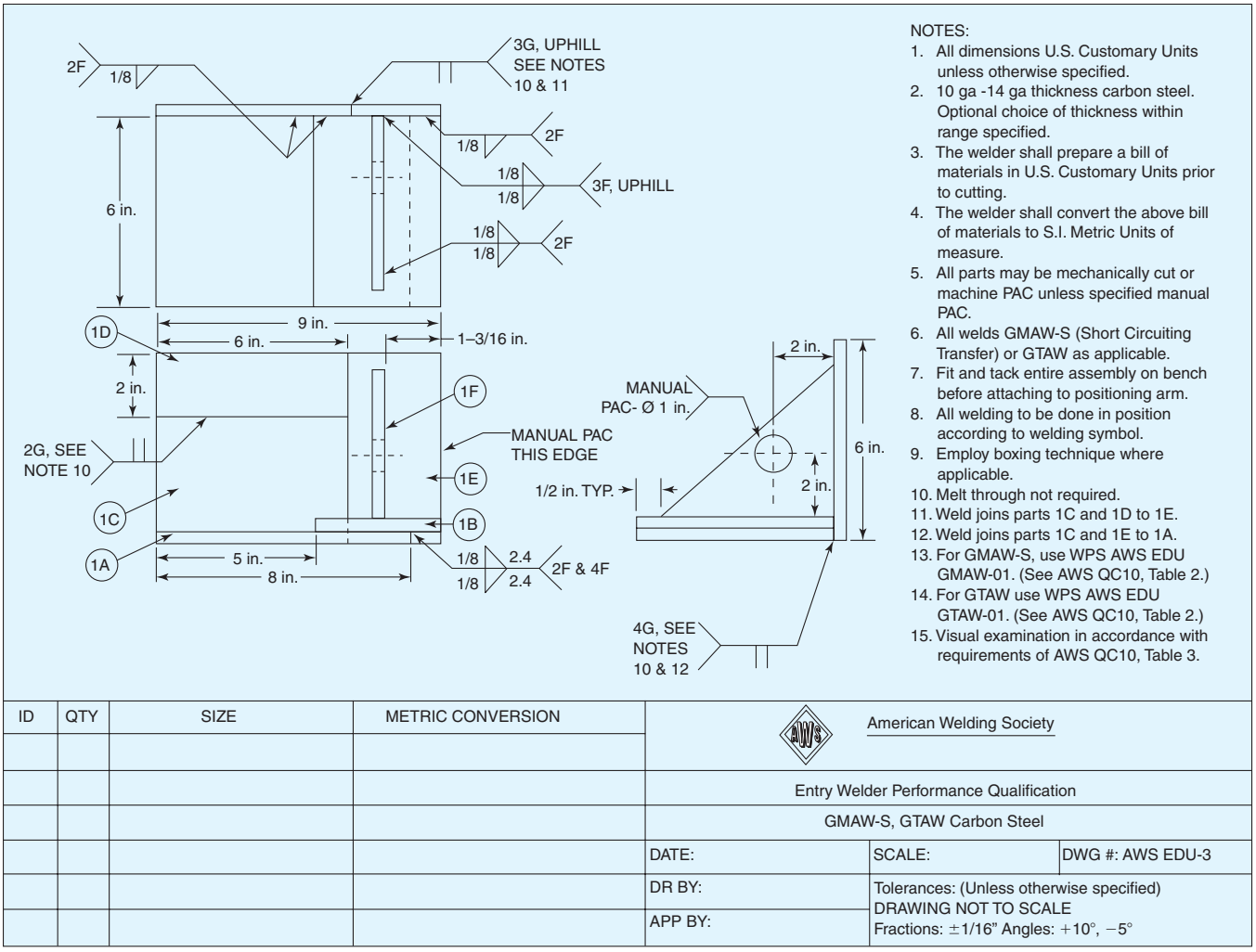

Figure 5.56
GTAW Plain Carbon Steel Workmanship Qualification Test
Courtesy of the American Welding Society

current until the arc starts and a molten weld pool is formed. As the weld progresses, add filler metal as required to maintain a flat or slightly convex weld face. If it is necessary to stop the weld or to reposition yourself or if the weld is completed, the current must be lowered slowly so that the molten weld pool can be tapered down in size.

Interpass Temperature:

The plate should not be heated to a temperature higher than 120°F (49°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:

Recleaning may be required if the parts or filler metal becomes contaminated or reoxides to a degree that the weld quality will be affected. Reclean using the same procedure used for the original metal preparation. Any slag must be cleaned off between passes.

Visual Inspection:

Visual inspection criteria for entry welders*:

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration groove welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm).
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.
6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

Sketches:

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-37

Gas Tungsten Arc Welding (GTAW) on Stainless Steel Workmanship Sample

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 5-37. Date:

Module 1
Key Indicator 1, 3, 4

Title:

Welding GTAW of sheet to sheet.

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Scope:

This procedure is applicable for square groove and fillet welds within the range of 18 gauge through 10 gauge.

Welding may be performed in the following positions: 1G and 2F.

Module 7
Key Indicator 1, 2
Austenitic Stainless Steel
Key Indicator 8, 9, 10, 11, 12

Base Metal:

The base metal shall conform to austenitic stainless steel M-8 or P-8. Backing material specification: none.

Module 9
Key Indicator 1, 2

Filler Metal:

The filler metal shall conform to AWS specification no. ER3XX from AWS specificationA5.9.ThisfillermetalfallsintoF-numberF-6andA-numberA-8.

Electrode:

The tungsten electrode shall conform to AWS specification no. EWTh-2, EWCe-2 or EWLa from AWS specification A5.12. The tungsten diameter shall be 1/8 in. (3.2 mm) maximum. The tungsten end shape shall be tapered at two to three times its length to its diameter.

Shielding Gas:

The shielding gas, or gases, shall conform to the following compositions and purity: welding grade argon.

*Courtesy of the American Welding Association.

Joint Design and Tolerances:

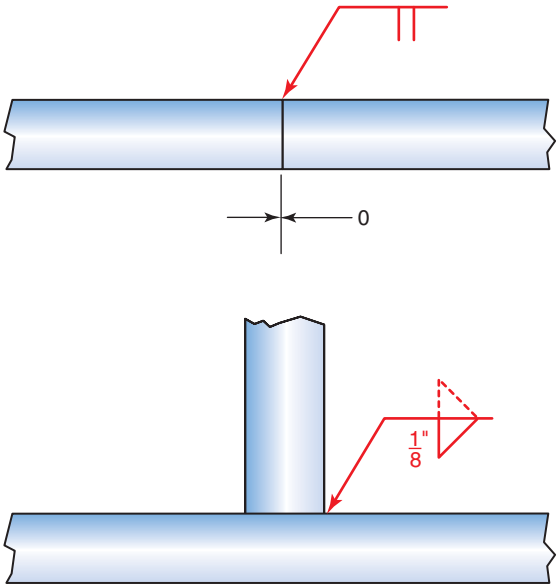

Preparation of Base Metal:

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The joint face and inside and outside plate surface within 1 in. (25 mm) of the joint must be cleaned of slag, oxide, and scale. Cleaning can be mechanical or chemical. Mechanical metal cleaning can be done by grinding, stainless steel wire brushing, scraping, machining, or filing. Chemical cleaning can be done by using acids, alkalies, solvents, or detergents. Cleaning must be done down to bright metal.

Electrical Characteristics:

The current shall be direct current electrode negative (DCEN). The base metal shall be on the positive side of the line.

Metal Specifications		Gas Flow			Nozzle Size in. (mm)	Amperage Min. Max.
Thickness	Diameter of ER3XX*	Rates cfm (L/min	Flow Times Preflow	Postflow		
18 ga	1/16 in. (2 mm)	15 to 20 (7 to 9)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	35 to 60
17 ga	1/16 in. (2 mm)	15 to 20 (7 to 9)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	40 to 65
16 ga	1/16 in. (2 mm)	15 to 20 (7 to 9)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	40 to 75
15 ga	1/16 in. (2 mm)	15 to 20 (7 to 9)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	50 to 80
14 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	50 to 90
13 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	55 to 100
12 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	60 to 110
11 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	65 to 120
10 ga	3/32 in. (2.4 mm)	20 to 25 (9 to 12)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	70 to 130

*Any ER3XX stainless steel A5.9 filler metal may be used.

Preheat:

The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:

None

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
TACK WELDS: With the parts securely clamped in place with the correct root gap, the tack welds are to be performed. Holding the electrode so that it is very close to the root face but not touching, slowly increase the current until the arc starts and a molten weld pool is formed. Add filler metal as required to maintain a slightly convex weld face and a flat or slightly concave root face. When it is time to end the tack weld, lower the current slowly so that the molten weld pool can be tapered down in size. When all tack welds are complete, allow the parts to cool as needed before assembling the remaining parts. Repeat the tack welding procedure until the entire part is assembled.

SQUARE GROOVE AND FILLET WELDS: Holding the electrode so that it is very close to the metal surface but not touching, slowly increase the current until the arc starts and a molten weld pool is formed. As the weld progresses, add filler metal as required to maintain a flat or slightly convex weld face. If it is necessary to stop the weld or to reposition yourself or if the weld is completed, the current must be lowered slowly so that the molten weld pool can be tapered down in size.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (180°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
Recleaning may be required if the parts or filler metal become contaminated or oxidized to a degree that the weld quality will be affected. Reclean using the same procedure used for the original metal preparation.

Visual Inspection:
Visual inspection criteria for entry welders*:

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration groove welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm).
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.
6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

*Courtesy of the American Welding Association.

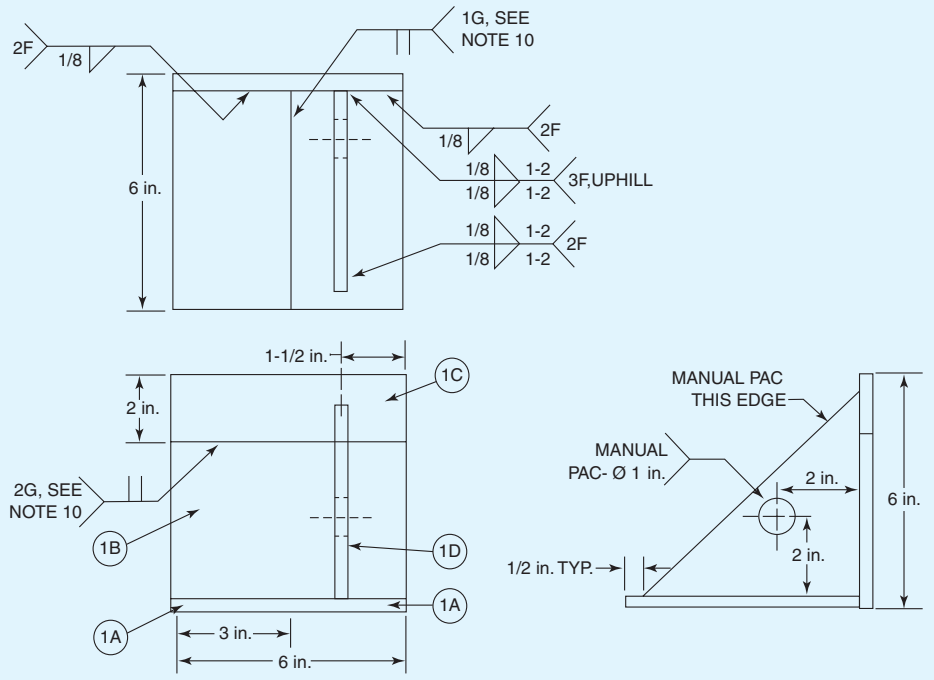

NOTES:
1. All dimensions U.S. Customary Units unless otherwise specified.
2. 10 ga -14 ga thickness austenitic stainless steel. Optional choice of thickness within range specified.
3. The welder shall prepare a bill of materials in U.S. Customary Units prior to cutting.
4. The welder shall convert the above bill of materials to S.I. Metric Units of measure.
5. All parts may be mechanically cut or machine PAC unless specified manual PAC.
6. All welds GMAW.
7. Fit and tack entire assembly on bench before attaching to positioning arm.
8. All welding to be done in position according to welding symbol.
9. Employ boxing technique where applicable.
10. Melt through not required.
11. Use WPS AWS EDU GTAW-04. (See AWS QC10, Table 2.)
12. Visual examination in accordance with requirements of AWS QC10, Table 3.

ID	QTY	SIZE	METRIC CONVERSION	American Welding Society		
				Entry Welder Performance Qualification		
				GTAW Austenitic Stainless Steel		
				DATE:	SCALE:	DWG #: AWS EDU-3
				DR BY:	Tolerances: (Unless otherwise specified) DRAWING NOT TO SCALE Fractions: ±1/16" Angles: +10°, −5°	
				APP BY:		

Figure 5.57
GTAW Stainless Steel Workmanship Qualification Test
Courtesy of the American Welding Society

Sketches:
Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 5-38

Gas Tungsten Arc Welding (GTAW) on Aluminum Workmanship Sample

Module 1
Key Indicator 1, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 6, 7

Module 7
Key Indicator 1, 2
Aluminium
Key Indicator 13, 14, 15, 16, 17

Module 9
Key Indicator 1, 2

Welding Procedure Specification (WPS)
Welding Procedure Specification No.: Practice 5-38. Date:

Title:
Welding GTAW of sheet to sheet.

Scope:
This procedure is applicable for square groove and fillet welds within the range of 18 gauge through 10 gauge.

Welding may be performed in the following positions: 1G and 2F.

Base Metal:
The base metal shall conform to aluminum M-22 or P-22.
Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. ER4043 from AWS specification A5.10. This filler metal falls into F-number F-23 and A-number.

Electrode:
The tungsten electrode shall conform to AWS specification No. EWCe-2, EWZr, EWLa, or EWP from AWS specification A5.12. The tungsten diameter shall be 1/8 in. (3.2 mm) maximum. The tungsten end shape shall be rounded.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity: welding grade argon.

Joint Design and Tolerances:

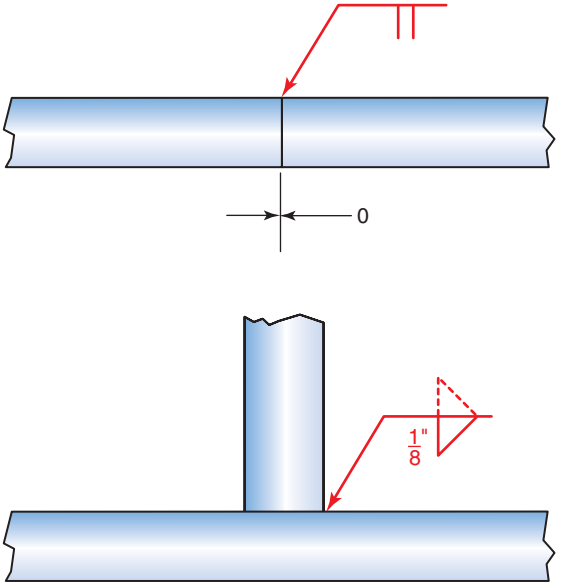

Preparation of Base Metal:
All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The joint face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically or chemically cleaned of oxides. Mechanical cleaning may be done by stainless steel wire brushing, scraping, machining, or filing. Chemical cleaning may be done by using acids, alkalies, solvents, or detergents. Because the oxide layer may reform quickly and affect the weld, welding should be started within 10 minutes of cleaning.

Electrical Characteristics:
The current shall be alternating current high-frequency stabilized (balanced wave preferably). The base metal shall be on the N/A side of the line.

Gas Flow

Preheat:

The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Metal Specifications		Gas Flow			Nozzle Size in. (mm)	Amperage Min. Max.
Thickness	Diameter of ER4043*	Rates cfm (L/min)	Flow Times Preflow	Postflow		
18 ga	3/32 in. (2.4 mm)	20 to 30 (9 to 14)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	40 to 60
17 ga	3/32 in. (2.4 mm)	20 to 30 (9 to 14)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	50 to 70
16 ga	3/32 in. (2.4 mm)	20 to 30 (9 to 14)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	60 to 75
15 ga	3/32 in. (2.4 mm)	20 to 30 (9 to 14)	10 to 15 sec	10 to 25 sec	1/4 to 3/8 (6 to 10)	65 to 85
14 ga	3/32 in. (2.4 mm)	20 to 30 (9 to 14)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	75 to 90
13 ga	1/8 in. (3 mm)	25 to 40 (12 to 19)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	85 to 100
12 ga	1/8 in. (3 mm)	25 to 40 (12 to 19)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	90 to 110
11 ga	1/8 in. (3 mm)	25 to 40 (12 to 19)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	100 to 115
10 ga	1/8 in. (3 mm)	25 to 40 (12 to 19)	10 to 20 sec	10 to 30 sec	3/8 to 5/8 (10 to 16)	100 to 125

*Other aluminum AWS A5.10 filler metal may be used if needed.

Backing Gas:

N/A

Safety:

Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:

The welder's hands or gloves must be clean and oil free to prevent contamination of the metal or filler rods.

TACK WELDS: With the parts securely clamped in place with the correct root gap, the tack welds are to be performed. Holding the electrode so that it is very close to the root face but not touching, slowly increase the current until the arc starts and a molten weld pool is formed. Add filler metal as required to maintain a slightly convex weld face and a flat or slightly concave root face. When it is time to end the tack weld, lower the current slowly so that the molten weld pool can be tapered down in size. When all tack welds are complete, allow the parts to cool as needed before assembling the remaining parts. Repeat the tack welding procedure until the entire part is assembled.

SQUARE GROOVE AND FILLET WELDS: Holding the electrode so that it is very close to the metal surface but not touching, slowly increase the current until the arc starts and a molten weld pool is formed. As the weld progresses, add filler metal as required to maintain a flat or slightly convex weld face. If it is necessary to stop the weld or to reposition yourself or the weld is completed, the current must be lowered slowly so that the molten weld pool can be tapered down in size.

Interpass Temperature:

The plate should not be heated to a temperature higher than 120°F (49°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

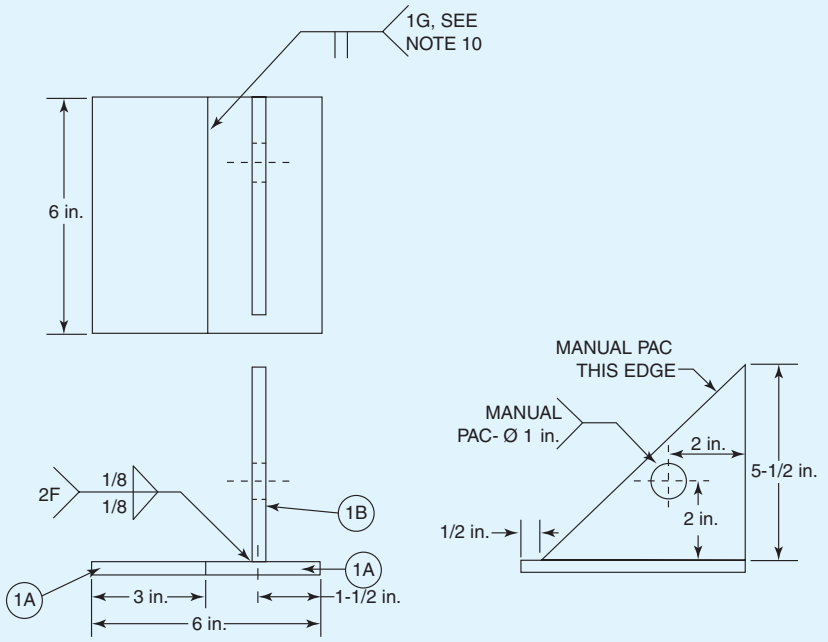

NOTES:
1. All dimensions U.S. Customary Units unless otherwise specified.
2. 10 ga -14 ga thickness aluminium. Optional choice of thickness within range specified.
3. The welder shall prepare a bill of materials in U.S. Customary Units prior to cutting.
4. The welder shall convert the above bill of materials to S.I. Metric Units of measure.
5. All parts may be mechanically cut or machine PAC unless specified manual PAC.
6. All welds GTAW.
7. Fit and tack entire assembly on bench before attaching to positioning arm.
8. All welding to be done in position according to welding symbol.
9. Employ boxing technique where applicable.
10. Melt through not required.
11. Use WPS AWS EDU GTAW-03 for 4000/5000 Series aluminium. (See AWS QC10, Table 2.)
12. Visual examination in accordance with requirements of AWS QC10, Table 3.

ID	QTY	SIZE	METRIC CONVERSION	American Welding Society	
				Entry Welder Performance Qualification	
				GTAW Aluminium	
				DATE:	SCALE: DWG #: AWS EDU-5
				DR BY:	Tolerances: (Unless otherwise specified) DRAWING NOT TO SCALE Fractions: ±1/16" Angles: +10°, −5°
				APP BY:	

Figure 5.58
GTAW Aluminum Workmanship Qualification Test
Courtesy of the American Welding Society

Cleaning:

Recleaning may be required if the parts or filler metal becomes contaminated or oxidizes to a degree that the weld quality will be affected. Reclean using the same procedure used for the original metal preparation.

Visual Inspection:

Visual inspection criteria for entry welders*:

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration groove welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm)
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.

*Courtesy of the American Welding Association.

6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

Sketches:
Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

SUMMARY

One of the most difficult aspects of learning to produce gas tungsten arc welds is positioning yourself so that you can control the electrode filler metal and see the joint both at the same time. Beginning welders assume they must see the tungsten tip as they make the weld. Experienced welders, however, realize that they need to see only the leading edge of the molten weld pool to know how the weld is being produced. A good view of the leading edge will tell you how the base metal is being melted, or depth of penetration. You can even tell from this small portion whether the filler metal is being added at an appropriate rate. As you learn how to control the weld and develop your skills, it is a good idea to gradually reduce your need for seeing 100% of the molten weld pool. Increasing this part of your skill will be a significant advantage in the field because, unlike welding in the classroom, which is typically done on a comfortable-height table, welding in the field may have to be done out of position.

REVIEW

1. What effect does torch angle have on the shielding gas protective zone?
2. Why must the end of the filler rod be kept in the shielding gas protective zone?
3. What can cause tungsten contamination?
4. What determines the correct current setting for a GTA weld?
5. What is the lowest acceptable amperage setting for GTA welding?
6. List the factors that affect the gas flow setting for GTA welding.
7. When should the minimum gas flow rates be increased?
8. What is the minimum gas flow rate for a nozzle size?
9. What is the maximum gas flow rate for a nozzle size?
10. Which incorrect welding parameters does stainless steel show clearly?
11. Using **Table 5.4,** determine the approximate temperature of metal that has formed a dark blue color.
12. Using **Table 5.3, Table 5.5,** and **Table 5.6,** list the filler metals for the following metals.
    a. 1020 low-carbon steel
    b. 309 stainless steel
13. Why is it possible to control a large aluminum weld bead?
14. What may happen to the end of the aluminum welding rod if it is held too close to the arc?
15. What should be done if someone comes in contact with a cleaning chemical?

16. Using **Table 5.7,** determine the suggested setting for GTA welding of mild steel using a 3/32-in. (2.4-mm) tungsten.
17. What can be done to limit oxide formation on stainless steel?
18. How should the filler metal be added to the molten weld pool?
19. How can the rod be freed if it sticks to the plate?
20. How is an outside corner joint assembled?
21. What must be done with the weld craters when back stepping a weld? Why?
22. How is the lap joint tested for 100% penetration?
23. What can prevent both sides of a stainless steel tee joint from being welded?
24. How is the filler metal added for a 3F weld?
25. What can cause undercutting on a 3F tee joint?
26. What helps hold the weld in place on a 2F lap joint?
27. What helps hold the weld in place on a 4G weld?

APPENDIX I—STUDENT WELDING REPORT

I. STUDENT WELDING REPORT

Student Name: _____ Date: _____

Instructor: _____ Class: _____

Experiment or Practice #: _____ Process: _____

Briefly describe task: _____

_____

_____

_____

INSPECTION REPORT			
Inspection	Pass/Fail	Inspector's Name	Date
Safety:			
Equip. Setup:			
Equip. Operation:			
Welding	Pass/Fail	Inspector's Name	Date
Accuracy:			
Appearance:			
Overall Rating:			
Comments:			

Student Grade: _____ Instructor Initials: _____ Date: _____

Jeffus: Welding Principles and Applications, Sixth Edition. Copyright © 2008 by Delmar Cengage Learning, Inc.

This page intentionally left blank

Glossary

amperage A measurement of the rate of flow of electrons; amperage controls the size of the arc.
amperaje Una medida de la proporción de la corriente de electrones; el amperaje controla el tamaño del arco.
amperage range The lower and upper limits of welding power, in amperage, that can be produced by a welding machine or used with an electrode or by a process.
rango de amperaje Los límites máximos y mínimos de poder de soldadura (en amperaje) que puede tener una máquina para soldar o que pueden usarse con un electrodo o a través de un proceso.
anode Material with a lack of electrons; thus, it has a positive charge.
ánodo Un material que carece electrones; por eso tiene una carga positiva.
arc length The length from the tip of the welding electrode to the adjacent surface of the weld pool.
largura del arco La distancia de la punta del electrodo a la superficie que colinda con el charco de la soldadura.
back gouging The removal of weld metal and base metal from the weld root side of a welded joint to facilitate complete fusion and complete joint penetration upon subsequent welding from that side.
gubia trasera Quitar el metal soldado y el metal base del lado de la raíz de una junta soldada para facilitar una fusión completa y penetración completa de la junta soldada subsecuente a soldar de ese lado.
burn-through Burning out of molten metal on the back side of the plate.
metal quemado que pasa al otro lado Metal derretido que se quema en el lado de atrás del plato.
cathode A natural curve material with an excess of electrons, thus having a negative charge.
cátodo Un material de curva natural con un exceso de electrones, por eso tiene una carga negativa.
cellulose-based fluxes Fluxes that use an organic-based cellulose ($C_6H_{10}O_5$) (a material commonly used to make paper) held together with a lime binder. When this flux is exposed to the heat of the arc, it burns and forms a rapidly expanding gaseous cloud of CO_2 that protects the molten weld pool from oxidation. Most of the fluxing material is burned, and little slag is deposited on the weld. E6010 is an example of an electrode that uses this type of flux.
fundentes para electrodos celulósicos Fundentes que usan celulosa de base orgánica ($C_6H_{10}O_5$) (un material normalmente utilizado para fabricar papel), y que se mantienen unidos con un aglomerante de cal. Cuando a este fundente se lo expone al calor del arco, se consume y forma una nube gaseosa de CO_2 que se expande rápidamente y protege de la oxidación al charco de soldadura derretido. La mayor parte del material del fundente se consume, y se deposita poca escoria en la soldadura. El E6010 es un ejemplo de un electrodo que utiliza este tipo de fundente.
chill plate A large piece of metal used in welding to correct overheating.
plato desalentador Una pieza de metal grande que se usa para corregir el sobrecalentamiento.
cleaning action A phenomenon occurring during DCEP in which oxides are removed from the surface to be welded by accelerated ions or electrons. These particles cause surface erosion, also called etching, that assist in the cleaning of the surface to be welded. Cleaning action requires an argon-rich atmosphere.
acción limpiadora Fenómeno que se presenta durante el proceso de DCEP (corriente directa electrodo positivo) en el que los óxidos se eliminan de la superficie para ser soldados por iones acelerados o electrones. Estas partículas provocan que la superficie se erosione (un proceso también conocido como decapado), lo cual auxilia en la limpieza de la superficie que se va a soldar. La acción limpiadora requiere una atmósfera con cantidades abundantes de argón.
collet A cone-shaped sleeve that holds a GTAW electrode in place within a GTA welding torch.
collarín Un manguito cónico que mantiene al electrodo en su lugar dentro del soplete para soldar por arco de tungsteno con gas (en inglés, GTAW).
contamination Undesirable foreign substances found on base metal surfaces, on filler wires, or in gas shielding atmospheres that inhibit welding operations.
contaminación Sustancias extrañas e indeseables en las superficies de metales básicos, alambres de relleno o atmósferas de protección por gas que inhiben las operaciones de soldadura.
cover pass The last layer of weld beads on a multiple pass weld. The final bead should be uniform in width and reinforcement, not excessively wide, and free of any visual defects.
pasada para cubrir La última capa de cordónes soldadura de pasadas múltiples. La pasada final debe ser uniforme en anchura y refuerzo, no excesivamente ancha, y libre de defectos visuales.

duty cycle The percentage of time during an arbitrary test period that a power source or its accessories can be operated at rated output without overheating.
ciclo de trabajo El porcentaje de tiempo durante un período a prueba arbitraria de una fuente de poder y sus accesorios que pueden operarse a la capacidad de carga de salida sin sobrecalentarse.
electrode angle The angle between the electrode and the surface of the metal; also known as the direction of travel (leading angle or trailing angle); leading angle pushes molten metal and slag ahead of the weld; trailing angle pushes the molten metal away from the leading edge of the molten weld pool toward the back, where it solidifies.
ángulo del electrodo El ángulo en medio del electrodo y la superficie del metal; también conocido como la dirección de avance (apuntado hacia adelante o apuntado hacia atras); el ángulo apuntado empuja el metal derretido y la escoria enfrente de la soldadura; y el ángulo apuntado hacia atrás empuja el metal derretido lejos de la orilla delantera del charco del metal derretido hacia atrás, donde se solidifica.
electrons A negatively charged component of an atom. Electrons exist outside of and surrounding the atom's nucleus. Each electron carries one unit of negative charge and has a very small mass as compared with that of a neutron or proton.
electrones Uno de los componentes con carga negativa de un átomo. Los electrones existen en el exterior y alrededor del núcleo del átomo. Cada electrón cuenta con una unidad de carga negativa así como con una masa pequeña en comparación con la de un neutrón o protón.
filler pass One or more weld beads used to fill the groove with weld metal. The bead must be cleaned after each pass to prevent slag inclusions.
pasada para rellenar Uno o más cordones de soldadura usados para llenar la ranura con el metal de soldadura. El cordón debe ser limpiado después de cada pasada para prevenir inclusiones de escoria.
flowmeter A control device used to regulate shield and backing gas flow-in. A flowmeter may control flow on a manifold system or it may be part of a combination flow and pressure regulator device.
flujómetro Un dispositivo de control empleado para regular el flujo del gas de protección y respaldo. Un flujómetro puede controlar el flujo en un sistema colector o bien puede ser parte de un dispositivo con regulación combinada de flujo y presión.
frequency The rate at which alternating electrical current (AC) is measured as it switches back and forth between polarities. Each cycle back and forth is counted as one hertz (Hz).
frecuencia La velocidad registrada por la corriente eléctrica alterna (CA) durante los cambios de polaridad alternados. Cada ciclo de intercambio se cuenta como un hertz (Hz).
gas coverage The protective atmosphere delivered on top of or underneath a weld. Gas coverage may be supplied from the welding torch or from a secondary trailing or backup device.
cobertura de gas La atmósfera protectora aplicada a la parte superior o inferior de una soldadura. La cobertura de gas puede provenir del soplete para soldar o bien de un dispositivo de escape secundario o de respaldo.
guided bend specimen Any bend specimen that will be bend-tested in a fixture that controls the bend radii, such as the AWS bend-test fixture.
probeta de dobléz guiada Cualquier probeta de dobléz en la cual se va a hacer un dobléz guiado en una máquina que controla el radio del dobléz, como la máquina de dobléz guiado del AWS.
hot pass The welding electrode is passed over the root pass at a higher than normal amperage setting and travel rate to reshape an irregular bead and turn out trapped slag. A small amount of metal is deposited during the hot pass so the weld bead is convex, promoting easier cleaning.
pasada caliente El electrodo de soldadura se pasa sobre la pasada de raíz poniendo el amperaje más alto que lo normal y proporción de avance para reformar un cordón irregular y sacar la escoria atrapada. Una cantidad pequeña de metal es depositada durante la pasada caliente para que el cordón soldado sea convexo, promoviendo más fácil la limpieza.
inert gas A gas that normally does not combine chemically with materials.
gas inerte Un gas que normalmente no se combina químicamente con materiales.
interpass temperature In a multiple pass weld, the temperature of the weld area between weld passes.
temperatura de pasada interna En una soldadura de pasadas multiples, la temperatura en la área de la soldadura entre pasadas de soldaduras.
inverter A welding machine that is much smaller than other types of machines of the same amperage range.
inversor Una máquina soldadora más pequeña que otros tipos de máquinas similares con el mismo rango de amperaje.
keyhole A welding technique in which a concentrated heat source penetrates completely through a workpiece, forming a hole at the leading edge of the weld pool. As the heat source progresses, the molten metal fills in behind the hole to form the weld bead.
soldadura con pocillo Una técnica en la cual una fuente de calor concentrado se penetra completamente a través de la pieza de trabajo, formando un agujero en la orilla del frente del charco de la soldadura. Asi como progresa la potencia de calor, el metal derretido rellena detrás del agujero para formar un cordón de soldadura.
lap joint A joint between two overlapping members.
junta de solape Una junta entre dos miembros traslapadas.

magnetic flux lines Parallel lines of force that always go from the north pole to the south pole in a magnet and surround a wire carrying DC current.
líneas magnéticas de flujo Líneas paralelas de fuerza que siempre van del polo norte al polo sur en un magneto, y rodea un alambre que lleva corriente DC.
mineral-based fluxes Fluxes that use inorganic compounds, such as the rutile-based flux (titanium dioxide, TiO_2). These mineral compounds do not contain hydrogen, and electrodes that use these fluxes are often referred to as low hydrogen electrodes. Less smoke is generated with this welding electrode than with cellulose-based fluxes, but a thicker slag layer is deposited on the weld. E7018 is an example of an electrode that uses this type of flux.
fundentes para electrodos de base mineral Fundentes que usan compuestos inorgánicos, como por ejemplo, el fundente a base de rutilo (bióxido de titanio, TiO_2). Estos compuestos minerales no contienen hidrógeno, y a los electrodos que usan estos fundentes se los llama con frecuencia electrodos de bajo hidrógeno. En la soldadura con electrodos se producen menos humos que en la que se realiza con fundentes celulósicos, pero se deposita una capa de escoria más gruesa en la soldadura. El F7018 es un ejemplo de un electrodo que usa este tipo de fundente.
molten weld pool The liquid state of a weld prior to solidification as weld material.
charco de soldadura derretido El estado líquido de una soldadura antes de solidificarse como material de soldadura.
multiple pass weld A weld requiring more than one pass to ensure complete and satisfactory joining of the metal pieces.
soldadura de pasadas múltiples Una soldadura que requiere más de una pasada para asegurar una completa y satisfactoria unión de las piezas de metal.
open circuit voltage The voltage between the output terminals of the power source when no current is flowing to the torch or gun.
voltaje de circuito abierto El voltaje entre los terminales de salida de una fuente de poder cuando la corriente no está corriendo a la antorcha o pistola.
operating voltage Also called closed circuit voltage, it is the voltage measured during welding.
voltaje operativo También conocido como voltaje de circuito cerrado, es el voltaje registrado durante el proceso de soldadura.
output The welding current a particular welding power supply is rated to produce, or the measure of current being produced during welding operations.
potencia de salida La capacidad de la corriente de soldadura de un suministro eléctrico para soldar, o bien la cantidad de corriente producida durante las operaciones de soldadura.
oxide layer An oxide layer is a thin coating that forms on the surface of aluminum alloys. Once the oxide layer has formed, the progress of corrosion is significantly reduced. However, the oxide layer must be removed prior to welding operations because it inhibits penetration. The melting point of aluminum oxide is approximately 3500ºF (1926ºC), while the melting point of aluminum itself is about 1200ºF (650ºC).
capa de óxido Una capa de óxido es una película delgada presente en la superficie de las aleaciones de aluminio. Una vez que se forma la capa de óxido el avance de la corrosión se reduce de manera significativa. Sin embargo, la capa de óxido debe eliminarse antes de realizar la soldadura ya que evita la penetración. El punto de fusión del óxido de aluminio es aproximadamente a 3500ºF (1926ºC), mientras que el punto de fusión del aluminio es a unos 1200ºF (650ºC).
postflow time The time interval from current shut-off to shielding gas and/or cooling water shut-off.
tiempo de poscorriente El intervalo de tiempo de cuando se cierra la corriente a cuando se cierra el gas de protección y o cuando se cierra el agua para enfriar.
postheating The application of heat to an assembly after welding, brazing, soldering, thermal spraying, or thermal cutting.
poscalentamiento La aplicación de calor a una asamblea después de la soldadura, soldadura fuerte, soldadura blanda, rociado termal, o corte termal.
preflow time The time interval that shield gas is delivered to a welding gun or torch before current is applied.
tiempo de preflujo El intervalo en el que el gas de protección se deposita en un soplete antes de aplicar corriente eléctrica.
preheating The heat applied to the base metal or substrate to attain and maintain preheat temperature.
precalentamiento El calor aplicado al metal base o substrato para obtener y mantener temperatura de precalentamiento.
protective zone The area covered by shield gas during welding operations.
zona protectora El área cubierta con gas de protección durante las operaciones de soldadura.
rectification Arc rectification is a phenomenon that occurs when the surface oxide of a nonferrous metal, such as aluminum, acts as a barrier that obstructs the flow of electrons between the electrode and the work during alternating current GTAW operations. Traditional square wave power supplies are equipped with superimposed high-frequency current, which helps establish and maintain the arc. Inverter power supplies with advanced circuitry switch back and forth so quickly between DCEN and DCEP that rectification is rarely experienced and superimposed high frequency is not needed.
rectificación Un arco de rectificación es un fenómeno que se presenta cuando el óxido superficial de un metal

no ferroso, tal como el aluminio, actúa como una barrera que obstruye el flujo de electrones entre el electrodo y el proyecto durante la realización de la soldadura TIG (en inglés, GTAW) con corriente alterna. Los suministros eléctricos tradicionales de ondas cuadradas están equipados con una corriente de alta frecuencia superpuesta que ayuda a establecer y mantener el arco. El suministro eléctrico de los inversores con circuitos avanzados cambia a una velocidad tal entre DCEN y DCEP que rara vez se presenta la rectificación, por lo que el uso de alta frecuencia superpuesta no es necesaria.

rectifier Allows current to flow in one direction only.

rectificador Permite que la corriente fluya en una sola dirección.

root pass The first weld of a multiple pass weld. The root pass fuses the two pieces together and establishes the depth of weld metal penetration.

pasada de raíz La primera soldadura de una soldadura de pasadas múltiples. La pasada de raíz funde las dos piezas juntas y establece la profundidad de la penetración del metal soldado.

rutile-based fluxes This flux system produces a smooth and stable medium-penetrating arc with easily removeable slag. Rutile-based fluxes are used in several types of electrodes and in all welding positions for SMAW and FCAW.

flujos con base de rutilo Este sistema de flujo produce un arco homogéneo y estable para la penetración de medios con escoria fácil de eliminar. Los flujos con base de rutilo se utilizan en varios tipos de electrodos y en todas las posiciones de soldadura SMAW y FCAW.

spark gap oscillator A component in GTAW power supplies that helps to deliver superimposed high-frequency current that assists in arc initiation and alternating current GTAW welding of aluminum alloys.

oscilador de chispa Un componente en suministros eléctricos para soldar por arco de tungsteno con gas (en inglés, GTAW) que ayuda a administrar una corriente de alta frecuencia superpuesta que auxilia en la iniciación del arco y en el intercambio de corriente en soldaduras GTAW de aleaciones de aluminio.

square butt joint A joint made when two flat pieces of metal face each other with no edge preparation.

junta escuadra de tope Una junta hecha cuando dos piezas planas de metal se enfrentan una a la otra sin preparación de orilla.

step-down transformer A transformer-type welding machine that is quieter, more energy efficient, requires less maintenance, and is less expensive than electric-motor or gas-powered engine welding power supplies.

transformador reductor Máquina para soldar tipo transformador que es más silenciosa, consume menos energía, requiere menos mantenimiento y es menos costosa que los dispositivos para soldar con motor eléctrico o de gas.

stringer bead A type of weld bead made without appreciable weaving motion.

cordón encordador Un tipo de cordón de soldadura sin movimiento del tejido apreciable.

surface tension Surface tension is an attractive property of the surface of a liquid. Surface tension causes the surface portion of liquid to be attracted to another surface, such as that of another portion of liquid (as in connecting bits of water or a drop of mercury that forms a cohesive ball). In welding, surface tension keeps the molten weld bead in the proper shape during overhead joining operations.

tensión superficial La tensión superficial es una propiedad de atracción en la superficie de un líquido. La tensión superficial provoca que la superficie de un líquido se anexe a otra superficie, por ejemplo, la porción de otro líquido (como gotas de agua o una gota de mercurio, las cuales forman una esfera cohesiva). En soldaduras, la tensión superficial mantiene la forma correcta del cordón de soldadura fundido durante la realización de operaciones de acoplamiento elevado.

tee joint A joint between two members located approximately at right angles to each other in the form of a "T."

junta en T Una junta en medio de dos miembros que están localizados aproximadamente a ángulos rectos de uno al otro en la forma de "T."

tungsten The primary element used in making GTAW electrodes.

tungsteno El elemento primario empleado en la creación de electrodos GTAW.

voltage The measurement of electrical pressure.

voltaje La medición de presión eléctrica.

wagon tracks A pattern of trapped slag inclusions in the weld that show up as discontinuities in X rays of the weld.

huellas de carreta Una muestra de inclusiones de escoria atrapadas en la soldadura que enseña que hay discontinuidades en los rayos-x de la soldadura.

wattage A measurement of the amount of power in the arc; the wattage of the arc controls the width and depth of the weld bead.

número de vatios Una medida de la cantidad de poder en el arco; el número de vatios del arco controla lo ancho y hondo del cordón de la soldadura.

weave pattern The movement of the welding electrode as the weld progresses; common weave patterns include circular, square, zigzag, stepped, "C," "J," "T," and figure eight.

muestra de tejido El movimiento del electrodo para soldar a como progresa la soldadura; las muestras de tejidos comunes incluyen circular, de cuadro, zigzag, de pasos, "C," "J," "T" y la figura ocho.

weld groove A channel in the surface of a workpiece or an opening between two joint members that provides space to contain a weld.

soldadura de ranura Un canal en la superficie de una pieza de trabajo o una abertura entre dos miembros de junta que provee espacio para contener una soldadura.

welding cables The work cable and electrode cable of an arc welding circuit. Refer to Figure 1-7 for direct current electrode positive.

cables para soldar Los cables de pieza de trabajo y el portelectrodo de un circuito de soldadura de arco. Refierase al dibujo orriente directa con el electrodo positivo.

welding leads The work lead and electrode lead of an arc welding circuit.

cables para soldar Los cables de pieza de trabajo y el portelectrodo de un circuito de soldadura de arco.

weld specimen A welded component or section of a welded component that is to be inspected.

muestra de soldadura El componente o sección soldado de un componente también soldado que se va a analizar.

This page intentionally left blank

Index

Italic page numbers indicate material in tables or figures.

AC (alternating current), 5, *6,* 8–9, 111–113, 123–129
advanced square wave, 126–127
air-cooled torches, 117–119
alloy not specified tungsten electrodes (EWG), 113
alternating current (AC), 5, *6,* 8–9, 111–113, 123–129
alternators, 13–15
aluminum base metal, 107, 112, 146, 151–154, 158t
aluminum deposits, 114
American Bureau of Ships, 87
American Society of Mechanical Engineers (ASME), 87, 88
American Welding Society (AWS), 71, 87, 88
amperage, 4, 32, 146–147
  estimating and calculating, 11–12
  range charts, *29, 135,* 147
angle, torch, 143–144
angle, trailing, 35–36
anode, 6
arc, *3,* 7
  cleaning, 165
  reestablishing the, 125
  striking the, 15, 27–29, 136–138
  temperature of, 4–5
arc blow, 7–8, *9,* 21
arc length, 32–34, 78
arc strike defects, 28–29
argon, 107, 129
ASME Boiler and Pressure Vessel (BPV) Section IX, 87, 88
ASTM standards, 87, 88
AWS Certified Welder program, 71
AWS standards, 88

back caps, 118–119
back edge, 52, 81, 85
back gouging, 72, *73,* 86, 87f
backing gas, 130, 171, 177
backing strips, 72, 74f
backing tapes, 72
bead color, 153
belt sander, 113
breaking and remelting tungsten electrodes, 114–115
bridge rectifiers, 16, 17f
burn back, 83f
burn-through, 31, 45, 72, 75–77, 148–149

cathode, 6
CC (constant current), 6, 7f
cellulose-based fluxes, 41
cerium, 110, 112
cerium tungsten electrodes (EWCe-2), *111,* 112, 152, 153
checklist, portable welder, 16t
chemical cleaning and pointing, 116
chill plate, 156
clamps, 21
cleaning action, 124, 129
cold lap, 34–35
collet, 108–109, 136f
condensation, 121
constant current (CC), 6, 7f
constant voltage (CV), 6, 7f
contamination, 107
  and max cleaning setting, 127
  atmospheric, 143–144, 171, 177
  carbon, 115
  from dirt, 154, 181, 182, 189
  from grinding, 113–114
  from high gas flow rate, 132, 149
  from oxygen, 106
  from vaporized droplets, 112
  from weld pool, 112, 146
  of inert gas, 129
  of shielding gas, 120
  of stainless steel, 152–153
  on tungsten electrode, 115–116, 138, 146
  radioactive, 112
copper, 18–19, 116, 130
cover passes, 84–85
cracking the valve, 134
crater cracking, 145
cups (nozzles), 121–122, 132, *133,* 136f
current settings, 29–31, 146–149
CV (constant voltage), 6, 7f

DAV (dropping-arc voltage), 6
DC (direct current), 14–16, *17,* 112–113
DCEN (direct current electrode negative), 5, 107, 108–110, 123–126
DCEP (direct current electrode positive), 5, *6,* 106–107, 109–110, 116, 123–126
DCRP (direct current reverse polarity). *See* DCEP (direct current electrode positive)
DCSP (direct current straight polarity). *See* DCEN (direct current electrode negative)
defects
  arc strikes, 28–29
  overlaps, 73
  root surface, 72–73
  undercuts, 39, 63, 73, 79, 84–85, 88
deoxidizers, 152
deposition rate, 31
direct current (DC), 14–16, *17,* 112–113
direct current electrode negative (DCEN), 5, 107, 108–110, 123–126
direct current electrode positive (DCEP), 5, *6,* 106–107, 109–110, 116, 123–126
direct current reverse polarity (DCRP). *See* direct current electrode positive (DCEP)
direct current straight polarity (DCSP). *See* direct current electrode negative (DCEN)
drag electrodes (drag rods), 33
dropping-arc voltage (DAV), 6
duty cycles, 17–18

E6010 electrodes, 41
E6011 electrodes, 41
E6012 electrodes, 41
E6013 electrodes, 41
E7016 electrodes, 41
E7018 electrodes, *35,* 41, *44,* 71

E7024 electrodes, 33, 40–41
E7028 electrodes, 40–41
edge weld, 49
electrode amperage range table, *29, 135,* 147
electrode angle, 34–37
electrode holders, *5,* 20–21, 28
electrode lead, 5f
electrode manipulation, 37–39
electrodes, 2, 3f. *See also* tungsten electrodes
  and arc length, 32–33
  F-number classes, 40–41
  size and heat of, 31, 32f
electrons, 3
ER1100 filler metal, 154t
ER308, ER308L filler metal, 153t
ER309 filler metal, 153t
ER310 filler metal, 153t
ER316L filler metal, 153t
ER4043 filler metal, 154t
ER410 filler metal, 153t
ER70S-3 filler metal, 152t
erosion, electrode, 108–110
EWCe-2 (cerium tungsten electrodes), *111,* 112, 152, 153
EWG (alloy not specified tungsten electrodes), 113
EWLa-X (lanthanum tungsten electrodes), *111,* 112–113, 152, 153
EWP (pure tungsten electrodes), 111, 153
EWTh-X (thoriated tungsten electrodes), 111–112, 152
EWZr (zirconium oxide tungsten electrodes), 112, 153

F-number electrode classes, 40–41
face bends, 88–89
filler metals, 106, 114, 129, 152–153
filler passes, 81–84
filler rod manipulation, 144–145
fitup, 101–102
flame-cut specimens, 88–89
flowmeters, 122–123
flux, 2, 28–29, 33, 40–41
frequency, 125

gas coverage, 143
gas flow rate, 132, *133,* 149–151
gas metal arc welding (GMAW), 107
gas tungsten arc welding (GTAW). *See also* tungsten electrodes
  metal preparation, 154–155
  overview, 106–107, 143
  setup, *120,* 133–136
  types of current, 123–126
generators, 13–15
gouging, back, 72, *73,* 86, 87f
grain growth, 107, 112
grinding, 113–114
grinding stone wheel, 114
guided bend specimen, 89f

hard spatter, 30, 33, 41, 132
heat. *See* amperage
heliarc welding, 106
helium, 106, 130
high frequency arc starting, 125–126, 137
hoses, 118, 119–121
hot pass, 71, 79–81
hot start, 131
hydrogen additives, 130

icicles, 53, 54f
inclusions, 82, *94,* 108
inert gases, 129–130
interpass temperatures, 100–101
inverters, 13, *14,* 125–126
iron core, in transformer, 9, 13
iron-powder-based fluxes, 40–41

keyhole, 73–74
knobs, adjustment, 11–12

lanthanum, 110
lanthanum tungsten electrodes (EWLa-X), *111,* 112–113, 152, 153
lap joints, 58–63
leading angle, 34–35
lugs, 19, 20f

magnesium base metal, 107, 112
magnetic flux lines, 7–8
max cleaning setting, 127–128
max penetration setting, 127–128
metals, reactive, 107
mild steel, 151–152
mineral-based fluxes, 41
molten weld pool, 2–4
movable-coil machines, 12, 13f
multiple pass weld, 71
multiple-coil machines, 10–11

nameplate, 18
nickel alloy base metal, 111
nitrogen additives, 130–131
nozzles, 121–122, 132, *133,* 136f

open circuit voltage, 7
open root, 72–73
operating voltage, 7
output, 7
outside corner joints, 54–58
overlaps, 73
oxide layers, 146, 153–154, 165f

patterns, weave, 37–39
penetration, minimizing, 35
plasma arc welding (PAW), 13
point gap setting, 125
pointing tungsten electrodes, 116
poor fitup, 101–102
porosity, 41, 78, *94,* 130, 152
portable welders, 15, 16t
postflow time, 131–132, 149
postheating temperatures, 92, 100
power cable safety fuse, 118
power supplies, 6–7, 9–13
practice welds, 40–41, 151–155
preflow time, 131, 149
preheating temperatures, 92
protective zone, 143–144
pure tungsten electrodes (EWP), 111, 153

reactive metals, 107
rectification, 125
rectifier, 15–16, *17,* 126–128
remelting
  filler, 145
  tungsten electrodes, 116
remote controls, *120,* 133, 147
return water hose, 120
RG45 gas welding rods, 152
RG60 filler metal, 152t
root bends, 88–89
root fusion, *59,* 166f
root opening (root gap), 44, 45f
root pass, 71–78, 143
root penetration problems, 33–34
root surface defects, 72–73
run-off tabs, 83f
rutile-based fluxes, 41

safety
  setup considerations for, 22
  SMAW, 27
  when grinding tungsten electrodes, 114
  when using chemical cleaners, 154
safety fuse, 118

semiautomatic operation, 117f
settings, suggested
aluminum, 158t
mild steel, 157t
stainless steel, 157t
setup
gas tungsten arc welding (GTAW), *120,* 133–136
shielded metal arc welding (SMAW), 22
shelf, support, *38,* 39, *44,* 47
shielded metal arc welding (SMAW)
overview, 2–3
power supply requirements, 7
safety, 27
setup, 22
to join plate, 27
shielding gas, *3, 90,* 129–131
argon, 107, 129
coverage, 143
effect of low amperage, 30f
flow rate, 132, 133t
helium, 106, 130
hydrogen additives, 130
nitrogen additives, 130–131
protective zone, 143–144
shielding gas hose, 120
slag, 2, *3, 47,* 79–81
slag inclusions, 30, 35
spark gap oscillator, 125
spatter, hard, 30, 33, 41, 132
splices (welding cables), 19, 20f
square butt joints, 44–49
square wave, 126–127
stabilizers, arc, 4
steel base metal
low-carbon and mild, 92, 151–152
stainless, 107, 111, 130, 151–153, 157t
step-down transformers, 9–10
stick welding. *See* shielded metal arc welding (SMAW)
sticking, 66–67
stone grinding wheel, 114
striking the arc, 27–29, 136–138
stringer beads, 41–43
surface tension, 146

tack welding, 44–45, *161f–162f*
tap-type machines, 10–11
tee joints, 63–66
testing, 82, 88–89
thermal conductivity, 108, 113, 131, 154
thoriated tungsten electrodes (EWTh-X), 111–112, 152
thorium, 110–111
tip color chart for tungsten electrodes, 111t
titanium base metal, 107
torches, 108, 117–119, 143–144
trailing angle, 35–36
trailing edge, 31, *52,* 59, *61,* 85
transformers, 9–10
tungsten (W), properties of, 107
tungsten electrodes
amperage chart for, 135t
erosion of, 108
properties of, 106–110
shapes and shaping of, *110,* 113–117
types of, 110–113
tungsten inert gas welding (TIG). *See* gas tungsten arc welding (GTAW)

under-fill, *83,* 88f
undercuts, 39, 63, 73, 79, 84–85, 88
valve, cracking the, 134
voltage, 4
closed circuit, 7
constant, 4, 6, 7f
open circuit, 7
operating, 7
rising-arc, 6, 7f

wagon tracks, 80
water-cooled torches, 117–119
water-in hose, 120
wattage, 4
waveforms, AC, 126–129
weave pattern, 37–39
weld beads, 89–92
weld failure from
backing strip, 72
inclusions, 82, 108, 146
inert gas contamination, 129
poor surface preparation, 89
rapid cooling, 92
weld groove, 85, 86f
weld pool, 2–4
weld specimen, 88–89
weld toes, 76f
welders, vision of, 42
welding cables (leads), 13, 18–19, 22
welding current, types of, 5–6
welding position, 40
welds, practice, 40–41, 151–155
work (base metal), 106
work clamps, 21
work lead cable, 8–9f

zip-on hose protection, 121
zirconium, 110–112
zirconium oxide tungsten electrodes (EWZr), 112, 153

针灸

Claudia Focks

Atlas of Acupuncture

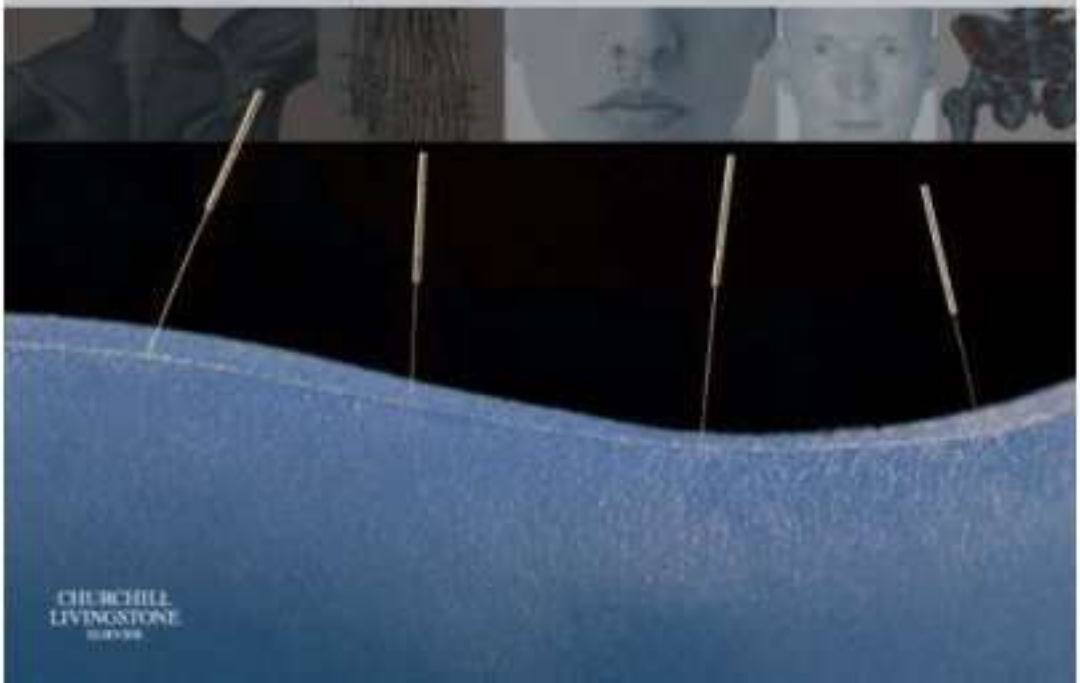

CHURCHILL
LIVINGSTONE

Dedication

To my parents Paula and Alfons Focks who have always supported my siblings and me in finding and pursuing our own paths, with love and gratitude

Commissioning Editor: Karen Morley, Claire Wilson
Development Editor: Kerry McGechie, Martin Mellor
Project Manager: Morven Dean
Designer: Stewart Larking
Photographs: Anja Messerschmidt
Graphics: Henriette Rintelen

CHURCHILL
LIVINGSTONE
ELSEVIER

CHURCHILL LIVINGSTONE
An imprint of Elsevier Limited

First and second edition published in German under the title *Atlas Akupunktur*
1. First edition 1998
2. Second edition 2006
© Elsevier GmbH, München

First edition published in English
© 2008, Elsevier Limited. All rights reserved.

The right of Claudia Focks to be identified as editor of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the Publishers. Permissions may be sought directly from Elsevier's Health Sciences Rights Department, 1600 John F. Kennedy Boulevard, Suite 1800, Philadelphia, PA 19103-2899, USA: phone: (+1) 215 239 3804; fax: (+1) 215 239 3805; or, e-mail: healthpermissions@elsevier.com. You may also complete your request on-line via the Elsevier homepage (http://www.elsevier.com), by selecting 'Support and contact' and then 'Copyright and Permission'.

First edition	**1998**
Second edition	**2006**
English edition	**2008**

ISBN-13: 9780443100284
ISBN-10: 0443100284

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book is available from the Library of Congress

Notice
Neither the Publisher nor the Editor assumes any responsibility for any loss or injury and/or damage to persons or property arising out of or related to any use of the material contained in this book. It is the responsibility of the treating practitioner, relying on independent expertise and knowledge of the patient, to determine the best treatment and method of application for the patient.

The Publisher

ELSEVIER your source for books, journals and multimedia in the health sciences
www.elsevierhealth.com

Working together to grow libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

Printed in China

The Publisher's policy is to use **paper manufactured from sustainable forests**

Preface

'An acupuncture point is located where it is located' – like veins, arteries or nerves, the location of acupuncture points may vary. Consequently, locating and stimulating acupuncture points is an individual process in each patient, similar to locating veins or arteries for puncture. The goal of this atlas is therefore to help acupuncture students and therapists find the correct location of acupuncture points. Ultimately, the exact location of an 'acupuncture hole' will be where it can be palpated, addressed energetically in a safe way, and – most importantly – where it will be therapeutically effective.

Acupuncture therapy requires not only the knowledge to locate individual points but also an understanding of the context of these points. This atlas therefore includes both a description of the individual points as well as the central idea – the channel system.

The core chapters 4, 5 and 6 contain single page portraits of all the channel points as well as extra points used in body acupuncture. The clearly structured page layout, complemented by pictograms, allows for quick access to the clinically relevant information for the location and indication of individual points. Clear step-by-step instructions guide the reader through the surrounding surface anatomy to the correct location of the point. Figures with the relevant anatomical structures, as well as drawings showing channel pathways pertaining to the relevant region of the body, provide a further practical aid for correct point location. The text and the photo details also contain information regarding other points located in the vicinity or points in comparable locations in other parts of the body. This not only draws attention away from the individual points to the larger anatomical orientation, but also fosters an understanding of the context between the location and the action of a given point.

Chapter 7 presents the channel points according to the anatomical region – including illustrations of anatomical overview as well as text/picture details. This format will deepen the understanding of the anatomical relationship between the channel points and complete this academic concept. Chapter 2 'Location Methods and Cun Measurements' and Chapter 3 'Anatomical Orientation' provide further support in locating the individual points. The modified illustrations in these chapters are based on the well-known Sobotta Atlas.

Chinese medicine practitioners work with the Qi and its flow. In this context the individual 'acupuncture hole' has both an anatomical and an energetic component and is an intrinsic part of the channel and vessel network, which provides an exchange between the body's Exterior and Interior. In this way Chapter 1 describes the network of channels and vessels, and Chapter 8 the point categories and point combinations, explaining them in terms of their effects with relation to channel energetics, and presenting them in a clearly defined and visually organised manner. Chapter 9 contains up-to-date information and data about the scientifically proven effects or otherwise of individual points.

I hope the new edition of this atlas will provide a valuable help to students and therapists alike in studying acupuncture and applying the information in clinical practice. I look forward to hearing your constructive criticism and ideas.

Claudia Focks, March 2008

Acknowledgements

This Acupuncture Picture Atlas is a joint project of many very helpful contributors who have helped to complete it.

First of all my thanks go to the models and especially to the photographer Anja Messerschmidt. Her understanding and photographic skills provided the visual foundation for the Atlas to also be an artistic 'feast for the eyes'.

I would like to thank the graphic designer Henriette Rintelen for the successful graphic design and converting my ideas into visual images.

Thanks go to my colleague Ulrich März for the succint arrangement and editing of the chapter on Anatomical Orientation as well as his contributions to the core chapters. I would like to thank my colleague Ingolf Hosbach for his contributions to the chapter on Research as well as his idea about the visually concise pictograms and for partially converting them by computer.

I want to thank all the staff at Elsevier involved in this project for their, once again, very productive co-operation. My thanks go particularly to Rolf Lenzen for agreeing to do the atlas, Christine Kosel for dealing with the gruelling organisation during production, as well as Petra Münzel-Kaiser and Christl Kiener for editing. My thanks extend particularly to Christl Kiener for her numerous ideas and fruitful exchange of concepts.

My heartfelt gratitude also includes my teachers who have sown the nourishing spark of enthusiasm for Chinese and energetic medicine. I thank my patients for their trust and for allowing me to learn from them.

I would also like to thank the strong women in my family and my circle of friends for their inspiration and being a source of vibrant female power. Thank you to my sister Annette Focks for her friendship and wonderful music, which helped me overcome some of the hurdles of this project. I extend particular thanks to Anneliese Solenski and Christa Skopp for their helpful support with my private affairs.

Above all I want to thank my inspirational husband Christoph Ranzinger and our wonderful children Tabea and Lasse for their love and generous tolerance, without which the Atlas would not have been possible.

For the English edition I would like to extend my thanks to the staff at Elsevier in London, particularly Karen Morley for accepting the atlas and to Kerry McGechie and Martin Mellor for the extensive editing work. Thank you also to Johanna Schuster who has translated the Atlas into English with great care and knowledge.

Claudia Focks, Rottweil, March 2008

About the editor and contributors

Dr med. Claudia Focks, Rottweil: anaesthesiologist, additional qualification in naturopathy and homeopathy; training in Chinese Medicine in China, Austria, and Germany, postgraduate training at the ETHNOMED Institut für Ethnomedizin e.V. (Institute for Ethnomedicine, registered charity) in Munich. Clinical focus on Chinese Medicine and energy therapies. Co-editor and co-author of *Leitfaden Chinesische Medizin* (*Chinese Medicine: An Introduction*), co-author of *Leitfaden Akupunktur* (*Acupuncture: An Introduction*), both published by Elsevier Munich. Lectures on Chinese Medicine, both in Germany and abroad.

Dr med. Ulrich März, Ulm: General practitioner with his own practice, focus on acupuncture and Traditional Chinese Medicine. Adjunct lecturer at the University of Ulm for acupuncture and TCM. Project manager for the postgraduate programme for acupuncture and TCM at the Akademie für Wissenschaft, Wirtschaft und Technik (Academy for Science, Economy and Technology) at the University of Ulm. Co-author of *Leitfaden Akupunktur* (*Acupuncture: An Introduction*), published by Elsevier Munich.

Dr med. Ingolf Hosbach, Bochum: Anaesthesiologist, additional qualification in palliative therapy. Employed at the University Clinic Bergmannsheil Bochum. Adjunct lecturer at the Universities of Witten/Herdecke and Bochum. Scientific publications and lectures on palliative therapy, acupuncture and TCM research. Webmaster of the homepage of the Department for Chinese Medicine at the University Witten/Herdecke and of the DWGTCM e.V. (www.dwgtcm.com).

Frontcover-F10028 2/23/08 11:37 AM Page 1

Pictogrammes

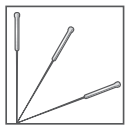

Angle of insertion: The angle of insertion recommended for a particular point (see figure below) is shown in colour. (Note that occasionally more than one angle is possible – this is signified by colouring as appropriate).

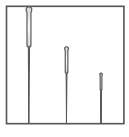

Depth of insertion: Shallow: up to 0.5 cun, **Medium:** 0.5–1 cun, **Deep:** >1 cun. (In some cases more than one depth is possible depending on the angle of insertion – this is signified by colouring as appropriate).

Caution! Organs and structures such as the eyes, nerves, blood vessels, peritoneum (symbolised by intestinal tissue) etc. that might be injured by needling a particular point are shown by a symbol together with a blue exclamation mark. A lightning bolt signifies a point that is painful when needled.

Moxibustion: In principle all points can be treated with moxibustion. The following options are pointed out: **Blue colour below moxa cigar:** moxibustion particularly recommended, **Question mark next to moxa cigar:** moxibustion questionable (contraindicated according to some classical texts, but contradictory information), **Moxa cigar crossed out:** moxibustion contraindicated.

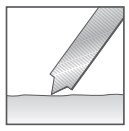

Bloodletting: In principle bloodletting may be applied to all points. **Neutral pictogramme:** bloodletting possible, **Blue droplet of blood:** bloodletting particularly recommended, **Blue question mark:** bloodletting questionable

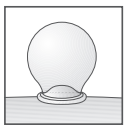

Blue cup: Cupping particularly recommended.

Anatomical orientation: A blue hand signifies cases where anatomical structures are of particular importance for point location (for more detail ➞ chapter 3).

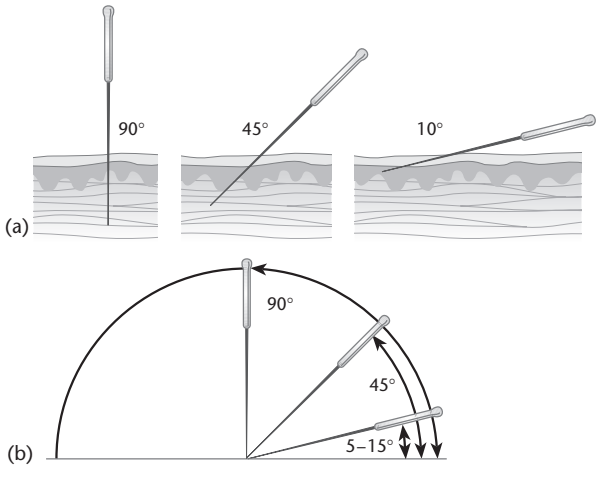

Angle of insertion
a) Angle of insertion and layers of tissue: All needles are inserted to the same 'depth' (=the same proportion of the needle is inserted into the tissue), but reaching different levels of tissue.
b) Angle of insertion: transverse (subcutaneously): 5–15°; oblique: 15–45°; perpendicular: 90°.

Point hierarchy (Chapters 4–6)

In the introductions to the channels (chapters 4–6) the importance of points are marked with the following symbols:
■■: very important, universal point
■■: important point
This hierarchy is of course subjective and influenced by the authors' clinical experience but has proved very helpful for beginners.

1 *Jing luo* (Channel and Network Vessel System)

Claudia Focks

1.1 Introduction

1.1.1 Two Models of Qi Flow in the Channels – Historical Overview

There exist essentially two views about the direction of the Qi flow in the channels, which are based on different understandings of the Qi movement within the body (→ Pirog 1995, Manaka 1995/2004). For a better understanding of either model, imagine a person extending his/her arms towards the sky (→ Figs. 1.1 and 1.2).

Centripetal circulation model

The centre of classical Chinese cosmology, which describes humans as standing between heaven (Yang) and earth (Yin), includes the mutual relationship, influence, and dependence between cosmos and man. Accordingly, early records of the pathways (→ see appendix) depict the (11/12) channels as pathways that originate at the extremities, flow towards the centre of the body and terminate either on the head or the torso. This is the distinguishing feature of the **centripetal circulation model** (→ Fig. 1.1).

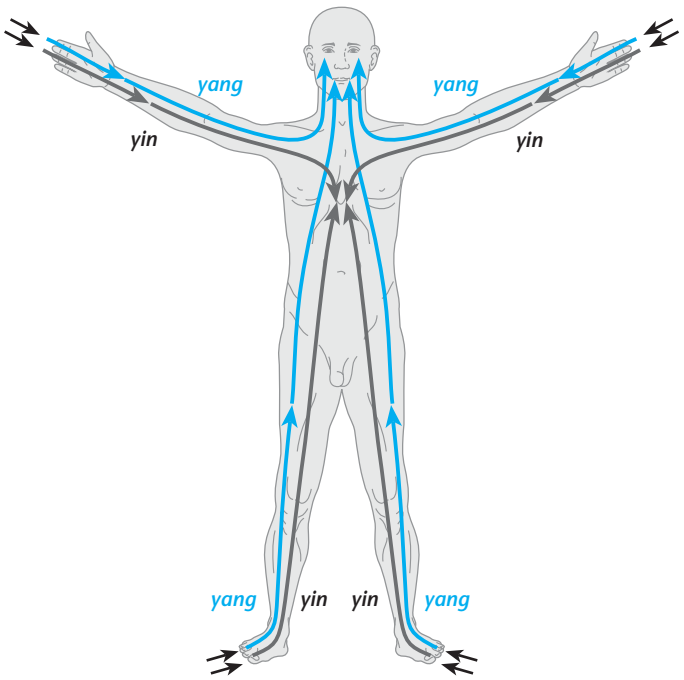

Fig. 1.1 Centripetal circulation model (modified according to Pirog 1996)

This model is also present in the theory of the five *shu*-Transporting points (→ 8.1.6). The macrocosmic Qi enters the body at the tips of the extremities and can be compared to the course of a river. It begins very dynamically as a well, spring and stream, expands to form a river and flows into the wide, deep sea at the elbows and the knees and then further to the internal organs via the channels.

According to this model, the functions of the channels can be compared to antennae, which receive the cosmic influence, transmitting it into the body. The Qi flow in the channels is always from distal (coming from the Exterior, entering at the tips of the extremities) to proximal (towards the centre, flowing towards the internal organs). Each channel connects man with a different part of the cosmos, which can be identified by a numerological structure. Thus a relationship was assumed between the eight extraordinary vessels and the eight trigrams of the *Yijing* (→ 1.7). The primary channels reflect the 12 earthly branches and the 10 heavenly branches. Since the early records only mention 11 channels (→ Appendix 2), they were counted as 10 channels on the arms and 12 channels on the legs. Later, the 12 primary channels were mainly associated with the 12 earthly branches – the 10 heavenly branches rather representing the Five Phases model (*note*: the earthly branches can be included in the latter as well).

Self-contained circulation model

In the course of the development of modern Chinese society, the idea of a close relationship between man and cosmos (as a heaven-earth-man model) became weaker. The channel system was now increasingly compared to phenomena created by man himself such as **canals** and **drainage ditches**. The importance of the original connection with the cosmos declined while that of relationships within society was on the rise. Chinese society became more complex, forming a self-contained unit. Similarly, the concept of the channels as a connection to the macrocosm was partially abandoned and the Qi flow is described as independent and self-contained (→ Fig. 1.2). According to this **self-contained circulation model**, Qi can flow forward as well as backward – from the outer extremities to the Interior of the body and from the Interior to the Exterior.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 2

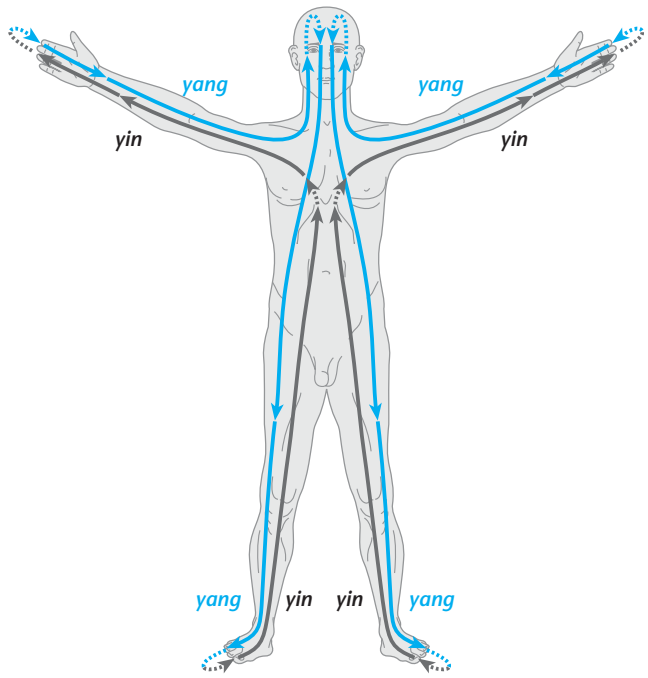

Fig. 1.2 Self-contained circulation model (modified according to Pirog 1996)

Thus the direction of the Qi flow in the primary channels can be from proximal to distal and vice versa, depending on the Yin/Yang polarity of the respective channel, as well as the respective extremity.

This second, probably more recent concept describes the Qi **as circulating continuously** through the body (→ Appendix 2): from the thorax to the hand, to the head, to the foot and back to the thoracic region. These ideas reflect the development of the Chinese civilisation, its agriculture and in particular its **water control** and water **storage** in reservoirs, drainage canals, ditches, etc., which largely form the foundation of the theory of the **channels as water conduits** circulating Qi and Blood. The connections among the channels are considered to be anastomoses (→ 1.2.2), which facilitate the ceaseless, circular flow of Qi from one channel to the next, allowing the exchange of the Qi flow with the Interior.

The self-contained circulation model can thus be described as follows:

- Hand Yin channels run from the thorax to the hand: **LU, HE, P**
- Hand Yang channels run from the hand to the head: **L.I., S.I., T.B.**
- Foot Yang channels run from the head to the foot: **ST, BL, G.B.**
- Foot Yin channels run from the foot to the thorax: **SP, KID, LIV**

More importance is attached to this model of Qi flow in the Western acupuncture tradition, which can in part explain the use of numbers in naming the acupuncture points, rather than names as in China.

However, the structure of this new circulatory model was probably also too rigid to sufficiently explain some of the effects of acpuncture. According to Pirog (1996), this might be the reason why the secondary channels, for example the sinew channels (→ 1.4) and the divergent channels (→ 1.3), with their rather primitive and natural pathways, were integrated into the channel and network vessel system (*jing luo*) in accordance with the **centripetal** circulation model.

Comparison of the two circulation models (modified according to Pirog 1996)

	Centripetal model	Self-contained model
Direction of Qi-flow	Always from distal to proximal	Either from proximal to distal or vice versa, depending on the polarity of the channel (Yin/Yang)
Origin of Qi	From outside the body. The distal end of the channel is open in order to receive cosmic Qi	Originates in the inside of the body. The distal end of the channels is connected to the channel that follows next.
Function of the channels	Transporting the Qi from the outer cosmos to the Interior of the body. Supporting the relationship between man and nature/ cosmos	Circulation of the Qi inside the body. Supporting man's relationship with himself

1.1.2 Overview of the Channel and Network Vessel System (*jing luo* System)

In *Ling Shu*, chapter 11, it says: 'Man lives, diseases occur … both the beginner and the experienced master always have to start with the channels and network vessels (*jing luo*).'

In Chinese Medicine, the *jing luo* are considered to be a network of channels and vessels in which the Qi and Blood (*xue*) flow. They are connected to the Organ systems (*zangfu*) and 'water' the whole organism, supplying the body with Qi and Blood (*xue*) on the surface (Exterior) and deep inside the body (Interior), above as well as below.

From a functional point of view, the channels and network vessels (*jing luo*) govern the distribution of Qi and Blood (*xue*), they regulate Yin and Yang and they protect the body. However, they also enable the spreading of diseases. Reactions to any disorders may therefore manifest along these channel pathways. These could be either disorders of the channels themselves, or external reflections of *zangfu* disorders. In clinical practice, the channels and network vessels (*jing luo*) can be utilised in order to send Qi to the diseased parts of the body (for an overview of the classification and nomenclature of the *jing luo* system → Fig. 1.3).

Ch01-F10028.qxd 2/22/08 2:30 PM Page 3

a)

- *Jing luo* (channel and network vessels)
  - *Jing mai* (channels)
    - 12 *jing zheng* (primary channels)
      - *shou* (Hand)
        - 3 *yin*
          - *tai yin* LU channel
          - *shao yin* HE channel
          - *jue yin* P channel
        - 3 *yang*
          - *tai yang* S.I. channel
          - *shao yang* T.B. channel
          - *yang ming* L.I. channel
      - *zu* (Foot)
        - 3 *yin*
          - *tai yin* SP channel
          - *shao yin* KID channel
          - *jue yin* LIV channel
        - 3 *yang*
          - *tai yang* BL channel
          - *shao yang* G.B. channel
          - *yang ming* ST channel
    - 12 *jing bie* (divergent channels)
    - 12 *jing jin* (sinew channels)
    - 8 *qi jing ba mai* (extraordinary vessels)
      - *ren mai du mai*
      - *chong mai dai mai*
      - *yin qiao mai yang qiao mai*
      - *yin wei mai yang wei mai*
  - *Luo mai* (*luo*-vessels)
    - 12 *luo*-connecting vessels (begin at the extremities) — Spread from the primary channels
    - 4 *luo*-connecting vessels (begin at the torso)
      - 1 spreads from the SP primary channel (great *luo*-connecting vessel of the Spleen)
      - 2 spread from the extraordinary vessels (*ren mai* and *du mai*)
      - 1 spreads from a *fu*-Organ (great *luo*-connecting vessel of the Stomach)
    - Small *luo*-connecting vessels (superficial *luo*)
      - *sun luo* — Vertical branches of the *luo mai*
      - *fu luo* — Sub-branches of the *sun luo*
      - *xue luo* — Horizontal branches of the *fu luo*
    - 12 *pi bu* (cutaneous zones) — Superficial aspect of the *jing luo* system (channel and network vessel system), covering it towards the Exterior (outer layer)

b)

- 12 *jing zheng* (primary channels) = 6 *liu jing* (great channels)
  - 3 *yin* (*tai yin*, *shao yin*, *jue yin*)
    - 3 *shou* (Hand)
      - *tai yin* LU channel
      - *shao yin* HE channel
      - *jue yin* P channel
    - 3 *zu* (Foot)
      - *tai yin* SP channel
      - *shao yin* KID channel
      - *jue yin* LIV channel
  - 3 *yang* (*tai yang*, *shao yang*, *yang ming*)
    - 3 *shou* (Hand)
      - *tai yang* S.I. channel
      - *shao yang* T.B. channel
      - *yang ming* L.I. channel
    - 3 *zu* (Foot)
      - *tai yang* BL channel
      - *shao yang* G.B. channel
      - *yang ming* ST channel

Fig. 1.3 (a) Overview and classification of the *jing luo* system (channel and network vessel system); (b) the 12 primary channels and the six great channels *(liu jing)*

Ch01-F10028.qxd 2/22/08 2:30 PM Page 4

1.1.3 Distribution and Organisation of the Channel and Network Vessel (*jing luo*) System

According to the law of the Exterior–Interior (*biao-li*) the 'Exterior' communicates with the 'Interior'.

Exterior (*biao*) corresponds to the skin, the muscles and the pathways of the channel and network vessel superficial (*jing luo*) system. The deeper pathways of the channels and the Organ systems (*zangfu*) are attributed to the Interior (*li*). A particular organisational structure within the *jing luo* system is necessary to safeguard the circulation of Qi and the communication between the Exterior and the Interior. In this respect, the extraordinary vessels play a special role. While they play a major role in coordinating and regulating the primary channels and the *jing luo* system in general, they do not directly connect the Interior and the Exterior. There is also no direct connection between the extraordinary vessels and the *zangfu*-Organs (see → 1.7 and Chapter 5).

Depth organisation of the *jing luo* system

Differing ideas exist regarding the depth at which the various channels and vessels are located within the body. The table below and Fig. 1.4 show the ideas according to Solinas et al. (1998) and Deadman et al. (1998) in a modified way. While these authors regard the deep, Interior pathways of the primary and divergent channels as the deepest channel structures, other authors attribute this role to the eight extraordinary vessels (for the eight extraordinary vessels, see → 1.7).

Overview over the possible location with regard to depth of the *jing luo* system

Levels	Channel system
Superficial levels (skin and muscles)	• 12 cutaneous regions • Superficial *luo*-connecting vessels (*xue luo, fu luo, sun luo*) • 12 sinew channels (*jing jin*)
Mid-levels	• 15/16 *luo*-connecting vessels (*luo mai*) • External pathways of the 12 primary channels • External pathways of the 12 divergent channels (*jing bie*) • Eight extraordinary vessels
Deep levels (*zangfu*-Organs)	• Deep (internal) pathways of the 12 primary channels • Deep (internal) pathways of the 12 divergent channels

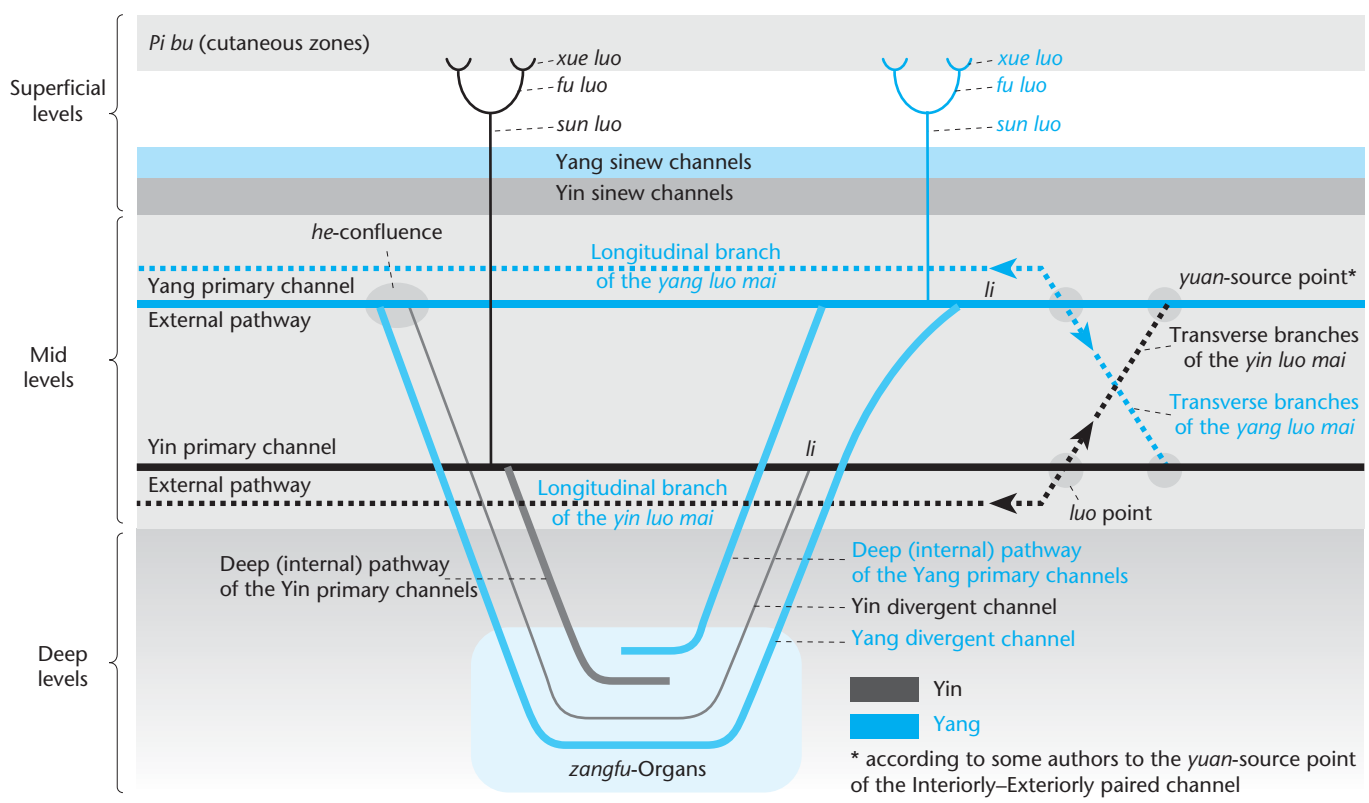

Fig. 1.4 Illustration of the possible organisation of the *jing luo* system (channel and network vessel system)

Ch01-F10028.qxd 2/22/08 2:30 PM Page 5

Channel systems

Each primary channel, together with its pertaining 'secondary channels', forms a complex, multi-layered organisational unit, a 'system'. These systems comprise the various relationships and connections among each other as well as with the eight extraordinary vessels (→ 1.7, → Chapter 5). They support the harmonious regulation within the body. There are 12 channel systems (hand *taiyin*, hand *yangming*, etc.), each of which includes the following:

- a **primary channel** (*jing zheng*) (→ 1.2) with an external and internal pathway, which enters its pertaining *zangfu*-Organ
- a **divergent channel** (*jing bie*) (→ 1.3)
- a **sinew channel** (*jing jin*) (→ 1.4)
- a great ***luo*-connecting vessel** (*luo mai*) (→ 1.5).

In turn, the **cutaneous zones** (*pi bu*) (→ 1.6) wrap around the whole channel system.

1.1.4 Qi Circulation and the Channel System

What moves with and within the channels?

The channel system represents the 'pathway of the Qi'. In acupuncture many schools of thought work with the **True Qi** (***zhen qi***), which, according to Maciocia (1989), represents the final stage of a process of Qi refinement and transformation: the Gathering Qi (*zong qi*), acted upon by the Original Qi (*yuan qi*) as a catalyst, becomes True Qi (*zhen qi*), which manifests in its two aspects, the Defensive Qi (*wei qi*) and the Nutritive Qi (*ying qi*). According to Larre and Rochat de la Vallée (1986), the True Qi is regarded as the **sum of all Qi mechanisms and/or forms of Qi** in the body. In other words: if there is a balanced and harmonious flow in the channels, this can be called True Qi (*zhen qi*). It means that which circulates in the 'here and now'. The sum total of all forms of Qi in their correct alignment is called **Upright Qi** (***zheng qi***) and forms the counterpart to Evil Qi (*xie qi*), for example to pathogenic factors or counterflow Qi.

Forms of Qi

Essence (Qi) (*jing qi*)

The Essence (Qi) acts within the organism and, according to some authors, to some extent also in the extraordinary vessels (→ 1.7). It represents the combination of Early Heaven Essence inherited from the parents, which, according to many schools of thought, is stored in the Kidneys and has a relationship to the Life Gate (*mingmen*), and the Later Heaven Essence of the Middle Burner. They support and complement each other.

Original Qi (*yuan qi*)

The Original Qi (*yuan qi*) is often described as the **active form of Essence** (***jing***), circulating in the channels and spreading to the whole organism with the help of the Triple Burner. Like a catalyst, it is the dynamic driving force which wakens and sustains the functional activity of all organs and structures. It has a pre- and a postnatal component. Therefore its proper functioning is dependent on the supply of acquired Qi derived from water and food supplied by the Middle Burner. The Original Qi can be directly accessed and influenced at the *yuan*-source points (→ 8.1.1) or at the points **Ren-17** (*danzhong/shanzhong*), **Ren-12** (*zhongwan*) and **Ren-6** (*qihai*).

Gathering Qi (*zong qi*)

According to Larre and Rochat de la Vallée (1986), the Gathering Qi (*zong qi*) is formed in the centre of the thorax when the essences derived from respiration (Great Qi, *da qi*) meet with the food (*shui gu*, from Grain Qi (*gu qi*) and water) and are set in motion through ancestral mechanisms by the first postnatal breath. Like an 'engine', it dictates the rhythm and circulation, it controls respiration and it regulates the heartbeat. It does not circulate itself, but is like a 'Sea of Qi' that 'collects' in the centre of the thorax behind **Ren-17** (*danzhong/shanzhong*). This sea contains the water from all the rivers (e.g. the vessels), resembling an inexhaustible reservoir that does not overflow but redistributes all its water supplies.

Defensive Qi (*wei qi*)

The *wei qi* is the body's **Defensive Qi**. Thus it defends the body at the level of the skin, the fascia and the muscles. In relation to the Nutritive Qi (*Ying qi*), it is a more slippery and less pure form of Qi that moves dynamically and quickly through the body like a **guard**. Its root lies in the **Lower Burner**, where it is produced by the ***mingmen***-Fire (Life Gate fire), so that it has a considerable genetic or constitutional component. Essence (*jing*) and Original Qi (*yuan qi*), which are stored in the Lower Burner (according to many schools in the Kidneys), are involved in the formation of Defensive Qi (*wei qi*), therefore also playing a role in the defence against Exterior pathogens. Additionally, the *wei qi* is constantly being replenished with pure essence, formed from food by the Spleen and Stomach in the **Middle Burner**. It is dispersed throughout the body by the **Upper Burner**. Proper functioning of the *wei qi* therefore depends on all three Burners.

Nutritive Qi (*ying qi*)

The Nutritive Qi (*ying qi*) is the result of a purification and/or distillation process of pure, clear origin. Compared to Defensive Qi (*wei qi*), it is more Yin in nature. It **nourishes** the whole body and is its '**building substance**', thus often translated as 'constructive Qi' or 'building energy'.

Circulation of Defensive Qi (*wei qi*) and Nutritive Qi (*ying qi*)

Circulation of Defensive Qi (*wei qi*)

In chapter 43 of the *Su Wen* the *wei qi* is described as flowing 'outside of the *mai*' (the channels and vessels). However, it circulates in part along the channel pathways, moving in the space between the skin and the muscles – the ***cou li***.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 6

The *cou li* – often unsatisfactorily translated as 'pores' – are the striae or compartments between the skin and the muscles. They have the function of a gateway for the entry and exit of Qi and fluids and serve as protection against the invasion of Exterior pathogenic factors. According to Larre and Rochat de la Vallée (1986), the *cou li*, as the outermost of the wrappings of the Triple Burner, cover the whole body. They connect the surface of the body with its inner organs. This explains the effectiveness of acupuncture and other manual therapies for the treatment of internal disorders as well as showing up the presence of diseases of the inner organs on the Exterior of the body.

In the **superficial layers of the body**, the Defensive Qi (*wei qi*) circulates through the skin and the superficial musculature, warming, nourishing and strengthening them. Through these actions it supports the general defence against Exterior pathogenic factors, acting mainly in the realm of the sinew channels (*jing jin*) (→ 1.4). In the **deep layers of the body**, it plays an important role in the functioning of the 'diaphragm'. According to Larre and Rochat de la Vallée (1986), this represents not only a membranous barrier between the thorax and the abdomen, but can be regarded as a 'sac of membranes', including and connecting the peritoneum, the pleura and the pericardium. Nielsen (1995) interprets this 'network of caves' as the **inner aspect of the Triple Burner**, which connects with its **outer aspect, the *cou li***. Following this interpretation, the Defensive Qi (*wei qi*) would also be involved with the **mesenteric defence** and the protection of the inner organs.

According to chapter 75 of the *Ling Shu*, the **circulation of the Defensive Qi (*wei qi*)** is cyclical, changing from day to night and vice versa. At dawn, when the Yin Qi is exhausted, the Yang Qi pours itself from the eyes and the eyes are opened. As a result, the *wei qi* rises from the heel via the (Yin) *qiao mai* upwards to the eye to **BL-1** (*jingming*) and flows into the whole body like a waterfall by following the six great Yang channels – 'it moves 25 times in the Yang'. At dusk, when the Yang Qi is exhausted, the *wei qi* enters the Interior of the body and moves '25 times in the Yin', following the controlling cycle (*ke* cycle → 8.2.5): from the Kidneys to the Heart, from the Heart to the Lungs, from the Lungs to the Liver, from the Liver to the Spleen, and back to the Kidneys. Therefore the *wei qi* has a relationship with the sleep–wake rhythm: during sleep it withdraws deeper into the body – during daytime it circulates in the outer layers of the body. The two extraordinary vessels, the *yin qiao mai* and the *yang qiao mai* (→ 1.7, → Chapter 5), play an important role in this respect. If their cycle is blocked, the circulation of the *wei qi* will be interrupted and disorders will arise. The *yin qiao mai* ascends, while the *yang qiao mai* descends, both meeting at the eyes at **BL-1** (*jingming*) and forming a cycle, similar to the small heavenly cycle of the *ren mai* and *du mai*. If there is an excess of Yang Qi, this will be drained from the channels into the *yang qiao mai*, 'the eyes cannot close' and insomnia will occur. If there is an excess of Yin Qi, this will be drained from the channels into the *yin qiao mai*, 'the eyes cannot open' and somnolence will occur. In both cases, a relative disharmony develops between the two opponents.

Circulation of the Nutritive Qi (*ying qi*)

The Nutritive Qi (*ying qi*) circulates in the *mai*. The *mai* include both the primary channels (*jing mai*) and the secondary channels such as the Connecting Vessels (*luo mai*, *sun luo*), the divergent channels (*jing bie*), the extraordinary vessels and the Blood Vessels. Wherever there are channels and vessels, the *ying qi* will circulate. According to chapter 16 and 18 of the *Ling Shu*, the *ying qi* circulates constantly and continuously.

First circulation of the *ying qi* (→ Fig. 1.5)

The first circulation of the *ying qi* encompasses the 12 primary channels. The cycle **starts** at the **primary LU channel** and **ends** at the **primary LIV channel**, which in turn connects with the primary LU channel, thus closing the cycle (→ Fig. 1.5). The **Middle Burner** provides and **distributes** the *ying qi* throughout the whole body. According to some of the classics, the Grain Qi (*gu qi*) derived from food is only transformed into Nutritive Qi (*ying qi*) in the Lungs, not already in the Middle Burner. Since the internal pathway of the primary Lung channel begins in the Middle Burner, these two statements do not contradict each other in principle.

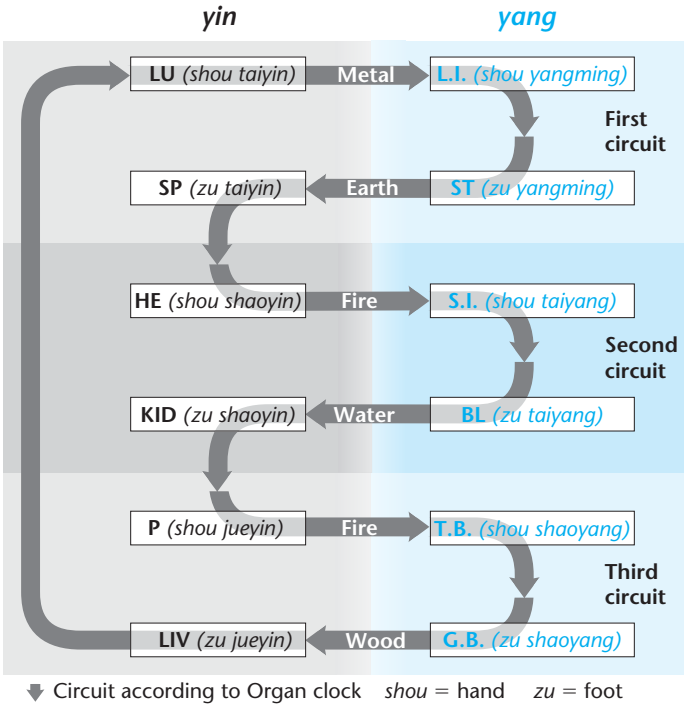

Fig. 1.5 First circulation of the Nutritive Qi (*ying qi*)

The **changes in Yin/Yang polarisation** – the changes from a Yin channel to a Yang channel and vice versa – always take place in the hand or the foot (also → 1.2.2). For example, the Yang energy potential rises from the Yin channel to the Yang channel, and then descends again from the Yang channel to the Yin channel (→ Fig. 1.6).

Ch01-F10028.qxd 2/22/08 2:30 PM Page 7

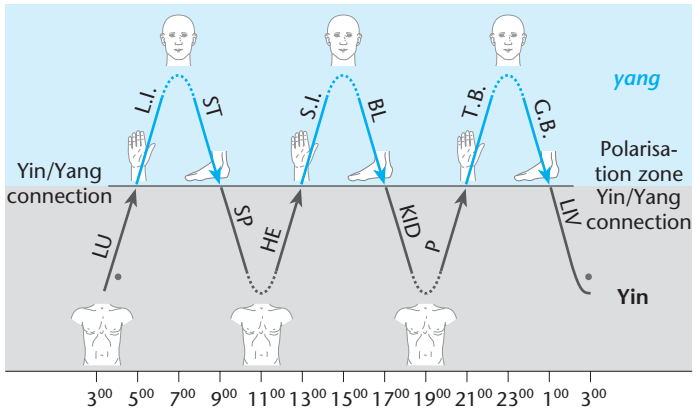

Fig. 1.6 Changes in Yin/Yang polarisation, shown as a sinus curve

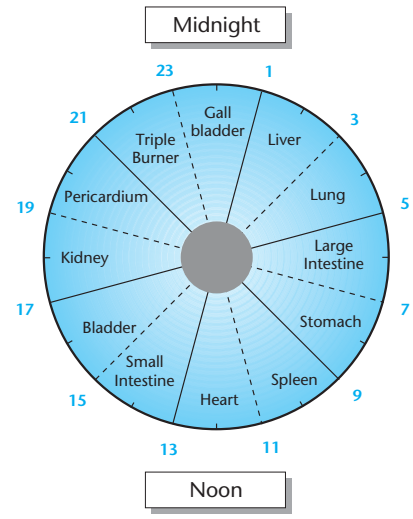

Fig 1.7 Organ Clock

First circulation of the Nutritive Qi (*ying qi*) and the Organ clock

The Nutritive Qi (*ying qi*) circulates in the 12 primary channels in a **circadian 24-hour rhythm** (→ Fig. 1.7) with each **Chinese hour**, named after one of the 'earthly branches', corresponding to two Western hours (→ expanded Organ clock, Fig. 1.8).

Each two-hour slot favours a particular channel, in other words: for two (Western) hours per day, **energy flow peaks** in a particular channel. During this period, the respective channel Qi increases, falling again during the following two hours. However, it will never fall below a particular level, so that there is always a certain portion of energy flowing continuously through the channels.

Example: The channel Qi of the primary Stomach channel will begin to appear more 'powerful', when the channel Qi of the Large Intestine is very strong. Therefore the 'high tide' of the Stomach channel takes place between 5–7 a.m., its peak is between 7–9 a.m. and its 'low tide' between 9–11 a.m. During the peak period of the Stomach channel, the channel on the opposite side of the Organ clock is at its lowest point (→ 8.3.7); while the Stomach channel is peaking, the Pericardium channel is at its lowest point.

In figure 1.8 the 'expanded Organ clock' with its 24-hour circulation through the 12 primary channels is shown in relation to the earthly branches and the hexagrams of the *Yijing*.

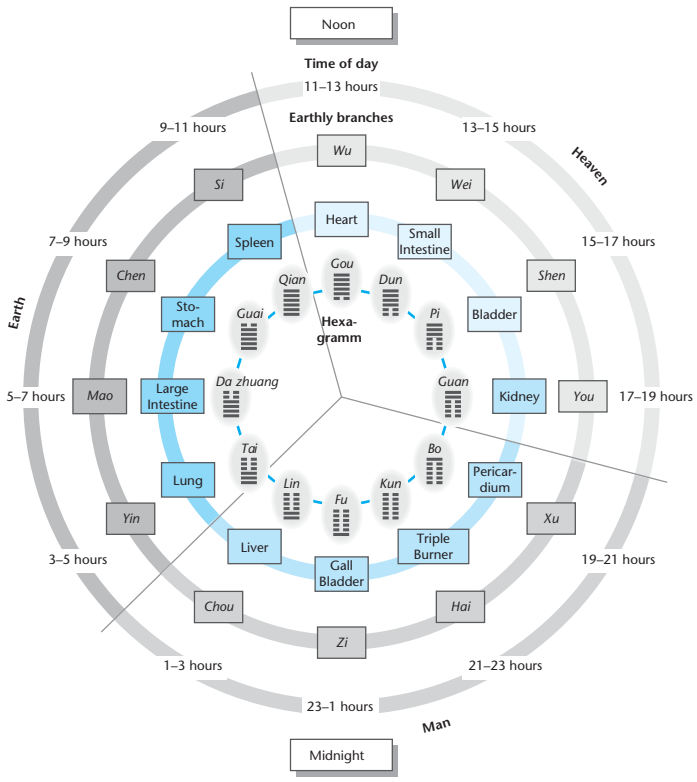

Fig. 1.8 Expanded Organ clock

Second circulation of the Nutritive Qi (*ying qi*)

According to chapter 16 of the *Ling Shu*, the Nutritive Qi (*ying qi*), after having circulated through the 12 primary channels, takes a deep pathway of the primary Liver channel, which originates at **LIV-14** (*qimen*), passes the Lungs, the posterior aspect of the neck, the inner nasal passages and reaches **Du-20** (*baihui*), then follows along the *du mai* on the back and the *ren mai* along the midline of the abdomen. At the approximate level of **Ren-22** (*tiantu*) the *ying qi* again reaches the Lung channel, which traverses the supraclavicular fossa (near **ST-12**). Here, a new circulation of the *ying qi* through the 12 primary channels starts at the beginning of the primary LU channel.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 8

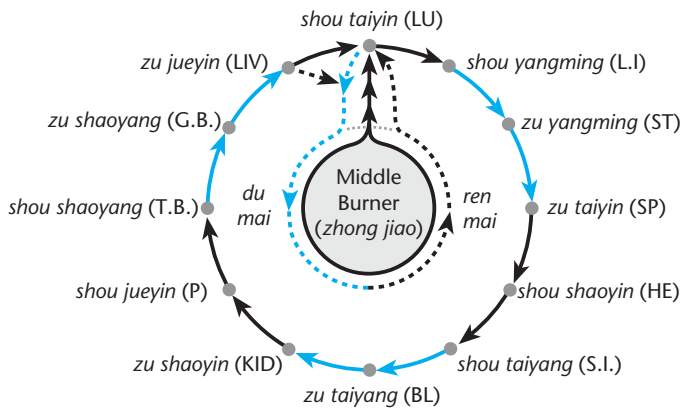

Fig. 1.9 First and second circulation of the Nutritive Qi (*ying qi*) with the Middle Burner

1.2 The 12 Primary Channels (*jing zheng*)

1.2.1 Background Information

Synonyms: Meridians, regular channels. '*zheng*' can be translated as 'mainly' (primary channels), but also relates to terms such as 'straight' and 'direct'.

Chinese terms for the channels

Meanings (also → 1.2.4, the six great channels (*liu jing*)

Tai: means greater, highest. The polarised energy (Yang or Yin) develops and reaches the maximum extent of its activity in the *taiyang* and *taiyin* channels. Here, the maximum extent of their respective polarity is reached and begins to fall again.

Shao: means less or younger. The polarised energy (Yang or Yin) is less in the *shaoyang* and *shaoyin* channels than in the *taiyang* or *taiyin* channels.

Ming: means clear, bright, radiant, shining (for more detail see → 1.2.4, the six great channels).

Jue: means 'absolute', 'at the end', 'exhausted', and according to Wiseman also 'inverted'. The polarised energy (Yin) in the *jueyin* channel represents the terminal phase of the Yin; here polarity changes to Yang.

Composition of the channel names

In the Chinese language, the name for each channel comprises

- the nature or **polarity of the energy** (Yin or Yang) that flows in the channel
- the **quality** or **intensity** (*tai*, *shao*, *jue*, *ming*) of the Yin or Yang energy and
- the **extremity** at which the channel originates or terminates.

Thus the Yin or Yang channels that begin or terminate at the hand carry the name of the primary hand channels (***shou*** *jing zheng*). The Yin and Yang channels that terminate or begin at the feet carry the name of the foot primary channels (***zu*** *jing zheng*) (→ Fig. 1.2).

1.2.2 Communication and Connections

Principle of the primary channel system

The 12 primary channels cover the body **bilaterally**. Each channel has its **individual regular course** with a deep, **internal** and a more superficial **external** pathway.

One can distinguish between Yin and Yang channels, which are Interiorly/Exteriorly paired. While each channel is connected with its pertaining *zang* or *fu* Organ, it is also connected with the Organ of its Interiorly/Exteriorly coupled channel.

All hand Yin channels begin in the region of the thorax and flow to the hand. All hand Yang channels start at the hands and travel to the head, where they meet the foot Yang channels. These descend down to the toes, where they meet the foot Yin channels, which ascend to the thorax and there meet the hand Yin channels.

For a better understanding of this model, imagine a person standing with his/her hands raised to the sky. In this position, all **Yin channels are ascending** (a **Yang phenomenon**), while all **Yang channels are descending** (a **Yin phenomenon**) (→ Fig. 1.10).

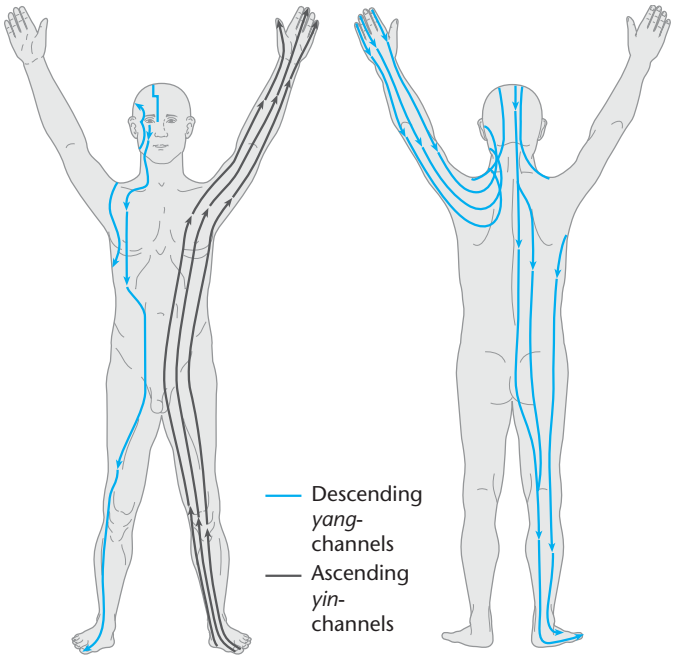

Fig. 1.10 Ascending and descending primary channels (according to Pirog 1996)

Ch01-F10028.qxd 2/22/08 2:30 PM Page 9

Each channel is associated with its own specific pathological symptoms, which are an important diagnostic tool in clinical practice (for channel-specific pathologies see → chapter 4).

Communication between channels and *zangfu*-Organs

The **connections (anastomoses)** between the channels have the function of allowing communication within the channel system as well as between the channels and the Organ systems. This external/internal (*biao-li*) relationship comprises the following aspects:

Communication from the Interior to the Exterior

- The **internal pathways of the hand Yin channels** connect with their pertaining Organ as well as with their Yin/Yang-paired *fu*-Organ. They flow from the thoracic region to the Exterior and connect with their paired Yang channel at the finger tips or on the hand (→ Fig. 1.11 a).
- The **internal pathways of the foot Yang channels** connect with their pertaining *fu*-Organ as well as with their Yin/Yang-paired *zang*-Organ. They then flow from the centre of the body to its Exterior and connect with their paired Yin–channel at the foot (→ Fig. 1.11 b).

Communication from the Exterior to the Interior

- **Each hand Yang channel** begins on the hand, runs to the thoracic region, penetrates the thorax and connects in the Interior of the body with its pertaining *fu*-Organ and its Yin/Yang-paired *zang*-Organ (→ Fig. 1.11 c).
- **Each foot Yin channel** starts at the foot, penetrates the abdomen, reaches the Interior of the body and its pertaining *zang*-Organ as well as its Yin/Yang-paired *fu*-Organ (→ Fig. 1.11 d).

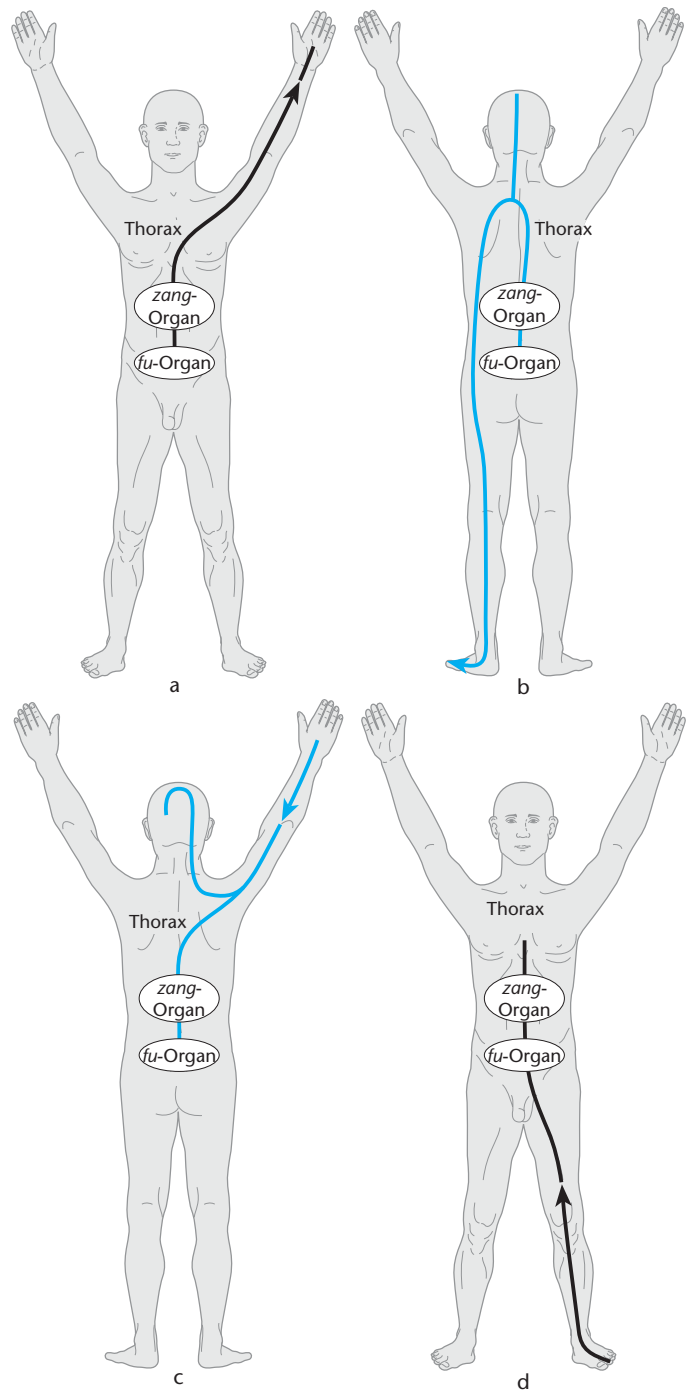

Fig. 1.11 Communication Interior → Exterior, hand-Yin channels (a). Communication Interior → Exterior, foot-Yang channels (b). Communication Exterior → Interior, hand-Yang channels (c). Communication Exterior → Interior, foot-Yin channels (d).

Ch01-F10028.qxd 2/22/08 2:30 PM Page 10

Channel circuits

Within the primary channel system we can distinguish three circuits, each circuit comprising four primary channels. The trajectories of two of the channels in one circuit are located on the Yin side of the body, the more anterior side (Interior), and two are found on the Yang side, the more posterior side of the body (Exterior). The Stomach channel (foot *yangming*, for more detail ➞ 1.2.3) presents the exception to this rule. Despite its major part running along the anterior side of the body, it is considered a Yang channel (➞ Fig. 1.12 and 1.3).

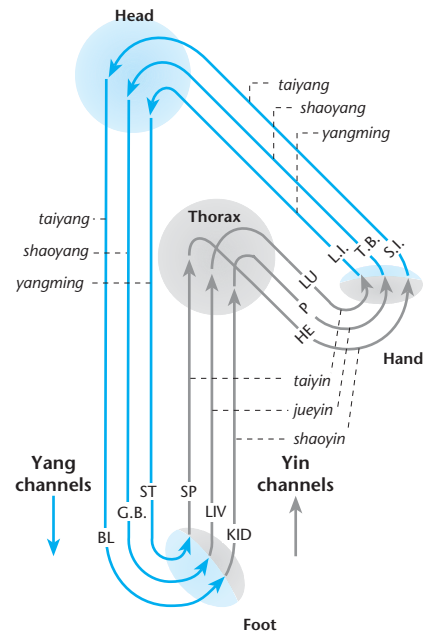

Fig. 1.12 Principle of the channel circuits

		First circuit			
Channel	**Time**	**Starts**	**Terminates**	**Hand/foot**	
LU	3–5 a.m.	Thorax	Finger tips	Hand (*shou*)	*taiyin*
L.I.	5–7 a.m.	Finger tips	Face	Hand (*shou*)	*yangming*
ST	7–9 a.m.	Face	Tips of the toes	Foot (*zu*)	*yangming*
SP	9–11 a.m.	Tips of the toes	Thorax (Heart)	Foot (*zu*)	*taiyin*
		Second circuit			
H	11 a.m.–1 p.m.	Thorax	Finger tips	Hand (*shou*)	*shaoyin*
S.I.	1–3 p.m.	Finger tips	Face	Hand (*shou*)	*taiyang*
BL	3–5 p.m.	Face	Tips of the toes	Foot (*zu*)	*taiyang*
KID	5–7 p.m	Tips of the toes	Thorax (Pericardium)	Foot (*zu*)	*shaoyin*
		Third circuit			
P	7–9 p.m.	Thorax	Finger tips	Hand (*shou*)	*jueyin*
T.B.	9–11 p.m.	Finger tips	Face	Hand (*shou*)	*shaoyang*
G.B.	11 p.m.–1 a.m.	Face	Tips of the toes	Foot (*zu*)	*shaoyang*
LIV	1–3 a.m.	Tips of the toes	Thorax (Lung)	Foot (*zu*)	*jueyin*

Ch01-F10028.qxd 2/22/08 2:30 PM Page 11

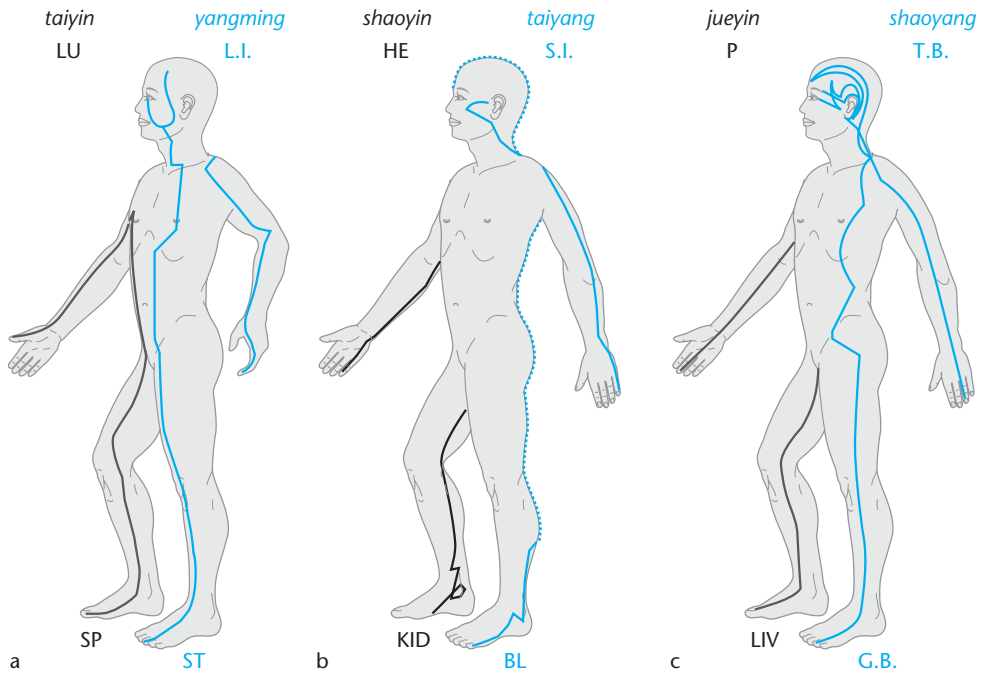

Fig. 1.13 Overview of the channel circuits

Connections among the primary channels

In order to ensure the continuous circulation (also see → 1.1.4) within the primary channel system, there need to be connections among the individual primary channels. Fig. 1.5 illustrates these in a schematic graphic overview.

According to some schools, the connections among the primary channels are facilitated by so-called entry/exit points. Various authors (for example Worsley, Jarrett, Pirog, Hicks et al., for more detail → 8.1.16) describe these as the shunting points between successive channels according to the Organ clock. The exit point marks the point on a channel from which the (internal) flow moves to a point (the entry point) on the channel that succeeds it on the Organ clock. There are some discrepancies regarding the location of the connections as described by the above authors and those described by Solinas et al. (1998), which are marked below as appropriate.

Yin–Yang connections

The connections – anastomoses – between the Yin and Yang channels are located on the hands and feet. The channels are most commonly connected at their respective final points, but sometimes also by branches separating from the main channel. These Yin–Yang connections are places where the Qi can change its polarity: Yin Qi becomes Yang Qi, and Yang Qi becomes Yin Qi. This creates a dynamic balance and allows for a better circulation of the Qi. The Yin–Yang connections follow the order of the basic structure of the circadian rhythm (24-hour cycle).

Hand Yin–Yang connections (→ Fig. 1.14)

The **hand Yin channels** are in most cases connected to the **hand Yang channels** at the tips of the fingers; sometimes they are connected on the hand by a branch that separates from the primary Yin channel and runs to the hand Yang channel:

FIRST CIRCUIT

LU channel (*shou taiyin*) ➡ L.I. channel (*shou yangming*)

LU-7 (*lieque*) ➡ **L.I.-1** (*shangyang*)

A branch separates from the primary LU channel at → **LU-7** (*lieque*) and travels to the primary L.I. channel at → **L.I.-1** (*shangyang*).

Note: According to Hicks et al. (2004, → 8.1.16) **L.I.-4** (*hegu*) is considered to be an entry point.

SECOND CIRCUIT

HE channel (*shou shaoyin*) ➡ S.I. channel (*shou taiyang*)

HE-9 (*shaoyang*) ➡ **S.I.-1** (*shaoze*)

THIRD CIRCUIT

P channel (*shou jueyin*) ➡ T.B. channel (*shou shaoyang*)

P-8 (*laogong*) ➡ **T.B.-1** (*guanchong*)

Ch01-F10028.qxd 2/22/08 2:30 PM Page 12

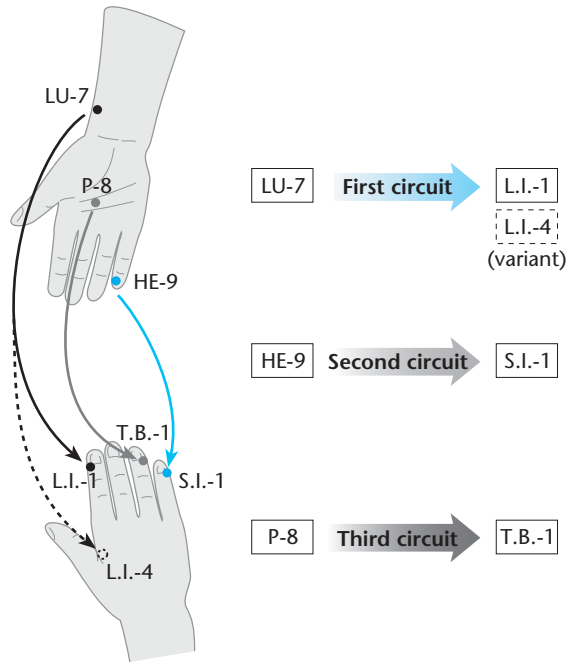

Fig. 1.14 Hand Yin–Yang connections (primary channels)

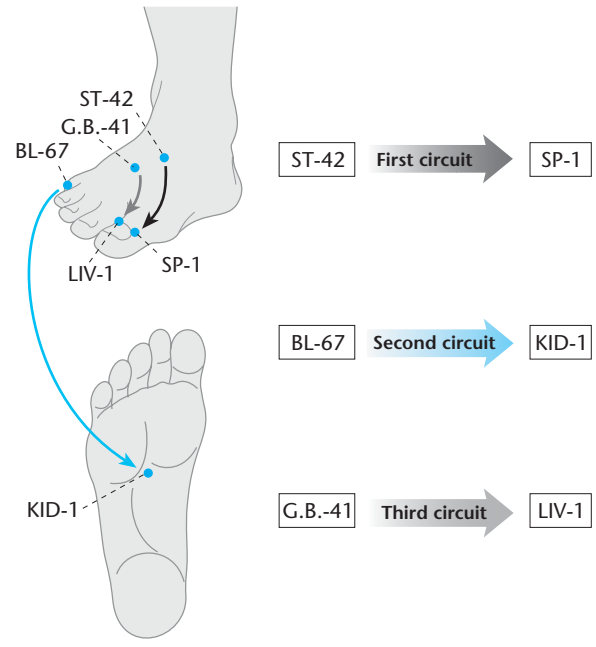

Fig. 1.15 Foot Yang–Yin connections (primary channels)

Foot Yang–Yin connections (→ Fig. 1.15)

The **foot Yang channels** are connected to the **foot Yin channels** at the toes by branches that separate from the primary Yang channels and run to the foot Yin channels (→ Fig. 1.15):

FIRST CIRCUIT

ST channel (*zu yangming*) ➡ SP channel (*zu taiyin*)

ST-42 (*chongyang*) ➡ **SP-1** (*yinbai*)

At → **ST-42** a branch separates from the primary ST channel and travels to → **SP-1**.

SECOND CIRCUIT

BL channel (*zu taiyang*) ➡ KID channel (*zu shaoyin*)

BL-67 (*zhiyin*) ➡ **KID-1** (*yongquan*)

THIRD CIRCUIT

G.B. channel (*zu shaoyang*) ➡ LIV channel (*zu jueyin*)

G.B.-41 (*zulinqi*) ➡ **LIV-1** (*dadun*)

On the instep of the foot, at → **G.B.-41**, a branch separates from the G.B. channel and travels to → **LIV-1**

Yang–Yang connections (→ Fig. 1.16)

The Yang–Yang connections between the primary channels are located on the **head**. They are rather superficial and follow the order of the channels within the **circadian cycle** (according to the Organ clock → Fig. 1.7). These connections support the communication between the hand Yang and foot Yang primary channels in the cranial region, which convey Qi of the same nature and quality (*yangming*, *taiyang*, *shaoyang*). They also mark the connections between the Yang axes, also referred to as hand–foot pairing (→ 1.2.3), and serve as a link between **above** and **below**.

FIRST CIRCUIT

L.I. channel (*shou yangming*) ➡ ST channel (*zu yangming*)

L.I.-20 (*yingxiang*) ➡ **ST-1** (*chengqi*)

SECOND CIRCUIT

S.I. channel (*shou taiyang*) ➡ BL channel (*zu taiyang*)

S.I.-18 (*quanliao*) ➡ **BL-1** (*yingming*)

THIRD CIRCUIT

T.B. channel (*shou shaoyang*) ➡ G.B. channel (*zu shaoyang*)

T.B.-23 (*sizhukong*) ➡ **G.B.-1** (*tongziliao*). Comment: according to Hicks et al. (2004, → entry/exit points, → 8.1.16) **T.B.-22** (*erheliao*) is considered an exit point.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 13

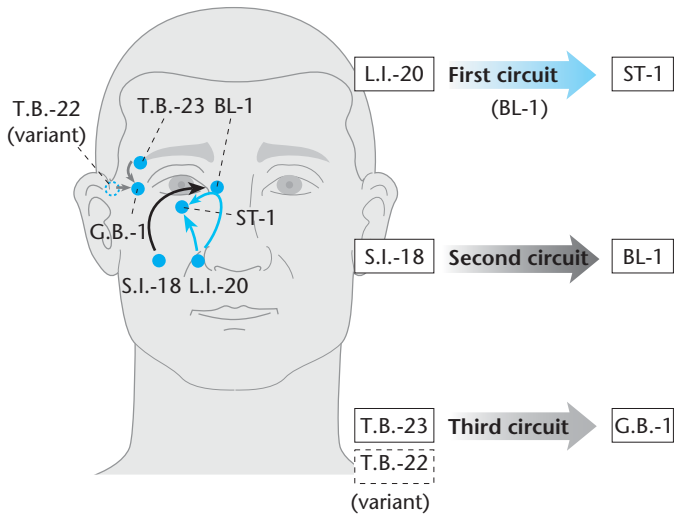

Fig. 1.16 Yang–Yang connections of the primary channels

Yin–Yin connections

The Yin–Yin connections between the primary channels are located in the **thoracic region** (Fig. 1.17).

In contrast to the more superficial Yang–Yang connections, they are located deep inside the body (internal pathways of the channels).

There are two types of Yin–Yin connection:

- the Yin axes connections or hand–foot pairings (*taiyin*, *jueyin*, *shaoyin*), which are **non-circadian** (do not flow according to the Organ clock) and
- the (deep) Yin–Yin connections, whose Qi flow follows the circadian rhythm of the Organ clock.

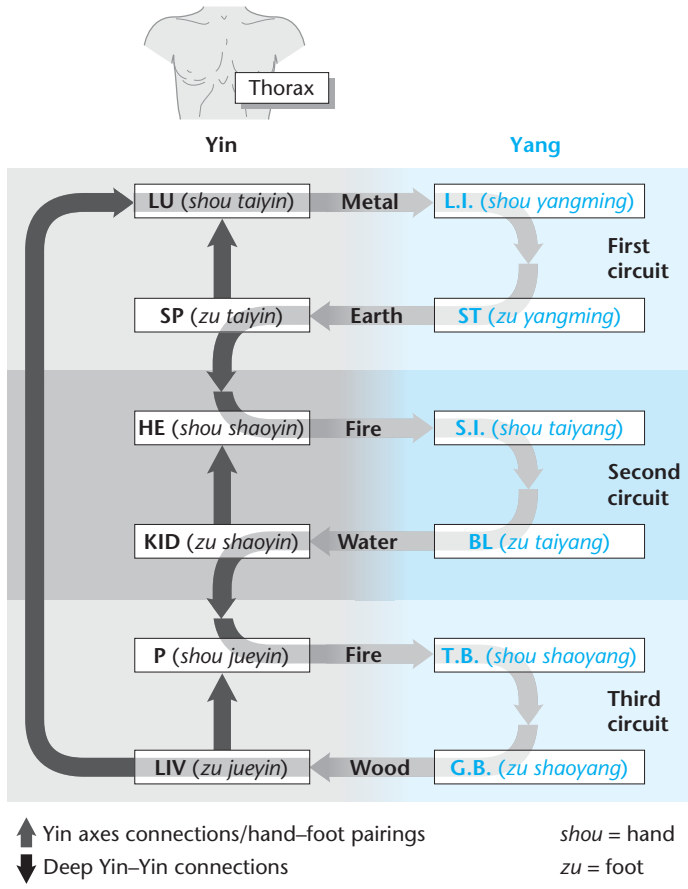

Fig. 1.17 Overview of the Yin–Yin connections (primary channels)

Connections of the Yin axes (hand–foot pairings)

These Yin–Yin connections support the communication between the hand Yin and the foot Yin channels, which convey Qi of the same quality (*taiyin*, *jueyin*, *shaoyin*) and serve as a link between **above** and **below**. Their Qi flow does **not follow the circadian rhythm** of the Organ clock (→ Fig. 1.17).

FIRST CIRCUIT

SP channel (*zu taiyin*) ➡ LU channel (*shou taiyin*)

SP-20 (*zhourong*) ➡ **LU-1** (*zhongfu*)

According to Solinas et al. (1998), a superficial branch separates from the primary SP channel at → **SP-20** and travels to the primary LU channel at → **LU-1**.

SECOND CIRCUIT

LIV channel (*zu jueyin*) ➡ P channel (*shou jueyin*)

LIV ➡ P

An internal branch ascends from the Liver, passes through the diaphragm, connects with the primary P channel below → **P-1** (*tianchi*).

THIRD CIRCUIT

KID channel (*zu shaoyin*) ➡ HE channel (*shou shaoyin*)

KID ➡ HE

An internal branch of the primary KID channel runs to the Liver, penetrates the diaphragm and spreads into the Lungs. From the

Ch01-F10028.qxd 2/22/08 2:30 PM Page 14

Lung a branch travels to the Heart, where it connects with the primary HE channel.
Many authors give **HE-1** (*jiquan*) as the site of connection.

Deep Yin–Yin connections

These Yin–Yin connections facilitate the **circadian cycle** according to the Organ clock and mark the transitions of the channel circuits. They are located deep in the thorax area on the level of the *zang*-Organs.

CONNECTION OF THE FIRST AND SECOND CIRCUIT
SP channel (*zu taiyin*) ➡ HE channel (*shou shaoyin*)
SP ➡ HE
An internal branch of the primary SP channel spreads in the Heart and connects with the HE channel. Note: according to Hicks et al. (2004, → entry/exit points, → 8.1.16) **SP-21** is considered to be the exit point of the SP channel and **HE-1** the entry point of the HE channel.

CONNECTION OF THE SECOND AND THIRD CIRCUIT
KID channel (*zu shaoyin*) ➡ P channel (*shou jueyin*)
KID ➡ P
An internal branch of the primary KID channel travels to the Kidneys, then to the Liver, penetrates the diaphragm and spreads into the Lung. From the Lung, an internal branch runs to the Heart, where it meets the primary P channel and also reaches → **Ren-17** (*shanzhong/danzhong*). Note: according to Hicks et al. (2004, → entry/exit points → 8.1.16) **KID-22** is considered to be the exit point of the KID channel and **P-1** the entry point of the P channel.

CONNECTION OF THE THIRD AND FIRST CIRCUIT
LIV channel (*zu jueyin*) ➡ LU channel (*shou taiyin*)
LIV ➡ P
An internal branch of the LIV channel originates in the Liver, passes the diaphragm, spreads in the Lungs and connects with the LU channel through a network of branches. Note: according to Hicks et al. (2004, → entry/exit points → 8.1.16) **LIV-14** is considered to be the exit point of the LIV channel and **LU-1** the entry point of the LU channel.

1.2.3 The Six Great Channels (*liu jing*)

Based on the Yang–Yang and Yin–Yin connections (→ 1.2.2, → Fig. 1.16, → Fig. 1.17), primary channels of the same nature and quality can be divided into **six great channels** (*liu jing*) or **hand–foot pairings (or axes).** They represent the three great Yang channels and the three great Yin channels.

Channels	Trajectory		Amount of Qi and Blood
Exterior			
Yang channels (hand–foot pairings)			
taiyang **(S.I., BL)**	Latero-posterior aspect of the four extremities, posterior aspect of the head and body	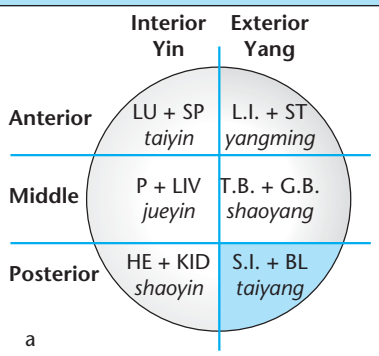	Less Qi, more Blood **Qi < Blood**
shaoyang **(T.B., G.B.)**	Lateral aspect of the four extremities and torso	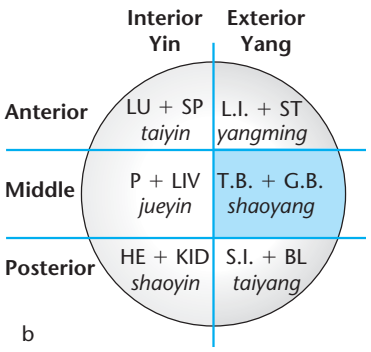	More Qi, less Blood **Qi > Blood**

Fig. 1.18 Overview of the six great channels (*liu jing*)

Ch01-F10028.qxd 2/22/08 2:30 PM Page 15

Yang channels (hand–foot pairings)			
yangming **(ST, L.I.)**	Latero-anterior aspect of the four extremities, frontal aspect of the head (face, forehead), anterior aspect of the trunk	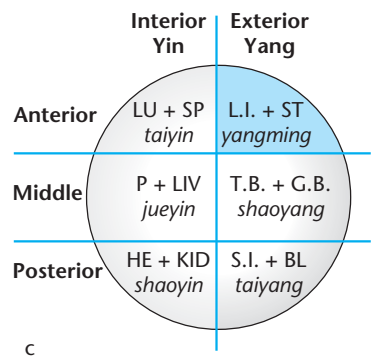 c	More Qi, more Blood **Qi = Blood**
Yin channels (hand–foot pairings)			
taiyin **(LU, SP)**	Medio-anterior aspect of the four extremities as well as the thorax and abdomen	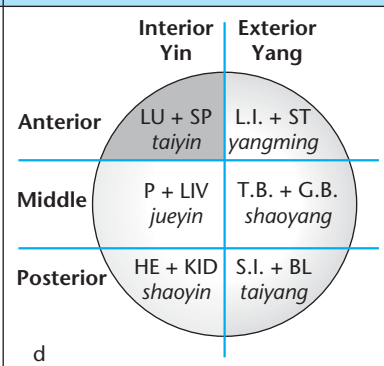 d	More Qi, less Blood **Qi > Blood**
jueyin **(P, LIV)**	Middle of the medial aspect of the four extremities as well as the thorax and abdomen	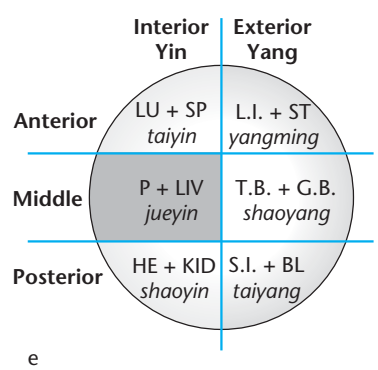 e	Less Qi, more Blood **Qi < Blood**
shaoyin **(HE, KID)**	Medio-anterior aspect of the four extremities as well as the thorax and abdomen	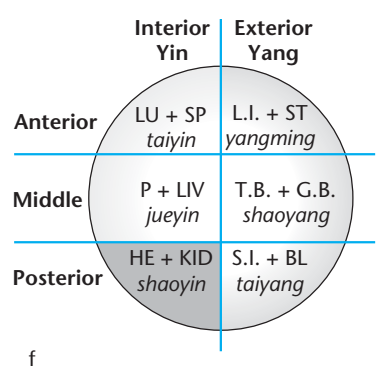 f	More Qi, less Blood **Qi > Blood**
Interior			

Fig. 1.18 (cont'd) Overview of the six great channels (*liu jing*)

Ch01-F10028.qxd 2/22/08 2:30 PM Page 16

Names and functions of the six great channels (*liu jing*)

Tai means 'greater', 'highest'

The polarised energy (Yang or Yin) develops and reaches its maximum activity in the *taiyang* or *taiyin* channels. Here, **the maximum level of the respective polarity** is reached and begins to fall again. '*Taiyang* opens to the outside' means that it spreads to the Exterior (→ see comments below under *yangming*, also → Fig. 1.19).

Shao means 'less' or 'younger'

The polarised energy (Yang or Yin) is less in the *shaoyang* or *shaoyin* channels than in the *taiyang* or *taiyin* channels. This also manifests in their location: the *shaoyang* channels are located midway between the **inside** (or anterior aspect) and the **outside** (or posterior aspect) of the body, acting like a hinge. According to Larre and Rochat de la Vallée (1986), *shaoyang* should be translated as 'young Yang'. In their view it is positioned between the outside (the other two Yang channels) and the inside (the other three Yin channels).

Depending on the interpretation, the *shaoyin* channels are considered either to be located between the *taiyin* and the *jueyin* channels or to be the deepest of the six channels. Thus they are either seen as a hinge or as a fixed central point (for more detail see → Depth organisation of the channels).

Ming means 'clear', 'bright', 'radiant', 'shining'

'*Yangming* closes towards the inside – *yangming* contracts towards the inside.'

The *yangming* channel has some unique features. Based on its course alone, it is **not 'Yang'** like the other two Yang channels, but it develops as a Yang phenomenon within the Yin aspect of the body (partially anterior pathway of the ST channel).

The character for *ming* is composed of the radicals for the sun and the moon. The character for Yang also contains the sun radical. According to the *Su Wen*, the *yangming* is where the two Yang shine together. Thus, the two suns can be symbolically interpreted as a doubling of Yang energy. *Taiyang* and *shaoyang* are 'Yang' simply due to their physical location. *Yangming*, however, is 'Yang' due to the content of its stored energy. This raises the question as to why the *yangming* channel is so abundant, why it carries so much Qi and Blood. Pirog (1996) explains this fact based on embryonic development. The 'closing' of the *yangming*' can be equated with the image of a curled-up **embryo** (→ Fig. 1.19). Because of its contracting, 'closing' position, the energy of the *yangming* channel can be compared to the compressed energy in a pressure cooker. Thus the 'Yin' process of contracting towards the Interior results in producing Yang or, in other words, compression of the energy in the Interior.

This naturally compressed energy is utilised in clinical practice. The *yangming* channels, but especially the ST channel, can be accessed, either by massage or by acupuncture, to supply the whole body with the energy stored and accumulated here. By the same token, the *yangming* channels can be needled in order to drain excess Heat – a procedure comparable to a controlled release of steam from a pressure cooker.

Jue means 'absolute', 'at the end', 'exhausted' (also, according to Wiseman, 'to revert')

The polarised energy (Yin) in the *jueyin* channel represents the terminal phase of the Yin; here polarity changes to Yang (→ see below for more detail).

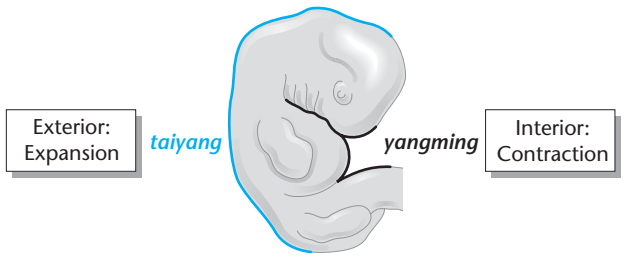

Fig. 1.19 Embryonic development of the *yangming* and *taiyang* channels (according to Pirog 1996)

The six great channels (*liu jing*): relationships and depth organisation

The systemic division of the channels into six levels or stages is mainly associated with the *Shang Han Lun*, a herbal classic. But according to Pirog (1996), this system can also be used as a paradigm for acupuncture, since it provides fundamental ideas regarding the functions of the points and channels as well as the relationships among the channels. It focuses particularly on changes of characteristics of the channel Qi, depending on which level it is acting within the body. The channels can be described as being stacked one upon each other, similar to layered rocks in a quarry. The depth of a channel or a point will be indicative of its function and significance. The postulate is that the more deeply a channel is located in the body, the more deep-seated its functions are and the more deep-seated illnesses it will be able to treat. They are not static, however. Depending on their location in the human body, they are indeed living structures with specific activities such as openers, closers, hinges or pivots, reflected in the meaning of the Chinese names. Fig 1.20 illustrates the arrangement of the six great channels from the outside to the inside, resembling the layers of an onion (→ Fig. 1.20).

Ch01-F10028.qxd 2/22/08 2:30 PM Page 17

Nguyen Van Nghi (1996) compares the three Yin and three Yang channels to doors. The door (*taiyang* or *taiyin*) is closed to prevent the entry of unwelcome guests such as external pathogenic factors. But the door can also be opened to let in friends. Whether the door opens and closes well depends largely on the hinges (*shaoyang*, *shaoyin*). Only if the hinges function smoothly can the energy circulate adequately. The doorframe (or closing mechanism) (*yangming*, *jueyin*) closes towards the Interior; in other words, it protects and supports what is contained on the inside. (→ Fig. 1.21).

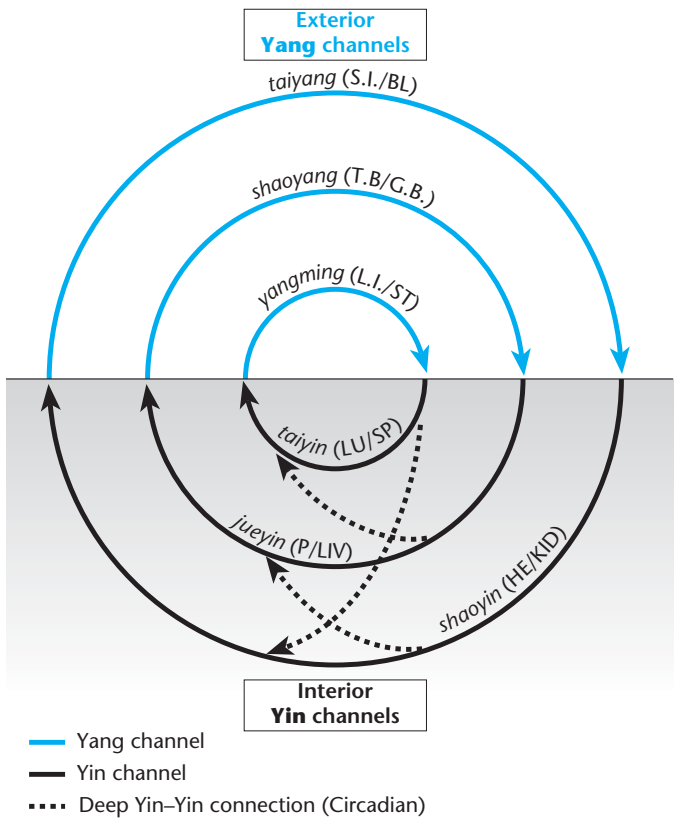

Fig. 1.20 Relationships of the three Yin and three Yang channels (according to Nguyen Van Nghi, 1996)

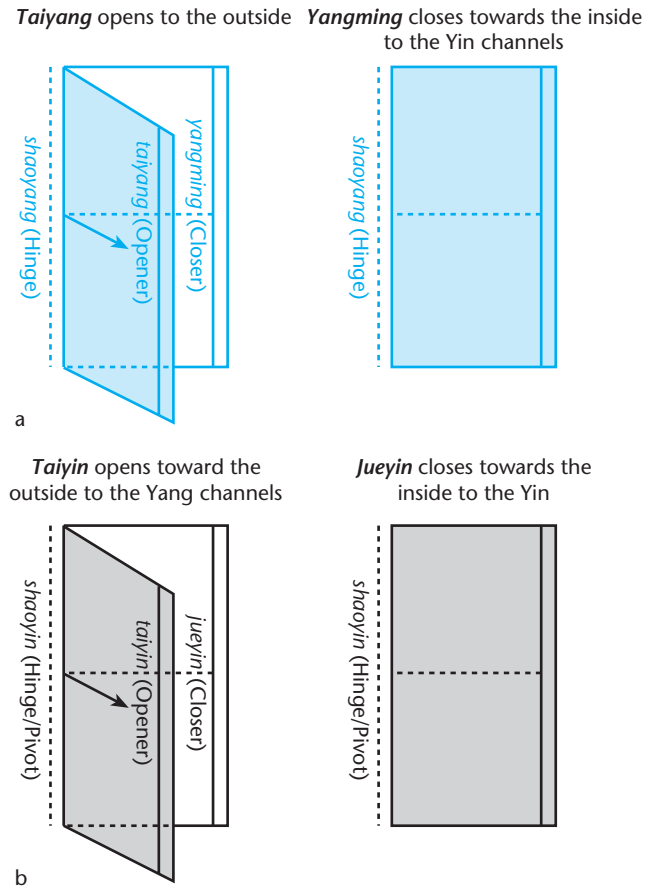

Fig. 1.21 The six great channels as opener, hinge and frame of a door

The literature is somewhat contradictory about which are the deeper of the six great channels, the *shaoyin* or the *jueyin* channels. Pirog (1996) has contributed an interesting theory in suggesting that the *shaoyin* channels occupy the deepest position in the body. They are located in such a medial location (near the middle of the body) that, when a person is standing in a normal position, the *shaoyin* channels are covered and not directly visible. According to Pirog, the *shaoyin* does not function like a hinge, but rather like a fixed pivot, similar to the axis of a mill wheel, which itself is immobile, but is able to cause movement. When the Qi has reached the deepest point of the *shaoyin*, and the 'Yin of the *jueyin* is exhausted', it has to revert. In other words: it has to return to a more superficial level on the outside or it has to transform into Yang. In that sense, it is the P and LIV channels that again start the flow of the Qi towards the Yang, from the inside to the outside and from below to above (→ Fig. 1.22).

Ch01-F10028.qxd 2/22/08 2:30 PM Page 18

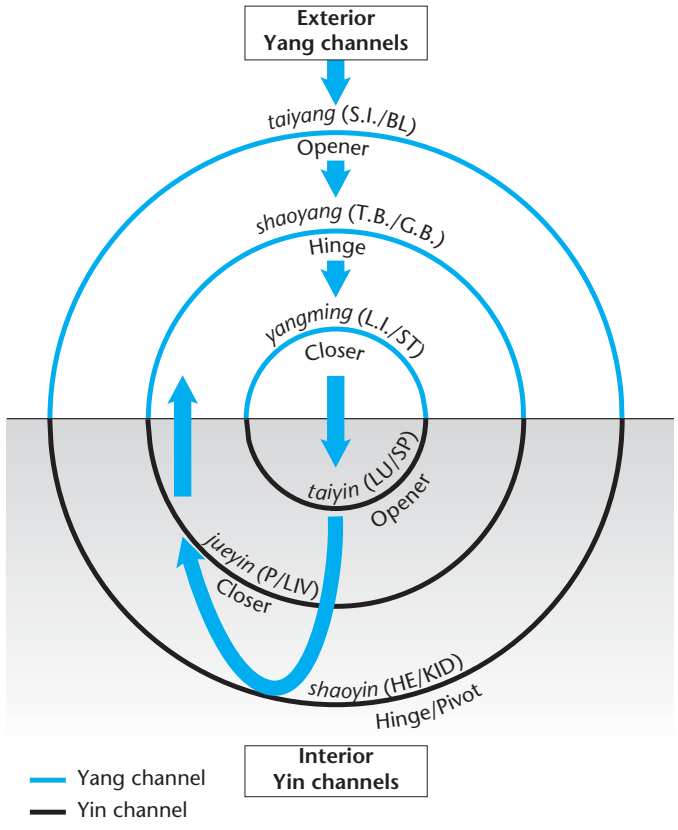

Fig. 1.22 Depth organisation of the six great channels (according to Pirog, 1996)

Confluent or connecting points of the six great channels (*liu jing*)

The **primary Yin channels** (*liu jing*) connect in the **thoracic region** (see also → Yin axes or hand–foot pairings → 1.2.2, Fig. 1.17). In contrast, the **primary Yang channels** connect in the **cranial region** (see also → Yang axes or hand–foot pairings → 1.2.2, Fig. 1.16). Many schools of thought suggest the connections of the channels of the same nature and quality at the following **six confluent points**:

Yin channels	**LU-1** (*zhongfu*), **P-1** (*tianchi*), **H-1** (*jiquan*)
Yang channels	**BL-1** (*jingming*), **G.B.-1** (*tongziliao*), **ST-1** (*chengqi*)

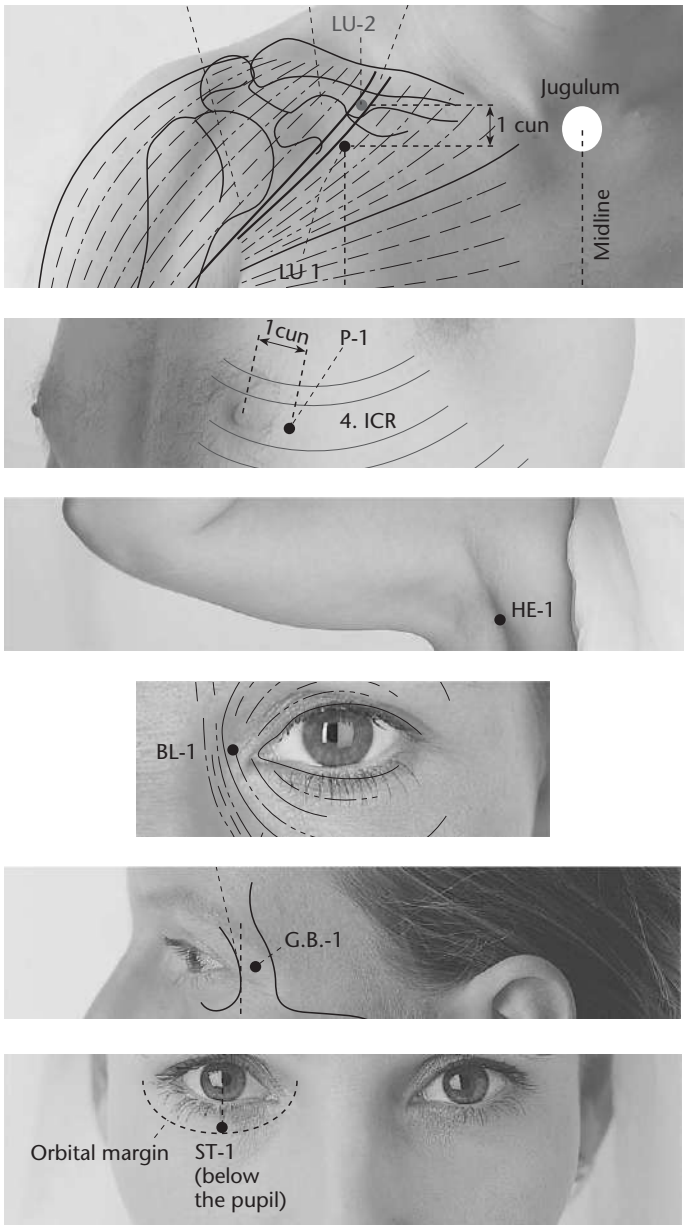

Fig. 1.23 Confluent points of the six great channels

Ch01-F10028.qxd 2/22/08 2:30 PM Page 19

Starting and end points of the six great channels (*liu jing*)

Fig. 1.24 shows the starting and end points of the six great channels on the feet or hands respectively.

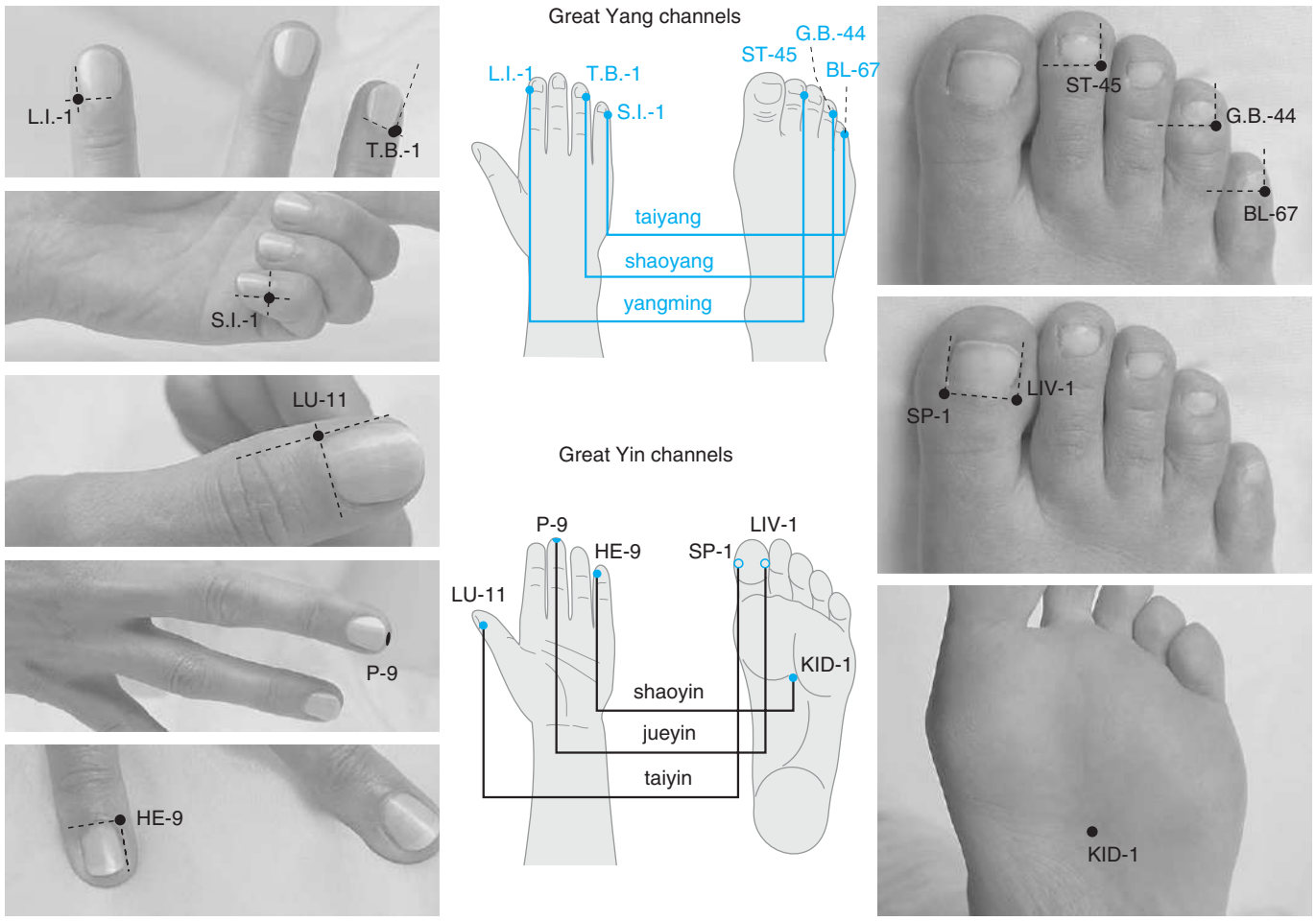

Fig. 1.24 Starting and end points of the *jiu jing*

Root points and binding points of the six great channels (*liu jing*)

According to Chapter 5 of the *Ling Shu*, each of the six great channels (*liu jing*) has a root point (*gen*) and a binding point (*jie*).

Root points (*gen*)

At the root point, the polarised energy (Yin/Yang) of a great channel is at its minimum – this point marks the end of the Yang or the beginning of the Yin. The root points are all located on the foot (→ Fig. 1.24).

Root points of the three Yang channels (→ Fig. 1.24)

These always represent the **terminal point** of each great **Yang** channel.

taiyang (S.I. ➡ BL)	**BL-67** (*zhiyin*)
shaoyang (T.B. ➡ G.B.)	**G.B.-44** (*zuqiaoyin*)
yangming (L.I. ➡ ST)	**ST-45** (*lidui*)

Root points of the three Yin channels (→ Fig. 1.24)

These always represent the **first point** of each great **Yin** channel.

taiying (SP ➡ LU)	**SP-1** (*yinbai*)
jueyin (LIV ➡ P)	**LIV-1** (*dadun*)
shaoyin (KID ➡ H)	**KID-1** (*yongquan*)

Binding points (*jie*)

The binding point is the point at which the polarised energy (Yin/Yang) of a great channel reaches its maximum level. It is at this point that the hand channel and the foot channel are connected.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 20

Binding points of the three Yang channels

These are the **switching points** of the great Yang channels (hand to foot) and are located in the **cranial region**.

taiyang (BL ➡ S.I.)	**BL-1** (*jingming*)
shaoyang (G.B. ➡ T.B.)	**T.B.-21** (*ermen*)*
yangming (ST ➡ L.I.)	**ST-1** (*chengqi*)**

Binding points of the three Yin channels

These are the **switching points** of the great Yin channels (foot to hand) and are located in the **thoracic or abdominal region**.

taiying (SP ➡ LU)	**Ren-12** (*zhongwan*)
jueyin (LIV ➡ P)	**Ren-18** (*yutang*)
shaoyin (KID ➡ H)	**Ren-23** (*lianqian*)

1.3 Divergent Channels (*jing bie*)

Synonyms: Channel divergence (Wiseman), separate pathways (Larre and Rochat de la Vallée).
Character: *bie* means diverging, separating, branching off. In this sense, the 12 divergent channels can be seen as branches of the primary channels.

General pathways of the divergent channels (→ Fig. 1.25)

The 12 bilateral divergent channels, which have no specific points of their own, are named after and are dependent on their corresponding primary channel.
They can be divided into **six great Yin and six great Yang pairs** and have their own system of six confluences (reunions): BL/KID, G.B./LIV, ST/SP, S.I./H, T.B./P, L.I./LU. According to Ramakers (course material, 2003) the divergent channels follow their own (non-cosmic) daily biorhythm:
BL ➡ G.B. ➡ ST ➡ S.I. ➡ T.B. ➡ L.I. ➡ KID ➡ LIV ➡ SP ➡ H ➡ P ➡ LU.
All divergent channels separate (*li*) from their corresponding primary channel on the extremities, except for the divergent T.B. channel, which branches off on the head. But many authors suspect that their pathways begin before the actual branching off. Based on this assumption, the divergent channels begin at the distal ends of the primary channels, running parallel to them, but as separate entities. The circulation in the divergent channels is of a centripetal nature (→ 1.1.1); in other words, the flow is always from a distal in a proximal direction, towards the trunk and the head. After separating from the primary channels, they enter deeper (*ru*) into the body and, after completing their internal pathway, they re-emerge close to the surface of the body (*chu*). The divergent channel pairs (Yin/Yang) pass together through the Interior of the body to join (*he*) their pertaining primary Yang channel in the occipital or cranial region. It is this joining of the courses that is referred to as the six confluences (or reunions) (*liu he*).

Pathways of the divergent Yin channels

After separating from their primary channels, they connect with their pertaining *zang*-Organ, then with the Interiorly-Exteriorly paired *fu*-Organ. They continue their flow in the upper half of the body together with their paired divergent Yang channel to join the paired primary Yang channel. In contrast to the divergent Yang channels, the divergent Yin channels will not return to their pertaining primary Yin channel after having completed their internal pathway (→ Fig. 1.25).

Pathways of the divergent Yang channels

After separating from their primary channels, the divergent Yang channels travel to their pertaining *fu*-Organ, then to the Interiorly-Exteriorly paired *zang*-Organ. They connect again with their pertaining primary channel and also with their Interiorly-Exteriorly paired divergent Yin channel in the upper half of the body (→ Fig. 1.25).

The divergent channels and the Heart

All divergent channels pass the thoracic region and the Heart, except for the divergent channels of the LU, L.I. and KID. According to Shima and Chase (2001), this means that all divergent channels passing the thorax will communicate with the True Qi (*zhen qi*), representing the synthesis of the Gathering Qi (*zong qi*) and Original Qi (*yuan qi*) (→ 1.1.4). They suggest that this hypothesis is in accordance with all the information available about the divergent channels, and explains furthermore why the divergent channels are so useful in the treatment of deep-seated Organ disorders as well as for disharmonies of the Defensive Qi (*wei qi*). It supports the image of a divergent channel as a fundamental axis between the Interior and the Exterior of the body. In the opinion of Shima and Chase (2001), the divergent channels are an invaluable connection to facilitate the return of the Defensive Qi (*wei qi*) to the Interior of the body and to transfer the fundamental inner Original Qi (*yuan qi*) to the periphery. In that respect the Original Yang (*yuan yang*) not only touches upon, but also strengthens the Defensive Yang (*wei-yang*).

* According to Nguyen Van Nghi (1996) and Solinas et al. (1998), this is G.B.-2 (*tinghui*)
** According to Nguyen Van Nghi (1996) and Solinas et al. (1998), this is ST-8 (*touwei*)

Ch01-F10028.qxd 2/22/08 2:30 PM Page 21

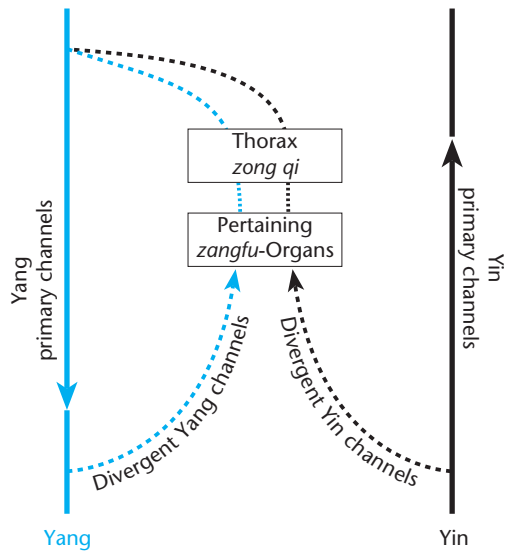

Fig. 1.25 Schematic illustration of the divergent channel pathways (according to Shima and Chase 2001)

Functions

The divergent channels reinforce the circulation of and communication between the primary channels in the Interior of the body. On the one hand, this is achieved due to their connection with the primary Yang channels (→ pathways). But their organisational structure also supplements the primary channels in a functional way – for example they also control areas of the body not covered by the primary channels. Thus points on the primary Yin channels can affect the head and face, in spite of their 'external' pathways ending in the thoracic region. This effect can be explained by the fact that the Yin divergent channels, after branching off from their pertaining primary Yin channel, pour into the Yang divergent channels flowing to the head. A deeper understanding of the trajectory and functions of the divergent channels will in turn lead to a better understanding of the clinical application of some commonly used acupuncture points. Since their pathway passes the Heart and thorax, the divergent channels further strengthen the connection to the Heart (→ explanations above). According to some schools, they can therefore also be used for the treatment of psychosomatic disorders. Many interpretations and variants exist regarding the pathways and clinical application of the divergent channels (for a good overview see → Shima and Chase 2001).

1.4 The 12 Sinew Channels (*jing jin*)

Synonyms and translations: muscle channels, muscular meridians/the muscles in connection with the meridians (Larre and Rochat de la Vallée), muscle regions, channel sinews.

Principles of the sinew channel system

The sinew channels mainly represent the muscles, sinews and ligaments located along the pathways of the primary channels and their pertaining connecting vessels. They do not have their own specific points and also no direct connection to the *zangfu*-Organs. The 12 bilateral sinew channels are connected to and named after their pertaining primary channel. Each sinew channel has its own pathway, which generally follows the trajectory of its pertaining primary channel. However, in contrast to the primary channels, the sinew channels always begin at the extremities, mostly separating from the primary channels at the *jing*-well points (→ 8.1.6) on the tips of the fingers and the toes. Their pathways always run from the extremities to the trunk or further to the head and face (→ centripetal circulation → 1.1.1). They cover a larger area than the primary channels, in the shape of a band and binding (*jie*) like a spindle at the joints and other areas of the body. This allows them to cover areas of the body which are not reached by the primary channels or the divergent channels. It further explains the actions of some of the points located on their pertaining primary channel. Under pathological conditions, the sinew channels will manifest as so-called *ashi*-points (*a shi xue*) or as trigger points.

Terms

Binding (*jie*) refers to spindle-shaped concentrations of the sinew channels, which tend to be located on the bigger muscles and joints (for more details on specific sinew channels see → Chapter 4).
The term *ju* (point of confluence, gathering) is used to signify an area where two or more sinew channels are bundled together. In the larger muscles, the Qi disperses or spreads (*san*), so that they can be divided into smaller muscle groups. The sinew channels connect (*luo*) the muscles, sinews and ligaments with the joints in order to facilitate mobility.

Rhythm of Qi flow

According to Ramakers (course material, 2003) the sinew channels are subject to their own daily rhythm of Qi flow:
BL ➡ G.B. ➡ ST ➡ S.I. ➡ T.B. ➡ L.I. ➡ SP ➡ LU ➡ LIV ➡ P ➡ KID ➡ HE.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 22

General pathways of the sinew channels (→ Fig. 1.26)

- The three hand Yang sinew channels (L.I./S.I./T.B.) flow from the finger tips towards the scapula, ascend to the ear and intersect with **G.B.-13** (*benshen*)
- The three foot Yang sinew channels (ST/BL/G.B.) run from the tips of the toes towards the head, join each other on the cheek and meet with **S.I.-18** (*quanliao*)
- The three hand Yin sinew channels (HE/P/LU) flow from the finger tips towards the thorax, joining within the thoracic musculature and meeting at **G.B.-22** (*yuanye*)
- The three foot Yin sinew channels (LIV/KID/SP) run from the tips of the feet towards the thorax, join each other in the genital region and, depending on the author, intersect in the area of **Ren-2** (*qugu*) or **Ren-3** (*zhongji*)

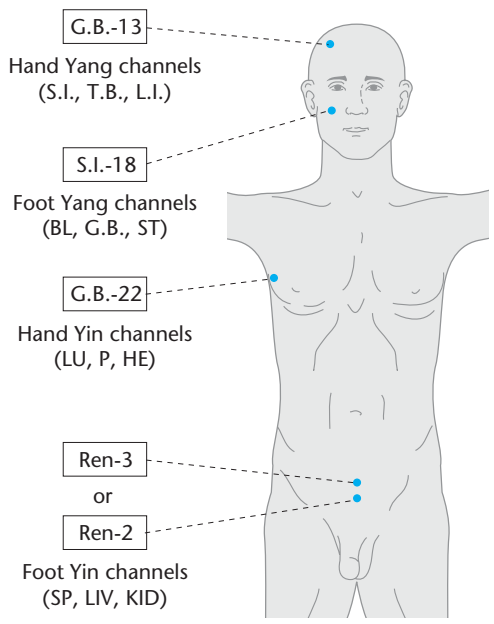

Fig. 1.26 Meeting points of the sinew channels

Functions

The main function of the sinew channels is to distribute Qi and Blood (*xue*) over the surface of the body as well as to integrate the muscles, sinews and ligaments with the joints. Generally, the sinew channels join groups of synergistic muscles, supporting their mobility and natural movement while also serving the function of integrating the surface of the body. At the same time they protect the bones and skeletal structure, linking the structures of the body.

Proper functioning of the muscles is dependent on the Spleen (*pi*) and the Liver (*gan*) and particularly on their Blood (*xue*). While the Yin (or material) aspect of the musculature, the 'flesh', is governed by the nourishing function of the Spleen, the functioning of the muscles and sinews is governed by the Liver. But the Liver is also connected to the mesenchymal structures of the body, especially those surrounding the organs. The sinew channels are not only composed of the muscular structures, but also of the mesenchymal formations, so that according to Larre and Rochat de la Vallée they are also involved in the structural aspect of the diaphragm as well as the thoracoabdominal serous membranes, in particular the pleura, peritoneum and pericardium. In the deeper parts of the body they act through the deep muscular fascia. One of the main functions of the sinew channels is that of a 'protective coat' that shields the body against attacks from the Exterior. In this respect the Defensive Qi (*wei qi*) is of particular importance ((→ 1.1.4): the muscle spindles are crossed by loose and diffuse flows of Qi, which also contain Defensive Qi (*wei qi*). Nutritive Qi (*ying qi*) is also present, particularly at the outer ends of the spindles (→ 1.1.4).

The sinew channels also have the task of barring access to the deeper channels. In other words, in a pathological situation, a deeper channel can only be reached if the resistance of the muscular layers can be broken. An important aim of Tuina therapy is to remove muscular tension. According to Larre and Rochat de la Vallée (1996), the sinew channels also play a role in psychosomatic disorders since they very effectively constitute the muscular carapace. The authors refer here to a bioenergetic interpretation of muscular tension.

Clinical and therapeutic importance

Disorders of the sinew channels will show the following symptoms along their pathway: muscular tension, pain, cramping, spasms, stiffness, but also paralyses, weakness and swellings of the muscles, tendons and ligaments. Any impairment of range of motion of the joints also indicates that the sinew channels are involved. In addition, each sinew channel also has its specific symptoms (→ chapter 4, individual channel descriptions).

The following factors can contribute to disorders of the sinew channels:

- *Bi* syndromes – external invasion of pathogenic Wind, Cold, Damp or Heat
- Traumatic injuries
- Muscle strain through over-use
- Muscle strain and contraction due to longstanding emotional and mental stress

While the sinew channels do not have their own specific points, they can easily be accessed by needling *ashi*-points, as well as through adjunctive therapies such as cupping, Tuina massage, plum blossom needling, moxibustion and *gua sha* (scraping technique). Distal points can also be used in treating disorders of the sinew channels.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 23

For traumatic injuries with limited range of motion, needling of distal points with simultaneous movement of the affected joint has proven to be very effective. The distal points (→ 8.2.1) can be chosen by first determining the affected area, then needling points in the corresponding area on the contralateral side.
For example, for a knee injury, the exact location of the pain is determined, then the corresponding point or area on the contralateral elbow is needled or massaged. Fig. 1.27 illustrates the areas that have been found to be effective. These areas are also referred to as reciprocal areas (for more detail also see → 8.2, 8.3).

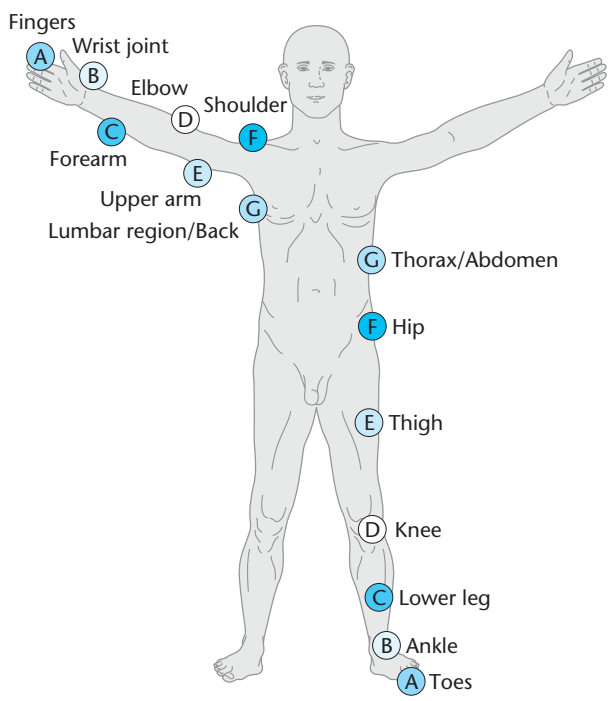

Fig. 1.27 Corresponding areas of the body

1.5 *Luo*-Connecting Vessels

Synonyms: secondary pathways (Larre and Rochat de la Vallée), network vessels, collaterals

Organisation

The *luo*-connecting vessels can be divided into 15/16 great *luo*-connecting vessels as well as into many smaller and more superficial *luo*-connecting vessels. Looking at the *luo*-connecting vessels altogether, they form a network that spans the entire body. And all *luo*-connecting vessels in turn connect with the great *luo*-connecting vessel of the Spleen (at → **SP-21**), which plays a major role within the network of the *luo*-connecting vessels (*Ling Shu*, Chapter 10). Together with the primary channel system, they form a network of vessels which distribute Qi and Blood not only within the body, but also to its surface (→ Fig. 1.3, 1.4).

The 16 great *luo*-connecting vessels

Traditionally there are 15 *luo*-connecting vessels, but if one includes the great *luo*-connecting vessel of the Stomach (*wei zhi da luo* or *xu li*), it raises the number to 16:

- 12 *luo*-connecting vessels (*luo mai*), related to the 12 primary channels and beginning at the extremities
- 4 *luo*-connecting vessels beginning at the thorax:
  - 2 *luo*-connecting vessels of the *ren mai* and *du mai*
  - 1 great *luo*-connecting vessel of the Spleen
  - 1 great *luo*-connecting vessel of the Stomach.

Except for the Great *luo*-connecting vessel of the Stomach (*xu li*) which originates in a *fu*-Organ, all other 15 *luo*-connecting vessels emerge at a specific point on their pertaining primary channel, the *luo*-connecting point (→ 8.1.2). 13 *luo*-connecting points are located on primary channels, two on extraordinary vessels (*ren mai* and *du mai*). The *luo*-connecting points are located on areas of the body, where the *luo*-connecting vessels that connect the Interiorly-Exteriorly pertaining channels with each other divide into smaller vessels. The 15/16 *luo*-connecting vessels thus fulfil a controlling function over all the other, smaller *luo*-connecting vessels. Each *luo*-connecting vessel has its own pathology according to its pertaining primary channel (→ introductions to the primary channels → Chapters 4 and 5).

Superficial *luo*-connecting vessels

The smaller superficial *luo*-connecting vessels are sub-branches of the 'great' *luo*-connecting vessels. They, in turn, are divided into the *sun luo*, vertical branches which are sometimes also referred to as tertiary branches or grandchild-vessels. The *sun luo* not only spread over the surface of the body, but also cover the internal Organs. The *sun luo* further divide into the more superficial *fu luo*. The *fu luo* again divide horizontally into small branches called the *xue luo* (blood vessel network).
Pathogenic factors tend to enter the body through the superficial *luo*-connecting vessels. By the same token, disorders such as Qi or Blood stagnation in a primary channel or an Organ may manifest in the *xue luo* as skin discolorations, spider veins, etc.

Clinical importance

The *luo*-connecting vessels can be accessed through superficial needling techniques as well as skin-sensitising techniques such as plum blossom needling, bloodletting, cupping, *gua sha* and moxibustion.

> New diseases are located in the primary channels; old (chronic) diseases lodge in the *luo*-connecting channels.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 24

1.6 Cutaneous Regions (*pi bu*)

Characteristics

The cutaneous regions form the most superficial aspect or outer cover of the primary channel system (*pi* = skin). According to Ramakers (course material, 2003), they can be compared to a 'big bag that holds everything together' and – besides the extraordinary vessels – they represent the oldest tissue in the body, which developed from the ectoderm. The cutaneous zones are not channels *per se*, but rather areas of skin connected to and covering the network of superficial channels and vessels. They are named after the six great channels (*liu jing* → 1.2.3) (Fig. 1.28).

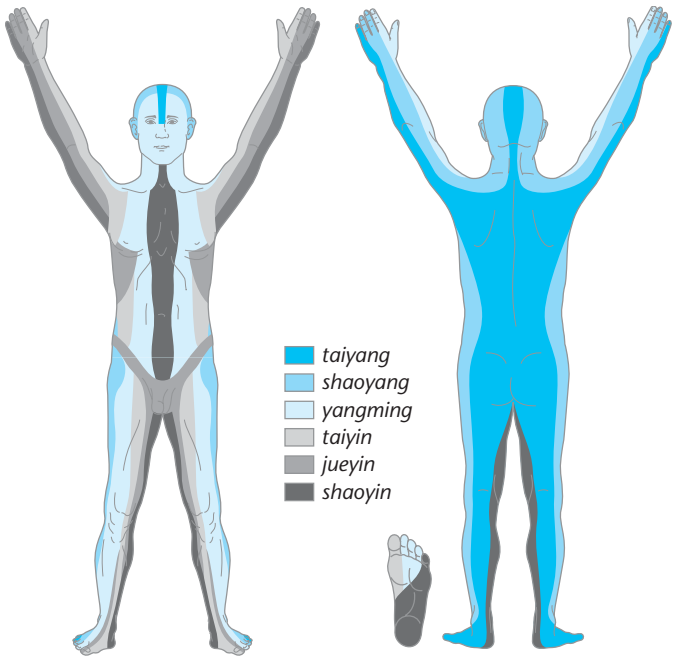

Fig. 1.28 Cutaneous zones (anterior and posterior aspect)

Functions

The cutaneous zones circulate Blood and Qi, especially Defensive Qi (*wei qi*), to the surface of the body and regulate the skin and pores, thus strengthening the body's defences. Proper functioning depends on the Lung Qi, which governs the skin, and the Defensive Qi (*wei qi*) (→ 1.1.4). Pathogenic factors can enter and exit the body through the cutaneous zones.

Clinical importance

Disorders of the primary channels and their pertaining *zangfu*-Organs may be reflected on the outside of the body by the cutaneous zones, for example pathological skin sensations, lesions or discolorations may occur. A bluish-greenish discoloration may indicate pain or obstructions, a reddish colour Heat and a pale-whitish colour deficiency or Cold. Therefore, the cutaneous zones can serve as a diagnostic tool for skin disorders or pain, specifically for chronic pain. Treatment techniques include superficial needling, plum blossom needling, cupping, massage, and *gua sha* (Chinese scraping technique). The author had good results in the treatment of chronic pain with the somatopic wrist/ankle acupuncture developed by Dr Zhang Xin Shu (→ JCM, No. 37,11/1991, → Focks and Hillenbrand 2003). The strictly subcutaneous needling used in this technique influences, of course, the cutaneous zones.

1.7 The Eight Extraordinary Vessels (*qi jing ba mai*)

Synonyms: Extraordinary meridians, Wondrous meridians
Meaning of *qi jing ba mai*:
Qi: extraordinary, strange, not paired, miraculous, wonderful
Jing: channels, meridians
Ba: eight (first differentiation)
Mai: vessels
The eight extraordinary vessels are not channels such as the *jing mai*, but they are – as the name implies – extraordinary. While they are often mentioned together with the primary channels (*jing mai* → Fig. 1.3), they still play a special role.

1.7.1 Theories about the Origin of the Extraordinary Vessels

The eight extraordinary vessels are often considered to be the deepest and most original structures of the body and channel system, developing very early on during the embryonic phase. They are considered to 'cover all of human activity' (Larre et al 1986). They also initiate all secondary functions and always remain operative. Therefore, their pathologies manifest with complex symptoms and include pre- and postnatal Essence problems. The *Nanjing* (Unschuld's translation, 1986) is the first to describe the extraordinary vessels as a complete system. Earlier records only mention them individually in the context of their respective functions.

Origins of the extraordinary vessels

In order for the manifestations of life to develop, the body requires structures that keep the rhythm of the Qi and its related processes in harmony and balance. The *Nanjing* introduces the concept of an energetic centre in the human body, the 'moving Qi between the Kidneys', which is often seen as connected with the *mingmen* and *qihai dantian* and is subject to the influences of the cosmic as well as of the prenatal and postnatal Qi.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 25

According to Matsumoto and Birch (1986), the 'moving Qi between the Kidneys' represents the *Taiji* of the body, the Great One, the origin of Yin and Yang. It is considered to be the root and origin of all human systems and structures and, according to some authors, it has a connection to the Original Qi (*yuan qi*) (→ 1.1.4). The *chong mai*, as the oldest vessel, is sometimes equated with the 'moving Qi'. The *ren mai* and *du mai* represent the basic division of the Yin and Yang in the body; in other words, they originate from the 'moving Qi' or *chong mai* as Yin and Yang branches. During the 'division', one movement ascends the front of the body (*ren mai*), the other movement the back (*du mai*).

Heaven and Earth have noon and midnight, the body has the *ren mai* and *du mai* as its polar axes. The *du mai*, *ren mai* and *chong mai* have different names, but in the end they are the same and have the same significance. The *chong mai* ensures the inseparability or unity of the *du mai* and *ren mai* – of Yin and Yang. If we tried to separate Yin and Yang, we would have to realise that they are an 'inseparable whole, a unit'.

Therefore, the extraordinary vessels function as a kind of matrix, which is the source of the energetic functioning and the cellular divisions in the body. This matrix is present from the moment of conception and organises the structuring of the embryonic development from the earliest stages.

Explanation of Fig. 1.29 (after Matsumoto and Birch)

In the beginning, there is the Great One, the *Taiji* (first level). At this level of 'no-form', the Great One begins to separate (second level) and becomes Yin and Yang (third level). At the fourth level, Yin and Yang separate twice (two branches):

- One branch **separates in a threefold way** to become the **Triple** Burner (*san jiao*) and its associated forms of Qi. Further separations lead to the formation of the **six** great channels (*liu jing* → 1.2), the **Five** Phases and **their source**, as well as the 12 primary channels.
- The other branch separates at the fourth level to form the **four** body areas (left, right, above, below) and then the **eight** extraordinary vessels.

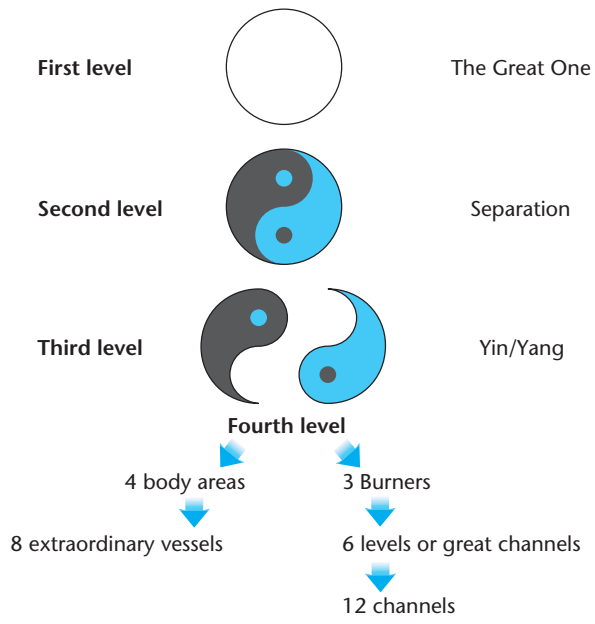

Fig. 1.29 Origin of the channels

Scars of the first cell division

Matsumoto and Birch (1986) comment on the work of Manaka and others, who interpret the extraordinary vessels with regard to embryonic development, regarding them as scars or 'lines of separation of the first cell divisions'.

The *du mai* and *ren mai* are the result of the first division of the fertilised egg, while the *dai mai* develops during the second division. At this stage, the embryo is essentially a torso, while the extremities and the more 'peripheral' extraordinary vessels – the *yang qiao mai* and the *yin qiao mai* as well as the *yang wei mai* and the *yin wei mai* – only form during subsequent divisions.

The eight extraordinary vessels and body structure

- According to Manaka et al (1995), a close connection exists between the eight extraordinary vessels and **body** structure/symmetry.
- The ***du mai*** and the ***ren mai*** separate the body **vertically** into **left** and **right**: the *ren mai* marks the anterior midline, the *du mai* ascends the posterior midline.
- The ***dai mai*** circles the waist **horizontally,** forming a **horizontal axis**. It separates **above** from **below** (→ Fig. 1.30 a).

In a further development of this idea, Manaka describes the division of the body into **eight areas,** using the structure of a topological octahedron for illustration. (→ Fig. 1.30 b). In this model, the eight extraordinary vessels directly relate to the lines 'separating' the faces of the octahedron:

- Due to their location on the anterior/posterior midline, the ***ren mai*** and the ***du mai*** separate the body into its right and left section.
- The ***dai mai*** separates above and below.
- The **G.B. and T.B. channels** compare to the anterior and posterior dividing lines at the lateral edges of the Yang aspect of the body.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 26

- The **P** and **SP channels** compare to the anterior and posterior dividing lines at the lateral edges of the Yin portion of the body.

The opening points of the extraordinary vessels, with the exception of **SP-4** (a point on the LIV channel would be more appropriate), represent this division:

- **P-6** and **T.B.-5** are the opening points of the *yin wei mai* and *yang wei mai* respectively
- **SP-4** and **G.B.-41** are the opening points of the *chong mai* and *dai mai* respectively.

These four channels divide the **anterior** and the **lateral** aspects of the body and connect the **upper** and **lower** parts, the four associated extraordinary vessels functioning accordingly. The *yin qiao mai* and the *yang qiao mai*, due to their connection to the KID and BL channels (opening points **KID-6** and **BL-62**) have a close relationship to the *ren mai* and *du mai*. Furthermore, the pathway of the KID channel is located close to the *ren mai* while the BL channel is close to the *du mai*.

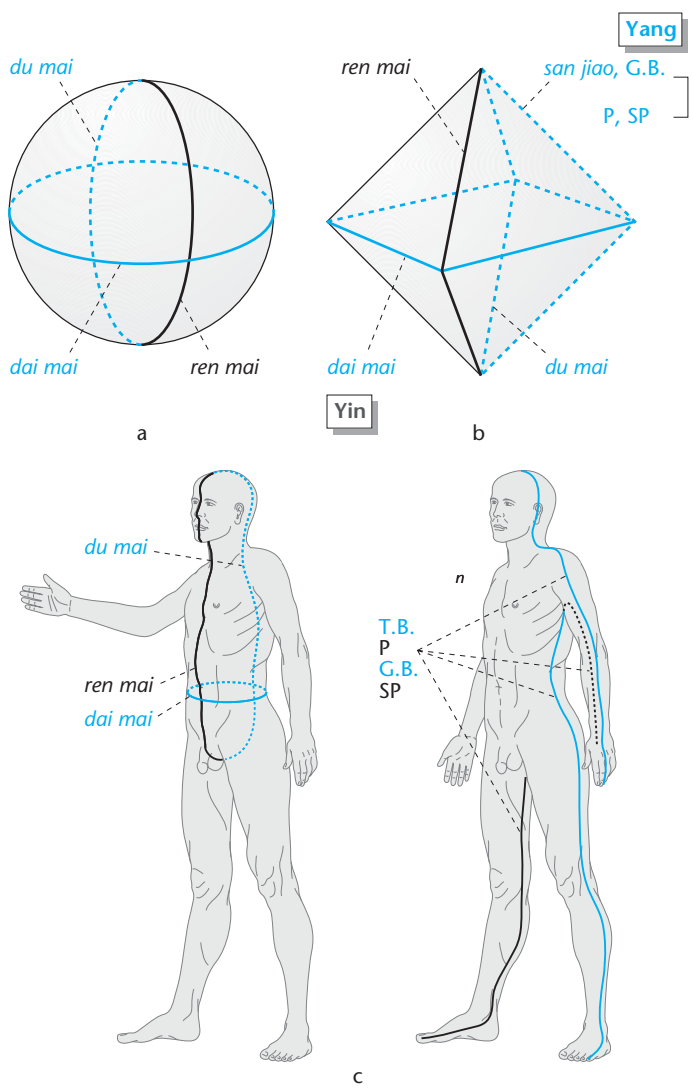

Fig. 1.30 Development of the octahedral structure based on body symmetry

1.7.2 Principles and Functions

'*The primary channels are the rivers, the extraordinary vessels are the lakes*.'

In some of the classics, the extraordinary vessels are compared to reservoirs, which store excess Yang, Yin, Qi or Blood, releasing it into the primary channel system in times of need. With the exception of the *ren mai* and *du mai* they do not have their own points, but they can be accessed through and are connected by points on the primary channels (see pathways of the primary channels → Chapter 5). All extraordinary vessels (with the exception of the *dai mai* with its horizontal trajectory) begin at the lower extremities or in the lower part of the torso and travel to the head. They can be activated by needling their opening and coupled points (differential therapy → 8.1.8).

Reservoir and compensation function

'*…when the network vessels (the primary channels) are filled to overflowing, none of the… conduits could seize any of their contents, and it is only then that the surplus contents of these vessels flow into the single conduit vessels (the extraordinary vessels)*' (*Nanjing*, translation by Unschuld 1986).

The *Nanjing* compares the extraordinary vessels to reservoirs, which are able to absorb a surplus of Qi and Blood in the primary channels, releasing them again during periods of deficiency. For example, if pathogenic factors penetrate the primary channel system, this excess can be absorbed by the extraordinary vessels. This will ensure that the *zangfu*-Organs remain unaffected. In that sense, the extraordinary vessels have a regulating compensation function.

Reproduction and fertility

The *chong mai* and the *ren mai* in particular, but also the *du mai*, play an important role in reproduction and fertility.

Connecting the primary channels

- The *chong mai*, also known as the '**sea of the 12 primary channels**', connects the primary ST and KID channels. It also strengthens the connection between the *ren mai* and the *du mai*.
- The *ren mai* as the **sea of Yin** connects all Yin channels
- The *du mai* as the **sea of Yang** connects all Yang channels at **Du-14** (*dazhui*).
- The *dai mai* circles the body at the waist and binds the vertical trajectories of all 12 primary channels in general, but especially the *chong mai*, the *ren mai* and the KID, LIV and SP channels.
- The *yin wei mai* **dominates the Interior** of the body and connects the KID, SP and LIV channels and the *ren mai*.
- The *yang wei mai* **controls the Exterior** of the body and connects the BL, G.B., T.B., S.I. and ST channels and the *du mai*.
- The *yin qiao mai* connects the KID and BL channels and controls balance.
- The *yang qiao mai* connects the BL, G.B., S.I., L.I. and ST channels and controls activity.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 27

In addition, the eight extraordinary vessels are regarded as **connecting channels for the extraordinary *fu*-**Organs.

Protection of the body

The *chong mai*, *ren mai* and *du mai* circulate the Defensive Qi (*wei qi*) in the thorax, abdomen and the back, and thus help to protect the body against external pathogenic factors. Due to their origin, they are also directly linked to the Kidneys, the pre-Heaven Qi and to the constitutional forces.

1.7.3 Comparison of the Eight Extraordinary Vessels with the Primary Channels

	Primary channels	Eight extraordinary vessels
Points, trajectory	Specific, bilateral points	Only the *ren mai* and *du mai* have their own points (and internal pathways); all other extraordinary vessels 'borrow' points from the primary channels. The four central vessels are unilateral (except for the *chong mai*, which also has a pathway on the extremities); the four peripheral vessels are bilateral.
Circulation	Continuously circulate *Nutritive Qi* and Blood	Circulate Essence Qi, Nutritive Qi and Defensive Qi, but according to some authors they do not have a continuous flow of Qi
Depth	Superficial, external pathway and a deeper, internal pathway; directly connected to their pertaining *zangfu*-Organs	Compared to the internal pathways of the primary channels they are more superficial, compared to the external pathways of the primary channels they are deeper. They have **no direct connection** to the *zangfu*-Organs, but their origin – especially of the four central vessels – lies deep within the body.
Connections	They connect the Exterior and Interior (through internal/ external pathways and through Yin/Yang channels) as well as above and below (hand–foot pairing, axes)	They do not connect the Interior with the Exterior, nor do they connect Yin and Yang.
Functions	Guide excess to the eight extraordinary vessels or to other secondary channels	Reservoir function: excess of Qi and Blood in the primary channels flows to the eight vessels and are stored there. In times of deficiency they are transported back into the primary channels.
Diagnosis	Diagnosis based on channel theory (meridian therapy), in TCM according to the *ba gang* (*zangfu*, pulse and tongue)	Diagnosis is characterised by specific arrays of symptoms and specific pulses.

Ch01-F10028.qxd 2/22/08 2:30 PM Page 28

1.7.4 Pairing and Special Functions

The eight extraordinary vessels can be paired in several ways (→ Fig. 1.31):

- The central vessels and the peripheral vessels can be coupled to form Yin/Yang pairs
- Or a central vessel can be paired with a peripheral vessel. These pairs support particular regions and are very important in acupuncture therapy (→ Opening points 8.1.8).

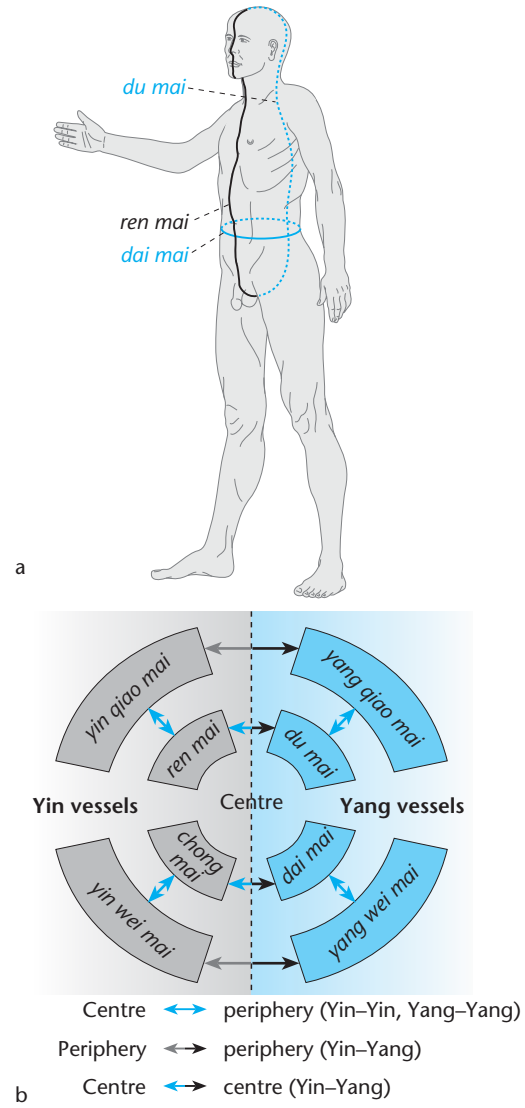

Fig. 1.31 Pairing of the extraordinary vessels (b, graphic overview modified according to Pirog 1996)

Ch01-F10028.qxd 2/22/08 2:30 PM Page 29

Pairing according to Yin/Yang polarity

Pairing of the extraordinary vessels according to Yin/Yang polarity

Yin vessels	Yang vessels	Functions of the Yin/Yang pair	General functions – Centre/periphery
Central vessels (pathways on the head and torso)			
ren mai	*du mai*	The *du mai* can be considered to be the first manifestation of the Qi and Yang, the *ren mai* the first manifestation of the Blood and Yin. This pair **stores the Yin Qi and Yang Qi**.	**Functions:** Storage and distribution (of Yin and Yang). Foundation of the eight extraordinary vessels, stores Essence Qi, is responsible for the support of the internal organs and extraordinary *fu* organs, as well as for the fundamental body structures
chong mai	*dai mai*	The *chong mai* is like a harmonious connection between Yin and Yang, it expresses vitality and sets impulses. The *dai mai* encircles; it is like a container and prepares the circulation of Yin and Yang. The pair **circulates the Essence (*jing*) and Qi between the Yin and Yang regions of the body**.	
Peripheral vessels (pathways from the lower extremities to the torso)			
yin qiao mai	*yang qiao mai*	This pair **creates balance between Yin and Yang** (for example balance between Essence (*jing*) and Qi, Blood and Qi, left and right, Nutritive Qi and Defensive Qi, also on the outside – between day and night), controls and regulates the ascending (*yin qiao mai*) and descending (*yang qiao mai*) movement in the body, is responsible for balance and rhythm between Yin and Yang.	**Functions:** Balance, circulation and connection (of the Yin and Yang aspects). Provides contact between the distal regions of the body and the torso and head.
yin wei mai	*yang wei mai*	The *yin wei mai* governs the Interior (*li*), it organises and creates connections within the *Yin* aspects of the body. The *yang wei mai* governs the Exterior (*biao*), it organises within the Yang aspects of the body. This pair **connects the Yin and Yang respectively.** 'That which moves and circulates between the *Yin* channels is called the *yin wei* – that which moves between the Yang channels is called the *yang wei*.'	

Central–peripheral pairing

Here, a central vessel is combined with a peripheral vessel. This pairing probably refers to regions of the body which are traversed and supplied by the pair.

Central–peripheral pairing of the extraordinary vessels

Pair	Vessel	Relationship	Opening point	Coupling point	Supported regions
Pair 1	*chong mai*	father	SP-4	P-6	Heart, thorax, Stomach
	yin wei mai	mother	P-6	SP-4	
Pair 2	*du mai*	husband	S.I.-3	BL-62	Inner canthus of the eye, occiput, shoulders and back, S.I. and BL channels
	ren mai	wife	BL-62	S.I.-3	
Pair 3	*dai mai*	man	G.B.-41	T.B.-5	Lateral eye and temple region, ears, cheeks, occiput and shoulders
	yang wei mai	woman	T.B.-5	G.B.-41	
Pair 4	*ren mai*	host	LU-7	KID-6	Face, throat, thorax, Lungs, diaphragm, abdomen
	yin qiao mai	guest	KID-6	LU-7	

2 Location Methods and Cun Measurements

Claudia Focks

One of the basic prerequisites for any successful acupuncture treatment, besides a differential Chinese Medicine diagnosis and the relevant point selection, is the correct point location.

2.1 Location Methods for Body Acupuncture

There are various methods for locating acupuncture points:
Anatomical approach: many acupuncture points are situated at clearly defined anatomical locations, for example in depressions, at muscle and tendon insertions, in grooves, at joint clefts, at bony prominences, etc. With some experience, the palpating finger will often automatically 'fall' into the right depressions and holes.
Various anatomical structures and zones form the basis for point location. These are described in more detail in Chapter 3. We can distinguish between the following:

- **Fixed markers** are not affected by the body position or by movement. These include bony landmarks such as depressions or prominences of the skeletal system, but also the finger- and toenails, the nipples, the umbilicus, etc. Most of the proportional cun measurements (→ 2.2) rely on fixed markers.
- **Movable markers** often become more clearly defined with a particular body posture or movement (→ 2.3.2). For example, by flexing the elbow, the cubital crease becomes much more clearly visible to locate **L.I.-11** precisely. Equally, the most distal palmar crease becomes much more visible when making a loose fist, in order to locate **S.I.-3**. Other parameters that might determine the location of a point include changed skin consistency, sensitivity to pressure, puffiness or a resistant sensation when palpating gently with a finger.

Proportional measurements: When trying to locate points not situated at any prominent structures, Chinese Medicine applies the proportional cun measurement (→ 2.2).
Electric tools: These measure the electric resistance of the skin in order to find the correct location of the points. Generally, electric resistance is lower in the immediate area around the point. This method is especially used in auricular acupuncture, but it has not proved to be a practical method for body acupuncture.
Very point technique (after Gleditsch 2005): The needle is held loosely and guided in a 'dancing' fashion over the skin. When the correct point – the 'very point' – is found, the needle will 'catch' on to it and penetrate the skin as if out of its own volition.

2.2 Cun Measurements of the Body

In Chinese medicine, distances on the body are traditionally measured in cun. In contrast to the official Chinese cun (1 cun = 2.5 cm), the cun used in a medical context is a **proportional unit of measurement** that takes the individual proportions of each body into account.
This relative unit of measurement is defined either by using the width of a particular finger or fingers (finger cun) or by the distances between clearly defined landmarks of the body (body cun).
In clinical practice, many points can be located by using **finger cun** measurements (→ Fig. 2.1). It is important, however, to use the **patient's fingers**, not the practitioner's fingers, for reference. If the fingers of both the practitioner and the patient have approximately the same width, the width of the practitioner's fingers can be applied for point location. Otherwise, measurements have to be adapted to correct minor deviations, for example by adjusting the spaces (narrow or wide) between the practitioner's fingers when using finger cun measurements.
The **body or proportional cun** is based on the proportions of particular sections of the body, which are divided into a specific number of units (→ Fig. 2.3). In clinical practice, the measurement of these divisions can be quite cumbersome. Therefore, the spreading hands technique is commonly applied (Koenig and Wancura 1979/1983). It allows the practitioner to quickly determine the midpoint of a given distance (→ 2.3.3) and, combined with finger cun measurements and palpation, is in most cases sufficient for correctly locating a point. For point location on the abdomen, a prepared elastic tape can be useful for measuring proportional cun (→ 2.3.1).

Finger cun

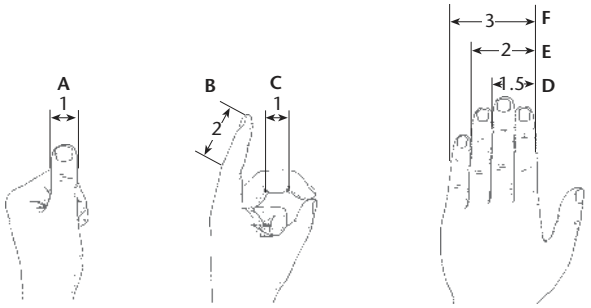

A:
1 cun:
thumb

B:
2 cun:
index finger

C:
1 cun:
middle finger

D:
1.5 cun:
finger-width

E:
2 cun:
finger-width

F:
3 cun:
finger-width

Fig 2.1

A	1 cun as thumb measure	The width of the interphalangeal joint of the thumb at its widest point
B	2 cun as index finger measure	The distance of the two distal phalanges of the index finger
C	1 cun as middle finger measure	With the tip of the middle finger placed on the thumb, the distance between the ends of the creases of the interphalangeal joints of the middle finger
D	1.5 cun as two-finger measure	The width of the index and middle finger when placed closely together, at the level of the distal interphalangeal joints
E	2 cun as index, middle and ring finger measure	The width of the index, middle and ring finger when placed closely together, at the level of the distal interphalangeal joints
F	3 cun as four finger measure	The width of the four fingers when placed closely together, at the level of the distal interphalangeal joints (=1 hand-breadth)

Body cun or proportional cun – the most commonly used proportional distances

Tip for locating the end of the axillary fold: The anterior or posterior axillary fold defines the end of the axilla, forming a fold in the skin (both anteriorly and posteriorly), where the arm joins the thorax. The axillary fold is the reference point for any cun measurements on the upper arm (→ Figs 2.2, 2.3, LU-3 and LU-4), which can be found by asking the patient to adduct his/her arm and to insert his/her flat hand into the axilla. The end of the axillary fold points in a superior direction, terminating at the lower border of the pectoralis major muscle, which in this area extends to its insertion in the humerus, below the deltoid muscle. The end of the axillary fold may vary considerably, depending on the individual proportions, positioning of the arm and flexing of the muscles. It should therefore be located with the patient's arms hanging down in a relaxed way.

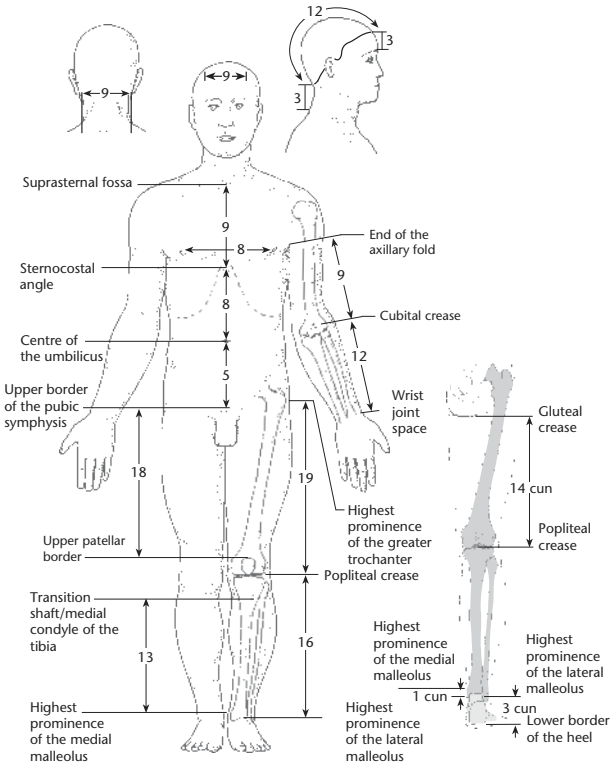

Fig. 2.2

Head		
Longitudinal	12 cun	From the anterior (→ 3.1.1) to the posterior hairline (→ 3.1.5)
	3 cun	From the midpoint of the eyebrows to the anterior hairline (→ 3.1.1)
	3 cun	From the lower border of the spinous process of C7 (→ 3.4.1) to the posterior hairline (→ 3.1.5)
	18 cun	From **Ex-HN-3** (***yintang***) to **Du-14**
Horizontal	9 cun	The distance between the mastoid processes (→ 3.1.4)
	9 cun	From corner to corner on the forehead (**ST-8** to **ST-8**)
Thorax and abdomen		
Longitudinal	9 cun	From the suprasternal depression (**Ren-22**) to the sternocostal angle (→ 3.5, xiphosternal synchondrosis)
	8 cun	From the sternocostal angle (→ 3.5, xiphosternal synchondrosis) to the centre of the umbilicus
	5 cun	From the centre of the umbilicus to the upper border of the pubic symphysis (→ 3.5)
Horizontal	8 cun	The distance between the two nipples
Laterally longitudinal	12 cun	From the end of the axillary fold to the tip of the free end of the 11th rib (**LIV-13**)
Back and lumbar region		
Both the spinous processes and the intercostal spaces are used for orientation. Main reference points are: the spinous process of C7, the inferior angle of the scapula (level with T7), the iliac crest (level with the spinous process of L4), the posterior superior iliac spine (PSIS) (level with S2 or the 2nd sacral foramen). Please note: the position of the patient (standing, sitting, lying) as well as individual anatomy and spinal structure (kyphosis, lordosis etc.) all influence the level of the vertebrae, which consequently may vary considerably in relation to other anatomical structures. For more detail → Chapter 3.		
Longitudinal	30 cun	From the spinous process of the T1 to the tip of the coccyx
Horizontal	3 cun	From the medial border of the scapula to the posterior midline (spinous processes), with the shoulder relaxed and hanging arms
Upper extremity		
Lateral	9 cun	From the end of the anterior axillary fold to the transverse cubital crease
	12 cun	From the transverse cubital crease to the wrist joint space (→ 3.3.3, transverse wrist crease)
Lower extremity		
Lateral	19 cun	From the highest prominence of the greater trochanter (→ 3.6) to the popliteal crease
	16 cun	From the popliteal crease to the highest prominence of the lateral malleolus (→ 3.6.2)
	14 cun	From **G.B.-34** (anterior and inferior to the head of the fibula) to the highest prominence of the lateral malleolus (→ 3.6.2)
	3 cun	From the highest prominence of the lateral malleolus (→ 3.6.2) to the lower border of the heel
Medial	18 cun	From the upper border of the pubic symphysis (3.5) to the upper border of the patella (→ 3.6.1)
	15 cun	From the medial end of the popliteal crease to the highest prominence of the medial malleolus (→ 3.6.2)
	13 cun	From the medial condyle of the tibia (→ 3.6.1, junction of the head and the shaft of the tibia) to the highest prominence of the medial malleolus (→ 3.6.2)
Dorsal	14 cun	From the gluteal crease to the popliteal crease
Difference between the medial and lateral malleolus		
	1 cun	Difference between the level of the highest prominence of the lateral and medial malleolus respectively (→ 3.6.2)

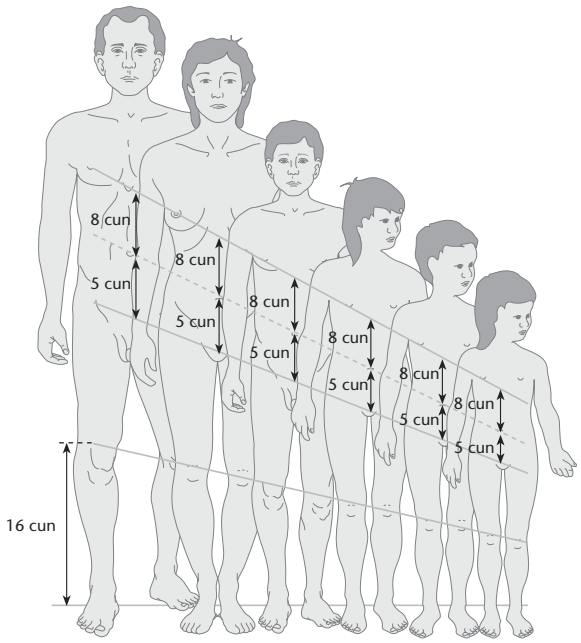

Fig. 2.3

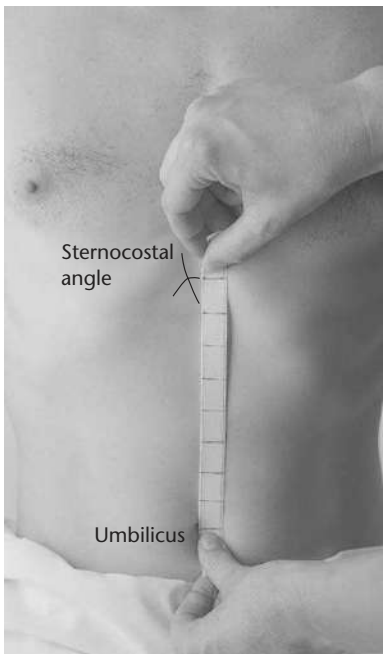

Fig. 2.4

2.3 Tips for Point Location

2.3.1 Elastic Tape

A marked elastic tape has proven a very useful tool for simplifying cun measurements and facilitating correct point location. It is especially recommended for measurements on the abdomen, the shape of which can vary considerably from patient to patient (Kitzinger 1995). An elastic tape, 1–2 cm wide and approximately 40 cm long, is marked at regular intervals of 2 cm, which can optionally be numbered from 1–20. The tape is placed onto the distance to be measured, stretching it according to the correct number of cun for that particular distance. This method allows the use of proportional cun for each individual patient.

Examples:

- For locating points on the upper abdomen, stretch the tape so that eight units cover the distance between the sternocostal angle and the centre of the umbilicus (→ Fig. 2.4).
- For locating points on the lower abdomen, stretch the tape so that five units cover the distance between the upper border of the pubic symphysis and the centre of the umbilicus (→ Fig. 2.5).

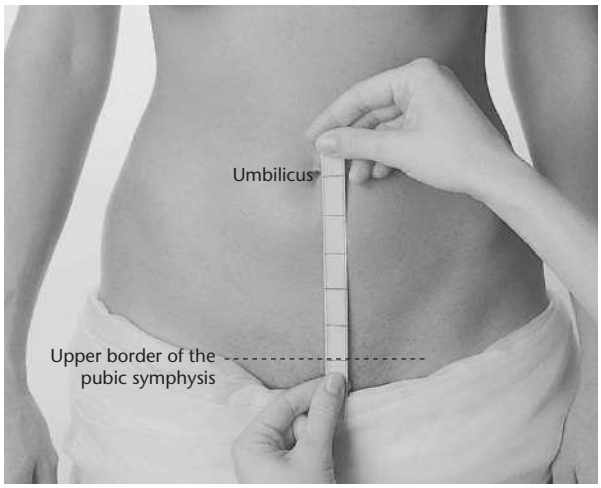

Fig. 2.5

2.3.2 Patient Positioning and Body Movements

To locate certain points, it can be very helpful to ask the patient to assume a particular position or to perform a specific movement.

Examples:

- For locating points on the **P** channel on the forearm, ask the patient to tightly press the thumb and little finger together. This will make the tendons more visible (→ Fig. 2.6)
- 'Tiger mouth grip' as location help for **LU-7** and **L.I.-6** (→ Fig. 2.7)
- For locating **L.I.-4**, ask the patient to press his/her thumb and index finger together. A muscle bulge will form; L.I.-4 is located on its highest point (→ Fig. 2.8).
- When turning the wrist joint from a pronated to a half-supinated position, the palpating finger will glide into a bony cleft just proximal to the styloid process of the ulna (**S.I.-6**, → fig 2.9).
- As the palpating finger glides along the skin, the skin bulge forming in front of it will 'arrest' the finger at the point to be located. This method is very useful for quick location of certain points such as **T.B.-5** or **P-6** (→ Fig. 2.10).

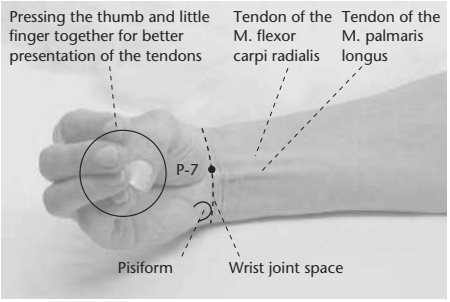

Fig. 2.6

Fig. 2.7

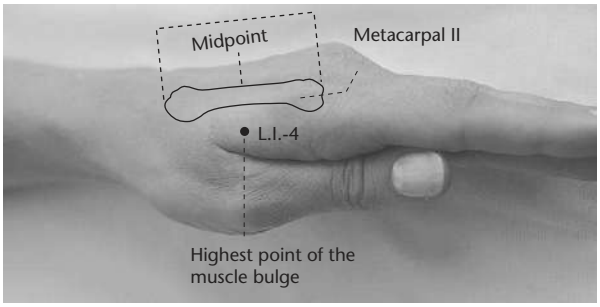

Fig. 2.8

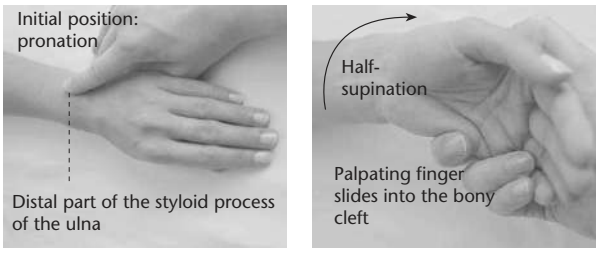

Fig. 2.9

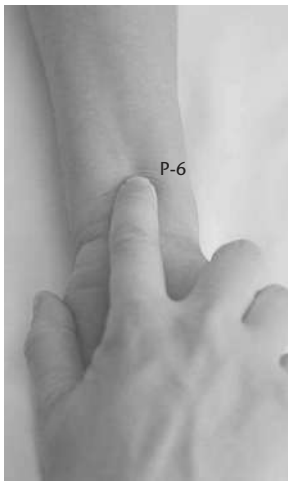

Fig. 2.10

2.3.3 Spreading Hands Technique According to König and Wancura

The spreading hands technique is particularly useful for quickly determining the midpoint of a given distance. For this purpose, the little fingers of each hand are placed on the endpoints of the distance to be measured. Both hands are then evenly spread over the distance. The thumbs are joined in the middle, indicating the midpoint of the distance.

Examples:

- To determine the midpoint of the 16 cun-distance between the popliteal crease and the prominence of the lateral malleolus, the little fingers are placed on both endpoints of this distance (→ Fig. 2.11)
- To locate **Du-20**, the hands are placed on the lateral aspects of the patient's head, with the little fingers touching the apices of the ears. The thumbs are then joined on the vertex of the head, and will be pointing to **Du-20** (→ Fig. 2.12).

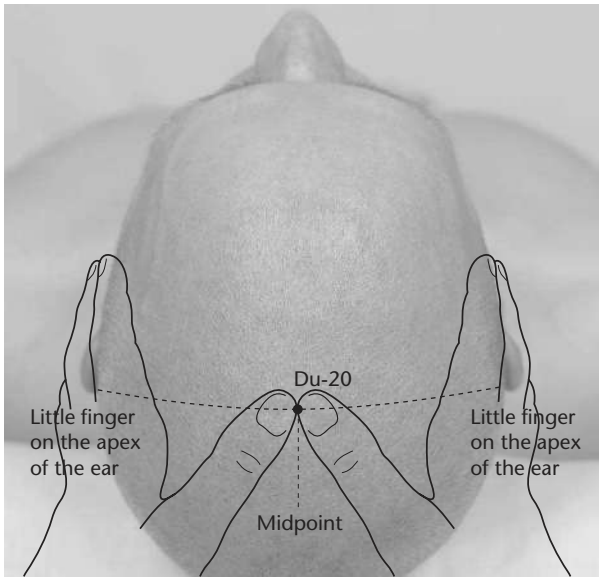

Fig. 2.12

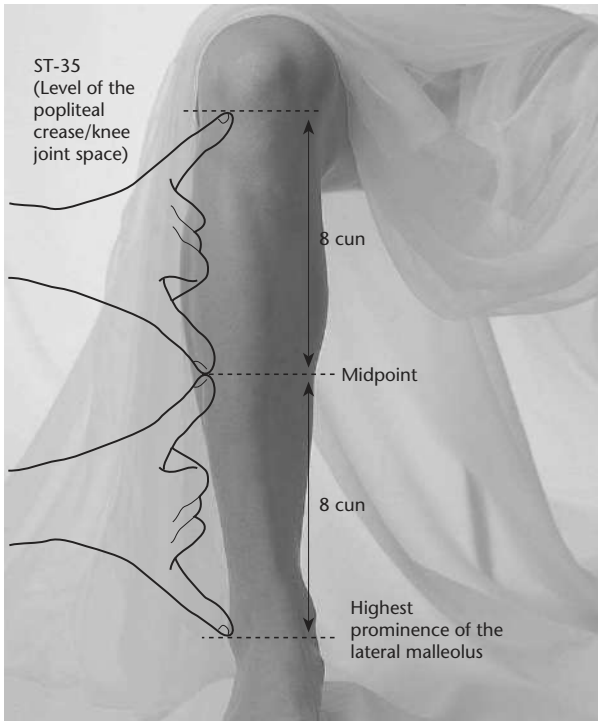

Fig. 2.11

3 Anatomical Orientation

Ulrich März

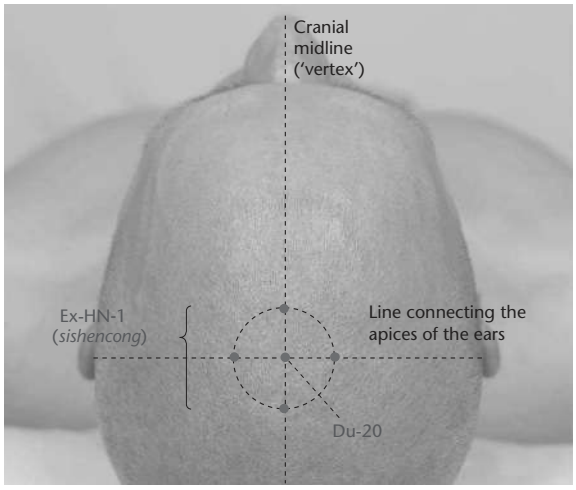

Fig. 3.1

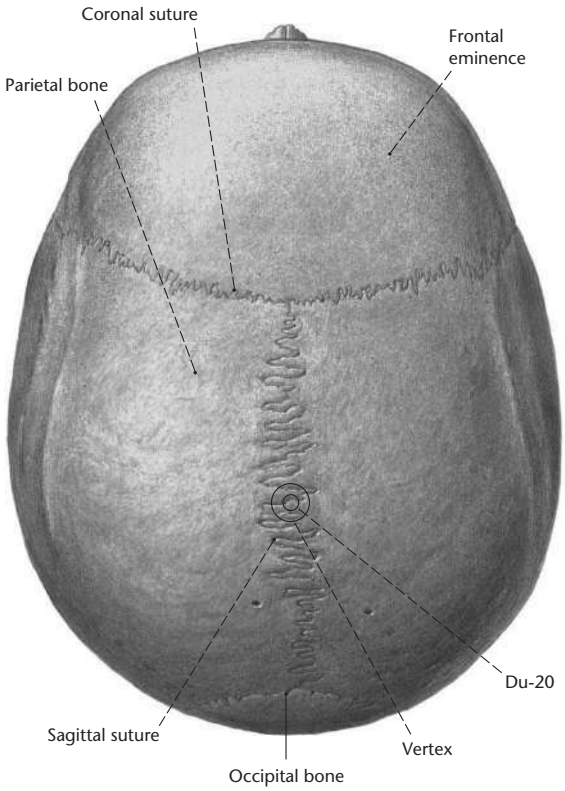

Fig. 3.2

3.1 Head

3.1.1 Vertex, Forehead and Supraorbital Region

Vertex (→ Figs 3.1, 3.2)

→ **Du-20** (*baihui*) is located at the junction of the cranial midline and an imaginary line connecting the apices of the ears. This point is at a distance of 5 cun from the anterior hairline and 7 cun from the posterior hairline, marking the highest point of the cranium ('vertex').

Please note: When locating **Du-20**, the patient's head should be in an upright position. Spreading hands technique (after König and Wancura, → Chapter 2): Place the hands on the lateral aspects of the patient's head, with the little fingers touching the apices of the ears. The thumbs are then joined at the cranial midline and will be pointing to → **Du-20** (*baihui*).

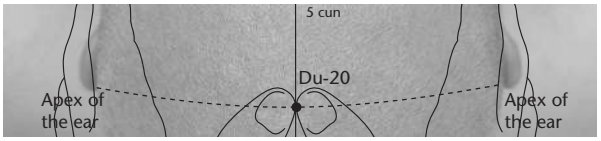

Fig. 3.3

Anterior hairline (→ Fig. 3.4)

The anterior hairline forms the superior border of the forehead and can vary considerably from person to person. The original line of hair growth is generally considered as the anterior hairline.
In men especially, the hairline may begin to recede early in life due to progressive baldness, so that the existing hairline may not be identical with the original line, which will now be located in an area not covered by hair. In these cases ask the patient to frown: the original hairline will be located superior to the wrinkled forehead, at the junction of the still slightly creased skin with the completely smooth skin. Points in this area are → **Du-20** to **Du-24**, **BL-3** to **BL-7, G.B.-4** to **G.B.-7, G.B.-13** to **G.B.-15, ST-8.**

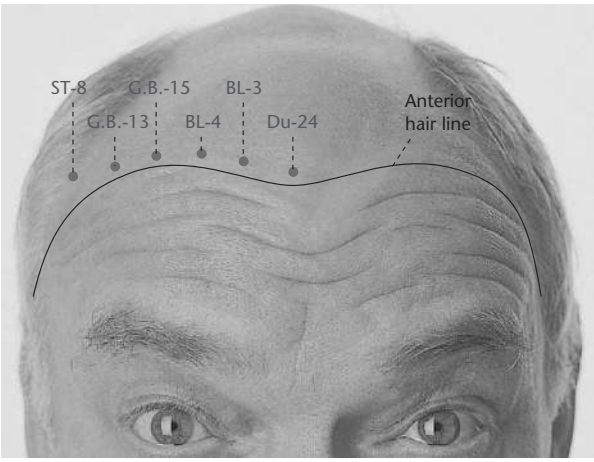

Fig. 3.4

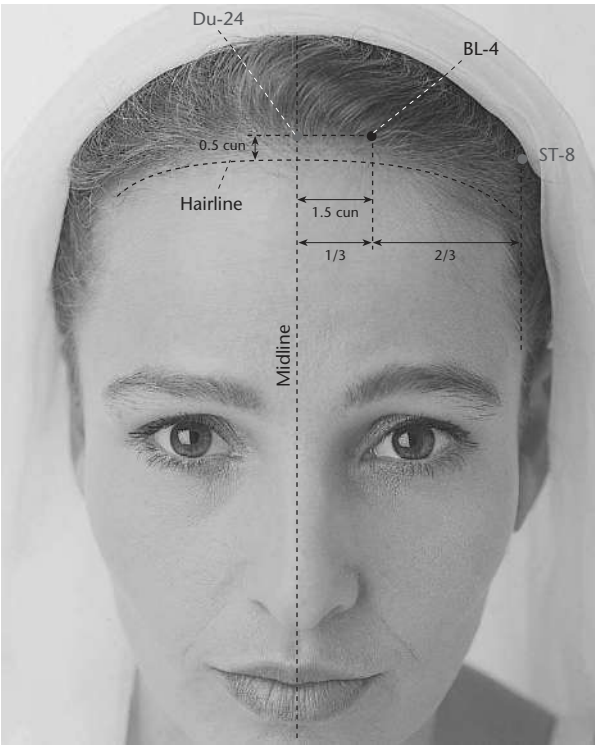

Fig. 3.5

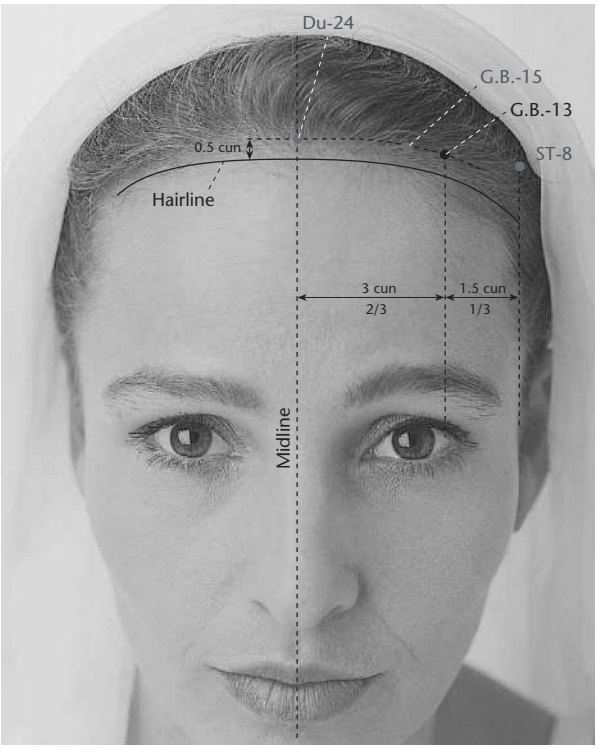

Fig. 3.6

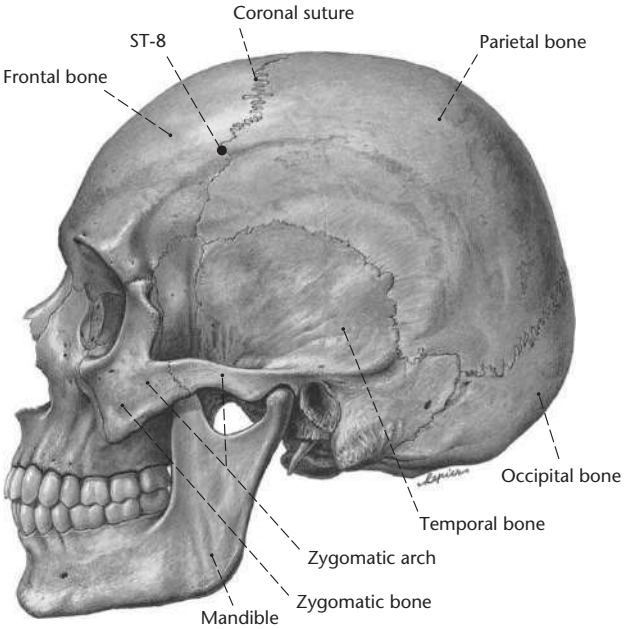

Fig. 3.7

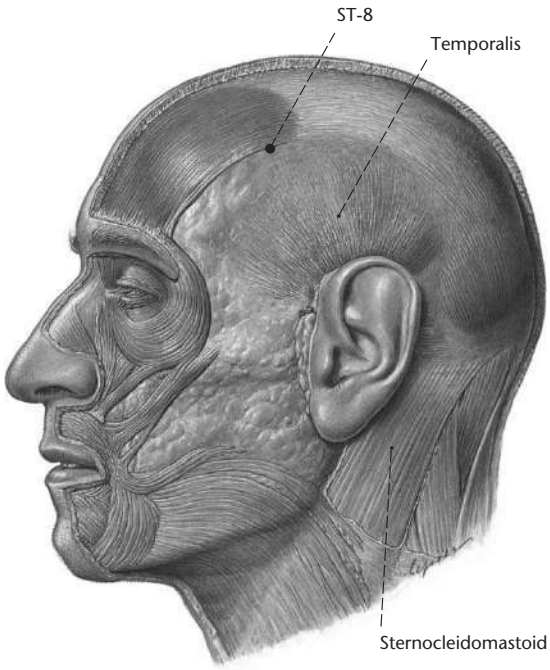

Fig. 3.8

Temporalis muscle and temporal corner of the forehead (→ Figs 3.7–3.9)

By asking the patient to firmly clench and release his/her teeth, the temporalis muscle can be palpated on the lateral aspect of the forehead. The corner between the forehead and the temples is marked by the junction of the anterior hairline with the temporal hairline, which tends to become more pronounced with a receding hairline.

→ **ST-8** is located in the centre of this area, at the insertion point of the temporalis muscle and 0.5 cun within the natural hairline. Often, the coronal suture can be palpated here.

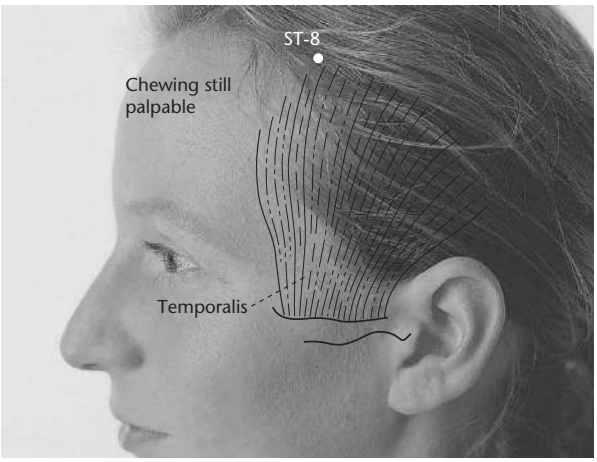

Fig. 3.9

Forehead (→ Figs 3.10, 3.11)

The more or less prominent bulge of the frontal eminences can be palpated on both sides of the forehead. → **G.B.-14** is located at their lower border, in a shallow depression at the junction of the eminences and the superciliary arches. With the patient looking straight ahead, G.B.-14 is located directly above the pupil, approximately 1 cun superior to the supraorbital margin.

Glabella, superciliary arches

The glabella is the central bony reference point on the forehead, forming a slightly convex, level area superior to the root of the nose and between the superciliary arches.

The extra point *yintang* (Ex-HN-3) is located at the centre of the glabella. Several points are located along the superciliary arches (from medial to lateral): → **BL-2, *yuyao*** (Ex-HN-4) and **T.B.-23.**

Please note: → **BL-2** is located superior to the inner canthus of the eye, in the area of the frontal (not the supraorbital) incisure, when this is present.

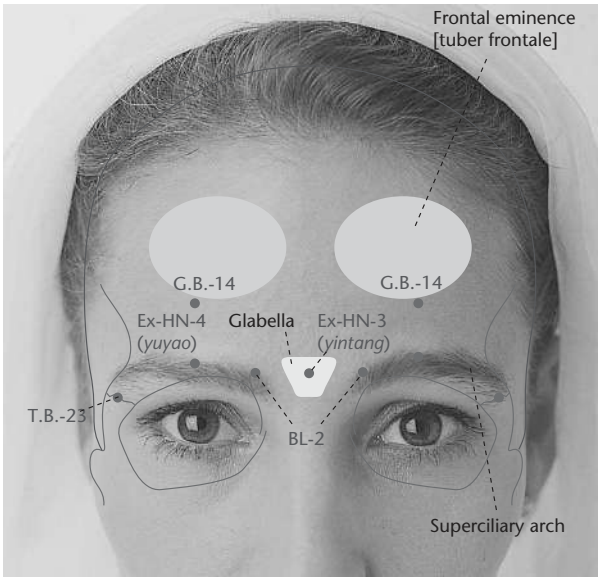

Fig. 3.10

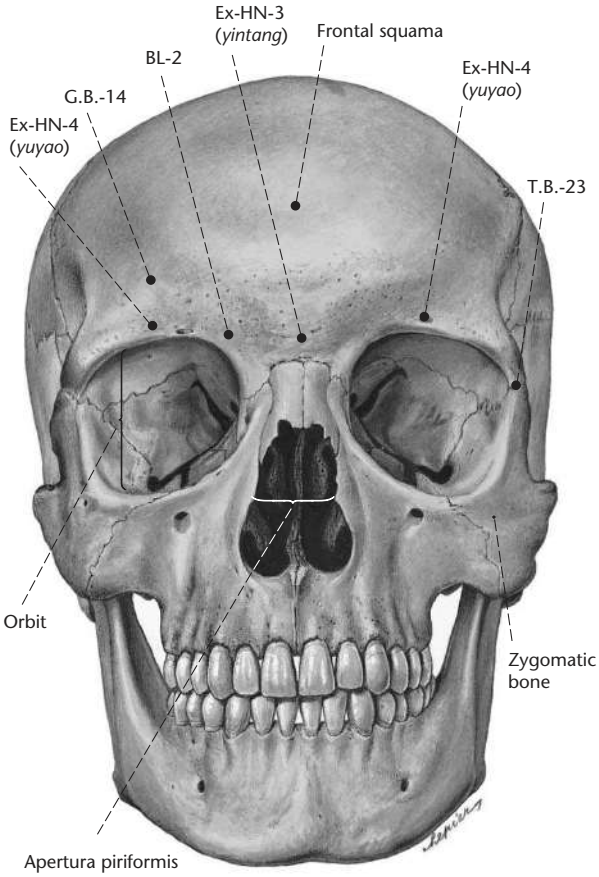

Fig. 3.11

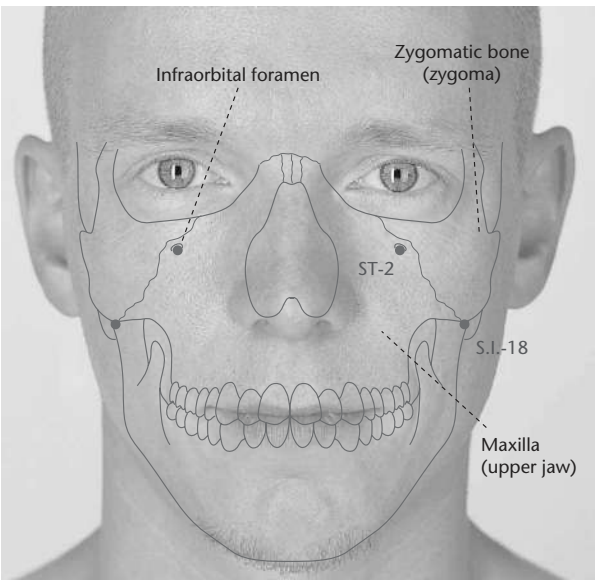

Fig. 3.12

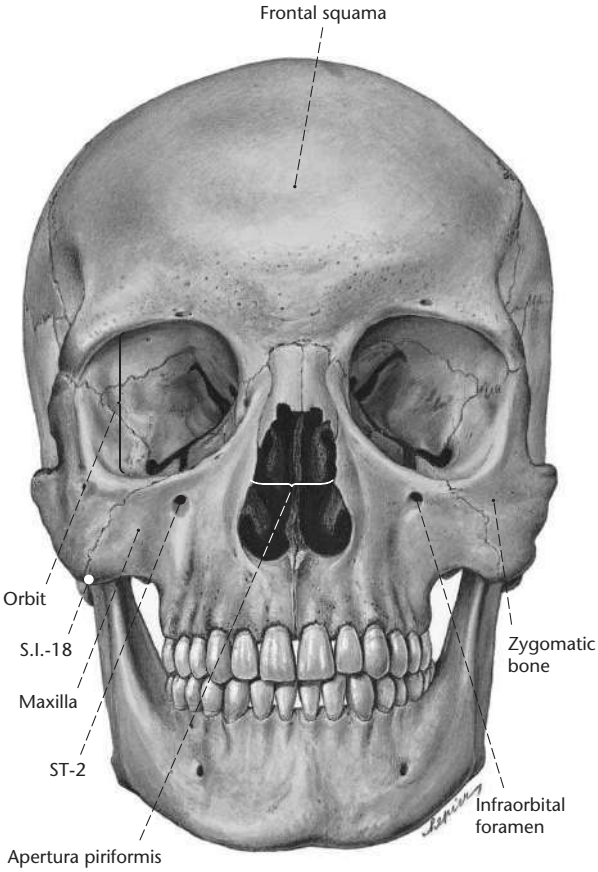

Fig. 3.13

3.1.2 Centre of the Face and Nasal Region

Cheek bones (→ Figs 3.12, 3.13)

Both the maxilla and the zygoma form the cheek bones below the eyes. Important anatomical landmarks are the infraorbital foramen, the lower border of the zygoma and the zygomatic arch of the temporal bone.

Infraorbital foramen (→ Figs 3.12, 3.13)

With the patient looking straight ahead, the infraorbital foramen with → **ST-2** is located on the maxilla, directly below the pupil. The foramen lies 1 finger-width below the infraorbital ridge. By palpating inferiorly from there, it can be felt below the maximal curvature of the ridge, in a small depression that is often sensitive to pressure.

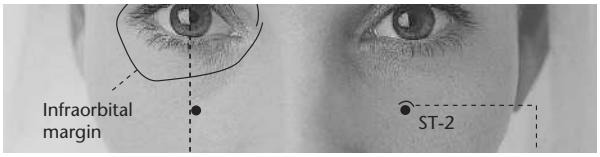

Fig. 3.14

Zygoma (→ Figs 3.15, 3.16)

The bony structures below the lateral orbit are formed by the zygoma. → **S.I.-18** is located directly below the outer canthus of the eye at the lower border of the zygoma, at the junction of the zygoma and the maxilla, on the anterior border of the masseter muscle. (→ Lower face and chin).

Zygomatic arch (→ Figs 3.15, 3.16)

On the lateral aspect of the face, the temporal process of the zygoma articulates with the zygomatic process of the temporal bone to form the zygomatic arch. This is an approximately horizontal bony structure which can be palpated laterally to the ear. → **G.B.-3** and → **T.B.-22** are both located on the upper border of the zygomatic arch, while → **ST-7** and → **T.B.-21** are situated at its lower border.

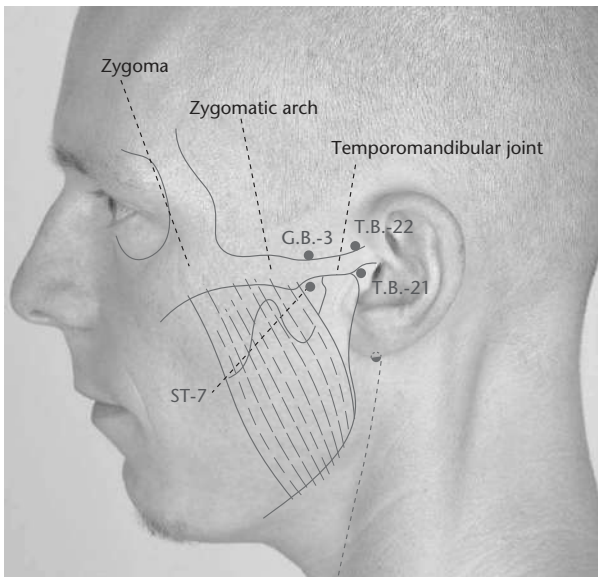

Fig. 3.15

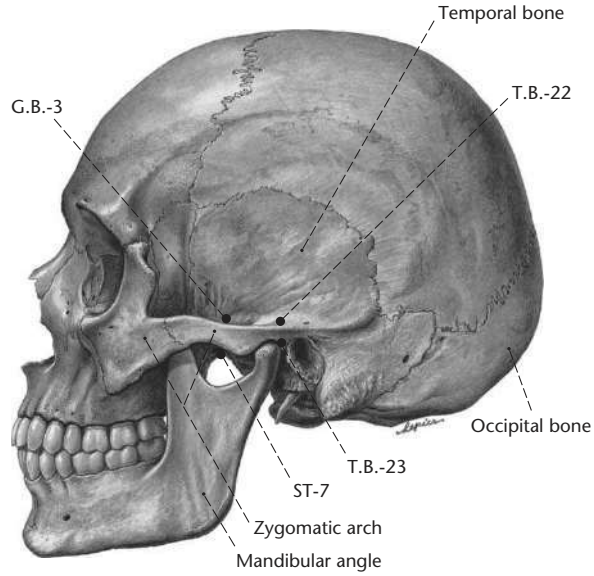

Fig. 3.16

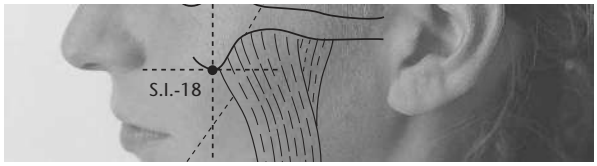

Fig. 3.17

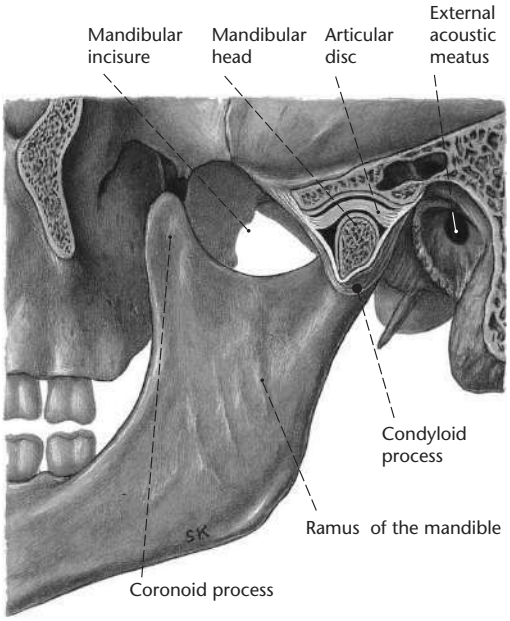

Fig. 3.18

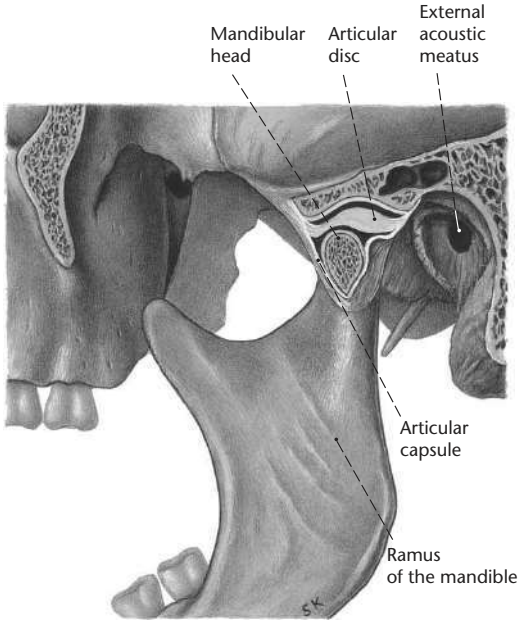

Fig. 3.19

Fig. 3.20

Temporomandibular joint (→ Figs 3.15, 3.18, 3.19)

The temporomandibular joint is located inferior to the zygomatic arch and anterior to the tragus of the ear. Its movement can easily be palpated by gently opening and closing the mouth. → **ST-7** is located in a depression anterior to the joint and directly inferior to the zygomatic arch. This depression is situated in the mandibular incisure, which is formed by the coronoid and condylar processes of the mandible. By opening the mouth, the palpating finger resting on → **ST-7** is 'lifted' out of the depression by the condyloid process sliding anteriorly.

Nasolabial groove (→ Fig. 3.21)

The nasolabial groove begins lateral to the nose and superior to the ala nasi, at the junction of the cartilage and the bone. From there it runs to the corners of the mouth. → **Ex-HN-8** (*shang-yingxiang, bitong*) is located at its upper end, at the junction of the upper jaw and the nasal cavities. → **L.I.-20** is located slightly below *bitong*, between the lateral border of the side of the nose and the nasolabial groove.

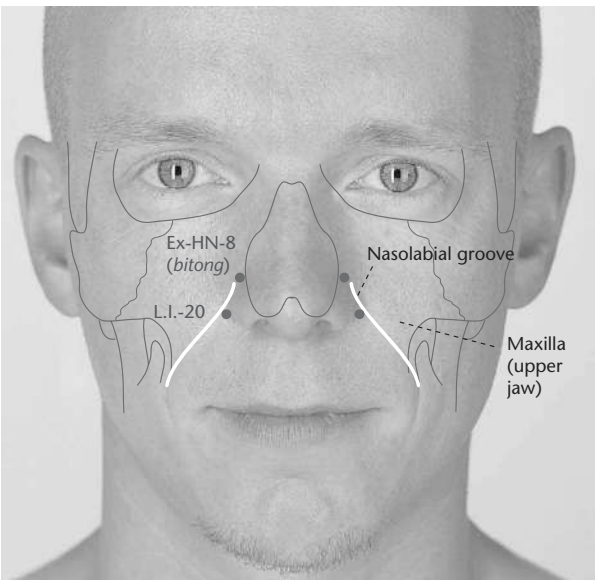

Fig. 3.21

3.1.3 Lower Face and Chin

Mandibular angle (→ Fig. 3.22)

The mandibular angle corresponds to the area where the lower jaw line changes its direction from horizontal to vertical. It is located markedly inferior and anterior to the ear lobe. This angle marks the location of → **ST-6** (1 finger-width anterior and superior to the mandibular angle) as well as of → **S.I.-17** and → **T.B.-16** (level with the mandibular angle).

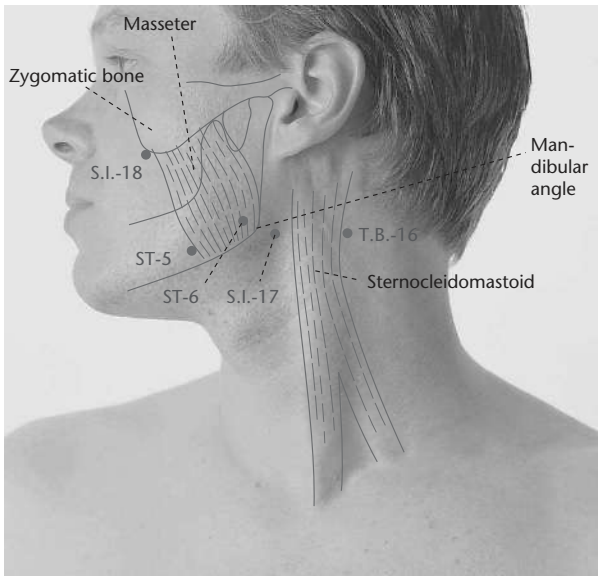

Fig. 3.22

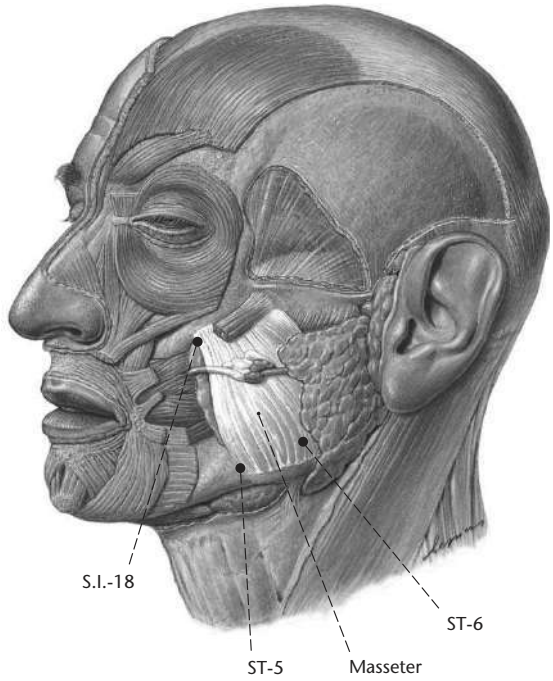

Fig. 3.23

Masseter muscle (→ Fig. 3.23)

By firmly clenching the jaw, the contours of the masseter muscle become well defined, stretching from the lateral aspect of the lower jaw to the zygomatic bone. It originates at the zygomatic arch (→ Fig. 3.15) and inserts at the mandibular angle (→ Fig. 3.22).

Mentolabial groove (→ Fig. 3.24)

This forms a transverse groove at the junction of the chin and the lower lip. → **Ren-24** is located in its centre.

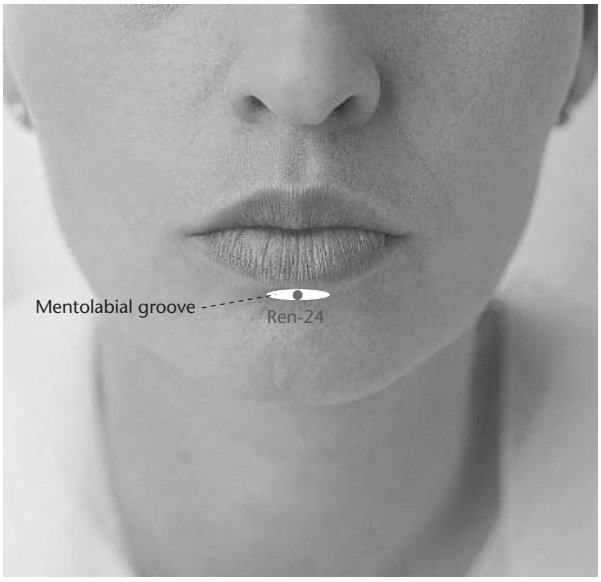

Fig. 3.24

3.1.4 Ear Region

Note: points used in auricular acupuncture are not discussed here.

Helix root (→ Figs 3.25, 3.26)

The helix root borders the upper part of the ear. This is the location of → **T.B.-22**.

Supratragic notch (→ Figs 3.25, 3.26)

The supratragic notch forms an incisure within the cartilage of the ear, separating the helix root from the tragus. → **T.B.-21** is located anterior to it.

Tragus (→ Figs 3.25, 3.26)

→ **S.I.-19** is located anterior to the midpoint of the tragus.

Intertragic notch (→ Figs 3.25, 3.26)

The intertragic notch forms an incisure in the auricular cartilage, separating the tragus from the lobe and antitragus. → **G.B.-2** is located anterior to this notch.

Ear lobe (→ Figs 3.25, 3.26)

Inferior to the earlobe and the ear canal is a depression, which is bordered posteriorly by the mastoid process and anteriorly by the lower jaw. The lower border is formed by the transverse process of the 1st cervical vertebra (atlas). → **T.B.-17** is located in this depression.

Border between the auricle and the face (→ Fig. 3.26)

The border between the auricle and the temple/cheek generally forms a more or less vertical line, which will become more clearly defined by bending the auricle and tragus towards the anterior. Several points are located along this line (from top to bottom): → **T.B.-22, T.B.-21, S.I.-19, G.B.-2.**

Mastoid process (→ Figs 3.26, 3.27)

The mastoid process is a cone-shaped bony structure which can be palpated posterior to the ear. → **G.B.-12** is located at its tip, while → ***anmian*** (Ex-HN) can be found at the border of the mastoid process and the occiput. → ***yiming*** (Ex-HN-14) is located somewhat more inferiorly.

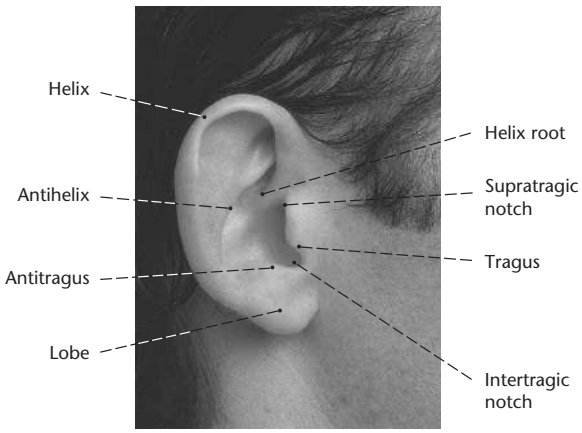

Fig. 3.25

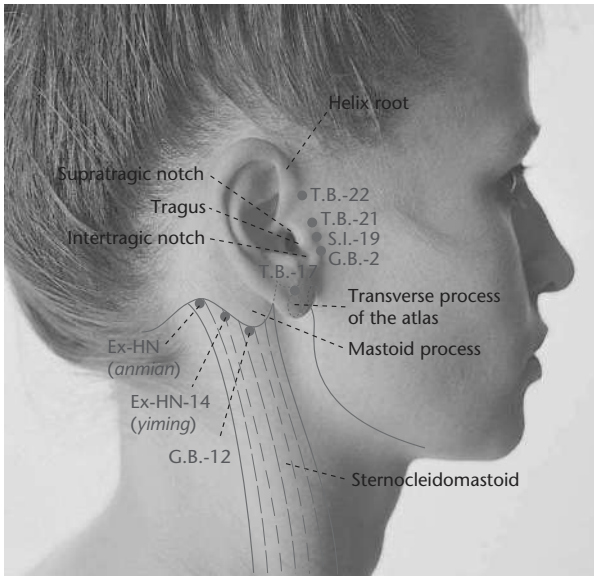

Fig. 3.26

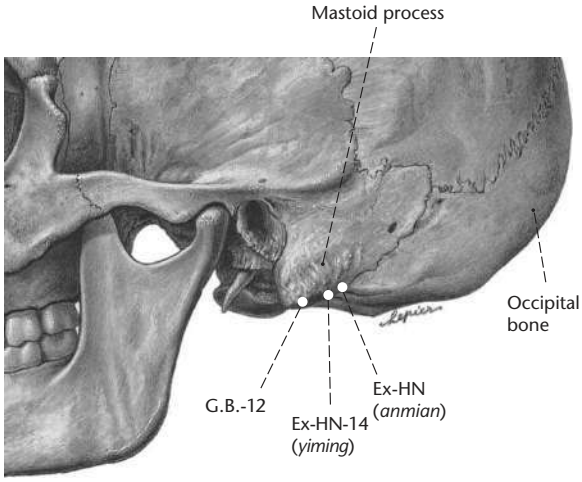

Fig. 3.27

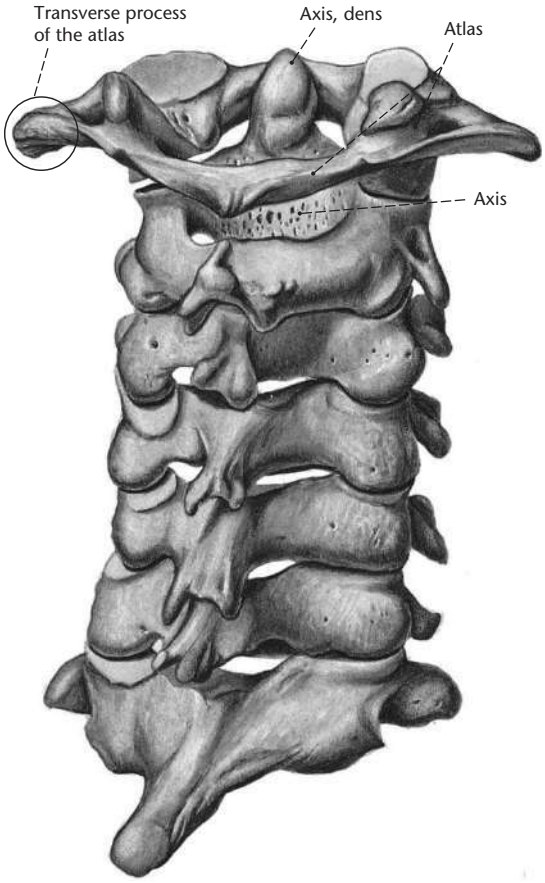

Fig. 3.28

Transverse process of the 1st cervical vertebra (atlas) (→ Fig. 3.28)

The transverse process of the atlas can be palpated inferior to the earlobe as a deep bony structure, which in most cases is very sensitive to pressure.

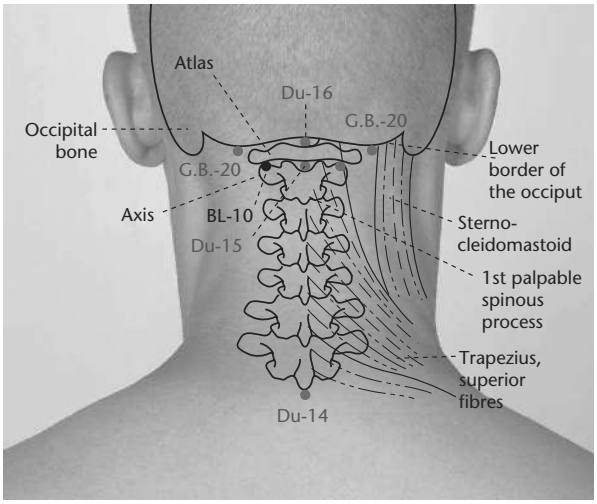

Fig. 3.29

3.1.5 Occiput

Craniocervical transition (→ Figs 3.28–3.32)

The transition from the head to the neck is formed by the mastoid process, the adjacent dorsal musculature and the occiput. The following points are located in this area (from lateral to medial) → **G.B.-12,** ***yiming*** **(Ex-HN-14),** ***anmian*** **(Ex-HN), G.B.-20, BL-10, Du-15** and **Du-16.**

External occipital protuberance (→ Figs 3.32, 3.33)

The external occipital protuberance forms a flat projection on the posterior midline of the occiput, which can be palpated slightly superior to the craniocervical transition. It defines the location of → **Du-16, Du-17** and **BL-9.** Sometimes, and more frequently in women, the protuberance may be defined only poorly or not at all.

Posterior hairline (no figure)

The posterior hairline is used as a reference point for locating points on the occiput, but due to its variable position it is not a very reliable landmark.

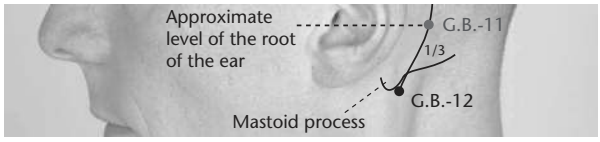

Fig. 3.30

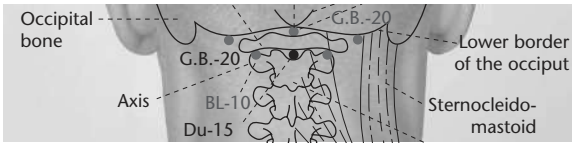

Fig. 3.31

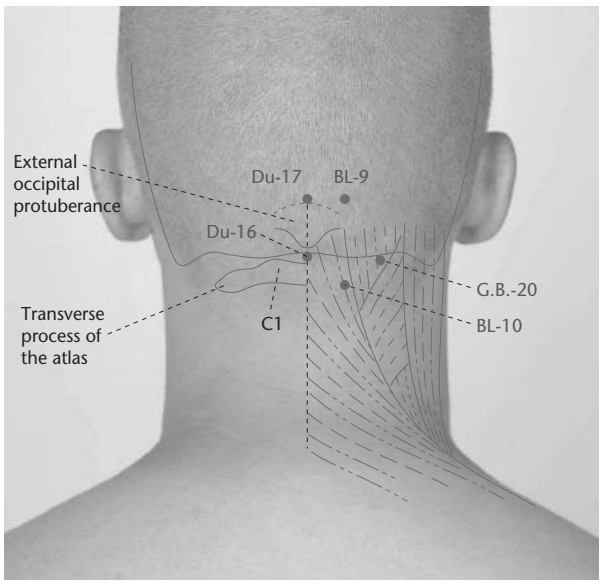

Fig. 3.32

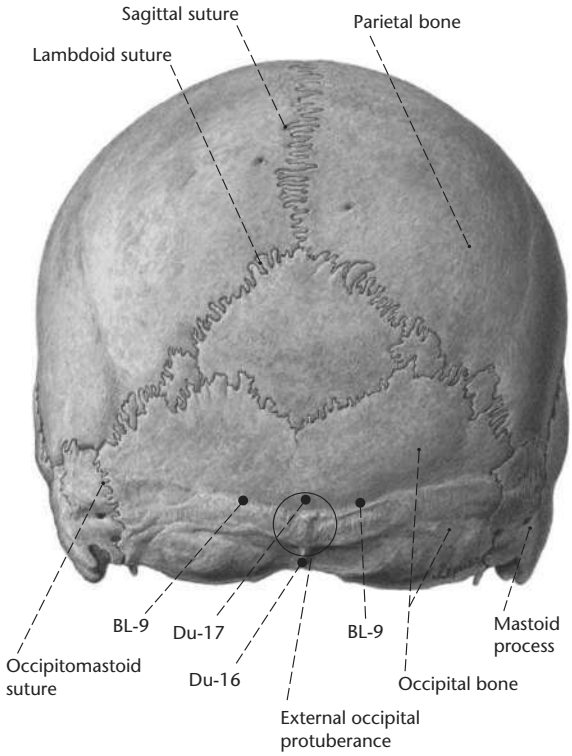

Fig. 3.33

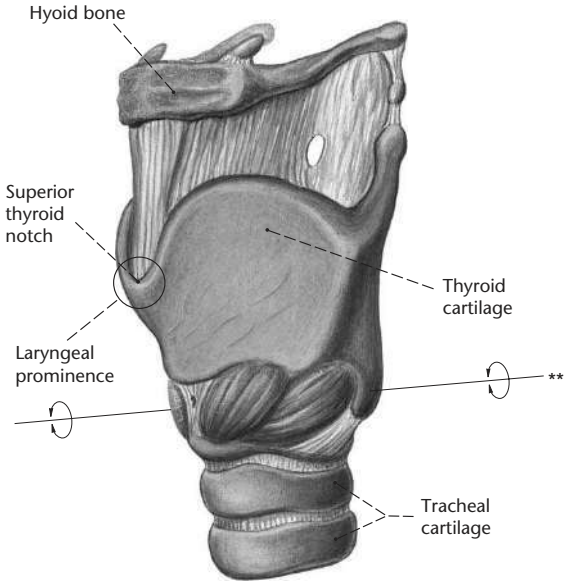

Fig. 3.34

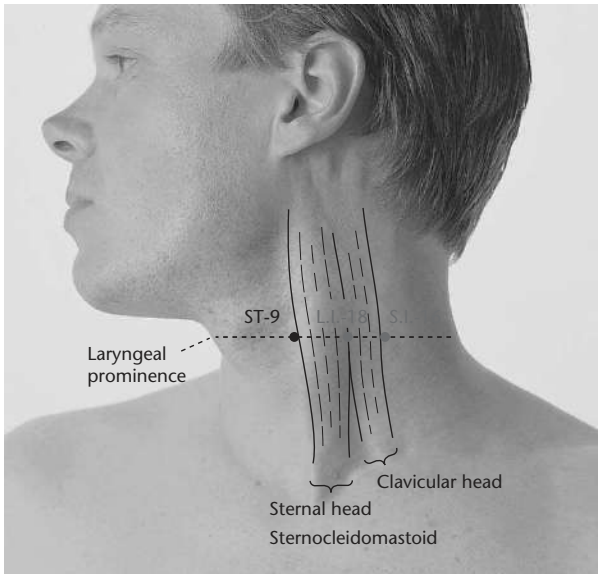

Fig. 3.35

3.2 Neck

Laryngeal prominence (→ Figs 3.34, 3.35)

In men, the laryngeal prominence as the highest point of the larynx ('Adam's apple') tends to form a clearly palpable and visible structure. In women, visual identification is more difficult, so that palpation is necessary in order to locate the V-shaped incisure on the upper border of the thyroid cartilage on the laryngeal midline. Located at the level of the laryngeal prominence are → **ST-9, L.I.-18** and **S.I.-16**.

Sternocleidomastoid (SCM) muscle (→ Figs 3.36, 3.37)

This muscle forms a distinct, easily identifiable structure on the anterior and lateral aspect of the neck, which becomes visible and palpable by rotating the neck in the opposite direction. This effect can be increased by rotating the head against resistance (by pushing against the chin for example). The more slender clavicular head originates from the anterior aspect of the manubrium, while the more flat clavicular head originates from the medial third of the clavicle, revealing a small triangle between the two heads near the sternoclavicular joint. This is the location of → **ST-11**. By palpating along the SCM muscle in a superior direction, the two heads can be distinguished for the major part of the muscle, before they merge to insert at the posterior part of the mastoid process and the lateral superior nuchal line of the occiput. Besides → **ST-11,** the following points are located alongside the muscle belly of the SCM: → **L.I.-17, L.I.-18, S.I.-16, ST-9, G.B-12** and ***anmian*** **(Ex-HN)**.

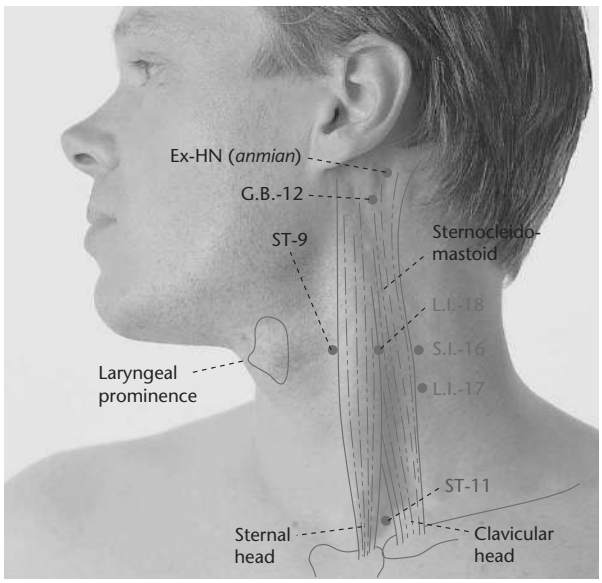

Fig. 3.36

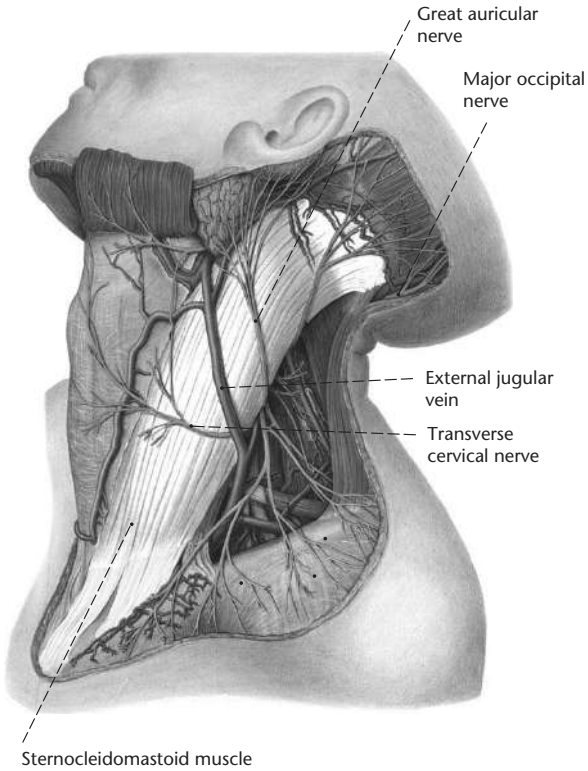

Fig. 3.37

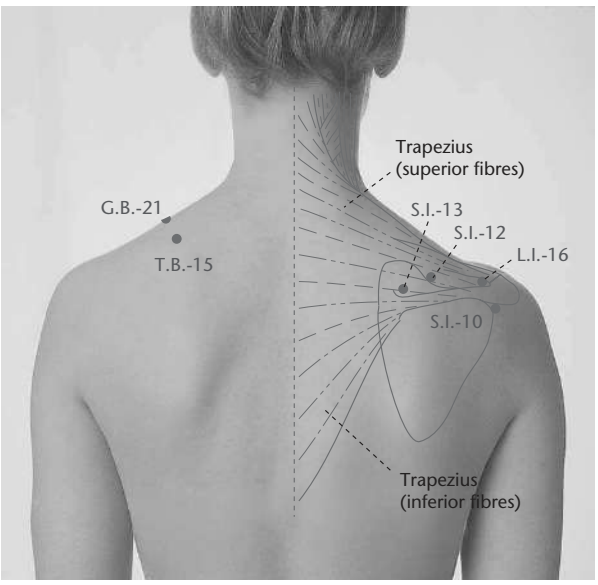

Fig. 3.38

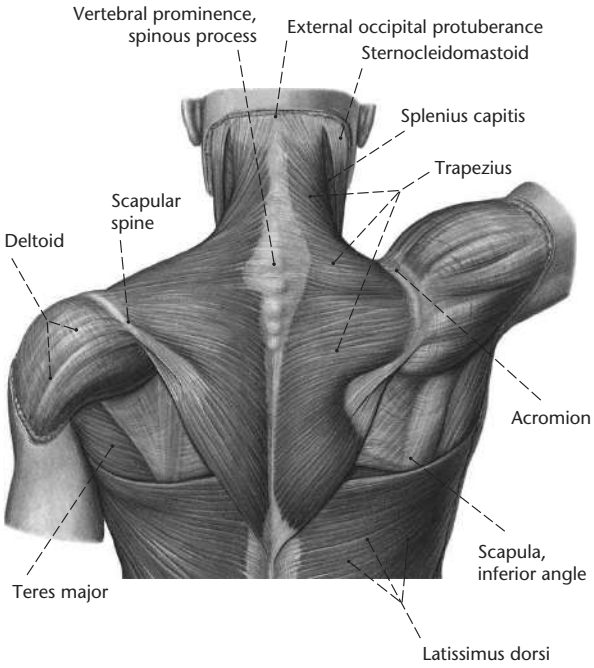

Fig. 3.39

3.3. Shoulder and Arm

3.3.1 Shoulder and Upper Arm

Trapezius muscle (→ Figs 3.38, 3.39)

This muscle covers the superior aspect of the shoulder, its superior portions extending from the cervical spine to the acromion at the lateral end of the scapula. Located at the midpoint of the superior border of the muscle are → **G.B.-21** and, slightly inferiorly, → **T.B.-15.**

Acromion (→ Figs 3.39, 3.42, 3.43)

When palpating the shoulder joint from a lateral direction, the acromion can be felt as a bony extension with a flat surface, slightly superior to the laterally protruding head of the humerus.

Scapular spine (→ Figs 3.39–3.44)

The scapular spine forms a bony ridge that originates at the acromion and runs in an oblique angle across the scapula, ending at its medial border in an upward opening curve. This is the location of → **S.I.-13**, → **L.I.-16** is located on the most lateral portion of the trapezius and supraspinatus muscles, in the angle between the acromial origin of the scapular spine and the acromioclavicular joint.

Head of the humerus, shoulder hollows (→ Figs 3.42, 3.44)

The head of the humerus is located inferior to the acromion and protrudes slightly on the lateral aspect of the shoulder. When abducting the arm horizontally, two hollows will form at the transition from the arm to the shoulder. The anterior hollow is the location of → **L.I.-15**, while → **T.B.-14** is located in the posterior one. The hollows mark the border between the clavicular and acromial portion and the border between the acromial and spinal portion of the deltoid muscle.

Fig. 3.40

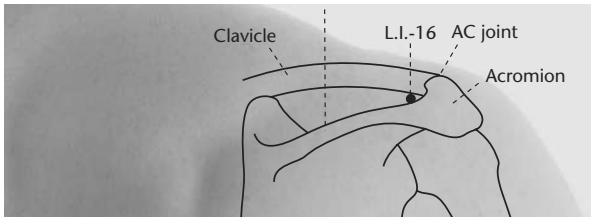

Fig. 3.41

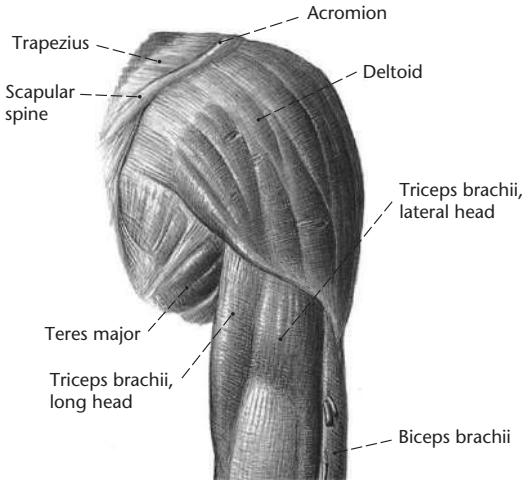

Fig. 3.42

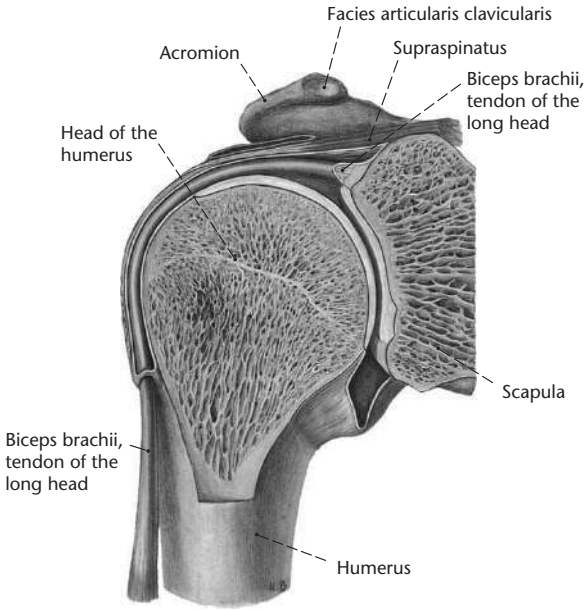

Fig. 3.43

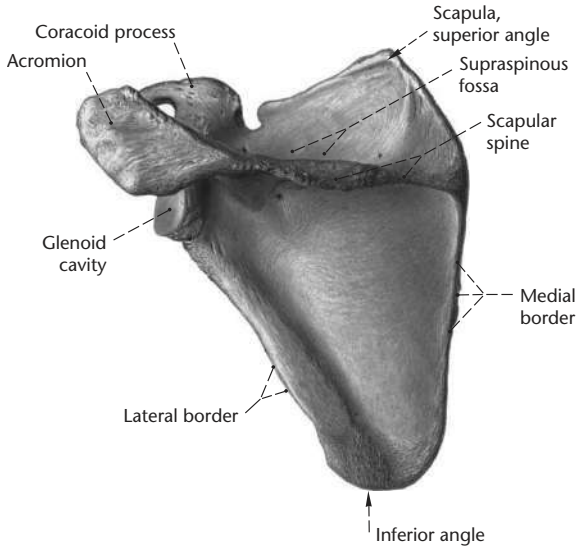

Fig. 3.44

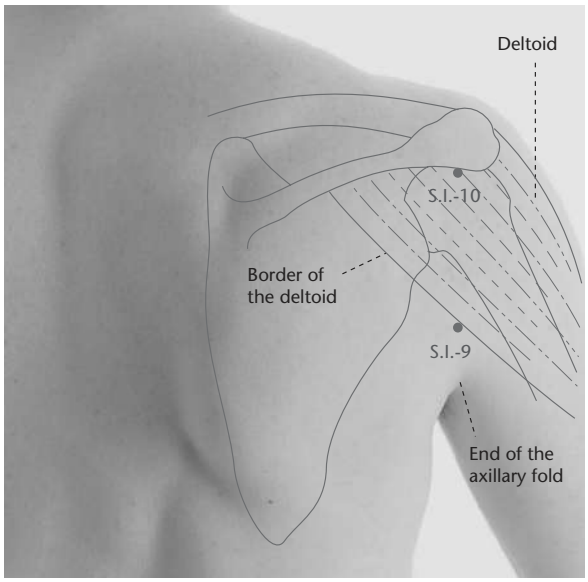

Fig. 3.45

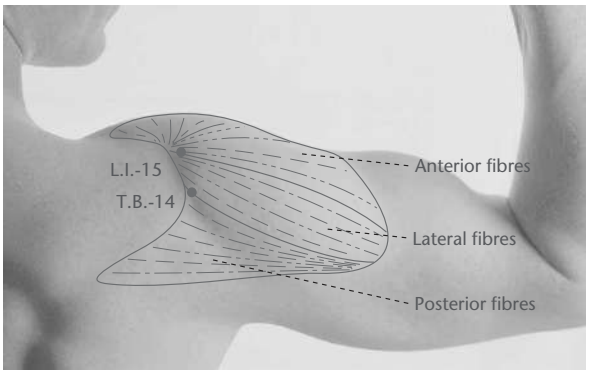

Fig. 3.46

Deltoid muscle (→ Figs 3.42, 3.45–3.51)

This muscle covers the anterior, posterior and lateral aspect of the head of the humerus. The following points are located either on the muscle or on its margins: **→ S.I.-9, S.I.-10, T.B.-14, L.I.-15, *jianquan* (Ex-UE), LU-1, LU-2, LU-3, L.I.-14**.

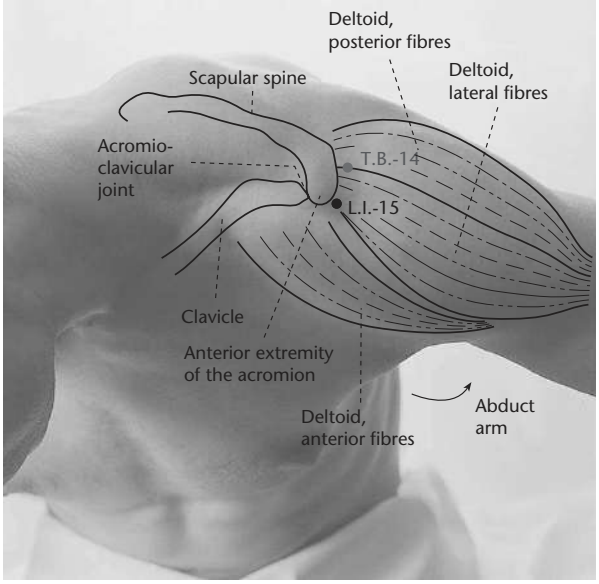

Fig. 3.47

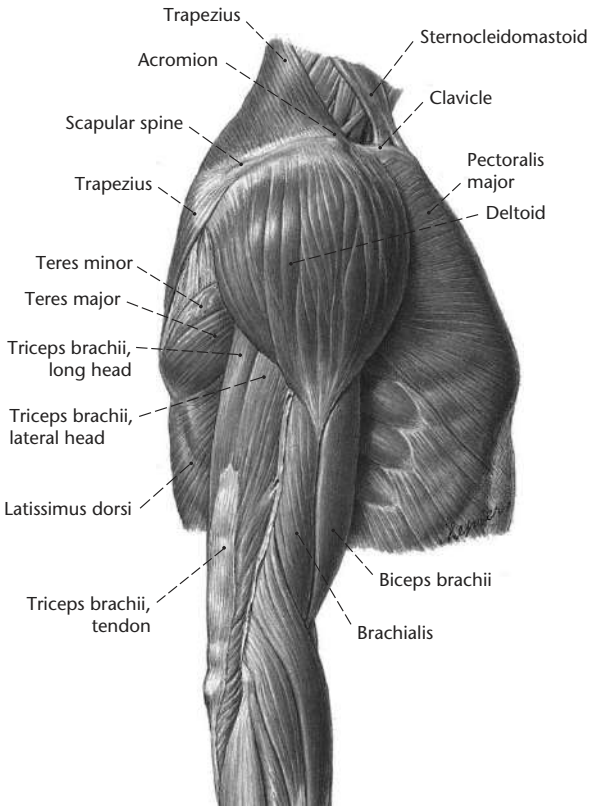

Fig. 3.48

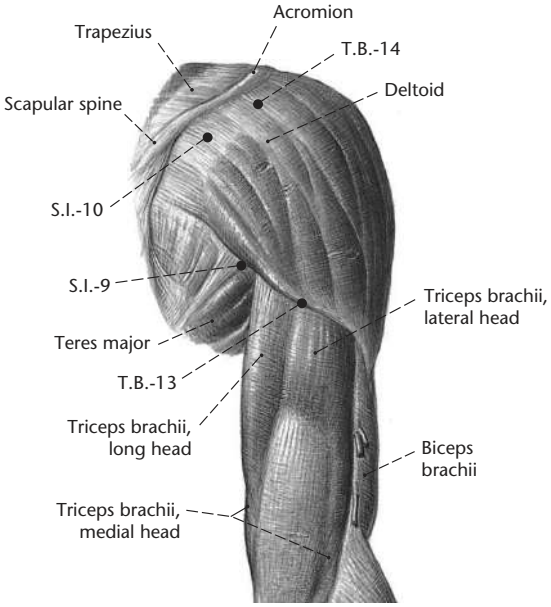

Fig. 3.49

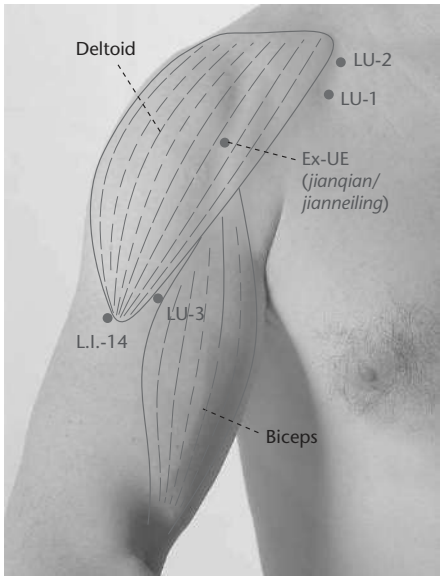

Fig. 3.50

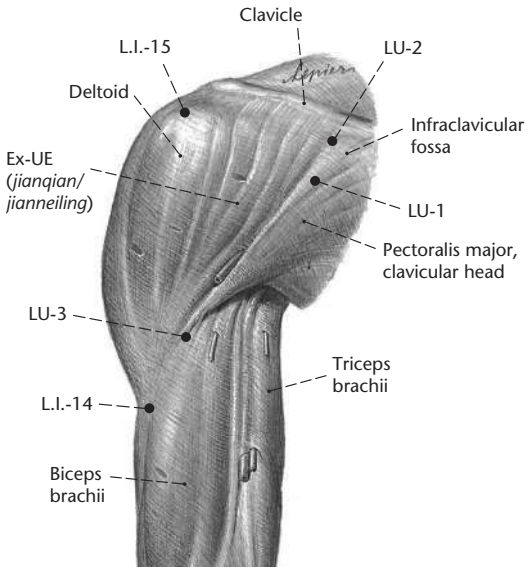

Fig. 3.51

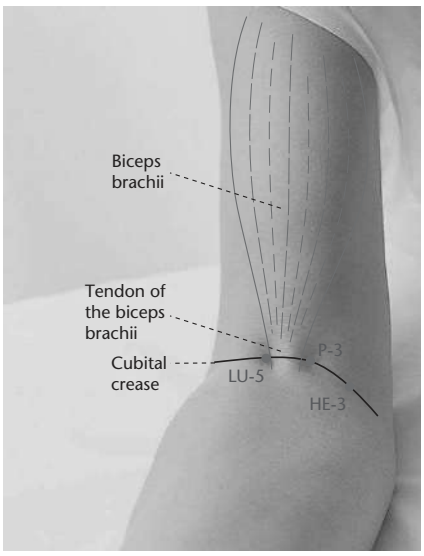

Fig. 3.52

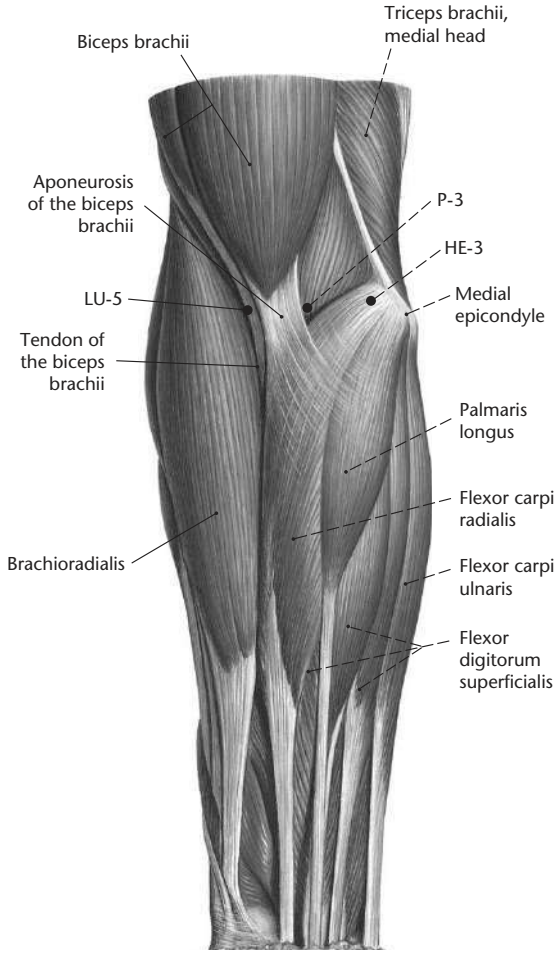

Fig. 3.53

3.3.2 Elbow

Elbow, tendon of the biceps muscle (→ Figs 3.52, 3.53)

The anterior aspect of the elbow is marked by the cubital crease, extending from the lateral to the medial epicondyle of the humerus. It is divided by the tendon of the biceps muscle. → **L.I.-11** and → **LU-5** are located lateral to the tendon while → **P-3** and → **HE-3** are located medial to it.

Lateral and medial epicondyles of the humerus (→ Figs 3.53, 3.55, 3.56)

The lateral and medial epicondyles of the humerus form the lateral aspects of the elbow. The olecranon is located between them, on the posterior aspect of the elbow. The medial epicondyle is the origin of the extensor muscles of the wrist joint, while the flexor muscles originate at the medial epicondyle.

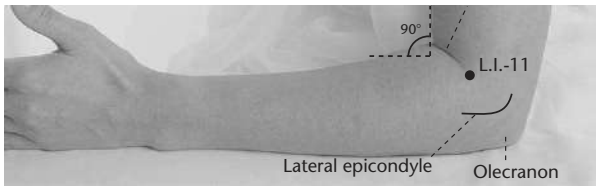

Fig. 3.54

Olecranon (→ Figs 3.55, 3.56)

The olecranon forms the tip of the elbow. Located in the ulnar sulcus, between the medial epicondyle and the tip of the olecranon, is → **S.I.-8**, while → **T.B.-10** is located in a depression proximal to the olecranon.

3.3.3 Forearm and Hand

For locating points on the forearm, it is important to note whether the arm is in a pronated or supinated position, since this can have a considerable effect on the relative location of some points (see also location details of specific points). For example, in supination the line connecting → **L.I.-5** and → **L.I.-11** runs along the lateral border of the forearm, while in pronation it traverses the forearm.

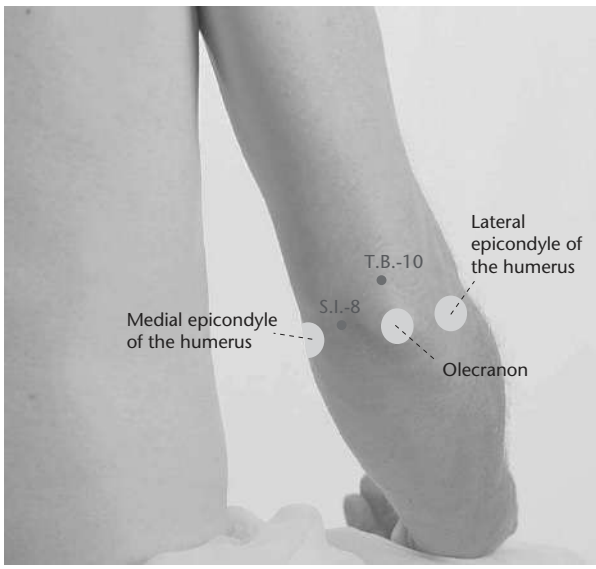

Fig. 3.55

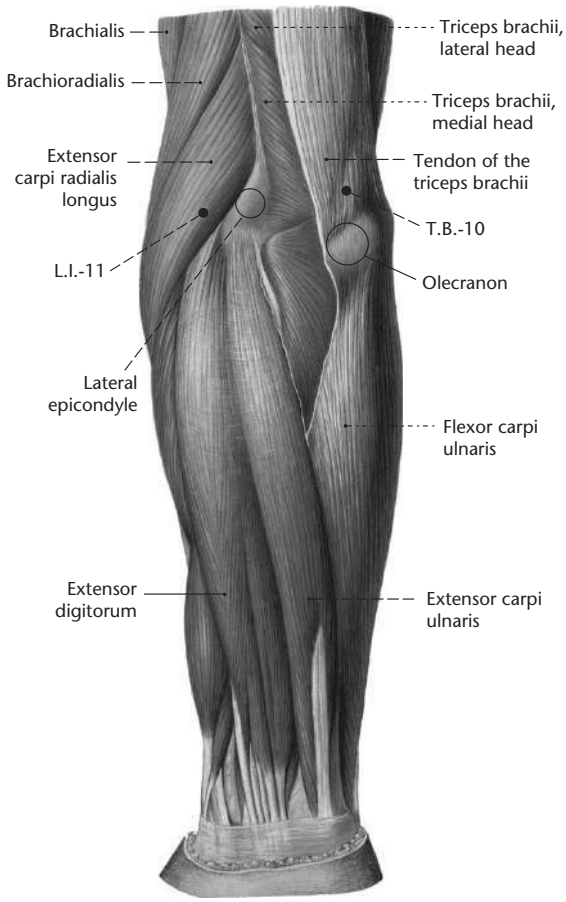

Fig. 3.56

[redacted]

[redacted]

[redacted]

Lumbosacral junction

Following orientation in the lumbar region with the help of the iliac crests (→ Iliac crest), the lumbosacral joint can be located by counting the spinous processes in an inferior direction. It can generally be felt as a marked 'kink' or depression, followed by the smaller spinous processes of the slightly convex median sacral crest directly below.

Caution: Note that (according to the literature in up to 15% of cases) there can be lumbarization of the 1st sacral vertebra or sacralization of L5. In these cases there will be either four or six lumbar vertebrae.

Posterior superior iliac spine (PSIS) (→ Figs 3.73, 3.74)

Another important landmark on the lower back is the posterior superior iliac spine (PSIS), which forms the posterior end of the iliac crest lateral to the upper sacral region. The PSIS can often be identified by two superficially visible dimples. It is best located by palpating from an inferior in a superior direction. If the dimples are not visible, they can be located by palpating 3 cun from the rima ani in a superior and lateral direction at an angle of 45°, until a marked bony ridge can be felt.

Generally, the PSIS is located at the level of the space between the 1st and 2nd sacral foramina, so that → **BL-27** is located directly medial or slightly superior and medial to the PSIS, while → **BL-28** is located just inferior and medial to the PSIS.

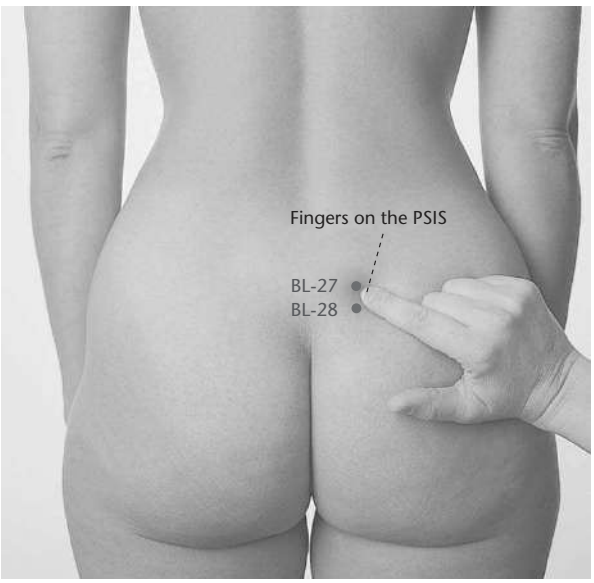

Fig. 3.73

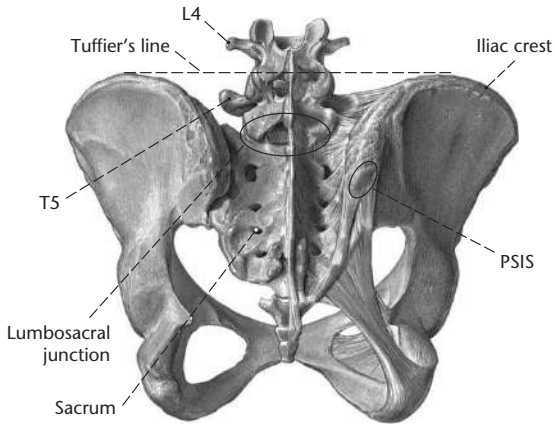

Fig. 3.74

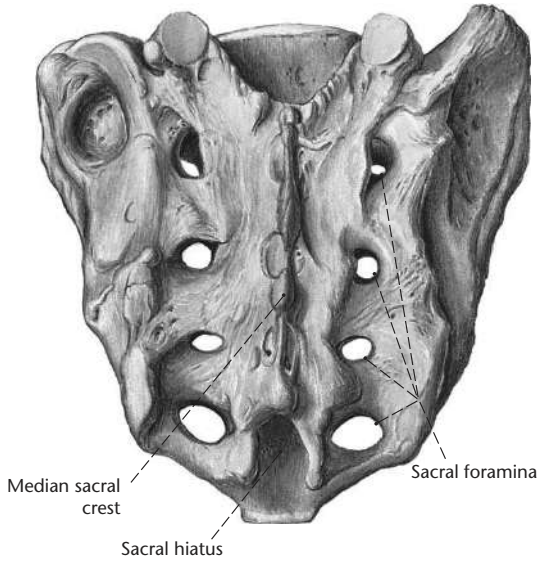

Fig. 3.75

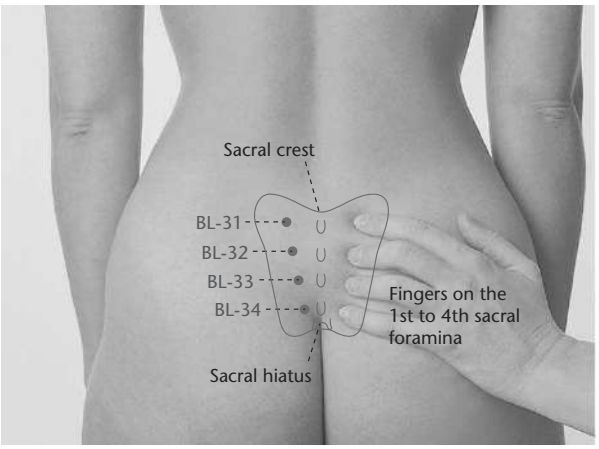

Fig. 3.76

3.4.4 Sacrum and Sacroiliac Joint

Sacrum (→ Figs 3.75, 3.76)

On the sacrum, landmarks for palpation are the median sacral crest and the sacral hiatus as well as the sacral foramina on its lateral aspects.

Sacral crest (→ Figs 3.74, 3.75)

The sacral crest is formed by the narrow, irregular spinous processes, which can be palpated on the sacral midline.

Sacral hiatus (→ Figs 3.75, 3.76)

The sacral hiatus can be palpated as a U-shaped depression at the inferior end of the sacral crest. → **Du-2** is located on the midline, just inferior to it. Both the sacral hiatus and the greater trochanter (→ Lower extremity) are reference points for locating → **G.B.-30.**

Sacral foramina (→ Figs 3.75, 3.76)

The sacral foramina are located between the lumbosacral joint and the sacral hiatus and can usually be palpated. They are in most cases evenly spaced and lie approximately 1 finger-width lateral to the midline, with the distance decreasing in a distal direction. The foramina are the location of → **BL-31–BL-34.**

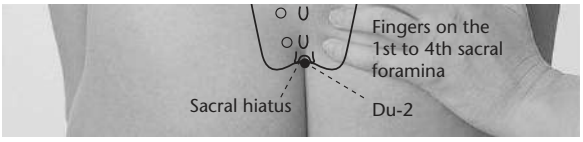

Fig. 3.77

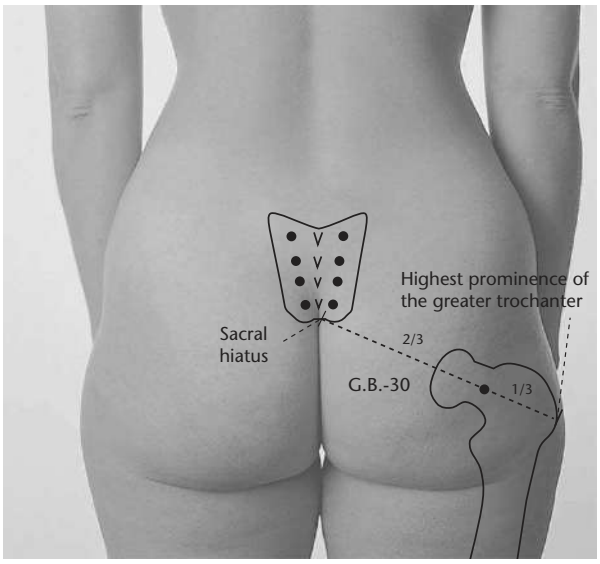

Fig. 3.78

Iliosacral joint (→ Figs 3.80, 3.81)

The iliosacral joint can be palpated bilaterally as a shallow groove between the sacrum, the PSIS (see above) and more inferiorly located ilial structures. → **BL-26–BL-29** are located in this area.

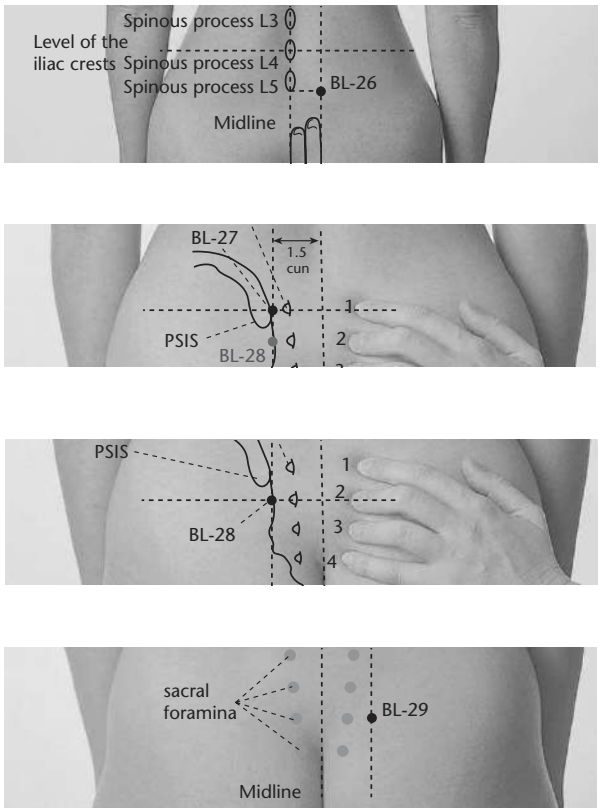

Fig. 3.79

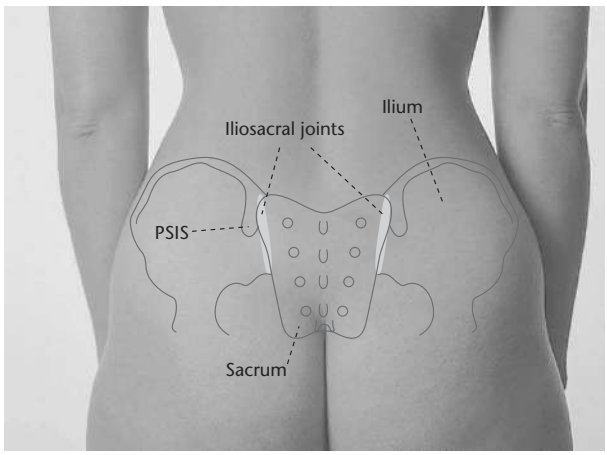

Fig. 3.80

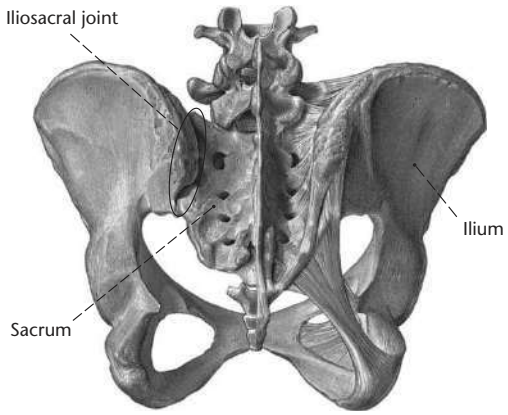

Fig. 3.81

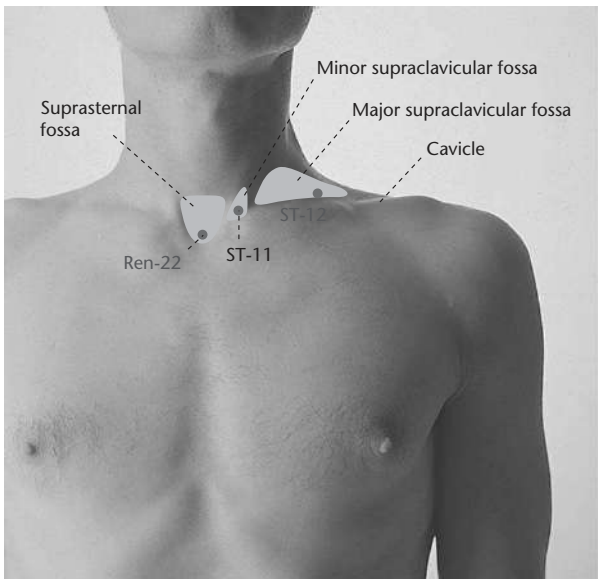

Fig. 3.82

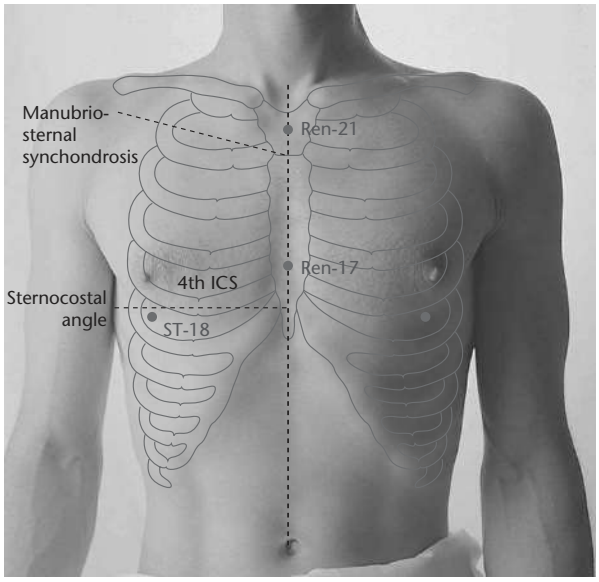

Fig. 3.83

3.5 Anterior and Lateral Aspect of the Thorax and Abdomen

Clavicle (→ Figs 3.82, 3.86)

The clavicle forms the connection between the shoulder girdle and the thorax. It separates the major and minor supraclavicular fossae with → **ST-11, ST-12** and ***jingbi*** (Ex-HN) from the ribcage. Located on the lower border of the clavicle are → **ST-13** (at its midpoint) and → **KID-27** (2 cun lateral to the midline).

Suprasternal fossa (→ Fig. 3.82)

The suprasternal fossa forms a depression superior to the sternum between the sternoclavicular joints. It is the location of → **Ren-22**.

Ribs (→ Figs 3.83, 3.86)

For anatomical orientation of the ribs, the midsection of the clavicle is a good starting point. The first rib lies directly below the clavicle, forming a narrow bow. It articulates with the manubrium just inferior to the sternoclavicular joint.

Please note the following when palpating the ribs and intercostal spaces (ICS) and locating the respective acupuncture points:

- From their articulation with the sternum, the ribs run initially horizontally or in a gentle curve in an inferior direction, while at the lateral aspect of the thorax, they assume a superior direction towards the posterior aspect of the body.
- Inferior to the 4th/5th intercostal spaces, the medial sections of the ribs merge with the costal cartilage, so that the intercostal spaces do not extend very far medially. Therefore it is necessary to palpate somewhat more laterally.

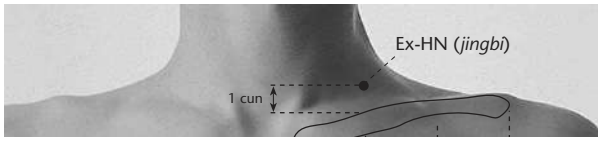

Fig. 3.84A

Fig. 3.84B

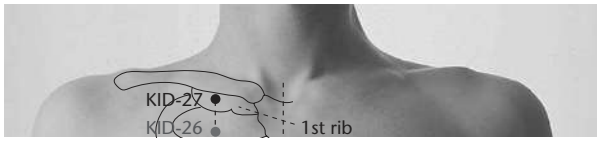

Fig. 3.84C

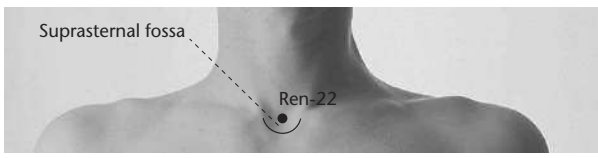

Fig. 3.84D

1st rib, 1st intercostal space (→ Fig. 3.86)

By palpating parasternally in an inferior direction, the first rib can generally be felt directly below the clavicle. Here, → **KID-27** is located 2 cun lateral to the anterior midline. Below follows the first intercostal space, then the second rib and so on. → **Ren-20** is located on the anterior midline, at the level of the first intercostal space, → **Ren-21** is located slightly superior to it.

2nd rib, manubriosternal synchondrosis (manubriosternal junction) (→ Figs 3.83, 3.86)

The synchondrosis is usually a clearly defined transverse bony structure on the superior section of the sternum, level with the articulation of the second rib. Directly below the synchondrosis is the 2nd intercostal space from where subsequent ribs and intercostal spaces can be counted by palpation.

Laterosuperior aspect of the thorax and coracoid process (→ Fig. 3.87)

The laterosuperior aspect of the thorax is bordered by the clavicle and the deltoid muscle with the coracoid process deep to the latter. The deltopectoral triangle can be located by palpating from the end of the anterior axillary fold along the border of the deltoid muscle towards the shoulder and the lower border of the clavicle. Here, in the centre of the triangle, lies → **LU-2**. Laterally, the triangle is bordered by a clearly palpable bony structure, the coracoid process. In order to distinguish the process from the lesser tubercle of the humerus, the arm is rotated externally with the elbow flexed at the same time. While the process will remain fixed, the tubercle will follow the movement. → **LU-1** is located slightly superior and lateral to → **LU-2** and medial to the lower border of the coracoid process. The contours of the deltoid muscle and the deltopectoral triangle can be emphasised through muscle flexion, for example by pressing the hands against each other in front of the chest.

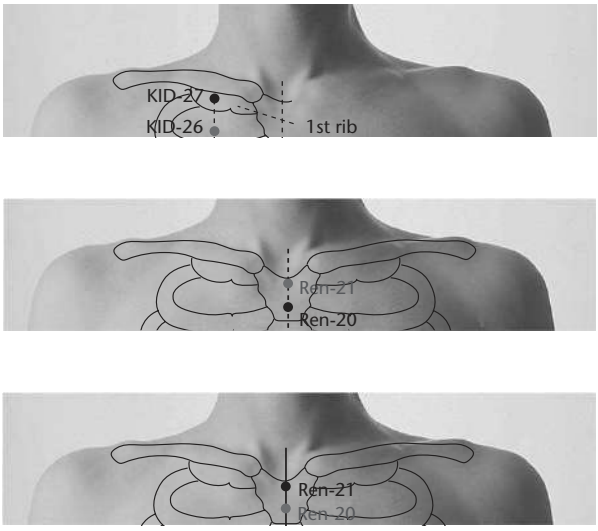

Fig. 3.85

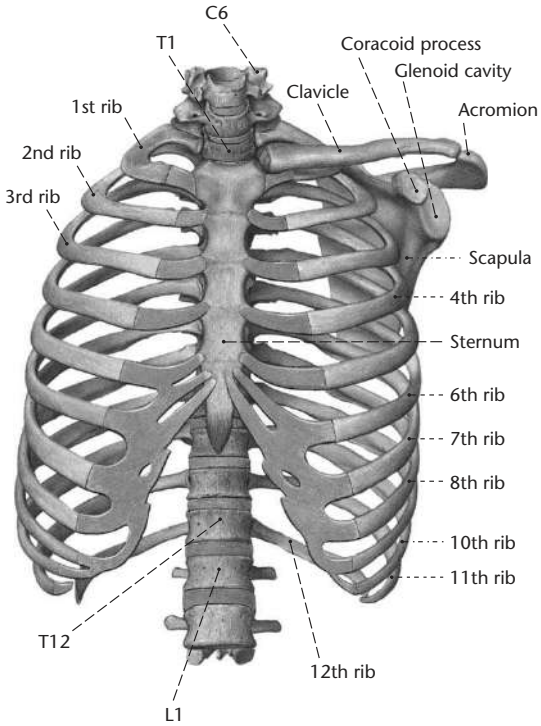

Fig. 3.86

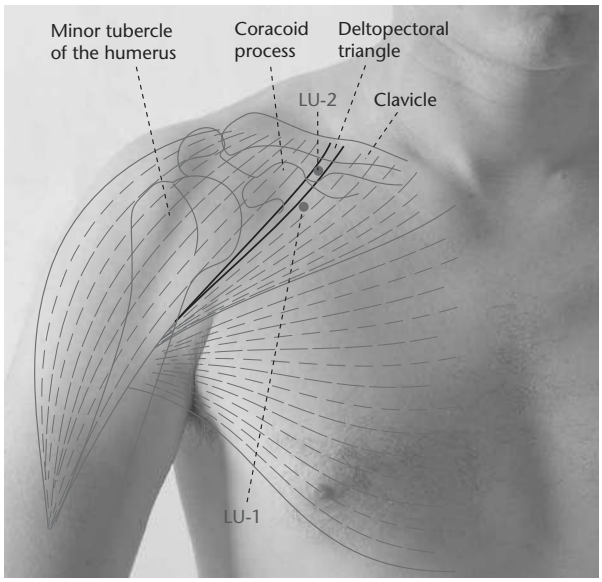

Fig. 3.87

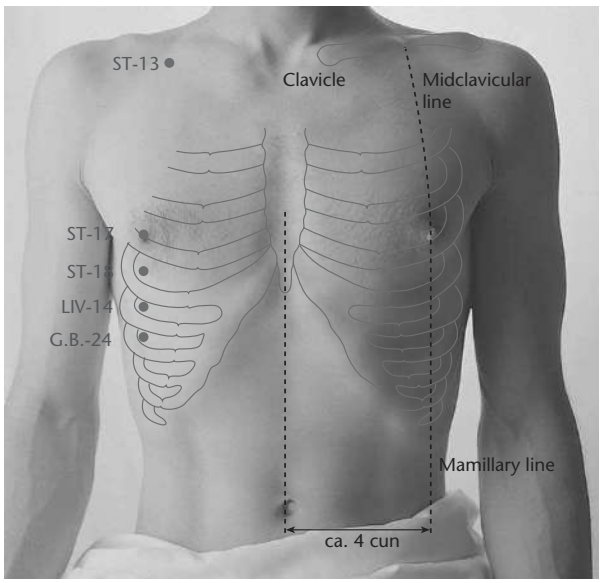

Fig. 3.88

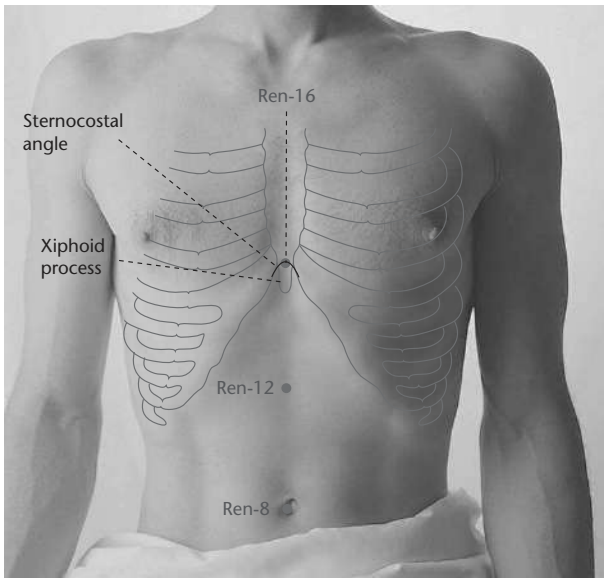

Fig. 3.89

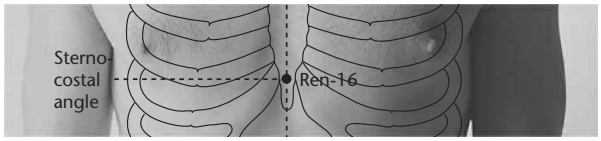

Fig. 3.90

Midclavicular line (→ Fig. 3.88)

The midclavicular line provides a further tool for reference on the upper thorax. It is an imaginary line that runs from the midpoint of the clavicle to the nipple. Since in most cases the nipple lies somewhat more laterally than the clavicular midpoint, the line tends to have a slightly oblique pathway. → **ST-13** to **ST-16** are all located on the line.

Nipple (→ Fig. 3.88)

In men, the nipple is located in the 4th intercostal space, lateral to the midline, while in women, the location can vary depending on the shape of the breast. → **ST-17** is located on the nipple.

Mamillary line (→ Fig. 3.88)

The mamillary line is an imaginary line for orientation on the mediolateral thorax and abdomen. It runs vertically from the nipple in an inferior direction. The points → **ST-18, LIV-14** and **G.B.-24** are all located on this line.

Sternocostal angle, xiphoid process and umbilicus (→ Figs 89, 3.91)

The lower borders of the ribcage unite at the lower sternal border, forming the sternocostal angle. This angle and the umbilicus are the two major anatomical landmarks for point location on the upper abdomen.

Caution: It is important not to mistake the sternocostal angle for the tip of the xiphoid process. The latter is attached to the sternum at the sternocostal angle, projecting inferiorly towards the abdomen. For definite identification, palpate along the costal cartilage in a superior direction to where it meets the sternocostal angle. This is the location of → **Ren-16**.

Axillary line (→ Fig. 3.92)

The axillary line is an imaginary line for orientation on the lateral aspect of the thorax and abdomen. It runs vertically from the middle of the axilla in an inferior direction. Both → **G.B.-22** and **SP-21** are located on this line.

Free end of the 11th rib (→ Figs 3.91, 3.92)

The free end of the 11th rib can be palpated below the costal cartilage, at the transitional zone from the thorax to the abdomen. Location tip: the patient places his/her flexed elbow (90° angle) alongside his/her thorax. The tip of the olecranon will then usually be close to the free end of the 11th rib. → **LIV-13** is located below its anterior lower border.

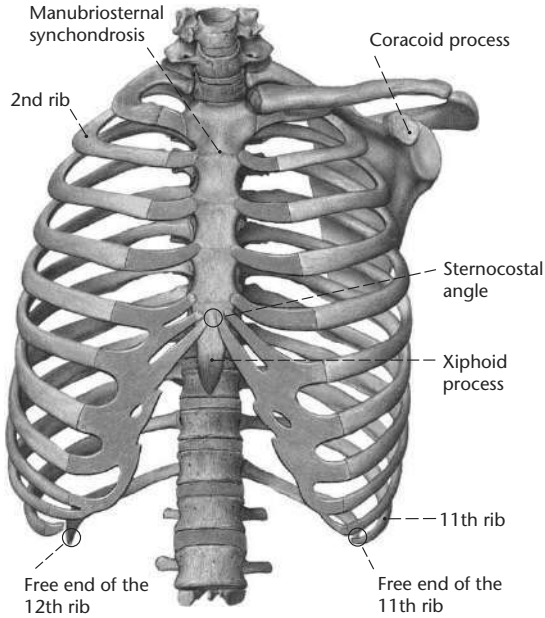

Fig. 3.91

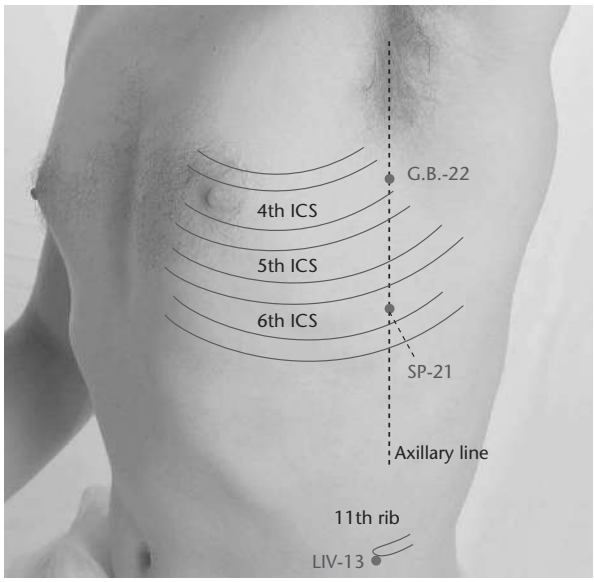

Fig. 3.92

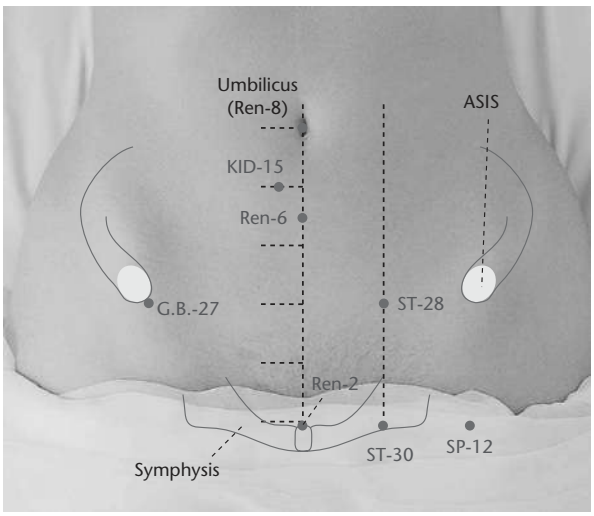

Fig. 3.93

Anterior superior iliac spine (ASIS) (→ Figs 3.93–3.95)

By palpating along the iliac crest in an anterior and inferior direction, the anterior superior iliac spine (ASIS) is reached at the lateral end of the inguinal groove. The ASIS is felt as a distinct bony ridge. → **G.B.-27** is located anterior to it, while → **G.B.-28** is located slightly more inferiorly and medially.

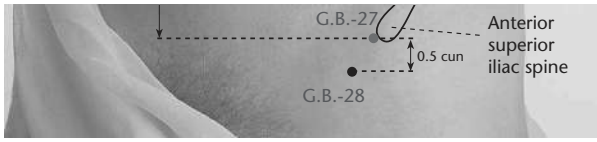

Fig. 3.96

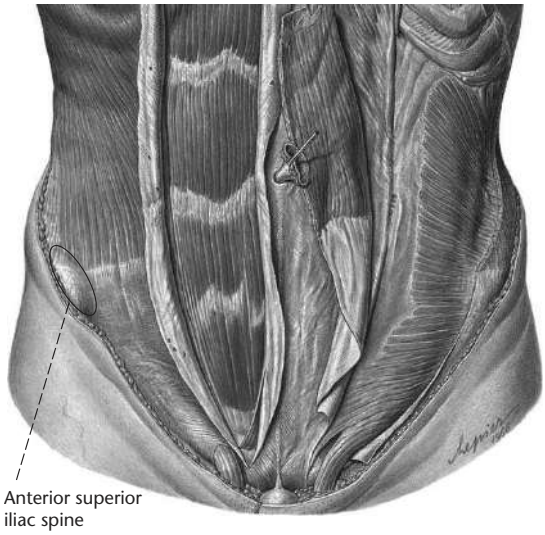

Fig. 3.94

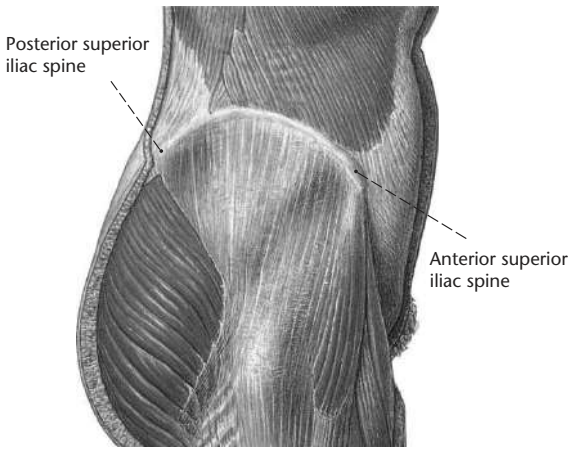

Fig. 3.95

Pubic symphysis, umbilicus (→ Fig. 3.98)

The upper border of the pubic symphysis is an important orientation point for point location on the lower abdomen. The distance between the centre of the umbilicus and the upper border of the symphysis can be divided into 5 proportional cun, which can differ significantly from the patient's finger cun (→ Chapter 2). → **Ren-2** is located on the anterior midline and on the upper border of the pubic symphysis.

3.6 Lower Extremity

Greater trochanter (→ Fig. 3.99)

The greater trochanter forms a distinct, projecting bony structure on the hip area. It is one of the endpoints on an imaginary line connecting it with the sacral hiatus. Located on this line is → **G.B.-30**, a third of the distance from the greater trochanter. Another line connecting the greater trochanter and the anterior superior iliac spine (ASIS) is used for locating → **G.B.-29**.

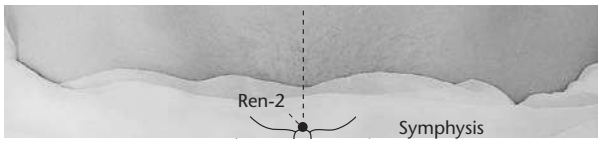

Fig. 3.97

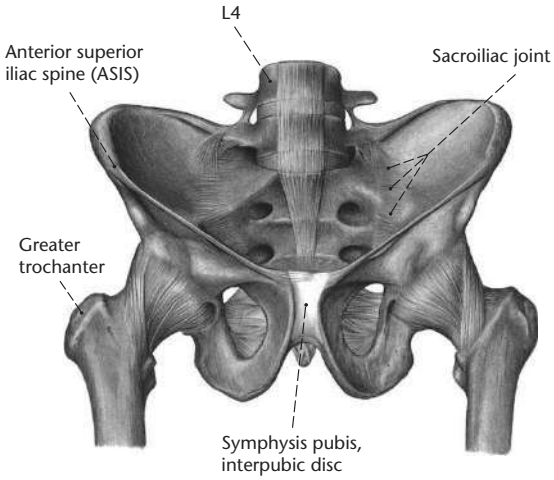

Fig. 3.98

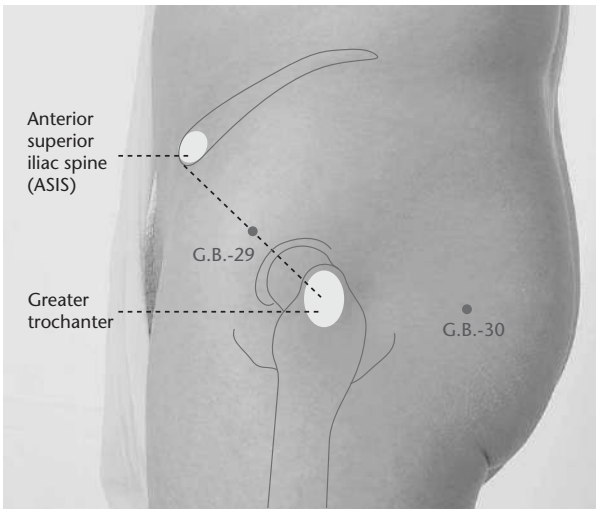

Fig. 3.99

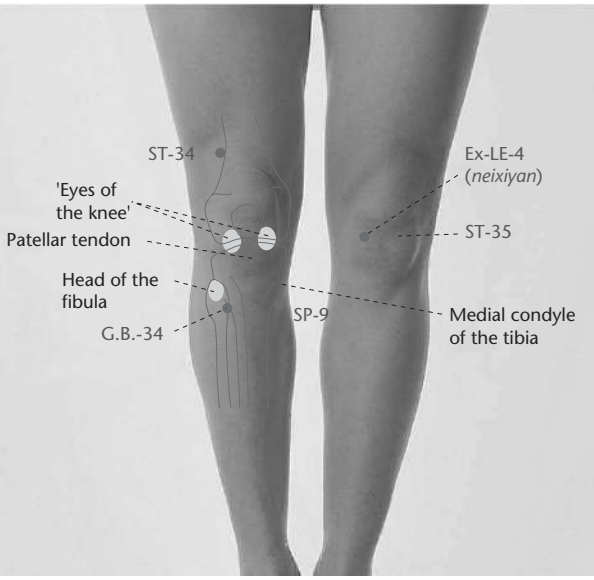

Fig. 3.100

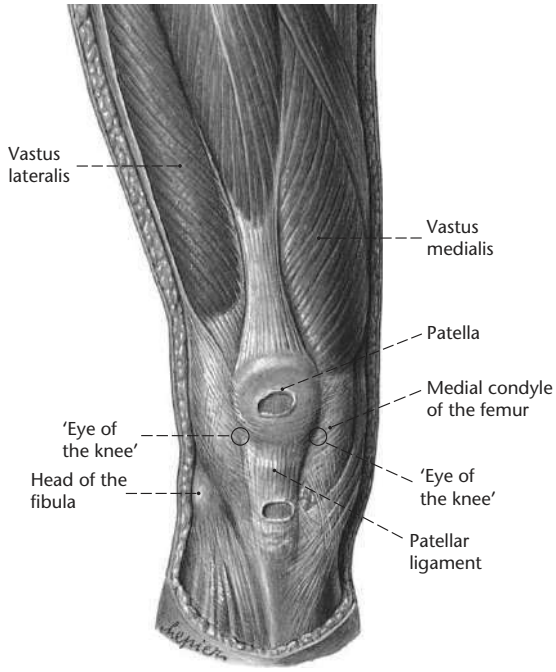

Fig. 3.101

3.6.1 Knee Region

For locating and needling points in the knee region, the patient should be in a relaxed supine position, with the knees slightly flexed and supported by pillows. This opens up the knee joint and allows for better palpation of the eyes of the knees. A prone position with a support below the feet and ankles is recommended for treating points in the popliteal crease.

Patella (→ Figs 3.100, 3.101)

The upper patellar border marks the location of → **SP-10** as well as of → **ST-32** to **ST-34**.

Eyes of the knee (→ Figs 3.100, 3.101)

The eyes of the knee are formed by the depressions on either side of the patellar tendon at the level of the lower patellar border. The lateral eye corresponds to the location of → **ST-35**, the medial eye is the location of the extra point → **Ex-LE-4** (*neixiyan*).

Head of the fibula (→ Figs 3.100, 3.101)

The head of the fibula forms a further important reference point inferior to the lateral aspect of the knee joint. For locating → **G.B.-34**, find the head of the fibula (approximately where the seam of a pair of trousers would be) and hold it with two fingers. By sliding downward, the anterior finger will 'drop' into the depression anterior and inferior to the head of the fibula, which marks the location of → **G.B.-34**.

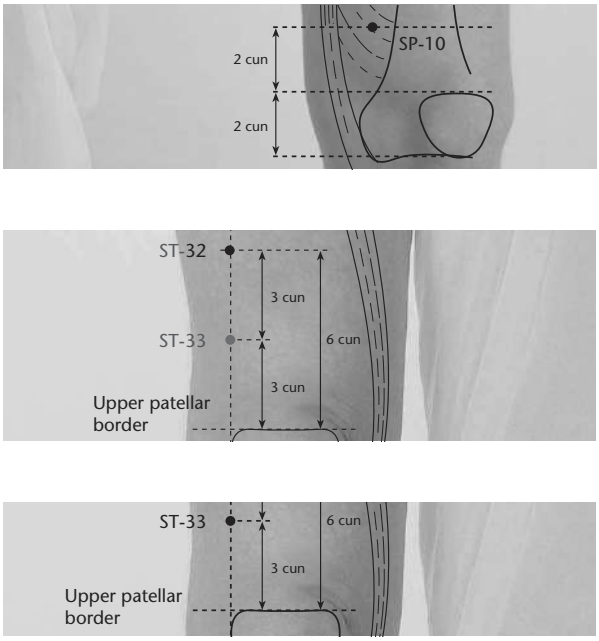

Fig. 3.102

Tibial shaft and medial condyle of the tibia (→ Fig. 3.103)

Below the medial aspect of the knee joint, the transition from the medial condyle of the tibia to the tibial shaft marks the location of → **SP-9**. This transition is best located by palpating in a superior direction along the posterior border of the medial aspect of the tibia. When using a knee roll, be aware that the soft tissue around this area may be pushed anteriorly, hindering palpation.

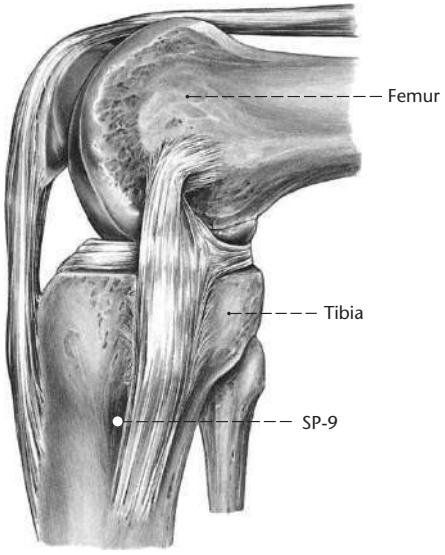

Fig. 3.103

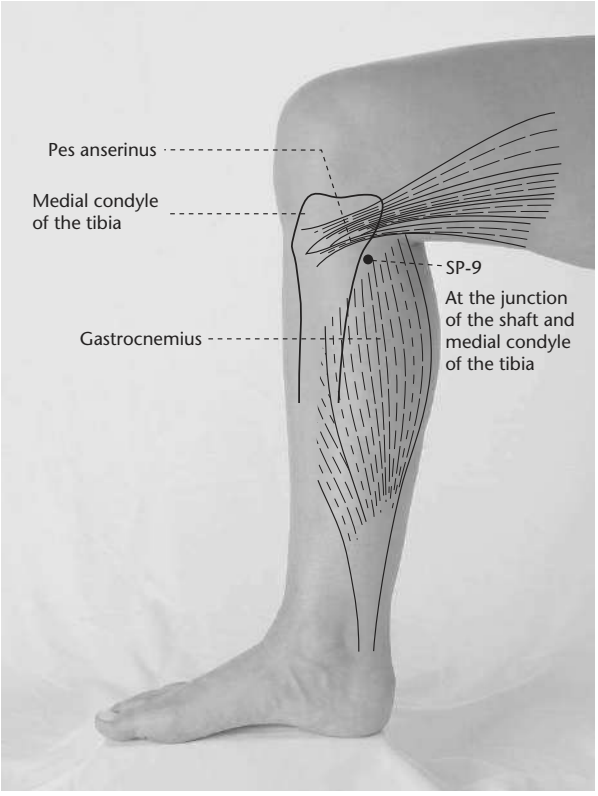

Fig. 3.104

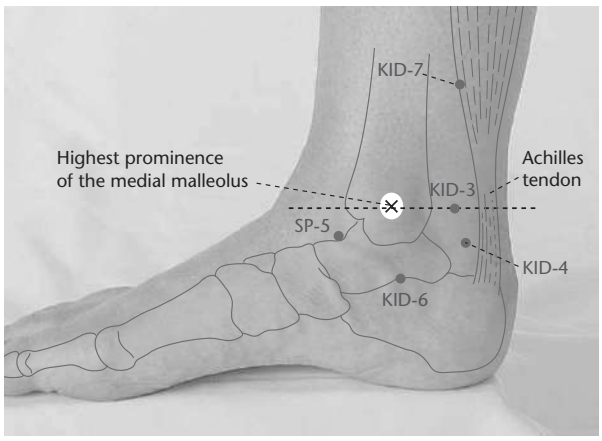

Fig. 3.105

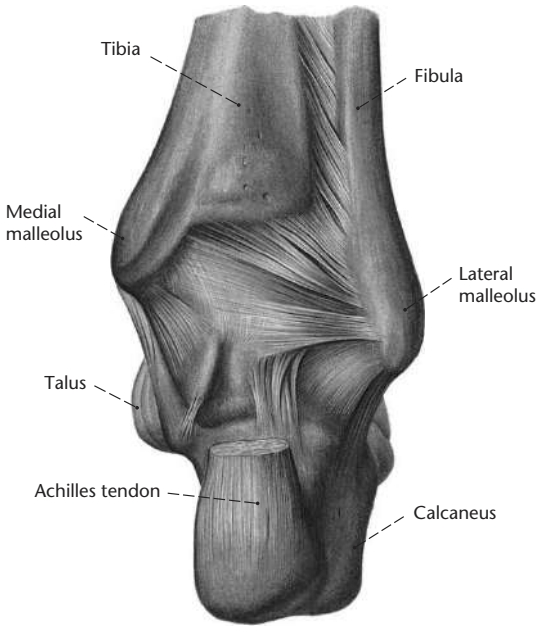

Fig. 3.106

3.6.2 Ankle and Foot

Highest prominence of the medial/lateral malleolus (→ Figs 3.105, 3.106)

Please note: The highest prominence of the malleolus does not refer to its lower border, but literally to its most protruding tip. The prominences are important reference points for the lower leg and ankle. → **Ex-LE-8** (*neihuajian*) is located on the highest prominence of the lateral malleolus, while → **Ex-LE-9** (*waihuajian*) is its medial counterpart.

Please also note: Contrary to some textbooks, the lateral and medial prominences are anatomically not located at the same level. There is approximately a 1 cun (→ Chapter 1, Fig. 1.2) difference in level. Thus, → **BL-60** and → **KID-3** are not located directly opposite each other, but rather occupy a comparable position.

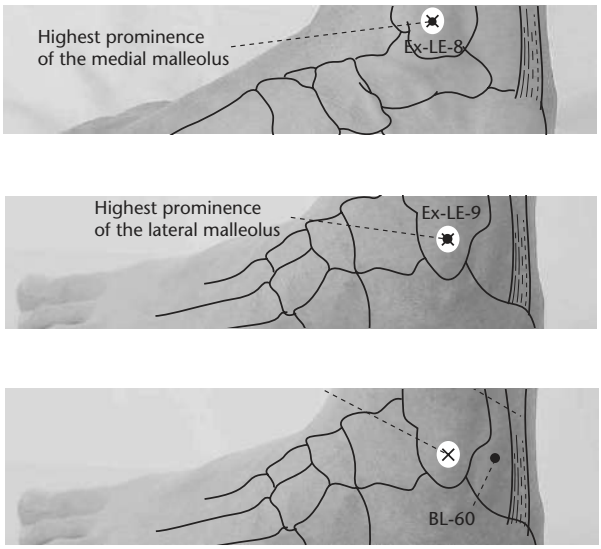

Fig. 3.107

Joint space on the anterior ankle (→ Figs 3.108, 3.109)

The joint space on the anterior aspect of the ankle (between the malleoli) is best palpated by passively flexing and extending the foot. → **LIV-4** and → **ST-41** are both located here.

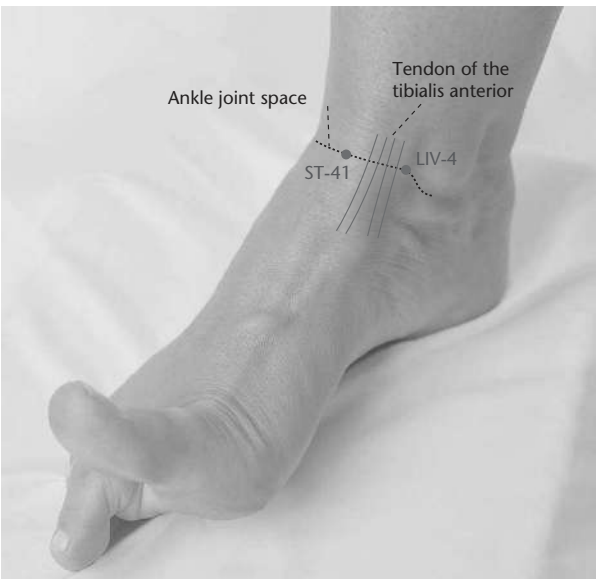

Fig. 3.108

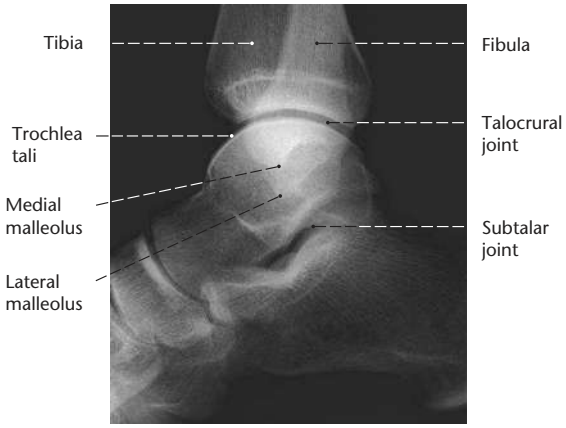

Fig. 3.109

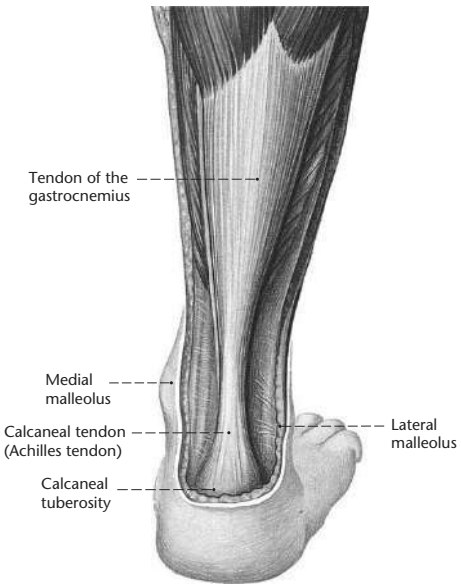

Fig. 3.110

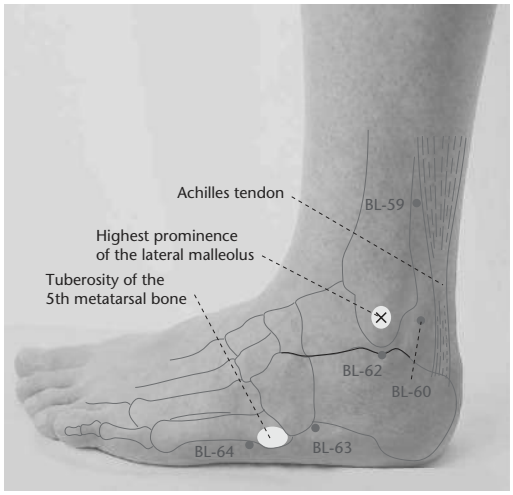

Fig. 3.111

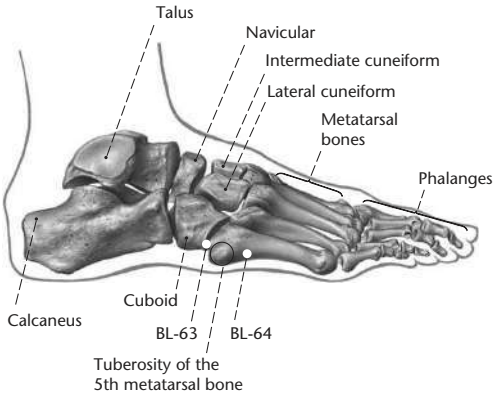

Fig. 3.112

Achilles tendon (→ Fig. 3.110)

On the distal aspect of the lower leg and on the heel, the Achilles tendon defines the location of → **KID-3, KID-4**, and **KID-7** on its medial aspect and of → **BL-59** and **BL-60** on its lateral aspect.

Tuberosity of the 5th metatarsal bone (→ Figs 3.111, 3.112)

The tuberosity at the proximal end of the 5th metatarsal bone is the most palpable orientation point on the lateral aspect of the midfoot. → **BL-63** is located in the depression proximal to the tuberosity, while → **BL-64** is located in the depression distal to it, at the junction of the head and the shaft.

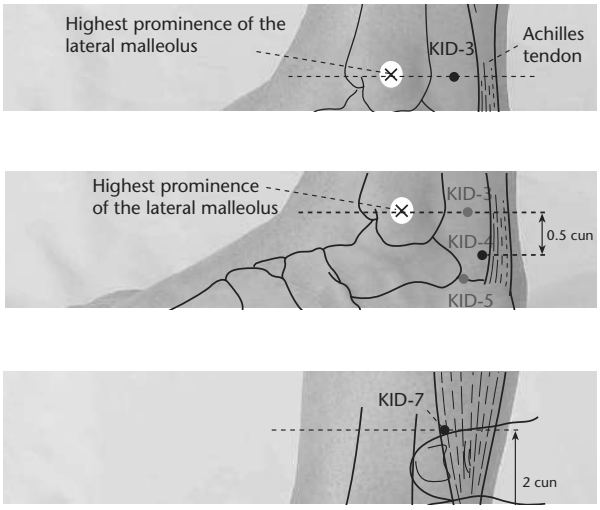

Fig. 3.113

Base of the 1st metatarsal bone (→ Figs 3.114, 3.115)

When palpating along the medial aspect of the midfoot in a proximal direction, the proximal end (base) of the 1st metatarsal bone forms the first distinct prominence. → **SP-4** is located distal to this bony orientation point, at the junction of the shaft and the base.

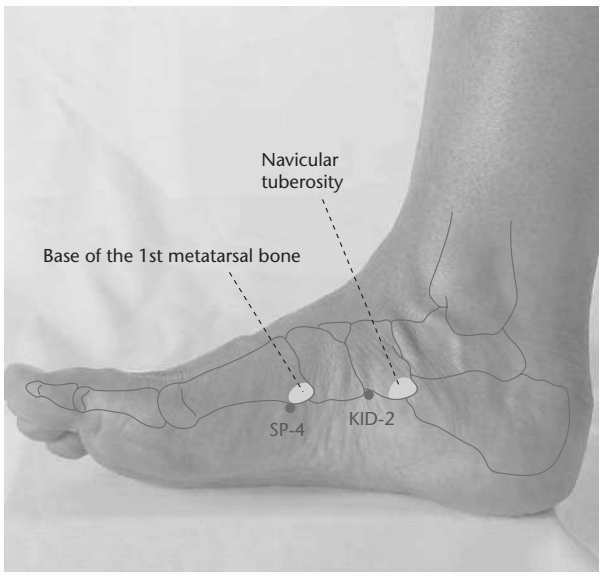

Fig. 3.114

Navicular tuberosity (→ Fig. 3.115)

When palpating along the medial side of the foot, past → **SP-4** and the base of the 1st metatarsal bone, a further bony structure can be clearly felt. This is the navicular tuberosity, where → **KID-2** is located.

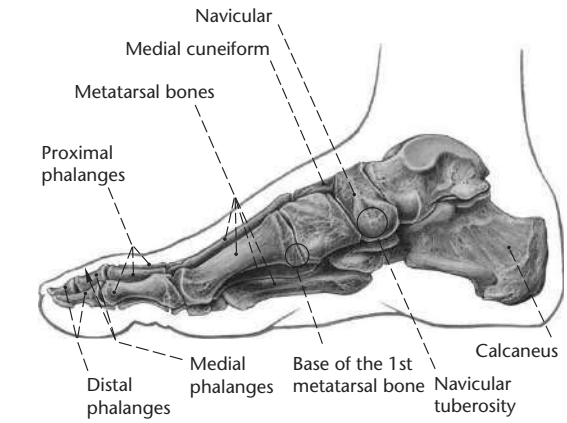

Fig. 3.115

4 Acupuncture Points of the Twelve Primary Channels

Claudia Focks, Ulrich März

4.1 The Lung Channel System – Hand-*taiyin* (*shou tai yin jing luo*)

4.1.1 The Lung Primary Channel (*shou tai yin jing*)

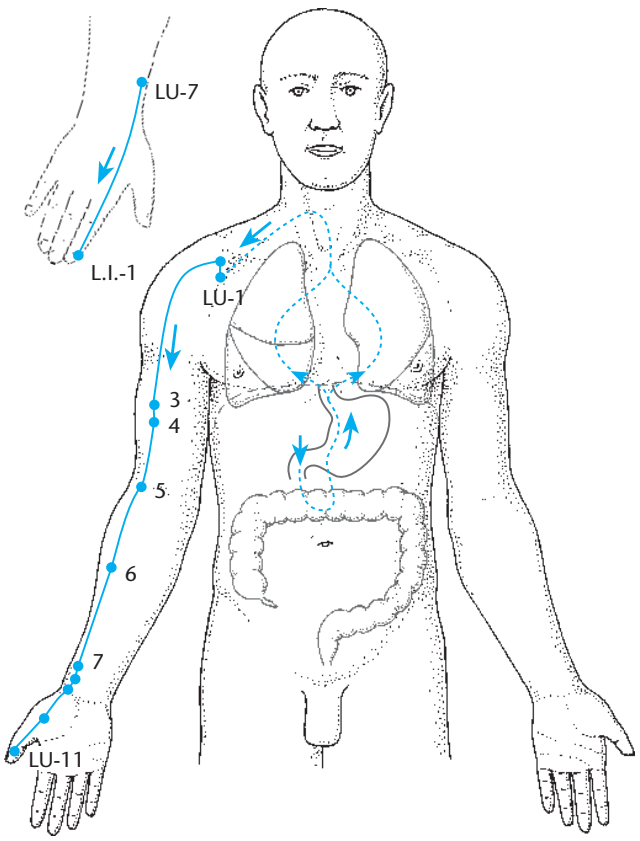

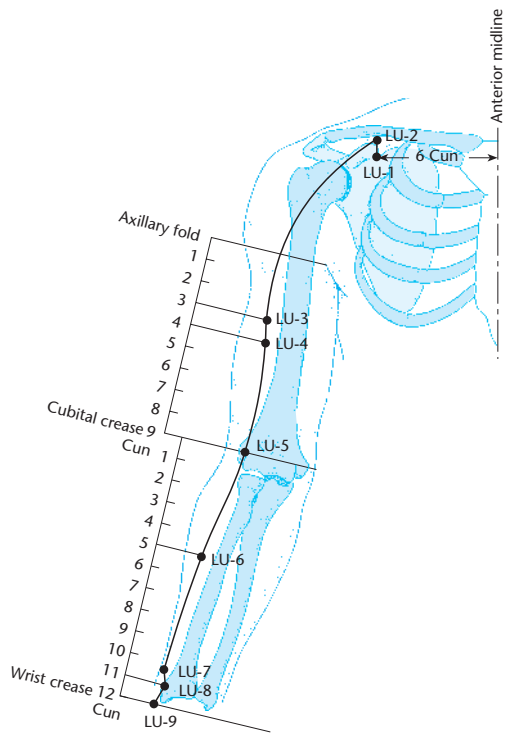

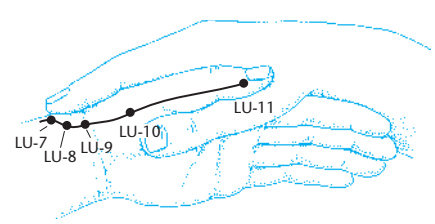

Pathway

The internal section of the Lung primary channel originates in the Middle Burner in the region of the Stomach (*wei*)

- ➡ descends to connect with the Large Intestine (*dachang*), its paired *fu*-Organ
- ➡ curves upwards again to the Stomach (*wei*)
- ➡ penetrates the diaphragm
- ➡ enters the Lung (*fei*), its pertaining *zang*-Organ
- ➡ ascends to the throat region
- ➡ and curves towards the lateral thoracic region.

The Lung primary channel emerges at LU-1 (*zhongfu*) on the lateral thoracic wall at the level of the first intercostal space

- ➡ travels along the anterolateral aspect of the upper arm and forearm
- ➡ ends at LU-11 (*shaoshang*) on the thumb, at the radial aspect of the corner of the nail.

A branch separates from the primary channel at LU-7 (at the styloid process of the radius) and travels along the radial side of the index finger to its tip, where it connects with → **L.I.-1** (*shangyang*) (hand Yin–Yang connection of the first great circuit).

Ch04.1-F10028.qxd 2/23/08 7:02 PM Page 80

Clinical importance (→ 1.2)

Exterior (*biao*) signs and symptoms: Fever, aversion to cold, nasal blockage, headache, pain in the region of the thorax, shoulder and back, pain with cold sensations along the channel
Interior (*li*) or *zangfu*-Organ signs and symptoms: Cough, asthma, wheezing, dyspnoea, shortness of breath, fullness and tightness of the chest, expectoration of phlegm, dry throat, changed colour of the urine, irritability, heat sensation in the palms, abdominal fullness and distension, diarrhoea
Excess (*shi*): Shoulder and back pain, Wind-Cold attack with spontaneous sweating, frequent urination, yawning
Deficiency (*xu*): Shoulder and back pain with aversion to cold, shortness of breath, changed colour of urine

Connections of the Lung primary channel (→ 1.2)

Connections with other channels

Large Intestine primary channel (*shou yang ming jing*)

Connection: Hand Yin–Yang connection of the first great circuit
Location: LU-7 → L.I.-1 (on the hand). A branch of the Lung primary channel separates at **LU-7** (*lieque*) and connects with the Large Intestine primary channel at **L.I.-1** (*shangyang*), according to some authors also with **→ L.I.-4** (*hegu*).
Circulation: Circadian (according to the Organ clock)
Importance: Interior–Exterior relationship

Spleen primary channel (*zu tai yin jing*)

Connection: Paired according to the six channel theory (hand–foot pairing): *tai yin* (Yin axes of the first great circuit)
Location: SP-20 → LU-1 (on the thorax). A branch of the Spleen primary channel separates at **→ SP-20** (*zhourong*) and connects with the Lung primary channel at **→ LU-1** (*zhongfu*).
Circulation: Non-circadian (not according to the Organ clock)
Importance: Above–below relationship

Liver primary channel (*zu jue yin jing*)

Connection: Deep Yin–Yin connection
Location: LIV → LU (on the thorax). An internal branch of the Liver primary channel emerges at the Liver, passes the diaphragm and spreads inside the Lung (*fei*), forming a network with the Lung primary channel.
Circulation: Circadian (according to the Organ clock)
Importance: Closes the circle of the first great circuit of the Nutritive Qi (*ying qi*, → 1.1.4)

Ren mai

Connection: *Ren mai* → Lung primary channel
Importance: Closes the circle of the second circuit of the Nutritive Qi (*ying qi*, → 1.1.4). The Lung primary channel as well as the great circuit (circulation through the 12 primary channels) receives Qi from the small circuit (*ren mai – du mai*).

Connections to the *zangfu*-Organ systems

Lung (*fei*), **Large Intestine (*dachang*)**, Stomach (*wei*)

4.1.2 The Lung Divergent Channel (*shou tai yin jing bie*)

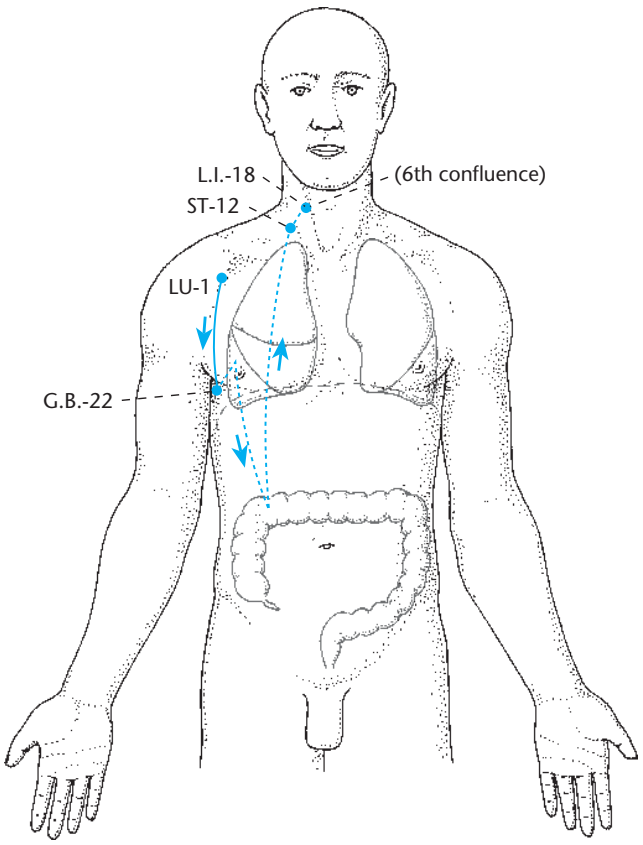

Pathway

The Lung divergent channel separates from the Lung primary channel in the axillary region

- travels anteriorly to → **G.B.-22** (*yuanye*) on the midaxillary line, 3 cun inferior to the axilla
- enters the thoracic region
- disperses in the Lung (*fei*), its pertaining *zang*-Organ
- descends to the Large Intestine (*dachang*), its paired *fu*-Organ
- ascends again and emerges in the region of the supraclavicular fossa at → **ST-12** (*quepen*)
- traverses the neck in a cranial direction and connects with the Large Intestine primary channel at → **L.I.-18** (*futu*) at one of the six *he*-confluences* (L.I./LU as 6th confluence. → 1.3).

There exist variations regarding the channel pathway due to differing interpretations of the *Ling Shu* (→ Solinas et al. 1998).

* According to some authors the 6th confluence is localised at → **ST-12** (*quepen*).

Ch04.1-F10028.qxd 2/23/08 7:02 PM Page 81

Clinical importance

- Strengthens the connection between the Lung and the Large Intestine (*zangfu*-Organ systems). Points on the Large Intestine primary channel can therefore be used for disorders of the Lung, and vice versa points on the Lung primary channel can treat disorders of the Large Intestine.
- Creates a network between the Lung primary channel and the neck region at the confluent point → **L.I.-18**: thus LU points such as → **LU-10** or → **LU-11** may be used to treat the neck region.

4.1.3 The Lung Sinew Channel (*shou tai yin jing jin*)

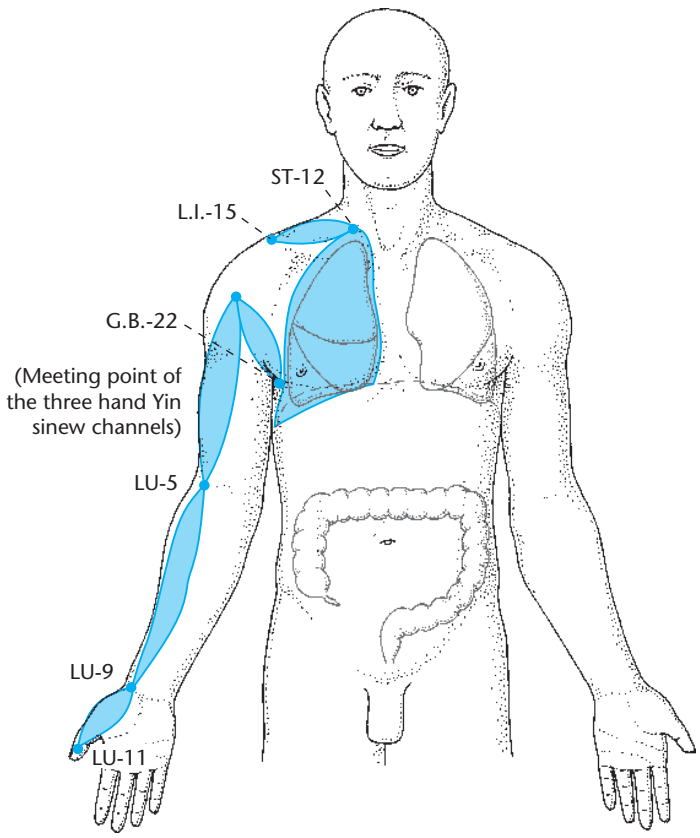

Pathway

The Lung sinew channel

- ➡ begins on the thumb at **LU-11** (*shaoshang*), at the radial aspect of the corner of the nail
- ➡ spreads along the radial aspect of the thumb and the 1st metacarpal bone as well as the thenar eminence
- ➡ meets **LU-9** (*taiyuan*) and binds (*jie*) at the radial wrist joint space
- ➡ travels proximally along the anterolateral aspect of the forearm
- ➡ reaches the elbow at **LU-5** (*chize*) and binds (*jie*) at the biceps tendon
- ➡ follows the biceps brachii muscle and the lateral portion of the deltoid muscle
- ➡ binds (*jie*) at the anterior shoulder region near the acromioclavicular joint
- ➡ continues to the axilla and meets the other hand Yin sinew channels (P, HE) at → **G.B.-22** (*yuanye*) and binds (*jie*) there
- ➡ continues deep to the pectoralis major muscle and enters the supraclavicular fossa at → **ST-12** (*quepen*)
- ➡ courses to → **L.I.-15** (*jianyu*) and returns to → **ST-12** (*quepen*)
- ➡ penetrates the supraclavicular fossa and disperses in the thoracic and hypochondriac region and the diaphragm (→ Fig.)

Clinical importance

Pathology: Stiffness, cramps and pain along the sinew channel. *Xi fen*: spasms and pain in the thorax and hypochondriac region. In severe cases formation of nodules below the right lateral costal region and distension along the lateral costal region.

Indication: Mainly for *bi*-syndromes (painful obstruction syndrome) along the LU channel.

The area covered by the Lung sinew channel is larger than that covered by the Lung primary channel. This explains why the indications of points on the Lung primary channel include disorders and diseases of the thoracic and hypochondriac region.

4.1.4 The Lung *luo*-Connecting Vessel System (*shou tai yin luo mai*)

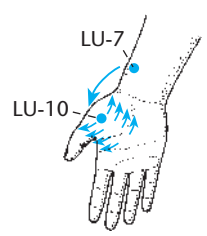

Pathway

The Lung *luo*-connecting channel system separates from the Lung primary channel at **LU-7** (*lieque*) (→ 8.1.2) and forms a three-dimensional reticular network, dividing into multiple branches and sub-branches (*sun luo*, *fu luo*, *xue luo* → 1.5) within the surrounding tissue.

- ➡ Horizontal divisions course to the Interiorly–Exteriorly paired Large Intestine primary channel; according to some schools (for example Van Nghi, → Appendix) they travel as a transverse Lung *luo* vessel to the *yuan*-source point → **L.I.-4** (*hegu*).
- ➡ A longitudinally orientated division spreads on the palm and thenar eminence.

Ch04.1-F10028.qxd 2/23/08 7:02 PM Page 82

Clinical importance (→ 8.1.2)

Pathology

- **Excess (*shi*)**: Hot and burning sensations in the palms and the wrist joints
- **Deficiency (*xu*)**: Yawning, shortness of breath, frequent urination, enuresis

4.1.5 Cutaneous Region (*tai yin pi bu*)

See description and figures → 1.6.

4.1.6 Points of the Lung Primary Channel (Overview)

Specific points according to function

- ***Yuan*-source point (→ 8.1.1): LU-9** (*taiyuan*) ■■
- ***Luo*-connecting point (→ 8.1.2.): LU-7** (*lieque*) ■■
- ***Xi*-cleft point (→ 8.1.3): LU-6** (*kongzui*)
- **Pertaining Back-*shu* point (→ 8.1.4): BL-13** (*feishu*) ■■
- **Pertaining Front-*mu* point (→ 8.1.5): LU-1** (*zhongfu*) ■■
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Wood): **LU-11** (*shaoshang*) ■■
  ying-spring point (Fire): **LU-10** (*yuji*)
  shu-stream point (Earth), tonification point: **LU-9** (*taiyuan*) ■■
  jing-river point (Metal), *ben* point: **LU-8** (*jingqu*) ■■
  he-sea point (Water), sedation point: **LU-5** (*chize*) ■■
- **Influential *hui*-meeting point (→ 8.1.7) of the vessels: LU-9** (*taiyuan*) ■■
- **Opening point (→ 8.1.8) of the *ren mai*: LU-7** (*lieque*) ■■
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the SP primary channel: **LU-1** (*zhongfu*) ■■
  – of other channels with the LU channel: –
- **Gao Wu Command point (→ 8.1.11) for the occiput and nape: LU-7** (*lieque*) ■■
- **Window of Heaven point (→ 8.1.12): LU-3** (*tianfu*)
- **Ma Dan Yang Heavenly Star points (→ 8.1.14): LU-7** (*lieque*) ■■
- **Sun Si Miao Ghost point (→ 8.1.15): LU-11** (*shaoshang*) ■■

Points according to region

- **Local points (→ 8.2.1):** Lung – **LU-1** (*zhongfu*) ■■; elbow – **LU-5** (*chize*) ■■; wrist – **LU-7** (*lieque*) ■■
- **Adjacent points (→ 8.2.1):** hands – **LU-7** (*lieque*) ■■
- **Distal points (→ 8.2.1):** occiput – **LU-7** (*lieque* ■■); throat – **LU-10** (*yuji*) – **LU-11** (*shaoshang*); Lung – **LU-5** (*chize*) ■■, **LU-7** (*lieque*) ■■; shoulder – **LU-7** (*lieque*) ■■

Specific points according to the course of the channel (in numeric order)

- **LU-1** (*zhongfu* ■■): associated Front-*mu* point (→ 8.1.5), *jiaohui*-meeting point with the SP channel (→ 8.1.10)
- **LU-3** (*tianfu*): Window of Heaven point (→ 8.1.2)
- **LU-5** (*chize*) ■■: *he*-sea point (Water) (→ 8.1.6); sedation point; distal point of the Lung (→ 8.2.1)
- **LU-6** (*kongzui*): *xi*-cleft point (→ 8.13)
- **LU-7** (*lieque*) ■■: *luo*-connecting point (→ 8.1.2); opening point (→ 8.1.8) of the *ren mai*; Gao Wu Command point (→ 8.1.11) of the nape and occiput; Ma Dan Yang Heavenly Star point (→ 8.1.14); distal point for the occiput, shoulder and Lungs (→ 8.2.1); adjacent point for the hands (→ 8.2.1)
- **LU-8** (*jingqu*): *jing*-river point (Metal) (→ 8.1.6); *ben* point (Five Phase point)
- **LU-9** (*taiyuan*) ■■: *yuan*-source point (→ 8.1.1), hui-meeting point (→ 8.1.7) of the vessels; *shu*-stream point (Earth) (→ 8.1.6); tonification point
- **LU-10** (*yuji*): *ying*-spring point (Fire) (→ 8.1.6); distal point for the neck (→ 8.2.1)
- **LU-11** (*shaoshang*) ■■: Sun Si Miao Ghost point (→ 8.1.15); *jing*-well point (Wood) (→ 8.1.6); distal point for the neck

Ch04.1-F10028.qxd 2/23/08 7:02 PM Page 83

Central Residence ZHONGFU LU-1

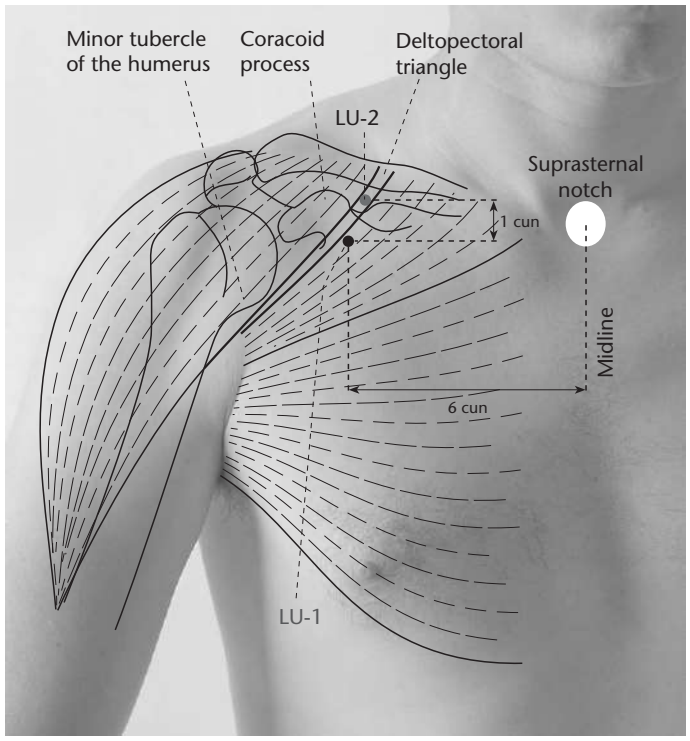

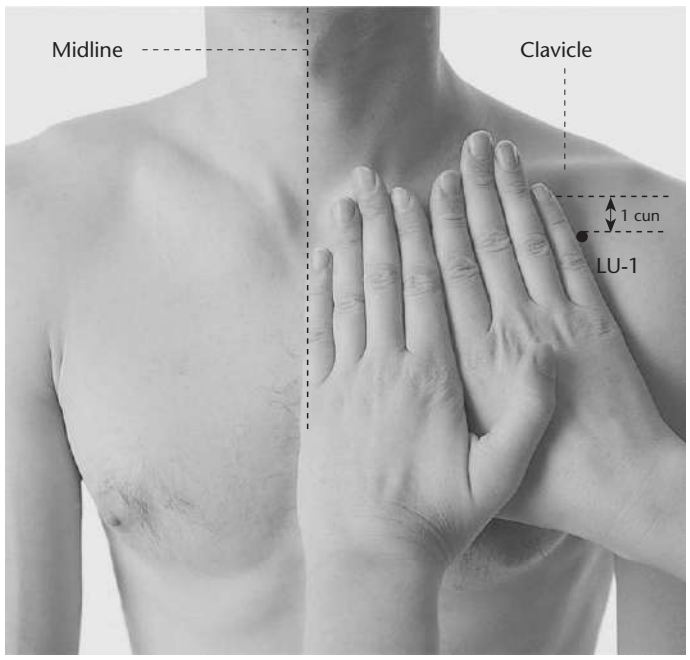

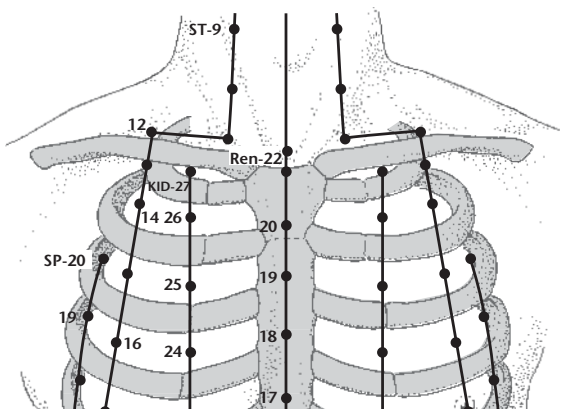

Location

6 cun lateral of the anterior midline and approximately 1 cun below → **LU-2,** slightly medial to the lower border of the coracoid process.

How to find

First, locate → **LU-2** in the deltopectoral triangle, which is bordered superiorly by the clavicle and laterally by the coracoid process (within the deltoid muscle). From there, palpate along the border of the deltoid for 1 cun and there locate **LU-1** slightly medial to the lower border of the deltoid on the thoracic wall, approximately 6 cun lateral to the midline. For orientation: when externally rotating the arm with the elbow flexed at the same time, the coracoid process will remain static, while the minor tuberculum of the humerus will follow the movement. → **SP-20** is also located 6 cun lateral to the midline on the level of the 2nd intercostal space. Located approximately on the same level are → **Ren-20,** → **KID-26** and → **ST-14**.

Needling

0.5–1 cun obliquely in a cranial (approximately 45°) and lateral direction towards the coracoid process.

Actions/Indications

- Regulates and descends the Lung Qi
- Clears Heat in the Upper Burner, transforms Phlegm
- Regulates the water passages
- Moves Qi locally and in the sinew channel

Special features

Front-*mu* point of the Lung, meeting point with the Spleen channel, entry point. Important point for disorders of the Lung.

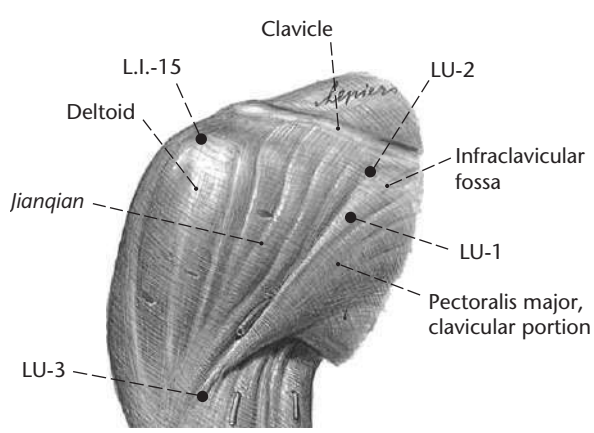

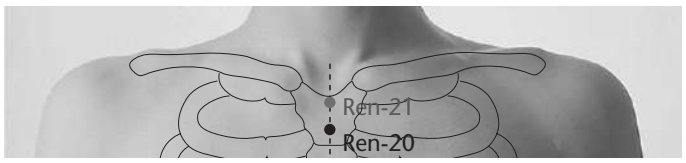

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 84

LU-2 *Cloud Gate* YUNMEN

Location

6 cun lateral to the anterior midline, below the clavicle, in the centre of the deltopectoral triangle.

How to find

Find the angle formed by the lateral lower border of the clavicle and the coracoid process (within the deltoid muscle). Locate **LU-2** at the junction of the clavicular insertions of the deltoid and pectoralis muscles (deltopectoral triangle), in the centre of a clearly palpable depression in the thoracic wall, 6 cun lateral to the anterior midline.

→ **ST-13** is also located at the lower border of the clavicle, 4 cun lateral to the anterior midline or at the midpoint of the clavicle. Also located on approximately the same level are → **KID-27** (2 cun lateral to the midline) and → **Ren-21** (on the midline). → **LU-1** is located 1 cun below and in most cases slightly lateral to **LU-2**.

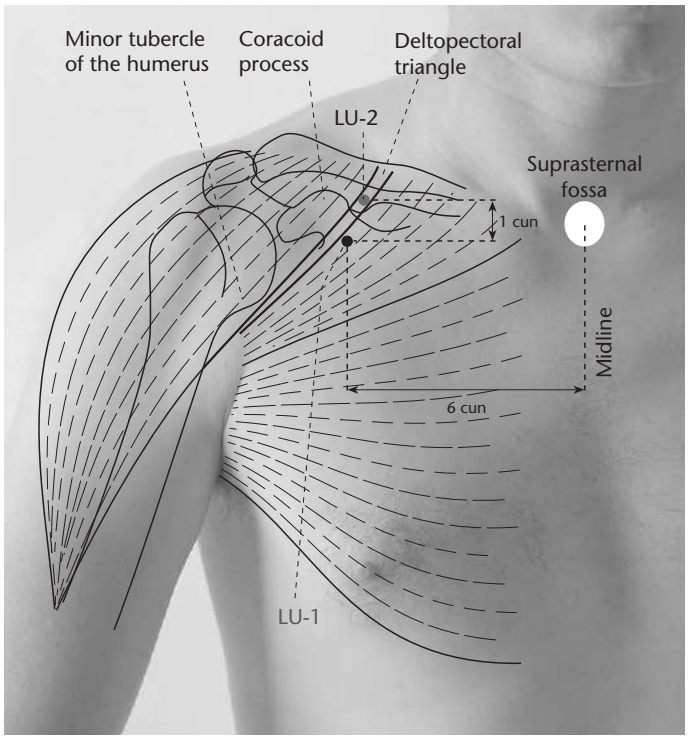

Needling

Obliquely to a maximal depth of 0.8 cun. No needling in a medial direction! Risk of pneumothorax!

Actions/Indications

- Clears Lung Heat, decends the Lung Qi
- Opens the channel and sinew channel
- Clears Heat from the extremities

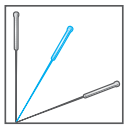

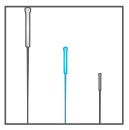

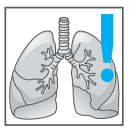

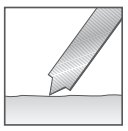

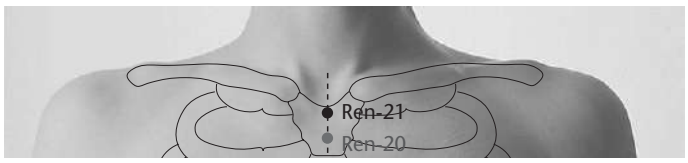

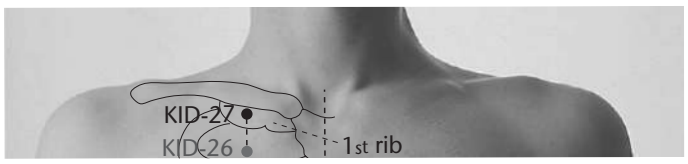

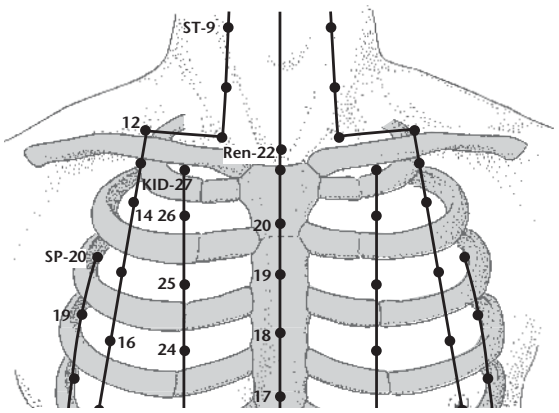

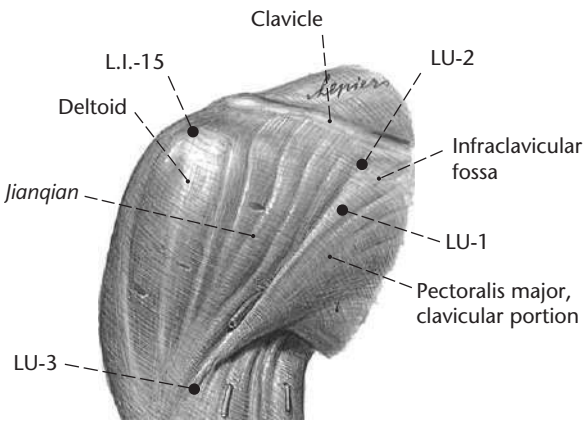

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 85

Heavenly Residence TIANFU LU-3

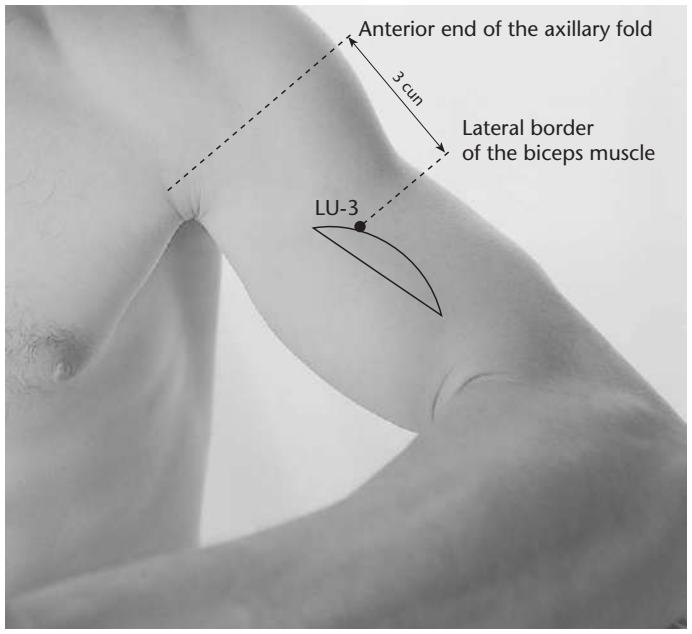

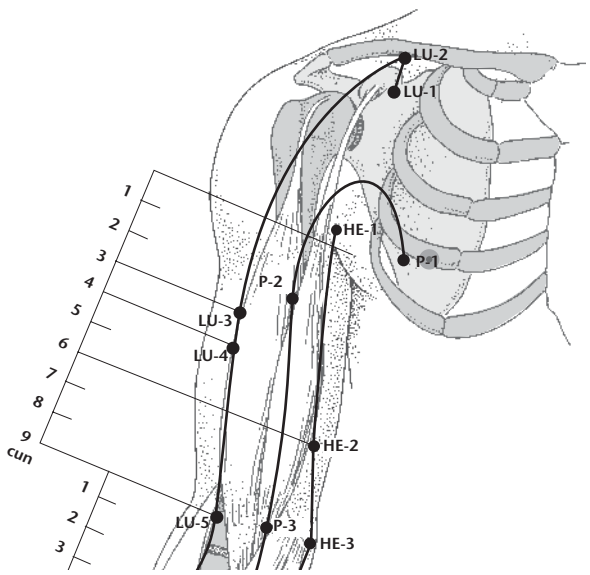

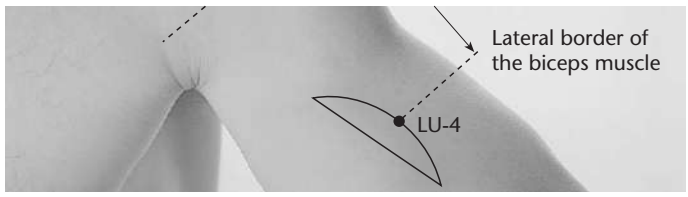

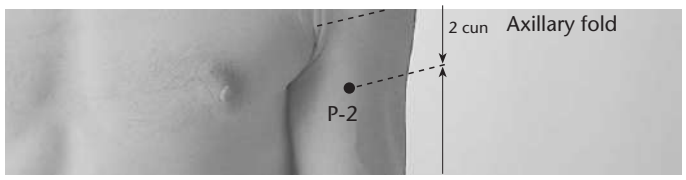

Location

On the medial aspect of the upper arm, 3 cun distal to the end of the anterior axillary fold, in the groove on the lateral border of biceps brachii.

How to find

Ask the patient to flex the biceps muscle against resistance. Locate **LU-3** 3 cun below the anterior end of the axillary fold (→ 2.2) in a groove on the lateral border of biceps brachii. Sometimes, you can feel the pulse of the brachial artery here. **Or:** Divide the distance of 9 cun (→ 2.2) between the end of the axillary fold and the cubital crease (→ **LU-5**) into thirds and locate **LU-3** a third of the distance from the axillary fold.
→ **LU-4** is located in the sulcus, 1 cun distal to **LU-3**. → **P-2** is located 2 cun below the axillary fold, between the two heads of the biceps brachii muscle.

Needling

Vertically 0.5–1 cun. Moxibustion is contraindicated according to some classics.

Actions/Indications

- Regulates and descends the Lung Qi
- Clears Lung Heat
- Cools the Blood and stops bleeding
- Calms the *po* (corporeal soul)
- Opens the channel and alleviates pain

Special features

Window of Heaven point

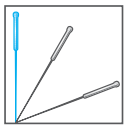
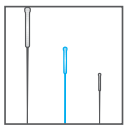
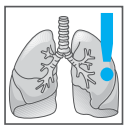

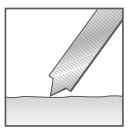

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 86

LU-4 *Clasping the White* XIABAI

Location

On the medial aspect of the upper arm, 4 cun distal to the end of the anterior axillary fold, in the depression on the lateral border of sulcus bicipitalis lateralis.

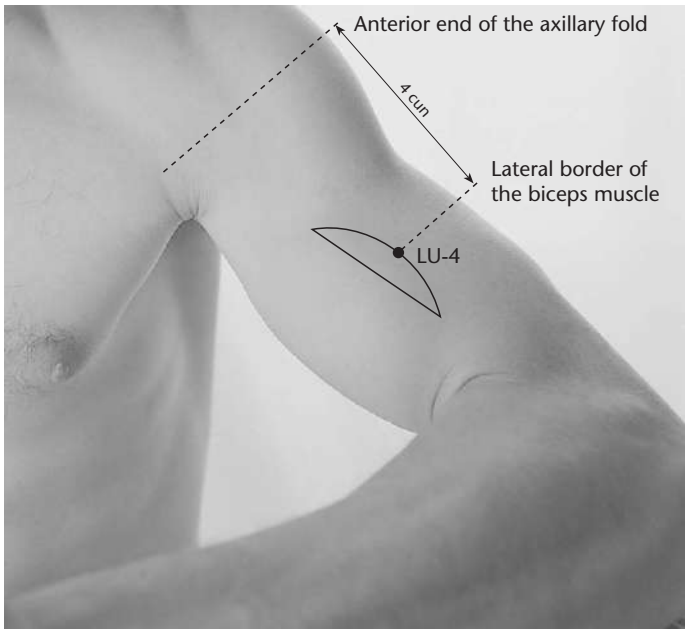

How to find

Ask the patient to flex their biceps muscle against resistance. Locate **LU-4** 4 cun below the anterior end of the axillary fold (→ 2.2) in the depression on the lateral border of the sulcus bicipitalis muscle. Sometimes you can feel the pulse of the brachial artery here. **Or quick method:** Spreading hands technique (→ 2.3.3): place the little fingers on the axillary fold and the cubital crease (at → **LU-5**) respectively (this distance corresponds to 9 proportional cun → 2.2) and determine the midpoint of this distance by joining the thumbs. Locate **LU-4** 0.5 proximal to the midpoint in the sulcus on the lateral border of the biceps muscle. → **LU-3** is also located in the sulcus, 1 cun proximal to **LU-4**.

Needling

Vertically 0.5–1 cun

Actions/Indications

- Regulates and moves Qi and Blood in the chest
- Descends the Lung Qi
- Opens the channel and alleviates pain

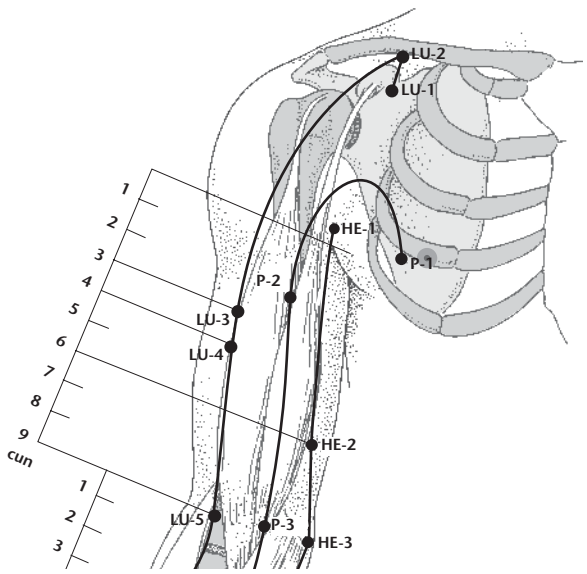

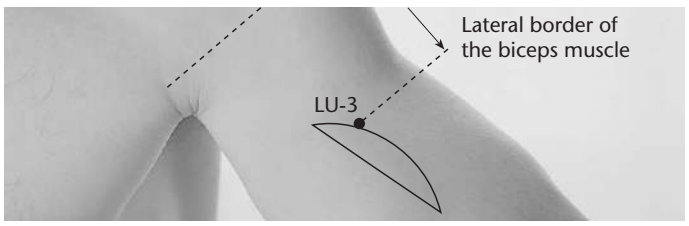

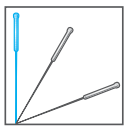

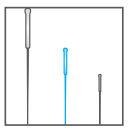

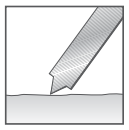

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 87

Cubit Marsh CHIZE LU-5

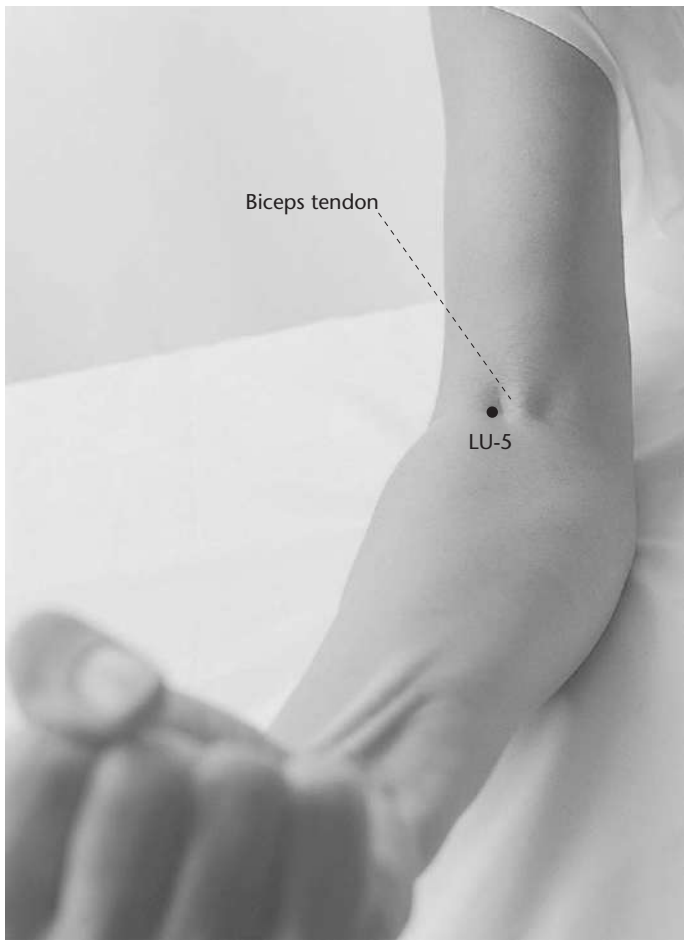

Right arm

Location
In the cubital crease, on the radial aspect of the tendon of the biceps.

How to find
By slightly flexing the forearm, the biceps tendon becomes more visible. **LU-5** is located on its radial aspect in the cubital crease. Also located in the area of the cubital crease are → **P-3** (ulnar to the tendon), → **HE-3** (at the ulnar end of the cubital crease) and → **L.I.-11** (between the radial end of the cubital crease and the lateral epicondyle).

Needling
Vertically 0.5–1 cun. Needle with the arm slightly flexed. Use bloodletting for Heat disorders. Caution: Avoid cubital vein.

Actions/Indications
- Clears Heat in the Upper Burner
- Descends the Lung Qi
- Regulates the water passages
- Opens the channel and *luo*-connecting vessels, alleviates pain

Special features
He-sea point, Water point, sedation point, distal point for the Lung, local point for the elbow. Use bloodletting for disorders of the skin and Lung due to excess and/or Heat.

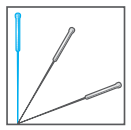

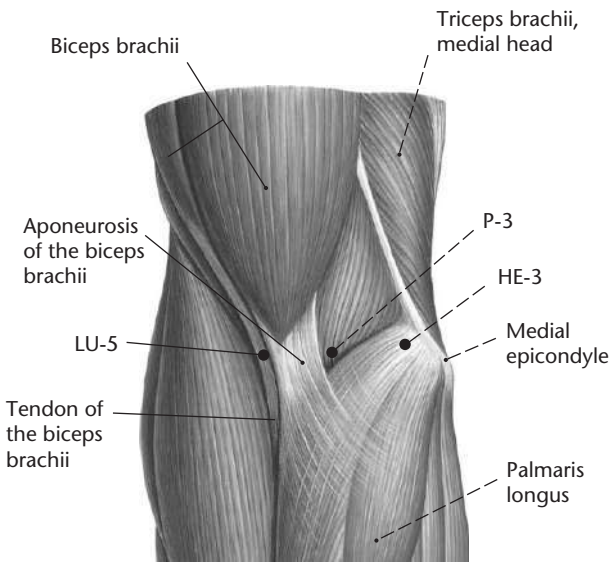

Left arm, anterior aspect

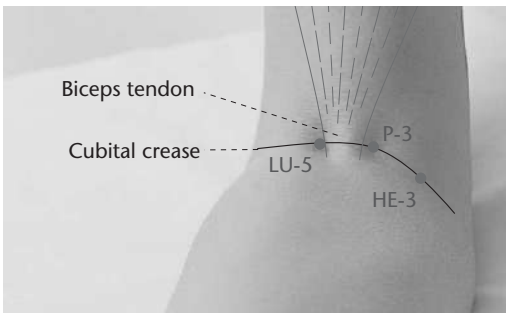

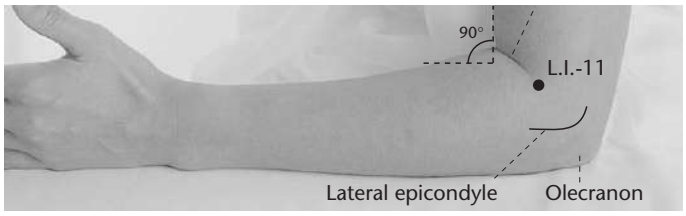

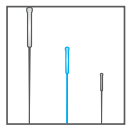

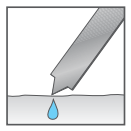

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 88

LU-6 *Biggest Hole* KONGZUI

Location

On the line connecting → **LU-5** (in the cubital crease) and → **LU-9** (on the wrist crease/joint space), 5 cun distal to → **LU-5** or 7 cun proximal to → **LU-9**.

How to find

The biceps tendon becomes more visible with the arm flexed. First, locate → **LU-5** radial to the biceps tendon in the cubital crease and → **LU-9** radial to the radial artery on the wrist joint space (→ 3.3.3) between the proximal row of carpals and the radius/ulna. By moving the hand, the joint space becomes more easily palpable. Next, find the midpoint of the distance between → **LU-5** and → **LU-9** using the spreading hands technique (→ 2.3.3). **LU-6** is located 1 cun proximal to the midpoint of this distance.

→ **P-4** is located more medially (between the tendons) and distally (1 cun distal to the midpoint of the distance).

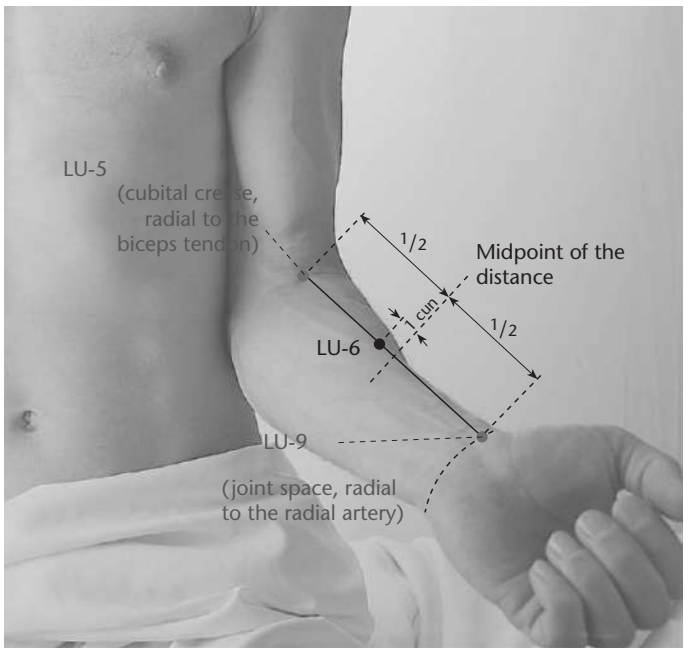

Needling

Vertically 0.5–1 cun

Actions/Indications

- Descends the Lung Qi
- Clears Lung Heat
- Cools Blood Heat, stops bleeding
- Expels Exterior pathogenic factors
- Opens the channel

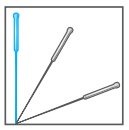

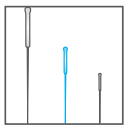

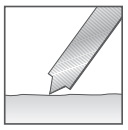

Special features

Xi-cleft point

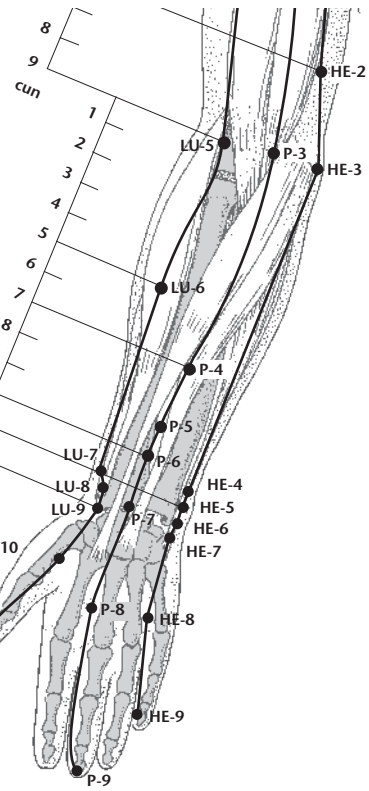

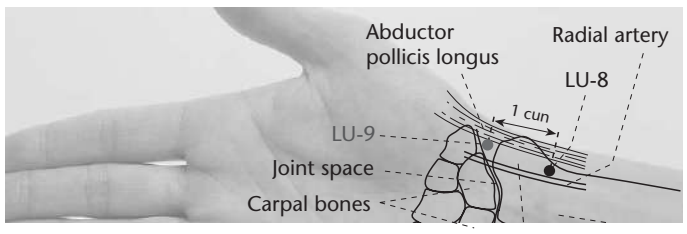

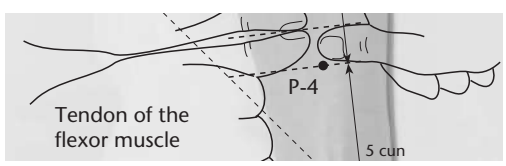

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 89

Broken Sequence LIEQUE LU-7

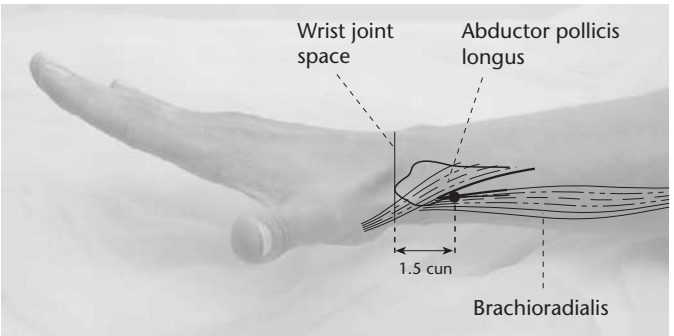

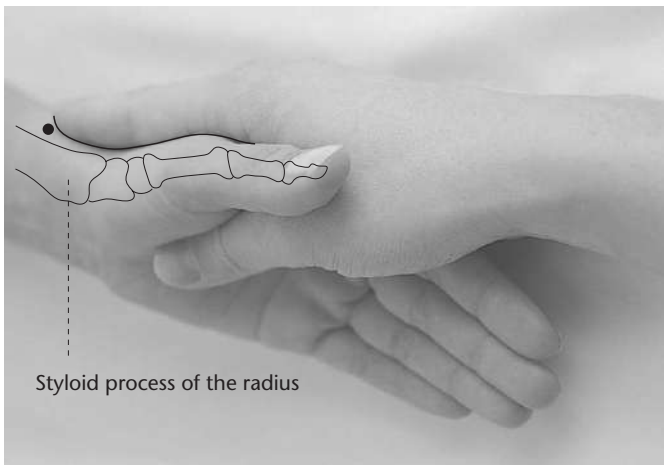

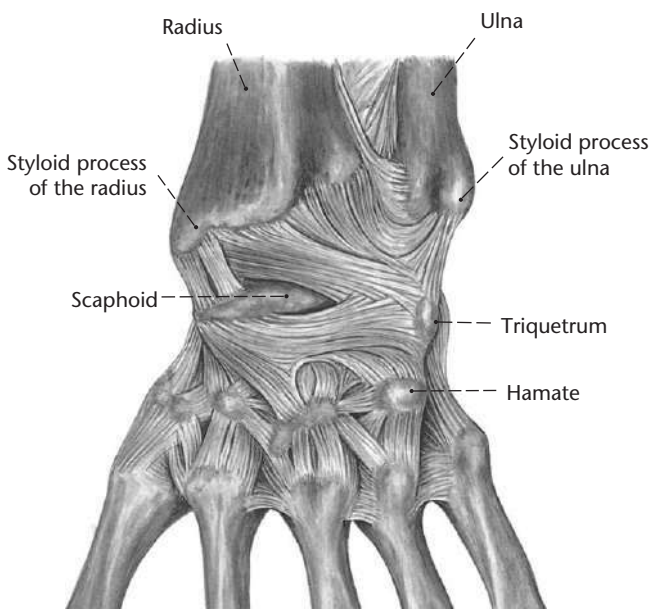

Left hand, dorsal aspect

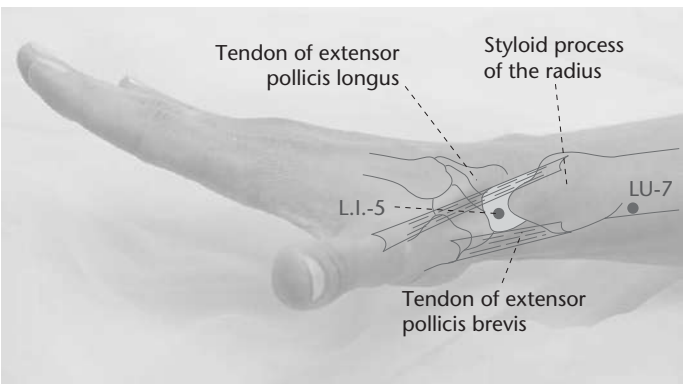

Location

On the radial aspect of the forearm, directly above the styloid process of the radius, approximately 1.5 cun proximal to the wrist joint space (wrist crease) in a V-shaped groove.

How to find

Place the palpating forefinger on the anatomical snuffbox (→ **L.I.-5**). From there, glide proximally over the styloid process of the radius and locate **LU-7** where the finger can palpate a cleft between two tendons (brachioradialis/abductor pollicis longus). **Or:** 'Tiger mouth grip' (→ Fig. 2.6): spread the thumb and forefinger of both hands, crossing them so that the forefinger of one hand comes to rest on the styloid process of the other, but avoiding bending the wrist. **LU-7** is located directly below the tip of the forefinger in a V-shaped groove

Needling

Lift the skin above the styloid process by pinching it. Insert the needle 0.5–1 cun obliquely and proximally (reducing) or distally (tonifying). Caution: Cephalic vein.

Actions/Indications

- Supports the occiput and head
- Releases the Exterior, expels Wind, descends the Lung Qi
- Opens and regulates the *ren mai*
- Regulates the water passages
- Opens the channel and *luo*-connecting channel, alleviates pain

Special features

Luo-connecting point, opening point of the *ren mai*, Ma Dan Yang Heavenly Star point, Gao Wu command point of the occiput, exit point.

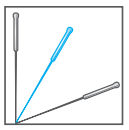
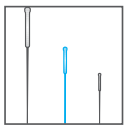

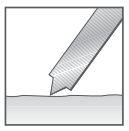
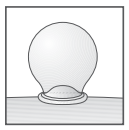

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 90

LU-8 Channel Gutter JINGQU

Location

Lateral to the radial artery, 1 cun proximal to the ventral wrist joint space (most distal wrist crease).

How to find

As the location of the wrist crease varies, use the joint space (→ 3.3.3) between the proximal row of carpals and the radius/ulna for orientation. By loosely moving the hand, the joint space is easily palpable. This is the level for palpating the radial pulse (using the finger tips, not the finger pads). → **LU-9** is located on the radial side of the artery, while **LU-8** is located 1 cun proximal to → **LU-9**.

→ **L.I.-5** is located close to **LU-8**, but on the dorsal aspect of the wrist joint in the anatomical snuffbox.

Needling

0.3–0.5 cun proximally or vertically. Avoid the radial artery by gently pushing it to the side when palpating for the point. Moxibustion is contraindicated according to some classics.

Actions/Indications

- Descends the Lung Qi, alleviates cough and wheezing
- Opens the pores and expels Wind in particular
- Opens the channel, especially locally

Special features

Jing-river point, Metal point, *ben* point (Five Phases).

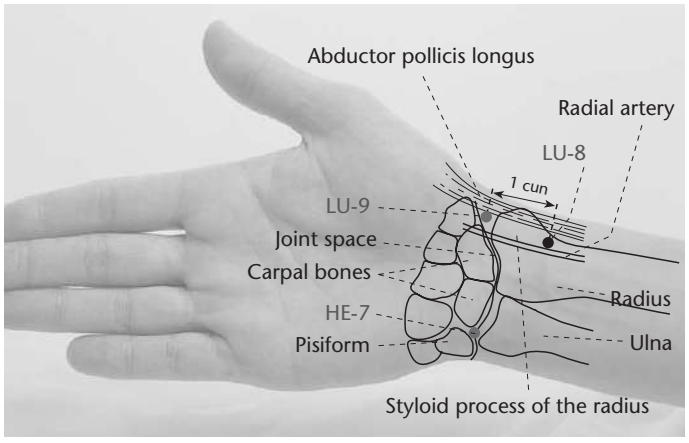

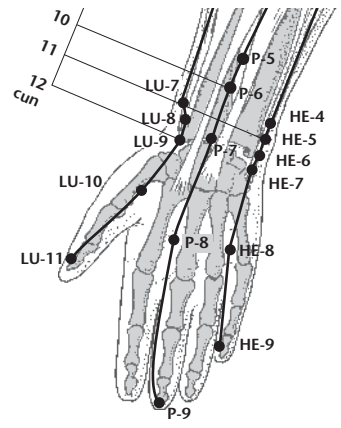

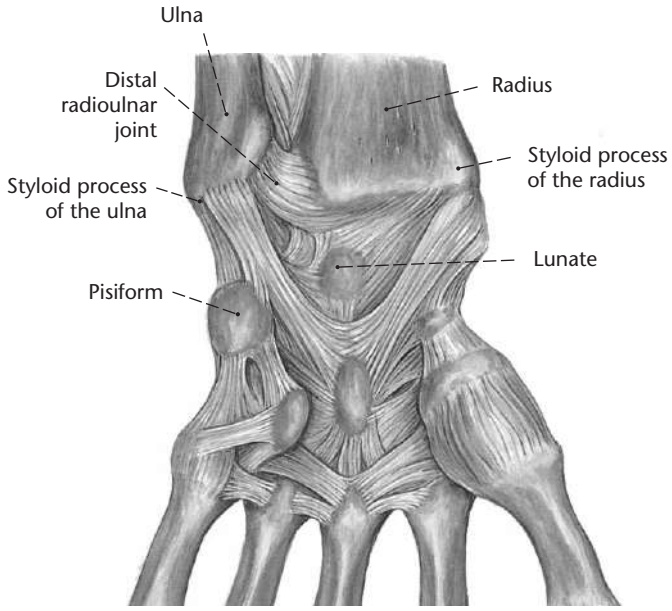

Left hand, palmar aspect

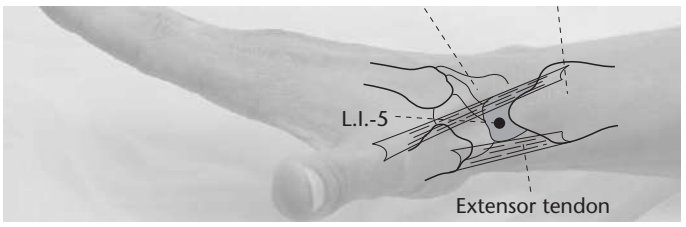

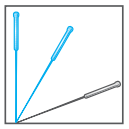
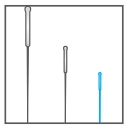

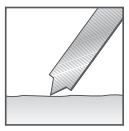

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 91

Supreme Abyss TAIYUAN LU-9

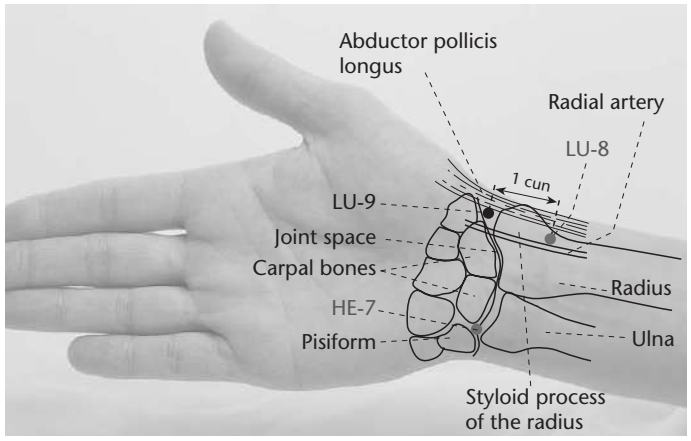

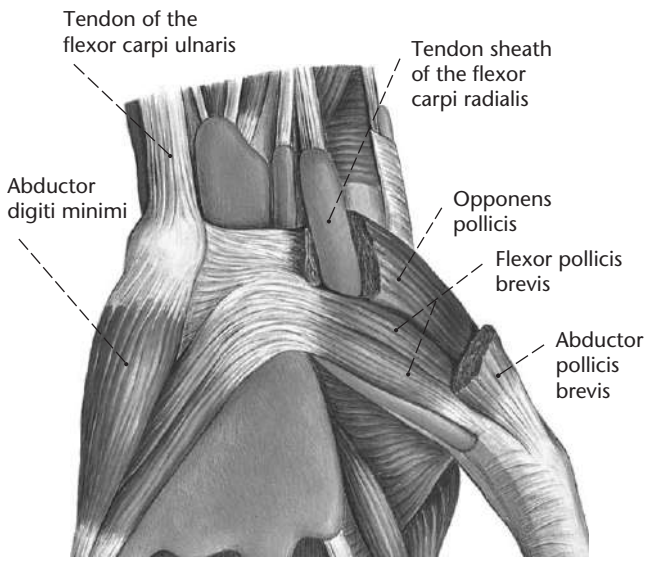

Left hand, palmar aspect

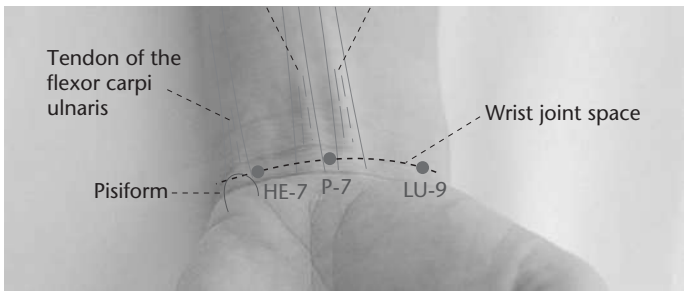

Location

On the ventral aspect of the wrist, at the level of the wrist joint space (most distal wrist crease), on the radial aspect of the radial artery and ulnar to the tendon of the abductor pollicis longus muscle.

How to find

As the location of the wrist crease varies, the ventral joint space (→ 3.3.3) between the proximal row of carpals and the radius/ulna is a more reliable landmark for orientation. By loosely moving the hand, the joint space is easily palpable. Palpate the radial artery (pulsation) on the radial side and locate **LU-9** on its radial aspect. Located on the same level are → **P-7** (between two tendons) and → **HE-7** (radial to the insertion of the tendon of the flexor carpi ulnaris on the pisiform bone).

Needling

Vertically 0.2–0.5 cun. Caution: Avoid radial artery. If the needle is pulsating (desired effect), no additional stimulation!

Actions/Indications

- Strengthens the Lung, transforms Phlegm, descends the Lung Qi
- Regulates and harmonises the vessels (relationship between *zong qi* and Blood circulation)
- Opens the channel, alleviates pain

Special features

Yuan-source point, *shu*-stream point, Earth point, tonification point, *hui*-meeting point of the vessels. Important point for tonifying the Lung Qi and Yin, especially with chronic disorders.

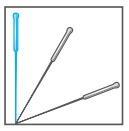
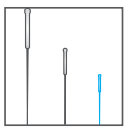

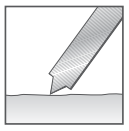

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 92

LU-10 *Fish Border* YUJI

Location
At the midpoint of the palmar border of the 1st metacarpal bone.

How to find
With the thumb relaxed, palpate the 'belly' of the thenar eminence from palmar to lateral (radial), then palpate the first metacarpal bone. **LU-10** is located at the midpoint of its palmar 'border'.
→ **L.I.-4** is located on the dorsal aspect of the hand, close to the 2nd metacarpal bone.

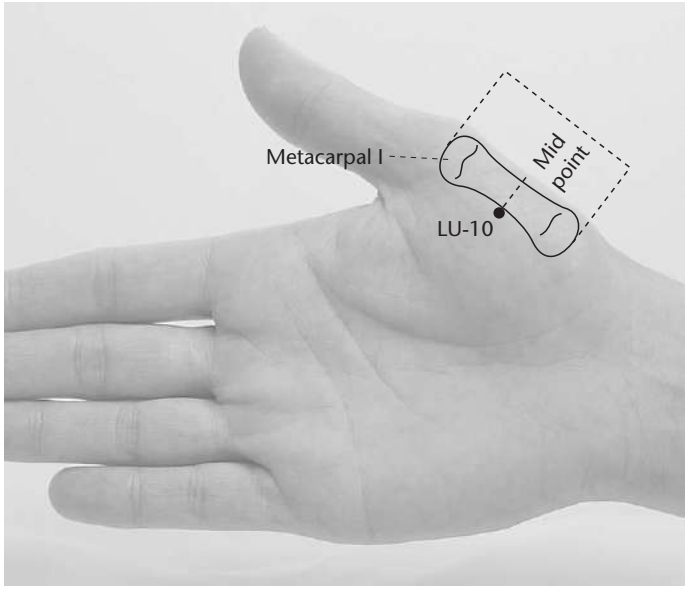

Needling
Vertically 0.5–1 cun. Caution: Painful!

Actions/Indications
- Regulates the Lung Qi
- Cools Blood Heat
- Expels pathogenic factors from the throat
- Descends rebellious Qi
- Harmonises the Stomach and Heart

Special features
Ying-spring point, Fire point, important distal point for sore throat caused by Wind-Heat.

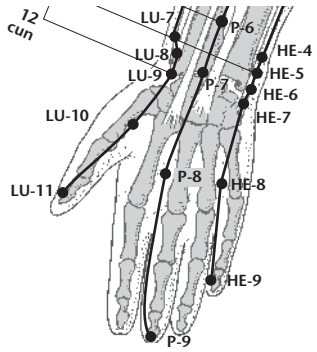

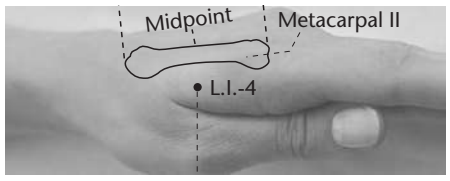

Ch04.1-F10028.qxd 2/23/08 7:03 PM Page 93

Lesser Shang SHAOSHANG LU-11

Location
On the thumb, 0.1 cun from the radial corner of the nail.

How to find
The point is located at the junction of two tangents running along the proximal and radial border of the thumb nail, approximately 0.1 cun from the actual border of the nail.

Needling
0.1–0.2 cun vertically or obliquely in a proximal direction or prick to bleed. Avoid needling the perionychium.

Actions/Indications
- Frees the senses
- Regulates the Lung Qi
- Clears Heat (especially from the throat)
- Opens the channel

Special features
Jing-well point, Wood point, Sun Si Miao Ghost point, important distal point for sore throat caused by Wind-Heat.

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 95

4.2 The Large Intestine Channel System – Hand *yangming* (*shou yang ming jing luo*)

4.2.1 The Large Intestine Primary Channel (*shou yang ming jing*)

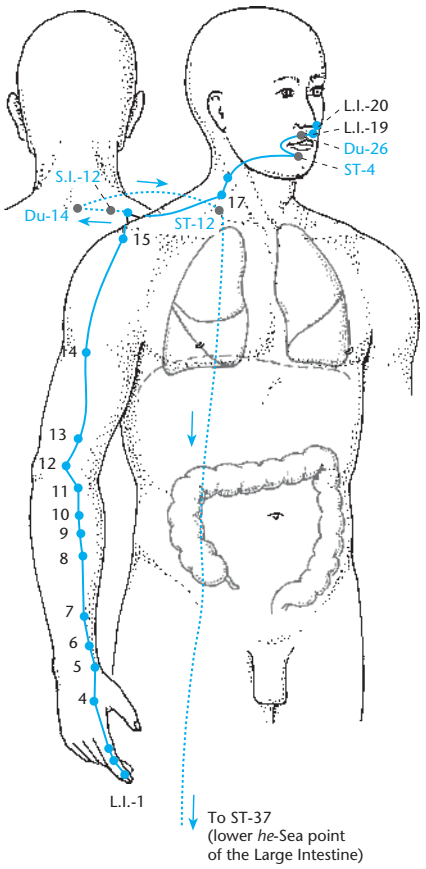

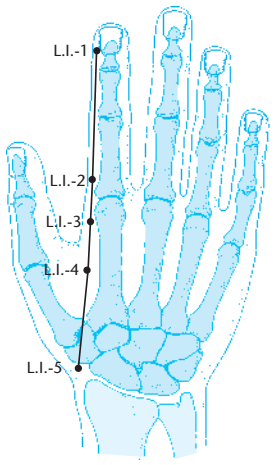

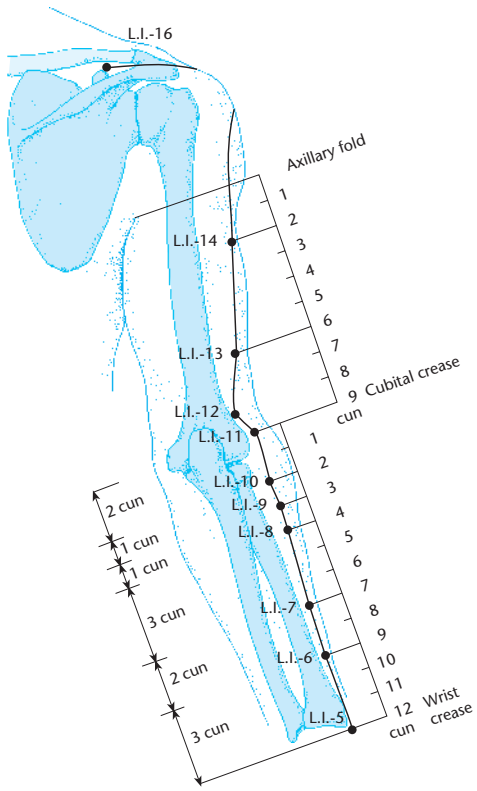

Pathway

The **external** part of the Large Intestine primary channel begins on the index finger at ➞ **L.I.-1** (*shangyang*, at the radial aspect of the corner of the nail). This point is reached by a branch which separates from the Lung primary channel at ➞ **LU-7** (*lieque*; hand Yin–Yang connection of the first great circuit).

The Large Intestine primary channel runs between the first and second metacarpal bones,

- then runs proximally along the radial aspect of the forearm towards the lateral aspect of the elbow
- travels in a proximal direction along the lateral aspect of the upper arm towards the shoulder
- crosses the Small Intestine primary channel at ➞ **S.I.-12** (*bingfeng*) in the centre of the suprascapular fossa
- continues to ➞ **Du-14** (*dazhui*) inferior to the spinous process of the 7th cervical vertebra, where it meets with the other five Yang primary channels
- crosses the neck and enters the supraclavicular fossa at ➞ **ST-12** (*quepen*).

Here the **internal** pathway branches off and connects with the Lung (*fei*), its paired *zang*-Organ, penetrates the diaphragm and enters the Large Intestine (*dachang*), its pertaining *fu*-Organ. From the Large Intestine (according to some authors: the Large Intestine primary channel) an **internal** branch runs in an inferior direction to ➞ **ST-37** (*shangjuxu*) on the lateral aspect of the lower leg, the lower *he*-sea point of the Large Intestine.

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 96

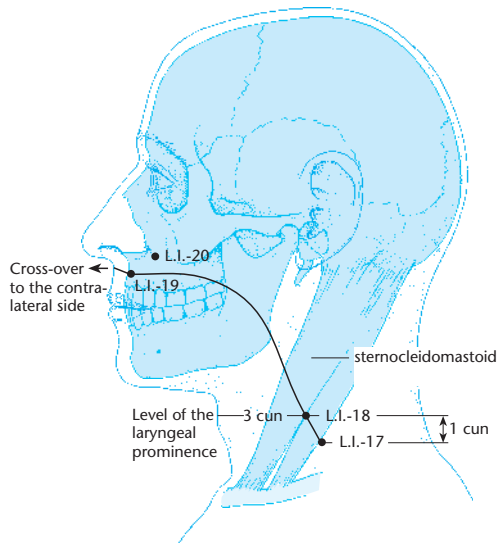

The **external** branch ascends from the supraclavicular fossa, travels along the lateral aspect of the neck and traverses the cheek.

- ➡ From here another **internal** branch enters the lower gums, where it extends further.
- ➡ The **external** course of the primary channel passes → **ST-4** (*dicang*) and curves around the mouth to the upper lip, crossing to the contralateral side of the body at → **Du-26** (*renzhong*), located at the philtrum. On the contralateral side of the body, it reaches → **L.I.-19** (*kouheliao*) and terminates **contralaterally** at → **L.I.-20** (*yingxiang*) on the side of the nose.

Note: Deadman et al. (1998) mention → **Ren-24** (*chengjiang*) as additional meeting point (variant → figure). They also mention → **G.B.-5** (*xuanlu*), → **G.B.-6** (*xuanli*) and → **G.B.-14** (*yangbai*) as classic meeting points with the Large Intestine channel but these are not normally shown in illustrations of the channel.

Clinical importance (→ 1.2)

Exterior (*biao*) signs and symptoms: Fever, dryness of the mouth, thirst, a sore and swollen throat, nosebleed, toothache, red, painful eyes, stiff fingers, painful cold or painful hot swellings of the upper arm and shoulder

Interior (*li*) or *zangfu*-Organ signs and symptoms: Pain in the lower abdomen, borborygmus, diarrhoea or yellow stools with mucus, shortness of breath, dyspnoea

Excess (*shi*): Heat sensations along the channel

Deficiency (*xu*): Cold sensations and chills along the channel

Connections of the Large Intestine primary channel (→ 1.2)

With other channels

Lung primary channel (*shou tai yin jing*)

Connection: Hand Yin–Yang connection of the first great circuit

Location: LU-7 → L.I.-1 (on the hand). A branch of the Lung primary channel connects with the Large Intestine primary channel at **L.I.-1** (*shangyang*)

Circulation: Circadian (according to the Organ clock)

Importance: Exterior–Interior relationship

Stomach primary channel (*zu yang ming jing*)

Connection: Paired according to the six channel theory (hand–foot pairing): *yangming* (Yang axes of the first great circuit)

Location: L.I.-20 → (BL-1) → ST-1 (on the head)

Circulation: Circadian (according to the Organ clock)

Importance: Above–below relationship

With the *zangfu*-Organ systems

Lung (*fei*), **Large Intestine (*dachang*)**, Stomach (*wei*)

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 97

4.2.2 The Large Intestine Divergent Channel (*shou yang ming jing bie*)

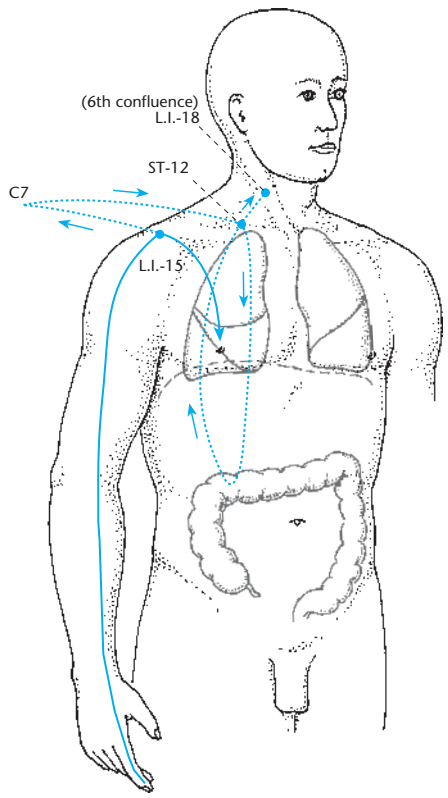

Pathway

The Large Intestine divergent channel separates from the Large Intestine primary channel on the hand,

- travels along the arm to the shoulder to → **L.I.-15** (*jianyu*)
- from **L.I.-15** a **branch** runs to the thorax and branches out into the breasts
- **another branch** runs to the cervical spine at the 7th cervical vertebra, returns from there to the supraclavicular fossa, descends to the Large Intestine (*dachang*), ascends to the Lung (*fei*) and **emerges** to the Exterior at → **ST-12** (*quepen*).*

It then travels along the anterolateral aspect of the neck and connects with the Large Intestine primary channel and the Lung divergent channel at → **L.I.-18** (*futu*) at one of the six *he*-confluences† (here: L.I./LU as 6th confluence → 1.3).

Clinical importance

- Strengthens the connection between the Lung and the Large Intestine (*zangfu*-Organs). Points on the Large Intestine primary channel can therefore be used for disorders of the Lung, and vice versa points on the Lung primary channel can treat disorders of the Large Intestine.
- A branch of the Large Intestine divergent channel spreads (from → **L.I.-15**) across the thorax and the breasts, thus extending the actions of the points on the Large Intestine primary channel to these regions, for example for the treatment of mastitis, myalgias and thoracic pain.

---

* Some authors describe the divergent channel as travelling directly from the spine to the Large Intestine, without meeting → **ST-12** (variant → Fig.).

† According to some authors the 6th *he*-Confluence is located at → **ST-12** (*quepen*).

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 98

4.2.3 The Large Intestine Sinew Channel (*shou yang ming jing jin*)

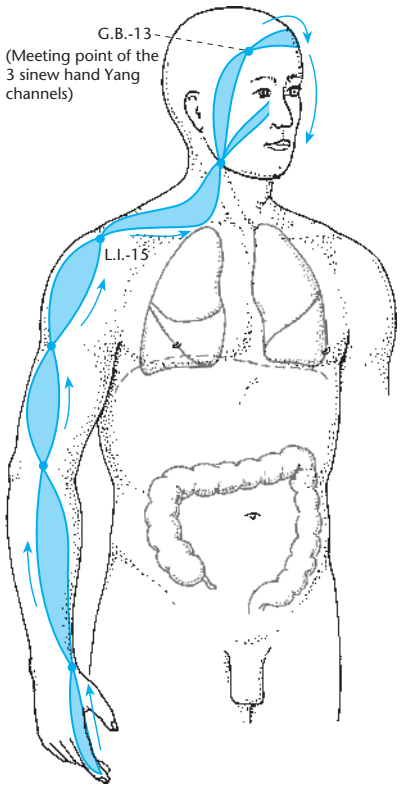

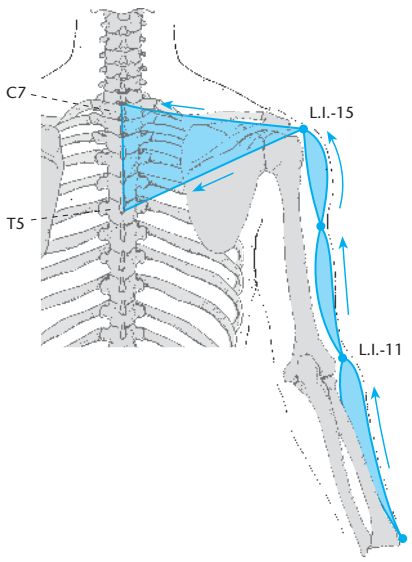

Pathway

The Large Intestine sinew channel

- begins on the index finger at → **L.I.-1** (*shangyang*)
- runs along the index finger and the 2nd metacarpal bone towards the wrist, where it binds (*jie*) at → **L.I.-5** (*yangxi*)
- travels along the posterolateral aspect of the forearm
- reaches the elbow and binds (*jie*) at → **L.I.-11** (*quchi*)
- continues along the lateral aspect of the upper arm and the deltoid muscle to the shoulder, where it binds (*jie*) again at → **L.I.-15** (*jianyu*)
- from **L.I.-15** a branch spreads across the scapula and then attaches to the spine between the 7th cervical vertebra and the 5th thoracic vertebra (or between → **Du-14** (*dazhui*) and → **Du-11** (*shendao*).
- The **main branch** ascends from **L.I.-15**, crosses the supraclavicular fossa and continues from there along the lateral aspect of the neck towards the angle of the jaw where it binds (*jie*).

Here it forks into two branches:

- **One branch** runs across the zygomatic bone towards the lateral side of the nose
- **The other branch** ascends the lateral aspect of the face and anterior to the Small Intestine sinew channel, connects at → **G.B.-13** (*benshen*) with the other sinew Yang channels, crosses over the head and ends at the lower jaw on the opposite side.

Note: According to some schools of thought, the sinew channel only ends at → **L.I.-17** (*tianding*; Solinas et al. 1998).

Clinical importance

Pathology: cramps, pain, tension and stiffness along the Large Intestine sinew channel. Limited range of motion of the upper limbs and shoulders. Occipital pain, stiffness and limited range of motion.

Indication: Mainly for *bi*-syndrome (painful obstruction syndrome) along the Large Intestine channel. The area covered by the Large Intestine sinew channel is larger than that covered by the Large Intestine primary channel. This explains why the indications of points on the Large Intestine primary channel include disorders and diseases of the upper neck and thoracic region as well as of the whole face (for example frontal and temporal headaches).

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 99

4.2.4 The Large Intestine *luo*-Connecting Vessel System (*shou yang ming luo mai*)

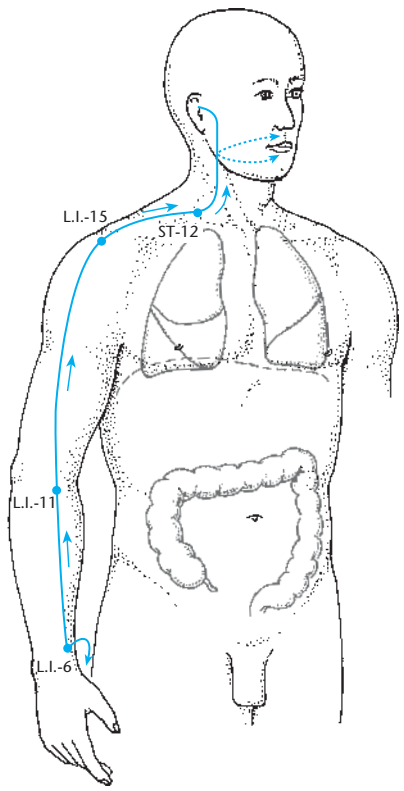

Pathway

The **Large Intestine** *luo*-connecting channel system begins at the *luo*-connecting point → **L.I.-6** (*pianli*) (→ 8.1.2), where it branches off from the primary channel and forms a three-dimensional reticular network, dividing into numerous branches and sub-branches (*sun luo*, *fu luo*, *xue luo* → 1.5) within the surrounding tissue.

➡ Horizontal divisions course to the Interiorly–Exteriorly paired Lung primary channel; according to some schools (for example, Van Nghi → Appendix) they travel as a **transverse** LU *luo*-connecting vessel to the *yuan*-source point → **LU-9** (*taiyuan*).

➡ A **longitudinally** orientated division follows the Large Intestine primary channel to **L.I.-15** (*jianyu*), continues via the supraclavicular fossa to → **ST-12** (*quepen*) and along the neck to the mandibular angle, where it spreads. One branch runs to the teeth; another branch reaches the ear, where it communicates with all channels reaching the ear.

Clinical importance (→ 8.1.2)

Pathology

- **Excess (*shi*):** Disorders of the teeth and gums, disorders of the ears
- **Deficiency (*xu*):** Sensitive teeth and gums (such as sensitivity to cold), sensation of tightness in the chest and diaphragm

4.2.5 Cutaneous Region (*yang ming pi bu*)

See description and figures → 1.6

4.2.6 Points of the Large Intestine Primary Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): L.I.-4** (*hegu*) ■■
- ***Luo*-Connecting point (→ 8.1.2.): L.I.-6** (*pianli*) ■■
- ***Xi*-cleft point (→ 8.1.3): L.I.-7** (*wenliu*)
- **Associated Back-*shu* point (→ 8.1.4): BL-25** (*dachangshu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): ST-25** (*tianshu*)
- **Five *shu*-transport point (→ 8.1.6):**
  jing-well point (Metal), *ben* point:. **L.-I.1** (*shangyang*)
  ying-spring point (Water), sedation point: **L.I.-2** (*erjian*)
  shu-stream point (Wood): **L.I.-3** (*sanjian*)
  jing-river point (Fire): **L.I.-5** (*yangxi*)
  he-sea point (Earth), tonification point: **L.I.-11** (*quchi*) ■■
- ***Hui*-meeting point (→ 8.1.7): –**
- **Opening point (→ 8.1.8): –**
- **Lower *he*-sea point (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the Small Intestine primary and BL channel, *yang wei mai*: **L.I.-14** (*binao*)*
  – with the *yang qiao mai*: **L.I.-15** (*jianyu*) ■■, **L.I.-16** (*jugu*)
  – with the primary ST channel: **L.I.-20** (*yingjiang*) ■■
  – of other channels with the L.I. channel: **ST-4, ST-12, S.I.-12, Du-14, Du-26, Ren-24***; according to some classics: **G.B.-5, G.B.-6, G.B.-14**
- **Gao Wu command point (→ 8.1.11) for the face and mouth: L.I.-4** (*hegu*) ■■
- **Window of Heaven point (→ 8.1.12): L.I.-18** (*futu*)
- **Point of the Four Seas (→ 8.1.13): –**
- **Ma Dan Yang Heavenly Star points (→ 8.1.14): L.I.-4** (*hegu*) ■■, **L.I.-11** (*quchi*) ■■
- **Sun Si Miao Ghost point (→ 8.1.15): L.I.-11** (*quchi*) ■■

* Mentioned by only some authors

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 100

Points according to region

- **Local points (→ 8.2.1):** nose – **L.I.-20** (*yingxiang*) ■■; shoulder – **L.I.-15** (*jianyu*) ■■; elbow – **L.I.-11** (*quchi*) ■■; hand – **L.I.-5** (*yangxi*); hand and fingers – **L.I.-4** (*hegu*) ■■; fingers (numbness and pain) – **L.I.-3** (*sanjian*)
- **Regional points (→ 8.2.1):** elbow – **L.I.-10** (*shousanli*), **L.I.-13** (*shouwuli*); shoulder and elbow – **L.I.-14** (*binao*)
- **Distal points (→ 8.2.1):** forehead, eyes, mouth/teeth – **L.I.-4** (*hegu*) ■■; nose – **L.I.-3** (*sanjian*), **L.I.-4** (*hegu*) ■■; neck – **L.I.-4** (*hegu*) ■■, **L.I.-11** (*quchi*) ■■; mouth/teeth – **L.I.-2** (*erjian*), **L.I.-4** (*hegu*) ■■; shoulder, elbow, hand, fingers – **L.I.-4** (*hegu*) ■■, **L.I.-1** (*shangyang*); knee and foot – **L.I.-5** (*yangxi*); foot – **L.I.-2** (*erjian*), **L.I.-3** (*sanjian*); toes – **L.I.-4** (*hegu*) ■■

Specific points according to the course of the channel (in numerical order)

- **L.I.-1** (*shangyang*): *jing*-well point (Metal) (→ 8.1.6); *ben* point (Five Phase point); distal point for the elbow and the hand (→ 8.2.1)
- **L.I.-2** (*erjian*): *ying*-spring point (Water) (→ 8.1.6); sedation point; distal point for the mouth/teeth (→ 8.2.1); distal point for the foot (→ 8.2.1)
- **L.I.-3** (*sanjian*): *shu*-stream point (Wood) (→ 8.1.6); local point for the fingers (numbness and pain) (→ 8.2.1); distal point for the nose (→ 8.2.1); distal point for the feet (→ 8.2.1)
- **L.I.-4** (*hegu*) ■■: *yuan*-source point (→ 8.1.1); Ma Dan Yang Heavenly Star point (→ 8.1.14); Gao Wu command point (→ 8.1.11) for the face and mouth; important point for acupuncture anaesthesia; distal point for the forehead, eyes, nose, mouth/teeth, jaw, shoulder, elbow, hand, fingers (→ 8.2.1); local point for the hand and fingers (stiffness) (→ 8.2.1)
- **L.I.-5** (*yangxi*): *shu*-stream point (Fire) (→ 8.1.6); local point for the hand (→ 8.2.1); distal point for the knees and feet (→ 8.2.1)
- **L.I.-6** (*pianli*): *luo*-connecting point (→ 8.1.2)
- **L.I.-7** (*wenliu*): *xi*-cleft point (→ 8.1.3)
- **L.I.-10** (*shousanli*): adjacent point for the elbow (→ 8.2.1)
- **L.I.-11** (*qichi*) ■■: Sun Si Miao Ghost point (→ 8.1.15); *he*-sea point (Earth) (→ 8.1.6); tonification point; Ma Dan Yang Heavenly Star point; distal point for the neck (→ 8.2.1); local point for the elbow (→ 8.2.1)
- **L.I.-13** (*shouwuli*): regional point for the elbow (→ 8.2.1)
- **L.I.-14** (*binao*)*: *jiaohui*-meeting point with the S.I. and BL channel and the *yang wei mai** (→ 8.1.10); regional point for the shoulder and elbow (→ 8.2.1)
- **L.I.-15** (*jianyu*) ■■: *jiaohui*-meeting point with the *yang qiao mai* (→ 8.1.10); local point for the shoulder (→ 8.2.1)
- **L.I.-16** (*jugu*): *jiaohui*-meeting point with the *yang qiao mai* (→ 8.1.10)
- **L.I.-18** (*futu*): Window of Heaven point (→ 8.1.12)
- **L.I.-20** (*yingxiang*) ■■: local point for disorders of the nose (→ 8.2.1); *jiaohui*-meeting point with the ST channel (→ 8.1.10)

General location help

- The course of the line connecting **L.I.-5** and **L.I.-11** (12 cun) changes depending on whether the forearm is in a supinated or a pronated position. For location purposes the forearm should be in a midposition with the elbow slightly flexed.
- Since the L.I. channel crosses to the opposite side of the body at **Du-26** (nasolabial groove), locate **L.I.-19** and **L.I.-20** contralaterally.

* Mentioned by only a few authors

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 101

Shang Yang SHANGYANG L.I.-1

Location
On the index finger, 0.1 cun from the radial corner of the nail.

How to find
This point is located where two tangents bordering the nail of the index finger medially and proximally cross, approximately 0.1 cun from the corner of the nail.

Needling
0.2–0.3 cun vertically or obliquely in a proximal direction. Avoid needling the perionychium. For acute disorders (pain, inflammation), prick to bleed with needle, lancet or three-edged needle.

Actions/Indications
- Expels pathogenic factors and opens the channel
- Frees the senses

Special features
Jing-well point, Metal point, *ben* point (Five Phases), entry point according to some authors.

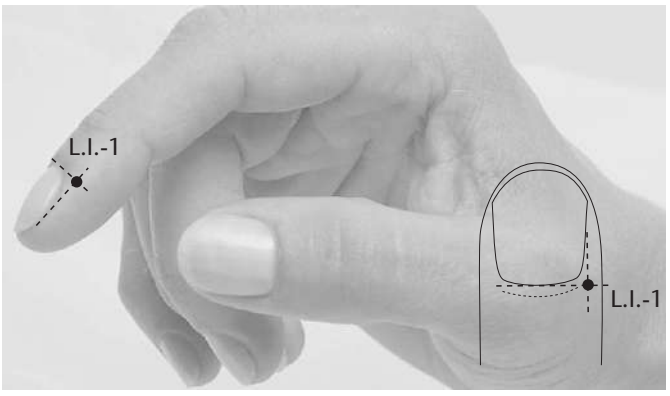

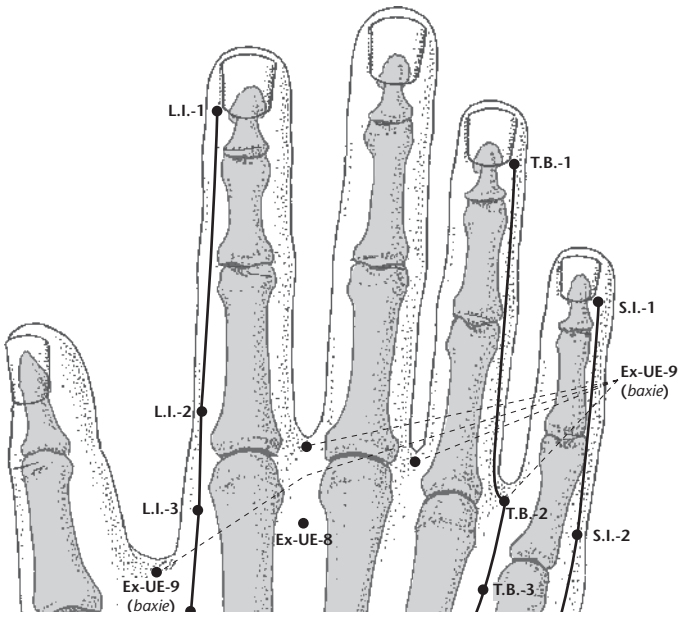

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 102

L.I.-2 *Second Space* ERJIAN

Location

On the radial aspect of the index finger, distal to the metacarpo-phalangeal joint, at the junction of the shaft and the basis of the proximal phalanx.

How to find

With the hand relaxed, palpate the border between the 'red to white' skin along the shaft of the second proximal phalangeal bone towards the metacarpo-phalangeal joint until the angle between the base and the shaft of the bone can be clearly felt. **L.I.-2** is located at the junction of the shaft and the base and slightly below (palmar to) the most external curvature of the bone.

→ **S.I.-2** is located in a similar position at the proximal phalanx of the little finger.

Needling

0.2–0.5 cun obliquely in a proximal direction or distal in a slightly palmar direction.

Actions/Indications

- Clears Heat and expels Wind

Special features

Ying-spring point, Water point, sedation point, distal point for the mouth and teeth

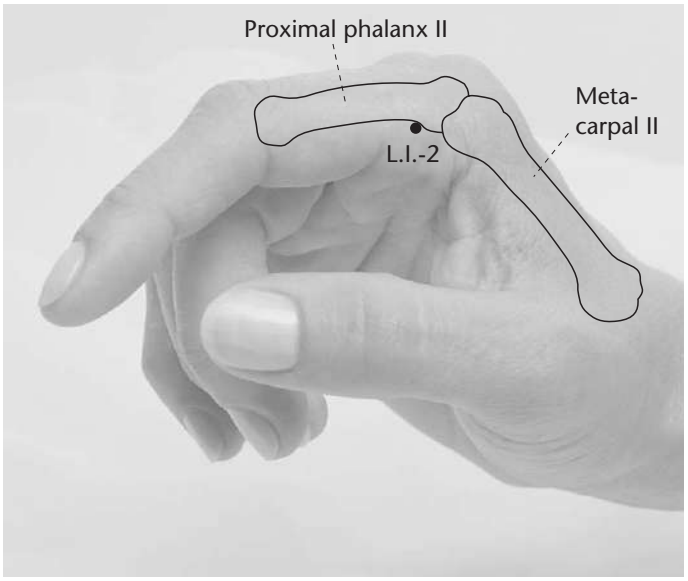

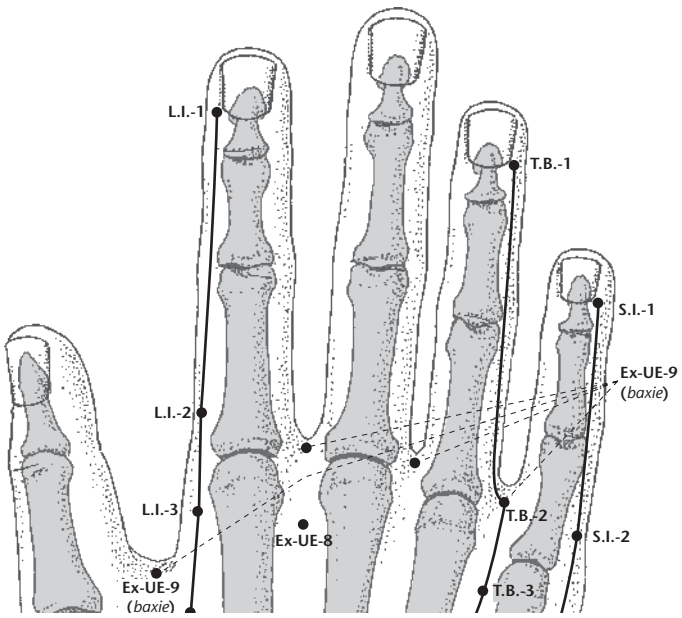

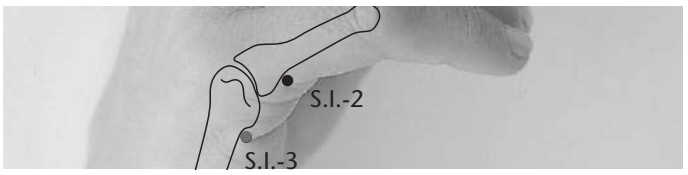

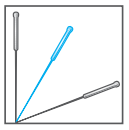
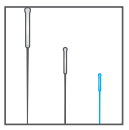

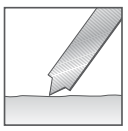

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 103

Third Space SANJIAN L.I.-3

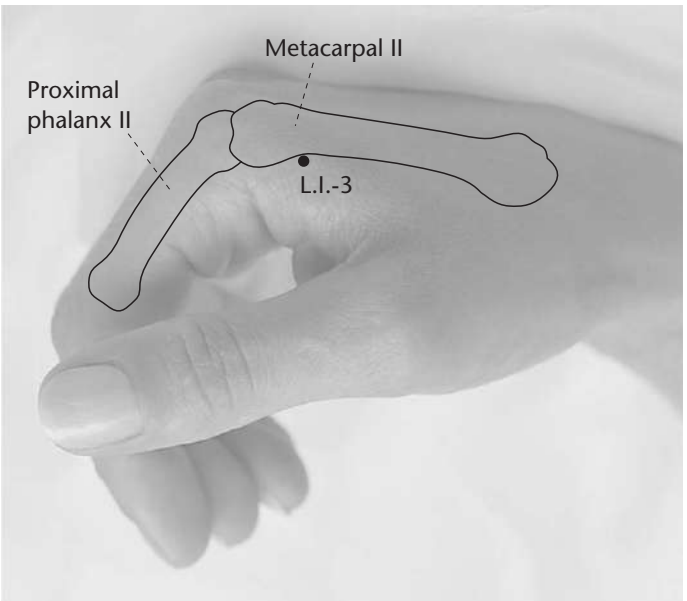

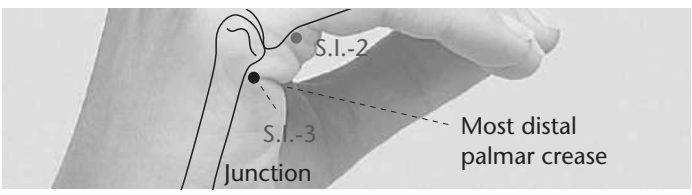

Location

On the radial aspect of the index finger, proximal to the metacarpo-phalangeal joint, at the junction of the shaft and the head of the 2nd metacarpal bone.

How to find

With the hand relaxed, palpate along the radial aspect of the 2nd metacarpal bone in a distal direction until the angle formed between the head and the shaft can be clearly felt. **L.I-3** is located at the junction of the shaft and the head and slightly below (palmar to) its most external curvature. → **S.I.-3** is located in a comparable position on the ulnar aspect of the 5th metacarpal bone (on the ulnar border of the hand). Located in a comparable location on the feet are → **SP-3** and → **BL-65**.

Needling

With the hand relaxed, 0.3–1 cun vertically directly below the lower border of the metacarpal bone towards → **S.I.-3**.

Actions/Indications

- Clears Heat and expels Wind
- Benefits the throat and teeth
- Alleviates pain

Special features

Shu-stream point, Wood point. Important point for pain in the fingers and the metacarpus.

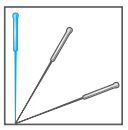

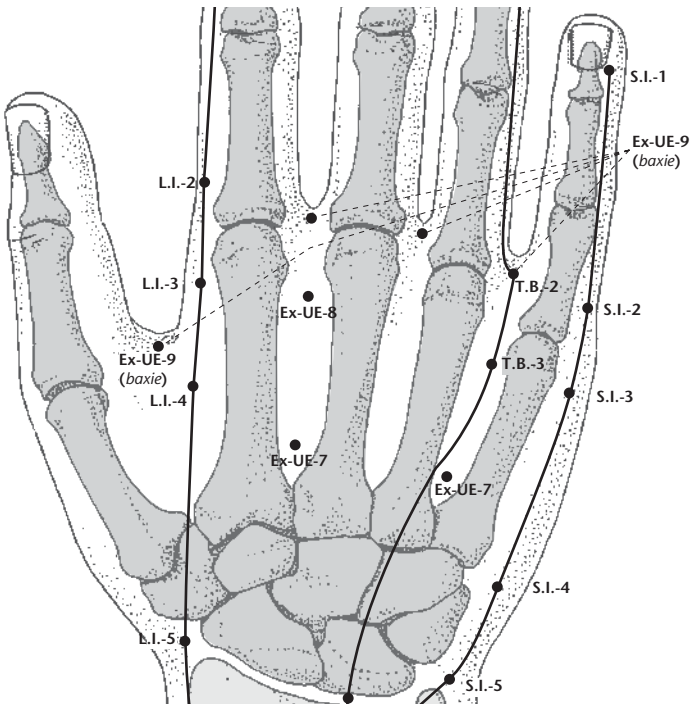

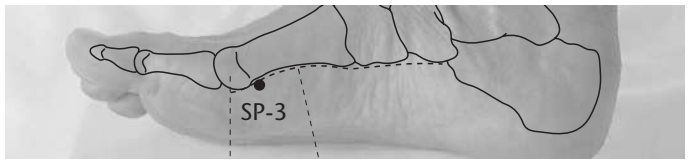

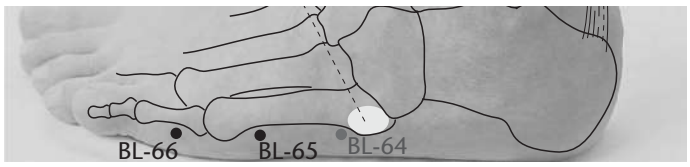

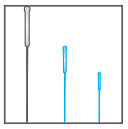

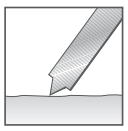

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 104

L.I.-4 *Joining Valley* HEGU

Location

On the radial aspect of the hand, between the 1st and 2nd metacarpal bones, closer to the 2nd metacarpal bone and approximately at its midpoint.

How to find

When pressing the thumb and index finger together, the adductor pollicis and the interosseus dorsalis muscles will form a bulge. Needle **L.I.-4** at the highest point of the bulge and push it further towards the inferior aspect of the second metacarpal bone. **Or**: Spread the thumb and forefinger and place the distal phalanx of the thumb of the other hand on the webbed border. When flexing the thumb, its tip will point to **L.I.-4. Or**: Insertion on the angle bisector between the 1st and 2nd metacarpal bones, needling in the direction of the midpoint of the second metacarpal bone.

Needling

0.5–1 cun vertically or slightly obliquely towards the palmar aspect of the hand. Caution: Reducing needle techniques are contraindicated during pregnancy; exception: to induce labour.

Actions/Indications

- Releases the Exterior (main point!)
- Regulates the face and head
- Regulates the Defensive Qi (*wei qi*) and sweating
- Opens the channel and *luo*-connecting vessels
- Alleviates pain
- Promotes labour

Special features

Yuan-source point, Gao Wu command point for the head and mouth, Ma Dan Yang Heavenly Star point, entry point according to some authors. Most important analgesic point for the whole body. Most important single distal point for disorders of the face and sensory organs. It is the most commonly used point in clinical practice. The combined bilateral needling of **L.I.-4** and **→ LIV-3** (between the 2nd and 3rd metatarsal bones) is known as *si guan* (Four Gates): they strongly regulate the Qi and Blood, stop pain and relieve spasms.

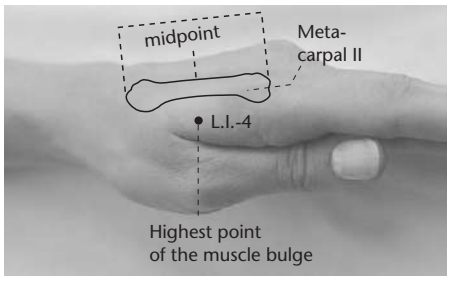

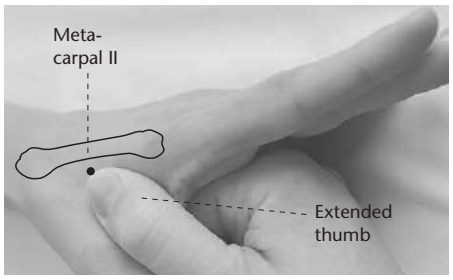

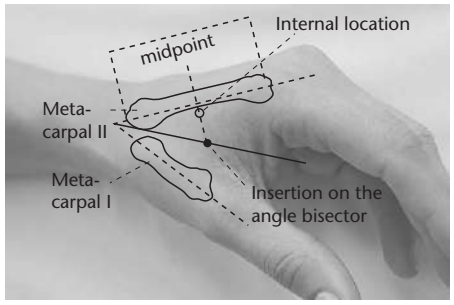

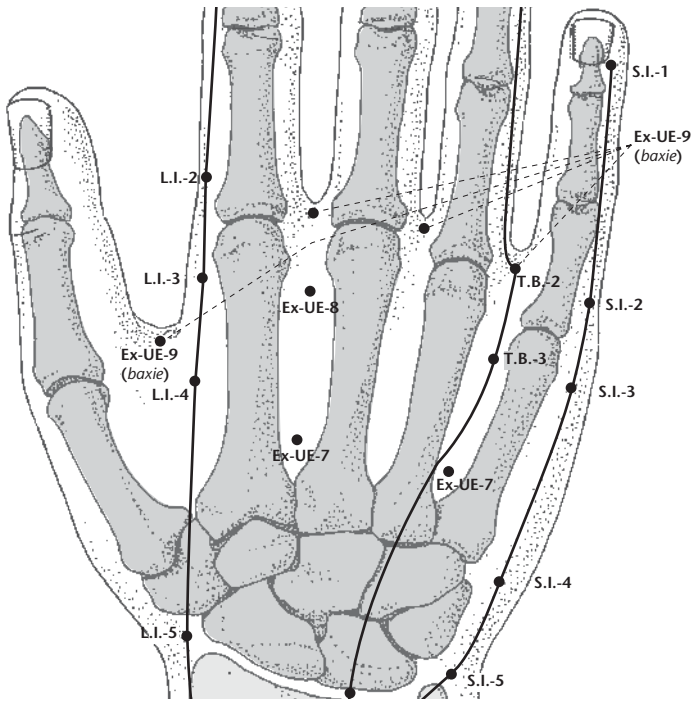

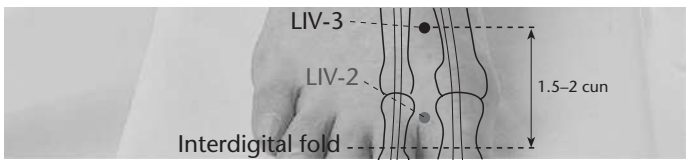

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 105

Yang Stream YANGXI L.I.-5

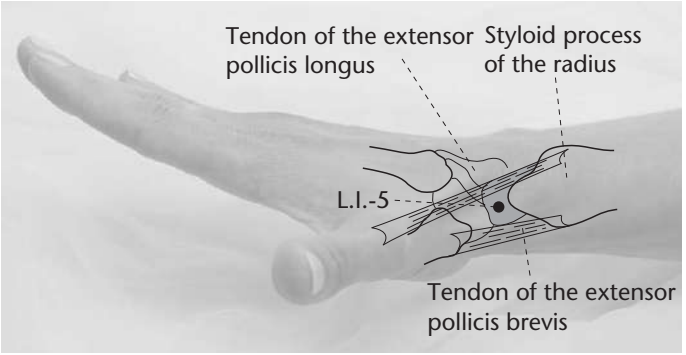

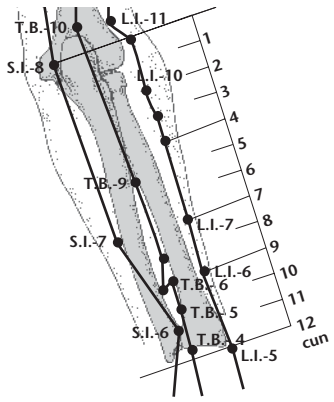

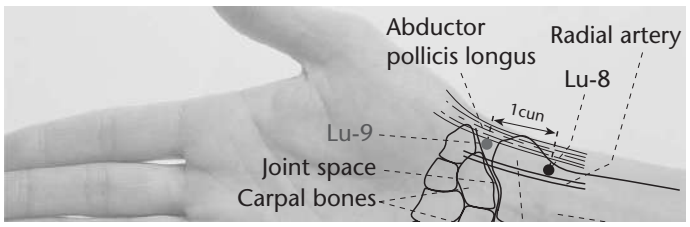

Location

With the thumb abducted, in a depression between the tendons of the extensor pollicis longus and brevis muscles ('anatomical snuffbox'), on the radial aspect of the wrist.

How to find

The anatomical snuffbox is a depression on the radial aspect of the wrist formed when abducting the thumb. Holding the hand in a horizontal position, it opens towards the body. **L.I.-5** is located on the wrist joint space, which becomes palpable by moving the wrist. → **LU-9** is also located on the wrist joint space and can be found by moving anteriorly from **L.I.-5** across the short tendon of the extensor pollicis muscle.

Needling

Vertically 0.5–1 cun. Caution: Avoid the superficial cephalic vein.

Actions/Indications

- Supports the wrist joint
- Clears Fire and expels Wind
- Clears Fire from the *yangming* (L.I., ST) and calms the *shen*

Special features

Jing-river point, Fire point. Important local point for disorders of the wrist, together with **T.B.-4, S.I.-5, HE-7, P-7** and **LU-9**.

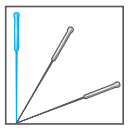
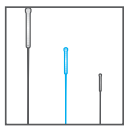

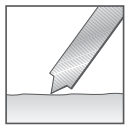

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 106

L.I.-6 *Veering Passage* PIANLI

Location

3 cun proximal to → **L.I.-5** (in the centre of the anatomical snuffbox), on the line connecting → **L.I.-5** and → **L.I.-11**, between the abductor pollicis longus and the extensor pollicis brevis muscles, at the level of the junction between the tendon and the muscle.

How to find

Please note: The course of the line connecting → **L.I.-5** and → **L.I.-11** depends on the position of the forearm. For location purposes, a midposition of the forearm with flexed elbow is recommended. To locate **L.I.-6**, measure 3 cun (1 handsbreath) from → **L.I.-5** (centre of the anatomical snuffbox) on the connecting line. **Or**: 'Tiger mouth grip' (→ Fig. 2.6): Spread thumb and index finger of both hands, crossing them so that the forefinger of one hand comes to rest on the styloid process of the other, but avoiding bending the wrist. Place the middle finger against the forefinger: **L.I.-6** is located directly below the tip of the middle finger.

Located at the same level (3 cun proximal to the wrist joint space) on the dorsal side of the forearm are → **T.B.-6** (in the depression between the radius and the muscle) and → **T.B.-7** (in the depression between the ulna and the muscle) as well as → **P-5** (on the ventral side of the forearm between the tendons).

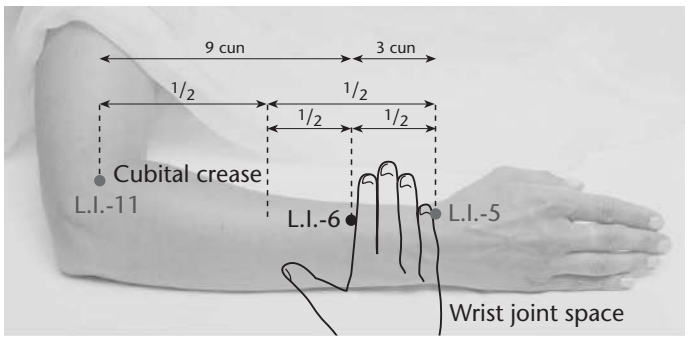

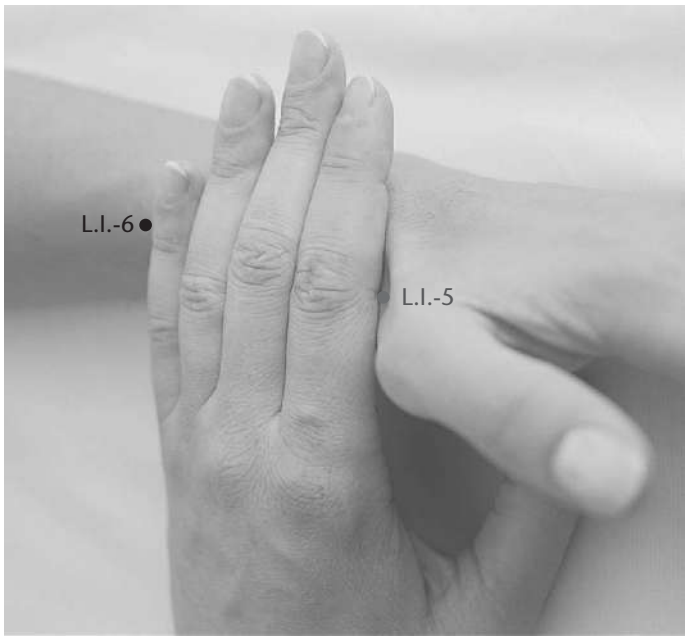

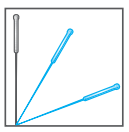

Needling

Obliquely or transversely (subcutaneously) 0.5–1 cun

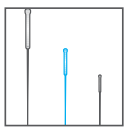

Actions/Indications

- Expels Wind, clears Heat
- Regulates and opens the water passages
- Opens the channel and *luo* vessels

Special features

Luo-connecting point. Major point of the L.I. channel for regulating the water passages.

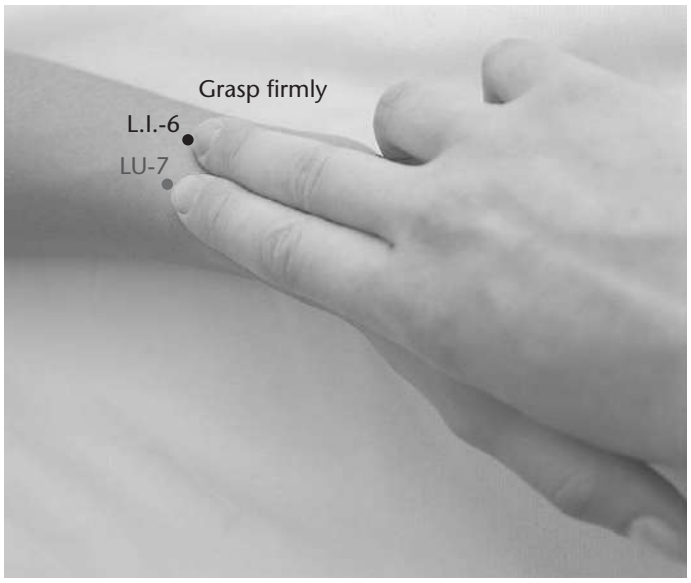

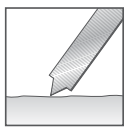

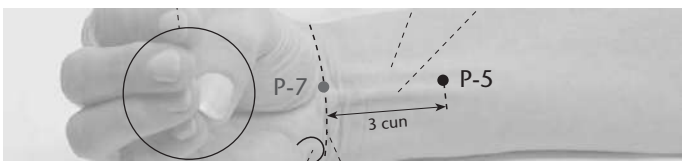

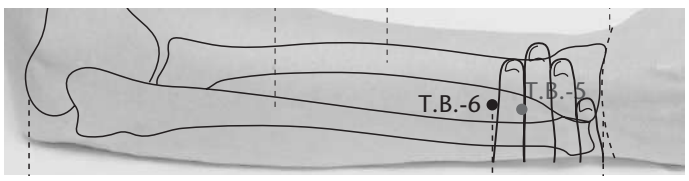

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 107

Warm Flow WENLIU L.I.-7

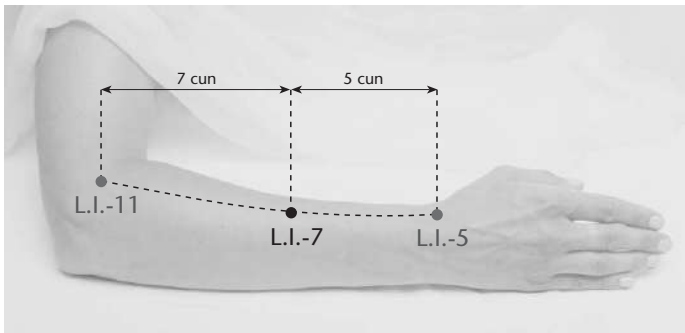

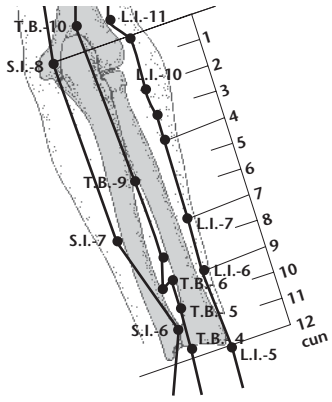

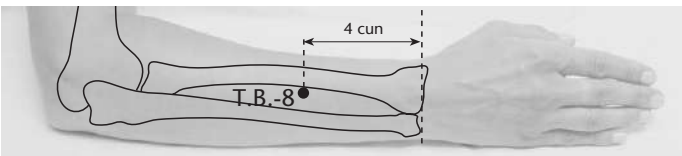

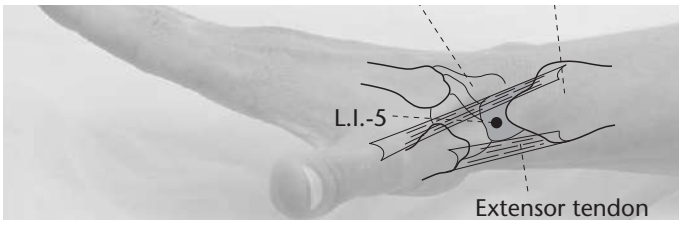

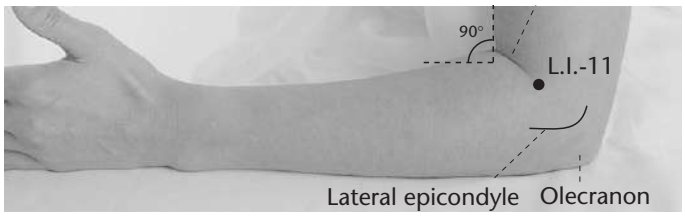

Location

5 cun proximal to the anatomical snuffbox in the direction of the lateral end of the elbow crease or 1 cun distal to the midpoint of the line connecting → **L.I.-5** and → **L.I.-11**.

How to find

Please note: With the hand supinated, the line connecting → **L.I.-5** and → **L.I.-11** runs alongside the radial margin of the forearm. However, with the arm pronated, it transverses the forearm. **L.I.-7** lies on the dorso-lateral aspect of the radius and is best located with the forearm in midposition and the elbow flexed. Spreading hands technique: Place the little fingers on → **L.I.-5** and → **L.I.-11** and span the hands evenly so that the thumbs join at the midpoint of the line. From there, measure 1 cun distally. → **T.B.-8** is located 4 cun proximal to the wrist crease between the radius and the ulna in the centre of the dorsal forearm or 1 cun distal to **L.I.-7**.

Needling

Vertically 0.5–1 cun

Actions/Indications

- Treats acute disorders
- Clears Heat and toxic Heat
- Regulates the Stomach and Large Intestine

Special features

Xi-cleft point

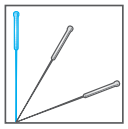
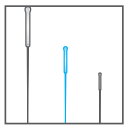

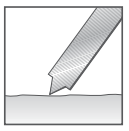

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 108

L.I.-8 *Lower Angle* XIALIAN

Location

4 cun distal to the lateral end of the elbow crease in the direction of the anatomical snuffbox and on a line connecting → **L.I.-5** and → **L.I.-11**.

How to find

Please note: In supination, the line connecting → **L.I.-5** and → **L.I.-11** runs along the radial margin of the forearm; in pronation, it transverses the forearm. **L.I.-8** lies on the dorsal aspect of the radius and is best located with the forearm in mid-position and the elbow flexed. **L.I.-8** can be found either 4 cun distal to → **L.I.-11** or 2 cun proximal to the midpoint of the line connecting → **L.I.-5** and → **L.I.-11** (spreading hands technique, → 2.3.3).

→ **T.B.-9** is 5 cun distal to the olecranon between the radius and the ulna, about 1 cun distal to **L.I.-8** in the middle of the dorsal forearm.

Needling

Vertically or obliquely 0.5–1 cun

Actions/Indications:

- Opens the channel
- Cools and expels Wind
- Harmonises the Small Intestine

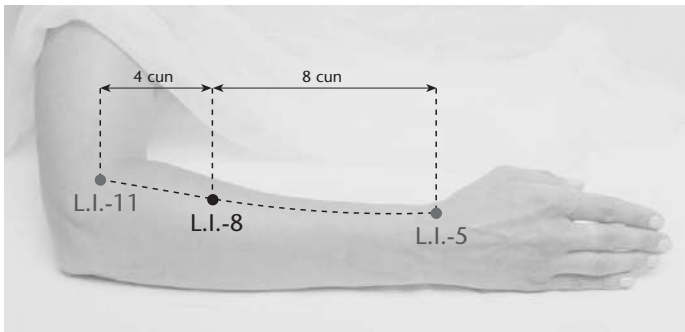

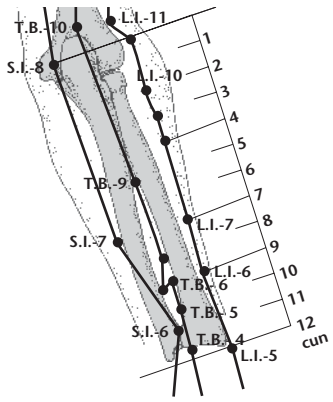

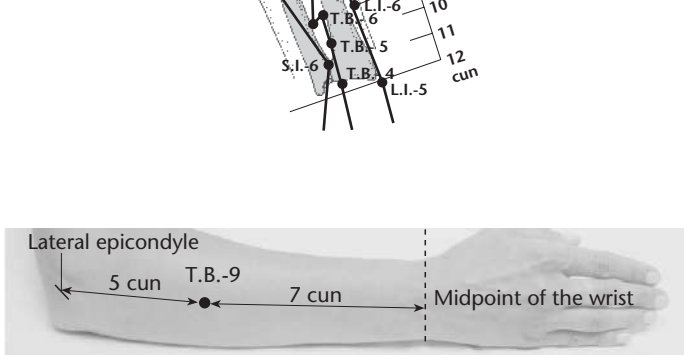

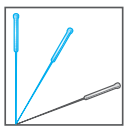
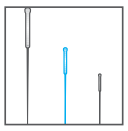

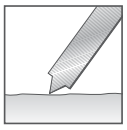

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 109

Upper Angle SHANGLIAN L.I.-9

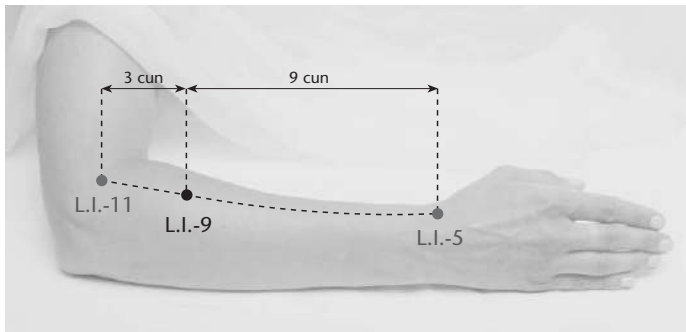

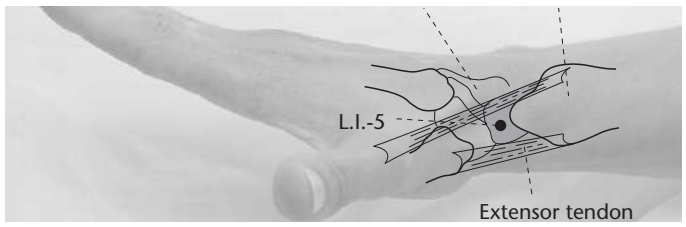

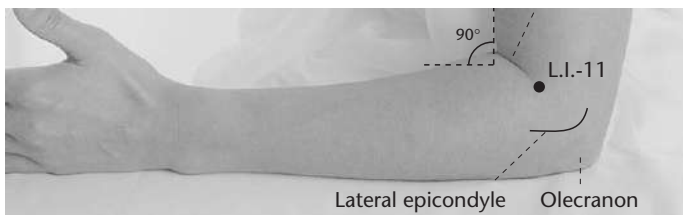

Location

3 cun distal to the lateral end of the elbow crease in the direction of the anatomical snuffbox, on the line connecting → **L.I.-5** and → **L.I.-11**.

How to find

Please note: With the hand supinated, the line connecting → **L.I.-5** and → **L.I.-11** runs along the radial margin of the forearm. However, with the arm pronated, it transverses the forearm. **L.I.-9** lies on the dorsal aspect of the radius and is best located with the forearm in midposition and the elbow flexed. From → **L.I.-11**, measure 3 cun distally. The point is located in an easily palpable depression between two muscle bellies and is often tender with pressure.

Needling

Vertically or obliquely 0.5–1 cun

Actions/Indications

- Opens the channel, cools and expels Wind
- Harmonises the Large Intestine

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 110

L.I.-10 *Arm Three Miles* SHOUSANLI

Location

2 cun distal to → **L.I.-11**, on the line connecting → **L.I.-5** and → **L.I.-11**, on the extensor carpi radialis longus muscle; a deeper insertion will reach the supinator muscle.

How to find

Please note: With the hand supinated, the line connecting → **L.I.-5** and → **L.I.-11** runs along the radial margin of the forearm. However, with the arm pronated, it transverses the forearm. **L.I.-10** lies on the dorsal aspect of the radius and is best located with the forearm in midposition and with the elbow flexed. First, palpate → **L.I.-11** in the depression lateral to the radial end of the cubital crease. **L.I.-10**, usually sensitive to pressure, is found by measuring 2 cun distally from → **L.I.-11** on the connecting line.

Needling

Vertically 1–2 cun

Actions/Indications

- Regulates the Qi and Blood (in the upper extremity), opens the channel and *luo* vessels, alleviates pain
- Regulates the Stomach and Intestines

Special features

Important local point, often used in chain-and-lock method with other channel points

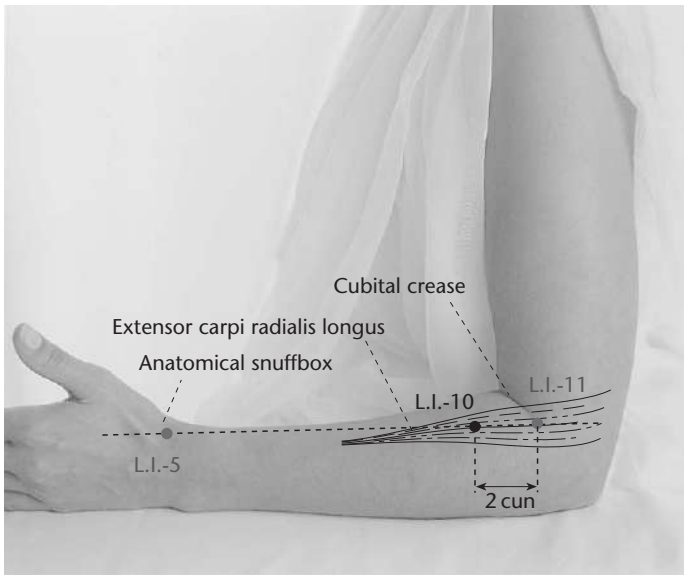

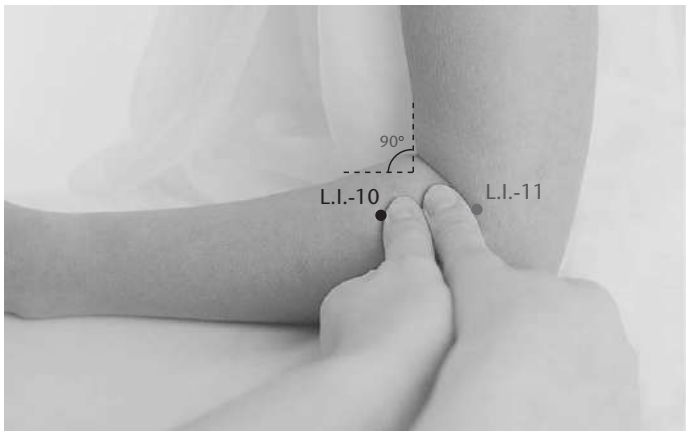

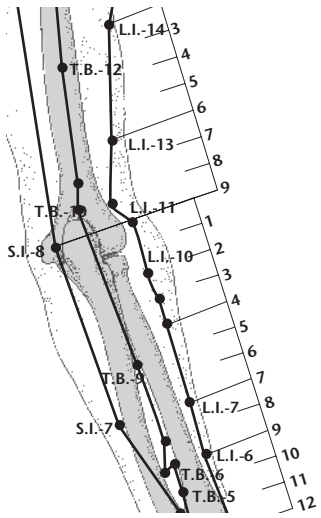

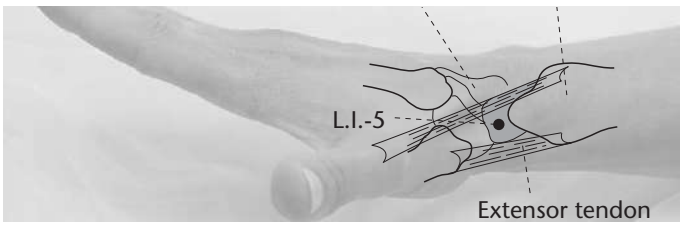

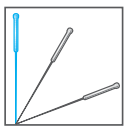

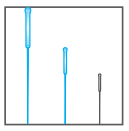

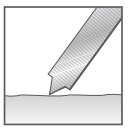

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 111

Pool at the Crook QUCHI L.I.-11

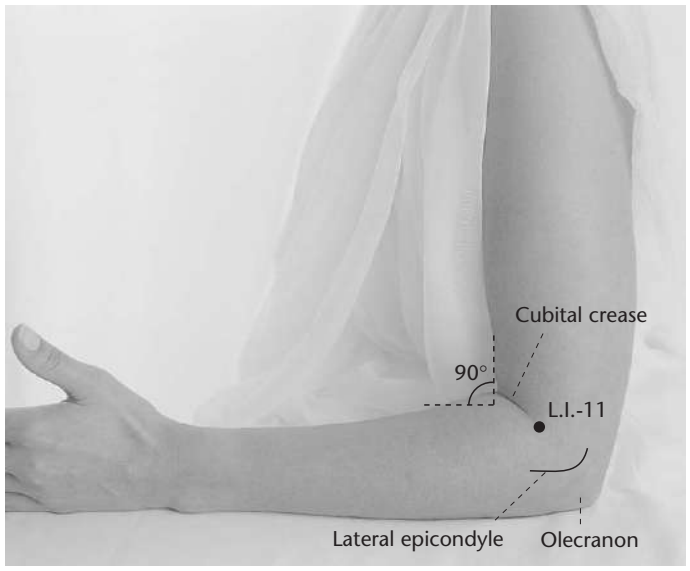

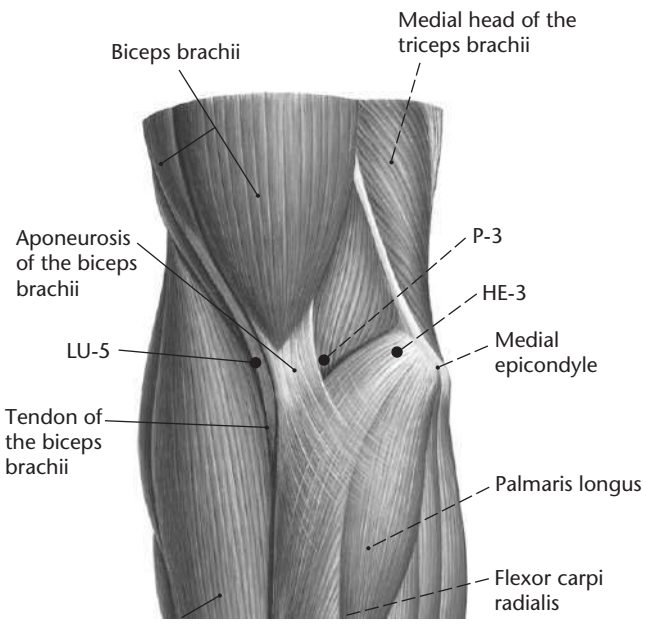

Left arm, anterior aspect

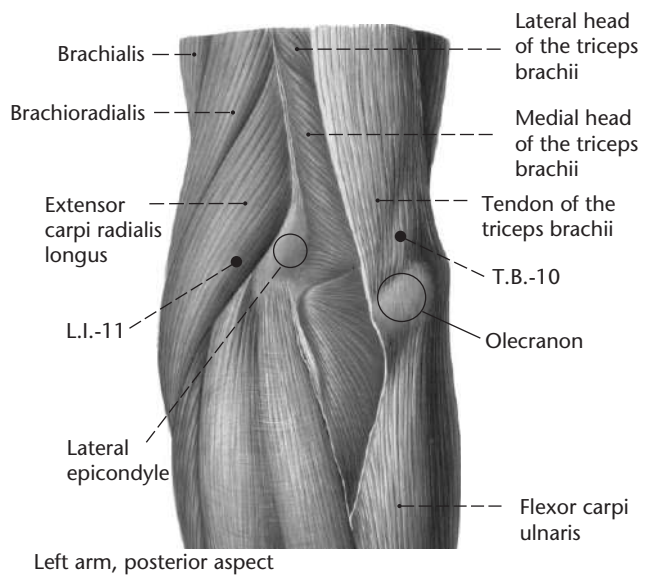

Left arm, posterior aspect

Location
With the elbow flexed, on the lateral end of the elbow crease, in a depression between the end of the crease and the lateral epicondyle of the humerus, on the extensor carpi radialis longus muscle.

How to find
First, with the elbow completely flexed, locate the lateral end of the elbow crease. Next, with the elbow flexed at approximately 90°, palpate for a pressure-sensitive depression in this area on the extensor carpi radialis longus muscle. **L.I.-11** is located close to the border of the proximal aspect of the ulna.
Also located at the level of the elbow crease are → **LU-5,** → **P-3** (radial/ulnar of the tendon of the biceps) and → **HE-3** (with the elbow completely flexed at the medial end of the cubital crease).

Needling
Vertically 1–1.5 cun

Actions/Indications
- Clears Heat, expels Wind
- Clears *yangming* – Fire
- Cools the Blood, drains Dampness, alleviates itching
- Opens the channel

Special features
He-sea point, Earth point, tonification point, Sun Si Miao Ghost point, Ma Dan Yang Heavenly Star point. Important point for Heat conditions and disorders of the upper extremity.

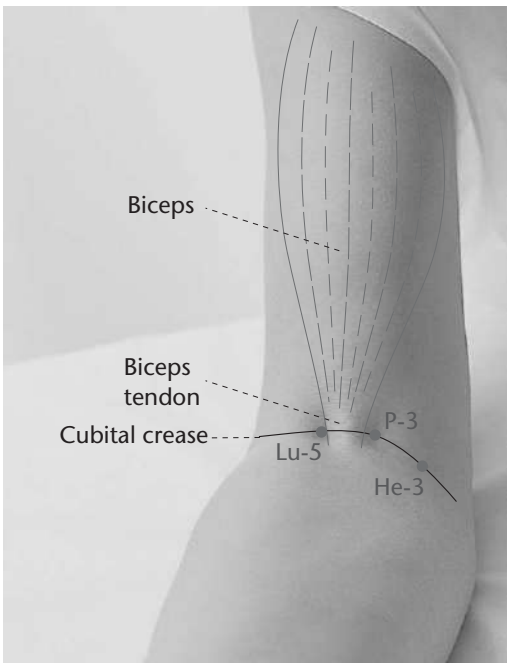

Right arm, anterior aspect

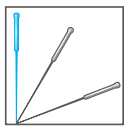
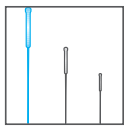

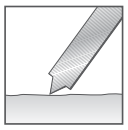
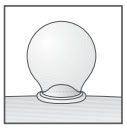

Ch04.2-F10028.qxd 2/23/08 7:05 PM Page 112

L.I.-12 *Elbow Crevice* ZHOULIAO

Location

1 cun proximal to the lateral end of the elbow crease (→ **L.I.-11**), on the anterior border of the humerus.

How to find

L.I.-12 is best located with the elbow flexed at 90°. From → **L.I.-11** (on the lateral end of the elbow crease), palpate 1 cun towards proximal. There, palpate for the anterior border of the humerus. **L.I.-12** is located at the junction of the shaft with the lateral epicondyle of the humerus and in front of the anterior border of the bone.

Needling

Vertically 0.5–1 cun, needle insertion between the border of the humerus and the flexor muscles in a medial direction.

Actions/Indications

- Opens the channel

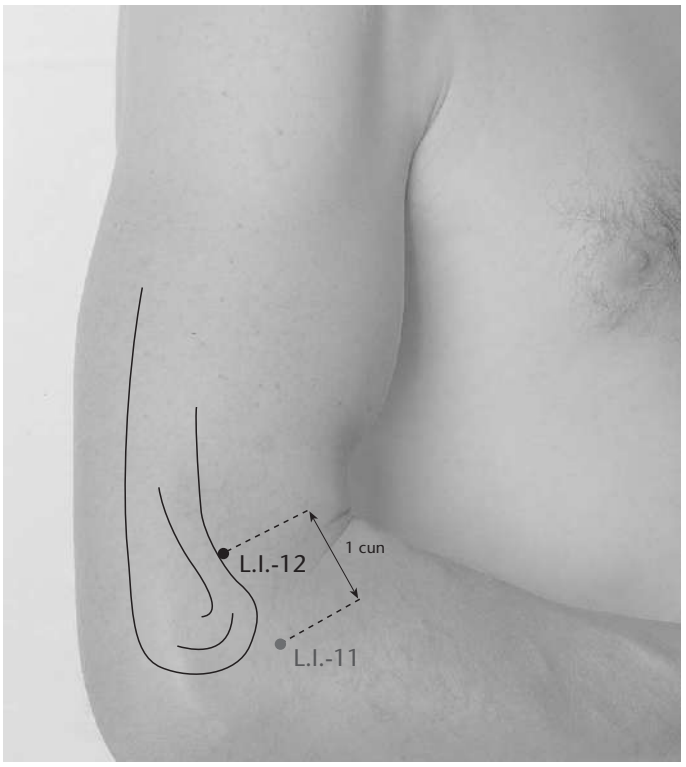

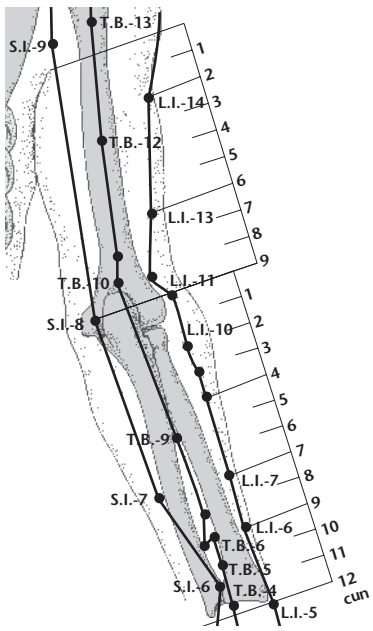

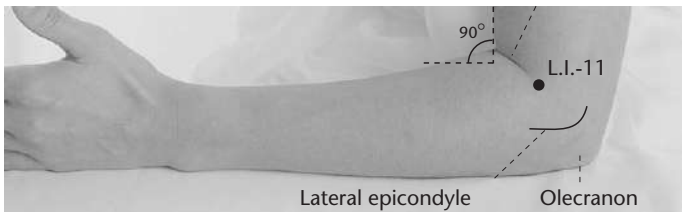

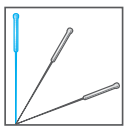
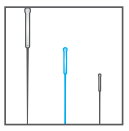

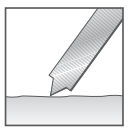

Ch04.2-F10028.qxd 2/23/08 7:06 PM Page 113

Arm Five Miles SHOUWULI L.I.-13

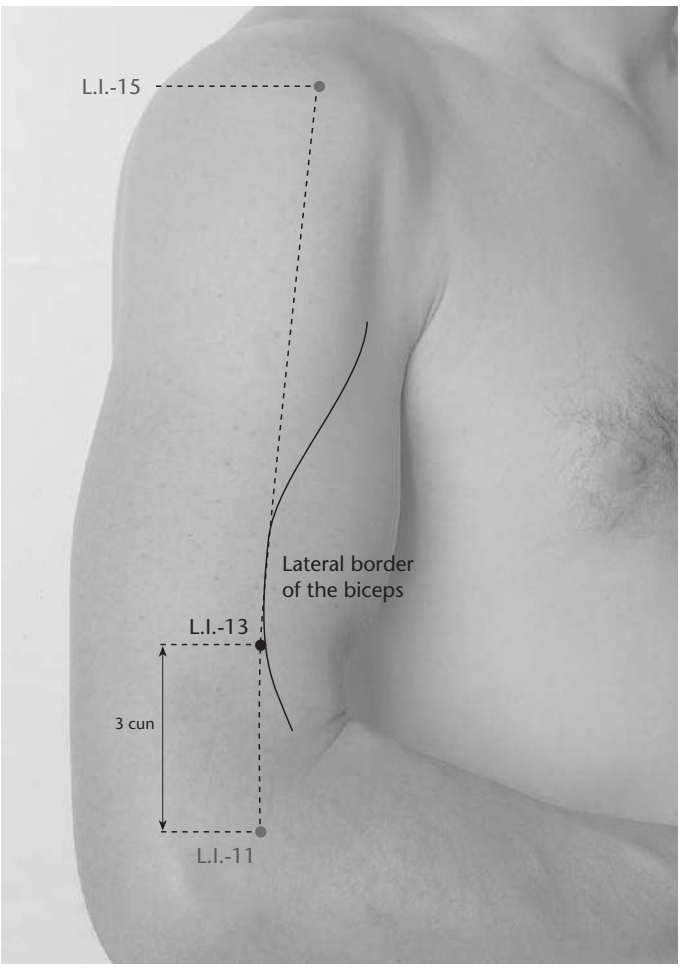

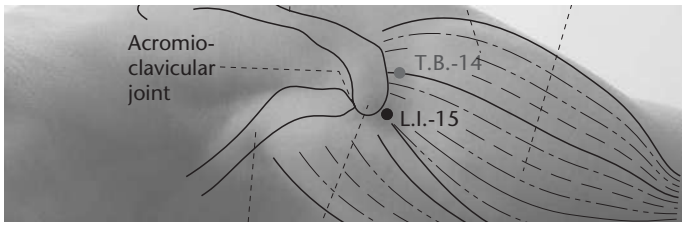

Location

On the lateral aspect of the upper arm, 3 cun proximal to the lateral end of the elbow crease (→ **L.I.-11**), in the direction of the head of the humerus.

How to find

This point is located on a line connecting → **L.I.-11** (lateral end of the elbow crease) and → **L.I.-15** (in a depression anterior and inferior to the acromion), on the lateral border of the biceps, in a groove between the biceps and the brachialis muscles. With the elbow flexed at 90°, measure 3 cun from → **L.I.-11** towards proximal. Ask the patient to flex their biceps.

Needling

Vertically 0.5–1 cun, alongside the anterior border of the humerus.

Actions/Indications

- Opens the channel

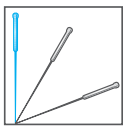
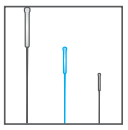

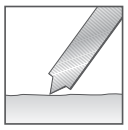
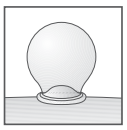

Ch04.2-F10028.qxd 2/23/08 7:06 PM Page 114

L.I.-14 *Upper Arm* BINAO

Location

On the lateral aspect of the upper arm, on a line connecting → **L.I.-11** and → **L.I.-15**, 7 cun proximal to → **L.I.-11** and slightly superior to the pointed insertion of the deltoid muscle.

How to find

This point is best located with the elbow flexed and the deltoid muscle flexed against resistance, making the lateral border of the muscle belly more visible. **L.I.-14** is located in a depression on the lower pointed insertion area of the deltoid. 'By moving the upper arm towards the tip of the nose, the patient will touch the approximate location of this point.' **For orientation:** This point lies approximately 2 cun inferior to the anterior end of the axillary fold (→ 2.2).

Needling

Vertically 0.5–1 cun; also obliquely towards the shoulder for disorders of the eyes.

Actions/Indications

- Opens the channel and *luo* vessels, alleviates pain
- Benefits the eyes

Special features

Important local point. According to some authors, meeting point with the *yang wei mai* and the S.I. and BL channels.

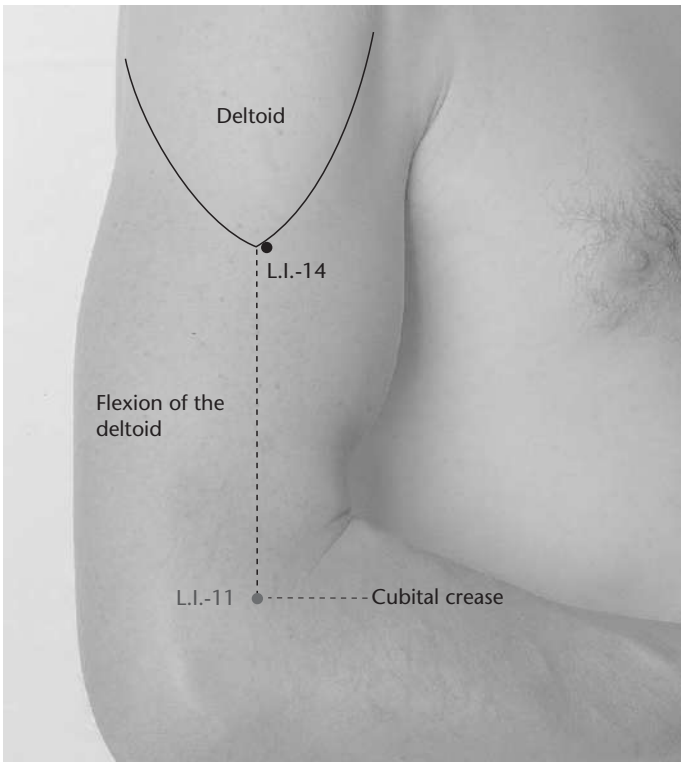

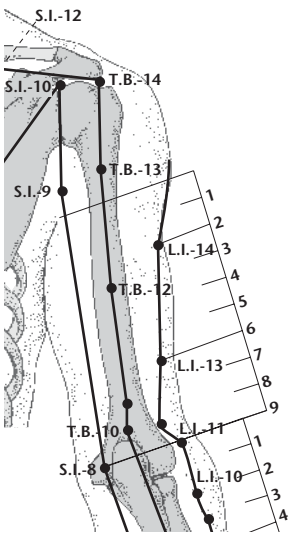

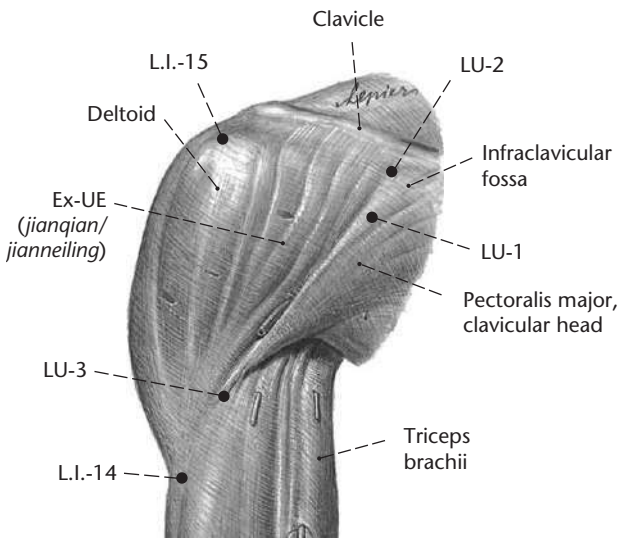

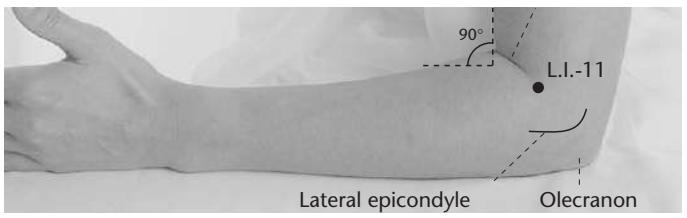

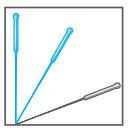
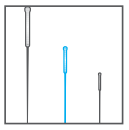

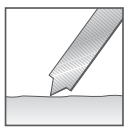

Ch04.2-F10028.qxd 2/23/08 7:06 PM Page 115

Shoulder Bone JIANYU L.I.-15

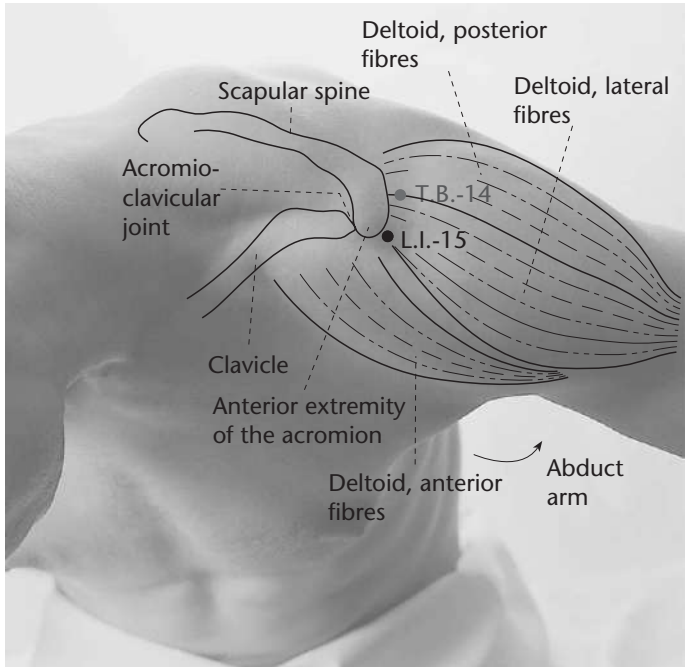

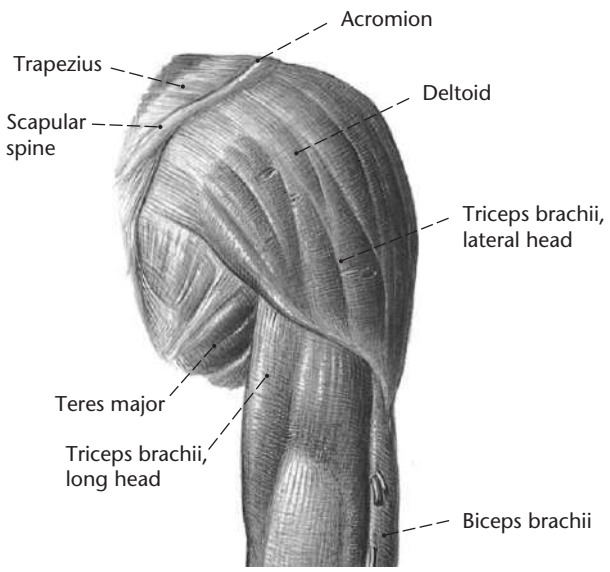

Location

In the depression distal and anterior to the acromion, between the clavicular and acromial portions of the deltoid muscle.

How to find

Ask the patient to abduct their arm. With the arm in a horizontal position, two depressions will form in the insertion area of the deltoid, distally to the acromion. By placing the thumb and index finger in those depressions in a pincer-like grip (with thumb and finger 1 thumb's breadth apart), the anterior finger will be in the anterior depression, where **L.I.-15** is located.
→ **T.B.-14** is located under the posterior finger in the dorsal depression.

Needling

With the arm slightly adducted, 1–1.5 cun obliquely in the bony cleft and in a posterior direction; with the arm abducted, 0.5–1 cun vertically. Caution: Shoulder joint.

Actions/Indications

- Expels Wind-Damp, opens the channel, alleviates pain, supports the shoulder joint
- Expels Wind
- Regulates the Qi, resolves Phlegm

Special features

Meeting point with the *yang qiao mai*. Important local point for disorders of the shoulder, often used with chain-and-lock method (→ 8.3.6) with other channel points of the upper extremity.

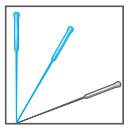
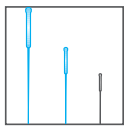

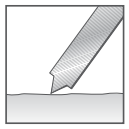
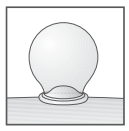

Ch04.2-F10028.qxd 2/23/08 7:06 PM Page 116

L.I.-16 *Great Bone* JUGU

Location
In a depression between the acromial extremity of the clavicle and the junction of the scapular spine and the acromion.

How to find
This point is located on the shoulder, at the angle between the acromio-clavicular joint and the junction of the scapular spine and the acromion (→ 3.3.1), above the most lateral portion of the trapezius and supraspinatus muscles. At this point, the tendon of the supraspinatus delves below the acromion, where it often causes problems owing to the cramped anatomical structure (for example, impingement syndrome).

Needling
Vertically 0.5–1 cun

Actions/Indications
- Opens the channel, supports the shoulder joint

Special features
Meeting point with the *yang qiao mai*

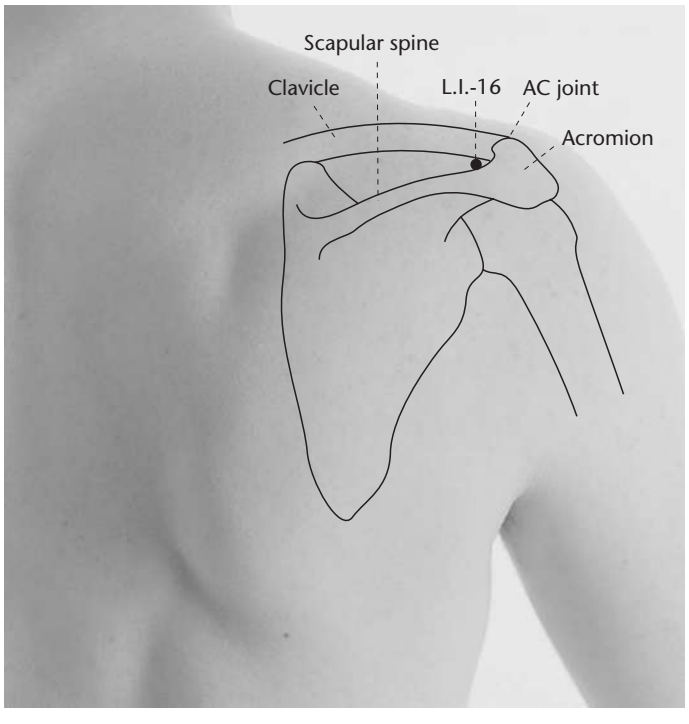

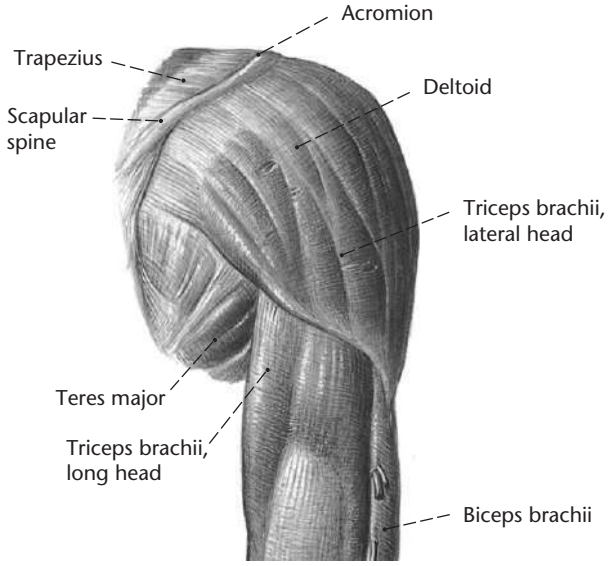

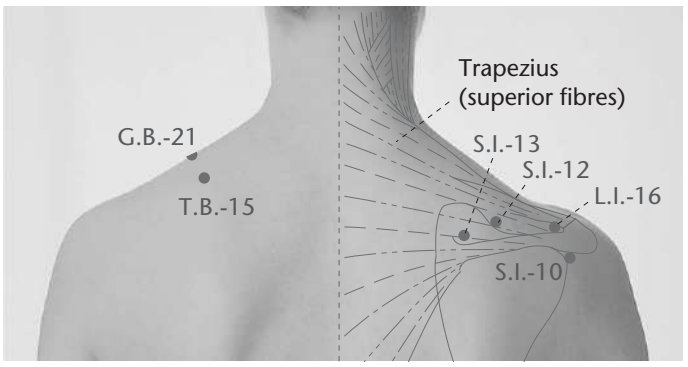

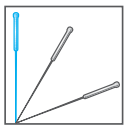
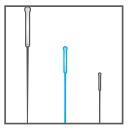

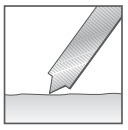

Ch04.2-F10028.qxd 2/23/08 7:06 PM Page 117

Heaven's Tripod TIANDING L.I.-17

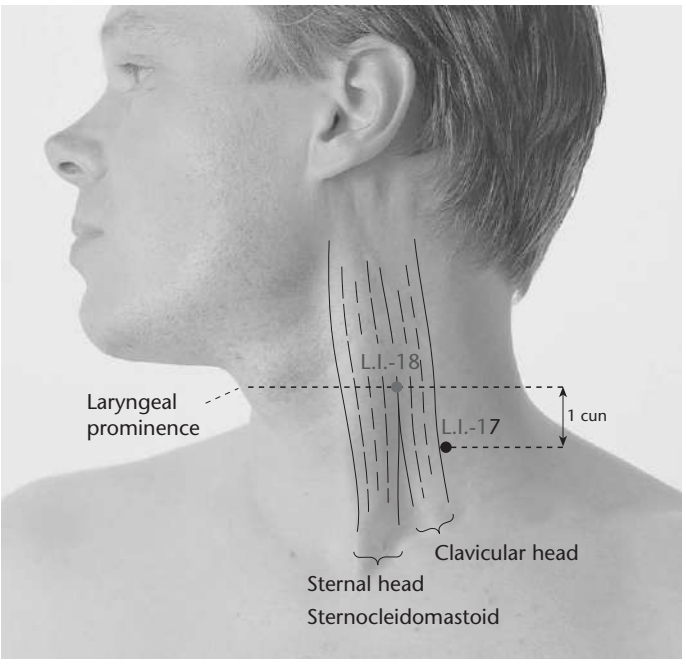

Location

On the posterior border of the sternocleidomastoid muscle, 1 cun below the laryngeal prominence (→ 3.2).

How to find

This point is located on the lateral musculature of the neck, directly posterior to the sternocleidomastoid muscle. By rotating the patient's head to the opposite side, this muscle will become more visible and palpable. → **L.I.-18** is located 1 cun superior to **L.I.-17**, between the two heads of the sternocleidomastoid.

Needling

Rotate the head back to its normal position before needle insertion. Vertically 0.3–0.5 cun or obliquely up to 0.8 cun. Caution: Carotid artery, jugular vein.

Actions/Indications

- Benefits the throat and larynx

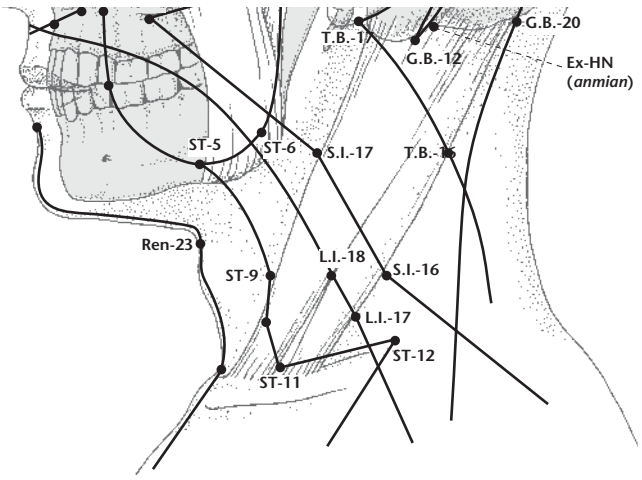

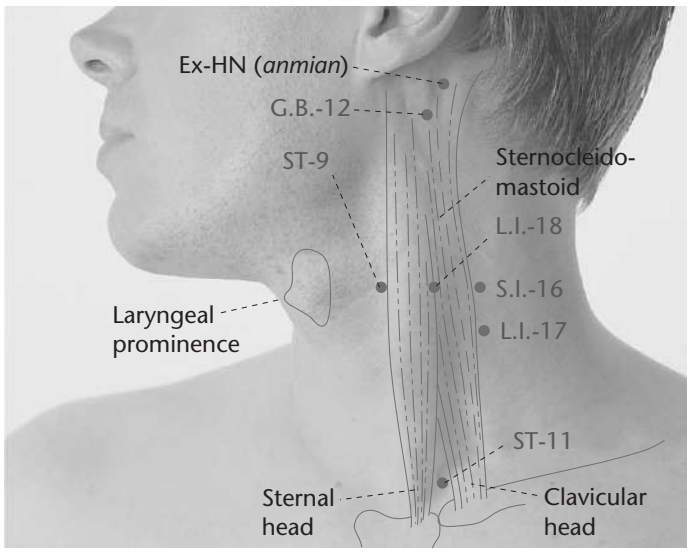

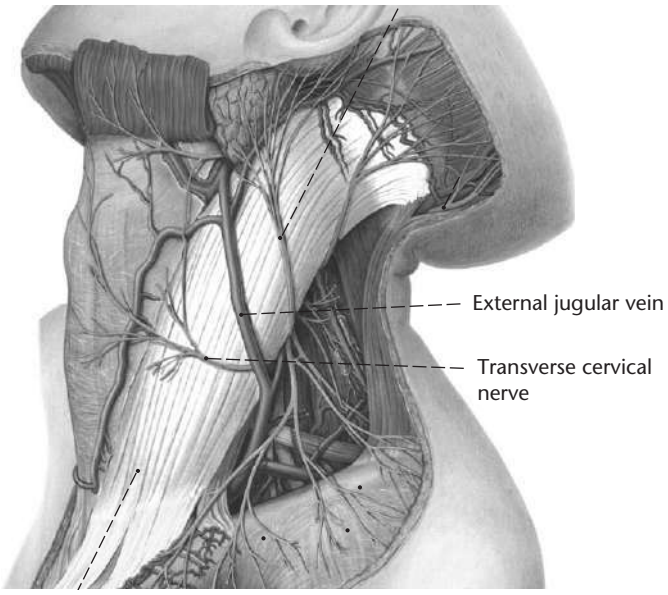

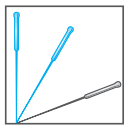
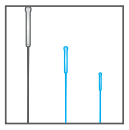

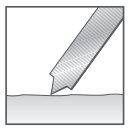

Ch04.2-F10028.qxd 2/23/08 7:06 PM Page 118

L.I.-18 *Support the Prominence* FUTU

Location

On the lateral aspect of the neck, at the level of the laryngeal prominence, between the two heads of the sternocleidomastoid muscle (→ 3.2).

How to find

From the laryngeal prominence (tip of the Adam's apple), draw a line across the sternocleidomastoid muscle to its posterior border. **L.I.-18** is located between the two heads of this muscle, which becomes more visible and palpable by rotating the head to the opposite side.

Located on the same level are → **ST-9** on the anterior border and → **S.I.-16** on the posterior border of the sternocleidomastoid muscle.

Needling

Rotate the head back to its normal position before inserting the needle. Vertically 0.3–0.5 cun or obliquely up to 0.8 cun. Caution: Carotid artery, jugular vein.

Actions/Indications

- Benefits the throat and larynx

Special features

Window of Heaven point. Often used when swallowing is difficult, for example after a stroke.

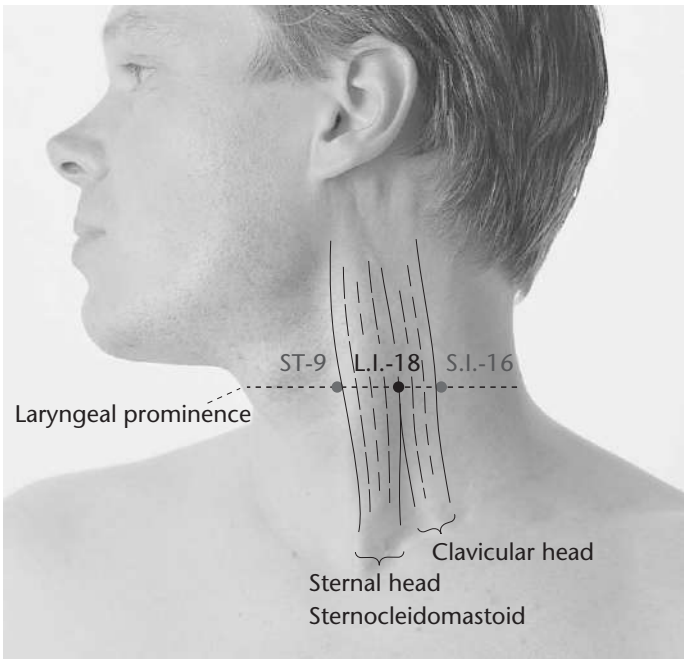

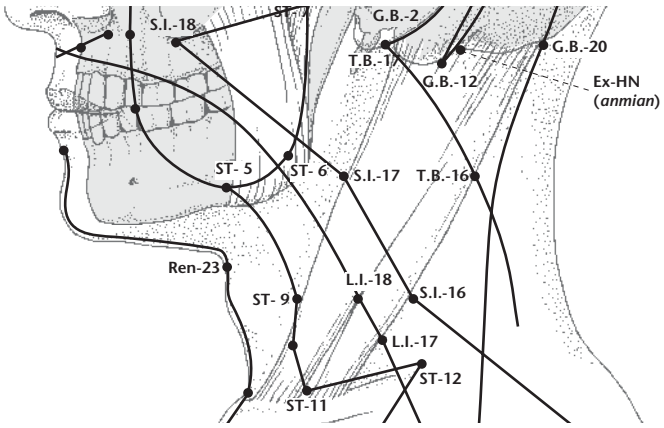

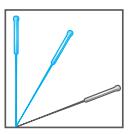

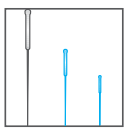

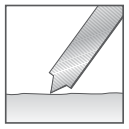

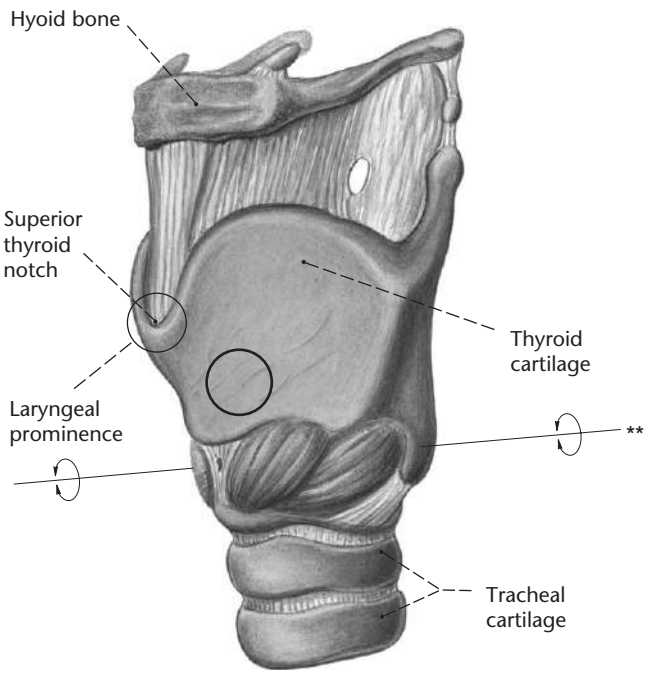

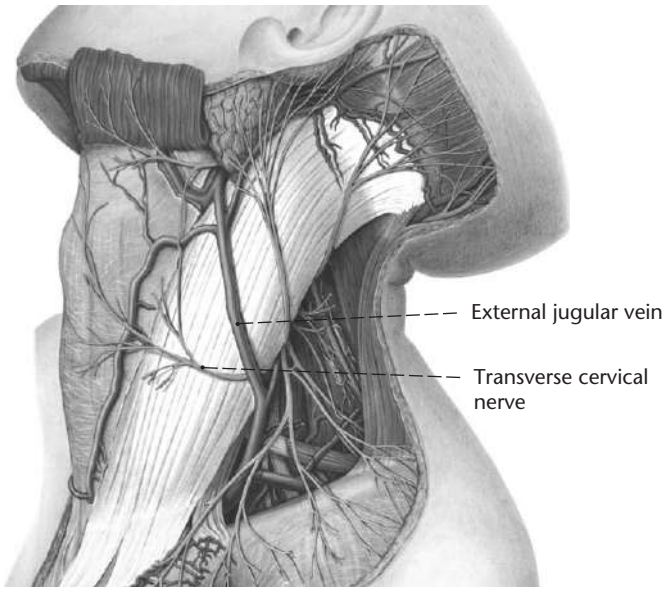

Ch04.2-F10028.qxd 2/23/08 7:06 PM Page 119

Mouth Grain Crevice KOUHELIAO L.I.-19

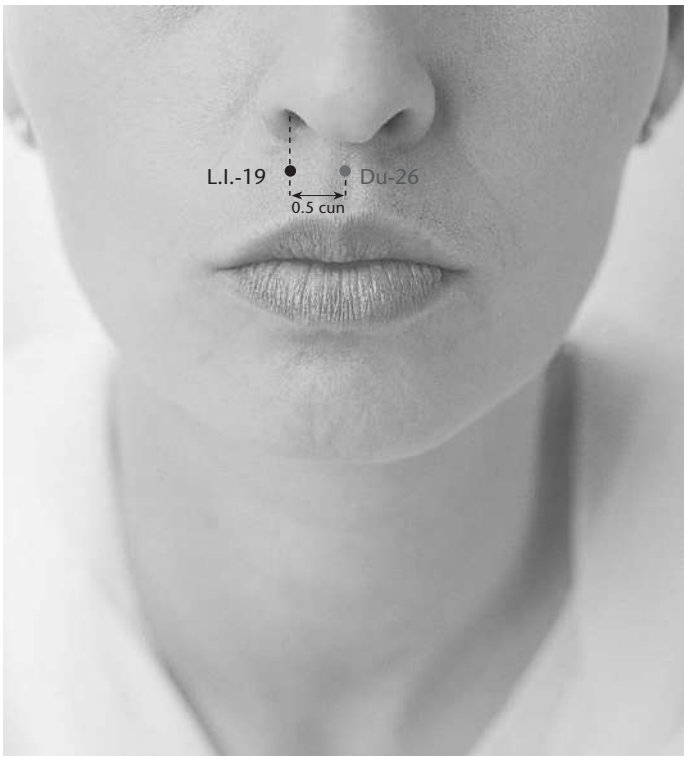

Location

On the maxilla, slightly below the lateral margin of the nostril.

How to find

By dividing the distance between the margin of the nostril and the margin of the upper lip into three sections, **L.I.-19** is located at the junction of the upper and the middle third.

→ **Du-26** is located on the same level, on the anterior midline and at a distance of approximately 0.5 cun.

Needling

Obliquely 0.3–0.5 cun. Painful point! According to some classics, moxibustion is contraindicated.

Actions/Indications

- Expels Wind and opens the nasal passages

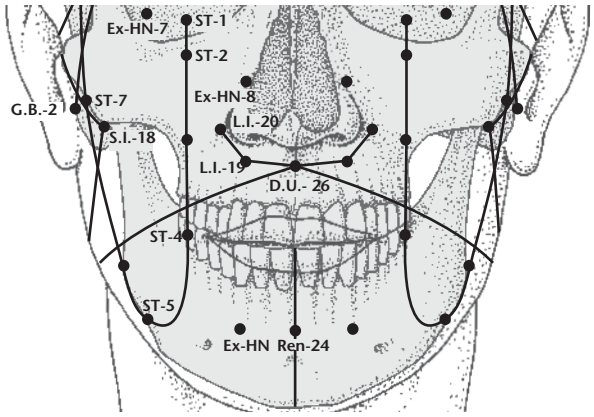

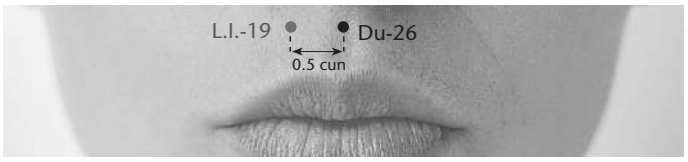

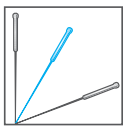
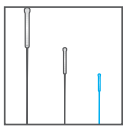

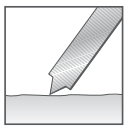

Ch04.2-F10028.qxd 2/23/08 7:06 PM Page 120

L.I.-20 *Welcome Fragrance* YINGXIANG

Location

In the nasolabial groove, on the level of the midpoint of the lateral border of the ala nasi.

How to find

L.I.-20 is located by finding the midpoint of the lateral border of the ala nasi (→ 3.1.2) and following it to the nasolabial groove. Tip: The nasolabial groove becomes more pronounced if you ask the patient to smile.

Needling

0.3–0.5 cun vertically, obliquely or transversely (subcutaneously) in a medial and superior direction towards the extra point → **EX-HN-8** (*shangyingxiang* or *bitong*: at the upper end of the nasolabial groove, at the transition between the nasal bone and the cartilage). According to some authors, moxibustion is contraindicated.

Actions/Indications

- Opens the nasal passages, expels Wind, clears Heat

Special features

Meeting point with the ST channel, exit point. Most important local point for disorders of the nose.

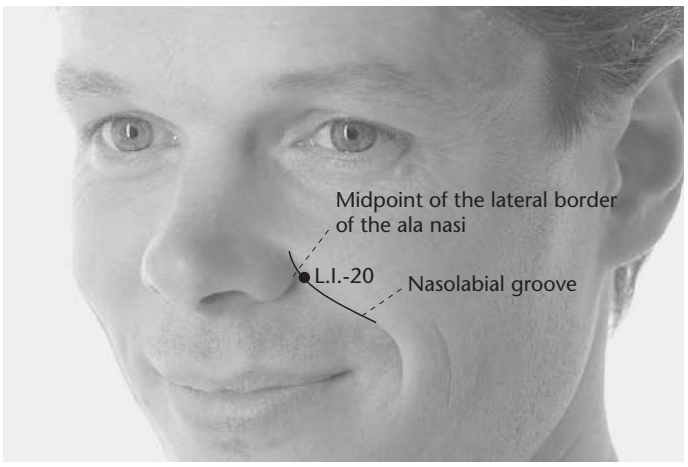

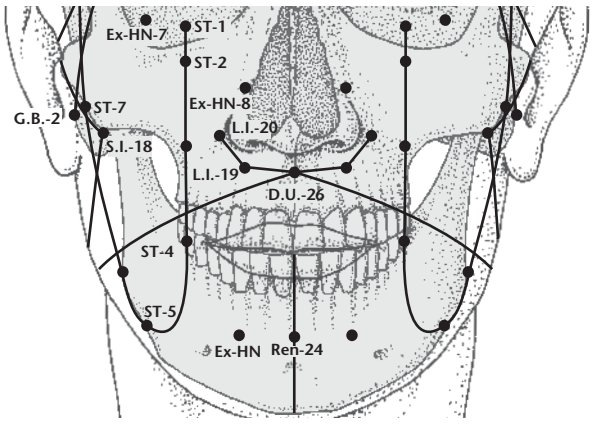

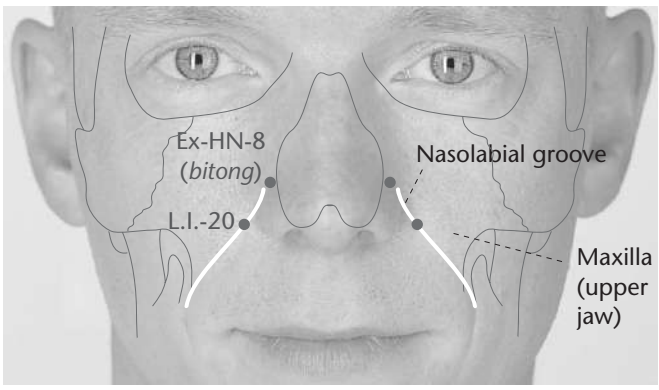

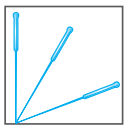
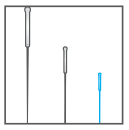

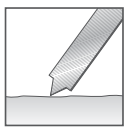

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 121

4.3 The Stomach Channel System – Foot *yangming* (*zu yang ming jing luo*)

4.3.1 The Stomach Primary Channel (*zu yang ming jing*)

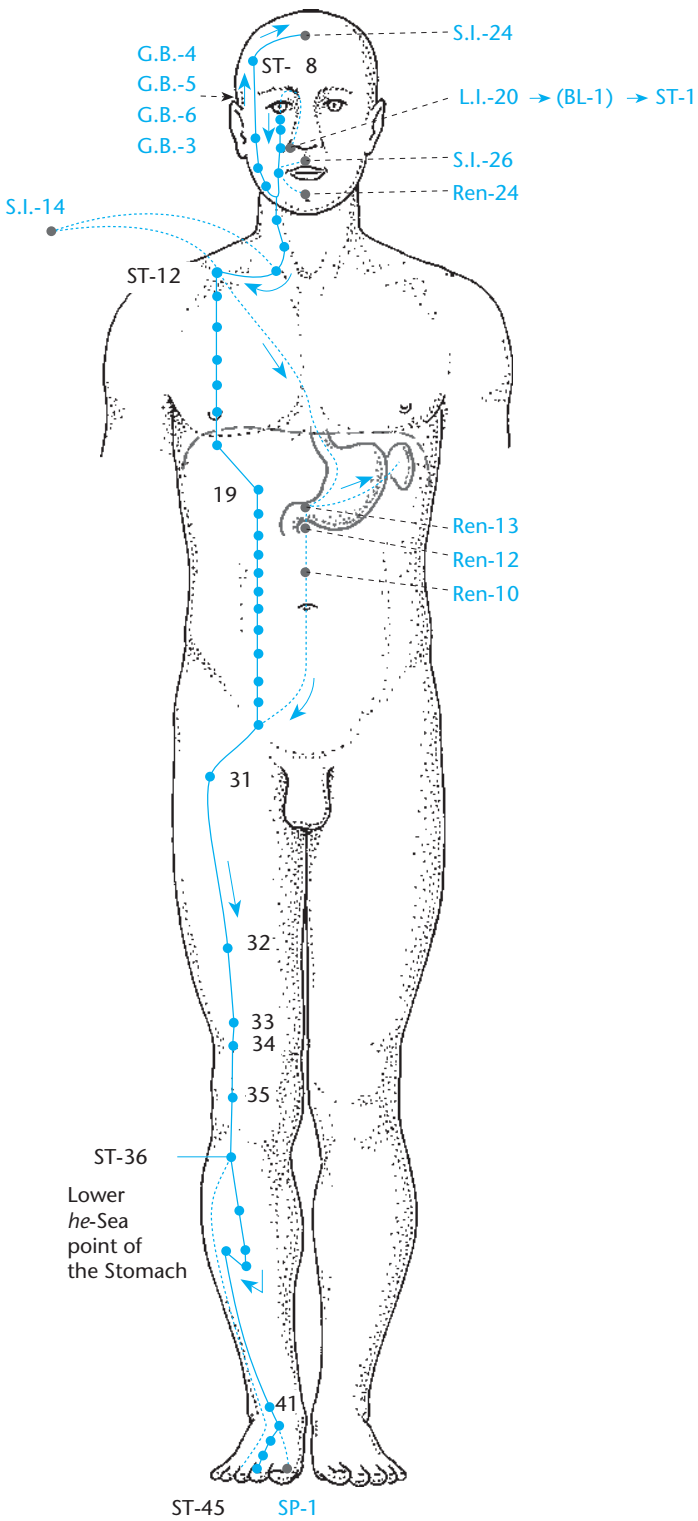

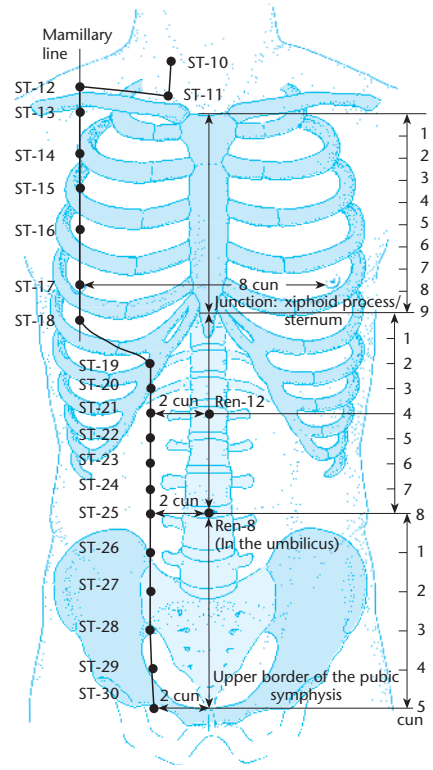

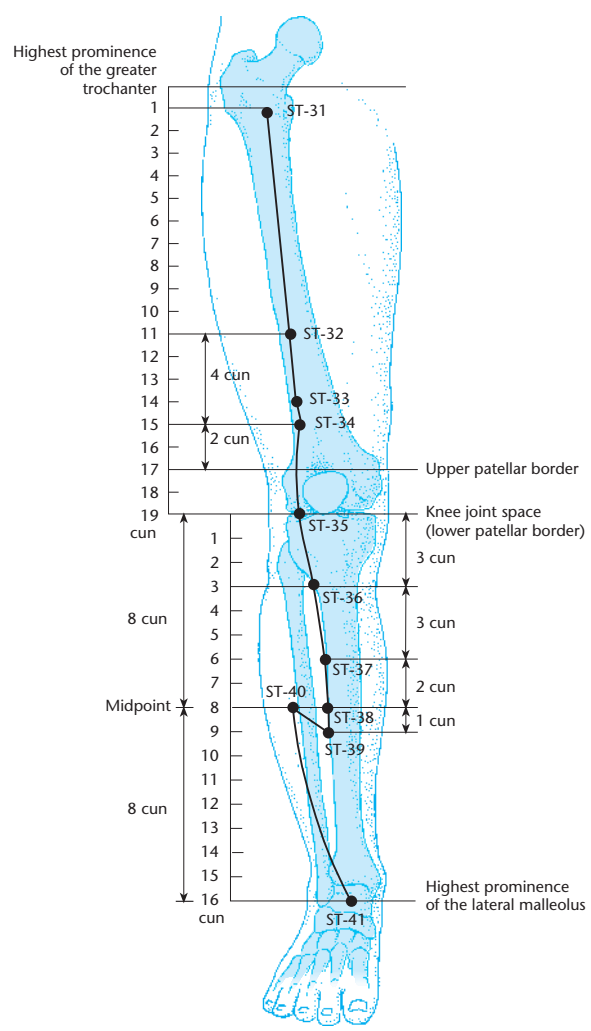

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 122

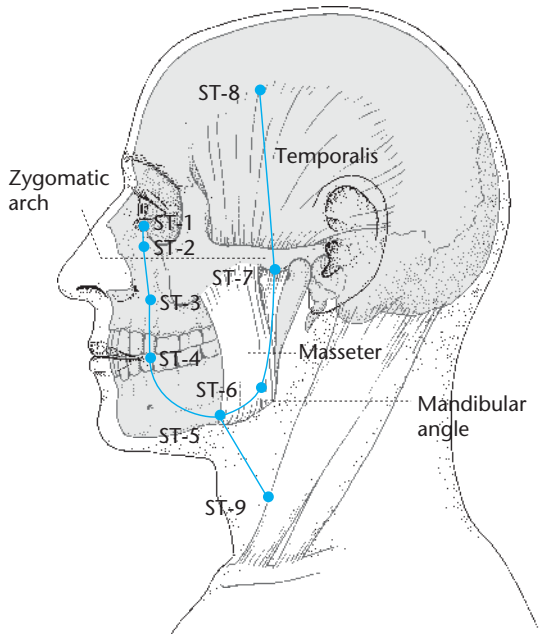

Pathway

The Stomach primary channel begins with an **internal** branch at the Yang axes connection of the first great circuit (*yangming*, hand–foot pairing) at → **LI-20** (*yingxiang*) lateral to the ala nasi

- travels **internally** to → **BL-1** (*jingming*) at the inner canthus of the eye
- emerges at the infraorbital ridge at **ST-1** (*chengqi*), the beginning of the external part of the primary channel
- descends laterally along the nose and **enters** the gums of the upper jaw
- meets the contralateral Stomach primary channel and the *du mai* at → **Du-26** (*renzhong*) and emerges again
- curves around the lips to meet the *ren mai* and again the contralateral Stomach channel at → **Ren-24** (*chengjiang*)
- traverses the lower cheek and winds around the lower angle of the mandible.

Here, at → **ST-5** (*daying*), the external channel **divides into two branches**:

- **One branch** ascends anterior to the ear, travels via → **G.B.-3** (*shangguan*), → **G.B.-6** (*xuanli*), → **G.B.-5** (*xuanlu*) and → **G.B.-4** (*hanyan*) to the temporal region to → **ST-8** (*touwei*) and continues to the forehead to meet → **Du-24** (*shenting*), where it terminates.
- The **other external branch** descends along the lateral aspect of the neck to the supraclavicular fossa and travels posteriorly over the back of the neck to → **Du-14** (*dazhui*) below the spinous process of the 7th cervical vertebra, where it meets the other five Yang primary channels.

An internal branch descends from the supraclavicular fossa, passes the diaphragm, enters first its pertaining *fu*-Organ, the Stomach (*wei*) and then connects with its paired *zang*-Organ, the Spleen (*pi*). Minor branches of the **internal** channel intersect with → **Ren-13** (*shangwan*), → **Ren-12** (*zhongwan*) and → **Ren-10** (*xiawan*). The internal branch then descends towards the inguinal region where it emerges slightly superior to **ST-30** (*qichong*) and reconnects with the **external** branch.

- The **external** channel descends from the supraclavicular fossa, crosses the nipple and laterally passes the umbilicus, travels in a curve to the leg and knee, continues along the anterior crest of the tibia and the dorsum of the foot and terminates on the second toe at → **ST-45** (*lidui*) at the lateral corner of the nail.
- An **internal** branch originates below the knee area at → **ST-36** (*zusanli*) and travels to the lateral aspect of the middle toe.
- A further branch separates from the primary channel at → **ST-42** (*chongyang*) on the dorsum of the foot and connects at → **SP-1** (*yinbai*) with its paired SP channel on the medial aspect of the big toe (foot Yin–Yang connection of the first great circuit).

Clinical importance (→ 1.2)

Exterior (*biao*) signs and symptoms: High fever, malaria, flushed face, sweating, confused mental state, aversion to cold, eye pain, dry nasal passages, nosebleeds, dry lips and mouth, lesions of the lips and mouth, sore and swollen throat, thoracic pain, red and swollen legs

Interior (*li*) or *zangfu*-Organ signs and symptoms: Abdominal distension, sensation of fullness, oedema, irritability during work and rest, mania and epilepsy, hyperpepsinia, constant hunger, yellow urine

Excess (*shi*): Heat sensations on the anterior aspect of the body, constant hunger, yellow urine

Deficiency (*xu*): Cold sensations on the anterior aspect of the body, chills, Stomach Cold with distension and fullness

Connections and meeting points of the Stomach primary channel (→ 1.2)

Connections with other channels

Spleen primary channel (*zu tai yin jing*)

Connection: Foot Yin–Yang connection of the first great circuit
Location: ST-42 → SP-1 (on the foot).
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Large Intestine primary channel (*shou yang ming jing*)

Connection: Paired according to the six channel theory (hand–foot pairing): *yang ming* (Yang axes of the first great circuit)
Location: L.I.-20 → (BL-1) → ST-1 (on the head)
Circulation: Circadian (according to the Organ clock)
Importance: Above–below relationship

Connections with other *zangfu*-Organ systems

Stomach (*wei*), Spleen (*pi*)

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 123

4.3.2 The Stomach Divergent Channel (*zu yang ming jing bie*)

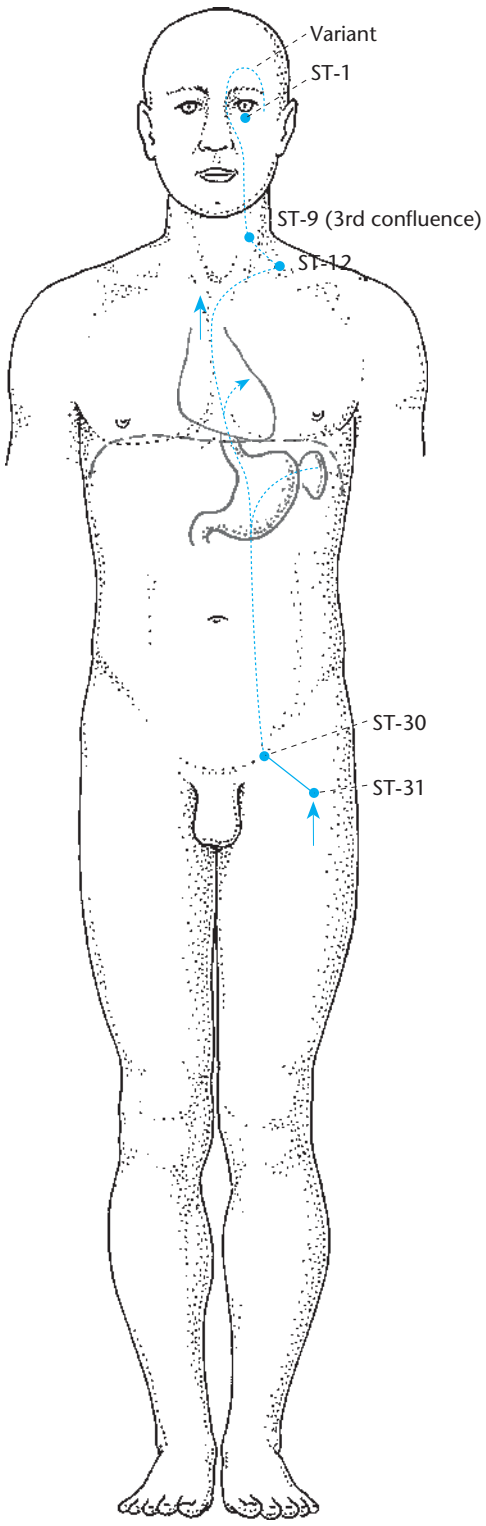

Pathway

The Stomach divergent Channel separates from the Stomach primary channel at ➞ **ST-31** (*biguan*) on the antero-lateral aspect of the thigh

- ➡ penetrates the body at the inguinal region at **ST-30** (*qichong*)
- ➡ continues to the Stomach (*wei*) and disperses in the Spleen (*pi*)
- ➡ ascends and penetrates the Heart (*xin*)
- ➡ ascends along the oesophagus to the suprasternal fossa and travels laterally to the supraclavicular fossa at **ST-12** (*quepen*).

From there it rises along the anterior border of the sternocleidomastoid muscle and meets its primary channel as well as the divergent Spleen channel at **ST-9** (*renying*) at one of the six *he*-confluences (here: ST/SP as third confluence ➞ 1.3).

It then **emerges** at the mouth, ascends along the nose, reaches the inner canthus of the eye and connects with the eye, circles the eye region and meets the Stomach primary channel at **ST-1** (*chengqi*) where it terminates (➞ fig.).

According to some authors, the Stomach divergent channel connects with the Stomach primary channel at **BL-1** (➞ variant, fig.). After intersecting with **BL-1**, the divergent channel could enter the head, disperse in the brain and re-emerge at **ST-1.**

Clinical importance

- Strengthens the relationship between the Stomach and the Spleen (*zangfu*-Organs). Points on the Stomach primary channel can therefore be used for disorders of the Spleen, and vice versa points on the Spleen primary channel can treat disorders of the Stomach.
- Spreads Qi to the face and sensory organs: many points on the ST channel treat disorders of the head and face.
- Strengthens the relationship between the Stomach and the eyes: Heat and excess in this region can be directed downward by using points on the ST channel.

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 124

4.3.3 The Stomach Sinew Channel (*zu yang ming jing jin*)

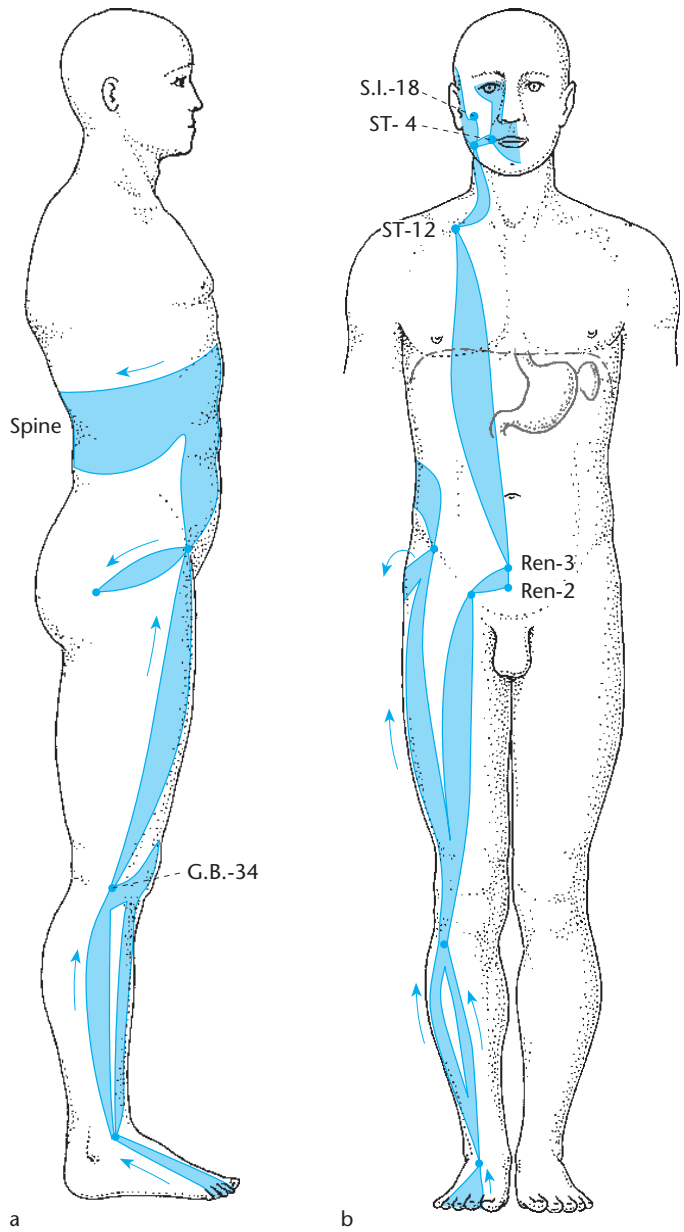

Pathway

The Stomach sinew channel begins as a wider band, encompassing the second, third and fourth toe, traverses the dorsum of the foot, binds (*jie*) at the tibial depression and divides into two branches:

- **One branch** obliquely ascends the **anterolateral aspect** of the leg and binds (*jie*) at the lateral aspect of the knee. It then ascends the anterolateral aspect of the thigh and binds (*jie*) at the anterior hip region. From there it sends **a branch** to → **G.B.-30** (*huantiao*) in the direction of the greater trochanter. The sinew channel travels from the anterior hip region across the lateral abdomen and spreads across the lower ribcage posteriorly to the spine.
- **Another branch** follows the tibia on the **anterior aspect** of the lower leg and binds (*jie*) at the level of the patella. From there a branch runs to the head of the fibula where it meets the G.B. sinew channel at → **G.B.-34** (*yanglingquan*).
- From the patella the channel ascends along the anterior aspect of the thigh, binds (*jie*) in the inguinal region, passes → **Ren-2** (*qugu*) and binds (*jie*) at → **Ren-3** (*zhongji*). It then ascends across the abdominal and thoracic region and reaches the supraclavicular fossa to bind (*jie*) at **ST-12** (*quepen*). It then continues along the anterolateral aspect of the neck and binds (*jie*) at the angle of the mandible.

At the angle of the mandible the channel divides into three branches:

- **One branch** terminates anterior to the ear.
- **A further branch** ascends to the zygomatic bone where it meets the other Yang sinew channels at → **S.I.-18** (*quanliao*).
- **The third branch** circles the mouth, ascends to the upper nose region, there connects with the Bladder sinew channel and then spreads along the lower eyelid. The Bladder sinew channel spreads along the upper eyelid, so that the two channels together form a network around the eyes.

Clinical importance

Pathology: Stiffness and aching of the toes, leg cramps (gastrocnemius and quadriceps muscles), stiffness and pain on the dorsum of the foot (at **ST-41**), swelling and tension in the inguinal region, *shan*-disorders, cramping in the abdomen as well as in the supraclavicular fossa and the face, facial paralysis, weakness and paralysis of the superior rectus muscle.

Indication: Mainly used for *bi*-syndromes (painful obstruction syndromes) along the Stomach channel. The area covered by the Stomach sinew channel is larger than that covered by the Stomach primary channel. This explains why the indications of points on the Stomach primary channel include disorders and diseases of the external genitalia (for example urinary tract disorders such as cystitis, hernia and orchitis, which are often treated in combination with LIV points) as well as disorders of the eyelids and of all sensory organs.

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 125

4.3.4 The Stomach *luo*-Connecting Vessel System (*zu yang ming luo mai*)

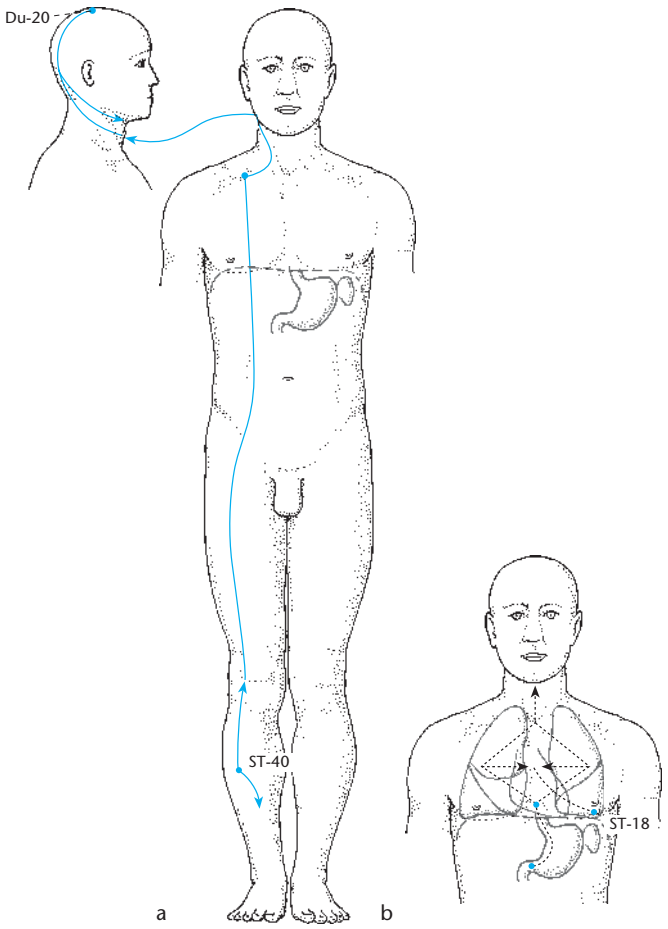

Pathway

The ST *luo*-connecting vessel separates from the Stomach primary channel at its *luo*-connecting point **ST-40** (*fenglong*) (→ 8.1.2). It forms a three-dimensional reticular network, dividing into multiple branches and sub-branches *(sun luo, fu luo, xue luo* → 1.5) within the surrounding tissue.

- ➡ Horizontal divisions run to the Interiorly–Exteriorly paired primary SP channel; according to some schools of thought (for example Van Nghi, → Appendix) they travel as a **transverse** Stomach *luo*-connecting vessel to the *yuan* point → **SP-3** (*taibai*).
- ➡ A **longitudinal** division ascends along the anterolateral aspect of the leg to the torso and reaches the lateral aspect of the neck at → **ST-12**, where it divides into two branches: **one branch** traverses the neck, the **other branch** runs to the occiput and ascends the head to reach → **Du-20** (*baihui*).

Clinical importance (→ 8.1.2)

Pathology

- **Counterflow Qi:** Swelling and pain of the throat, sudden loss of voice, sensation of tightness in the throat
- **Excess (*shi*):** Psychological disorders, also epilepsy and mania
- **Deficiency (*xu*):** *Wei*-syndrome of the legs (atrophy syndrome with muscle weakness and paralysis)

The great *luo*-connecting vessel of the Stomach (*wei zhi dao luo* or *xu li*)

Pathway

The great Stomach *luo*-connecting vessel begins at the Stomach (*wei*), passes through the diaphragm, intersects with → **Ren-17** (*shanzhong*) and spreads in the Lungs (*fei*), trachea and larynx. From the Lung it travels to the Heart (*xin*) and emerges on the left aspect of the thorax near → **ST-18** (*rugen*) which is located where the heartbeat is visible (→ 8.1.2).

Clinical importance (→ 8.1.2)

Pathology

- **Excess (*shi*):** Dyspnoea
- **Deficiency (*xu*):** Chest tightness, for example with asthma, coughing, angina pectoris, etc.

4.3.5 Cutaneous Region (*yang ming pi bu*)

See description and figures → 1.6

4.3.6 Points of the Stomach Primary Channel (Overview)

Specific points according to their function (→ 8.1)

- ***Yuan*-source point (→ 8.1.1): ST-42** (*chongyang*) ■■
- ***Luo*-connecting point (→ 8.1.2.): ST-40** (*fenglong*) ■■
- ***Xi*-cleft point (→ 8.1.3): ST-34** (*liangqiu*) ■■
- **Associated Back-*shu* point (→ 8.1.4): BL-21** (*weishu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): Ren-12** (*zhongwan*) ■■
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Metal), sedation point: **ST-45** (*lidui*) ■■
  ying-spring point (Water): **ST-44** (*neiting*) ■■
  shu-stream point (Wood): **ST-43** (*xiangu*) ■■
  jing-river point (Fire), tonification point: **ST-41** (*jiexi*) ■■
  he-sea point (Earth), *ben* point: **ST-36** (*zusanli*) ■■
- ***Hui*-meeting point (→ 8.1.7): –**
- **Opening point (→ 8.1.8): –**
- **Lower *he*-sea points (→ 8.1.9):**
  of the Stomach: **ST-36** (*zusanli*) ■■
  of the Large Intestine: **ST-37** (*shangjuxu*) ■■
  of the Small Intestine: **ST-39** (*xiajushu*) ■■
- ***jiaohui*-meeting points (→ 8.1.10):**

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 126

– with the *ren mai, yang qiao mai, du mai**: **ST-1** (*chengqi*) ■■
– with the *yang qiao mai**: **ST-2** (*sibai*) ■■
– with the *yang qiao mai*: **ST-3** (*juliao*)
– with the primary L.I. channel, *yang qiao mai, du mai*, ren mai*:* **ST-4** (*dicang*)
– with the G.B. channel*: **ST-5** (*daying*), **ST-6** (*jiache*)
– with the G.B. channel: **ST-7** (*xiaguan*) ■■, **ST-9** (*renying*)
– with the G.B. channel, *yang wei mai*: **ST-8** (*touwei*) ■■
– with the primary L.I., S.I., T.B., G.B. channels: **ST-12** (*quepen*)
– with the *chong mai*, G.B. channel*: **ST-30** (*qichong*) ■■
– other channels with the ST channel: **L.I.-20, BL-1, G.B.-3, G.B.-4, G.B.-5, G.B.-6; Du-14, Du-24; Du-26, Du-28*; Ren-10*, Ren-12, Ren-13, Ren-24; G.B.-14*, G.B.-21***

- **Gao Wu command point (→ 8.1.11) for the abdomen: ST-36** (*zusanli*) ■■
- **Window of Heaven point (→ 8.1.12): ST-9** (*renying*)
- **Points of the Four Seas (→ 8.1.13):**
  Sea of Qi point: **ST-9** (*renying*)
  Sea of Grain: **ST-30** (*qichong*) ■■, **ST-36** (*zusanli*) ■■
  Sea of Blood: **ST-37** (*shangjuxu*) ■■, **ST-39** (*xiajushu*) ■■
- **Ma Dan Yang Heavenly Star points (→ 8.1.14): ST-36** (*zusanli*) ■■, **ST-44** (*neiting*) ■■
- **Sun Si Miao Ghost point (→ 8.1.15): ST-6** (*jiache*) ■■
- **Other functional points:**
  – Front-*mu* point of the Large Intestine: **ST-25** (*tianshu*) ■■
  – most important acupuncture 'Phlegm'-point: **ST-40** (*fenglong*) ■■
- Points according to region
- **Local points (→ 8.2.1):** eyes – **ST-1** (*chengqi*) ■■; mouth and teeth – **ST-4** (*dicang*), **ST-6** (*jiache*) ■■; jaw – **ST-7** (*xiaguan*) ■■; Large Intestine – **ST-25** (*tianshu*) ■■; Small Intestine and Bladder – **ST-28** (*shuidao*); urogenital region – **ST-30** (*qichong*) ■■; knee – **ST-34** (*liangqiu*) ■■, **ST-35** (*dubi*) ■■, **ST-36** (*zusanli*) ■■; foot – **ST-41** (*jiexi*) ■■
- **Adjacent points (→ 8.2.1):** forehead – **ST-8** (*touwei*) ■■; eyes – **ST-2** (*sibai*) ■■; mouth/teeth – **ST-7** (*xiaguan*) ■■; throat – **ST-10** (*shuitu*); Spleen, Stomach, Gall Bladder – **ST-21** (*liangmen*); Kidney – **ST-29** (*guilai*); abdominal region – **ST-25** (*tianshu*) ■■, **ST-26** (*wailing*), **ST-30** (*qichong*) ■■; foot – **ST-34** (*liangqiu*) ■■, **ST-36** (*zusanli*) ■■; toes – **ST-41** (*jiexi*) ■■
- **Distal points (→ 8.2.1):** frontal headaches – **ST-44** (*neiting*) ■■; nose – **ST-44** (*neiting*) ■■; mouth/teeth – **ST-44** (*neiting*) ■■; throat – **ST-44** (*neiting*) ■■; shoulder region – **ST-38** (*tiakou*) ■■; Lung – **ST-40** (*fenglong*) ■■; Spleen/Stomach – **ST-36** (*zusanli*) ■■; Large Intestine – **ST-37** (*shangjuxu*) ■■; Small Intestine – **ST-39** (*xiajushu*) ■■; epigastrium – **ST-36** (*zusanli*) ■■; shoulder – **ST-36** (*zusanli*) ■■, **ST-38** (*tiakou*) ■■

Specific points according to the channel pathway (in numerical order)

- **ST-1** (*chengqi*) ■■: meeting point with the *ren mai, yang qiao mai*, du mai* (→ 8.1.10); local point for the eyes (→ 8.2.1)
- **ST-2** (*sibai*) ■■: meeting point with the *yang qiao mai** (→ 8.1.10); adjacent point for the eyes (→ 8.2.1)
- **ST-3** (*juliao*): meeting point with the *yang qiao mai* (→ 8.1.10)
- **ST-4** (*dicang*): meeting point with the Large Intestine channel, *yang qiao mai, du mai*, ren mai** (→ 8.1.10); local point for the mouth/teeth (→ 8.2.1)
- **ST-5** (*daying*): meeting point with the Gall Bladder channel
- **ST-6** (*jiache*) ■■: Sun Si Miao Ghost point (→ 8.1.15); meeting point with the Gall Bladder channel* (→ 8.1.10); local point for the mouth/teeth (→ 8.2.1)
- **ST-7** (*xiaguan*) ■■: meeting point with the Gall Bladder channel (→ 8.1.10); adjacent point for the teeth and jaw (8.2.1); local point for the jaw (→ 8.2.1)
- **ST-8** (*touwei*) ■■: meeting point with the Gall Bladder channel, *yang wei mai* (→ 8.1.10); adjacent point for frontal headaches (→ 8.2.1)
- **ST-9** (*renying*): meeting point with the Gall Bladder channel (→ 8.1.10); Sea of Qi point (→ 8.1.13); Window of Heaven point (→ 8.1.12)
- **ST-10** (*shuitu*): adjacent point for the neck (→ 8.2.1)
- **ST-12** (*quepen*): meeting point with the L.I., S.I., T.B. and G.B. channels (→ 8.1.10)
- **ST-21** (*liangmen*): adjacent point for the Spleen, Stomach and Gall Bladder (→ 8.2.1)
- **ST-25** (*tianshu*) ■■: Front-*mu* point of the Large Intestine (→ 8.1.5); adjacent point for the abdominal region (→ 8.2.1)
- **ST-26** (*wailing*): adjacent point for the abdominal region (→ 8.2.1)
- **ST-28** (*shuidao*): local point for the Small Intestine and Bladder (→ 8.2.1)
- **ST-29** (*guilai*): local point for the urogenital region (→ 8.2.1)
- **ST-30** (*qichong*) ■■: meeting point with the *chong mai*, G.B. channel* (→ 8.1.10); Sea of Water and Grain point (→ 8.1.13); local point for the urogenital region (8.2.1); adjacent point for the abdominal region (→ 8.2.1)
- **ST-34** (*liangqiu*) ■■: *xi*-cleft point, important adjacent point for disorders of the knee (→ 8.2.1)
- **ST-35** (*dubi*) ■■: local point for the knee region (→ 8.2.1)
- **ST-36** (*zusanli*) ■■: *he*-sea point (Earth) (→ 8.1.6), *ben*-point (Five Phase point); Lower *he*-sea point of the Stomach (→ 8.1.9); *Gao Wu* command point (→ 8.1.11) for the abdomen; Sea of Water and Grain point (→ 8.1.13); *Ma Dan Yang* Heavenly Star point (→ 8.1.14); distal point for the Spleen/Stomach and epigastrium (→ 8.2.1); local point for the knee (→ 8.2.1); distal point for the shoulder (→ 8.2.1); adjacent point for the foot (→ 8.2.1)

* Mentioned by only some authors.

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 127

- **ST-37** (*shangjuxu*) ■■: Lower *he*-sea point of the Large Intestine (→ 8.1.9); Sea of Blood point (→ 8.1.13); distal point for the Large Intestine (→ 8.2.1)
- **ST-38** (*tiakou*) ■■: distal point for the shoulder region (→ 8.2.1)
- **ST-39** (*xiajushu*) ■■: Lower *he*-sea point of the Small Intestine (→ 8.1.9); Sea of Blood point (→ 8.1.13); distal point for the Small Intestine (→ 8.2.1)
- **ST-40** (*fenglong*) ■■: *luo*-connecting point (→ 8.1.2); main 'Phlegm'-point; distal point for the Lung (→ 8.2.1)
- **ST-41** (*jiexi*) ■■: *jing*-river point (Fire) (→ 8.1.6); tonification point; local point for the foot (8.2.1); adjacent point for the toes (→ 8.2.1)
- **ST-42** (*chongyang*) ■■: *yuan*-source point (→ 8.1.1)
- **ST-43** (*xiangu*) ■■: *shu*-stream point (Wood) (→ 8.1.6)
- **ST-44** (*neiting*) ■■: *ying*-spring point (Water) (→ 8.1.6); distal point for frontal headaches, as well as for the nose, mouth/teeth and throat (→ 8.2.1); Ma Dan Yang Heavenly Star point (→ 8.1.14)
- **ST-45** (*lidui*) ■■: *jing*-well point (Metal) (→ 8.1.6); sedation point.

General location help

- **ST-1** to **ST-4**: with the patient looking straight ahead, these points are located on the vertical pupil line
- **ST12** to **ST-18** are all located on the midclavicular/mamillary line, 4 cun lateral to the anterior midline
  - **ST-12**: in the supraclavicular fossa
  - **ST-13**: on the lower border of the clavicle
  - **ST-14** to **ST-18**: in the first to fifth intercostal spaces
- **ST-19** to **ST-30** are all located on a vertical line midway between the mamillary line and the anterior midline, 2 cun lateral to the anterior midline:
  - **ST-19** to **ST-25** are located on the 8 cun section (→ 2.2) between the sternocostal angle and the umbilicus (**ST-19** to **ST-25** are evenly spread on this line, from 6 cun above the umbilicus to the same level as the umbilicus)
  - **ST-26** to **ST-30** are located on the 5 cun section (→ 2.2) between the umbilicus and the upper border of the pubic symphysis (1 cun steps from 4 cun superior to the upper border of the pubic symphysis to its upper border).

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 129

Container of Tears CHENGQI ST-1

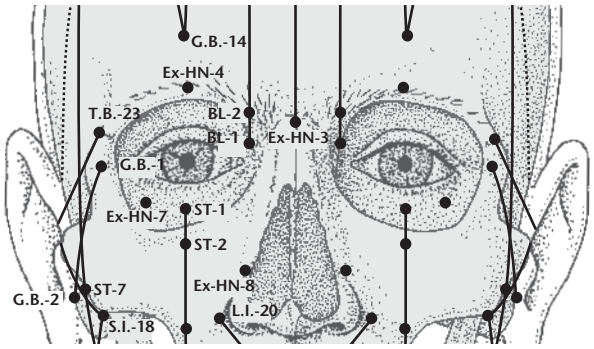

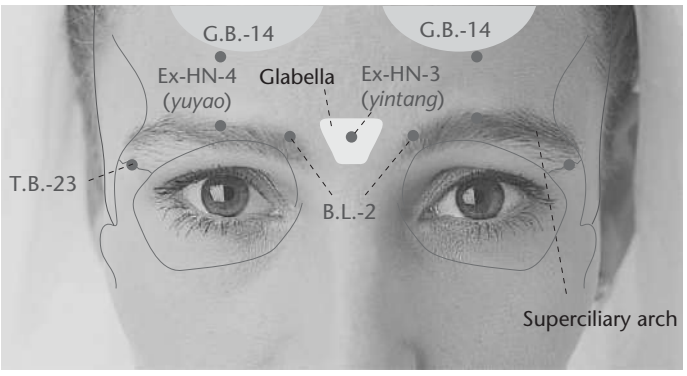

Location

With the eyes looking straight ahead, directly below the centre of the pupil, between the eyeball and the infraorbital ridge.

How to find

With the patient looking straight ahead, the first four points of the ST channel are located on a vertical line drawn through the centre of the pupil. Within the area of the lower eyelid, the infraorbital ridge presents itself as a distinct bony structure. **ST-1** is located on the pupil line, directly superior to the infraorbital ridge.

Needling

Ask the patient to look upward, gently push the eyeball upwards and insert the needle vertically along the orbital ridge in a dorsal direction. Caution: Venous plexus and arteries, avoid injuring the eyeball and periost! Needling recommended only by experienced practitioners! Pay attention to needling pain! No needle manipulation! After removing the needle, press firmly on the insertion site for about one minute to prevent haematoma. Inform the patient that it may still be possible for a haematoma to develop. Moxibustion is contraindicated! Points for disorders of the eyes with less risk of complications include **BL-2**, **T.B.-21**, **G.B.-1**, **ST-2**, **Ex-HN-5** (*taiyang*), **EX-HN-4** (*yuyao*).

Actions/Indications

- Expels Wind and Heat, clears and benefits the eyes

Special features

Meeting point with the *yang qiao mai* and *ren mai*, also with the *du mai* according to some authors, important point for the eyes, entry point.

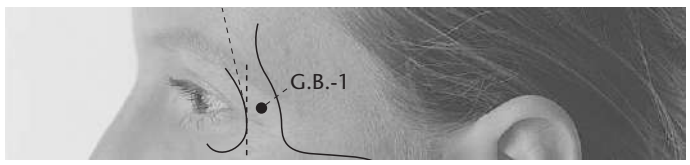

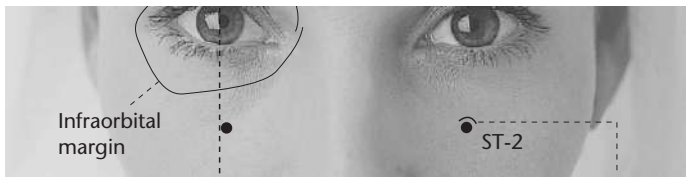

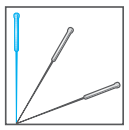
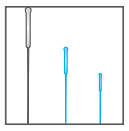
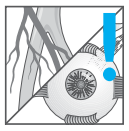

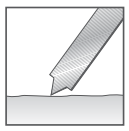

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 130

ST-2 *Four Whites* SIBAI

Location

With the eyes looking straight ahead, directly below the centre of the pupil, in the depression at the infraorbital foramen.

How to find

With the patient looking straight ahead, the first four points of the ST channel are located on a vertical line drawn through the centre of the pupil. Within the area of the lower eyelid, the infraorbital ridge presents itself as a distinct bony structure. Starting at the infraorbital ridge (→ 3.1.2), palpate inferiorly until you can feel the depression of the infraorbital foramen (this is often located slightly medial to the pupil line), where **ST-2** is located.

Needling

0.3–0.5 cun vertically or transversely (subcutaneously) towards → **S.I.-18** or → **L.I.-20** (for example, in cases of facial paralysis). Deep needling (in an oblique cranial direction) is contraindicated. Caution: Infraorbital nerve, eye injury. According to some classic texts, moxibustion is contraindicated.

Actions/Indications

- Benefits the eyes, clears Heat, eliminates Wind

Special features

Important local point for disorders of the eyes, pain and paralysis of the face; meeting point with the *yang qiao mai* according to some authors. A less risky substitute for → **ST-1**.

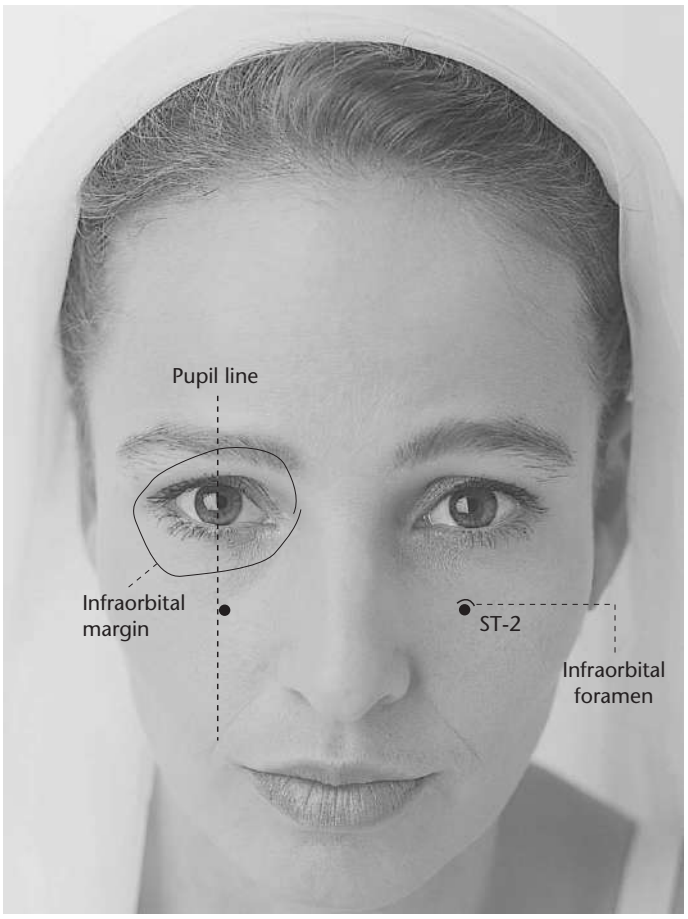

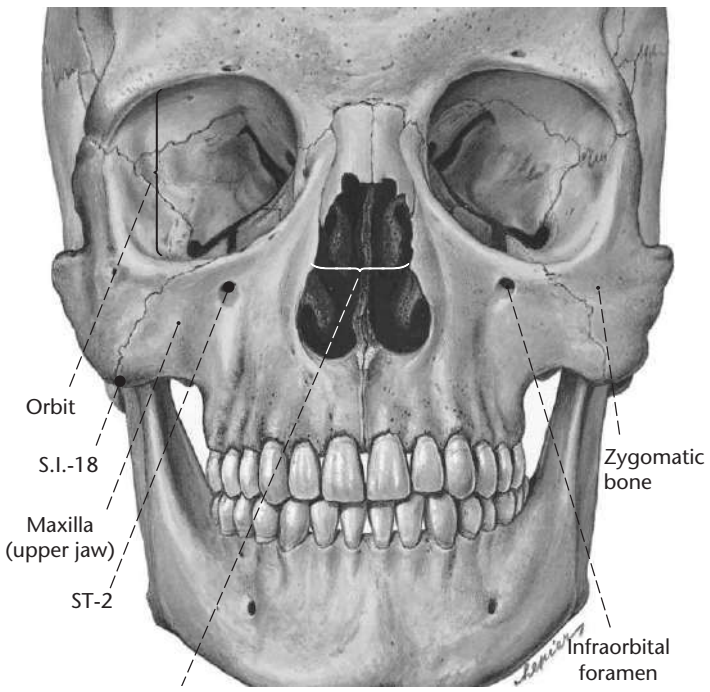

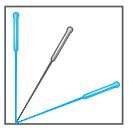
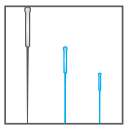
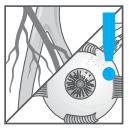

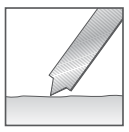

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 131

Great Crevice JULIAO ST-3

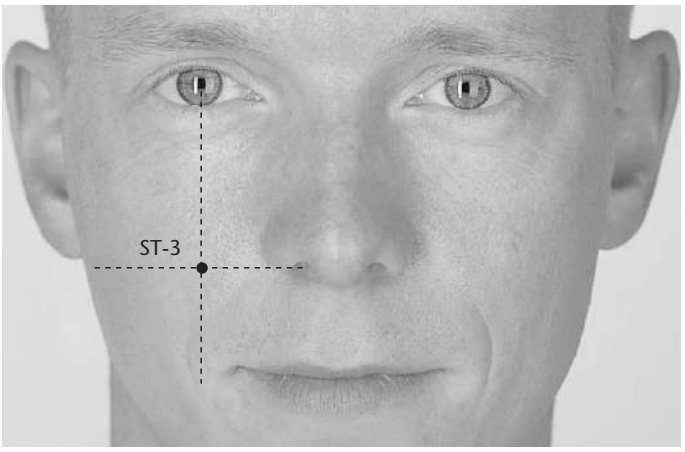

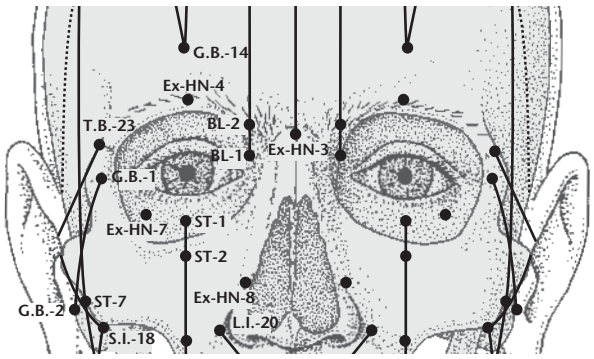

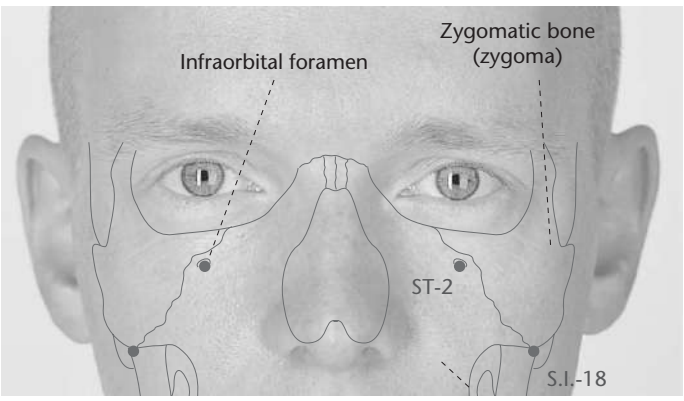

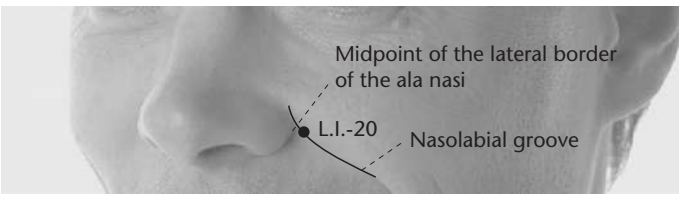

Location

With the patient looking straight ahead, directly below the centre of the pupil, on the level of the lower border of the ala nasi.

How to find

With the patient looking straight ahead, the first four points of the ST channel are located on a vertical line drawn through the centre of the pupil (pupil line). **ST-3** is located at the crossing point of the pupil line and a horizontal line on the level of the lower border of the ala nasi.

Needling

Vertically or obliquely 0.3–0.5 cun or penetration method: transversely (subcutaneously) towards → **ST-4,** → **S.I.-18**, etc, for example, in cases of peripheral facial paralysis.

Actions/Indications

- Expels Wind, opens the channel, disperses accumulations

Special features

Meeting point with the *yang qiao mai*

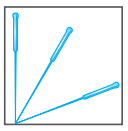
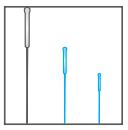

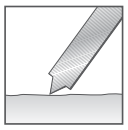
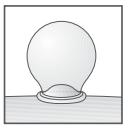

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 132

ST-4 *Earth Granary* DICANG

Location

With the patient looking straight ahead, directly below the centre of the pupil and approximately 0.4 cun lateral to the corner of the mouth.

How to find

With the patient looking straight ahead, the first four points of the ST channel are located on a vertical line drawn through the centre of the pupil (pupil line). **ST-4** is located at the crossing point of the pupil line and a horizontal line originating at the corner of the mouth. The point is located in the nasolabial groove (→ 3.1.2), approximately 0.4 cun lateral to the corner of the mouth. Ask the patient to smile in order to make the line more clearly visible.

Needling

Obliquely in a lateral direction or vertically 0.3–0.5 cun or penetration method: transversely (subcutaneously) approximately 1.5–2 cun in cases of facial paralysis towards → **ST-6**, for trigeminal neuralgia towards → **L.I.-20**. Caution: Facial artery/vein.

Actions/Indications

- Expels Wind (from the face), opens the channel, alleviates pain, relaxes the facial musculature
- Sometimes as distal point for disorders of the leg

Special features

Meeting point with the L.I. channel, the *ren mai* and *yang qiao mai* (also with the *du mai* according to some authors). Important local point for neuralgias and pareses affecting the mouth and cheeks.

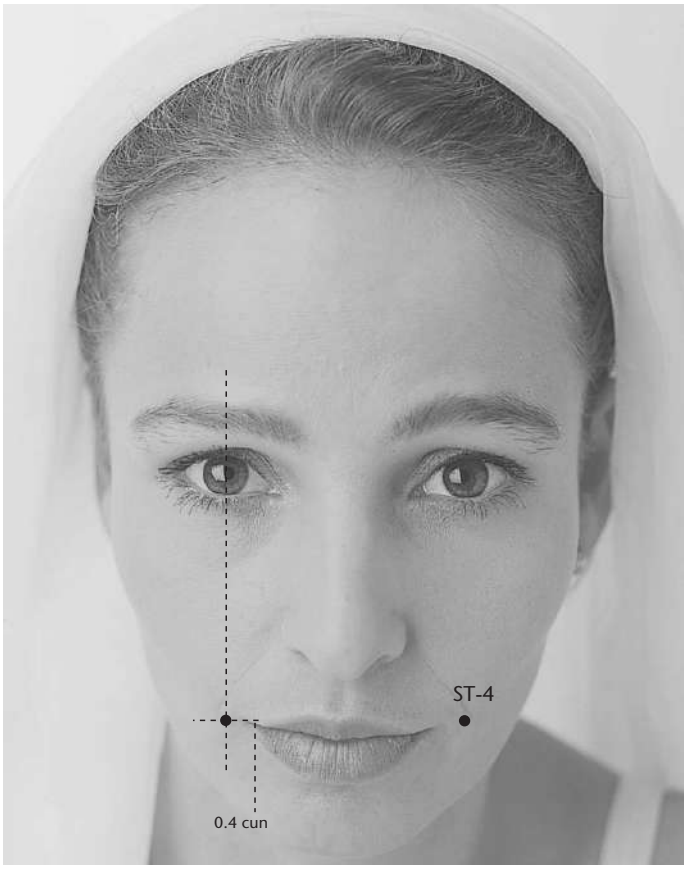

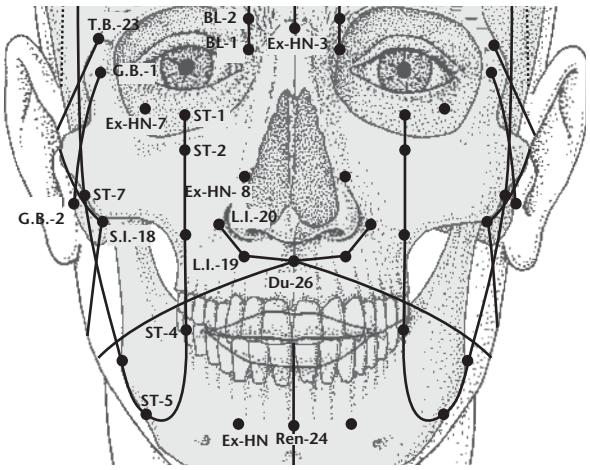

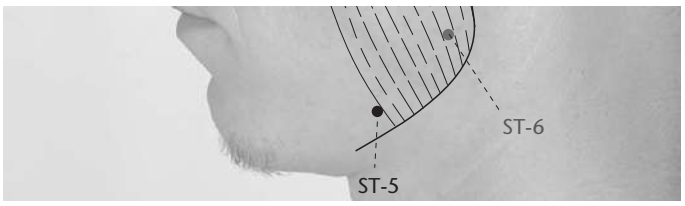

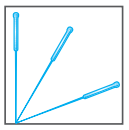
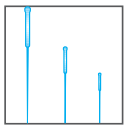

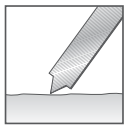

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 133

Great Welcome DAYING ST-5

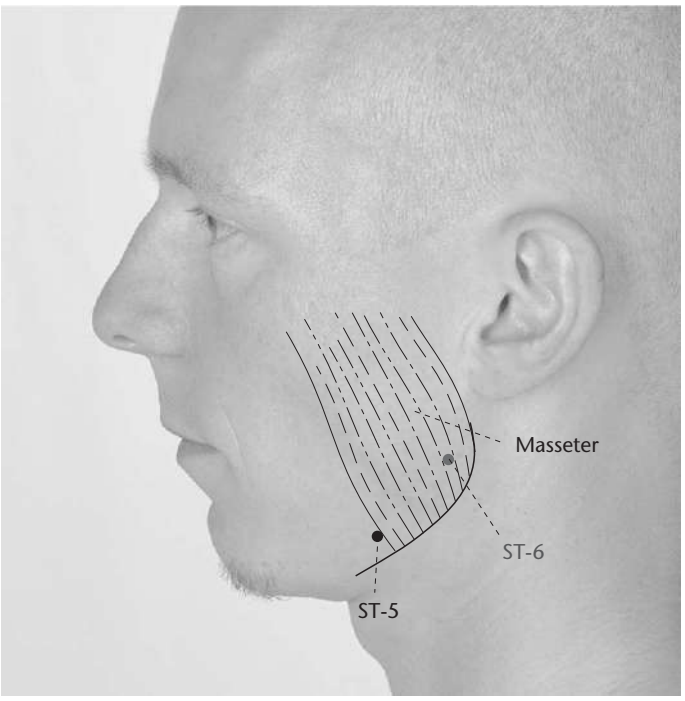

Location
On the lateral mandible, anterior to the masseter muscle.

How to find
By asking the patient to clench their teeth, the anterior border of the masseter muscle becomes clearly palpable. **ST-5** is located directly anterior to the border of the muscle, in a shallow depression on the mandible and slightly above the angle of the jaw. Here, the facial artery can be palpated.

Needling
0.3–0.5 cun obliquely or transversely (subcutaneously) in the direction of → **ST-6**. Caution: Facial artery/vein.

Actions/Indications
- Expels Wind, opens the channels

Special features
Meeting point with the G.B. channel according to some authors

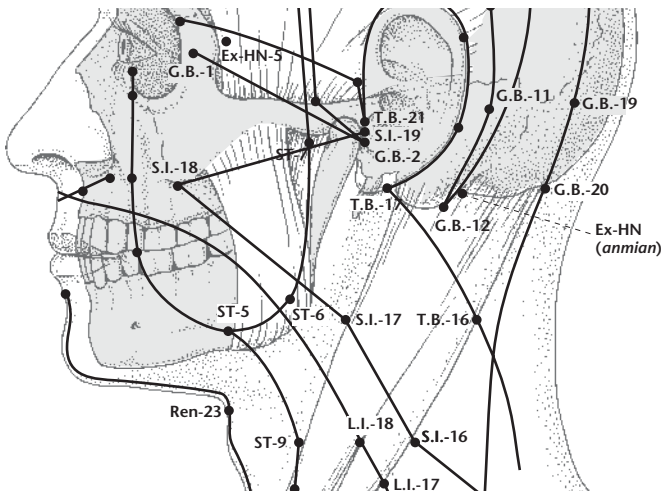

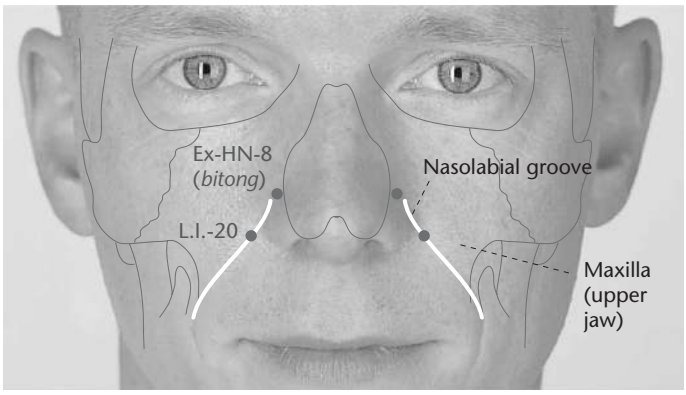

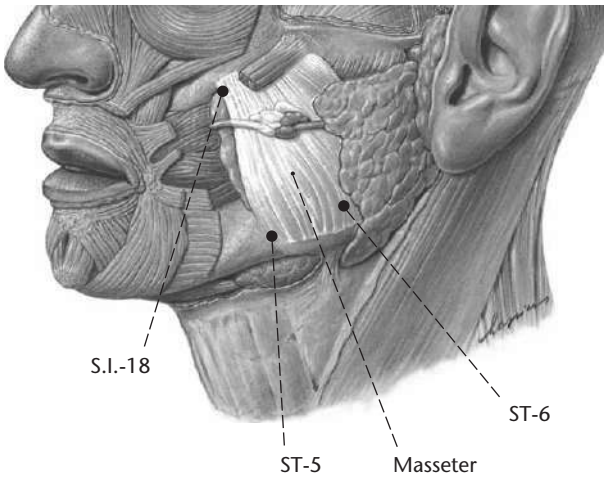

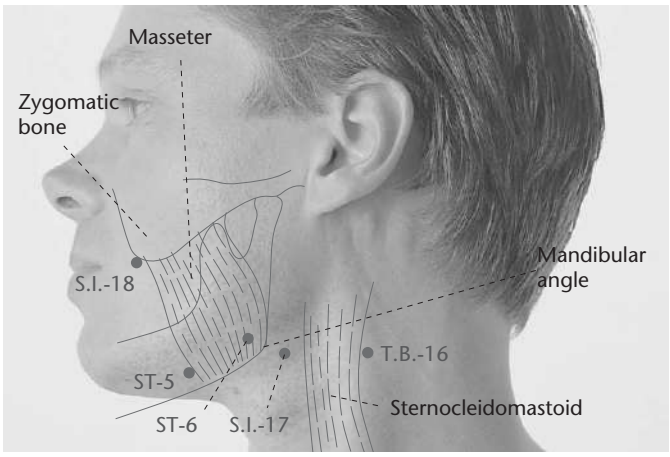

Ch04.3-F10028.qxd 2/22/08 8:41 PM Page 134

ST-6 *Jaw Bone* JIACHE

Location

With the teeth clenched, on the highest prominence of the masseter muscle, approximately one finger-width (middle finger) anterior and superior to the angle of the jaw.

How to find

Ask the patient to clench their teeth, which will allow the muscle belly of the masseter muscle to bulge. **ST-6** is located on its prominence, a little anterior and superior to the angle of the jaw (mandibular angle → 3.1.3). With chewing movements of the jaw ('biting and relaxing'), the palpating finger will naturally fall into the right depression when the muscle is relaxed.
→ **ST-5** is located more medially in the depression at the anterior border of the masseter muscle.

Needling

Insertion with the patient's mouth slightly opened (relaxed masseter muscle). Vertically 0.3–0.5 cun or obliquely or transversely (subcutaneously) 1–1.5 cun in the direction of → **ST-4**, **ST-5**, **ST-7**, for example, in cases of facial paralysis.

Actions/Indications

- Eliminates Wind, benefits the jaw and teeth, opens the channel and *luo* vessels, alleviates pain

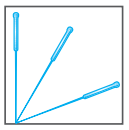

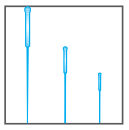

Special features

Sun Si Miao Ghost point. Important local point for the jaw region. Its location corresponds to a commonly used trigger point on the masseter muscle. According to some authors, **ST-6** is a meeting point with the G.B. channel.

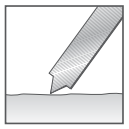

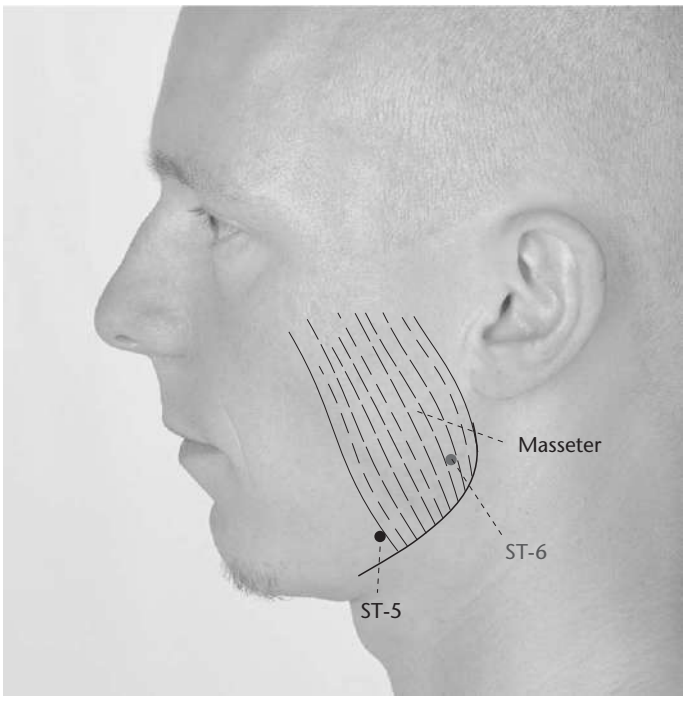

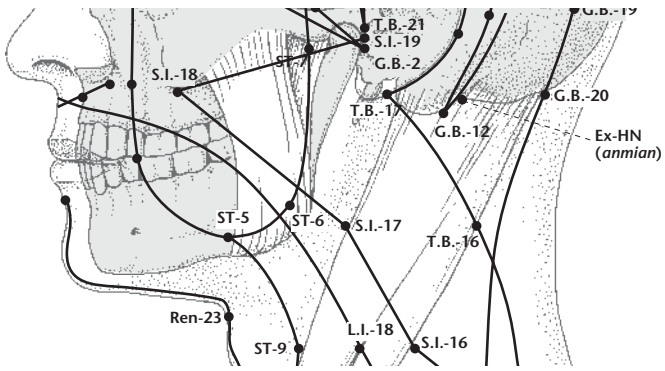

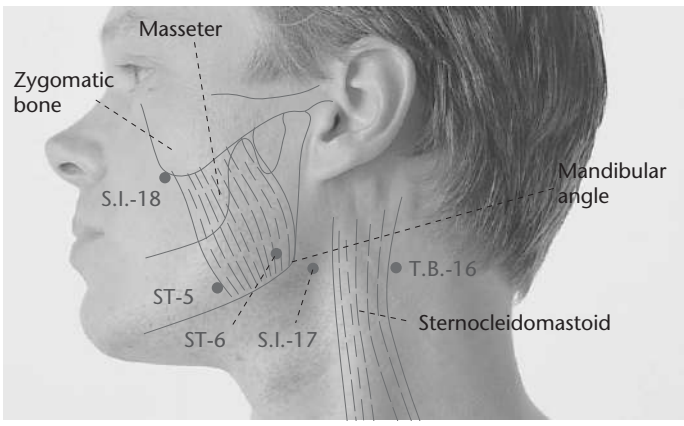

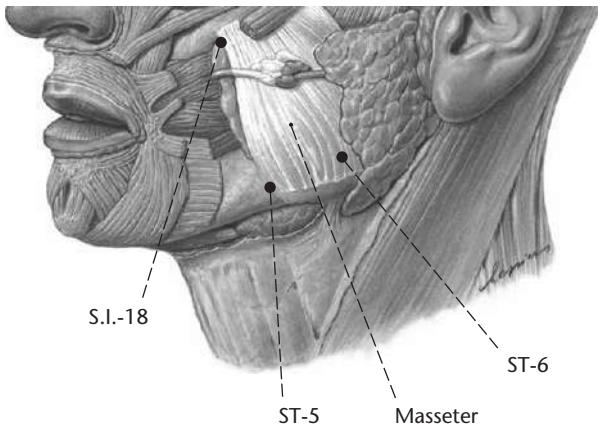

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 135

Below the Joint XIAGUAN ST-7

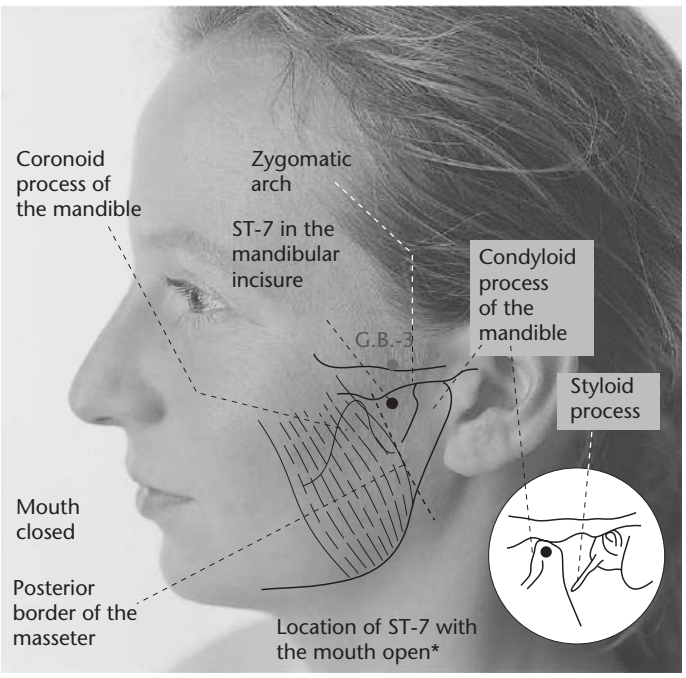

Location

With the mouth closed, at the lower border of the zygomatic arch (→ 3.1.2) in the centre of the depression of the mandibular fossa, between the coronoid process and the condyloid process of the mandible.

How to find

Palpate along the lower border of the zygomatic arch (→ 3.1.2) towards the ear. **ST-7** is located in a clearly palpable depression just before the temporomandibular joint and at the posterior border of the masseter muscle (clench teeth). To confirm: When the patient's mouth is fully opened, the condyloid process of the lower jaw slides anteriorly and the depression disappears (figure modified after Deadman et al).

→ **G.B.-3** is located directly superior to **ST-7**, on the upper border of the zygomatic arch.

Needling

Insert the needle with the patient's mouth closed. 0.3–0.5 cun vertically or 1–1.5 cun transversely (subcutaneously) in the direction of → **ST-6**, → **S.I.-19**, → **S.I.-18**.

Actions/Indications

- Opens the channel, benefits the jaw and teeth, alleviates pain
- Benefits the ears

Special features

Meeting point with the G.B. channel. Important local point and frequently used trigger point.

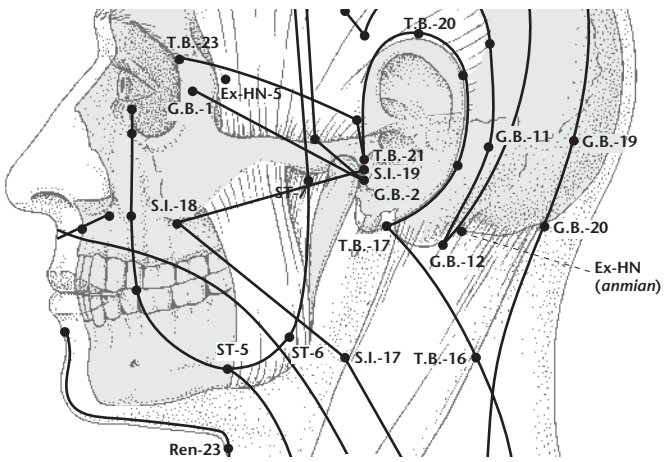

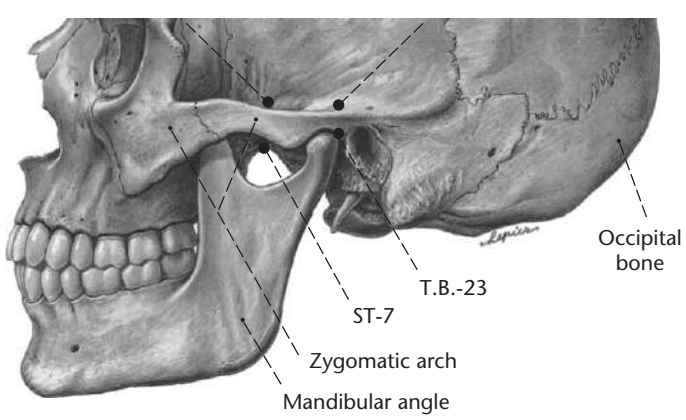

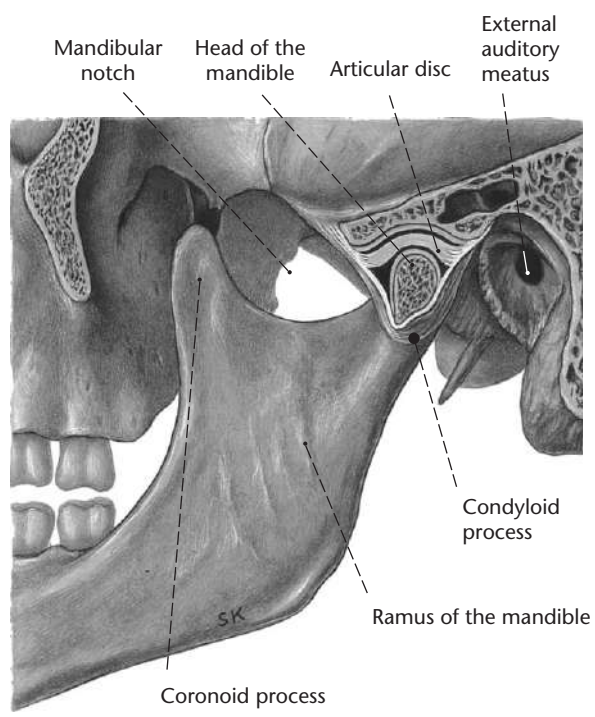

* According to Deadman et al 1998.

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 136

ST-8 *Head's Binding* TOUWEI

Location

At the temporal corner of the forehead, on the border of the temporalis muscle and 0.5 cun within the anterior hairline or 4.5 cun lateral to the anterior midline (→ **Du-24**).

How to find

First, find the anterior hairline (→ 3.1.1). Then locate **ST-8** at the temporal corner of the forehead 0.5 cun within the hairline. Chewing movements of the temporalis muscle will still be palpable here.

Also located 0.5 cun superior to the anterior hairline are → **Du-24/BL-3/BL-4/G.B.-15/G.B.-13** (anterior midline/superior to the inner canthus of the eye/1.5 cun lateral to the anterior midline/3 cun lateral to the anterior midline). These distances refer to the distance between → **Du-24** and **ST-8** (=4.5 cun, → 2.2). **ST-8** and → **G.B.-7** (a depression on the level of the apex of the ear, within the temporal hairline) mark the endpoints of a curved line, on which **ST-8 and → G.B.-4–7** are located at regular intervals (by dividing the line into five parts).

Needling

0.5–1 cun transversely (subcutaneously) or obliquely in the direction of the disorder. Moxibustion possible (contraindicated according to some classical texts).

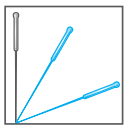

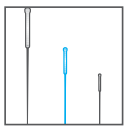

Actions/Indications

- Eliminates Wind from the head and eyes, benefits the eyes, alleviates pain

Special features

Meeting point with the *yang wei mai* and the G.B. channel. Important local point for headaches and disorders of the eyes.

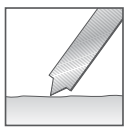

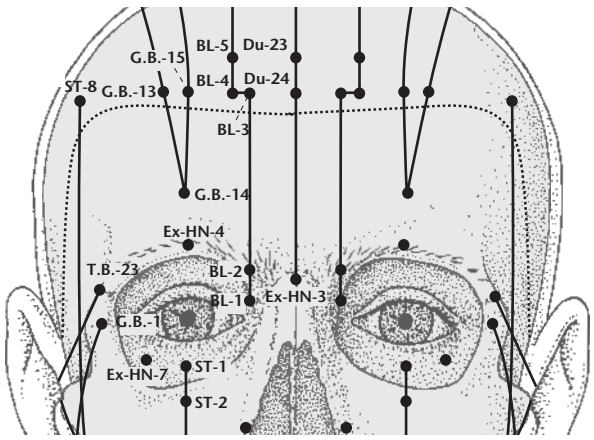

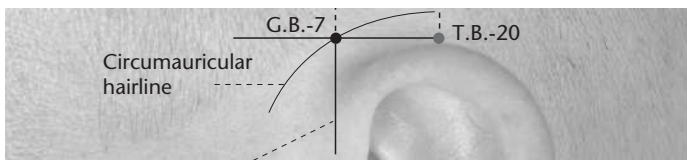

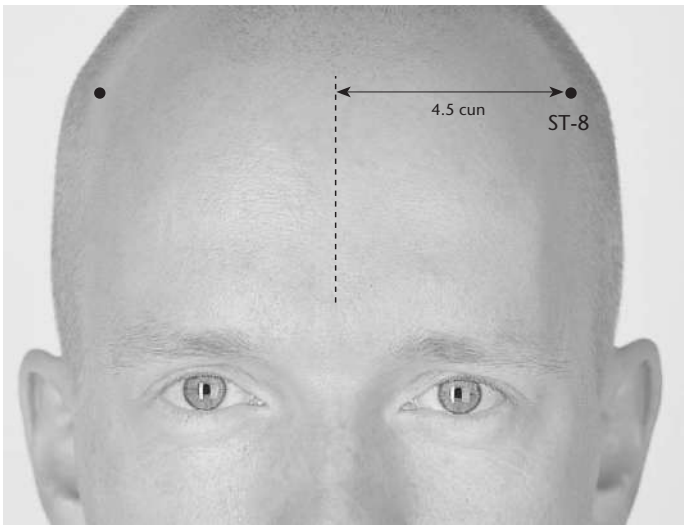

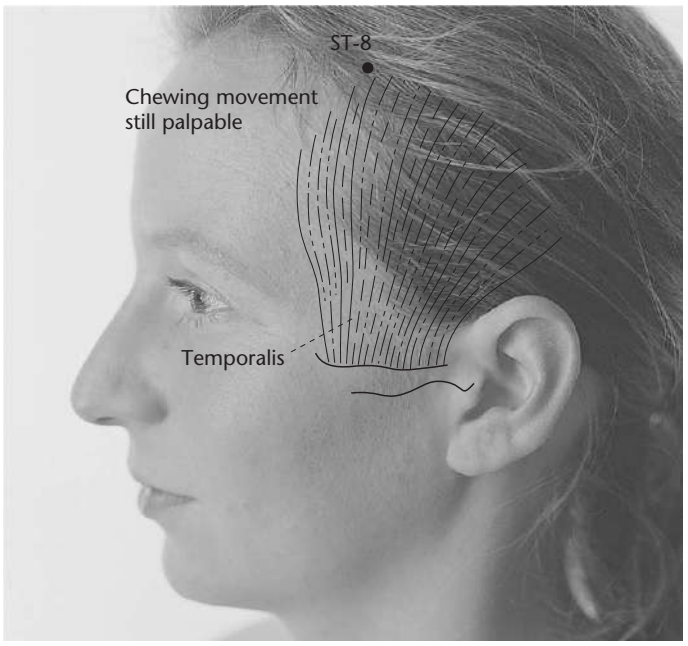

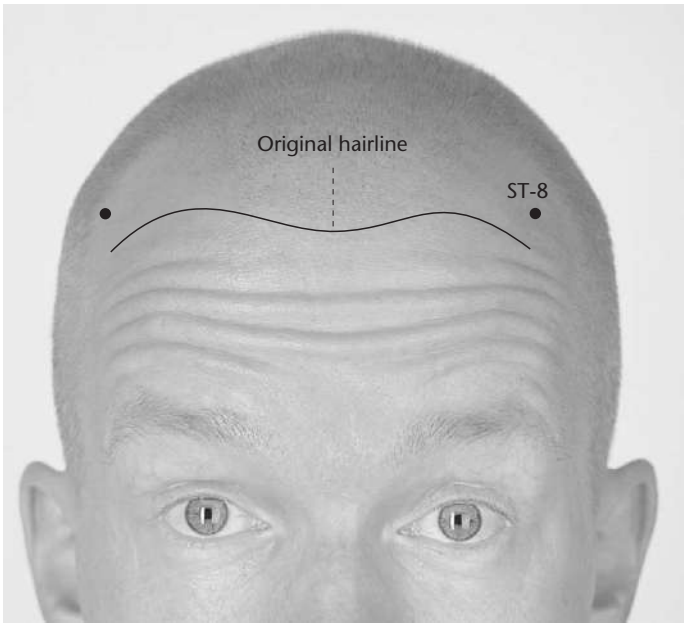

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 137

Man's Welcome RENYING ST-9

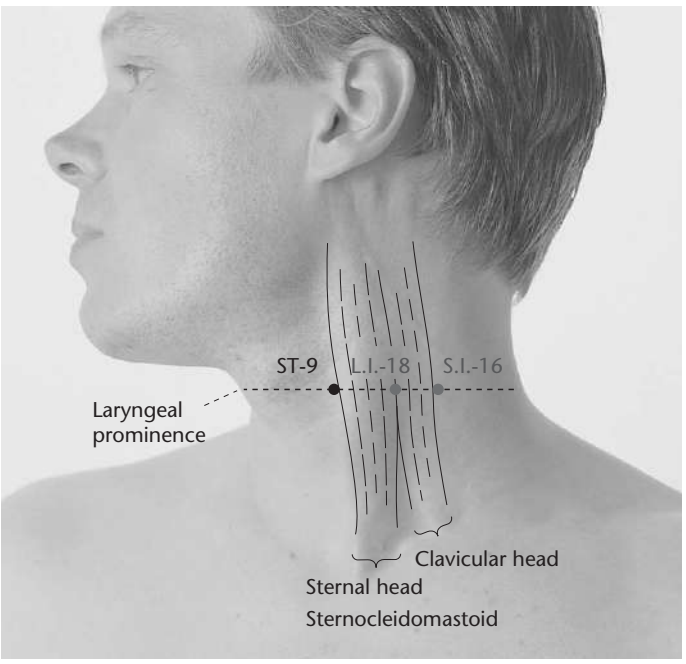

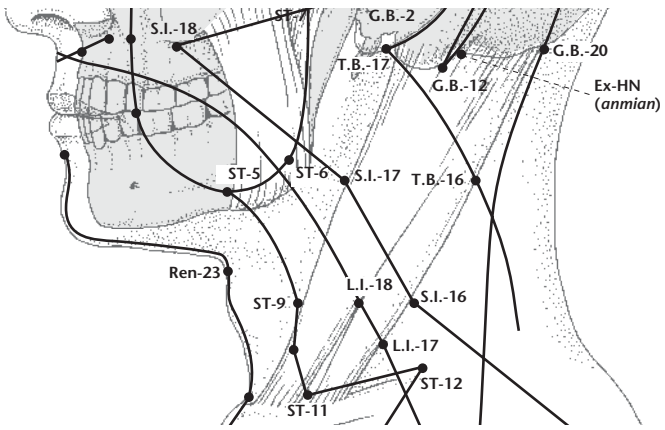

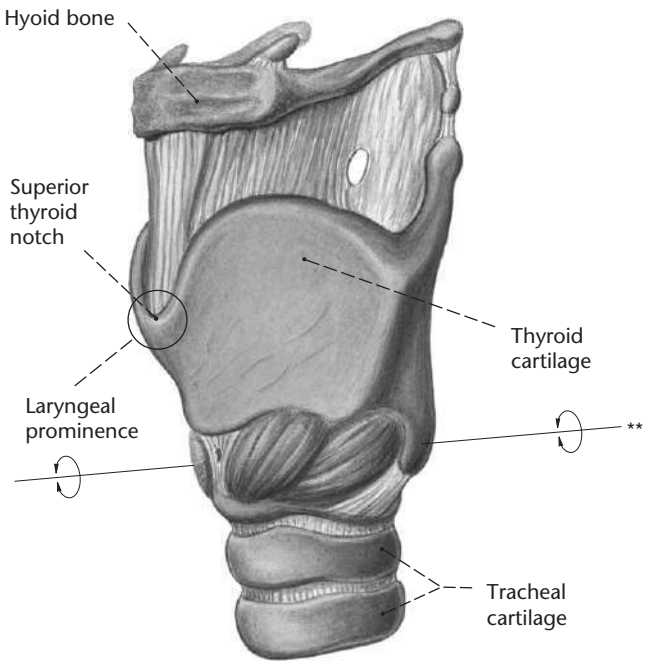

Location

1.5 cun lateral to the anterior midline, on the level of the laryngeal prominence and at the anterior border of the sternocleidomastoid muscle.

How to find

From the laryngeal prominence, palpate posteriorly to the anterior border of the sternocleidomastoid muscle (→ 3.2). **ST-9** is located at the anterior border of the muscle, which becomes more clearly visible and palpable if the patient's head is turned in the opposite direction.

Located on the same level are → **S.I.-16** on the posterior border of the sternocleidomastoid muscle and **L.I.-18** between the two heads of the sternocleidomastoid muscle.

Needling

If necessary, move the head to its normal position before needling. Caution: Carotid artery, carotid sinus, jugular veins. The carotid artery is located directly at the anterior border of the sternocleidomastoid muscle. The needle should be inserted anterior to where you can feel the pulsating of the carotid artery. Vertically 0.5–1 cun between the artery and the thyroid cartilage. For reasons of safety, the artery should be held in a lateral position with one finger of the palpating hand above and one finger below the insertion site. Irritation of the carotid sinus can lead to a lowering of the blood pressure and fainting!

Actions/Indications

- Regulates the Qi
- Opens the channel
- Harmonises excess or deficiency in the thorax

Special features

Window of Heaven point, Sea of Qi point, meeting point with the G.B. channel.

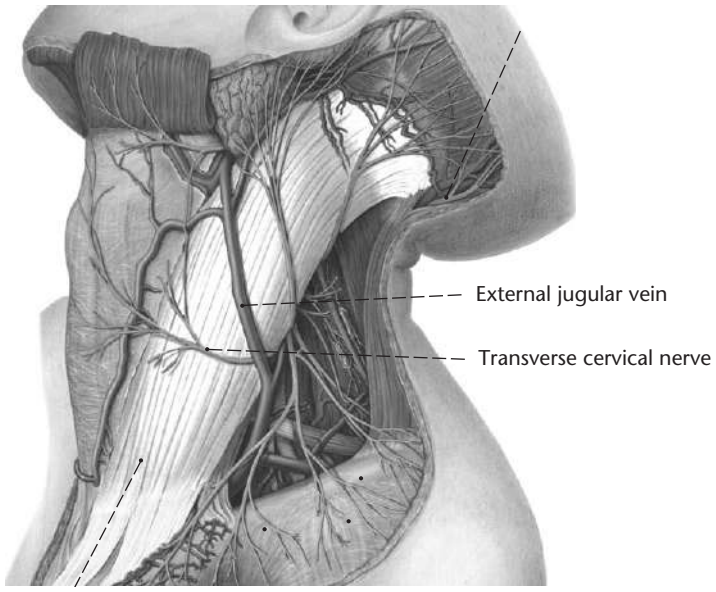

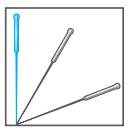
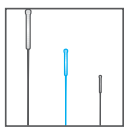

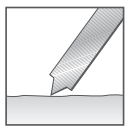

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 138

ST-10 *Water Prominence* SHUITU

Location

On the anterior border of the sternocleidomastoid muscle, at the midpoint of a line connecting → **ST-9** and **ST-11**.

How to find

→ **ST-9** and **ST-10** are both located on the anterior border of the sternocleidomastoid muscle (→ 3.2). → **ST-11** is located posterior to the sternal head. Place the middle finger of the left hand onto → **ST-9** on the level of the laryngeal prominence and the thumb on → **ST-11** posterior to the sternal insertion. **ST-10** is located at the midpoint between the two fingers. With the patient's head rotated in the opposite direction, the muscle will be more visible and can be more easily palpated.

Needling

0.5–1 cun obliquely in a medial direction, away from the artery. Caution: Carotid artery, jugular veins. If necessary, rotate the head back to a middle position before needling.

Actions/Indications

- Regulates the Lung Qi
- Benefits the throat

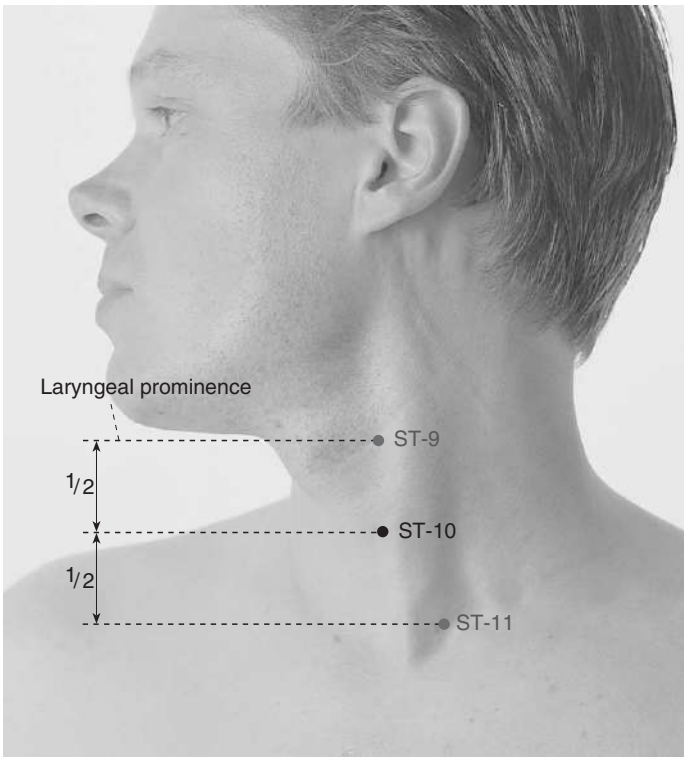

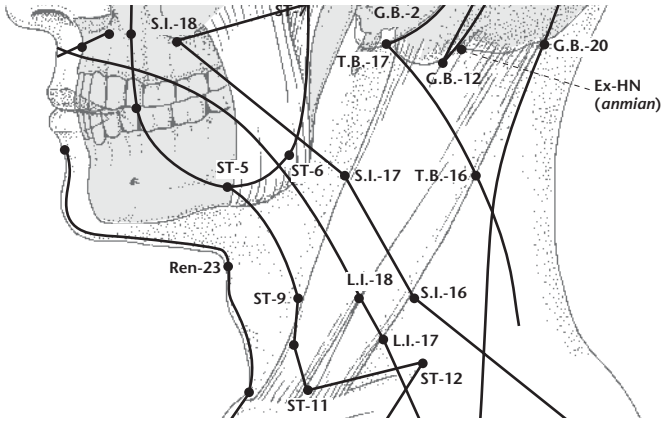

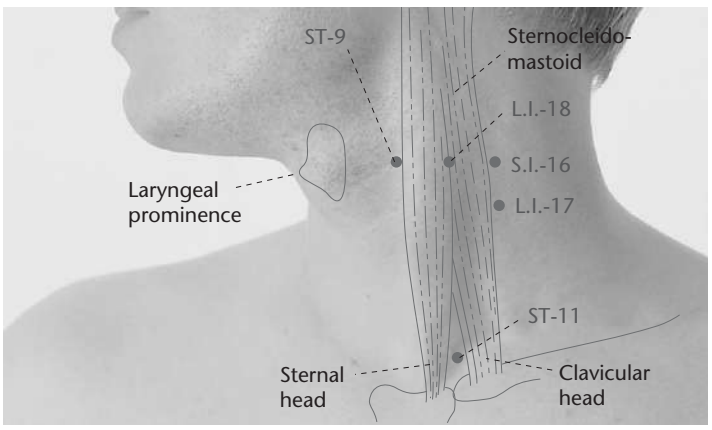

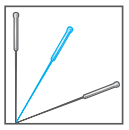
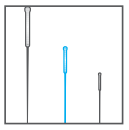

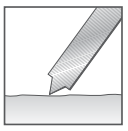

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 139

Qi Abode QISHE ST-11

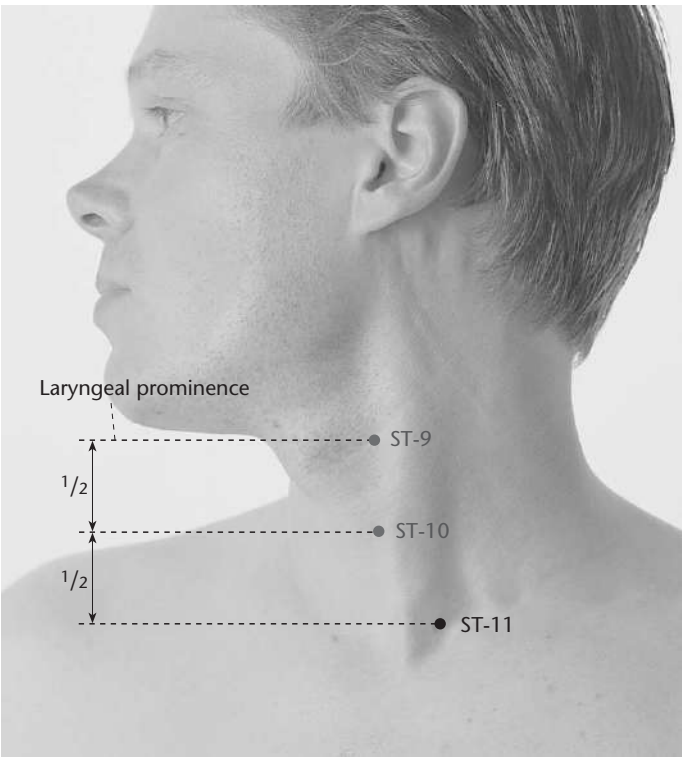

Location

On the upper border of the clavicle, between the tendons of the sternal and clavicular heads of the sternocleidomastoid muscle.

How to find

By rotating the patient's head in the opposite direction (especially against resistance), the ropey sternal insertion of the sternocleidomastoid muscle (→ 3.2) becomes clearly visible. **ST-11** is located in a shallow depression, the minor supraclavicular fossa, between the sternal insertion and the more extended insertion of the clavicular head.

Needling

Vertically 0.3–0.5 cun. Caution: Pneumothorax, jugular veins.

Actions/Indications

- Benefits the throat and neck
- Regulates counterflow Qi

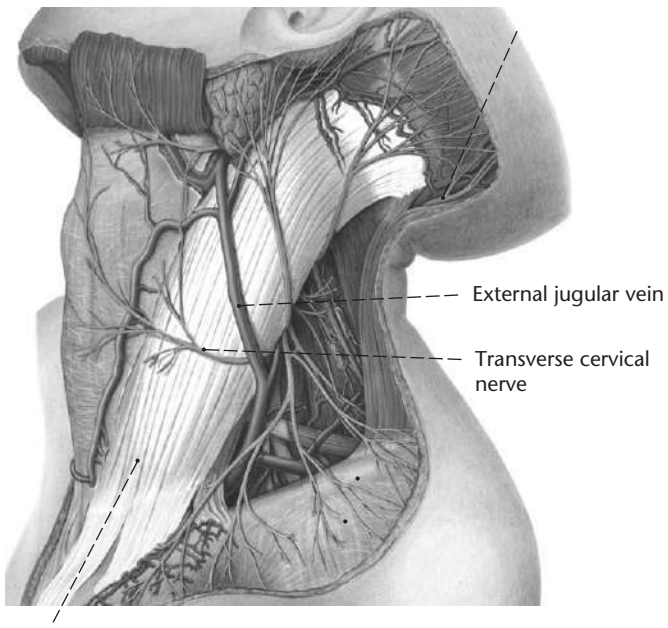

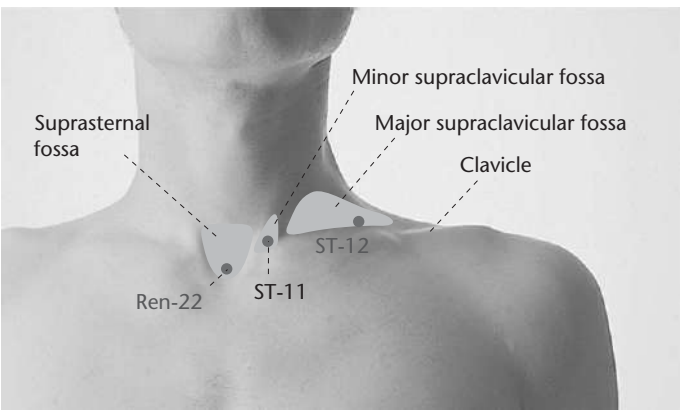

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 140

ST-12 *Empty Basin* QUEPEN

Location

In the supraclavicular fossa, superior to the midpoint of the clavicle, approximately 4 cun lateral to the anterior midline.

How to find

From the anterior midline, palpate to the midpoint of the clavicle (=4 cun in a lateral direction). There, an imaginary vertical line marks the midclavicular line. **ST-12** is located on the midclavicular line, at the upper border of the clavicle, in the supraclavicular fossa, lateral to the clavicular head of the sternocleidomastoid muscle.

Needling

Vertically 0.3–0.5 cun. The needle should be guided along the posterior border of the clavicle. Caution: Pneumothorax, especially with emphysema; superficial and deep cervical arteries. According to some authors, this point is contraindicated during pregnancy.

Actions/Indications

- Opens the channel
- Disperses and descends the Lung Qi
- Clears Heat in the throat

Special features

Meeting point with the L.I., S.I., G.B. and T.B. channel

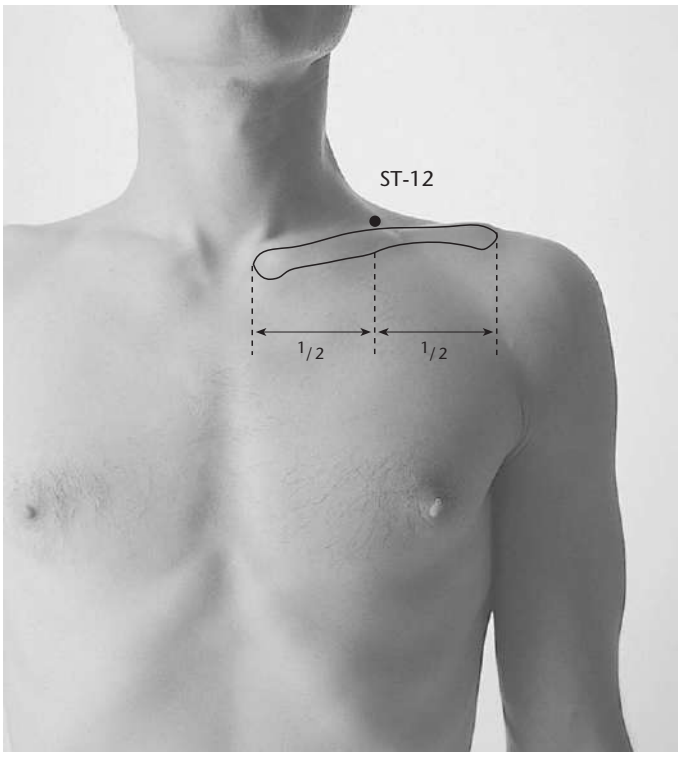

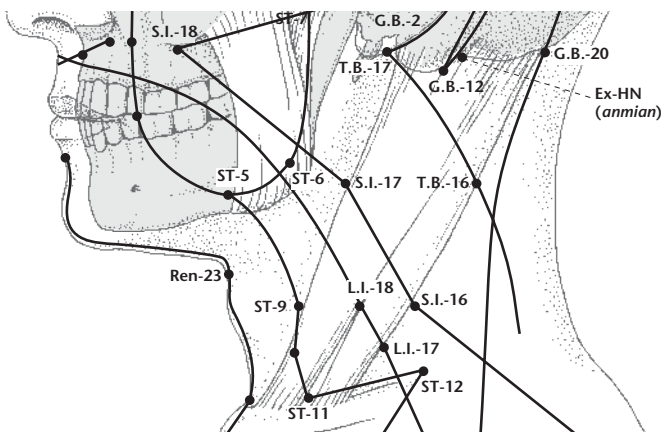

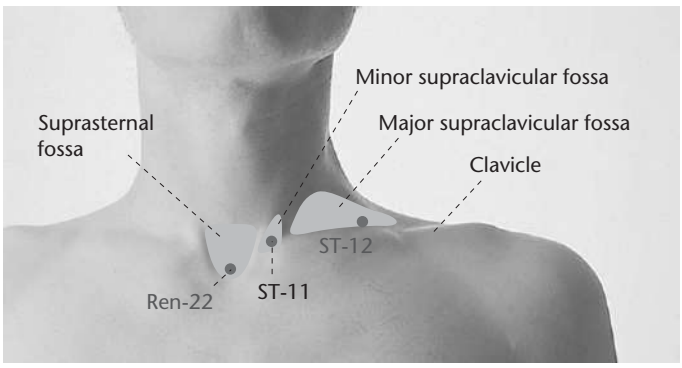

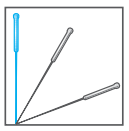
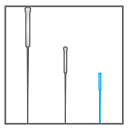

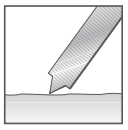

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 141

Qi Door QIHU ST-13

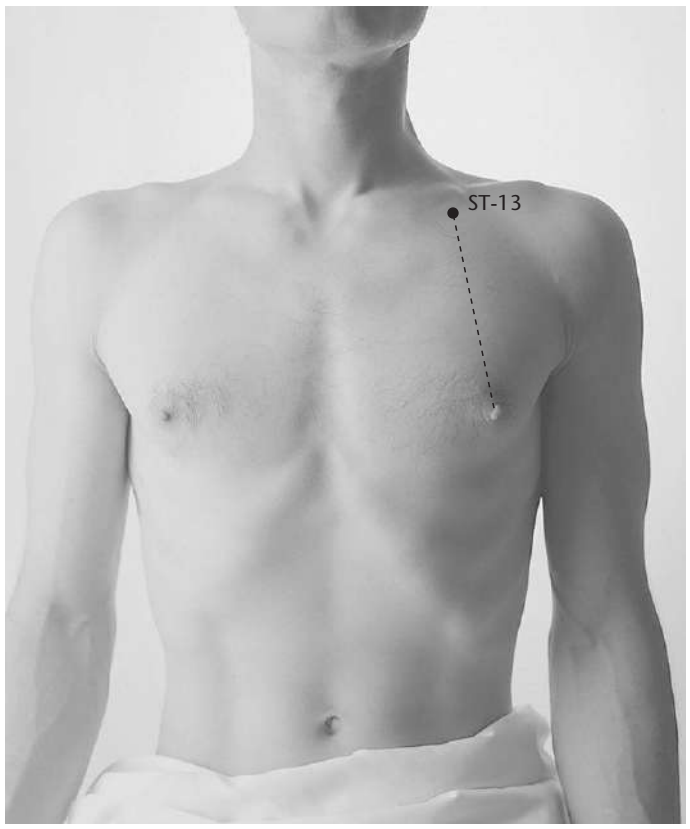

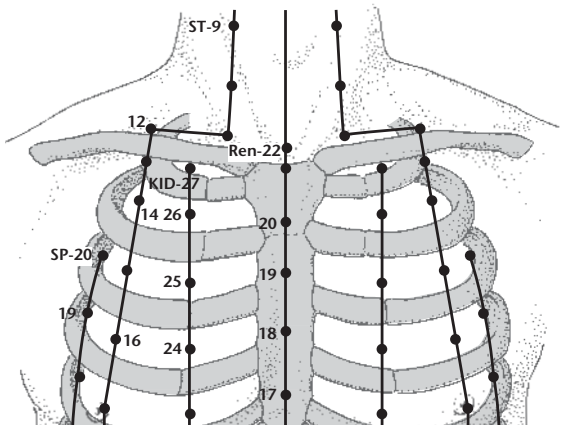

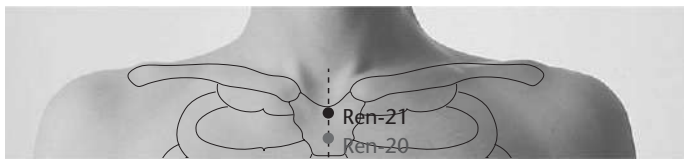

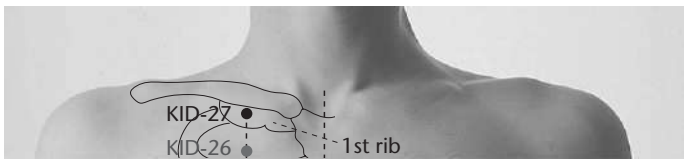

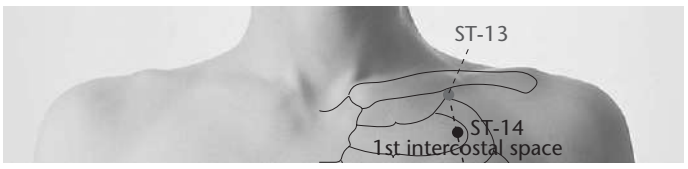

Location

At the midpoint of the clavicle and on its inferior border, 4 cun lateral to the anterior midline.

How to find

On the upper thorax region, the midclavicular line (4 cun lateral to the anterior midline) runs slightly obliquely from the clavicular midpoint to the generally more lateral nipple. Find the midpoint of the clavicle (or measure 4 cun lateral to the midline) and locate **ST-13** at the lower border of the clavicle. Since the 1st rib curves posteriorly in a short bow, **ST-13** is generally situated lateral to the 1st rib.

→ **KID-27** is located medial to **ST-13**, but inferior to the lower border of the clavicle and 2 cun lateral to the anterior midline. → **Ren-21** is located on the midline. → **ST-14** is located only slightly inferior to **ST-13** in the 1st intercostal space.

Needling

0.3–0.5 cun obliquely in a lateral or medial direction or transversely (subcutaneously) along the channel pathway. Caution: Pneumothorax, subclavicular artery/vein.

Actions/Indications

- Descends counterflow Qi
- Clears Heat
- Unbinds the chest

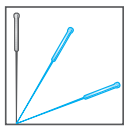
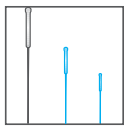
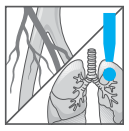

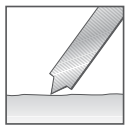

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 142

ST-14 *Store House* KUFANG

Location

On the midclavicular line, in the first intercostal space, 4 cun lateral to the anterior midline.

How to find

On the upper thorax, the midclavicular line (4 cun lateral to the anterior midline) runs slightly obliquely from the clavicular midpoint to the generally more lateral nipple. By palpating across the clavicle from superior to inferior, the first rib will generally be felt directly below the clavicle, although, in some cases, the 1st rib might be completely hidden under the clavicle. The 1st intercostal space follows directly inferior to the 1st rib, where **ST-14** is located.

→ **ST-13** is located only slightly superior to **ST-14** at the lower border of the clavicle. Also located on the level of the first intercostal space are → **Ren-20** (anterior midline), → **KID-26** (2 cun lateral to the anterior midline) and → **LU-1** (6 cun lateral to the midline). For more detail about orientation regarding intercostal spaces → 3.5.

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), parallel to the course of the intercostal space or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Regulates the Qi
- Unbinds the chest
- Benefits the breasts

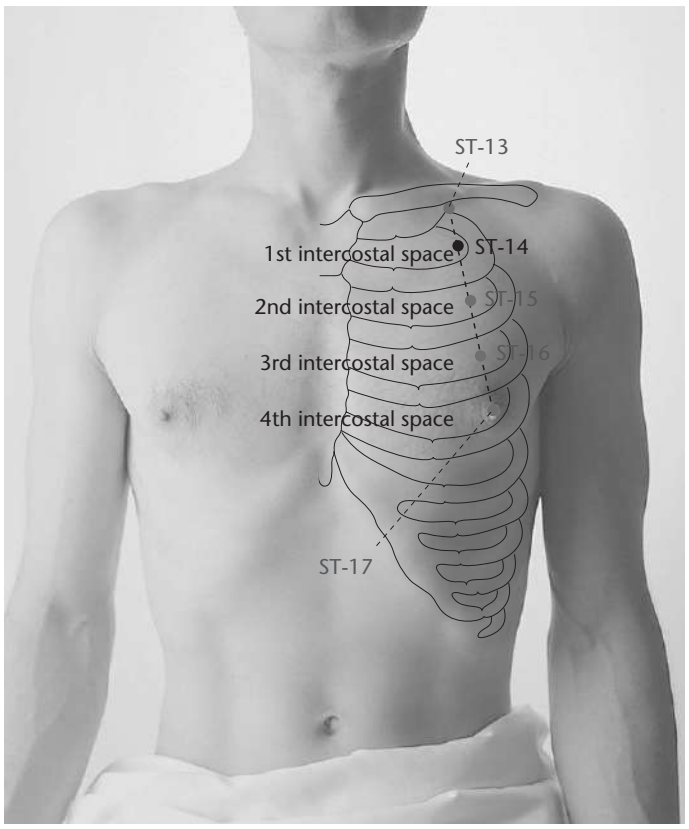

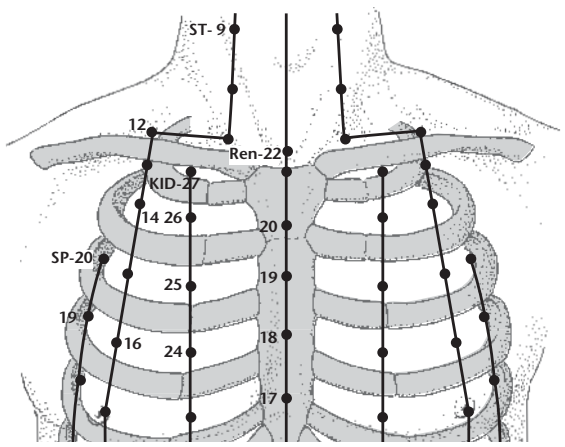

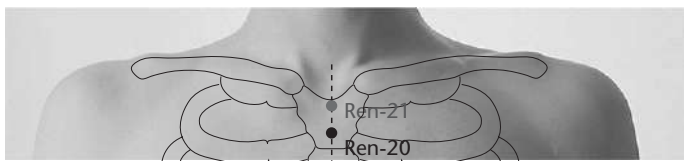

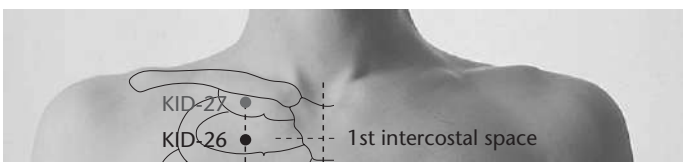

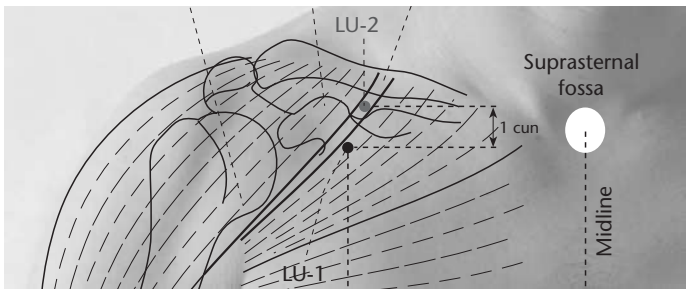

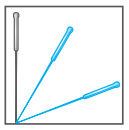
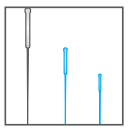
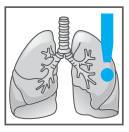

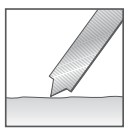

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 143

Room Screen WUYI ST-15

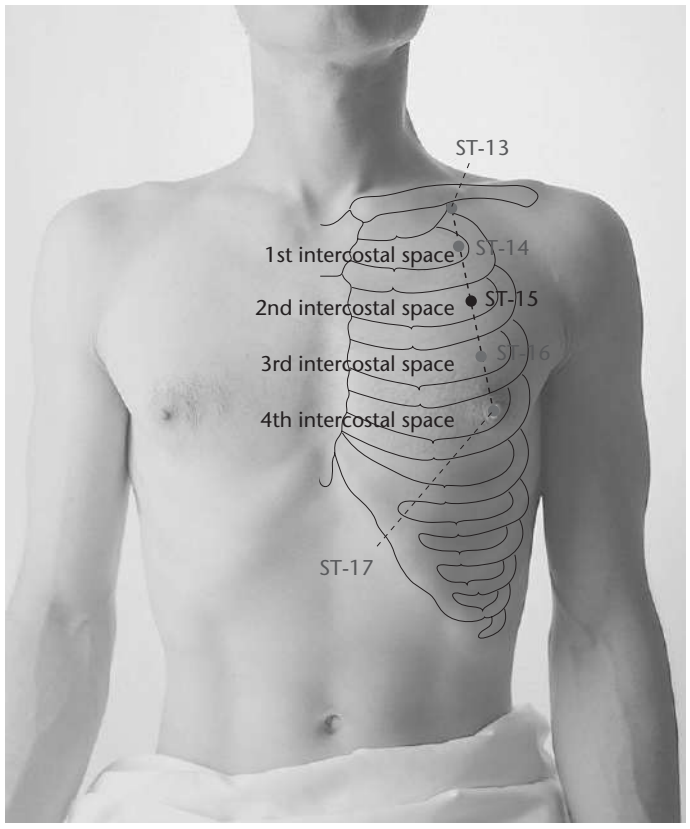

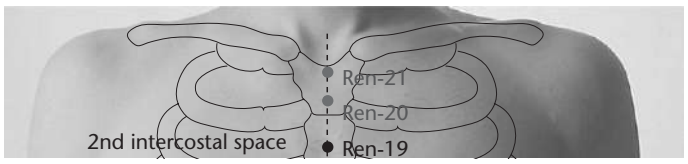

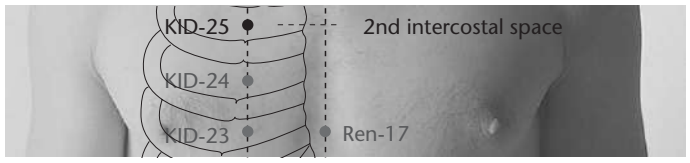

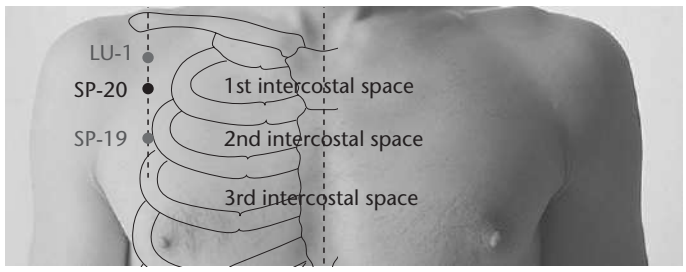

Location

In the 2nd intercostal space, on the midclavicular line, 4 cun lateral to the anterior midline.

How to find

On the upper thorax, the midclavicular line (4 cun lateral to the anterior midline) runs slightly obliquely from the clavicular midpoint to the generally more lateral nipple. Locate the 2nd intercostal space by palpating downward from the clavicle or by palpating from the manubriosternal synchondrosis (second rib) (→ 3.5). Then palpate along its course to the midclavicular line, where **ST-15** is located. Note the ascending course of the intercostal space! Located on the same level are → **Ren-19** (anterior midline), → **KID-25** (2 cun lateral to the anterior midline) and → **SP-20** (6 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), parallel to the course of the intercostal space or with or against the channel pathway. Caution: pneumothorax.

Actions/Indications

- Regulates the Lung Qi
- Unbinds the chest
- Benefits the breasts and clears Heat

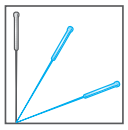
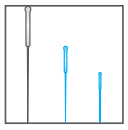
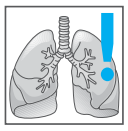

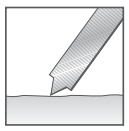

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 144

ST-16 *Breast Window* YINGCHUANG

Location

In the 3rd intercostal space, on the midclavicular line, 4 cun lateral to the anterior midline.

How to find

On the upper thorax, the midclavicular line (4 cun lateral to the anterior midline) runs slightly obliquely from the clavicular midpoint to the generally more lateral nipple. Locate the 3rd intercostal space by palpating downward from the clavicle or by palpating from the manubriosternal synchondrosis (second rib) (→ 3.5). Then palpate along its course to the midclavicular line, where **ST-16** is located. Note the ascending course of the intercostal space! In male patients, this point can also be located by counting one intercostal space upward from the level of the nipple.

Located on the same level are → **Ren-18** (anterior midline), → **KID-24** (2 cun lateral to the anterior midline) and → **SP-19** (6 cun lateral to the midline).

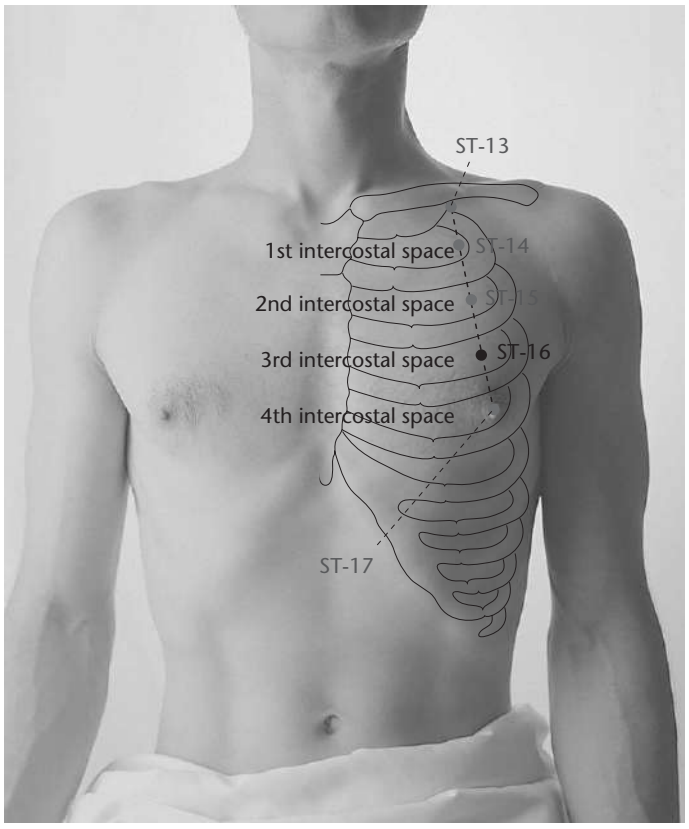

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), along the course of the intercostal space or with or against the flow of the channel. Caution: pneumothorax.

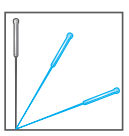

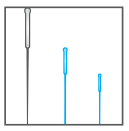

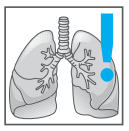

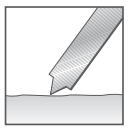

Actions/Indications

- Regulates the Qi
- Unbinds the chest
- Benefits the breasts

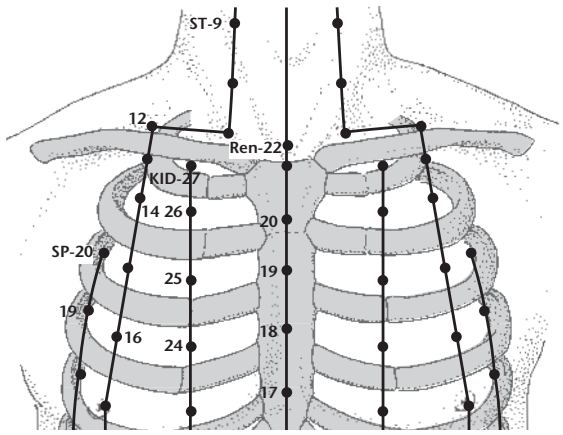

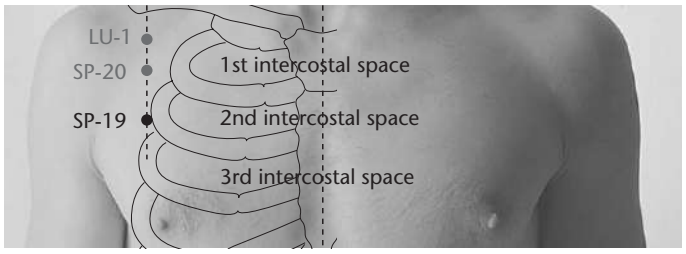

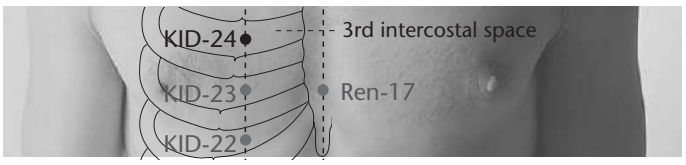

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 145

Breast Centre RUZHONG ST-17

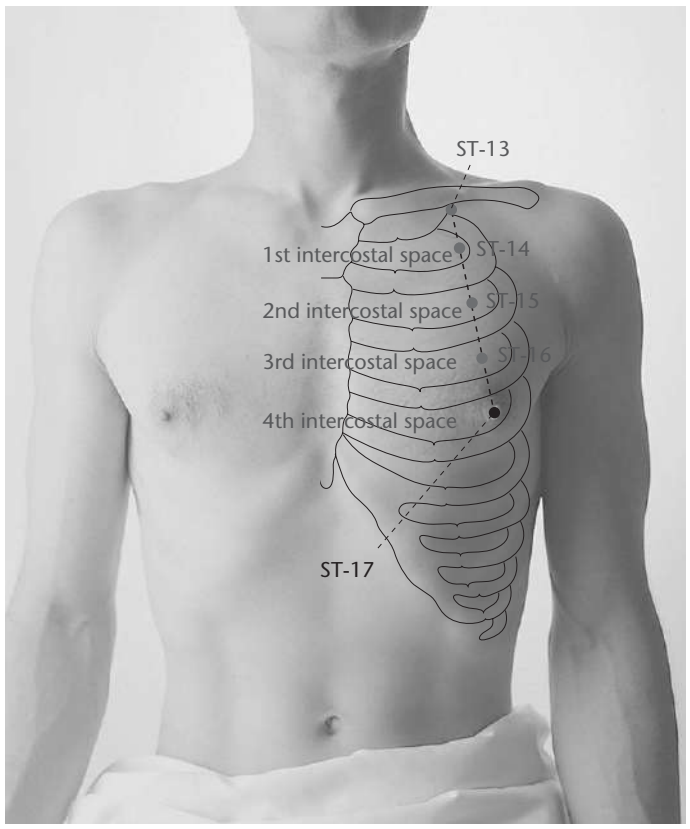

Location

In the centre of the nipple.

How to find

In the centre of the nipple. In men, this is usually located 4 cun lateral to the anterior midline, at the level of the 4th intercostal space and is also used for reference.

In men, also located on the same level are → **Ren-17** (anterior midline), → **KID-23** (2 cun lateral to the anterior midline), → **SP-18** (6 cun lateral to the midline), → **P-1** (1 cun lateral to the nipple), → **G.B.-22** and → **G.B.-23**. In women, the location of the nipples varies considerably and should therefore not be used for reference.

Needling

No therapy; this point is for reference only.

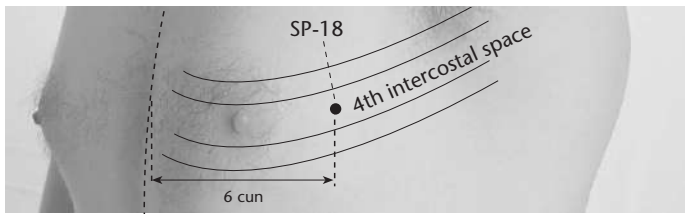

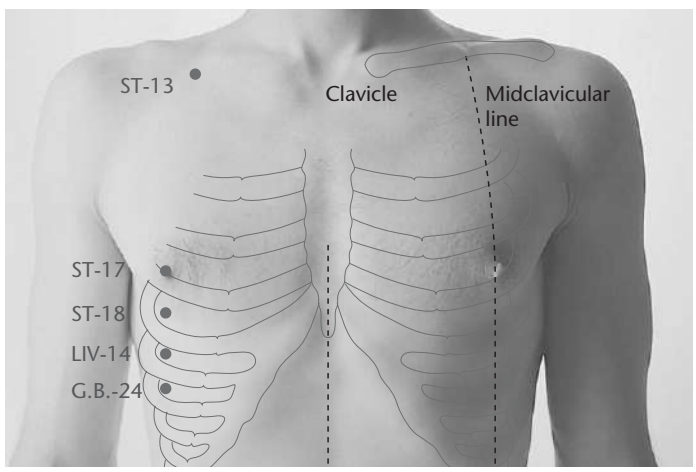

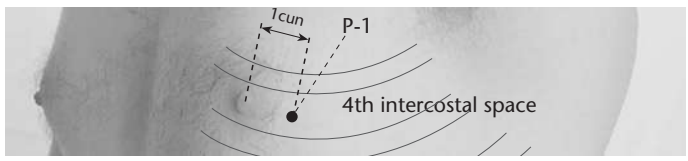

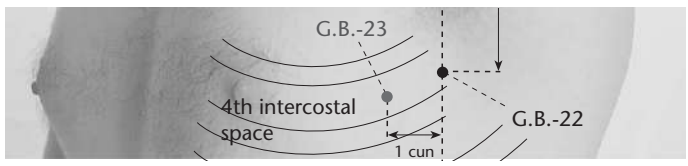

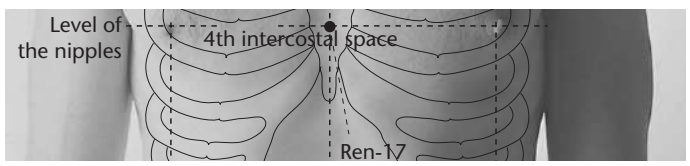

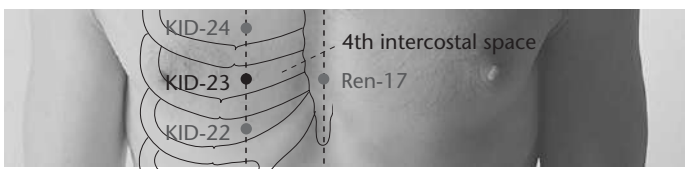

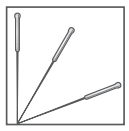
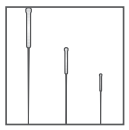

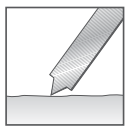

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 146

ST-18 *Breast Root* RUGEN

Location

In the 5th intercostal space, on the mamillary line, 4 cun lateral to the anterior midline.

How to find

The mamillary line, a vertical line used for reference, is located 4 cun lateral to the anterior midline (→ 3.5). In men, the nipple is generally located on the level of the 4th intercostal space, and **ST-18** is located in the intercostal space directly below. The point is often located at the lower border of the pectoralis major muscle; in women, it is often found at the root of the breast. Location help in women: Find the 5th intercostal space by palpating downward from the clavicle or from the manubriosternal synchondrosis (second rib) (→ 3.5). Then follow the course of the intercostal space laterally to the mamillary line and there locate **ST-18**. Note the ascending course of the intercostal spaces!

Located on the same level are → **Ren-16** (anterior midline), → **KID-22** (2 cun lateral to the anterior midline) and → **SP-17** (6 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), parallel to the course of the intercostal space or with or against the channel pathway. Caution: Pneumothorax.

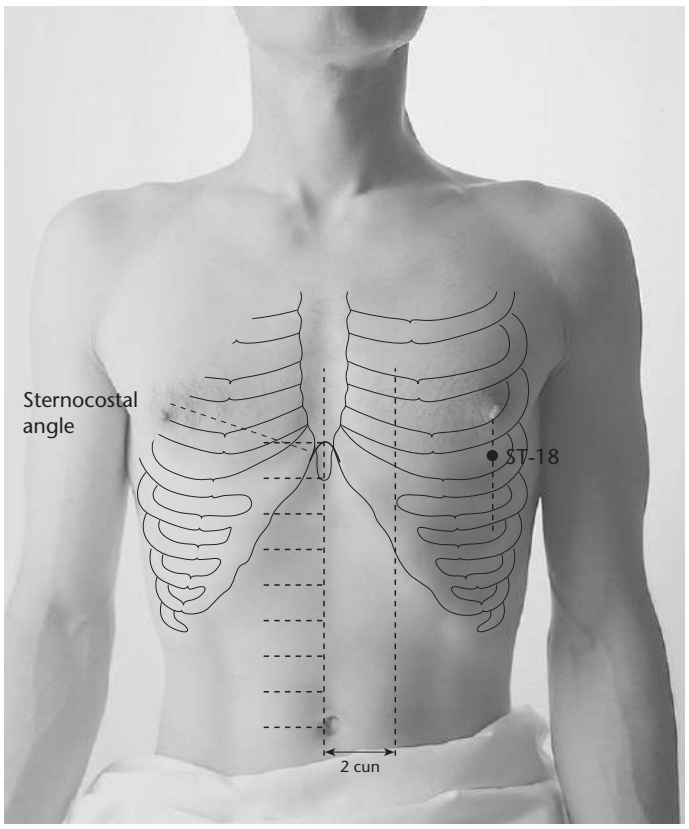

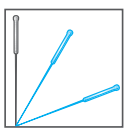

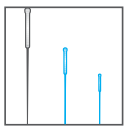

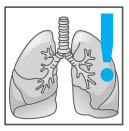

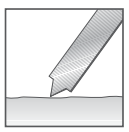

Actions/Indications

- Benefits the breasts, moves the Blood
- Unbinds the chest
- Regulates the Lung Qi

Special features

Important local point

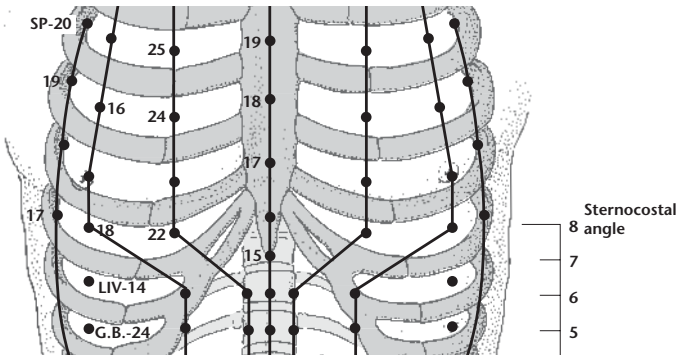

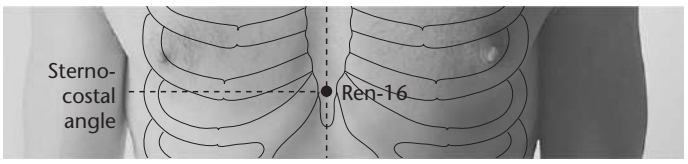

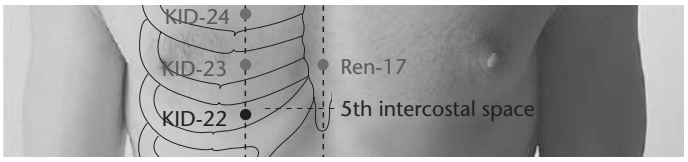

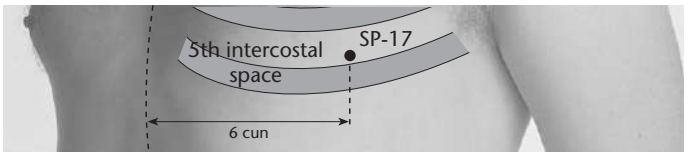

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 147

Not Contained BURONG ST-19

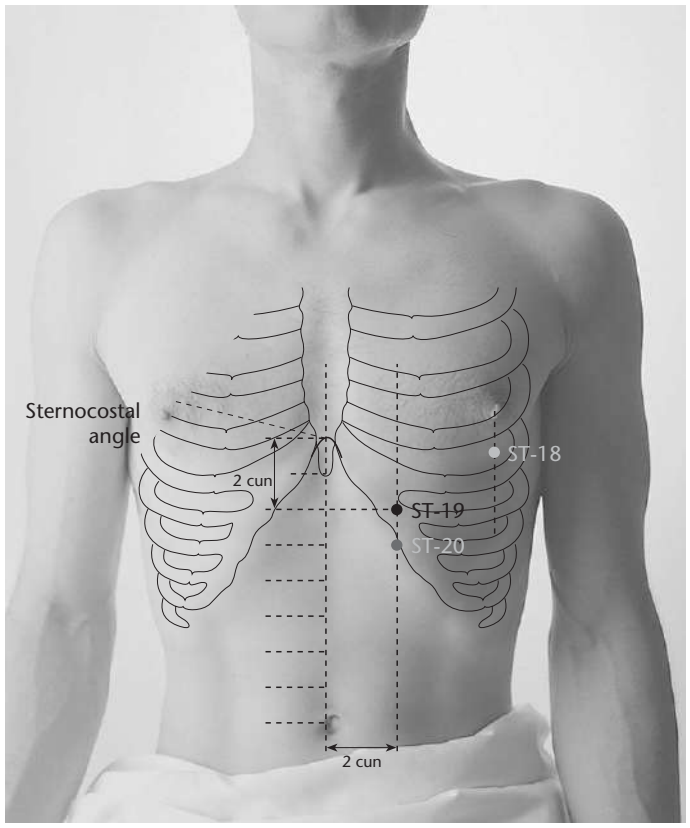

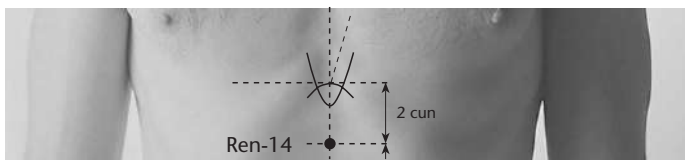

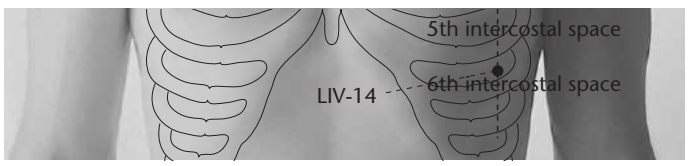

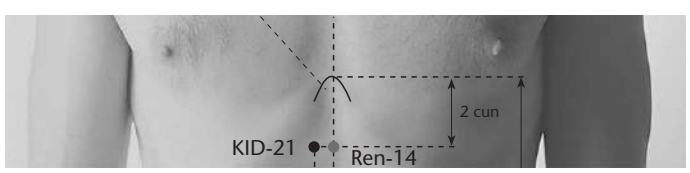

Location

2 cun below the sternocostal angle and 2 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape → 2.3.1). First, measure 2 cun down from the sternocostal angle. **Or:** Measure the midpoint of the upper half of the distance between the umbilicus and the subcostal angle (→ **Ren-12**: spreading hands technique → 2.3.3). From there, measure 2 cun in a lateral direction to locate **ST-19**. As its location depends on the shape of the thorax, it might be located either on the rib or on the abdomen. Located on the same level are → **Ren-14** (anterior midline), → **KID-21** (0.5 cun lateral to the anterior midline) and → **LIV-14** (on the mamillary line and in the 6th intercostal space). Please note: From **ST-19** onwards, the ST channel is located only 2 cun lateral to the anterior midline.

Needling

Vertically 0.5–1 cun. If the correct location projects this point on the lower ribcage, needle transversely on the rib or locate the point more medially (or choose an alternative point) Caution: Peritoneum, pericardium.

Actions/Indications

- Regulates the Stomach, Intestines and Middle Burner
- Regulates the Lung Qi

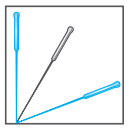
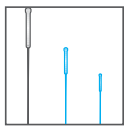
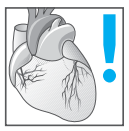

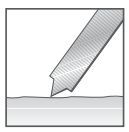

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 148

ST-20 *Supporting Fullness* CHENGMAN

Location

3 cun below the sternocostal angle (or 5 cun above the umbilicus) and 2 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape ➞ 2.3.1). First, measure 3 cun down from the sternocostal angle. **Or:** Measure 1 cun up from the midpoint of the distance between the umbilicus and the sternocostal angle (➞ **Ren-12**: spreading hands technique ➞ 2.3.3). From there, measure 2 cun in a lateral direction to locate **ST-20**. Depending on the shape of the thorax, **ST-20** might be located on the lower border of the ribcage in some cases (narrow subcostal angle).
Located on the same level are ➞ **Ren-13** (anterior midline) and ➞ **KID-20** (0.5 cun lateral to the anterior midline) and ➞ **G.B.-24** (on the mamillary line, in the 7th intercostal space).

Needling

Vertically 0.5–1 cun. If the correct location projects this point on the lower ribcage, needle transversely on the rib or locate the point more medially (or choose an alternative point). Caution: Peritoneum.

Actions/Indications

- Regulates the Middle Burner and the Stomach Qi
- Regulates the Lung Qi

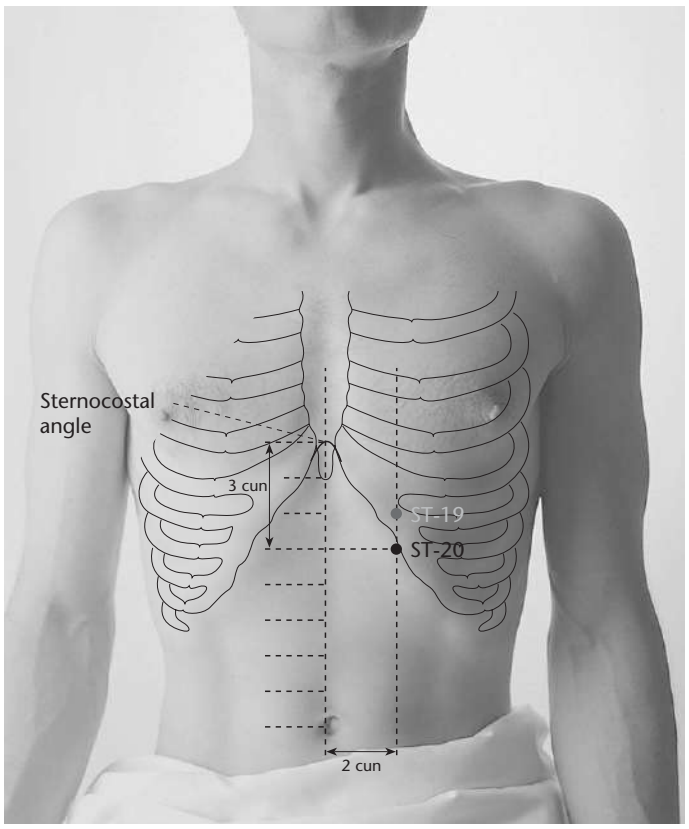

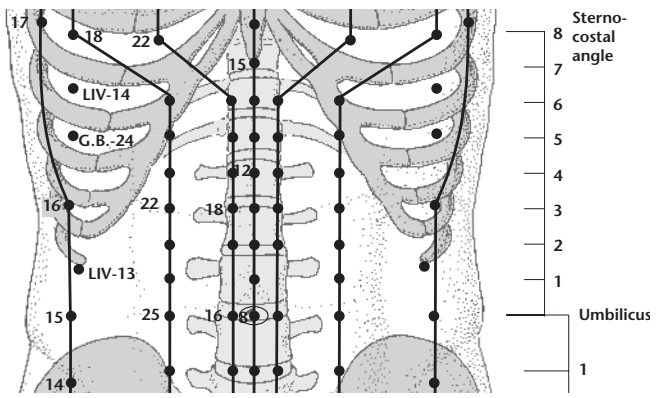

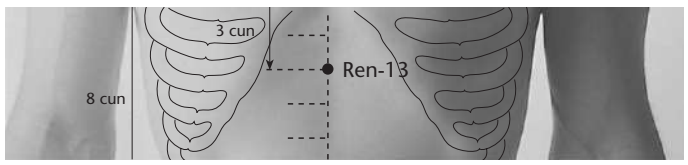

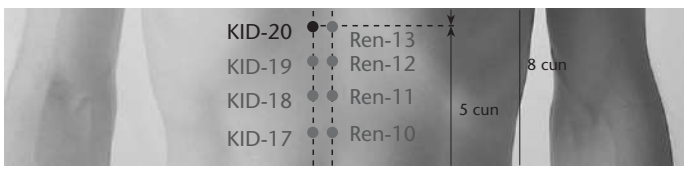

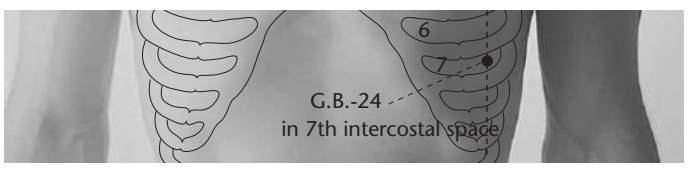

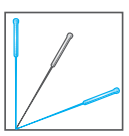
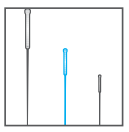
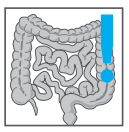

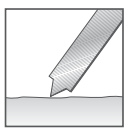

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 149

Beam Gate LIANGMEN ST-21

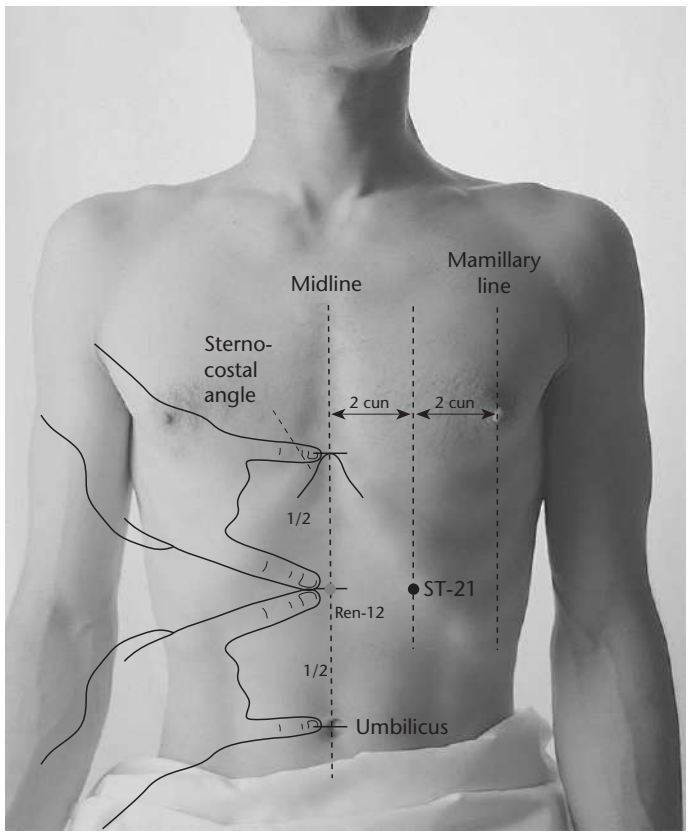

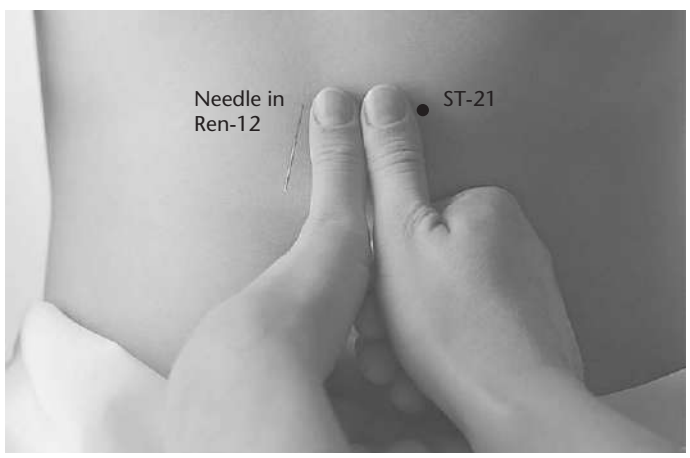

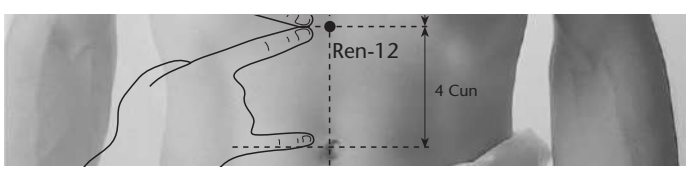

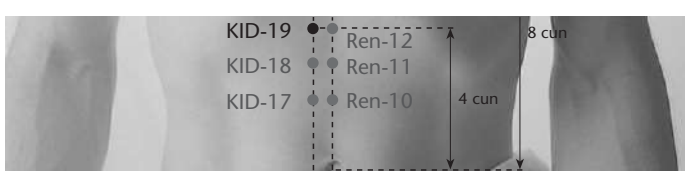

Location

2 cun lateral to the anterior midline, 4 cun above the umbilicus (or 4 cun below the subcostal angle).

How to find

The distance between the sternocostal angle and the centre of the umbilicus can be divided into 8 cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape → 2.3.1). **Quick method:** Find the midpoint between the umbilicus and the sternocostal angle (→ **Ren-12**) by using the spreading hands technique (→ 2.3.3), then measure 2 cun in a lateral direction to locate **ST-21**. Depending on the shape of the thorax, this can be located on the lower rib cage.
Located on the same level are → **Ren-12** (anterior midline) and → **KID-19** (0.5 cun lateral to the anterior midline).

Needling

Vertically 0.5–1 cun. If the correct location projects this point on the lower ribcage, needle transversely on the rib or locate the point more medially (or choose an alternative point). Caution: Peritoneum, enlarged liver (on the right), pregnancy.

Actions/Indications

- Regulates the Qi, harmonises the Middle Burner, eliminates stagnation, raises the Qi and stops diarrhoea

Special features

Important local point for distension and pain in the epigastrium and the lateral abdomen

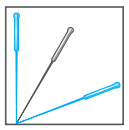
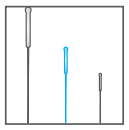
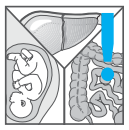

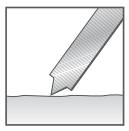

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 150

ST-22 *Pass Gate* GUANMEN

Location

3 cun above the umbilicus (or 5 cun inferior to the sternocostal angle) and 2 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle and the centre of the umbilicus can be divided into 8 cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape → 2.3.1). First, measure 3 cun up from the umbilicus. **Or:** Measure 1 cun downward from the midpoint of the distance between the umbilicus and the sternocostal angle (→ **Ren-12**: spreading hands technique → 2.3.3). **ST-22** is located 2 cun lateral from here.
Located on the same level are → **Ren-11** (anterior midline), → **KID-18** (0.5 cun lateral to the anterior midline) and → **SP-16** (4 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Regulates the Stomach and Intestines
- Regulates the water passages

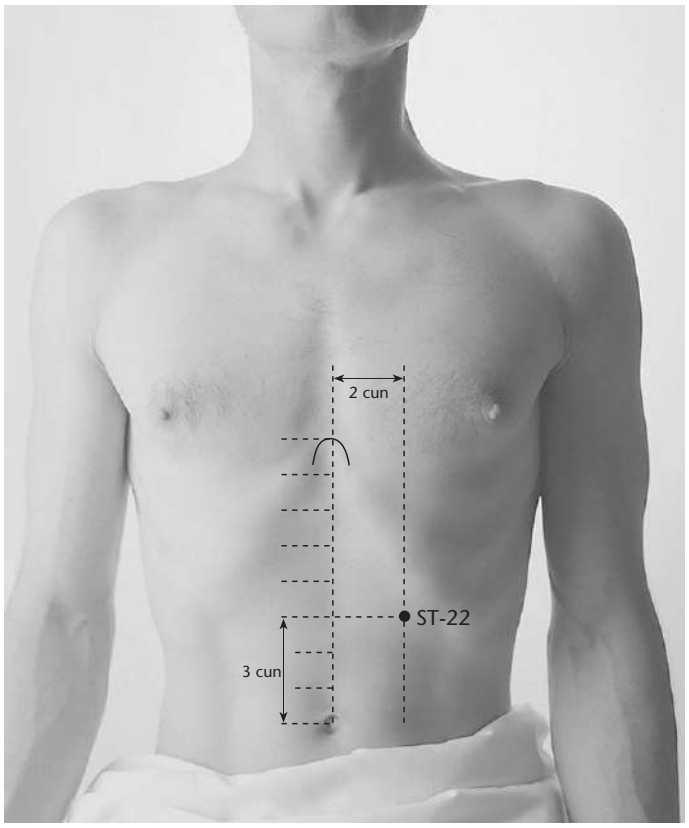

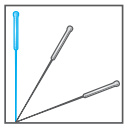

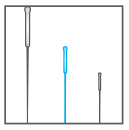

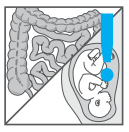

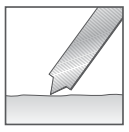

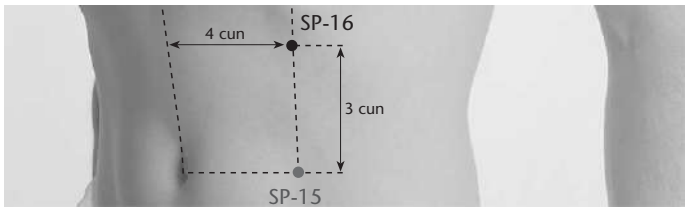

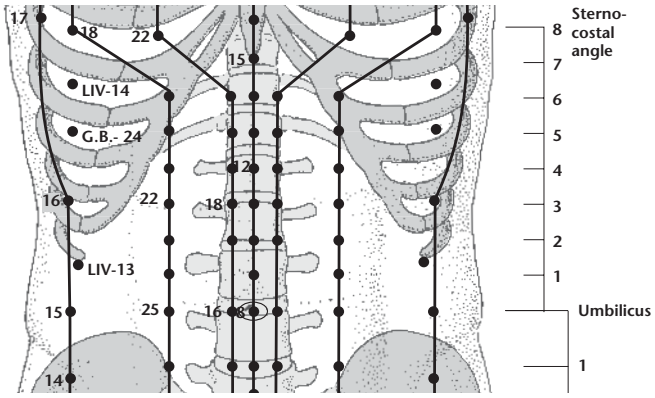

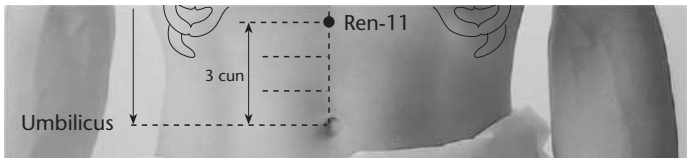

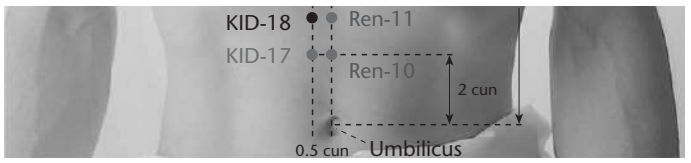

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 151

Supreme Unity TAIYI ST-23

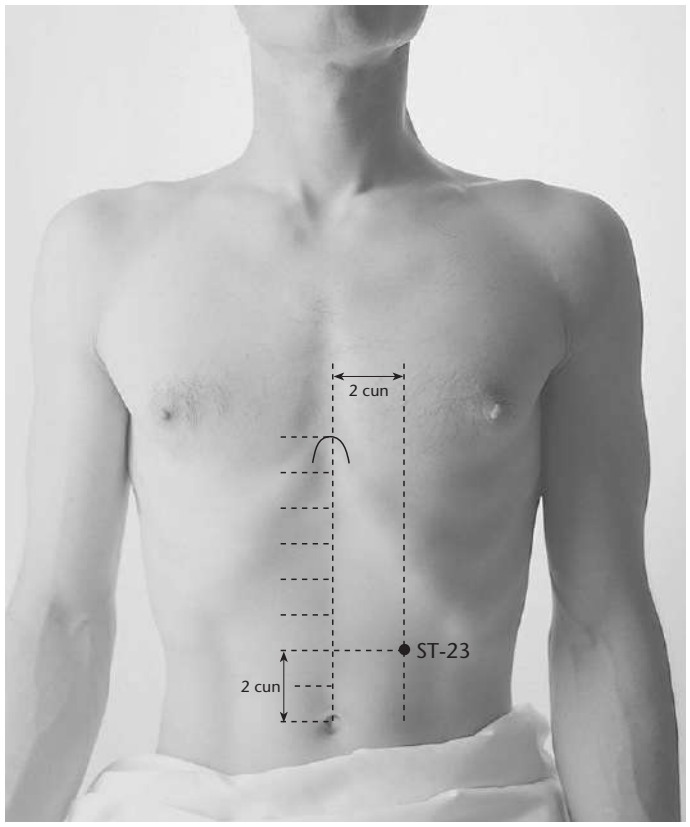

Location

2 cun above the umbilicus and 2 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape → 2.3.1). First, measure 2 cun up from the umbilicus. **Or:** Measure the midpoint of the lower half of the distance between the umbilicus and the subcostal angle (→ **Ren-12**: spreading hands technique → 2.3.3). **ST-23** is located 2 cun lateral from here.
Located on the same level are → **Ren-10** (anterior midline), → **KID-17** (0.5 cun lateral to the anterior midline) and → **Ex-CA** (*weishang*; 4 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Harmonises the Middle Burner, strengthens the Spleen
- Benefits the Heart and calms the *shen*

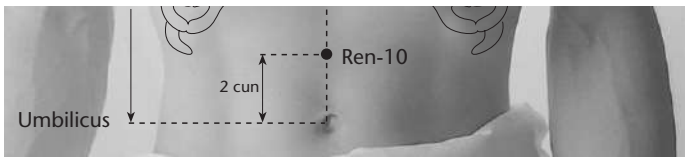

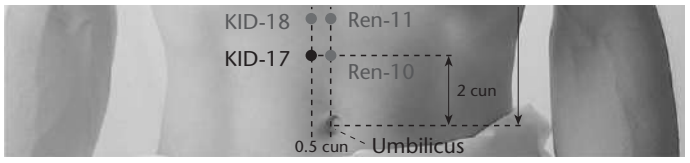

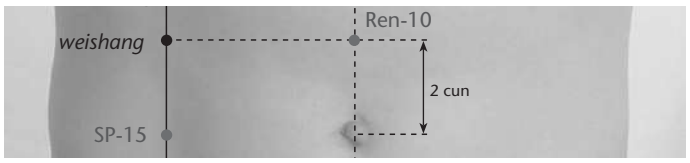

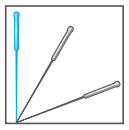
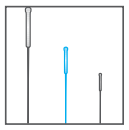
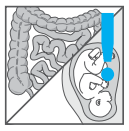

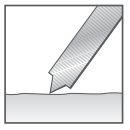
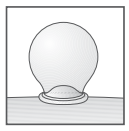

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 152

ST-24 *Slippery Flesh Gate* HUAROUMEN

Location
1 cun above the umbilicus and 2 cun lateral to the anterior midline.

How to find
The distance between the sternocostal angle and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape ➞ 2.3.1). **ST-24** is located 1 cun above the umbilicus and 2 cun lateral to the anterior midline.
Located on the same level is ➞ **Ren-19** (anterior midline); often, also ➞ **LIV-13** (at the free end of the 11th rib) will be located at this level.

Needling
Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications
- Regulates the Stomach
- Transforms Phlegm and calms the *shen*

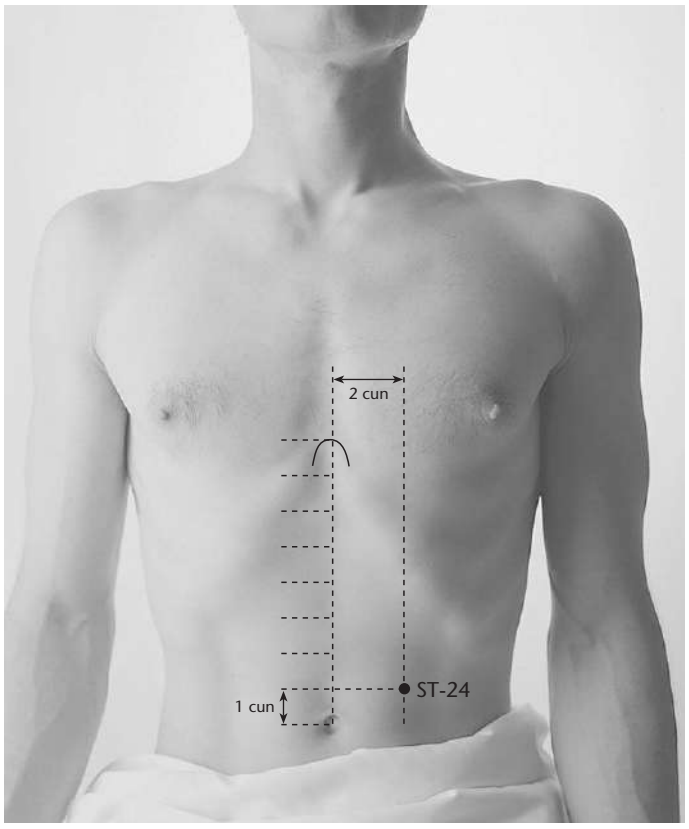

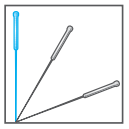
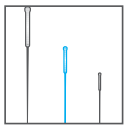
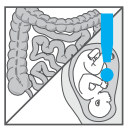

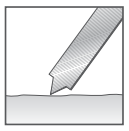

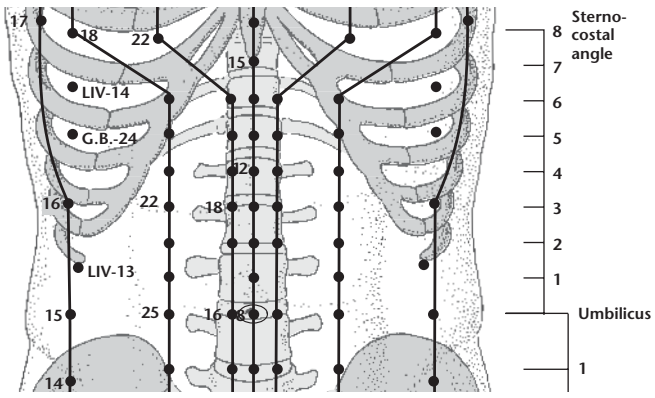

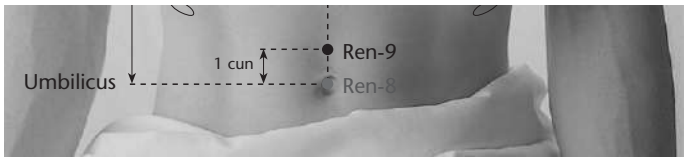

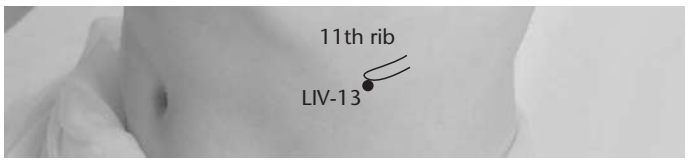

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 153

Heaven's Pivot TIANSHU ST-25

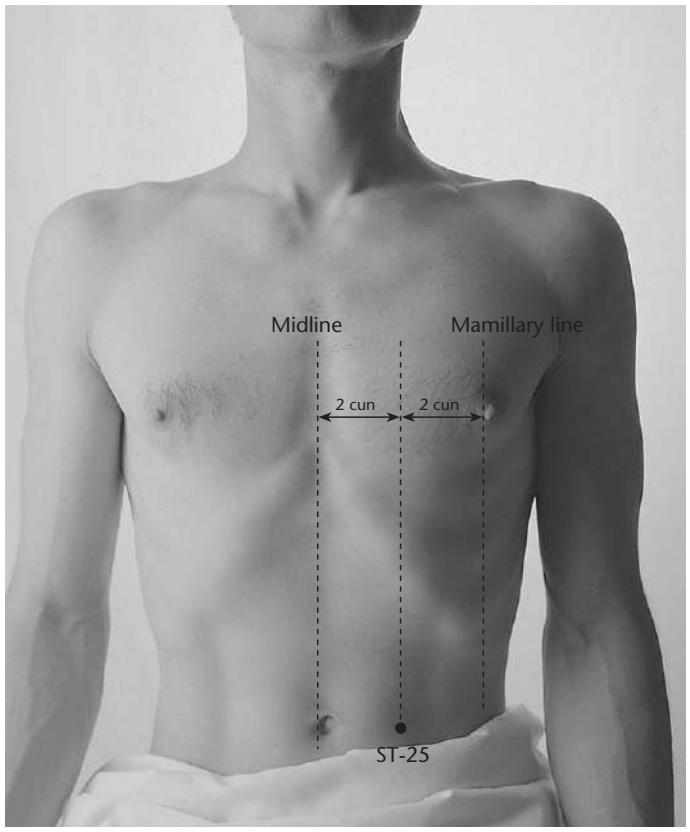

Location

2 cun lateral to the umbilicus.

How to find

Measure 2 cun lateral from the centre of the umbilicus. **For reference:** On the abdomen, 2 cun lateral corresponds to the midpoint between the anterior midline and the mamillary line (→ 3.5). Located on the same level are → **Ren-8/KID-16/SP-15/G.B.-26** (umbilicus/0.5 cun lateral to the umbilicus/mamillary line or 4 cun lateral to the umbilicus/at the level of the umbilicus and directly inferior to the free end of the 11th rib).

Needling

Vertically 0.5–1.5 cun. For gynaecological disorders, obliquely in an inferior direction towards the uterus. Caution: Peritoneum, pregnancy. The needle may reach the aponeuroses of the oblique abdominal musculature at the transition to the sheath of the rectus abdominis muscle.

Actions/Indications

- Regulates the Spleen, Stomach and Intestines, eliminates Dampness and clears Damp-Heat
- Eliminates stagnation

Special features

Front-*mu* point of the Large Intestine. Important point for disorders of the gastrointestinal tract.

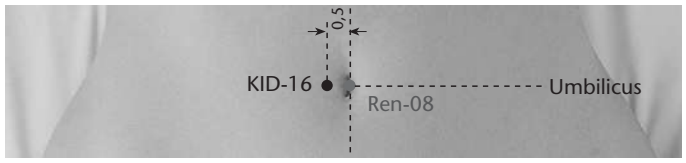

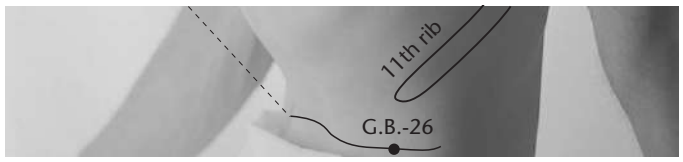

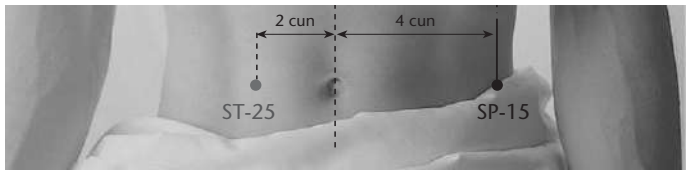

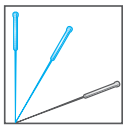
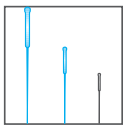
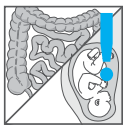
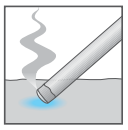
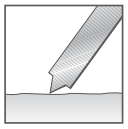
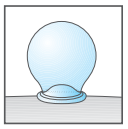

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 154

ST-26 *Outer Mound* WAILING

Location

1 cun below the umbilicus and 2 cun lateral to the anterior midline.

How to find

The distance between the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape → 2.3.1). From the umbilicus, measure 1 cun down: **ST-26** is located 2 cun laterally from there.
Located on the same level are → **Ren-7** (anterior midline) and → **KID-15** (0.5 cun lateral to the anterior midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Regulates the Qi, alleviates pain and Cold

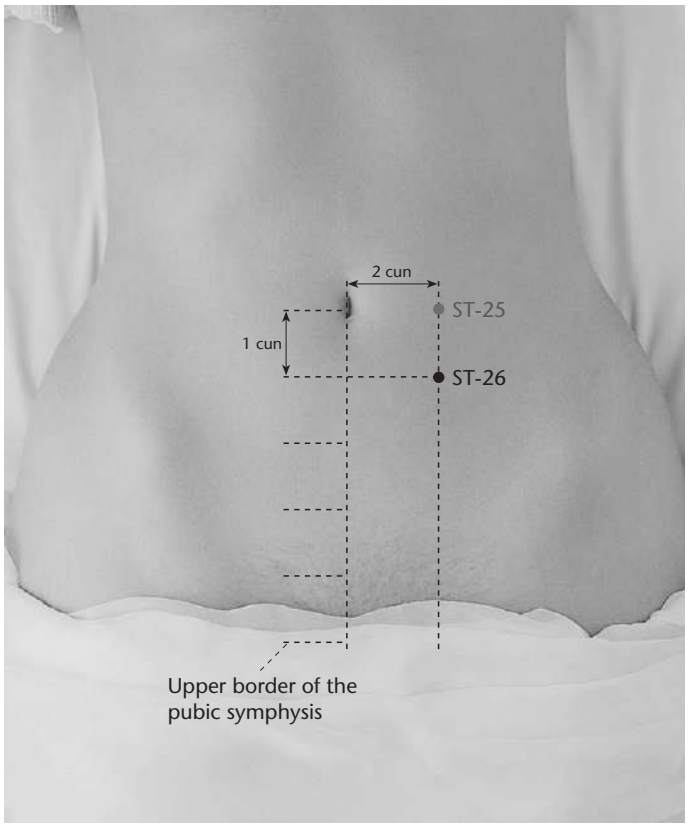

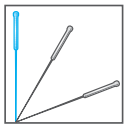

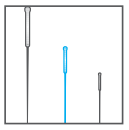

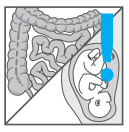

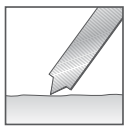

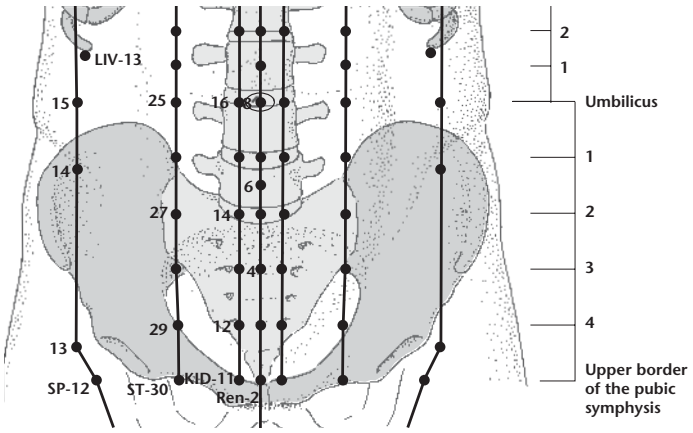

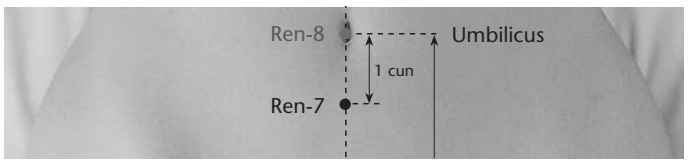

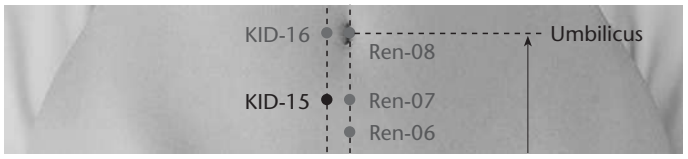

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 155

The Great DAJU ST-27

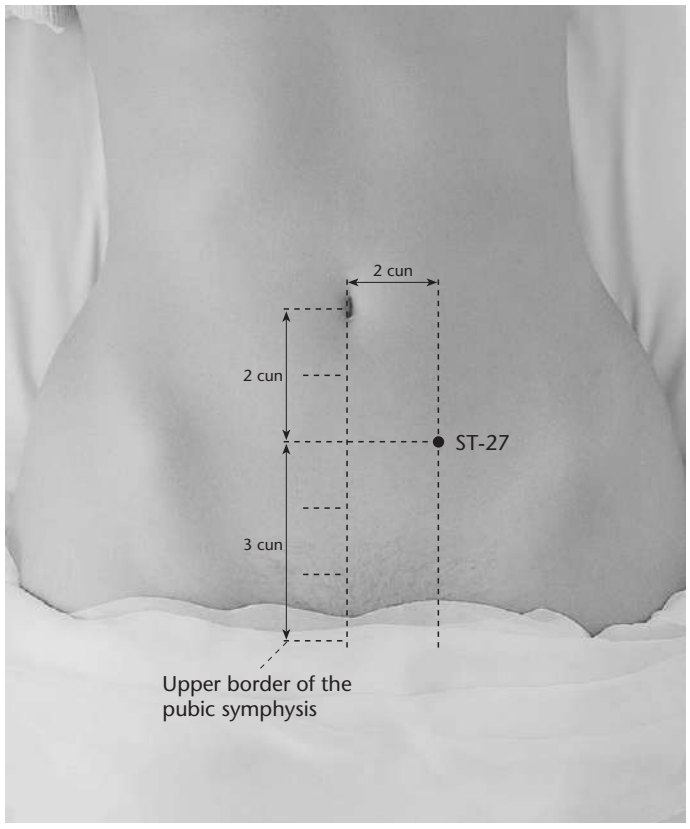

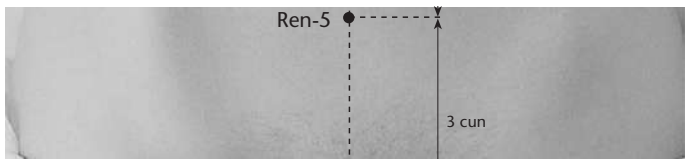

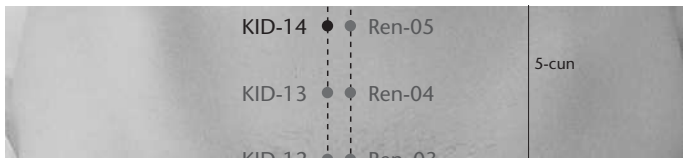

Location

2 cun below the umbilicus and 2 cun lateral to the anterior midline.

How to find

The distance between the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape → 2.3.1). From the umbilicus, measure 2 cun down and 2 cun laterally, where **ST-27** is located.
Located on the same level are → **Ren-5** (anterior midline) and → **KID-14** (0.5 cun lateral to the anterior midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Regulates and strengthens the Qi
- Benefits the Kidneys and promotes diuresis, strengthens the Yang and Essence (*jing*)

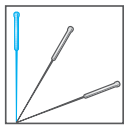
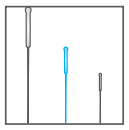
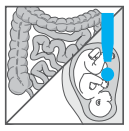

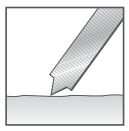

Ch04.3-F10028.qxd 2/22/08 8:42 PM Page 156

ST-28 *Water Passage* SHUIDAO

Location

3 cun below the umbilicus or 2 cun above the upper border of the pubic symphysis and 2 cun lateral to the anterior midline.

How to find

The distance between the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape → 2.3.1). From the upper border of the pubic symphysis, measure 2 cun up or measure 3 cun down from the umbilicus. **ST-28** is located 2 cun laterally from there.
Located on the same level are → **Ren-4** (anterior midline), → **KID-13** (0.5 cun lateral to the anterior midline) and → **Ex-CA** (*tituo/quimen/yijing*; 4/3/1 cun lateral to the anterior midline).

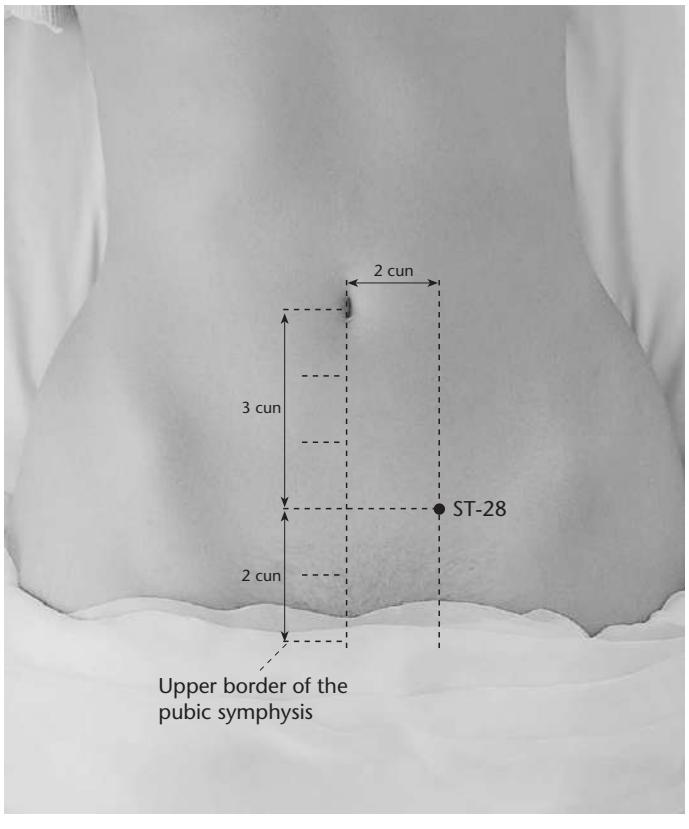

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, urinary bladder, pregnancy.

Actions/Indications

- Clears Damp-Heat, benefits the Lower Burner and eliminates stagnation

Special features

Important local point for the urogenital tract

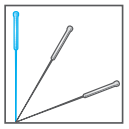
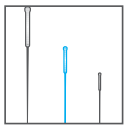
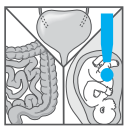

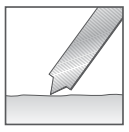

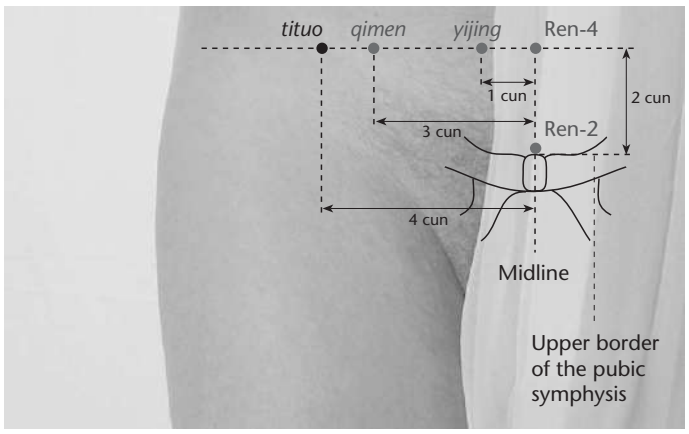

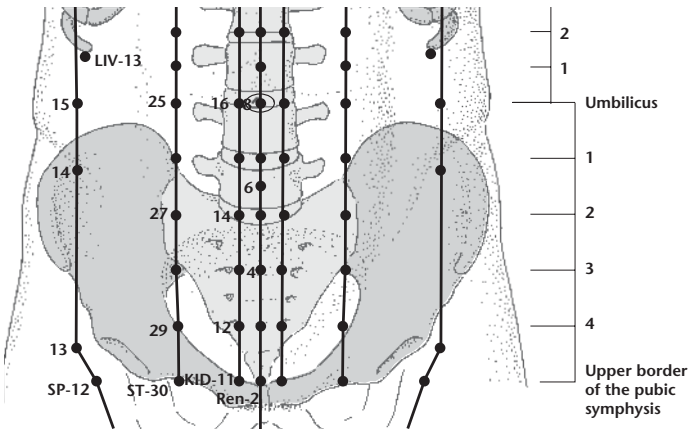

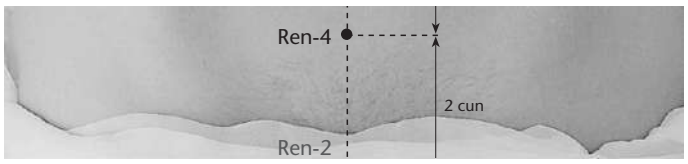

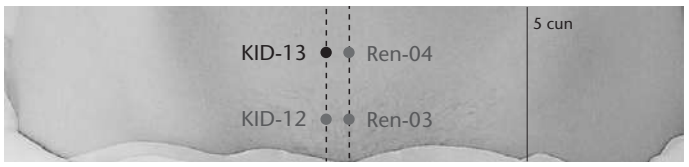

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 157

Return GUILAI ST-29

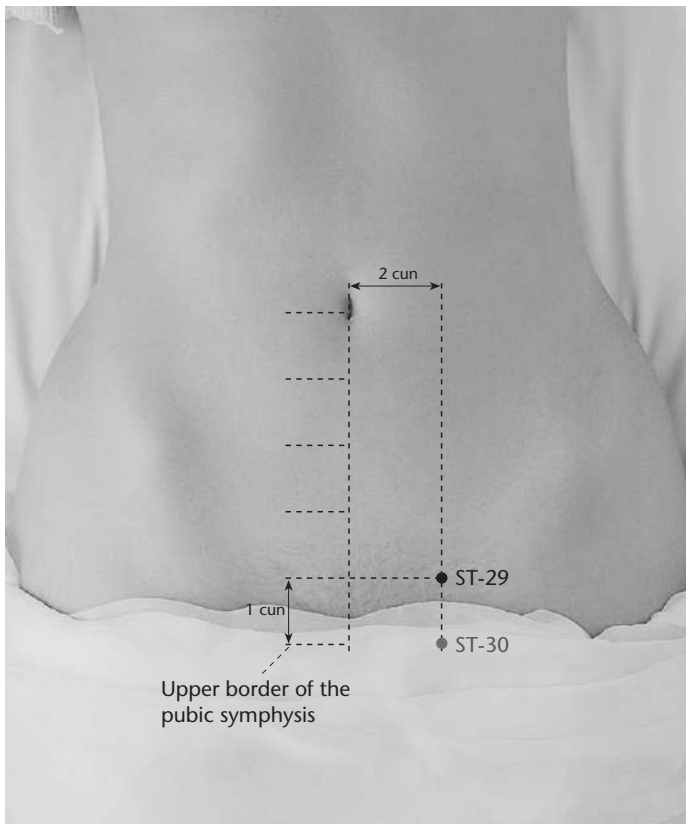

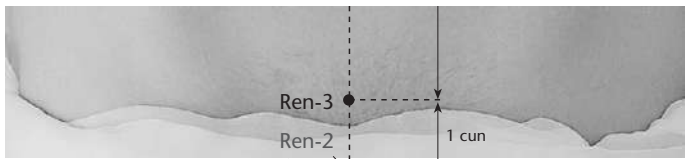

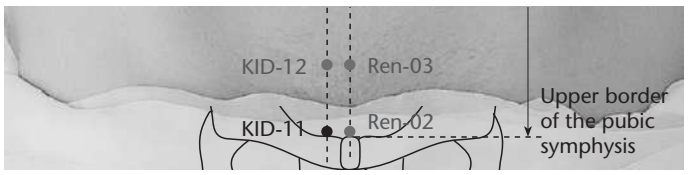

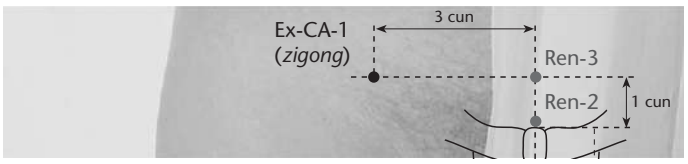

Location

1 cun above the upper border of the pubic symphysis or 4 cun below the umbilicus and 2 cun lateral to the anterior midline.

How to find

The distance between the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun; therefore, only proportional cun measurements should be used (helpful tool: elastic tape → 2.3.1). From the upper border of the pubic symphysis, measure 1 cun up and 2 cun lateral to **ST-29**.

Located on the same level are → **Ren-3** (anterior midline), → **KID-12** (0.5 cun lateral to the anterior midline) and → **Ex-CA** (*zigong*; 3 cun lateral to the anterior midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, urinary bladder, pregnancy.

Actions/Indications

- Warms and regulates the Lower Burner and menstruation

Special features

Commonly treated with moxibustion, the combination of acupuncture and moxibustion being particularly beneficial. Important local point for the urogenital tract.

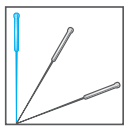
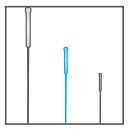
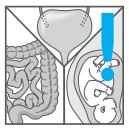

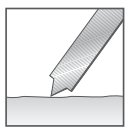
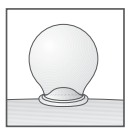

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 158

ST-30 *Rushing Qi* QICHONG

Location

2 cun lateral to the upper border of the pubic symphysis and medial to the femoral artery and vein, approximately 1 cun superior to the inguinal groove.

How to find

First, locate the upper border of the pubic symphysis (location of → **Ren-2**). Then locate **ST-30** 2 cun lateral from there (on the vertical line halfway between the anterior midline and the mamillary line). This often corresponds with a location 1 cun superior to the inguinal groove and medial to the femoral artery/vein.
Located on the same level are → **Ren-2/KID-11/SP-12** (anterior midline/0.5 cun lateral to the anterior midline/3.5 cun lateral to the anterior midline). → **LIV-12** is located in the inguinal groove, 1 cun inferior and 0.5 cun lateral to **ST-30**.

Needling

0.5–1 cun vertically or 1.5 cun slightly obliquely towards the external genitalia for disorders of the urogenital tract. Caution: Femoral artery, urinary bladder (quite superficial in a superior direction); in men, the spermatic ducts (quite superficial in an inferior direction), pregnancy. Ask the patient to empty their bladder before needling this point.

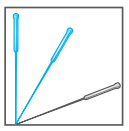

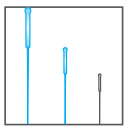

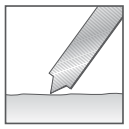

Actions/Indications

- Regulates the Qi in the Lower Burner
- Strengthens the Sea of Grain (later Heaven Qi)

Special features

Meeting point with the *chong mai* (according to some authors, also with the G.B. channel), Sea of Grain point. Wide usage for disorders of the Lower Burner.

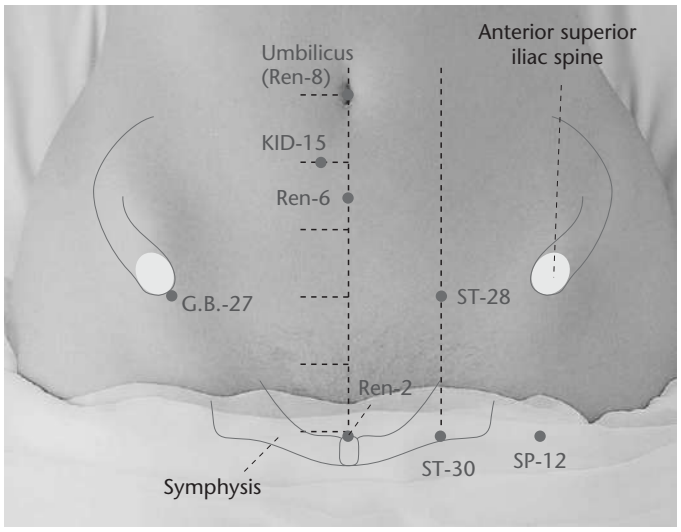

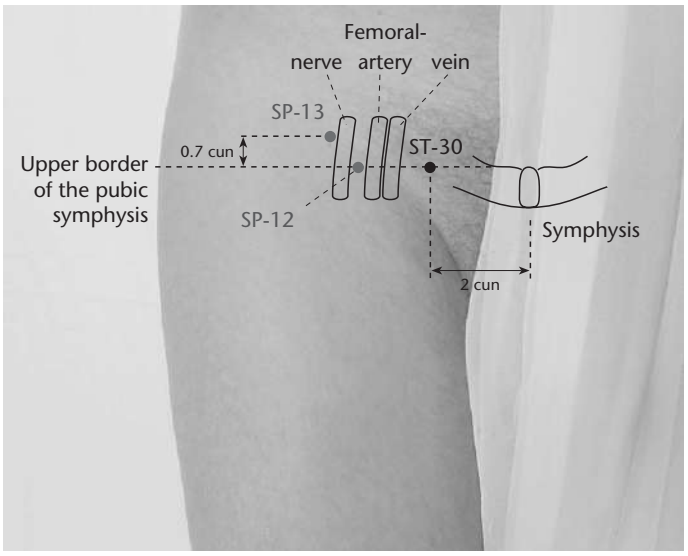

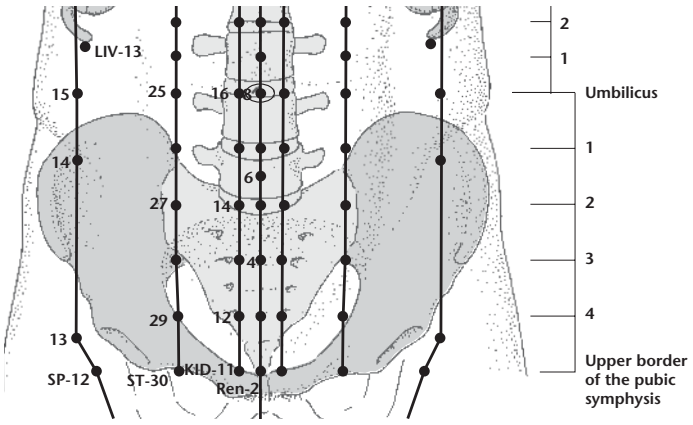

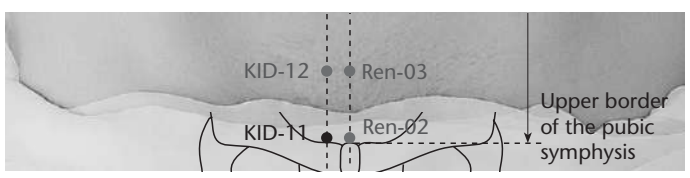

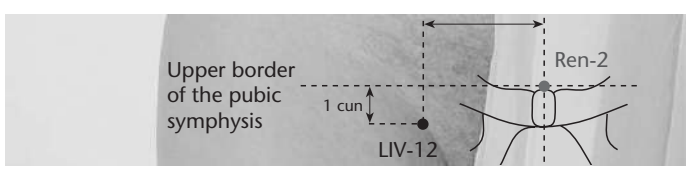

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 159

Thigh Gate BIGUAN ST-31

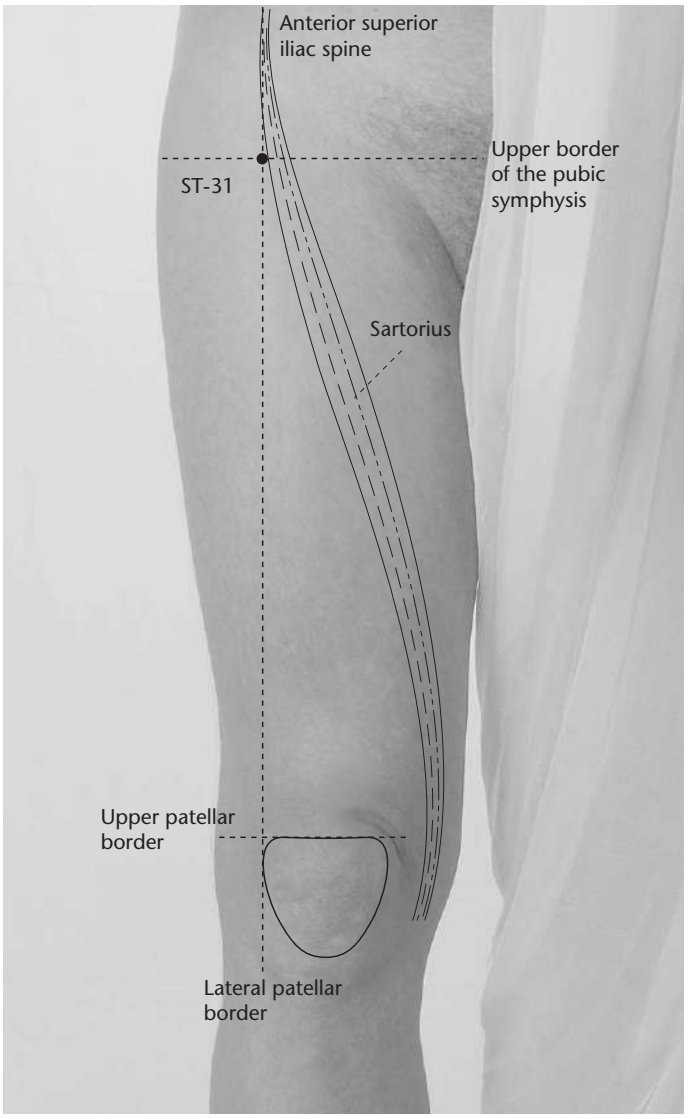

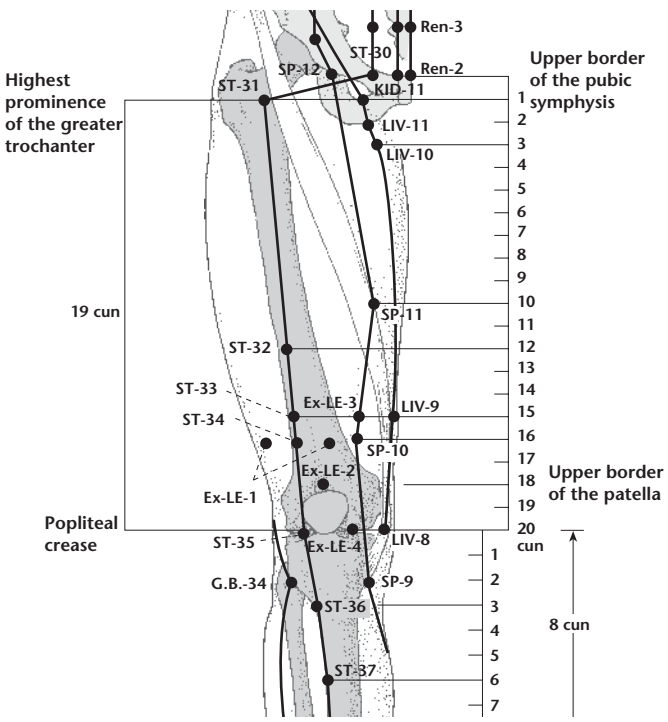

Location

Inferior to the anterior superior iliac spine and lateral to the sartorius muscle, at the level of the lower border of the pubic symphysis.

How to find

The sartorius muscle originates at the anterior superior iliac spine (ASIS) and crosses the thigh towards the medial aspect of the knee. With the knee and hip slightly flexed and with active lateral rotation, it can be isolated easily. **ST-31** is located on the lateral border of the muscle, at the crossing point of a vertical line connecting the ASIS and the lateral upper border of the patella and a horizontal line at the level of the lower border of the pubic symphysis.

Needling

Vertically 0.5–2 cun

Actions/Indications

- Opens the channel, warms and expels pathogenic factors

Special features

Particularly useful for Cold *bi*-syndrome, disorders of the knees and lower back. The chain-and-lock method in combination (→ 8) with → **ST-36** and → **ST-41** is often used for paraesthesia and pain in the hip and leg radiating along the channel.

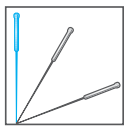
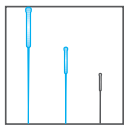

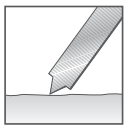
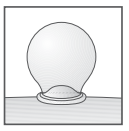

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 160

ST-32 *Crouching Rabbit* FUTU

Location

On the thigh, on a line connecting the lateral patellar border and the anterior superior iliac spine, 6 cun superior to the upper, lateral border of the patella.

How to find

Divide the distance between the tip of the greater trochanter and the lateral upper border of the patella into three sections. **ST-32** is located at the junction of the lower and the middle third of the thigh in a depression on the quadriceps muscle.

Needling

Vertically 0.5–2 cun

Actions/Indications

- Opens the channel and expels Wind, Dampness and Cold

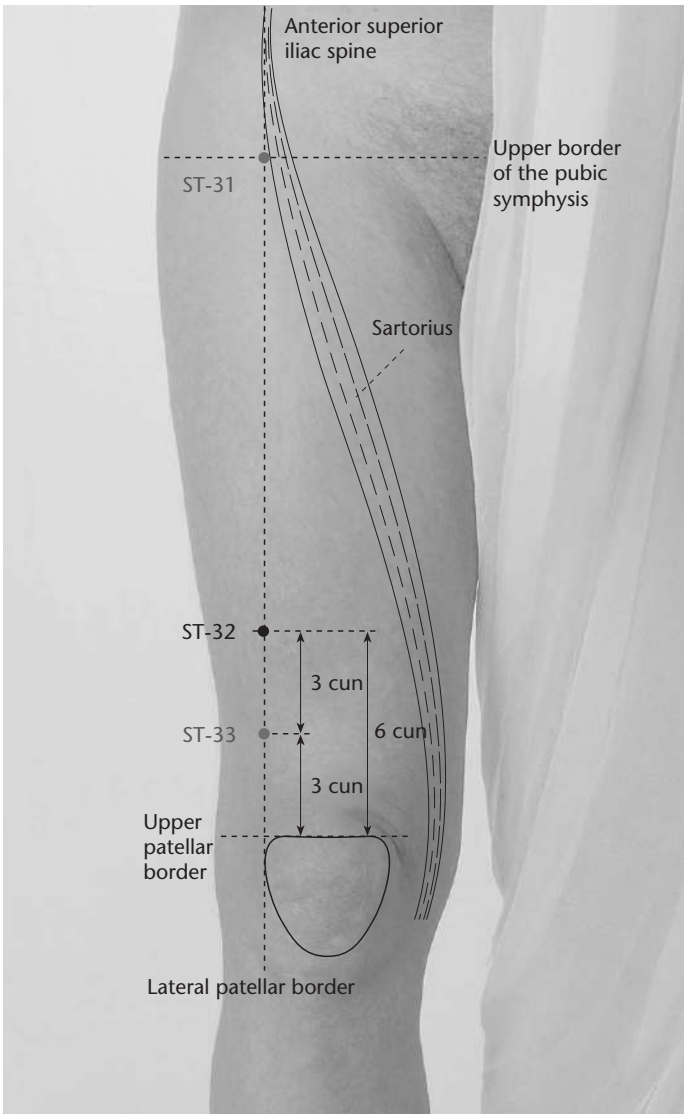

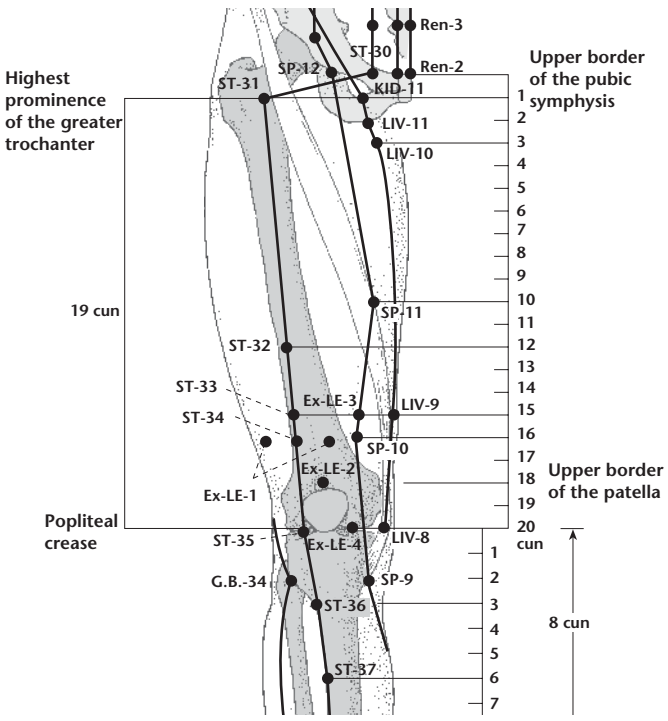

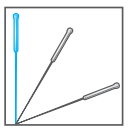
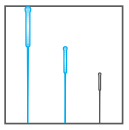

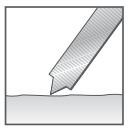

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 161

Yin Market YINSHI ST-33

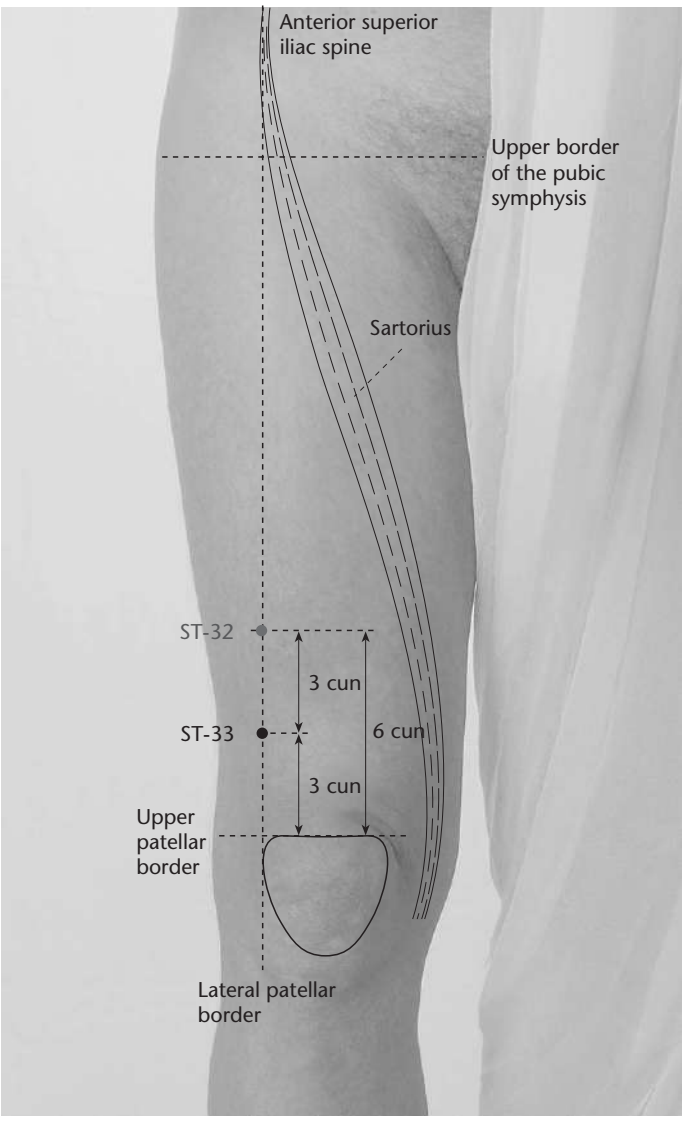

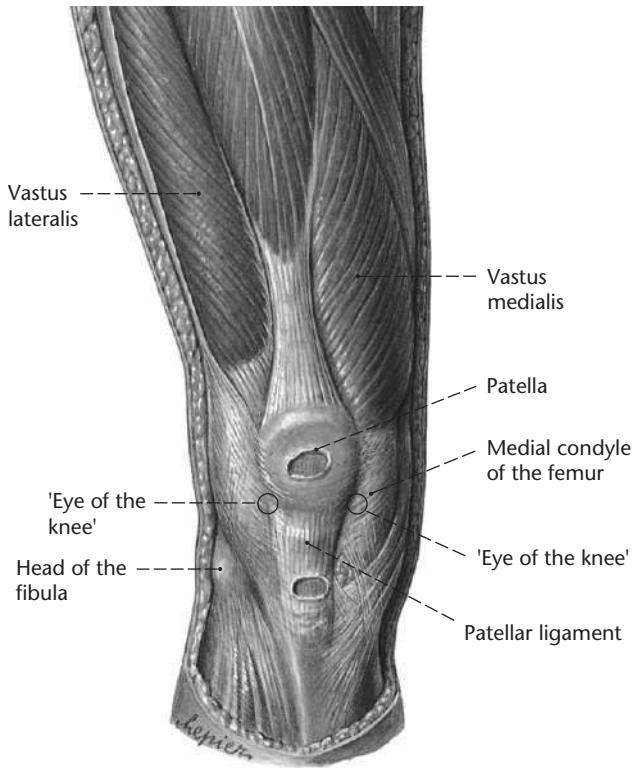

Location

On a line joining the lateral patellar border and the anterior superior iliac spine, 3 cun superior to the upper lateral border of the patella.

How to find

Measure 3 cun up from the upper lateral border of the patella, on the connecting line to the anterior superior iliac spine. **ST-33** is located in a groove at the junction of the rectus femoris and vastus lateralis muscles.

Needling

Vertically 0.5–1.5 cun

Actions/Indications

- Opens the channel and expels pathogenic factors

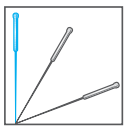
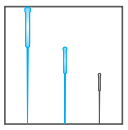

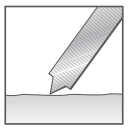
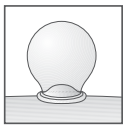

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 162

ST-34 *Ridge Mound* LIANGQIU

Location

On a line joining the lateral patellar border and the anterior superior iliac spine, 2 cun proximal to the upper lateral border of the patella, in a groove of the vastus lateralis muscle.

How to find

Locate the lateral upper border of the patella and from there measure 2 cun proximally. **ST-34** is located in a palpable groove on the vastus lateralis muscle of the quadriceps femoris group. This point projects on the line joining the lateral upper border of the patella and the anterior superior iliac spine (ASIS → 3.5). Located on the same level are the points → **Ex-LE-1** (each 1.5 cun lateral and medial to **ST-34**). → **SP-10** is located in a comparable position on the medial aspect of the thigh (2 cun proximal to the medial upper border of the patella and in a groove on the vastus medialis muscle).

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Regulates the Stomach Qi, alleviates acute conditions
- Opens the channel and alleviates pain

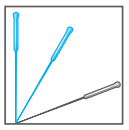

Special features

Xi-cleft point. Important local point for disorders of the knee. Modern application for acute epigastric pain – strongly reducing needling techniques have an immediate and calming effect on gastrointestinal peristalsis.

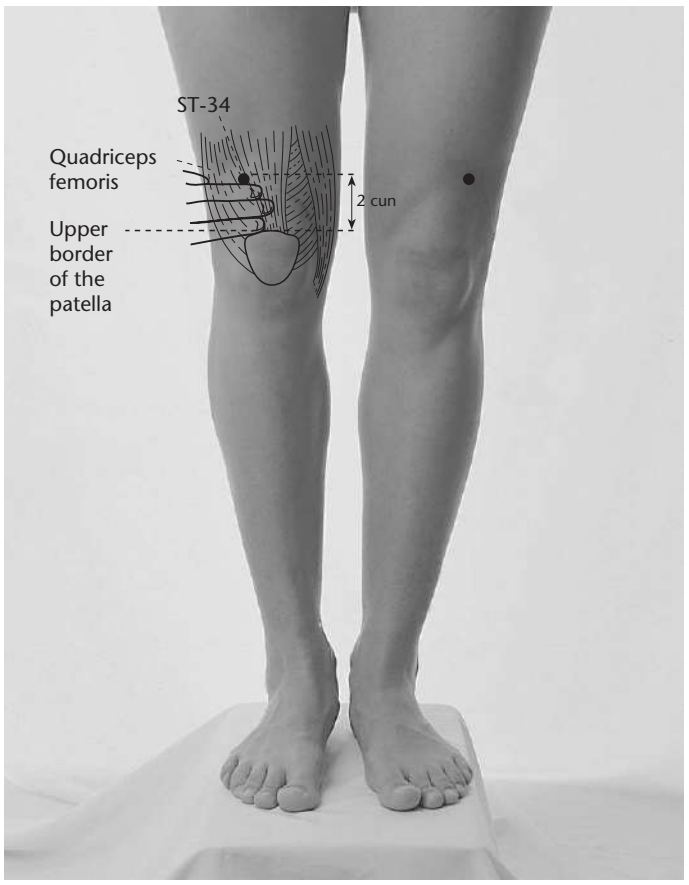

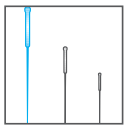

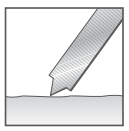

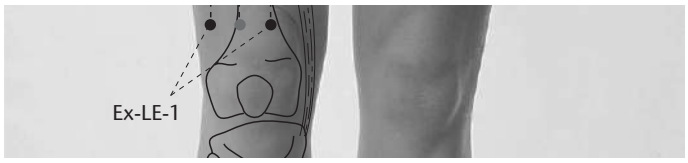

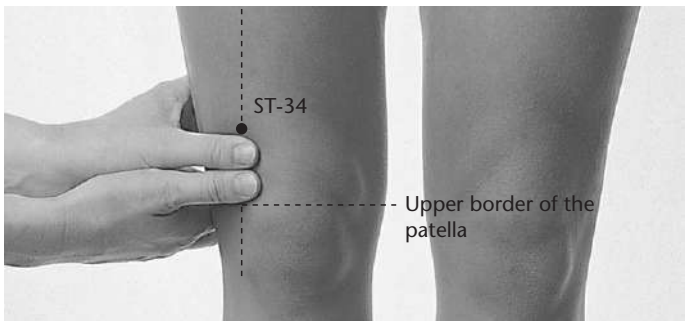

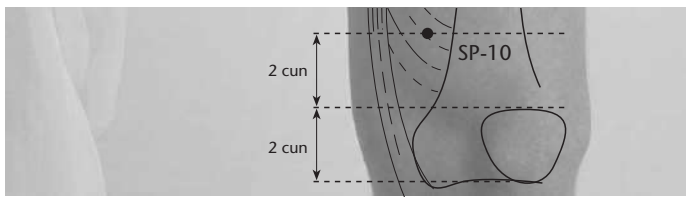

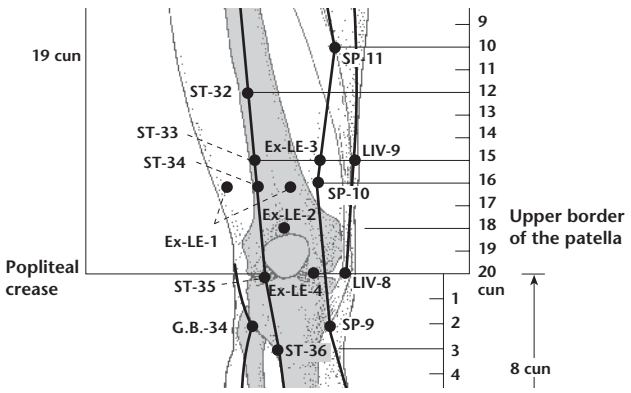

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 163

Calf's Nose DUBI ST-35

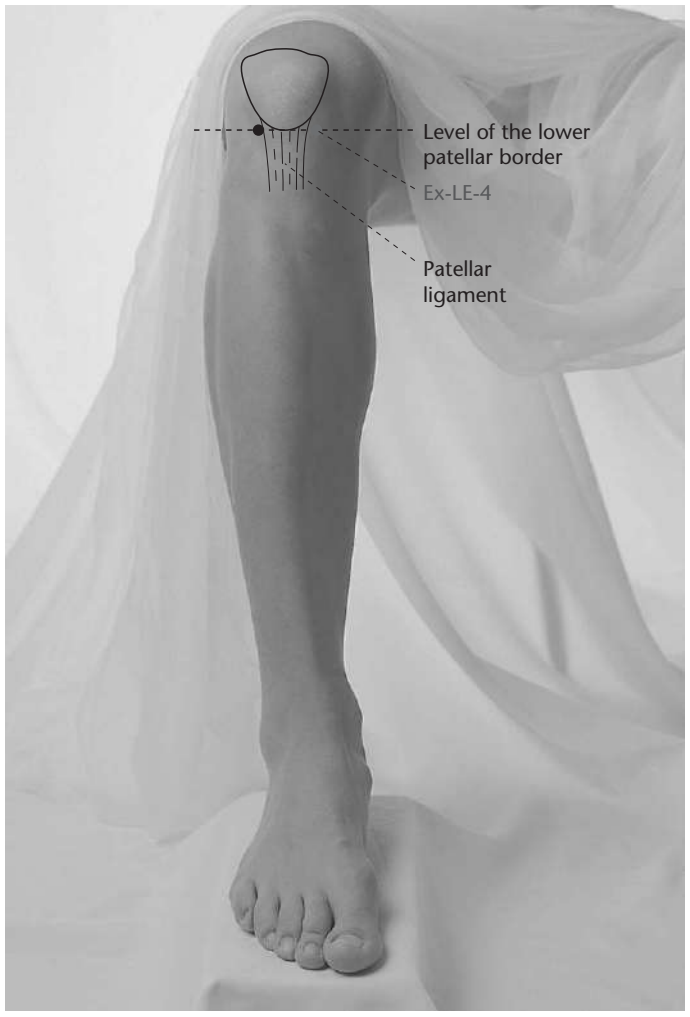

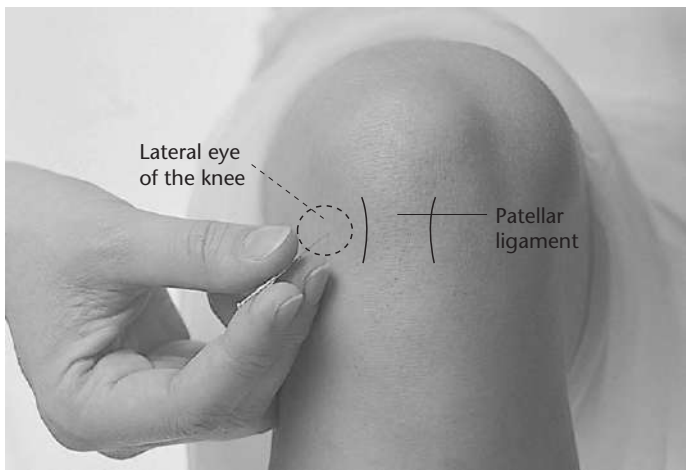

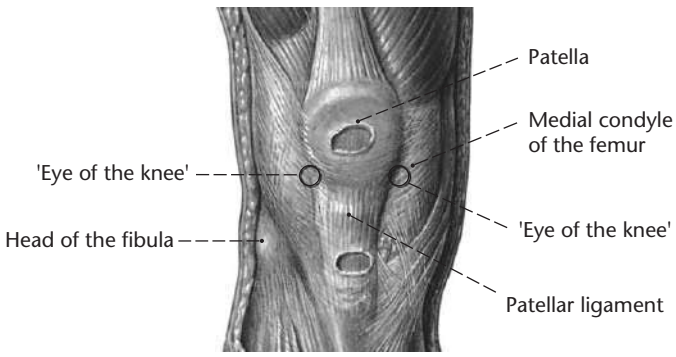

Location

With the knee flexed, in the depression inferior to the patella and lateral to the patellar ligament.

How to find

This point is best located and needled with the knee slightly flexed (use knee support). **ST-35** is located in a depression on the lower lateral border of the patella, in a groove lateral to the patellar ligament. It projects onto the level of the knee joint space and more or less corresponds with the arthroscopic access to the joint. **ST-35** is also known as 'lateral eye of the knee' and is part of the extra point → **Ex-LE-5** (*xiyan*).

Needling

0.5–1 cun slightly obliquely in a medial direction, towards the medial eye of the knee → **Ex-LE-4** (*neixiyan*). Caution: Knee joint (the location of this point corresponds approximately with the arthroscopic access to this joint!).

Actions/Indications

- Expels Wind and Dampness, reduces swellings, opens the channel, alleviates pain

Special features

Important local point for all disorders of the knee

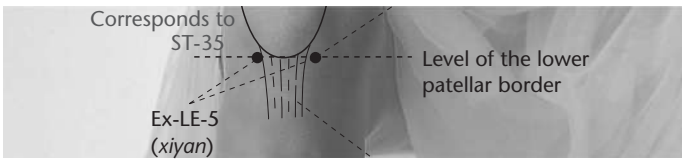

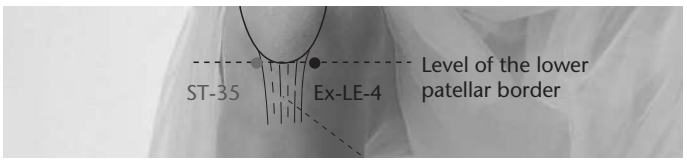

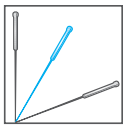
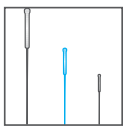

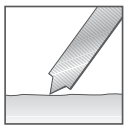

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 164

ST-36 *Leg Three Miles* ZUSANLI

Location

3 cun distal to → **ST-35** ('lateral eye of the knee') and 1 finger-breadth lateral to the anterior crest of the tibia, on the tibialis anterior muscle.

How to find

First, locate → **ST-35** ('lateral eye of the knee', at the level of the knee joint space). From → **ST-35**, palpate 3 cun (1 handbreadth) down and 1 fingerbreadth towards lateral by placing the middle finger adjacent to the tibial crest. There, **ST-36** is located in a depression, which can be palpated 'dynamically' (choose the point according to pressure sensitivity). **Or:** Palpate the lower border of the tibial tuberosity and 'dynamically' locate **ST-36** 1 fingerbreadth lateral to it.

Needling

Vertically 1–1.5 cun. Moxibustion is often used on this point.

Actions/Indications

- Regulates the Stomach, strengthens the Spleen, transforms Dampness
- Tonifies Qi and Yang, nourishes the Blood and Yin
- Calms the *shen*
- Opens the channel

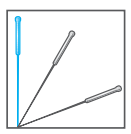

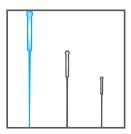

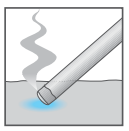

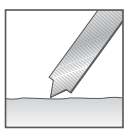

Special features

He-sea point, Earth point, *ben* point (Five Phases point), lower *he*-sea point of the Stomach, *Gao Wu* command point for disorders of the abdomen, Ma Dan Yang point, Sea of Water and Grain point. Main point to tonify Qi and Blood for general weakness. Very important point.

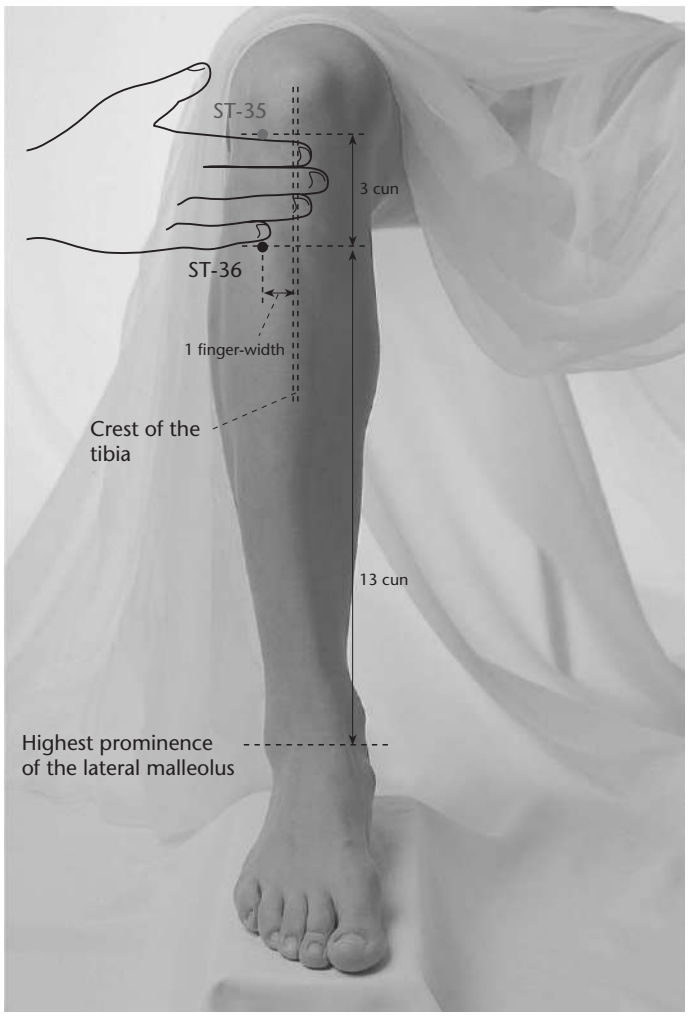

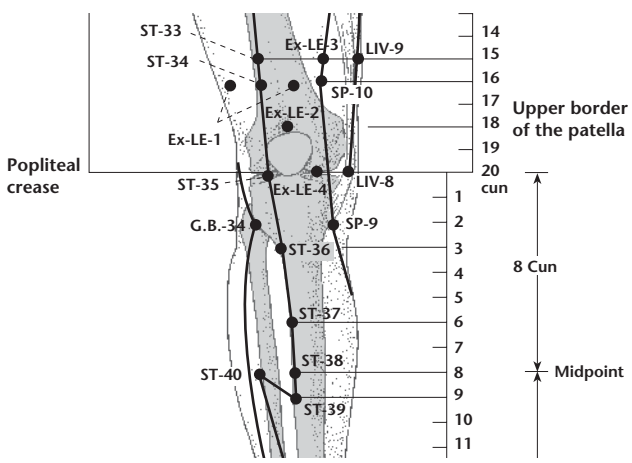

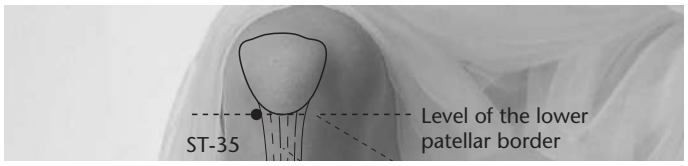

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 165

Upper Great Emptiness SHANGJUXU ST-37

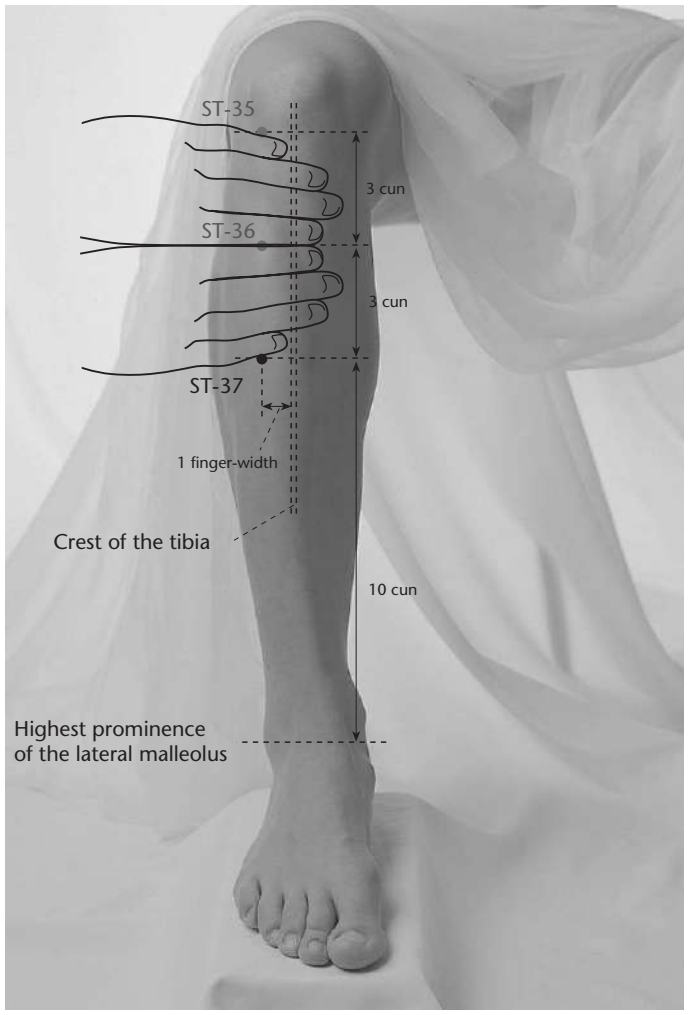

Location

6 cun distal to → **ST-35** (at the knee joint space) or 3 cun distal to → **ST-36** and 1 fingerbreadth lateral to the anterior crest of the tibia.

How to find

First, find → **ST-35** ('lateral eye of the knee', level with the knee joint space). From → **ST-35**, measure 6 cun (2 handbreadths) in a distal direction and, at that level, 1 fingerbreadth towards lateral (place the middle finger onto the anterior crest of the tibia). There, **ST-37** is located in a depression, which can be 'dynamically' palpated (choose the point according to pressure sensitivity). **Or:** Find the midpoint of the line joining → **ST-35** and → **ST-41** (in the depression between the tendons of the extensor digitorum longus and the hallucis longus, level with the highest prominence of the lateral malleolus) by using the spreading hands technique (→ 2.3.3). **ST-37** is located 2 cun proximal and 1 fingerbreadth lateral to this point.

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Regulates the Spleen, Stomach and Intestines, eliminates stagnation, clears Damp-Heat
- Opens the channel, alleviates pain

Special features

Lower *he*-sea point of the Large Intestine, Sea of Blood point.

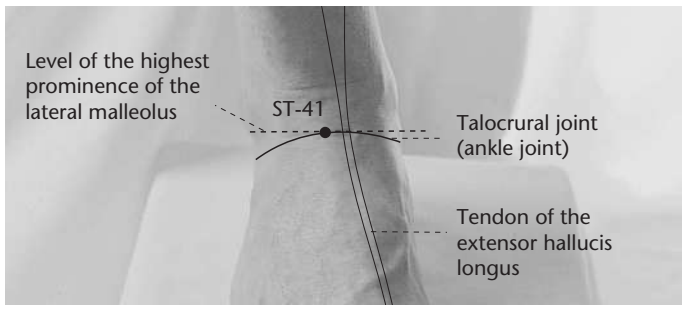

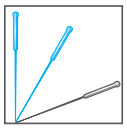
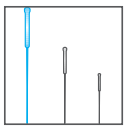

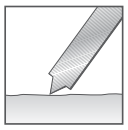

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 166

ST-38 *Lines Opening* TIAOKOU

Location

At the midpoint of the line joining → **ST-35** and → **ST-41** and 1 fingerbreadth lateral to the anterior crest of the tibia.

How to find

Quick method: Spreading-hands technique (→ 2.3.3): place the small fingers on → **ST-35** ('lateral eye of the knee', level with the knee joint space) and → **ST-41** (in the depression between the tendons of the extensor digitorum longus and the hallucis longus, level with the highest prominence of the lateral malleolus) and determine the midpoint of this line, using the thumbs. **ST-38** is located in a depression 1 fingerbreadth (middle fingerbreadth, measured from the anterior crest of the tibia) lateral to the midpoint. **ST-38** may be palpated 'dynamically' (according to pressure sensitivity). **Or:** From → **ST-35**, measure 8 cun (→ 2 handbreadths and 3 fingerbreadths) distally and 1 fingerbreadth laterally.
Located on the same level are → **ST-40** (1 fingerbreadth more laterally) and → **BL-57** (dorsal aspect of the lower leg, on the midpoint of the line joining → **BL-40** and → **BL-60**).

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Expels Wind-Damp, opens the channel, alleviates pain, benefits the shoulder

Special features

Very effective distal point for disorders of the shoulder. For limited range of motion of the shoulder joint: apply strong reducing needling techniques while the patient moves their shoulder joint.

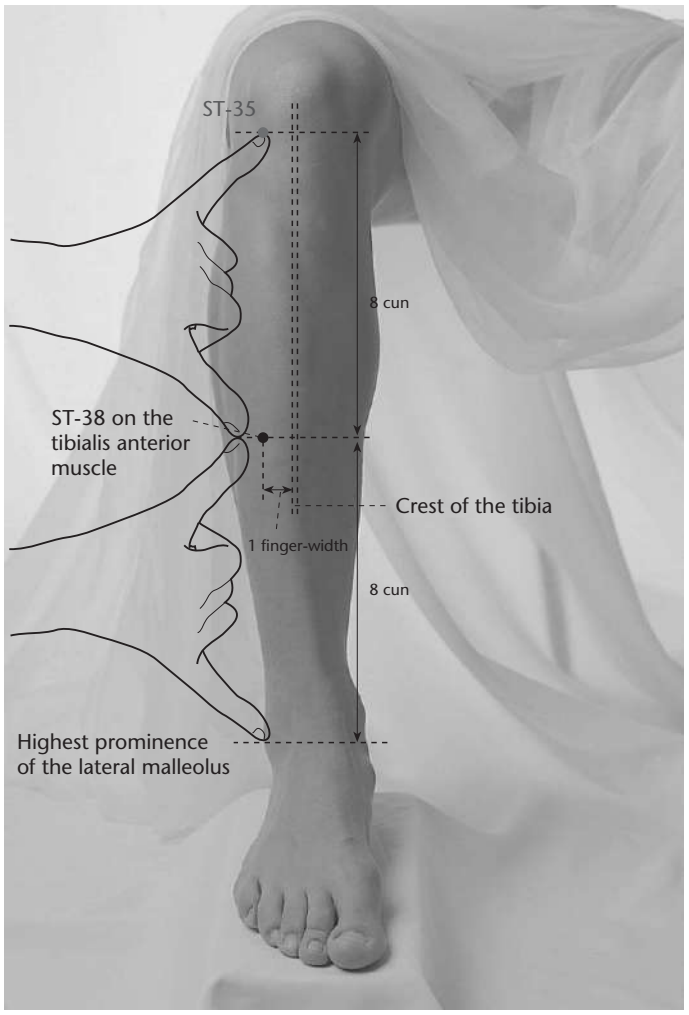

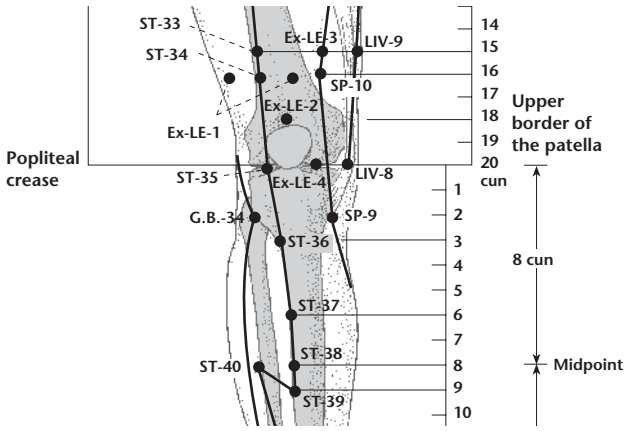

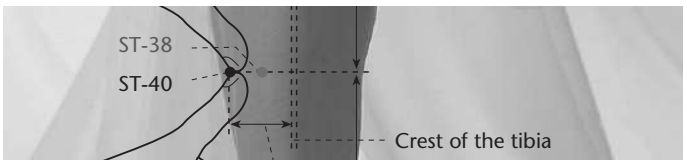

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 167

Lower Great Void XIAJUXU ST-39

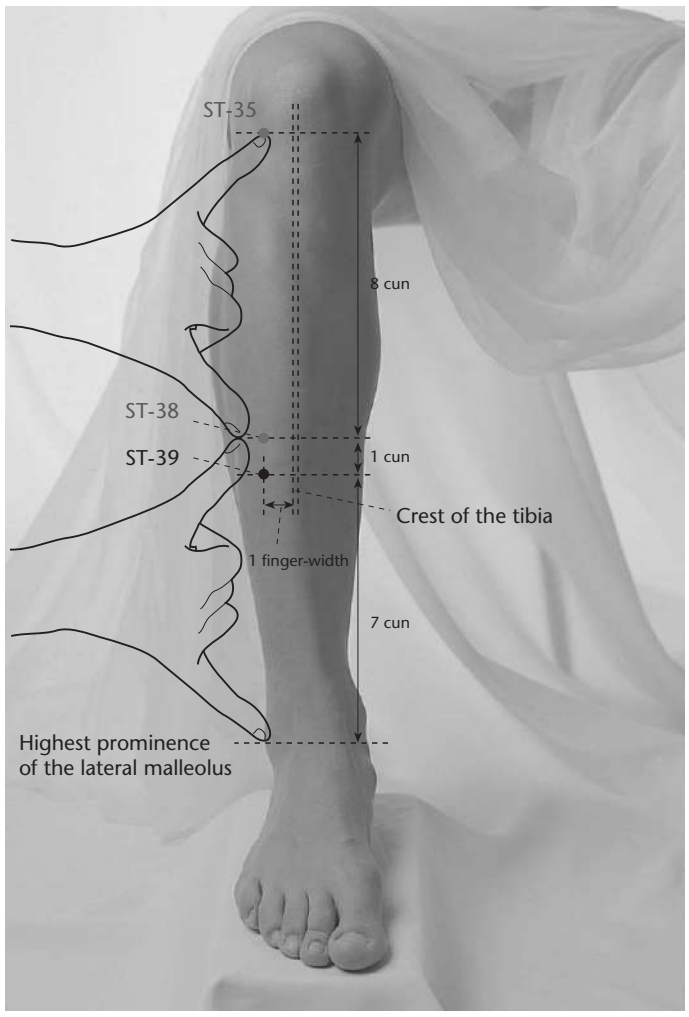

Location

1 cun distal to → **ST-38** (midpoint of the line joining → **ST-35** and → **ST-41**) and 1 fingerbreadth lateral to the anterior crest of the tibia.

How to find

Quick method: Spreading-hands technique (→ 2.3.3): Place the little fingers on → **ST-35** ('lateral eye of the knee', level with the knee joint space) and → **ST-41** (in the depression between the tendons of the extensor digitorum longus and the hallucis longus, level with the highest prominence of the lateral malleolus) and determine the midpoint of this line (level with → **ST-38**), using the thumbs. **ST-39** is located in a depression 1 cun distal and 1 fingerbreadth (middle fingerbreadth, measured from the anterior crest of the tibia) lateral to this point. **ST-39** may be palpated 'dynamically' (according to pressure sensitivity).

Located on the same level (7 cun proximal to the lateral malleolus) are → **G.B.-35** (on the posterior border of the fibula), → **G.B.-36** (on the anterior border of the fibula) and → **BL-58** (on the lateral border of the gastrocnemius muscle).

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Moves the Small Intestine Qi, eliminates stagnation, clears Damp-Heat
- Opens the channel, alleviates pain

Special features

Lower *he*-sea point of the Small Intestine, Sea of Blood point.

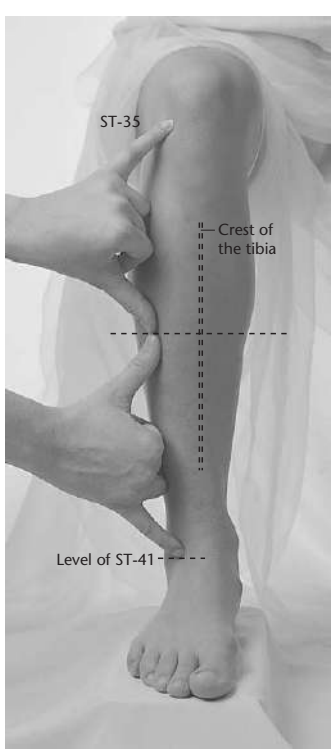

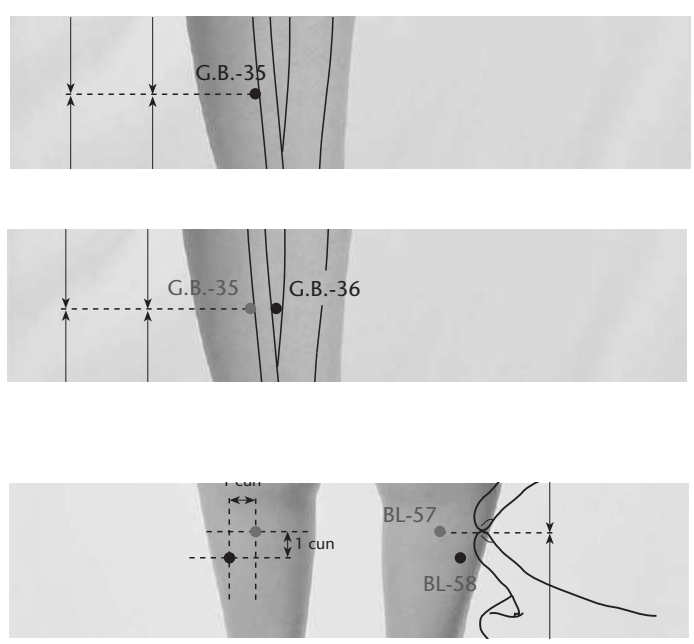

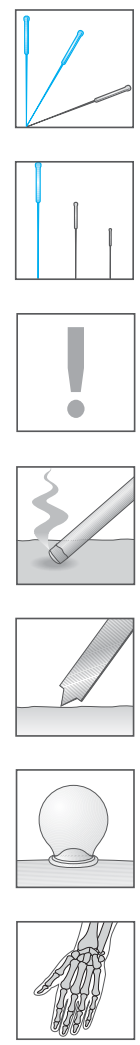

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 168

ST-40 *Abundant Bulge* FENGLONG

Location

At the midpoint of the line joining → **ST-35** and → **ST-41** and 2 fingerbreadths lateral to the anterior crest of the tibia or 1 fingerbreadth lateral to → **ST-38**, between the extensor digitorum longus and peroneus brevis muscles.

How to find

Quick method: Spreading-hands technique (→ 2.3.3): Place the little fingers on → **ST-35** ('lateral eye of the knee', level with the knee joint space) and → **ST-41** (in the depression between the tendons of the extensor digitorum longus and the hallucis longus, level with the highest prominence of the lateral malleolus) and determine the midpoint of this line, using the thumb. **ST-40** is located in a depression 2 fingerbreadths (2 middle fingerbreadths, measured from the anterior crest of the tibia) lateral to this point. **ST-40** may be palpated 'dynamically' (according to pressure sensitivity). Located on the same level (8 cun proximal to the highest prominence of the lateral malleolus) are → **ST-38** (1 fingerbreadth lateral to the tibia) and → **BL-57** (on the dorsal aspect of the lower leg, halfway between → **BL-40** and → **BL-60**).

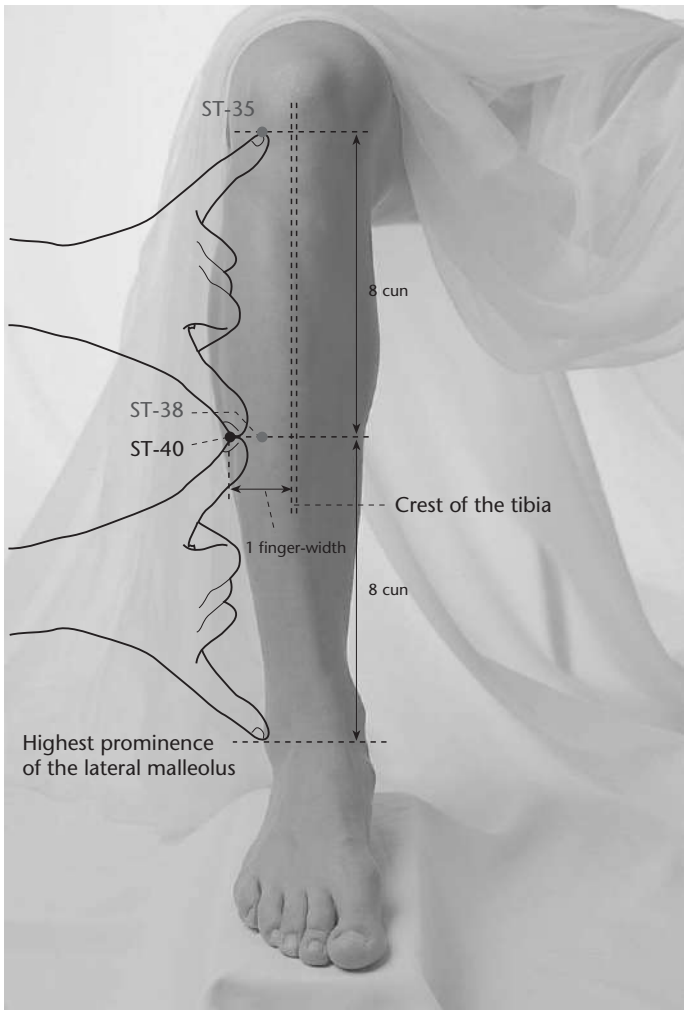

Needling

Vertically or obliquely 1–1.5 cun

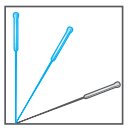

Actions/Indications

- Transforms Dampness and Phlegm, clears Phlegm in the Lung and Heart, calms the *shen*

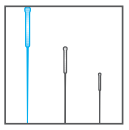

Special features

Luo-connecting point. Main acupuncture point for eliminating Phlegm.

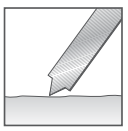

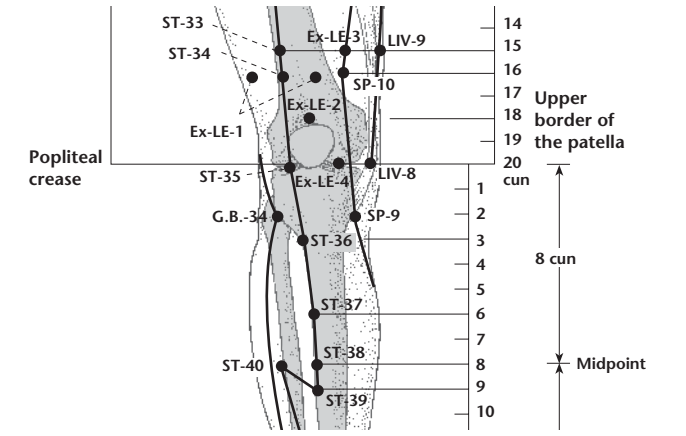

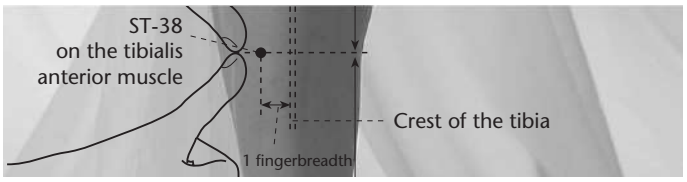

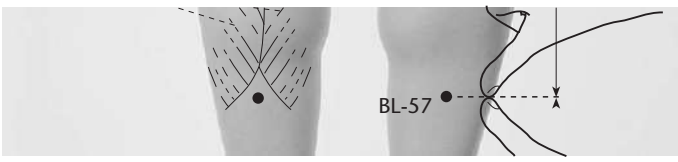

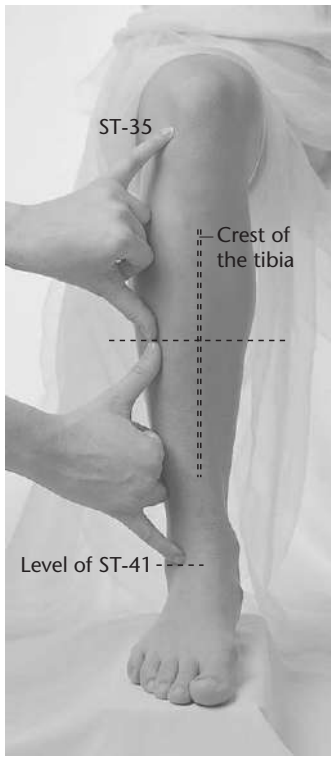

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 169

Stream Divide JIEXI ST-41

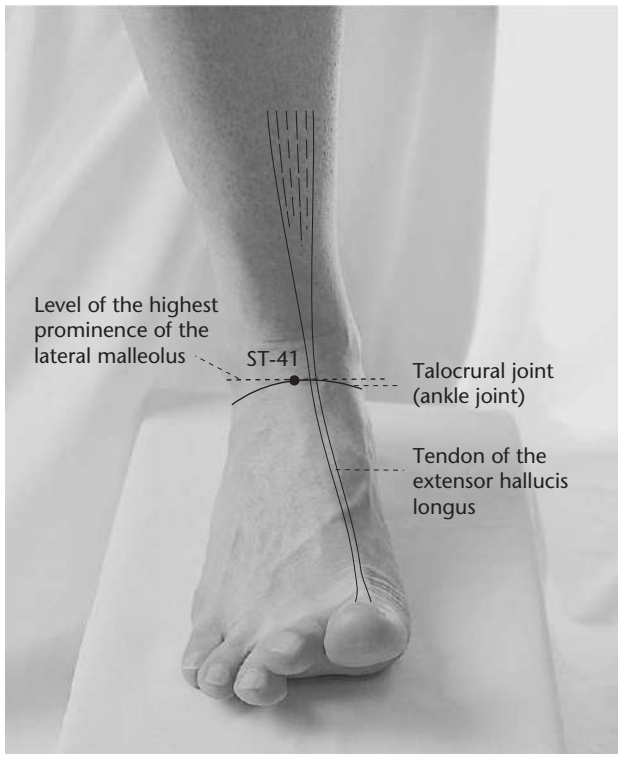

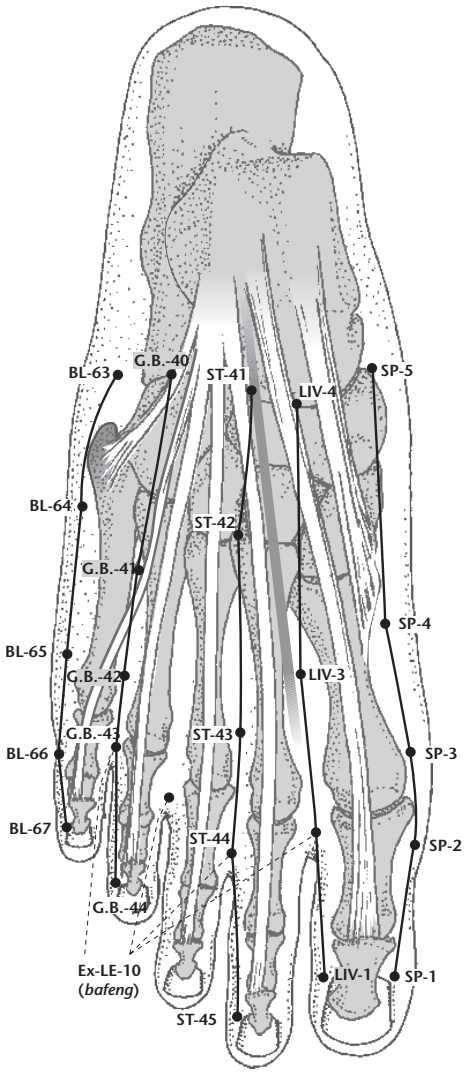

Location

On the ankle, on the level of the highest prominence of the lateral malleolus, in the depression between the tendons of the extensor digitorum and the extensor hallucis longus.

How to find

With the patient extending their big toe against resistance, the tendon of the extensor hallucis longus muscle will become more distinct. **ST-41** is located lateral to this tendon, in an obvious depression on the level of the highest prominence of the lateral malleolus (→ 3.6.2) and ankle joint space.
Located on the same level is → **LIV-4** (medial to the tendon, halfway between → **SP-5** and **ST-41**).

Needling

0.5–1 cun vertically or obliquely deep to the tendons in a medial direction towards → **SP-5** or in a lateral direction towards → **G.B.-40**. Caution: Tibial artery/vein/nerve lie deep to this point.

Actions/Indications

- Clears Heat from the Stomach and Stomach channel
- Calms the *shen*
- Opens the channel and *luo* vessels, alleviates pain

Special features

Jing-river point, Fire point, tonification point. Important local point for disorders of the ankle joint. Important distal point for frontal headaches caused by Stomach Heat or Stomach Fire.

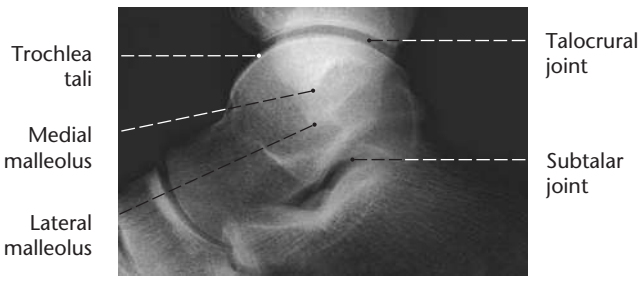

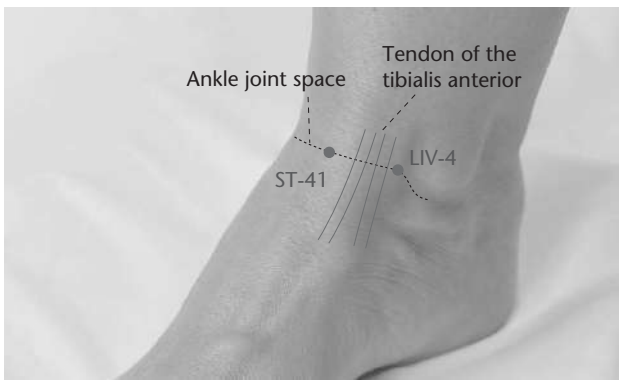

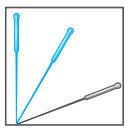
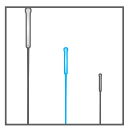

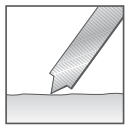

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 170

ST-42 *Rushing Yang* CHONGYANG

Location

On the highest point of the dorsum of the foot, between the tendons of the extensor hallucis longus and the extensor digitorum longus, directly lateral to the point where the dorsalis pedis artery may be palpated. The point is bordered proximally by the 2nd and 3rd metatarsal bones and distally by the 2nd and 3rd cuneiform bones. **Alternative location:** Sometimes, this point may be located lateral to the medial portion of the extensor digitorum longus tendon (joining the 2nd toe).

How to find

With the palpating finger, glide from → **ST-43** (in the depression between the base of the 2nd and 3rd metatarsal bones) proximally towards → **ST-41** (in the depression between the tendons of extensor digitorum longus and hallucis longus). Locate **ST-42** at the highest point of the dorsum of the foot, in a depression lateral to the pulse. In most cases, **ST-42** is located lateral to the tendon of the extensor hallucis longus, which becomes more distinct by extending the big toe.

Needling

Vertically 0.2–0.5 cun. Caution: Dorsalis pedis artery.

Actions/Indications

- Clears Heat from the Stomach channel, regulates the Stomach
- Calms the *shen*
- Opens the channel and *luo* vessels

Special features

Yuan-source point, exit point.

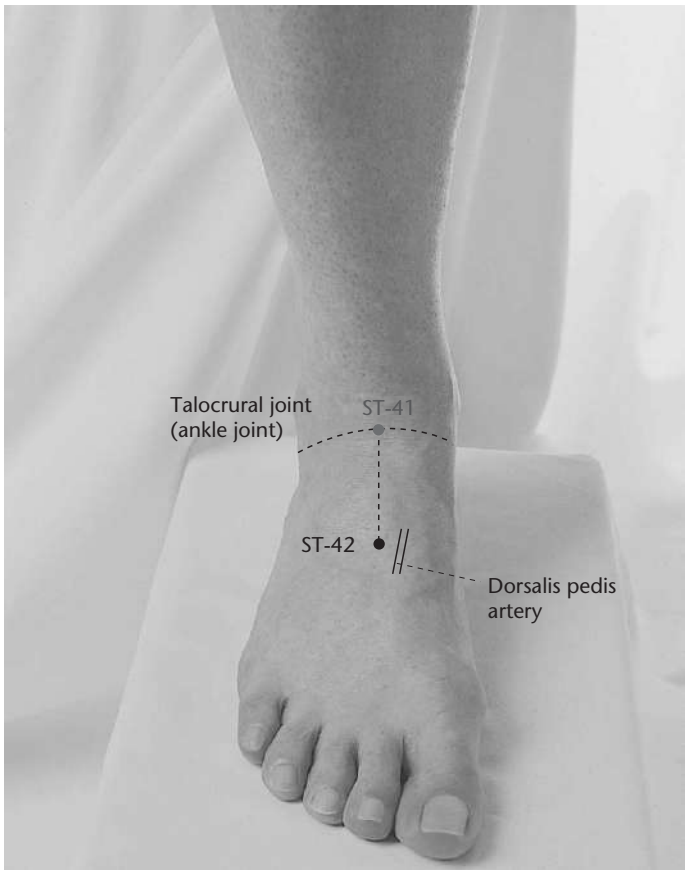

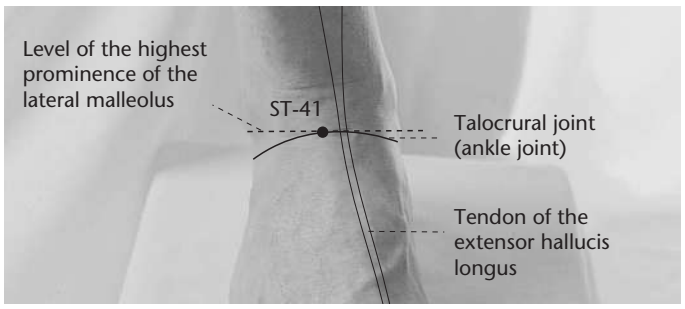

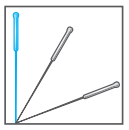
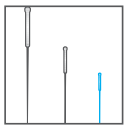

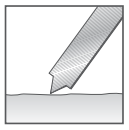

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 171

Sunken Valley XIANGU ST-43

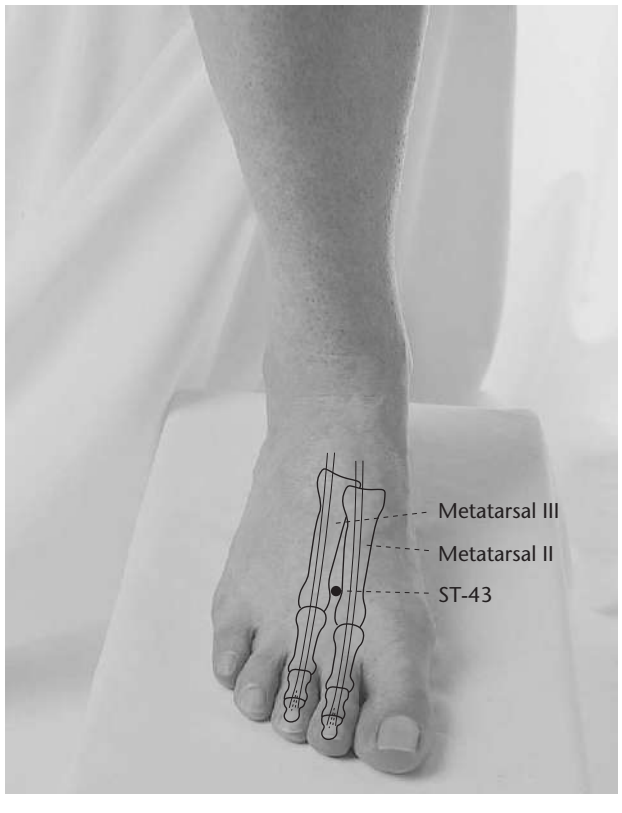

Location

On the dorsum of the foot, in the depression between the 2nd and 3rd metatarsal bones, at the level of the junction of the shafts and the heads of these metatarsal bones.

How to find

Palpate from distal (toes) towards proximal (ankle) between the 2nd and 3rd metatarsal bones, until the finger glides into a depression distal to the metatarsophalangeal joint, where **ST-43** is located. It is approximately on the same level as the junction of the shafts and the heads of the two metatarsal bones.

Needling

Vertically or obliquely 0.5–1 cun

Actions/Indications

- Regulates the Spleen, Stomach and Intestines, reduces oedema

Special features

Shu-stream point, Wood point, generally used for 'inflammatory' joint disorders (*bi*-syndromes).

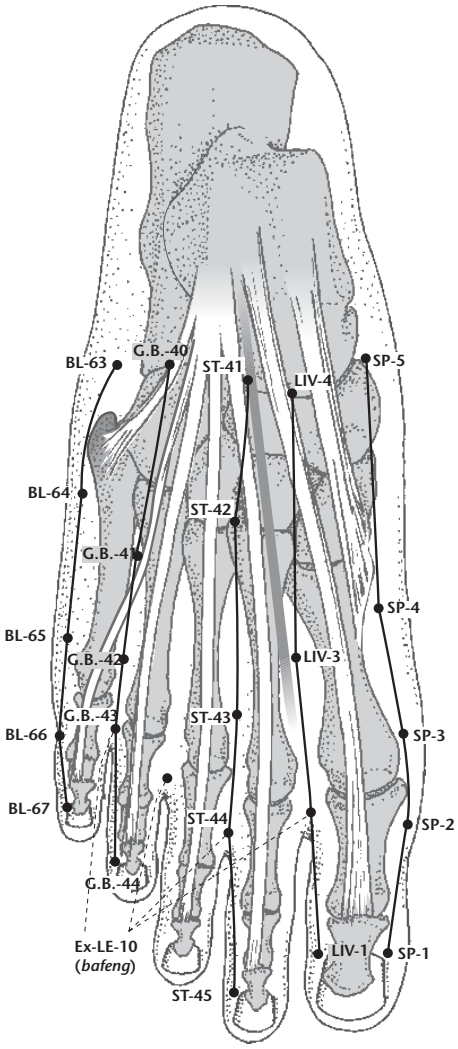

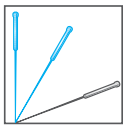
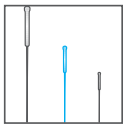

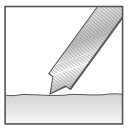
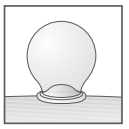

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 172

ST-44 *Inner Court* NEITING

Location
Between the 2nd and the 3rd toes, proximal to the interdigital fold.

How to find
Locate **ST-44** just proximal to the interdigital fold between the 2nd and the 3rd toes.

Needling
0.5–1 cun vertically or obliquely in a proximal direction

Actions/Indications
- Opens the channel, expels Wind and clears Heat, alleviates pain
- Regulates the Intestines, clears Damp-Heat
- Calms the *shen*

Special features
Ying-spring point, Water point, Ma Dan Yang Heavenly Star point. Major point for clearing Heat from the head and face.

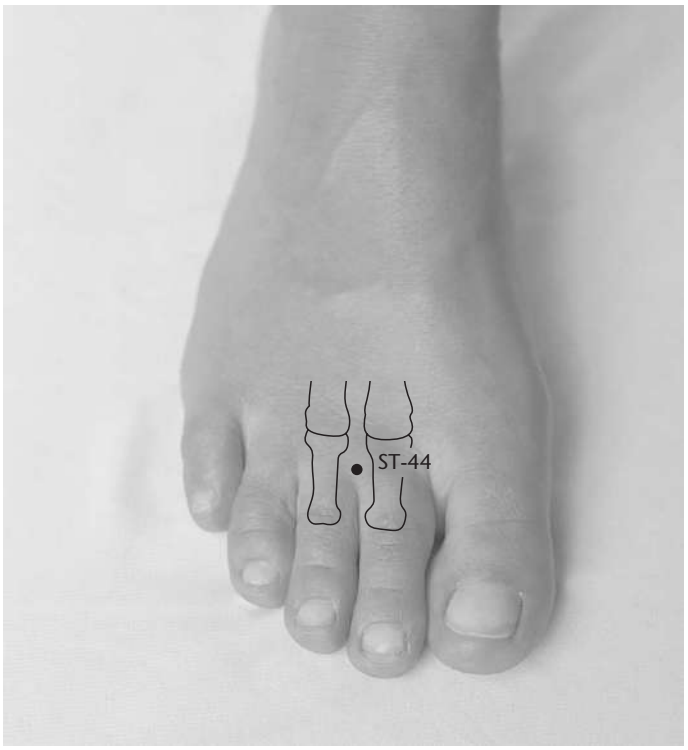

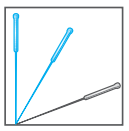

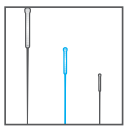

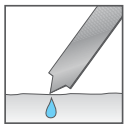

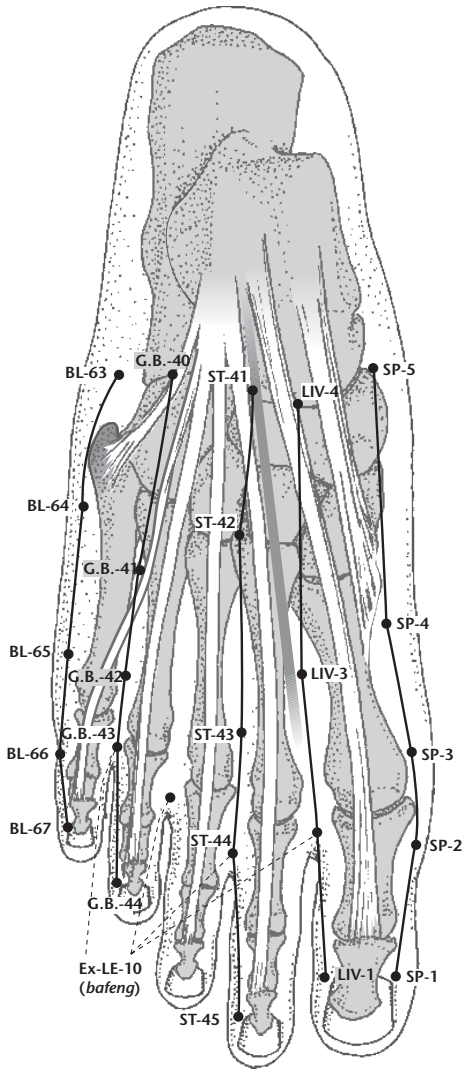

Ch04.3-F10028.qxd 2/22/08 8:43 PM Page 173

Strict Exchange LIDUI ST-45

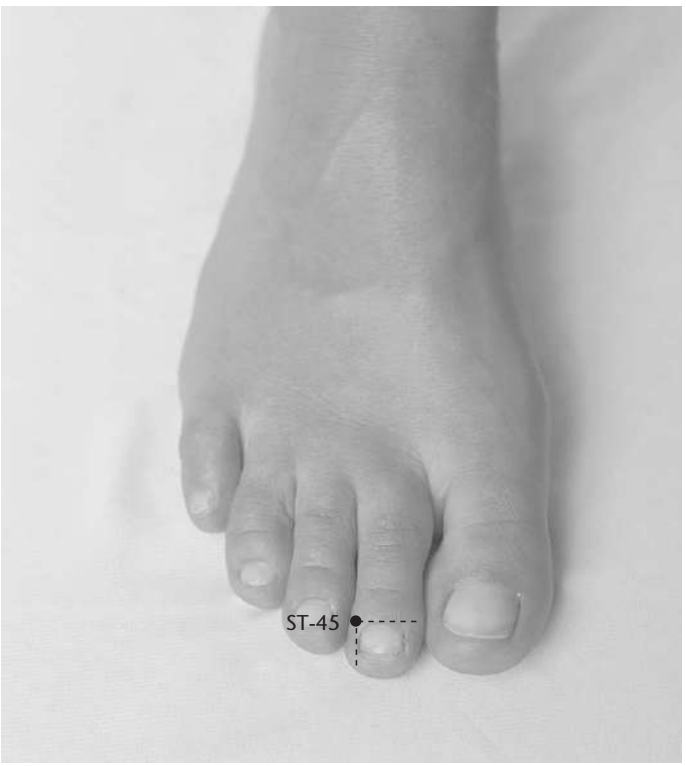

Location

On the 2nd toe, 0.1 cun from the lateral corner of the nail.

How to find

ST-45 is located at the junction of two tangents drawn along the lateral and proximal border of the nail of the 2nd toe, approximately 0.1 cun from the margin of the nail.

Needling

0.1 cun vertically or obliquely in a proximal direction or prick to bleed.

Actions/Indications

- Opens the channel and clears Heat from the *yangming* channels (L.I./ST)
- Calms and regulates the *shen*

Special features

Jing-well point, Metal point, sedation point, important distal point for all Heat disorders affecting the head. For severe insomnia due to Fire or Phlegm-Heat, apply moxibustion ('small Fire attracting big Fire' technique).

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 175

4.4 The Spleen Channel System – Foot *taiyin* (*zu tai yin jing luo*)

4.4.1 The Spleen Primary Channel (*zu tai yin jing*)

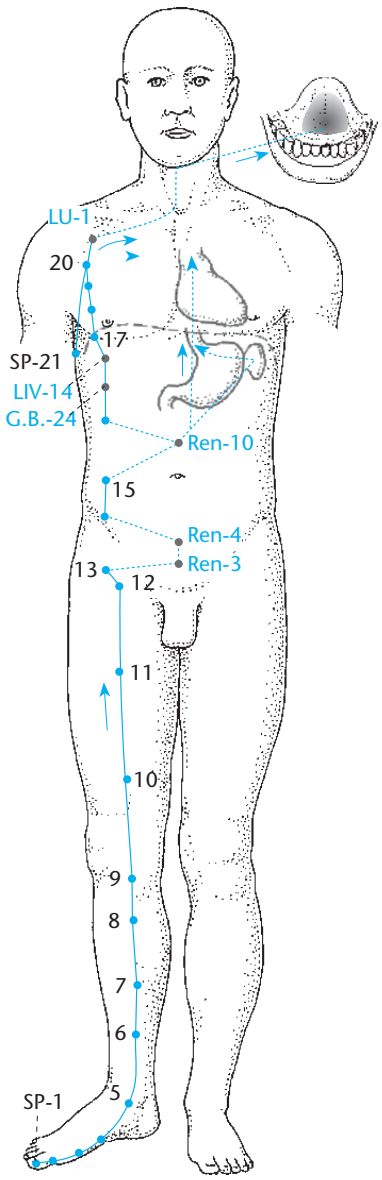

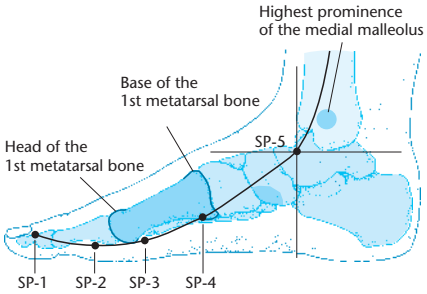

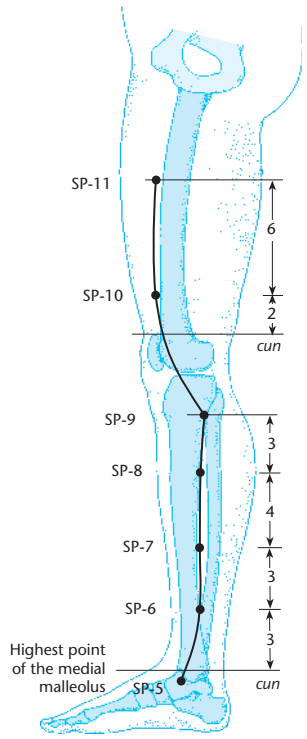

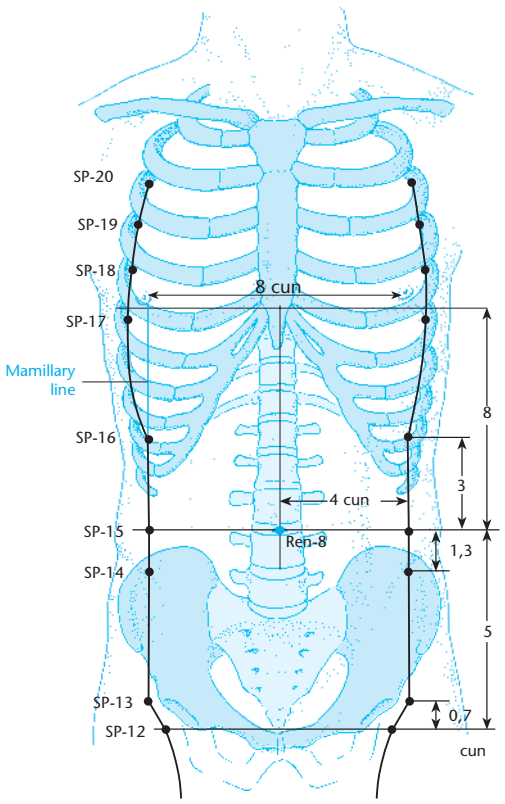

Pathway

The **external pathway** of the Spleen primary channel begins at **SP-1** (*yinbai*) on the big toe at the medial corner of the nail. This point is reached by a branch that separates from the Stomach primary channel at ➞ **ST-42** (*chongyang*) (foot Yin–Yang connection of the first great circuit).

➡ From **SP-1** the channel courses along the medial aspect of the foot

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 176

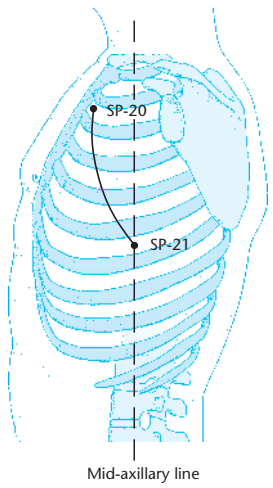

- ascends anterior to the medial malleolus and continues along the posterior border of the medial crest of the tibia
- passes the anteromedial region of the knee and thigh
- ascends the abdomen lateral to the midline to the second intercostal space, passing on its way → **Ren-3** (*zhongji*), → **Ren-4** (*guanyuan*), → **Ren-10** (*xiawan*), → **G.B.-24** (*riyue*) and → **LIV-14** (*qimen*)
- curves downward again and terminates 6 cun inferior to the axilla on the midaxillary line at **SP-21** (*dabao*).

According to Solinas et al (1998), its **inner pathway** branches off at → **Ren-10** (*xiawan*) and enters the Spleen (*pi*), its pertaining *zang*-Organ, and connects with the Stomach (*wei*), its paired *fu*-Organ. From the Stomach (*wei*), an **Interior branch** passes the diaphragm, ascends part of the oesophagus and enters the Heart (*xin*), where it connects with the Heart primary channel (*shou shao yin jing*) (deep Yin–Yin connection → 1.2). From **SP-20** (*zhourong*), a **further branch** ascends to → **LU-1** *(zhongfu)* at the level of the first intercostal space to form the hand–foot pairing (Yin axes of the first great circuit, *taiyin*). The branch ascends further to the tongue, reaches the root of the tongue and disperses on its lower surface.

Clinical importance (→ 1.2)

Exterior (*biao*) signs and symptoms: Heavy sensations of the head and body, sensation of weakness in the legs, general fever, pain in the lower cheek, muscular atrophy and weakness of the legs and medial hip region, cold knees, feet and legs, oedema

Interior (*li*) or *zangfu*-Organ signs and symptoms: Abdominal pain, diarrhoea or loose stools with undigested food, borborygmus, belching and dizziness, abdominal masses, loss of appetite, a sallow complexion, retention of urine

Excess (*shi*) → Spasms, foot pain

Deficiency (*xu*) → Abdominal fullness, borborygmus, diarrhoea with undigested food

Connections and meeting points of the Spleen primary channel (→ 1.2)

Connections with other channels

Stomach primary channel (*zu yang ming jing*)

Connection: Foot Yin–Yang connection of the first great circuit

Location: ST-42 SP-1 (on the foot).

Circulation: Circadian (according to the Organ clock)

Importance: Interior–Exterior relationship

Lung primary channel (*shou tai yin jing*)

Connection: Paired according to the six channel theory (hand–foot pairing) → *tai yin* (Yin axes of the first great circuit)

Location: SP-20 → LU-1 (on the thorax). A branch of the Spleen primary channel ascends from **SP-20** (*zhourong*) to **LU-1** (*zhongfu*)

Circulation: Non-circadian (**not** according to the Organ clock)

Importance: Above–below relationship

Heart primary channel (*shou shao yin jing*)

Connection: Deep Yin–*Yin* connection

Location: SP → HE (on the thorax). An internal branch of the Spleen primary channel spreads inside the Heart (*xin*) and connects with the Heart primary channel.

Circulation: Circadian (according to the Organ clock)

Importance: The Heart primary channel receives Nutritive Qi (*ying qi*) from the Spleen primary channel (first circulation of the *ying qi* → 1.1.4)

Connections with other *zangfu*-Organs

Spleen (*pi*), Stomach (*wei*), Heart (*xin*)

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 177

4.4.2 The Spleen Divergent Channel (*zu tai yin jing bie*)

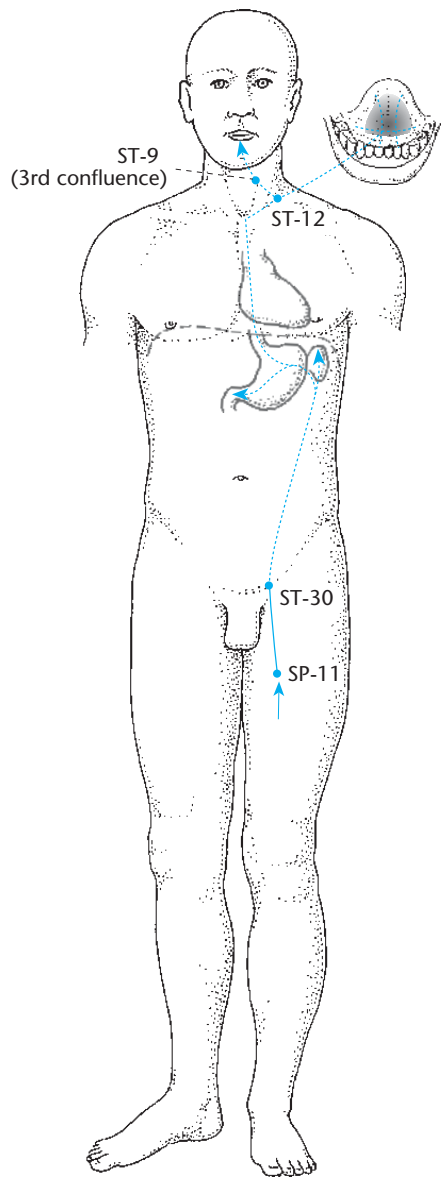

Pathway

The Spleen divergent channel separates from the Spleen primary channel near **SP-11** (*jimen*) on the anteromedial aspect of the thigh, at the medial border of the sartorius muscle. It

- penetrates the abdomen in the inguinal region at → **ST-30** (*qichong*)
- connects with the Stomach (*wei*), Spleen (*pi*) and Heart (*xin*)
- ascends along the oesophagus to the suprasternal fossa and travels laterally to → **ST-12** (*quepen*) in the supraclavicular fossa
- crosses the Stomach primary channel and the Stomach divergent channel at → **ST-9** (*renying*) where they meet to connect at one of the six *he*-confluences (here: ST/SP as third confluence → 1.3).

Clinical importance

- Strengthens the relationship between the Spleen and the Stomach (*zangfu*-Organs). Points on the Spleen primary channel can therefore be used for disorders of the Stomach, and vice versa points on the Stomach primary channel can treat disorders of the Spleen. In clinical practice, Spleen points are used for the whole gastrointestinal tract.
- Meets the Stomach primary channel at → **ST-30**; this is also a Sea of Water and Grain point, which can be used for many digestive disorders.
- Links the Spleen primary channel with the Heart; in combination with other points, SP points can be used for Qi and Blood deficiency as well as for psycho-emotional disorders.
- Connects the Spleen primary channel and the neck region and can therefore be used for chronic disorders of the neck and thyroid caused by retention of Phlegm (for example **SP-3** with **ST-40**, **P-5**).
- Connects the Spleen primary channel with the mouth and tongue and can therefore be used for disorders/ulcers of the tongue and mouth caused by retention of Phlegm (for example **SP-2, SP-3, SP-6**).

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 178

4.4.3 The Spleen Sinew Channel (*zu tai yin jing jin*)

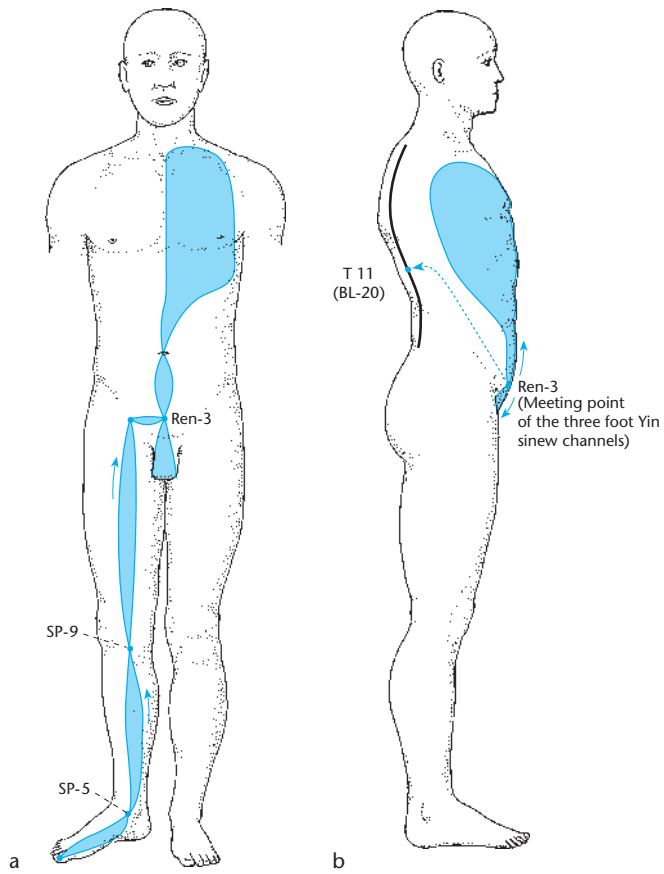

Pathway

The Spleen sinew channel begins on the medial aspect of the big toe at **SP-1** (*yinbai*)

➡ runs along the medial border of the foot

➡ binds (*jie*) anterior to the medial malleolus at **SP-5** (*shangqiu*)

➡ ascends along the medial aspect of the tibia to **SP-9** (*yinlingquan*) and binds (*jie*) anterior to the medial head of the tibia

➡ runs along the medial aspect of the thigh and binds (*jie*) on the medial aspect of the inguinal region

➡ continues to → **Ren-3** (*zhongji*) where it meets the other sinew channels of the foot Yin channels

At → **Ren-3** the sinew channel separates into three branches:

- **One branch** runs to the genital region
- **A further branch** ascends along the anterior midline, covering the *ren mai* and binding (*jie*) at the umbilical region. It penetrates the abdomen and spreads in the costal and thoracic region.
- **A third branch** penetrates deeper into the body and ends at the spinal region at → **BL-20** (*pishu*).

Clinical importance

- **Pathology:** Distending pain of the big toe, pain at the medial malleolus, cramps and pain in the calves, pain of the medial aspect of the knees and thighs, pain in the inguinal and genital region (radiating to the umbilicus, thorax and back).
- **Indication:** Mainly for *bi*-syndromes (Painful Obstruction Syndrome) and *wei*-syndrome (atrophy syndrome) along the Spleen channel, especially along the medial aspect of the leg. The area covered by the Spleen sinew channel is larger than that covered by the Spleen primary channel. This explains why Spleen primary channel points can also be used for disorders and diseases of the external genitalia and the umbilical region.

4.4.4 The Spleen *luo*-Connecting Vessel System (*zu tai yin luo mai*)

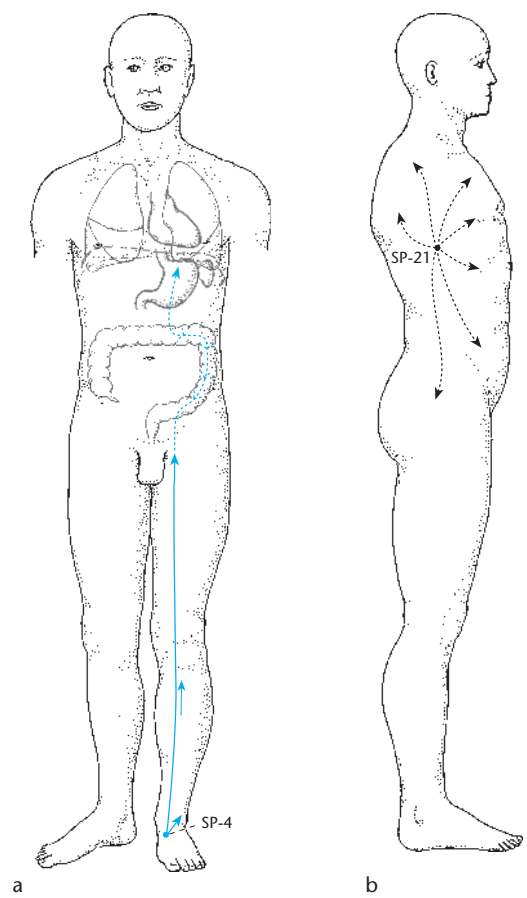

Pathway

The Spleen *luo*-connecting channel separates from the Spleen primary channel at its *luo*-connecting point **SP-4** (*gongsun*) (→ 8.1.2). It forms a three-dimensional reticulate network, dividing into multiple branches and sub-branches (*sun luo, fu luo, xue luo* → 1.5) within the surrounding tissue.

➡ Horizontal divisions run to the Interiorly–Exteriorly paired Stomach primary channel, according to some schools of thought

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 179

(for example Ngyen Van Nghi → Appendix) they travel as a **transverse** Spleen *luo*-connecting vessel to the *yuan*-source point → **ST-42** (*chongyang*).

- A **longitudinal** division follows the pathway of the Spleen primary channel, enters the abdomen in the inguinal region and connects with the Large Intestine (*dachang*) and the Stomach (*wei*).

Clinical importance (→ 8.1.2)

Pathology

- **Excess (*shi*)**: Colicky pain in the Stomach and Large Intestine
- **Deficiency (*xu*)** → *Gu zhang* (drum-like abdominal distension) such as digestive disorders in children, abdominal fullness and distension, ascites
- **Counterflow Qi:** Acute vomiting, diarrhoea, severe abdominal pain

The great *luo*-connecting vessel of the Spleen (*zu tai yin luo mai*)

Pathway

The great *luo*-connecting vessel of the Spleen begins on the lateral aspect of the thorax at → **SP-21** (*dabao*). It contains the Blood of the *luo*-connecting vessel and disperses to the vessels of the whole body, covering them in a net-like fashion (→ 1.5, → 8.1.2).

Clinical importance (→ 8.1.2)

- **Excess (*shi*)**: Pain in the whole body, polyarthritis, *bi*-Syndrome (Painful Obstruction Syndrome)
- **Deficiency (*xu*)**: Muscular atrophy and weakness, weakness of the joints, weakness in the whole body.

4.4.5 Cutaneous Region (*tai yin pi bu*)

See description and figures → 1.6

4.4.6 Points of the Spleen Primary Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): SP-3** (*taibai*) ■■
- ***Luo*-connecting point (→ 8.1.2.): SP-4** (*gongsun*) ■■/ SP-21 (*dabao*) ■■
- ***Xi*-cleft point (→ 8.1.3): SP-8** (*diji*) ■■
- **Associated Back-*shu* point (→ 8.1.4): BL-20** (*pishu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): LIV-13** (*zhangmen*) ■■
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Wood): **SP-1** (*yinbai*) ■■
  ying-spring point (Fire), tonification point: **SP-2** (*dadu*)
  shu-stream point (Earth), *ben* point: **SP-3** (*taibai*) ■■
  jing-river point (Metal), sedation point: **SP-5** (*shangqiu*) ■■
  he-sea point (Water): **SP-9** (*yinlingquan*) ■■
- ***Hui*-meeting point (→ 8.1.7): –**
- **Opening point (→ 8.1.8) of the *chong mai*: SP-4** (*gongsun*) ■■
- **Lower *he*-sea points (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the Kidney and Liver primary channels (meeting point of the three foot *Yin* channels and the *yin qiao mai**: **SP-6** (*sanyinjiao*) ■■
  – with the LIV and T.B.* channels and the *yin wei mai:* **SP-12** (*chongmen*)
  – with the LIV channel and the *yin wei mai*: **SP-13** (*fushe*)
  – with the *yin wei mai*: **SP-15** (*daheng*), **SP-16** (*fuai*)
  – of other channels with the SP channel: **LU-1**, **G.B.-24**, **LIV-14**, **Ren-3**, **Ren-4**, **Ren-10**, **Ren-17***
- **Gao Wu command point (→ 8.1.11):** –
- **Window of Heaven point (→ 8.1.12):** –
- **Points of the Four Seas (→ 8.1.13):** –
- **Ma Dan Yang Heavenly Star point(→ 8.1.14):** –
- **Sun Si Miao Ghost point (→ 8.1.15): SP-1** (*yinbai*) ■■
- **Other functional points:**
  – important point for diarrhoea: **SP-4** (*gongsun*) ■■
  – important point for Damp *bi*-syndrome: **SP-5** (*shangqiu*) ■■
  – important point for eliminating Dampness: **SP-9** (*yinlingquan*) ■■
  – important point for regulating the Blood: **SP-10** (*xuehai*) ■■
  – important point for general pain: **SP-21** (*dabao*) ■■

Points according to region

- **Local points (→ 8.2.1):** knee region – **SP-9** (*yinlingquan*) ■■; Large Intestine – **SP-15** (daheng) ■■; knee – **SP-9** (*yinlingquan*) ■■; foot – **SP-5** (*shangqiu*) ■■; toes – **SP-3** (*taibai*) ■■
- **Adjacent points (→ 8.2.1):** Spleen/Stomach – **SP-15** (*daheng*) ■■; hypochondrium – **SP-21** (*dabao*) ■■; knee – **SP-10** (*xuehai*) ■■; toes – **SP-4** (*gongsun*) ■■, **SP-5** (*shangqiu*) ■■, **SP-9** (*yinlingquan*) ■■
- **Distal points (→ 8.2.1):** for the Spleen/Stomach, upper abdominal region – **SP-4** (*gongsun*) ■■; for the female genital organs, urogenital region – **SP-6** (*sanyinjiao*) ■■; lower abdomen – **SP-1** (*yinbai*) ■■; for the hand and knee – **SP-5** (*shangqiu*) ■■

Specific points according to the channel pathway (in numerical order):

- **SP-1** (*yinbai*) ■■: Sun Si Miao Ghost point (→ 8.1.15); distal point for the lower abdomen (→ 8.2.1); *jing*-well point (Wood) (→ 8.1.6)
- **SP-2** (*dadu*): *ying*-spring point (Fire) (→ 8.1.6), tonification point
- **SP-3** (*taibai*) ■■: *yuan*-source point (→ 8.1.1), *shu*-stream point (→ 8.1.6), *ben* point (Five Phase point) (→ 8.1.6)

* Mentioned by only some authors.

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 180

- **SP-4** (*gongsun*) ■■: *luo*-connecting point of the Spleen (→ 8.1.2); opening point of the *chong mai* (→ 8.1.8); distal point for the Spleen/Stomach, upper abdomen (→ 8.2.1); important point for diarrhoea; adjacent point for the toes (→ 8.2.1)
- **SP-5** (*shangqiu*) ■■: *jing*-river point (Metal) (→ 8.1.6); Sedation point; important point for Damp *bi*-syndrome; local point for the foot (→ 8.2.1); distal point for the hand and knee (→ 8.2.1)
- **SP-6** (*sanyinjiao*) ■■: *jiaohui*-meeting point with the KID and LIV channels and the *yin qiao mai* (→ 8.1.10) (meeting point of the three foot Yin channels); distal point for the female genital system, the urogenital region (→ 8.2.1)
- **SP-8** (*diji*) ■■: *xi*-cleft point (→ 8.1.3)
- **SP-9** (*yinlingquan*) ■■: *he*-sea point (Water) (→ 8.1.6); important point for eliminating Dampness; local point for the knee region (→ 8.2.1); local point for the knees (→ 8.2.1)
- **SP-10** (*xuehai*) ■■: important point for regulating the Blood; adjacent point for the knees (→ 8.2.1)
- **SP-12** (*chongmen*): *jiaohui*-meeting point with the LIV and T.B.* channels and the *yin wei mai* (→ 8.1.10)
- **SP-13** (*fushe*): *jiaohui*-meeting point with the LIV channel and the *yin wei mai* (→ 8.1.10)
- **SP-15** (*daheng*) ■■: local point for the Large Intestine (→ 8.2.1); adjacent point for the Spleen/Stomach (→ 8.2.1); *jiaohui*-meeting point with the *yin wei mai* (→ 8.1.10)
- **SP-16** (*fuai*): *jiaohui*-meeting point with the *yin wei mai* (→ 8.1.10)
- **SP-21** (*dabao*) ■■: adjacent point for the hypochondrium (→ 8.2.1); Main-*luo*-connecting point of the Spleen (→ 8.1.2); important point for generalised pain

General location help

- **SP-13 – SP-16**: these points are all located 4 cun lateral to the anterior midline (= mamillary line) → **SP-13**: 0.7 cun superior to **SP-12; SP-14**: 1.3 cun inferior to the umbilicus; **SP-15**: at the level of the umbilicus; **SP-16**: 3 cun superior to the umbilicus.
- **SP-17 – SP-20**: these points are all located 6 cun lateral to the anterior midline in the fifth to second intercostal spaces.

*Mentioned by only some authors.

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 181

Hidden White YINBAI SP-1

Location
On the big toe, 0.1 cun from the medial corner of the nail.

How to find
SP-1 is located at the junction of two tangents drawn along the medial and proximal border of the nail of the big toe, approximately 0.1 cun from the margin of the nail.

Needling
0.1–0.2 cun vertically or obliquely in a proximal direction or prick to bleed with a needle, lancet or three-edged needle.

Actions/Indications
- Regulates the Blood
- Regulates the Spleen
- Unbinds the chest
- Calms the *shen*
- Revives consciousness

Special features
Jing-well point, Wood point, Sun Si Miao Ghost point, entry point, important distal point for the lower abdomen.

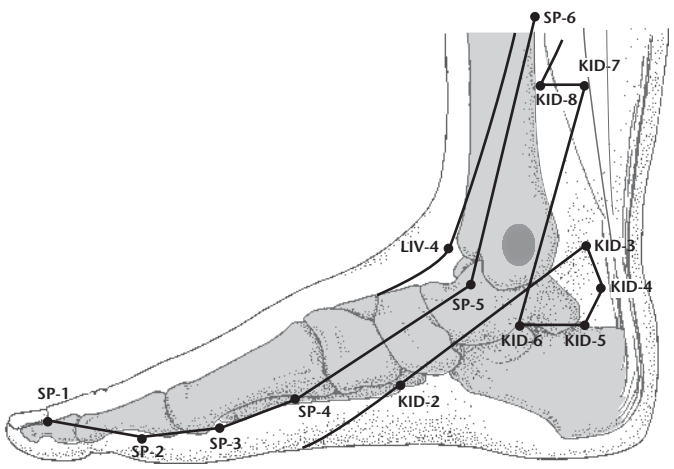

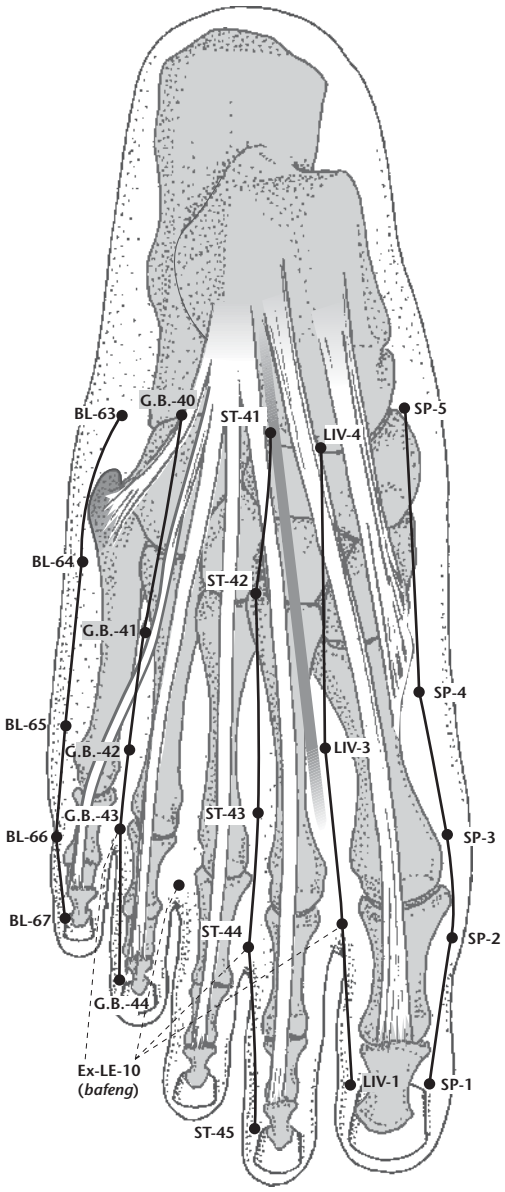

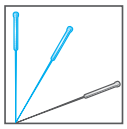
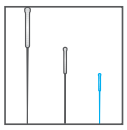

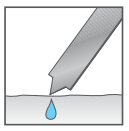
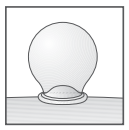

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 182

SP-2 *Great Metropolis* DADU

Location

On the medial aspect of the big toe, at the junction of the shaft and the base of the proximal phalanx, distal to the 1st metatarsophalangeal joint.

How to find

On the medial aspect of the big toe, palpate from distal to proximal towards the metatarsophalangeal joint, until you can feel the junction of the shaft with the base of the proximal phalanx. **SP-2** is located distally to the base of the bone, which forms a distinct step, and slightly inferiorly to its outer curvature, at the border of the red and white skin (sole/dorsum of the foot).
Located in comparable positions are → **BL-66** on the lateral border of the foot, → **L.I.-2** on radial aspect and → **S.I.-2** on the ulnar aspect of the hand.

Needling

0.2–0.5 cun vertically or obliquely, just below the lower border of the bone. Caution: Painful point.

Actions/Indications

- Regulates and supports the Spleen, harmonises the Lower Burner and clears Heat

Special features

Ying-spring point, Fire point, tonification point.

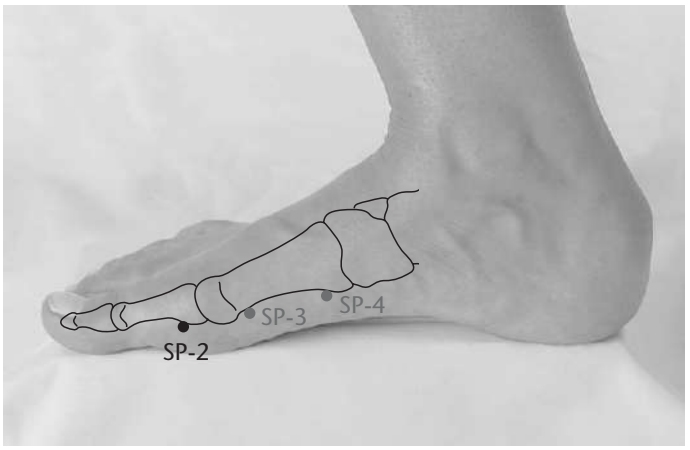

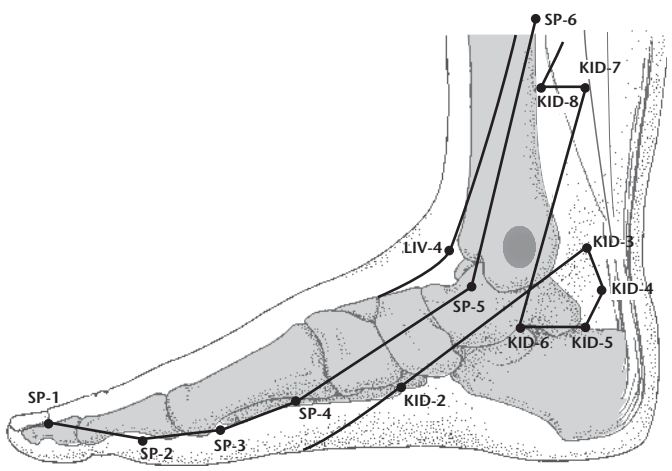

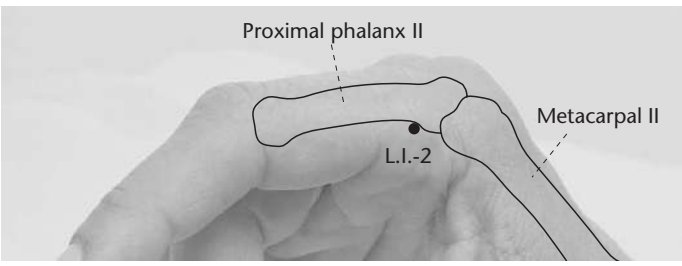

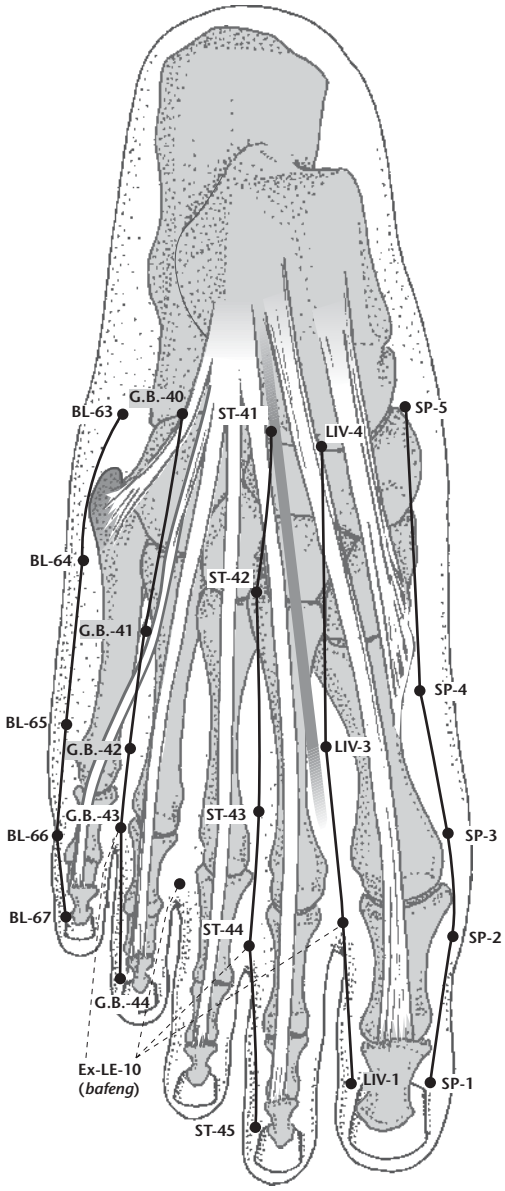

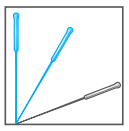
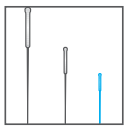

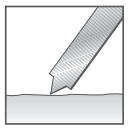

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 183

Supreme White TAIBAI SP-3

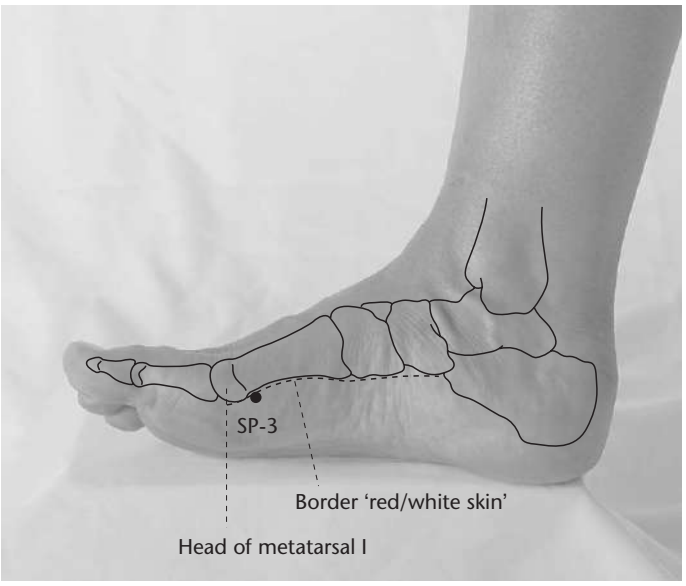

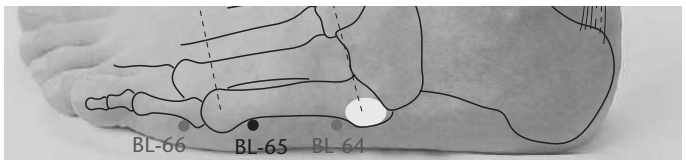

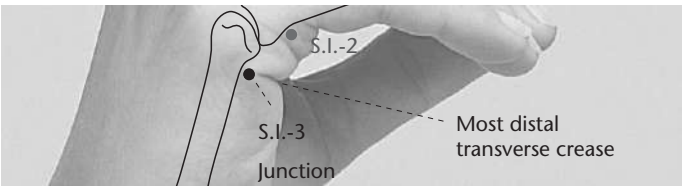

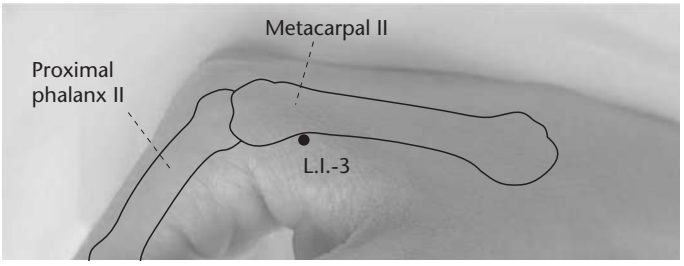

Location

On the medial aspect of the foot, in the depression proximal to the head of the 1st metatarsal bone, at the border of the red and white skin.

How to find

When palpating along the border of the red and white skin on the medial aspect of the foot from distal (toes) to proximal (ankle), the metatarsophalangeal joint forms a prominent landmark. **SP-3** is located just proximal to this joint in a clearly palpable depression, which marks the junction of the head and the shaft of the 1st metatarsal bone.

Located in a comparable position is → **BL-65** (proximal to the head of the 5th metatarsal bone) on the lateral border of the foot. In comparable locations on the borders of the hand are → **L.I.-3** on the radial aspect and → **S.I.-3** on the ulnar aspect of the hand.

Needling

0.5–1 cun vertically or obliquely from medial towards lateral. Caution: Needling this point tends to be painful.

Actions/Indications

- Tonifies the Spleen, harmonises the Lower Burner
- Eliminates Dampness and Damp-Heat
- Regulates Qi locally

Special features

Yuan-source point, *shu*-stream point, Earth point, *ben* point (Five Phases). Important point for tonifying the Spleen.

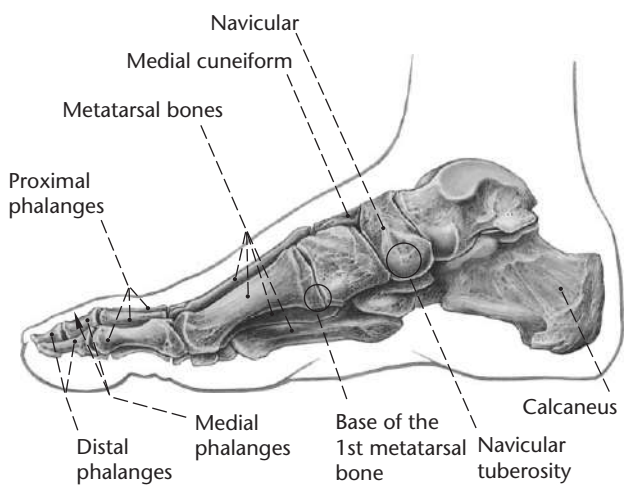

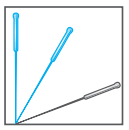
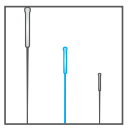

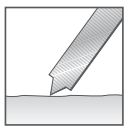
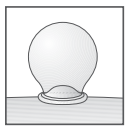

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 184

SP-4 *Grandfather Grandson* GONGSUN

Location

In the depression distal to the base of the 1st metatarsal bone, on the border of the red and white skin.

How to find

When palpating along the border of the red and white skin (sole/dorsum of the foot) on the medial aspect of the midfoot from proximal towards medial, the base of the 1st metatarsal bone (→ 2.6.2) forms a distinct bony prominence. **SP-4** is located in the depression distal to this prominence, at the junction of the shaft and the base of the bone.

→ **SP-3** is located slightly more distally at the junction of the head and the shaft of the 1st metatarsal bone. → **BL-64** is located in a comparable position on the lateral aspect of the midsection of the foot (at the junction of the shaft and the base of the 5th metatarsal bone).

Needling

0.5–1 cun vertically or obliquely from medial towards lateral. Caution: Needling may be painful.

Actions/Indications

- Strengthens and regulates the Spleen Qi, eliminates Dampness
- Calms the *shen*
- Benefits the Heart and chest
- Regulates the *chong mai*
- Local point, for example for metatarsalgia

Special features

Luo-connecting point, opening (master) point of the *chong mai*. Important point for distension and pain in the upper and lower abdomen caused by Qi stagnation, Blood stasis or obstruction due to Dampness.

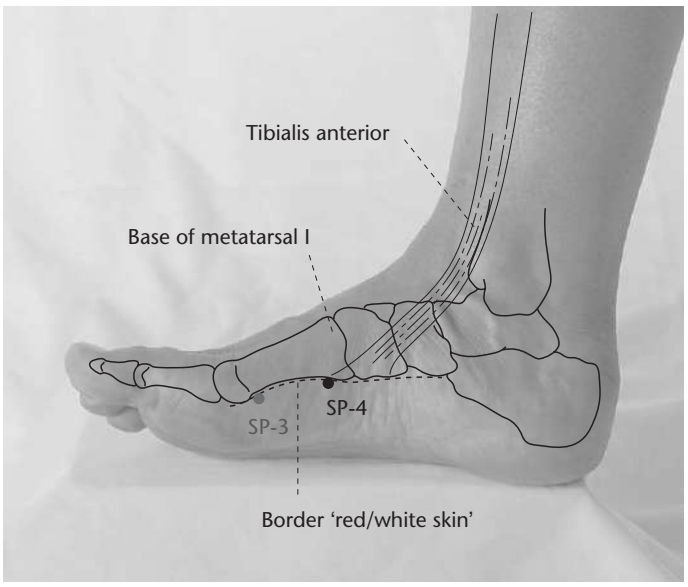

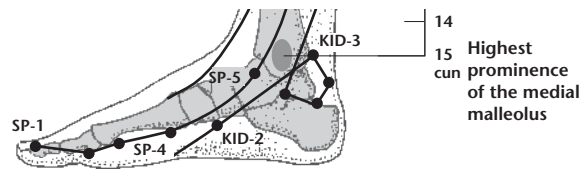

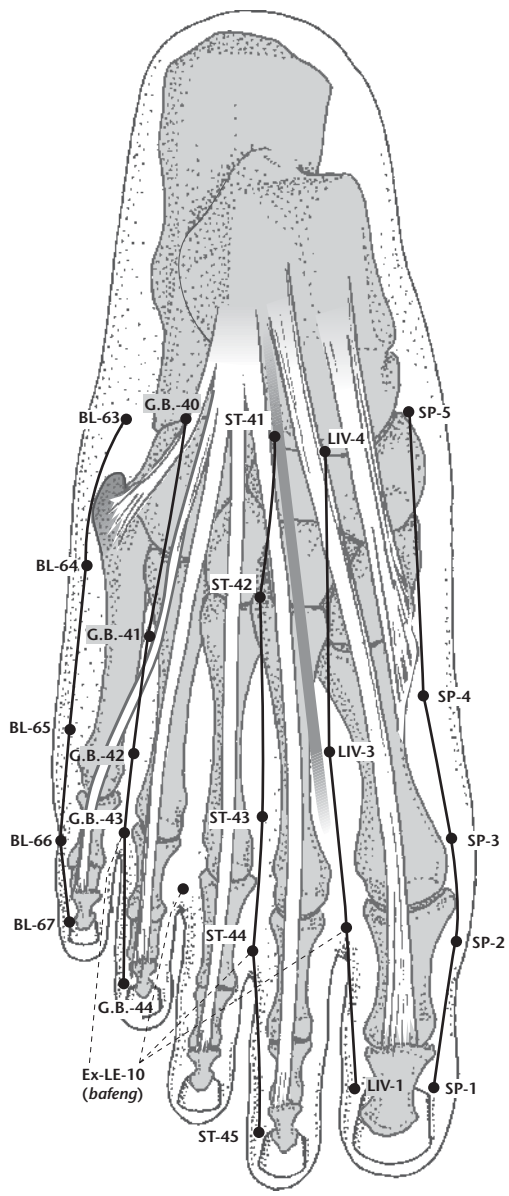

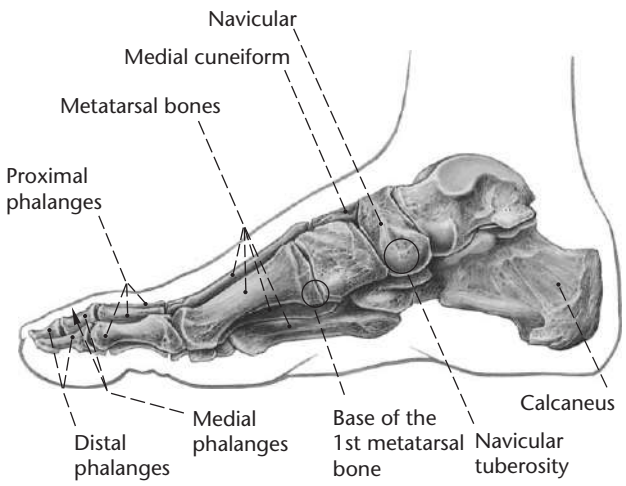

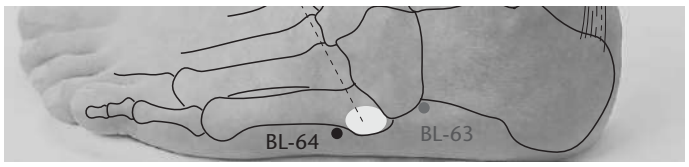

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 185

Shang Mound SHANGQIU SP-5

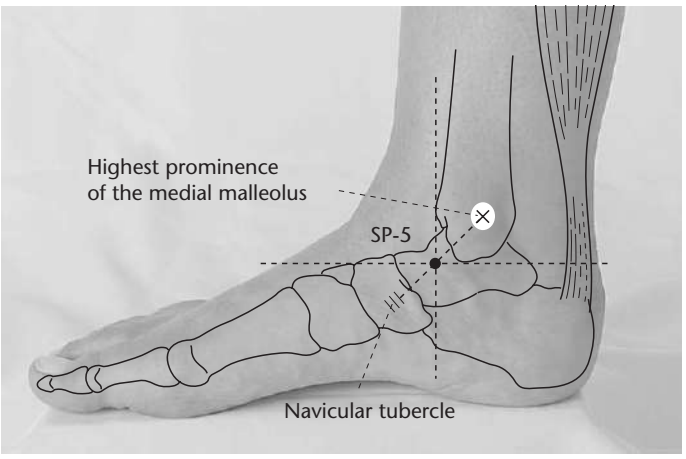

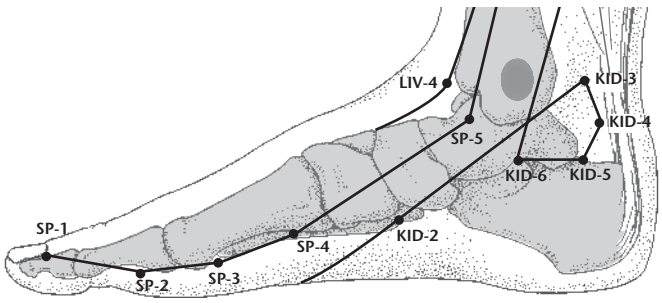

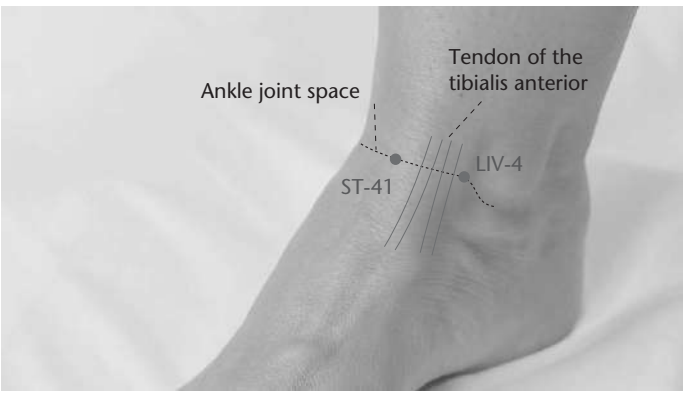

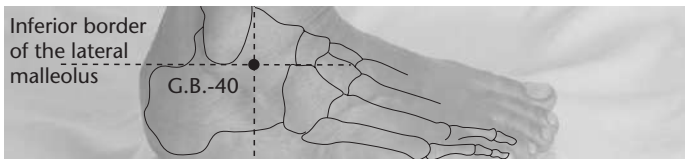

Location

In the depression at the junction of a vertical line drawn along the anterior border and a horizontal line drawn along the lower border of the medial malleolus. **Or**: In the depression halfway between the highest prominence of the medial malleolus and the tubercle of the navicular bone.

How to find

SP-5 is located in a depression anterior and inferior to the medial malleolus, at the junction of a vertical line drawn along the anterior border and a horizontal line drawn along the inferior border of the medial malleolus. **SP-5** is located distal to the tendon of the tibialis anterior muscle.

Located in a comparable position on the lateral aspect of the foot is → **G.B.-40** (in a depression at the junction of a vertical line drawn along the anterior border and a horizontal line drawn along the inferior border of the lateral malleolus).

Needling

0.3–0.6 cun vertically, obliquely or transversely (subcutaneously) deep to the tendons towards → **ST-41**.

Actions/Indications

- Tonifies the Spleen, eliminates Dampness, benefits the tendons and bones
- Moves Qi locally
- Calms the *shen*

Special features

Jing-river point, Metal point, sedation point. Important point for treating Damp *bi*-syndrome.

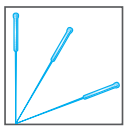
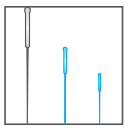

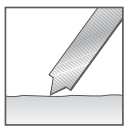

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 186

SP-6 *Three Yin Intersection* SANYINJIAO

Location

3 cun proximal to the highest prominence of the medial malleolus, on the posterior border of the medial crest of the tibia.

How to find

From the highest prominence of the medial malleolus (→ 3.6.2), measure 3 cun (1 handbreadth) in a superior direction, where **SP-6** is located in a depression at the posterior border of the medial crest of the tibia. This point is often sensitive to pressure. Sometimes, this depression is located more anteriorly, and locating **SP-6** should be based on pressure-sensitivity.
Located in a comparable position is → **G.B.-39** (3 cun proximal to the prominence of the lateral malleolus, on the anterior border of the fibula, intersecting point of the three foot Yang channels).

Needling

Vertically or obliquely 1–2 cun. Caution: Contraindicated during pregnancy, particularly with reducing needle techniques; exception: to promote labour.

Actions/Indications

- Tonifies the Spleen and the Stomach, eliminates Dampness
- Nourishes the Blood and Yin, regulates menstruation, promotes labour
- Regulates urination, benefits the genitals, harmonises the Lower Burner
- Calms the *shen*

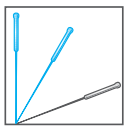

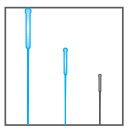

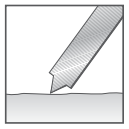

Special features

Meeting point with the Liver and Kidney channels (group *luo*-connecting point) and the *yin qiao mai*. Very important point for disorders of the Lower Burner (especially for gynaecological disorders, birth and urinary disorders), as well as for many digestive disorders and psychological disorders (caused by Blood and Yin deficiency).

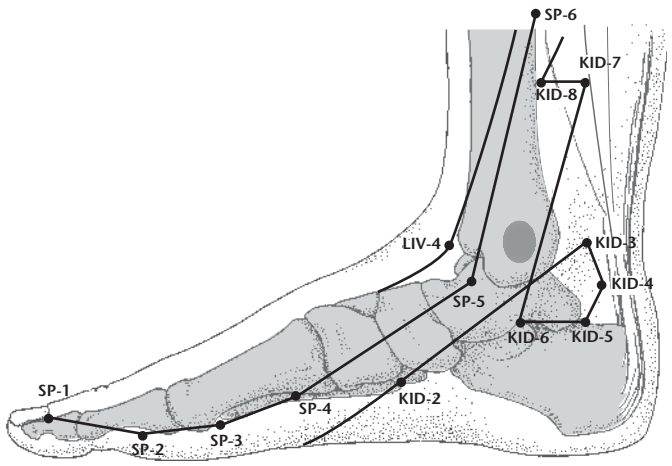

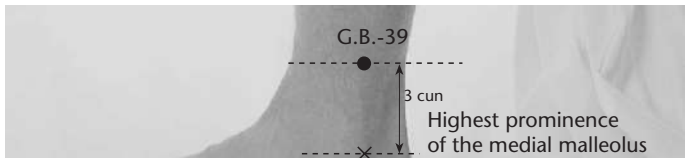

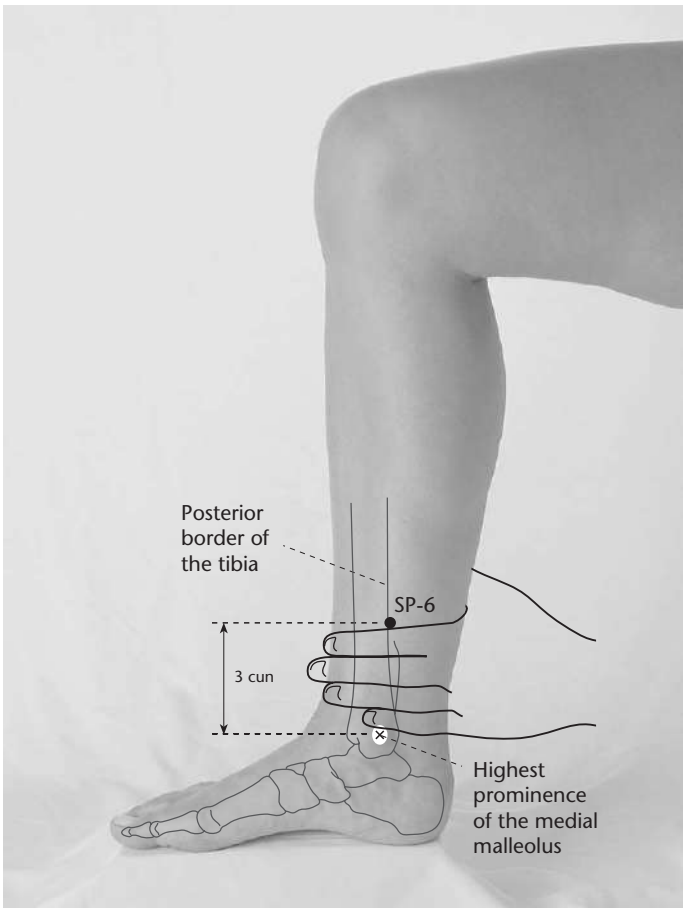

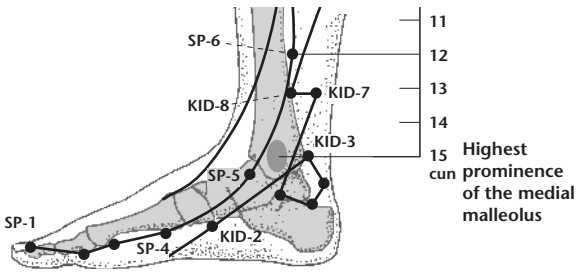

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 187

Dripping Valley LOUGU SP-7

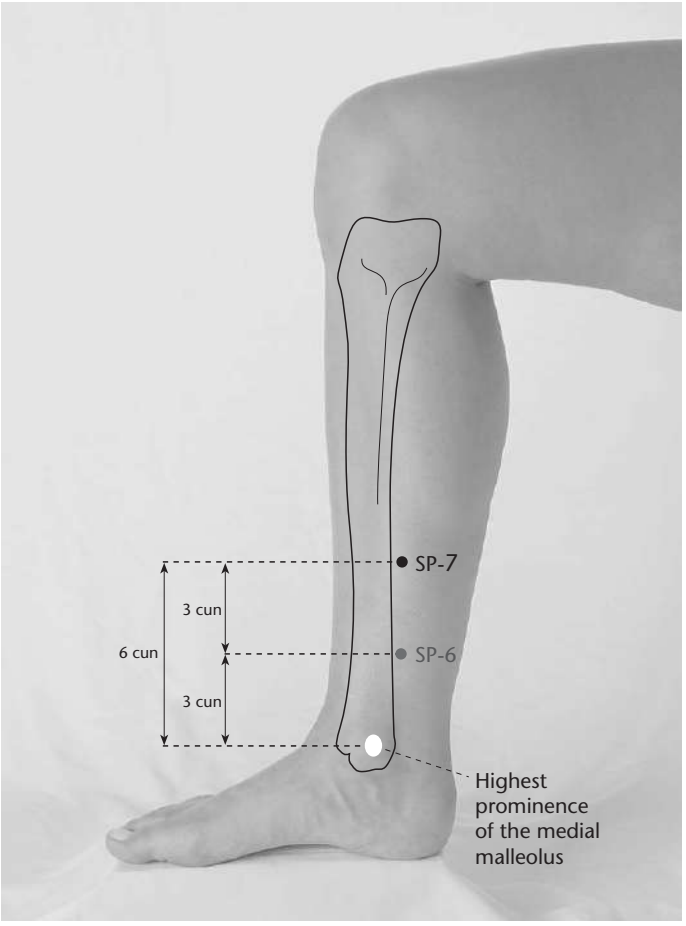

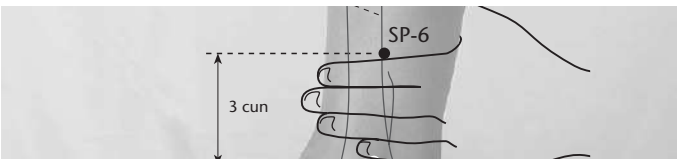

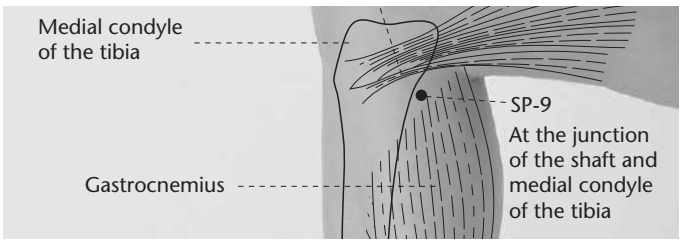

Location

6 cun proximal to the highest prominence of the medial malleolus, on the posterior border of the medial crest of the tibia.

How to find

From the highest prominence of the medial malleolus, measure 6 cun in a superior direction, where **SP-7** is located at the posterior border of the medial crest of the tibia. **Or**: Spreading hands technique (→ 2.3.3): **SP-7** is located 0.5 cun distal to the midpoint of a line connecting → **SP-9** (distal to the tibial condyle) and the highest prominence of the medial malleolus.
→ **SP-6** is located midway between the tip of the malleolus and **SP-7** (3 cun superior/inferior respectively).

Needling

Vertically 1–1.5 cun

Actions/Indications

- Strengthens the Spleen, drains Dampness, opens the channel
- Promotes urination

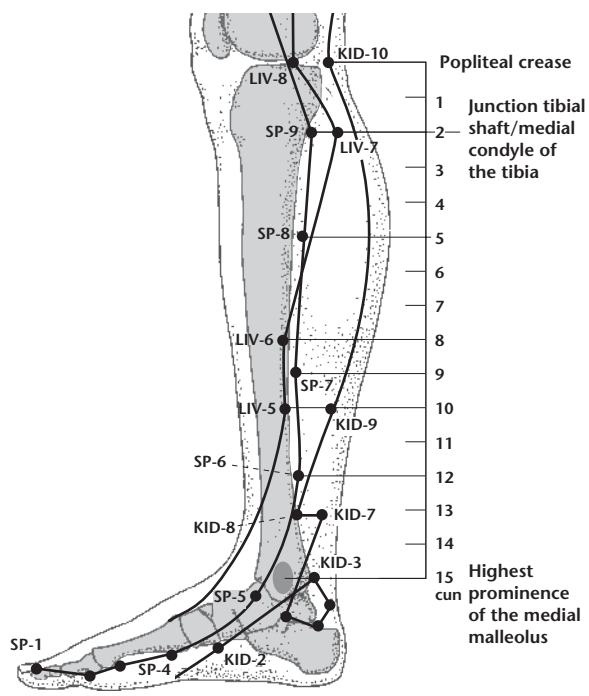

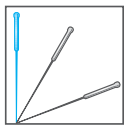
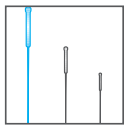

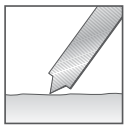

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 188

SP-8 *Earth Pivot* DIJI

Location

3 cun distal to the junction of the shaft and the medial condyle of the tibia (→ **SP-9**), at the posterior border of the medial crest of the tibia.

How to find

Measure 1 handbreadth directly inferior to the medial condyle of the tibia (→ **SP-9**). **SP-8** is located 3 cun distal to → **SP-9** on the posterior border of the medial crest of the tibia, on a line connecting → **SP-9** and the highest prominence of the medial malleolus.

Needling

Vertically 1–1.5 cun

Actions/Indications

- Regulates the uterus and the Blood
- Regulates the Spleen and eliminates Dampness

Special features

Xi-cleft point. Important point for acute dysmenorrhoea, often used in combination with → **L.I.-4**.

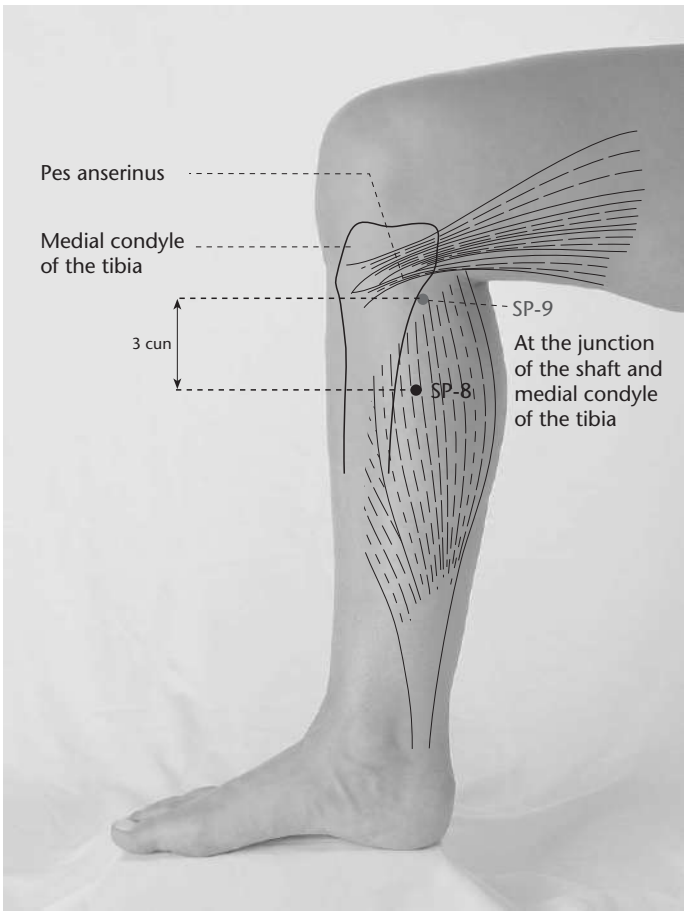

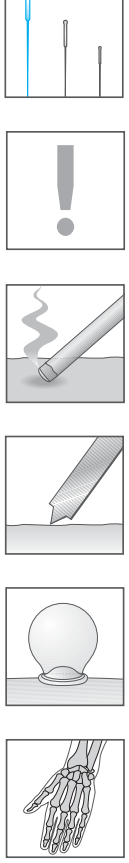

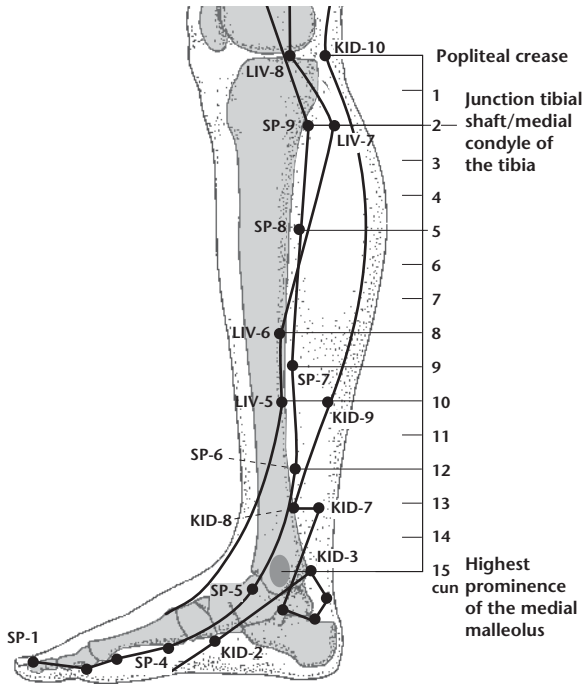

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 189

Yin Mound Spring YINLINGQUAN SP-9

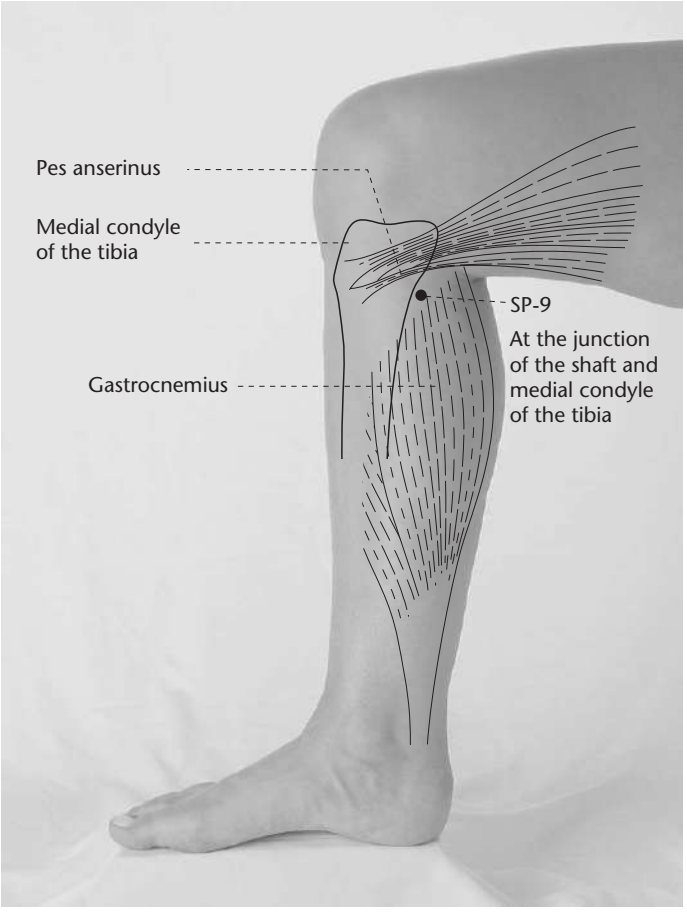

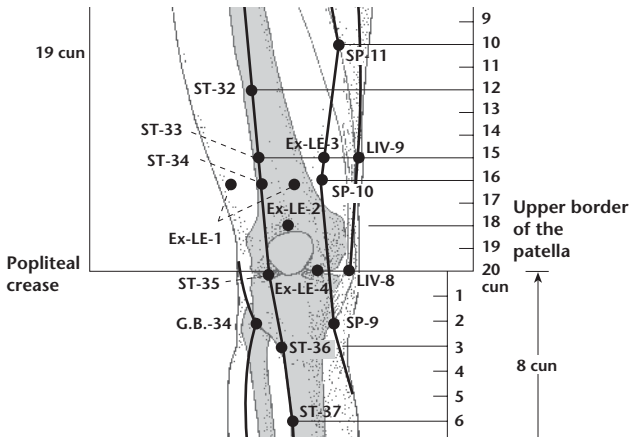

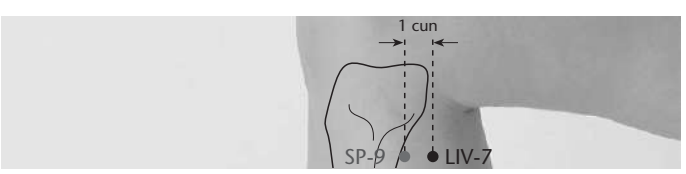

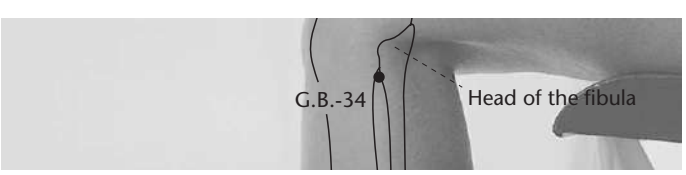

Location

With the knee flexed, this point is located in a depression distal to the medial condyle of the tibia, at the junction of the shaft and the medial condyle.

How to find

Patient's position: Preferably locate and needle **SP-9** with the knee bent and the hip slightly externally rotated, so that the knee slightly lifts from the surface, or rest the knee in an externally rotated position on a knee support. **Location**: Palpate along the posterior border of the medial crest of the tibia towards the head of the tibia to locate **SP-9** in a pressure-sensitive depression at the junction of the shaft and the head of the tibia. This depression is located between the pes anserinus superficialis and the belly of the gastrocnemius muscle (→ 3.6.1).
Located on the same level but 1 cun posterior to **SP-9** is → **LIV-7**. Located on the lateral aspect of the lower leg is → **G.B.-34** (anterior and inferior to the head of the fibula).

Needling

1–1.5 cun vertically, along the posterior border of the tibia.

Actions/Indications

- Regulates the Spleen, transforms Dampness, regulates the water passages, benefits the Lower Burner
- As local point for disorders of the knee, especially if there are swellings

Special features

He-sea point, Water point. One of the major points for eliminating Dampness anywhere in the body. Important local point for disorders of the knee. **SP-9** is located approximately on the same level as → **G.B.-34**. If indicated, this point tends to be very tender upon pressure.

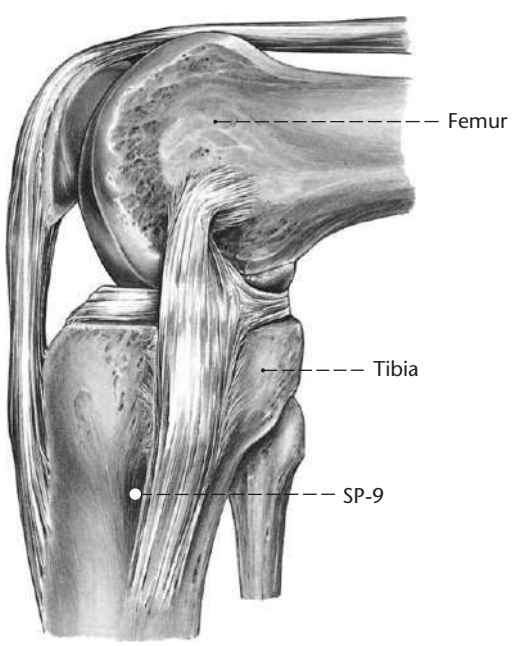

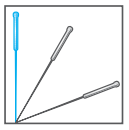
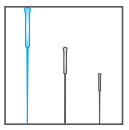

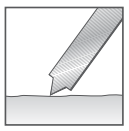
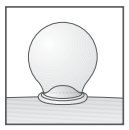

Ch04.4-F10028.qxd 2/23/08 7:08 PM Page 190

SP-10 *Sea of Blood* XUEHAI

Location

With the knee flexed, **SP-10** is located 2 cun proximal and slightly medial to the medial superior border of the patella, in a depression on the vastus medialis muscle.

How to find

Measure 2 cun proximal to the medial superior border of the patella and locate **SP-10** by palpating slightly medially for a soft depression. **Or**: The practitioner places their left hand on the patient's right patella, with the heel of the palm level with the inferior patellar border and the extended fingers pointing in a superior direction. With the thumb and index finger forming an angle of 45°, the tip of the thumb will point to **SP-10**.

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Strengthens the Blood, dispels Blood stasis, cools the Blood, stops bleeding, regulates menstruation, benefits the skin
- Local point for disorders of the knee

Special features

Important point for regulating the Blood

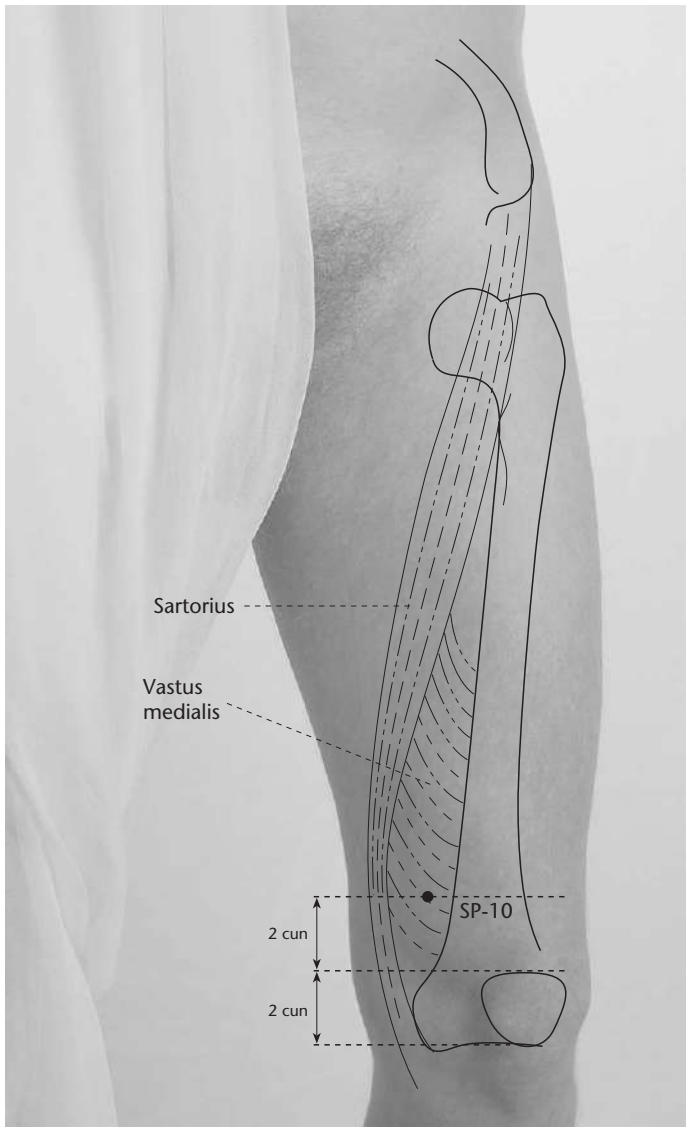

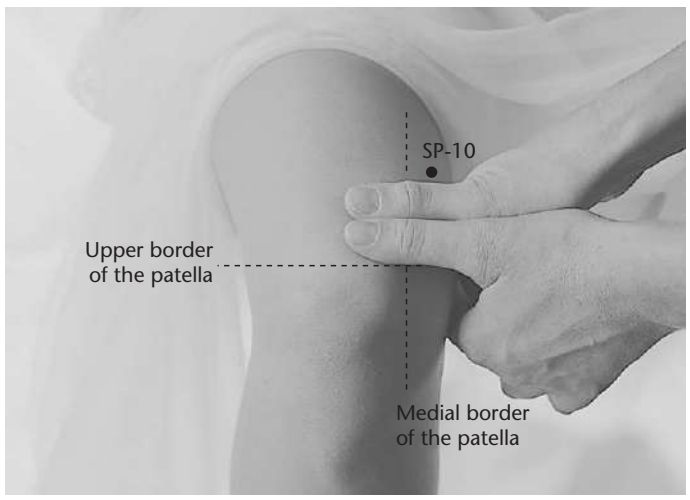

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 191

Winnowing Gate JIMEN SP-11

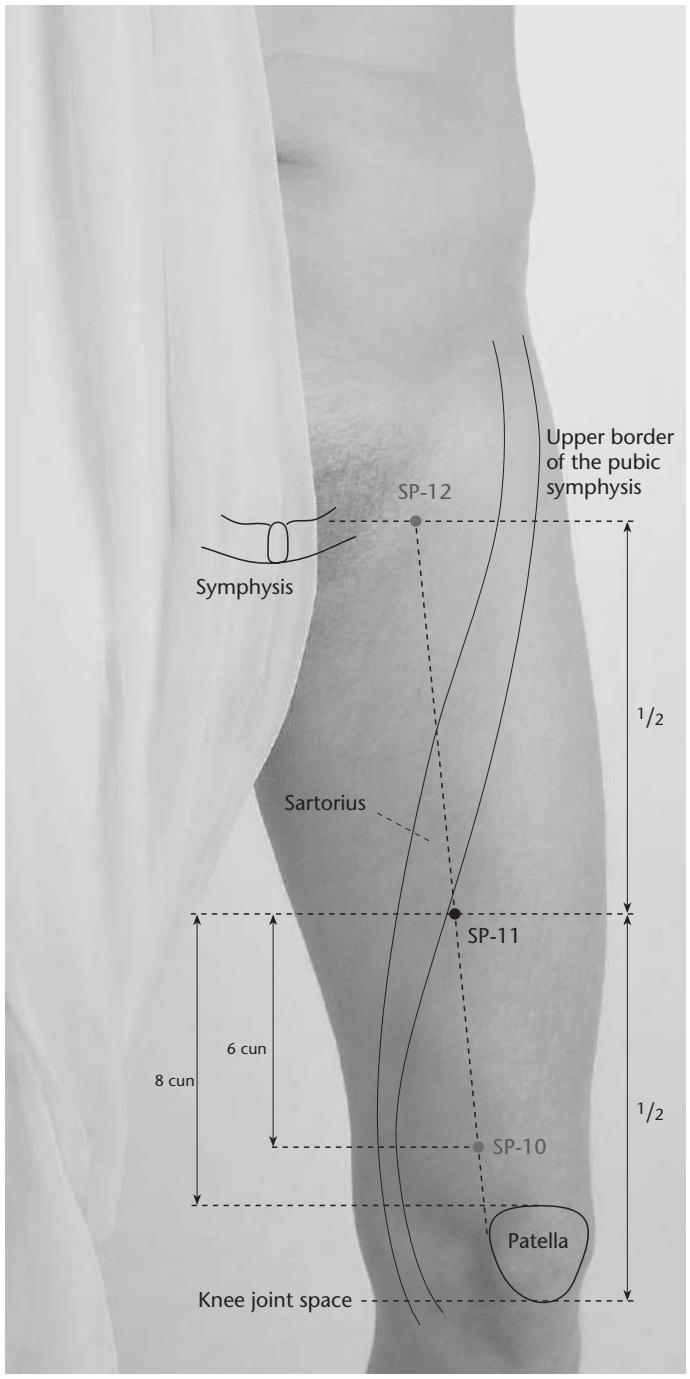

Location

6 cun proximal to → **SP-10** or 8 cun proximal to the medial upper border of the patella, at the midpoint of the femur between the sartorius and vastus lateralis muscles.

How to find

First, locate → **SP-10** by measuring 2 cun proximal and 1 cun medial from the medial upper border of the patella. From there, measure 6 cun vertically in a proximal direction to locate **SP-11** in a depression on the lateral border of the sartorius muscle. **Or**: Spreading hands technique (→ 2.3.3): **SP-11** is located midway between the knee joint space and → **SP-12** (2 cun lateral to the anterior midline, on a level with the upper border of the pubic symphysis).

Needling

Vertically 0.5–1 cun. Caution: Femoral artery/vein.

Actions/Indications

- Regulates the Lower Burner and urination, drains Dampness, clears Heat

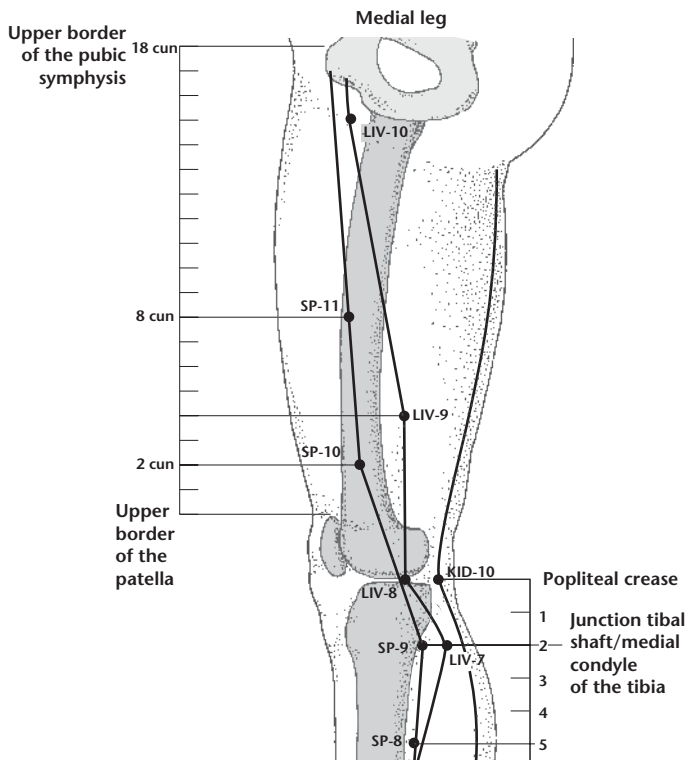

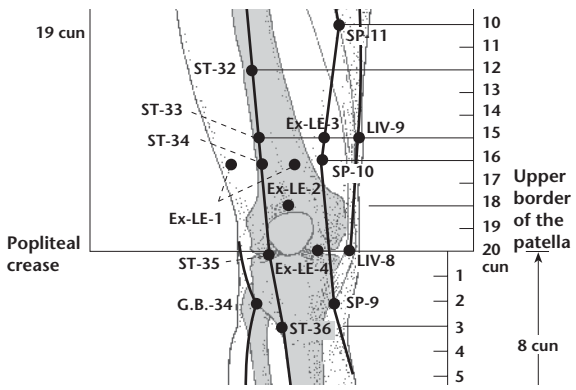

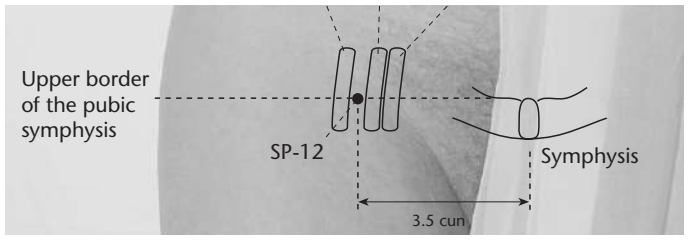

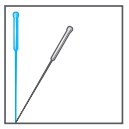
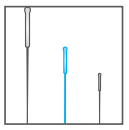

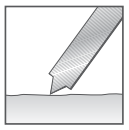
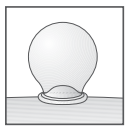

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 192

SP-12 *Rushing Gate* CHONGMEN

Location

3.5 cun lateral to the midline, at the level of the upper border of the pubic symphysis, lateral to the femoral artery.

How to find

From the upper border of the pubic symphysis (→ **Ren-2**), measure 3.5 cun in a lateral direction. There, you can feel the pulsation of the femoral artery. **SP-12** is located in a depression immediately lateral to the artery.

Located on the same level are **Ren-2** (on the midline), → **KID-11** (0.5 cun lateral to the midline) and → **ST-30** (2 cun lateral to the midline). → **LIV-12** is located 2.5 cun lateral to the midline, but 1 cun inferior to the upper border of the pubic symphysis.

Needling

Vertically 0.5–1 cun. Caution: Needling in a medial direction may puncture the artery; needling in a lateral direction may injure the femoral nerve.

Actions/Indications

- Regulates Qi and Blood in the Lower Burner
- Descends rising fetal Qi
- Clears Heat and drains Dampness

Special features

Meeting point with the LIV and T.B.* channels and the *yin wei mai*

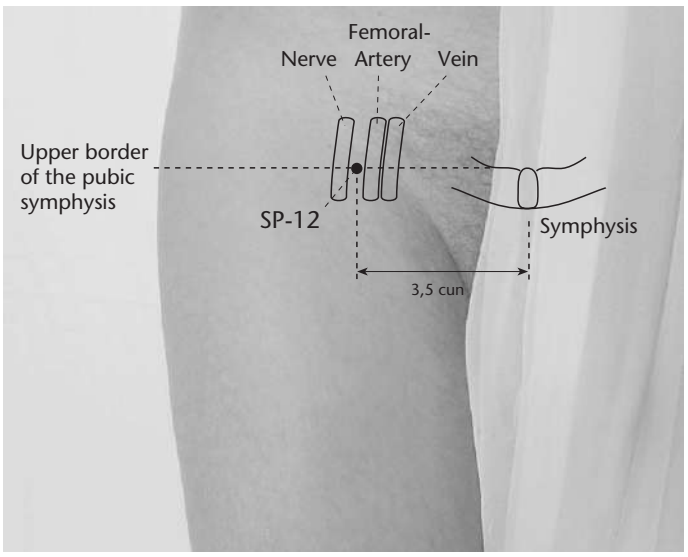

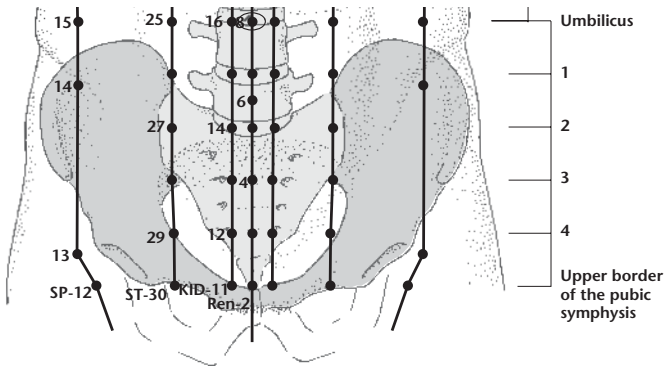

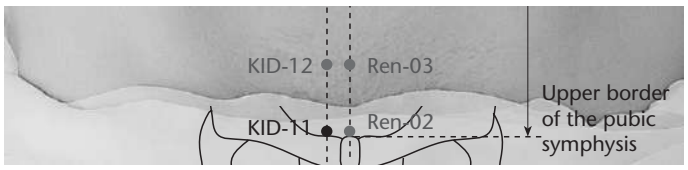

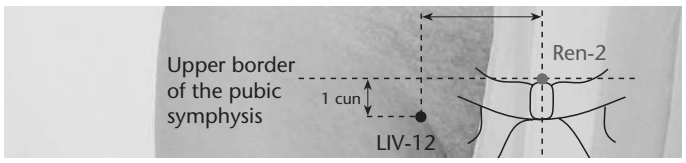

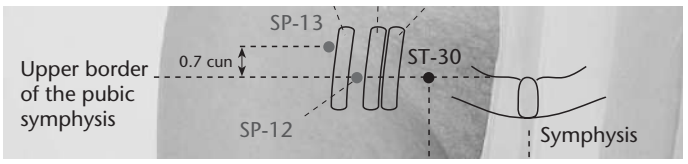

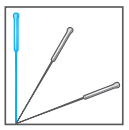

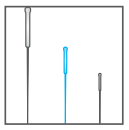

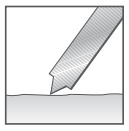

* According to some authors

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 193

Abode of the Fu FUSHE SP-13

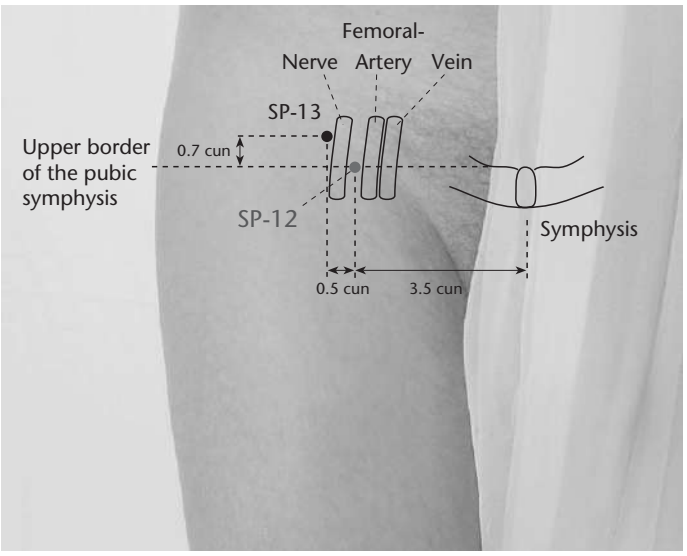

Location

4 cun lateral to the anterior midline and 0.7 cun superior to the upper border of the pubic symphysis.

How to find

First, measure 4 cun in a lateral direction from the upper border of the pubic symphysis and from there 0.7 cun in a superior direction to locate **SP-13** in a depression in the inguinal groove. **SP-13** is located 0.5 cun lateral and 0.7 cun superior to → **SP-12** (→ 3.5 cun lateral to the upper border of the pubic symphysis).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum.

Actions/Indications

- Disperses accumulations
- Regulates the Qi

Special features

Meeting point with the *yin wei mai* and the Liver channel

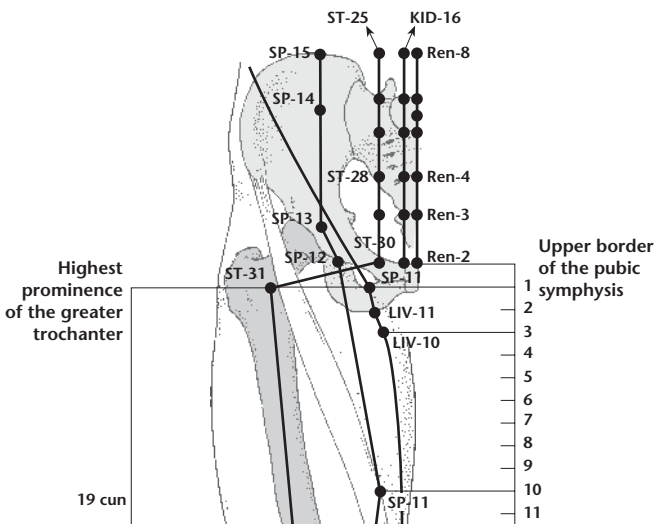

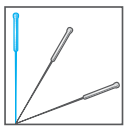
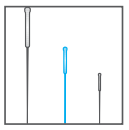
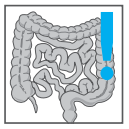

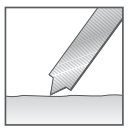
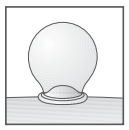

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 194

SP-14 *Abdomen Knot* FUJIE

Location
4 cun lateral to the anterior midline, on the mamillary line, 3 cun superior to → **SP-13** or 1.3 cun inferior to → **SP-15**.

How to find
First, measure 4 cun from the umbilicus in a lateral direction in order to locate the mamillary line. This is the point → **SP-15**. From there, measure 1.3 cun in an inferior direction to locate **SP-14**.

Needling
Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications
- Moves Qi and Blood, warms the middle, dispels Cold
- Regulates the Lower Burner and descends counterflow Qi

Special features
Sensitive to pressure with appendicitis

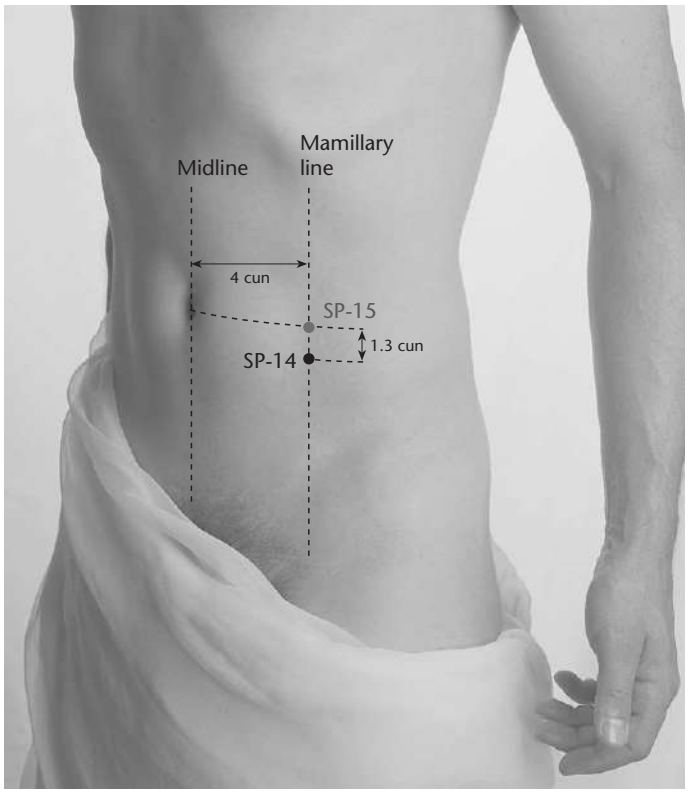

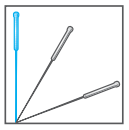
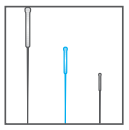
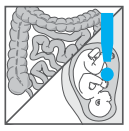

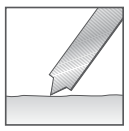

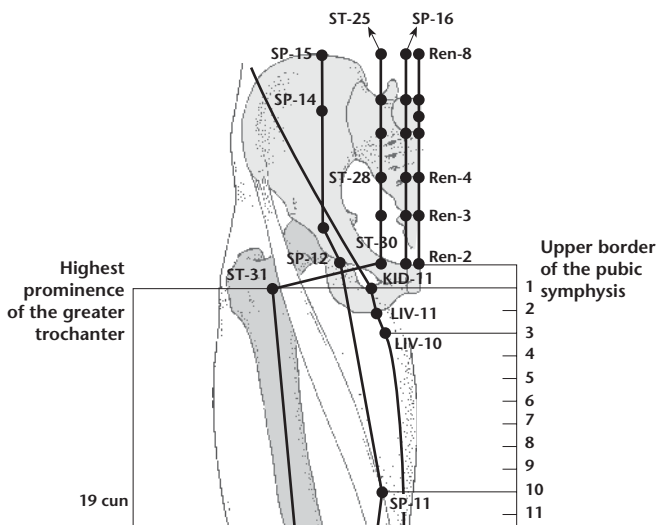

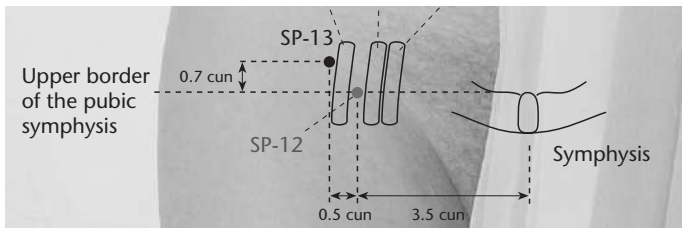

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 195

Great Horizontal DAHENG SP-15

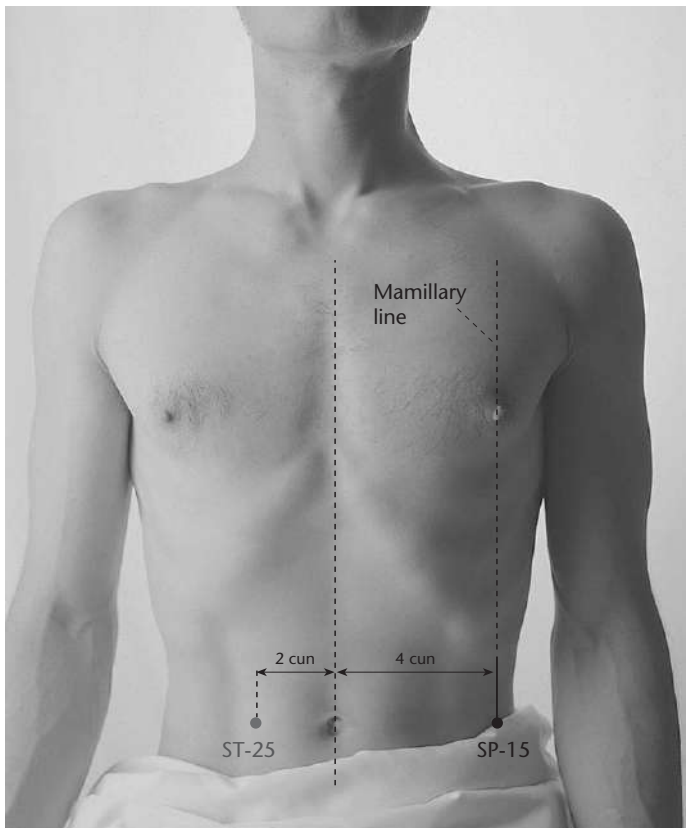

Location

4 cun lateral to the centre of the umbilicus, on the mamillary line.

How to find

Locate **SP-15** 4 cun lateral to the umbilicus, using the mamillary line for reference, especially in male patients.
Located on the same level are → **Ren-8/KID-16/ST-25/G.B.-26** (umbilicus/0.5 cun lateral to the umbilicus/2 cun lateral to the umbilicus/below the free end of the 11th rib at the level of the umbilicus).

Needling

Vertically 1–1.5 cun. Caution: Peritoneum, pregnancy. Deep needling may reach the aponeuroses of the transverse abdominal musculature.

Actions/Indications

- Regulates and moves Qi in the Intestines (promotes bowel movement by stimulating peristalsis)

Special features

Meeting point with the *yin wei mai*. Important point for regulating the Large Intestine Qi.

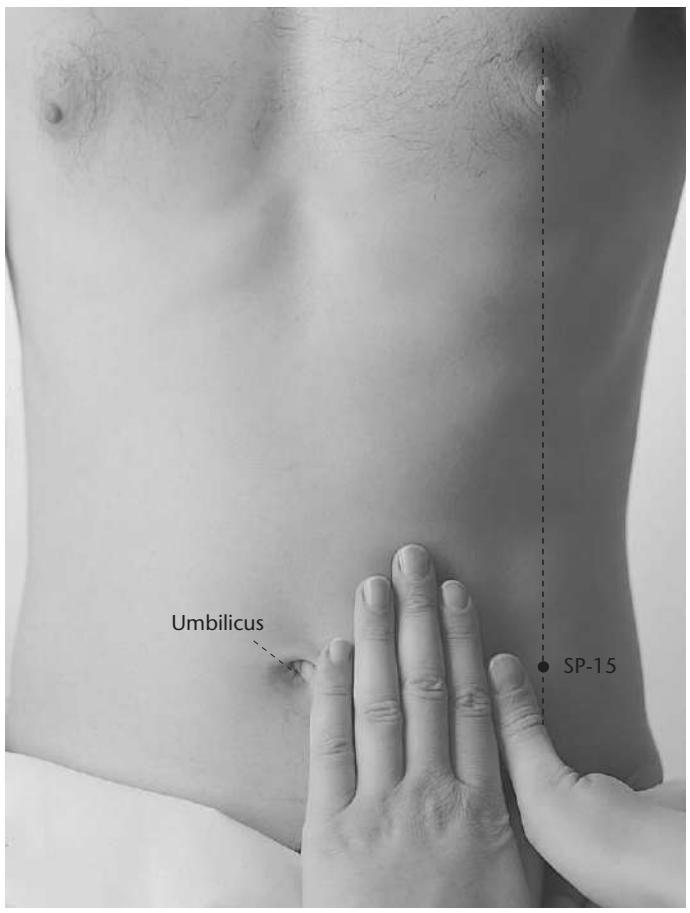

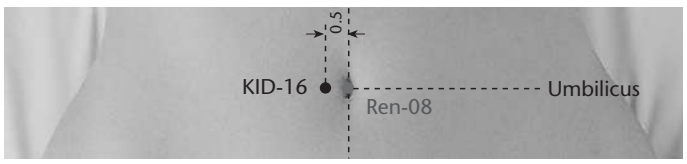

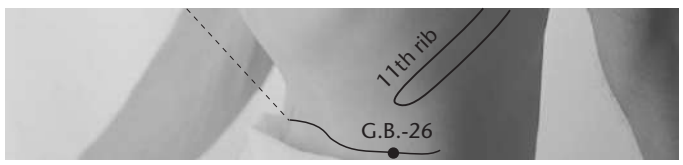

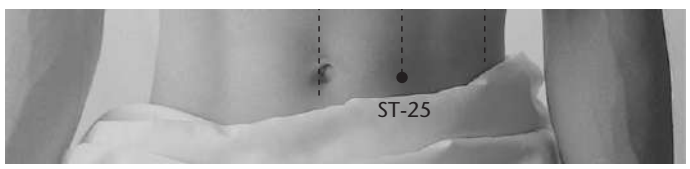

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 196

SP-16 *Abdomen Sorrow* FUAI

Location

3 cun above the centre of the umbilicus and 4 cun lateral to the anterior midline, on the mamillary line.

How to find

The distance between the sternocostal angle and the centre of the umbilicus can be divided into 8 proportional cun, which can differ considerably from the patient's finger cun; therefore, only proportional cun measurements should be applied (elastic tape → 2.3.1). First, measure 3 cun from the umbilicus in a superior direction, then 4 cun in a lateral direction, where **SP-16** is located on the mamillary line. Depending on the shape of the thorax, this point will either be located on the abdomen or it will project onto the lower border of the ribcage.

Located on the same level are → **Ren-11** (anterior midline), → **KID-18** (0.5 cun lateral to the anterior midline) and → **ST-22** (2 cun lateral to the anterior midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy. Depending on the shape of the thorax, the correct location of this point might be on the costal cartilage. In this case, either transverse (subcutaneous) needling or locate the point more medially (or choose an alternative point).

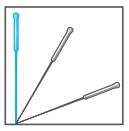

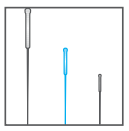

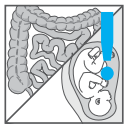

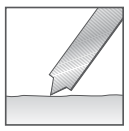

Actions/Indications

- Regulates Qi in the Intestines
- Drains Dampness

Special features

Meeting point with the *yin wei mai*

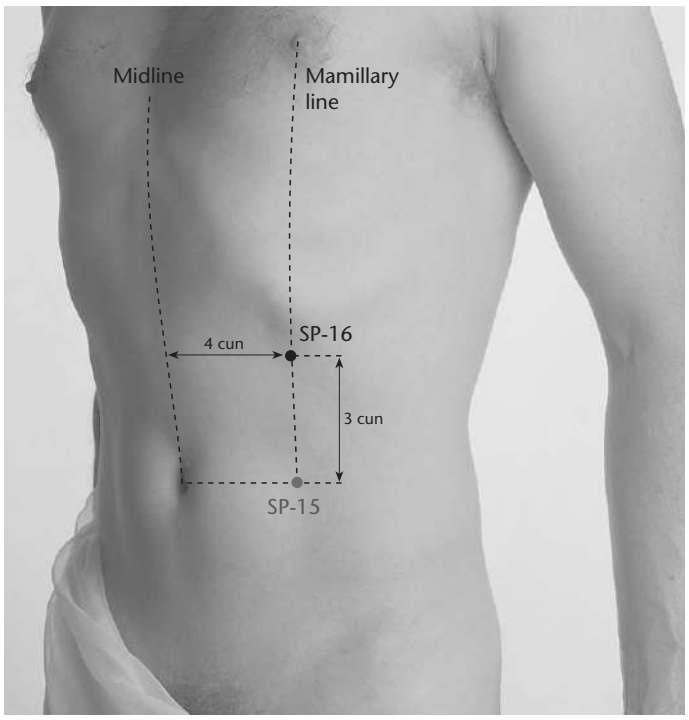

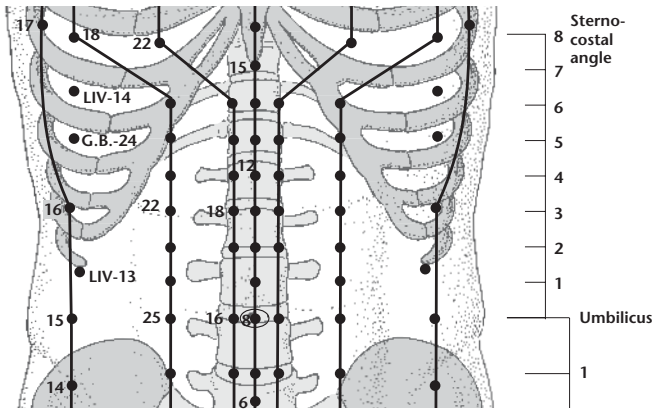

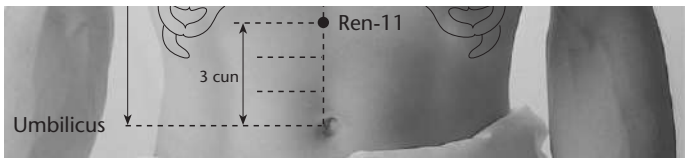

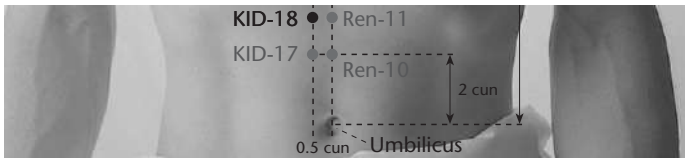

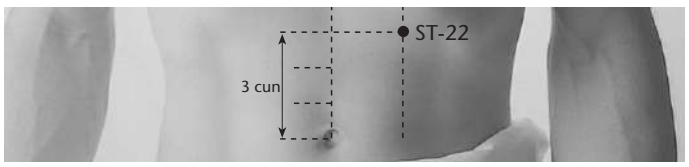

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 197

Food Cavity SHIDOU SP-17

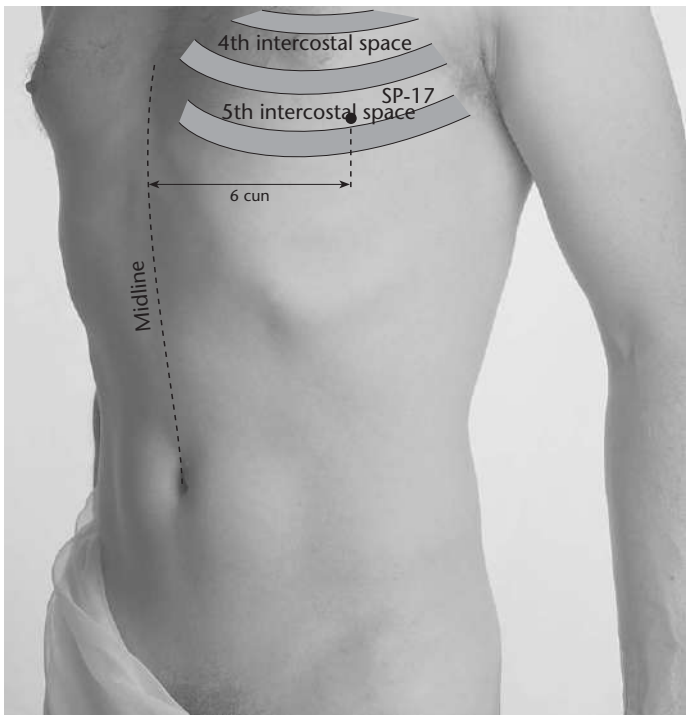

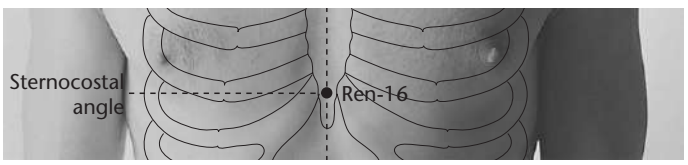

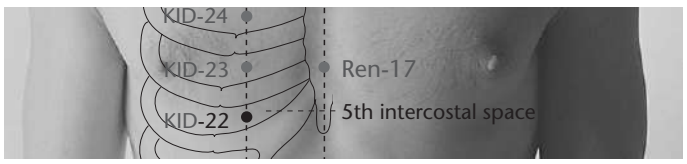

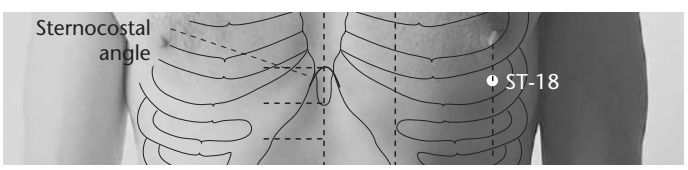

Location

In the 5th intercostal space, 6 cun lateral to the anterior midline (in male patients, 2 cun lateral to the nipple).

How to find

In men, the nipple is generally located on the level of the 4th intercostal space, while in women, when in a supine position, it is not always at this level. Locate **SP-17** in the 5th intercostal space, 2 cun lateral to the mamillary line (note that the intercostal space laterally curves upward). Both the clavicle and the manubriosternal synchondrosis are good reference points for locating the 5th intercostal space. **SP-17** can be located by following the course of the intercostal space in a lateral direction and measuring 2 cun lateral to the mamillary line.
Also located on the level of the 5th intercostal space are → **Ren-16** (anterior midline), → **KID-22** (2 cun lateral to the anterior midline) and → **ST-18** (4 cun lateral to the anterior midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), parallel to the intercostal space or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Regulates the Qi
- Harmonises the Stomach and eliminates stagnation
- Drains Dampness

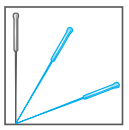
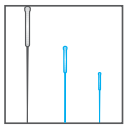
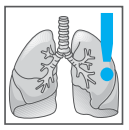

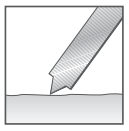

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 198

SP-18 *Heavenly Stream* TIANXI

Location

In the 4th intercostal space, 6 cun lateral to the anterior midline or 2 cun lateral to the midclavicular line.

How to find

In men, the nipple is generally located on the level of the 4th intercostal space, while in women, when in a supine position, it is not always at this level. Follow the course of the 4th intercostal space in a lateral direction and locate **SP-18** 2 cun lateral to the nipple (note that the intercostal space laterally curves upward). In women, either the clavicle or the manubriosternal synchondrosis are good reference points for locating the 4th intercostal space (→ 3.5). **SP-18** can be located by following the course of the intercostal space in a lateral direction while measuring 6 cun laterally from the anterior midline or 2 cun laterally from the nipple.

Also located on the level of the 4th intercostal space are → **Ren-17** (anterior midline), → **KID-23** (2 cun lateral to the anterior midline), → **ST-17** (nipple), → **P-1** (1 cun lateral to the nipple) and → **G.B.-23** (1 cun anterior to **G.B.-22**).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), parallel to the intercostal space or with or against the channel pathway. Caution: Pneumothorax.

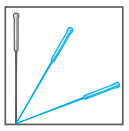

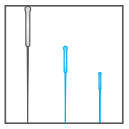

Actions/Indications:

- Unbinds the chest
- Descends the Qi
- Benefits the breasts

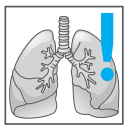

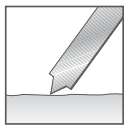

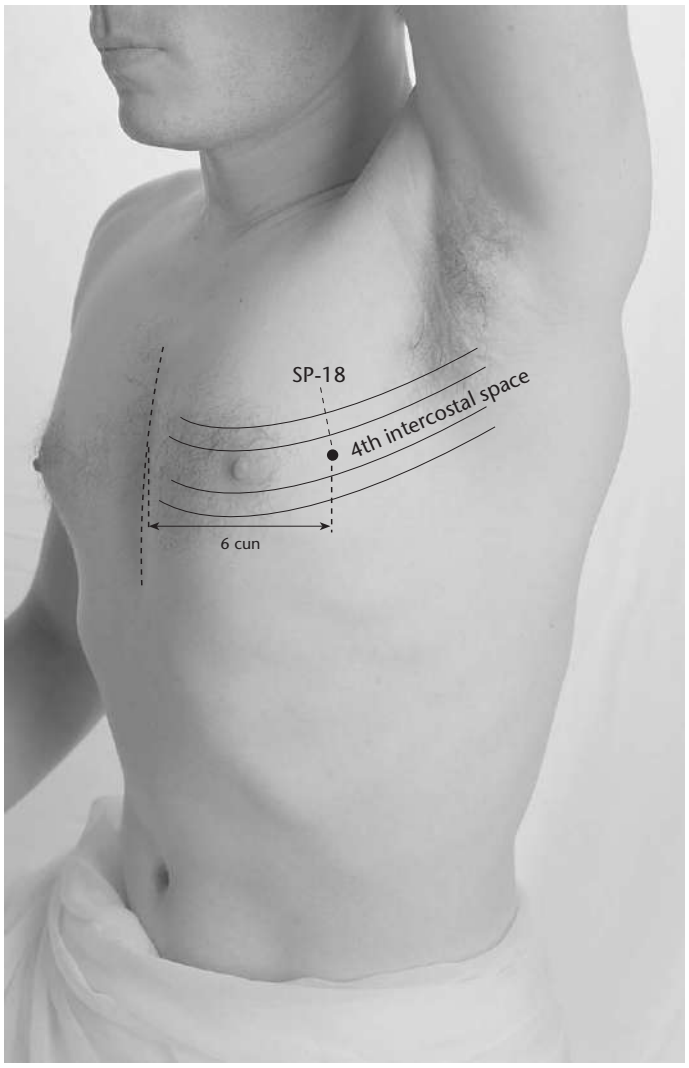

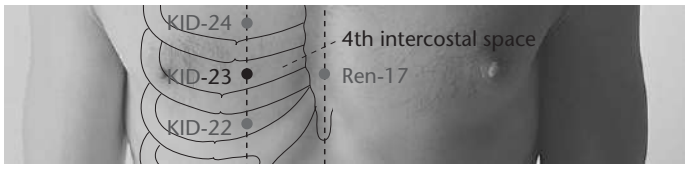

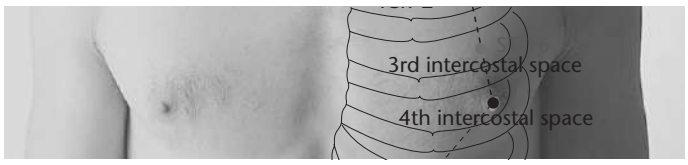

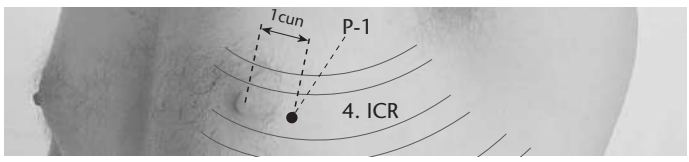

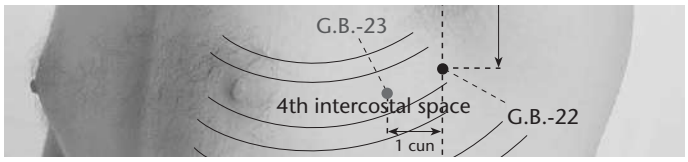

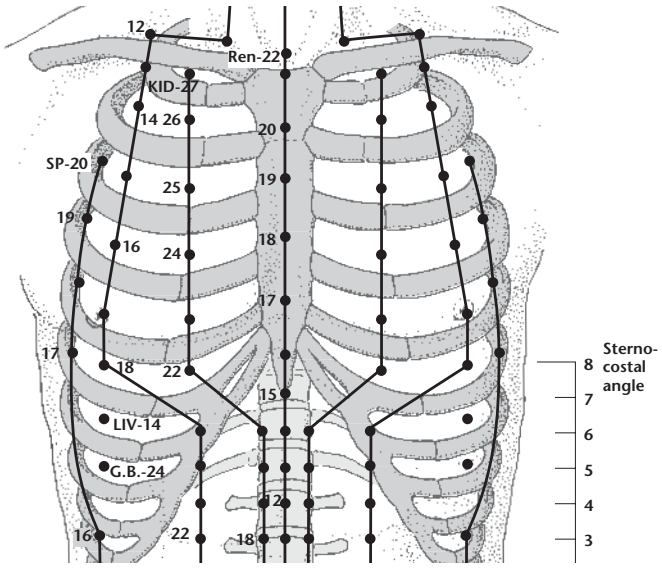

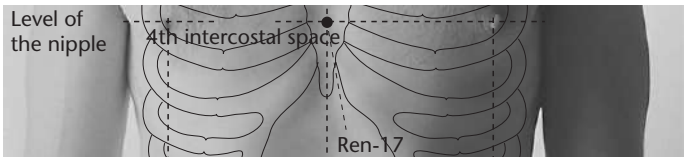

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 199

Chest Village XIONGXIANG SP-19

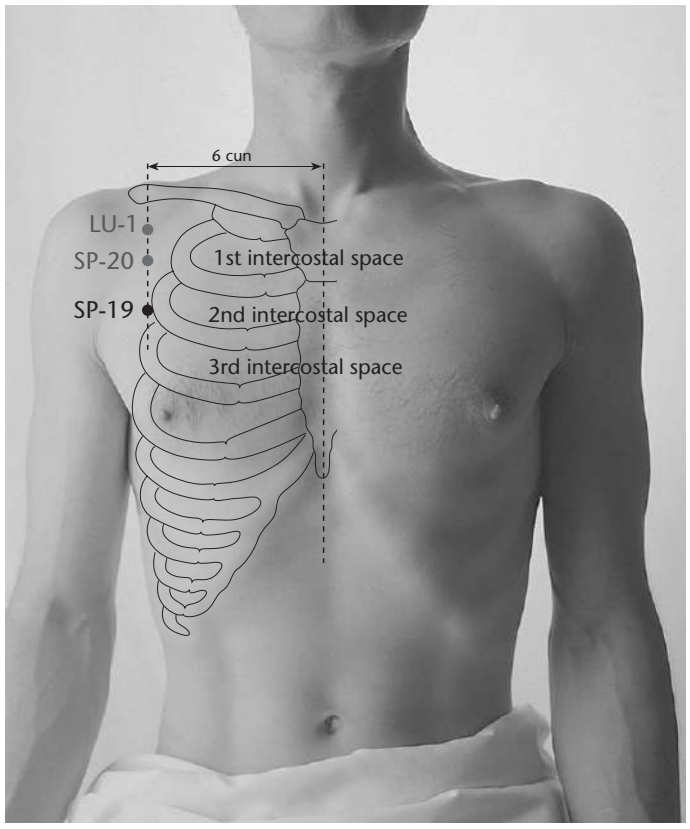

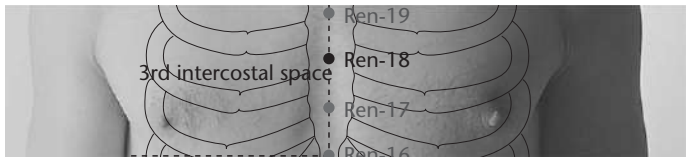

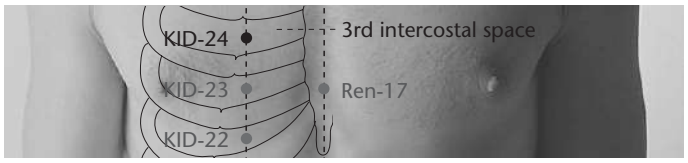

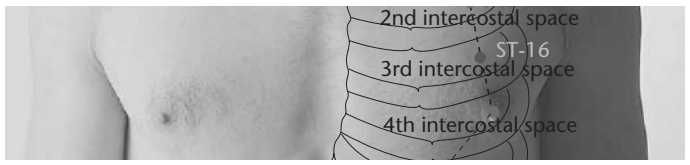

Location

In the 3rd intercostal space, 6 cun lateral to the anterior midline or 2 cun lateral to the midclavicular line.

How to find

Locate the 3rd intercostal space by counting parasternally from the clavicle or from the manubriosternal synchondrosis (→ 3.5). Follow the course of the intercostal space in a lateral direction and locate **SP-19** 6 cun lateral to the anterior midline or 2 cun lateral to the midclavicular line (note: the intercostal space laterally curves upward).

Located on the same level (note: course of the intercostal space) are → **Ren-18** (anterior midline), → **KID-24** (2 cun lateral to the midline) and → **ST-16** (4 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), parallel to the intercostal space, or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Disperses and descends the Lung Qi
- Unbinds the chest

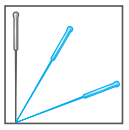
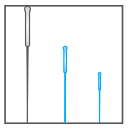
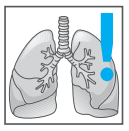

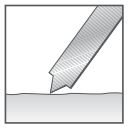

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 200

SP-20 *Encircling Glory* ZHOURONG

Location

In the 2nd intercostal space, 6 cun lateral to the anterior midline or 2 cun lateral to the midclavicular line.

How to find

Locate the 2nd intercostal space by counting parasternally from the clavicle or from the manubriosternal synchondrosis (→ 3.5). Follow the course of the intercostal space in a lateral direction and locate **SP-20** 6 cun lateral to the anterior midline or 2 cun lateral to the midclavicular line (note: the intercostal space laterally curves upward). **SP-20** is located 1 cun inferior to → **LU-1**. Located on the same level (note: course of the intercostal space) are → **Ren-19** (anterior midline), → **KID-25** (2 cun lateral to the midline) and → **ST-15** (4 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), parallel to the intercostal space or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Unbinds the chest
- Regulates the Lung Qi

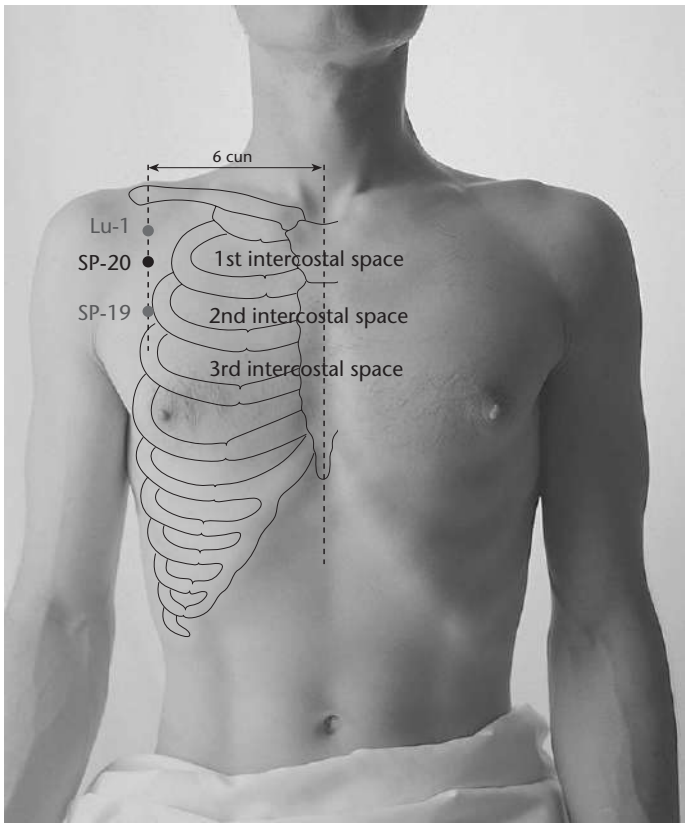

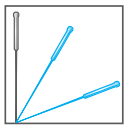

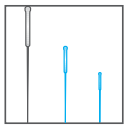

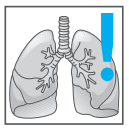

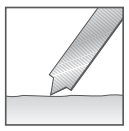

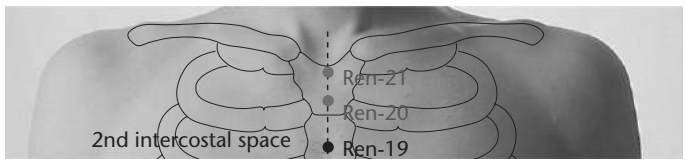

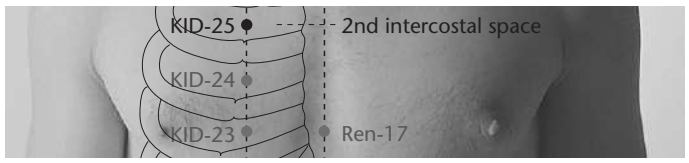

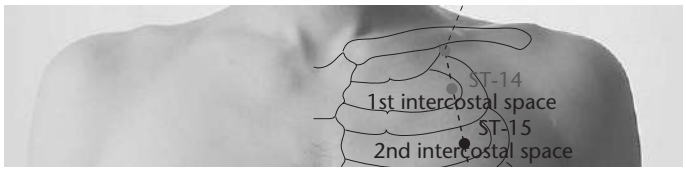

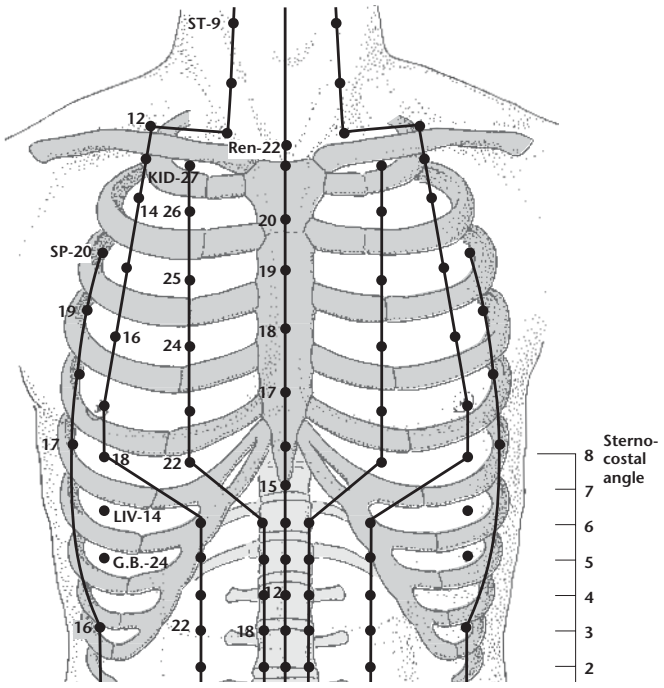

Ch04.4-F10028.qxd 2/23/08 7:09 PM Page 201

Great Wrapping DABAO SP-21

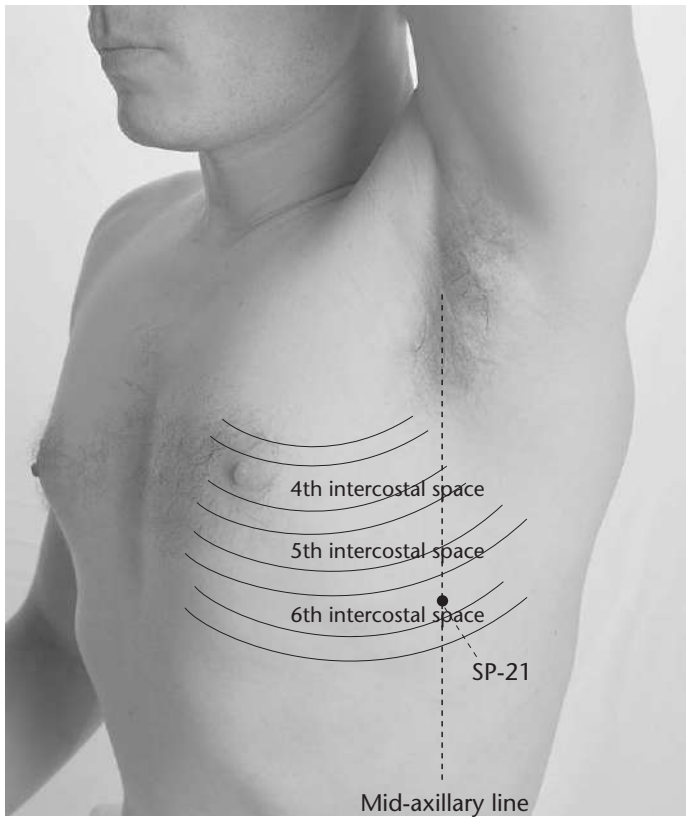

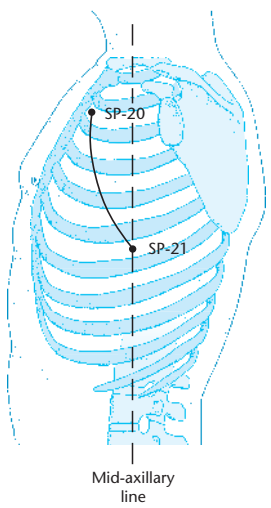

Location

On the midaxillary line, in the 6th intercostal space. Note: According to some texts, this point is located in the 7th intercostal space.

How to find

Locate the 6th intercostal space by counting parasternally from the clavicle or from the manubriosternal synchondrosis (→ 3.5). Follow the course of the intercostal space in a lateral direction and locate **SP-21** on the axillary midline (note: the intercostal space laterally curves upward). **SP-21** is often sensitive to pressure. Or: Locate **SP-21** on the midaxillary line, midway between the axilla and the end of the 11th rib.

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously), parallel to the intercostal space or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Regulates Qi and Blood, governs and wraps all the *luo*-connecting vessels of the body
- Unbinds the chest

Special features

Great *luo*-connecting point of the spleen (→ 8.1.2), exit point.

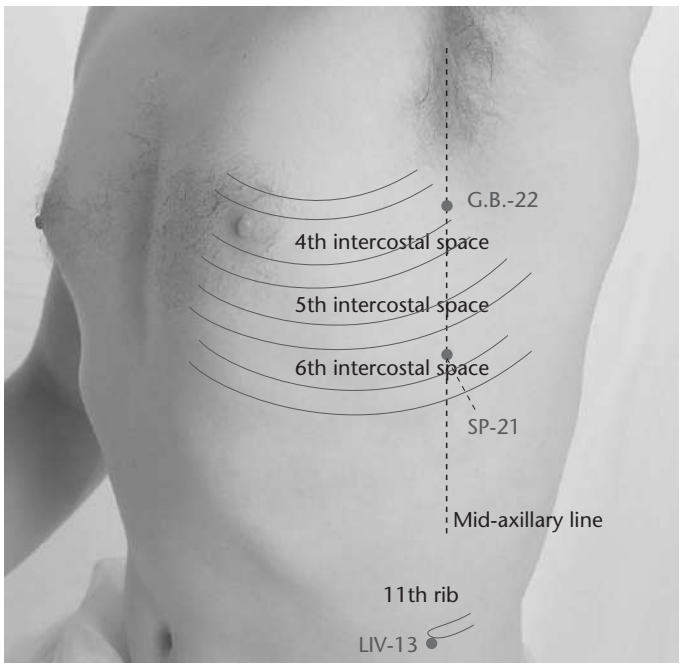

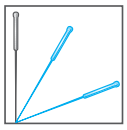
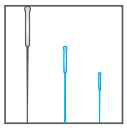
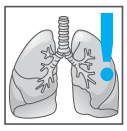

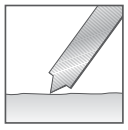

Ch04.5-F10028.qxd 2/22/08 4:37 PM Page 203

4.5 The Heart Channel System – Hand *shaoyin* (*shou shao yin jing luo*)

4.5.1 The Heart Primary Channel (*shou shao yin jing*)

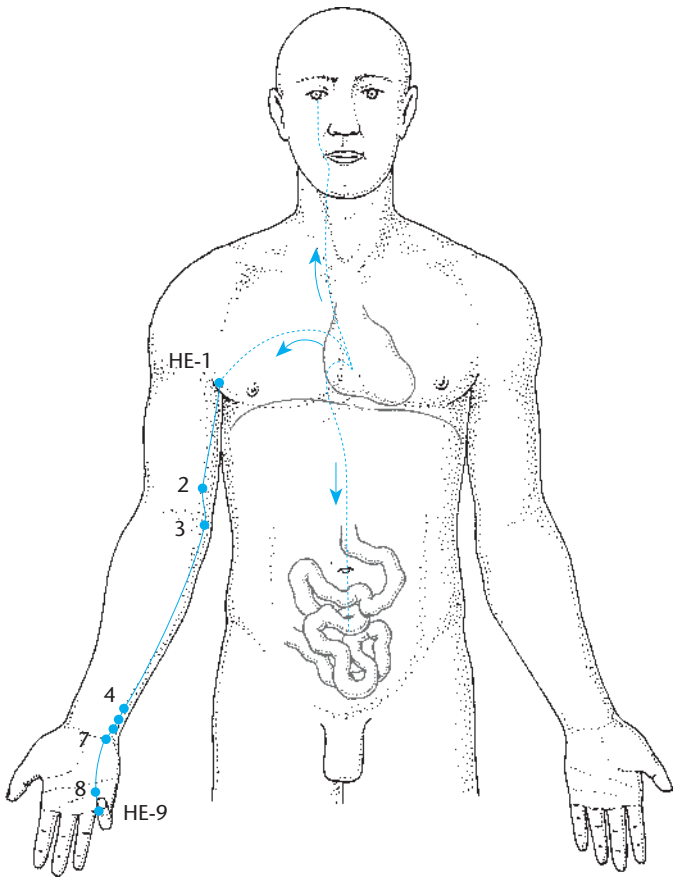

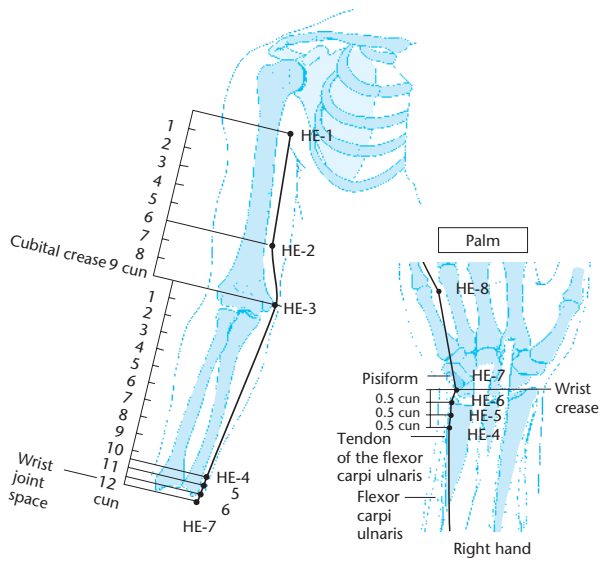

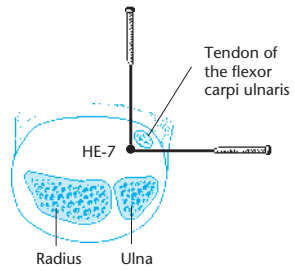

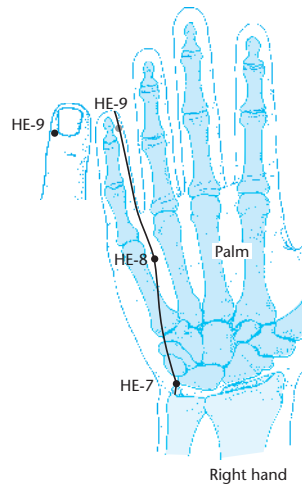

Pathway

The **internal** pathway of the Heart primary channel originates in its related *zang*-Organ, the Heart (*xin*). An **internal branch** of the Spleen primary channel reaches the Heart and connects with the Heart primary channel (deep Yin–Yin connection ➞ 1.2).

The **internal pathway** of the Heart primary channel splits into three branches:

- **the first branch** penetrates the diaphragm and connects with the internally–externally related *fu*-Organ, the Small Intestine (*xiaochang*)
- **the second branch** ascends along the oesophagus and the neck region, disperses in the tissues around the eyes and, according to some authors, also spreads inside the brain.
- **the main branch** runs from the Heart to the Lung (*fei*) and descends again to emerge in the centre of the axilla at ➞ **HE-1** (*jiquan*) (beginning of the external pathway).

From here the **external pathway** runs along the anteromedial aspect of the upper arm and forearm, traverses the palm between the fourth and fifth metacarpal bones, continues along the radial aspect of the little finger and ends at the radial corner of its nail at **HE-9** (*shaoyang*).

From **HE-9** (*shaoyang*) a branch crosses to the ulnar border of the little finger's nail to ➞ **S.I.-1** (*shaoze*) (hand Yin–Yang connection of the second great circuit).

Clinical importance (➞ 1.2)

Exterior (*biao*) signs and symptoms: General fever, headaches, pain in the thorax and back, thirst with a desire to drink, Heat sensation in the palms, Cold invasion in the extremities, pain in the shoulder and the medial aspect of the upper arm

Interior (*li*) or *zangfu*-Organ signs and symptoms: Heart pain, pain and fullness in the thorax and lateral costal region,

Ch04.5-F10028.qxd 2/22/08 4:37 PM Page 204

hypochondriac pain, irritability, shortness of breath, restlessness, dizziness, mental confusion

Connections of the Heart primary channel (→ 1.2)

Connections with other channels

Small Intestine primary channel (*shou tai yang jing*)

Connection: Hand Yin–Yang connection of the second great circuit
Location: H-9 → S.I.-1 (on the hand).
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Kidney primary channel (zu shao yin jing)

Connection: Paired according to the six channel theory (hand–foot pairing): *shao yin* (Yin axes of the second great circuit)
Location: KID → HE (on the thorax). An internal branch of the Kidney primary channel travels to the Liver (*gan*), penetrates the diaphragm and spreads in the Lung (*fei*). From the Lung a branch runs to Heart (*xin*) and there connects with the Heart primary channel.
Circulation: Non-circadian (**not** according to the Organ clock)
Importance: Above–below relationship

Spleen primary channel (*zu tai yin jing*)

Connection: Deep Yin–Yin connection
Location: SP → HE (on the thorax). An **internal** branch of the Spleen primary channel spreads in the Heart (*xin*) and connects with the Heart primary channel.
Circulation: Circadian (according to the Organ clock)
Importance: The Heart primary channel receives Nutritive Qi (*ying qi*) from the Spleen primary channel (first circulation of the *ying qi* → 1.1.4)

Connections with other *zangfu*-Organ systems

Small Intestine (*xiaochang*), Heart (*xin*), Lung (*fei*)

4.5.2 The Heart Divergent Channel (*zu shao yin jing bie*)

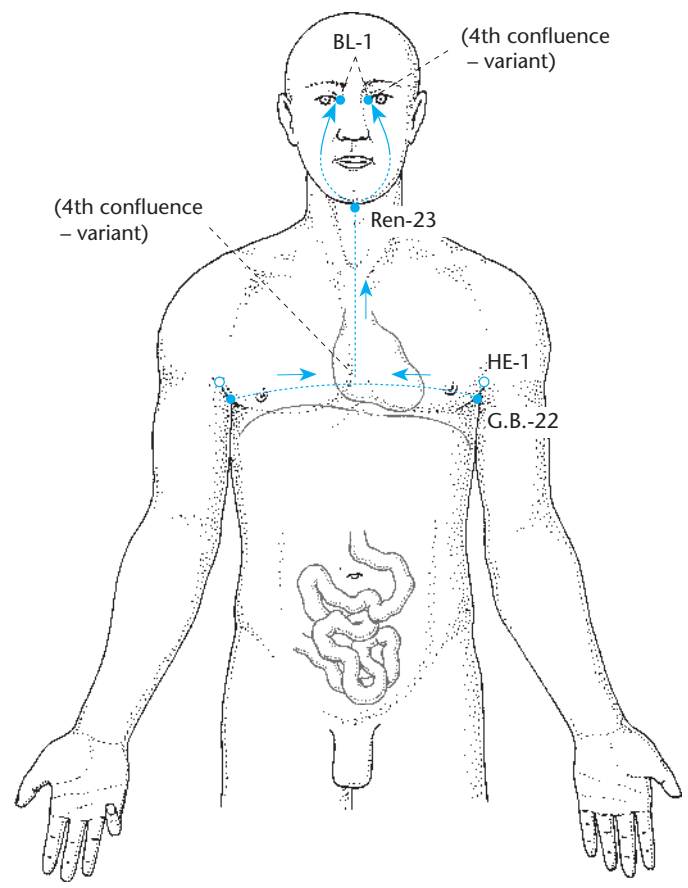

Pathway

The Heart divergent channel separates from the Heart primary channel at **HE-1** (*jiquan*) in the axilla

- penetrates the thorax at → **G.B.-22** (*yuanyue*) on the mid-axillary line and 3 cun distal to the axilla
- and reaches the Heart (*xin*) (see commentary below).
- It then ascends to the neck and crosses → **Ren-23** (*lianquan*)
- disperses in the cheek and reaches the medial canthus of the eye at → **BL-1** (*jingming*), where, according to most authors, it connects with a small branch coming from → **S.I.-18** (*quanliao*) to form one of the six he-confluences (here: S.I./BL as fourth confluence → 1.3).

Commentary: according to Solinas et al (1998) the Heart divergent channel meets the Small Intestine divergent channel and the Small Intestine primary channel at the Heart (*xin*) in order to bind with one of the six he-confluences.

Clinical importance

- Strengthens the relationship between the Heart and the thoracic region. Points on the Heart primary channel mainly treat disorders of the Heart and the thorax. There is no connection with the Small Intestine (*xiaochang*).

Ch04.5-F10028.qxd 2/22/08 4:37 PM Page 205

- Connects the Heart primary channel with the neck. Points on the Heart primary channel can be used to treat disorders of the throat and vocal chords such as laryngitis, pharyngitis, etc. Use points such as **HE-5, HE-8**.
- Strengthens the connection between the Heart and the inner canthus of the eye. Heart points can thus be used for disorders of the eyes.

4.5.3 The Heart Sinew Channel (*shou shao yin jing jin*)

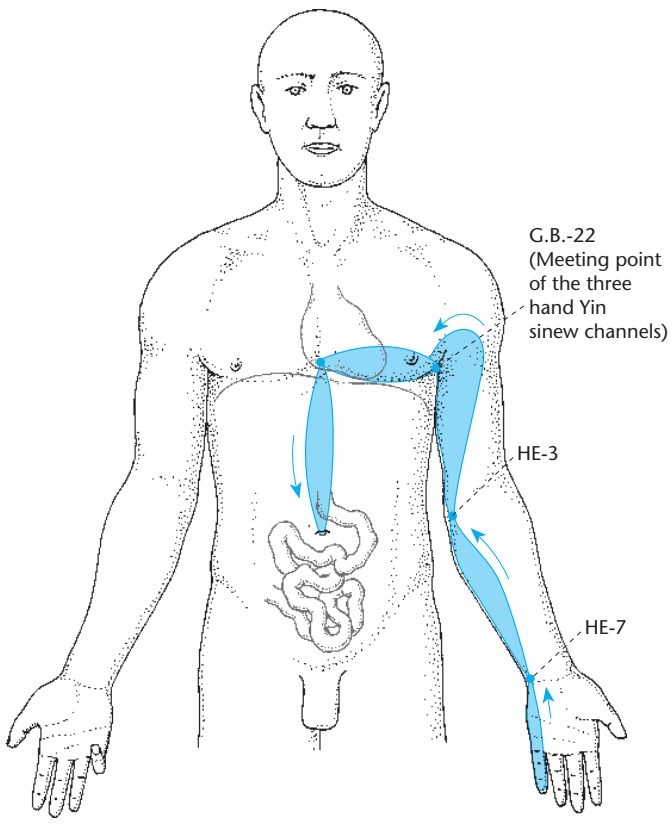

Pathway

The Heart sinew channel begins on the radial aspect of the little finger

- ➡ continues along the anterior and radial aspect of the little finger and the hand
- ➡ binds (*jie*) at the pisiform bone at **HE-7** (*shenmen*)
- ➡ ascends the anterior and ulnar aspect of the forearm, binds (*jie*) at the medial epicondyle at **HE-3** (*shaohai*) and continues to the axillary region, where it meets the Lung sinew channel, binds (*jie*) in the axilla and meets the other hand Yin sinew channels near → **G.B.-22** (*yuanye*)
- ➡ from here it continues at a deeper level, crosses the thoracic region, binds (*jie*) at → **Ren-17** (*shanzhong*), descends to the upper cardiac orifice, penetrates the diaphragm and terminates in the umbilical region.

Clinical importance

Pathology: Stiffness, cramps and pain along the pathway of the Heart sinew channel. Internal cramping in the cardiac region, for example disorders with angina-like pain due to stress, hiatal hernia and disorders of the stomach.

Indication: Mainly for *bi*-syndromes (painful obstruction syndrome) along the pathway of the Heart channel. The area covered by the Heart sinew channel is larger than that covered by the Heart primary channel. This explains why Heart primary channel points can also be used for disorders and diseases of the thorax and the chest, for example pain and tightness in the chest, mastitis and mastopathy. Since the Heart sinew channel also penetrates the diaphragm, it is also indicated for disorders such as hiccups, acid reflux and oesophagitis. Due to its connection with the umbilical region, the Heart sinew channel has a direct link to the Original Qi (*yuan qi*): if the Heart (*shen*) has suffered a shock, for example due to fainting or collapse, direct moxibustion at **Ren-8** can be very helpful.

4.5.4 The Heart *luo*-Connecting Vessel System (*shou shao yin luo mai*)

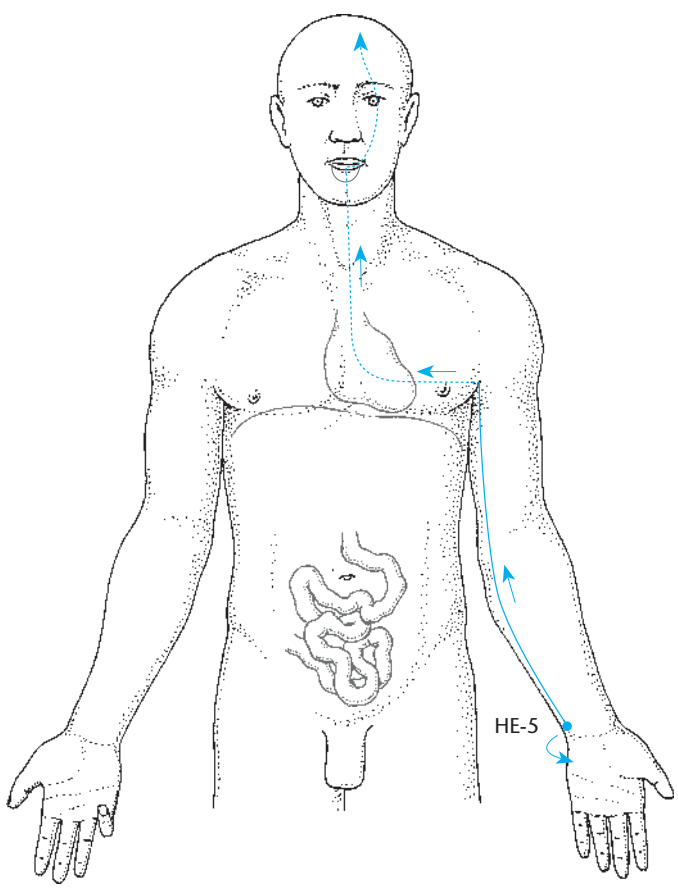

Pathway

The HE *luo*-connecting vessel separates from the Heart primary channel at its *luo*-connecting point **HE-5** (*tongli*) (→ 8.1.2). It forms a three-dimensional reticulate network, dividing into

Ch04.5-F10028.qxd 2/22/08 4:37 PM Page 206

multiple branches and sub-branches (*sun luo, fu luo, xue luo* → 1.5) within the surrounding tissue.

- Horizontal divisions run to the Interiorly–Exteriorly paired Small Intestine primary channel; according to some schools of thought (for example Ngyen Van Nghi → Appendix) they travel as a **transverse** HE *luo*-connecting vessel to the *yuan*-source point → **S.I.-4** (*wangu*).
- A **longitudinal** division follows the pathway of the Heart primary channel along the anteromedial aspect of the arm, ascends to the axilla, reaches the Heart, ascends to the root of the tongue and terminates in the eye while further branches penetrate the brain.

Clinical importance (→ 8.1.2)

Pathology

- **Excess (*shi*)**: Distension and fullness of the thorax and diaphragm
- **Deficiency (*xu*)**: Aphasia, disorders of the vocal cords

4.5.5 Cutaneous Region (*shao yin pi bu*)

See description and figures → 1.6.

4.5.6 Points of the Heart Primary Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): HE-7** (*shenmen*) ■■
- ***Luo*-connecting point (→ 8.1.2.): HE-5** (*tongli*) ■■
- ***Xi*-cleft point (→ 8.1.3): HE-6 (*yinxi*)** ■■
- **Associated Back-*shu* point (→ 8.1.4): BL-15** (*xinshu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): Ren-14** (*juque*) ■■
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Wood): **HE-9** (*shaochong*) ■■
  ying-spring point (Fire), *ben* point (Five Phases): **HE-8** (*shaofu*) ■■
  shu-stream point (Earth), sedation point: **HE-7** (*shenmen*) ■■
  jing-river point (Metal): **HE-4** (*lingdao*)
  he-sea point (Water): **HE-3** (*shaohai*) ■■
- ***Hui*-meeting point (→ 8.1.7): –**
- **Opening point (→ 8.1.8): –**
- **Lower *he*-sea points (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10): –**
- **Gao Wu command point (→ 8.1.11): –**
- **Window of Heaven point (→ 8.1.12): –**
- **Points of the Four Seas (→ 8.1.13): –**
- **Ma Dan Yang Heavenly Star point (→ 8.1.14): HE-5** (*tongli*) ■■
- **Sun Si Miao Ghost point (→ 8.1.15): –**
- **Other functional points: –**

Points according to region

- **Local point (→ 8.2.1):** elbow – **HE-3** (*shaohai*) ■■
- **Distal points (→ 8.2.1):** tongue – **HE-5** (*tongli*) ■■; Heart – **HE-7** (*shenmen*) ■■, **HE-5** (*tongli*) ■■, **HE-3** (*shaohai*) ■■

Specific points according to the channel pathway (in numeric order)

- **HE-3** (*shaohai*) ■■: *he*-sea point (Water); local point for the elbow (→ 8.2.1); distal point for the Heart (→ 8.2.1)
- **HE-4** (*lingdao*): *jing*-river point (Metal)
- **HE-5** (*tongli*) ■■: *luo*-connecting point (→ 8.1.2); Ma Dan Yang Heavenly Star point (→ 8.1.14); distal point for the tongue and Heart (→ 8.2.1)
- **HE-6** (*yinxi*) ■■: *xi*-cleft point (→ 8.1.3)
- **HE-7** (*shenmen*) ■■: *yuan*-source point (→ 8.1.1); distal point for the Heart (→ 8.2.1); *shu*-stream point (Earth), sedation point
- **HE-8** (*shaofu*) ■■: *ying*-spring point (Fire); *ben* point (Five Phases)
- **HE-9** (*shaochong*) ■■: *jing*-well point (Wood), tonification point

Ch04.5-F10028.qxd 2/22/08 4:37 PM Page 207

Summit Spring JIQUAN HE-1

Location

With the arm abducted, in the centre of the axilla, medial to the axillary artery.

How to find

With the arm abducted, palpate along the thoracic wall in a superior direction towards the centre of the axilla until you reach the highest point.

Needling

Palpate for the artery before needling! 0.5–1 cun vertically in the direction of → **G.B.-21** (border of the trapezius, middle of the shoulder). Caution: Axillary plexus and artery. Needling in a medial direction may puncture the lung!

Actions/Indications

- Unbinds the chest
- Opens the channel
- Nourishes the Heart Yin

Special features

Entry point. Important point in Qigong practice: During practice (and in everyday life), the axillary region should always be slightly open to allow for the free circulation of Qi and Blood to and from the upper extremity.

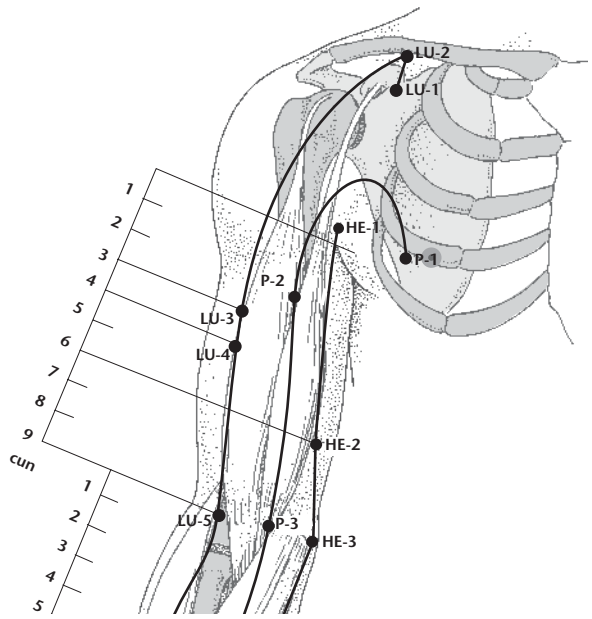

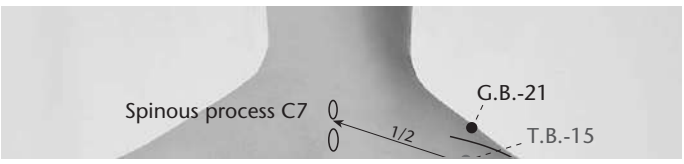

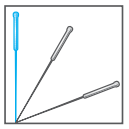
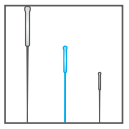
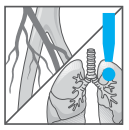

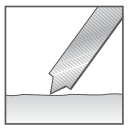

Ch04.5-F10028.qxd 2/22/08 4:37 PM Page 208

HE-2 *Green Spirit* QINGLING

Location

3 cun proximal to the cubital crease, on the medial border of the biceps brachii muscle.

How to find

With the elbow flexed, measure 3 cun from the ulnar end of the cubital crease (→ **HE-3**) towards the axilla. **HE-2** is located in a groove on the medial border of the biceps brachii muscle. Ask the patient to flex this muscle for easier location.

Needling

Vertically or obliquely 0.5–1 cun. Caution: Brachial artery! According to classic texts, needling is contraindicated, only moxibustion is permitted.

Actions/Indications

- Opens the channel

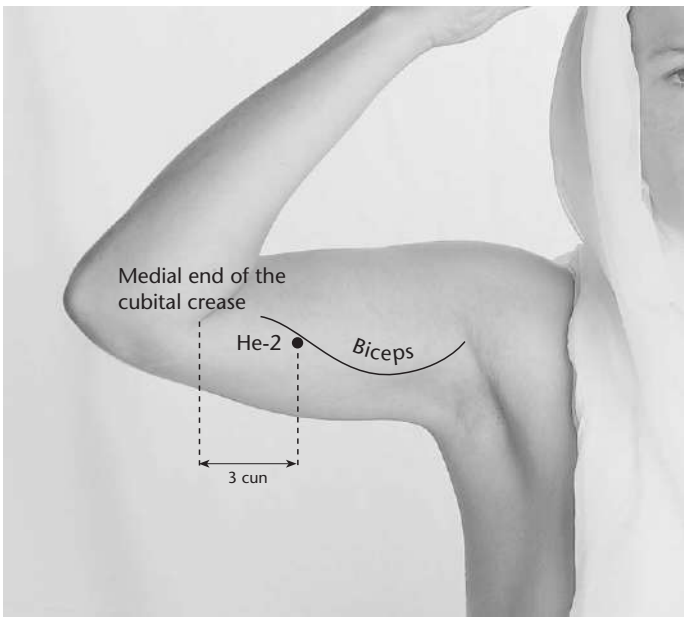

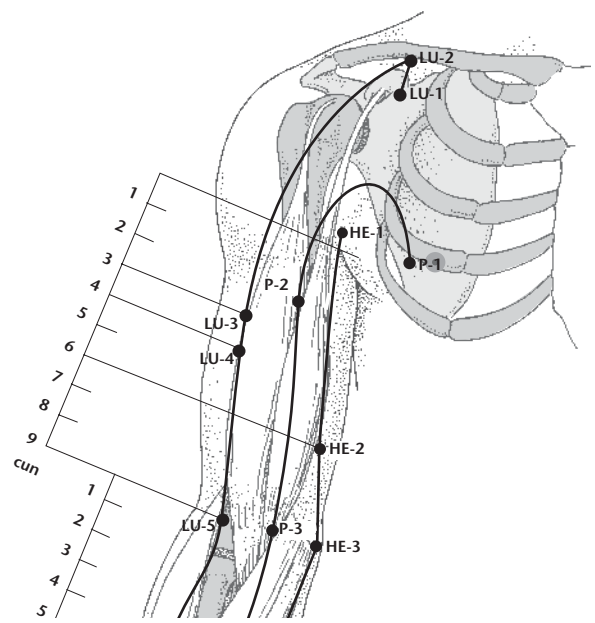

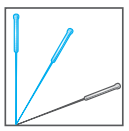
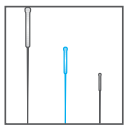

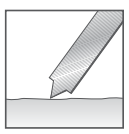

Ch04.5-F10028.qxd 2/22/08 4:37 PM Page 209

Lesser Sea SHAOHAI HE-3

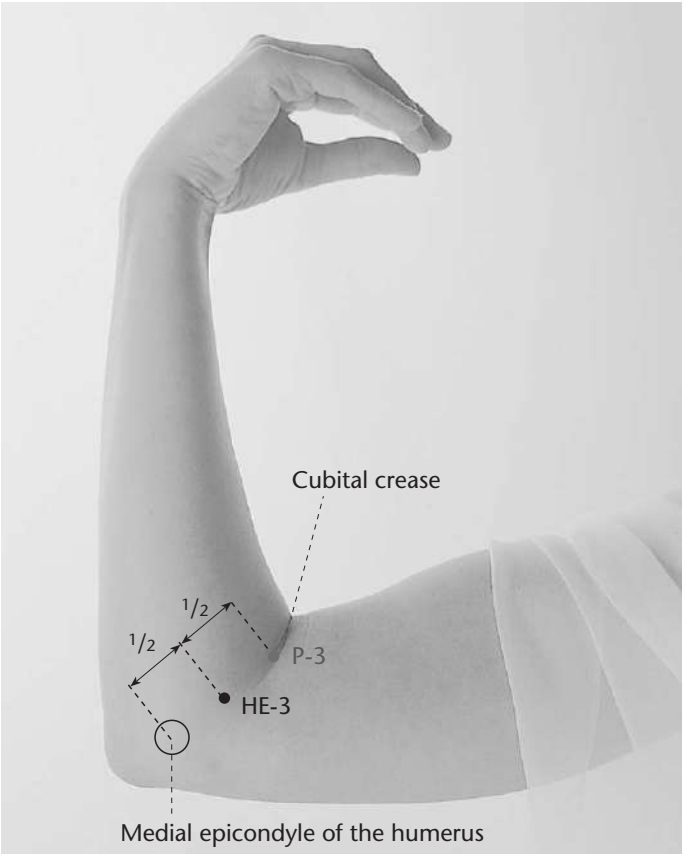

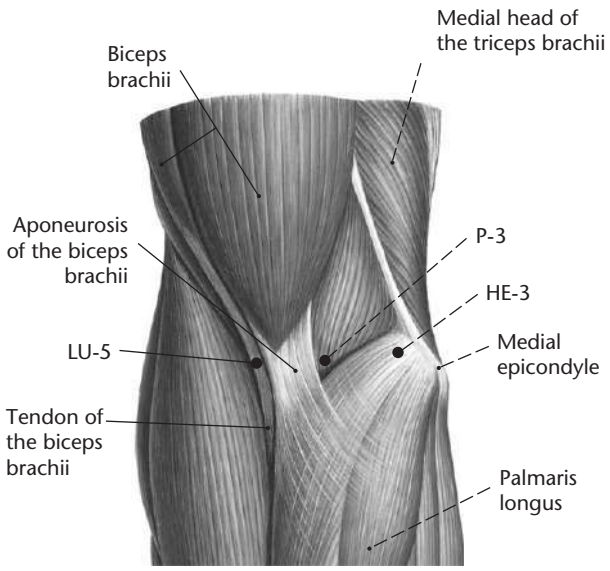

Right arm anterior aspect

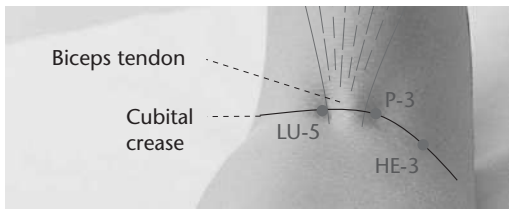

Right arm

Location

With the elbow flexed, this point is in a depression between the ulnar end of the transverse cubital crease and the medial epicondyle of the humerus.

How to find

This point should be located with the elbow slightly flexed and the forearm in a supinated position. Find the ulnar end of the transverse cubital crease and palpate towards the medial epicondyle of the humerus (→ 3.2.2). **HE-3** is located in a depression approximately midway between → **P-3** and the medial epicondyle. **Or**: With the elbow fully flexed, **HE-3** is located directly at the medial end of the transverse crease.
Located on the same level are → **P-3** (ulnar to the tendon of the biceps muscle), → **LU-5** (radial to the tendon of the biceps muscle) and → **L.I.-11** (at the radial end of the crease).

Needling

0.5–1.5 cun obliquely in a distal or proximal direction or vertically towards → **L.I.-11**.

Actions/Indications

- Transforms Phlegm, clears Heat (from the heart), calms the *shen*
- Opens the channel and *luo*-vessels

Special features

He-sea point, Water point. Important local point, important calming point.

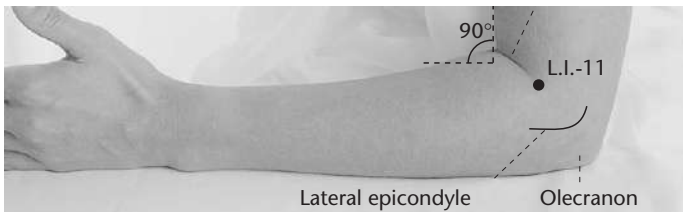

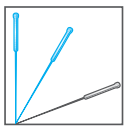
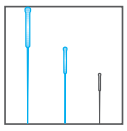

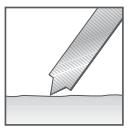

Ch04.5-F10028.qxd 2/22/08 4:38 PM Page 210

HE-4 *Spirit Path* LINGDAO

Location

1.5 cun proximal to the transverse wrist crease, on the radial side of the tendon of the flexor carpi ulnaris muscle.

How to find

The flexor carpi radialis muscle has a distinct tendon which can be easily palpated ulnar and proximal to the wrist crease. It attaches to the pisiform bone (→ 3.3.3). **HE-4** is located directly on the radial side of the tendon.

→ **HE-5**, → **HE-6** and → **HE-7** are located at 0.5 cun intervals distal to **HE-4**. → **LU-7** is located approximately 1.5 cun proximal to the wrist crease on the radial aspect of the forearm.

Needling

0.3–1 cun vertically or obliquely in a proximal or distal direction.

Actions/Indications

- Opens the channel
- Nourishes the Heart, calms the *shen*
- Strengthens the voice

Special features

Jing-river point, Metal point.

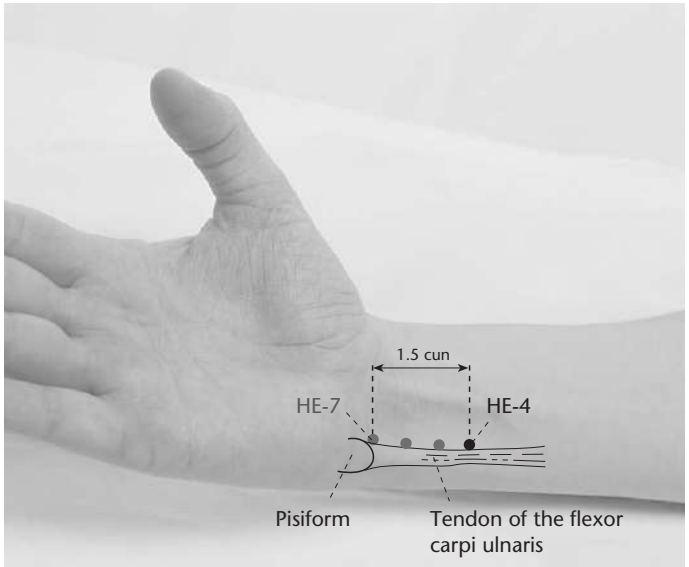

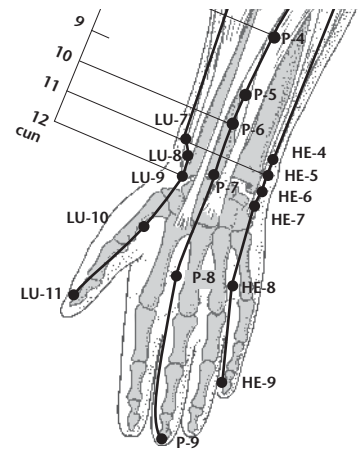

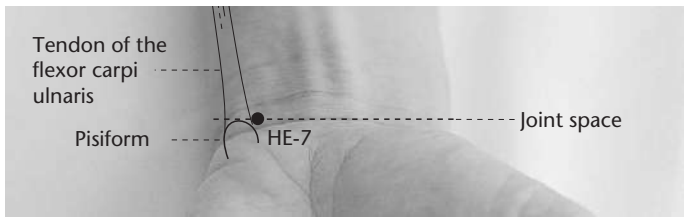

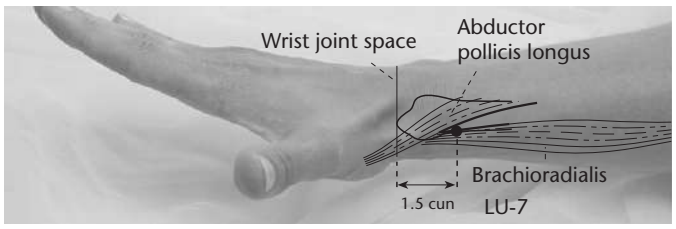

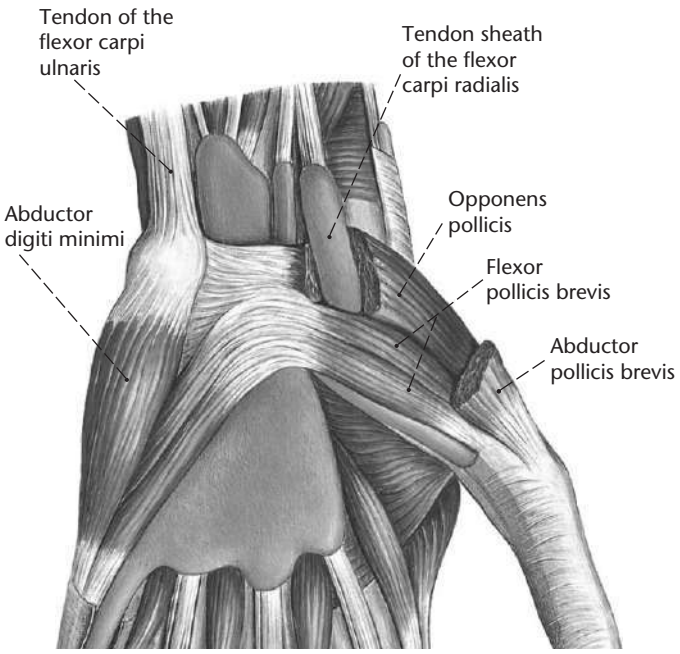

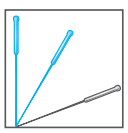
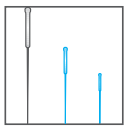

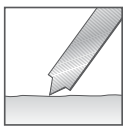

Ch04.5-F10028.qxd 2/22/08 4:38 PM Page 211

Penetrating the Interior TONGLI HE-5

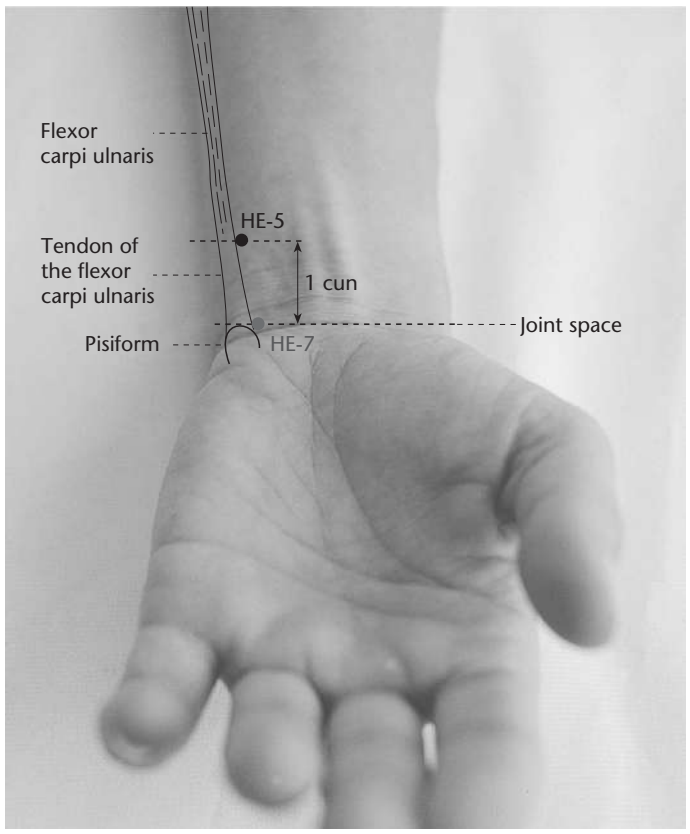

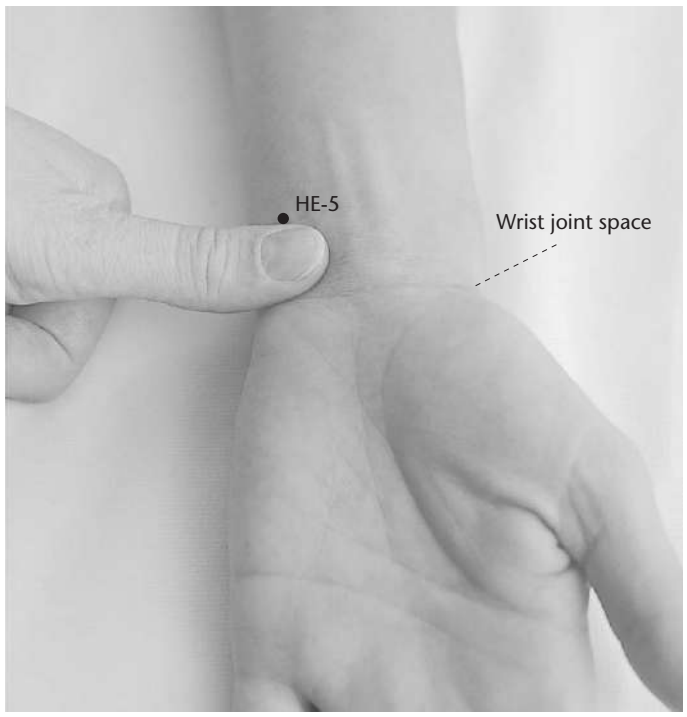

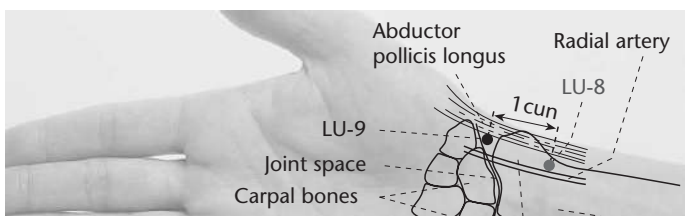

Location

1 cun proximal to the ventral wrist joint space ('most distal transverse wrist crease'), on the radial side of the tendon of the flexor carpi ulnaris muscle.

How to find

Position the forearm in a relaxed fashion. The location of the ventral wrist crease is variable; therefore, the proximal border of the pisiform bone (→ 3.3.2) provides a more reliable landmark. → **HE-7** is located on the wrist joint space, radial to the distinct tendon of the flexor carpi ulnaris muscle. Locate **HE-5** 1 cun proximal to → **HE-7**, also on the radial side of the tendon.
Located along a line on the radial aspect of the tendon are → **HE-6** (0.5 cun proximal to → **HE-7**) and → **HE-4** (1.5 cun proximal to **HE-7**). → **LU-8** is also located 1 cun proximal to the ventral wrist joint space, but on the radial aspect of the forearm.

Needling

Vertically 0.3–0.5 cun or obliquely 0.5–1 cun in a proximal or distal direction. Caution: Needling may be painful.

Actions/Indications

- Regulates and strengthens the Heart Qi (main point!) and the Heart rhythm
- Calms the *shen*
- Benefits the tongue
- Benefits the Bladder (via the *taiyang* channel)
- Opens the channel and *luo*-connecting vessels

Special features

Luo-connecting point, Ma Dan Yang Heavenly Star point. Main point for regulating and strengthening the Heart Qi.

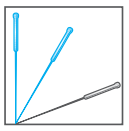
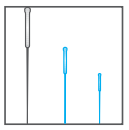

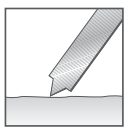

Ch04.5-F10028.qxd 2/22/08 4:38 PM Page 212

HE-6 *Yin Cleft* YINXI

Location
0.5 cun proximal to the transverse wrist crease, on the radial side of the tendon of the flexor carpi ulnaris muscle.

How to find
The flexor carpi ulnaris muscle has a distinct tendon on the ventral aspect of the forearm that can be easily palpated ulnar and proximal to the wrist. **HE-6** is located on the radial aspect of the tendon, 0.5 cun proximal to → **HE-7** (located on the joint space).
→ **HE-5** and → **HE-4** are located proximally on the same line at 0.5 cun intervals.

Needling
0.3–1 cun vertically or obliquely in a proximal or distal direction. Caution: Ulnar artery/nerve. Painful point!

Actions/Indications
- Tonifies and regulates the Heart Yin and Blood, clears empty Heat in the Heart, calms the *shen*

Special features
Xi-cleft point

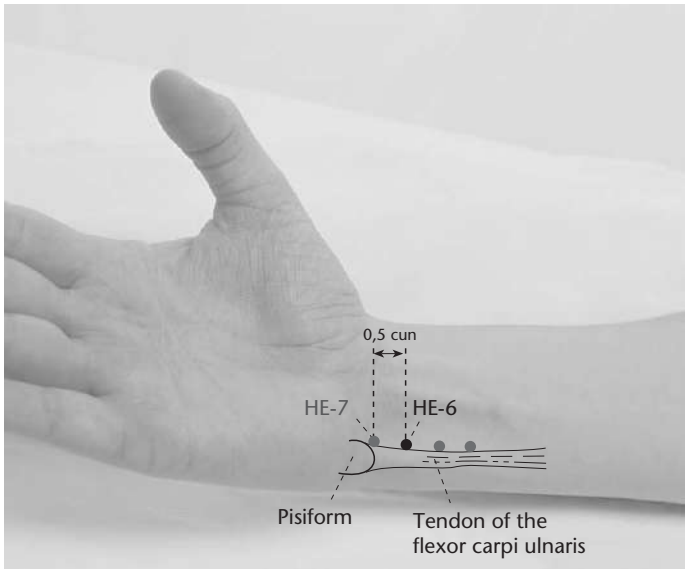

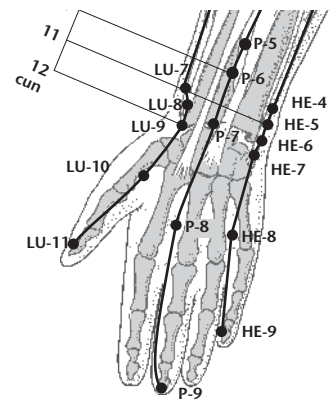

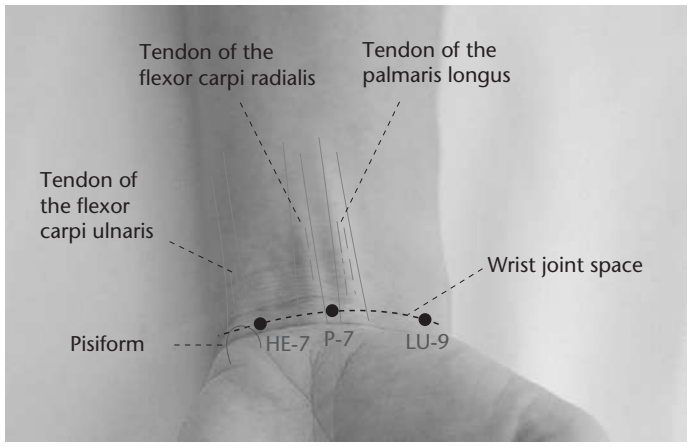

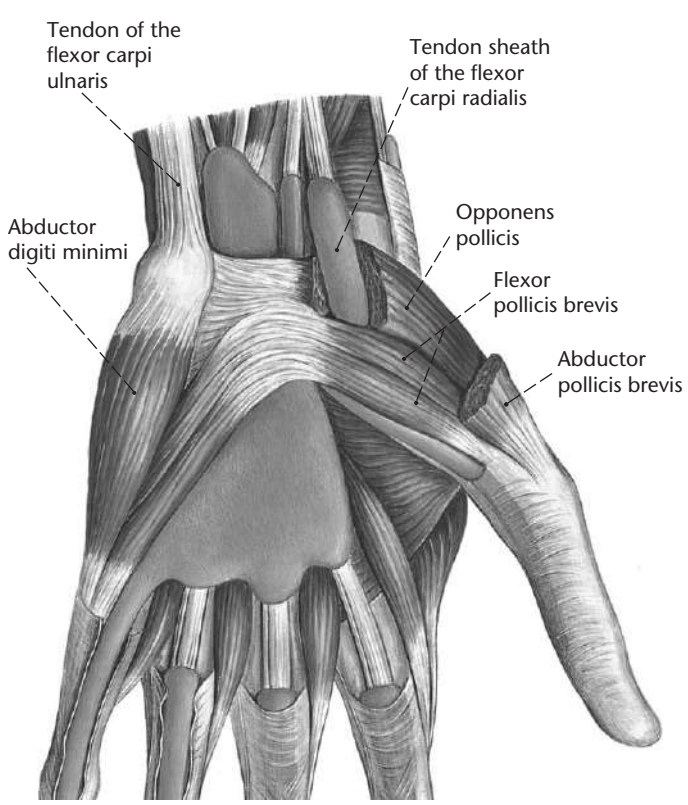

Ch04.5-F10028.qxd 2/22/08 4:38 PM Page 213

Spirit Gate SHENMEN HE-7

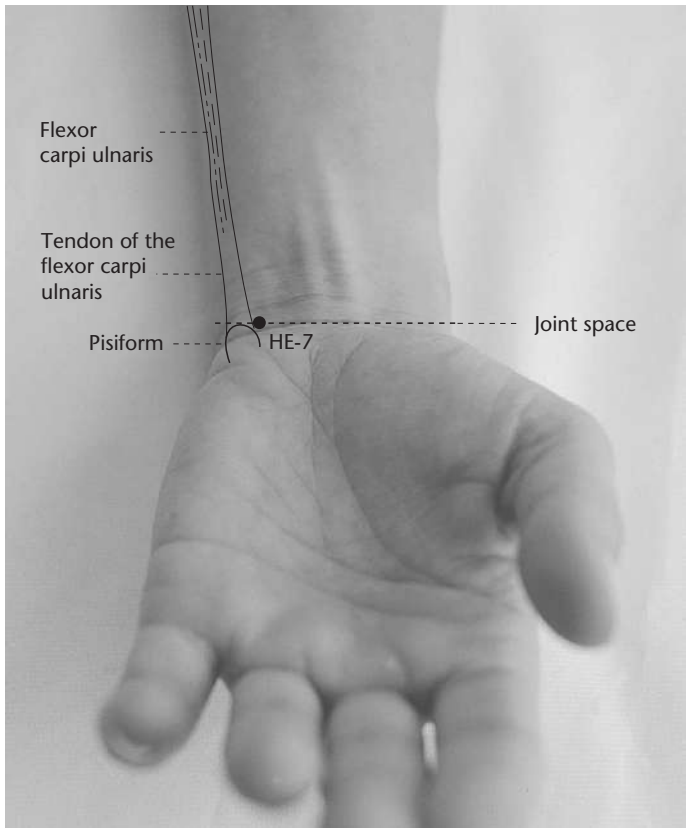

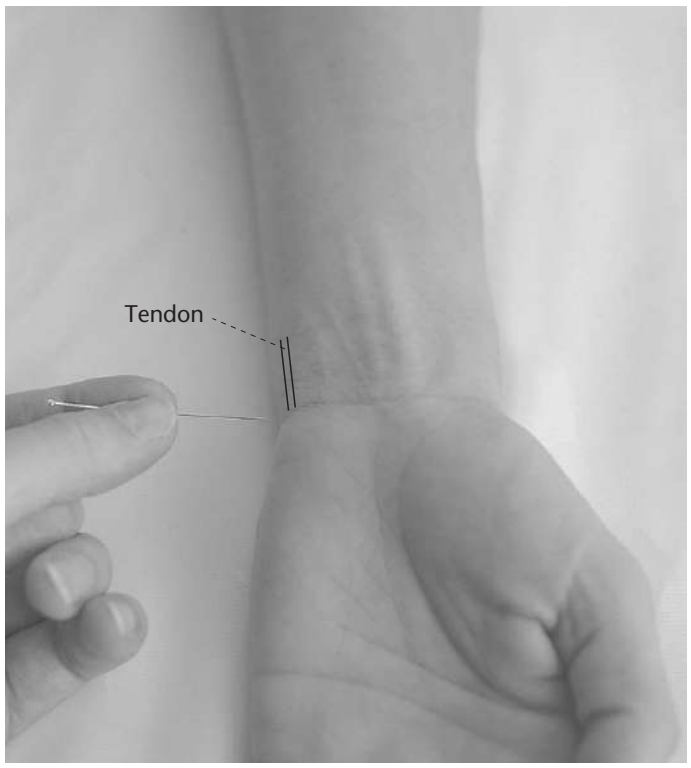

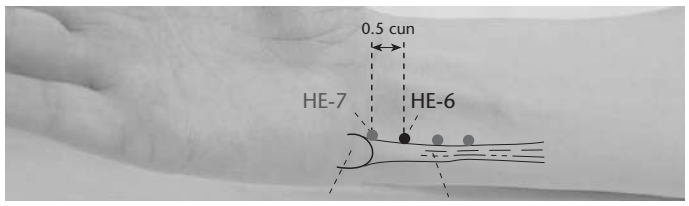

Location

On the ventral aspect of the wrist joint space ('most distal wrist crease'), radial to the insertion of the tendon of the flexor carpi ulnaris muscle.

How to find

Position the forearm in a relaxed supinated position. As the location of the wrist crease can vary, the proximal border of the pisiform bone (→ 3.3.3) should be used for reference instead. **HE-7** is located radial to the insertion of the tendon of the flexor carpi ulnaris muscle, in a depression proximal to the pisiform bone.

→ **HE-6**, → **HE-5** and → **HE-4** are all located on a line on the radial aspect of the tendon at 0.5 cun intervals. Also located on the wrist crease are → **P-7** (between the tendons of the palmaris longus and the flexor carpi radialis muscles) and → **LU-9** (on the radial aspect, lateral to the radial artery).

Needling

0.3–0.5 cun vertically or transversely deep to the tendon (parallel to the wrist crease 0.5–0.8 cun). Caution: Ulnar artery/nerve. Painful point!

Actions/Indications

- Regulates and strengthens the Heart (Blood and Yin), calms the *shen*
- Clears Heat from the Heart channel
- Moves Qi locally for disorders of the wrist

Special features

Yuan-source point, *shu*-stream point, Earth point, sedation point. Major point for calming the *shen*.

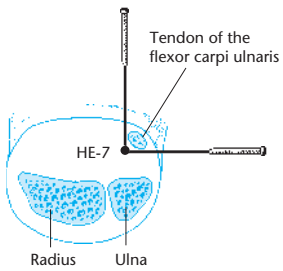

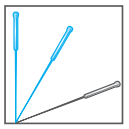
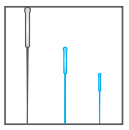

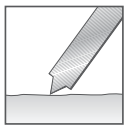
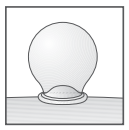

Ch04.5-F10028.qxd 2/22/08 4:38 PM Page 214

HE-8 *Lesser Palace* SHAOFU

Location

On the palm, between the 4th and 5th metacarpal bones.

How to find

When making a fist, **HE-8** is generally located below the tip of the little finger, in the depression between the 4th and 5th metacarpal bones. **HE-8** often lies between the two transverse palmar creases. It is often sensitive to pressure.

→ **P-8** is also located on the palm, generally below the tip of the middle finger when making a fist (between the 2nd and 3rd metacarpal bones).

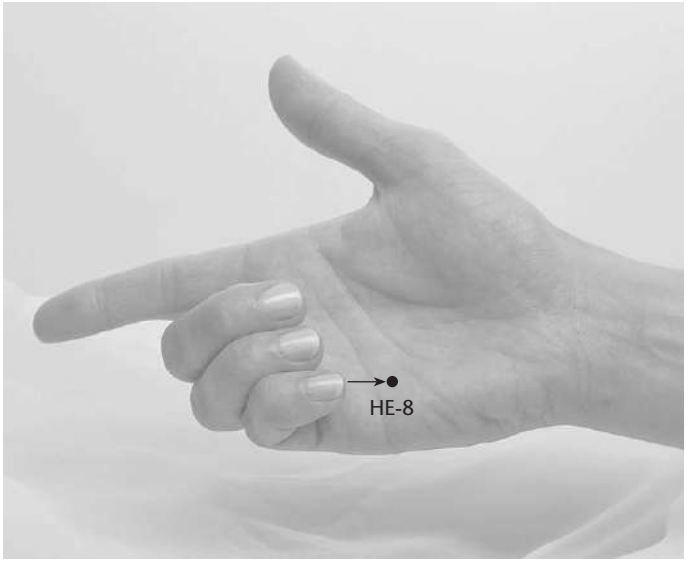

Needling

Vertically 0.5–1 cun. Caution: Needling this point may be painful. Provide a competitive stimulus by pressing the palm with the fingernail close to the point before inserting the needle. Ask the patient to breathe out during needle insertion.

Actions/Indications

- Clears Heat in the Heart and Small Intestine
- Regulates the Heart Qi for Liver Qi stagnation
- Calms the *shen* and strengthens the Heart Qi
- Opens the channel, especially for contractures of the fingers

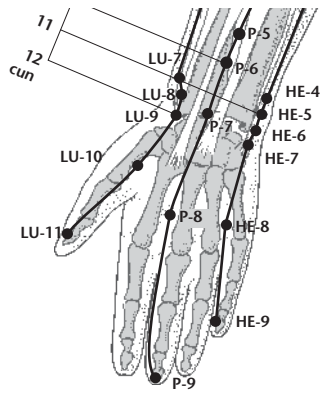

Special features

Ying-spring point, Fire point, *ben* point (Five Phases).

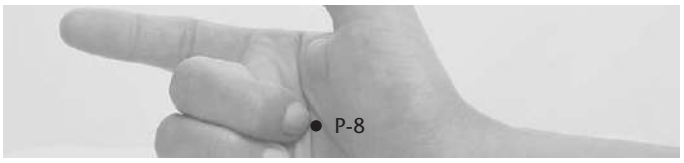

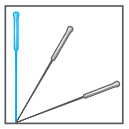

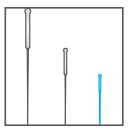

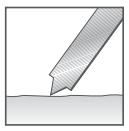

Ch04.5-F10028.qxd 2/22/08 4:38 PM Page 215

Lesser Rushing SHAOCHONG HE-9

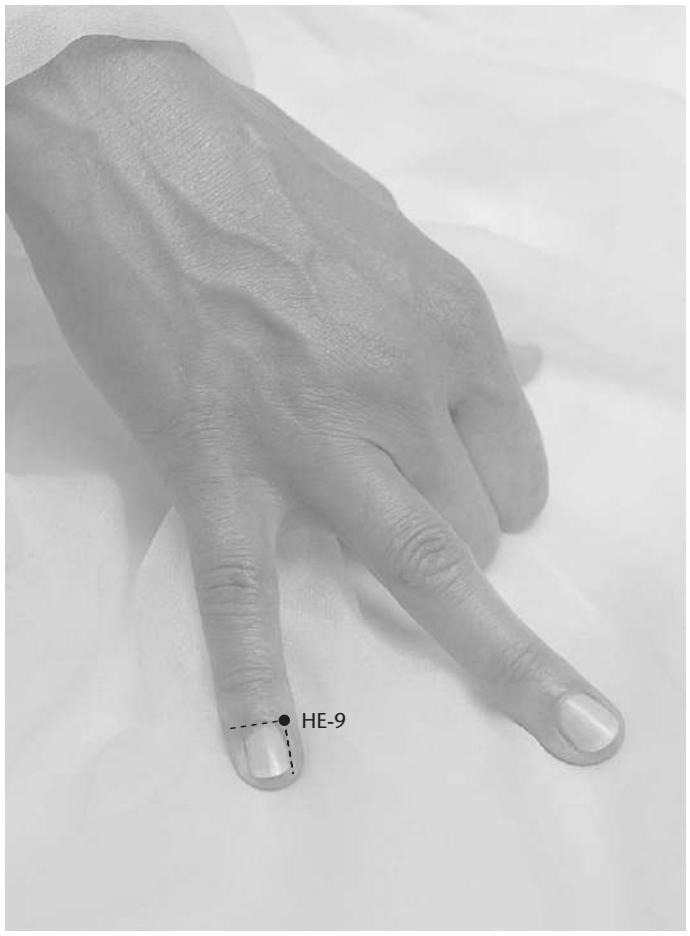

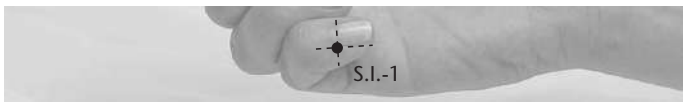

Location

On the little finger, 0.1 cun lateral to the radial corner of the nail.

How to find

This point is located at the junction of two tangents that border the radial and proximal aspect of the nail, approximately 0.1 cun from the actual corner of the nail. → **S.I.-1** is located on the opposite (ulnar) aspect of the nail.

Needling

0.2–0.3 cun vertically or obliquely in a proximal direction. For acute disorders (pain, inflammation), prick to bleed with a needle, lancet or three-edged needle. Caution: Painful point!

Actions/Indications

- Revives consciousness
- Clears Heat from the opposite end of the channel
- Regulates Qi in the chest and clears Heat in the Heart
- Opens the channel

Special features

Jing-well point, Wood point, tonification point, exit point.

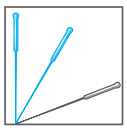
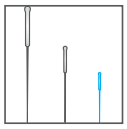

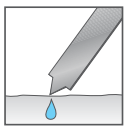
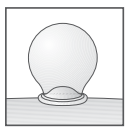

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 217

4.6 The Small Intestine Channel System – Hand *taiyang* (*shou tai yang jing luo*)

4.6.1 The Small Intestine Primary Channel (*shou tai yang jing*)

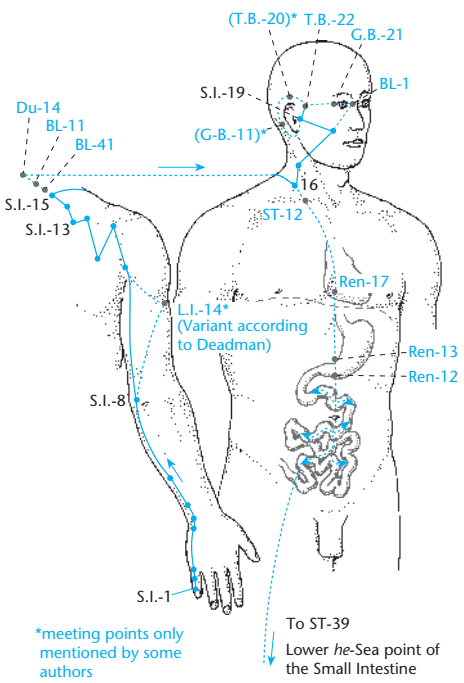

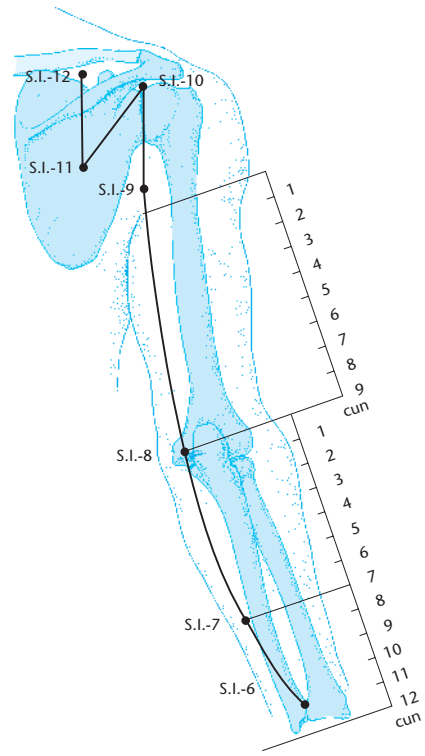

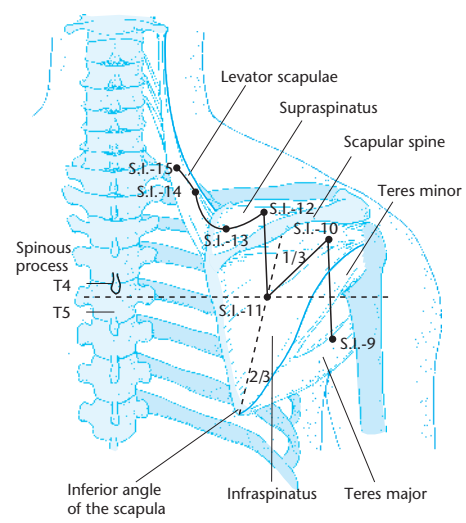

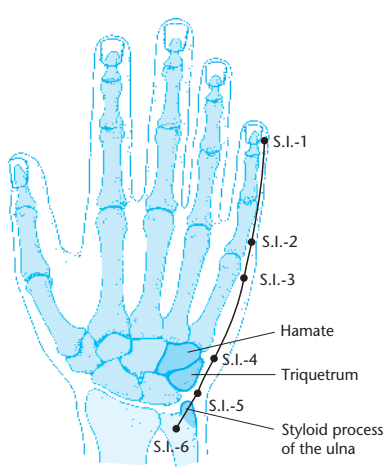

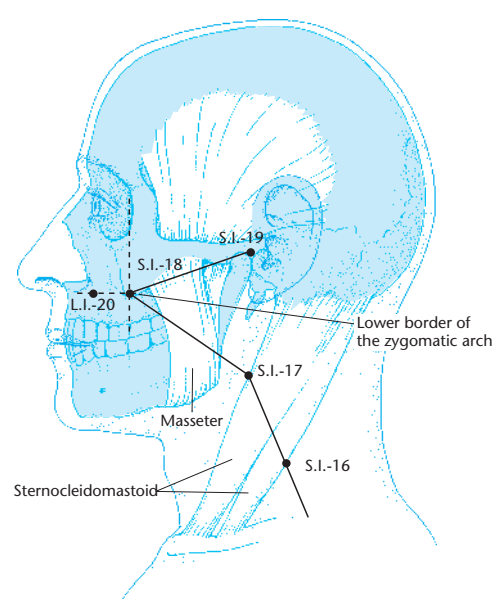

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 218

Pathway

The **external** pathway of the Small Intestine primary channel begins on the little finger, on the ulnar corner of the nail, at **S.I.-1** (*shaoze*). This point is reached by a branch that separates from the Heart primary channel at → **HE-9** (*shaochong*) (hand Yin–Yang connection of the second great circuit).

The channel runs along the ulnar border of the hand, passes the wrist and continues along the ulna to the medial aspects of the elbow.

- ➡ passes between the olecranon and the medial epicondyle, ascends the posterior aspect of the humerus to the shoulder joint and, according to some authors, meets → **L.I.-14**
- ➡ zig-zags across the scapula, passes → **BL-41** (*fufen*) and → **BL-11** (*dazhu*) and meets the extraordinary vessel *du mai* and the other Yang primary channels at → **Du-14** (*dazhui*), below the spinous process of C7
- ➡ continues anteriorly to the supraclavicular fossa to → **ST-12** (*quepen*).

An **internal** branch originates at → **ST-12** and runs inferiorly to the Heart (*xin*), the paired *zang*-Organ, continues to → **Ren-17** and descends along the oesophagus, penetrates the diaphragm and continues to the Stomach (*wei*). Here it crosses → **Ren-13** (*shangwan*) and → **Ren-12** (*zhongwan*), further descends and spreads in its pertaining *fu*-Organ, the Small Intestine (*xiaochang*). According to the *Neijing Ling Shu*, an **internal** branch courses from the Small Intestine (according to some authors from the Small Intestine primary channel) to → **ST-39** (*xiajuxu*), the lower *he*-sea point of the Small Intestine.

The **external** pathway of the Small Intestine primary channel continues from the supraclavicular fossa along the neck and ascends to the cheek to **S.I.-18** (*quanliao*).

From **S.I.-18**

- ➡ **a branch** courses to → **BL-1** (*jingming*) at the inner canthus of the eye, where it connects with the Bladder primary channel (six channel theory, Yang axes: *taiyang*)
- ➡ **a further branch** travels to the outer canthus of the eye, intersects with → **G.B.-1** (*tongziliao*), according to some authors also with → **T.B.-20** and → **G.B.-11**, continues to → **T.B-22** (*erheliao*) anterior to the ear and terminates anterior to the ear deep to **S.I.-19** (*tinggong*)

Clinical importance (➡ 1.2)

Exterior (*biao*) signs and symptoms: Mouth and tongue ulcers, cheek pain, neck pain, occipital stiffness, pain on the lateral aspect of the shoulder and the upper arm

Interior (*li*) or *zangfu*-Organ signs and symptoms: Pain and distension of the upper abdomen, radiating to the genital region, diarrhoea, epigastric pain with constipation

Connections of the Small Intestine primary channel

Connections with other channels

Heart primary channel (*shou shao yin jing*)

Connection: Hand Yin–Yang connection of the second great circuit
Location: **H-9** → **S.I.-1** (on the hand)
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Bladder primary channel (*zu tai yang jing*)

Connection: Paired according to the six channel theory (hand–foot pairing): *tai yang* (Yang axes of the second great circuit)
Location: **S.I.-18** → **BL-1** (on the head)
Circulation: Circadian (according to the Organ clock)
Importance: Above–below relationship

Connections with other *zangfu*-Organ systems

Heart (*xin*), Stomach (*wei*), **Small Intestine (*xiaochang*)**

4.6.2 The Small Intestine Divergent Channel (*shou tai yang jing bie*)

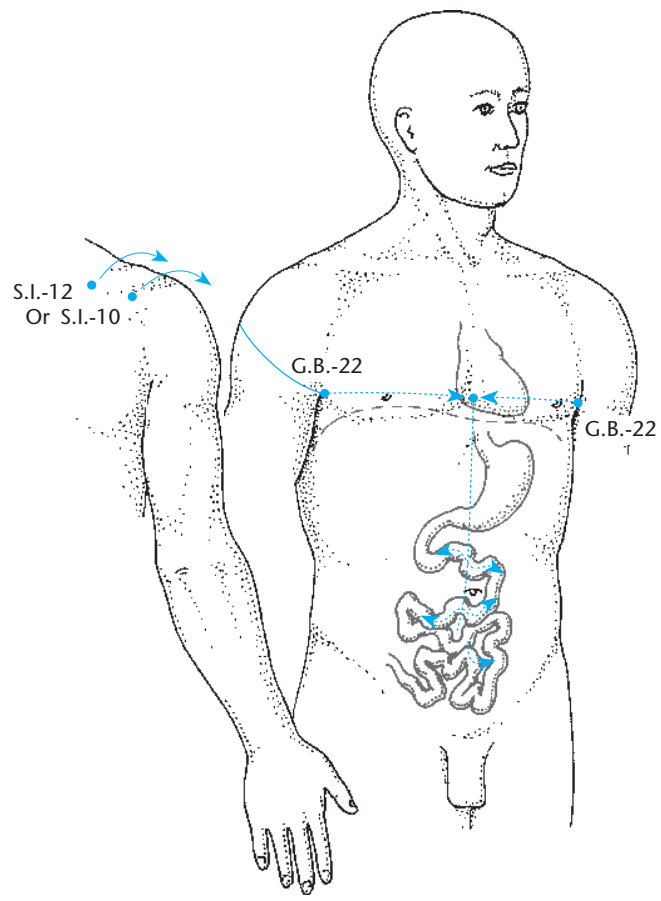

Pathway

The Small Intestine divergent channel separates from the Small Intestine primary channel at the shoulder, depending on the school of thought in the area of → **S.I.-12** (*bingfeng*) or in the area of → **S.I.-10** (*naoshu*)

- ➡ descends towards the axilla and continues to → **G.B.-22** (*yuanye*)
- ➡ penetrates the thorax and spreads in the Heart (*xin*) where, according to most authors, it forms the fourth confluence
- ➡ penetrates the diaphragm to connect with its pertaining *fu*-Organ, the Small Intestine

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 219

Clinical importance

- Strengthens the relationship between the Heart and the Small Intestine (*zangfu*-Organ system). It provides an explanation for the pattern of Heart Fire transferring to the Small Intestine and the Bladder.

4.6.3 The Small Intestine Sinew Channel (*shou tai yang yin jing jin*)

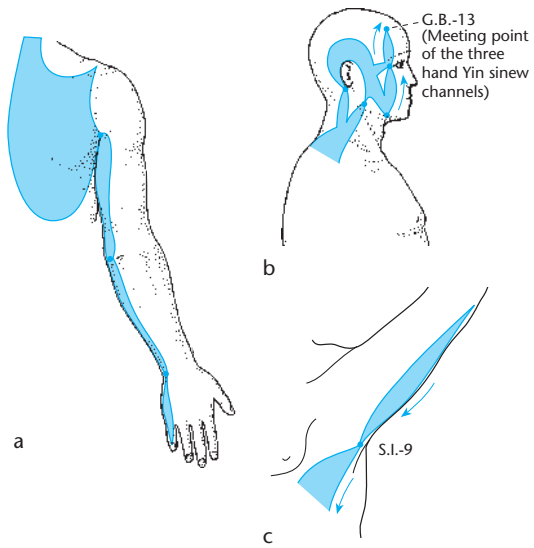

Pathway

The Small Intestine sinew channel begins on the ulnar aspect of the little finger at ➞ **S.I.-1** (*shaoze*)

- ➡ continues along the ulnar aspect of the little finger and the ulnar border of the hand
- ➡ runs to **S.I.-5** (*yanggu*) distal to the styloid process of the ulna and binds (*jie*) at the wrist
- ➡ travels along the ulnar aspect of the forearm to the elbow, continues to the axilla, binding at **S.I.-8** (*xiaohai*) and the posterior aspect of the axilla.

From here a branch courses into the axilla where it binds (*jie*). Another branch spreads across the scapula and ascends anterior to the Bladder and Gall Bladder sinew channels and posterior to the Stomach sinew channel to the lateral aspect of the neck, where it meets the Triple Burner sinew channel.

At the neck it divides into **two branches:**

- ➡ the **anterior branch** runs to the mandibular angle where it binds (*jie*), continues along the mandible to the ear and the zygomatic bone to bind (*jie*) at the lateral aspect of the orbit
- ➡ the **posterior branch** binds (*jie*) at the mastoid process (from here a small branch runs to the ear). The branch curves around the ear, descends to the cheek and binds (*jie*) at the insertion of the masseter muscle on the mandible, continues to the lateral aspect of the orbit and reaches ➞ **G.B.-13** (*benshen*) on the frontoparietal aspect of the head, where it meets the other hand Yang sinew channels.

Clinical importance

Pathology: Distending sensations, stiffness and pain of the little finger, impaired range of motion of the upper extremity and shoulder. Pain, stiffness and impaired range of motion of the cervical spine and occiput. Tinnitus and ear pain radiating to the mandible.

Indication: Mainly for *bi*-syndromes (painful obstruction syndrome) along the pathway of the S.I. channel. The area covered by the Small Intestine sinew channel is larger than that covered by the Small Intestine primary channel. This explains why Small Intestine primary channel points can also be used for disorders and diseases of the cervical and thoracic spine and the whole face (for example frontal and lateral headaches).

4.6.4 The Small Intestine *luo*-Connecting Vessel System (*shou tai yang luo mai*)

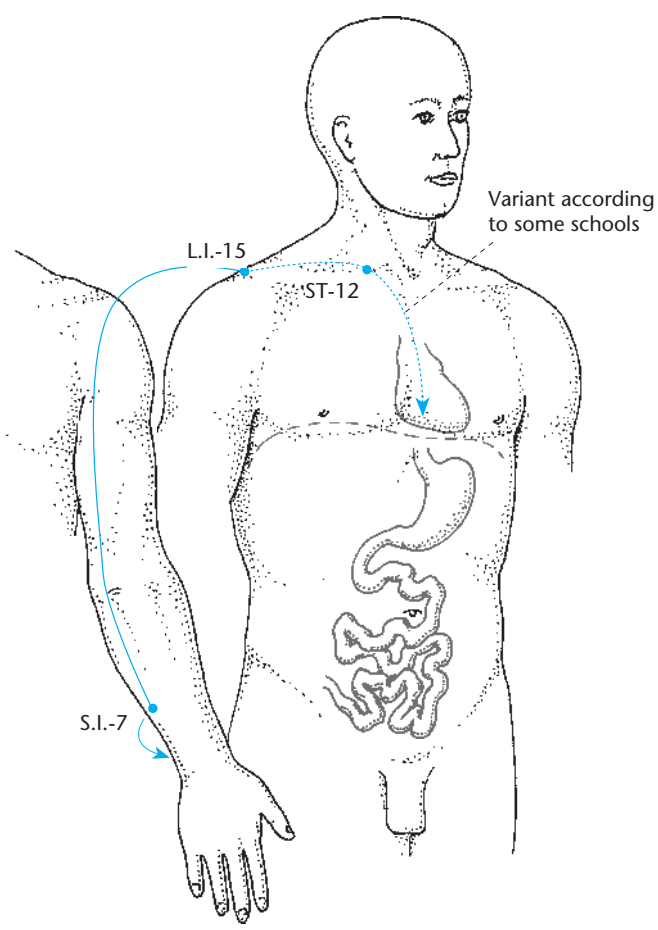

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 220

Pathway

The S.I. *luo*-connecting vessel separates from the Small Intestine primary channel at its *luo*-connecting point **S.I.-7** (*zhizheng*) (→ 8.1.2). It forms a three-dimensional reticular network, dividing into multiple branches and sub-branches (*sun luo*, *fu luo*, *xue luo* → 1.5) within the surrounding tissue.

- Horizontal divisions run to the Interiorly–Exteriorly paired Heart primary channel, according to some schools of thought (for example Ngyen Van Nghi → Appendix) they travel as a **transverse** S.I. *luo*-connecting vessel to the *yuan*-source point → **HE-7** (*shenmen*).
- A **longitudinal** division runs to the medial condyle of the humerus, ascends to the shoulder to → **L.I.-15** (*jianyu*). According to some schools of thought the vessel spreads and terminates here, while according to other schools it traverses the acromion, continues to → **ST-12** (*quepen*) in the supraclavicular fossa, penetrating the thorax in order to connect with the Heart.

Clinical importance (➡ 8.1.2)

Pathology

- **Excess (*shi*)**: Weak and unstable joints, weakness and paralysis of the elbow and the arm
- **Deficiency (*xu*)**: Warts

4.6.5 Cutaneous Region (*tai yang pi bu*)

See description and figures → 1.6.

4.6.6 Points of the Small Intestine Primary Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): S.I.-4** (*wangu*)
- ***Luo*-connecting point (→ 8.1.2.): S.I.-7** (*zhizheng*)
- ***Xi*-cleft point (→ 8.1.3): S.I.-6** (*yanglao*) ■■
- **Associated Back-*shu* point (→ 8.1.4): BL-27** (*xiaochangshu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): Ren-4** (*guanyuan*) ■■
- **Five *shu*-transporting point (→ 8.1.6):**
  jing-well point (Metal): **S.I.-1** (*shaoze*)
  ying-spring point (Water): **S.I.-2** (*qiangu*)
  shu-stream point (Wood), tonification point: **S.I.-3** (*houxi*) ■■
  jing-river point (Fire), *ben* point: **S.I.-5** (*yanggu*)
  he-sea point (Earth), tonification point: **S.I.-8** (*xiaohai*)
- ***Hui*-meeting point (→ 8.1.7): –**
- **Opening point (→ 8.1.8) of the *du mai*: S.I.-3** (*houxi*) ■■
- **Lower *he*-sea points (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the BL channel, the *yang wei mai, yang qiao mai*: **S.I.-10** (*naoshu*)
  – with the L.I., T.B., G.B. channels: **S.I.-12** (*bingfeng*)
  – with the T.B. channel: **S.I.-18** (*quanlao*) ■■
  – with the T.B., G.B.: **S.I.-19** (*tinggong*) ■■
  – of other channels with the S.I. channel: **L.I.-14*, Du-14, BL-41, BL-11, ST-12, Ren-17*, Ren-13, Ren-12, G.B.-1, G.B.-11*, T.B.-20*, T.B.-22, BL-1**
- **Gao Wu command point (→ 8.1.11): –**
- **Window of Heaven point (→ 8.1.12): S.I.-16** (*tianchuang*), **S.I.-17** (*tianrong*) ■■
- **Points of the Four Seas (→ 8.1.13): –**
- **Ma Dan Yang Heavenly Star point (→ 8.1.14): –**
- **Sun Si Miao Ghost point (→ 8.1.15): –**

Points according to region

- **Local points (→ 8.2.1):** face and cheek – **S.I.-18** (*quanlao*) ■■; jaw – **S.I.-19** (*tinggong*) ■■; neck – **S.I.-17** (*tianrong*) ■■; ears – **S.I.-19** (*tinggong*) ■■; shoulder – **S.I.-9** (*jianzhen*), **S.I.-10** (*naoshu*), **S.I.-11** (*tianzong*), **S.I.-12** (*bingfeng*) ■■, **S.I.-13** (*quyuan*), **S.I.-14** (*jianwaishu*) ■■; elbow – **S.I.-8** (*xiaohai*); hand – **S.I.-4** (*wangu*), **S.I.-5** (*yanggu*); finger (numbness and pain) – **S.I.-3** (*houxi*) ■■; finger (stiffness) – **S.I.-3** (*houxi*) ■■, **S.I.-5** (*yanggu*)
- **Adjacent points (→ 8.2.1):** shoulder – **S.I.-9–S.I.-15**; finger – **S.I.-5** (*yanggu*)
- **Distal points (→ 8.2.1):** for the occiput – **S.I.-3** (*houxi*) ■■; for the eyes – **S.I.-6** (*yanglao*) ■■; for the cervical and thoracic spine – **S.I.-3** (*houxi*) ■■; for the thoracic spine – **S.I.-6** (*yanglao*) ■■; for the lumbar region – **S.I.-3** (*houxi*) ■■, **S.I.-6** (*yanglao*) ■■

Specific points according to the channel pathway (in numerical order)

- **S.I.-1** (*shaoze*): *jing*-well point (Metal)
- **S.I.-2** (*qiangu*): *ying*-spring point (Water)
- **S.I.-3** (*houxi*) ■■: *shu*-stream point (Wood), tonification point, opening point (→ 8.1.8) of the *du mai*, distal point for the occiput and the lumbar region (→ 8.2.1); local point for the fingers (→ 8.2.1); distal point for the cervical and thoracic spine (→ 8.2.1)
- **S.I.-4** (*wangu*): *yuan*-source point (→ 8.1.1); local point for the hand (→ 8.2.1)
- **S.I.-5** (*yanggu*): *jing*-river point (Fire), *ben* point (Five Phases); local point for the hand and fingers (stiffness) (→ 8.2.1); regional point for the fingers (→ 8.2.1)
- **S.I.-6** (*yanglao*) ■■: *xi*-cleft point (→ 8.1.3); distal point for the eyes, the thoracic and lumbar spine (→ 8.2.1)
- **S.I.-7** (*zhizheng*): *luo*-connecting point (→ 8.1.2)
- **S.I.-8** (*xiaohai*): *he*-sea point (Earth), sedation point; local point for the elbow (→ 8.2.1)
- **S.I.-9** (*jianzhen*): local point for the shoulder (often sensitive to pressure) (→ 8.2.1)
- **S.I.-9 – S.I.-15**: adjacent points for the shoulder (often sensitive to pressure) (→ 8.2.1)

* According to Deadman et al 1998.

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 221

- **S.I.-10** (*naoshu*): *jiaohui*-meeting point with the BL channel, the *yang wei mai, yang qiao mai* (→ 8.1.10); local point for the shoulder (often sensitive to pressure) (→ 8.2.1)
- **S.I.-11** (*tianzong*) ■■: local point for the shoulder (often sensitive to pressure) (→ 8.2.1)
- **S.I.-12** (*bingfeng*): *jiaohui*-meeting point with the L.I., T.B. and G.B. channels (→ 8.1.10); local point for the shoulder (often sensitive to pressure) (→ 8.2.1)
- **S.I.-13** (*quyuan*): local point for the shoulder (often sensitive to pressure) (→ 8.2.1)
- **S.I.-14** (*jianwaishu*) ■■: local point for the shoulder (often sensitive to pressure) (→ 8.2.1)
- **S.I.-16** (*tianchuang*): Window of Heaven point (→ 8.1.12)
- **S.I.-17** (*tianrong*) ■■: Window of Heaven point (→ 8.1.12); local point for the neck (→ 8.2.1)
- **S.I.-18** (*quanlao*) ■■: local point for the face and cheek (→ 8.2.1); *jiaohui*-meeting point with the T.B. channel (→ 8.1.10)
- **S.I.-19** (*tinggong*) ■■: *jiaohui*-meeting point with the T.B., G.B. channels (→ 8.1.10); local point for disorders of the ear and jaw (→ 8.2.1)

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 223

Lesser Marsh SHAOZE S.I.-1

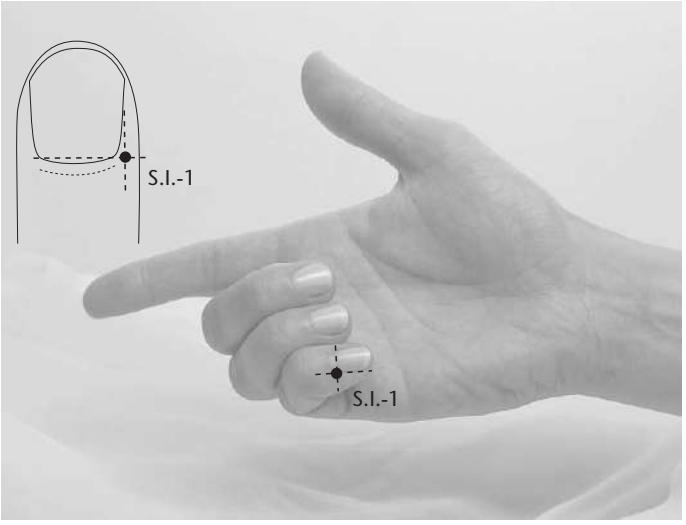

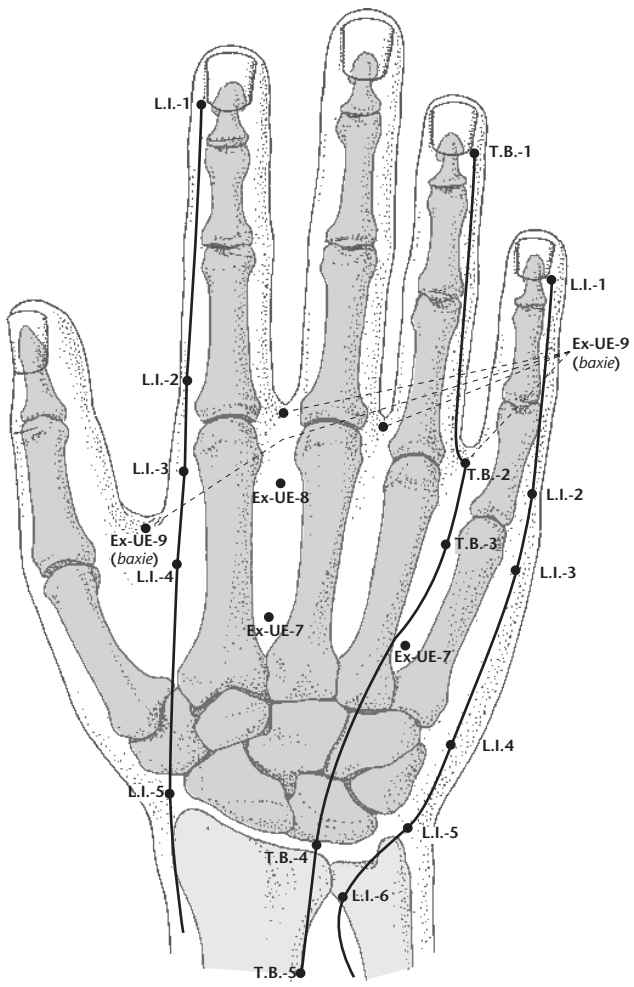

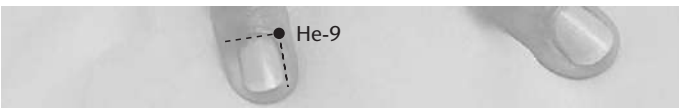

Location

On the little finger, 0.1 cun from the ulnar corner of the nail.

How to find

This point is located at the junction of two tangents drawn along the lateral and proximal borders of the nail, approximately 0.1 cun from the actual corner of the nail. → **HE-9** is located at the radial corner of the nail.

Needling

0.1–0.2 cun vertically or obliquely in a proximal direction or prick to bleed with a needle, lancet or three-edged needle. Caution: Painful point.

Actions/Indications

- Expels Wind and Heat
- Opens the orifices
- Revives consciousness
- Opens the channel
- Benefits the breasts

Special features

Jing-well point, Metal point, entry point.

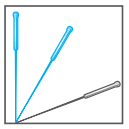
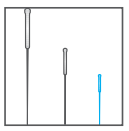

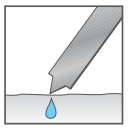

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 224

S.I.-2 *Front Valley* QIANGU

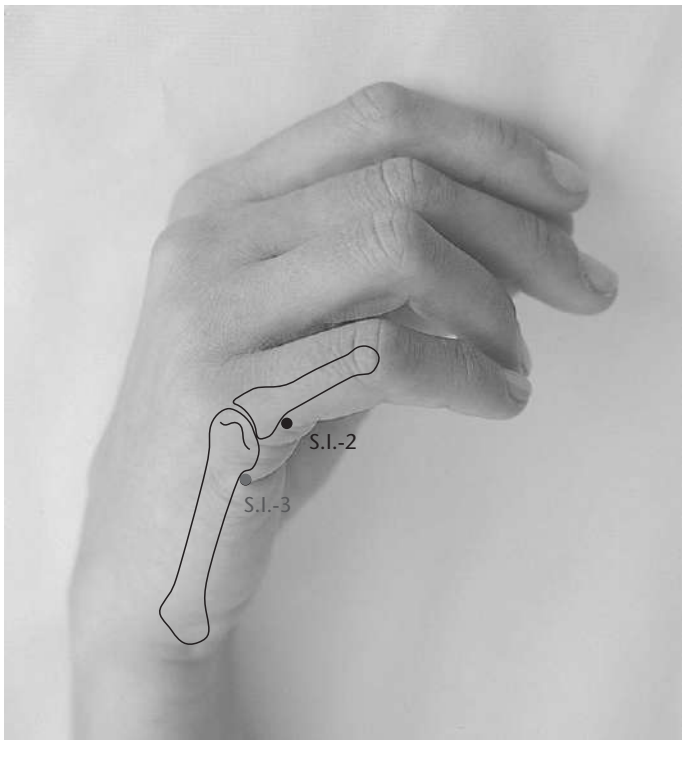

Location

On the ulnar aspect of the little finger, distal to the metacarpophalangeal joint, at the junction of the shaft and the base of the proximal phalanx.

How to find

On the ulnar aspect of the little finger, palpate along the border of the 'red and white' skin in the direction of the metacarpophalangeal joint, until you can clearly feel the junction between the shaft and the base. **S.I.-2** is located at the junction of the shaft and the base and slightly inferior (palmar) to the exterior curvature of the bone. → **L.I.-2** is located in a comparable position on the proximal phalanx of the index finger. In a comparable location on the foot are → **BL-66** (on the lateral aspect) and → **SP-2** (on the medial aspect).

Needling

0.2–0.5 cun obliquely in a proximal or distal direction and slightly towards the palm. Caution: Painful point.

Actions/Indications

- Drains Heat, expels Wind
- Disperses accumulations

Special features

Ying-spring point, Water point.

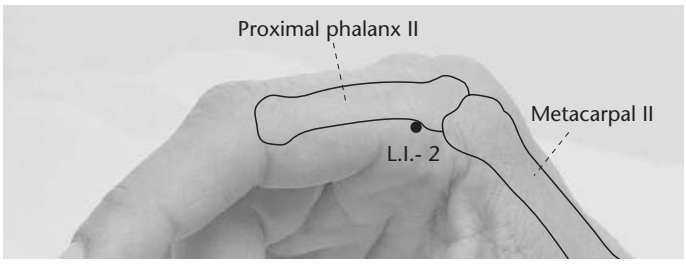

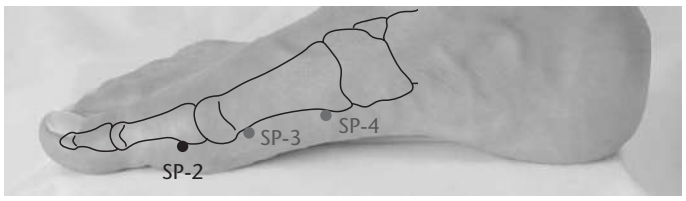

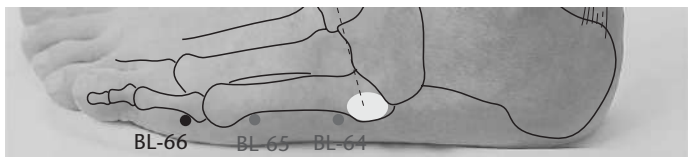

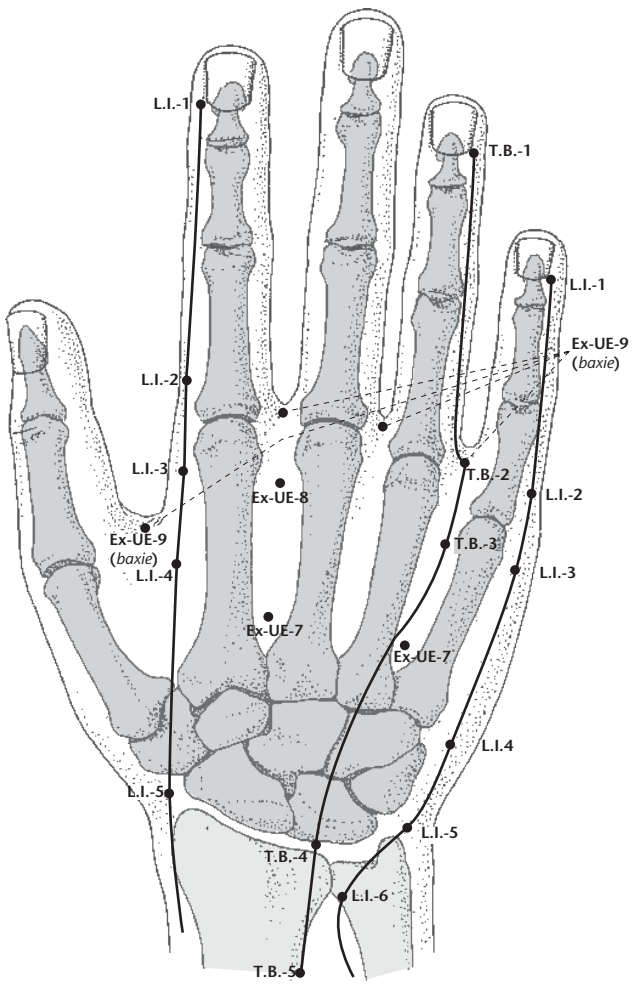

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 225

Back Stream HOUXI S.I.-3

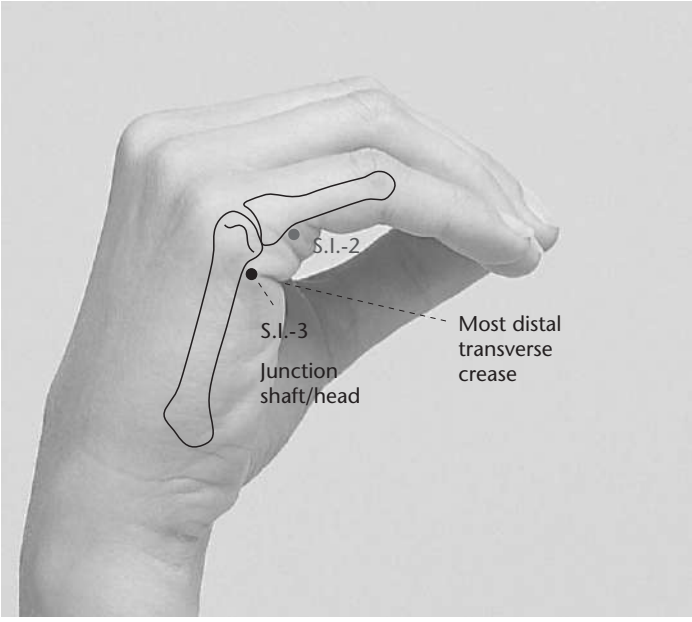

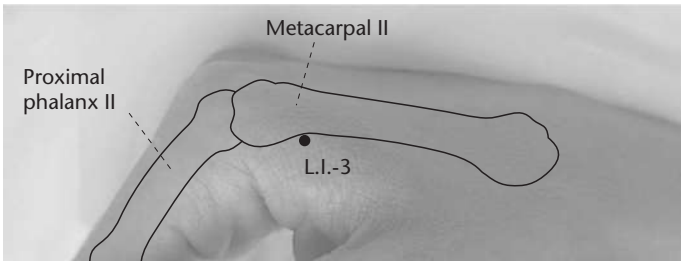

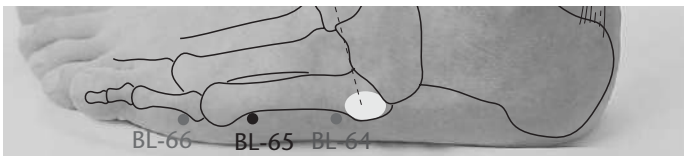

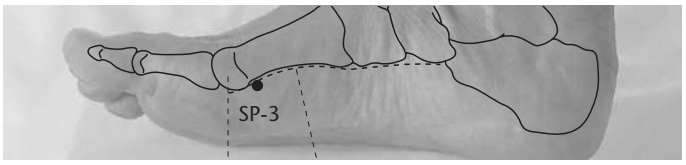

Location

On the ulnar border of the hand, in the depression proximal to the metacarpophalangeal joint, at the border of the 'red and white' skin.

How to find

Quick method: Slide from proximal (wrist) to distal (little finger) along the ulnar border of the hand, until the palpating finger comes to rest at the junction between the base and the head of the 5th metacarpal bone proximal to the metacarpophalangeal joint. Here, locate **S.I.-3. Or:** Locate, with the patient making a loose fist. Locate **S.I.-3** at the end of the most distal transverse crease (this often originates between the index and middle fingers and runs to the little finger). When making a fist, a small bulge forms at the end of the crease. There, locate **S.I.-3** in a palpable depression slightly proximal and dorsal to the bulge.

→ **L.I.-3** is located in a comparable position on the radial border of the hand. Located in a comparable position on the foot are → **SP-3** (on the medial aspect) and → **BL-65** (on the lateral aspect).

Needling

0.5–1 cun horizontally below the bone – that is, vertically to the skin towards the palm – or 0.2–0.3 cun obliquely in a distal direction. Caution: Painful point.

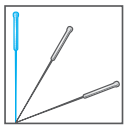

Actions/Indications

- Eliminates Wind and Heat from the *taiyang* (S.I., BL)
- Opens the channel and the *luo* vessels
- Clears Heat and benefits the orifices
- Regulates the *du mai*, calms the *shen*

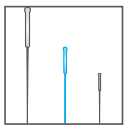

Special features

Shu-stream point, Wood point, tonification point, opening (master) point of the *du mai*. Important distal point for pain, stiffness and contractures along the course of the channel and for disorders of the cervical spine.

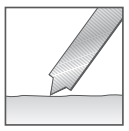

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 226

S.I.-4 *Wrist Bone* WANGU

Location
On the ulnar border of the hand, between the 5th metacarpal bone and the carpal bones, on the border of the red and white skin.

How to find
On the ulnar aspect of the hand, palpate along the shaft of the 5th metacarpal bone in a proximal direction, until the finger glides over the curvature of the base and into the cleft between the metacarpal and carpal bones.
→ **S.I.-5** is located on the same line, proximal to the pisiform bone and at the level of the wrist crease.

Needling
Vertically 0.3–0.5 cun. Caution: Painful point.

Actions/Indications
- Clears pathogenic factors from the *taiyang* channel
- Clears Damp-Heat in the Small Intestine

Special features
Yuan-source point

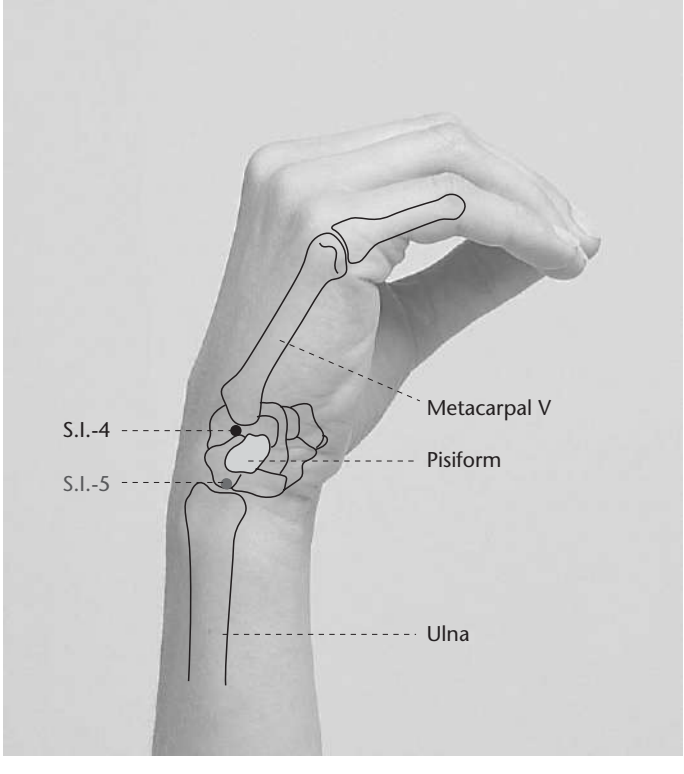

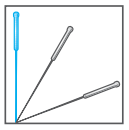
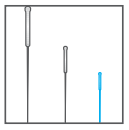

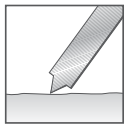

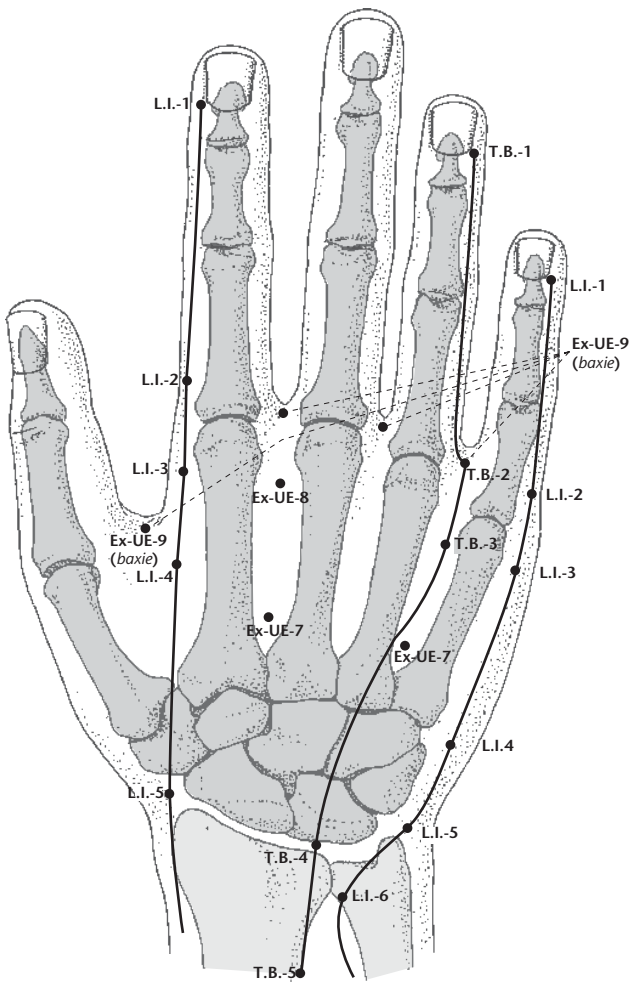

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 227

Yang Valley YANGGU S.I.-5

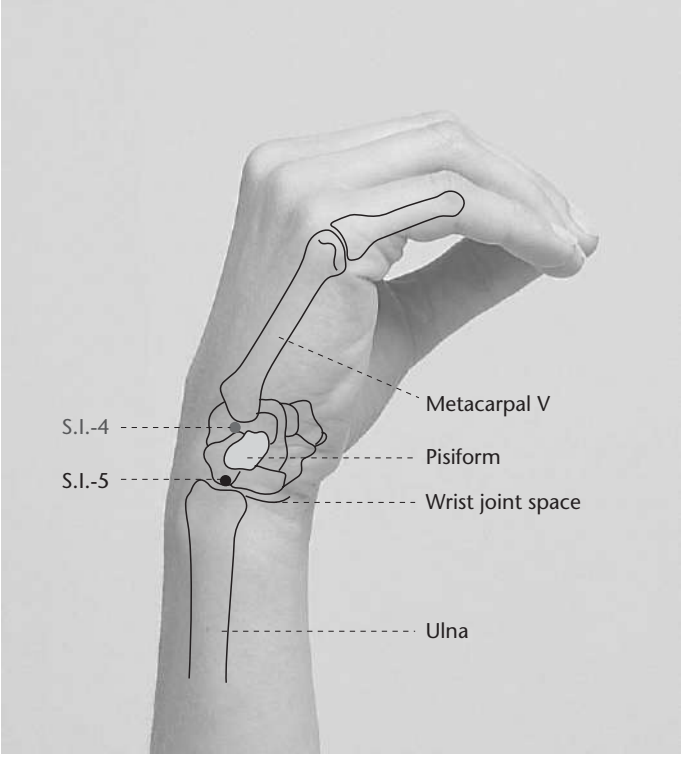

Location

On the ulnar aspect of the wrist, at the level of the lateral joint space.

How to find

By moving the wrist in a relaxed way, the wrist joint space can be palpated on the ulnar side. **S.I.-5** is located directly distal to the styloid process of the ulna (→ 3.3.3), on a line running along the ulnar border of the hand and continuing to the wrist. Located on the same line and distal to the prominence of the pisiform bone is → **S.I.-4**.

Needling

Vertically 0.3–0.5 cun

Actions/Indications

- Clears Heat and reduces swellings

Special features

Jing-river point, Fire point, *ben* point (Five Phases point).

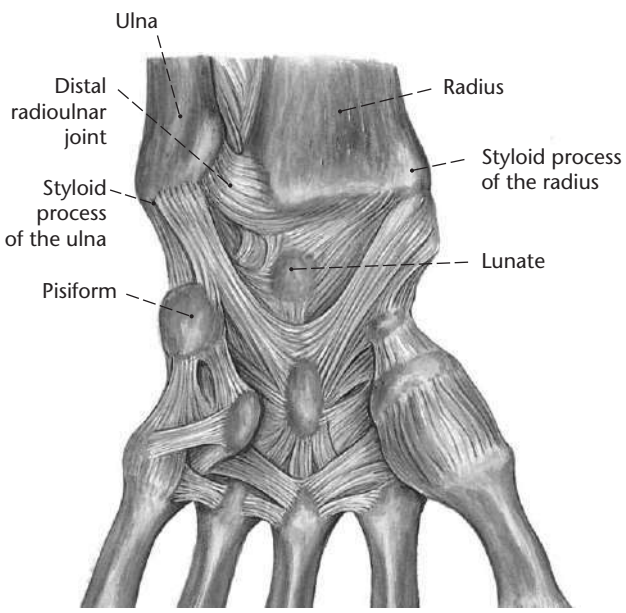

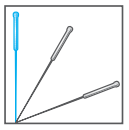
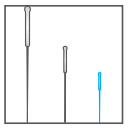

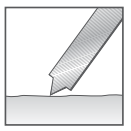
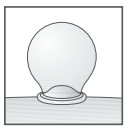

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 228

S.I.-6 *Support the Aged* YANGLAO

Location

In the depression radial and proximal to the styloid process of the ulna, which forms when the hand is turned from a pronated to a supinated position.

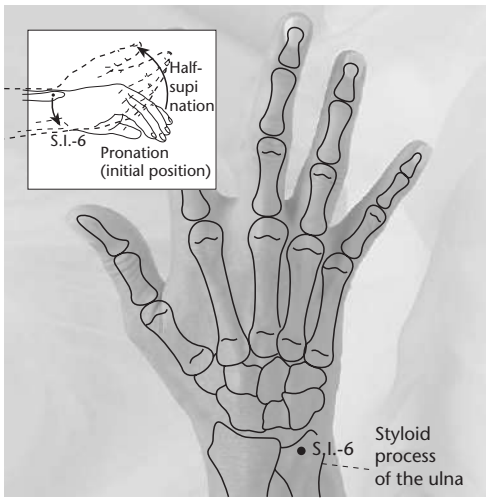

How to find

This point is best located with the elbow slightly flexed. Place the palpating finger on the distal portion of the styloid process of the ulna (→ 3.3.3). When turning the hand from a pronated to a supinated position (mnemonic: cut **br**ead and eat **soup**), the palpating finger will glide into a bony cleft at the proximal aspect of the styloid process (the tendon of the extensor carpi ulnari muscle runs along this groove). This cleft also becomes palpable when the patient rests the hand on the chest (half-supinated position). Locate **S.I.-6** in this bony groove.

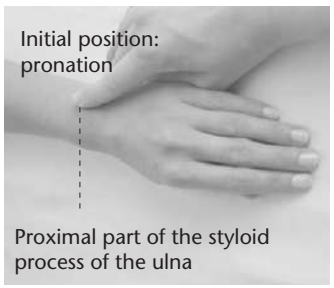

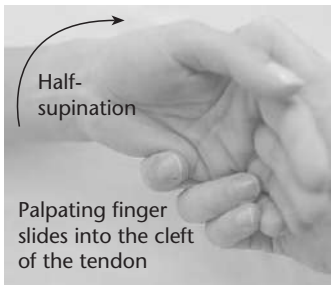

Needling

0.3–0.8 cun vertically or slightly obliquely in a proximal direction towards → **P-6** or along the tendon towards the elbow joint.

Actions/Indications

- Opens the channel, alleviates pain, benefits the shoulder and arm, alleviates acute conditions
- Benefits the eyes

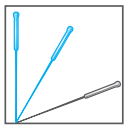

Special features

Xi-cleft point. Important distal point for pain and limited range of motion of the occiput and shoulder (for example, acute torticollis) and the lumbar region (for example, acute sciatica). For acute and painful impairment of motion, use strong reducing needling technique while the patient gently moves the affected region.

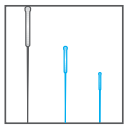

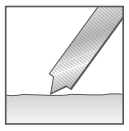

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 229

Branch of the Upright ZHIZHENG S.I.-7

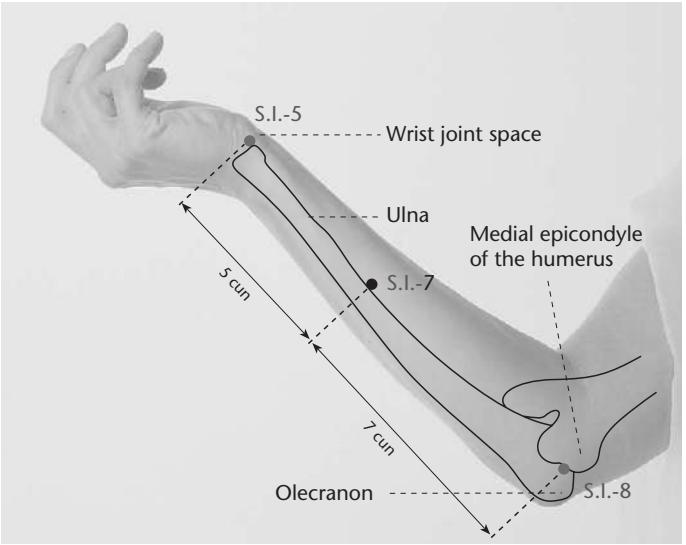

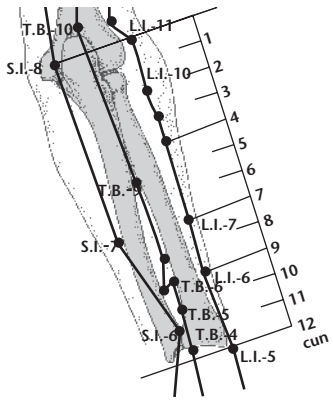

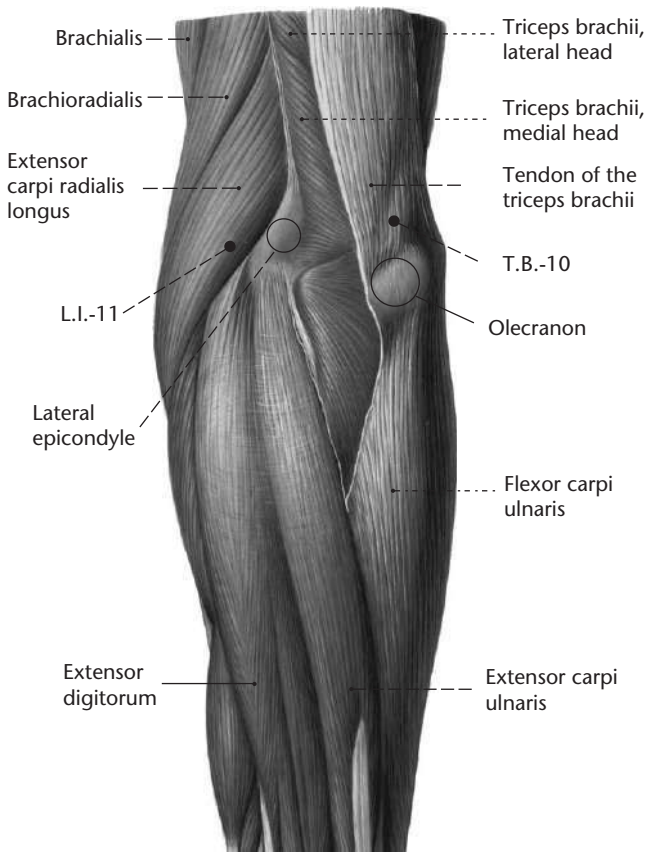

Location

5 cun proximal to the wrist joint, on the line connecting the ulnar wrist joint space with the ulnar sulcus of the elbow (→ **S.I.-5** to → **S.I.-8**) or 1 cun distal to the midpoint of this line.

How to find

Find the midpoint of the distance between → **S.I.-5** and → **S.I.-8** by using the spreading hands technique (→ 2.3.3): Place the little fingers on → **S.I.-5** and → **S.I.-8** respectively and span the hands evenly, with the thumbs joining at the midpoint. From the midpoint, measure 1 cun in a distal direction. The point lies between the palpable border of the ulna and the more anteriorly located flexor carpi ulnaris muscle.

Needling

Vertically 0.5–1 cun

Actions/Indications

- Opens the channel
- Releases the Exterior
- Calms the *shen*

Special features

Luo-connecting point

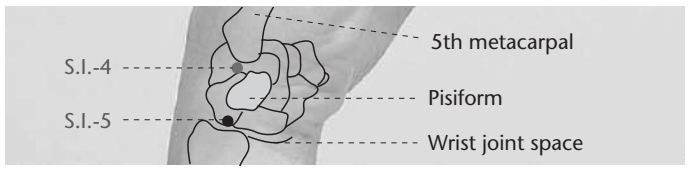

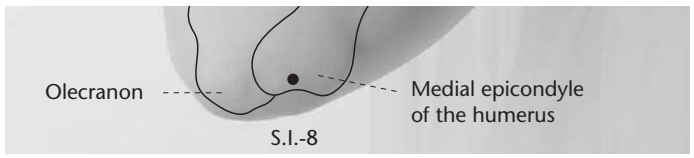

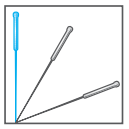
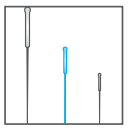

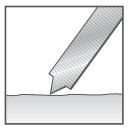

Ch04.6-F10028.qxd 2/22/08 9:47 PM Page 230

S.I.-8 *Small Sea* XIAOHAI

Location
With the elbow flexed, in the depression between the olecranon process of the ulna and the medial epicondyle of the humerus.

How to find
S.I.-8 is located on a line connecting the tip of the olecranon and the tip of the medial epicondyle, at the deepest point of the groove between those two bony protrusions.

Needling
0.3–0.5 cun vertically or obliquely with or against the flow of the channel. Caution: The ulnar nerve is located in this groove.

Actions/Indications
- Opens the channel

Special features
He-sea point, Earth point, sedation point.

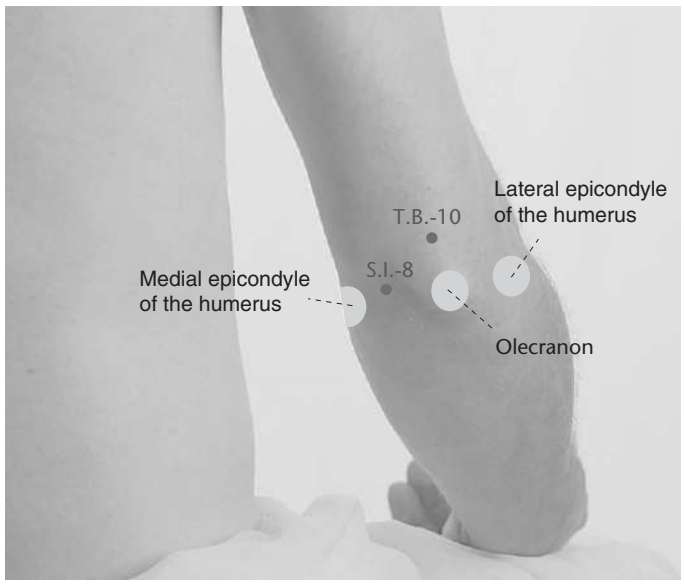

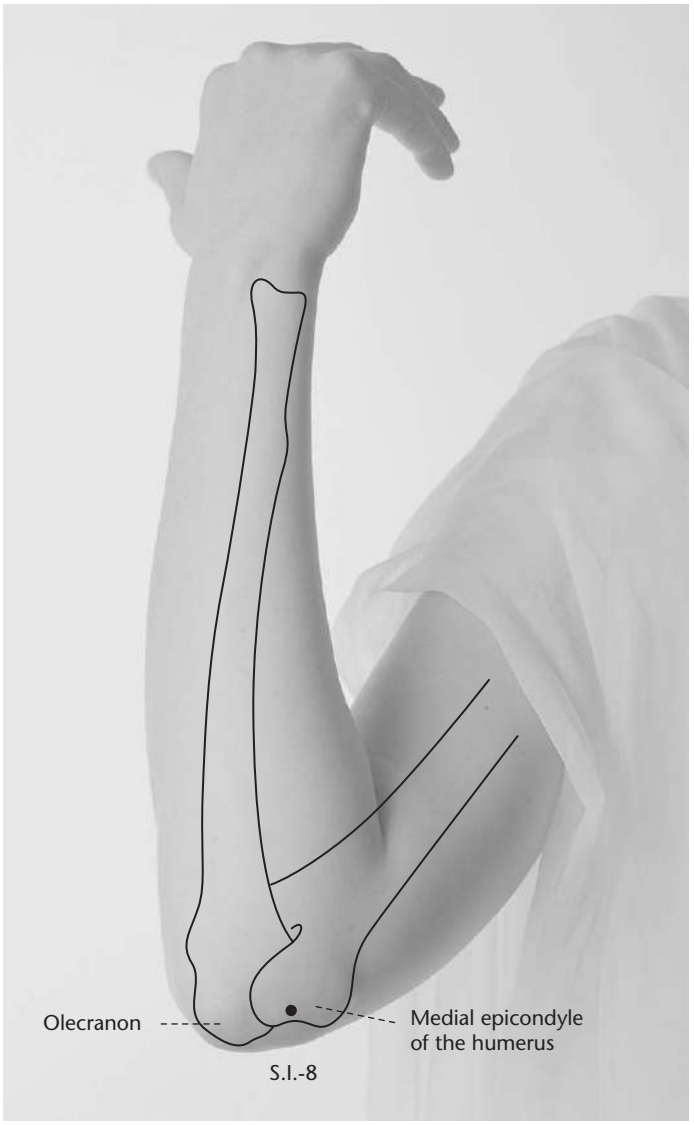

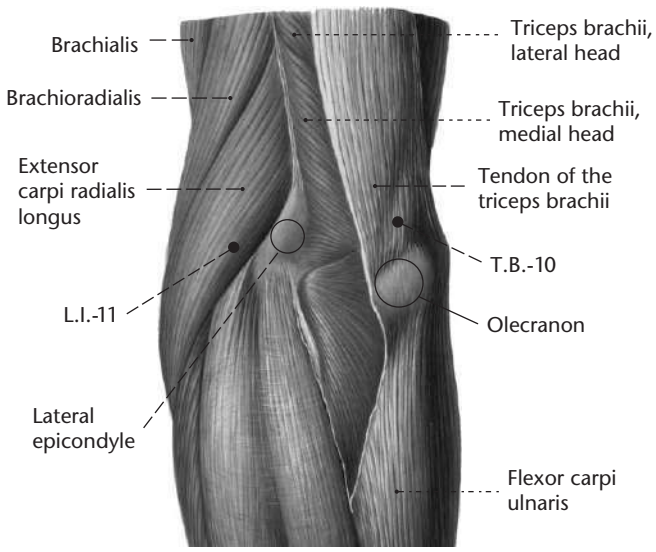

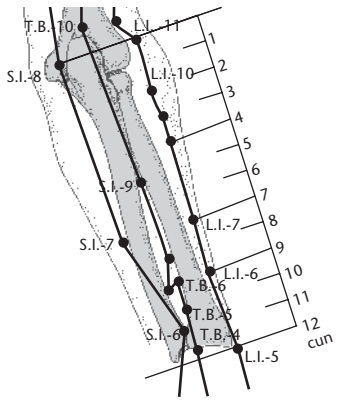

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 231

True Shoulder JIANZHEN S.I.-9

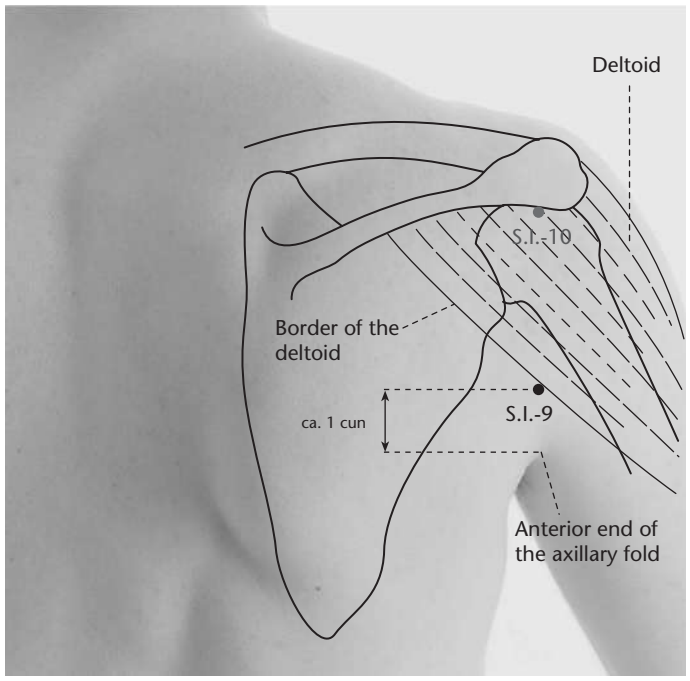

Location

With the arm adducted, 1 cun superior to the posterior axillary fold, on the lower border of the deltoid muscle.

How to find

With the patient sitting upright, palpate upward from the end of the posterior axillary fold, until you can feel the lower border of the deltoid muscle. If in doubt, ask the patient to flex their muscle. **S.I.-9** is located on its lower border.

Needling

Vertically 1–1.5 cun

Actions/Indications

- Opens the channel, expels Wind and benefits the shoulder

Special features

S.I.-9 corresponds to a trigger point that is often sensitive to pressure.

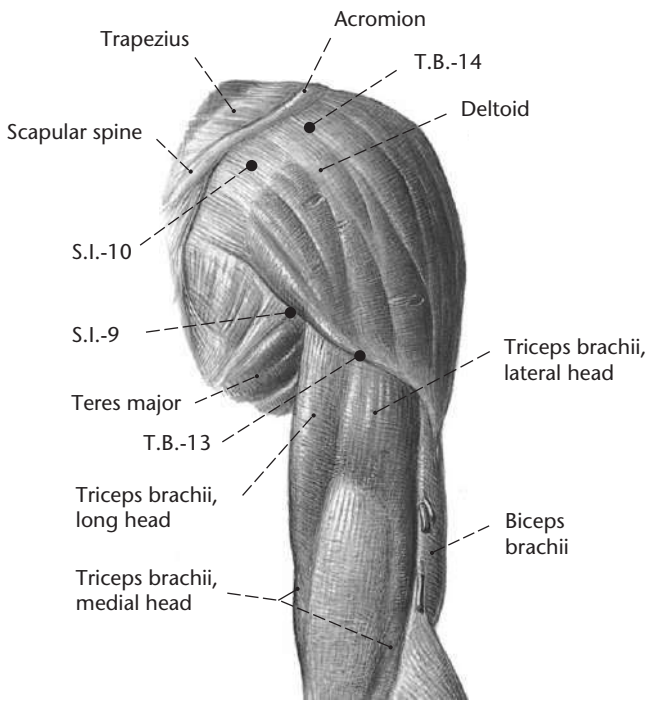

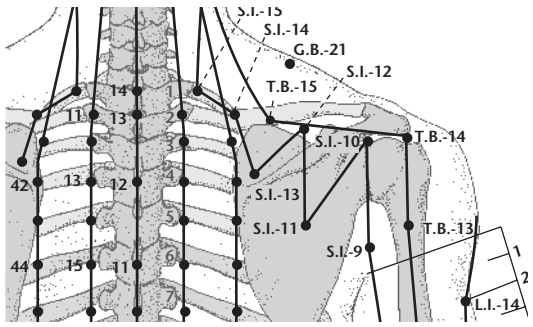

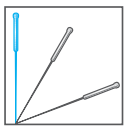
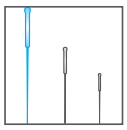

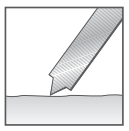
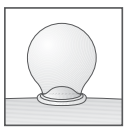

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 232

S.I.-10 *Upper Arm* Shu NAOSHU

Location

With the arm adducted, on an imaginary line extending in a superior direction from the posterior axillary fold, on the lower border of the scapular spine.

How to find

With the patient sitting upright, palpate from the posterior axillary fold upward, past the lower border of the deltoid muscle (→ **S.I.-9**) to the bony resistance of the scapular spine and acromion, which here form a curve that opens towards the inferior. **S.I.-10** is located below its bony border.

Needling

Vertically 0.5–1.5 cun

Actions/Indications

- Opens the channel and relaxes the sinews

Special features

Meeting point of the *yang wei mai* and the *yang qiao mai*. S.I.10 corresponds to a trigger point that is often sensitive to pressure. It also corresponds to the posterior access for puncturing the shoulder joint.

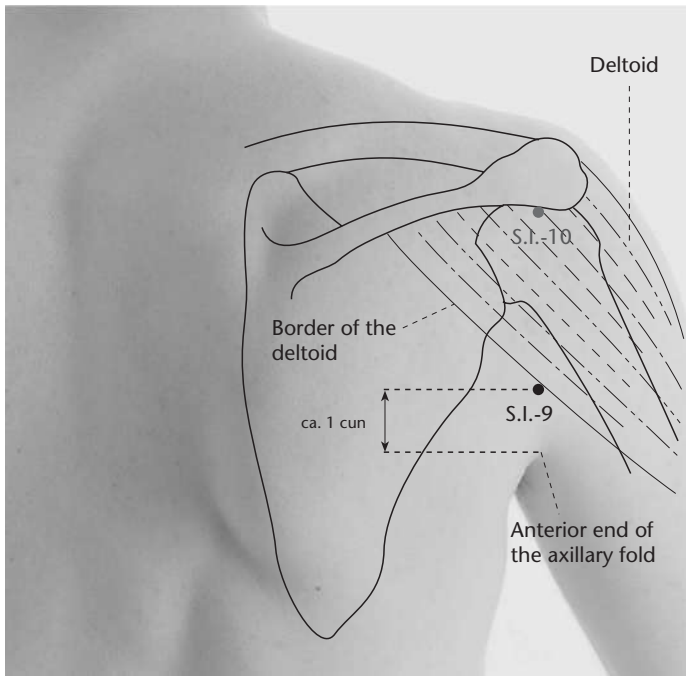

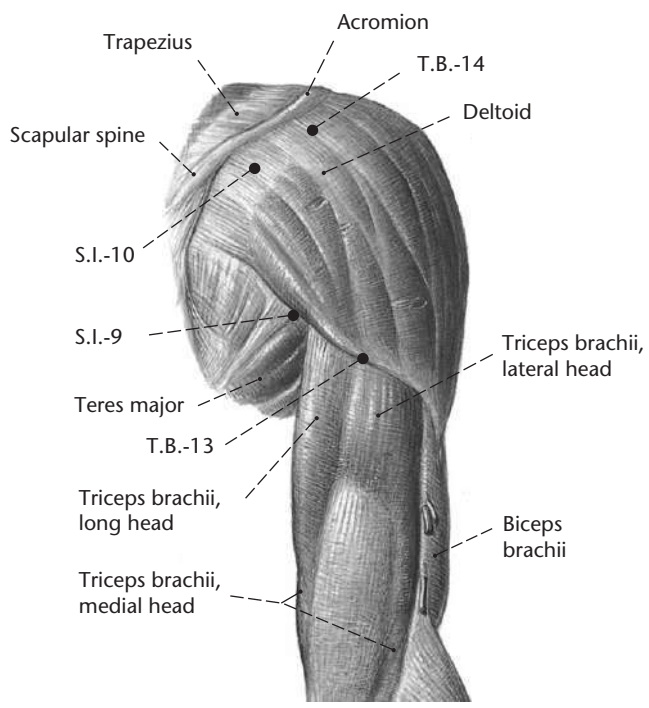

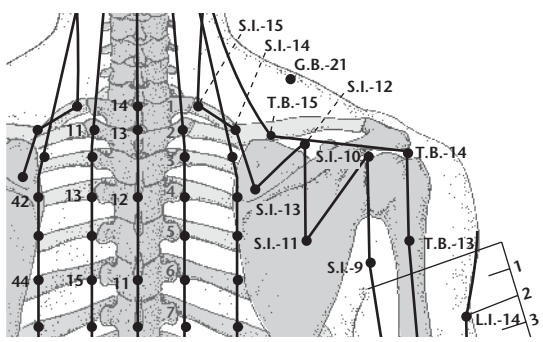

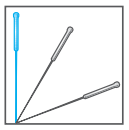
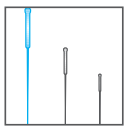

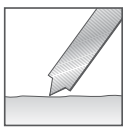

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 233

Heavenly Gathering TIANZONG S.I.-11

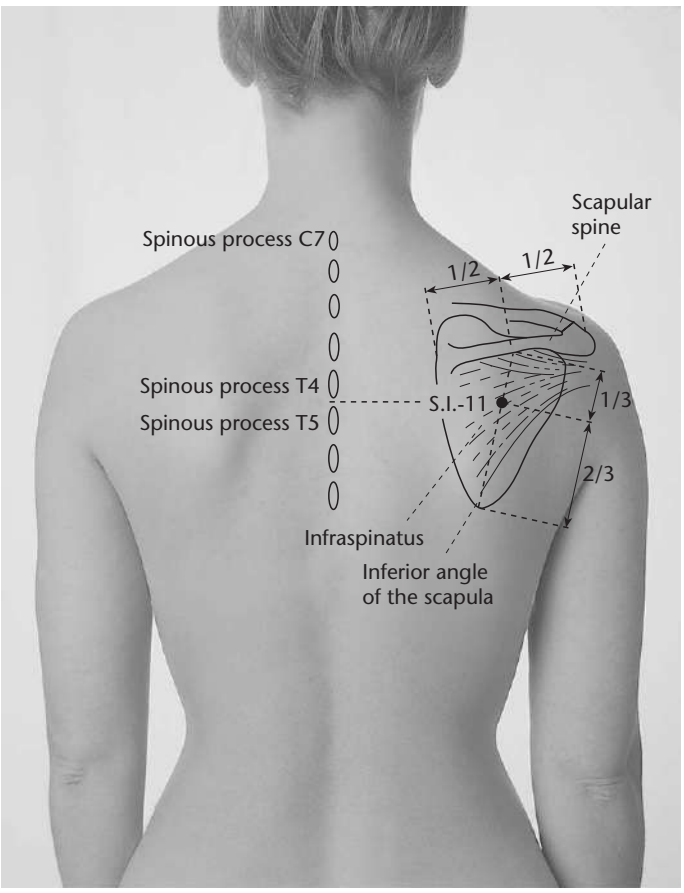

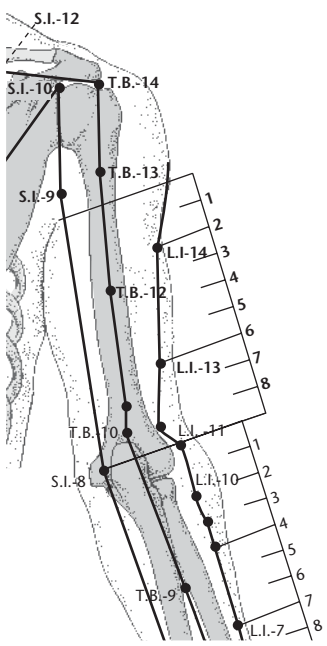

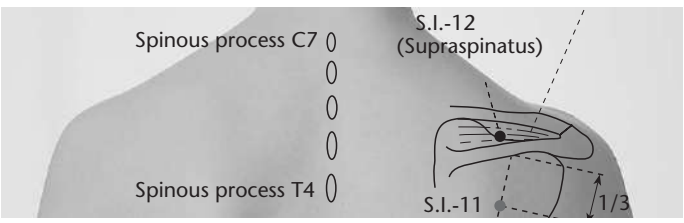

Location

On the scapula, in a depression on the infraspinatus muscle, one third of the distance from the midpoint of the scapular spine and the inferior angle of the scapula.

How to find

Patient's position: Prone position or better seated with relaxed shoulders. Draw an imaginary line from the midpoint of the scapular spine (→ 3.3.1) to the inferior angle of the scapula. **S.I.-11** is located at the junction of the upper and middle third of this line, in a depression on the infraspinatus muscle, which is often sensitive to pressure. In a seated or standing position with the arms hanging down, it will generally be level with the spinous process of the 4th thoracic vertebra or the body of the 5th thoracic vertebra.

→ **S.I.-12** is located directly above **S.I.-11**, in the centre of the supraspinous fossa.

Needling

0.5–1.5 cun vertically or obliquely into the muscle tissue

Actions/Indications

- Opens the channel, alleviates pain, moves Qi locally, unbinds the chest and lateral costal region
- Benefits the breasts

Special features

S.I.-11 is often a trigger point on the infraspinatus muscle.

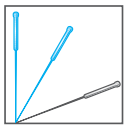
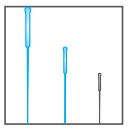

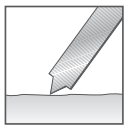
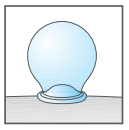

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 234

S.I.-12 *Grasping the Wind* BINGFENG

Location

Directly above → **S.I.-11**, in the centre of the supraspinous fossa.

How to find

Patient's position: Prone position or better seated with relaxed shoulders. First, locate → **S.I.-11** (one third of the distance between the midpoint of the scapular spine and the inferior angle of the scapula). **S.I.-12** is located directly above → **S.I.-11** in the centre of the supraspinous fossa.

Needling

0.5–1 cun vertically or obliquely into the supraspinatus muscle in the direction of the shoulder. Caution: Pneumothorax.

Actions/Indications

- Eliminates Wind, benefits the shoulder and scapula, opens the channel and the *luo* vessels

Special features

Meeting point with the Gall Bladder, Triple Burner and Large Intestine channels. **S.I.-12** is often a trigger point in the supraspinatus muscle.

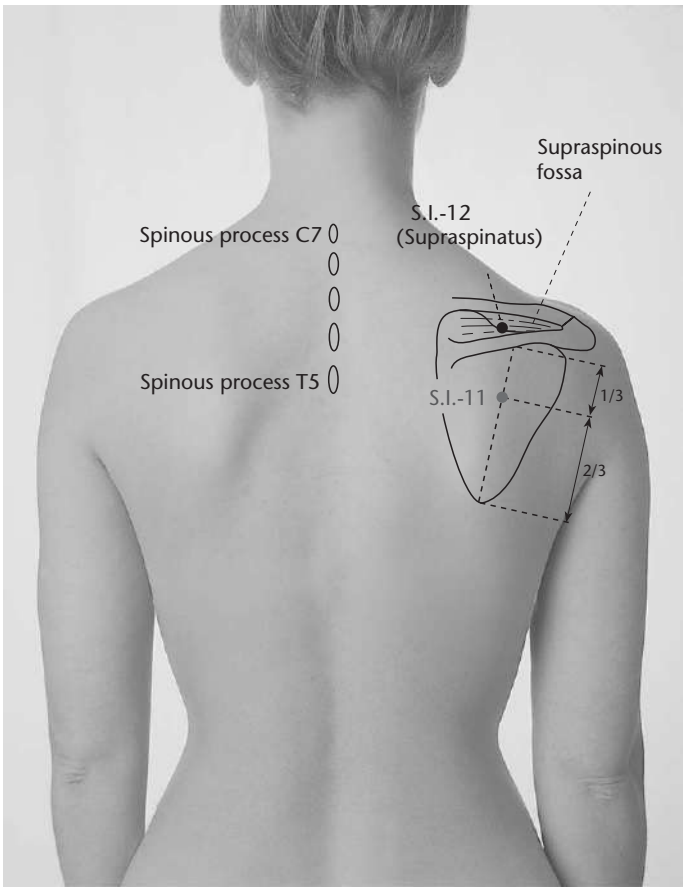

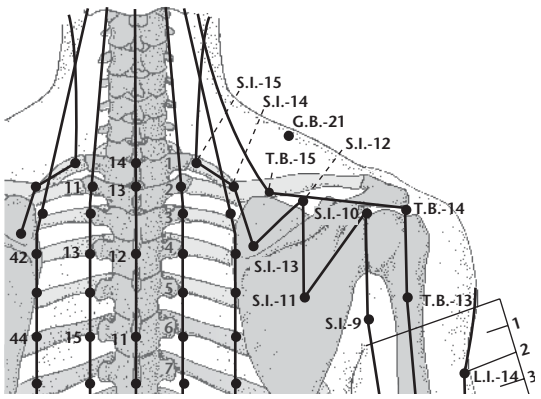

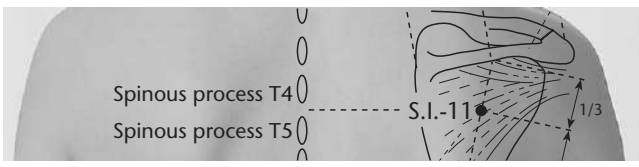

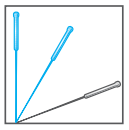

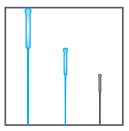

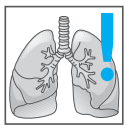

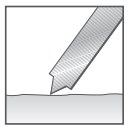

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 235

Crooked Wall QUYUAN S.I.-13

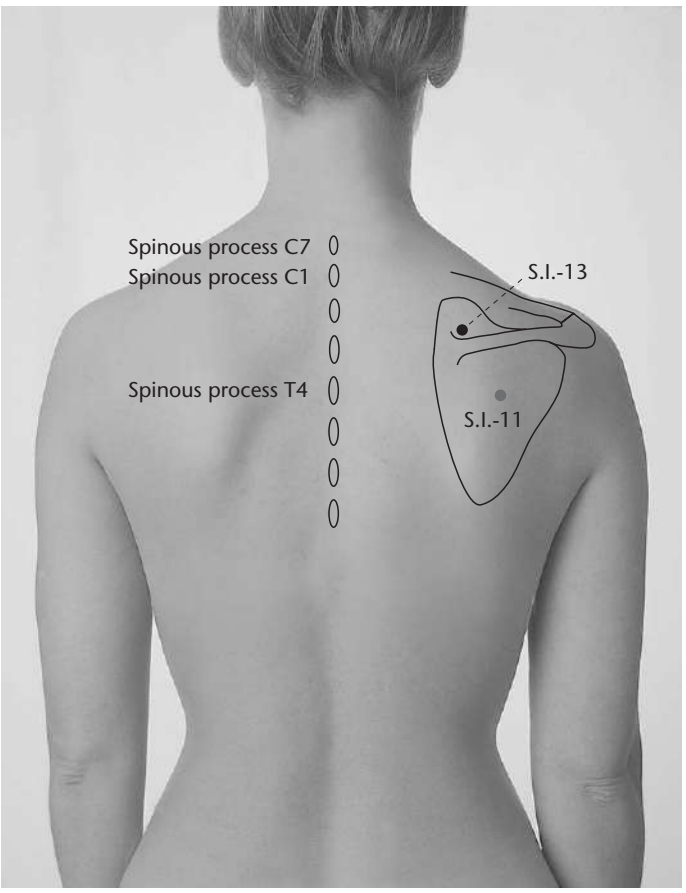

Location

At the medial end of the supraspinous fossa.

How to find

At its medial end, the scapular spine fans out in a deltoid shape, curving round in a superior direction. When palpating, this feels like a 'crooked wall', hence its name. **S.I.-13** is located on the scapula, in the slightly concave bend, at the medial origin of the supraspinatus muscle.

Needling

0.5–1 cun vertically or obliquely in a lateral direction. No danger of puncturing the pleura if located correctly.

Actions/Indications

- Opens the channel

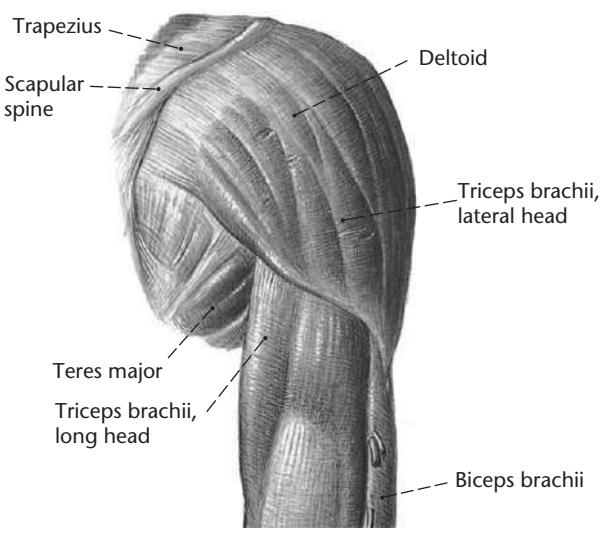

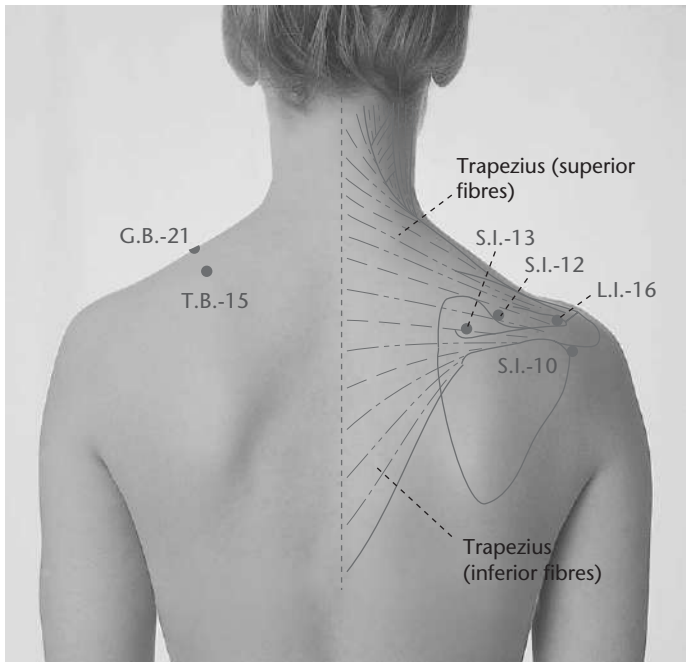

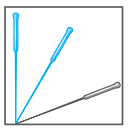
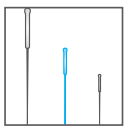

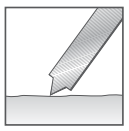

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 236

S.I.-14 *Outer Shoulder* Shu JIANWAISHU

Location

3 cun lateral to the lower border of the spinous process of the 1st thoracic vertebra, at the insertion of the levator scapulae muscle.

How to find

The levator scapulae muscle originates on the upper aspect of the medial border and the superior angle of the scapula. In this area, it is often myogelotic and therefore well defined and often sensitive to pressure (trigger point). **S.I.-14** is located just superior to the actual muscle insertion; in contrast to → **S.I.-13**, it is not located on the scapula.

Located on approximately the same level are → **Du-13** (below the spinous process of the 1st thoracic vertebra) and → **BL-11** (1.5 cun distal to the posterior midline).

Needling

0.5–1 cun vertically or obliquely in a medial direction. Caution: Pneumothorax.

Actions/Indications

- Opens the channel
- Expels Wind-Cold

Special features

Trigger point on the levator scapulae muscle

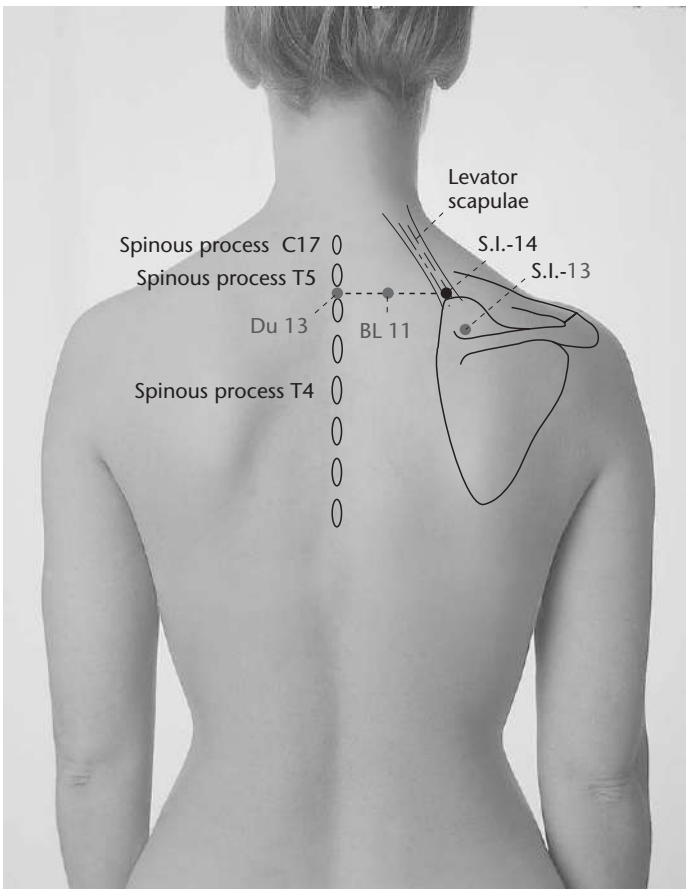

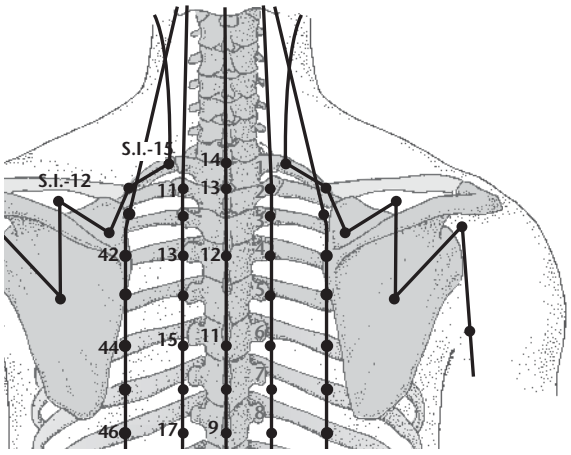

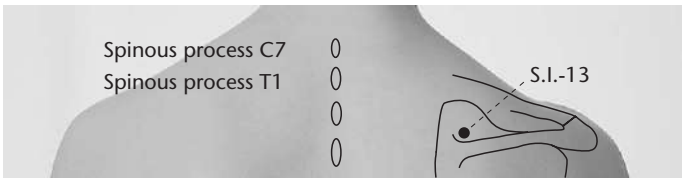

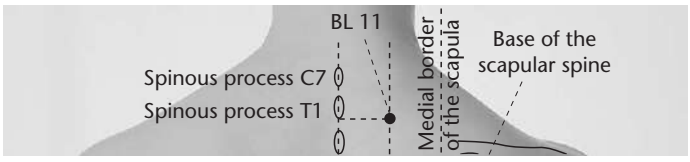

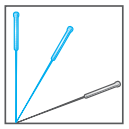

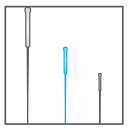

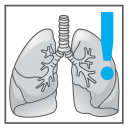

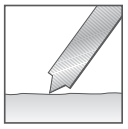

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 237

Middle Shoulder Shu JIANZHONGSHU S.I.-15

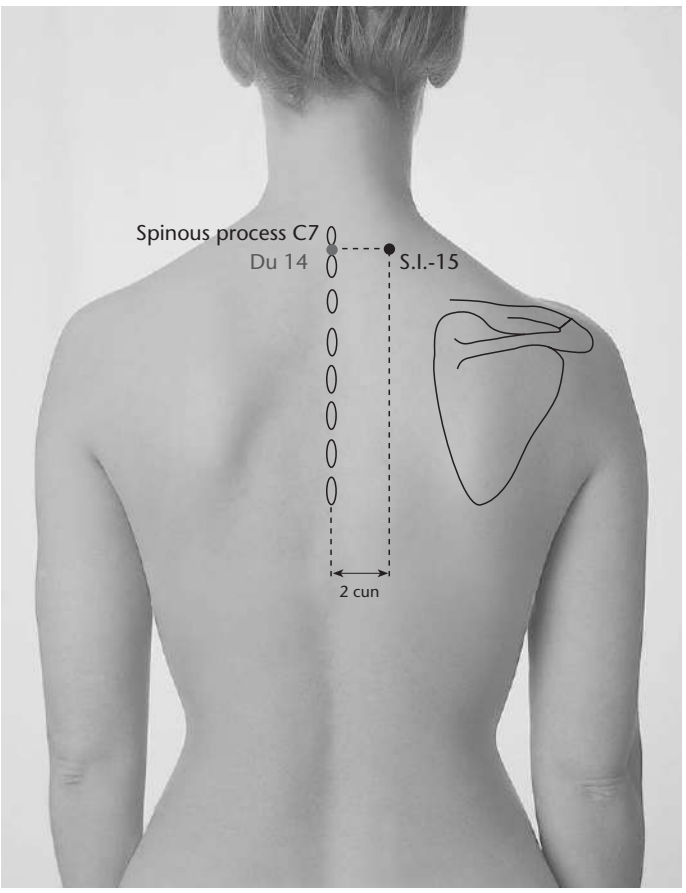

Location

2 cun lateral to the lower border of the spinous process of the 7th thoracic vertebra.

How to find

For locating the 7th cervical vertebra, see ➞ 3.4.
Located on the same level is ➞ **Du-14** below the spinous process of the 7th cervical vertebra. Locate **S.I.-15** by measuring 2 cun in a lateral direction.

Needling

0.5–1 cun vertically or obliquely in a medial direction. Caution: Pneumothorax.

Actions/Indications

- Opens the channel

Special features

Trigger point on the levator scapulae muscle, effective point for disorders of the cervicodorsal area.

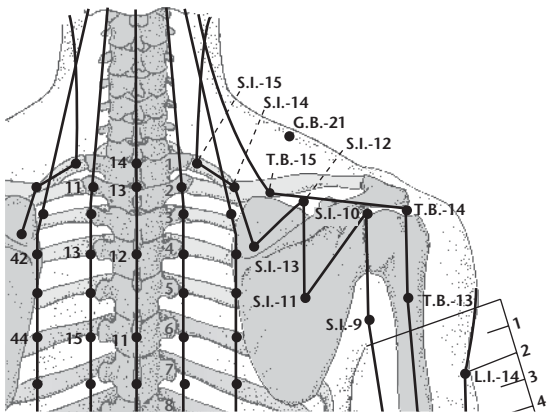

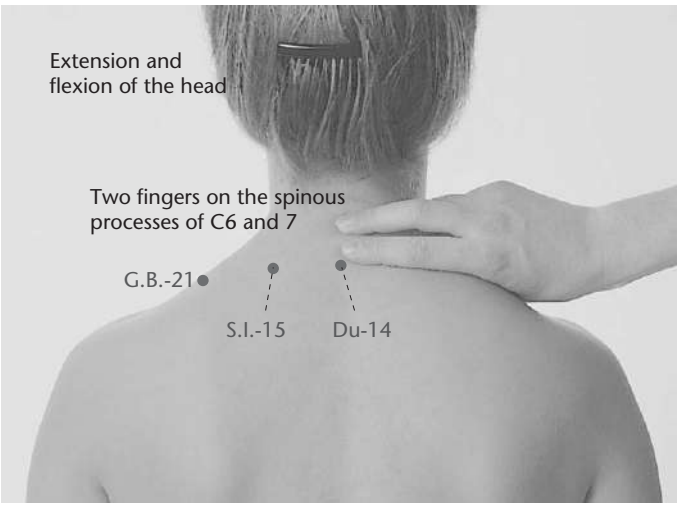

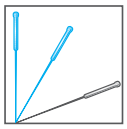
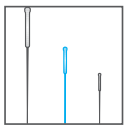
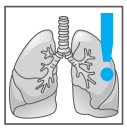

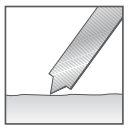

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 238

S.I.-16 *Heavenly Window* TIANCHUANG

Location

Approximately 3.5 cun lateral to the anterior midline, at the level of the laryngeal prominence, on the posterior border of the sternocleidomastoid muscle.

How to find

From the laryngeal prominence (→ 3.2), draw a line to the posterior border of the sternocleidomastoid muscle. **S.I.-16** is located directly at its posterior border, which can be made more clearly visible and palpable by rotating the patient's head in the opposite direction.

Located on the same level are → **ST-9** (on the anterior border of the sternocleidomastoid muscle) and → **L.I.-18** (between the two heads of the SCM muscle).

Needling

Vertically 0.5–1 cun. If necessary, rotate the patient's head back to its normal position before needle insertion.

Actions/Indications

- Expels Wind and Heat, benefits the throat, ears and voice
- Opens the channel

Special features

Window of Heaven point

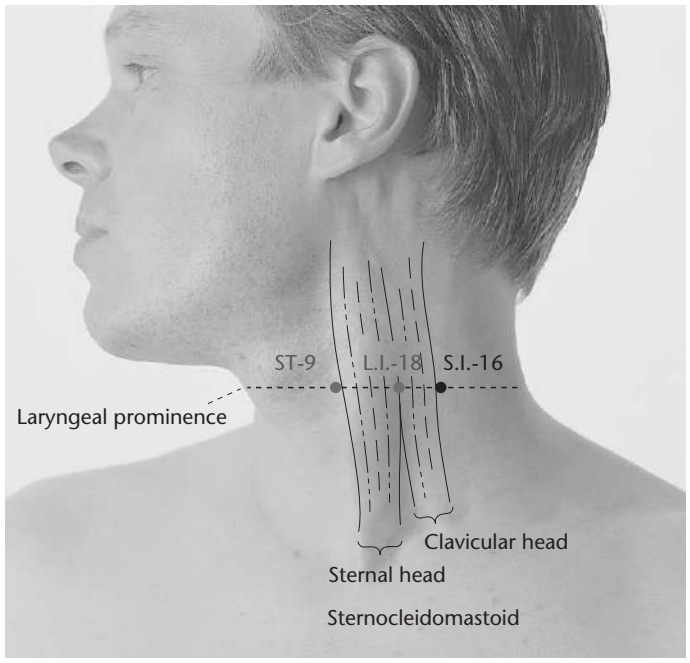

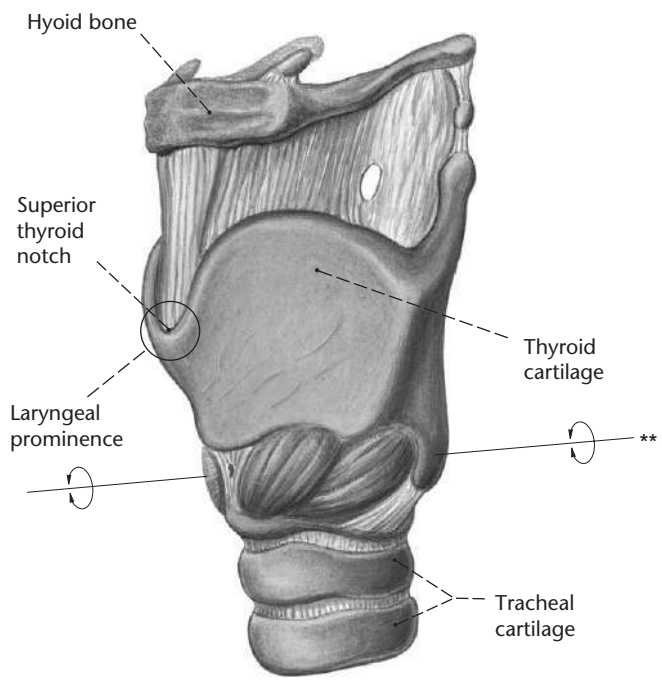

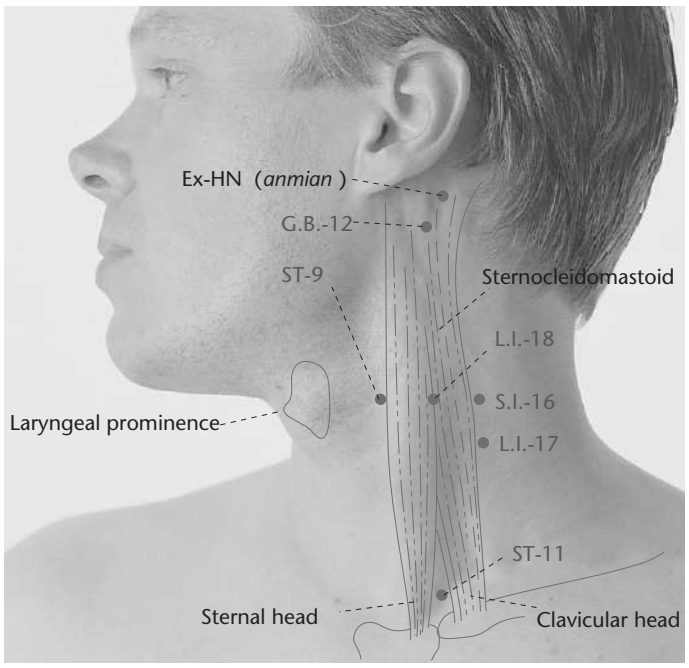

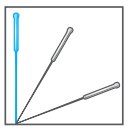
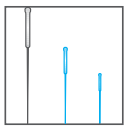

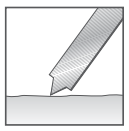

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 239

Heavenly Appearance TIANRONG S.I.-17

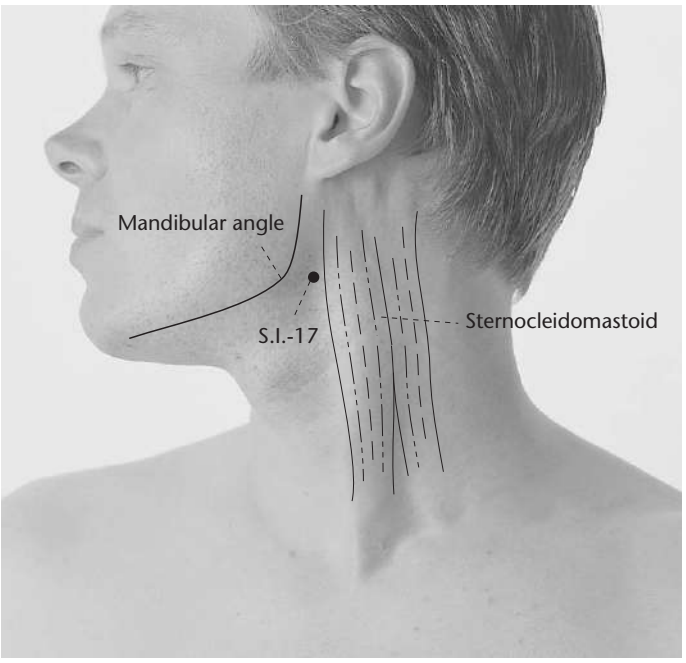

Location

Posterior to the mandibular angle, on the anterior border of the sternocleidomastoid muscle.

How to find

By rotating the patient's head in the opposite direction, the border of the sternocleidomastoid muscle becomes more pronounced. **S.I.-17** is located halfway between the angle of the mandible and the anterior border of the sternocleidomastoid muscle.

Needling

Vertically 0.5–1 cun. Caution: Internal carotid artery, internal and external jugular veins.

Actions/Indications

- Expels Wind, opens the channel, disperses accumulations, benefits the ears and throat

Special features

Window of Heaven point

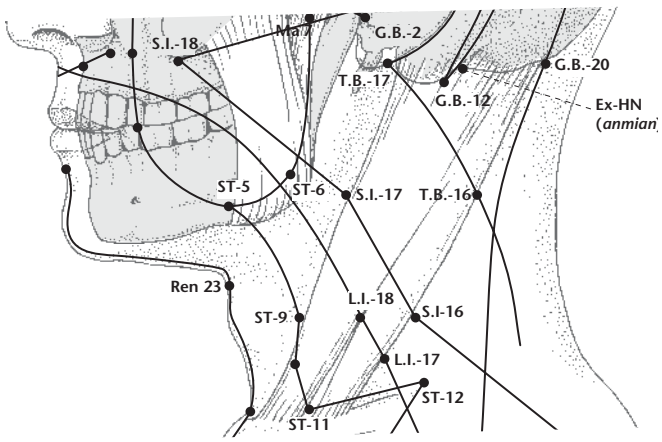

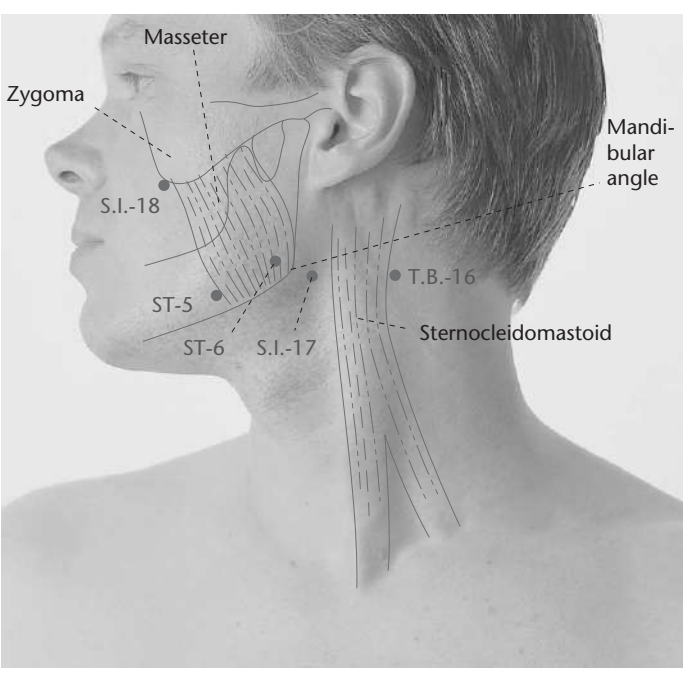

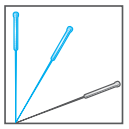
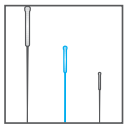
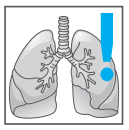

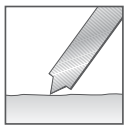

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 240

S.I.-18 *Cheekbone Crevice* QUANLIAO

Location

At the intersection of a vertical line through the outer canthus of the eye with the lower border of the zygomatic bone, on the anterior border of the masseter muscle.

How to find

Draw an imaginary vertical line from the outer canthus of the eye to the lower border of the zygomatic bone. **S.I.-18** is located in a depression on the anterior border of the masseter muscle, approximately on the level of the lateral border of the ala nasi (→ **L.I.-20**). By making chewing movements, the anterior border of the masseter muscle becomes easily palpable.

Needling

0.3–0.5 cun vertically or 1–1.5 cun transversely (subcutaneously) in the direction of → **ST-4,** → **ST-7** and → **L.I.-20**. According to some texts, moxibustion is contraindicated.

Actions/Indications

- Eliminates Wind, alleviates pain, clears Heat, reduces swellings

Special features

Meeting point with the Triple Burner channel. Important local point for disorders of the face (caused by internal or external Wind).

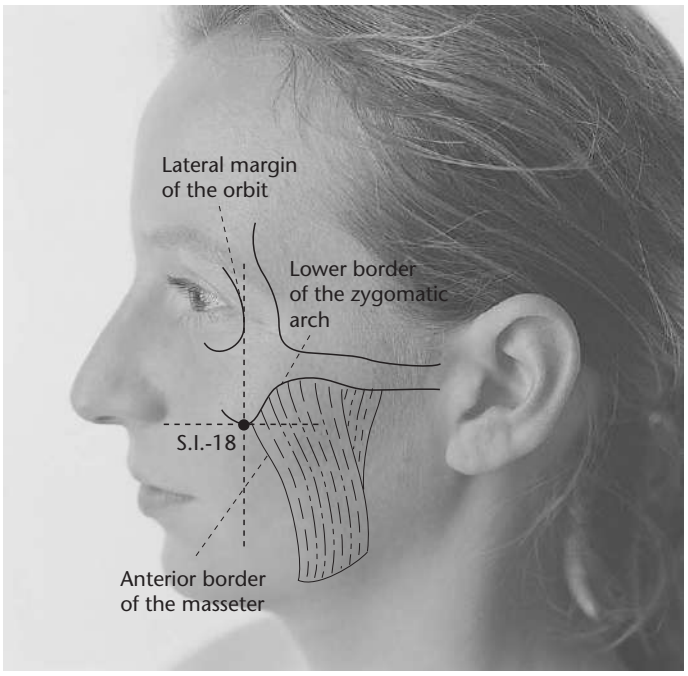

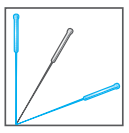

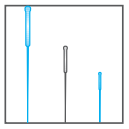

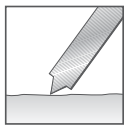

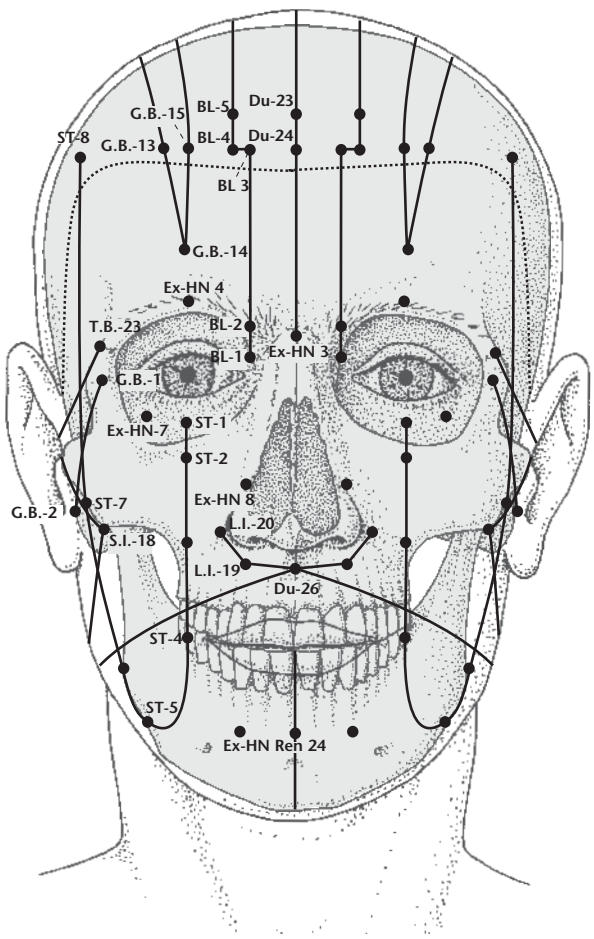

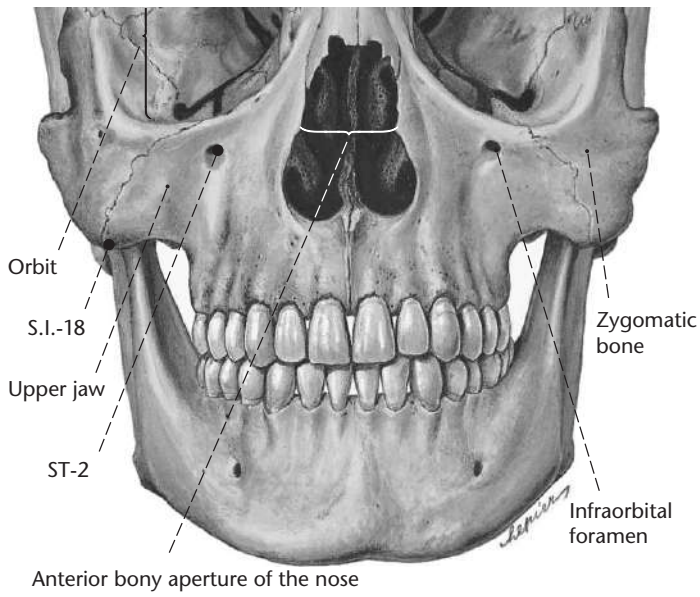

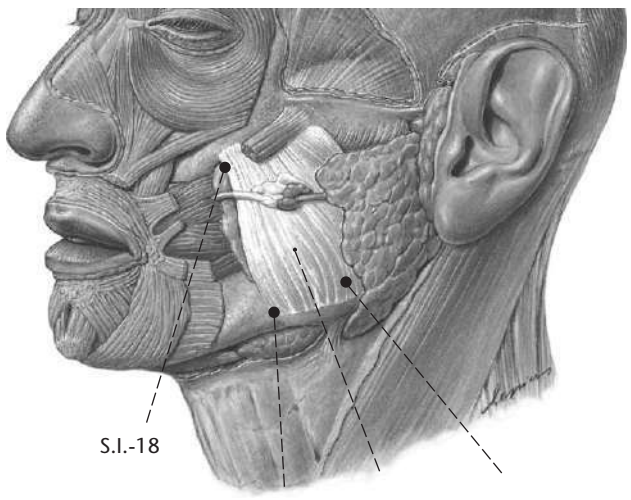

Ch04.6-F10028.qxd 2/22/08 9:48 PM Page 241

Palace of Hearing TINGGONG S.I.-19

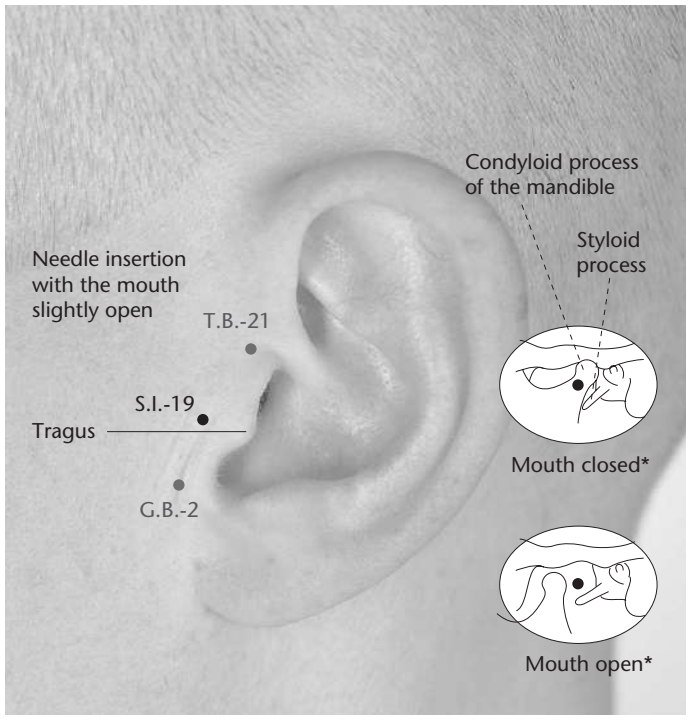

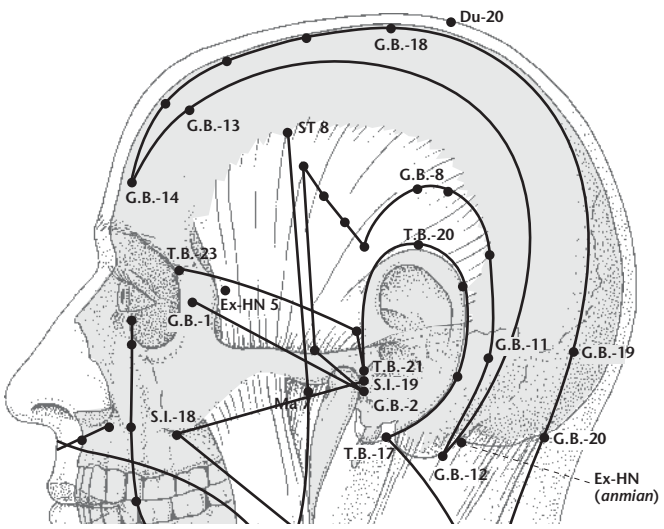

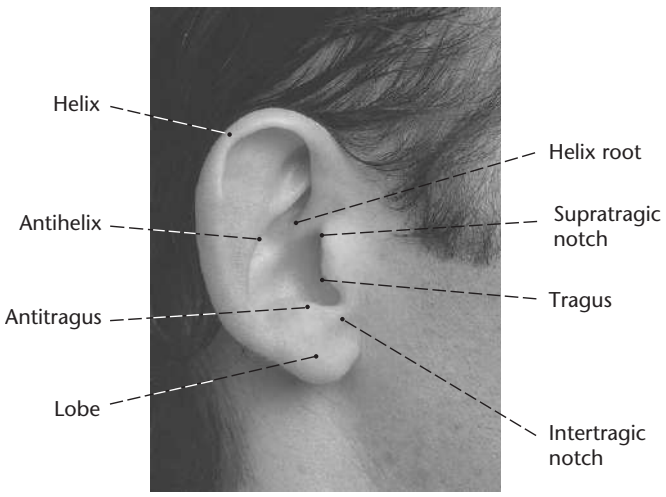

Location

Anterior to the ear, with the mouth open at the level of the middle of the tragus, in a depression between the tragus and condyloid process of the mandible.

How to find

Locate and needle this point with the mouth open. This allows the condyloid process of the mandible to slide anteriorly and reveals the depression. Locate the vertical sulcus at the junction between the ear and the cheek (this is not always very pronounced). Then locate **S.I.-19** on the level of the tragus on the sulcus. If you cannot clearly define the sulcus (it becomes more pronounced with increasing age), an (ear) point locator can be used to locate the point (the patient should have their mouth open in order to reveal the depression).

S.I.-19 is the centre point of three points located anteriorly to the ear (→ **T.B.-21** is located proximal and → **G.B.-2** is located distal to **S.I.-19**).

Needling

0.5–1 cun vertically or slightly obliquely in an inferior direction. Needle with the mouth open (to avoid intra-articular needling). Caution: This point is located close to the superficial temporal artery and the auriculotemporal nerve (as are → **T.B.-21** and → **G.B.-2**).

Actions/Indications

- Benefits the ears
- Calms the *shen*

Special features

Meeting point with the Gall Bladder and Triple Burner channels, exit point. Important local point for disorders of the ears. For temporomandibular joint dysfunction and pain, rather use → **G.B.-2** as a local point.

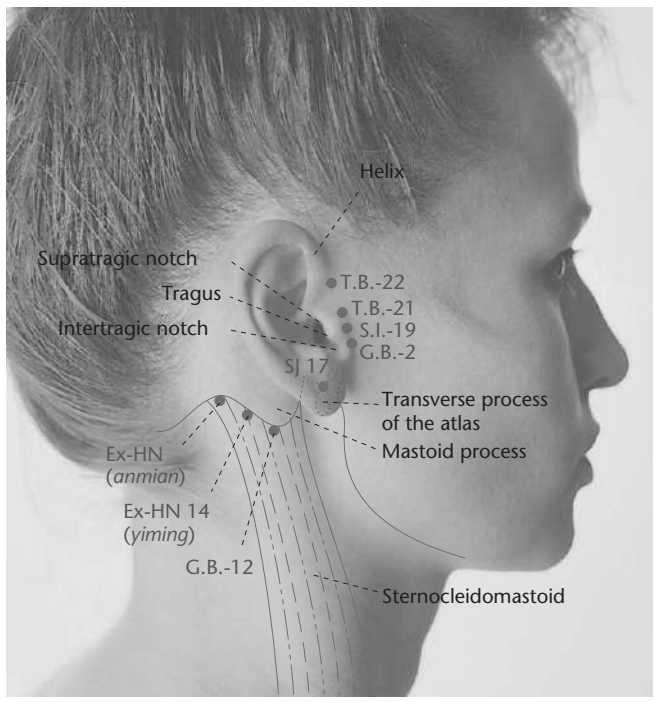

* According to Deadman et al 1998.

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 243

4.7 The Bladder Channel System – Foot *taiyang* (*zu tai yang jing luo*)

4.7.1 The Bladder Primary Channel (*zu tai yang jing*)

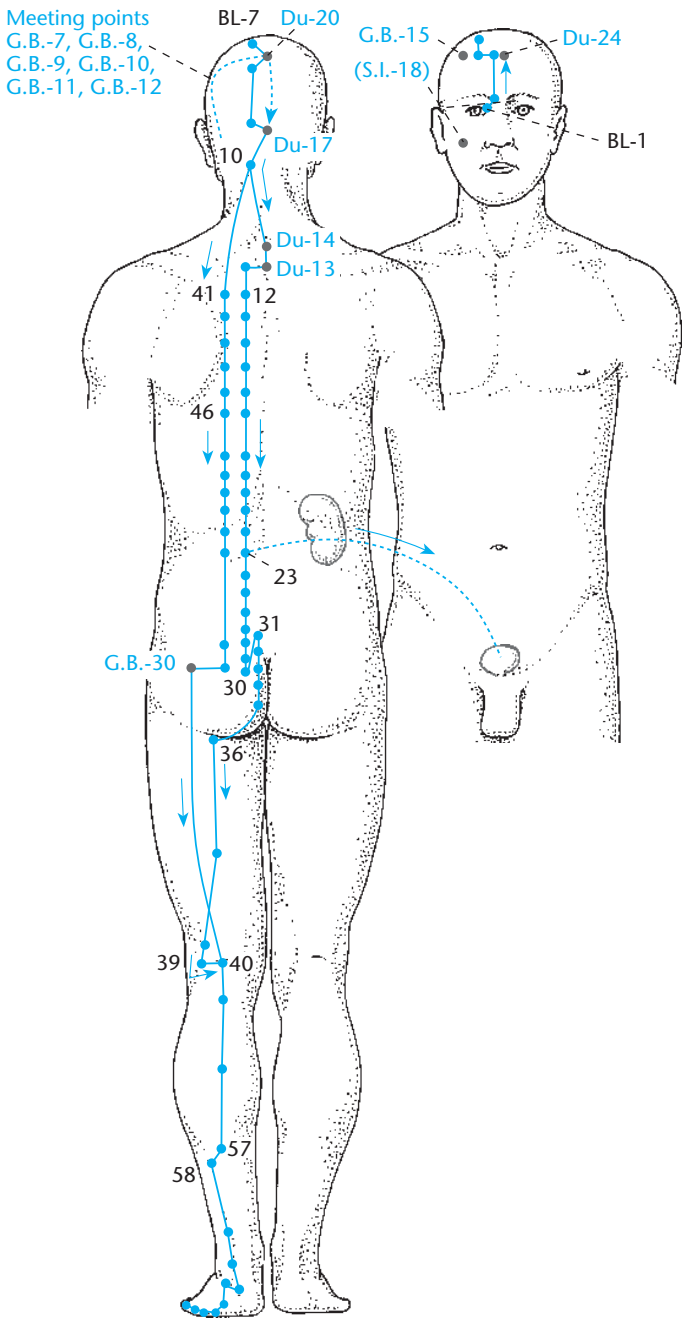

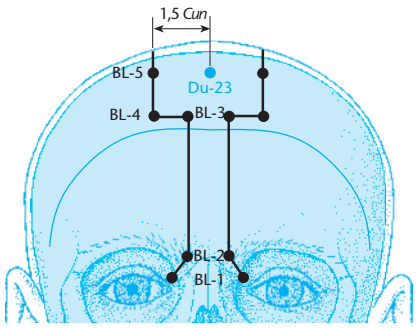

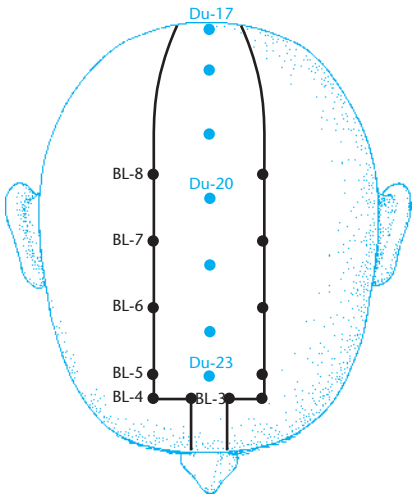

Pathway

The **external** pathway of the Bladder primary channel begins at the inner canthus of the eye at → **BL-1** (*jingming*). This point is reached by a branch that separates from the Small Intestine primary channel at → **S.I.-18** (*quanliao*) (hand–foot pairing of the second great circuit, Yang axes: *taiyang*).

The **external** pathway ascends the forehead, there intersects with → **Du-24** (*shenting*) and → **G.B.-15** (*toulinqi*), continues to the vertex and there connects with the *du mai* at → **Du-20** (*baihui*).

At → **Du-20 two branches** separate from the primary channel:

- **one branch** travels to the temporal region and intersects the Gall Bladder primary channel at **G.B.-8** (*shuaigu*), **G.B.-7** (*qubin*), **G.B.-9** (*tianchong*), **G.B.-10** (*fubai*), **G.B.-11** (*touqiaoyin*) and **G.B.-12** (*wangu*)
- the **other branch** penetrates deeper into the brain and re-emerges at → **Du-17** (*naohu*) or, according to some authors, at **BL-8** (*luoque*).

From → **Du-20**, the **external pathway** runs to **BL-8** (*luoque*) and **BL-9** (*yuzhen*), passes → **Du-17** (*naohu*) and continues to **BL-10** (*tianzhu*) in the occipital region. Here, the **external pathway** of the channel separates into **two branches**:

- the **medial** branch runs to → **Du-14** (*dazhui*) below the spinal process of C7 where it meets the other Yang primary channels, then continues to → **Du-13** (*taodao*). From T1 it runs parallel to the posterior midline, 1.5 cun lateral to it.

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 244

4 Acupuncture Points of the Twelve Primary Channels

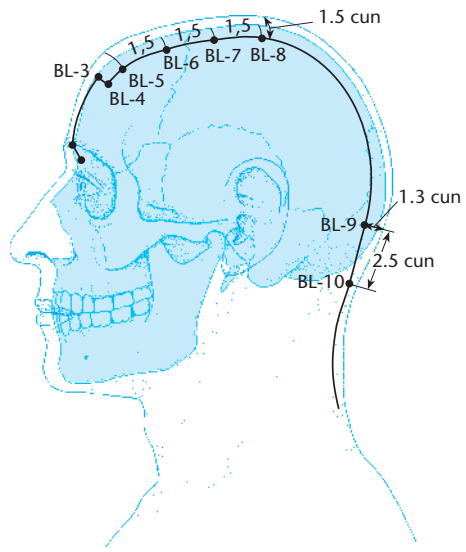

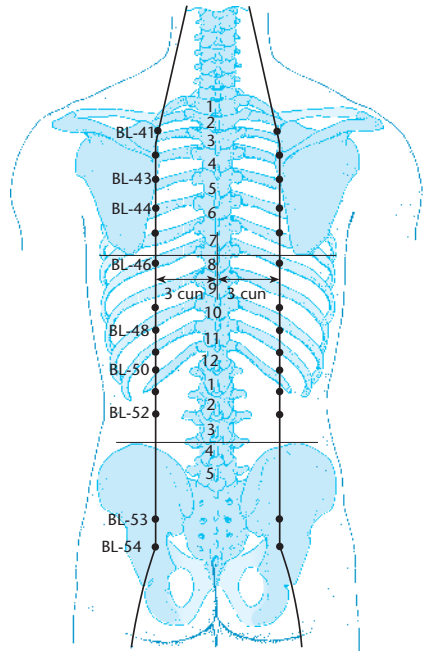

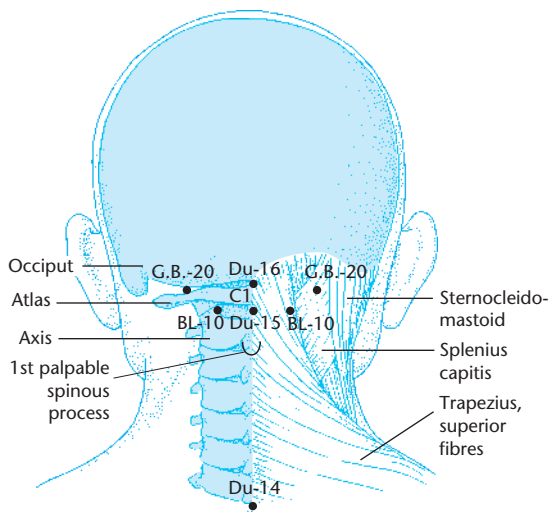

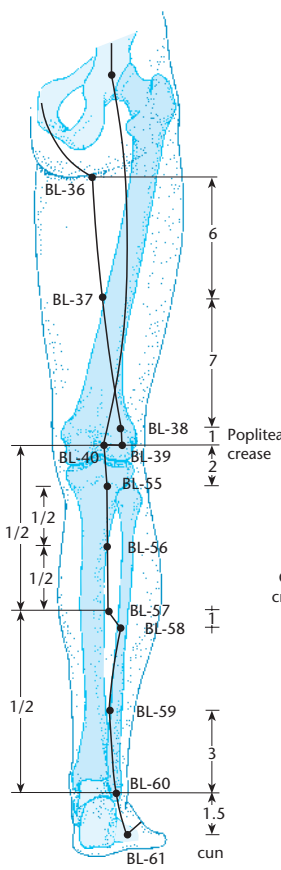

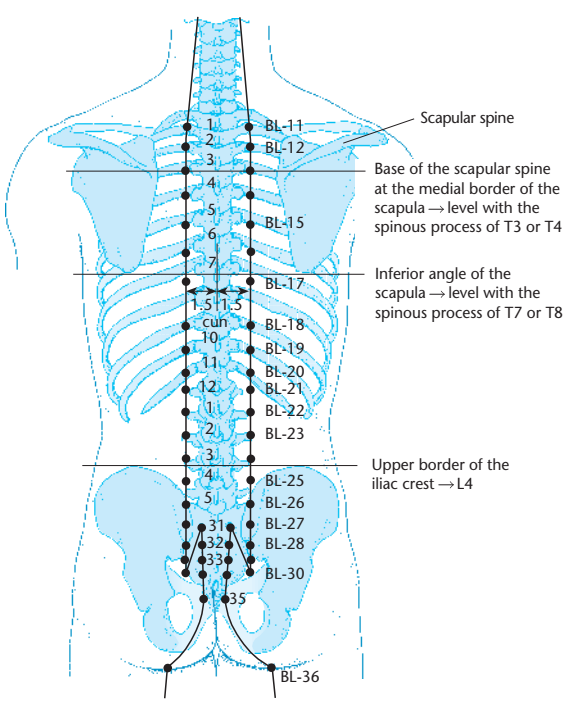

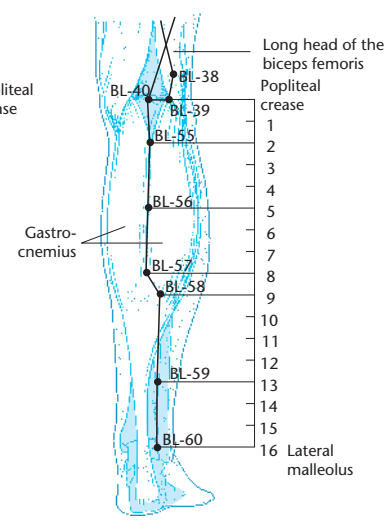

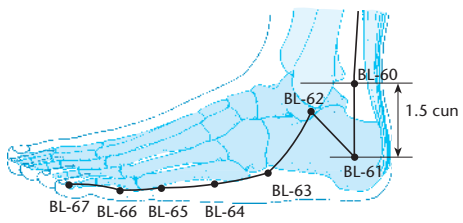

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 245

At the level of L2, it **enters** the abdomen and reaches its paired *zang*-Organ, the Kidney (*shen*) and its pertaining *fu*-Organ, the Bladder (*pangguang*). The **external** branch runs through the gluteal region and the thigh to the popliteal fossa (**BL-40**), where it re-unites with the lateral external branch.

- the **lateral branch** runs from **BL-10** (*tianzhu*) to **BL-41** (*fufen*) at the level of T2, from where it descends down the back, 3 cun lateral to the midline to the level of the fourth sacral foramen. From there it traverses the gluteal region to → **G.B.-30** (*huantiao*), continues along the posterolateral aspect of the thigh and meets the medial external branch at the popliteal fossa.

From the popliteal fossa (**BL-40**) the channel descends along the posterolateral aspect of the foreleg to the foot

- passes the lateral malleolus
- travels along the lateral border of the foot to the lateral corner of the little toenail to **BL-67** (*zhiyin*). From here, a branch travels to → **KID-1** (*yongquan*) located on the Interiorly–Exteriorly paired Kidney primary channel (foot Yin–Yang pairing of the second great circuit).

Clinical importance (→ 1.2)

Exterior (*biao*) signs and symptoms: Chills and fever, headaches, occipital stiffness, pain in the lumbar region, obstruction of the nasal passages, eye pain with tearing, pain along the posterior aspect of the hip, knee and foreleg, foot pain

Interior (*li*) or *zangfu*-Organ signs and symptoms: Pain and tension in the lower abdomen, retention of urine, enuresis, psycho-emotional disorders, opisthotonus

Connections and meeting points of the Bladder primary channel (→ 1.2)

Connections with other channels

Kidney primary channel (*zu shao yin jing*)

Connection: Foot Yin–Yang pairing of the second great circuit
Location: BL-67 → KID-1 (on the foot).
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Small Intestine primary channel (*shou tai yang jing*)

Connection: Paired according to the six channel theory (hand–foot pairing): *tai yang* (Yang axes of the second great circuit)
Location: S.I.-18 → BL-1 (on the head).
Circulation: Circadian (according to the Organ clock)
Importance: Above–below relationship

Connections with other *zangfu*-Organ systems

Kidney (*shen*), Bladder (*pangguang*)

4.7.2 The Bladder Divergent Channel (*zu tai yang jing bie*)

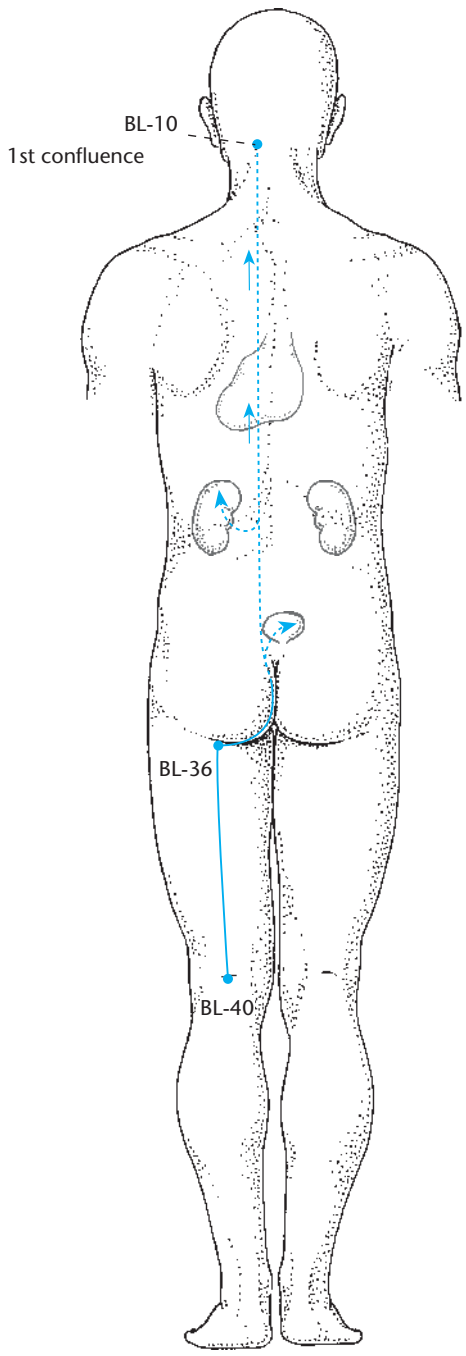

Pathway

The Bladder divergent channel separates from the Bladder primary channel in the popliteal fossa at **BL-40** (*weizhong*)

- ascends to the centre of the gluteal crease at **BL-36** (*chengfu*)
- ascends further and penetrates the anus, continues to the Bladder (*pangguang*), its pertaining *fu*-Organ, and the Kidney (*shen*), its paired *zang*-Organ
- rises along the spine and spreads in the Heart (*xin*)

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 246

➡ from here it ascends to the occiput, where it unites at **BL-10** (*tianzhu*) with the Kidney divergent channel and the Bladder primary channel to form one of the six *he*-confluences (here: KID/BL as first confluence ➞ 1.3)

Clinical importance

- Strengthens the relationship between the Bladder and the Kidney (*zangfu*-Organs). Points on the Bladder primary channel can therefore be used for disorders of the Kidney, and vice versa points on the Kidney primary channel can treat disorders of the Bladder.
- Supports the connection between the Bladder primary channel and the Heart
- Supports the connection between the Bladder primary channel and the anus
- Creates a connection between the popliteal region (**BL-40**) and the sacral region (**BL-32**, **BL-31**) as well as with the occipital region (**BL-10**).

4.7.3 The Bladder Sinew Channel (*zu tai yang jing jin*)

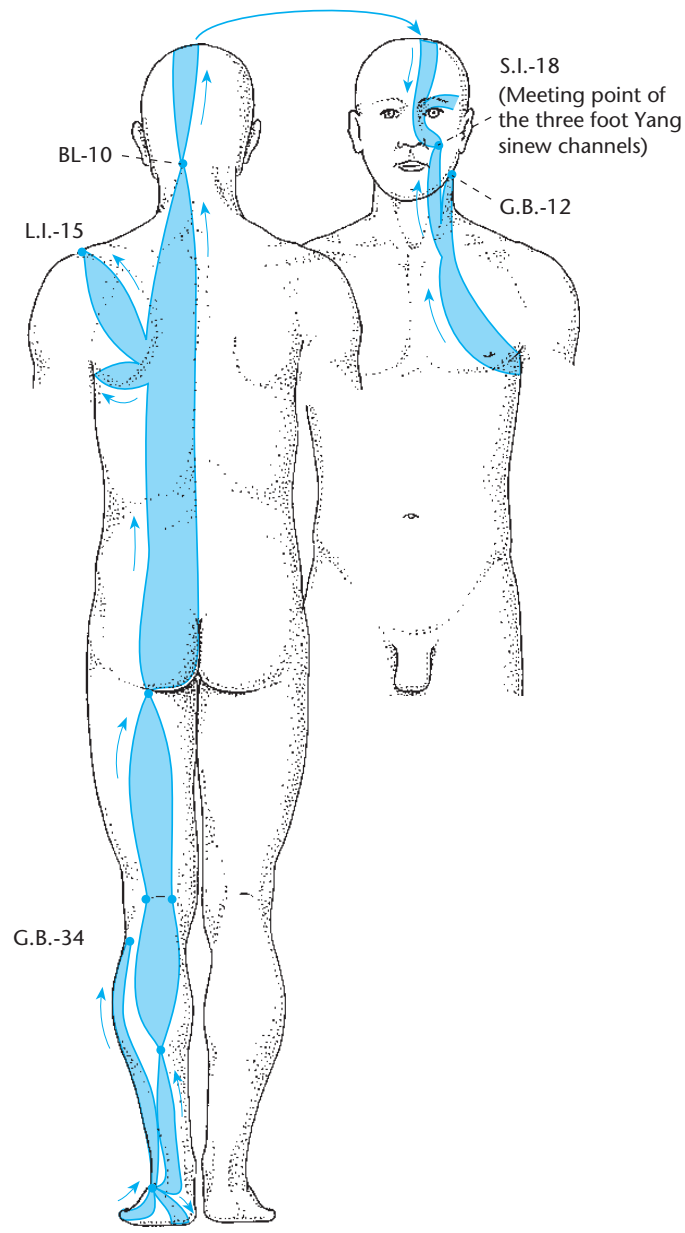

Pathway

The Bladder sinew channel begins on the lateral aspect of the little toe and travels along the lateral border of the foot to the lateral malleolus, where it binds (*jie*).

At the lateral malleolus **three branches** separate from the sinew channel:

➡ **one branch** runs to the calcaneus where it binds (*jie*) and disperses over the heel

➡ **a further branch** ascends the lateral aspect of the lower leg and binds (*jie*) at the head of the fibula, where it also meets the Gall Bladder and Stmach sinew channels at ➞ **G.B.-34** (*yanglingquan*)

➡ **the main branch** travels to the Achilles tendon, binds (*jie*) in the area of the two heads of the gastrocnemius muscle, ascends the posterior aspect of the leg along the gastrocnemius

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 247

and soleus muscles and binds (*jie*) on either side of the popliteal crease, ascends the posterior aspect of the thigh and binds (*jie*) at the centre of the gluteal crease, ascends as a broad band along the spine and binds (*jie*) at **BL-10** (*tianzhu*).

Two branches separate from the main branch in the thoracic region at the level of the axilla:

- **one branch** ascends to the shoulder and reaches → **L.I.-15** (*jianyu*)
- **the other branch** penetrates the axilla, ascends the thorax and runs to the supraclavicular fossa, where again **two sub-branches** separate:
  - **one sub-branch** ascends to the mastoid process to → **G.B.-12** (*wangu*)
  - **the other sub-branch** crosses the cheek to → **S.I.-18** (*quan-liao*) where it meets with the other foot Yang sinew channels

The **main branch** divides at **BL-10** (*tianzhu*) into **two branches:**

- **one branch** penetrates deeper to the root of the tongue
- **the other branch** crosses the head lateral to the midline, reaches the inner canthus of the eye and separates into **two branches:**
  - **one branch** binds (*jie*) lateral to the nose and descends the cheek
  - **the other branch** crosses the upper eyelid and connects with the Stomach sinew channel, together forming a network around the eye. It then penetrates deeper into the upper orbital region and ends at the roof of the mouth.

Clinical importance

Pathology: Muscular cramping in the region of the little toe. Tension and pain in the heel. Stiffness and limited range of motion of the joints. Tension and stiffness in the back and along the spine, and difficulty when bending forward. Inability to lift the arms above the shoulder region. Stiffness or distending pain in the axillary region. Pain and tense musculature in the supra-clavicular fossa. Limited range of motion of the cervical spine.

Indication: Mainly with *bi*-syndromes (painful obstruction syndrome) along the Bladder channel. The area covered by the Bladder sinew channel is larger than that covered by the Bladder primary channel. This explains why Bladder primary channel points can also be used for disorders of the scapula and the shoulder region as well as for disorders of the lower eyelid and the nasal region.

4.7.4 The Bladder *luo*-Connecting Vessel System (*zu tai yang luo mai*)

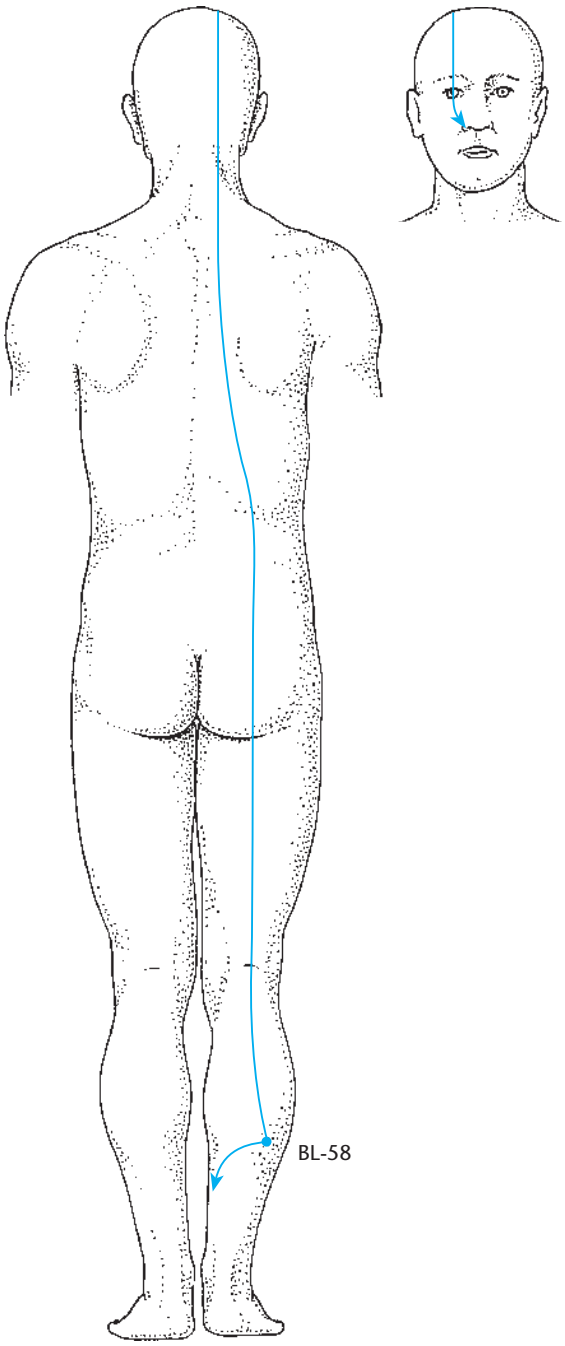

Pathway

The Bladder *luo*-connecting channel separates from the Bladder primary channel at **BL-58** (*feiyang*) (→ 8.1.2). It forms a three-dimensional reticular network, dividing into multiple branches and sub-branches (*sun luo*, *fu luo*, *xue luo* → 1.5) within the surrounding tissue.

- Horizontal divisions run to the Interiorly–Exteriorly paired Kidney primary channel; according to some schools of thought (for example Nguyen Van Nghi → Appendix) they travel as a **transverse** BL *luo*-connecting vessel to the *yuan*-source point → **KID-3** (*taixi*).

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 248

- Solinas et al. (1998, variant → fig.) describe a **longitudinal** division following the pathway of the Bladder primary channel, reaching the head and penetrating the nose. While this division is not mentioned in the classics, according to Solinas et al it becomes evident based on the classic descriptions of the pathologies of the *luo*-connecting vessels.

Clinical importance (→ 8.1.2)

Pathology

- **Excess (*shi*):** Obstruction of the nasal passages, clear watery nasal discharge, headaches, back pain
- **Deficiency (*xu*):** Nosebleeds, chronic clear nasal discharge, sinusitis

4.7.5 Cutaneous Region (*tai yang pi bu*)

See description and figures → 1.6.

4.7.6 Points of the Bladder Primary Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): BL-64** (*jinggu*)
- ***Luo*-connecting point (→ 8.1.2): BL-58** (*feiyang*)
- ***Xi*-cleft point (→ 8.1.3): BL-63** (*jinmen*)
- ***Xi*-cleft point of the *yang qiao mai*: BL-59** (*fuyang*)
- **Associated Back-*shu* point (→ 8.1.4): BL-28** (*pangguangshu*) ■■
  – Back-*shu* point of the Lung: **BL-13** (*feishu*) ■■
  – Back-*shu* point of the Pericardium: **BL-14** (*jueyinshu*)
  – Back-*shu* point of the Heart: **BL-15** (*xinshu*) ■■
  – Back-*shu* point of the *du mai*: **BL-16** (*dushu*)
  – Back-*shu* point of the diaphragm: **BL-17** (*geshu*) ■■
  – Back-*shu* point of the Liver: **BL-18** (*ganshu*) ■■
  – Back-*shu* point of the Gall Bladder: **BL-19** (*danshu*) ■■
  – Back-*shu* point of the Spleen: **BL-20** (*pishu*) ■■
  – Back-*shu* point of the Stomach: **BL-21** (*weishu*) ■■
  – Back-*shu* point of the Triple Burner: **BL-22** (*sanjiaoshu*) ■■
  – Back-*shu* point of the Kidneys: **BL-23** (*shenshu*) ■■
  – Back-*shu* point of the Large Intestine: **BL-25** (*dachangshu*) ■■
  – Back-*shu* point of the Small Intestine: **BL-27** (*xiaochangshu*)
  – Back-*shu* point of the Bladder: **BL-28** (*pangguangshu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): Ren-3** (*zhongji*) ■■
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Metal), tonification point: **BL-67** (*zhiyin*) ■■
  ying-spring point (Water), *ben* point (Five Phases): **BL-66** (*zutonggu*)
  shu-stream point (Wood), sedation point: **BL-65** (*shugu*)
  jing-river point (Fire): **BL-60** (*kunlun*) ■■
  he-sea point (Earth): **BL-40** (*weizhong*) ■■
- ***Hui*-meeting point (→ 8.1.7):**
  – of the Blood: **BL-17** (*geshu*) ■■
  – of the Bones: **BL-11** (*dazhu*) ■■
- **Opening point (→ 8.1.8) of the *yang qiao mai*: BL-62** (*shenmai*) ■■
- **Lower *he*-sea points (→ 8.1.9):**
  of the Triple Burner: **BL-39** (*weiyang*) ■■
  of the Bladder: **BL-40** (*weizhong*) ■■
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the ST, T.B.* channels, the *yin qiao mai*, *yang qiao mai*, S.I., G.B.* channels, *du mai*: **BL-1** (*jingming*) ■■
  – with the S.I. channel, (T.B., G.B. channels, *du mai**): **BL-11** (*dazhu*) ■■
  – with the *du mai*: **BL-12** (*fengmen*) ■■
  – with the *du mai**: **BL-23** (*shenshu*) ■■
  – with the G.B. channel: **BL-31–BL-34** ■■
  – with the S.I. channel: **BL-41** (*fufen*)
  – with the *yang qiao mai*: (**BL-59** (*fuyang*)*), **BL-61** (*pucan*), **BL-62*** (*shenmai*) ■■
  – with the *yang wei mai*: **BL-63** (*jinmen*)
  – of other channels with the Bladder channel: **Du-24, G.B.-15, Du-20, (G.B.-6*), G.B.-7, G.B.-8, G.B.-9, G.B.-10, G.B.-11, G.B.-12, Du-17, Du-16*, Du-14, Du-13, G.B.-30**
- **Gao Wu command point (→ 8.1.11) for the Back and lumbar region: BL-40** (*weizhong*) ■■
- **Window of Heaven point (→ 8.1.12): BL-10** (*tianzhu*) ■■
- **Points of the Four Seas (→ 8.1.13):**
  – Sea of Blood point: **BL-11** (*dazhu*) ■■
- **Ma Dan Yang Heavenly Star points (→ 8.1.14): BL-40** (*weizhong*) ■■, **BL-57** (*chengshan*) ■■, **BL-60** (*kunlun*) ■■
- **Sun Si Miao Ghost point (→ 8.1.15): BL-62** (*shenmai*) ■■
- **Other functional points:**
  – *xi*-cleft point of the *yang qiao mai*: **BL-59** (*fuyang*)

Points according to region

- **Local points (→ 8.2.1):** occiput – **BL-10** (*tianzhu*) ■■; eyes – **BL-1** (*jingming*) ■■; Stomach/Spleen – **BL-20** (*pishu*) ■■, **BL-21** (*weishu*) ■■; Kidney – **BL-23** (*shenshu*) ■■, **BL-62** (*shenmai*) ■■; rectum – **BL-35** (*huiyang*); cervical spine – **BL-10** (*tianzhu*) ■■; thoracic spine – **BL-17** – **BL-23**; lumbar spine – **BL-23** – **BL-26**, **BL-52** (*zhishi*); iliosacral region – **BL-27** (*xiaochangshu*), **BL-28** (*pangguangshu*); hip – **BL-49** (*yishe*), **BL-50** (*weicang*), **BL-54** (*zhibian*); knee – **BL-40** (*weizhong*) ■■; foot – **BL-60** (*kunlun*) ■■, **BL-62** (*shenmai*) ■■
- **Adjacent points (→ 8.2.1):** eyes – **BL-2** (*zanzhu*) ■■; nose – **BL-7** (*tongtian*); Lung – **BL-13** (*feishu*) ■■, **BL-43** (*gaohuang*); Heart – **BL-14** (*jueyinshu*), **BL-15** (*xinshu*) ■■; Liver – **BL-18** (*ganshu*) ■■, **BL-20** (*pishu*) ■■; Gall Bladder – **BL-19** (*danshu*) ■■; Large Intestine – **BL-25** (*dachangshu*) ■■; Small Intestine – **BL-27** (*xiaochangshu*) ■■; rectum – **BL-30** (*baihuanshu*), **BL-34** (*xialiao*) ■■; Bladder – **BL-28** (*pangguangshu*) ■■, **BL-23** (*shenshu*) ■■; urogenital region – **BL-23** (*shenshu*) ■■, **BL-32** (*ciliao*); cervical spine – **BL-11** (*dazhu*) ■■;

* Mentioned by only some authors

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 249

iliosacral region – **BL-23** (*shenshu*) ■■; thoracic spine – **BL-38** – **BL-47**; toes – **BL-60** (*kunlun*) ■■, **BL-**65 (*shugu*)

- **Distal points (→ 8.2.1):** occipital region – **BL-65** (*shugu*), **BL-60** (*kunlun*); rectum – **BL-57** (*chengshan*) ■■; thoracic spine – **BL-57** (*chengshan*) ■■; cervical, thoracic, lumbar spine – **BL-60** (*kunlun*) ■■; shoulder and iliosacral region – **BL-58** (*feiyang*); for the thoracic spine, lumbar spine and iliosacral region – **BL-40** (*weizhong*) ■■; lumbar spine – **BL-40** (*weizhong*) ■■, **BL-59** (*fuyang*), **BL-60** (*kunlun*) ■■; lumbar spine and hip – **BL-62** (*shenmai*) ■■

Specific points according to the channel pathway (in numerical order):

- **BL-1** (*jingming*) ■■: *jiaohui*-meeting point with the ST, T.B.* channels, the *yin qiao mai, yang qiao mai*, S.I., G.B.* channels, *du mai* (→ 8.1.10): local point for the eyes (→ 8.2.1)
- **BL-2** (*zanzhu*) ■■: adjacent point for the eyes (→ 8.2.1)
- **BL-7** (*tongtian*): adjacent point for the nose (→ 8.2.1)
- **BL-10** (*tianzhu*) ■■: Window of Heaven point (→ 8.1.12); important point for eliminating Wind; local point for the occiput and the cervical spine (→ 8.2.1)
- **BL-11** (*dazhu*) ■■: *hui*-meeting point (→ 8.1.7) of the bones; Sea of Blood point (→ 8.1.13); *jiaohui*-meeting point with the S.I. channel (T.B., G.B. channels, *du mai**) (→ 8.1.10); adjacent point for the cervical spine (→ 8.1.10)
- **BL-12** (*fengmen*) ■■: *jiaohui*-meeting point with the *du mai* (→ 8.1.10)
- **BL-13** (*feishu*) ■■: Back-*shu* point of the Lung (→ 8.1.4); adjacent point for the Lung (→ 8.2.1)
- **BL-14** (*jueyinshu*): Back-*shu* point of the Pericardium (→ 8.1.4); adjacent point for the Heart (→ 8.2.1)
- **BL-15** (*xinshu*) ■■: Back-*shu* point of the Heart (→ 8.1.4); adjacent point for the Heart (→ 8.2.1)
- **BL-16** (*dushu*): Back-*shu* point of the *du mai* (→ 8.1.4);
- **BL-17** (*geshu*) ■■: Back-*shu* point of the diaphragm (→ 8.1.4); *hui*-meeting point of the Blood (→ 8.1.7)
- **BL-17 to BL-23**: local points for the thoracic spine (→ 8.2.1)
- **BL-18** (*ganshu*) ■■: Back-*shu* point of the Liver (→ 8.1.4); adjacent point for the Liver (→ 8.2.1)
- **BL-19** (*danshu*) ■■: Back-*shu* point of the Gall Bladder (→ 8.1.4); adjacent point for the Gall Bladder (→ 8.2.1)
- **BL-20** (*pishu*) ■■: Back-*shu* point of the Spleen (→ 8.1.4); adjacent point for the Liver (→ 8.2.1)
- **BL-21** (*weishu*) ■■: local point for the Stomach (→ 8.2.1); Back-*shu* point of the Stomach (→ 8.1.4)
- **BL-22** (*sanjiaoshu*) ■■: Back-*shu* point of the Triple Burner (→ 8.1.4)
- **BL-23** (*shenshu*) ■■: *jiaohui*-meeting point with the *du mai** (→ 8.1.10); local point for the Kidneys and lumbar region (→ 8.2.1); adjacent point for the Bladder, urogenital region and iliosacral region (→ 8.2.1); Back-*shu* point of the Kidney (→ 8.1.4)
- **BL-24** (*qihaishu*): local point for the lumbar spine (→ 8.2.1)
- **BL-25** (*dachangshu*) ■■: Back-*shu* point of the Large Intestine (→ 8.1.4); adjacent point for the lumbar spine (→ 8.2.1)
- **BL-26** (*guanyuanshu*): local point for the lumbar spine (→ 8.2.1)
- **BL-27** (*xiaochangshu*) ■■: Back-*shu* point of the Small Intestine (→ 8.1.4); adjacent point for the Small Intestine (→ 8.2.1); local point for the iliosacral region (→ 8.2.1)
- **BL-28** (*pangguangshu*) ■■: Back-*shu* point of the Bladder (→ 8.1.4); adjacent point for the Bladder (→ 8.2.1); local point for the iliosacral region (→ 8.2.1)
- **BL-30** (*baihuanshu*): regional point for the rectum (→ 8.2.1)
- **BL-31** – **BL-34**: *jiaohui*-meeting points with the G.B. channel (→ 8.1.10)
- **BL-32** (*ciliao*): regional point for the urogenital region (→ 8.2.1); local point for the iliosacral region (→ 8.2.1)
- **BL-34** (*xiaoliao*): regional point for the rectum (→ 8.2.1)
- **BL-35** (*huiyang*): local point for the rectum (→ 8.2.1)
- **BL-38** – **BL-47**: adjacent points for the thoracic spine (→ 8.2.1)
- **BL-39** (*weiyang*) ■■: lower *he*-sea point of the Triple Burner (→ 8.1.9)
- **BL-40** (*weizhong*) ■■: *he*-sea point (Earth); lower *he*-sea point of the Bladder (→ 8.1.9); Gao Wu command point (→ 8.1.11) of the back and lumbar region; Ma Dan Yang Heavenly Star point (→ 8.1.14); distal point for the thoracic and lumbar spine as well as for the iliosacral region (→ 8.2.1); local point for the knee (→ 8.2.1)
- **BL-41** (*fufen*): *jiaohui*-meeting point with the S.I. channel (→ 8.1.10)
- **BL-43** (*gaohuang*): adjacent point for the Lung (→ 8.2.1)
- **BL-49** (*yishe*): local point for the hip (→ 8.2.1)
- **BL-50** (*weicang*): local point for the hip (→ 8.2.1)
- **BL-52** (*zhishi*): local point for the lumbar region (→ 8.2.1)
- **BL-54** (*zhibian*): local point for the hip (→ 8.2.1)
- **BL-57** (*chengshan*) ■■: Ma Dan Yang Heavenly Star point (→ 8.1.14); distal point for the rectum (→ 8.2.1); distal point for the thoracic spine (→ 8.2.1)
- **BL-58** (*feiyang*): *luo*-connecting point (→ 8.1.2); distal point for the shoulder and iliosacral region (→ 8.2.1)
- **BL-59** (*fuyang*): *xi*-cleft point of the *yang qiao mai; jiaohui*-meeting point with the *yang qiao mai* (→ 8.1.10)*; distal point for the lumbar spine (→ 8.2.1)
- **BL-60** (*kunlun*) ■■: *jing*-river point (Fire); Ma Dan Yang Heavenly Star point (→ 8.1.14); distal point for the cervical, thoracic and lumbar spine (→ 8.2.1); distal point for the occipital and lumbar region (→ 8.2.1); local point for the foot (→ 8.2.1); regional point for the toes (→ 8.2.1)
- **BL-61** (*pucan*): *jiaohui*-meeting point with the *yang qiao mai* (→ 8.1.10)
- **BL-62** (*shenmai*) ■■:Opening point of the *yang qiao mai* (→ 8.1.8); Sun Si Miao Ghost point (→ 8.1.15); *jiaohui*-meeting point with the *yang qiao mai* (→ 8.1.10); local point

* Mentioned by only some authors

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 250

for the Kidney (→ 8.2.1); local point for the foot (→ 8.2.1); distal point for the lumbar spine and hip (→ 8.2.1)

- **BL-63** (*jinmen*): *xi*-cleft point (→ 8.1.3); *jiaohui*-meeting point with the *yang wei mai* (→ 8.1.10)
- **BL-64** (*jinggu*): *yuan*-source point (→ 8.1.1)
- **BL-65** (*shugu*): *shu*-stream point (Wood); sedation point; distal point for the occiput (→ 8.2.1); adjacent point for the toes (→ 8.2.1)
- **BL-66** (*zutongu*): *ying*-spring point (Water); *ben* point (Five Phases point)
- **BL-67** (*zhiyin*) ■■: *jing*-well point; tonification point

General location help

- **The medial branch of the Bladder channel** runs from **BL-11** (below the spinal process of the T1) to **BL-30** (level with the fourth sacral foramen), 1.5 cun lateral to the midline

Memorisation help for the medial branch of the BL channel:

- **BL11 to BL-17**: the second digit of the point number corresponds to the level of the respective thoracic vertebra. Example: **BL-13** is located below the spinal process of T3.
- **BL-18 to BL-21**: second digit of the point number + 1. *Example*: **BL-19** (9 + 1) is located below the spinal process of the T10.

The outer branch of the Bladder channel runs from **BL-41** (below the spinal process of the T2) to **BL-54** (below the spinal process of S4/level with the 4th sacral foramen) 3 cun lateral to the midline.

Memorisation help for the lateral branch of the BL channel:

- **BL41 to BL-46**: second digit of the point number + 1. *Example*: **BL-42** (2 + 1) corresponds to T3
- **BL-47 to BL-49**: second digit of the point number + 2. *Example*: **BL-47** (7 + 2) corresponds to the level of T9.
- **BL-50**: second digit of the point number + 12.

Needling: The Back-*shu* points are also commonly needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape).

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 251

Bright Eyes JINGMING BL-1

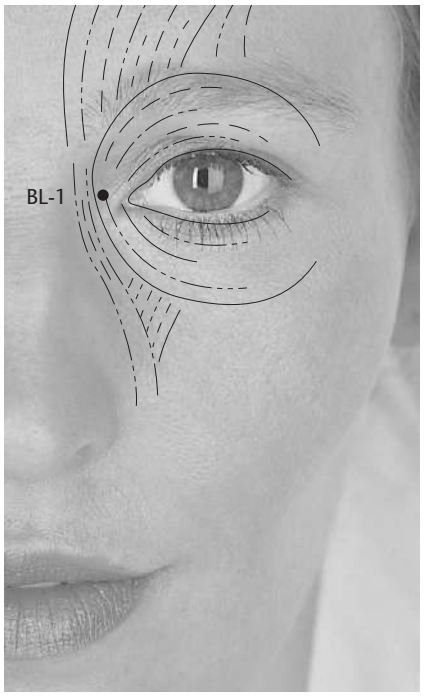

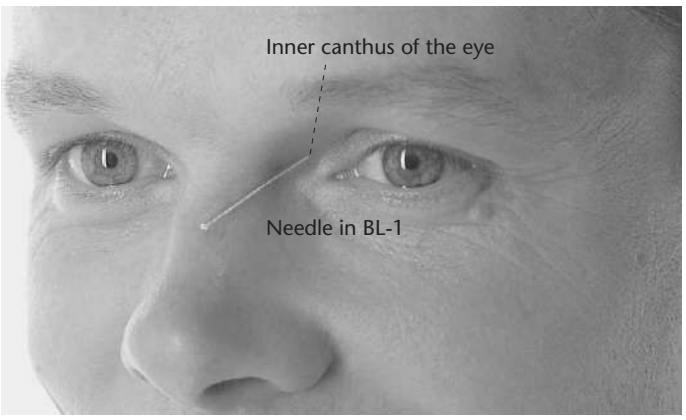

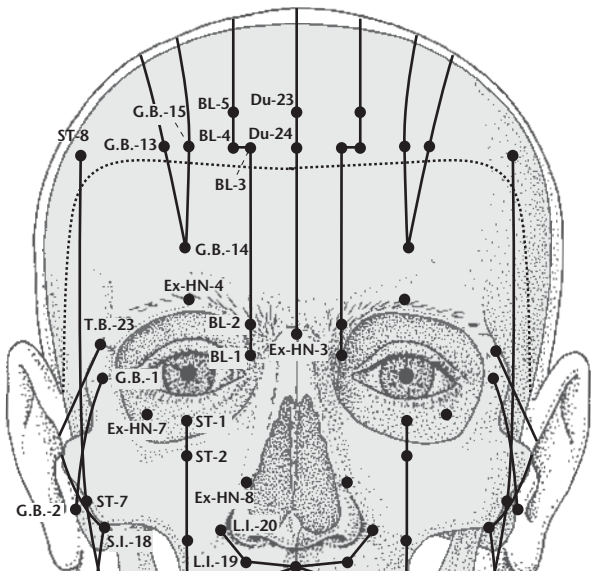

Location

0.1 cun superior and medial to the inner canthus of the eye, in a depression.

How to find

BL-1 is located in a small depression near the inner canthus of the eye, superior to the insertion of the upper eyelid. Medially, it is bordered by the bony root of the nose. → **BL-2** is located directly above **BL-1** in a depression at the medial end of the eyebrow. → **ST-1** is located on the pupil line between the eyeball and the border of the infraorbital ridge, → **G.B.-1** and → **T.B.-23** are located on the outer canthus of the eye.

Needling

Ask the patient to look downward and outward, with their eyes closed. Gently push the eyeball downward and outward from above. Needle perpendicularly along the supraorbital ridge in a posterior direction. Caution: Angular artery and vein, eyeball, periosteum. This point should only be needled by experienced practitioners. After removing the needle, press on the insertion site for 10 minutes. Despite this precaution, a bruise may still form (inform the patient). **No** needle stimulation! Moxibustion is contraindicated. Alternative point: → **BL-2** or transverse (subcutaneous) insertion at **BL-1** in the direction of → **BL-2**.

Actions/Indications

- Expels Wind and clears Heat, benefits the eyes

Special features

Meeting point with the *du mai*, *yin qiao mai*, *yang qiao mai* as well as the ST, S.I., G.B. channels and, according to some authors, also the T.B. channel. Entry point.

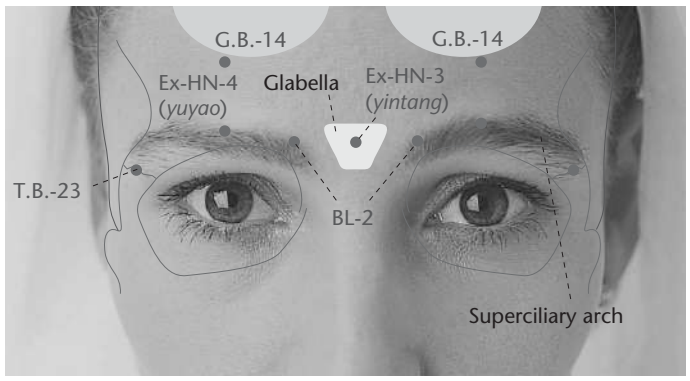

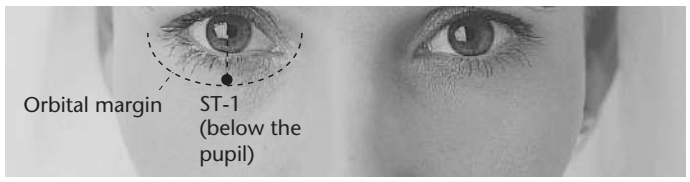

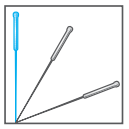
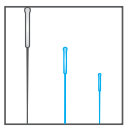
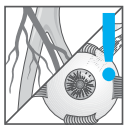

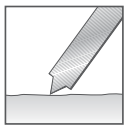

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 252

BL-2 *Gathered Bamboo* ZANZHU

Location

In a depression at the medial end of the eyebrow, directly above the inner canthus of the eye.

How to find

As the location of the eyebrow can vary, the medial canthus of the eye should be used for reference (→ **BL-1**). Palpate for a generally pressure-sensitive depression in this area of the orbital ridge. Caution: According to Sobotta, except in rare cases, **BL-2** is not located on the supraorbital foramen (mainly located more laterally), but in the area where the supratrochlear artery and the supraorbital nerve emerge (frontal notch).
→ **BL-1** is located inferior to **BL-2**.

Needling

0.3–0.5 cun obliquely from lateral to medial or transversely (subcutaneously) in the direction of the disorder. Caution: Branches of the frontal nerve, facial nerve, blood vessels! According to the majority of authors, moxibustion is contraindicted. Considered a less risky alternative to → **BL-1**.

Actions/Indications

- Benefits the eyes
- Expels external pathogenic factors such as Wind, clears Heat
- Clears the head
- Moves Qi in the Bladder divergent channel (for example, for painful haemorrhoids)

Special features

Important and frequently used local point for the eye region. Good point for excess and Heat syndromes of the eye region: prick to bleed (puncture the point and let it bleed), but avoid facial haematoma.

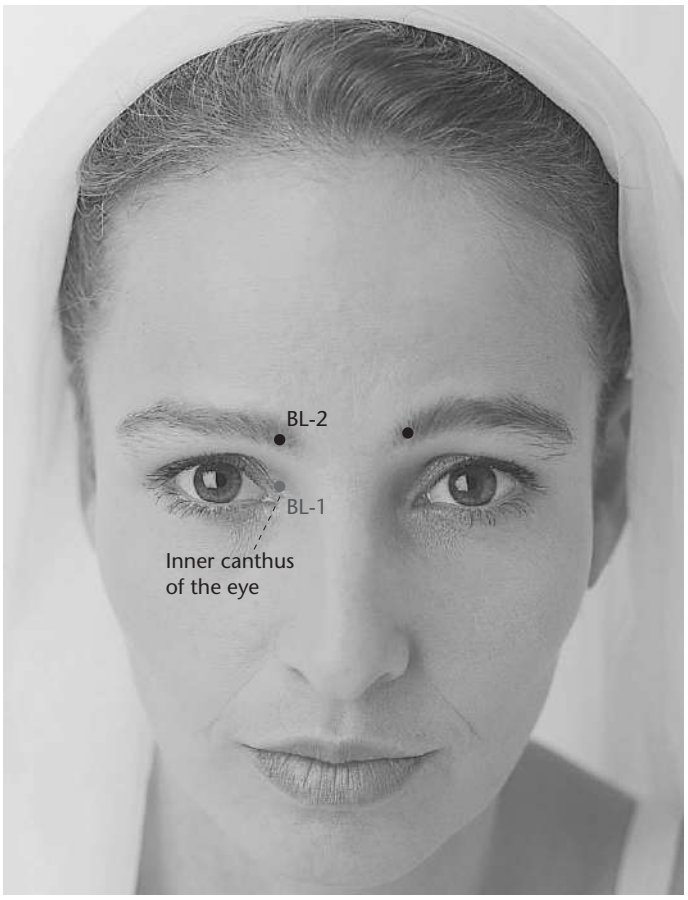

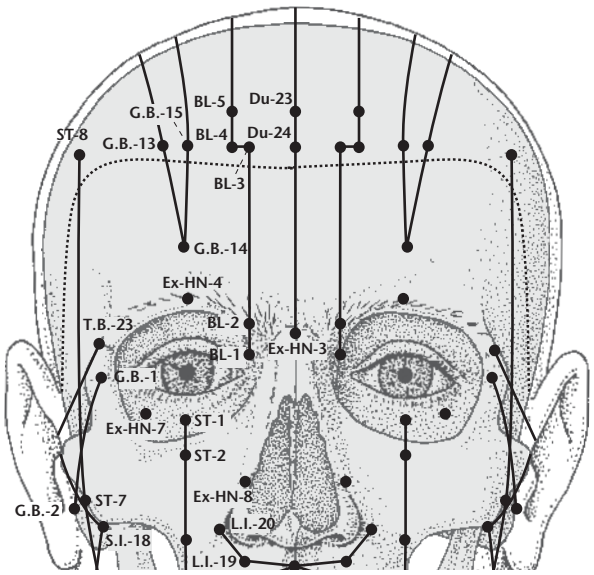

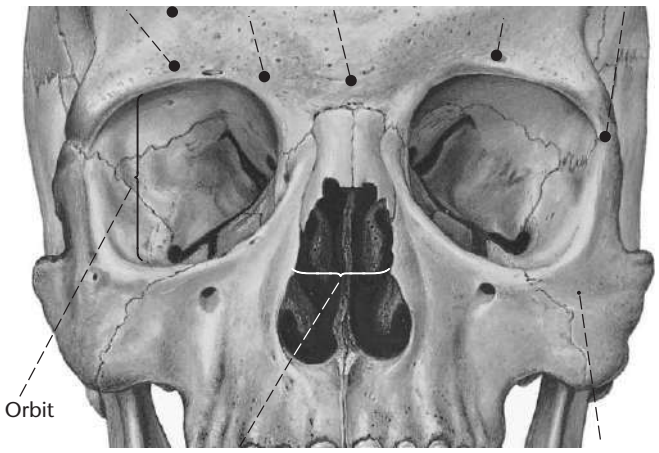

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 253

Eyebrows' Pouring MEICHONG BL-3

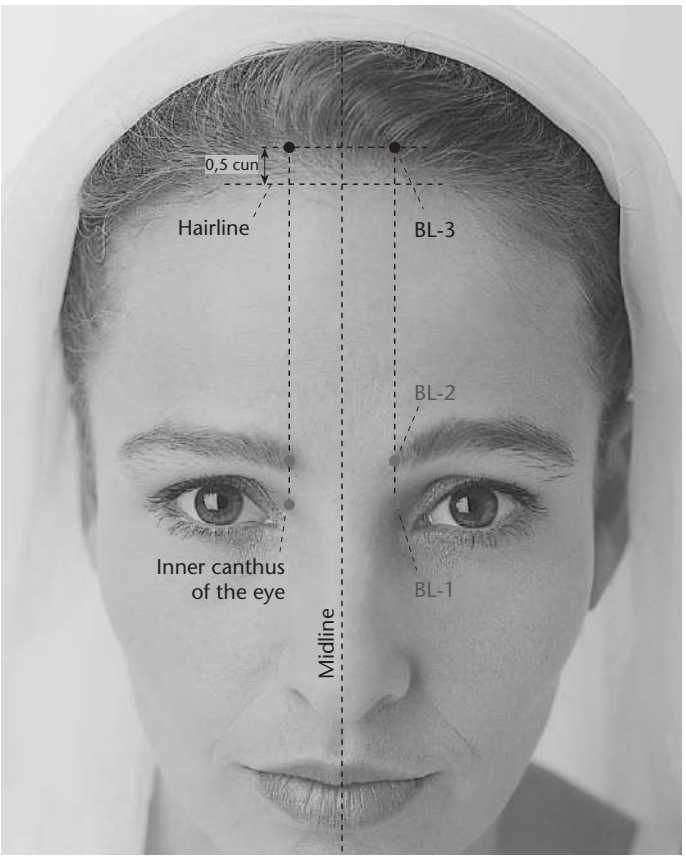

Location

0.5 cun within the anterior hairline, vertically above the medial canthus of the eye.

How to find

BL-3 is located directly above the medial canthus of the eye (→ **BL-1**) and 0.5 cun within the anterior hairline (→ 3.1.1). Located on the same level (0.5 cun within the anterior hairline) are → **Du-24** (on the midline), → **G.B.-15** (on the pupil line), → **G.B.-13** (3 cun lateral to the midline) and → **ST-8** (on the corner of the forehead).

Needling

0.5–1 cun obliquely in a dorsal direction or transversely (subcutaneously). According to some classic texts, moxibustion is contraindicated.

Actions/Indications

- Expels Wind and clears Heat from the head and forehead

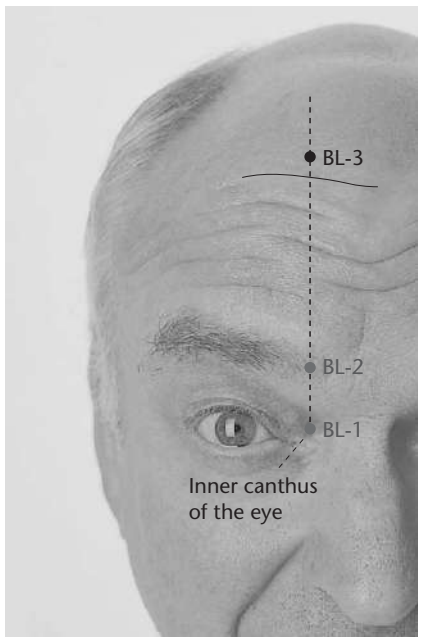

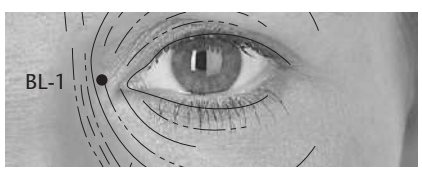

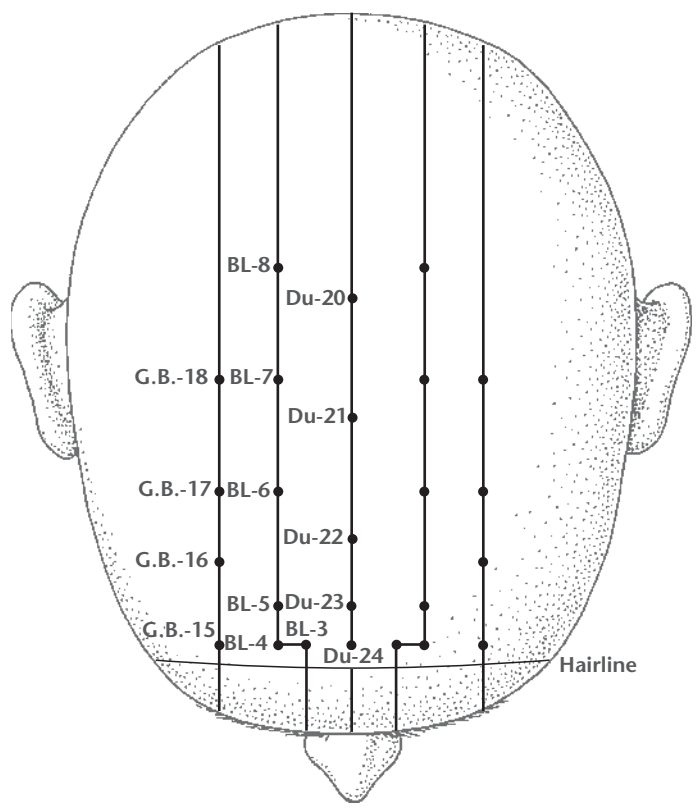

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 254

BL-4 *Crooked Curve* QUCHA

Location

0.5 cun superior to the anterior hairline and 1.5 cun lateral to the midline (or one third of the distance between → **Du-24** and → **ST-8**).

How to find

While → **BL-3** is still located on a vertical line through the inner canthus of the eye, **BL-4** is located slightly more laterally (hence its name). The distance of 1.5 cun from the midline refers to the proportional cun measurements along the anterior hairline (→ 3.1.1): the distance between → **Du-24** (on the midline) and → **ST-8** (on the corner of the forehead) is 4.5 proportional cun; therefore, the distance between **BL-4** and the midline is one third of this distance.

Located on the same level (0.5 cun within the anterior hairline) are → **Du-24** (on the midline), → **BL-3** (superior to the inner canthus of the eye), → **G.B.-15** (on the pupil line), → **G.B.-13** (3 cun lateral to the midline) and → **ST-8** (on the corner of the forehead).

Needling

Transversely (subcutaneously) 0.5–1 cun

Actions/Indications

- Expels Wind and Wind-Heat, especially from the eyes and the nose
- Clears the head, opens the orifices, alleviates pain

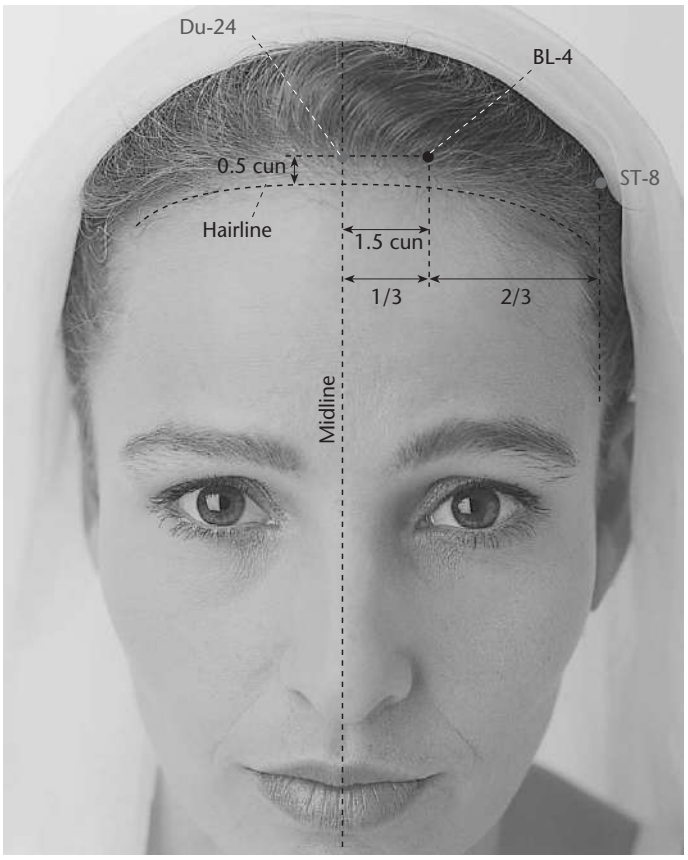

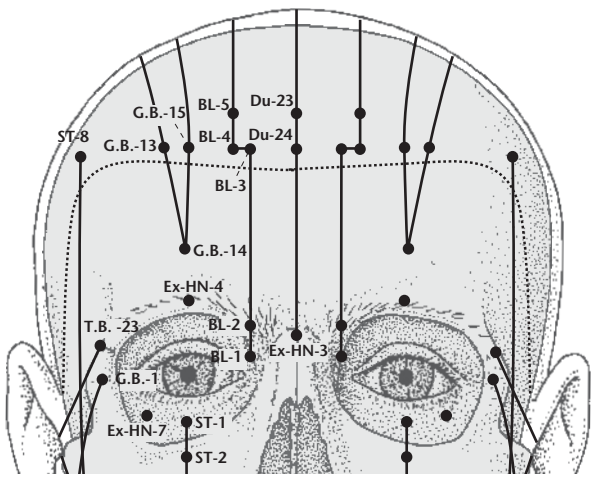

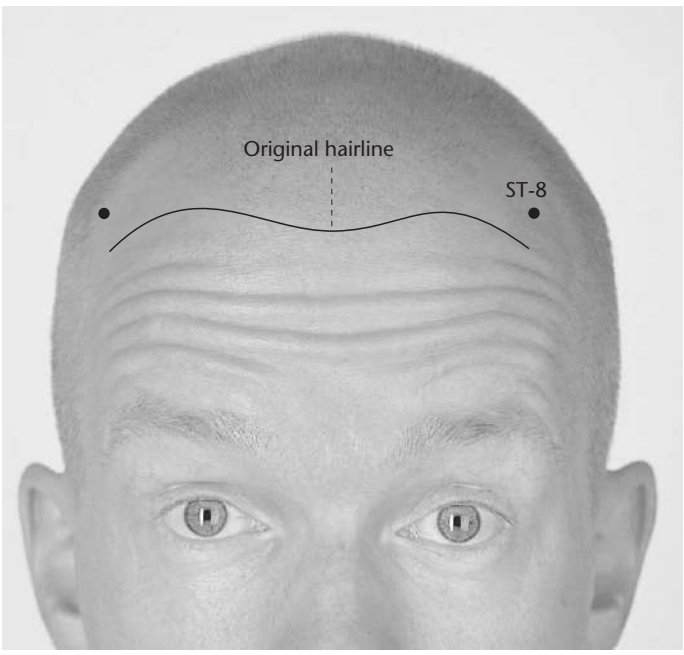

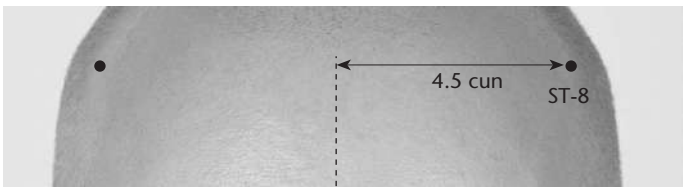

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 255

Fifth Place WUCHU BL-5

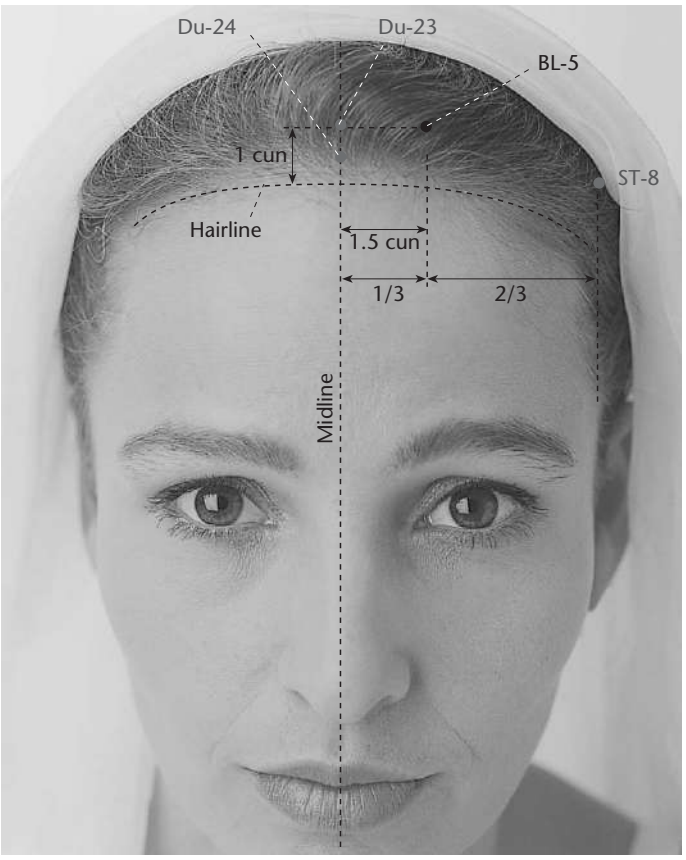

Location

1 cun superior to the anterior hairline and 1.5 cun lateral to the midline or one third of the distance between → **Du-24** and → **ST-8**.

How to find

The distance of 1.5 cun refers to the proportional cun measurement for the distance of 4.5 cun between → **Du-24** (midline, 0.5 cun superior to the anterior hairline) and → **ST-8** (corner of the forehead), (→ 2.2). By dividing this distance into thirds, **BL-5** can be located one third of the distance from → **Du-24** and 1 cun superior to the anterior hairline.
Located on the same level is → **Du-23** (on the midline).

Needling

Transversely (subcutaneously) 0.5–1 cun. Moxibustion is contraindicated according to some texts.

Actions/Indications

- Releases the Exterior, expels Wind and Heat from the head and eyes
- Opens the channel and the *luo* vessels
- Opens the orifices (especially the eyes) and revives consciousness

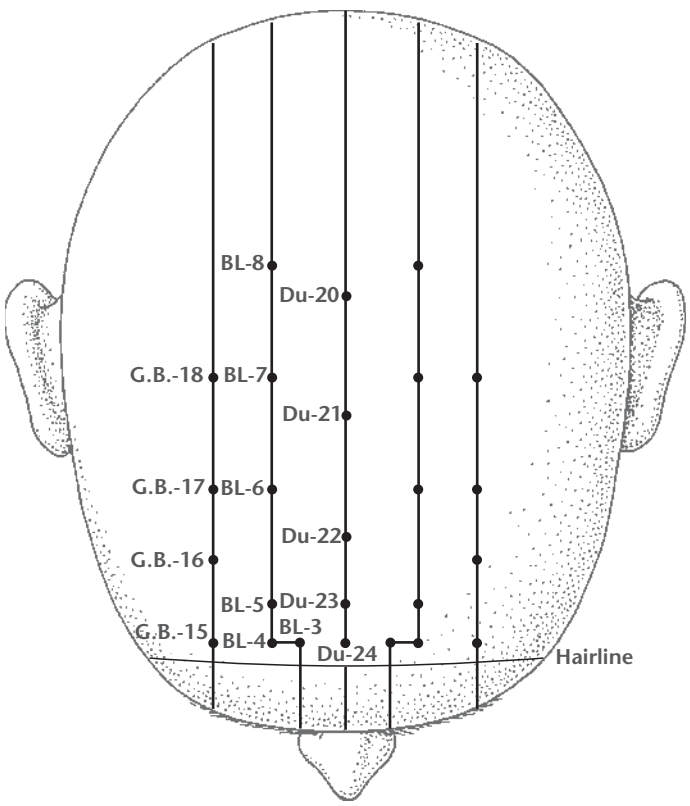

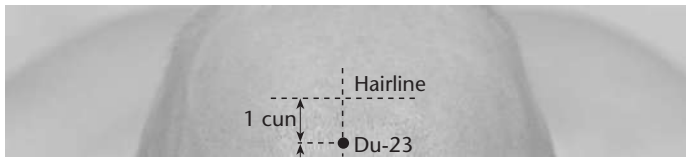

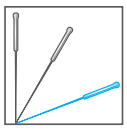
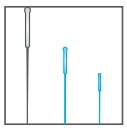

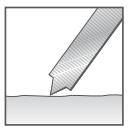

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 256

BL-6 *Receiving Light* CHENGGUAN

Location

2.5 cun superior to the anterior hairline and 1.5 cun lateral to the midline or one third of the distance between → **Du-24** and → **ST-8**.

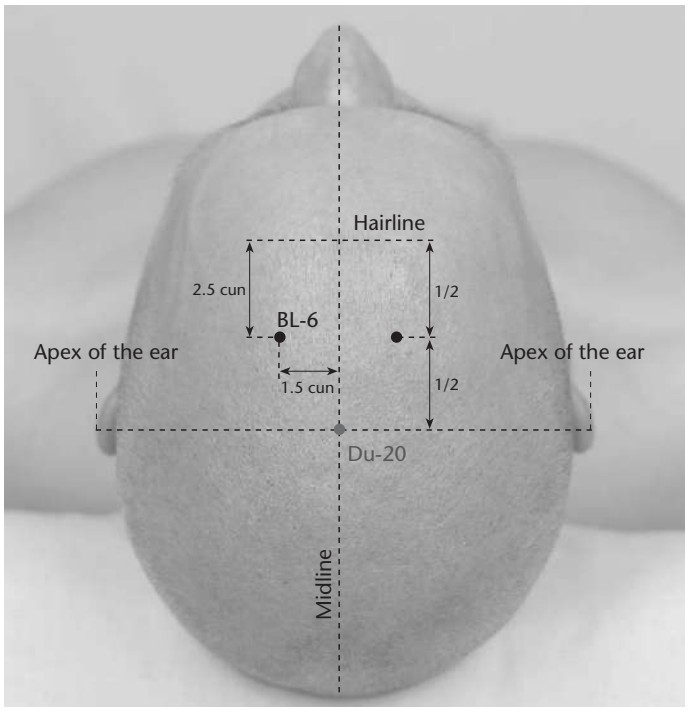

How to find

Spreading hands technique (→ 2.3.3): First, locate the anterior midline at the level of the anterior hairline (→ 3.1.1) as well as → **Du-20** (5 cun superior to the anterior hairline, on the vertex, at the junction of the midline with a line connecting the apices of the ears). Next, determine the midpoint between those two reference points (→ 2.5 cun superior to the anterior hairline). From this point, measure 1.5 cun laterally (or one third of the distance between → **Du-24** and → **ST-8**) to locate **BL-6**.
Located on the same level (→ 2.5 cun superior to the anterior hairline), but more laterally (pupil line), is → **G.B.-17**.

Needling

Transversely (subcutaneously) 0.5–1 cun. Moxibustion only following careful diagnosis.

Actions/Indications

- Expels Wind, especially from the head and eyes
- Clears Heat from the head, especially from the eyes and nose

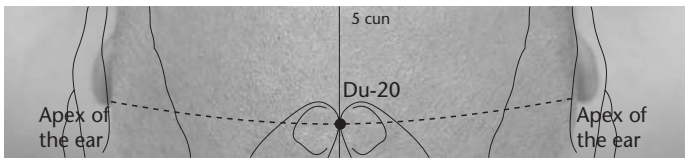

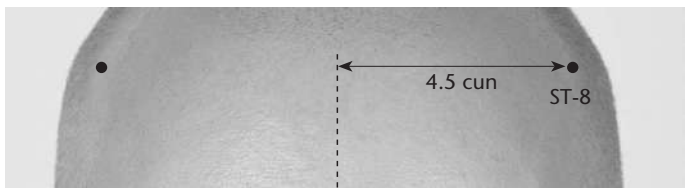

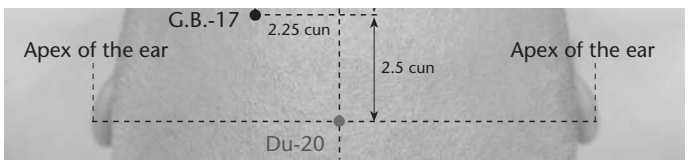

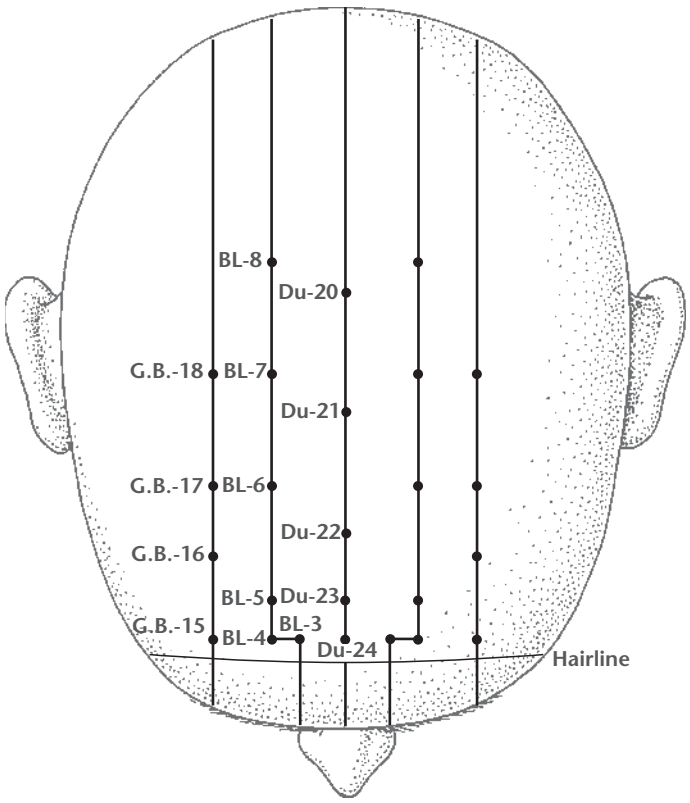

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 257

Heavenly Connection TONGTIAN BL-7

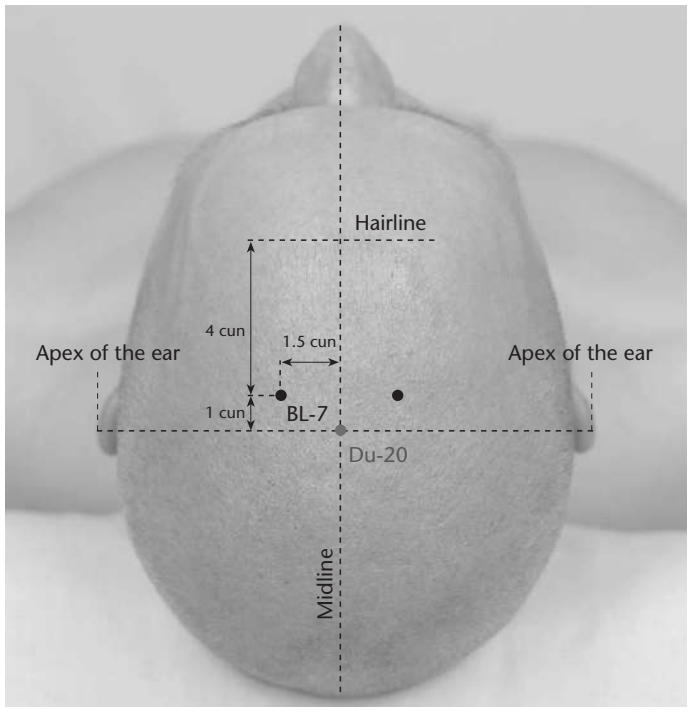

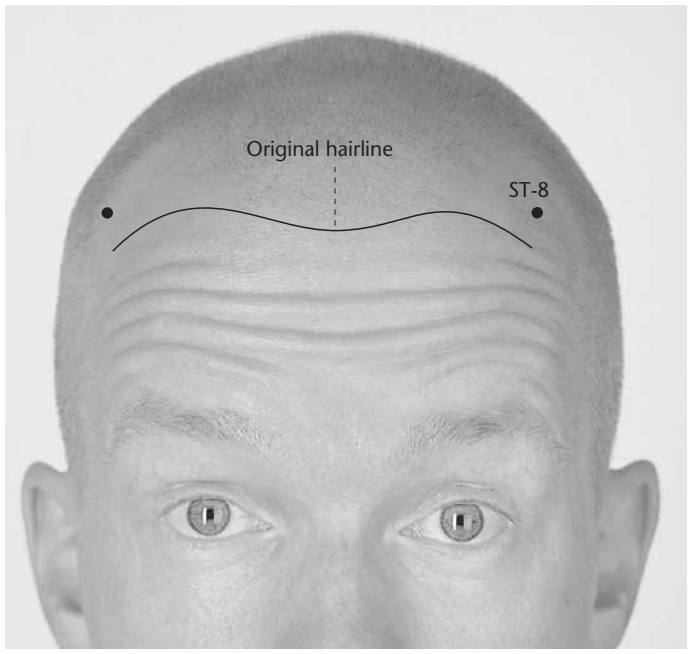

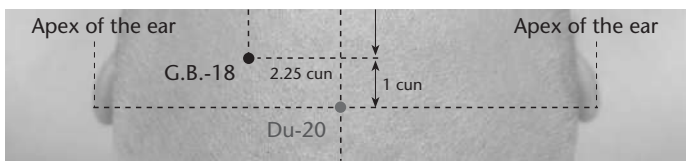

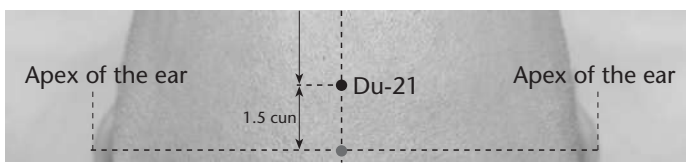

Location

1.5 cun lateral to the midline and 4 cun superior to the anterior hairline or 1 cun anterior to → **Du-20**.

How to find

BL-7 is best located by using → **Du-20** (5 cun superior to the anterior hairline, on the vertex, at the junction of the midline with a line connecting the apices of the ears) as a reference point. Measure 1 cun anterior from → **Du-20** and 1.5 cun lateral to the midline (or one third of the distance between → **Du-24** and **ST-8,** → 2.2) to locate **BL-7**. Located on the same level (4 cun superior to the anterior hairline), but more laterally (pupil line), is → **G.B.-18**. → **Du-21** is located on the midline, 1.5 cun anterior to → **Du-20**.

Needling

Transversely (subcutaneously) 0.5–1 cun

Actions/Indications

- Clears the head, benefits especially the nose

Special features

Important point for treating disorders of the nose

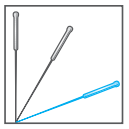
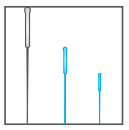

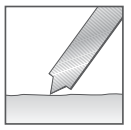
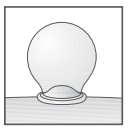

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 258

BL-8 *Declining Connection* LUOQUE

Location

1.5 cun lateral to the midline and 5.5 cun superior to the anterior hairline or 0.5 cun posterior to → **Du-20**.

How to find

BL-8 is best located by using → **Du-20** as a reference point. → **Du-20** is located at the junction of the vertical midline with a line connecting the apices of the ears, 5 cun superior to the anterior hairline (→ 3.1.1). To locate **BL-8**, measure 0.5 cun from → **Du-20** in a posterior direction and 1.5 cun in a lateral direction.

Needling

Transversely (subcutaneously) 0.5–1 cun. Moxibustion possible.

Actions/Indications

- Opens the orifices
- Subdues Wind
- Transforms Phlegm
- Calms the *shen*

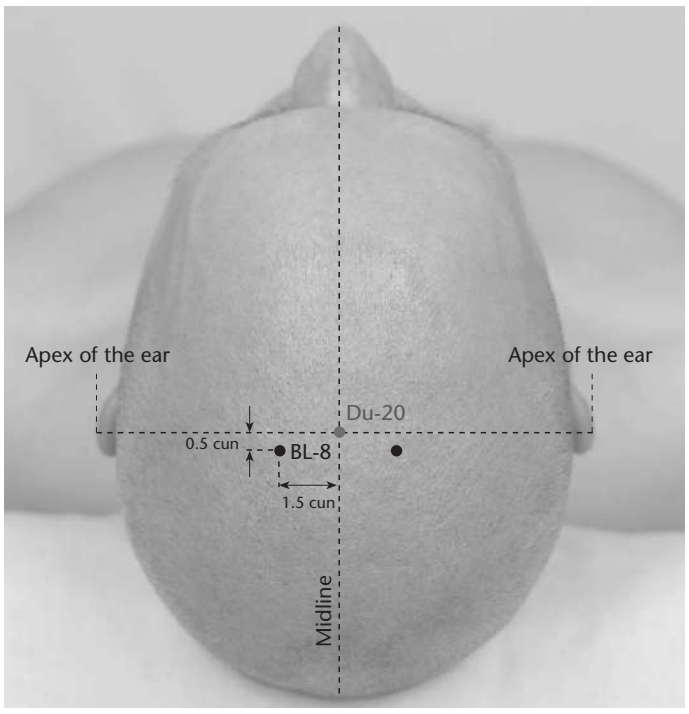

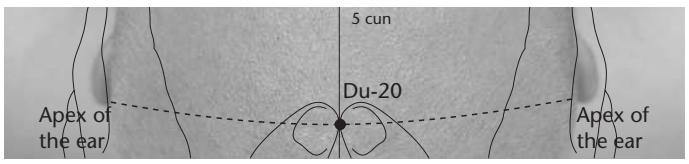

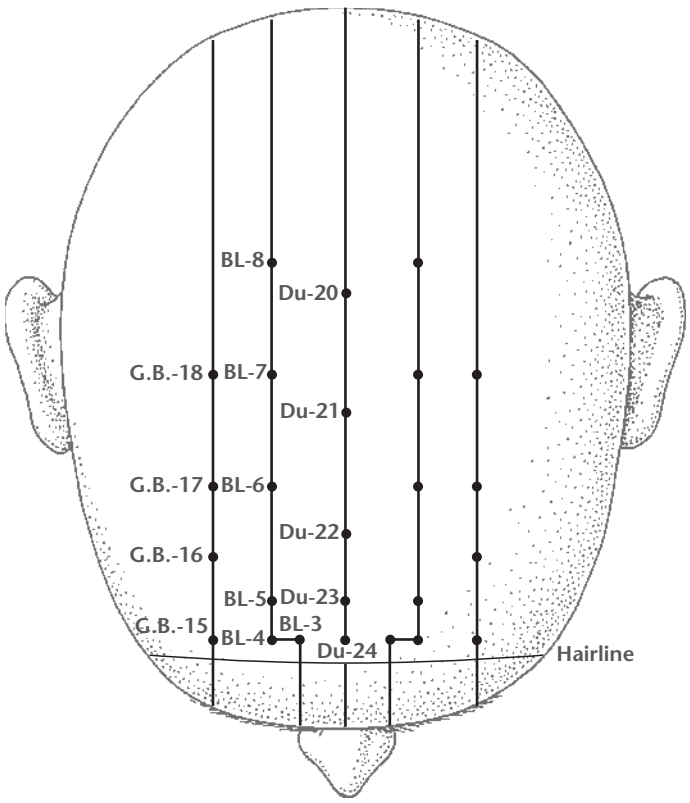

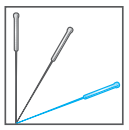
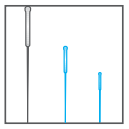

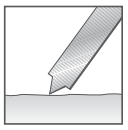

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 259

Jade Pillow YUZHEN BL-9

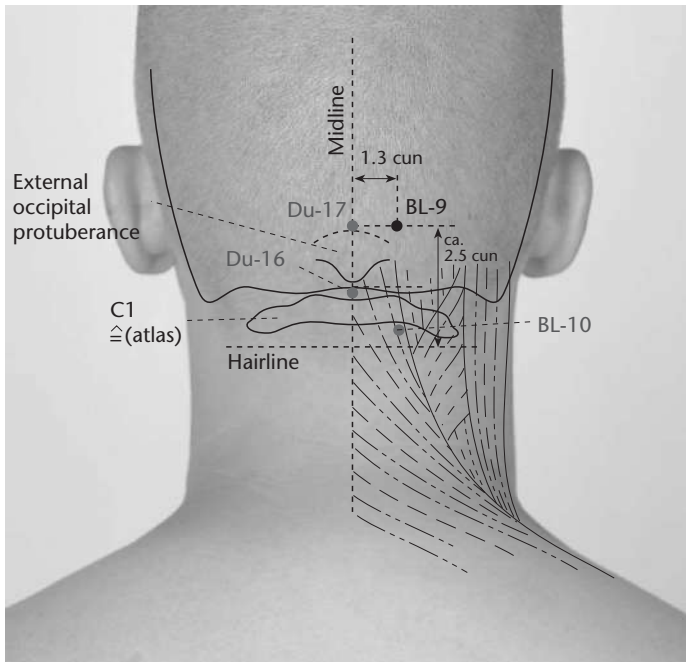

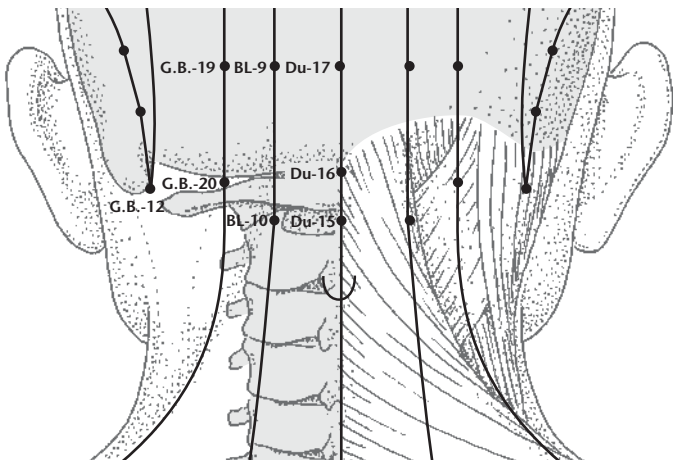

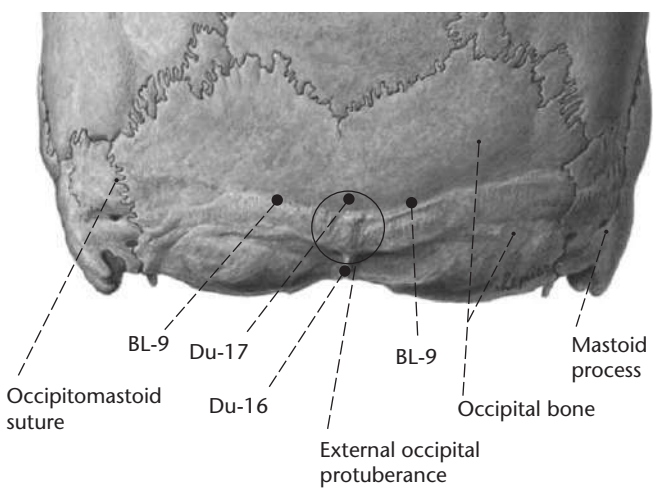

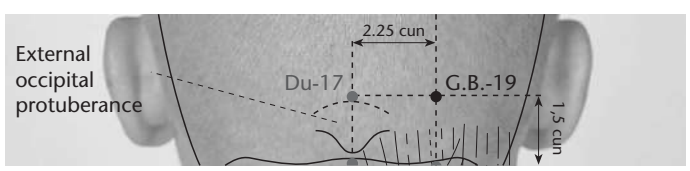

Location

2.5 cun superior to the posterior hairline and 1.3 cun lateral to the midline or ➞ **Du-17** (directly superior to the external occipital protuberance).

How to find

To locate **BL-9,** ➞ **Du-17** is a good reference point, located on the midline, in the depression directly superior to the upper border of the external occipital protuberance (➞ 3.1.5). From there, measure 1.3 cun in a lateral direction to locate **BL-9**, which is generally located on an imaginary vertical line through ➞ **BL-10. Or:** Locate the posterior hairline (often not very accurate ➞ 3.1.5) and, on the midline, measure 2.5 cun in a superior direction, then 1.3 cun towards lateral.
Located on the same level are ➞ **Du-17** (midline, in the depression at the upper border of the external occipital protuberance) and ➞ **G.B.-19** (more laterally, on a vertical line through ➞ **G.B.-20**).

Needling

Transversely (subcutaneously) 0.5–1 cun. Moxibustion possible.

Actions/Indications

- Expels Wind and Cold
- Opens the channel and *luo* vessels, alleviates pain
- Benefits the eyes and nose

Special features

In Qigong practice, **BL-9** plays an important role as one of the three gates (*san guan*).

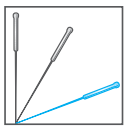
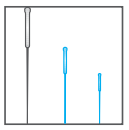

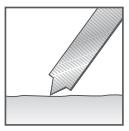

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 260

BL-10 *Heavenly Pillar* TIANZHU

Location

Approximately 1.3 cun lateral to → **Du-15**, where the trapezius muscle inserts on the lower border of the occiput, close to where the major occipital nerve emerges.

How to find

With the patient's head in an upright and relaxed position, first locate → **G.B.-20** below the lower border of the occiput and between the insertions of the sternocleidomastoid and trapezius muscles. From there, palpate in a medial and inferior direction at a 45° angle, until the finger comes to a halt against the bulge of the trapezius muscle. This is the location of **BL-10**. For reference: If the practitioner places their left middle finger on → **G.B.-20**, the (shorter) index finger will be pointing to **BL-10**. Located on the same level is → **Du-15** (approximately 1.3 cun lateral to the midline and superior to the axis (first palpable spinous process)).

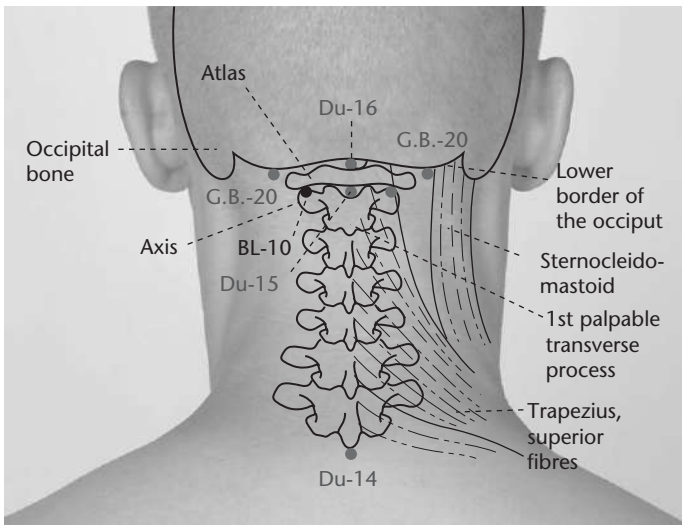

Needling

0.5–1 cun vertically or slightly obliquely in a dorsal direction. Caution: Do not needle in a superior direction.

Actions/Indications

- Regulates the Qi, calms Wind and the *shen*, benefits the head and sensory orifices
- Opens the channel
- Strengthens the lower back
- Expels Wind

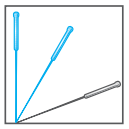

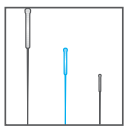

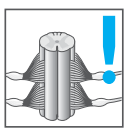

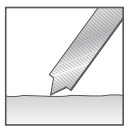

Special features

Window of Heaven point. Important point for expelling (internal and external) Wind. In contrast to → **G.B.-20**, **BL-10** is used in clinical practice mainly for pacifiying internal Wind.

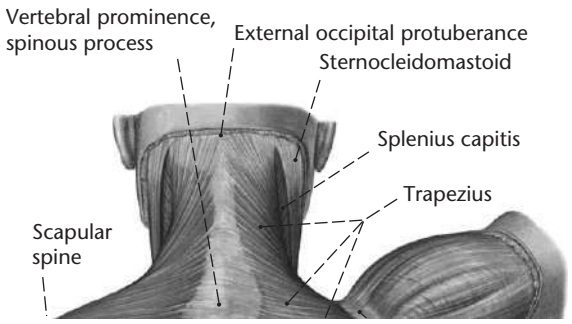

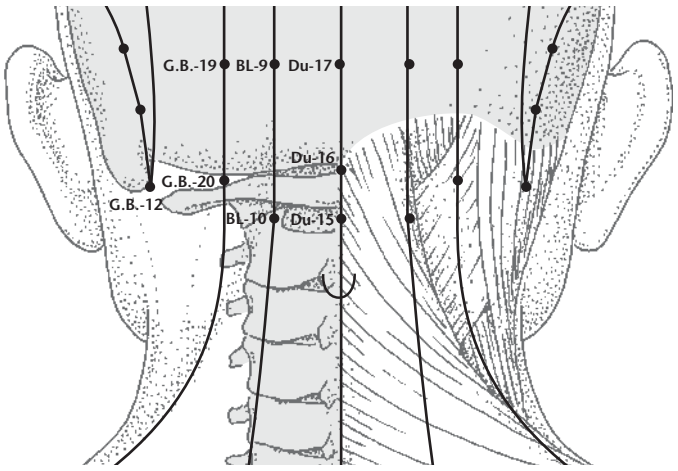

Ch04.7a-F10028.qxd 2/23/08 3:53 PM Page 261

Great Shuttle DAZHU BL-11

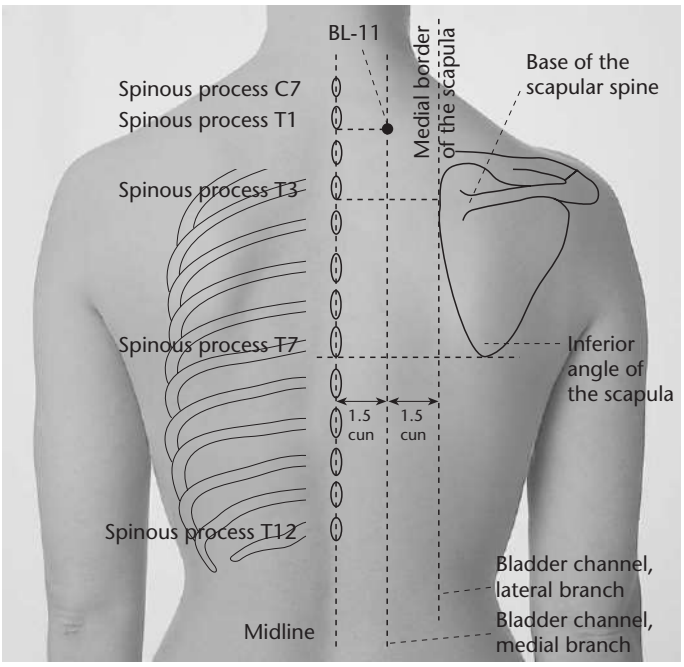

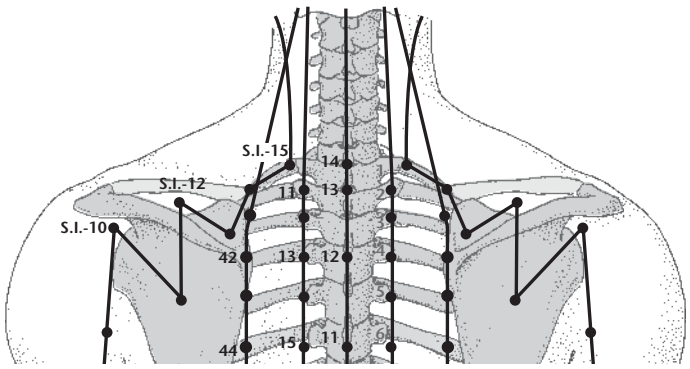

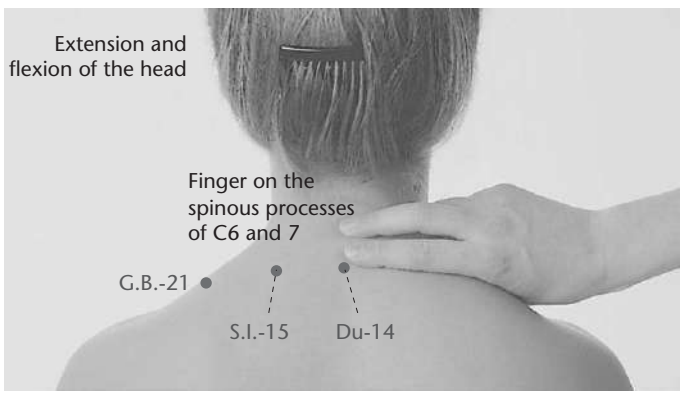

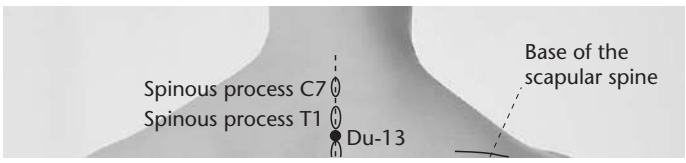

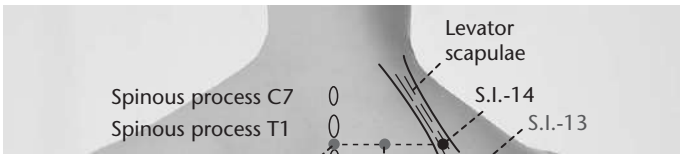

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 1st thoracic vertebra (T1, with hanging shoulders on the level of the acromion).

How to find

First, locate the spinous process of C7 (→ 3.4.1). The next spinous process inferior to this belongs to the T1. Locate **BL-11** 1.5 cun lateral to the lower border of T1, at the highest point of the paraspinal musculature.

Located on the same level are → **Du-13** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **S.I.-14** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine or transversely (subcutaneously). Caution: Pneumothorax.

Actions/Indications

- Expels pathogenic factors (Wind-Heat)
- Regulates the Lung Qi
- Opens the channel, benefits the bones and joints

Special features

Hui-meeting point with the bones; meeting point with the S.I. channel, according to some authors also with the T.B., G.B. channels and the *du mai*; Sea of Blood point.

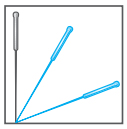
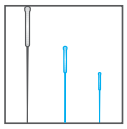
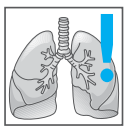
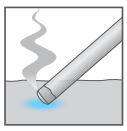
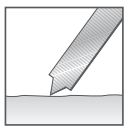
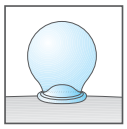

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 262

BL-12 *Wind Gate* FENGMEN

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 2nd thoracic vertebra (T2).

How to find

First, locate the spinous process of C7 (→ 3.4.1). From there, count down two spinous processes to the lower border of the spinous process of T2. Measure 1.5 cun laterally to locate **BL-12** on the highest point of the paraspinal musculature.
Located on the same level are a point of → **Ex-B-2** (0.5 cun lateral to the midline), → **BL-41** (3 cun lateral to the midline) and → **S.I.-13** (on the scapula, on the medial end of the scapular spine).

Needling

0.5–1 cun obliquely towards the spine. In clinical practice, this point is often needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape). Caution: Pneumothorax. For acute febrile colds, use reducing needling techniques or cupping. For susceptibility to catching colds, use tonifying needling techniques and/or moxibustion.

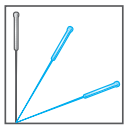
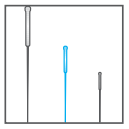
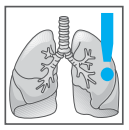

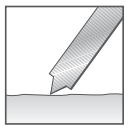
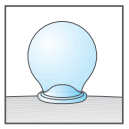

Actions/Indications

- Expels (external) Wind and releases the Exterior
- Spreads and descends the Lung Qi
- Tonifies the defensive Qi (*wei qi*), stabilises the Exterior
- Benefits the nose
- Opens the channel

Special features

Meeting point with the *du mai*. Important point for expelling pathogenic factors, especially Wind.

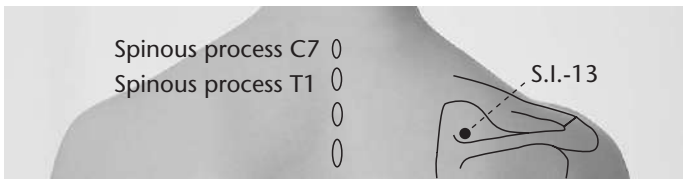

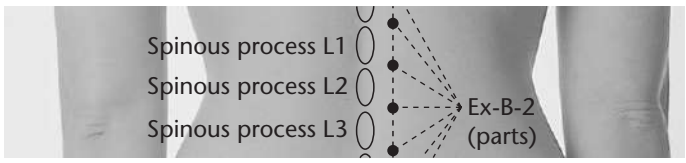

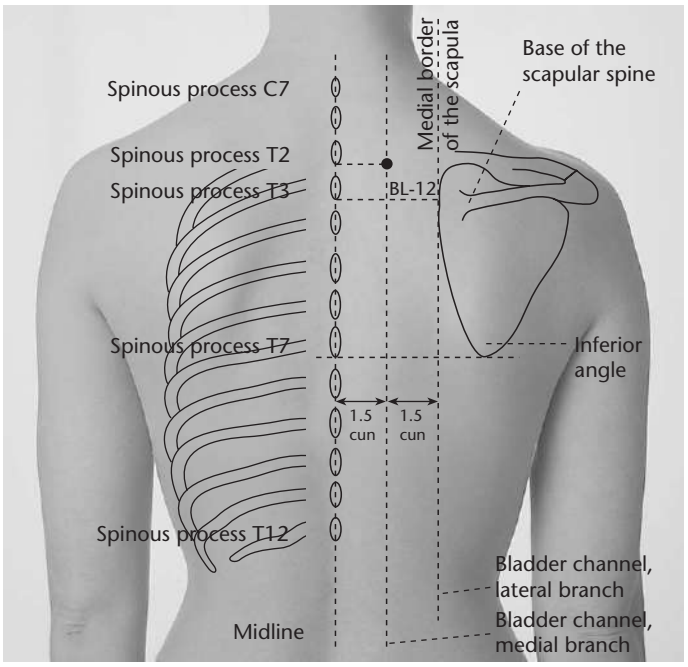

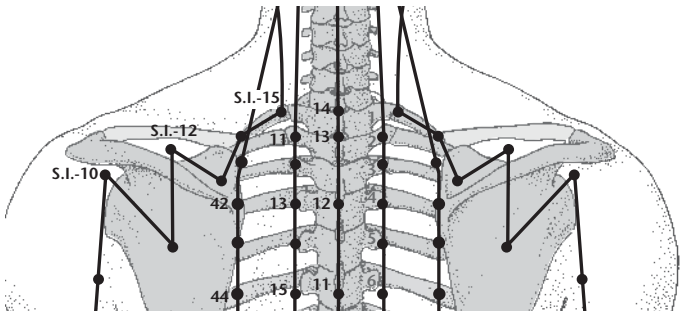

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 263

Lung Shu FEISHU BL-13

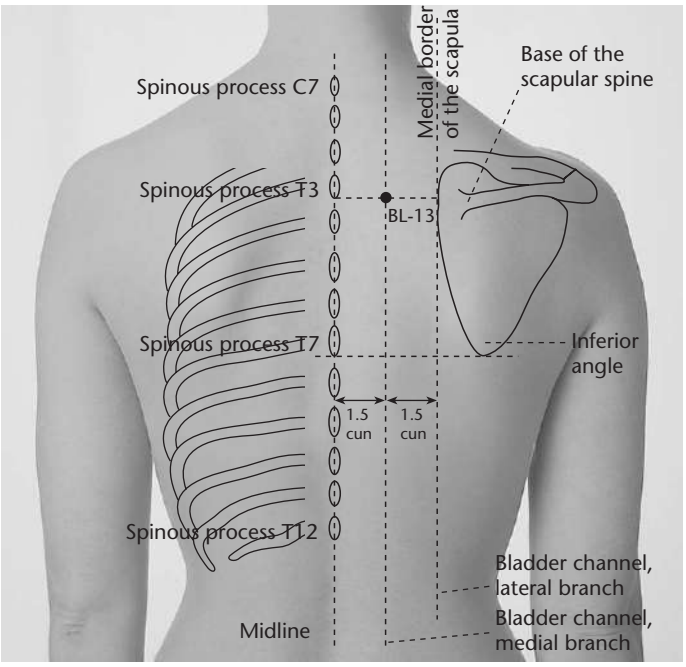

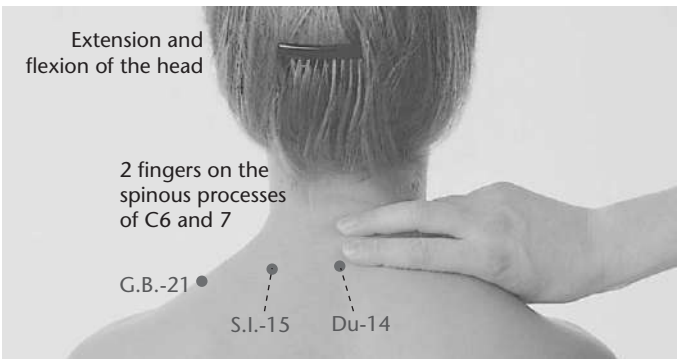

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 3rd thoracic vertebra (T3).

How to find

First, locate the spinous process of C7 (→ 3.4.1). From there, count down three spinous processes to the lower border of the spinous process of the T3. Measure 1.5 cun laterally to locate **BL-13** on the highest point of the paraspinal musculature. **Or:** With the patient seated and their arms hanging down, the spinous process of the T3 is generally on the level of the medial end of the scapular spine (→ 3.4.2). Locate **BL-13** 1.5 cun lateral to the midline.

Located on the same level are → **Du-12** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and **BL-42** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine. In clinical practice, this point is often needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape). Caution: Pneumothorax. For chronic Lung Qi deficiency without Heat signs, use moxibustion.

Actions/Indications

- Tonifies, spreads and descends the Lung Qi, nourishes the Lung Yin
- Clears Heat from the Lung
- Releases the Exterior
- Opens the channel, alleviates pain

Special features

Back-*shu* point of the lung. Major point for all disorders of the Lung (Lung deficiency and excess patterns).

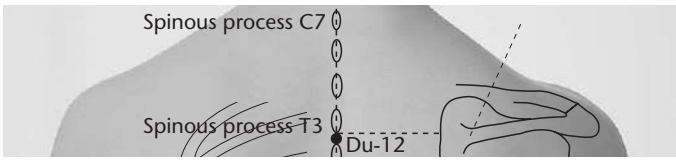

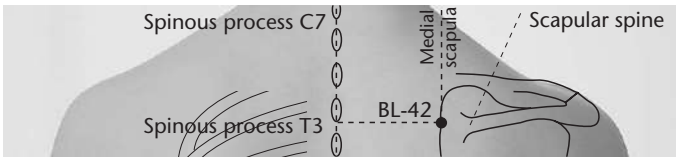

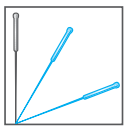
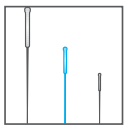
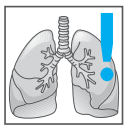
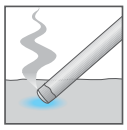
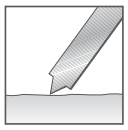
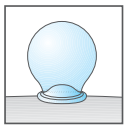

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 264

BL-14 Jueyin Shu JUEYINSHU

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 4th thoracic vertebra (T4).

How to find

First, locate the spinous process of C7 (→ 3.4.1). From there, count down four spinous processes to the lower border of the spinous process of T4. Measure 1.5 cun laterally to locate **BL-14** on the highest point of the paraspinal musculature. **Or:** With the patient seated and their arms hanging down, the spinous process of T3 is generally on the level of the medial end of the scapular spine (→ 3.4.2). The next spinous process in an inferior direction belongs to the T4. Locate **BL-14** 1.5 cun lateral to the midline. Located on the same level are a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-43** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine or transversely (subcutaneously). Caution: Pneumothorax.

Actions/Indications

- Regulates and tonifies the Heart
- Unbinds the chest, regulates the Qi
- Calms the *shen*

Special features

Back-*shu* point of the Pericardium

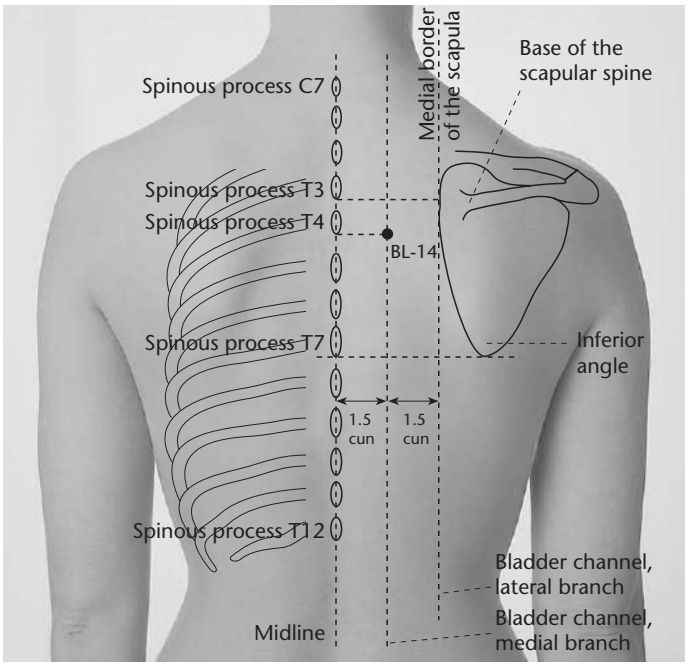

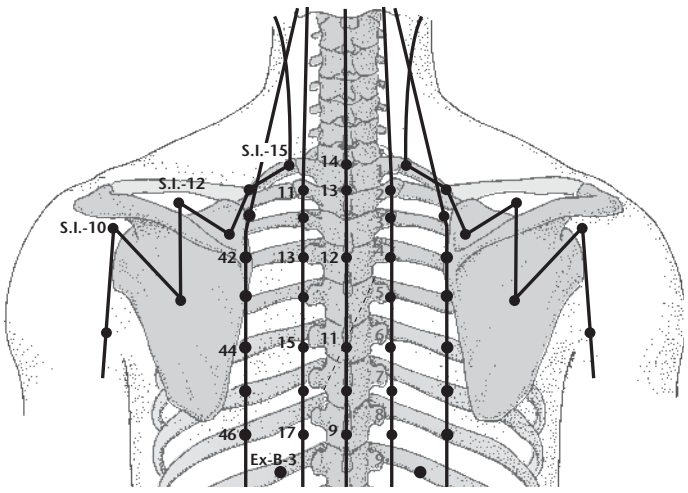

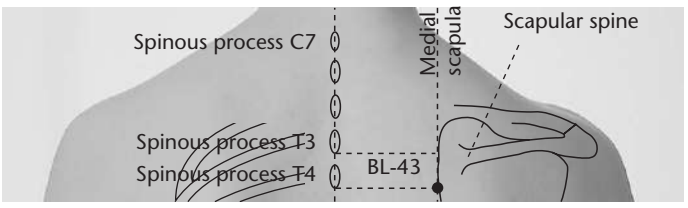

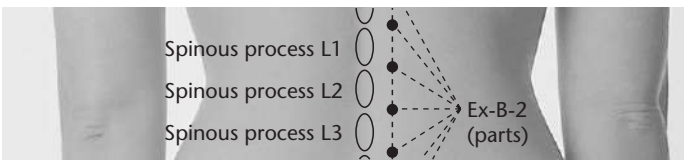

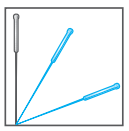
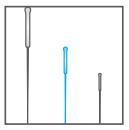
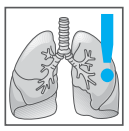

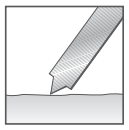
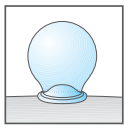

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 265

Heart Shu XINSHU BL-15

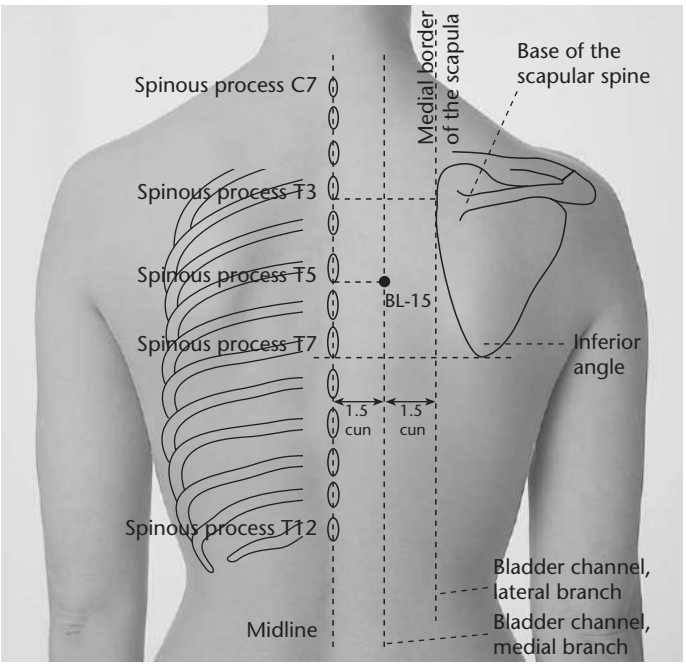

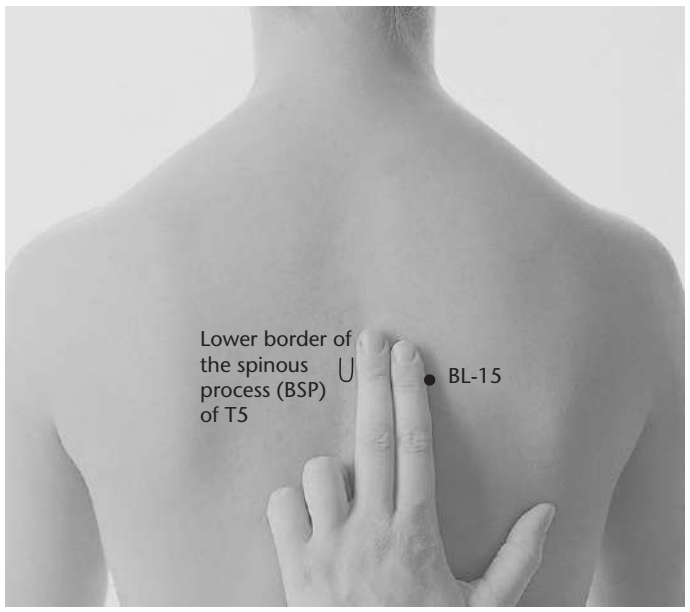

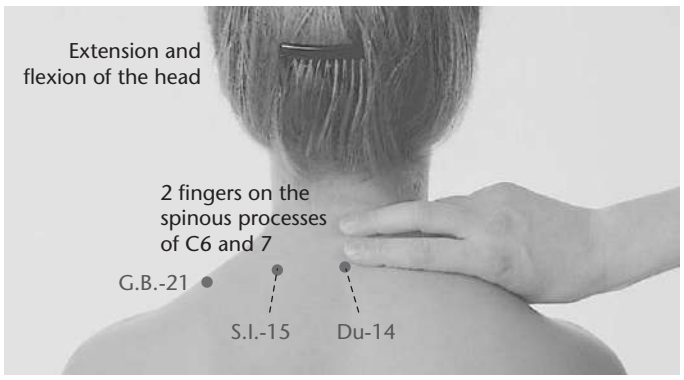

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 5th thoracic vertebra (T5).

How to find

First, locate the spinous process of C7 (→ 3.4.1). From there, count down five spinous processes to the lower border of the spinous process of T5. Measure 1.5 cun laterally to locate **BL-15** on the highest point of the paraspinal musculature. **Or:** With the patient seated and their arms hanging down, the spinous process of the T3 is generally on the level of the easily palpable medial end of the scapular spine (→ 3.4.2). From there, count down two spinous processes in an inferior direction to the T5. Locate **BL-15** 1.5 cun lateral to the midline.

Located on the same level are → **Du-11** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-44** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine. In clinical practice, this point is often needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape). Caution: Pneumothorax.

Actions/Indications

- Tonifies and nourishes the Heart, regulates the Heart Qi, unbinds the chest and removes Blood stasis, calms the *shen*, clears Fire, moves Qi locally and along the channel pathway: for example, for disorders of the thoracic spine and intercostal region

Special features

Back-*shu* point of the Heart

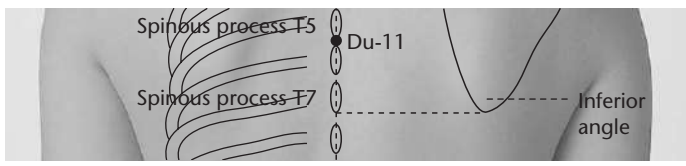

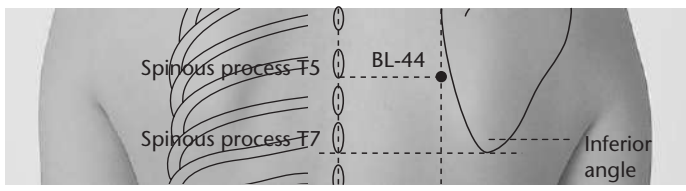

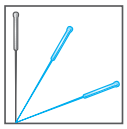
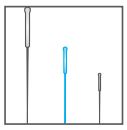
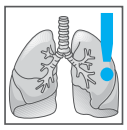
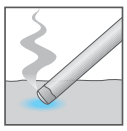
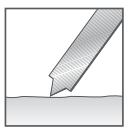
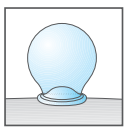

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 266

BL-16 *Governing* Shu DUSHU

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 6th thoracic vertebra (T6).

How to find

First, locate the spinous process of the C7 (→ 3.4.1). From there, count down six spinous processes to the lower border of the spinous process of T6. Measure 1.5 cun laterally to locate **BL-16** on the highest point of the paraspinal musculature. **Or:** With the patient standing and their arms hanging down, the spinous process of T7 is level with the inferior angle of the scapula (→ 3.4.2). The next spinous process in a superior direction belongs to T6. Locate **BL-16** 1.5 cun lateral to the midline.
Located on the same level are → **Du-10** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-45** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine or transversely (subcutaneously). Caution: Pneumothorax.

Actions/Indications

- Unbinds the chest and regulates the Qi movement in the thorax and the abdomen

Special features

Despite its designation as Back-*shu* point of the *du mai*, this point is hardly mentioned in classical or modern texts.

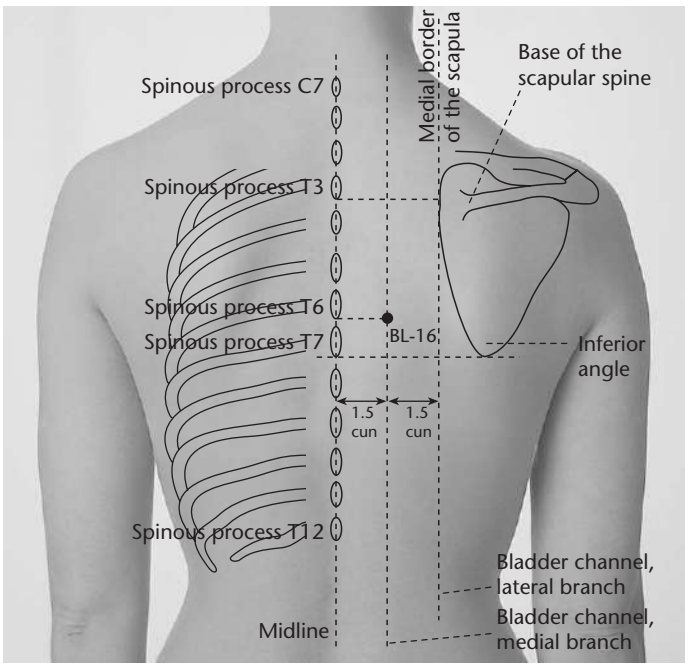

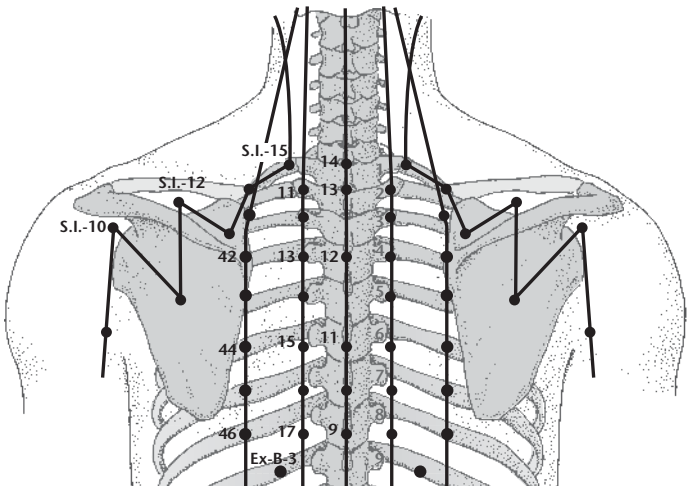

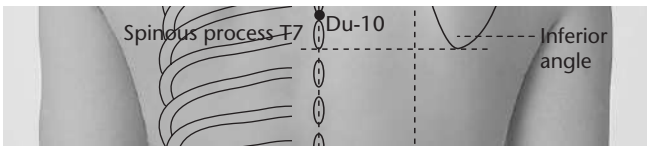

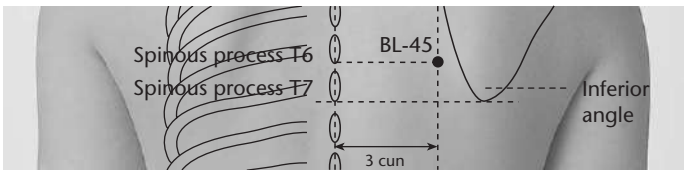

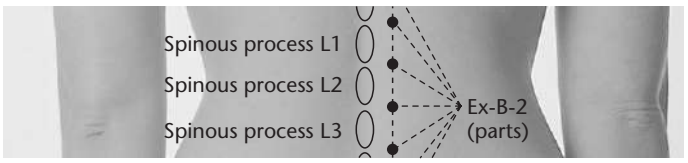

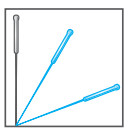

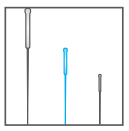

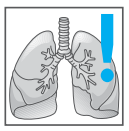

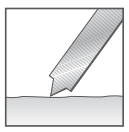

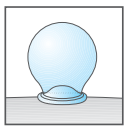

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 267

Diaphragm Shu GESHU BL-17

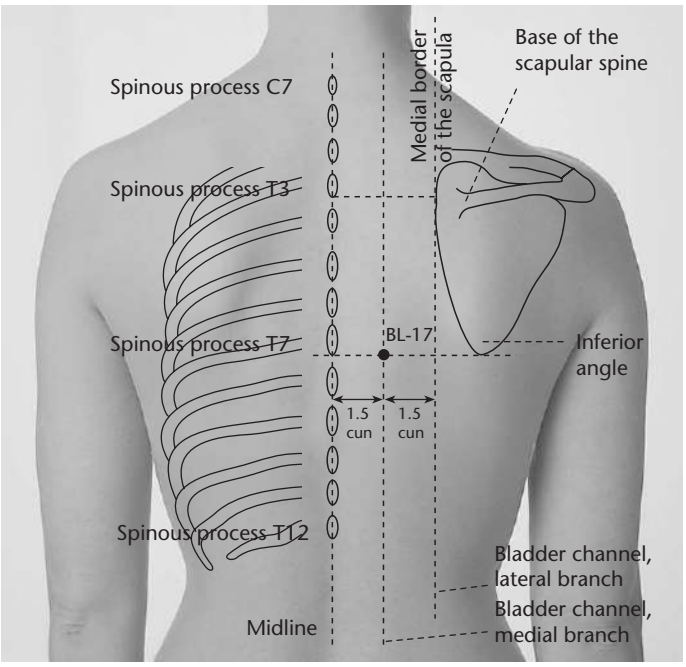

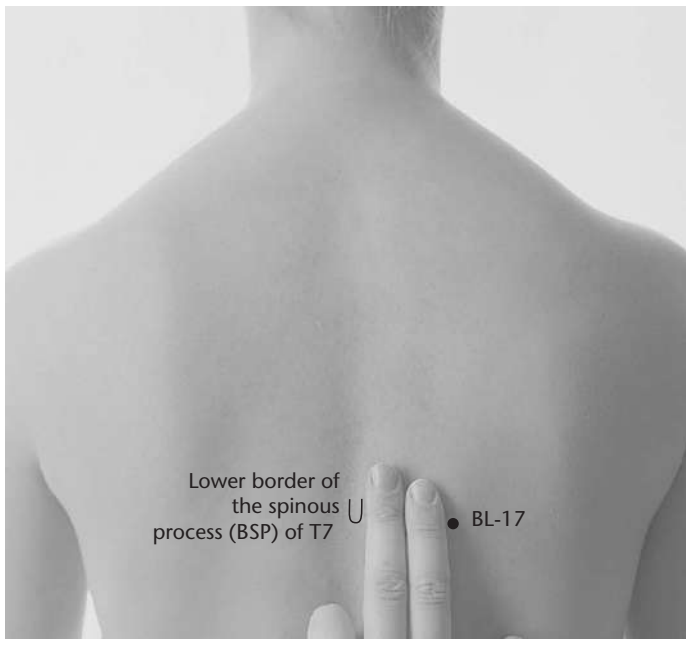

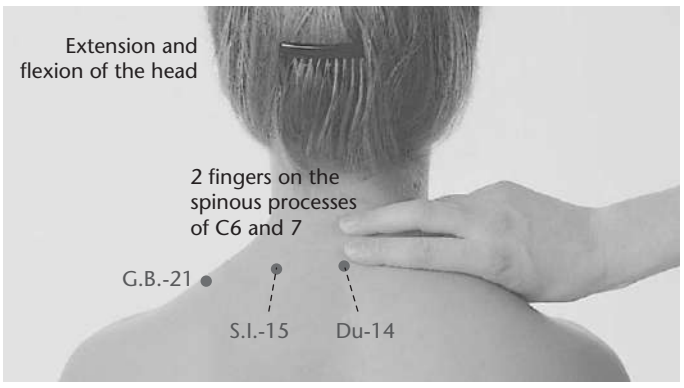

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 7th thoracic vertebra (T7).

How to find

The spinous process of the T7 is approximately on the level of the inferior angle of the scapula (→ 3.4.2). **BL-17** is located on the level of the lower border of its spinous process and 1.5 cun lateral to the midline at the highest point of the paraspinal musculature. **Or:** Use the spinous process of the T3 for reference (generally on the level of the medial end of the scapular spine) (→ 3.4.2). **Or:** Use C7 as a reference point (→ 3.4.1).
Located on the same level are → **Du-9** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-46** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine. In clinical practice, this point is often needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape). Caution: Pneumothorax.

Actions/Indications

- Cools Blood-Heat, stops bleeding, removes Blood stasis, nourishes and harmonises Blood (and Yin)
- Regulates the diaphragm, descends counterflow Qi
- Moves Qi locally and along the channel pathway

Special features

Back-*shu* point of the Diaphragm, *hui*-meeting point of the Blood.

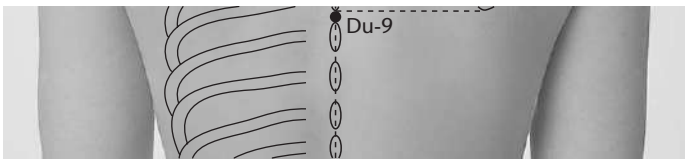

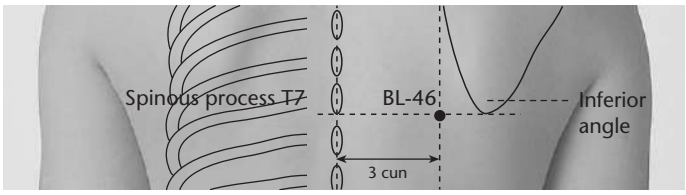

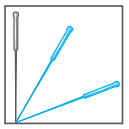
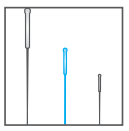
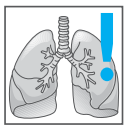

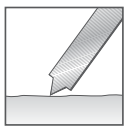
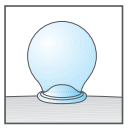

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 268

BL-18 *Liver* Shu GANSHU

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 9th thoracic vertebra (T9).

How to find

In a standing position with the arms hanging down, the spinous process of T7 is approximately on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down two spinous processes to the spinous process of T9. **BL-18** is located on the level of the lower border of its spinous process and 1.5 cun lateral to the midline at the highest point of the paraspinal musculature. **Or:** Use the C7 as a reference point (→ 3.4.1). From there, count down nine spinous processes to T9 and locate **BL-18** 1.5 cun lateral to the lower border of its spinous process.
Located on the same level are → **Du-8** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-47** (3 cun lateral to the midline).

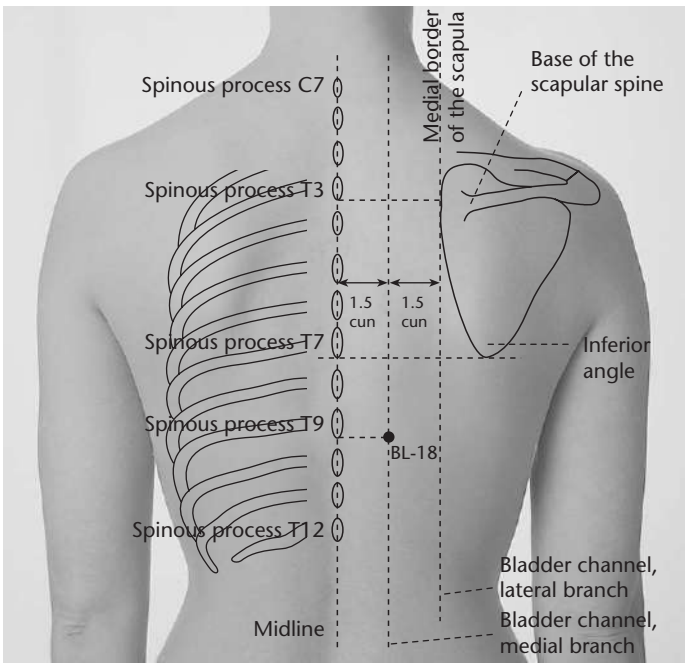

Needling

0.5–1 cun obliquely towards the spine. In clinical practice, this point is often needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape). Caution: Pneumothorax.

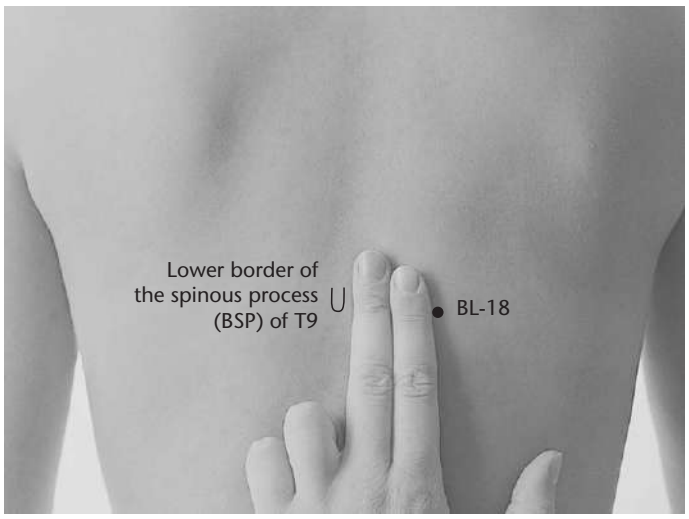

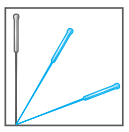

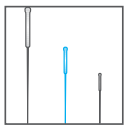

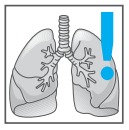

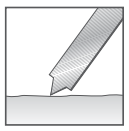

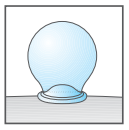

Actions/Indications

- Spreads Liver Qi, cools Fire, clears Damp-Heat, regulates and nourishes Liver Blood, subdues (internal) Wind
- Benefits the eyes
- Benefits the sinews

Special features

Back-*shu* point of the Liver

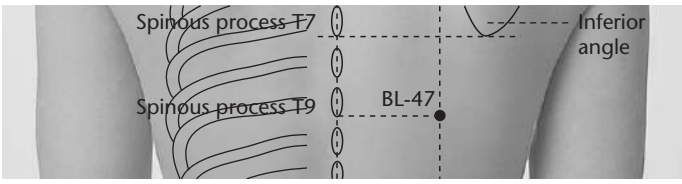

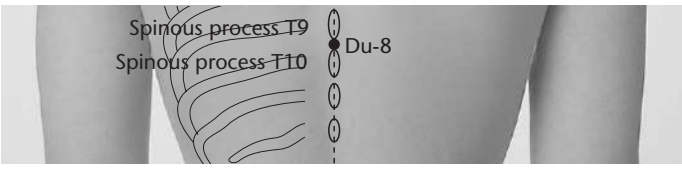

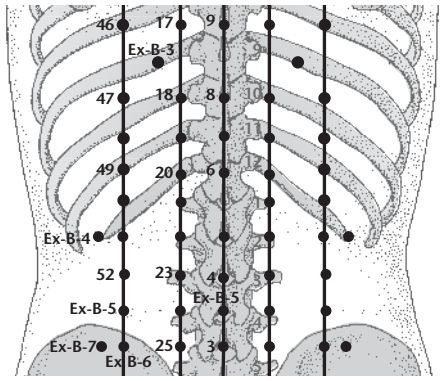

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 269

Gall Bladder Shu DANSHU BL-19

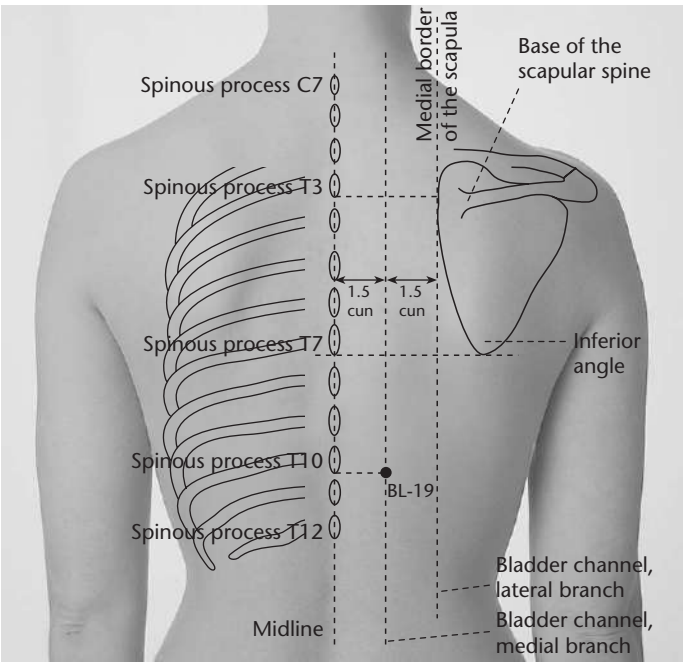

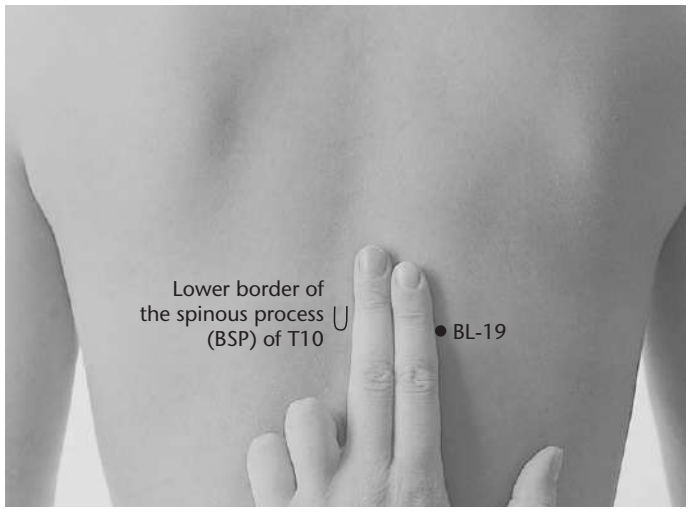

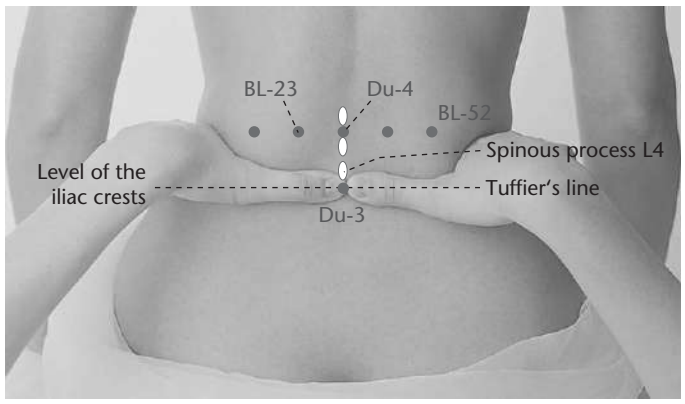

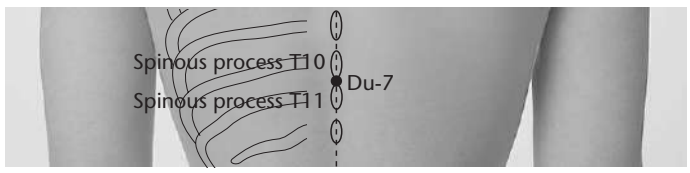

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 10th thoracic vertebra (T10).

How to find

In a standing position with the patient's arms hanging down, the spinous process of T7 is approximately on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down three spinous processes to the spinous process of T10. **BL-19** is located on the level of the lower border of its spinous process and 1.5 cun lateral to the midline at the highest point of the paraspinal musculature. **Or:** Use C7 as a reference point (→ 3.4.1). From there, count down 10 spinous processes to T10 and locate **BL-19** 1.5 cun lateral to the lower border of its spinous process. **Or:** Use the lumbar spine as a reference point (→ 3.4.3).
Located on the same level are → **Du-7** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-48** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine. In clinical practice, this point is often needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape). Caution: Pneumothorax.

Actions/Indications

- Clears Damp-Heat from the Liver and Gall Bladder
- Expels pathogenic factors from the *shaoyang*
- Tonifies and regulates the Gall Bladder Qi
- Unbinds the chest and relaxes the pathway of the channel

Special features

Back-*shu* point of the Gall Bladder. Important point for clearing Damp-Heat from the Liver and Gall Bladder.

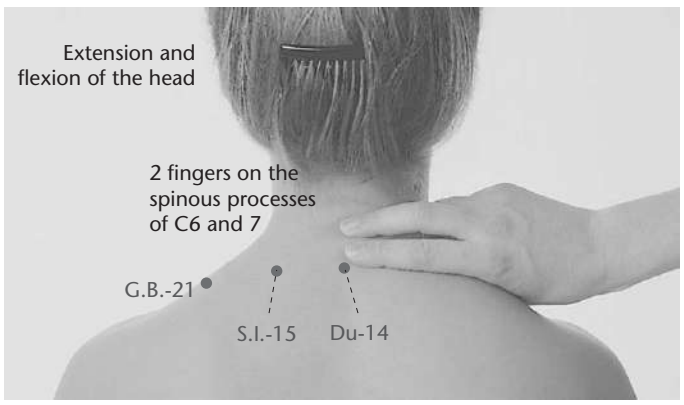

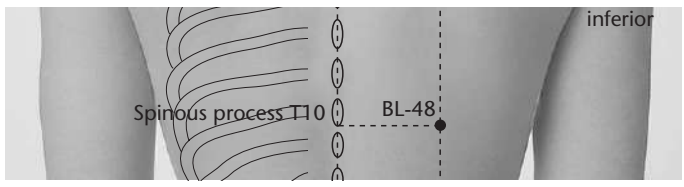

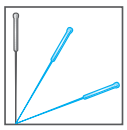
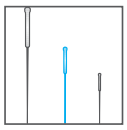
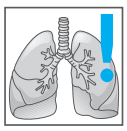
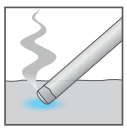
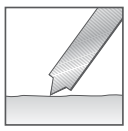
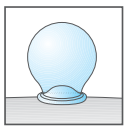

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 270

BL-20 *Spleen* Shu PISHU

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 11th thoracic vertebra (T11).

How to find

In a seated position with the arms hanging down, the spinous process of T7 is approximately on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down four spinous processes to the spinous process of T11. **BL-20** is located on the level of the lower border of its spinous process and 1.5 cun lateral to the midline at the highest point of the paraspinal musculature. **Or:** Locate the articulation of the most inferior rib (T12) and palpate in a superior direction to T11. **Or:** Use the lumbar spine for reference (→ 3.4.3).

Located on the same level are → **Du-6** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and **BL-49** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine. In clinical practice, this point is often needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape). Caution: Pneumothorax.

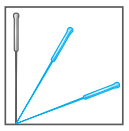

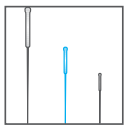

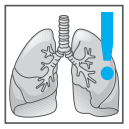

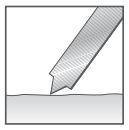

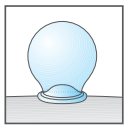

Actions/Indications

- Tonifies the Spleen Qi and Yang, regulates Qi in the Middle, raises Qi
- Transforms Dampness
- Holds and nourishes Blood

Special features

Back-*shu* point of the spleen. Major point for tonifying the Middle, especially for Spleen deficiency syndromes. Use moxibustion or tonifying needle techniques for nourishing the Blood.

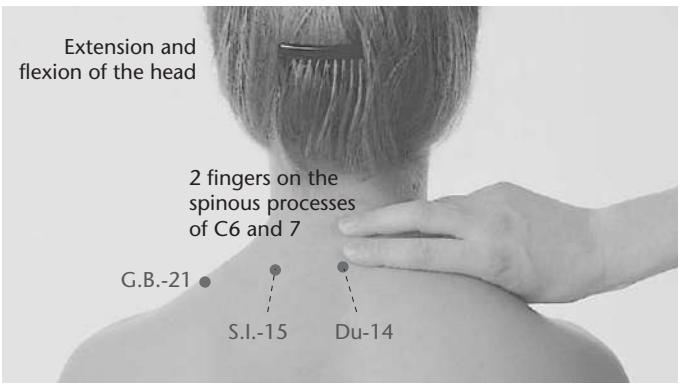

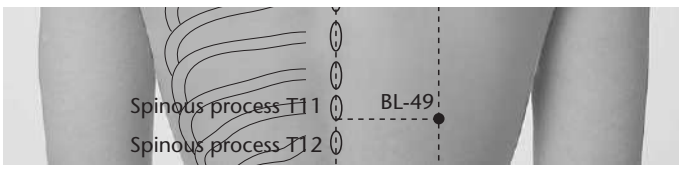

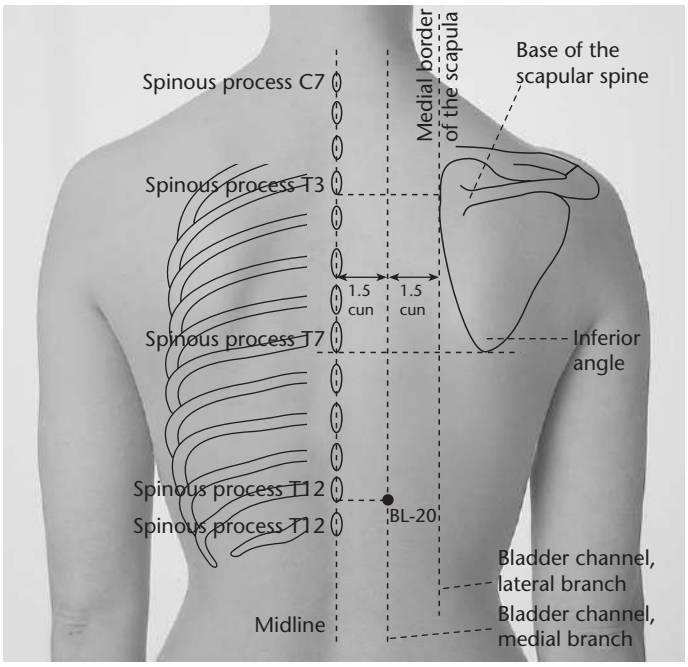

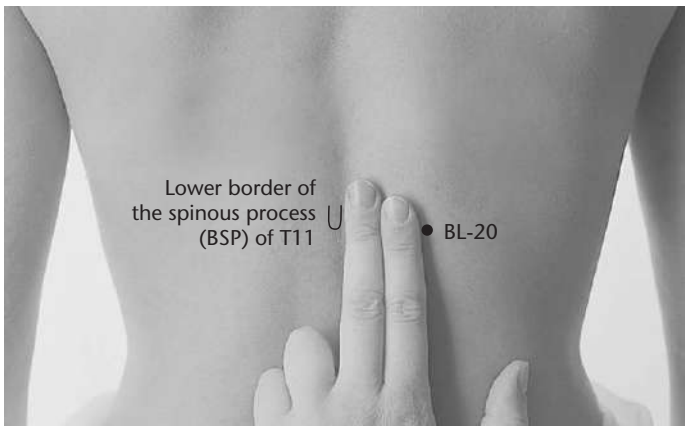

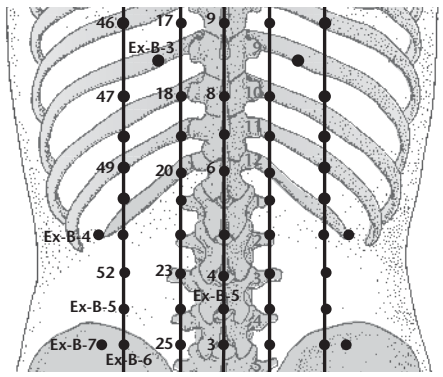

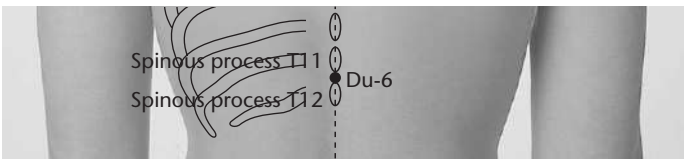

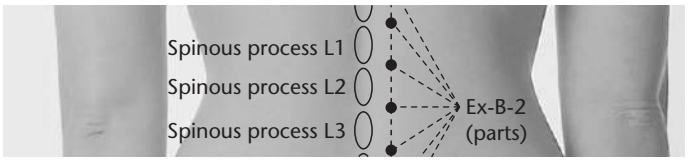

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 271

Stomach Shu WEISHU BL-21

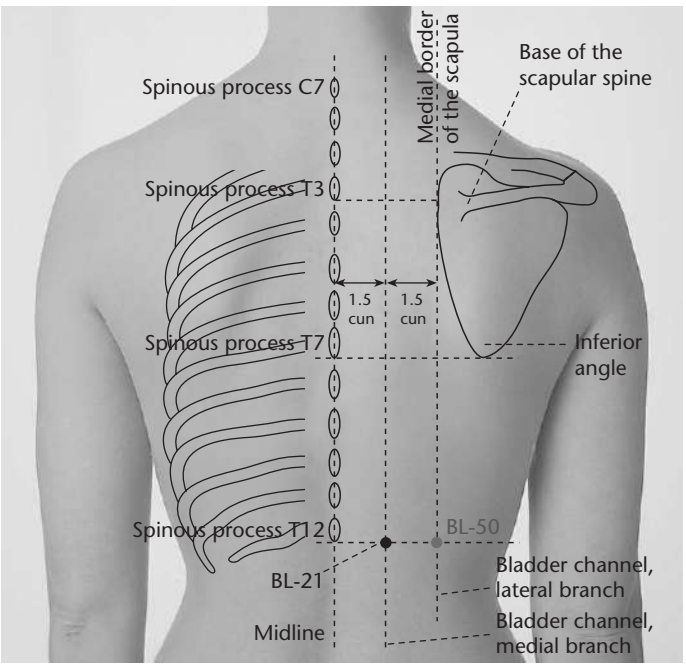

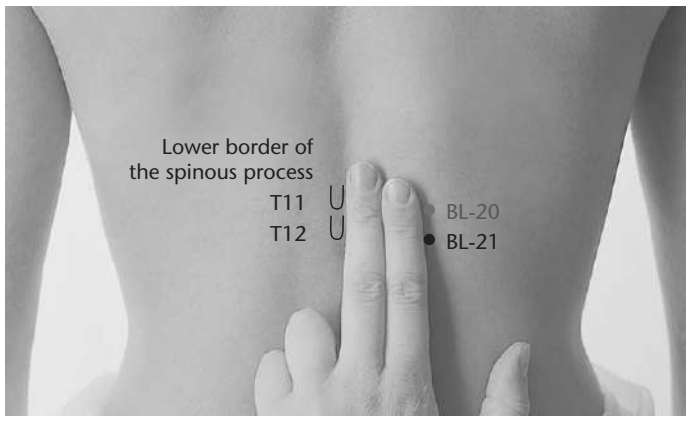

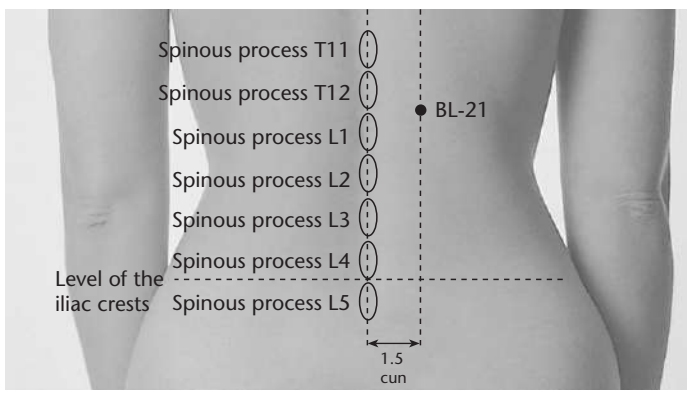

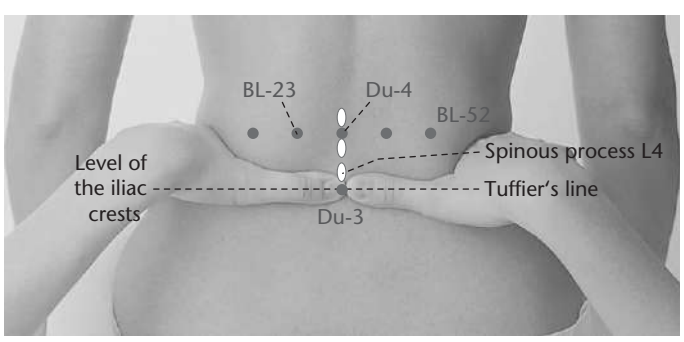

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 12th thoracic vertebra (T12).

How to find

In a seated position with the arms hanging down, the spinous process of T7 is approximately on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down five spinous processes to the spinous process of T12. **BL-21** is located on the level of the lower border of its spinous process and 1.5 cun lateral to the midline at the highest point of the paraspinal musculature. **Or:** Palpate for the articulation of the most inferior rib with T12 and locate **BL-21** 1.5 cun lateral to the lower border of its spinous process. **Or:** Use the lumbar spine for reference (→ 3.4.3). Located on the same level are a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-50** (3 cun lateral to the midline).

Needling

0.5–1 cun obliquely towards the spine. In clinical practice, this point is often needled transversely (subcutaneously) towards the spine. This allows the patient to lie in a supine position during needle retention (the needles can be secured with tape). Caution: Pneumothorax.

Actions/Indications

- Regulates the Stomach, descends counterflow Qi, harmonises the Middle Burner, eliminates Dampness and food stagnation
- Opens the channel

Special features

Back-*shu* point of the Stomach, main point for regulating any dysfunction of the Stomach.

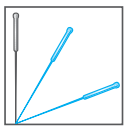
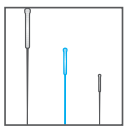
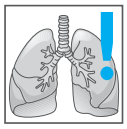
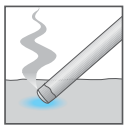
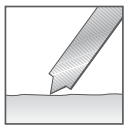
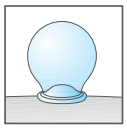

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 272

BL-22 San Jiao Shu SANJIAOSHU

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 1st lumbar vertebra (L1).

How to find

For orientation in the lumbar region (→ 3.4.3), first, determine the Tuffier's line (a line connecting the highest points of the iliac crests), which in most cases intersects with the spinous process of L4 (note: this varies depending on the patient's position → 3.4.3). From there, count upward to the lower border of the spinous process of L1 and, on this level, locate **BL-22** 1.5 cun in a lateral direction, on the highest point of the paraspinal musculature. Located on the same level are → **Du-5** (on the midline), a point of → **Ex-B-2/BL-51/Ex-B-4** (0.5 cun/3 cun/3.5cun lateral of the midline).

Needling

Obliquely 0.5–1.5 cun. Caution: Kidneys.

Actions/Indications

- Regulates the Triple Burner
- Opens the water passages and promotes urination
- Has an effect on the *shaoyang*

Special features

Back-*shu* point of the Triple Burner. Especially regulates the relationship between the Middle and Lower Burner.

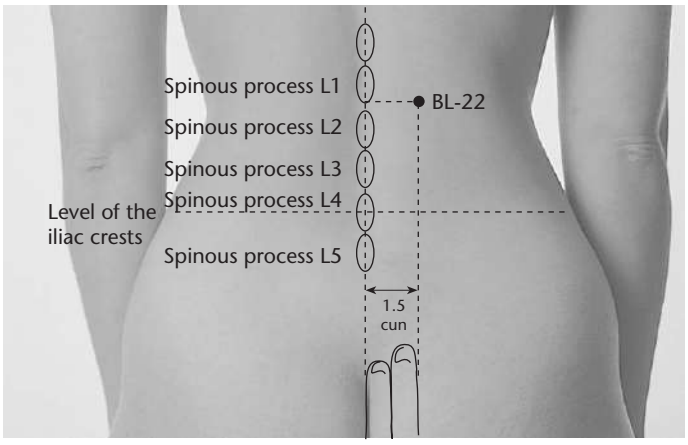

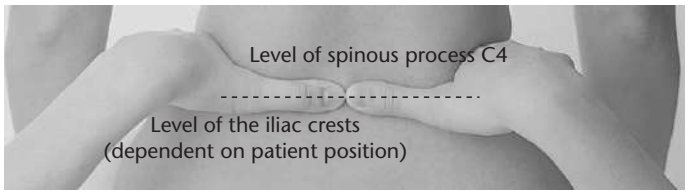

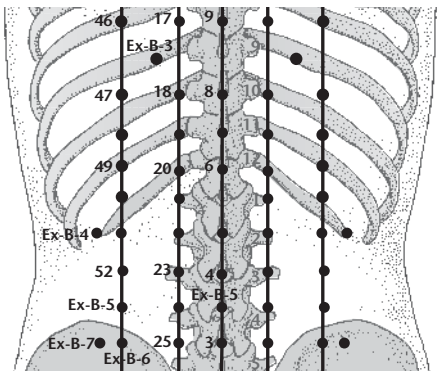

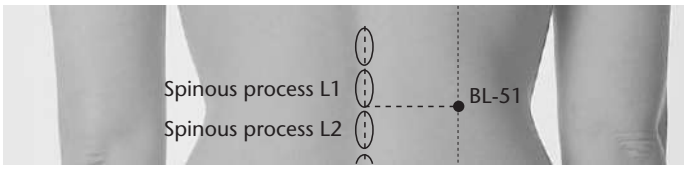

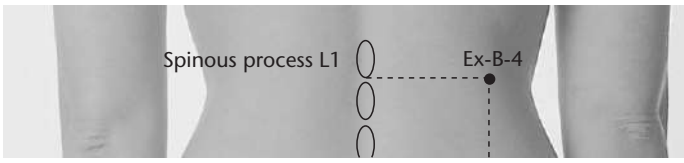

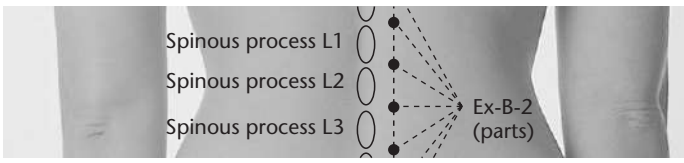

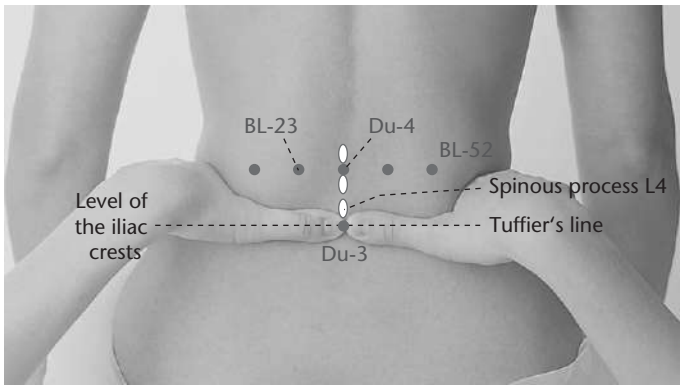

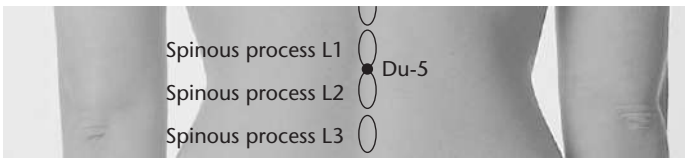

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 273

Kidney Shu SHENSHU BL-23

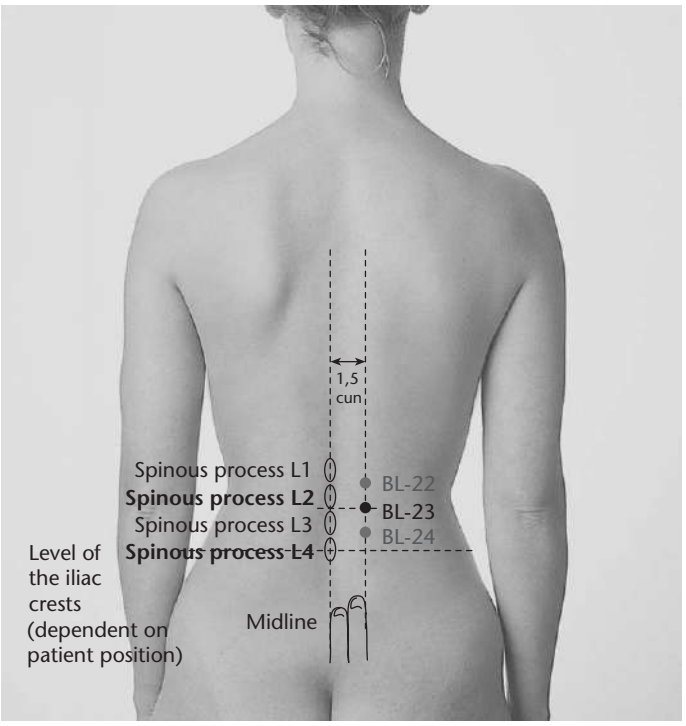

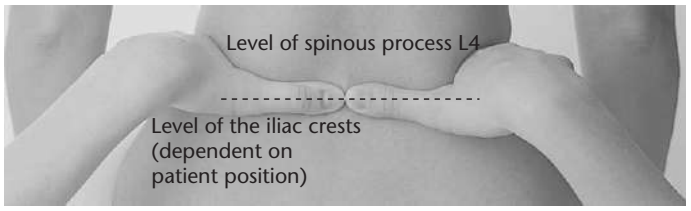

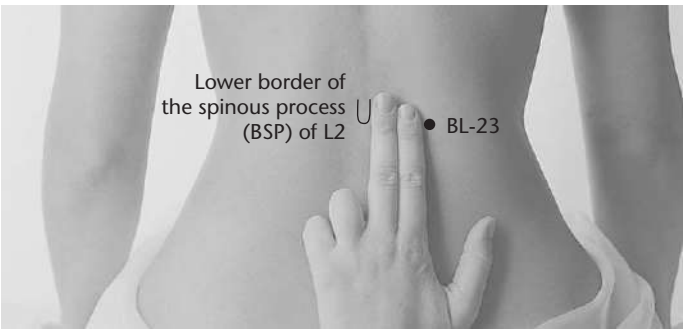

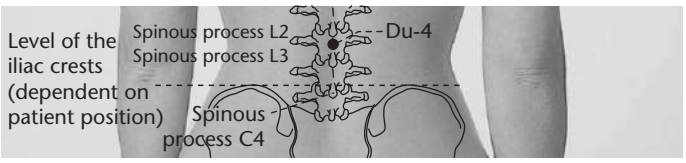

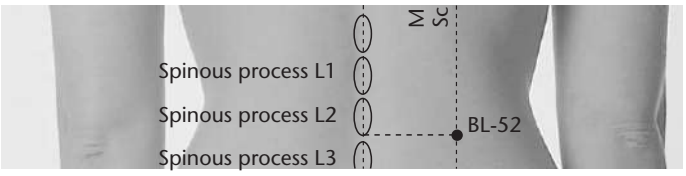

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 2nd lumbar vertebra (L2).

How to find

For orientation in the lumbar region (→ 3.4.3), first, determine the Tuffier's line (a line connecting the highest points of the iliac crests), which in most cases intersects with the spinous process of L4 (note: this varies depending on the patient's position → 3.4.3). From there, count upward to the lower border of the spinous process of L2 and, on this level, locate **BL-23** 1.5 cun in a lateral direction, on the highest point of the paraspinal musculature. Located on the same level are → **Du-4** (on the midline), a point of → **Ex-B-2/BL-52** (0.5 cun/3 cun lateral of the midline).

Needling

Vertically or obliquely 0.5–1.5 cun. Caution: Kidneys.

Actions/Indications

- Strengthens the Kidneys, tonifies the Kidney Qi and Yang, benefits the Essence (*jing*), nourishes the Kidney Yin
- Regulates the Lower Burner, benefits the uterus
- Benefits the bones and the marrow
- Benefits the eyes and ears
- Strengthens the lower back

Special features

Back-*shu* point of the Kidneys, according to some authors, meeting point with the *du mai*. Major point for tonifying the Kidneys. Use tonifying needle techniques for all Kidney deficiency syndromes. Moxibustion tonifies especially the Kidney Yang. **Caution**: Moxibustion is contraindicated for Kidney Yin deficiency with empty Heat (Heat in the Five Hearts (Heat in the thorax, the palms and the soles of the feet), night sweats, restlessness, often a red tongue).

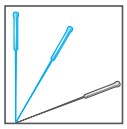
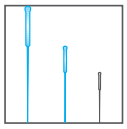

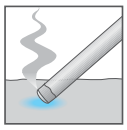
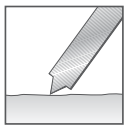
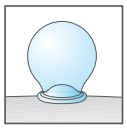

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 274

BL-24 *Sea of Qi* QIHAISHU

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 3rd lumbar vertebra (L3).

How to find

For orientation in the lumbar region (→ 3.4.3), first, determine the Tuffier's line. Place both hands directly superior or lateral to the highest points of the iliac crests, joining the thumbs on the midline at the lumbar spine. In most cases, this line intersects with the spinous process of L4 (note: this varies depending on the patient's position → 3.4.3). From there, count upward to the lower border of the spinous process of L3 and, on this level, locate **BL-24** 1.5 cun in a lateral direction.

Located on the same level are a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **Ex-B-5** (*xiazhishi*; 3 cun lateral to the midline according to the WHO location notes).

Needling

Vertically 0.5–1.5 cun

Actions/Indications

- Benefits the back and strengthens the knees
- Regulates and tonifies Qi and Blood of the Lower Burner

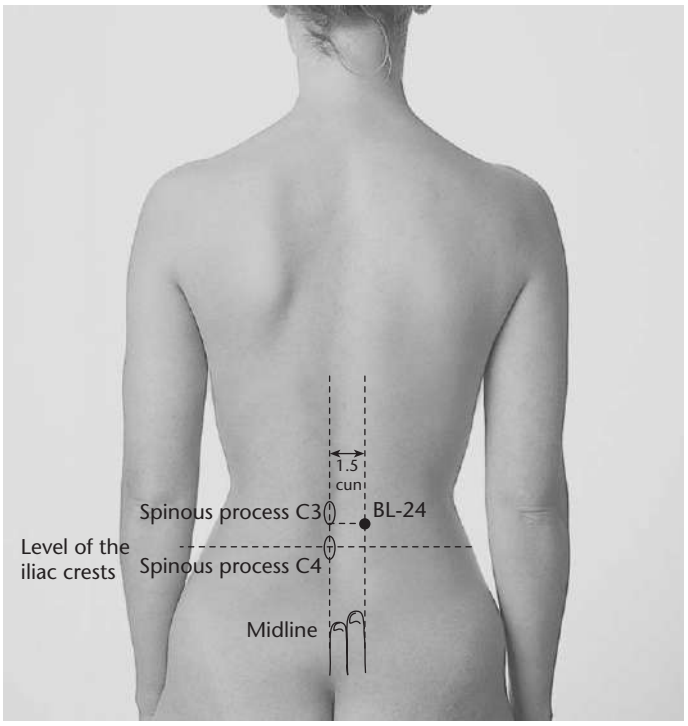

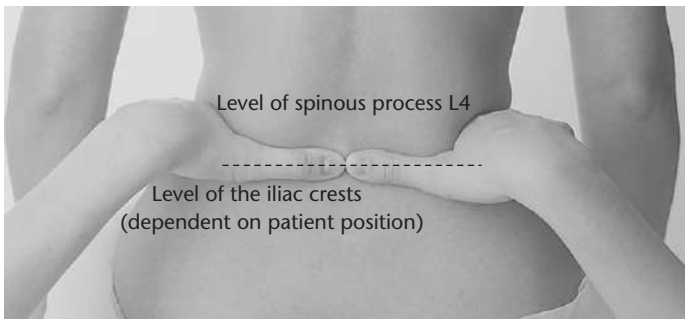

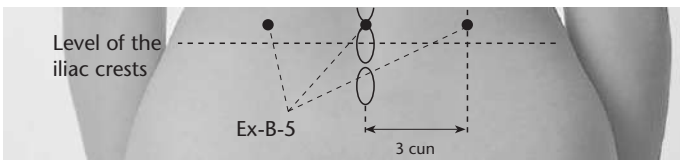

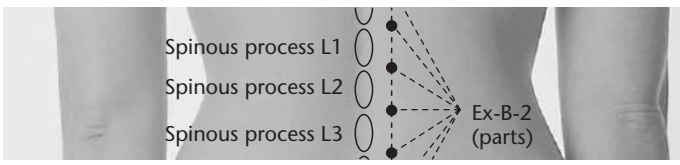

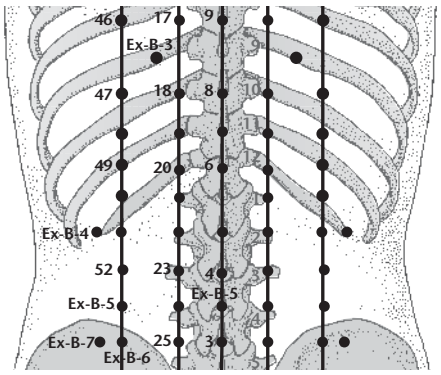

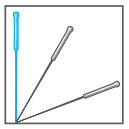
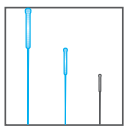

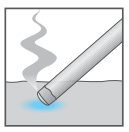
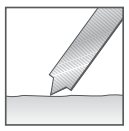
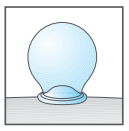

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 275

Large Intestine Shu DACHANGSHU BL-25

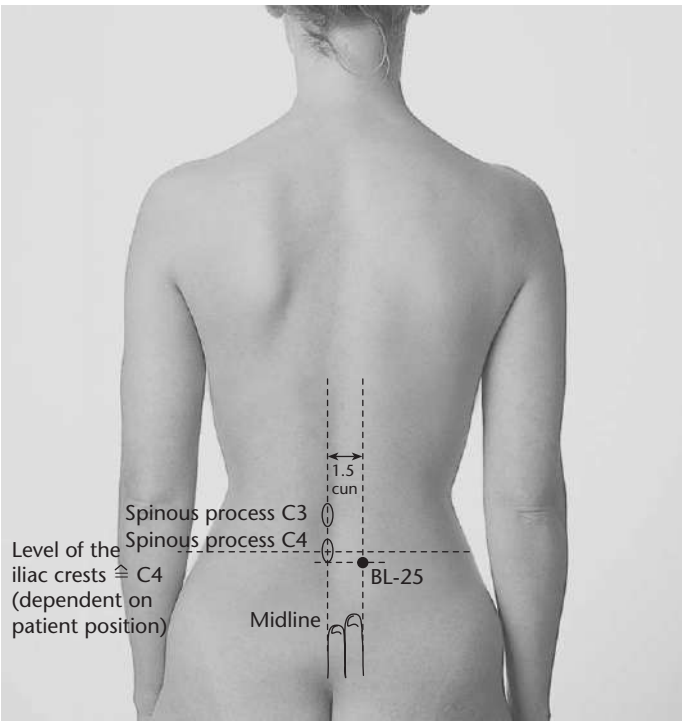

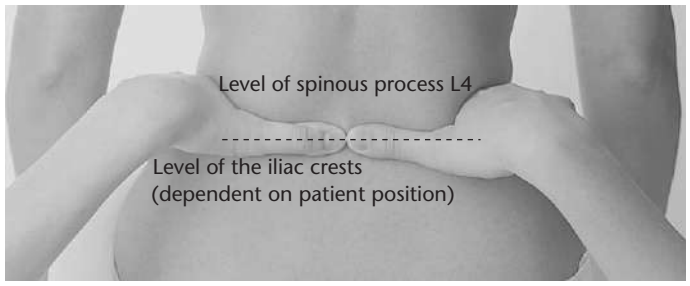

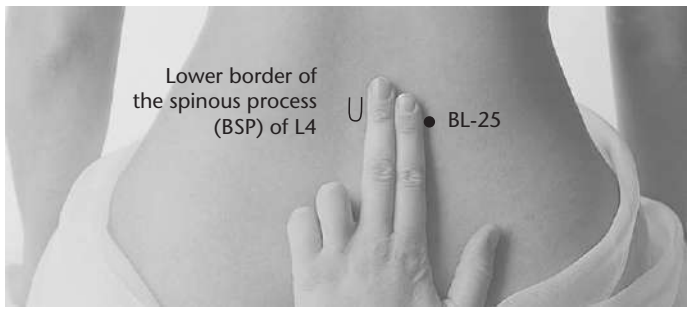

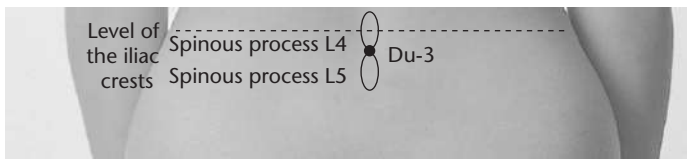

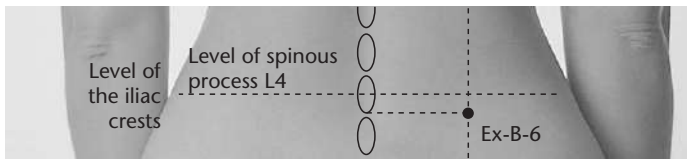

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 4th lumbar vertebra (L4).

How to find

For orientation in the lumbar region (→ 3.4.3), first, determine the Tuffier's line. Place both hands directly superior or lateral to the highest points of the iliac crests, joining the thumbs on the midline at the lumbar spine. In most cases, this line intersects with the spinous process of L4 (note: this varies depending on the patient's position → 3.4.3). On this level, locate **BL-25** 1.5 cun in a lateral direction.

Located on the same level are → **Du-3** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline), → **Ex-B-6** (3 cun lateral to the midline) and → **Ex-B-7** (3.5 cun lateral to the midline).

Needling

Vertically 1–1.5 cun

Actions/Indications

- Regulates the intestines and promotes the Qi flow
- Strengthens the lower back

Special features

Back-*shu* point of the Large Intestine. Important point for regulating the Large Intestine Qi. Important local point for acute and chronic lumbar disorders.

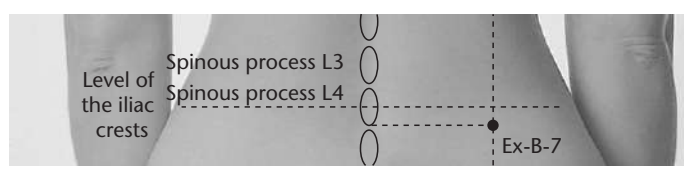

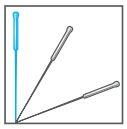
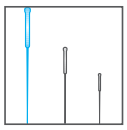

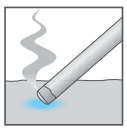
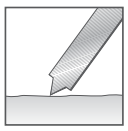
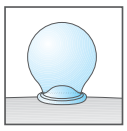

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 276

BL-26 *Gate of Origin* Shu GUANYUANSHU

Location

1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 5th lumbar vertebra (L5).

How to find

For orientation in the lumbar region (→ 3.4.3), first, determine the Tuffier's line. Place both hands directly superior or lateral to the highest points of the iliac crests, joining the thumbs on the midline at the lumbar spine. In most cases, this line intersects with the spinous process of L4 (note: this varies depending on the patient's position; for more detail on orientation in the lumbar/sacral region → 3.4.3, 3.4.4). Next, locate the lower border of the spinous process of L5 and, on this level, locate **BL-26** 1.5 cun in a lateral direction.

Located on the same level are → **Ex-B-8** (on the midline) and a point of → **Ex-B-2** (0.5 cun lateral to the midline).

Needling

Vertically 0.5–1.5 cun

Actions/Indications

- Strengthens the back, especially for Kidney deficiency syndromes
- Regulates the Lower Burner

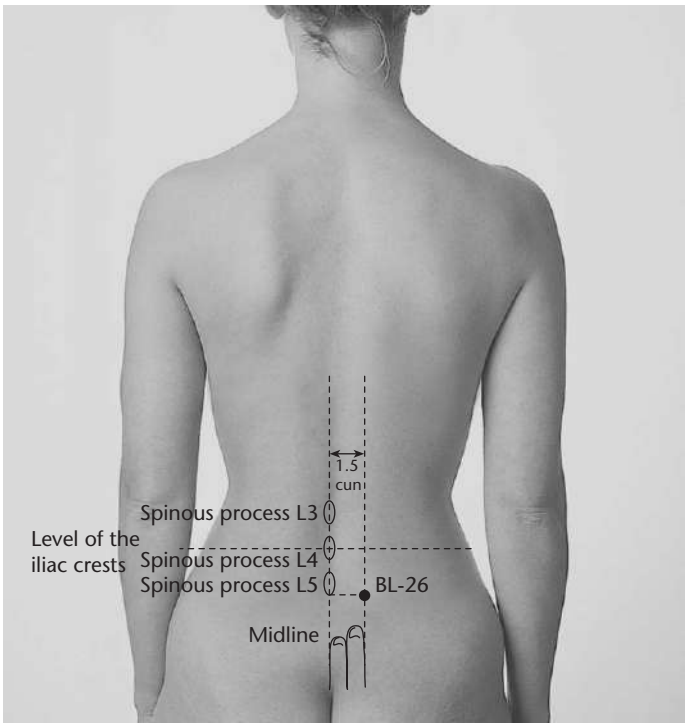

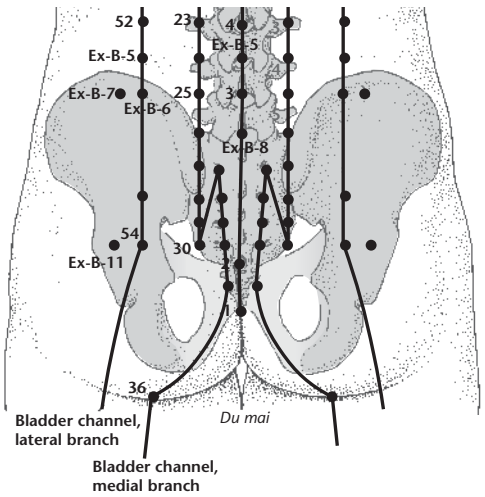

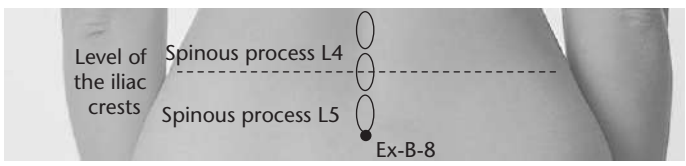

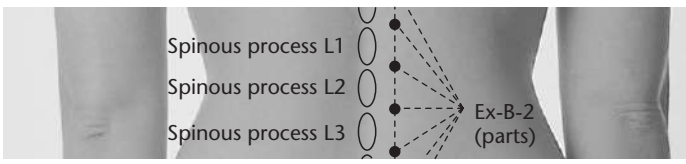

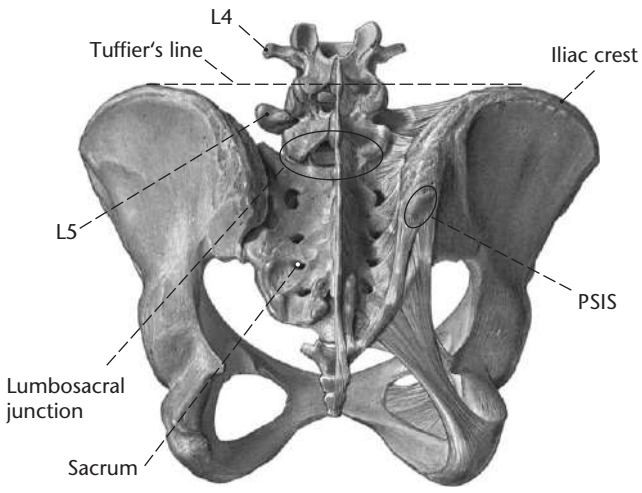

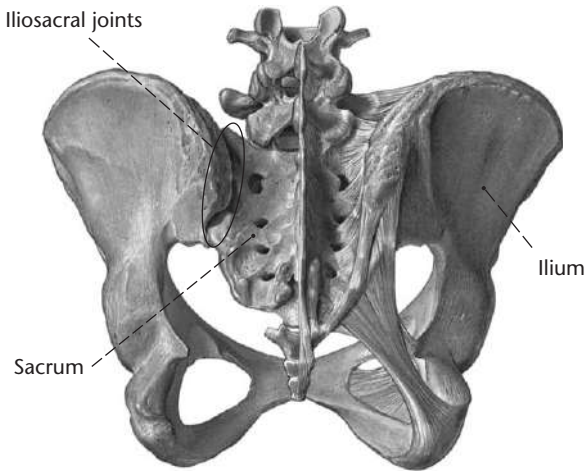

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 277

Small Intestine Shu XIAOCHANGSHU BL-27

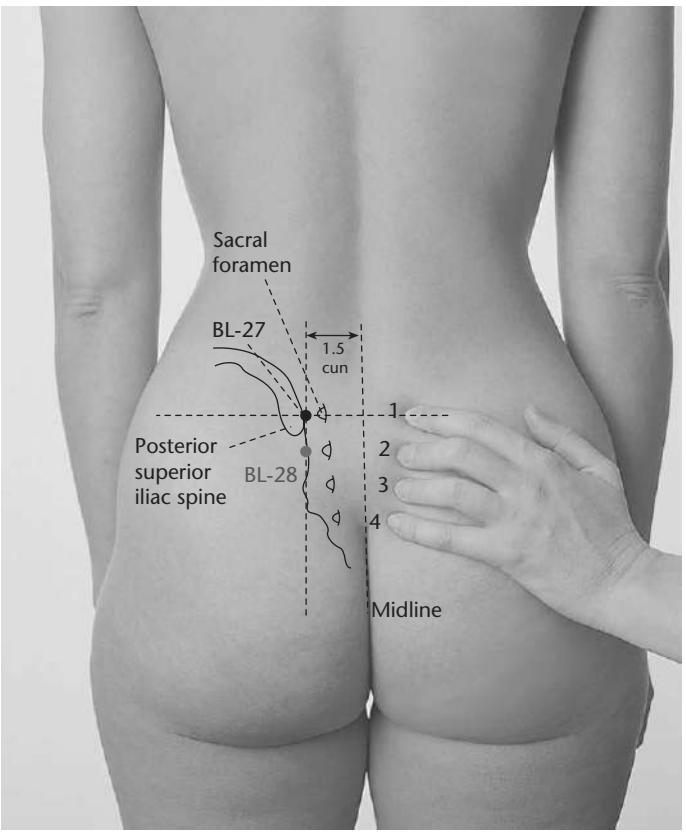

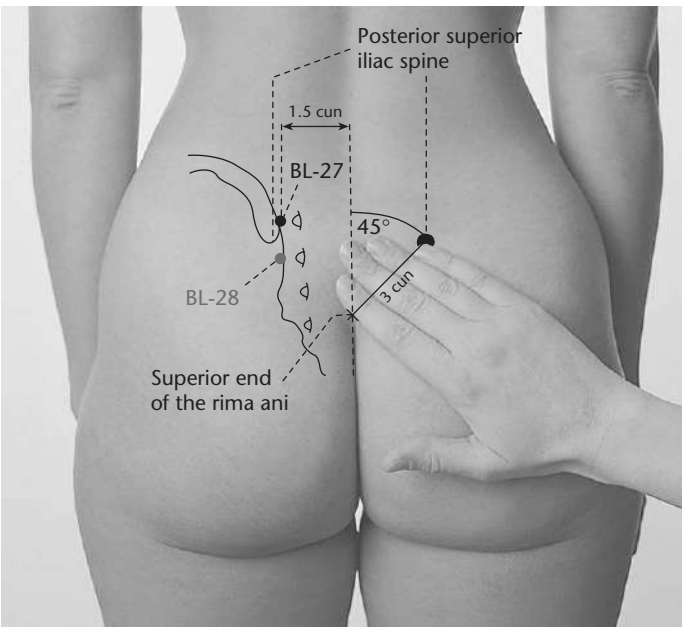

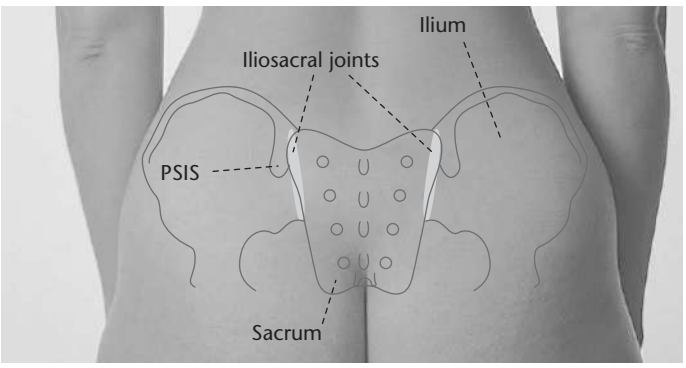

Location

1.5 cun lateral to the posterior midline, on the level of the first sacral foramen.

How to find

Quick method: Locate the posterior superior iliac spine (PSIS → 3.4.3): from the rima ani, palpate 3 cun (1 handbreadth) in a superior and lateral direction at a 45° angle, until you can feel a distinct bony ridge (often superficially visible by a dimple). **BL-27** is located slightly superiorly and medially to the PSIS, 1.5 cun lateral to the midline, on the level of the first sacral foramen. For more detail on orientation in the lumbar and sacral region → 3.4.3, 3.4.4.

Located on the same level are → **BL-31** (over the 1st sacral foramen) and → **BL-28** (slightly below and medial to the PSIS at the level of the 2nd sacral foramen).

Needling

Vertically 0.5–1 cun

Actions/Indications

- Tonifies the Small Intestine Qi
- Drains Dampness and Damp-Heat
- Regulates the water passages

Special features

Back-*shu* point of the Small Intestine

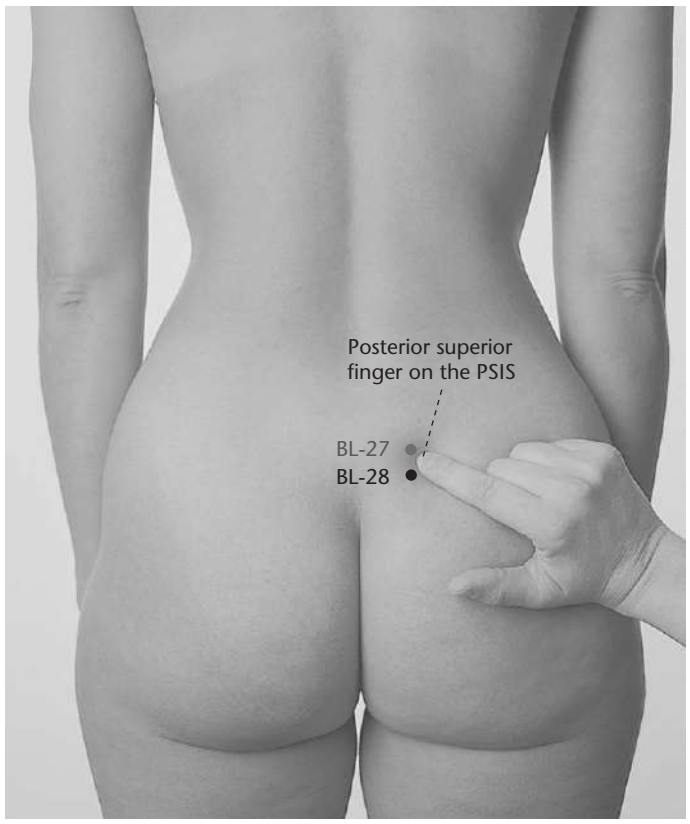

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 278

BL-28 *Bladder* Shu PANGGUANGSHU

Location

1.5 cun lateral to the posterior midline, on the level of the second sacral foramen.

How to find

Quick method: Locate the posterior superior iliac spine (PSIS → 3.4.3): from the rima ani, palpate 3 cun (1 handbreadth) in a superior and lateral direction at a 45° angle, until you can feel a distinct bony ridge (often superficially visible by a dimple). **BL-28** is located slightly inferiorly and medially to the PSIS, 1.5 cun lateral to the midline on the level of the 2nd sacral foramen. For more detail on orientation in the lumbar and sacral region → 3.4.3, 3.4.4.

Located on the same level are → **BL-32** (over the 2nd sacral foramen) and → **BL-53** (3 cun lateral to the midline). → **BL-27** is located slightly superiorly and medially to the PSIS on the level of the 1st sacral foramen.

Needling

Vertically 0.5–1.5 cun. For local disorders, slightly oblique insertion towards the iliosacral joint.

Actions/Indications

- Regulates the Bladder and the Lower Burner, clears Damp-Heat from the Lower Burner, removes stagnation and transforms masses
- Benefits the lumbar region and the legs

Special features

Back-*shu* point of the Bladder

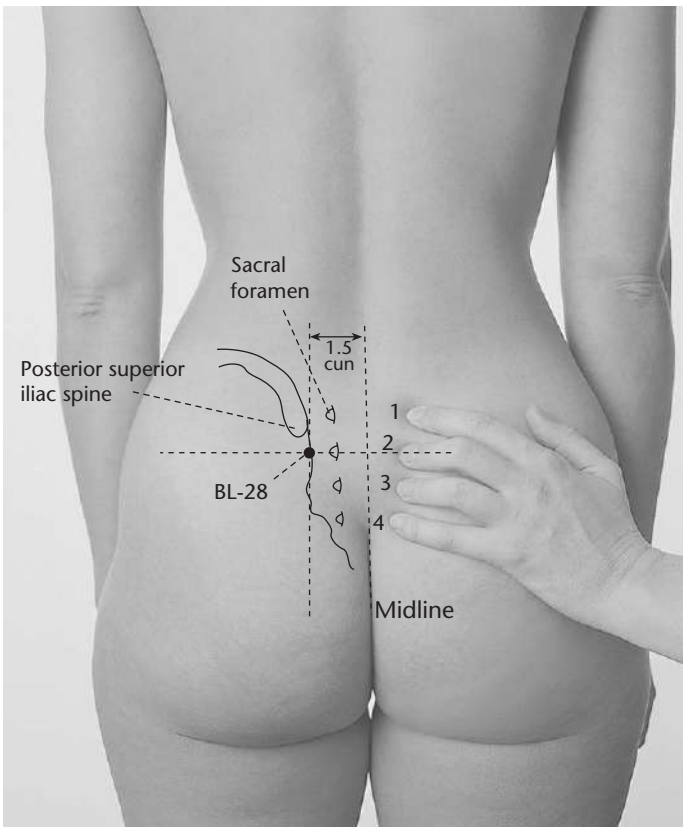

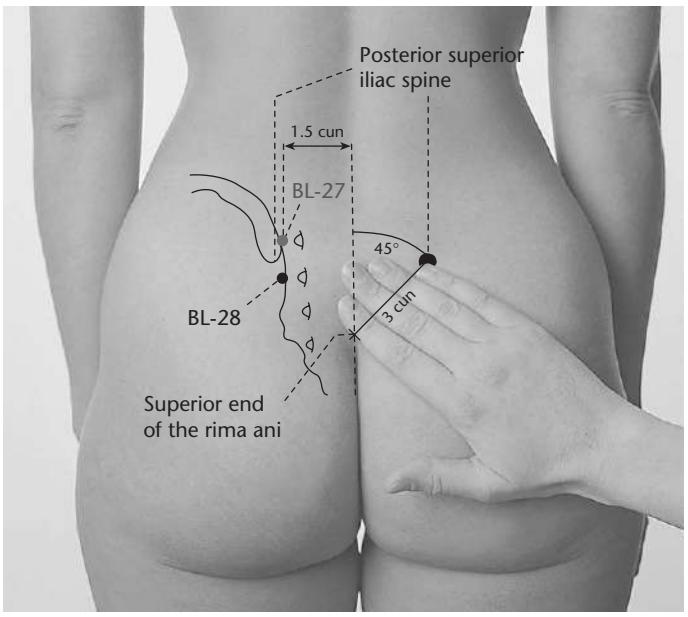

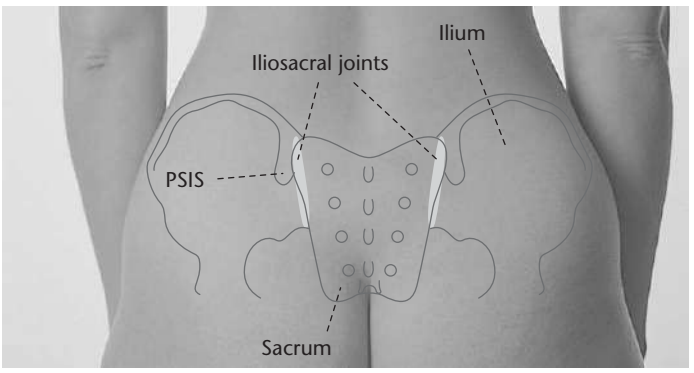

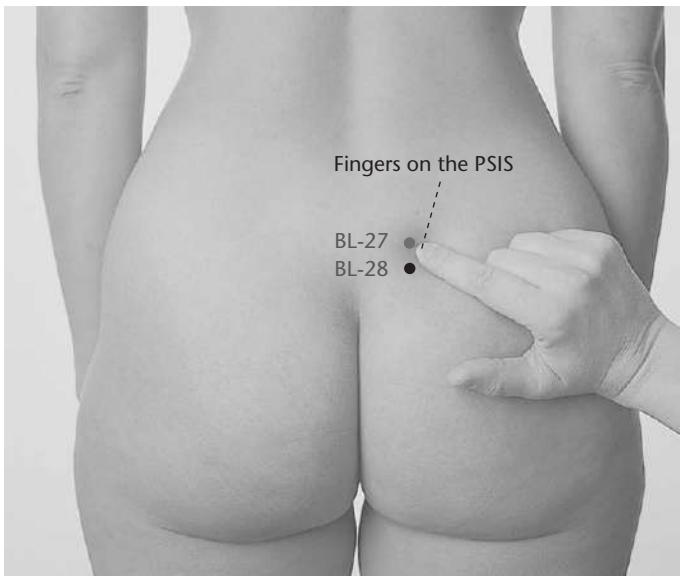

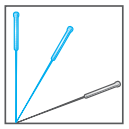
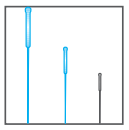

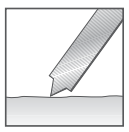
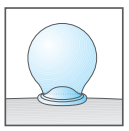

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 279

Mid Spine Shu ZHONGLÜSHU BL-29

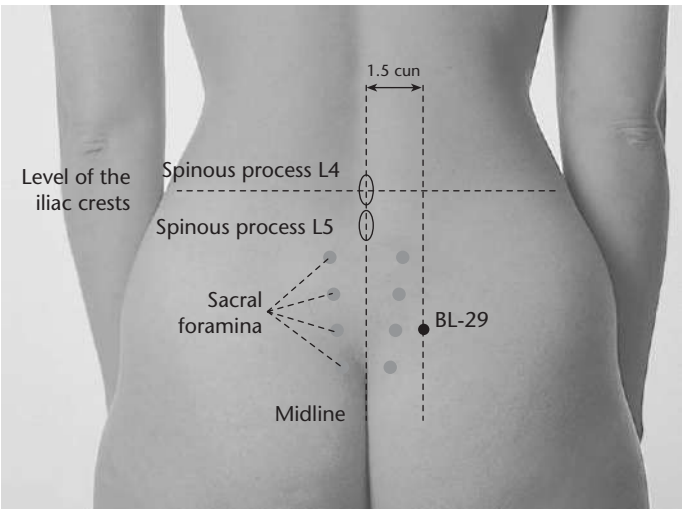

Location

1.5 cun lateral to the posterior midline, on the level of the 3rd sacral foramen.

How to find

For more detail on orientation in the lumbar and sacral region, see ➞ 3.4.3, 3.4.4. First of all find L5. The next palpable structure in an inferior direction is the most superior sacral spinous process. Palpate in an inferior direction for two further sacral processes and locate the 3rd sacral foramen in a depression less than 1 cun lateral to the 3rd process. **BL-29** is located on the same level (3rd sacral foramen) and 1.5 cun lateral to the midline. Also located on this level is ➞ **BL-33** (over the 3rd sacral foramen).

Needling

Vertically 0.5–1.5 cun

Actions/Indications

- Strengthens the lower back and the Kidneys
- Expels Cold
- Regulates the Lower Burner and stops diarrhoea

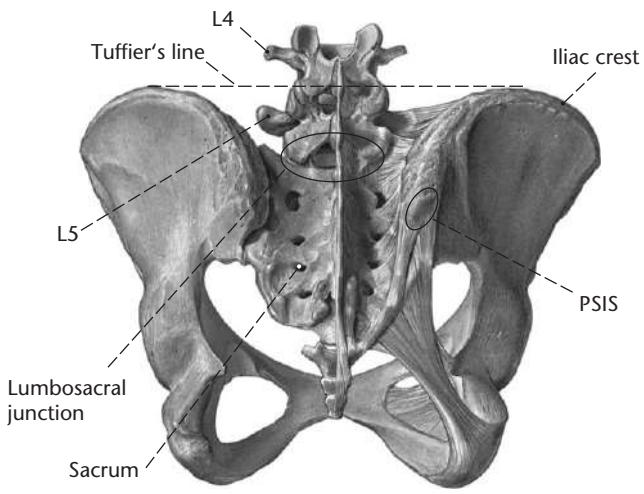

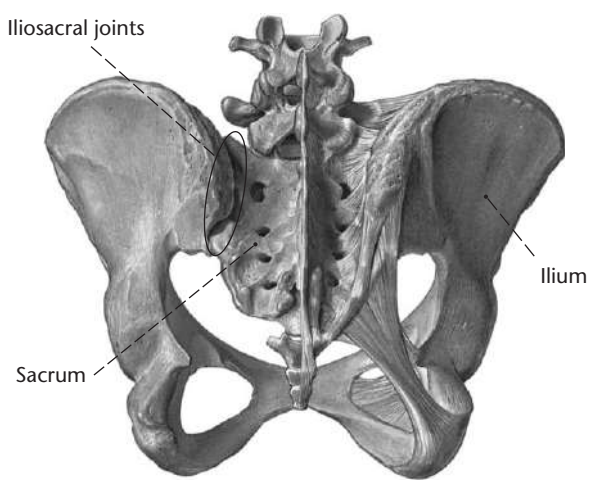

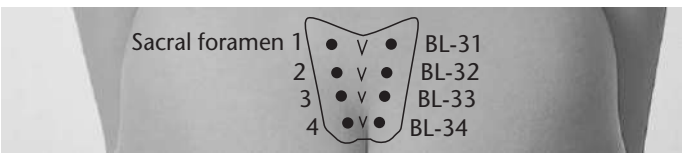

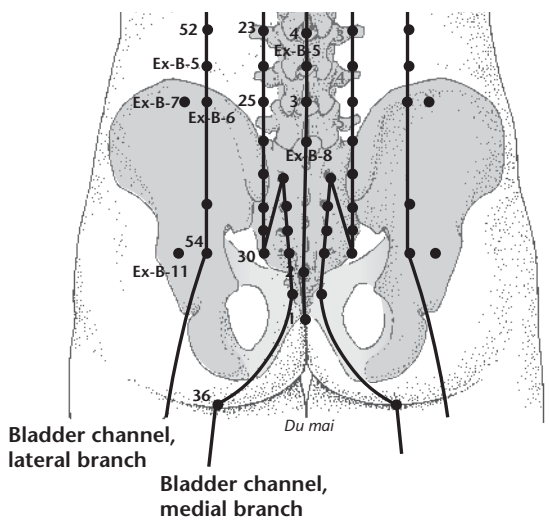

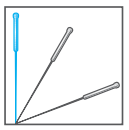
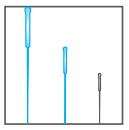

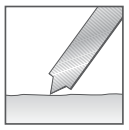

Ch04.7a-F10028.qxd 2/23/08 3:54 PM Page 280

BL-30 *White Ring* Shu BAI HUAN SHU

Location

1.5 cun lateral to the posterior midline, on the level of the 4th sacral foramen.

How to find

For more detail on orientation in the lumbar and sacral region, see ➞ 3.4.3, 3.4.4. First, find L5. The next palpable structure in an inferior direction is the most superior sacral spinous process. Palpate in an inferior direction for three further sacral processes and locate the 4th sacral foramen in a depression less than 1 cun lateral to the 4th process. **BL-30** is located on the level of the 4th foramen, 1.5 cun lateral to the midline.
Located on the same level are ➞ **BL-34** (over the 4th sacral foramen) and ➞ **BL-54** (3 cun lateral to the midline).

Needling

Vertically 0.5–1.5 cun

Actions/Indications

- Strengthens the lower back and the knees
- Eliminates Damp-Heat from the Lower Burner
- Regulates menstruation

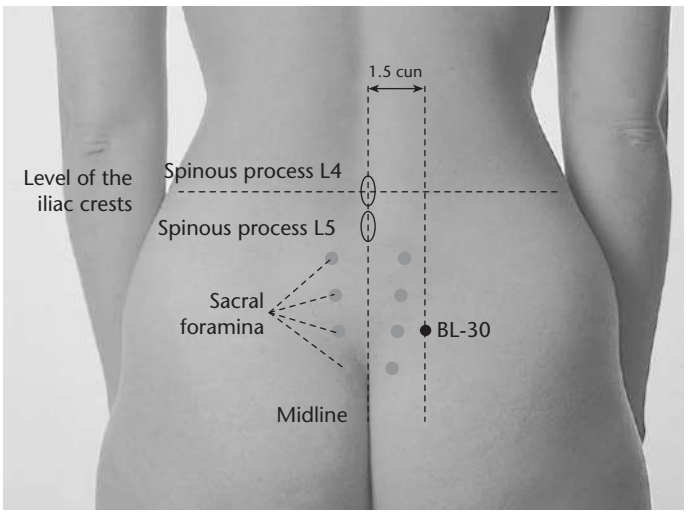

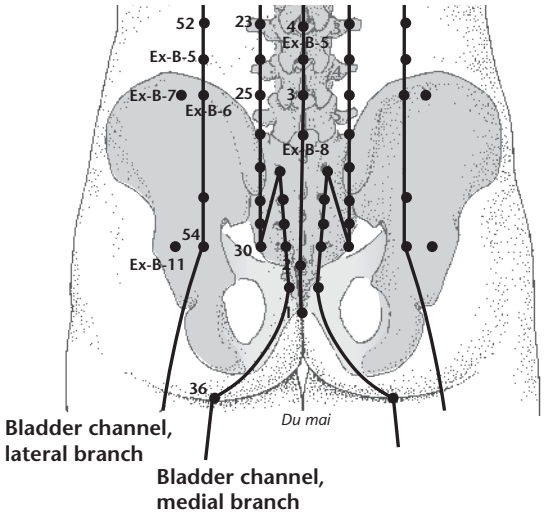

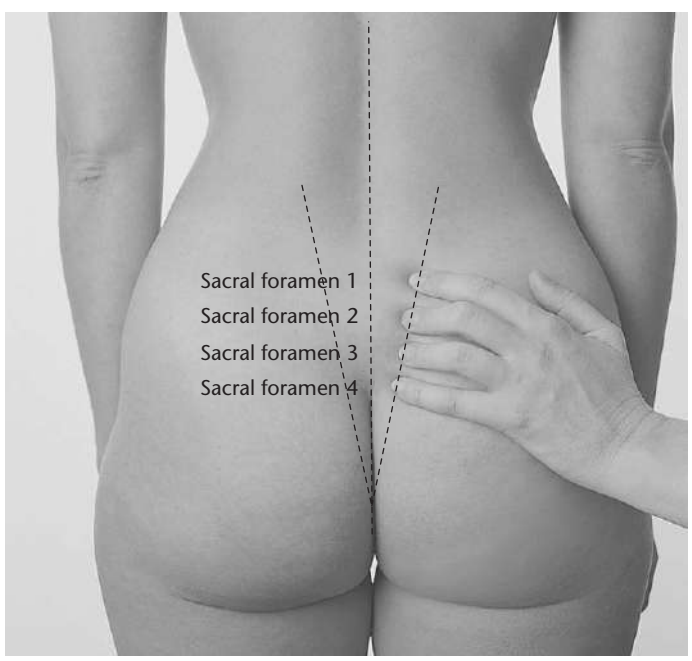

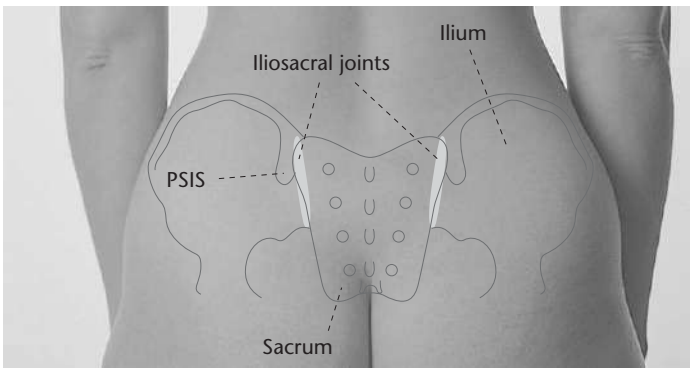

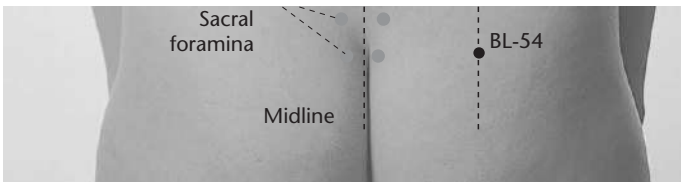

Ch04.7a-F10028.qxd 2/23/08 3:55 PM Page 281

Eight Liao BALIAO BL 31 – BL 34

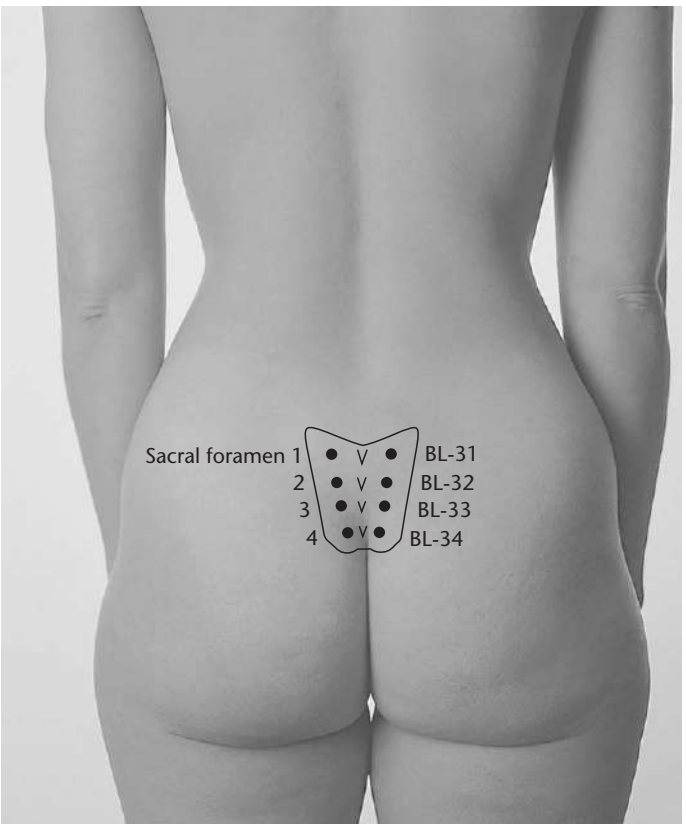

These points are described together because of their similar actions and indications.

BL-31 (*shangliao*) 'Upper Crevice', 1st sacral foramen
BL-32 (*ciliao*) 'Second Crevice', 2nd sacral foramen
BL-33 (*zhongliao*) 'Middle Crevice', 3rd sacral foramen
BL-34 (*xialiao*) 'Lower Crevice', 4th sacral foramen

Location

BL-31: Over the 1st sacral foramen
BL-32: Over the 2nd sacral foramen
BL-33: Over the 3rd sacral foramen
BL-34: Over the 4th sacral foramen

How to find

For more detail on orientation in the lumbar and sacral region, ➞ 3.4.3, 3.4.4. First, locate L5. The next palpable structure in an inferior direction is the most superior sacral spinous process. Palpate in a lateral direction on the level of the sacral processes for the depressions of the sacral foramina. **Or:** Locate the 4th sacral foramen on the level of the superior end of the rima ani (sometimes a bit more superiorly or inferiorly). Place the little finger on the 4th foramen, then the ring, middle and index fingers on a slightly V-shaped line in a superior direction, palpating for the depressions of the sacral foramina. With the fingers in this position, the index finger will be resting on the 1st foramen (**BL-31**), the middle finger on the 2nd foramen (**BL-32**), the ring finger on the 3rd foramen (**BL-33**) and the little finger on the 4th foramen (**BL-34**). Located on the same level as **BL-31–BL-34** are the points ➞ **BL-27–BL-30** on the inner branch of the BL channel (1.5 cun lateral to the midline) as well as ➞ **BL-53** (on the level of **BL-32**) and ➞ **BL-54** (on the level of **BL-34**), both 3 cun lateral to the midline. Also on the level of **BL-34** is ➞ **Ex-B-11** (3.5 cun lateral to the midline).

Needling

Vertically 0.7–1.5 cun. Caution: Pregnancy (especially reducing needling techniques are contraindicated); exception: to promote labour.

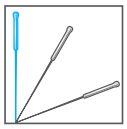

Actions/Indications

- Regulates the Lower Burner, promotes urination (all eight points) (**BL-32** and **BL-33** have the strongest action in cases of urinary disorders, while **BL-34** has the strongest action in cases of genital disorders)
- Strengthens the Kidneys and Essence (*jing*)
- Benefits the Intestines (**BL-34** has the strongest action in cases of intestinal disorders)
- Promotes labour (for example, use **BL-32** with electro-acupuncture)
- Benefits the lumbar region (**BL-32** and **BL-33** have the strongest action)

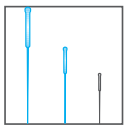

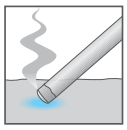

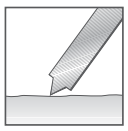

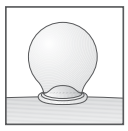

Special features

According to some authors, meeting point with the G.B. channel. Clinically, **BL-32** is the most commonly used point owing to its widest range of indications. Electro-acupuncture at **BL-32** is useful for alleviating pain during labour: after needling this point, bend the handle of the needle in a superior direction at a 90° angle and tape it to the skin. Now attach the electro-stimulation. This method allows the woman to lie down if desired.
Moxibustion of the *baliao* is indicated for difficult urination or defaecation due to Kidney deficiency in elderly people.

Ch04.7a-F10028.qxd 2/23/08 3:55 PM Page 282

BL-35 *Meeting of Yang* HUIYANG

Location

0.5 cun lateral to the posterior midline, on the level of the tip of the coccyx.

How to find

Locate the coccyx superior to the anus. In contrast to the sacrum, the coccyx is movable. **BL-35** is located 0.5 cun lateral to the midline on the level of the tip of the coccyx.

→ **Du-2** is located on the midline, on the level of the sacral hiatus (→ 3.4.4) and superior to **BL-35** and the tip of the coccyx.

Needling

Vertically 1–1.5 cun

Actions/Indications

- Clears Damp-Heat from the Lower Burner
- Treats haemorrhoids and rectal prolapse
- Benefits the coccyx

Special features

Especially in patients with a weakened immune system, this point should be needled only if strictly necessary and after using a disinfecting wipe three times.

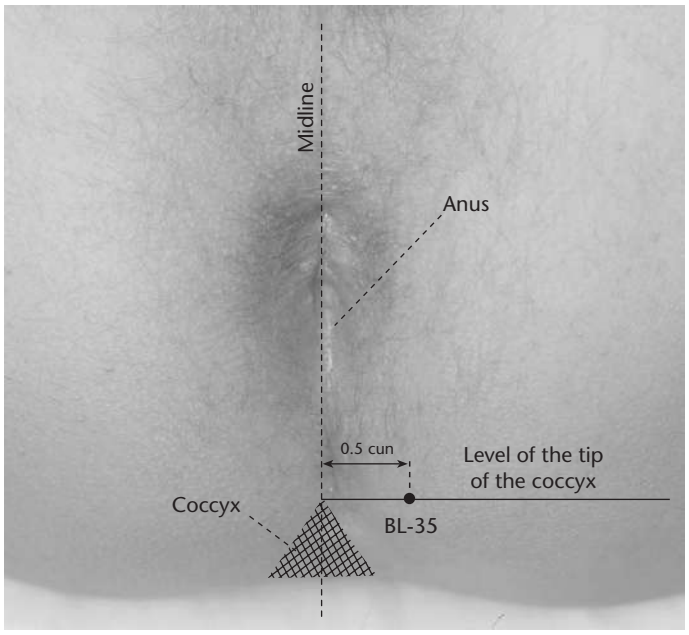

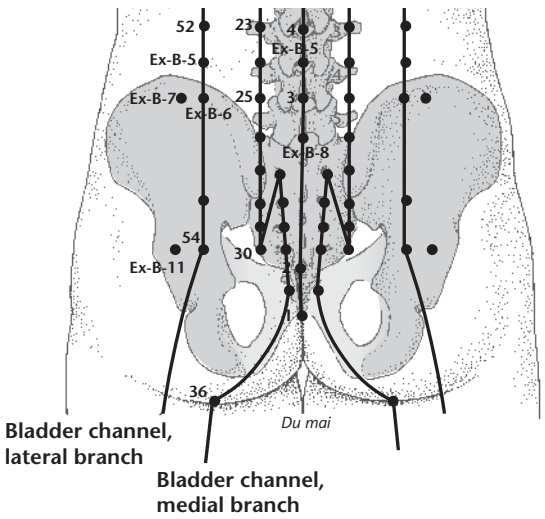

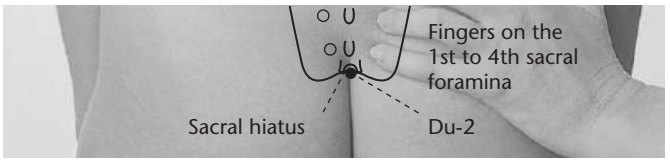

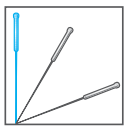
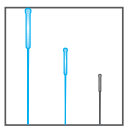

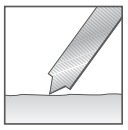

Ch04.7a-F10028.qxd 2/23/08 3:55 PM Page 283

Hold and Support CHENGFU BL-36

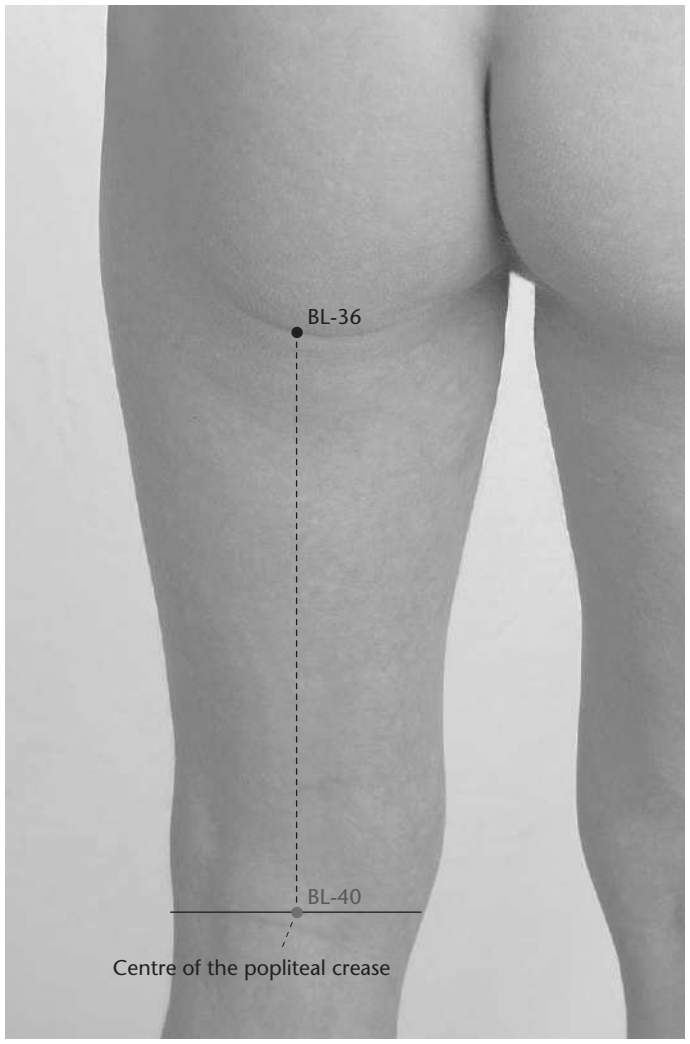

Location

In the gluteal crease, superior to the midpoint of the popliteal crease (→ **BL-40**).

How to find

Locate the midpoint of the popliteal crease (in obese patients, use bony/muscular structures to find the midpoint). **BL-36** is located directly vertically above the midpoint of the popliteal crease, in the gluteal crease at the transition between the buttocks and the posterior aspect of the thigh (palpate for a depression).

Needling

Vertically 1–2 cun

Actions/Indications

- Opens the channel, alleviates pain
- Relaxes the tendons
- Regulates the Lower Burner and treats haemorrhoids

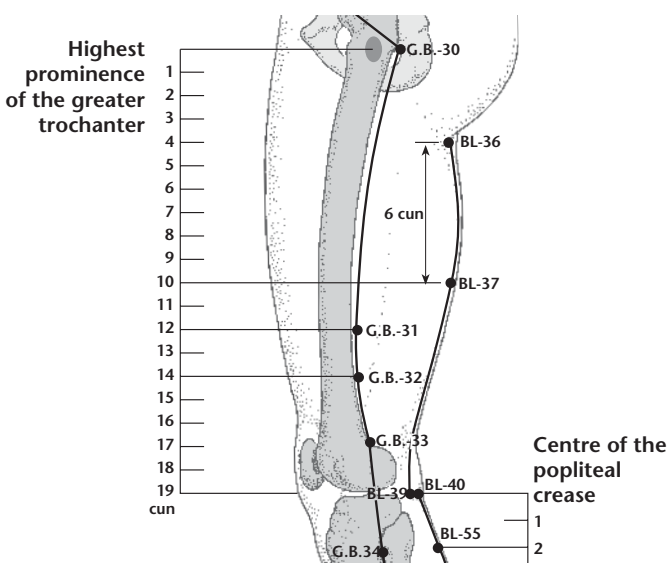

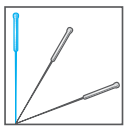
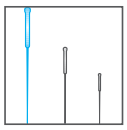

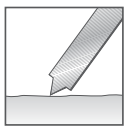
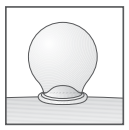

Ch04.7a-F10028.qxd 2/23/08 3:55 PM Page 284

BL-37 *Gate of Abundance* YINMEN

Location

6 cun distal to → **BL-36** (gluteal crease), on a line connecting → **BL-36** and → **BL-40** (in the popliteal crease), in a gap in the musculature.

How to find

First, locate → **BL-36** in the centre of the gluteal crease and, from there, palpate 6 cun along the posterior aspect of the thigh in an inferior direction towards the centre of the popliteal crease (→ **BL-40**). This is the location of **BL-37**, which in slim patients can be located where the long head of the biceps femoris muscle and the semitendinosus muscle converge (usually 8 cun superior to the centre of the popliteal crease. **Or:** Spreading hands technique (→ 2.3.3 or elastic tape): Place the little fingers on the centre of the popliteal crease (→ **BL-40**) and the centre of the gluteal crease (→ **BL-36**). Locate **BL-37** 1–2 cun proximal to the midpoint of the distance, in a depression between the two muscle bellies.

Needling

Vertically 0.5–2 cun

Actions/Indications

- Relaxes the tendons
- Opens the channel and *luo* vessels
- Benefits the lower back

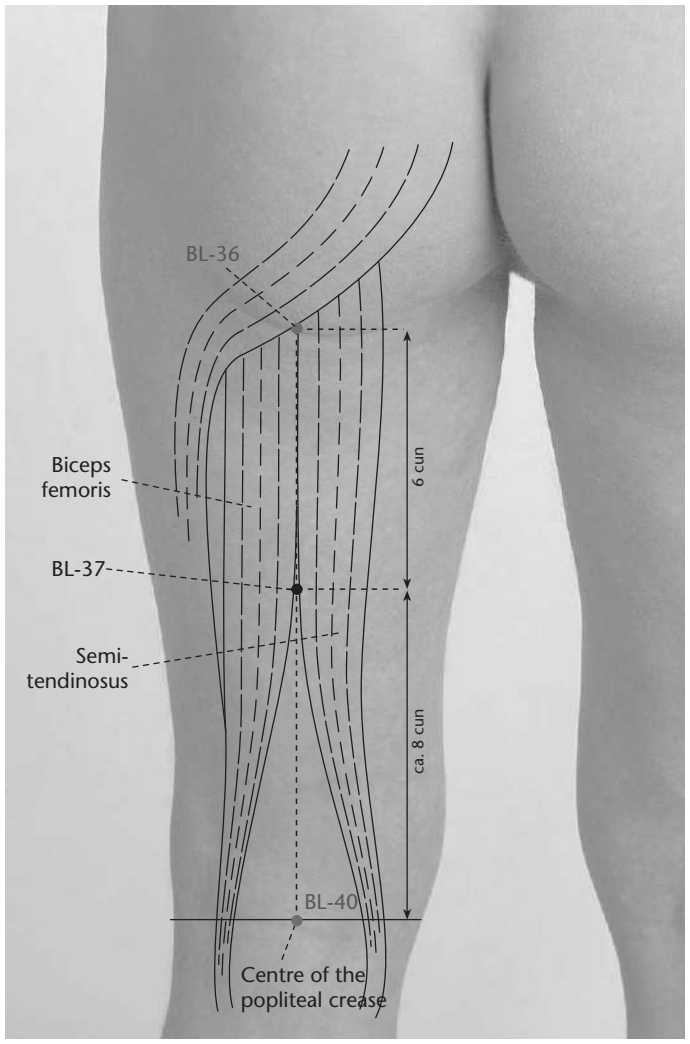

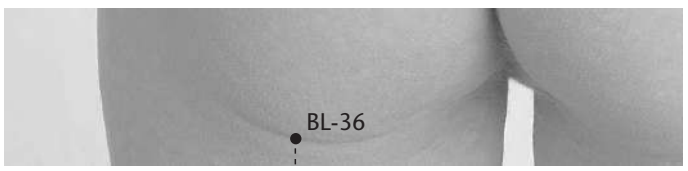

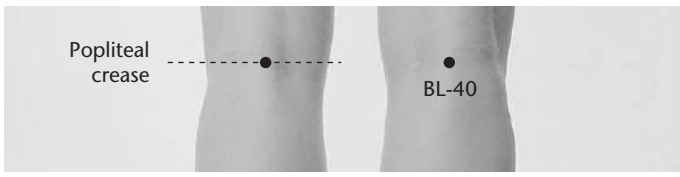

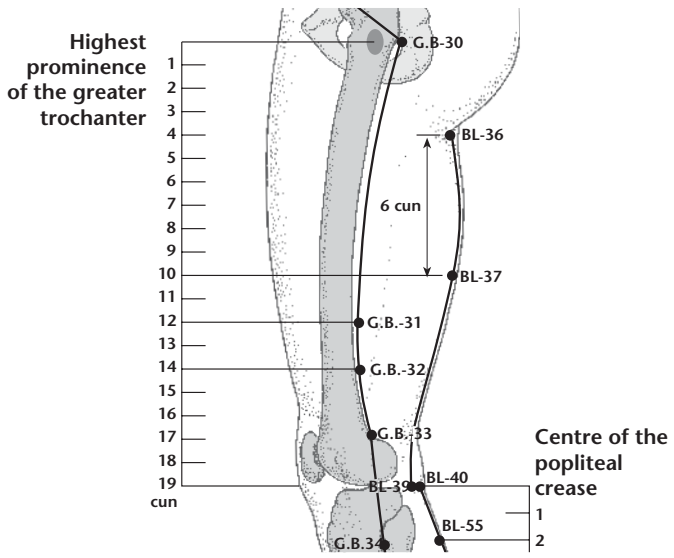

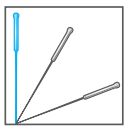
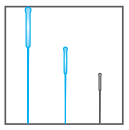

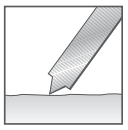

Ch04.7a-F10028.qxd 2/23/08 3:55 PM Page 285

Floating Cleft FUXI BL-38

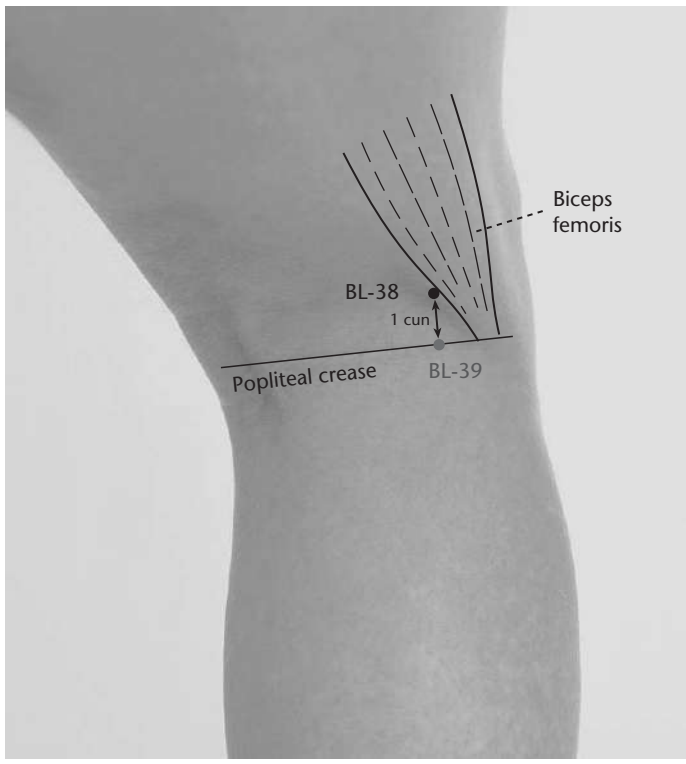

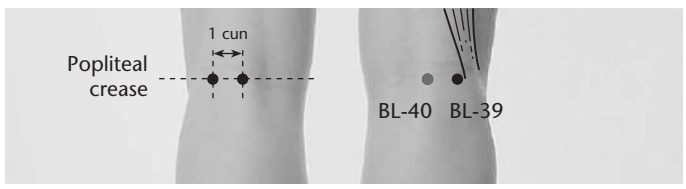

Location

On the laterodorsal aspect of the knee, 1 cun superior and lateral to the centre of the popliteal crease (→ **BL-40**), medial to the biceps femoris muscle, or 1 cun proximal to → **BL-39**.

How to find

To locate the popliteal crease, ask the patient to flex their knee slightly and palpate for the knee joint space. By flexing the knee against resistance, the tendon of the biceps femoris muscle will become more pronounced and visible. Locate **BL-38** 1 cun superior to the knee joint on the medial border of the tendon.
→ **BL-39** is located 1 cun distal to **BL-38** on the level of the popliteal crease.

Needling

Vertically 1–2 cun

Actions/Indications

- Relaxes the muscles and tendons
- Opens the channel and *luo* vessels
- Clears Heat (from the Small Intestine)

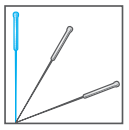
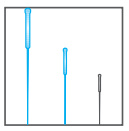

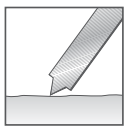

Ch04.7a-F10028.qxd 2/23/08 3:55 PM Page 286

BL-39 *Outside of the Crook* WEIYANG

Location

At the lateral end of the popliteal crease, on the medial side of the tendon of the long head of the biceps femoris muscle, 1 cun lateral to → **BL-40** (in the centre of the popliteal crease).

How to find

This point is best located with the patient's knee slightly flexed. Find the midpoint of the popliteal crease (→ **BL-40**) and locate **BL-39** approximately 1 cun lateral in a depression medial to the tendon of the long head of the biceps femoris muscle. Located on the same level is → **BL-40** at the centre of the popliteal crease.

Needling

Vertically 0.5–1.5 cun. Caution: N. fibularis communis.

Actions/Indications

- Harmonises the Triple Burner and regulates the water passages
- Opens the channel and the *luo* vessels, alleviates pain

Special features

Lower *he*-sea point of the Triple Burner

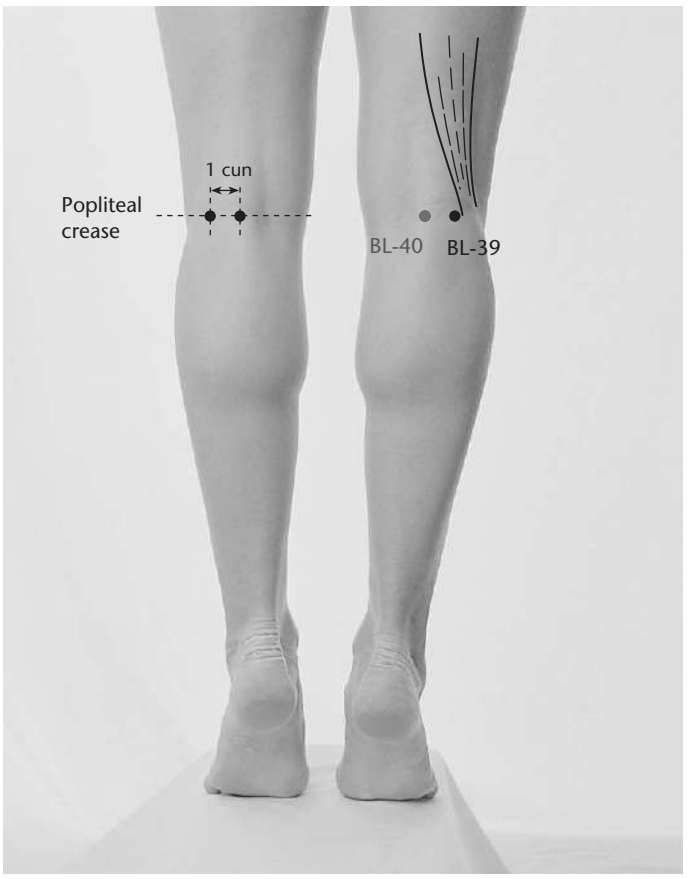

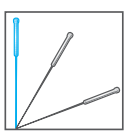

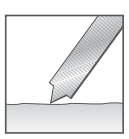

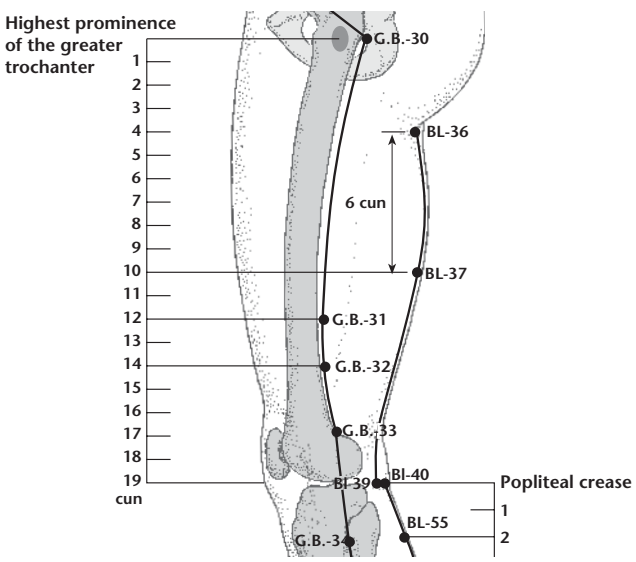

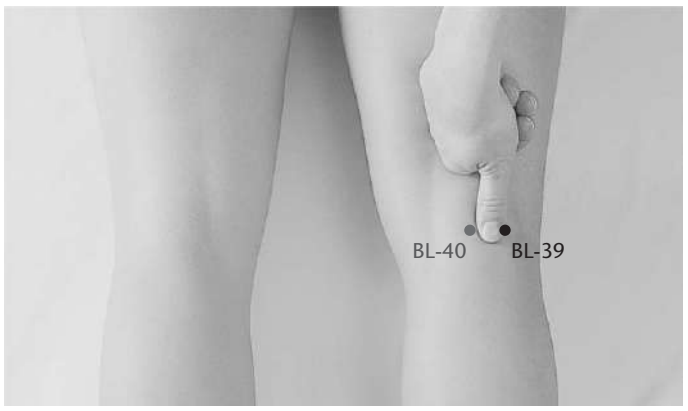

Ch04.7a-F10028.qxd 2/23/08 3:55 PM Page 287

Middle of the Crook WEIZHONG BL-40

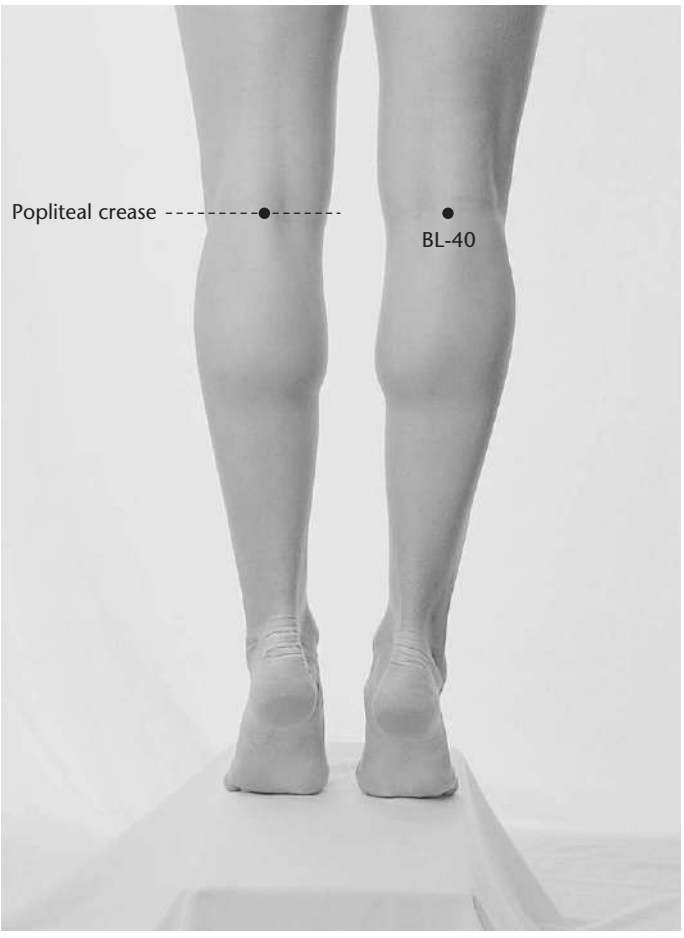

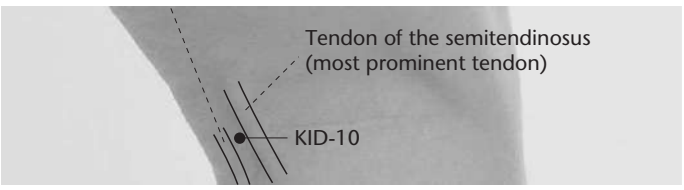

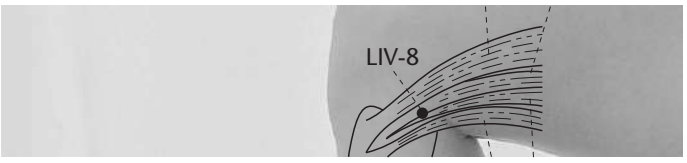

Location

In the centre of the popliteal crease, between the tendons of the biceps femoris and semitendinosus muscles.

How to find

This point is best located with the patient's knee slightly flexed. Locate the centre of the popliteal crease (a pulse may be palpable) and there locate **BL-40**.

Located on the same level are → **BL-39** (1 cun laterally), → **KID-10** (more medially, between the tendons of the semimembranosus and semitendinosus muscles) and → **LIV-8** (more medially, anterior to the tendons of the semimembranosus and semitendinosus muscles).

Needling

Vertically 0.5–1.5 cun. Caution: Popliteal nerve, artery and vein lie deep to this point. Use moxibustion with caution (contraindicated according to some authors). To clear Heat and eliminate stasis: bleed the superficial veins, possibly followed by brief cupping. Caution: Only use for excess conditions and in constitutionally robust patients. **BL-40** is also indicated as a distal point for the lower back: needle with the patient standing (for safety and balance, use a treatment couch or chair as a support for the hands) and ask them to move their lower back gently while manipulating the needle (only short needle retention). When needling this point with the patient in a prone position, place a supporting roll or pillow under the ankles to relax the popliteal crease.

Actions/Indications

- Clears (summer) Heat, stops vomiting and diarrhoea
- Cools the Blood
- Benefits the lower back and knees, opens the channel and *luo* vessels, alleviates pain

Special features

He-sea point, Earth point, lower *he*-sea point of the Bladder, Ma Dan Yang Heavenly Star point, Gao Wu command point of the lumbar region. Important distal point for the lower back, point with a wide range of indications.

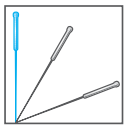
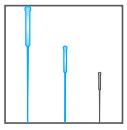

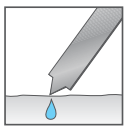
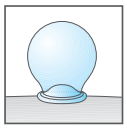

Ch04.7a-F10028.qxd 2/23/08 3:55 PM Page 288

BL-41 *Attached Branch* FUFEN

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 2nd thoracic vertebra (T2).

How to find

After locating the spinous process of C7 (→ 3.4.1), count down 2 spinous processes to the lower border of the spinous process of T2. Locate **BL-41** 3 cun lateral from the midline (with the shoulders relaxed and the arms hanging loosely down, 3 cun corresponds on this level to the distance from the midline to the medial border of the scapula). Note: From **BL-41** to → **BL-54**, all points on the BL channel are located on its outer branch.
Located on the same level are a point of → **Ex-B-2** (0.5 cun lateral to the midline), → **BL-12** (1.5 cun lateral to the midline) and → **S.I.-13** (more laterally, on the scapula).

Needling

Obliquely 0.3–0.5 cun. Caution: Pneumothorax.

Actions/Indications

- Expels pathogenic factors such as Wind and Cold
- Opens the channel and the *luo* vessels, alleviates pain

Special features

Meeting point with the S.I. channel

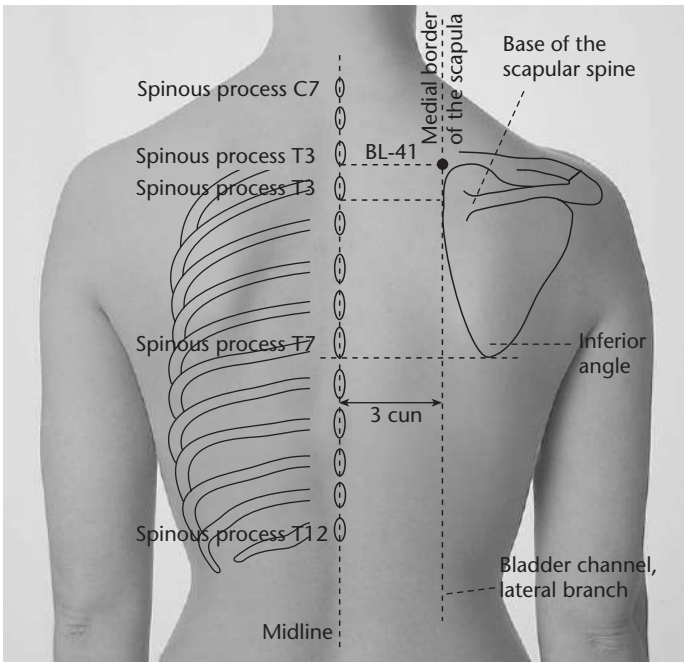

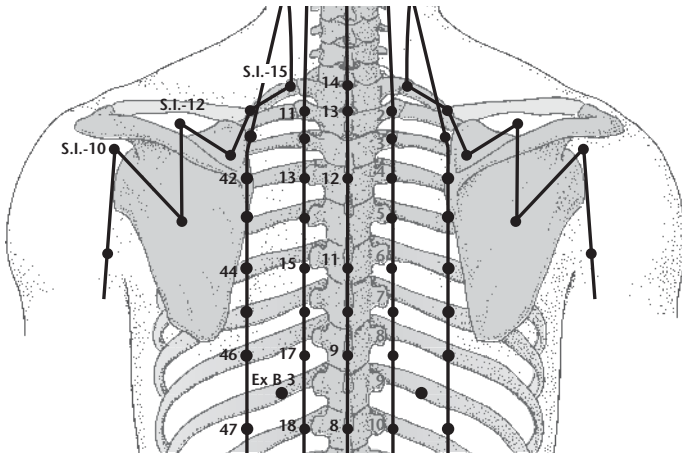

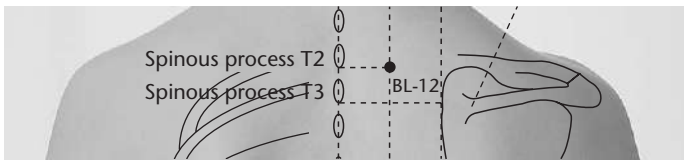

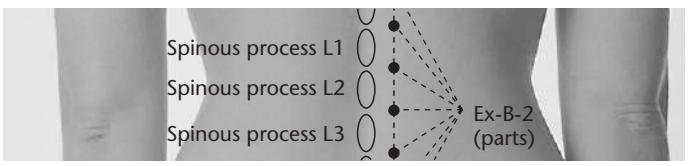

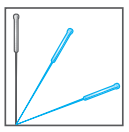

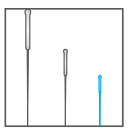

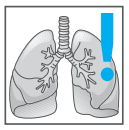

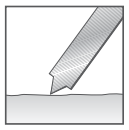

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 289

Door of the Corporeal Soul POHU BL-42

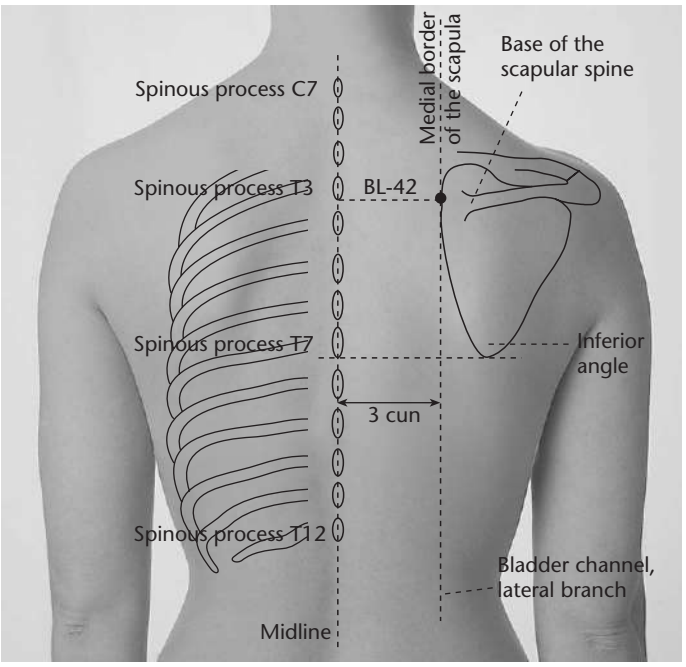

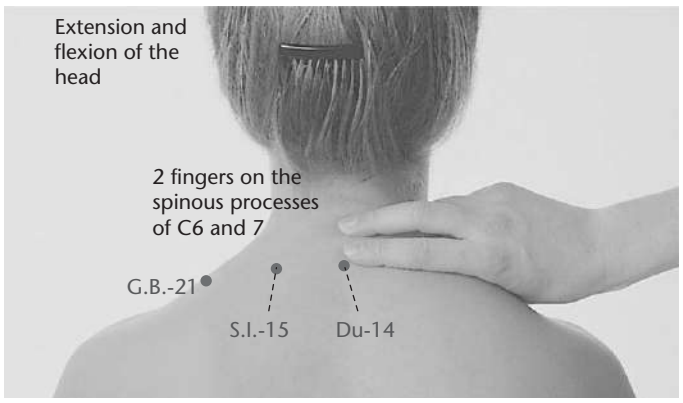

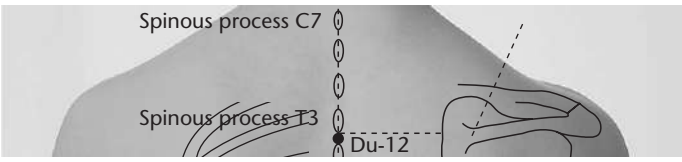

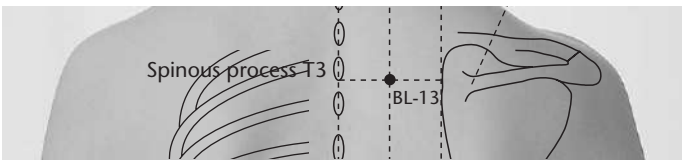

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 3rd thoracic vertebra (T3).

How to find

After locating the spinous process of C7 (→ 3.4.1), count down 3 spinous processes to the lower border of the spinous process of T3. Located on this level and 3 cun lateral to the midline is **BL-42.**
Or: With the patient seated and their shoulders in a relaxed position, the spinous process of T3 is generally on the level of the medial end of the scapular spine, an easily palpable structure (→ 3.4.2). Locate **BL-42** on the level of the lower border of the spinous process of T3, 3 cun lateral to the midline. With the patient's shoulders relaxed, 3 cun corresponds on this level to the distance from the midline to the medial border of the scapula.
Located on the same level are → **Du-12** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-13** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.5 cun. Caution: Pneumothorax.

Actions/Indications

- Tonifies and nourishes the Lung
- Alleviates wheezing and coughing, calms the corporeal soul (*po*)
- Opens the channel and alleviates pain
- Clears Heat from the Lung

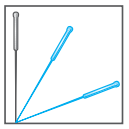
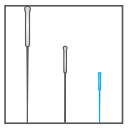
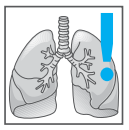

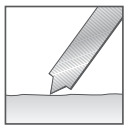

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 290

BL-43 *Vital Region* Shu GAOHUANG

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 4th thoracic vertebra (T4).

How to find

After locating the spinous process of C7 (→ 3.4.1), count down 4 spinous processes to the lower border of the spinous process of T4. Located on this level and 3 cun lateral to the midline is **BL-43. Or:** Locate **BL-43** by using the spinous process of T3 (generally, on the level of the easily palpable medial end of the scapular spine, 3.4.2) as a reference point.
Located on the same level are a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-14** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.5 cun. Moxibustion is often used on this point. Caution: Pneumothorax.

Actions/Indications

- Tonifies and nourishes the five *zang*-Organs: Lung, Heart, Kidneys, Stomach and Spleen
- Nourishes Yin, clears Heat
- Calms the *shen*
- Strengthens the Original Qi (*yuan qi*)
- Eliminates Phlegm

Special features

Important point for deficiency syndromes (classic indication)

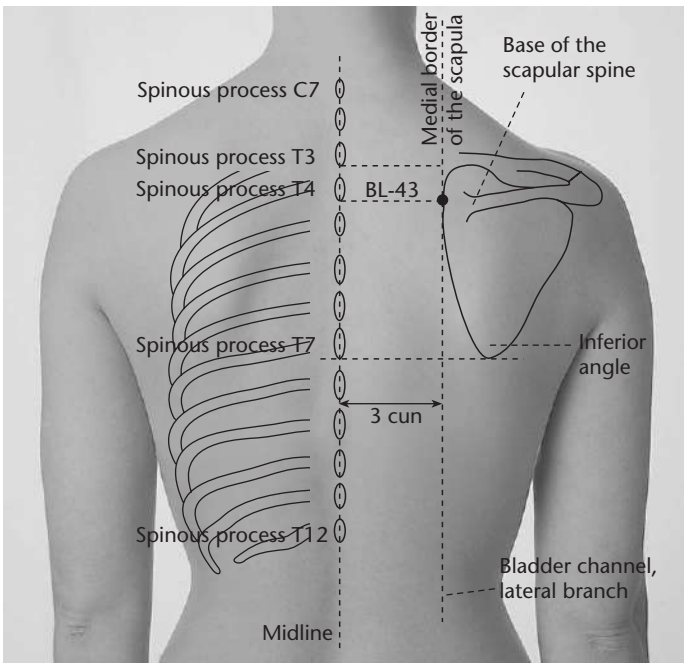

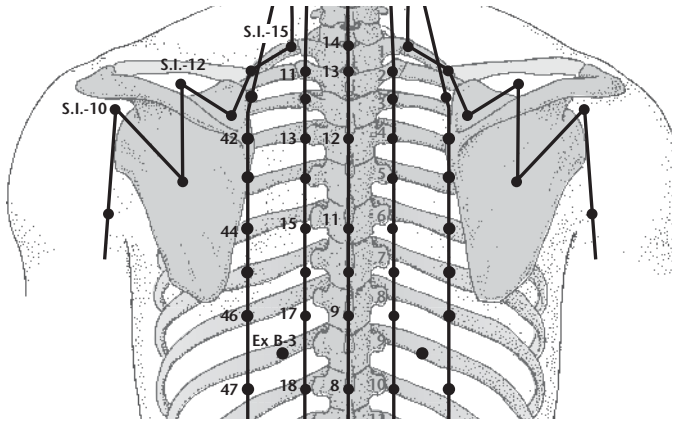

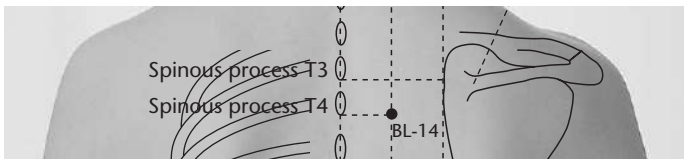

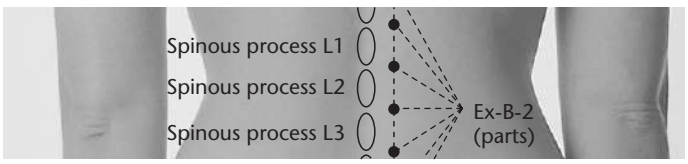

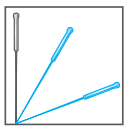

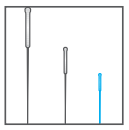

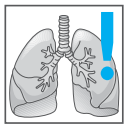

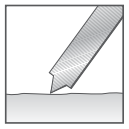

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 291

Spirit Hall SHENTANG BL-44

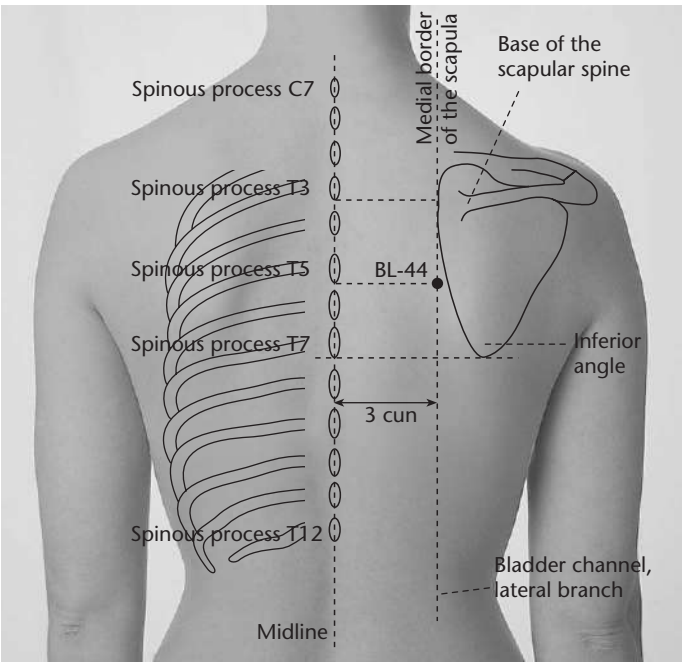

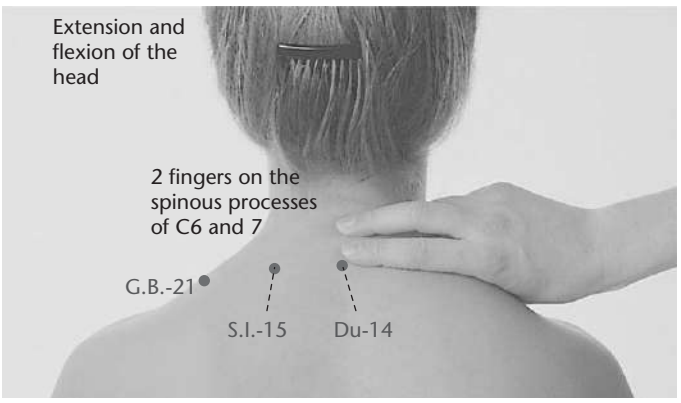

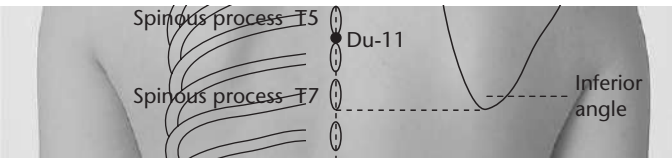

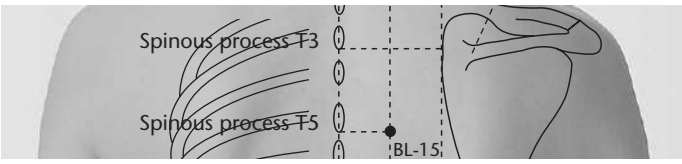

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 5th thoracic vertebra (T5).

How to find

After locating the spinous process of C7 (→ 3.4.1), count down 5 spinous processes to the lower border of the spinous process of T5. Located on this level and 3 cun lateral to the midline is **BL-44. Or:** With the patient seated and their shoulders in a relaxed position, the spinous process of T3 is generally on the level of the medial end of the scapular spine, an easily palpable structure (→ 3.4.2). Locate **BL-44** by counting down to the level of the lower border of the spinous process of T5 and measure 3 cun lateral to the midline.

Located on the same level are → **Du-11** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-15** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.8 cun. Caution: Pneumothorax.

Actions/Indications

- Regulates the Upper Burner, unbinds the chest
- Opens the channel and alleviates pain

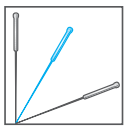
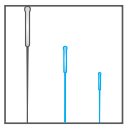
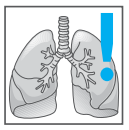

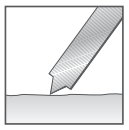

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 292

BL-45 *Cry of Pain* YIXI

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 6th thoracic vertebra (T6).

How to find

After locating the spinous process of C7 (→ 3.4.1), count down 6 spinous processes to the lower border of the spinous process of T6. Located on this level and 3 cun lateral to the midline is **BL-45. Or:** With the patient seated and their shoulders in a relaxed position, the spinous process of T7 is generally on the level of the inferior angle of the scapula (→ 3.4.2). Locate **BL-45** on the level of the spinous process of T6, 3 cun lateral to the midline. Located on the same level are → **Du-10** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-16** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.8 cun. Caution: Pneumothorax.

Actions/Indications

- Expels pathogenic factors, especially Wind and Heat
- Descends the Lung Qi and unbinds the chest
- Strengthens the Qi and Blood and alleviates pain

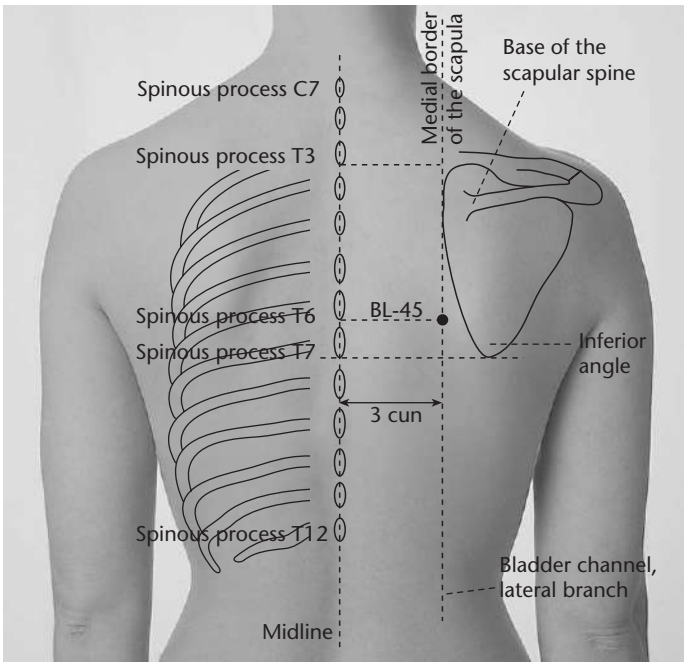

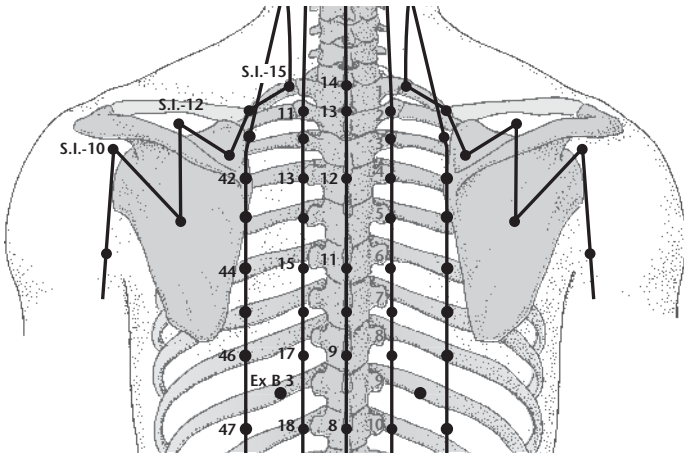

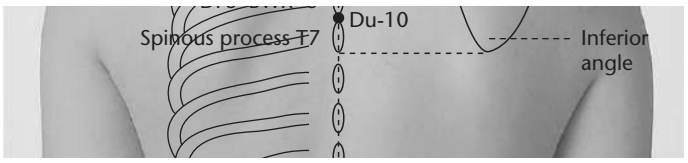

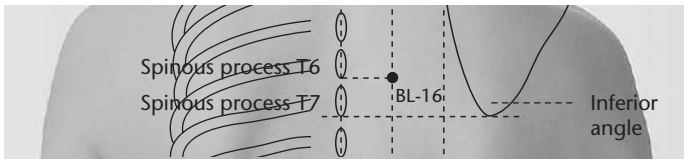

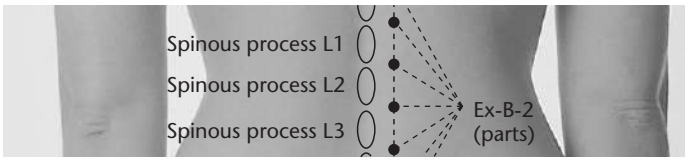

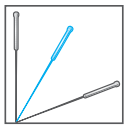

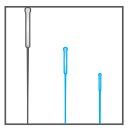

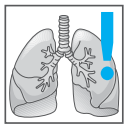

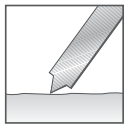

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 293

Diaphragm's Gate GEGUAN BL-46

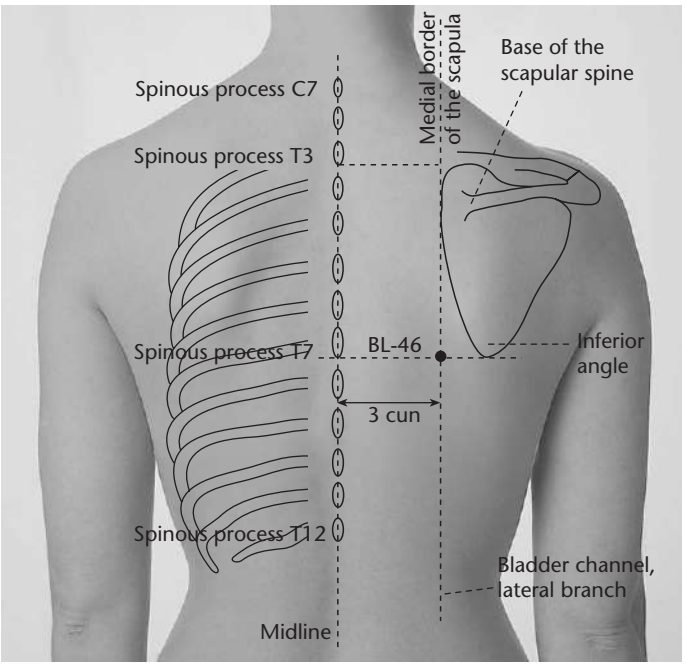

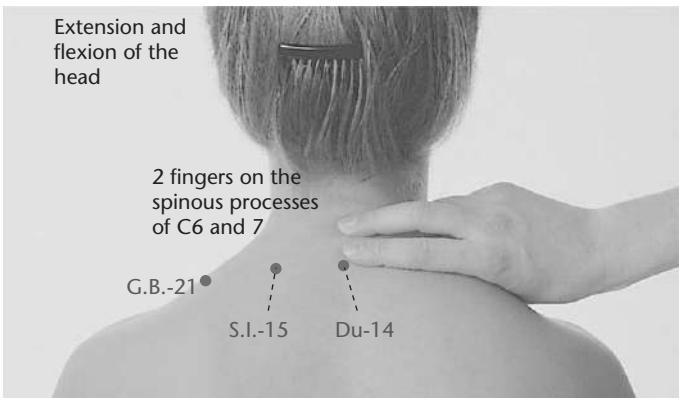

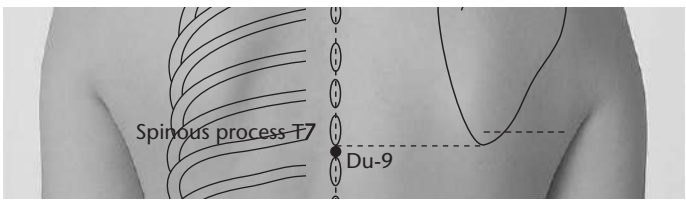

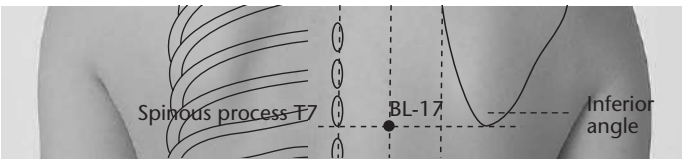

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 7th thoracic vertebra (T7).

How to find

With the patient standing and their shoulders in a relaxed position, the spinous process of T7 is generally on the level of the inferior angle of the scapula (→ 3.4.2). Locate **BL-46** on the level of the spinous process of T7, 3 cun lateral to the midline. **Or:** With the patient seated, the spinous process of T3 is generally on the level of the medial end of the scapular spine, an easily palpable bony protrusion at the medial border of the scapula (→ 3.4.2). From there, count down 4 spinous processes to the spinous process of T7. Locate **BL-46** on this level, 3 cun lateral to the midline. **Or:** Find the spinous process of C7 (→ 3.4.1). From there, count down 7 spinous processes in an inferior direction to the spinous process of T7. Locate **BL-46** on this level, 3 cun lateral to the midline.

Located on the same level are → **Du-9** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-17** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.8 cun. Caution: Pneumothorax.

Actions/Indications

- Regulates the diaphragm and descends counterflow Qi
- Harmonises the Middle Burner
- Opens the channel and alleviates pain

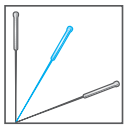
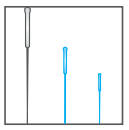
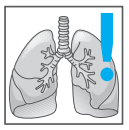

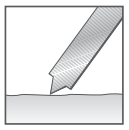

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 294

BL-47 *Gate of the Ethereal Soul* HUNMEN

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 9th thoracic vertebra (T9).

How to find

In a sitting position with the arms hanging down, the spinous process of T7 is generally on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down 2 processes to the lower border of the spinous process of T9. **BL-47** is located 3 cun lateral to this point. **Or:** Locate the spinous process of C7 (→ 3.4.1). From there, count down 9 processes to the lower border of the spinous process of T9 and locate **BL-47** 3 cun lateral to it.
Located on the same level are → **Du-8** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-18** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.5 cun. Caution: Pneumothorax.

Actions/Indications

- Spreads the Liver Qi
- Harmonises the Spleen and the Stomach
- Relaxes the tendons

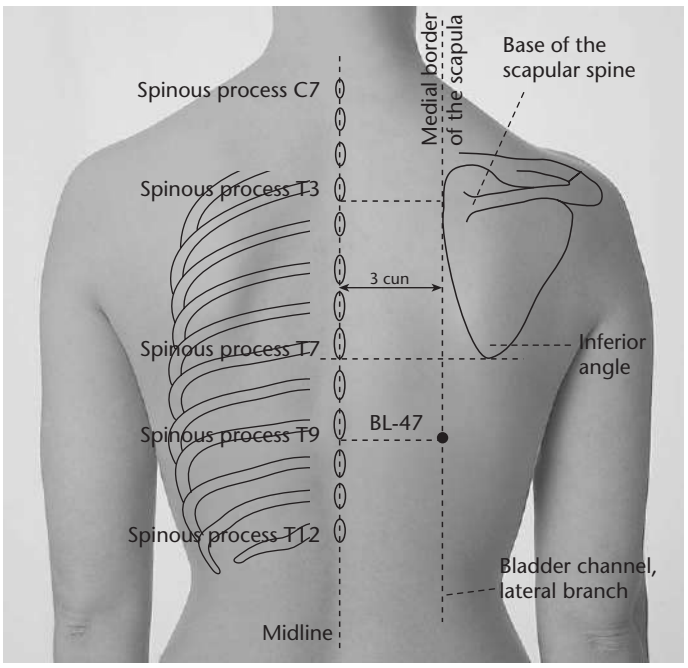

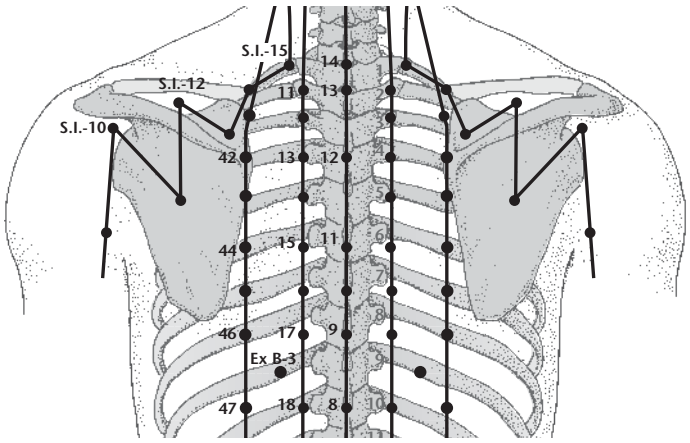

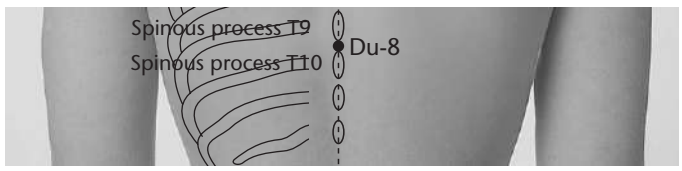

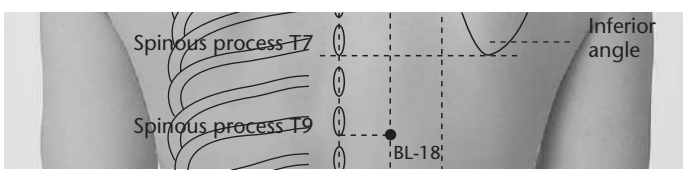

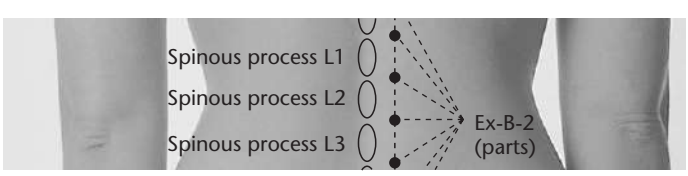

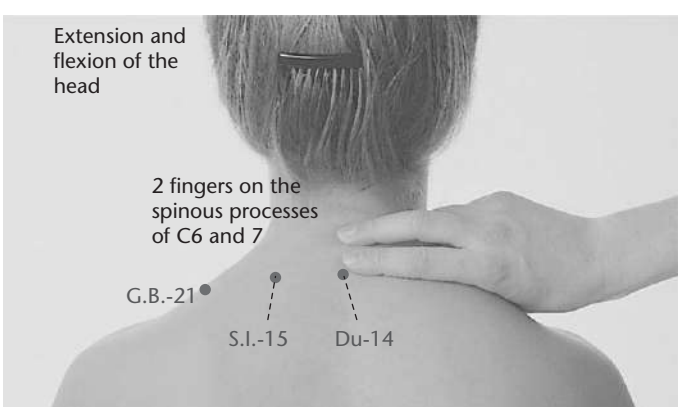

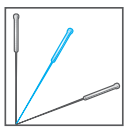
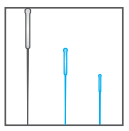
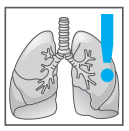
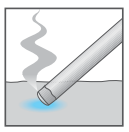
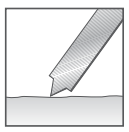

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 295

Yang's Key Link YANGGANG BL-48

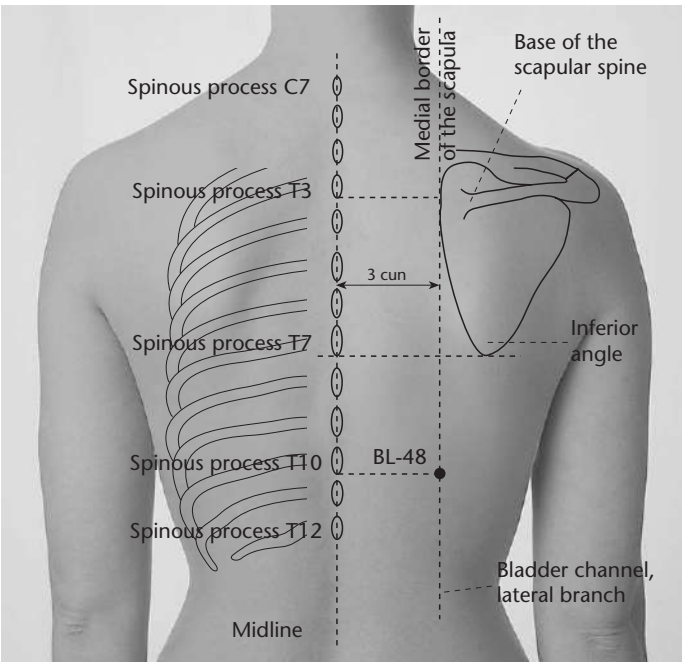

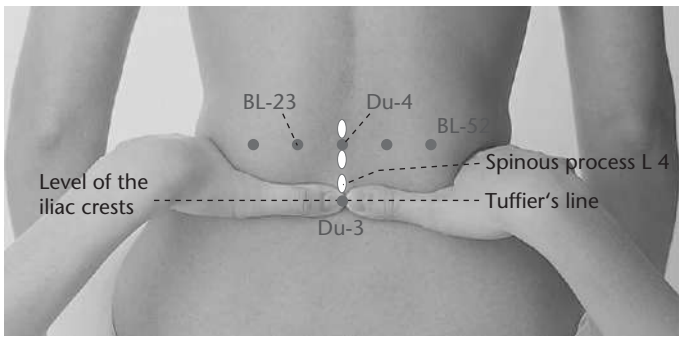

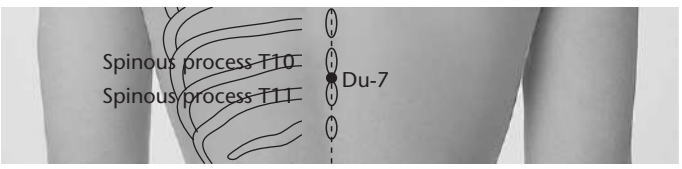

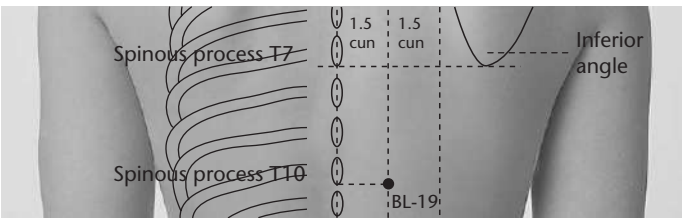

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 10th thoracic vertebra (T10).

How to find

In a sitting position with the arms hanging down, the spinous process of T7 is generally on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down 3 processes to the lower border of the spinous process of T10. **BL-48** is located 3 cun lateral to this point. **Or:** Locate the spinous process of C7 (→ 3.4.1). From there, count down 10 processes to the lower border of the spinous process of T10 and locate **BL-48** 3 cun lateral to it.

Located on the same level are → **Du-7** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-19** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.8 cun. Caution: Pneumothorax.

Actions/Indications

- Regulates the Gall Bladder
- Clears Damp-Heat
- Harmonises the Middle Burner

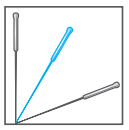
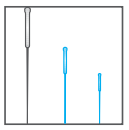
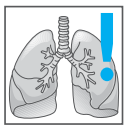

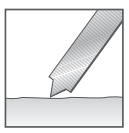

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 296

BL-49 *Abode of Thought* YISHE

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 11th thoracic vertebra (T11).

How to find

In a sitting position with the arms hanging down, the spinous process of T7 is generally on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down 4 spinous processes to the lower border of the spinous process of T11 and locate **BL-49** 3 cun lateral to this point. **Or:** From the articulation of the most inferior rib, palpate in a superior direction to T11. **Or:** Use the lumbar spine for reference (→ 3.4.3).
Located on the same level are → **Du-6** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-20** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.8 cun. Caution: Pneumothorax.

Actions/Indications

- Clears Damp-Heat
- Harmonises the Spleen and the Stomach

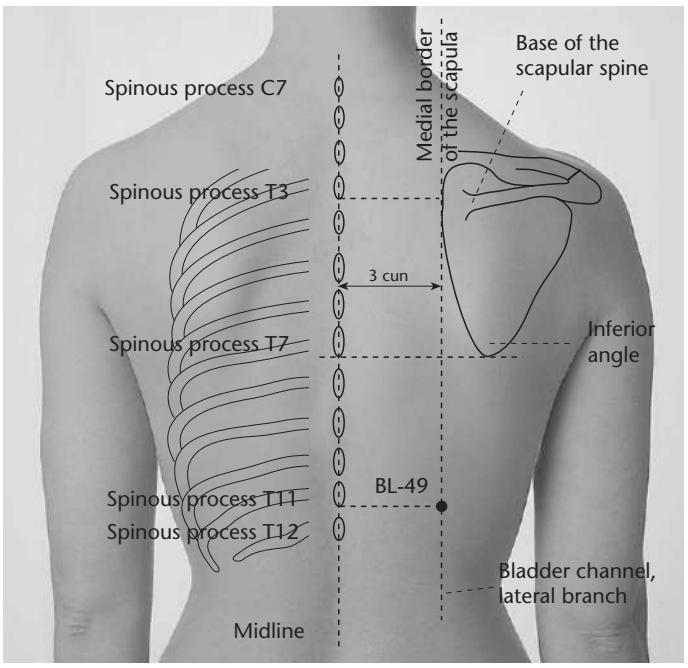

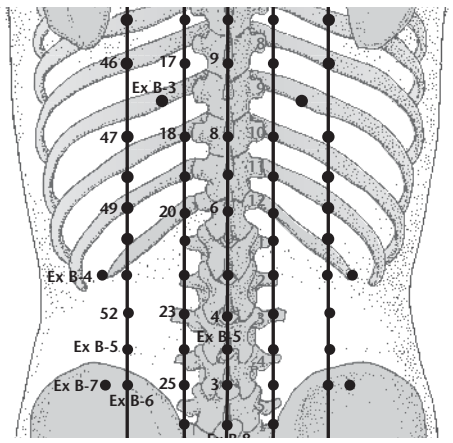

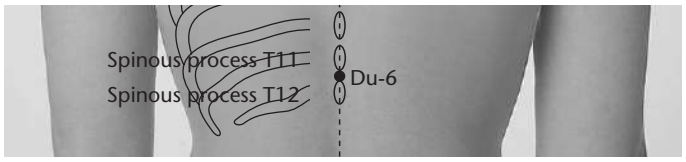

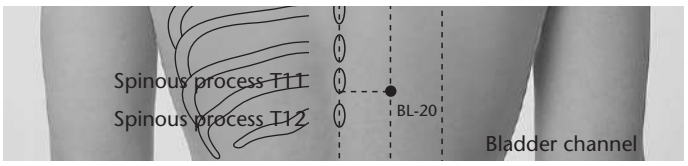

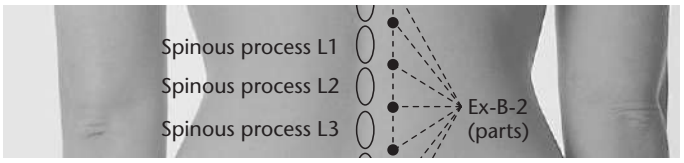

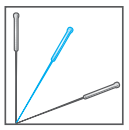
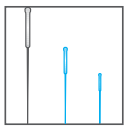
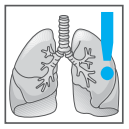

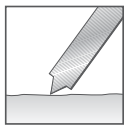

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 297

Stomach Granary WEICANG BL-50

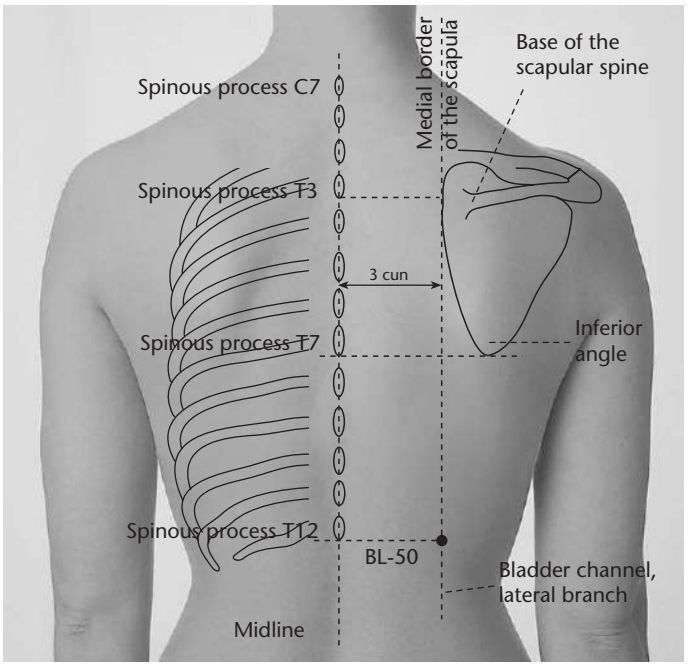

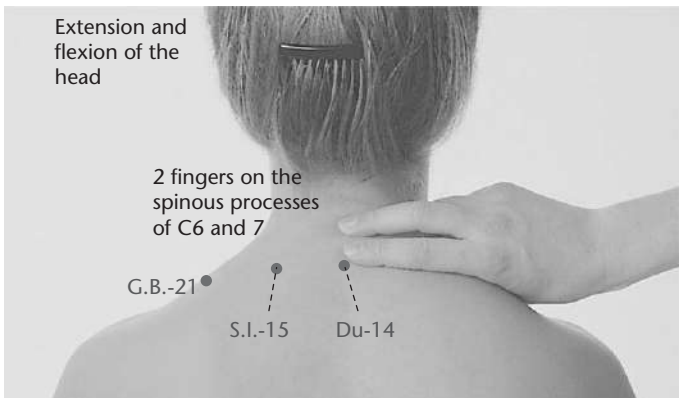

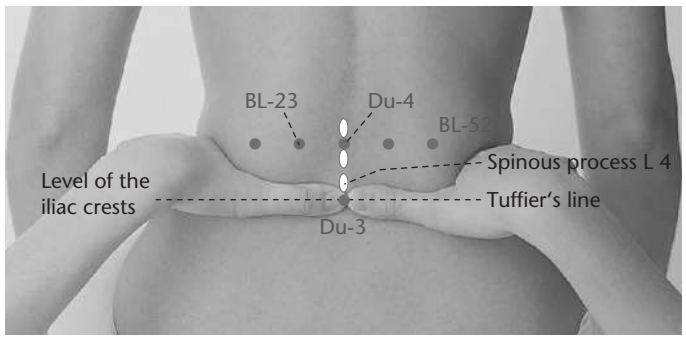

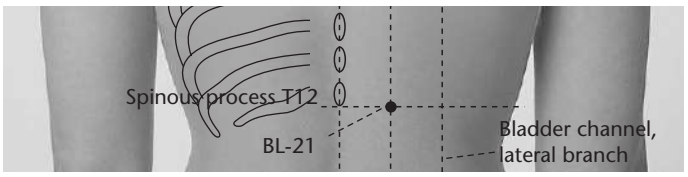

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 12th thoracic vertebra (T12).

How to find

In a sitting position with the arms hanging down, the spinous process of T7 is generally on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down 5 spinous processes to the lower border of the spinous process of the T12 and locate **BL-50** 3 cun lateral to this point. **Or:** Locate the articulation of the most inferior rib with T12 and palpate 3 cun in a lateral direction for **BL-50. Or:** Orientation from the lumbar spine (→ 3.4.3). Located on the same level are a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-21** (1.5 cun lateral to the midline).

Needling

Obliquely 0.3–0.8 cun. Caution: Pneumothorax, especially with asthenic patients or patients with emphysema.

Actions/Indications

- Harmonises the Middle Burner
- Opens the channel

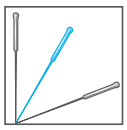

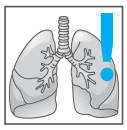
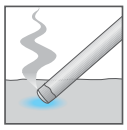
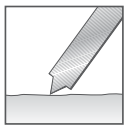

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 298

BL-51 *Vitals Gate* HUANGMEN

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 1st lumbar vertebra (L1).

How to find

For orientation in the lumbar region (→ 3.4.3), first, the Tuffier's line: place both hands superior or lateral to the highest points of the iliac crest, joining the thumbs in the middle above the spine. In most cases, this line intersects with the spinous process of L4 (note: this varies depending on the patient's position → 3.4.). From there, count upward to the lower border of the spinous process of L1 and, on this level, locate **BL-51** 3 cun in a lateral direction. **Or:** With the patient seated and their arms hanging down in a relaxed way, the spinous process of T7 is generally on the level of the inferior angle of the scapula (→ 3.4.2). From there, count down six spinous processes and locate **BL-51** 3 cun lateral to the spinous process of L1.

Located on the same level are → **Du-5** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline), → **BL-22** (1.5 cun lateral to the midline) and → **Ex-B-4** (→ 3.5 cun lateral to the midline).

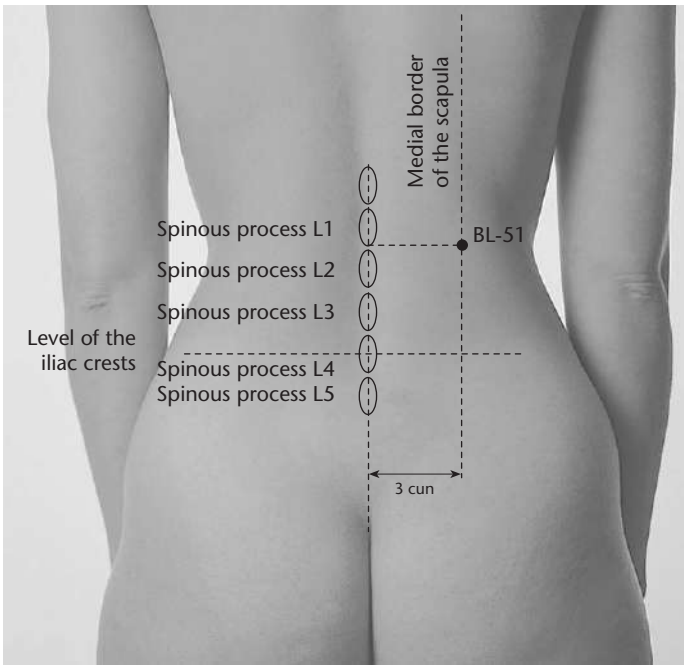

Needling

Vertically or obliquely 0.5–1 cun. Caution: Kidneys.

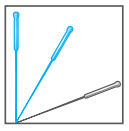

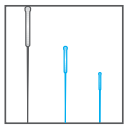

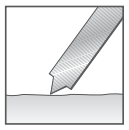

Actions/Indications

- Moves the Qi and resolves stagnation
- Benefits the breasts

Special features

Distal point for disorders of the breasts

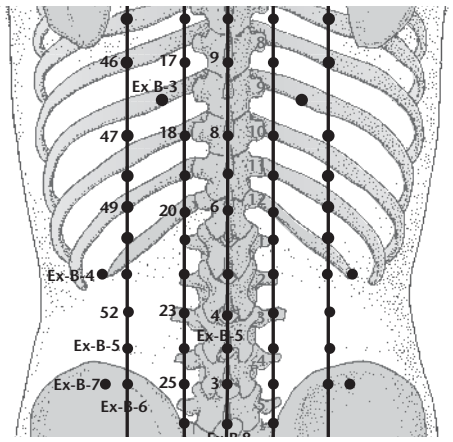

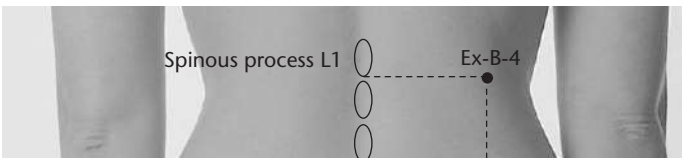

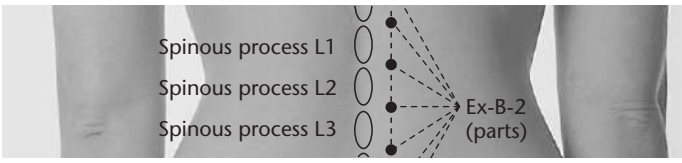

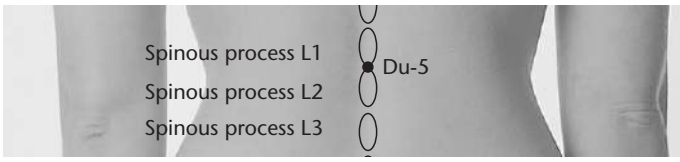

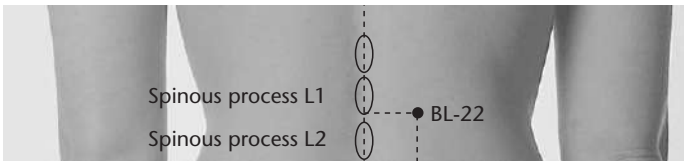

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 299

Residence of the Will ZHISHI BL-52

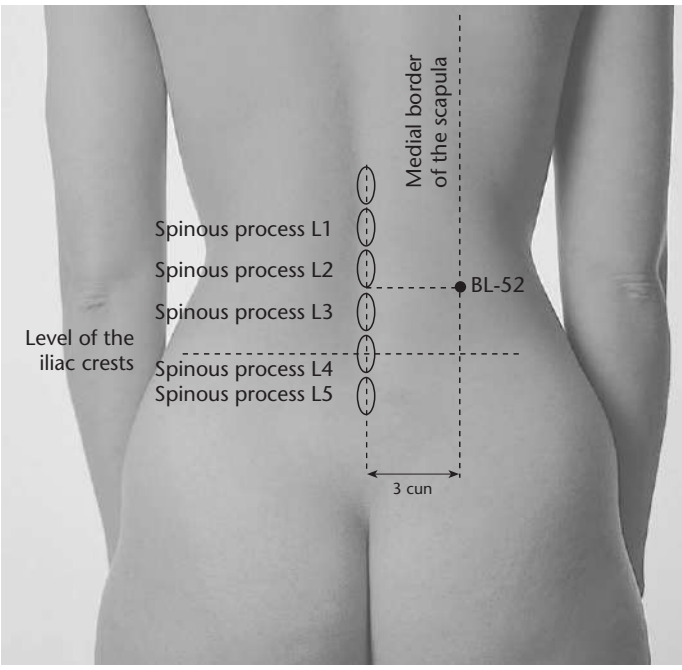

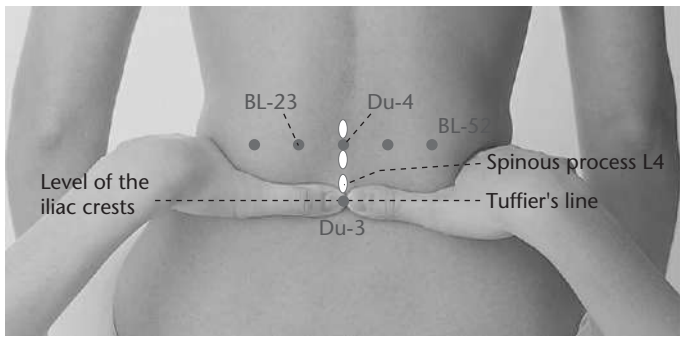

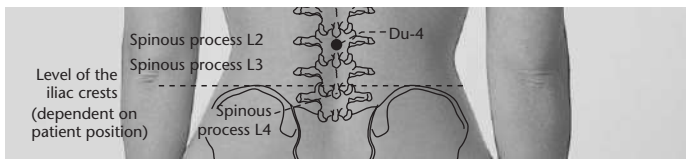

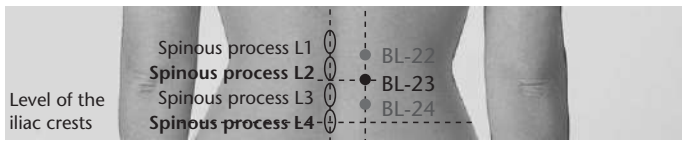

Location

3 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 2nd lumbar vertebra (L2).

How to find

For orientation in the lumbar region (→ 3.4.3), first, locate the Tuffier's line: place both hands superior or lateral to the highest points of the iliac crest, joining the thumbs in the middle above the spine. In most cases, this line intersects with the spinous process of L4 (note: this varies depending on the patient's position → 3.4.). From there, count upward to the lower border of the spinous process of L2 and, on this level, locate **BL-52** 3 cun in a lateral direction.

Located on the same level are → **Du-4** (on the midline), a point of → **Ex-B-2** (0.5 cun lateral to the midline) and → **BL-23** (1.5 cun lateral to the midline).

Needling

Vertically or obliquely 0.5–1 cun. Caution: Kidneys.

Actions/Indications

- Tonifies the Kidneys and Essence (*jing*), regulates urination
- Benefits the lumbar region

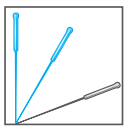
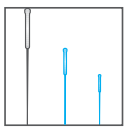

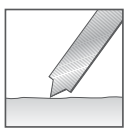

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 300

BL-53 *Bladder Vitals* BAOHUANG

Location

3 cun lateral to the posterior midline, on the level of the 2nd sacral foramen.

How to find

For details on orientation in the lumbar and sacral region, ➞ 3.4.3, 3.4.4. First, locate L5. The next palpable bony structure inferior to L5 is the highest process of the medial sacral crest. Palpate inferiorly for the 2nd process and locate the 2nd sacral foramen slightly less than 1 cun lateral to the process. Next, locate **BL-53** 3 cun lateral to the midline on the level of the 2nd foramen. **Or:** Use the sacral foramina as a reference point (➞ 3.4.4). **BL-53** is located on the level of the 2nd foramen, 3 cun lateral to the midline.

Located on the same level are ➞ **BL-32** (over the 2nd sacral foramen) and ➞ **BL-28** (1.5 cun lateral to the midline).

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Benefits the lumbar region, opens the channel and alleviates pain
- Regulates the Lower Burner

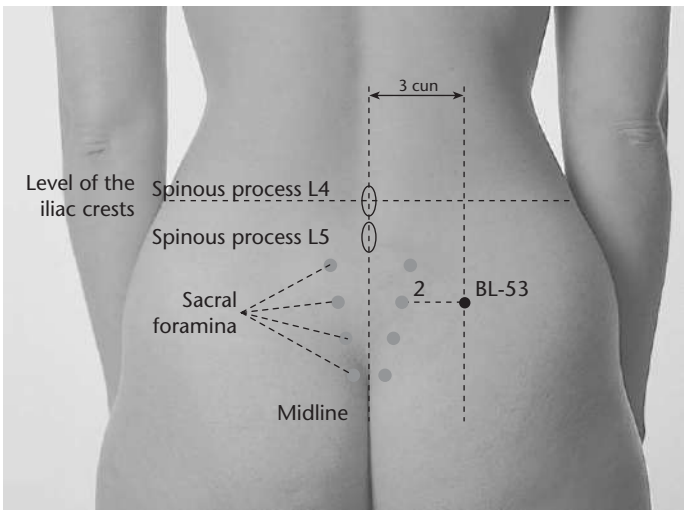

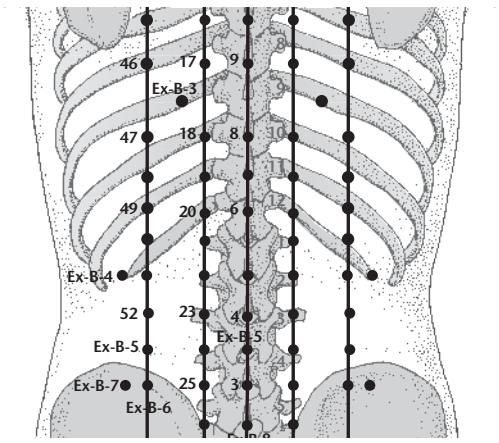

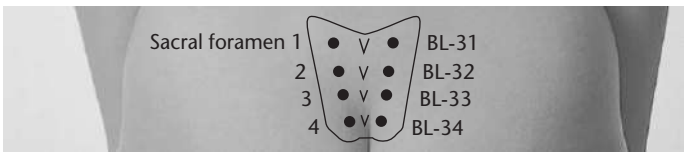

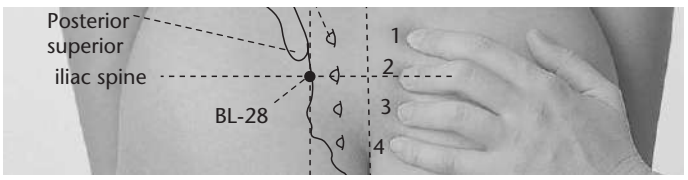

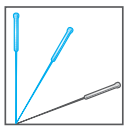
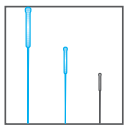

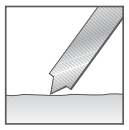
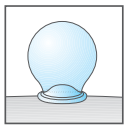

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 301

Order's Limit ZHIBIAN BL-54

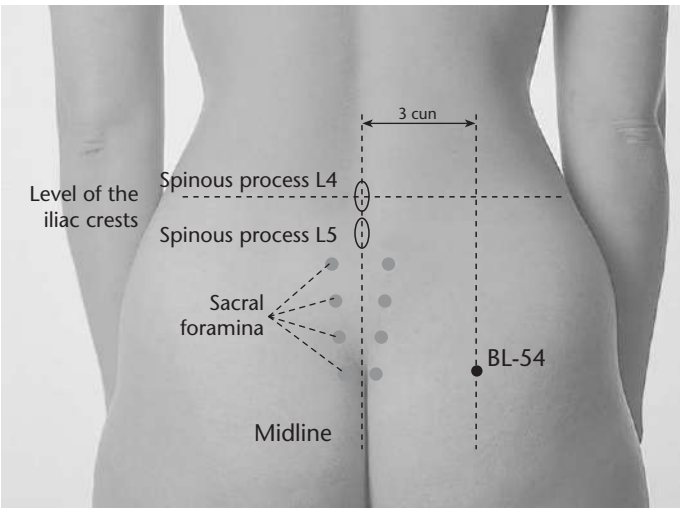

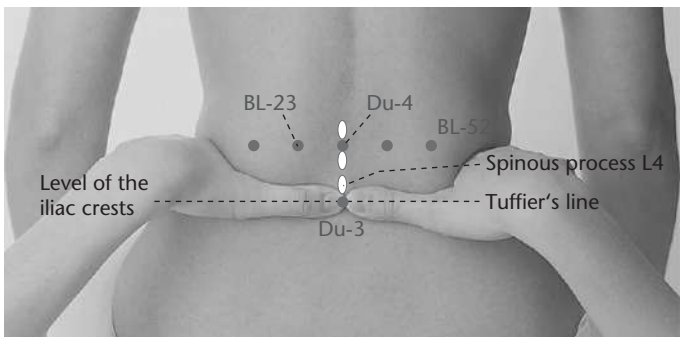

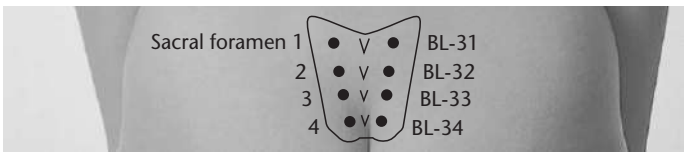

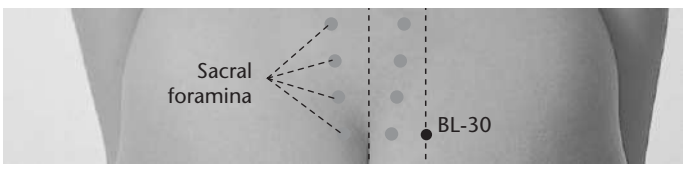

Location

3 cun lateral to the posterior midline, on the level of the 4th sacral foramen, approximately at the centre of the buttock.

How to find

For details on orientation in the lumbar and sacral region, see ➞ 3.4.3, 3.4.4. First, locate L5. Inferior to it, the next palpable bony structure is the highest point on the medial sacral crest (= sacral spinous processes). Palpate downward for 3 more bony elevations and locate the depression of the 4th sacral foramen less than 1 cun lateral to it. This is on the level of **BL-54**, located 3 cun lateral to the midline. For further reference: The 4th sacral foramen is mostly level with the superior end of the rima ani. Located on the same level are ➞ **BL-34** (over the 4th sacral foramen) and ➞ **BL-30** (1.5 cun lateral to the midline).

Needling

Vertically 1.5–2 cun or 2–3 cun in the direction of the anus or the genital region, depending on the indication.

Actions/Indications

- Opens the channel and *luo* vessels, eliminates stagnation
- Regulates urination
- Eliminates haemorrhoids

Special features

Important local point

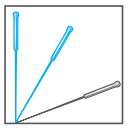
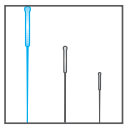

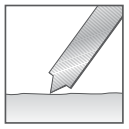
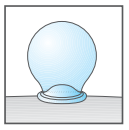

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 302

BL-55 *Confluence of Yang* HEYANG

Location

2 cun inferior to the midpoint of the popliteal crease, in a depression between the two bellies of the gastrocnemius muscle.

How to find

Locate the popliteal crease by flexing the knee. In slim patients, palpate for the joint space. From the centre of the popliteal crease, palpate 2 cun in an inferior direction and locate **BL-55** in a depression between the two bellies of the gastrocnemius muscle.

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Opens the channel, alleviates pain (also in the urogenital region)
- Stops uterine bleeding

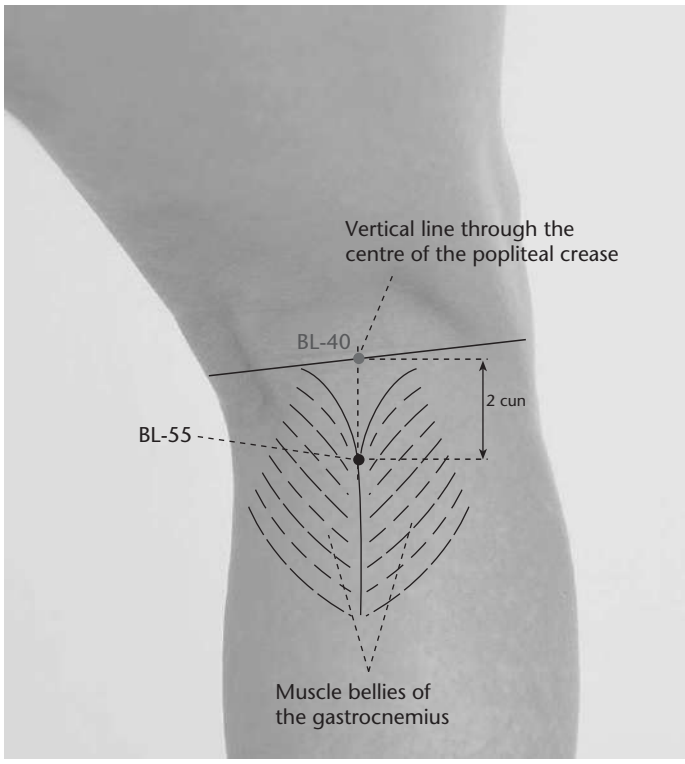

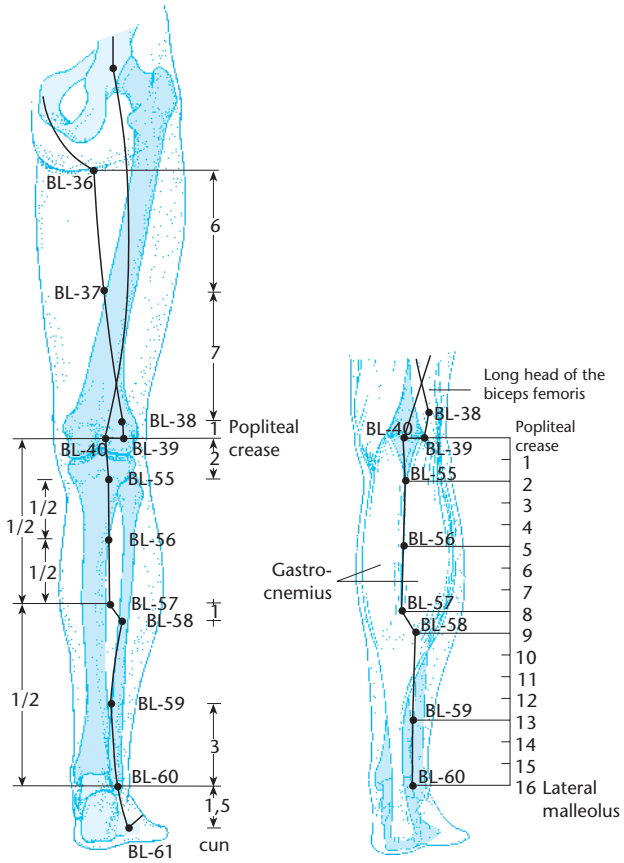

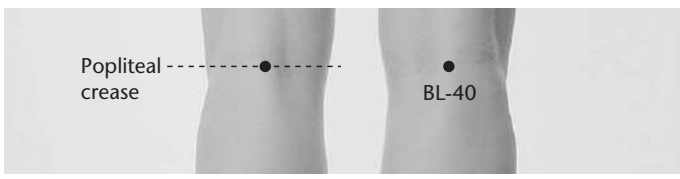

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 303

Support the Sinews CHENGJIN BL-56

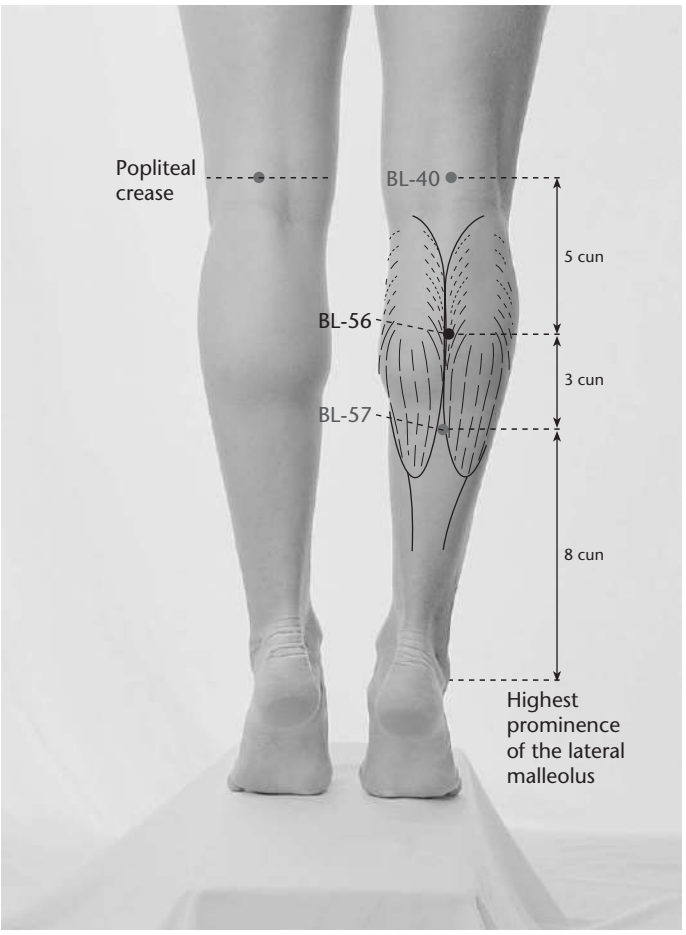

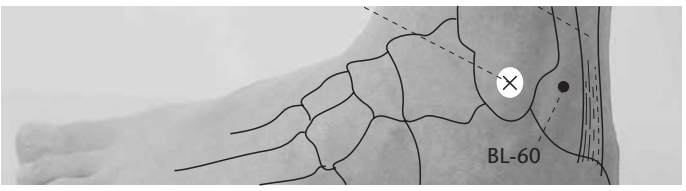

Location

5 cun inferior to the midpoint of the popliteal crease, between the two bellies of the gastrocnemius muscle.

How to find

Locate the popliteal crease by flexing the knee. In slim patients, palpate for the joint space. **BL-56** is located 5 cun distal to the popliteal crease, in a depression between the two bellies of the gastrocnemius muscle.

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Relaxes the muscles and tendons
- Opens the channel
- Eliminates haemorrhoids

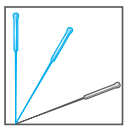
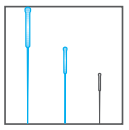

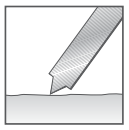
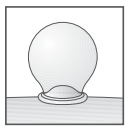

Ch04.7b-F10028.qxd 2/22/08 5:00 PM Page 304

BL-57 *Supporting Mountain* CHENGSHAN

Location

In the middle of the calf, between the two heads of the gastrocnemius muscle, on a line connecting → **BL-40** and → **BL-60**, approximately 8 cun distal to → **BL-40**.

How to find

Slide along the dorsal aspect of the Achilles tendon in a superior direction until you can feel the depression between the two bellies of the gastrocnemius muscle. This will become more defined when the patient's muscles are flexed. **Or:** Spreading hands technique (→ 2.3.3): Place the little fingers on → **BL-40** (centre of the popliteal crease) and → **BL-60** (depression between the Achilles tendon and the highest prominence of the lateral malleolus). The midpoint of this distance, where the thumbs meet, marks the location of **BL-57**.

Located on the same level on the anterolateral aspect of the lower leg are → **ST-38** (1 finger-width lateral to the tibia) and → **ST-40** (2 fingerbreadths lateral to the tibia).

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Relaxes the muscles and tendons, opens the channel
- Eliminates haemorrhoids (pathway of the Bladder divergent channel)

Special features

Ma Dan Yang Heavenly Star point. Important local point with distal effect on the lumbar and anal region.

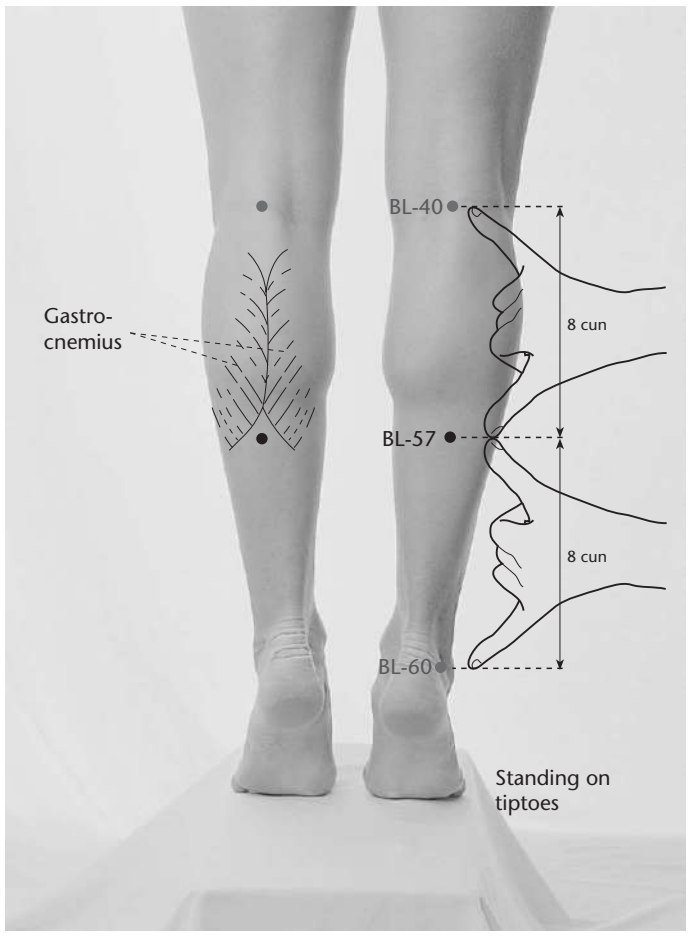

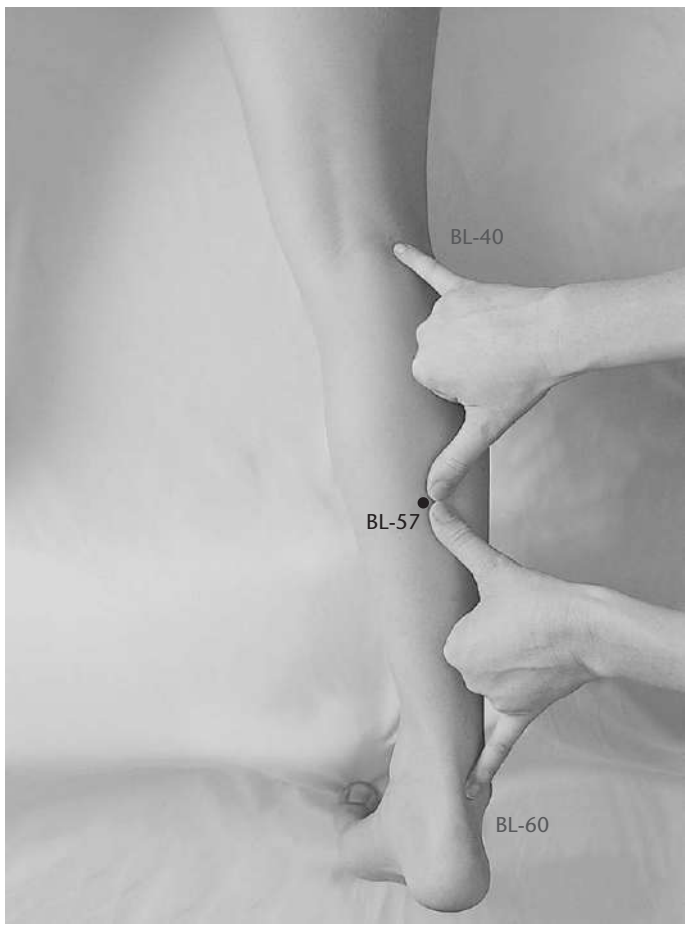

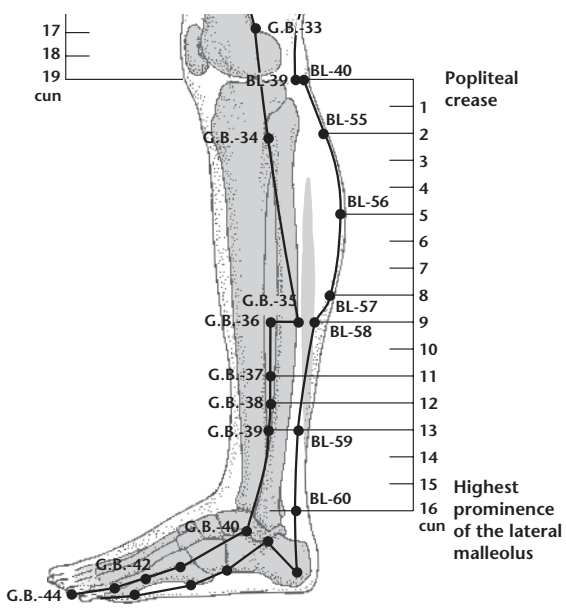

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 305

Soaring Upward FEIYANG BL-58

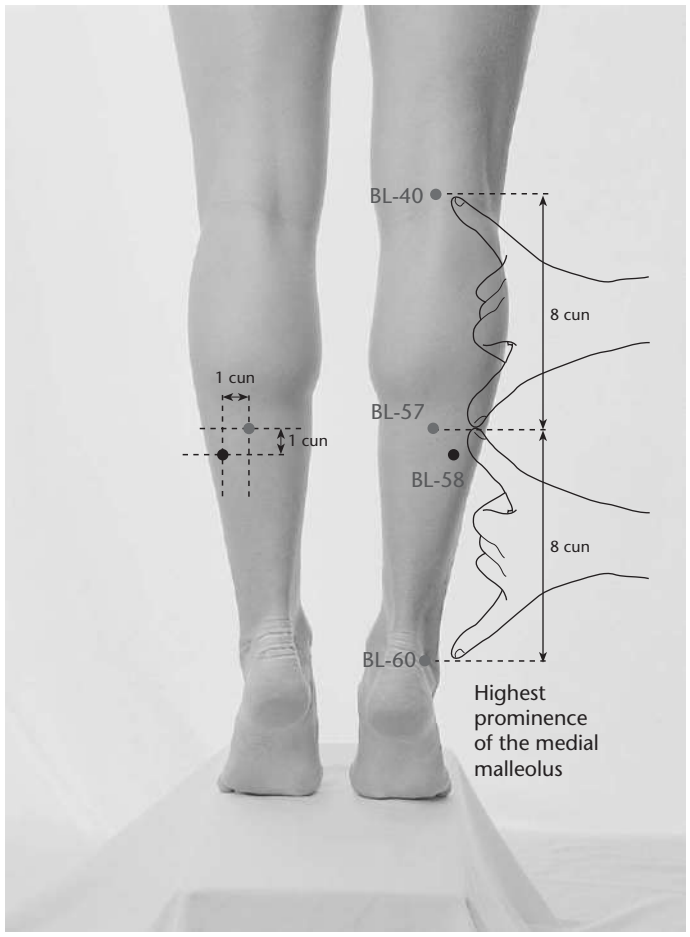

Location

1 cun distal and 1 cun lateral to → **BL-57** or 7 cun proximal to → **BL-60**, at the posterior border of the fibula and at the lower border of the gastrocnemius muscle.

How to find

First, locate → **BL-57** (in the centre of the calf, between the two bellies of the gastrocnemius muscle). From → **BL-57**, palpate 1 cun in a distal direction and 1 cun in a lateral direction. There, locate **BL-58** on the lower border of the gastrocnemius muscle. Located on the same level (7 cun superior to the highest prominence of the lateral malleolus) are → **G.B.-35** (at the posterior border of the fibula), → **G.B.-36** (at the anterior border of the fibula) and → **ST-39** (1 cun distal to the midpoint of the line connecting → **ST-35** and → **ST-41** and 1 fingerbreadth lateral to the anterior crest of the tibia).

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Opens the channel and *luo* vessels, alleviates pain
- Expels pathogenic factors from the *taiyang* channels (S.I., BL)
- Harmonises above and below
- Treats haemorrhoids

Special features

Luo-connecting point

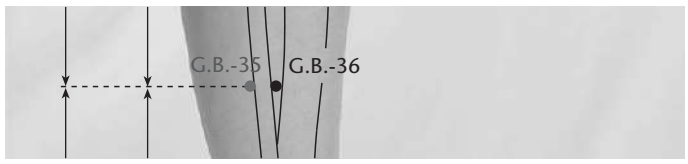

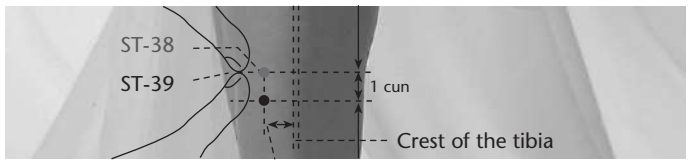

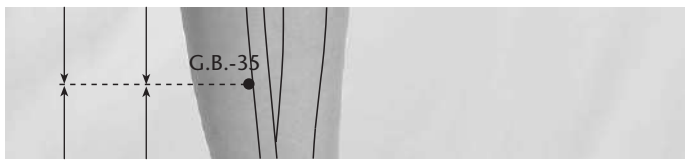

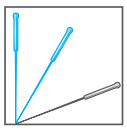

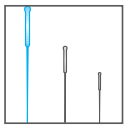

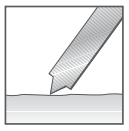

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 306

BL-59 *Instep Yang* FUYANG

Location

On the lateral aspect of the lower leg, 3 cun superior to → **BL-60** (in the depression between the highest prominence of the lateral malleolus and the Achilles tendon).

How to find

First, locate → **BL-60** in the depression between the highest prominence of the lateral malleolus and the Achilles tendon. With oedema of the lower leg, this point is not visible but palpable. From → **BL-60**, measure 3 cun (1 handbreadth) in a superior direction. There, locate **BL-59** in a depression between the Achilles tendon and the tendons of the peroneous longus and brevis muscles.

Located on the same level is → **G.B.-39**, 3 cun directly above the prominence of the lateral malleolus.

Needling

Vertically or obliquely 1–1.5 cun

Actions/Indications

- Opens the channel and *luo* vessels
- Benefits the lower back
- Benefits the head, activates the *yang qiao mai*

Special features

Xi-cleft point of the *yang qiao mai*, meeting point with the *yang qiao mai* according to some authors.

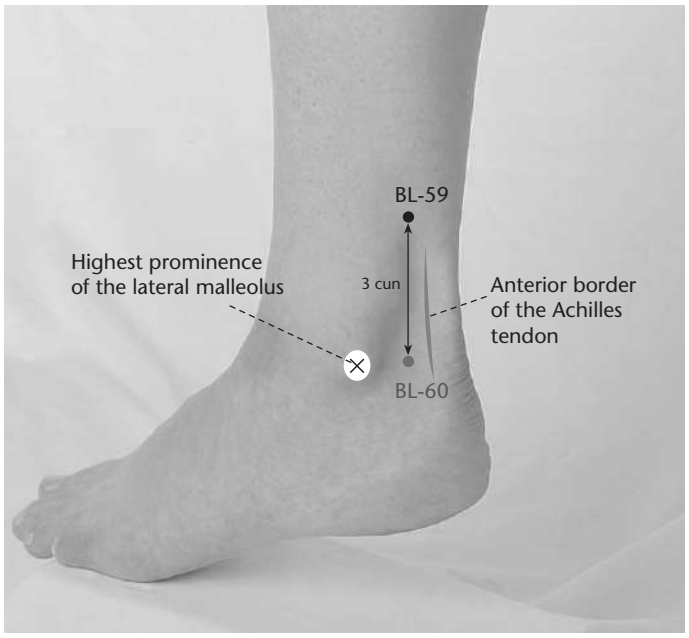

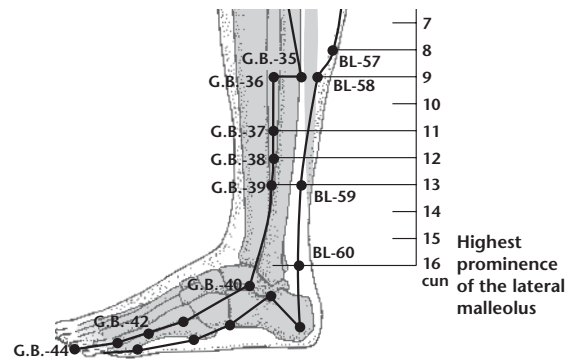

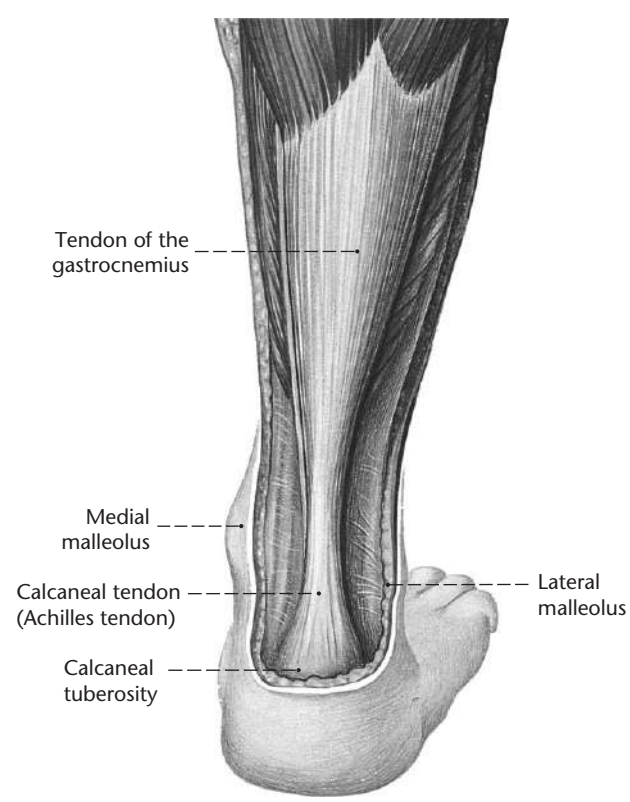

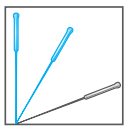
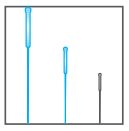

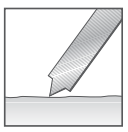
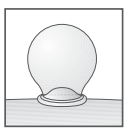

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 307

Kunlun Mountains KUNLUN BL-60

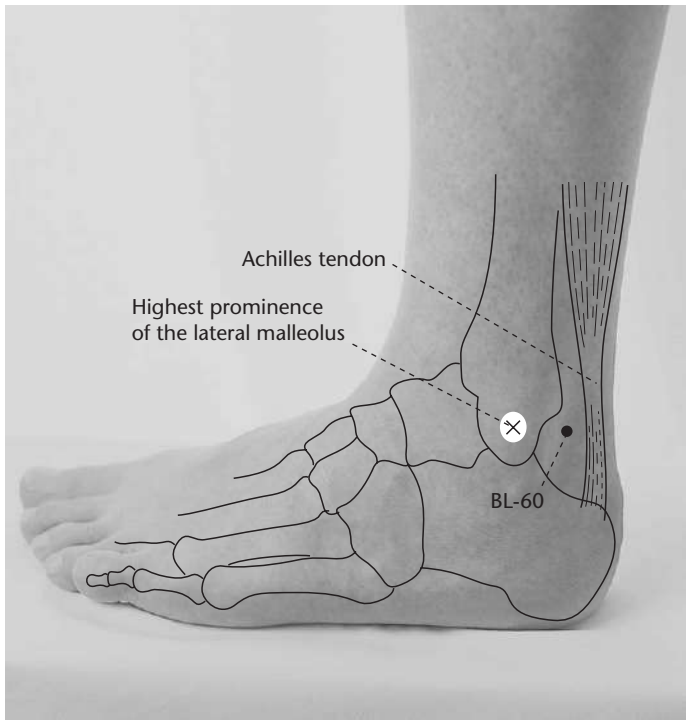

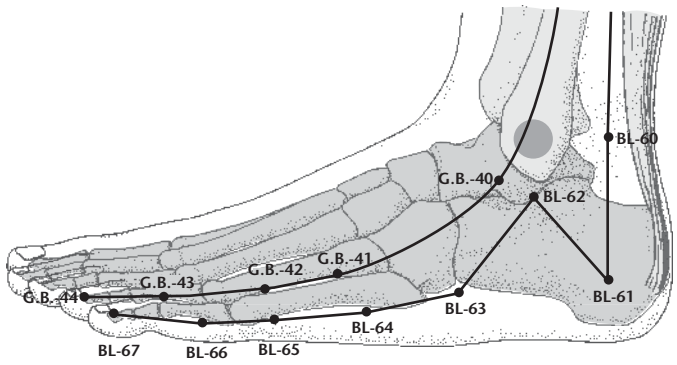

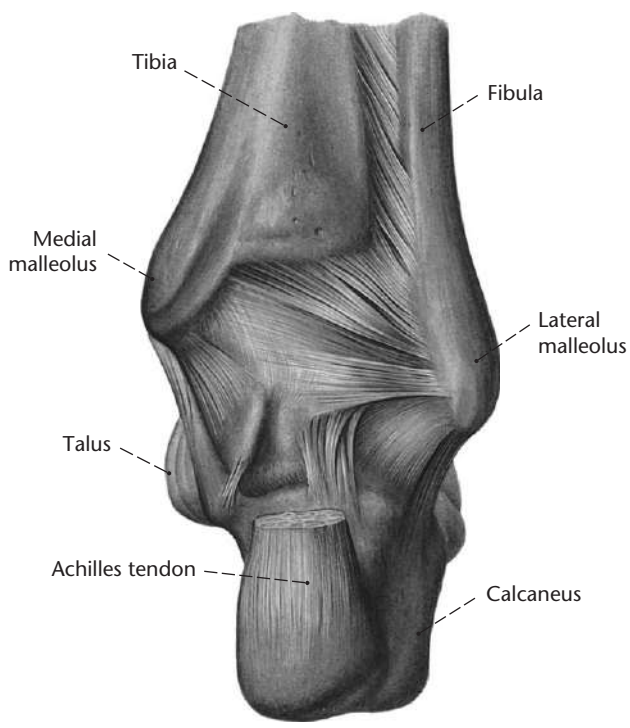

Location

In the depression on the line connecting the Achilles tendon and the highest prominence of the lateral malleolus.

How to find

Locate the highest prominence of the lateral malleolus (→ 3.6.2). From there, palpate horizontally towards the Achilles tendon and locate **BL-60** in a depression anterior to the tendon.

Needling

Vertically 0.5–1 cun. Caution: Reducing needle techniques contraindicated during pregnancy. Exception: A supportive point during labour. If the patient has cold feet, the therapeutic effect can be increased by using a heatlamp or a hot cherry-pit bag under the soles of the feet.

Actions/Indications

- Clears Heat, subdues Yang and Wind and eliminates excess, especially in the head
- Opens the channel, alleviates pain, relaxes the tendons, strengthens the lumbar region
- Promotes labour

Special features

Jing-river point, Fire point, Ma Dan Yang Heavenly Star point. Important distal point for the cervical, thoracic and lumbar spine, especially for chronic cases.

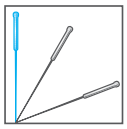
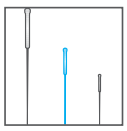

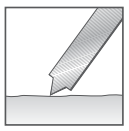
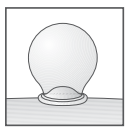

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 308

BL-61 *Servant's Respect* PUCAN

Location

On the lateral aspect of the heel, 1.5 cun inferior to → **BL-60** (in the depression between the highest prominence of the lateral malleolus and the Achilles tendon), in a depression on the calcaneus.

How to find

First, locate → **BL-60** in the depression between the highest prominence of the lateral malleolus and the Achilles tendon. With oedema of the lower leg, the depression is not visible but palpable. From → **BL-60**, measure 1.5 cun in a distal direction. There, locate **BL-61** in a depression on the calcaneus, approximately on the midpoint of a line connecting the sole of the foot and → **BL-60**.

Needling

Vertically or obliquely 0.3–0.5 cun

Actions/Indications

- Relaxes the tendons
- Opens the channel, alleviates pain

Special features

Meeting point with the *yang qiao mai*

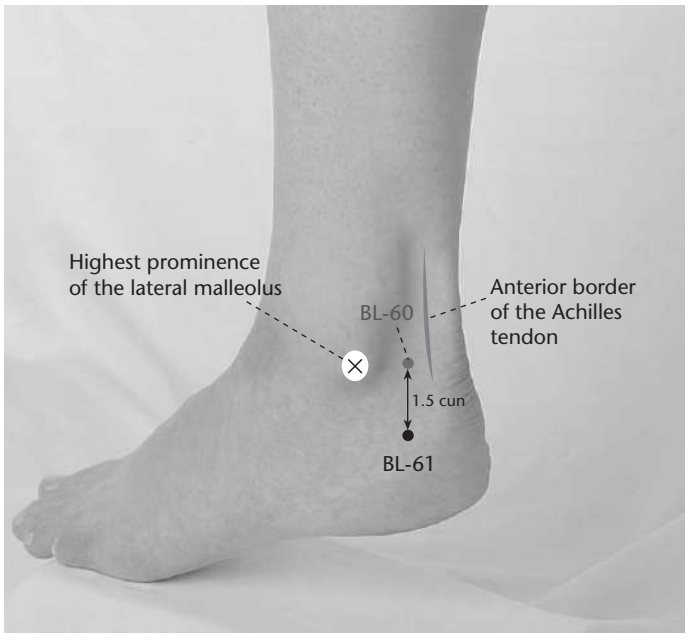

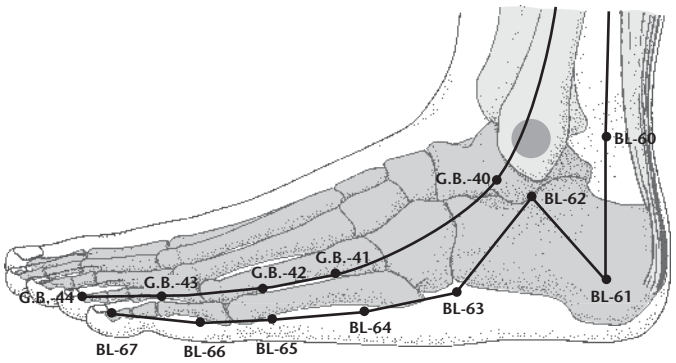

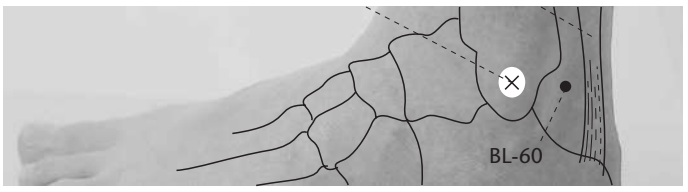

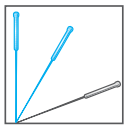

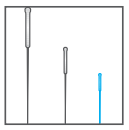

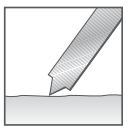

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 309

Extending Vessel SHENMAI BL-62

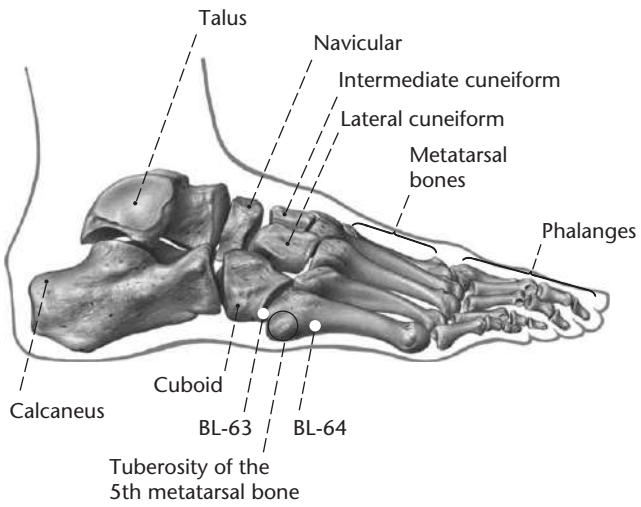

Location

In a depression directly inferior to the highest prominence of the lateral malleolus, above the joint space between the talus and calcaneus.

How to find

First, locate the highest prominence of the lateral malleolus (→ 3.6.2). **BL-62** is located directly below it, in a depression inferior to the lower border of the malleolus and superior to the tendons of the peroneus muscle, in the joint space between the talus and calcaneus.

Needling

Vertically or obliquely 0.3–0.5 cun. The needle may reach the calcaneofibular ligaments superior to the tendons of the peroneus longus and brevis muscles, possibly also the joint space.

Actions/Indications

- Subdues (internal) Wind, clears Heat from the head, calms the *shen*, benefits the head and eyes
- Expels external Wind
- Opens and regulates the *yang qiao mai*
- Opens the channel, alleviates pain
- Moves the Qi in the Bladder sinew channel

Special features

Opening (master) point of the *yang qiao mai*, Sun Si Miao Ghost point.

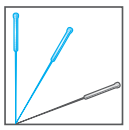
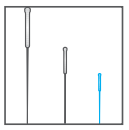

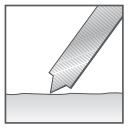

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 310

BL-63 *Golden Gate* JINMEN

Location

On the lateral aspect of the foot, proximal to the tuberosity of the 5th metatarsal bone, in a depression anterior and inferior to → **BL-62** between the calcaneus and the cuboid bone. Note: Some authors locate **BL-63** between the cuboid bone and the tuberosity of the 5th metatarsal bone (see figure on **p. 311**); the more tender point should be selected.

How to find

When palpating along the lateral aspect of the foot on the level of the border of the 'red and white' skin, a bony structure can be felt approximately at the midpoint of the foot. This is the tuberosity of the 5th metatarsal point (→ 3.6.2). Slightly proximal to it (towards the heel), you can palpate a depression between the calcaneus and the cuboid bone. There, locate **BL-63**.
→ **BL-64** is located distal (towards the toes) to the tuberosity of the 5th metatarsal bone.

Needling

Vertically 0.3–0.5 cun

Actions/Indications

- Opens the channel, alleviates pain and eliminates stagnation (*xi*-cleft point)
- Subdues internal Wind and calms the *shen*

Special features

Xi-cleft point, meeting point with the *yang wei mai*.

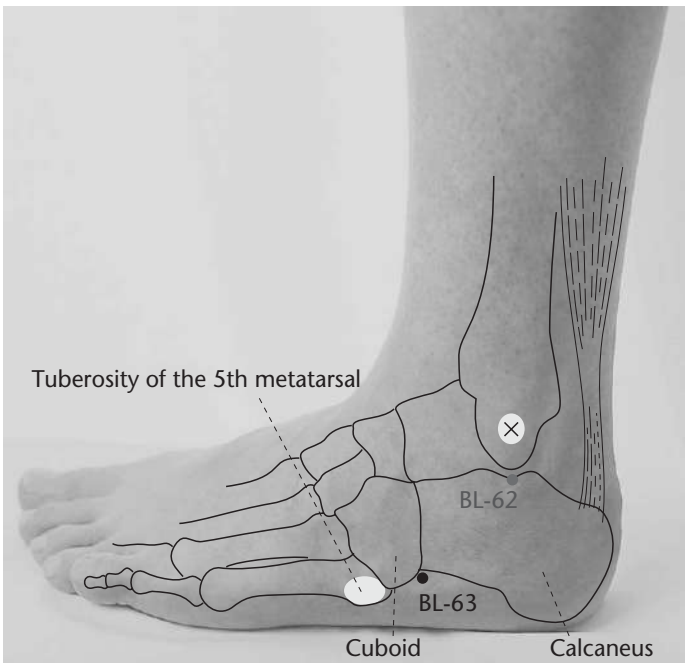

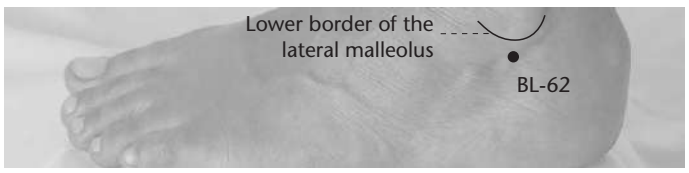

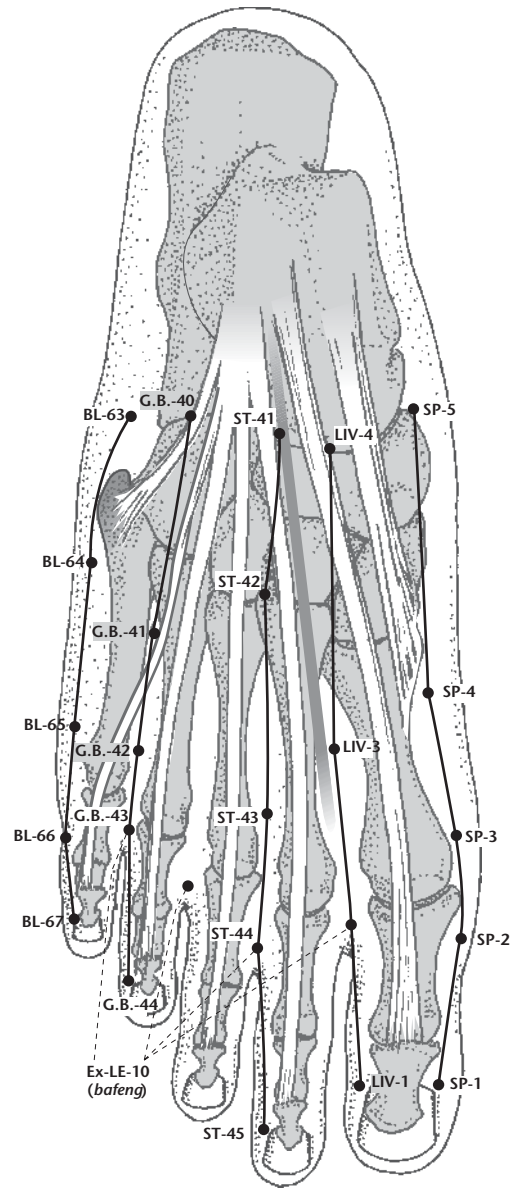

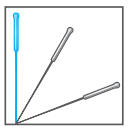
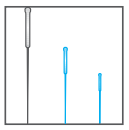

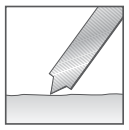

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 311

Capital Bone JINGGU BL-64

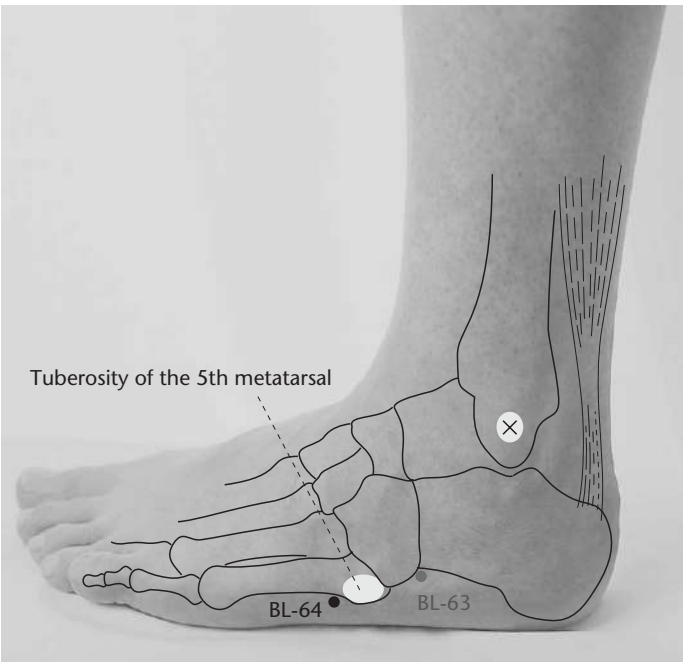

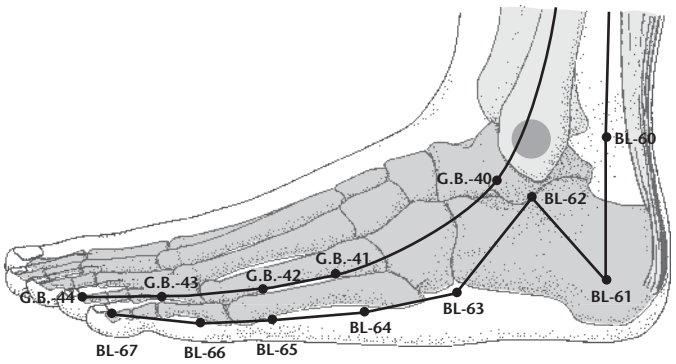

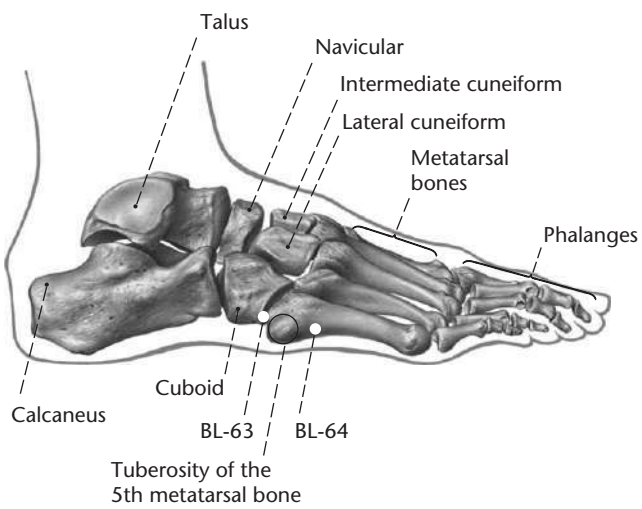

Location

On the lateral aspect of the foot, distal to the tuberosity of the 5th metatarsal bone.

How to find

When palpating along the lateral aspect of the foot on the level of the border of the 'red and white' skin, a bony structure can be felt approximately at the midpoint of the foot. This is the tuberosity of the 5th metatarsal bone (→ 3.6.2). **BL-64** is located directly distal to it (towards the toes), at the junction between the base and the shaft of the 5th metatarsal bone.

→ **BL-63** is located proximal to the tuberosity of the 5th metatarsal bone, in a depression between the calcaneus and cuboid bone (or between the tuberosity and the cuboid bone). → **SP-4** is located in a comparable position on the medial aspect of the foot, in a depression distal to the base of the 1st metatarsal bone.

Needling

Vertically 0.3–0.5 cun

Actions/Indications

- Clears Wind and Heat from the head and eyes
- Calms the *shen*
- Opens the channel

Special features

Yuan-source point

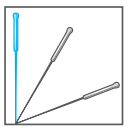
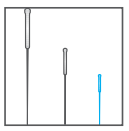

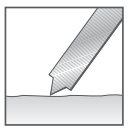

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 312

BL-65 *Restraining Bone* SHUGU

Location

On the lateral aspect of the foot, in the depression proximal to the head of the 5th metatarsal bone.

How to find

When palpating along the bony structures on the lateral aspect of the foot, a prominent bony landmark can be felt approximately at the midpoint of the foot. This is the tuberosity of the 5th metatarsal bone. Distal to it (towards the toes), you can feel a further bony structure, the head of the 5th metatarsal bone. Directly proximal to it, at the border of the 'red and white' skin (border dorsum/sole of the foot), **BL-65** can be palpated in a depression. → **BL-66** is located in a depression distal to the head of the 5th metatarsal bone (distal to the metatarsophalangeal joint), at the junction of the base and the shaft of the proximal phalanx of the little toe.
→ **SP-3** is located in a comparable position on the medial aspect of the foot (proximal to the head of the 1st metatarsal bone). → **S.I.-3** and → **L.I.-3** are located in comparable positions at the ulnar/lateral border of the hand.

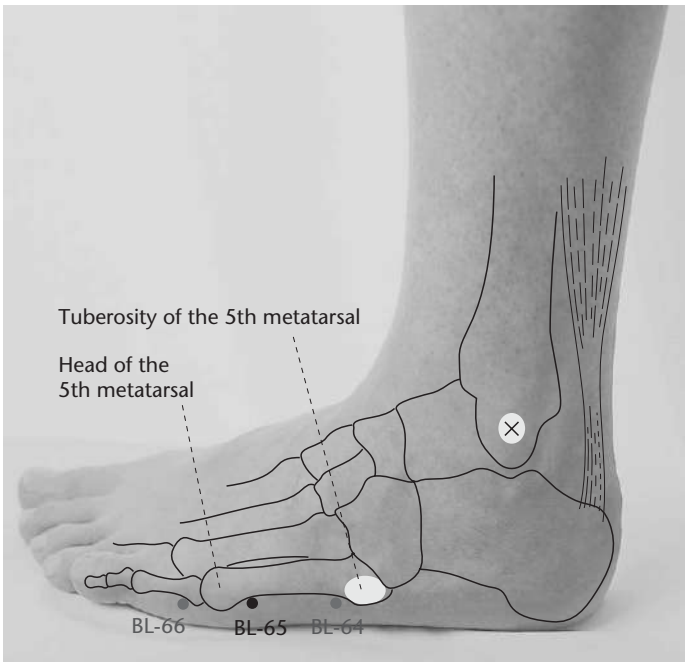

Needling

Vertically 0.3–0.5 cun. Caution: Painful point.

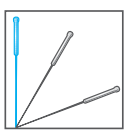

Actions/Indications

- Expels pathogenic factors (especially from the head)
- Clears Heat
- Opens the channel and *luo* vessels, alleviates pain
- Calms the *shen*

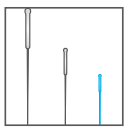

Special features

Shu-stream point, Wood point, sedation point.

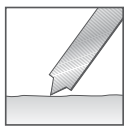

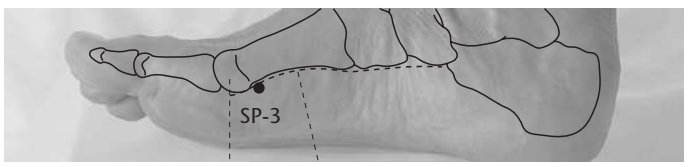

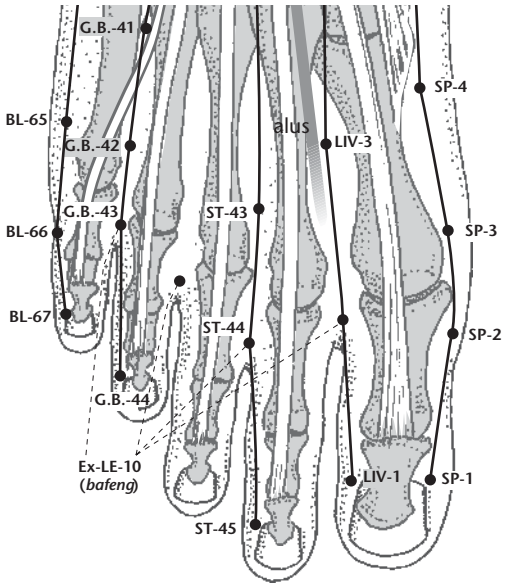

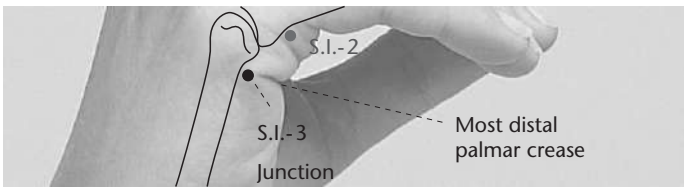

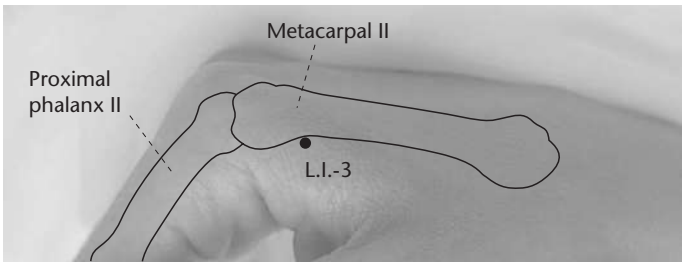

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 313

Foot Connecting Valley ZUTONGGU BL-66

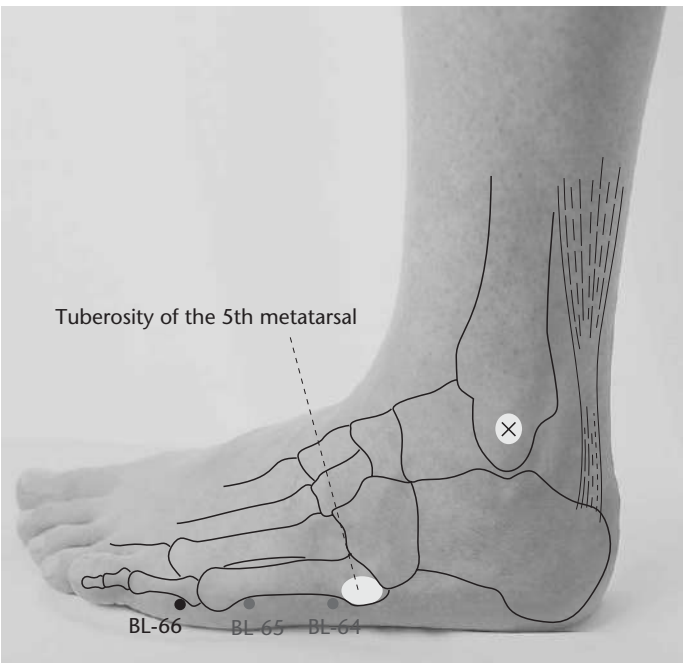

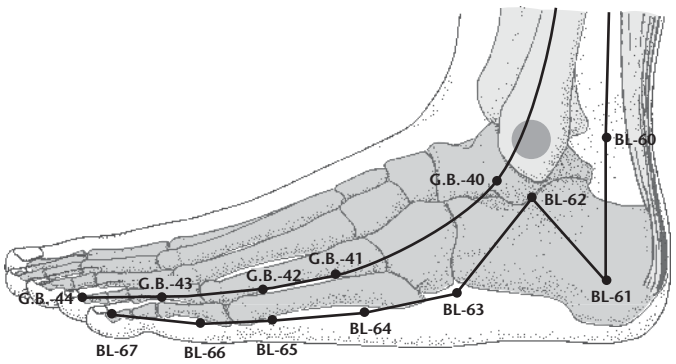

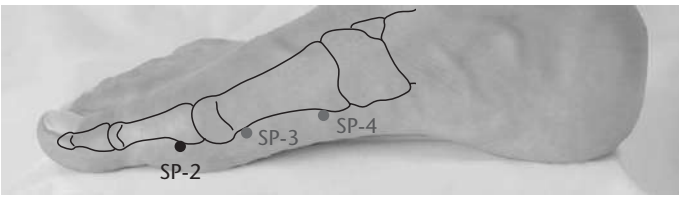

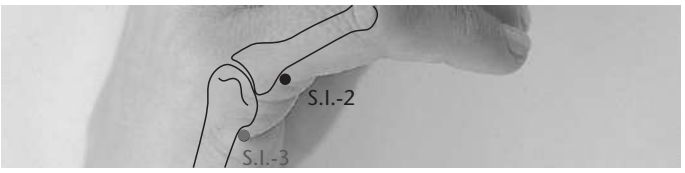

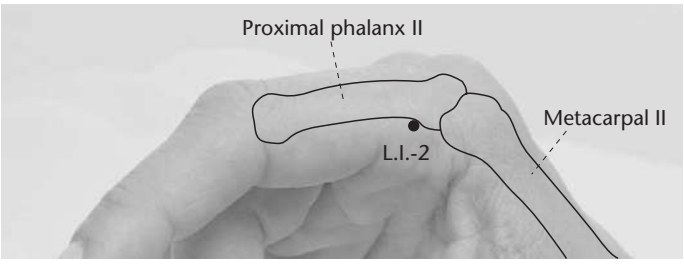

Location

At the lateral border of the foot, in the depression distal to the metatarsophalangeal joint of the little toe.

How to find

When palpating along the border of the 'red and white' skin on the lateral aspect of the foot, at approximately the midpoint of the length of the foot, a clearly defined bony structure can be felt. This is the tuberosity of the 5th metatarsal bone. Distal to it (towards the toes), you can feel a further prominent bony structure, the head of the 5th metatarsal bone (or the 5th metatarsophalangeal joint). **BL-66** is located directly distal to it, in a depression at the junction of the base and the shaft of the proximal phalanx of the little toe.

→ **BL-65** is located proximal to the head of the 5th metatarsal bone. → **SP-2** is located in a comparable position on the medial aspect of the foot. Located in comparable positions on the hand are → **S.I.-2** (on the ulnar aspect) and → **L.I.-2** (on the radial aspect).

Needling

Vertically 0.3–0.5 cun. Caution: Painful point.

Actions/Indications

- Clears the head
- Descends the Lung and Stomach Qi
- Calms the *shen*

Special features

Ying-spring point, Water point, *ben* point (Five Phases point).

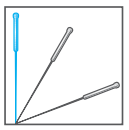
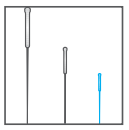

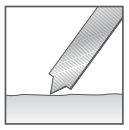
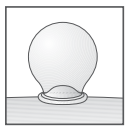

Ch04.7b-F10028.qxd 2/22/08 5:01 PM Page 314

BL-67 *Reaching Yin* ZHIYIN

Location

On the dorsal aspect of the little toe, 0.1 cun from the lateral corner of the nail.

How to find

This point is located at the junction of two tangents that border the nail of the little toe laterally and proximally, approximately 0.1 cun from the actual corner of the nail.

Needling

Vertically 0.1 cun or obliquely 0.2 cun in the direction of the foot. Prick to bleed. Caution: Painful point.

Actions/Indications

- Scientifically established: conversion of breech presentations prior to birth: moxibustion and electro-acupuncture at this point seem to be most effective (approximately 70–80%)
- Facilitates labour
- Expels pathogenic factors from the opposite end of the channel
- Regulates Yin and Yang of the Water element

Special features

Jing-Well point, Metal point, tonification point, exit point. An isolated case has been reported regarding a pregnant woman who was treated with moxibustion at **BL-67**. This resulted in a fetomaternal transfusion of 300 ml of blood, endangering the fetus. For this reason, treatments should be controlled by cardiotocography.

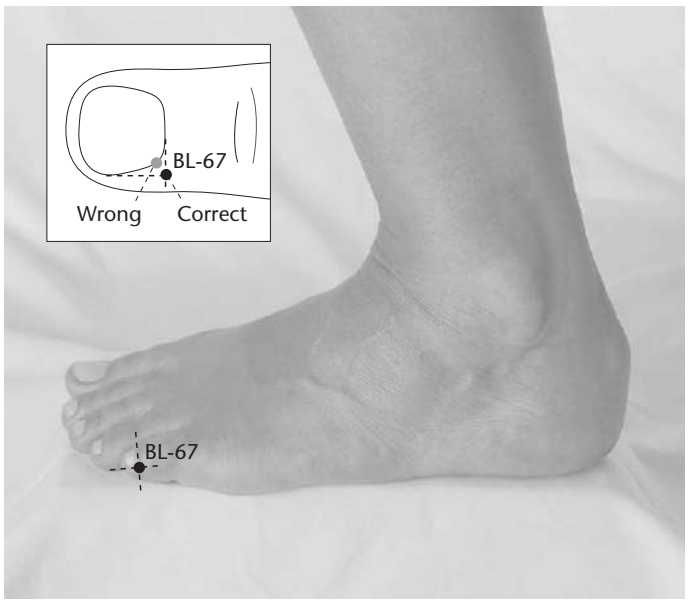

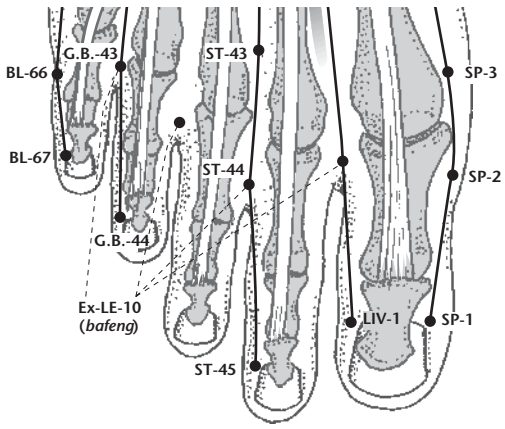

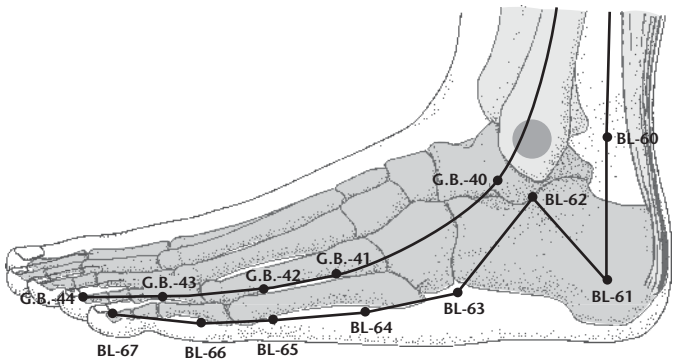

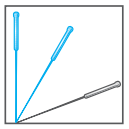
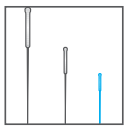

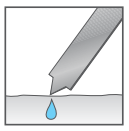

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 315

4.8 The Kidney Channel System – Foot *shaoyin* (*zu shao yin jing luo*)

4.8.1 The Kidney Primary Channel (*zu shao yin jing*)

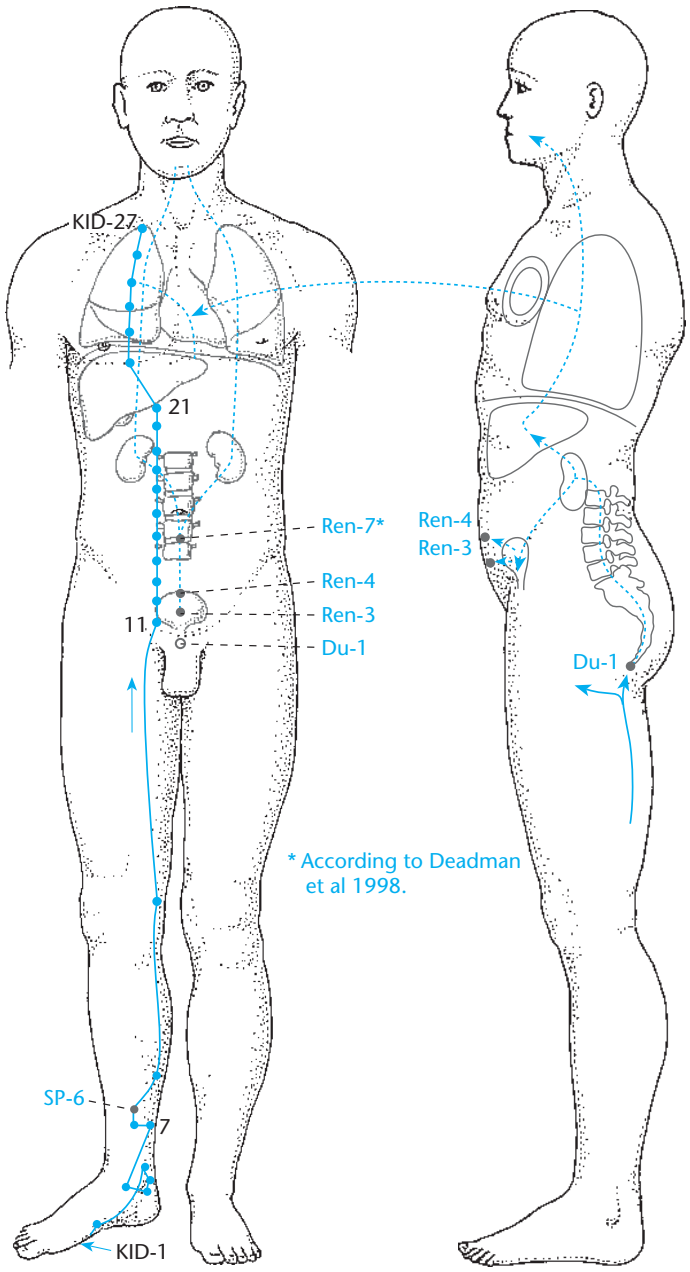

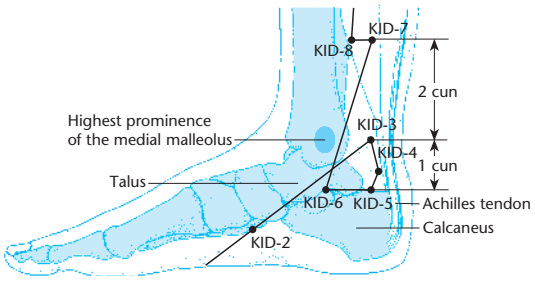

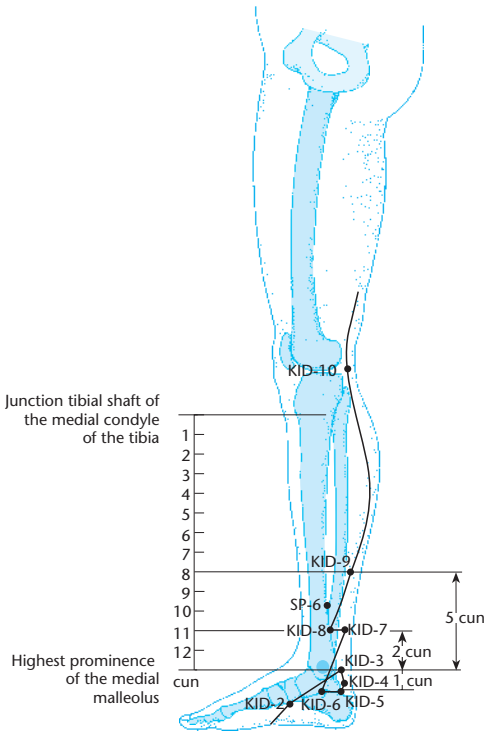

Pathway

The Kidney primary channel begins beneath the little toe, which is reached by **a branch** separating from the Bladder primary channel at its endpoint at → **BL-67** (*zhiyin*) (foot Yin–Yang connection of the second great circuit).

The Kidney primary channel diagonally crosses the sole of the foot to **KID-1** (*yongquan*) and to **KID-2** (*rangu*) below the navicular tuberosity, continues to the posterior aspect of the medial malleolus and descends to enter the heel. From here it ascends the medial aspect of the lower leg, intersecting with → **SP-6** (*yinlingquan*) and continuing to ascend the postero-medial aspect of the thigh towards the perineum.

In the perineal region the primary channel divides into **two branches**:

- the **deep, internal** branch travels to → **Du-1** (*changqiang*) and ascends alongside the spine to connect with its pertaining *zang*-Organ, the Kidneys (*shen*), and its paired *fu*-Organ, the Bladder (*pangguang*). Here it divides into smaller branches which connect with → **Ren-4** (*guanyuan*) and → **Ren-3** (*zhongji*), and, according to some authors, also with → **Ren-7** (*yinjiao*). From the Kidney (*shen*) a branch ascends to the Liver (*gan*), runs to and spreads in the Lung (*fei*), continues to the trachea and terminates at the root of the tongue. From the Lung (*fei*), an **inner branch** travels to the Heart (*xin*), where it meets the Pericardium primary channel (deep Yin–Yin connection). It then disperses in the chest and reaches → **Ren-17** (*shanzhong*).
- the **external** branch separates in the perineal region and runs to the lower abdomen at **KID-11** (*henggu*). From the pubic

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 316

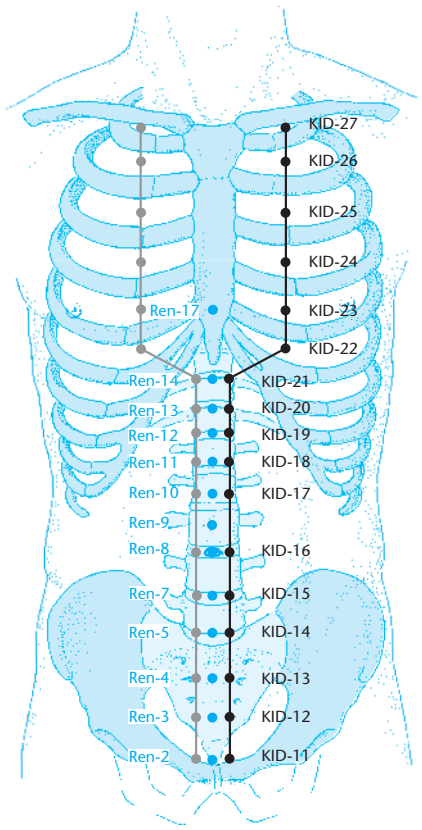

bone it ascends initially 0.5 cun lateral to the anterior midline, from the 5th intercostal space (**KID-22**) 2 cun lateral to the midline to the infraclavicular fossa.

Clinical importance (→ 1.2)

Exterior (*biao*) signs and symptoms: Pain in the lumbar region, counterflow Cold or weakness of the legs, dry mouth, sore throat, pain in the lateral gluteal region and on the posterior thigh, pain on the soles of the feet.

Interior (*li*) or *zangfu*-Organ signs and symptoms: Pain, facial oedema, dark rings under the eyes, shortness of breath, somnolence, restlessness, diarrhoea, thin or dry stools, abdominal distension, nausea and vomiting, impotence.

Connections of the Kidney primary channel

Connections with other channels

Bladder primary channel (*zu tai yang jing*)

Connection: Foot Yin–Yang connection of the second great circuit
Location: BL-67 → KID-1 (on the foot)
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Heart primary channel (*shou shao yin jing*)

Connection: Paired according to the Six-Channel theory (hand–foot pairing): *shao yin* (Yin axes of the second great circuit)
Location: KID → HE (on the thorax). A branch of the internal pathway of the Kidney primary channel travels from the Kidney to the Liver, penetrates the diaphragm and spreads in the Lung (*fei*). From the Lung a branch runs to the Heart (*xin*), there connecting with the Heart primary channel.
Circulation: Non-circadian (**not** according to the Organ clock)
Importance: Above–below relationship

Pericardium primary channel (*shou jue yin jing*)

Connection: Deep Yin–Yin connection
Location: KID → P. A branch of the internal pathway of the Kidney primary channel travels from the Kidney to the Liver, penetrates the diaphragm and spreads in the Lung (*fei*). From the Lung an internal branch courses to the Heart (*xin*), where it joins the Pericardium primary channel (deep Yin–Yin connection) and also reaches → **Ren-17** (*shanzhong*).
Circulation: Circadian (according to the Organ clock)
Importance: The Pericardium primary channel receives Nutritive Qi (*ying qi*) from the Kidney primary channel (first circulation of the *ying qi* → 1.1.4).

Connections with other *zangfu*-Organ systems

Kidney (***shen***), **Bladder** (***pangguang***), Liver (*gan*), Lung (*fei*), Pericardium (*xin bao*), Heart (*xin*)

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 317

4.8.2 The Kidney Divergent Channel (*zu shao yin jing bie*)

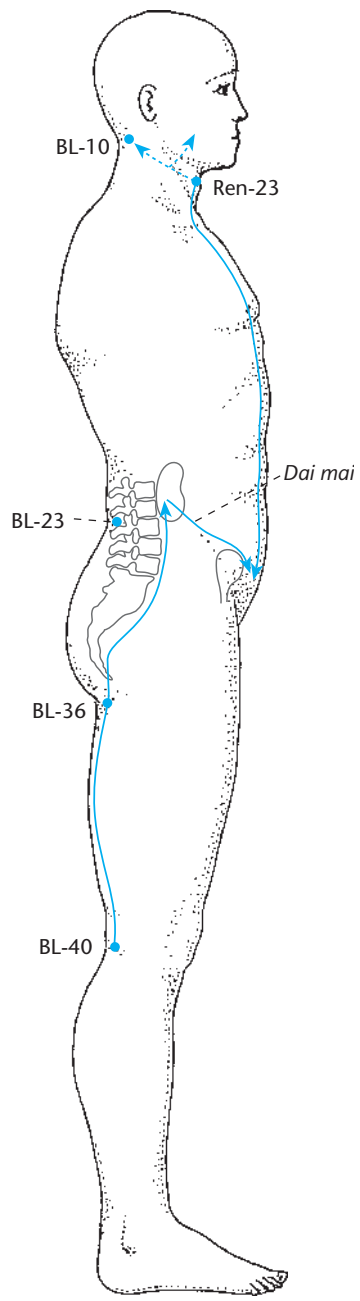

Pathway

The Kidney divergent channel separates from the Kidney primary channel in the popliteal fossa at ➞ **KID-10** (*yinggu*). It

- ➡ meets ➞ **BL-40** (*weizhong*) in the popliteal fossa
- ➡ ascends to ➞ **BL-36** (*chengfu*) in the centre of the gluteal crease
- ➡ penetrates the anus and reaches the Kidney (*shen*) and the Bladder (*pangguang*)
- ➡ ascends to ➞ **BL-23** (*shenshu*) at the level of the lower border of the spinous process of L2, where it also meets the *dai mai*, following it round the waist to the abdomen
- ➡ ascends lateral to the midline to the infraclavicular fossa
- ➡ reaches the neck at ➞ **Ren-23** (*liangquan*)
- ➡ emerges at the nape of the neck at ➞ **BL-10** (*tianzhu*) and converges with the Bladder primary channel and the Bladder divergent channel to form one of the 6 *he*-confluences (here: BL/KID as first confluence ➞ 1.3).

Clinical importance

- Strengthens the relationship between the Kidneys and the Bladder (*zangfu*-Organ systems). Points on the Kidney primary channel can therefore be used for disorders of the Bladder, and vice versa points on the Bladder primary channel can treat disorders of the Kidneys.
- Through the connection with the *dai m*ai, KID points can be used for disorders of the *dai mai*.
- Kidney Qi and Essence (*jing*) are sent to the Brain through the Kidney divergent channel via the spinal canal, thus strengthening the connection between the Kidneys, the Marrow and the Brain.
- Supports the connection to the tongue.

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 318

4.8.3 The Kidney Sinew Channel (*zu shao yin jing jin*)

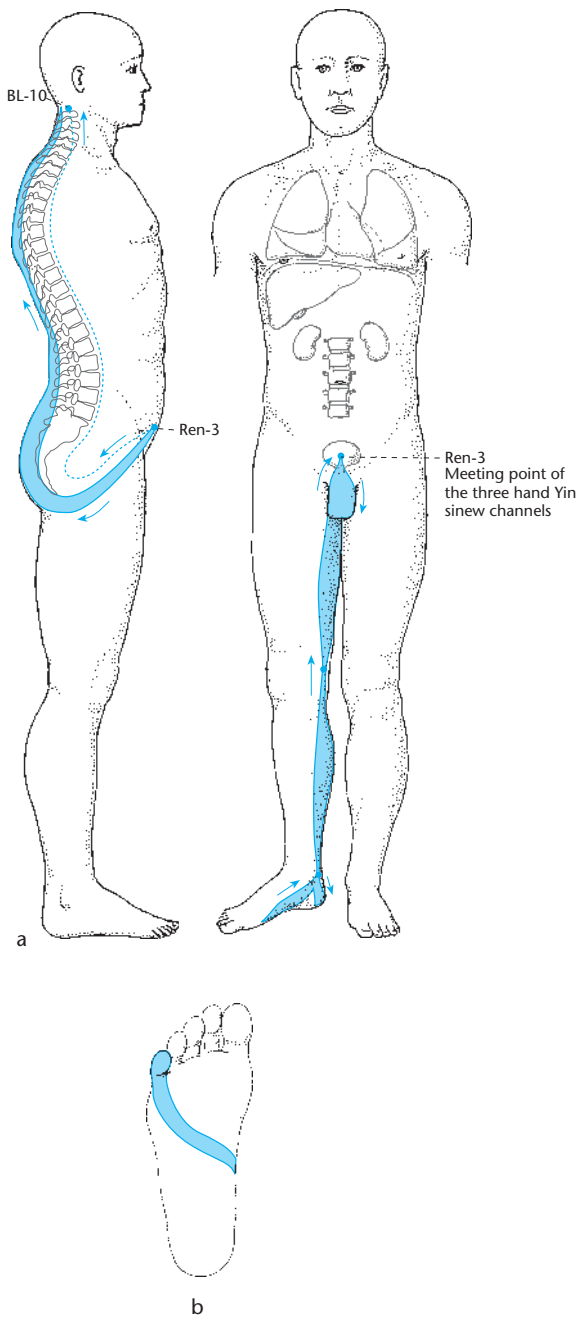

Pathway

The Kidney sinew channel originates beneath the little toe

- diagonally crosses the sole of the foot, meets the Spleen sinew channel and continues to the medial malleolus where it binds (*jie*). From here a **minor branch** travels to the medial portion of the calcaneus. The **main branch** ascends the posteromedial aspect of the leg, binds (*jie*) at the medial condyle of the femur, crosses the Bladder sinew channel, ascends the posteromedial aspect of the thigh to the genital region to → **Ren-3** (*zhongji*) and → **Ren-2** (→ 1.3), where it meets the other Yin sinew channels. It then penetrates the abdomen, runs along the spine to the nape of the neck and there meets the Bladder sinew channel.

From the genital region **an internal branch** traverses the gluteal region and follows the deep musculature along the spine to the occiput, where it meets the Bladder sinew channel.

Clinical importance

Pathology: Spasms on the sole of the foot, pain or spasms along the posteromedial aspect of the lower extremity, the back and the occiput. Chronic *bi*-syndromes. Epileptic seizures, limited range of motion with spinal flexion and extension. Degenerative disorders of the bones and joints, spondylosis, osteoarthritis.

Indication: Mainly for chronic pain, spasms and Cold sensations of the lower back and spine, accompanied by limited range of motion.

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 319

4.8.4 The Kidney *luo*-Connecting Vessel System (*zu shao yin luo mai*)

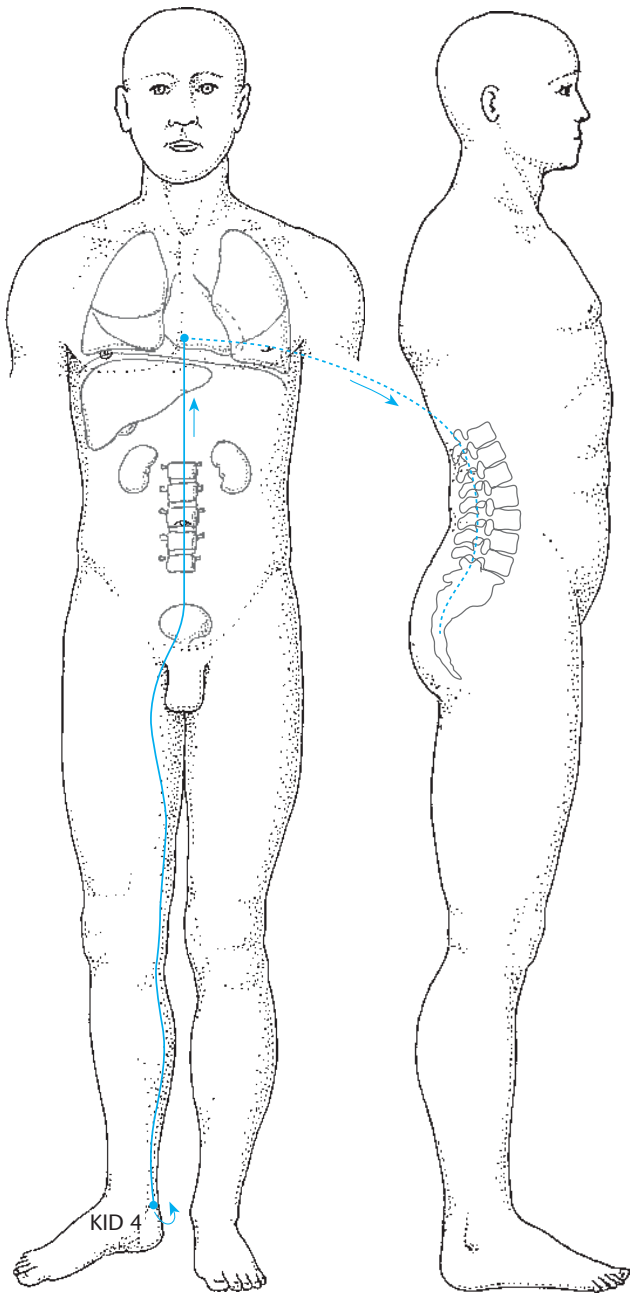

Pathway

The Kidney *luo*-connecting channel separates from the Kidney primary channel at **KID-4** (*dazhong*). It forms a three-dimensional reticular network, dividing into multiple branches and sub-branches (*sun luo*, *fu luo*, *xue luo* → 1.5) within the surrounding tissue.

➡ Horizontal divisions run to the Interiorly–Exteriorly paired Bladder primary channel; according to some schools of thought (for example Ngyen Van Nghi → Appendix) they travel as a **transverse** KID *luo*-connecting vessel to the *yuan*-source point **BL-64** (*jinggu*).

➡ A **longitudinal** division follows the Kidney primary channel to → **Ren-17** (*shanzhong*), penetrates the thorax and reaches the spine in the lumbar region.

Clinical importance (→ 8.1.2)

Pathology
Excess (*shi*): Retention of urine
Deficiency (*xu*): Pain in the lumbar region
Counterflow Qi: Restlessness, anxiety, fear, depression, fullness in the chest and epigastrium.

4.8.5 Cutaneous Region (*shao yin pi bu*)

See description and figures → 1.6.

4.8.6 Points of the Kidney Primary Channel (Overview)

Specific points according to their functions

- ***Yuan*-source point (→ 8.1.1): KID-3** (*taixi*) ■■
- ***Luo*-connecting point (→ 8.1.2.): KID-4** (*dazhong*)
- ***Xi*-cleft point (→ 8.1.3): KID-5** (*shuiquan*)
- **Associated Back-*shu* point (→ 8.1.4): BL-23** (*shenshu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): G.B.-25** (*jingmen*) ■■
- **Five *shu*-transporting point (→ 8.1.6):**
  jing-well point (Wood), sedation point: **KID-1** (*yongquan*)
  ying-spring point (Fire): **KID-2** (*rangu*)
  shu-stream point (Earth): **KID-3** (*taixi*) ■■
  jing-river point (Metal), tonification point: **KID-7** (*fuliu*) ■■
  he-sea point (Water), *ben* point (Five Phases): **KID-10** (*yingu*) ■■
- *Hui*-meeting point (→ 8.1.7): –
- Opening point (→ 8.1.8) of the *yin qiao mai*: **KID-6** (*zhaohai*) ■■
- Lower *he*-sea points (→ 8.1.9): –
- *Jiaohui*-meeting points (→ 8.1.10):
  – with the *yin qiao mai*: **KID-2** (*rangu*)*, (**KID-6** (*zhaohai*)*, **KID-8** (*jiaoxin*)*)
  – with the *yin wei mai*: **KID-9** (*zhubin*)
  – with the *chong mai*: **KID-11–KID-21**
  – of other channels with the Kidney channel: **SP-6**, **Du-1**, **Ren-4**, **Ren-3**, **Ren-7***, **Ren-17**
- **Gao Wu command point (→ 8.1.11): –**
- **Window of Heaven point (→ 8.1.12): –**
- **Points of the Four Seas (→ 8.1.13): –**

* Mentioned by only some authors

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 320

- **Ma Dan Yang Heavenly Star point (→ 8.1.14): –**
- **Sun Si Miao Ghost point (→ 8.1.15): –**
- **Other functional points:**
  – *xi*-cleft point of the *yin qiao mai*: **KID-8** (*jiaoxin*)
  – *xi*-cleft point of the *yin wei mai*: **KID-9** (*zhubin*)

Points according to region

- **Local points** (→ 8.2.1): foot – **KID-6** (*zhaohai*) ■■; knee – **KID-10** (*yingu*) ■■
- **Adjacent points** (→ 8.2.1): foot – **KID-7** (*fuliu*) ■■; toes – **KID-6** (*zhaohai*) ■■
- **Distal points** (→ 8.2.1): for the throat – **KID-6** (*zhaohai*) ■■; for the tongue – **KID-6** (*zhaohai*) ■■; for the Kidneys and the Bladder – **KID-3** (*taixi*) ■■, **KID-7** (*fuliu*) ■■; for the genito-urinary region – **KID-3** (*taixi*) ■■

Specific points according to the channel pathway (in numerical order)

- **KID-1** (*yongquan*) ■■: *jing*-well point (Wood); sedation point
- **KID-2** (*rangu*): *ying*-spring point (Fire); *jiaohui*-meeting point with the *yin qiao mai* (→ 8.1.10)
- **KID-3** (*taixi*) ■■: *shu*-stream point (Earth); distal point for the Kidneys, Bladder, urogenital region (→ 8.2.1); *yuan*-source point (→ 8.1.1)
- **KID-4** (*dazhong*): *luo*-connecting point (→ 8.1.2)
- **KID-5** (*shuiquan*): *xi*-cleft point (→ 8.1.3)
- **KID-6** (*zhaohai*) ■■: opening point (→ 8.1.8) of the *yin qiao mai*; distal point for the tongue and throat (→ 8.2.1); *jiaohui*-meeting point with the *yin qiao mai** (→ 8.1.10); local point for the foot (→ 8.2.1)
- **KID-7** (*fuliu*) ■■: distal point for the Kidneys and the Bladder (→ 8.2.1); *jing*-river point (Metal); tonification point; regional point for the foot (→ 8.2.1)
- **KID-8** (*jiaoxin*): *xi*-cleft point of the *yin qiao mai*; *jiaohui*-meeting point with the *yin qiao mai** (→ 8.1.10)
- **KID-9** (*zhubin*): *xi*-cleft point of the *yin wei mai*; *jiaohui*-meeting point with the *yin wei mai** (→ 8.1.10)
- **KID-10** (*yingu*) ■■: *he*-sea point (Water), *ben* point (Five Phases); local point for the knee (→ 8.2.1)
- **KID-11** to **KID-21**: *jiaohui*-meeting points with the *chong mai* (→ 8.1.10)

General location help

- **KID-11** to **KID-21** are all located 0.5 cun lateral to the midline
  – **KID-11** to **KID-15** are spread on the 5 cun distance (→ 1.2) between the upper border of the pubic symphysis and the umbilicus (they are spaced at 1 cun intervals from the upper border of the pubic symphysis to 4 cun above it)
  – **KID-16** to **KID-21** are spread on the 8 cun distance (→ 1.2) between the umbilicus and the sternocostal angle (they are spaced at 1 cun intervals from the level of the umbilicus to 6 cun superior to it).
- **KID-22** to **KID-27** are all located 2 cun lateral to the midline
  – **KID-22** to **KID-26** are located in the 5th to 1st intercostal spaces
  – **KID-27** is located on the lower border of the sternoclavicular joint

* Mentioned by only some authors

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 321

Gushing Spring YONGQUAN KID-1

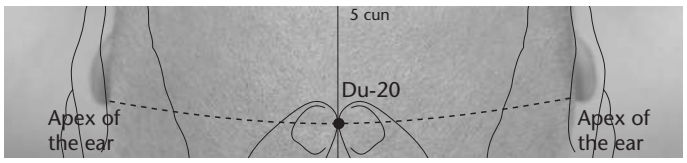

Location

On the sole of the foot, in a depression between the 2nd and 3rd metatarsal bones, at the junction of the anterior third and the posterior two-thirds of the sole.

How to find

Measure one third from the anterior border of the sole of the foot. Palpate for a pressure-sensitive point dorsal to the balls of the foot, at the midpoint of the width of the foot.

Needling

Needling this point is very painful and is therefore only recommended for severe disorders or in an emergency. Acupressure is commonly used instead; moxibustion is possible. For the treatment of collapse, unconsciousness, shock or a severe excess condition, this point should be strongly stimulated; in pronounced deficiency conditions, stimulation should be applied more cautiously.

Actions/Indications

- Revives collapsed Yang
- Clears excess, Heat and Wind from the head, lowers Yang
- Calms the *shen*

Special features

Jing-well point, Wood point, sedation point, entry point. **KID-1** is the lowest acupuncture point on the body and the only point on the sole of the foot. For this reason, it plays an important role in Qigong, forming a grounding point in the centre of the foot, where the energies of the Earth and Man unite. During practice, the centre of gravity should be above this point.

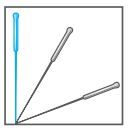
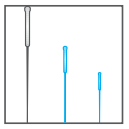

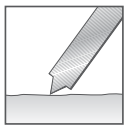

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 322

KID-2 *Blazing Valley* RANGU

Location

At the medial border of the foot, in a depression at the anterior border of the navicular bone, at the border of the 'red and white' skin.

How to find

Palpate from distal to proximal along the tarsal section of the medial aspect of the foot, past the shaft and the head of the 1st metatarsal bone (→ **SP-4**) and the medial cuneiform bone, finally reaching the prominent navicular bone. Locate **KID-2** anterior to the navicular bone, at the inferior angle of the joint between the medial cuneiform bone and the navicular bone.

Needling

From the medial aspect of the foot, 0.5–1 cun vertically below the border of the bone. Use reducing needling techniques for empty Heat. Use tonifying techniques for Kidney Yang/Qi deficiency.

Actions/Indications

- Clears empty Heat
- Regulates the Lower Burner and the Kidneys
- Local point

Special features

Ying-spring point, Fire point, meeting point with the *yin qiao mai*.

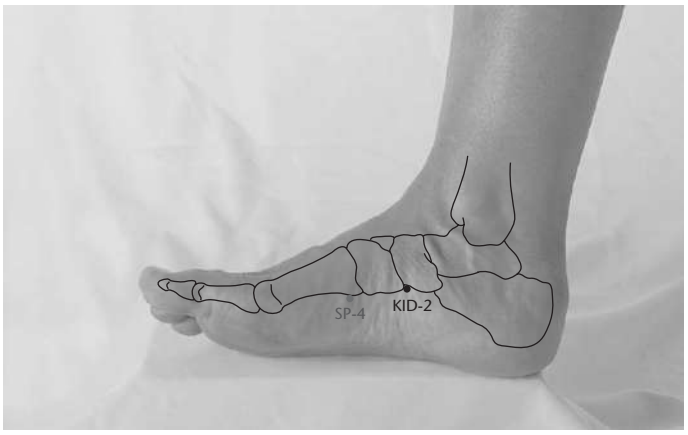

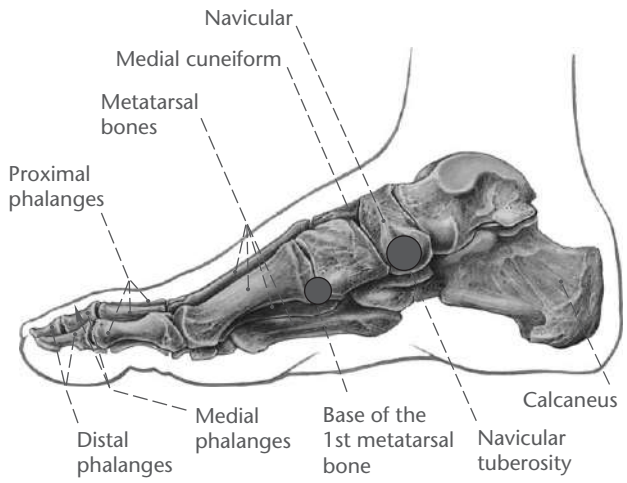

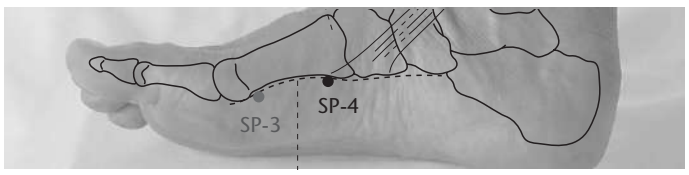

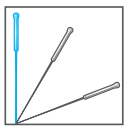
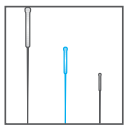

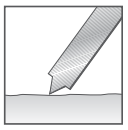

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 323

Supreme Stream TAIXI KID-3

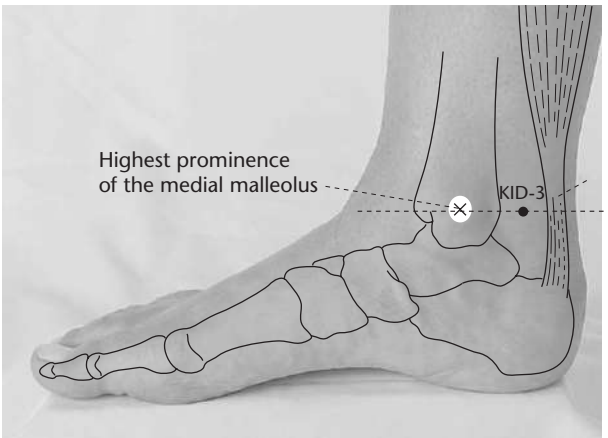

Location

In the depression between the highest prominence of the medial malleolus and the Achilles tendon.

How to find

Locate the highest prominence of the medial malleolus (→ 3.6.2). From there, palpate horizontally towards the Achilles tendon. **KID-3** is located in a depression that can be palpated anterior to the tendon.

Needling

Vertically 0.3–1 cun

Actions/Indications

- Nourishes Kidney Yin and clears deficiency Heat, tonifies Kidney Yang, stabilises the Kidney Qi and the Lungs (helping the Kidneys absorb the Qi), regulates menstruation
- Strengthens the lower back, local point

Special features

Yuan-source point, *shu*-stream point, Earth point. Important point for tonifying the Kidneys (especially Kidney Yin).

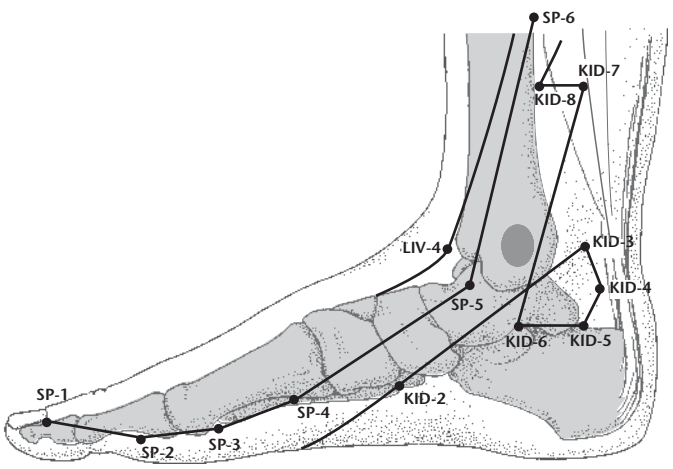

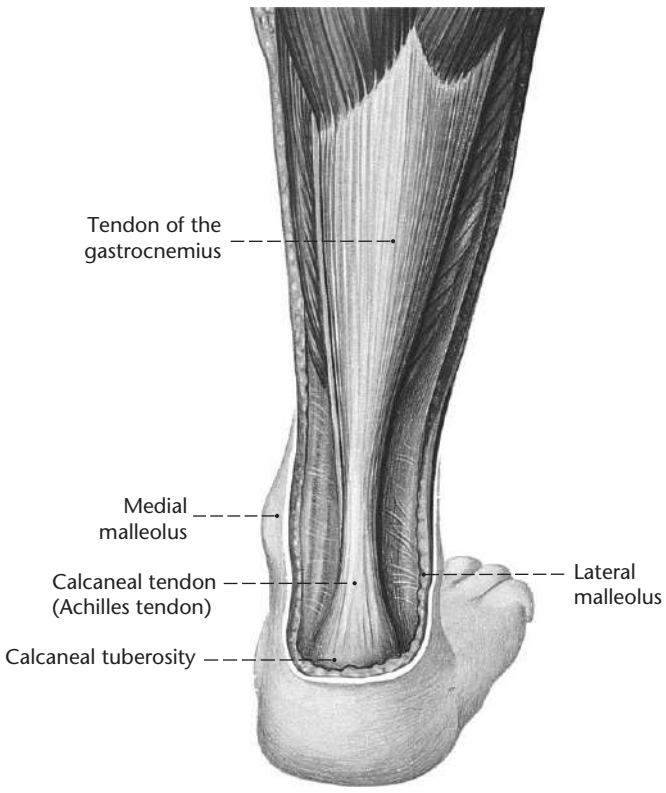

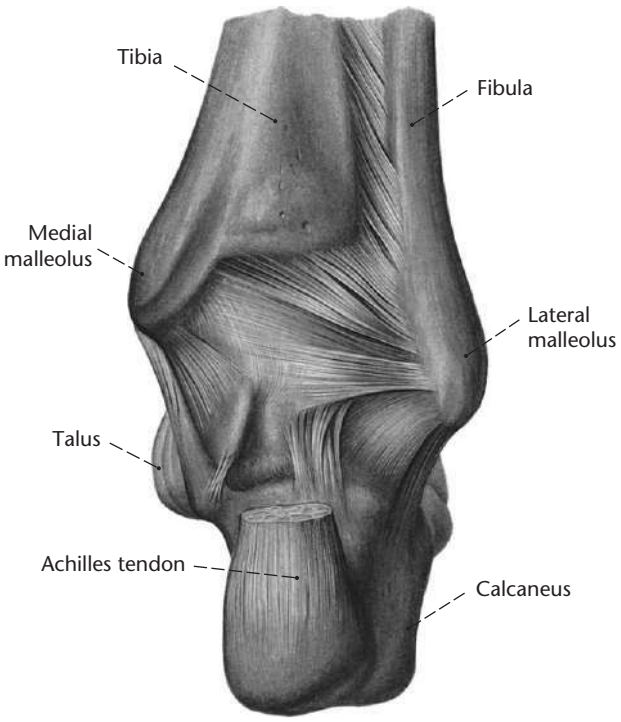

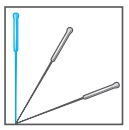
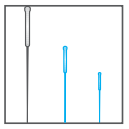

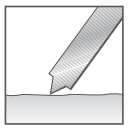
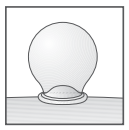

Ch04.8-F10028.qxd 2/22/08 6:13 PM Page 324

KID-4 *Great Bell* DAZHONG

Location

Anterior to the medial border of the Achilles tendon, superior to its insertion at the calcaneus.

How to find

From the highest prominence of the medial malleolus, draw a horizontal line to the medial border of the Achilles tendon. From there, measure 0.5 cun in a distal direction. **KID-4** is located in a depression anterior to the Achilles tendon, slightly superior to its insertion at the calcaneus. **Or: KID-4** is located posterior to the midpoint of a line connecting → **KID-3** and → **KID-5** anterior to the Achilles tendon.

Needling

Vertically 0.3–0.5 cun. Avoid puncturing the tendon.

Actions/Indications

- Firms the Kidney Qi (helping the Kidneys absorb the Qi) and supports the Lungs
- Strengthens the Kidneys and cools empty Heat
- Strengthens the will, dispels fear
- Local point

Special features

Luo-connecting point

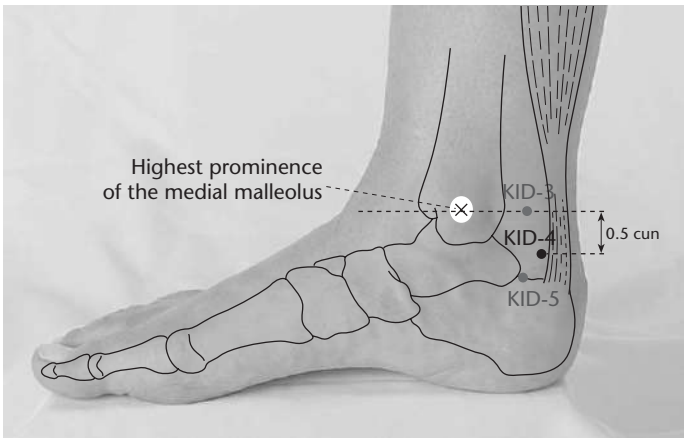

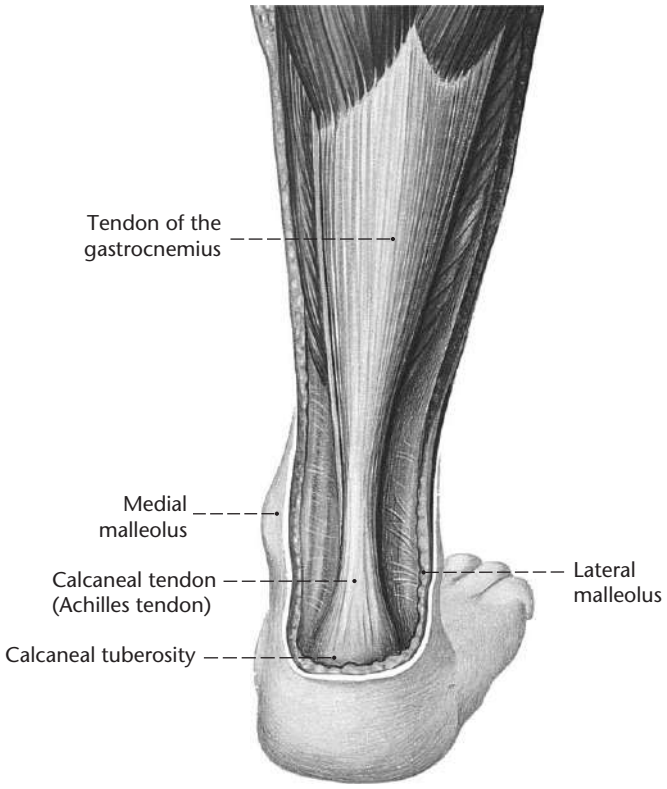

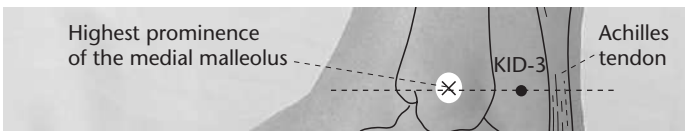

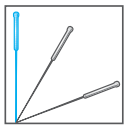
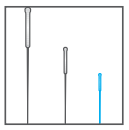

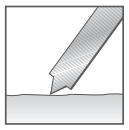

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 325

Water Spring SHUIQUAN KID-5

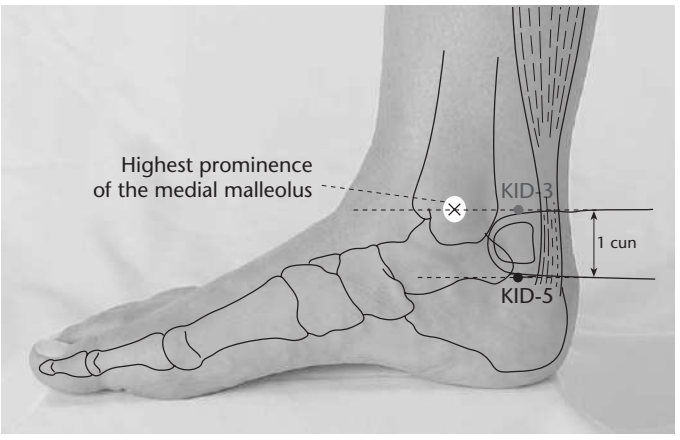

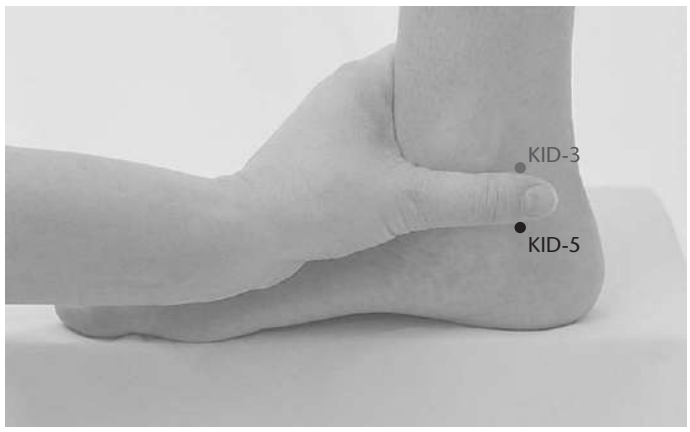

Location

1 cun distal to ➞ **KID-3**, in a depression over the joint space between the talus and calcaneus.

How to find

First, locate ➞ **KID-3** on the level of the highest prominence of the medial malleolus (➞ 3.6.2), in a depression between the malleolus and the Achilles tendon. From **KID-3**, palpate 1 cun in an inferior direction to a palpable depression over the joint space between the talus and calcaneus. This is the location of **KID-5**.

Needling

0.3–0.5 cun obliquely to vertically superior to the margin of the bone.

Actions/Indications

- Regulates the *chong mai* and *ren mai*, benefits menstruation and urination

Special features

Xi-cleft point. Especially indicated for acute disorders and painful conditions.

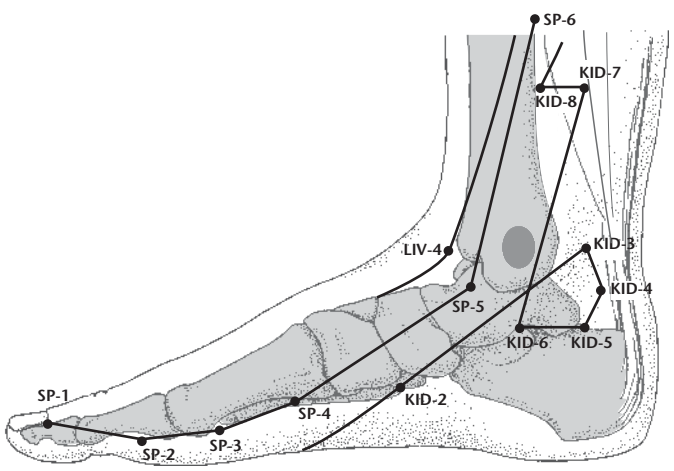

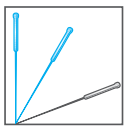
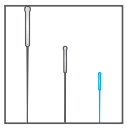

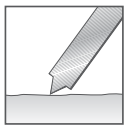

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 326

KID-6 *Shining Sea* ZHAOHAI

Location

Approximately 1 cun inferior to the highest prominence of the medial malleolus, over the joint space between the talus and the calcaneus.

How to find

From the highest prominence of the medial malleolus (→ 3.6.2), palpate distally, until the palpating finger can feel the depression of the joint space between the talus and the calcaneus. (**KID-6** is often located between the tendons of the tibialis and flexor digitorum longus muscles.) Directly distal to this point, you can palpate a small bony prominence, the sustentaculum tali. With the patient's foot in a supinated position, a visible skin crease forms on the level of the point. Note: There are several variations regarding the exact location of this point. However, the determining aspect for its location is not the cun measurement but the joint space below the prominence.

Needling

0.3–0.5 cun vertically or obliquely in a proximal direction. The needle may reach the deltoid ligament.

Actions/Indications

- Nourishes Kidney Yin, clears empty Heat, benefits the throat, regulates the Lower Burner
- Calms the *shen*
- Moves Qi locally

Special features

Opening (master) point of the *yin qiao mai*. Important point, especially for strengthening the Kidney Yin.

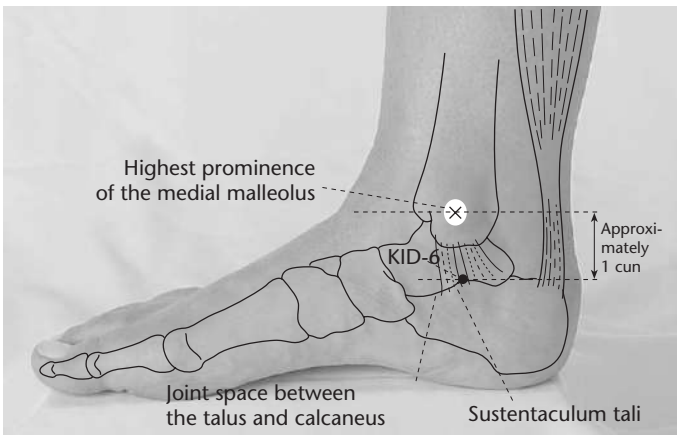

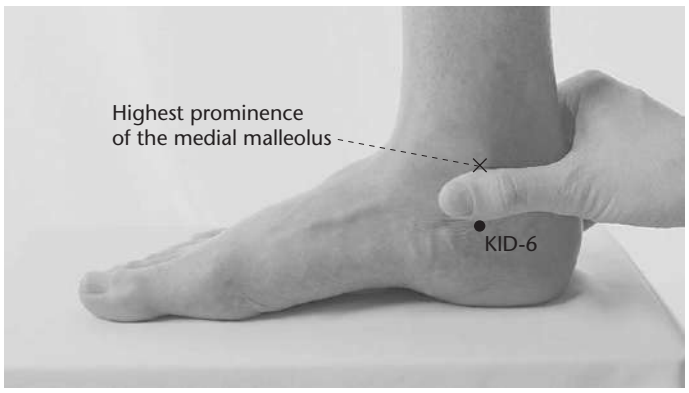

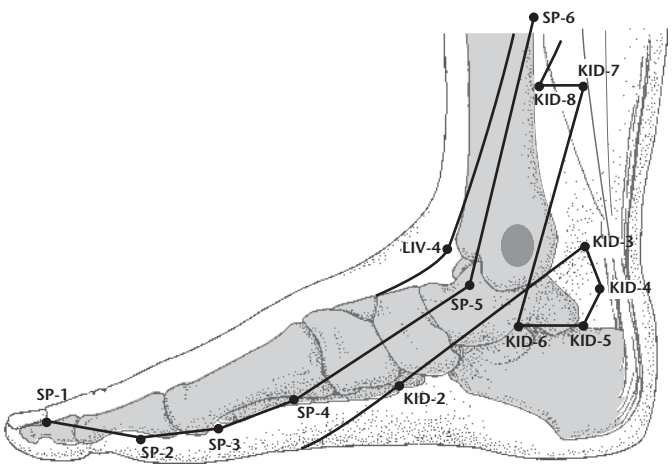

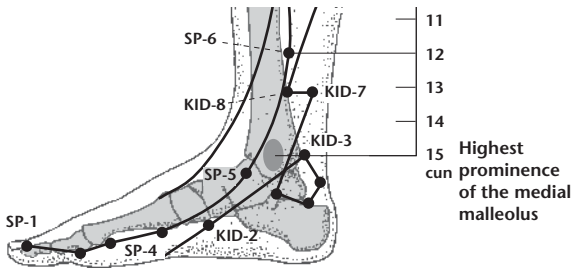

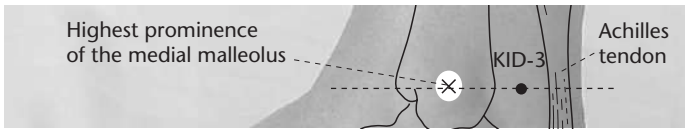

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 327

Returning Current FULIU KID-7

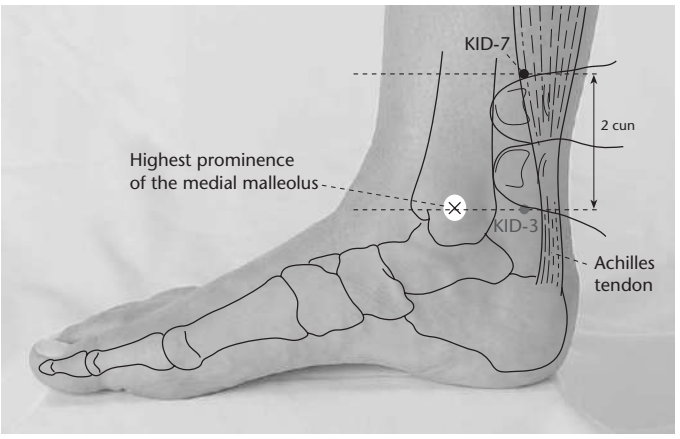

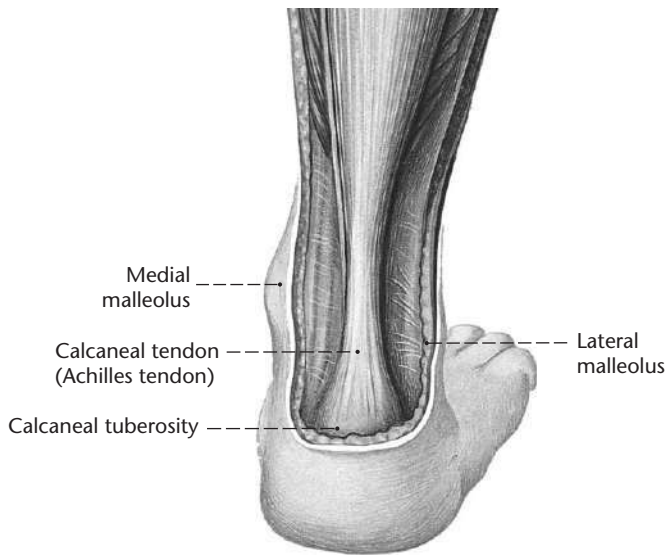

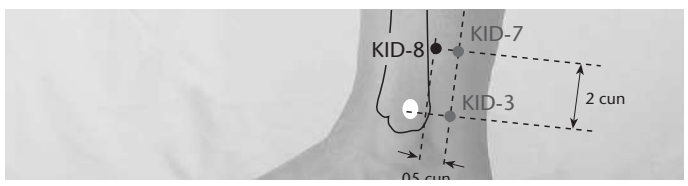

Location

2 cun proximal to → **KID-3**, at the anterior border of the Achilles tendon.

How to find

First, locate → **KID-3** on the level of the higest prominence of the medial malleolus (→ 3.6.2), in the depression between the malleolus and the Achilles tendon. From → **KID-3**, measure 2 cun in a proximal direction (towards the knee joint) and locate **KID-7** in a depression at the anterior border of the Achilles tendon.
Located on the same level, but more medially, is → **KID-8** (2 cun directly proximal to the medial malleolus).

Needling

Vertically 0.5–1 cun

Actions/Indications

- Regulates the water passages and treats oedema, strengthens the Kidneys (especially Kidney Yang), drains Dampness and clears Damp-Heat
- For disorders of the Intestines due to Damp-Heat
- Strengthens the lumbar region

Special features

Jing-river point, Metal point, tonification point

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 328

KID-8 *Exchange Belief* JIAOXIN

Location
2 cun proximal to the highest prominence of the medial malleolus, posterior to the medial border of the tibia.

How to find
KID-8 is located on the same level as → **KID-7**, but slightly closer to the border of the tibia (approximately 0.5 cun).

Needling
Vertically 0.5–1 cun

Actions/Indications
- Regulates menstruation, regulates the *ren mai* and *chong mai*
- Clears Heat and eliminates Dampness from the Lower Burner
- Treats blockages of the *yin qiao mai*

Special features
Xi-cleft point of the *yin qiao mai*

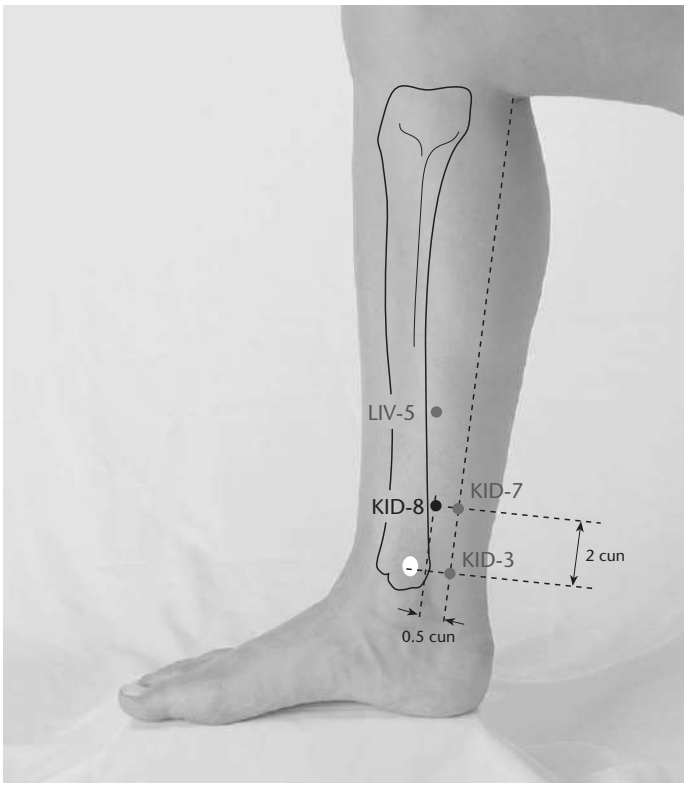

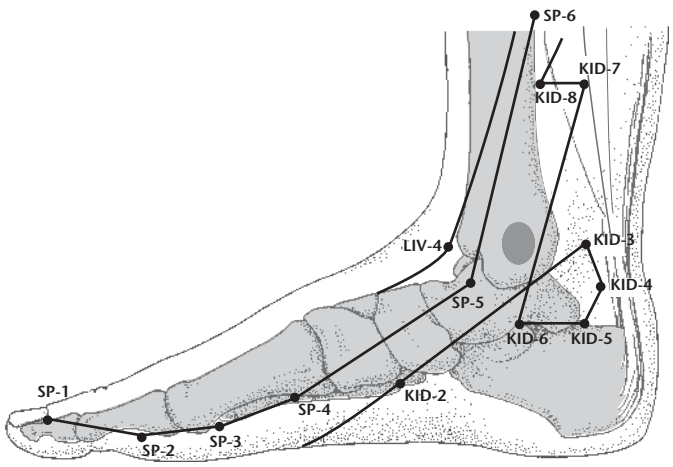

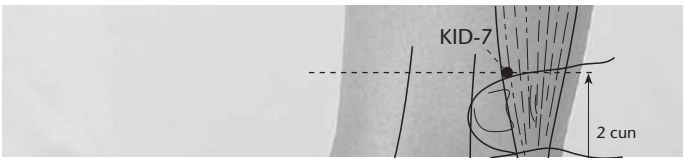

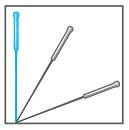
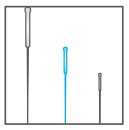

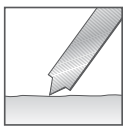

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 329

Guesthouse ZHUBIN KID-9

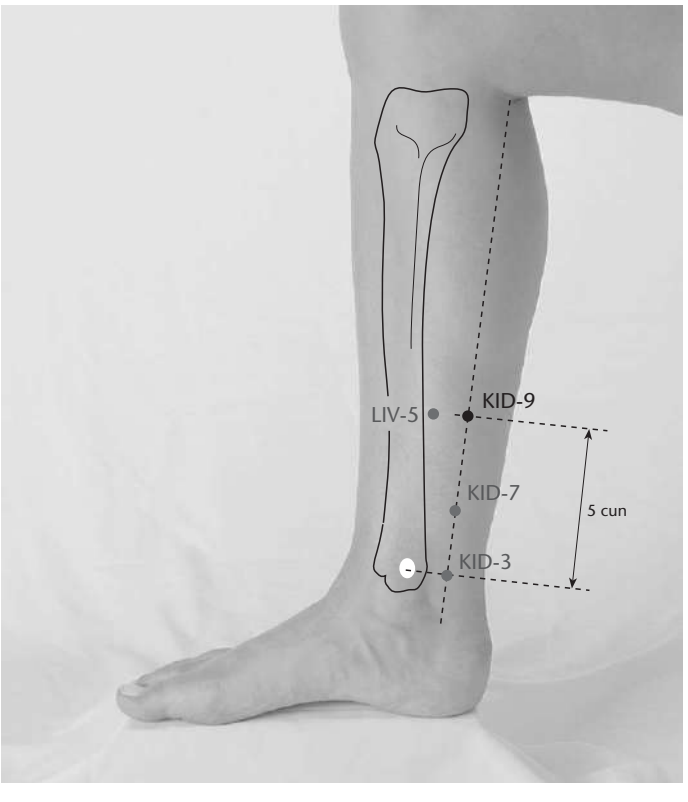

Location

5 cun proximal to the highest prominence of the medial malleolus, 2 cun posterior to the medial border of the tibia.

How to find

On a line connecting → **KID-3** (posterior to the medial malleolus) and → **KID-10** (medial aspect of the popliteal crease), measure 5 cun from the level of the prominence of the medial malleolus. Here, **KID-9** is located 2 cun posterior to the medial border of the tibia.

Located on the same level, but directly posterior to the medial border of the tibia, is → **LIV-5**.

Needling

Vertically 1–1.5 cun

Actions/Indications

- Clears the Heart and transforms Phlegm
- Regulates the Qi and alleviates pain

Special features

Xi-cleft point of the *yin wei mai*

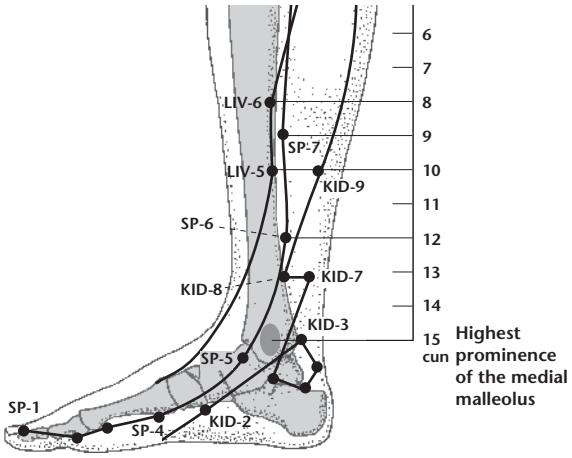

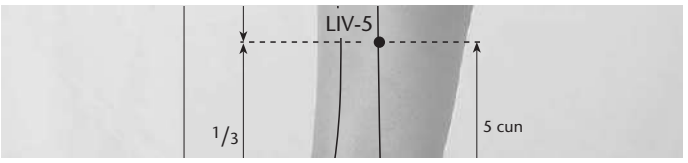

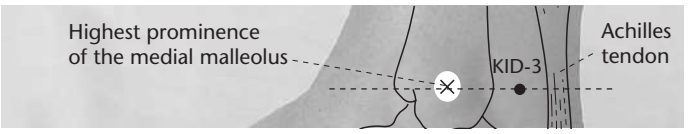

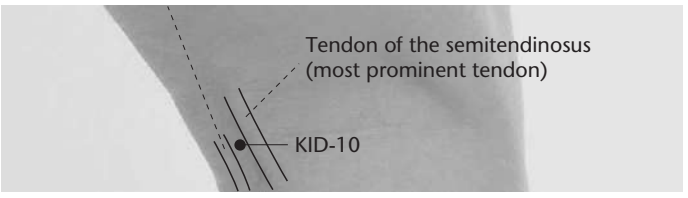

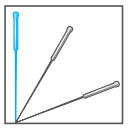
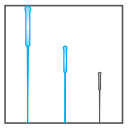

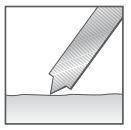

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 330

KID-10 *Yin Valley* YINGU

Location

At the medial end of the popliteal crease, between the tendons of the semimembranosus and semitendinosus muscles, on the level of the knee joint space.

How to find

With the knee flexed less than 90°, ask the patient to press their heel against the treatment couch so that the two tendons will become more pronounced in the popliteal crease. From a medial direction, **KID-10** is located in a small gap anterior to the more prominent tendon of the semitendinosus muscle and posterior to the less well-defined tendon of the semimembranosus muscle.
→ **BL-40** is located on the same level, in the centre of the popliteal crease.

Needling

Needle 1–1.5 cun vertically from a dorsomedial aspect towards the tibial tuberosity. Avoid puncturing the tendon.

Actions/Indications

- Drains Damp-Heat from the Lower Burner and benefits the Kidneys
- Opens the channel

Special features

He-sea point, *ben* point (Five Phases point); for recurring urinary disorders with Kidney Yin deficiency.

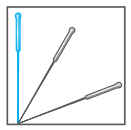

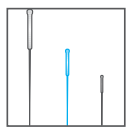

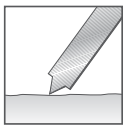

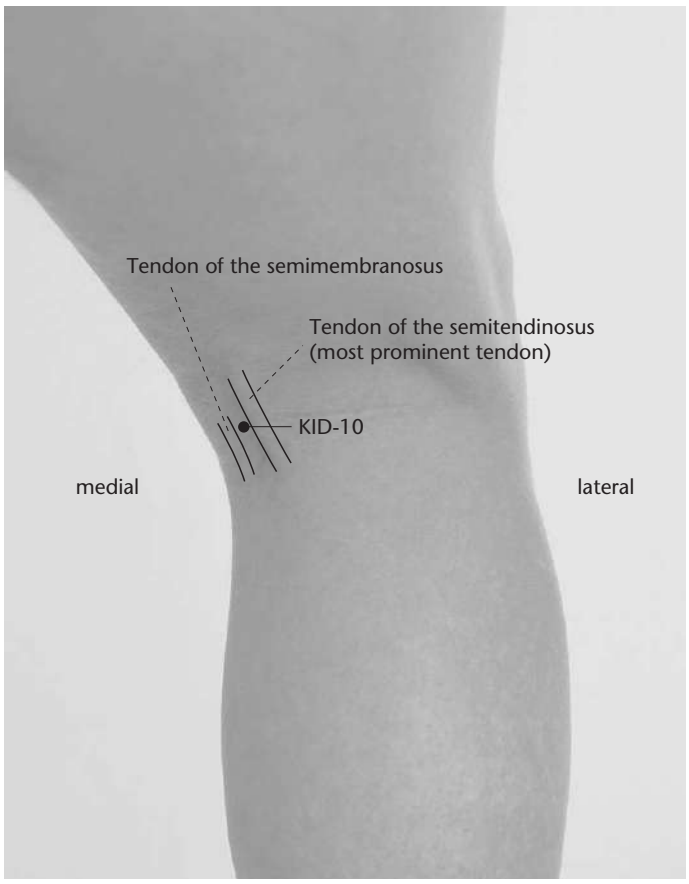

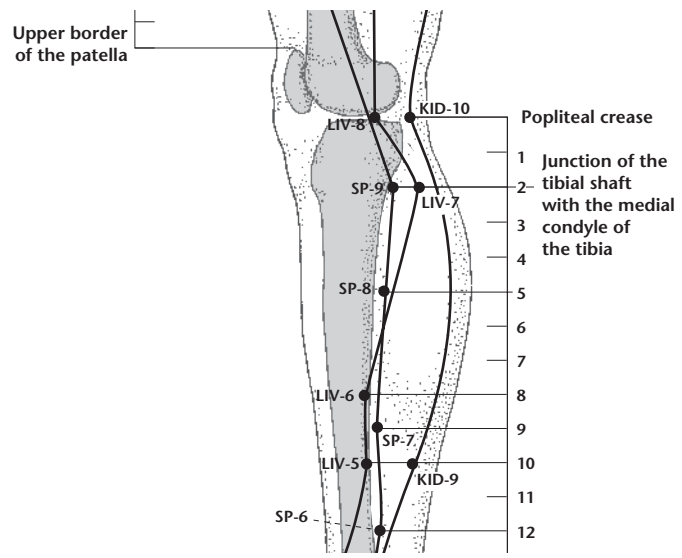

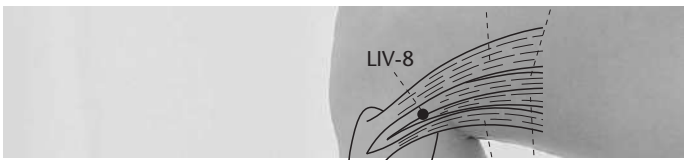

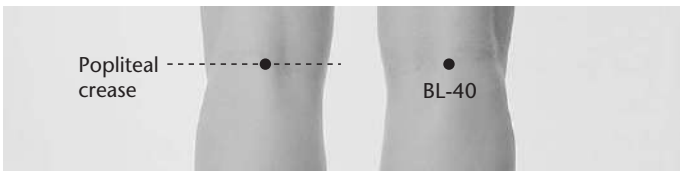

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 331

Pubic Bone HENGGU KID-11

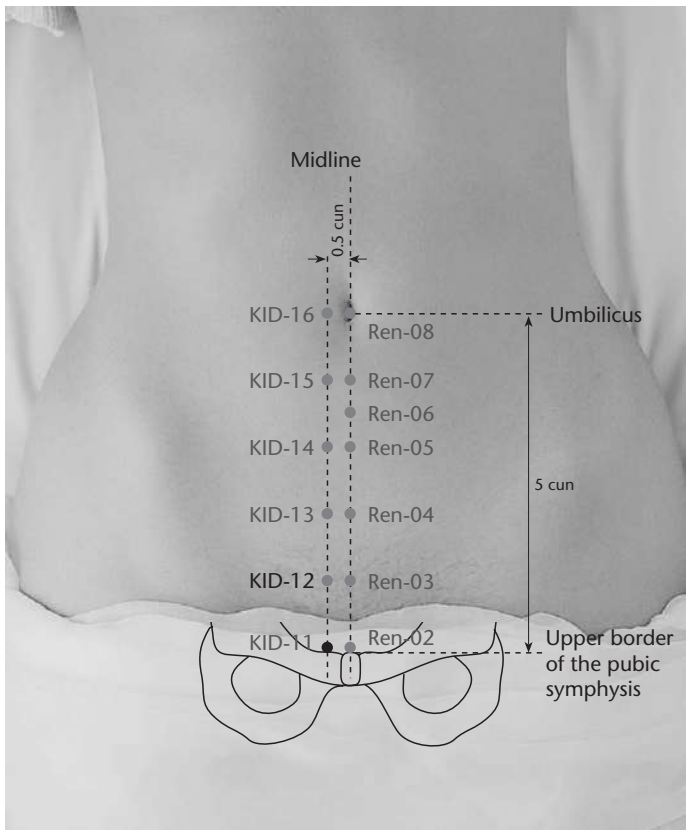

Location

At the upper border of the pubic symphysis, 0.5 cun lateral to the anterior midline.

How to find

Close to the midline, the upper border of the pubic symphysis is generally easily palpable. **KID-11** is located directly superior to its bony margin and 0.5 cun lateral to the midline.
Located on the same level are → **Ren-2** (on the midline), → **ST-30** (2 cun lateral to the midline) and → **SP-12** (3.5 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, full bladder, pregnancy.

Actions/Indications

- Regulates the Lower Burner and the water passages, strengthens the Kidneys

Special features

Meeting point with the *chong mai*

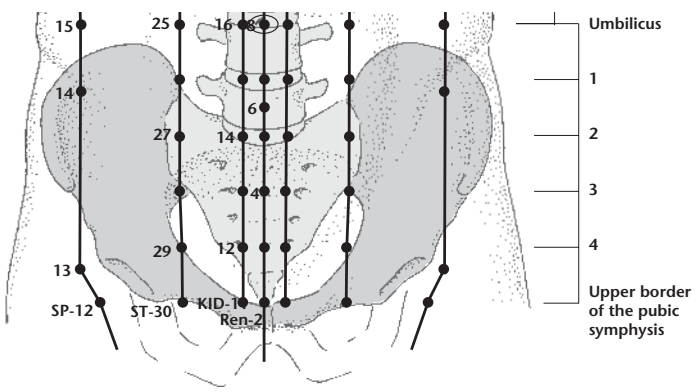

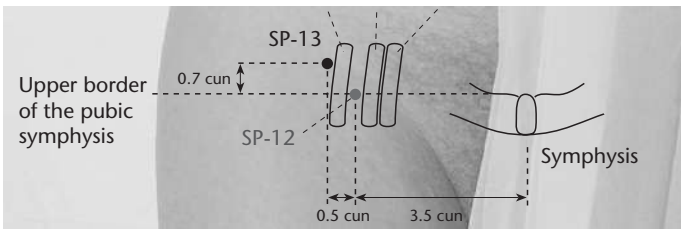

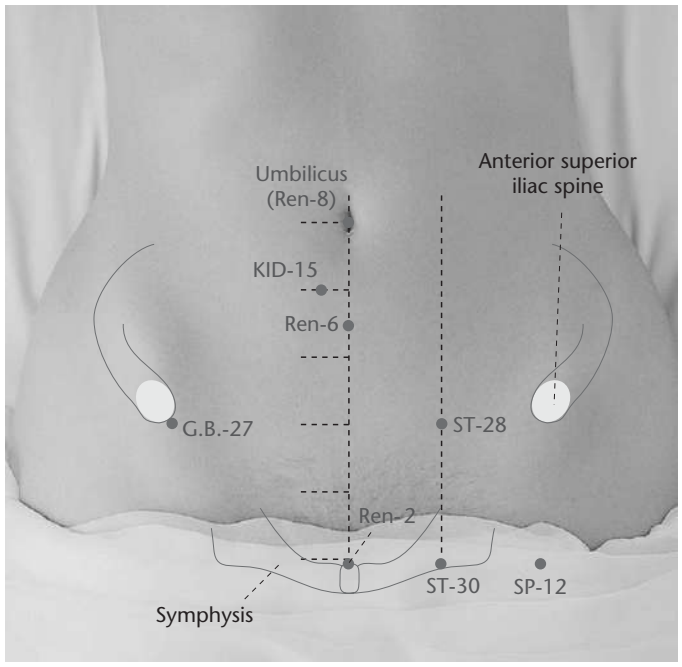

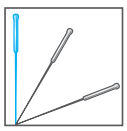
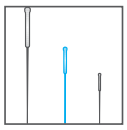
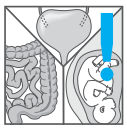

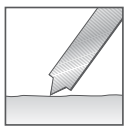
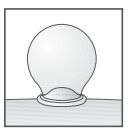

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 332

KID-12 *Great Luminance* DAHE

Location

1 cun superior to the upper border of the pubic symphysis, 0.5 cun lateral to the anterior midline.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun. As these can vary considerably from finger cun measurements, only proportional cun measurements should be applied (elastic tape → 2.3.1). From the upper border of the pubic symphysis, measure 1 cun in a superior direction and on this level locate **KID-12** 0.5 cun lateral to the midline. Located on the same level are → **Ren-3** (on the midline), → **ST-29** (2 cun lateral to the midline) and → **Ex-CA-1** (*zigong*; 3 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, full bladder, pregnancy.

Actions/Indications

- Regulates the Lower Burner, strengthens the Kidneys and the Essence (*jing*)

Special features

Meeting point with the *chong mai*

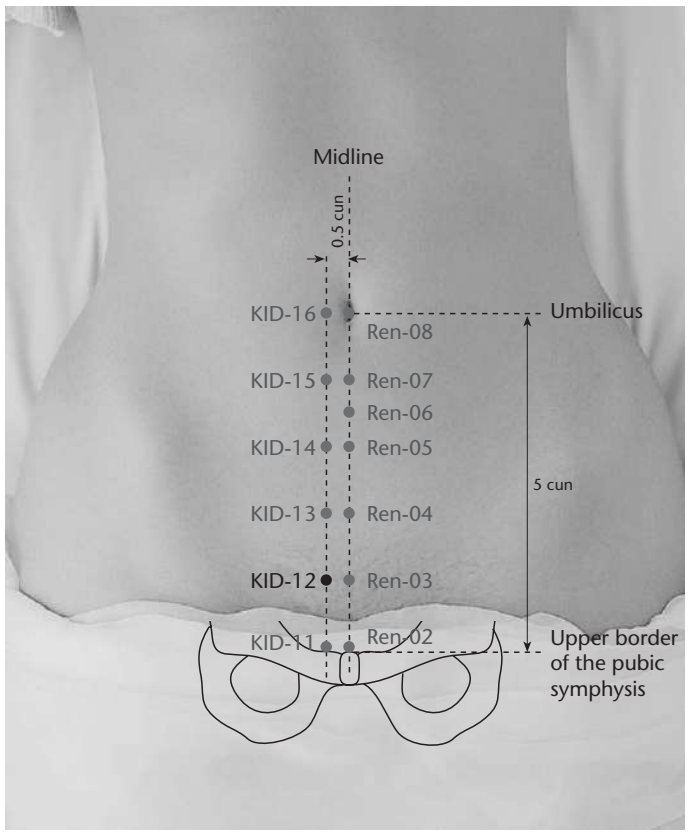

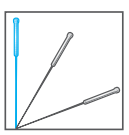

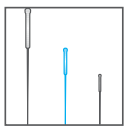

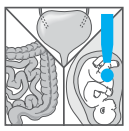

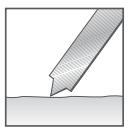

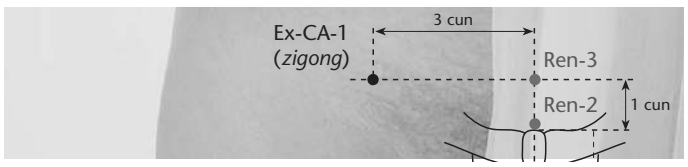

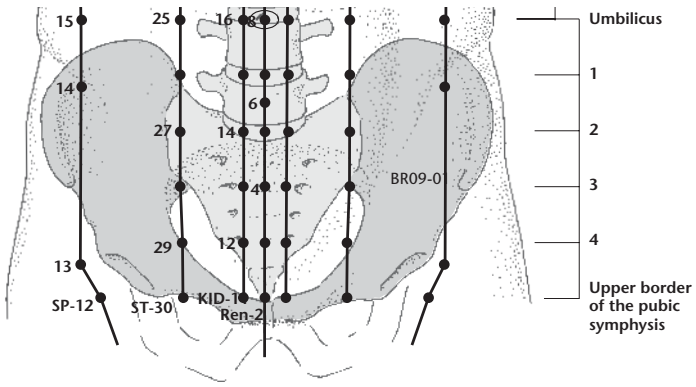

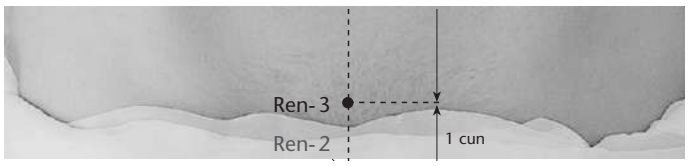

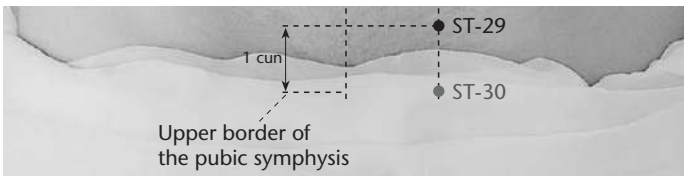

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 333

Qi Cave QIXUE KID-13

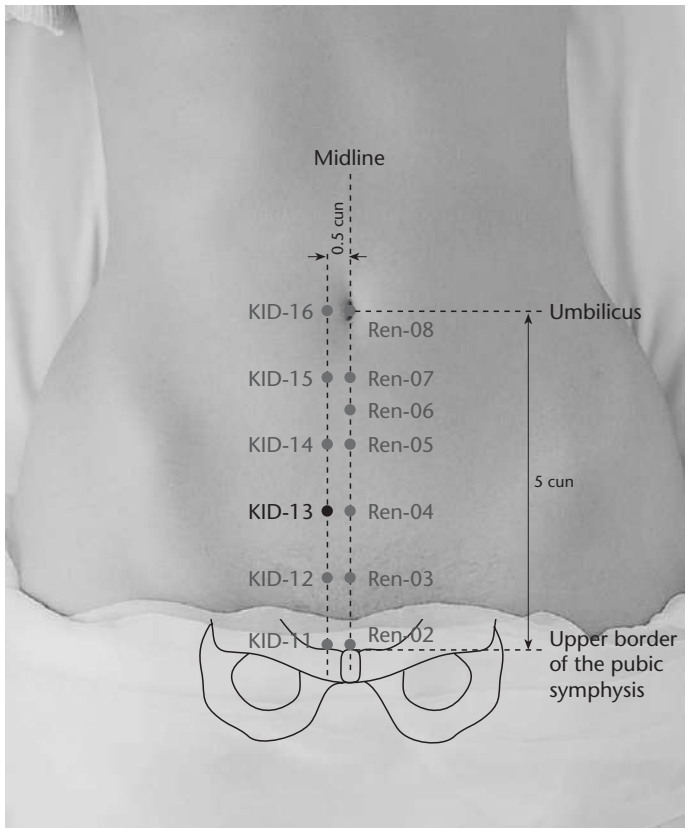

Location

2 cun superior to the upper border of the pubic symphsis, 0.5 cun lateral to the anterior midline.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun. As these can vary considerably from finger cun measurements, only proportional cun measurements should be applied (elastic tape ➞ 2.3.1). From the upper border of the pubic symphysis, measure 2 cun in a superior direction and on this level locate **KID-13** 0.5 cun lateral to the midline.

Located on the same level are ➞ **Ren-4** (on the midline), ➞ **ST-28** (2 cun lateral to the midline), ➞ **Ex-CA** (*tituo/qimen/yijing*: 4/3/1 cun lateral to the midline) and ➞ **G.B.-27** (anterior and medial to the ASIS).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, full bladder, pregnancy.

Actions/Indications

- Regulates the Lower Burner, regulates the *chong mai* and *ren mai*, strengthens the Kidneys and the essence (*jing*)

Special features

Meeting point with the *chong mai*

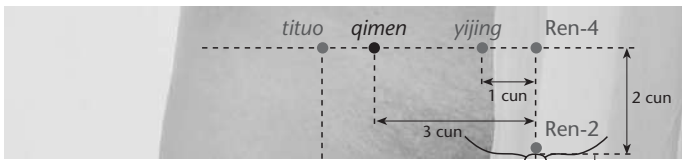

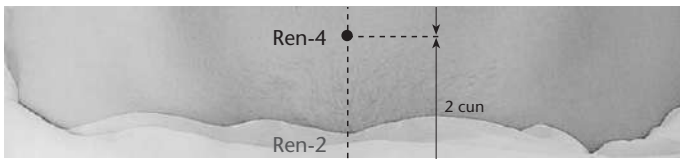

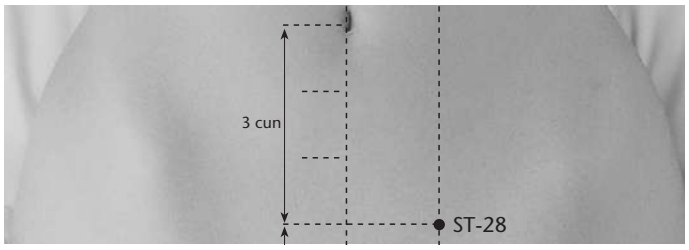

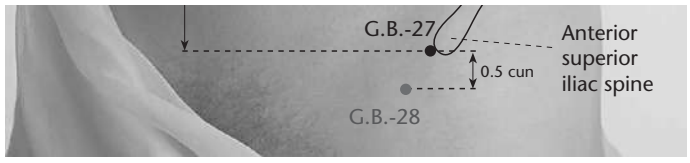

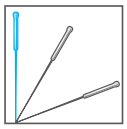
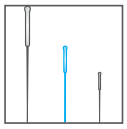
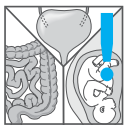

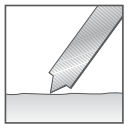

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 334

KID-14 *Four Fullnesses* SIMAN

Location

2 cun inferior to the umbilicus, 0.5 cun lateral to the anterior midline.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun. As these can vary considerably from finger cun measurements, only proportional cun measurements should be applied (elastic tape → 2.3.1). From the centre of the umbilicus, measure 2 cun in an inferior direction and on this level locate **KID-14** 0.5 cun lateral to the midline.

Located on the same level are → **Ren-5** (on the midline) and → **ST-27** (2 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Regulates the Lower Burner and benefits the Kidneys
- Regulates the water passages
- Regulates the Qi and moves Blood stasis

Special features

Meeting point with the *chong mai*

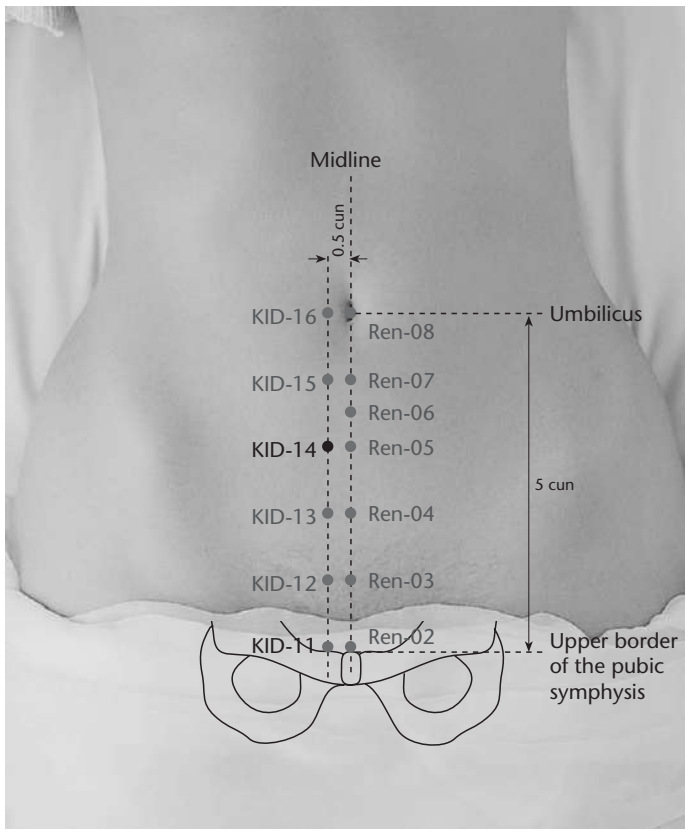

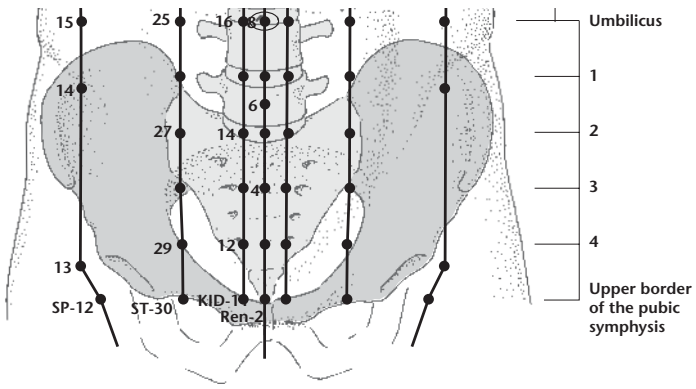

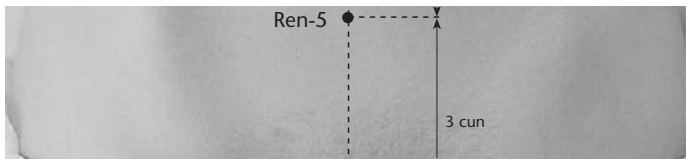

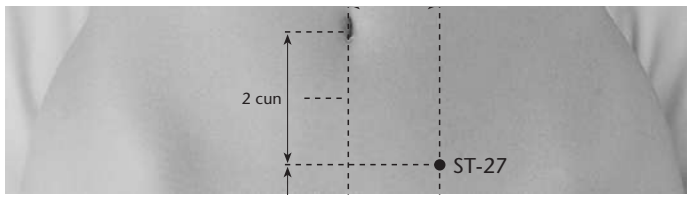

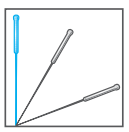

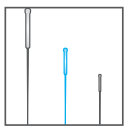

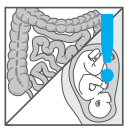

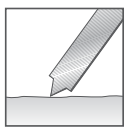

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 335

Middle Flow ZHONGZHU KID-15

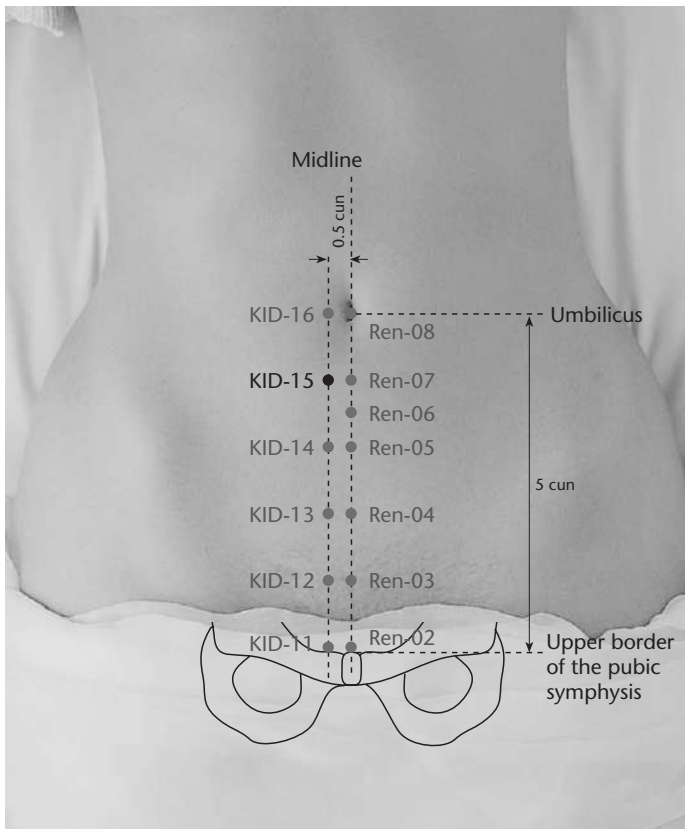

Location

1 cun inferior to the umbilicus, 0.5 cun lateral to the anterior midline.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun. As these can vary considerably from finger cun measurements, only proportional cun measurements should be applied (elastic tape → 2.3.1). From the centre of the umbilicus, measure 1 cun in an inferior direction and on this level locate **KID-15** 0.5 cun lateral to the midline.

Located on the same level are → **Ren-7** (on the midline) and → **ST-26** (2 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Regulates the Intestines
- Regulates the Lower Burner

Special features

Meeting point with the *chong mai*

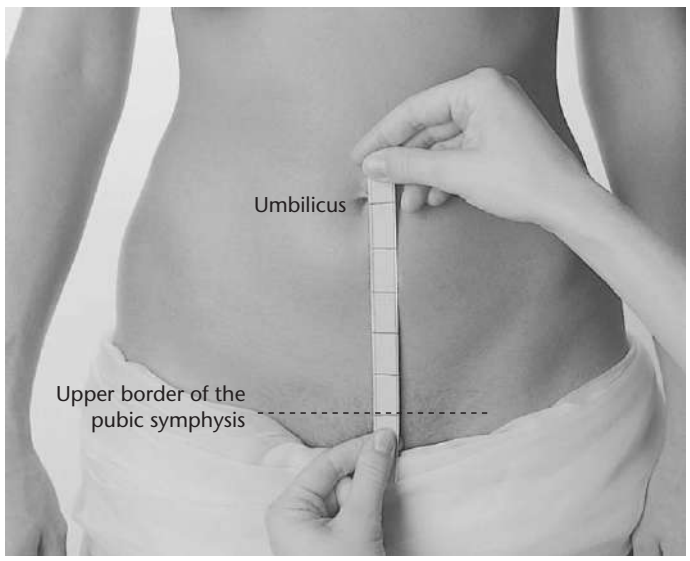

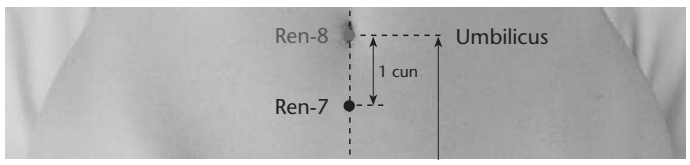

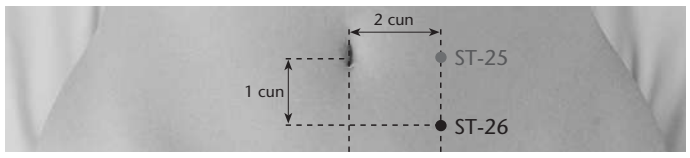

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 336

KID-16 *Vitals* Shu HUANGSHU

Location
0.5 cun lateral to the centre of the umbilicus.

How to find
From the centre of the umbilicus, measure 0.5 cun in a lateral direction. Located on the same level are → **Ren-8** (in the centre of the umbilicus), → **ST-25** (2 cun lateral to the midline), → **SP-15** (4 cun lateral to the midline) and → **G.B.-26** (on the level of umbilicus, on a vertical line drawn through the free end of the 11th rib).

Needling
Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy. Avoid the umbilicus.

Actions/Indications
- Regulates the Qi, regulates and warms the Stomach and Intestines

Special features
Meeting point with the *chong mai*

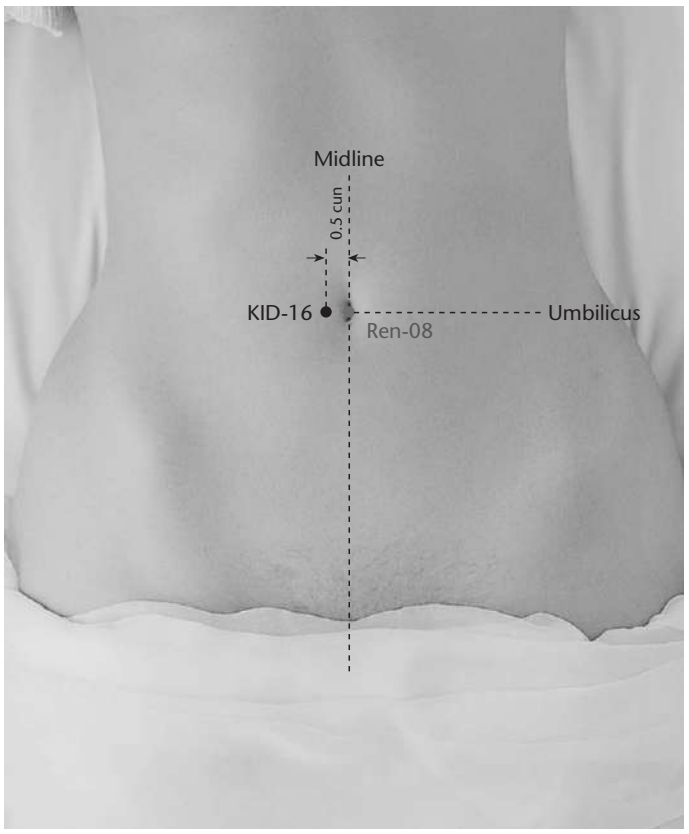

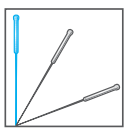

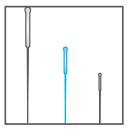

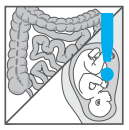

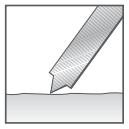

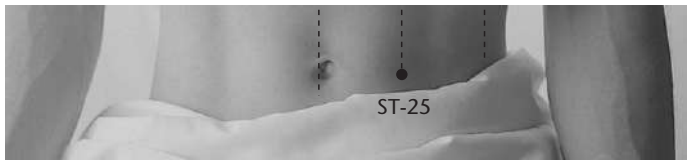

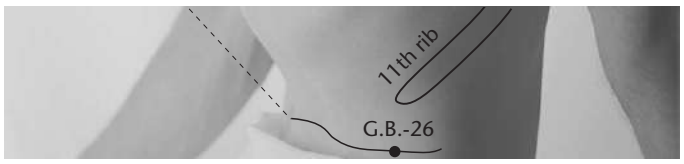

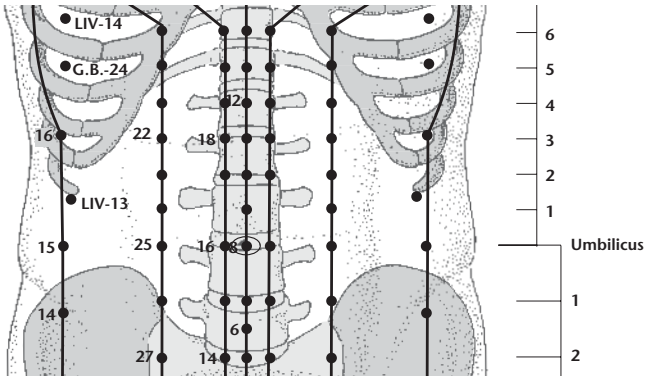

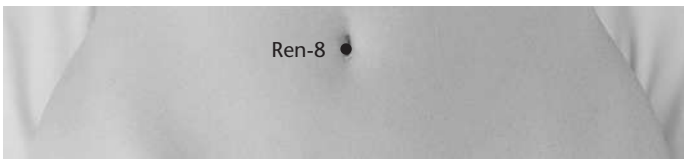

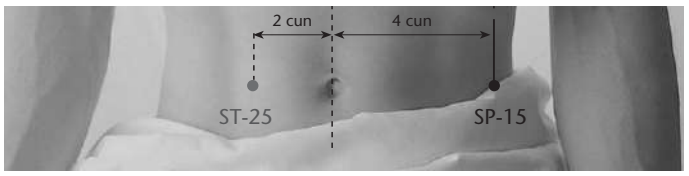

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 337

Shang Bend SHANGQU KID-17

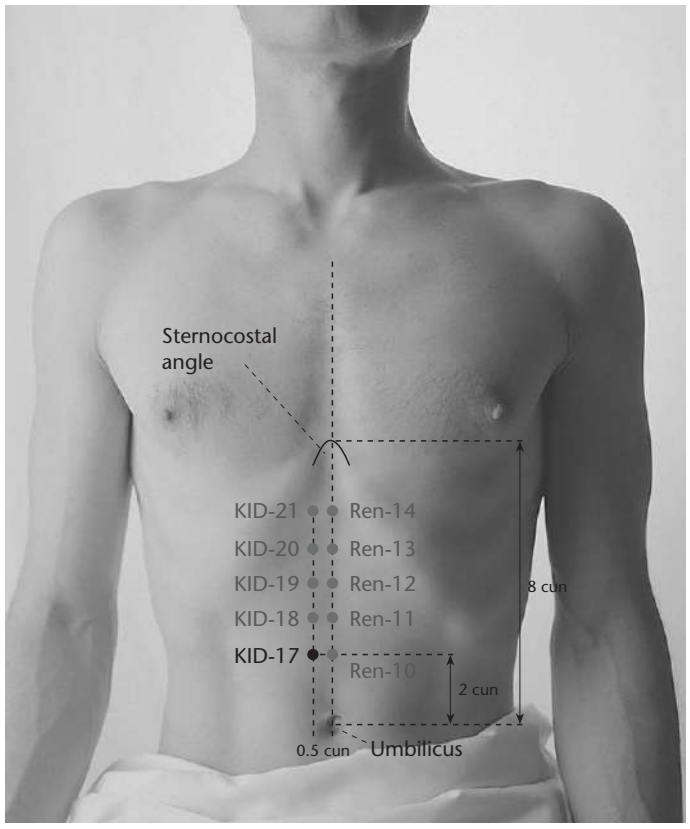

Location

2 cun superior to the umbilicus, 0.5 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun should be used (elastic tape → 2.3.1). From the centre of the umbilicus, measure 2 cun in a superior direction and on this level locate **KID-17** 0.5 cun lateral to the midline.
Located on the same level are → **Ren-10** (on the midline), → **ST-23** (2 cun lateral to the midline) and **Ex-CA** (*weishang*, 4 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Regulates the Stomach and Intestines, eliminates stagnation

Special features

Meeting point with the *chong mai*

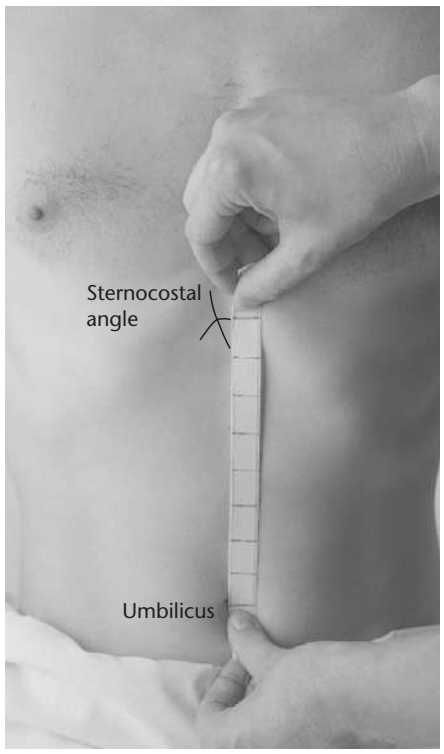

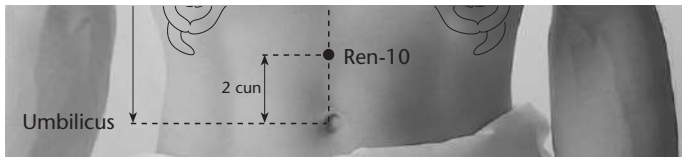

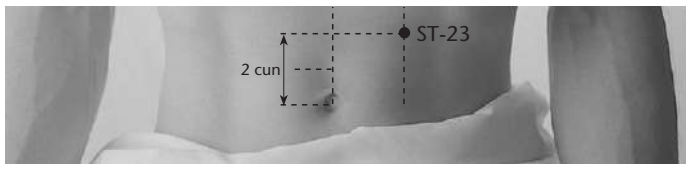

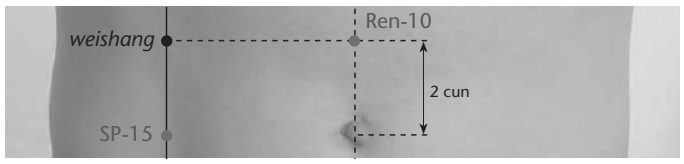

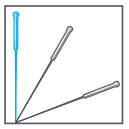
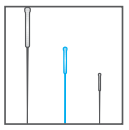
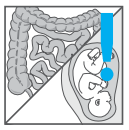

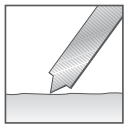

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 338

KID-18 *Stone Gate* SHIGUAN

Location

3 cun superior to the umbilicus, 0.5 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used (elastic tape → 2.3.1). From the centre of the umbilicus measure 3 cun in a superior direction and on this level locate **KID-18** 0.5 cun lateral to the midline.

Located on the same level are → **Ren-11** (on the midline), → **ST-22** (2 cun lateral to the midline) and → **SP-16** (on the mamillary line).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Strengthens the Middle, regulates the Stomach and Intestines, removes Blood stagnation

Special features

Meeting point with the *chong mai*

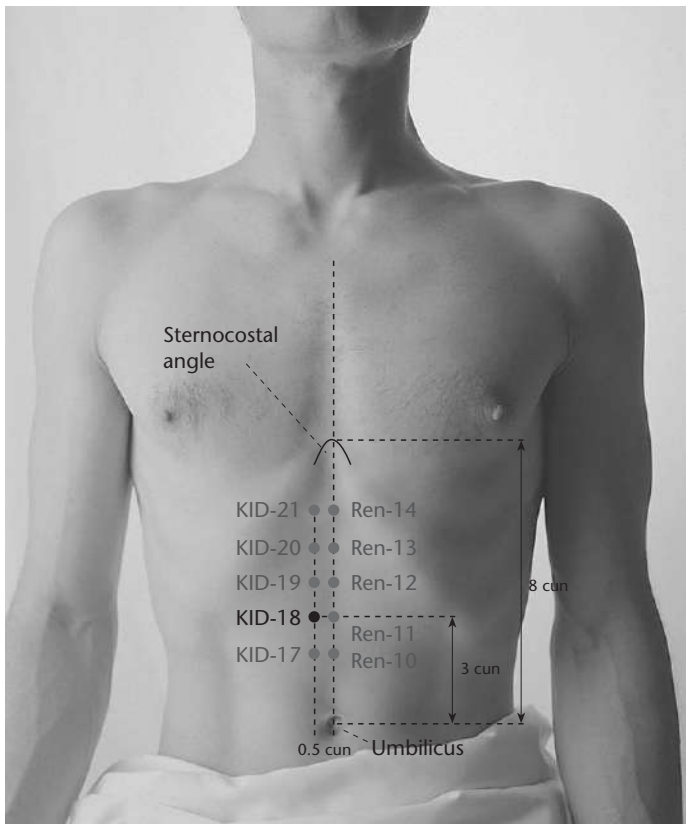

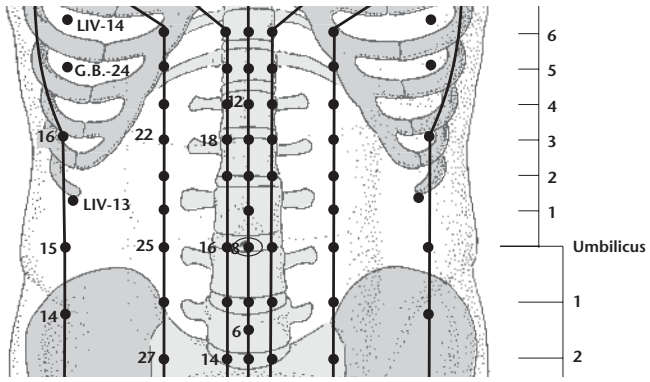

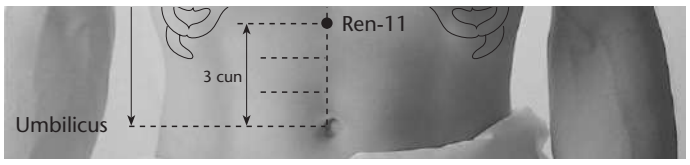

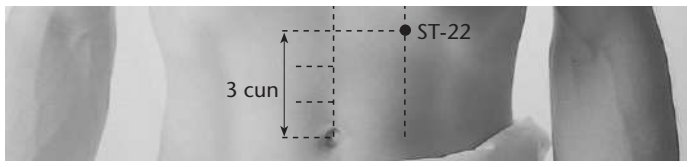

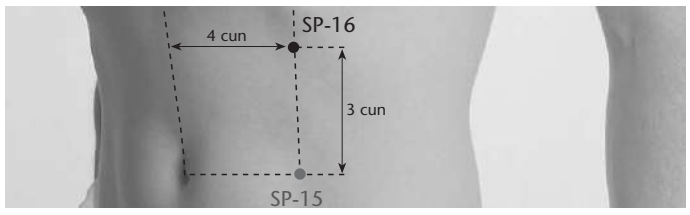

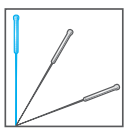

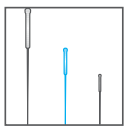

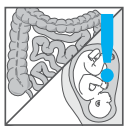

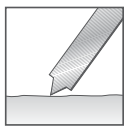

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 339

Yin Metropolis YINDU KID-19

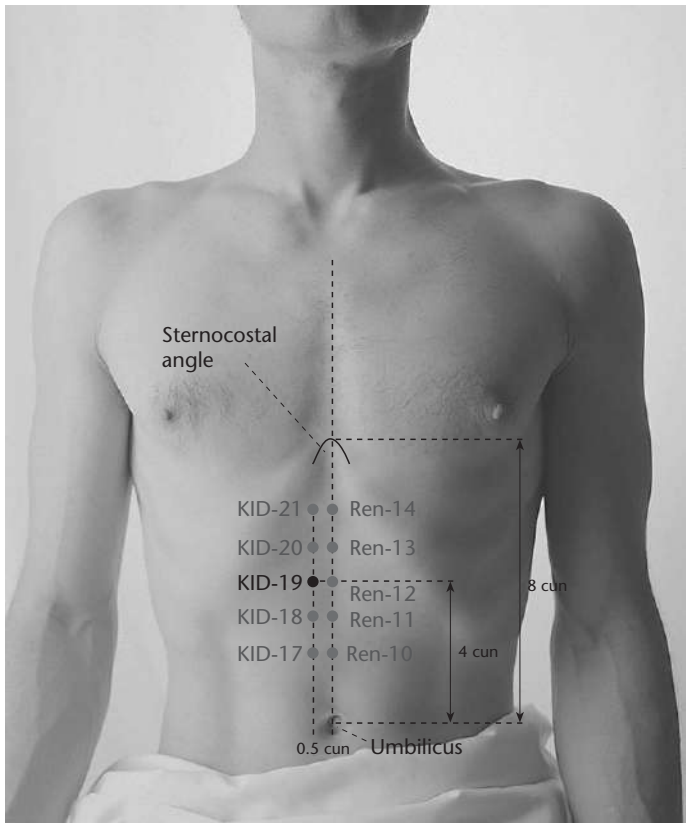

Location

Midway between the sternocostal angle and the centre of the umbilicus, 0.5 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun should be used (elastic tape → 2.3.1). Spreading hands technique (→ 2.3.3): Place the little fingers on the centre of the umbilicus and the sternocostal angle respectively and join the thumbs at the midpoint of this distance by evenly spreading both hands. **KID-19** is located on this level, 0.5 cun lateral to the midline.

Located on the same level are → **Ren-12** (on the midline) and → **ST-21** (2 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Regulates the Qi, regulates the Stomach and Intestines
- Regulates counterflow Qi

Special features

Meeting point with the *chong mai*

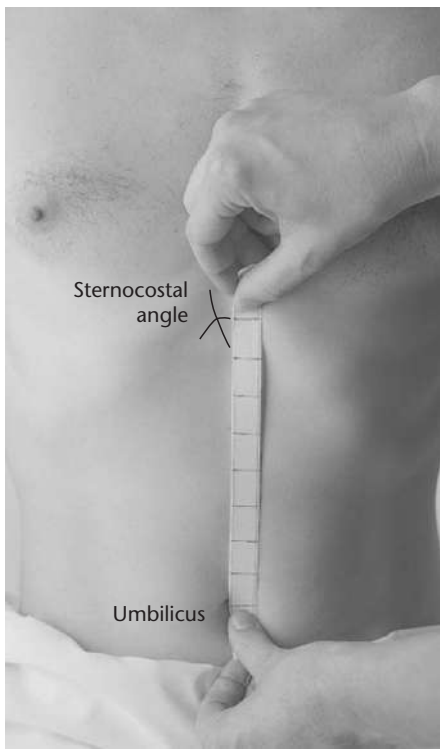

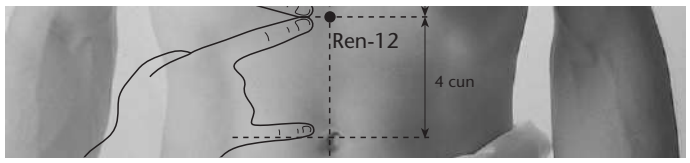

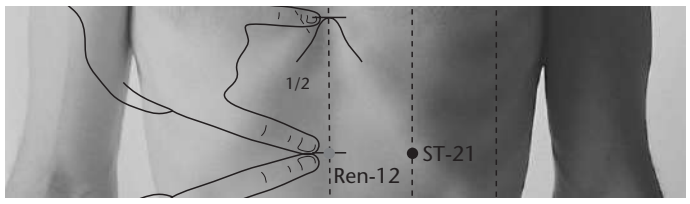

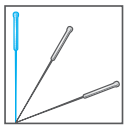
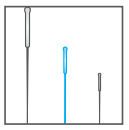
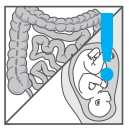

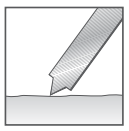

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 340

KID-20 *Abdomen Connecting Valley* FUTONGGU

Location

5 cun superior to the umbilicus or 3 cun inferior to the sternocostal angle, 0.5 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used (elastic tape → 2.3.1). From the sternocostal angle, measure 3 cun in an inferior direction and on this level locate **KID-20** 0.5 cun lateral to the midline.

Located on the same level are → **Ren-13** (on the midline), → **ST-20** (2 cun lateral to the midline) and → **G.B.-24** (on the mamillary line, in the 7th intercostal space).

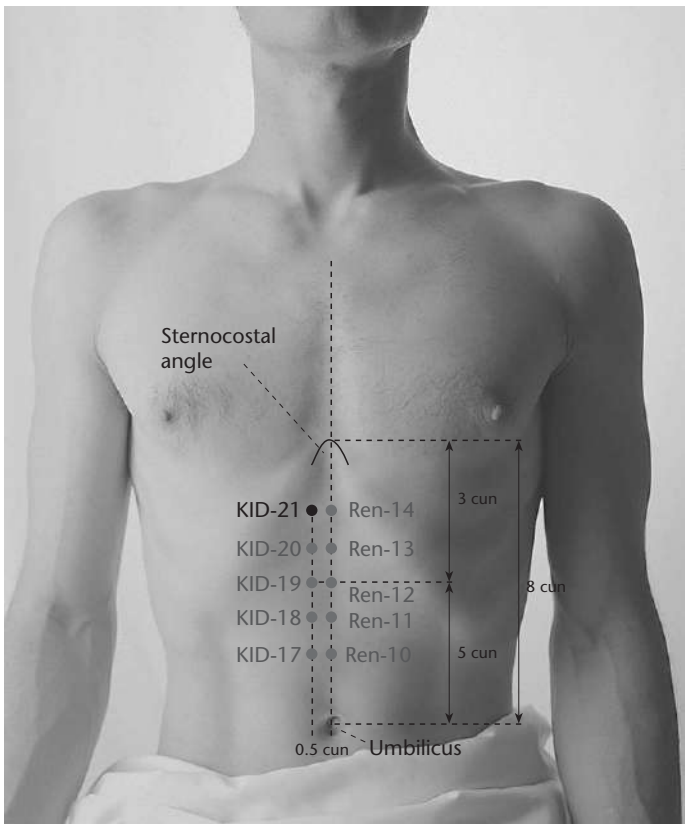

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Strengthens the Spleen, harmonises the Stomach and Intestines, regulates counterflow Qi

Special features

Meeting point with the *chong mai*

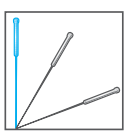

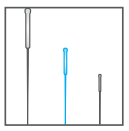

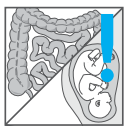

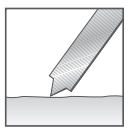

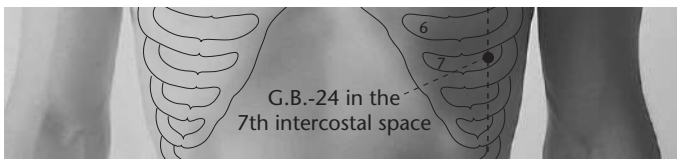

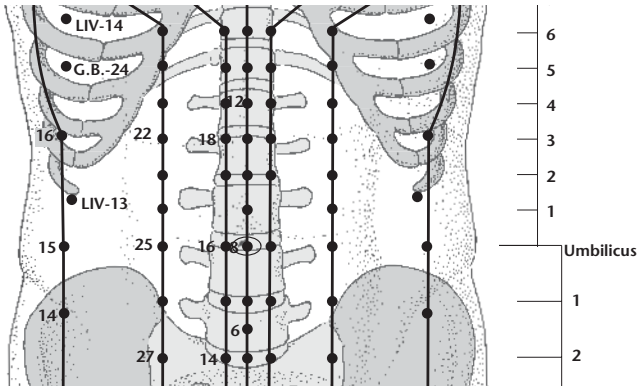

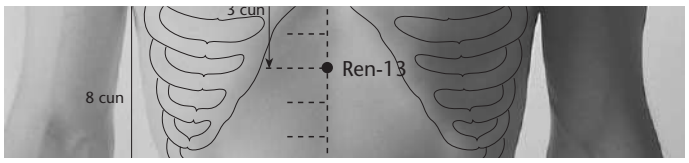

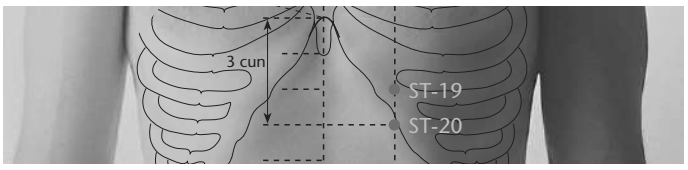

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 341

Hidden Gate YOUMEN KID-21

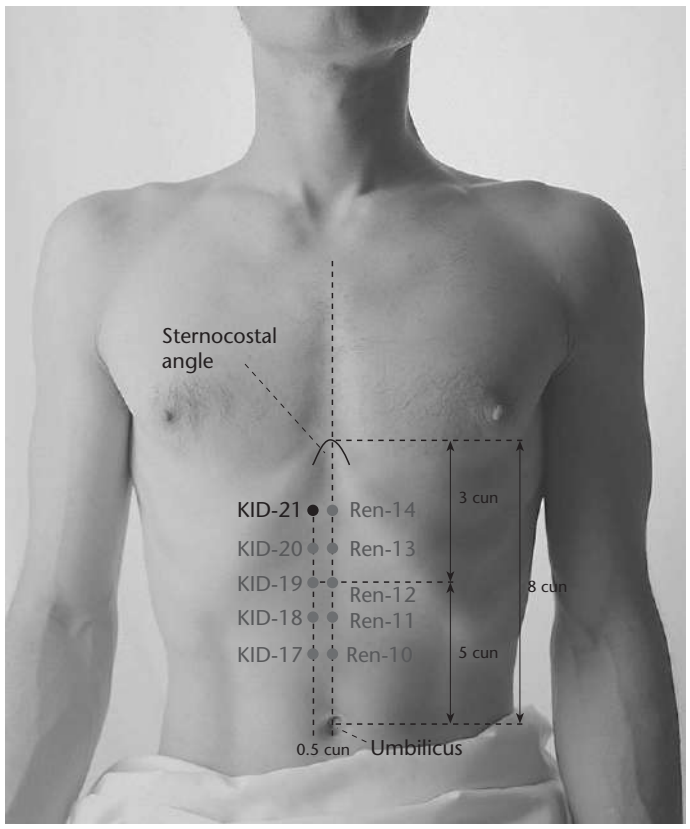

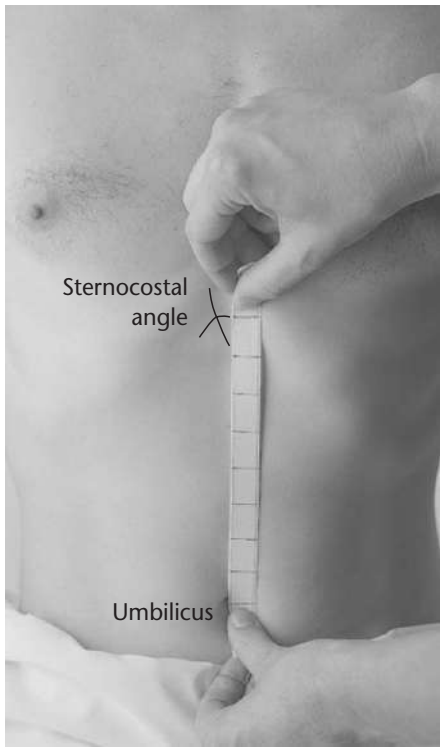

Location

2 cun inferior to the sternocostal angle, 0.5 cun lateral to the anterior midline.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun should be used (elastic tape → 2.3.1). From the sternocostal angle, measure 2 cun in an inferior direction and on this level locate **KID-21** 0.5 cun lateral to the midline.
Located on the same level are → **Ren-14** (on the midline), → **ST-19** (2 cun lateral to the midline) and → **LIV-14** (on the mamillary line, in the 6th intercostal space).

Needling

Vertically 0.5–1 cun. Caution: The liver is located on the right, the peritoneum on the left. Caution during pregnancy.

Actions/Indications

- Harmonises the Stomach, regulates counterflow Qi

Special features

Meeting point with the *chong mai*. Last point of the *chong mai* located on the Kidney channel.

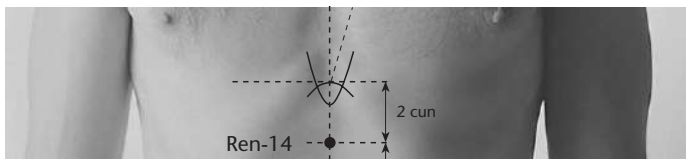

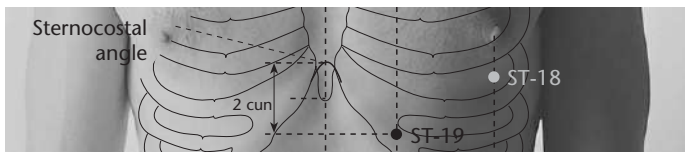

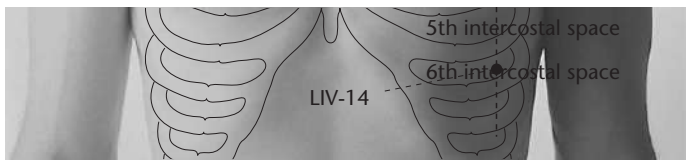

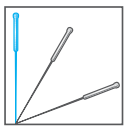
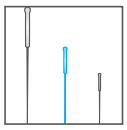
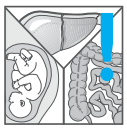

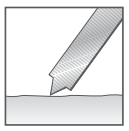

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 342

KID-22 *Walking Corridor* BULANG

Location

In the 5th intercostal space (ICS), 2 cun lateral to the anterior midline.

How to find

Locate the 5th ICS either by counting parasternally from the clavicle or from the manubriosternal synchondrosis (→ 2nd rib) (→ 3.5), then locate **KID-22** 2 cun lateral to the anterior midline. Located on the same level (5th ICS) are → **Ren-16** (on the midline), → **ST-18** (on the mamillary line) and → **SP-17** (6 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously) with or against the channel pathway. Caution: Avoid injuring the liver (on the right side), the heart (on the left side) and the pleura.

Actions/Indications

- Regulates the Lung Qi and counterflow Qi, unbinds the chest

Special features

Exit point. The *Su Wen* describes points **KID-22** to **KID-27** as *shu* points of the chest, indicated for difficult breathing and disorders of the thorax. They are particularly indicated for breathing difficulties due to 'excess above, deficiency below', which refers to the inability of the Kidneys to absorb the Qi that has descended from the Lung.

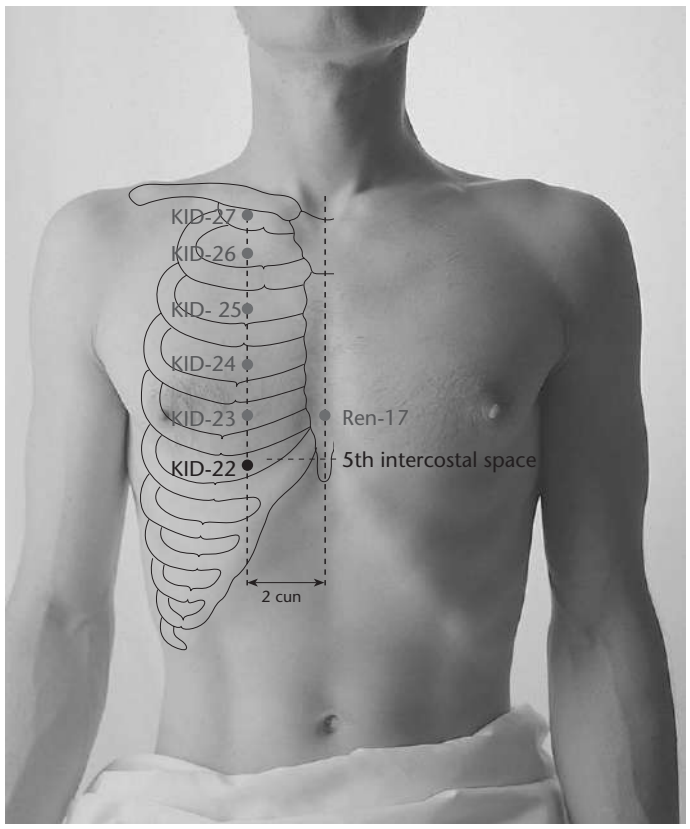

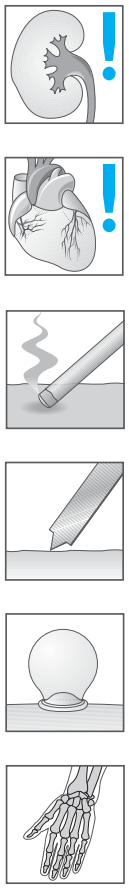

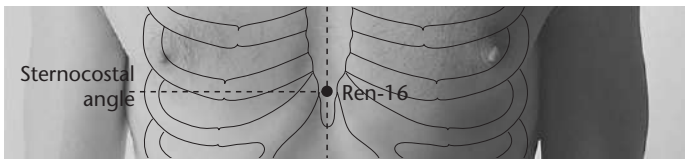

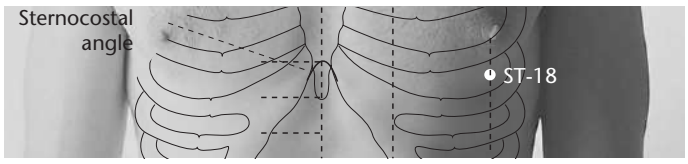

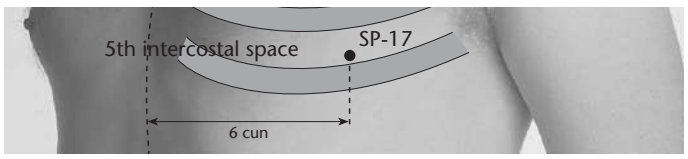

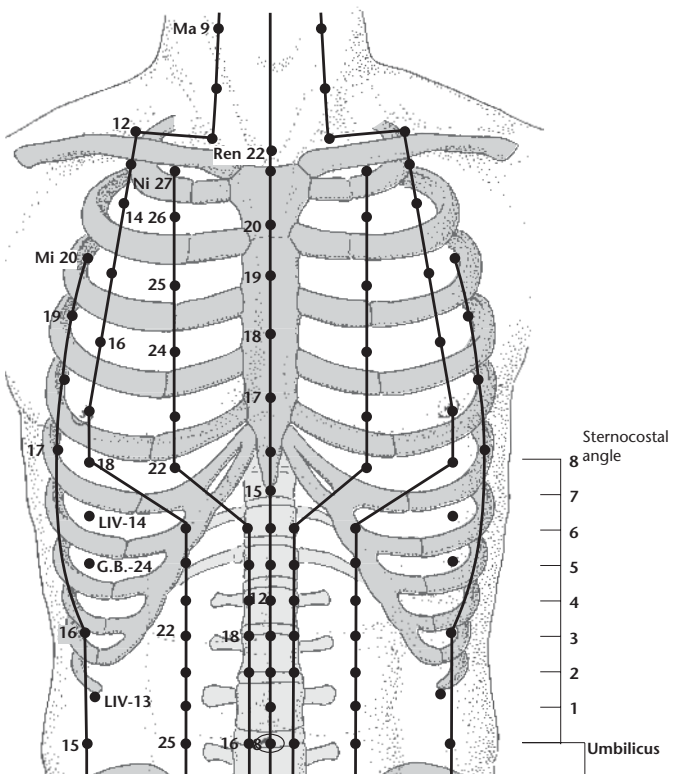

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 343

Spirit Seal SHENFENG KID-23

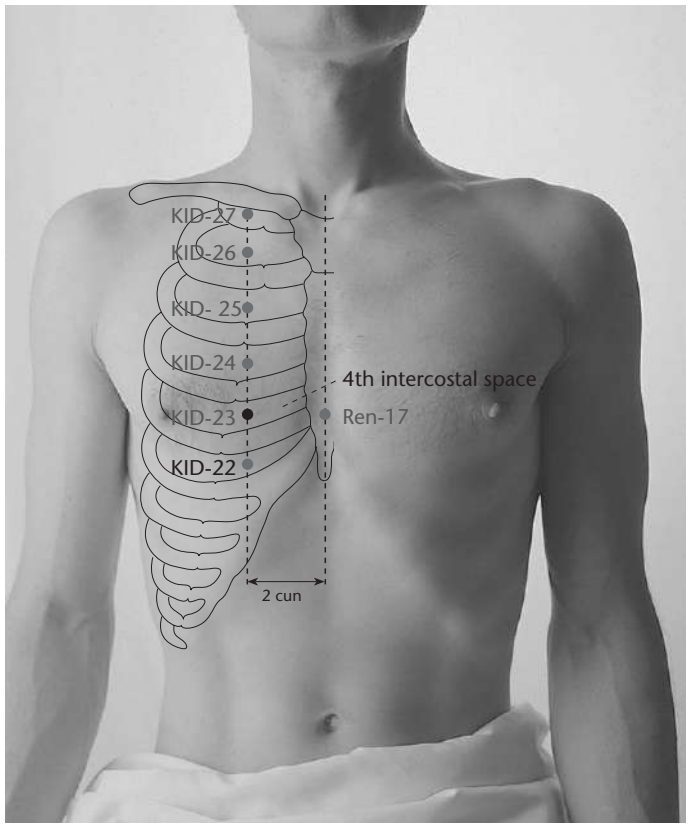

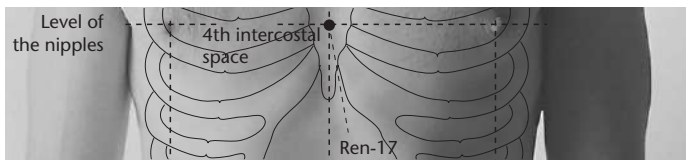

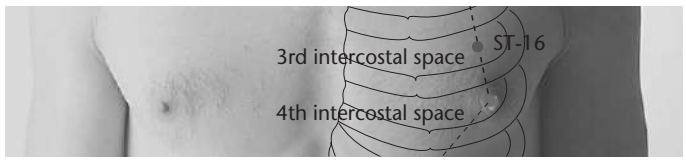

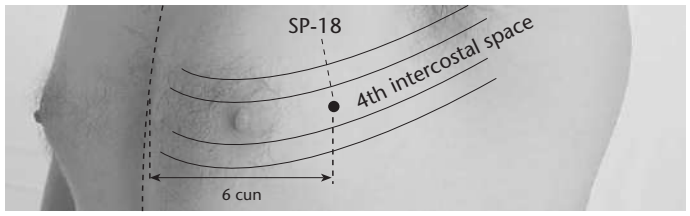

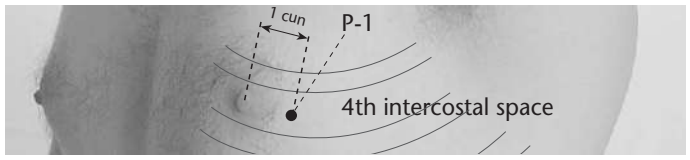

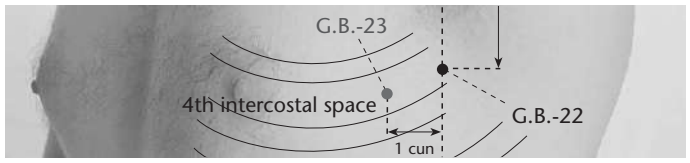

Location

In the 4th intercostal space (ICS), 2 cun lateral to the anterior midline.

How to find

Locate the 4th ICS either by palpating parasternally from the clavicle or by counting from the manubriosternal synchondrosis (→ 2nd rib) (→ 3.5), then locate **KID-23** 2 cun lateral to the anterior midline.

Located on the same level (4th ICS) are → **Ren-17** (on the midline), → **ST-17** (on the nipple), → **SP-18** (6 cun lateral to the midline), → **P-1** (1 cun lateral to the nipple) and → **G.B.-22** (3 cun inferior to the axilla) as well as → **G.B.-23** (1 cun anterior to **G.B.-22**).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously) along the course of the ICS or with or against the channel pathway. Caution: Heart, pneumothorax.

Actions/Indications

- Regulates the Lung Qi and counterflow Qi, unbinds the chest
- Harmonises the Stomach
- Benefits the breasts

Special features

The *Su Wen* describes points **KID-22** to **KID-27** as *shu* points of the chest, indicated for difficult breathing and disorders of the thorax. They are particularly indicated for breathing difficulties due to 'excess above, deficiency below', which refers to the inability of the Kidneys to absorb the Qi that has descended from the Lung.

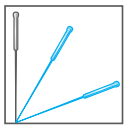
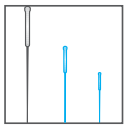
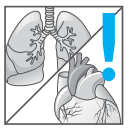

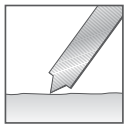

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 344

KID-24 *Spirit Ruin* LINGXU

Location

In the 3rd intercostal space (ICS), 2 cun lateral to the anterior midline.

How to find

Locate the 3rd ICS by palpating parasternally from the clavicle or by counting from the manubriosternal synchondrosis (→ 2nd rib) (→ 3.5), then locate **KID-24** 2 cun lateral to the anterior midline. Located on the same level (→ 3rd ICS) are → **Ren-18** (on the midline), → **ST-16** (4 cun lateral to the midline) and → **SP-19** (6 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously) along the course of the ICS or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Regulates the Lung and Stomach Qi
- Unbinds the chest
- Benefits the breasts

Special features

The *Su Wen* describes points **KID-22** to **KID-27** as *shu* points of the chest, indicated for difficult breathing and disorders of the thorax. They are particularly indicated for breathing difficulties due to 'excess above, deficiency below', which refers to the inability of the Kidneys to absorb the Qi that has descended from the Lung.

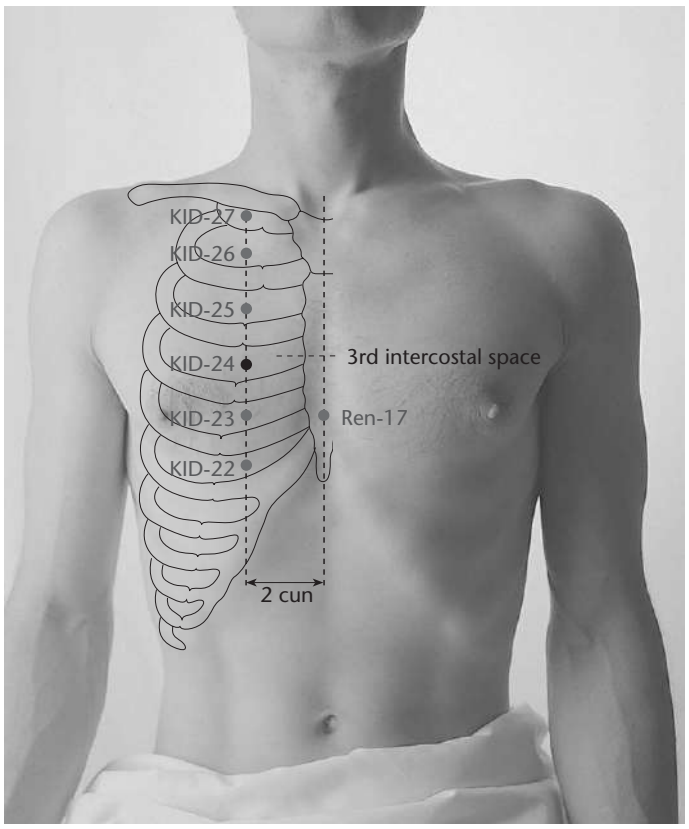

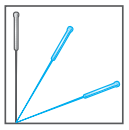
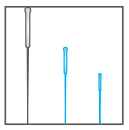
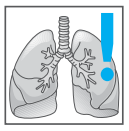

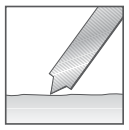

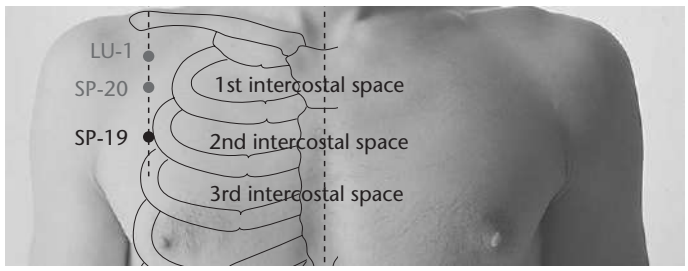

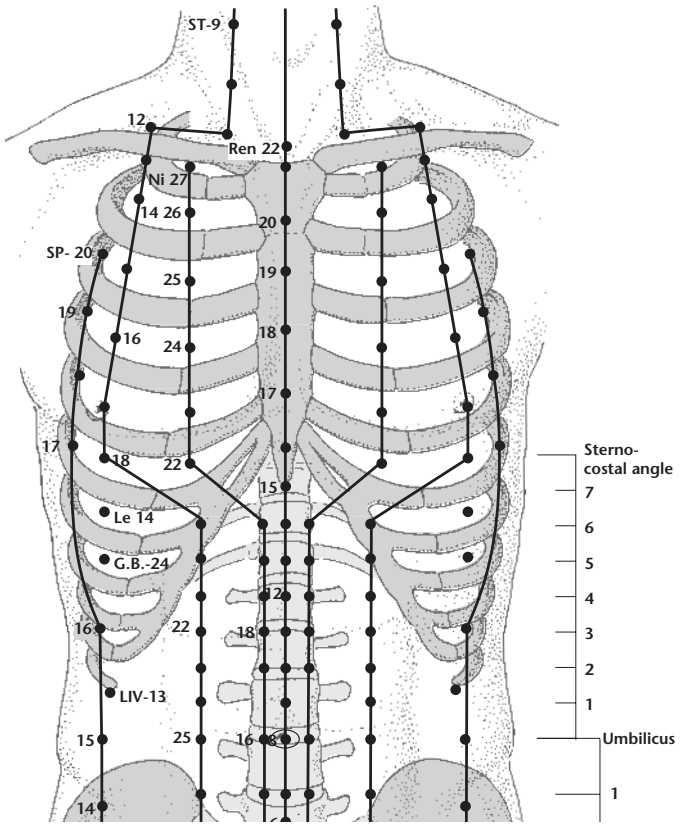

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 345

Spirit Storehouse SHENCANG KID-25

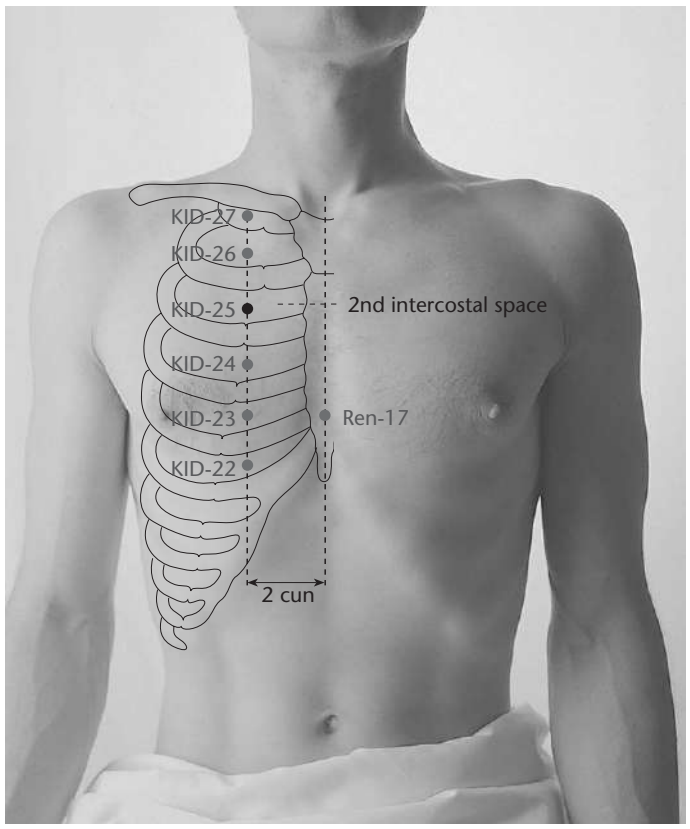

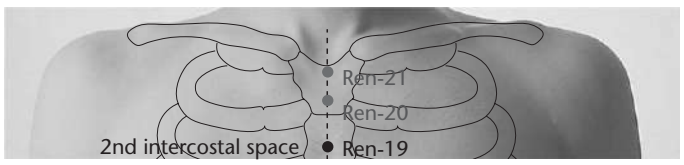

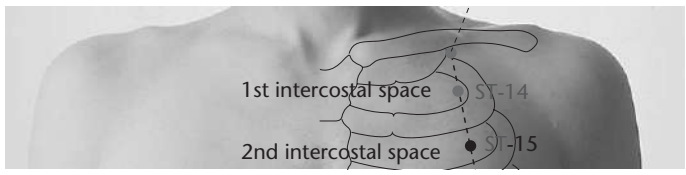

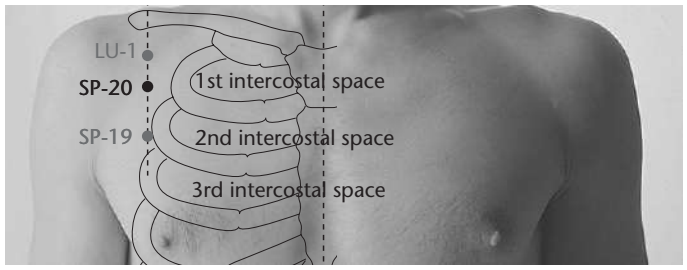

Location

In the 2nd intercostal space (ICS), 2 cun lateral to the anterior midline.

How to find

Locate the 2nd ICS by palpating parasternally from the clavicle or by counting from the manubriosternal synchondrosis (→ 2nd rib) (→ 3.5), then locate **KID-24** 2 cun lateral to the anterior midline. Located on the same level (→ 2nd ICS) are → **Ren-19** (on the midline), → **ST-15** (4 cun lateral to the midline) and → **SP-20** (6 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously) along the course of the ICS or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Regulates the Lung Qi
- Unbinds the chest

Special features

The *Su Wen* describes points **KID-22** to **KID-27** as *shu* points of the chest, indicated for difficult breathing and disorders of the thorax. They are particularly indicated for breathing difficulties due to 'excess above, deficiency below', which refers to the inability of the Kidneys to absorb the Qi that has descended from the Lung.

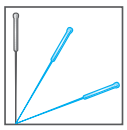
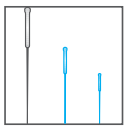
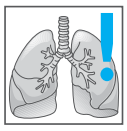

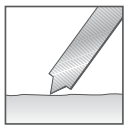

Ch04.8-F10028.qxd 2/22/08 6:14 PM Page 346

KID-26 *Comfortable Chest* YUZHONG

Location

In the 1st intercostal space (ICS), 2 cun lateral to the anterior midline.

How to find

Palpating parasternally from superior to inferior, the 1st rib can be palpated directly below the clavicle; in some cases, it can be completely hidden under the clavicle. Locate **KID-26** below the 1st rib in the 1st ICS, 2 cun lateral to the midline. Located on the same level (1st ICS) are → **Ren-20** (on the midline), → **ST-14** (4 cun lateral to the midline) and → **LU-1** (6 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously) along the course of the ICS or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Regulates the Lung Qi and transforms Phlegm
- Unbinds the chest
- Benefits the breasts

Special features

The *Su Wen* describes points **KID-22** to **KID-27** as *shu* points of the chest, indicated for difficult breathing and disorders of the thorax. They are particularly indicated for breathing difficulties due to 'excess above, deficiency below', which refers to the inability of the Kidneys to absorb the Qi that has descended from the Lung.

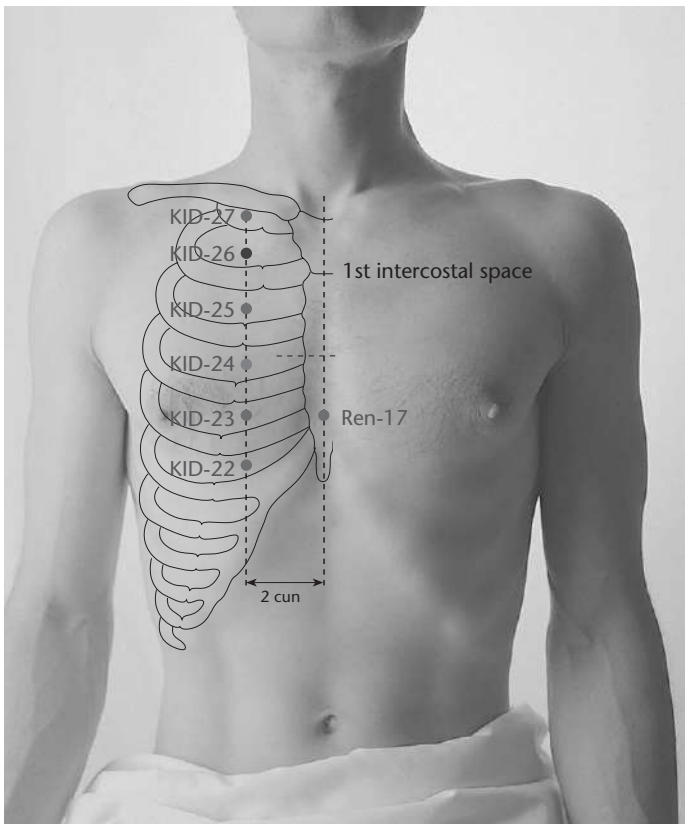

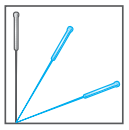
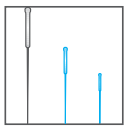
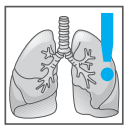

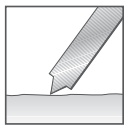

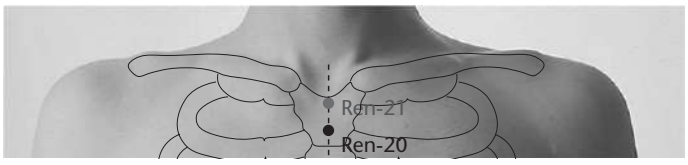

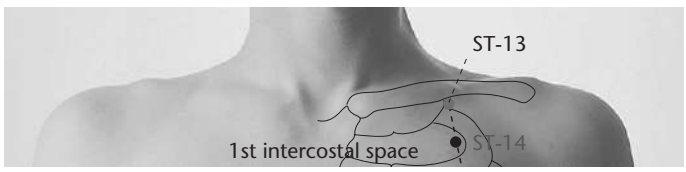

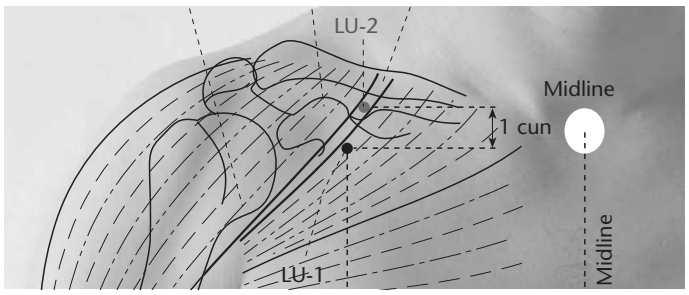

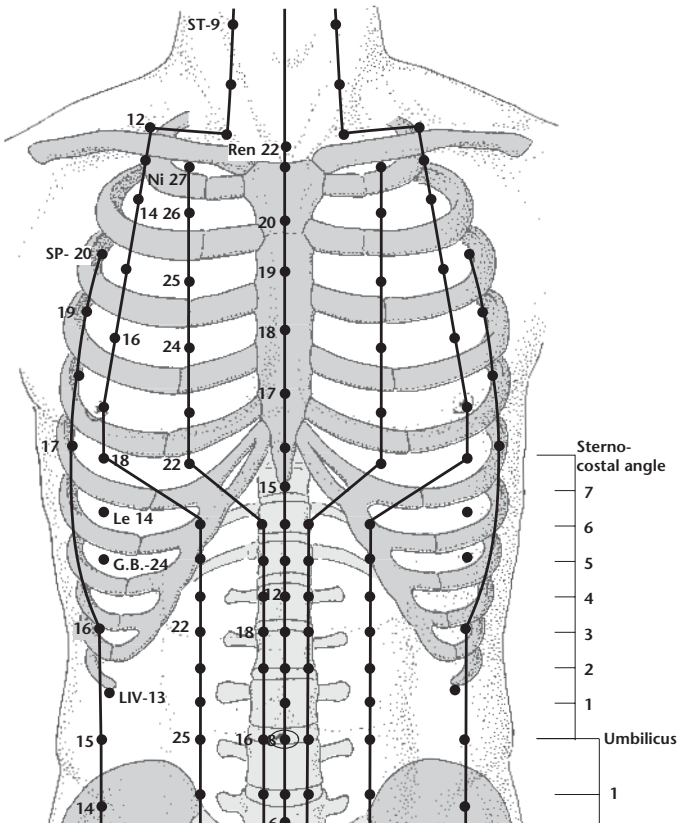

Ch04.8-F10028.qxd 2/22/08 6:15 PM Page 347

Shu *Mansion* SHUFU KID-27

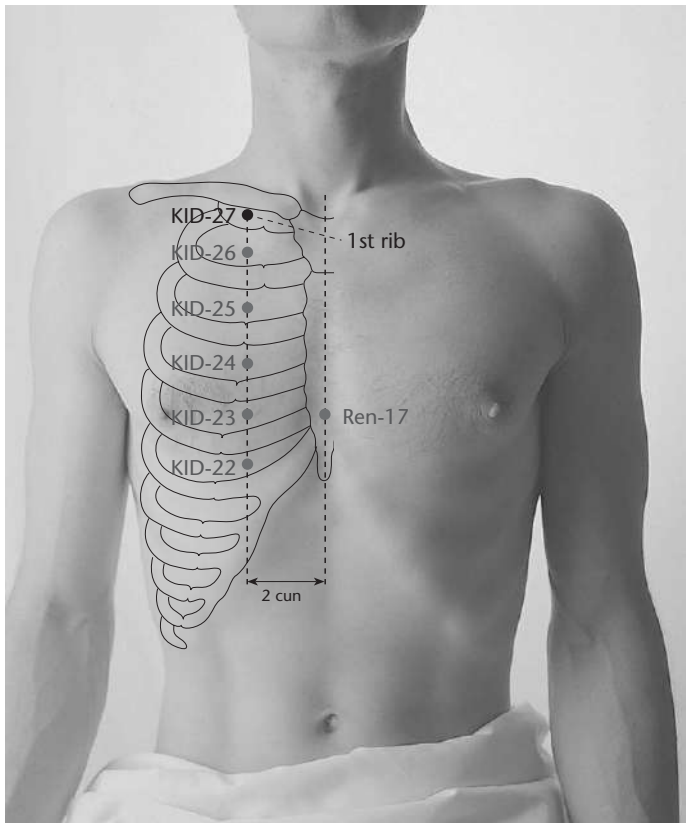

Location

At the lower border of the clavicle, 2 cun lateral to the anterior midline

How to find

When palpating parasternally from superior to inferior, the 1st rib can be felt directly below the clavicle; in some cases, it can be completely hidden under the clavicle. Therefore, **KID-27** will be located either on the 1st rib or, in some cases, at its lower border. → **ST-13** is located lateral to the short curvature of the 1st rib, below the midpoint of the clavicle.
Located on approximately the same level are → **Ren-21** (on the midline) and → **LU-2** (6 cun lateral to the midline).

Needling

0.5–0.8 cun obliquely or transversely (subcutaneously) along the course of the ICS or with or against the channel pathway. Caution: Pneumothorax.

Actions/Indications

- Regulates the Lung Qi and transforms Phlegm
- Unbinds the chest

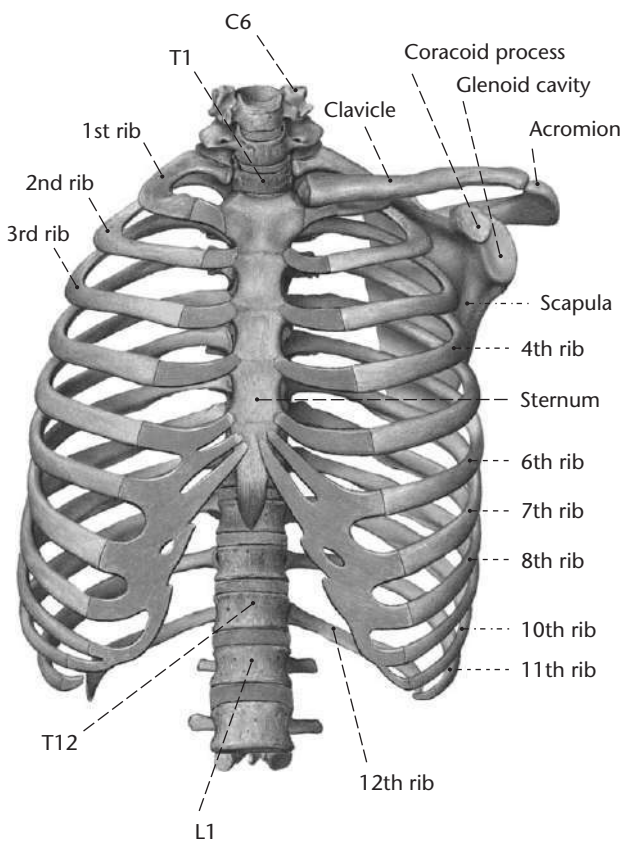

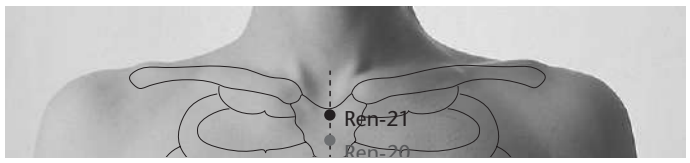

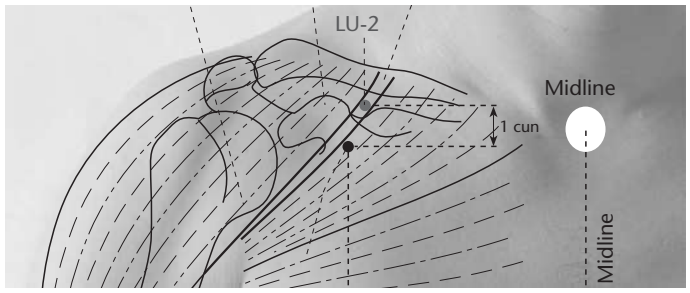

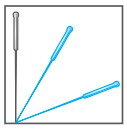
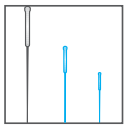
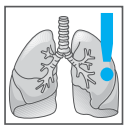

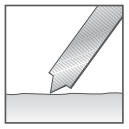

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 349

4.9 The Pericardium Channel System – Hand *jueyin* (*shou jue yin jing luo*)

4.9.1 Primary Pericardium Channel (*shou jue yin jing*)

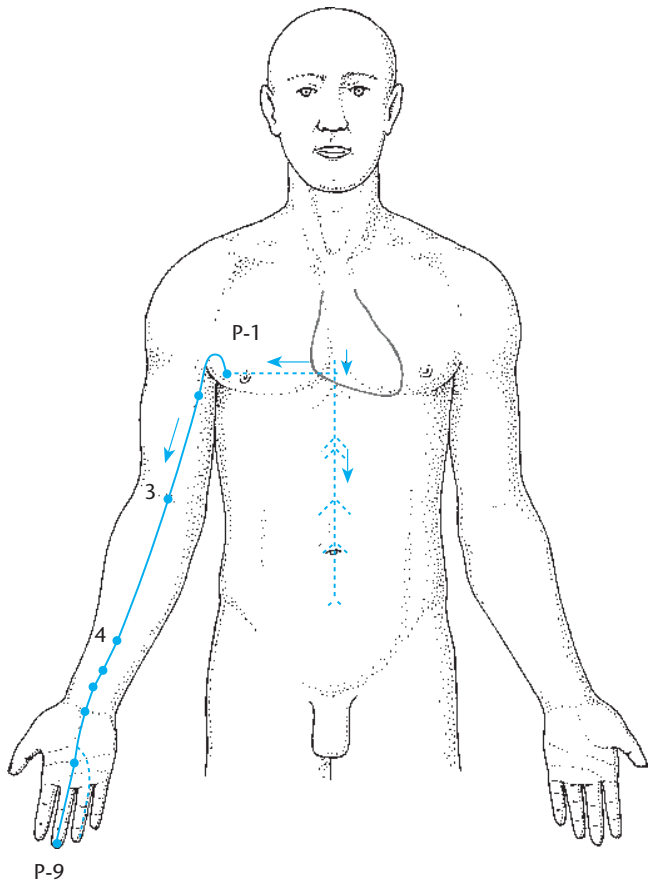

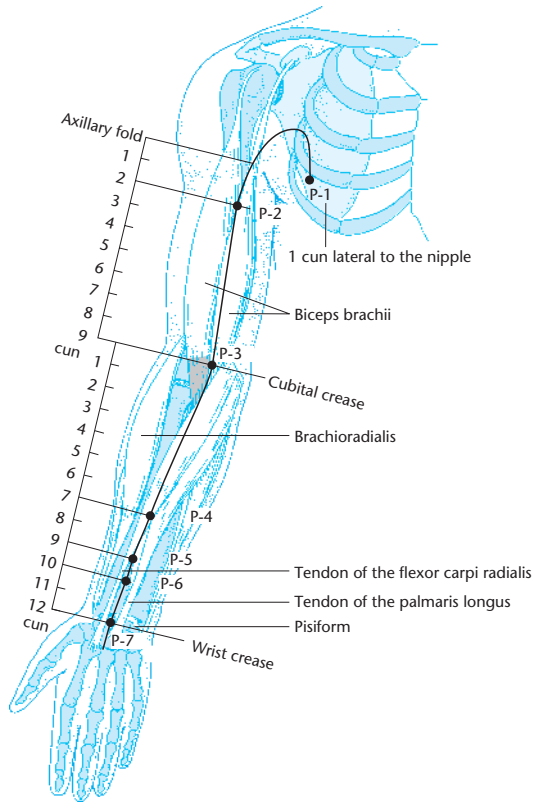

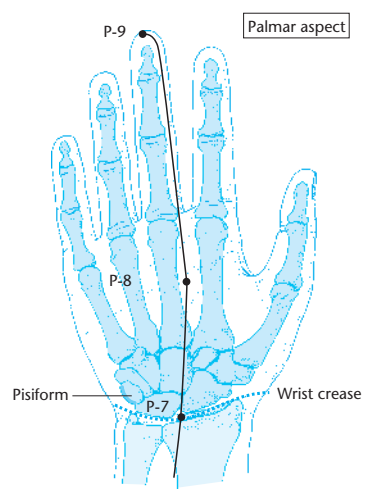

Pathway

The Pericardium primary channel begins in the centre of the thorax on the level of → **Ren-17** (*shanzhong*), where an internal branch of the Kidney primary channel ends, having also connected with the Liver, Lung and Heart (deep Yin–Yin connection).

The Pericardium primary channel then divides into two branches:

- The **first branch** penetrates the diaphragm, descends to the abdomen and connects with the Upper, Middle and Lower Burner.
- A **further internal branch** also originates in the thoracic region, continues to the breast and emerges approximately 1 cun lateral to the nipple in the 4th intercostal space at **P-1** (*tianchi*). Beneath **P-1**, an internal branch of the Liver primary channel emerges from the Liver and joins the Pericardium primary channel (Yin axes of the third great circuit: *jueyin*).

The **external** pathway ascends to the axilla, follows the medial aspect of the upper arm between the Lung and Heart primary channels, passes the elbow and continues along the forearm between the tendons of the palmaris longus and flexor carpi ulnaris muscles, ending at the tip of the middle finger.

Another **internal** branch originates in the centre of the palm at **P-8** (*laogong*), runs along the ring finger and connects at its tip with the Triple Burner primary channel (hand Yin–Yang connection of the third circuit).

Clinical importance (→ 1.2)

Exterior *(biao)* signs and symptoms: Occipital stiffness, leg cramps, red complexion, eye pain, swellings in the subaxillary region, hypertonic musculature with limited range of motion of the arm and elbow, Heat sensations in the palms

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 350

Interior (*li*) or *zangfu*-Organ signs and symptoms: Delirious speech, confusion, irritability, fullness and distension in the thorax and lateral costal region, aphasia, palpitations

Connections of the Pericardium primary channel (→ 1.2)

Connections with other channels

Triple Burner primary channel (*shou shao yang jing*)

Connection: Hand Yin–Yang pairing of the third great circuit
Location: P-8 → T.B.-1 (on the hand)
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Liver primary channel (*zu jue yin jing*)

Connection: Six-channel pairing of the third great circuit (Yin axes): *jue yin*
Location: LIV → P. An internal branch of the Liver primary channel emerges at the Liver, penetrates the diaphragm and connects with the Pericardium primary channel beneath **P-1** (*tianchi*).
Circulation: Non-circadian (**not** according to the Organ clock)
Importance: Above–below relationship

Kidney primary channel (*zu shao yin jing*)

Connection: Deep Yin–Yin connection
Location: KID → P (on the thorax). A branch of the internal pathway of the Kidney primary channel travels from the Kidney to the Liver, penetrates the diaphragm and spreads in the Lung (*fei*). From the Lung, an internal branch runs to the Heart (*xin*), where it joins the Pericardium primary channel (deep Yin–Yin connection) and further reaches → **Ren-17** (*shanzhong*).
Circulation: Circadian (according to the Organ clock)
Importance: The Pericardium primary channel receives Nutritive Qi (*ying qi*) from the Kidney primary channel (first circulation of the *ying qi* → 1.1.4).

Connections with *zangfu*-Organ systems

Pericardium (*xin bao*), Triple Burner (*san jiao*)

4.9.2 The Pericardium Divergent Channel (*shou jue yin jing bie*)

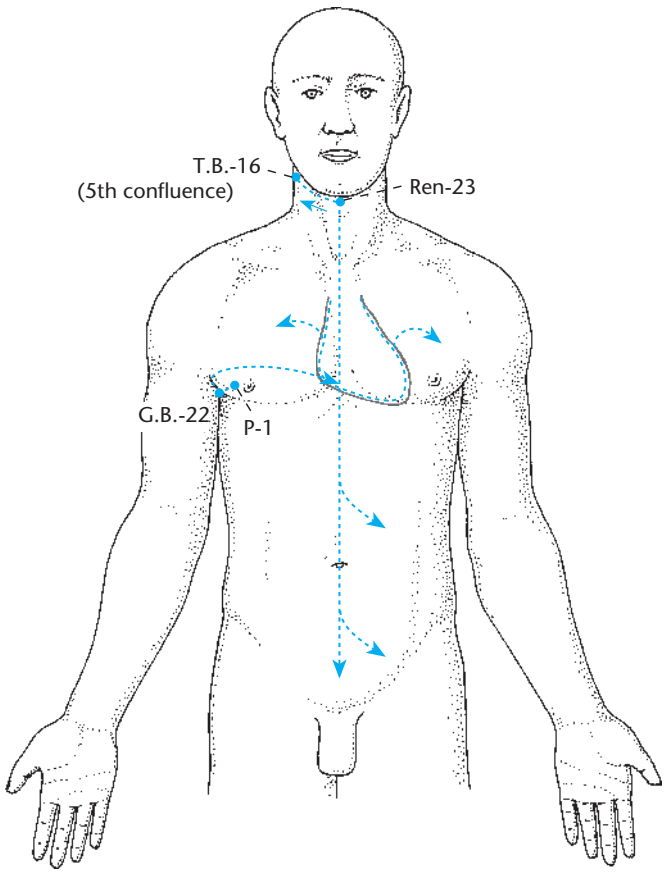

Pathway

The Pericardium divergent channel separates from the Pericardium primary channel near **P-1** (*tianchi*), 1 cun lateral to the nipple on the level of the 4th intercostal space. It

- ➡ runs in a lateral direction to → **G.B.-22** (*yuanye*)
- ➡ penetrates the thorax, continues to the Heart (*xin*) and divides into **two branches:**
  - **one branch** connects with the Upper, Middle and Lower Burner
  - the **other branch** ascends to the neck to → **Ren-23** (*lianquan*) and continues to the lateral aspect of the neck, emerges at → **T.B.-16** (*tianyou*) on the posterior border of the sternocleidomastoid muscle on the level of the mandibular angle, connecting with the Triple Burner primary channel and the Triple Burner divergent channel to form one of the 6 *he*-confluences (here: P/T.B. as 5th Confluence → 1.3)

Clinical importance

- Strengthens the relationship between the Pericardium and the Triple Burner (*zangfu*-Organ systems). Points on the Pericardium primary channel can therefore be used for disorders of the Triple Burner; vice versa, points on the Triple Burner primary channel can treat disorders of the Pericardium.

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 351

- In contrast to the Pericardium primary channel, the Pericardium divergent channel reaches the occiput and the neck. This explains why several points on the P channel can be used for disorders of the neck and ears, chronic throat pain, etc.
- By reaching the thorax, breast and hypochondrium, the Pericardium divergent channel supports the relationship between the Pericardium primary channel and these regions.

4.9.3 The Pericardium Sinew Channel (*shou jue yin jing jin*)

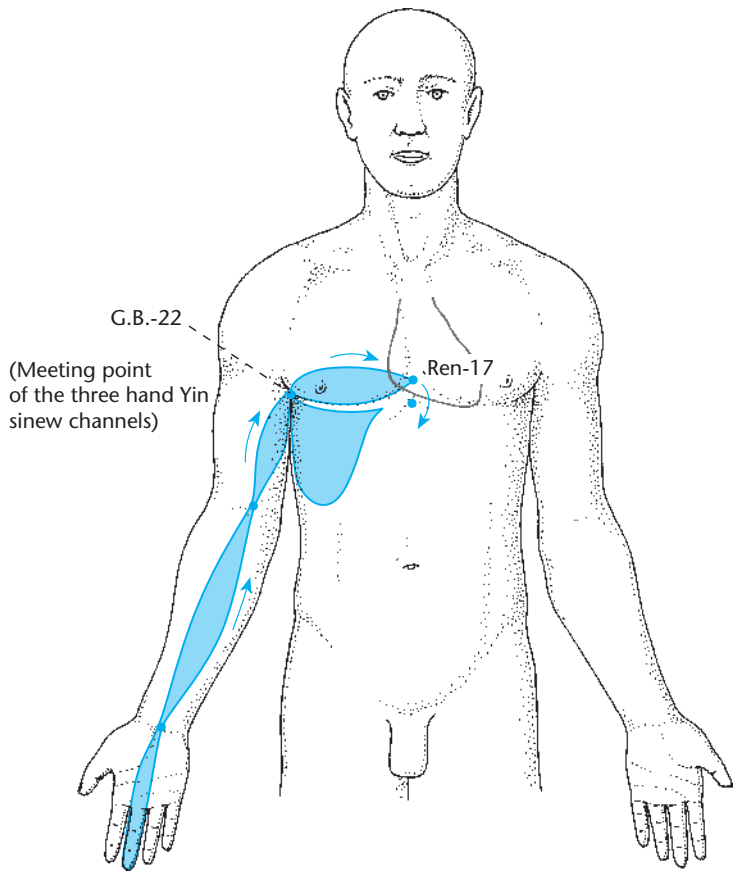

Pathway

The Pericardium sinew channel begins at the tip of the middle finger

- ➡ runs along the anterior aspect of the middle finger and along the 3rd metacarpal bone
- ➡ spreads in the centre of the palm and meets the Lung sinew channel
- ➡ binds (*jie*) in the centre of the wrist
- ➡ ascends the anterior aspect of the forearm as well as the anteromedial aspect of the upper arm to the axilla, binding (*jie*) at **P-3** (*quze*) on the elbow and at ➞ **G.B.-22** (*yuanye*), where it meets the other hand Yin sinew channels.

At ➞ **G.B.-22**, the channel separates into two branches:

- **one branch** disperses on the anterior, lateral and posterior aspects of the costal region
- **the other branch** penetrates the thorax below the axilla and reaches the midthoracic region on the level of ➞ **Ren-17** (*shanzhong*), from where it spreads in the thoracic cavity to bind (*jie*) at the diaphragm.

Clinical importance

Pathology: Stiffness, pain, spasms and distending sensations along the pathway of the Pericardium sinew channel. Thoracic pain and tightness, *xi fen* syndrome (one of the five *ji*-syndromes, an 'accumulation syndrome') due to Lung Qi stagnation with accumulation of Phlegm and Heat manifesting with symptoms such as masses in the right hypochondriac region, possibly accompanied by thoracic or back pain, haemoptysis, chills and fever, shortness of breath and dyspnoea, etc.

Indication: Mainly for pain, spasms and distending sensations along the pathway of the sinew channel, limited range of motion along the medial aspect of the arm and elbow, swellings and pain in the axilla and the breast.

4.9.4 The Pericardium *luo*-Connecting Vessel System (*shou jue yin luo mai*)

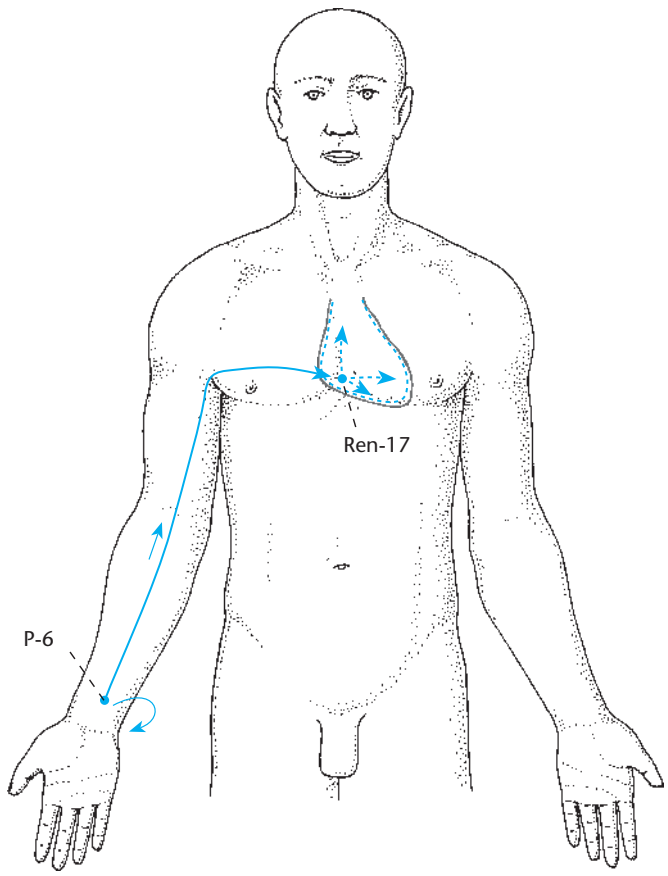

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 352

Pathway

The *luo*-connecting Pericardium vessel system separates from the Pericardium primary channel at its associated *luo*-connecting point **P-6** (*neiguan*) (→ 8.1.2). It forms a three-dimensional reticular network, dividing into multiple branches and sub-branches (*sun luo, fu luo, xue luo* → 1.5) within the surrounding tissue.

- Horizontal divisions run to the Interiorly–Exteriorly paired Triple Burner primary channel; according to some schools of thought (for example, Ngyen Van Nghi → Appendix), they travel as a **transverse** Pericardium *luo*-connecting vessel to the *yuan*-source point → **T.B.-4** (*yangchi*).
- **A longitudinal division** follows the Pericardium primary channel along the anterior aspect of the arm, crosses → **Ren-17** (*shanzhong*) on the midline on the level of the 4th intercostal space, penetrates the thorax and terminates at the Pericardium (*xin bao*) and the Heart (*xin*).

Clinical importance (→ 8.1.2)

Pathology

Excess (*shi*): Heart pain, angina pain, thoracic pain
Deficiency (*xu*): Restlessness, irritability

4.9.5 Cutaneous Region (*jue yin pi bu*)

See description and figures → 1.6.

4.9.6 Points of the Primary Pericardium Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): P-7** (*daling*) ■■
- ***Luo*-connecting point (→ 8.1.2.): P-6** (*neiguan*) ■■
- ***Xi*-cleft point (→ 8.1.3): P-4** (*ximen*) ■■
- **Associated Back-*shu* point (→ 8.1.4): BL-14** (*jueyinshu*)
- **Associated Front-*mu* point (→ 8.1.5): Ren-17** (*shanzhong*) ■■
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Wood), tonification point: **P-9** (*zhongchong*) ■■
  ying-spring point (Fire), *ben* point (Five Phases): **P-8** (*laogong*) ■■
  shu-stream point (Earth), sedation point: **P-7** (*daling*) ■■
  jing-river point (Metal): **P-5** (*jianshi*) ■■
  he-sea point (Water): **P-3** (*quze*) ■■
- ***Hui*-meeting point (→ 8.1.7): –**
- **Opening point (→ 8.1.8) of the *yin wei mai:* P-6** (*neiguan*) ■■
- **Lower *he*-sea point (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the LIV, T.B., G.B.* channels: **P-1** (*tianchi*)
  – of other channels with the Pericardium channel: –
- **Gao Wu command point (→ 8.1.11) for the thoracic region: P-6** (*neiguan*) ■■
- **Window of Heaven point (→ 8.1.12): P-1** (*tianchi*)
- **Points of the Four Seas (→ 8.1.13): –**
- **Ma Dan Yang Heavenly Star point (→ 8.1.14): –**
- **Sun Si Miao Ghost point (→ 8.1.15): P-7** (*daling*) ■■, **P-8** (*laogong*) ■■
- **Other functional points:**
  – important point for eliminating Phlegm misting the orifices: **P-5** (*jianshi*) ■■
  – important point for calming the *shen*: **P-6** (*neiguan*) ■■

Points according to region

- **Local points (→ 8.2.1):** hand – **P-7** (*daling*) ■■
- **Adjacent points (→ 8.2.1): –**
- **Distal points (→ 8.2.1):** tongue – **P-8** (*laogong*) ■■; Lung – **P-6** (*neiguan*) ■■; Heart – **P-4** (*ximen*) ■■; **P-6** (*neiguan*) ■■; Spleen – **P-6** (*neiguan*) ■■; thorax and epigastrium – **P-6** (*neiguan*) ■■; foot – **P-7** (*daling*) ■■

Specific points according to the channel pathway (in numeric order)

- **P-1** (*tianchi*) ■■: *jiaohui*-meeting point with the LIV, T.B., G.B.* channels (→ 8.1.10); Window of Heaven point (→ 8.1.12)
- **P-3** (*quze*) ■■: lower *he*-sea point (Water) (→ 8.1.6)
- **P-4** (*ximen*) ■■: *xi*-cleft point (→ 8.1.3); distal point for the Heart (→ 8.2.1)
- **P-5** (*jianshi*) ■■: *jing*-river point (Metal) (→ 8.1.6); important point for eliminating Phlegm misting the orifices
- **P-6** (*neiguan*) ■■: *luo*-connecting point (→ 8.1.2); Gao Wu command point (→ 8.1.11) for the thorax; Opening point (→ 8.1.8) of the *yin wei mai*; distal point for the Heart, Spleen, thorax and epigastrium (→ 8.2.1); important point for calming the *shen*
- **P-7** (*daling*) ■■: *yuan*-source point (→ 8.1.1); *shu*-stream point (Earth) (→ 8.1.6); sedation point; Sun Si Miao Ghost point (→ 8.1.15); local point for the hand (→ 8.2.1); distal point for the foot (→ 8.2.1)
- **P-8** (*laogong*) ■■: *ying*-spring point (Fire) (→ 8.1.6); *ben* point (Five Phases); Sun Si Miao Ghost point (→ 8.1.15); distal point for the tongue (→ 8.2.1)
- **P-9** (*zhongchong*) ■■: *jing*-well point (Wood) (→ 8.1.6); tonification point

* Mentioned by only some authors

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 353

Heavenly Pool TIANCHI P-1

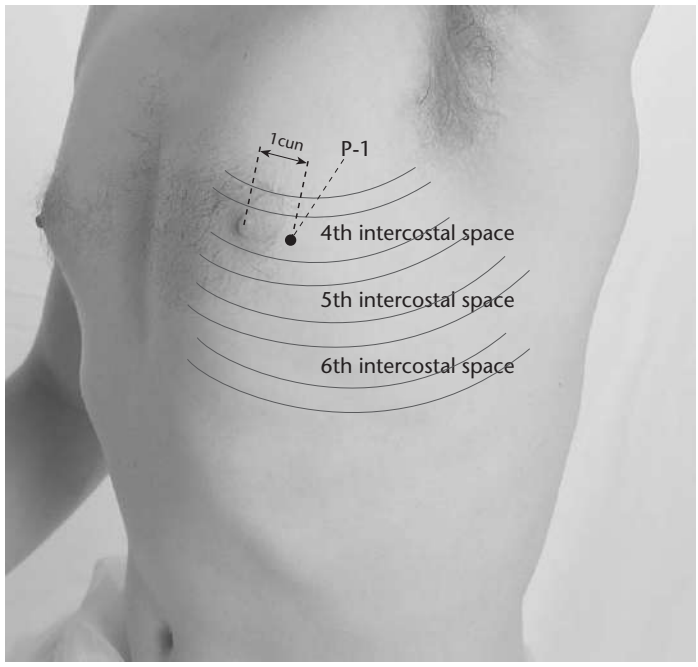

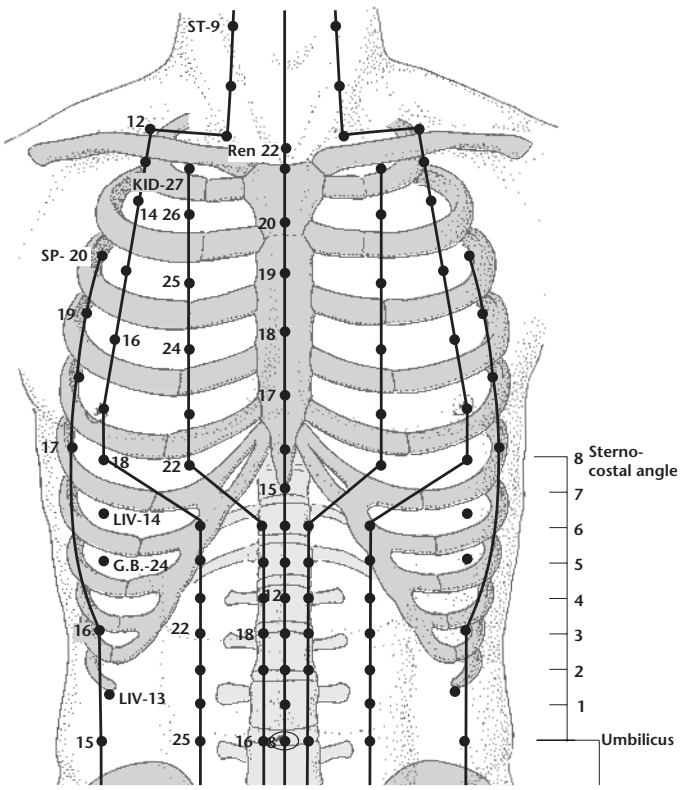

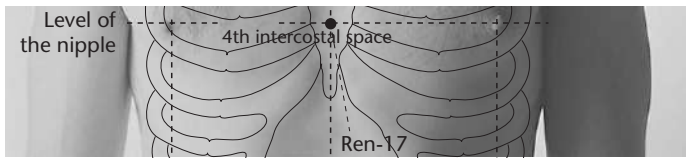

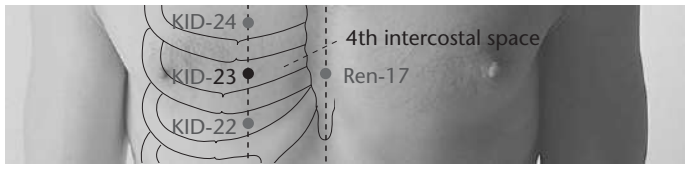

Location

In the 4th intercostal space (ICS), 1 cun lateral to the nipple.

How to find

Quick method, especially in male patients: The 4th ICS is located on the same level as the nipple. Locate **P-1** 1 cun lateral to the nipple in the 4th ICS (note: ascending course of the ICS towards lateral). **Or:** For better orientation in the intercostal region, first, locate the manubriosternal synchondrosis (→ 3.4.2.). This is on the level of the costal cartilage of the 2nd rib, followed by the 2nd ICS below it. Count down to ICS to the 4th ICS and measure 5 cun in a lateral direction. There, locate **P-1**.
Located on the same level (4th ICS) are → **Ren-17** (on the midline), → **KID-23/ST-17/SP-18** (2/4/6 cun lateral to the midline), as well as → **G.B.-22** and → **G.B.-23**.

Needling

0.5–1 cun transversely (subcutaneously) along the course of the ICS. Especially in women, transverse needling only, avoid puncturing the mamillary tissue. Caution: Pneumothorax.

Actions/Indications

- Descends counterflow Qi
- Eliminates excess (especially Heat and Phlegm) from the thorax
- Promotes lactation and eliminates local Qi stagnation

Special features

Meeting point with the LIV and T.B. channels, also with the G.B. channel according to some authors. Window of Heaven point, entry point.

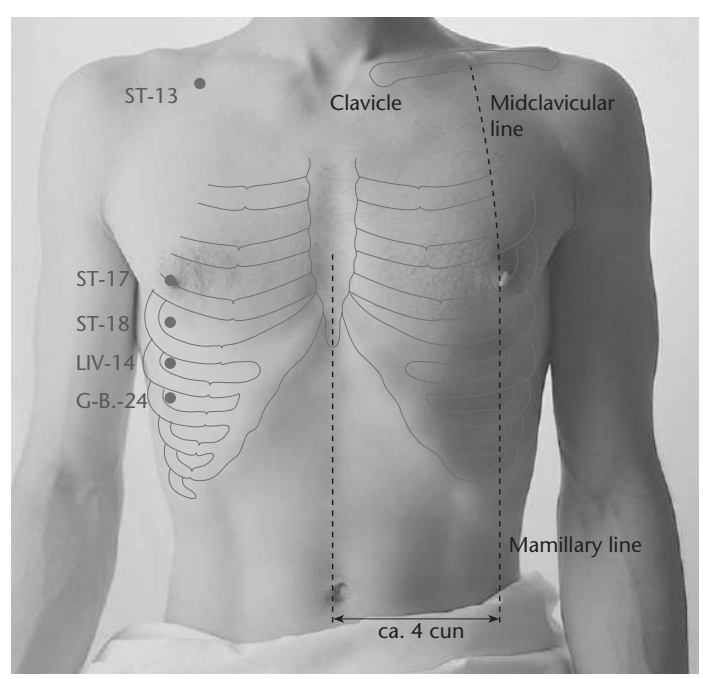

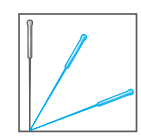
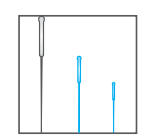
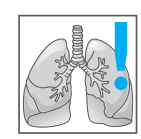

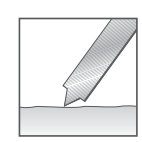
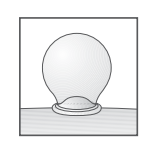

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 354

P-2 *Heavenly Spring* TIANQUAN

Location

Between the two heads of the biceps brachii muscle, 2 cun inferior to the axillary fold.

How to find

Locate the biceps brachii muscle through slight external rotation; in patients with less muscle definition, ask the patient to flex their elbow against resistance. **P-2** is located between the two heads of the biceps brachii muscle, 2 cun inferior to the upper end of the anterior axillary fold (→ 2.2).

→ **LU-3** is located radial to **P-2,** on the lateral aspect of the biceps brachii muscle and 1 cun more distally.

Needling

1–1.5 cun obliquely in a proximal or lateral direction.

Actions/Indications

- Unbinds the chest.
- Nourishes and invigorates the Blood
- Calms the *shen*

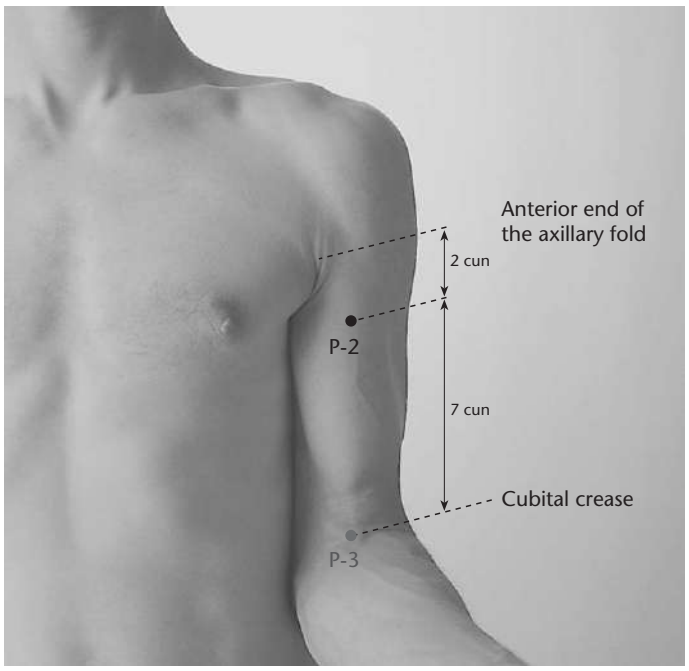

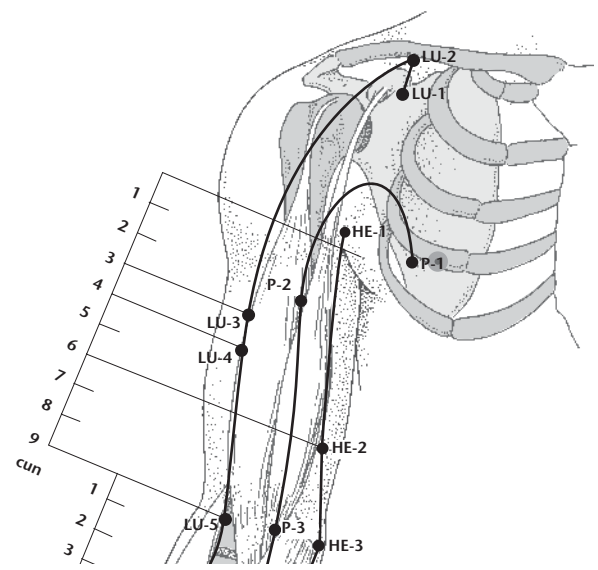

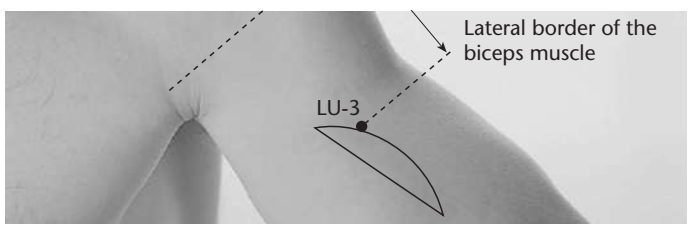

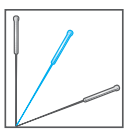
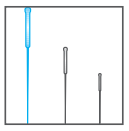

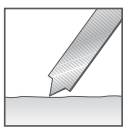
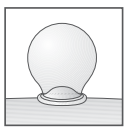

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 355

Marsh at the Crook QUZE P-3

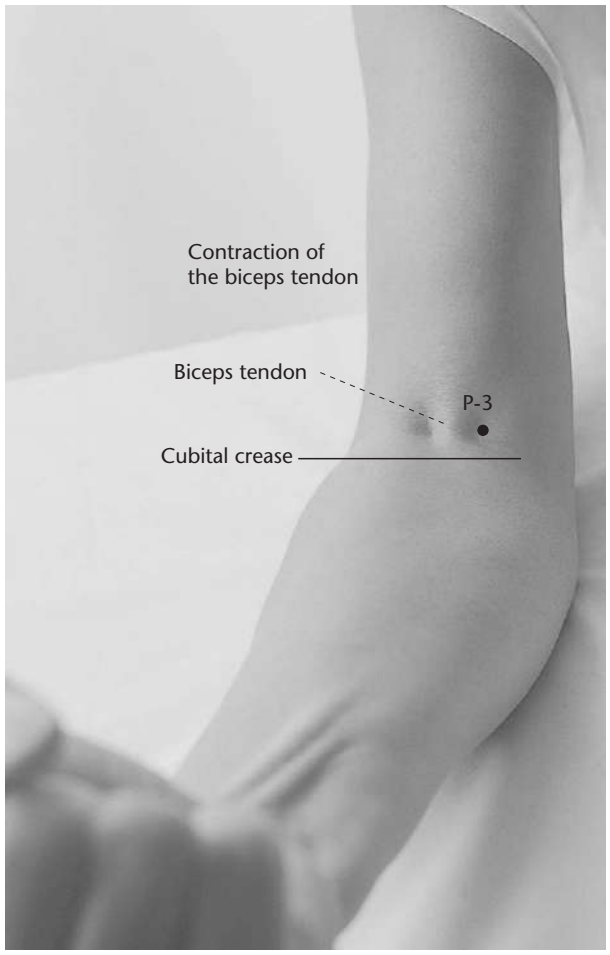

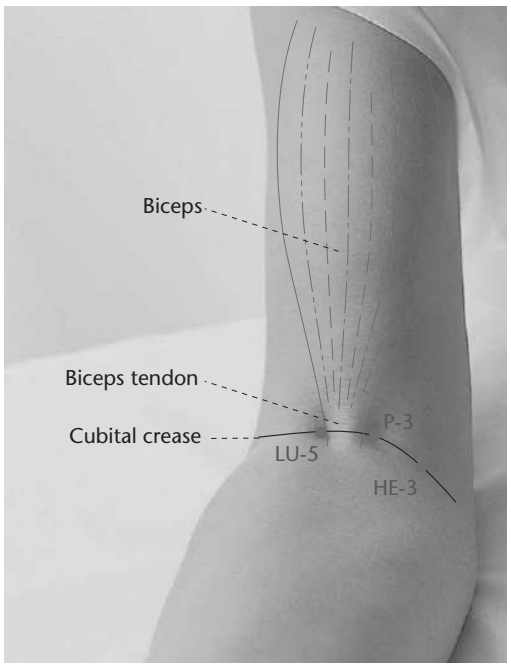

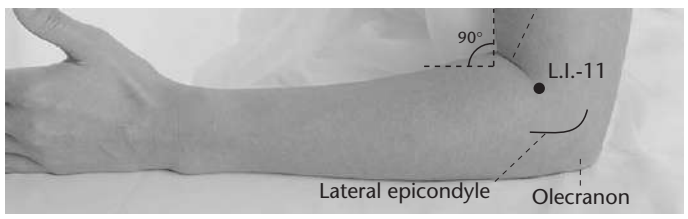

Location

On the cubital crease, on the ulnar side of the biceps tendon, between the tendon and the brachial artery.

How to find

Place the patient's arm in a relaxed supinated position. This point is best located with the elbow bent and the biceps flexed in order to make the tendon and elbow crease more clearly visible and palpable. **P-3** is located on the cubital crease, on the ulnar aspect of the tendon.

Also located in the area of the cubital crease are → **LU-5** (on the radial side of the tendon), → **L.I.-11** (between the radial end of the crease and the lateral epicondyle of the humerus) and → **HE-3** (with the elbow fully flexed, at the ulnar end of the cubital crease).

Needling

0.5–1 cun vertically or prick to cause bleeding of the vein. Caution: Brachial artery/vein. Palpate for the pulse before needling.

Actions/Indications

- Clears Heat, drains Fire
- Harmonises the Stomach and Intestines, stops vomiting
- Opens the channel, alleviates pain

Special features

He-sea point, Water point. Important local point and important point to drain Heat. For Heat conditions, apply bloodletting.

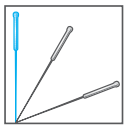
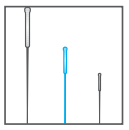

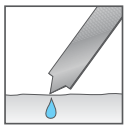
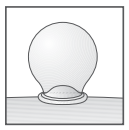

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 356

P-4 Xi *Cleft Gate* XIMEN

Location

5 cun proximal to the anterior space of the wrist joint (most distal wrist crease), between the tendons of the palmaris longus and flexor carpi radialis muscles.

How to find

Spreading hands technique (→ 2.3.3): Place the little fingers in the centre of the anterior wrist joint space (→ 3.3.3, → **P-7**) and the cubital crease respectively (→ **P-3**). From the midpoint of this distance, measure 1 cun in a distal direction. On this level, locate **P-4** between the two tendons, which become more pronounced when the fingertips are pressed together. If only one tendon is visible, this will be the tendon of the flexor carpi radialis muscle. **P-4** is located on its ulnar aspect.

Needling

0.5–1 cun vertically or 1–1.5 cun obliquely in a proximal direction.

Actions/Indications

- Tonifies the Blood, dispels Blood stasis
- Calms the *shen*
- Cools the Blood
- Opens the channel

Special features

Xi-cleft point. Major point for acute, painful Blood stasis in the region of the thorax and the Heart

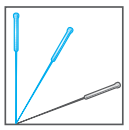
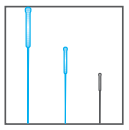

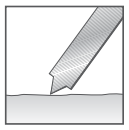

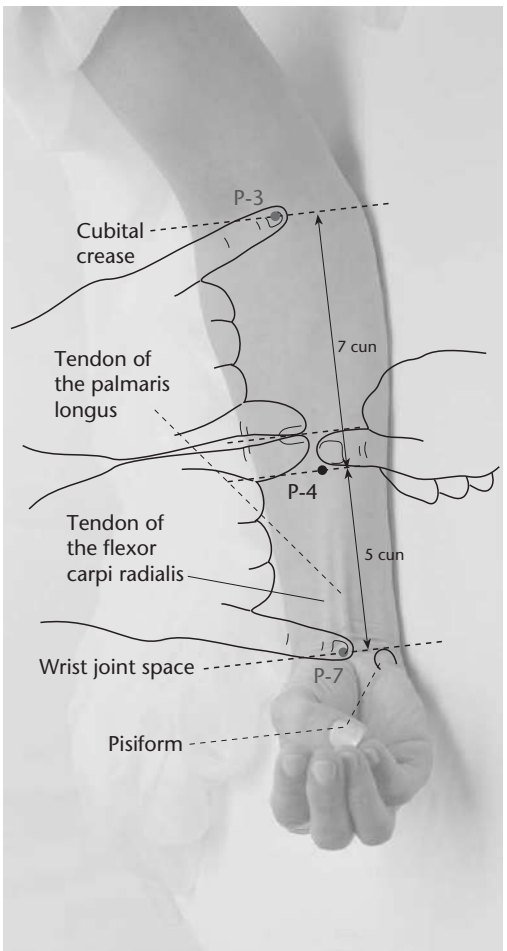

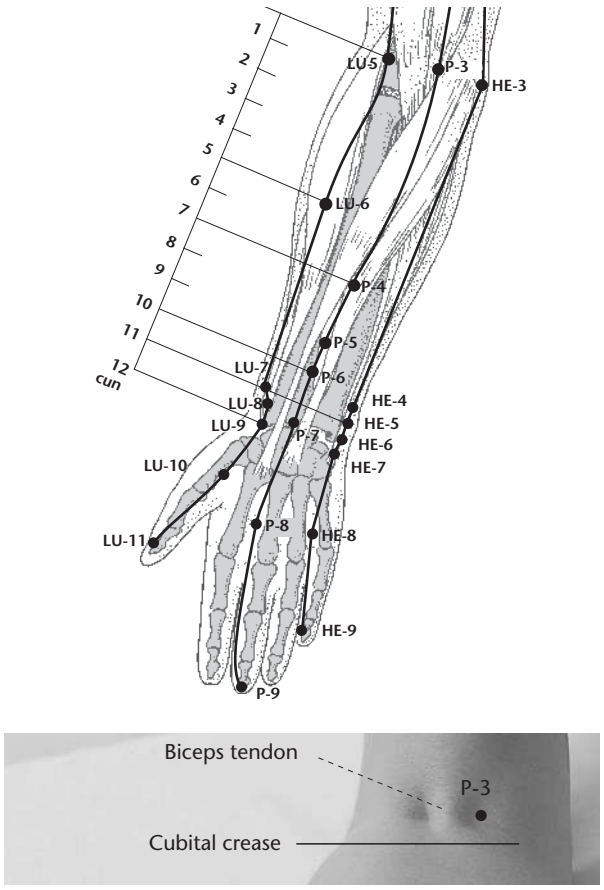

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 357

Intermediary Messenger JIANSHI P-5

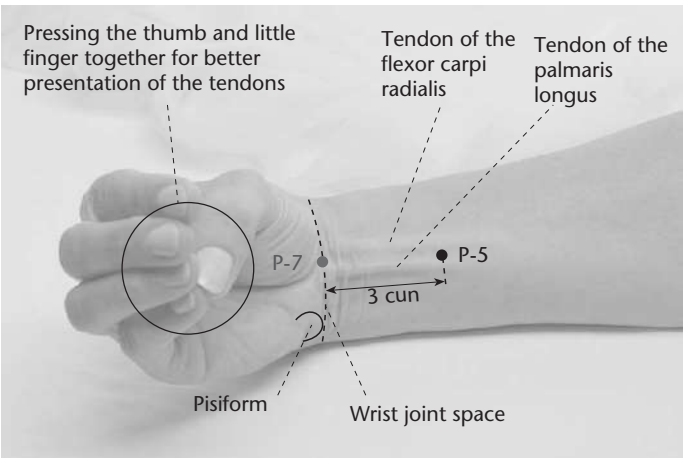

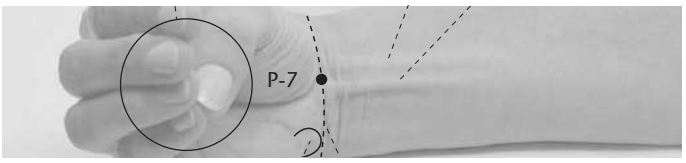

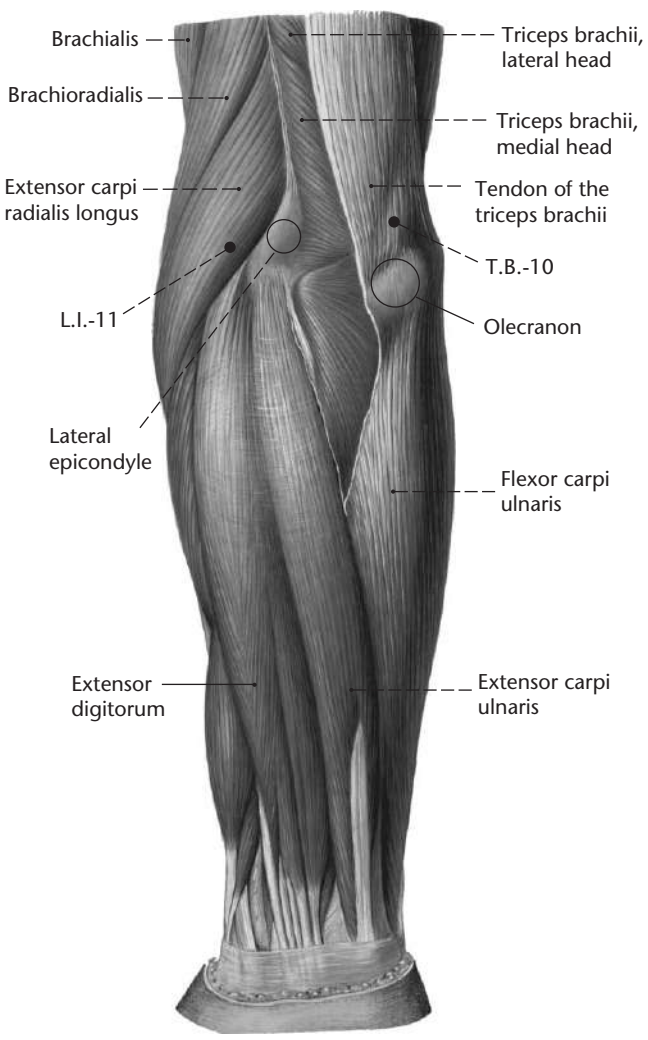

Location

3 cun proximal to the anterior wrist joint space ('most distal wrist crease'), between the tendons of the palmaris longus and flexor carpi radialis muscles.

How to find

As the location of the wrist crease varies, the wrist joint space is a more reliable reference point (→ 3.3.3). It can easily be palpated by moving the patient's hand in a relaxed way. From the centre of the joint space (→ **P-7**), measure 3 cun in a proximal direction and locate **P-5** between the two tendons, which become more pronounced when the fingertips are pressed together. If only one tendon is visible, this will be the tendon of the flexor carpi radialis muscle. **P-5** is located at its ulnar aspect.

Needling

0.5–1 cun vertically or obliquely in a proximal direction. Caution: The median nerve here is only covered by the radial head of the flexor digitorum superficialis. Vertical needling should not exceed 0.5 cun; no strong needle stimulation.

Actions/Indications

- Calms the *shen* and the Heart
- Harmonises the Middle Burner
- Clears Heart Fire
- Unbinds the chest
- Transforms invisible Phlegm

Special features

Jing-river point, Metal point. Important point for transforming Phlegm blocking the orifices.

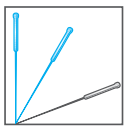
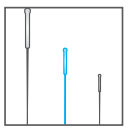

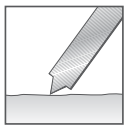

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 358

P-6 *Inner Gate* NEIGUAN

Location

2 cun proximal to the anterior wrist joint space ('most distal wrist crease'), between the tendons of the palmaris longus and flexor carpi radialis muscles.

How to find

As the location of the wrist crease varies, the wrist joint space is a more reliable reference point (→ 3.3.3). It can easily be palpated by moving the patient's hand in a relaxed way. From the centre of the wrist joint space (→ **P-7**), measure 2 cun in a proximal direction and locate **P-6** between the two tendons, which become more pronounced when the fingertips are pressed together. If only one tendon is visible, this will be the tendon of the flexor carpi radialis muscle. **P-6** is located on its ulnar aspect.
Or quick method: From the centre of the wrist joint space, glide in a proximal direction between the two tendons. The skin will begin to form a bulge in front of the sliding finger, which will come to a halt at **P-6.**
→ **T.B.-5** (2 cun proximal to the dorsal wrist joint space) is located opposite **P-6**.

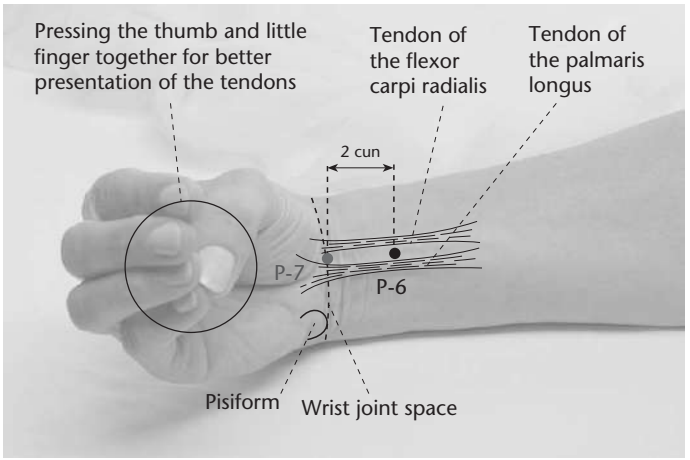

Needling

0.5–1 cun vertically or obliquely in a proximal or lateral direction. Caution: The median nerve is located directly below this point.

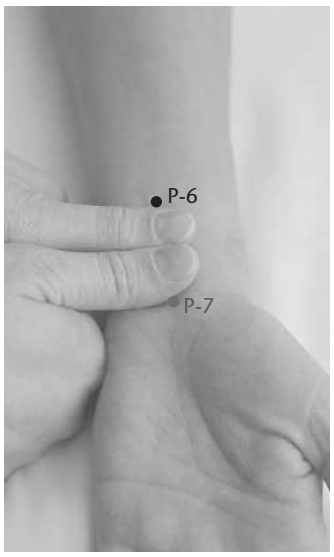

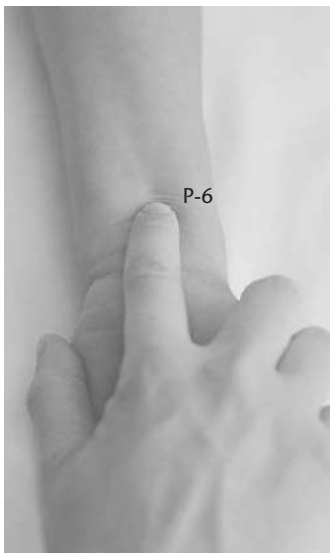

Actions/Indications

- Calms the *shen*
- Unbinds the chest
- Harmonises the Stomach
- Opens the channel

Special features

Luo-connecting point, opening (master) point of the *yin wei mai*, Gao Wu command point for the thorax. Main point for nausea and vomiting. Very important point.

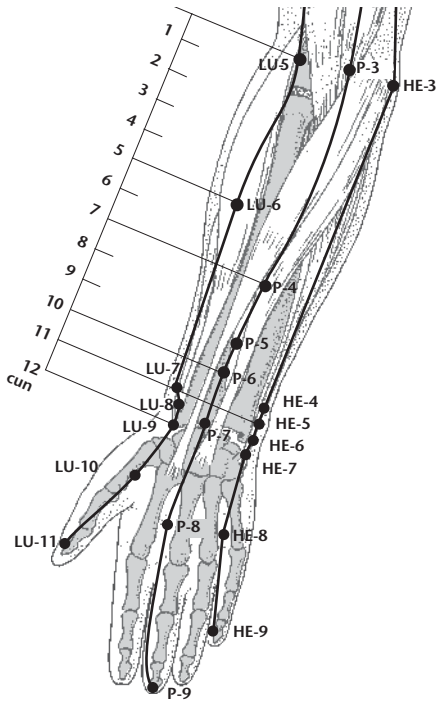

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 359

Great Mound DALING P-7

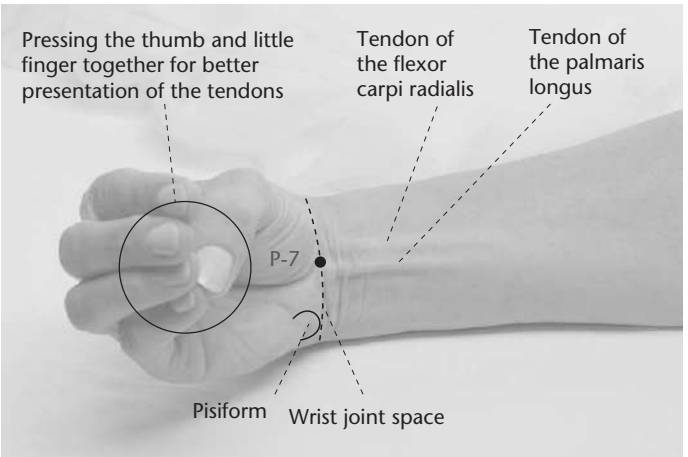

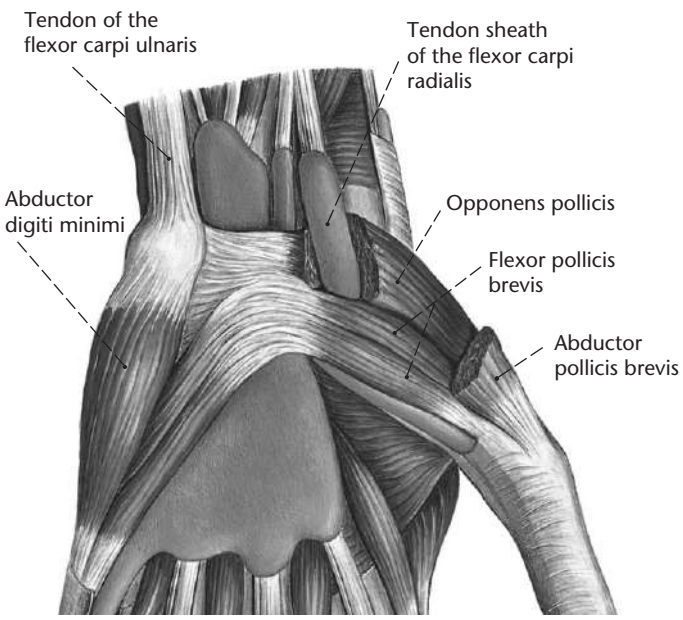

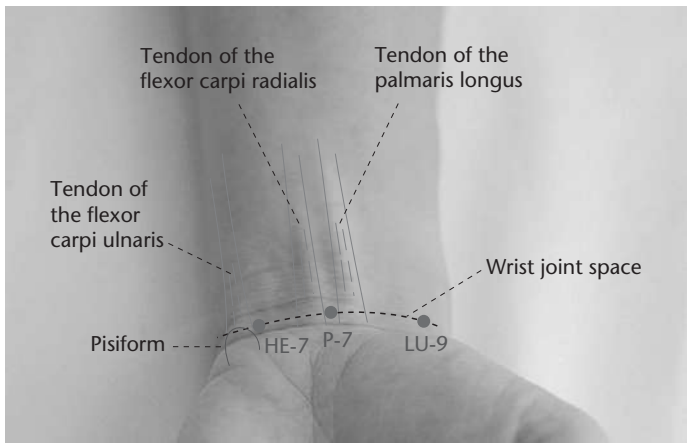

Location

On the anterior aspect of the wrist joint space ('most distal wrist crease'), between the tendons of the palmaris longus and flexor carpi radialis muscles.

How to find

As the location of the wrist crease varies, the wrist joint space is a more reliable reference point (→ 3.3.3). By moving the patient's hand in a relaxed way, it can easily be palpated. Locate **P-7** on this level between the two tendons, which become more pronounced when the fingertips are pressed together. If only one tendon is visible, this will be the tendon of the flexor carpi radialis muscle. **P-7** is located on its ulnar aspect.

Located on the same level are → **HE-7** (on the ulnar aspect of the wrist joint, radial to the tendon of the flexor carpi ulnaris muscle) and → **LU-9** (on the radial aspect of the wrist joint, lateral to the radial artery.

Needling

Vertically 0.3–0.5 cun, for carpal tunnel syndrome also 0.5–1 cun obliquely in a lateral direction along the tunnel. Caution: The median nerve is located directly under this point, thus needling can cause a significant electric sensation; in this case, no further needle manipulation to avoid possible injury.

Actions/Indications

- Clears Heat from the Heart, calms the *shen*
- Harmonises the Stomach and Intestines
- Cools Heart Fire affecting the Bladder
- Unbinds the chest
- Cools the Blood
- Opens the channel, alleviates pain

Special features

Yuan-source point, *shu*-stream point, Earth point, sedation point, Sun Si Miao Ghost point. Important point for calming, main point for carpal tunnel syndrome.

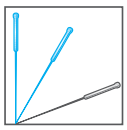
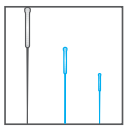

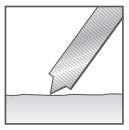

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 360

P-8 *Palace of Toil* LAOGONG

Location

In the centre of the palm, between the 2nd and 3rd metacarpal bones, slightly closer to the 3rd metacarpal bone.

How to find

When making a loose fist, the tip of the middle finger will rest on **P-8,** between the 2nd and 3rd metacarpal bones, slightly closer to the 3rd metacarpal bone.

→ **HE-8** is located more radially, between the 4th and 5th metacarpal bones.

Needling

Vertically 0.3–0.5 cun. Thinner needles should be used as needling this point can be painful.

Actions/Indications

- Clears Heart Fire and calms the *shen*
- Harmonises the Middle Burner
- Clears the *ying*-nutritive level and cools the Blood

Special features

Ying-spring point, Fire point, *ben* point (Five Phases), Sun Si Miao Ghost point, exit point.

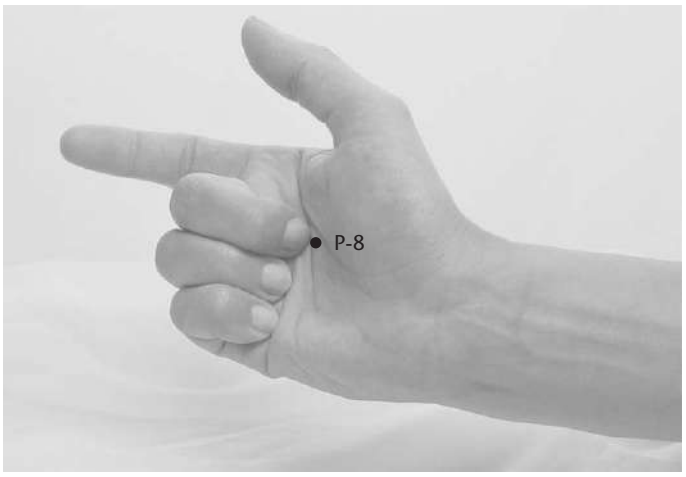

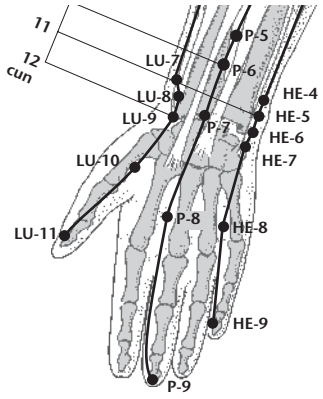

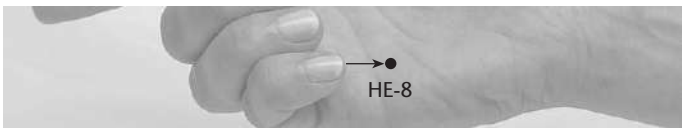

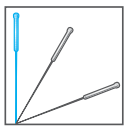

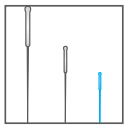

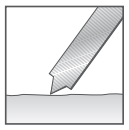

Ch04.9-F10028.qxd 2/23/08 3:57 PM Page 361

Middle Rushing ZHONGCHONG P-9

Location
On the most distal point of the middle finger.

How to find
Find the most distal point on the tip of the middle finger and there locate **P-9**.

Needling
Vertically 0.1–0.2 cun or prick to bleed. Caution: Painful point.

Actions/Indications
- Clears Heat, especially from the Heart and Pericardium
- Strengthens collapsed Yang
- Expels internal Wind

Special features
Jing-well point, Wood point, tonification point.

Ch04.10-F10028.qxd 2/23/08 2:26 PM Page 363

4.10 The Triple Burner Channel System – Hand *shaoyang* (*shou shao yang jing luo*)

4.10.1 Triple Burner Primary Channel (*shou shao yang jing*)

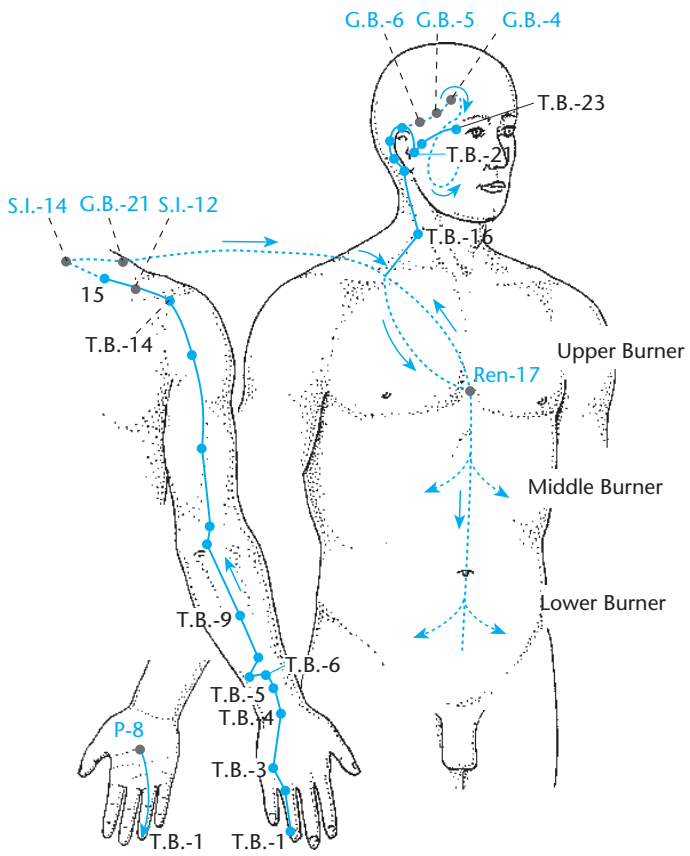

Pathway

The **external pathway** of the Triple Burner primary channel begins on the ring finger, at the ulnar corner of the nail at **T.B.-1** (*guanchong*). **T.B.-1** is reached by a branch of the Pericardium primary channel that separates on the palm of the hand at → **P-8** (*laogong*) (hand Yin–Yang connection of the third great circuit). The channel runs along the ulnar side of the ring finger

- continues between the 4th and 5th metacarpal bones
- travels along the posterior aspect of the forearm between the radius and the ulna towards the lateral aspect of the arm
- passes the olecranon and continues along the upper arm to the shoulder
- traverses the shoulder, intersects with → **S.I.-12** (*bingfeng*) and meets the other Yang primary channels at → **Du-14** (*dazhu*)
- runs to the anterior aspect of the body, crossing → **G.B.-21** (*jianjing*) and → **ST-12** (*quepen*), the latter in the supraclavicular fossa.

At **ST-12,** the channel enters the body to begin its **internal** section. It meets the Pericardium (*xin bao*) at → **Ren-17** (*shanzhong*), penetrates the diaphragm and enters the abdomen, connecting with the Upper, Middle and Lower Burners.

From the Lower Burner, an **internal branch** (according to some authors, the T.B. channel) descends to the popliteal fossa to → **BL-39** (*weiyang*), the lower *he*-sea point of the Triple Burner.

A further **inner branch** originates in the thorax at → **Ren-17** (*shanzhong*), emerges at the supraclavicular fossa and ascends to the posterior border of the sternocleidomastoid muscle to **T.B.-16**, follows via **T.B.-17, T.B.-18, T.B.-19** the temporal bone along its border behind the ear and reaches **T.B.-20** (*jiasun*) above the apex of the ear. It crosses the Gall Bladder primary channel at **G.B.-6** (*xuanli*), **G.B.-5** (*xuanlu*) and **G.B.-4** (*hanyan*), according to Deadman et al 1998) also at **G.B.-11** (*touqiaoyin*) and **G.B.-14** (*yangbai*), descends the cheek to the lower jaw and ascends again to intersect with → **S.I.-18** (*quanliao*) on the lower border of the zygomatic arch and continues to the infraorbital region.

From **T.B.-17** (*yifeng*), a branch penetrates the ear, emerges to the **Exterior** anterior to the tragus, intersects with → **S.I.-19** (*tinggong*), continues past **T.B.-21** (*ermen*) and **T.B.-22** (*erheliao*), meets → **G.B.-3** (*shangguan*), crosses its own cheek branch and ends at **T.B.-23** (*sizhukong*) on the frontozygomatic suture. From **T.B.-23**, or, according to some authors from **T.B.-22**, a small branch runs to → **G.B.-1** (*tongziliao*) at the lateral canthus of the eye (Yang axes of the third great circuit: *shao yang*).

Clinical importance (→ 1.2)

Exterior (***biao***) signs and symptoms: Neck pain, cheek pain, red and painful eyes, deafness, pain behind the ears, pain on the posterior aspect of the shoulder and the upper arm

Interior (***li***) or *zangfu*-Organ signs and symptoms: Abdominal fullness and distension, difficult urination, skin oedema and skin swellings, enuresis

Connections of the Triple Burner primary channel (→ 1.2)

Connections with other channels

Pericardium primary channel (*shou jue yin jing*)

Connection: Hand Yin–Yang pairing of the third great circuit
Location: P-8 → T.B.-1 (on the hand).
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Gall Bladder primary channel (*zu shao yang jing*)

Connection: Paired according to the six-channel theory (hand–foot pairing): *shao yang* (Yang axes of the third great circuit)
Location: T.B.-23 (or **T.B.-22**) → **G.B.-1** (on the head).
Circulation: Circadian (according to the Organ clock)
Importance: Above–below relationship

Ch04.10-F10028.qxd 2/23/08 2:26 PM Page 364

Connections with *zangfu*-Organ systems
Pericardium (*xin bao*), Triple Burner (*san jiao*)

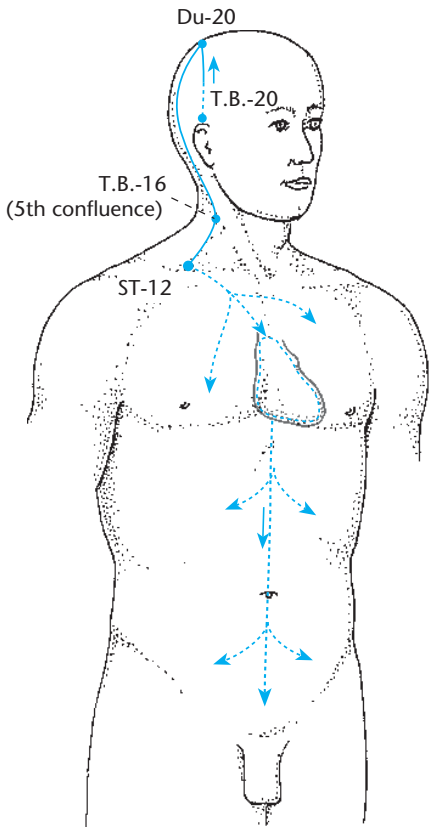

4.10.2 The Triple Burner Divergent Channel (*shou shao yang jing bie*)

Pathway

The Triple Burner divergent channel separates from the Triple Burner primary channel near **T.B.-20** (*jioasun*)

- ➡ ascends to → **Du-20** (*baihui*)
- ➡ descends along the posterior aspect of the ear to the neck to **T.B.-16** (*tianyou*) on the posterior border of the sternocleidomastoid muscle, where it connects with the Triple Burner primary channel and the Pericardium primary channel to form one of the six *he*-confluences (here: P/T.B. as fifth confluence → 1.3)
- ➡ The Triple Burner divergent channel further descends to → **ST-12** (*quepen*) in the supraclavicular fossa, penetrates the thorax, disperses in the Pericardium (*xin bao*) and reaches the Upper, Middle and Lower Burners.

Clinical importance

- Strengthens the relationship between the Upper, Middle and Lower Burners. T.B. points can be used for disorders affecting these areas and Organs.
- The Triple Burner divergent channel reaches the vertex at → **Du-20**. Therefore, T.B. points can be used for headaches, dizziness, etc, but also to raise the Yang Qi for prolapse, etc.
- The Triple Burner divergent channel spreads in the thorax, so that T.B. points can be used to support the treatment of thoracic problems such as pain, coughing etc.

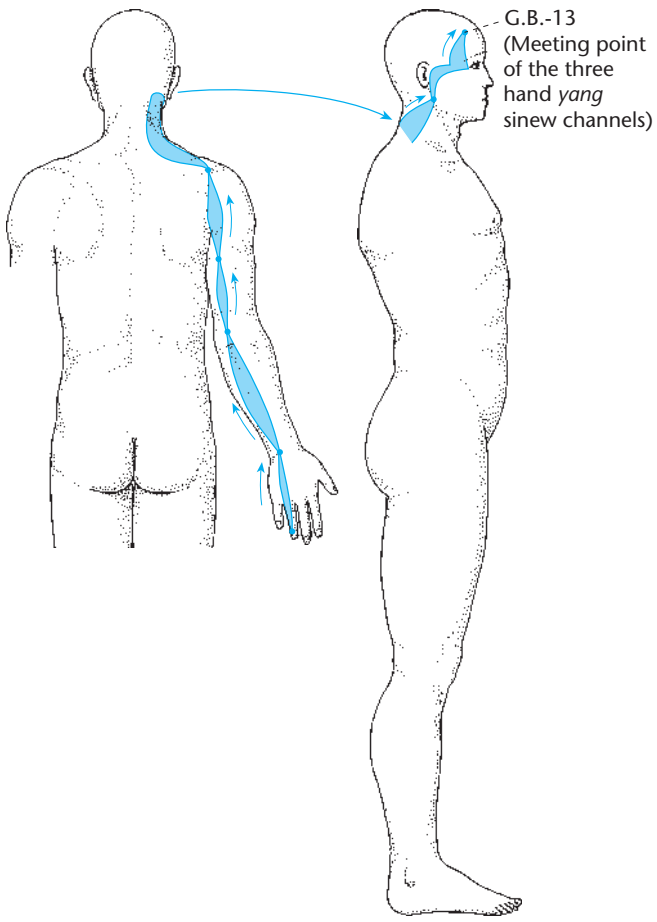

4.10.3 The Triple Burner Sinew Channel (*shou shao yang jing jin*)

Pathway

The Triple Burner sinew channel begins on the ring finger, at the ulnar corner of the nail at **T.B.-1** (*guanchong*)

- ➡ travels between the 4th and 5th metacarpal bones to the wrist
- ➡ binds (*jie*) at **T.B.-4** (*yangchi*)
- ➡ continues along the posterior aspect of the forearm between the Small Intestine and Large Intestine sinew channels, binds (*jie*) at the olecranon, ascends the posterior aspect of the upper arm to bind (*jie*) at the posterior corner of the acromion, traverses the posterior aspect of the shoulder and the lateral aspect of the occiput and there meets the Small Intestine sinew channel. It then reaches the mandibular angle, where it divides into **two branches**:
- ➡ **One branch** enters internally and penetrates the root of the tongue

Ch04.10-F10028.qxd 2/23/08 2:26 PM Page 365

- **The other branch** ascends in front of the ear, binds (*jie*) at the outer canthus of the eye, crosses the temple and terminates at → **G.B.-13** (*benshen*) in the frontoparietal region. There, it also meets the other hand Yang sinew channels.

Clinical importance

Pathology: Stiffness, swellings and distending sensations along the course of the sinew channel. Curling or contraction of the tongue.

Indication: Predominantly for pain, stiffness, spasms, tension and distending sensations along the course of the channel, including the lateral aspects of the head and the body as well as the lateral sections of the extremities. T.B. points can also be used for muscle tension caused by emotional disorders. Owing to the connection of the sinew channel with the tongue, some T.B. points are also indicated for motor disorders of the tongue.

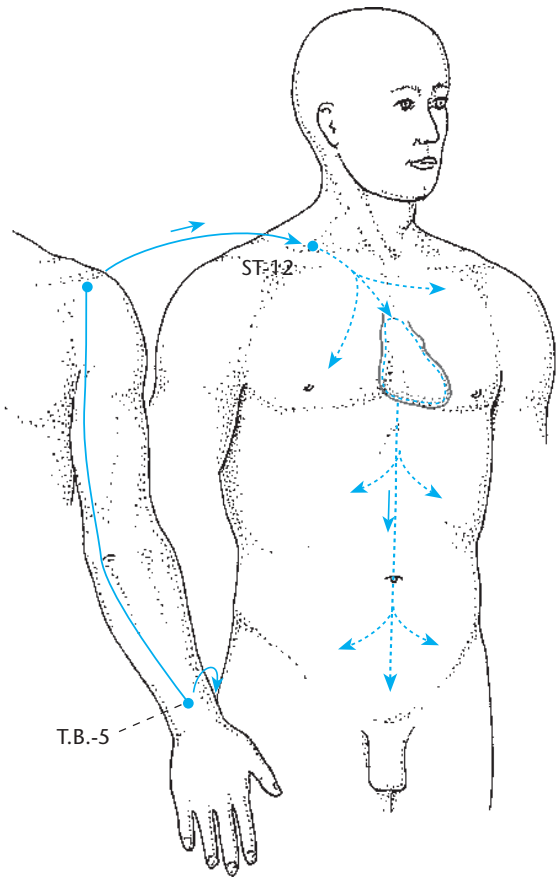

4.10.4 The Triple Burner *luo*-Connecting Vessel System (*shou shao yang luo mai*)

Pathway

The Triple Burner *luo*-connecting vessel system separates from the Triple Burner primary channel at its associated *luo*-connecting point **T.B.-5** (*waiguan*) (→ 8.1.2). It forms a three-dimensional reticulate network, dividing into multiple branches and sub-branches (*sun luo, fu luo, xue luo* → 1.5) within the surrounding tissue.

- Horizontal divisions run to the Interiorly–Exteriorly paired Pericardium primary channel; according to some schools of thought (for example Ngyen Van Nghi → Appendix), they travel as a **transverse** Triple Burner *luo*-connecting vessel to the *yuan*-source point **P-7** (*daling*).
- **A longitudinal division** follows the posterior aspect of the arm, crosses the posterior shoulder region, continues to the front of the body and, according to some authors (for example, Solinas et al 1998), reaches → **ST-12** (*quepen*) in the supraclavicular fossa, penetrates the thorax and reaches the Pericardium primary channel, both dispersing in the Upper, Middle and Lower Burners.

Clinical importance

Pathology

Excess (*shi*): Tension of the elbow joint

Deficiency (*xu*): Muscle weakness in the arm, difficulty flexing the elbow

4.10.5 Cutaneous Region (*shao yang pi bu*)

See description and figures → 1.6.

4.10.6 Points of the Triple Burner Primary Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): T.B.-4** (*yangchi*)
- ***Luo*-connecting point (→ 8.1.2.): T.B.-5** (*waiguan*) ■■
- ***Xi*-cleft point (→ 8.1.3): T.B.-7** (*huizong*)
- **Associated Back-*shu* point (→ 8.1.4): BL-22** (*sanjiaoshu*)
- **Associated Front-*mu* point (→ 8.1.5): Ren-5** (*shimen*)
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Metal): **T.B.-1** (*guanchong*)
  ying-spring point (Water): **T.B.-2** (*yemen*)
  shu-stream point (Wood), tonification point: **T.B.-3** (*zhongzhu*)
  jing-river point (Fire), *ben point* (Five Phases): **T.B.-6** (*zhigou*)
  he-sea point (Earth), sedation point: **T.B.-10** (*tianjing*)
- ***Hui*-meeting point (→ 8.1.7): –**
- **Opening point (→ 8.1.8) of the *yin wei mai*: T.B.-5** (*waiguan*) ■■
- **Lower *xiahe-s*ea points (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the *yang wei mai* and the G.B. channel: **T.B.-13*** (*naohui*)
  – with the *yang wei mai* and the G.B. channel: **T.B.-15** (*tianliao*) ■■
  – with the G.B. channel: **T.B.-17** (*yifeng*) ■■
  – with the S.I.* and G.B. channel: **T.B.-20** (*jiaosun*)
  – with the S.I. and G.B. channels: **T.B.-22** (*erheliao*)

* Mentioned by only some authors

Ch04.10-F10028.qxd 2/23/08 2:26 PM Page 366

– of other channels with the T.B. channel: **S.I.-12, BL-11, (S.I.-18, S.I.-19*), Du-14, Ren-12, Ren-17, ST-12, G.B.-21, G.B.-14, G.B.-11, G.B.-6, G.B.-5, G.B.-4, G.B.-3, G.B.-1, (S.I.-9*)**

- **Gao Wu command point (→ 8.1.11): –**
- **Window of Heaven point (→ 8.1.12): T.B.-16** (*tianyou*)
- **Points of the Four Seas (→ 8.1.13): –**
- **Ma Dan Yang Heavenly Star point (→ 8.1.14): –**
- **Sun Si Miao Ghost point (→ 8.1.15): –**
- **Other functional points: –**

Points according to region

- **Local points (→ 8.2.1):** ears – **T.B.-21** (*ermen*) ■■; shoulder – **T.B.-14** (*jianliao*) ■■, **T.B.-15** (*tianliao*) ■■
- **Adjacent points (→ 8.2.1):** temporal aspect of the head – **T.B.-17** (*yifeng*) ■■; eyes – **T.B.-23** (*sizhukong*) ■■; ears – **T.B.-17** (*yifeng*) ■■; jaw – **T.B.-17** (*yifeng*) ■■; shoulder – **T.B.-13** (*naohui*), **T.B.-15** (*tianliao*) ■■
- **Distal points (→ 8.2.1):** temporal aspect of the head – **T.B.-3** (*zhongzhu*) ■■, **T.B.-5** (*waiguan*) ■■; ears – especially **T.B.-3** (*zhongzhu*), **T.B.-5** (*waiguan*) ■■; hypochondrium – **T.B.-5** (*waiguan*) ■■, **T.B.-6** (*zhigou*); cervical spine – **T.B.-8** (*sanyangluo*); cervical spine and shoulder – **T.B.-5** (*waiguan*) ■■; shoulder – **T.B.-1** (*guanchong*)

Specific points according to the channel pathway (in numeric order)

- **T.B.-1** (*guanchong*): *jing*-well point (Metal) (→ 8.1.6); distal point for the shoulder (→ 8.2.1)
- **T.B.-2** (*yemen*): *ying*-spring point (Water) (→ 8.1.6)
- **T.B.-3** (*zhongzhu*): *shu*-stream point (Wood) (→ 8.1.6); tonification point; distal point for the temporal region and the ears (→ 8.2.1); local point for the fingers (numbness and pain) (→ 8.2.1)
- **T.B.-4** (*yangchi*): *yuan*-source point (→ 8.1.1); local point for the hand (→ 8.2.1)
- **T.B.-5** (*waiguan*) ■■: *luo*-connecting point (→ 8.1.2); Opening point (→ 8.1.8) of the *yang wei mai*; distal point for the temporal region, the ears, the cervical spine, the shoulder, the elbow and the hypochondrium (→ 8.2.1); regional point for the hand and fingers (→ 8.2.1)
- **T.B.-6** (*zhigou*): *jing*-river point (Fire) (→ 8.1.6); *ben* point (Five Phases); distal point for the hypochondrium (→ 8.2.1)
- **T.B.-7** (*huizong*): *xi*-cleft point (→ 8.1.3)
- **T.B.-8** (*sanyangliao*): distal point for the cervical spine (→ 8.2.1)
- **T.B.-10** (*tianjing*): *he*-sea point (Earth) (→ 8.1.6); sedation point; local point for the elbow (→ 8.2.1)
- **T.B.-13** (*naohui*): *jiaohui*-meeting point (→ 8.1.10) of the *yang wei mai** and the G.B. channel*; adjacent point for the shoulder (→ 8.2.1)
- **T.B.-14** (*jianliao*) ■■: local point for the shoulder (→ 8.2.1)
- **T.B.-15** (*tianliao*) ■■: *jiaohui*-meeting point of the *yang wei mai* and the G.B. channel; local point for the shoulder (→ 8.2.1)
- **T.B.-16** (*tianyou*): Window of Heaven point (→ 8.1.12)
- **T.B.-17** (*yifeng*) ■■: *jiaohui*-meeting point of the G.B. channel; adjacent point for the temporal region, the ears and the jaw (→ 8.2.1)
- **T.B.-20** (*jiaosun*): *jiaohui*-meeting point of the S.I.* and G.B. channels
- **T.B.-21** (*ermen*) ■■: local point for the ears (→ 8.2.1)
- **T.B.-22** (*erheliao*): *jiaohui*-meeting point with the S.I. and G.B. channels
- **T.B.-23** (*sizhukong*) ■■: adjacent point for the eyes (→ 8.2.1)

*Mentioned only by some authors

Ch04.10-F10028.qxd 2/23/08 2:26 PM Page 367

Rushing Pass GUANCHONG T.B.-1

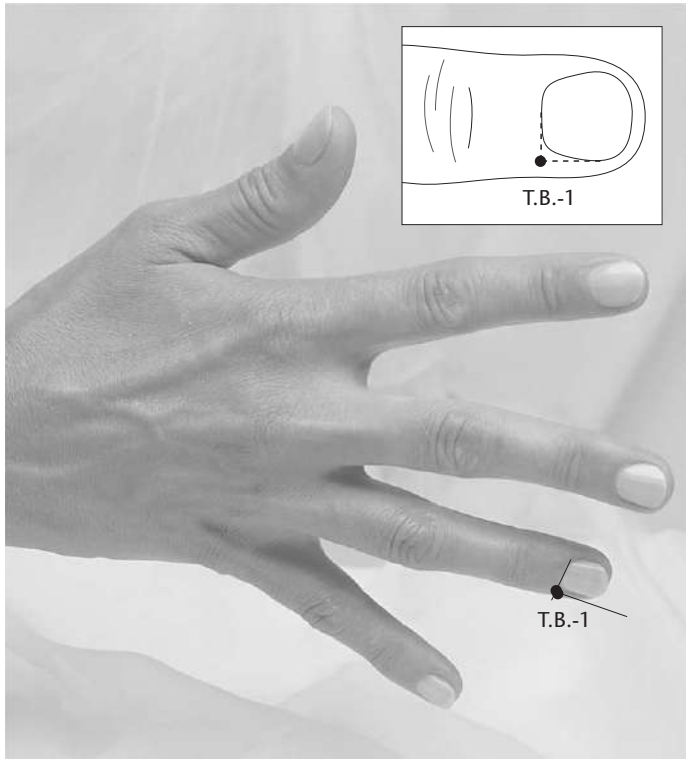

Location

On the ring finger, 0.1 cun from the ulnar corner of the nail.

How to find

This point is located at the junction of two tangents bordering the ulnar and proximal margins of the nail, approximately 0.1 cun from the corner of the actual nail.

Needling

Approximately 0.2 cun vertically or obliquely in a proximal direction or prick to bleed. Caution: Avoid needling too close to the nail; painful point.

Actions/Indications

- Clears Heat from the Upper Burner
- Benefits the ears and tongue
- Opens the channel and alleviates pain

Special features

Jing-well point, Metal point, entry point. Especially indicated for acute disorders.

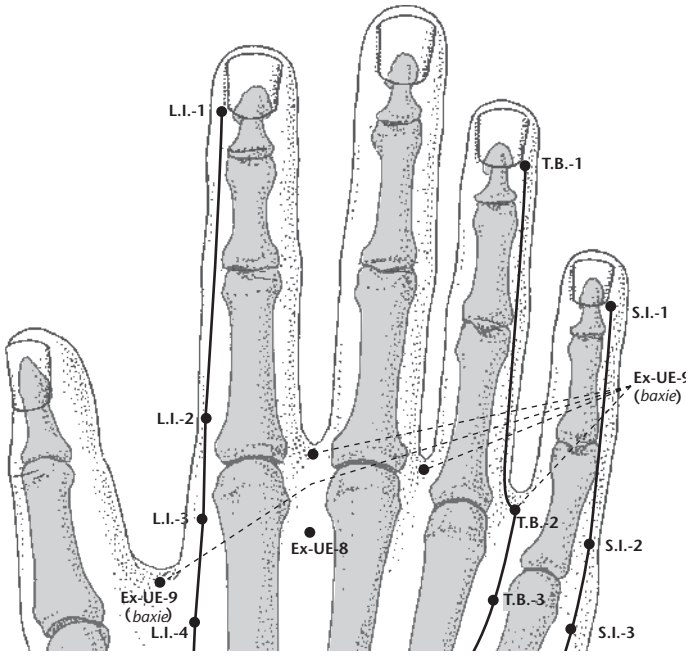

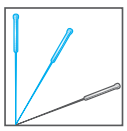
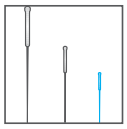

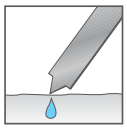

Ch04.10-F10028.qxd 2/23/08 2:26 PM Page 368

T.B.-2 *Fluid Gate* YEMEN

Location
Between the little finger and ring finger, proximal to the margin of the web.

How to find
This point is best located when making a loose fist. Locate the web between the little finger and ring finger (4th and 5th fingers) and locate **T.B.-2** proximal to its margin. **T.B.-2** is part of → **Ex-UE-9** (*baxie*: proximal to the webs between the fingers). Located in a comparable position on the foot is → **G.B.-43** (on the web between the 4th and 5th toes), which is also part of → **Ex-LE-10** (*bafeng*).

Needling
Vertically up to 0.5 cun

Actions/Indications
- Clears Heat from the Upper Burner
- Benefits the ears and calms the *shen*
- Opens the channel and alleviates pain

Special features
Ying-spring point, Water point. Especially indicated for psychoneurological disorders caused by Heat and local channel problems.

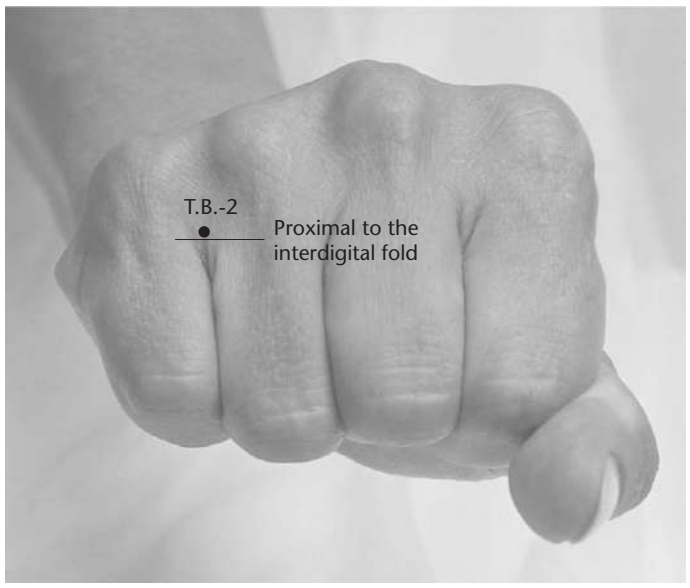

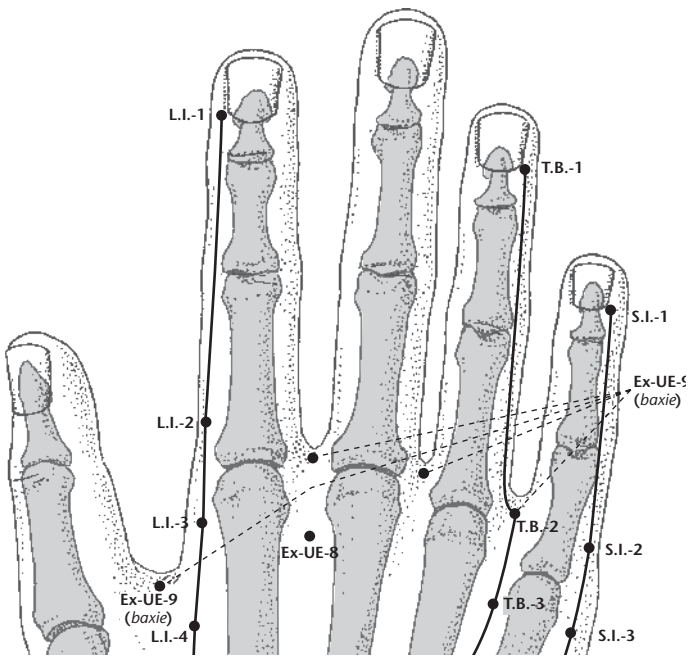

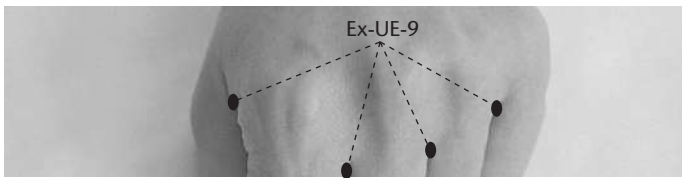

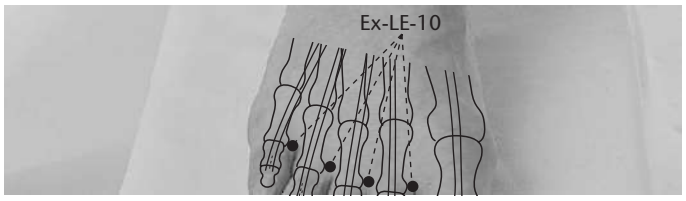

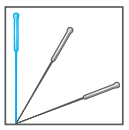

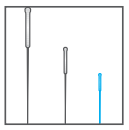

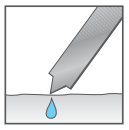

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 369

Central Islet ZHONGZHU T.B.-3

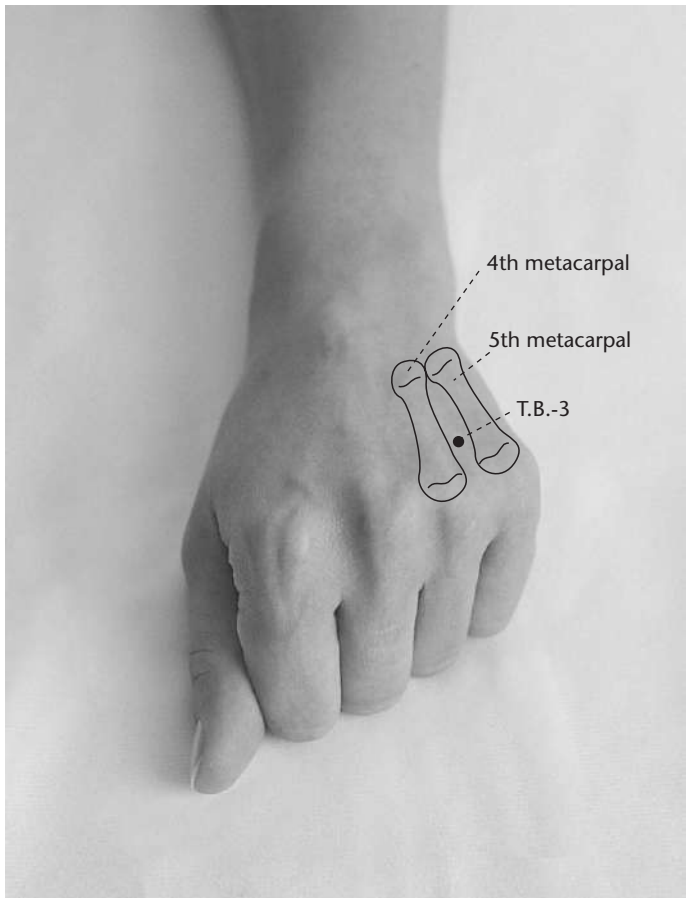

Location

On the dorsum of the hand, in a depression between the 4th and 5th metacarpal bones, proximal to the metacarpophalangeal joints. **T.B.-3** is located at the junction between the heads and shafts of the two metacarpal bones.

How to find

The hand should be relaxed or a loose fist should be made. Palpate from the metacarpophalangeal joints along the groove between the 4th and 5th metacarpal bones towards proximal. **T.B.-3** is located at the widest/deepest point of the groove, slightly distal to the carpometacarpal joints.
Located in comparable positions are → **Ex-UE-8** (*wailaogong*: between the 2nd and 3rd metacarpal bones), → **S.I.-3** (on the ulnar aspect of the 5th metacarpal bone) and → **L.I.-3** (on the radial aspect of the 2nd metacarpal bone).

Needling

0.5–1 cun vertically or obliquely in a proximal direction.

Actions/Indications

- Benefits the ears, clears Heat (and drains Fire), clears the head and the eyes
- Opens the channel and *luo*-connecting vessels, alleviates pain

Special features

Shu-stream point, Wood point, tonification point. One of the most important distal points for any disorder of the ears.

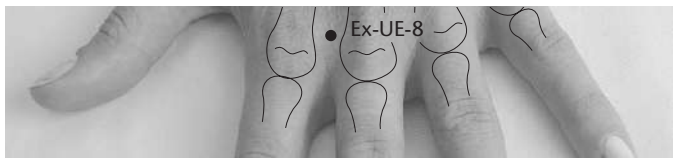

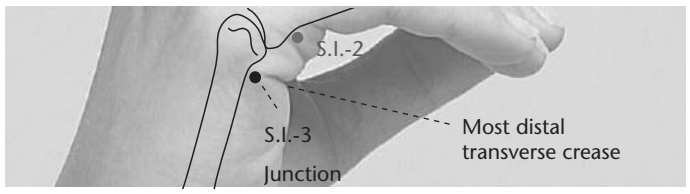

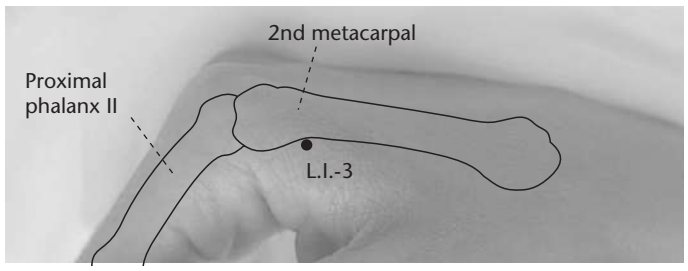

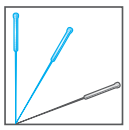
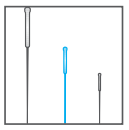

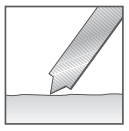

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 370

T.B.-4 *Yang Pool* YANGCHI

Location

On the dorsum of the wrist (above the wrist joint space, 'dorsal wrist crease'), in the gap between the tendons of the extensor digitorum (on the ulnar aspect) and the extensor digiti minimi muscles (on the radial aspect).

How to find

As the location of the dorsal wrist crease varies, the wrist joint space should be used for orientation instead (→ 3.3.3). By moving the patient's wrist in a relaxed way, the joint space becomes well defined. **T.B.-4** is located slightly laterally to the midpoint of the joint space, in the depression between the tendons of the extensor digiti minimi muscle (continuing to the little finger) on its radial aspect and the tendon of the extensor digitorum muscle on its ulnar aspect. The tendons of the extensor digitorum muscle will become more pronounced by moving the three middle fingers.

Located on the same level but radial to the tendons of the extensor digitorum muscle is → **Ex-UE-3** (*zhongquan*).

Needling

0.3–0.5 cun vertically or slightly obliquely in a proximal direction or transversely (subcutaneously) deep to the tendons towards the radial aspect of the wrist.

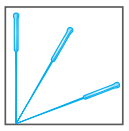

Actions/Indications

- Clears Heat, opens the channel, relaxes the tendons. In Japanese acupuncture tradition, **T.B.-4**, in its function as *yuan*-source point, is used for deficiency syndromes.

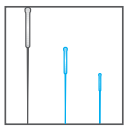

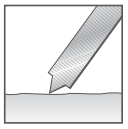

Special features

Yuan-source point

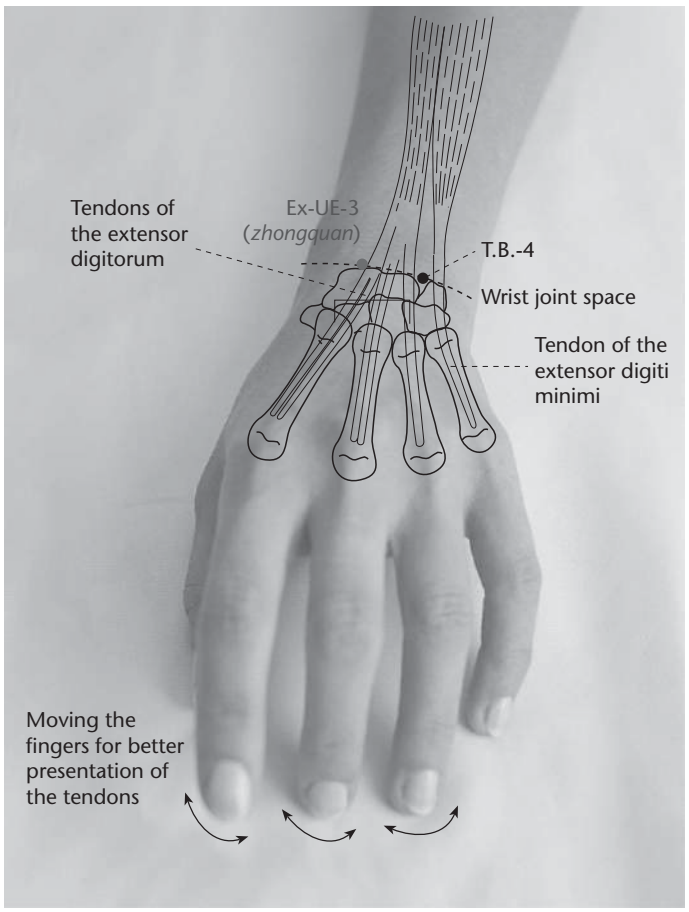

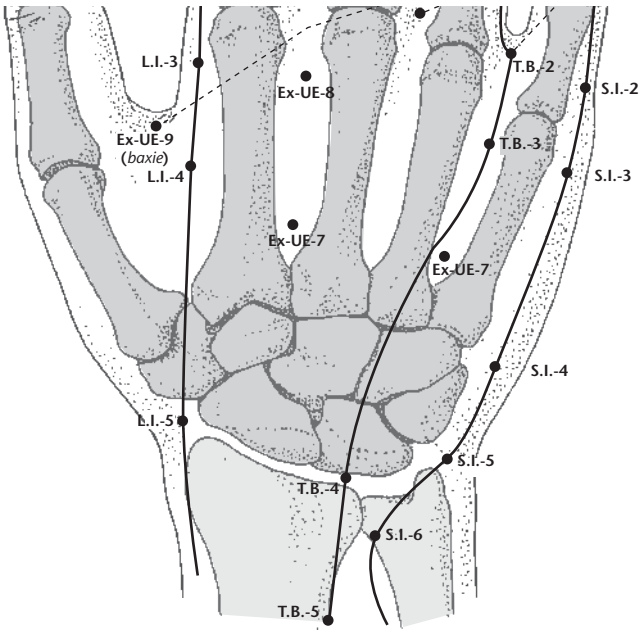

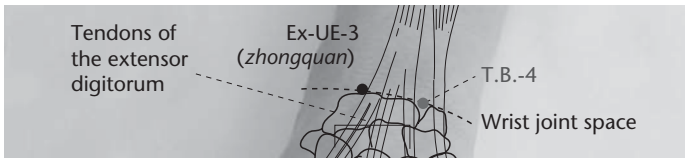

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 371

Outer Pass WAIGUAN T.B.-5

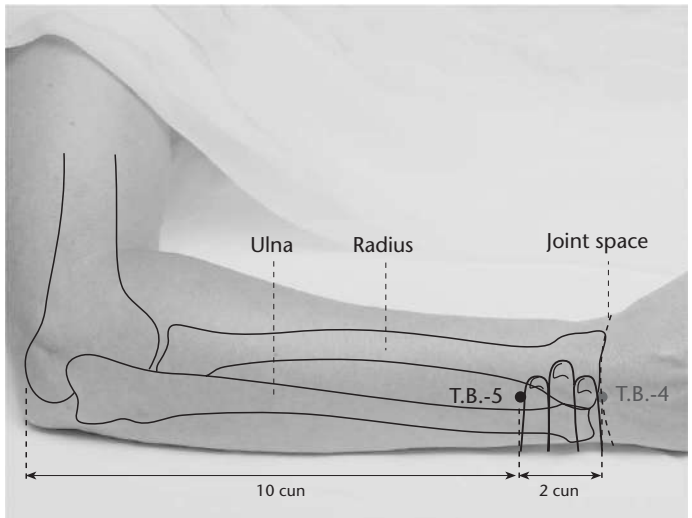

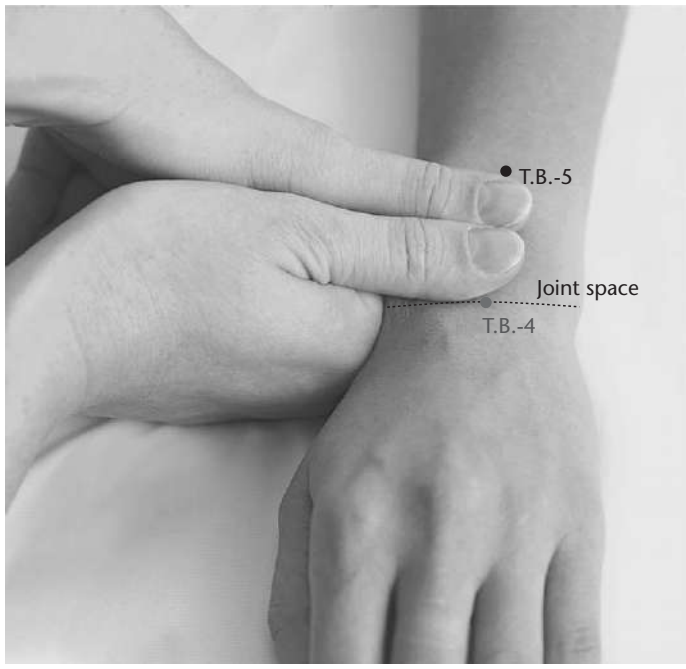

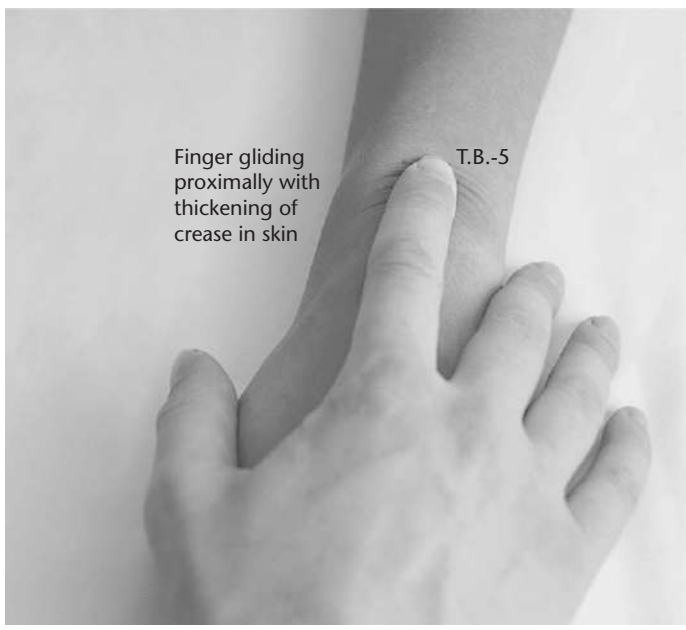

Location

2 cun proximal to the dorsal wrist joint space ('dorsal wrist crease'), between the radius and the ulna.

How to find

As the location of the dorsal wrist crease varies, the wrist joint space should be used for orientation instead (→ 3.3.3). By moving the patient's wrist in a relaxed way, the joint space becomes more easily palpable. **T.B.-5** is located 2 cun proximal to the midpoint of the joint space, in the depression midway between the radius and the ulna. **Or quick method:** With the palpating finger, slide with slight pressure from the dorsal aspect of the wrist joint in a proximal direction between the radius and the ulna, until the finger is brought to rest at **T.B.-5** by the increasing skin bulge building up in front of it.

→ **P-6** is located opposite **T.B.-5,** on the anterior aspect of the forearm.

Needling

Vertically or obliquely 0.5–1.5 cun. Caution: Movement of the hand/arm during needle retention may cause bending of the needle.

Actions/Indications

- Expels Wind and benefits the head and ears, clears Heat, opens and regulates the *yang wei mai*
- Opens the channel and alleviates pain

Special features

Luo-connecting point, opening (master) point of the *yang wei mai*. Major point for expelling Wind-Heat, analgesic point for the upper extremity.

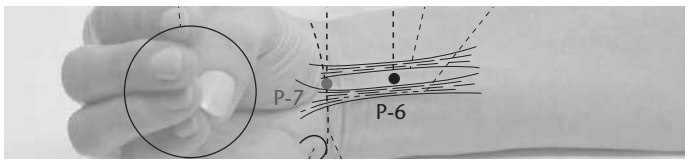

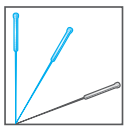
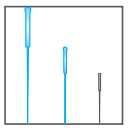

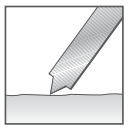

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 372

T.B.-6 *Branching Ditch* ZHIGOU

Location

3 cun proximal to the dorsal wrist joint space ('dorsal wrist crease'), in a depression between the radius and the ulna, radial to the tendon of the extensor digitorum communis muscle.

How to find

As the location of the dorsal wrist crease varies, the wrist joint space should be used for orientation instead (→ 3.3.3). By moving the patient's wrist in a relaxed way, the joint space becomes more easily palpable. From the joint space, measure 3 cun in a proximal direction. At this point, the extensor digitorum communis muscle will often lie midway between the ulna and the radius. **T.B.-6** is located in a depression close to the border of the radius, on the radial side of the muscle.

Located on the same level is → **T.B.-7,** in a depression between the ulna and the extensor digitorum communis muscle. → **P-5** is also located on this level, but on the anterior aspect of the forearm.

Needling

0.5–1.5 cun slightly obliquely towards the ulna or obliquely in a proximal (elbow) or distal (wrist) direction. Or through-needling to **P-5** (*jianshi*). Caution: Movement of the hand/arm during needle retention may cause bending of the needle.

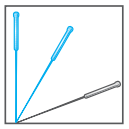

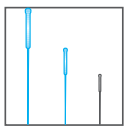

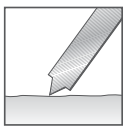

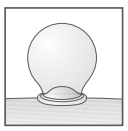

Actions/Indications

- Regulates the Qi, clears Heat in the Triple Burner, benefits the lateral costal region, promotes bowel movements
- Benefits the voice
- Opens the channel

Special features

Jing-river point, Fire point, *ben* point (Five Phases).

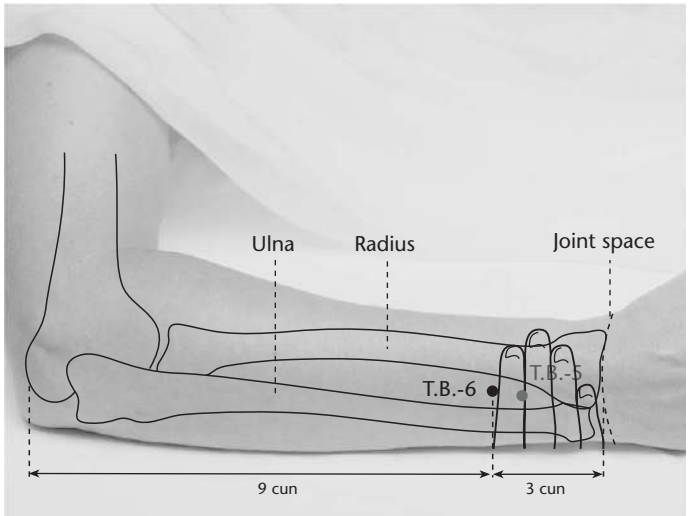

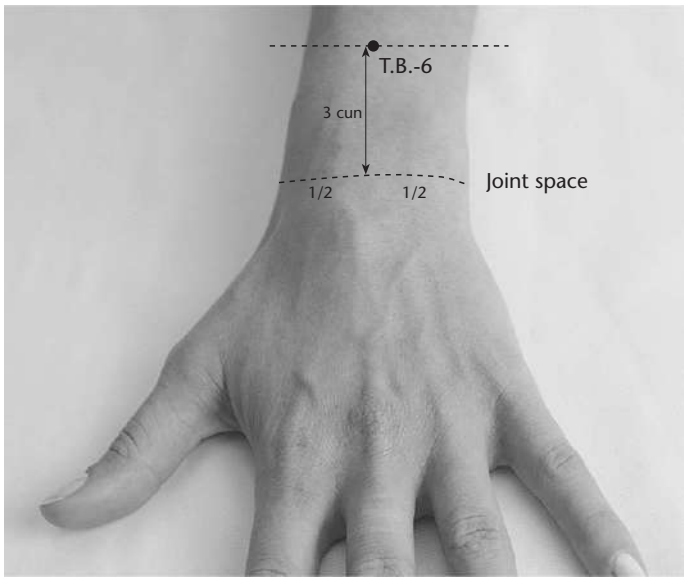

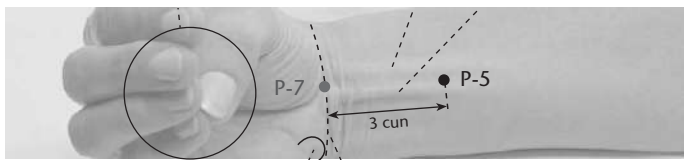

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 373

Ancestral Meeting HUIZONG T.B.-7

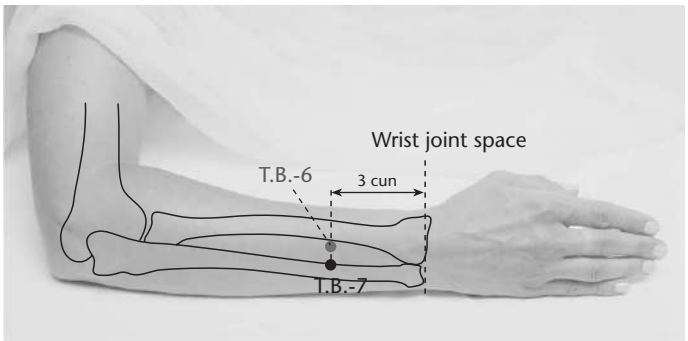

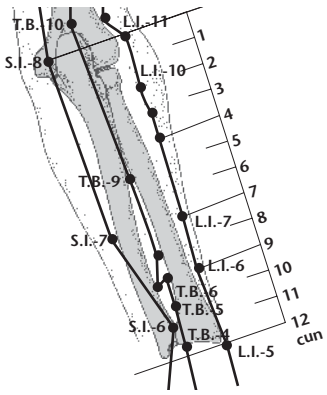

Location

3 cun proximal to the dorsal wrist joint space ('dorsal wrist crease') and 0.5 cun ulnar to the centre of the forearm.

How to find

As the location of the dorsal wrist crease varies, the wrist joint space should be used for orientation instead (→ 3.3.3). By moving the patient's wrist in a relaxed way, the joint space becomes more easily palpable. From the joint space, measure 3 cun in a proximal direction and locate **T.B.-7** in a depression between the ulna and the extensor digitorum communis muscle. **Or:** Spreading hands technique (→ 2.3.3): Place the little fingers on the elbow crease and the wrist joint space respectively (this distance corresponds to 12 cun). Divide this distance into quarters and locate **T.B.-7** one quarter of the distance from the wrist joint and 0.5 cun ulnar to the centre of the forearm (where → **T.B.-6** is located) on the border of the ulna.

Located on the same level is → **T.B.-6** (0.5 cun towards the radius).

Needling

Vertically or obliquely up to 1.5 cun

Actions/Indications

- Opens the channel and alleviates pain
- Benefits the ears

Special features

Xi-cleft point

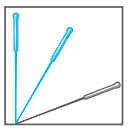

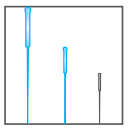

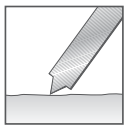

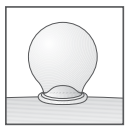

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 374

T.B.-8 *Three Yang Connection* SANYANGLUO

Location

4 cun proximal to the dorsal wrist joint space ('dorsal wrist crease'), between the radius and the ulna, radial to the tendon of the extensor digitorum communis muscle.

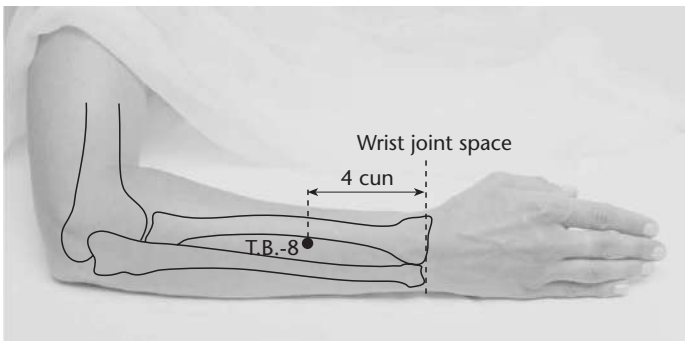

How to find

As the location of the dorsal wrist crease varies, the wrist joint space should be used for orientation instead (→ 3.3.3). By moving the patient's wrist in a relaxed way, the joint space becomes more easily palpable. From the joint space, measure 4 cun in a proximal direction and locate **T.B.-8** in a depression between the ulna and radial to the extensor digitorum communis muscle. **Or:** Spreading hands technique (→ 2.3.3): Divide the distance between the elbow crease and the wrist joint space (=12 cun) into thirds and locate **T.B.-8** one third of the distance from the wrist joint between the radius and the ulna.

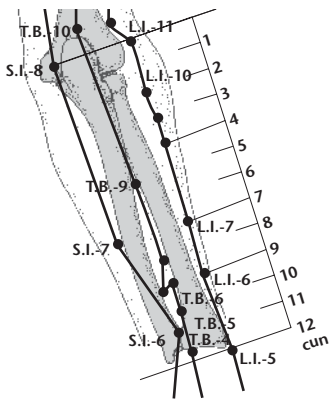

Needling

Vertically or obliquely up to 1.5 cun

Actions/Indications

- Benefits the ears and the voice
- Opens the channel and the *luo*-connecting vessels, alleviates pain

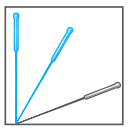

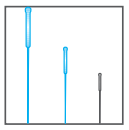

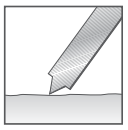

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 375

Four Rivers SIDU T.B.-9

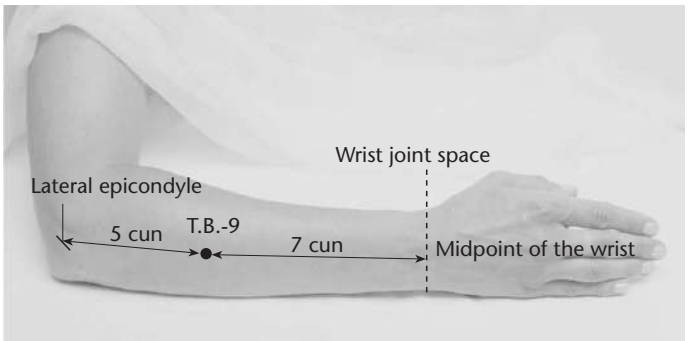

Location

7 cun proximal to the dorsal wrist joint space ('dorsal wrist crease'), between the radius and the ulna.

How to find

From the wrist joint space (→ 3.3.3, more accurate reference point than the variable dorsal wrist crease), measure 7 cun in a proximal direction and locate **T.B.-9** in a depression between the extensor digitorum communis muscle and the extensor carpi ulnaris muscle. Alternatively, measure 5 cun distally from the lateral epicondyle of the humerus. **T.B.-9** is located on a line connecting the midpoint of the dorsal wrist joint space and the lateral epicondyle of the humerus. **Or:** Spreading hands technique (→ 2.3.3): Place the little fingers on the cubital crease and the wrist joint space (→ 2.2) respectively and join the thumbs at the midpoint of this distance (=12 cun). From the midpoint of this line, measure 1 cun in a proximal direction and there locate **T.B.-9** between the radius and the ulna.

Needling

Vertically or obliquely up to 2 cun

Actions/Indications

- Opens the channel and *luo*-connecting vessels
- Benefits the ears and throat

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 376

T.B.-10 *Heavenly Well* TIANJING

Location
On the lateral aspect of the upper arm, with the elbow flexed in a depression approximately 1 cun proximal to the olecranon.

How to find
This point is best located with the patient's elbow flexed to 90°, which will reveal an easily palpable depression 1 cun superior to the olecranon. **T.B.-10** is located in this depression, on the tendon of the triceps brachii muscle.

Needling
Up to 1 cun vertically or obliquely in a proximal direction.

Actions/Indications
- Transforms Phlegm and disperses accumulations
- Regulates and descends the Qi
- Calms the *shen*
- Clears Heat from the channel
- Opens the channel and *luo*-connecting vessels, alleviates pain

Special features
He-sea point, Earth point, sedation point.

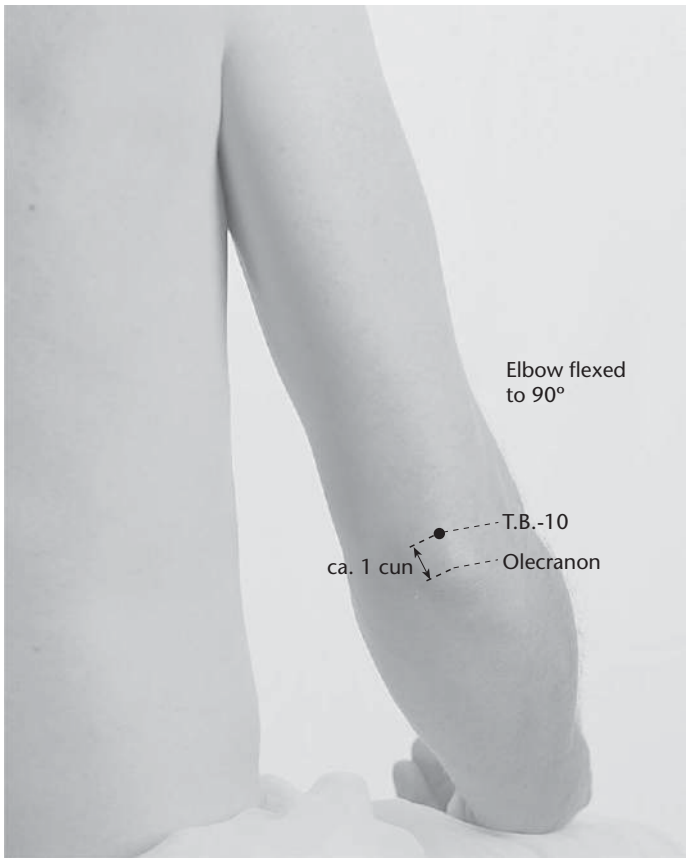

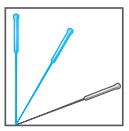

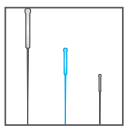

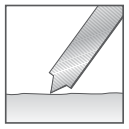

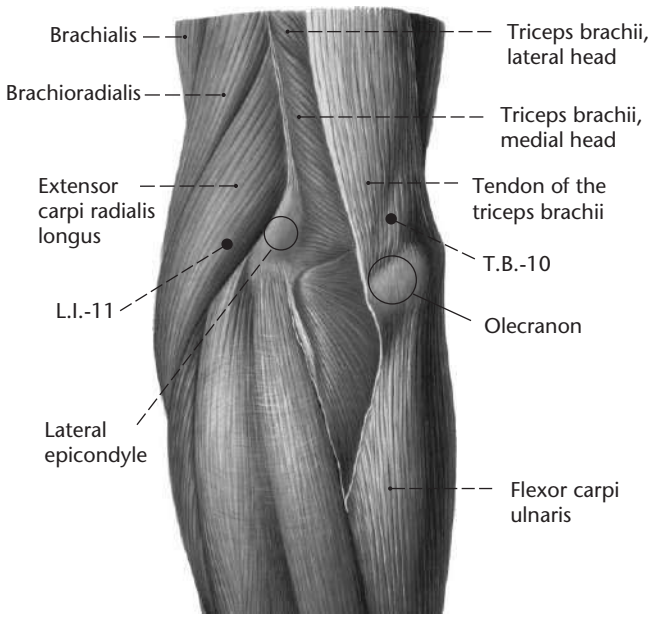

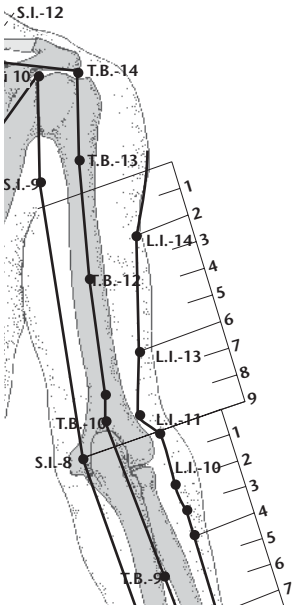

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 377

Clear Cold Abyss QINGLENGYUAN T.B.-11

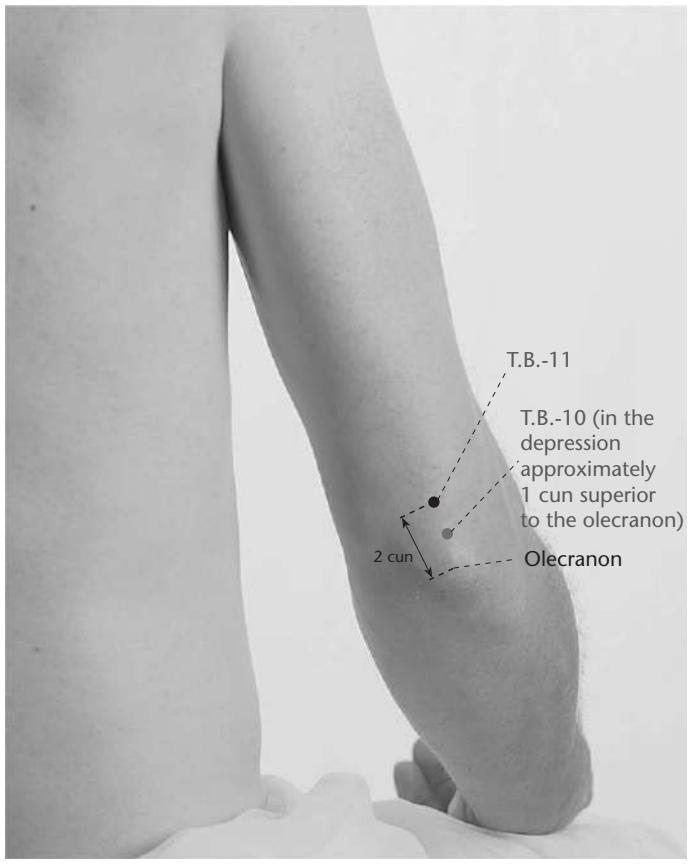

Location

On the lateral aspect of the upper arm, with the elbow flexed 1 cun proximal to → **T.B.-10** or 2 cun proximal to the olecranon, on the triceps brachii muscle.

How to find

This point is best located with the patient's elbow flexed to 90°, using → **T.B.-10** as a reference point. This is located in an easily palpable depression 1 cun superior to the olecranon. From → **T.B.-10**, palpate approximately 1 cun in a proximal direction and locate **T.B.-11** in a depression on the triceps brachii muscle.

Needling

Vertically up to 1 cun

Actions/Indications

- Dispels Wind-Damp
- Opens the channel and the *luo*-connecting vessels
- Clears Damp-Heat

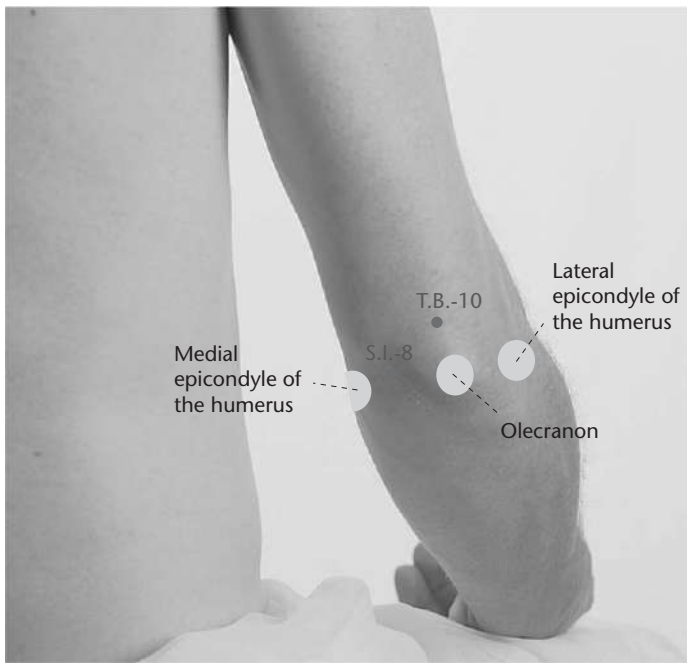

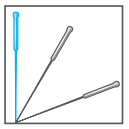
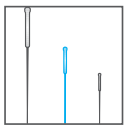

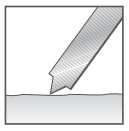

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 378

T.B.-12 *Dispersing Luo River* XIAOLUO

Location

4 cun proximal to → **T.B.-10** (with the elbow flexed in a depression superior to the olecranon) or 5 cun proximal to the olecranon on a line connecting the olecranon and the lateral extremity of the acromion (location of → **T.B.-14**).

How to find

First, locate the acromion (→ 3.3.1). When the patient's arm is in abduction, two depressions will form below the acromion, in the area of origin of the deltoid muscle. → **T.B.-14** is located in the posterior depression, below the posterior aspect of the lateral extremity of the acromion. Next, on a line connecting → **T.B.-14** and the olecranon, measure 4 cun from → **T.B.-10** (1 cun proximal to the olecranon, in the depression that forms when the elbow is flexed) in a proximal direction and there locate **T.B.-12**. **Or quick method:** Spreading hands technique (→ 2.3.3): Place the little fingers on → **T.B.-14** and → **T.B.-10** respectively and determine the midpoint of this distance with the thumbs. From the midpoint, measure 1 cun in a distal direction and there locate **T.B.-12. For orientation: T.B.-12** is midway between → **T.B.-11** (1 cun proximal to **T.B.-10** or 2 cun proximal to the olecranon) and → **T.B.-13** (3 cun distal to → **T.B.-14**, on the margin of the deltoid muscle).

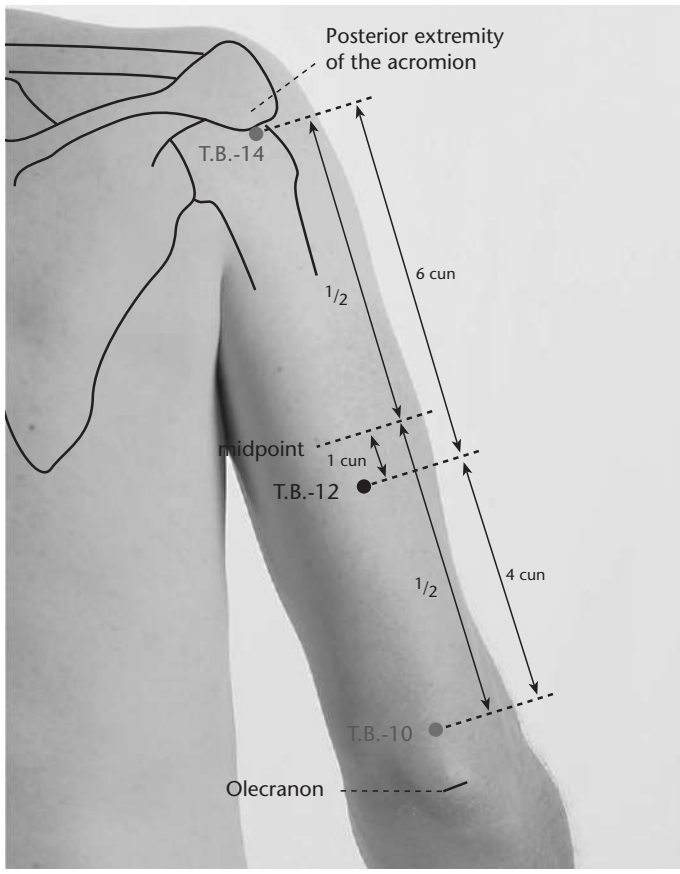

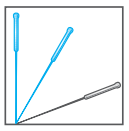

Needling

Vertically or obliquely up to 2 cun

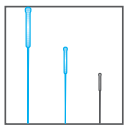

Actions/Indications

- Opens the channel and alleviates pain

Special features

T.B.-12 represents an area extending from 5–7 cun proximal to the tip of the elbow rather than a discrete acupuncture point. Determine the exact location by tenderness.

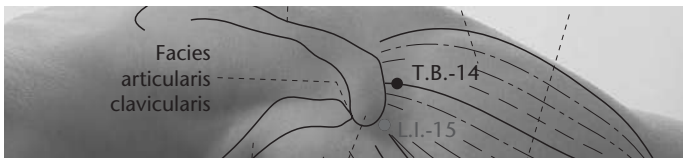

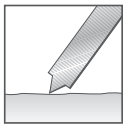

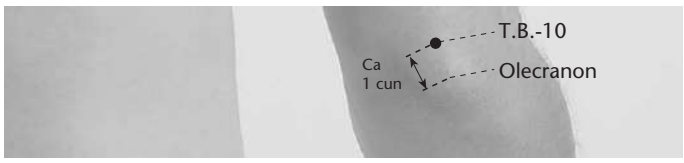

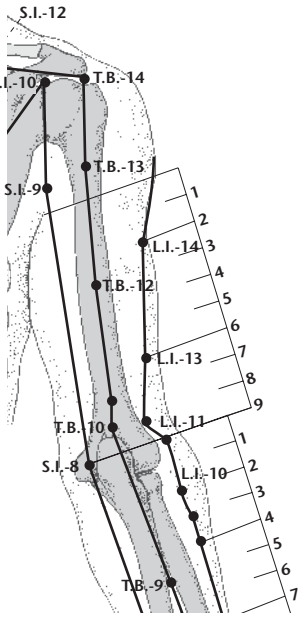

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 379

Upper Arm Meeting NAOHUI T.B.-13

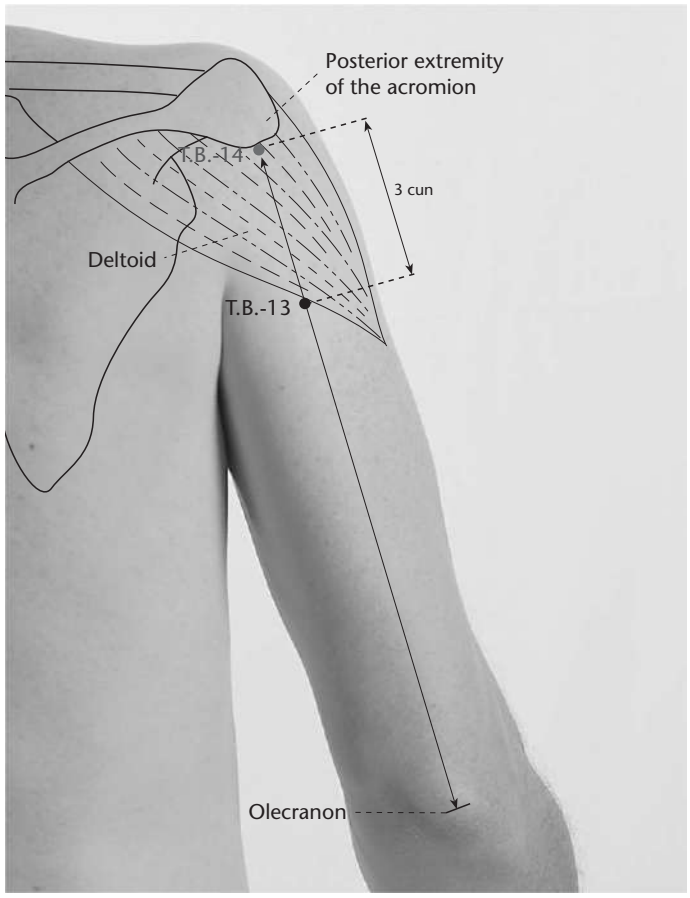

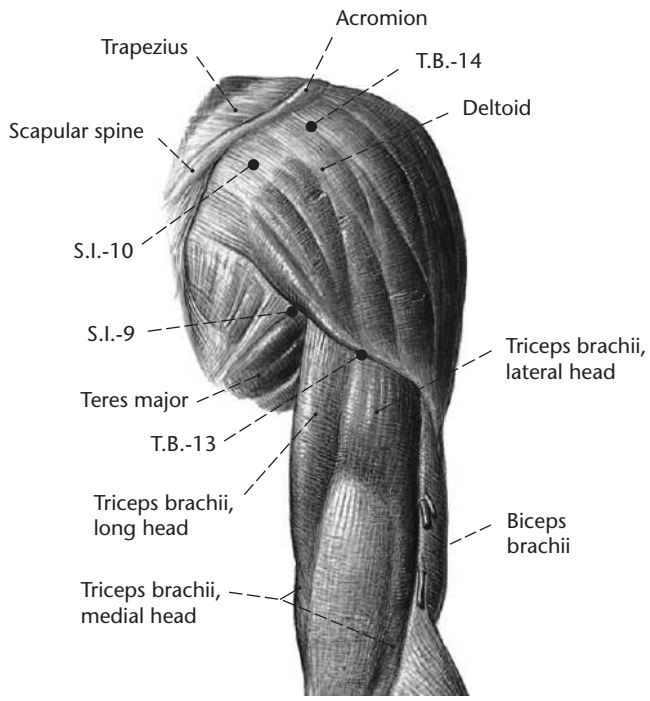

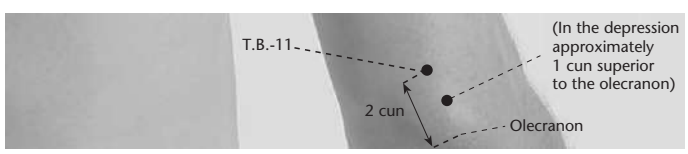

Location

3 cun distal to the lateral extremity of the acromion (location of → **T.B.-14**), on a line connecting → **T.B.-14** and the olecranon, at the junction of this line with the margin of the deltoid muscle.

How to find

First, locate the acromion (→ 3.3.1). When the patient's arm is in abduction, two depressions will form inferior to the acromion, in the area of origin of the deltoid muscle. → **T.B.-14** is located in the posterior depression, below the dorsal aspect of the lateral extremity of the acromion. Next, locate **T.B.-13** 3 cun distal to → **T.B.-14**, on a line connecting → **T.B.-14** and the olecranon. **T.B.-13** is located at the junction of this line with the margin of the deltoid muscle, approximately on the level of the end of the posterior axillary fold.

Needling

Vertically or obliquely up to 2 cun

Actions/Indications

- Opens the channel and the *luo*-connecting vessels, alleviates pain
- Regulates the Qi flow and transforms Phlegm

Special features

According to some authors, meeting point with the *yang wei mai* and the G.B. channel

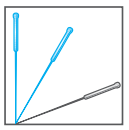
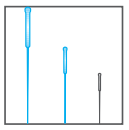

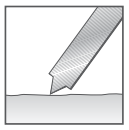
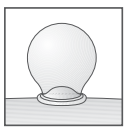

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 380

T.B.-14 *Shoulder Crevice* JIANLIAO

Location

Inferior to the lateral extremity of the acromion between the acromial and spinal portions of the deltoid muscle or, with the arm abducted, in the posterior of the two depressions on the shoulder joint.

How to find

Ask the patient to abduct their arm to a 90° angle, which will reveal two depressions anterior and posterior to the acromion. **T.B.-14** is located in the posterior depression, directly below the dorsal aspect of the lateral extremity of the acromion, between the acromial and spinal fibres of the deltoid muscle. The dorsal aspect of the lateral extremity of the acromion becomes more easily palpable when the upper arm is gently pulled (extension of the upper arm), allowing the head of the humerus to glide anteriorly.

→ **L.I.-15** is located in the depression anterior to **T.B.-14.**

Needling

With the arm adducted, 0.5–1 cun perpendicularly towards the centre of the axilla or 1–2 cun transversally to obliquely towards the elbow.

Actions/Indications

- Disperses Wind and Dampness, benefits the shoulder joint, opens the channel and the *luo*-connecting vessels, alleviates pain

Special features

Important local point for shoulder pain, especially for the dorsal aspect of the shouler

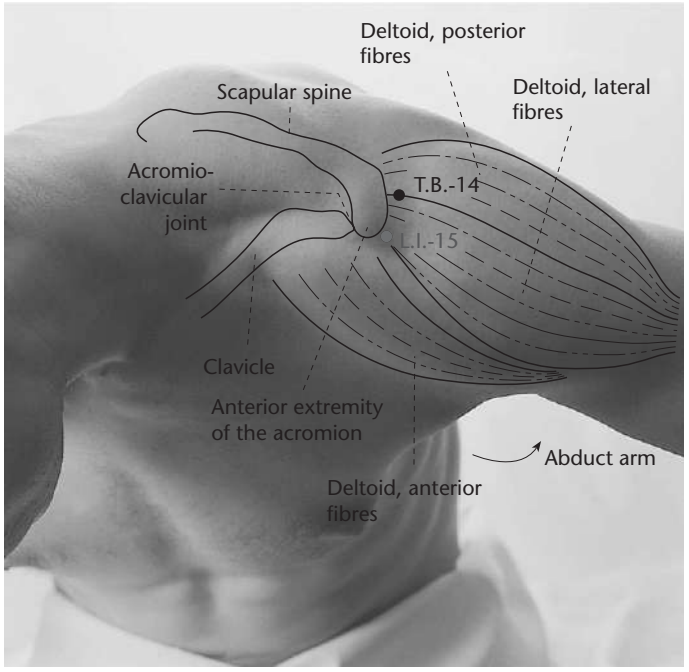

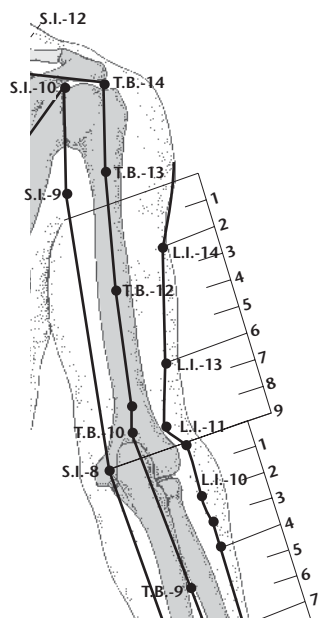

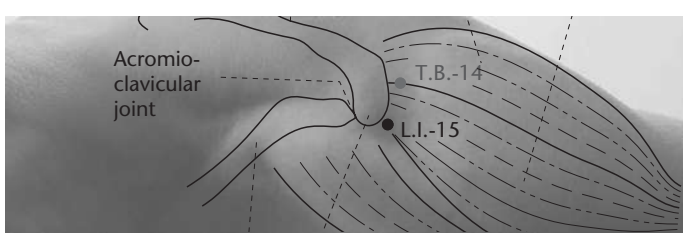

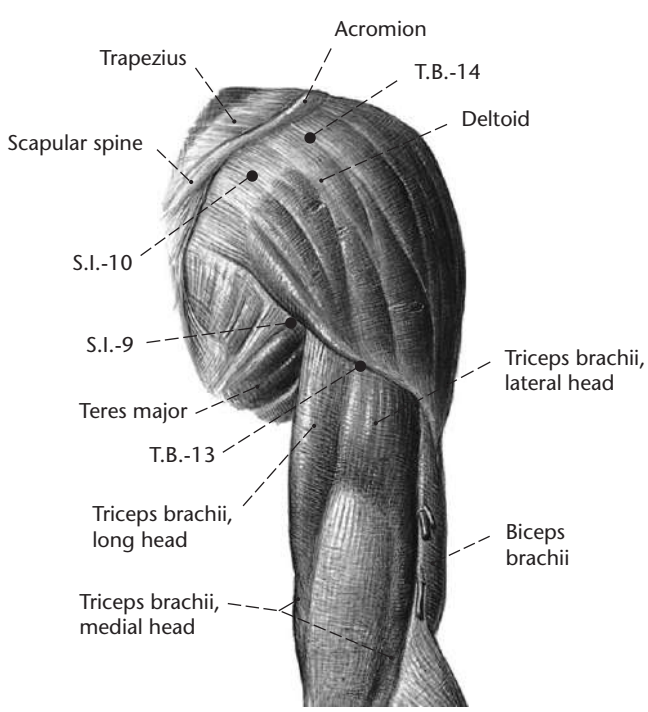

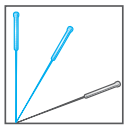
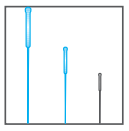

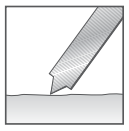

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 381

Heavenly Crevice TIANLIAO T.B.-15

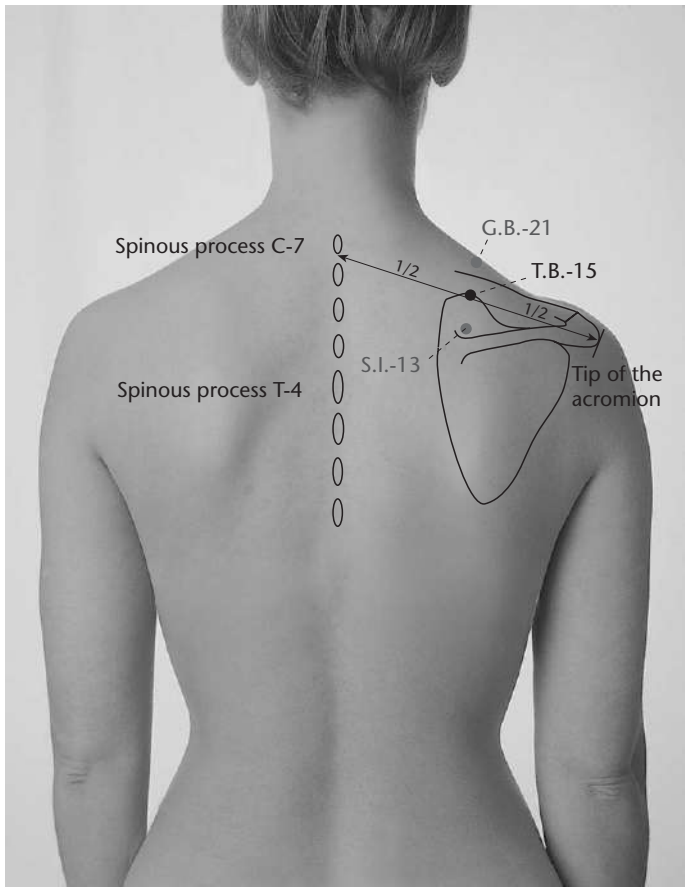

Location

At the midpoint of an imaginary line between the spinous process of the 7th cervical vertebra (C7) and the tip of the acromion (lateral extremity of the acromion).

How to find

First, locate C7 (→ 3.4.1) and the acromion (→ 3.3.1), which represents the lateral extension of the scapular spine (→ 3.3.1) and, at its most lateral end, forms a flat surface superior to the head of the humerus (acromial angle). Using, for example, the spreading hands technique (→ 2.3.3), determine the midpoint between C7 and the acromial angle. Located at the highest point of the shoulder is → **G.B.-21. T.B.-15** is located approximately 1 cun inferior to → **G.B.-21**, on the superior angle of the scapula.
→ **S.I.-13** is located inferior to **T.B.-15**, directly superior to the medial end of the scapular spine.

Needling

0.5–1 cun vertically or obliquely towards the neck or shoulder. Caution: Pneumothorax.

Actions/Indications

- Opens the channel and alleviates pain
- Dispels especially Wind-Damp, unbinds the chest

Special features

Meeting point with the *yang wei mai* and the G.B. channel

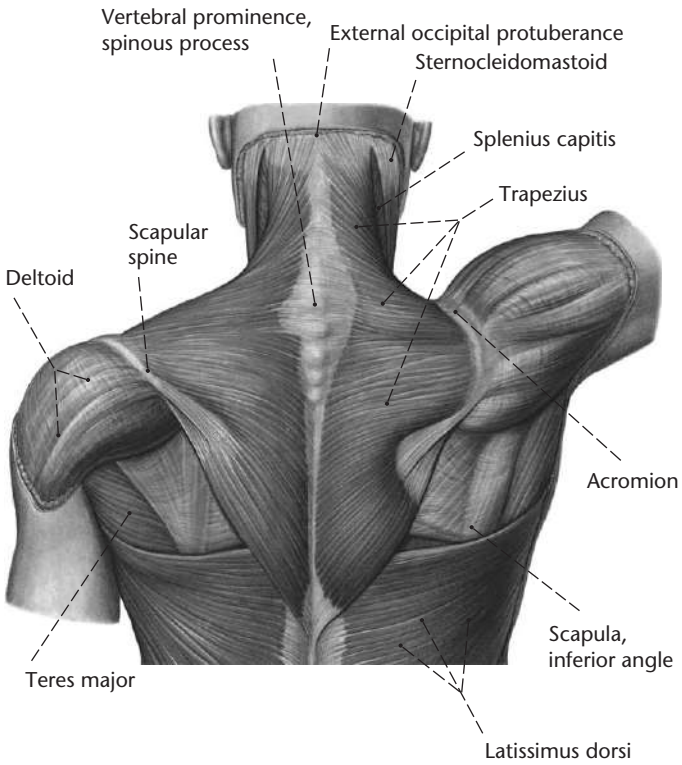

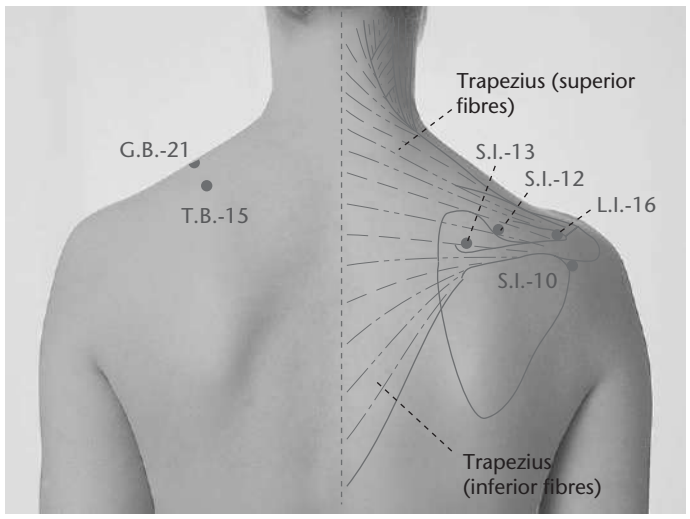

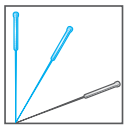
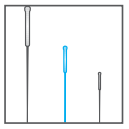
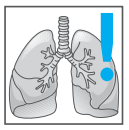

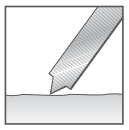

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 382

T.B.-16 *Window of Heaven* TIANYOU

Location

At the posterior border of the sternocleidomastoid muscle, directly inferior to the mastoid process, on the level of the mandibular angle.

How to find

For a better definition of the sternocleidomastoid muscle, ask the patient to rotate their head against resistance towards the side to be needled. **T.B.-16** is located on the level of the mandibular angle, on the dorsal aspect of the sternocleidomastoid muscle. This point is located inferior to the posterior border of the mastoid (→ 3.1.4). → **G.B.-12** is located directly posterior and inferior to the mastoid; → **S.I.-17** is also located on the level of the mandibular angle, but on the anterior border of the sternocleidomastoid muscle.

Needling

Vertically 0.5–1 cun. Caution: Carotid artery.

Actions/Indications

- Benefits the head and sensory organs
- Descends the Qi, expels Wind
- Opens the channel and alleviates pain

Special features

Window of Heaven point

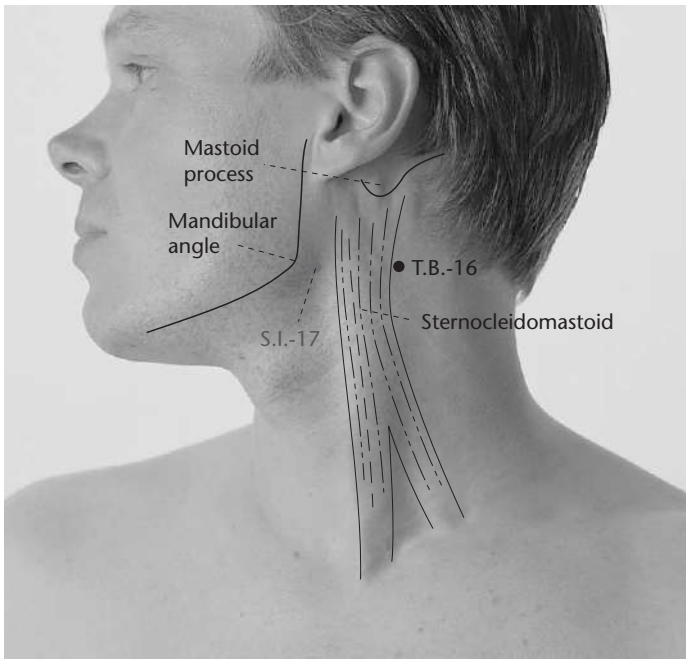

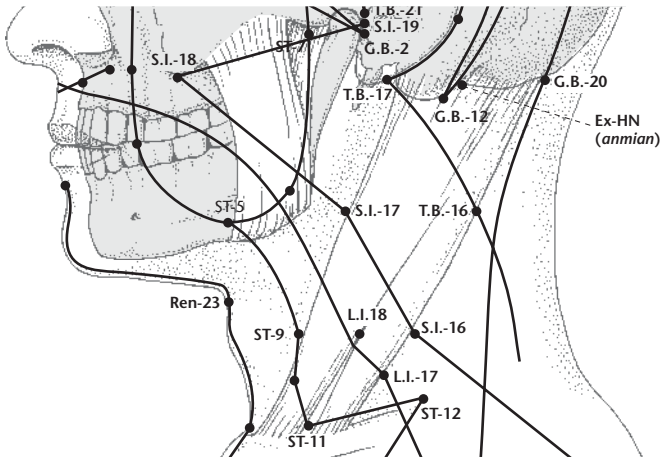

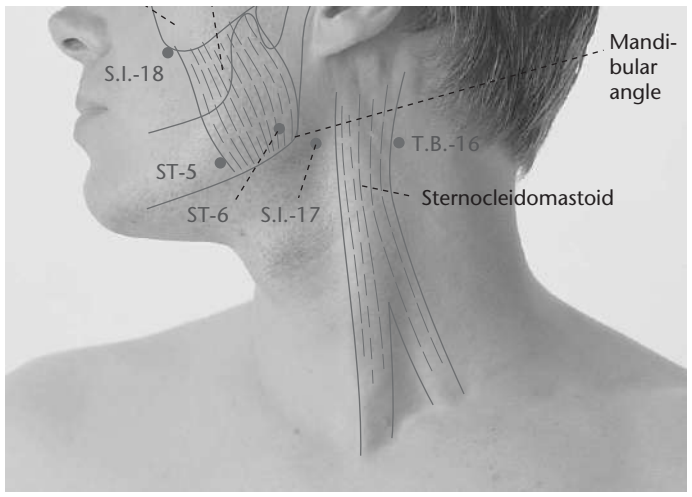

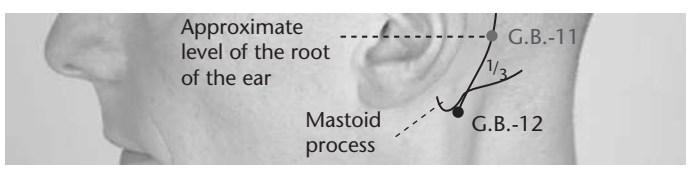

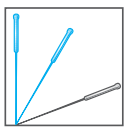

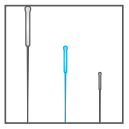

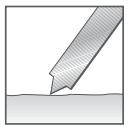

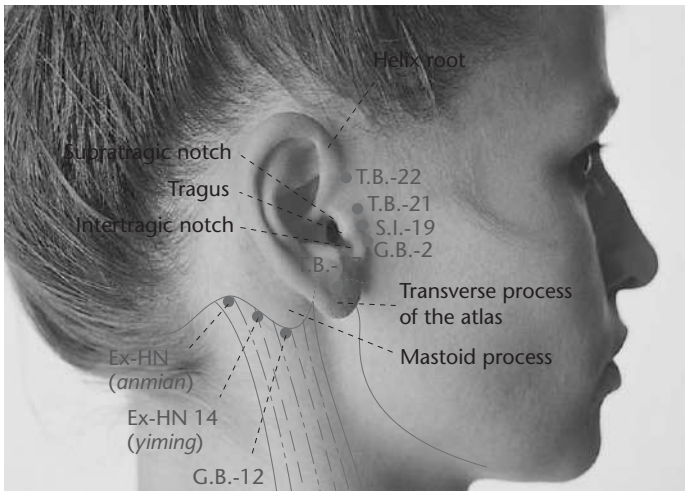

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 383

Wind Screen YIFENG T.B.-17

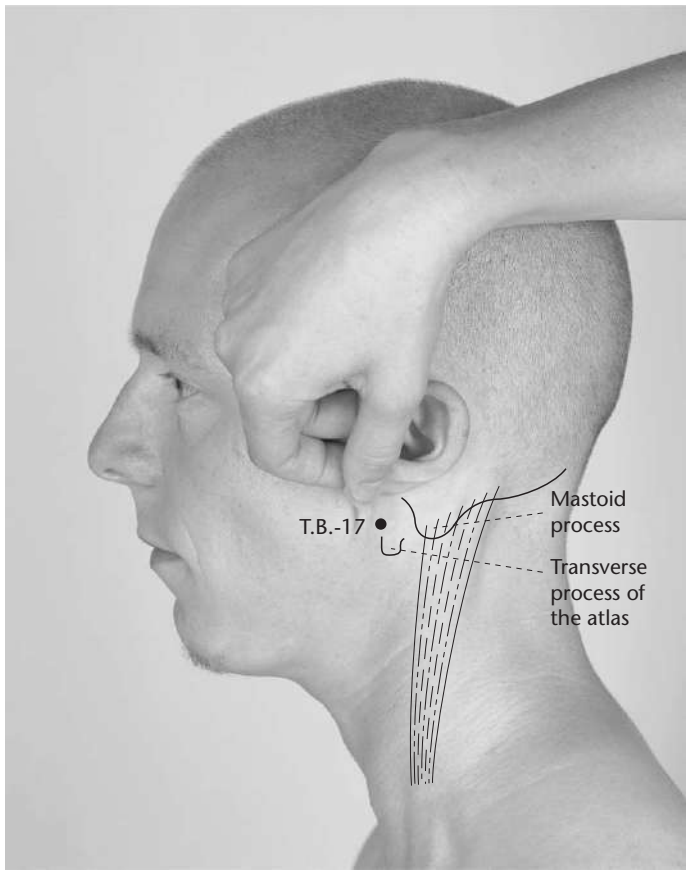

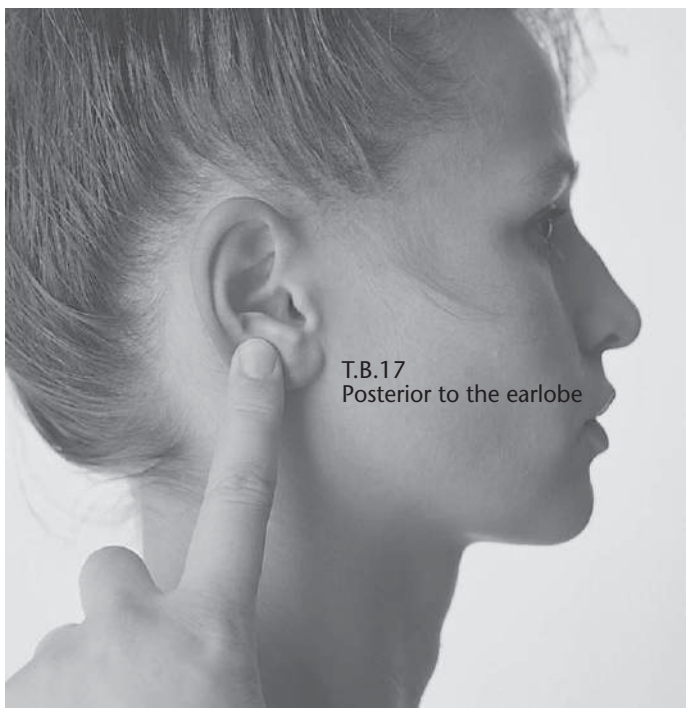

Location

With the patient's mouth open, in the depression behind the earlobe, between the mastoid process and the mandible.

How to find

This point should be located with the patient's mouth open. **Quick method:** When pressing onto the ear lobe, it will touch **T.B.-17**. **Or:** Gently fold the earlobe anteriorly. **T.B.-17** is located posterior to the lobe, in a pressure-sensitive depression which can be palpated between the mandible and the mastoid process. The tip of the needle will get close to the transverse process of the 1st cervical vertebra (transverse process of the atlas, → 3.1.4), which, below the earlobe, can be palpated as a deep bony structure often sensitive to pressure.

Needling

This point should be needled with the patient's mouth slightly open. Caution: This point is close to the facial nerve, therefore no deep needling. Superior to the transverse process, the vertebral artery curves round posteriorly, therefore strictly needle 0.5–1 cun in an anterior direction.

Actions/Indications

- Expels (external) Wind, benefits the ears, clears Heat, opens the channel and *luo*-connecting vessels, alleviates pain

Special features

Meeting point with the G.B. channel. Important local point for disorders of the ears and the temporomandibular joint.

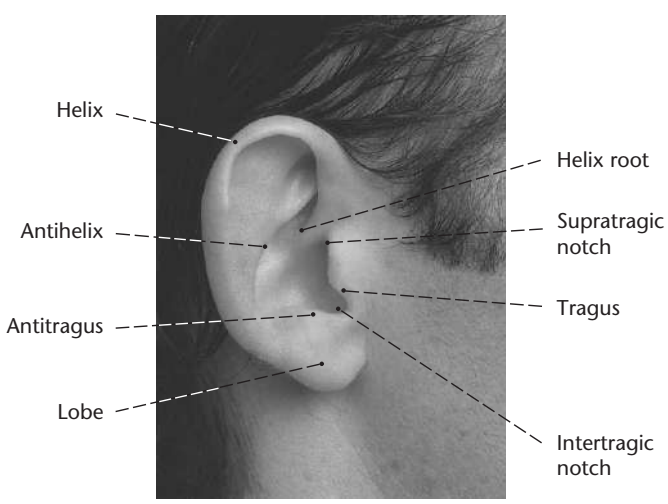

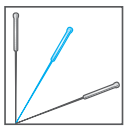
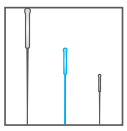

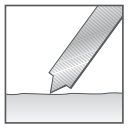

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 384

T.B.-18 *Spasm Vessel* QIMAI

Location

Posterior to the ear, in a well-defined depression in the centre of the mastoid.

How to find

On an imaginary clock face superimposed over the auricle (12 o'clock = apex of the ear, 6 o'clock = earlobe), you can find an easily palpable depression at 8 o'clock (on the right side) and at 4 o'clock (on the left side). It is located directly behind the margin of the auricle and is the location of **T.B.-18**.

→ **G.B.-11** is located slightly superior to **T.B.-18** and approximately 0.3 cun from the ear; → **G.B.-12** is located inferior to **T.B.-18**, directly posterior to the mastoid.

Needling

0.5 cun transversely (subcutaneously) in an inferior direction or prick to bleed.

Actions/Indications

- Dispels Wind
- Opens the channel, alleviates pain
- Benefits the ears

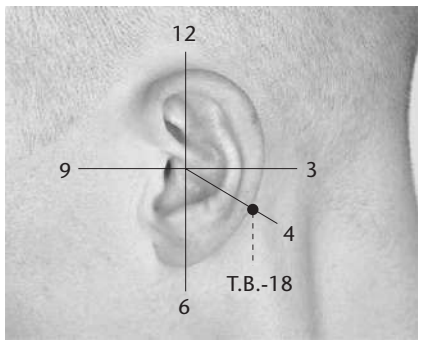

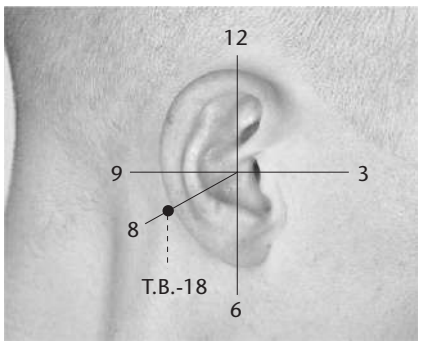

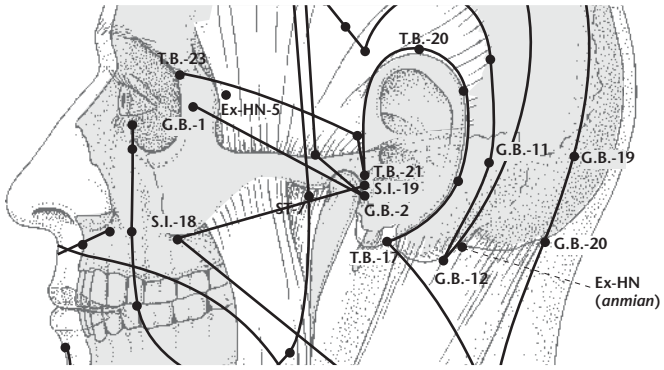

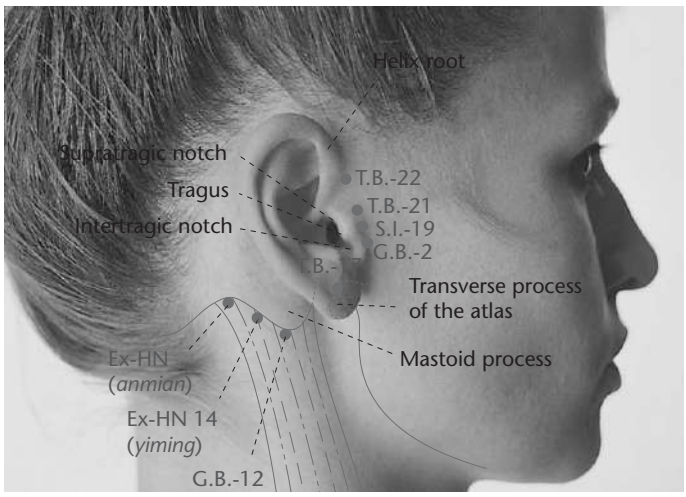

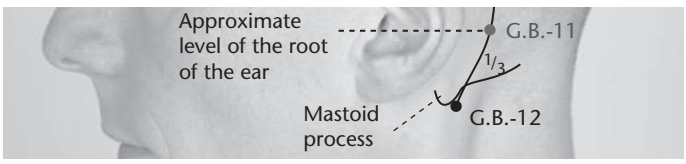

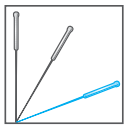
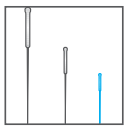

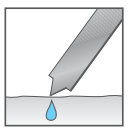
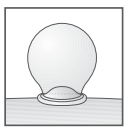

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 385

Skull's Rest LUXI T.B.-19

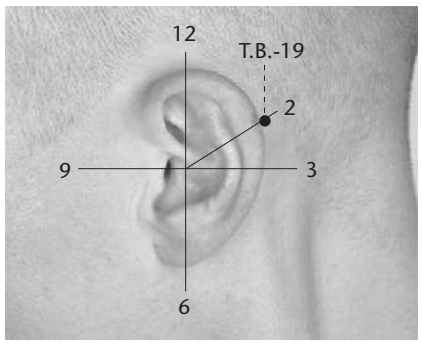

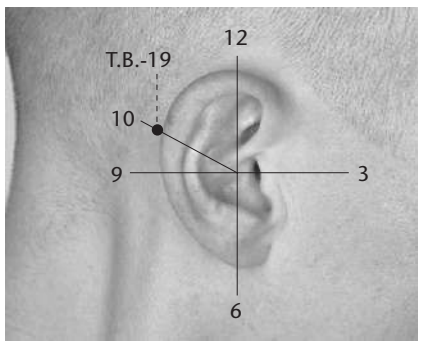

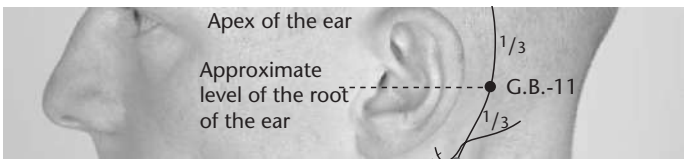

Location

Posterior to the ear, in a well-defined depression superior to the centre of the ear.

How to find

On an imaginary clock face superimposed over the auricle (12 o'clock = apex of the ear, 6 o'clock = earlobe), you can find an easily palpable depression at approximately 10 o'clock (on the right side) and at 2 o'clock (on the left side). It is located directly behind the margin of the auricle and is the location of **T.B.-19**.

→ **G.B.-11** is located slightly inferior to **T.B.-19** and approximately 0.3 cun from the margin of the ear.

Needling

Up to 0.5 cun transversely (subcutaneously) in an inferior direction or prick to bleed.

Actions/Indications

- Dispels Wind
- Opens the channel
- Benefits the ear, clears Heat

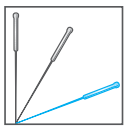
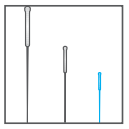

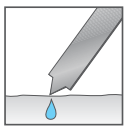

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 386

T.B.-20 *Minute Angle* JIAOSUN

Location
Directly superior to the apex of the ear.

How to find
Directly superior to the apex of the ear. → **G.B.-7** is also located on the apex of the ear, but anterior to it, within the circumauricular hairline. → **G.B.-8** is located 1.5 cun superior to the apex of the ear.

Needling
Up to 1.5 cun transversely (subcutaneously) in the direction of the disorder.

Actions/Indications
- Benefits the ears and the eyes
- Clears Heat, especially from the mouth

Special features
Meeting point with the G.B. channel, also with the S.I. channel according to some authors.

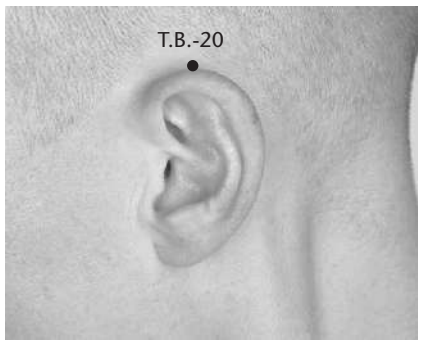

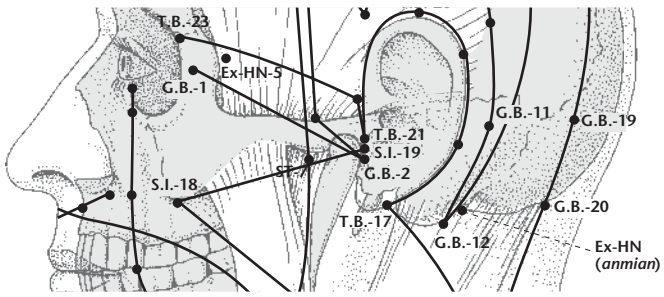

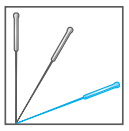
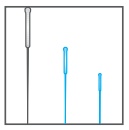

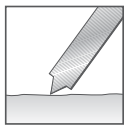

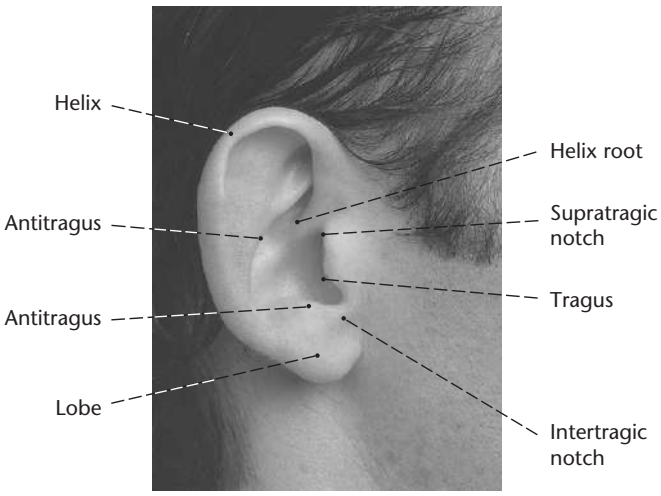

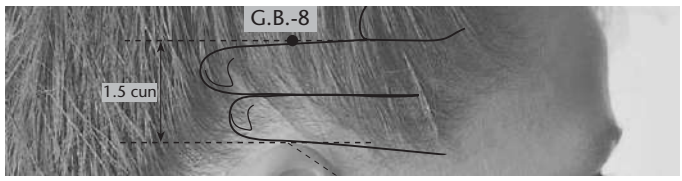

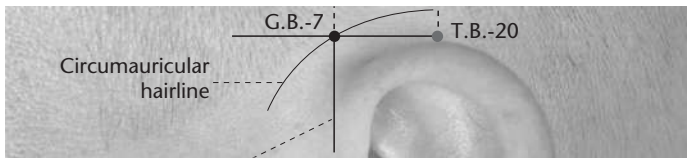

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 387

Ear Gate ERMEN T.B.-21

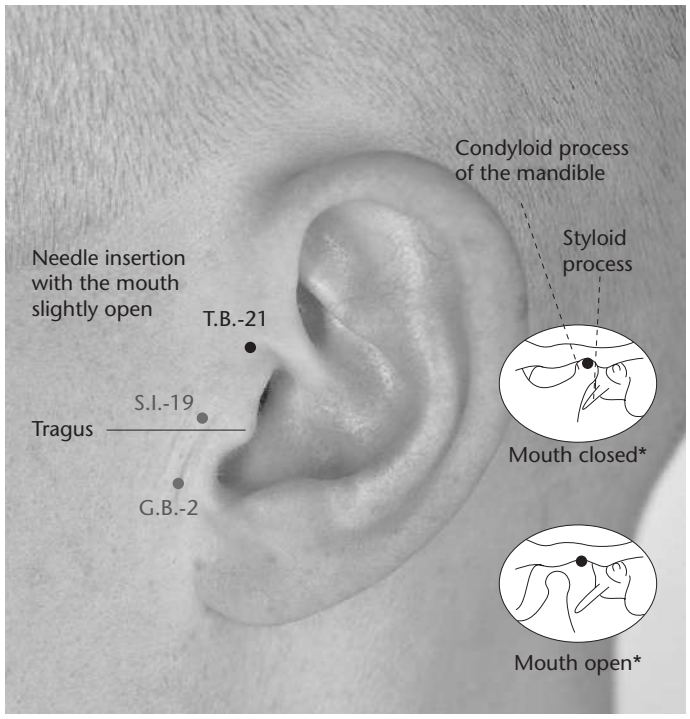

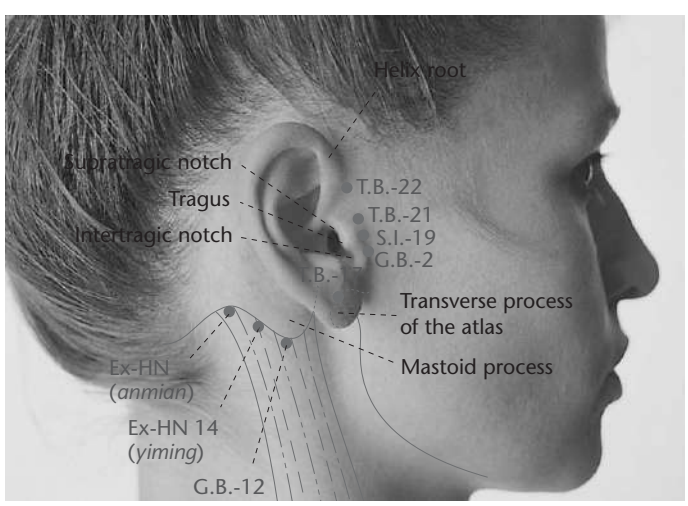

Location

Anterior to the ear, with the patient's mouth open in the depression on the level of the supratragic notch and slightly superior to the condyloid process of the mandible.

How to find

This point should be both located and needled with the patient's mouth open. This allows the condyloid process of the mandible to slide anteriorly to reveal the depression where **T.B.-21** is located. Locate the sulcus between the auricle and the cheek anterior to the tragus (this can be more or less well defined). Then locate **T.B.-21** on the level of the supratragic notch on the sulcus. If the latter cannot be clearly identified (it becomes more pronounced with increasing age), the depression can be located with an (ear) point locator when the patient's mouth is opened. **T.B.-21** is the most superior of three points located anterior to the ear (→ **S.I.-19** and → **G.B.-2** are located more distally).

Needling

0.5–1 cun vertically or slightly obliquely in an inferior direction. Needle with the patient's mouth open (to avoid intra-articular insertion). After insertion, patients can close their mouth again. Caution: Like S.I.-19 and G.B.-2, this point is close to the superficial temporal artery and the auriculotemporal nerve.

Actions/Indications

- Clears Heat, benefits the ears

Special features

Important local point for disorders of the ears

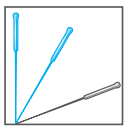
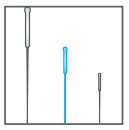

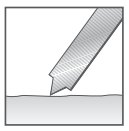
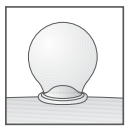

* Accoring to Deadman et al 1998.

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 388

T.B.-22 *Ear Harmony Crevice* ERHELIAO

Location

In a depression at the border of the circumauricular temporal hairline, anterior to and on the level of the root of the auricle.

How to find

First, locate the anterior aspect of the root of the ear. From there, palpate approximately the breadth of the little finger in the direction of the eye. There, **T.B.-22** is located superior to the zygomatic arch (→ 3.1.2), which forms a bony ridge when sliding inferiorly from **T.B.-22**.
Located inferior to the zygomatic arch is → **T.B.-21**.

Needling

Transversely (subcutaneously) 0.5 cun. Caution: Superficial temporal artery.

Actions/Indications

- Dispels Wind
- Opens the channel and the *luo*-connecting vessels

Special features

Meeting point with the S.I. and G.B. channels, exit point according to some authors.

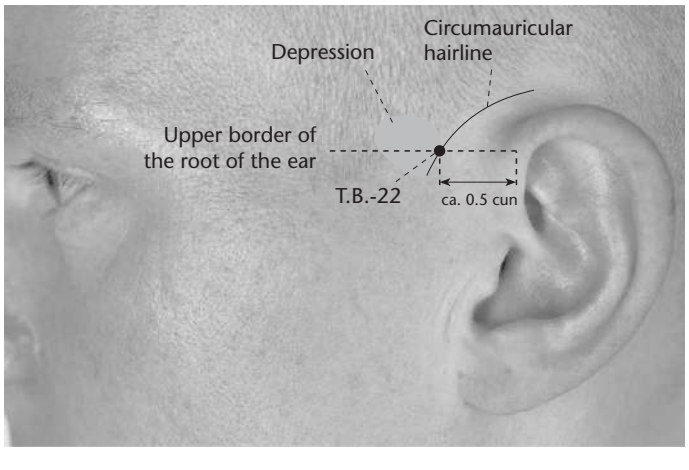

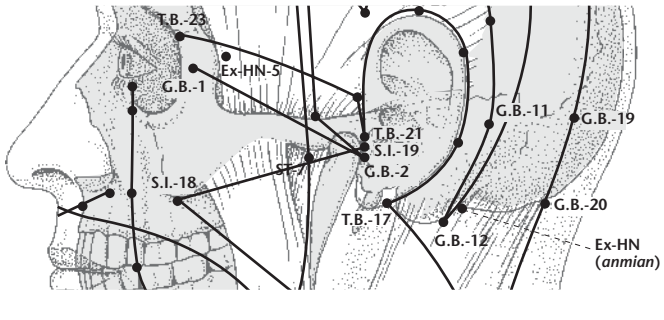

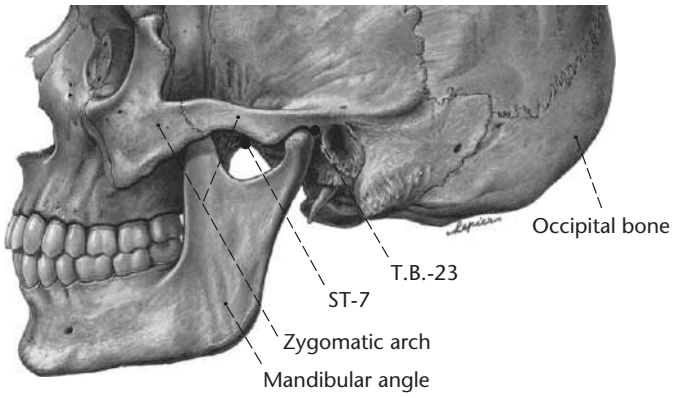

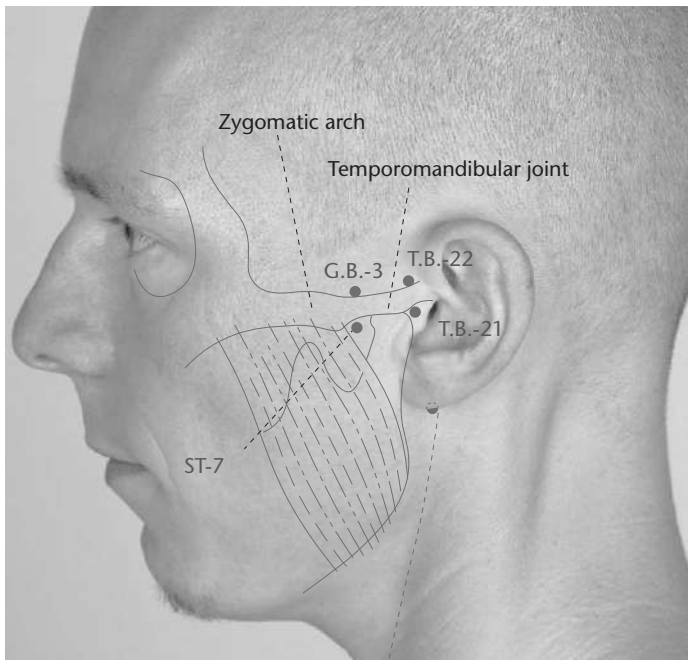

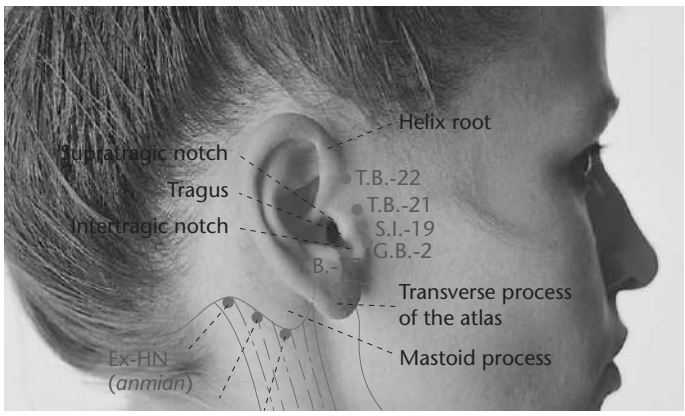

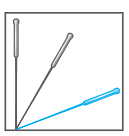
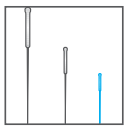

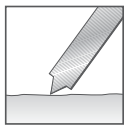

Ch04.10-F10028.qxd 2/23/08 2:27 PM Page 389

Silken Bamboo Hollow SIZHUKONG T.B.-23

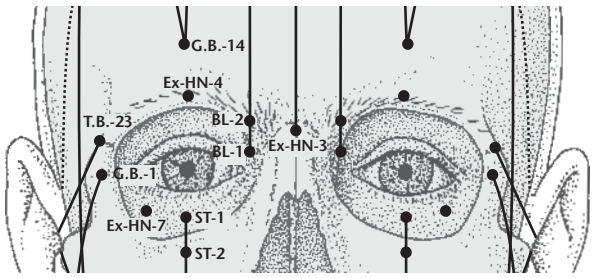

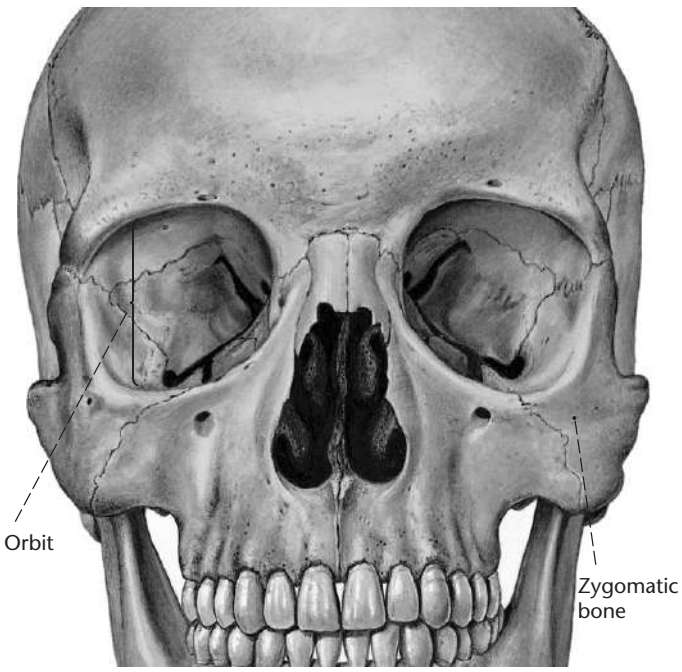

Location

On the lateral end of the eyebrow, in the bony depression of the frontozygomatic suture, between the frontal and zygomatic bones.

How to find

The frontozygomatic suture (between the zygoma and the frontal bone) is generally located at the lateral end of the eyebrow. As the position of the latter can vary considerably, the suture is a more suitable reference point. In order to locate it, palpate from the outer canthus of the eye superiorly along the orbital margin, from its zygomatic to the frontal section, until you can feel a bony depression in the area of the suture. Locate **T.B.-23** in this mainly pressure-sensitive depression.

Needling

0.5–1 cun obliquely or transversely (subcutaneously) in a posterior direction towards **Ex-HN-5** (*taiyang*). In China, this point is also needled along the eyebrow to **Ex-HN-4** (*yuyao*: in the centre of the eyebrow). Moxibustion is contraindicated according to some classic texts.

Actions/Indications

- Benefits the eyes, alleviates pain, eliminates Wind

Special features

Important local point for headaches and disorders of the eyes

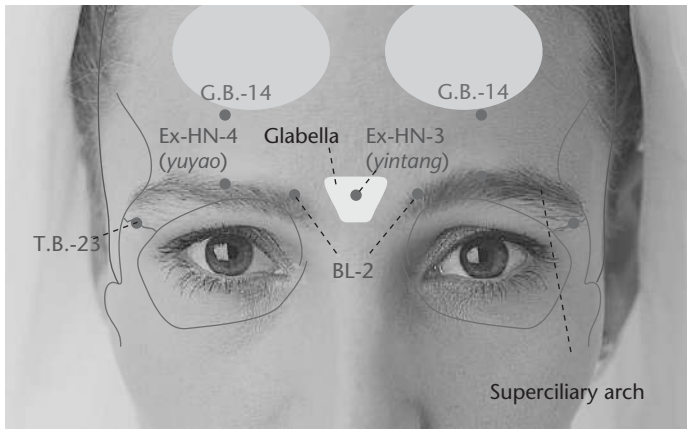

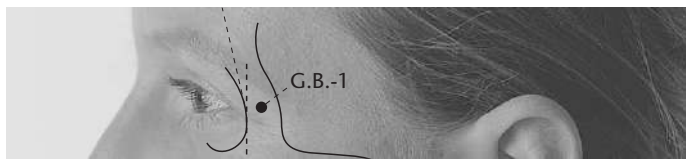

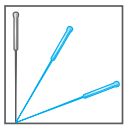
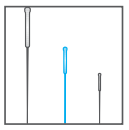

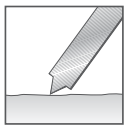

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 391

4.11 Gall Bladder Channel System – Foot *shaoyang* (*zu shao yang jing luo*)

4.11.1 The Gall Bladder Primary Channel (*zu shao yang jing*)

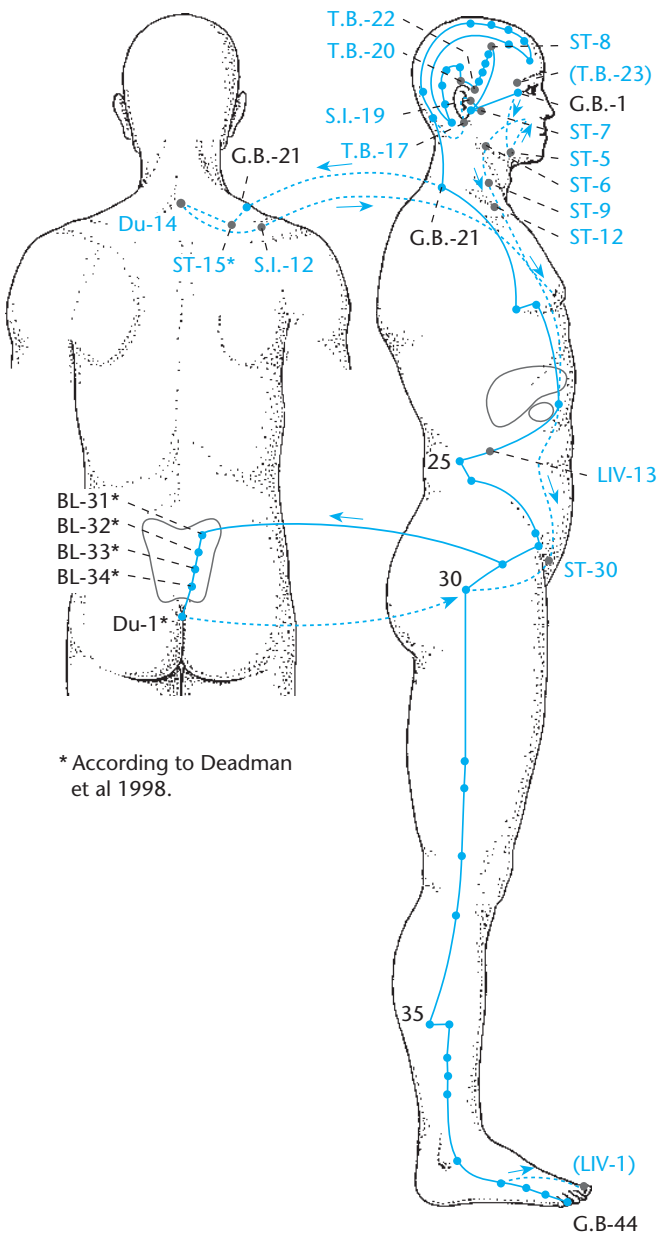

* According to Deadman et al 1998.

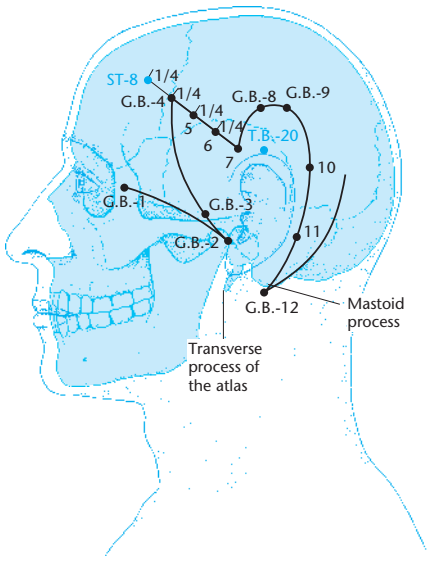

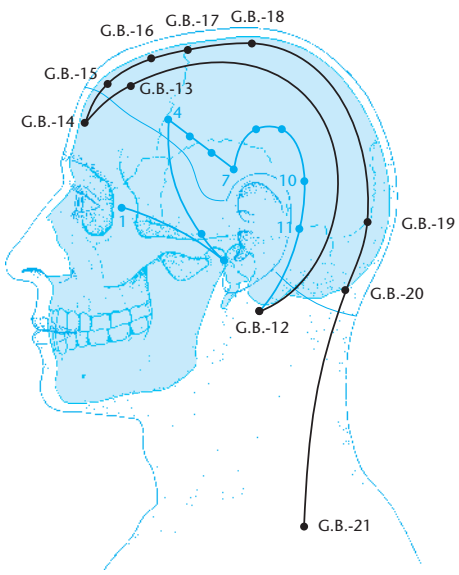

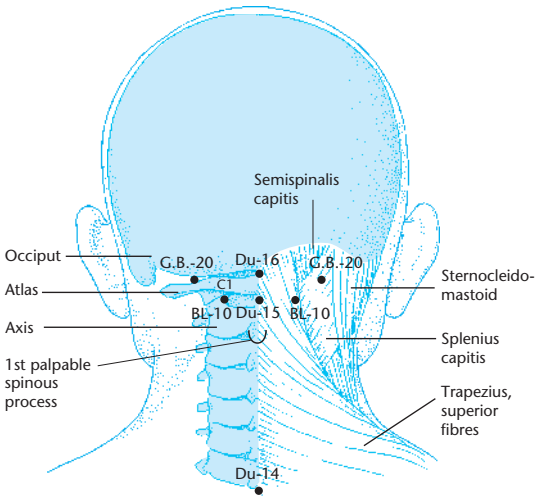

Pathway

The Gall Bladder primary channel begins at **G.B.-1** (*tongziliao*) at the outer canthus of the eye. This point is reached by a **small branch** which separates from the Triple Burner primary channel at ➞ **T.B.-23** (*sizhukong*) on the frontozygomatic suture, according to some authors at ➞ **T.B.-22** (hand–foot pairing of the third great circuit, Yang axes: *shaoyang*).

From **G.B.-1**, the **external pathway** runs to **G.B.-2** (*tinghui*) in front of the ear

➡ ascends to the forehead and **ST-8** (*touwei*)

➡ traverses the temporal region in a slight curve, passing

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 392

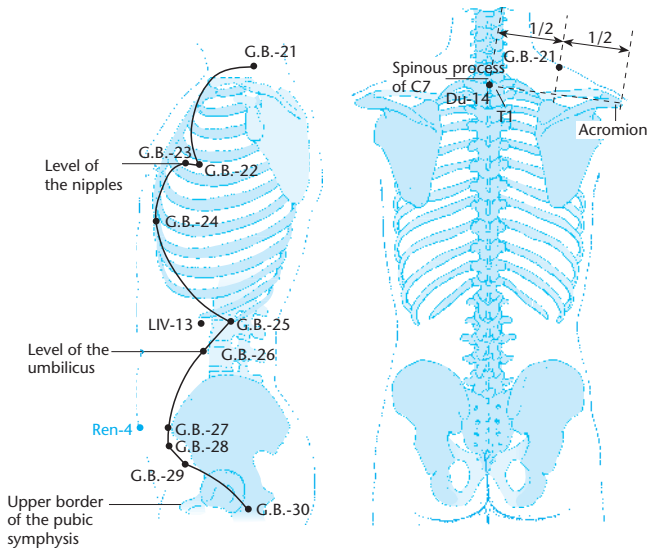

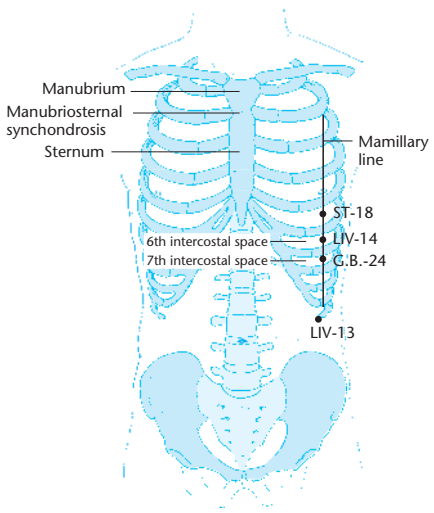

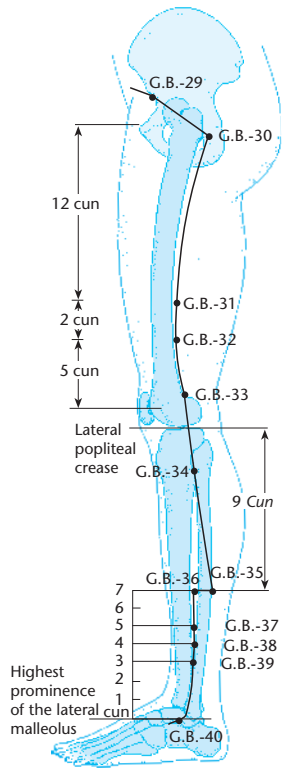

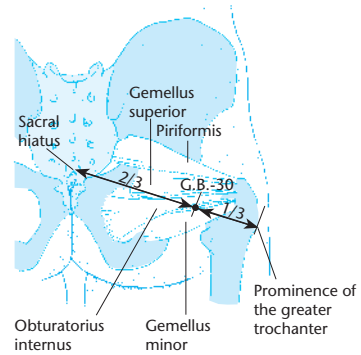

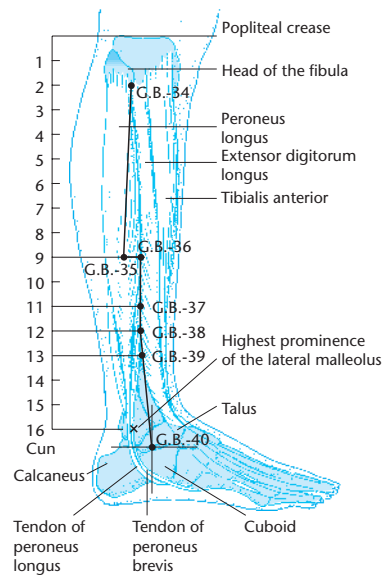

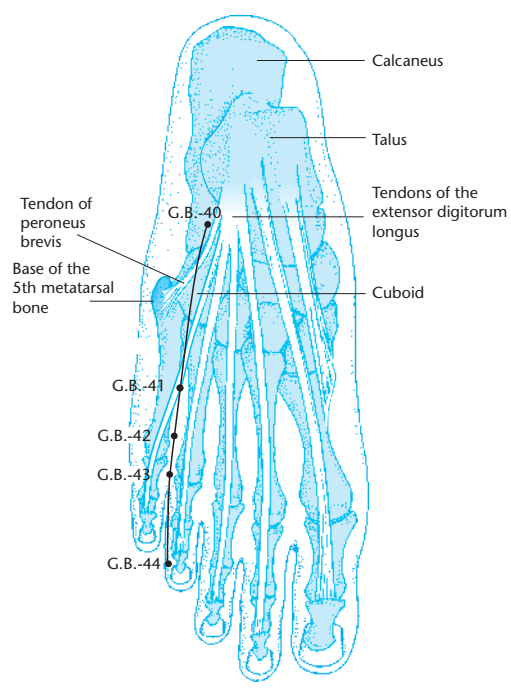

G.B.-4 to **G.B.-7**, and continues to ➞ **T.B.-22** (*erheliao*) anterior to the root of the ear

- ➡ passes ➞ **T.B.-20** (*jiaosun*) superior to the apex of the ear and curves posterior to the ear from **G.B.-8** to **G.B.-12**
- ➡ traverses the lateral aspect of the head to **G.B.-13** (*benshen*) and **G.B.-14** (*yangbai*) on the forehead
- ➡ again traverses the side of the head to **G.B.-20** (*fengchi*)
- ➡ crosses the superior aspect of the shoulder and passes **G.B.-21** (*jianjing*) and ➞ **T.B.-15** (*tianliao*) to reach ➞ **Du-14** (*dazhu*), where it meets the other Yang primary channels
- ➡ continues to **S.I.-12** (*bingfeng*), according to some authors (Deadman et al 1998) previously passing ➞ **BL-11** (*dashu*) and **ST-12** (*quepen*) in the supraclavicular fossa.

One branch runs from **G.B.-20** to ➞ **T.B.-17** (*yifeng*), enters the ear and travels to ➞ **S.I.-19** (*tinggong*), passes ➞ **ST-7** (*xiaguan*) and continues to **G.B.-1** (*tongziliao*) at the lateral canthus of the eye, descends to ➞ **ST-5** (*daying*), again ascends to the infraorbital region, where it meets the Triple Burner primary channel; then descends to ➞ **ST-6** (*jiache*) on the lower jaw, traverses the lateral aspect of the neck, passing ➞ **ST-9** (*renying*) at the sternocleidomastoid, and meets the main branch again at

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 393

→ **ST-12** (*quepen*) in the supraclavicular fossa, where the latter divides into **2 branches**:

- **The inner branch** enters the thorax, meets the Pericardium primary channel at → **P-1** (*tianchi*), runs across the diaphragm, connects with the Liver (*gan*) and the Gall Bladder (*dan*), traverses the hypochondrium and continues to the lower abdomen, enters the lumbar region near the femoral artery, according to Solinas et al (1998) via → **ST-30** (*qichong*), and continues along the pubic hairline to the hip.
- **The outer branch** descends superficially from the supraclavicular fossa, crosses the midaxillary line and continues along the lateral thoracic wall, passing → **LIV-13** (*zhangmen*) at the free end of the 11th rib and continuing to the hip to **G.B.-29** (*juliao*). It then traverses the sacral region, passing → **BL-31** (*shangliao*), → **BL-32** (*ciliao*), → **BL-33** (*zhongliao*) and → **BL-34** (*xialiao*) over the sacral foramina and continues to → **Du-1** (*changqiang*). From there, it travels laterally to **G.B.-30** (*huantiao*), where it reconnects with the main branch (→ see comment below*).

The **external pathway** descends along the lateral aspect of the thigh to the knee and further along the anterior/posterior border of the fibula, passing the anterior aspect of the lateral malleolus and ending at the ulnar corner of the nail of the fourth toe.

A further branch originates at **G.B.-41** (*zulinqi*) on the dorsum of the foot and travels between the 1st and 2nd metatarsal bones to the tip of the big toe and **LIV-1** (*dadun*), where it joins the LIV channel.

Clinical importance (→ 1.2)

Exterior (*biao*) signs and symptoms: Alternating fevers and chills, headaches, eye pain, pain in the cheek and on the chin, subaxillary swellings, deafness, lateral knee and leg pain

Interior (*li*) or *zangfu*-Organ signs and symptoms: Pain in the lateral costal region, vomiting, bitter taste in the mouth, thoracic pain

Connections of the Gall Bladder primary channel (→ 1.2)

Connections with other channels (→ 1.2)

Liver primary channel (*zu jue yin jing*)

Connection: Foot Yin–Yang pair of the third great circuit
Location: G.B.-41 → LIV-1 (on the foot)
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Triple Burner primary channel (*shou shao yang jing*)

Connection: Hand–foot pairing of the third great circuit: *shaoyang* (Yang axes)
Location: T.B.-23 (or **T.B.-22**, according to some authors) → **G.B.-1** (on the head)
Circulation: Circadian (according to the Organ clock)
Importance: Above–below relationship

Connections with *zangfu*-Organ systems

Liver (*gan*), Gall Bladder (*dan*)

4.11.2 The Gall Bladder Divergent Channel (*zu shao yang jing bie*)

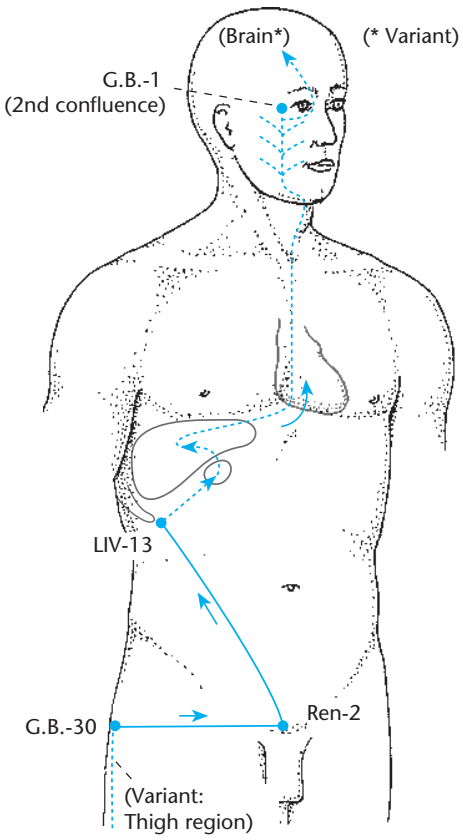

Pathway

The Gall Bladder Divergent channel separates from the Gall Bladder primary channel near **G.B.-30** (*huantiao*); according to some authors, it separates on the thigh (→ variant in the illustration)

- traverses the hip region to the anterior aspect of the body
- reaches the inguinal region and crosses the Liver divergent channel at → **Ren-2** (*qugu*)
- ascends the abdomen to **LIV-13** (*zhangmen*) below the free end of the 11th rib
- continues internally along the thoracic wall
- connects with the Gall Bladder (*dan*) and the Liver (*gan*) and passes the Heart (*xin*)
- ascends to the neck
- emerges on the lower jaw

* Not all authors describe the pathway of the G.B. channel as passing the points **BL-31** to **BL-34** and **Du-1** (see figure for variant pathway).

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 394

- disperses on the face
- connects with the Gall Bladder primary channel and the Liver divergent channel to form one of the six *he*-confluences (here: G.B./LIV as 2nd confluence → 1.3)
- continues to the eye system and ends in the Brain.

Clinical importance

- Strengthens the relationship between the Gall Bladder and the Liver (*zangfu*-Organ systems). Points on the Gall Bladder primary channel can therefore be used to treat disorders of the Liver and vice versa.
- Supports the relationship between the oesophagus and the Heart.
- The Gall Bladder Divergent channel covers the area around the eyes, thus supporting the action of some G.B. points for disorders of the eyes.

4.11.3 The Gall Bladder Sinew Channel (*zu shao yang jing jin*)

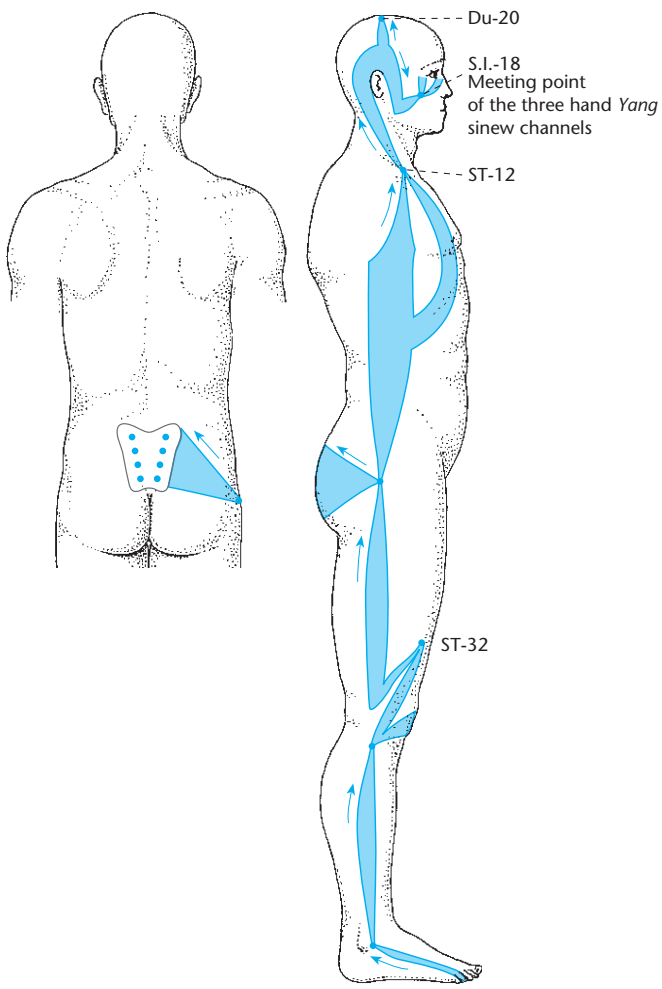

Pathway

The Gall Bladder sinew channel begins at the 4th toe, binds (*jie*) anterior and inferior to the lateral malleolus in the area of **G.B.-40** (*qiuxu*), follows the lateral aspect of the leg, binds (*jie*) at the fibula and the lateral aspect of the knee.

From the fibula, it continues to ascend the lateral aspect of the thigh

- while a **branch** runs obliquely to → **ST-32** (*futu*), where it binds (*jie*).

The main branch further ascends the leg to bind (*jie*) at the greater trochanter. Here, a **branch** separates and disperses over the gluteal and sacral regions.

From the hip, the channel continues to ascend the flanks to the lower costal region, where it divides into **two branches**:

- **one branch** traverses the lateral aspect of the thorax and ascends to bind (*jie*) at → **ST-12** (*quepen*)
- **the other branch** follows the midaxillary line to the thorax, meeting the other branch in the supraclavicular fossa.

From **ST-12**, the channel ascends the lateral aspect of the neck

- curves around the ear
- a branch continues to the apex of the ear and to → **Du-20** (*baihui*)
- from the temporal region, **a branch** descends and traverses the cheek to the lower jaw
- crosses the zygomatic arch to reach **S.I.-18** (*quanliao*), where it meets the other foot Yang sinew channels and divides into **two branches**. One branch travels laterally to the root of the nose, the other to the outer canthus of the eye.

Clinical importance

Pathology: Stiffness and distending sensations in the area of the 4th toe as well as on the lateral aspect of the knee. Limited range of motion of the knee joint. Pain, tension and distending sensations in the popliteal crease radiating to the thigh and the sacral region and vice versa. Pain and distending sensations in the sacral region radiating to the hypochondrium and vice versa. Pain and tension in the supraclavicular fossa, thorax, breast and neck region. Pain along the left side of the channel and inability to open the right eye and vice versa.

Indication: Predominantly for pain, muscle tension, stiff joints and distending sensations on the lateral aspect of the body. Used for disorders of the anterior, lateral and posterior aspects of the legs, especially those affecting the knees and the thighs. The Gall Bladder Sinew channel spreads in the thorax and the breasts so that points on the G.B. channel can be used for disorders in those areas. Also for headaches on the vertex and temporal regions. The channel also reaches the lateral aspect of the nose, which supports the relationship with the nose. Therefore, G.B. points can be used for chronic nasal disorders, chronic sinusitis, etc.

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 395

4.11.4 The Gall Bladder *luo*-Connecting Channel (*zu shao yang luo mai*)

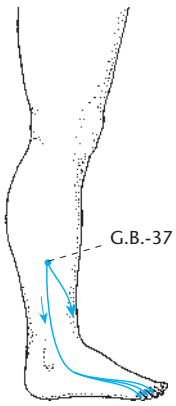

Pathway

The Gall Bladder *luo*-connecting channel separates from the Gall Bladder primary channel at its *luo*-connecting point **G.B.-37** (*guangming*) (→ 8.1.2). It forms a three-dimensional reticulate network, dividing into multiple branches and sub-branches (*sun luo, fu luo, xue luo* → 1.5) within the surrounding tissue.

- Horizontal divisions run to the Interiorly–Exteriorly paired Liver primary channel; according to some schools of thought (for example, Ngyen Van Nghi → Appendix), they travel as a **transverse** G.B. *luo* vessel to the *yuan*-source point **LIV-3** (*taichong*).
- A **longitudinal** division descends to the dorsum of the foot and spreads to the 3rd, 4th and 5th toes.

Clinical importance (→ 8.1.2)

Pathology

Excess (*shi*): Cold sensations in the feet

Deficiency (*xu*): Weakness, weak musculature of the foot, including difficulty in standing, paralysis of the lower extremity

4.11.5 Cutaneous Region (*shao yang pi bu*)

See description and figures → 1.6.

4.11.6 Points of the Gall Bladder Primary Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): G.B.-40** (*qiuxu*) ■■
- ***Luo*-connecting point (→ 8.1.2.): G.B.-37** (*guangming*) ■■
- ***Xi*-cleft point (→ 8.1.3): G.B.-36** (*waiqiu*)
- **Associated Back-*shu* point (→ 8.1.4): BL-19** (*danshu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): G.B.-24** (*riyue*) ■■
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Metal): **G.B.-44** (*zuqiaoyin*) ■■
  ying-spring point (Water), tonification point: **G.B.-43** (*xiaxi*)
  shu-stream point (Wood), *ben* point (Five Phases): **G.B.-41** (*zulinqi*) ■■
  jing-river point (Fire), sedation point: **G.B.-38** (*yangfu*)
  he-sea point (Earth): **G.B.-34** (*yanglingquan*) ■■
- ***Hui*-meeting point (→ 8.1.7)**
  – with the sinews: **G.B.-34** (*yanglingquan*) ■■
  – with the Marrow: **G.B.-39** (*xuanzhong*) ■■
- **Opening point (→ 8.1.8) of the *dai mai:* G.B.-41** (*zulinqi*) ■■
- **Lower *he*-sea point (→ 8.1.9) of the Gall Bladder: G.B.-34** (*yanglingquan*) ■■
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the S.I. and T.B. channels: **G.B.-1** (*tongziliao*)
  – with the T.B. and ST channels: **G.B.-3** (*shangguan*), **G.B.-4** (*hanyan*)
  – with the T.B., L.I. and ST channels: **G.B.-5** (*xuanlu*), **G.B.-6** (*xuanli*)
  – with the BL channel: **G.B.-7** (*qubin*); **G.B.-8** (*shuaigu*) ■■, **G.B.-9** (*tianchong*), **G.B.-10** (*fubai*)
  – with the BL, (S.I., T.B.*) channels: **G.B.-11** (*touqiaoyin*)
  – with the BL channel: **G.B.-12** (*wangu*) ■■
  – with the *yang wei mai*: **G.B.-13** (*benshen*) ■■
  – with the *yang wei mai*, (T.B., ST, L.I. channels*): **G.B.-14** (*yangbai*) ■■
  – with the *yang wei mai* and the BL channel: **G.B.-15** (*toulinqi*)
  – with the *yang wei mai*: **G.B.-16** (*muchuang*). **G.B.-17** (*zhengying*), **G.B.-18** (*chengling*), **G.B.-19** (*naokong*)
  – with the *yang wei mai, yang qiao mai* and (T.B. channel*): **G.B.-20** (*fengchi*) ■■
  – with the *yang wei mai*, T.B. and (ST*) channels: **G.B.-21** (*jianjing*) ■■
  – with the BL channel*: **G.B.-23** (*zhejin*)
  – with the SP channel and (*yang wei mai**): **G.B.-24** (*riyue*) ■■
  – with the *dai mai*: **G.B.-26** (*daimai*), **G.B.-27** (*wushu*), **G.B.-28** (*weidao*)
  – with the *yang qiao mai* and (*dai mai**): **G.B.-29** (*juliao*)
  – with the BL channel: **G.B.-30** (*huantiao*) ■■
  – with the *yang wei mai**: **G.B.-35** (*yangjiao*)
  – of other channels with the G.B. channel: **ST-7, ST-8, ST-9, ST-12, S.I.12, S.I.-19, (BL-1, BL-11*), BL-31–34, P-1, (T.B.-15*), T.B-17, T.B.-20, T.B.-22, LIV-13, Du-1, Du-14, (Du-20, ST-5, ST-6, ST-30*)**
- **Gao Wu command point (→ 8.1.11): –**
- **Window of Heaven point (→ 8.1.12): G.B.-9** (*tianchong*)
- **Points of the Four Seas (→ 8.1.13): –**
- **Ma Dan Yang Heavenly Star points (→ 8.1.14): G.B.-30** (*huantiao*) ■■, **G.B.-34** (*yanglingquan*) ■■

* Mentioned by only some authors.

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 396

- **Sun Si Miao Ghost point (→ 8.1.15): –**
- **Other functional points:**
  - main point for all disorders of the head, sensory organs, Brain: **G.B.-20** (*fengchi*) ■■
  - front-*mu* point of the Kidneys (→ 8.1.5): **G.B.-25** (*jingmen*) ■■
  - *xi*-cleft point (→ 8.1.3) of the *yang wei mai:* **G.B.-35** (*yangjiao*)

Points according to region

- **Local points (→ 8.2.1):** forehead – **G.B.-14** (*yangbai*) ■■; head, temporal region – **G.B.-8** (*shuaigu*) ■■; head, occipital region – **G.B.-20** (*fengchi*) ■■; ears – **G.B.-2** (*tingshui*) ■■; cervical spine – **G.B.-20** (*fengchi*) ■■; neck and shoulder region – **G.B.-20** (*fengchi*) ■■, **G.B.-21** (*jianjing*) ■■; Gall Bladder – **G.B.-24** (*riyue*) ■■; hypochondrium – **G.B.-25** (*jingmen*) ■■, **G.B.-26** (*daimai*), **G.B.-27** (*wushu*); hip – **G.B.-29** (*juliao*), **G.B.-30** (*huantiao*) ■■; knee – **G.B.-34** (*yanglingquan*) ■■, **G.B.-35** (*yangjiao*); foot – **G.B.-40** (*qiuxu*) ■■
- **Adjacent points (→ 8.2.1):** head, temporal region – **G.B.-20** (*fengchi*) ■■; ears – **G.B.-8** (*shuaigu*) ■■, **G.B.-20** (*fengchi*) ■■; cervical spine and shoulders – **G.B.-21** (*jianjing*) ■■; lumbar region – **G.B.-25** (*jingmen*) ■■, **G.B.-30** (*huantiao*) ■■; hip – **G.B.-31** (*fengshi*); knee – **G.B.-33** (*xiyangguan*); foot – **G.B.-34** (*yanglingquan*) ■■; toes – **G.B.-38** (*yangfu*)
- **Distal points (→ 8.2.1):** head, temporal region – **G.B.-43** (*xiaxi*), **G.B.-41** (*zulinqi*) ■■; eyes – **G.B.-37** (*guangming*) ■■; ears – **G.B.-41** (*zulinqi*) ■■; Liver – **G.B.-34** (*yanglingquan*) ■■; Gall Bladder – **G.B.-34** (*yanglingquan*) ■■, **G.B.-40** (*qiuxu*) ■■; hypochondrium – **G.B.-38** (*yangfu*), **G.B.-34** (*yanglingquan*) ■■, **G.B.-43** (*xiaxi*); cervical spine and hip – **G.B.-39** (*xuanzhong*) ■■; hip – **G.B.-41** (*zulinqi*) ■■

Specific points according to the channel pathway (in numeric order)

- **G.B.-1** (*tongziliao*): *jiaohui*-meeting point with the S.I. and T.B. channels (→ 8.1.10)
- **G.B.-2** (*tinghui*) ■■: local point for the ears (→ 8.2.1)
- **G.B.-3** (*shangguan*): *jiaohui*-meeting point with the T.B. and ST channels (→ 8.1.10)
- **G.B.-4** (*hanyan*): *jiaohui*-meeting point with the T.B. and ST channels (→ 8.1.10)
- **G.B.-5** (*xuanlu*): *jiaohui*-meeting point with the T.B., L.I. and ST channels (→ 8.1.10)
- **G.B.-6** (*xuanli*):): *jiaohui*-meeting point with the T.B., L.I. and ST channels (→ 8.1.10)
- **G.B.-7** (*qubin*):): *jiaohui*-meeting point with the BL channel (→ 8.1.10)
- **G.B.-8** (*shuaigu*) ■■: *jiaohui*-meeting point with the BL channel (→ 8.1.10); local point for the lateral aspect of the head (→ 8.2.1); adjacent point for the ears (→ 8.2.1)
- **G.B.-9** (*tianchong*): *jiaohui*-meeting point with the BL channel (→ 8.1.10)*; Window of Heaven point (→ 8.1.12)*
- **G.B.-10** (*fubai*): *jiaohui*-meeting point with the BL channel (→ 8.1.10)
- **G.B.-11** (*touqiaoyin*): *jiaohui*-meeting point with the BL, (S.I., T.B.)* channels (→ 8.1.10)
- **G.B.-12** (*wangu*): *jiaohui*-meeting point with the BL channel (→ 8.1.10)
- **G.B.-13** (*benshen*) ■■: *jiaohui*-meeting point with the *yang wei mai* (→ 8.1.10)
- **G.B.-14** (*yangbai*) ■■: *jiaohui*-meeting point with the *yang wei mai* and (T.B., ST, L.I.*) channels (→ 8.1.10); local point for the forehead (→ 8.2.1)
- **G.B.-15** (*toulinqi*): *jiaohui*-meeting point with the *yang wei mai* and BL channel (→ 8.1.10)
- **G.B.-16** (*muchuang*): *jiaohui*-meeting point with the *yang wei mai* (→ 8.1.10)
- **G.B.-17** (*zhengying*): *jiaohui*-meeting point with the *yang wei mai* (→ 8.1.10)
- **G.B.-18** (*chengling*): *jiaohui*-meeting point with the *yang wei mai* (→ 8.1.10)
- **G.B.-19** (*naokong*): *jiaohui*-meeting point with the *yang wei mai* (→ 8.1.10)
- **G.B.-20** (*fengchi*) ■■: *jiaohui*-meeting point with the *yang wei mai, yang qiao mai* and (T.B. channel*) (→ 8.1.10); local point for the occiput (→ 8.2.1); local point for the neck and shoulders (→ 8.2.1); local point for all disorders of the head, sensory organs, Brain; adjacent point for the lateral aspect of the head and the ears (→ 8.2.1)
- **G.B.-21** (*jianjing*) ■■: *jiaohui*-meeting point with the *yang wei mai*, T.B. and (ST*) channels (→ 8.1.10); adjacent point for the neck and shoulders (→ 8.2.1)
- **G.B.-23** (*zhejin*): *jiaohui*-meeting point with the BL channel (→ 8.1.10)
- **G.B.-24** (*riyue*) ■■: associated Front-*mu* point (→ 8.1.5); *jiaohui*-meeting point with the SP channel and (*yang wei mai**) (→ 8.1.10); local point for the Gall Bladder (→ 8.2.1)
- **G.B.-25** (*jingmen*) ■■: Front-*mu* point of the Kidneys (→ 8.1.5); local point for the hypochondrium (→ 8.2.1); adjacent point for the lumbar region (→ 8.2.1)
- **G.B.-26** (*daimai*): *jiaohui*-meeting point with the *dai mai* (→ 8.1.10); local point for the hypochondrium (→ 8.2.1)
- **G.B.-27** (*wushu*): *jiaohui*-meeting point with the *dai mai* (→ 8.1.10); local point for the hypochondrium (→ 8.2.1)
- **G.B.-28** (*weidao*): *jiaohui*-meeting point with the *dai mai* (→ 8.1.10);
- **G.B.-29** (*juliao*): *jiaohui*-meeting point with the *yang qiao mai* and (*dai mai**) (→ 8.1.10); local point for the hip (→ 8.2.1)
- **G.B.-30** (*huantiao*) ■■: *jiaohui*-meeting point with the BL channel (→ 8.1.10); Ma Dan Yang Heavenly Star point (→ 8.1.14); adjacent point for the lumbar region and the hip (→ 8.2.1); local point for the hip (→ 8.2.1)
- **G.B.-31** (*fengshi*): adjacent point for the hip (→ 8.2.1)
- **G.B.-33** (*xiyangguan*): adjacent point for the knee (→ 8.2.1)

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 397

- **G.B.-34** (*yanglingquan*) ■■: *he*-sea point (Earth) (→ 8.1.6); *hui*-meeting point with the sinews (→ 8.1.7); lower *he*-sea point of the Gall Bladder (→ 8.1.9); Ma Dan Yang Heavenly Star point (→ 8.1.14); distal point for the Liver, Gall Bladder and hypochondrium (→ 8.2.1); local point for the knee (→ 8.2.1); adjacent point for the foot (→ 8.2.1)
- **G.B.-35** (*yangjiao*): *xi*-cleft point of the *yang wei mai* (→ 8.1.3); *jiaohui*-meeting point with the *yang wei mai** (→ 8.1.10); local point for the knee (→ 8.2.1)
- **G.B.-36** (*waiqiu*): *xi*-cleft point (→ 8.1.3)
- **G.B.-37** (*guangming*) ■■: *luo*-connecting point (→ 8.1.2); distal point for the eyes (→ 8.2.1)
- **G.B.-38** (*yangfu*): *jing*-river point (Fire) (→ 8.1.6); sedation point; distal point for the hypochondrium (→ 8.2.1)
- **G.B.-39** (*xuanzhong*) ■■: *hui*-meeting point with the Marrow (→ 8.1.7); distal point for the cervical spine and the hip (→ 8.2.1)
- **G.B.-40** (*qiuxu*) ■■: *yuan*-source point (→ 8.1.1); local point for the foot (→ 8.2.1); distal point for the Gall Bladder (→ 8.2.1)
- **G.B.-41** (*zulinqi*) ■■: *shu*-stream point (Wood); *ben* point (Five phases) (→ 8.1.6); Opening point (→ 8.1.8) of the *dai mai*; distal point for the temporal region and the ears (→ 8.2.1); distal point for the hip (→ 8.2.1)
- **G.B.-43** (*xiaxi*): *ying*-spring point (Water) (→ 8.1.6); tonification point; distal point for the lateral aspect of the head and the hypochondrium (→ 8.2.1)
- **G.B.-44** (*zuqiaoyin*) ■■: *jing*-well point (Metal) (→ 8.1.6)

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 399

Pupil Crevice TONGZILIAO G.B.-1

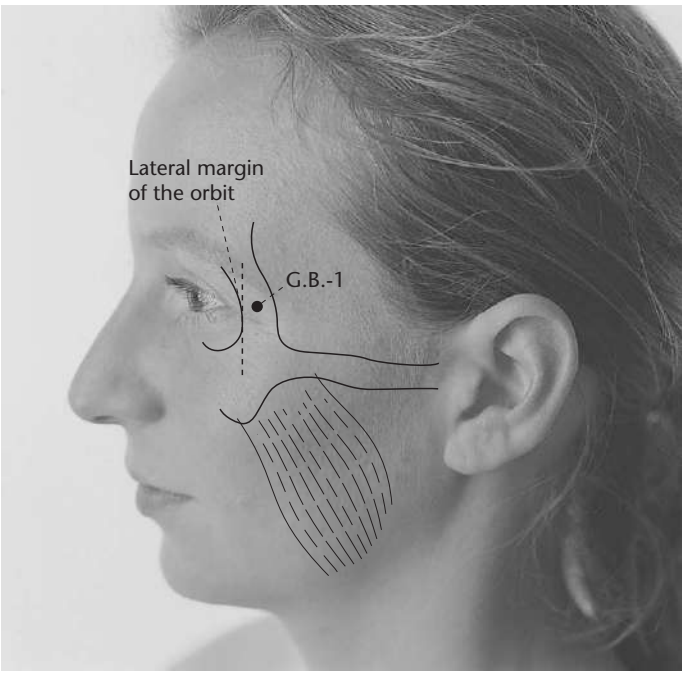

Location

In a bony depression on the lateral aspect of the orbital margin, on the level of the outer canthus of the eye.

How to find

From the outer canthus of the eye, palpate in a lateral direction. Locate **G.B.-1** on the outer aspect of the orbit (diagonally inferior to the temple), where a bony depression can be palpated. → **T.B.-23** is located more superiorly on the lateral end of the eyebrow, in a depression on the frontozygomatic suture. → **BL-1** is located at the inner canthus of the eye.

Needling

0.2–0.3 cun obliquely in a posterior direction or up to 1 cun transversely (subcutaneously) towards → **Ex-HN-5** (*taiyang*). According to some texts, moxibustion is contraindicated.

Actions/Indications

- Expels Wind and Heat from the eyes

Special features

Meeting point with the S.I. and T.B. channels, entry point.

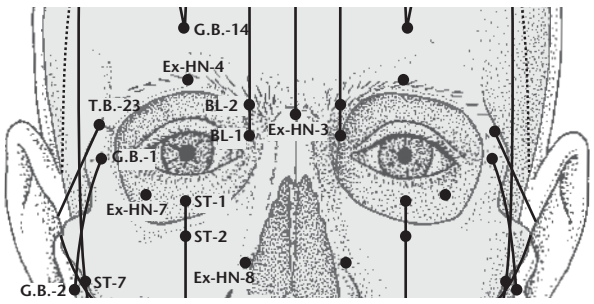

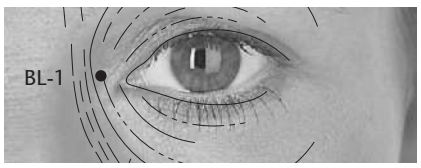

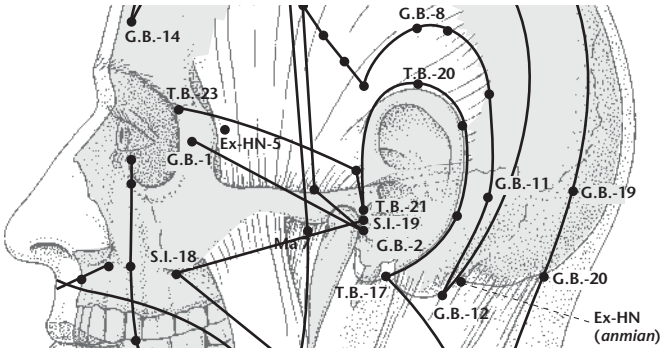

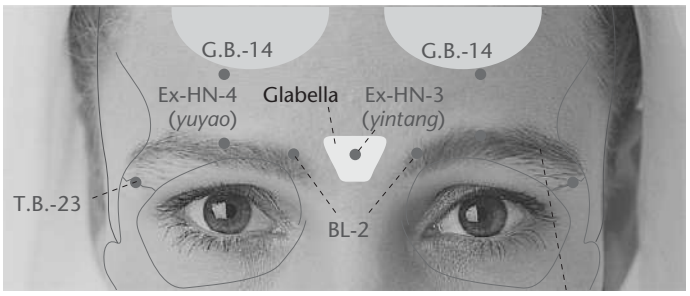

Ch04.11-F10028.qxd 2/23/08 11:23 AM Page 400

G.B.-2 *Meeting of Hearing* TINGHUI

Location

Anterior to the ear, with the patient's mouth open in the depression on the level of the intertragic notch, on the lower border of the condyloid process of the mandible.

How to find

Locate and needle with the patient's mouth open as this allows the condyloid process of the mandible to slide towards the anterior, revealing the depression where **G.B.-2** is located. Find the vertical sulcus anterior to the ear (at the ear/cheek junction), which may be more or less pronounced. Then locate **G.B.-2** on the sulcus, on the level of the intertragic notch. If the sulcus is not clearly defined (it becomes more clearly visible with increasing age), the depression can be located by using an (ear) point locator. **G.B.-2** is the most distal of three points located anterior to the ear (→ **S.I.-19** and → **T.B.-21** are both located more superiorly).

Needling

0.5–1 cun vertically or slightly obliquely in an inferior direction. Needle with the patient's mouth open to avoid intra-articular insertion. Caution: Like **T.B.-21** and **S.I.-19**, this point is located close to the superficial temporal artery and the auriculotemporal nerve.

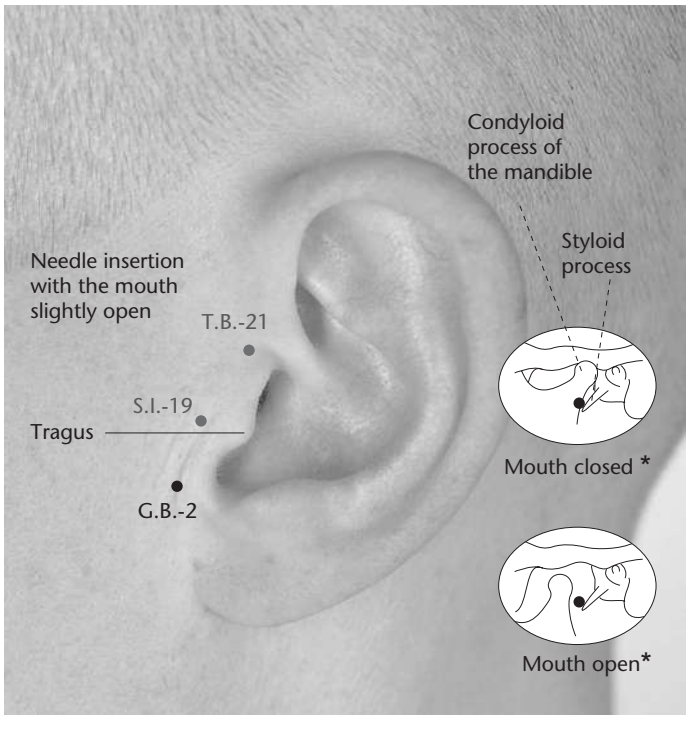

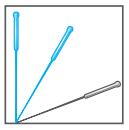

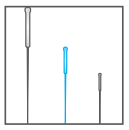

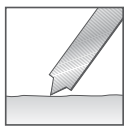

Actions/Indications

- Eliminates Wind, clears Heat, benefits the ears and the temporomandibular joint
- Opens the channel and the *luo*-connecting vessel, alleviates pain

Special features

Important local point for disorders of the ears and the jaw. This point is often used alternately with **T.B.-21** and **S.I.-19**.

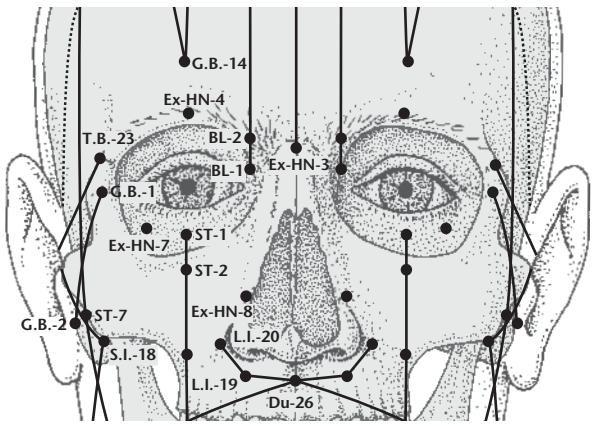

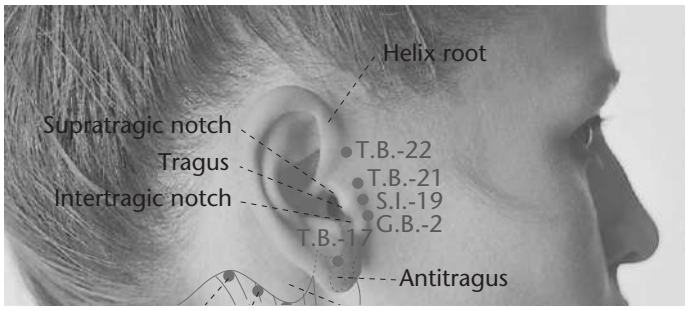

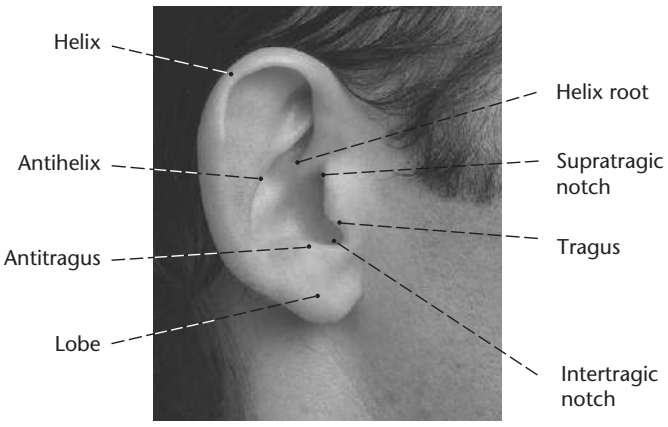

* According to Deadman et al 1998.

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 401

Above the Joint SHANGGUAN G.B.-3

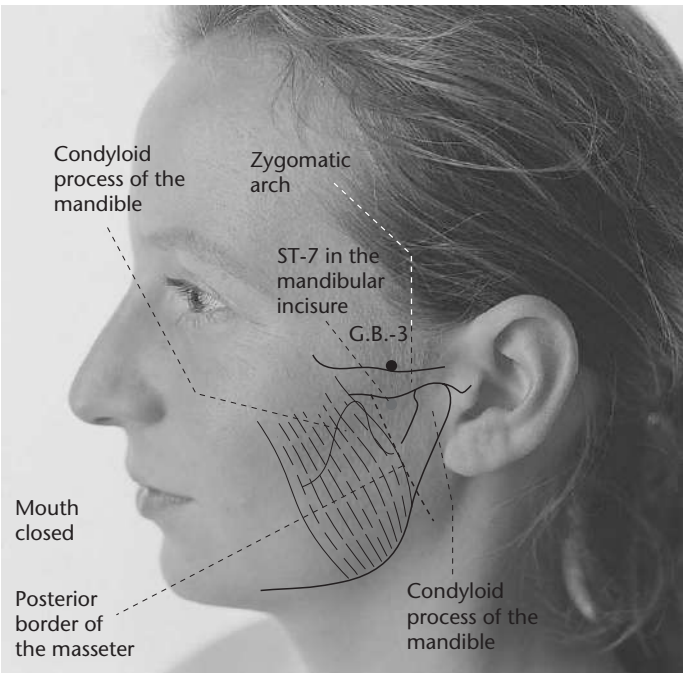

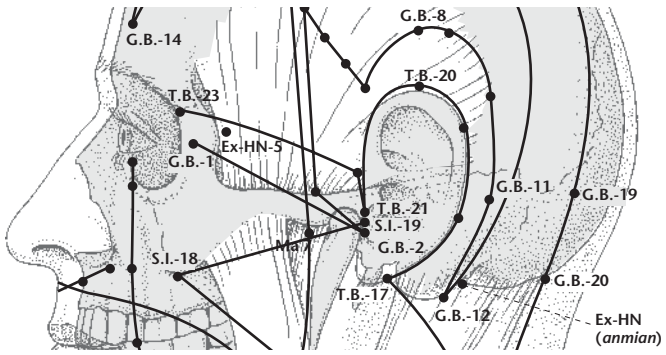

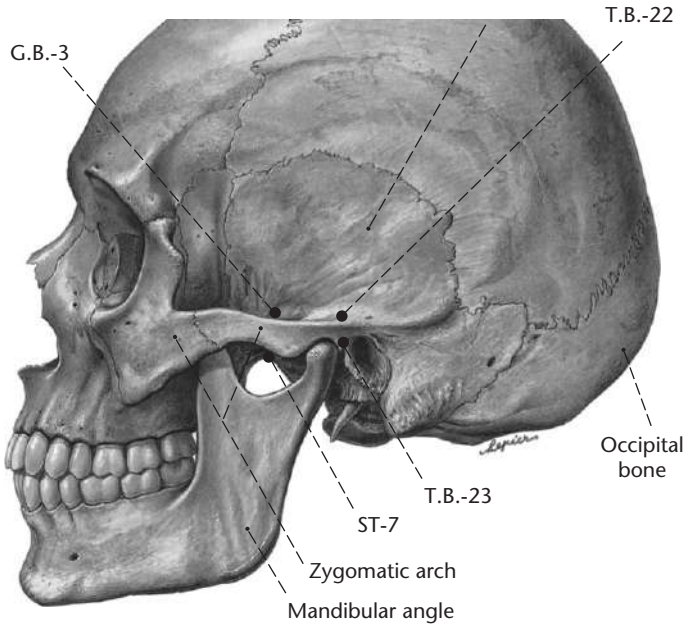

Location

In a depression on the upper border of the zygomatic arch, approximately 1 cun anterior to the root of the ear, superior to → **ST-7**.

How to find

Locate the zygomatic arch (→ 3.1.2) by palpating from the root of the ear (root of the helix) approximately 1 cun towards the orbit. Follow its course by placing one finger above and the other finger below the arch. As soon as the lower finger has reached a clearly palpable depression anterior to the temporomandibular joint and posterior to the masseter muscle (→ **ST-7**), the upper finger will be resting on **G.B.-3**, which is located directly superior to → **ST-7** in a shallow depression on the superior border of the zygomatic arch.

Needling

Vertically 0.3–0.5 cun, no strong stimulation. Caution: Branches of the temporal, transverse facial and masseteric arteries! Traditionally, deep needling is prohibited.

Actions/Indications

- Opens the channel, alleviates pain
- Eliminates Wind
- Benefits the ears

Special features

Meeting point with the T.B. and ST channels

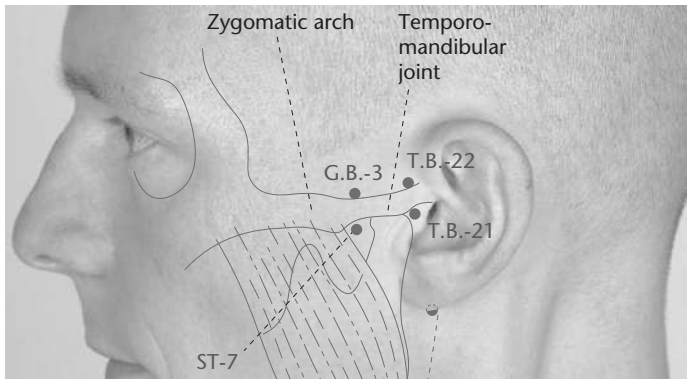

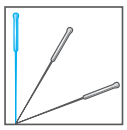
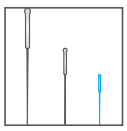

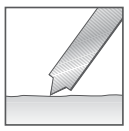
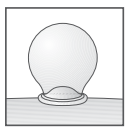

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 402

G.B.-4 *Jaw Serenity* HANYAN

Location

At the junction of the upper quarter and the second quarter of a line connecting → **ST-8** and → **G.B.-7**.

How to find

First, locate the two reference points: → **ST-8** (4.5 cun lateral to the midline and 0.5 cun within the anterior hairline, at the corner of the forehead) and → **G.B.-7** (in the depression on the level of the apex of the ear, within the circumauricular temporal hairline). Then divide the slightly curved line between these two points into quarters and locate **G.B.-4** at the junction of the upper quarter with the lower three quarters. Generally, **G.B.-4** is located on the temporal hairline and the anterior portion of the temporalis muscle, which can be felt when chewing.

Needling

0.3–1.5 cun transversely (subcutaneously), tangentially along the skull, in the direction of the occiput or towards the disorder/pain.

Actions/Indications

- Eliminates Wind and Heat
- Opens the channel

Special features

Meeting point with the T.B. and ST channel

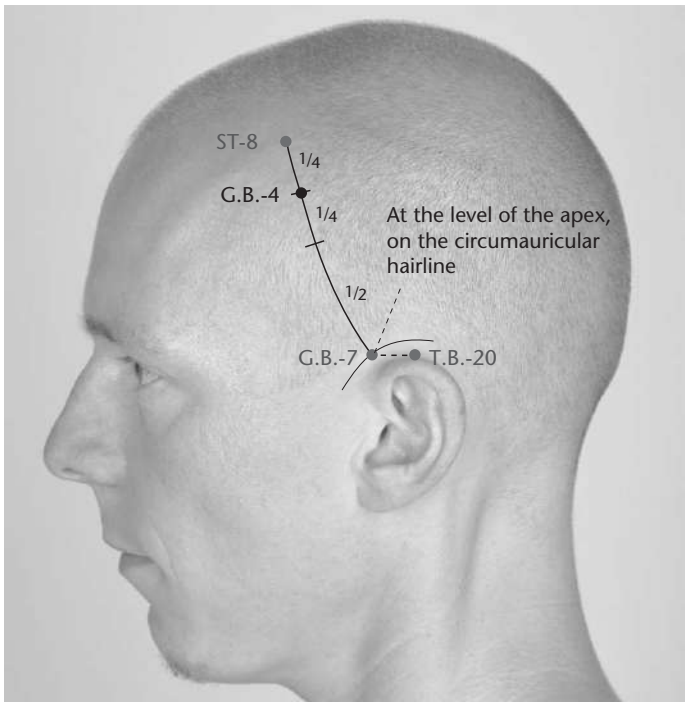

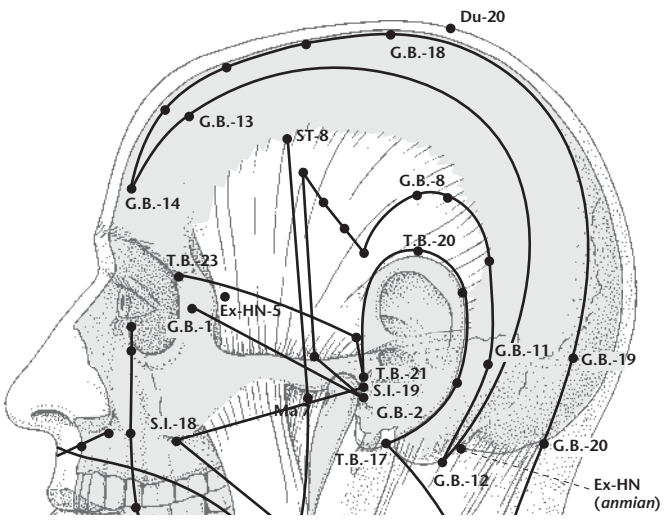

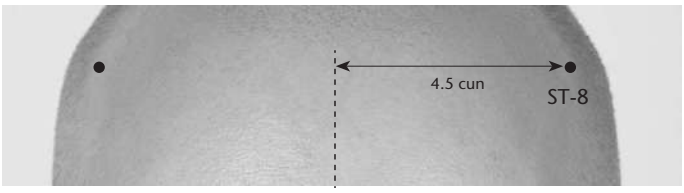

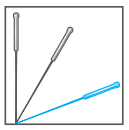
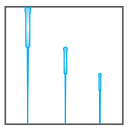

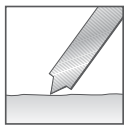

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 403

Suspended Skull XUANLU G.B.-5

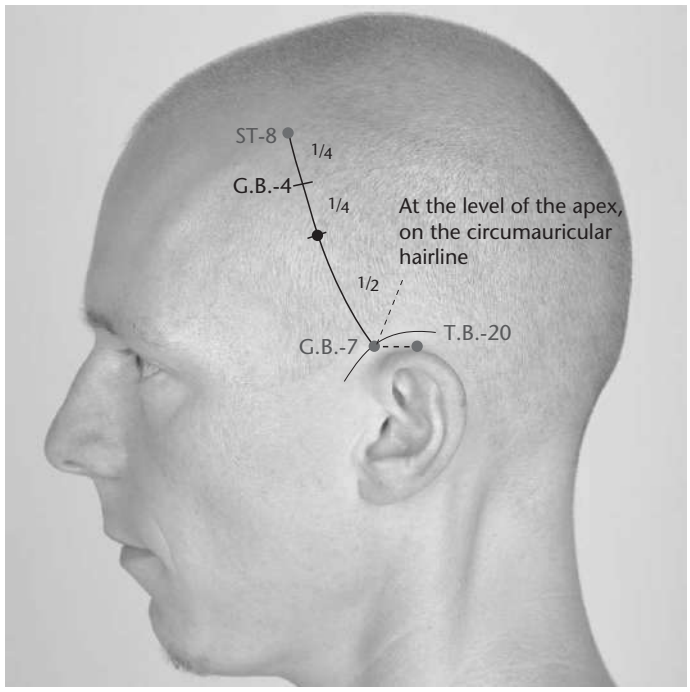

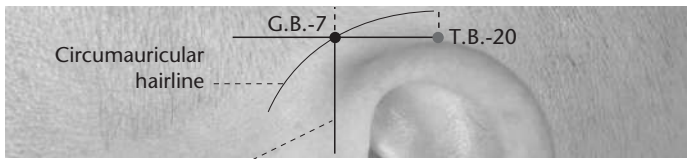

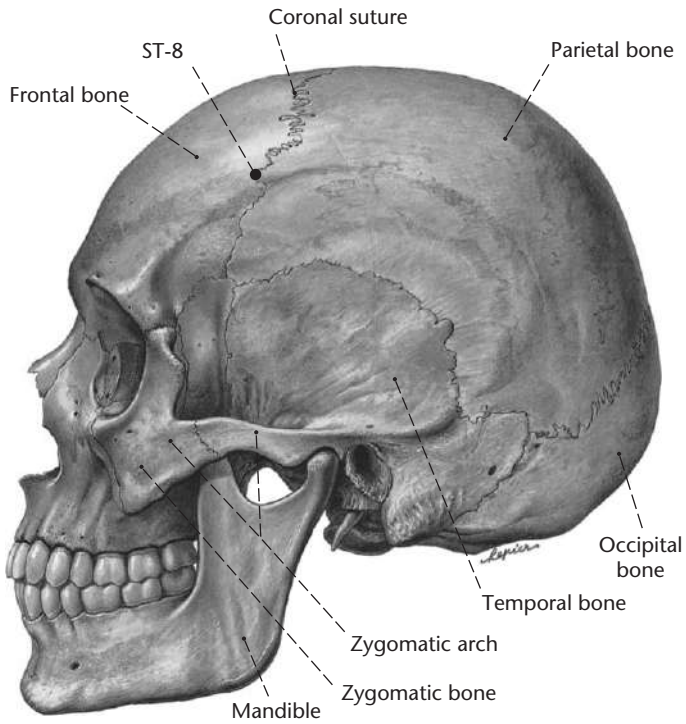

Location

At the junction of the 2nd and 3rd quarter of an imaginary line connecting ➞ **ST-8** and ➞ **G.B.-7**.

How to find

First, locate the two reference points: ➞ **ST-8** (4.5 cun lateral to the midline and 0.5 cun within the anterior hairline, at the corner of the forehead) and ➞ **G.B.-7** (in the depression on the level of the apex of the ear, within the circumauricular temporal hairline). Then divide the slightly curved line between these two points into quarters and locate **G.B.-5** at the junction of the upper two quarters with the lower two quarters. Generally, this point is located on the level of the parietal suture, just within the hairline.

Needling

0.3–1.5 cun transversely (subcutaneously), tangentially along the skull, in the direction of the occiput or towards the disorder/pain.

Actions/Indications

- Expels Wind and Heat
- Opens the channel

Special features

Meeting point with the T.B., L.I. and ST channels

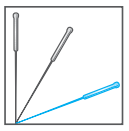
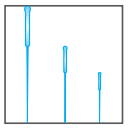

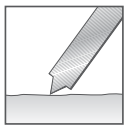

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 404

G.B.-6 *Suspended Hair* XUANLI

Location

At the junction of the lower quarter and the upper three quarters of a line connecting → **ST-8** and → **G.B.-7**.

How to find

First, locate the two reference points: → **ST-8** (4.5 cun lateral to the midline and 0.5 cun within the anterior hairline, at the corner of the forehead) and → **G.B.-7** (in the depression on the level of the apex of the ear, within the circumauricular temporal hairline). Then divide the slightly curved line between these two points into quarters and locate **G.B.-6** at the junction of the upper three quarters with the lower quarter.

Needling

0.3–1.5 cun transversely (subcutaneously) in the direction of the occiput or towards the disorder/pain.

Actions/Indications

- Expels Wind and Heat
- Opens the channel

Special features

Meeting point with the L.I., ST and T.B. channels

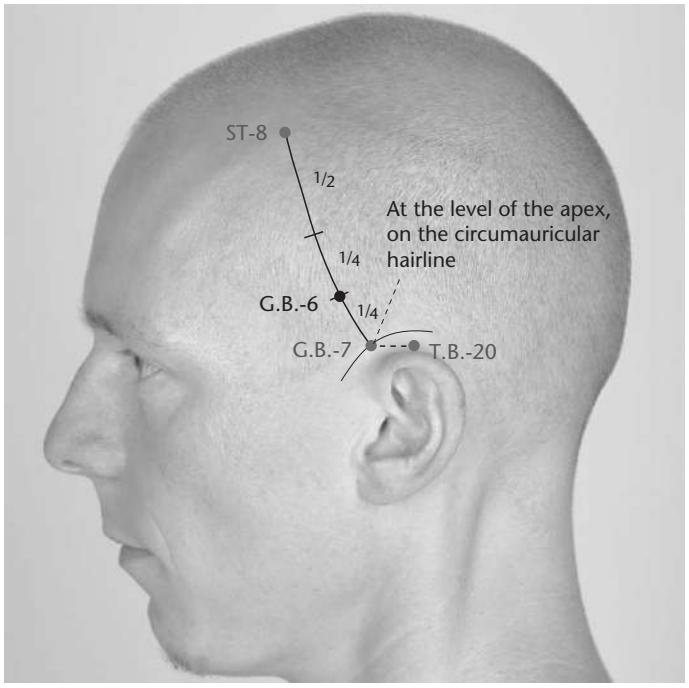

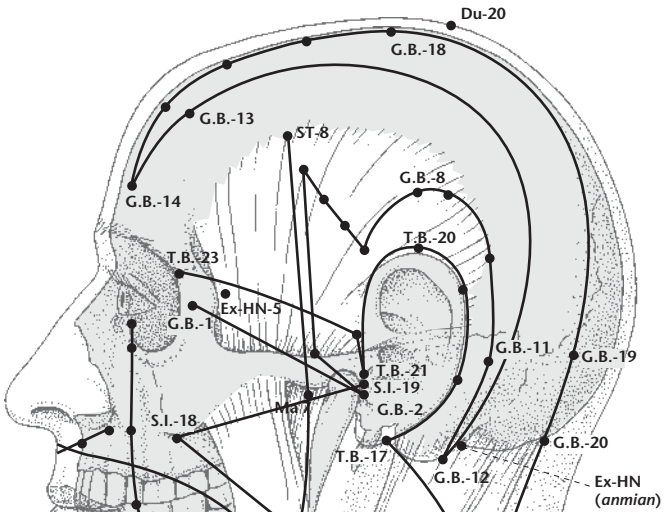

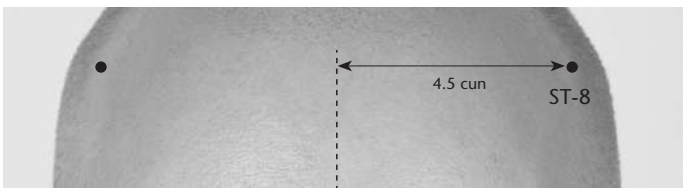

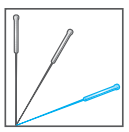
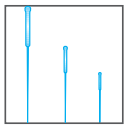

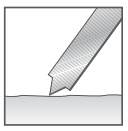

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 405

Crook of the Temple QUBIN G.B.-7

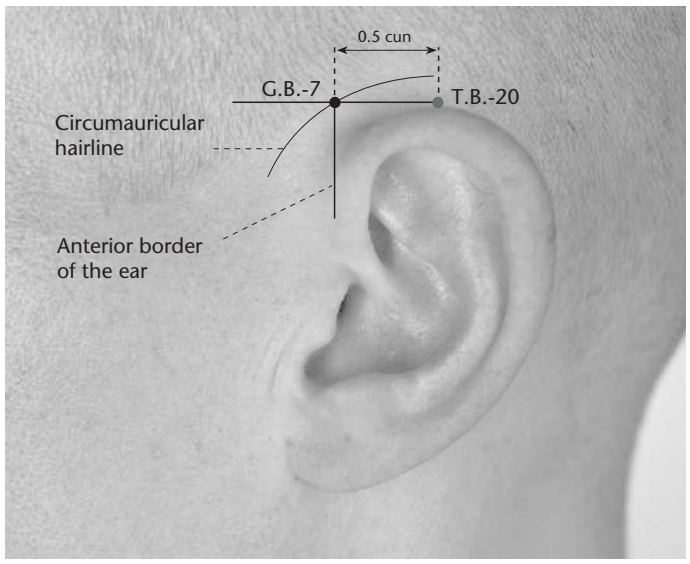

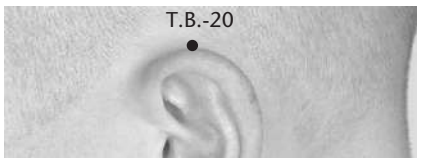

Location

In a depression on the level of the apex of the ear, within the circumauricular temporal hairline, approximately at the junction of a horizontal line through the apex of the ear and a vertical line along the posterior border of the temple anterior to the ear.

How to find

Locate **G.B.-7** by palpating for a small depression anterior to the apex of the ear at the circumauricular temporal hairline. The point is located approximately at a junction between a horizontal line through the apex of the ear and a vertical line along the posterior border of the temple anterior to the ear.

For orientation: → **T.B.-20** is located directly superior to the apex of the ear.

Needling

0.3–1.5 cun transversely (subcutaneously) in the direction of the occiput or towards the site of the disorder/pain.

Actions/Indications

- Expels Wind
- Benefits the mouth and jaw

Special features

Meeting point with the BL channel

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 406

G.B.-8 *Leading Valley* SHUAIGU

Location

1.5 cun directly superior to the apex of the ear, in a depression on the upper border of the temporalis muscle.

How to find

Locate the apex of the ear, which becomes more clearly defined by folding the auricle towards the anterior so that the posterior part of the upper helix covers its anterior part. From the apex of the ear, measure 1.5 cun (2 fingerbreadths) in a superior direction. There, the palpating finger will glide into a bony depression (**G.B.-8**) which is often sensitive to pressure. For orientation: When a chewing movement is made, this can just about be felt at this point.

→ **G.B.-9** is located on the same level, 0.5 cun dorsal to **G.B.-8**.

Needling

0.3–1.5 cun transversely (subcutaneously), mostly from anterior to posterior or towards the site of the pain.

Actions/Indications

- Expels Wind, benefits the head and ears, alleviates pain, harmonises the diaphragm and Stomach

Special features

Meeting point with the BL channel. Important local point for parietal and temporal headaches.

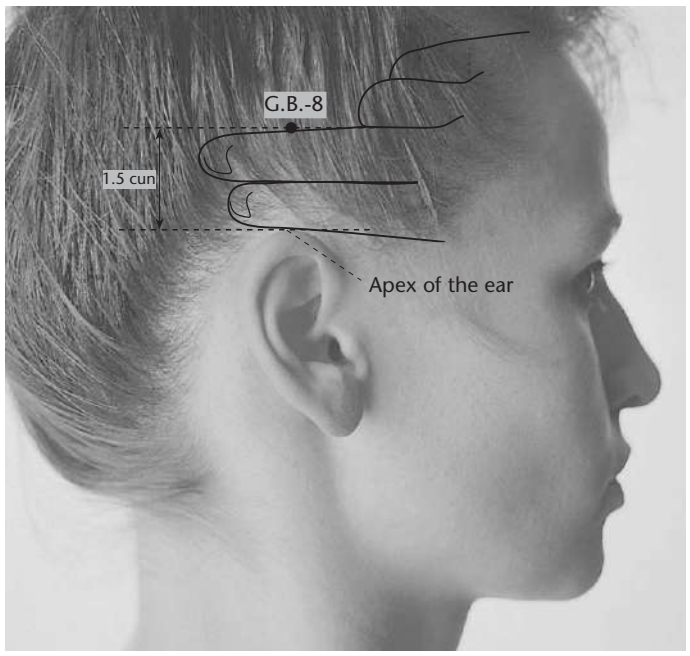

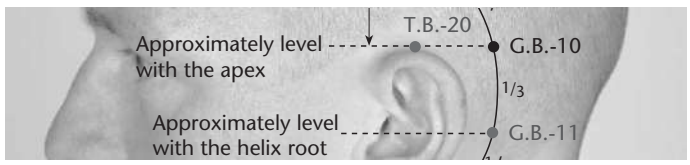

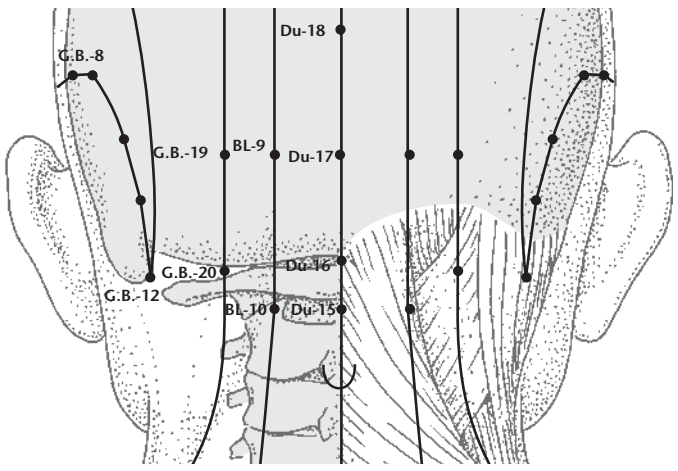

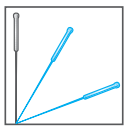
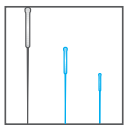

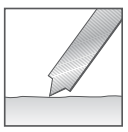

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 407

Heavenly Rushing TIANCHONG G.B.-9

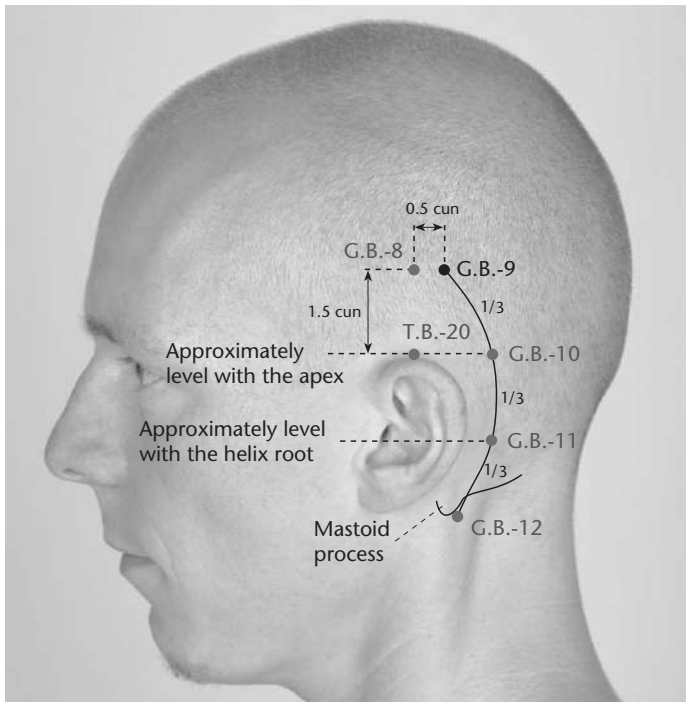

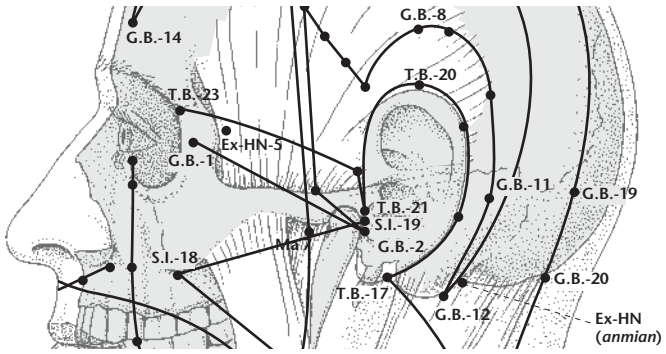

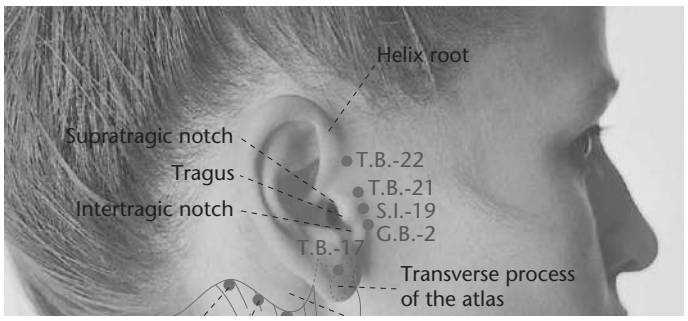

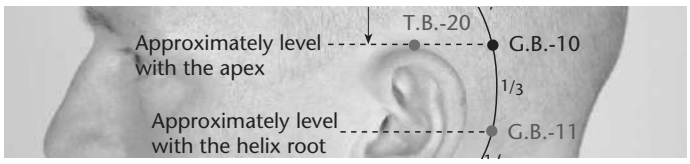

Location

1.5 cun directly superior to the apex of the ear (→ **T.B.-20**) and 0.5 cun posterior to → **G.B.-8**. The point is approximately superior to the posterior margin of the ear.

How to find

First, locate → **G.B.-8** 1.5 cun superior to the apex of the ear. The latter becomes more clearly defined by folding the auricle towards the anterior so that the posterior part of the upper helix covers its anterior part. From → **G.B.-8**, measure 0.5 cun in a posterior direction and there locate **G.B.-9**. Similar to → **G.B.-8**, **G.B.-9** is located in a slight 'dip' in the bone.

Both **G.B.-9** and → **G.B.-12** (in the depression posterior and inferior to the mastoid process) are reference points for a curved line that runs approximately parallel to the posterior margin of the ear within the hairline. When dividing this line into thirds, → **G.B.-10** and → **G.B.-11** are located at the junctions of the thirds.

Needling

0.5–1.5 cun transversely (subcutaneously) towards the occiput or the site of the pain.

Actions/Indications

- Expels Wind and Heat
- Opens the *luo*-connecting vessels
- Calms the *shen*

Special features

Meeting point with the BL channel; Window of Heaven point according to some authors.

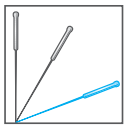
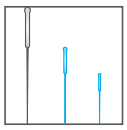

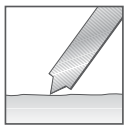
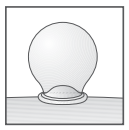

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 408

G.B.-10 *Floating White* FUBAI

Location

Posterior to the ear, at the junction of the upper third with the two lower thirds of the curved line connecting → **G.B.-9** and → **G.B.-12**.

How to find

First, locate → **G.B.-9** (1.5 cun superior to the apex of the ear and 0.5 cun in a posterior direction). Then locate → **G.B.-12** in a depression posterior and inferior to the mastoid process (→ 3.1.4). These two points are the endpoints of a curved line that runs approximately parallel to the posterior margin of the ear within the hairline. Divide this line into thirds and locate **G.B.-10** at the junction of the upper and the middle third, where often a small 'dip' can be felt on the bone. **G.B.-10** tends to be on the level of the apex of the ear.

→ **T.B.-20** is located directly above the apex of the ear.

Needling

Transversely (subcutaneously) 0.5–0.8 cun

Actions/Indications

- Expels Wind
- Opens the *luo*-connecting vessels

Special features

Meeting point with the BL channel

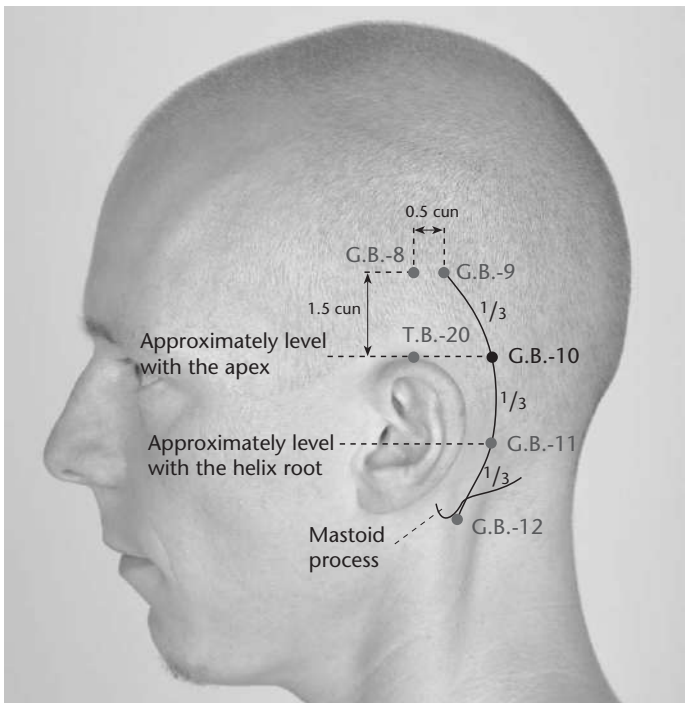

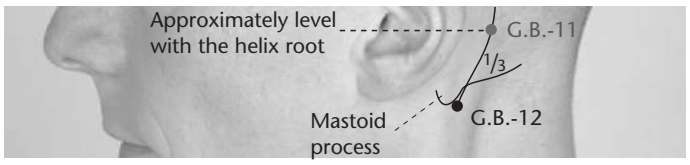

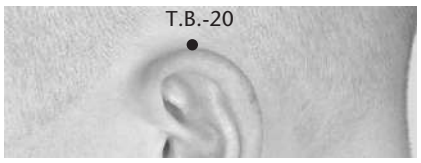

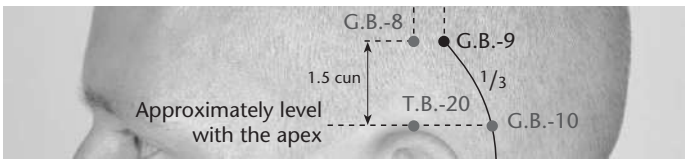

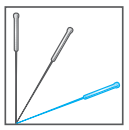

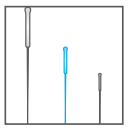

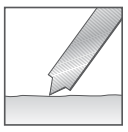

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 409

Head Portal Yin TOUQIAOYIN G.B.-11

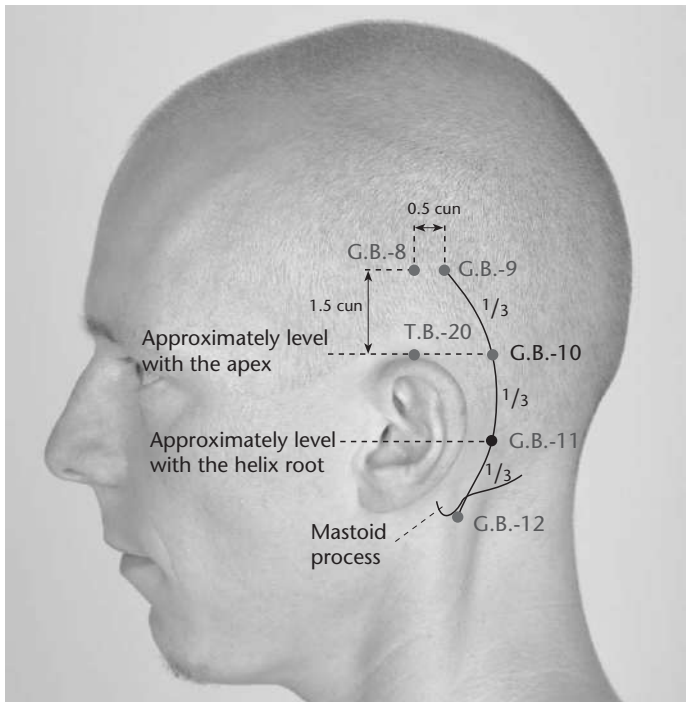

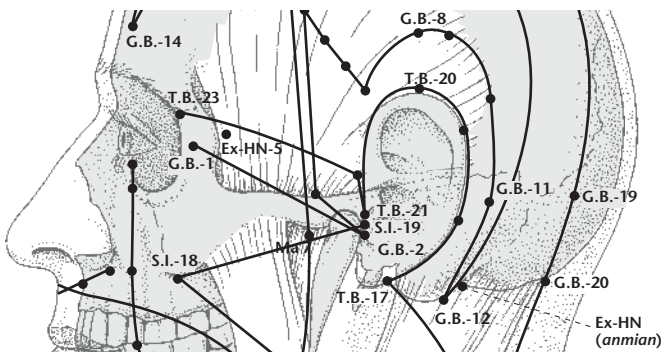

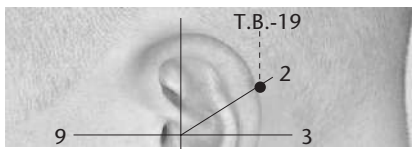

Location

Posterior to the ear, at the junction of the lower third with the two upper thirds of the curved line connecting → **G.B.-9** and → **G.B.-12**.

How to find

First, locate → **G.B.-9** (1.5 cun superior to the apex of the ear and 0.5 cun in a posterior direction). Then locate → **G.B.-12** in a depression posterior and inferior to the mastoid process (→ 3.1.4). These two points are the endpoints of a curved line that runs approximately parallel to the posterior margin of the ear within the hairline. Divide this line into thirds and locate **G.B.-11** at the junction of the middle and the lower third. At this point, a small 'dip' often can be felt in the bone. For reference: **G.B.-11** is located at the midpoint between a line connecting → **G.B.-10** and → **G.B.-12** and is generally on the level of the root of the helix.

→ **T.B.-19** is located on the same level, directly posterior to the helix.

Needling

Transversely (subcutaneously) 0.5–0.8 cun

Actions/Indications

- Expels Wind and Damp-Heat
- Opens the ears and the eyes
- Moves the (Liver) Qi and Blood

Special features

Meeting point with the BL channel, also with the S.I. and T.B. channels according to some authors.

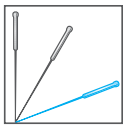
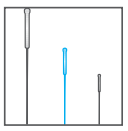

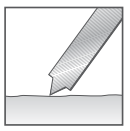
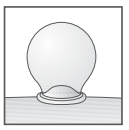

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 410

G.B.-12 *Mastoid Process* WANGU

Location

In a depression directly posterior and inferior to the mastoid process.

How to find

First, locate the mastoid process posterior to the ear (→ 3.1.4). This can be palpated at the junction of the cranium and the neck as a cone-shaped, bony structure. With the palpating finger, find the inferior aspect of the mastoid process and locate **G.B.-12** on its lower border posterior to the tip of the process.

→ **T.B.-17** is located more anteriorly, in the depression posterior to the earlobe, between the mastoid process and the mandible.

Needling

0.5–1 cun obliquely in an inferior direction.

Actions/Indications

- Expels Wind, Heat and Dampness/Phlegm
- Opens the ears
- Calms the *shen*

Special features

Meeting point with the BL channel

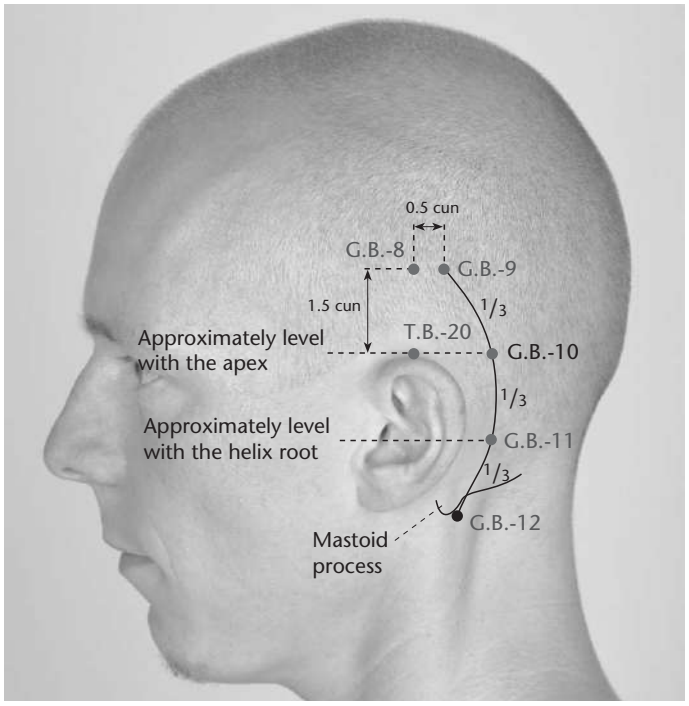

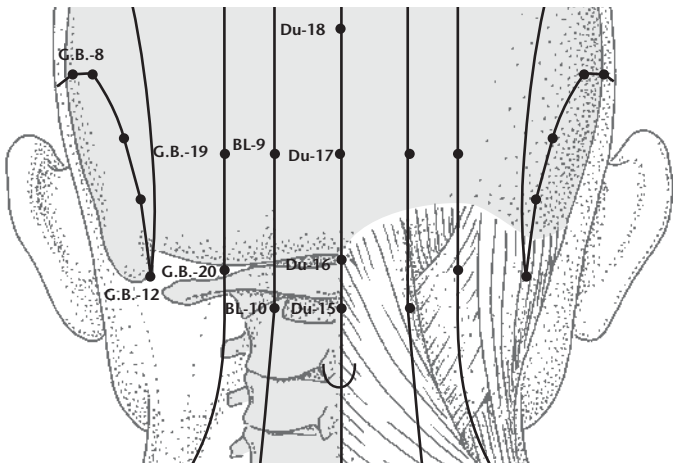

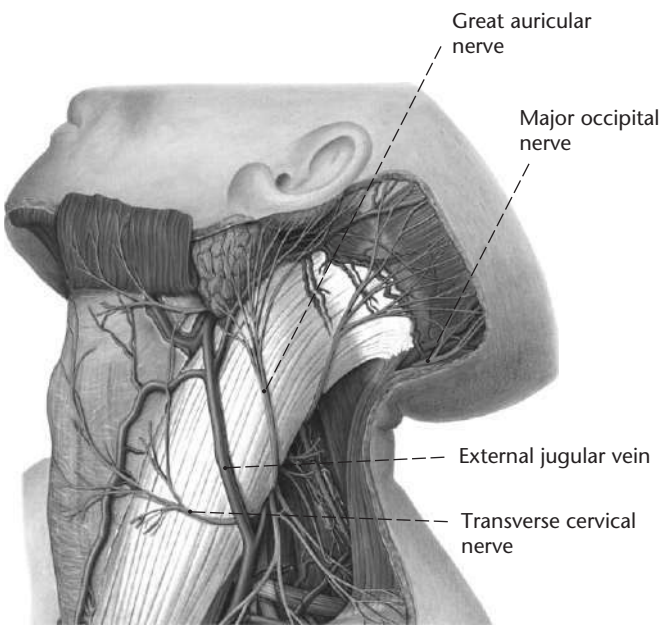

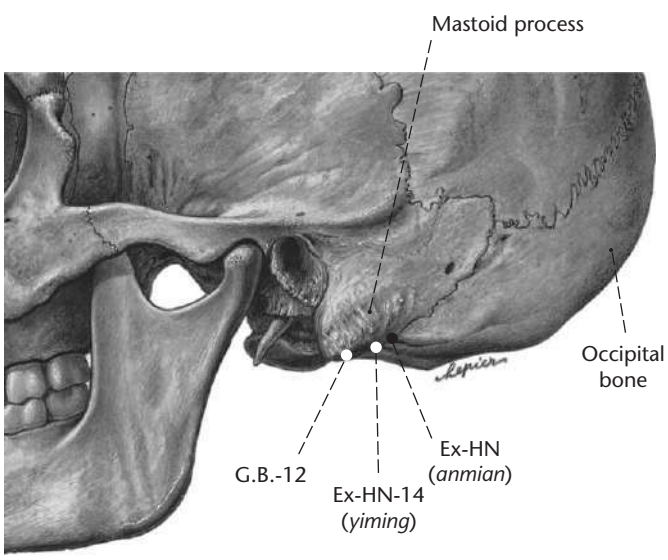

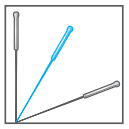
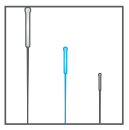

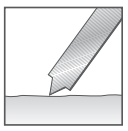

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 411

Spirit Root BENSHEN G.B.-13

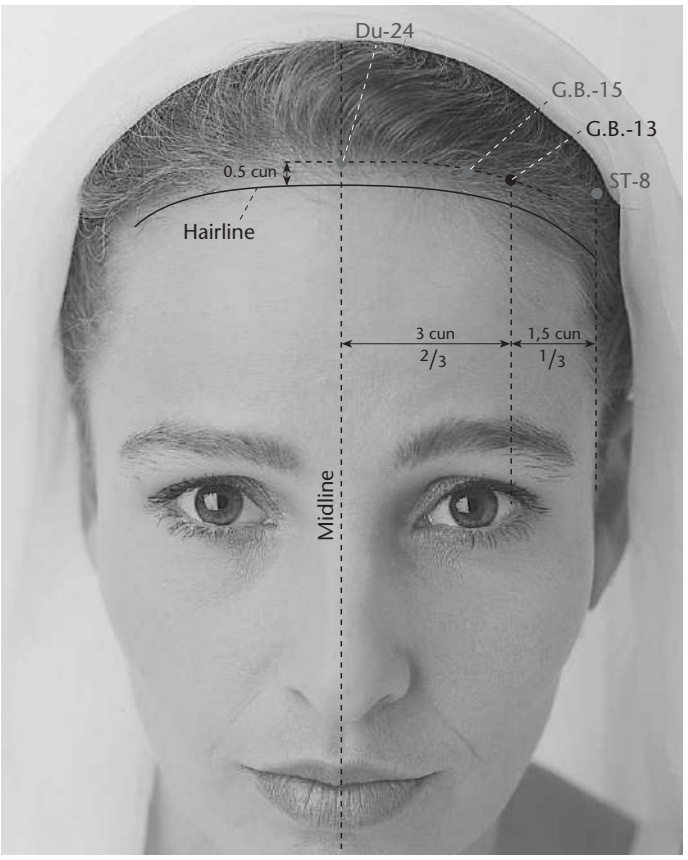

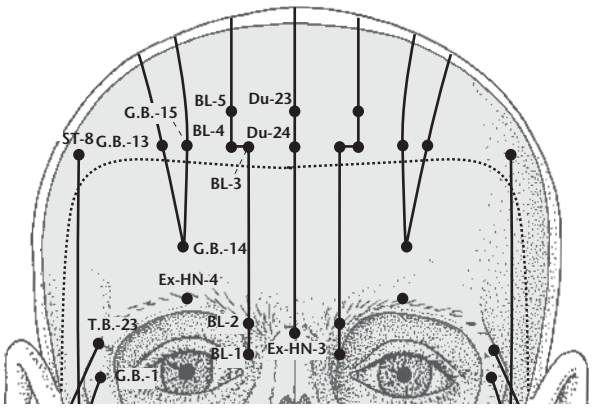

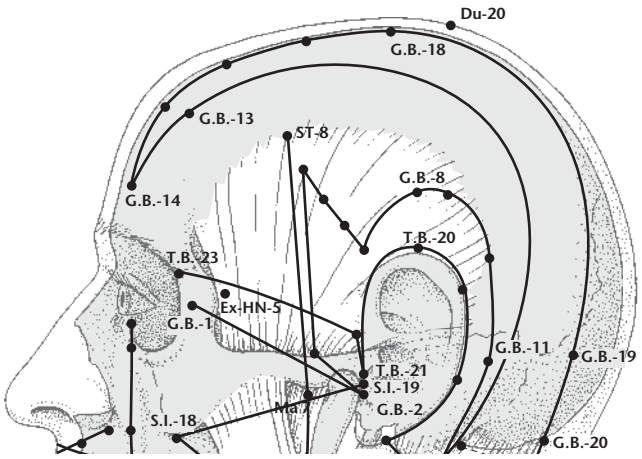

Location

3 cun lateral to → **Du-24** (on the midline, 0.5 cun superior to the anterior hairline).

How to find

First, locate the anterior hairline (→ 3.1.1, with hairloss: when the patient frowns, the anterior hairline is marked by the border between the creased forehead and the smooth skin above) and there locate → **Du-24** on the midline and 0.5 cun within the hairline. **G.B.-13** is located on the same level, 3 cun lateral to the midline, measured in proportional cun based on the distance → **ST-8** to → **Du-24** (=4.5 cun, 2.2). Divide this distance into thirds and locate **G.B.-13** one third of the distance from **ST-8**. Located at the same level (0.5 cun within the anterior hairline) are → **Du-24/BL-3/BL-4/G.B.-15/ST-8** (on the midline/superior to the inner canthus of the eye/1.5 cun lateral to the midline/on the pupil line or 2.25 cun lateral to the midline/on the corner of the forehead).

Needling

0.5–0.8 cun transversely (subcutaneously) towards the occiput.

Actions/Indications

- Expels Wind
- Calms the *shen*
- Benefits the eyes

Special features

Meeting point with the *yang wei mai*

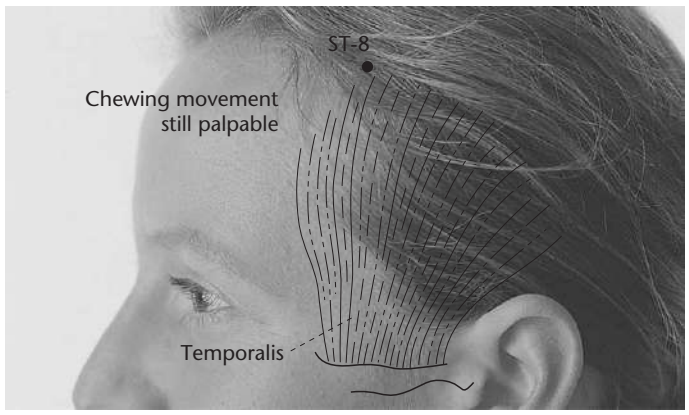

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 412

G.B.-14 *Yang White* YANGBAI

Location

With the patient looking straight ahead, on the pupil line, approximately 1 cun superior to the midpoint of the eyebrow, at the junction of the frontal eminence and the superciliary arch.

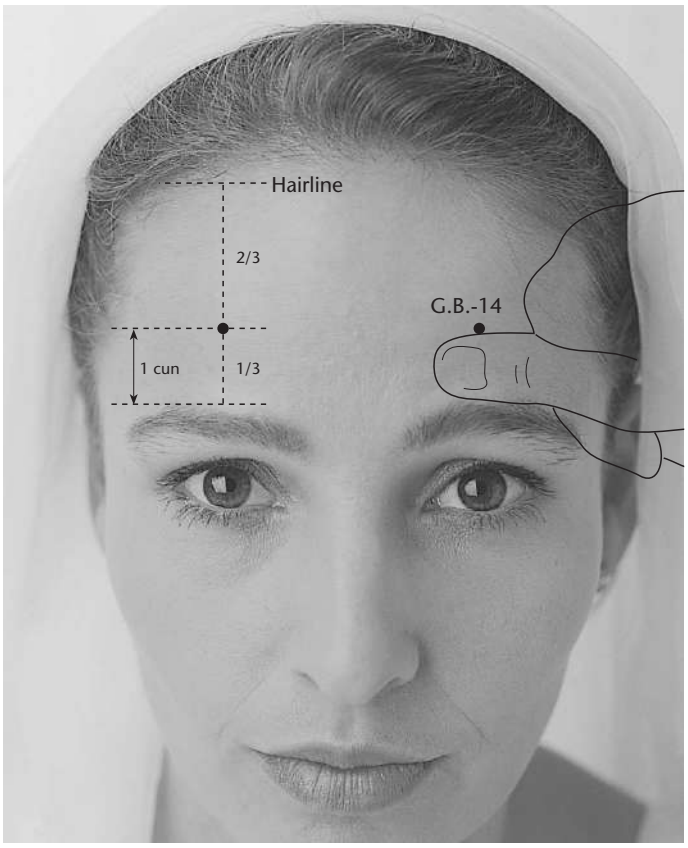

How to find

On the pupil line, palpate from the anterior hairline in an inferior direction, past the frontal eminence, and locate **G.B.-14** at the deepest point between the frontal eminence and the superciliary arch. The distance between the midpoint of the eyebrow and the anterior hairline (→ 3.1.1) is 3 proportional cun (→ 2.2). **G.B.-14** is located one third of the distance or 1 cun superior to the midpoint of the eyebrow. With the patient looking straight ahead, **G.B.-14** is located on the pupil line.

Needling

0.3–1 cun transversely (subcutaneously) towards the centre of the eyebrow or towards the site of the pain. The pinching-skin method might be used for needling: pinch the skin between the thumb and index finger so that a skin fold forms. Insert the needle subcutaneously into this fold, directing it towards the site of the pain.

Actions/Indications

- Expels (internal and external) Wind and Wind-Heat, benefits the head, alleviates pain
- Benefits the eyes

Special features

Meeting point with the *yang wei mai*, also with the T.B., L.I. and ST channels according to some authors. Important local point for frontal headache regardless of the pathology.

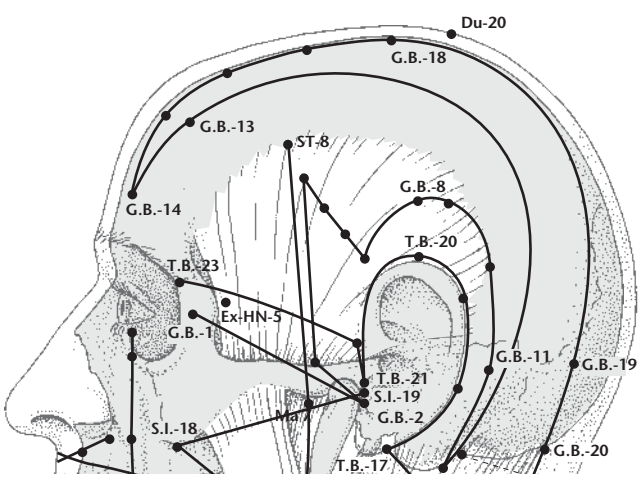

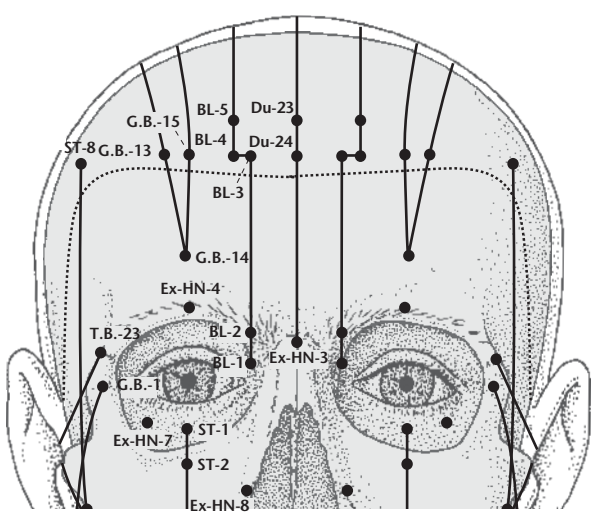

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 413

Head Governor of Tears TOULINQI G.B.-15

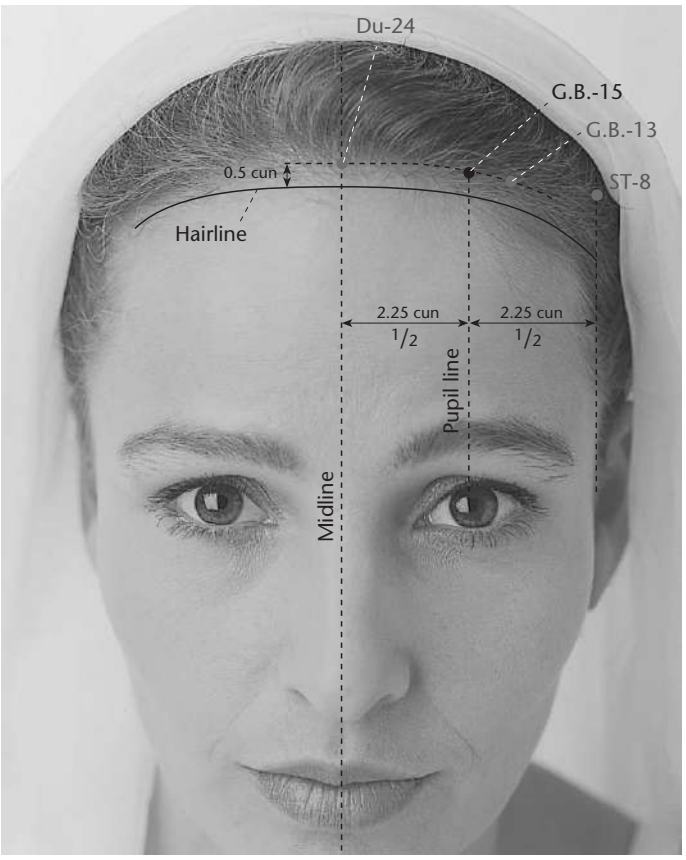

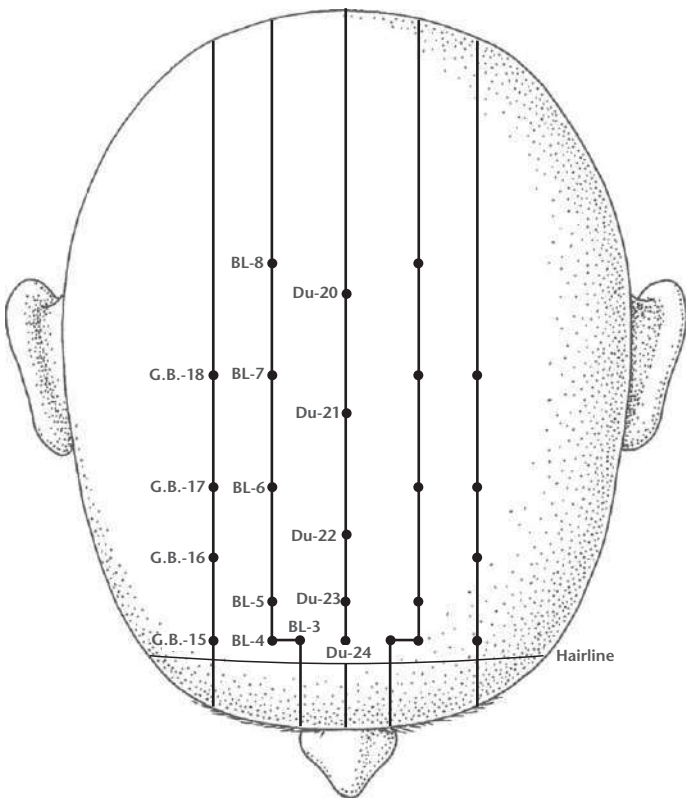

Location

With the patient looking straight ahead, directly superior to the pupil and 0.5 cun superior to the anterior hairline.

How to find

First, locate the anterior hairline (→ 3.1.1, with hairloss: when the patient frowns, the anterior hairline is marked by the border between the creased forehead and the smooth skin above). With the patient looking straight ahead, **G.B.-15** is located 0.5 cun superior to the anterior hairline and directly superior to the pupil. **G.B.-15** can also be described as being located midway between → **Du-24** (on the midline) and → **ST-8** (4.5 cun lateral to the midline, on the corner of the forehead).

Located on the same level (0.5 cun within the anterior hairline) are → **Du-24/BL-3/BL-4/G.B.-13/ST-8** (on the midline/ superior to the inner canthus of the eye/1.5 cun lateral to the midline/3 cun lateral to the midline or one third of the distance from → **ST-8** (on the corner of the forehead) to → **Du-24**.

Needling

Transversely (subcutaneously) 0.3–0.5 cun

Actions/Indications

- Expels Wind
- Opens the eyes and nasal passages
- Calms the *shen*

Special features

Meeting point with the *yang wei mai*, also with the BL channel according to some authors.

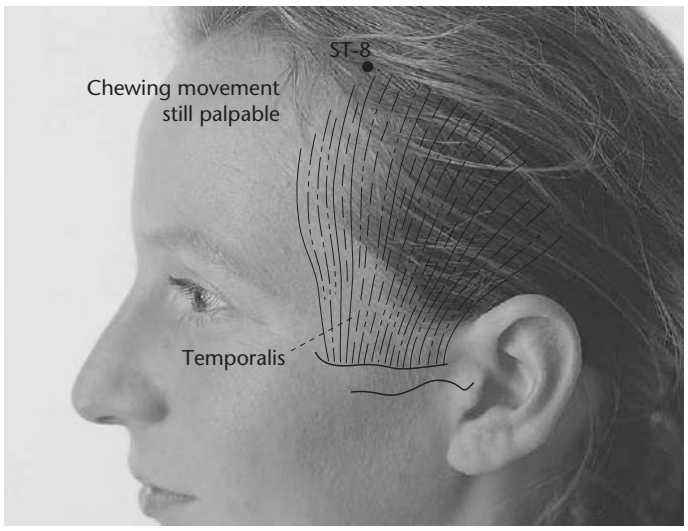

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 414

G.B.-16 *Window of the Eye* MUCHUANG

Location

1.5 cun superior to the anterior hairline, on the pupil line or 2.25 cun lateral to the midline (midway between → **Du-24** and → **ST-8**).

How to find

Locate the anterior hairline (→ 3.1.1, with hairloss: when the patient frowns, the anterior hairline is marked by the border between the creased forehead and the smooth skin above). With the patient looking straight ahead, **G.B.-16** is located 1.5 cun superior to the anterior hairline, on a vertical line through the centre of the pupil. This line is located 2.25 cun lateral to the midline (this corresponds to the midpoint of the distance between → **Du-24** and → **ST-8**). For further orientation: The distance between the anterior hairline and → **Du-20** (at the junction of the vertical midline and a line connecting the apices of the ears) is 5 cun. → **G.B.-16** is located 3.5 cun anterior to → **Du-20**, on the curved line connecting → **G.B.-15** and → **G.B.-20**.
→ **G.B.-15** is located on the same vertical line, but 0.5 cun superior to the anterior hairline.

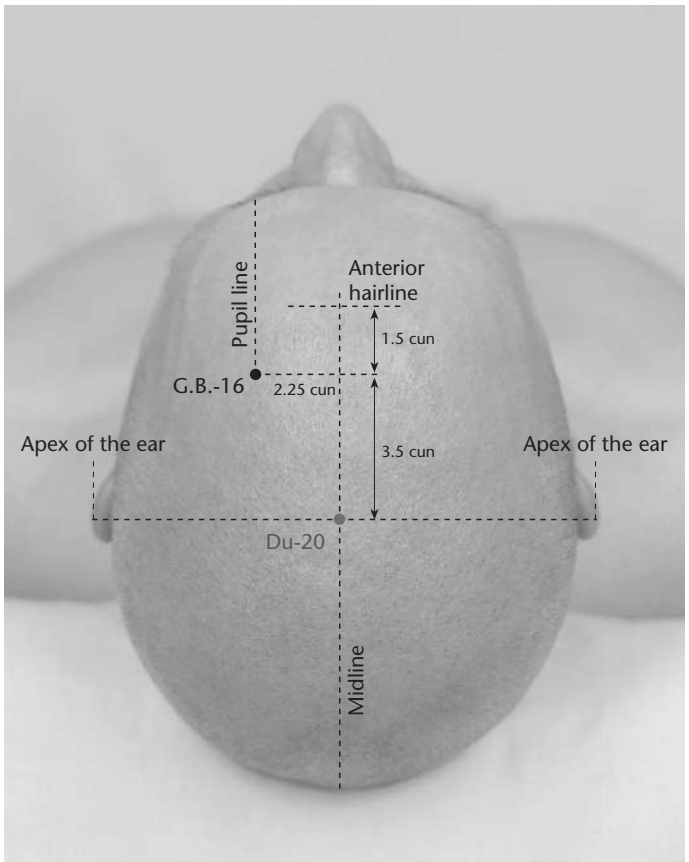

Needling

0.3–1.5 cun transversely (subcutaneously) towards the occiput or the site of the disorder/pain.

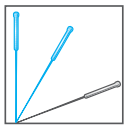
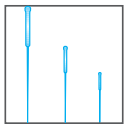

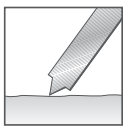

Actions/Indications

- Expels pathogenic factors (especially Wind) from the head and eyes
- Opens the *luo*-connecting vessels

Special features

Meeting point with the *yang wei mai*

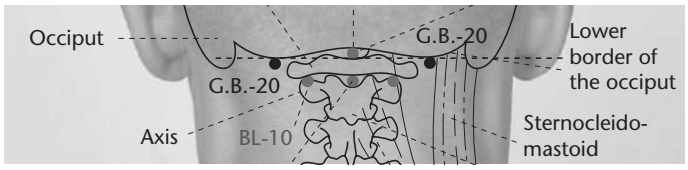

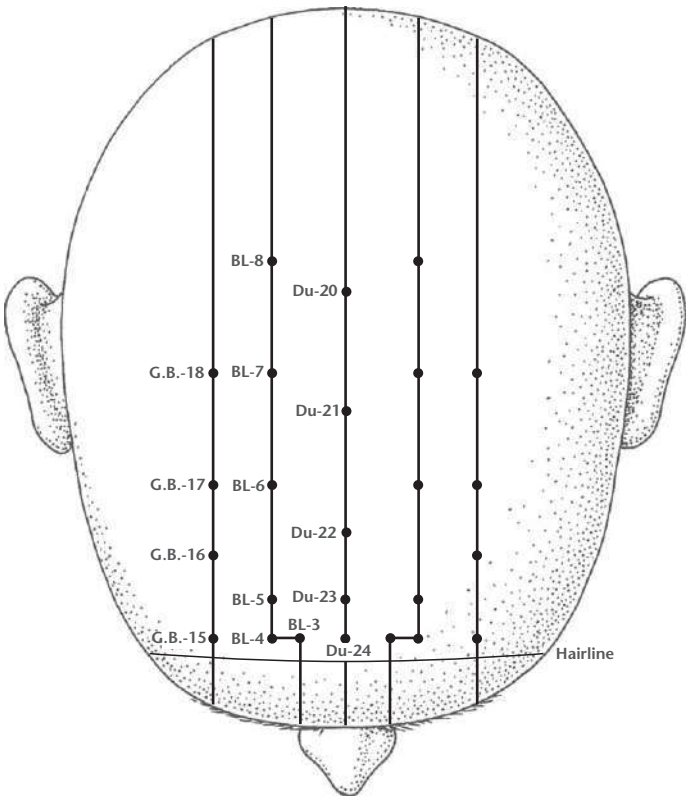

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 415

Upright Nutrition ZHENGYING G.B.-17

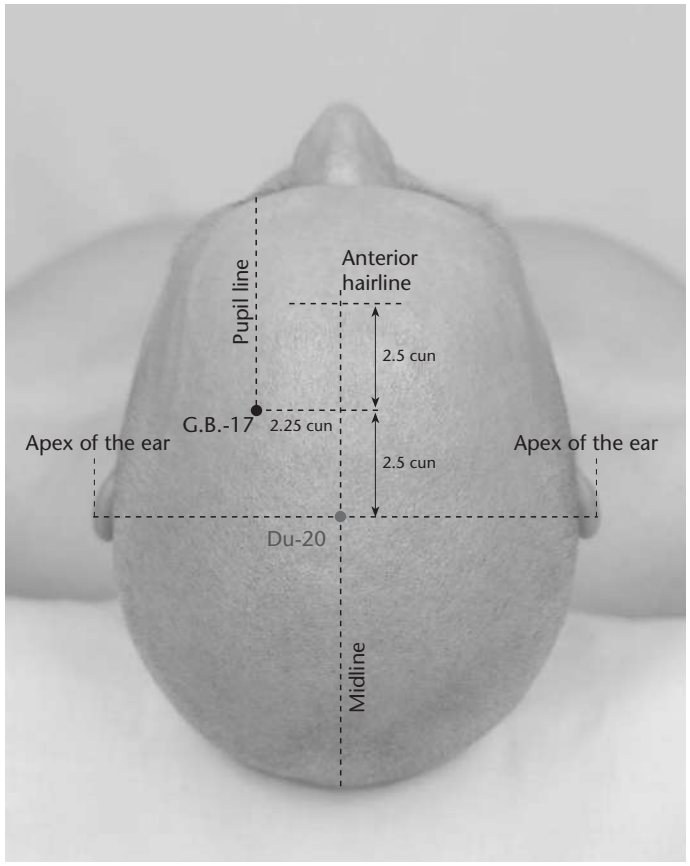

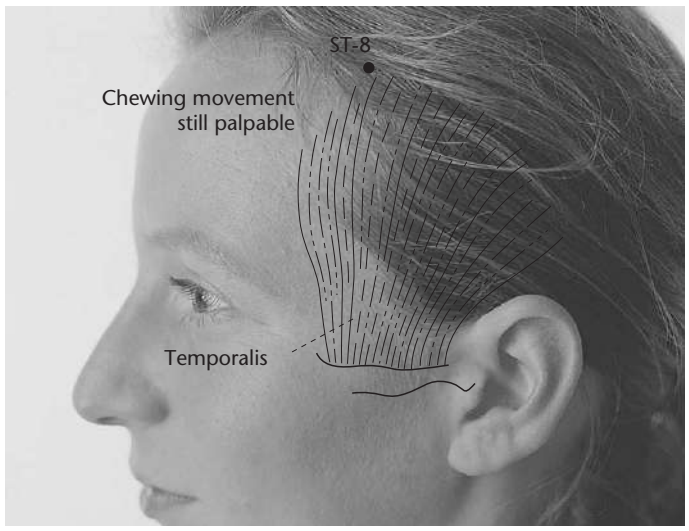

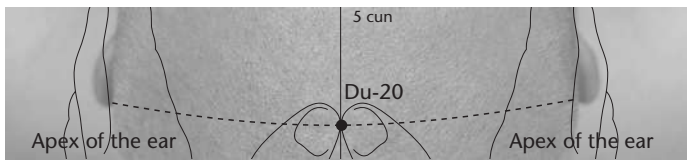

Location

2.5 cun superior to the anterior hairline and 2.25 cun lateral to the midline.

How to find

First, on the midline, locate the anterior hairline (→ 3.1.1, with hairloss: when the patient frowns, the anterior hairline is marked by the border between the creased forehead and the smooth skin above) and → **Du-20** (at the junction of the vertical midline and a line connecting the apices of the ears). The distance between these two points is 5 proportional cun (→ 2.2). Locate the mid-point (for example by using the spreading hands technique → 2.3.3). **G.B.-17** is located on the curved line connecting → **G.B.-15** and → **G.B.-20** (an extension of the pupil line, 2.25 cun lateral to the midline).

→ **BL-6** is also located 2.5 cun anterior to → **Du-20**, but 1.5 cun lateral to the midline.

Needling

0.3–1.5 cun transversely (subcutaneously) towards the occiput or the site of the disorder/pain.

Actions/Indications

- Expels Wind
- Opens the *luo*-connecting vessels
- Descends counterflow Stomach Qi

Special features

Meeting point with the *yang wei mai*

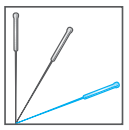
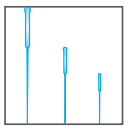

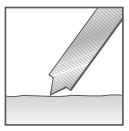
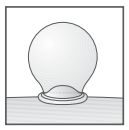

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 416

G.B.-18 *Support Spirit* CHENGLING

Location

4 cun superior to the anterior hairline or 1 cun anterior to → **Du-20** and 2.25 cun lateral to the midline.

How to find

First, locate → **Du-20** (at the junction of the vertical midline and a line connecting the apices of the ears, distance to the anterior hairline = 5 cun). Next, locate **G.B.-18** 1 cun anterior to → **Du-20** on the curved line connecting → **G.B.-15** and → **G.B.-20** (= extension of the pupil line, midway between → **ST-8** and → **Du-24**).
→ **BL-7** is located on the same level (1 cun anterior to → **Du-20**), but 1.5 cun lateral to the midline.

Needling

0.3–1.5 cun transversely (subcutaneously) towards the occiput or the site of the disorder/pain.

Actions/Indications

- Expels Wind, especially from the nose

Special features

Meeting point with the *yang wei mai*

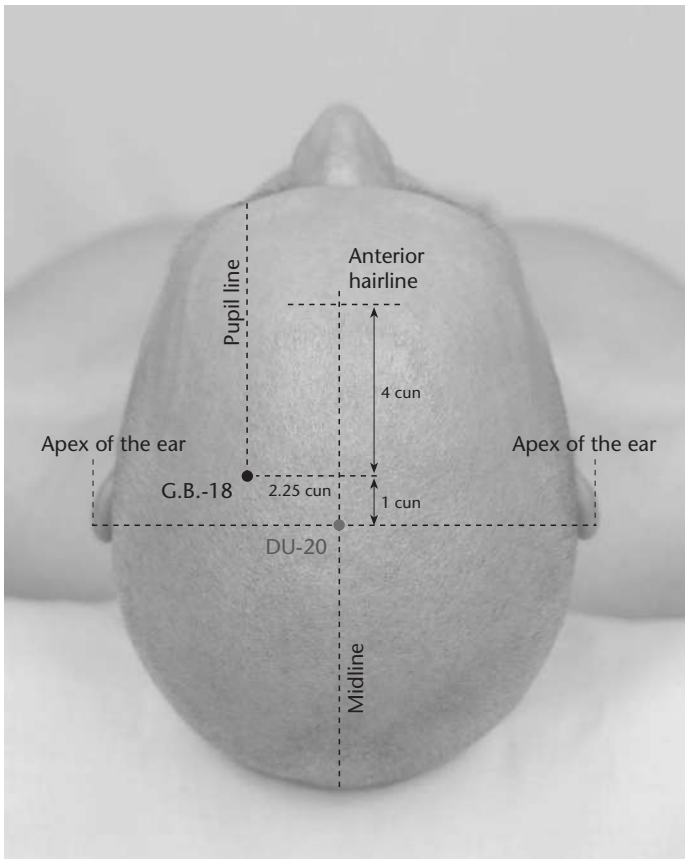

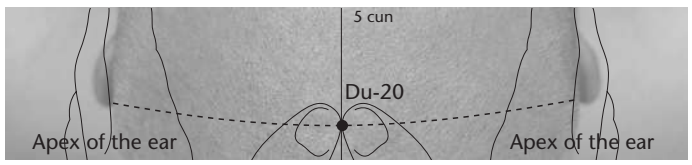

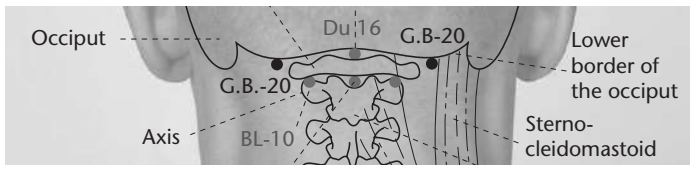

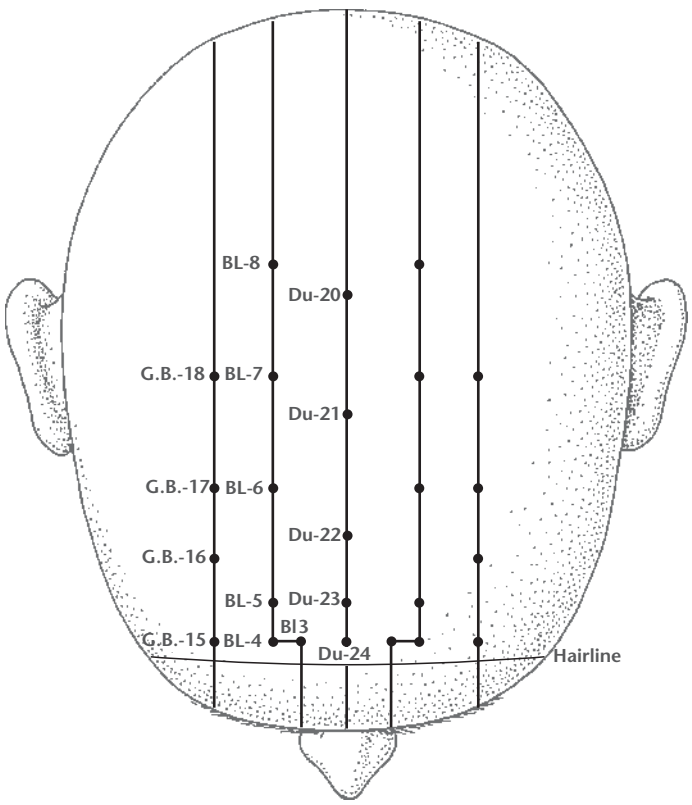

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 417

Brain Hollow NAOKONG G.B.-19

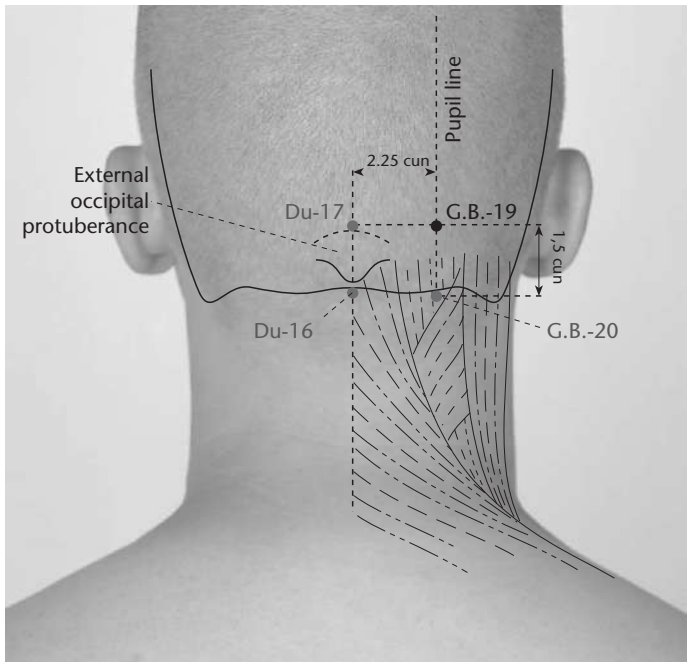

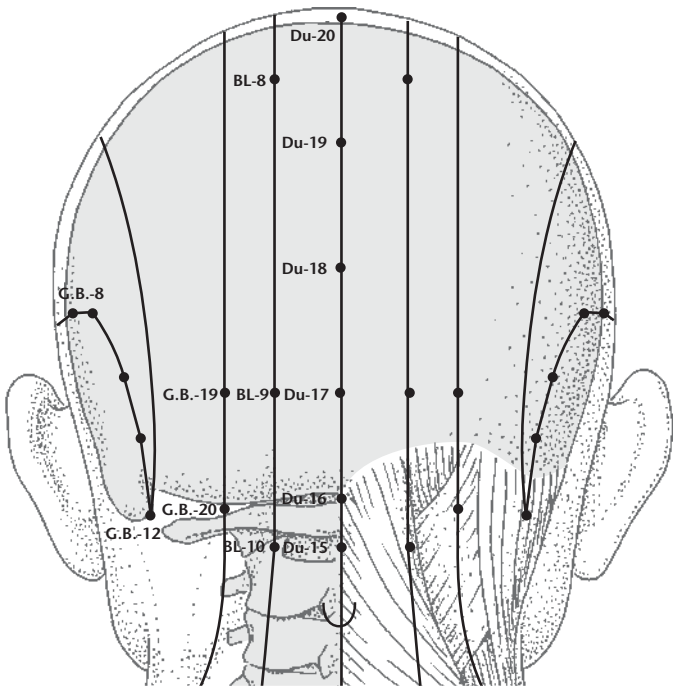

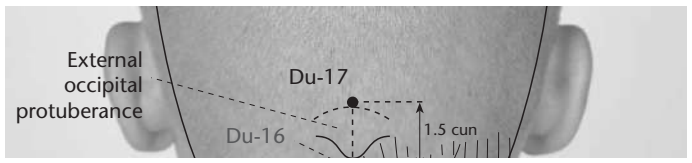

Location

On the posterior aspect of the head, at the upper border of the external occipital protuberance (→ **Du-17**) and 2.25 cun lateral to the midline.

How to find

First, locate the external occipital protuberance (→ 3.1.5), which can be palpated as a flat protruding area on the midline. Next, find → **Du-17** in the depressoin just superior to the protuberance. From there, measure 2.25 cun in a lateral direction and locate **G.B.-19** on the curved line connecting → **G.B.-15** and → **G.B.-20**, an extension of the pupil line (→ 2.25 cun lateral to the midline or midway between → **Du-24** and → **ST-8**). **For orientation: G.B.-19** is located approximately 2.5 cun superior to the posterior hairline (→ 3.1.5) and approximately 1.5 cun superior to → **G.B.-20** (on the lower border of the occiput, between the origins of the sternocleidomastoid and trapezius muscles).

Needling

0.3–1.5 cun transversely (subcutaneously) towards the occiput or the site of the disorder/pain.

Actions/Indications

- Expels Wind
- Clears the sensory orifices
- Opens the *luo*-connecting vessels

Special features

Meeting point with the *yang wei mai*

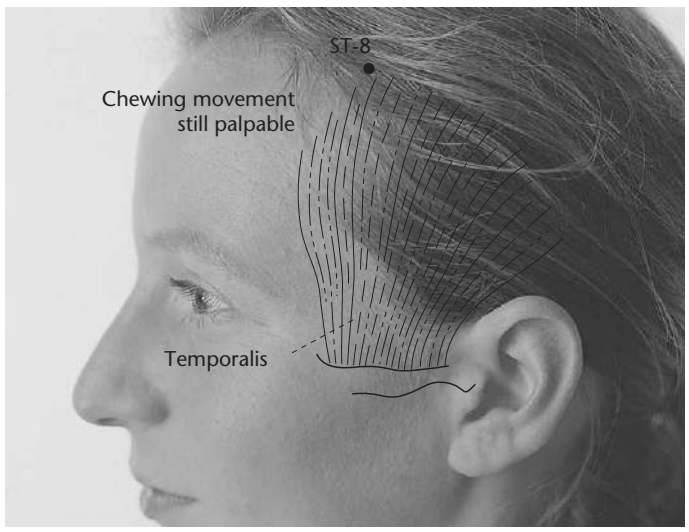

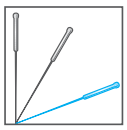
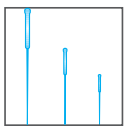

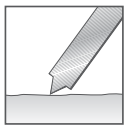

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 418

G.B.-20 *Wind Pool* FENGCHI

Location

At the lower border of the occipital bone, in the depression between the origins of the sternocleidomastoid and trapezius muscles.

How to find

Patient's position: Prone, sitting or supine (supported by a pillow, so that the occipital region is accessible). Starting at the midline, glide with the palpating finger along the lower border of the occiput, crossing the bulge of the origin of the trapezius muscle, until you reach a depression the size of a finger pad. Locate **G.B.-20** in its centre.

Located on the same level is → **Du-16**. → **BL-10** is located more medially and inferiorly.

Needling

Needle with the patient's head bent forward and the tip of the needle pointing inferiorly. Depending on the position of the head, insertion towards the tip of the nose or the contralateral orbit, 0.5–1.2 cun. Caution: In slim patients, do not needle deeper than 2 cm (the vertebral artery is located at a depth of approximately 4 cm).

Actions/Indications

- Eliminates Wind, benefits the head, clears the sensory organs
- Opens the channel
- With tonifying needle technique: strengthens the Marrow and the Brain (according to Maciocia)

Special features

Meeting point with the T.B. channel*, *yang wei mai, yang qiao mai*. Major point for all 'Wind disorders', very important point for disorders of the head and eyes.

*According to Deadman et al 1998.

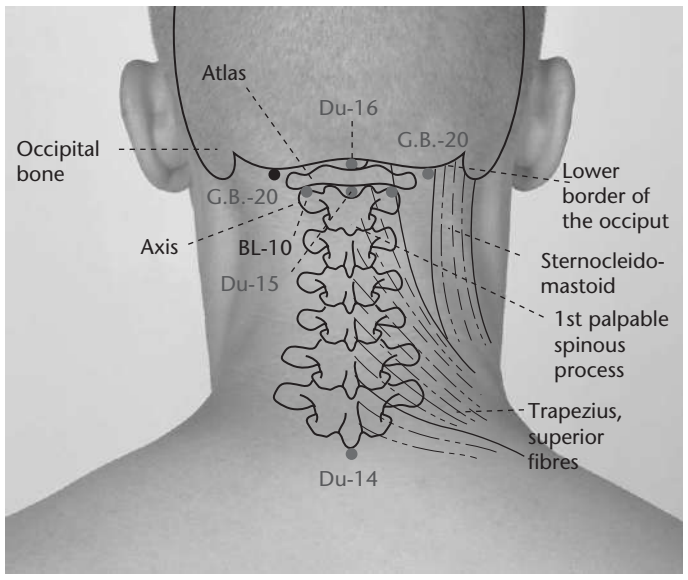

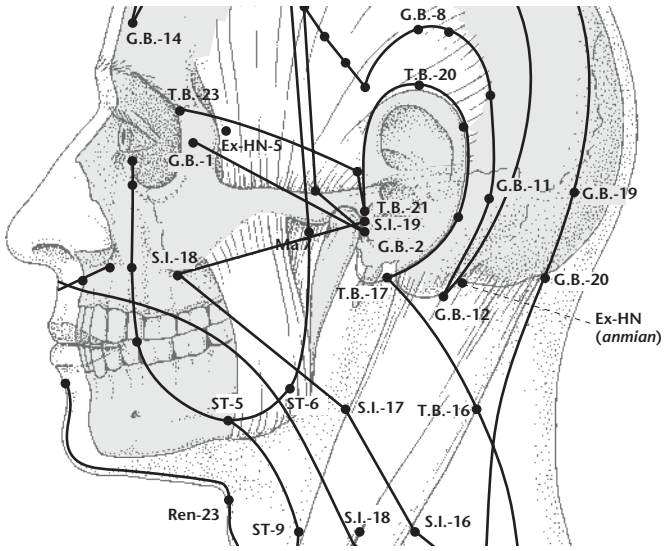

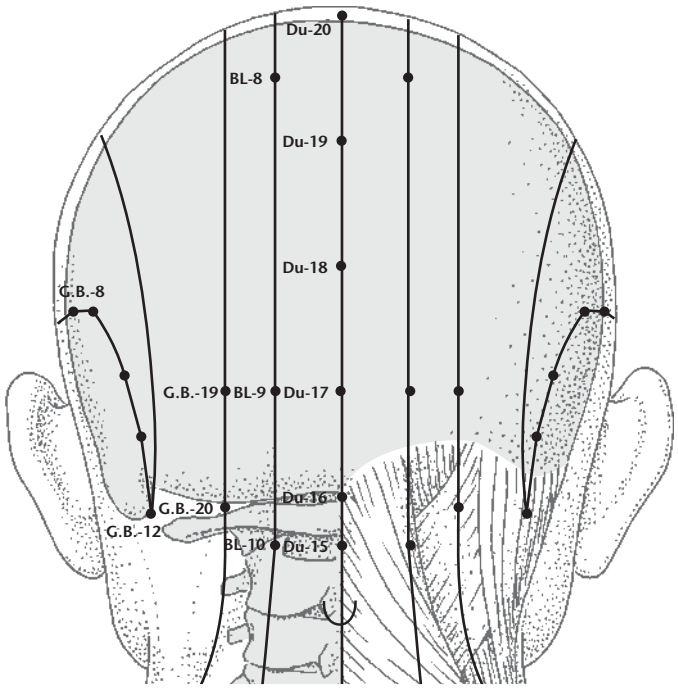

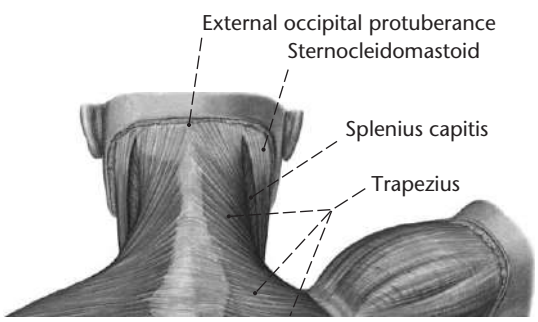

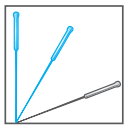
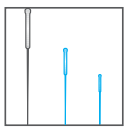

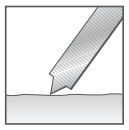

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 419

Shoulder Well JIANJING G.B.-21

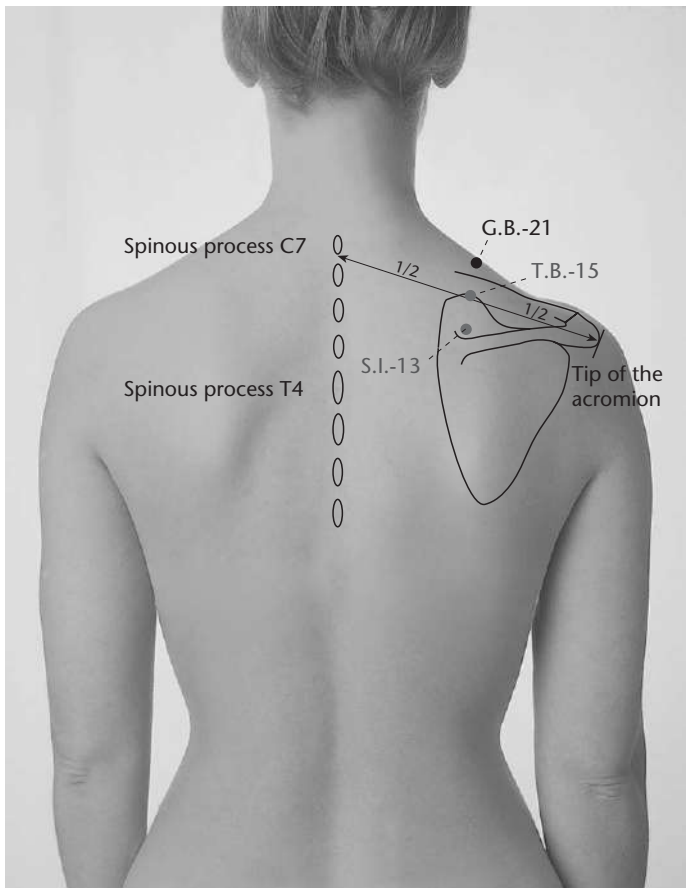

Location

At the highest point of the shoulder, at the midpoint of a line connecting the 7th cervical vertebra (C7) and the lateral extremity of the acromion.

How to find

First, locate C7 (→ 3.4.1) and the lateral extremity of the acromion (→ 3.3.1). Next, locate **G.B.-21** at the midpoint of a line connecting these two reference points, on the highest point of the trapezius muscle on the sagittal plane.
→ **T.B.-15** is located 1 cun inferior to **G.B.-21**, midway between **G.B.-21** and → **S.I.-13** (medial to the supraspinous fossa).

Needling

Vertically approximately 0.3–0.5 cun (lift the muscle). An alternative, safer method: Lift the muscle and insert the needle approximately 1 cun anteriorly or posteriorly into the muscle belly. Caution: Contraindicated during pregnancy! Pneumothorax.

Actions/Indications

- Opens the channel
- Descends the Qi
- Regulates the Qi flow, transforms Phlegm, dissipates nodules
- Promotes labour, benefits the breasts

Special features

Meeting point with the *yang wei mai* and the T.B. and ST* channels. Important local point. Trigger point for the shoulder with a distal effect on the Uterus.

*According to Deadman et al 1998.

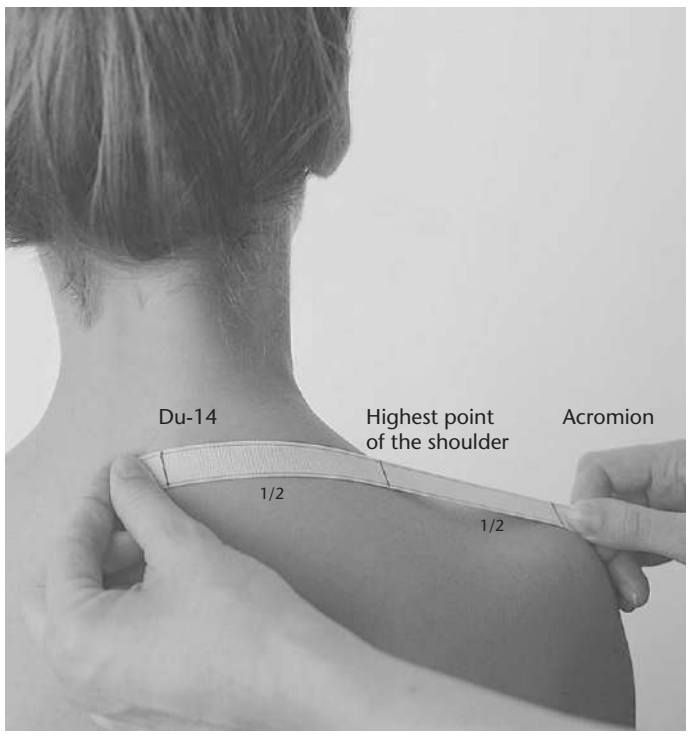

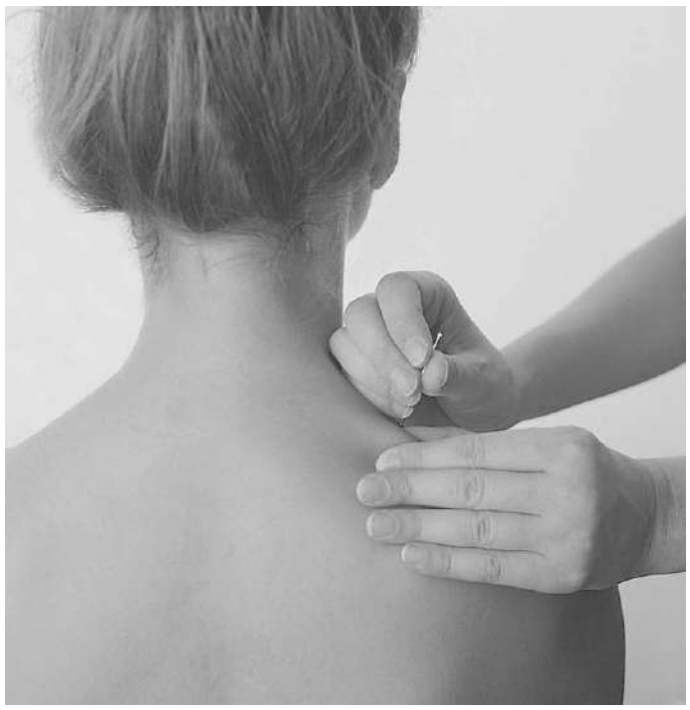

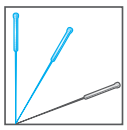
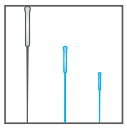

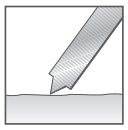
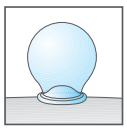

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 420

G.B.-22 *Armpit Abyss* YUANYE

Location

On the midaxillary line, with the arm abducted approximately 3 cun inferior to the apex of the axilla, in the 4th intercostal space (according to some authors, in the 5th intercostal space).

How to find

With the patient's arm slightly abducted, locate the apex of the axilla. Locate **G.B.-22** 3 cun inferior to it in the 4th intercostal space. **For orientation:** In men, the nipple is located in the 4th intercostal space; in women, in a supine position, its location may vary. Therefore, in women, the manubriosternal synchondrosis (→ 3.5) is a more reliable reference point for the costal region. Note: The intercostal space curves in a superior direction towards lateral.

Also located on the level of or in the 4th intercostal space, but more medially, are → **Ren-17/KID-23/ST-17/P-1/SP-18/G.B.-23** (on the midline/2 cun lateral to the midline/in the centre of the nipple or 4 cun lateral to the midline/1 cun lateral to the nipple or 5 cun lateral to the midline/6 cun lateral to the midline/1 cun anterior to **G.B.-22**).

Needling

0.5–1 cun obliquely or transversely (subcutaneously) along the intercostal space. Caution: Pneumothorax.

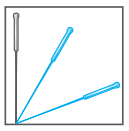

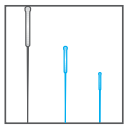

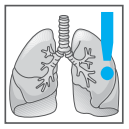

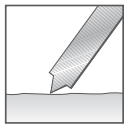

Actions/Indications

- Regulates the Qi in the Upper Burner
- Opens the *luo*-connecting vessels of the axilla

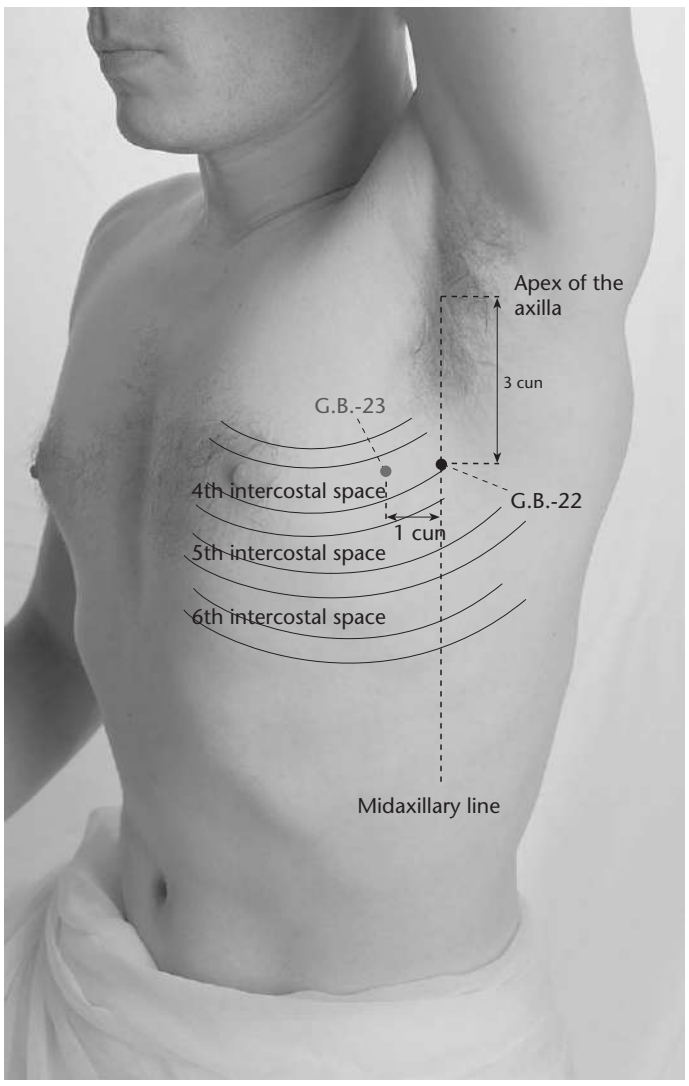

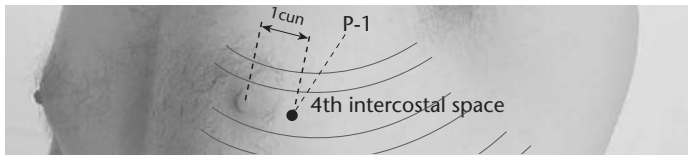

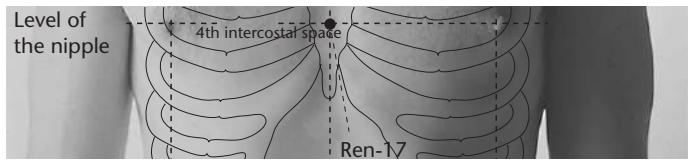

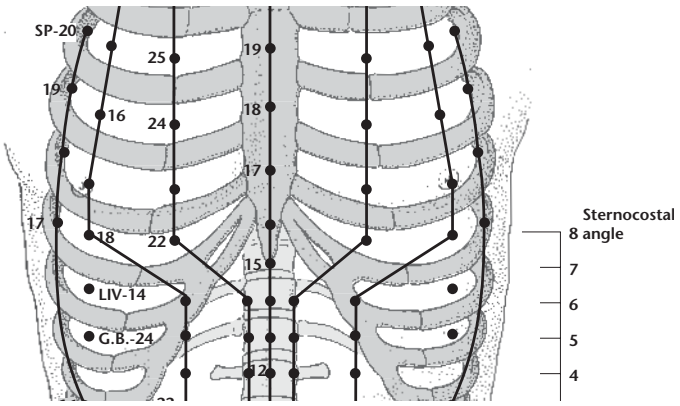

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 421

Flank Sinews ZHEJIN G.B.-23

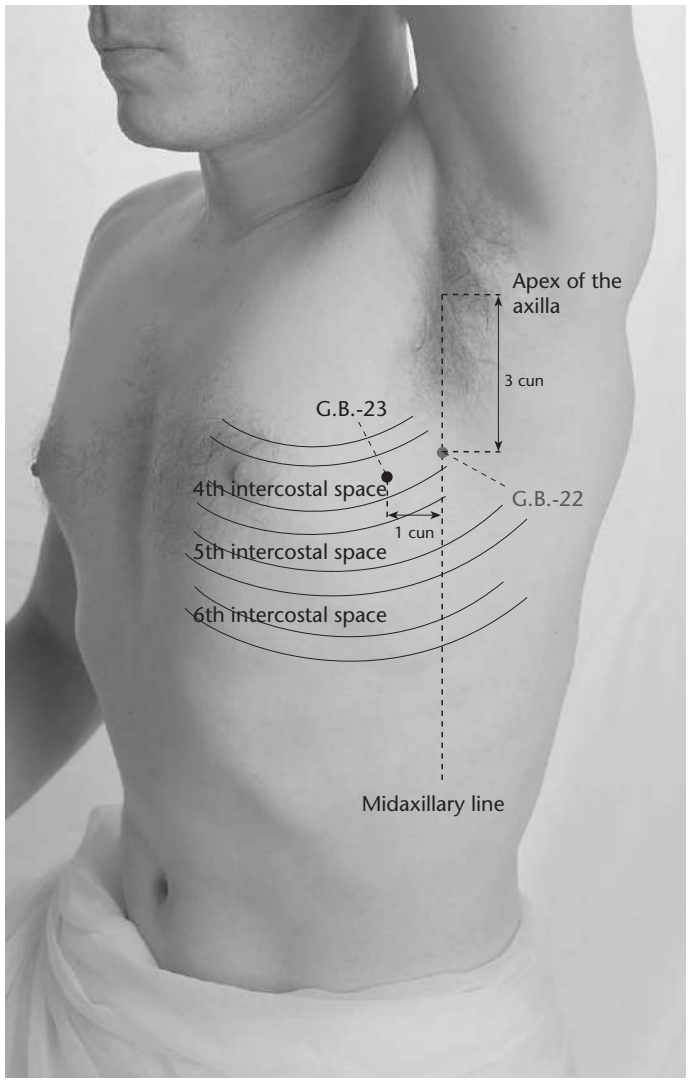

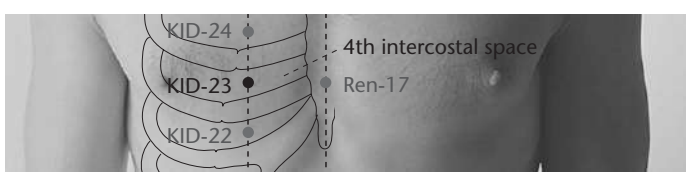

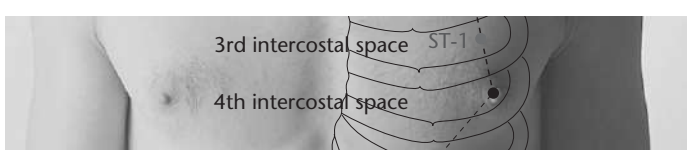

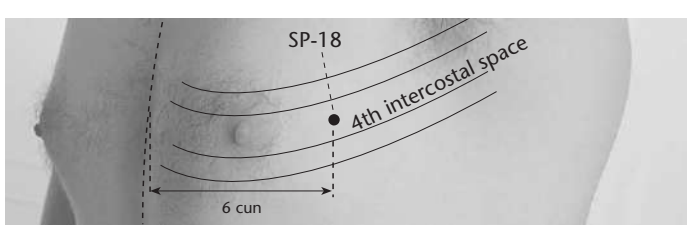

Location

1 cun anterior to → **G.B.-22** (on the midaxillary line, 3 cun inferior to the apex of the axilla in the 4th intercostal space).

How to find

With the patient's arm slightly abducted, locate the apex of the axilla. First, locate → **G.B.-22** 3 cun inferior to it in the 4th intercostal space, then locate **G.B.-23** 1 cun anterior to → **G.B.-22** and also in the 4th intercostal space. **For orientation:** In men, the nipple is located in the 4th intercostal space; in women, in a supine position, its location may vary. Therefore, in women, the manubriosternal synchondrosis (→ 3.5) is a more reliable reference point for the costal region. Note: The intercostal space curves in a superior direction towards lateral.

Also located on the level of or in the 4th intercostal space, but more medially, are → **Ren-17/KID-23/ST-17/P-1/SP-18** (on the midline/2 cun lateral to the midline/in the centre of the nipple or 4 cun lateral to the midline/1 cun lateral to the nipple or 5 cun lateral to the midline/6 cun lateral to the midline).

Needling

0.5–1 cun obliquely or transversely (subcutaneously) along the 4th intercostal space. Caution: Pneumothorax.

Actions/Indications

- Regulates the Qi between the Upper and Middle Burners
- Opens the channel

Special features

Meeting point with the BL channel according to some authors

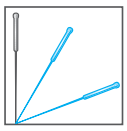
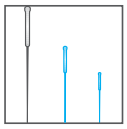
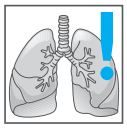

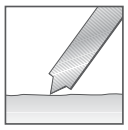

Ch04.11-F10028.qxd 2/23/08 11:24 AM Page 422

G.B.-24 *Sun and Moon* RIYUE

Location

In the 7th intercostal space, on the mamillary line (4 cun lateral to the anterior midline).

How to find

The mamillary line, a vertical line 4 cun lateral to the midline, is used for reference in the thoracic region (→ 3.5). **Quick method, especially in men:** In men, the nipple is usually located on the level of the 4th intercostal space. From the nipple, count downward to the 7th intercostal space. **Or:** A more reliable reference, especially in women, is the manubriosternal synchondrosis, a horizontal bony structure on the sternum. The costal cartilage of the second rib is lateral to the synchondrosis, with the 2nd intercostal space below. From there, count downward to the 7th intercostal space and locate **G.B.-24** on the mamillary line. → **LIV-14** is located directly superior to **G.B.-24** in the 6th intercostal space.

Located approximately on the same level (1 cun inferior to the sternocostal angle) are → **Ren-13/KID-20/ST-20** (on the midline/0.5 cun lateral to the midline/2 cun lateral to the midline).

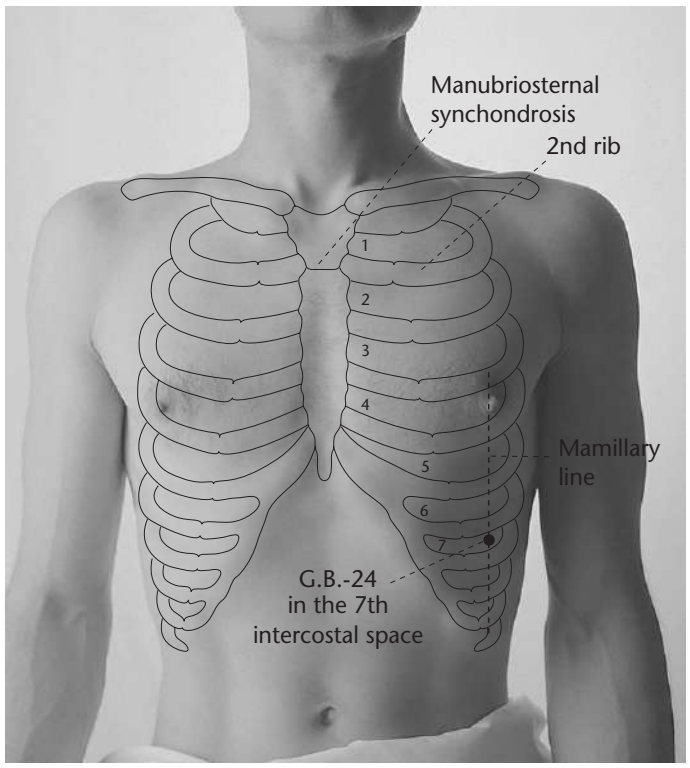

Needling

0.3–0.8 cun obliquely in a lateral direction along the intercostal space. Caution: Pneumothorax.

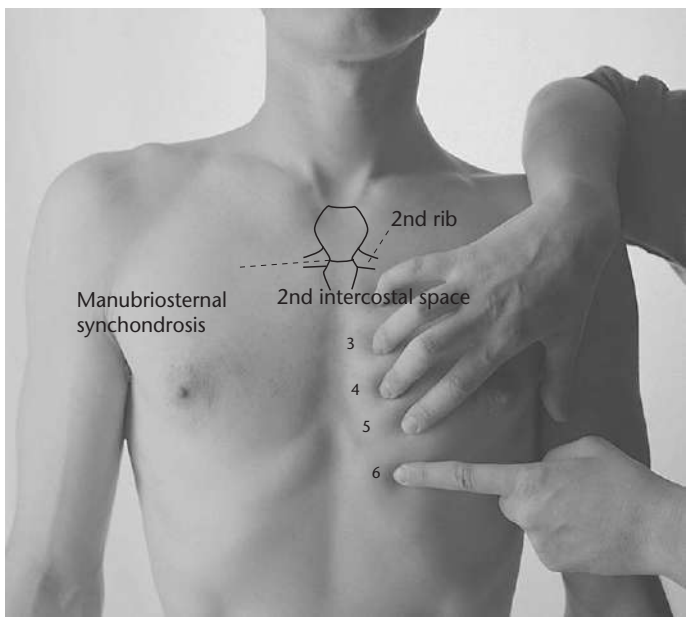

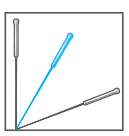

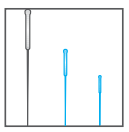

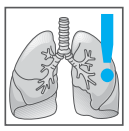

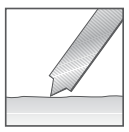

Actions/Indications

- Benefits the Gall Bladder, eliminates Damp-Heat, regulates and spreads Liver Qi, descends counterflow Qi, harmonises the Middle Burner

Special features

Front-*mu* point of the Gall Bladder, meeting point with the SP channel, also with the *yang wei mai* according to some authors. Major point for disorders of the Gall Bladder.

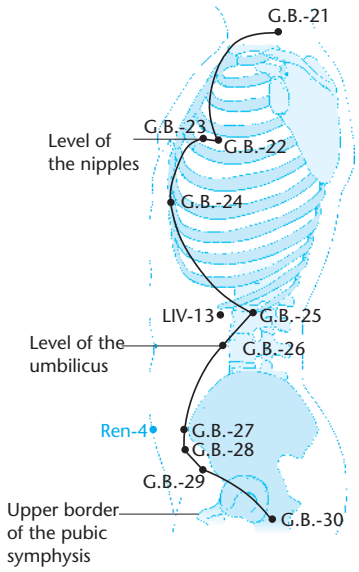

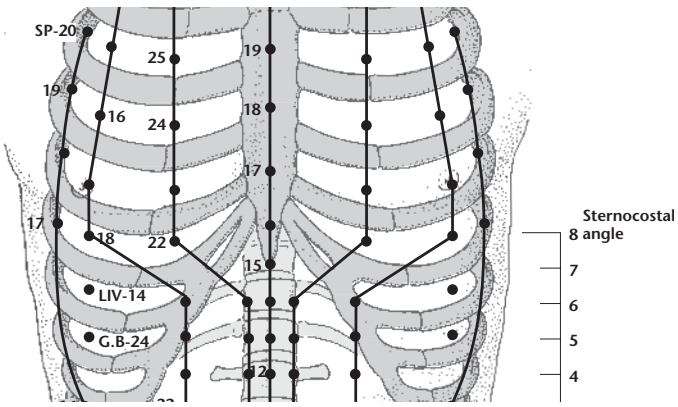

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 423

Capital Gate JINGMEN G.B.-25

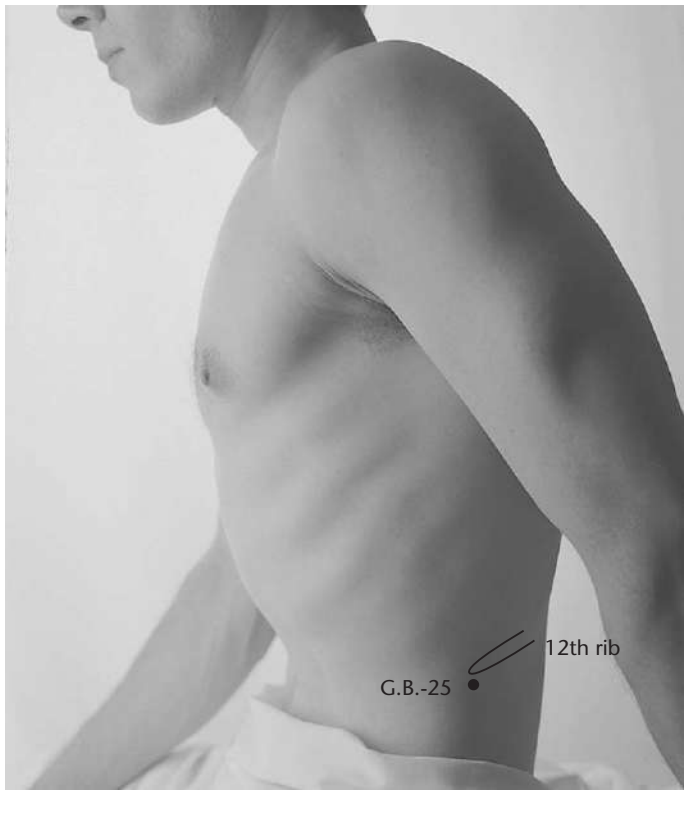

Location

On the lateral aspect of the ribcage, at the lower border of the free end of the 12th rib.

How to find

Place the hand on the upper abdomen and, with gentle pressure, glide along the lower border of the ribcage until you can feel the free end of the 11th rib (→ **LIV-13**) slightly superior to the umbilicus. By continuing to palpate along the lower border of the ribcage, you will feel the free end of the 12th rib on the lateral aspect of the waist. Locate **G.B.-25** on its lower border. **For orientation:** When pressing the flexed elbow onto the thorax, the tip of the olecranon will rest on the area around the free end of the 11th rib (→ **LIV-13**).

Needling

Vertically or obliquely 0.3–1 cun. Caution: Peritoneum; the needle should be inserted into the obliquus externus or internus muscles.

Actions/Indications

- Tonifies the Kidneys and regulates the water passages
- Strengthens the Spleen, regulates the Intestines
- Benefits the lumbar region

Special features

Front-*mu* point of the Kidneys

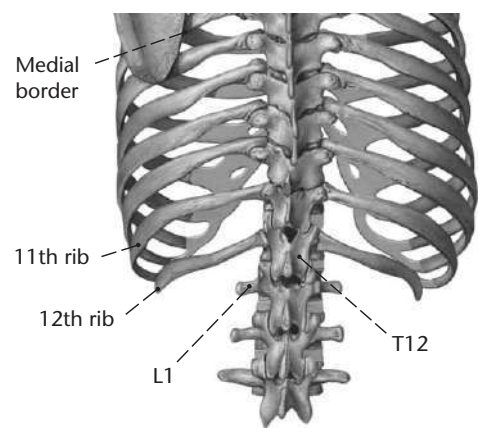

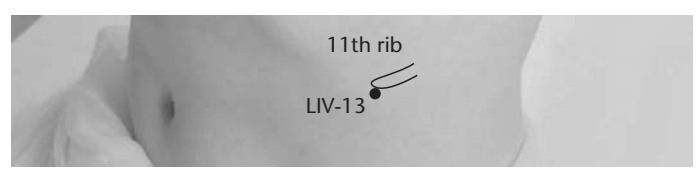

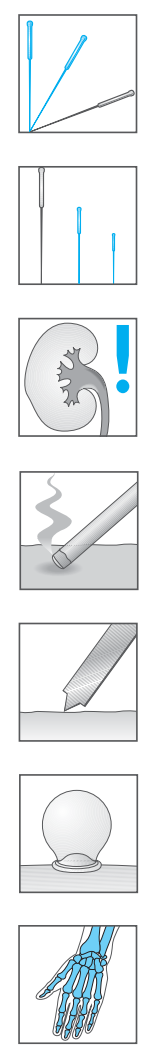

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 424

G.B.-26 *Girdling Vessel* DAIMAI

Location

On the lateral aspect of the waist, at the junction of a vertical line through the free end of the 11th rib and a horizontal line through the umbilicus, approximately 1.8 cun inferior to → **LIV-13**.

How to find

First, palpate the lower border of the ribcage to locate the free end of the 11th rib (→ **LIV-13**). Next, locate **G.B.-26** inferior to the free end of the 11th rib, level with the umbilicus. **Quick method for locating → LIV-13:** When pressing the flexed elbow onto the thorax, the tip of the olecranon will rest on the area around the free end of the 11th rib (→ **LIV-13**).
Located on the same level are → **Ren-8** (in the centre of the umbilicus), → **KID-16/ST-25/SP-15** (0.5/2/4 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Be careful with slim patients.

Actions/Indications

- Regulates the *dai mai*
- Regulates the Uterus
- Drains Damp-Heat

Special features

Meeting point with the *dai mai*

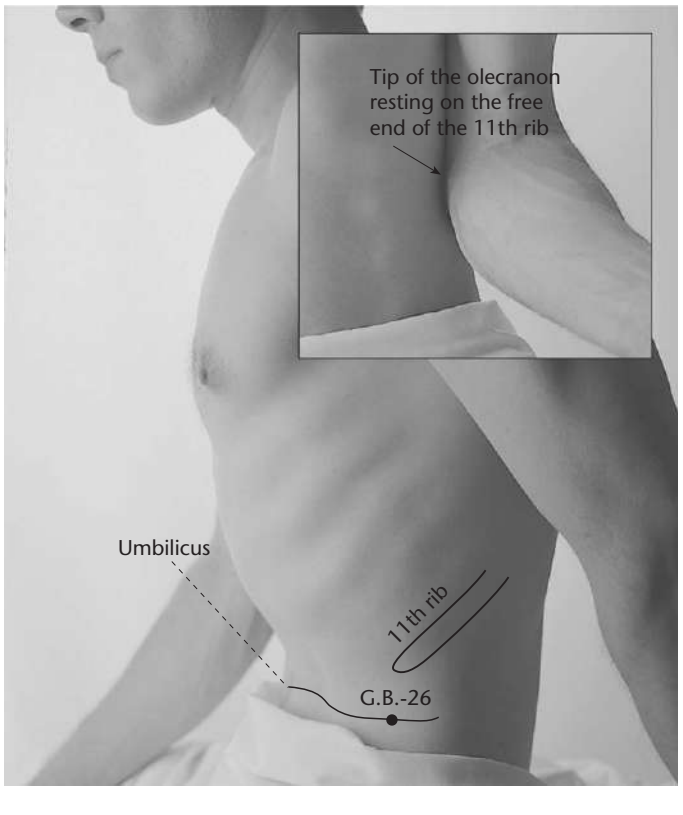

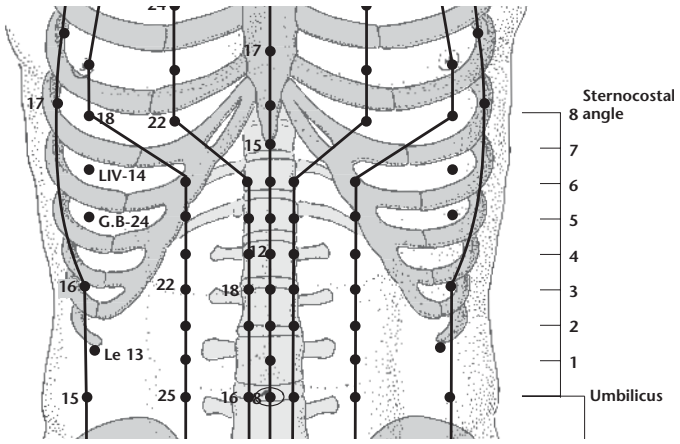

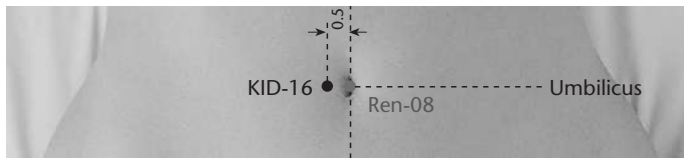

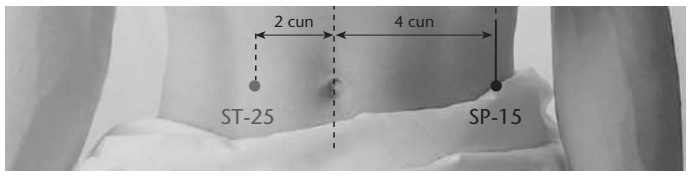

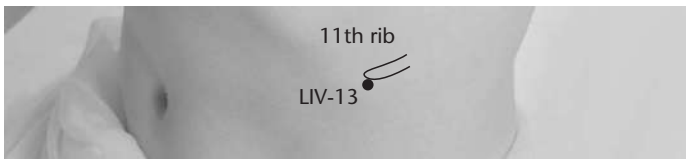

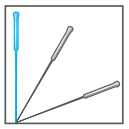
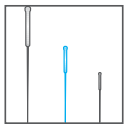

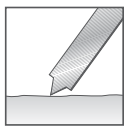

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 425

Fifth Pivot WUSHU G.B.-27

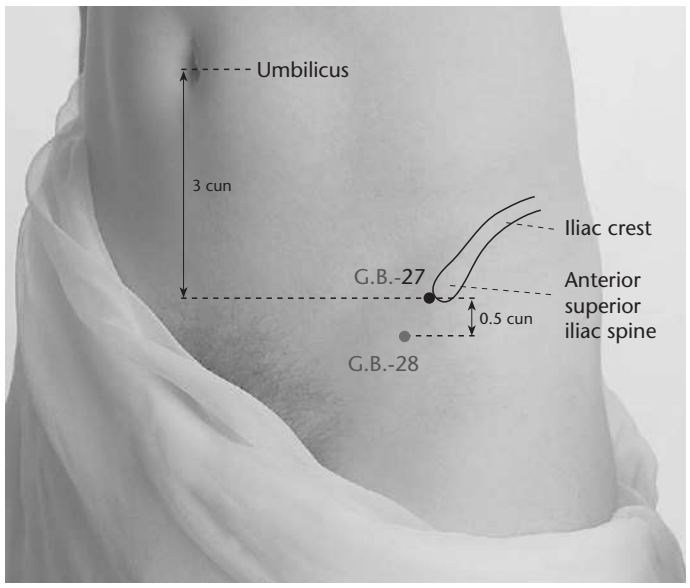

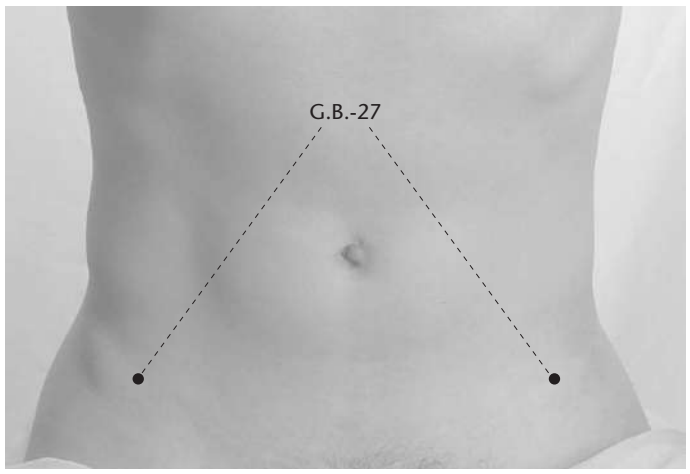

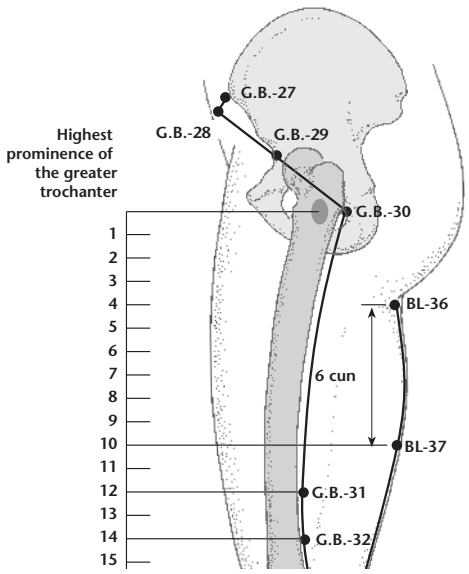

Location

In the depression medial to the anterior superior iliac spine (ASIS), approximately 3 cun inferior to the umbilicus.

How to find

First, locate the anterior superior iliac spine (ASIS → 3.5) by palpating along the upper border of the iliac crest in an anterior and inferior direction. At its anterior end, the ASIS can be palpated as a bony ridge on the lateral aspect of the lower abdomen. Locate **G.B.-27** in a depression anterior and medial to the ASIS. → **G.B.-28** is located approximately 0.5 cun inferior and medial to **G.B.-27**.

Located approximately on the same level (3 cun inferior to the umbilicus) are → **Ren-4/KID-13/ST-28** (on the midline/0.5 cun lateral to the midline/2 cun lateral to the midline).

Needling

Vertically 1–1.5 cun. Caution during pregnancy.

Actions/Indications

- Strengthens the Kidneys
- Regulates the *dai mai*
- Regulates the Triple Burner

Special features

Meeting point with the *dai mai*

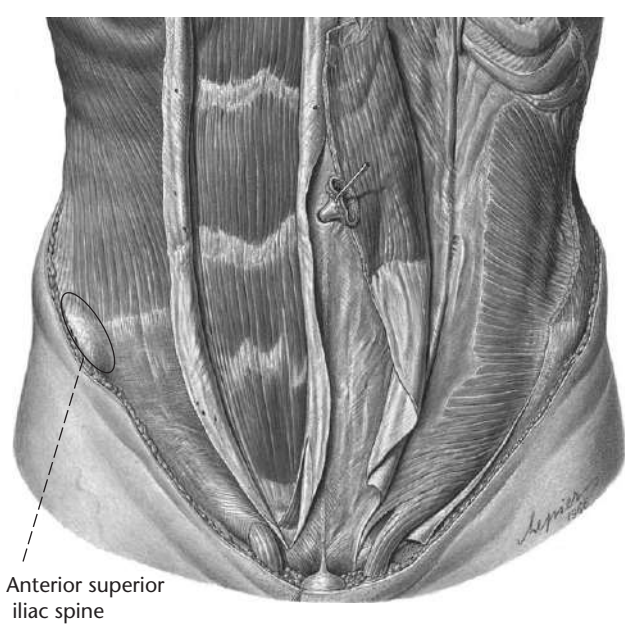

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 426

G.B.-28 *Linking Path* WEIDAO

Location

On the lateral aspect of the abdomen, anterior and inferior to the anterior superior iliac spine (ASIS), approximately 0.5 cun anterior and inferior to → **G.B.-27**.

How to find

First, locate the ASIS (→ 3.5) by palpating along the upper border of the iliac crest in an anterior and inferior direction. On its anterior border, the ASIS can be palpated as a bony ridge on the lateral aspect of the lower abdomen, with → **G.B.-27** located in a depression anterior and medial to the ASIS. Locate **G.B.-28** approximately 0.5 cun anterior and inferior to → **G.B.-27**.

Needling

Vertically 1–1.5 cun. Caution during pregnancy.

Actions/Indications

- Regulates the *dai mai*
- Regulates the Triple Burner
- Eliminates Dampness

Special features

Meeting point with the *dai mai*

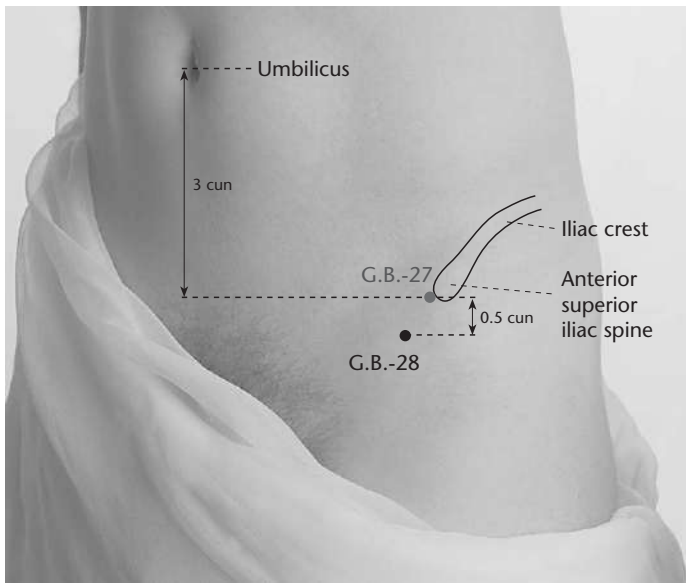

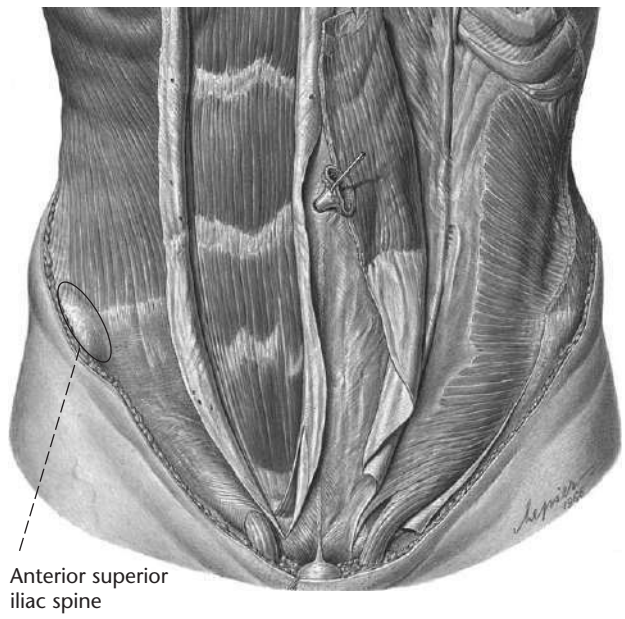

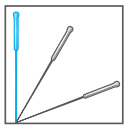

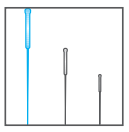

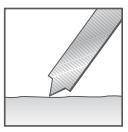

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 427

Stationary Crevice JULIAO G.B.-29

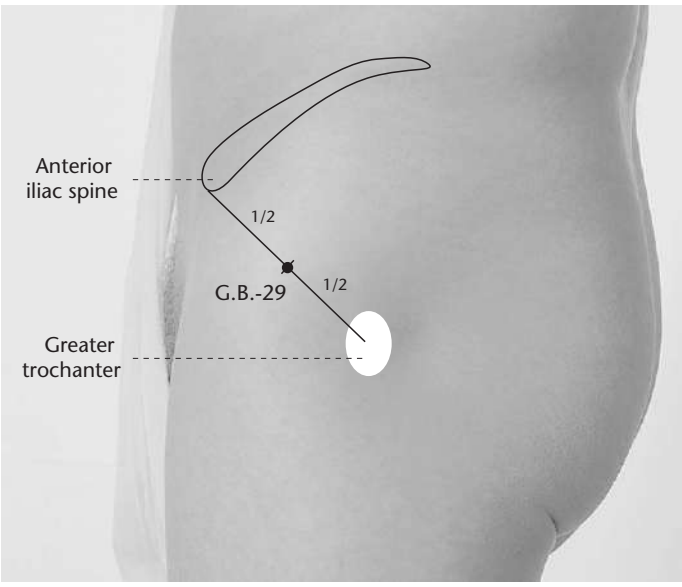

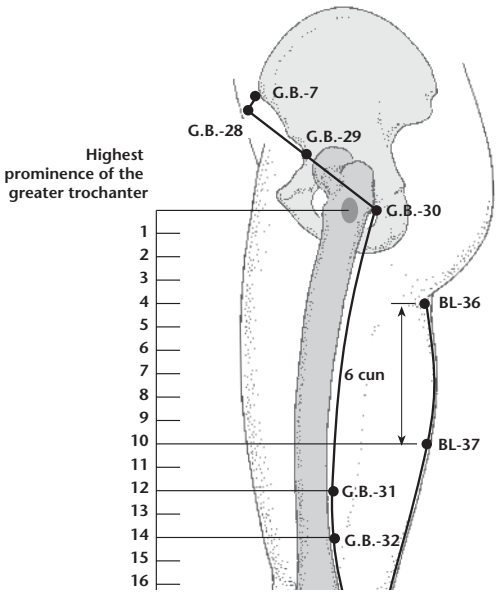

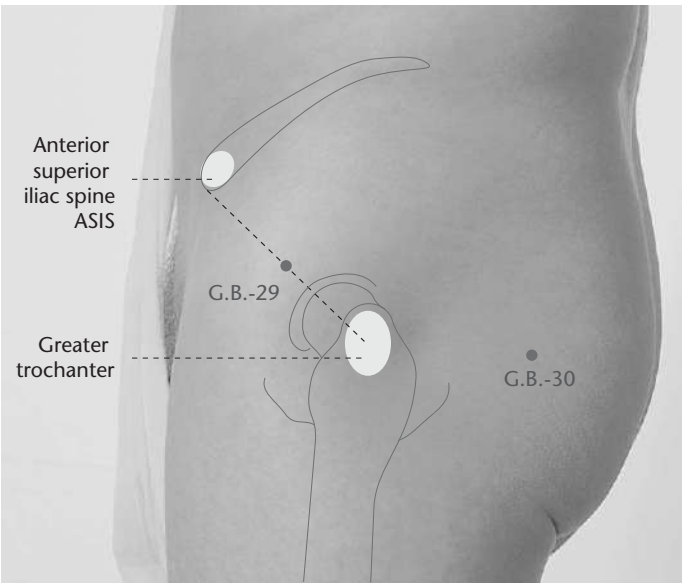

Location

At the midpoint of a line connecting the anterior superior iliac spine (ASIS) and the greater trochanter, at the anterior border of the iliac crest.

How to find

First, locate the ASIS (→ 3.5), the highest point on the anterior aspect of the iliac crest, by palpating along the upper border of the iliac crest in an anterior and inferior direction. At its anterior end, the ASIS can be palpated as a bony ridge on the lateral aspect of the lower abdomen. With the patient lying on their side and their leg slightly flexed, locate the greater trochanter (→ 3.6), a clearly marked bony structure in the region of the hip joint. **G.B.-29** is located at the midpoint of a line connecting these two reference points.

Needling

Vertically 1–2 cun

Actions/Indications

- Expels Cold, Dampness and Wind, opens the channel

Special features

Meeting point with the *yang qiao mai*, also with the *dai mai* according to some authors.

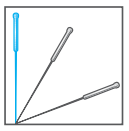
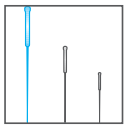

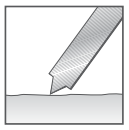

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 428

G.B.-30 *Jumping Circle* HUANTIAO

Location

With the patient lying on their side, the point is at the junction between the medial two-thirds and the lateral third of a line connecting the greater trochanter and the sacral hiatus.

How to find

Patient's position: Supine or, better, lying on their side, preferably with the hip and knee joints of the side to be needled flexed and the lower leg straight. Use pillows, etc, for a comfortable position. Reference points: the sacral hiatus (→ 3.4.4) and the lateral prominence of the greater trochanter (→ 3.6). **G.B.-30** is located on a line connecting these two points, one third of the distance from the greater trochanter.

Needling

1.5–3 cun vertically towards the genital region. The needle may reach the fascia of the obturator internus muscle as well as intermuscular connective tissue. Long 3 cun needles (50mm) should be used. Caution: Needling is often painful. Needling can result in an electric *de qi* sensation radiating to the toes. This sensation is particularly common with sciatic disorders and if **G.B.-30** is needled 1 cun inferior to its normal location.

Actions/Indications

- Opens the channel and the *luo*-connecting vessels, alleviates pain, benefits the hip and legs, eliminates Wind-Damp.

Special features

Meeting point with the BL channel, Ma Dan Yang Heavenly Star point. Important point for disorders of the hips.

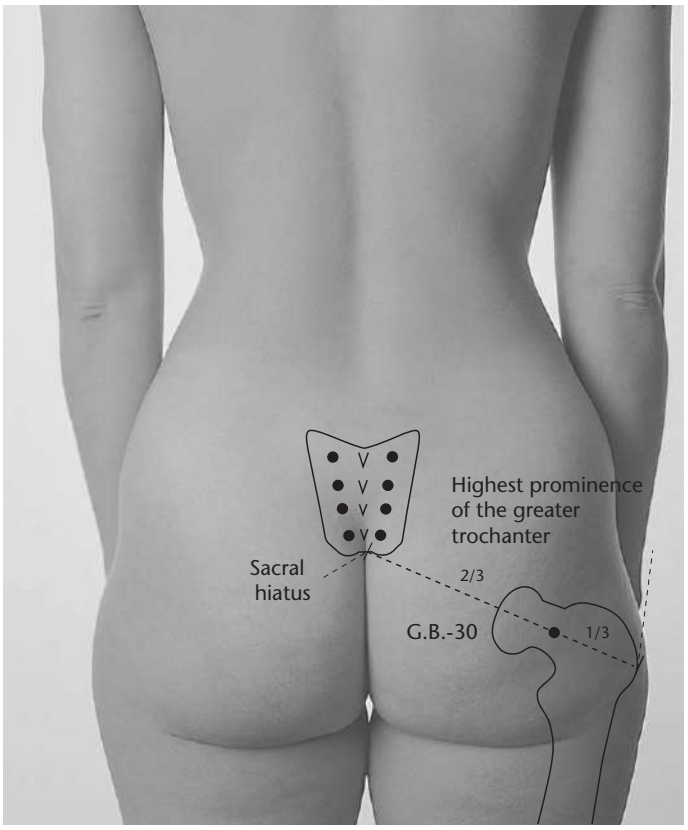

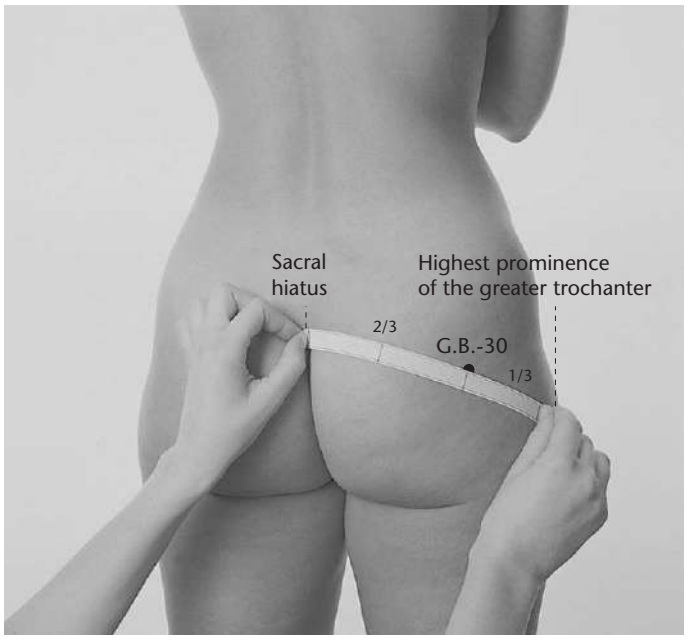

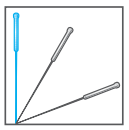
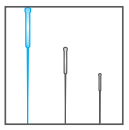

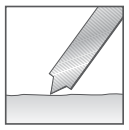
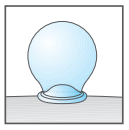

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 429

Wind Market FENGSHI G.B.-31

Location

On the lateral aspect of the thigh, inferior to the greater trochanter, approximately 7 cun proximal to the popliteal crease.

How to find

Ask the patient to place their hands on the imaginary seam of their trousers (this is best done with the patient standing). **G.B.-31** is located where the middle finger touches the lateral aspect of the thigh. **Or:** The distance between the highest prominence of the greater trochanter (→ 3.6) to the popliteal crease is 19 cun (→ 2.2). Divide this distance into thirds and locate **G.B.-31** 1 cun proximal to a third of the distance from the popliteal crease. Sensitivity to pressure should help determine the location of this point.

Needling

Vertically 1–2 cun

Actions/Indications

- Eliminates Wind, Dampness and Heat
- Opens the channel and alleviates pain

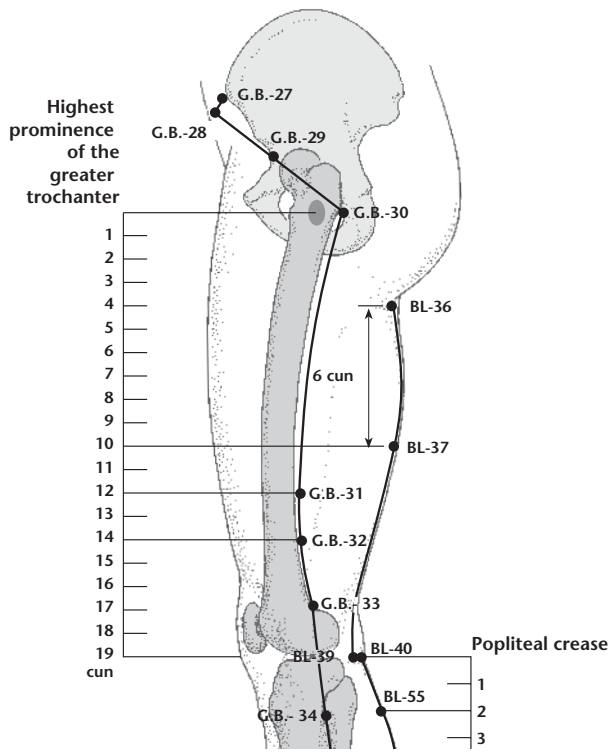

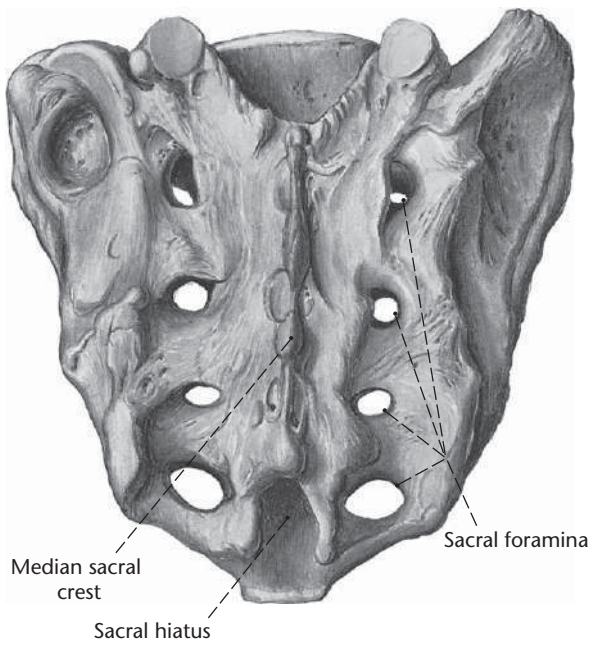

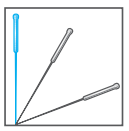
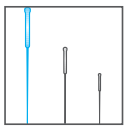

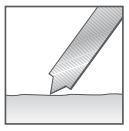

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 430

G.B.-32 *Middle Ditch* ZHONGDU

Location

On the lateral aspect of the thigh, 5 cun proximal to the popliteal crease, between the vastus lateralis and biceps femoris muscles.

How to find

First, ask the patient to place their hands on the imaginary seam of their trousers (this is best done with the patient standing) and locate → **G.B.-31** where the middle finger touches the thigh. **G.B.-32** can now be located 2 cun distal to → **G.B.-31. Or:** The distance between the highest prominence of the greater trochanter (→ 3.6) to the popliteal crease is 19 cun (→ 2.2). Divide this distance into quarters (use an elastic tape or the spreading hands technique) and locate **G.B.-32** slightly proximal to a quarter of the distance from the popliteal crease. Sensitivity to pressure should help determine the location of this point.

Needling

Vertically 1–2 cun

Actions/Indications

- Eliminates Wind, Dampness and Cold, opens the channel

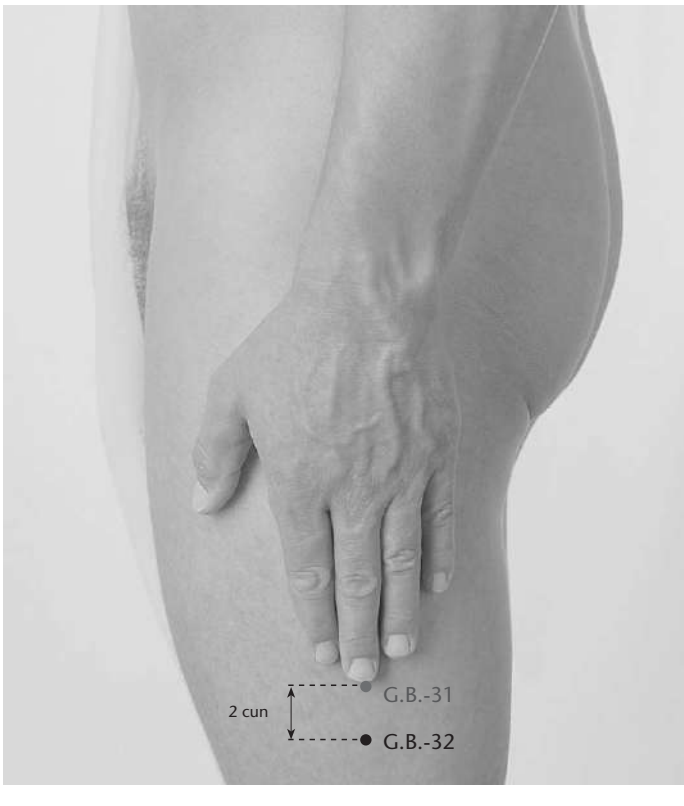

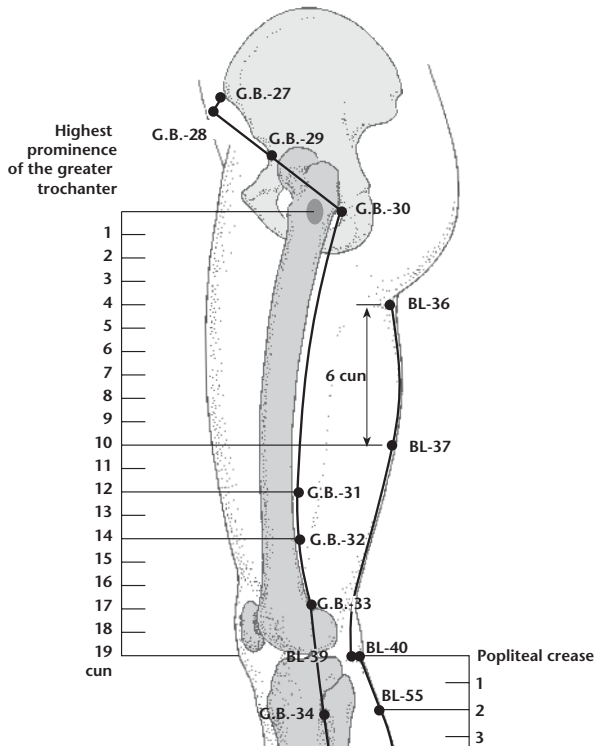

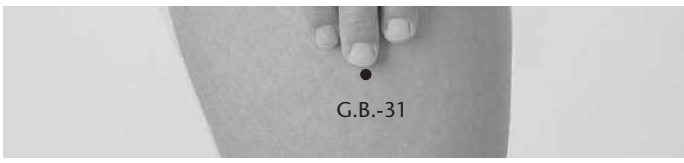

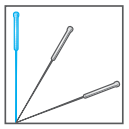
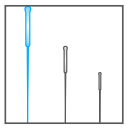

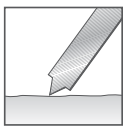
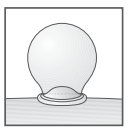

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 431

Knee Yang Gate XIYANGGUAN G.B.-33

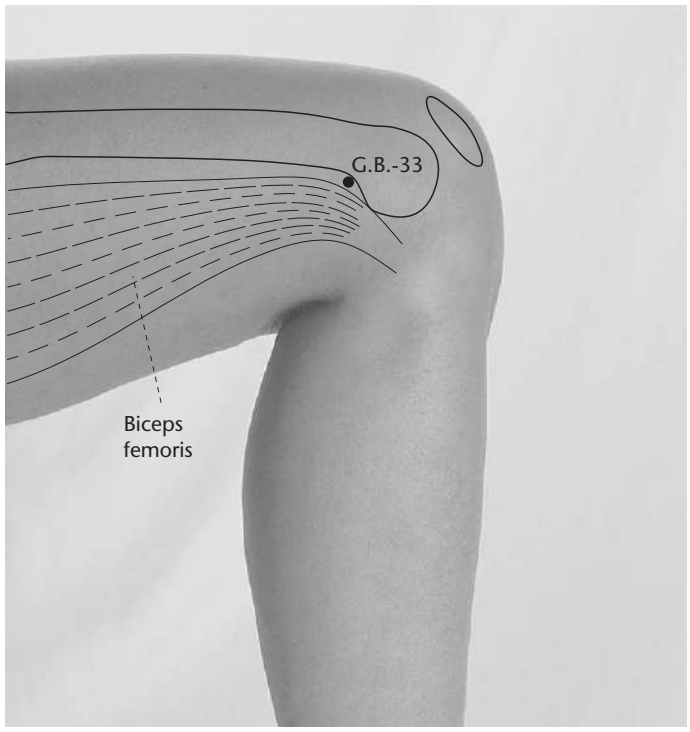

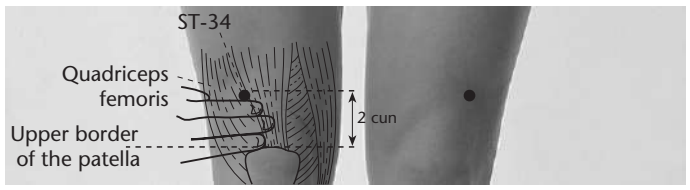

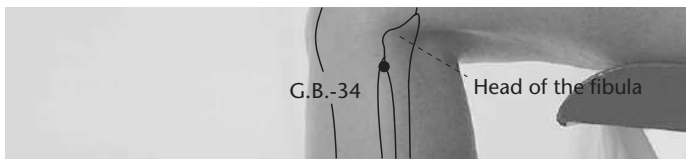

Location

On the lateral aspect of the knee. With the knee flexed, in the depression between the shaft and the lateral epicondyle and the tendon of the biceps femoris muscle, approximately 3 cun proximal to ➞ **G.B.-34**.

How to find

This point is best located with the patient's knee flexed. On the level of the lateral upper border of the patella, palpate in a lateral direction towards the thigh and palpate for the distal end of the lateral epicondyle of the femur. Then locate **G.B.-33** in a clearly palpable depression between the epicondyle and the tendon of the biceps femoris muscle. The latter runs along the lateral aspect of the leg (along the line of an imaginary trouser seam) and attaches inferior to the knee joint at the head of the fibula.
➞ **G.B.-34** is located 3 cun more distal, ➞ **ST-34** is located 2 cun proximal to the lateral upper border of the patella.

Needling

Vertically 1–2 cun

Actions/Indications

- Opens the channel
- Relaxes the tendons

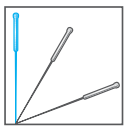
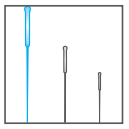

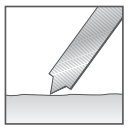

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 432

G.B.-34 *Yang Mound Spring* YANGLINGQUAN

Location

In the depression anterior and inferior to the head of the fibula, between the peroneus longus and extensor digitorum longus muscles.

How to find

This point is best located with the patient's knee flexed (use knee support). Palpate for the head of the fibula on the lateral aspect of the leg and hold it with the index and middle fingers in a tweezer-like way. Gliding distally with both fingers, the more medial finger will drop into a depression directly anterior and inferior to the head of the fibula, the location of **G.B.-34**.
Located on the same level but on the medial aspect of the lower leg is → **SP-9** (at the junction of the shaft and the medial condyle of the tibia).

Needling

1–1.5 cun vertical insertion between the tibia and fibula towards the interosseous membrane. Caution: Deep peroneal nerve with deep needling, in some cases also common peroneal nerve. The needle may reach the interosseous membrane as well as the epineural tissue of the peroneal nerve.

Actions/Indications

- Benefits the tendons and joints
- Opens the channel, alleviates pain, benefits the lateral costal region
- Clears Damp-Heat from the Liver and Gall Bladder
- Spreads Liver Qi

Special features

He-sea point, Earth point, *hui*-meeting point of the tendons (coordination/movement), lower *he*-sea point of the Gall Bladder, Ma Dan Yang Heavenly Star point. Major point for disorders of the tendons and musculature.

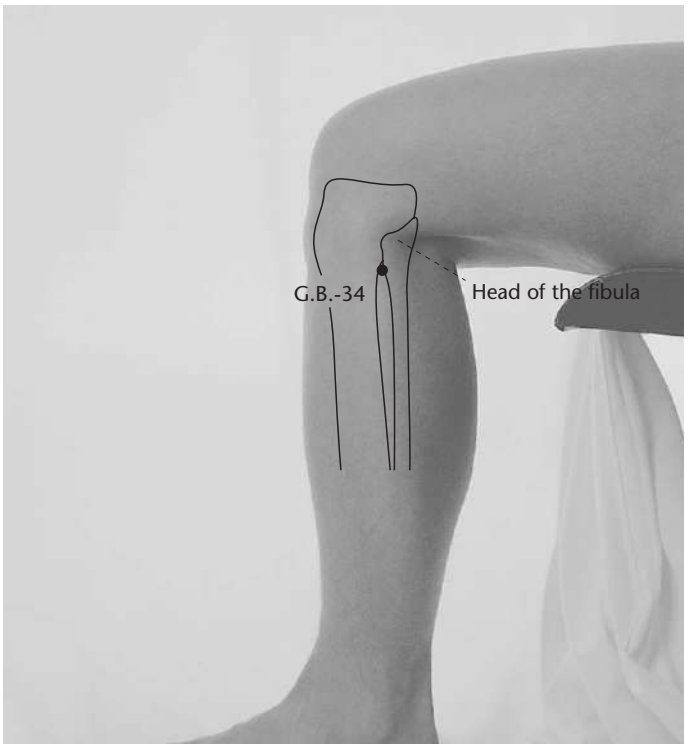

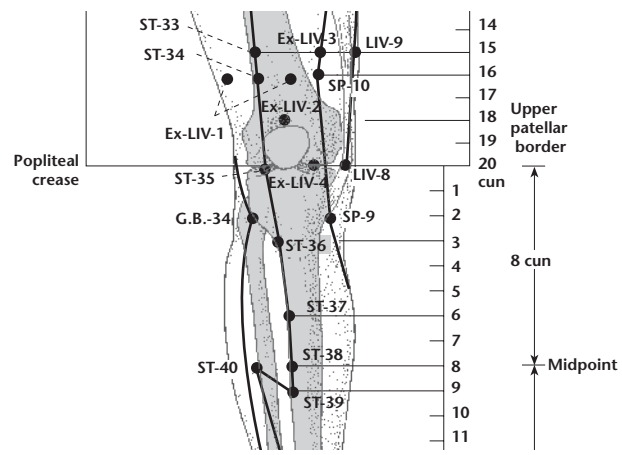

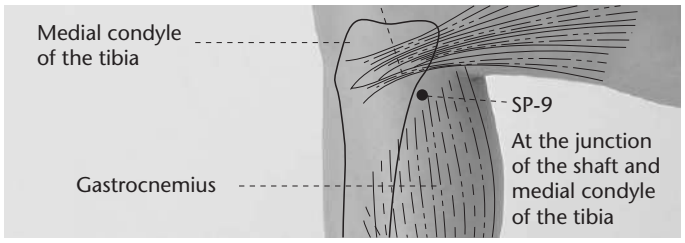

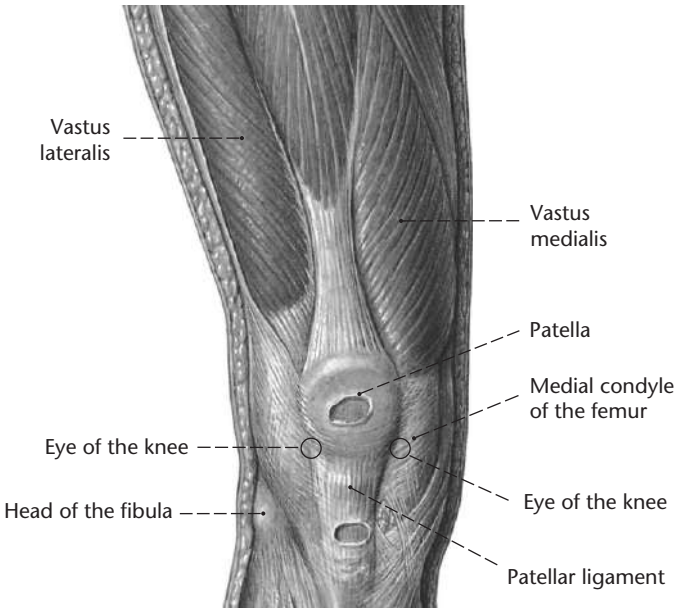

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 433

Yang Intersection YANGJIAO G.B.-35

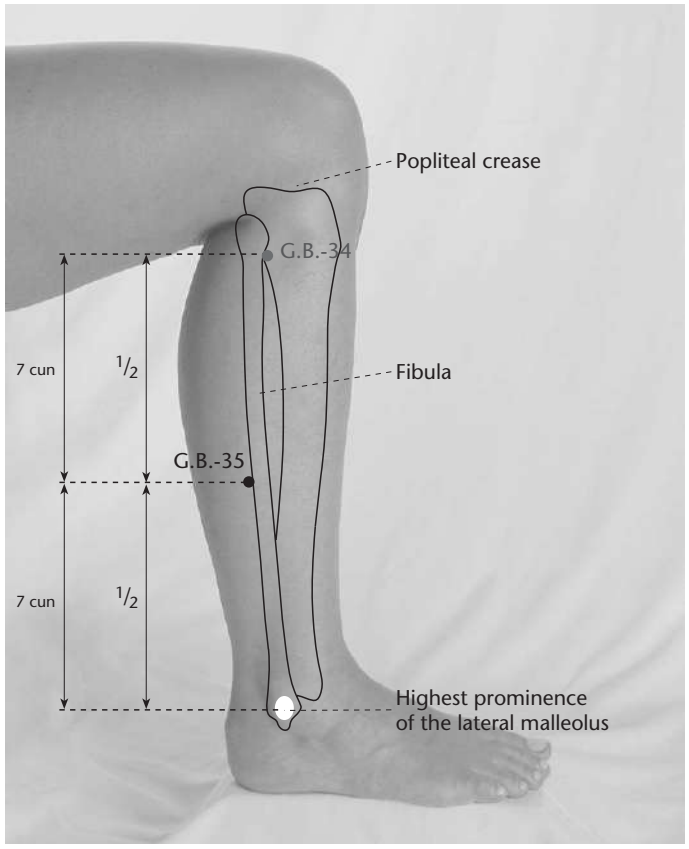

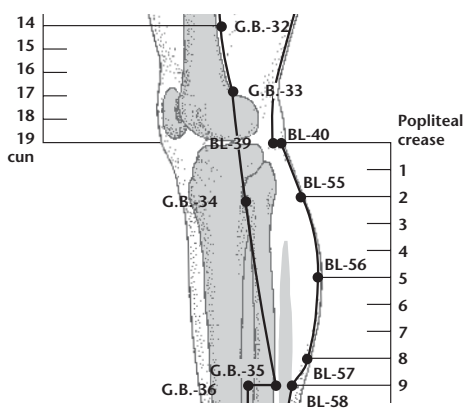

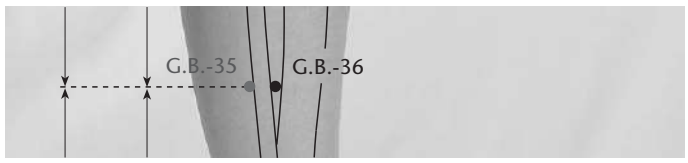

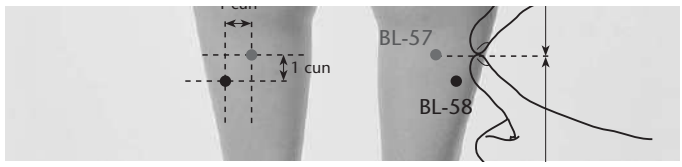

Location

7 cun proximal to the highest prominence of the lateral malleolus, on the **posterior** border of the fibula.

How to find

Quick method: Spreading hands technique (→ 2.3.3): Locate **G.B.-35** at the midpoint of a line connecting → **G.B.-34** (in the depression anterior and inferior to the head of the fibula) and the highest prominence of the lateral malleolus (→ 3.6.2), on the **posterior** border of the fibula (distance = 14 cun). **G.B.-35** is located 7 cun proximal to the prominence of the lateral malleolus. **Or:** Spreading hands technique (→ 2.3.3): Locate **G.B.-35** 1 cun distal to the midpoint of a line connecting the popliteal crease and the highest prominence of the lateral malleolus, on the **posterior** border of the fibula (distance = 16 cun, → 2.2). **For orientation:** The borders of the fibula are deep to the peroneus brevis muscle and are often not easily palpable. For this reason, it is suggested to palpate the posterior border of the fibula just superior to the lateral malleolus and then locate the point on an imaginary line running to the head of the fibula.
Located on the same level are → **G.B.-36** (on the **anterior** border of the fibula) and → **BL-58** (7 cun proximal to → **BL-60**).

Needling

Vertically 0.5–1.5 cun

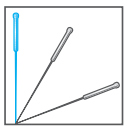

Actions/Indications

- Clears Heat, opens the channel, relaxes the tendons
- Regulates Qi of the Gall Bladder
- Calms the *shen*

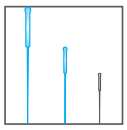

Special features

Xi-cleft point of the *yang wei mai*, meeting point with the *yang wei mai* according to some authors.

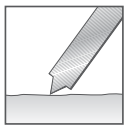

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 434

G.B.-36 *Outer Mound* WAIQIU

Location

7 cun proximal to the highest prominence of the lateral malleolus, on the **anterior** border of the fibula.

How to find

Quick method: Spreading hands technique (→ 2.3.3): Locate → **G.B.-35** at the midpoint of a line connecting → **G.B.-34** (in the depression anterior and inferior to the head of the fibula) and the highest prominence of the lateral malleolus (→ 3.6.2), on the **anterior** border of the fibula (distance = 14 cun). **G.B.-36** is located 7 cun proximal to the prominence of the lateral malleolus. **Or:** Spreading hands technique (→ 2.3.3): locate **G.B.-36** 1 cun distal to the midpoint of a line connecting the popliteal crease and the highest prominence of the lateral malleolus, on the **anterior** border of the fibula (distance = 16 cun, → 2.2). **For orientation:** The borders of the fibula are deep to the peroneus brevis muscle and are often not easily palpable. For this reason, it is suggested to palpate the anterior border of the fibula just superior to the ankle and then locate the point on an imaginary line running to the head of the fibula.
Located on the same level are → **G.B.-35** (on the **posterior** border of the fibula) and → **BL-58** (7 cun proximal to → **BL-60**).

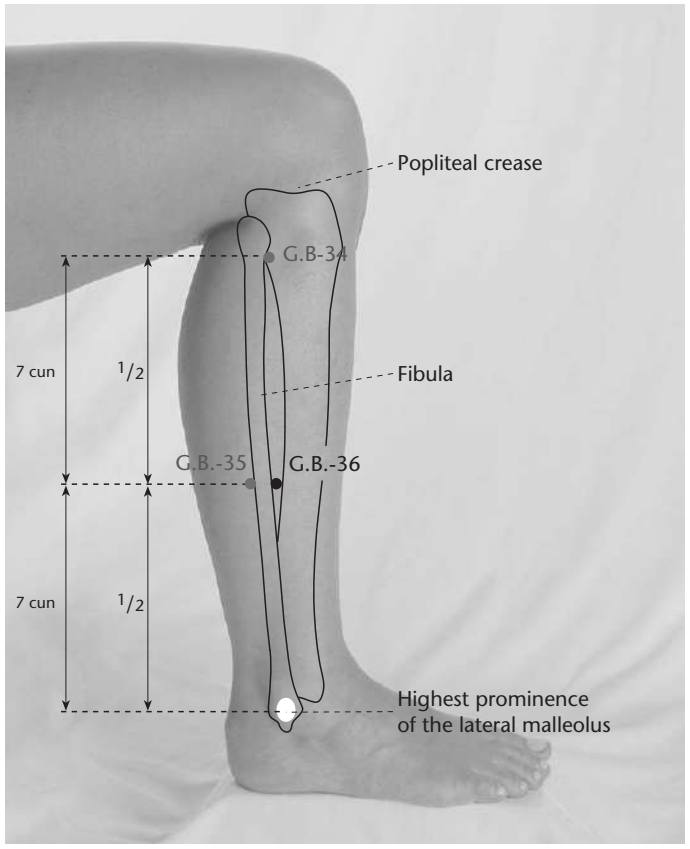

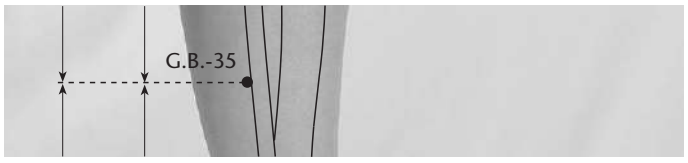

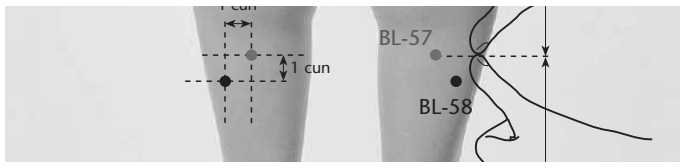

Needling

Vertically 0.5–1.5 cun

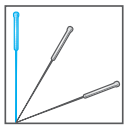

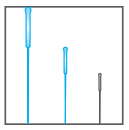

Actions/Indications

- Regulates Gall Bladder and Liver Qi
- Drains Damp-Heat
- Relaxes the tendons and muscles
- Calms the *shen*

Special features

Xi-cleft point

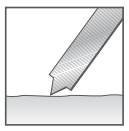

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 435

Bright Light GUANGMING G.B.-37

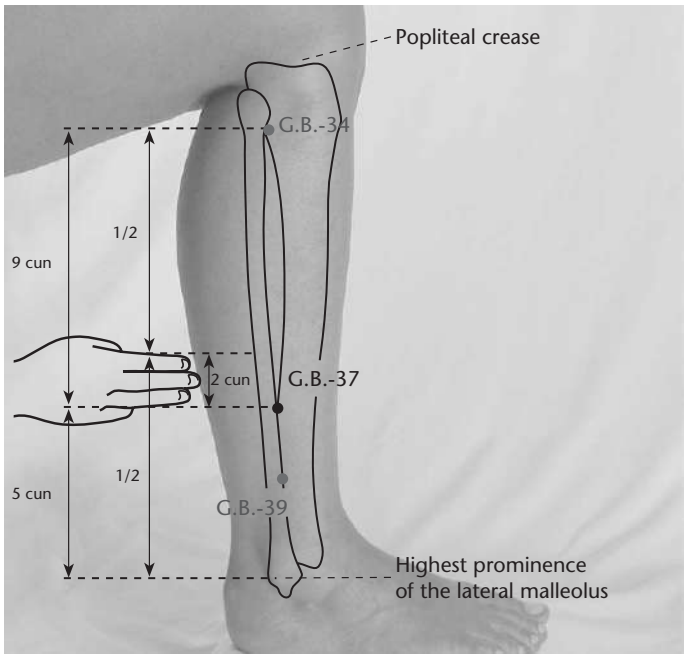

Location

5 cun proximal to the highest prominence of the lateral malleolus, on the anterior border of the fibula, between the peroneus longus and extensor digitorum longus muscles.

How to find

Quick method: Spreading hands technique (→ 2.3.3): Place the little fingers on → **G.B.-34** (in the depression anterior and inferior to the head of the fibula) and the highest prominence of the lateral malleolus (→ 3.6.2). This distance is 14 cun. From the midpoint of this distance, measure 2 cun in a distal direction and locate **G.B.-37** on this level in a depression on the **anterior** border of the fibula. This depression is located 5 cun proximal to the prominence of the lateral malleolus (the width of 1 hand and 2 thumbs). **For orientation:** The borders of the fibula are deep to the peroneus brevis muscle and are often not easily palpable. For this reason, it is suggested to palpate the anterior border of the fibula just superior to the ankle and then locate the point on an imaginary line running to the head of the fibula.

Needling

Vertically 0.5–1.5 cun

Actions/Indications

- Benefits the eyes
- Eliminates Wind-Damp, opens the channel, alleviates pain

Special features

Luo-connecting point. Important distal point for disorders of the eyes.

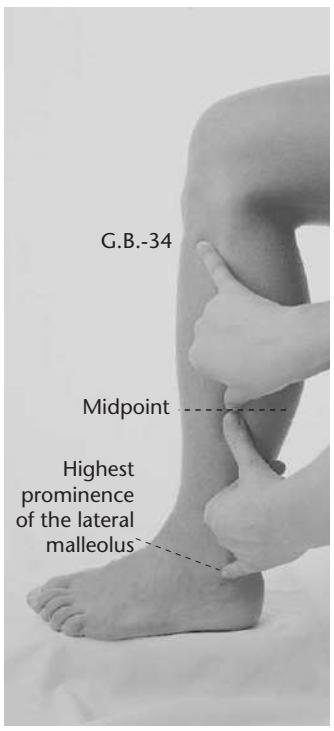

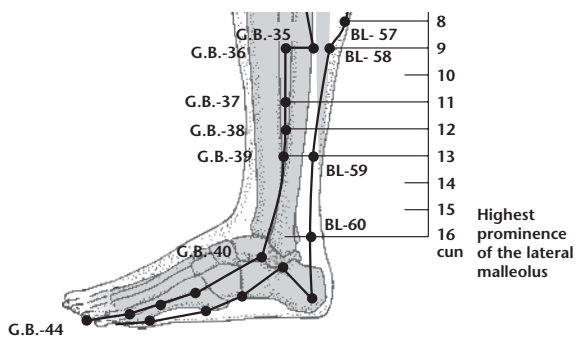

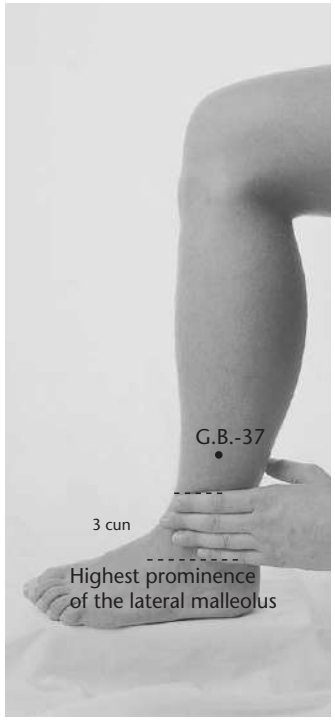

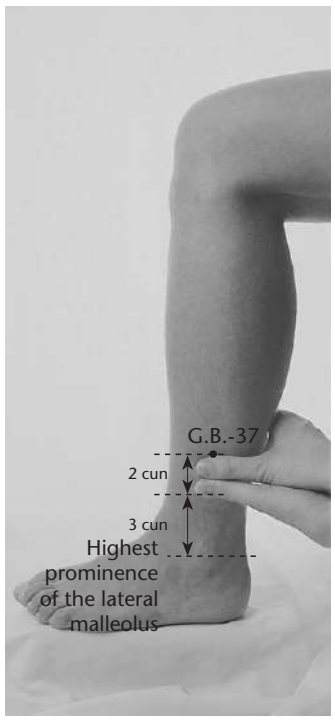

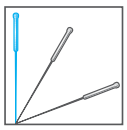
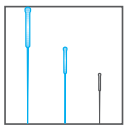

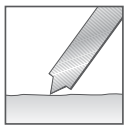

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 436

G.B.-38 *Yang Assistance* YANGFU

Location

4 cun proximal to the highest prominence of the lateral malleolus, on the **anterior** border of the fibula.

How to find

First, locate the prominence of the lateral malleolus (→ 3.6.2) and, from there, measure 4 cun in a proximal direction. There, locate **G.B.-38** on the **anterior** border of the fibula. **Or:** On the lateral aspect of the lower leg, the distance between the prominence of the lateral malleolus (→ 3.6.2) and the popliteal crease is 16 cun (→ 2.2). Divide this distance into quarters and locate **G.B.-38** one quarter of the distance from the lateral malleolus (aid: spreading hands technique or elastic tape → 2). **For orientation:** The borders of the fibula are deep to the peroneus brevis muscle and are often not easily palpable. For this reason, it is suggested to palpate the anterior border of the fibula just superior to the ankle and then locate the point on an imaginary line running to the head of the fibula.

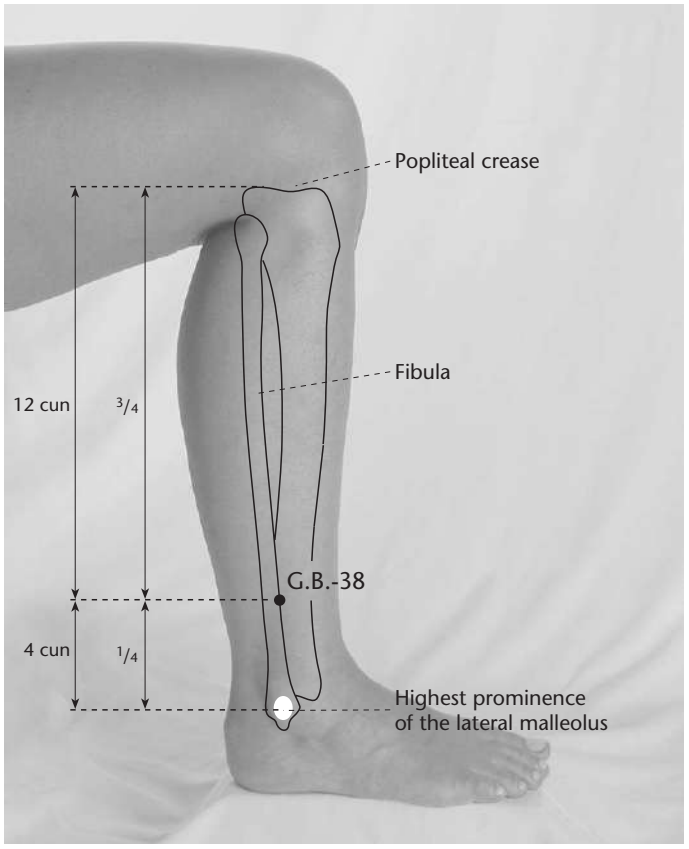

Needling

Vertically 0.5–1.5 cun

Actions/Indications

- Opens the channel
- Clears Wind and Heat

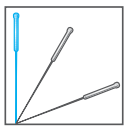

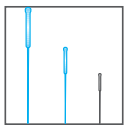

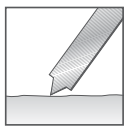

Special features

Jing-river point, Fire point, sedation point.

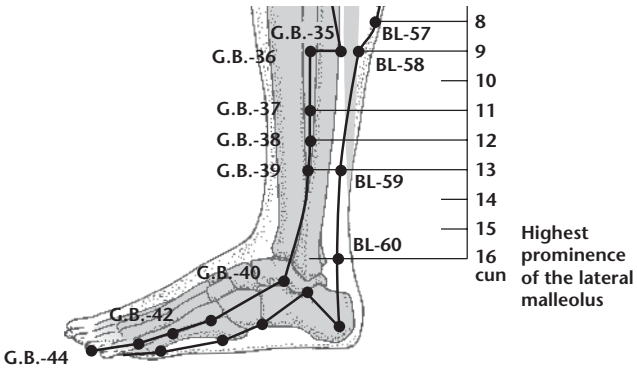

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 437

Hanging Bell XUANZHONG G.B.-39

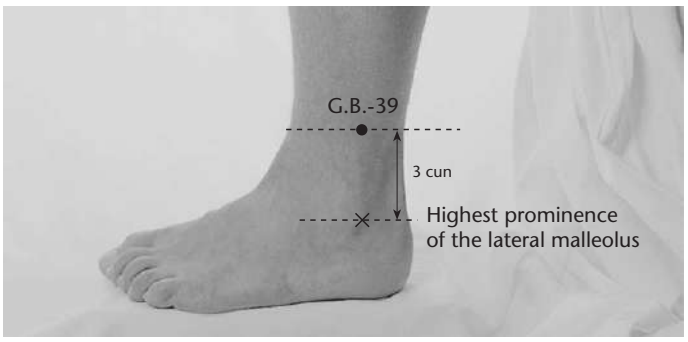

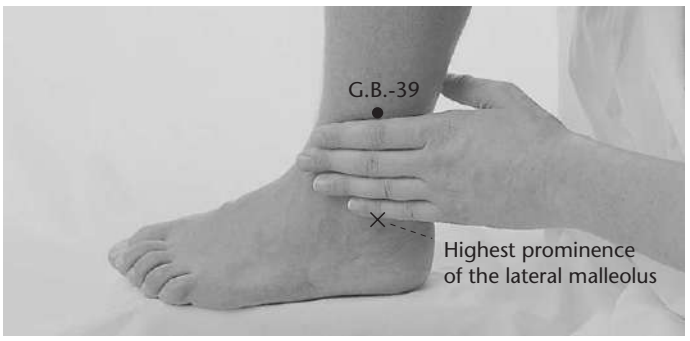

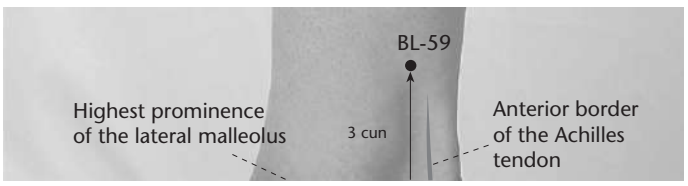

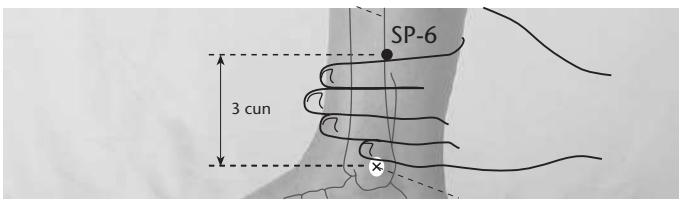

Location

3 cun proximal to the highest prominence of the lateral malleolus, on the **anterior** border of the fibula.

How to find

From the highest prominence of the lateral malleolus (→ 3.6.2), measure 3 cun (1 handbreadth) in a proximal direction and locate **G.B.-39** in a depression on the **anterior** border of the fibula. According to some authors, it is located between the posterior border of the fibula and the tendons of the peroneus longus and brevis musles. When in doubt, choose the more pressure-sensitive point.

Located on the same level but 3 cun proximal to → **BL-60** (depression between the malleolus and the Achilles tendon) is → **BL-59**. In a comparable position but on the medial aspect of the leg is → **SP-6** (3 cun proximal to the highest prominence of the medial malleolus, meeting point of the three leg Yin channels).

Needling

Vertically 1–1.5 cun

Actions/Indications

- Opens the channel, benefits the tendons and bones
- Benefits the Marrow, clears Wind-Dampness
- Clears Heat from the Gall Bladder

Special features

Hui-meeting point of the Marrow. Important distal point for disorders of the cervical spine.

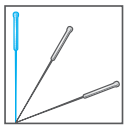
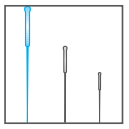

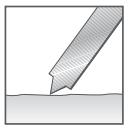

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 438

G.B.-40 *Mound of Ruins* QIUXU

Location

In the depression anterior and inferior to the lateral malleolus, lateral to the tendons of the extensor digitorum longus muscle.

How to find

For easier location, ask the patient to flex their ankle at a 90° angle. From the lateral malleolus (→ 3.6.2), let the palpating finger glide into a well-defined depression anterior and inferior to the malleolus. By lifting the toes, the tendons of the extensor digitorum longus muscle as well as the depression (**G.B.-40**) will become more pronounced. **For orientation: G.B.-40** is located at the junction of a vertical line along the anterior border of the lateral malleolus and a horizontal line along its lower border.
→ **SP-5** is located in a comparable position but on the medial aspect of the ankle (in the depression anterior and inferior to the medial malleolus). On a line connecting → **SP-5** and **G.B.-40**, → **LIV-4** is located medial to and → **ST-41** is located lateral to the tendon of the extensor hallucis longus muscle, which runs to the big toe.

Needling

0.5–1.5 cun vertically or slightly obliquely in the area of the fibulotarsal ligaments towards the medial arch of the foot.

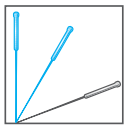

Actions/Indications

- Opens the channel, benefits the joints
- Spreads Liver Qi, clears Heat and Dampness from the Gall Bladder

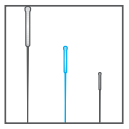

Special features

Yuan-source point

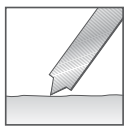

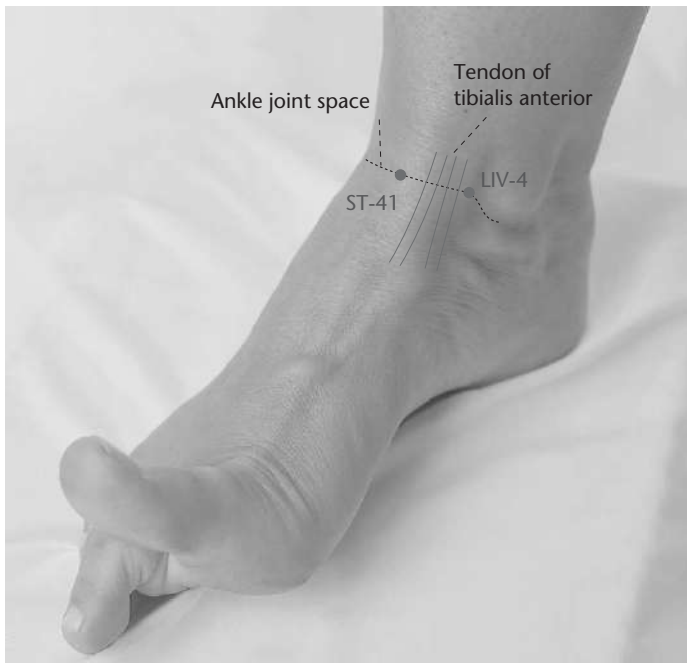

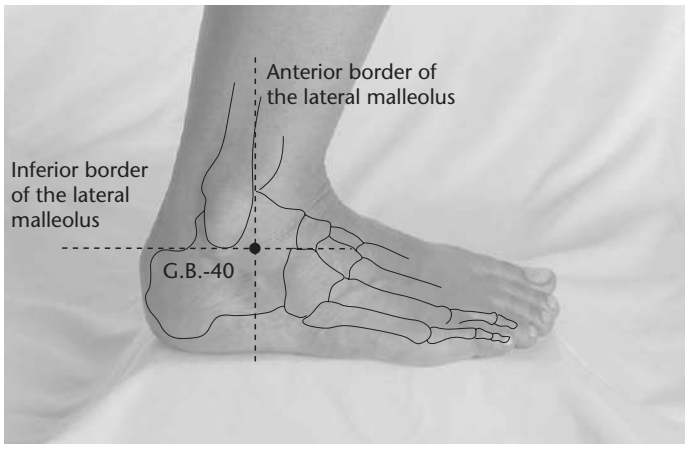

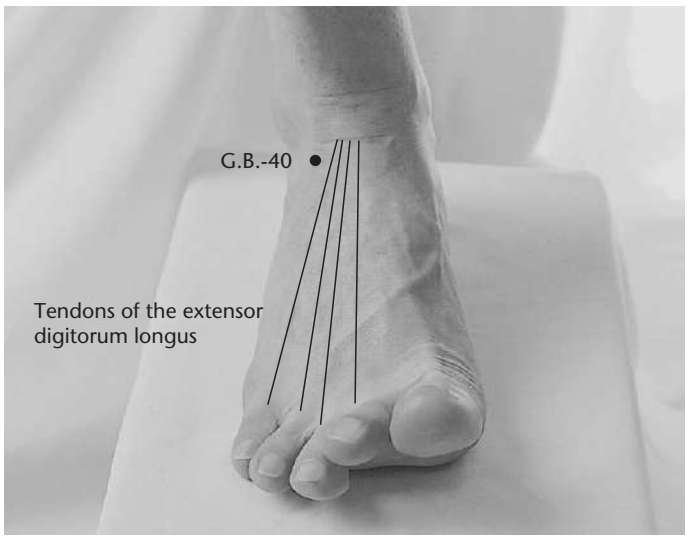

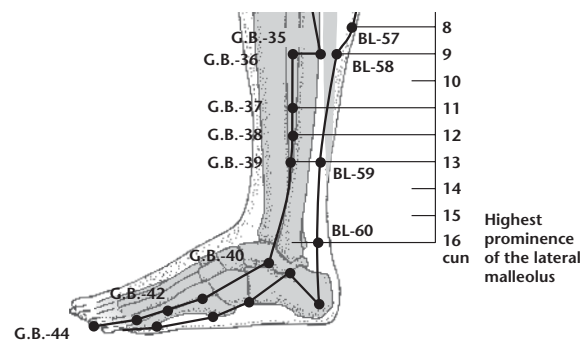

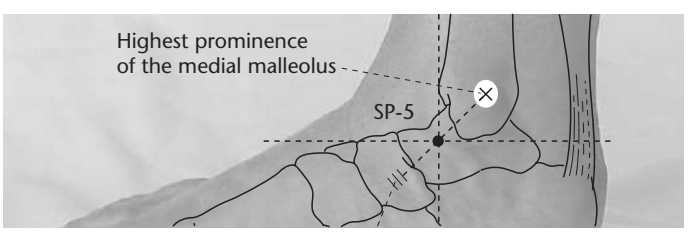

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 439

Foot Governor of Tears ZULINQI G.B.-41

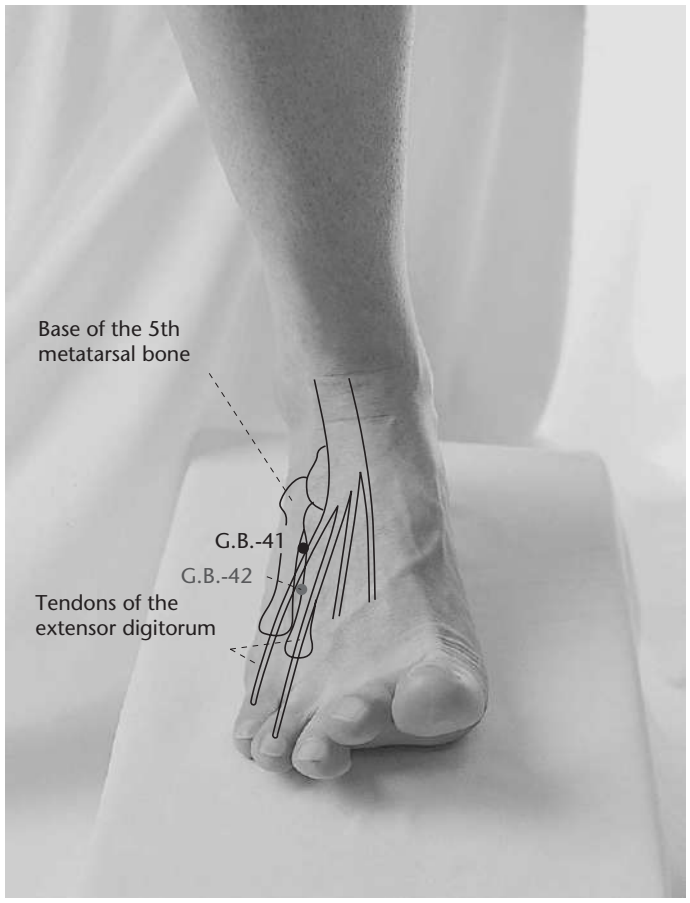

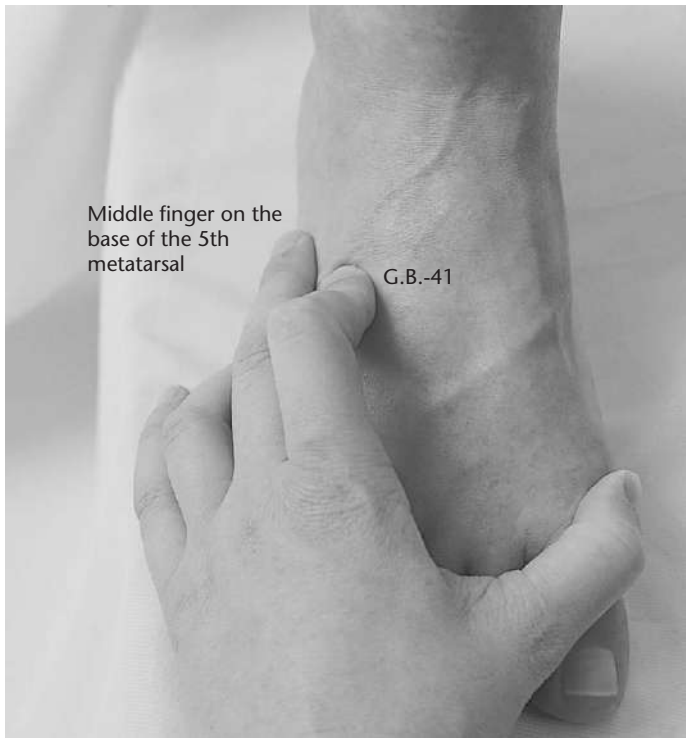

Location

In the depression at the junction of the shafts and the bases of the 4th and 5th metatarsal bones, lateral to the tendon of the extensor digitorum longus muscle.

How to find

Palpate for the distinct tuberosity of the 5th metatarsal bone (→ 3.6.2) on the lateral aspect of the midfoot. With the palpating finger, glide on from the tuberosity to the dorsum of the foot, into the groove between the 4th and 5th metatarsal bones. There, palpate in a distal direction. **G.B.-41** is located at the junction of the heads and the shafts of the two bones. **Or:** Ask the patient to abduct their toes so that the branch of the tendon of the extensor digitorum longus muscle extending to the little toe becomes more pronounced. Then palpate in the groove between the 4th and 5th metatarsal bones from distal to proximal to where the tendon crosses the groove. → **G.B.-42** is located medial (or anterior) to it, while **G.B.-41** is located in the depression lateral (or posterior) to the tendon.

Needling

Perpendicularly or obliquely approximately 0.3–0.8 cun

Actions/Indications

- Spreads Liver Qi, benefits the thorax and the lateral costal region, transforms Phlegm, dissipates nodules, benefits the breasts
- Clears the head, benefits the eyes

Special features

Shu-stream point, Wood point, opening point of the *dai mai*, *ben* point (Five Phases), exit point.

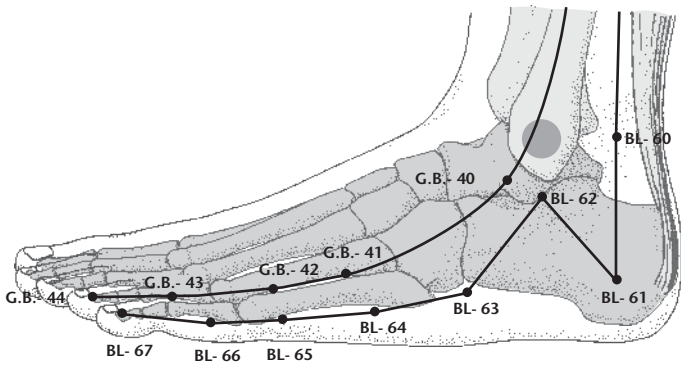

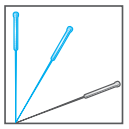
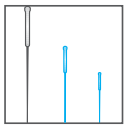

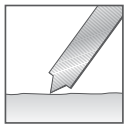

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 440

G.B.-42 *Earth Five Meetings* DIWUHUI

Location

Between the 4th and 5th metatarsal bones, proximal to the metatarsophalangeal joints and medial to the tendon of the extensor digitorum longus muscle.

How to find

From the space between the 4th and 5th toes, palpate towards the ankle. Locate **G.B.-42** in the groove between the two bones, proximal to the metatarsophalangeal joints (and proximal to the heads of the two metatarsal bones). **Or:** Ask the patient to abduct their toes so that the branch of the tendon of the extensor digitorum longus muscle extending to the little toe becomes more pronounced. Then palpate from the space between the toes along the groove between the 4th and 5th metatarsal bones from distal to proximal to where the tendon crosses the groove. **G.B.-42** is located medial (or anterior) to the tendon.

→ **G.B.-41** is located lateral (or posterior) to the tendon when continuing to palpate along the groove. → **BL-65** is located approximately on the same level on the lateral border of the foot, proximal to the head of the 5th metatarsal bone.

Needling

Vertically or obliquely 0.3–0.8 cun

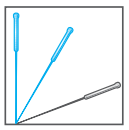

Actions/Indications

- Moves Liver Qi
- Clears Heat from the Gall Bladder

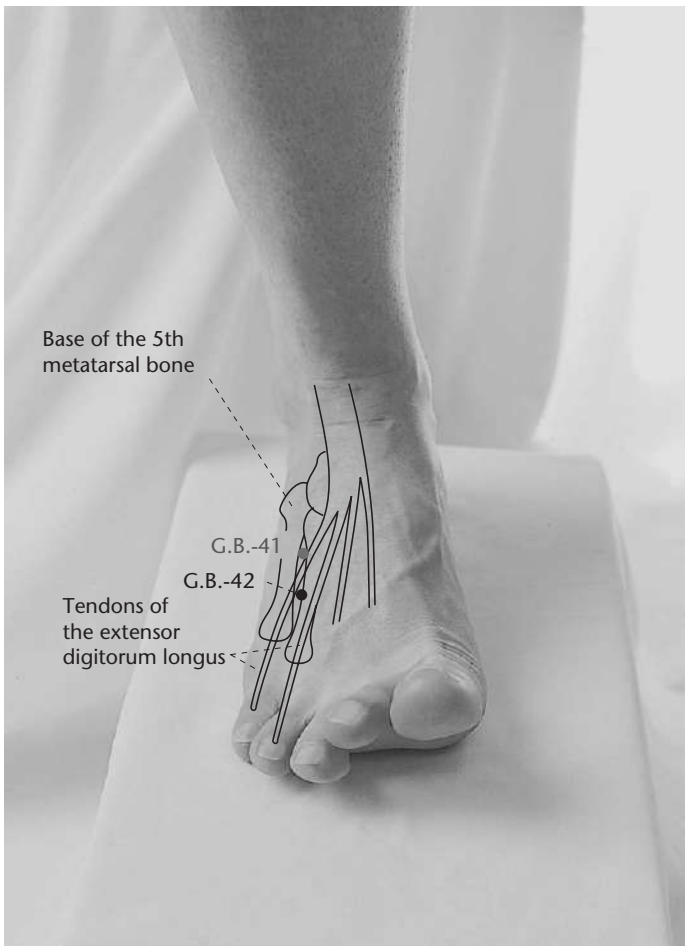

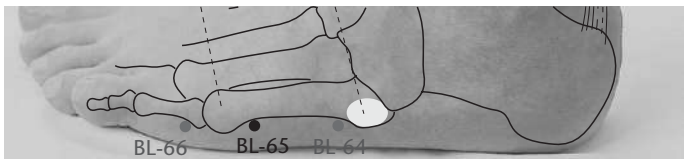

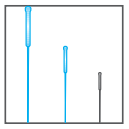

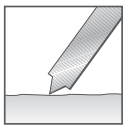

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 441

Clamped Stream JIAXI/XIAXI G.B.-43

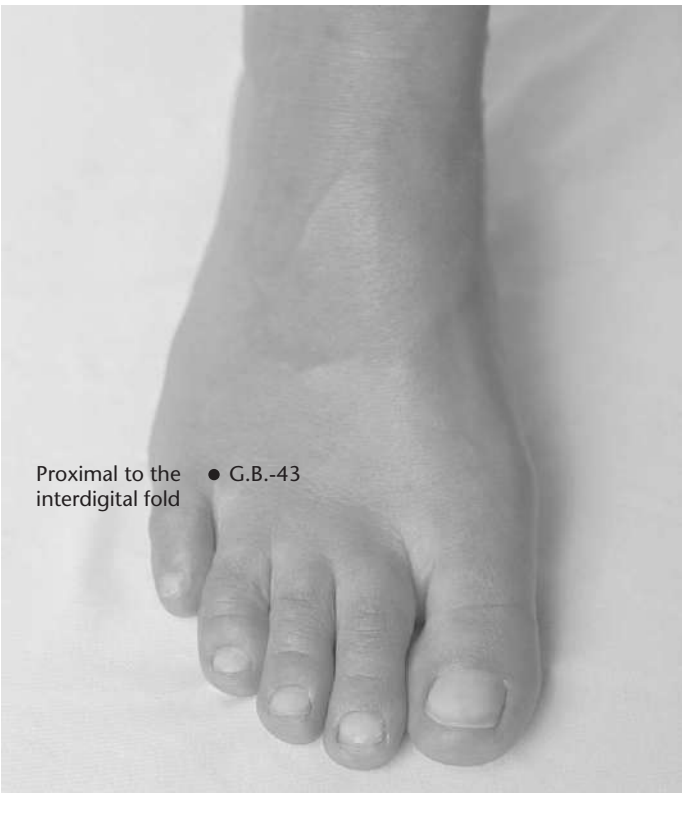

Location

Between the 4th and 5th toes, proximal to the margin of the interdigital web.

How to find

Locate the interdigital web between the 4th and 5th toes. Then locate **G.B.-43** slightly proximal to the margin of the web. **G.B.-43** as well as → **LIV-2** and → **ST-44** are part of → **Ex-UE-10** (*bafeng*: proximal to the margins of the webs between the toes). → **T.B.-2** is located in a comparable position on the hand, between the ring finger and little finger. It is also part of → **Ex-UE-9** (*baxie*).

Needling

Up to 0.5 cun vertically or 1 cun obliquely in a proximal direction.

Actions/Indications

- Clears Heat, Damp-Heat and Wind, especially from the opposite end of the channel
- Calms Liver Yang

Special features

Ying-spring point, Water point, tonification point.

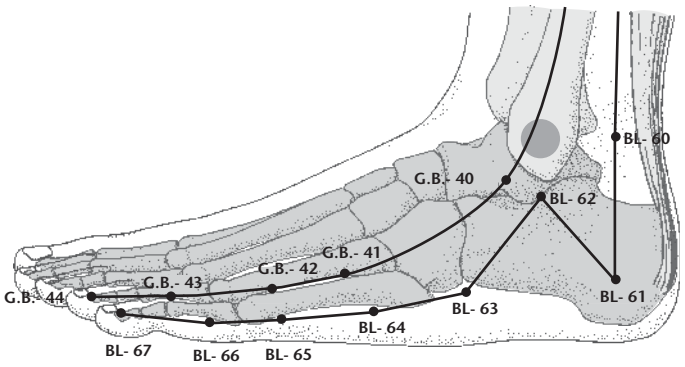

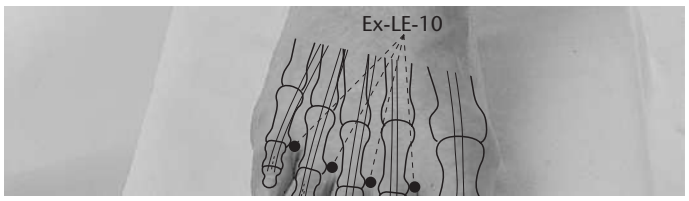

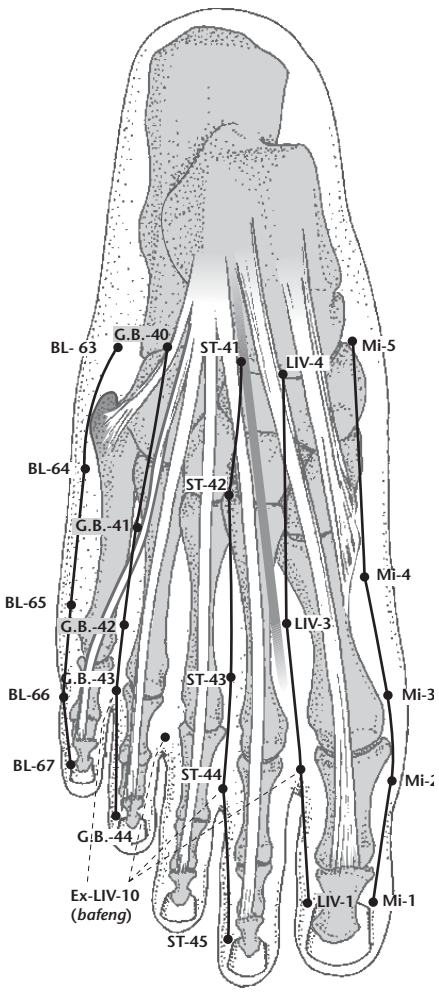

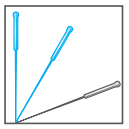
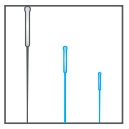

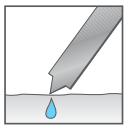
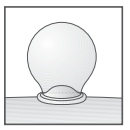

Ch04.11-F10028.qxd 2/23/08 11:25 AM Page 442

G.B.-44 *Foot Portal Yin* ZUQIAOYIN

Location
On the 4th toe, 0.1 cun from the lateral corner of the nail.

How to find
G.B.-44 is located at the junction of two tangents along the proximal and lateral borders of the 4th toe.
→ **BL-67** is located at the lateral corner of the little toe.

Needling
0.1 cun vertically or 0.2 cun obliquely in a proximal direction or prick to bleed. Caution: Painful point.

Actions/Indications
- Clears Wind, Heat and Fire
- Harmonises the Liver and Gall Bladder
- Calms the *shen*

Special features
Jing-well point, Metal point

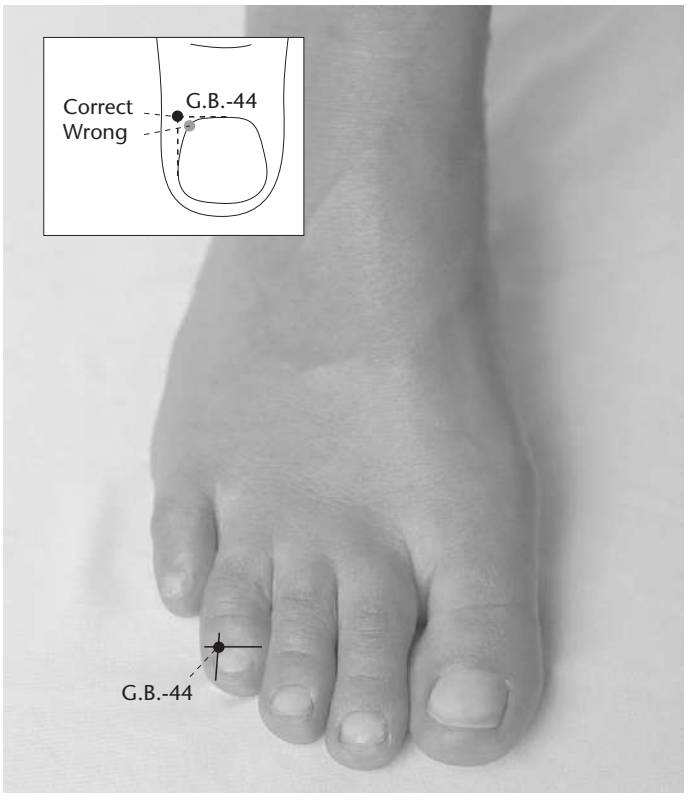

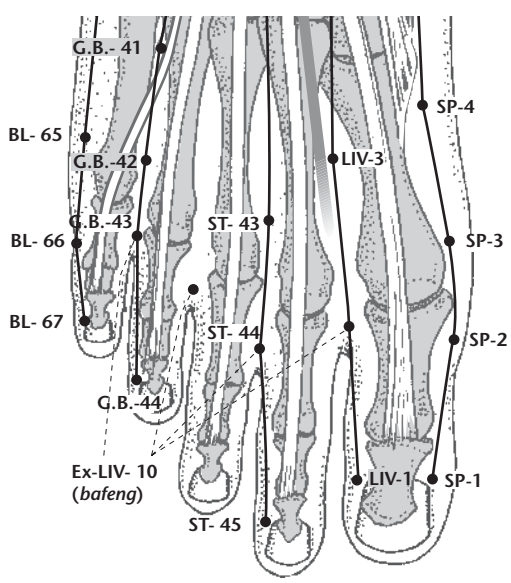

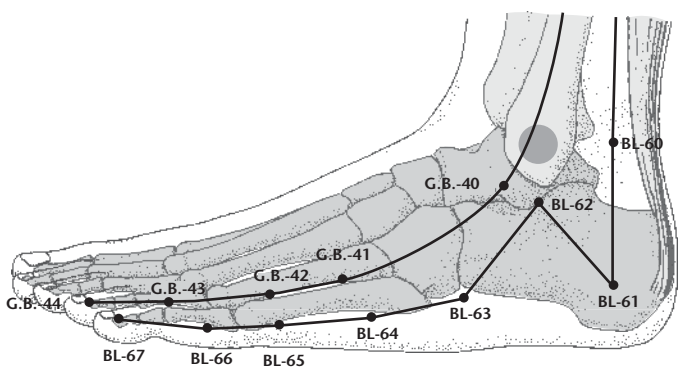

Ch04.12-F10028.qxd 2/23/08 11:31 AM Page 443

4.12 The Liver Channel System – Foot *jueyin* (*zu jue yin jing luo*)

4.12.1 The Liver Primary Channel (*zu jue yin jing*)

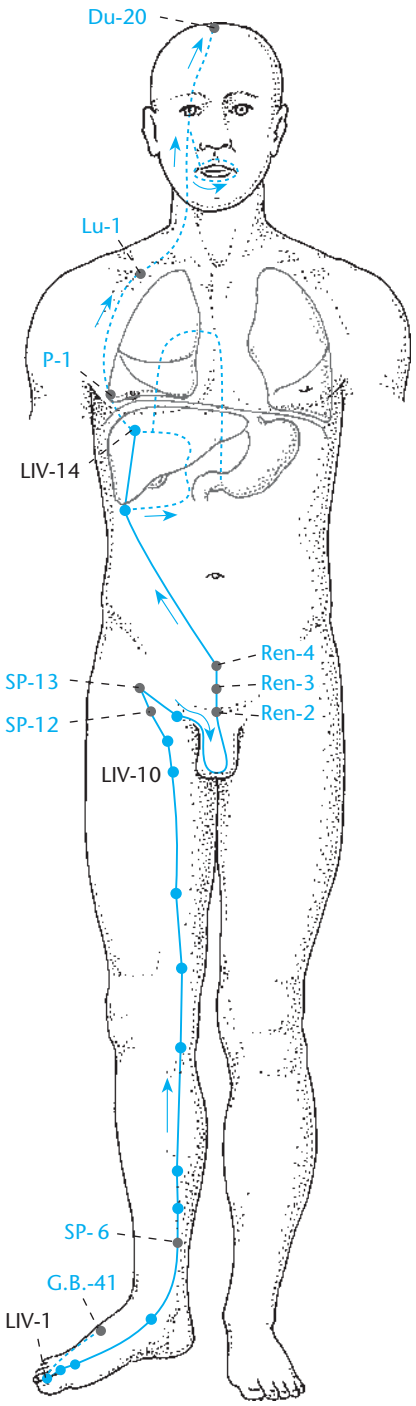

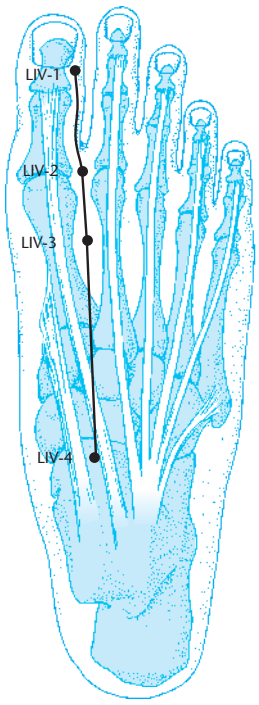

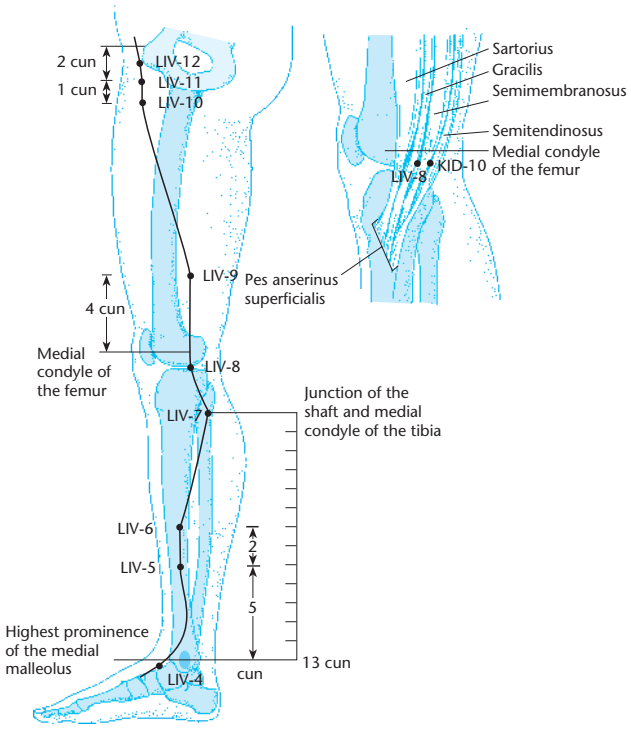

Pathway

The Liver primary channel begins at **LIV-1** (*dadun*) on the lateral corner of the nail of the big toe. This point is reached by a small branch of the G.B. channel, which separates from the Gall Bladder primary channel on the dorsum of the foot at → **G.B.-41** (*zulinqi*) (foot Yin–Yang connection of the third great circuit).

Ch04.12-F10028.qxd 2/23/08 11:31 AM Page 444

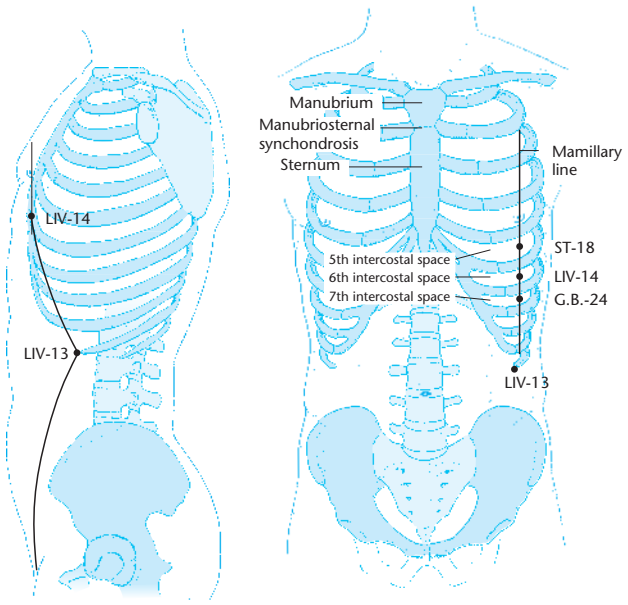

The **external pathway** runs proximally between the 1st and 2nd metatarsal bones

- ➡ ascends anterior to the medial malleolus
- ➡ crosses the SP channel at → **SP-6** (*yinlingquan*) on the medial aspect of the lower leg
- ➡ ascends the medial aspect of the leg towards the knee anterior to the SP channel
- ➡ continues along the medial aspect of the thigh to the pubic region, where it passes → **SP-12** (*chongmen*) and → **SP-13** (*fushe*)
- ➡ circles the external genitalia
- ➡ ascends the lower abdomen, passing → **Ren-2** (*qugu*), → **Ren-3** (*zhongji*) and → **Ren-4** (*guanyuan*)
- ➡ obliquely traverses the abdomen to **LIV-13** (*zhangmen*) at the free end of the 11th rib and to **LIV-14** (*qimen*), where the **external pathway** terminates.

At **LIV-13**, the channel enters the abdomen, marking the beginning of the **internal pathway**. It

- ➡ circles the Stomach (*wei*)
- ➡ connects with its pertaining *zang*-Organ, the Liver (*gan*) and its paired *fu*-Organ, the Gall Bladder (*dan*)
- ➡ penetrates the diaphragm
- ➡ spreads in the lateral hypochondrium and thoracic region.

The channel then ascends along the posterior aspect of the trachea to the throat and nasopharynx and connects with the eye system and the brain. It crosses the forehead and ascends to the vertex, where it connects with the extraordinary vessel *du mai* at → **Du-20** (*baihui*).

An **internal branch** descends from the maxillary sinus to the cheek and circles the **inner surface** of the lips.

An **internal branch** emerges from the Liver, penetrates the diaphragm, disperses in the Lung (*fei*) and meets the Lung primary channel (deep Yin–Yin connection), closing the first circuit of the Nutritive Qi (*ying qi*) (→ 1.1.4). This branch further connects with the Pericardium primary channel beneath → **P-1** (*tianchi*) (hand–foot pairing of the third great circuit: Yin axes, *jue yin*).

Clinical importance (→ 1.2)

Exterior (*biao*) signs and symptoms: Headaches, dizziness, blurred vision, tinnitus, fever, spastic extremities

Interior (*li*) or *zangfu*-Organ signs and symptoms: Distension, fullness and pain in the costal region, tightness and fullness in the chest, abdominal pain, vomiting, jaundice, diarrhoea, *shan* disorder, enuresis, urinary retention, yellow urine

Connections of the Liver primary channel (→ 1.2)

Connections with other channels

Gall Bladder primary channel (*zu shao yang jing*)

Connection: Foot Yin–Yang pairing of the third great circuit
Location: G.B.-41 → LIV-1 (on the foot)
Circulation: Circadian (according to the Organ clock)
Importance: Exterior–Interior relationship

Pericardium primary channel (*shou jue yin jing*)

Connection: Paired according to the six-channel theory (hand–foot pairing): *jue yin* (Yin axes of the third great circuit)
Location: LIV → P. An internal branch originating in the Liver penetrates the diaphragm and connects with the Pericardium primary channel beneath → P-1.
Circulation: Non-circadian (not according to the Organ clock)
Importance: Above–below relationship

Lung primary channel (*shou tai yin jing*)

Connection: Deep Yin–Yin connection
Location: LIV → LU (in the thorax). An internal branch originating in the Liver penetrates the diaphragm and disperses in the Lung to connect with the Lung primary channel.
Circulation: Circadian (according to the Organ clock)
Importance: The Lung primary channel receives part of its Nutritive Qi (*ying qi*) from the Liver primary channel (first circulation of the *ying qi* → 1.1.4).

Connections with zangfu-Organ systems

Stomach (*wei*), **Liver (*gan*)**, **Gall Bladder (*dan*)**, Lung (*fei*)

Ch04.12-F10028.qxd 2/23/08 11:31 AM Page 445

4.12.2 The Liver Divergent Channel (*zu jue yin jing bie*)

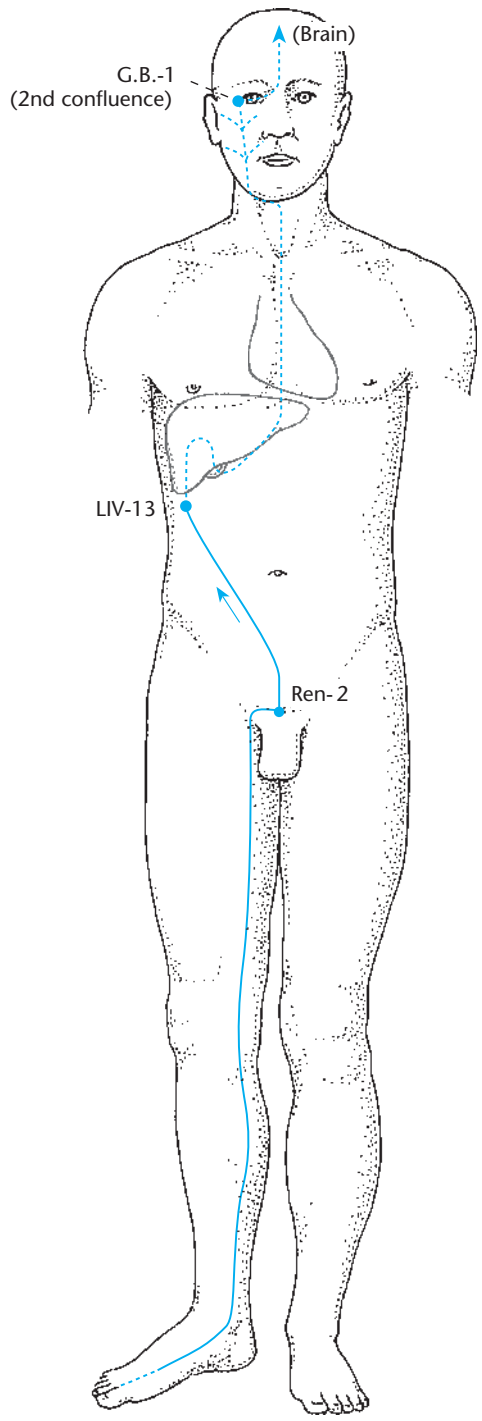

Pathway

The Liver divergent channel separates from the Liver primary channel on the dorsum of the foot

- ➡ ascends the medial aspect of the leg to the inguinal region
- ➡ meets the Gall Bladder divergent channel near → **Ren-2** (*qugu*)
- ➡ traverses the abdomen and enters internally at **LIV-13** (*zhangmen*)
- ➡ continues inside the thoracic cavity, spreads in the Liver (*gan*), connects with the Gall Bladder (*dan*) and runs to the Heart (*xin*)
- ➡ ascends to the neck
- ➡ emerges at the mandibular angle and disperses over the face
- ➡ connects with the Gall Bladder primary channel and the Gall Bladder divergent channel at → **G.B.-1** (*tongziliao*) on the outer canthus of the eye to form one of the six *he*-confluences (here: G.B./LIV as 2nd confluence → 1.3)
- ➡ continues to the eye system and ends in the Brain.

Clinical importance

- Strengthens the relationship between the Liver and the Gall Bladder (*zangfu*-Organ systems). Points on the Liver primary channel can therefore be used to treat disorders of the Gall Bladder and vice versa.
- Based on the pathways of the Gall Bladder and Liver divergent channels, points on both channels can be used for disorders of the hips and the lower extremities as well as for disorders of the eye system.

Ch04.12-F10028.qxd 2/23/08 11:31 AM Page 446

4.12.3 The Liver Sinew Channel (*zu jue yin jing jin*)

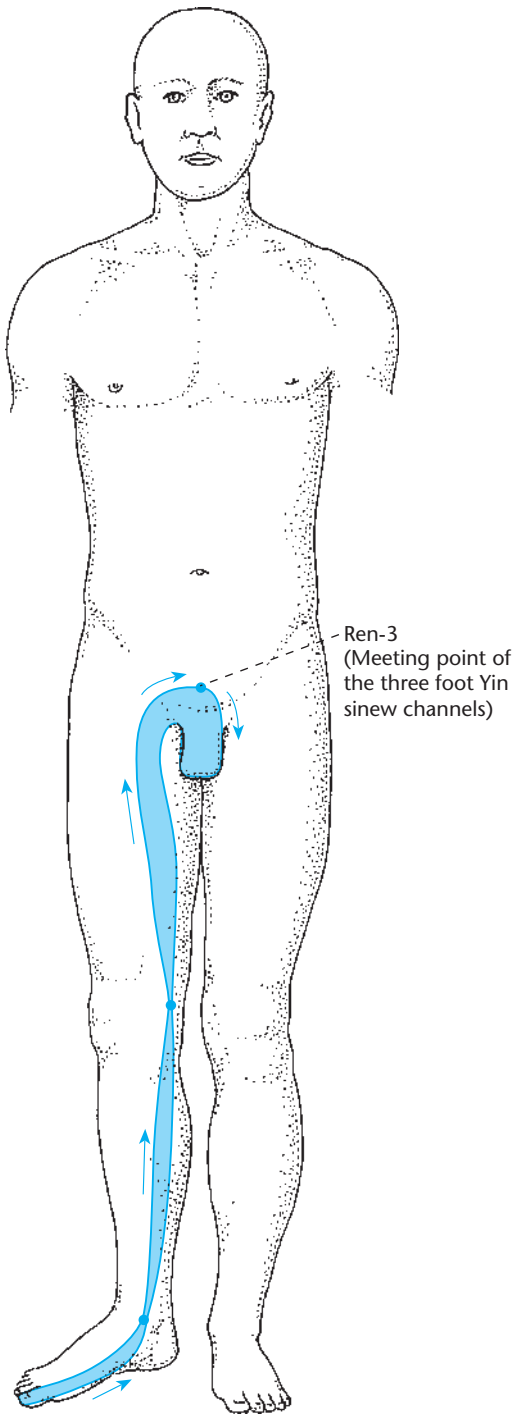

Pathway

The Liver sinew channel begins on the dorsum of the big toe

- ➡ follows the dorsolateral aspect of the big toe to the anterior aspect of the medial malleolus, where it binds (*jie*)
- ➡ ascends the medial aspect of the tibia and binds (*jie*) at the medial aspect of the knee
- ➡ further ascends the medial aspect of the thigh between the Kidney and Spleen sinew channels
- ➡ reaches the inguinal region, meeting the other foot Yin sinew channels at → **Ren-3** (*zhongji*) and continues to the genitalia.

Clinical importance

Pathology: Stiffness, tension and distending sensations along the big toe. Pain around the medial malleolus and medial aspect of the knee. Pain and muscular tension along the medial aspect of the thigh. Disorders of the genital region.

Indication: Pain, tension, muscle contractions and distending sensations along the pathway of the channel. Disorders of the genital region.

Ch04.12-F10028.qxd 2/23/08 11:31 AM Page 447

4.12.4 The Liver *luo*-Connecting Vessel System (*zu jue yin luo mai*)

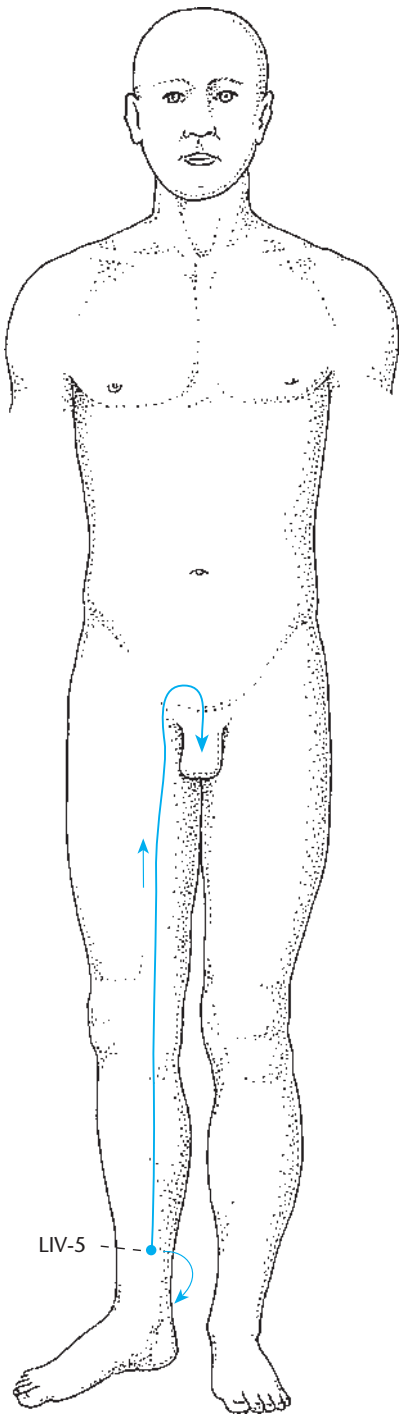

Pathway

The Liver *luo*-connecting vessel system separates from the Liver primary channel at its *luo* point **LIV-5** (*ligou*) (→ 8.1.2). It forms a three-dimensional reticular network, dividing into multiple branches and sub-branches (*sun luo, fu luo, xue luo* → 1.5) within the surrounding tissue.

- Horizontal divisions run to the Interiorly–Exteriorly paired Gall Bladder primary channel; according to some schools of thought (for example, Ngyen Van Nghi → Appendix), they travel as a **transverse** Liver *luo*-connecting vessel to the *yuan*-source point → **G.B.-40** (*qiuxu*).
- **A longitudinal division** ascends the medial aspect of the leg to the genital region, where it branches out.

Clinical importance (→ 8.1.2)

Pathology

Counterflow Qi: Swellings and pain of the testicles and scrotum, hernia
Excess (*shi*): Priapism
Deficiency (*xu*): Itching of the external genitalia

4.12.5 Cutaneous Region (*jue yin pi bu*)

See description and figures → 1.6.

4.12.6 Points of the Liver Primary Channel (Overview)

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): LIV-3** (*taichong*) ■■
- ***Luo*-connecting point (→ 8.1.2.): LIV-5** (*ligou*) ■■
- ***Xi*-cleft point (→ 8.1.3): LIV-6** (*zhongdu*) ■■
- **Associated Back-*shu* point (→ 8.1.4): BL-18** (*ganshu*) ■■
- **Associated Front-*mu* point (→ 8.1.5): LIV-14** (*qimen*) ■■
- **Five *shu*-transporting points (→ 8.1.6):**
  jing-well point (Wood): **LIV-1** (*dadun*) ■■, *ben* point (Five Phases)
  ying-spring point (Fire): **LIV-2** (*xingjian*) ■■
  shu-stream point (Earth): **LIV-3** (*taichong*) ■■
  jing-river point (Metal): **LIV-4** (*zhongfeng*)
  he-sea point (Water): **LIV-8** (*ququan*) ■■, tonification point
- ***Hui*-meeting point (→ 8.1.7) of the *zang*-Organs: LIV-13** (*zhangmen*) ■■
- **Opening point (→ 8.1.8): –**
- **Lower *he*-sea point (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10):**
  – with the G.B. channel and the *dai mai**: **LIV-13** (*zhangmen*) ■■

*Mentioned by only some authors

Ch04.12-F10028.qxd 2/23/08 11:31 AM Page 448

– with the *yin wei mai* and the ST channel: **LIV-14** (*qimen*) ■■
– of other channels with the LIV channel: **SP-6, SP-12, SP-13, Ren-2, Ren-3, Ren-4, P-1, (Du-20*)**

- **Gao Wu command point (→ 8.1.11): –**
- **Window of Heaven point (→ 8.1.12): –**
- **Points of the Four Seas (→ 8.1.13): –**
- **Ma Dan Yang Heavenly Star point (→ 8.1.14): LIV-3** (*taichong*) ■■
- **Sun Si Miao Ghost point (→ 8.1.15): –**
- **Other functional points:**
  – Front-*mu* point of the Spleen: **LIV-13** (*zhangmen*) ■■

Points according to region

- **Local points** (→ 8.2.1): Liver – **LIV-13** (*zhangmen*) ■■; knee – **LIV-7** (*xiguan*), **LIV-8** (*ququan*) ■■; foot – **LIV-4** (*zhongfeng*)
- **Adjacent points (→ 8.2.1):** Stomach/Spleen – **LIV-13** (*zhangmen*) ■■; hypochondrium – **LIV-13** (*zhangmen*) ■■, **LIV-14** (*qimen*)
- **Distal points** (→ 8.2.1): vertex – **LIV-3** (*taichong*) ■■; eyes – **LIV-2** (*xingjian*) ■■, **LIV-3** (*taichong*) ■■; Liver – **LIV-3** (*taichong*) ■■; Gall Bladder – **LIV-3** (*taichong*) ■■; urogenital region – **LIV-3** (*taichong*) ■■; genital region – **LIV-5** (*ligou*) ■■; lower abdomen – **LIV-8** (*ququan*) ■■

Specific points according to the channel pathway (in numeric order)

- **LIV-1** (*dadun*) ■■: *jing*-well point (Wood), *ben* point (Five Phases) (→ 8.1.6)
- **LIV-2** (*xingjian*) ■■: *ying*-spring point (Fire) (→ 8.1.6), sedation point; distal point for the eyes (→ 8.2.1)
- **LIV-3** (*taichong*) ■■: *yuan*-source point (→ 8.1.1); *shu*-stream point (Earth) (→ 8.1.6); Ma Dan Yang Heavenly Star point (→ 8.1.4); distal point for the vertex, eyes, Liver, Gall Bladder and urogenital region (→ 8.2.1)
- **LIV-4** (*zhongfeng*): *jing*-river point (Metal) (→ 8.1.6); local point for the foot (→ 8.2.1)
- **LIV-5** (*ligou*) ■■: *luo*-connecting point (→ 8.1.2); distal point for the genital region (→ 8.2.1)
- **LIV-6** (*zhongdu*) ■■: *xi*-cleft point (→ 8.1.3)
- **LIV-7** (*xiguan*): local point for the knee (→ 8.2.1)
- **LIV-8** (*ququan*) ■■: *he*-sea point (Water) (→ 8.1.6); tonification point; distal point for the abdomen (→ 8.2.1); local point for the knee (→ 8.2.1)
- **LIV-13** (*zhangmen*) ■■: Front-*mu* point of the Spleen (→ 8.1.5); *hui*-meeting point of the *zang*-Organs (→ 8.1.7); *jiaohui*-meeting point with the G.B. channel and the (*dai mai**) (→ 8.1.10); adjacent point for the Spleen/Stomach and the hypochondrium (→ 8.2.1)
- **LIV-14** (*qimen*) ■■: associated Front-*mu* point(→ 8.1.5); *jiaohui*-meeting point with the *yin wei mai* and the SP channel (→ 8.1.10); regional point for the hypochondrium (→ 8.2.1)

* Mentioned by only some authors

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 449

Big Mound DADUN LIV-1

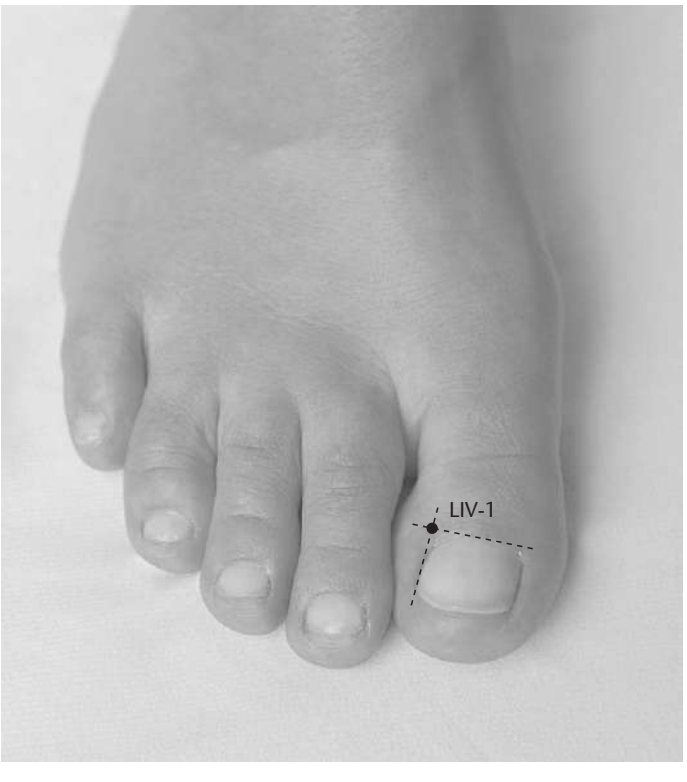

Location

On the lateral aspect of the big toe, 0.1 cun proximal and lateral to the corner of the nail.

How to find

LIV-1 is located at the junction of two tangents along the proximal and lateral borders of the big toe, 0.1 cun from the actual margin of the nail. → **SP-1** is located at the medial corner of the big toe.

Needling

Vertically or obliquely 0.1–0.2 cun. Avoid needling into the perionychium. For excess conditions, prick to bleed. Caution: Painful point.

Actions/Indications

- Regulates the Lower Burner, the genitals and Liver Qi
- Eliminates Dampness
- Opens the sensory orifices

Special features

Jing-well point, Wood point, *ben* point (Five Phases), entry point.

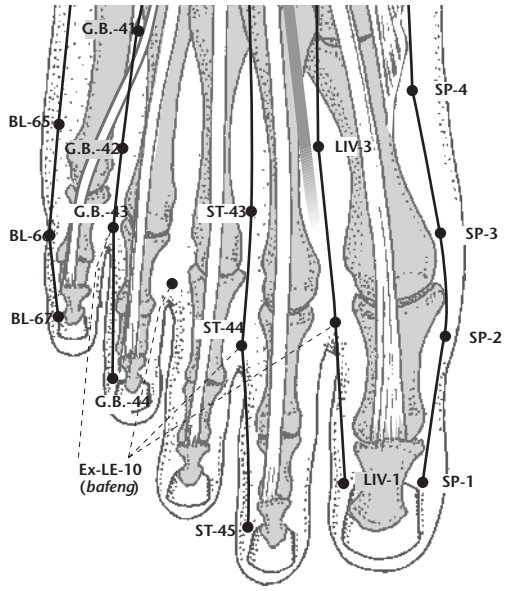

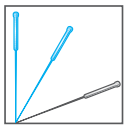
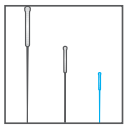

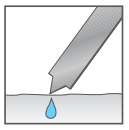

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 450

LIV-2 *Moving Between* XINGJIAN

Location

Between the 1st and 2nd toes, proximal to the margin of the interdigital web.

How to find

Locate **LIV-2** slightly proximal to the margin of the interdigital web between the 1st and 2nd toes.

Located in comparable positions are → **ST-44** (between the 2nd/3rd toes) and → **G.B.-43** (between the 4th/5th toes). → **LIV-2,** → **ST-44** and → **G.B.-43** are all part of the extra point → **Ex-LE-10** (*bafeng*). Its counterpart on the hand is the extra point → **Ex-UE-9** (*baxie*).

Needling

0.3–0.5 cun obliquely towards the heel or vertically.

Actions/Indications

- Clears Liver Fire, spreads Liver Qi, calms (internal) Liver Wind, clears Heat and Blood Heat, stops bleeding, benefits the Lower Burner

Special features

Ying-spring point, Fire point, sedation point. Major point for excess syndromes of the Liver (especially Liver Fire).

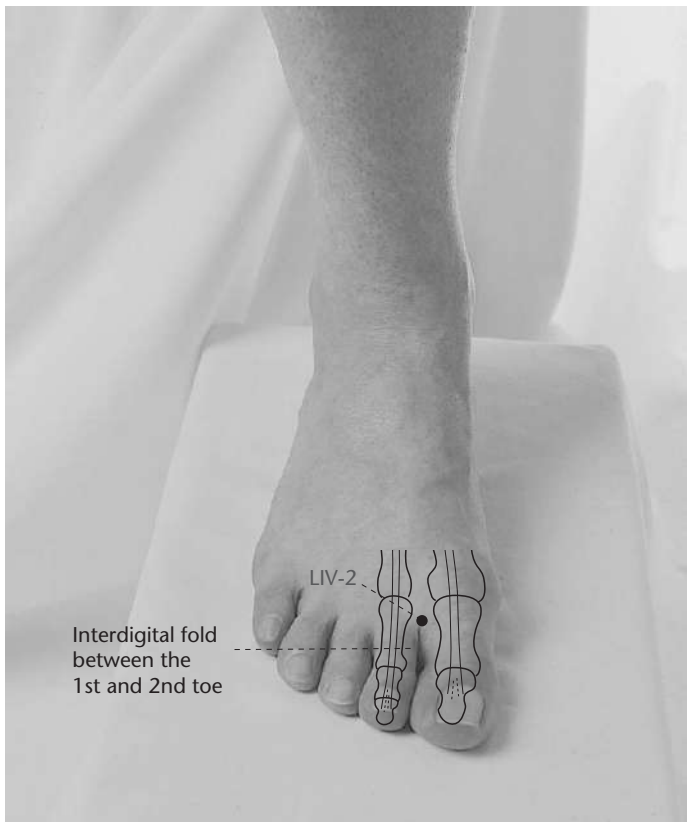

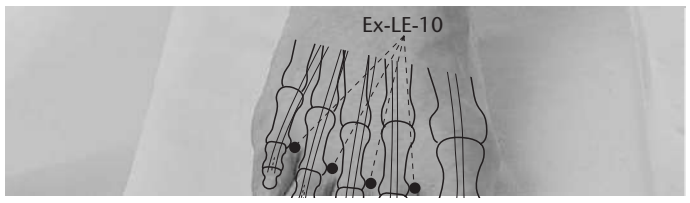

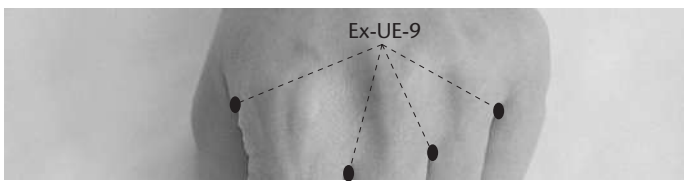

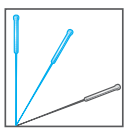

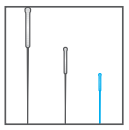

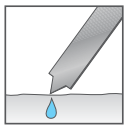

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 451

Great Rushing TAICHONG LIV-3

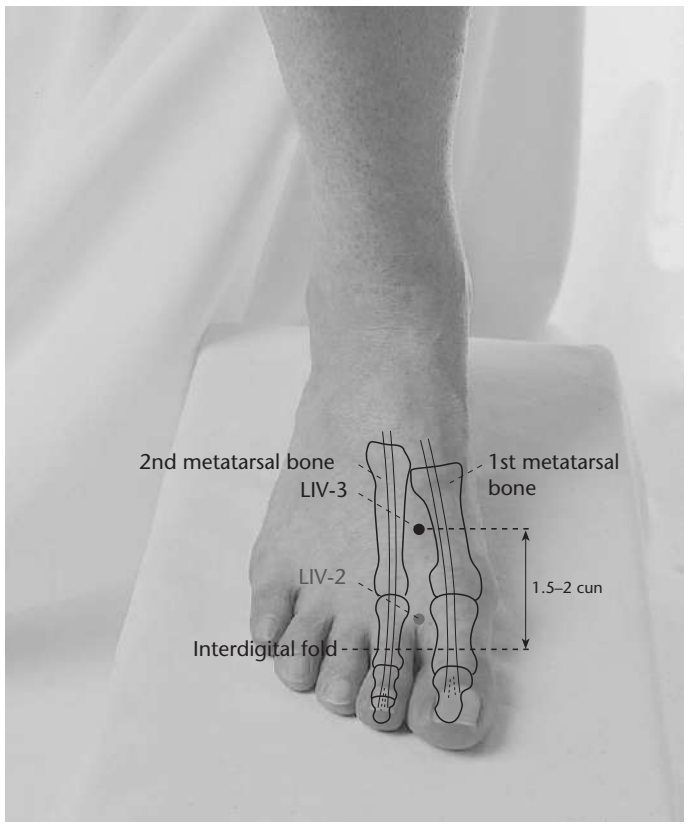

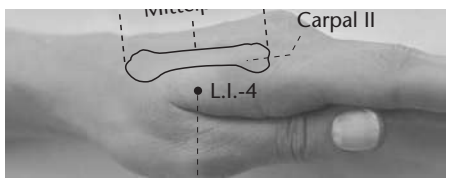

Location

On the dorsum of the foot, between the 1st and 2nd metatarsal bones, in the depression proximal to the metatarsophalangeal joints and the proximal angle between the two bones.

How to find

From the web between the 1st and 2nd toes, palpate proximally past the metatarsophalangeal joints and along the groove between the 1st and 2nd metatarsal bones, until you reach the widest and deepest part of the groove. There, locate **LIV-3**, which is often sensitive to pressure. To double-check: When continuing to palpate along the groove, it will become more narrow and shallow again.

Located in a similar position on the hand is → **L.I.-4** (between the 1st and 2nd metacarpal bones).

Needling

Vertically or slightly obliquely 0.5–1 cun. Caution during pregnancy.

Actions/Indications

- Spreads Liver Qi, clears the head, benefits the eyes, nourishes Liver Blood and Liver Yin, calms spasms, alleviates pain, regulates menstruation and the Lower Burner
- Eliminates (internal) Wind, pacifies Liver Yang

Special features

Yuan-source point, *shu*-stream point, Earth point, Ma Dan Yang Heavenly Star point (later added by Xu Feng). Major point for spreading Liver Qi.

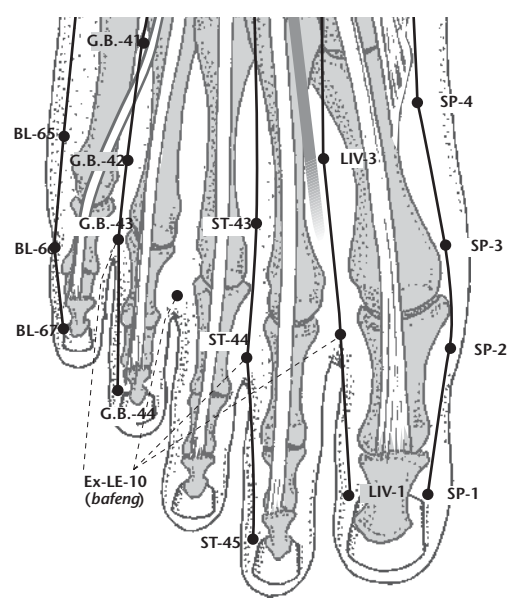

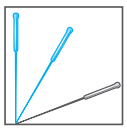
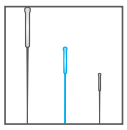

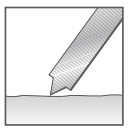

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 452

LIV-4 *Middle Seal* ZHONGFENG

Location

1 cun anterior to the prominence of the medial malleolus, medial to the tendon of the tibialis anterior muscle.

How to find

Even with slight dorsiflexion of the ankle, the tendon of the tibialis anterior muscle will become more pronounced on the medial aspect of the ankle. **LIV-4** is located between the tendon and the prominence of the medial malleolus, in a depression over the palpable joint space.

→ **ST-41** is also located over the joint space (on the midpoint of its anterior aspect); → **SP-5** lies only slightly distal to → **LIV-4** at the junction of a vertical line along the anterior border and a horizontal line along the lower border of the medial malleolus.

Needling

Vertically 0.3–0.5 cun

Actions/Indications

- Opens the channel, promotes the flow of Liver Qi, regulates the Lower Burner

Special features

Jing-river point, Metal point.

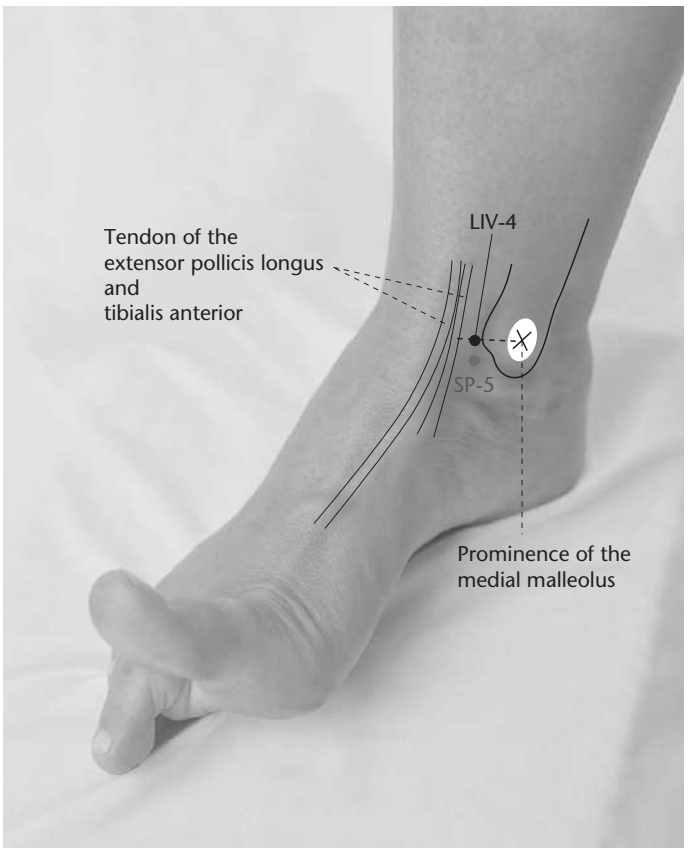

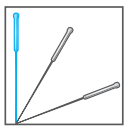

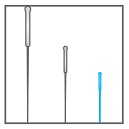

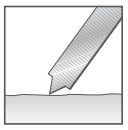

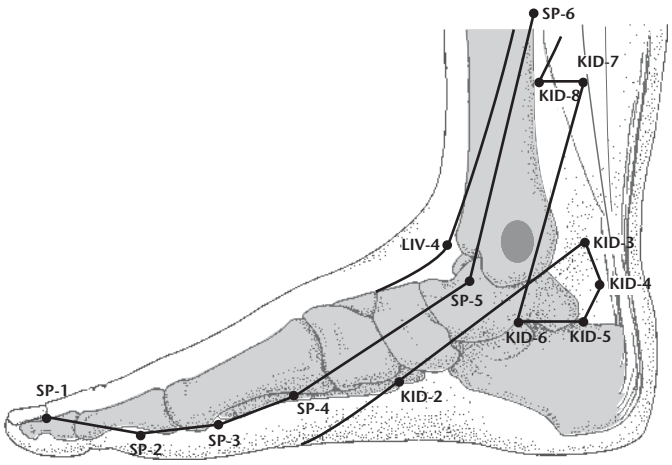

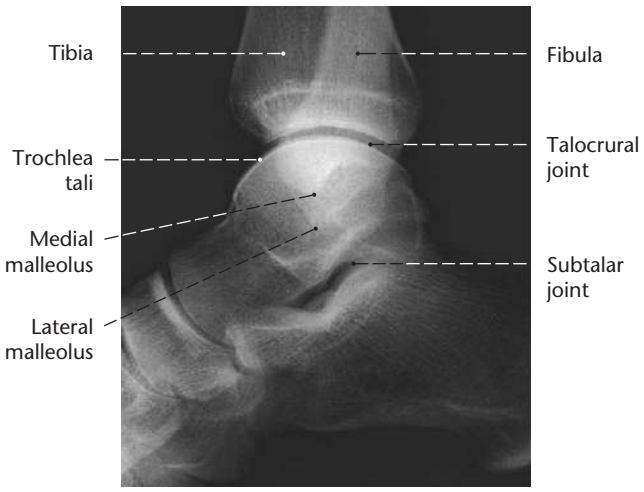

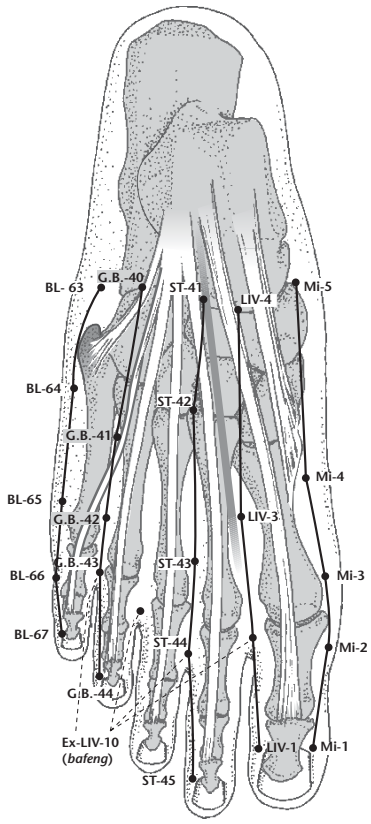

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 453

Woodworm Canal LIGOU LIV-5

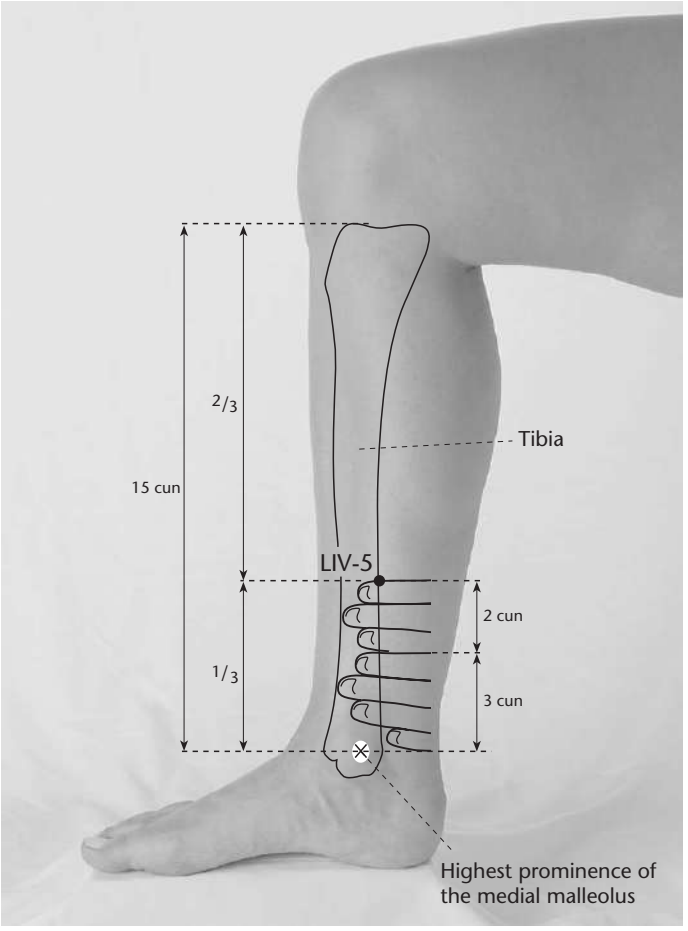

Location

On the medial aspect of the lower leg, 5 cun proximal to the highest prominence of the medial malleolus, just posterior to the medial crest of the tibia, between the crest of the tibia and the gastrocnemius muscle.

How to find

First, locate the highest prominence of the medial malleolus (→ 3.6.2). From there, palpate 5 cun in a proximal direction and there locate **LIV-5** in a depression directly posterior to the medial crest of the tibia. **Or quick method:** Divide the distance between the medial end of the popliteal crease and the highest prominence of the medial malleolus (=approximately 15 cun, → 2.2) into thirds and locate **LIV-5** a third of the distance from the medial malleolus. Note: Some texts locate **LIV-5 on** the tibia, others on the **posterior** border of the tibia; sensitivity to pressure should be the determining factor. → **KID-9** is located on the same level but directly superior to → **KID-3** (between the medial malleolus and the Achilles tendon).

Needling

0.5–1 cun vertically or obliquely in a posterior direction towards the fibula or transversely (subcutaneously) along the tibia towards the abdomen.

Actions/Indications

- Regulates Liver Qi, benefits the genitals, clears Damp-Heat from the Lower Burner, regulates menstruation
- Eliminates 'plum-pit' Qi
- Moves Qi locally

Special features

Luo-connecting point. Important distal point for the urogenital region (especially with Liver Qi stagnation and Damp-Heat).

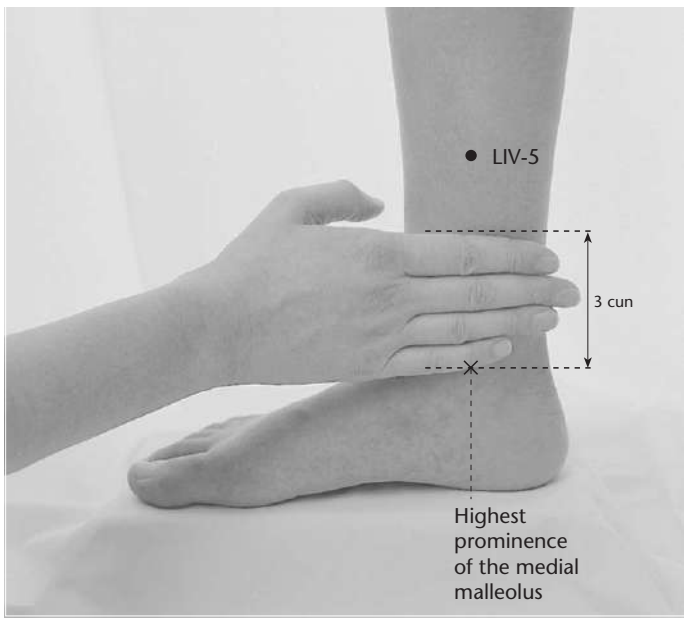

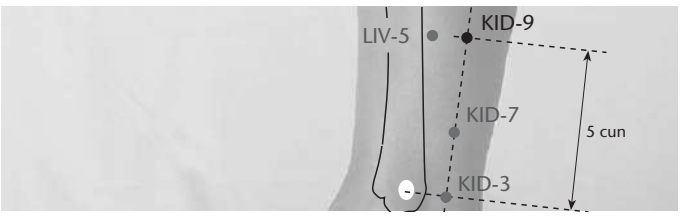

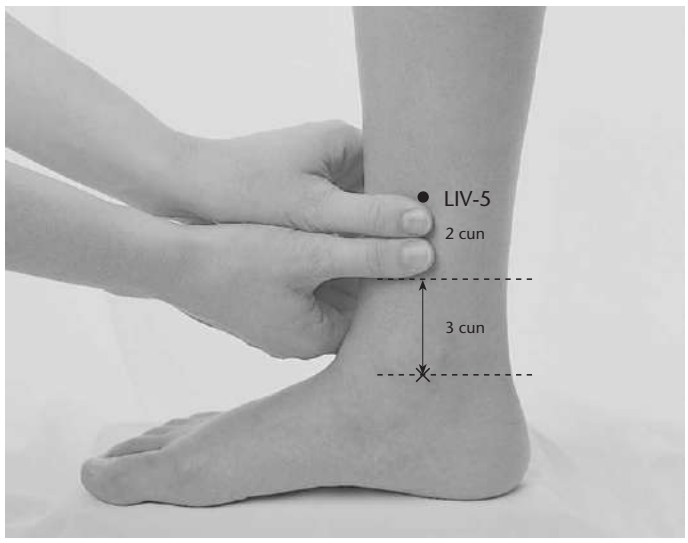

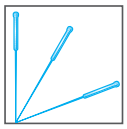
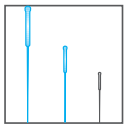

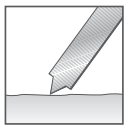

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 454

LIV-6 *Central Capital* ZHONGDU

Location

7 cun proximal to the highest prominence of the medial malleolus, just posterior to the medial crest of the tibia.

How to find

First, locate the highest prominence of the medial malleolus (→ 3.6.2). From there, palpate 7 cun (= the width of 2 hands and 1 thumb) in a proximal direction and there locate **LIV-6** in a depression directly posterior to the medial crest of the tibia. **Or quick method:** Spreading hands technique (→ 2.3.3): Place the little fingers on the medial end of the popliteal crease and the highest prominence of the medial malleolus respectively and join the thumbs on the posterior border of the tibia at the midpoint of this distance (= approximately 15 cun, 2.2). From there, measure 0.5 cun in a distal direction and here locate **LIV-6** posterior to the medial crest of the tibia. Note: Some texts locate **LIV-6 on** the tibia, others on the **posterior** border of the tibia; sensitivity to pressure should be the determining factor.

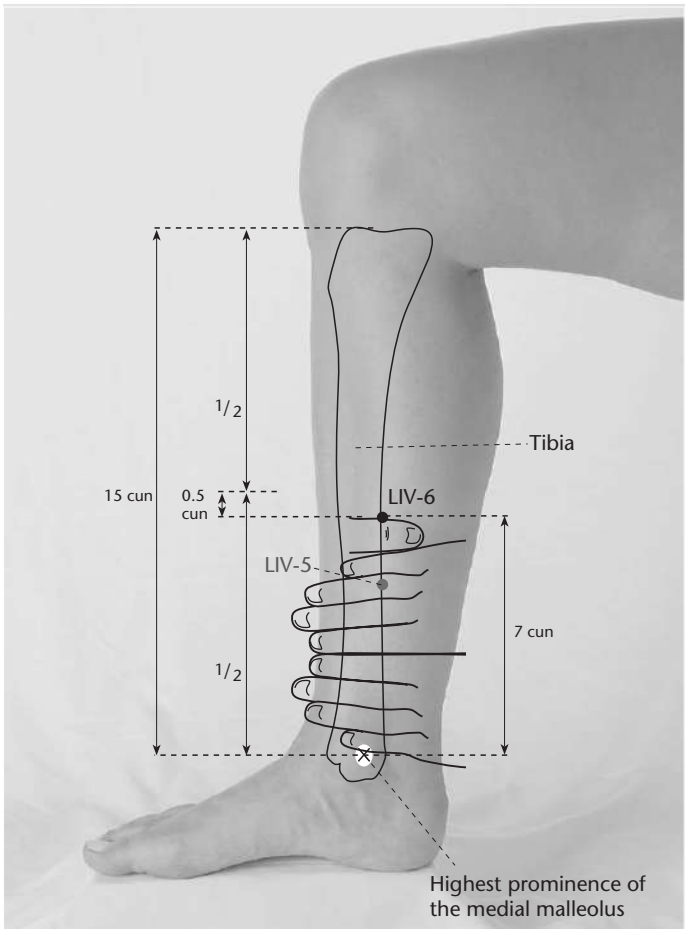

Needling

0.5–1 cun vertically or obliquely in a posterior direction towards the fibula or transversely (subcutaneously) in a proximal direction along the tibia.

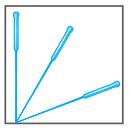

Actions/Indications

- Spreads Liver Qi, regulates the Lower Burner, regulates Blood, eliminates Dampness

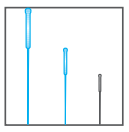

Special features

Xi-cleft point

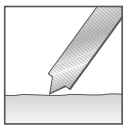

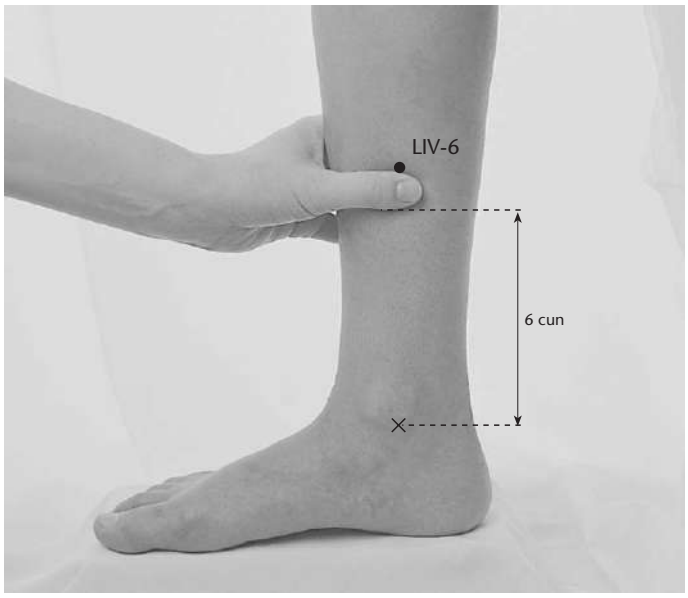

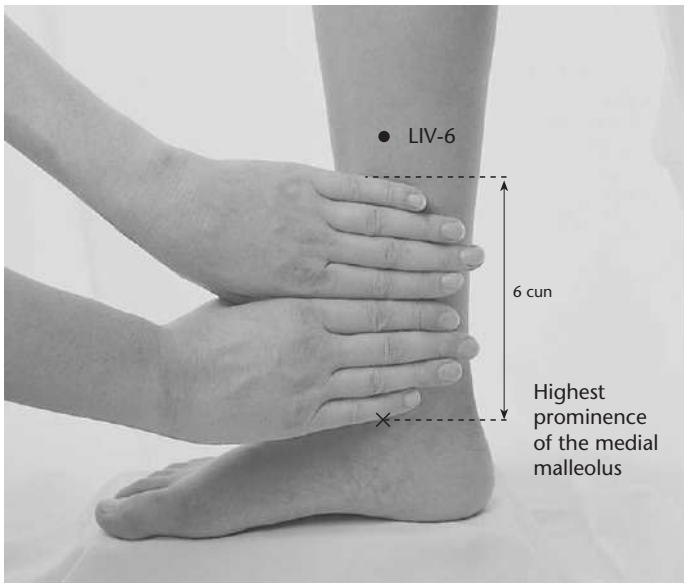

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 455

Knee Joint XIGUAN LIV-7

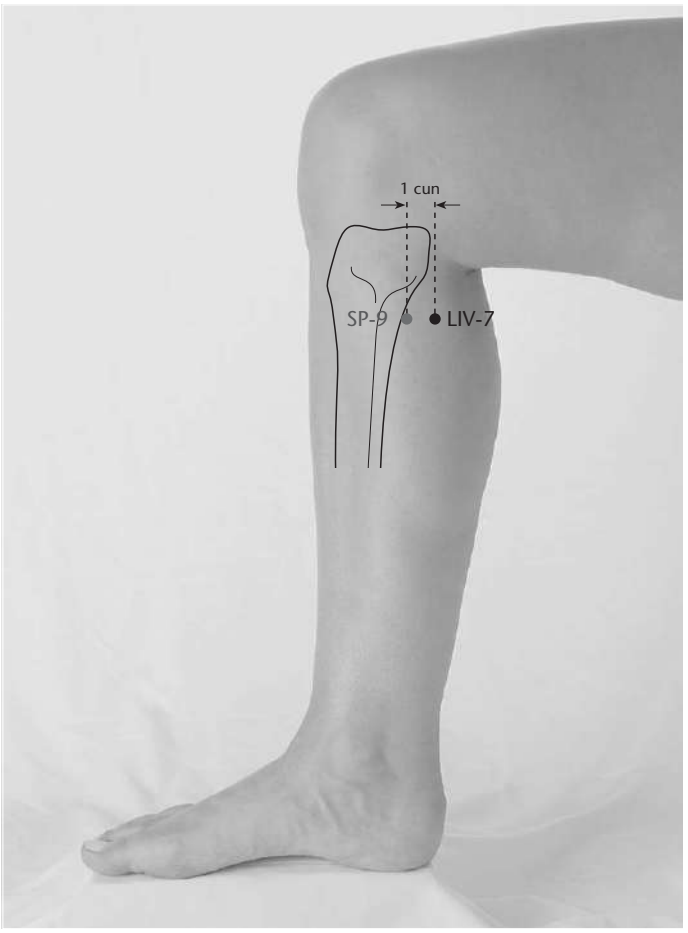

Location

At the junction of the shaft and the medial condyle of the tibia, 1 cun posterior to ➞ **SP-9**.

How to find

The junction of the shaft and the medial condyle of the tibia can be easily palpated. There, locate ➞ **SP-9** and **LIV-7** 1 cun posterior to it. Both points tend to be sensitive to pressure if indicated.

Needling

Vertically 1–2 cun

Actions/Indications

- Opens the channel
- Expels Wind and Dampness

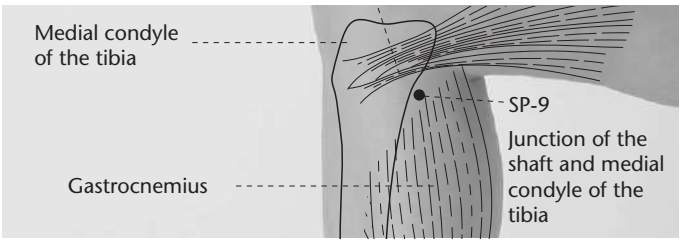

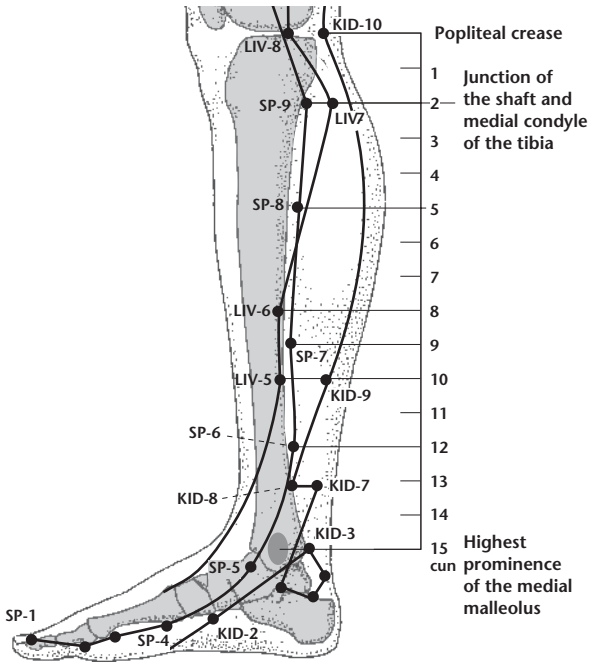

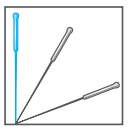

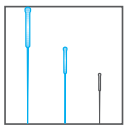

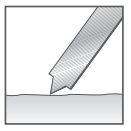

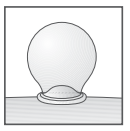

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 456

LIV-8 *Spring at the Crook* QUQUAN

Location

With the knee flexed, **LIV-8** is located directly proximal to the medial end of the popliteal crease, in a depression anterior to the tendons of the semitendinosus and semimembranosus muscles.

How to find

LIV-8 is best located with the patient's knee flexed and slight external rotation of the hip. First, palpate the prominent, ropey tendon of the semitendinosus muscle, which becomes more pronounced with the knee flexed. The tendon of the semimembranosus muscle lies deep to the semitendinosus and is often more difficult to isolate. With the palpating finger, glide from the tendon of the semitendinosus towards the patella. Approximately after 1 cun, you can feel a depression between the muscle bellies. **LIV-8** is located anterior to both tendons. Some authors locate **LIV-8** 1 cun superior and anterior to → **KID-10**. Sensitivity to pressure should determine the location.

→ **KID-10** is located approximately 1 cun in the direction of the popliteal crease and, from a medial perspective, posterior to the tendon of the semimembranosus and anterior to the tendon of the semitendinosus.

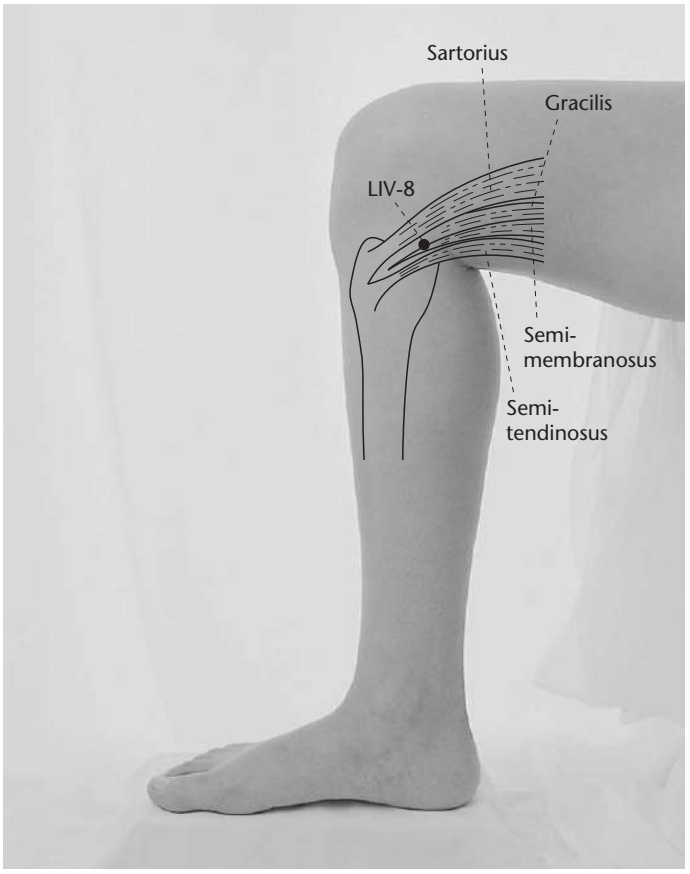

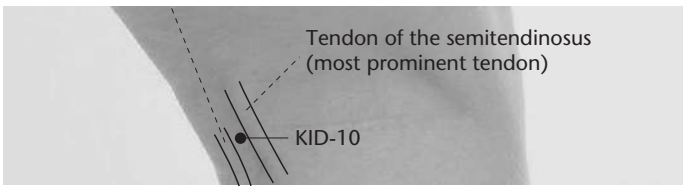

Needling

Vertically 0.5–1.5 cun

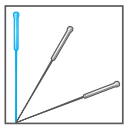

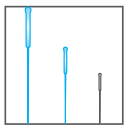

Actions/Indications

- Clears and eliminates Dampness and Damp-Heat from the Lower Burner (major action), benefits the genitals and the Uterus
- Nourishes Blood and Yin
- Moves Qi locally

Special features

he-sea point, Water point, tonification point.

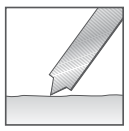

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 457

Yin Wrapping YINBAO LIV-9

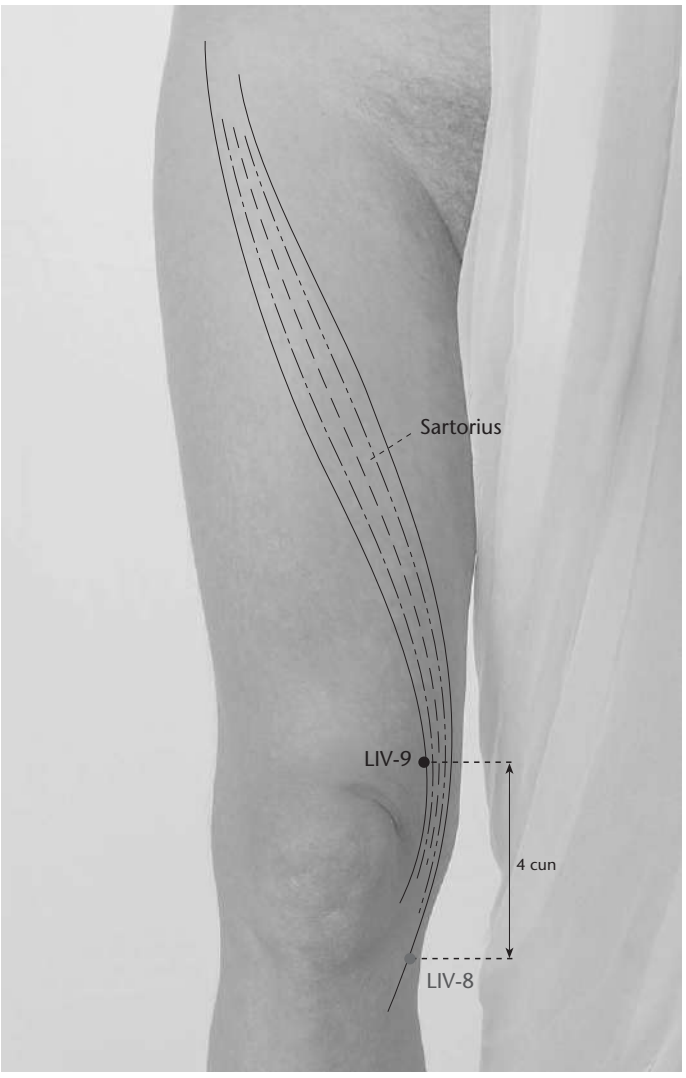

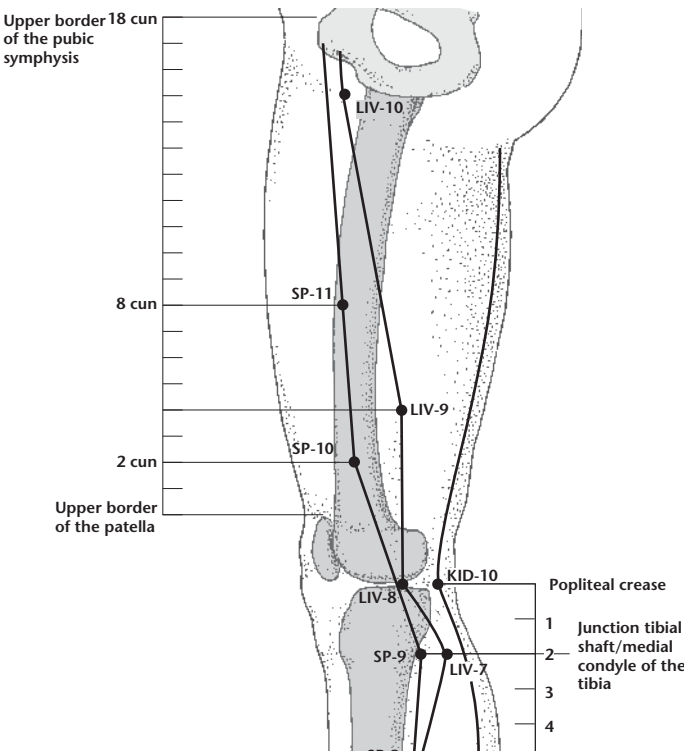

Location

4 cun proximal to the medial condyle of the femur, between the sartorius and vastus medialis muscles. Alternative locations: 4 cun proximal to → **LIV-8** (for example, Deadman et al 1998, → photo), 4 cun proximal to the junction of the shaft and the condyle of the femur (for example, Hecker et al → see line drawing) or 4 cun proximal to the midpoint of the condyle of the femur (for example, Ogal and Stör 1999, → illustration on p. 443).

How to find

From → **LIV-8**, measure 4 cun in a proximal direction. There, locate **LIV-9** on the border between the more medial, narrow sartorius muscle and the anteromedial vastus medialis muscle. Sensitivity to pressure should determine the location of this point.

Needling

Vertically 1–2 cun

Actions/Indications

- Benefits the Lower Burner
- Opens the channel

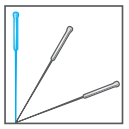
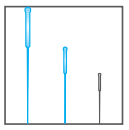

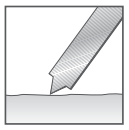

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 458

LIV-10 *Leg Five Miles* ZUWULI

Location

On the anterior aspect of the thigh, 3 cun inferior to the upper border of the symphysis, on the lateral border of the adductor longus muscle.

How to find

Ask the patient to flex their adductors by pressing their feet against each other. **LIV-10** is located in a groove on the medial border of the adductor longus muscle, where it forms an angle with the sartorius muscle.

Needling

Vertically 0.5–1.5 cun. Caution: Great saphenous vein, femoral artery/nerve.

Actions/Indications

- Relaxes the tendons and muscles
- Clears Damp-Heat from the Lower Burner
- Regulates the water passages

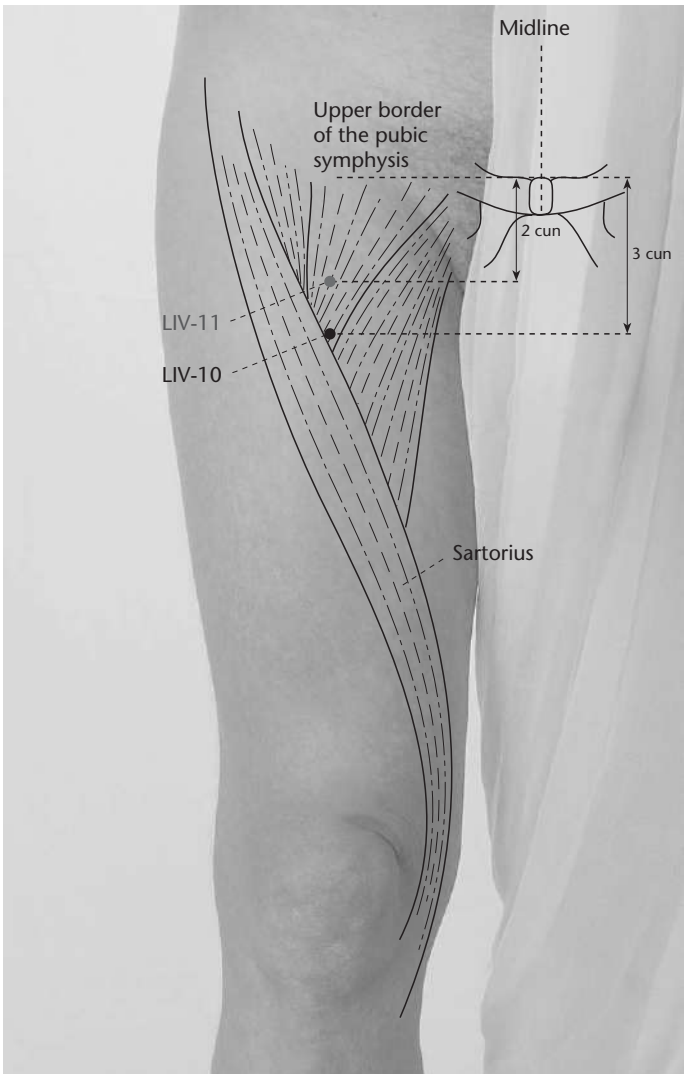

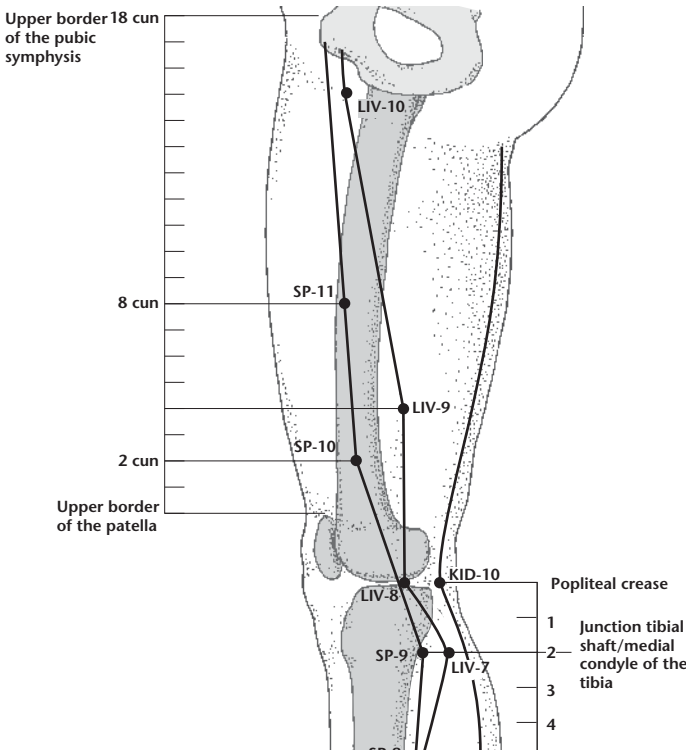

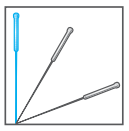
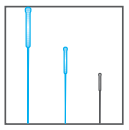

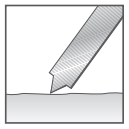

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 459

Yin Corner YINLIAN LIV-11

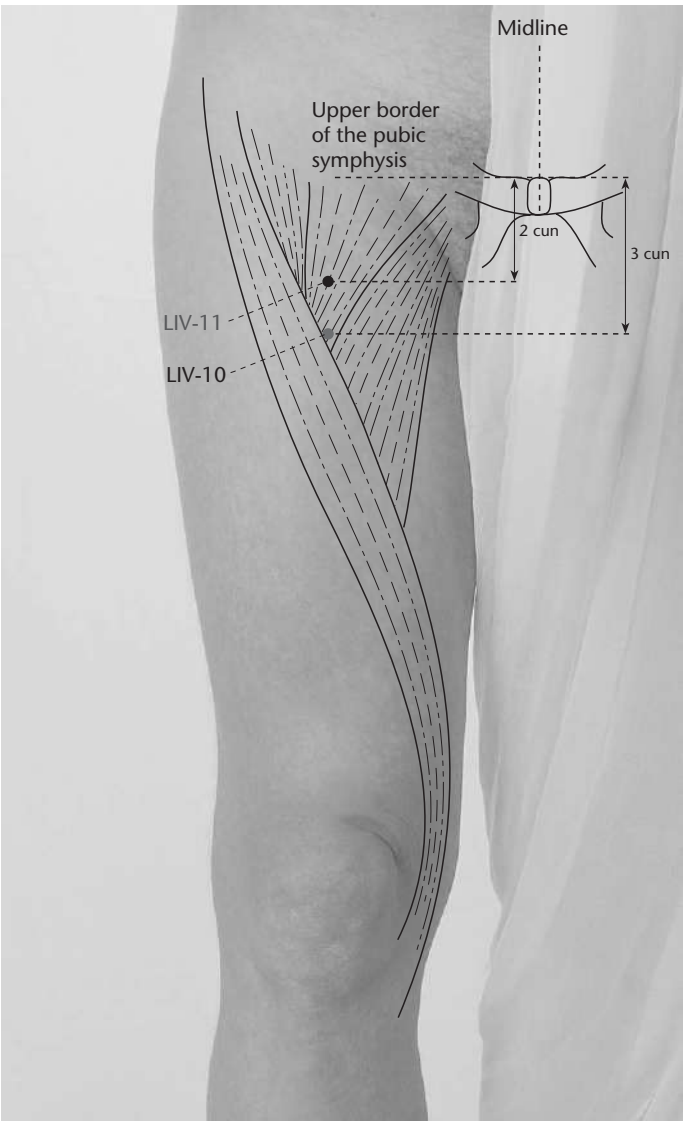

Location

On the anterior aspect of the thigh, 2 cun inferior to the upper border of the symphysis and on the lateral border of the adductor longus muscle.

How to find

Ask the patient to flex their adductors by pressing their feet against each other. **LIV-11** is located 2 cun inferior to the upper border of the pubic symphysis, in a groove on the medial border of the adductor longus muscle.

LIV-11 is located approximately 1 cun inferior to where the femoral artery passes under the inguinal ligament.

Needling

Vertically 0.5–1.5 cun. Caution: Great saphenous vein, femoral artery/vein/nerve. Classical texts recommend moxibustion for female sterility.

Actions/Indications

- Relaxes the tendons
- Benefits the Uterus and regulates menstruation
- For disorders of the lumbar spine

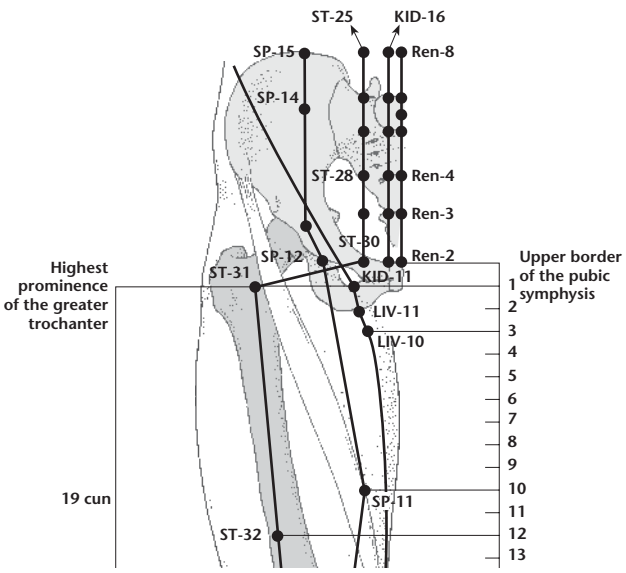

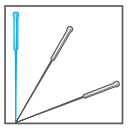
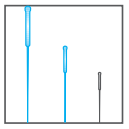

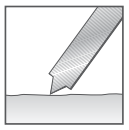

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 460

LIV-12 *Urgent Pulse* JIMAI

Location

In the inguinal groove, 2.5 cun lateral to the anterior midline and 1 cun inferior to the upper border of the symphysis.

How to find

From the upper border of the pubic symphysis, measure 2.5 cun in a lateral and 1 cun in an inferior direction. In this area, you can palpate the pulse of the femoral artery. **LIV-12** is located medial to the pulsating of the artery. As the femoral vein also runs medial to the artery, **LIV-12** should be needled in a medial direction, at least 1 fingerbreadth from the artery.

→ **ST-30** is located 1 cun superior and 0.5 cun medial to **LIV-12**.

Needling

Slightly obliquely 0.5–0.8 cun. Caution: Femoral artery/vein. The femoral vein is located medial to the artery and is approximately the breadth of a finger. To avoid injury to the vein, **LIV-12** should not be needled medial to the artery. Owing to its tricky location, classical texts recommend moxibustion only, while in modern texts moxibustion is contraindicated for the same reason.

Actions/Indications

- Dispels Cold from the Liver channel and benefits the Lower Burner

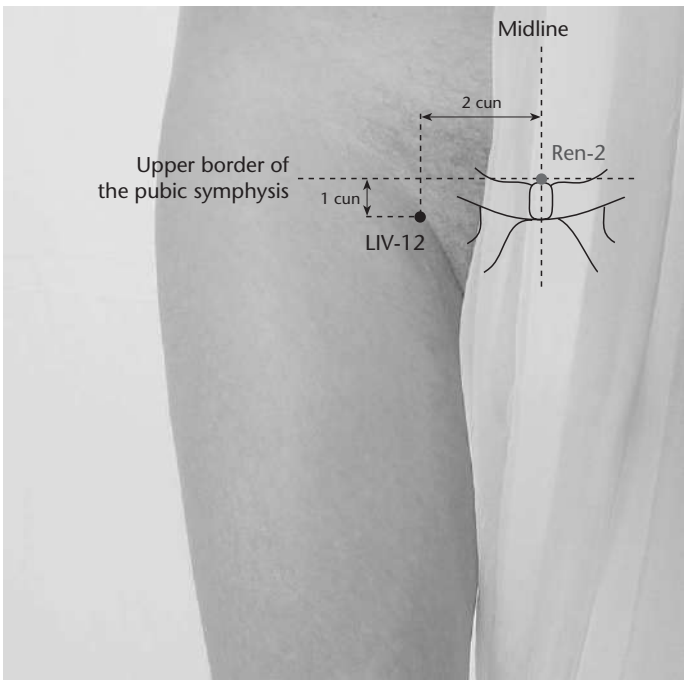

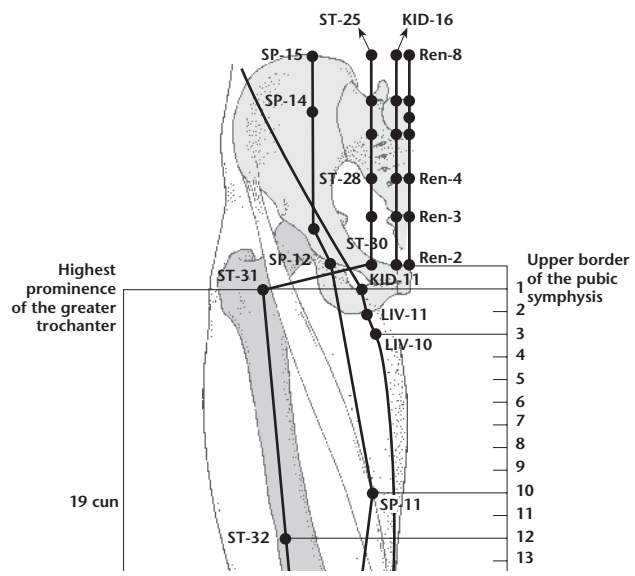

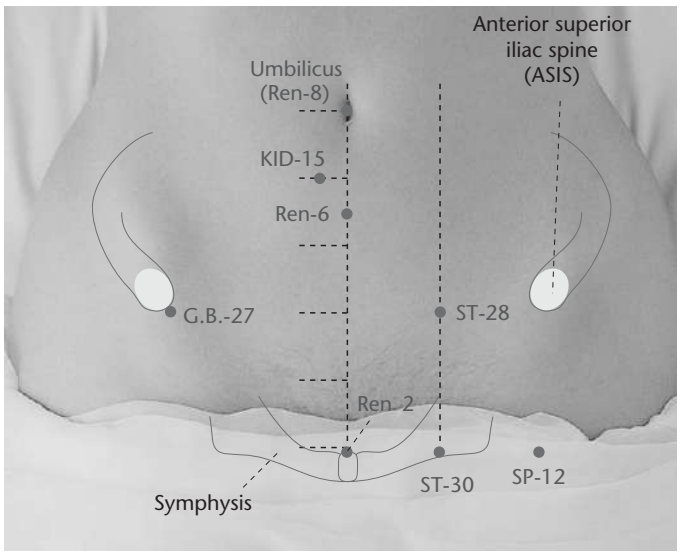

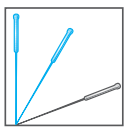
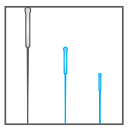

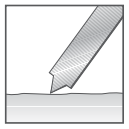

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 461

Completion Gate ZHANGMEN LIV-13

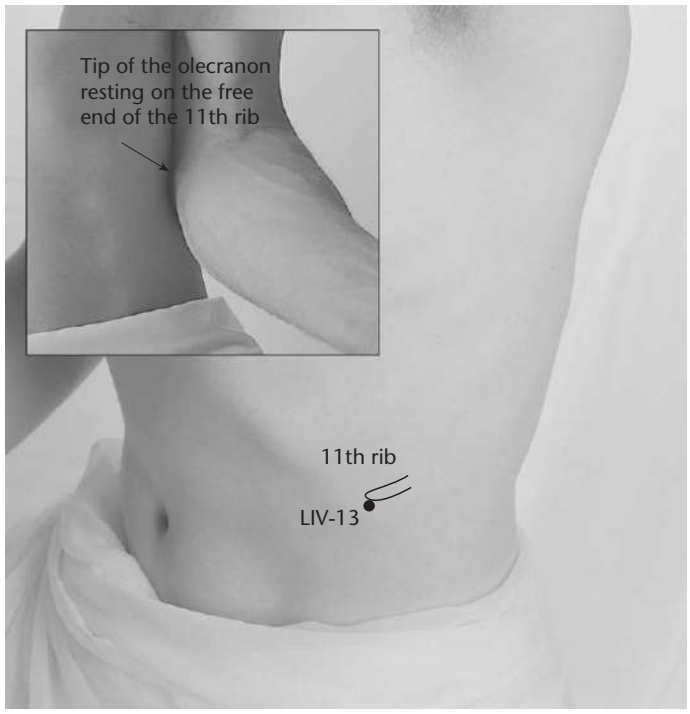

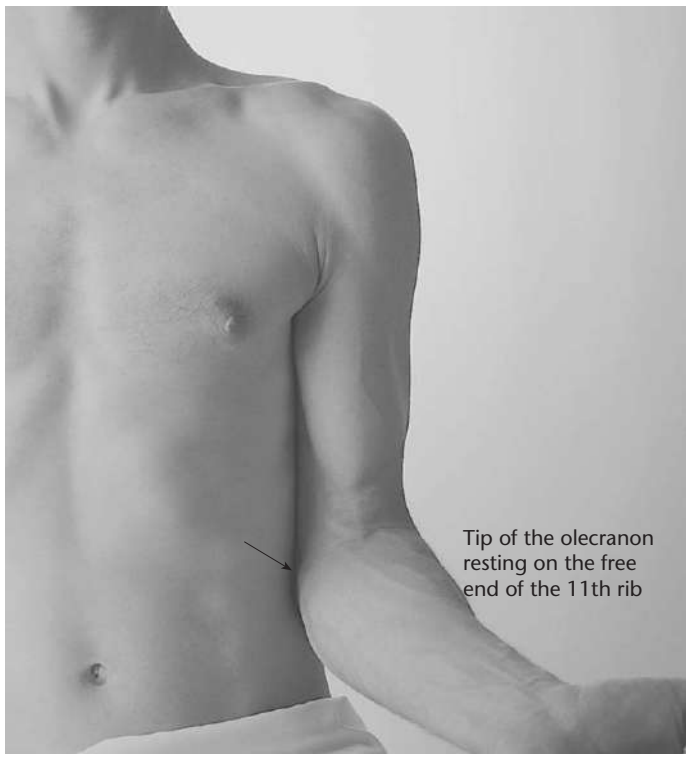

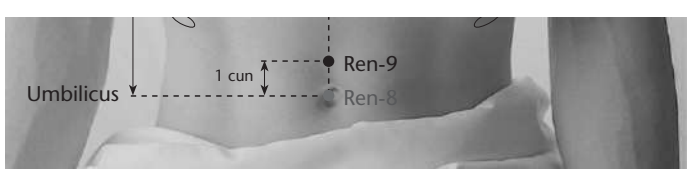

Location

Anterior and inferior to the free end of the 11th rib, approximately 1 cun superior to the umbilicus.

How to find

Quick method: Ask the patient to press their flexed elbow (90°) against the lateral aspect of the thorax and the tip of the olecranon will be pointing to the free end of the 11th rib. **LIV-13** can then be located on its anterior and inferior border. **Or:** Place the hand on the upper abdomen and, with gentle pressure, palpate along the lower border of the rib cage until you can feel the free end of the 11th rib just superior to the umbilicus. Locate **LIV-13** on its anterior and inferior border.

Located on approximately the same level (1 cun superior to the umbilicus) are → **Ren-9** (on the midline) and → **ST-24** (2 cun lateral to the midline). → **G.B.-25** is located more laterally and inferiorly, at the free end of the 12th rib.

Needling

0.5–1 cun vertically or better obliquely in a lateral direction. Caution: Peritoneum, hypertrophied organs (right: liver, left: spleen).

Actions/Indications

- Harmonises the Liver and the Spleen, regulates Liver Qi (especially in the Middle and Lower Burner), strengthens the Spleen
- Opens the channel

Special features

Front-*mu* point of the Spleen, meeting point with the *dai mai* and the G.B. channel. *Hui*-meeting point of the *zang*-Organs.

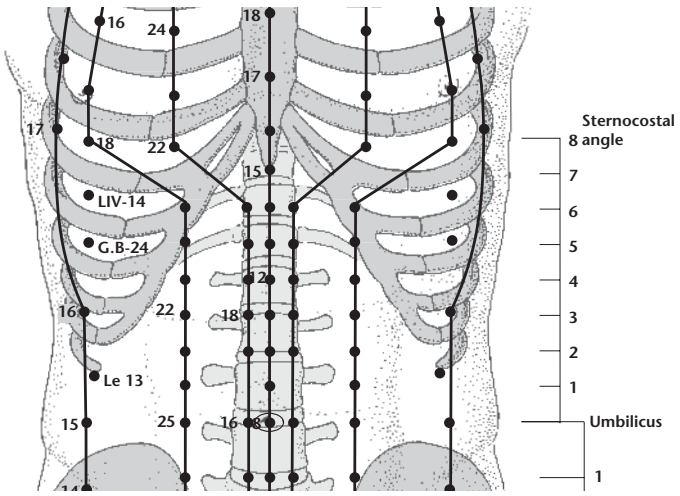

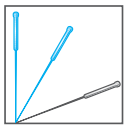
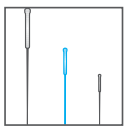

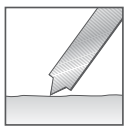
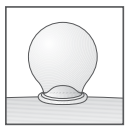

Ch04.12-F10028.qxd 2/23/08 11:32 AM Page 462

LIV-14 *Cycle Gate* QIMEN

Location

In the 6th intercostal space, on the mamillary line or 4 cun lateral to the midline.

How to find

Quick method, especially in men: The nipple is generally located on the level of the 4th intercostal space. From there, palpate downward 2 intercostal spaces and locate **LIV-14** in the 6th intercostal space. **Or:** For more precise orientation on the ribcage (→ 3.5), palpate for the manubriosternal synchondrosis, a well-defined horizontal bony structure on the sternum. Lateral to it, locate the 2nd rib and below it the 2nd intercostal space. From there, count downward 4 intercostal spaces to the 6th intercostal space and there locate **LIV-14** on the mamillary line (4 cun lateral to the midline).

→ **G.B.-24** is located in the 7th intercostal space. Located on the same level (2 cun inferior to the sternocostal angle → 3.5) are → **Ren-14** (on the midline), → **KID-21** (0.5 cun lateral to the midline) and → **ST-19** (2 cun lateral to the midline).

Needling

0.5–1 cun obliquely, either in a lateral or medial (along the course of the ICS) direction. The latter is recommended for disorders of the breasts. Caution: Pneumothorax.

Actions/Indications

- Regulates Liver Qi and Liver Blood (especially in the Upper and Middle Burner), cools the Blood, disperses accumulations, harmonises the Liver and the Stomach

Special features

Front-*mu* point of the Liver, meeting point with the *yin wei mai* and the Spleen channel, exit point.

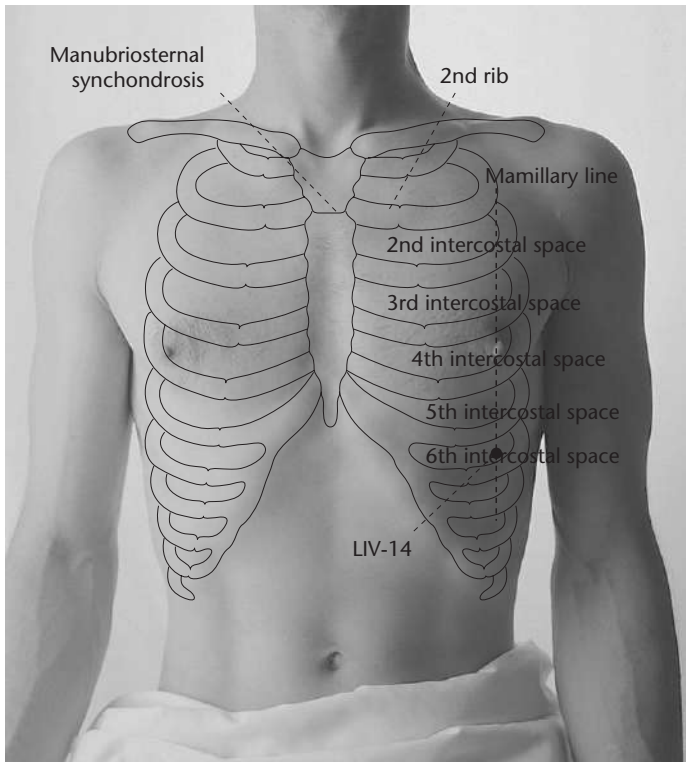

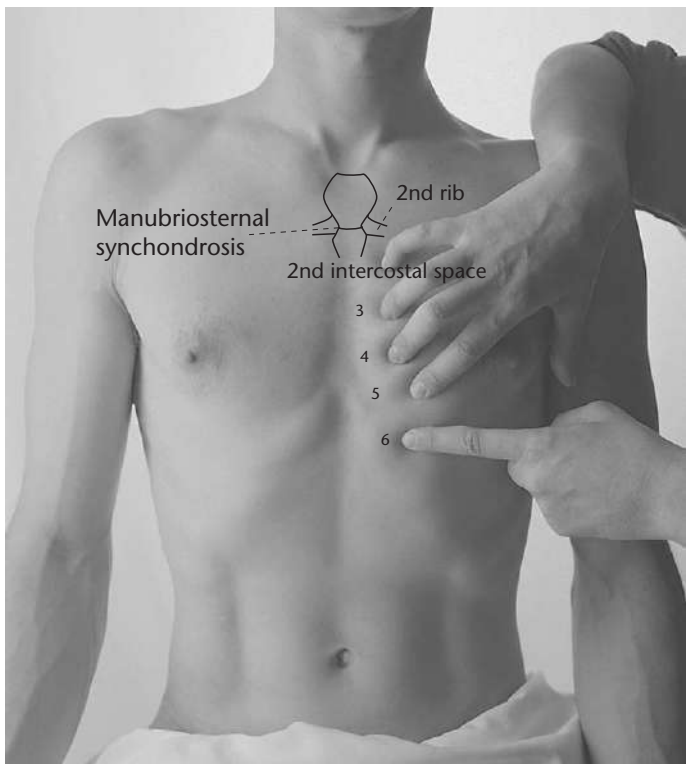

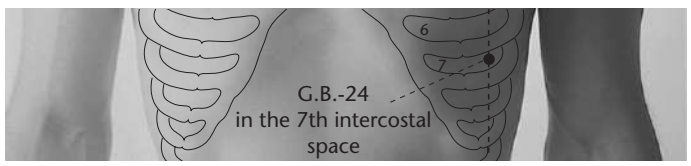

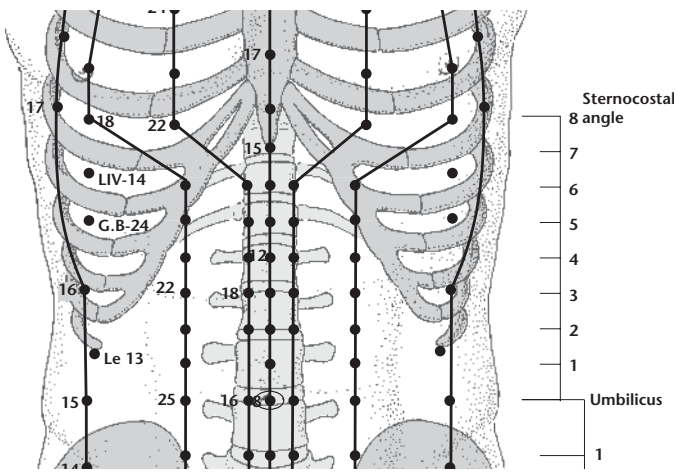

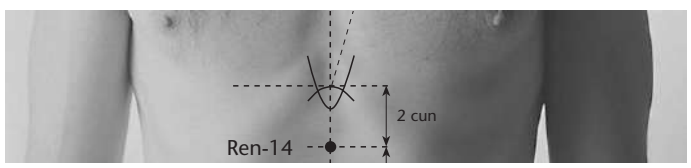

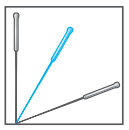
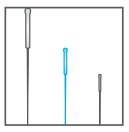
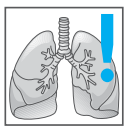

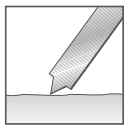

5 The Eight Extraordinary Vessels (*qi jing ba mai*) with Points

Claudia Focks, Ulrich März

For an overview of the theoretical background of the eight extraordinary vessels ➞ 1.7.

5.1 *Chong mai*

Synonym: Penetrating Vessel

Relationships (➞ 1.7.3)

- **Yin/Yang:** *chong mai/dai mai*
- **Central/peripheral:** *chong mai/yin wei mai*. Regions supported by this pairing: Heart, thorax, Stomach. Opening point: SP-4 (*gongsun*), coupled point: P-6 (*neiguan*)

Pathway

The extraordinary vessel *chong mai* originates like the *du mai* and *ren mai* in the lower abdomen at *bao zhong* (in women in the Uterus) or, according to some authors, at the Kidneys. It crosses the genital area and divides into two branches:

➡ A **deep branch** reaches the perineum at ➞ **Ren-1** (*huiyin*), travels to the spine, penetrates the spinal canal and ascends to the level of T1 and T2. (According to some authors the vessel ascends the anterior aspect of the spine.)

➡ The **other branch** emerges at ➞ **ST-30** (*qichong*) to the surface of the body and divides into **two external branches**, one descending and one ascending branch:

- the **ascending** branch reaches ➞ **KID-11** (*henggu*) 0.5 cun lateral to the midline at the level of the pubic symphysis and from there follows the KID channel to ➞ **KID-21** (*youmen*). The KID channel initially ascends the abdomen at a distance of 0.5 cun from the midline, passing ➞ **KID-12** (*dahe*), ➞ **KID-13** (*qixue*), ➞ **KID-14** (*siman*) and ➞ **KID-15** (*zhongzhu*), runs to the midline, meeting ➞ **Ren-7** (*yinjiao*) 1 cun inferior to the umbilicus, and again returns to its pathway 0.5 cun lateral to the midline, passing ➞ **KID-16** (*huangshu*), ➞ **KID-17** (*shangqu*), ➞ **KID-18** (*shiguan*), ➞ **KID-19** (*yindu*), ➞ **KID-20** (*futonggu*) and ➞ **KID-21** (*youmen*). The vessel then disperses into the intercostal spaces, ascends the neck, circles the inner surface of the lips and spreads in the upper pharynx and the nasal cavities.

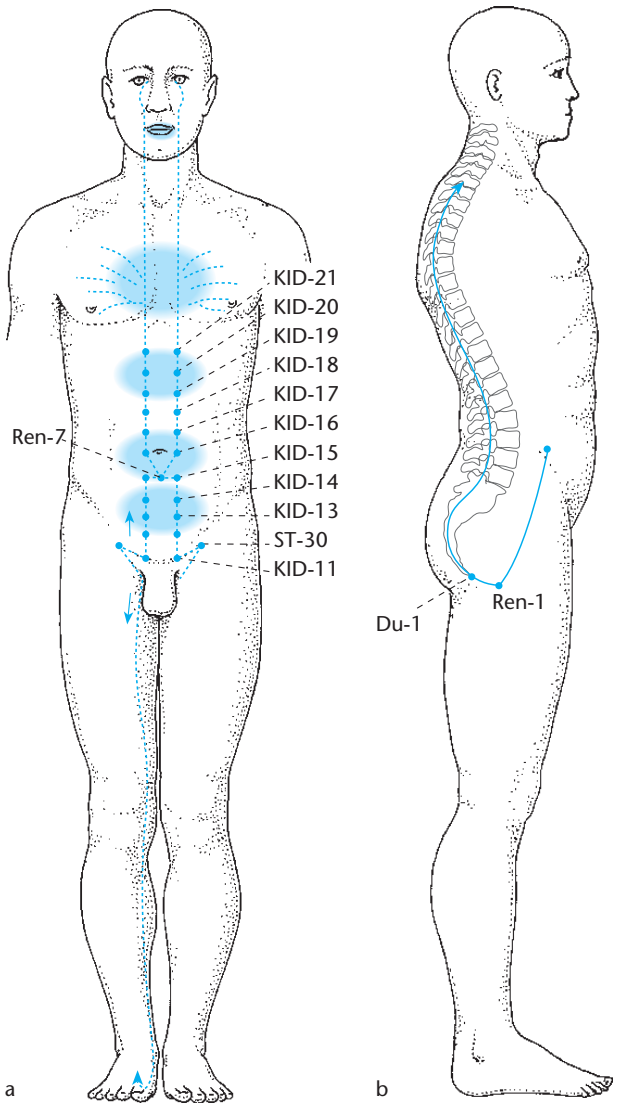

- the **descending branch** meets the Kidney primary channel and descends the medial aspect of the thigh (according to the *Ling Shu* it divides into the three foot Yin channels), penetrates the popliteal fossa, follows the medial border of the tibia and emerges posterior to the medial malleolus, where it divides into two sub-branches: one branch travels along the anteromedial aspect of the foot to the big toe, the other branch reaches the sole of the foot.

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 464

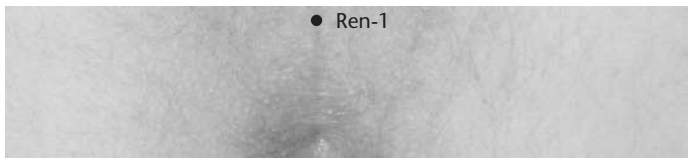

Ren-1 (*huiyin*): In the centre of the perineum.

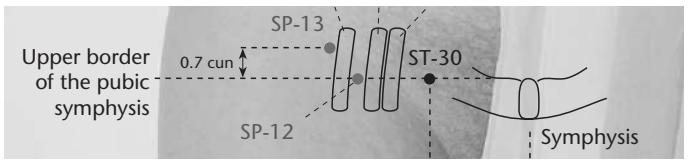

ST-30 (*qichong*): 2 cun lateral to the upper border of the pubic symphysis, medial to the femoral artery; at the level of Ren-2

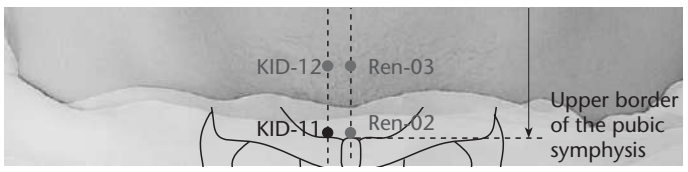

KID-11 (*henggu*): On the upper border of the pubic symphysis, 0.5 cun lateral to the anterior midline

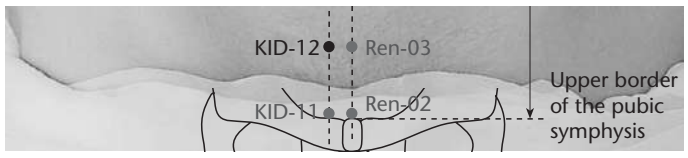

KID-12 (*dahe*): 1 cun superior to the upper border of the pubic symphysis, 0.5 cun lateral to the anterior midline

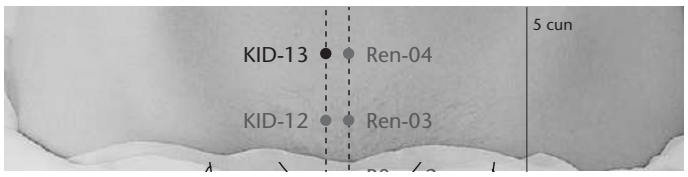

KID-13 (*qixue*): 2 cun superior to the upper border of the pubic symphysis, 0.5 cun lateral to the anterior midline

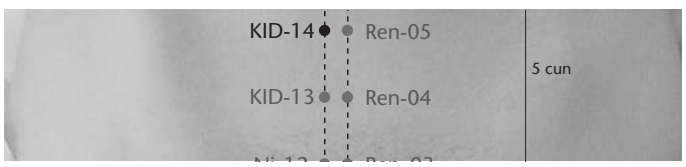

KID-14 (*siman*): 2 cun inferior to the umbilicus, 0.5 cun lateral to the anterior midline

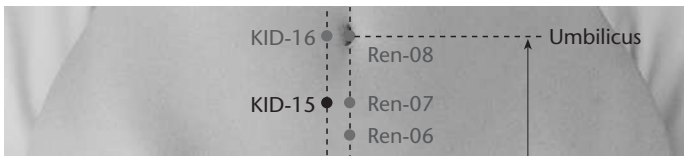

KID-15 (*zhongzhu*): 1 cun inferior to the umbilicus, 0.5 cun lateral to the anterior midline

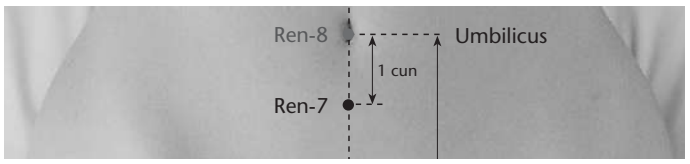

Ren-7 (*yinjiao*): On the anterior midline, 1 cun inferior to the umbilicus

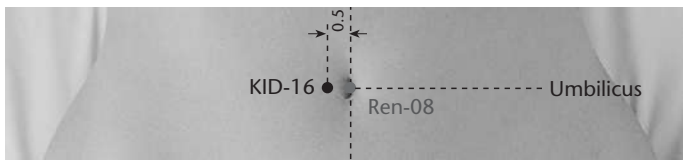

KID-16 (*huangshu*): 0.5 cun lateral to the centre of the umbilicus

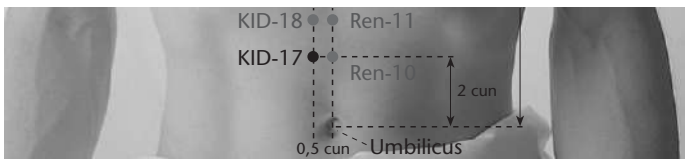

KID-17 (*shangqu*): 2 cun superior to the umbilicus, 0.5 cun lateral to the anterior midline

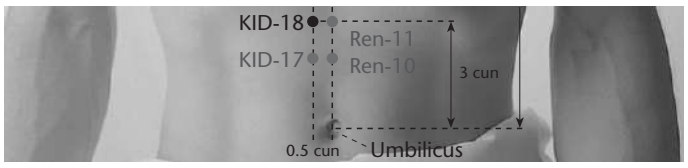

KID-18 (*shiguan*): 3 cun superior to the umbilicus, 0.5 cun lateral to the anterior midline

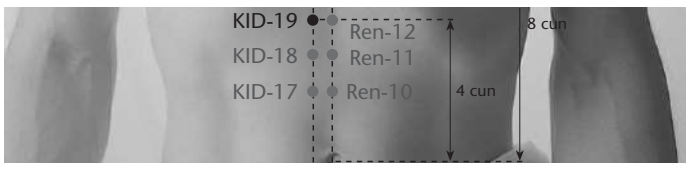

KID-19 (*yindu*): Midway between the sternocostal angle and the centre of the umbilicus, 0.5 cun lateral to the anterior midline

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 465

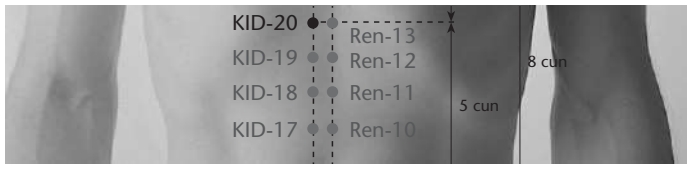

KID-20 (*futonggu*): 5 cun superior to the umbilicus or 3 cun inferior to the sternocostal angle, 0.5 cun lateral to the anterior midline

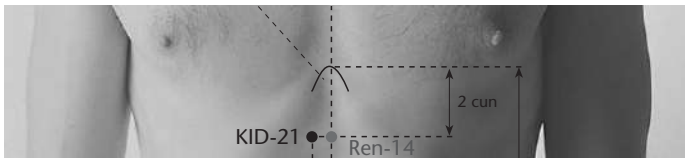

KID-21 (*youmen*): 2 cun inferior to the sternocostal angle, 0.5 cun lateral to the anterior midline

Connections with other channels and organs

- Kidney primary channel, *ren mai*, *du mai*, Stomach and Liver primary channels
- Kidney, Uterus

Clinical importance (→ 1.7.2, 1.7.3)

- **Sea of Blood**, strong connection to the various functions of the Blood
- Strong relationship to the Uterus, regulates the menses
- Regulates the 12 primary channels, effectively regulates the circulation of the Blood (*xue*) in the whole body
- Regulates counterflow Qi
- Connects the Stomach and Kidney primary channels and strengthens the relationship between the *ren mai* and *du mai*

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 467

5.2 Ren mai

Synonym: Conception Vessel

Relationships (→ 1.7.3)

- **Yin/Yang:** *ren mai/du mai*
- **Central/peripheral:** *ren mai/yin qiao mai*. Regions supported by this pairing: face, throat, thorax, Lungs, diaphragm, abdomen. Opening point: **LU-7** (*lieque*); coupled point: **KID-6** (*zhaohai*)

Pathway

The extraordinary vessel *ren mai* originates like the *chong mai* and *du mai* in the lower abdomen (*bao zhong*) or, according to some authors, at the Kidneys, connects with the urogenital area and emerges at the perineum at **Ren-1** (*huiyin*). From there the **external** pathway ascends the anterior midline, passing the abdomen, the chest, and the throat. It terminates on the chin in the mentolabial groove at **Ren-24** (*chengjiang*), where it meets the Stomach primary channel and the extraordinary vessel *du mai*. From here the vessel continues internally, encircles the inner surface of the lips and connects with the *du mai* at → **Du-28** (*yinjiao*) beneath the upper frenulum. It divides into **two branches**, which both reach the infraorbital ridge at → **ST-1** (*chengqi*), where they cross the Stomach primary channel as well as the extraordinary vessel *yin qiao mai*, and terminate near the eyes.

- **An internal branch** runs from the perineum to the spine, penetrates the spinal canal at **Du-1** (*changqiang*) and ascends in the spine.

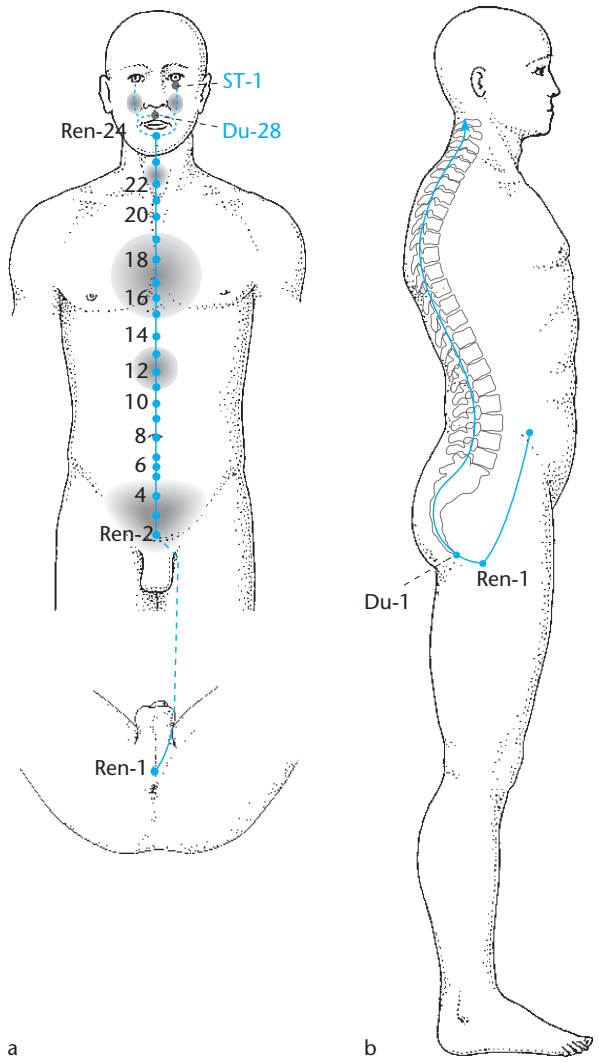

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 468

Meeting Points of other channels with the *ren mai*

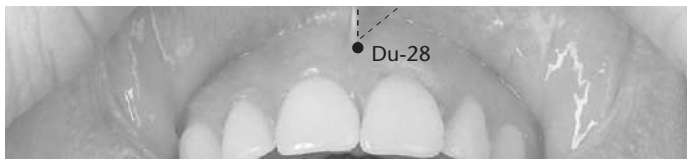

Du-28 (*yinjiao*): On the inside of the upper lip, at the junction of the frenulum and the upper gum

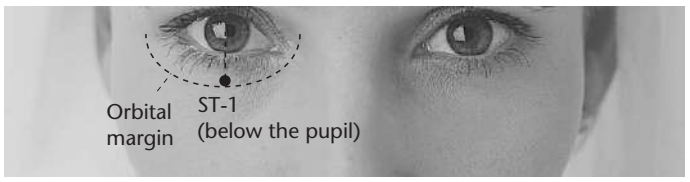

ST-1 (*chengqi*): With the eyes looking straight ahead, directly below the centre of the pupil, between the eyeball and the infraorbital ridge

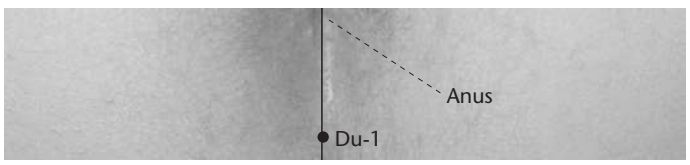

Du-1 (*changqiang*): On the midline, midway between the tip of the coccyx and the anus

Connections with other channels and organ systems

- *Du mai*, *chong mai*, Kidney and Stomach primary channels
- Uterus, Kidney

Clinical importance (→ 1.7.2, 1.7.3)

- **Sea of the Yin Qi**, regulates and circulates the Yin Qi (including the Blood, Essence (*jing*) and body fluids) in order to nourish and moisten the body
- Connects all Yin channels, especially those of the lower abdomen
- Controls the anterior thoracic region and the abdomen
- Close relationship to the functions of the abdominal organs, particularly the Kidney, Bladder, Uterus and Lung
- Regulates reproduction
- Stores the Essence Qi (*jing qi*) overflowing from the Yin channels
- Absorbs excess from the Yin channels, especially stagnant Qi and Blood

Notes about the *ren mai*

Both the *ren mai* and *du mai* have a unique position among the eight extraordinary vessels. They are the only two extraordinary vessels with their own specific points – all other extraordinary vessels can only be accessed through points on the primary channels. For this reason the *ren mai* and the *du mai* together with the primary channels are often categorised as 14 channels. However, from the point of view of channel energetics this is not correct, since the extraordinarily vessels occupy a special position, both evolutionary as well as functionally (→ 1.7).

The pathway of the *ren mai*, like that of the *du mai*, is limited to the torso and the head. While it does not course along the extremities and does not have any points on the limbs (neither general points nor specific points such as the five *shu*-transporting points, *yuan*-source point, etc.) it does have an opening point (or Master point) (→ 8.1.8). The *ren mai* is also referred to as the Sea of Yin and governs the Yin in the whole body. Along its pathway are located the anterior aspects of the energy centres (or chakras according to Yogi tradition). Daoist tradition describes three energy centres, the three cinnabar fields (*dantian*) on the anterior aspect of the body, which play an important role in *Qigong* practice as centres where the Qi pools and originates. While all three centres are located on the pathway of the *ren mai*, they do not correspond to specific points but rather to certain areas.

The *ren mai* runs along the Yin side of the body. It contains not only several Front-*mu* points acting directly on the *zangfu*-Organs, but also multiple meeting points with other channels. This clearly denotes the outstanding importance of the *ren mai*.

Specific points according to their function

- ***Yuan*-Source point (→ 8.1.1): –**
- ***Luo*-Connecting point (→ 8.1.2.): Ren-15** (*jiuwei*) ■■
- ***Xi*-Cleft point (→ 8.1.3): –**
- **Associated Back-*shu* point (→ 8.1.4): –**
- **Associated Front-*mu* point (→ 8.1.5): –**
- **Five *shu*-transporting points (→ 8.1.6): –**
- **Hui-meeting point (→ 8.1.7)**
  - of the Qi: **Ren-17** (*danzhong*) ■■
  - of the *fu*-Organs: **Ren-12** (*zhongwan*) ■■
- **Pertaining opening point (→ 8.1.8): LU-7** (*lieque*)
- **Pertaining coupled point (→ 8.1.8): KID-6** (*zhaohai*)
- **Lower *he*-Sea point (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10):**
  - with the *chong mai* and the *du mai*: **Ren-1** (*huiyin*)
  - with the LIV channel: **Ren-2** (*qugu*) ■■
  - with the SP, LIV, KID channels: **Ren-3** (*zhongji*) ■■, **Ren-4** (*guanyuan*) ■■
  - with the *chong mai* and the KID channel*: **Ren-7** (*yinjiao*)
  - with the SP and ST* channels: **Ren-10** (*xiawan*)
  - with the S.I., T.B.* and ST channels: **Ren-12** (*zhongwan*) ■■
  - with the ST and S.I. channels: **Ren-13** (*shangwan*)
  - with the SP*, KID, S.I.* and T.B. channels: **Ren-17** (*danzhong*) ■■
  - with the *yin wei mai*: **Ren-22** (*tiantu*) ■■, **Ren-23** (*lianquan*)

* Mentioned by only some authors

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 469

–468 with the *du mai*, the L.I. and ST channels: **Ren-24** (*chengjiang*) ■■
– of other channels with the *ren mai*: **Du-28, ST-1, Du-1**

- **Gao Wu command point (→ 8.1.11):** –
- **Window of Heaven point (→ 8.1.12): Ren-22** (*tiantu*) ■■
- **Points of the Four Seas (→ 8.1.13) Qi: Ren-17** (*danzhong*) ■■
- **Ma Dan Yang Heavenly Star point (→ 8.1.14):** –
- **Sun Si Miao Ghost points (→ 8.1.15): Ren-1** (*huiyin*), **Ren-24** (*chengjiang*) ■■
- **Other functional points:**
  – Front-*mu* point of the Bladder: **Ren-3** (*zhongji*) ■■
  – Front-*mu* point of the Small Intestine: **Ren-4** (*guanyuan*) ■■
  – Front-*mu* point of the Triple Burner: **Ren-5** (*shimen*)
  – important, general tonification point: **Ren-6** (*qihai*) ■■
  – Front-*mu* point of the Stomach: **Ren-12** (*zhongwan*) ■■
  – Front-*mu* point of the Heart: **Ren-14** (*juque*) ■■
  – important calming point: **Ren-15** (*jiuwei*) ■■
  – Front-*mu* point of the Pericardium, point of the Sea of Qi: **Ren-17** (*danzhong*) ■■

Points according to region

- **Local points (→ 8.2.1):** throat – **Ren-23** (*lianquan*); Lungs – **Ren-22** (*tiantu*) ■■, **Ren-17** (*danzhong*) ■■; Heart – **Ren-17** (*danzhong*) ■■; Spleen/Stomach – **Ren-12** (*zhongwan*) ■■; Small Intestine – **Ren-4** (*guanyuan*) ■■, **Ren-9** (*shuifen*); Bladder – **Ren-3** (*zhongji*) ■■; urogenital region – **Ren-3** (*zhongji*), **Ren-4** (*guanyuan*) ■■; epigastrium – **Ren-12** (*zhongwan*) ■■, **Ren-13** (*shangwan*); abdomen – **Ren-3** (*zhongji*) ■■, **Ren-4** (*guanyuan*) ■■, **Ren-6** (*qihai*) ■■
- **Adjacent points (→ 8.2.1):** throat – **Ren-22** (*tiantu*) ■■; tongue – **Ren-23** (*lianquan*); Heart – **Ren-14** (*juque*) ■■; Gall Bladder – **Ren-11** (*jianli*); Kidney – **Ren-4** (*guanyuan*) ■■; epigastrium – **Ren-8** (*shenque*)
- **Distal points (→ 8.2.1):** –

Specific points according to the vessel pathway (in numerical order)

- **Ren-1** (*huiyin*): *jiaohui*-meeting point with the *chong mai* and *du mai* (→ 8.1.10); Sun Si Miao Ghost point (→ 8.1.15)
- **Ren-2** (*qugu*) ■■: *jiaohui*-meeting point with the LIV channel (→ 8.1.10)
- **Ren-3** (*zhongji*) ■■: *jiaohui*-meeting point with the SP, LIV, KID channels (→ 8.1.10); Front-*mu* point of the Bladder; local point for the Bladder, urogenital region and lower abdomen (→ 8.2.1)
- **Ren-4** (*guanyuan*) ■■: Front-*mu* point of the Small Intestine (→ 8.1.5); *jiaohui*-meeting point with the SP, LIV, KID channels (→ 8.1.10); local point for the Small Intestine, urogenital region and lower abdomen (→ 8.2.1); adjacent point for the Kidneys
- **Ren-5** (*shimen*): Front-*mu* point of the Triple Burner
- **Ren-6** (*qihai*) ■■: important general tonification point; local point for the lower abdomen (→ 8.2.1); adjacent point for the Lung (→ 8.2.1)
- **Ren-7** (*yinjiao*): *jiaohui*-meeting point with the *chong mai* and the KID channel (→ 8.1.10)
- **Ren-8** (*shenque*): adjacent point for the epigastrium (→ 8.2.1)
- **Ren-9** (*shuifen*): local point for the Small Intestine (→ 8.2.1)
- **Ren-10** (*xiawan*): *jiaohui*-meeting point with the SP and ST* channels (→ 8.1.10)
- **Ren-11** (*jianli*): adjacent point for the Gall Bladder (→ 8.2.1)
- **Ren-12** (*zhongwan*) ■■: *hui*-meeting point (→ 8.1.7) of the *fu*-Organs; *jiaohui*-meeting point with the S.I., T.B.* and ST channels (→ 8.1.10); Front-*mu* point of the Stomach (→ 8.1.5); local point for the SP/ST and the epigastrium (→ 8.2.1)
- **Ren-13** (*shangwan*): *jiaohui*-meeting point with the ST and S.I. channels (→ 8.1.10); local point for the epigastrium (→ 8.2.1)
- **Ren-14** (*juque*) ■■: Front-*mu* point of the Heart (→ 8.1.5); adjacent point for the Heart
- **Ren-15** (*jiuwei*) ■■: *luo*-connecting point (→ 8.1.2); important point for calming
- **Ren-17** (*danzhong*) ■■: *hui*-meeting point (→ 8.1.7) of the Qi; *jiaohui*-meeting point with the SP, KID, S.I. and T.B. channels (→ 8.1.10); Front-*mu* point of the Pericardium, point of the Sea of Qi; local point for the Lung and Heart (→ 8.2.1)
- **Ren-22** (*tiantu*) ■■: *jiaohui*-meeting point with the *yin wei mai* (→ 8.1.10); Window of Heaven point (→ 8.1.12); local point for the Lungs (→ 8.2.1); adjacent point for the throat (→ 8.2.1)
- **Ren-23** (*lianquan*): *jiaohui*-meeting point with the *yin wei mai* (→ 8.1.10); local point for the throat (→ 8.2.1); adjacent point for the tongue (→ 8.2.1)
- **Ren-24** (*chengjiang*) ■■: *jiaohui*-meeting point with the *du mai*, the L.I. and ST channels (→ 8.1.10); Sun Si Miao Ghost point

* Mentioned by only some authors

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 470

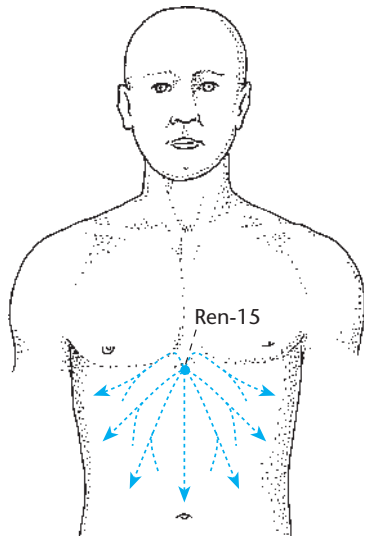

Luo-Connecting Vessel of the *Ren Mai* (→ 1.6)

Pathway

The *luo*-connecting vessel of the *ren mai* originates at **Ren-15** (*jiuwei*), below the xiphoid process (→ 3) and disperses over the skin of the abdomen.

Clinical importance

Pathology

Excess (*shi*): pain of the abdominal skin
Deficiency (*xu*): itching of the abdominal skin

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 471

Meeting of the Yin HUIYIN Ren-1

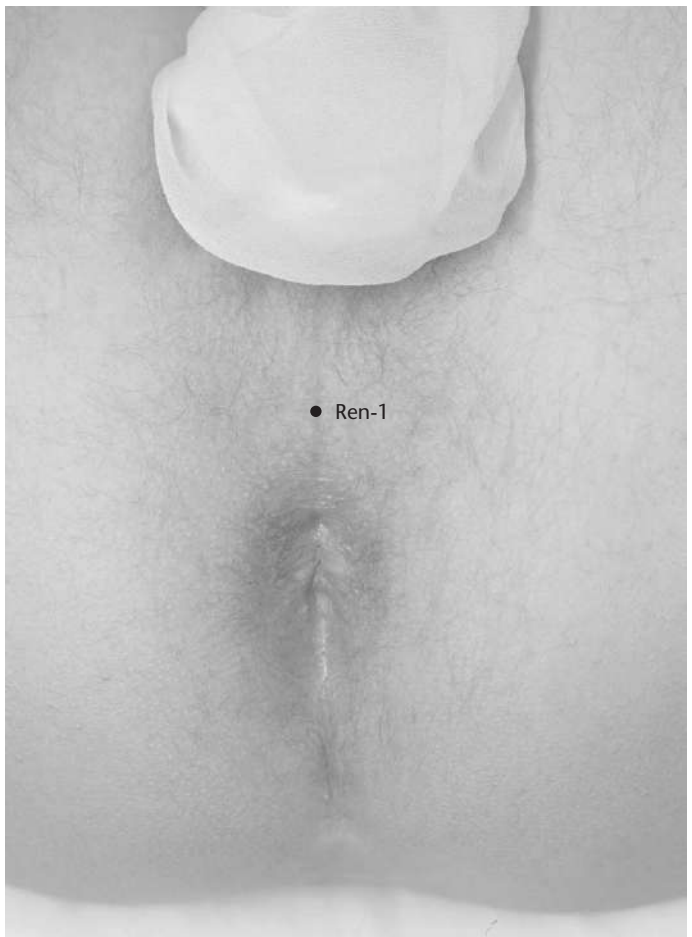

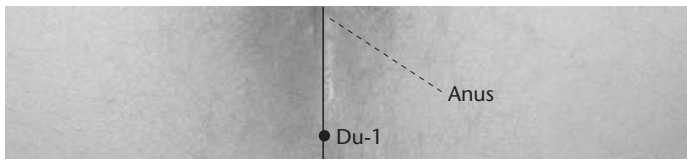

Location
In the centre of the perineum.

How to find
In women: midway between the anus and the posterior labial commissure. In men: midway between the anus and the scrotum.

Needling
Vertically 0.5–1 cun. Caution: Contraindicated during pregnancy. Some classic texts prohibit needling this point. Particular attention should be paid to the prevention of infection.

Actions/Indications
- Clears Heat, strengthens the Kidneys and drains Dampness
- Regulates menstruation

Special features
Meeting point with the *du mai* and the *chong mai*. As a Sun Si Miao Ghost point, **Ren-1** is traditionally indicated for treating epilepsy and psychoses.

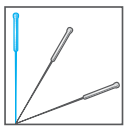
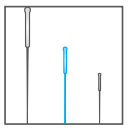

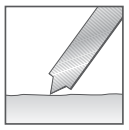
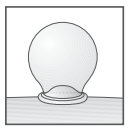

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 472

Ren-2 *Curved Bone* QUGU

Location
On the anterior midline, on the upper border of the pubic symphysis.

How to find
In the pubic area, locate the upper border of the pubic symphysis. **Ren-2** is located directly superior to its upper border, on the anterior midline.
Located on the same level are → **KID-11** (0.5 cun lateral to the midline), → **ST-30** (2 cun lateral to the midline) and → **SP-12** (3.5 cun lateral to the midline).

Needling
Vertically 0.5–1 cun. Caution: Peritoneum. Also caution with a full bladder and during pregnancy.

Actions/Indications
- Benefits urination, warms the Yang and strengthens the Kidneys
- Regulates the Lower Burner

Special features
Meeting point with the LIV channel

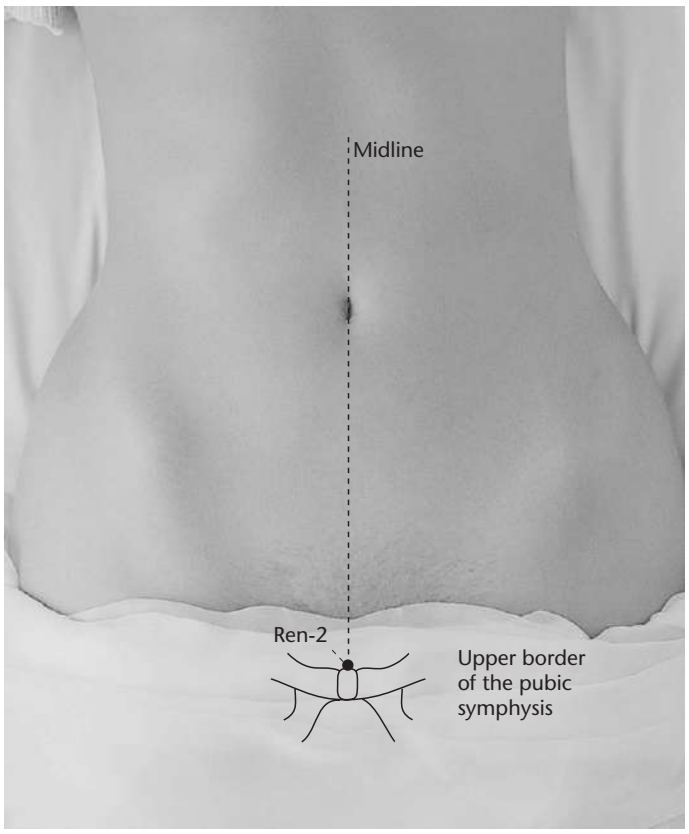

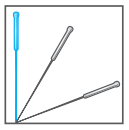
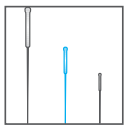
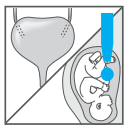

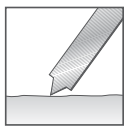

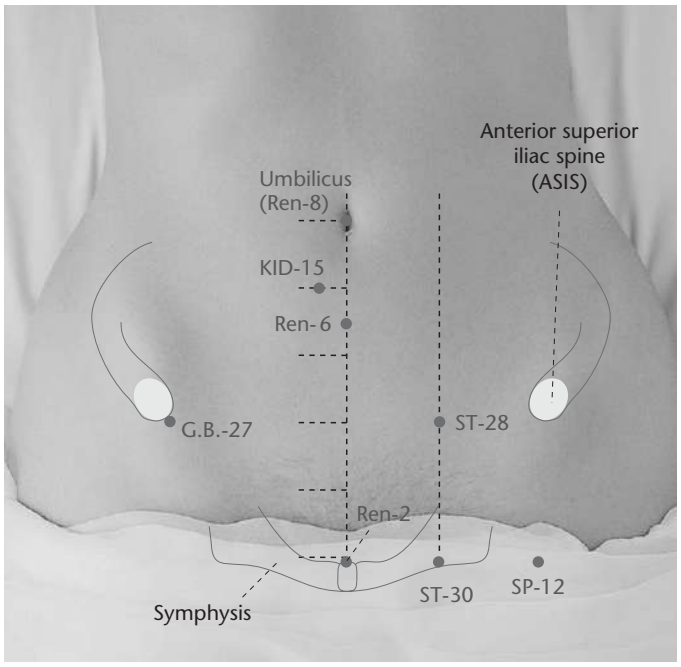

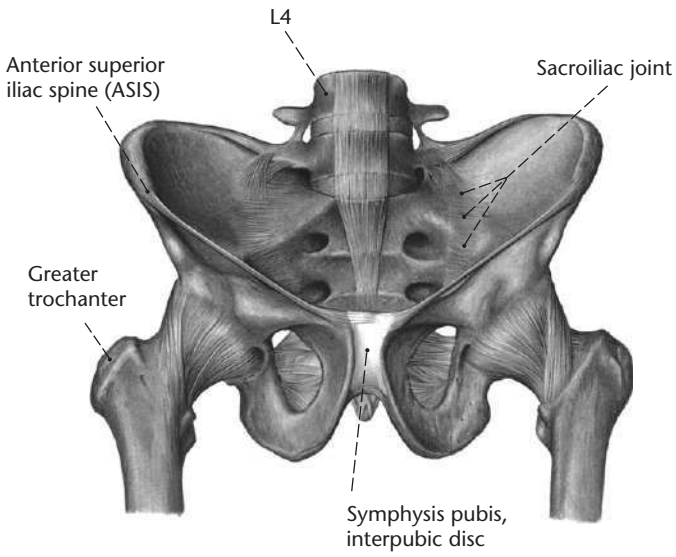

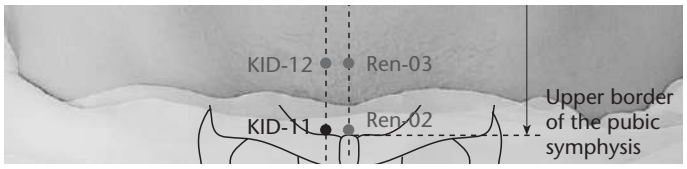

Ch05a-F10028.qxd 2/22/08 9:43 PM Page 473

Middle Pole ZHONGJI Ren-3

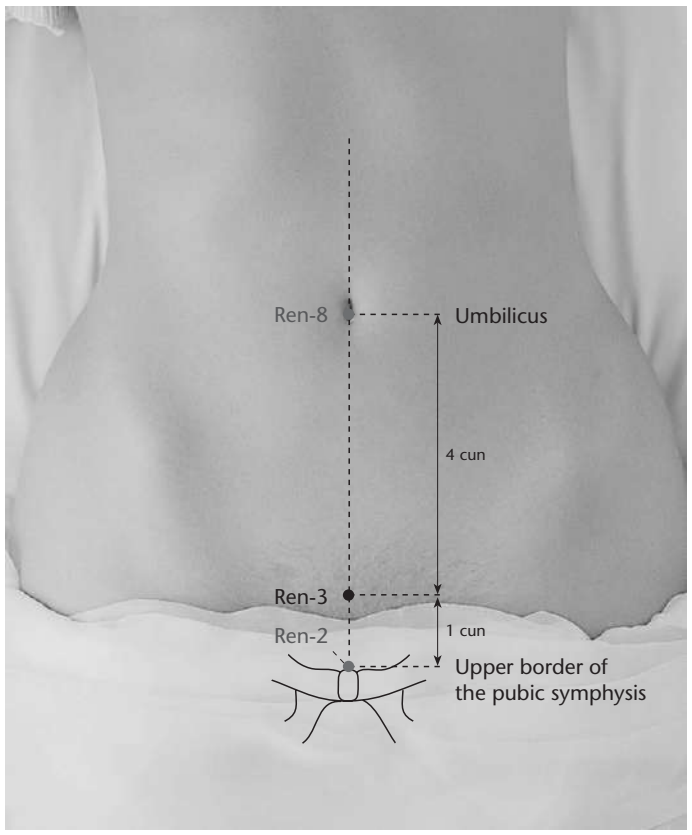

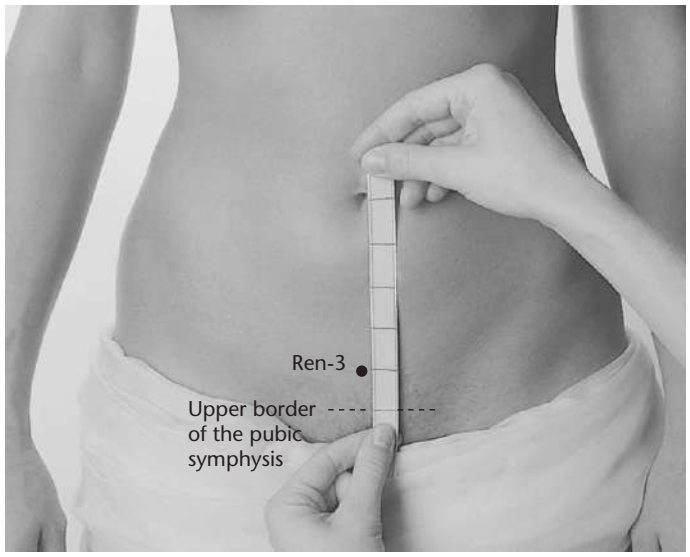

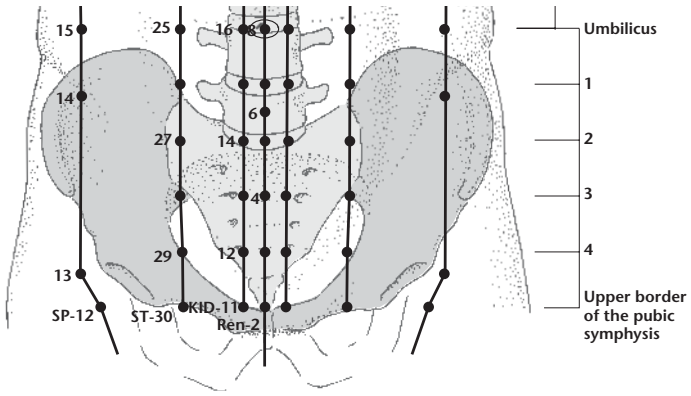

Location

On the anterior midline, 1 cun superior to the upper border of the pubic symphysis or 4 cun inferior to the umbilicus.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 cun. As these can vary considerably from the patient's finger cun, only proportional cun measurements should be used here (helpful device: elastic tape → 2.3.1). Locate **Ren-3** on the midline, 1 cun superior to the midpoint of the upper border of the pubic symphysis.

Located on the same level are → **KID-12** (0.5 cun lateral to the midline) and → **ST-29** (2 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, bladder. Contraindicated during pregnancy. Ask the patient to void their bladder before needling.

Actions/Indications

- Benefits the Bladder, drains Dampness and Damp-Heat, dispels stagnation, benefits the Lower Burner
- Regulates menstruation
- Strengthens the Kidneys (→ **Ren-4** is more commonly used for this purpose)

Special features

Front-*mu* point of the Bladder, meeting point with the LIV, KID and SP channels. Important point for disorders of the urogenital tract, with its main action on the Bladder (especially for acute cases).

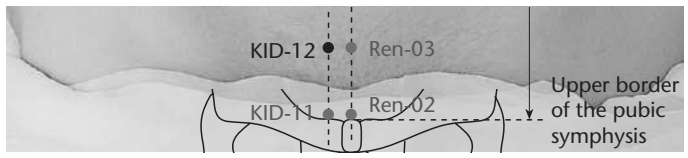

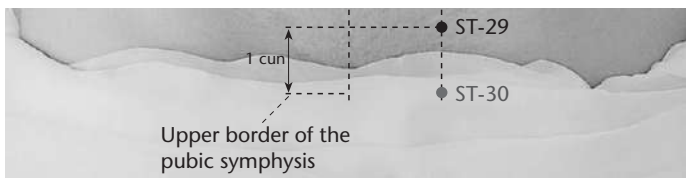

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 474

Ren-4 *Gate of the Source* GUANYUAN

Location

On the anterior midline, 2 cun superior to the upper border of the pubic symphysis or 3 cun inferior to the umbilicus.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used here (helpful device: elastic tape → 2.3.1). **Ren-4** is located on the midline, either 2 cun superior to the midpoint of the upper border of the pubic symphysis or 3 cun inferior to the umbilicus.

Located on the same level are → **KID-13** (0.5 cun lateral to the midline), → **ST-28** (2 cun lateral to the midline) and three extra points → **Ex-CA** (*yijing*: 1 cun lateral to the midline; *qimen*: 3 cun lateral to the midline; *tituo*: 4 cun lateral to the midline). Also located on approximately this level is → **G.B.-27** (medial to the ASIS).

Needling

Vertically 0.5–1.2 cun. Caution: Peritoneum. Needling contraindicated during pregnancy.

Actions/Indications

- Tonifies the Original Qi (*yuan qi*), benefits the Essence (*jing*), strengthens and nourishes the Kidneys as well as the Blood and Yin
- Eliminates Cold and Dampness from the Lower Burner (especially with moxibustion)

Special features

Front-*mu* point of the Small Intestine, meeting point with the SP, KID and LIV channels. Important point for tonification. Major point for urogenital disorders (especially gynaecological disorders).

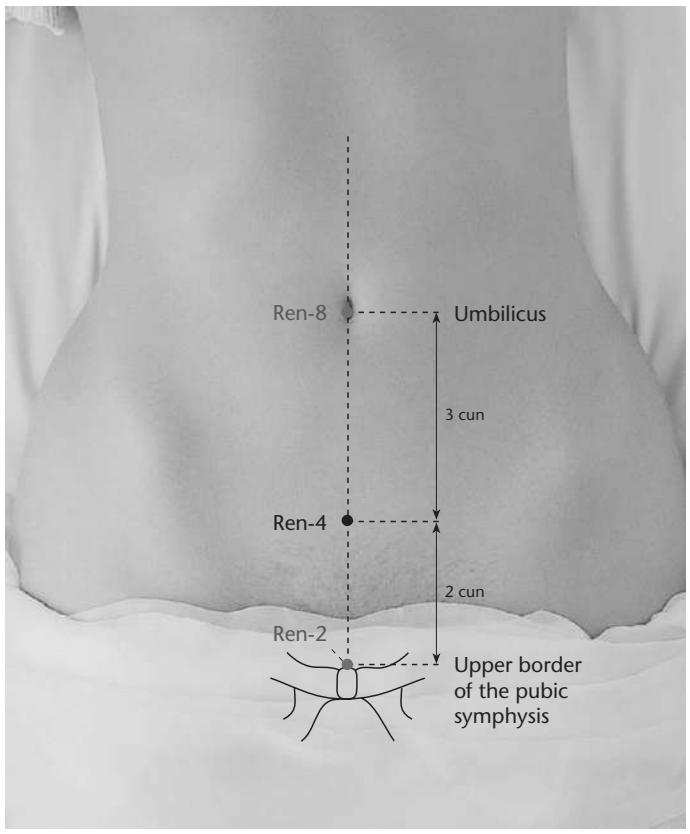

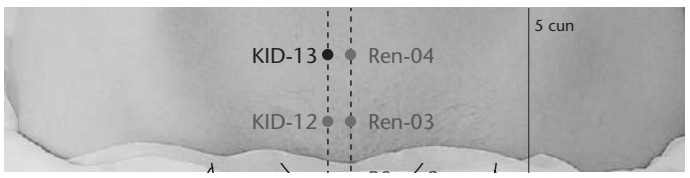

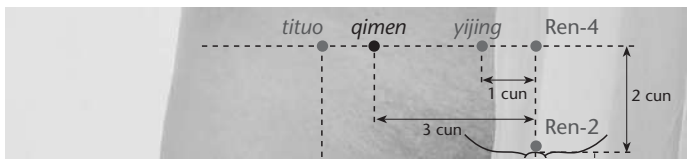

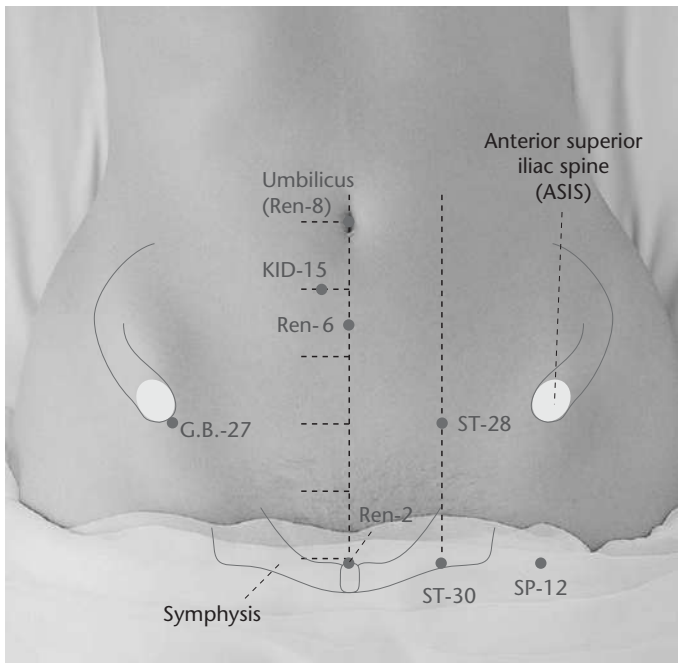

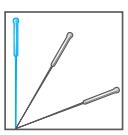
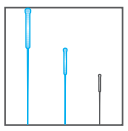
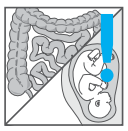
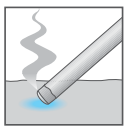
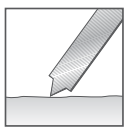
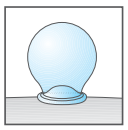

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 475

Stone Gate SHIMEN Ren-5

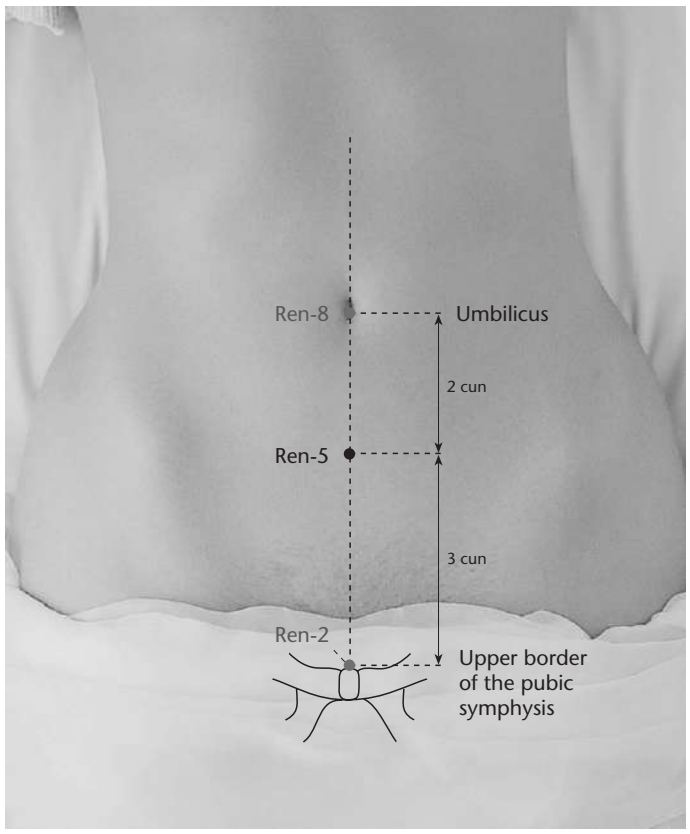

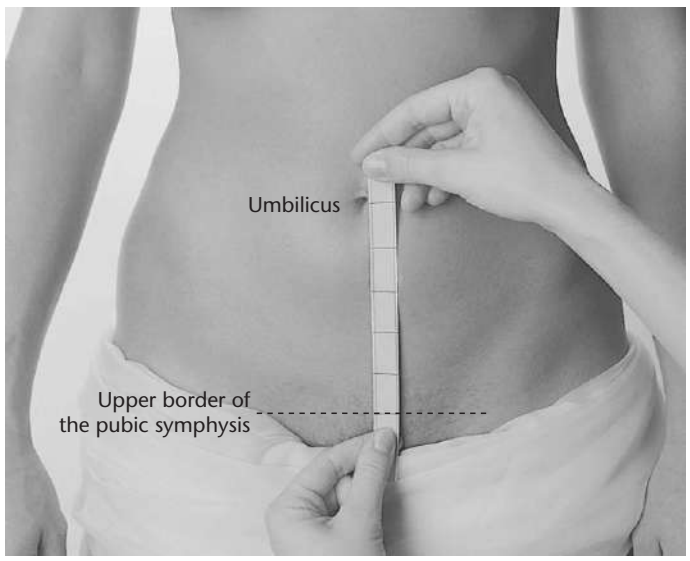

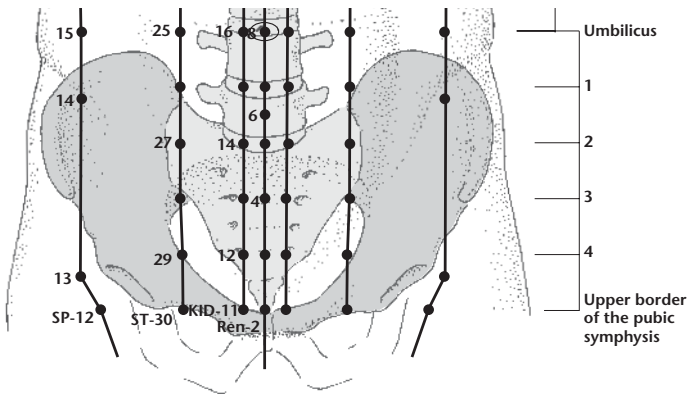

Location

On the anterior midline, 2 cun inferior to the umbilicus or 3 cun superior to the upper border of the pubic symphysis.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used here (helpful device: elastic tape → 2.3.1). **Ren-5** is located on the midline, 2 cun inferior to the umbilicus. Located on the same level are → **KID-14** (0.5 cun lateral to the midline) and → **ST-27** (2 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Moves and benefits the water passages
- Regulates the Qi and Blood in the Lower Burner
- Regulates the Uterus

Special features

Front-*mu* point of the Triple Burner. Some classic texts warn that needling this point in women can lead to infertility (hence the name: infertile women were referred to as 'stone women'). There are no such warnings in modern texts.

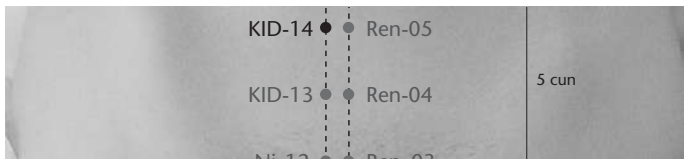

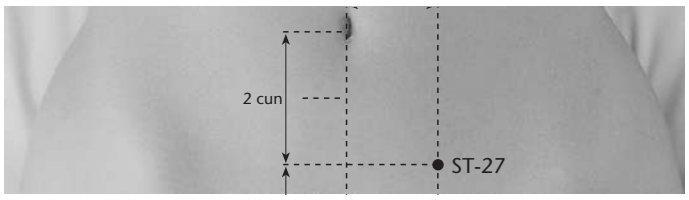

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 476

Ren-6 *Sea of Qi* QIHAI

Location

On the anterior midline, 1.5 cun inferior to the umbilicus or 3.5 cun superior to the upper border of the pubic symphysis.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used here (helpful device: elastic tape → 2.3.1). **Ren-6** is located on the midline, 1.5 cun inferior to the umbilicus or 3.5 cun superior to the upper border of the pubic symphysis.

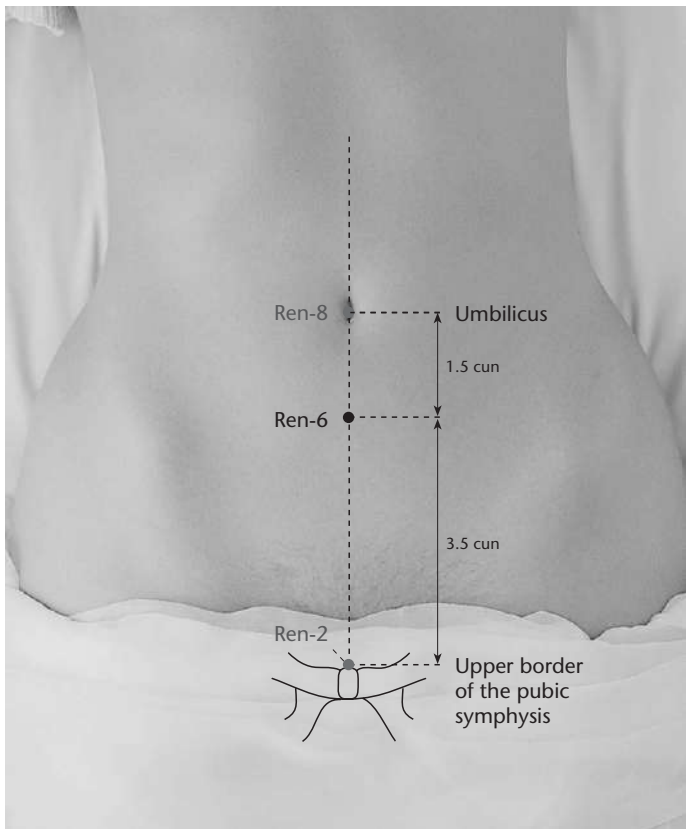

Needling

Vertically 0.8–1.5 cun. Caution: Peritoneum, pregnancy. According to some authors, this point should never be needled with strong reducing techniques owing to its strong Qi tonifying action. Moxibustion and cupping are recommended for this point.

Actions/Indications

- Tonifies the Original Qi (*yuan qi*) as well as Qi in general, strengthens the Kidneys (especially Kidney Yang), regulates Qi, harmonises Blood and eliminates Dampness

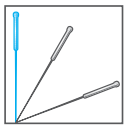
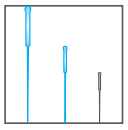
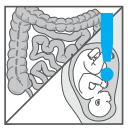

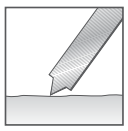
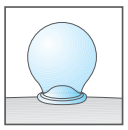

Special features

Important tonification point for mental and physical exhaustion, strengthens especially the Qi and Yang with moxibustion. One of the most important acupuncture points.

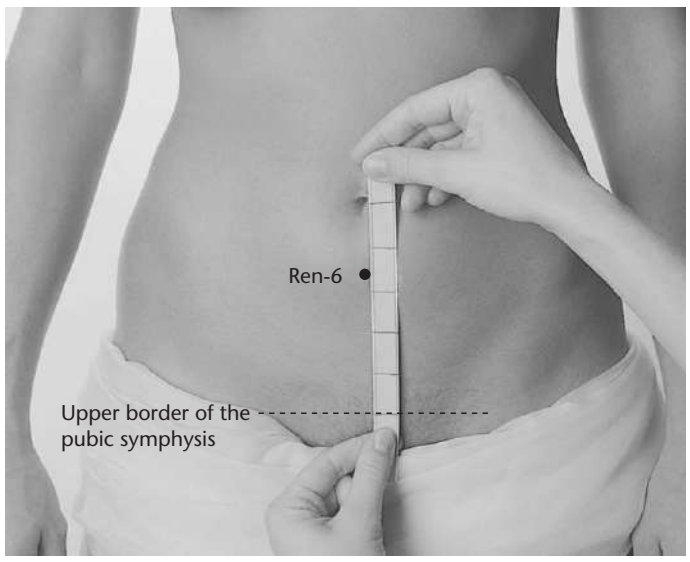

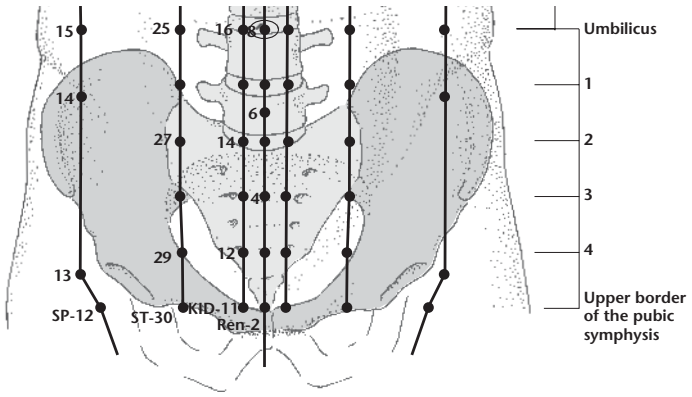

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 477

Yin Intersection YINJIAO Ren-7

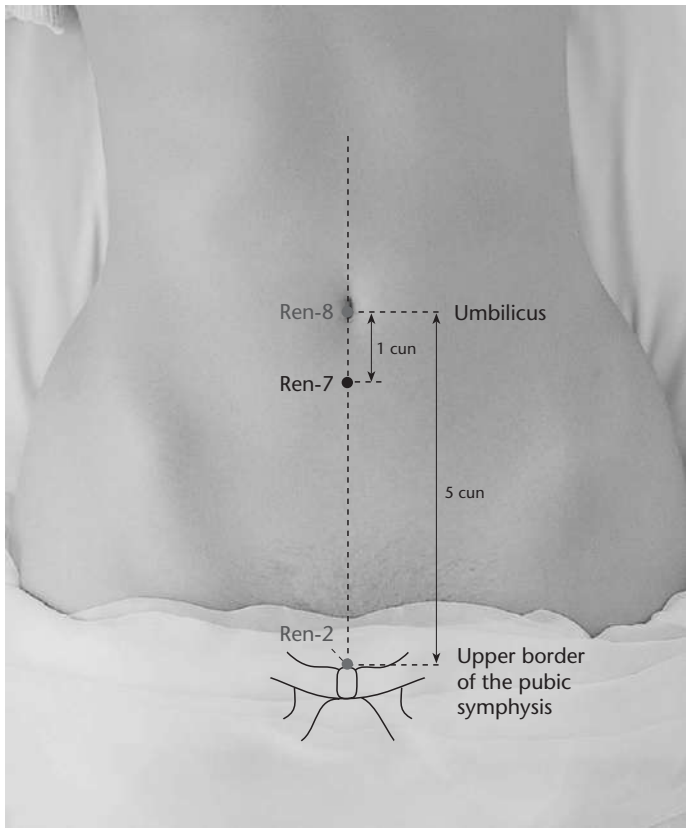

Location
On the anterior midline, 1 cun inferior to the umbilicus.

How to find
The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used here (helpful device: elastic tape → 2.3.1). **Ren-7** is located on the midline, 1 cun inferior to the umbilicus.
Located on the same level are → **KID-15** (0.5 cun lateral to the midline) and → **ST-26** (2 cun lateral to the midline).

Needling
Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications
- Regulates menstruation
- Regulates the Qi flow in the abdomen and the genital area

Special features
Meeting point with the *chong mai* and the KID channel

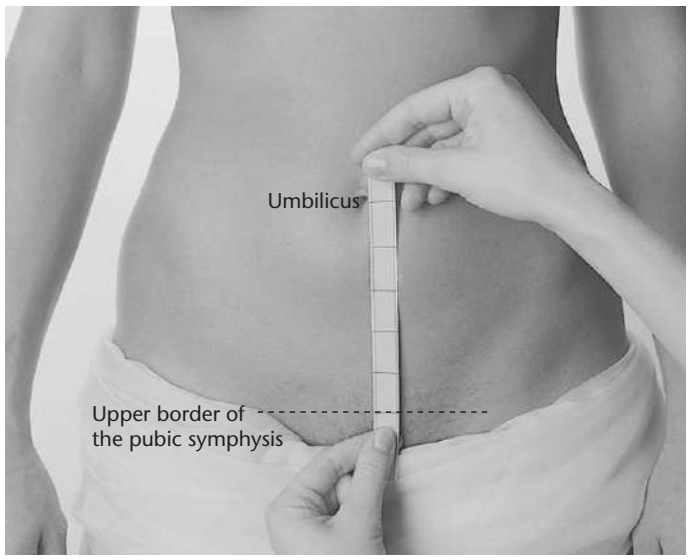

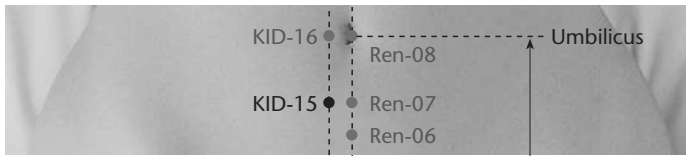

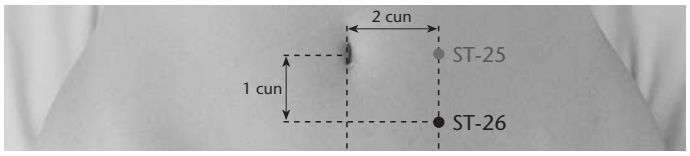

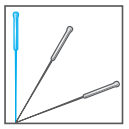
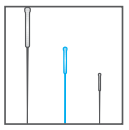
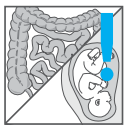

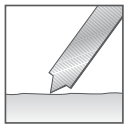
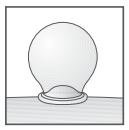

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 478

Ren-8 *Spirit Gateway* SHENQUE

Location

In the centre of the umbilicus.

How to find

Located on the same level are → **KID-16** (0.5 cun lateral to the midline), → **ST-25** (2 cun lateral to the midline), → **SP-15** (4 cun lateral to the midline) as well as → **G.B.-26** (on a vertical line through the free end of the 11th rib).

Needling

Needling contraindicated. Only indirect moxibustion (on salt or ginger, moxa pole, moxa box).

Actions/Indications

- Warms and stabilises the Yang and the Intestines

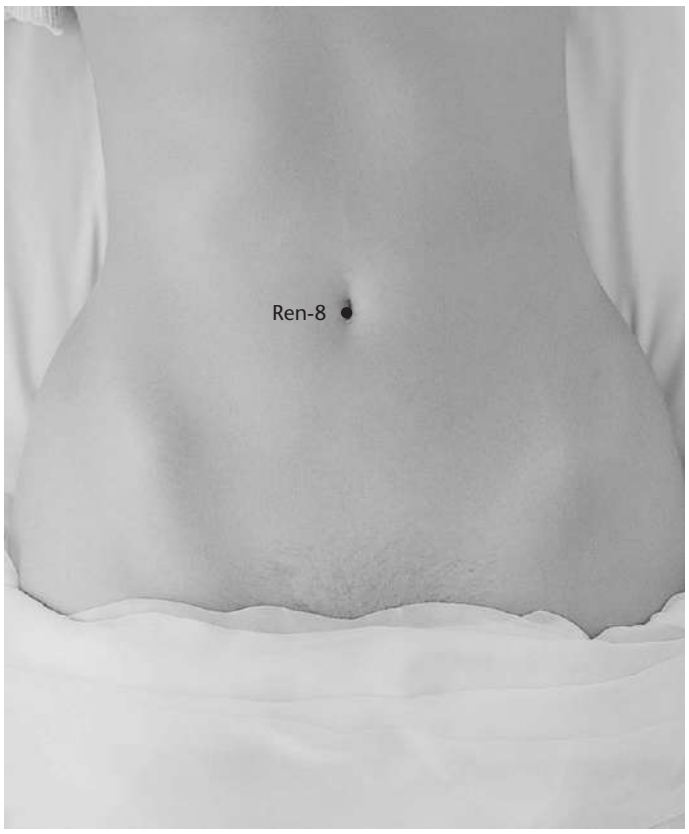

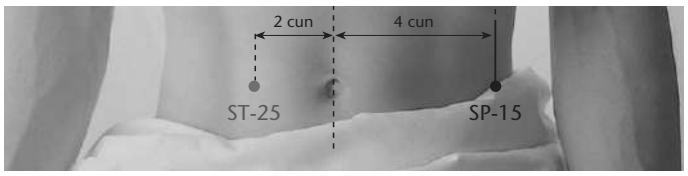

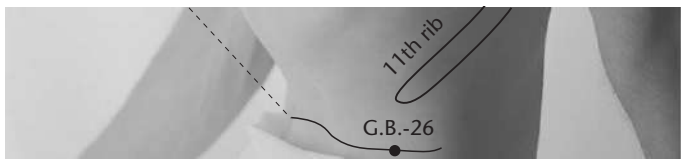

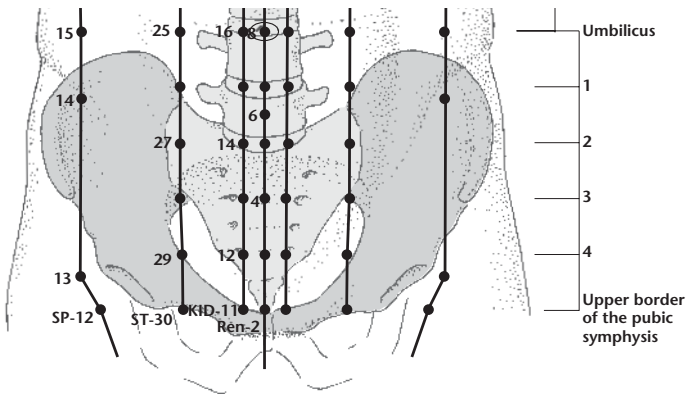

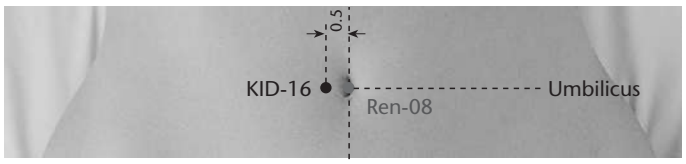

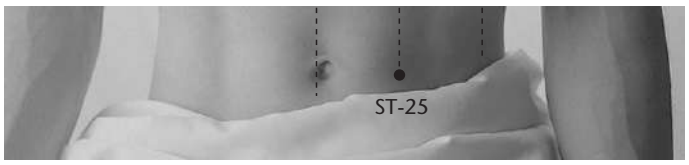

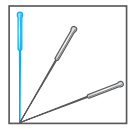
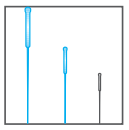

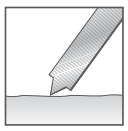
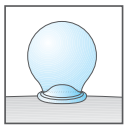

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 479

Water Separation SHUIFEN Ren-9

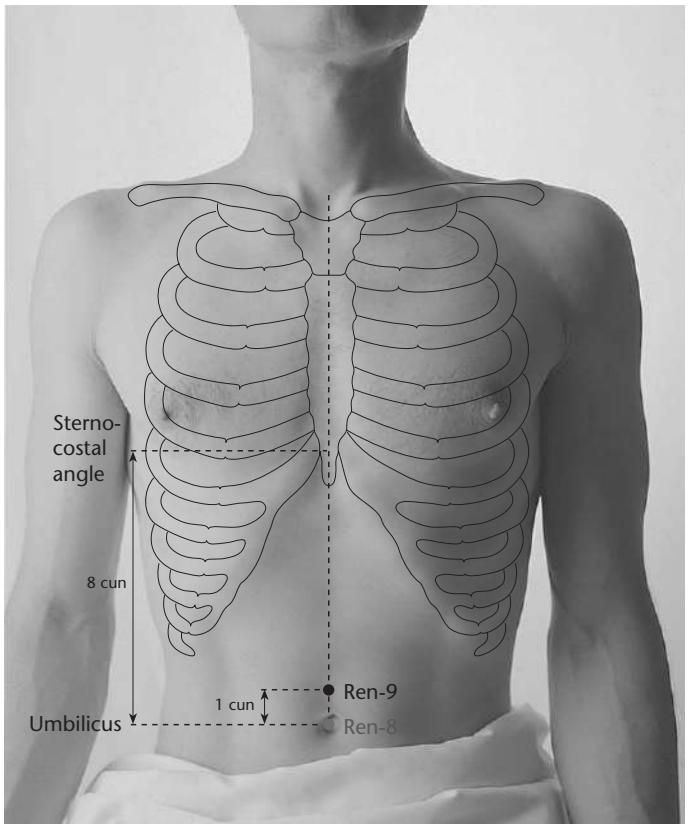

Location

On the anterior midline, 1 cun superior to the umbilicus.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be applied here (helpful device: elastic tape → 2.3.1). From the centre of the umbilicus, measure 1 cun in a superior direction and there locate **Ren-9** on the anterior midline.

Located on the same level is → **ST-24** (2 cun lateral to the midline). → **LIV-13** is also located on approximately this level (at the free end of the 11th rib).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy

Actions/Indications

- Eliminates water accumulations, drains Dampness and controls the water passages Regulates the Intestines and disperses accumulations
- Delayed closure of the fontanelles

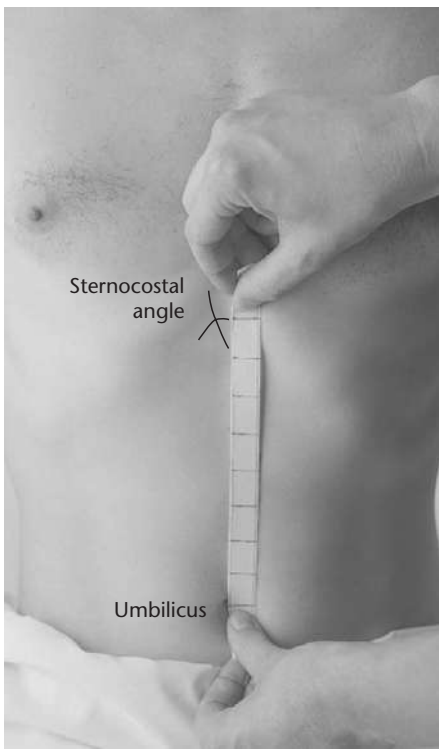

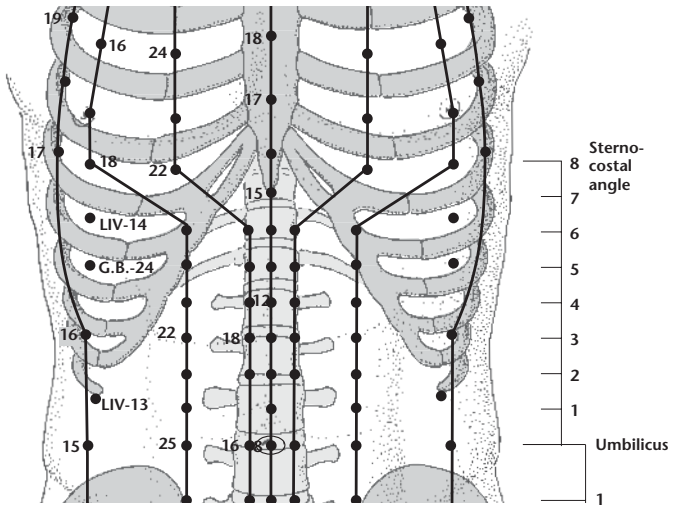

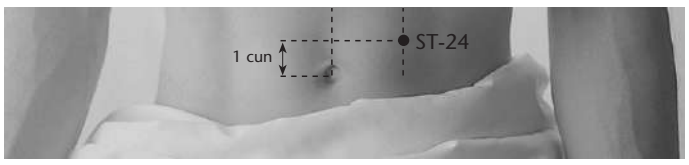

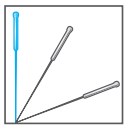
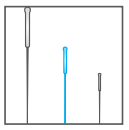
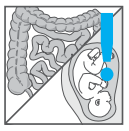

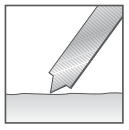

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 480

Ren-10 *Lower Cavity* XIAWAN

Location

On the anterior midline, 2 cun superior to the centre of the umbilicus.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun. Therefore only proportional cun measurements should be applied here (helpful device: elastic tape → 2.3.1). From the centre of the umbilicus, measure 2 cun in a superior direction and there locate **Ren-10** on the anterior midline.

Located on the same level are → **KID-17** (0.5 cun lateral to the midline), → **ST-23** (2 cun lateral to the midline) and **Ex-CA** (*weishang*; 4 cun lateral to the midline).

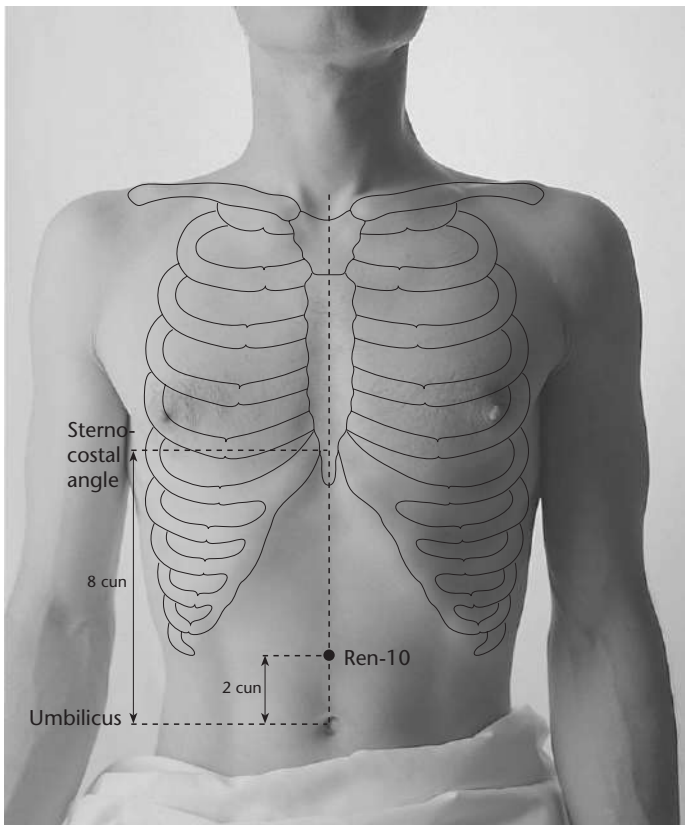

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Tonifies and regulates the Spleen and Stomach Qi, dispels food stagnation

Special features

Meeting point with the SP channel and according to some authors with the ST channel. According to some classical texts, this point is contraindicated during pregnancy (for both needling and moxibustion).

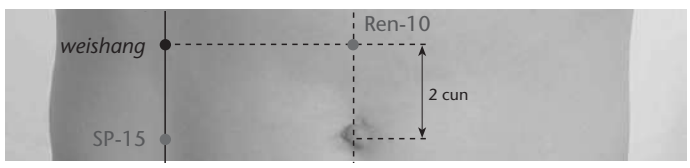

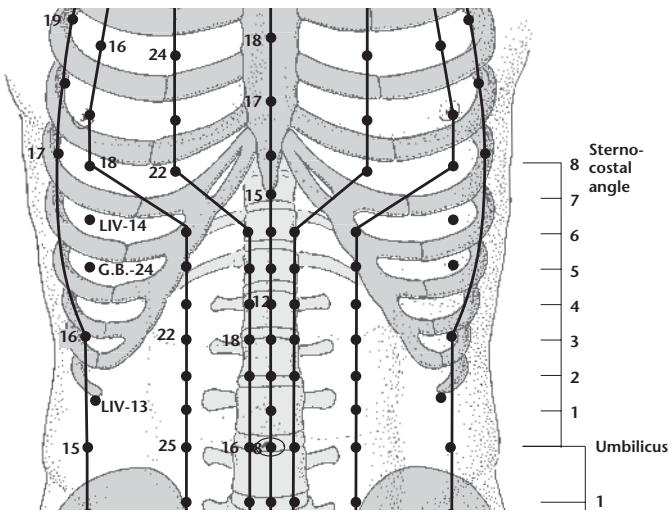

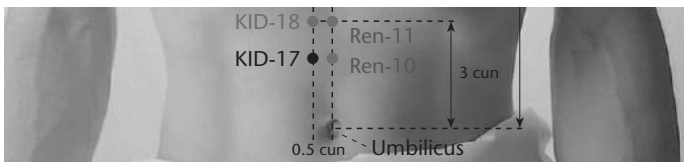

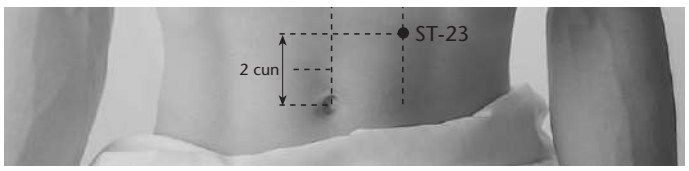

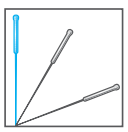
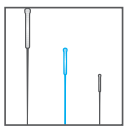
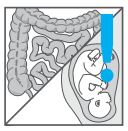

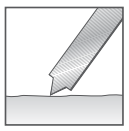

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 481

Interior Strengthening JIANLI Ren-11

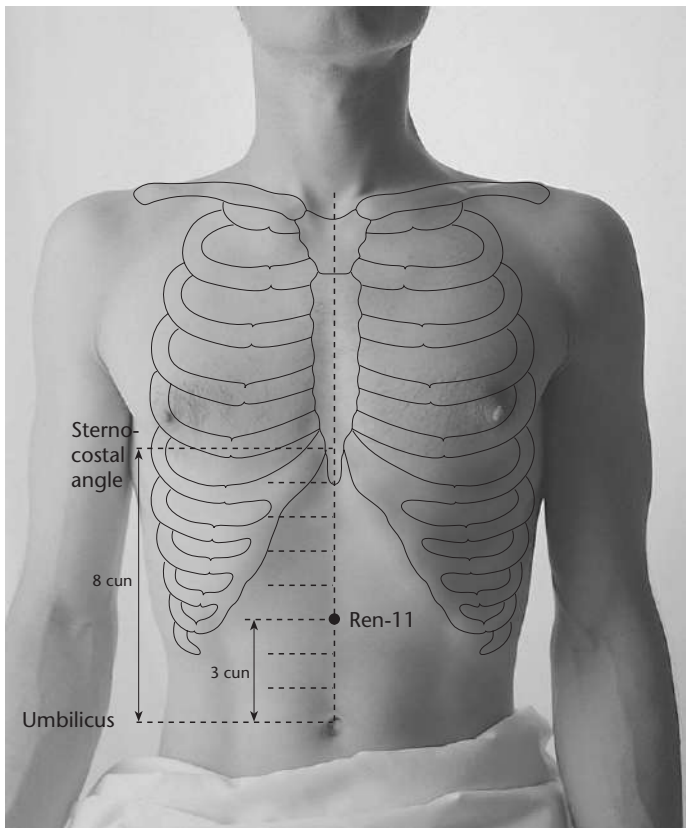

Location

On the anterior midline, 3 cun superior to the umbilicus.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be applied here (helpful device: elastic tape → 2.3.1). From the centre of the umbilicus, measure 3 cun in a superior direction and there locate **Ren-11** on the anterior midline. Located on the same level are → **KID-18** (0.5 cun lateral to the midline), → **ST-22** (2 cun lateral to the midline) and → **SP-16** (4 cun lateral to the midline).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Harmonises the Middle Burner, regulates the Qi flow

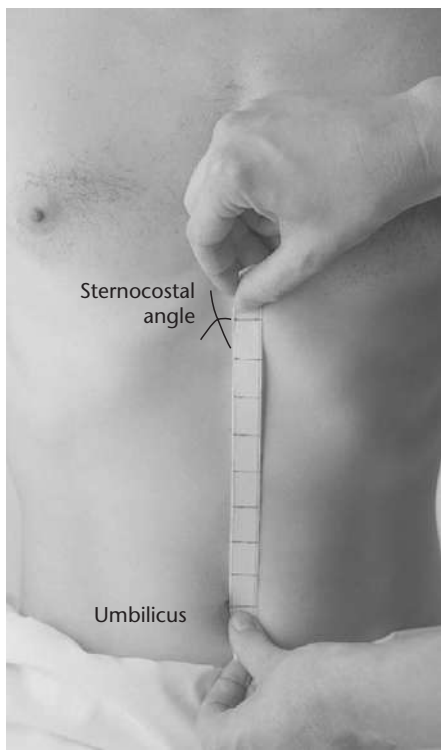

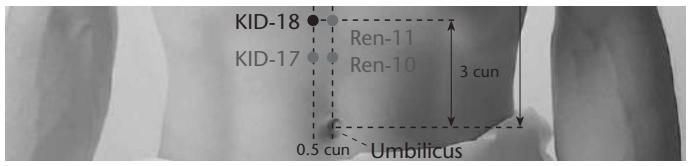

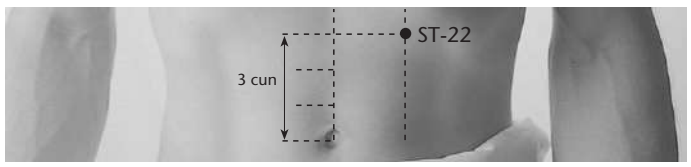

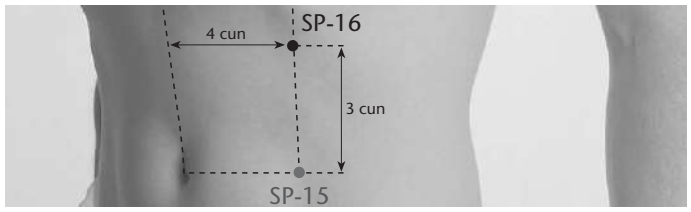

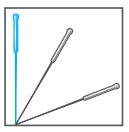

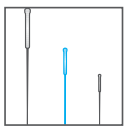

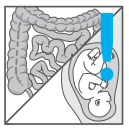

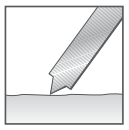

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 482

Ren-12 *Middle Cavity* ZHONGWAN

Location

On the anterior midline, 4 cun superior to the umbilicus or 4 cun inferior to the sternocostal angle.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus can be divided into 8 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be applied here (helpful device: elastic tape → 2.3.1). From the centre of the umbilicus, measure either 4 cun in a superior direction or from the sternocostal angle 4 cun in an inferior direction. There, locate **Ren-12** on the anterior midline. **Or:** Spreading hands technique (→ 2.3.3): Place the little fingers on the sternocostal angle and the centre of the umbilicus respectively and determine the midpoint (=**Ren-12**) of this distance by joining the index fingers. Located on the same level are → **KID-19** (0.5 cun lateral to the midline) and → **ST-21** (2 cun lateral to the midline).

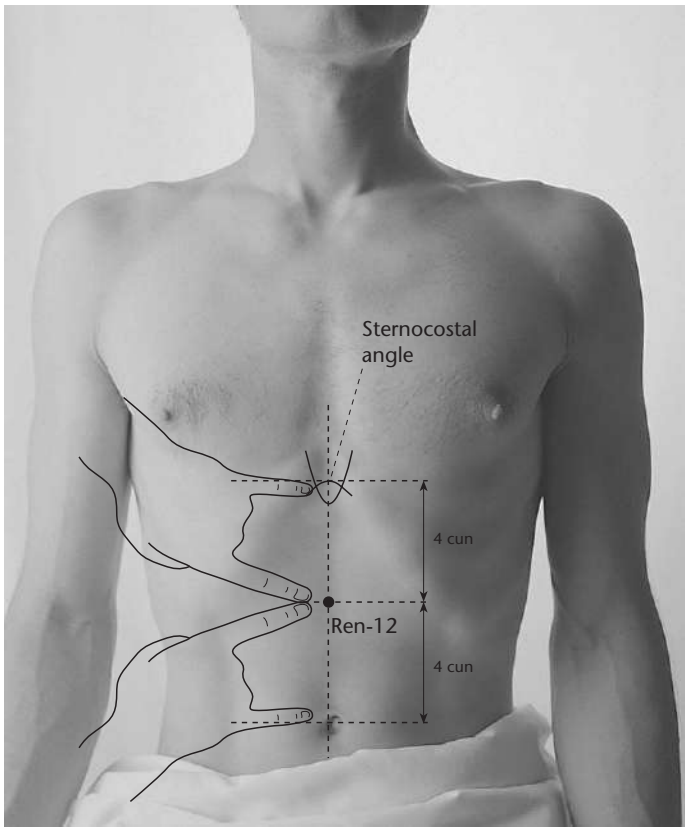

Needling

0.8–1.5 cun vertically or obliquely towards surrounding points (→ **ST-21**, **Ren-10**, **Ren-15**). Caution: Peritoneum, pregnancy. If indicated, moxa is recommended.

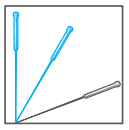

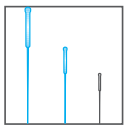

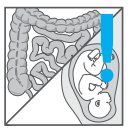

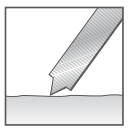

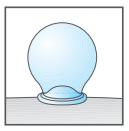

Actions/Indications

- Harmonises and tonifies the Middle Burner, descends counterflow Qi, regulates the Qi flow and alleviates pain, transforms Dampness

Special features

Front-*mu* point of the stomach, meeting point with the S.I., T.B.* and ST channels, *hui*-meeting point of the *fu*-Organs. Major point for (functional) stomach disorders.

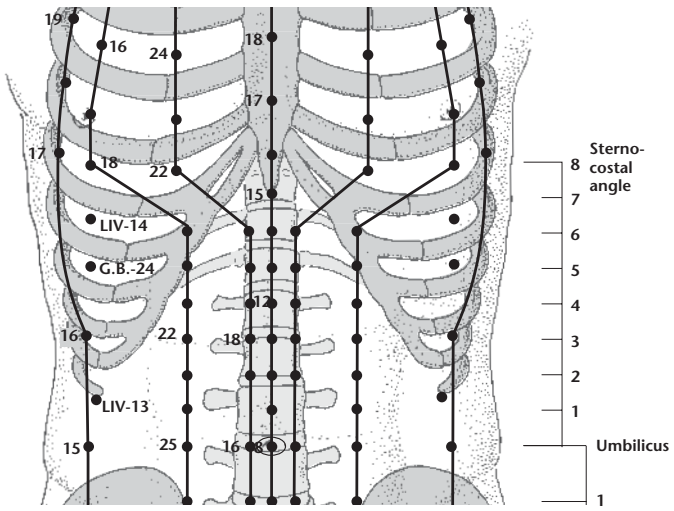

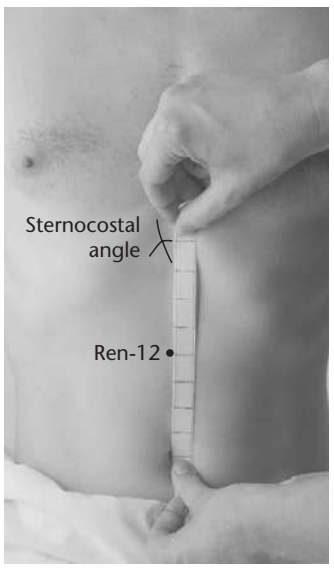

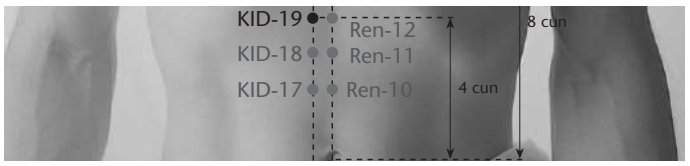

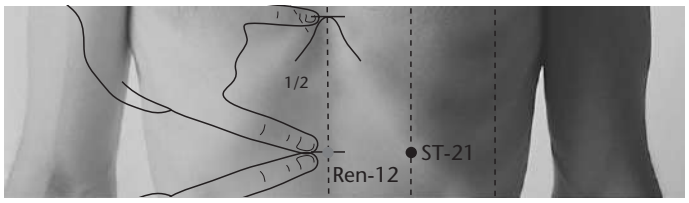

* Mentioned by only some authors

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 483

Upper Cavity SHANGWAN Ren-13

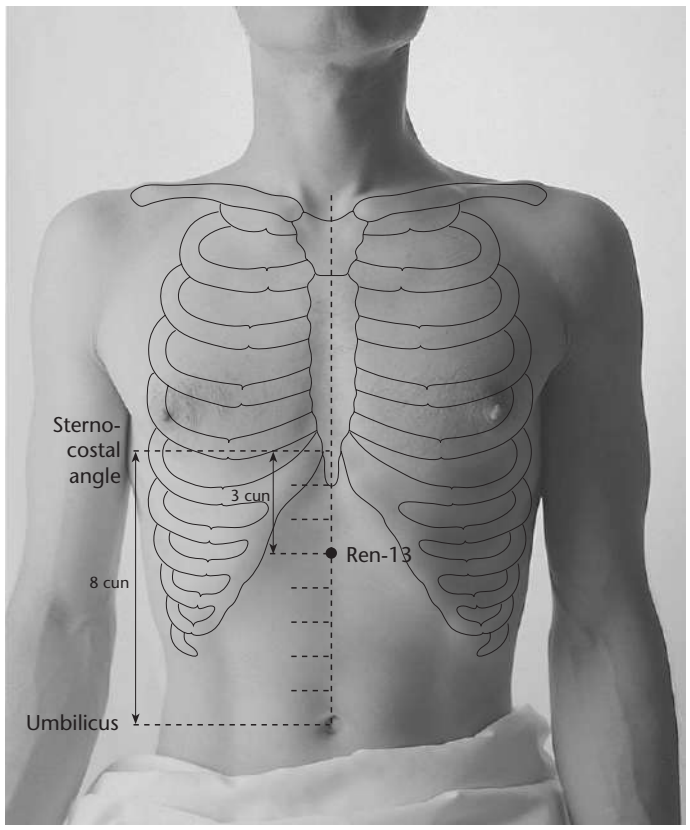

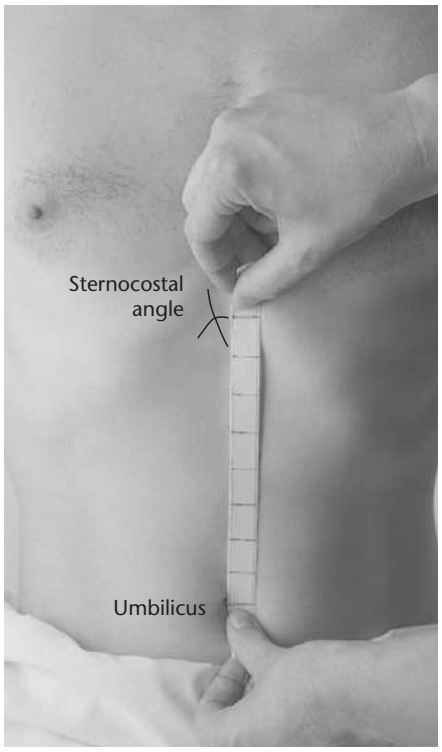

Location

On the anterior midline, 3 cun inferior to the sternocostal angle.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be applied here (helpful device: elastic tape → 2.3.1). From the sternocostal angle, measure 3 cun in an inferior direction and there locate **Ren-13** on the anterior midline.

Located on the same level are → **KID-20** (0.5 cun lateral to the midline), → **ST-20** (2 cun lateral to the midline) and → **G.B.-24** (approximately on this level, in the 7th intercostal space, on the mamillary line).

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Regulates the Stomach and counterflow Qi
- Regulates the Heart

Special features

Meeting point with the ST and S.I. channel

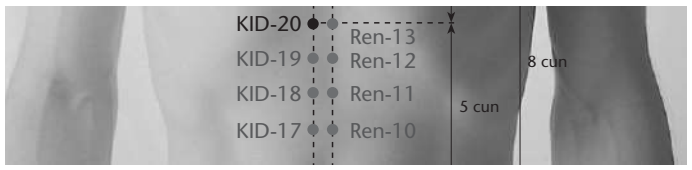

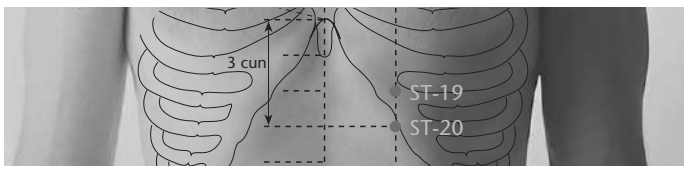

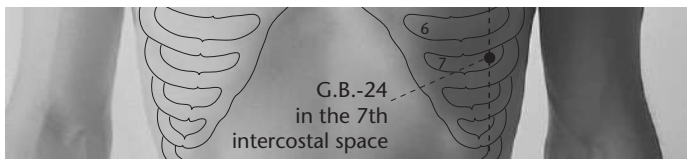

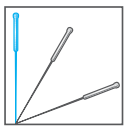
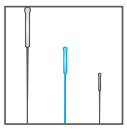
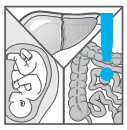

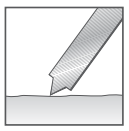

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 484

Ren-14 *Great Gateway* JUQUE

Location

On the anterior midline, 2 cun inferior to the sternocostal angle or 6 cun superior to the umbilicus.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 cun (note: these are proportional cun → 2.2). **Ren-14** is located on the midline, 2 cun inferior to the sternocostal angle or 6 cun superior to the umbilicus. **Or:** Use spreading hands technique (→ Chapter 2): Place the little fingers on the sternocostal angle and the umbilicus respectively and determine the midpoint of this distance (→ **Ren-12**). Then locate **Ren-14** on the midpoint of the distance between → **Ren-12** and the sternocostal angle.
Located on the same level are → **KID-21** (0.5 cun lateral to the midline), → **ST-19** (2 cun lateral to the midline) and → **LIV-14** (approximately on this level, in the 6th intercostal space, on the mamillary line).

Needling

0.3–0.8 cun vertically or often obliquely in an inferior direction or up to 1.5 cun transversely (subcutaneously). Depending on individual anatomy, this point can be located on the xiphoid process. Caution: Peritoneum, injury to hypertrophied organs (left: heart, right: liver) depending on the thickness of the abdominal wall.

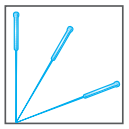

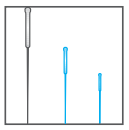

Actions/Indications

- Regulates the Heart, disperses stagnant Phlegm, alleviates pain, unbinds the thorax, descends the Qi
- Harmonises the Stomach, descends Stomach Qi
- Transforms Phlegm, calms the *shen*

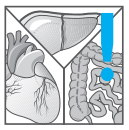

Special features

Front-*mu* point of the Heart. Major local point for thoracic pain caused by Blood stasis or Phlegm blockages.

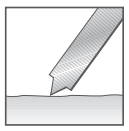

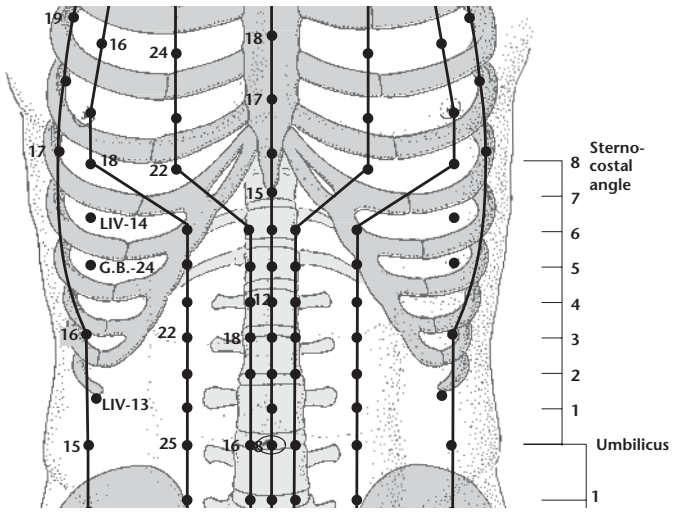

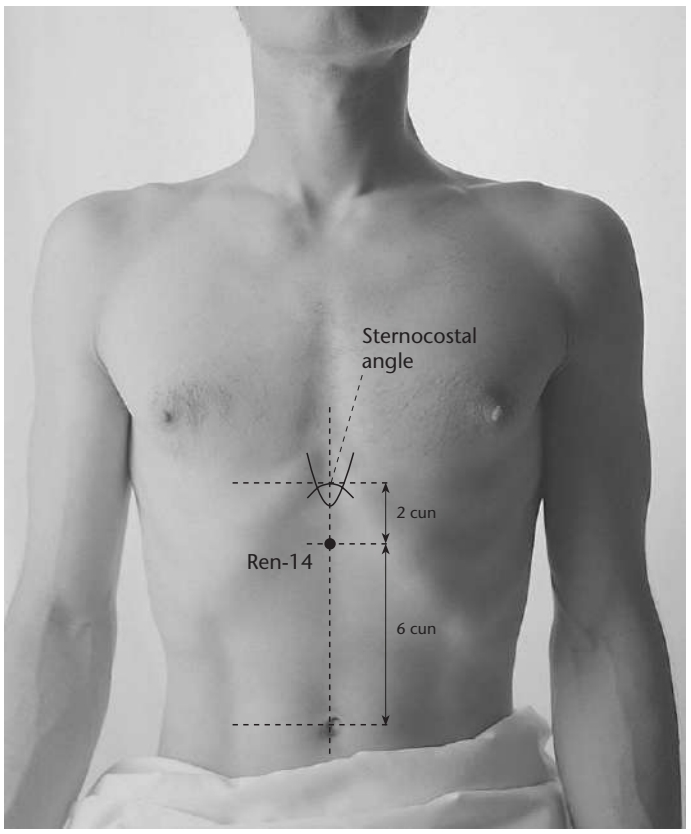

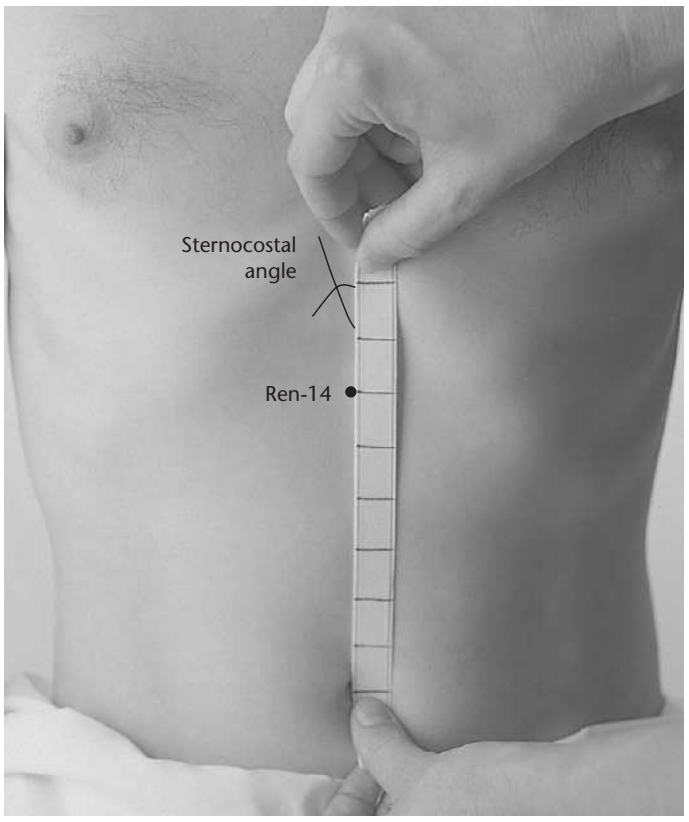

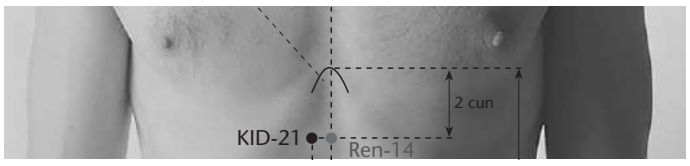

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 485

Turtledove Tail JIUWEI Ren-15

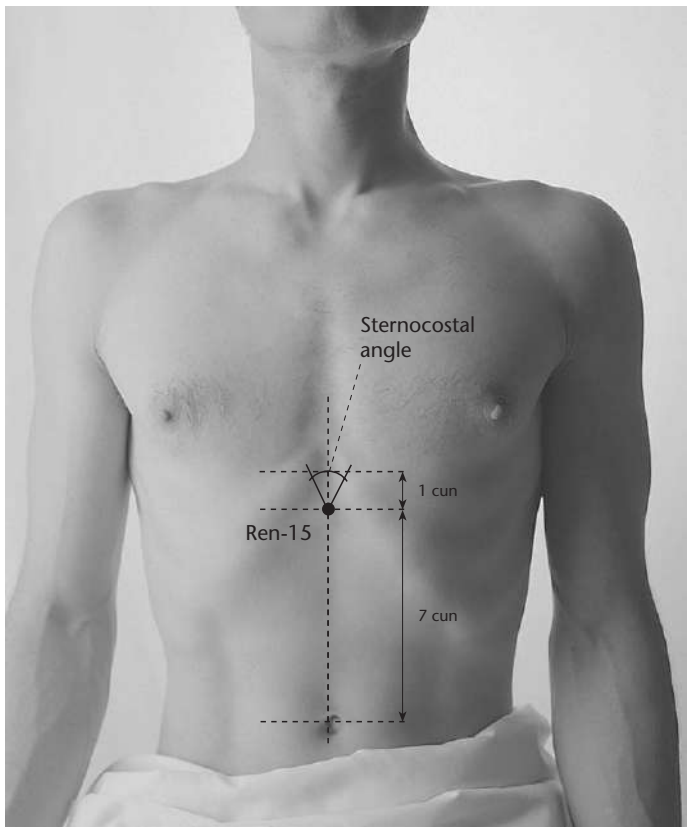

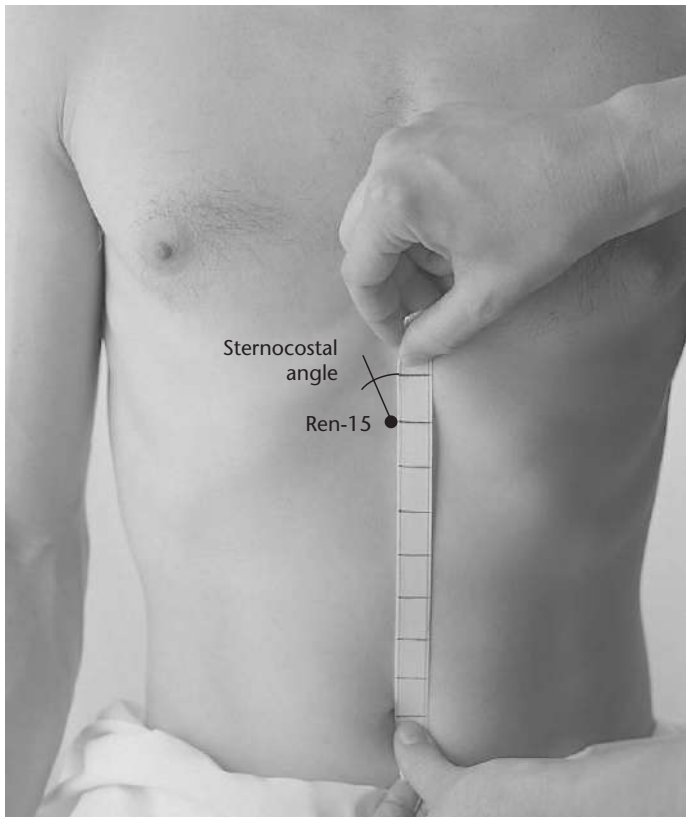

Location

On the anterior midline, 1 cun inferior to the sternocostal angle or 7 cun superior to the umbilicus.

How to find

The distance between the sternocostal angle (→ 3.5) and the centre of the umbilicus is divided into 8 cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be applied here (helpful device: elastic tape → 2.3.1). From the sternocostal angle, measure 1 cun in an inferior direction and there locate **Ren-15** on the anterior midline. It is often located directly below or, depending on individual anatomy, on the xiphoid process.

Needling

0.5–1 cun obliquely in an inferior direction. The patient should be in an extended position. Caution: Possible injury to organs with deep needling or oblique insertion in a superior direction, especially in patients with hypertrophied organs (left: Liver, right: Heart). According to some classical texts, moxibustion is contraindicated.

Actions/Indications

- Regulates the Heart, calms the *shen*
- Descends counterflow lung and stomach Qi, unbinds the chest
- Regulates the *luo*-connecting vessel of the *ren mai*

Special features

Luo-connecting point of the *ren mai*. Important point for calming.

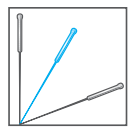

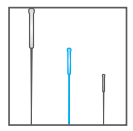

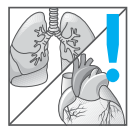

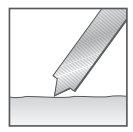

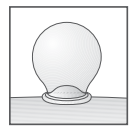

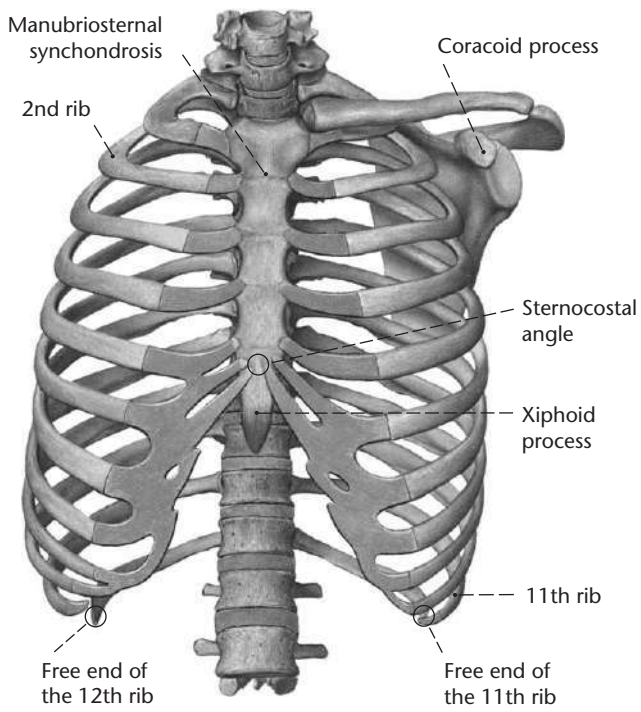

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 486

Ren-16 *Central Courtyard* ZHONGTING

Location

On the anterior midline, on the level of the sternocostal angle.

How to find

Ren-16 is located on the level of the sternocostal angle, where the two lower borders of the ribcage (→ 3.5) meet. To locate this point, palpate upward along the lower borders of the ribcage with one hand on either side until the palpating fingers meet at the sternocostal angle. Here, locate **Ren-16** in a shallow depression between the sternum and the xiphoid process.

Located on the same level (approximately the 5th intercostal space) are → **KID-22** (2 cun lateral to the midline), → **ST-18** (4 cun lateral to the midline) and → **SP-17** (6 cun lateral to the midline).

Needling

0.5 cun transversely (subcutaneously) in a superior or inferior direction. Caution: Danger of heart injury in patients with a split xiphoid process or a sternal foramen. Therefore, only transverse needling is recommended.

Actions/Indications

- Unbinds the chest
- Regulates the Stomach and counterflow Stomach Qi

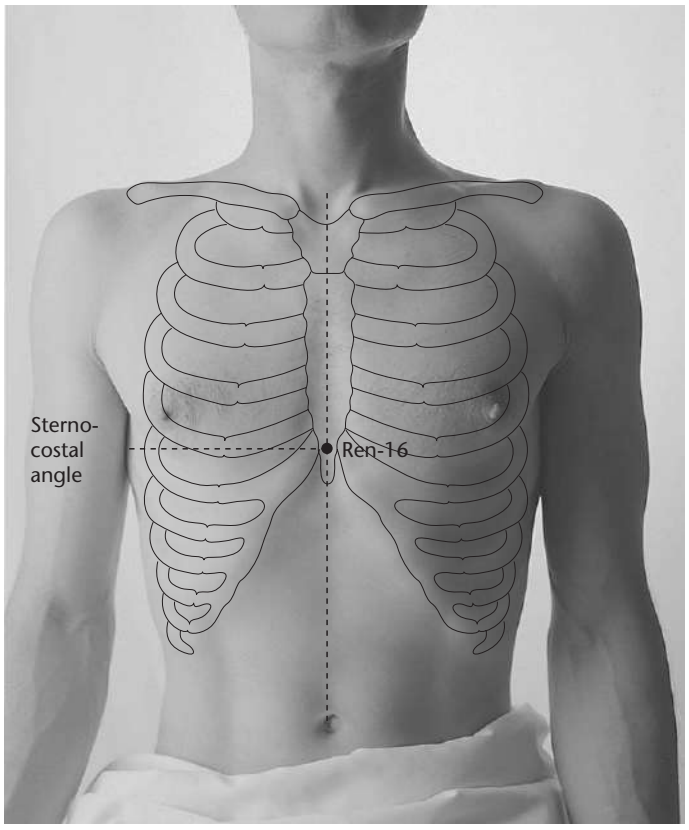

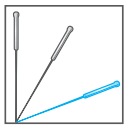

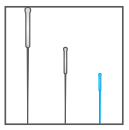

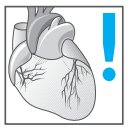

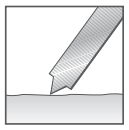

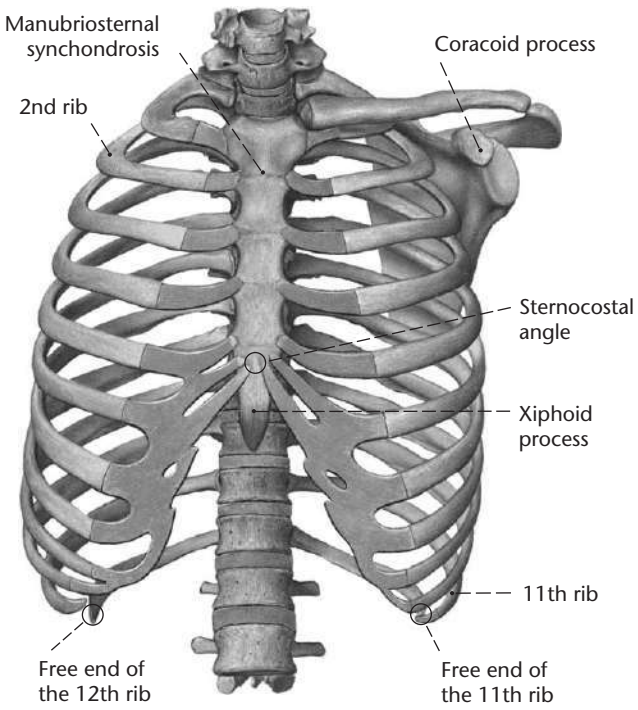

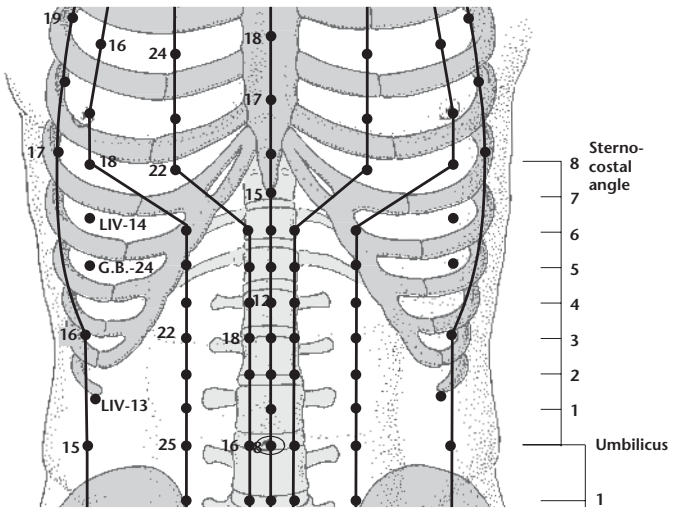

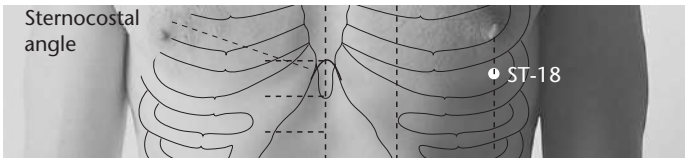

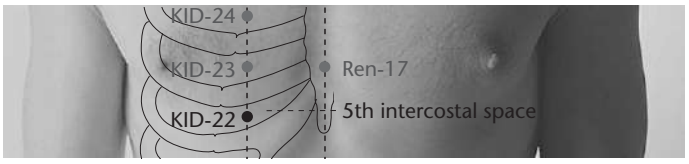

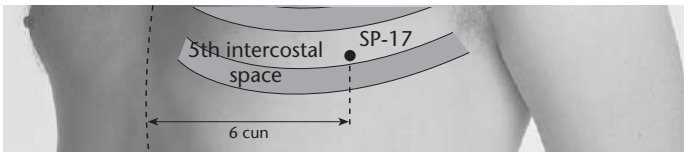

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 487

Chest Centre DANZHONG Ren-17

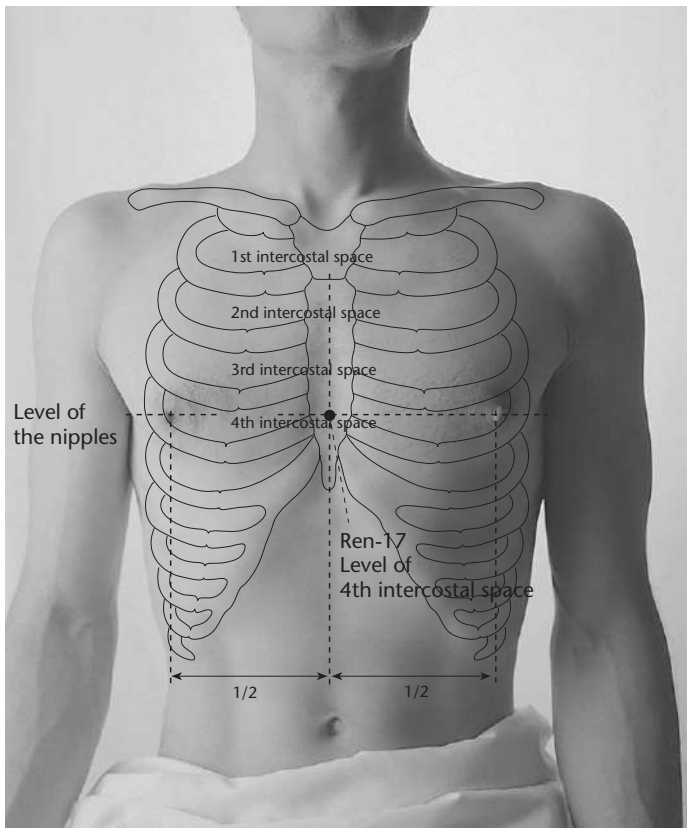

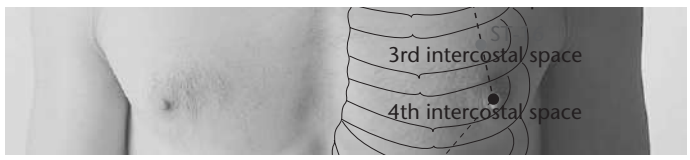

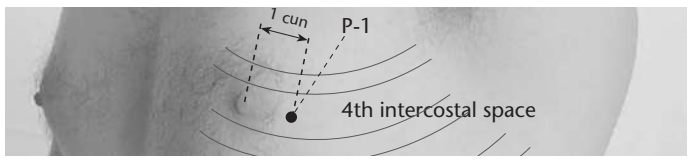

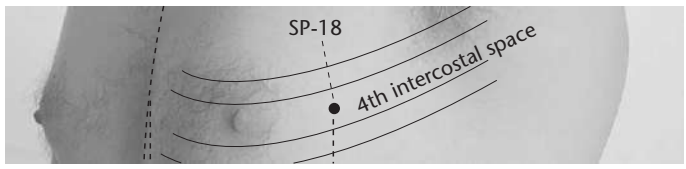

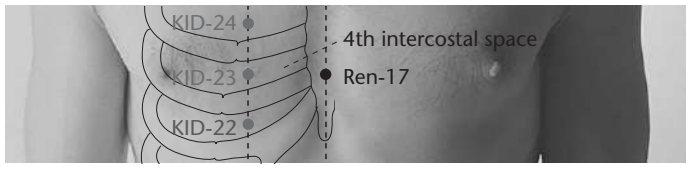

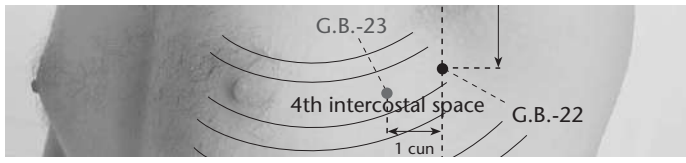

Location

On the anterior midline, on the level of the 4th intercostal space.

How to find

Quick method, especially in men: Locate **Ren-17** on the sternum, between the two nipples. For more detailed orientation on the intercostal region (→ 3.5), first, locate the manubriosternal synchondrosis, which can be felt on the sternum as a well-defined horizontal bony structure on the level of the costal cartilage of the 2nd rib. The 2nd intercostal space is below it. From there, count in an inferior direction to the 4th intercostal space. Locate **Ren-17** on this level on the midline of the sternum. Located on the same level in the 4th intercostal space are → **KID-23/ST-17/P-1/SP-18/G.B.-22/G.B.-23** (2 cun lateral to the midline/on the nipple/1 cun lateral to the nipple/6 cun lateral to the midline/on the axillary line/1 cun anterior to → **G.B.-22**).

Needling

Depending on the indication, 0.5–1 cun transversely (subcutaneously) in a superior or inferior direction or towards the nipples (for mamillary disorders). Caution: Intracardiac needling, pneumothorax. The sternum can be very thin in this area owing to embryonic ossification disorders! Moxibustion is recommended in order to tonify the Qi, but caution with asthmatic patients.

Actions/Indications

- Regulates and benefits the Qi, unbinds the chest, descends counterflow Lung and Stomach Qi
- Benefits the breasts

Special features

Front-*mu* point of the Pericardium, meeting point with the SP, KID, S.I. and T.B. channels. *Hui*-meeting point of the Qi. Sea of Qi. Important point for respiratory disorders.

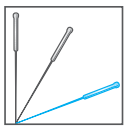
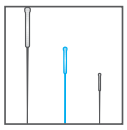
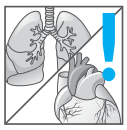

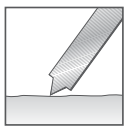
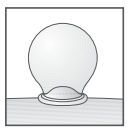

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 488

Ren-18 *Jade Hall* YUTANG

Location

On the anterior midline, on the level of the 3rd intercostal space.

How to find

Locate the 3rd intercostal space either by parasternal palpation starting at the clavicle or by counting downward from the manubriosternal synchondrosis (→ 2nd rib) (→ 3.5).
Located on the same level (→ 3rd intercostal space) are → **KID-24/ST-16/SP-19** (2/4/6 cun lateral to the midline).

Needling

0.5–1 cun transversely (subcutaneously) in a superior or inferior direction. Caution: Danger of injury to the heart or the pleura in patients with a sternal foramen, therefore only transverse needling with or against the flow of the channel.

Actions/Indications

- Unbinds the chest, regulates and descends the Qi

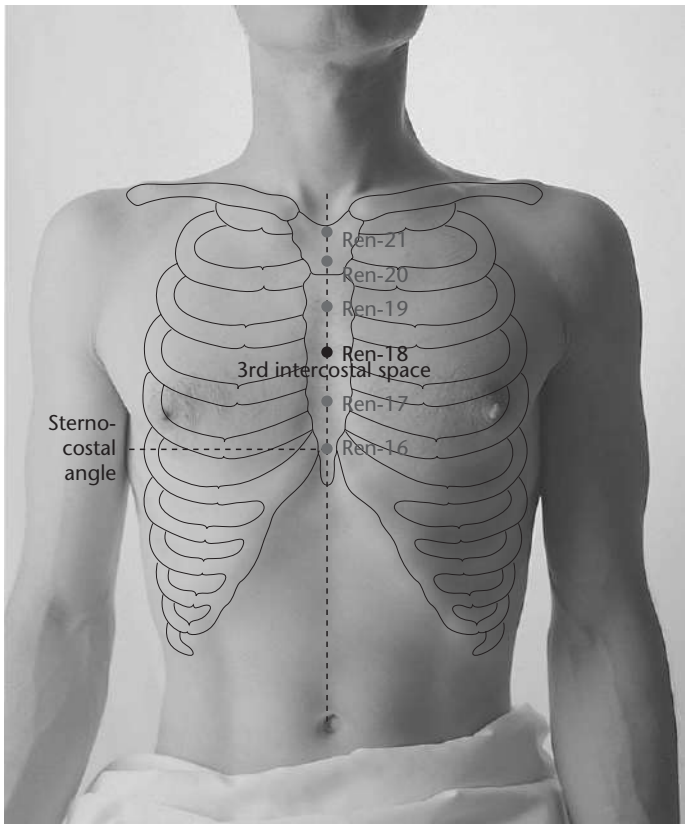

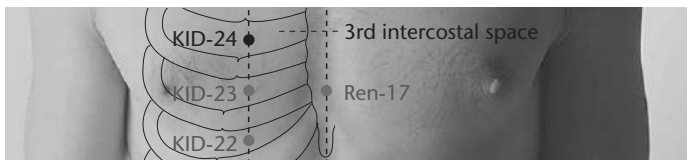

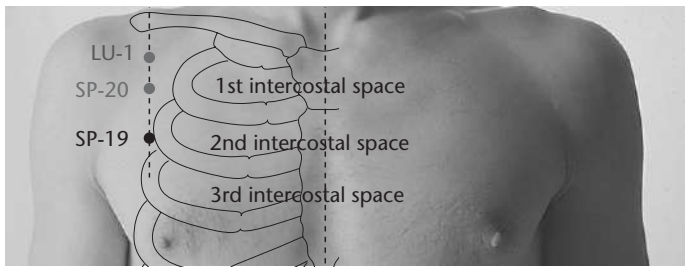

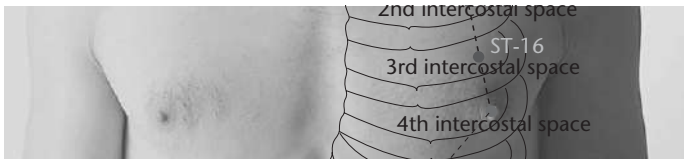

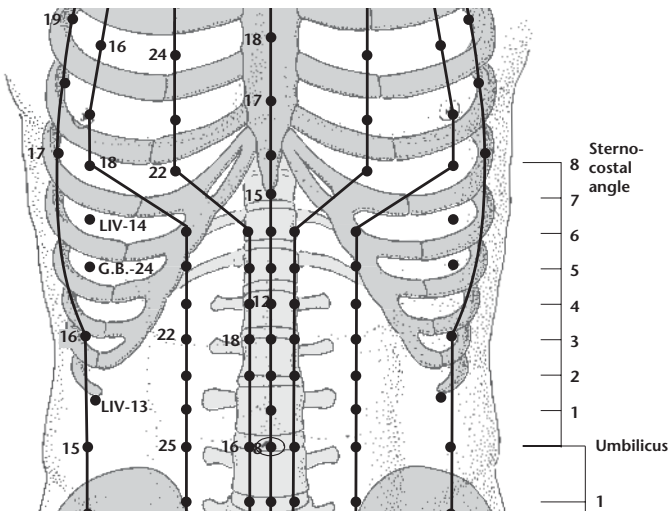

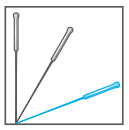

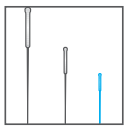

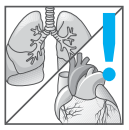

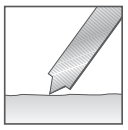

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 489

Purple Palace ZIGONG Ren-19

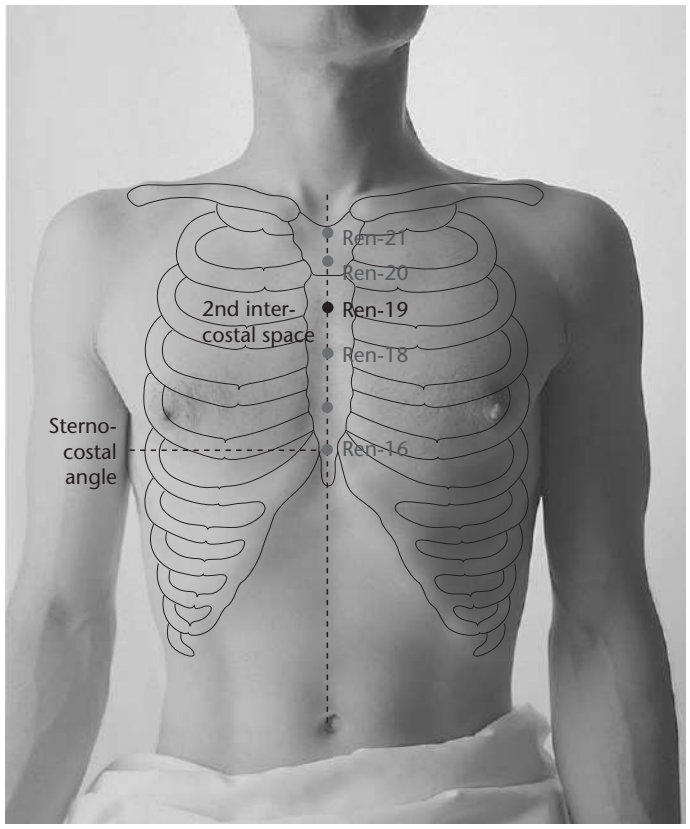

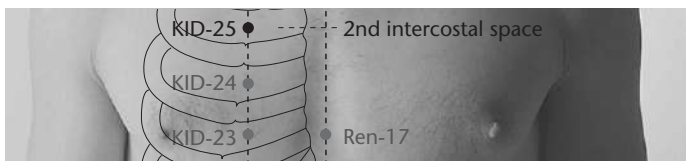

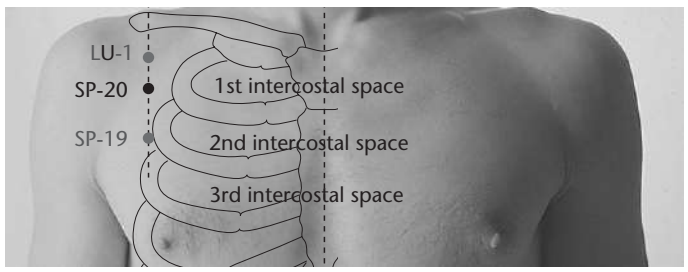

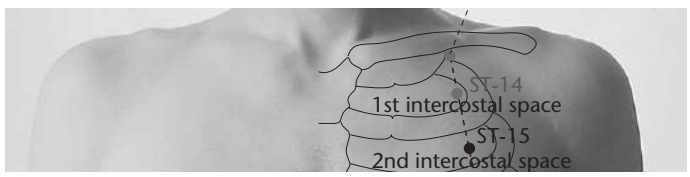

Location

On the anterior midline, on the level of the 2nd intercostal space.

How to find

Locate the 2nd intercostal space by parasternal palpation starting at the clavicle or by palpating from the manubriosternal synchondrosis (→ 2nd rib) (→ 3.5).

Located at the same level (2nd intercostal space) are → **KID-25/ST-15/SP-20** (2/4/6 cun lateral to the midline).

Needling

0.5 cun transversely in a superior or inferior direction. Caution: In patients with a sternal foramen, danger of injury to the pleura or the mediastinum, therefore only transverse needling with or against the flow of the channel.

Actions/Indications

- Unbinds the chest
- Regulates and descends the Qi

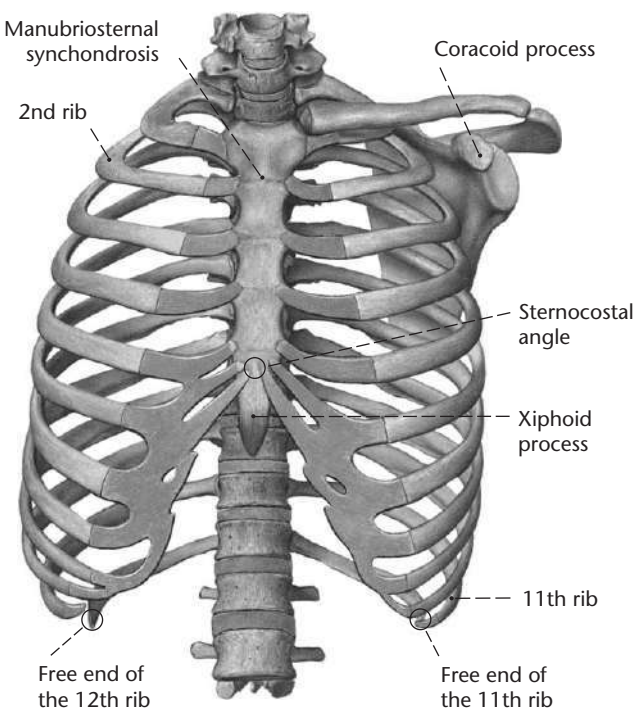

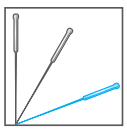
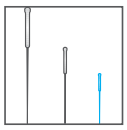
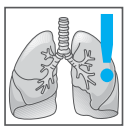

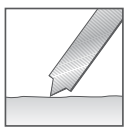
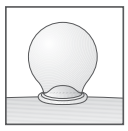

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 490

Ren-20 *Magnificent Canopy* HUAGAI

Location

On the anterior midline, on the lower part of the manubrium sterni, on the level of the 1st intercostal space.

How to find

The manubriosternal synchondrosis generally forms a distinct horizontal bony structure on the upper part of the sternum. Locate **Ren-20** slightly superior to the synchondrosis, on the midline and on the level of the 1st intercostal space.

Located at the same level (1st intercostal space) are → **KID-26/ST-14/LU-1** (2/4/6 cun lateral to the midline).

Needling

0.5 cun transversely in a superior or inferior direction. Caution: In patients with a sternal foramen, danger of injury to the pleura or the mediastinum, therefore only transverse needling with or against the flow of the channel.

Actions/Indications

- Unbinds the chest
- Regulates and descends the Qi

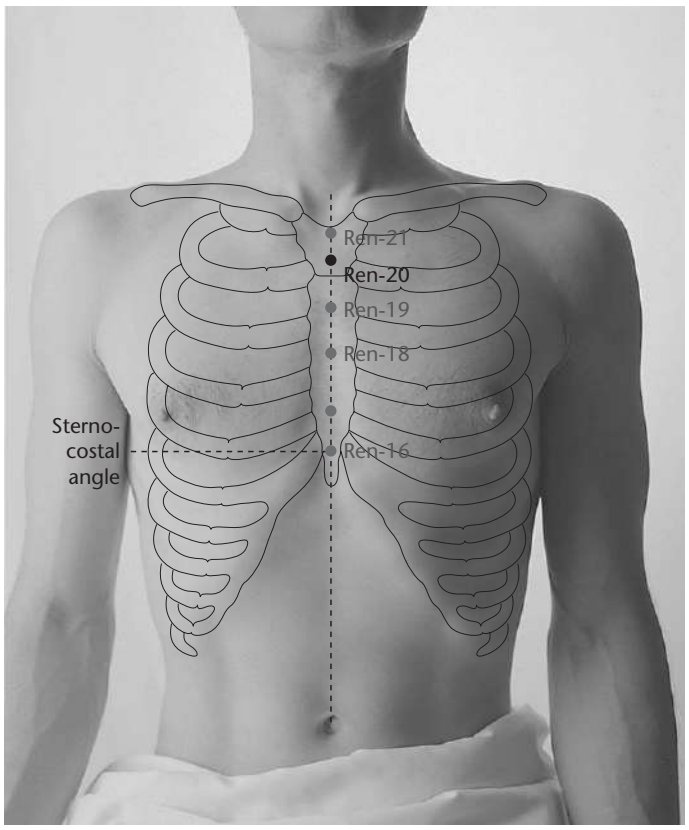

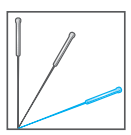

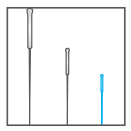

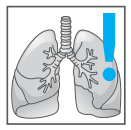

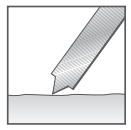

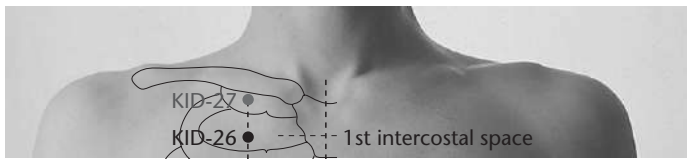

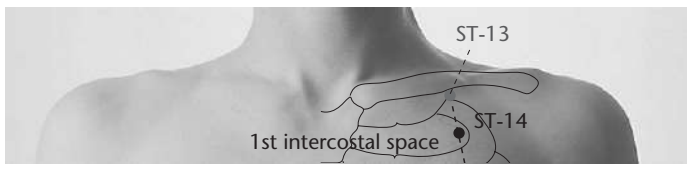

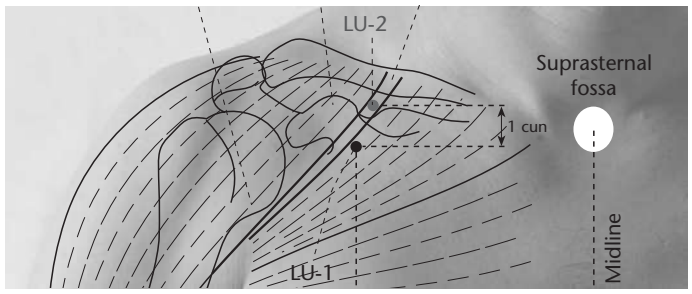

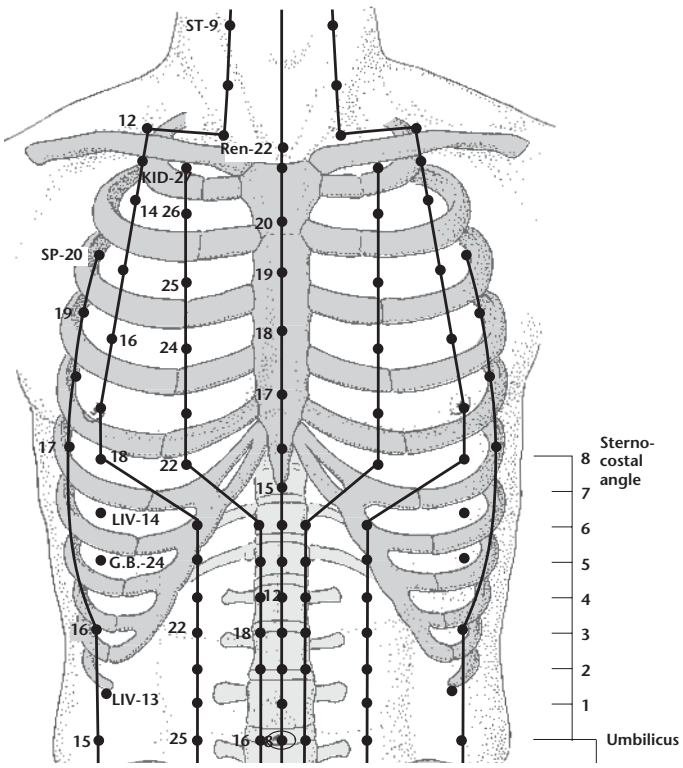

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 491

Jade Pivot XUANJI Ren-21

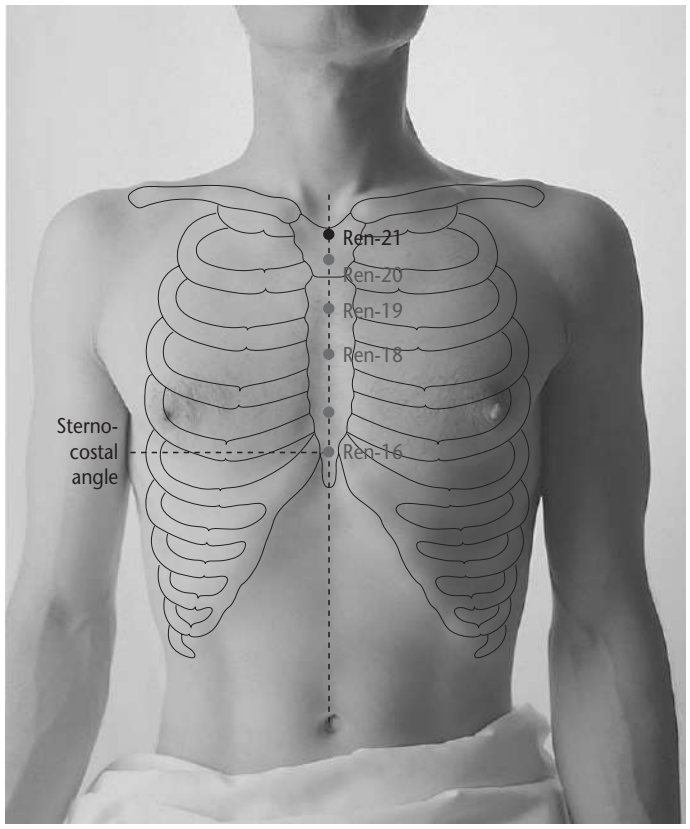

Location

On the anterior midline, below the upper border of the manubrium sterni.

How to find

The upper border of the manubrium sterni forms the lower border of the suprasternal fossa. Locate **Ren-21** on the midline, inferior to the bony margin of the sternum.
Located on approximately the same level (below the clavicle) are → **KID-27/ST-13/LU-2** (2/4/6 cun lateral to the midline).

Needling

0.5 cun transversely in a superior or inferior direction. Caution: In patients with a sternal foramen, danger of injury to the pleura or the mediastinum, therefore only transverse needling with or against the flow of the channel.

Actions/Indications

- Unbinds the chest
- Benefits the throat
- Descends counterflow Qi

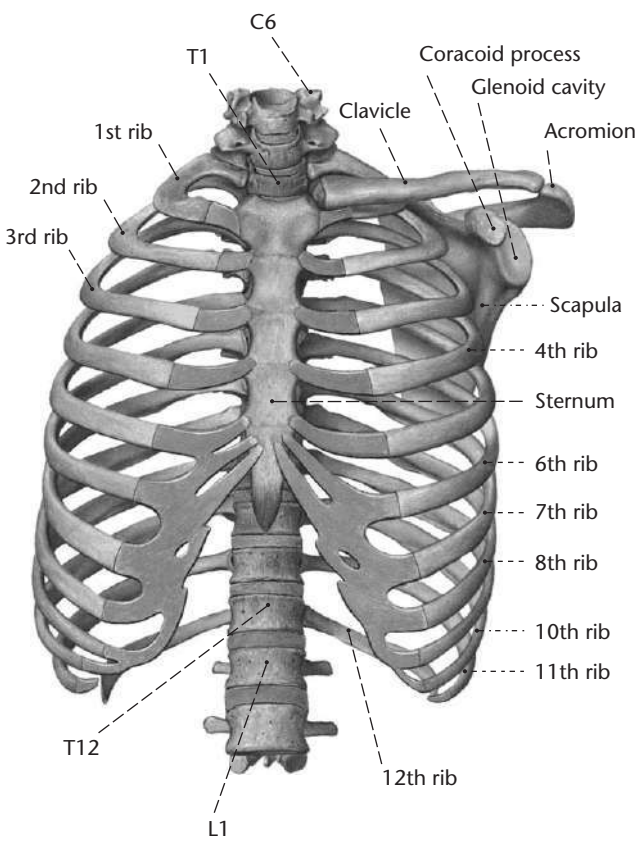

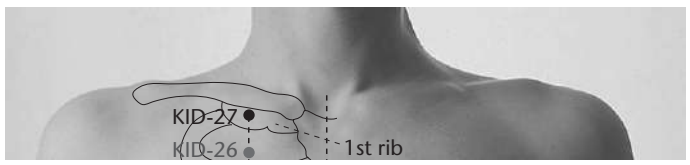

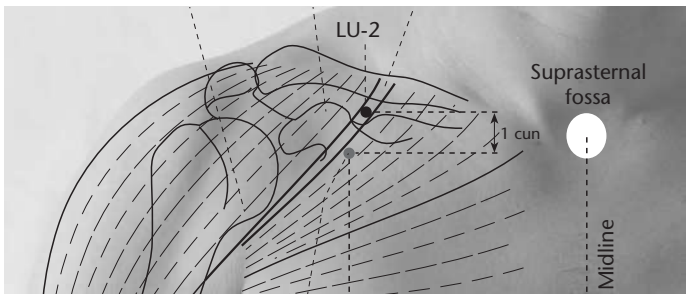

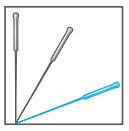
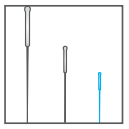
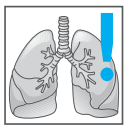

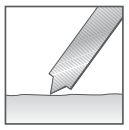

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 492

Ren-22 *Heavenly Prominence* TIANTU

Location

0.5 cun superior to the sternum, in the centre of the suprasternal fossa.

How to find

This point is best located with the patient in a supine position and a pillow under the shoulders or seated with a comfortable and safe head support. Locate the centre of the suprasternal fossa, approximately 0.5 cun superior to the sternum and there locate **Ren-22**.

Needling

Initial vertical insertion 0.2 cun (in order to penetrate the skin), then direct the tip of the needle inferiorly and continue to insert the needle parallel to the posterior aspect of the sternum up to 0.5–1 cun. Caution: Dangerous point. Wrong insertion technique (for example, vertical or oblique retrosternal needling) may lead to injury of the big vessels and organs inside the mediastinum. This point should only be needled by experienced practitioners.

Actions/Indications

- Descends counterflow Lung Qi, alleviates coughing and dyspnoea, benefits the throat

Special features

Meeting point with the *yin wei mai*, Window of Heaven point. Important emergency point for acute asthma attacks.

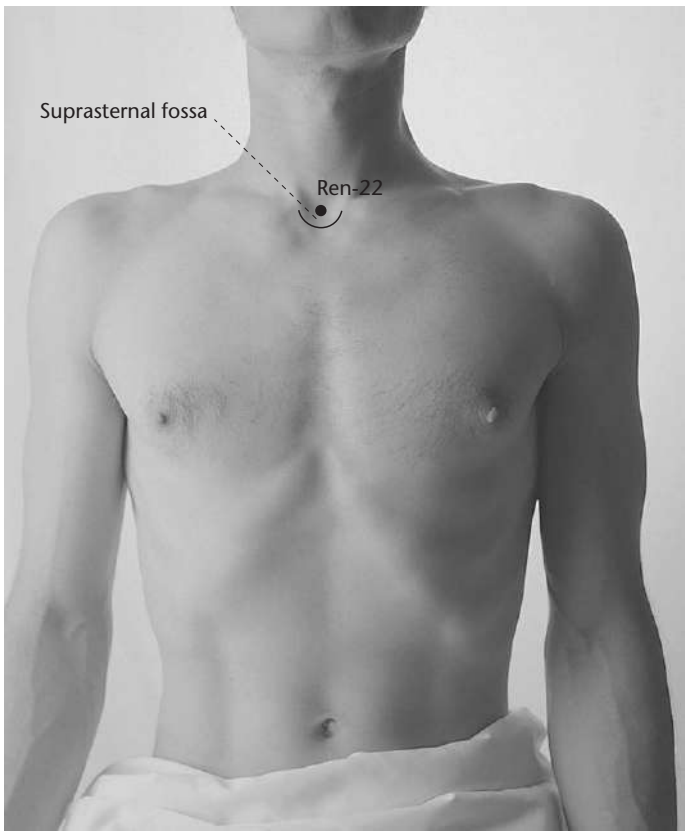

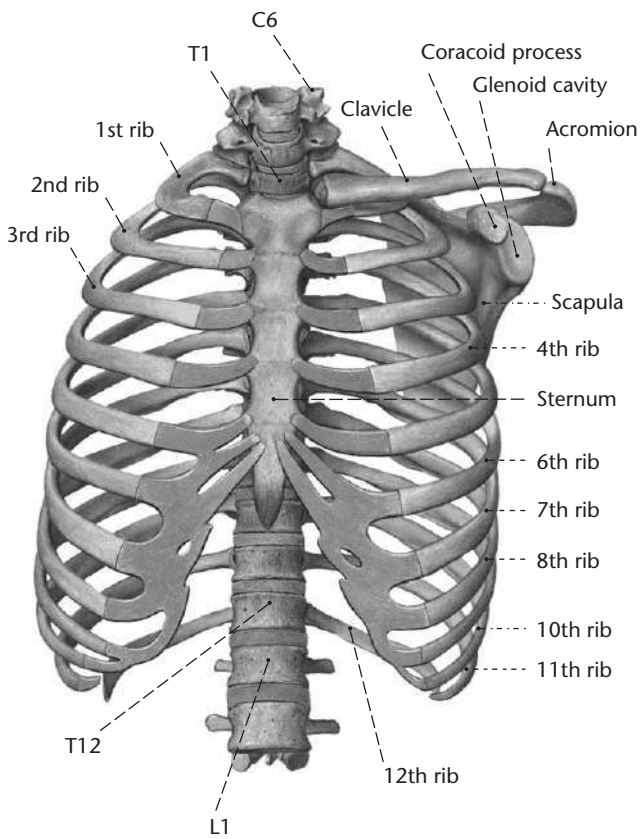

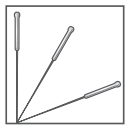
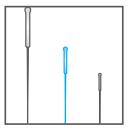

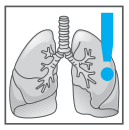
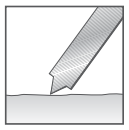

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 493

Corner Spring LIANQUAN Ren-23

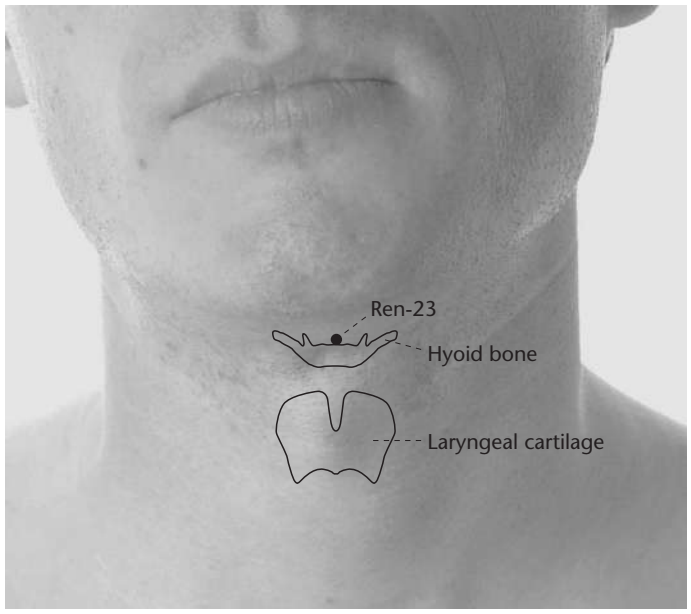

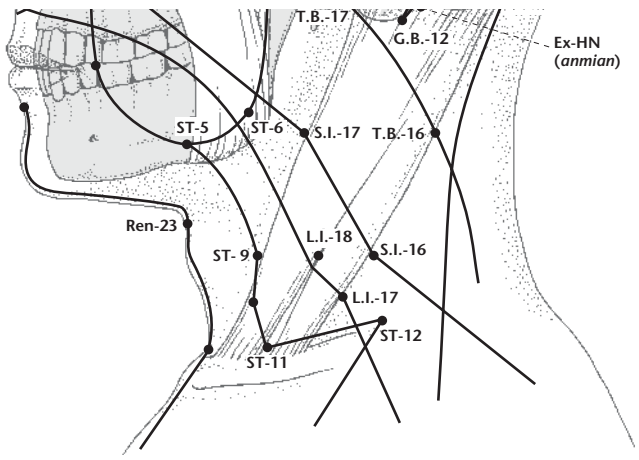

Location

On the anterior midline, superior to the upper border of the hyoid bone.

How to find

The hyoid bone can be palpated as a delicate bony structure slightly superior to the upper border of the laryngeal prominence. The patient should not recline their head too much, as this will cause the soft tissue to become stretched, making palpation of the bone difficult. **Ren-23** is located on the midline, on the upper border of the hyoid bone and at the junction of the vertical aspect of the neck to the horizontal floor of the mouth (with normal tissue distribution).

Needling

0.3–1 cun obliquely in a superior direction towards the vertex (→ **Du-20**). Moxibustion is contraindicated according to some texts.

Actions/Indications

- Benefits the tongue
- Benefits the throat, eliminates Phlegm and Heat
- Strengthens the voice

Special features

Meeting point with the *yin wei mai*

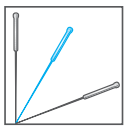
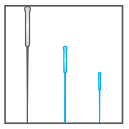

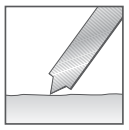

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 494

Ren-24 *Container of Fluids* CHENGJIANG

Location
On the anterior midline, below the lower lip, in the mentolabial groove.

How to find
On the chin, locate the mentolabial groove (→ 3.1.3), which forms a horizontal depression at the transition from the chin to the lower lip. Locate **Ren-24** in this groove, on the midline.

Needling
0.2–0.3 cun obliquely in a superior direction.

Actions/Indications
- Eliminates (external) Wind, opens the *luo*-connecting vessels, alleviates pain and swellings, regulates the *ren mai*
- As Sun Si Miao Ghost point for epilepsy and mania

Special features
Meeting point with the *du mai*, ST and L.I. channels, Sun Si Miao Ghost point. Important local point.

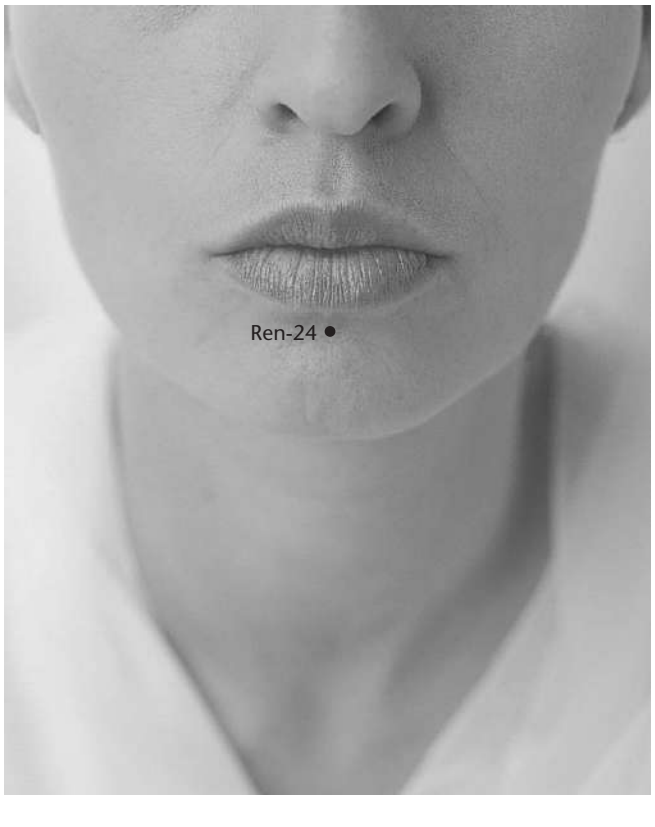

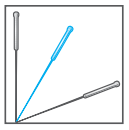
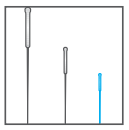

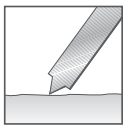

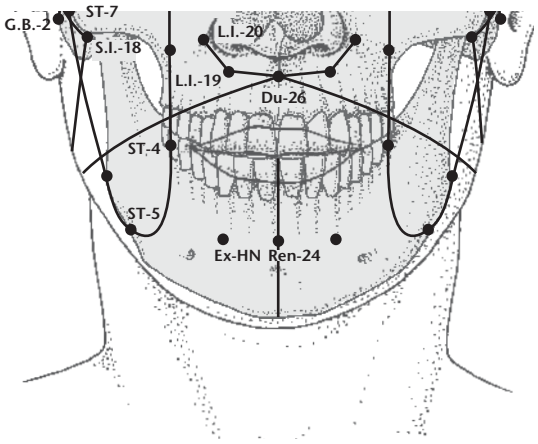

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 495

5.3 Du mai

Synonym: Governing Vessel

Relationships (→ 1.7.3)

- **Yin/Yang:** *ren mai/du mai*
- **Central/peripheral:** *du mai/yang qiao mai*. Regions supported by this pairing: inner canthus of the eye, occiput, shoulders, back, S.I. and BL channels, opening point: **S.I.-3** (*houxi*), coupled point: **BL-62** (*shenmai*)

Pathway

The extraordinary vessel *du mai* originates like the *chong mai* and *ren mai* in the lower abdomen (*bao zhong*), according to some authors at the Kidneys, emerges at the perineum and divides into several branches:

➡ The **main pathway** of the *du mai* begins at → **Ren-1** (*huiyin*) at the perineum, reaches **Du-1** (*changqiang*) between the anus and the coccyx and ascends the posterior midline along the spine. At **Du-12** (*shenzhu*), below the spinous process of T3, a branch separates on either side to reach → **BL-12** (*fengmen*), rejoining the main branch at **Du-13** (*taodao*) below the spinous process of T1. It then continues to **Du-16** (*fengfu*) on the lower border of the occipital bone. Here, an **internal branch** enters the Brain. The external pathway continues to ascend the head to **Du-20** (*baihui*), from there beginning its descent along the midline to the forehead and the nose to **Du-26** (*renzhong*) where it crosses the ST and L.I. channels. It terminates inside the mouth below the frenulum at **Du-28** (*yinjiao*) where it meets the ST channel and the extraordinary vessel *ren mai*.

➡ **one branch** of the *du mai* (the spinal branch) originates at → **Ren-1** (*huiyin*), courses to the tip of the coccyx where it meets the Kidney primary channel, ascends **within** the spinal canal to the space between L2 and L3 and enters the Kidneys.

➡ **The second branch** of the *du mai* (the abdominal branch) also originates at → **Ren-1** (*huiyin*), encircles the external genitalia and ascends to the umbilicus, passes through the Heart (*xin*), continues to the throat, circles the lips, travels to the cheek, passes → **ST-1** (*chengqi*) at the infraorbital ridge and reaches → **BL-1** (*jingming*).

➡ **The third branch** of the *du mai* travels bilaterally alongside the BL channel from → **BL-1** (*jingming*) to the vertex. From here, it branches out **internally** into the Brain. The **external** pathway follows the medial branches of the BL channel in an inferior direction and enters the body at the level of the lower border of the spinous process of L2 at → **BL-23** (*shenshu*) to terminate at the Kidneys (*shen*).

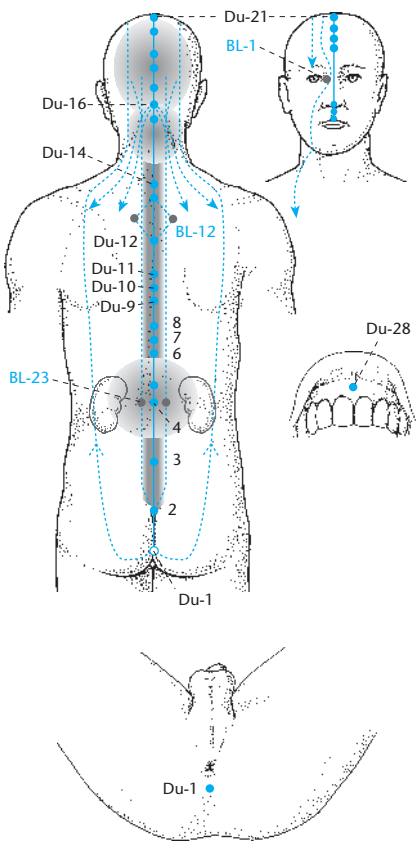

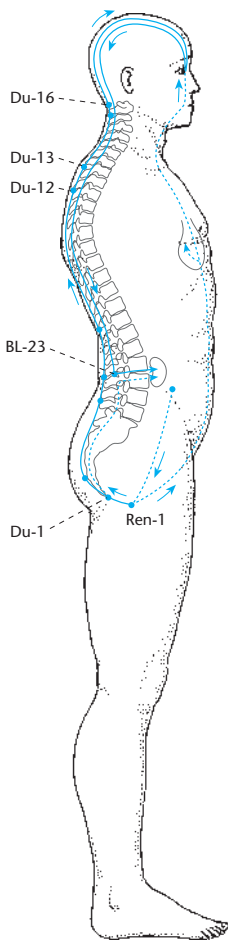

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 496

Meeting points of other channels with the *du mai*

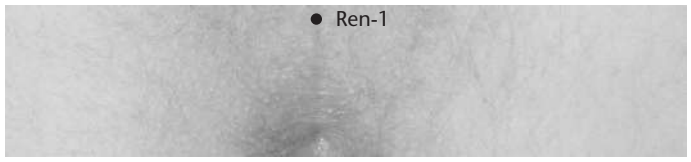

Ren-1 (*huiyin*): In the centre of the perineum

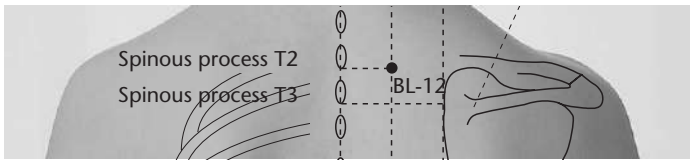

BL-12 (*fengmen*): 1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 2nd thoracic vertebra (T2)

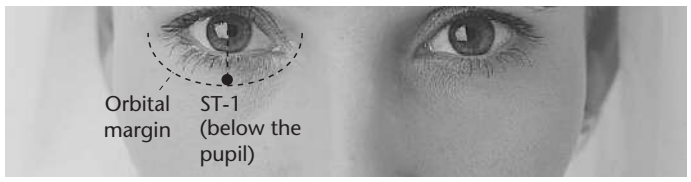

ST-1 (*chengqi*): With the eyes looking straight ahead, directly below the centre of the pupil, between the eyeball and the infraorbital ridge

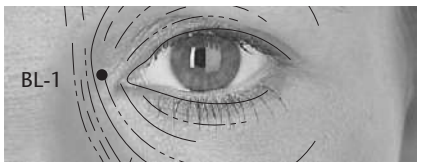

BL-1 (*jingming*): 0.1 cun superior and medial to the inner canthus of the eye, in a depression

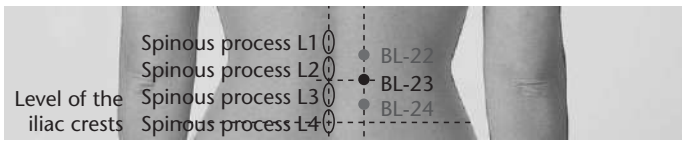

BL-23 (*shenshu*): 1.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 2nd lumbar vertebra (L2)

Connections with other channels/Organs

- *Ren mai*, *chong mai*, Bladder and Kidney primary channels
- Uterus, Kidneys, Brain

Clinical importance (→ 1.7.2, 1.7.3)

- **Sea of the Yang Qi**, supplies Yang Qi in order to warm the Organs and channels
- Supports and regulates all Yang channels, especially those of the upper back and head
- Influences and strengthens the functions of the Brain, Marrow and sensory organs
- Strengthens the body constitutionally and supports the immune system (circulation of the Defensive Qi (*wei qi*))
- Stores the Essence Qi (*jing qi*) overflowing from the Yang channels
- Absorbs excess from the Yang channels, especially excess Heat and internal Wind

Notes about the *du mai*

Both the *du mai* and *ren mai* have a unique position among the eight extraordinary vessels. They are the only two extraordinary vessels with their own specific points – all the other extraordinary vessels can only be accessed through points on the primary channels. For this reason the *ren mai* and the *du mai*, together with the primary channels, are often categorised as 14 channels. However, from the point of view of channel energetics this is not correct, since the extraordinary vessels occupy a special position, both evolutionarily and functionally (→ 1.7).

The pathway of the *du mai*, like that of the *ren mai*, is limited to the torso and the head. While it does not course along the extremities and does not have any points on the limbs (neither general points nor specific points such as the five *shu*-transporting points, *yuan*-source points, etc.) it does have a pertaining opening point (or master point) (→ 8.1.8).

The *du mai* is also referred to as the Sea of Yang and governs the Yang in the whole body. Located along its pathway are the posterior aspects of the energy centres (or chakras according to Yogi tradition) while their anterior aspects are located along the pathway of the *ren mai*. In Daoist tradition the *du mai* plays an important role in the cultivation of Qi. This concerns the development of a 'spirit child', which can be influenced by opening and activating energy centres located along the spine.

Furthermore, the *du mai* contains many meeting points with other channels. This clearly denotes the outstanding importance of the *du mai*.

Specific points according to their function

- ***Yuan*-source point (→ 8.1.1): –**
- ***Luo*-Connecting point (→ 8.1.2.): Du-1** (*changqiang*)
- ***Xi*-cleft point (→ 8.1.3): –**
- **Back-*shu* point (→ 8.1.4): –**
- **Front-*mu* point (→ 8.1.5): –**
- **Five *shu*-transporting points (→ 8.1.6): –**
- ***Hui*-meeting point (→ 8.1.7): –**
- **Pertaining opening point (→ 8.1.8): S.I.-3** (*houxi*)
- **Pertaining coupled point (→ 8.1.8): BL-62** (*shenmai*)
- **Lower *he*-Sea point (→ 8.1.9): –**
- ***Jiaohui*-meeting points (→ 8.1.10):**

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 497

– with the G.B. and KID channels and the *ren mai*: **Du-1** (*changqiang*)
– with the BL channel: **Du-13** (*taodao*)
– with all Yang channels: **Du-14** (*dazhui*) ■■
– with the *yang wei mai*: **Du-15** (*yamen*) ■■
– with the *yang wei mai* and the *yang qiao mai**: **Du-16** (*fengfu*) ■■
– with the BL channel: **Du-17** (*naohu*)
– with the BL, G.B., T.B. and LIV channels: **Du-20** (*baihui*) ■■
– with the BL and ST channels: **Du-24** (*shenting*) ■■
– with the L.I. and ST channels: **Du-26** (*renzhong*) ■■
– with the *ren mai* and ST channel*: **Du-28** (*yinjiao*)
– of other channels with the *du mai*: **Ren-1, BL-12**, **ST-1, BL-1, BL-23**

- **Gao Wu command point (→ 8.1.11):** –
- **Window of Heaven point (→ 8.1.12): Du-16** (*fengfu*) ■■
- **Points of the Four Seas (→ 8.1.13):**
  – Sea of Qi: **Du-14** (*dazhui*) ■■, **Du-15** (*yamen*) ■■
  – Sea of Marrow: **Du-16** (*fengfu*) ■■; **Du-20** (*baihui*) ■■
- **Ma Dan Yang Heavenly Star point (→ 8.1.14):**
- **Sun Si Miao Ghost point (→ 8.1.15): Du-16** (*fengfu*) ■■, **Du-26** (*renzhong*) ■■
- **Other functional points:**
  – very important point for strengthening the Yang (with moxibustion): **Du-4** (*mingmen*) ■■
  – important emergency point: **Du-26** (*renzhong*) ■■

Points according to region

- **Local points (→ 8.2.1):** Temporal region – **Du-20** (*baihui*) ■■; nose – **Du-25** (*suliao*); rectum – **Du-1** (*changqiang*); lumbar spine – **Du-3** (*yaoyangguan*), **Du-4** (*mingmen*) ■■
- **Adjacent points (→ 8.2.1):** occipital region – **Du-14** (*dazhui*) ■■; temporal region, eyes – **Du-23** (*shangxing*); nose – **Du-23** (*shangxing*), **Du-24** (*shenting*); Lungs – **Du-14** (*dazhui*) ■■; Heart – **Du-11** (*shendao*); cervical spine – **Du-14** (*dazhui*) ■■
- **Distal points (→ 8.2.1):** tongue – **Du-15** (*yamen*) ■■; rectum – **Du-20** (*baihui*) ■■; lumbar spine – **Du-12** (*shenzhu*), – **Du-26** (*renzhong*) ■■

Specific points according to the vessel pathway (in numerical order)

- **Du-1** (*changqiang*): *luo*-connecting point (→ 8.1.2); *jiaohui*-meeting point with the G.B., KID channels and the *ren mai* (→ 8.1.10); local point for the rectum (→ 8.2.1)
- **Du-3** (*yaoyangguan*): local point for the lumbar spine (→ 8.2.1)
- **Du-4** (*mingmen*) ■■: very important point for strengthening the Yang (with moxibustion)
- **Du-11** (*shendao*): adjacent point for the Heart (→ 8.2.1)
- **Du-12** (*shenzhu*): distal point for the lumbar spine (→ 8.2.1)
- **Du-13** (taodao): *jiaohui*-meeting point with the BL channel (→ 8.1.10)
- **Du-14** (*dazhui*) ■■: *jiaohui*-meeting point with all Yang channels (→ 8.1.10); point of the Sea of Qi (→ 8.1.13); adjacent point for the occiput and Lung (→ 8.2.1); adjacent point for the cervical spine (→ 8.2.1)
- **Du-15** (*yamen*) ■■: *jiaohui*-meeting point with the *yang wei mai* (→ 8.1.10); Sea of Qi point (→ 8.1.13); distal point for the tongue (→ 8.2.1)
- **Du-16** (*fengfu*) ■■: point of the Sea of Marrow (→ 8.1.13); Window of Heaven point (→ 8.1.12); Sun Si Miao Ghost point (→ 8.1.15); *jiaohui*-meeting point with the *yang wei mai* and the *yang qiao mai** (→ 8.1.10)
- **Du-17** (*naohu*): *jiaohui*-meeting point with the BL channel (→ 8.1.10)
- **Du-20** (*baihui*) ■■ *jiaohui*-meeting point with the BL, G.B., T.B. and LIV channels (→ 8.1.10); point of the Sea of Marrow (→ 8.1.13); local point for the temporal aspect of the head (→ 4.2.1); distal point for the rectum (→ 8.2.1)
- **Du-23** (*shangxing*): regional point for the temporal region, the eyes and the nose (→ 8.2.1)
- **Du-24** (*shenting*) ■■: *jiaohui*-meeting point with the BL and ST channels (→ 4.1.10); adjacent point for the nose (→ 8.2.1)
- **Du-25** (*suliao*): local point for the nose (→ 4.2.1)
- **Du-26** (*renzhong*) ■■: *jiaohui*-meeting point with the L.I. and ST channels (→ 8.1.10); Sun Si Miao Ghost point (→ 8.1.15); distal point for the lumbar spine (→ 8.2.1); important emergency point
- **Du-28** (*yinjiao*): *jiaohui*-meeting point with the *ren mai* and ST channel (→ 8.1.10)

* Mentioned by only some authors

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 498

luo-Connecting Vessel of the *Du Mai*

Pathway

The *luo*-connecting vessel of the *du mai* originates at **Du-1** (*changqiang*) between the anus and the tip of the coccyx, bilaterally ascends the spine to the neck and disperses in the occipital region, where it connects with the Bladder primary channel. Both course together to the shoulder region. There the *luo*-connecting vessel enters deeper into the body at the level of the scapula and spreads in the paravertebral musculature.

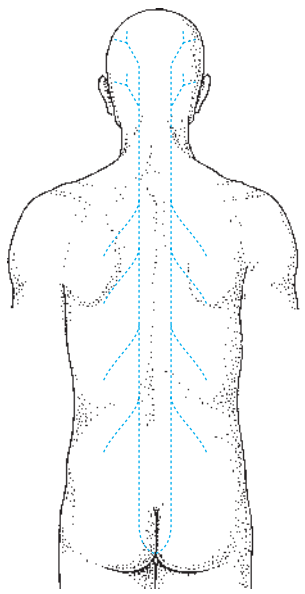

Clinical importance

Pathology

Excess (*shi*): limited range of motion and stiffness of the spine
Deficiency (*xu*): heavy sensations of the head, dizziness, balance disorders

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 499

Long Strong CHANGQIANG Du-1

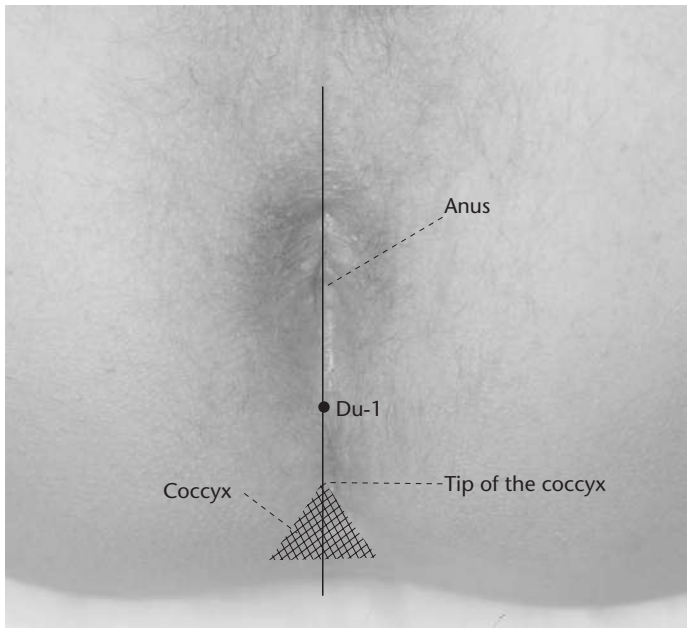

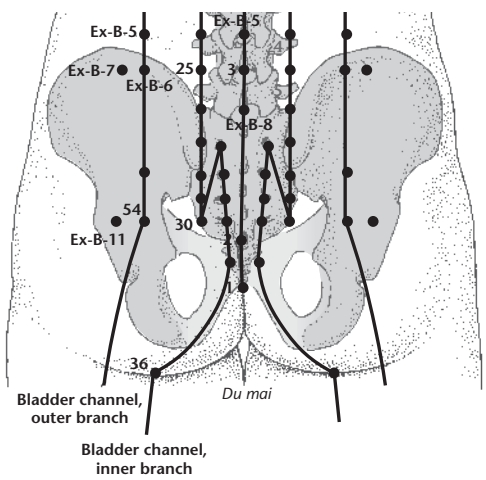

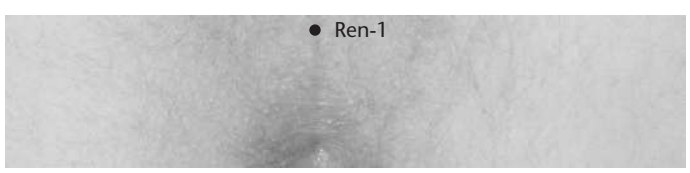

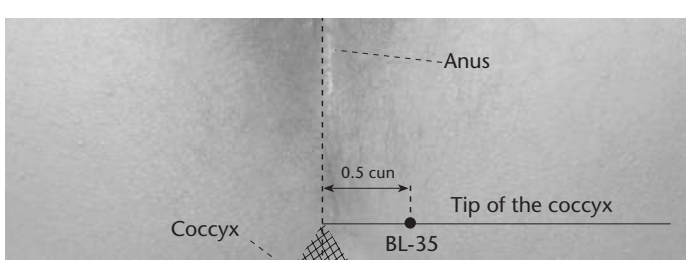

Location

On the midline, midway between the tip of the coccyx and the anus.

How to find

First, locate the coccyx superior to the anus, then palpate for its tip. From the tip of the coccyx, palpate with pressure towards the anus. **Du-1** can often be located based on a strong *de* Qi sensation.

→ **Ren-1** is located anterior to the anus, posterior to the genitals.

→ **BL-35** is located 0.5 cun lateral to the midline, at the level of the tip of the coccyx.

Needling

After thorough disinfection, vertically 0.5–1 cun. Avoid puncturing the rectum.

Actions/Indications

- Clears Damp-Heat, lifts the Qi
- Opens the channel and alleviates pain
- Calms the *shen*

Special features

Luo-connecting point, meeting point with the *ren mai*, G.B. and KID channels.

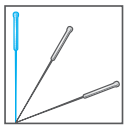
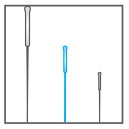
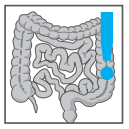

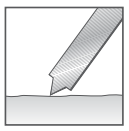
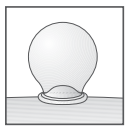

Ch05a-F10028.qxd 2/22/08 9:44 PM Page 500

Du-2 *Lumbar Shu* YAOSHU

Location
On the midline, in the sacral hiatus.

How to find
Palpate for the sacral hiatus (→ 3.4.4) from the sacrum. It can be felt at the inferior end of the iliac median crest as a U-shaped depression opening towards the inferior. **Du-2** is located on the midline, directly below the hiatus.
→ **BL-35** is located slightly more inferiorly, 0.5 cun lateral to the midline, on the level of the tip of the coccyx.

Needling
Up to 1 cun vertically or obliquely in a superior direction.

Actions/Indications
- Strengthens the lumbar region and the legs
- Dispels Wind-Dampness
- Warms the Lower Burner

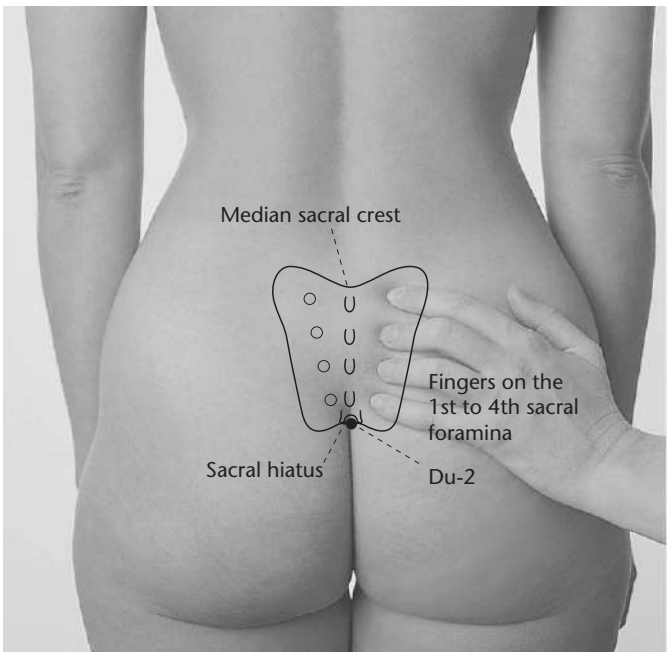

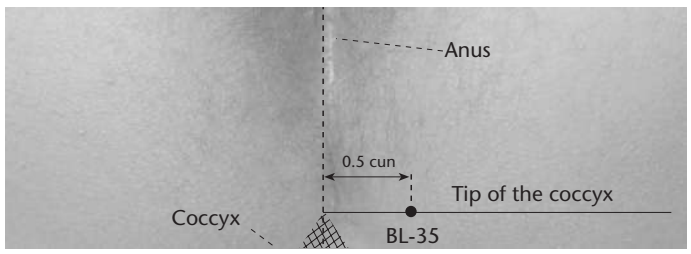

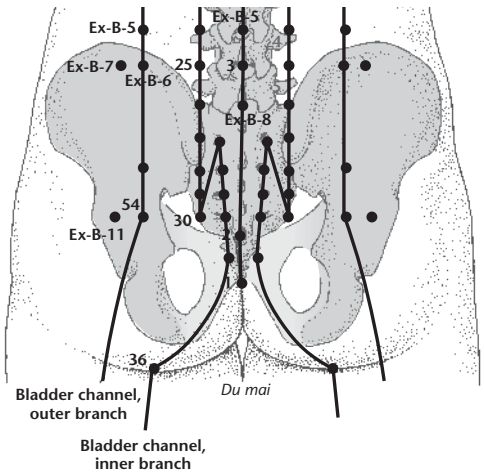

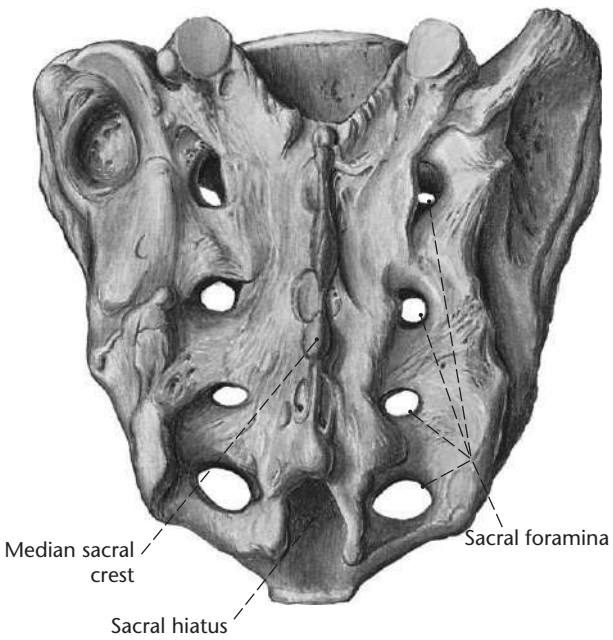

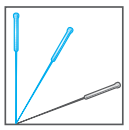
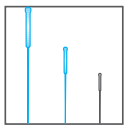

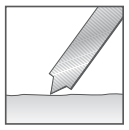

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 501

Lumbar Yang Gate YAOYANGGUAN Du-3

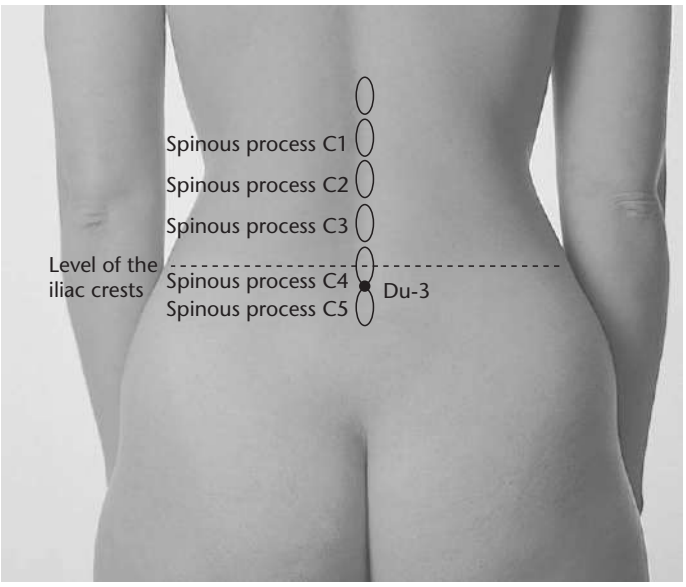

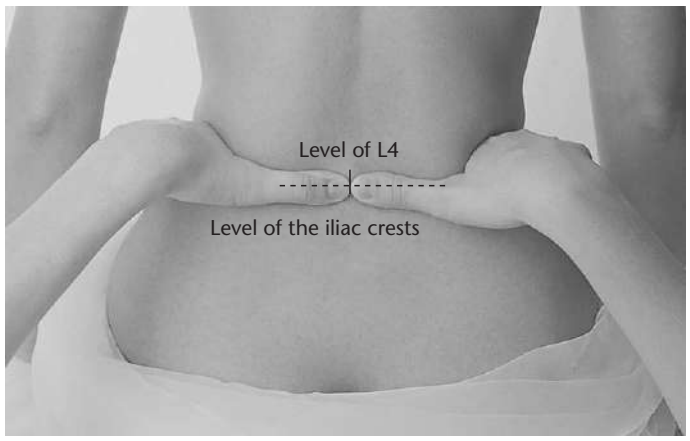

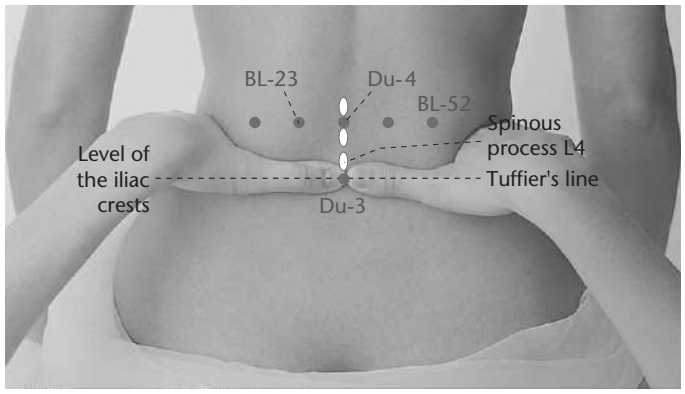

Location

On the midline, below the spinous process of the 4th lumbar vertebra (L4).

How to find

For orientation in the lumbar region (→ 3.4.3), first, locate the Tuffier's line (a line connecting the two most superior points on the iliac crest), which generally intersects with the spinous process of L4. (Note: Point location in the lumbar region is dependent on the patient's position; for more detail on orientation → 3.4). Locate **Du-3** on the midline, below the spinous process of L4.
Located on the same level are a point of → **Ex-B-2/BL-25/Ex-B-6/Ex-B-7** (0.5/1.5/3/3.5 cun lateral to the midline).

Needling

0.5–1 cun strictly vertically to the skin or obliquely in an inferior direction (generally, there is only danger of accidental puncture of the spinal canal from the level of L2 in a cranial direction).

Actions/Indications

- Dispels Wind, Cold and Dampness
- Regulates Qi in the Lower Burner, strengthens the Original Qi (*yuan qi*)

Special features

Important local point for combined back and leg pain

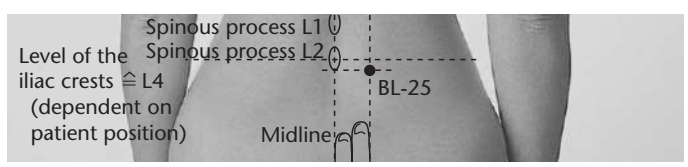

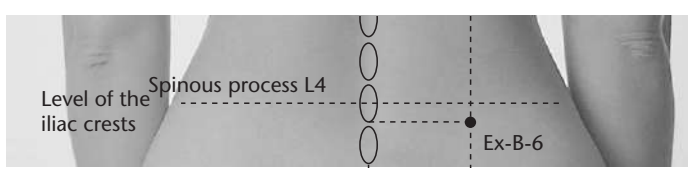

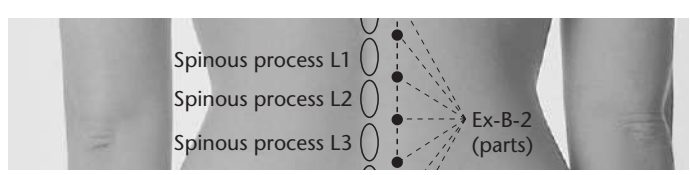

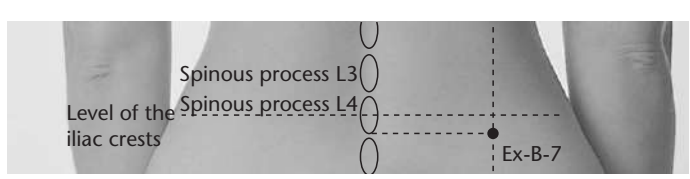

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 502

Du-4 *Gate of Life* MINGMEN

Location

On the posterior midline, below the spinous process of the 2nd lumbar vertebra (L2).

How to find

For orientation in the lumbar region (→ 3.4.3), first, locate the Tuffier's line by placing both hands directly superior and lateral to the two highest points on the iliac crest, joining the thumbs on the midline. The line will generally intersect with the spinous process of L4. (Note: Orientation on the lumbar region depends on the patient's position; for more detail on differential orientation → 3.4). From there, count up to the spinous process of L2 and locate **Du-4** on the midline, below the spinous process. Located on the same level are a point of → **Ex-B-2/BL-23/BL-52** (0.5/1.5/3 cun lateral to the midline).

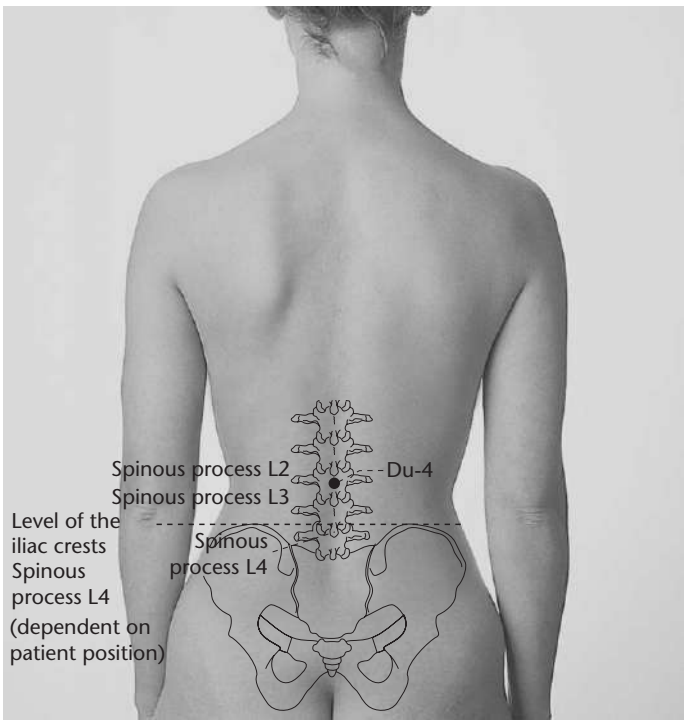

Needling

0.5–1 cun vertically to slightly obliquely in an inferior direction, with the patient's back straight or flexed. The spinal cord generally ends between the 1st and 2nd lumbar vertebrae; still, there is hardly any mention in the literature of spinal punctures with deep needling and/or insertions in a superior direction.

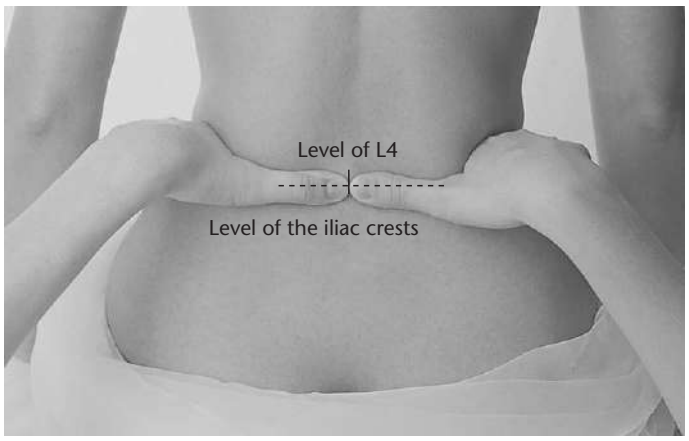

Actions/Indications

- Tonifies Kidney Yang (especially with moxibustion), warms the *mingmen*, strengthens the Kidneys, regulates the *du mai*, benefits the lumbar region
- Pacifies Wind in the *du mai*

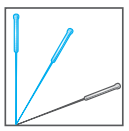

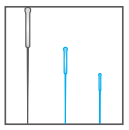

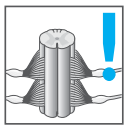

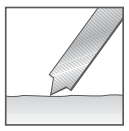

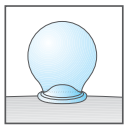

Special features

Major point for tonifying Yang, especially Kidney Yang

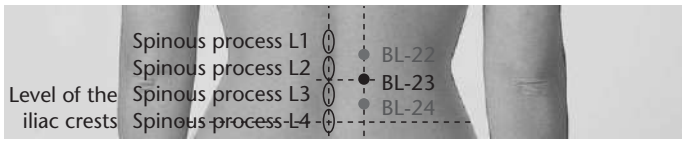

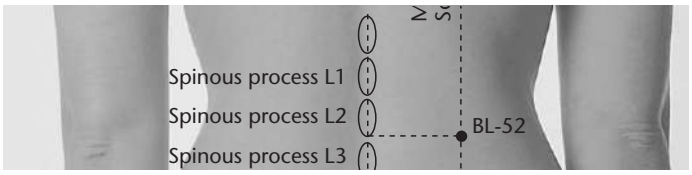

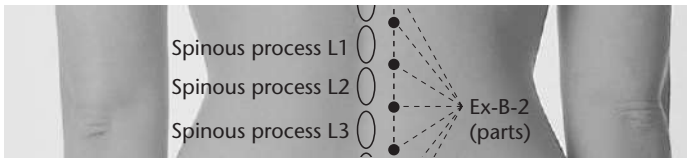

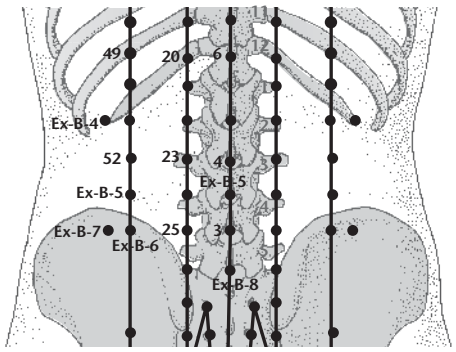

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 503

Suspended Pivot XUANSHU Du-5

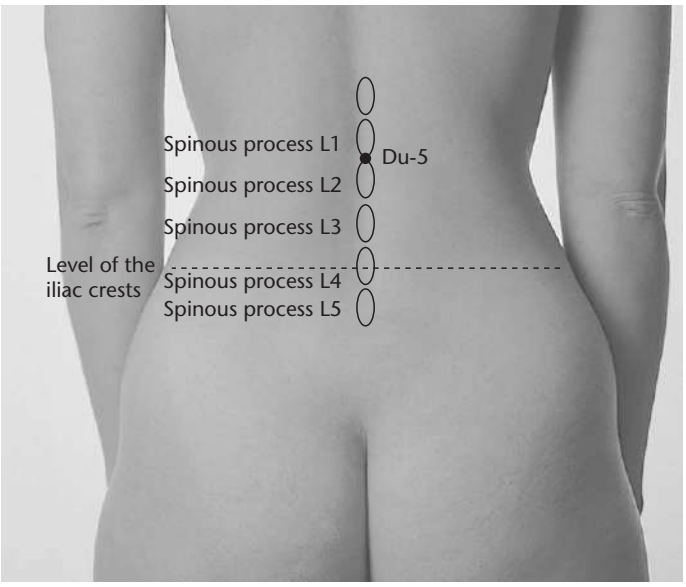

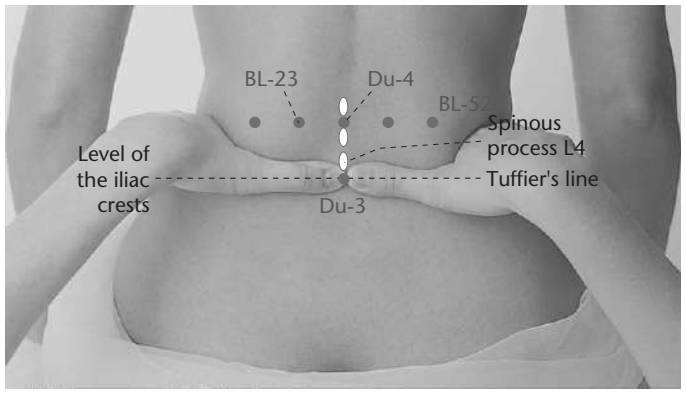

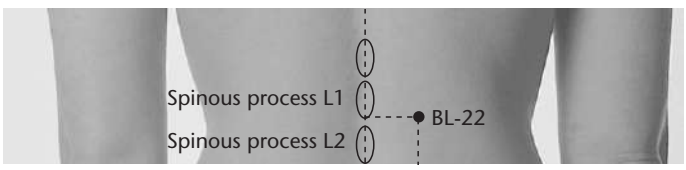

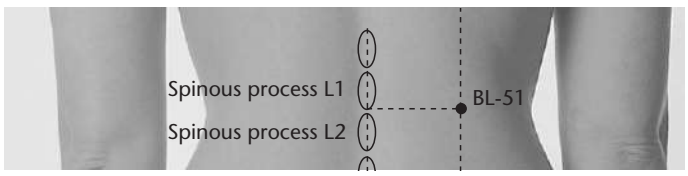

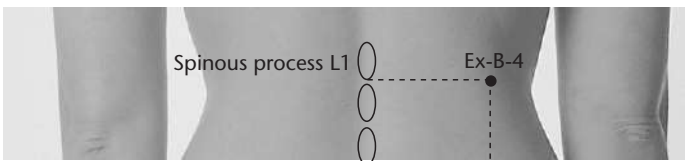

Location

On the midline, below the spinous process of the 1st lumbar vertebra (L1).

How to find

For orientation in the lumbar region (→ 3.4.3), first, locate the Tuffier's line by placing both hands directly superior and lateral to the two highest points on the iliac crest, joining the thumbs on the midline. The Tuffier's line generally intersects with the spinous process of L4. (Note: Anatomical orientation in the lumbar region is dependent on the patient's position; for more detail on differential orientation → 3.4.) From there, count up to the spinous process of L1 and locate **Du-5** on the midline, below its spinous process. **Or:** Use the spinous process of T7 (inferior angle of the scapula → 3.4.2) for reference.

Located on the same level are a point of → **Ex-B-2/BL-22/BL-51/Ex-B-4** (0.5/1.5/3/3.5 cun lateral to the midline).

Needling

0.5–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Opens the channel
- Regulates Qi in the Lower Burner

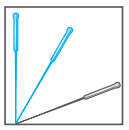
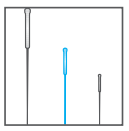
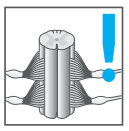

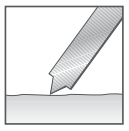

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 504

Du-6 *Centre of the Spine* JIZHONG

Location

On the midline, below the spinous process of the 11th thoracic vertebra (T11).

How to find

Locate T7 (→ 3.4.2) and, from there, count down 4 spinous processes to T11. Locate **Du-6** on the midline, below its spinous process. **Or:** Use the Tuffier's line for reference (→ 3.4.3). Locate T11 by counting up 5 spinous processes from L4 and there locate **Du-6** on the midline, below the spinous process. Located on the same level are a point of → **Ex-B-2/BL-20/BL-49** (0.5/1.5/3 cun lateral to the midline).

Needling

0.5–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Strengthens the Spleen and its function of transforming Dampness
- Benefits the spine

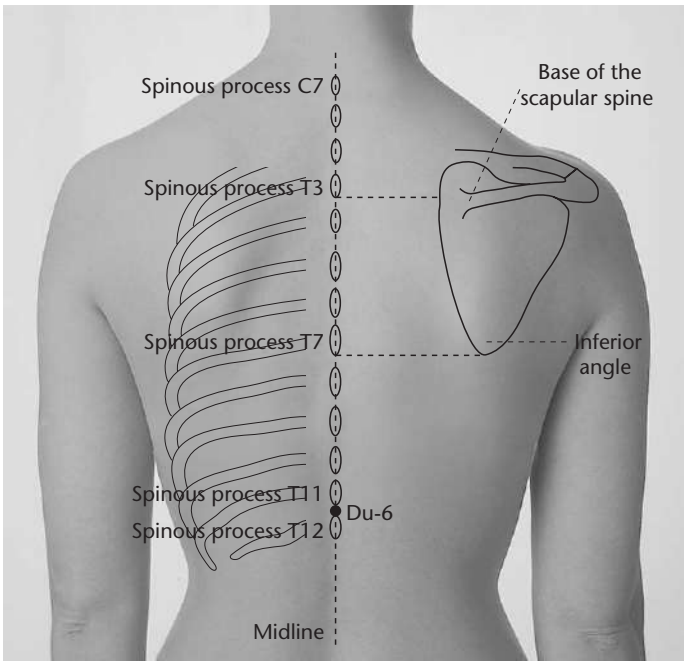

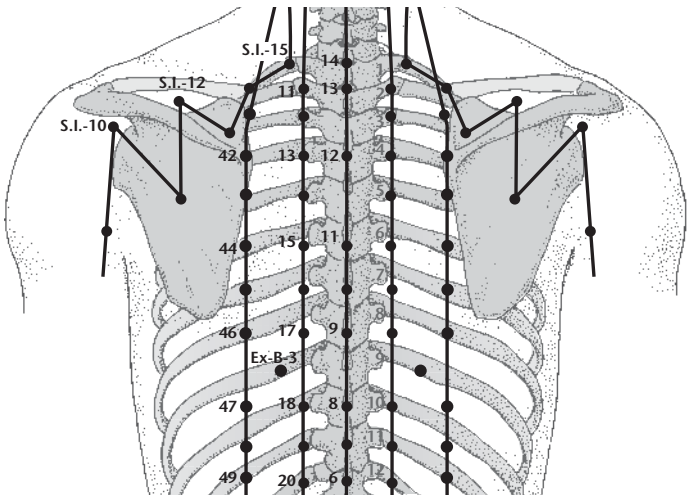

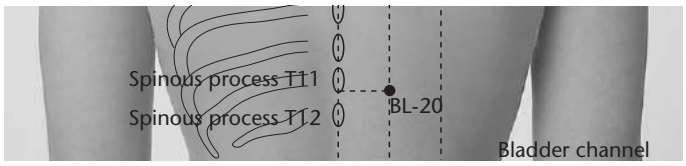

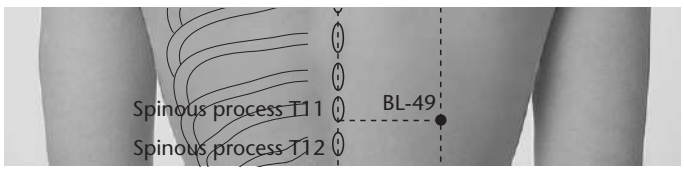

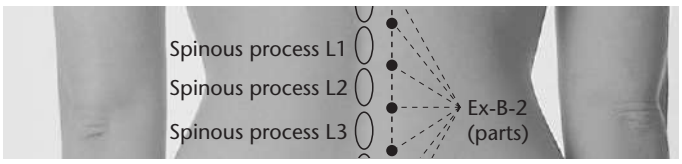

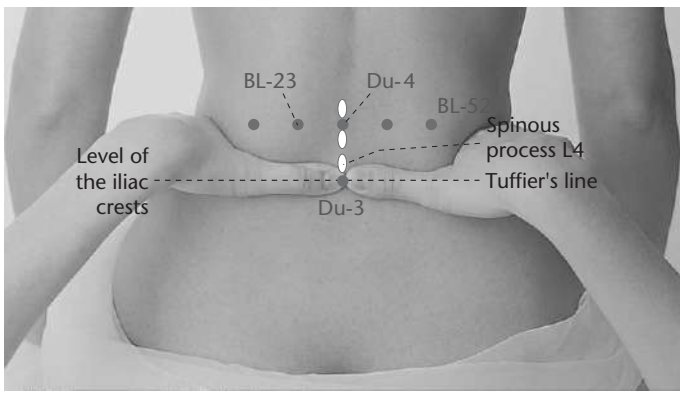

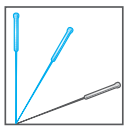
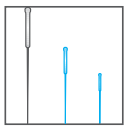
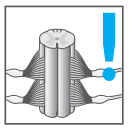

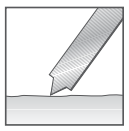
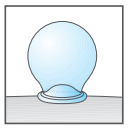

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 505

Central Pivot ZHONGSHU Du-7

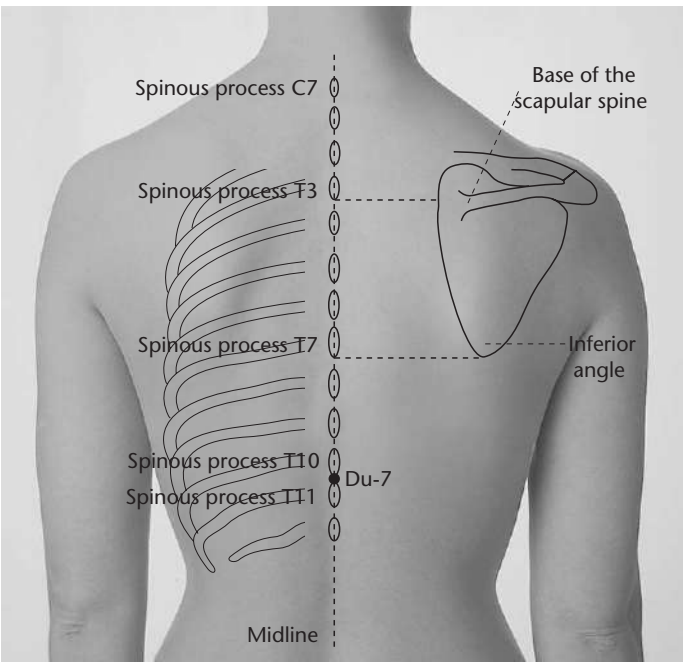

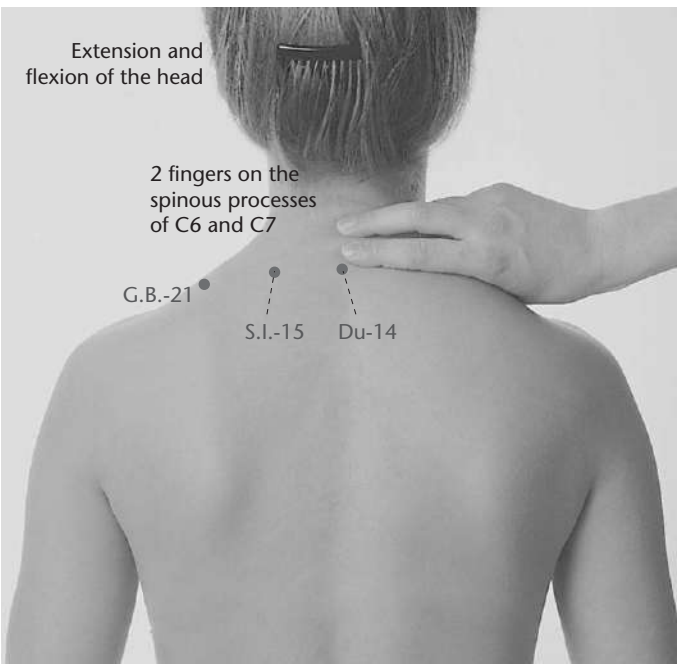

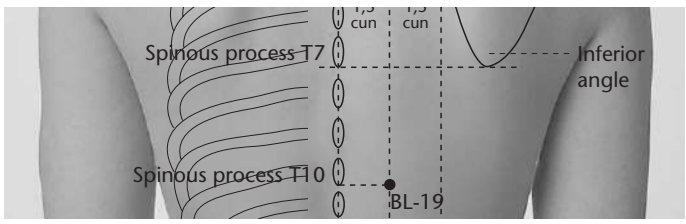

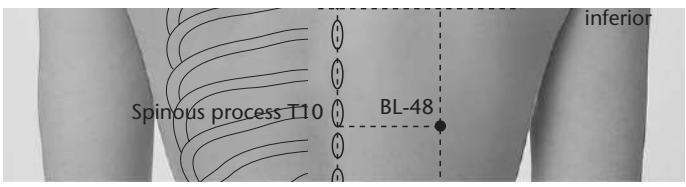

Location

On the midline, below the spinous process of the 10th thoracic vertebra (T10).

How to find

Locate T7 (→ 3.4.2) and, from there, count down 3 spinous processes to T10. Locate **Du-7** on the midline, below its spinous process. **Or:** Use the Tuffier's line for reference (→ 3.4.3). Locate T11 by counting up 5 spinous processes from L4 and there locate **Du-7** on the midline, below the spinous process. Located on the same level are a point of → **Ex-B-2/BL-19/ BL-48** (0.5/1.5/3 cun lateral to the midline).

Needling

0.5–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Tonifies the Middle Burner
- Benefits the spine

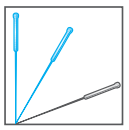
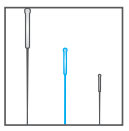
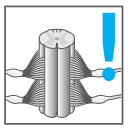

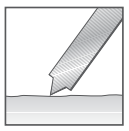

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 506

Du-8 *Sinew Contraction* JINSUO

Location

On the midline, below the spinous process of the 9th thoracic vertebra (T9).

How to find

Locate T7 (→ 3.4.2) and, from there, count down 2 spinous processes to T9. Locate **Du-8** on the midline, below its spinous process. **Or:** Use the Tuffier's line for reference (→ 3.4.3). Locate T9 by counting up from L4 and there locate **Du-8** on the midline, below the spinous process.
Located on the same level are a point of → **Ex-B-2/BL-18/BL-47** (0.5/1.5/3 cun lateral to the midline).

Needling

0.5–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Soothes the Liver
- Dispels Wind
- Calms the *shen*

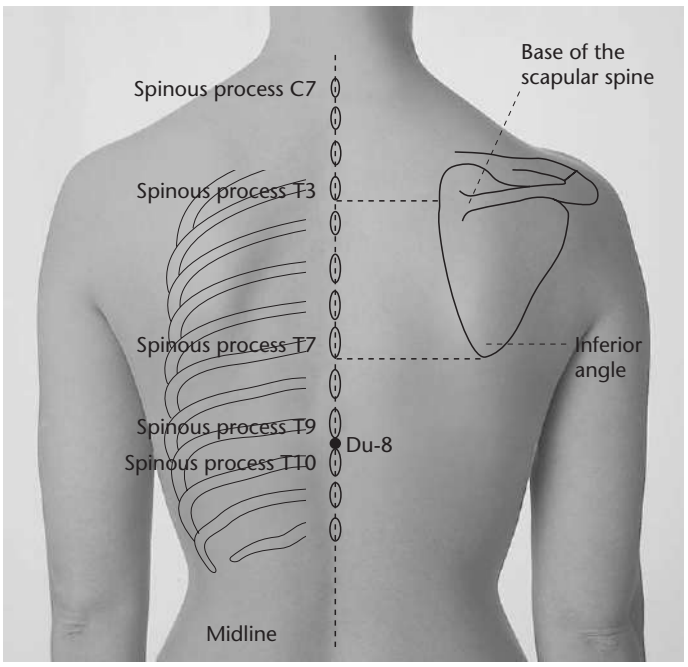

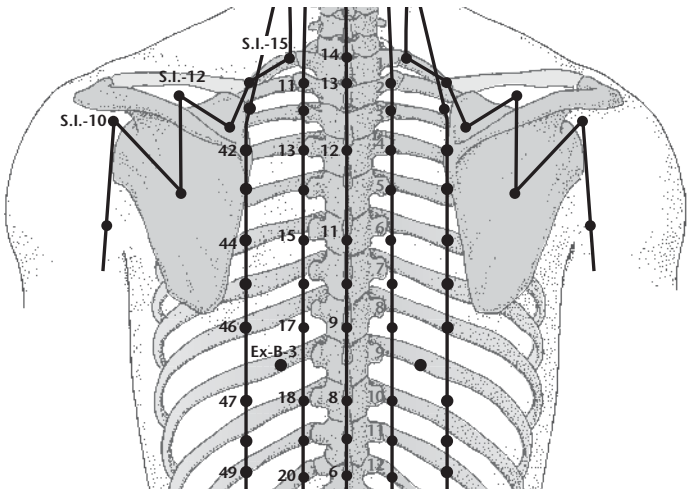

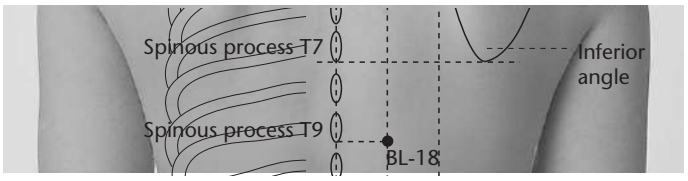

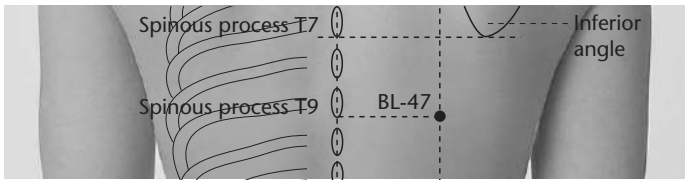

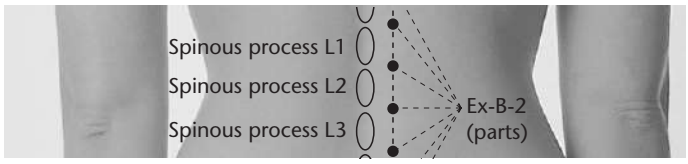

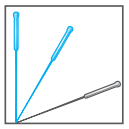

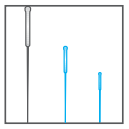

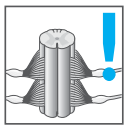

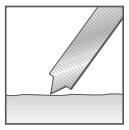

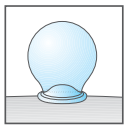

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 507

Reaching Yang ZHIYANG Du-9

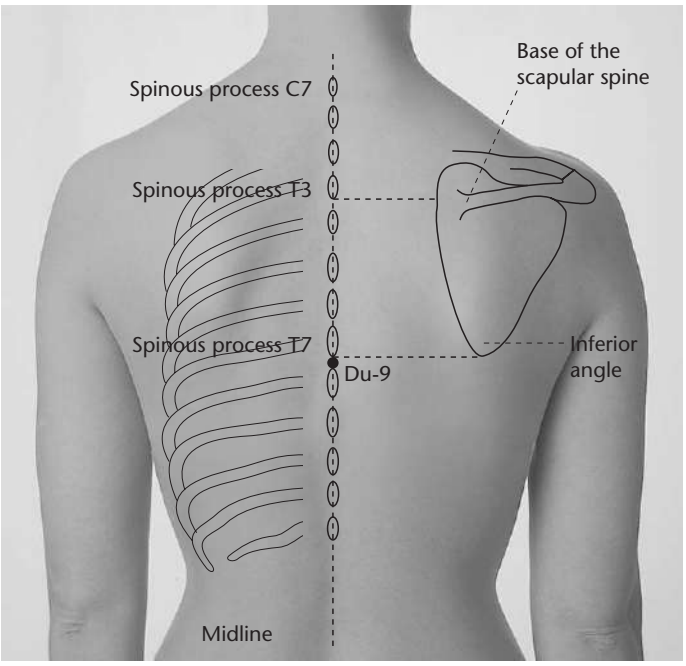

Location

On the midline, below the spinous process of the 7th thoracic vertebra (T7).

How to find

Find T7 (→ 3.4.2) and locate **Du-9** on the midline, below its spinous process. **Or:** Use the Tuffier's line for reference (→ 3.4.3). Locate T7 by counting up from L4 and there locate **Du-9** on the midline, below the spinous process.
Located on the same level are a point of → **Ex-B-2/BL-17/BL-46** (0.5/1.5/3 cun lateral to the midline).

Needling

0.5–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Strengthens the Spleen and eliminates Dampness
- Clears Damp-Heat from the Middle Burner
- Harmonises the Upper Burner

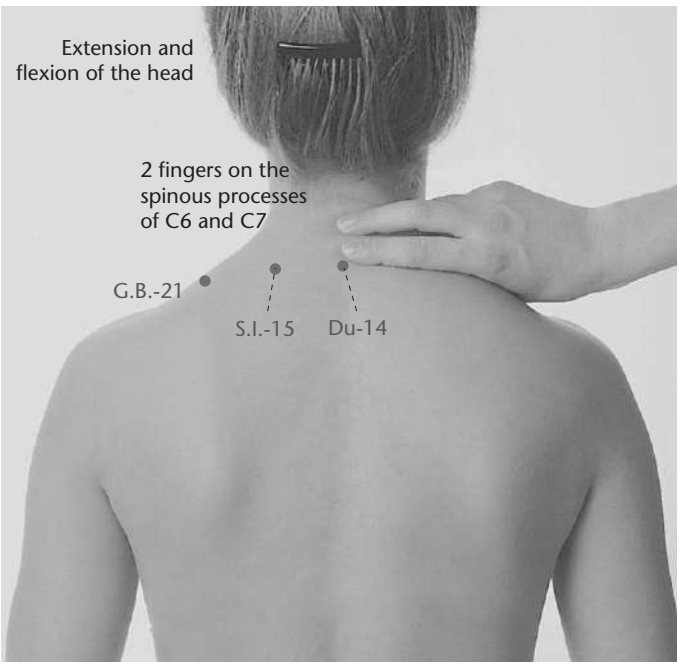

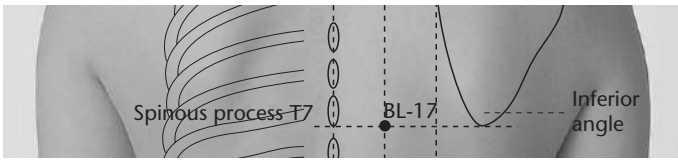

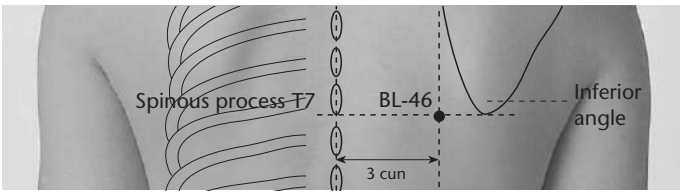

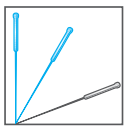
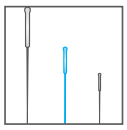
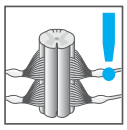

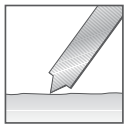

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 508

Du-10 *Spirit Tower* LINGTAI

Location

On the midline, below the spinous process of the 6th thoracic vertebra (T6).

How to find

Locate T7 (→ 3.4.2). **Du-10** is located on the midline, below the spinous process of the T6. **Or:** Locate C7 (→ 3.4.1). **Du-10** can then be located by counting down 6 spinous processes (T1–T6). Located on the same level are a point of → **Ex-B-2/BL-16/BL-45** (0.5/1.5/3 cun lateral to the midline).

Needling

0.5–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Harmonises the Upper Burner
- Clears Heat and Fire toxins

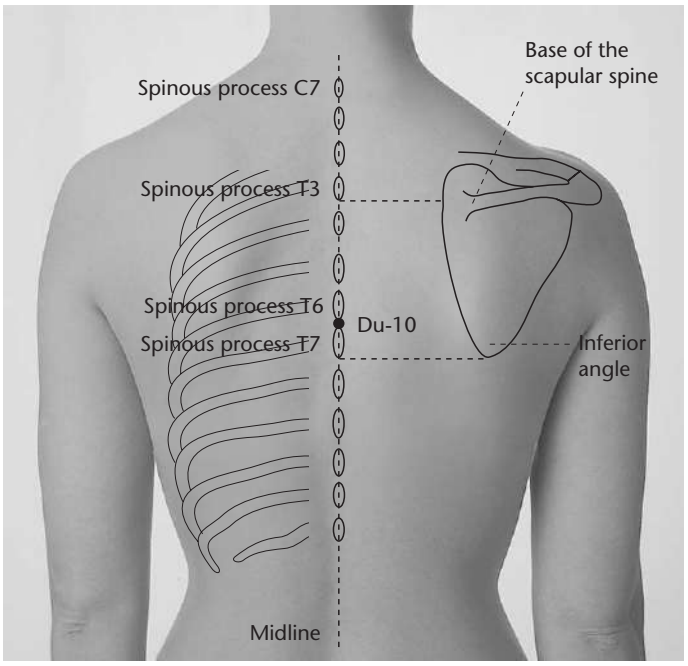

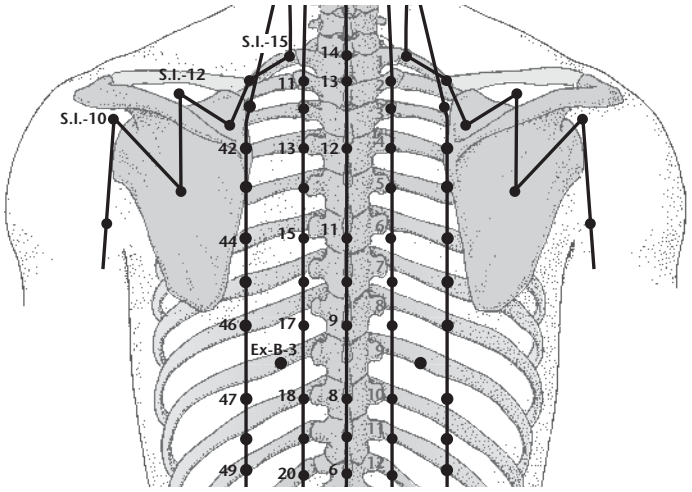

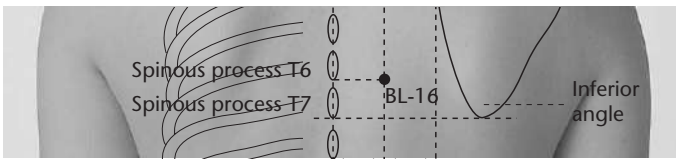

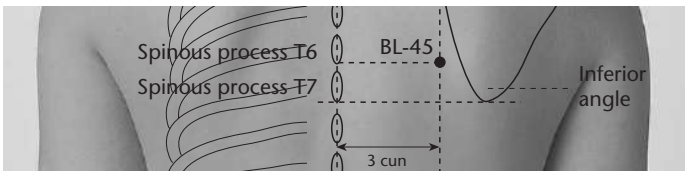

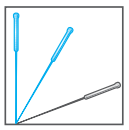

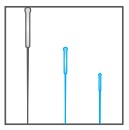

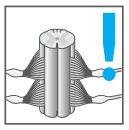

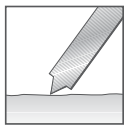

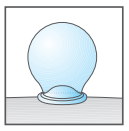

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 509

Spirit Pathway SHENDAO Du-11

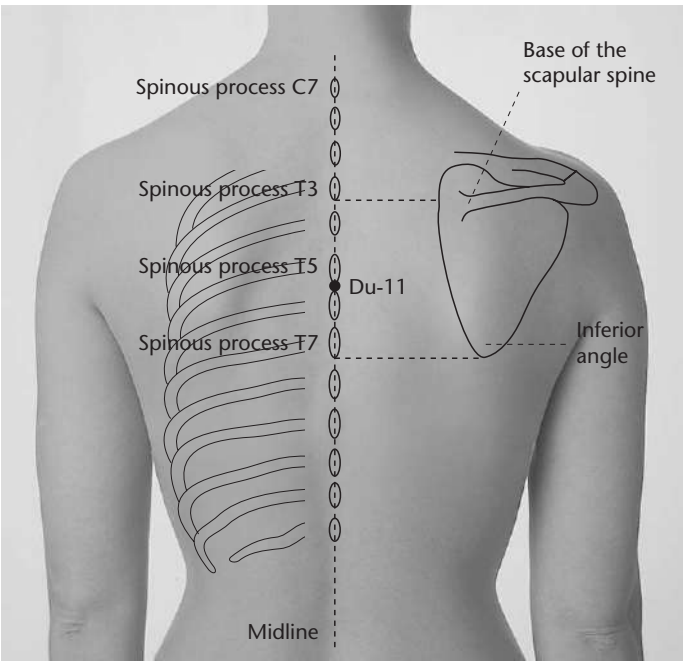

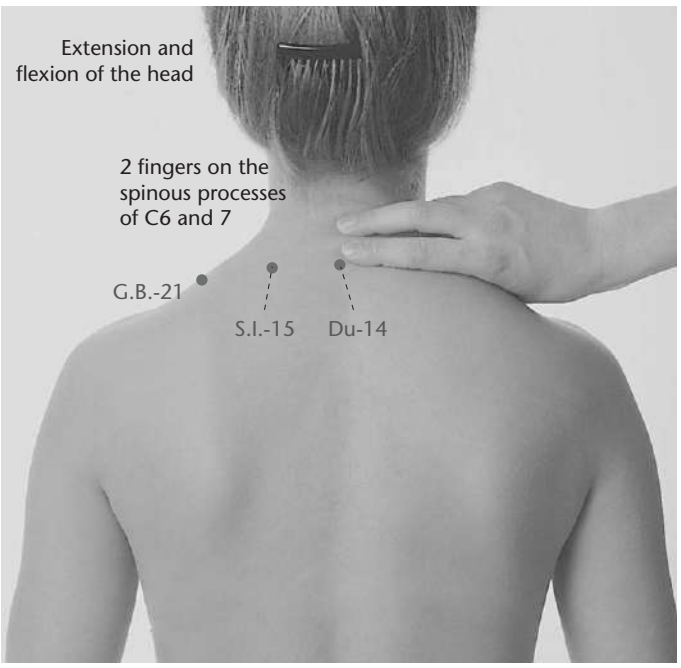

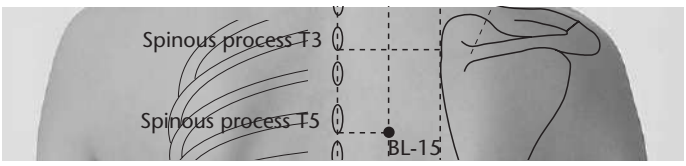

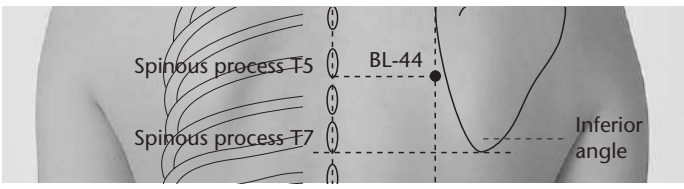

Location

On the midline, below the spinous process of the 5th thoracic vertebra (T5).

How to find

Locate T7 (→ 3.4.2). From there, count up 2 spinous processes and find **Du-11** on the midline, below the spinous process of T5.
Or: Locate C7 (→ 3.4.1). **Du-11** can then be located by counting down 5 spinous processes (T1–T5).
Located on the same level are a point of → **Ex-B-2/BL-15/BL-44** (0.5/1.5/3 cun lateral to the midline).

Needling

0.5–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Tonifies the Heart and the Lungs
- Calms the *shen*
- Dispels Wind and Heat

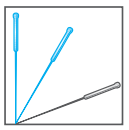
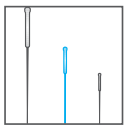
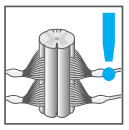

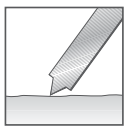
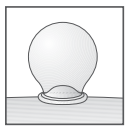

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 510

Du-12 *Body Pillar* SHENZHU

Location

On the midline, below the spinous process of the 3rd thoracic vertebra (T3).

How to find

Locate T7 (→ 3.4.2). From there, count up 4 spinous processes and locate **Du-12** on the midline, below the spinous process of T3. **Or:** Locate C7 (→ 3.4.1). **Du-12** can then be located by counting down 3 spinous processes (T1–T3).
Located on the same level are a point of → **Ex-B-2/BL-13/BL-42** (0.5/1.5/3 cun lateral to the midline).

Needling

0.3–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Dispels internal and external Wind
- Calms the *shen*
- Tonifies the Lungs

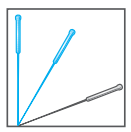

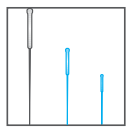

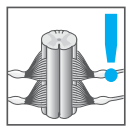

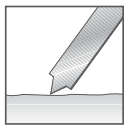

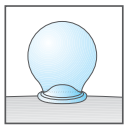

Special features

Meeting point with the BL channel

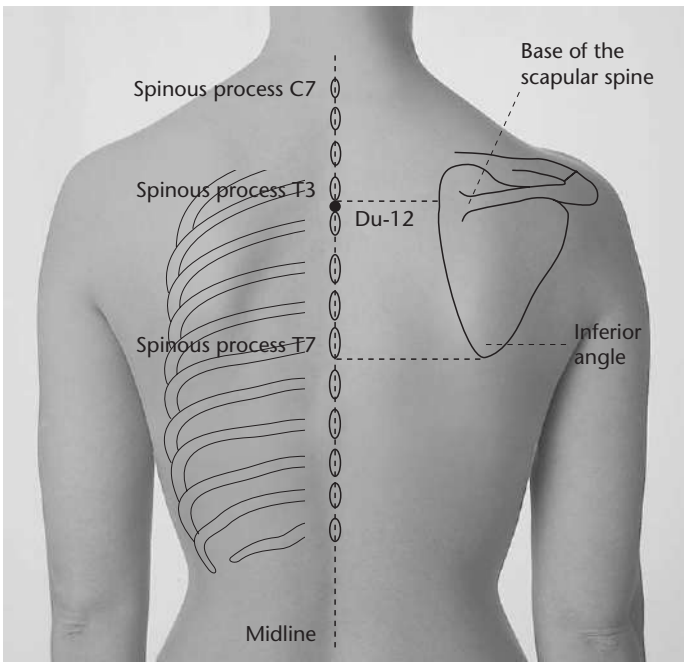

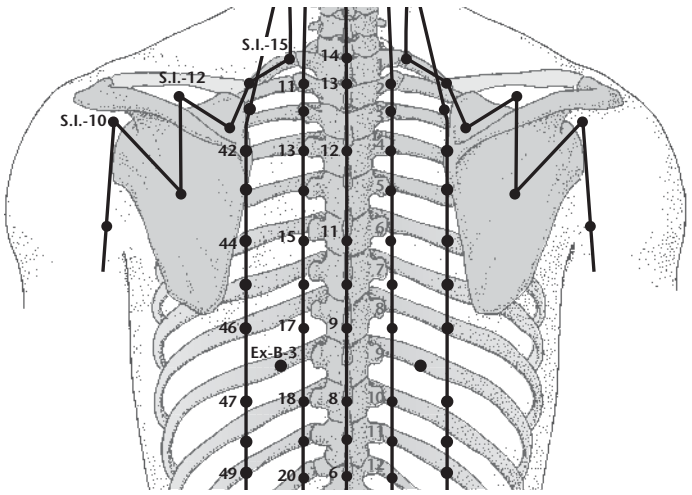

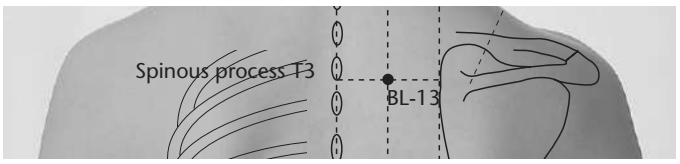

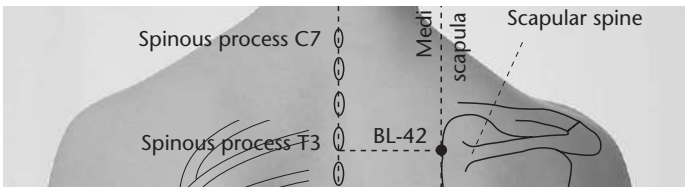

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 511

Way of Happiness TAODAO Du-13

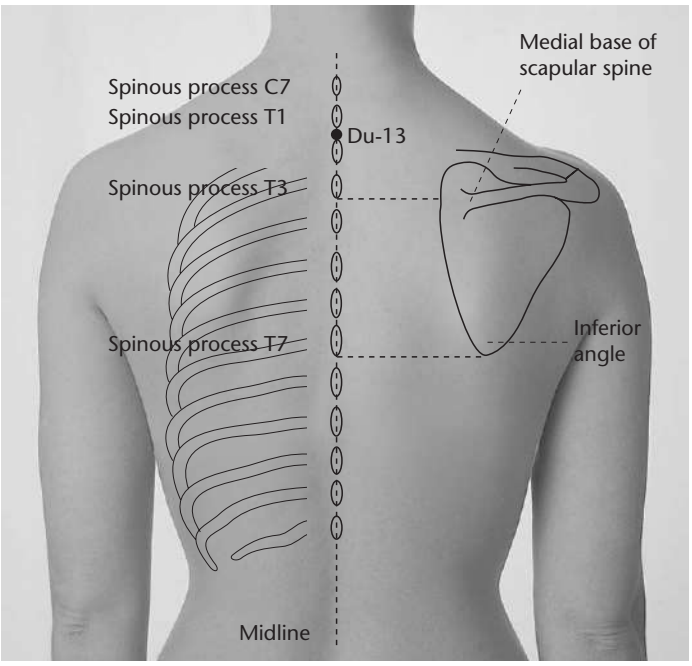

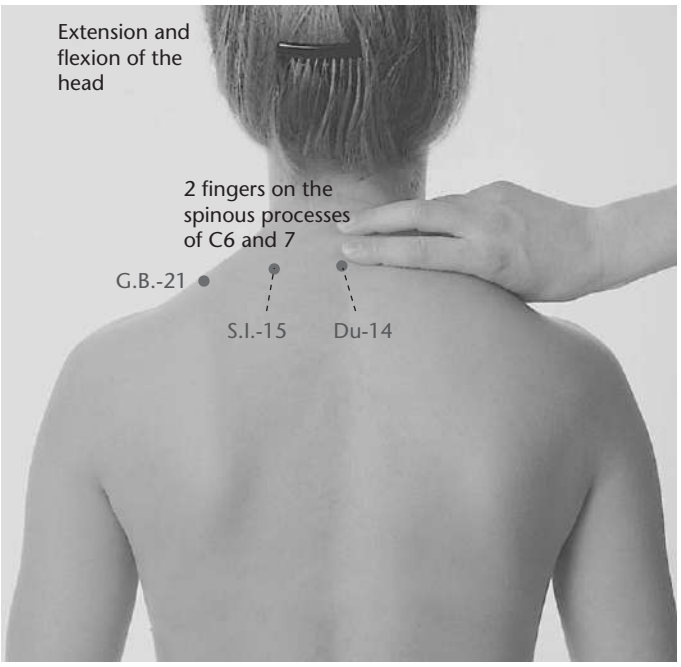

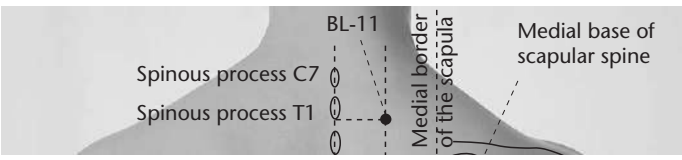

Location

On the midline, below the spinous process of the 1st thoracic vertebra (T1).

How to find

Locate C7 (→ 3.4.2). Palpate downward to the next spinous process (T1) and locate **Du-13** on the midline, below its spinous process.

Located on the same level are a point of → **Ex-B-2/BL-11/S.I.-14** (0.5/1.5/3 cun lateral to the midline).

Needling

0.3–1 cun strictly perpendicularly to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun.

Actions/Indications

- Dispels pathogenic factors
- Opens the channel and *luo*-connecting vessels

Special features

Meeting point with the BL channel

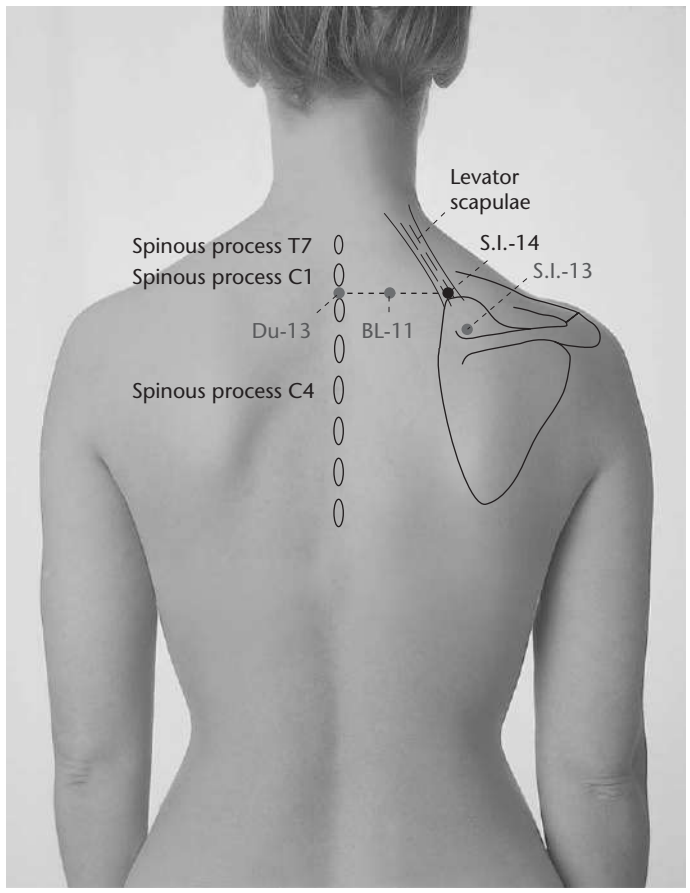

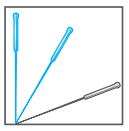
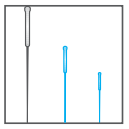
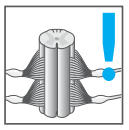

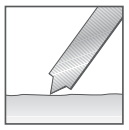

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 512

Du-14 *Great Vertebra* DAZHUI

Location

On the midline, below the spinous process of the 7th cervical vertebra (C7).

How to find

First, identify C7 (→ 3.4.1): Place 2 fingers on the spinous processes believed to belong to C6 and C7 and ask the patient to flex and extend their head. In a fully functional spine and with correct finger placement, C6 will glide anteriorly with neck extension, while C7 will remain fixed. However, if the vertebra below the upper finger does not move with neck extension, the fingers will most probably have been placed on C7 and T1. Locate **Du-14** below the spinous process of C7.

Located on the same level are → **Ex-B-1** (*dingchuan*)/ **S.I.-15/Ex-B** (*jiehexue*) (0.5/2/3.5 cun lateral to **Du-14**).

Needling

0.3–1 cun strictly vertically to the skin or obliquely in an inferior direction, with the patient's back straight or flexed in order to avoid puncturing the spinal canal. Oblique insertions should only be carried out by experienced practitioners, as in small persons (irrespective of their body weight), the spinal canal may be reached after only 1.25 cun. For persistent epistaxis, place a cold metal object (for example, a spoon) or a cold cloth on **Du-14**. Cupping on this point for acute colds.

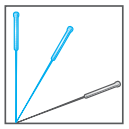

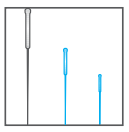

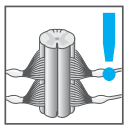

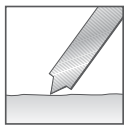

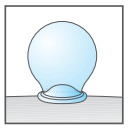

Actions/Indications

- Clears Wind, firms and regulates the surface, dispels pathogenic factors
- Clears Heat
- Calms (internal) Wind and the *shen*
- Strengthens Qi and Yang in deficiency conditions
- Benefits the spine, especially the cervical spine

Special features

Meeting point with all Yang channels, Sea of Qi point.

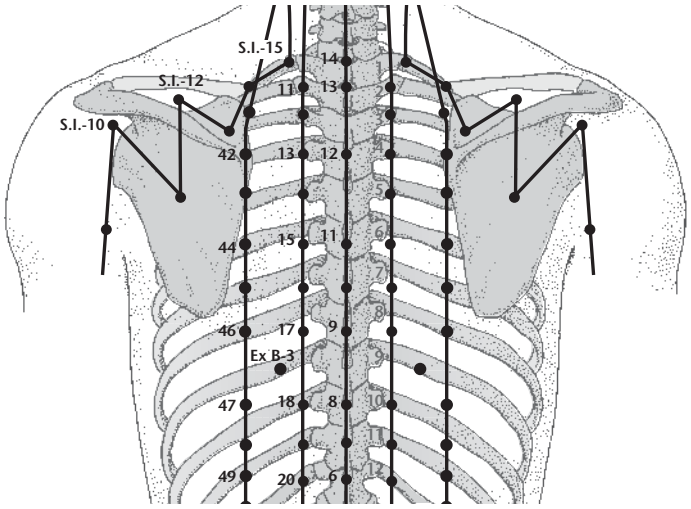

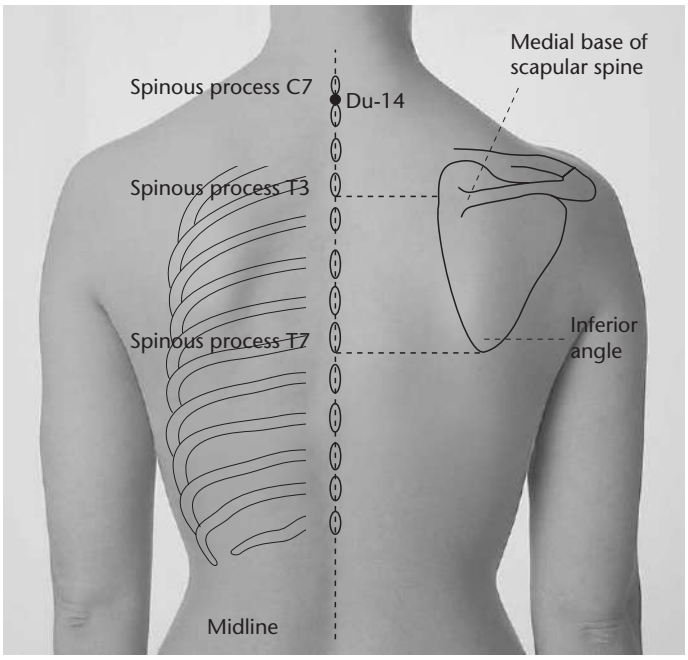

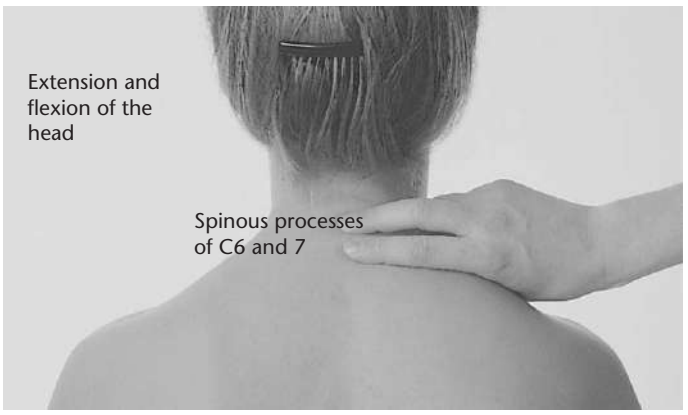

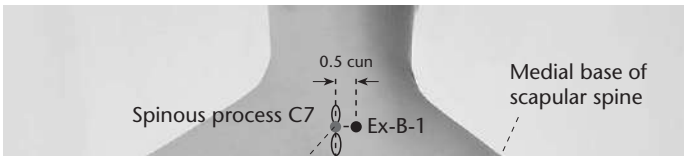

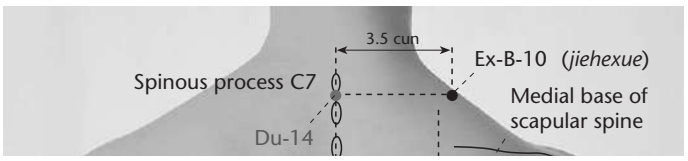

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 513

Gate of Muteness YAMEN Du-15

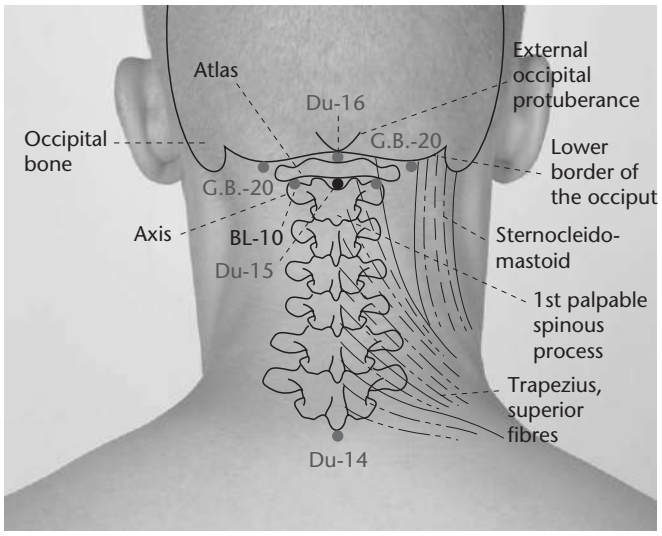

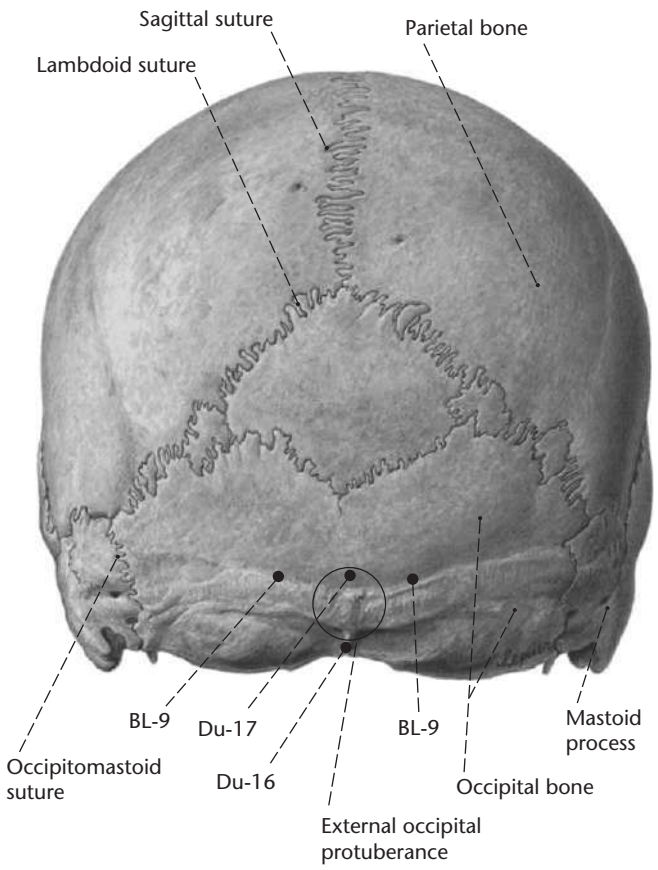

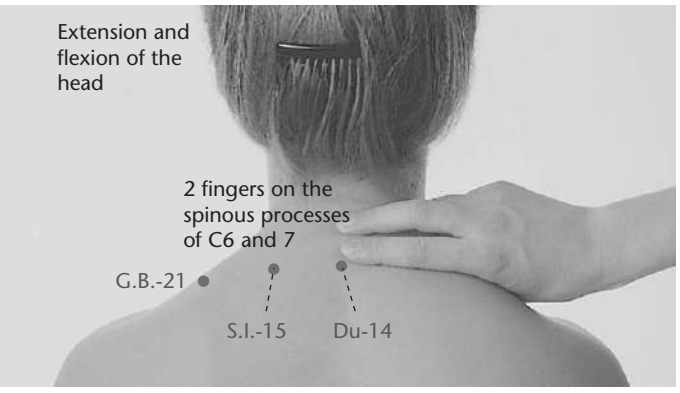

Location

On the occiput, on the posterior midline, in the depression between the 1st (atlas) and 2nd (axis) cervical vertebrae, approximately 0.5 cun inferior to ➞ **Du-16** (directly below the external occipital protuberance).

How to find

The external occipital protuberance (➞ 3.1.5) is a hump-shaped projection on the posterior surface of the occipital bone, on the midline and slightly superior to the craniocervical junction (➞ 3.1.5). ➞ **Du-16** is located in a depression on the posterior midline that you can palpate directly inferior to the protuberance. From there, glide 0.5 cun in an inferior direction and locate **Du-15** superior to the first palpable spinous process (belonging to the axis; the atlas has no spinous process) and approximately 0.5 cun superior to the posteror hairline.
Located on the same level is ➞ **BL-10** (on the lateral aspect of the trapezius muscle).

Needling

0.5–1 cun strictly perpendicularly in an inferior direction. Caution: Injury to the cervical marrow. According to classic texts, moxibustion is contraindicated.

Actions/Indications

- Benefits the tongue, ears, neck and spine
- Dispels Wind
- Clears Yang Heat/excess

Special features

Meeting point with the *yang wei mai*. Sea of Qi point.

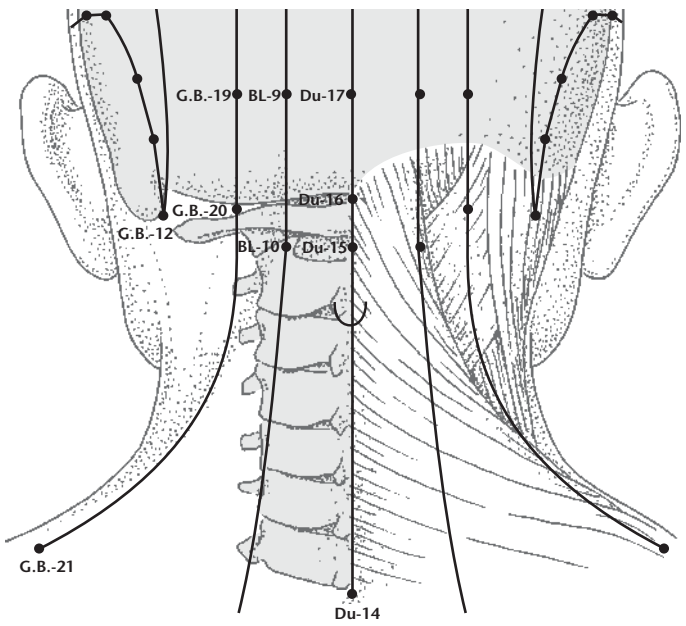

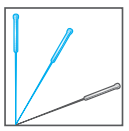
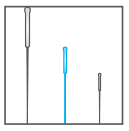
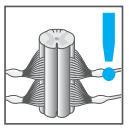

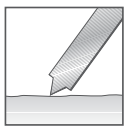

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 514

Du-16 *Palace of Wind* FENGFU

Location

On the posterior midline, directly below the external occipital protuberance, in the depression between the origins of the trapezius muscle.

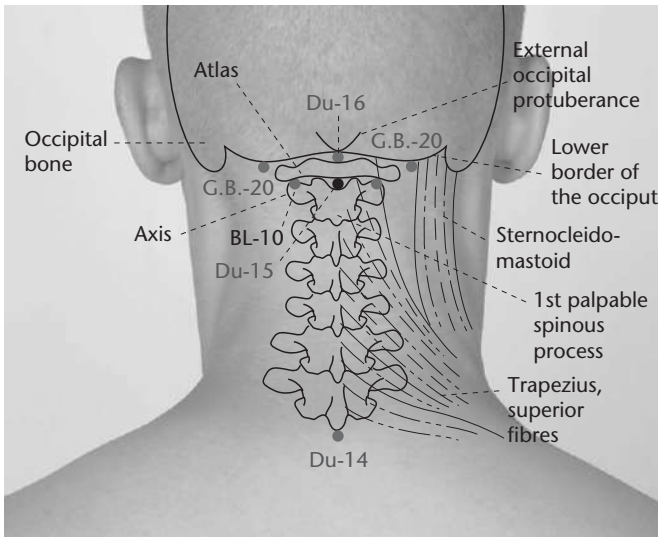

How to find

The external occipital protuberance (→ 3.1.5) is a hump-shaped projection on the posterior surface of the occipital bone, on the midline and slightly superior to the craniocervical junction (→ 3.1.5). Locate **Du-16** in a depression on the posterior midline that can be palpated directly inferior to the protuberance and between the two origins of the trapezius muscle. Generally, **Du-16** is located approximately 1 cun superior to the posterior hairline. → **G.B.-20** is located on the same level below the occiput, in a depression between the origins of the trapezius and sternocleido-mastoid muscles.

Needling

0.5–1 cun slightly obliquely in an inferior direction. Bend the patient's head slightly forward when needling. Caution: Dangerous point with deep needling! No stimulation. Strictly no needling in a superior direction, danger of puncturing the cerebellomedullary cistern (also site for suboccipital puncture). The needle should be inserted into the nuchal ligament. According to some authors, moxibustion is contraindicated.

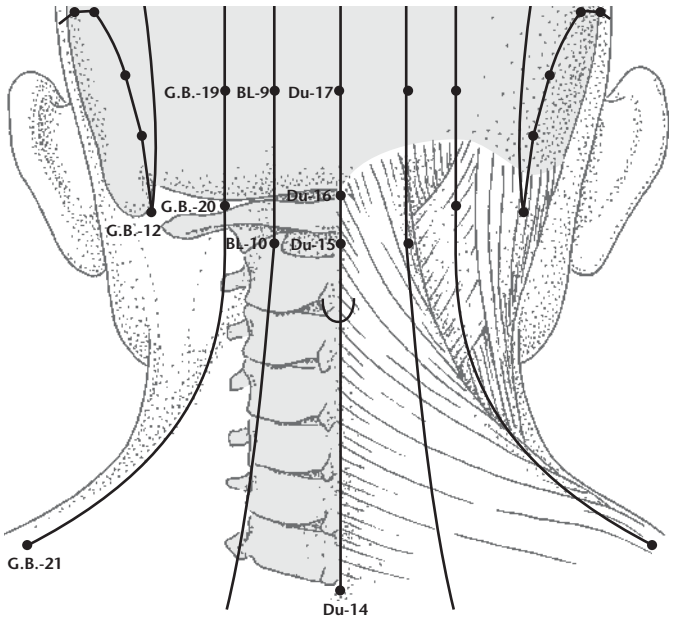

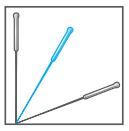

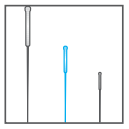

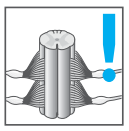

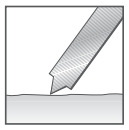

Actions/Indications

- Dispels (external) Wind
- Calms internal Wind and the *shen*
- Nourishes the Sea of Marrow, lightens the *shen*

Special features

Meeting point with the *yang wei mai* and, according to some authors, the *yang qiao mai*. Sea of Marrow point, Window of Heaven point, Sun Si Miao Ghost point. Used for all disorders which are caused by 'external or internal' Wind and which are characterised by sudden onset, changing location and/or intensity of the disorder, involuntary movements.

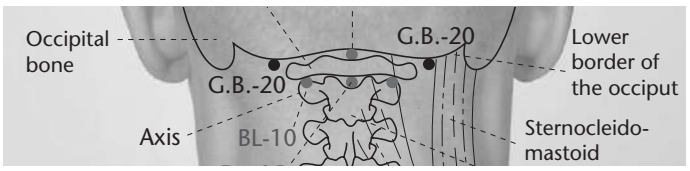

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 515

Brain's Door NAOHU Du-17

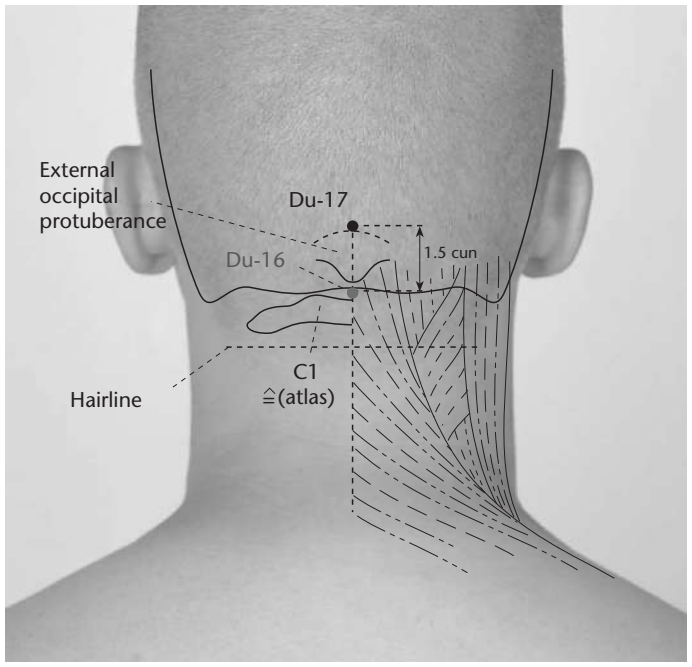

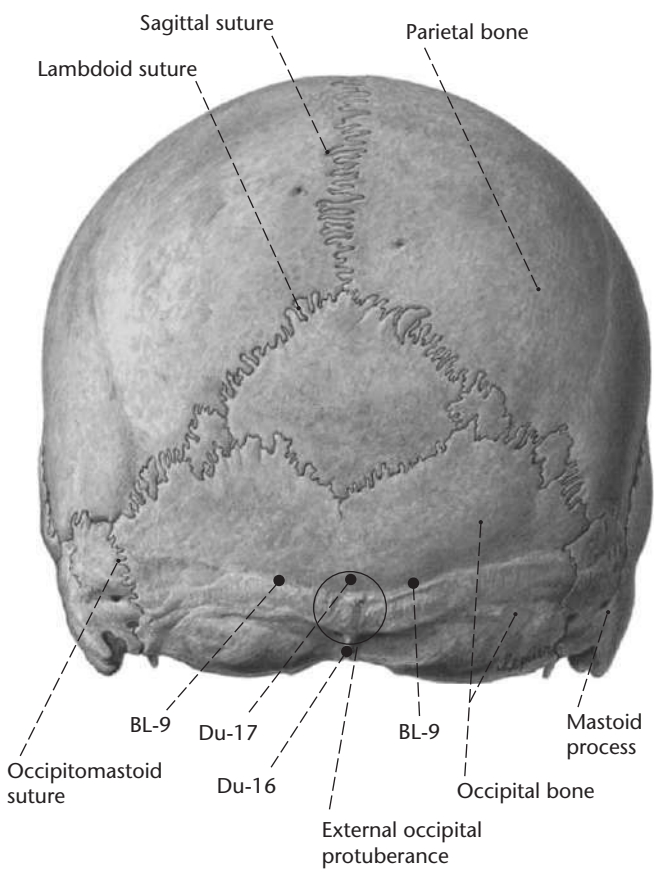

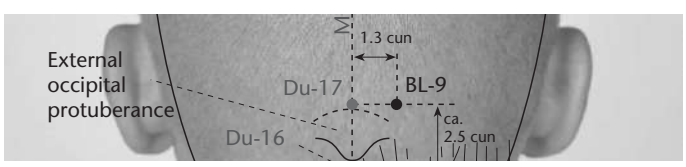

Location

In a depression superior to the external occipital protuberance, approximately 2.5 cun superior to the posterior hairline or 1.5 cun superior to → **Du-16**.

How to find

First, locate the external occipital protuberance (→ 3.1.5), which forms a hump-shaped projection on the posterior surface of the occipital bone. Next, locate **Du-17** on the posterior midline, in a depression directly superior to the protuberance. This corresponds to a distance of approximately 2.5 cun superior to the posterior hairline (→ 3.1.5).

→ **Du-16** is located directly inferior to the protuberance, whereas → **BL-9** is located 1.3 cun lateral to **Du-17**.

Needling

0.5–1 cun transversely (subcutaneously) in an inferior or superior direction. Moxibustion controversial.

Actions/Indications

- Dispels Wind
- Opens the channel
- Benefits the eyes
- Calms the *shen*

Special features

Meeting point with the BL channel

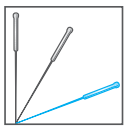
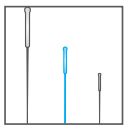

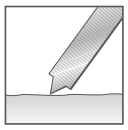

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 516

Du-18 *Unyielding Space* QIANGJIAN

Location

On the posterior midline, 1.5 cun superior to **Du-17** (directly superior to the external occipital protuberance) or 3 cun inferior to ➞ **Du-20** (on the vertex).

How to find

First, locate ➞ **Du-17** in a depression directly superior to the external occipital protuberance (➞ 3.1.5), which forms a hump-shaped projection on the posterior surface of the occipital bone. From there, palpate 1.5 cun in a superior direction and there locate **Du-18** on the posterior midline.

➞ **Du-20** is located 3 cun in a superior direction, on the vertex.

Needling

Transversely (subcutaneously) 0.5–1 cun

Actions/Indications

- Dispels Wind, especially internal Wind
- Opens the channel
- Calms the *shen*

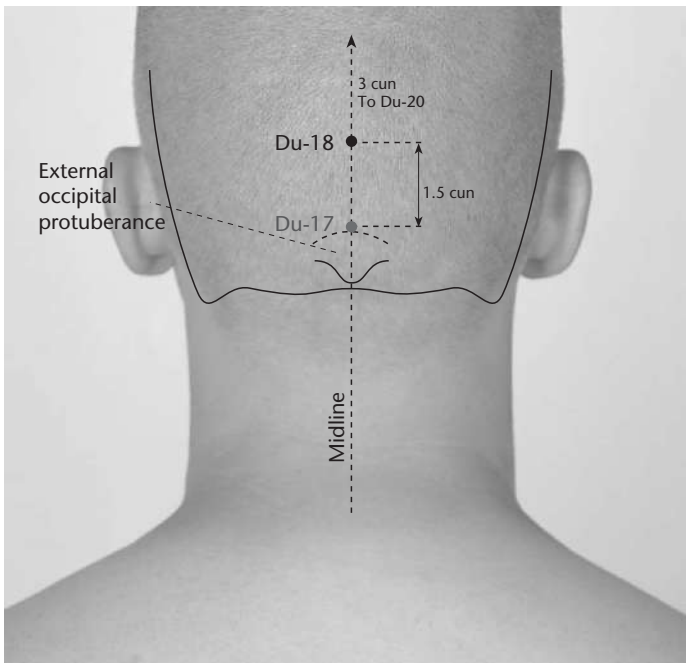

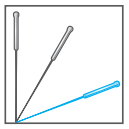
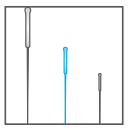

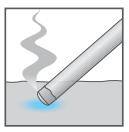
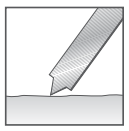

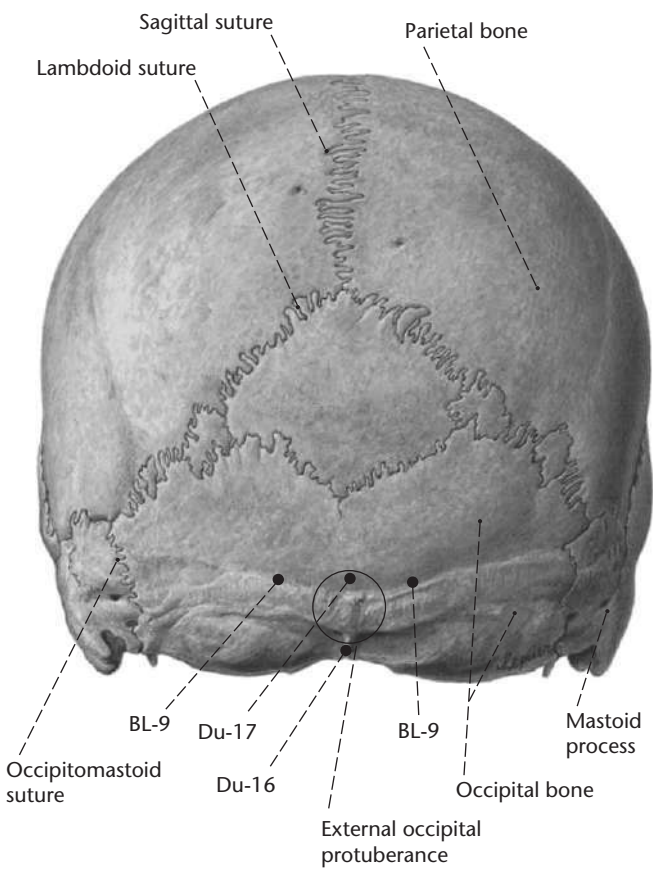

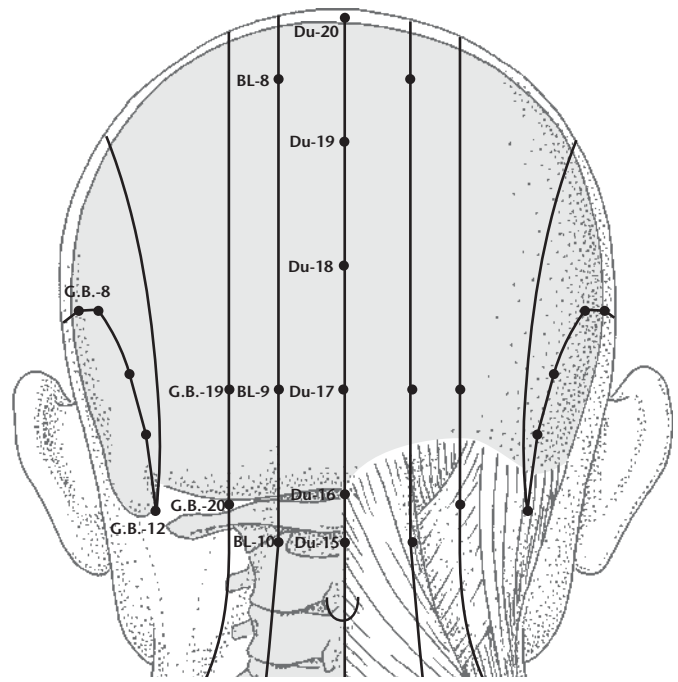

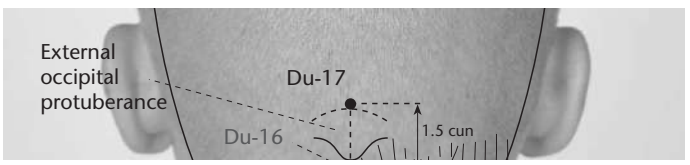

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 517

Behind the Crown HOUDING Du-19

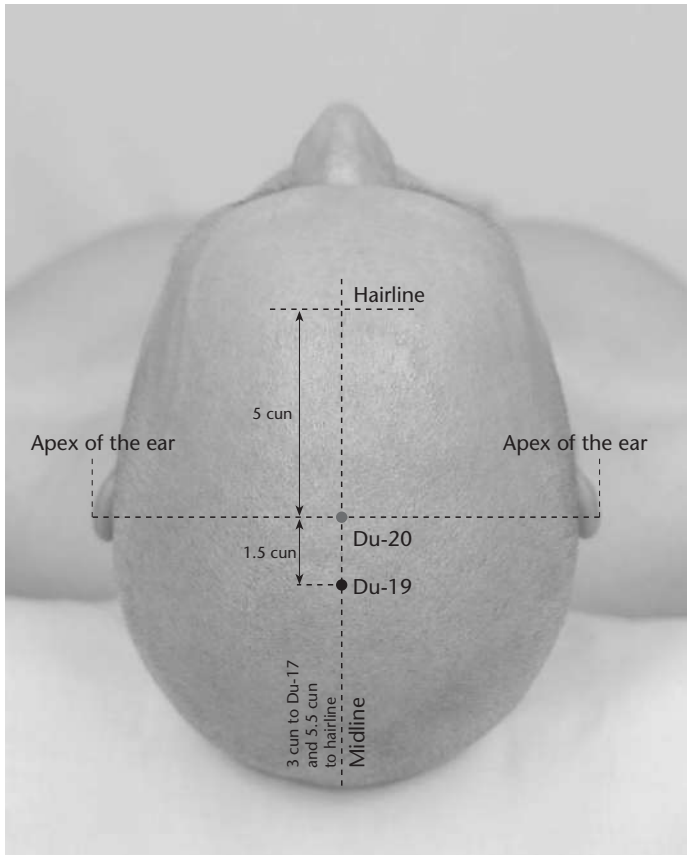

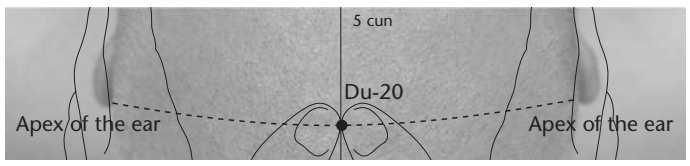

Location

On the midline, 3 cun superior to → **Du-17** (directly superior to the external occipital protuberance) or 1.5 cun posterior to → **Du-20**.

How to find

This point is best located by using → **Du-20** as a reference point (at the junction of the vertical midline and a line connecting the apices of the ears). **Du-19** can then be located by palpating 1.5 cun in a posterior direction from → **Du-20**. **Or:** First, locate → **Du-17** superior the external occipital protuberance (→ 3.1.5), then locate **Du-19** on the midline, 3 cun superior to → **Du-17**.

Needling

Transversely (subcutaneously) 0.5–1 cun.

Actions/Indications

- Dispels Wind
- Opens the channel
- Calms the *shen*

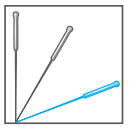
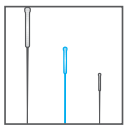

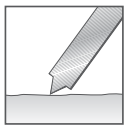
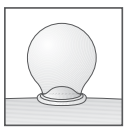

Ch05b-F10028.qxd 2/22/08 9:55 PM Page 518

Du-20 *Hundred Meetings* BAIHUI

Location

At the junction of a line connecting the apices of the ears and the midline, 5 cun from the anterior or 7 cun from the posterior hairline respectively.

How to find

Spreading hands technique (→ 2.3.3): Place the hands on both sides of the head, with the little fingers touching the apices of the ears. Join the thumbs on the midline and locate **Du-20** in shallow depression on the vertex of the head (alternative location method: use an elastic tape, with the midpoint marked on it). The symbolic counterpart to **Du-20** (as the highest point on the body and in contact with heaven) is → **KID-1**, the lowest part of the body and in contact with the earth.

Needling

0.5–1 cun transversely (subcutaneously) towards the posterior (reducing effect) or the anterior (tonifying effect). Caution: Apply pressure to head points after needle removal to avoid bleeding. Remember to remove the needle after treatment.

Actions/Indications

- Calms Wind, pacifies Yang, benefits the Brain and sensory organs, calms the *shen*
- Nourishes the Sea of Marrow; applying moxibustion raises the Yang

Special features

Meeting point with the BL, G.B., T.B. and LIV channels; Sea of Marrow point. Important point for descending the Yang (reducing needle techniques) or raising the Yang (tonifying needle techniques, moxibustion).

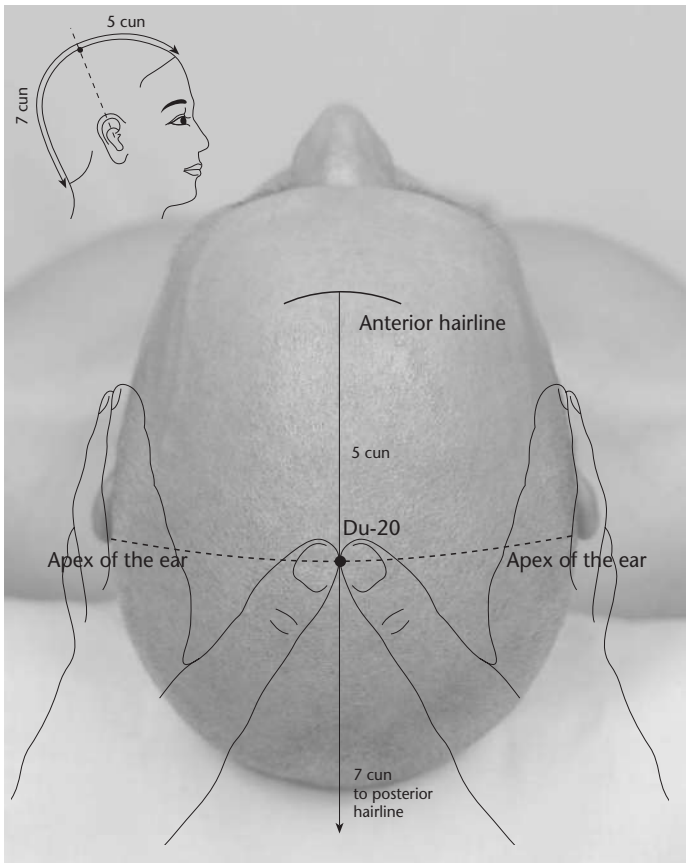

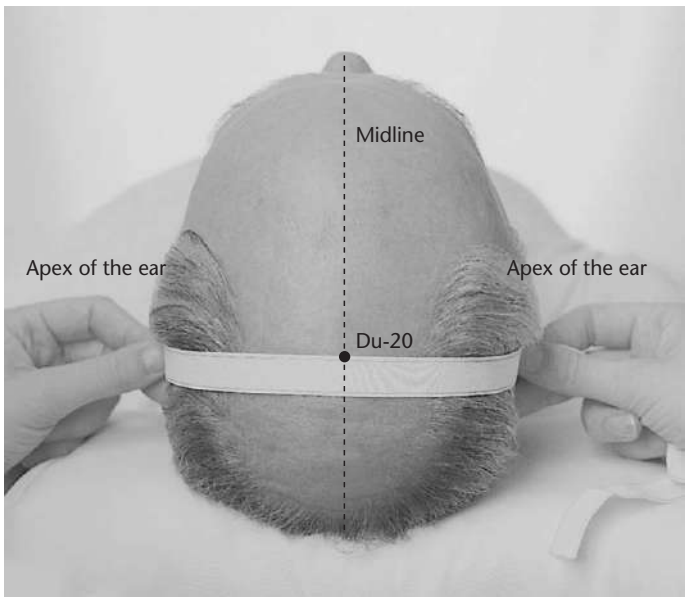

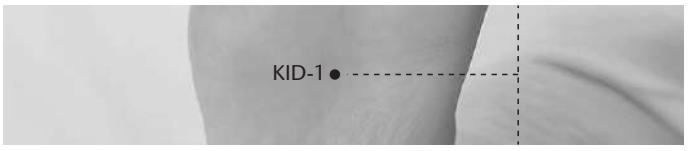

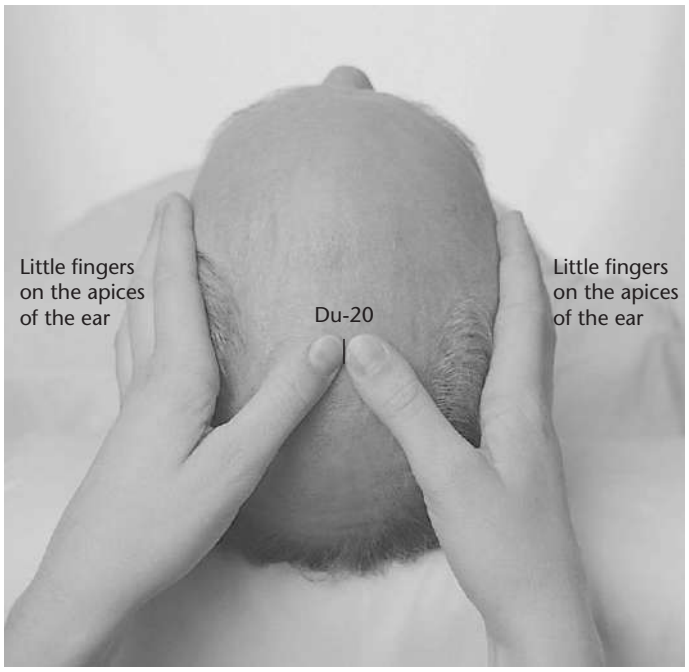

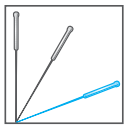
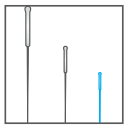

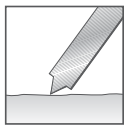

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 519

In Front of the Crown QIANDING Du-21

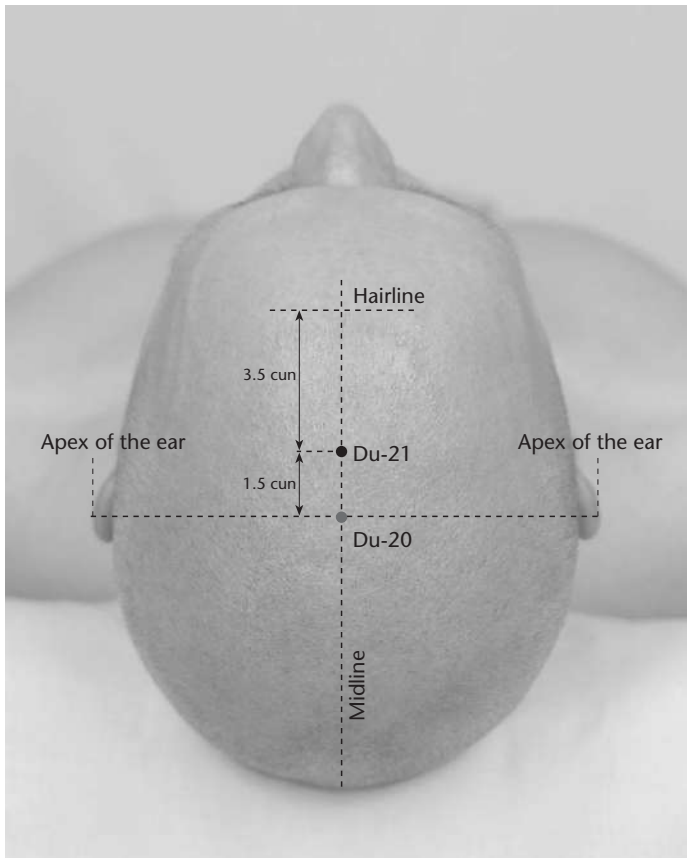

Location

On the midline, 3.5 cun superior to the anterior hairline or 1.5 cun anterior to → **Du-20**.

How to find

First, locate → **Du-20** (at the junction of the vertical midline and a line connecting the apices of the ears) and, from there, measure 1.5 cun in an anterior direction. **Or:** Locate the anterior hairline (→ 3.1.1; the distance from the anterior hairline to → **Du-20** is 5 cun) and palpate 3.5 cun in a superior/posterior direction. There, locate **Du-21** on the midline.

Needling

Transversely (subcutaneously) 0.5–1 cun. Caution with infants whose fontanelle has not yet closed. Moxibustion possible.

Actions/Indications

- Dispels Wind and Dampness
- Opens the channel
- Calms the *shen*

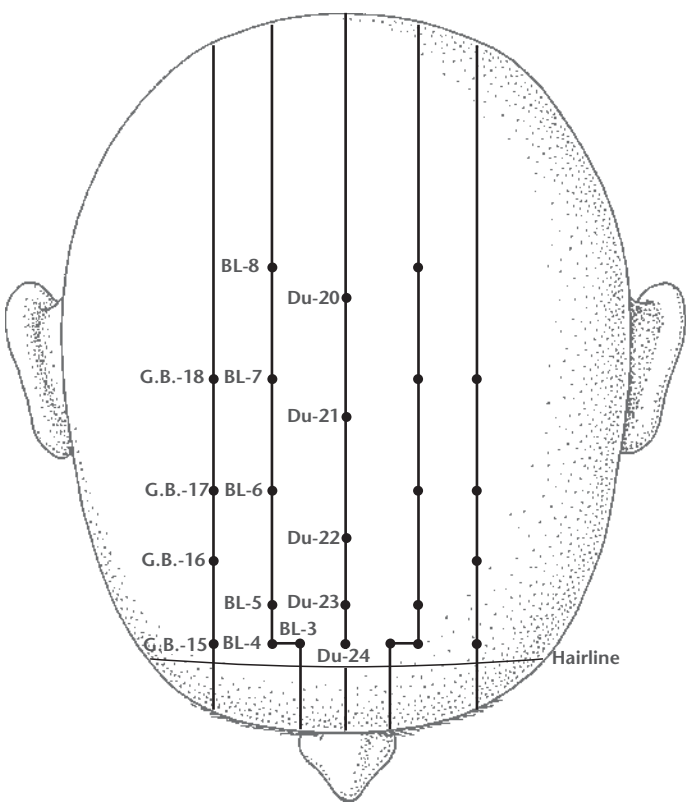

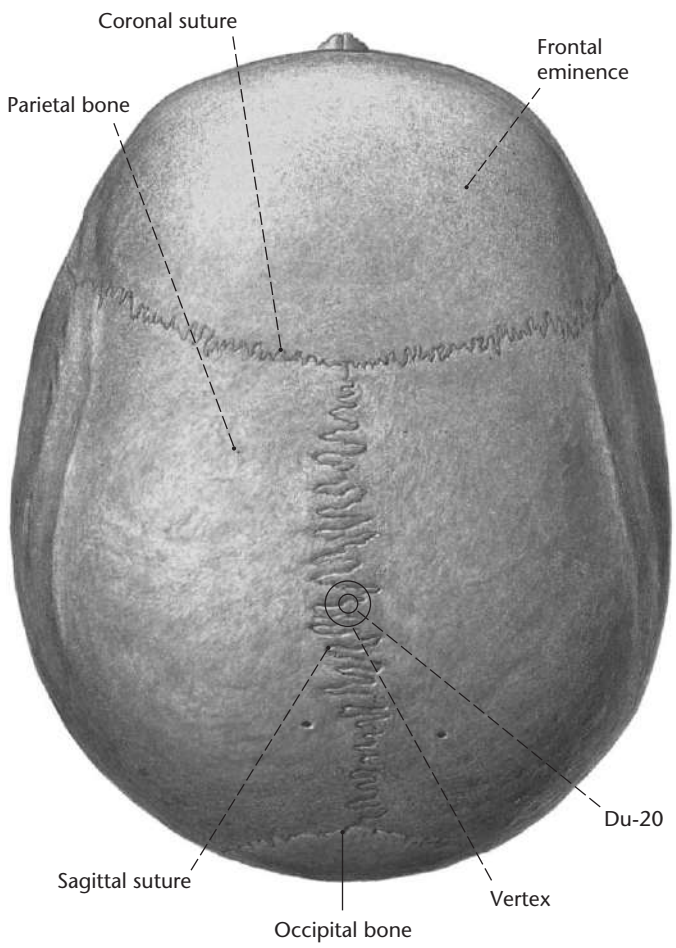

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 520

Du-22 *Fontanelle Meeting* XINHUI

Location

On the midline, 2 cun superior to the anterior hairline.

How to find

First, locate the anterior hairline (→ 3.1.1) and, from there, palpate 2 cun in a superior direction. There, locate **Du-22** on the midline. For reference: The distance from the anterior hairline to → **Du-20** is 5 cun. **Or:** From → **Du-20** (at the junction of the vertical midline and a line connecting the apices of the ears), measure 3 cun in an anterior direction and there locate **Du-22** on the midline.

Needling

Transversely (subcutaneously) 0.5–1 cun. Caution with infants whose fontanelle has not yet closed (**Du-22** is located on the anterior border of the fontanelle). Moxibustion possible.

Actions/Indications

- Dispels Wind
- Calms the *shen*

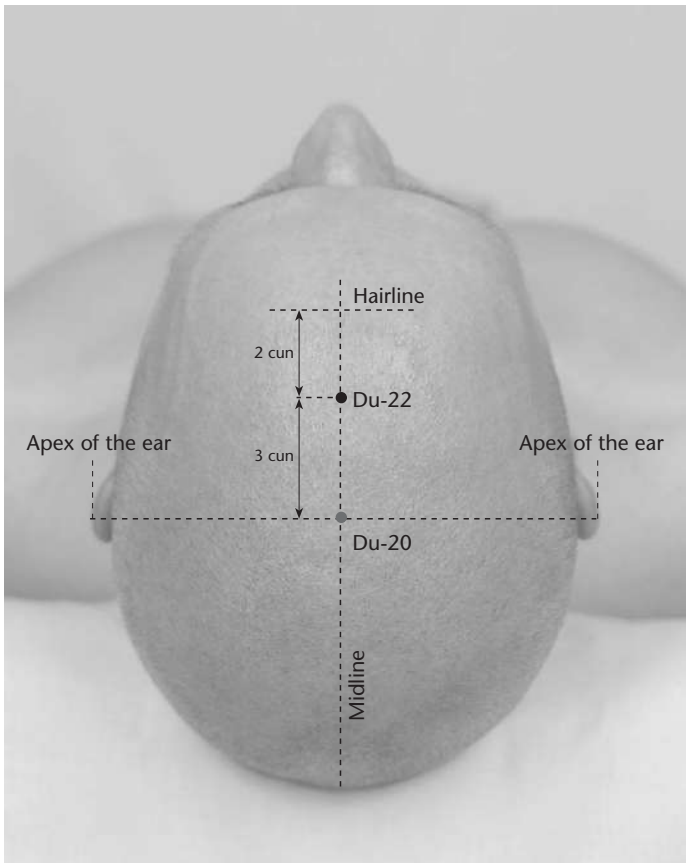

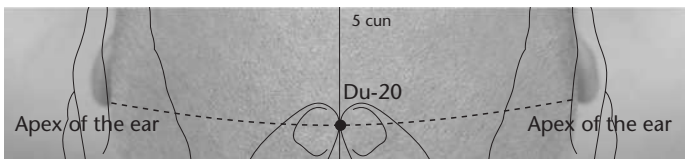

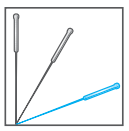

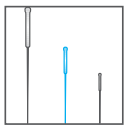

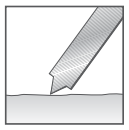

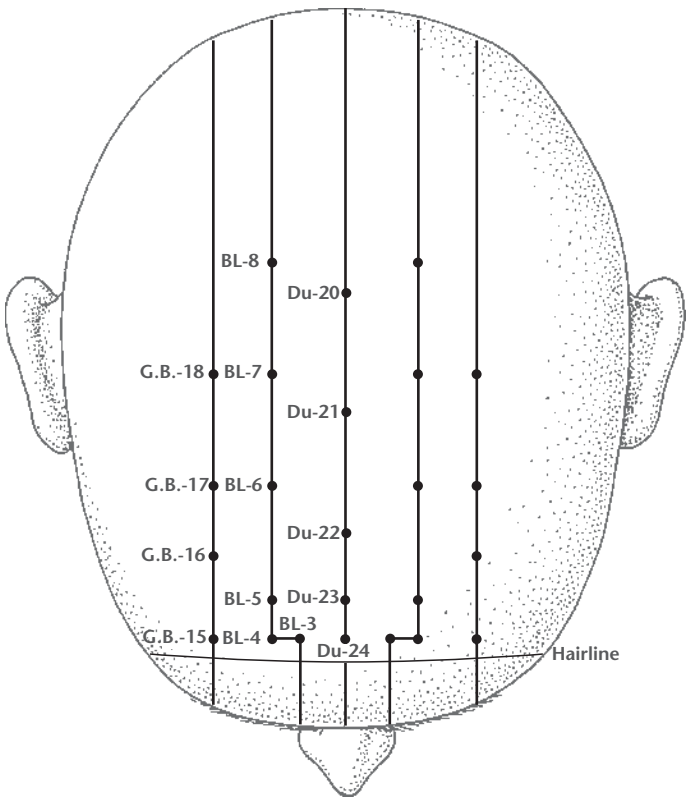

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 521

Upper Star SHANGXING Du-23

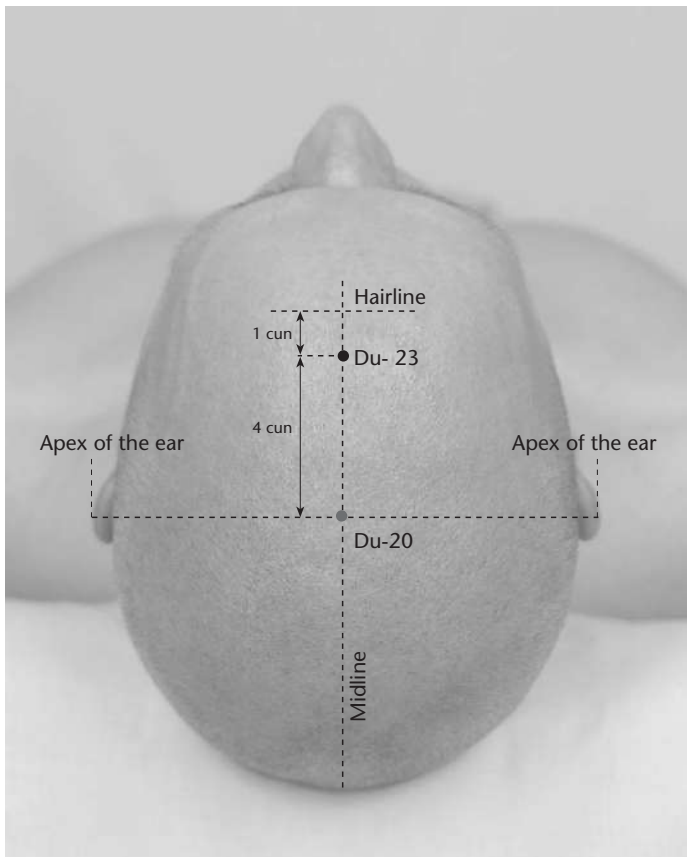

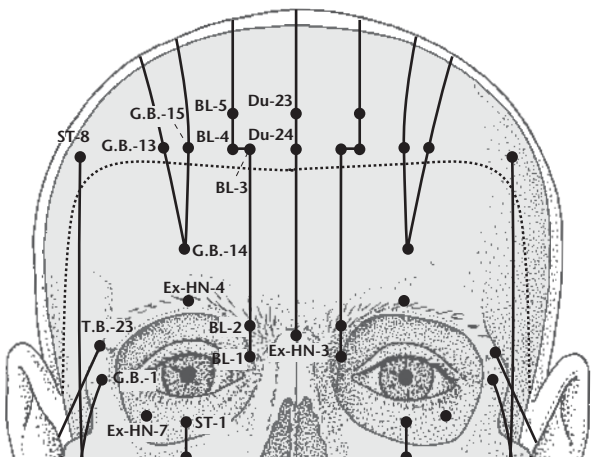

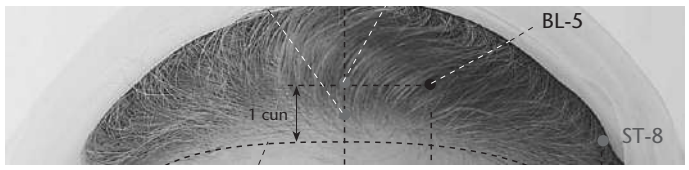

Location

On the midline, 1 cun superior to the anterior hairline or 4 cun anterior to → **Du-20**.

How to find

First, locate the anterior hairline (→ 3.1.1) and, from there, palpate 1 cun in a superior direction. There, locate **Du-23** on the midline. For reference: The distance from the anterior hairline to → **Du-20** is 5 cun.

Located on the same level is → **BL-5** (1.5 cun lateral to the midline).

Needling

Transversely (subcutaneously) 0.3–0.5 cun. Moxibustion possible.

Actions/Indications

- Dispels pathogenic factors (especially Wind and Heat) from the nose
- Dispels pathogenic factors (especially Heat) from the face
- Calms the *shen*

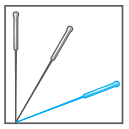
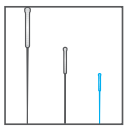

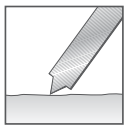
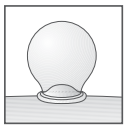

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 522

Du-24 *Courtyard of the Spirit* SHENTING

Location

On the midline, 0.5 cun superior to the anterior hairline or 4.5 cun anterior to → **Du-20**.

How to find

First, locate the anterior hairline (→ 3.1.1) and, from there, palpate 0.5 cun in a superior direction. There, locate **Du-24** on the midline. For reference: The distance from the anterior hairline to → **Du-20** (at the junction of the vertical midline and a line connecting the apices of the ears) is 5 cun.

Located on the same level (0.5 cun superior to the anterior hairline) are → **BL-3/BL-4/G.B.-15/G.B.-13** (superior to the inner canthus of the eye/1.5 cun lateral to the midline/on the pupil line or 2.25 cun lateral to the midline/3 cun lateral to the midline). These distances refer to the proportional cun distance (→ 2.2) between **Du-24** and → **ST-8** (on the corner of the forehead), which equals 4.5 cun (1.5 cun lateral to the midline).

Needling

Transversely (subcutaneously) 0.3–0.5 cun. Moxibustion possible.

Actions/Indications

- Calms the *shen*
- Dispels (internal) Wind
- Clears pathogenic factors from the eyes and nose

Special features

Meeting point with the BL and ST channels. One of the major points for calming the *shen*.

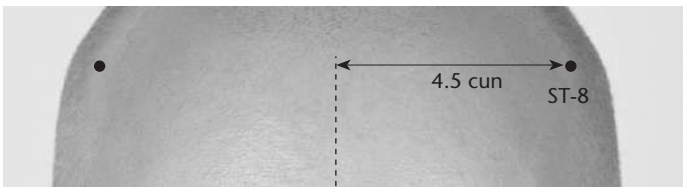

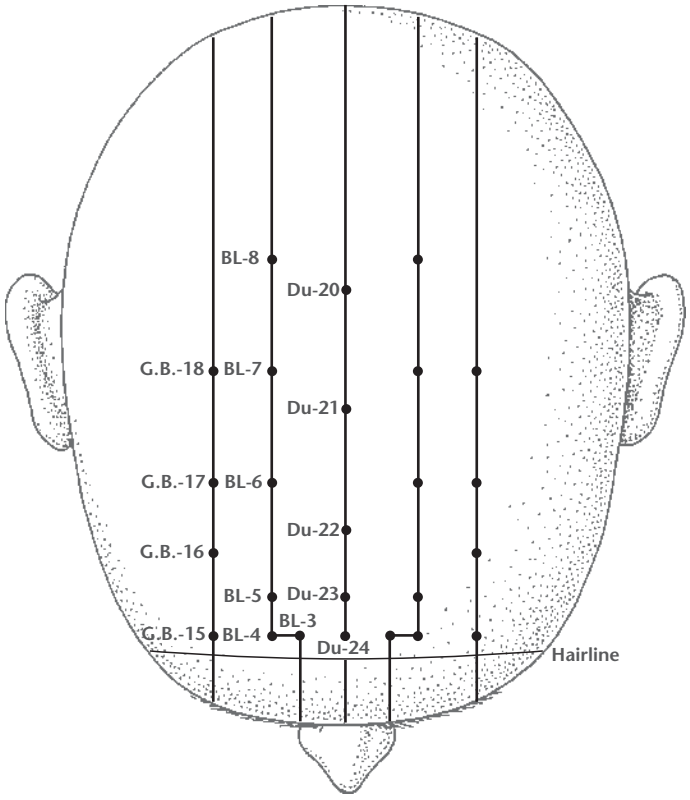

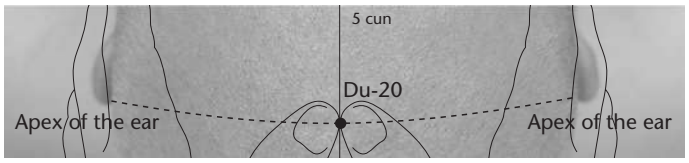

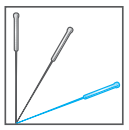
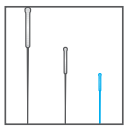

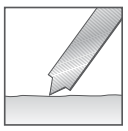

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 523

White Crevice SULIAO Du-25

Location

In a depression on the tip of the nose.

How to find

As the name implies, **Du-25** can be located in a depression on the tip of the nose, which generally can be easily palpated.

Needling

0.2 cun vertically or transversely (subcutaneously) in a superior direction up to 1 cun. Or prick to bleed. Moxibustion controversial.

Actions/Indications

- Benefits the nose

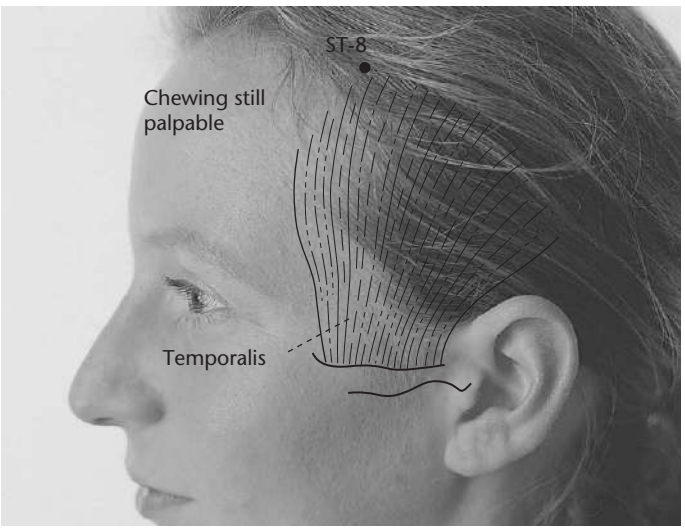

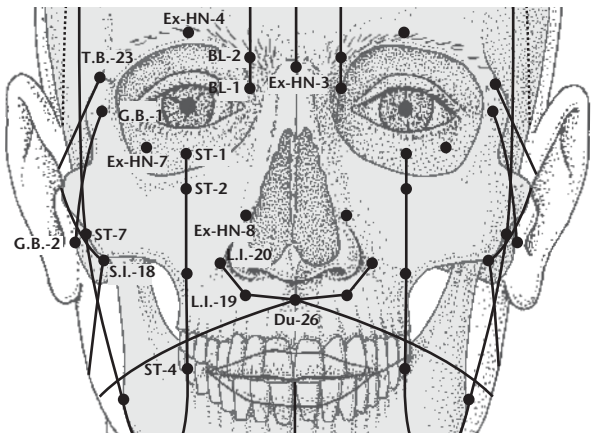

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 524

Du-26 *Man's Middle* RENZHONG

Location

Below the nose, on the upper third of the philtrum. Note: *shuigou* (Water Grave) is an alternative name for this point.

How to find

Locate the philtrum, which forms a distinct groove on the midline between the root of the nose and the margin of the upper lip. Locate **Du-26** sligthly superior to the midpoint of the philtrum.
→ **L.I.-19** is located on the same level, 0.5 cun lateral to the midline.

Needling

0.3–0.5 cun obliquely in a superior direction. Caution: Needling this point may be painful.

Actions/Indications

- Revives consciousness
- Benefits the face and nose, eliminates (external) Wind
- Benefits the spine

Special features

Meeting point with the L.I. and ST channels; Sun Si Miao Ghost point. Major point for acute emergencies: Needle with strongly reducing technique. If no acupuncture needle is available, use a syringe or perform strong acupressure with the nail of the thumb or index finger, supporting the patient's chin with the same hand. Can be used in cases of needle collapse – first remove all other needles.

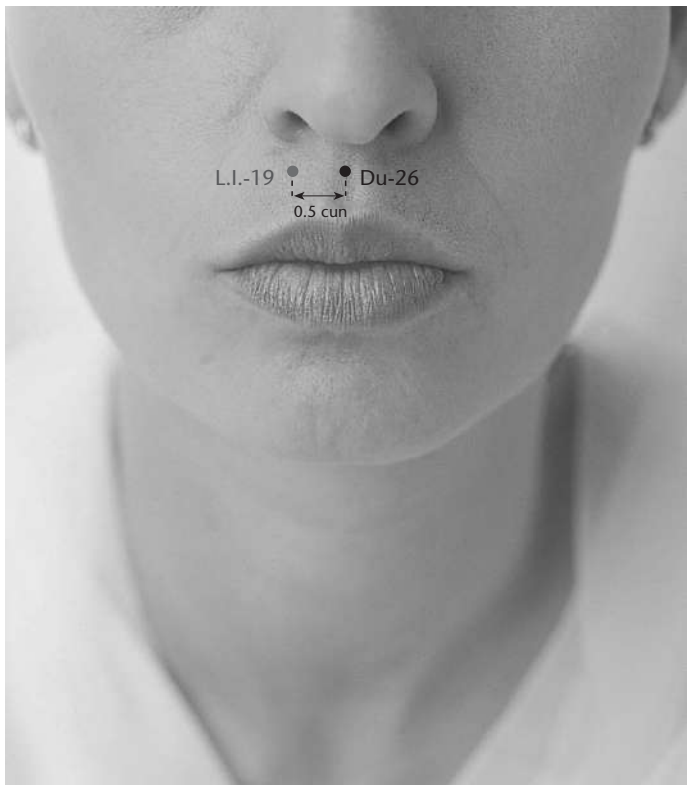

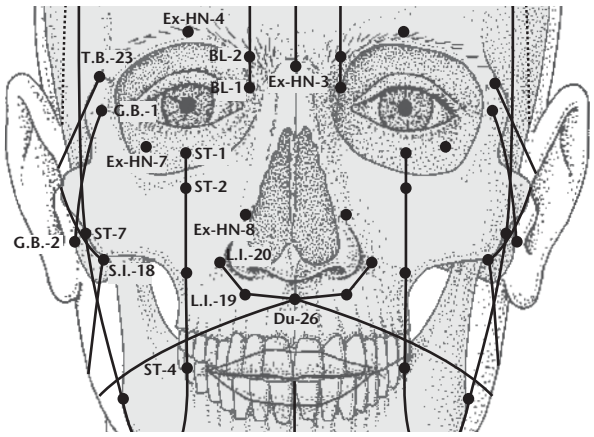

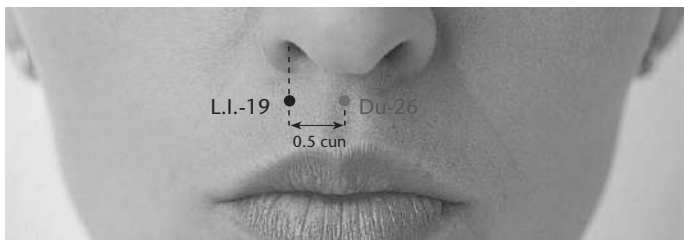

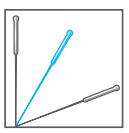
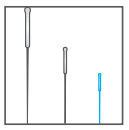

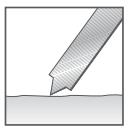

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 525

Extremity of the Mouth DUIDUAN Du-27

Location

On the midline, on the margin of the upper lip and the philtrum.

How to find

Location help: As the name implies, **Du-27** is located on the midline, at the junction of the upper lip and the philtrum.

Needling

Up to 0.3 cun obliquely in a superior direction. No moxibustion.

Actions/Indications

- Clears Heat
- Moistens the body
- Local point for the mouth
- Calms the *shen*

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 526

Du-28 *Gum Intersection* YINJIAO

Location

On the inside of the upper lip, at the junction of the frenulum and the upper gum.

How to find

Hold the upper lip in a raised position in order to reveal the frenulum and the upper gum. Locate **Du-28** on the midline, at the junction of the frenulum with the upper gum.

Needling

0.1–0.2 cun obliquely in a superior direction or prick to bleed. Do not puncture the frenulum.

Actions/Indications

- Clears Heat (especially from the eyes and the mouth)

Special features

As the name implies, meeting point with the *ren mai* and the ST channel

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 527

5.4 Dai mai

Synonym: Girdling Vessel

The extraordinary vessel *dai mai* is the only vessel (or channel) with a horizontal pathway (except for *luo*-connecting vessels).

Relationships (→ 1.7.3)

- **Yin/Yang:** *chong mai/dai mai*
- **Central/peripheral:** *dai mai/yang wei mai*. Regions supported by this pairing: lateral eye region as well as temporal region, ears, cheeks, occiput and shoulders. Opening point: **G.B.-41** (*zulinqi*), coupled point: **T.B.-5** (*waiguan*).

Pathway

The extraordinary vessel *dai mai* originates in the hypochondrium at the level of L2; according to many other authors it originates at → **LIV-13** (*zhangmen*) at the free end of the 11th rib. It encircles the waist of the body like a belt. As it does so it passes → **G.B.-26** (*dai mai*) at the level of the umbilicus and below the free end of the 11th rib, → **G.B.-27** (*wushu*) and → **G.B.-28** (*weidao*).

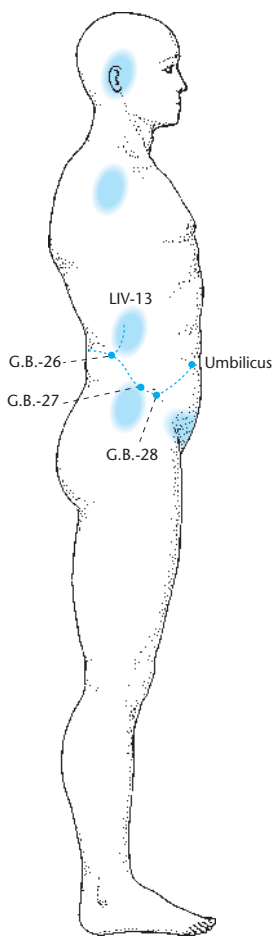

Meeting points with other channels

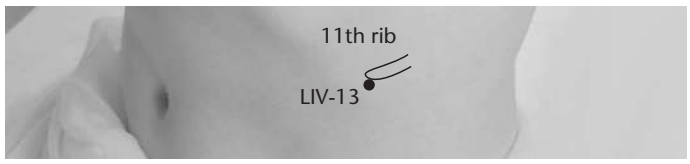

LIV-13 (*zhangmen*): Anterior and inferior to the free end of the 11th rib

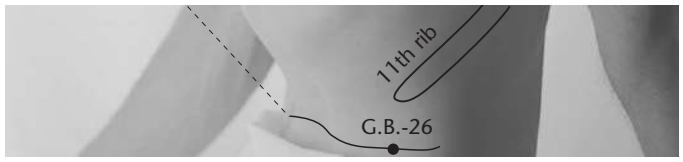

G.B.-26 (*daimai*): On the lateral aspect of the waist, at the junction of a vertical line through the free end of the 11th rib and a horizontal line through the umbilicus, approximately 1.8 cun inferior to → **LIV-13**

G.B.-27 (*wushu*): In the depression medial to the anterior superior iliac spine, approximately level with Ren-4

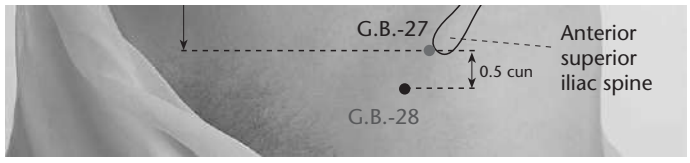

G.B.-28 (*weidao*): Anterior and inferior to the anterior superior iliac spine, 0.5 cun anterior and inferior to G.B.-27

Connections with other channels/Organs

- Gall Bladder and Liver primary channels, Kidney divergent channel
- Kidney, Uterus

Clinical importance (→ 1.7.2, 1.7.3)

- **Controls** and **encircles** the horizontal channels like a belt
- **Connects** the upper and lower halves of the body at the waist
- **Regulates** vaginal discharge
- **Eliminates** Damp-Heat from the Lower Burner, especially from the genitalia
- **Controls** the *shaoyang* channels, especially the G.B. channel

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 528

5.5 *Yin wei mai*

Synonym: Yin Motility Vessel (Deadman), **Yin Heel Vessel** (Maciocia)

Relationships (→ 1.7.3)

- **Yin/Yang:** *yin wei mai/yang wei mai*
- **Central/peripheral:** *chong mai/yin wei mai*. Regions supported by this pairing: Heart, thorax, Stomach. Opening point: **P-6** (*neiguan*), coupled point: **SP-4** (*gongsun*).

Pathway

The extraordinary vessel *yin wei mai* originates in the hypochondrium at → **KID-9** (*zhubin*), 5 cun proximal to the prominence of the medial malleolus, ascends the medial aspect of the leg to the inguinal region, follows the Spleen and Liver primary channels, passing the points → **SP-12** (*chongmen*) and → **SP-13** (*fushe*). It then passes → **SP-15** (*daheng*) and → **SP-16** (*fuai*) and again meets with the Spleen and Liver primary channels at → **LIV-14** (*qimen*). The vessel then traverses the thoracic region, passes → **Ren-22** (*tiantu*) in the suprasternal fossa and follows the extraordinary vessel *ren mai* to → **Ren-23** (*lianqiao*) where it terminates.

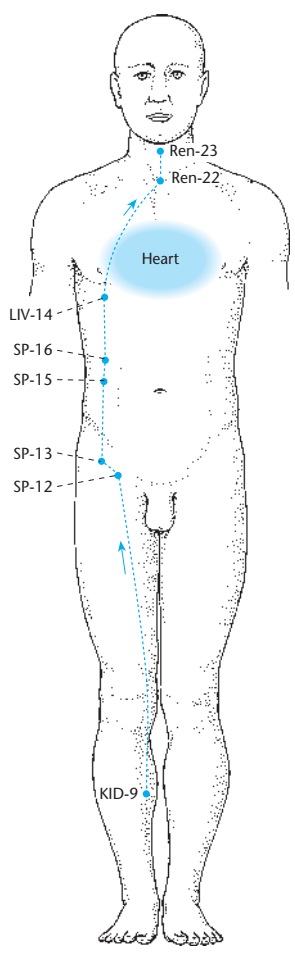

Meeting points with other channels

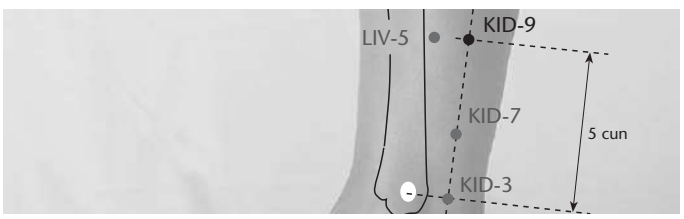

KID-9 (*zhubin*): 5 cun proximal to KID-3 and 2 cun posterior to the medial border of the tibia

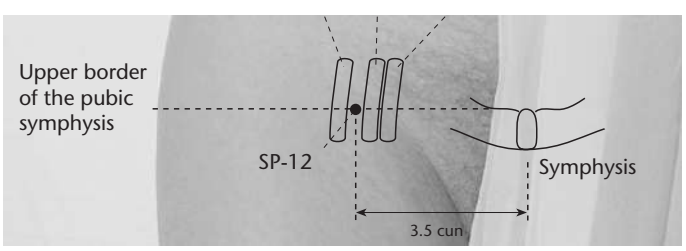

SP-12 (*chongmen*): 3.5 cun lateral to the midline, level with the upper border of the pubic symphysis, lateral to the femoral artery

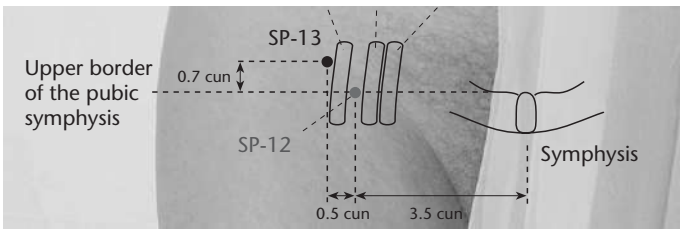

SP-13 (*fushe*): 4 cun lateral to the anterior midline (mamillary line) and 0.7 cun superior to the upper border of the pubic symphysis

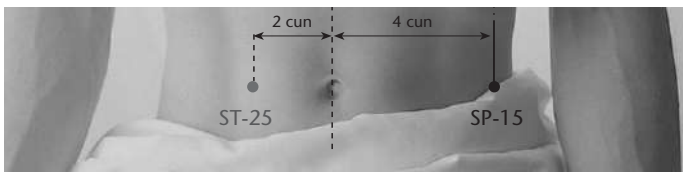

SP-15 (*daheng*): 4 cun lateral to the centre of the umbilicus

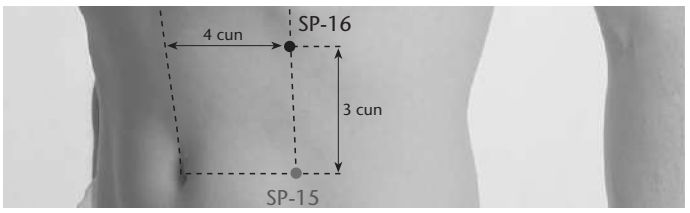

SP-16 (*fuai*): 3 cun superior to the centre of the umbilicus and 4 cun lateral to the anterior midline, on the mamillary line

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 529

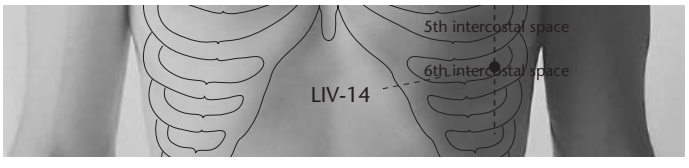

LIV-14 (*qimen*): In the 6th intercostal space, on the mamillary line, 4 cun lateral to the midline.

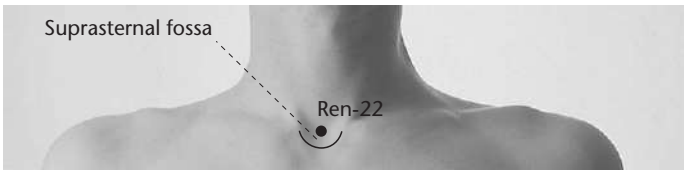

Ren-22 (*tiantu*): 0.5 cun superior to the sternum, in the centre of the suprasternal fossa

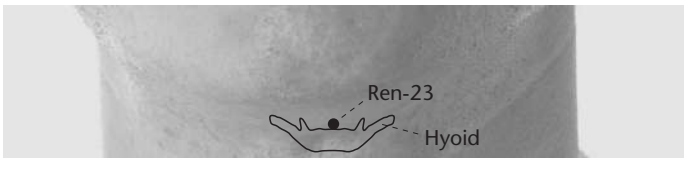

Ren-23 (*lianquan*): On the anterior midline, superior to the upper border of the hyoid bone

Connections with other channels/organs

- Kidney, Spleen and Liver primary channels, *ren mai*
- Internal organs

Clinical importance (→ 1.7.2, 1.7.3)

- **Connects** and **regulates** all **Yin channels** of the body. It connects the KID, SP and LIV channels and the *ren mai* and dominates the **Interior** of the body (the internal organs).
- **Strengthens** Yin and Heart-Blood, especially in women
- **Balances** the emotions

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 530

5.6 *Yang wei mai*

Synonym: Yang Motility Vessel (Deadman), **Yang Heel Vessel** (Maciocia)

Relationships (→ 1.7.3)

Yin/Yang: *yin wei mai/yang wei mai*
Central/peripheral: *dai mai/yang wei mai*. Regions supported by this pairing: Lateral eye region, temporal region, ears, cheeks, neck and shoulders. Opening point: **T.B.-5** (*waiguan*), coupled point: **G.B.-41** (*zulinqi*).

Pathway

The extraordinary vessel *yang wei mai* originates at → **BL-63** (*jinmen*) in the depression posterior to the tuberosity of the 5th metatarsal bone, runs anterior to the lateral malleolus, passes → **G.B.-35** (*yangjiao*), ascends the lateral aspect of the leg, passes the hip, ascends the posterior aspect of the costal region to the shoulder, passes → **S.I.-10** (*naoshu*), → **T.B.-15** (*tianliao*), → **G.B.-21** (*jianjing*) and ascends the neck. The vessel then ascends anterior to the ear (according to some authors, for example Solinas et al. (1998), posterior to the ear) to the forehead, passing → **ST-8** (*touwei*), → **G.B.-13** (*benshen*) and → **G.B.-14** (*yangbai*). It then runs with the G.B. channel from **G.B.-15** to **G.B.-20** on the occiput, passing → **G.B.-15** (*toulinqi*), → **G.B.-16** (*muchuang*), → **G.B.-17** (*zhengying*), → **G.B.-18** (*chengling*), → **G.B.-19** (*naokong*) and → **G.B.-20** (*fengchi*). The vessel then continues from **G.B.-20** to the posterior midline, passing → **Du-16** (*fengfu*) and → **Du-15** (*yamen*), where the vessel terminates.

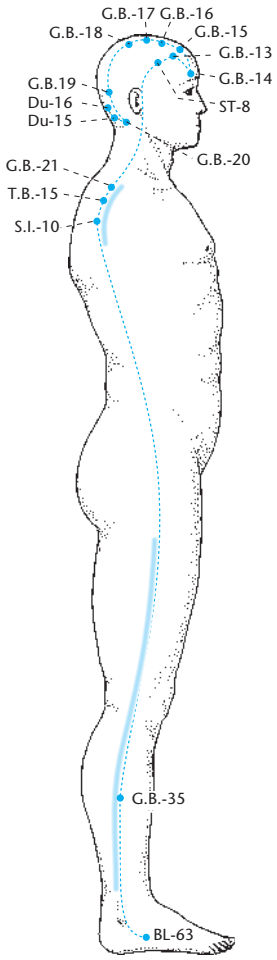

Meeting points with other channels

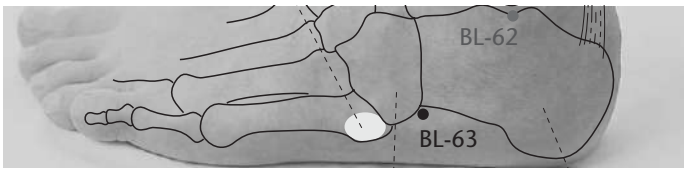

BL-63 (*jinmen*): In a depression anterior and inferior to → **BL-62** between the calcaneus and the cuboid bone, proximal to the tuberosity of the 5th metatarsal bone

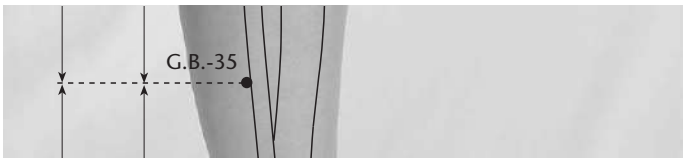

G.B.-35 (*yangjiao*): 7 cun proximal to the highest prominence of the lateral malleolus, on the posterior border of the fibula

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 531

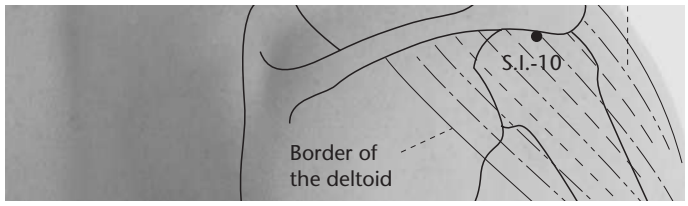

S.I.-10 (*naoshu*): With the arm adducted on a line extending in a superior direction from the posterior axillary fold, on the lower border of the scapular spine

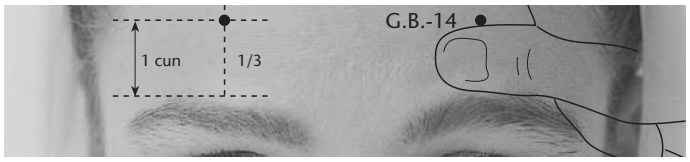

G.B.-14 (*yangbai*): When looking straight ahead on the pupil line, approximately 1 cun superior to the midpoint of the eyebrow, at the junction of the eminence and the superciliary arch

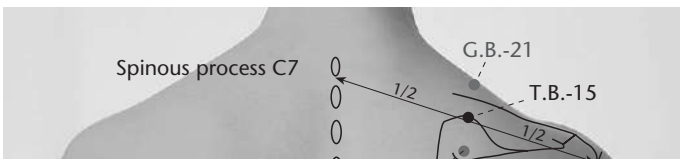

T.B.-15 (*tianliao*): At the midpoint of an imaginary line between the spinous process of the C7 and the lateral extremity of the acromion

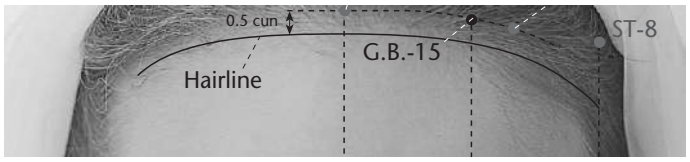

G.B.-15 (*toulinqi*): When looking straight ahead superior to the pupil, 0.5 cun superior to the anterior hairline

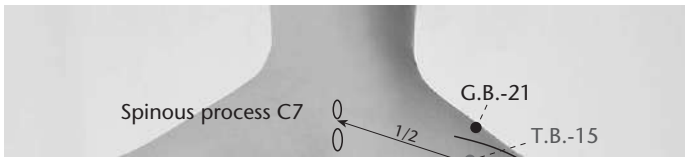

G.B.-21 (*jianjing*): At the highest point of the shoulder, midway between the lower border of the spinous process of C7 and the lateral extremity of the acromion

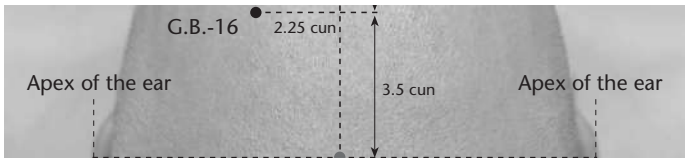

G.B.-16 (*muchuang*): 1.5 cun superior to the anterior hairline, on the pupil line, or 2.25 cun lateral to the midline (= midway between → **Du-24** and → **ST-8**)

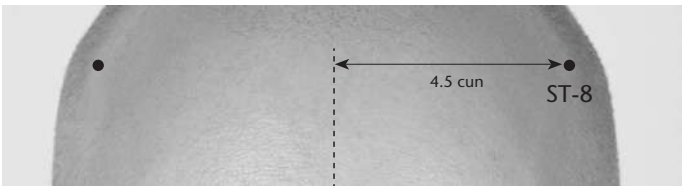

ST-8 (*touwei*): At the temporal corner of the forehead, on the border of the temporalis muscle and 0.5 cun within the anterior hair line or 4.5 cun lateral to the anterior midline (Du-24)

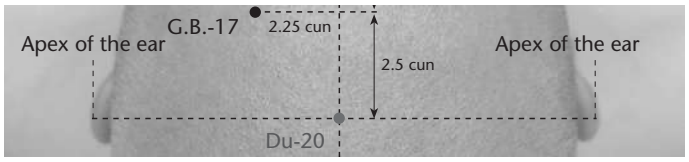

G.B.-17 (*zhengying*): 2.5 cun superior to the anterior hairline and 2.25 cun lateral to the midline

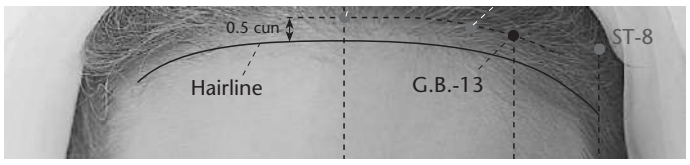

G.B.-13 (*benshen*): 3 cun lateral to → **Du-24** (on the midline, 0.5 cun superior to the anterior hairline)

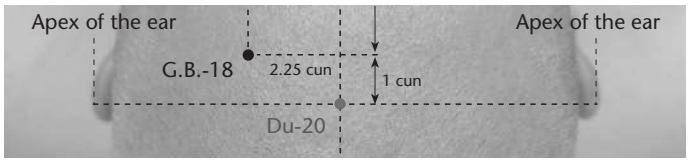

G.B.-18 (*chengling*): 4 cun superior to the anterior hairline or 1 cun anterior to → **Du-20**, 2.25 cun lateral to the midline

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 532

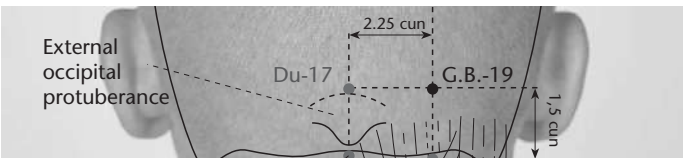

G.B.-19 (*naokong*): On the occiput, level with the upper border of the external occipital protuberance (→ **Du-17**), 2.25 cun lateral to the midline

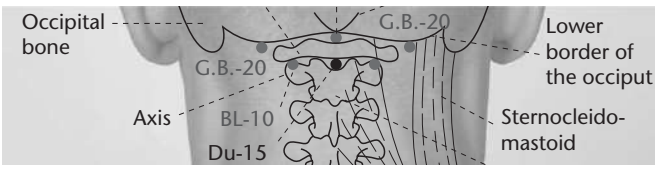

Du-15 (*yamen*): On the occiput, on the posterior midline, in the depression between the 1st (atlas) and 2nd cervical vertebrae, approximately 0.5 cun below → **Du-16**

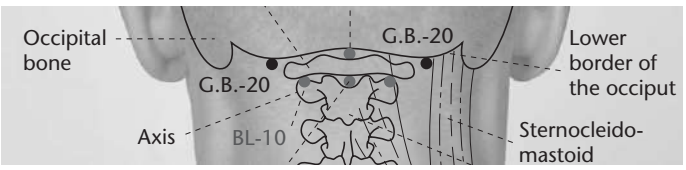

G.B.-20 (*fengchi*): On the lower border of the occipital bone, in the depression between the origins of the sternocleidomastoid and trapezius

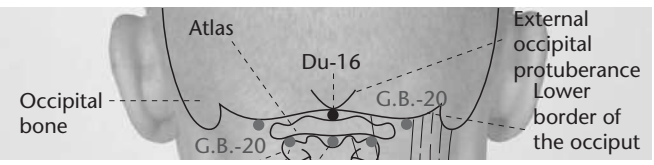

Du-16 (*fengfu*): On the posterior midline, directly below the external occipital protuberance, in the depression between the origins of the trapezius

Connections with other channels/Organs

- BL, G.B., ST, S.I., T.B. and L.I. primary channels, *du mai*

Clinical importance

- **Connects** and **regulates** all **Yang channels** of the body. It connects the BL, G.B., ST, S.I., T.B. and L.I. channels as well as the *du mai* and controls the Exterior of the whole body (especially the *taiyang* and *shaoyang* channels)
- Harmonises the Nutritive Qi (*ying qi*) and the Defensive Qi (*wei qi*)
- Expels Exterior Wind-Cold, especially with *shaoyang*-syndrome.

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 533

5.7 Yin qiao mai

Synonym: Yin Motility Vessel (Deadman), **Yin Heel Vessel** (Maciocia)

Relationships (→ 1.7.2, 1.7.3)

- **Yin/Yang:** *yin qiao mai/yang qiao mai*
- **Central/peripheral:** *ren mai/yin qiao mai*. Regions supported by this pairing: Face, throat, thorax, Lung, diaphragm, abdomen. Opening point: **KID-6** (*zhaohai*), coupled point: **LU-7** (*lieque*).

Pathway

The extraordinary vessel *yin qiao mai* originates at → **KID-2** (*rangu*) inferior to the navicular tuberosity (according to some authors it only begins at → **KID-6**), runs to → **KID-6** (*zhaohai*) and → **KID-8** (*jiaoxin*), ascends the posteromedial aspect of the leg to the external genital region, ascends the abdomen to the thorax, crosses the supraclavicular fossa and continues to the throat and face. It then runs to the inner canthus of the eye where it connects with the Bladder primary channel and the *yang qiao mai* at → **BL-1** (*jingming*). It then ascends to enter the Brain.

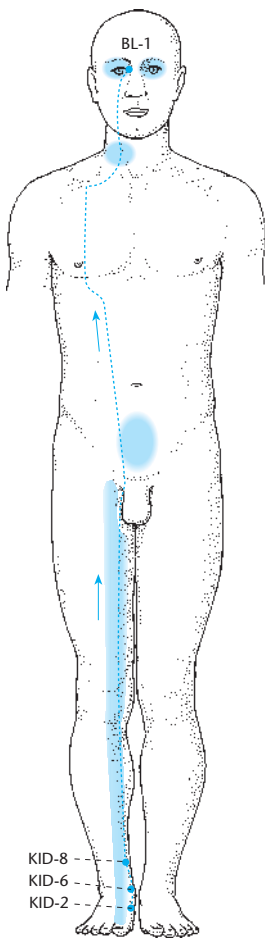

Meeting points with other channels

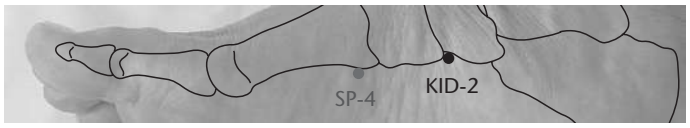

KID-2 (*rangu*): In a depression at the anterior/inferior border of the navicular bone, at the border of the 'red and white' skin

KID-6 (*zhaohai*): In the depression inferior to the highest prominence of the medial malleolus, over the joint space between the talus and calcaneus

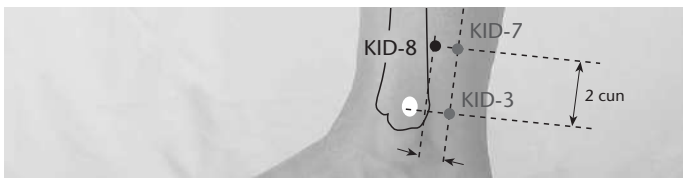

KID-8 (*jiaoxin*): 2 cun proximal to the highest prominence of the medial malleolus, posterior to the border of the tibia

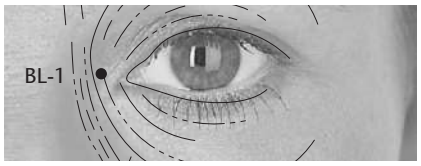

BL-1 (*jingming*): In a depression 0.1 cun superior and medial to the inner canthus of the eye

Connections with other channels/Organs

- Kidney and Bladder primary channels
- Brain

Clinical importance (→ 1.7.2, 1.7.3)

- Together with the *yang qiao mai* controls the musculature of the lower extremities
- Together with the *yang qiao mai* regulates the opening and closing of the eyes
- Regulates the functioning of the Brain
- Removes stagnation (of Qi, Blood or Dampness) in the Lower Burner, especially in the Uterus in women

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 534

5.8 *Yang qiao mai*

Synonym: Yang Motility Vessel (Deadman), **Yang Heel Vessel** (Maciocia)

Relationships (→ 1.7.3)

- **Yin/Yang:** *yin qiao mai/yang qiao mai*
- **Central/peripheral:** *du mai/yin qiao mai*. Regions supported by this pairing: Inner canthus of the eye, occiput, shoulders and back. Opening point: **BL-62** (*shenmai*), coupled point: **S.I.3** (*houxi*)).

Pathway

The extraordinary vessel *yang qiao mai* originates at → **BL-62** (*shenmai*) below the prominence of the lateral malleolus, curves briefly around the malleolus to → **BL-61** (*pucan*) and ascends anterior to the Achilles tendon to → **BL-59** (*fuyang*), travels up the lateral aspect of the leg, crosses the hip, passing → **G.B.-29** (*juliao*), ascends the flanks and the posterior aspect of the shoulders, passes → **S.I.10** (*naoshu*), continues to → **L.I.-15** (*jianyu*) and runs in a curve to → **L.I.-16** (*jugu*), crosses the supraclavicular fossa, ascends the throat, according to some authors passing → **ST-9** (*renying*), and reaches the face. Here the vessel passes the points → **ST-4** (*dicang*) and → **ST-3** (*juliao*); according to some authors also → **ST-2** (*sibai*) and → **ST-1** (*chengqi*) in the infraorbital region. It then reaches → **BL-1** (*jingming*), where it meets the extraordinary vessel *yin qiao mai*, ascends the forehead to the vertex and descends posteriorly to → **G.B.-20** (*fengchi*); according to the *Nan Jing* and the *Nei Jing* it also reaches → **Du-16** (*fengfu*), where it enters the Brain.

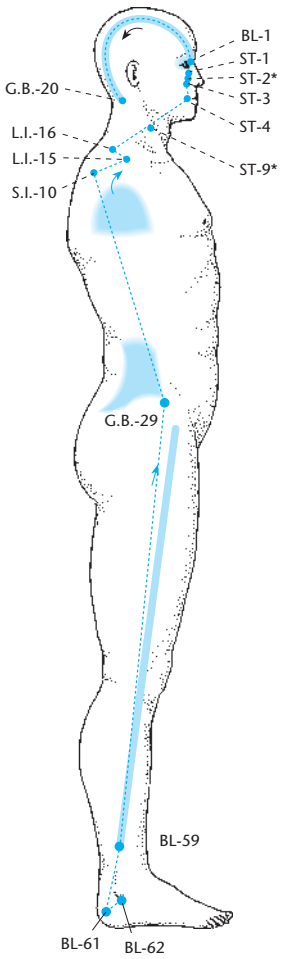

* According to Deadman et al 1998.

Intersection points with other channels

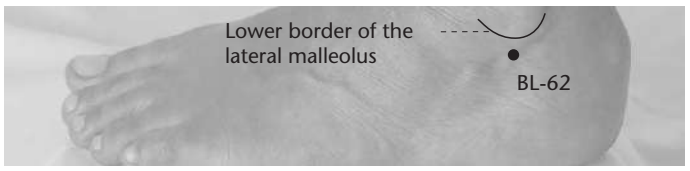

BL-62 (*shenmai*): In a depression directly inferior to the highest prominence of the lateral malleolus, over the joint space between the talus and calcaneus

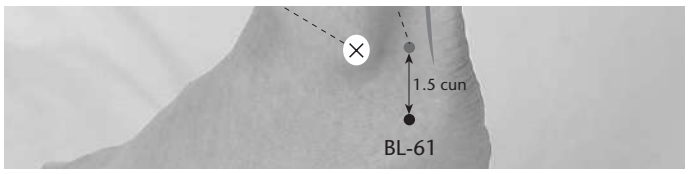

BL-61 (*pucan*): In a depression on the calcaneus, approximately 1.5 cun inferior to → **BL-60**

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 535

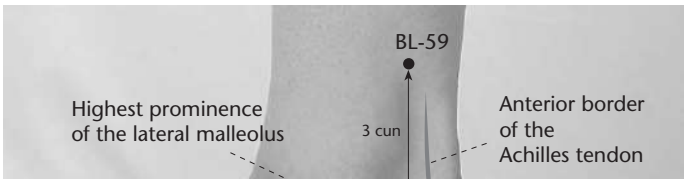

BL-59 (*fuyang*): 3 cun superior to → BL-60 (in the depression between the highest prominence of the lateral malleolus and the Achilles tendon)

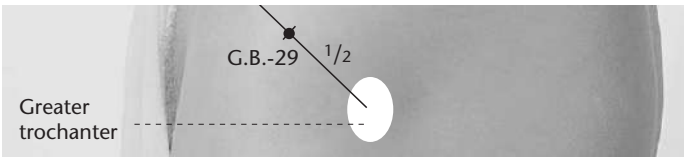

G.B.-29 (*juliao*): At the midpoint of a line connecting the anterior superior iliac spine and the greater trochanter, at the anterior border of the iliac crest

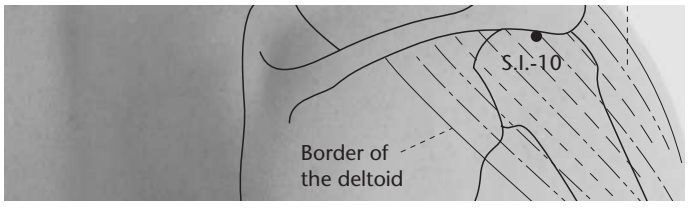

S.I.-10 (*naoshu*): With the arm adducted on a line extending in a superior direction from the posterior axillary fold, on the lower border of the scapular spine

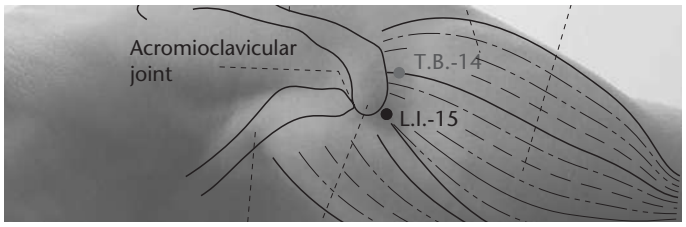

L.I.-15 (*jianyu*): In the depression distal and anterior to the acromion, between the clavicular and acromial portions of the deltoid muscle

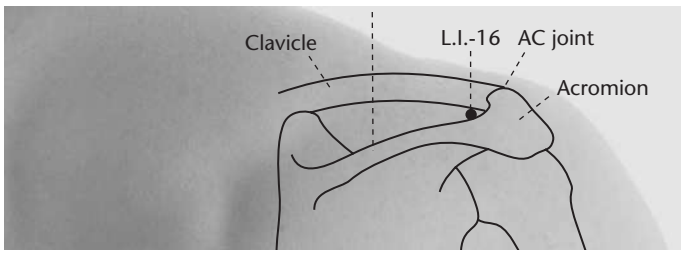

L.I.-16 (*jugu*): In a depression between the acromial extremity of the clavicle and the junction of the scapular spine and the acromion

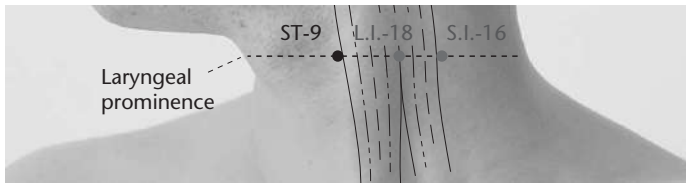

ST-9 (*renying*): Approximately 1.5 cun lateral to the anterior midline, level with the laryngeal prominence, at the anterior border of the sternocleidomastoid

ST-4 (*dicang*): When looking straight ahead on the pupil line, approximately 0.4 cun lateral to the corner of the mouth

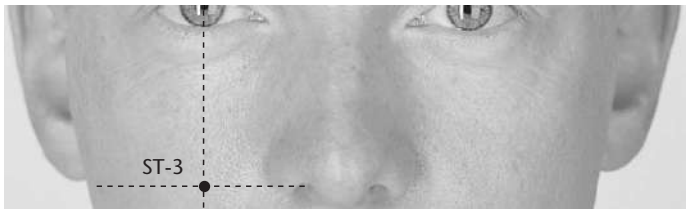

ST-3 (*juliao*): When looking straight ahead, directly below the centre of the pupil, level with the lower border of the ala nasi

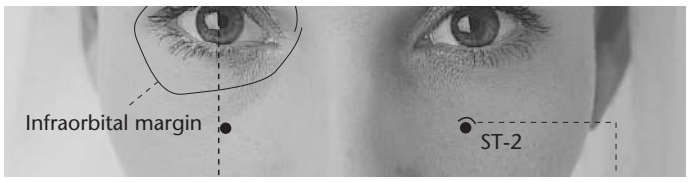

ST-2 (*sibai*): When looking straight ahead, directly below the centre of the pupil, in the depression at the infraorbital foramen

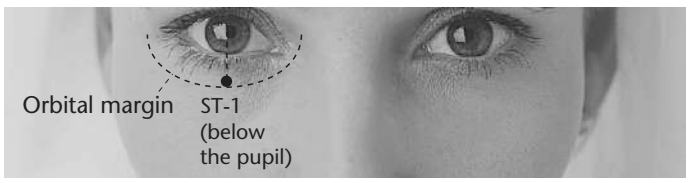

ST-1 (*chengqi*): With the eyes looking straight forward, directly below the centre of the pupil, between the eyeball and the infraorbital ridge

Ch05b-F10028.qxd 2/22/08 9:56 PM Page 536

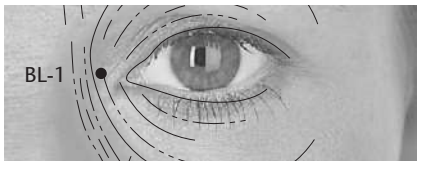

BL-1 (*jingming*): In a depression 0.1 cun superior and medial to the inner canthus of the eye

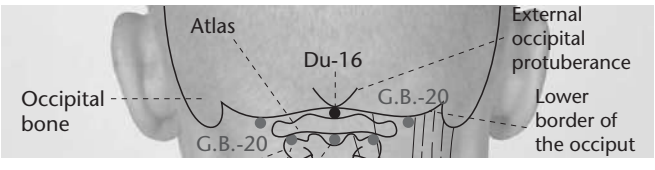

Du-16 (*fengfu*): On the posterior midline, directly below the external occipital protuberance, in the depression between the origins of the trapezius

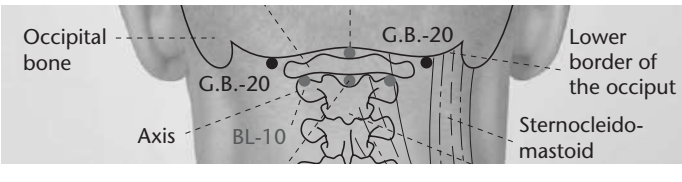

G.B.-20 (*fengchi*): On the lower border of the occipital bone, in the depression between the origins of the sternocleidomastoid and trapezius

Connections with other channels/Organs

- Kidney, Bladder, Gall Bladder, Stomach, Large Intestine and Small Intestine primary channels
- Brain

Clinical importance (→ 1.7.2, 1.7.3)

- Together with the *yin qiao mai* controls the musculature of the lower extremities
- Together with the *yang qiao mai* regulates the opening and closing of the eyes
- Regulates the functioning of the Brain, eliminates internal and external Wind from the head
- Removes obstructions and stagnation from the spine, especially after traumatic injuries

6 Extra Points

Claudia Focks, Ulrich März

Besides the 361 classic acupuncture points located on the channels there are a number of **Ex**tra Points (**Ex**), located with a few exceptions outside the pathways of the primary channels. In 1991, a new nomenclature for these extra points was confirmed in the People's Republic of China, listing 48 points (*The Location of Acupoints, State Standard of the People's Republic of China*, Foreign Languages Press, Beijing, 1990).

In the past, authors have developed ways of identifying the extra points by names or numbers:

- Nguyen Van Nghi; König and Wancura: '**P**unkte **a**ußerhalb der **M**eridiane' (points not on meridians) (PaM) and '**N**eu-**P**unkte' (new points) (NP); Schnorrenberger (wall charts translated from the Chinese language) uses the same numbering system as Nguyen Van Nghi but refers to the **PaM** as '**Z**usatz**p**unkte' (additional points) (ZP) and also as '**N**eu-**P**unkte' (NP)
- Shanghai College for Traditional Medicine (*Acupuncture – a Comprehensive Text*, English translation by O'Connor J, Bensky D, which also forms the foundation for Deadman et al. 1998 and Ellis and Wiseman 1991): **M**iscellaneous (M) and **N**ew (N) Points
- Hempen 1995: '**E**xtrapunkte' (extra points) (Ex).

The following tables give an overview of the most commonly used extra points also listed in this atlas.

6.1 Extra Points: Head and Neck (EX-HN)

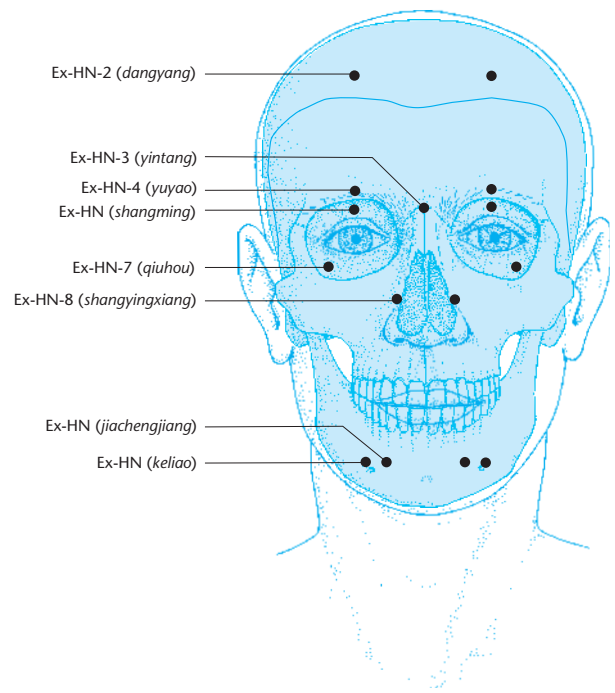

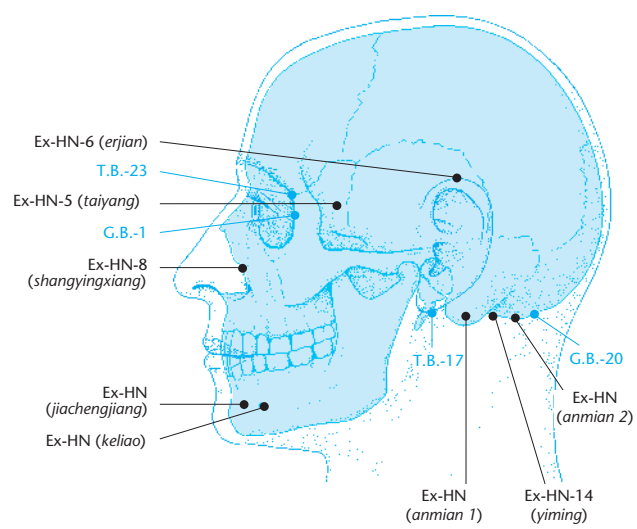

Ch06a-F10028.qxd 2/23/08 1:24 PM Page 538

English abbreviation (standard)	*Pinyin* name	Nguyen Van Nghi, König and Wancura, Schnorrenberger	Shanghai College	Ex (Hempen)
Ex-HN-1	*sishencong*	PaM or ZP1	M-HN-1	Ex6
Ex-HN-2	*dangyang*		M-HN	
Ex-HN-3	*yintang*	PaM or ZP3	M-HN-3	Ex1
Ex-HN-4	*yuyao*	PaM or ZP6	M-HN-6	
Ex-HN-5	*taiyang*	PaM or ZP9	M-HN-9	Ex2
Ex-HN-6	*erjian*	PaM or ZP10	M-HN-10	
Ex-HN-7	*qiuhou*	PaM or ZP8	M-HN-8	
Ex-HN-8	*shangyingxiang/bitong*	NP12 (PaM or ZP14*)	M-HN-14	Ex3
Ex-HN-9	*neiyingxiang*		M-HN-35	
Ex-HN-10	*juquan*		M-HN-36	
Ex-HN-11	*haiquan*		M-HN-37	
Ex-HN-12	*jinjin*†	PaM or ZP20	M-HN-20	
Ex-HN-13	*yuye*†	PaM or ZP20	M-HN-20	
Ex-HN-14	*yiming*	PaM or ZP13	M-HN-13	Ex4
Ex-HN-15	*jing(bailao)*	PaM or ZP30	M-HN-30	
Further extra points				
Ex-HN	*shangming*		N-HN-4	
Ex-HN	*anmian*‡	NP27 and 28 (*anmian*‡1 und 2)	N-HN-54 M-HN-54 (Deadman)	Ex5
Ex-HN	*jiabi*	PaM or ZP15		
Ex-HN	*jiachengjiang*	PaM or ZP18 (*heliao/keliao*)	M-HN-18	Ex7
Ex-HN	*chonggu/zhuidong*	PaM or ZP31	M-HN-31	
Ex-HN	*jingbi*		M-HN-41	

* Nguyen Van Nghi and Schnorrenberger describe a different location for the point with this *pinyin* name (location: 0.5 cun inferior to the inner canthus of the eye); the 'standard' location is described under NP12 as '*bitong*' or '*bicong*'.

† The Shanghai nomenclature and also Nguyen Van Nghi, König and Wancura and Schnorrenberger describe the extra points *jinjin* and *yuye* as a pair of points.

‡ There exists differing information regarding the extra point *anmian*; for example the Shanghai College (and Wiseman) describe *anmian* as N-HN-54 (in Deadman: M-HN-54) with the following location: midway between G.B.-20 and T.B.-17; Nguyen Van Nghi and Schnorrenberger describe *anmian* as two new points with the following locations: *anmian* 1 is located midway between T.B.-17 and *yiming* (PaM13 or Ex-HN-14), *anmian* 2 midway between G.B.-20 and *yiming* (*anmian* according to the Shanghai College is shown on the point page; the location of *anmian* 1 and 2 according to Nguyen and Schnorrenberger in figure 6.1).

Ch06a-F10028.qxd 2/23/08 1:24 PM Page 539

Four Alert Spirit SISHENCONG Ex-HN-1

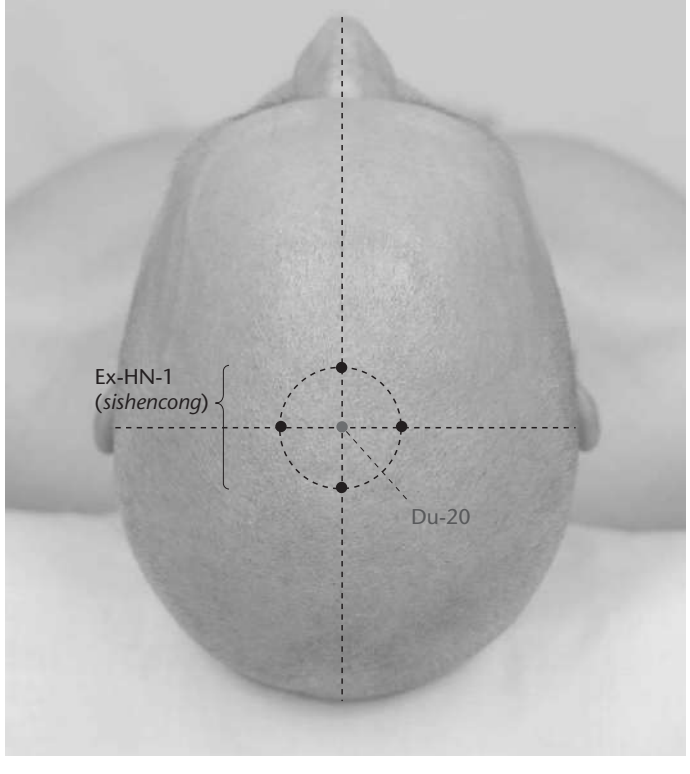

Location

A group of four points, each located 1 cun from ➞ **Du-20** (anterior, posterior and lateral) (➞ 3.1.1).

How to find

First, locate ➞ **Du-20** on the highest point of the head, at the junction of a line connecting the apices of the ears and the vertical midline. Alternative location method: Locate ➞ **Du-20** 5 cun superior to the anterior and 7 cun superior to the posterior hairline. **Ex-HN-1** (*sishencong*) forms a star-shaped group of points around ➞ **Du-20**: two points are located at a distance of 1 cun each on the *du mai*, whereas the other two are located 1 cun each from ➞ **Du-20** in a lateral direction.

Needling

0.5–1 cun obliquely towards ➞ **Du-20**

Actions/Indications

- Calm the *shen*
- Alleviate pain
- Calm Wind
- Benefit the eyes and ears

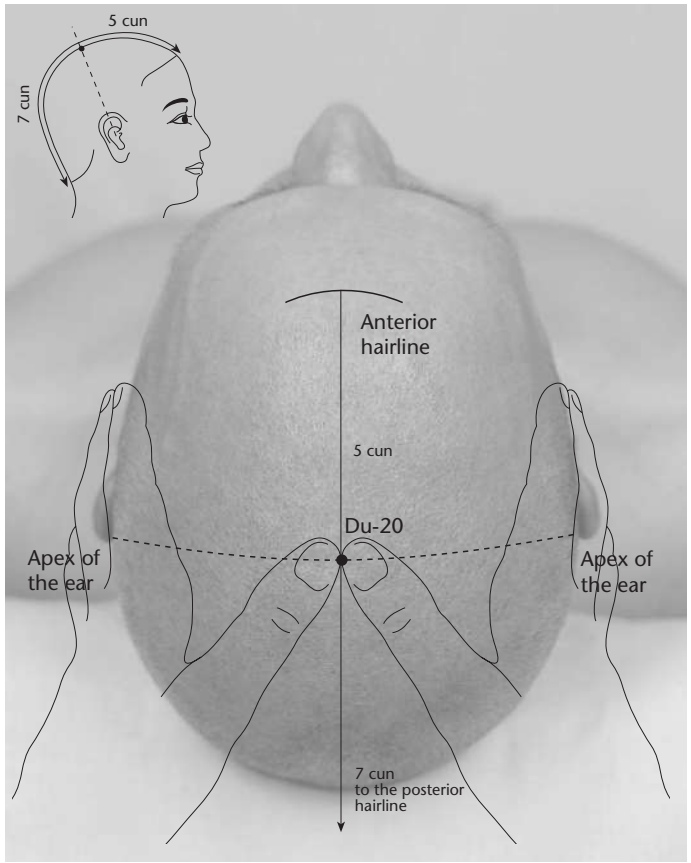

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 540

Ex-HN-2 *Above the Yang* DANGYANG

Location

With the patient looking straight ahead, vertically above the pupil, 1 cun above the anterior hairline.

How to find

In patients with a receding hairline, the original hairline can often be recognized by a change in skin texture. Location help (→ 3.1.1): Ask the patient to frown, as this often reveals the site of the original hairline. Ask the patient to look straight ahead, then locate **Ex-HN-2** by measuring 1 cun from the hairline, on a vertical line through the pupil.

→ **G.B.-14** is also located on the pupil line (on the lower border of the frontal eminence). Located on the same level (1 cun above the hairline) are → **BL-5** (more medial, 1.5 cun from the anterior midline) and → **Du-23** (on the anterior midline).

Needling

0.5 cun transversely (subcutaneously) towards the site of the disorder/pain.

Actions/Indications

- Dispels Wind and Heat, alleviates pain

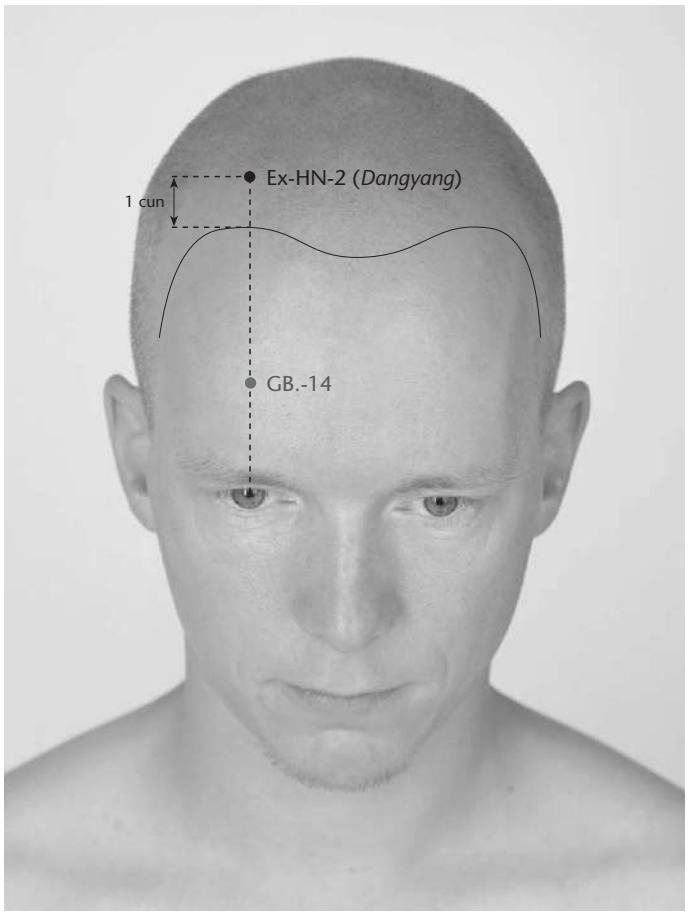

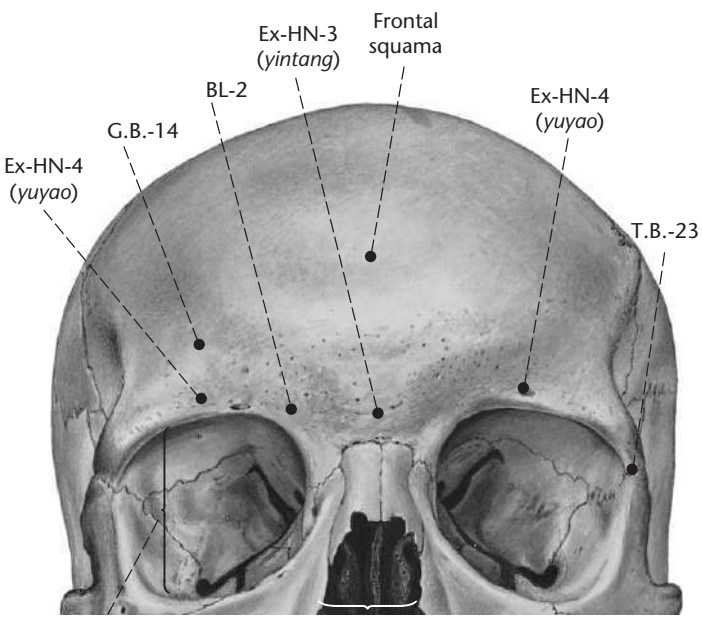

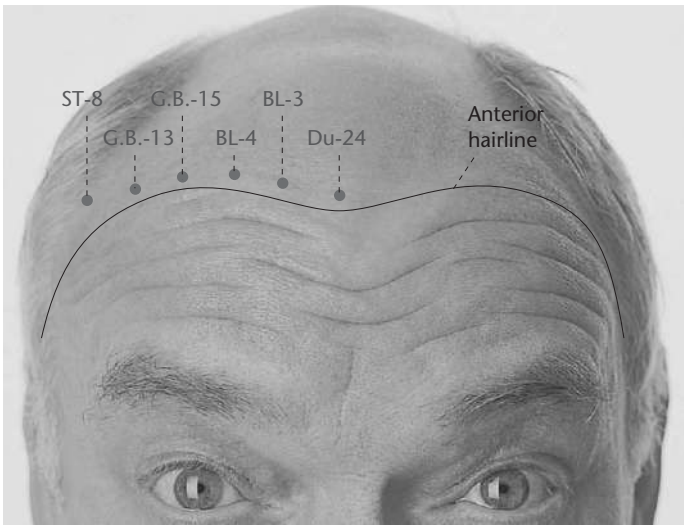

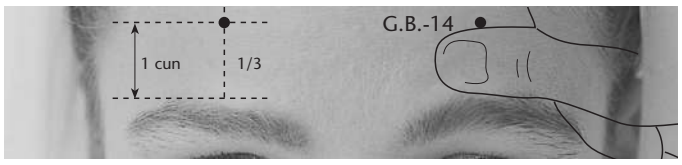

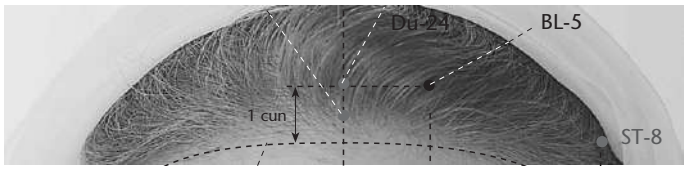

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 541

Hall of Impression YINTANG Ex-HN-3

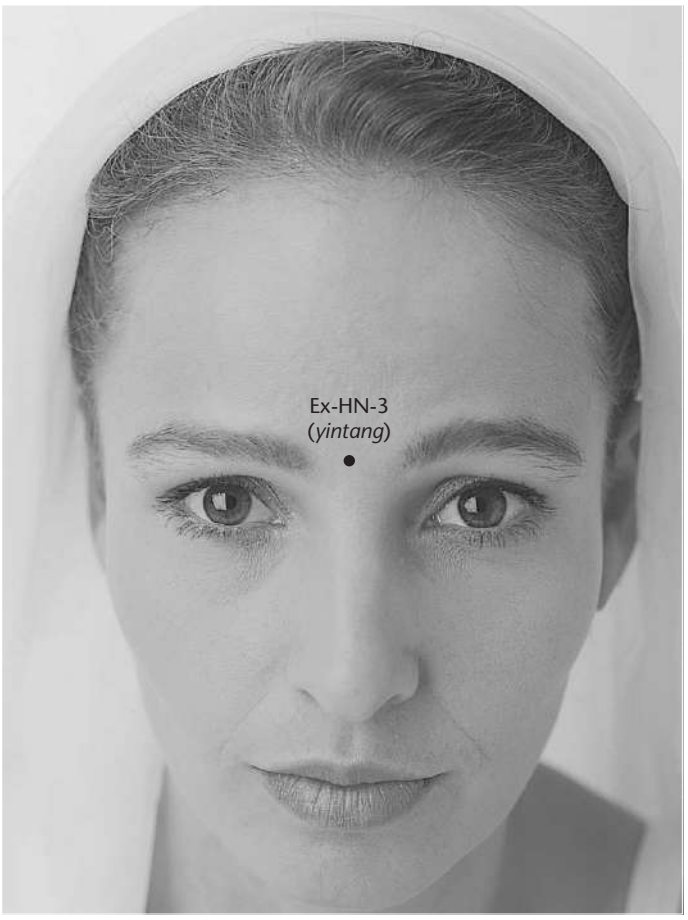

Location

On the anterior midline (*du mai*), between the eyebrows.

How to find

Locate the glabella (→ 3.1.1), a smooth surface on the frontal bone, directly above the root of the nose, between the superciliary arches. Then locate **Ex-HN-3** (*yintang*) in its centre, on the midline and between the medial extremities of the eyebrows. Located along the superciliary arches (from medial to lateral) are → **BL-2,** → **Ex-HN-4** (*yuyao*) and → **T.B.-23**.

Needling

Pinching-skin method: With the thumb and index finger, form a skin fold above the point and insert the needle 0.3–0.5 cun transversely from superior to inferior into the fold, towards the root of the nose. Alternative methods: Oblique or transverse (subcutaneous) insertion towards the eyebrows (→ **BL-2**) or prick to bleed.

Actions/Indications

- Calms the *shen*
- Benefits the nose
- Eliminates (internal) Wind
- Opens the channel, alleviates pain

Special features

Commonly used extra point with a harmonising effect

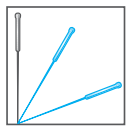

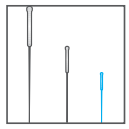

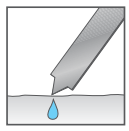

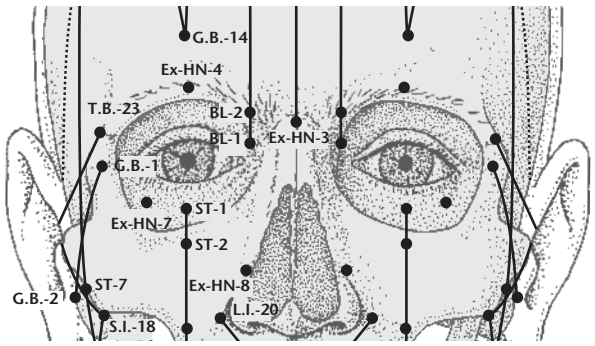

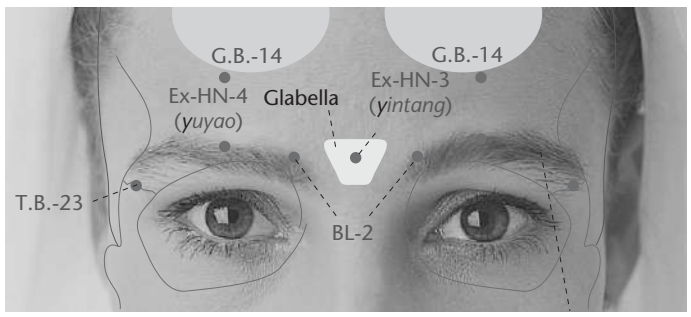

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 542

Ex-HN-4 *Fish Waist* YUYAO

Location

In the centre of the eyebrow, with the eyes looking straight ahead directly above the pupil.

How to find

With the patient looking straight ahead, palpate the eyebrow directly above the pupil, superior to the supraorbital ridge. There, locate **Ex-HN-4** (*yuyao*) in a small depression on the superciliary arch.

Needling

0.3–0.5 cun vertically or obliquely towards the eye.

Actions/Indications

- Benefits the eyes
- Dispels pain

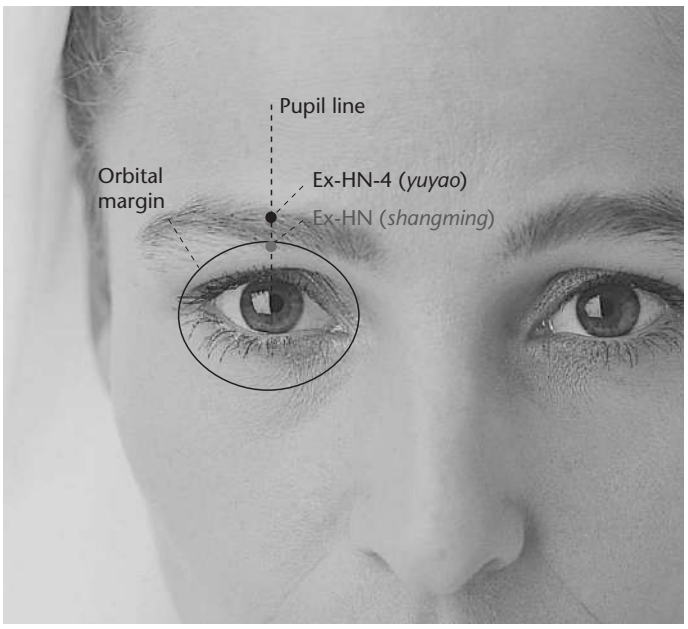

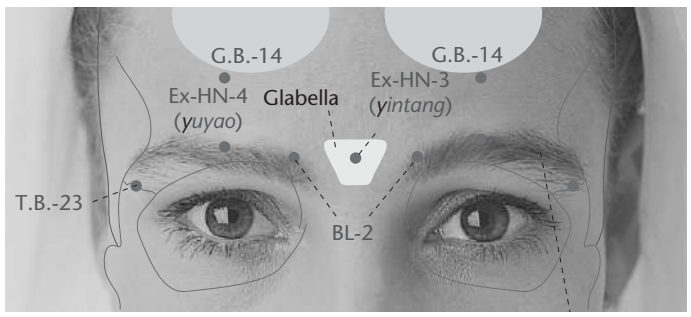

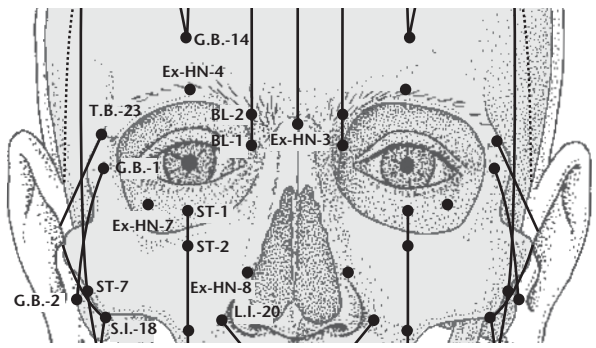

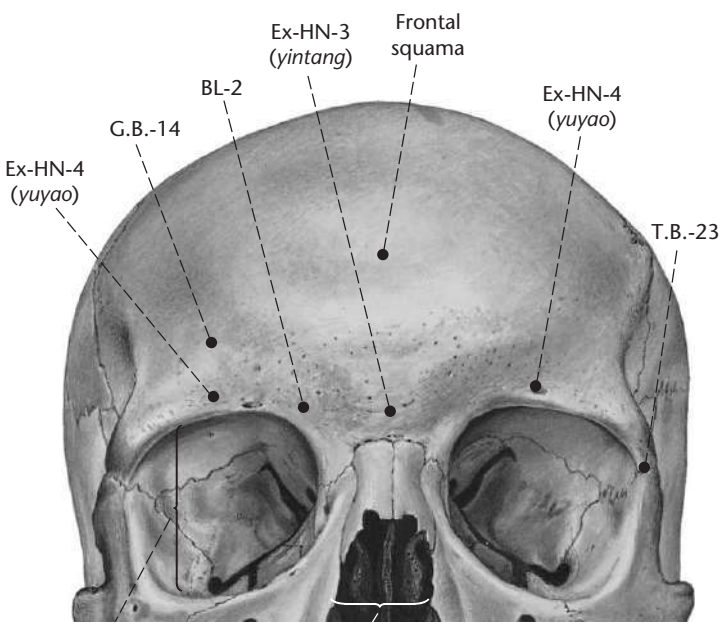

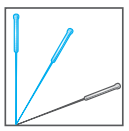
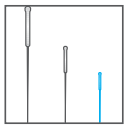

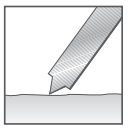

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 543

Supreme Yang TAIYANG Ex-HN-5

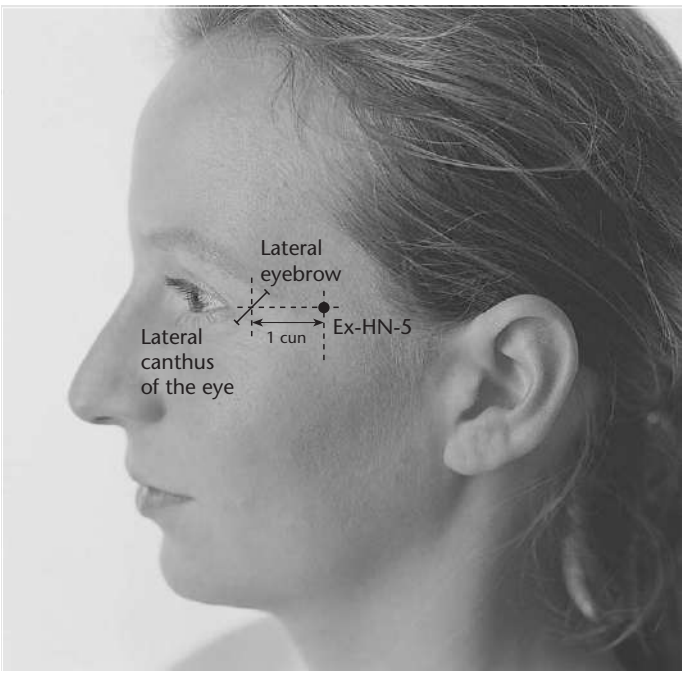

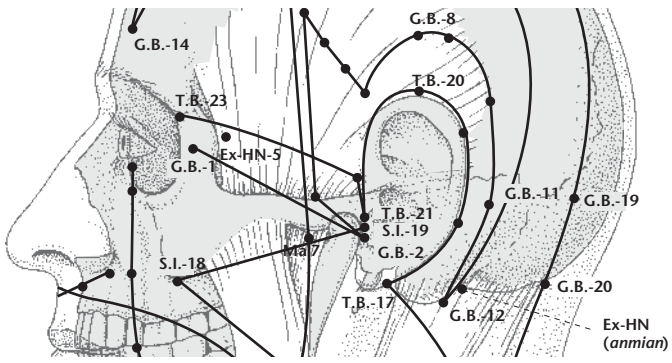

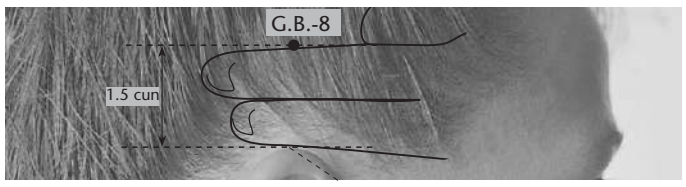

Location

On the temple, in a depression approximately 1 cun lateral to the midpoint of a line connecting the lateral extremity of the eyebrow and the outer canthus of the eye.

How to find

On the temple, first, draw a line between the lateral extremity of the eyebrow and the outer canthus of the eye. Next, palpate from the midpoint of this line in a lateral direction, until you can feel a distinct depression in the temporal bone. **Ex-HN-5** (*taiyang*) is located in the centre of this depression. This point tends to be pressure-sensitive, especially with temporal headaches. Patients often get relief by using acupressure on this point.

Needling

0.5–0.6 cun vertical insertion into the temporal muscle or obliquely in a lateral direction or transversely (subcutaneously) towards → **G.B.-8.** This point often bleeds when needled. Prick to bleed to drain Heat.

Actions/Indications

- Eliminates Wind, clears Heat, reduces swellings, opens the channel, alleviates pain

Special features

One of the most important and most commonly used extra points, especially with headaches

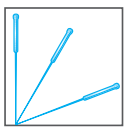
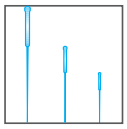

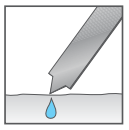

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 544

Ex-HN-6 *Tip of the Ear* ERJIAN

Location

On the apex of the ear, on the helix.

How to find

Fold the ear anteriorly, so that the posterior portion of the upper helix covers its anterior part. The point is located at the highest point of the fold. This is also the location of the auricular point **78**.

Needling

Vertically 0.1–0.2 cun. Some authors advise to bleed this point for Heat conditions, but this is not recommended owing to the danger of causing an otic haematoma. Indirect moxibustion for vision disorders.

Actions/Indications

- Clears Heat, alleviates pain, benefits the eyes and the throat

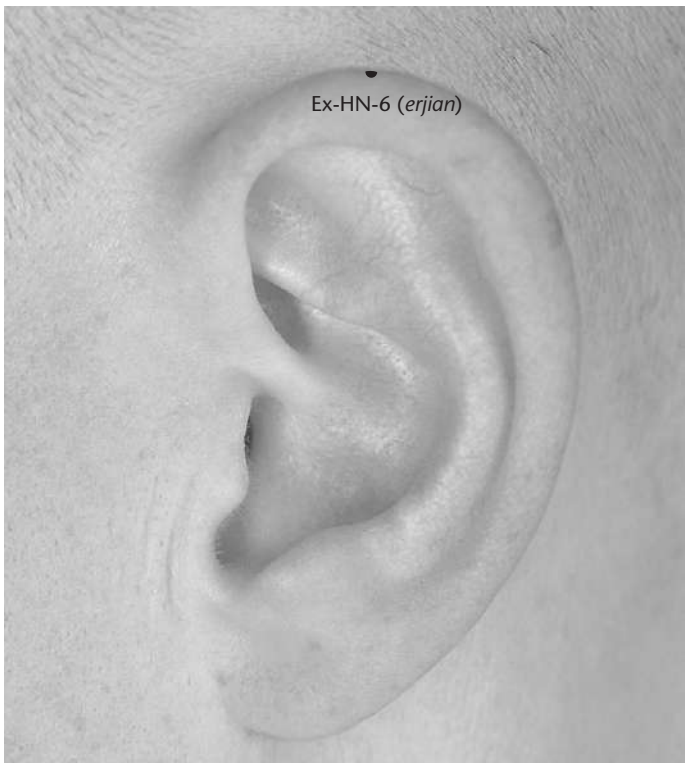

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 545

Behind the Ball QIUHOU Ex-HN-7

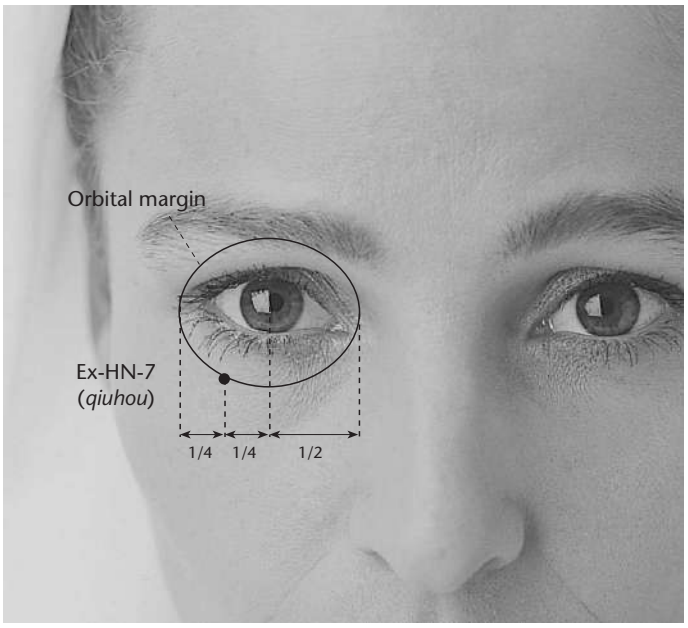

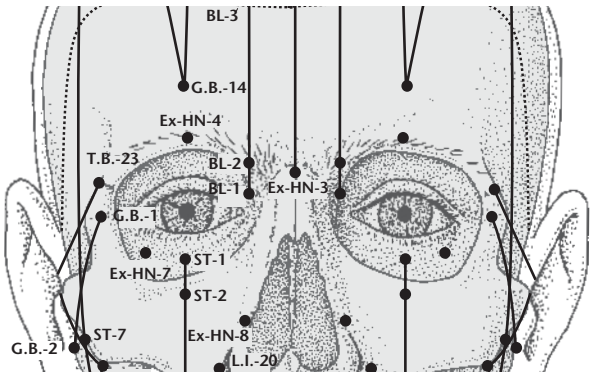

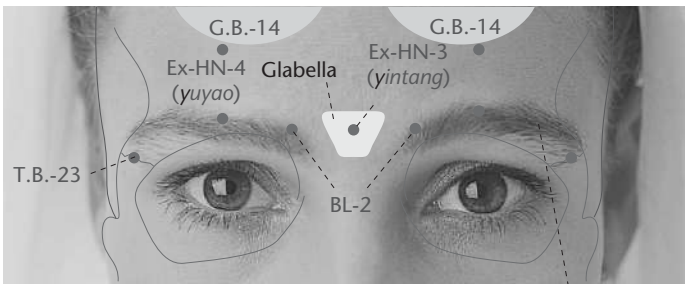

Location

On the lower border of the orbit, at the junction of the lateral quarter and the medial three quarters.

How to find

Divide the horizontal extension of the diameter of the orbit in quarters. Then locate the point at the junction of the first and second lateral quarters, slightly superior to the border of the orbit.

Needling

Ask the patient to look upward and gently push the eyeball upward, away from the lower eyelid. Slowly insert the needle 0.5–1 cun into the fatty tissue immediately above the bone (orbit). Caution: Do not injure the eyeball and the periosteum. Avoid the venous plexus and arteries! Pay attention to any pain from needling. No needle manipulation! After removal of the needle, compress the site for 10 minutes. Haematomas may still occur (inform the patient prior to needling). Moxibustion is contraindicated. Alternative, less difficult points for eye disorders: **BL-2, T.B.-23**, **G.B.-1**, **ST-2, Ex-HN-5** (*taiyang*), **Ex-HN-4** (*yuyao*).

Actions/Indications

- Disorders of the eyes

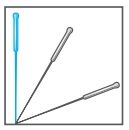
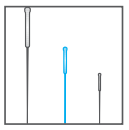
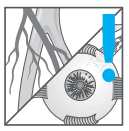

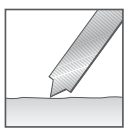

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 546

Ex-HN-8 *Upper* Yingxiang SHANGYINGXIANG

Location

At the upper end of the nasiolabial groove, at the junction of the maxilla and the nasal cavity. Note: *bitong* (Clear Nose) is an alternative name for this point.

How to find

At the upper end of the nasiolabial groove, palpate for the bony margin of the nasal cavity. Here, **Ex-HN-8** (*shangyingxiang* or *bitong*) is located at the junction of the nose and the cheek.

Needling

0.3–0.5 cun vertically towards the centre of the nasal cavity.

Actions/Indications

- Clears Heat, opens the nasal passages

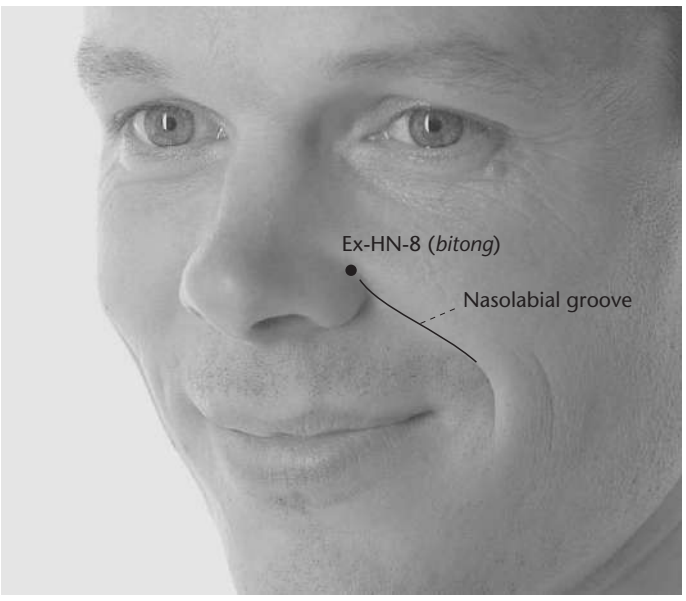

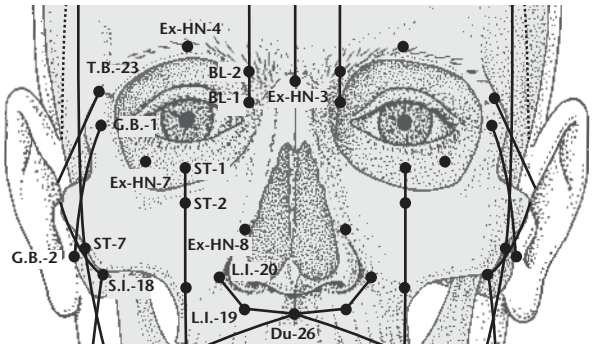

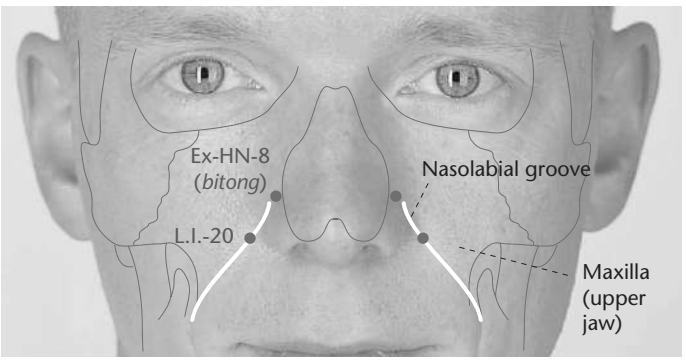

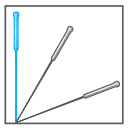
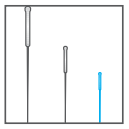

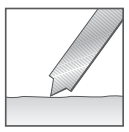

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 547

Inner Yingxiang NEIYINGXIANG Ex-HN-9

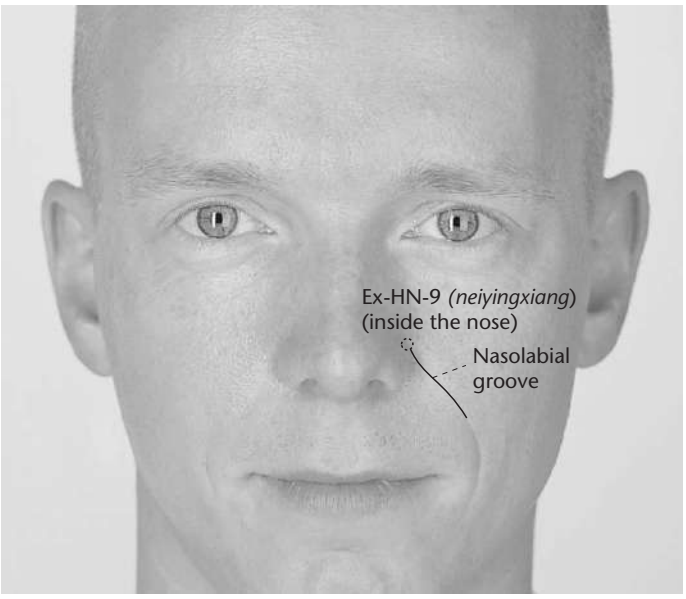

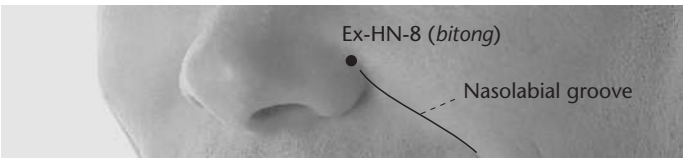

Location

In the nasal cavity, at the junction of the nasal bone and the nasal cartilage.

How to find

This point lies 'opposite' the external point → **Ex-HN-8** (*shang-yingxiang* or *bitong*), which is located at the upper end of the nasolabial groove.

Needling

Prick to bleed with a needle, lancet or three-edged needle. Caution: Contraindicated in patients with bleeding disorders (or taking anticoagulants). Needling may be painful!

Actions/Indications

- Clears Heat, drains Fire

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 548

Ex-HN-10 *Gathering Spring* JUQUAN

Location

With maximal extension of the tongue, in the centre of the tongue body.

How to find

With maximal extension of the tongue in the centre of the tongue body.

Needling

Vertically 0.2 cun. Needling may be painful!

Actions/Indications

- Deviation of the tongue, impaired mobility or atrophy of the musculature of the tongue: for example, after a stroke or with loss of sense of taste

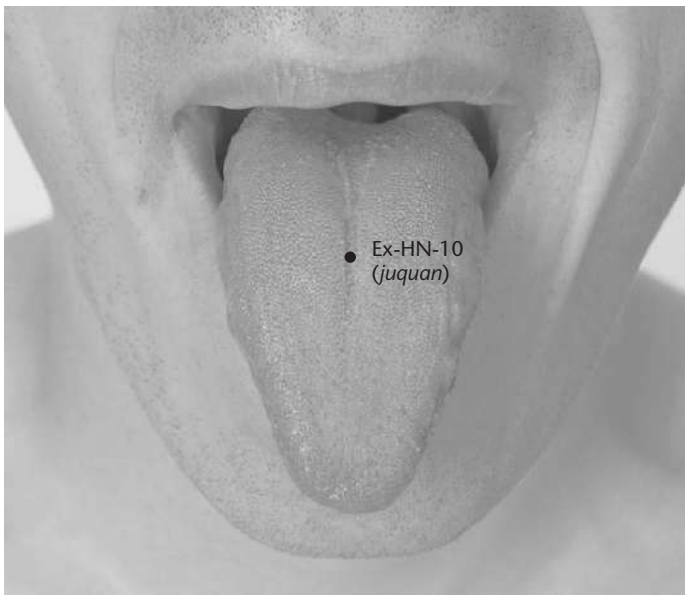

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 549

Sea Spring HAIQUAN Ex-HN-11

Location

Below the tongue, in the centre of the frenulum, between the points → **Ex-HN-12** (*jinjin*) and → **Ex-HN-13** (*yuye*).

How to find

Ask the patient to roll their tongue upward to reveal the frenulum. To avoid injury to the frenulum, this point should be located at the base of the frenulum.

Needling

Vertically 0.2 cun. Remove needle immediately after short stimulation. Needling may be painful!

Actions/Indications

- Mouth and tongue ulcers, hiccups

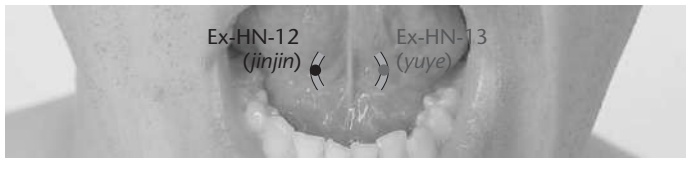

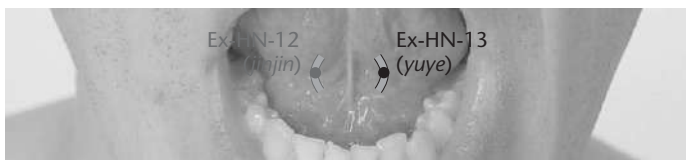

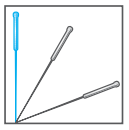
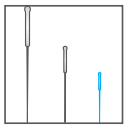

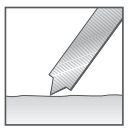

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 550

Ex-HN-12 *Golden Liquid* JINJIN

Location

On the underside of the tongue, on the lingual vein to the left of the frenulum.

How to find

Ask the patient to roll their tongue upward to reveal the lingual veins.

Needling

Prick to bleed: Briefly puncture with a needle, lancet or three-edged needle and let bleed. Often used together with → **Ex-HN-13** (*yuye*). Caution: Contraindicated with bleeding disorders or in patients taking anticoagulant medication.

Actions/Indications

- Dispels Heat and Wind

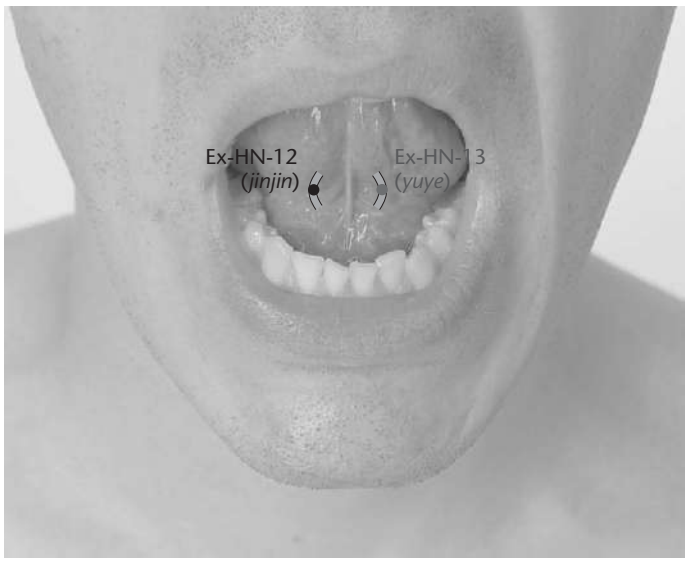

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 551

Jade Fluid YUYE Ex-HN-13

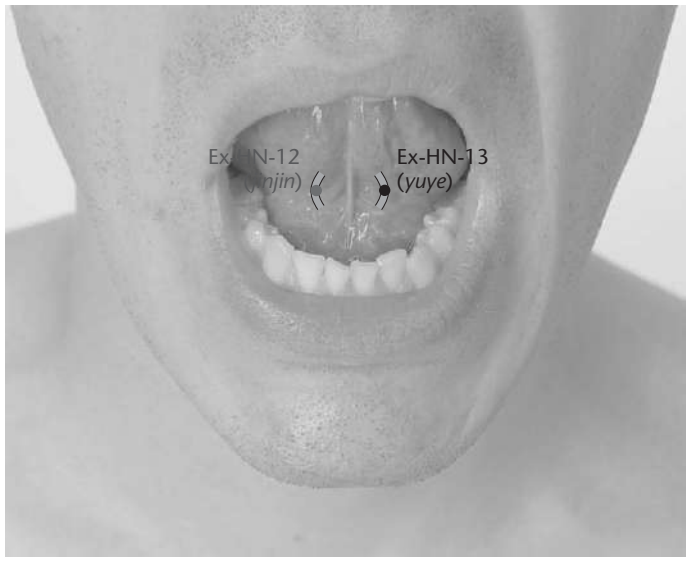

Location

On the underside of the tongue, on the great lingual vein to the right of the frenulum.

How to find

Ask the patient to roll their tongue upward to reveal the lingual veins.

Needling

Prick to bleed: Briefly puncture with a needle, lancet or three-edged needle and let bleed. Often used together with → **Ex-HN-12** (*jinjin*). Caution: Contraindicated with bleeding disorders or in patients taking anticoagulant medication.

Actions/Indications

- Dispels Heat and Wind

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 552

Ex-HN-14 *Eye Brightening* YIMING

Location

At the junction of the head and the occiput, posterior to the mastoid bone.

How to find

Ex-HN-14 (*yiming*) is located on a line connecting → **T.B.-17** (below the earlobe, between the maxilla and the mastoid) and → **G.B.-20** (at the junction of the occiput and the neck, in the centre of the dorsolateral depression). From → **T.B.-17**, measure 1 cun towards the posterior and there locate *yiming* posterior to the mastoid.

Located slightly superior and posterior to *yiming* is → **Ex-HN** (*anmian*), in the corner between the mastoid and the occiput.

Needling

Vertically 0.5–1 cun

Actions/Indications

- Disorders of the eyes and ears, dizziness, insomnia

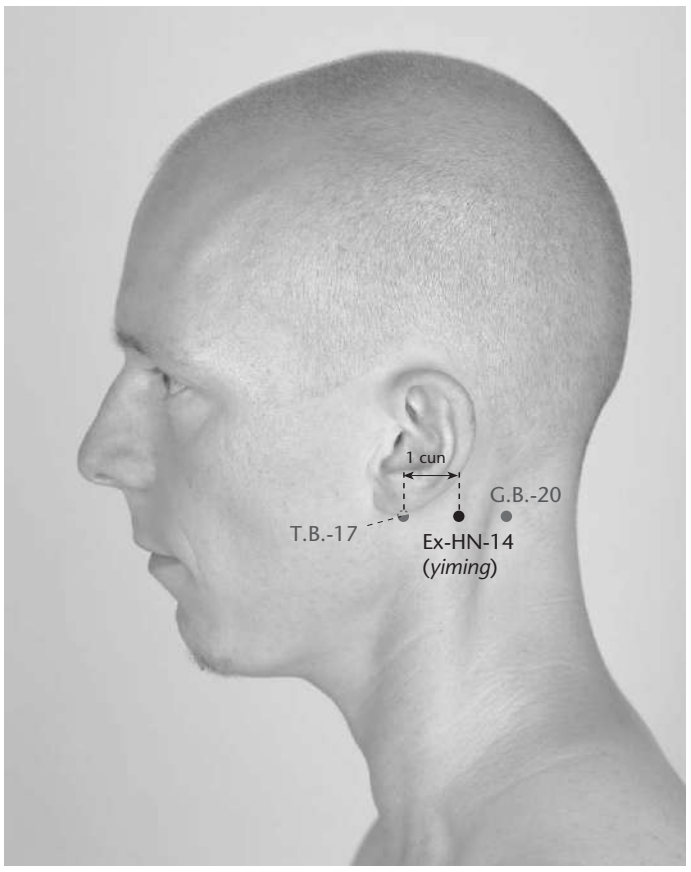

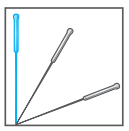

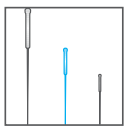

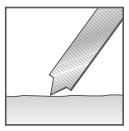

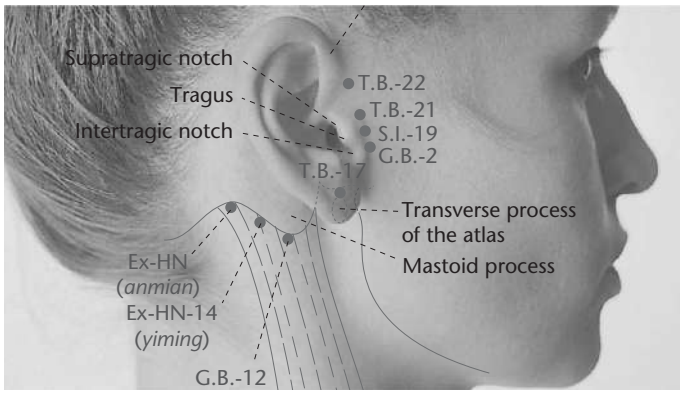

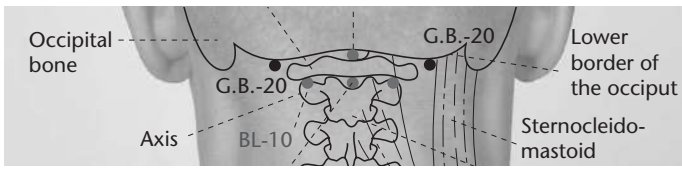

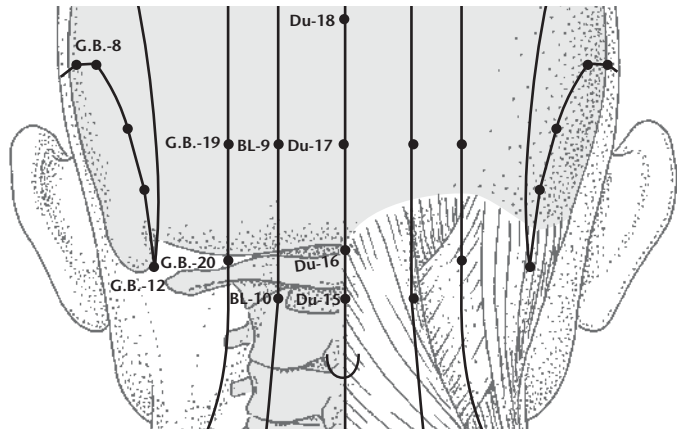

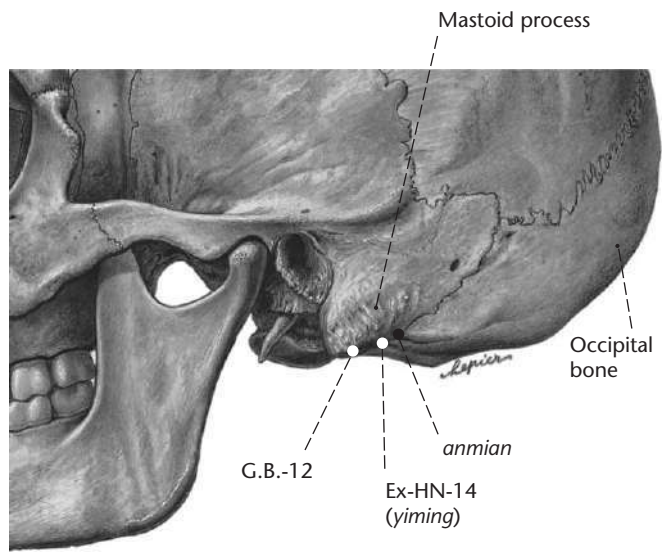

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 553

Hundred Taxations JINGBAILAO/BAILAO Ex-HN-15

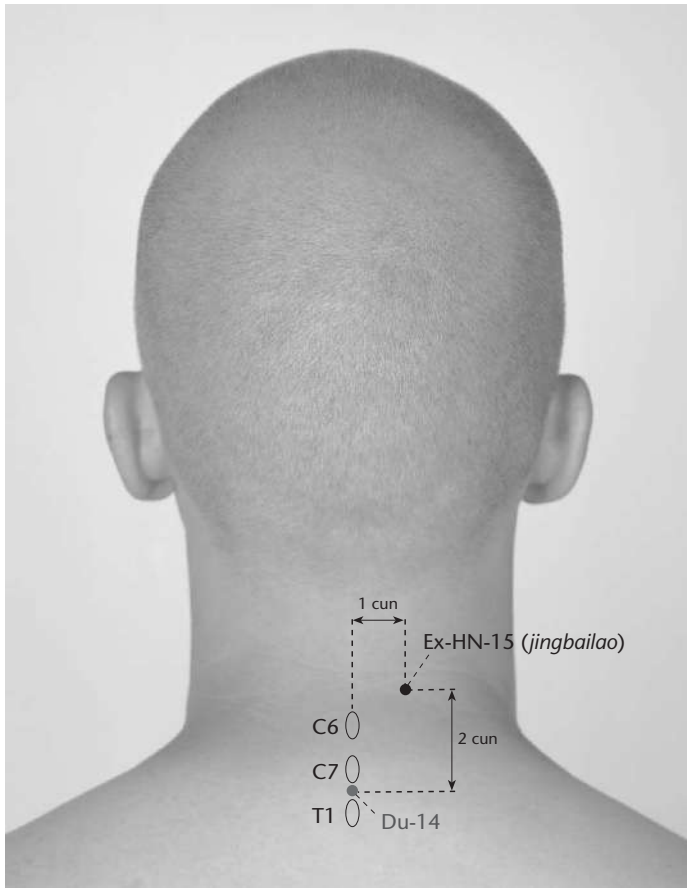

Location

2 cun superior to the lower border of the spinous process of the 7th cervical vertebra (C7) and 1 cun lateral to the midline.

How to find

For locating C7, see ➞ 3.4.1. ➞ **Du-14** (*dazhui*) is located below the spinous process of C7. From there, measure 2 cun in a superior and 1 cun in a lateral direction.

Needling

Vertically 0.5–0.8 cun

Actions/Indications

- Dispels Wind and Dampness
- Opens the channel

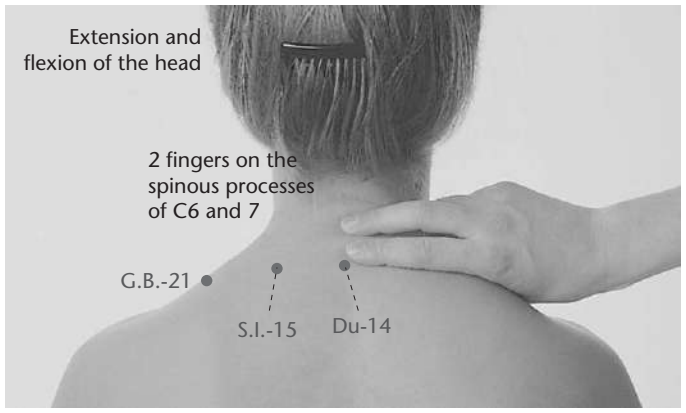

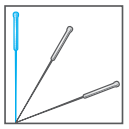
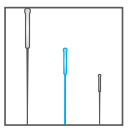

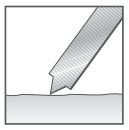

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 554

Ex-HN Upper Brightness SHANGMING

Location

Directly superior to the pupil, below the margin of the orbit.

How to find

With the patient looking straight ahead, palpate for the margin of the orbit directly above the pupil. **Ex-HN** (*shangming*) is located between the upper margin of the orbit and the eyeball.

Needling

Gently push the eyeball downward. Slowly insert the needle 0.5–1 cun vertically into the fatty tissue immediately below the bone (orbit). Caution: Pay attention to any pain from needling. No manipulation! After removal of the needle, compress the site for 10 minutes. Haematomas may still occur (inform the patient prior to needling). Alternative points with less risk of complications are **BL-2, T.B.-23**, **G.B.-1, ST-2, Ex-HN-5** (*taiyang*), **Ex-HN-4** (*yuyao*).

Actions/Indications

- Disorders of the eyes

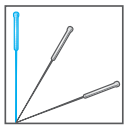

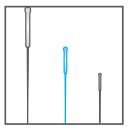

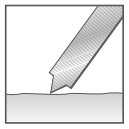

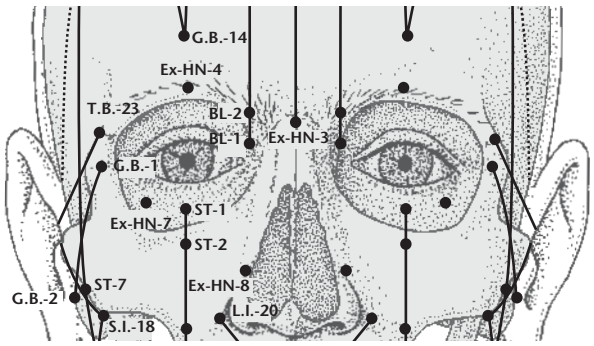

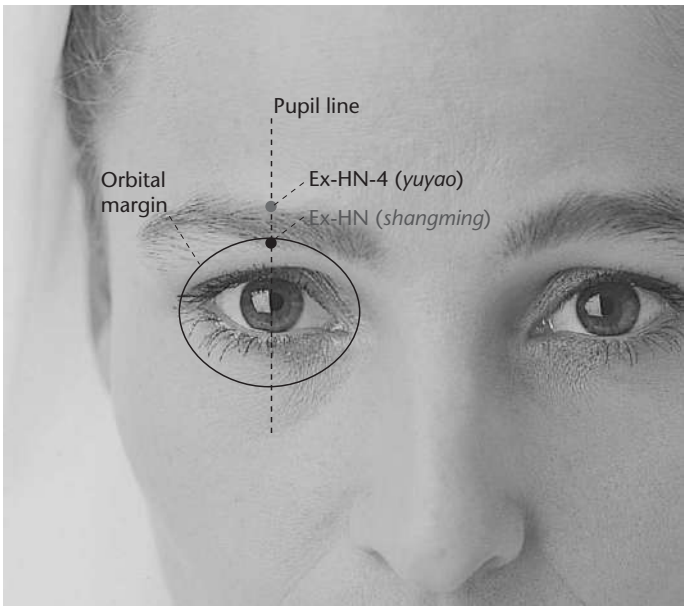

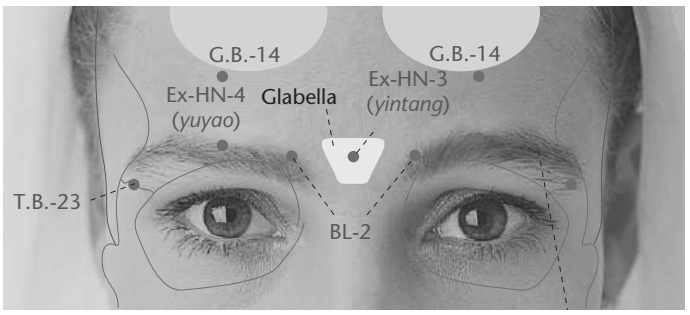

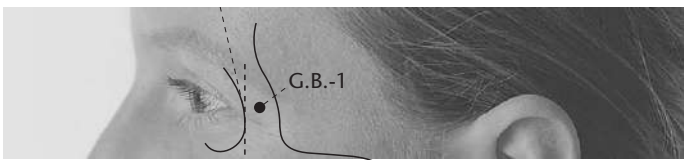

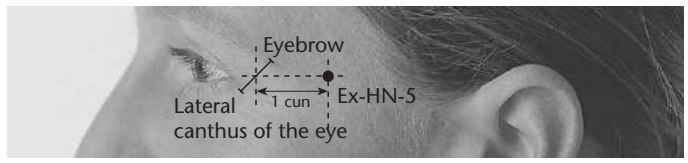

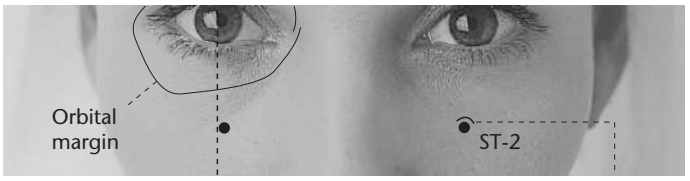

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 555

Peaceful Sleep ANMIAN Ex-HN

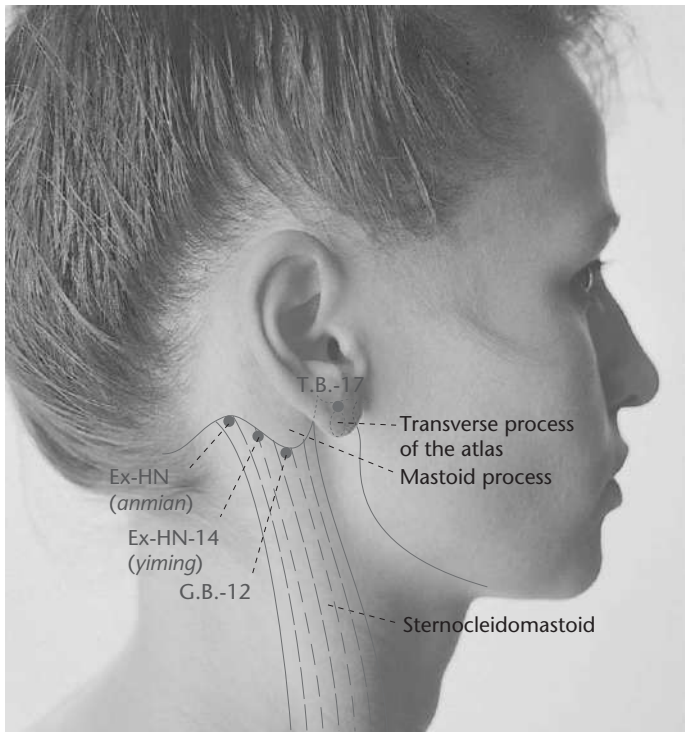

Location

Posterior to the ear, between → **T.B.-17** and → **G.B.-20**, posterior to the mastoid process.

How to find

First, locate → **T.B.-17** (directly below the earlobe, in the depression between the lower jaw and the mastoid process (→ 3.1) that forms when the mouth is open). Next, locate → **G.B.-20** on the lower border of the occiput, in the depression between the origins of the sternocleidomastoid and trapezius muscles. Locate **Ex-HN** (*anmian*) approximately midway between these two points, slightly posterior to the mastoid bone and superior to → **G.B.-12**, in a depression on the lower border of the occiput.

Needling

0.5–1 cun vertically or obliquely towards → **T.B.-17** (*yifeng*) or → **G.B.-20** (*fengchi*).

Actions/Indications

- Calms the *shen*

Special features

Important calming point for sleeping disorders. This extra point is not part of the WHO classification of extra points.

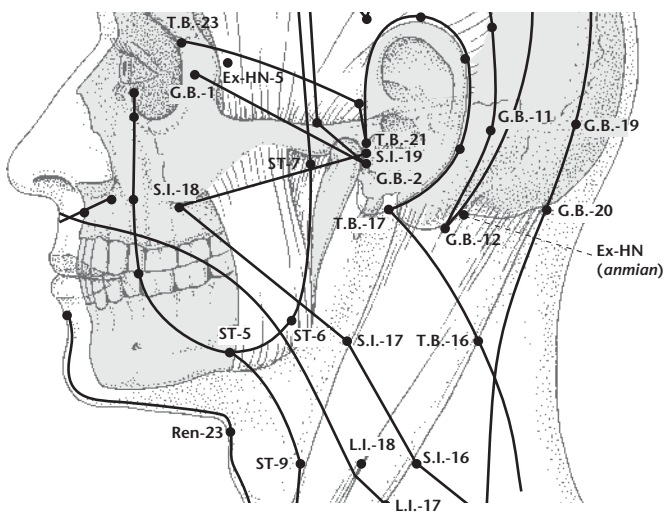

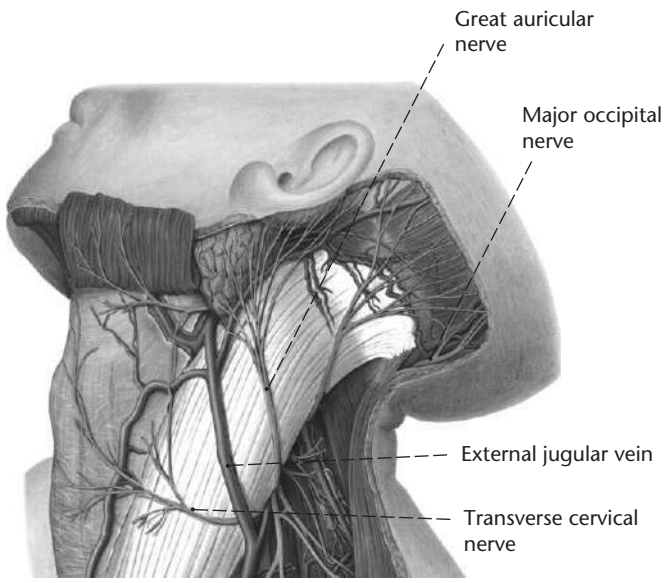

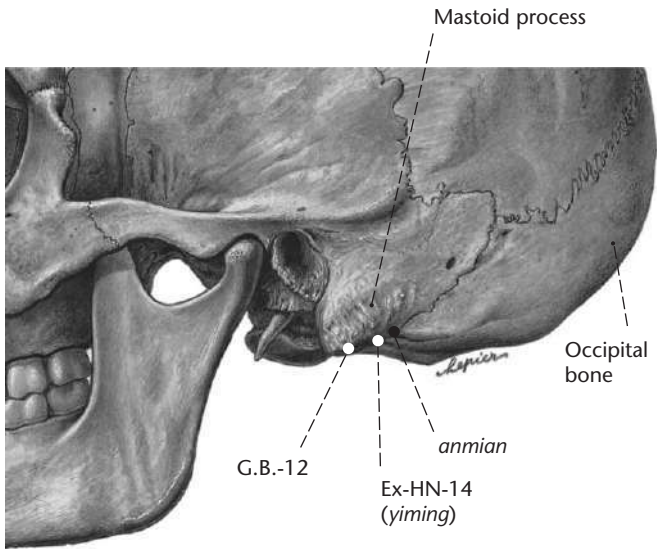

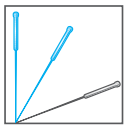
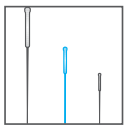

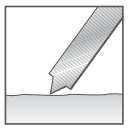
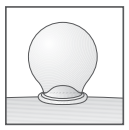

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 556

Ex-HN *Cheek Centre* JIALI

Location
Inside the mouth, on the mucosa of the cheek, 1 cun posterior to the corner of the mouth.

How to find
Locate this point on the mucosa of the cheek, 1 cun posterior to the corner of the mouth.

Needling
0.3–0.5 cun obliquely in a posterior direction, prick to bleed.

Actions/Indications
- Clears Heat
- For inflammations of the mouth and throat, for gastritis

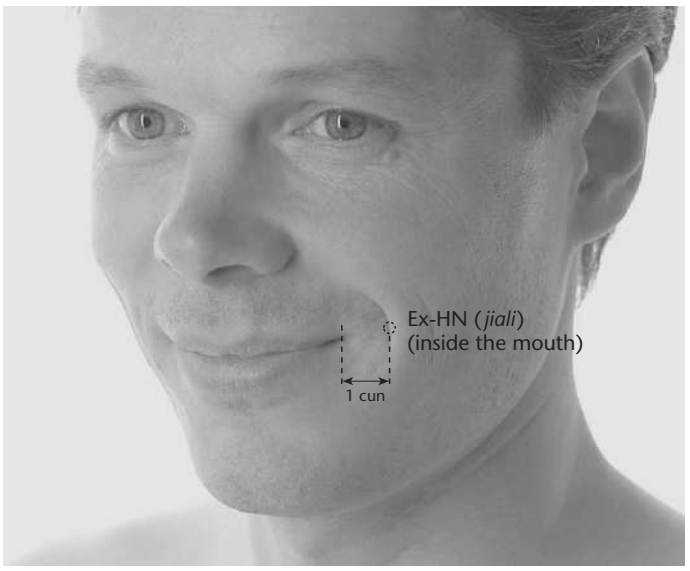

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 557

Adjacent to Container of Fluids JIACHENGJIANG

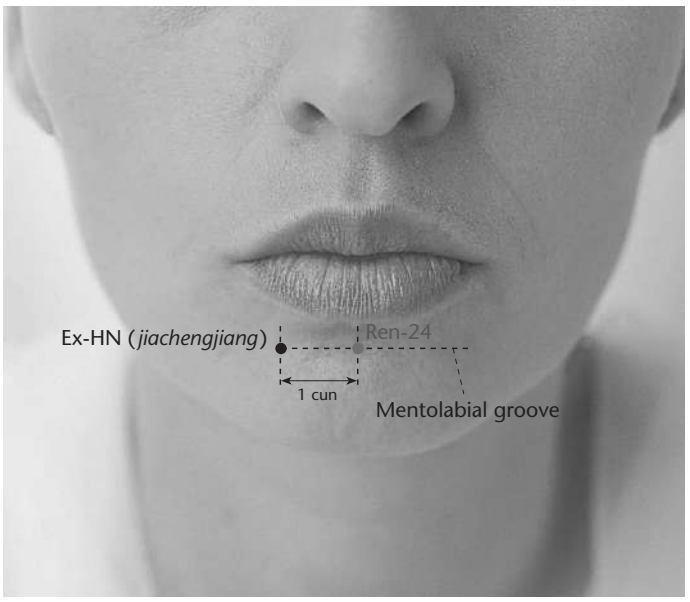

Location

1 cun lateral to the centre of the mentolabial groove.

How to find

This point is located on the mental foramen. → **Ren-24** (*chengjiang*) is located in the centre of the mentolabial groove.

Needling

0.3–0.5 cun vertically or up to 1 cun obliquely towards the mental foramen.

Actions/Indications

- Dispels Wind, opens the channel

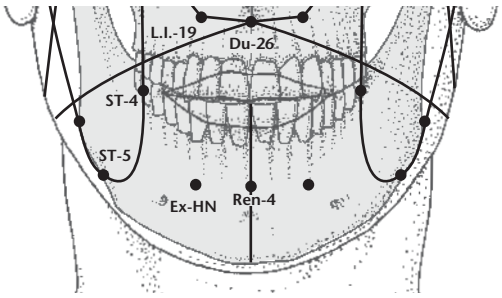

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 558

Ex-HN *Prominent Bone* CHONGGU ZHUIDONG

Location
Below the spinous process of the 6th cervical vertebra (C6).

How to find
This point is located on the posterior midline and thus on the *du mai*. In fact, the 'prominent bone' (vertebra prominens) inferior to *chonggu zhuidong* is the spinous process of either C7 or T1. ➞ **Du-14** is located between the two. For the correct method of locating C6 and C7 ➞ 3.4.1.

Needling
0.5–1 cun obliquely in a superior direction.

Actions/Indications
- Dispels external pathogenic factors
- Harmonises the *shen*

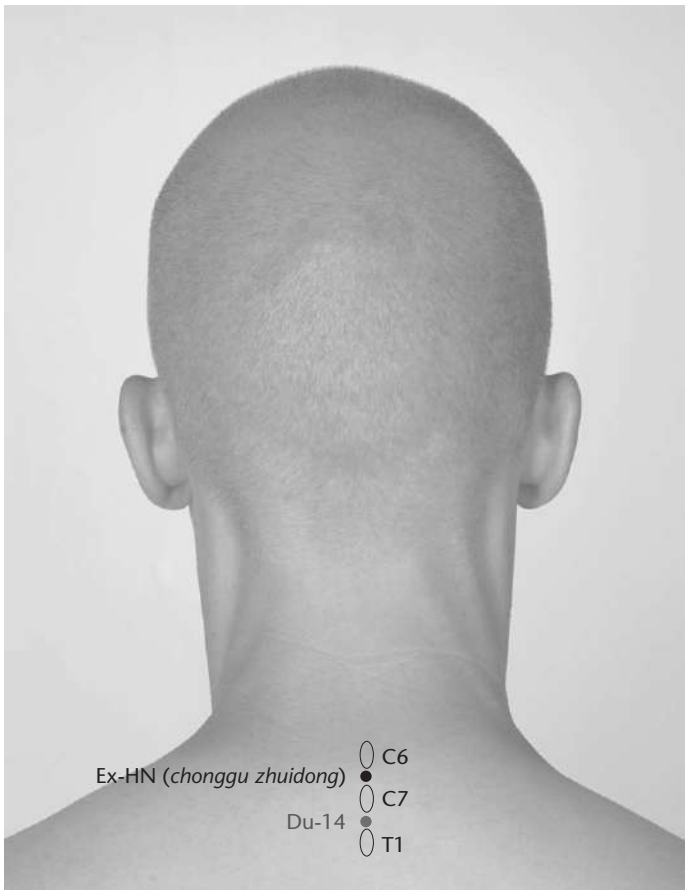

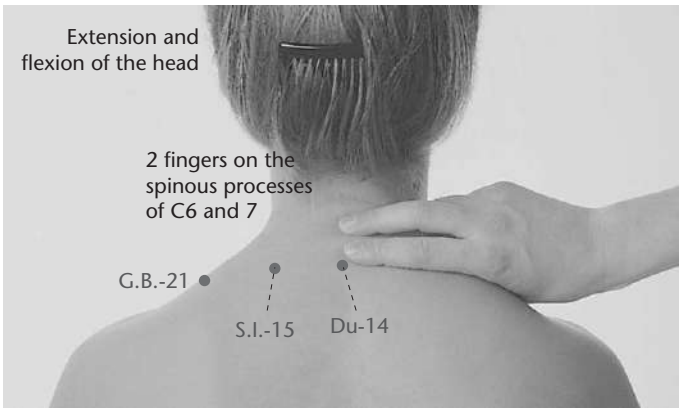

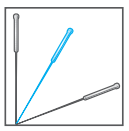
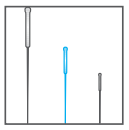

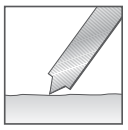

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 559

Upper Arm JINGBI

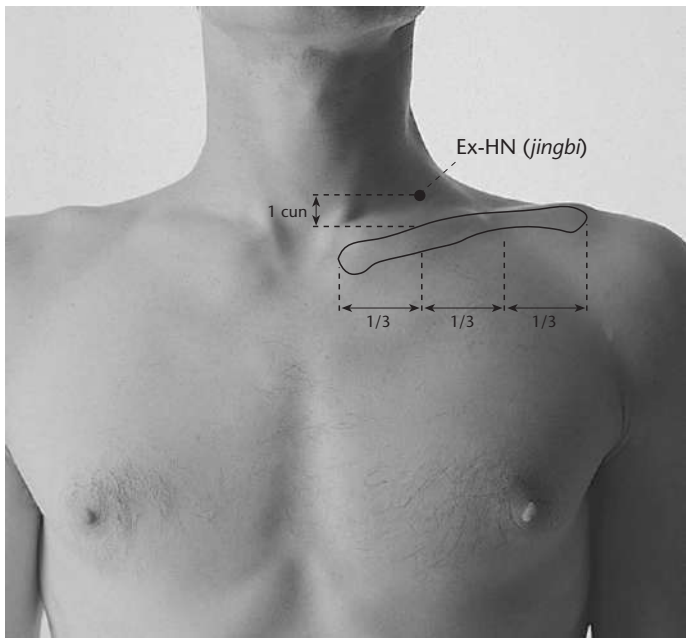

Location

1 cun superior to the junction of the proximal and middle third of the clavicle.

How to find

Measure a third of the clavicle from its medial end, then measure 1 cun in a superior direction and there locate *jingbi* in the supraclavicular fossa. The point is located above the brachial plexus.
→ **ST-12** is very close to this point, directly superior to the midpoint of the clavicle.

Needling

Vertically 0.3–0.5 cun. During insertion or stimulation, a tingling or warm sensation should be felt radiating to the fingers. Caution: Pneumothorax.

Actions/Indications

- Paraesthesia and paralysis of the upper extremity

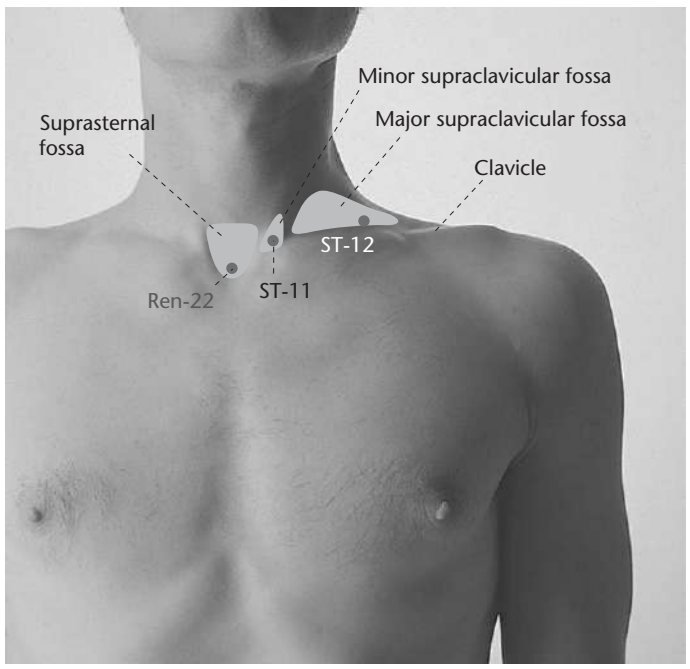

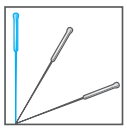
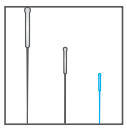
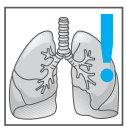

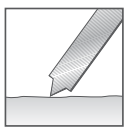

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 561

6.2 Extra Points: Chest and Abdomen (EX-CA)

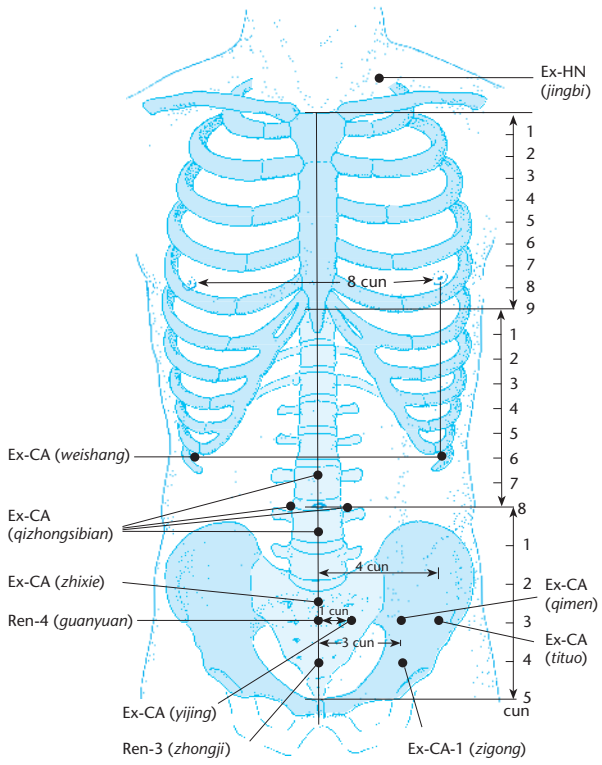

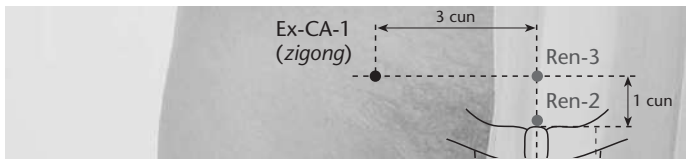

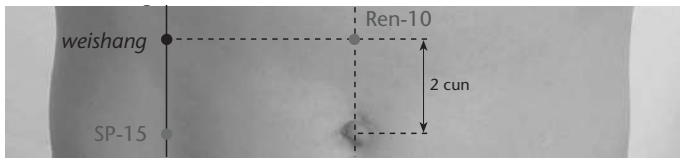

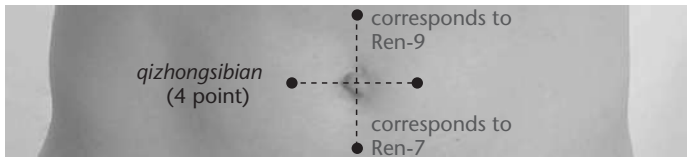

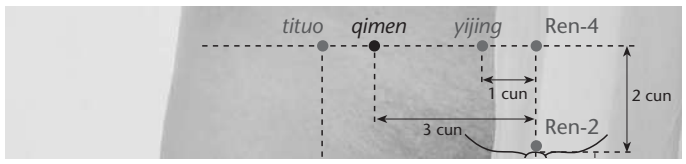

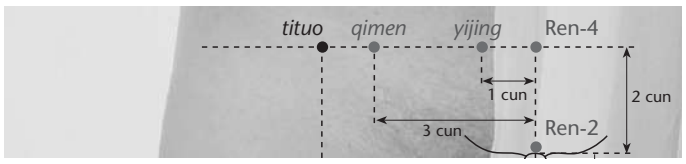

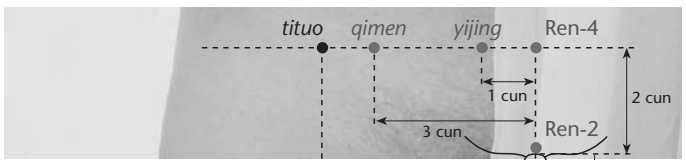

English abbreviation (standard)	*Pinyin* name	Nguyen Van Nghi, König and Wancura, Schnorrenberger	Shanghai College	Ex (Hempen)
Ex-CA-1	*zigong*	PaM or ZP49	M-CA-18	Ex9
Further extra points				
Ex-CA	*weishang*			
Ex-CA	*qizhongsibian*			
Ex-CA	*yijing*	PaM or ZP45		
Ex-CA	*qimen*	PaM or ZP46		
Ex-CA	*tituo*	NP39	M-CA-4	
Ex-CA	*zhixie*	NP38	N-CA31	

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 563

Palace of the Child ZIGONG Ex-CA-1

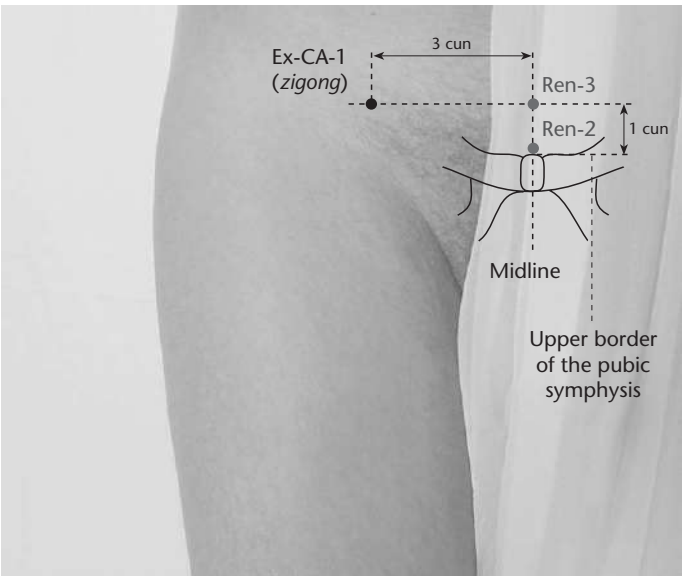

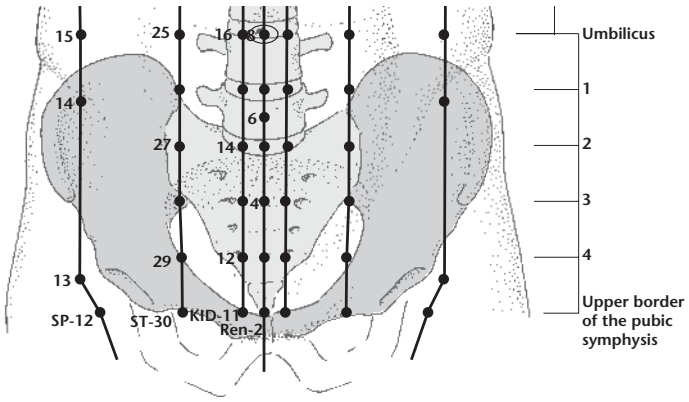

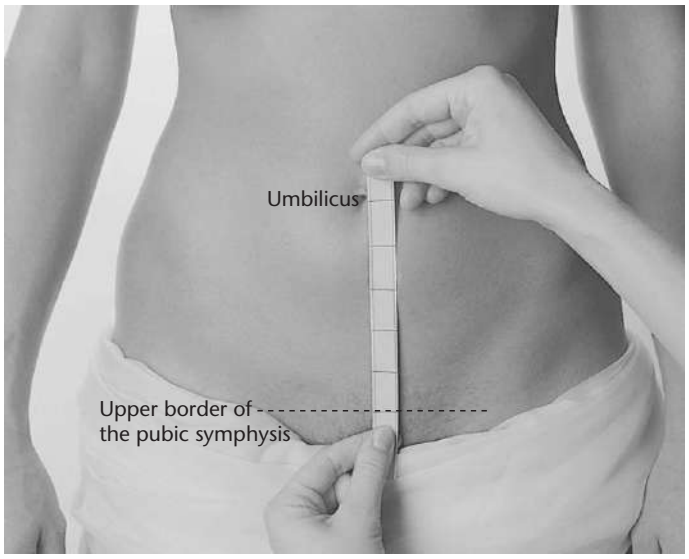

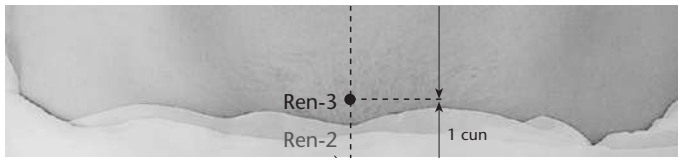

Location

3 cun lateral to the anterior midline and 1 cun superior to the upper border of the pubic symphysis.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used (helpful device: elastic tape ➞ 2.3.1). From the upper border of the pubic symphysis, measure 1 cun in a superior direction – this is the location of ➞ **Ren-3**. From there, measure 3 cun in a lateral direction to locate *zigong*.

Located on the same level are ➞ **KID-12** (0.5 cun from the midline) and ➞ **ST-29** (2 cun lateral to the midline).

Needling

0.5–1 cun vertically or 1–2 cun obliquely towards the upper border of the pubic symphysis. Caution: Peritoneum, pregnancy, full bladder.

Actions/Indications

- Strengthens and raises Qi
- Regulates menstruation
- Alleviates pain in the abdomen and the lumbar region

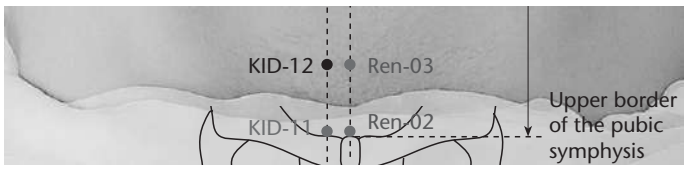

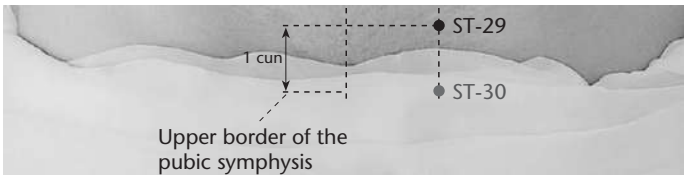

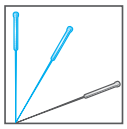
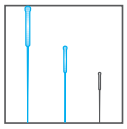
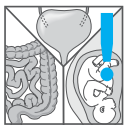

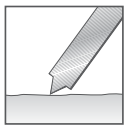

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 564

Ex-CA *Lifting the Stomach* WEISHANG

Location

On the Spleen channel, 4 cun lateral and 2 cun superior to the umbilicus.

How to find

The distance between the sternocostal angle and the centre of the umbilicus is divided into 8 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used (helpful device: elastic tape → 2.3.1). From the centre of the umbilicus, measure 2 cun in a superior direction (location of → **Ren-10**) and then locate *weishang* 4 cun lateral to the midline. 1 cun superior to → **Ex-CA** (*weishang*) is → **SP-16**.

Located on the same level are → **Ren-10/KID-17/ST-23** (on the midline/0.5 cun lateral to the midline/2 cun lateral to the midline).

Needling

2–3 cun obliquely in the direction of the umbilicus. Caution: Peritoneum, pregnancy.

Actions/Indications

- Gastroptosis, abdominal pain

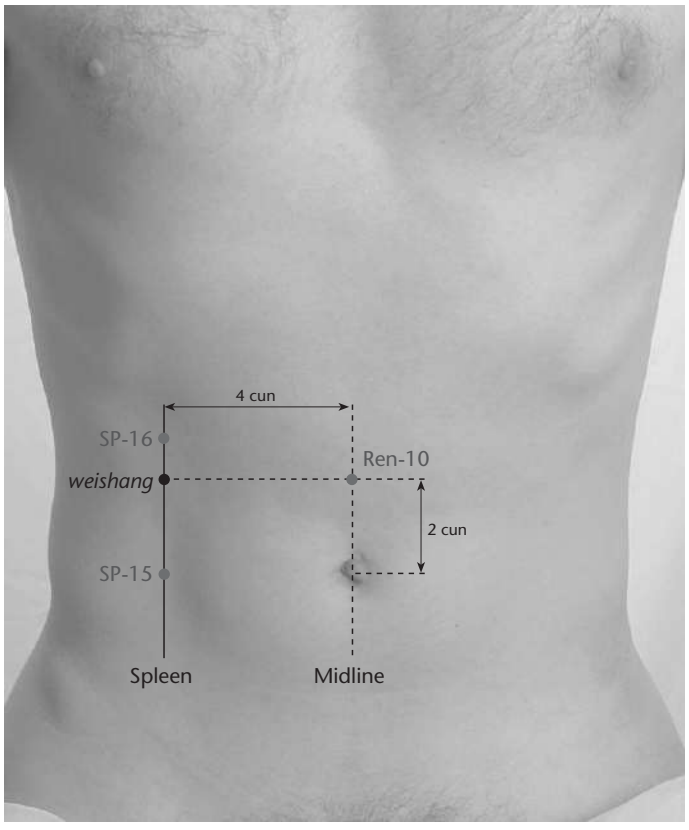

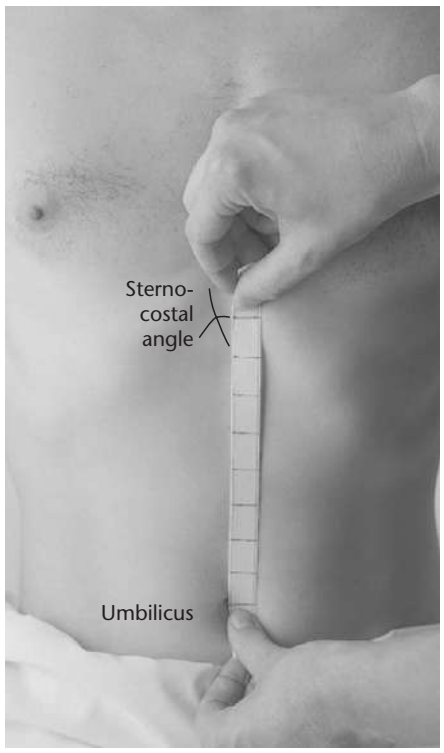

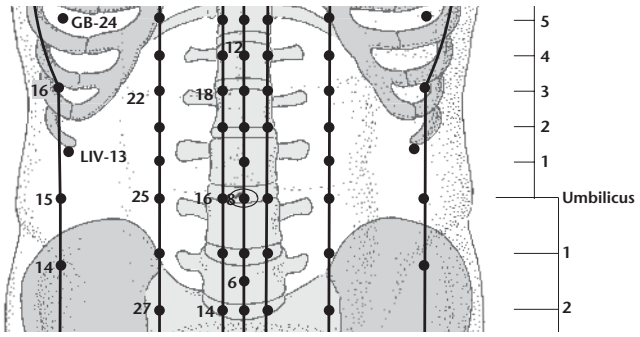

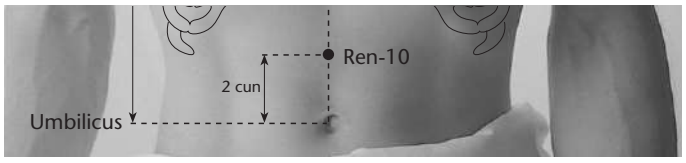

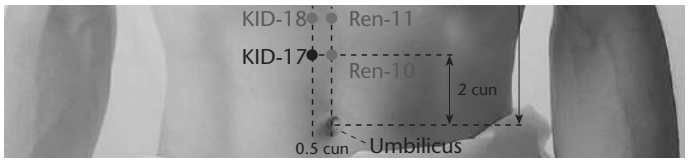

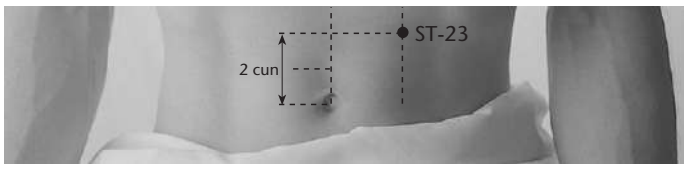

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 565

Four Points Around the Umbilicus QIZHONGSIBIAN

Ex-CA

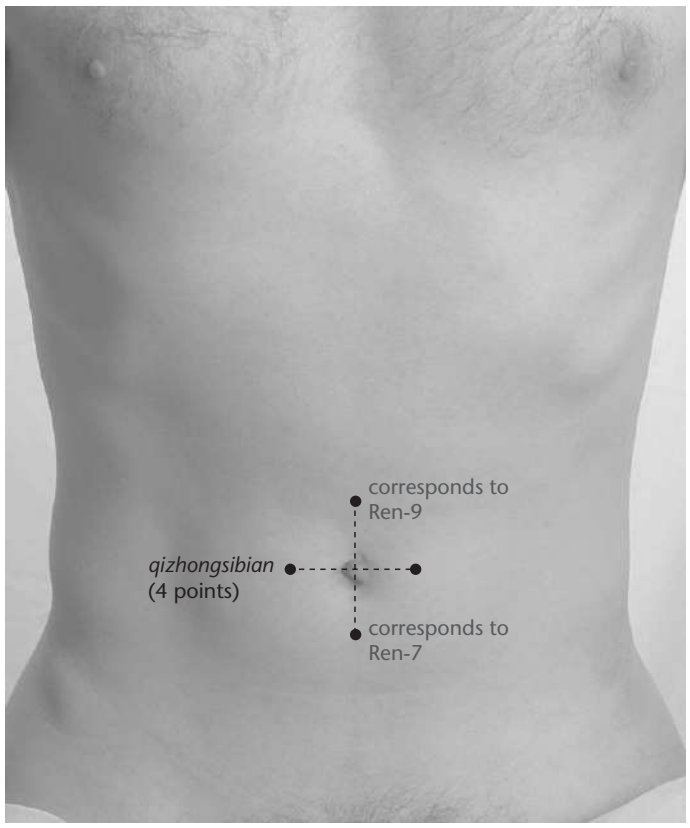

Location

Four points, 1 cun lateral, superior and inferior to the umbilicus.

How to find

These four points are arranged in a star shape around the umbilicus.

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Distension, diarrhoea, dyspepsia, dysmenorrhoea

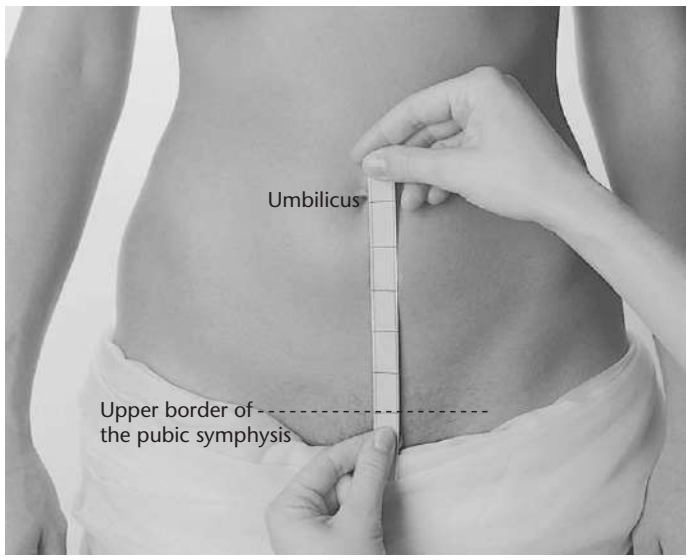

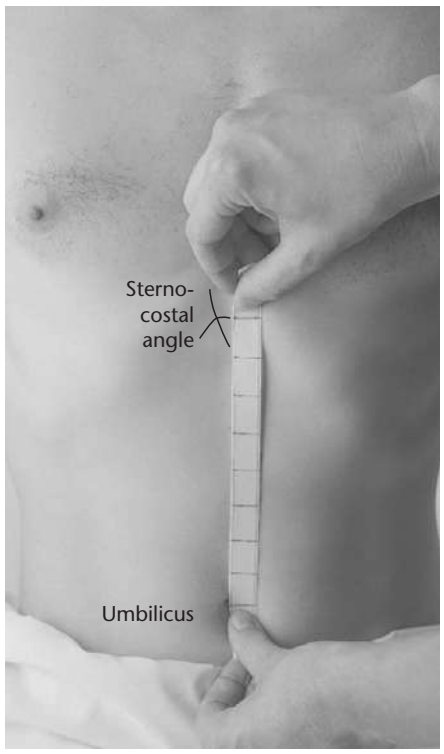

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 566

Ex-CA Loss of Semen YIJING

Location

1 cun lateral to the anterior midline and 2 cun superior to the upper border of the pubic symphysis.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used (helpful device: elastic tape → 2.3.1). From the upper border of the pubic symphysis, measure 2 cun in a cranial direction (→ **Ren-4**) and, from there, measure 1 cun in a lateral direction. There, locate *yijing*.

Located on the same level are → **KID-13** (0.5 cun lateral to the midline), **Ex-CA** (*qimen*, 3 cun lateral to the midline), → **ST-28** (2 cun lateral to the midline) and **Ex-CA** (*tituo*, 4 cun lateral to the midline). → **G.B.-27** is located medial to the ASIS, approximately on the same level.

Needling

0.5–1 cun. Caution: Peritoneum, full bladder, pregnancy.

Actions/Indications

Ejaculation disorders, impotence, scrotal eczema

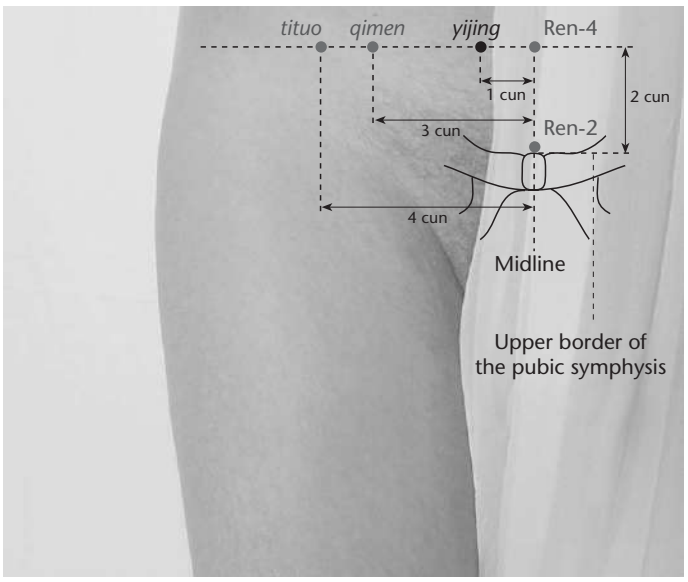

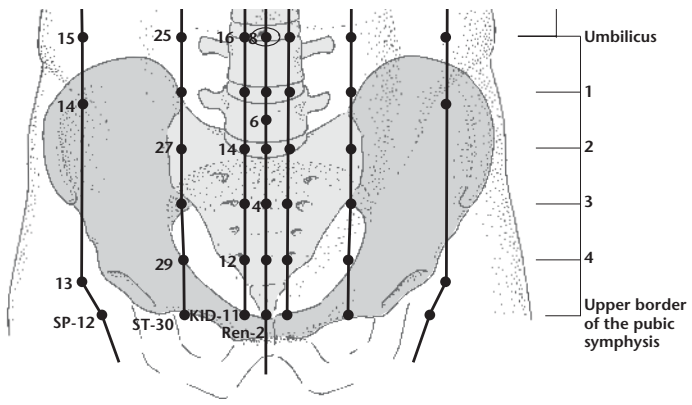

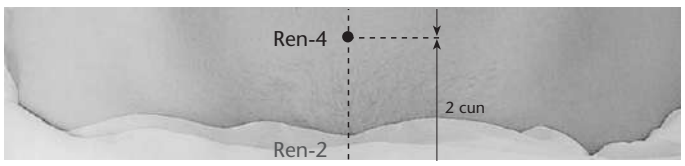

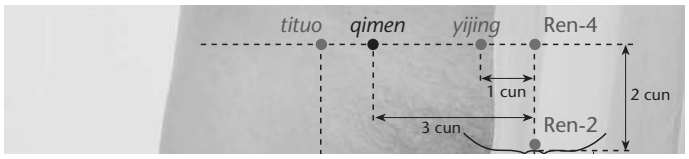

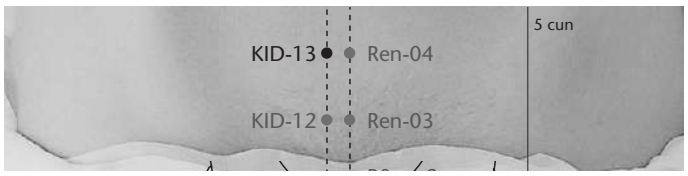

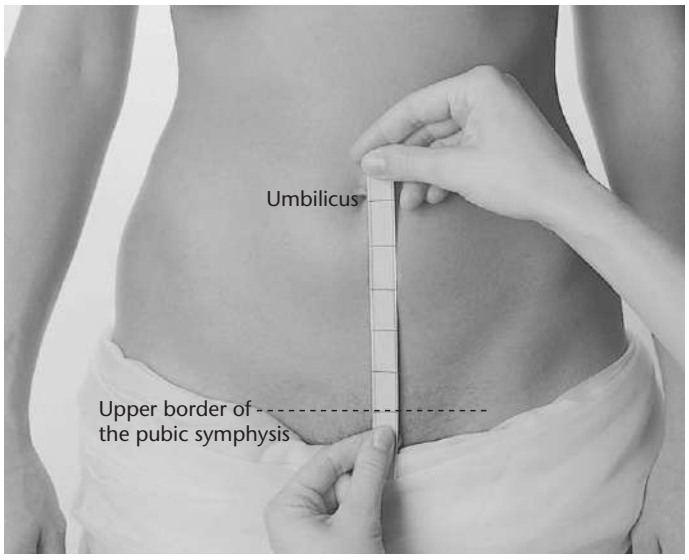

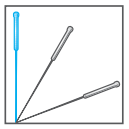

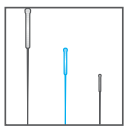

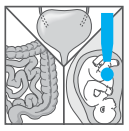

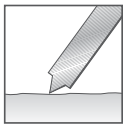

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 567

Qi Gate QIMEN Ex-CA

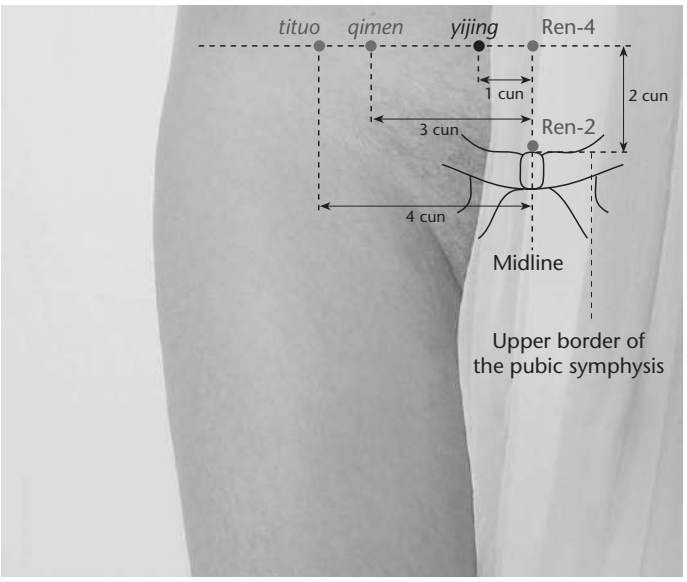

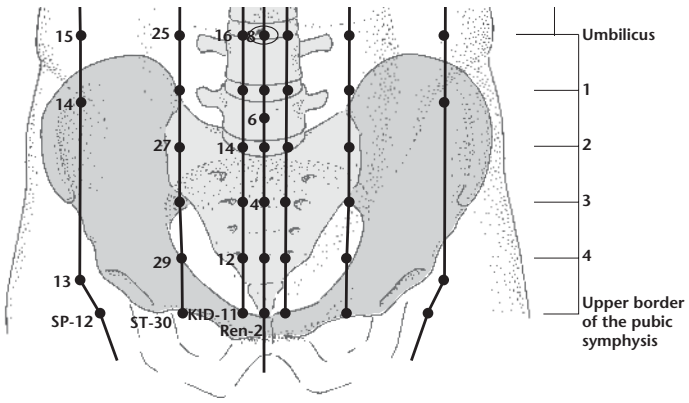

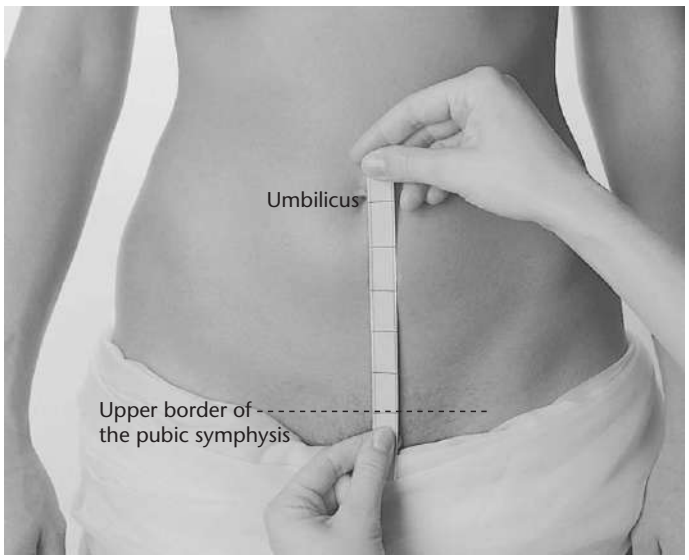

Location

3 cun lateral to the anterior midline and 2 cun superior to the upper border of the pubic symphysis.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used (helpful device: elastic tape → 2.3.1). From the upper border of the pubic symphysis, measure 2 cun in a cranial direction (→ **Ren-4**) and, from there, measure 3 cun in a lateral direction. There, locate *qimen*.

Located on the same level are → **KID-13** (0.5 cun lateral to the midline), **Ex-CA** (*yijing*, 1 cun lateral to the midline), → **ST-28** (2 cun lateral to the midline) and **Ex-CA** (*tituo*, 4 cun lateral to the midline). → **G.B.-27** is located medial to the ASIS, approximately on the same level.

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Metrorrhagia, female infertility, orchitis, urinary tract infections, persistent lochial discharge

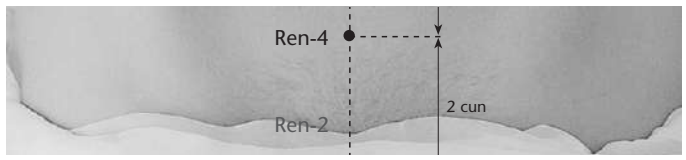

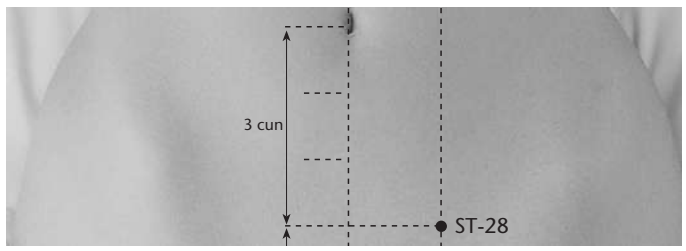

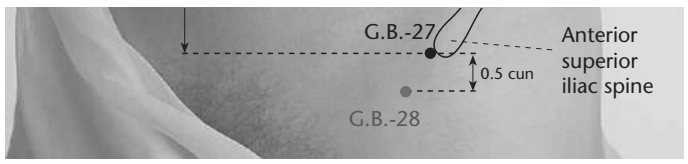

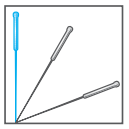
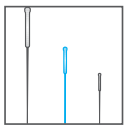
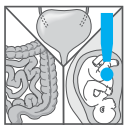

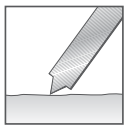

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 568

Ex-CA *Lift and Support* TITUO

Location

4 cun lateral to the anterior midline and 2 cun superior to the upper border of the pubic symphysis.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used (helpful device: elastic tape → 2.3.1). From the upper border of the pubic symphysis, measure 2 cun in a cranial direction (→ **Ren-4**) and from there measure 4 cun in a lateral direction. There, medial to the ASIS, locate *tituo*.

Located on the same level are → **KID-13** (0.5 cun lateral to the midline), → **Ex-CA** (*yijing*, 1 cun lateral to the midline), → **ST-28** (2 cun lateral to the midline) and → **Ex-CA** (*qimen*, 3 cun lateral to the midline). → **G.B.-27** is located medial to the ASIS, approximately on the same level.

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, pregnancy.

Actions/Indications

- Strengthens the rising Qi and alleviates organ prolapse

Special features

In modern TCM, *tituo* is considered one of the most important points for prolapse of the uterus.

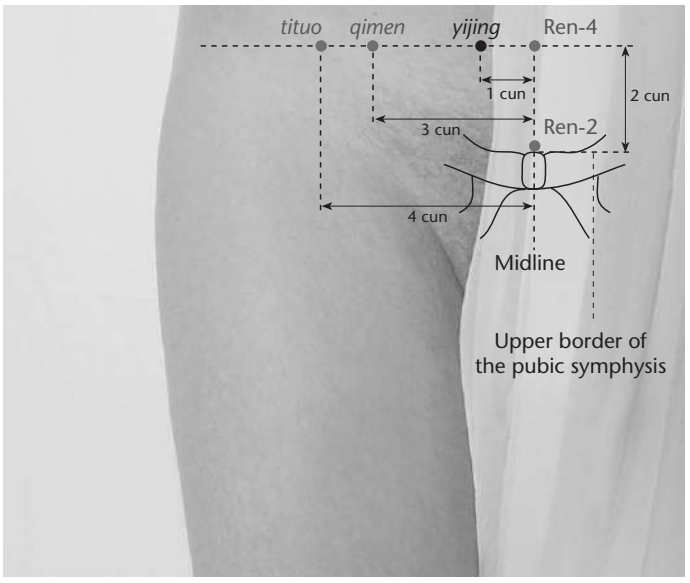

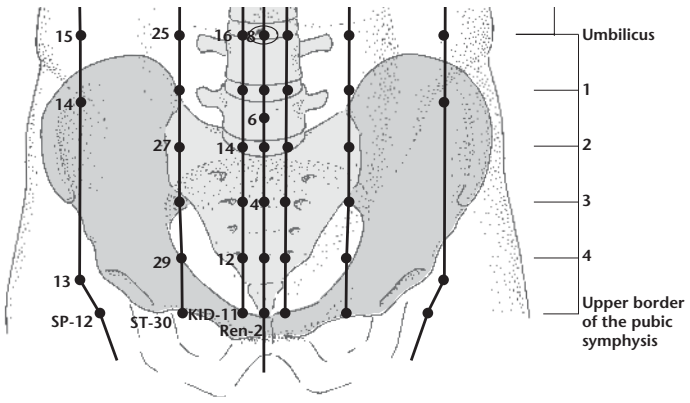

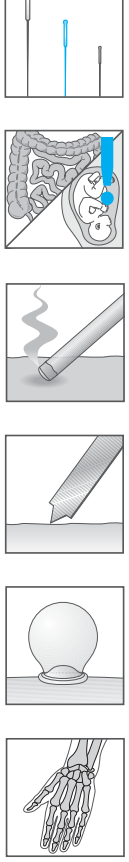

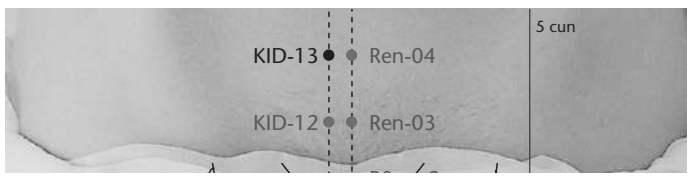

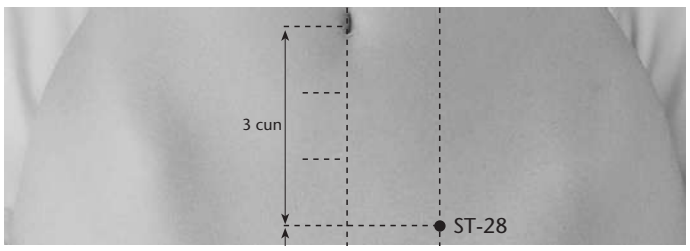

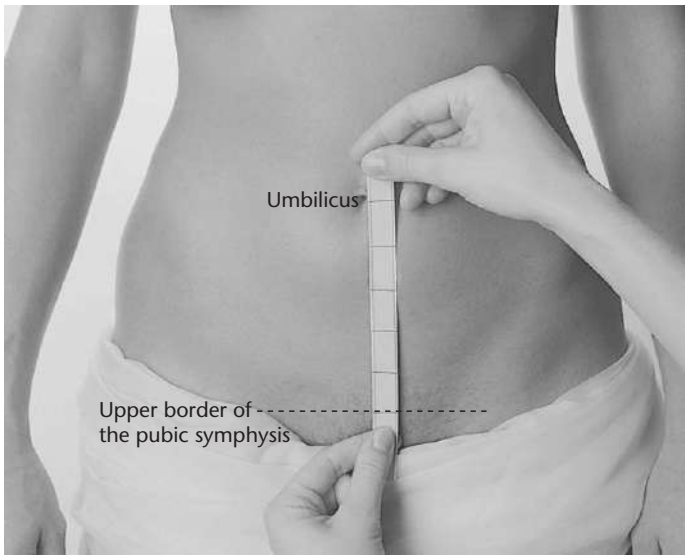

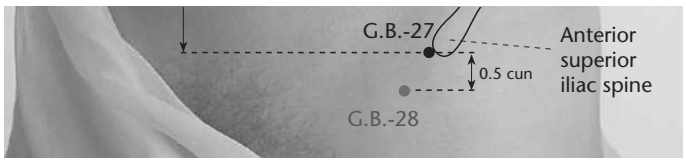

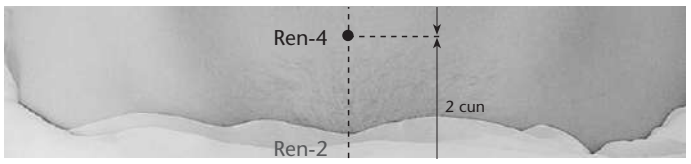

Ch06a-F10028.qxd 2/23/08 1:25 PM Page 569

End Diarrhoea ZHIXIE

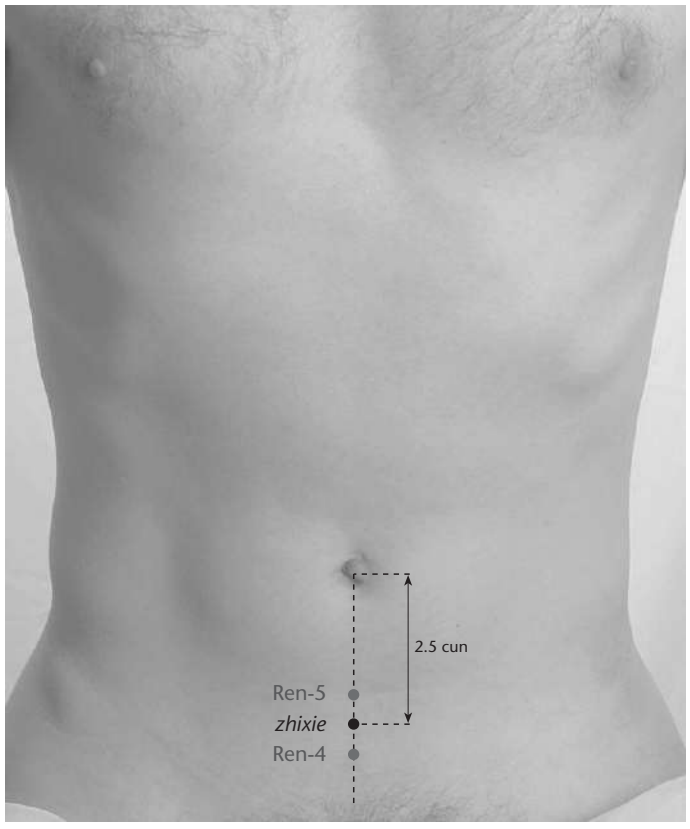

Location

On the anterior midline, 2.5 cun inferior to the umbilicus.

How to find

The distance between the centre of the umbilicus and the upper border of the pubic symphysis is divided into 5 proportional cun, which can vary considerably from the patient's finger cun. Therefore, only proportional cun measurements should be used (helpful device: elastic tape → 2.3.1). *Zhixie* is located at the midpoint of this distance, between → **Ren-4** (2 cun superior to the upper border of the pubic symphysis) and → **Ren-5** (2 cun inferior to the umbilicus). **Or:** Spreading hands technique (→ 2.3.3): Place the little fingers on the umbilicus and the upper border of the pubic symphysis respectively and spread the hands evenly, joining the thumbs at the midpoint of this distance.

Needling

Vertically 0.5–1 cun. Caution: Peritoneum, full bladder, pregnancy.

Actions/Indications

- Stops diarrhoea

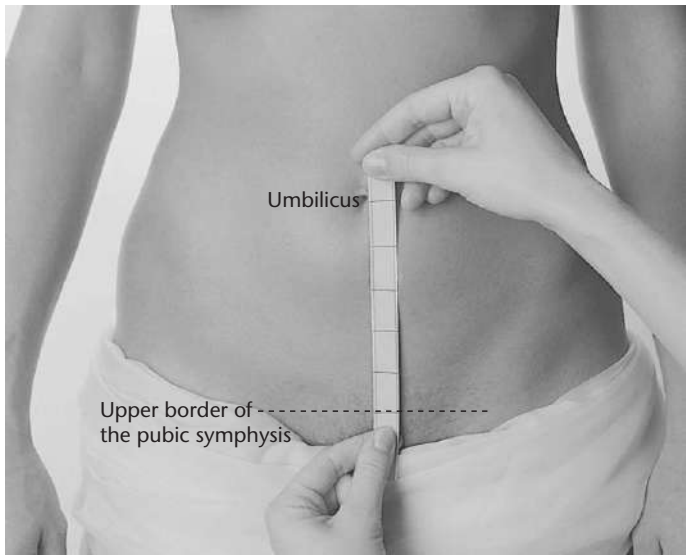

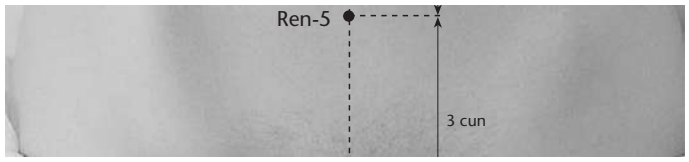

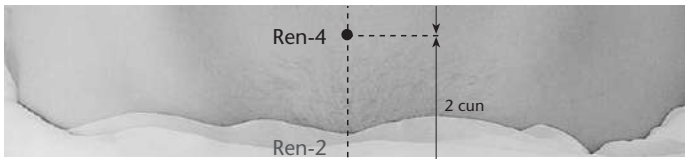

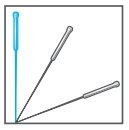
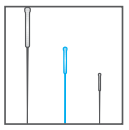
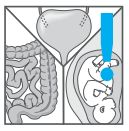

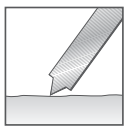
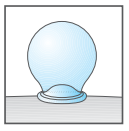

Ch06a-F10028.qxd 2/23/08 1:26 PM Page 570

Ex-CA *Triangle Moxibustion* SANJIAOJIU

Location

These three points are located on the corners of an equilateral triangle, the apex of which is formed by the umbilicus, while the base forms a horizontal line on the abdomen. The sides are equal to the patient's smile.

How to find

It is helpful to explain the location of this point to the patient. Invariably, this will make them smile, allowing the practitioner to determine the length of the sides of the triangle.

Needling

Moxibustion only, no needling.

Actions/Indications

- Regulate Qi and stop diarrhoea.

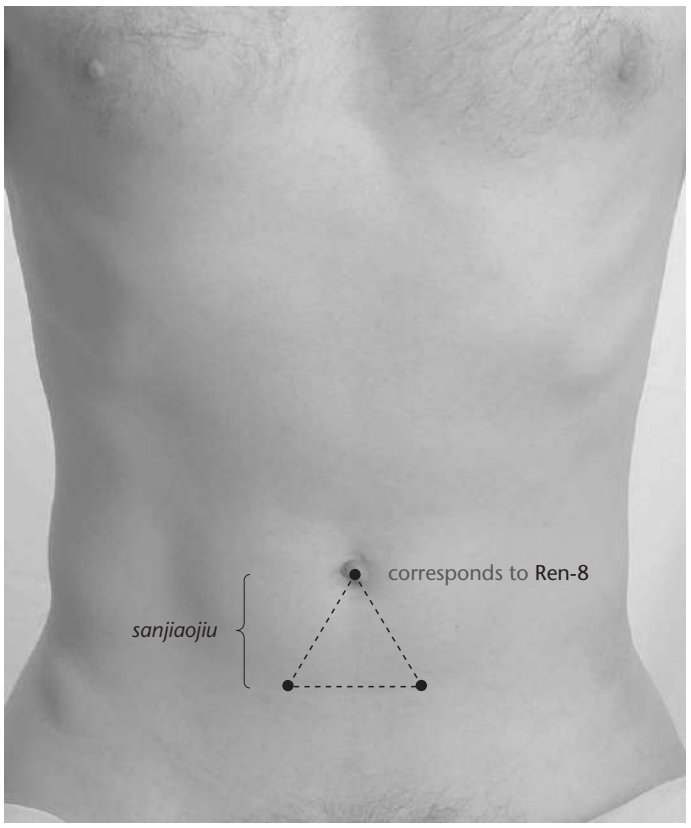

Ch06a-F10028.qxd 2/23/08 1:26 PM Page 571

6.3 Extra Points: Back (EX-B)

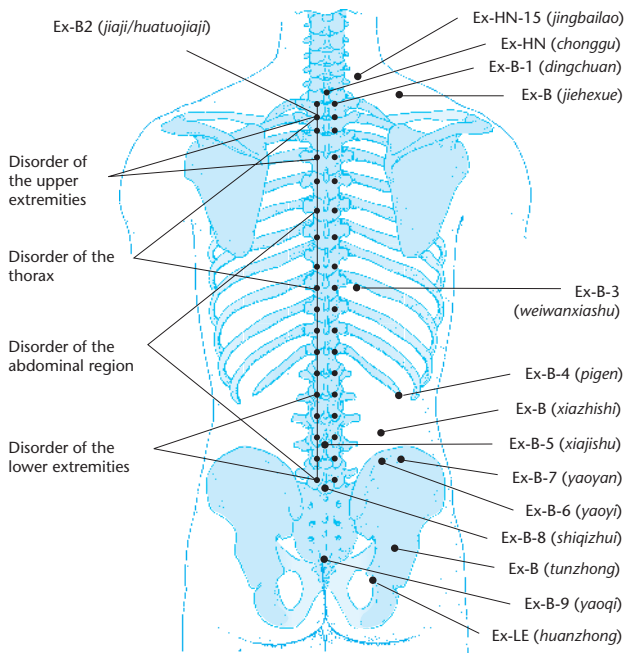

English abbreviation (standard)	*Pinyin* name	Nguyen Van Nghi, König and Wancura, Schnorrenberger	Shanghai College	Ex (Hempen)
Ex-B-1	*dingchuan*	NP45	M-BW-1	Ex10
Ex-B-2	*huatuojiaji/jiaji*	PaM or ZP85	M-BW-35	Ex12
Ex-B-3	*weiwanxiashu/ waiguanxiashu/ bashu* or *yishu*	PaM or ZP62	M-BW-12	
Ex-B-4	*pigen*	PaM or ZP66	M-BW-16	
Ex-B-5	*xiajishu*	PaM or ZP71	M-BW-21	
Ex-B-6	*yaoyi*	PaM or ZP73	M-BW-23	
Ex-B-7	*yaoyan*	PaM or ZP74	M-BW-24	
Ex-B-8	*shiqizhui/shiqizhuixia*	PaM or ZP75	M-BW-25	Ex11
Ex-B-9	*yaoqi*	PaM or ZP79	M-BW-29	
Further extra points				
Ex-B	*jiehexue*	NP47	N-BW-6	
Ex-B	*tunzhong*	PaM or ZP83	M-BW-33	

Ch06a-F10028.qxd 2/23/08 1:26 PM Page 573

Calm Dyspnoea DINGCHUAN Ex-B-1

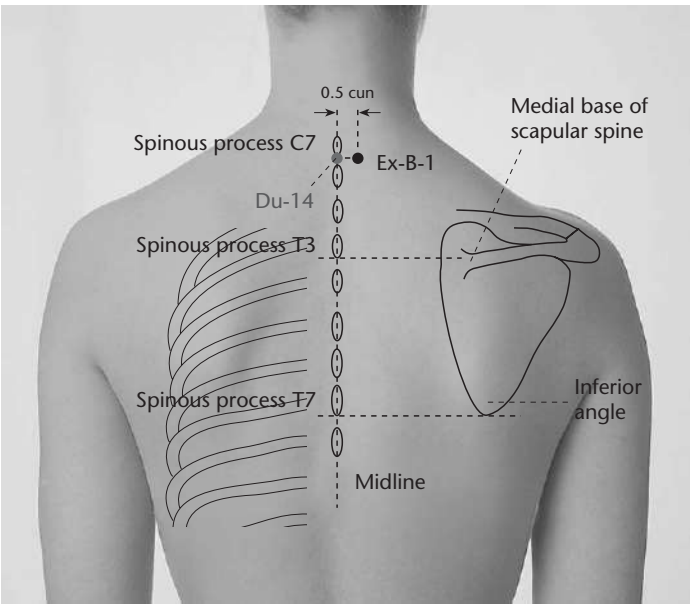

Location

0.5 cun lateral to the lower border of the spinous process of the 7th cervical vertebra (C7).

How to find

First, identify C7 (→ 3.4.1). Next, measure 0.5 cun from the lower border of its spinous process in a lateral direction and there locate **Ex-B-1.**

Located on the same level are → **Du-14** (on the midline), → **S.I.-15** (2 cun lateral to the midline) and **Ex-B** (*jiehexue*, 3.5 cun lateral to the midline).

Needling

0.5–1 cun in a medial direction.

Actions/Indications

- Harmonises Lung Qi
- Local point for the neck and shoulders

Special features

Modern main point for dyspnoea

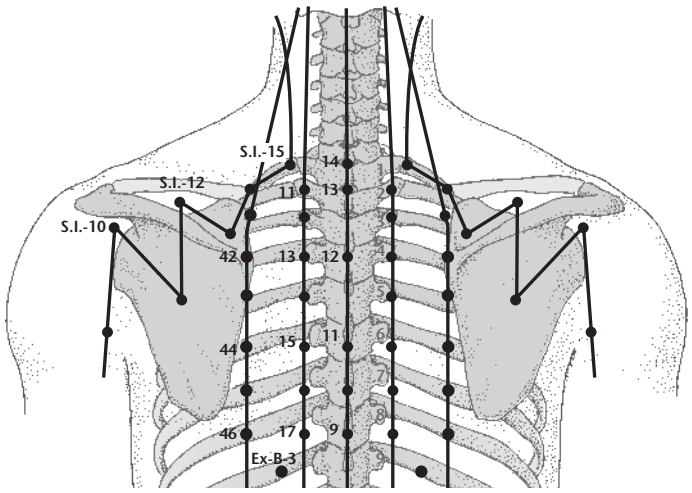

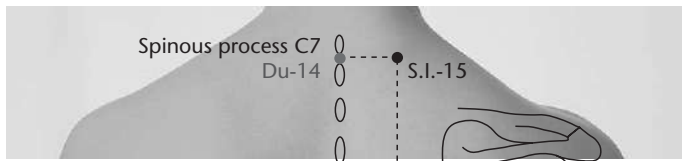

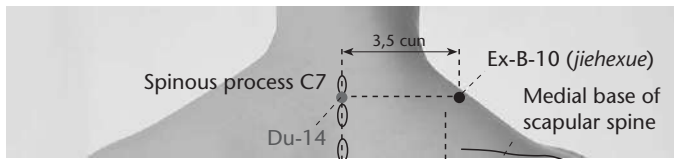

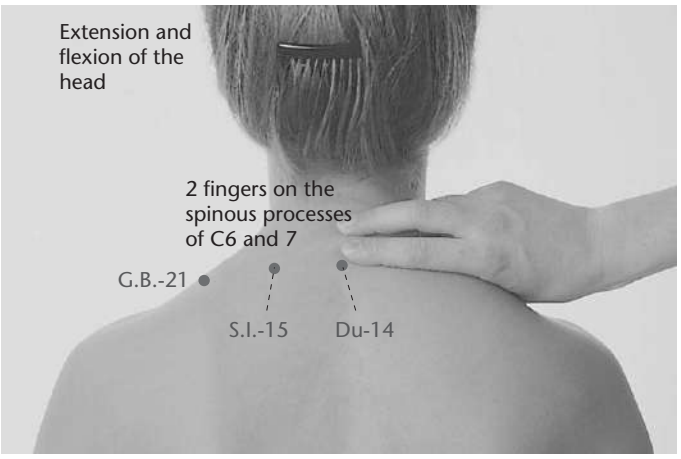

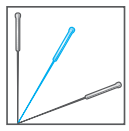
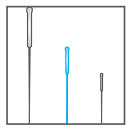
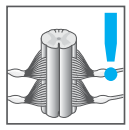
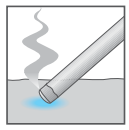
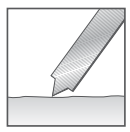
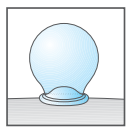

Ch06a-F10028.qxd 2/23/08 1:26 PM Page 574

Ex-B-2 *Hua Tuo's Paravertebral Points* HUATUOJIAJI

Location

17 point pairs, 0.5 cun lateral to the lower borders of the spinous processes, close to the spinal facet joints:

- 12 thoracic point pairs (*xiongjiaji*): between T1 and T12
- 5 lumbar point pairs (*yaojiaji*): between L1 and L5, Depending on the school of thought, corresponding points lateral to the cervical spine are described as 'additional *huatuojiaji*'.

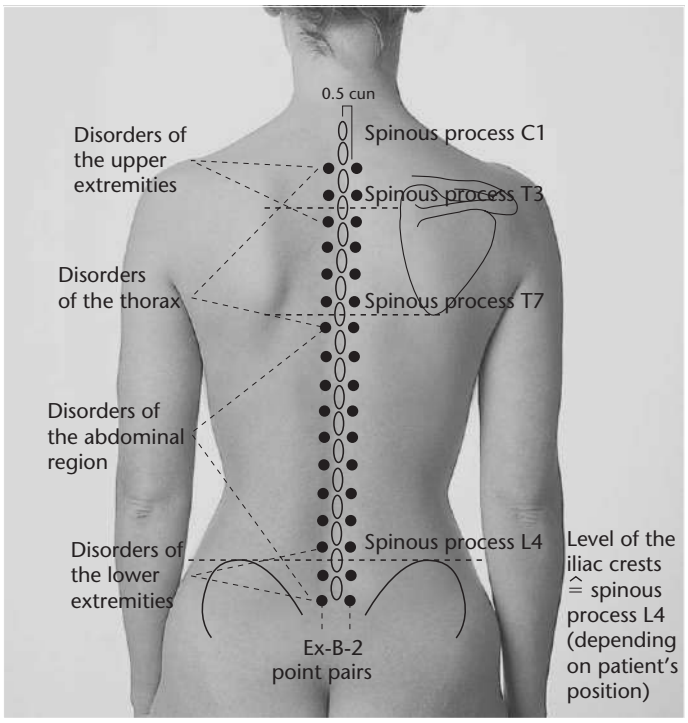

How to find

Select points of **Ex-B-2** depending on the disorder. Determine the relevant spinous process and locate the point pairs 0.5 cun lateral to its lower border.

Needling

0.5–1 cun vertically or better obliquely in a medial direction towards the spine, up to 1.5 cun in the lumbar region. The purpose of needling these points is to stimulate the local nerve roots. This means that the insertion angle has to be adapted to the patient's anatomy. Never needle in a lateral direction. Once *deqi* is obtained, do not further stimulate the needle; it is possible to apply electrostimulation.

Actions/Indications

- Regulate and harmonise the five *zang* and six *fu* Organs, depending on the points selected

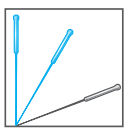
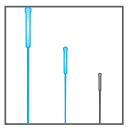

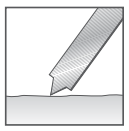
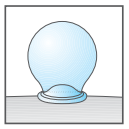

Special features

With disorders of the facet joints of the cervical spine, there are often tender points 0.5 cun lateral to the lower borders of the spinous processes. In China, these are needled for chronic disorders of the cervical spine. If oblique insertions (approximately 45°) in a medial direction are used for points on the inner branch of the Bladder channel, the tip of the needle will reach the area of the *huatuojiaji* points. This enhances the therapeutic effect.

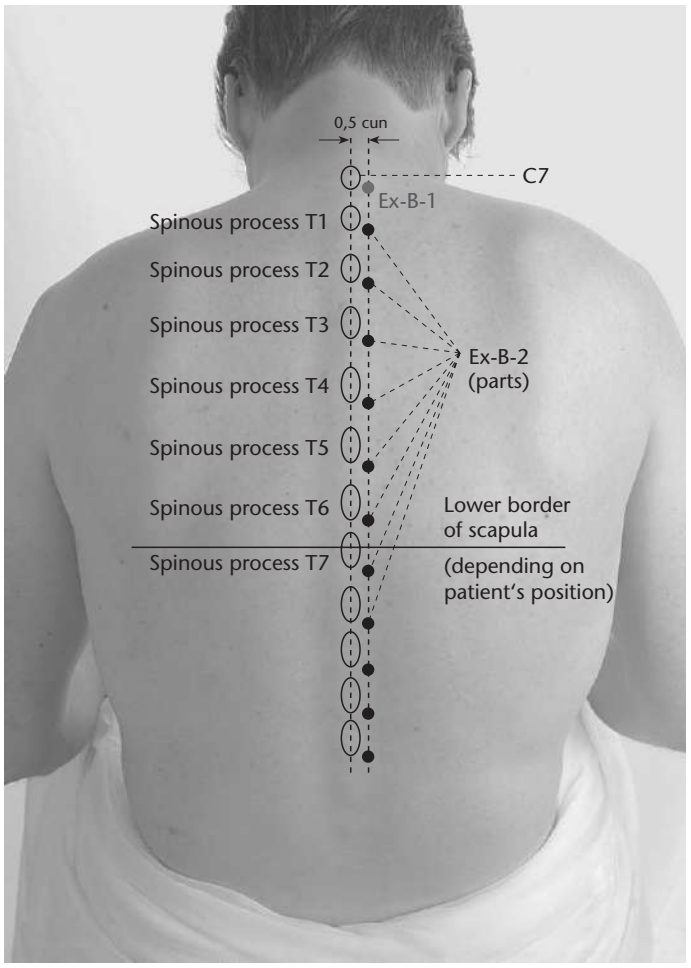

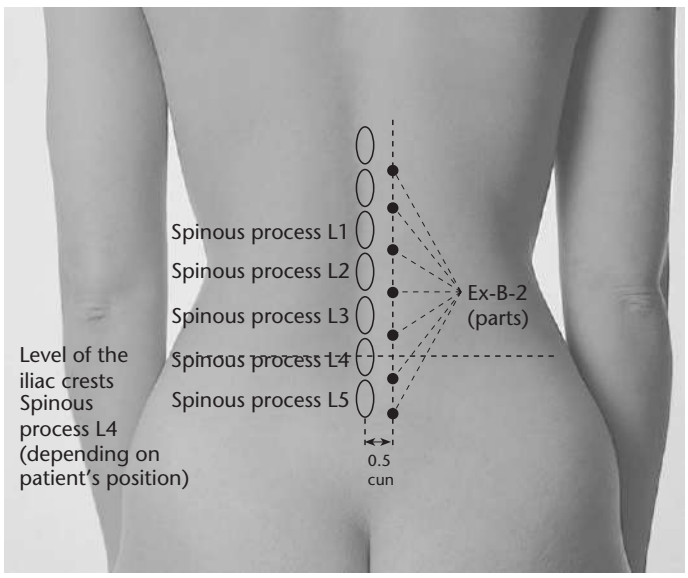

Ch06a-F10028.qxd 2/23/08 1:26 PM Page 575

Stomach Controller Lower Shu WEIWANXIASHU Ex-B-3

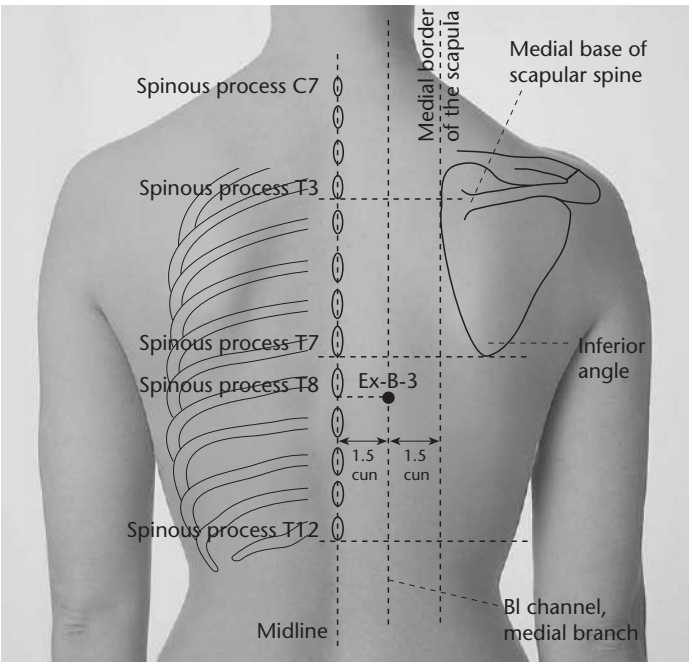

Location

1.5 cun lateral to the lower border of the spinous process of the 8th thoracic vertebra (T8).

How to find

Locate T7 (→ 3.4.2). From there, count down one spinous process (to T8) and locate **Ex-B-3** 1.5 cun lateral from the midline. Use the Tuffier's line (L4, 3.4.3) to confirm the location of this point: from L4, count up the spinous processes of L3 to L1 and of T12 to T8.

Located on the same level is a point of → **Ex-B-2** (0.5 cun lateral to the midline).

Needling

0.5–1 cun vertically or up to 1.5 cun obliquely in a medial direction. Do not needle in a lateral direction. Caution: Pneumothorax.

Actions/Indications

- Opens the channel locally
- Moistens the body and clears Heat

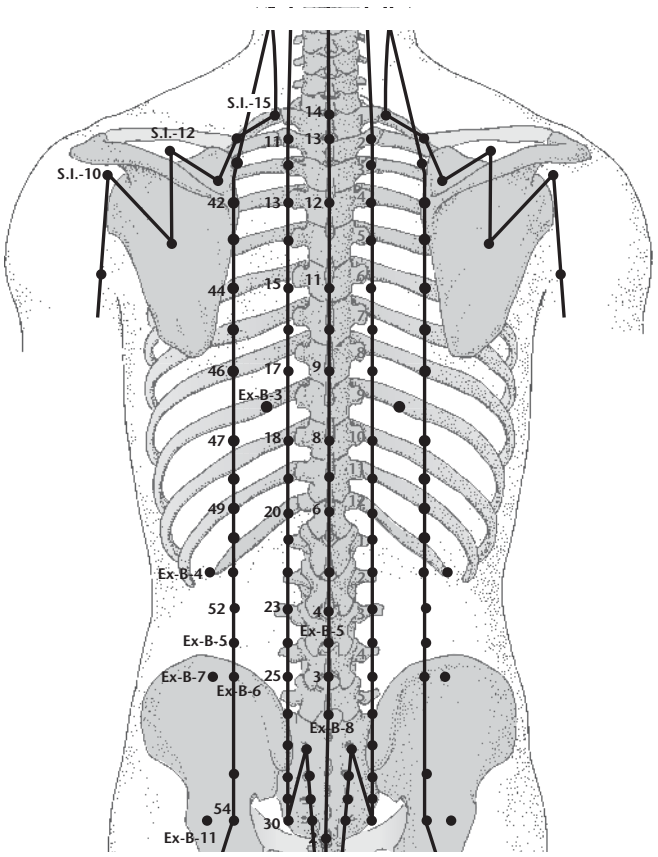

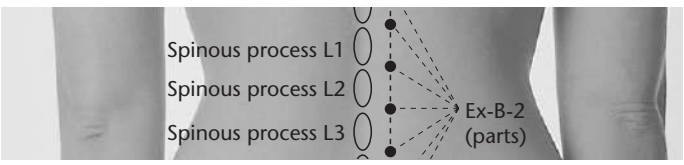

Ch06a-F10028.qxd 2/23/08 1:26 PM Page 576

Ex-B-4 *Fullness Root* PIGEN

Location

3.5 cun lateral to the posterior midline, on the level of the lower border of the spinous process of the 1st lumbar vertebra (L1).

How to find

Locate the T7 (→ 3.4.2) and count down 6 spinous processes to L1. Locate **Ex-B-4** on the level of the lower border of its spinous process, 3.5 cun lateral to the midline. Use the Tuffier's line (L4, → 3.4.3) to confirm the location of this point: from L4, count up the spinous processes of L3 to L1.
Located on the same level are → **Du-5** (on the midline), a point of → **Ex-B-2/BL-22/BL-51** (0.5/1.5/3 cun lateral to the midline) as well as → **G.B.-25** (on the lower border of the free end of the 12th rib).

Needling

0.8–1 cun obliquely in a medial direction. Caution: Injury to the kidneys!

Actions/Indications

- Improves the Qi flow between the Upper and Middle Burner
- Eliminates local Qi and Blood stagnation

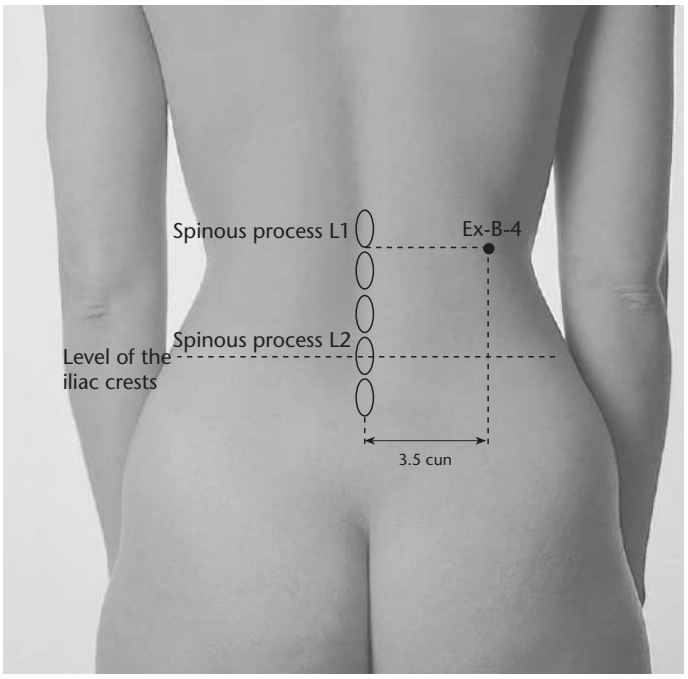

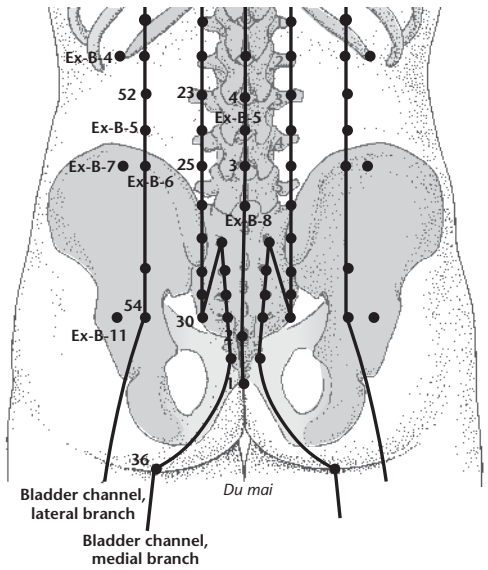

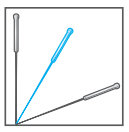

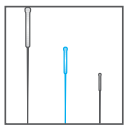

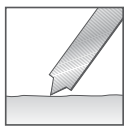

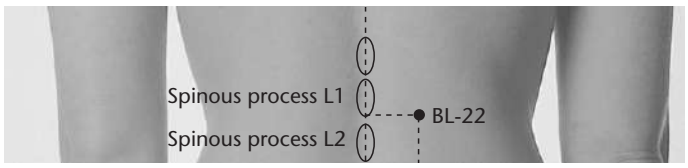

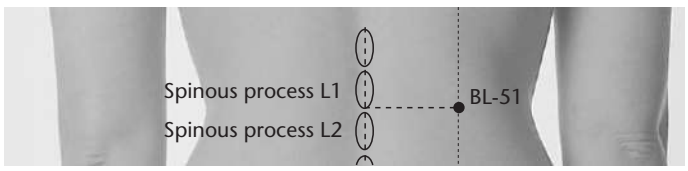

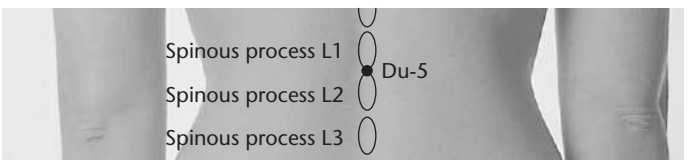

Ch06a-F10028.qxd 2/23/08 1:26 PM Page 577

Lower Shu *Point* XIAJISHU Ex-B-5

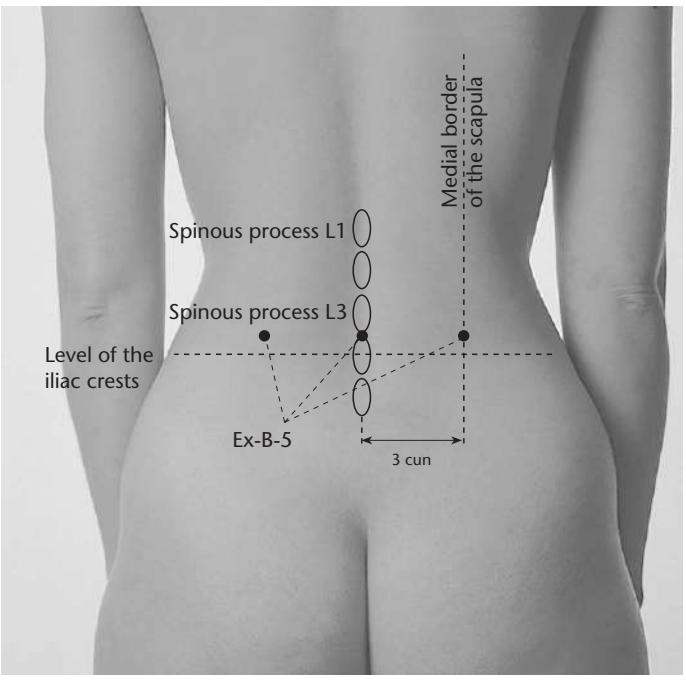

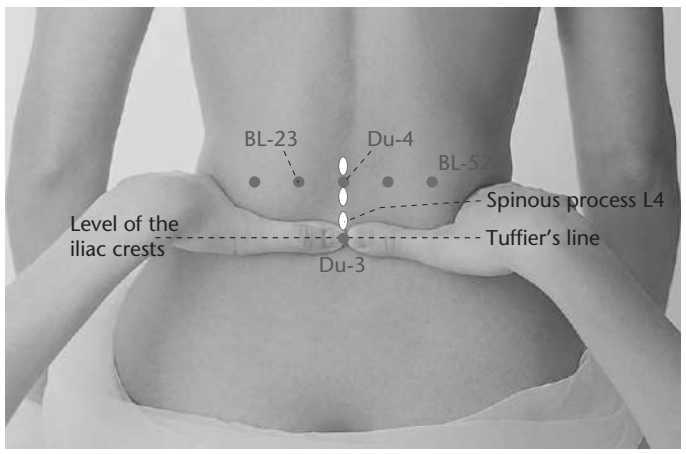

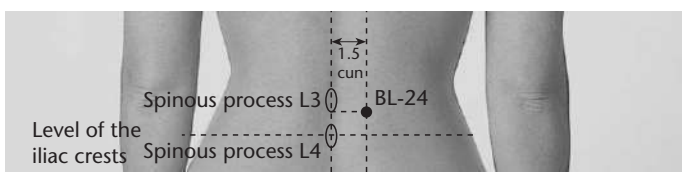

Location

Some authors consider **Ex-B-5** (*xiajishu*) to comprise three points: one point on the midline below the spinous process of the 3rd lumbar vertebra (L3) (this is the more common and single location of this extra point), complemented by two lateral points 3 cun lateral to the centre point on the midline.

How to find

Locate the Tuffier's line and L4 (→ 3.4.3). Locate **Ex-B-5** on the midline, on the lower border of the spinous process of L3. Then measure 3 cun in a lateral direction and there locate the two lateral points of **Ex-B-5.**

Located on the same level are a point of → **Ex-B-2/BL-24** (0.5/1.5 cun lateral to the midline).

Needling

0.5–0.8 cun vertically to the skin or obliquely in an inferior direction. The patient's back should be straight or overextended to avoid a spinal puncture. Oblique insertions in a superior direction should only be carried out by experienced practitioners, as in small persons (regardless of their weight), the spinal canal can be reached after only 1.25 cun. Caution: In pregnant women about to go into labour, this point might have a labour-promoting effect.

Actions/Indications

- Tonifies Kidney Yang
- Opens the channel locally

Special features

Strictly speaking, this extra point should be considered as a single point and a pair of points with differing indications: depending on the school of thought and the clinical requirements, the lateral points are used for localised Qi and Blood stagnation, whereas the central point is indicated for Kidney Yang deficiency.

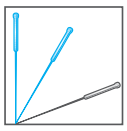
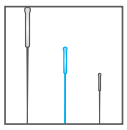
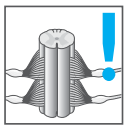

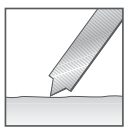
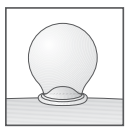

Ch06a-F10028.qxd 2/23/08 1:26 PM Page 578

Ex-B-6 *Back Pain Point* YAOYI

Location

3 cun lateral to the midline, on the level of the lower border of the spinous process of the 4th lumbar vertebra (L4).

How to find

Locate the Tuffier's line and L4 (→ 3.4.3). On the level of the lower border of the spinous process of L4, measure 3 cun in a lateral direction and there locate **Ex-B-6.** Located on the same level are → **Du-3** (on the midline), a point of → **Ex-B-2/BL-25/ Ex-B-7** (0.5/1.5/3.5 cun lateral to the midline).

Needling

Vertically 0.5–0.8 cun

Actions/Indications

- Opens the channel and *luo* vessels

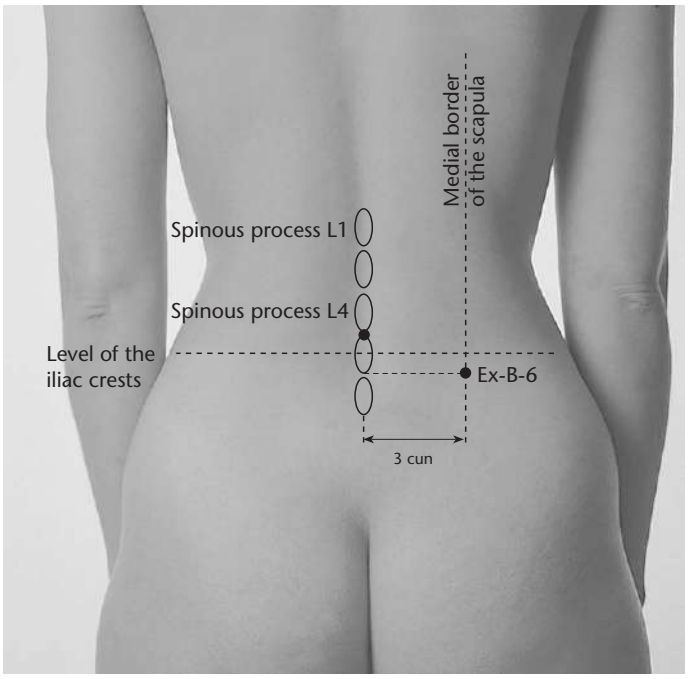

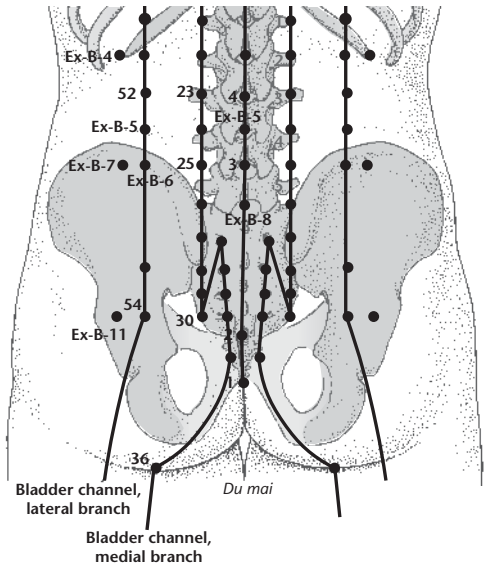

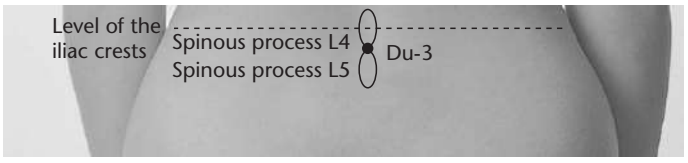

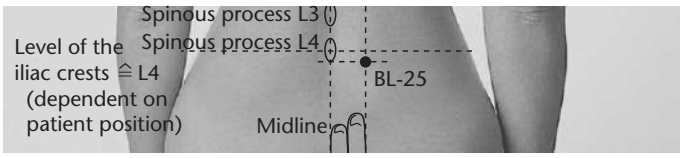

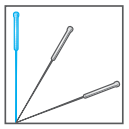
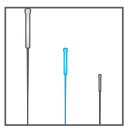

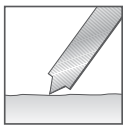

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 579

Lumbar Eyes YAOYAN Ex-B-7

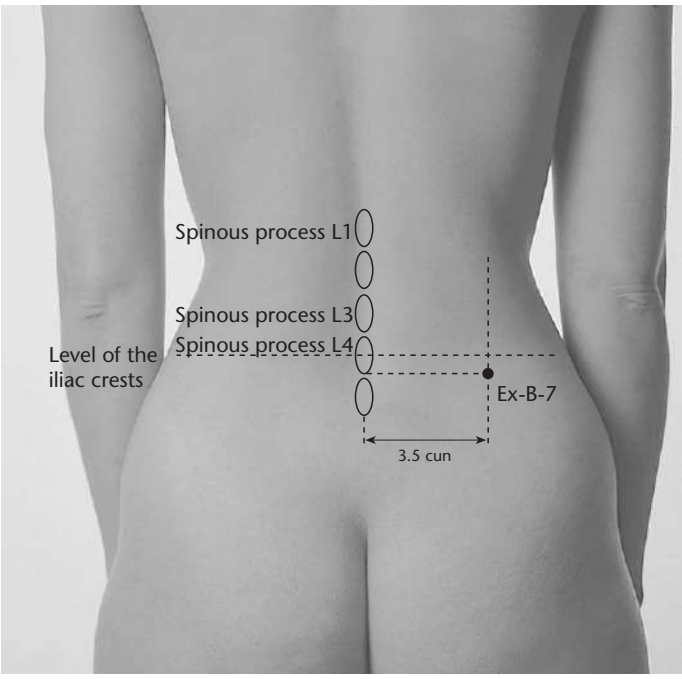

Location

3.5 cun lateral to the midline, on the level of the lower border of the spinous process of the 4th lumbar vertebra (L4).

How to find

Locate the Tuffier's line and L4 (→ 3.4.3). **Ex-B-7** is located 3.5 cun lateral to it. Located on the same level are → **Du-3** (on the midline), a point of → **Ex-B-2/BL-25/Ex-B-6** (0.5/1.5/3 cun lateral to the midline).

Needling

Vertically 0.5–0.8 cun

Actions/Indications

- Opens the channel locally
- Tonifies the Kidneys

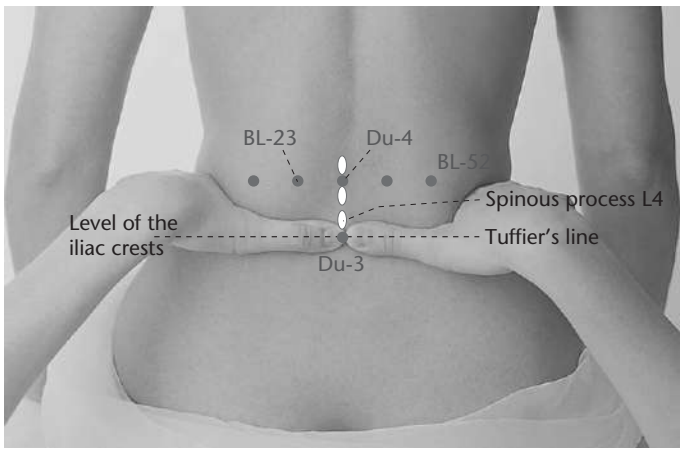

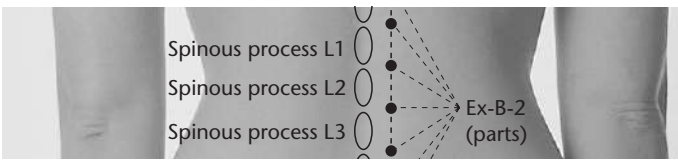

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 580

Ex-B-8 *Below the 17th Vertebra* SHIQIZHUI/SHIQIZHUIXIA

Location

On the midline, below the spinous process of the 5th lumbar vertebra (L5).

How to find

Locate the Tuffier's line and L4 (→ 3.4.3). Next, palpate in an inferior direction to the next vertebra (L5) and locate **Ex-B-8** below the border of its spinous process, at the lumbosacral joint. Located on the same level are a point of → **Ex-B-2/BL-26** (0.5/1.5 cun lateral to the midline).

Needling

0.5–1 cun vertically to the skin or obliquely in an inferior direction. The patient's back should be straight or slightly flexed to avoid a spinal puncture. Oblique insertions in a superior direction should only be carried out by experienced practitioners, as in small persons (regardless of their weight), the spinal canal can be reached after only 1.25 cun. Caution: In pregnant women about to go into labour, this point might have a labour-promoting effect.

Actions/Indications

- Tonifies the Kidneys
- Opens the channel locally

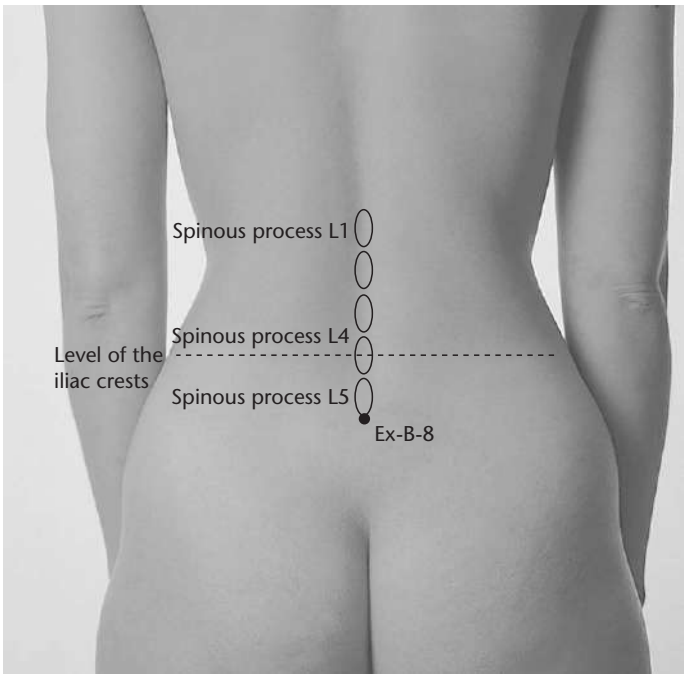

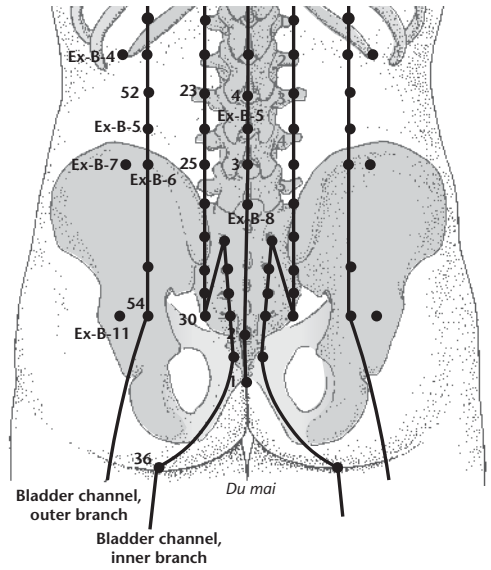

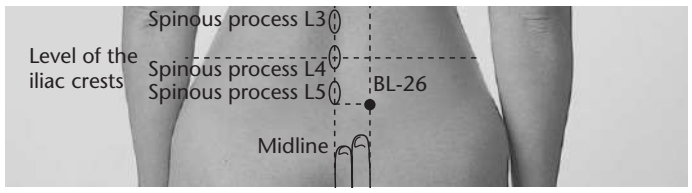

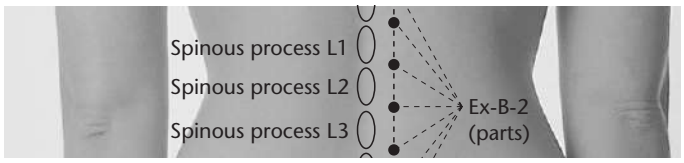

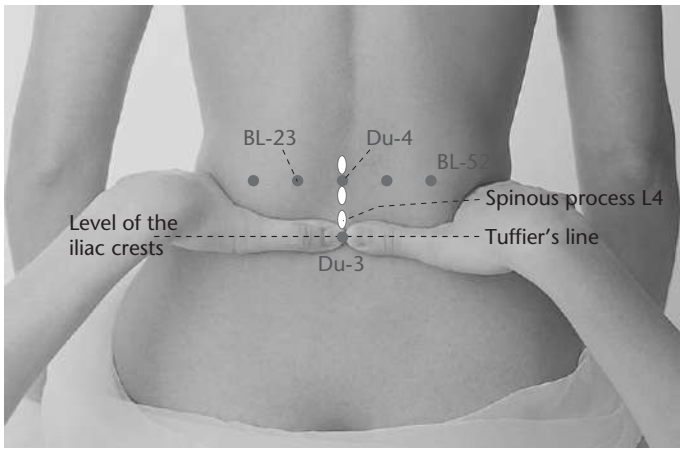

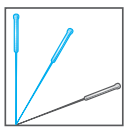
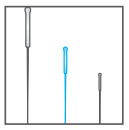
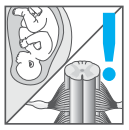

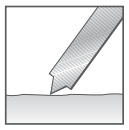

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 581

Miraculous Lumbar Point YAOQI Ex-B-9

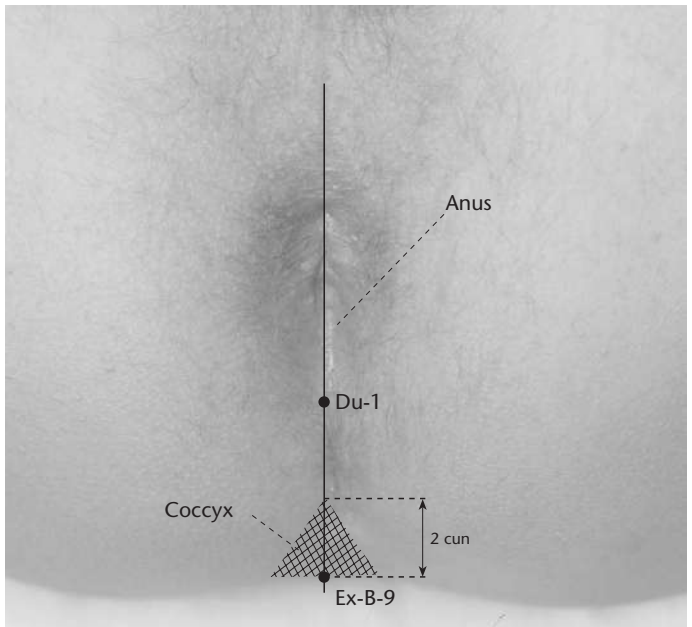

Location

2 cun superior to the tip (inferior end) of the coccyx.

How to find

Locate the tip of the coccyx superior to the anus and palpate for the sacral hiatus. From there, palpate 2 cun in a superior direction and locate **Ex-B-9** in a depression between the sacral processes.

→ **Du-1** is located on the midline, between the tip of the coccyx and the anus.

Needling

Up to 1.5 cun obliquely in a superior direction.

Actions/Indications

- Calms the *shen*

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 582

Ex-B-10 *Tuberculosis Point* JIEHEXUE

Location

3.5 cun lateral to the lower border of the spinous process of the 7th cervical vertebra (C7).

How to find

Locate the spinous process of C7 and, on this level, palpate 3.5 cun in a lateral direction. There, locate **Ex-B** (*jiehexue*).
Located on the same level are → **Du-14** (on the midline), → **Ex-B-2/S.I.-15** (0.5/2 cun lateral to the midline).

Needling

Vertically 0.5–0.8 cun. Caution: Pneumothorax.

Actions/Indications

- Tonifies the Lung
- Opens the channel locally

Special features

This is an unofficial extra point.

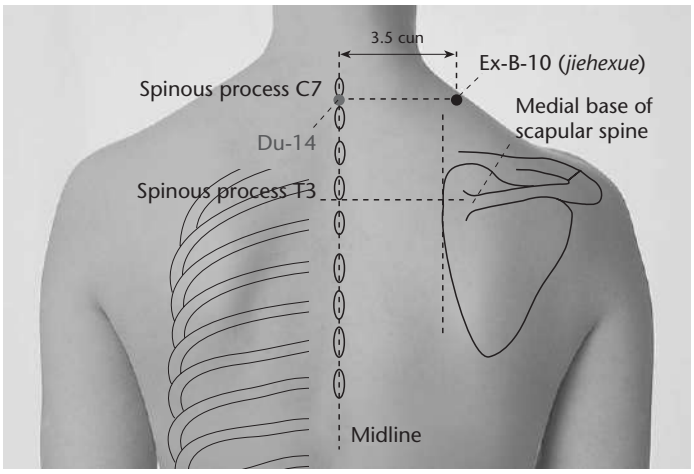

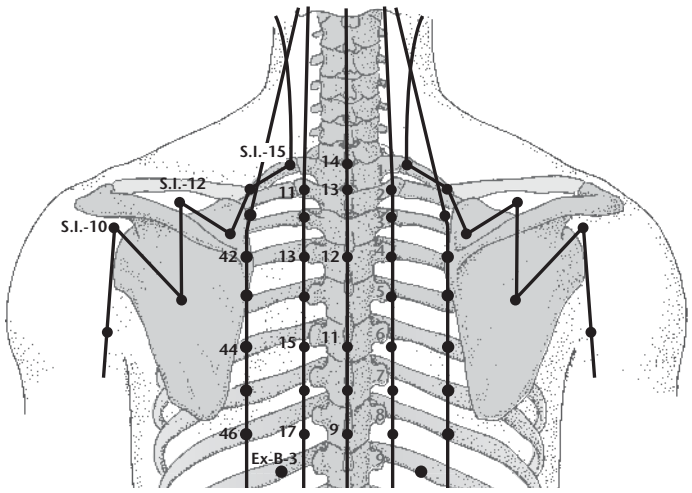

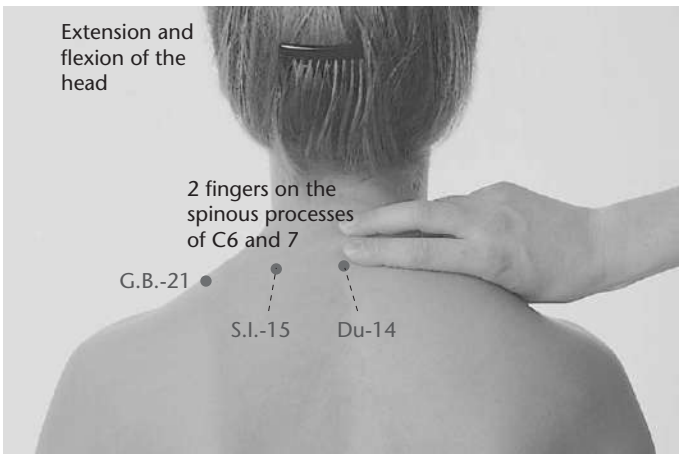

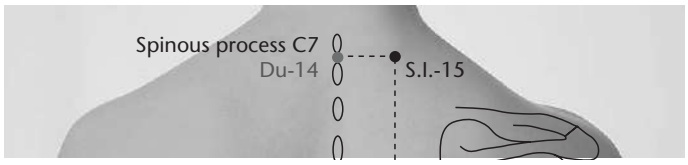

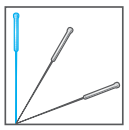

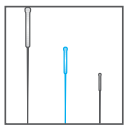

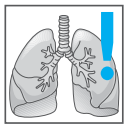

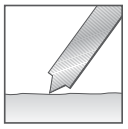

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 583

Buttock Centre TUNZHONG

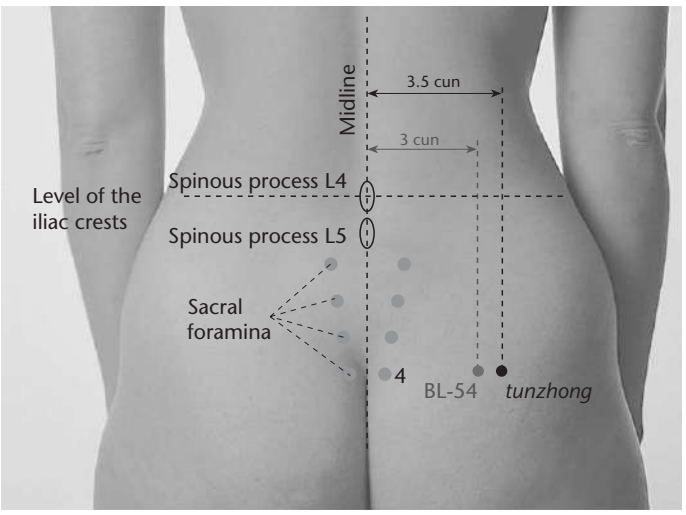

Location

At the centre of the buttock, 3.5 cun lateral to the posterior midline, on the level of the 4th sacral foramen.

How to find

Locate the lumbosacral joint, which often forms an easily palpable 'kink' between the L5 and the sacrum. Alternatively, locate the Tuffier's line (→ 3.4.3) for reference. Then palpate the inferior aspect of the sacrum for the sacral hiatus (→ 3.4.4). Between these two reference points, the four pairs of sacral foramina are arranged on either side of the midline in a slight V-shaped form and can be palpated as shallow depressions. **Ex-B** (*tunzhong*) is located 3.5 cun lateral to the lowest depression (=4th foramen), approximately at the centre of the buttock.
Located on the same level are → **BL-34/BL-30/BL-54** (over the 4th foramen/1.5/3 cun lateral to the midline).

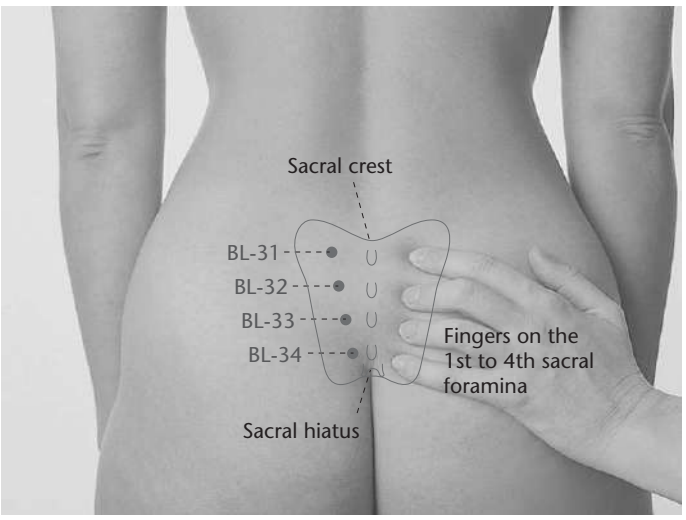

Needling

Vertically 2–3 cun

Actions/Indications

- Opens the channel and *luo* vessels locally

Special features

This is an unofficial extra point.

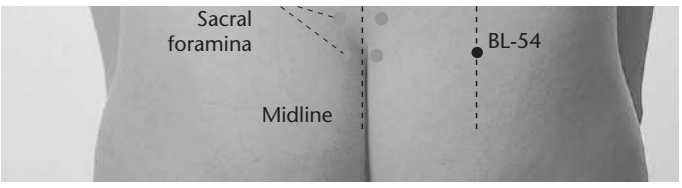

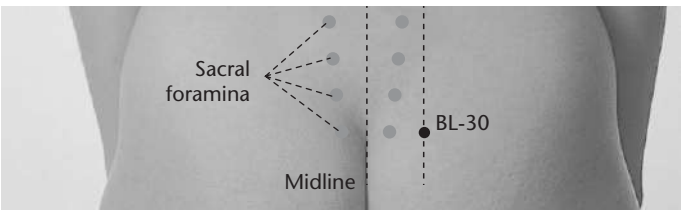

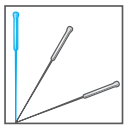
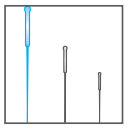

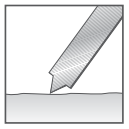
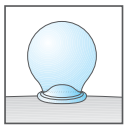

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 585

6.4 Extra Points: Upper Extremities (EX-UE)

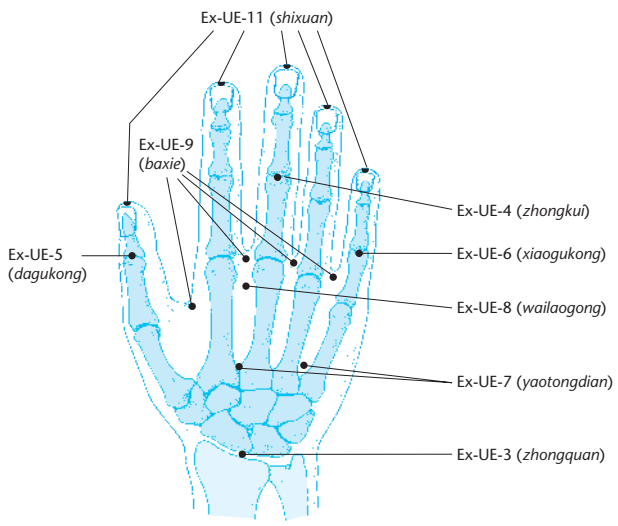

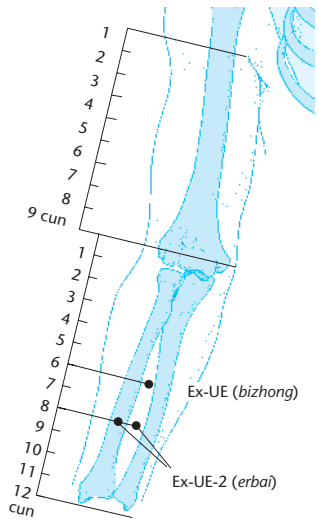

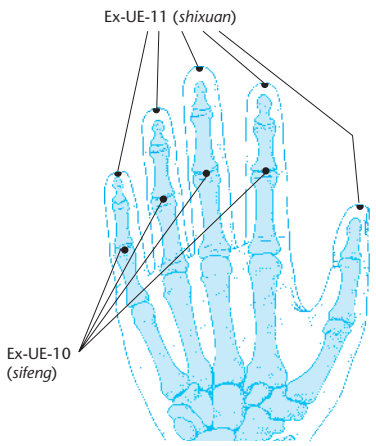

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 586

English abbreviation (standard)	*Pinyin* name	Nguyen Van Nghi, König and Wancura, Schnorrenberger	Shanghai College	Ex (Hempen)
Ex-UE-1	*zhoujian*		M-UE-46	
Ex-UE-2	*erbai*	PaM or ZP114	M-UE-29	
Ex-UE-3	*zhongquan*	PaM or ZP118	M-UE-33	
Ex-UE-4	*zhongkui*	PaM101	M-UE-16	
Ex-UE-5	*dagukong*	PaM or ZP100	M-UE-15	
Ex-UE-6	*xiaogukong*	PaM or ZP102	M-UE-17	
Ex-UE-7	*yaotongxue/yaotongdian*	Handpoint 2 PaM or ZP110/111*	M-UE	Ex18
Ex-UE-8	*wailaogong*†/*luozhen*	PaM or ZP108†	M-UE-24	Ex17
Ex-UE-9	*baxie*	PaM or ZP107	M-UE-22	Ex14
Ex-UE-10	*sifeng*		M-UE-9	Ex13
Ex-UE-11	*shixuan*	PaM or ZP86	M-UE-1	Ex15
Further Extra points				
Ex-UE	*jianqian/jianneiling*	Part of NP74‡ (*sanjian*): *jianyu* (L.I.-15), *jianqian*, *jianhou*	M-UE-48§	Ex16
Ex-UE	*bizhong*	PaM or ZP115	M-UE-30	

* The 'standard' point Ex-UE-7 (*yaotongdian* or *yaotongxue*) is described by Nguyen Van Nghi and Schnorrenberger as two separate points with different *pinyin* names: PaM or ZP110 (*weiling*) and 111 (*jingling*). Their location also corresponds to handpoint 2.

† Nguyen Van Nghi and Schnorrenberger distinguish between *luozhen* (PaM or ZP108) and *wailaogong* (PaM or ZP109). They describe the location of *luozhen* as identical with the 'standard' location of Ex-UE-8 (*wailaogong*): on the dorsum of the hand, between the 2nd and 3rd metacarpal bones, approximately 0.5 cun proximal to the metacarpophalangeal joint. 'Their' *wailaogong* is located at the midpoint of a line connecting the wrist and the head of the 3rd metacarpal bone, between the metacarpals, directly opposite to the point *laogong* (P-8) – hence the name *wailaogong* ('outer' *laogong*).

‡ **NP74** (*jiansanzhen*/Three pinpricks at the Shoulder) according to Schnorrenberger: a set of three points: the first is identical with **L.I.-15** (*jianyu*), the second (*jianquan*/ Anterior to the Shoulder) is located 1 cun superior to the end of the anterior axillary fold, and the third (*jianhou*/Posterior to the Shoulder) is located 1 cun superior to the end of the posterior axillary fold.

§ According to Deadman et al. (1998), **M-UE-48** (*jianqian*) is located midway between the end of the anterior axillary fold and **L.I.-15** (this point is not mentioned by the Shanghai College). See the point page for the location of this point.

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 587

Elbow Tip ZHOUJIAN Ex-UE-1

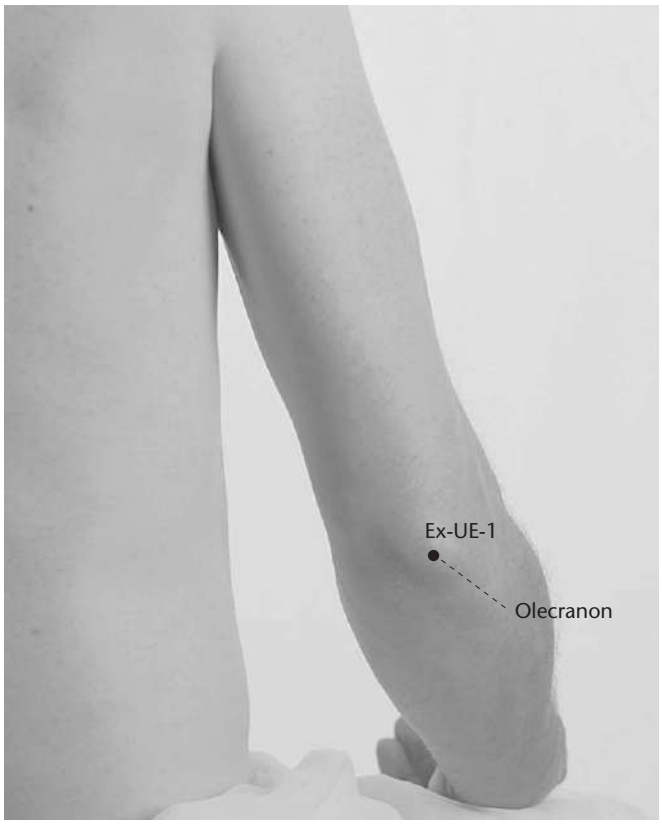

Location
On the tip of the olecranon.

How to find
On the tip of the olecranon. This point is best located with the elbow flexed.

Needling
Moxibustion only!

Actions/Indications
- Opens the channel and *luo* vessels locally
- Disperses Phlegm locally

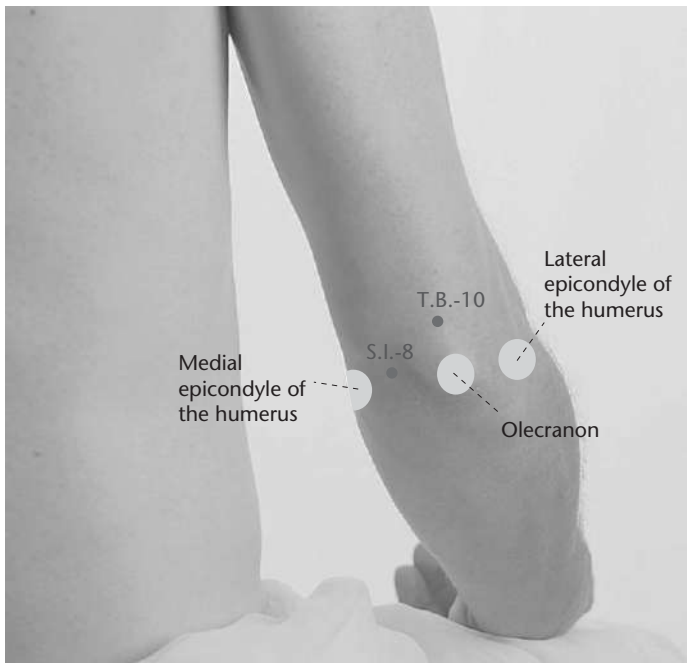

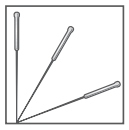
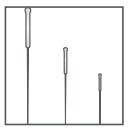

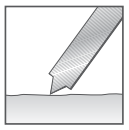

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 588

Ex-UE-2 *Two Whites* ERBAI

Location

A pair of points on the palmar aspect of the forearm, 4 cun proximal to the wrist joint space (most distal wrist crease), on either side of the tendon of the flexor carpi radialis muscle.

How to find

Place the patient's arm in a relaxed supinated position. As the location of the wrist crease can vary, the wrist joint space provides a more reliable point of reference (→ 3.3.3). From the centre of the wrist joint space, measure 4 cun in a proximal direction and, on this level, locate one of the points lateral to and the other point medial to the tendon of the flexor carpi radialis muscle. **Or:** Divide the distance between the cubital crease (at → **P-3**) and the wrist joint space (at → **P-7**) into thirds, then locate → **Ex-UE-2** one third of the distance from the wrist joint (helpful tool: elastic tape → 2).

Needling

Up to 1 cun vertically or up to 1.5 cun obliquely towards proximal.

Actions/Indications

- Raises Qi to treat prolapse of the rectum and haemorrhoids

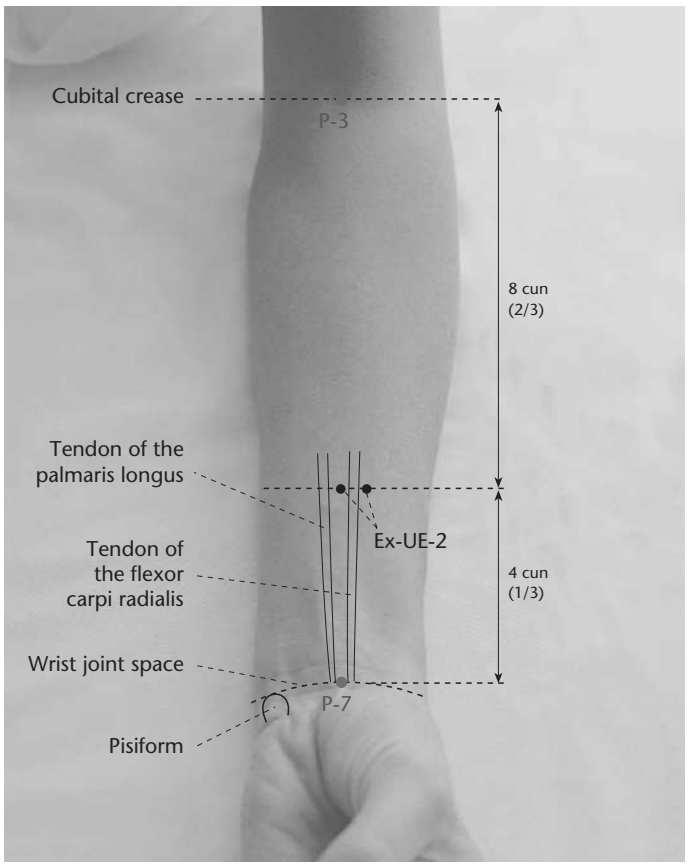

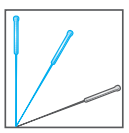

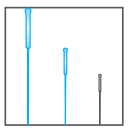

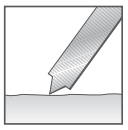

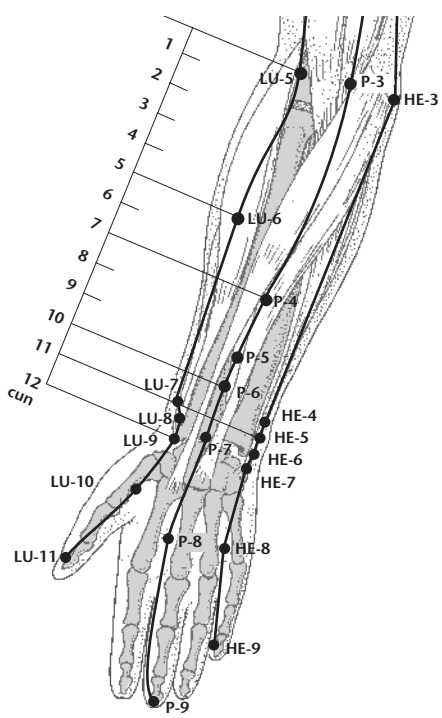

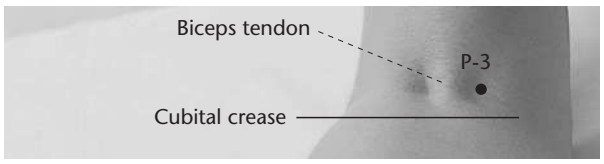

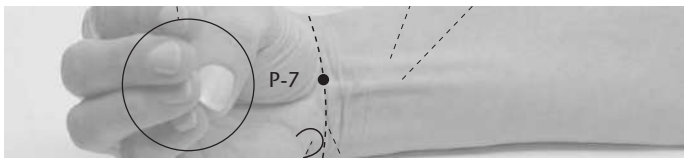

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 589

Posterior Spring ZHONGQUAN Ex-UE-3

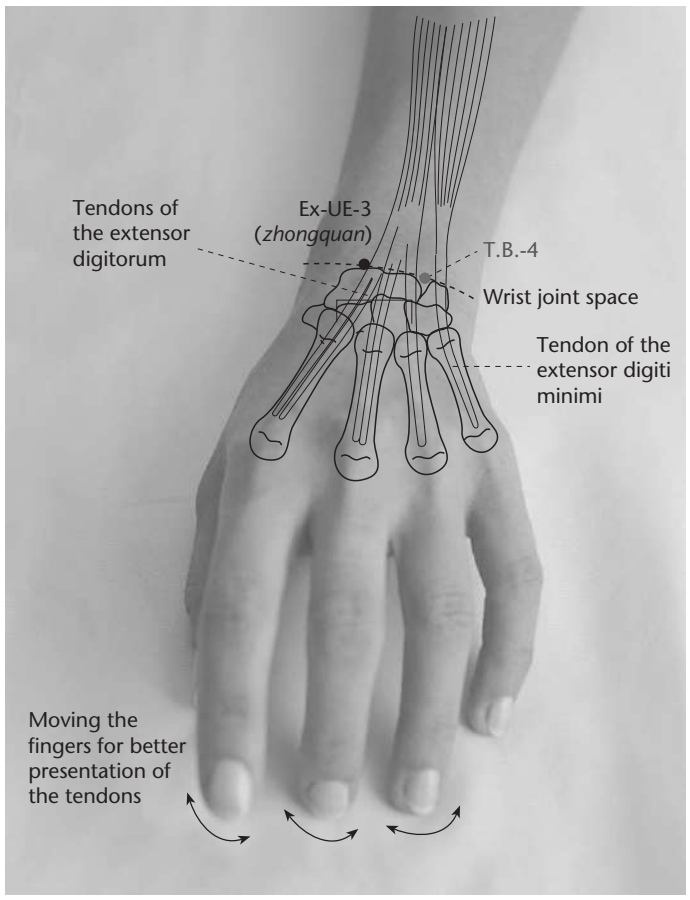

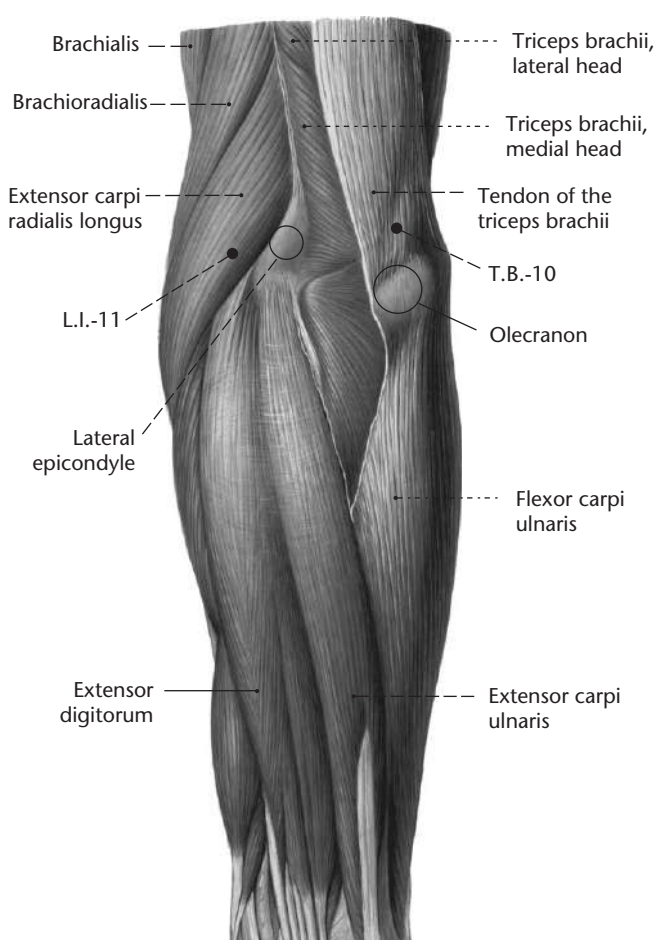

Location

On the dorsal aspect of the wrist joint space (dorsal wrist crease), radial to the tendon of the extensor digitorum communis muscle.

How to find

Place the patient's arm in a relaxed pronated position. As the location of the dorsal wrist crease can vary, the joint space between the proximal row of carpals and the radius/ulna provides a more reliable point of reference (→ 3.3.3). By moving the wrist in a relaxed way, you can easily palpate the joint space. Locate **Ex-UE-3** in the joint space, radial to the tendon of the extensor digitorum communis muscle. The tendons become more pronounced by moving and overextending the three medial fingers. → **T.B.-4** is located on the same level, ulnar to the tendon of the extensor digitorum communis muscle.

Needling

Vertically 0.3–0.5 cun

Actions/Indications

- Harmonises the Qi flow between the Upper and Middle Burner

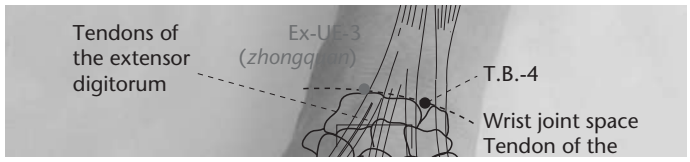

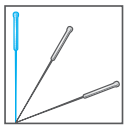
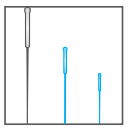

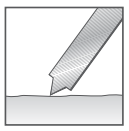

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 590

Ex-UE-4 *Back of the Middle Finger* ZHONGKUI

Location

As the name implies, on the dorsal aspect of the middle finger, in the centre of the transverse creases of the proximal interphalangeal joint (PIP).

How to find

Slightly bend the PIP of the middle finger. The point is located at the vertex of the flexed joint.
Located in a comparable position on the little finger is → **Ex-UE-6** (*xiaokugong*, in the centre of the proximal PIP of the little finger). Also located in a comparable position is → **Ex-UE-10** (*sifeng*, four points on the palmar aspect of the fingers of each hand, in the centre of the creases of the PIP joints of the 2nd to 5th fingers).

Needling

Prick to bleed or moxibustion.

Actions/Indications

- Descends Stomach Qi
- Clears Heat

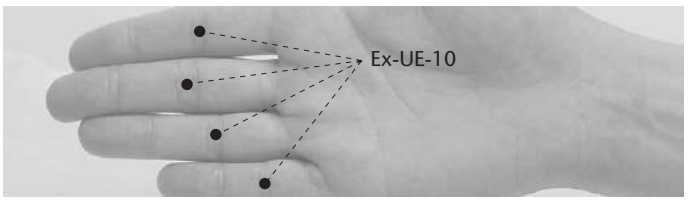

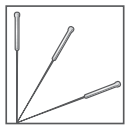

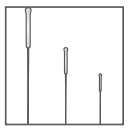

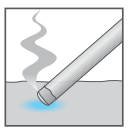

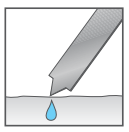

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 591

Thumb Joint DAGUKONG Ex-UE-5

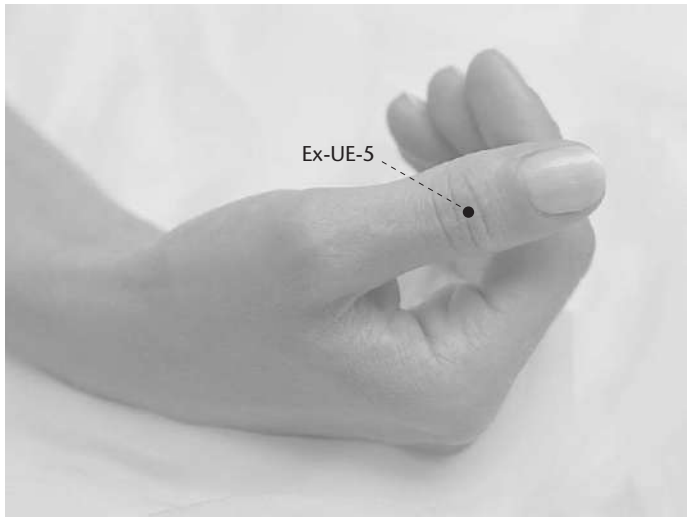

Location

As the name implies, on the dorsal aspect of the thumb, in the centre of the transverse creases of the interphalangeal joint.

How to find

Ask the patient to slightly flex the interphalangeal joint of the thumb. The point is located at the vertex of the bent joint.

Needling

Prick to bleed or moxibustion.

Actions/Indications

- Clears Heat
- Harmonises the Middle Burner

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 592

Ex-UE-6 *Little Finger Joint* XIAOGUKONG

Location

As the name implies, on the dorsal aspect of the little finger, in the centre of the transverse creases of the proximal interphalangeal joint (PIP).

How to find

Slightly bend the PIP of the little finger. The point is located at the vertex of the flexed joint.

Located in a comparable position on the middle finger are → **Ex-UE-4** (in the centre of the proximal PIP of the middle finger) and, on the palmar aspect of the little finger, → **Ex-UE-10** (four points on the palmar aspect of the fingers of each hand, in the centre of the creases of the PIP joints of the 2nd to 5th fingers).

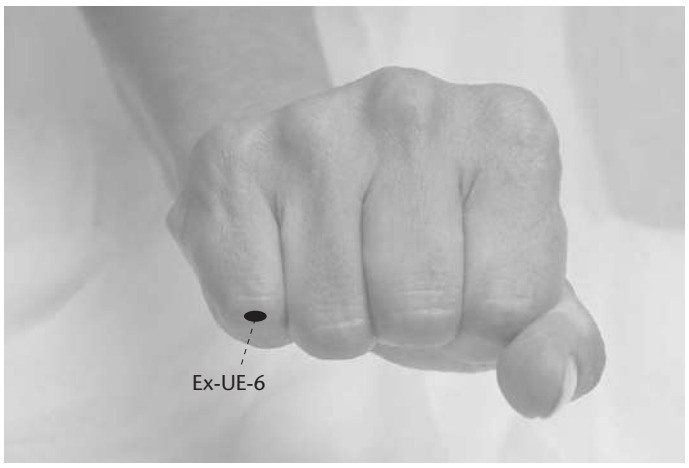

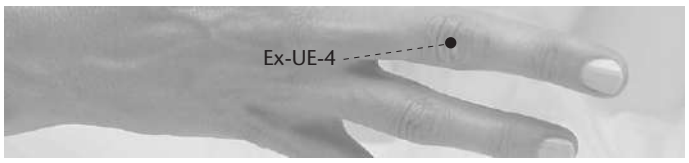

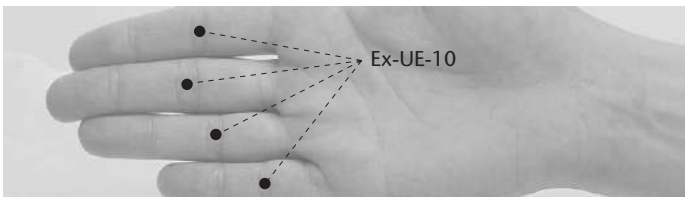

Needling

Prick to bleed

Actions/Indications

- Clears Heat

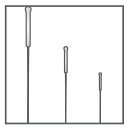

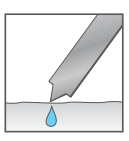

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 593

Lumbar Pair Point YAOTONGDIAN Ex-UE-7

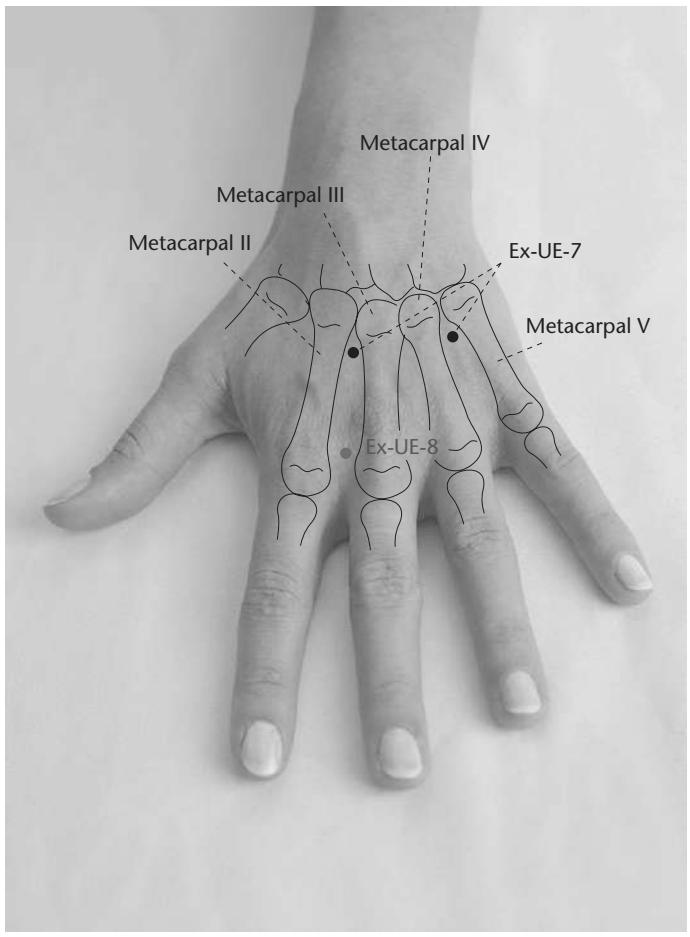

Location

Two points on the dorsal aspect of the hand, between the 2nd/3rd and 4th/5th metacarpal bones, on the level of the junctions of the shaft and the base of the respective metacarpal bones.

How to find

On the dorsal aspect of the hand, glide with the palpating fingers in the grooves between the 2nd and 3rd as well as between the 4th and 5th metacarpal bones towards the wrist joint, until the fingers come to rest in the depressions just distal to the bases of the metacarpal bones. Here, locate the points of **Ex-UE-7** (*yaotongdian* or *yaotongxue*).

Needling

0.5–0.8 cun vertically or slightly obliquely towards the centre of a loose fist. Caution: Painful point! Do not stimulate these distal points in patients with a weak constitution: danger of needle collapse!

Actions/Indications

- Strengthen the Qi and Blood in the lumbar region, alleviate pain

Special features

Very effective distal point for acute disorders of the lumbar spine

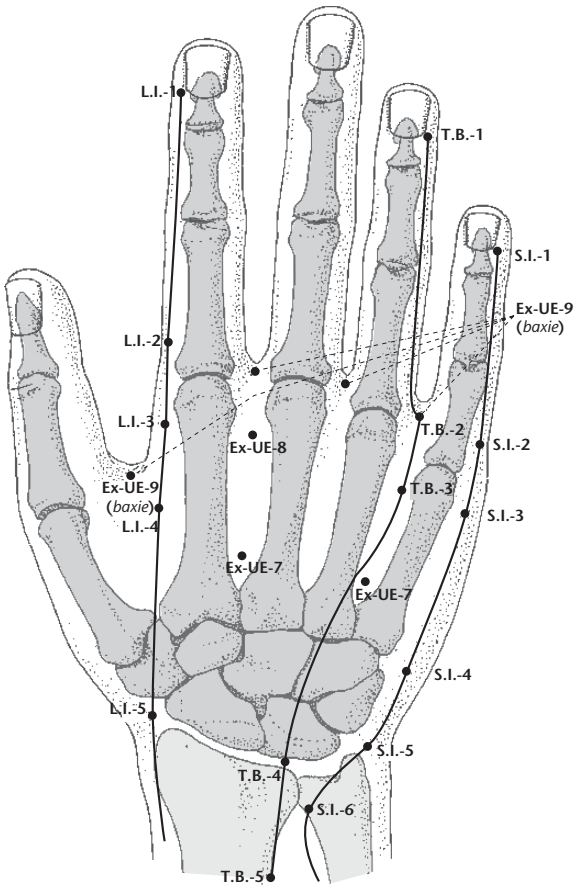

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 594

Ex-UE-8 *Stiff Neck* WAILAOGONG/LUOZHEN/XIANQIANG

Location

On the dorsal aspect of the hand, between the 2nd and 3rd metacarpal bones, proximal to the metacarpophalangeal joints, at the junction of the heads and the shafts of the metacarpal bones.

How to find

This point is best located with the patient making a loose fist. With the palpating finger, glide in the groove between the 2nd and 3rd metacarpal bones from the wrist joint towards the fingers until the finger comes to rest in the depression just proximal to the heads of the two metacarpal bones. Here, locate **Ex-UE-8** (*wailaogong*). Located in a comparable position between the 4th and 5th metacarpal bones is → **T.B.-3**.

Needling

0.5–1 cun vertically or slightly obliquely. Caution: Painful point! Do not stimulate this point in patients with a weak constitution: danger of needle collapse!

Actions/Indications

- Strengthens the Qi and Blood of the neck and alleviates pain

Special features

Very effective distal point for acute disorders of the cervical spine. Use strongly reducing needling techniques (for 1–2 minutes) on the affected or contralateral side while asking the patient to gently move their neck/cervical spine. The practitioner should keep the patient's hand in a fixed position and ensure that the hand and forearm are sufficiently supported.

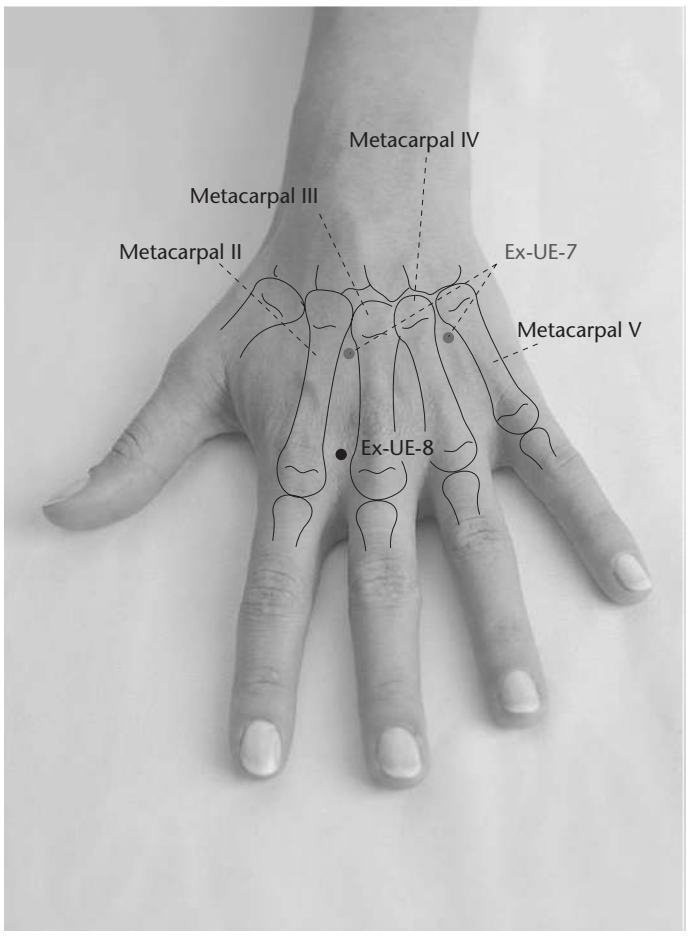

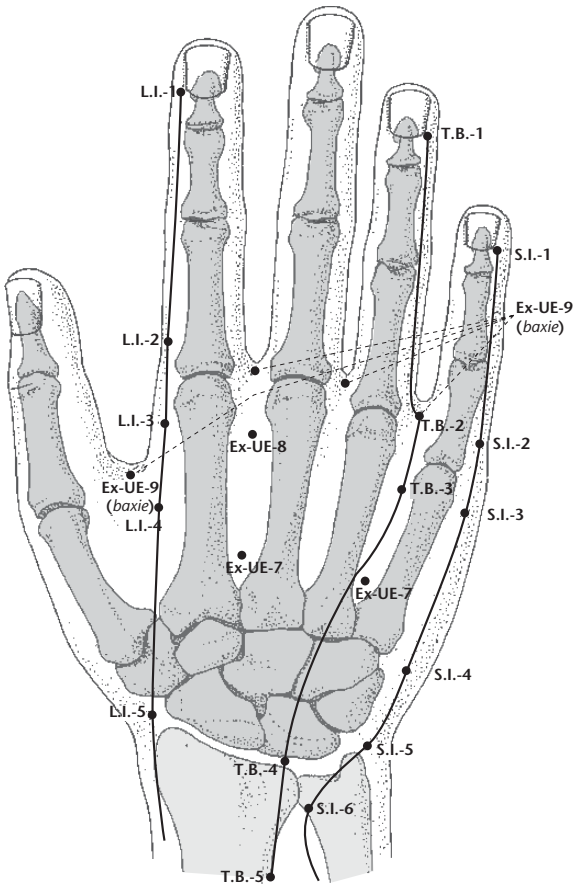

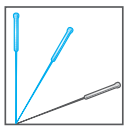
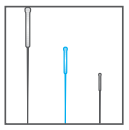

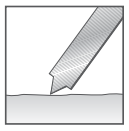

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 595

Eight Pathogens BAXIE Ex-UE-9

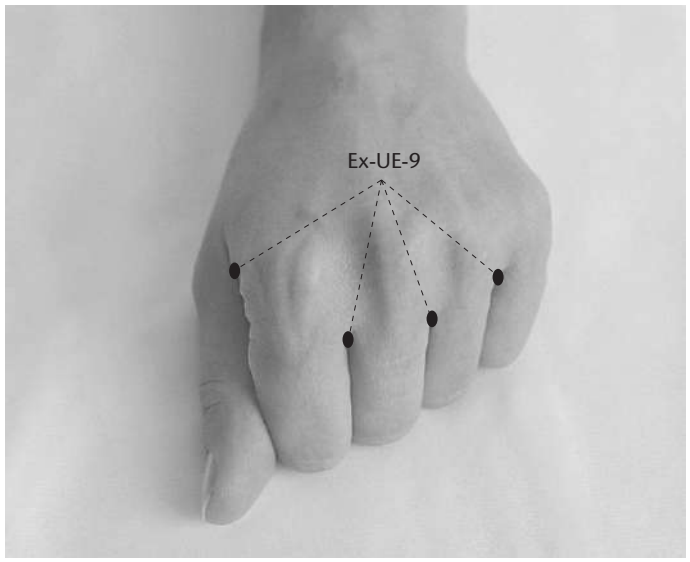

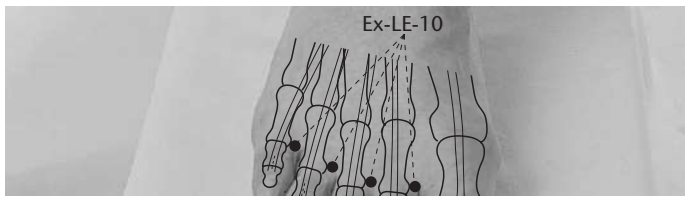

Location

Slightly proximal to the margins of the webs between the fingers, on the border of the red and white skin.

How to find

These points are best located when a loose fist is made. Locate the points on the border between the red and white skin. → **T.B.-2** (between the 4th and 5th fingers) is a point of **Ex-UE-9** (*baxie*). The points of → **Ex-LE-10** (*bafeng*) are located in a comparable position on the foot (proximal to the margins of the webs between the toes). → **LIV-2,** → **ST-44** and → **G.B.-43** are all part of → **Ex-LE-10**.

Needling

Up to 1 cun parallel to the metacarpal bones.

Actions/Indications

- Clears Heat
- Expels Cold and Dampness

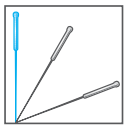
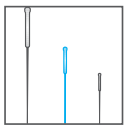

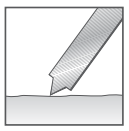

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 596

Ex-UE-10 *Four Seams* SIFENG

Location

On the palmar aspect of the 2nd to 5th fingers, at the midpoint of the transverse creases of the proximal interphalangeal joints (PIP).

How to find

Locate these points on the palmar aspect of the 2nd to 5th fingers, at the midpoint of the transverse creases of the proximal interphalangeal joints (PIP).
Located in a similar position on the dorsal aspect of the little and middle fingers are → **Ex-UE-6** and → **Ex-UE-4** respectively.

Needling

Prick to bleed

Actions/Indications

- Harmonises the Qi flow between the Upper and Middle Burner

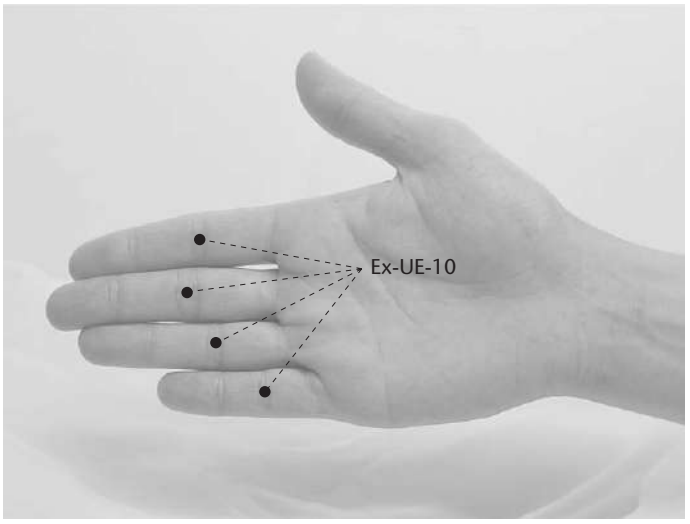

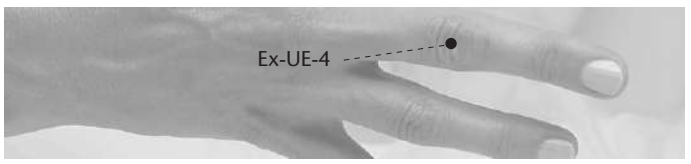

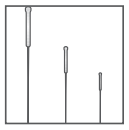

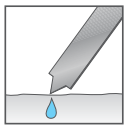

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 597

Ten Diffusions SHIXUAN Ex-UE-11

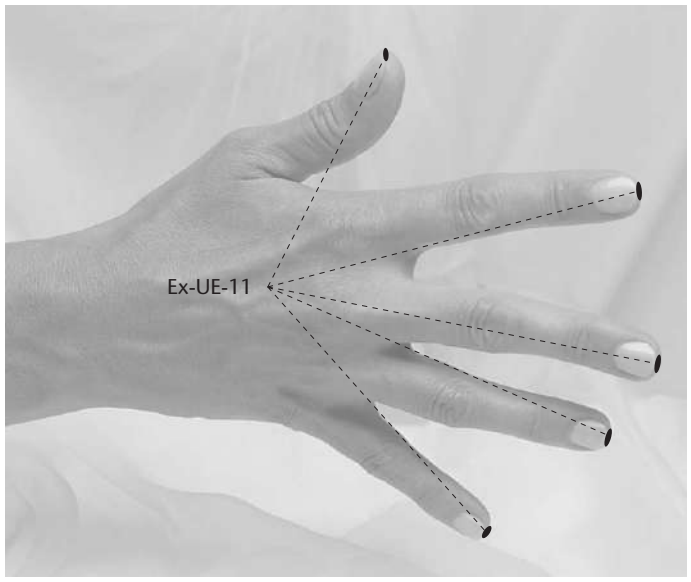

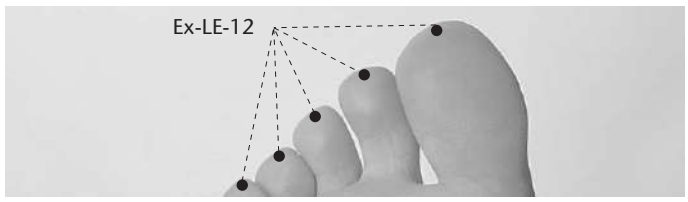

Location

On the tips of the 10 fingers.

How to find

These 10 points are located on the tips of the 10 fingers, approximately 0.1 cun from the free margin of the fingernail.
Located in a comparable position on the foot are the points of → **Ex-LE-12** (*qiduan*, on the tips of the 10 toes).

Needling

Prick to bleed

Actions/Indications

- Expels pathogenic factors

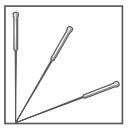
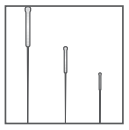

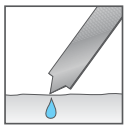

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 598

Ex-UE Front of the Shoulder JIANQIAN/JIANNEILING

Location

On the midpoint of a line connecting the end of the anterior axillary fold and ➞ **L.I.-15**.

How to find

To locate ➞ **L.I.-15**: Ask the patient to actively abduct their arm and locate **L.I.-15** in the depression anterior and inferior to the acromion (=ventrolateral extremity of the acromion). The end of the anterior axillary fold is formed by the easily palpable lower border of the pectoralis major muscle. *Jianqian* is located on the anterior aspect of the shoulder joint, midway between these two reference points.

Needling

Vertically up to 1.5 cun

Actions/Indications

- Opens the channel and *luo* vessels locally

Addendum: For more details, ➞ Table 6.4

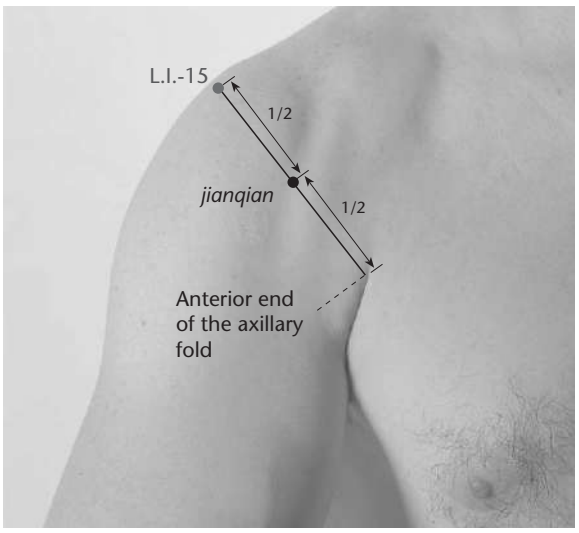

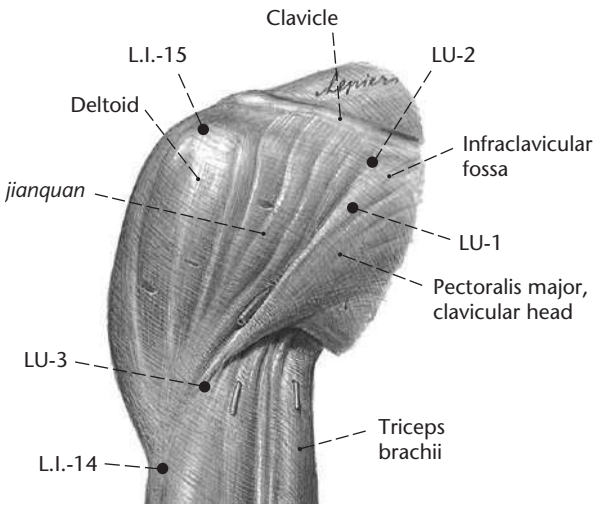

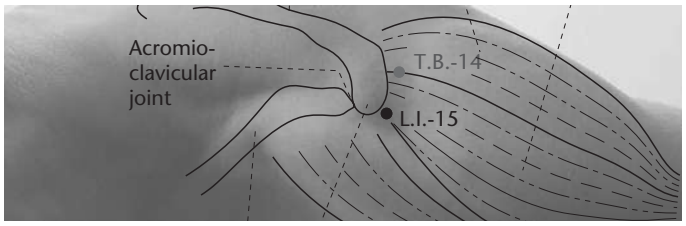

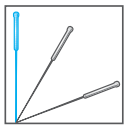
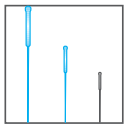

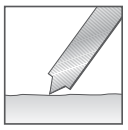

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 599

Arm Centre BIZHONG

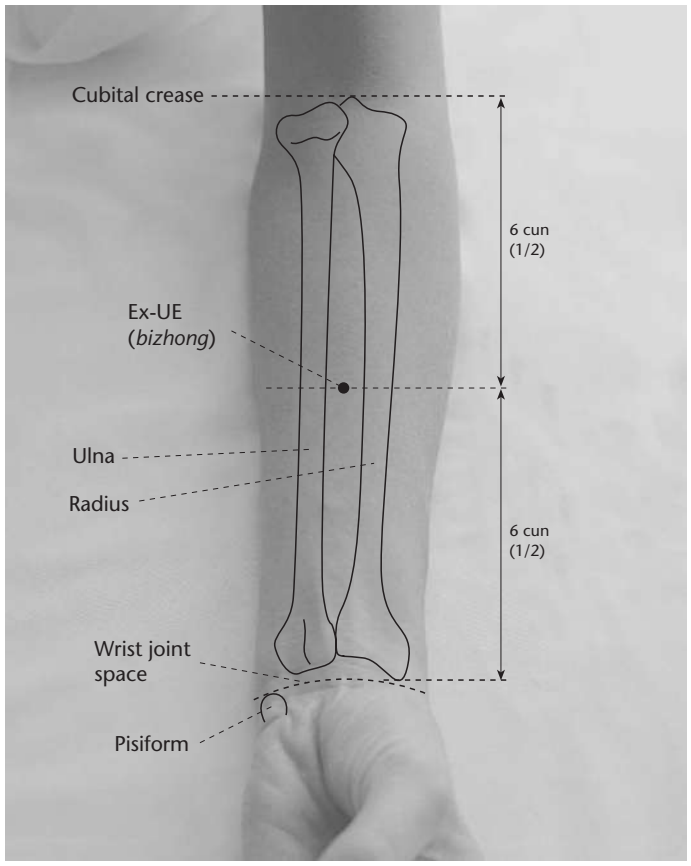

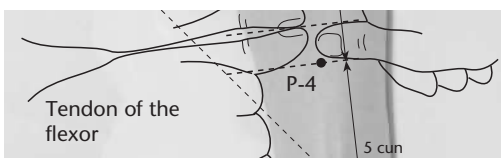

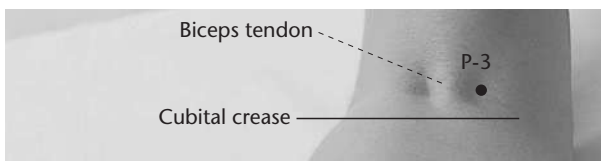

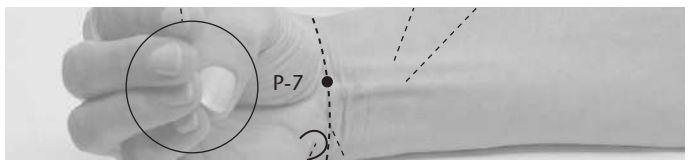

Location

On the palmar aspect of the forearm, midway between the cubital crease and the wrist joint space ('most distal wrist crease').

How to find

Place the patient's forearm in a relaxed supinated position. As the location of the wrist crease may vary, the wrist joint space presents a more reliable reference point (→ 3.3.3). The spreading hands technique is the best method for locating this point (→ 2.3.3): Place the little fingers on the centre of the wrist joint (→ **P-7**) and the cubital crease (→ **P-3**) respectively and determine the midpoint of this distance by joining the thumbs. Locate **Ex-UE** (*bizhong*) on this level between the radius and the ulna.
→ **P-4** is located 1 cun distal to the midpoint of the distance between → **P-3** and → **P-7** or 5 cun proximal to → **P-7**.

Needling

Vertically 1–1.5 cun

Actions/Indications

- Opens the channel and the *luo* vessels locally

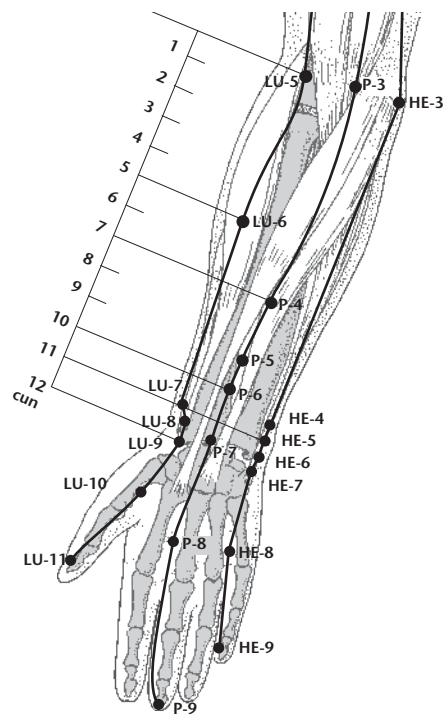

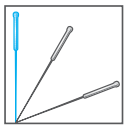
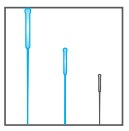

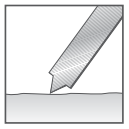

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 601

6.5 Extra Points: Lower Extremities (EX-LE)

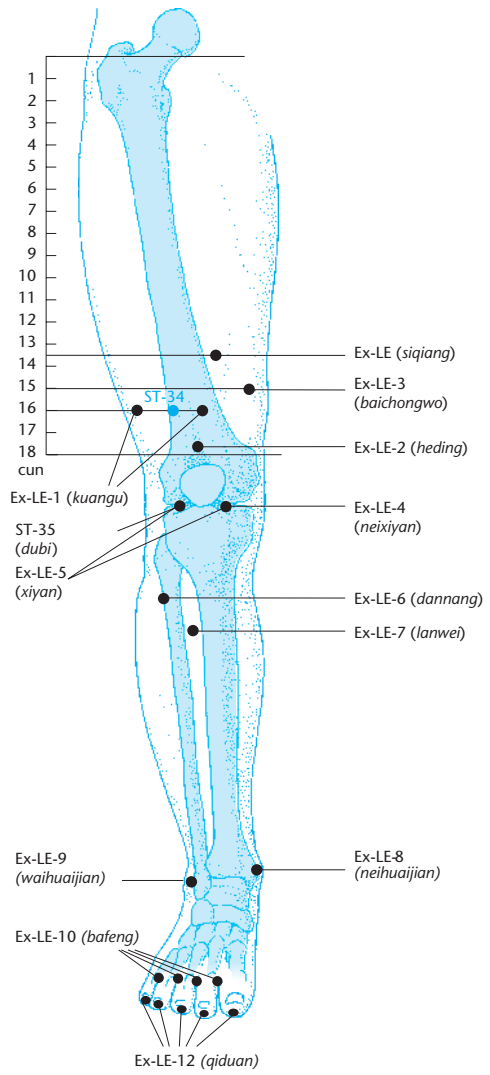

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 602

English abbreviation (standard)	*Pinyin* name	Nguyen Van Nghi, König and Wancura, Schnorrenberger	Shanghai College	Ex (Hempen)
Ex-LE-1	*kuangu*	PaM or ZP165 (*changgu*)	M-LE-28	
Ex-LE-2	*heding/xiding*	PaM or ZP156	M-LE-27	
Ex-LE-3	*baichongwo*	PaM or ZP163	M-LE-34	Ex21
Ex-LE-4	*neixiyan*	(PaM or ZP145)	M-LE	(Ex23)
Ex-LE-5	*xiyan*	PaM or ZP145	M-LE	Ex23
Ex-LE-6	*dannangxue*	PaM or ZP152	M-LE-23	
Ex-LE-7	*lanweixue*	PaM or ZP142	M-LE-13	Ex22
Ex-LE-8	*neihuaijian*	PaM or ZP146	M-LE-17	
Ex-LE-9	*waihuaijian*	PaM or ZP151	M-LE-22	
Ex-LE-10	*bafeng*	PaM or ZP137	M-LE-8	Ex19
Ex-LE-11	*duyin*		M-LE	
Ex-LE-12	*qiduan*		M-LE-6	
Further extra points				
Ex-LE	*huanzhong*	PaM or ZP84	M-**BW-34**	Ex20
Ex-LE	*siqiang*	NP94	**N**-LE-19	
Ex-LE	*lineiting*	PaM or ZP130	M-LE-1	

Ch06b-F10028.qxd 2/22/08 10:04 PM Page 603

Hip Bone KUANGU Ex-LE-1

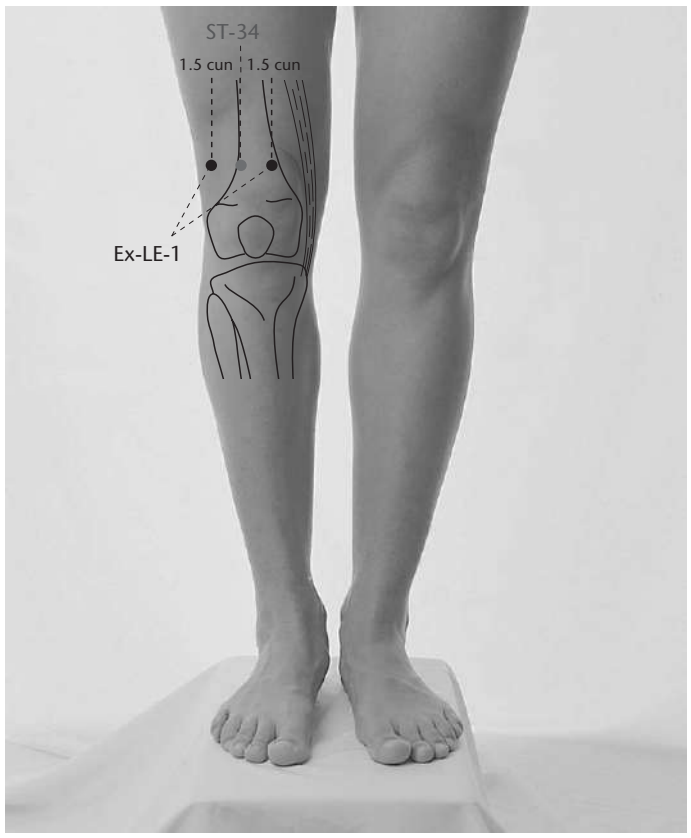

Location

A pair of points 2 cun superior to the patella and 1.5 cun lateral and medial to ➞ **ST-34** (*liangqiu*).

How to find

First, locate ➞ **ST-34**: measure 2 cun superiorly from the lateral upper border of the patella and there locate ➞ **ST-34** in a depression on the vastus lateralis muscle. Then locate **Ex-LE-1** 1.5 cun lateral and medial to ➞ **ST-34**.

Needling

Vertically 1–1.5 cun

Actions/Indications

- Alleviate pain

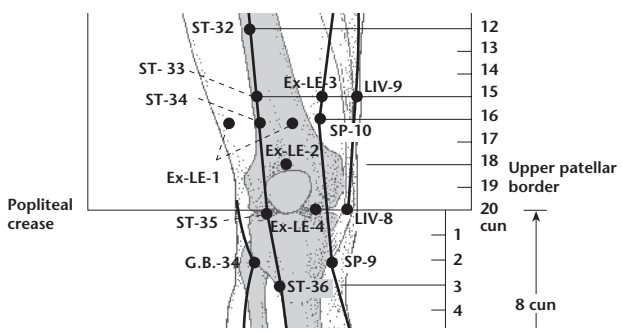

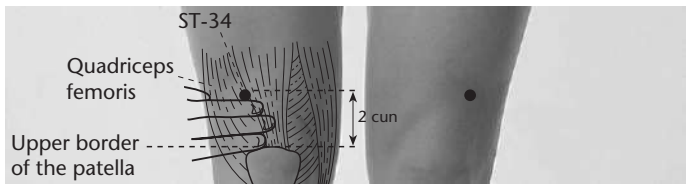

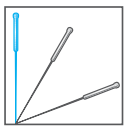
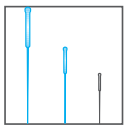

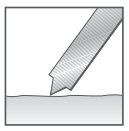

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 604

Ex-LE-2 *Crane's Summit* HEDING/XIDING

Location
In the centre of the upper border of the patella.

How to find
This point is best located and needled with the patient's knee slightly flexed (knee support). Locate the midpoint of the upper patellar border and there locate **Ex-LE-2** (*heding*).

Needling
Vertically 0.5–0.8 cun

Actions/Indications
- Benefits the knee joint

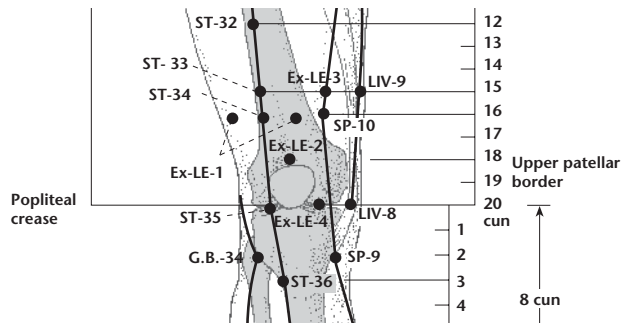

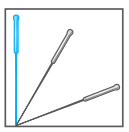
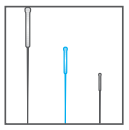

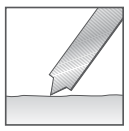

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 605

Hundred Insect Burrow BAICHONGWO Ex-LE-3

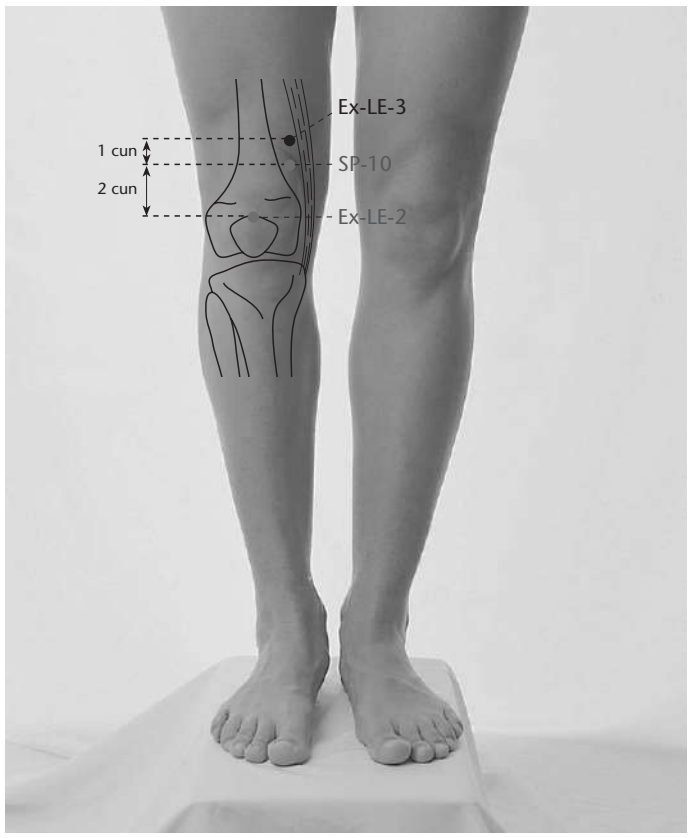

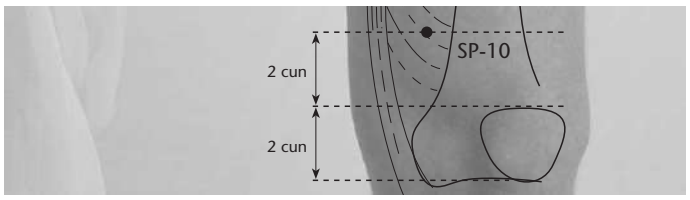

Location

3 cun superior and 1 cun medial to the upper medial border of the patella, in a small depression on the vastus medialis muscle or 1 cun superior to → **SP-10**.

How to find

The patient should be in a supine or preferably in a seated position, with their knee flexed. Locate the upper medial border of the patella and, from there, measure 3 cun (1 handsbreadth) towards proximal. There, locate **Ex-LE-3** in a small depression on the vastus lateralis muscle.

Needling

Vertically 1–2 cun

Actions/Indications

- Cools Blood-Heat, eliminates Wind, drains Dampness

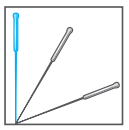
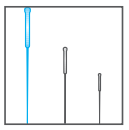

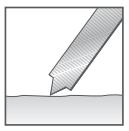

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 606

Ex-LE-4 *Inner Eye of the Knee* NEIXIYAN

Location
With the knee flexed, inferior to the patella, in a depression medial to the patellar ligament.

How to find
This point is best located with the patient's knee flexed (use knee support). The medial eye of the knee can be located in a distinct depression on the level of the lower border of the patella, medial to the patellar ligament. The lateral eye of the knee corresponds to → **ST-35** (*dubi*). Together, these two points form the extra point → **Ex-LE-5** (*xiyan*).

Needling
0.5–1 cun vertically or obliquely towards the lateral eye of the knee. Caution: Knee joint.

Actions/Indications
- Disorders of the knee joint

Special features
Important local point

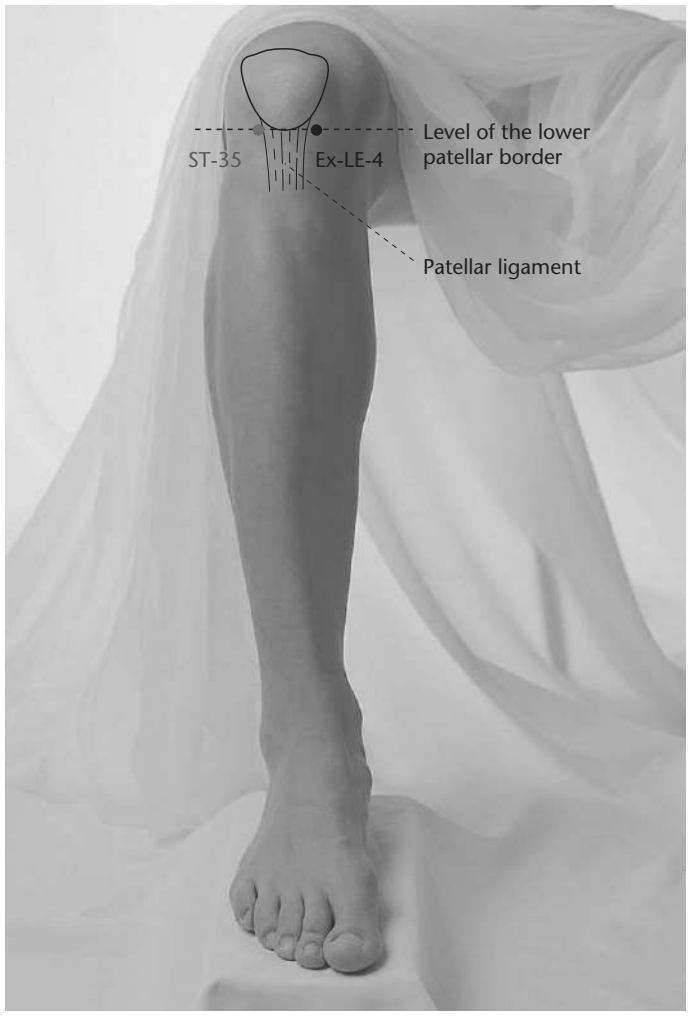

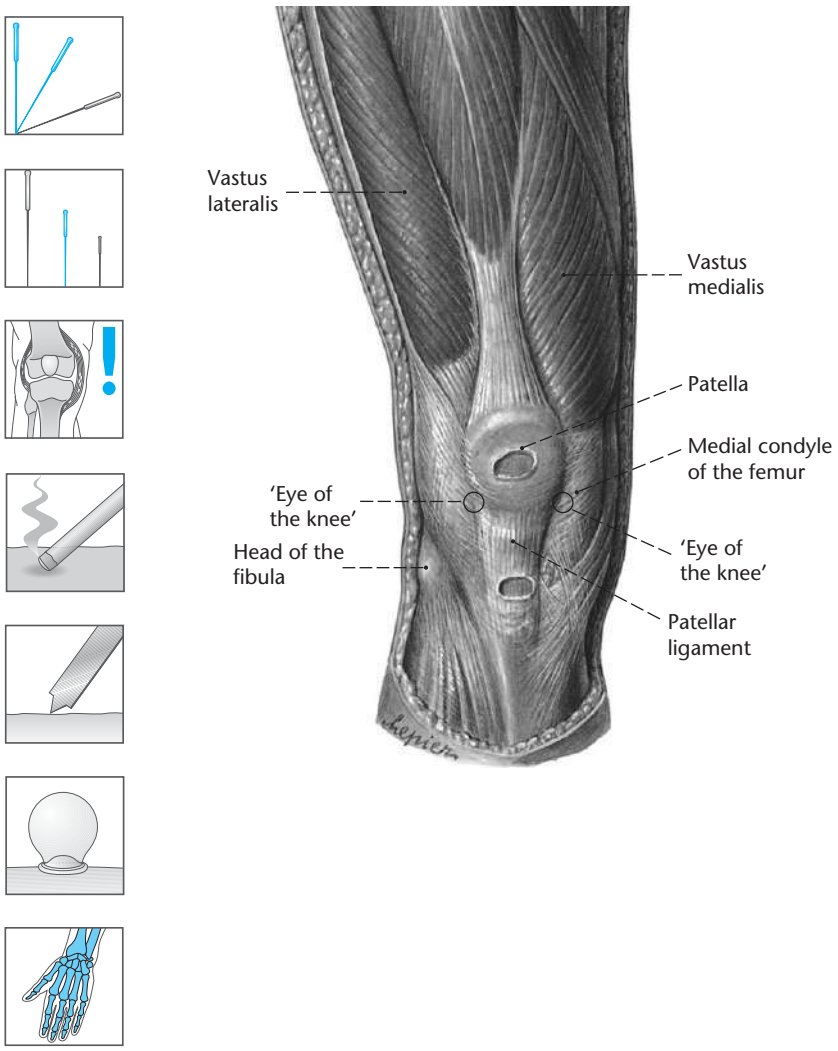

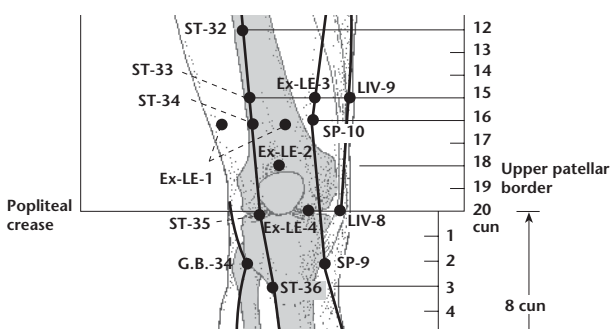

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 607

Eyes of the Knee XIYAN Ex-LE-5

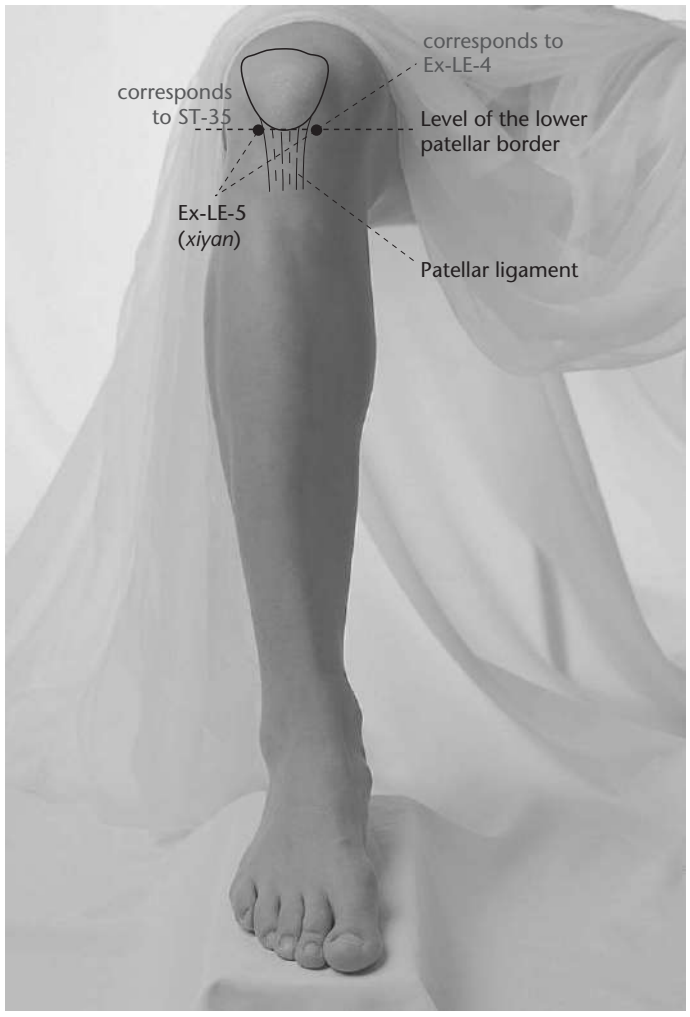

Location

With the knee flexed, this pair of points is located inferior to the patella, medial and lateral to the patellar ligament. **Ex-LE-5** (*xiyan*) includes two points: the medial eye of the knee corresponds to → **Ex-LE-4** (*neixiyan*), the lateral eye of the knee to **ST-35** (*dubi*).

How to find

This point is best located with the patient's knee flexed (use knee support). The eyes of the knee can be located on the level of the lower border of the patella, lateral and medial to the patellar ligament. The lateral eye of the knee corresponds to → **ST-35** (*dubi*), the medial eye of the knee to → **Ex-LE-4** (*neixiyan*). Together, these two points form the extra point **Ex-LE-5** (*xiyan*).

Needling

Medial eye of the knee: 0.5–1 cun vertically or obliquely towards the lateral eye of the knee; lateral eye of the knee: 0.5–1 cun vertically or obliquely towards the medial eye of the knee. Caution: Do not needle too deeply to avoid intra-articular puncture (these points correspond to the incision sites for knee arthroscopy).

Actions/Indications

- Eliminate Wind-Dampness, reduce swellings and alleviate pain

Special features

Important local points for disorders of the knee

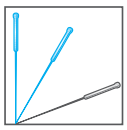
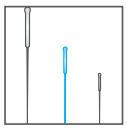

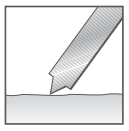

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 608

Ex-LE-6 *Gall Bladder Point* DANNANG/DANNANGXUE/DANNANGDIAN

Location

On the G.B. channel of the **right** leg, approximately 1–2 cun distal to ➞ **G.B.-34**, palpate for the most tender point.

How to find

First, locate ➞ **G.B.-34** on the right leg: locate the head of the fibula on the lateral aspect of the leg (where the seam of the trousers would be) and hold it with the index and middle fingers (as with tweezers). By gliding distally with both fingers, the index finger will come to rest in a depression directly inferior and anterior to the head of the fibula (➞ **G.B.-34**). From there, continue to palpate slowly along the G.B. channel, until you find the most tender point approximately 1–2 cun inferior to ➞ **G.B.-34**. This is the location of **Ex-LE-6** (*dannangxue*).

Needling

Vertically 1–1.5 cun

Actions/Indications

- Clears Heat and drains Dampness for acute and chronic disorders of the Gall Blader such as cholecystitis and cholelithiasis (tenderness on this point is diagnostically significant).

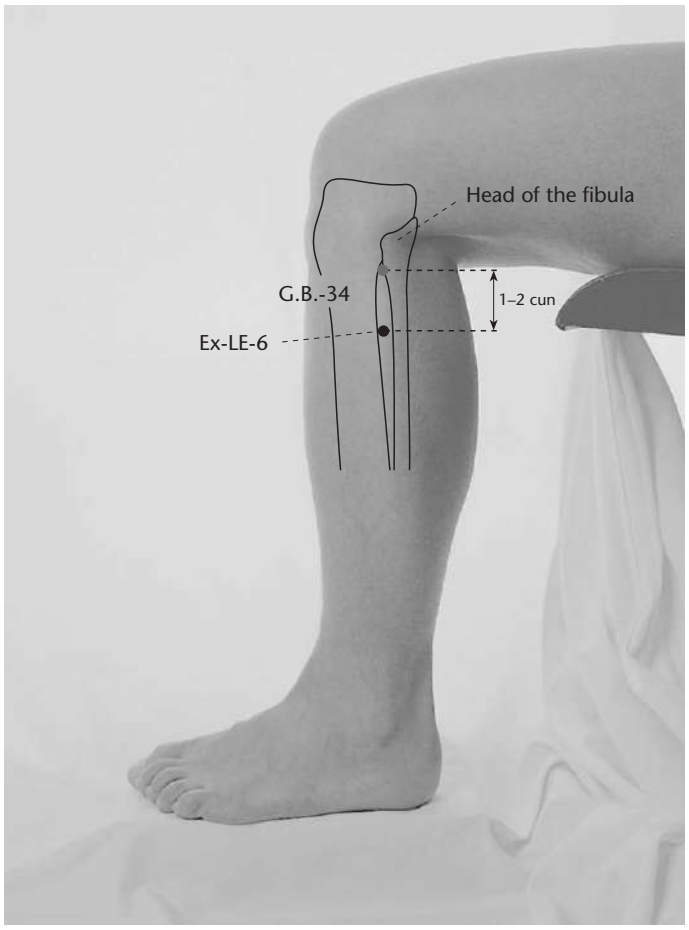

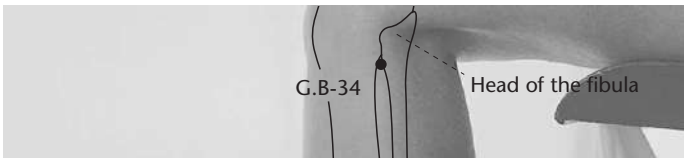

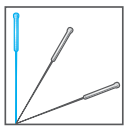

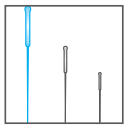

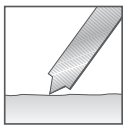

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 609

Appendix Point LANWEI/LANWEIXUE Ex-LE-7

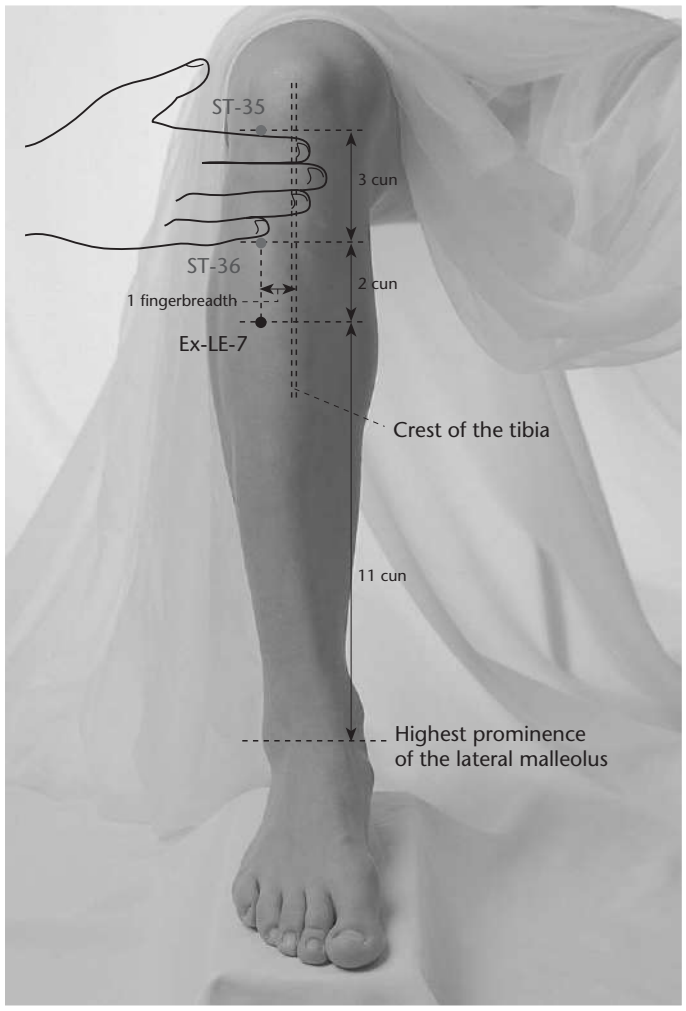

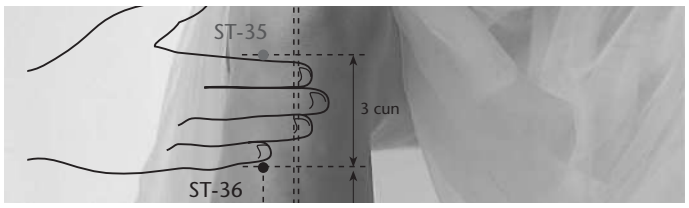

Location

On the ST channel of the **right** leg, the most tender point approximately 2 cun distal to → **ST-36**.

How to find

Locate this point on the right leg only (this corresponds to the unilateral location of the appendix on the abdomen). First, locate → **ST-36** 3 cun inferior to the knee joint space (→ **ST-35**) and 1 cun lateral to the anterior crest of the tibia. Then continue to palpate along the ST channel for approximately 2 cun (towards → **ST-37**, 3 cun inferior to → **ST-36**) and locate **Ex-LE-7** (*lanweixue*) on the most tender point.

Needling

Vertically 1–1.5 cun

Actions/Indications

- Activates the Qi and Blood flow in the Large Intestine, clears Heat and Heat toxins from the Large Intestine

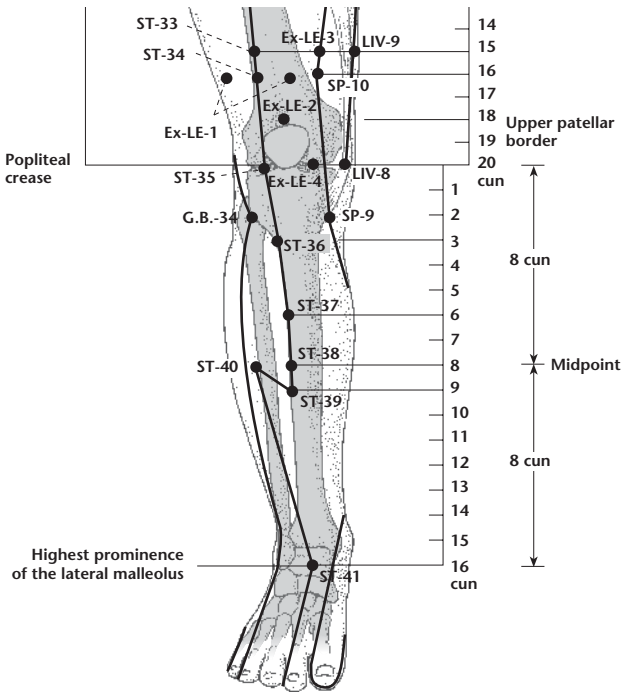

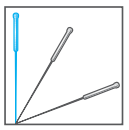

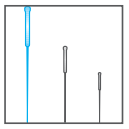

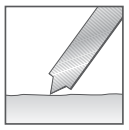

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 610

Ex-LE-8 *Medial Malleolus Tip* NEIHUAIJIAN

Location

On the highest point of the medial malleolus.

How to find

Find the highest prominence of the medial malleolus and there locate **Ex-LE-8** (*neihuaijian*). Located in a comparable position but on the lateral malleolus is → **Ex-LE-9** (*waihuaijian*).

Needling

0.1 cun transversely (subcutaneously) or prick to bleed.

Actions/Indications

- Pain in the medial ankle region

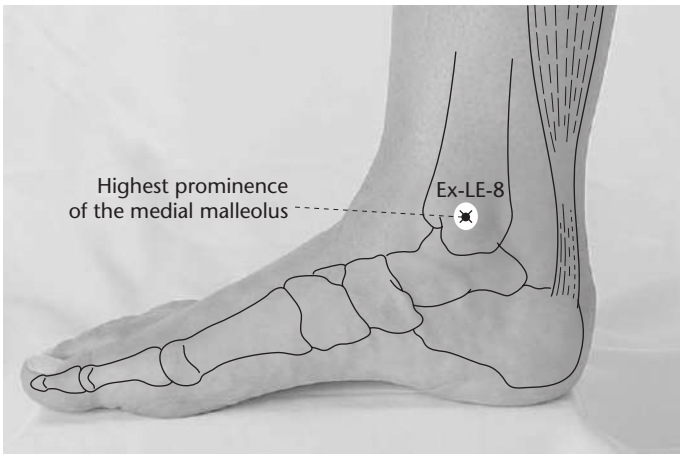

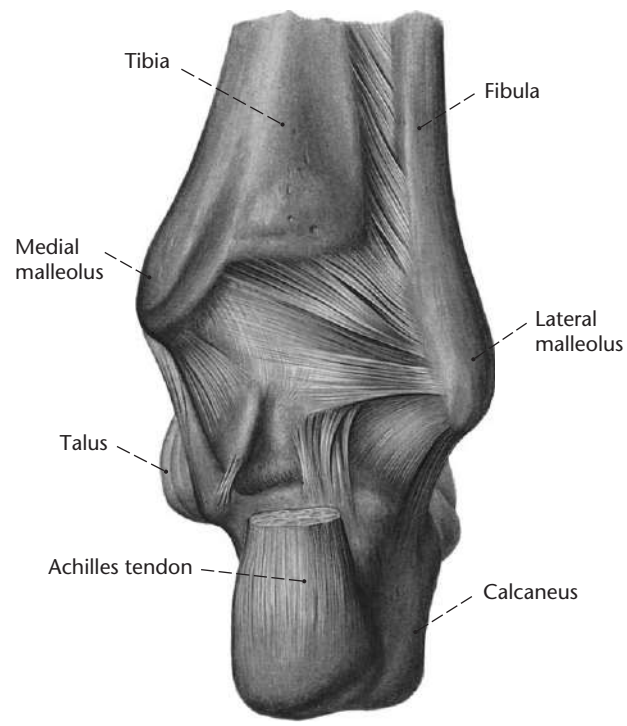

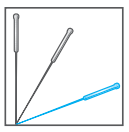
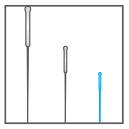

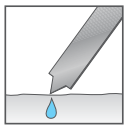

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 611

Outer Malleolus Tip WAIHUAIJIAN Ex-LE-9

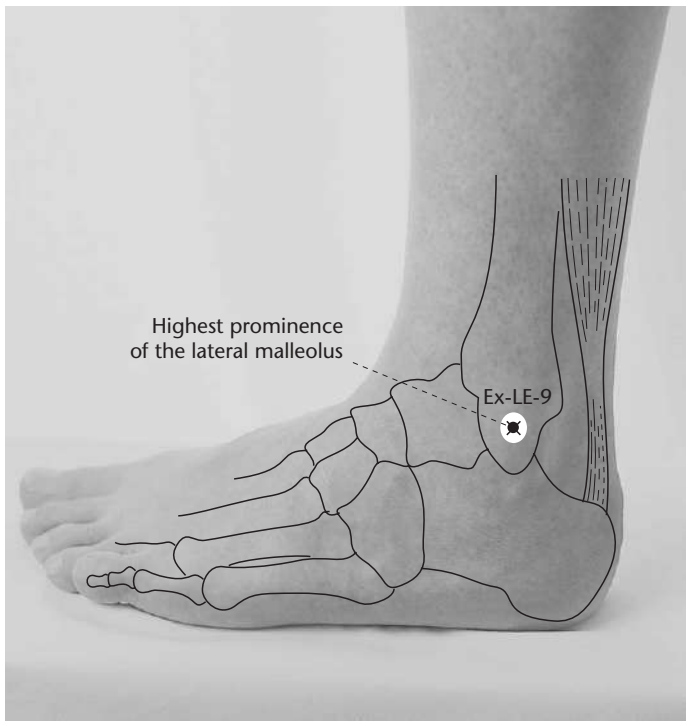

Location

On the highest prominence of the lateral malleolus

How to find

Find the highest prominence of the lateral malleolus and there locate **Ex-LE-9** (*waihuaijian*). Located in a comparable position but on the medial malleolus is → **Ex-LE-8** (*neihuaijian*).

Needling

0.1 cun transversely (subcutaneously) or prick to bleed.

Actions/Indications

- Pain in the lateral ankle region

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 612

Ex-LE-10 *Eight Winds* BAFENG

Location

Eight points located on the dorsum of the foot, slightly proximal to the margins of the interdigital webs.

How to find

Locate the points of **Ex-LE-10** (*bafeng*) on the dorsum of the foot, proximal to the margins of the webs between the toes, on the border of the red and white skin. → **LIV-2** (*xingjian*), → **ST-44** (*neiting*) and → **G.B.-43** (*xiabai*) are all part of the *bafeng* points. The extra points → **Ex-UE-9** (*baxie*) are located in a comparable position on the dorsum of the hand, proximal to the margins of the webs between the fingers (→ **T.B.-2** is part of the → **Ex-UE-9**).

Needling

0.3–1 cun obliquely in a proximal direction or prick to bleed.

Actions/Indications

- Clears Heat and reduces swelling

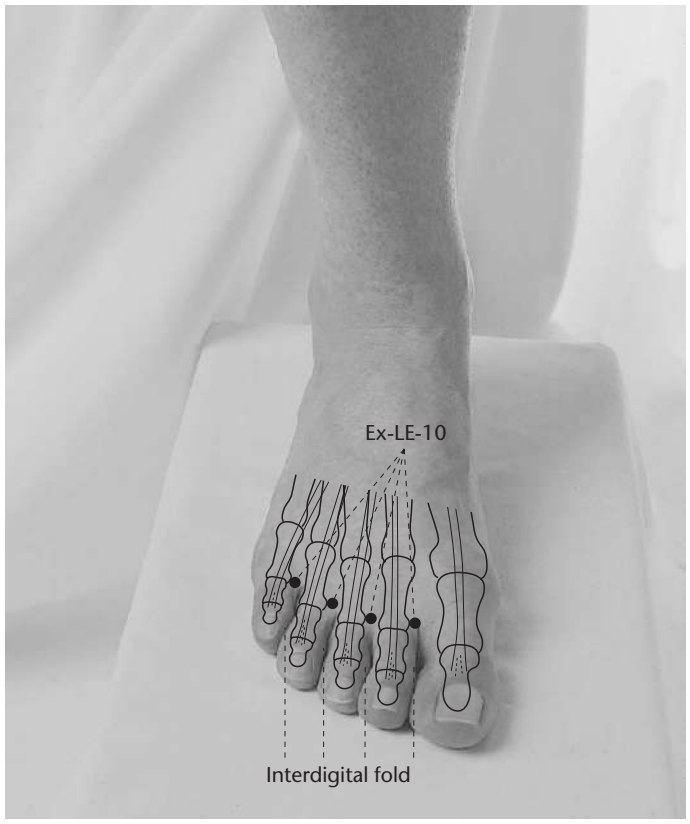

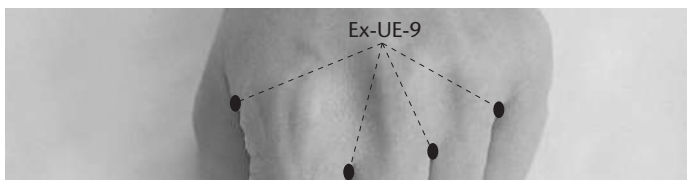

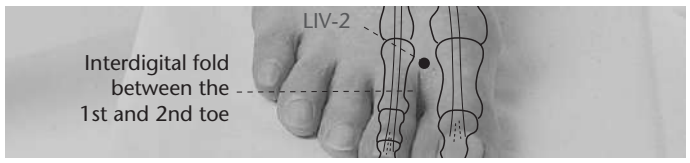

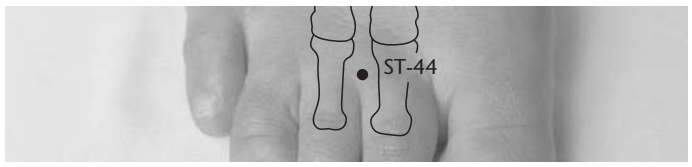

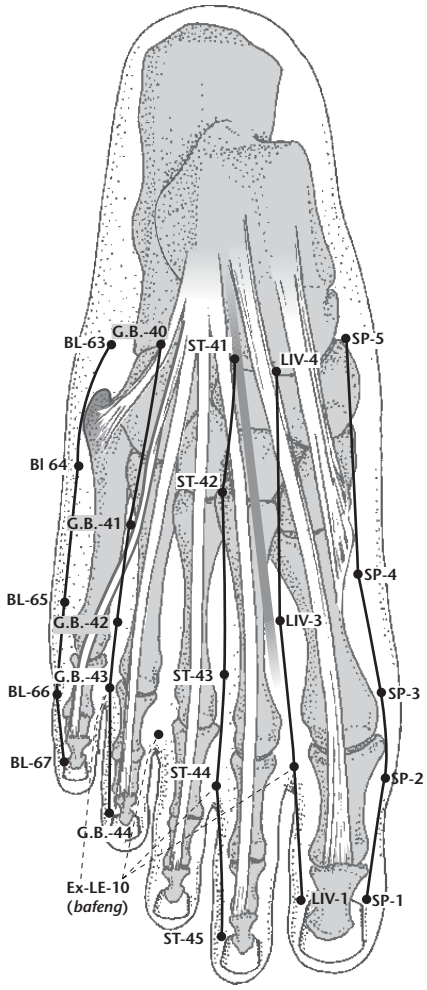

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 613

Solitary Yin DUYIN Ex-LE-11

Location

On the plantar aspect of the 2nd toe, at the midpoint of the transverse crease of the distal interphalangeal joint.

How to find

Locate the plantar aspect of the 2nd toe. There, locate **Ex-LE-11** (*duyin*) at the midpoint of the transverse crease of the distal interphalangeal joint.

Needling

0.2–0.3 cun vertically or transversely (subcutaneously) or prick to bleed or moxibustion.

Actions/Indications

- Acute angina, thoracic and hypochondriac pain, nausea, vomiting, retention of the lochia, irregular menstruation, inguinal hernia

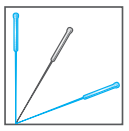
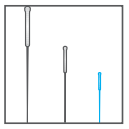

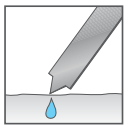
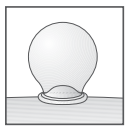

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 614

Ex-LE-12 *End of Qi* QIDUAN

Location
On the tips of the 10 toes.

How to find
Locate the points of **Ex-LE-12** (*qiduan*) on the tips of the 10 toes, 0.1 cun from the free margin of the nail.
The points of → **Ex-UE-11** (*shixuan*) are located in a comparable position on the hand, on the tips of the 10 fingers.

Needling
Prick to bleed

Actions/Indications
- Syncope, oedema of the feet, acute abdominal pain

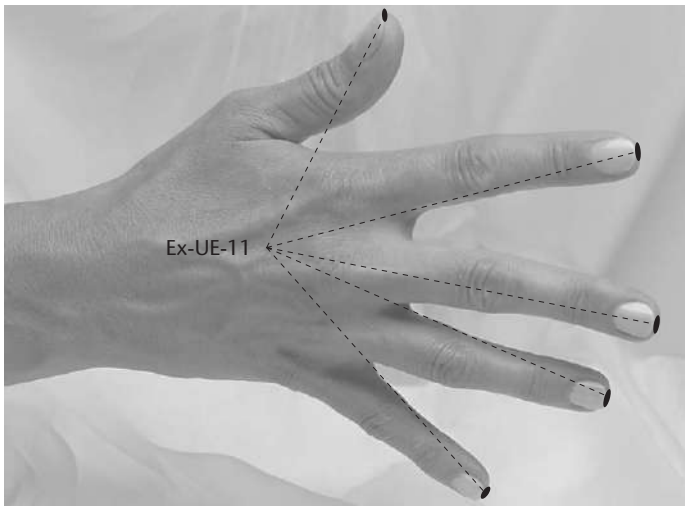

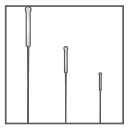

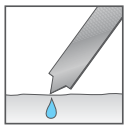

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 615

Circle Centre HUANZHONG

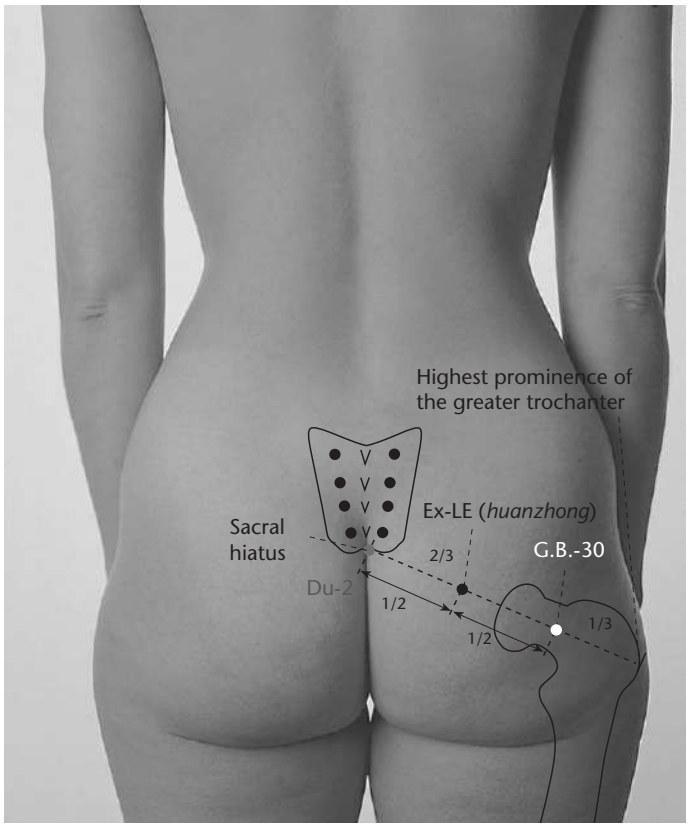

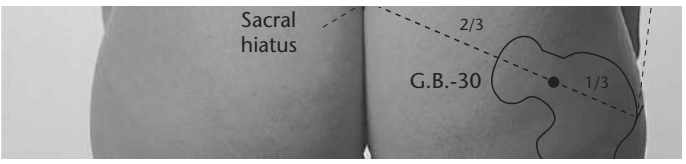

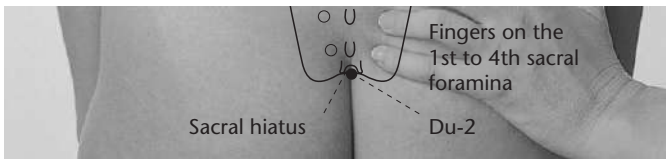

Location

Midway between ➞ **G.B.-30** (*huantiao*) and ➞ **Du-2** (*yaoshu*).

How to find

First, locate ➞ **G.B.-30**: divide the distance between the prominence of the greater trochanter and the sacral hiatus into thirds (➞ 3.4.4) and locate ➞ **G.B.-30** a third of the distance from the greater trochanter. Then locate ➞ **Du-2** directly below the sacral hiatus. Next, locate **Ex-LE** (*huanzhong*) midway between these two points.

Needling

Vertically 2–2.5 cun

Actions/Indications

- Sciatica, urinary tract infection, haemorrhoids, paralysis of the lower extremities

Special features

- This point is not part of the WHO classification of extra points.

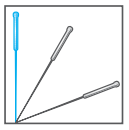
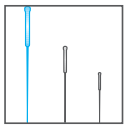

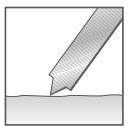
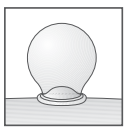

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 616

Ex-LE *Four Muscles Strengthening Point* SIQIANG

Location

4.5 cun superior to the centre of the upper patellar border.

How to find

Find the centre of the upper patellar border. From there, measure 4.5 cun (1 handbreadth and 2 fingerbreadths) in a superior direction. **Or:** The distance between the upper border of the pubic symphysis and the upper border of the patella is 18 cun (→ 2.2). Divide this distance into quarters (spreading hands technique, elastic tape → Chapter 2) and **Ex-LE** (*siqiang*) is located on the centre of the thigh, one quarter of the distance from the upper patellar border.

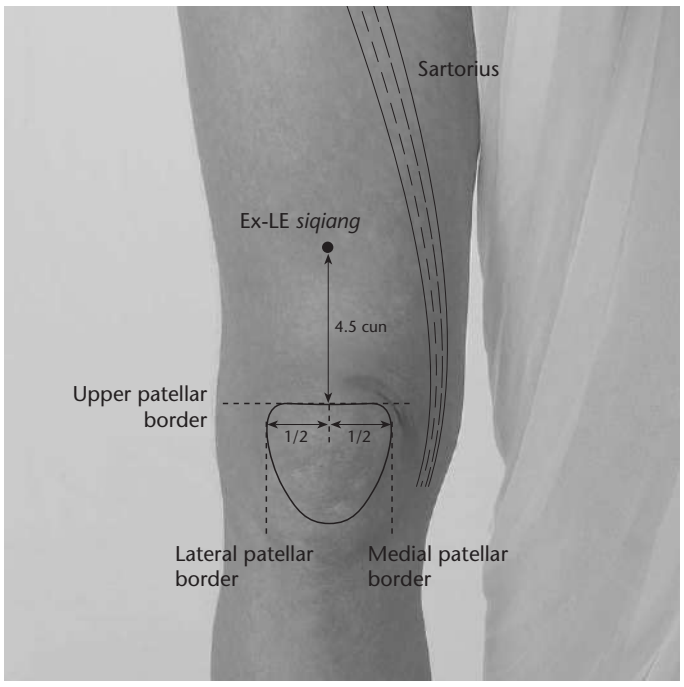

Needling

Vertically 1–2 cun

Actions/Indications

- Paralysis and atrophy of the muscles of the lower extremities, especially of the quadriceps femoris muscle

Special features

'Four muscles' refers to the four portions of the quadriceps femoris muscle. This point is not part of the WHO classification of extra points.

Ch06b-F10028.qxd 2/22/08 10:05 PM Page 617

Plantar Inner Space LINEITING

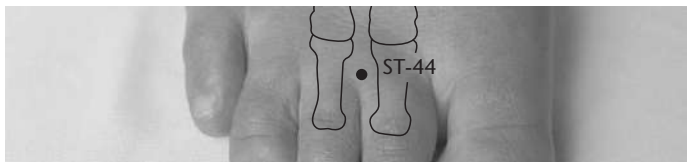

Location

On the plantar aspect of the foot, between the 2nd and 3rd metatarsal bones, opposite → **ST-44** (*neiting*).

How to find

Locate the space between the 2nd and 3rd toes on the plantar aspect of the foot. The extra point *lineiting* lies opposite → **ST-44** (on the dorsal aspect of the foot, proximal to the margin of the web between the 2nd and 3rd toe).

Needling

0.2–0.3 cun vertically towards → **ST-44**. Caution: Painful point.

Actions/Indications

- Acute epigastric pain, local pain, epilepsy, restlessness

Special features

This point is not part of the WHO classification of extra points.

7 Important Points According to Region

Claudia Focks

7.1 Anterior Aspect of the Head

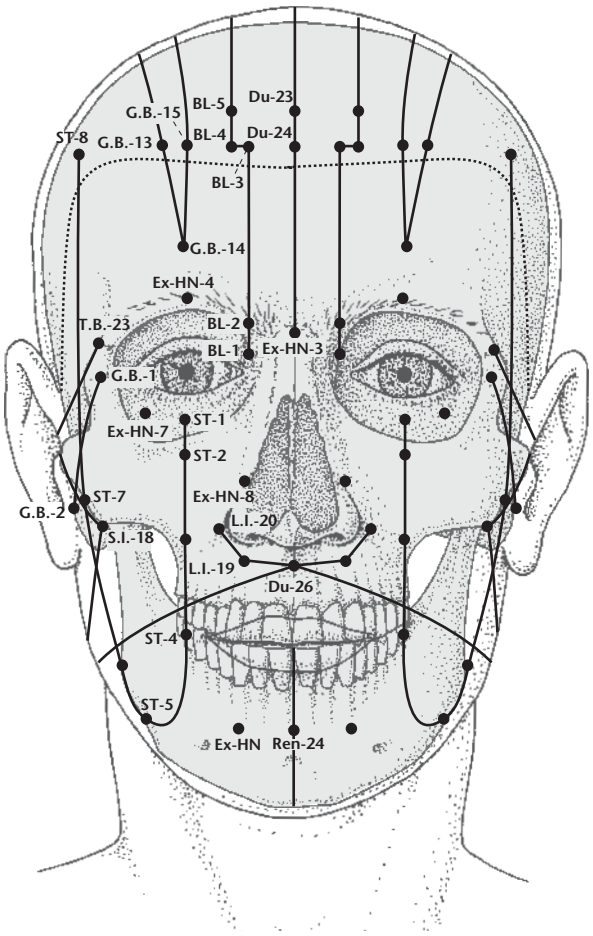

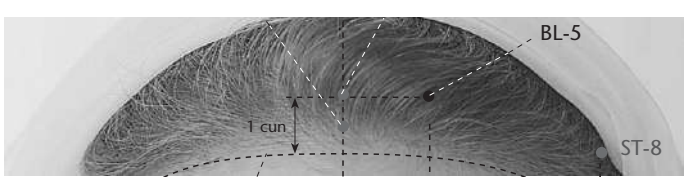

BL-5 (*wuchu*) 'Fifth Place' 1 cun superior to the anterior hairline and 1.5 cun lateral to the midline (directly above **BL-4**, level with **Du-23**)

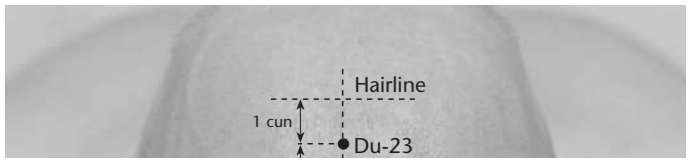

Du-23 (*shangxing*) 'Upper Star' On the midline, 1 cun superior to the anterior hairline

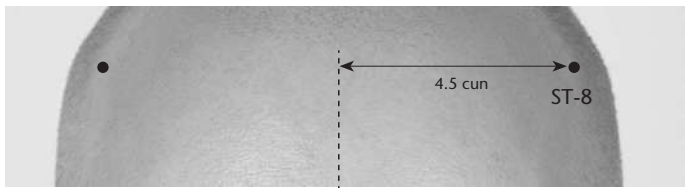

ST-8 (*touwei*) 'Head's Binding' At the temporal corner of the forehead, on the border of the temporalis muscle and 0.5 cun within the anterior hair line or 4.5 cun lateral to the anterior midline (**Du-24**)

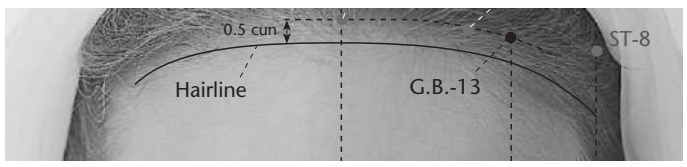

G.B.-13 (*benshen*) 'Root of the Spirit' 0.5 cun superior to the anterior hairline and 3 cun lateral to the midline, on the line between **Du-24** and **ST-8** (= 4.5 cun) at the junction of the lateral and medial third

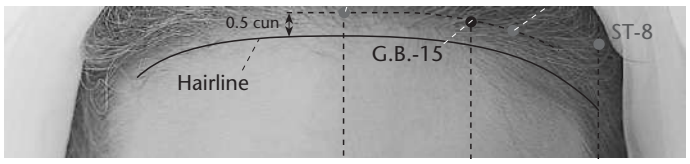

G.B.-15 (*toulinqi*) 'Head Governor of Tears' On the pupil line when looking straight ahead, 0.5 cun superior to the anterior hairline, at the midpoint of the line between **Du-24** and **ST-8** (= 4.5 cun)

Ch07-F10028.qxd 2/22/08 10:08 PM Page 620

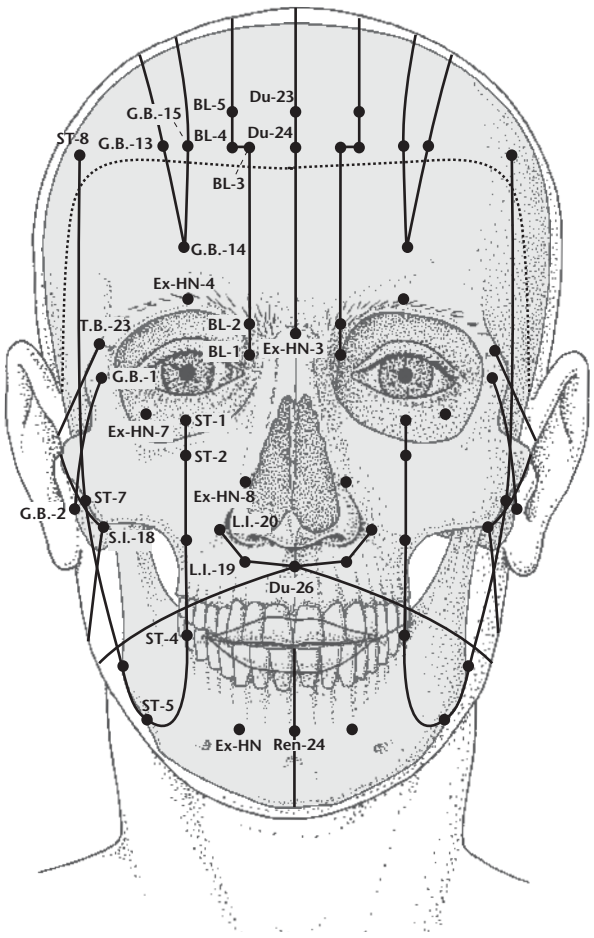

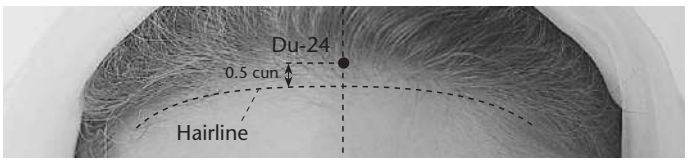

Du-24 (*shenting*) 'Courtyard of the Spirit' On the midline, 0.5 cun superior to the anterior hairline

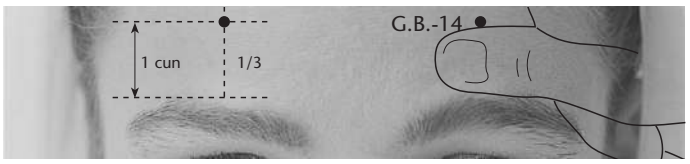

G.B.-14 (*yangbai*) 'Yang White' When looking straight ahead on the pupil line, approximately 1 cun superior to the midpoint of the eyebrow (centre of the eyebrow to anterior hairline = 3 cun)

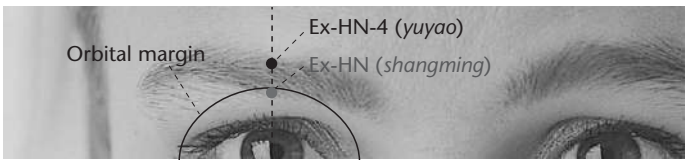

Ex-HN-4 (*yuyao*) 'Fish Waist' In the centre of the eyebrow, with the eyes looking straight ahead directly above the pupil

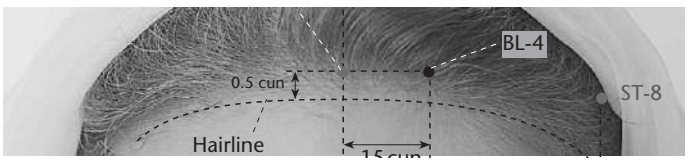

BL-4 (*qucha*) 'Crooked Curve' 0.5 cun superior to the anterior hairline and 1.5 cun lateral to the midline, or one third of the distance between **Du-24** and **ST-8** (= 4.5 cun)

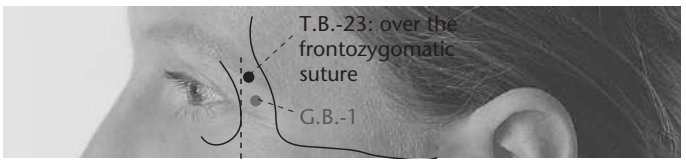

T.B.-23 (*sizhukong*) 'Silken Bamboo Hollow' In the depression (frontozygomatic suture) at the lateral end of the eyebrow

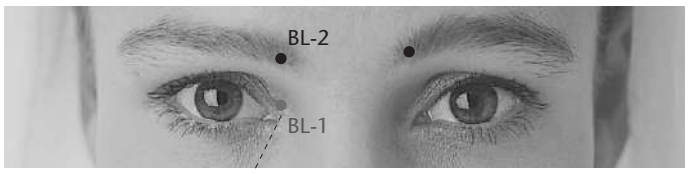

BL-2 (*zanzhu*) 'Gathered Bamboo' In a depression at the medial end of the eyebrow, directly superior to the inner canthus of the eye

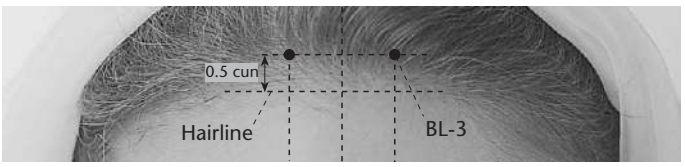

BL-3 (*meichong*) 'Eyebrows' Pouring' 0.5 cun within the anterior hairline, superior to the medial canthus of the eye (**BL-1**)

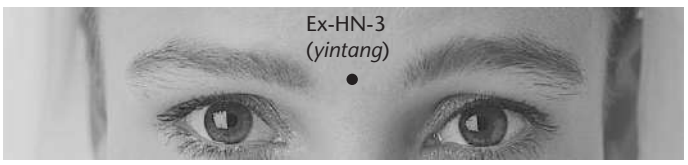

Ex-HN-3 (*yintang*) 'Hall of Impression' On the anterior midline between the eyebrows

Ch07-F10028.qxd 2/22/08 10:08 PM Page 621

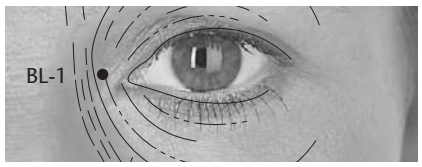

BL-1 (*jingming*) 'Bright Eyes' 0.1 cun superior to the inner canthus of the eye, where spectacles tend to rest

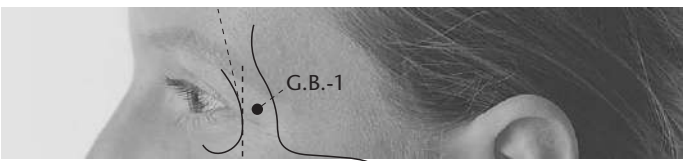

G.B.-1 (*tongziliao*) 'Pupil Crevice' 0.5 cun lateral to the outer canthus of the eye

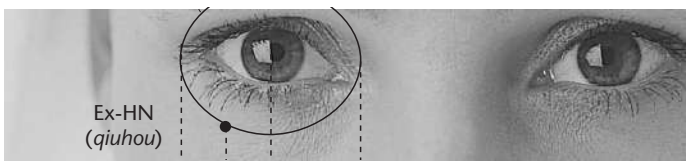

Ex-HN-7 (*qiuhou*) 'Behind the Ball' On the lower border of the orbit, at the junction of the lateral quarter and the medial three-quarters of the orbital margin

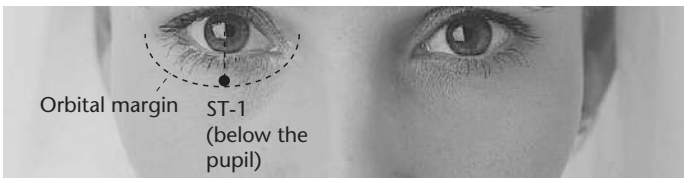

ST-1 (*chengqi*) 'Container of Tears' With the eyes looking straight forward, directly below the centre of the pupil, between the eyeball and the infraorbital ridge

ST-2 (*sibai*) 'Four Whites' When looking straight ahead, directly below the centre of the pupil, in the depression at the infraorbital foramen

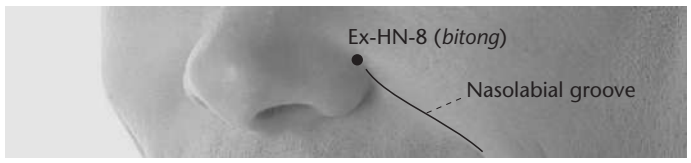

Ex-HN-8 (*shangyingxiang or bitong*) 'Upper Yingxiang' or 'Free Nose' At the upper end of the nasolabial groove, at the junction of the maxilla and the nasal cavity (at the border between the nasal bone and cartilage)

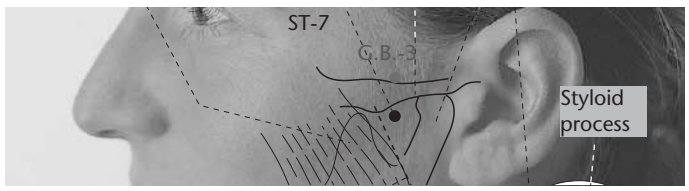

ST-7 (*xiaguan*) 'Below the Joint' With the mouth closed in the depression of the mandibular incisure, between the coronoid and condyloid process of the mandible

G.B.-2 (*tinghui*) 'Meeting of Hearing' Anterior to the ear, with the mouth open in the depression at the level of the intertragic notch, on the lower border of the condyloid process of the mandible.

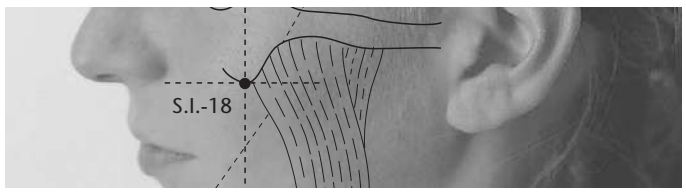

S.I.-18 (*quanliao*) 'Cheek Bone Crevice' In a depression at the intersection of a vertical line through the outer canthus of the eye with the lower border of the zygomatic bone

ST-3 (*juliao*) 'Great Crevice' When looking straight ahead, directly below the centre of the pupil, level with of lower border of the ala nasi.

Ch07-F10028.qxd 2/22/08 10:08 PM Page 622

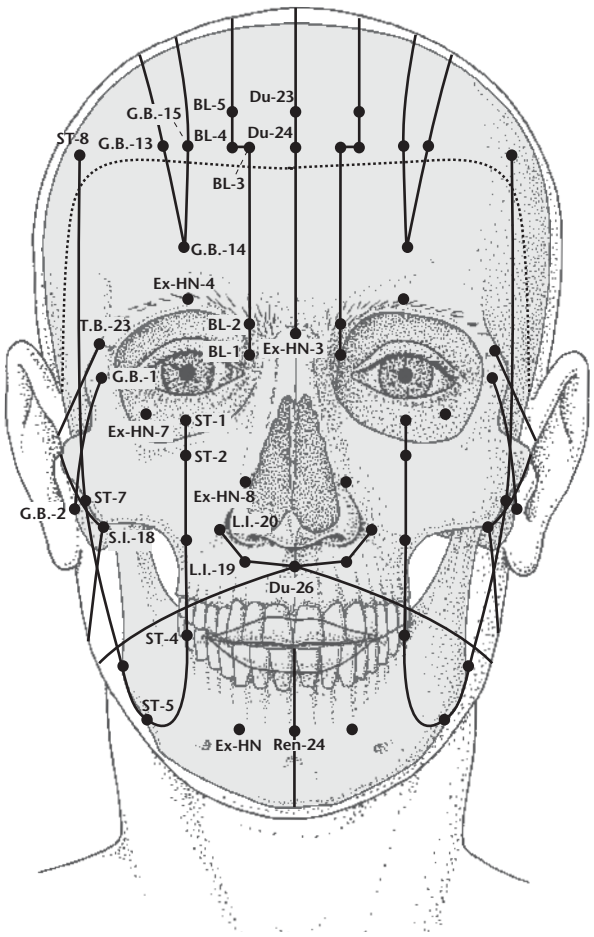

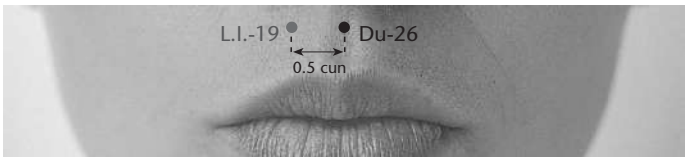

Du-26 (*renzhong* or *shuigou*) 'Man's Middle' Below the nose, on the midline, at the junction of the upper and middle third of the philtrum

ST-4 (*dicang*) 'Earth Granary' When looking straight ahead on the pupil line, approximately 0.4 cun lateral to the corner of the mouth

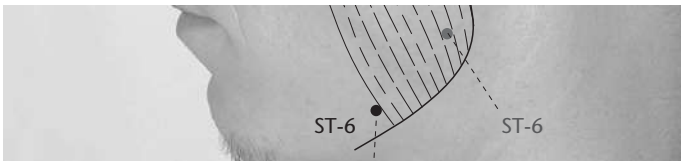

ST-5 (*daying*) 'Great Welcome' On the lateral mandible, anterior to the masseter (ask the patient to clench the teeth); a branch of the facial artery is palpable here

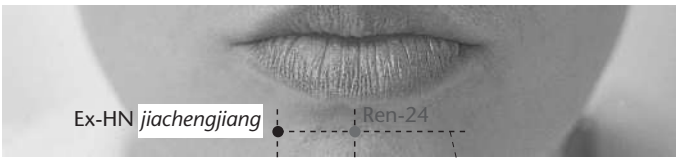

Ex-HN (*jiachengjiang*) 'Adjacent to Container of Fluids' At the same level and 1 cun lateral to **Ren-24** (chengjiang)

Ren-24 (*chengjiang*) 'Container of Fluids' On the midline, in the mentolabial groove below the lower lip

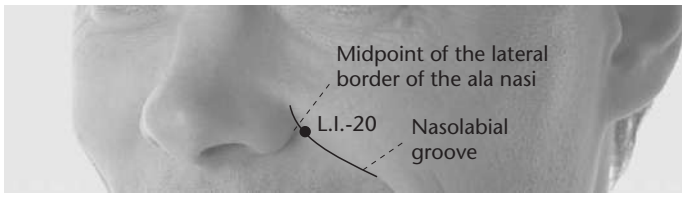

L.I.-20 (*yingxiang*) 'Welcome Fragrance' In the nasolabial groove, at the level of the midpoint of the lateral border of the ala nasi.

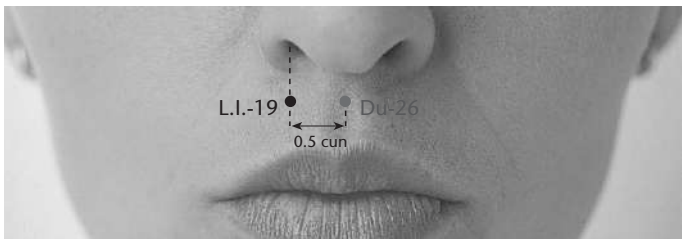

L.I.-19 (*kouheliao*) 'Mouth Grain Crevice' Directly inferior to the lateral margin of the nostril, level with **Du-26**, which is located slightly superior to the midpoint of the philtrum (meeting point of the L.I. channel)

Ch07-F10028.qxd 2/22/08 10:08 PM Page 623

7.2 Head and Lateral Aspect of the Neck

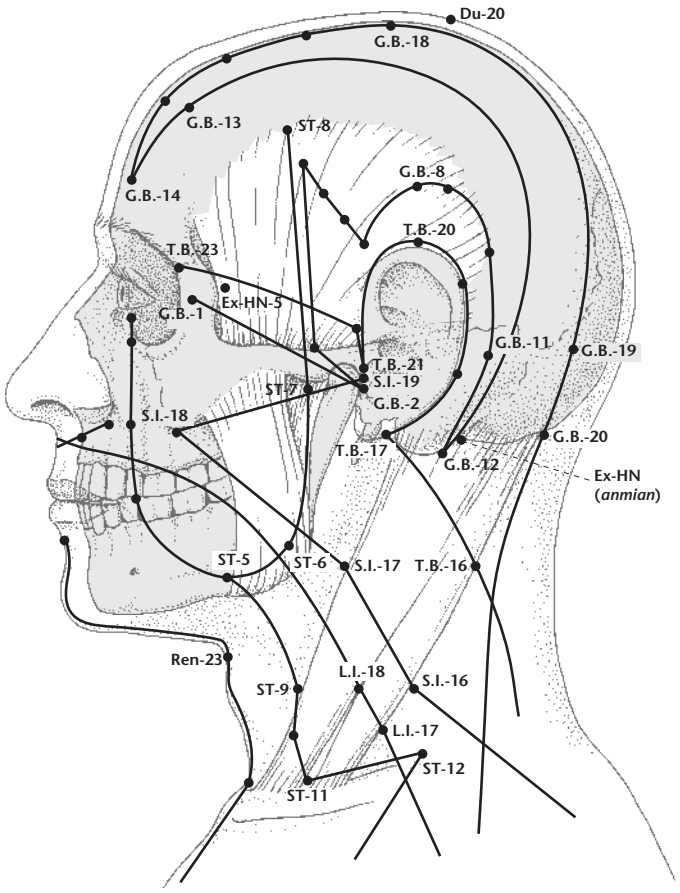

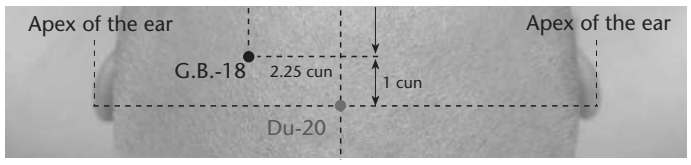

G.B.-18 (*chengling*) 'Support Spirit' 4 cun superior to the anterior hairline or 1 cun anterior to **Du-20**, on the line connecting **G.B.-15** and **G.B.-20**, or 2.25 cun lateral to the midline (midpoint of the line connecting **ST-8** and **Du-24**)

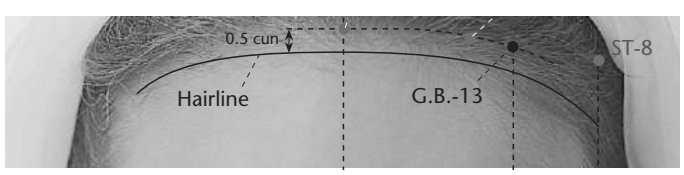

G.B.-13 (*benshen*) 'Root of the Spirit' 0.5 cun superior to the anterior hairline and 3 cun lateral to the midline, on the line between **Du-24** and **ST-8** (= 4.5 cun) at the junction of the lateral and medial third

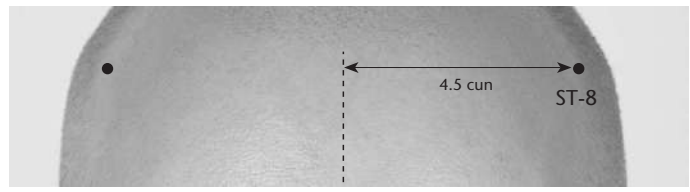

ST-8 (*touwei*) 'Head's Binding' At the temporal corner of the forehead, on the border of the temporalis muscle and 0.5 cun within the anterior hair line or 4.5 cun lateral to the anterior midline (**Du-24**)

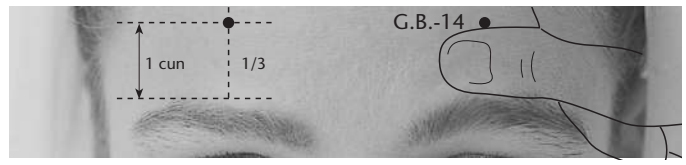

G.B.-14 (*yangbai*) 'Yang White' When looking straight ahead on the pupil line, approximately 1 cun superior to the midpoint of the eyebrow (centre of the eyebrow to the anterior hairline = 3 cun)

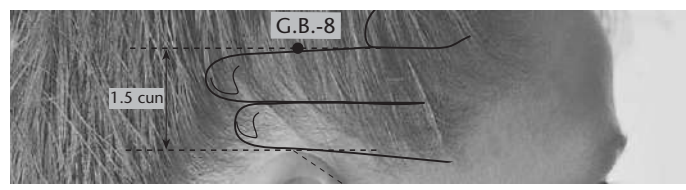

G.B.-8 (*shuaigu*) 'Leading Valley' 1.5 cun directly superior to the apex of the ear

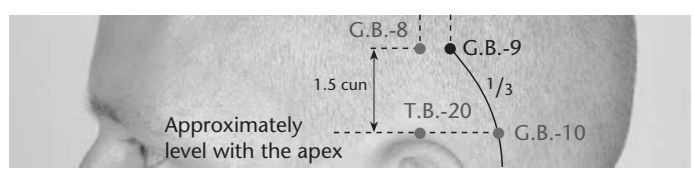

G.B.-9 (*tianchong*) 'Heavenly Rushing' 1.5 cun superior and 0.5 cun posterior to the apex of the ear or 0.5 cun posterior to **G.B.-8**

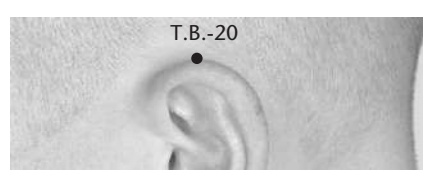

T.B.-20 (*jiaosun*) 'Minute Angle' Directly superior to the apex of the ear

Ch07-F10028.qxd 2/22/08 10:08 PM Page 624

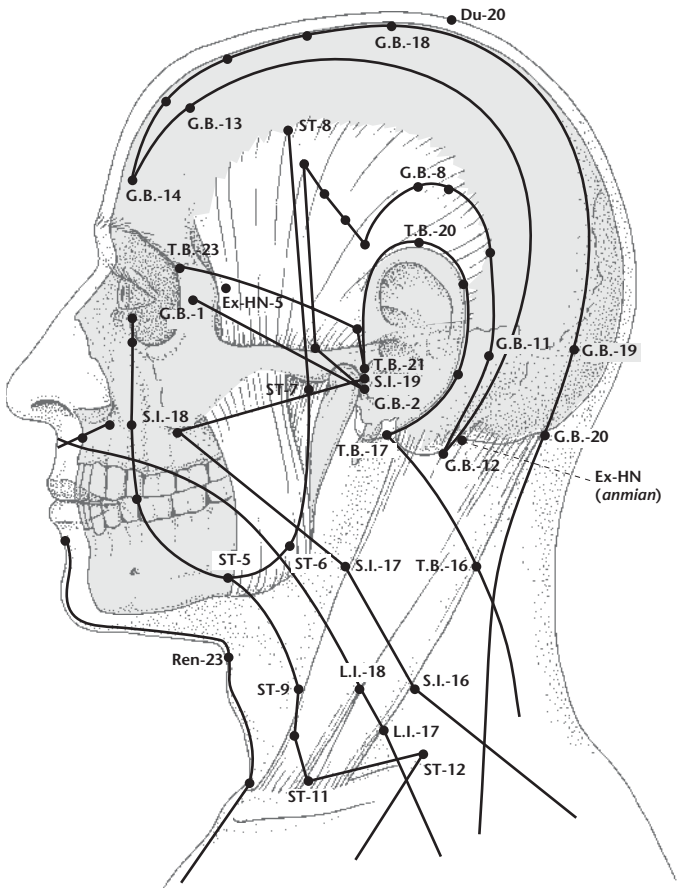

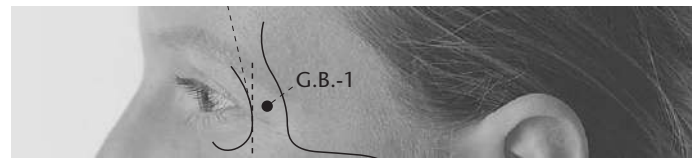

G.B.-1 (*tongziliao*) 'Pupil Crevice' 0.5 cun lateral to the outer canthus of the eye

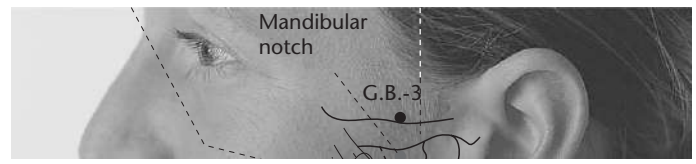

G.B.-3 (*shangguan*) 'Above the Joint' In a depression on the upper border of the zygomatic arch, approximately 1 cun anterior to the root of the ear, directly superior to **ST-7** (lower border of the zygomatic arch)

T.B.-21 (*ermen*) 'Ear Gate' Anterior to the ear, with the mouth open in the depression at the level of the supratragic notch, slightly superior to the condyloid process of the mandible

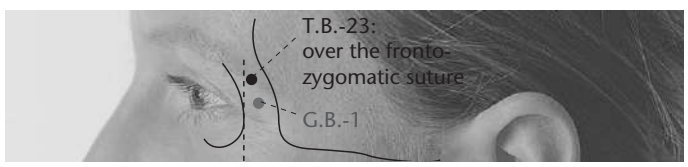

T.B.-23 (*sizhukong*) 'Silken Bamboo Hollow' In the depression (frontozygomatic suture) at the lateral end of the eyebrow

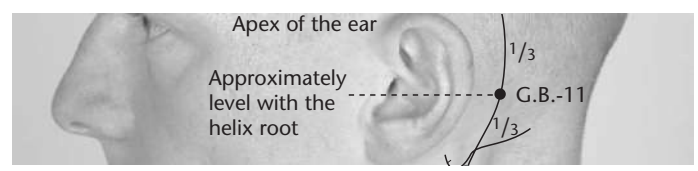

G.B.-11 (*touqiaoyin*) 'Yin Portals of the Head' At the junction of the lower third with the upper two-thirds of the curved line connecting **G.B.-9** and **G.B.-12**

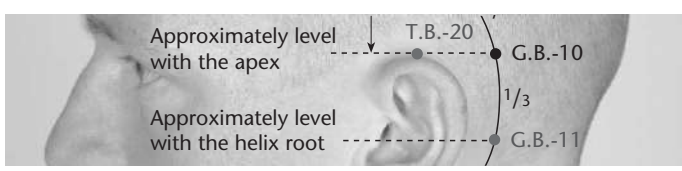

G.B.-10 (*fubai*) 'Floating White' At the junction of the upper third with the lower two-thirds of the curved line connecting **G.B.-9** and **G.B.-12**

S.I.-19 (*tinggong*) 'Palace of Hearing' Anterior to the ear with the mouth open, at the level of the middle of the tragus, in a depression between the tragus and condyloid process of the mandible

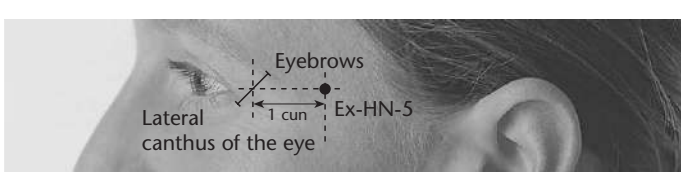

Ex-HN-5 (*taiyang*) 'Supreme Yang' In the depression approximately 1 cun posterior to the midpoint between the lateral end of the eyebrow and the outer canthus of the eye

* According to Deadman et al 1998.

Ch07-F10028.qxd 2/22/08 10:08 PM Page 625

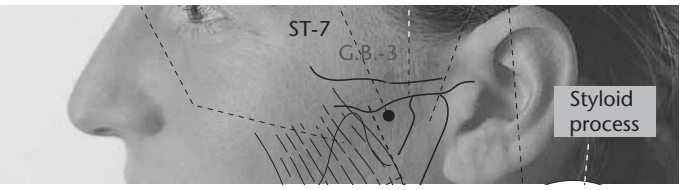

ST-7 (*xiaguan*) 'Below the Joint' With the mouth closed in the depression of the mandibular notch, between the coronoid and condyloid process of the mandible

G.B.-2 (*tinghui*) 'Meeting of Hearing' Anterior to the ear with the mouth open, in the depression at the level of the intertragic notch, on the lower border of the condyloid process of the mandible

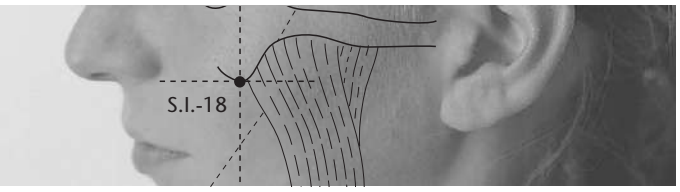

S.I.-18 (*quanliao*) 'Cheek Bone Crevice' In a depression at the intersection of a vertical line through the outer canthus of the eye with the lower border of the zygomatic bone

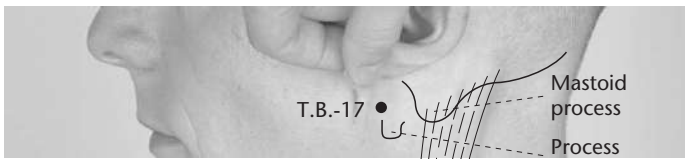

T.B.-17 (*yifeng*) 'Wind Screen' With the mouth open in the depression behind the earlobe, between the mastoid process and the mandible

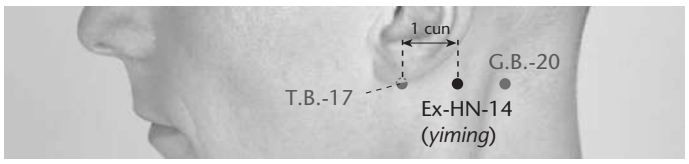

Ex-HN-14 (*yiming*) 'Eye Brightening' On the occiput, 1 cun posterior to **T.B.-17**, on a line connecting **T.B.-17** and **G.B.-20**

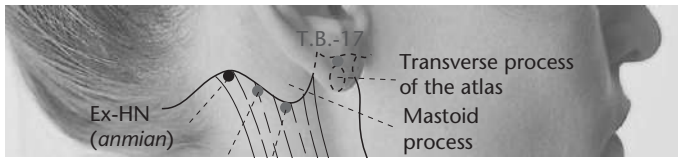

Ex-HN (*anmian*) 'Peaceful Sleep' Posterior to the ear, midway between **T.B.-17** and **G.B.-20**, posterior to the mastoid process

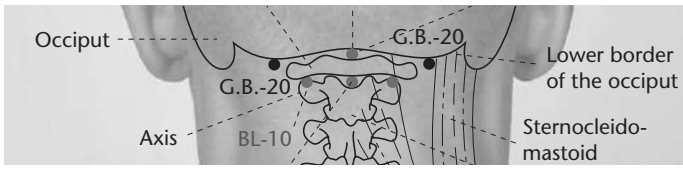

G.B.-20 (*fengchi*) 'Wind Pool' On the lower border of the occipital bone, in the depression between the origins of the sternocleidomastoid and trapezius

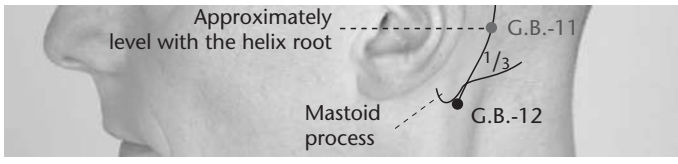

G.B.-12 (*wangu*) 'Mastoid Process' In a depression posterior and inferior to the mastoid process

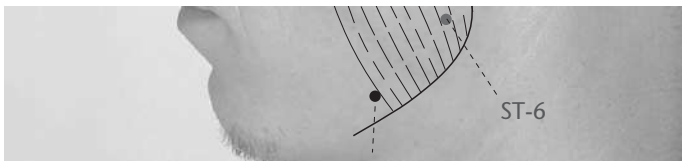

ST-6 (*jiache*) 'Jaw Bone' With the teeth clenched, on the highest prominence of the masseter, approximately one fingerbreadth anterior and superior to the mandibular angle

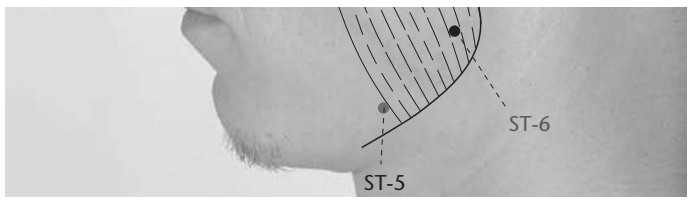

ST-5 (*daying*) 'Great Welcome' On the lateral mandible, anterior to the masseter (ask the patient to clench the teeth); a branch of the facial artery is palpable here

Ch07-F10028.qxd 2/22/08 10:08 PM Page 626

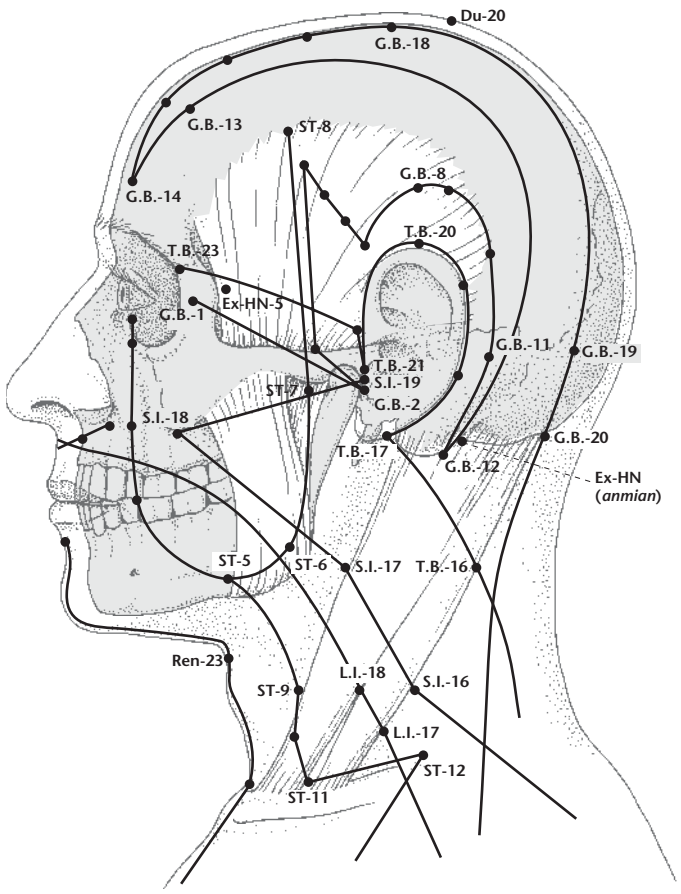

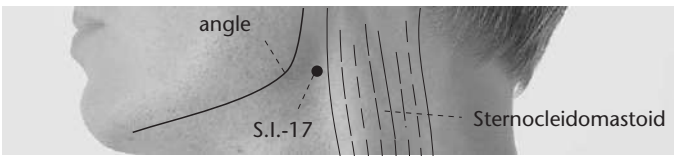

S.I.-17 (*tianrong*) 'Heavenly Appearance' Level with and posterior to the mandibular angle, on the anterior border of the sternocleidomastoid

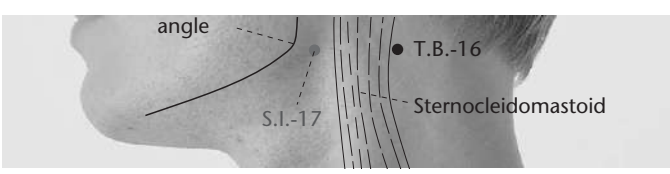

T.B.-16 (*tianyou*) 'Window of Heaven' posterior and inferior to the mastoid process, on the posterior border of the sternocleidomastoid, directly level with the mandibular angle

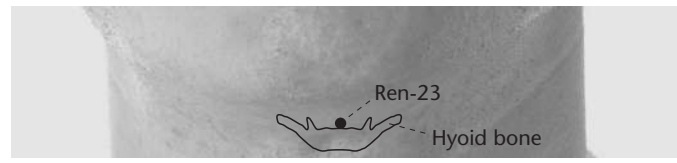

Ren-23 (*lianquan*) 'Corner Spring' On the anterior midline, superior to the upper border of the hyoid bone

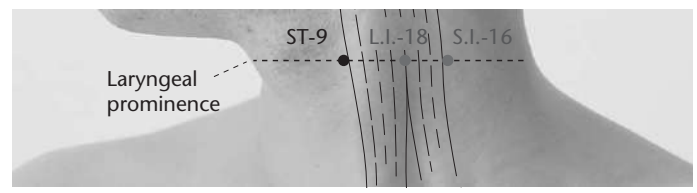

ST-9 (*renying*) Man's Welcome' Approximately 1.5 cun lateral to the anterior midline, level with the laryngeal prominence, at the anterior border of the sternocleidomastoid

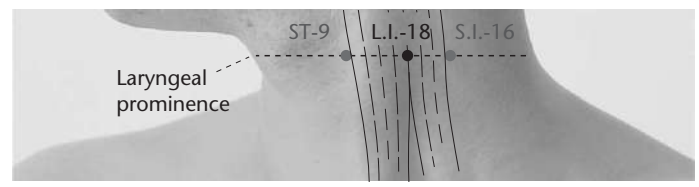

L.I.-18 (*futu*) 'Support the Prominence' At the level of the laryngeal prominence, between the sternal and clavicular heads of the sternocleidomastoid muscle

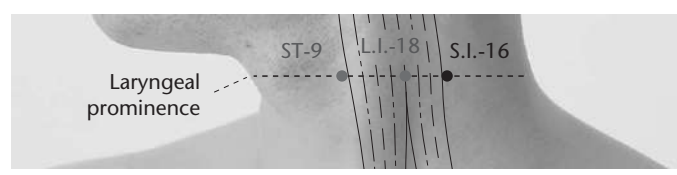

S.I.-16 (*tianchuang*) 'Heavenly Window' Approximately 3.5 cun lateral to the anterior midline, level with the laryngeal prominence, on the posterior border of the sternocleidomastoid

ST-10 (*shuitu*) 'Water Prominence' On the anterior border of the sternocleidomastoid, at the midpoint of a line connecting **ST-9** and **ST-11**

Ch07-F10028.qxd 2/22/08 10:08 PM Page 627

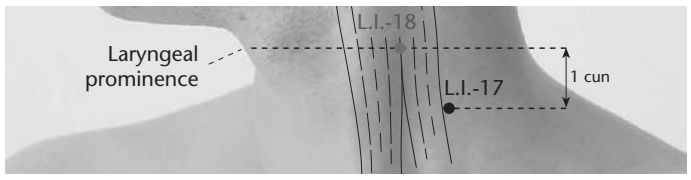

L.I.-17 (*tianding*) 'Heaven's Tripod' 1 cun inferior to the laryngeal prominence, on the posterior border of the sternocleidomastoid

ST-11 (*qishe*) 'Abode of Qi' On the upper border of the clavicle, between the tendons of the sternal and clavicular heads of the sternocleidomastoid

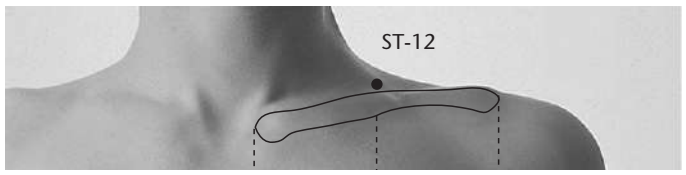

ST-12 (*quepen*) 'Empty Basin' In the supraclavicular fossa, superior to the midpoint of the clavicle, approximately 4 cun lateral to the midline

7.3 Occipital Region

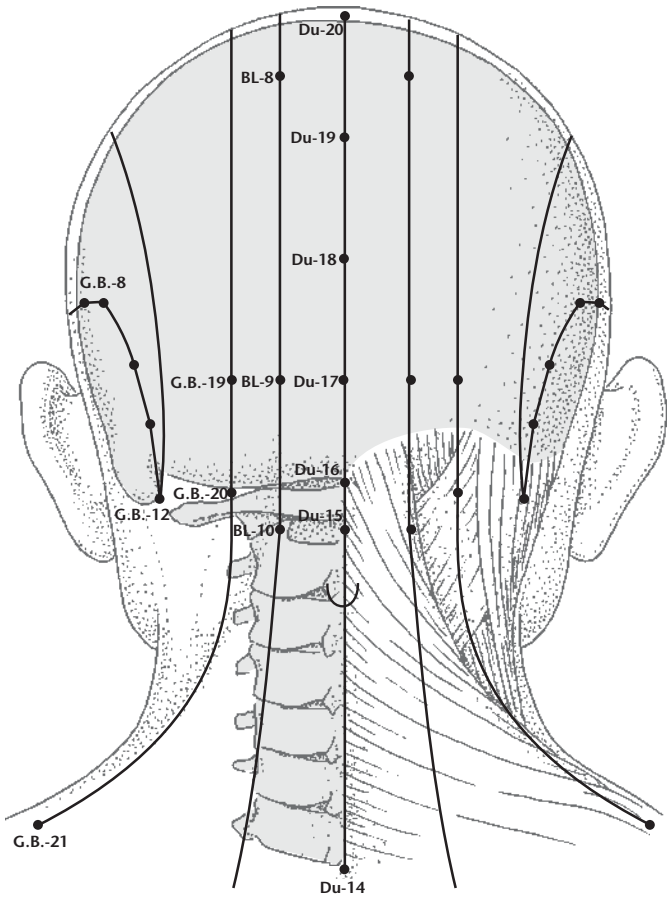

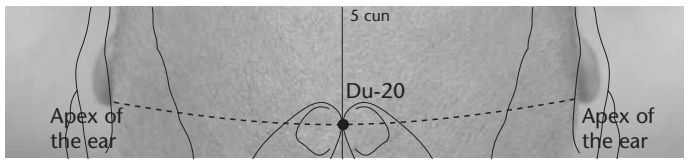

Du-20 (*baihui*) 'Hundred Meetings' At the junction of a line connecting the apices of the ears and the midline, 5 cun from the anterior or 7 cun from the posterior hairline respectively

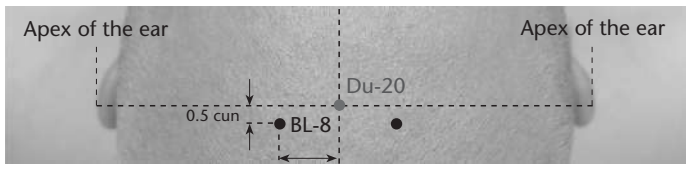

BL-8 (*luoque*) 'Declining Connection' 1.5 cun lateral to the midline and 5.5 cun superior to the anterior hairline, or 0.5 cun posterior to **Du-20**

Ch07-F10028.qxd 2/22/08 10:08 PM Page 628

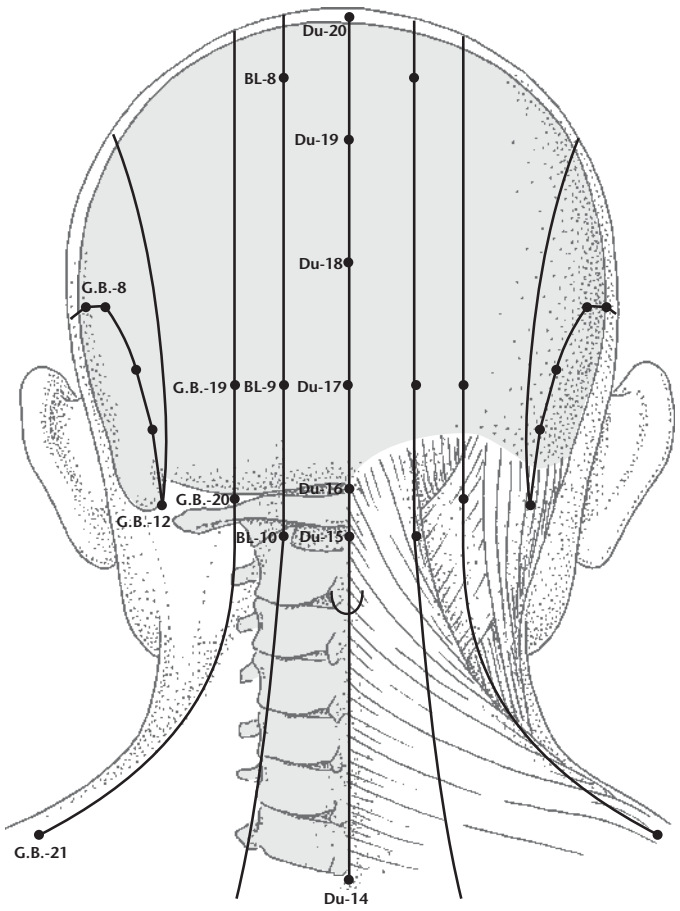

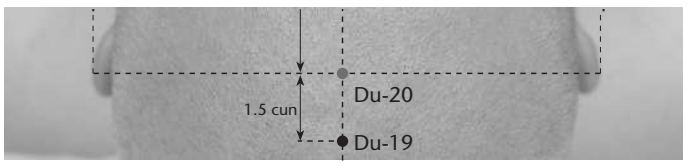

Du-19 (*houding*) 'Behind the Crown' On the midline, 1.5 cun posterior to **Du-20** or 3 cun superior to **Du-17**

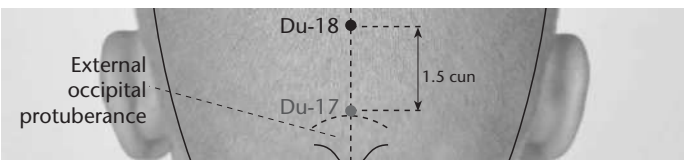

Du-18 (*qiangjian*) 'Unyielding Space' On the posterior midline, 1.5 cun superior to **Du-17** or 3 cun posterior to **Du-20**

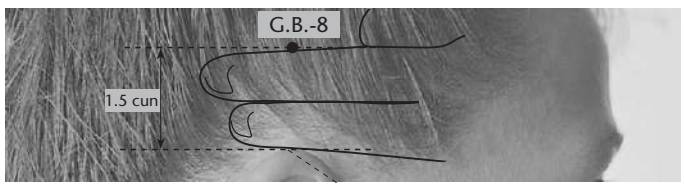

G.B.-8 (*shuaigu*) 'Leading Valley' 1.5 cun directly superior to the apex of the ear

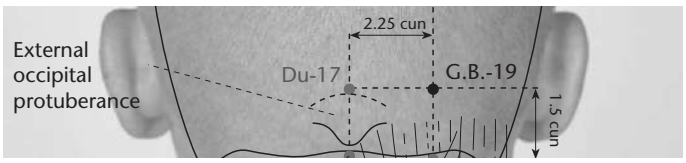

G.B.-19 (*naokong*) 'Brain Hollow' Level with the upper border of the external occipital protuberance (**Du-17**), on the line connecting **G.B.-15** and **G.B.-20**, and 2.25 cun lateral to the midline

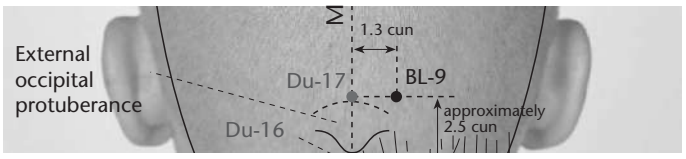

BL-9 (*yuzhen*) 'Jade Pillow' 2.5 cun superior to the posterior hairline and 1.3 cun lateral to the midline or **Du-17** (directly superior to the external occipital protuberance)

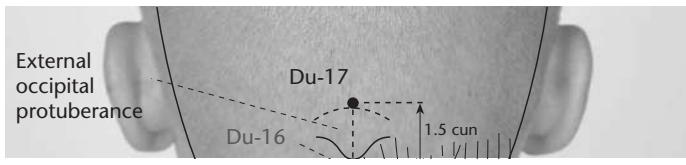

Du-17 (*naohu*) 'Brain's Door' On the midline, in a depression superior to the external occipital protuberance, approximately 2.5 cun superior to the posterior hairline, 1.5 cun superior to **Du-16**

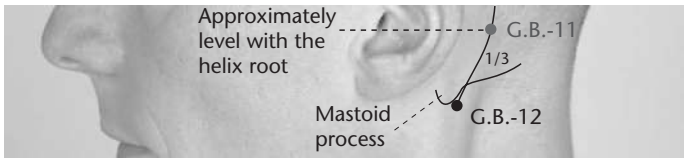

G.B.-12 (*wangu*) 'Mastoid Process' In a depression posterior and inferior to the mastoid process

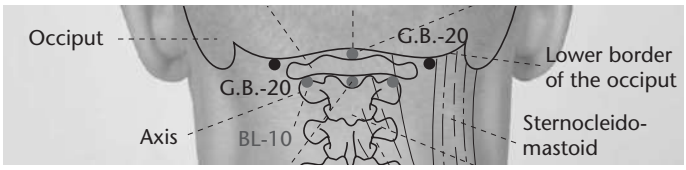

G.B.-20 (*fengchi*) 'Wind Pool' On the lower border of the occipital bone, in the depression between the origins of the sternocleidomastoid and trapezius

Ch07-F10028.qxd 2/22/08 10:08 PM Page 629

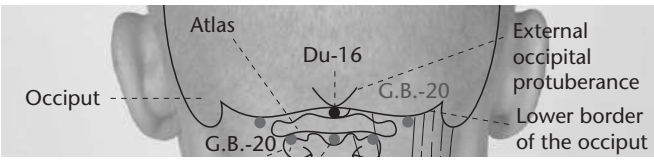

Du-16 (*fengfu*) 'Palace of Wind' On the posterior midline, directly below the external occipital protuberance, in the depression between the origins of the trapezius

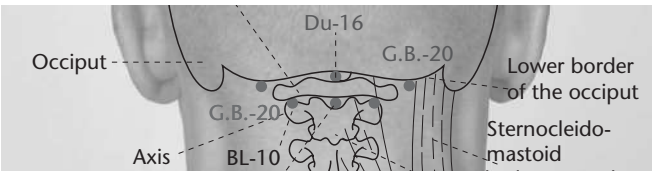

BL-10 (*tianzhu*) 'Celestial Pillar' Below the lower border of the occiput, approximately 1.3 cun lateral to the midline (**Du-15**), at the insertion of the trapezius

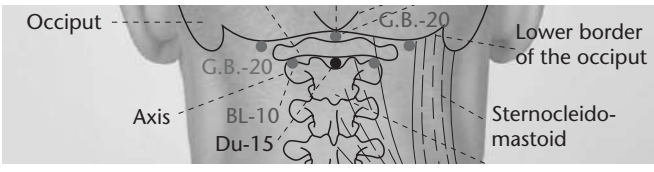

Du-15 (*yamen*) 'Gate of Muteness' On the posterior midline, in the depression between the 1st (atlas) and 2nd cervical vertebrae (axis = first palpable spinous process), approximately 0.5 cun below **Du-16**

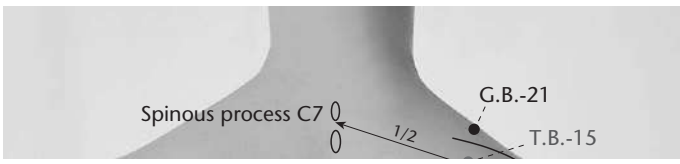

G.B.-21 (*jianjing*) 'Shoulder Well' At the highest point of the shoulder, midway between the lower border of the spinous process of C7 and the lateral extremity of the acromion

Du-14 (*dazhui*) 'Great Vertebra' Below the spinous process of C7

7.4 Vertex

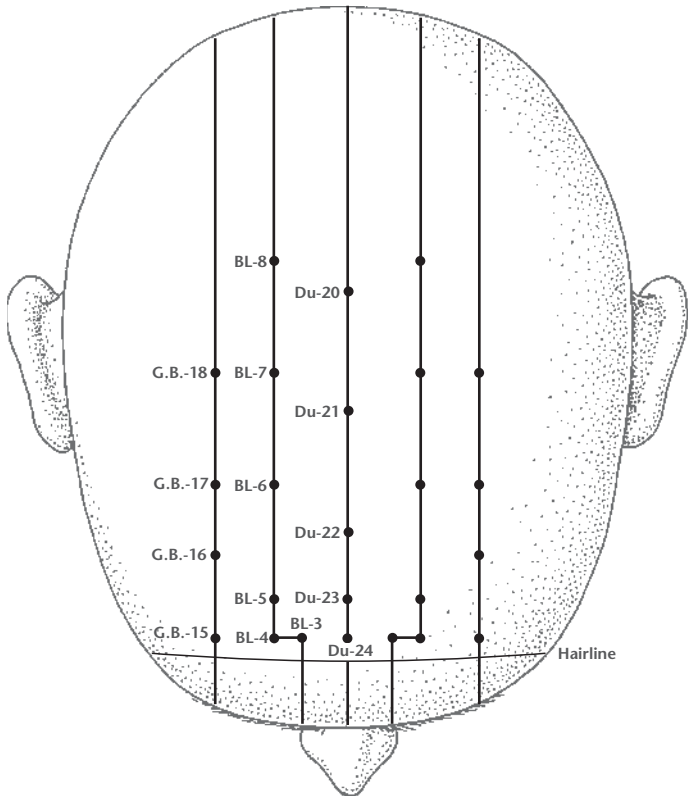

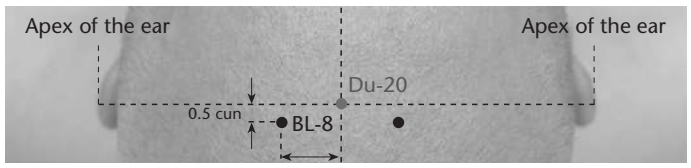

BL-8 (*luoque*) 'Declining Connection' 1.5 cun lateral to the midline and 5.5 cun superior to the anterior hairline, or 0.5 cun posterior to **Du-20**

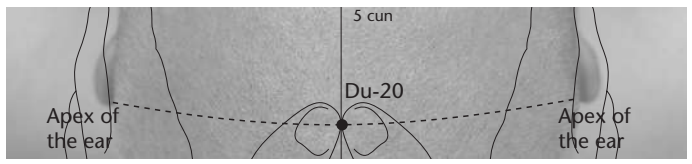

Du-20 (*baihui*) 'Hundred Meetings' At the junction of a line connecting the apices of the ears and the midline, 5 cun from the anterior or 7 cun from the posterior hairline respectively

Ch07-F10028.qxd 2/22/08 10:08 PM Page 630

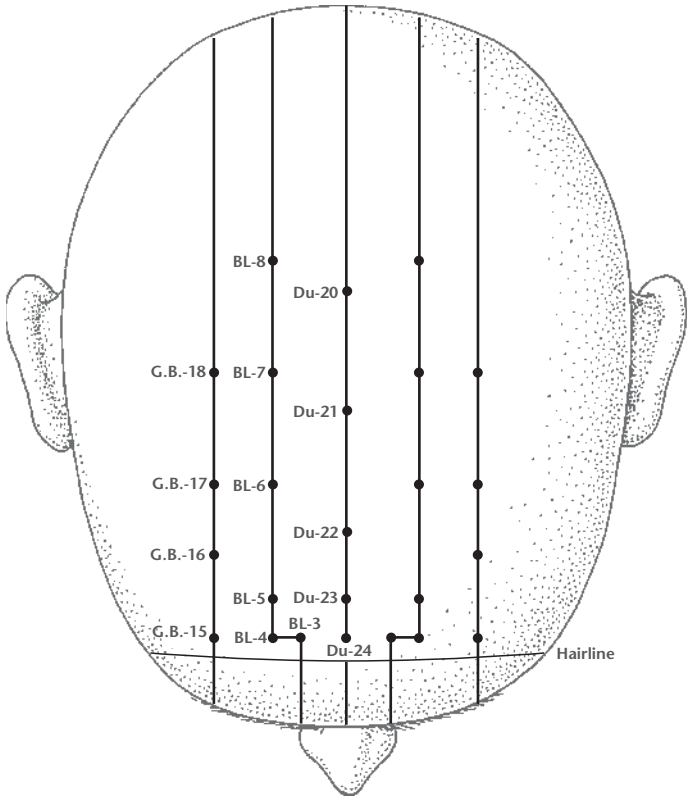

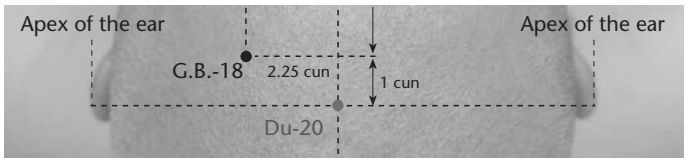

G.B.-18 (*chengling*) 'Support Spirit' 4 cun superior to the anterior hairline or 1 cun anterior to **Du-20**, on the line connecting **G.B.-15** with **G.B.-20**, or 2.25 cun lateral to the midline (midpoint of the line connecting **ST-8** and **Du-24**)

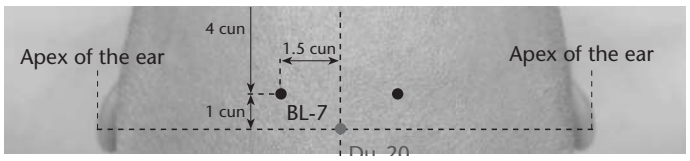

BL-7 (*tongtian*) 'Heavenly Connection' 1.5 cun lateral to the midline and 4 cun superior to the anterior hairline or 1 cun anterior to **Du-20**

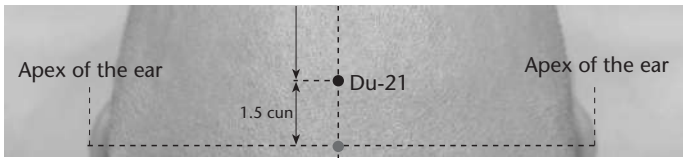

Du-21 (*qianding*) 'In Front of the Crown' On the midline, 1.5 cun anterior to **Du-20** or 3.5 cun above the anterior hairline

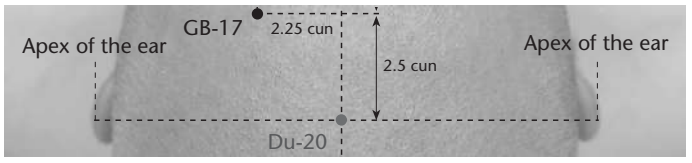

G.B.-17 (*zhengying*) 'Upright Nutrition' 2.5 cun superior to the anterior hairline, on the pupil line, on the line connecting **G.B.-15** and **G.B.-20** (= 2.25 cun lateral to the midline or midway between **Du-23** and **ST-8**)

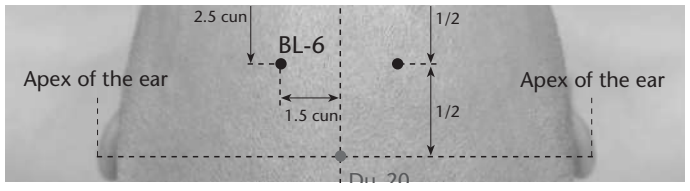

BL-6 (*chengguang*) 'Receiving Light' 2.5 cun superior to the anterior hairline and 1.5 cun lateral to the midline, or one-third of the distance between **Du-24** and **ST-8**

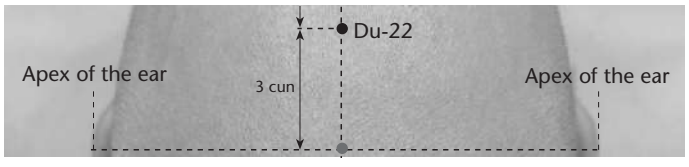

Du-22 (*xinhui*) 'Fontanelle Meeting' On the midline, 2 cun superior to the anterior hairline or 3 cun anterior to **Du-20**

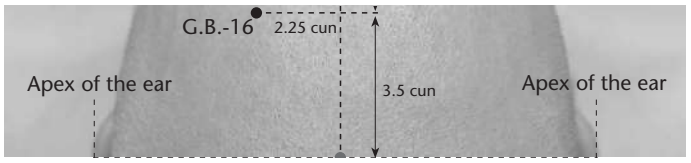

G.B.-16 (*muchuang*) 'Window of the Eye' 1.5 cun superior to the anterior hairline, on the pupil line, on the line connecting **G.B.-15** and **G.B.-20** (= 2.25 cun lateral to the midline or midway between **Du-24** and **ST-8**)

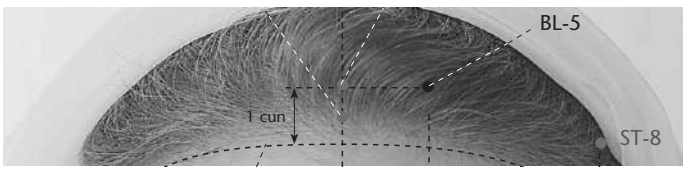

BL-5 (*wuchu*) 'Fifth Place' 1 cun superior to the anterior hairline and 1.5 cun lateral to the midline (directly above **BL-4**, level with **Du-23**)

Ch07-F10028.qxd 2/22/08 10:08 PM Page 631

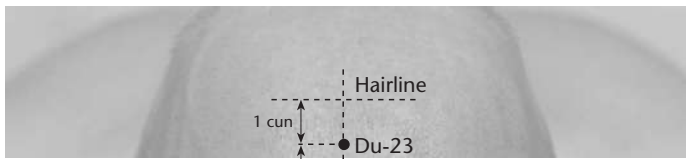

Du-23 (*shangxing*) 'Upper Star' On the midline, 1 cun superior to the anterior hairline

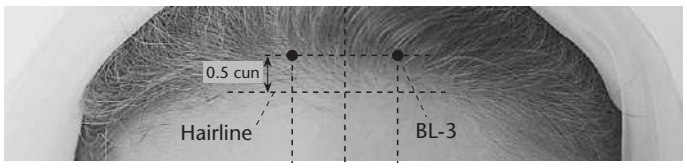

BL-3 (*meichong*) 'Eyebrows' Pouring' 0.5 cun within the anterior hairline, superior to the medical canthus of the eye (**BL-1**)

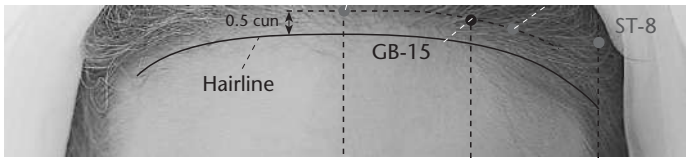

G.B.-15 (*toulinqi*) 'Head Governor of Tears' On the pupil line when looking straight ahead, 0.5 cun superior to the anterior hairline, at the midpoint of the line between **Du-24** and **ST-8** (= 4.5 cun)

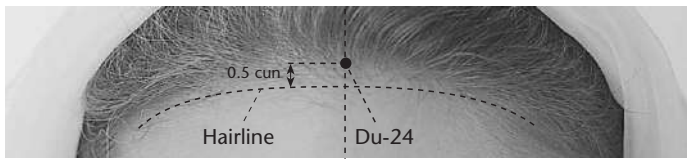

Du-24 (*shenting*) 'Courtyard of the Spirit' On the midline, 0.5 cun superior to the anterior hairline

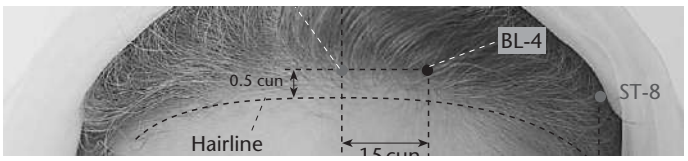

BL-4 (*qucha*) 'Crooked Curve' 0.5 cun superior to the anterior hairline and 1.5 cun lateral to the midline, or one-third of the distance between **Du-24** and **ST-8** (= 4.5 cun)

Ch07-F10028.qxd 2/22/08 10:08 PM Page 632

7.5 Back

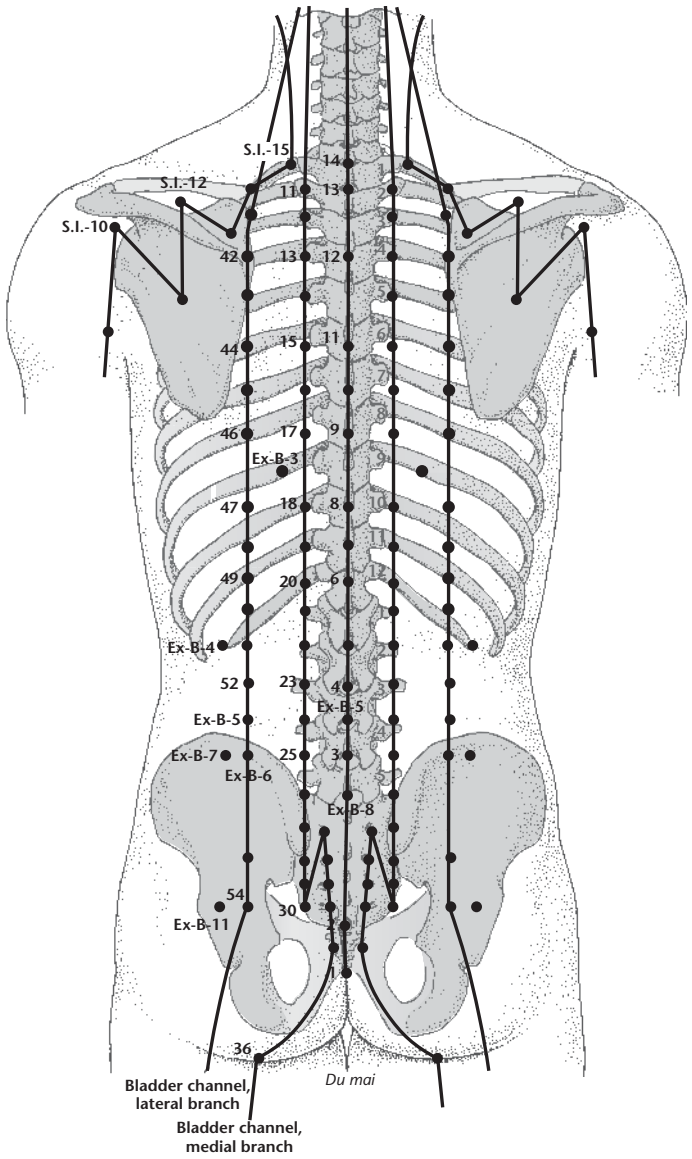

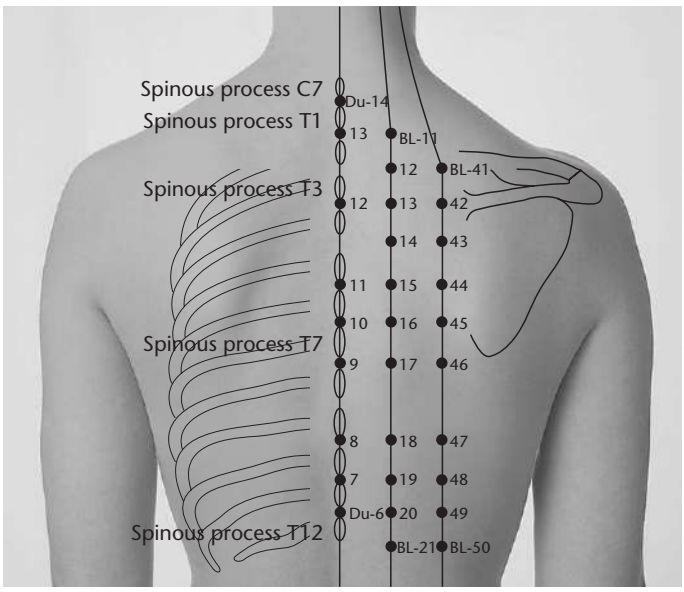

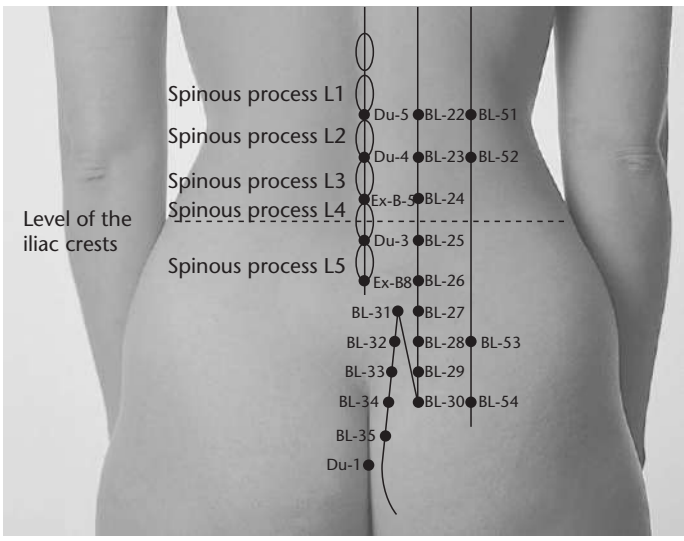

Ch07-F10028.qxd 2/22/08 10:08 PM Page 633

Vertebra	Below the spinous process or in the sacral foramen	0.5 cun lateral to the midline, level with the lower border of the spinous process or sacral foramen	1.5 cun lateral to the midline, level with the lower border of the spinous process or sacral foramen	3 cun lateral to the midline, level with the lower border of the spinous process or sacral foramen
C7	Du-14	Ex-B-1 *(dingchuan)*	**2 cun lateral** S.I.-15	**3.5 cun lateral** Ex-B (*jiehexue*)
T1	Du-13	A point of Ex-B-2 (*huatuojiaji*)	BL-11	S.I.-14
T2		A point of Ex-B-2 (*huatuojiaji*)	BL-12	BL-41
T3	Du-12	A point of Ex-B-2 (*huatuojiaji*)	BL-13	BL-42
T4		A point of Ex-B-2 (*huatuojiaji*)	BL-14	BL-43
T5	Du-11	A point of Ex-B-2 (*huatuojiaji*)	BL-15	BL-44
T6	Du-10	A point of Ex-B-2 (*huatuojiaji*)	BL-16	BL-45
T7	Du-9	A point of Ex-B-2 (*huatuojiaji*)	BL-17	BL-46
T8		A point of Ex-B-2 (*huatuojiaji*)	Ex-B-3 (*weiwanxiashu*)	
T9	Du-8	A point of Ex-B-2 (*huatuojiaji*)	BL-18	BL-47
T10	Du-7	A point of Ex-B-2 (*huatuojiaji*)	BL-19	BL-48
T11	Du-6	A point of Ex-B-2 (*huatuojiaji*)	BL-20	BL-49
T12		A point of Ex-B-2 (*huatuojiaji*)	BL-21	BL-50
L1	Du-5	A point of Ex-B-2 (*huatuojiaji*)	BL-22	BL-51/ Ex-B-4 (**3.5 cun lateral**)
L2	Du-4	A point of Ex-B-2 (*huatuojiaji*)	BL-23	BL-52
L3	Ex-B-5 (*xiajishu*)	A point of Ex-B-2 (*huatuojiaji*)	BL-24	Ex-B-5 (according to WHO a set of 3 points)
L4	Du-3	A point of Ex-B-2 (*huatuojiaji*)	BL-25	Ex-B-6/Ex-B-7 (**3.5 cun lateral**)
L5	Ex-B-8	A point of Ex-B-2 (*huatuojiaji*)	BL-26	
1st sacral foramen	BL-31		BL 27	
2nd sacral foramen	BL-32		BL 28	BL-53
3rd sacral foramen	BL-33		BL 29	
4th sacral foramen	BL-34		BL 30	BL-54/Ex-B-11 (3.5 cun lateral)
Sacral hiatus	Du- 2			
		BL-35 (0.5 cun lateral and inferior to Du-2)		
Between the coccyx and the anus	Du-1			

Ch07-F10028.qxd 2/22/08 10:08 PM Page 634

7.6 Frontal and Lateral Thorax and Abdomen

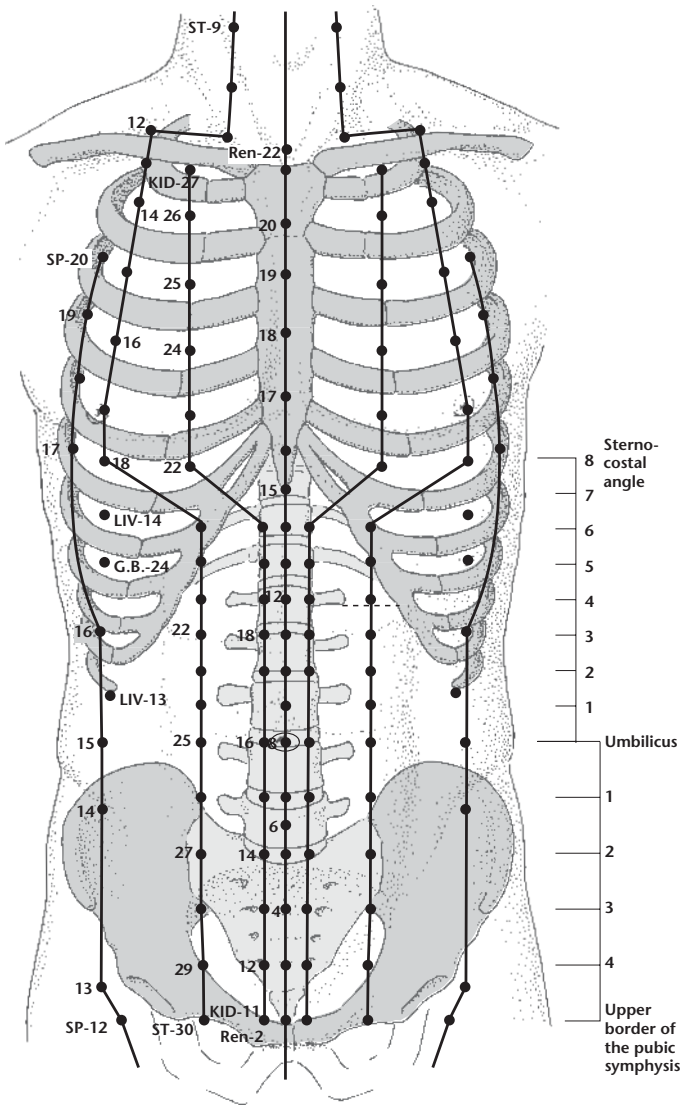

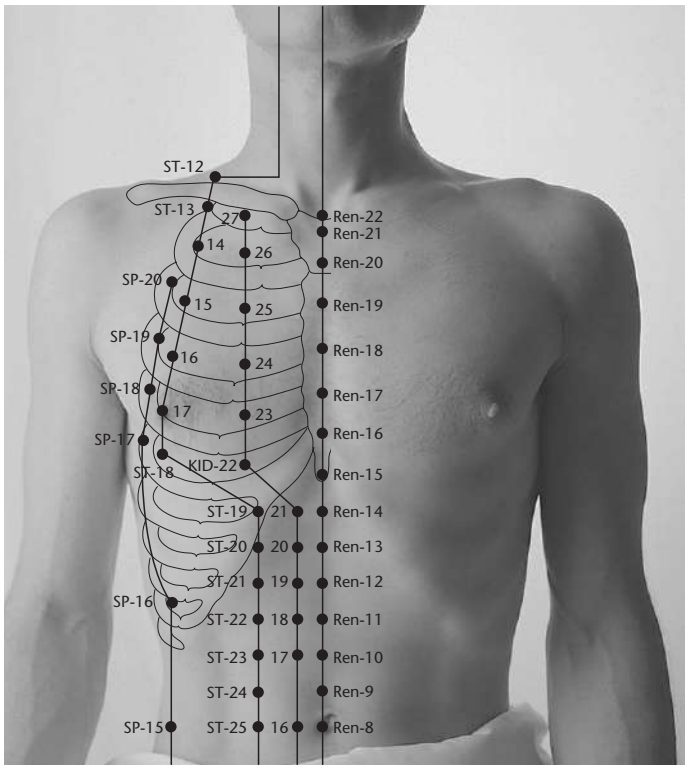

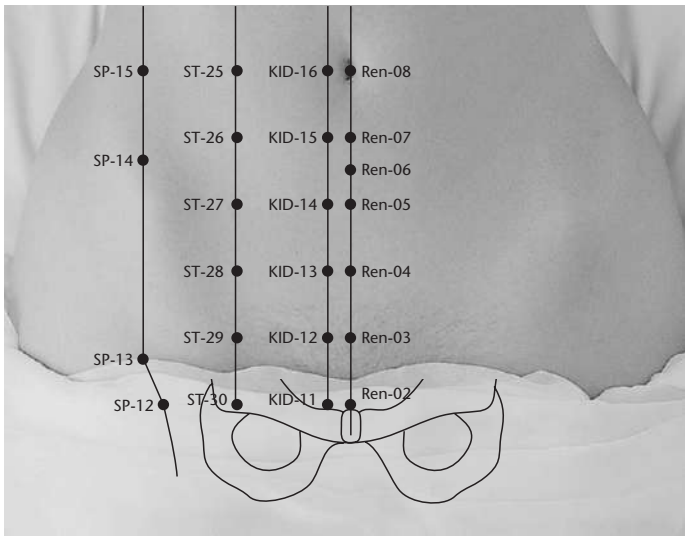

Ch07-F10028.qxd 2/22/08 10:08 PM Page 635

Anatomical level	***Ren Mai* (midline)**	**KID channel (2 cun lateral to the midline)**	**ST channel (4 cun lateral to the midline)**	**Other (6 cun lateral to the midline)**
Below the clavicle	≈ Ren-21	KID-27	ST-13	LU-2
1st ICS	Ren-20	KID-26	ST-14	LU-1
2nd ICS	Ren-19	KID-25	ST-15	SP-20
3rd ICS	Ren-18	KID-24	ST-16	SP-19
4th ICS	Ren-17	KID-23	ST-17 (nipple)	SP-18/P-1 (1 cun lateral to the nipple)/G.B.-22 (3 cun inferior to the apex of the axilla)/G.B.-23 (1 cun anterior to G.B.-22)
5th ICS	Ren 16	KID-22	ST-18	SP-17
Anatomical level	***Ren mai* (midline)**	**KID-channel (0.5 cun lateral to the midline)**	**ST channel (2 cun lateral to the midline)**	**Other (4 cun lateral to the midline)**
6 cun superior/6th ICS	Ren-14	KID-21	ST-19	LIV-14 (6th ICS)
5 cun superior/7th ICS	Ren-13	KID-20	ST-20	G.B.-24 (7th ICS)
4 cun superior	Ren-12	KID-19	ST-21	
3 cun superior	Ren-11	KID-18	ST-22	SP-16
2 cun superior	Ren-10	KID-17	ST-23	Ex-CA (*weishang*)
1 cun superior	Ren-9		ST-24	≈ LIV-13 (free end of the 11th rib)
Level of umbilicus	Ren-8	KID-16	ST-25	SP-15, G.B.-26 (inferior to the free end of the 12th rib, level with the umbilicus)
1 cun inferior	Ren-7	KID-15	ST-26	
(1.5 cun inferior)	Ren-6			
2 cun inferior	Ren-5	KID-14	ST-27	
(2.5 cun inferior)	Ex-CA (*zhishi*)			
3 cun inferior	Ren-4	KID-13	ST-28	Ex-CA (*tituo*)/Ex-CA (*qimen*)/ Ex-CA (*yijing*)/(4/3/1 cun lateral to the midline)/≈ G.B.-27 (anterior and medical to the ASIS → 3)
4 cun inferior	Ren-3	KID-12	ST-29	Ex-CA (*zigong*), 3 cun lateral to the midline
5 cun inferior **symphysis**	Ren-2	KID-11	ST-30	SP-12 (3.5 cun lateral to the midline)

Note: The intercostal spaces and the ribs curve upwards laterally!

Ch07-F10028.qxd 2/22/08 10:08 PM Page 636

7.7 Posterior Shoulder and Arm

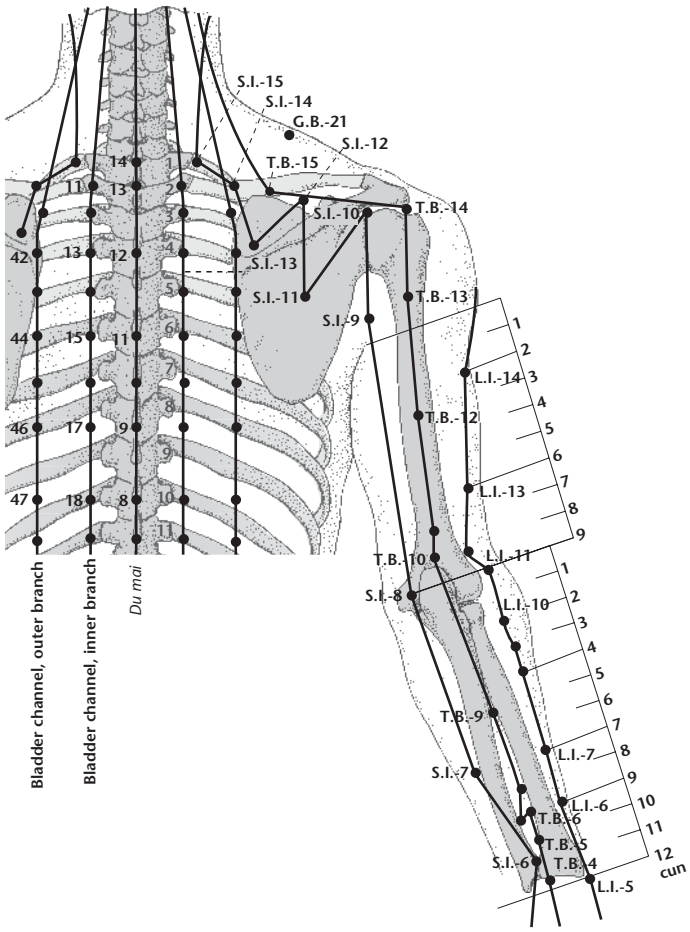

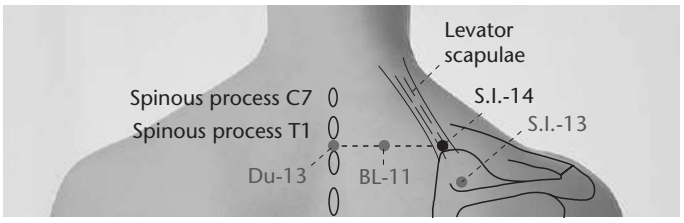

S.I.-14 (*jianwaishu*) 'Outer Shoulder Shu' 3 cun lateral to the lower border of the spinous process of the T1 (**Du-13**), at the insertion of the levator scapulae

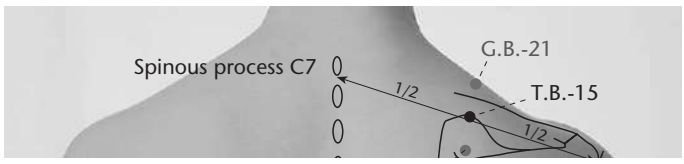

T.B.-15 (*tianliao*) 'Heavenly Crevice' At the midpoint of an imaginary line between the spinous process of the C7 and the lateral extremity of the acromion

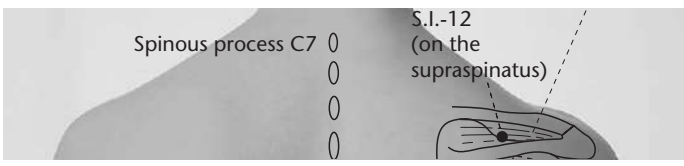

S.I.-12 (*bingfeng*) 'Grasping the Wind' Directly superior to **S.I.-11**, in the centre of the supraspinous fossa

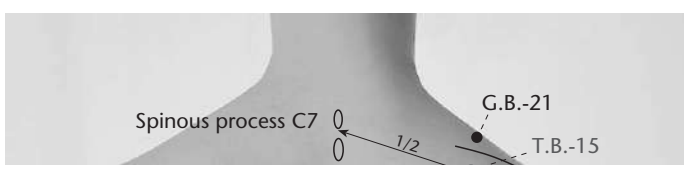

G.B.-21 (*jianjing*) 'Shoulder Well' At the highest point of the shoulder, midway between the lower border of the spinous process of C7 and the lateral extremity of the acromion

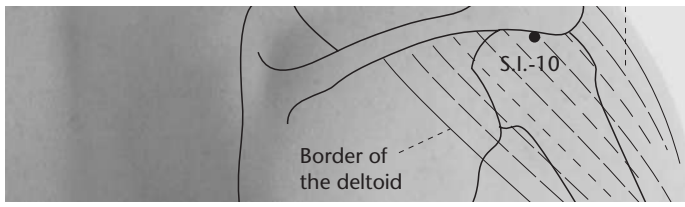

S.I.-10 (*naoshu*) 'Upper Arm Shu' With the arm adducted on a line extending in a superior direction from the posterior axillary fold, on the lower border of the scapular spine

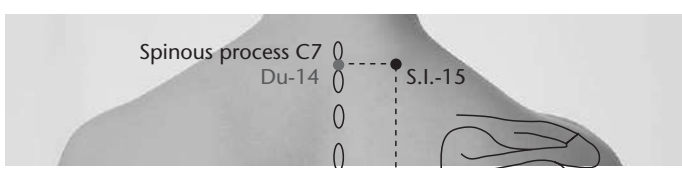

S.I.-15 (*jianzhongshu*) 'Middle Shoulder Shu' 2 cun lateral to the lower border of the spinous process of the T7 (**Du-14**)

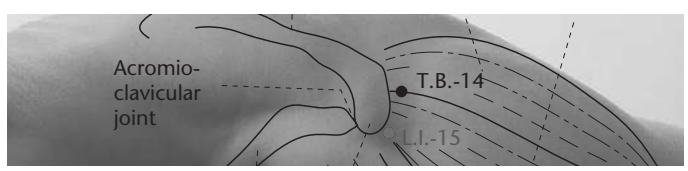

T.B.-14 (*jianliao*) 'Shoulder Crevice' Inferior to the lateral extremity of the acromion between the acromial and spinal portions of the deltoid, or with the arm abducted in the posterior of the two depressions on the shoulder joint

Ch07-F10028.qxd 2/22/08 10:08 PM Page 637

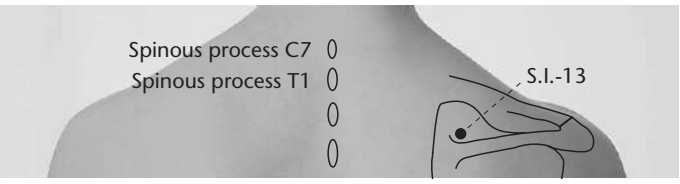

S.I.-13 (*quyuan*) 'Crooked Wall' At the medial end of the supraspinous fossa

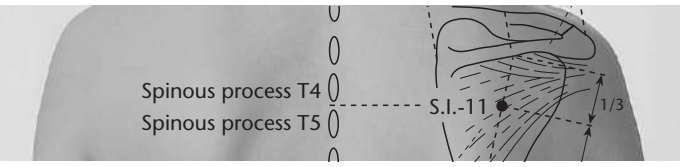

S.I.-11 (*tianzong*) 'Heavenly Gathering' On the scapula, in a depression on the infraspinatus, one third of the distance from the midpoint of the scapular spine and the inferior angle of the scapula

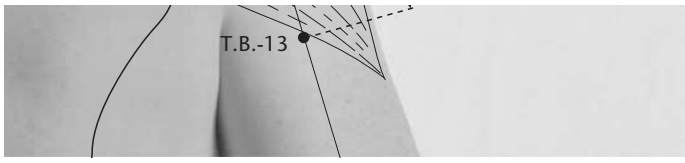

T.B.-13 (*naohui*) 'Upper Arm Meeting' 3 cun distal to the lateral extremity of the acromion (location of **T.B.-14**), on a line connecting **T.B.-14** and the olecranon, at the junction of this line with the margin of the deltoid

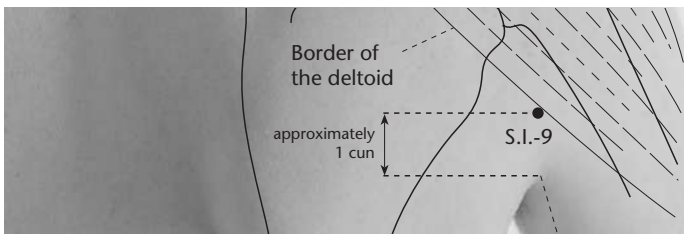

S.I.-9 (*jianzhen*) 'True Shoulder' With the arm adducted on a line extending in a superior direction from the posterior axillary fold, on the lower border of the deltoid, 1 cun superior to the posterior axillary fold

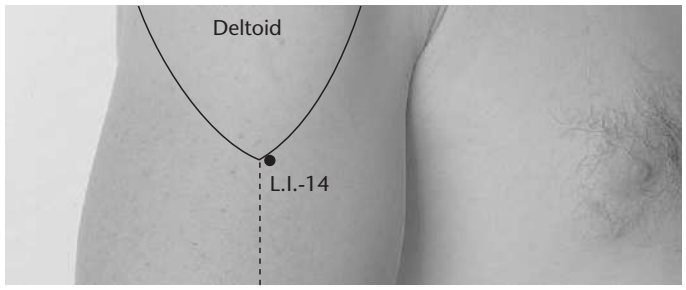

L.I.-14 (*binao*) 'Upper Arm' 7 cun proximal to **L.I.-11**, at the distal end of the deltoid, on a line connecting **L.I.-11** and **L.I.-15**

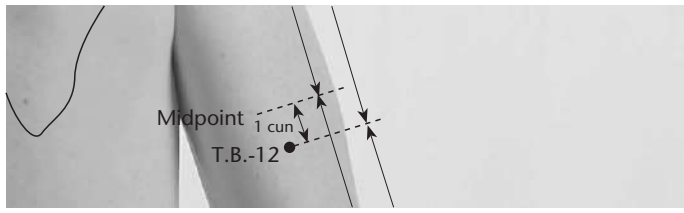

T.B.-12 (*xiaoluo*) 'Dispersing Luo River' 4 cun proximal to **T.B.-10** on a line connecting the olecranon and the lateral extremity of the acromion (location of **T.B.-14**)

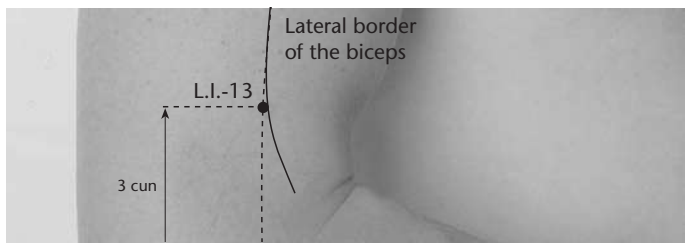

L.I.-13 (*shouwuli*) 'Arm Five Miles' On the lateral aspect of the upper arm, 3 cun proximal to **L.I.-11**

T.B.-10 (*tianjing*) 'Heavenly Well' With the elbow flexed in a depression approximately 1 cun proximal to the olecranon

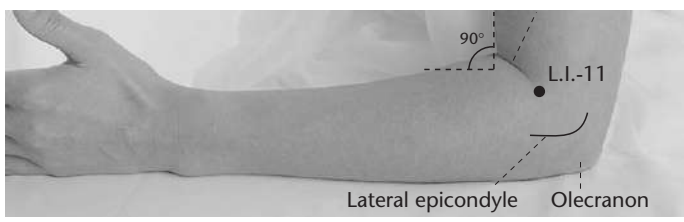

L.I.-11 (*quchi*) 'Pool at the Crook' With the elbow flexed, on the lateral end of the elbow crease, in a depression between the end of the crease and the lateral epicondyle of the humerus, on the extensor carpi radialis longus muscle

Ch07-F10028.qxd 2/22/08 10:08 PM Page 638

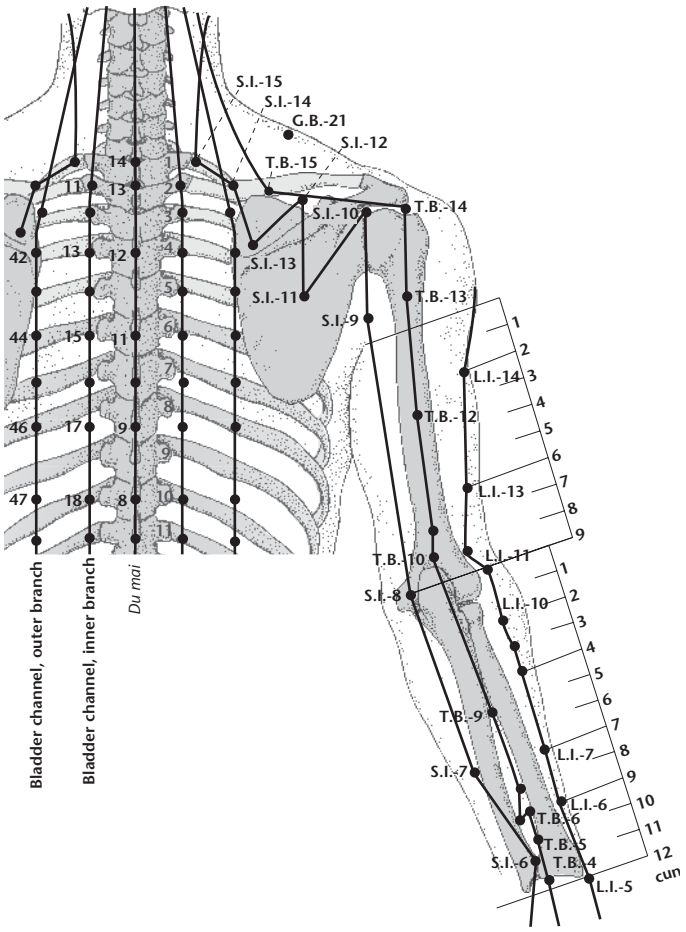

L.I.-9 (*shanglian*) 'Upper Angle' 3 cun distal to **L.I.-11**, on the line connecting **L.I.-5** and **L.I.-11**

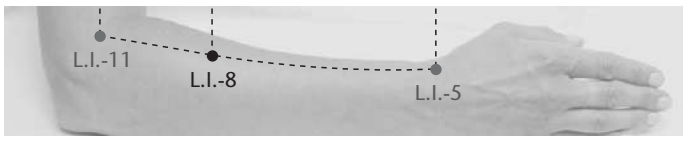

L.I.-8 (*xialian*) 'Lower Angle' 4 cun distal L.I.-11, on the line connecting **L.I.-5** and **L.I.-11**

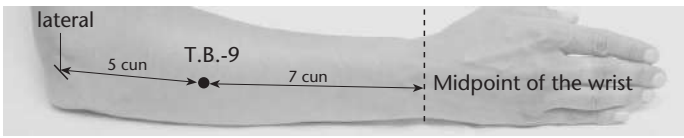

T.B.-9 (*sidu*) 'Four Rivers' 5 cun distal to the olecranon or 7 cun proximal to **T.B.-4**, between the radius and the ulna

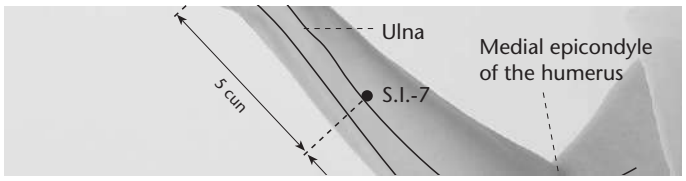

S.I.-7 (*zhizheng*) 'Branch of the Upright' 5 cun proximal to the wrist joint space, on the line connecting **S.I.-5** and **S.I.-8**

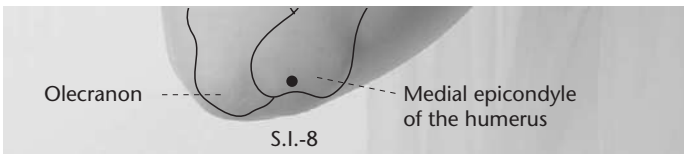

S.I.-8 (*xiaohai*) 'Small Sea' With the elbow flexed in the depression between the olecranon process of the ulna and the medial epicondyle of the humerus

L.I.-7 (*wenliu*) 'Warm Flow' 5 cun proximal to **L.I.-5** or midway between **L.I.-5** and **L.I.-11**

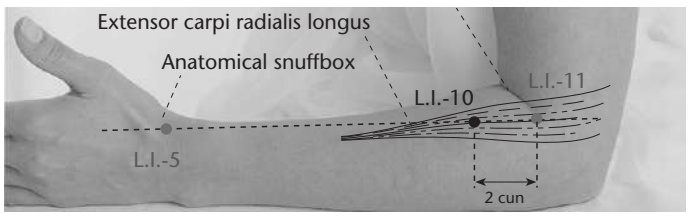

L.I.-10 (*shousanli*) 'Arm Three Miles' 2 cun distal to **L.I.-11**, on the line connecting **L.I.-5** and **L.I.-11**

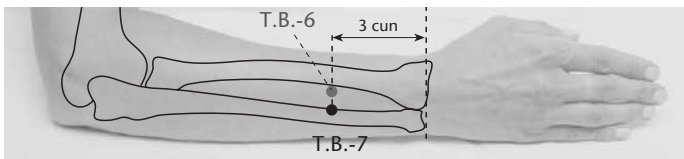

T.B.-7 (*huizhong*) 'Ancestral Meeting' 3 cun proximal to **T.B.-4**, between radius and ulna, ulnar to extensor digitorum communis

Ch07-F10028.qxd 2/22/08 10:09 PM Page 639

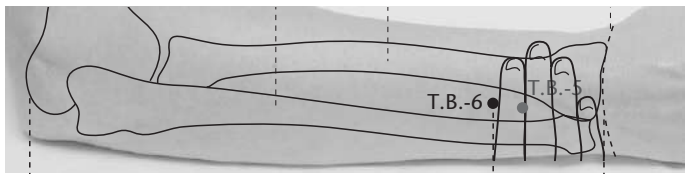

T.B.-6 (*zhigou*) 'Branch Ditch' 3 cun proximal to **T.B.-4**, between radius and ulna, radial to the tendon of extensor digitorum communis

S.I.-6 (*yanglao*) 'Support the Aged' In the depression radial and proximal to the styloid process of the ulna that forms when turning the hand from a pronated to a supinated position.

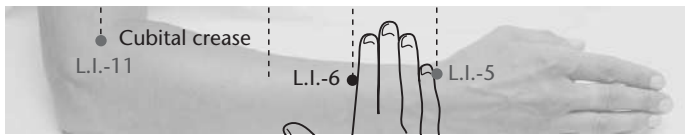

L.I.-6 (*pianli*) 'Veering Passage' 3 cun proximal to **L.I.-5**, on the line connecting **L.I.-5** and **L.I.-11**

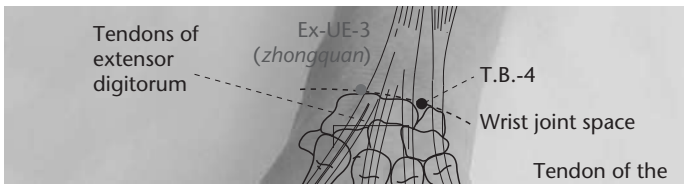

T.B.-4 (*yangchi*) 'Yang Pool' On the dorsum of the wrist, in the gap ulnar to the tendons of the extensor digitorum

T.B.-5 (*waiguan*) 'Outer Pass' 2 cun proximal to **T.B.-4** (on the dorsal wrist joint space, between radius and ulna.

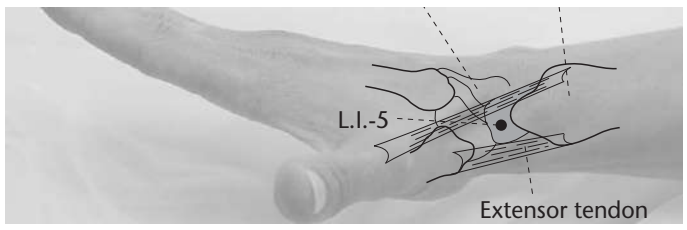

L.I.-5 (*yangxi*) 'Yang Stream' With the thumb abducted in the depression between the tendons of the extensor pollicis longus and brevis muscles ('anatomical snuff box')

Ch07-F10028.qxd 2/22/08 10:09 PM Page 640

7.8 Anterior Shoulder and Arm

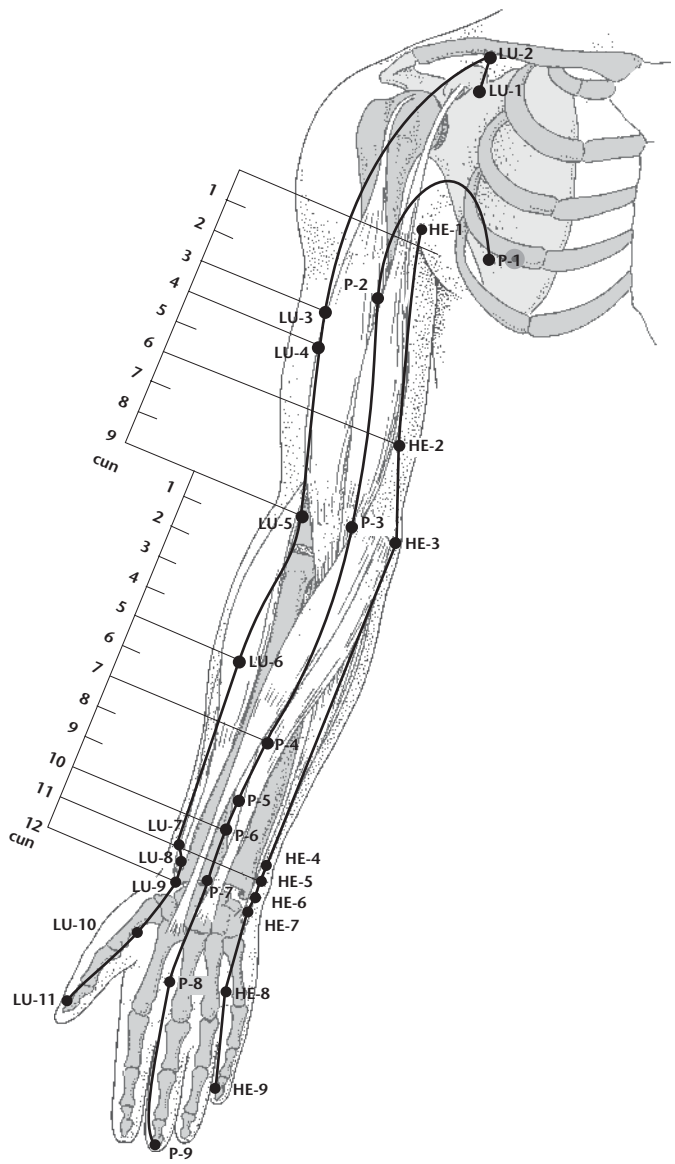

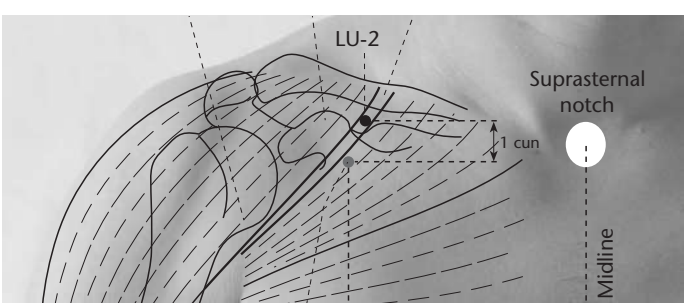

LU-2 (*yunmen*) 'Cloud Gate' 6 cun lateral to the anterior midline, below the clavicle, in the centre of the deltopectoral triangle.

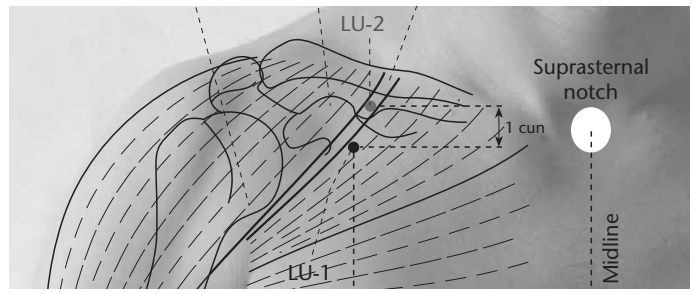

LU-1 (*zhongfu*) 'Middle Palace' 6 cun lateral to the anterior midline and approximately 1 cun inferior to **LU-2**, slightly medial to the lower border of the coracoid process

HE-1 (*jiquan*) 'Summit Spring' With the arm abducted, in the centre of the axilla, medial to the axillary artery

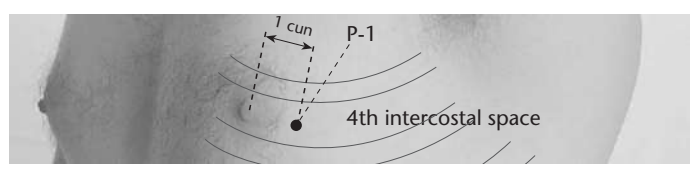

P-1 (*tianchi*) 'Heavenly Pool' In the 4th ICS, 1 cun lateral to the nipple

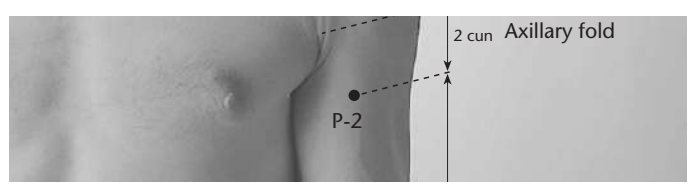

P-2 (*tianquan*) 'Heavenly Spring' Between the two heads of the biceps brachii muscle, 2 cun inferior to the axillary fold.

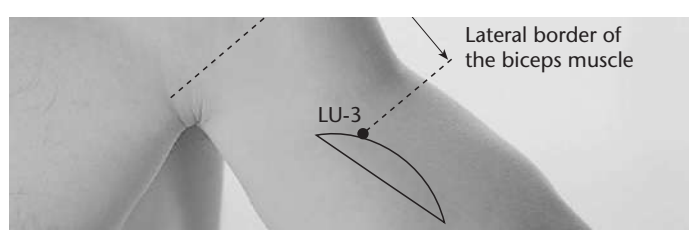

LU-3 (*tianfu*) 'Palace of Heaven' 3 cun distal to the end of the anterior axillary fold, on the medial aspect of the upper arm, radial to the biceps brachialis

Ch07-F10028.qxd 2/22/08 10:09 PM Page 641

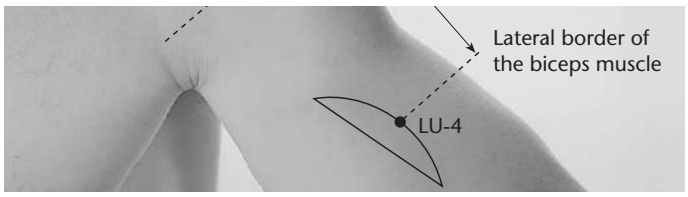

LU-4 (*xiabai*) 'Clasping the White' 4 cun distal to the end of the anterior axillary fold, on the medial aspect of the upper arm, radial to the biceps brachii

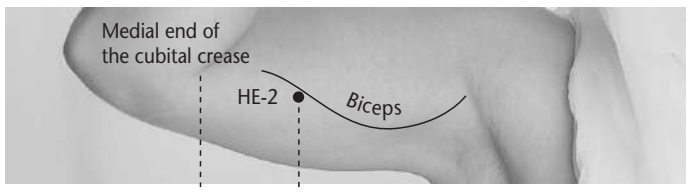

HE-2 (*qingling*) 'Green Spirit' 3 cun proximal to the cubital crease, on the medial border of the biceps brachii

LU-5 (*chize*) 'Cubit Marsh' In the cubital crease, radial to the tendon of the biceps brachii

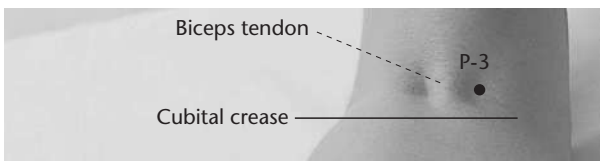

P-3 (*quze*) 'Marsh at the Crook' In the cubital crease, ulnar to the biceps tendon

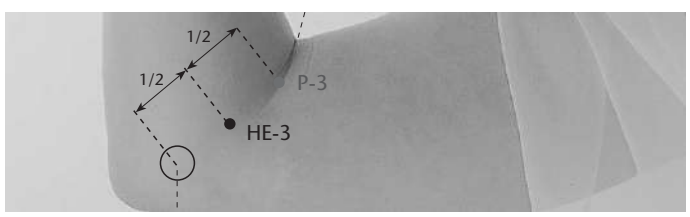

HE-3 (*shaohai*) 'Lesser Sea' With the elbow flexed, in the depression between the ulnar end of the cubital crease and the medial epicondyle of the humerus

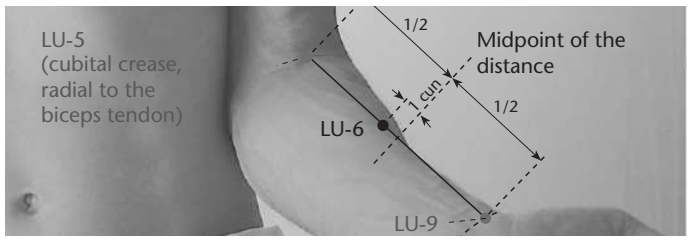

LU-6 (*kongzui*) 'Maximum Opening' 5 cun distal to **LU-5** or 7 cun proximal to the wrist joint space, on a line connecting **LU-5** and **LU-9**

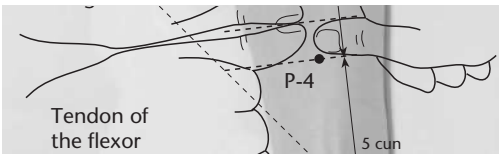

P-4 (*ximen*) 'Xi-Cleft Gate' 5 cun proximal to the palmar wrist joint space, between the tendons of palmaris longus and flexor carpi radialis, or 1 cun distal to the midpoint of the line connecting **P-3** and **P-7**

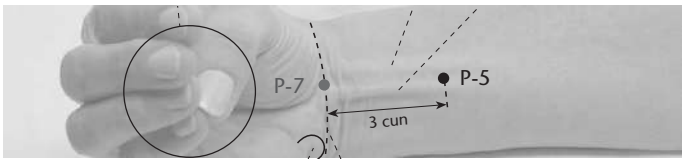

P-5 (*jianshi*) 'Intermediate Messenger' 3 cun proximal to the palmar wrist joint space, between the tendons of palmaris longus and flexor carpi radialis

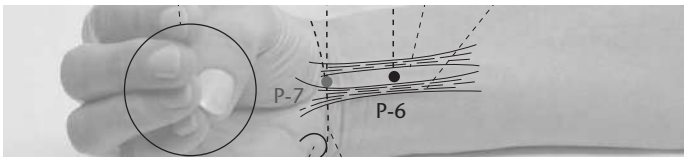

P-6 (*neiguan*) 'Inner Pass' 2 cun proximal to the palmar wrist joint space, between the tendons of palmaris longus and flexor carpi radialis

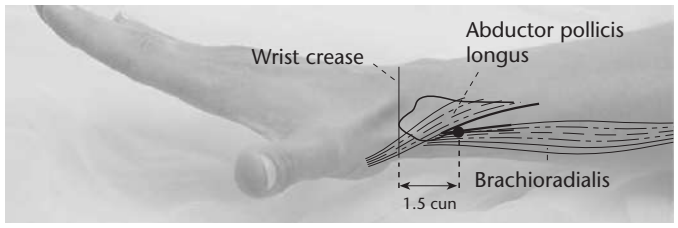

LU-7 (*lieque*) 'Broken Sequence' On the radial aspect of the forearm, immediately above the styloid process of the radius, approximately 1.5 cun proximal to the wrist joint space in a V-shaped groove

Ch07-F10028.qxd 2/22/08 10:09 PM Page 642

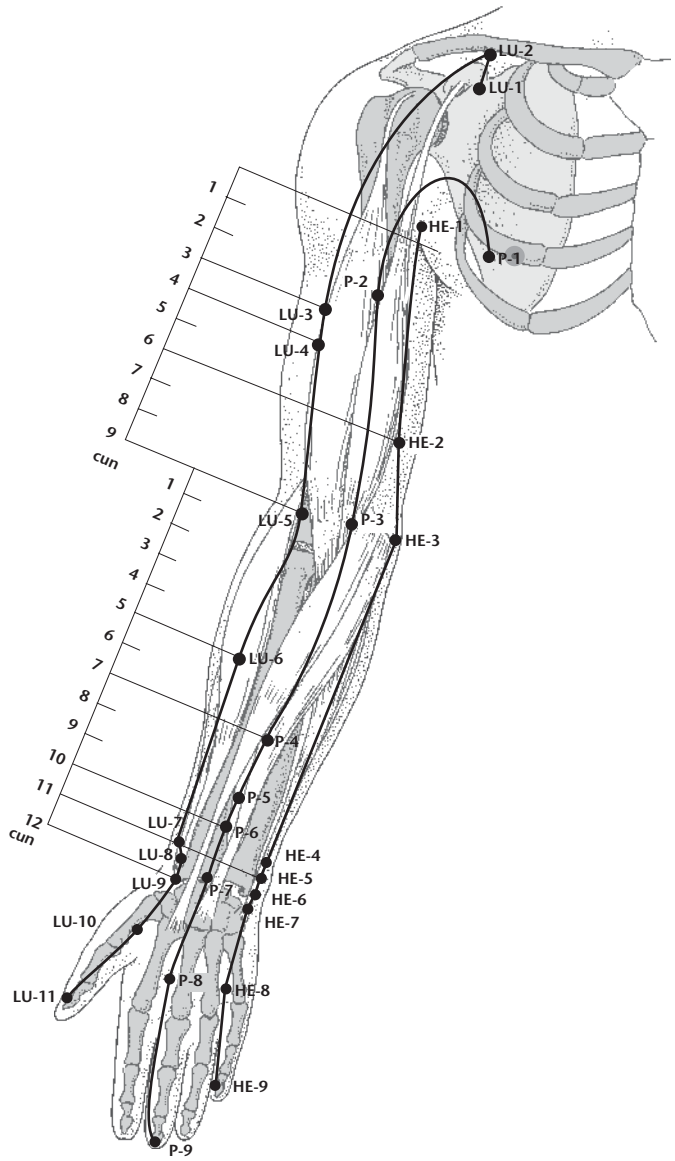

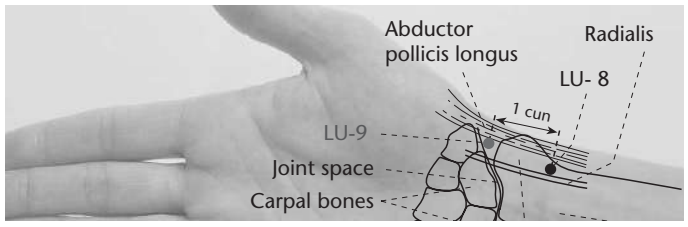

LU-8 (*jingqu*) 'Channel Gutter' 1 cun proximal to the palmar wrist joint space (**LU-9**), lateral to the radial artery

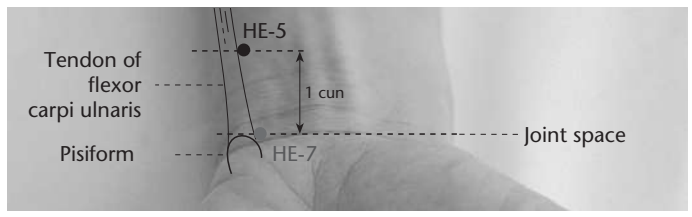

HE-5 (*tongli*) 'Penetrating the Interior' 1 cun proximal to the palmar wrist joint space, on the radial side of the flexor carpi ulnaris tendon

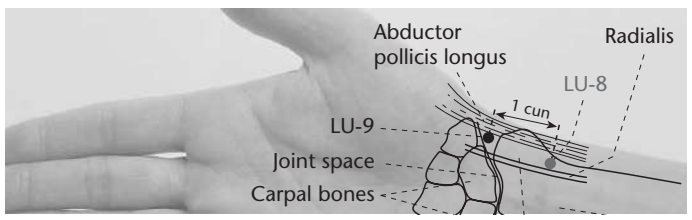

LU-9 (*taiyuan*) 'Supreme Abyss' On the wrist joint space, lateral to the radial artery and ulnar to the tendon of abductor pollicis longus

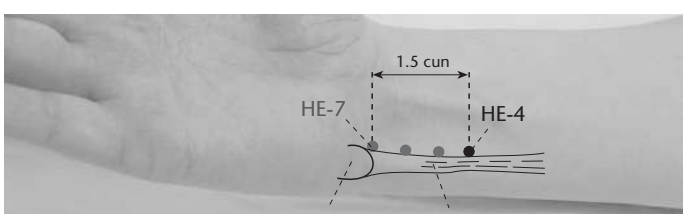

HE-4 (*lingdao*) 'Spirit Path' 1.5 cun proximal to the palmar wrist joint space, radial to the flexor carpi ulnaris tendon

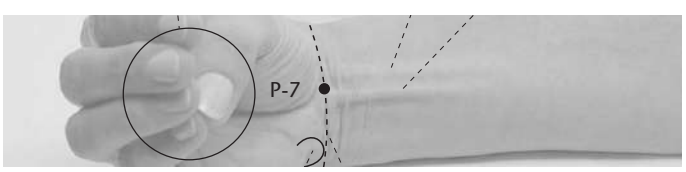

P-7 (*daling*) 'Great Mound' In the centre of the palmar aspect of the wrist joint space ('most distal wrist crease'), between the tendons of palmaris longus and flexor carpi radialis

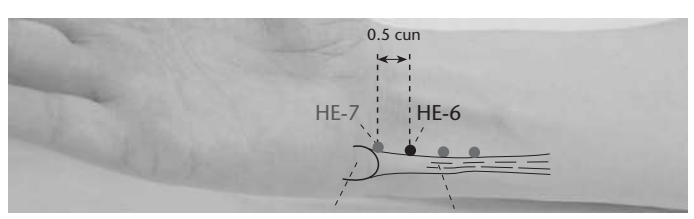

HE-6 (*yinxi*) 'Yin Cleft' 0.5 cun proximal to the palmar wrist joint space, radial to the flexor carpi ulnaris tendon

Ch07-F10028.qxd 2/22/08 10:09 PM Page 643

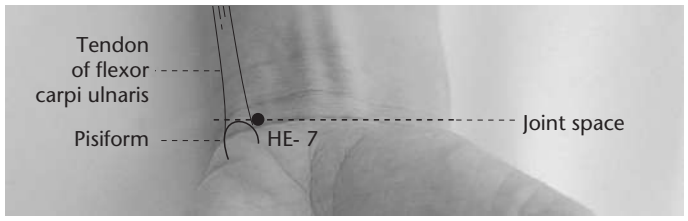

HE-7 (*shenmen*) 'Spirit Gate' On the palmar aspect of the wrist joint space ('most distal wrist crease'), in the depression radial to the insertion of the flexor carpi ulnaris tendon

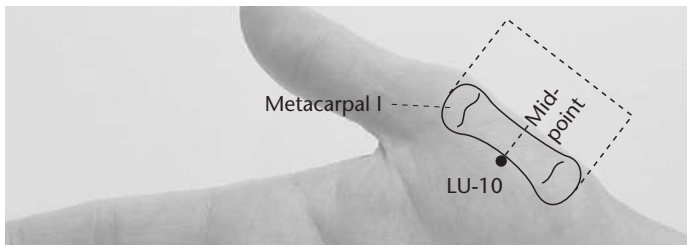

LU-10 (*yuji*) 'Fish Border' At the midpoint of the palmar border of the 1st metacarpal bone

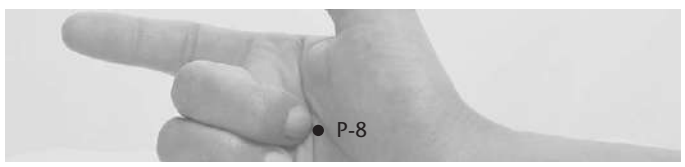

P-8 (*laogong*) 'Palace of Toil' In the centre of the palm, between the 2nd and 3rd metacarpal bones, slightly closer to the 3rd metacarpal bone. When a fist is made, the point is located beneath the tip of the middle finger

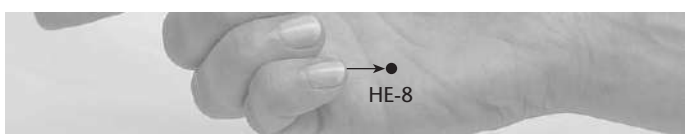

HE-8 (*shaofu*) 'Lesser Palace' When a fist is made, beneath the tip of the little finger, on the palm, between the 4th and 5th metacarpal bones

LU-11 (*shaoshang*) 'Lesser Shang' 0.1 cun from the radial corner of the thumb nail

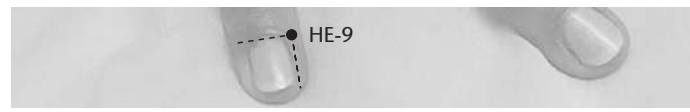

HE-9 (*shaochong*) 'Lesser Rushing' 0.1 cun lateral to the radial corner of the nail of the little finger

P-9 (*zhongchong*) 'Middle Rushing' On the most distal point of the middle finger

Ch07-F10028.qxd 2/22/08 10:09 PM Page 644

7.9 Dorsum of the Hand

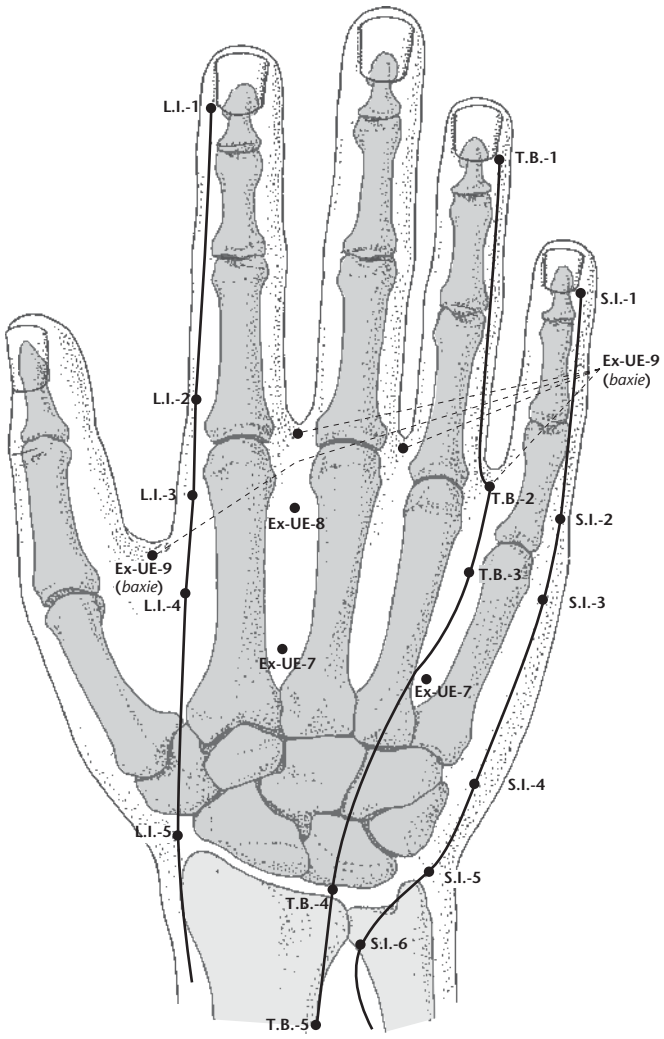

L.I.-1 (*shangyang*) 'Shang Yang' On the index finger, 0.1 cun from the radial corner of the nail

T.B.-1 (*guanchong*) 'Rushing Pass' On the ring finger (4th finger), 0.1 cun from the ulnar corner of the nail

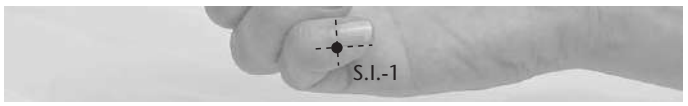

S.I.-1 (*shaoze*) 'Lesser Marsh' On the little finger, 0.1 cun from the ulnar corner of the nail

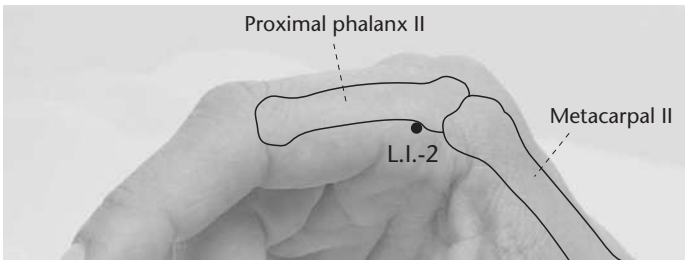

L.I.-2 (*erjian*) 'Second Space' On the radial aspect of the index finger, distal to the metacarpophalangeal joint at the junction of the shaft and the basis of the proximal phalanx

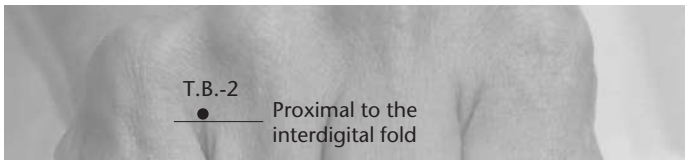

T.B.-2 (*yemen*) 'Fluid Gate' Between the little finger and ring finger, proximal to the margin of the web; locate with the patient making a fist

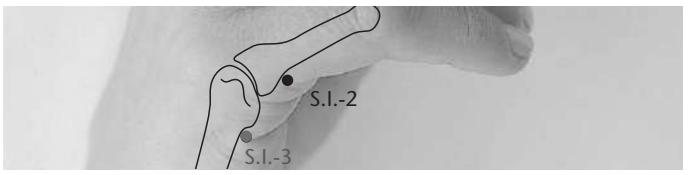

S.I.-2 (*qiangu*) 'Front Valley' On the ulnar aspect of the little finger, at the junction of the shaft and the base of the proximal phalanx

Ch07-F10028.qxd 2/22/08 10:09 PM Page 645

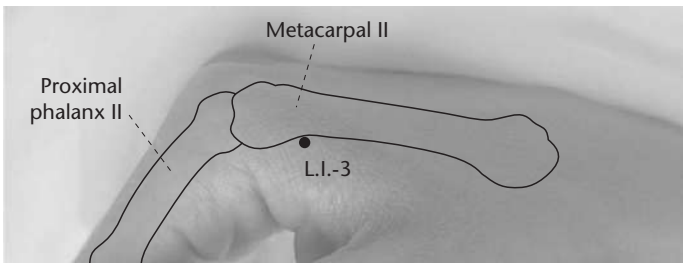

L.I.-3 (*sanjian*) 'Third Space' On the radial aspect of the index finger, proximal to the metacarpophalangeal joint at the junction of the shaft and the head of the 2nd metacarpal bone

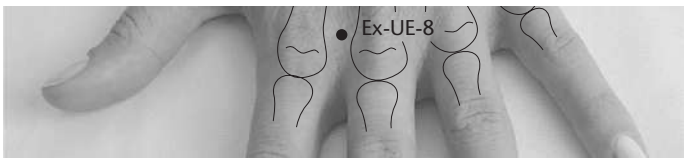

Ex-UE-8 (*wailaogong/luozhen/xianqiang*) 'Stiff Neck' On the dorsum of the hand, between the 2nd and 3rd metacarpal bones, approximately 0.5 cun proximal to the metacarpophalangeal joints

T.B.-3 (*zhongzhu*) 'Central Islet' On the dorsum of the hand, in the depression between the 4th and 5th metacarpal bones, proximal to the metacarpophalangeal joints. **T.B.-3** is located at the junction between the heads and shafts of the two metacarpals

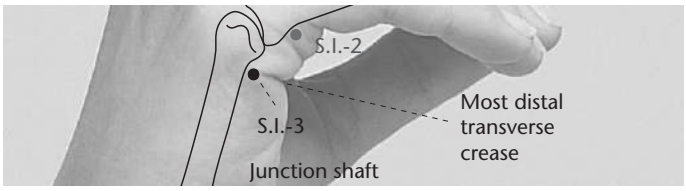

S.I.-3 (*houxi*) 'Back Stream' On the ulnar border of the hand, in the depression proximal to the metacarpophalangeal joint, at the junction between the head and shaft of the 5th metacarpal

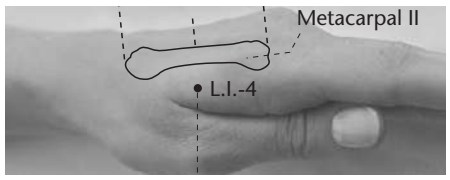

L.I.-4 (*hegu*) 'Joining Valley' On the radial aspect of the hand, between the 1st and 2nd metacarpal bone, closer to the 2nd metacarpal bone and approximately at its midpoint

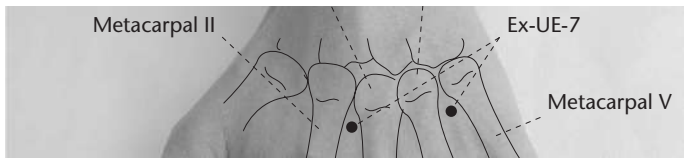

Ex-UE-7 (*yaotongdian/yaotongxue*) 'Lumbar Pair Point' Two points on the dorsum of the hand, between the 2nd/3rd and 4th/5th metacarpal bones

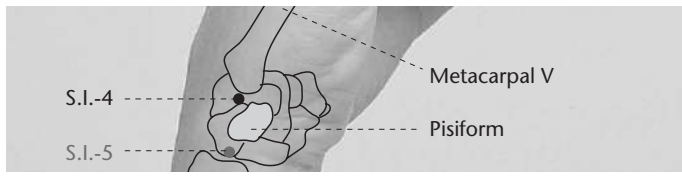

S.I.-4 (*wangu*) 'Wrist Bone' On the ulnar border of the hand, between the 5th metacarpal and the carpal bones, at the border of the red and white skin

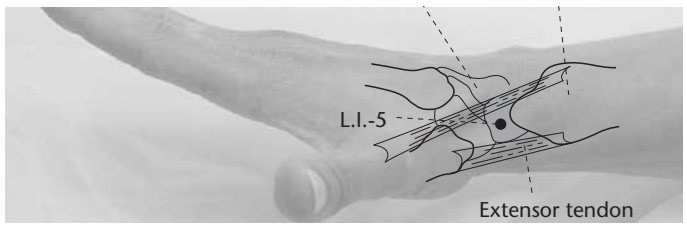

L.I.-5 (*yangxi*) 'Yang Stream' With the thumb abducted in the depression between the tendons of the extensor pollicis longus and brevis muscles ('anatomical snuff box')

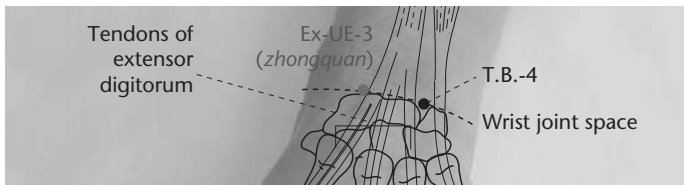

T.B.-4 (*yangchi*) 'Yang Pool' On the dorsum of the wrist, in the gap ulnar to the tendons of extensor digitorum

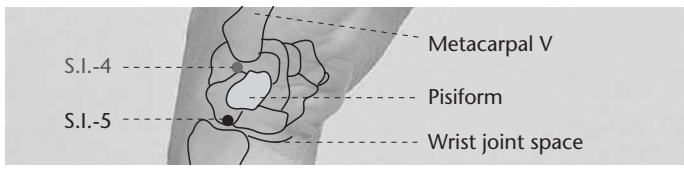

S.I.-5 (*yanggu*) 'Yang Valley' On the ulnar aspect of the wrist, at the level of the lateral joint space

Ch07-F10028.qxd 2/22/08 10:09 PM Page 646

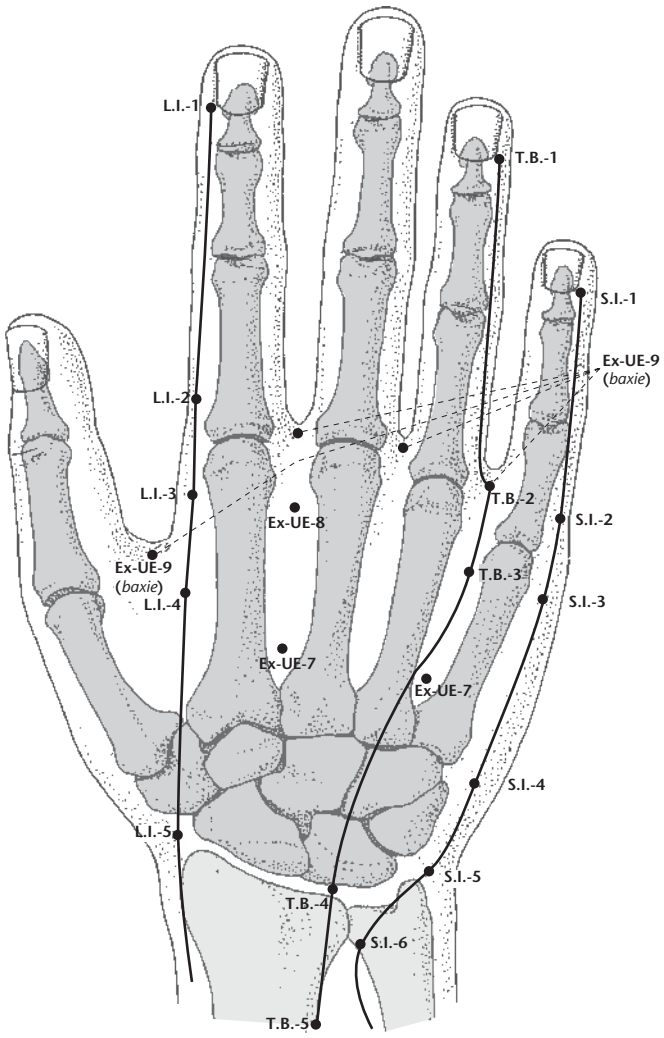

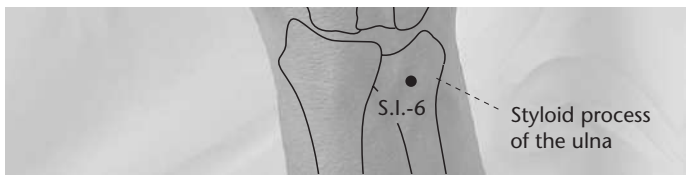

S.I.-6 (*yanglao*) 'Nourishing the Old' In the depression radial and proximal to the styloid process of the ulna that forms when the hand is turned from a pronated to a supinated position

T.B.-5 (*waiguan*) 'Outer Pass' 2 cun proximal to **T.B.-4** (on the dorsal wrist joint space, between radius and ulna)

7.10 Anterior Leg

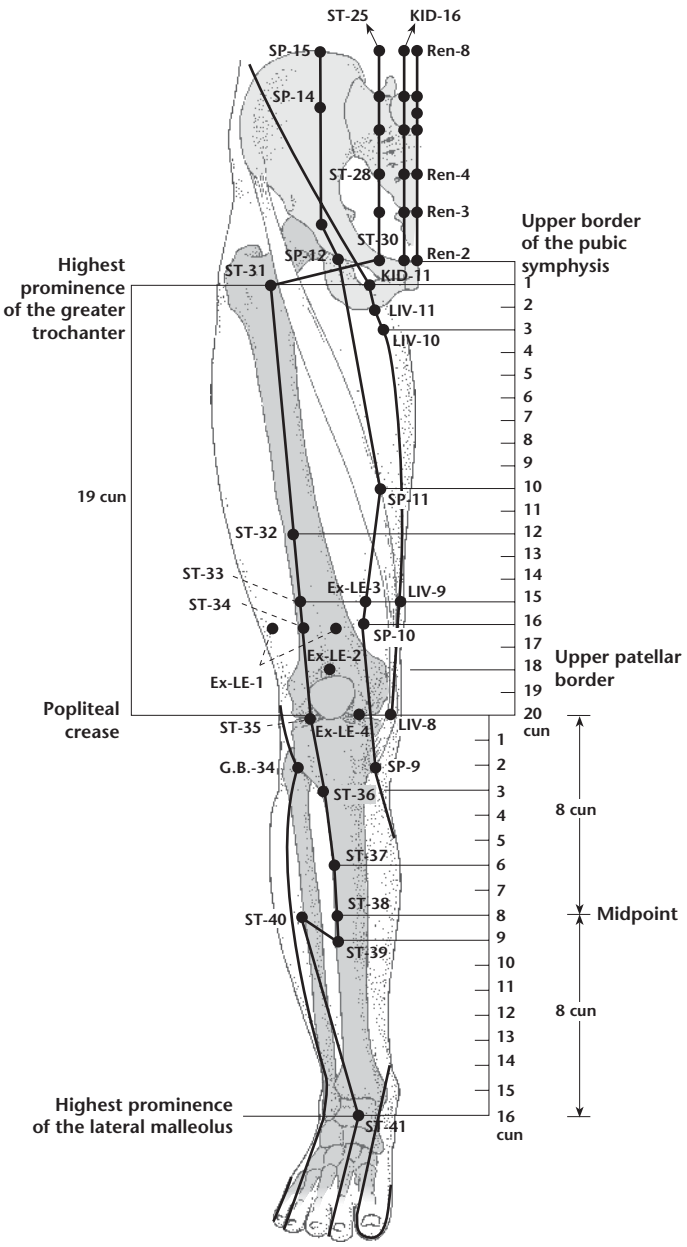

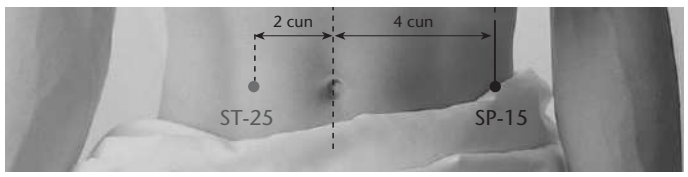

SP-15 (*daheng*) 'Great Horizontal' 4 cun lateral to the centre of the umbilicus, on the mamillary line

Ch07-F10028.qxd 2/22/08 10:09 PM Page 647

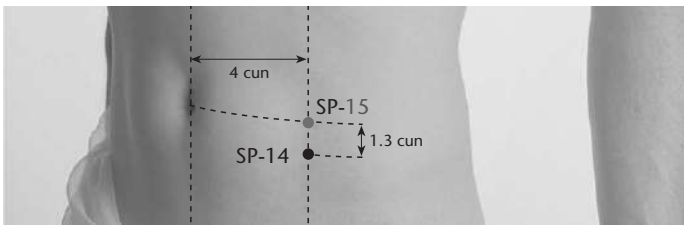

SP-14 (*fujie*) 'Abdomen Knot' 4 cun lateral to the anterior midline (mamillary line), 3 cun superior to **SP-13** or 1.3 cun inferior to **SP-15**

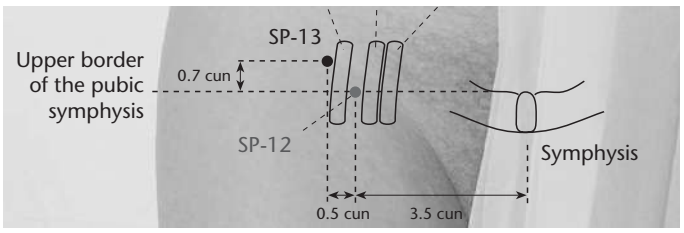

SP-13 (*fushe*) 'Meeting of the Fu Organs' 4 cun lateral to the anterior midline (mamillary line) and 0.7 cun superior to the upper border of the pubic symphysis

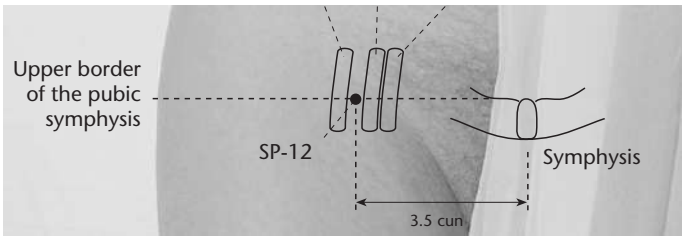

SP-12 (*chongmen*) 'Rushing Gate' 3.5 cun lateral to the midline, level with the upper border of the pubic symphysis, lateral to the femoral artery

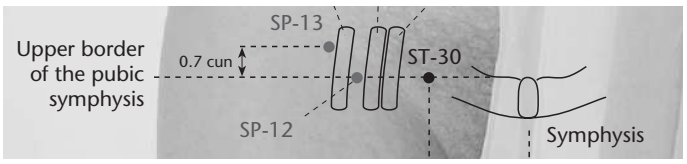

ST-30 (*qichong*) 'Rushing Qi' 2 cun lateral to the upper border of the pubic symphysis, medial to the femoral artery; at the level of **Ren-2**

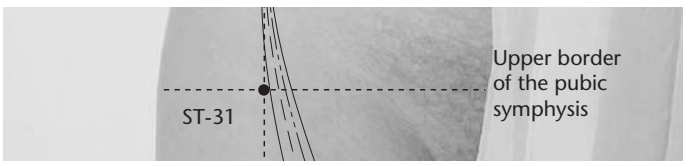

ST-31 (*biguan*) 'Thigh Gate' With the hip joint flexed, in the depression below the anterior superior iliac spine, lateral to the sartorius

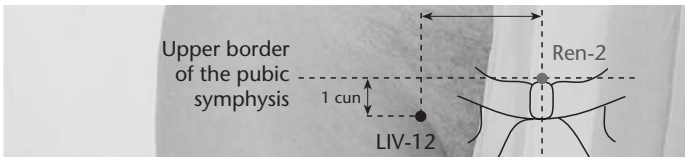

LIV-12 (*jimai*) 'Urgent Pulse' In the inguinal groove, 2.5 cun lateral to the midline and 1 cun inferior to the upper border of the symphysis, inferior to **ST-30** and medial to the femoral artery

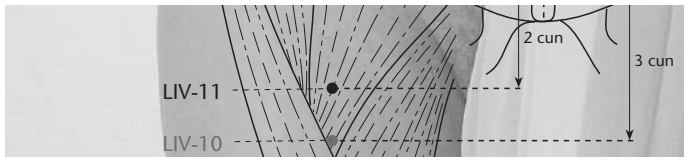

LIV-11 (*yinlian*) 'Yin Corner' 2 cun inferior to **ST-30** (level with the upper border of the pubic symphysis), on the lateral border of the adductor longus muscle or approximately 1 cun inferior to where the femoral artery passes the inguinal ligament

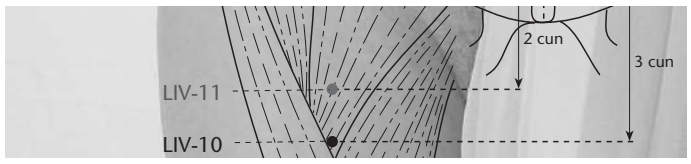

LIV-10 (*zuwuli*) 'Leg Five Miles' On the anterior aspect of the thigh, 3 cun inferior to the upper border of the symphysis, on the lateral border of adductor longus

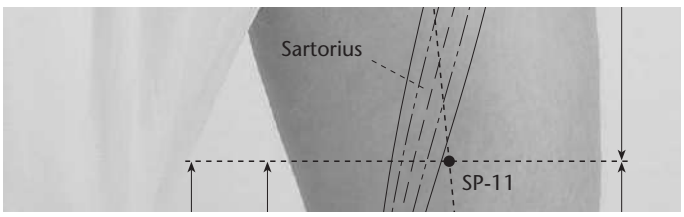

SP-11 (*jimen*) 'Winnowing Gate' 6 cun proximal to **SP-10** at the midpoint of the femur, in the depression between the sartorius and vastus lateralis

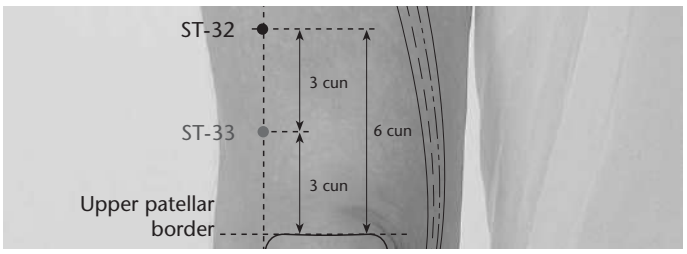

ST-32 (*futu*) 'Crouching Rabbit' 6 cun proximal to the lateral upper border of the patella, on a line connecting with the anterior superior iliac spine

Ch07-F10028.qxd 2/22/08 10:09 PM Page 648

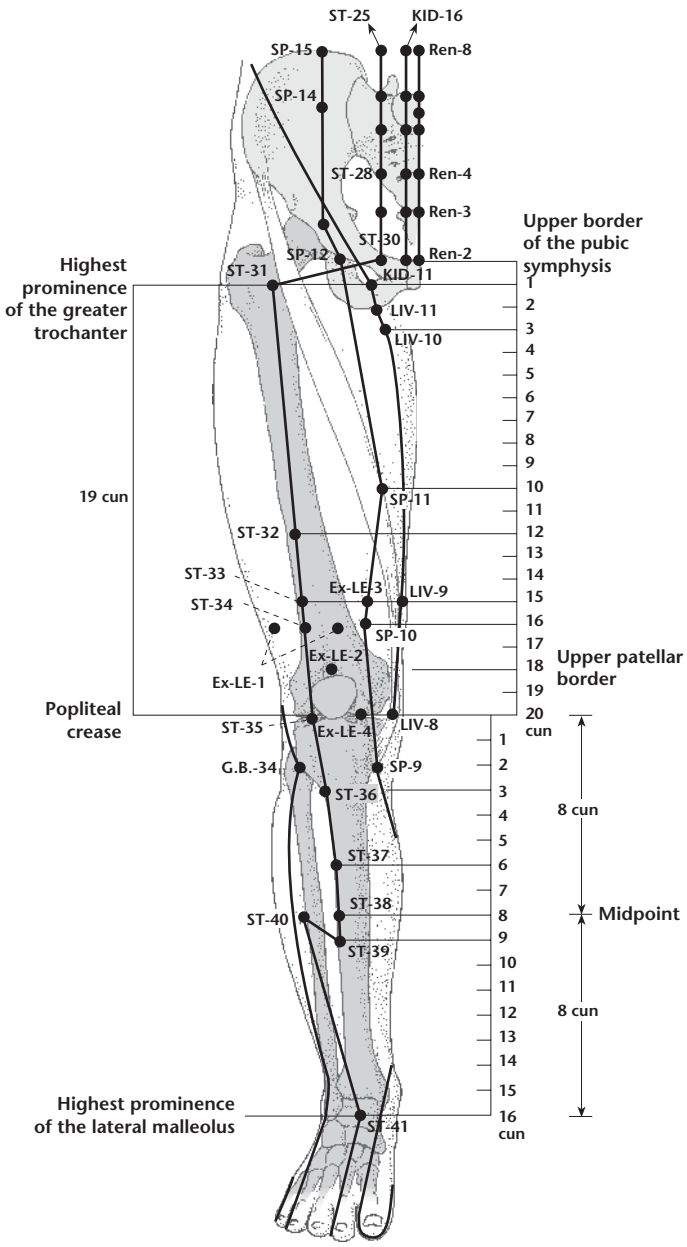

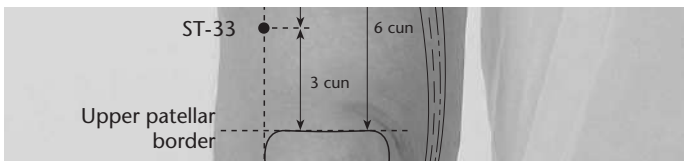

ST-33 (*yinshi*) 'Yin Market' 3 cun superior to the upper lateral border of the patella

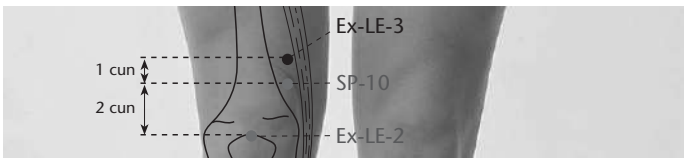

Ex-LE-3 (*baichongwo*) 'Hundred Insect Burrow' 3 cun superior and 1 cun medial to the upper medial border of the patella, in a small depression on the vastus medialis muscle or 1 cun superior to **SP-10**

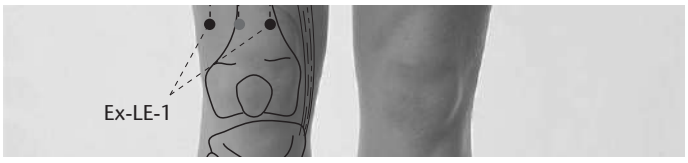

Ex-LE-1 (*kuangu*) 'Hip Bone' A pair of points 2 cun superior to the patella and 1.5 cun lateral and medial to **ST-34**

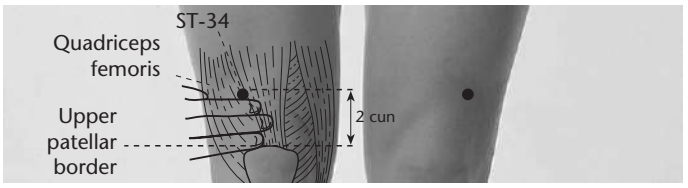

ST-34 (*liangqiu*) 'Ridge Mound' 2 cun proximal to the upper lateral border of the patella, in a depression on vastus lateralis

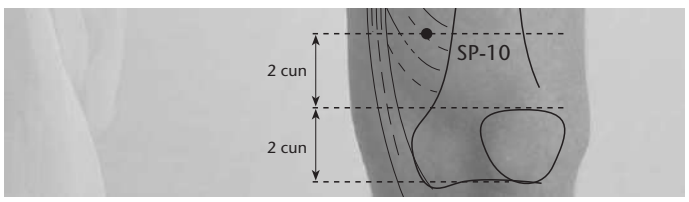

SP-10 (*xuehai*) 'Sea of Blood' With the knee flexed, 2 cun proximal to the medial superior border of the patella. To find: place the right palm on the patient's left patella and locate the point at the tip of the thumb

Ex-LE-2 (*heding/xiding*) 'Crane's Summit' With the knee flexed, at the midpoint of the upper patellar border

Ch07-F10028.qxd 2/22/08 10:09 PM Page 649

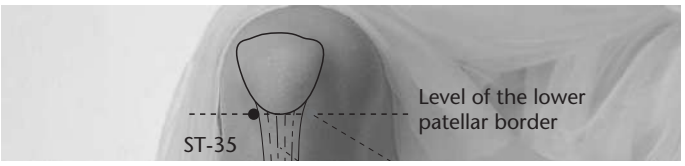

ST-35 (*dubi*) 'Calf's Nose' With the knee flexed, in the depression inferior to the patella and lateral to the patellar ligament; corresponds to the lateral eye of the knee; **Ex-LE-5** (*xiyan*) is the combination of this point and **Ex-LE-4**

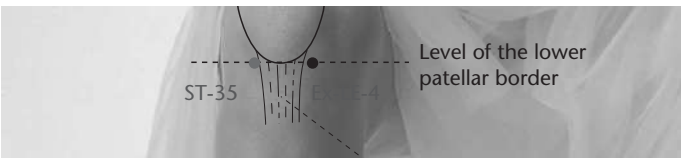

Ex-LE-4 (*neixiyan*) 'Inner Eye of the Knee' Inferior to the patella, in a depression medial to the patellar ligament; **Ex-LE-5** (*xiyan*) is the combination of this point and **ST-35**

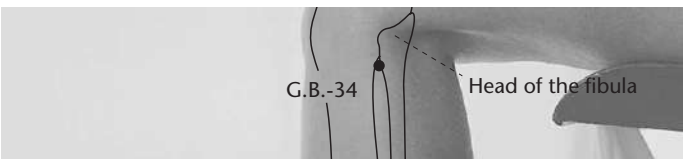

G.B.-34 (*yanglingquan*) 'Yang Mound Spring' In the depression anterior and inferior to the head of the fibula

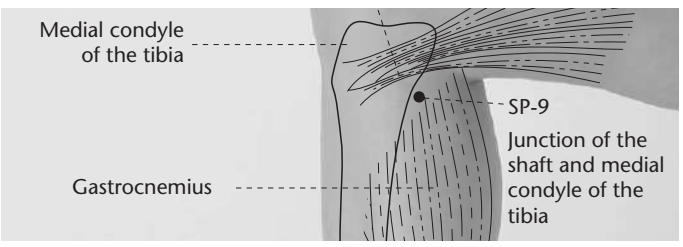

SP-9 (*yinlingquan*) 'Yin Mound Spring' With the knee flexed in a depression distal to the medial condyle of the tibia. To find: palpate from the medial posterior border of the tibia towards proximal, at roughly the same level as **G.B.-34**

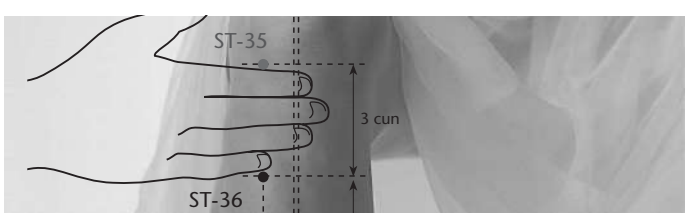

ST-36 (*zusanli*) 'Leg Three Miles' one fingerbreadth (middle finger) lateral to the anterior crest of the tibia, 3 cun distal to **ST-35**

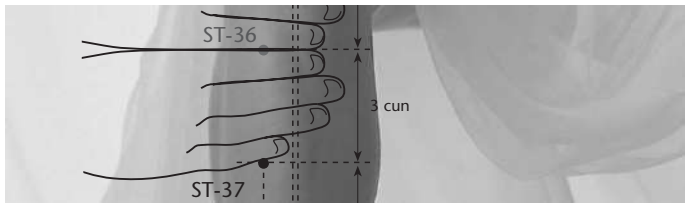

ST-37 (*shangjuxu*) 'Upper Great Void' 1 fingerbreadth lateral to the anterior crest of the tibia, 6 cun distal to **ST-35** or 3 cun distal to **ST-36**

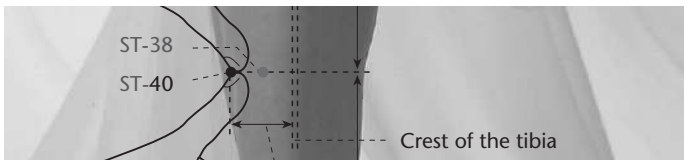

ST-40 (*fenglong*) 'Abundant Bulge' 2 fingerbreadths lateral to the anterior crest of the tibia, at the midpoint of the line joining **ST-35** ('lateral eye of the knee', level with the knee joint space) and the highest prominence of the lateral malleolus

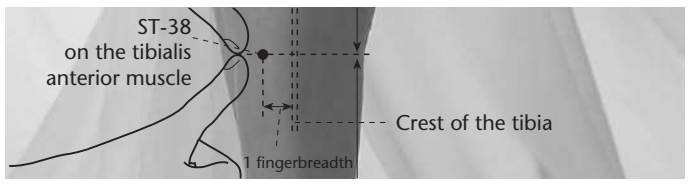

ST-38 (*tiaokou*) 'Lines Opening' 1 fingerbreadth lateral to the anterior crest of the tibia, at the midpoint of the line joining **ST-35** ('lateral eye of the knee', level with the knee joint space) and the highest prominence of the lateral malleolus

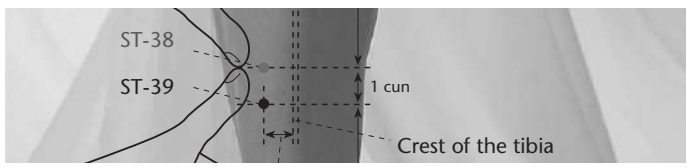

ST-39 (*xiajuxu*) 'Lower Great Void' 1 cun distal to **ST-38**

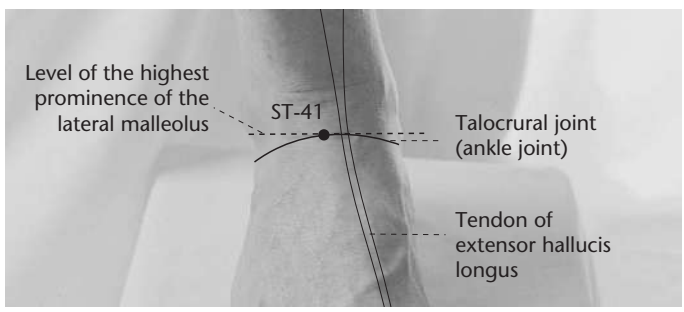

ST-41 (*jiexi*) 'Stream Divide' On the ankle, between the tendons of the extensor digitorum and the extensor hallucis longus, at the level of the highest prominence of the lateral malleolus

Ch07-F10028.qxd 2/22/08 10:09 PM Page 650

7.11 Medial Leg

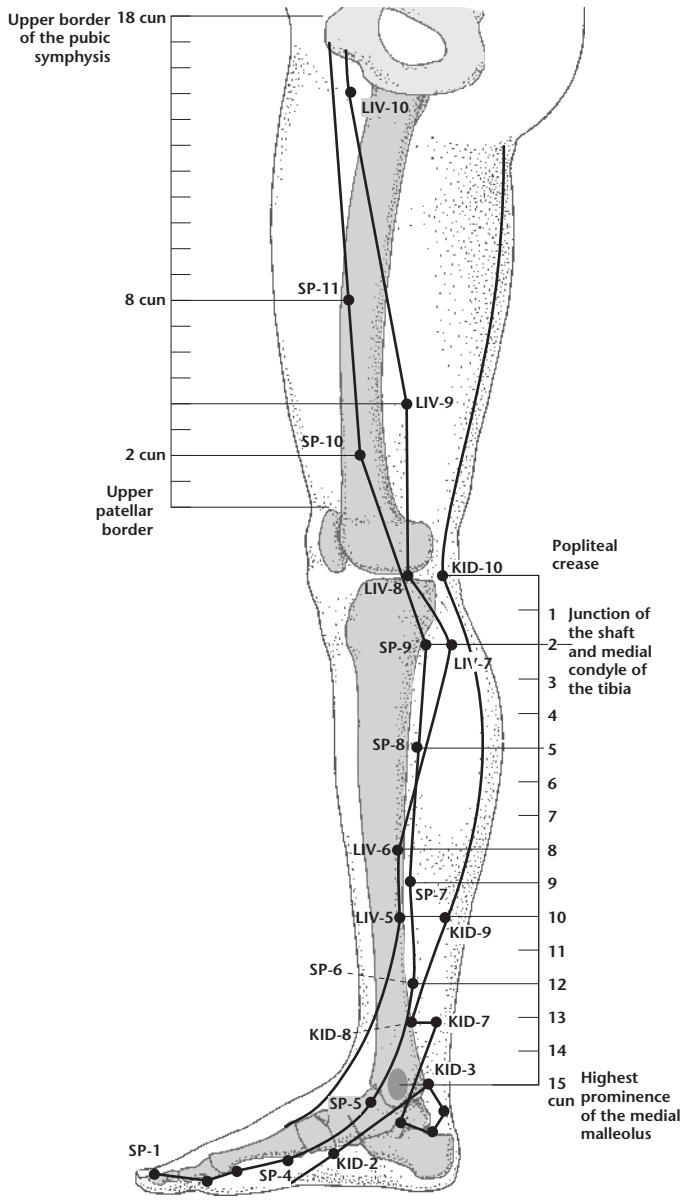

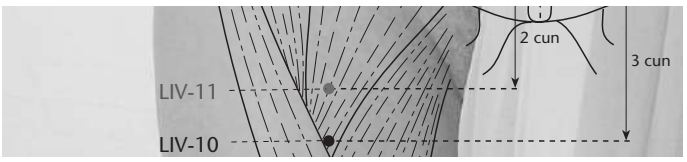

LIV-10 (*zuwuli*) 'Leg Five Miles' On the anterior aspect of the thigh, 3 cun inferior to the upper border of the symphysis, on the lateral border of the adductor longus

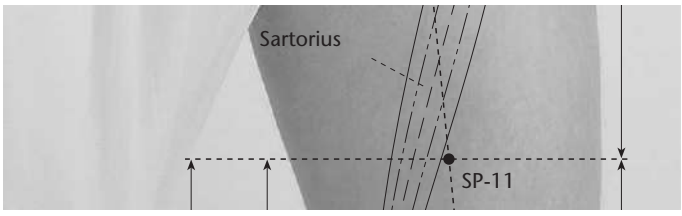

SP-11 (*jimen*) 'Winnowing Gate' 6 cun proximal to **SP-10** at the midpoint of the femur, in the depression between the sartorius and vastus lateralis

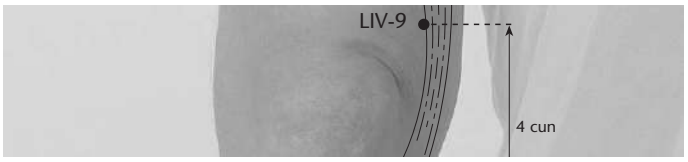

LIV-9 (*yinbao*) 'Yin Wrapping' 4 cun proximal to the medial condyle of the femur, between the sartorius and vastus medialis

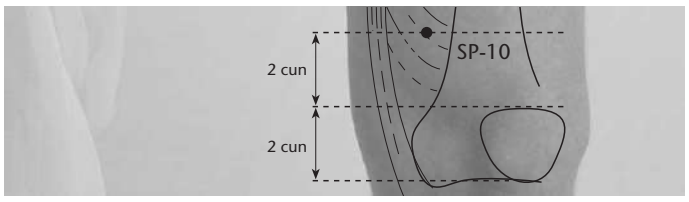

SP-10 (*xuehai*) 'Sea of Blood' With the knee flexed, 2 cun proximal to the medial superior border of the patella. To find: place the right palm on the patient's left patella and locate the point at the tip of the thumb

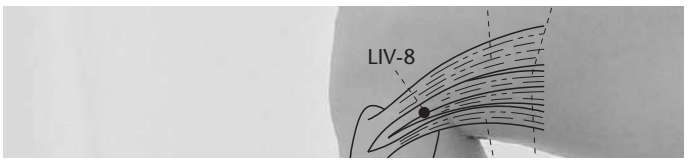

LIV-8 (*ruquan*) 'Spring at the Crook' With the knee flexed, at the medial end of the popliteal crease, in the depression anterior to the tendons of semitendinosus and semimembranosus

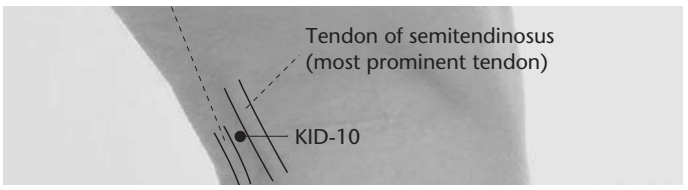

KID-10 (*yingu*) 'Yin Valley' At the medial end of the popliteal crease, between the tendons of semimembranosus and semitendinosus, at the level of the knee joint space

Ch07-F10028.qxd 2/22/08 10:09 PM Page 651

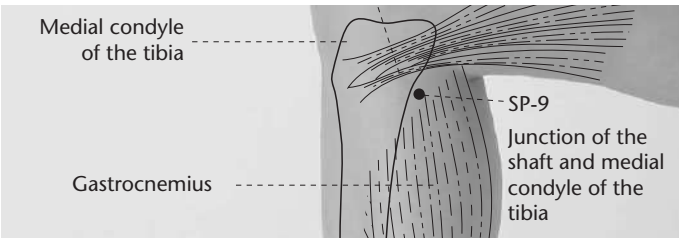

SP-9 (*yinlingquan*) 'Yin Mound Spring' With the knee flexed, in a depression distal to the medial condyle of the tibia. To find: palpate from the medial posterior border of the tibia towards proximal, at roughly the same level as **G.B.-34**

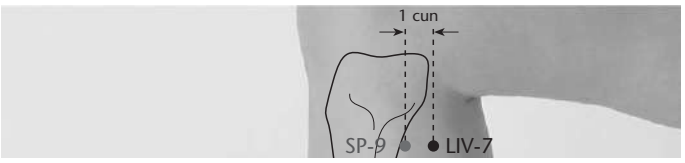

LIV-7 (*xiguan*) 'Knee Joint' At the junction of the shaft and the medial condyle of the tibia, 1 cun posterior to **SP-9**

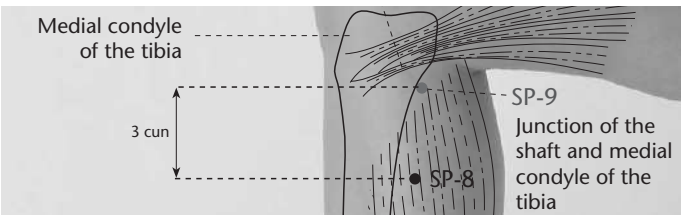

SP-8 (*diji*) 'Earth Pivot' 3 cun distal to **SP-9** (at the junction of the shaft and the medial condyle of the tibia) on the posterior border of the tibia

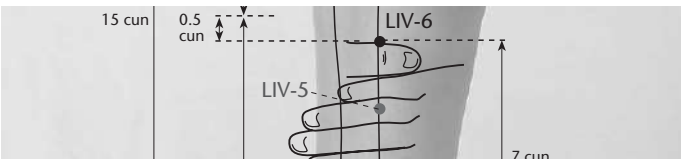

LIV-6 (*zhongdu*) 'Central Capital' 7 cun proximal to the highest prominence of the medial malleolus, just posterior to the medial crest of the tibia

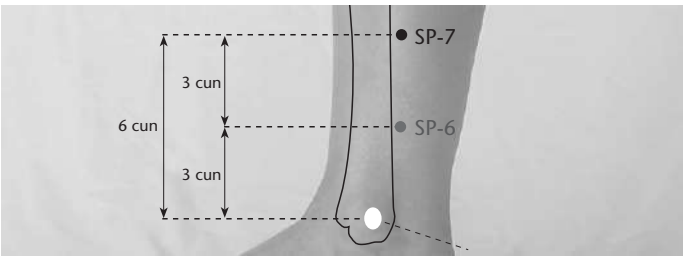

SP-7 (*lougu*) 'Dripping Valley' 6 cun proximal to the highest prominence of the medial malleolus or 3 cun proximal to **SP-6**, on the medial posterior border of the tibia

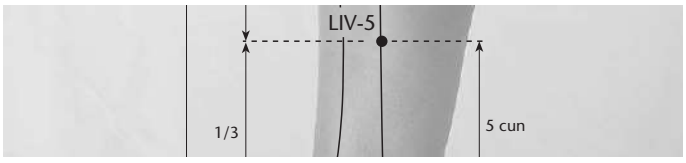

LIV-5 (*ligou*) 'Woodworm Canal' 5 cun proximal to the highest prominence of the medial malleolus, just posterior to the medial crest of the tibia

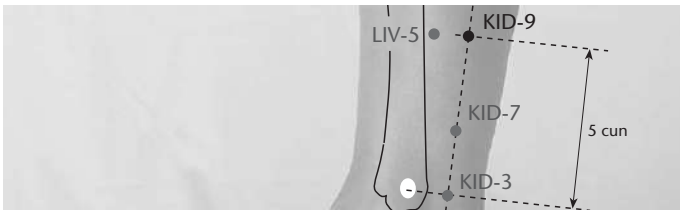

KID-9 (*zhubin*) 'Guest House' 5 cun proximal to KID-3 and 2 cun posterior to the medial border of the tibia

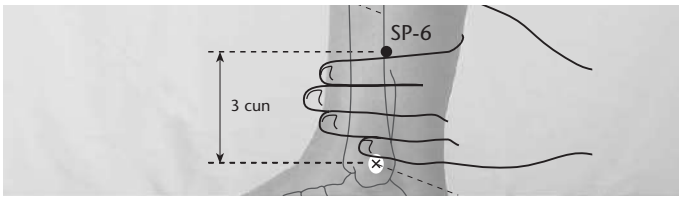

SP-6 (*sanyinjiao*) 'Three Yin Intersection' 3 cun proximal to the highest prominence of the medial malleolus, posterior border to the medial border of the tibia

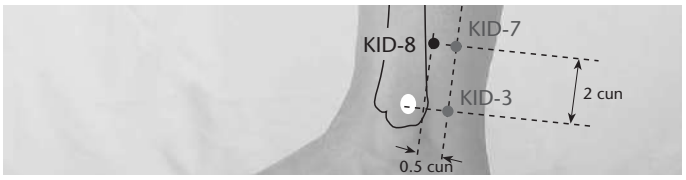

KID-8 (*jiaoxin*) 'Exchange Belief' 2 cun proximal to the highest prominence of the medial malleolus, posterior to the medial border of the tibia

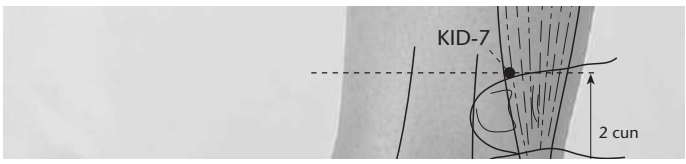

KID-7 (*fuliu*) 'Returning Current' 2 cun proximal to KID-3, in a depression on the anterior border of the Achilles tendon

Ch07-F10028.qxd 2/22/08 10:09 PM Page 652

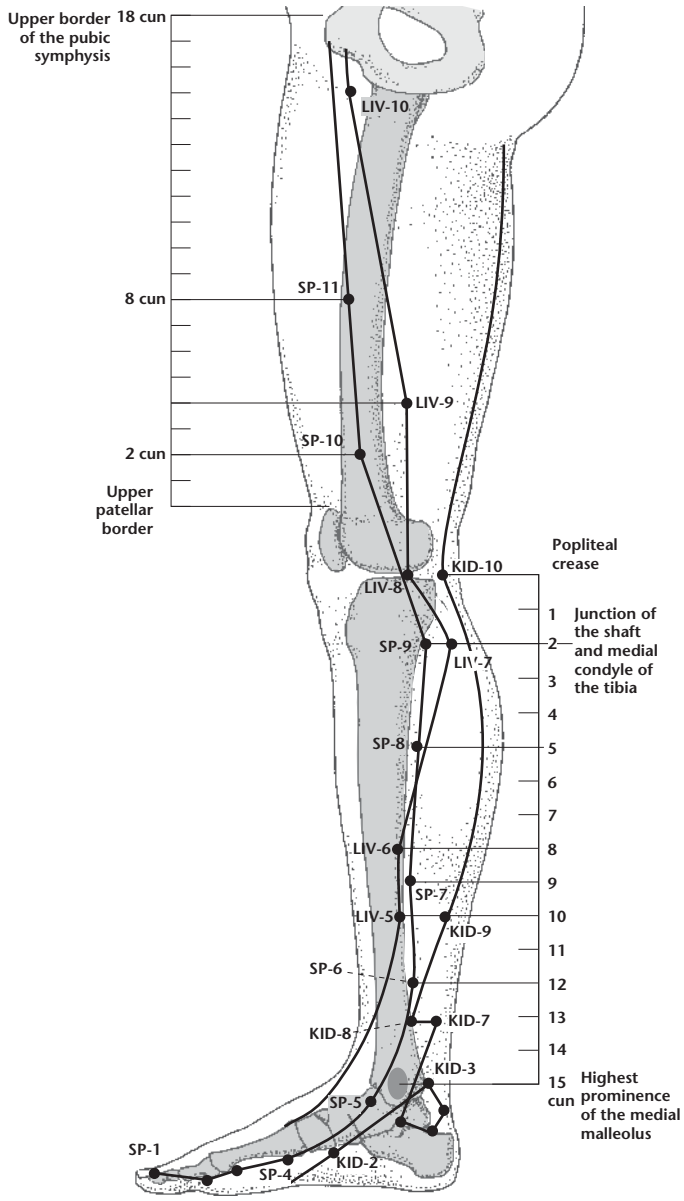

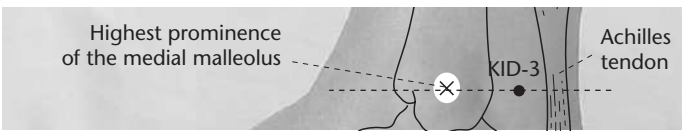

KID-3 (*taixi*) 'Supreme Stream' In the depression between the highest prominence of the medial malleolus and the Achilles tendon

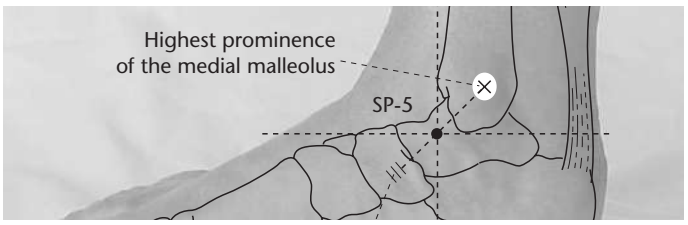

SP-5 (*shangqiu*) 'Shang Mound' In the depression at the junction of a vertical line along the anterior border and a horizontal line along the lower border of the medial malleolus. Or: in the depression halfway between the highest prominence of the medial malleolus and the navicular tuberosity

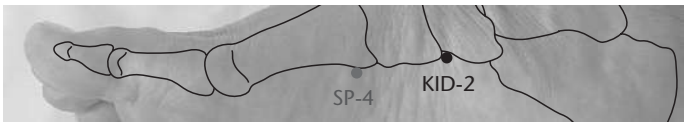

KID-2 (*rangu*) 'Blazing Valley' In a depression at the anterior inferior border of the navicular bone, at the border of the red and white skin

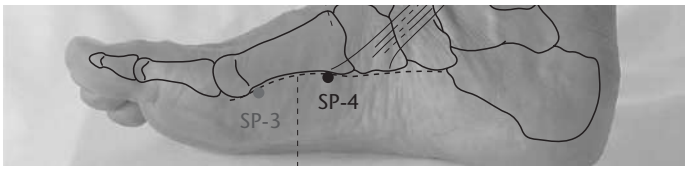

SP-4 (*gongsun*) 'Grandfather Grandson' In a depression distal and inferior to the base of the 1st metatarsal bone, at the border of the red and white skin

SP-1 (*yinbai*) 'Hidden White' On the big toe, 0.1 cun from the medial corner of the nail

Ch07-F10028.qxd 2/22/08 10:09 PM Page 653

7.12 Lateral Leg

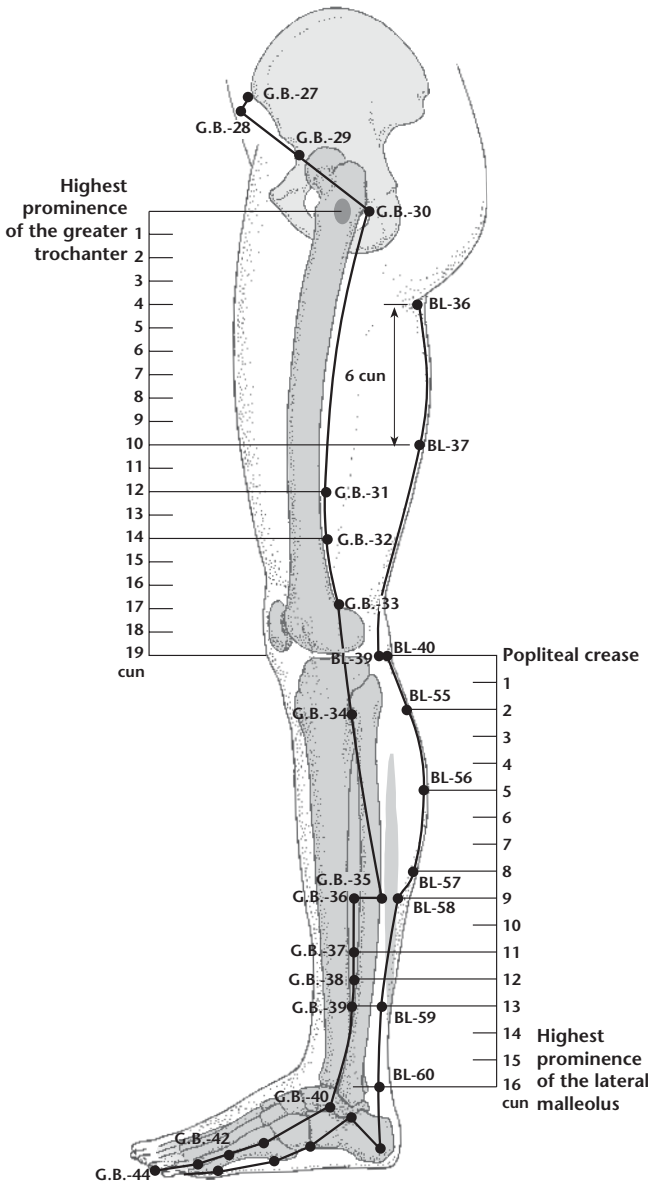

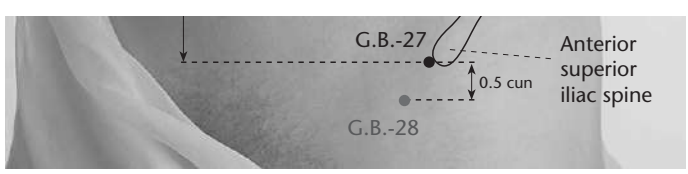

G.B.-27 (*wushu*) 'Five Pivots' In the depression medial to the anterior superior iliac spine, approximately level with **Ren-4**

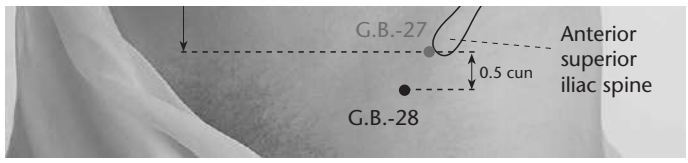

G.B.-28 (*weidao*) 'Linking Path' Anterior and inferior to the anterior superior iliac spine, 0.5 cun anterior and inferior to **G.B.-27**

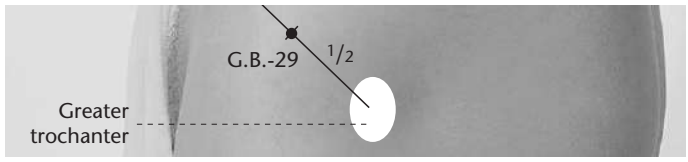

G.B.-29 (*juliao*) 'Stationary Crevice' At the midpoint of a line connecting the anterior superior iliac spine and the greater trochanter, at the anterior border of the iliac crest

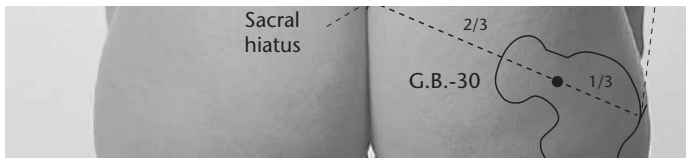

G.B.-30 (*huantiao*) 'Jumping Circle' With the patient lying on his/her side, at the junction between the medial two-thirds and the lateral third on a line connecting the greater trochanter and the sacral hiatus

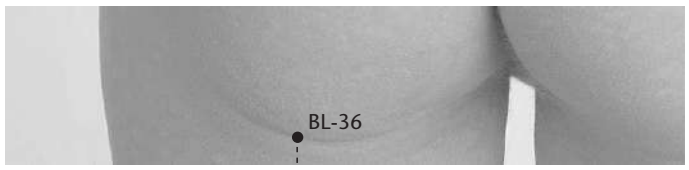

BL-36 (*chengfu*) 'Hold and Support' In the gluteal crease, superior to the midpoint of the popliteal crease

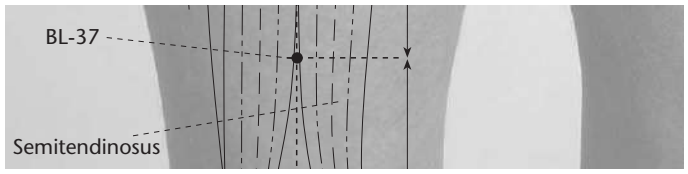

BL-37 (*yinmen*) 'Gate of Abundance' 6 cun distal to **BL-36** (gluteal crease), on a line connecting BL-36 and BL-40 (centre of the popliteal crease), in a gap in the musculature

Ch07-F10028.qxd 2/22/08 10:09 PM Page 654

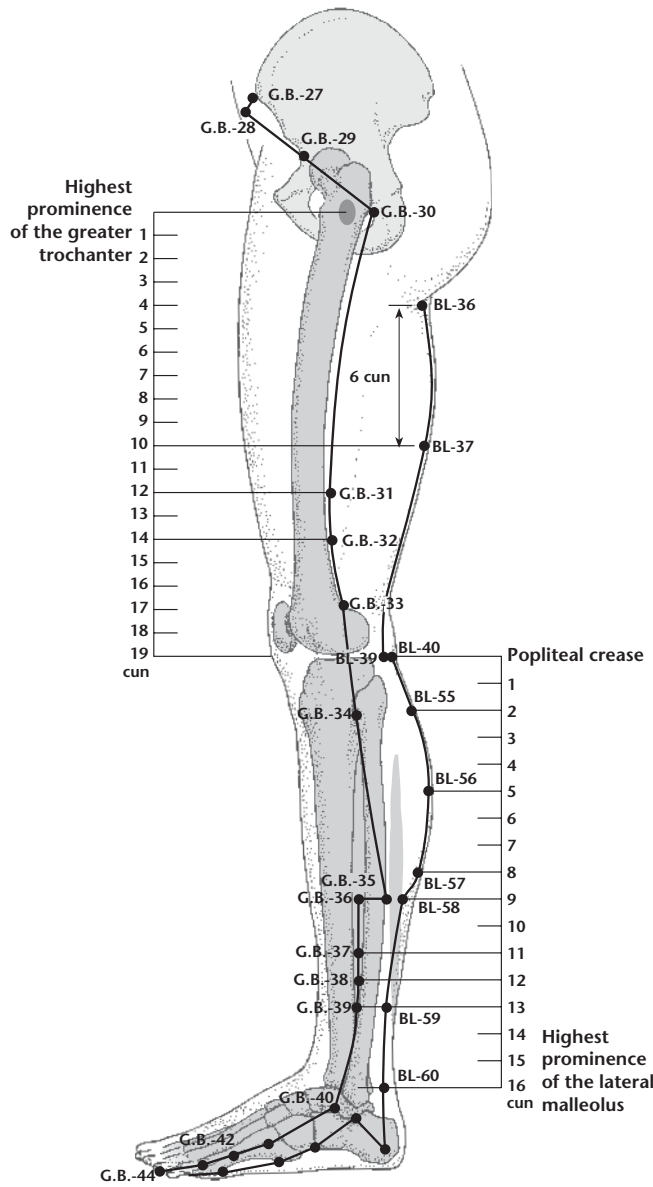

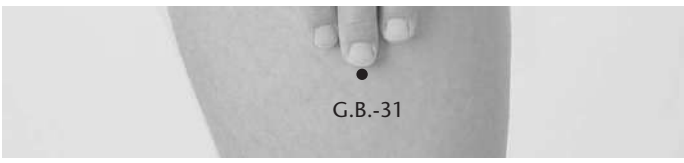

G.B.-31 (*fengshi*) 'Wind Market' On the lateral aspect of the thigh, approximately 7 cun proximal to the popliteal crease. With the patient standing upright, the point is where the patient's middle fingertip touches the thigh

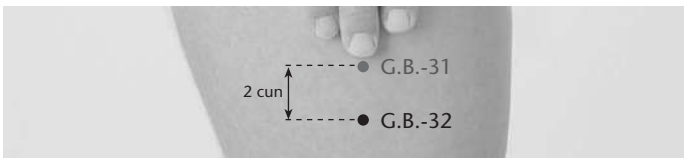

G.B.-32 (*zhongdu*) 'Middle Ditch' On the lateral aspect of the thigh, 2 cun inferior to **G.B.-31** or 5 cun superior to the popliteal crease, between vastus lateralis and biceps femoris

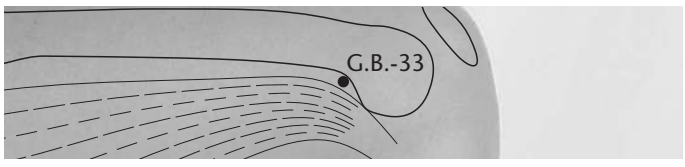

G.B.-33 (*xiyangguan*) 'Knee Yang Gate' On the lateral aspect of the knee, with the knee flexed, in the depression between the lateral epicondyle of the femur and the tendon of biceps femoris, approximately 3 cun proximal to **G.B.-34**

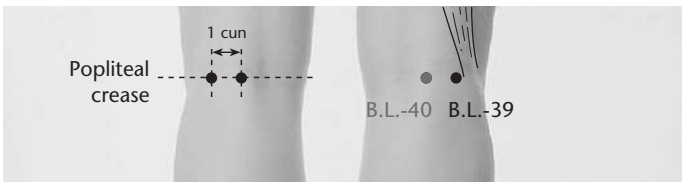

BL-39 (*weiyang*) 'Outside of the Crook' At the lateral end of the popliteal crease, on the medial side of the tendon of the long head of the biceps femoris, 1 cun lateral to **BL-40** (in the centre of the popliteal crease).

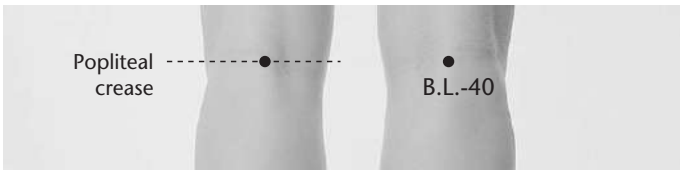

BL-40 (*weizhong*) 'Middle of the Crook' In the centre of the popliteal crease, between the tendons of the biceps femoris and semitendinosus.

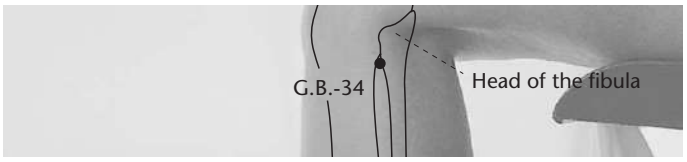

G.B.-34 (*yanglingquan*) 'Yang Mound Spring' In the depression anterior and inferior to the head of the fibula

Ch07-F10028.qxd 2/22/08 10:09 PM Page 655

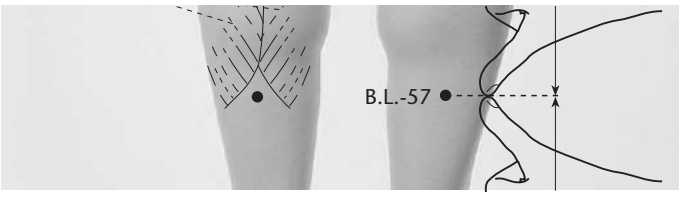

BL-57 (*chengshan*) 'Support of the Mountain' In the middle of the calf, between the two heads of the gastrocnemius, on a line connecting **BL-40** and **BL-60**, approximately 8 cun distal to **BL-40** (spreading hands technique → 2)

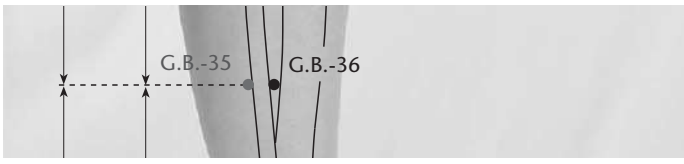

G.B.-36 (*waiqiu*) 'Outer Hill' 7 cun proximal to the highest prominence of the lateral malleolus, on the anterior border of the fibula

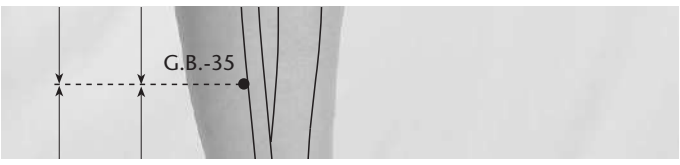

G.B.-35 (*yangjiao*) 'Yang Intersection' 7 cun proximal to the highest prominence of the lateral malleolus, on the posterior border of the fibula

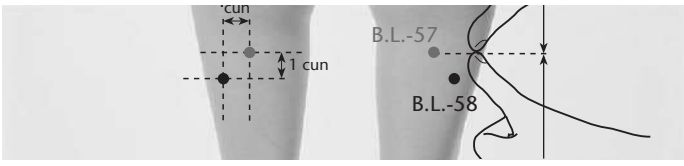

BL-58 (*feiyang*) 'Soaring Upwards' 1 cun distal and 1 cun lateral to **BL-57** or 7 cun proximal to **BL-60**, at the posterior border of the fibula and on the lower border of the gastrocnemius

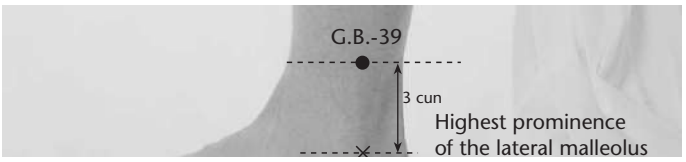

G.B.-39 (*xuanzhong*) 'Suspended Bell' 3 cun proximal to the highest prominence of the lateral malleolus, on the anterior border of the fibula

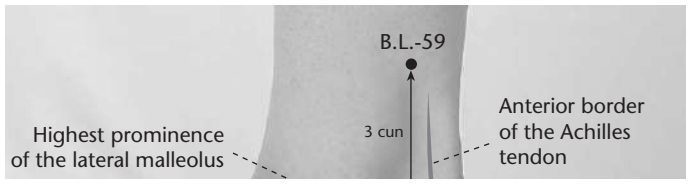

BL-59 (*fuyang*) 'Instep Yang' 3 cun superior to **BL-60** (in the depression between the highest prominence of the lateral malleolus and the Achilles tendon)

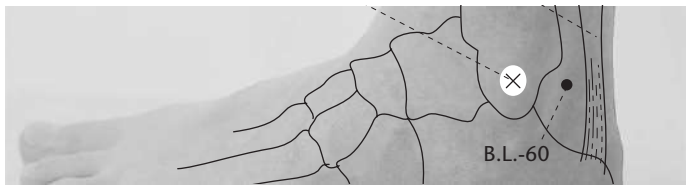

BL-60 (*kunlun*) 'Kunlun Mountains' In the depression between the Achilles tendon and the highest prominence of the lateral malleolus

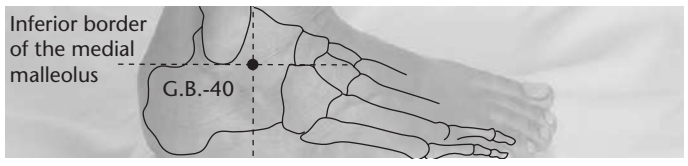

G.B.-40 (*qiuxu*) 'Large Mound' In the depression anterior and inferior to the lateral malleolus, lateral to the tendons of the extensor digitorum longus muscle.

G.B.-42 (*diwuhui*) 'Earth Five Meetings' Between the 4th and 5th metatarsal bones, proximal to the metatarsophalangeal joints and medial to the tendon of extensor digitorum longus

G.B.-43 (*jiaxi/xiaxi*) 'Clamped Stream' Between the 4th and the 5th toes, proximal to the margin of the interdigital web

Ch07-F10028.qxd 2/22/08 10:09 PM Page 656

G.B.-44 (*zuqiaoyin*) 'Yin Portals of the Foot' On the 4th toe, 0.1 cun from the lateral corner of the nail

7.13 Dorsum of the Foot

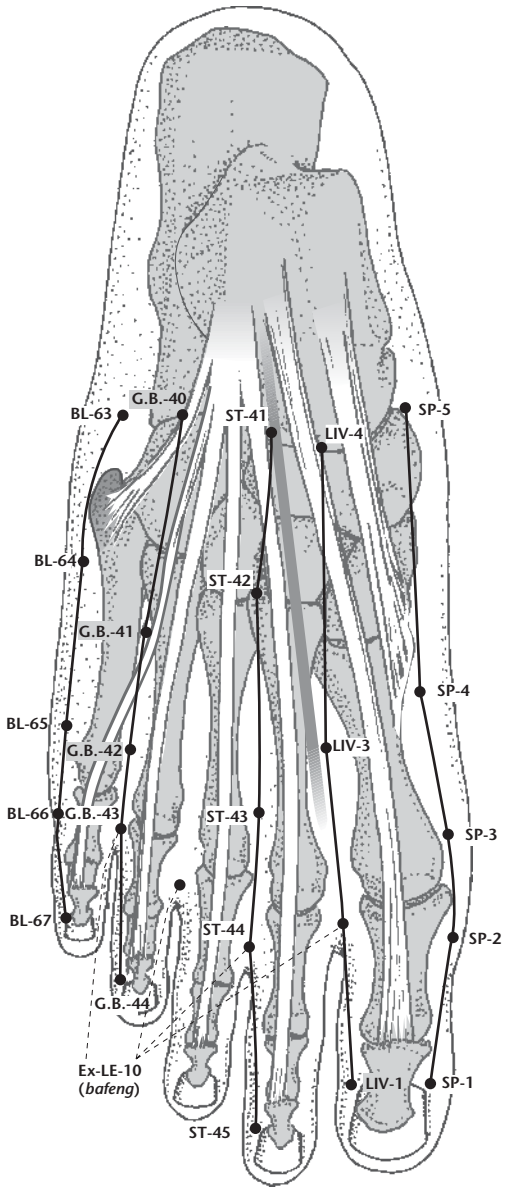

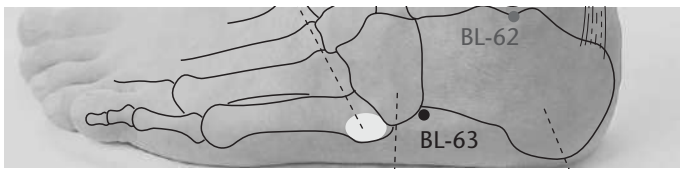

BL-63 (*jinmen*) 'Golden Gate' In a depression anterior and inferior to **BL-62** between the calcaneus and the cuboid bone, proximal to the tuberosity of the 5th metatarsal bone

Ch07-F10028.qxd 2/22/08 10:09 PM Page 657

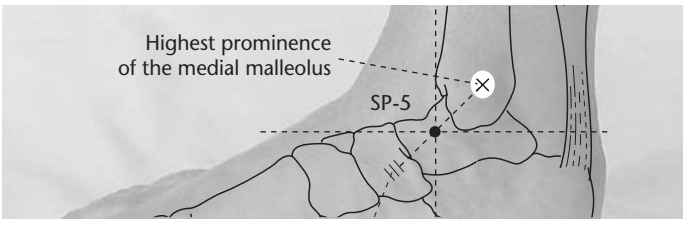

SP-5 (*shangqiu*) 'Shang Mound' In the depression at the junction of a vertical line along the anterior border and a horizontal line along the lower border of the medial malleolus. Or: in the depression halfway between the highest prominence of the medial malleolus and the navicular tuberosity

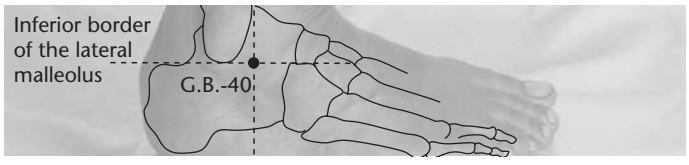

G.B.-40 (*qiuxu*) 'Large Mound' In the depression anterior and inferior to the lateral malleolus, lateral to the tendons of extensor digitorum longus muscle.

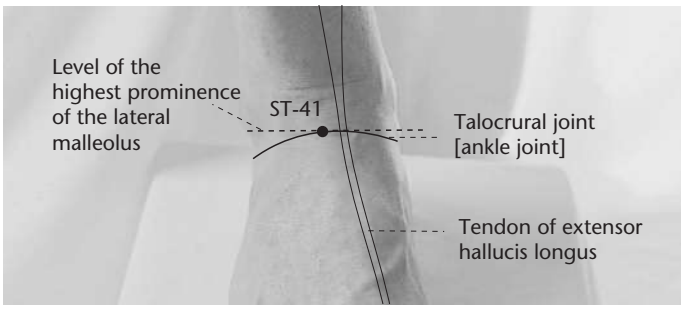

ST-41 (*jiexi*) 'Stream Divide' On the ankle, between the tendons of extensor digitorum and extensor hallucis longus, at the level of the highest prominence of the lateral malleolus

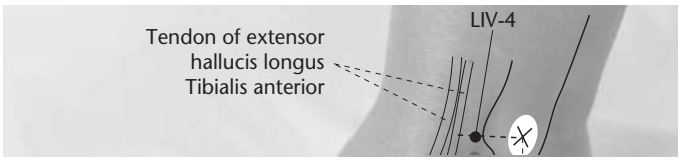

LIV-4 (*zhongfeng*) 'Middle Seal' 1 cun anterior to the prominence of the medial malleolus, medial to the tendon of the tibialis anterior

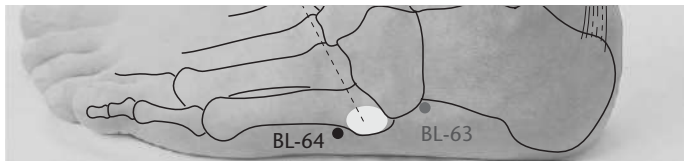

BL-64 (*jinggu*) 'Capital Bone' Distal to the tuberosity of the 5th metatarsal bone, at the border of the red and white skin

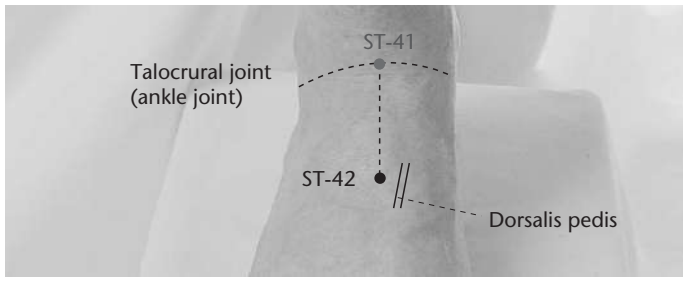

ST-42 (*chongyang*) 'Rushing Yang' On the highest point of the dorsum of the foot, between the tendons of extensor hallucis longus and extensor digitorum longus, directly lateral to where the dorsalis pedis artery may be palpated. The point is bordered proximally by the second and third metatarsal bones and distally by the second and third cuneiform bones. **Alternative location:** sometimes this point may be located lateral to the medial portion of the extensor digitorum longus tendon (joining the second toe)

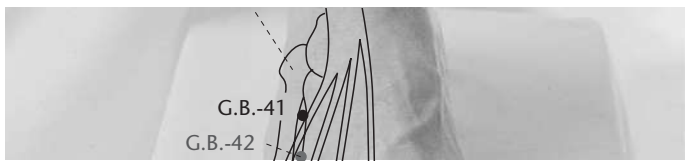

G.B.-41 (*zulinqi*) 'Foot Governor of Tears' In the depression at the junction of the shafts and the bases of the 4th and 5th metatarsal bones, lateral to the branch of the extensor digitorum longus tendon running to the little toe

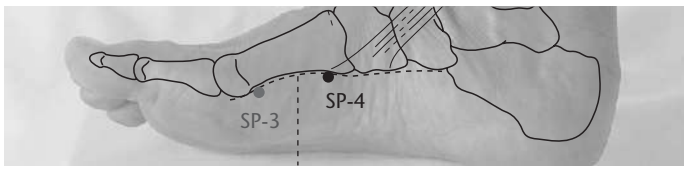

SP-4 (*gongsun*) 'Grandfather Grandson' In a depression distal and inferior to the base of the 1st metatarsal bone, at the border of the red and white skin

Ch07-F10028.qxd 2/22/08 10:10 PM Page 658

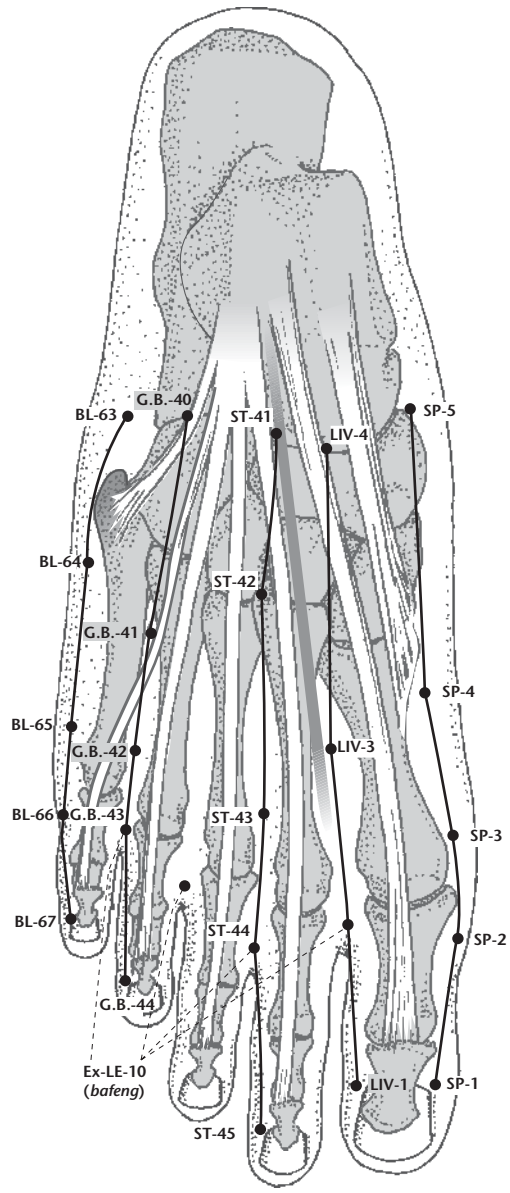

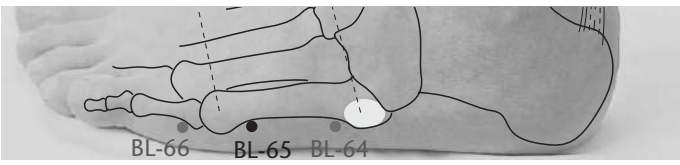

BL-65 (*shugu*) 'Restraining Bone' In the depression proximal to the head of the 5th metatarsal bone, at the border of the red and white skin

G.B.-42 (*diwuhui*) 'Earth Five Meetings' Between the 4th and 5th metatarsal bones, proximal to the metatarsophalangeal joints and medial to the tendon of the extensor digitorum longus

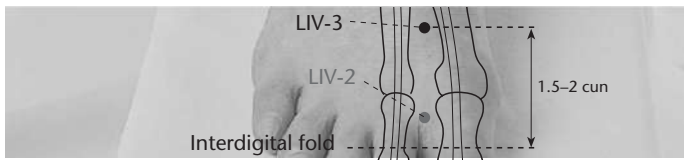

LIV-3 (*taichong*) 'Great Rushing' On the dorsum of the foot, in the depression between the 1st and 2nd metatarsal bones, approximately 1.5–2 cun proximal to the interdigital web

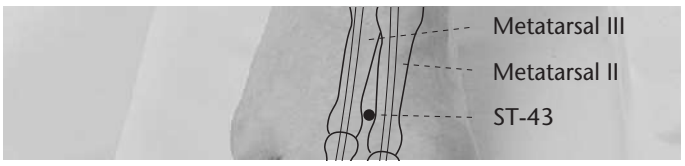

ST-43 (*xiangu*) 'Sunken Valley' In the depression between the second and third metatarsal bones, at the junction of the shafts and the heads of these metatarsal bones

G.B.-43 (*xiaxi*) 'Clamped Stream' Between the 4th and the 5th toes, proximal to the margin of the interdigital web

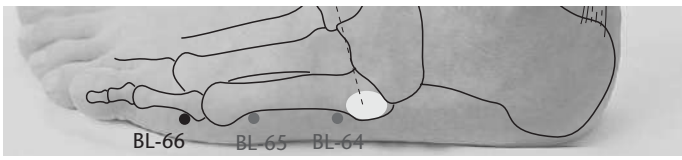

BL-66 (*zutonggu*) 'Foot Connecting Valley' In a depression distal to the metatarsophalangeal joint of the little toe

Ch07-F10028.qxd 2/22/08 10:10 PM Page 659

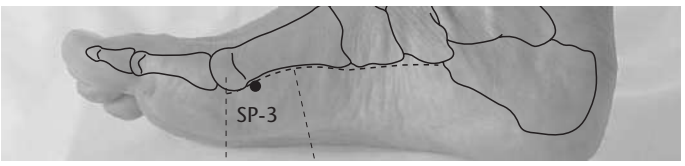

SP-3 (*taibai*) 'Supreme White' In the depression proximal to the head of the 1st metatarsal bone, at the border of the red and white skin

G.B.-44 (*zuqiaoyin*) 'Yin Portals of the Foot' On the 4th toe, 0.1 cun from the lateral corner of the nail

BL-67 (*zhiyin*) 'Reaching Yin' On the little toe, 0.1 cun from the lateral corner of the nail

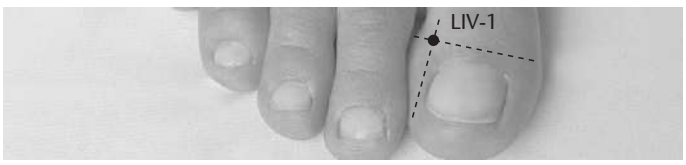

LIV-1 (*dadun*) 'Big Mound' On the lateral aspect of the big toe, 0.1 cun proximal and lateral to the corner of the nail

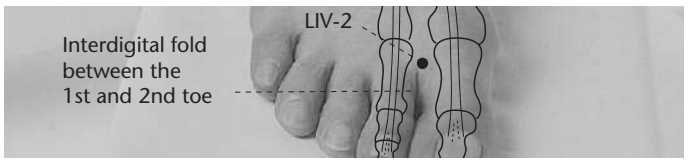

LIV-2 (*xingjian*) 'Moving Between' Between the 1st and 2nd toe, proximal to the margin of the interdigital web

SP-1 (*yinbai*) 'Hidden White' On the big toe, 0.1 cun from the medial corner of the nail

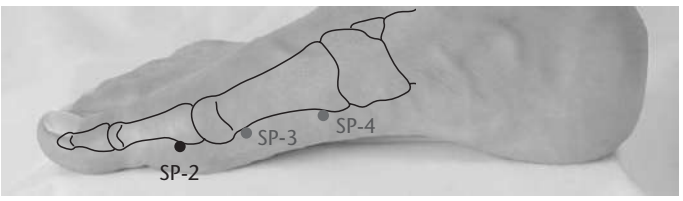

SP-2 (*dadu*) 'Great Metropolis' On the medial aspect of the big toe, in the depression distal to the first metatarsophalangeal joint, on the border of the red and white skin

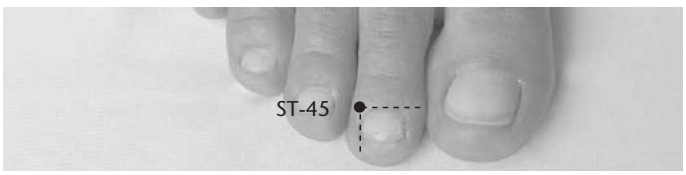

ST-45 (*lidui*) 'Strict Exchange' On the 2nd toe, 0.1 cun from the lateral corner of the nail

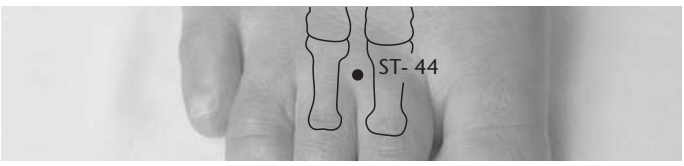

ST-44 (*neiting*) 'Inner Court' Between the 2nd and the 3rd toe, proximal to the interdigital fold

Ch07-F10028.qxd 2/22/08 10:10 PM Page 660

7.14 Medial Foot

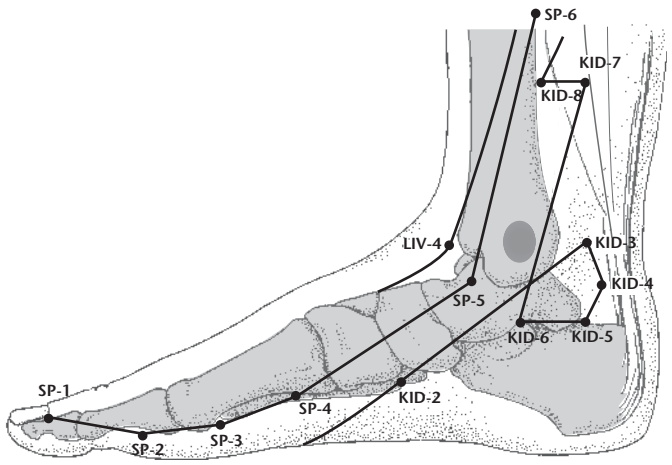

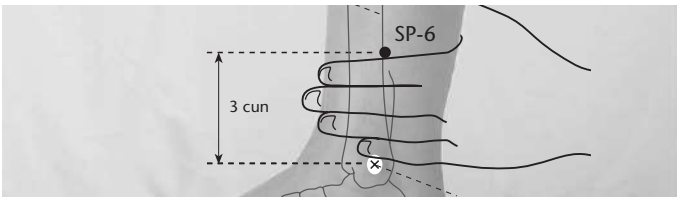

SP-6 (*sanyinjiao*) 'Three Yin Intersection' 3 cun proximal to the highest prominence of the medial malleolus, on the posterior border of the medial crest of the tibia

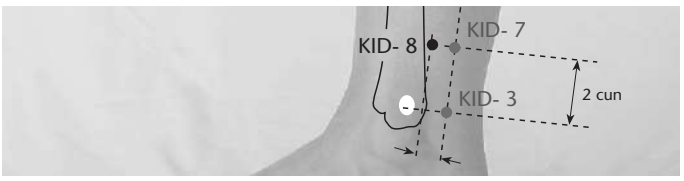

KID-8 (*jiaoxin*) 'Exchange Belief' 2 cun proximal to the highest prominence of the medial malleolus, posterior to the medial border of the tibia.

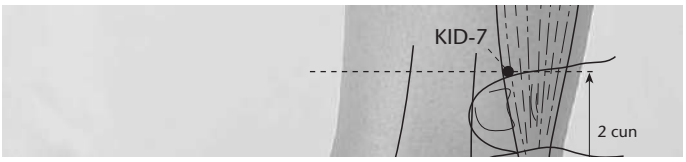

KID-7 (*fuliu*) 'Returning Current' 2 cun proximal to **KID-3**, in a depression on the anterior border of the Achilles tendon

LIV-4 (*zhongfeng*) 'Middle Seal' 1 cun anterior to the prominence of the medial malleolus, medial to the tendon of tibialis anterior

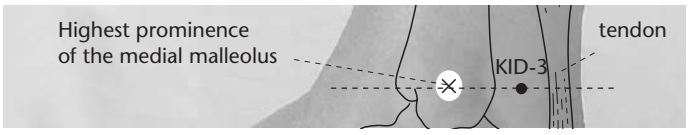

KID-3 (*taixi*) 'Supreme Stream' In the depression between the highest prominence of the medial malleolus and the Achilles tendon

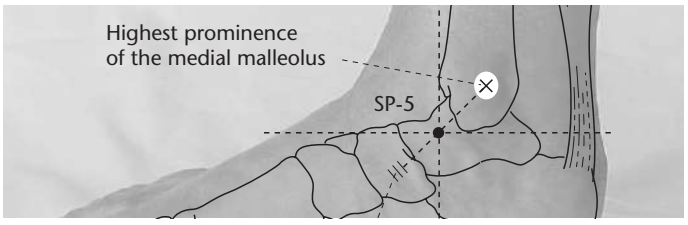

SP-5 (*shangqiu*) 'Shang Mound' In the depression at the junction of a vertical line along the anterior border and a horizontal line along the lower border of the medial malleolus. Or: in the depression halfway between the highest prominence of the medial malleolus and the navicular tuberosity

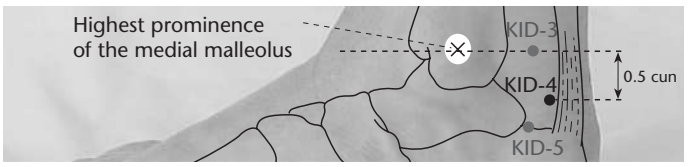

KID-4 (*dazhong*) 'Great Bell' Anterior to the medial border of the Achilles tendon, superior to its insertion at the calcaneus

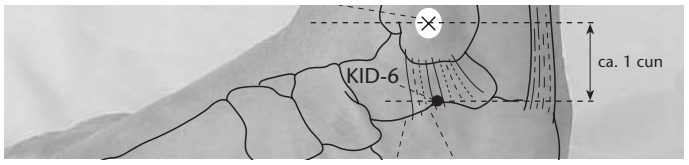

KID-6 (*zhaohai*) 'Shining Sea' In the depression inferior to the highest prominence of the medial malleolus, over the joint space between the talus and the calcaneus

Ch07-F10028.qxd 2/22/08 10:10 PM Page 661

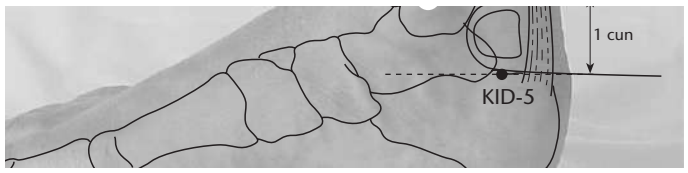

KID-5 (*shuiquan*) 'Water Spring' 1 cun distal to **KID-3**, in a depression over the joint space between the talus and calcaneus

SP-1 (*yinbai*) 'Hidden White' On the big toe, 0.1 cun from the medial corner of the nail

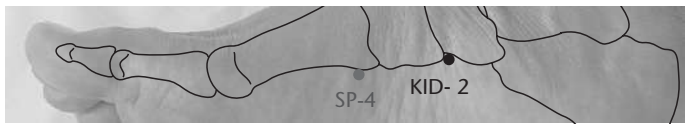

KID-2 (*rangu*) 'Blazing Valley' In a depression at the anterior inferior border of the navicular bone, at the border of the red and white skin

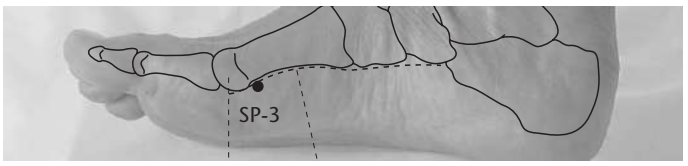

SP-3 (*taibai*) 'Supreme White' In the depression proximal to the head of the 1st metatarsal bone, at the border of the red and white skin

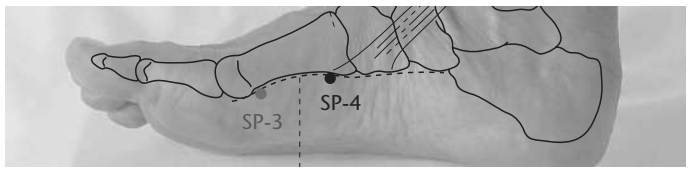

SP-4 (*gongsun*) 'Grandfather Grandson' In a depression distal and inferior to the base of the 1st metatarsal bone, at the border of the red and white skin

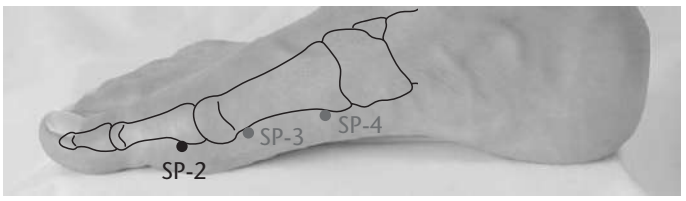

SP-2 (*dadu*) 'Great Metropolis' On the medial aspect of the big toe, in the depression distal to the first metatarsophalangeal joint, on the border of the red and white skin

Ch07-F10028.qxd 2/22/08 10:10 PM Page 662

7.15 Lateral Foot

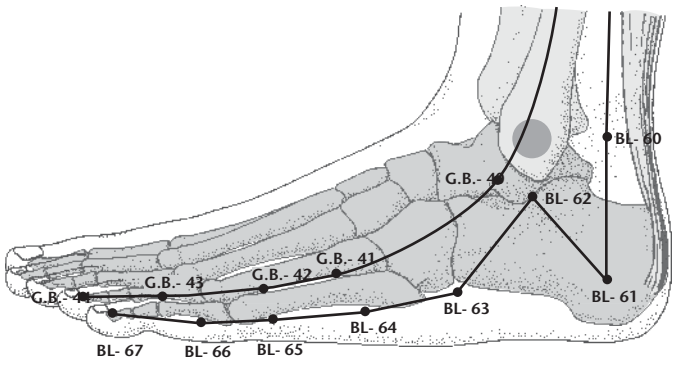

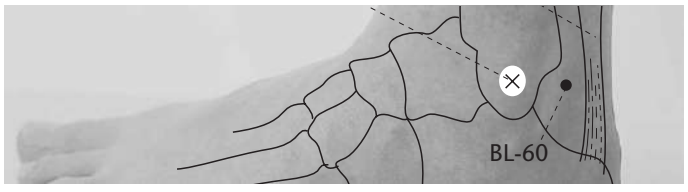

BL-60 (*kunlun*) 'Kunlun Mountains' In the depression between the Achilles tendon and the highest prominence of the lateral malleolus

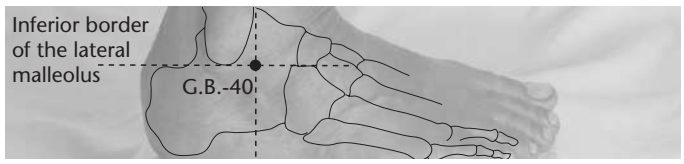

G.B.-40 (*qiuxu*) 'Large Mound' In the depression anterior and inferior to the medial malleolus, lateral to the tendons of extensor digitorum longus muscle.

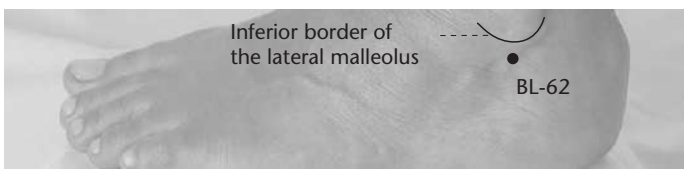

BL-62 (*shenmai*) 'Extending Vessel' In a depression directly inferior to the highest prominence of the lateral malleolus, over the joint space between the talus and calcaneus

G.B.-41 (*zulinqi*) 'Foot Governor of Tears' In the depression at the junction of the shafts and the bases of the 4th and 5th metatarsal bones, lateral to the branch of the extensor digitorum longus tendon running to the little toe

G.B.-42 (*diwuhui*) 'Earth Five Meetings' Between the 4th and 5th metatarsal bones, proximal to the metatarsophalangeal joints and medial to the tendon of extensor digitorum longus

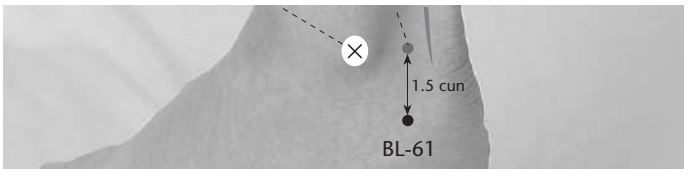

BL-61 (*pucan*) 'Servant's Respect' In a depression on the calcaneus, approximately 1.5 cun inferior to **BL-60**

G.B.-43 (*xiaxi*) 'Clamped Stream' Between the 4th and the 5th toes, proximal to the margin of the interdigital web

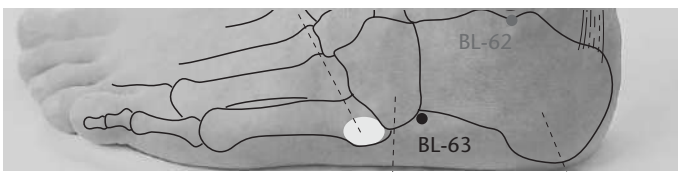

BL-63 (*jinmen*) 'Golden Gate' On the lateral aspect of the foot, proximal to the tuberosity of the 5th metatarsal bone, in a depression anterior and inferior to **BL-62** between the calcaneus and the cuboid bone

G.B.-44 (*zuqiaoyin*) 'Yin Portals of the Foot' On the 4th toe, 0.1 cun from the lateral corner of the nail

Ch07-F10028.qxd 2/22/08 10:10 PM Page 663

BL-*67* (*zhiyin*) 'Reaching Yin' On the little toe, 0.1 cun from the lateral corner of the nail

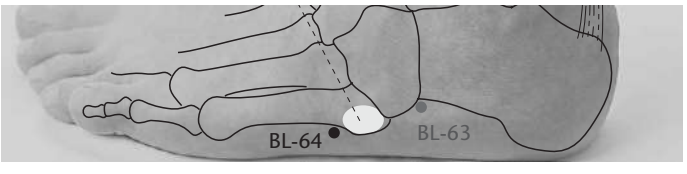

BL-*64* (*jinggu*) 'Capital Bone' Distal to the tuberosity of the 5th metatarsal bone, at the border of the red and white skin

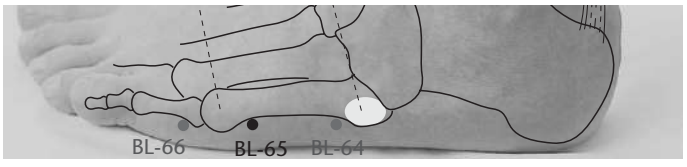

BL-*65* (*shugu*) 'Restraining Bone' In the depression proximal to the head of the 5th metatarsal bone, at the border of the red and white skin

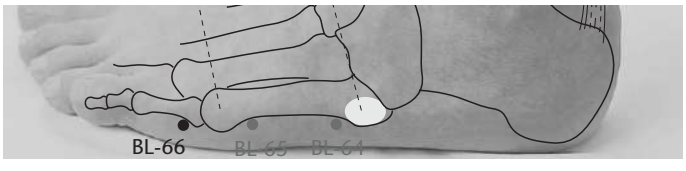

BL-*66* (*zutonggu*) 'Foot Connecting Valley' In a depression distal to the metatarsophalangeal joint of the little toe

8 Point Categories and Point Selection

Claudia Focks

There are a total of 361 acupuncture points located on the channels (→ Chapter 1). Translated literally from the Chinese language, the meaning of the word for 'point' is rather 'opening' or 'hole'. Stimulation of the points can be achieved through massage (in acupressure) or through the insertion of needles (in acupuncture). Besides the regular acupuncture points, there exist a wide variety of so-called extra points, located outside the pathways of the channels. The most commonly used of these extra points are described in Chapter 6, as classified by the WHO.

The acupuncture points are effective

- **locally** (for example, points on the elbow treat disorders of the elbow)
- **regionally** (the channel pathway – for example, points on the arm treat disorders of the upper extremity)
- some points have an effect on their **pertaining** ***zangfu***-Organ
- some points have **additional actions** such as calming the *shen* or eliminating Heat.

The following table provides an overview for the general indications of the channel points (modified after Ellis and Wiseman 1991)

Indications of the channel points

Channel	Indications		
Hand-Yin channels			
LU	Lungs, throat		Disorders of the thoracic region
P	Heart, Stomach	Spirit (*shen*) disorders	
HE	Heart		
Hand-Yang channels			
L.I.	Face, mouth, occiput, teeth, nose		Eyes, disorders of the throat, febrile diseases
T.B.	Head/lateral costal region	Disorders of the ears	
S.I.	Occipital region, shoulder, scapula, *shen* disorders		
Foot-Yin channels			
SP	Spleen, Stomach, intestines		Urogenital disorders, menstrual disorders, vaginal discharge, spirit (*shen*) disorders
LIV	Liver, genitalia		
KID	Kidneys, Lungs, throat		
Foot-Yang channels			
ST	Frontal aspect of the head, face, mouth, teeth, throat, Stomach, intestines		Psychoemotional disorders (*shen* disorders), febrile colds
G.B.	Lateral aspect of the head, ears, lateral costal region	Disorders of the eyes	
BL	Posterior aspect of the head, lower back		

8.1 Point Categories

Certain acupuncture points or groups of points are categorised according to their actions or location on their respective channels. (For an overview of these points and point groups → inside back cover.)

8.1.1 *Yuan*-Source Points

At the *yuan*-source points, the Original *Qi* (*yuan qi* → 1.1.4) of the respective Organ system emerges to the surface of the body.
Location: near the hand or ankle.

Ch08-F10028.qxd 2/23/08 4:40 PM Page 666

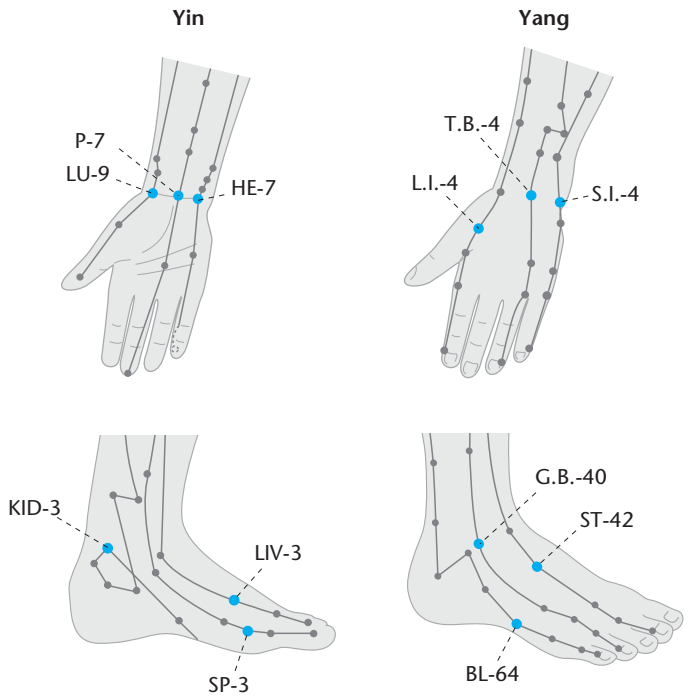

Points

Yuan-source points of the Yin channels

LU-9 (*taiyuan*) **SP-3** (*taibai*) **HE-7** (*shenmen*)
KID-3 (*taixi*) **P-7** (*daling*) **LIV-3** (*taichong*)

- On the Yin channels the *yuan*-source point is always the **3rd point** from the distal end of the channel
- They are identical with the *shu*-stream points and the Earth point of the channel

Function

- They tonify deficiency or weakness of their pertaining *zang*-Organ
- They regulate the Yin/Yang balance, restoring homeostasis
- As Earth points they have a 'grounding' effect on the body, the emotions and the mind.

Yuan-source points of the Yang channels

L.I.-4 (*hegu*) **ST-42** (*chongyang*) **S.I.-4** (*wangu*)
BL-64 (*jinggu*) **T.B.-4** (*yangchi*) **G.B.-40** (*qiuxu*)

- On the Yang channels the *yuan*-source point is always the **4th point** from the distal end of the channel, except for the G.B. channel (5th point)
- Regarding the energetics of the channel, they are all located between the *shu*-stream point and the *jing*-river point.

Function

- Their main function is to expel pathogenic factors in excess syndromes
- They tonify their pertaining *fu*-Organ.

Clinical application

Diagnosis

Disorders of a particular channel or its pertaining *zangfu*-Organ will often manifest themselves with tenderness, skin changes such as discolorations, swellings, redness, etc. in the area of the respective *yuan*-source point.

Therapy

- For deficiency of the relevant channel or Organ, use tonifying needling techniques; for excess, use reducing techniques
- Combine with the *luo*-connecting point of its Interiorly–Exteriorly paired channel (→ 1.2.2) in order to increase the therapeutic effect and to balance Yin and Yang (→ 8.3.3)
- Combine with the respective Back-*shu* point (→ 8.1.4) for disorders of the *zangfu*-Organs

8.1.2 *Luo*-Connecting Points

The *luo*-connecting points are located in those regions of the body where the *luo*-network vessels (→ 1.5) branch out, connecting the Interiorly–Exteriorly paired Yin and Yang channels (→ 1.2.2).

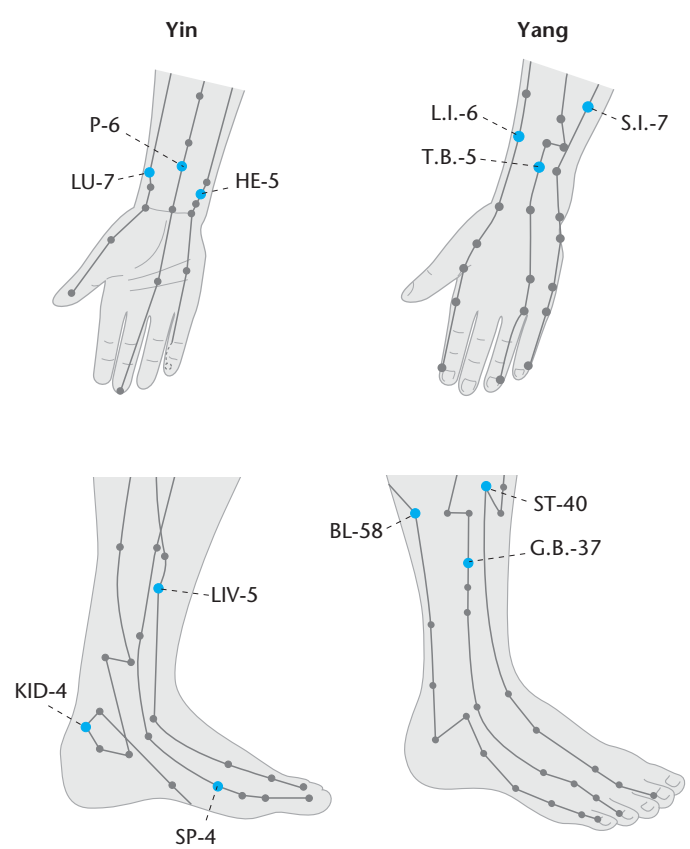

Ch08-F10028.qxd 2/23/08 4:40 PM Page 667

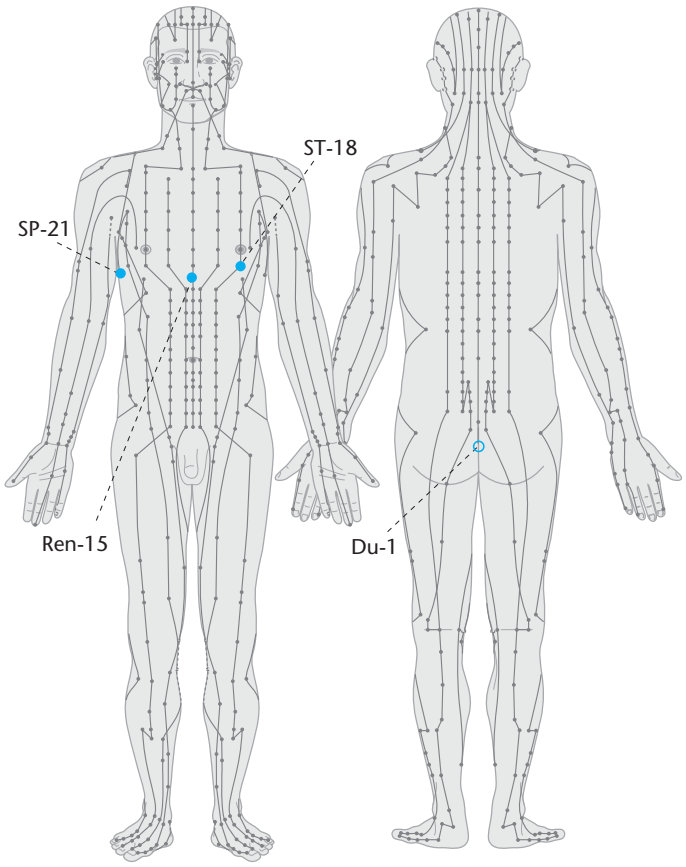

Points

The 12 *luo*-connecting points of the primary channels are located between the fingers and the elbow on the upper extremity and between the toes and the knee on the lower extremity.

- *Luo*-connecting points of the primary Yin channels
  - hand Yin channels: **LU-7** (*lieque*), **P-6** (*neiguan*), **HE-5** (*tongli*)
  - foot Yin channels: **SP-4** (*gongsun*), **LIV-5** (*ligou*), **KID-4** (*dazhong*)
- *Luo*-connecting points of the primary Yang channels
  - hand Yang channels: **L.I.-6** (*pianli*), **T.B.-5** (*waiguan*), **S.I.-7** (*zhizheng*)
  - foot Yang channels: **ST-40** (*fenglong*), **G.B.-37** (*guangming*), **BL-58** (*feiyang*)
- *Luo*-connecting points of the *ren mai:* **Ren-15** (*jiuwei*); action: regulates the *luo*-connecting vessels of the Yin channels; indication: abdominal diseases and disorders
- *Luo*-connecting points of the *du mai:* **Du-1** (*changqiang*); action: regulates the *luo*-connecting vessels of the Yang channels; indication: disorders of the back
- Great *luo*-connecting point of the Spleen: **SP-21** (*dabao*); action: connects with all *luo*-vessels and governs all *luo* and Blood vessels of the body; indication: disorders of the joint, general weakness and pain
- Great *luo*-connecting point of the Stomach: ***xu li****, near the left **ST-18**; moxibustion is contraindicated, only superficial needling is recommended. Indication: breathing and heart disorders such as asthma, dyspnoea, emphysema and bradycardia, also for shortness of breath due to anxiety attacks.

Clinical application

Diagnosis

- In excess syndromes the *luo* vessels often become visible near the *luo*-connecting points, for example skin discolorations and swellings may occur. With pain due to Cold the predominant colour may be bluish to greenish, with Heat syndromes reddish, with chronic stagnation and stasis dark to dark purple.
- In deficiency syndromes there are often no colour changes. However, in severe chronic cases (for example severe Qi deficiency) there will more likely be distinct muscular flaccidity, presenting as a dip or depression in the area of the point.

Therapy

- For disorders of the pertaining *zangfu*-Organ or channel
- For diseases of the Interiorly–Exteriorly paired Organ or channel
- For disorders due to the accumulation of pathogenic factors or substances, for example Qi stagnation, Blood stasis, retention of Dampness or Phlegm
- For psychoemotional problems, especially *luo*-connecting points of the Yin channels, for example classically **P-6** (*neiguan*) and **HE-5** (*tongli*)
- Bloodletting (three-edged needle, lancet or plum-blossom needle) on the *luo*-connecting point of the affected channel: this is particularly indicated for excess syndromes with Qi stagnation and Blood stasis. For more details on point selection → 8.3.3

8.1.3 *Xi*-Cleft Points

Synonyms: Accumulation points (Maciocia)

The *xi*-cleft points are all located between the fingers and the elbow on the upper extremity and between the toes and the knee on the lower extremity. Only **ST-34** is located more proximally on the thigh.

* *Xu li* is only rarely mentioned in Western literature. The location and significance of this point is based on *Su Wen*, chapter 18: 'The geat *luo* vessel of the Stomach is called *xu li* and emerges at **ST-18** under the left breast, passes through the diaphragm and ascends to connect with the Lung. One can feel its pulsation with the hand. It is the place where the *zong qi* (→ 1.1.4) of the channels resides.' (Ngyen Van Nghi, 1989). *Xu li* is the only *luo*-connecting vessel to directly emerge from a *fu*-Organ. It plays an important role in Korean acupuncture.

Ch08-F10028.qxd 2/23/08 4:40 PM Page 668

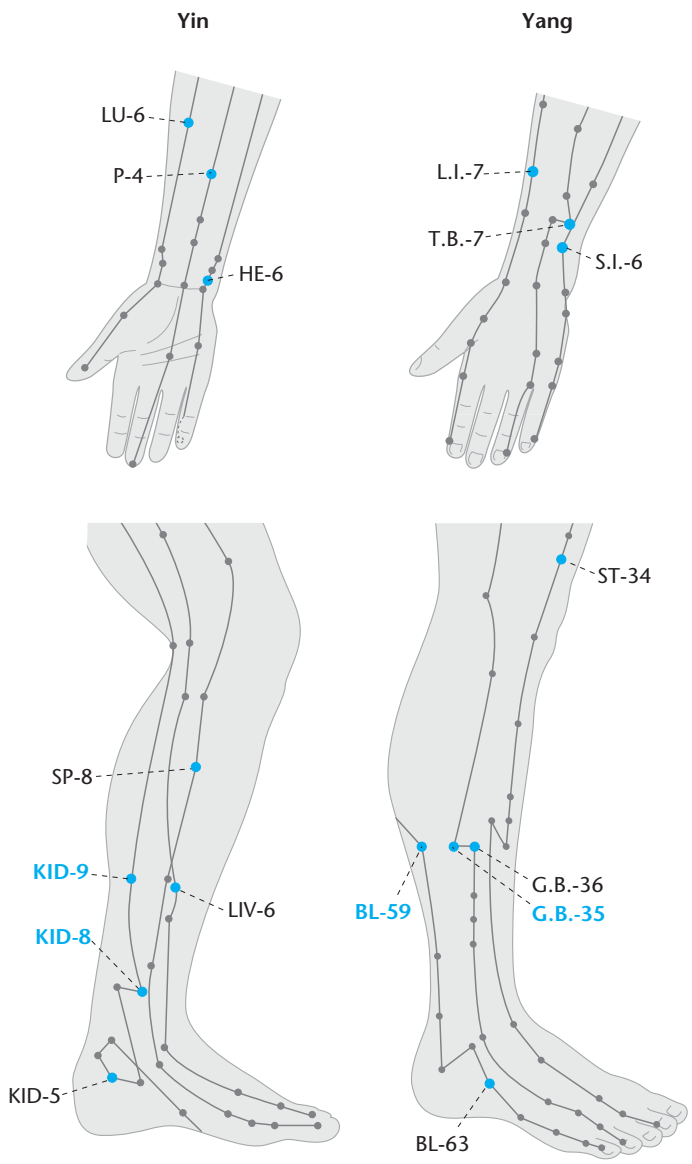

Points

Xi-cleft points of the primary Yin channels

Hand Yin channels

LU-6 (*kongzui*) **HE-6** (*yinxi*) **P-4** (*ximen*)

Foot Yin channels

SP-8 (*diji*) **KID-5** (*shuiquan*) **LIV-6** (*zhongdu*)

Xi-cleft points of the primary Yang channels

Hand Yang channels

L.I.-7 (*wenliu*) **S.I.-6** (*yanglao*) **T.B.-7** (*huizhong*)

Foot Yang channels

ST-34 (*liangqiu*) **BL-63** (*jinmen*) **G.B.-36** (*waiqiu*)

Xi-cleft points of the extraordinary vessels (point names printed in blue)

- *Xi*-cleft point of the *yang qiao mai*: **BL-59** (*fuyang*)
- *Xi*-cleft point of the *yin qiao mai*: **KID-8** (*jiaoxin*)
- *Xi*-cleft point of the *yang wei mai*: **G.B.-35** (*yangjiao*)
- *Xi*-cleft point of the *yin wei mai*: **KID-9** (*zhubin*)

Clinical application

Diagnosis

- They provide information about disorders affecting their pertaining channel or *zangfu*-Organ
  - Excess: acute, intense pain upon pressure or redness and swelling in the area of the point
  - Deficiency: dull, mild pain upon pressure or presence of a depression in the area of the point

Therapy

- Mainly used for acute and/or therapy-resistant disorders, especially for excess conditions and pain of the affected channel and/or pertaining *zangfu*-Organ. In patients with a robust constitution use reducing needling techniques.
- Add the *xi*-cleft points of the Yin channels for disorders of the Blood.
- In order to increase the therapeutic effect, the *xi*-cleft point may be combined with a relevant *hui*-meeting point (→ 8.1.7).

8.1.4 Back-*Shu* Points

Synonym: Back-transporting points

The Back-*shu* points are located on the inner branch of the BL channel at the level of their pertaining *zangfu*-Organ. They are mostly at the same level as the corresponding Front-*mu* point (→ 8.1.5).

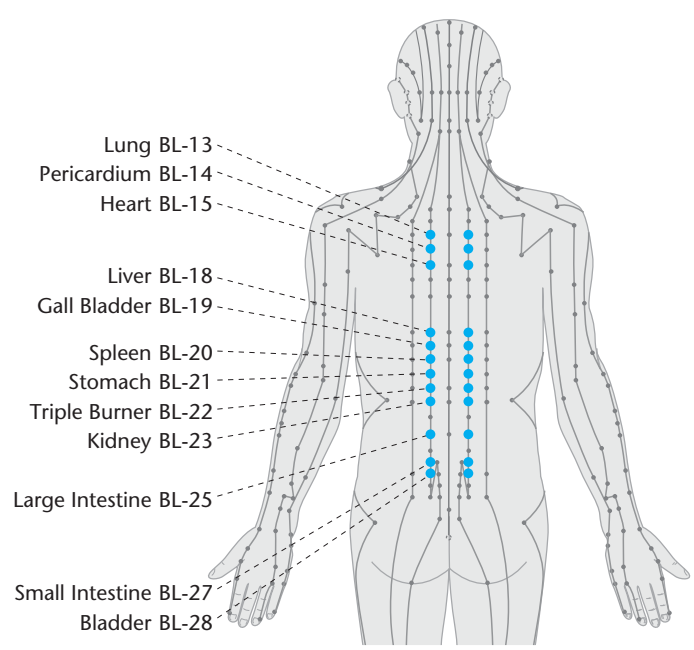

Ch08-F10028.qxd 2/23/08 4:40 PM Page 669

Points

Organ	Back-*shu* point
Lung	**BL-13** (*feishu*)
Large Intestine	**BL-25** (*dachangshu*)
Stomach	**BL-21** (*weishu*)
Spleen	**BL-20** (*pishu*)
Heart	**BL-15** (*xinshu*)
Small Intestine	**BL-27** (*xiaochangshu*)
Bladder	**BL-28** (*pangguangshu*)
Kidney	**BL-23** (*shenshu*)
Pericardium	**BL-14** (*jueyinshu*)
Triple Burner	**BL-22** (*sanjiaoshu*)
Gall Bladder	**BL-19** (*danshu*)
Liver	**BL-18** (*ganshu*)

Clinical application

Diagnosis

- Often spontaneously tender and/or sensitive upon pressure in disorders of the pertaining *zangfu*-Organ

Therapy

- Mainly used for disorders of the corresponding *zangfu*-Organ. These points have a tonifying effect for chronic disorders and deficiency syndromes, but can also be used for acute disorders of their pertaining *zangfu*-Organ.
- For disorders of the sensory organs, use the Back-*shu* point of the pertaining *zangfu*-Organ.
  Example: for disorders of the eyes, needle the Back-*shu* point of the Liver **BL-18** (*ganshu*).

For excess syndromes, the Back-*shu* points can be needled with reducing needling techniques; for deficiency syndromes, tonifying needling techniques should be applied. When Cold signs are present, moxibustion can also be used. The needles should not be retained longer than 10 minutes – longer needle retention can have a sedating effect and lead to tiredness.

8.1.5 Front-*Mu* Points

Synonym: Alarm points

The Front-*mu* points are located on the thorax and abdomen in the area of their pertaining *zangfu*-Organs.

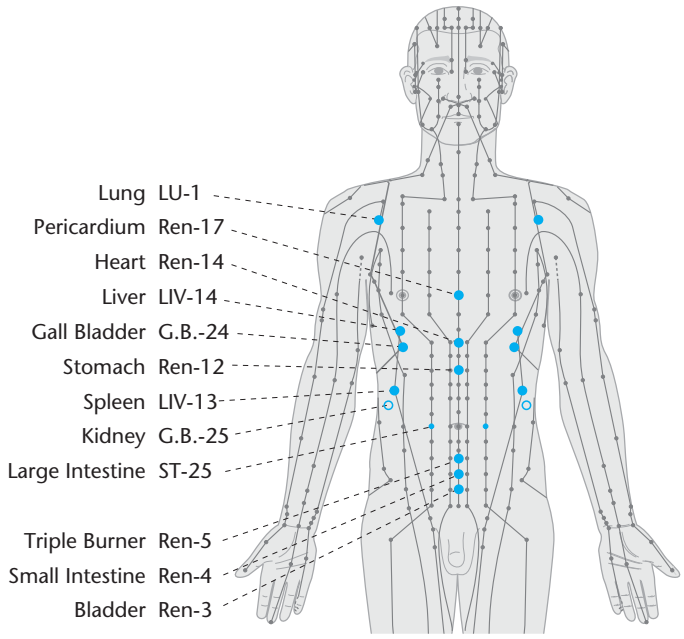

Points

Organ	Back-*shu* point
Lung	**LU-1** (*zhongfu*)
Large Intestine	**ST-25** (*tianshu*)
Stomach	**Ren-12** (*zhongwan*)
Spleen	**LIV-13** (*zhangmen*)
Heart	**Ren-14** (*juque*)
Small Intestine	**Ren-4** (*guanyuan*)
Bladder	**Ren-3** (*zhongji*)
Kidney	**G.B.-25** (*jingmen*)
Pericardium	**Ren-17** (*danzhong*)
Triple Burner	**Ren-5** (*shimen*)
Gall Bladder	**G.B.-24** (*riyue*)
Liver	**LIV-14** (*qimen*)

Clinical application

Diagnosis

- Often spontaneously tender or sensitive upon pressure in disorders of the pertaining *zangfu*-Organ

Therapy

- Used to regulate the pertaining *zangfu*-Organ. For excess syndromes use reducing needling techniques; for deficiency syndromes use tonifying methods.
- Commonly used point combinations:
  – *Shu–mu* method (→ 8.3.2)
  – For disorders of the *fu*-Organs: simultaneous needling of the lower *he-sea* point (→ 8.1.9) and the Front-*mu* point of the affected *fu*-Organ.

Ch08-F10028.qxd 2/23/08 4:40 PM Page 670

8.1.6 The Five *Shu*-Transporting Points

The five *shu-t*ransporting points are located **between the fingers and the elbow on the upper extremity, between the toes and the knee on the lower extremity**. The classification and application of these points is based on two theories: on the one hand they are applied according to the Qi flow in the channel from distal to proximal, on the other hand they are used according to the Five Phases (→ 8.2.5).

While the location of the five *shu-t*ransporting points is identical to the Five Phase points, their dynamics and range of application based on the Qi flow in the channels differs from or is even contrary to their actions as Five Phase points. The theoretical foundations of the Qi flow in the channels as applied in the instance of the *shu*-transporting point are based on a rather earlier circulatory concept (→ 1.1.1).

The therapeutic application of the Five Phase points will be explained in more detail in → 8.2.5, while the clinical application according to channel flow is described below.

Theoretical background of the Qi circulation in the channels

According to the more original centripetal model (→ 1.1.1), the Qi flows from the peripheral parts of the body to the elbows and knees, similar to a river coursing from its spring to its estuary and finally into the sea:

- The Qi flow begins at the anastomosis at the *jing-w*ell point. Here the water surfaces, it is dynamic and unstable (change of polarity from Yin to Yang)
- The water course then changes to a spring (*ying*) and a stream (*shu*) – both are at this point still quite dynamic features, the water flowing rapidly and superficially.
- The water then becomes increasingly deeper and slower, taking on the characteristics of a river (*jing*).
- Eventually it flows as a big river into the sea (*he*).

This image of a river system is often used to explain disorders caused by pathogenic factors (*xie qi*). Due to the superficial position of these points on the extremities, pathogens can here easily penetrate the body. By the same token, they can be removed again with relatively little effort. Barbara Kirschbaum (course material 1999) provides a useful visual image: she suggests imagining a ship carrying a cargo of pathogens that begins its journey at the tips of the extremities. The further a point is located peripherally in the body, the easier it is for the ship to get rid of its dangerous cargo (or to expel the pathogen according to the dynamics of the Qi flow). However, once the ship has reached the *he*-sea point, the ship and its cargo will disappear in the big, wide sea, having the potential to injure the internal *zangfu*-Organ. Thus it is the therapeutic goal to eliminate pathogens from the body as soon as possible.

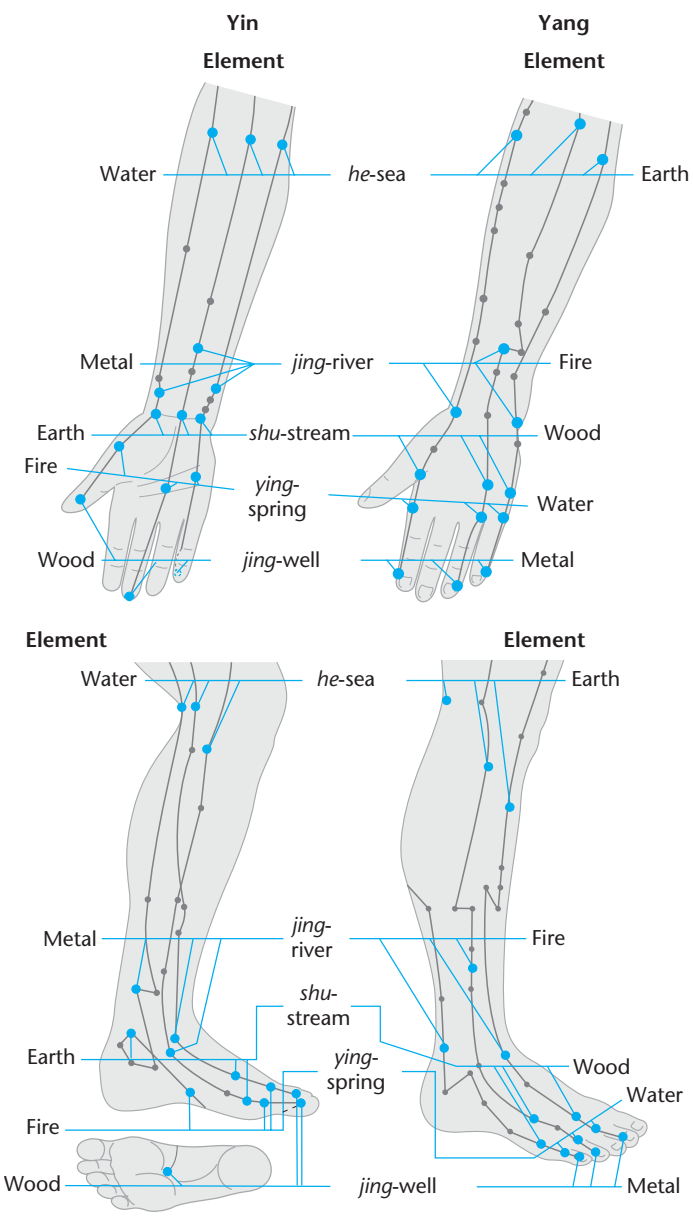

Ch08-F10028.qxd 2/23/08 4:40 PM Page 671

Overview: *shu*-transporting points

Shu-transporting points of the Yin channels					
Phase/ channel	Wood point	Fire point	Earth point	Metal point	Water point
LU	LU-11	LU-10	LU-9 +	LU-8	LU-5 −
SP	SP-1	SP-2 +	SP-3	SP-5 −	SP-9
HE	HE-9 +	HE-8	HE-7 −	HE-4	HE-3
KID	KID-1 −	KID-2	KID-3	KID-7 +	KID-10
P	P-9 +	P-8	P-7 −	P-5	P-3
LIV	LIV-1	LIV-2 −	LIV-3	LIV-4	LIV-8 +
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point
***Shu*-transporting points of the Yang channels**					
Phase/ channel	Metal point	Water point	Wood point	Fire point	Earth point
L.I.	L.I.-1	L.I.-2 −	L.I.-3	L.I.-5	L.I.-11 +
ST	ST-45 −	ST-44	ST-43	ST-41+	ST-36
S.I.	S.I.-1	S.I.-2	S.I.-3 +	S.I.-5	S.I.-8 −
BL	BL-67 +	BL-66	BL-65 −	BL-60	BL-40
T.B.	T.B.-1	T.B.-2	T.B.-3 +	T.B.-6	T.B.-10 −
G.B.	G.B.-44	G.B.-43+	G.B.-41	G.B.-38 −	G.B.-34
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point

+, tonification point; − , sedation point

Clinical application of the *shu*-transporting points

Jing-well points (*jing xue*)

Shu-transporting Points of the Yin channels					
Phase/ channel	Wood point	Fire point	Earth point	Metal point	Water point
LU	LU-11	LU-10	LU-9 +	LU-8	LU-5 −
SP	SP-1	SP-2 +	SP-3	SP-5 −	SP-9
HE	HE-9 +	HE-8	HE-7 −	HE-4	HE-3
KID	KID-1 −	KID-2	KID-3	KID-7 +	KID-10
P	P-9 +	P-8	P-7 −	P-5	P-3
LIV	LIV-1	LIV-2 −	LIV-3	LIV-4	LIV-8 +
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point
***Shu*-transporting points of the Yang channels**					
Phase/ channel	Metal point	Water point	Wood point	Fire point	Earth point
L.I.	L.I.-1	L.I.-2 −	L.I.-3	L.I.-5	L.I.-11 +
ST	ST-45 −	ST-44	ST-43	ST-41 +	ST-36
S.I.	S.I.-1	S.I.-2	S.I.-3 +	S.I.-5	S.I.-8 −
BL	BL-67 +	BL-66	BL-65 −	BL-60	BL-40
T.B.	T.B.-1	T.B.-2	T.B.-3 +	T.B.-6	T.B.-10 −
G.B.	G.B.-44	G.B.-43+	G.B.-41	G.B.-38 −	G.B.-34
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point

+, tonification point; − , sedation point

According to the Five Phases (→ 8.2.5) the *jing*-well points of the **Yin channels** correspond to the **Wood points**, while those of the **Yang channels** correspond to the **Metal points**.
Location: as the **last or first point of each channel**, they are located on the **tips of the fingers or the toes** (exception: **KID-1** of the sole of the foot). The course of the channel is here at its **most superficial**. It is also here that a change of polarity is taking place in the energy: from Yin to Yang or vice versa. **Caution**: needling these points can be quite intense and painful, presenting a strong stimulus to the body.

Actions

- Clear Heat, restore consciousness, eliminate Heat and excess from the upper end of the channel: commonly pricked to bleed.
- Treat fullness and Cold below the Heart
- Treat disorders of the spirit (*shen*) for psychological disorders

Indications

- For acute conditions and emergencies: the *jing*-well points can quickly expel Exterior pathogenic factors (*xie qi*), especially Heat. They have a strong effect on the psyche, for example for coma, collapse and shock as well as for severe irritability and restlessness.
- The points on the Yin channels also have an effect on their pertaining *zang*-Organs and can eliminate internal Wind.

Ch08-F10028.qxd 2/23/08 4:40 PM Page 672

Examples

- **LU-11**: prick to bleed for acute sore throat
- **KID-1**: for spasms and unconsciousness
- **HE-9**: for collapse
- **SP-1**: for uterine bleeding

For psychological disorders:

- **KID-1**: for severe restlessness
- **P-9** and **HE-9**: for extreme restlessness and insomnia

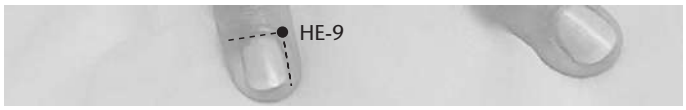

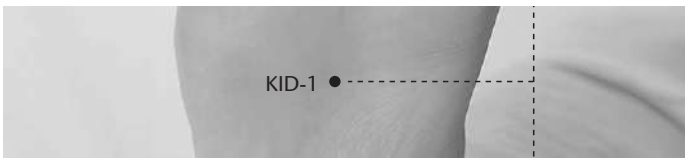

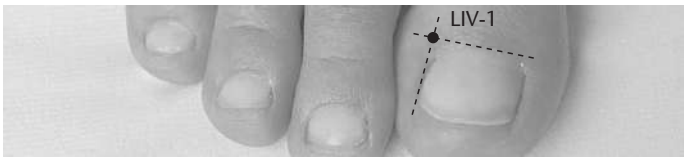

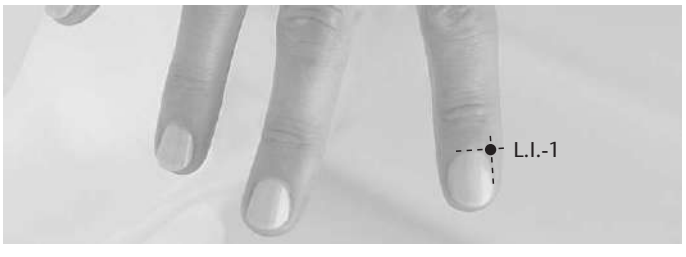

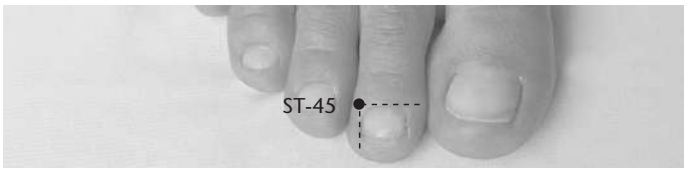

Ch08-F10028.qxd 2/23/08 4:40 PM Page 673

Ying-spring points (*ying xue*)

Shu-transporting points of the Yin channels					
Phase/ channel	Wood point	Fire point	Earth point	Metal point	Water point
LU	LU-11	LU-10	LU-9 +	LU-8	LU-5 −
SP	SP-1	SP-2 +	SP-3	SP-5 −	SP-9
HE	HE-9 +	HE-8	HE-7 −	HE-4	HE-3
KID	KID-1 −	KID-2	KID-3	KID-7 +	KID-10
P	P-9 +	P-8	P-7 −	P-5	P-3
LIV	LIV-1	LIV-2 −	LIV-3	LIV-4	LIV-8 +
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point
***Shu*-transporting points of the Yang channels**					
Phase/ channel	Metal point	Water point	Wood point	Fire point	Earth point
L.I.	L.I.-1	L.I.-2 −	L.I.-3	L.I.-5	L.I.-11 +
ST	ST-45 −	ST-44	ST-43	ST-41 +	ST-36
S.I.	S.I.-1	S.I.-2	S.I.-3 +	S.I.-5	S.I.-8 −
BL	BL-67 +	BL-66	BL-65 −	BL-60	BL-40
T.B.	T.B.-1	T.B.-2	T.B.-3 +	T.B.-6	T.B.-10 −
G.B.	G.B.-44	G.B.-43+	G.B.-41	G.B.-38 −	G.B.-34
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point

+, tonification point; − , sedation point

According to the Five Phases (→ 8.2.5) the *ying*-spring points of the Yin channels correspond to the Fire points, while those on the Yang channels correspond to the Water points. **Caution**: needling some of these points can be painful, for example **LU-10, P-8, HE-8**.

Location: as the second most distal point of each channel, they are located on the upper extremity between the phalanges and metacarpal bones and on the lower extremity between the phalanges with the exception of KID-2.

Actions

- They are very dynamic points that eliminate both external and internal pathogenic factors.
- They clear Heat from their pertaining *zangfu*-Organ and channel, especially in the upper half of the body.

Indications

- For example for febrile diseases: Exterior Heat syndromes are mainly treated using points on the Yang channels while Interior Heat or deficient Heat syndromes are treated with points on the Yin channels.

Examples

- **LU-10:** for acute inflammations of the throat
- **ST-44:** for acute frontal headaches or acute inflammations affecting the face
- **LIV-2:** for severe temporal headaches, for example due to blazing Liver Fire

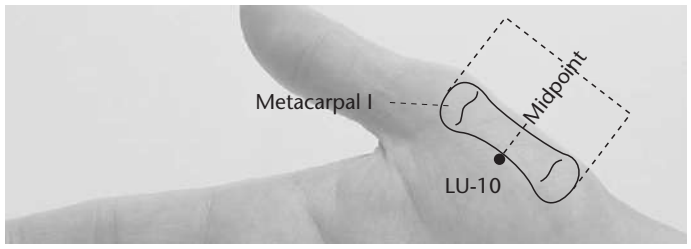

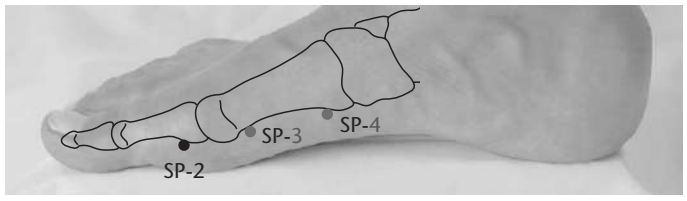

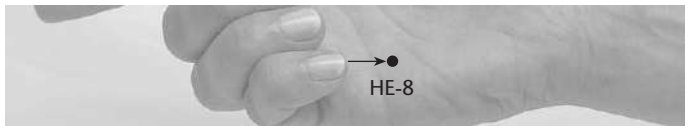

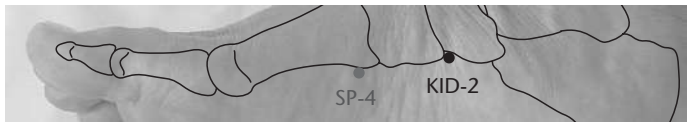

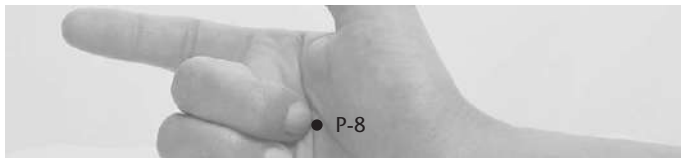

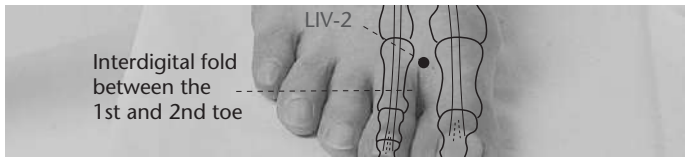

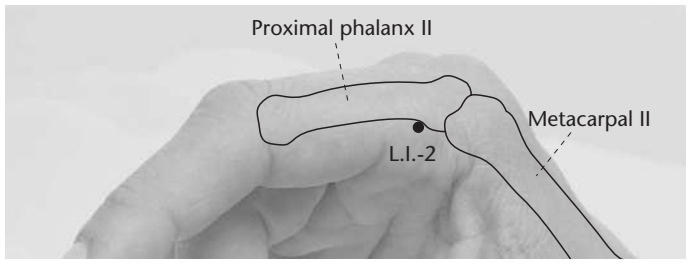

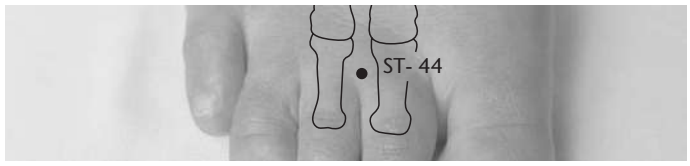

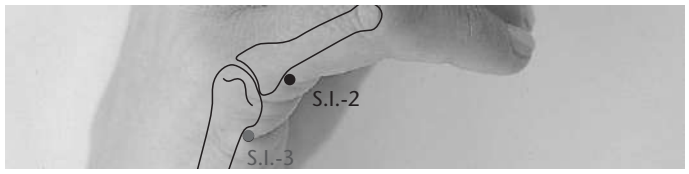

Ch08-F10028.qxd 2/23/08 4:40 PM Page 674

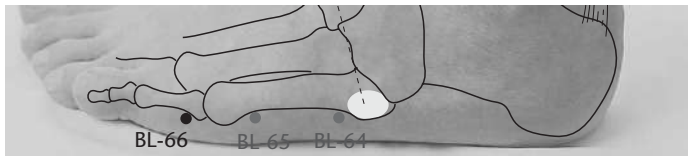

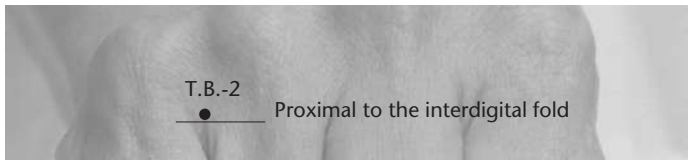

Shu-stream points (*shu xue*)

Shu-transporting points of the Yin channels					
Phase/ channel	Wood point	Fire point	Earth point	Metal point	Water point
LU	LU-11	LU-10	LU-9 +	LU-8	LU-5 −
SP	SP-1	SP-2 +	SP-3	SP-5 −	SP-9
HE	HE-9 +	HE-8	HE-7 −	HE-4	HE-3
KID	KID-1 −	KID-2	KID-3	KID-7 +	KID-10
P	P-9 +	P-8	P-7 −	P-5	P-3
LIV	LIV-1	LIV-2−	LIV-3	LIV-4	LIV-8 +
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point
***Shu*-transporting points of the Yang channels**					
Phase/ channel	Metal point	Water point	Wood point	Fire point	Earth point
L.I.	L.I.-1	L.I.-2 −	L.I.-3	L.I.-5	L.I.-11+
ST	ST-45 −	ST-44	ST-43	ST-41+	ST-36
S.I.	S.I.-1	S.I.-2	S.I.-3 +	S.I.-5	S.I.-8 −
BL	BL-67 +	BL-66	BL-65 −	BL-60	BL-40
T.B.	T.B.-1	T.B.-2	T.B.-3 +	T.B.-6	T.B.-10 −
G.B.	G.B.-44	G.B.-43+	G.B.-41	G.B.-38−	G.B.-34
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point

+, tonification point; − , sedation point

According to the Five Phases (→ 8.2.5) the *shu*-stream points of the **Yin channels** correspond to the **Earth points** and ***yuan*-Source points** (→ 8.1.1), while those on the **Yang channels** correspond to the **Wood points**.

Location: counting from the periphery, they are the **third point of each channel**, the only exception being the G.B. channel (4th point on the channel). They are located proximal to the metacarpophalangeal or metatarsophalangeal joints, except for **KID-3** (posterior to the medial malleolus), **LIV-3** and **SP-3** (both located at the distal ends of the metatarsal bones) as well as **LU-9, HE-7** and **P-7** (all located on the wrist joint space).

Actions

They are considered to be the entry points for pathogenic factors; needling these points can strengthen the Defensive Qi (*wei qi*) (→ 1.1.4) and eliminate pathogenic factors from the respective channel. At the *shu*-stream points, the Qi flow is already somewhat deeper: the bed of the river is widening and deepening.

Indications

For *bi*-syndromes of the joints, especially due to Damp or pathogenic Wind. They are generally used for heaviness of the body as well as for *shaoyang* syndrome with intermittent symptoms.

- Points on the Yang channels are predominantly used for Exterior syndromes: they have a specific action on their respective channel.
- Points on the Yin channels are more commonly used for Interior syndromes; they strengthen and harmonise their respective *zang*-Organs (this is in part due to these points also being the *yuan*-source points).

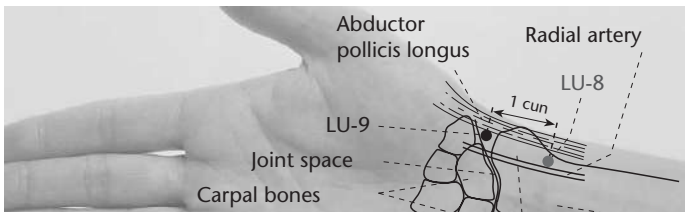

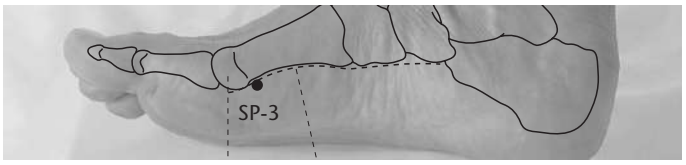

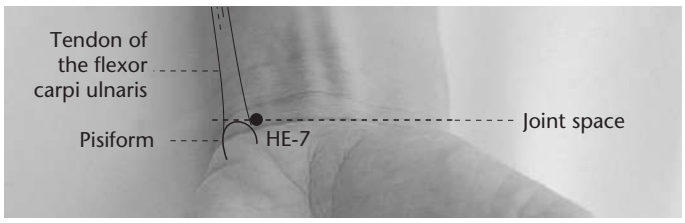

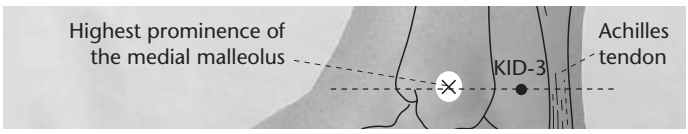

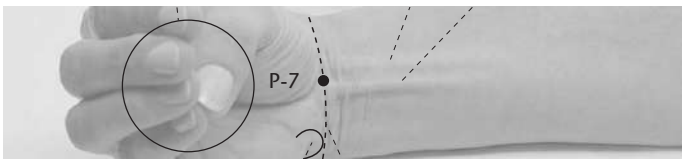

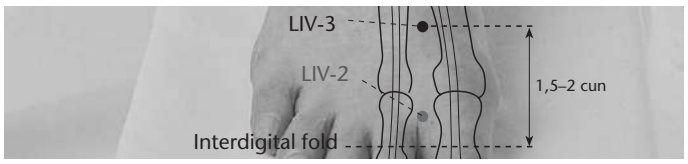

Ch08-F10028.qxd 2/23/08 4:40 PM Page 675

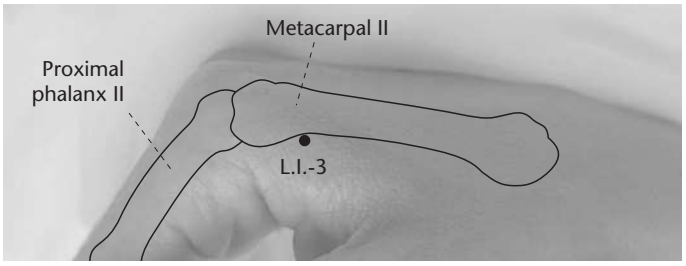

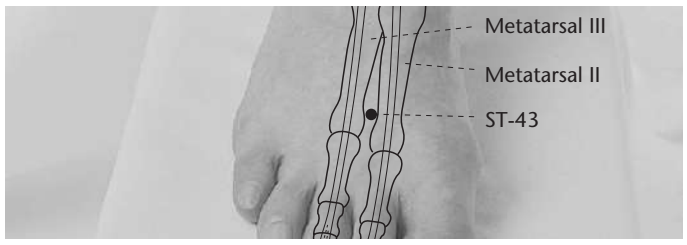

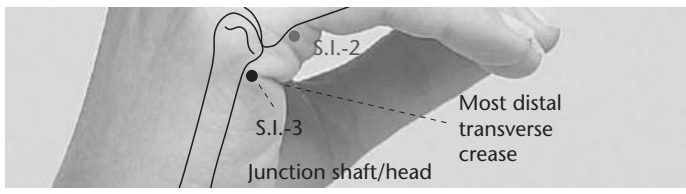

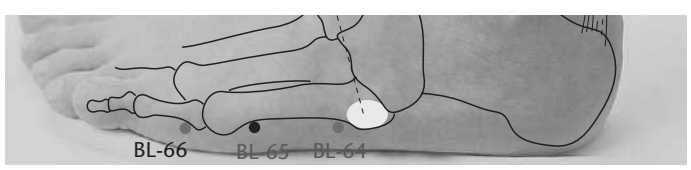

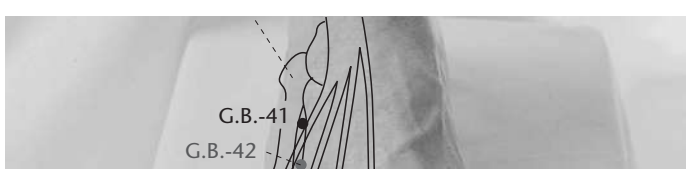

Jing-river points (*jing xue*)

***Shu*-transporting points of the Yin channels**					
Phase/ channel	Wood point	Fire point	Earth point	Metal point	Water point
LU	LU-11	LU-10	LU-9 +	LU-8	LU-5 −
SP	SP-1	SP-2 +	SP-3	SP-5 −	SP-9
HE	HE-9 +	HE-8	HE-7 −	HE-4	HE-3
KID	KID-1 −	KID-2	KID-3	KID-7 +	KID-10
P	P-9 +	P-8	P-7−	P-5	P-3
LIV	LIV-1	LIV-2 −	LIV-3	LIV-4	LIV-8 +
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point
***Shu*-transporting points of the Yang channels**					
Phase/ channel	Metal point	Water point	Wood point	Fire point	Earth point
L.I.	L.I.-1	L.I.-2 −	L.I.-3	L.I.-5	L.I.-11 +
ST	ST-45 −	ST-44	ST-43	ST-41 +	ST-36
S.I.	S.I.-1	S.I.-2	S.I.-3 +	S.I.-5	S.I.-8 −
BL	BL-67 +	BL-66	BL-65 −	BL-60	BL-40
T.B.	T.B.-1	T.B.-2	T.B.-3 +	T.B.-6	T.B.-10 −
G.B.	G.B.-44	G.B.-43+	G.B.-41	G.B.-38 −	G.B.-34
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point

+, tonification point; − , sedation point

According to the Five Phases (→ 8.2.5), the *jing*-river points of the **Yin channels** correspond to the **Metal points**, while those on the **Yang channels** correspond to the **Fire points**.
Location: they are located **between** the ***shu*-stream points** and the ***he*-sea points**. On the upper extremity, they are located either at, or distal or proximal to the wrist, while on the lower extremity they are located either at, or distal or proximal to the ankle (see figure on p. 676).

Actions

They divert external pathogenic factors away from joints, bones and tendons towards the Exterior. At the *jing*-river points the Qi flow becomes increasingly wider, slower, deeper and larger – thus the actions of these points are also less dynamic compared to the first three, more peripheral, *shu*-transporting points. At the *jing*-river points external pathogens have already penetrated deeper into the body, especially into the joints, bones and tendons.

Indications

- Dyspnoea, disorders of the throat such as inflammations, cough, disorders manifesting in the voice.
- Alternating fever and chills (especially points on the Yin channels)
- Also for *bi*-syndromes such as painful joint and tendon disorders.

Ch08-F10028.qxd 2/23/08 4:40 PM Page 676

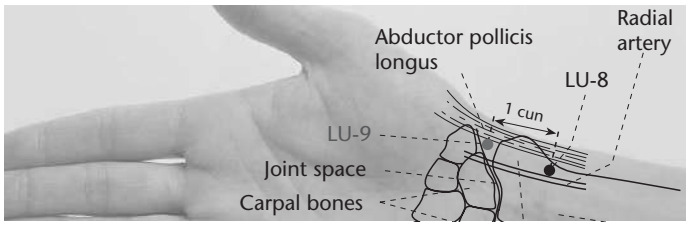

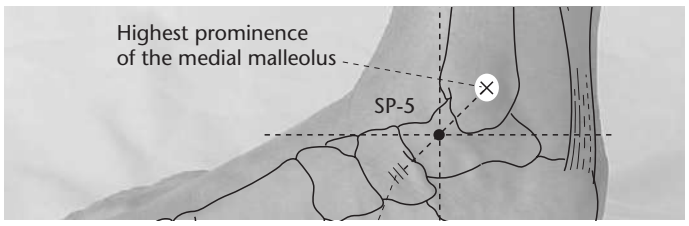

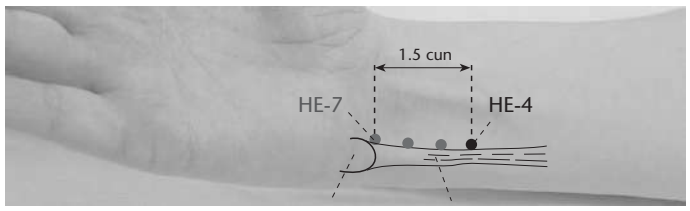

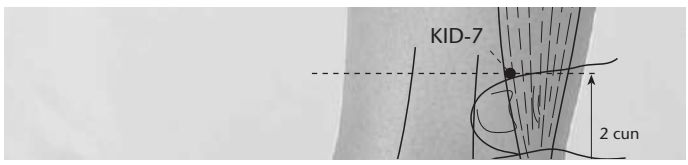

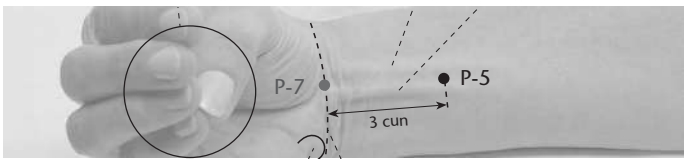

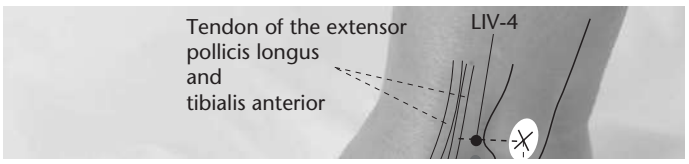

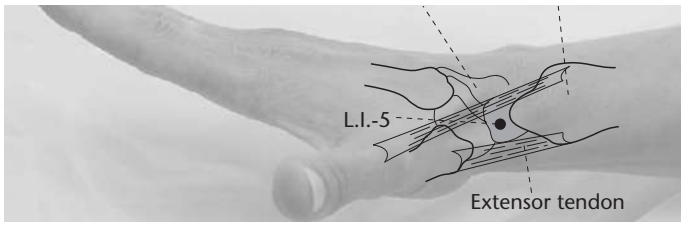

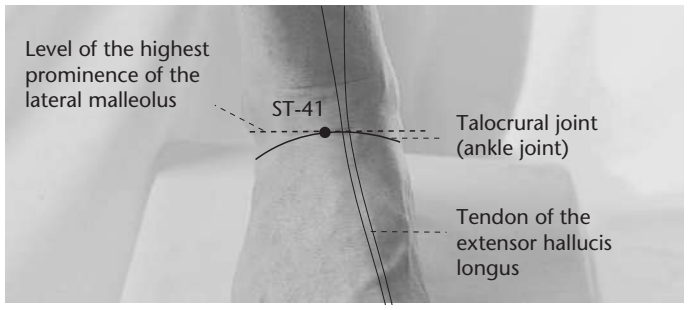

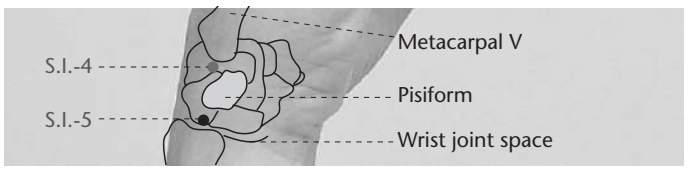

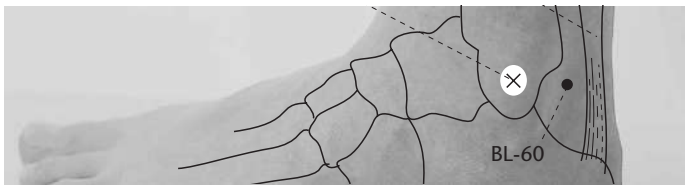

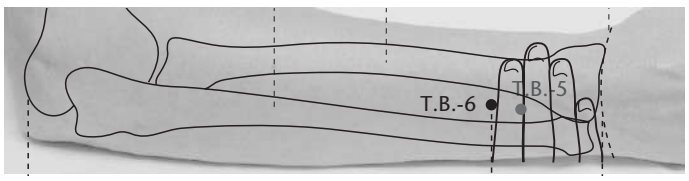

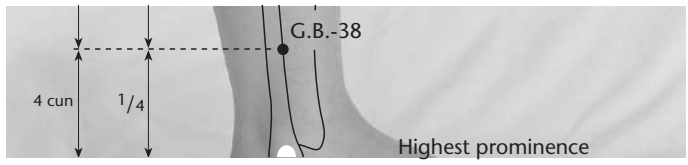

He-sea points (*he xue*)

Shu-transporting points of the Yin channels					
Phase/ channel	Wood point	Fire point	Earth point	Metal point	Water point
LU	LU-11	LU-10	LU-9 +	LU-8	LU-5 −
SP	SP-1	SP-2 +	SP-3	SP-5 −	SP-9
HE	HE-9 +	HE-8	HE-7−	HE-4	HE-3
KID	KID-1 −	KID-2	KID-3	KID-7 +	KID-10
P	P-9+	P-8	P-7−	P-5	P-3
LIV	LIV-1	LIV-2 −	LIV-3	LIV-4	LIV-8 +
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point
***Shu*-transporting points of the Yang channels**					
Phase/ Channel	Metal point	Water point	Wood point	Fire point	Earth point
L.I.	L.I.-1	L.I.-2 −	L.I.-3	L.I.-5	L.I.-11 +
ST	ST-45 −	ST-44	ST-43	ST-41 +	ST-36
S.I.	S.I.-1	S.I.-2	S.I.-3 +	S.I.-5	S.I.-8 −
BL	BL-67 +	BL-66	BL-65 −	BL-60	BL-40
T.B.	T.B.-1	T.B.-2	T.B.-3 +	T.B.-6	T.B.-10 −
G.B.	G.B.-44	G.B.-43+	G.B.-41	G.B.-38−	G.B.-34
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point

+, tonification point; − , sedation point

According to the Five Phases (→ 8.2.5) the *he*-sea points of the **Yin channels** correspond to the **Water points**, while those on the **Yang channels** correspond to the **Earth points**.

Location: on the upper extremity these points are located near the **elbow**, on the lower extremity near the **knee**.

Ch08-F10028.qxd 2/23/08 4:40 PM Page 677

Actions

- They descend counterflow Qi and stop diarrhoea. At the *he*-sea points the Qi flow is very wide, deep and slow, 'terminating in the deep sea' where the 'toxic cargo' can 'disappear' into the body unless treated adequately. External pathogens (*xie qi*) which have not been eliminated from the channel beforehand can here penetrate the Interior of the body and injure the *zangfu*-Organs.
- Diseases of the *fu*-Organs (here especially points on the lower extremities are indicated).

Indications

- Disorders of the *fu*-Organs, especially gastro-intestinal disorders with symptoms such as vomiting and diarrhoea
- Some *he*-sea points are also indicated for skin diseases, for example **LI-11, BL-40**.

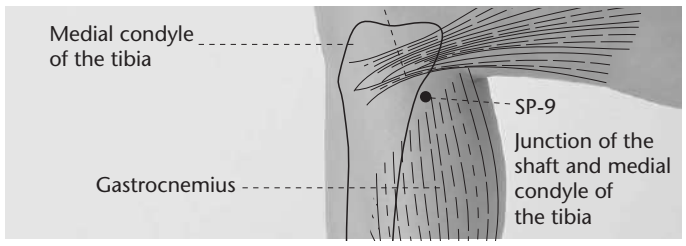

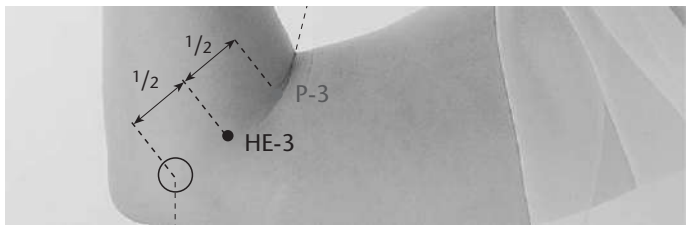

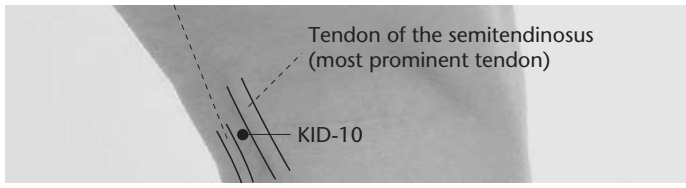

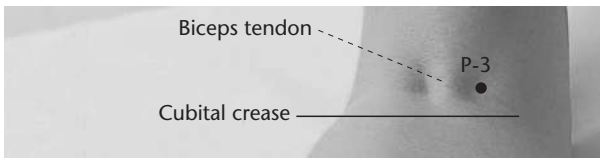

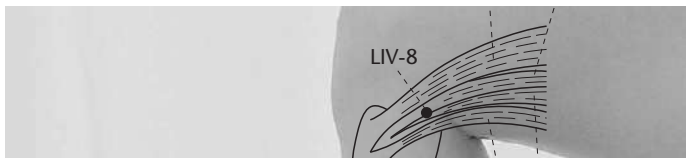

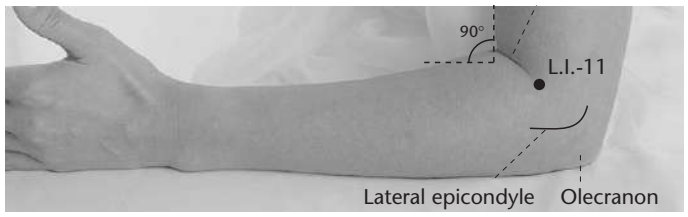

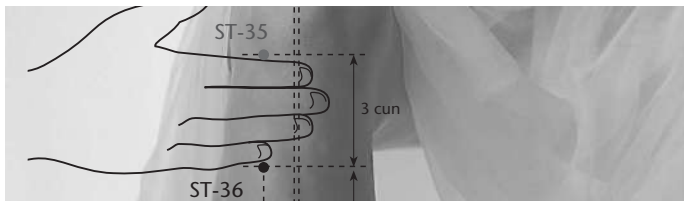

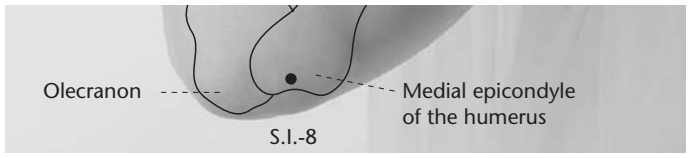

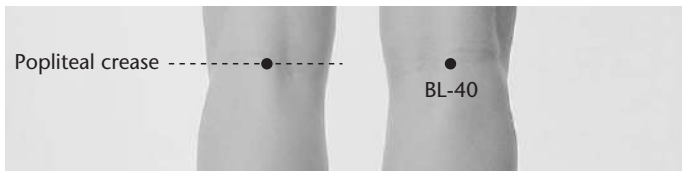

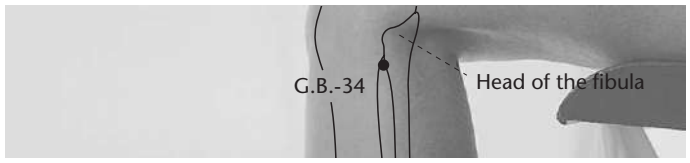

8.1.7 The Eight *Hui*-Meeting Points

Synonym: Gathering Points (Maciocia)

The Qi of a particular tissue, Organ system or substance gathers at the *hui*-meeting points and can be influenced through these points.

Ch08-F10028.qxd 2/23/08 4:40 PM Page 678

Points and clinical application

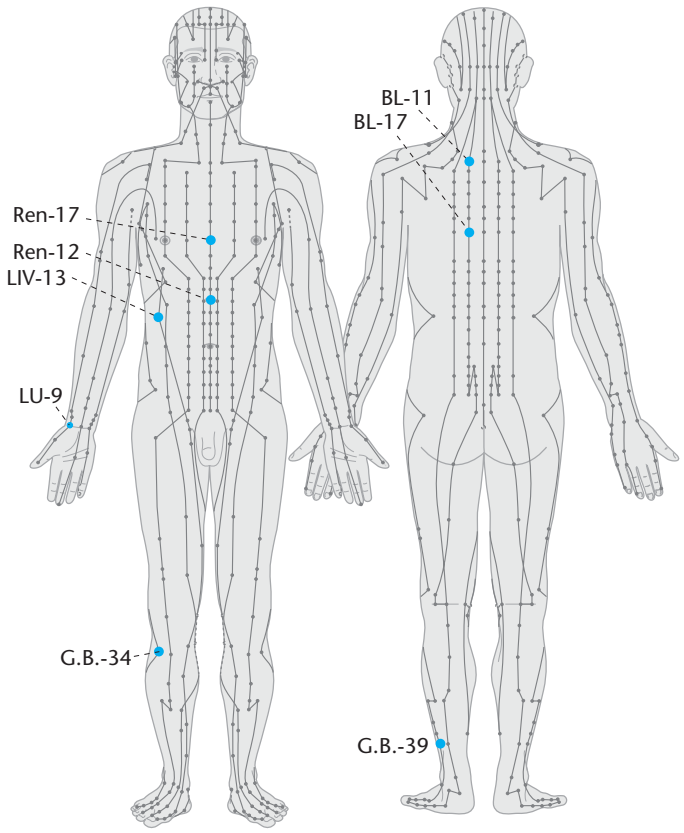

The Eight *hui*-Meeting Points		
Point	Tissue/organ system/substance	Indications
LIV-13 (*zhangmen*)	*zang*-Organs	Disorders of the *zang*-Organs (tonifies the Spleen and indirectly all *zang*-Organs)
Ren-12 (*zhongwan*)	*fu*-Organs	Disorders of the *fu*-Organs, for example of the gastro-intestinal tract
Ren-17 (*danzhong or shanzhong*)	Qi, has a strong effect on the *zong qi*	Respiratory diseases and disorders of the Qi flow such as hiccups; controls the *zong qi* and therefore functioning of the Lungs and Heart
BL-17 (*geshu*)	Blood (*xue*)	Disorders of the Blood such as anaemia, Blood stasis, haemorrhagia and gynaecological disorders
G.B.-34 (*yangling-quan*)	Sinews	Disorders of the joints, sinews and muscles
LU-9 (*taiyuan*)	Blood vessels	Stimulates the peripheral Blood circulation for disorders of the blood vessels such as vasculitis and arteriosclerosis
BL-11 (*dazhu*)	Bones	Disorders of the Bones such as shoulder and spine problems, joint and bone pain
G.B.-39 (*xuanzhong*)	Marrow	Disorders of the Bones, the Marrow and Brain (nerves)

8.1.8 Opening Points of the Eight Extraordinary Vessels

Synonyms: Confluent points (Deadman et al 1998), Master points (Pirog 1996)

The table below shows the opening points of the eight extraordinary vessels. For more detail on their origin and pathways → 1.7, Chapter 5.

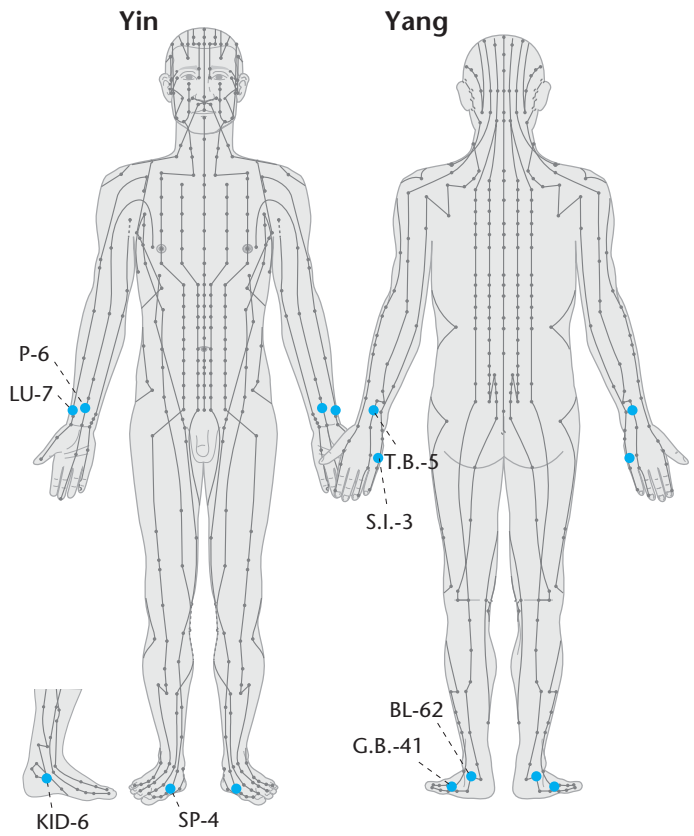

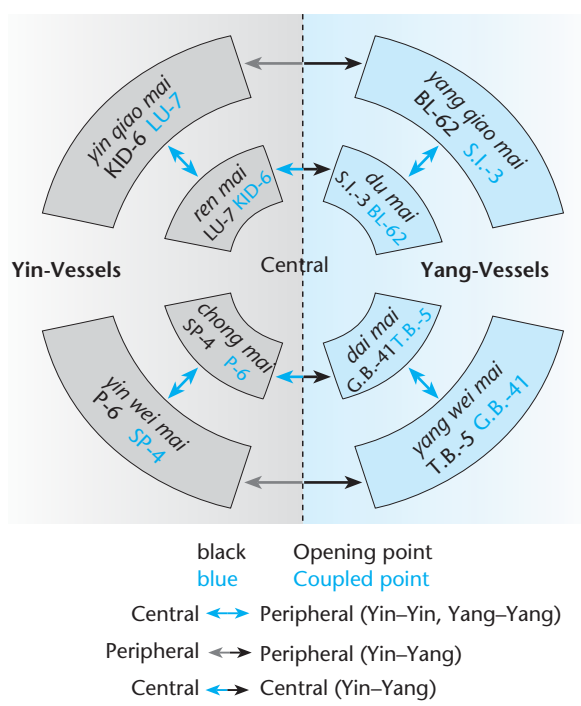

black Opening point
blue Coupled point
Central ↔ Peripheral (Yin–Yin, Yang–Yang)
Peripheral ↔ Peripheral (Yin–Yang)
Central ↔ Central (Yin–Yang)

Ch08-F10028.qxd 2/23/08 4:40 PM Page 679

Points

Pairs	Extraordinary Vessel	Opening point	Coupled point (opening point of the paired vessel)	Supported body regions
Pair 1	*chong mai*	**SP-4** (*gongsun*)	**P-6** (*neiguan*)	Heart, chest, Stomach
	yin wei mai	**P-6** (*neiguan*)	**SP-4** (*gongsun*)	
Pair 2	*du mai*	**S.I.-3** (*houxi*)	**BL-62** (*shenmai*)	Occiput, shoulder, back
	yang qiao mai	**BL-62** (*shenmai*)	**S.I.-3** (*houxi*)	
Pair 3	*dai mai*	**G.B.-41** (*zulinqi*)	**T.B.-5** (*waiguan*)	Temples, ears, lateral aspects of the body
	yang wei mai	**T.B.-5** (*waiguan*)	**G.B.-41** (*zulinqi*)	
Pair 4	*ren mai*	**LU-7** (*lieque*)	**KID-6** (*zhaohai*)	Face, throat, chest, Lungs, abdomen
	yin qiao mai	**KID-6** (*zhaohai*)	**LU-7** (*lieque*)	

Clinical application

Various approaches exist for needling the opening/coupled points of the extraordinary vessels. The following method **for 'opening' or regulating the extraordinary vessels** has been used with good therapeutic results in clinical practice:

- Needle the **opening point** of the relevant channel first, then needle the **contralateral coupled point** (opening point of the paired extraordinary vessel). This will mobilise and regulate the Qi of the extraordinary vessel, making it more dynamic.
- During a subsequent session, needle **either the *xi*-cleft point** of the previously treated extraordinary vessel (→ 8.1.3) or add **other channel points of the vessel** (see channel pathways, chapter 5) during the same or subsequent session. This will further support the dynamics of the Qi flow in the vessel.

In addition Maciocia (1989) and Kirschbaum (1995) suggest the following procedure for 'opening' the extraordinary vessels:

- **In women:** First needle the opening point on the *right* side with even method, then the coupled point on the *left* side with even method. Afterwards other points on the respective vessel may be added.
- **In men:** First needle the opening point on the *left* side with even method, then the coupled point on the *right* side with even method. Afterwards other points on the respective vessel may be added.
- **Needle retention:** the needles should be retained for 20–25 minutes and should be removed in the opposite order.

8.1.9 Lower *He-S*ea Points

The lower *he-s*ea points are used as the main points for disorders of their pertaining *fu*-Organ.

Points and clinical application

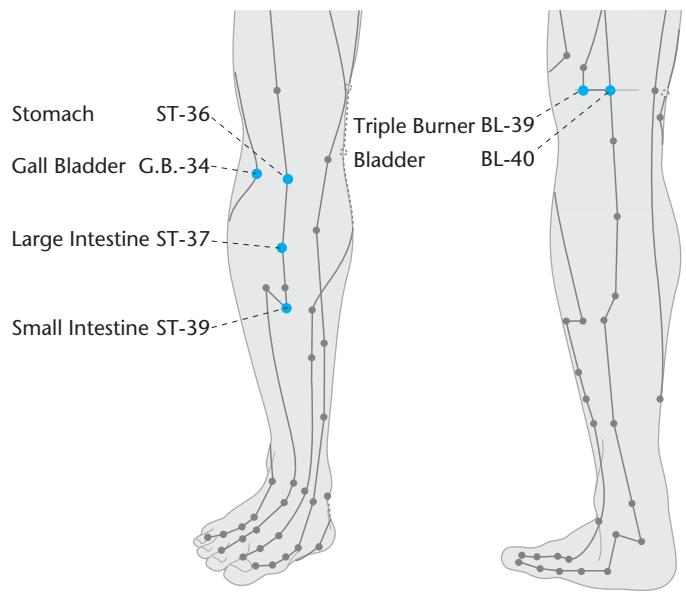

- **ST-36** (*zusanli*): Lower *he-s*ea point of the Stomach; indication: loss of appetite, epigastric pain and fullness, acid reflux, abdominal pain, constipation, diarrhoea
- **ST-37** (*shangjuxu*): Lower *he-s*ea point of the Large Intestine; indication: appendicitis, diarrhoea
- **ST-39** (*xiajuxu*): Lower *he-s*ea point of the Small Intestine; indication: diarrhoea, abdominal pain
- **G.B.-34** (*yanglingquan*): Lower *he*-sea point of the Gall Bladder; indication: cholecystitis, vomiting
- **BL-40** (*weizhong*): Lower *he*-sea point of the Bladder; indication: urinary incontinence, urinary retention
- **BL-39** (*weiyang*): Lower *he*-sea point of the Triple Burner; indication: urinary incontinence, urinary retention

8.1.10 Meeting Points (*Jiaohui* Points)

Synonym: Intersecting points (Practical Dictionary)

Meeting points (*jiaohui* points) represent intersections between channels and vessels. Needling them allows more than one channel to be influenced and widens the range of actions of a given point. By the same token, the number of needles may be reduced while maintaining a good therapeutic effect.

Ch08-F10028.qxd 2/23/08 4:41 PM Page 680

	Lung	Large Intestine	Stomach	Spleen	Heart	Small Intestine	Bladder	Kidney	Pericardium	Triple Burner	Gall Bladder	Liver	*Du mai*	*Ren mai*	*Yang wei mai*	*Yin wei mai*	*Yang qiao mai*	*Yin qiao mai*	*Chong mai*	*Dai mai*
LU 1	■			■																
L.I. 14		■				▨	▨								▨					
L.I. 15		■															■			
L.I. 16		■															■			
L.I. 20		■	■																	
ST 1			■										▨	■			■			
ST 2			■														▨			
ST 3			■														■			
ST 4		■	■										▨	▨			■			
ST 5			■								▨									
ST 6			■								▨									
ST 7			■								■									
ST 8			■								■				■					
ST 9			■								■									
ST 12		■	■			■				■	■									
ST 30			■								▨								■	
SP 6				■				■				■						▨		
SP 12				■						▨		■				■				
SP 13				■								■				■				
SP 15				■												■				
SP 16				■												■				
S.I. 10						■	■								■		■			
S.I. 12		■				■				■	■									
S.I. 18						■				■										
S.I. 19						■				■	■									
BL 1			■			■	■			▨	▨		■				■	■		
BL 11						■	■			▨	▨		▨							
BL 12							■						■							
BL 23							■						▨							
BL 31							■				■									
BL 32							■				■									
BL 33							■				■									
BL 34							■				■									
BL 41						■	■													
BL 59							■										▨			
BL 61							■										■			
BL 62							■										▨			
BL 63							■								■					
KID 2								■										▨		
KID 6								■										▨		
KID 8								■										▨		
KID 9								■								■				
KID 11								■											■	
KID 12								■											■	
KID 13								■											■	
KID 14								■											■	
KID 15								■											■	
KID 16								■											■	
KID 17								■											■	
KID 18								■											■	

	Lung	Large Intestine	Stomach	Spleen	Heart	Small Intestine	Bladder	Kidney	Pericardium	Triple Burner	Gall Bladder	Liver	*Du mai*	*Ren mai*	*Yang wei mai*	*Yin wei mai*	*Yang qiao mai*	*Yin qiao mai*	*Chong mai*	*Dai mai*
KID 19								■											■	
KID 20								■											■	
KID 21								■											■	
P 1									■	▨	■	■								
T.B. 13										■					▨					
T.B. 15										■	■				■					
T.B. 17										■	■									
T.B. 20						▨				■	■									
T.B. 22						■				■	■									
G.B. 1						■				■	■									
G.B. 3			■							■	■									
G.B. 4			■							■	■									
G.B. 5		■	■							■	■									
G.B. 6		■	■							■	■									
G.B. 7							■				■									
G.B. 8							■				■									
G.B. 9							■				■									
G.B. 10							■				■									
G.B. 11						▨	■			▨	■									
G.B. 12							■				■									
G.B. 13											■				■					
G.B. 14		▨	▨							▨	■				■					
G.B. 15							■				■				■					
G.B. 16											■				■					
G.B. 17											■				■					
G.B. 18											■				■					
G.B. 19											■				■					
G.B. 20										▨	■				■		■			
G.B. 21			▨							■	■				■					
G.B. 23							▨				■									
G.B. 24				■							■				▨					
G.B. 26											■									■
G.B. 27											■									■
G.B. 28											■									■
G.B. 29											■						■			▨
G.B. 30							■				■									
G.B. 35											■						▨			
LIV 13											■	■								▨
LIV 14				■								■				■				
Du 1								■			■		■	■						
Du 13							■						■							
Du 14		■	■			■	■			■	■		■							
Du 15													■		■					
Du 16													■		■		▨			
Du 17							■						■							
Du 20							■			■	■	■	■							
Du 24			■				■						■							
Du 26		■	■										■							
Du 28			▨										■	■						
Ren 1													■	■					■	

Ch08-F10028.qxd 2/23/08 4:41 PM Page 681

	Lung	Large Intestine	Stomach	Spleen	Heart	Small Intestine	Bladder	Kidney	Pericardium	Triple Burner	Gall Bladder	Liver	*Du mai*	*Ren mai*	*Yang wei mai*	*Yin wei mai*	*Yang qiao mai*	*Yin qiao mai*	*Chong mai*	*Dai mai*
Ren 2												■								
Ren 3				■				■				■								
Ren 4				■				■				■								
Ren 7								▨											■	
Ren 10			▨	■																
Ren 12			■			■				▨										
Ren 13			■			■														
Ren 17				■		■		■		■										
Ren 22																■				
Ren 23																■				
Ren 24		■	■										■							

▨ only mentioned by some authors

8.1.11 The Gao Wu Command Points

Synonym: none known

These points have the ability to affect certain regions of the body. By combining these points with local and other specific points they can increase the therapeutic effect in a particular area of the body.

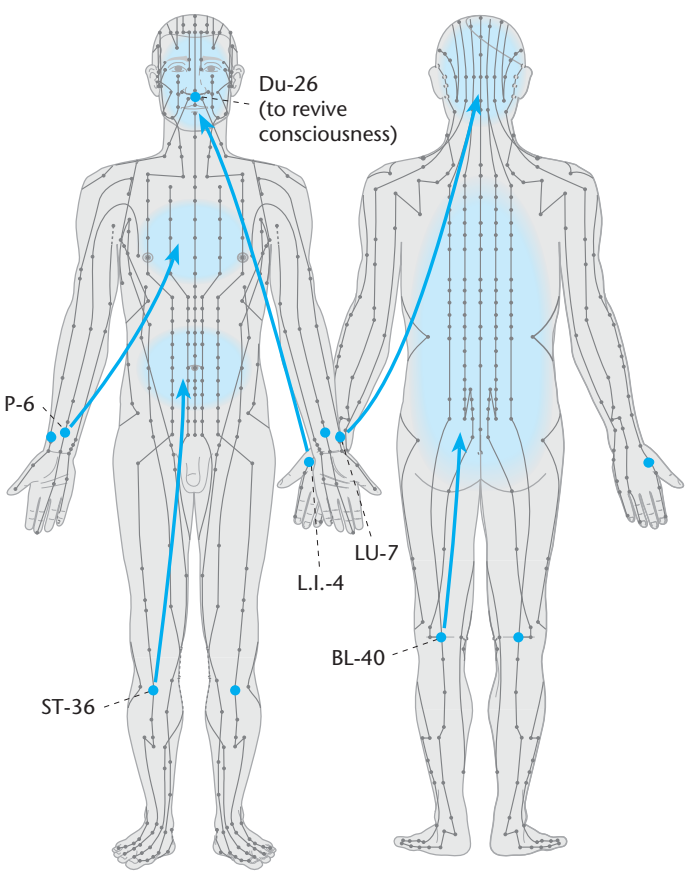

Points and clinical application

- **ST-36** (*zusanli*): for all disorders of the abdomen
- **L.I.-4** (*hegu*): for disorders of the face and mouth
- **LU-7** (*lieque*): for disorders of the occiput
- **BL-40** (*weizhong*): for disorders of the back and lumbar region
- **P-6** (*neiguan*): for disorders of the thorax
- **Du-26** (*renzhong*): for restoring consciousness (The last two points were added later, → Deadman et al 1998)

8.1.12 The Window of Heaven Points

In Chinese classical literature there are only a few references to these points. In modern times, both Ross (1995) and Deadman et al (1998) mention this group of points.

Points

LU-3 (*tianfu*) **L.I.-18** (*futu*) **P-1** (*tianchi*)
T.B.-16 (*tianyou*) **S.I.-16** (*tianchuang*) **S.I.-17** (*tianrong*)
ST-9 (*renying*) **BL-10** (*tianzhu*) **Ren-22** (*tiantu*)
Du-16 (*fengfu*)

Location: Eight of the Window of Heaven points are located in the neck region (exceptions are **LU-3** on the upper arm and **P-1** on the thorax). This might provide a clue about the action of these points in regulating the Qi flow between the head and the body.

Comment: According to Deadman, the Ming dynasty physician Ma Shi remarked that **S.I.-17** should be replaced by **G.B.-9** (*tianchong*). If that was the case all six Yang channels running to the head would be represented by a Window of Heaven point.

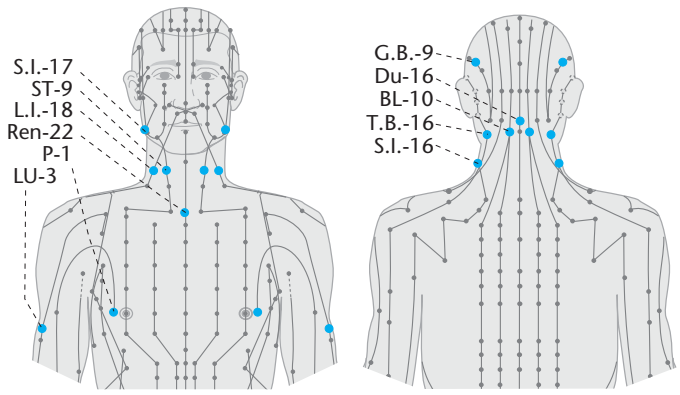

Clinical application

- For disorders of the Qi flow:
  - **LU-3, L.I.-18, Ren-22** for counterflow Qi causing cough, wheezing, etc.
  - **P-1** for counterflow Lung Qi with copious Phlegm
  - **ST-9** for counterflow Lung and Stomach Qi
  - **BL-10** treats dizziness, headaches and stiffness of the occiput, etc.
  - **Du-16** for internal Liver Wind

Ch08-F10028.qxd 2/23/08 4:41 PM Page 682

- For goitre, swellings, pain and Qi stagnation in the occiput and neck: especially indicated if these occur locally
- For disorders with an acute onset, for example
  - **L.I.-18** for acute aphonia
  - **ST-9** for acute diarrhoea
  - **S.I.-16** for acute aphonia, for example following a stroke
  - **BL-10** for epilepsy
  - **T.B.-16** for acute hearing loss
  - **Ren-22** for acute dyspnoea
  - **Du-16** for acute aphasia following a stroke
- For psychological disorders:
  - **LU-3** for somnolence, grief, disorientation, forgetfulness, insomnia
  - **S.I.-16** for manic agitation and bipolar disorder
  - **BL-10** for mania, hallucinations, epilepsy, childhood convulsions
  - **T.B.-16** for vivid dreaming
  - **Du-16** for mania, grief, anxiety with anxiety-induced palpitations
- For disorders of the sensory organs:
  - **LU-3** for nose bleeding, blurry vision, myopia
  - **ST-9** for blurry vision
  - **S.I.-16** for deafness, tinnitus, ear pain
  - **S.I.-17** for tinnitus and deafness
  - **BL-10** for eye pain, blurry vision, excessive tearing, speech disorders, blocked nasal passages, loss of sense of smell
  - **T.B.-16** for hearing disorders, visual disorders, eye pain, excessive tearing, loss of sense of smell, blocked nasal passages
  - **Du-16** for speech disorders (for example following a stroke), blurry vision, nosebleeds
  - **Ren-22** for speech disorders

8.1.13 Points of the Four Seas

These points have a particular, supportive effect on their respective 'Sea'.

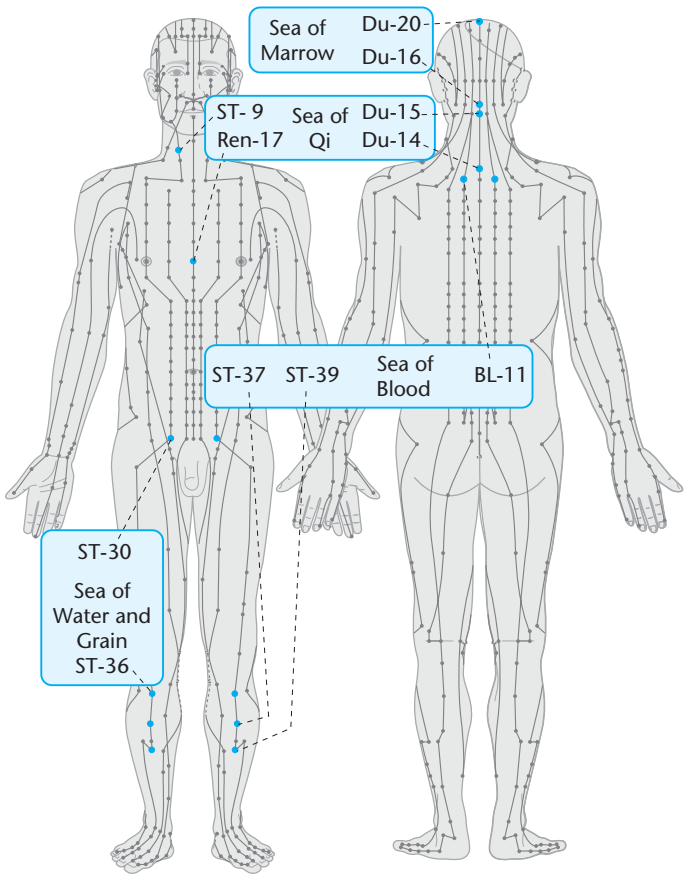

Sea of Qi	**ST-9** (*renying*), **Ren-17** (*danzhong or shanzhong*), **Du-15** (*yamen*), **Du-14** (*dazhui*)
*Ling Shu**: 'When the Sea of Qi is in excess there is fullness in the chest, urgent breathing and a red complexion. When the Sea of Qi is insufficient, there is scanty energy, insufficient for speech.'	
Sea of Blood	**BL-11** (*dazhu*), **ST-37** (*shangjuxu*), **ST-39** (*xiajuxu*)
*Ling Shu**: 'The Penetrating vessel is the sea of the twelve channels. In the upper it is conveyed to BL-11 (*dazhu*), and in the lower it emerges at both the ST-37 (*shangjuxu*) and ST-39 (*xiajuxu*) …When the Sea of Blood is in excess, then one has a sensation of the body being big; one feels disquiet, but does not know what the illness is; when the Sea of Blood is insufficient, one has the sensation of one's body being small; one feels reduced but does not know what the illness is.'	
Sea of Water and Grain	**ST-30** (*qichong*) as upper point, **ST-36** (*zusanli*) as lower point
*Ling Shu**: 'When the Sea of Water and Grain is in excess, there is abdominal fullness, and when it is deficient, there is hunger with inability to eat.'.	
Sea of Marrow	**Du-20** (*baihui*), **Du-16** (*fengfu*)
*Ling Shu**: 'When the Sea of Marrow is in excess, there is lightness of the body and much strength, and a person's feeling of 'self' exceeds the normal level; when the Sea of Marrow is insufficient there is a whirling sensation of the brain, dizziness, tinnitus, pain of the lower legs, impairment of vision, indolence and desire to sleep.'	

* Quotes from the *Ling Shu* according to Deadman et al. (1998).

Ch08-F10028.qxd 2/23/08 4:41 PM Page 683

8.1.14 The 12 Heavenly Star Points of Ma Dan Yang

Ma Dan Yang, a famous physician of the Jin dynasty, considered these initial 11 points as the most important acupuncture points. It was the physician Xu Feng who later added **LIV-3** as the 12th point.

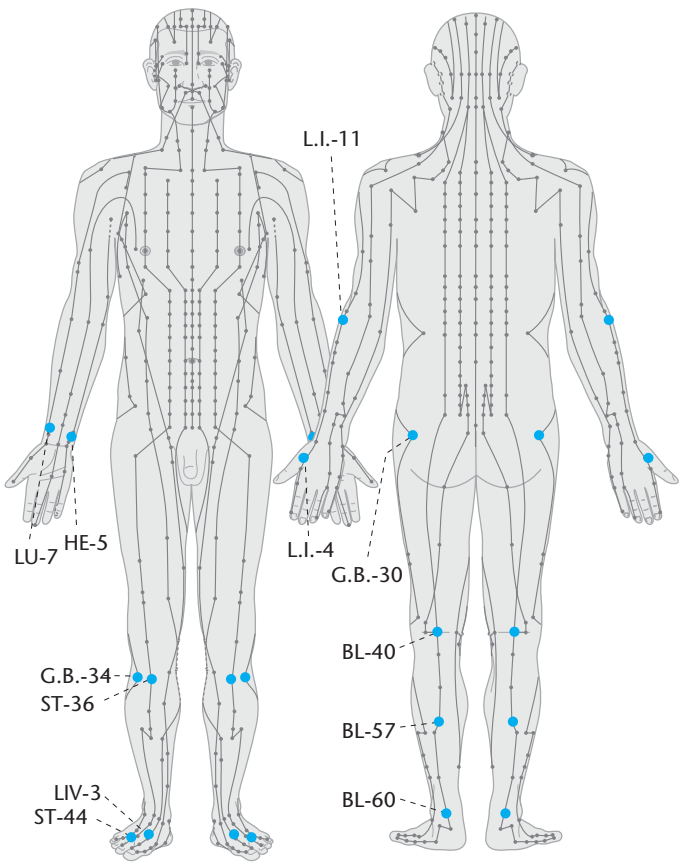

LU-7 (*lieque*) **L.I.-4** (*hegu*) **L.I.-11** (*quchi*)
ST-36 (*zusanli*) **ST-44** (*neiting*) **HE-5** (*tongli*)
BL-40 (*weizhong*) **BL-57** (*chengshan*) **BL-60** (*kunlun*)
G.B.-30 (*huantiao*) **G.B.-34** (*yanglingquan*) **LIV-3** (*taichong*)

8.1.15 The 13 Ghost Points of Sun Si Miao

Sun Si Miao, a famous physician of the Tang Dynasty, applied these 13 points (also called demon points) for treating disorders that today would be considered severe, manic, psychological disorders and/or epilepsy. Each *Sun Si Miao* Ghost point has an alternative Chinese name containing the component 'demon' or 'ghost' (see list according to Deadman et al. 2000). According to Deadman et al., some authors assumed that *guixin* is **LU-9** (*taiyuan*) rather than **P-7** (*daling*), while *guilu* is believed to be **P-5** (*jianshi*) or **P-8** (*laogong*) rather than **BL-62** (*shenmai*). There also exist different versions of this list of points; for example the physician Gao Wu omitted **BL-62, Du-23, Ren-1*** (in the original text of *Sun Si Miao: yumentou/yinxiafeng*) and **L.I.-11**, adding **Du-24, ST-17, LIV-2 and G.B.-34** instead.

	Name	Alternative name	Translation
Du-26	*renzhong*	*guigong*	Ghost Palace
Du-16	*fengfu*	*guizhen*	Ghost Pillow
Du-23	*shangxing*	*guitang*	Ghost Hall
Ren-24	*chengjiang*	*guishi*	Ghost Market
LU-11	*shaoshang*	*guixin*	Ghost Faith
L.I.-11	*quchi*	*guitui*	Ghost Leg
ST-6	*jiache*	*guichuang*	Ghost's Bed
SP-1	*yinbai*	*guilei*	Ghost Fortress
P-7	*daling*	*guixin*	Ghost Heart
P-8	*laogong*	*guicu*	Ghost Cave
BL-62	*shenmai*	*guilu*	Ghost Path
Ex-HN-11	*haiquan*	*guifeng*	Ghost Seal
Ren-1*	*huiyin*	*guicang*	Ghost Store

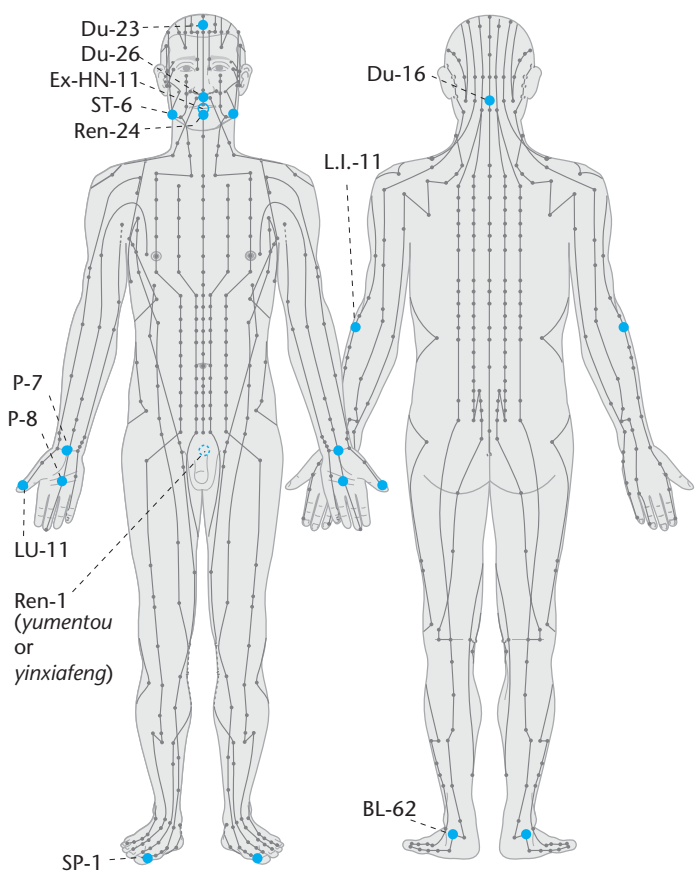

* *Yumentou* (in women) and *yinxiafeng* (in men): these two extra points approximately correspond to Ren-1 (*huiyin/guicang*) (Ghost Store).

Ch08-F10028.qxd 2/23/08 4:41 PM Page 684

8.1.16 Entry/Exit Points

While in modern Chinese literature entry/exit points are not mentioned as a separate point category, they are listed as such in some Western literature (for example Jarrett 2003, Hicks et al. 2004, Pirog 1996). However, Hicks et al. (→ 2004, p. 250) postulate that original Chinese sources exist for these points. The concept of entry/exit points is based on a continuous flow of Qi from one channel to the next, the sequence corresponding to the Organ clock (→ 1.1.4). Shunts are suggested in order to facilitate the flow of Qi from the end of one channel to the beginning of the next. At an entry point the Qi therefore flows from the preceding channel into the following one, while at the exit points it leaves the current channel in order to enter the next one.
However, it should be noted that the entry/exit points do not always correspond to the first or last point on a channel. In the following table these points are marked by grey shading.

Channel (sequence according to Organ clock)	Time of maximum Qi flow	Entry/exit points in sequential order		
LU (LU-1–LU-11)	3–5am	LU-1 (*zhongfu*)	➡	LU-7 (*lieque*)
				⬇
L.I. (L.I.-1–L.I.-20)	5–7am	L.I.-20 (*yingxiang*)	⬅	L.I.-4 (*hegxu*)
		⬇		
ST (ST-1–ST-45)	7–9am	ST-1 (*chengqi*)	➡	ST-42 (*chongyang*)
				⬇
SP (SP-1–SP-21)	9–11am	SP-21 (*dabao*)	⬅	SP-1 (*yinbai*)
		⬇		
HE (HE-1–HE-9)	11am–1pm	HE-1 (*jiquan*)	➡	HE-9 (*shaochong*)
				⬇
S.I. (S.I.-1–S.I.-19)	1–3pm	S.I.-19 (*tinggong*)	⬅	S.I.-1 (*shaoze*)
		⬇		
BL (BL-1–BL-67)	3–5pm	BL-1 (*jingming*)	➡	BL-67 (*zhiyin*)
				⬇
KID (KID-1–KID-27)	5–7pm	KID-22 (*bulang*)	⬅	KID-1 (*yongquan*)
		⬇		
P (P-1–P-9)	7–9pm	P-1 (*tianchi*)*	➡	P-8 (*laogong*)*
				⬇
T.B. (T.B.-1–T.B.-23)	9–11pm	T.B.-22 (*erheliao*)	⬅	T.B.-1 (*guanchong*)
		⬇		
G.B. (G.B.-1–G.B.-44)	11pm–1am	G.B.-1 (*tongziliao*)	➡	G.B.-41 (*zulinqi*)
				⬇
LIV (LIV-1–LIV-14)	1–3am	LIV-14 (*qimen*)	⬅	LIV-1 (*dadun*)

* In women P-2 is often substituted for P-1 due to its anatomical location close to the breast.

Ch08-F10028.qxd 2/23/08 4:41 PM Page 685

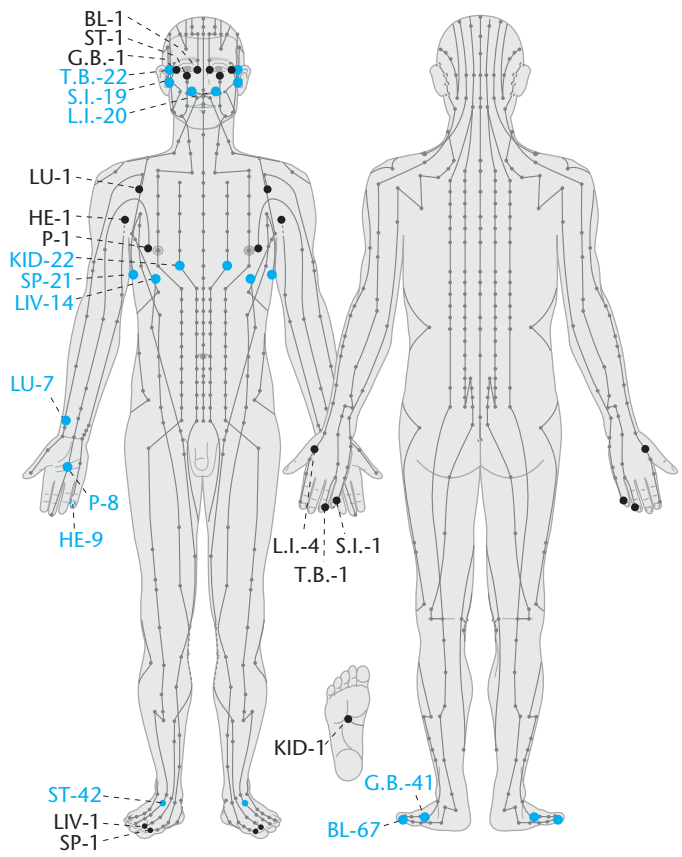

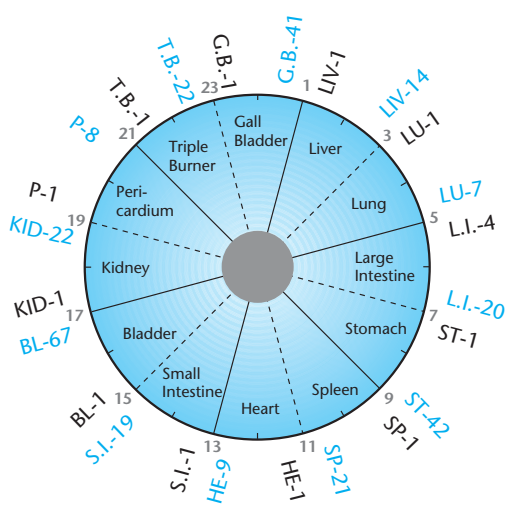

Entry/exit block and therapy

Pulse diagnosis can reveal the partial or complete block between entry and exit points. An entry/exit block will inhibit the Qi flow from one channel to the next. In order to restore the blocked flow, the exit point of the blocked channel and the entry point of the following channel are needled at the same time. According to Hicks et al. (2004), further exit and entry points preceding or following the affected channel can be added in order to regulate the Qi flow over a greater distance. Needling should always be **bilateral**, even if the disorder occurs only on one side of the body. If it is not only the shunt between two channels, but the circulation in a channel as a whole that is affected, the flow in the whole channel can be strengthened and stimulated by simultaneous needling of both its entry point and its exit point.

Pirog (1996) compares the entry/exit points to input and output valves regulating the flow in a series of pipes (the channels) by 'opening' or 'closing' them.

- Tonifying needling techniques at the entry point (opening the input valve) will increase the flow into the channel. For example: tonifying **LU-1** opens the input valve. This will increase Qi flow in the LU channel for deficiency in this channel.
- Tonifying needling techniques at the exit point (opening the output valve) will decrease the flow in the channel. For example: Tonifying **LIV-14** opens the output valve: excess in the LIV channel will be drained.
- Reducing needling techniques at the entry point (closing the input valve) will reduce the flow into the channel. For example: Reducing **LU-1** closes the input valve and reduces the flow from the LIV channel into the LU channel, for example if there is pre-existing excess in the LU channel.
- Reducing needling techniques at the exit point (closing the output valve) will strengthen the flow in the channel. For example: Reducing **LIV-14** closes the output valve and prevents a further loss of Qi if there is a pre-existing deficiency in the LIV channel.

8.2 Strategies for Point Selection

8.2.1 Local, Adjacent and Distal Points

Local points

Local points are located in the immediate vicinity of the affected region. Each point that is tender with pressure (*ashi* point) can be considered to be a local point.

Adjacent points

Adjacent points are located near the affected or painful region.

Distal points

Distal points, despite being located at a distance from the affected area, still have a therapeutic effect, either directly or by being connected to the affected area by a channel (→ 1, also → 8.2.2, selection of distal points). The most effective and dynamic distal points are located distal to the knee (for the leg) or the elbow (for the arm). Generally the distal points of the foot channels tend to be more dynamic and effective than those on the arm channels.

Ch08-F10028.qxd 2/23/08 4:41 PM Page 686

Needling technique: In acute conditions or with severe pain the distal points should be needled with a reducing technique in order to activate the Qi flow more strongly in the blocked channels.

Stimulation of distal points: For a limited range of motion accompanied by pain the relevant distal point is strongly stimulated with a reducing technique while the patient moves the affected joint.

Enhancing the therapeutic effect: The therapeutic effect can be enhanced by combining distal points on the hand channels with those on the foot channels or by combining distal points with relevant local and adjacent points (→ 8.3.1).

8.2.2 Point Selection According to Affected Channel

Here the method of selecting points is based on the channel system (→ 1):

- Selecting points on the channel which **traverses the affected area**
- Selecting points on the channel that is **Interiorly–Exteriorly pertaining** (→ 1.2) to the affected channel
- Selecting points from the channel which is connected to the affected channel by **hand–foot pairing** (for example, *taiyang* axis, etc. → 1.2.3)

Differential selection of distal points

The following possibilities for selecting distal points are, among others, based on course material about the 'one point therapy' by R. Thambirajah (1990, 1991), R. Tan (2003) and my own clinical experience.

Selection of distal points according to corresponding areas

The distal points can be selected according to corresponding areas. The affected area is carefully examined, then the corresponding area is needled contralaterally (→ Fig.).

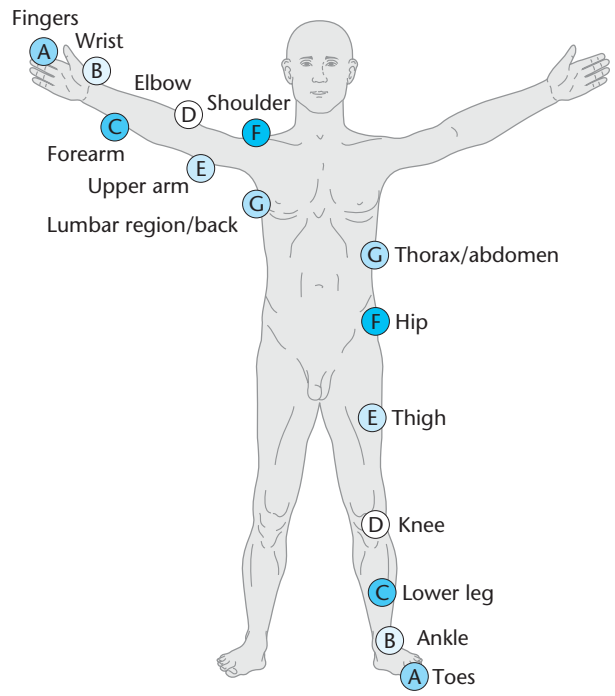

Selection of distal points according to Interiorly–Exteriorly connected channels

Here points are selected on the Interiorly–Exteriorly connected channel (Yin/Yang). The affected area is carefully located, then the corresponding point (mirror point) is needled with reducing technique on the contralateral side on the Interiorly–Exteriorly pertaining channel. Another specific example for this method of point selection is the *yuan–luo* combination (→ 8.3.3).

Example: For a disorder affecting the ST channel (foot *yangming*) select a point on the Interiorly–Exteriorly pertaining SP channel (foot *taiyin*). For example, for knee pain at **ST-35**, needle **SP-9** contralaterally (→ Fig.).

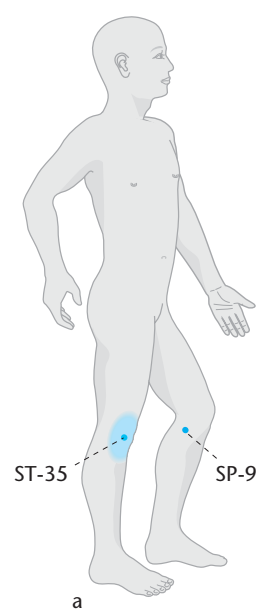

a

Ch08-F10028.qxd 2/23/08 4:41 PM Page 687

Selection of distal points according to the six great channels (*liu jing*, hand–foot pairing, → 1.2.3)

The point selection with this method is based on the six great channels (hand–foot pairing: *taiyin, taiyang*, etc., → 1.2.3). Disorders affecting the area of the hand channel are therefore treated with points on the corresponding foot channel and vice versa.

Example: Disorders in the area of the hand *taiyin* channel (LU channel) can be balanced by contralateral needling of points on the foot *taiyin* channel (SP channel). For example, shoulder pain with limited range of motion and maximum pain near **LU-1** (*zhongfu*) can be improved by contralateral needling of **SP-9** (*yinlingquan*) with reducing technique.

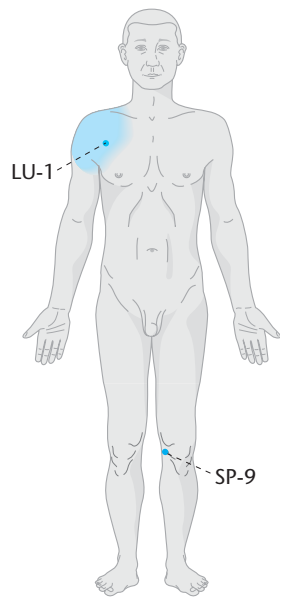

b

Selection of distal points according to the Organ clock (→ 1.4)

Distal points are selected on the channel that lies chronologically opposite to the affected channel on the Organ clock (→ 1.4). With this method, hand channels will treat disorders affecting the foot channels and vice versa.

Example: For a disorder affecting the G.B. channel (foot *shaoyang*), needle a point on the HE channel (hand *shaoyin*) located opposite the G.B. channel on the Organ clock. Needling can be contralateral or ipsilateral.

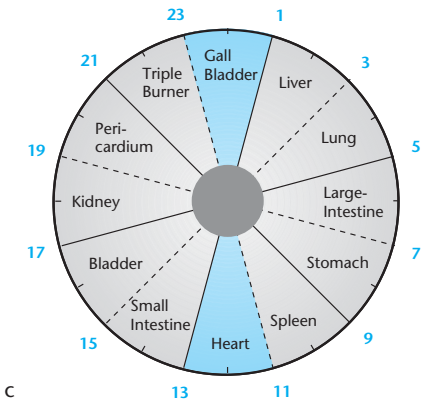

c

8.2.3 Point Selection Based on Symptoms

Points can also be selected based on traditional or empirical therapeutic experiences. While this method cannot replace a root treatment based on a differential diagnosis, it can be very helpful in situations where quick action is required.

Examples:

- **Du-14** or **L.I.-11** for fever
- **ST-36, Ren-12** for acute gastrointestinal disorders

8.2.4 Point Selection Based on the Qi Flow of the Organs

Descend counterflow Qi:

- Stomach: **Ren-10, Ren-13, ST-34, ST-44, ST-45, L.I.-4**
- Lung: **LU-1, LU-5, LU-7**
- Heart: **HE-5, HE-8, Ren-15**
- Kidney: **KID-1, KID-7, Ren-4**
- Liver: **LIV-2, LIV-3, LIV-1, LIV-14**

Raise Qi:

Spleen: **Ren-6, Ren-12, Du-20, BL-20**

Ch08-F10028.qxd 2/23/08 4:41 PM Page 688

8.2.5 Point Selection According to the Five Phases (Elements)

While the Five Phase points are identical to the *shu*-transporting points in location, their function and application differ from the latter (for more detail → 8.1.6). There exist various very specific methods regarding diagnosis, point selection and therapy based on the Five Phases, which are practised by several different schools of thought. Within the framework of this Atlas, only a simplified overview of the possible clinical applications is listed below.

***Shu*-transporting points of the Yin channels**					
Phase/ channel	Wood point	Fire point	Earth point	Metal point	Water point
LU	LU-11	LU-10	LU-9 +	LU-8	LU-5 −
SP	SP-1	SP-2 +	SP-3	SP-5 −	SP-9
HE	HE-9 +	HE-8	HE-7 −	HE-4	HE-3
KID	KID-1 −	KID-2	KID-3	KID-7 +	KID-10
P	P-9 +	P-8	P-7 −	P-5	P-3
LIV	LIV-1	LIV-2 −	LIV-3	LIV-4	LIV-8 +
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point
***Shu*-transporting points of the Yang channels**					
Phase/ channel	Metal point	Water point	Wood point	Fire point	Earth point
L.I.	L.I.-1	L.I.-2 −	L.I.-3	L.I.-5	L.I.-11 +
ST	ST-45 −	ST-44	ST-43	ST-41 +	ST-36
S.I.	S.I.-1	S.I.-2	S.I.-3 +	S.I.-5	S.I.-8 −
BL	BL-67 +	BL-66	BL-65 −	BL-60	BL-40
T.B.	T.B.-1	T.B.-2	T.B.-3 +	T.B.-6	T.B.-10 −
G.B.	G.B.-44	G.B.-43 +	G.B.-41	G.B.-38 −	G.B.-34
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point

+, tonification point; − , sedation point

Cycles

The Five Phases influence each other both physiologically (*sheng* and *ke* cycle) as well as pathologically (*cheng*, *wu* and to some extent also *sheng* cycle). If the balance between the Five Phases is disturbed, pathological syndromes will appear.

- *Sheng* cycle (generating cycle, mother–son cycle): a phase generates and nourishes the following phase. Each phase nourishes (as mother) and is nourished (as son) at the same time. Fire nourishes Earth, Earth nourishes Metal, Metal nourishes Water, Water nourishes Wood, and Wood nourishes Fire. In a pathological condition either the mother is too weak to nourish the son sufficiently or the son is too strong, draining his mother and weakening her.
- *Ke* cycle (controlling cycle): one phase controls another and is itself controlled by yet another phase.
- *Cheng* cycle (overacting cycle): The controlled phase is pathologically suppressed or weakened.
- *Wu* cycle (insulting cycle): A phase is pathologically stronger than its controlling phase.

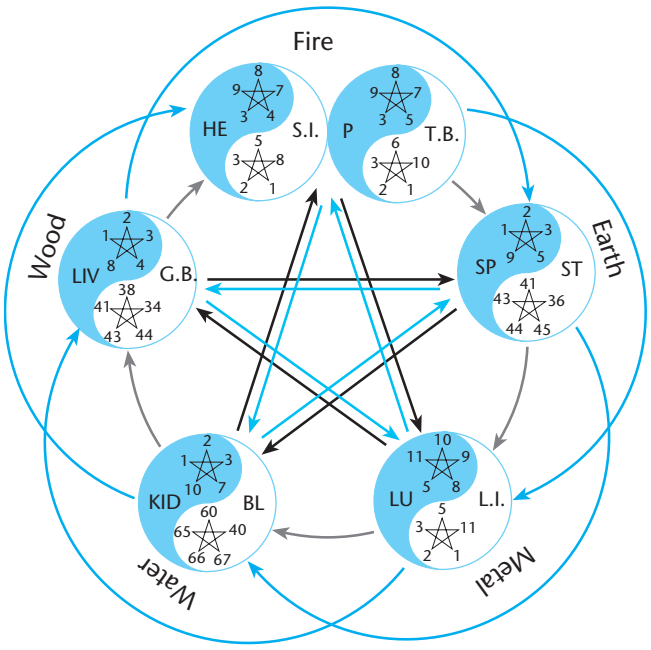

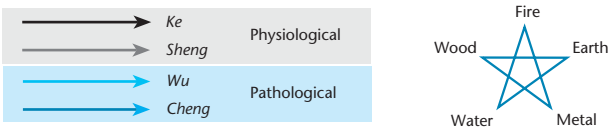

Ch08-F10028.qxd 2/23/08 4:41 PM Page 689

Simple point selection according to the *sheng* cycle

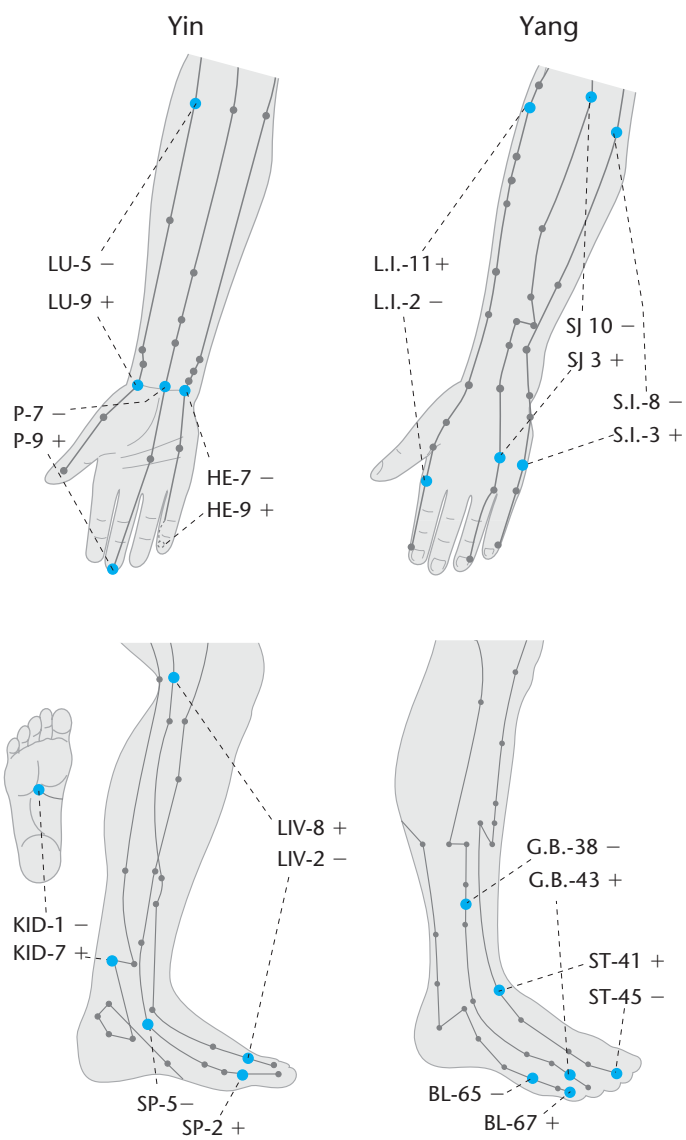

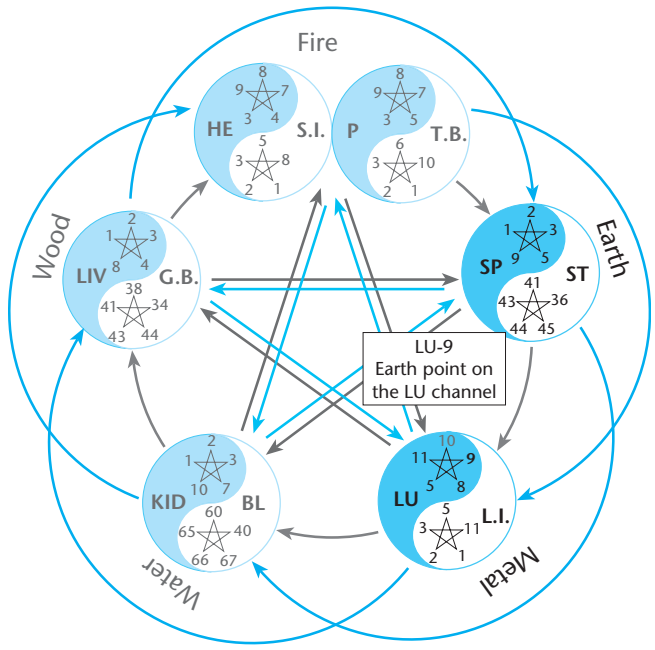

Treatment principle: For deficiency tonify the mother, for excess drain (sedate) the son. Mother–son points are identical to the tonification and sedation points of a channel. The correct needling technique is important in order to obtain the desired result:

- tonifying method for tonification points
- reducing method for sedation points

Example for the clinical application of the *sheng* cycle

For deficiency: In cases of deficiency of a particular channel/Organ, tonify that point on the affected channel that corresponds to the phase of the mother.

Example: Earth is the mother of Metal. If the Lung (Metal) is deficient, needle the Earth point (**LU-9**) on the Lung channel with tonifying technique.

For excess: If there is excess in a particular channel/Organ, select that point on the channel which corresponds to the phase of the son and needle it with reducing technique.

Example: Water is the son of Metal. If the Lung (Metal) is in excess, needle the Water point (LU-5) on the Lung channel with reducing technique.

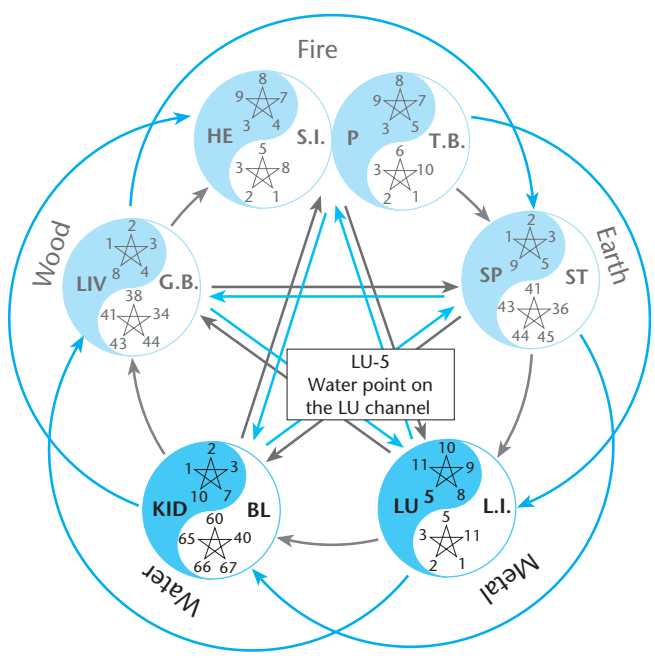

Ch08-F10028.qxd 2/23/08 4:41 PM Page 690

***Addendum*:** The terms tonification/sedation points are somewhat controversial, since the properties of a point can often be overshadowed by its other features. For example, according to the Five Phase theory, **P-9** and **HE-9** are considered tonification points, since they correspond to the phase of the mother. However, in clinical practice they are much more frequently used as *jing*-well points for acute situations – thus they are considered from the point of view of the Qi flow in the channels, especially in order to drain Heat.

Application of the Five Phase points for the elimination of external pathogenic factors

A connection exists between the Five Phases and pathogenic factors (Maciocia 1994; according to Ross 1998, however, these points should only be used for internal disorders):

- Wood corresponds to Wind
- Fire corresponds to Heat
- Earth corresponds to Dampness
- Metal corresponds to Dryness. According to Maciocia this correlation presents an exception: the Metal point should not be needled to treat Dryness; a moistening treatment is here the preferred choice.
- Water corresponds to Cold.

In accordance with these interrelations, Five Phase points can be applied to eliminate pathogenic factors.

Example: Acute sore throat with fever, redness and swelling of the throat due to Wind-Heat

Affected phases: Wood (Wind), Fire (Heat)

Affected Organ/channel: Lungs

Therapy: Needling of the Wood point on the LU channel (**LU-11**) and the Fire point on the LU channel (**LU-10**).

The application of the ***ke* cycle (controlling cycle)** allows a more differentiated approach in the therapy and elimination of pathogenic factors:

- On the Yin channel, needle the point corresponding to the pathogenic factor with reducing technique (for example, for Heat needle the Fire point)
- In addition, needle the point which corresponds to the controlling phase (according to the *ke* cycle/controlling cycle → 8.2.5) of the pathogenic factor on the paired Yang channel with tonifying technique.

Ben points

***Shu*-transporting points of the Yin channels**					
Phase/ channel	Wood point	Fire point	Earth point	Metal point	Water point
LU	LU-11	LU-10	LU-9 +	LU-8	LU-5 −
SP	SP-1	SP-2 +	SP-3	SP-5 −	SP-9
HE	HE-9 +	HE-8	HE-7 −	HE-4	HE-3
KID	KID-1 −	KID-2	KID-3	KID-7 +	KID-10
P	P-9+	P-8	P-7 −	P-5	P-3
LIV	LIV-1	LIV-2−	LIV-3	LIV-4	LIV-8 +
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point
***Shu*-transporting points of the Yang channels**					
Phase/ channel	Metal point	Water point	Wood point	Fire point	Earth point
L.I.	L.I.-1	L.I.-2 −	L.I.-3	L.I.-5	L.I.-11 +
ST	ST-45 −	ST-44	ST-43	ST-41 +	ST-36
S.I.	S.I.-1	S.I.-2	S.I.-3 +	S.I.-5	S.I.-8 −
BL	BL-67 +	BL-66	BL-65 −	BL-60	BL-40
T.B.	T.B.-1	T.B.-2	T.B.-3 +	T.B.-6	T.B.-10 −
G.B.	G.B.-44	G.B.-43+	G.B.-41	G.B.-38−	G.B.-34
Channel Qi flow	*Jing*-well point	*Ying*-spring point	*Shu*-stream point	*Jing*-river point	*He*-sea point

+, tonification point; – , sedation point

Ben points (root points, element points) represent the same phase as its pertaining channel. **Example:** The Spleen corresponds to Earth; thus the Earth point on the Spleen channel is its *ben* point.

Ch08-F10028.qxd 2/23/08 4:41 PM Page 691

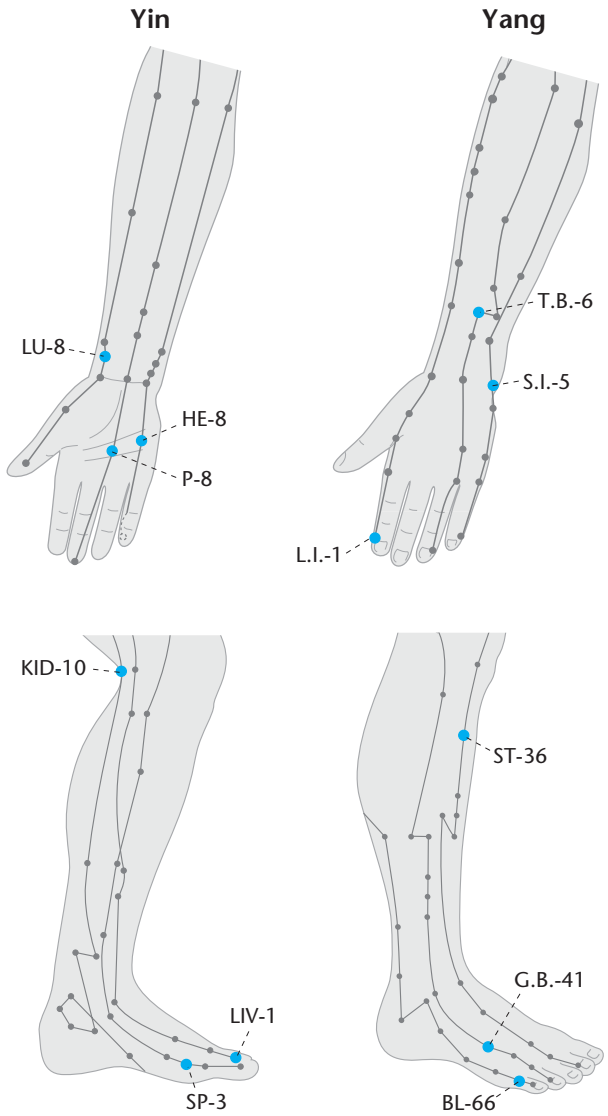

8.3 Point Combinations

8.3.1 Combining Local and Distal Points

Point selection and combination can be carried out according to affected areas and channels (→ 8.2.1, → 8.2.2).

For disorders affecting the channel: For example, for *bi*-syndromes affecting the sinews and joints, distal points can be stimulated first. Then local and adjacent points can be selected based on tenderness. These are needled and/or subsequently cupped.

For disorders of the *zangfu*-Organs: Generally, only distal points are applied in acute conditions. Once the condition (for example, acute pain) has improved, local points can be selected. In chronic conditions the combination of local and distal points – and especially Back-*shu* and Front-*mu* points as adjacent points (→ 8.3.2) – is commonly used. Another possibility is the combination of the opening point of an extraordinary vessel (→ 8.1.8) with local points in the affected area.

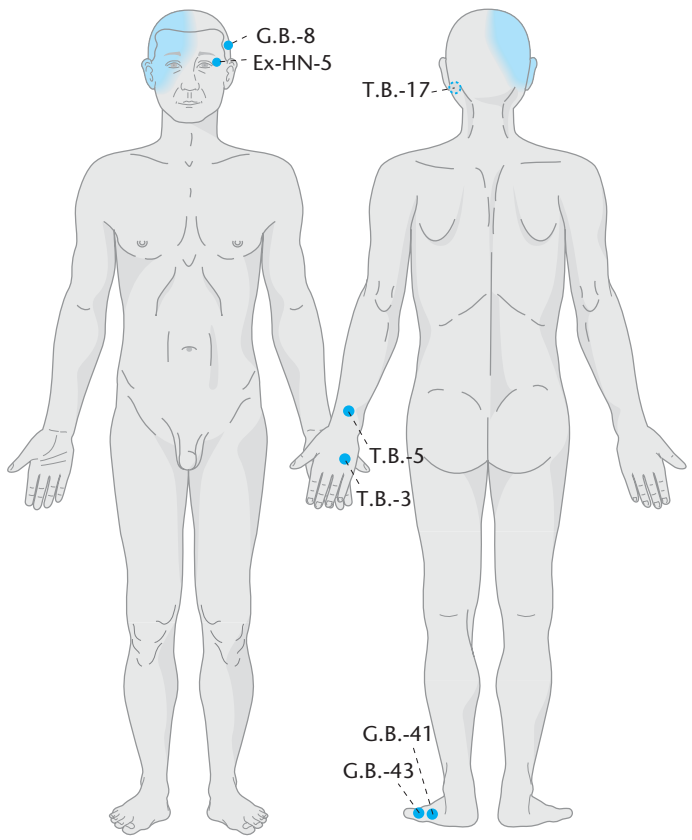

Clinical application

- **For deficiency conditions:** needling the *ben* point with tonifying technique will tonify its corresponding Organ or channel
- **For excess conditions:** needling the *ben* point with reducing technique will drain the excess from its corresponding Organ or channel
- **Spiritual aspect:** the *ben* points on the Yin channels also affect the spiritual aspect of their corresponding phases. Thus the LU *ben* point **LU-8** allows the *po* (corporeal soul) to unfold its potential. Similarly the HE *ben* point **HE-8** will empower the *shen* (spirit), the SP *ben* point **SP-3** the *yi* (thinking), the LIV *ben* point **LIV-1** the *hun* (ethereal soul) and the KID *ben* point **KID-10** the *zhi* (will).

Ch08-F10028.qxd 2/23/08 4:41 PM Page 692

8.3.2 Combining Points on the Front and Back of the Body

Here, points on the anterior and posterior aspect of the body are combined:

- **Anterior points**: mainly located on the thorax and abdomen
- **Posterior points**: mainly located on the back and lumbar region

Back-*shu*/Front-*mu* method

The combination of the Back-*shu* point (→ 8.1.4) and the Front-*mu* point (→ 8.1.5) of the affected Organ is the most important way of combining points on the front and the back of the body. The Back-*shu*/Front-*mu* method enhances and extends the therapeutic effect, compared to using either of those points only. It has a strong Yin/Yang balancing effect and it is particularly beneficial for chronic disorders of the *zangfu*-Organs.

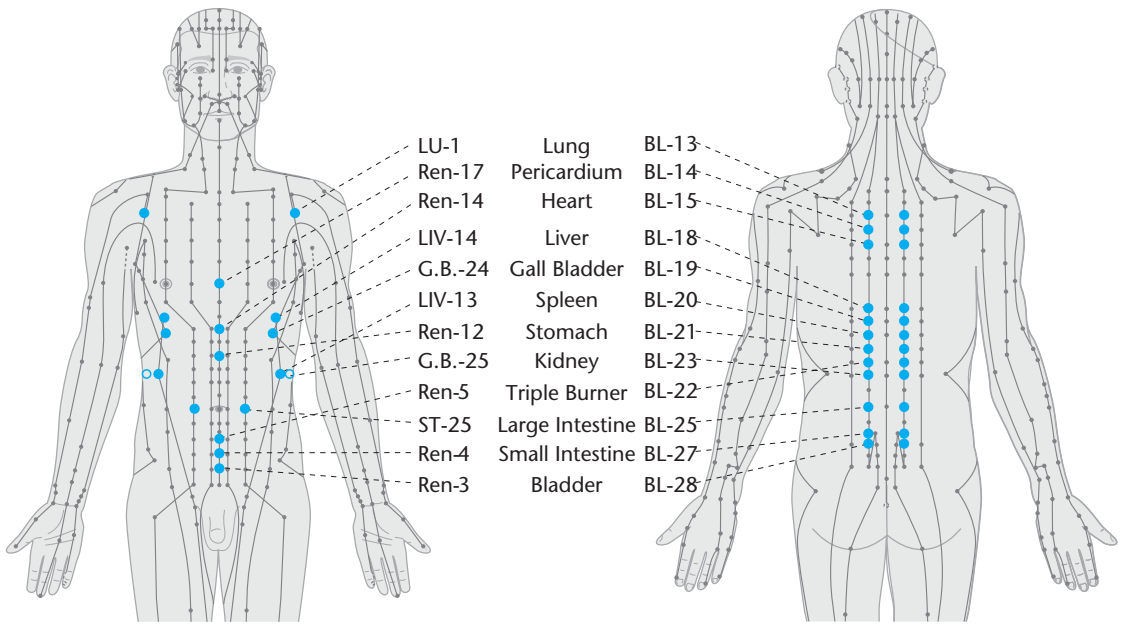

Clinical application of the Back-*shu*/Front-*mu* method

- Needling of the Back-*shu* and Front-*mu* points during one session
- For repeated treatments within a short period of time, alternate needling of the Back-*shu* and Front-*mu* points (for example, during the first treatment select the relevant Back-*shu* point, in the next treatment select the corresponding Front-*mu* point, etc.)
- The Back-*shu*/Front-*mu* combination can also be applied to balance an incorrect treatment. For example, if the needles have been retained too long in the Back-*shu* points and the patient is tired, Front-mu points can be needled to restore balance.

Balancing the *du mai* and the *ren mai*

The points on the *ren mai* are located on the anterior aspect of the body, while those of the *du mai* are predominantly located on the posterior aspect. Combining points on the *ren mai* and *du mai* has a balancing affect on the Yin and Yang and can regulate the ascending and descending flow of Qi. This combination therefore not only balances the front and the back of the body, but also the Yin and the Yang. Furthermore, it has a strong effect on the psyche and, depending on the needling technique and point selection, will have a calming or stimulating effect (figure modified according to Ross 1998).

Ch08-F10028.qxd 2/23/08 4:41 PM Page 693

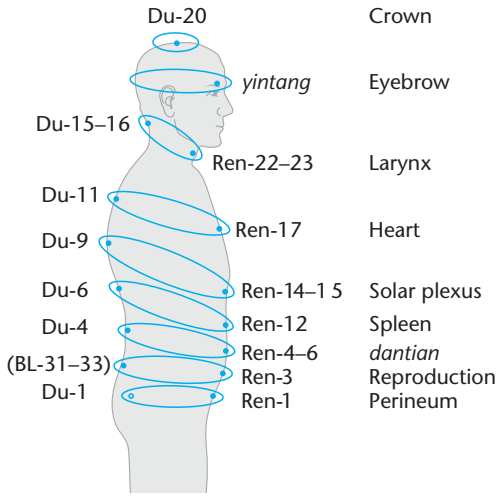

8.3.3 Yin–Yang Combination

Synonym: Interior/Exterior combination

The Yin–Yang combination balances the flow of Yin and Yang in the channels. It is based on the concepts of channel energetics (→ channel circuits, Fig. 1.12).

Needling too many points on the Yang channels can make the patient nervous and restless. In this case, needling points on the Yin channels can restore balance and have a calming effect.

Needling too many points on the Yin channels can lead to tiredness. In this case, needling points on the Yang channels has a balancing, stimulating effect.

Balancing Yin and Yang in paired channels

- Points on the Interiorly–Exteriorly paired channels are combined in order to balance Yin and Yang. For example the LU and L.I. channels are the Interiorly–Exteriorly paired channels of the hand, while the ST and SP channels are the Interiorly–Exteriorly paired channels of the foot of the first circuit.
- This method increases the therapeutic effect compared to using only Yin or only Yang points.

Yuan/luo combination

Most important Yin–Yang combination (Synonym: host–guest combination): The *yuan*-source point (host) of the primarily affected channel/Organ is combined with the *luo*-connecting point of the Interiorly–Exteriorly paired channel.

Important *yuan/luo* combinations

- **L.I.-4** (*hegu*) and **LU-7** (*lieque*) support the descending function of the Lung Qi, expel pathogenic factors, tonify the Defensive Qi (*wei qi*) and calm the *shen*
- **SP-3** (*taibai*) and **ST-40** (*fenglong*) tonify the Spleen and transform Phlegm
- **LIV-3** (*taichong*) and **G.B.-37** (*guangming*) benefit the eyes for Liver syndromes
- **T.B.-4** (*yangchi*) and **P-6** (*neiguan*) regulate the Triple Burner, move Liver Qi, calm the *shen*, relax the muscles in the occiput and shoulders.

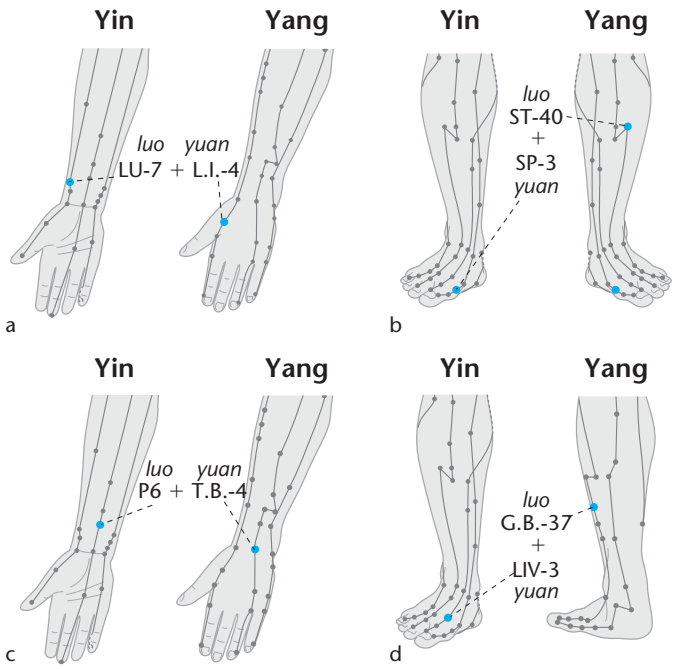

Further methods of combining points of Interiorly–Exteriorly paired channels and modified *yuan/luo* combinations

a Combining the *luo*-connecting points of Interiorly–Exteriorly paired channels tends to enhance the therapeutic effect; for example combine **ST-40** and **SP-4** for abdominal pain

b For chronic disorders, combine the *yuan*-source point and the *luo*-connecting point of the same Yin channel in order to increase the therapeutic effect; for example needle **LU-9** and **LU-7** for chronic cough.

c For disorders affecting a Yin channel, only needle the *yuan*-source point of the Interiorly–Exteriorly paired Yang channel. For example for an acute cold (a Lung syndrome), needle only **L.I.-4** (*yuan*-source point).

d For unilateral musculoskeletal disorders affecting a Yang channel: in addition to local and adjacent points on the affected channel, needle the contralateral *luo*-connecting point of the Interiorly–Exteriorly paired Yin channel. This method, which promotes balance between Yin and Yang, is mainly used for channel-pertaining disorders. For example: for unilateral arm or shoulder pain along the L.I. channel, needle points on the L.I. channel on the affected side as well as **LU-7** on the opposite side of the body.

Ch08-F10028.qxd 2/23/08 4:41 PM Page 694

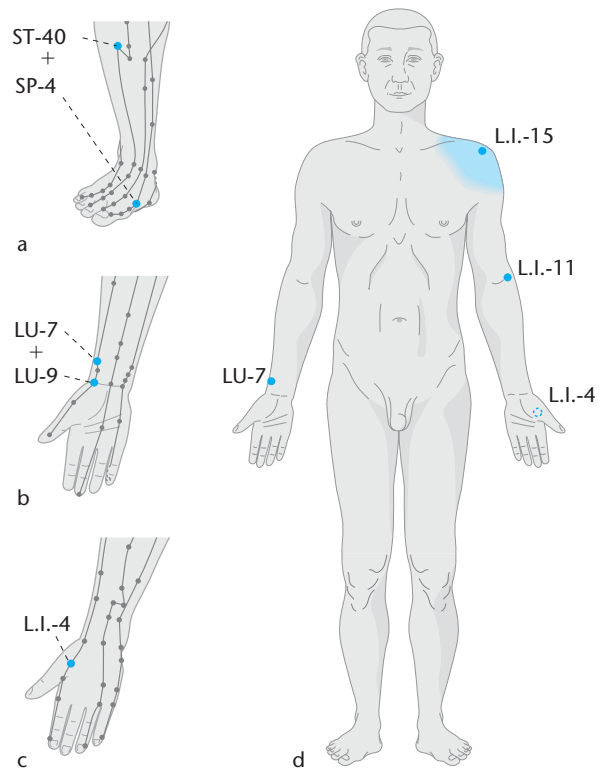

8.3.4 Combining Points from Above and Below

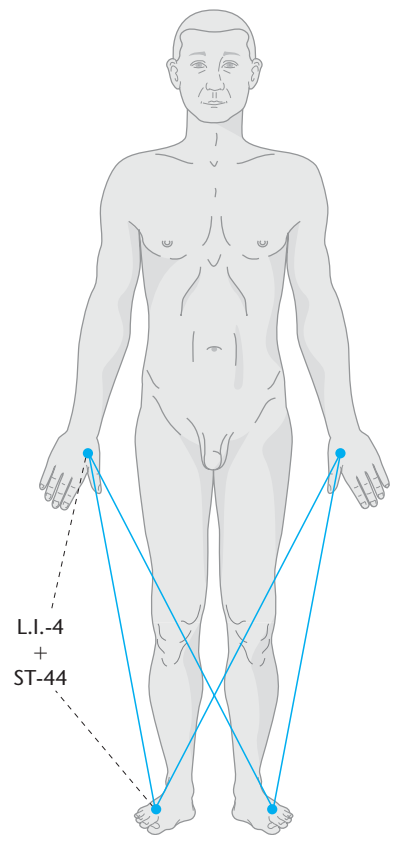

Yin–Yang balance for unpaired channels

Examples:

ST-36 + P-6	Harmonises the Middle Burner and descends counterflow Qi
L.I.-4 + LIV-3	Most important combination for regulating the Qi flow in the whole body

- A balance between the **upper and lower** half of the body is achieved by selecting points that are evenly distributed. For headaches, therefore, **L.I.-4** on the upper extremity is combined with **ST-44** on the lower extremity.
- Combining points above with points below **promotes the smooth flow of Qi**. For example: combining the opening point and coupled point of an extraordinary vessel (→ 8.1.8).

When points from above and below should not be combined

- for acute disorders of the joints and the back – here only distal points should be needled with the reducing method.
- if there is a pre-existing energetic disharmony between above and below. For example: for Heat sensations of the head (for example due to Kidney Yin deficiency with Empty Fire) needle **KID-1** in order to cause the Qi to descend. On the other hand, for a uterine prolapse use moxibustion **on Du-20** in order to raise the Qi.

8.3.5 Left–Right Combination

Often the balance between the right and left half of the body can be restored through bilateral needling of the relevant points. This will increase the therapeutic effect as opposed to unilateral needling.

Ch08-F10028.qxd 2/23/08 4:41 PM Page 695

Clinical application for Interiorly–Exteriorly paired channels (→ 1.2.3)

- For acute unilateral channel disorders caused by the invasion of pathogenic factors, combine **local points** on the affected side with the ***luo*-connecting point** on the contralateral side.
- For chronic unilateral channel disorders caused by the invasion of pathogenic factors, combine **local points** on the affected side with the *luo*-**connecting point** of the **Interiorly–Exteriorly paired channel** on the **contralateral** side (with tonifying technique).

Indications for unilateral and/or contralateral needling

- **For acute, painful disorders:** contralateral or diagonal needling (corresponding hand and foot channel, also see → 8.2.1 and → 8.2.2)
- **For chronic joint disorders:** ipsilateral needling of several points balanced by needling of contralateral points.
- **In paediatric acupuncture** (use fewer needles)
- **For ongoing treatment sessions:** occasional balancing contralateral needling of the healthy side of the body

8.3.6 Chain and Lock Point Association Method

Two, three or more points on the same channel are selected and needled one after the other (in a row). This method is predominantly applied for disorders of the musculoskeletal system or the nervous system.

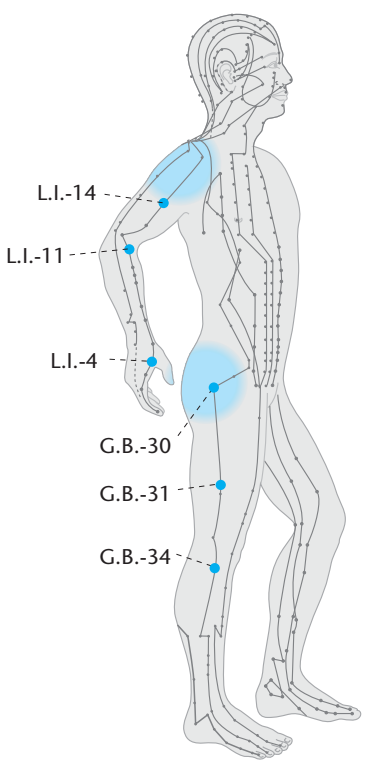

8.3.7 Point Selection According to the Organ Clock

Within a 24-hour cycle each channel experiences a two-hour period of maximum Qi flow (→ Fig. 1.17, Chinese Organ clock). Some modern (Western) schools extrapolate the circadian Qi flow to the Organs (internal channel pathways and divergent channels). If one Organ is at its maximum period, the Organ on the opposite side of the clock will be at its minimum period, in other words the Qi flow will at that time be at its lowest. Symptoms of a disorder occurring at a particular time may provide a hint about the Organ affected, which at that time will be during its **maximal period in excess syndromes** or during its **minimal period in deficiency syndromes**.

For example: Always waking up between 1 and 3 am can be a sign of a Liver syndrome. Thus the tonification point and/or *luo*-connecting point of the channel opposite to the Liver on the Organ clock could be selected, in this example **S.I.-3** (*houxi*, tonification point) or **S.I.-7** (*zhizheng*, *luo*-connecting point).

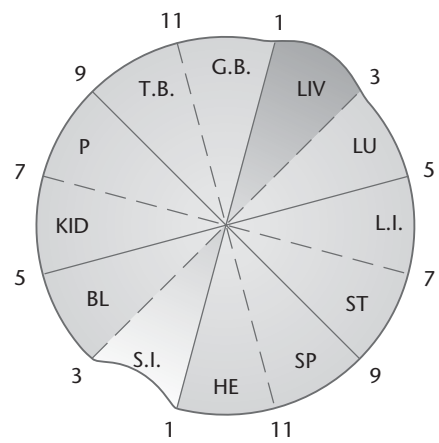

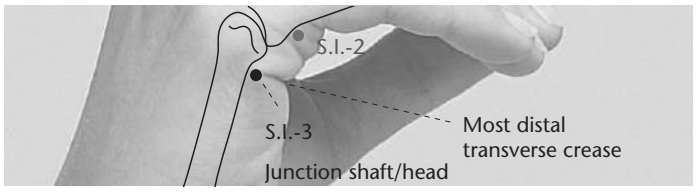

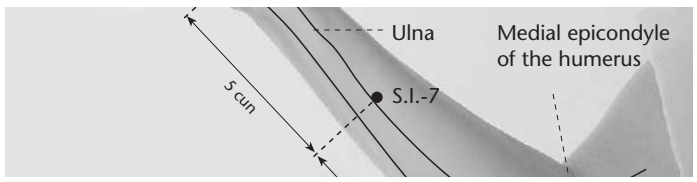

Addendum: There are some very specific methods for chronoacupuncture, for exampel the *zi wu liu zhu* method and the *ling gui ba fa* method, where points are needled at specific times (→ Kubiena and Ramakers 2002).

Ch08-F10028.qxd 2/23/08 4:41 PM Page 696

8.3.8 Clinical Application

The following illustration shows the clinical application of several point selection methods (8.3.1 – 8.3.5) for the treatment of dysmenorrhoea.

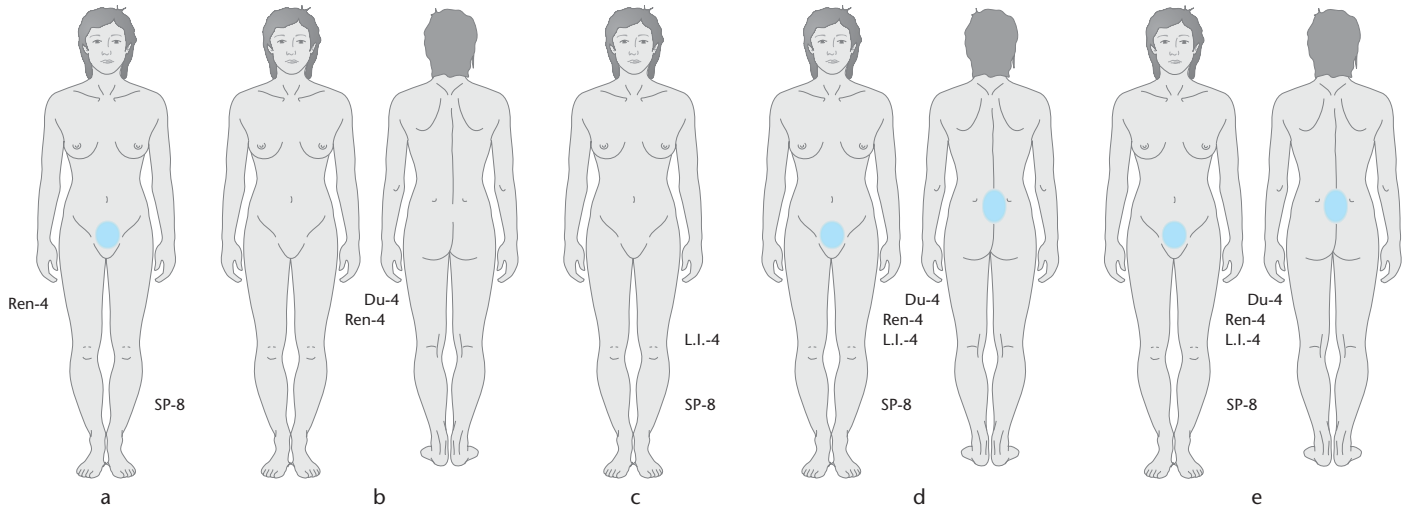

Possible point combinations for the treatment of dysmenorrhoea:

a. combining local and distal points (→ 8.3.1)
b. combining points on the front and the back of the body (→ 8.3.2) and Yin–Yang (→ 8.3.3)
c. Yin–Yang combination (→ 8.3.3) and above–below (→ 8.3.4)
d. Combination of all of the above methods (→ 8.3.1–8.3.4)
e. Additionally combining points on the left and right side of the body (→ 8.3.5)

9 Scientific Research

Ingolf Hosbach

9.1 Introduction

The aim of this chapter is to give a comprehensive overview about the scientifically proven or disproved effectiveness of certain points. It is not the aim of this chapter to prove the effectiveness of acupuncture as such. Based on the latest research, and 1000 recent human studies with mostly positive results, it is assumed that acupuncture is effective, despite criticism. If applied properly it is a well-established, low-risk method with minimal side effects.

Accordingly, the scope of acupuncture research has developed from simple questions, such as whether acupuncture works, to more specific questions such as which acupuncture points can be used for which conditions and with which stimulation methods. Furthermore, studies are increasingly designed for precise diagnoses and treatments based on the principles of Traditional Chinese Medicine. By the same token, the quality of the biomathematical methods used for the design and analysis of the studies is increasing. The number of randomised, controlled studies (RCTs) is also on the rise (Fig. 1).

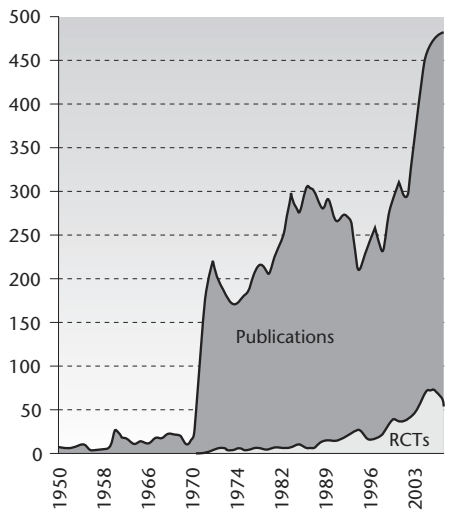

Fig. 1: Number of all publications and RCTs per year between 1954 and 2005. (Source: PubMed, US National Library of Medicine, Bethesda, Maryland)

The percentage of RCTs of all publications is nearly 7%, a figure that compares well with most conventional forms of therapy (for example: arthroscopy 5%, beta-blockers 7%, pain medication 7%, lowering cholesterol medication 8–11%).

Fundamental principles for scientific assessment

One of the prerequisites for including a study in this chapter was the mention of specific acupuncture points in the abstract. Out of 1092 studies listed in PubMed until May 2006 under the keyword 'acupuncture', 259 were therefore chosen for this chapter. In some exceptional cases, studies based on animal experiments were also considered. Regarding the presentation of the results, particular emphasis was placed on the role of biomathematical methods as a parameter for reliability. The terms which allow assessment of a study in terms of its biomathematical quality are explained below:

Case study: Usually a retrospective (=in hindsight) summary of results of a particular treatment method in patients with a particular disorder; no control group or control therapy. Most basic form of clinical trial.

Controlled study: Usually a prospective (=planned in advance) scientific comparison of a treatment method with a group of untreated patients or of patients treated with a placebo or conventional therapy.

Placebo: A form of 'mock' or sham therapy. This can take the form of a pill containing no active ingredient or a therapy without any authentic foundation. The administration of a placebo tries to eliminate the non-specific improvement of the researched disorder (such as the therapist's care of the patient and expectations of the patient) compared to the study group receiving authentic treatment. While the possibilities for placebo-acupuncture are plentiful, none is free of disadvantages. The effectiveness of acupuncture as a holistic therapy can be seen as the sum of many psychological as well as non-specific and specific physiological effects. Depending on the choice of the acupuncture placebo, certain effects can be excluded. Based on the concept of the sinew channels, however, there is an ongoing debate whether non-acupuncture points exist at all. The highest level of placebo acupuncture is the so-called 'sham acupuncture' where actual deep needling takes place at non-acupuncture points. The value of a study increases considerably if it is controlled by a placebo therapy.

Randomisation: The random assignment of a patient to a treatment group, without any influence of either the patient or the therapist. Randomisation significantly increases the value of a study.

Waitlist: A control group formed by patients receiving no treatment as a form of control. While a patient is waiting for his/her treatment, all his/her data relevant to the treatment are collected and later compared to those data obtained during treatment. Since there is neither randomisation nor a placebo treatment, the value of a waitlist group as a control group is very low.

Cross-over study: A particular design of controlled study. Each patient will participate in two different courses of treatment (for example treatment 1: medication; treatment 2: acupuncture). The order in which the treatments are administered can differ from patient to patient and be random (=randomised cross-over study). However, any long-term effect of the first form of treatment poses a problem for the final result. Therefore so-called washout phases during which patients receive no treatments at all are implemented to eliminate these effects and prevent a distortion of the data to be collected during the second phase. Since the effects of some acupuncture treatments have been shown to last for up to a year, cross-over studies are not very suitable for acupuncture research and particularly non-significant results should be questioned.

The 'n of 1' study is another form of research: a patient will receive various different treatments in order to find out which form of treatment led to the best improvement for the relevant disorder. However, the same limitations as discussed above will affect this type of study.

Significance: In simple terms, the significance of a study indicates if the difference of a particular parameter which has been observed between two groups is caused by the treatment (and is therefore a true difference) or whether the difference is the result of random deviation. For example, if in a large number of treated patients the observed data differ 'significantly' from a group of untreated patients, it is **highly probable** that this difference is not due to a random deviation. This probability can be expressed mathematically. If, for example, the significance is expressed as $p = 0.03$ this means that, considering a random error of 3%, the difference noted between the two groups is statistically significant. The convention is to accept a random error of up to 5% as acceptable. The significance of a study is based on the number of patients treated and the extent of the difference in the treatment result (for example by assuming the same random error, the difference in treatment results has to be much larger in a group of 10 patients ($n = 10$) than if the same study was carried out with 100 patients ($n = 100$)). Only differences classified as 'significant' are accepted for research studies.

Power of the study: This refers to the ability of a study to be statistically significant. The higher the number of participants, the higher the power of the study. By international standards, 80% is an acceptable level of power. Depending on the difference between the two study groups, the number of patients required can therefore easily be in the 100s or 1000s.

Experimental studies: Through standardising and controlling conditions of the experiment strictly, every attempt is made to avoid random results, often leading to just a small number of participants. Generally these tend to be healthy volunteers, who are subjected to a clearly defined stimulus. In this case acupuncture as a form of therapy will achieve a high level of validity. More questionable are studies of healthy subjects without a well-defined stimulus, since in that case acupuncture as a balancing form of treatment is not used according to its principles (for example changes in peristalsis caused by acupuncture in healthy subjects).

Blinding: The gold standard of clinical trials are the so-called double blind studies in which both the patient and the therapist remain ignorant as to which treatment (for example the tablet with or without the active ingredient) is administered to the patient. However, blinding of an acupuncturist is not possible. Even in a single blind study (only the patient is 'blinded'), his/her behaviour will distort the result of the trial. This effect can be counterbalanced to some extent by blinding the assessor of the treatment results (for example changes in the range of motion) so that he/she does not know which treatment (authentic or sham acupuncture) was received by the patient.

Standardisation: In pharmaceutical trials the amount of the active ingredient per tablet can easily be stated (for example standardisation of a 10 mg tablet). However, acupuncture cannot be standardised. The acupuncturist's point selection will largely be influenced by experience, prior knowledge, specialisation and cultural background. Strictly speaking, acupuncture trials therefore do not investigate acupuncture per se but acupuncture as performed by an acupuncturist or a group of acupuncturists.

The fact that acupuncture is to a large extent centred on the therapist has in the past led to a paradoxical situation: experienced Chinese acupuncturists performing the 'best' acupuncture produced biomathematically 'bad' trials due to their lack of training, while 'bad' Western acupuncturists produced 'good' trials from a biomathematical point of view. 'Importing' Chinese acupuncturists into Western trials did not solve the problem either, since the Chinese clinicians were now treating Europeans with diseases not common in China and which reacted differently to acupuncture. Only as experience grows on both sides – which would have to be evaluated individually for each trial – will the value of acupuncture as a researched form of therapy increase accordingly.

9.2 Scientific Research According to Channels

Lung Channel

LU-5 → Ren-23; L.I.-10

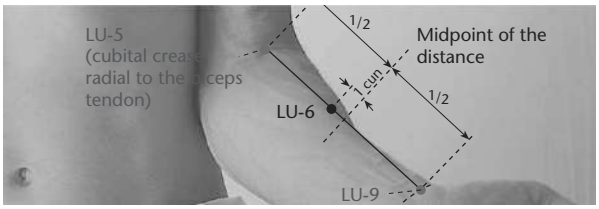

LU-6 192 patients with bronchial asthma were treated with **LU-6** and **LU-10**. 98.9% of the patients experienced an immediate improvement; the rate of clinical remission plus marked improvement was 76.5%. Best therapeutic results were achieved for allergic asthma according to Western medicine (asthma due to Cold according to TCM) and with longer needle retention (approximately 40 minutes). The lack of blinding and control groups reduce the value of this precise therapeutic case study. **Zang, J Tradit Chin Med 1990**

LU-7 → ST-8

LU-10 → LU-6

A Chinese prospective, randomised, controlled, non-blinded trial investigated the effect of bilateral bloodletting at **LU-10**, **P-8**, **HE-8**, **L.I.-2**, **T.B.-2** and **S.I.-3** on heart rate and consciousness in patients with **apoplexy** that had occurred within 3 days. Only patients with mild injuries showed improved consciousness while the heart rated increased independently of the size of injury. **Yi et al**, **J Tradit Chin Med 2005**

LU-11 Three Russian animal trials demonstrated the antipyretic action of **LU-11**. LU-11 can also strengthen the effect of hypothermic substances, suggesting a combination with antipyretic Chinese herbs. **Nezhenzev et al**, **Biull Eksp Bio Med 1992, and Fizion Zh SSSR Im I M Sechenova 1991**

A Taiwanese prospective, randomised, non-blinded, controlled trial (n = 76) investigated whether bloodletting at **LU-11** and **L.I.-1** can prevent or treat the occurrence of **laryngospasm** after tracheal extubation in children. The incidence of laryngospasm was significantly less in the bloodletting group (5% vs 24%). If laryngospasm developed, this could be relieved with the treatment of LU-11 or L.I.-1 within one minute in all patients. **Lee et al, Anaesthesia 1998**

Large Intestine Channel

L.I.-1 → LU-11

L.I.-2 → LU-10

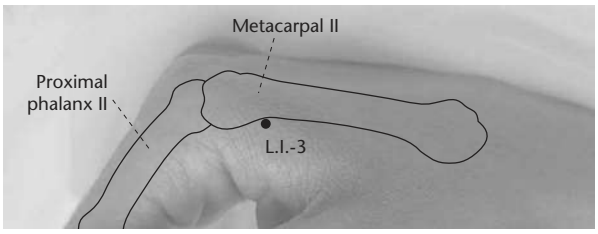

L.I.-3 A Chinese prospective, randomised, controlled study (n = 210) reports the significantly better effects of needling **L.I.-3** and ***lingxia*** in the treatment of **periarthritis of the humerus** compared to the control group treated with local points. Of interest is the application of *lingxia*, an unofficial extra point 2 cun inferior to G.B.-34. **Feng, J Tradit Chin Med 2003**

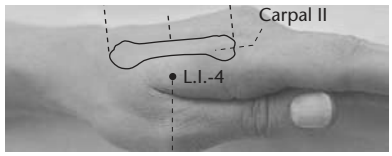

L.I.-4 → ST-36; P-5, P-6; ST-8; G.B.-43; Du-20; LIV-3; S.I.-18; SP-9; L.I.-11; ST-2

A Turkish prospective, controlled, non-blinded, group-comparison trial (n = 55) investigated the effect of electro-acupuncture (EA) at **L.I.-4**, **L.I.-11**, **ST-25**, **ST-36**, **ST-44** and **LIV-3** (combined with auricular points *shenmen* and *sanjiao*) on **obesity in women**. Treatments were given daily for 20 days. One control group received a 1425 kcal diet for 20 days while the other control group received no treatment. There was a 4.8% weight reduction in the EA group, whereas women on the restricted diet had a 2.5% weight loss. Both groups showed a significant reduction in total cholesterol and triglyceride levels. In the EA group only, LDL levels were also reduced significantly. **Cabioglu et Ergene**, **Am J Chin Med 2005**

A Chinese (Hong Kong) prospective, randomised, controlled, non-blinded study (n = 29) investigated the effect of **L.I.-4** and **P-6** on the heart rate variability (HRV) in normal subjects in **fatigue** and **non-fatigue states** respectively. Only fatigued subjects showed a significant change in HRV. The authors concluded that the effect of acupuncture depends not only on the points selected but also on the functional state of the subject, one of the principles of Chinese Medicine. **Li et al, Eur J Appl Physiol 2005**

An unusual veterinary case study investigated the effects of acupuncture on the **immune system**. According to the authors, immunostimulant points **include LI-4**, **LI-11**, **ST-36**, **GB-39**, **SP-6**, **GV-14**, **BL-11**, **BL-20**, **BL-23**, **BL-24**, **BL-25**, **BL-26**, **BL-27**, **BL-28,** and **Ren-12**. **BL-47** is considered to be immunosuppressive. Anti-febrile points include **ST-36** and **Du-14**. **Rogers et al, Probl Vet Med 1992**

A Chinese prospective, randomised, single-blinded, experimental trial ($n = 20$) compared the effect of transcutaneous electric nerve stimulation (TENS) at **L.I.-4** on the **heat pain threshold** and **vibration threshold** in humans with a placebo control point. In the study group, after TENS at **L.I.-4**, the pain threshold for heat was higher for up to half an hour compared to the control group. There was no effect on the threshold for feeling vibration. **Wang et al, Chin Med J (Engl) 2003**

A Korean prospective, randomised, placebo-controlled trial ($n = 35$) evaluated the efficacy of electro-acupuncture (EA) and moxibustion on **L.I.-4**, **L.I.-10**, **L.I.-11** and **T.B.-5** on **spasticity due to stroke**. The control group was treated with routine acupuncture. Only the EA group experienced a significant, immediate and lasting reduction in spasticity (using a modified Ashworth scale). **Moon et al, Am J Chin Med 2003**

A German prospective, randomised, controlled, non-blinded trial ($n = 29$) investigated if acupuncture at **L.I.-4**, **T.B.-5**, **ST-36**, **P-6**, **Du-20** and **Ex-UE-9** (*baxie*) had a vasospasmolytic effect on patients with **Raynaud's syndrome** during the winter season. There was no significant difference regarding the number and intensity of episodes, nor was there a significant effect on skin microcirculation between the acupuncture group and the placebo group before and after treatment. **Hahn et al, J Dtsch Dermatol Ges 2004**

A Japanese prospective, randomised, controlled study ($n = 21$) investigated the effect of **L.I.-4**, **ST-36 and SP-6** on objective and subjective parameters regarding **exhaustion** in elite female soccer players during a competition period. Only the subjects treated with acupuncture experienced a better rating for both subjective and objective parameters. **Akimoto et al, Med Sci Sports Exerc 2003**

A Chinese prospective, randomised, controlled trial ($n = 89$) researched the effect of **L.I.-4**, **ST-36**, **BL-18** and **BL-23** combined with CT positioning scalp circum-needling (SCN) and herbal therapy in patients with **poly-infarctional vascular dementia**. The control group received herbal treatment only. The treated group had a significantly better short-term result (96.5%) than the control group (75%) regarding changes of clinical symptoms, intelligence and haemorrheological characteristics. **Lun et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 2003**

A Chinese prospective, randomised, controlled experimental trial ($n = 11$) investigated the changes in **brain activation patterns** evoked by stimulating **L.I.-4** manually versus stimulation with electro-acupuncture (EA). These were assessed by functional magnetic resonance imaging (fMRI). Results showed that EA mainly produced fMRI signal increases in precentral gyrus, postcentral gyrus/inferior parietal lobule and putamen/insula; in contrast, manual needle manipulation produced prominent decreases of fMRI signals in posterior cingulate, superior temporal gyrus, putamen/insula. These results suggest that different brain mechanisms may be recruited during manual and EA. **Kong et al, J Altern Complement Med 2002**

An American prospective, randomised, controlled, non-blinded experimental trial ($n = 13$) evaluated the effect of transcutaneous electric acupoint stimulation (TEAS) on **L.I.-4** and **P-6** on **tonic pain** induced with the cold-pressor test. While TEAS alone had a significant analgesic effect, this could be increased by combining TEAS analgesia with low-dose morphine. **Yuan et al, J Clin Pharmacol 2002**

A Taiwanese prospective, randomised, placebo-controlled experimental study ($n = 13$) investigated the effect of manual acupuncture as well as transcutaneous electrical nerve stimulation (TENS) with 2 Hz and 100 Hz respectively at **L.I.-4** on the spinal cord by measuring the **H-reflex**. While the application of TENS with both 2 Hz and 100 Hz changed the amplitude of the H-reflex, it was retained longer with the 100 Hz stimulation. These findings suggest that TENS enhances the excitability of the motoneuron pool in the spinal cord. **Chang et al, Acupunct Electrother Res 2001**

An Austrian prospective, randomised, controlled study ($n = 45$) evaluated whether acupuncture at **L.I.-4** and **SP-6** at term can influence cervical ripening, **induce labour** and thus reduce the need for postdate induction. Compared to the control group, the study group showed a significantly shorter time span (three days) between estimated date of confinement (EDC) and actual time of delivery. Medical induction of labour 10 days after EDC was also significantly lower in the acupuncture group (20%) than in the control group (35%). **Rabl et al, Wien Klin Wochenschr 2001**

An Israeli prospective, randomised, double-blinded controlled trial ($n = 25$) investigated the effect of acupuncture in the treatment of **irritable bowel syndrome**. True acupuncture was performed at **LI-4** and sham acupuncture at **BL-60** during a total of two sessions. There was no significant difference between the two groups. However, the choice of acupuncture points seems somewhat peculiar. While both points are Ma Dan Yang points (the 12 most important points), they have no specific action for intestinal disorders. Obviously the authors chose the points according to the channel (Large Intestine). Furthermore, only two treatments does not seem sufficient. **Fireman et al, Digestion 2000**

A British prospective experimental trial ($n = 14$) investigated the effect of acupuncture at **L.I.-4** on the **resting EEG** of healthy subjects. Acupuncture did not induce any changes in the resting EEG. **Rosted et al, Complement Ther Med 2001**

An Irish prospective, randomised, controlled trial ($n = 30$) applied transcutaneous electric nerve stimulation (TENS) either at **L.I.-4** or over the median nerve, in order to investigate **cutaneous blood flow**. The control group received no treatment. Only TENS over the median nerve showed a significant increase in cutaneous blood flow. **Cramp et al, Acupunct Electrother Res 2001**

A Japanese prospective, randomised, controlled, non-blinded trial ($n = 22$) investigated the effect of acupuncture at **L.I.-4**, **ST-6** and **ST-7** on pain after mandibular **wisdom tooth extraction**. Compared to the control group there was a significant reduction in pain in the treatment group. **Kitade et al, Acupunct Electrother Res 2000**

A Swedish prospective, randomised, controlled experimental trial ($n = 12$) investigated if acupuncture at **L.I.-4** induced **changes in the sympathetic and/or parasympathetic nervous system** in healthy subjects. For this purpose, regular and superficial

needling of **L.I.-4** was compared. Changes were measured by using power spectral analysis and heart rate frequency analysis. Only deeper needling at L.I.-4 led to a significant activation of the sympathetic and parasympathetic nervous system during and after treatment as well as to a significant reduction of the heart rate frequency. **Haker et al, J Auton Nerv Syst 2000**

A US prospective, randomised, sham-controlled, single-blinded trial (n = 101) assessed the effect of varying intensities of transcutaneous electric acupoint stimulation (TEAS) at **L.I.-4** on the **postoperative patient-controlled analgesia** (PCA) requirement for **hydromorphine** (HM), the incidence of opioid-related side effects, and the recovery profile after lower abdominal surgery in women. In the sham-TEAS group HM-requirement was reduced by 23%, low-intensity TEAS produced a 34% decrease in the HM requirement while the high-intensity TEAS reduced requirement significantly by 65%, accompanied by a reduction in the occurrence of dizziness, nausea and pruritus. **Wang et al, Anesth Analg 1997**

A Chinese prospective, randomised, controlled trial (n = 110) investigated the effect of transcutaneous electric acupoint stimulation TEAS) at **L.I.-4**, **Ex-HN-4** ***(yuyao)*** and **G.B.-31** during enflurane anaesthesia in patients undergoing craniotomy. In comparison to the control group (anaesthesia maintained with enflurane only) the **minimum alveolar concentration** (MAC) of enflurane decreased in the TEAS-supplemented group by 38–47%. If anaesthesia was supplemented by TEAS plus scalp infiltration with procaine, there was an even higher reduction of 42–66%. The haemodynamic balance was also more stable during the operation, and postoperative recovery was faster. **Wang et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1994**

A German prospective, waitlist-controlled trial (n = 6) investigated the effect of electro-acupuncture at **L.I.-4**, **ST-25**, **LIV-3** and **BL-25** on stool frequency and colonic transit time in **chronic constipation**. There were no significant differences in all the parameters between the control group and the waitlist group. **Klauser et al, Z Gastroenterol 1993**

A Japanese experimental trial researched the influence of **L.I.-4 on vibration induced finger flexion reflex**. Unilateral acupuncture at L.I.-4 suppressed the reflex in both hands. **Takakura et al, Am J Chin Med 1992**

A Chinese prospective, randomised, matched controlled trial (n = 12) investigated the analgesic effect of aqueous acupuncture (injection of glucose solution) at **L.I.-4** and **G.B.-34** in **postoperative pain control**. In comparison to the control group, the intensity of postoperative pain as well as the requirement for analgesic medication were significantly lower in the study group. **Chen et al, Gaoxiong Yi Xue Ke Xue Za Zhi 1991**

A German prospective, non-randomised, sham-controlled trial (n = 36) investigated the **analgesic effect** of acupuncture at **L.I.-4**, **P-6**, **ST-36** and **SP-4** prior to a **colonoscopy**. The pain experienced by the acupuncture patients was significantly lower compared to the group receiving sham acupuncture. The analgesics and sedatives required were also significantly less. **Li et al, Dtsch Med Wochenschr 1991**

A prospective, randomised, single-blinded experimental study (n = 39) investigated the effect of acupuncture and laser acupuncture at **L.I.-4** and **Ex-UE** ***jianqian*** on the **pain threshold**. Compared to laser acupuncture, acupuncture significantly increased the pain threshold. **Brockhaus et al, Pain 1990**

A British prospective, controlled experimental cross-over study investigated the effectiveness of electro-acupuncture at **L.I.-4** on experimental **tooth pain**. After 30 minutes of electric stimulation the pain threshold increased by 27%. This increase could be partially blocked by naloxone. **Ernst et al, Acupunct Electrother Res 1987**

A British prospective, randomised, placebo-controlled study (n = 51) investigated the effect of acupuncture at **L.I.-4** as an **analgesic for operative dentistry**. There were no significant differences between the treatment group and the placebo group. **Taub et al, Oral Surg Oral Med Oral Pathol 1979**

A prospective non-randomised, placebo-controlled experimental trial (n = 40) investigated the effect of acupuncture at **L.I.-4** on the **pain threshold of postoperative dental pain** compared to codeine. Both **L.I.-4** and codeine had a significantly better effect than the placebo. The combination of both had the strongest effect. **Sung et al, Anesth Analg 1977**

An Austrian prospective, sham-controlled, single-blinded experimental trial (n = 12) measured the effect of **L.I.-4** and **P-6** on **pain threshold and pain tolerance**. In contrast to sham acupuncture, true acupuncture significantly raised the pain threshold. **Stacher et al, Am J Chin Med 1975**

L.I.-5 → P-5

L.I.-6 → P-5

L.I.-7 → P-5

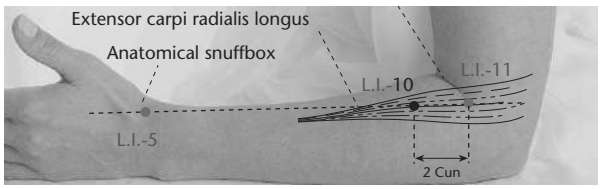

L.I.-10 → L.I.-4; Du-12; ST-36

A Swedish prospective, randomised, controlled, double-blinded trial (n = 49) investigated the effect of laser acupuncture on **L.I.-10**, **L.I.-11**, **L.I.-12**, **LU-5** and **T.B.-5** on **lateral humeral epicondalgia (tennis elbow)**. No significant difference occurred at a dose of 0.36 J/point. **Haker et al, Pain 1990**

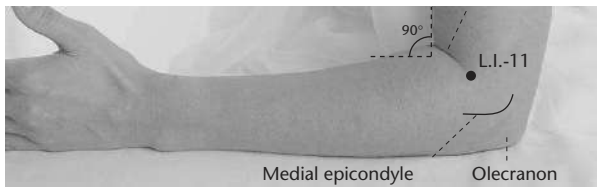

L.I.-11 → L.I.-4; ST-36; LIV-3; L.I.-10; L.I.-15

A Taiwanese prospective, randomised, placebo-controlled, single-blinded trial (n = 40) investigated the effect of acupuncture

at **L.I.-11** on **uraemic pruritus**. The results of the pruritus score questionnaires demonstrated that only true acupuncture at **L.I.-11** resulted in a significantly reduced pruritus lasting for a period of three months. Uraemia levels remained unchanged. **Che-yi et al, Nephro Dial Transplant 2005**

A German prospective, randomised, placebo-controlled, single-blinded experimental trial ($n = 22$) investigated the effect of **L.I.-11** on the **phagocytic immune system**. Based on laboratory parameters, the **activity of neutrophils** was significantly higher in the study group. **Karst et al, Complement Ther Med 2003**

A Swedish prospective, non-randomised, placebo-controlled, single-blinded experimental longitudinal trial investigated the influence of electro-acupuncture at **L.I.-11** and **L.I.-4** on sympathetic nerve activity and the pain threshold. Only electro-acupuncture at L.I.-11 and L.I.-4 increased the pain threshold in combination with increased sympathetic nerve activity. **Knardahl et al, Pain 1998**

A Swedish prospective, randomised, placebo-controlled, single-blinded trial ($n = 58$) investigated the effect of laser acupuncture at **L.I.-11** and **L.I.-12** for **lateral humeral epicondalgia (tennis elbow)** compared to laser application on the painful area. There were no significant differences between the laser and placebo treatments after the treatment period. **Haker et al, Arch Phys Med Rehabil 1991**

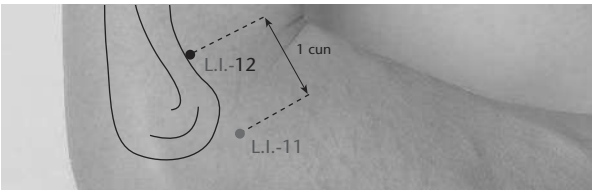

L.I.-12 → L.I.-11; L.I.-10

A Chinese prospective, randomised, controlled trial ($n = 64$) investigated the effect of electro-acupuncture at **L.I.-12**, **T.B.-5**, **G.B.-30** and **ST-36** accompanied by medication for **acute cerebral infarction** with regard to the plasma and CSF somatostatin levels. The control group received medication only. Both plasma and CSF somatostatin levels increased significantly in the acupuncture group. **Zhang et al, J Tradit Chin Med 1999**

A Chinese prospective, randomised, controlled trial ($n = 64$) investigated the effect of electro-acupuncture at **L.I.-12**, **T.B.-5**, **G.B.-30** and **ST-36** on the plasma and CSF levels of VIP (vasoactive intestinal peptide), somatostatin and pancreatic polypeptide in patients with **acute cerebral infarction**. The CSF VIP level dropped significantly while the plasma pancreatic polypeptide level increased significantly. In patients with a good response both plasma and CSF somatostatin levels increased significantly after acupuncture. **Zhang et al, Zhen Ci Yan Jiu 1996**

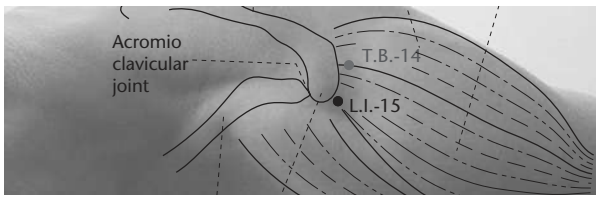

L.I.-15 A Taiwanese prospective, randomised, controlled, non-blinded trial ($n = 150$) investigated the effect of electro-acupuncture at **L.I.-15**, **G.B.-21** and **Ex-UE *jianneiling*** with or without regional nerve block (RNB) (stellate ganglion and suprascapular plexus) for **'frozen shoulder' (adhesive capsulitis)**. This was compared to RNB treatment alone. The combination of electro-acupuncture and RNB produced the most significant reduction of pain and duration of pain relief as well as improvement of range of motion. **Lin et al, Acta Anaesthesiol Sin 1994**

A Chinese prospective, randomised, controlled, non-blinded trial ($n = 63$) investigated the effect of acupuncture at **L.I.-15**, **L.I.-11**, **L.I.-4**, **G.B.-30**, **G.B.-34**, **G.B.-37**, **LIV-3**, **G.B.-20**, **ST-25**, **ST-40**, **ST-36**, **SP-6** and **KID-3** on symptoms of **cerebral infarction** compared to the treatment with the calcium channel blocker verapamil. The effectiveness in the acupuncture group was 94%, in the verapamil group 84%. This difference is statistically significant. However, the choice of a calcium channel blocker seems unusual for this indication. **Zou et al, Zhong Xi Yi Jie He Za Zhi 1990**

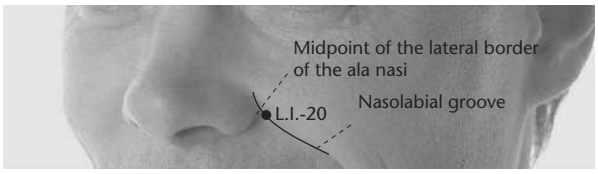

L.I.-20 → G.B.-14

A British prospective, randomised, controlled trial ($n = 20$) investigated the effect of a 30 second massage at **L.I.-20** on **nasal airflow** compared to an untreated control group. Symptoms improved significantly in the massage group. **Takeuchi et al, Am J Rhinol 1999**

Stomach channel

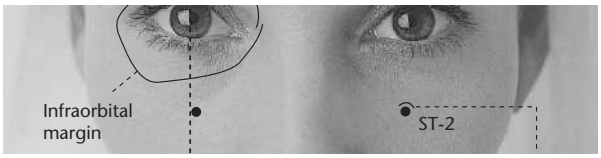

ST-2 → G.B.-14

A prospective, non-randomised experimental trial ($n = 33$) investigated the effect of acupuncture and electro-acupuncture at **ST-2**, **L.I.-4** and **ST-44** on experimentally induced **tooth pain**.

While both procedures increased the pain threshold, the increase was not sufficient to justify acupuncture as a means of pain control in conservative dentistry. **Bakke, Scand J Dent Res 1976**

ST-4 → G.B.-14

ST-6 → G.B.-14; L.I.-4

ST-7 → L.I.-4

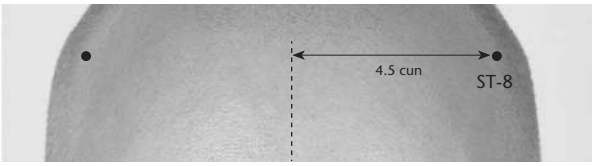

ST-8 A large prospective, randomised, multi-centre, group-comparison trial (n = 120) in patients suffering from **migraine without aura** demonstrated that acupuncture at **ST-8**, **G.B.-5**, **G.B.-20**, **Du-14** and **LU-7** was significantly more effective than conventional drug therapy, regarding both frequency and duration of the migraine attacks as well as number of sick days. The authors calculated that by treating migraine with acupuncture the Italian health system could save approximately 0.5 billion Euros per year. **Liguori et al., J Tradit Chin Med 2000**

A Bulgarian case study (n = 35) reports the successful treatment of headaches following caesarian sections with spinal anaesthesia with acupuncture at ST-8, L.I.-4, G.B.-11, G.B.-14, BL-10, Du-14 and Du-20 during 1–3 treatments. **Tsenov, Akush Ginekol (Sofia) 1996**

A Swedish prospective, placebo-controlled trial investigated the effect of intrasegmental (**ST-8**) and extrasegmental (**L.I.-4**) electro-acupuncture on sensory thresholds. Only the pain threshold could be increased, and this only with intrasegmental acupuncture at ST-8. **Lundeberg et al, Am J Chin Med 1989**

ST-18 → G.B.-14; Du-1

ST-20 → T.B.-8

ST-21 → Du-20

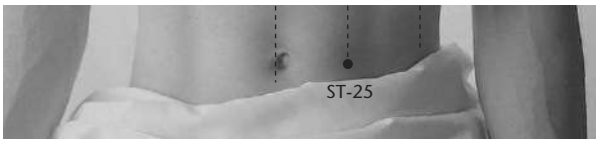

ST-25 → Du-1; Du-20; L.I.-4; L.I.-15

A Chinese prospective, randomised, controlled trial (n = 62) investigated the effect of acupuncture and moxibustion at ST-25 and Ren-4 for chronic colitis. This was compared to modern drug therapy. The acupuncture group showed the same therapeutic results but with fewer side effects. **Yang et al, J Tradit Chin Med 1999**

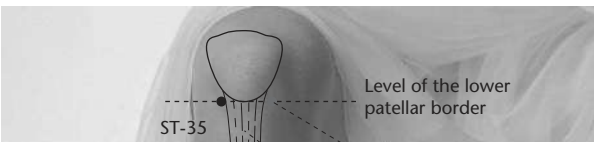

ST-32 → G.B.-31

ST-34 → ST-36; SP-9

ST-35 A Chinese prospective, randomised, single-blinded, controlled, long-term trial (n = 24) investigated the effect of electro-acupuncture (EA) and transcutaneous electrical nerve stimulation (TENS) at **ST-35** and **Ex-LE-4** (eight treatments over a period of two weeks) in older patients (average age: 85 years) suffering from **painful osteoarthritic knees**. This was compared to conventional therapy for this disorder. Both TENS and EA significantly reduced the knee pain. EA also had an effect on range of motion. **Ng et al, Altern Complement Med 2003**

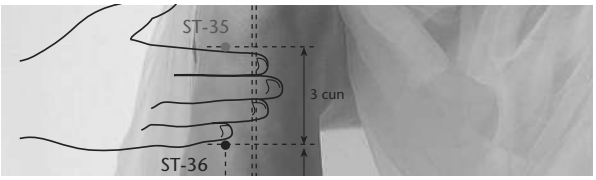

ST-36 → L.I.-4; BL-27; P-6; T.B.-8; G.B.-26; G.B.-31; Du-1, Ex-B-3; Du-4; Du-20; LIV-3; SP-9; BL-60; L.I.-12; L.I.-15; Ren-17

An Iranian prospective, randomised, placebo-controlled, single-blinded trial (n = 51) investigated the effect of transcutaneous electric nerve stimulation (TENS) at **ST-36** and **ST-38** on nitroglycerin-induced **hypotension under general anaesthesia with halothane**. With true acupuncture at the above points, the mean arterial blood pressure (MAP) was significantly lower (60 vs 66 mmHg). The effect of acupuncture was significantly greater with lower dosages of nitroglycerin compared with higher dosage. In addition, the time to reach steady state MAP was significantly shorter (10 vs 15.8 minutes) and the quality of operative ischaemia was excellent in 83.3% subjects compared to 25% in the sham group. **Saghaei M. et al, Acta Anaesthesiol Taiwan 2005**

A US prospective non-blinded experimental trial investigated the response to acupuncture at **ST-36** as evidenced by **fMRI**. The limbic and paralimbic structures of the cortical and subcortical regions in the telencephalon, diencephalon, brainstem and cerebellum were activated in response to acupuncture. **Hui et al, Neuroimage 2005**

A prospective, randomised placebo-controlled single-blinded trial (n = 13) investigated the effect of electro-acupuncture at **ST-36** on the **heart rate variability** (HRV) in healthy subjects. There was no acupuncture-specific effect on HRV. **Chang et al, Am J Chin Med 2005**

A German prospective randomised placebo-controlled single-blinded experimental crossover study (n = 10) investigated the effect of acupuncture at **ST-36**, **L.I.-11**, **SP-10 and Du-14** on

leukocyte circulation in healthy young subjects. There was a significant decrease in leukocyte and lymphocyte values in the acupuncture group while cortisol and norepinephrine plasma levels remained unchanged. **Kou et al, Brain Behav Immun 2005**

A US prospective, non-randomised, placebo-controlled, single-blinded trial ($n = 7$) investigated if electro-acupuncture at **ST-36, SP-6**, **ST-34** and **L.I.-4** before and during induction of general anaesthesia (desflurane) influenced the **anaesthetic requirement** compared to a control group. There was no reduction in pain in the treatment group. **Chernyak et al Anesth Analg 2005**

A Taiwanese prospective, randomised, placebo-controlled, single-blinded experimental cross-over study ($n = 15$) investigated if electro-acupuncture (EA) at **ST-36** can normalise **atropine-induced gastric dysrhythmia**. While there was a significant increase in the percentage of normal frequency, atropine-induced gastric dysrhythmia was not normalised by EA (no significant differences in the EGG). **Chang et al, Dig Dis Sci 2002**

A Taiwanese prospective, randomised, controlled experimental trial ($n = 15$) investigated the effect of **ST-36** on **gastric myoelectrical regularity**. There was a significant decrease in the tachygastric and bradygastric rhythm during electro-acupuncture on ST-36. **Chang et al, Digestion 2002**

A Taiwanese prospective, randomised, sham-controlled trial ($n = 100$) investigated the effect of electro-acupuncture at **ST-36** on **postoperative pain and opioid-related side effects**. The postoperative pain, the requirement for morphine (via PCA) and the incidence of nausea and dizziness were all significantly reduced. **Lin et al, Pain 2002**

A US prospective, randomised, double-blinded, sham-controlled experimental trial ($n = 14$) investigated if electro-acupuncture at **ST-36**, **G.B.-34** and **BL-60** could reduce the anaesthetic requirement (desflurane) for **experimental pain stimuli**. Electro-stimulation of these points did not reduce desflurane requirements compared to the control group. **Morioka et al, Anesth Analg 2002**

A Taiwanese prospective case study ($n = 15$) investigated the effect of bilateral electro-acupuncture **at ST-36** on gastric slow waves in **diabetic patients** with symptoms suggesting **gastric motor dysfunction**. Changes of gastric activity were measured by EGG. During and after acupuncture there was a significant increase in the percentages of normal frequency. In addition the percentage of tachygastric frequency was decreased significantly during and after acupuncture. **Chang et al, Digestion 2001**

A Chinese case study ($n = 104$) reports the successful injection-acupuncture at **ST-36** for **chemotherapy-induced leukopenia**. **Yin et al, J Tradit Chin Med 2001**

A Chinese prospective, randomised, controlled, non-blinded trial ($n = 26$) investigated the effect of acupuncture at **ST-36** plus auricular plaster therapy on **peristalsis** for postoperative recovery of intestinal function after abdominal surgery. 92% of patients in the treatment group showed recovery of normal peristalsis within 72 hours, while in the control group only 46% patients recovered within the same time-period. **Wan et al, J Tradit Chin Med, 2000**

A US prospective, randomised, controlled pilot study ($n = 17$) investigated the effect of acupressure at **ST-36** and **P-6** on the intensity and frequency of **nausea** in patients undergoing **chemotherapy due to breast cancer**. Both the frequency and intensity of the nausea was significantly lower in the treatment group than in the control group. **Dibble et al, Oncol Nurs Forum 2000**

An Italian prospective, randomised, controlled experimental trial ($n = 120$) investigated the effect of acupuncture at **ST-36** and **L.I.-4** regarding changes in the levels of beta-endorphins and other parameters (VIP, lymphocyte subsets, NK cells and monocyte phagocytosis) in patients suffering from various painful disorders. These changes were compared to an untreated control group. Only in the acupuncture group were the endorphin levels, the CD3 and CD4 values and monocyte phagocytosis increased for at least 24 hours after treatment. At the same time there was an increase of the CD8 values. **Petti et al, J Tradit Chin Med 1998**

A Chinese retrospective case study reports good results in treating **leukopenia** with acupuncture at **ST-36**. **Wei, J Tradit Chin Med 1998**

A Taiwanese prospective, randomised, sham-controlled experimental trial ($n = 18$) investigated the effect of acupuncture at **ST-36** and **L.I.-4** on central nervous system pathways with functional MR imaging of the brain. Besides a significant reduction of the heart rate frequency, acupuncture at both points resulted in activation of the hypothalamus and nucleus accumbens (structures of the descending antinociceptive pathway) and deactivation of the rostral part of the anterior cingulate cortex, amygdala formation, and hippocampal complex (limbic areas of pain recognition). Therefore this study provides explanations for the **pain-reducing effect** of acupuncture. **Wu et al, Radiology 1999**

A prospective, sham-controlled, single-blinded experimental longitudinal trial ($n = 13$) investigated the effect of acupuncture at ST-36 only, compared to acupuncture at **ST-36 plus L.I.-10** on **auditory endogenous potentials** (P300). The effect of the different forms of acupuncture was identical. In contrast to the sham group, it led to a decrease of P300 amplitude in the acupuncture group. **Hsieh et al, Am J Chin Med 1998**

A US prospective, randomised, sham-controlled, single-blinded trial ($n = 100$) investigated the effect of transcutaneous electric nerve stimulation (TENS) at **ST-36** on **opioid analgesic requirement** by PCA following hysterectomy or myomectomy compared to peri-incisional dermatomal stimulation by TENS. TENS applied at the dermatomal level of the incision is as effective as stimulation at ST-36, and both were more effective than stimulation at a sham location. **Chen et al, Anesth Analg 1998**

A Chinese prospective, randomised, controlled trial ($n = 69$) compared the so-called immunotherapy at **ST-36** to conventional desensitisation therapy in patients suffering from **anaphylactic asthma**. Results in the acupuncture group were significantly better, regarding both clinical and immunohistochemical parameters. **Chen et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1996**

A Mongolian prospective, randomised, controlled study ($n = 48$) compared the effect of acupuncture at **ST-36** and **P-6** combined with epidural anaesthesia to simple epidural anaesthesia for subtotal gastrectomies. The acupuncture group, while requiring a lower dosage of anaesthetic, experienced a stronger analgesic effect, the abdominal muscles were more relaxed and the haemodynamic disturbances were smaller. **Sun, Zhen Ci Yan Jiu 1996**

A US prospective, randomised, controlled, experimental trial ($n = 11$) researched the effect of transcutaneous electric nerve stimulation (TENS) at **ST-36** and **ST-37** on **peripheral blood circulation** and **haemodynamics**. Twenty minutes of TENS on healthy subjects did not produce a significant change in the measured parameters. **Balogun et al, Disabil Rehabil 1996**

A Taiwanese prospective, randomised, placebo-controlled, single-blinded trial investigated the influence of acupuncture at **ST-36** on the **pulse spectrum**. The results indicated that acupuncture at ST-36 has a specific effect on the Fourier components of the pulse: C2 and C4 decreased while C5, C6, C8 and C9 increased. This specific frequency effect was not found when acupuncture was applied on a non-acupuncture point. **Wang et al, Am J Chin Med 1995**

A Chinese prospective, randomised, controlled trial ($n = 45$) investigated the effect of **ST-36**, **L.I.-11** and **KID-6** on the immunoactivity of natural killer cells and the interleukin-2 level in **malignant tumour** patients. After receiving one treatment of 30 minutes daily for 10 days, parameters in the treatment group had increased significantly compared to the control group. **Wu et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1994**

A Chinese prospective, randomised, controlled trial ($n = 80$) compared **anaesthesia with electro-acupuncture** at **ST-36** and **L.I.-4** to epidural anaesthesia in **appendectomies**. The operations were equally successful in both groups, without any significant differences regarding the success rate. **Sun et al, Zhen Ci Yan Jiu 1992**

A Chinese prospective, randomised, controlled trial ($n = 39$) investigated the effect of acupuncture at **ST-36** and **SP-6** on **bowel movements following abdominal surgery**. In the acupuncture group the first postoperative bowel movement occurred after approximately 58 hours, but only after 86 hours in the untreated control group. **Liu et al, Zhong Xi Yi Jie He Za Zhi 1991**

ST-37 → ST-36

ST-38 → G.B.-34; ST-36

ST-40 → L.I.-15

ST-44 → L.I.-4; ST-2

Spleen channel

SP-1 A US prospective experimental trial ($n = 13$) investigated the effect of electro-acupuncture (EA) at **SP-1** and **LIV-1** on **thermal pain thresholds**. The pain threshold increased significantly 30 seconds after onset of EA. The authors postulate that EA at the above points has an inhibitory effect on the C-fibre afferents; the analgesic benefit observed is most probably A-delta afferent mediated. **Leung et al, J Altern Complement Med 2005**

SP-4 → Du-4; P-6; L.I.-4

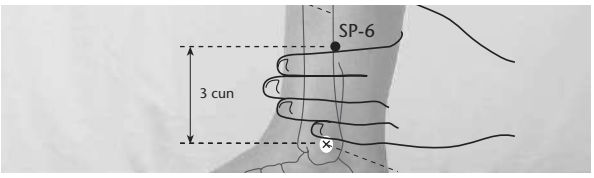

SP-6 → ST-36; L.I.-4; P-6; G.B.-26; Du-4; Du-20; LIV-3; L.I.-15

In a prospective, randomised, double-blinded group-comparison trial ($n = 56$) patients suffering from **interstitial cystitis** performed daily laser therapy on SP-6 at home for 30 seconds over a period of 12 weeks. There were no significant differences between the treatment and control cohorts. **O'Reilly et al, J Urol 2004**

An Australian prospective, randomised, placebo-controlled trial ($n = 20$) investigated the effect of transcutaneous electrical stimulation (TENS) at **SP-6** and **LIV-3** on **uterine contractions** in post-date pregnant women. A significant increase in frequency and strength of contractions was found in the TENS group compared with the placebo group. **Dunn et al, Obstet Gynecol 1989**

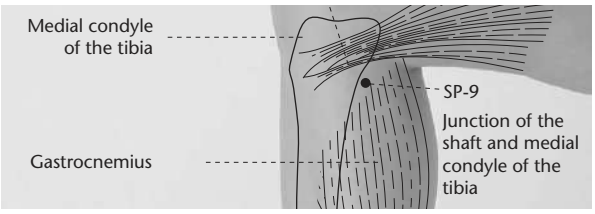

SP-9 A British prospective, randomised, controlled, assessor-blinded trial ($n = 44$) investigated the influence of **SP-9**, **SP-10**, **ST-34**, **ST-36** and **L.I.-4** on unilateral versus bilateral acupuncture in patients with **advanced osteoarthritis of the knee**. The symptoms improved significantly in both groups. This improvement was sustained for six months. There was no statistically significant difference between the groups. **Tillu et al, Acupunct Med 2001**

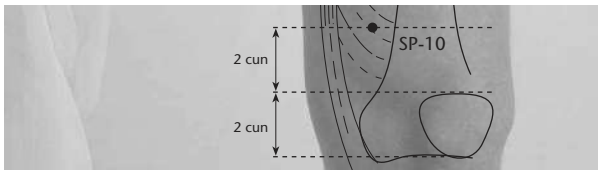

SP-10 → SP-9; ST-36; Ren-17

A Chinese prospective, randomised, controlled trial ($n = 62$) investigated the effect of Q-wave millimetre microwave application at **SP-10** and **BL-17** on **chemotherapy-induced leukopenia in gastrointestinal cancer patients**. One group received

irradiation before the start of chemotherapy, the other group only after the onset of the chemotherapy-induced leukopenia. Beginning the irradiation before starting chemotherapy produced a significantly better result (86% versus 73% with later application). **Wu et al, Zhongguo Zhong Xi Yi Jie He za Zhi 1997**

SP-12 → Du-4

Heart channel

HE-3 → P-6

HE-5 → P-6

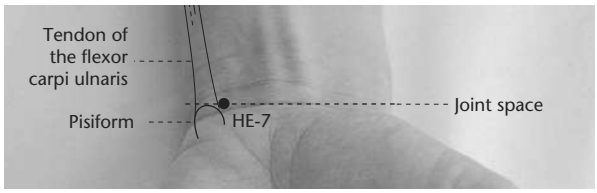

HE-7 → G.B.-26; Du-4; Du-20; S.I.-3; P-6

A German prospective, randomised, placebo-controlled, single-blinded trial (n = 36) investigated the effect of **H-7**, **P-6**, **Du-20**, **BL-62** and **Ex-HN-6** ***erjian*** on the cardiac autonomous nervous system in patients with **minor depression and anxiety disorders**. Only the acupuncture group demonstrated a relative increase of cardiovagal modulation of heart rate. **Agelink et al, Fortschr Neurol Psychiatr 2003**

A British pilot study (n = 17) investigated the effect of **HE-7** on certain **stress parameters**. There was an average decrease of 44% (based on the Edinburgh Postnatal Depression Scale (EPDS). **Chan et al, Acupunct Med 2002**

HE-8 → LU-10

HE-9 → Du-4

Small Intestine channel

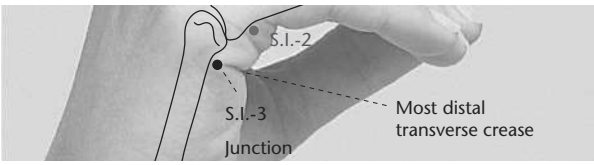

S.I.-3 → LU-10; Du-12

A larger-scale Taiwanese prospective, controlled, randomised trial (n = 100) demonstrated a significant improvement of neurological functioning in patients with **severe spinal cord injuries** (ASIA status A and B) after electro-acupuncture at **S.I.-3** and **BL-62** plus auricular acupuncture versus the control group (conventional and rehabilitation therapy only). **Wong et al, Am J Phys Med Rehabil 2003**

A non-controlled pilot study (n = 18) found that the treatment of **acute torticollis** with a single treatment of ipsilateral acupuncture at **S.I.-3** and **M-UE-8** resulted in a mean improvement of lateral head rotation of 53%. **Samuels N, Am J Chin Med 2003**

A Chinese prospective, randomised, controlled trial (n = 62) investigated the effect of acupuncture at **S.I.-3** and **HE-7** in patients with **cerebral traumatic dementia** compared to conventional physiotherapy. There were significant changes of MMSE scores and auditory P300 values in the acupuncture group only. **Zhang et al, Zhen Ci Yan Jiu 1996** and **Zhang et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1995**

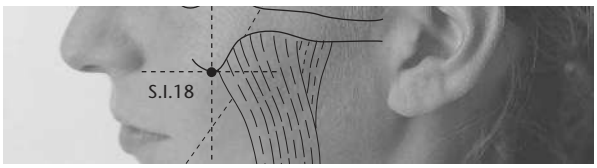

S.I.-18 A Chinese prospective, randomised, controlled, single-blinded trial (n = 42) investigated the effect of electro-acupuncture at **S.I.-18**, **Ex-HN-04**, **G.B.-20** and **L.I.-4** on controlled hypotension induced by isoflurane during a cerebral operation in order to avoid blood loss. The concentration of isoflurane necessary to achieve controlled hypotension was significantly reduced by 31–42% compared to the control group. **Wang et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 2000**

Bladder channel

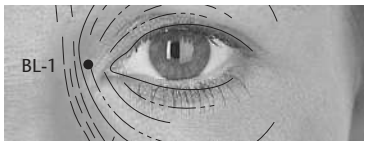

BL-1 → G.B.-14

A non-randomised group-comparison trial (n = 34) investigated the effect of warmed needle acupuncture at **BL-1** in patients suffering from **epiphora** due to dysfunction of the lacrimal duct. The effect in the treatment group was significantly better in comparison to the control group treated with lacrimal duct irrigation and norfloxacin eye drops (92% vs 54%). **Ni et al, J Tradit Chin Med 2002**

A Chinese case study reports the treatment of **epiphora** due to **insufficiency of lacrimal passage** with acupuncture at **BL-1**. Of the 68 eyes treated in 42 patients, 28 eyes were cured and 35 improved. 34 eyes improved after only one course of treatment. **Ni Y et al, J Tradit Chin Med 1999**

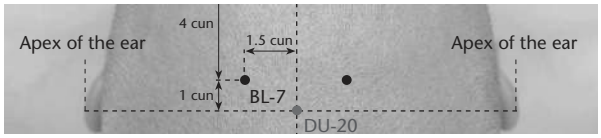

BL-7 A Chinese experimental trial carried out in 1988 was the first to discuss the possible effect of **BL-7**, **G.B.-6** and **Ex-HN-1** on **hemiplegia** through changes of the micro-circulation in the nail bed. **Shun et al, Zhen Ci Yan Jiu 1988**

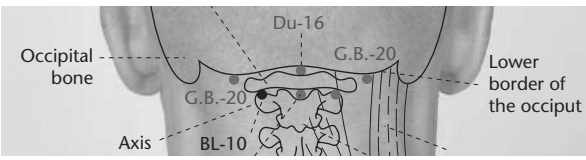

BL-10 → ST-8
A placebo-controlled, non-blinded group-comparison trial (n = 65) investigated the effect of treating children with an acu-plaster at **BL-10**, **BL-11** and **G.B.-34** the night before undergoing **strabismus surgery,** for the prevention of **post-operative nausea and emesis**. Nausea and emesis were reduced by 50% during the early and later postoperative stage compared to the placebo-group. The well-researched point P-6 often failed to prevent nausea in this type of surgery. **Chu et al, Acta Anaesthesiol sin 1998**
BL-11 → Du-12; BL-10; L.I.-4; Ex-B-1
BL-12 → Ex-B-1
BL-13 → Ex-B-1; Du-14

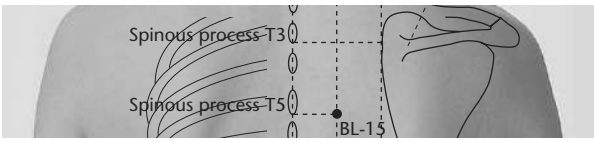

BL-15 → Ex-B-1; P-6
In an animal experiment, moxibustion at **BL-27** was able to reduce the renal excretion of Na^+ while at the same time decreasing urinary volume. Systolic blood pressure remained unchanged while plasma levels of aldosterone and ANP (atrial natriuretic peptide) decreased. In contrast, moxibustion at BL-15 increased urinary volume while reducing excretion of Na^+ and decreasing systolic blood pressure. Plasma levels of aldosterone and ANP decreased. These results suggest that BL-15 and BL-27 might be applied for certain forms of **hypertension**. **Lee et al, Am J Chin Med 1997**
BL-17 → SP-10

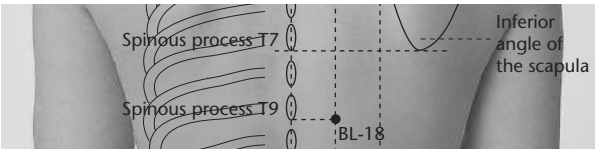

BL-18 → BL-23
A Japanese prospective, randomised, placebo-controlled, double-blinded trial (n = 189) investigated the effect of Japanese intradermal acupuncture at **BL-18**, **BL-19**, **BL-20**, **BL-21**, **BL-22**, **BL-23**, **BL-24**, **BL-25** and **BL-26** on **analgesic requirement, nausea and emesis** as well as stress markers after **abdominal surgery**. Starting from the recovery room and including the second postoperative day incisional pain at rest and during coughing was significantly lower in the verum group than in the control group. The requirement for epidural morphine was also significantly lower (only 50%). Nausea and emesis were also significantly lower (20–30%). Plasma cortisol and epinephrine concentrations were reduced by 30–50% in the acupuncture group during recovery and on the first postoperative day. This is one of the most valuable trials due to its excellent methodology and design. **Kotani et al, Anesthesiology 2001**
BL-19 → BL-18
BL-20 → L.I.-4; Du-1; BL-18; Du-14; P-6
BL-21 → BL-18
BL-22 → BL-18; Ex-B-2
BL-23 → L.I.-4; Du-4; Du-20; BL-18; Du-14
BL-24 → L.I.-4; BL-18
BL-25 → L.I.-4; BL-18; Ex-B-2
BL-26 → L.I.-4; BL-18

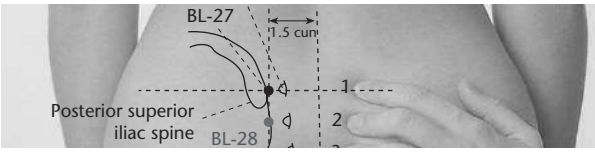

BL-27 → L.I.-4; BL-15
A German randomised, placebo-controlled experimental trial (n = 42) compared the effect of acupuncture at **BL-27** on pain reduction with a point of the Yamamoto New Scalp Acupuncture (YNSA). Experimental pain stimuli were set on the **upper calcaneus edge**. The difference in pain reduction between the acupuncture group and the scalp acupuncture group was highly significant. There are also highly significant differences concerning the verum and the placebo treatment. BL-27 could therefore be a useful point for treating calcaneal pain. **Ogal et al, Anasthesiol Intensivmed Notfallmed Schmerzther 2002**
An animal experiment in 15 healthy dogs investigated the effect of acupuncture on intestinal (duodenum) motility. Electro-acupuncture at BL-27 decreased the frequency of intestinal motility by 31% during stimulation, while showing an increase of 18% after stimulation. Electro-stimulation of ST-36 led to the opposite result: a 20% increase during stimulation followed by a decrease of 7% after stimulation. Based on this result, BL-27 could be applied for **paralytic ileus** or **constipation** while ST-36 could be used for treating **bloating, tenesmus and diarrhoea**. **Choi et al, J Vet Sci 2001**
BL-28 → L.I.-4; Du-4

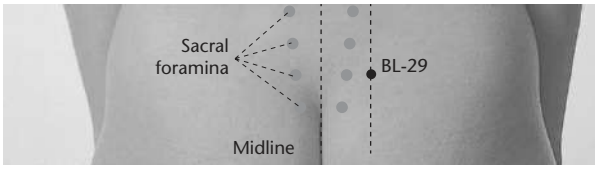

BL-29 There exists only a small case study describing **various disorders**. This is of educational value rather than anything else. **Chen Y., J Tradit Chin Med 2002**

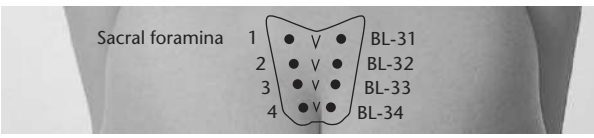

BL-32 → Du-4; Ren-3
A medium-scale Chinese case study ($n = 30$) reports the successful treatment of **urinary and faecal incontinence** due to nerve damage with electro-acupuncture at **BL-35** and **BL-32**. **Yang et al, J Tradit Chin Med 2003**

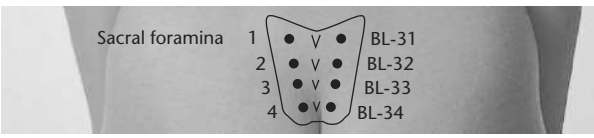

BL-33 A Japanese case study ($n = 10$) documents a significant reduction of pain for **chronic pelvic pain syndrome** with **intrapelvic venous congestion** by weekly treatments with bilateral acupuncture at **BL-33** (needle retention 10 minutes, manual rotation). **Honjo et al, Int J Urol 2004**
A Japanese prospective case study ($n = 13$) investigated the effect of bilateral acupuncture at **BL-33** on cystometry and **incontinence** in **spinal-cord-injured patients**. Incontinence completely disappeared in 15% of cases and decreased to 50% or less compared to baseline in a further 46% of patients. Maximum cystometric bladder capacity increased significantly from 75 ml to 148 ml. When cystometry was repeated in six patients one month after the last acupuncture treatment, bladder capacity showed a further average increase to 187 ml. **Honjo et al, Urol Int 2000**

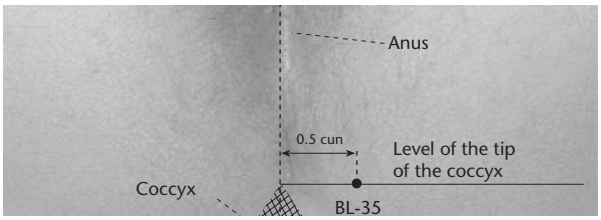

BL-35 → BL-29; BL-32
A larger-scale Chinese case study ($n = 103$) reports the successful treatment of **post-partum urinary retention** with classic acupuncture at **BL-35**. **Li et al, J Tradit Chin Med 1996**
BL-47 → L.I.-4

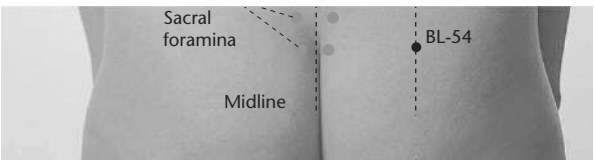

BL-54 → Du-3
A larger-scale Chinese case study ($n = 100$) reports the successful treatment of **lumbar pain** with acupuncture at **BL-54**. **Cui, J Tradit Chin Med 1992**

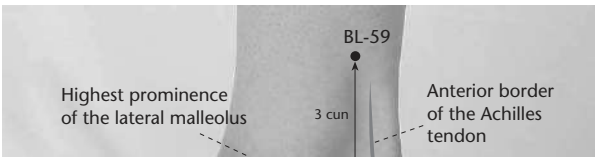

BL-59 A smaller-scale Chinese case study reports the treatment of **acute lumbar pain** with acupuncture at **BL-59**. **Hu, J Tradit Chin Med 1993**

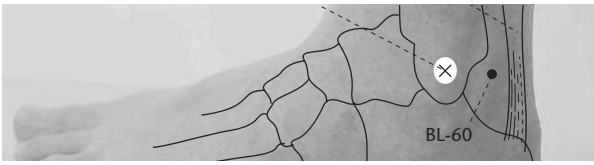

BL-60 → Du-3; ST-36; L.I.-4
A US prospective case study ($n = 7$) reports the positive subjective (questionnaire) and objective (tibial H-reflex) effect of non-invasive electro-acupuncture at **BL-60**, **ST-36**, **KID-1** and **LIV-3** on **HIV-related neuropathy**. Results showed a significant overall improvement in functional activities. **Galantino et al, J Altern Complement Med 1999**
BL-62 → S.I.-3; HE-7; Du-20

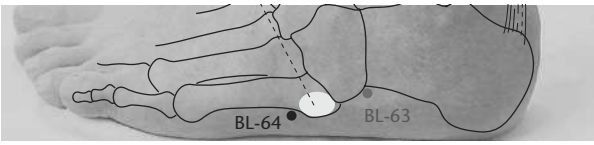

BL-64 An animal experiment demonstrated that electro-acupuncture at BL-64 and BL-65 increased the activity of nitric oxide synthase in the brain stem nuclei (nucleus gracilis). **Ma at al, Acupunct Electrother Res 2002**
BL-65 → BL-64

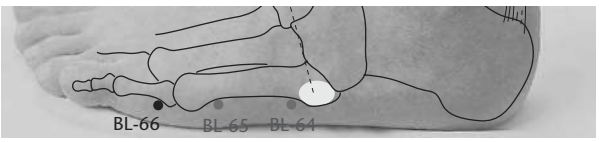

BL-66 An experimental study in rats investigated the antipyretic effect of manual acupuncture at **BL-66** on lipopolysaccharide-induced **fever**. Acupuncture at this point reduced the fever and also reduced the production of the pro-inflammatory cytokines interleukin-1 beta and interleukin-6 in the hypothalamus of the treated rats. **Son et al, Neurosci Lett 2002**

BL-67 An Italian prospective, randomised, controlled, single-blinded trial (n = 123) investigated the effect of moxibustion at **BL-67** on the conversion rate in **breech presentations**. The trial had to be terminated prematurely due to a high drop-out rate of participants. At that time, there were no significant differences between the moxibustion group and the control group. The authors conclude that there is too little acceptance for moxibustion treatment in non-Asian cultures. **Cardini et al, Br J Obstet Gynaecol 2005**

A randomised, controlled, prospective larger-scale trial (n = 67) demonstrated a higher rate of conversion of fetal **breech presentations** in a group treated with 30 minutes of manual acupuncture at **BL-67** (76.4%) compared to an untreated control group (45.4%). **Habek et al, Fetal Diagn Ther 2003**

A large-scale (n = 260), randomised, controlled, prospective trial demonstrated a higher rate of conversion of fetal **breech presentations** in women treated with moxibustion at **BL-67** for 7–14 days (75.4%) than the conventionally treated control group (47.7%). **Cardini et al, JAMA 1998**

A medium-scale case study (n = 48) with retrospective control groups demonstrated that electro-acupuncture at **BL-67** for the conversion of fetal **breech presentations** had a 81.3% success rate. **Li et al, J Tradit Chin Med 1996**

Acupuncture at **BL-67** resulted in a significant increase in the number of **c-Fos-positive cells in the primary visual cortex** of binocularly deprived rat pups to levels suggesting that BL-67 has an influence on the activity of the primary visual cortex. **Lee et al, Am J Chin Med 2002**

A recent Italian prospective, randomised trial investigated the effect of acupuncture and moxibustion at **BL-67** on the conversion of **breech presentations** in 240 pregnant women. At delivery, breech presentation was significantly lower in the active-treatment group (37%) than in the observation group (54%). Therefore the proportion of caesarean sections indicated for breech presentation was also significantly lower in the treatment group than in the observation group (52% vs 67%). **Neri et al, J Matern Neonatal Med 2004**

A Croatian prospective, randomised, controlled, non-blinded group-comparison trial (n = 67) assessed the value of acupuncture at **BL-67** for 30 minutes in the conversion of **breech presentation**. The success rate of the acupuncture correction of the breech presentation was 76% while the rate in the untreated control group was only 45%. This is a highly significant difference. **Habek et al, Fetal Diagn Ther 2003**

An Austrian prospective, randomised, placebo-controlled experimental trial (n = 10) investigated the effect of laser acupuncture at **BL-67** on the visual cortex using fMRI to monitor **neuroradiological changes**. Only verum acupuncture resulted in an activation in the cuneus (Brodmann Area BA 18) and the medial occipital gyrus (BA 19). These results confirm the traditional application of this point for ophthalmic disorders. **Siedentopf et al, Neurosci Lett 2002**

An Italian prospective, non-randomised, sham-controlled, single-blinded trial (n = 12 pregnant women with **breech presentation**) demonstrated that acupuncture at **BL-67** significantly reduced the fetal heart rate while increasing fetal movement and producing more accelerations. **Neri et al, J Soc Gynecol Invest 2002**

Kidney channel

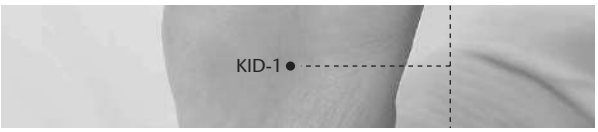

KID-1 → BL-60

A Chinese prospective, randomised, controlled, non-blinded trial (n = 150) investigated the effect of acupuncture at **KID-1** in the treatment of **children with bronchial asthma** compared to a control group treated with conventional drugs. The clinical effectiveness of acupuncture at KID-1 was significantly higher than treatment with drugs (89% vs 64%). In addition, in the acupuncture group the level of eosinophils and IgE was highly significant and significantly lower respectively than in the control group. **Gao et Zhu, Zhongguo Zhong Xi Yi Jie He Za Zhi 2005**

A Chinese retrospective case study reports the successful local stimulation of **KID-1** by applying a paste in the treatment of **essential hypertension**. **Gong et al, J Tradit Chin Med 1995**

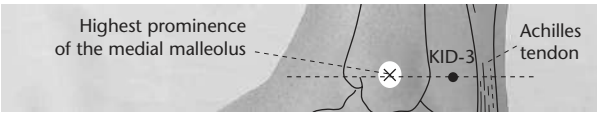

KID-3 → Du-4; L.I.-15

A Taiwanese prospective, randomised, controlled experimental trial (n = 157) compared the absorption of Tc-99 m (a radioactive technetium marker) at **KID-3** with non-acupuncture points. Absorption at KID-3 was significantly better than at non-acupuncture points. **Wu et al, Am J Chin Med 1994**

KID-4 → P-6
KID-5 → Du-4
KID-6 → ST-36

Pericardium channel

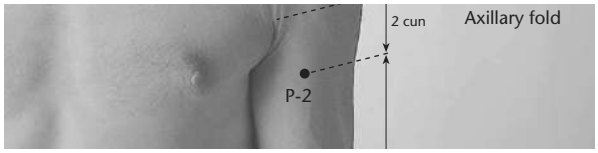

P-2 A Chinese experimental trial (n = 100) examined cardiac function before and during acupuncture at **P-2**, **P-3**, **P-4** and **P-6** as well as at control points (some arbitrary points and **G.B.-37**) in patients with **coronary heart disease**. In contrast to the control points the verum points had an effect – even if only a minor one – on the measured parameters (PEP, LVET, P/L, HI, SV, CO, ST segment and T wave of ECG). **You et al, Zhen Ci Yan Jiu 1993**
P-3 → P-2
P-4 → P-2

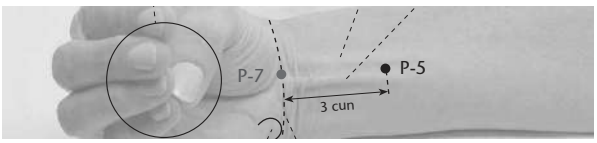

P-5 An earlier Chinese trial interpreted the synchronising effect of electro-acupuncture at **P-5** on EEG activity as having a **homeostatic effect on the whole body**. **Huang et al, Zhen Ci Yan Jiu 1990**
Electro-acupuncture at **P-5** improved the ST segment of ECG in rabbits with induced **acute myocardial infarction**. **Cao et al, Zhen Ci Yan Jiu 1990**
An earlier Chinese experimental trial (n = 10) found that electro-acupuncture at **P-5** and **P-6** shortened sino-atrial conduction in healthy subjects and increased frequency at the sinus node (**positive chronotropic and bathmotropic effect**). **Xi et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1993**
Previous work of the team led by Li and Tjen-A-Looi suggests that the inhibitory effect of electro-acupuncture (EA) on the pressor reflex induced by bradykinin (BK) applied to the gallbladder is in part due to the activation of opioid receptors most likely located in the rostral ventrolateral medulla (rVLM). This trial investigated the specific opioid receptor subtypes and neurotransmitters responsible for this inhibition. Therefore BK was applied to the gallbladder of anaesthetised cats to induce an increase of arterial blood pressure. It was found that applying EA at P-5 and P-6 activated the mu and delta opioid receptors located in the rVLM, thus preventing a rise of the arterial blood pressure for a longer period of time through the activating influence of the splanchnic nerve (for example due to stretching of the gallbladder or stomach). This experiment substantiated the positive effect of EA at P-5 and P-6, especially for myocardial ischaemia due to **coronary heart disease**. **Li et al, Auton Neurosci 2001; Li et al, Am J Physiol Regul Integr Comp Physiol 2002; Tjen-A-Looi et al, Auton Neurosci 2003**
In a US study 17 healthy subjects were subjected to weekly bicycle training for 3–4 weeks. Subjects were asked to perform the training with and without electro-acupuncture (EA) at **P-5**, **P-6**, **L.I.-4**, **L.I.-5**, **L.I.-6**, **L.I.-7**, **G.B.-37**, **G.B.-38** and **G.B.-39**. In 70% of subjects, EA at points on the Pericardium and Large Intestine channels led to an increase in maximal workload with decreased blood pressure. Points on the Gall Bladder channel had no effect. Further trials may investigate the application of these points in the treatment of hypertension induced by exercise stress. **Li et al, Clin Auton Res 2004**

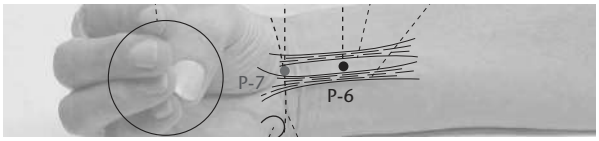

P-6 → P-2; P-5; Du-20; Du-14; HE-7; LIV-3; L.I.-4; ST-36
A Taiwanese prospective, randomised, placebo-controlled, double-blinded trial (n = 110) investigated the effect of acupressure bands at **P-6** on **nausea and vomiting during spinal anaesthesia for caesarean delivery**. The reduction and occurrence of nausea (64% vs 71%) and vomiting (22% vs 27%) was not significant. **Ho et al, Anesth Analg 2006**
A US prospective, randomised, placebo-controlled, single-blinded trial (n = 94) investigated the effect of acupressure bands at **P-6** on the occurrence of **nausea and vomiting during caesarean delivery under spinal anaesthesia**. There was no statistically significant difference between the active and sham control groups in the incidence of intraoperative nausea (30% vs 43%), postoperative nausea (23% vs 41%), intraoperative vomiting (13% vs 9%), postoperative vomiting (26% vs 34%), intraoperative antiemetic requirement (23% vs 18%), postoperative antiemetic requirement (34% vs 39%), complete intraoperative response (55% vs 57%) or complete postoperative response (51% vs 34%). These results did not reach the selected level of significance. There were also no significant differences between the groups in nausea scores, number of vomiting episodes and patient satisfaction. **Habib et al, Anesth. Analg 2006**
A Swedish prospective, randomised placebo-controlled experimental double-blinded trial (n = 60) investigated the effect of acupressure at **P-6** on **nausea induced by eccentric rotation**. Mean time to nausea was significantly the longest in the P-6 group (352 seconds), compared to the placebo-acupressure group (280 seconds) and the untreated control group (151 seconds). **Alkaissi et al, Can J Anaesth 2005.**
A Turkish prospective randomised controlled non-blinded trial (n = 90) investigated the effect of transcutaneous electrical acupoint stimulation (TEAS) at **P-6** and **Ren-13** compared to ondansetron on **postoperative nausea following paediatric**

tonsillectomy. There was the same incidence of nausea episodes in the TEAS group and in the ondansetron group, but significantly fewer episodes than in the untreated control group. Side-effects were significantly higher in the ondansetron group. **Kabalak et al, J Altern Complement Med 2005**

An Indian prospective, randomised, placebo-controlled, double-blinded trial ($n = 120$) investigated the effect of capsicum-plasters on **P-6** on **postoperative nausea and vomiting** (PONV) compared to conventional treatment with ondansetron. For the duration of the plaster application (6 hours after the operation) the incidence of PONV and the requirement for antiemetics were significantly lower in both groups. **Misra et al, Can J Anaesth 2005**

A Turkish prospective, randomised, placebo-controlled, single-blinded trial ($n = 127$) investigated the effect of transcutaneous electric nerve stimulation (TENS) at **P-6** on the incidence and severity of **nausea and gagging during gastroscopies**. There were no significant differences between the treatment group and placebo group. **Tarcin et al, Turk J Gastroenterol 2004**

A Chinese prospective, non-blinded, randomised, group-comparison trial ($n = 30$) investigated the effect of acupuncture at **P-6** plus mexiletin on **intermittent ventricular extrasystole**. The treatment group had a total effective rate of 90% compared with the control group, who was administered mexiletin only (80% total effective rate). Only the acupuncture group showed no increase in extrasystole. **Zhang et al, J Tradit Chin Med 2004.**

A Chinese group-comparison trial investigated the effect of acupuncture at **P-6** in the treatment of patients with angina pectoris and **acute myocardial infarct**. The effectiveness of acupuncture was significantly better (91%) than conventional treatment with isosorbide dinitrate and nifedipine. **Meng et al, J Tradit Chin Med 2004**

A US prospective, controlled, group-comparison trial ($n = 77$) investigated the effect of stimulation at **P-6** on the symptoms of **motion sickness** during exposure to optokinetic drum rotation. Despite the extreme test conditions, stimulating P-6 delayed the onset of symptoms considerably. **Miller et al, Aviat Space Environ Med 2004**

A Croatian prospective, single-blinded group-comparison trial ($n = 36$) investigated the efficiency of acupuncture at **P-6** in the treatment of **hyperemesis gravidarum**. Treatment results were measured by the requirement for anti-emetic medication. Acupuncture at P-6 had an efficiency rate of 90%, acupressure at P-6 64%, placebo acupuncture 12.5% and placebo acupressure 0%. **Habek et al, Forsch Komplementarmed Klass Naturheilkd 2004**

In a German prospective, randomised, placebo-controlled double-blinded group-comparison trial acupuncture at **P-6** significantly reduced the incidence of **emesis after gynaecological and breast surgery** (40% placebo, 25% P-6). Postoperative emesis was only reduced for gynaecological surgery, not for breast surgery. **Streitberger et al, Anaesthesia 2004**

A Chinese prospective, randomised, single-blinded, placebo-controlled group-comparison trial ($n = 41$) investigated the effect of twice daily acupressure at **P-6**, **ST-36** and **SP-6** on **gastrointestinal motility following trans-abdominal hysterectomy**. In comparison to the control group receiving acupressure on sham points, gastrointestinal motility (measured with a multifunctional stethoscope) increased significantly in the acupressure group. **Chen et al, Am J Chin Med 2003**

A German prospective, randomised, placebo-controlled, single-blinded trial ($n = 80$) investigated the anti-emetic effect of acupuncture at **P-6** in addition to administration of ondansetron in patients with **chemotherapy-related nausea**. There was no significant difference between the treatment and control group. **Streitberger et al, Clin Cancer Res 2003**

A US prospective, randomised, placebo-controlled, double-blinded trial ($n = 230$) investigated the effect of TENS at **P-6** for the relief of **nausea and vomiting in pregnancy**. Based on the Rhodes Index, results in the verum group were significantly better than in the control group. **Rosen et al, Obstet Gynecol 2003**

A British prospective, partially randomised, partially blinded, placebo-controlled trial ($n = 301$ patients with **acute myocardial infarction**, 125 of whom acted as a non-randomised control group) compared the effect of wristband-acupressure at **P-6** with placebo-acupressure. P-6 led to a significant reduction in **nausea and vomiting** during the last 20 hours of the 24-hour study phase (18%), compared with the placebo group (32%) and the control group (43%). **Dent et al, Complement Ther Med 2003**

A Swedish prospective case study ($n = 39$) investigated the effect of **P-6** combined with ondansetron for **nausea and vomiting** associated with **cyclophosphamide chemotherapy**. The authors stated that, compared with ondansetron treatment alone, the combined acupuncture-ondansetron treatment was significantly more effective, but they do not mention a control group. **Josefson et al, Rheumatology (Oxford) 2003**

In a US prospective, randomised, single-blinded trial ($n = 53$) the application of 'minute sphere' acupressure at **P-6**, **ST-36**, **SP-6** and **SP-4** showed no decrease of **postoperative pain and morphine requirement after abdominal surgery**. **Sakurai et al**, **Anesth Analg 2003**

A Swedish prospective, randomised, placebo-controlled, double-blinded, multi-centre trial ($n = 410$) investigated the effect of **P-6** for the treatment of **postoperative nausea and vomiting after gynaecological surgery**. The incidence of nausea and vomiting was significantly lower in the acupuncture group (33%) than in the control group (46%). **Alkaissi et al, Can J Anaesth 2002**

A US prospective, randomised, placebo- and sham-controlled, double-blinded trial ($n = 120$) investigated the effect of the prophylactic application of a 'ReliefBand' at **P-6** on the incidence of **nausea and vomiting after plastic surgery,** compared to the application of the ReliefBand in addition to 4 mg ondansetron. The occurrence of nausea and vomiting as well as the need for antiemetic 'rescue' medication was significantly the lowest in the P-6 plus ondansetron group. **White et al, Anesthesiology 2002**

A Korean prospective, randomised, placebo-controlled, double-blinded trial ($n = 160$) investigated among other parameters the

effect of a capsicum plaster applied at **P-6** on **postoperative nausea and vomiting**. The incidence of nausea and vomiting was significantly less (P-6: 26% within 24 hours, placebo: 57%). **Kim et al, Anesth Analg 2002**

A Japanese case study investigated the effect of different stimulation techniques at **P-6** on the coronary arteries in patients with **coronary heart disease**. The mean coronary dilation with acupuncture was 69% of that caused by isosorbide dinitrate. **Kurono et al, Am J Chin Med 2002**

A US prospective, randomised, sham-controlled, single-blinded trial ($n = 187$) investigated the effect of **P-6** on the incidence of **postoperative nausea and vomiting** in children. P-6 produced significantly better results than the sham points and was equally effective as administration of droperidol, but without the sedative, hypotensive side effects. **Wang et al, Anesthesiology 2002**

A US prospective, randomised, sham-controlled, single-blinded trial ($n = 27$) investigated the effect of **P-6** on the incidence of **nausea and vomiting during chemotherapy**. Compared to the control group, the anti-emetic requirement was significantly lower in the P-6 group. **Roscoe et al, Altern Ther Health Med 2002**

An Austrian prospective, randomised, placebo-controlled, experimental cross-over study ($n = 51$) investigated the effect of acupuncture at **P-6** on the **skin blood perfusion** measured by laser Doppler perfusion imaging. This was compared to needling a sham point. Changes in the skin blood perfusion occurred significantly earlier in the acupuncture group than in the sham group, with a basically more pronounced reduction in skin blood perfusion at the verum point. **Litscher et al, Lasers Med Sci 2002**

A large-scale Australian prospective, randomised, sham-controlled trial ($n = 593$) investigated the effect of acupuncture at **P-6** on the incidence **of nausea and vomiting in early pregnancy**. Women receiving traditional acupuncture reported significantly less nausea after the second treatment and significantly less dry retching after the third week compared with women in the control group. Women treated with sham acupuncture showed significantly less nausea and retching after the third treatment compared with untreated women. Individualised acupuncture resulted in significantly less nausea and retching already after the first treatment. In this study none of the treatments had a significant influence on the frequency of vomiting. **Smith et al, Birth 2002**

A US prospective, randomised, sham-controlled trial ($n = 120$) investigated the influence of electro-acupuncture (EA) at **P-6** on postoperative **nausea and vomiting** in paediatric patients who had undergone **ENT surgery** and were awake. EA at P-6 significantly decreased the incidence of nausea (P-6: 60%, sham: 85%, control group: 93%). While vomiting had the lowest incidence in the P-6 group, the difference from the other groups was not significant. **Rusy et al, Anesthesiology 2002**

A Swedish prospective, randomised, placebo-controlled pilot study ($n = 60$) investigated the effect of tuina (acupressure) at **P-6** on **pregnancy-related nausea and vomiting**. The frequency of both nausea and vomiting was significantly lower in the acupressure group compared to the sham acupressure group and untreated control group. **Werntoft et al, J Reprod Med 2001**

A US prospective, randomised, controlled trial ($n = 25$) investigated the effect of an acupressure wrist band at **P-6** on **motion sickness**. Both the symptoms of motion sickness and abnormal gastric activity as recorded via EEG were significantly lower in the group treated with wristbands at P-6. **Stern et al, Altern Ther Health Med 2001**

A US prospective, randomised, placebo-controlled, single-blinded trial investigated the effect of 'sea-bands' (acupressure wristbands) at **P-6** on the incidence of **postoperative nausea and vomiting**. There were no significant differences. **Windle et al, J Perianesth Nurs 2001**

A Swedish prospective, randomised, sham-controlled, single-blinded cross-over study ($n = 33$) investigated the effect of acupuncture at **P-6** on pregnant women with **hyperemesis gravidarum** (vomiting in pregnancy). The verum group showed a significantly faster reduction of the nausea, and also a higher number of the patients experienced no vomiting. **Carlsson et al, J Pain Symptom Manage 2000**

A Chinese prospective, controlled, randomised, group-comparison study ($n = 181$) and a Chinese case study ($n = 33$) demonstrated that electro-acupuncture at **P-6** and **P-8** as well as at **L.I.-4**, **T.B.-5**, **ST-36** and **SP-6** with de-escalating treatment frequency from four times daily to once weekly for a period of 15 and 28 days had a significant effect on the **withdrawal symptoms of heroin addicts**. **Zhang et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 2000; Wu et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 2000**

A Taiwanese prospective, non-randomised, placebo-controlled, single-blinded experimental cross-over study ($n = 44$) investigated the effect of acupuncture at **P-6** on the left ventricular ejection fraction (LVEF) in 22 healthy subjects and 22 patients with **coronary artery disease** (CAD). There was no change of LVEF in the healthy subjects while LVEF significantly increased in the CAD patients. **Ho et al, Am J Chin Med 1999**

A US prospective, randomised, placebo-controlled, single-blinded trial ($n = 100$) investigated the effect of pre-operative acupressure (verum group) and intra-operative acupuncture (verum and control group) at **P-6** on **postoperative nausea and vomiting** after **tonsillectomy** in children. There were no significant differences. **Shenkman et al, Anesthesiology 1999**

A Swedish prospective, randomised, placebo-controlled, double-blinded trial ($n = 60$) investigated the effect of acupressure at **P-6** on **nausea and vomiting** after **minor gynaecological surgery**. Only the group receiving acupressure at P-6 experienced no vomiting and also required no anti-emetic medication. **Alkaissi et al, Acta Anaesthesiol Scand 1999**

An Austrian prospective, randomised, placebo-controlled, double-blinded trial investigated the effect of laser acupuncture at **P-6** on **nausea and vomiting in children undergoing strabismus surgery**. The laser acupuncture was administered 15 minutes before induction of anaesthesia and 15 min after arriving in the recovery room. In the laser stimulation group, the incidence of vomiting was significantly lower (25%) than that in the placebo group (85%). **Schlager et al, Br J Anaesth 1998**

A Chinese prospective, randomised, placebo-controlled, double-blinded trial ($n = 163$) investigated the effect of acupoint stickers at **P-6**, **Ren-8** and **T.B.-18** on **motion sickness**. All tested acupuncture points showed a significantly better result than the active control points (scopolamine plaster) and the placebo control (lactose plaster). T.B.-18 was the most effective point (symptoms 100% reduced), followed by Ren-8 (81%) and P-6 (46%). **Pei et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1998**

A Chinese prospective, randomised, controlled trial ($n = 50$) compared two forms of anaesthesia for the **anterior approach cervical discectomy**: local anaesthesia plus intravenous anaesthesia (IVA) in the control group and electro-acupuncture at **P-6** and **L.I.-4** plus IVA. The effect of the anaesthesia was the same in both groups. The authors therefore recommend electro-acupuncture at P-6 and L.I.-4 as a less risky alternative. **Li et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1997**

A British prospective, randomised, placebo-controlled, double-blinded trial ($n = 81$) investigated the effect of intra-operative acupuncture at **P-6** in the prevention of **postoperative nausea and vomiting** in patients undergoing **gynaecological laparoscopic surgery**. The use of acupuncture significantly reduced the incidence of postoperative nausea and vomiting by 30–38% compared with placebo. **Al-Sadi et al, Anaesthesia 1997**

A US prospective, randomised, placebo- and sham-controlled, double-blinded trial ($n = 75$) investigated the effect of acupressure at **P-6** on **nausea and vomiting during caesarean section under spinal anaesthesia** compared with administration of 10 mg metoclopramide. Acupressure at P-6 is as effective as 10 mg metoclopramide. **Stein et al, Anesth Analg 1997**

An Australian prospective, randomised, placebo-controlled, double-blinded trial ($n = 84$) investigated the effect of TENS at **P-6** and **L.I.-4** on **postoperative nausea and vomiting** in **paediatric patients** undergoing **minor urological surgery**. The differences between the treatment and control group were not statistically significant. **Schwager et al, Anaesth Intensive Care 1996**

A Taiwanese prospective, randomised, sham-controlled, single-blinded experimental trial ($n = 48$) evaluated the relationship between electro-acupuncture at **P-6** and **ST-36** and cardiopulmonary function in healthy subjects. In the P-6/ST-36 group **resting heart rate**, **carbon dioxide production** and **oxygen requirement** were significantly decreased, indicating that acupuncture can lower the metabolic rate. **Lin et al, Chin Med J (Engl) 1996**

A Taiwanese prospective, randomised, placebo-controlled, double-blinded trial ($n = 60$) investigated the effect of acupressure wristbands at **P-6** on the incidence of **nausea and vomiting after epidural morphine for post-caesarean section pain relief**. Compared to the control group, the incidence of nausea significantly decreased from 43% to 3%, vomiting significantly decreased from 27% to 0%. **Ho et al, Acta Anaesthesiol Scand 1996**

A US prospective, randomised, placebo-controlled, single-blinded cross-over study ($n = 9$) investigated the effect of acupressure wristbands at **P-6** on nausea and vomiting for **seasickness**. Depending on the time of application (earlier or later during the trip) subjects were either symptom-free or complained about nausea and vomiting. The difference was highly significant. **Bertolucci et al, Aviat Space Environ Med 1995**

A US prospective, randomised, placebo- and sham-controlled, single-blinded experimental trial ($n = 9$) investigated the effect of acupressure at **P-6** on the incidence of visually-induced **motion sickness**. Acupressure at P-6 significantly reduced the incidence of nausea and abnormal gastric myoelectric activity. **Hu et al, Aviat Space Environ Med 1995**

A Chinese prospective, randomised, controlled trial ($n = 40$) investigated the effect of acupuncture on **P-6**, **L.I.-4**, **ST-36** and **KID-4** on the regulation of cellular immune function in patients with **malignant tumours**. There was a highly significant increase of the CD3+ and CD4+ levels, an increase in the CD4+/CD8+ ratio, a higher endorphin level and a decreased level of soluble interleukin-2 receptor (SIL-2R). **Wu, Zhen Ci Yan Jiu 1995**

A British prospective, randomised, controlled trial ($n = 46$) investigated the effect of acupressure at **P-6** on **nausea and vomiting** following **laparoscopy** for **gynaecological surgery** and the requirement for anti-emetic therapy. There was a significant reduction in the requests for anti-emetic therapy in the P-6-acupuncture group. **Allen et al, Anaesth Intensive Care 1994**

A Chinese prospective paired trial ($n = 40$) investigated the effect of electro-acupuncture (EA) at **P-6** during different times of the day (*chen*-time: 7–9am, *xu*-time: 7–9pm) on left ventricular function (LVF) in patients with **coronary heart disease**. EA performed at *chen*-time improved LVF, while EA administered under the same conditions during *xu*-time led to an impairment of LVF. **Li et al, J Tradit Chin Med 1994**

A British prospective case study ($n = 27$) reports that wearing a special wrist band decreased **pregnancy-related nausea and vomiting** more than 50%. **Stainton et al, Health Care Women Int 1994**

A US prospective, randomised, sham-controlled, single-blinded trial ($n = 60$) investigated the effect of acupressure at **P-6** on **pregnancy-related nausea and vomiting**. There was a significant decrease of nausea in the P-6 acupressure group. The incidence of vomiting did not differ from the sham control group. **Belluomini et al, Obstet Gynecol 1994**

A German prospective, randomised, placebo-controlled, single-blinded trial ($n = 60$) investigated the effect of acupressure at **P-6** on **nausea** in patients undergoing **gynaecological operations of longer duration** (6–8 h). Nausea was reduced from 53% in the placebo group to 23% in the acupressure group. **Gieron et al, Anaesthesist 1993**

A Chinese prospective, randomised, placebo-controlled, single-blinded trial ($n = 15$) investigated the effect of acupuncture at **P-6**, **HE-7** and **HE-3** on the frequency and duration of **angina pectoris** during exercise. The control groups received either sham acupuncture or no treatment at all. In the verum group, the anginal attack occurred significantly later and the duration of the attack after stopping the exercise was significantly shorter than

in the control groups. **Zhou et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1993**

A Taiwanese prospective, randomised, controlled trial ($n = 120$) investigated the effect of **P-6** acupoint injection with 0.2 ml 50% glucose in water compared to intravenous injection of 20 micrograms/kg droperidol for the prevention of **vomiting after gynaecological laparoscopy**. A non-controlled comparison-group received no treatment at all. The incidence of vomiting significantly decreased in both the P-6 group and the droperidol group. **Yang et al, Acta Anaesthesiol Scand 1993**

An Italian prospective, randomised, placebo-controlled, double-blinded cross-over study ($n = 60$) investigated the effect of unilateral and bilateral acupressure at **P-6** for the treatment of **morning sickness** compared to acupressure at a sham point. Compared to the placebo group both unilateral and bilateral acupressure at P-6 significantly reduced the frequency of morning sickness from 60% to 30%. **De Aloysio et al, Obstet Gynecol 1992**

A US 'n of 1' trial investigated the effect of **P-6** wristbands on **nausea and vomiting in hospice patients** ($n = 6$). There were no significant differences between the verum group, the placebo group and the group without wristbands. **Brown et al, Am J Hosp Palliat Care 1992**

A Canadian prospective, randomised, controlled, double-blinded trial ($n = 90$) investigated the effect of acupuncture at **P-6**, administered after induction of anaesthesia, on the frequency of **nausea and vomiting after strabismus surgery in children**. The control group received droperidol. There was no significant difference between the two groups. **Yentis et al, Can J Anaesth 1992**

A Canadian prospective, randomised, controlled, single-blinded trial ($n = 45$) investigated the effect of acupuncture at **P-6** administered after induction of anaesthesia on the frequency of **nausea and vomiting after tonsillectomy in children**. There was no significant difference between the verum group and the untreated control group. **Yentis et al, Br J Anaesth 1991**

A US prospective, randomised, controlled, double-blinded trial ($n = 66$) investigated the effect of acupuncture at **P-6** administered after induction of anaesthesia on the frequency of **nausea and vomiting after strabismus surgery in children**. There was no significant difference between the treatment group and placebo group. **Lewis et al, Br J Anaesth 1991**

An Irish case study ($n = 100$) reports that TENS at **P-6** reduced **chemotherapy-induced nausea** by 75%. **Dundee et al, J R Soc Med 1991**

A Swedish prospective, randomised, controlled cross-over study ($n = 21$) investigated the effect of acupuncture at **P-6**, **HE-5**, **BL-15**, **BL-20** and **ST-36** on the frequency of **angina pectoris** (AP) and the performance before onset of pain during exercise in patients suffering from coronary heart disease. The frequency of AP attacks decreased significantly from 11 to 6 attacks per week. Accordingly, the performance before onset of pain during exercise increased significantly from 82 W to 94 W. **Richter et al, Eur Heart J 1991**

A Chinese prospective, randomised, non-blinded, controlled trial ($n = 64$) investigated the effect of a 3 ml normal saline injection into **P-6** on the incidence of **postoperative vomiting**. Compared to the control group, vomiting was significantly reduced (2 vs 10 patients) in the treatment group. **Shyr et al, Ma Zui Xue Za Zhi 1990**

A British prospective, randomised, controlled experimental trial ($n = 18$) compared the effectiveness of acupressure bands at **P-6** on **motion sickness** to a placebo and 0.6 mg scopolamine. Only the subjects taking scopolamine showed a significant increase in tolerance to a laboratory nauseogenic cross-coupled motion challenge. **Bruce et al, Aviat Space Environ Med 1990**

A Taiwanese prospective, randomised, controlled trial ($n = 100$) compared the effect of electro-acupuncture (EA) and TENS at **P-6** on **nausea and vomiting after laparoscopy**. Compared to the control group (44%), nausea was significantly reduced in the EA and perchlorperazine group (12% each), but not in the TENS group (36%). **Ho et al, Anaesthesia 1990**

A British prospective, randomised, controlled trial ($n = 162$) compared the effect of acupressure wristbands at **P-6** on **postoperative nausea and vomiting** in the recovery room with conventional anti-emetic treatment. Acupressure at P-6 significantly reduced the intensity of the nausea. The reduced anti-emetic requirement and the reduced incidence of vomiting were not statistically significant. **Barsoum et al, J R Soc Med 1990**

A Chinese prospective, randomised, controlled trial ($n = 40$) compared the effect of electro-acupuncture (EA) at **L.I.-4** and **P-6** on intra-operative **analgesia** in patients undergoing **thyroidectomy** with superficial cervical plexus block. There was no significant difference between the two groups. **Ouyang et al, Zhen Ci Yan Jiu 1990**

An Irish prospective, randomised, sham-controlled, single-blinded trial investigated the effect of pre-operative electro-acupuncture and acupuncture at **P-6** on the frequency of **nausea and vomiting after minor gynaecological surgery** compared to sham acupuncture. The frequency of nausea and vomiting was significantly reduced in the P-6-group only. **Dundee et al, Br J Anaesth 1989**

A US prospective, randomised, controlled cross-over study ($n = 16$) investigated the efficacy of acupressure wristbands at **P-6** on **morning sickness**. Use of the acupressure wristbands significantly relieved morning sickness and also significantly reduced anxiety, depression and behavioural dysfunction. **Hyde, J Nurse Midwifery 1989**

An Irish prospective, randomised, sham-controlled cross-over study ($n = 10$) investigated the effect of electro-acupuncture on **P-6** on **chemotherapy-induced nausea**. In the verum group sickness was either completely absent or reduced considerably in 97% of patients and no side effects were encountered. The same research team confirmed these results in a larger-scale case study ($n = 105$). **Dundee et al, J R Soc Med 1988**

An Irish prospective, randomised, sham-controlled trial ($n = 350$) compared the effect of acupressure at P-6 on **morning sickness** to sham acupressure and a control group. Only acupressure at P-6 significantly relieved the symptoms of morning sickness. **Dundee et al, J R Soc Med 1988**

A prospective, randomised, placebo-controlled trial ($n = 75$) investigated the effect of acupuncture at **P-6** on **nausea and vomiting after minor gynaecological surgery**. Acupuncture at P-6 significantly reduced both nausea and vomiting. **Dundee et al, Br Med J (Clin Res Ed) 1986**

P-8 → P-6; LU-10

P-9 A larger-scale Chinese case study ($n = 100$) describes the positive effect of **P-9** on **infantile morbid night crying** in otherwise healthy children. **Zhao, J Tradit Chin Med 2002**

Triple Burner channel

T.B.-5 → P-6; L.I.-4; L.I.-12; L.I.-10

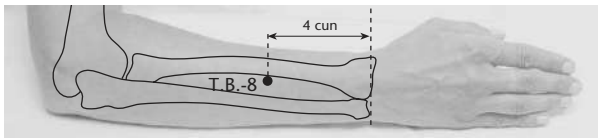

T.B.-8 Acupuncture at **T.B.-8** and **Du-15** activates areas important for **word generation** in the right inferior frontal gyrus, but not in the left inferior frontal gyrus. **Li et al, Hum Brain Mapp 2003**

A Russian case study reports the positive effects of acupuncture at **T.B.-8**, **T.B.-9**, **ST-20**, **ST-36**, **G.B.-14** and **Du-20** in 66 patients with **duodenal ulcer** accompanied by apparent shifts in psychovegetative correlations. **Kravtsova et al, Vopr Kurortol Fizioter Lech Fiz Kult 1994**

T.B.-9 → T.B.-8

T.B.-18 → P-6

Gall Bladder channel

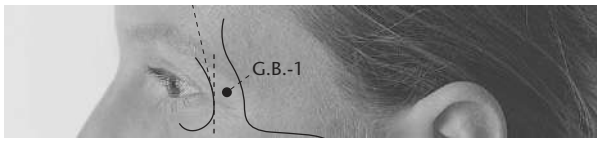

G.B.-1 An earlier Italian case study reports the successful application of **G.B.-1** for **heroin withdrawal** in nine male heroin addicts. After six months, 66% of these patients were classified as abstinent. Besides reservations regarding methodology of such a small-scale trial, the accompanying medication also poses problems in evaluating the effect of the acupuncture treatments. **Cocchi et al, Minerva Med 1979**

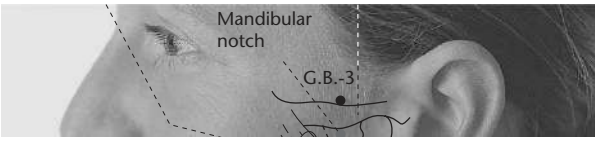

G.B.-3 A US prospective, randomised, double-blinded, experimental cross-over study ($n = 20$) investigated the effect of TENS at **G.B.-3** on the **anaesthetic requirement** (desflurane) to prevent purposeful movement of the extremities in response to noxious electrical stimulation. In comparison with the placebo group, bilateral application of TENS at G.B.-3 significantly reduced the desflurane requirement by 11%. **Greif et al, Anesthesiology 2002**

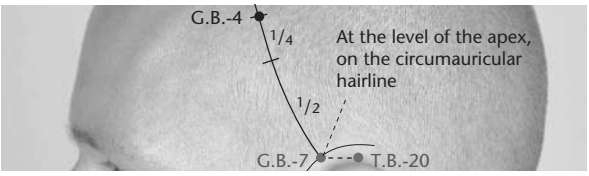

G.B.-4 Following the induction of an **ischaemic cerebral infarction** by occlusion of the middle cerebral artery in rats, animals treated with electro-acupuncture (EA) at **G.B.-4**, **G.B.-5**, **G.B.-6** and **G.B.-7** showed a significantly faster recovery than animals in the untreated control group or in the group treated with EA at **Du-20** and **Du-26**. Infarct volume was significantly reduced in the EA-treated groups. The authors discuss the acupuncture-induced production of vascular endothelial growth factor (VEGF) in astrocytes of the peri-infarct area as a causative factor. **Wang et al, Neurol Res 2003**

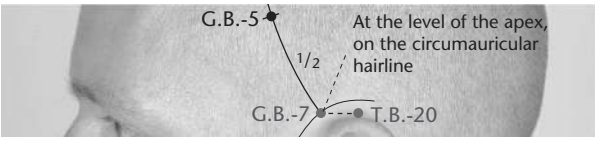

G.B.-5 → G.B.-4; ST-8

A Chinese case study recommends a particular needling technique (from **G.B.-5** or **Ex-HN-5** to **G.B.-8**) for the treatment of **migraine**. **Gan et al, J Tradit Chin Med 1986**

G.B.-6 → BL-7; G.B.-4

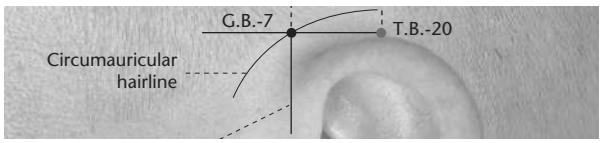

G.B.-7 → G.B.-4
A large-scale Chinese case study (n = 500) reports the successful application of acupuncture at **G.B.-7** in the treatment of **hemiplegia** after apoplexy. **Sun et al, J Tradit Chin Med 1985**
G.B.-8 → G.B.-5
G.B.-11 → ST-8

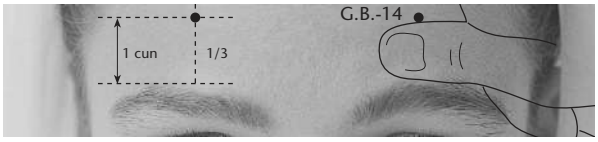

G.B.-14 → T.B.-8; ST-8; LIV-3
A case study reports that 63 patients with **facial paralysis** were treated with acupuncture at **G.B.-14**, **BL-1**, **ST-2**, **L.I.-20**, **ST-4**, **ST-6** and **ST-18** plus auxiliary points according to the underlying TCM pattern. Just over half of the patients were cured and all except one patient showed a marked improvement. **Liu et al, Zhen Ci Yan Jiu 1992**

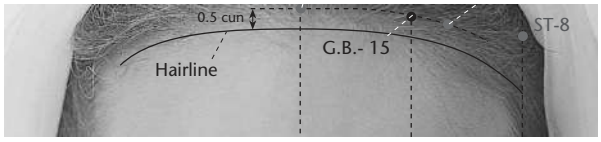

G.B.-15 An uncontrolled, non-randomised, group-comparison trial investigated the effect of acupuncture on **childhood diarrhoea**. Group I received scalp acupuncture (after Yamamoto) with points on line 3 starting at **G.B.-15**; Group II received body acupuncture; and Group III was treated with antibiotics. Group I showed the best results, followed by Group II and Group III, with results in Group I and II significantly better than in Group III. However, besides some methodological shortcomings, the authors did not seem to have a very firm grasp of their subject matter: while using the correct points for the correct indication, they refer to line 3 as line 2 and to G.B.-15 as BL-15! **Lin et al, J Tradit Chin Med 1993**
G.B.-20 → ST-8; LIV-3; S.I.-18; L.I.-15

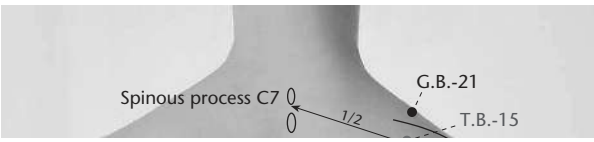

G.B.-21 → L.I.-15
A Chinese case study reports the successful treatment of **achalasia of the cardia** with acupuncture at **G.B.-21**. **Shi et al, J Tradit Chin Med 1994**
G.B.-24 → LIV-14

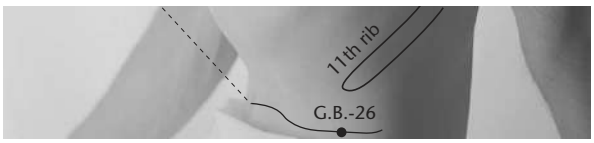

G.B.-26 An earlier prospective group-comparison trial investigated the **analgesic effect** of electro-acupuncture (EA) at **G.B.-26**, **ST-36**, **SP-6** and **HE-7** in patients who underwent hysterectomy by subumbilical midline incision. EA for 40 minutes had the same analgesic effect as 30 mg pentazocine. However, in contrast to pentazocine; EA also improved the vital capacity of the lung for 3–4 hours after the treatment. **Facco et al, Am J Chin Med 1981**
G.B.-30 → L.I.-12; L.I.-15

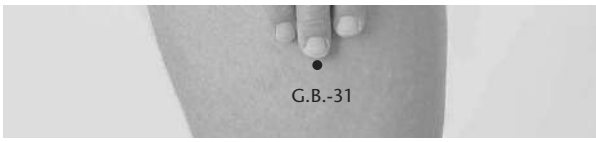

G.B.-31 → L.I.-4
Three publications by a Canadian research team led by Romita demonstrated that electro-acupuncture (EA) at **G.B.-31**, **ST-32** and **ST-36** resulted in a more pronounced inhibition of the **nociceptive tail withdrawal reflex** than classical acupuncture in slightly anaesthetised rats. **Romita et al, Brain Res 1997, Romita et al, Brain Res Bull 1997, Romita et al, Brain Res 1996**

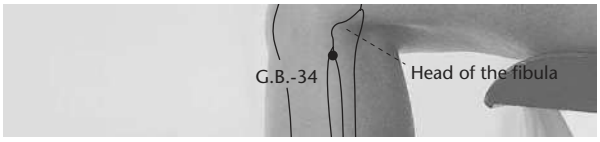

G.B.-34 → BL-10; Du-20; ST-36; L.I.-4; L.I.-15
A Korean prospective randomised placebo-controlled experimental trial (n = 10) investigated the effect of acupuncture at **G.B.-34** on motor cortex activities as evidenced by **fMRI**. Bilateral sensorimotor areas BA 3, 4, 6 and 7 were activated during acupuncture, providing a basis for future investigations regarding therapeutic interventions in **stroke** patients. **Jeun et al, Am J Chin Med 2005**
A Chinese prospective, randomised, controlled trial (n = 20) investigated the effectiveness of electro-acupuncture (EA) at **G.B.-34** and **ST-38** on **epicondylitis lateralis (tennis elbow)** compared to manual acupuncture at the same points. After six treatments over a period of two weeks, the EA group showed significantly better results in relation to pain relief (Pain visual

analogue scale) and pain free hand grip strength (PFG). **Tsui et al, Acupunct Electrother 2002**
A British, prospective, randomised, sham-controlled, double-blinded experimental trial (n = 240) investigated the effect of transcutaneous electric nerve stimulation (TENS) at **G.B.-34** on mechanical **pain thresholds**. Low frequency, high intensity stimulation showed a significant analgesic effect that was sustained post-stimulation. **Chesterton et al, Pain 2002**
G.B.-37 → P-2; P-5; L.I.-15
G.B.-38 → P-5
G.B.-39 → LIV-3
G.B.-40 → LIV-3

G.B.-43 An Austrian prospective, randomised, placebo-controlled, single-blinded trial investigated the effect of laser acupuncture at **G.B.-43** on cerebral cortical and subcortical activations as evidenced by **fMRI**. Only true laser acupuncture led to ipsilateral activations within the thalamus, nucleus subthalamicus, nucleus ruber, the brainstem, and the Brodmann areas 40 and 22. **Siedentopf et al, Laser Med Sci 2005**
A purely descriptive study investigated the effect of **G.B.-43** and **L.I.-4** on **somatosensory evoked potentials**. Particular attention was paid to the effect of the *deqi*-sensation. **Wu et al, Zhen Ci Yan Jiu 1993**

Liver channel

LIV-1 → SP-1

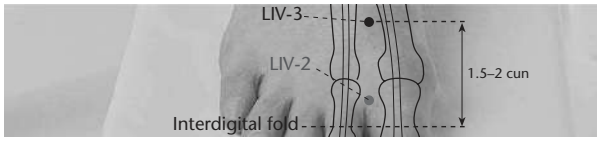

LIV-3 → Du-20; BL-60; L.I.-4; L.I.-15; SP-6
In a Chinese-German single-blinded experimental trial, acupuncture at **LIV-3** and **G.B.-40** resulted in the activation of the secondary somatosensory cortical areas, frontal areas, the right side of the thalamus and the left side of the cerebellum. Sham-points had no effect. **Fang et al, Neuroradiology 2004**
An Italian open, prospective, randomised group-comparison trial investigated the effect of acupuncture, laser acupuncture and TENS on **LIV-3**, **SP-6**, **L.I.-4**, **G.B.-20**, **Du-20** and **Ex-HN-5** on **transformed migraines**. The number of days with headache per month significantly decreased during treatment in all groups. Acupuncture showed the best effectiveness over time. **Allais et al, Neurol Sci 2003**
An Italian prospective, randomised, controlled, non-blinded trial (n = 160) investigated the effectiveness of acupuncture at **LIV-3**, **SP-6**, **ST-36**, **Ren-12**, **L.I.-4**, **P-6**, **G.B.-20**, **G.B.-14**, **Ex-HN-5** and **Du-20** on **migraine attacks** without aura compared to flunarizine. The number of attacks after two and four months of therapy was significantly lower in the acupuncture group than in the flunarizine group, and analgesic consumption was accordingly also significantly lower in the acupuncture group. Pain intensity was significantly reduced only by acupuncture treatment. **Allais et al, Headache 2002**
A British prospective, randomised, placebo-controlled cross-over study (n = 56) investigated the effect of **LIV-3** on **rheumatoid arthritis (RA)**. There were no significant changes between the treatment and placebo group. The treatment of only one point for a condition such as RA is unusual. **David et al, Rheumatology (Oxford) 1999**
A US prospective, randomised, sham-controlled, single-blinded trial (n = 10) investigated the effect of electro-acupuncture (EA) on **LIV-3**, **ST-36** and **L.I.-11** on lowering diastolic blood pressure in hypertensive subjects. Compared to the sham group, diastolic blood pressure showed a significant decrease in the EA-group. **Williams et al, Phys Ther 1991**
LIV-13 → Du-20

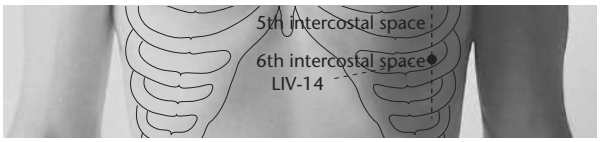

LIV-14 A Chinese prospective, non-randomised, controlled, non-blinded trial (n = 56 plus control group) investigated the effect of water injections at **LIV-14**, **G.B.-24** and **Ren-14** on biliary colic pain. The pain disappeared in 57% of cases and showed improvement in a further 39%. This result is significantly better compared to the control group receiving conventional treatment. **Jiang et al, J Tradit Chin Med 1995**

Du mai (governing vessel)

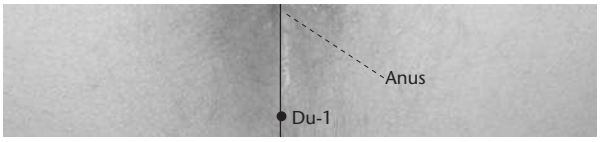

Du-1 A prospective, controlled, group-comparison trial in **hypogalactic postpartum sows** (n = 42) showed that apitherapy (acupuncture with bee venom) at **Du-1** and **ST-18** was superior to treatment with conventional drugs. **Choi et al, J Vet Sci 2001**
A prospective, controlled, group-comparison animal experiment showed that apitherapy at **Du-1** and **ST-25** in piglets with preweaning **diarrhoea** (n = 91) is as effective as conventional treatment with drugs. **Choi et al, Am J Chin Med 2003**

In an earlier prospective, controlled, group-comparison trial in piglets ($n = 44$) with induced **enteropathogenic *Escherichia coli*** diarrhoea, traditional acupuncture at **Du-1**, **ST-36**, **BL-20**, **Ren-12** and **ST-25** produced the best results in comparison to the groups treated with electro-acupuncture or antibiotics. There were too many groups given the number of cases. **Hwang et al, Am J Vet Res 1988**

Several animal experiments investigated the effect of acupuncture on signalling and modulating pain. Acupuncture at **Du-1** and **Du-2** inhibited the reaction on **experimental pain stimuli**. **Bing et al, Neuroscience 1991, Xu et al, Zhen Ci Yan Jiu 1989, Xu et al, Zhen Ci Yan Jiu 1989**

Du-2 → Du-1

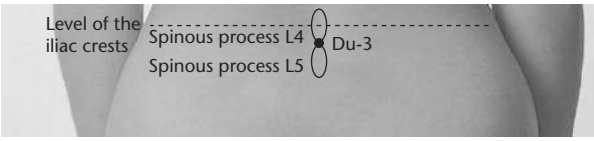

Du-3 Rats treated with electro-acupuncture at **Du-3**, **BL-54** and **BL-60** 15 minutes after a standardised **spinal cord injury** at T8, showed a marked improvement (regarding both morphology and functioning) three days post-operatively compared to rats not treated with acupuncture. However, none of the beneficial effects occurred in rats given acupuncture treatment 24 hours after spinal cord injury. **Politis et al, Acupunct Electrother Res 1990**

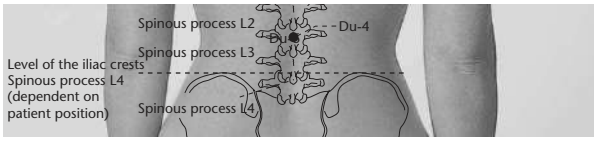

Du-4 A Turkish prospective, randomised, non-blinded, controlled group-comparison trial ($n = 24$) investigated the efficacy of acupressure at **Du-4**, **Du-15**, **Du-20**, **BL-23**, **BL-28**, **BL-32**, **HE-7**, **HE-9**, **ST-36**, **SP-4**, **SP-6**, **SP-12**, **Ren-2**, **Ren-3**, **Ren-6**, **KID-3** and **KID-5** on **enuresis** compared to oxybutynin given to the control group. Acupressure was administered to 12 patients by the parents. While treatments in both groups were successful, complete recovery was significantly higher in the acupressure group (83.3%) compared to the control group (58.3%). **Yuksek et al, J Int Med Res 2003**

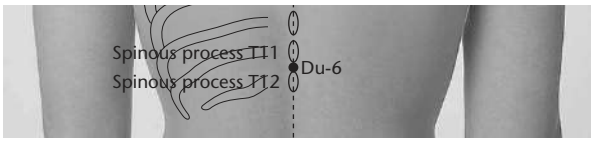

Du-6 An earlier Chinese case study reports the application of **Du-6** for **acupuncture anaesthesia for hysterectomies. Authors' Collective, Chin Med J (Engl) 1978**

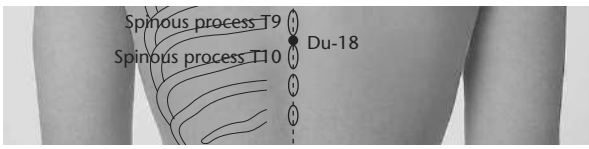

Du-8 An experiment in rats supports the effect of **Du-8** on **seizures** (by releasing melatonin) and on **cerebral infarction** (reduction of the cerebral infarction volume due to release of taurine and decrease of aspartame). **Chao et al, Acupunct Electrother Res 2001, Zhao et al, Acupunct Electrother Res 1997, Yin et al, Zhen Ci Yan Jiu 1994**

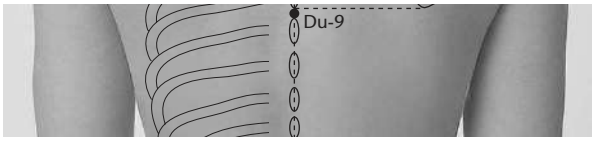

Du-9 The effect of electro-acupuncture at **Du-9** was investigated in 23 patients suffering from **coronary heart disease**. The results showed a slight dilation of the coronary arteries and a slight decrease of the heart rate frequency (monitored by coronary arteriography). **Yan et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1998**

Two earlier Chinese case studies report decreased symptoms in patients with coronary heart disease after acupuncture or tuina at Du-9. Wang et al, Zhong Xi Yi Jie He Za Zhi 1988, Wang et al, Zhong Xi Yi Jie He Za Zhi 1987

A Chinese case study reports the successful treatment of **angina pectoris** by implanting a micro-depressor at **Du-9**. **Wang et al, Zhong Xi Yi Jie He Za Zhi 1988**

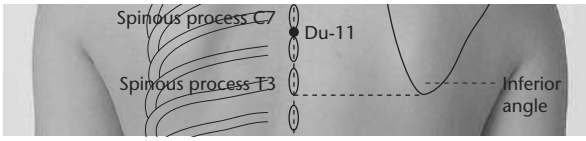

Du-11 Electro-acupuncture at **Du-11** and **Du-16** prior to, during and after an experimentally induced **transient ischaemia** in gerbils suggests the protective influence of these points. They seem to suppress both glutamate release and reperfusion injury following the ischaemic insult. **Pang et al, Am J Chin Med 2003**

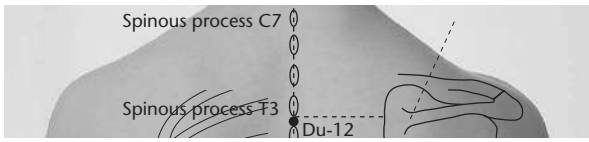

Du-12 A Chinese case study reports that classic acupuncture at **Du-12**, **L.I.-10**, **BL-11** and **S.I.-3** with *deqi* sensation in the direction of the disorder in 55 patients with **cervical vertebra**

disease and **impaired cerebral blood flow capacity** led to a significant improvement of blood flow capacity. **Qie et al, Zhong Xi Yi Jie He Za Zhi 1991**

Du-14 → L.I.-4, ST-8; ST-36

A Chinese experimental trial (n = 40) investigated the effect of acupuncture at **Du-14** and **P-6** on driving performance during a three-hour driving simulator test. The acupuncture group showed a significantly better driving performance and a decrease in **driving fatigue**, compared to the untreated control group. **Li et al, Accid Anal Prev 2004**

A Chinese prospective, randomised, controlled, non-blinded trial (n = 40) investigated the effect of acupuncture at **Du-14 and P-6** on heart rate and driving performance during a three-hour driving simulator test. The acupuncture group showed a significantly decreased **activation of the sympathetic nervous system** and increased **parasympathetic activity** compared to the untreated control group. **Li et al, Eur J Appl Physiol 2003**

A Chinese retrospective case study (n = 25) investigated the effect of cupping at **Du-14**, **BL-13**, **BL-20** and **BL-23** on **bronchial asthma** while reducing orally administered cortisone. The author reports that 56% of patients showed a marked improvement after 30 treatments. **Hu, J Tradit Chin Med 1998**

A Chinese prospective case study (n = 50) reports the successful treatment of **perennial allergic rhinitis** by applying 10% *Cantharides* extract at **Du-14** and **P-6**. The results showed an effective rate of 88%. The allergic nasal mucosa provocative test of the treated group improved after the treatment, the number of eosinophils and basophils in nasal secretion decreased and the serum total IgE also reduced significantly. **Tang et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 1995**

Du-15 → T.B.-8; Du-4

Du-16 → Du-11

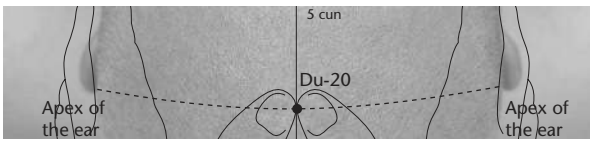

Du-20 → T.B.-8; G.B.-4, ST-8; Du-4; LIV-3; HE-7, L.I.-4

A Chinese prospective randomised controlled single-blinded group-comparison trial (n = 50) investigated the effect of a standardised acupuncture protocol, used either on its own or accompanied by needling of **Du-20**, **Du-26** and **HE-7** (either singly or in combination) on **vascular dementia** in hemiplegic patients. Clinical symptoms were observed and scales such as HDS-R, ADL and functional activities questionnaire (FAQ) used for assessment. The group additionally treated with **Du-20 and HE-7** showed marked improvement of memory, orientation and reaction. At the same time mental rigidity and trance decreased. **Du-20** on its own improved understanding, calculating and social adaptation, while **Du-26** was helpful in treating mental retardation, failure to perform daily activities, trance and poor memory. The combination of the three points showed overall the best results in improving intelligence and social adaptation. The particular design of this trial – combining a standard protocol used in both groups with the actually relevant points in the verum group – presents an interesting model for eliminating persistent problems regarding placebo methods and blinding. However, in this particular trial the number of participants (5 groups of 10 subjects) was too low to determine if the results were statistically significant. **Lai et Huang, Chin J Integr Med 2005**

A German prospective, non-blinded, pair-matched, controlled trial (n = 121 pregnant women between 30 and 39 weeks of gestation) investigated the effect of acupuncture at **Du-20** and **ST-36** on **fetal and maternal cardiotocographic parameters and maternal circulation**. The main difference was a transient increase of the Fisher score in the treatment group compared to the control group, with a persistent increase of the Fisher score. **Scharf et al, Z Geburtshilfe Neonatol 2003**

A Croatian prospective, randomised, placebo-controlled trial (n = 57) investigated the effectiveness of **Du-20**, **L.I.-4**, **Ren-3**, **Ren-4**, **Ren-6**, **G.B.-34**, **BL-23**, **SP-6** and **HE-7** on **primary dysmenorrhoea** compared to placebo-acupuncture. Subjects were observed for a period of two years. After one year, 93.3% of acupuncture patients were symptom-free and/or not taking medication compared to only 3.7% of the placebo patients. **Habek et al, Gynakol Geburtshilfliche Rundsch 2003**

A Polish prospective, randomised, controlled, non-blinded trial (n = 69) investigated the effect of acupuncture at **Du-20**, **Ren-12**, **ST-36**, **ST-21**, **ST-25**, **LIV-3**, **LIV-13**, **P-6** and **HE-7** in addition to a low calorie diet in the treatment of obesity (BMI approximately 33 kg/m^2). This diet consisted of a daily intake of about 1200–1500 kcal, which was about 1000 kcal below the daily energy requirement of the patients. The acupuncture group showed a significantly better result after only 12 treatments over a period of 6 weeks. The mean body weight decreased from 85.5 kg to 72.7 kg, while it only decreased to 77.1 kg in the diet group. **Wozniak et al, Ginekol Pol 2003**

A German prospective, randomised, placebo-controlled, double-blinded trial (n = 56) investigated the effect of acupuncture at **Du-20**, **Ex-HN-6**, **HE-7**, **P-6** and **BL-62** in the treatment of **depression** and **generalised anxiety**. Results were assessed using questionnaires. After 10 treatments, the verum acupuncture group showed a significantly larger clinical improvement compared to the placebo group (61% vs 21%). **Eich et al, Fortschr Neurol Psychiatr 2000**

Du-26 → G.B.-4; Du-20

Ren mai (conception vessel)

Ren-2 → Du-4

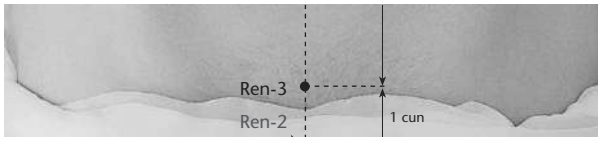

Ren-3 → Du-4; Du-20

A Taiwanese prospective, randomised, controlled trial ($n = 60$) investigated the influence of electro-acupuncture (EA) at **Ren-3**, **Ren-4** and **BL-32** on the **neurogenic bladder of spinal cord injured patients** compared to the conventional bladder training programme received by the control group. The acupuncture group took significantly less time to achieve balanced voiding than the control group (on average 28 days less). The best results were achieved in patients who received acupuncture within three weeks after injury (39 days less). **Cheng et al, Spinal Cord 1998**

Ren-4 → Du-20; ST-25; Ren-3

Ren-6 → Du-4; Du-20; Ren-17

Ren-8 → P-6

A Chinese prospective, randomised, controlled, non-blinded trial ($n = 56$) investigated the effect of moxibustion at **Ren-8** in addition to chemotherapy for **nasopharyngeal carcinoma** (stage III and IV). There were significantly fewer toxic and side effects in the moxibustion group compared to the control group. The five-year survival rate in the two groups were 50.0% and 35.7% respectively. **Chen et al, Zhongguo Zhong Xi Yi Jie He Za Zhi 2000**

A Chinese prospective, randomised, controlled trial ($n = 140$) investigated the effect of applying a topical ointment at **Ren-8** on **epigastric pain**. Results in the treatment group were significantly better than in the control group (94% vs 53%). **Ba et al, J Tradit Chin Med 1999**

Ren-12 → L.I.-4; Du-1; Du-20; LIV-3; Ren-17

Ren-13 → P-6

Ren-14 → LIV-14

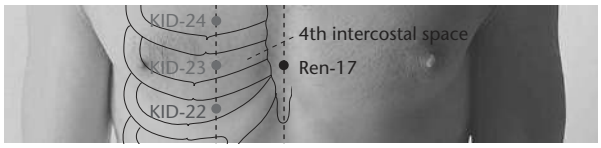

Ren-17 A Chinese prospective, randomised, controlled trial ($n = 60$) investigated the effect of **Ren-17**, **Ren-12**, **Ren-6**, **ST-36 and SP-10** on the symptoms of **vascular dementia** compared to conventional treatment. Both acupuncture and conventional treatment showed a highly significant increase of MMSE, HDS-R and ADL scores. However, the results of the acupuncture group were statistically more significant regarding MMSE, HDS-R and the total rate of effectiveness (80% vs 47%). **Yu et al, Neurol Res 2006**

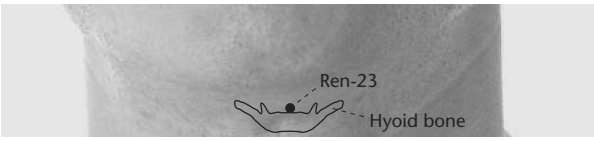

Ren-23 A Chinese retrospective case study ($n = 120$) reports the successful treatment of **pseudobulbar palsy** with acupuncture at **Ren-23** and **LU-5. Wang et al, J Tradit Chin Med 1998**

Ren-24 A German prospective, single-blinded, controlled trial ($n = 41$) reports that acupuncture at **Ren-24** significantly reduced the severity of the **gag reflex** during transoesophageal echocardiography compared to acupuncture at a sham point. **Rosler et al, J Altern Complement Med 2003**

Extra points

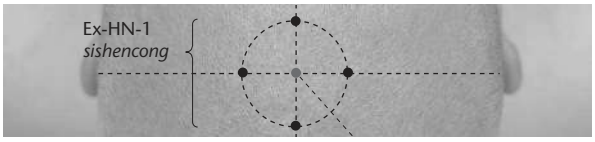

Ex-HN-1 → BL-7

A Taiwanese, prospective, randomised, placebo-controlled, single-blinded experimental trial ($n = 9$) investigated the effect of the *sishencong*-points (**Ex-HN-1**) on the variance of heart rate frequency. The authors observed an increase in the vagal activities but a suppression of the sympathetic activity. **Wang et al, Auton Neurosci 2002**

A Chinese case study ($n = 51$) reports the successful treatment of **hyperthyreosis** with acupuncture at Ex-HN-1. **Xie et al, J Tradit Chin Med 1994**

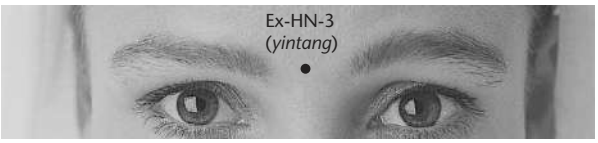

Ex-HN-3 An Austrian prospective, randomised, single-blinded experimental cross-over study ($n = 25$) showed that acupuncture

(but not laser acupuncture) at *yintang* monitored by EEG had a sedating effect. Clinical application of this point, for example for **sleeping disorders**, requires further investigation. **Litscher et al, Eur J Anaesthesiol 2004**

Ex-HN-4 → S.I.-18; L.I.-4
Ex-HN-5 → G.B.-5; LIV-3
Ex-HN-6 → HE-7; Du-20

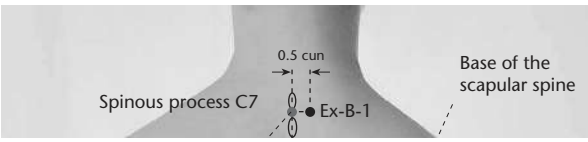

Ex-B-1 A Chinese prospective, randomised, controlled trial (n = 84) investigated the effect of a herbal plaster at **Ex-B-1**, **BL-11**, **BL-12**, **BL-13** and **BL-15** on **chronic bronchitis** compared to intramuscular injection therapy. In the plaster group the clinical total effective rate was 93.2% immediately after the treatment course and 91.5% after 18 months (control group: 80% and 80%). X-rays of the thorax showed the effectiveness of the plaster as 40.7% and 89.8% after 18 months (control group: 20% and 76%). **Huo et al, Zhongguo Zhong Xi Yi Jie He za Zhi 2001**

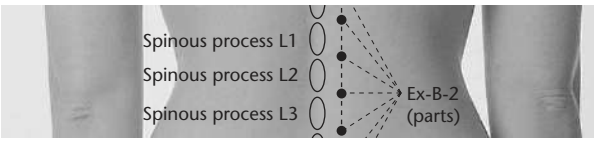

Ex-B-2 A Chinese case study (n = 43) reports the positive effects of treating **ulcerative colitis** with acupuncture at **Ex-B-2** as well as stimulation with a plum-blossom needle at **BL-22** and **BL-25. Yue and Zhenhui, J Tradit Chin Med 2005**

A Chinese retrospective case study (n = 168) reports the successful treatment of **sciatica** with acupuncture at the **Ex-B-2** points. **Pei et al, J Tradit Chin Med 1994**

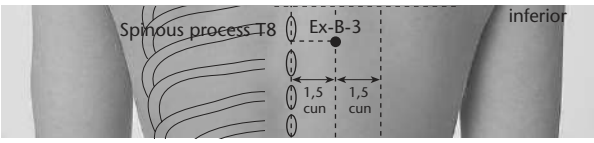

Ex-B-3 An experiment in diabetic rabbits demonstrated that electro-acupuncture at **Ex-B-3** decreased plasma glucose levels and inhibited the release of pancreatic glucagon. The effect was increased by adding **ST-36**. This latter point led to no significant changes when needled alone. **Zeng et al, J Tradit Chin Med 2002**

Ex-UE-9 → L.I.-4

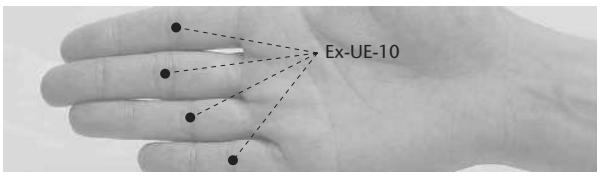

Ex-UE-10 A case study describes the successful treatment of **hiccups** with acupuncture at **Ex-UE-10**. **Qi, J Tradit Chin Med 1993**

Ex-UE (*JIANNEILING*) → L.I.-15
Ex-UE (*JIANQIAN*) → L.I.-4
Ex-LE-4 → ST-35
Ex-UE (*ZHONGPING*)

A Chinese retrospective case study (n = 115) reports the successful treatment of **shoulder and back pain** with acupuncture at ***zhongping***. The point *zhongping* is located midway between ST-36 and ST-37. **Wang et al, J Tradit Chin Med 1995**

9.3 Summary

The data presented clearly demonstrates that several actions of a number of acupuncture points are undoubtedly effective. These points (for example P-6) were tested according to the highest scientific standards, comparing standard acupuncture to deep acupuncture at non-acupuncture points (a methodology considered the most superior acupuncture placebo).

This proof is even more important since the results of the two largest acupuncture trials to date, ART (Acupuncture Randomized Trials) and GERAC (German Acupuncture Trial), may have led to the impression that, while acupuncture is indeed effective, its effect is independent of the location of the acupuncture points. The hitherto unsurpassed size of the trials and the undoubtedly intelligent and biomathematically impeccable design (especially of the GERAC trial) further contribute to this impression. Due to their flexible point prescriptions and their complexity, these studies have not been included in section 9.2.

But what are the basic results of these trials? (up to May 2006: **Witt et al, Lancet 2005; Linde et al, JAMA 2005; Melchart et al, Br Med J 2005; Brinkhaus, Arch Intern Med 2006**; **Diener et al, Lancet Neurol 2006**)

For both **gonarthrosis** and **lower back pain,** acupuncture had significantly better results than the untreated waitlist (ART) or conventional therapy (GERAC). For **migraine**, acupuncture is significantly better than the untreated waitlist (ART), and treatments with acupuncture for only six weeks had the same results as treatments with conventional medicine for six months (GERAC). For **tension headaches**, acupuncture provided significantly better results than the waitlist (ART). A comparison between acupuncture and conventional therapy in the treatment of tension headaches within the framework of GERAC was not possible, since all but two patients refused the stigmatised conventional treatment with amitriptyline due to the high incidence

Johannes W. Rohen
Chihiro Yokochi
Elke Lütjen-Drecoll

Color Atlas of Anatomy

FOURTH EDITION

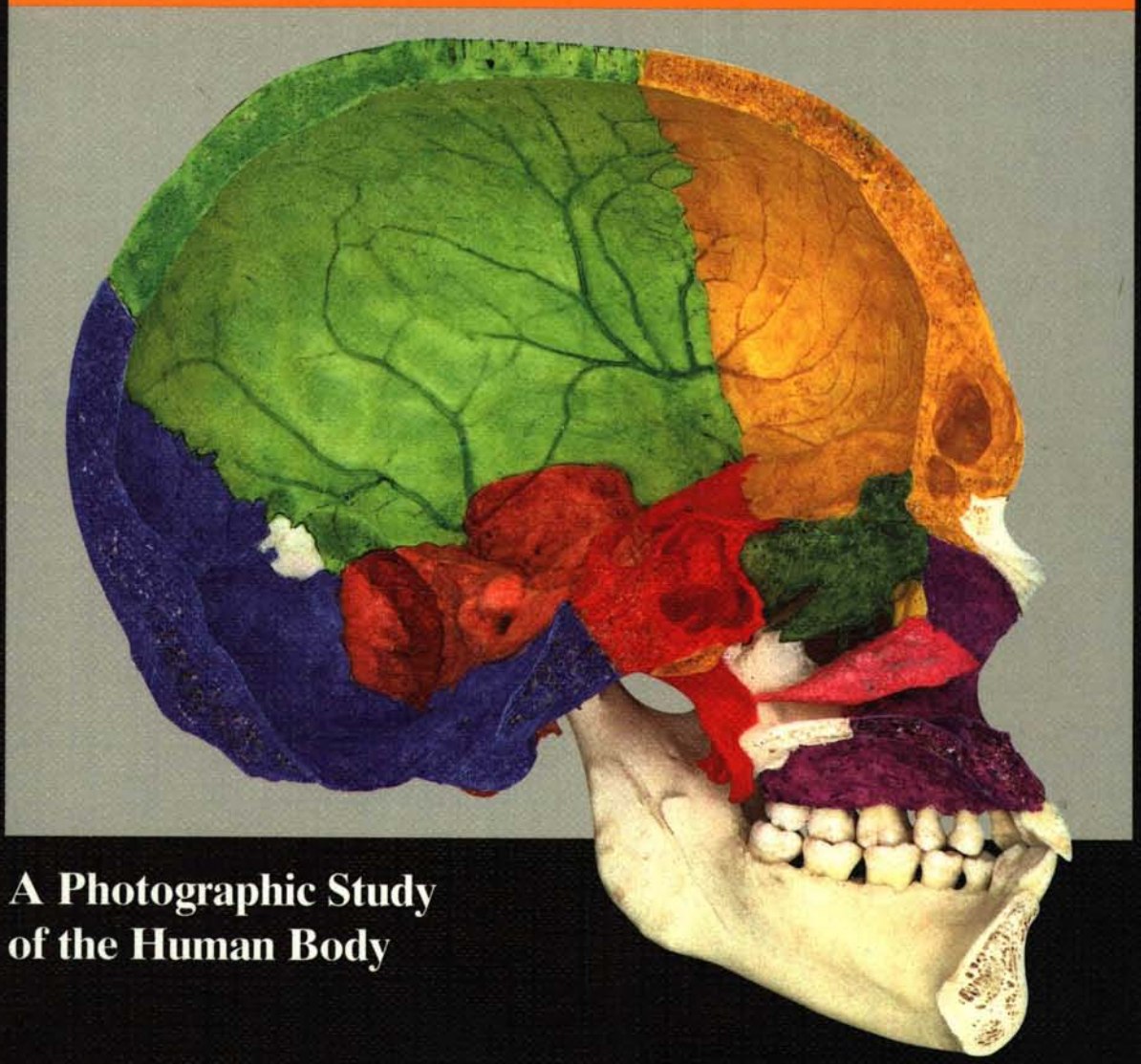

A Photographic Study of the Human Body

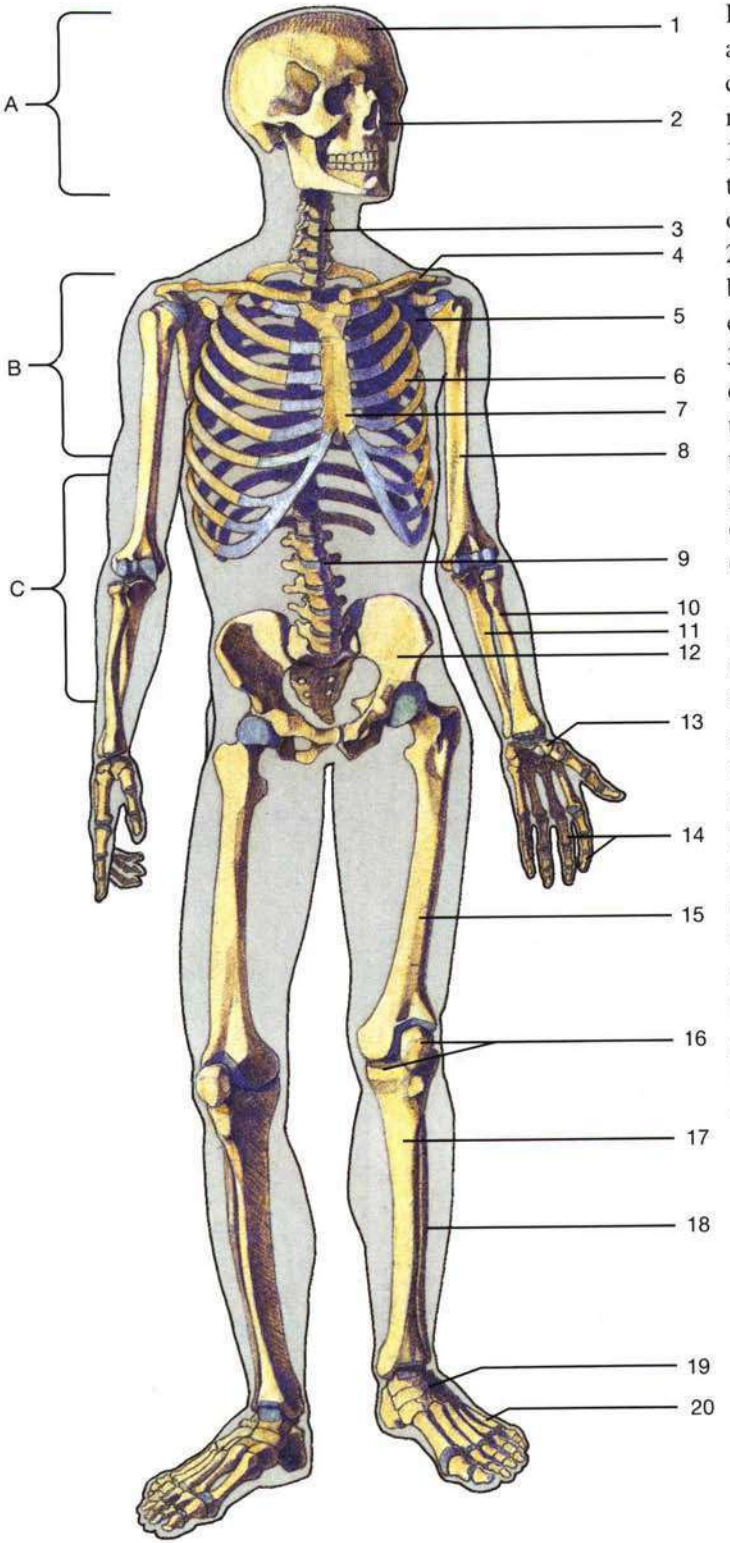

Structure of the human body and the skeleton. Blue = joints.
A Head (caput) B Thorax (thoracic cavity) C Abdominal and pelvic cavities

In contrast to most other mammals the human body is adapted for bipedal locomotion. Three general principles in the architecture of the human organism are recognizable:

1. The principle of **segmentation,** which dominates in the trunk. The vertebral column and the thorax consist of relatively equal, segmentally arranged elements.
2. The principle of **bilateral symmetry.** Both sides of the body are separated by a midsagittal plane and resemble each other like image and mirror-image.
3. The principle of **polarity** between the head at one end of the body and the lower extremities at the other. As the center of the information system the head contains the main sensory organs and the brain. The head has a predominantly spherical form while the extremities consist of radially formed skeletal elements, the number of which increases distally.

A. The **skull** consists of two parts: 1. a **cranial part** containing mainly the brain and the sensory organs and 2. a **facial part** which contains the nasal and oral cavity and the chewing apparatus. The cranial cavity is continuous with the vertebral canal which contains the spinal cord.

B. The **thorax** contains the respiratory and circulatory organs (lung, heart, etc.) but also some of the abdominal organs which are located underneath the diaphragm.

C. The **abdominal cavity** contains the organs of metabolism such as the liver, the stomach and the intestinal tract as well as the excretory and genital organs (kidney, uterus, urinary bladder, etc.). The latter are located primarily in the **pelvic cavity** with the exception of the testes.

1 Cranial part } of the skull
2 Facial part } of the skull
3 Vertebral column (cervical part)
4 Clavicle
5 Scapula
6 Ribs
7 Sternum
8 Arm (humerus)
9 Vertebral column (lumbar part)
10 Radius } forearm
11 Ulna } forearm
12 Pelvis
13 Wrist (carpals) } hand
14 Fingers (phalanges) } hand
15 Thigh (femur)
16 Patella and knee joint
17 Tibia } leg
18 Fibula } leg
19 Tarsals } foot
20 Metatarsals } foot

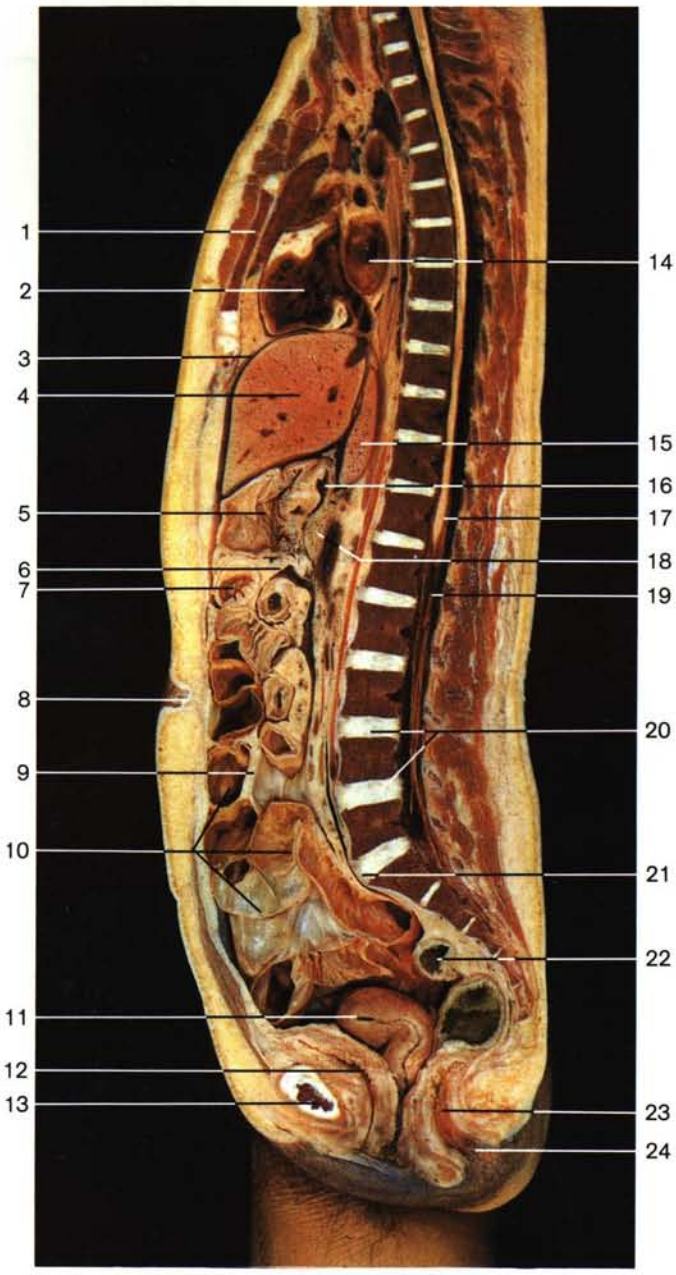

Median section through the trunk (female).

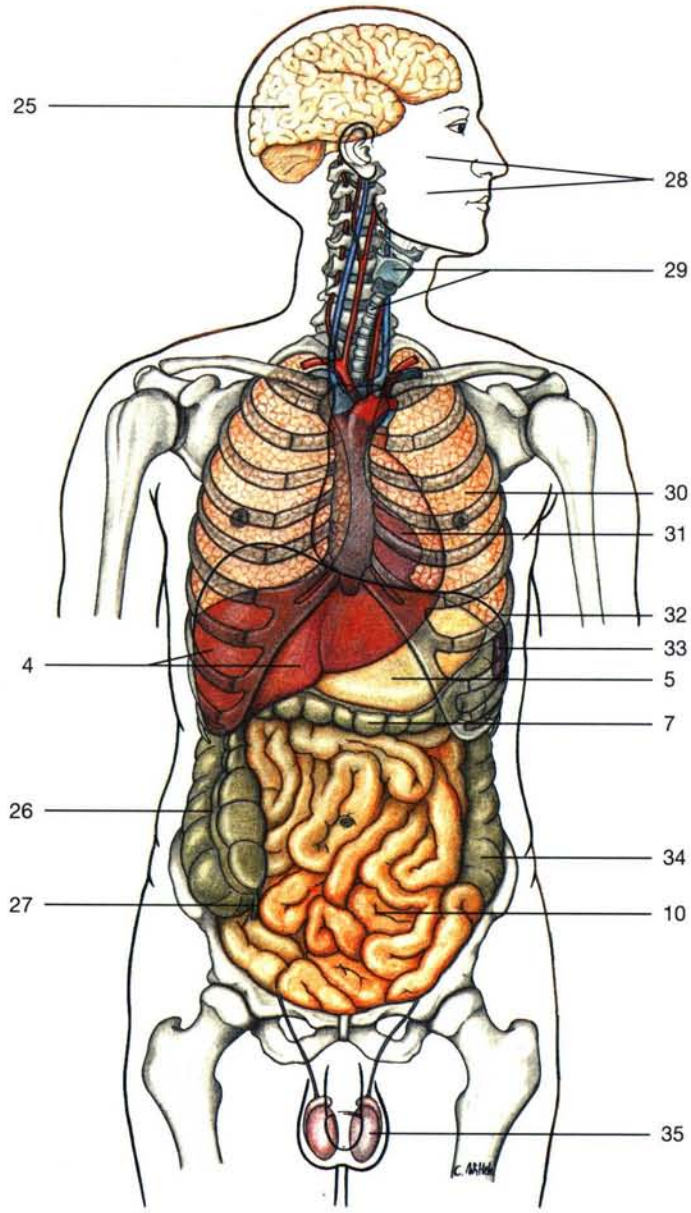

Positon of the inner organs of the human body (anterior aspect)
The main cavities of the body and their contents.

1 Sternum
2 Right ventricle of heart
3 Diaphragma
4 **Liver**
5 **Stomach**
6 Transverse mesocolon
7 Transverse colon
8 Umbilicus
9 Mesentery
10 **Small intestine**
11 Uterus
12 Urinary bladder
13 Pubic symphysis
14 Left atrium of heart
15 Caudate lobe of liver
16 Omental bursa or lesser sac
17 Conus medullaris
18 **Pancreas**
19 Cauda equina
20 Intervertebral discs (lumbar vertebral column)
21 Sacral promontory
22 Sigmoid colon
23 Anal canal
24 Anus
25 Head (neurocranium) with brain
26 Ascending colon
27 Appendix
28 Facial region (viscerocranium) with oral and nasal cavities
29 Trachea and larynx
30 Thorax with the lungs
31 **Heart**
32 Surface projection of the diaphragm
33 **Spleen**
34 Descending colon
35 Testis

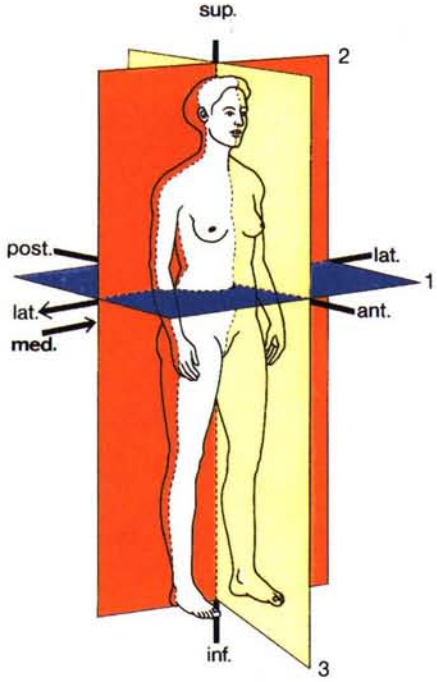

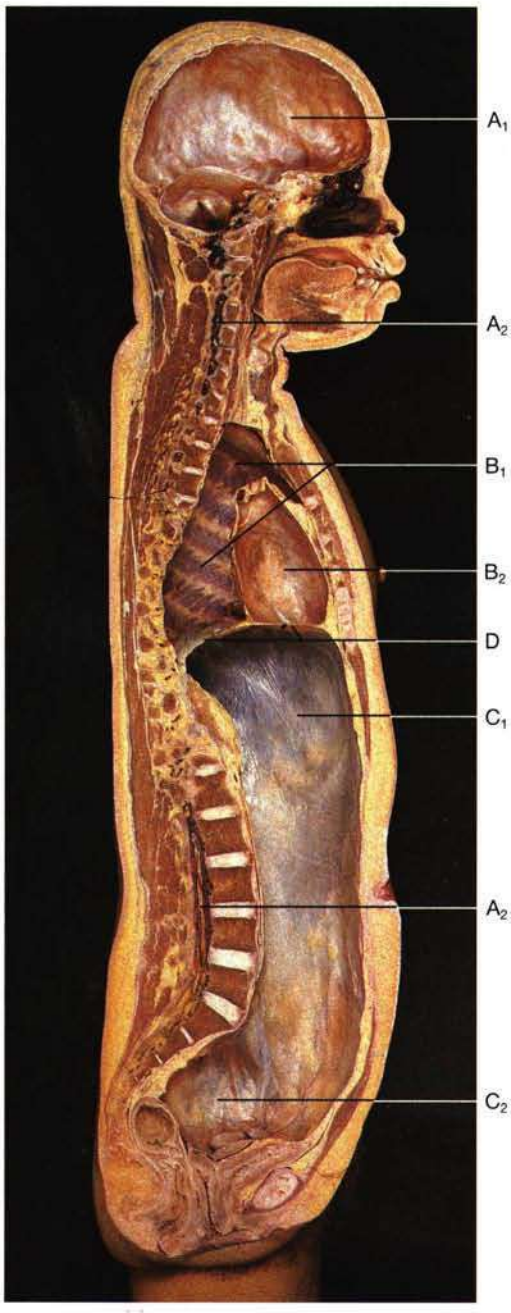

Sagittal section through the human body (female).
Demonstration of the main cavities of the body.
Internal organs are removed.

A_1 Cranial cavity
A_2 Vertebral canal
B_1 Thoracic cavity
B_2 Pericardial cavity
C_1 Abdominal cavity
C_2 Pelvic cavity
D Diaphragm

◁

Planes of the body

1 Transverse plane
2 Frontal plane
3 Sagittal plane (midsagittal)

Lines of direction

ant.	= anterior
inf.	= inferior
lat.	= lateral
med.	= medial
post.	= posterior
sup.	= superior

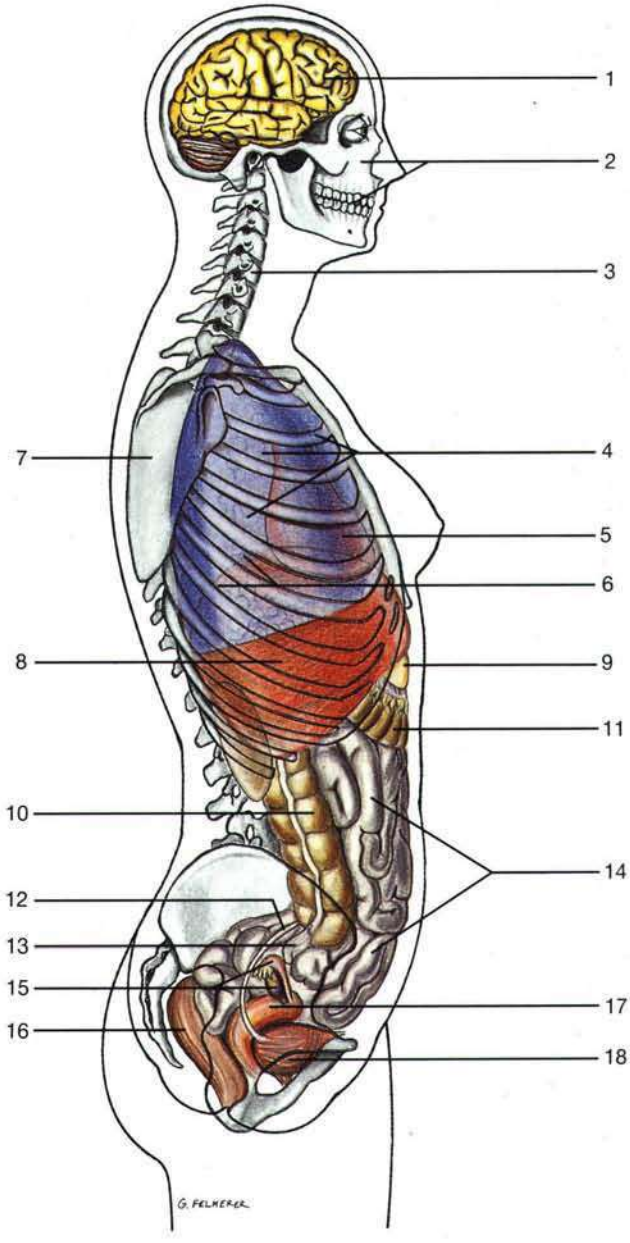

Position of the inner organs of the human body
(lateral aspect).
The three main cavities of the body and their contents.

1 Head (neurocranium) with the brain
2 Facial bones with oral and nasal cavities
3 Vertebral column (cervical part)
4 Thorax with the lungs
5 Heart
6 Surface projection of the diaphragm
7 Scapula
8 Liver
9 Stomach
10 Ascending colon
11 Transverse colon
12 Ureter
13 Appendix
14 Small intestine
15 Ovary, uterine tube
16 Rectum
17 Uterus
18 Urinary bladder

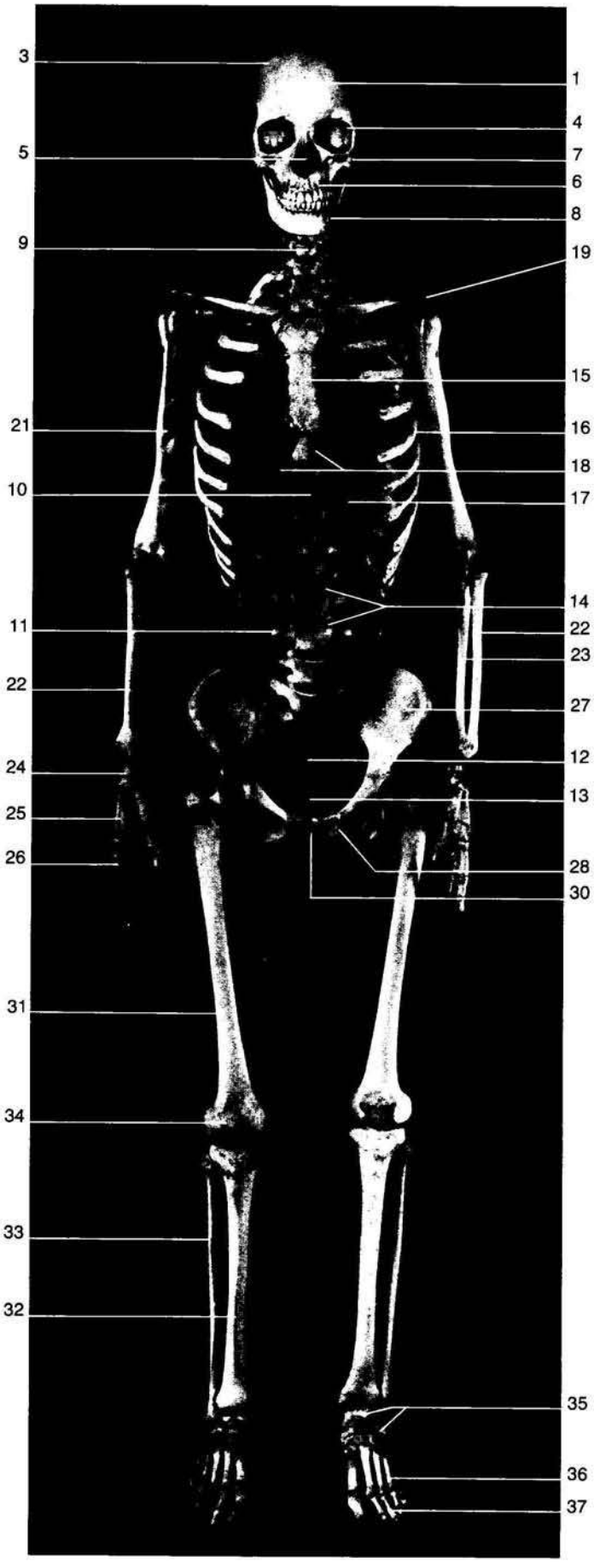

Skeleton of a female adult (anterior aspect).

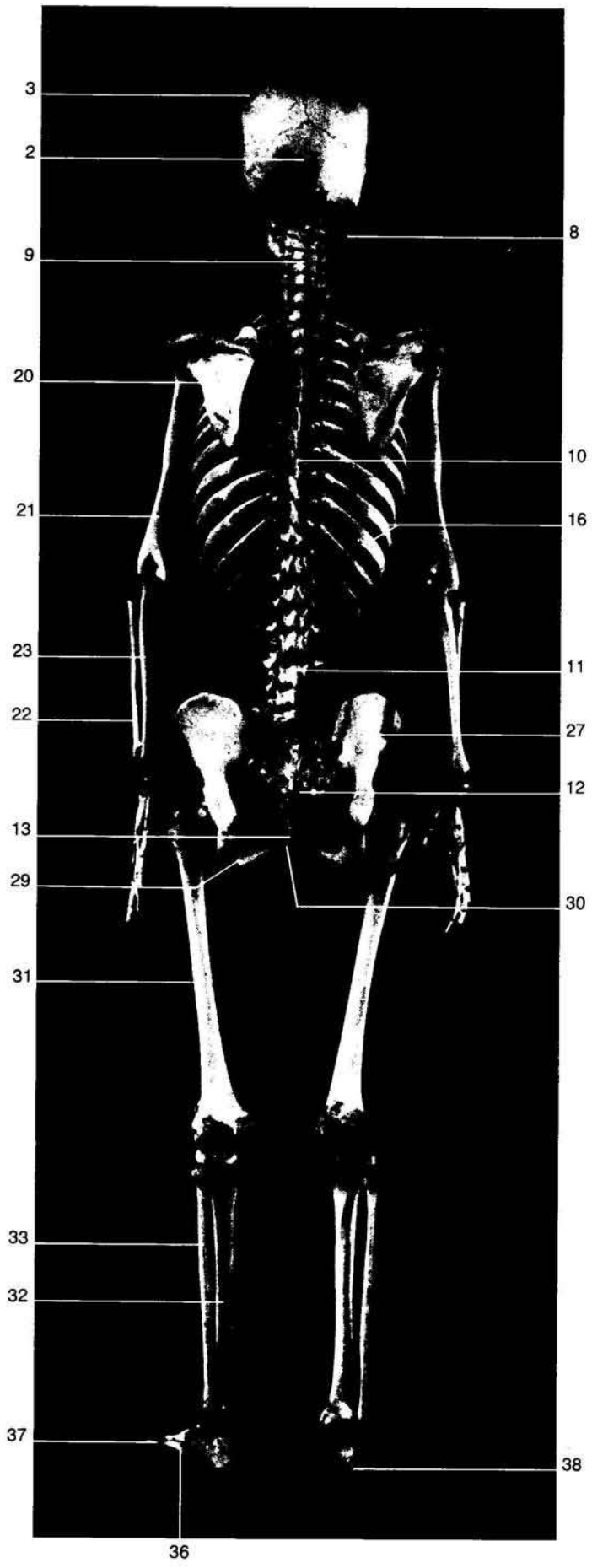

Skeleton of a female adult (posterior aspect).

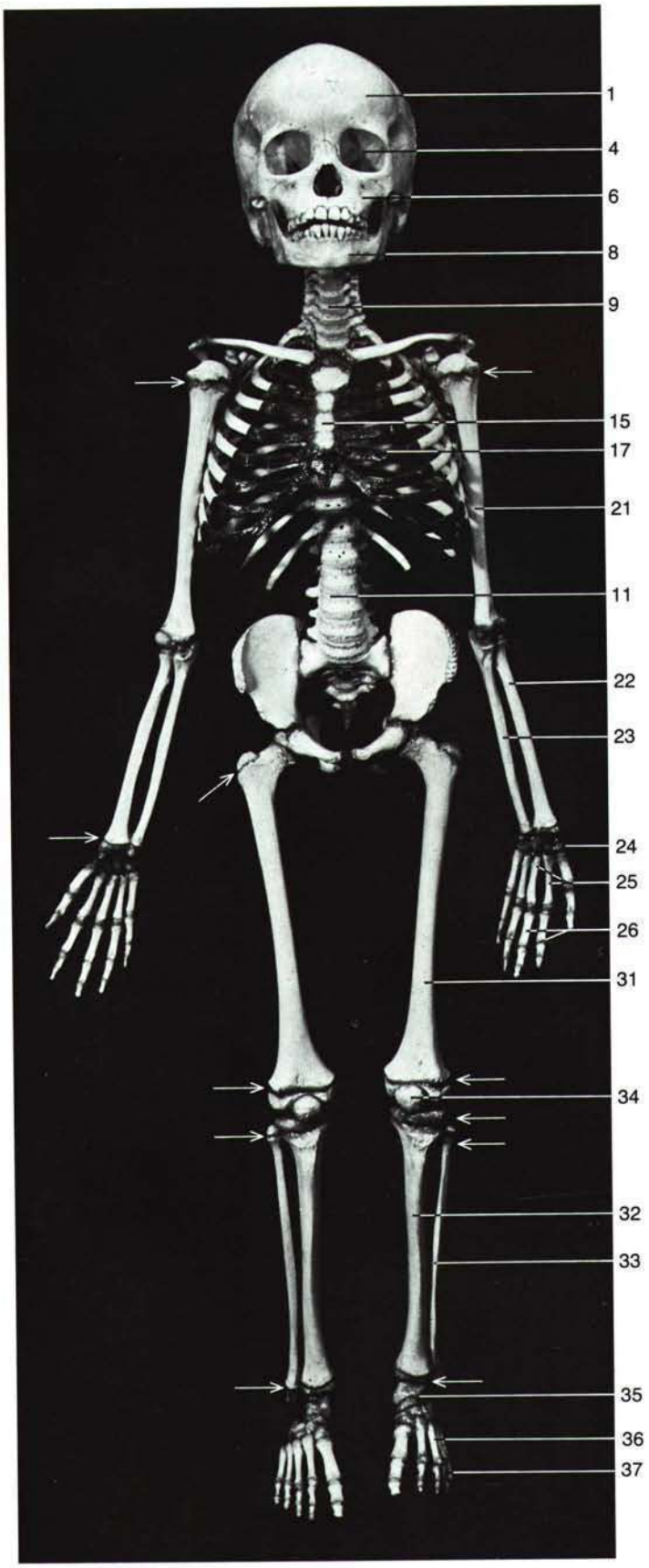

Axial skeleton

Head

1 Frontal bone
2 Occipital bone
3 Parietal bone
4 Orbit
5 Nasal cavity
6 Maxilla
7 Zygomatic bone
8 Mandible

Trunk and thorax

Vertebral column

9 Cervical vertebrae
10 Thoracic vertebrae
11 Lumbar vertebrae
12 Sacrum
13 Coccyx
14 Intervertebral discs

Thorax

15 Sternum
16 Ribs
17 Costal cartilage
18 Infrasternal angle

Appendicular skeleton

Upper limb and shoulder girdle

19 Clavicle
20 Scapula
21 Humerus
22 Radius
23 Ulna
24 Carpal bones
25 Metacarpal bones
26 Phalanges of the hand

Lower limb and pelvis

27 Ilium
28 Pubis
29 Ischium
30 Symphysis pubis
31 Femur
32 Tibia
33 Fibula
34 Patella
35 Tarsal bones
36 Metatarsal bones
37 Phalanges of the foot
38 Calcaneus

Skeleton of a 5-year-old child (anterior aspect).
The zones of the cartilaginous growth plates are seen (arrows).
In contrast to the adult, the ribs show a predominantly horizontal position.

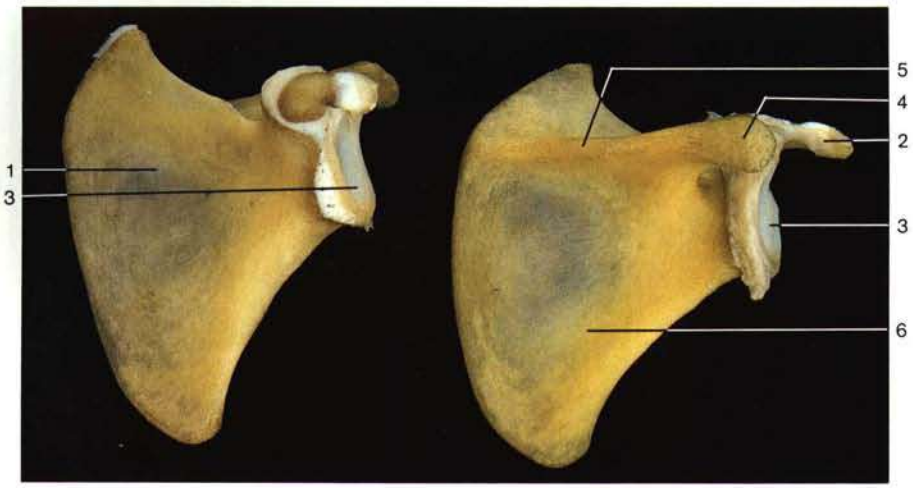

1 Subscapular fossa
2 Coracoid process
3 Glenoid fossa
4 Acromion
5 Spine of scapula
6 Infraspinous fossa

Ossification of the scapula
(left: anterior aspect, right: posterior aspect).

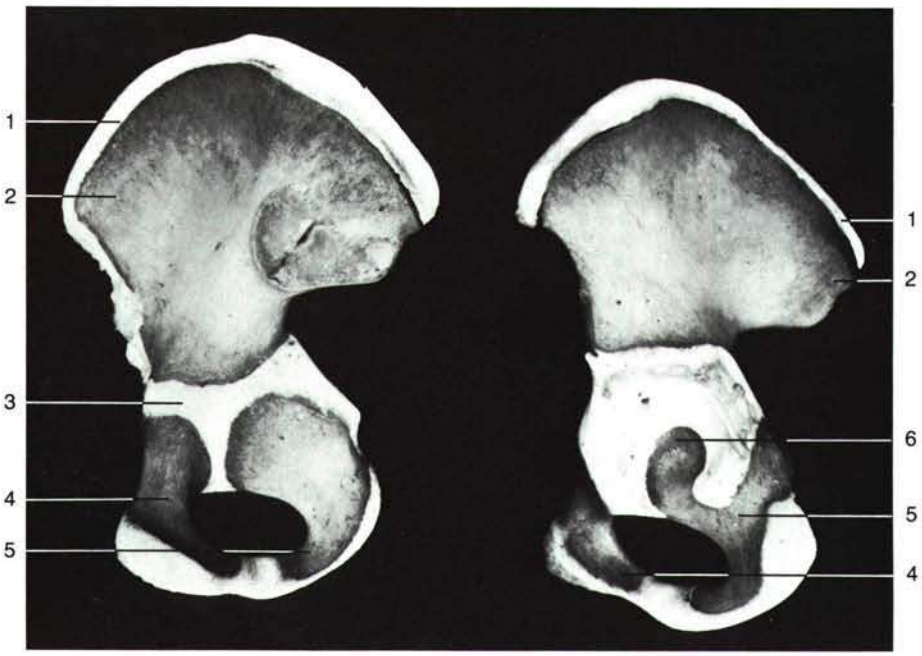

1 Cartilage of the iliac crest
2 Ilium
3 Cartilage
4 Pubis
5 Ischium
6 Acetabulum

Ossification of the hip bone
(left: medial aspect, right: lateral aspect).

1 Bone tissue (vertebral body)
2 Cartilaginous tissue (lateral epiphysis)
3 Intervertebral discs

Ossification of the sacrum (anterior aspect).
Note the five vertebral bones, which are still separated from each other.

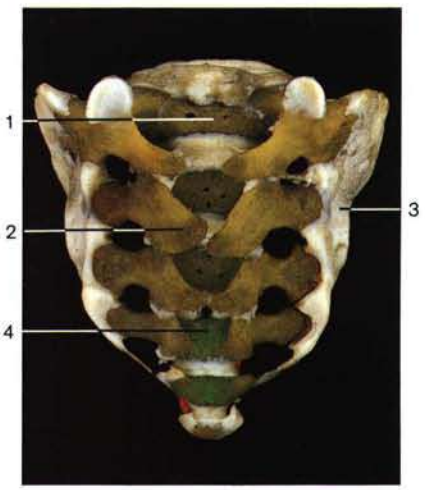

1 Bony tissue (center of ossification)
2 Vertebral arch (not completely united)
3 Cartilaginous tissue (lateral epiphysis)
4 Sacral canal

Ossification of the sacrum
(posterior aspect).

1 Ossification center in the head of the femur
2 Greater trochanter
3 Head of the femur
4 Neck of the femur
5 Lateral condyle
6 Medial condyle
7 Intercondylar notch
8 Diaphysis

Ossification of the femur (left: coronal section, right: posterior view of the femur). Arrows: distal epiphysis.

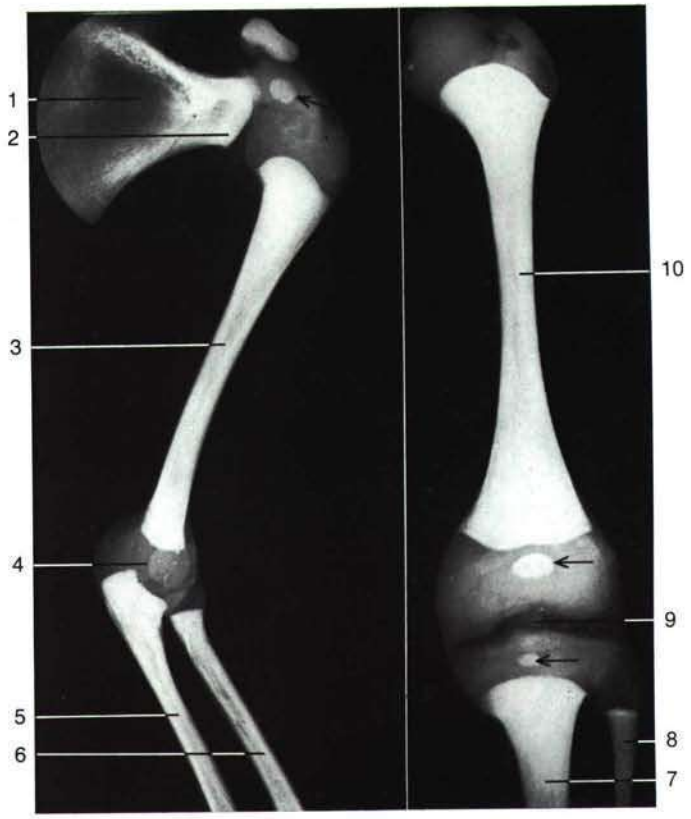

X-ray of the upper and lower limb of a newborn child.
Left: upper limb. Right: lower limb.
Arrows: ossification centers.

1 Scapula
2 Shoulder joint
3 Humerus
4 Elbow joint
5 Ulna
6 Radius
7 Tibia
8 Fibula
9 Knee joint
10 Femur

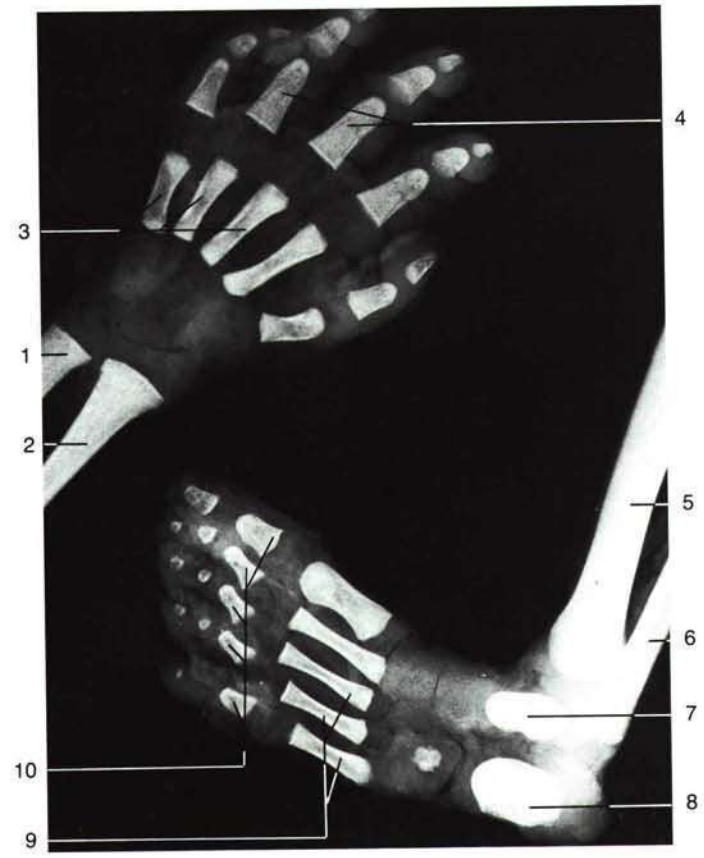

X-ray of hand and foot of a newborn.

1 Ulna
2 Radius
3 Metacarpals
4 Phalanges
5 Tibia
6 Fibula
7 Talus
8 Calcaneus
9 Metatarsals
10 Phalanges

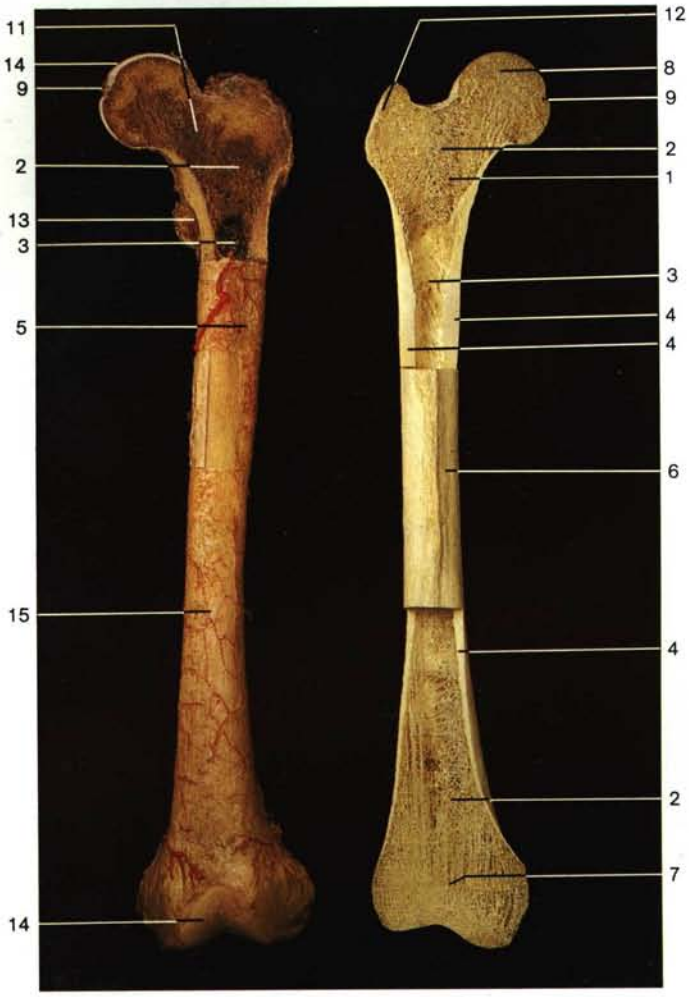

Femur of the adult. Left: the periosteum and the nutrient vessels are preserved. Right: coronal section of the proximal and distal epiphyses to display the spongy bone and the medullary cavity.

1 Metaphysis
2 Spongy bone
3 Medullary cavity in the diaphysis
4 Compact bone
5 Nutrient canal
6 Diaphysis
7 Epiphyseal line (remnants of the epiphyseal plate)
8 Epiphysis (head of the femur)
9 Fovea of head
10 Trabeculae of spongy bone
11 Neck of the femur
12 Greater trochanter
13 Lesser trochanter
14 Articular surface
15 Periosteum
16 Skin
17 Vastus medialis muscle
18 Sartorius muscle
19 Femoral artery and vein
20 Great saphenous vein
21 Gracilis muscle
22 Adductor longus muscle
23 Adductor magnus muscle
24 Semimembranosus muscle
25 Semitendinosus muscle
26 Rectus femoris muscle
27 Vastus lateralis muscle
28 Femur and medullary cavity
29 Vastus intermedius muscle
30 Sciatic nerve
31 Biceps femoris muscle
32 Spongy bone trabeculae containing bone marrow
33 Compact bone
34 Osteon with Haversian lamellae
35 Periosteum
36 Blood vessels and nerves for periosteum and bone

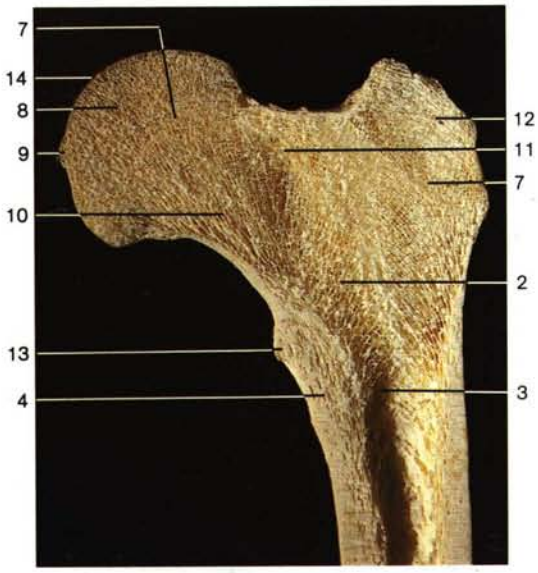

Coronal section through the proximal end of the adult femur, revealing the characteristic trajectorial structure of the spongy bone.

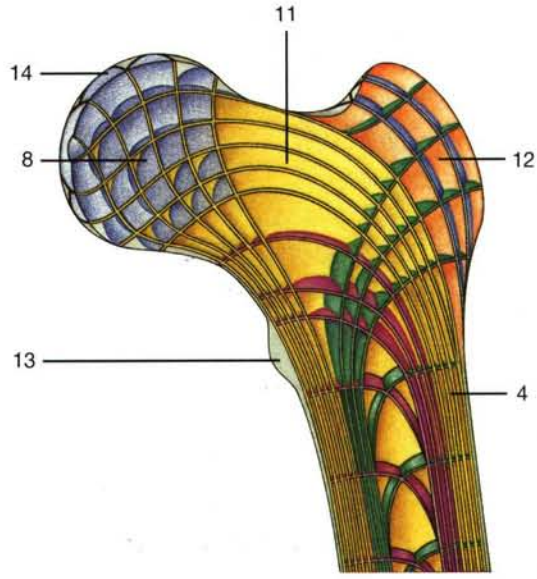

Three dimensional representation on the trajectorial lines of the femoral head (according to B. Kummer).

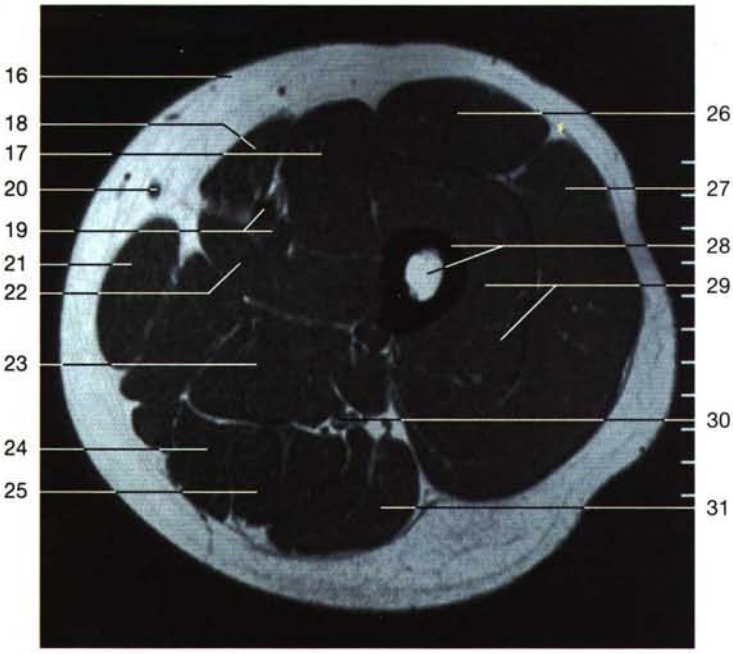

MR-Image of the thigh (axial section through the middle of the left thigh, the same level as the CT-Image).

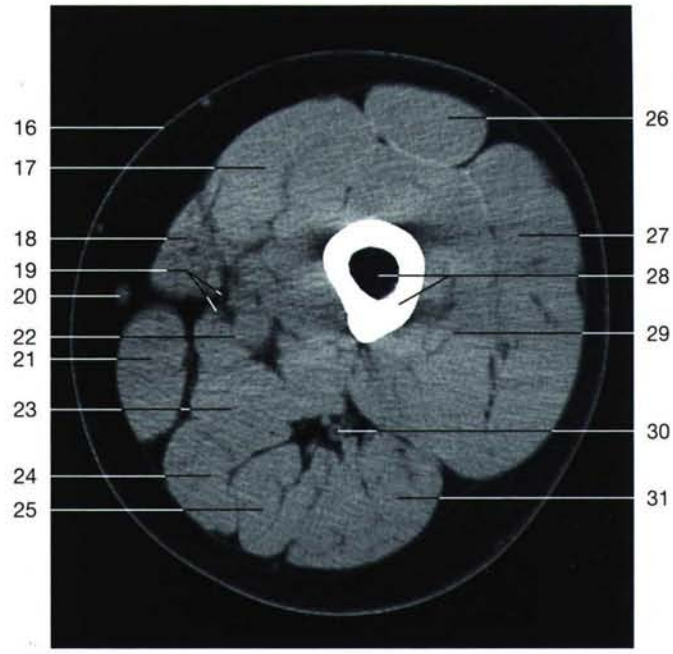

CT-Image of the section through the middle of the left thigh (axial section). Note the differences between the CT- and MR-Image (see text below).

The **bones** of the skeletal system consist of two different parts, the spongy and the compact bone. The spongy bone trabeculae are highly adapted to mechanical forces revealing a trajectorial structure. The intertrabecular spaces are filled with bone marrow, the site of blood formation. The appearance of bones, muscles and soft tisssues is quite different in **CT-** and **MR-Images**. The CT-images relate well to radiographs in that areas of great absorption such as bones are white, and those with little absorption such as fat appear black. In contrast, the intensity of signals in MR-Images, obtained without X-rays but by magnetic forces, is different so that dense areas of bones appear black and soft tissues such as bone marrow and fat appear white (for comparison see above figures).

A highly innervated **periosteum** is an essential structure for bone nutrition, blood supply, growth and bone repair.

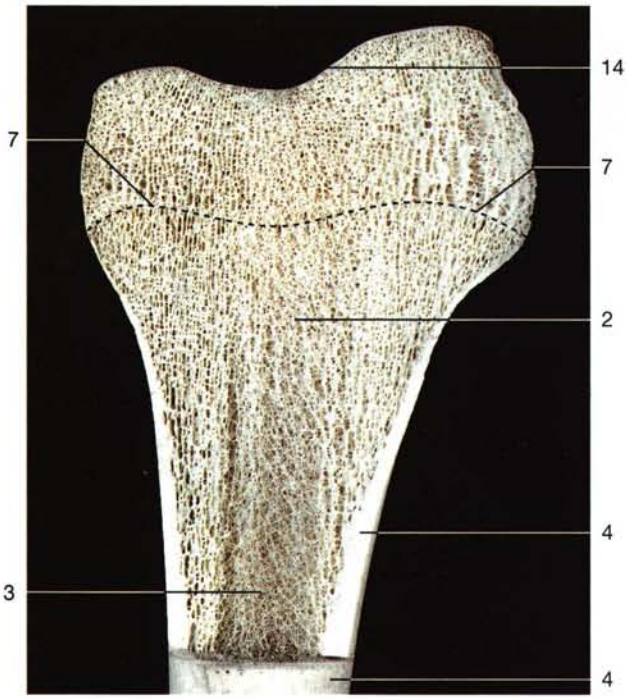

Coronal section through the proximal epiphysis of the adult tibia. Note the zone of dense bone at the site of the former epiphyseal plate (dotted line).

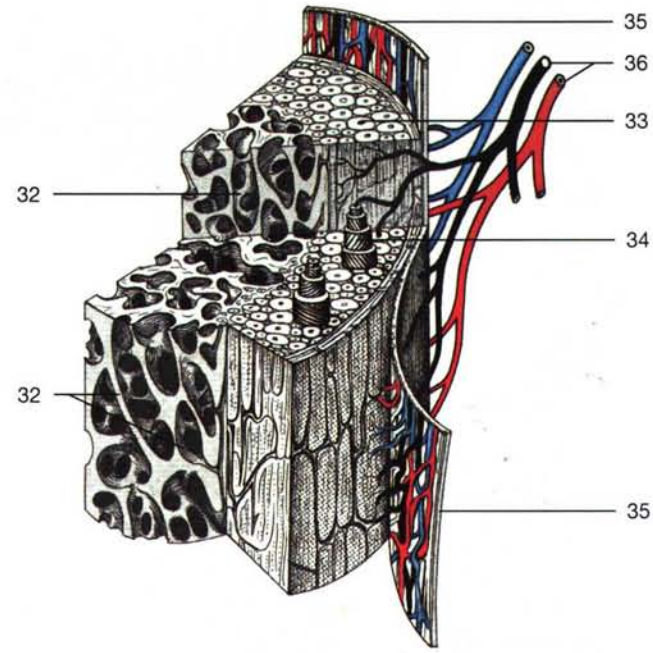

Structure of bones of the skeletal system (after Benninghoff). Note that the compact bone reveals a lamellar structure with Haversian lamellae and canals.

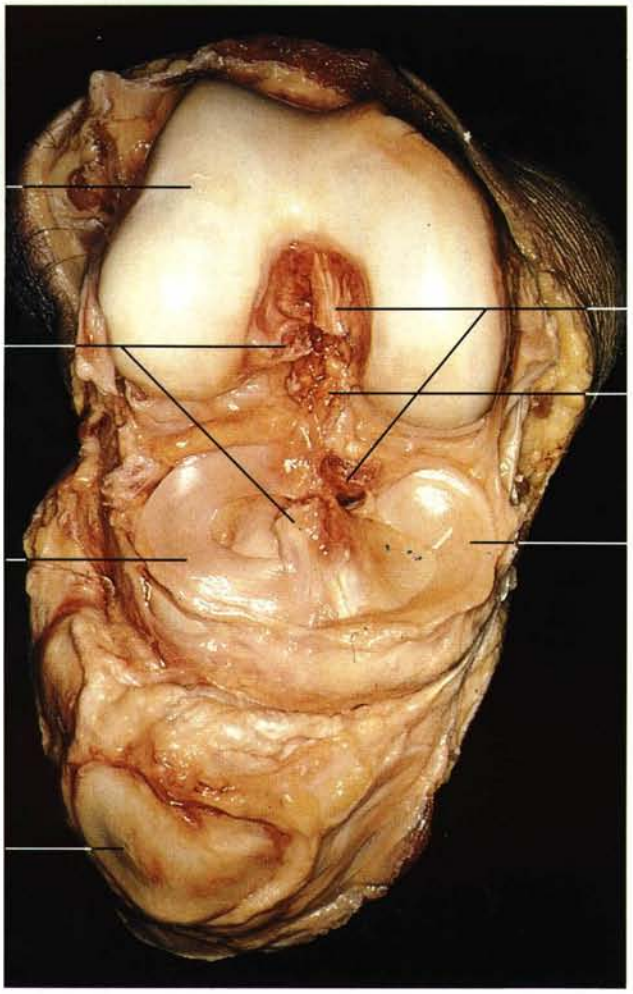

Knee joint. Anterior aspect, showing menisci and cruciate ligaments (cut). Quadriceps tendon cut and patella reflected distally.

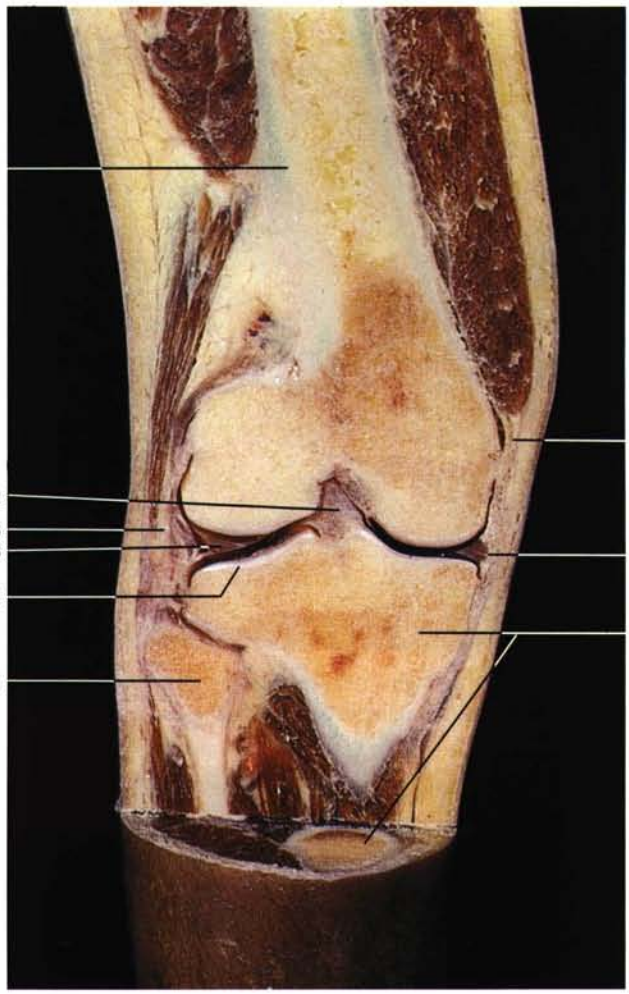

Coronal section through the knee joint. Anterior aspect of the right joint in extension.

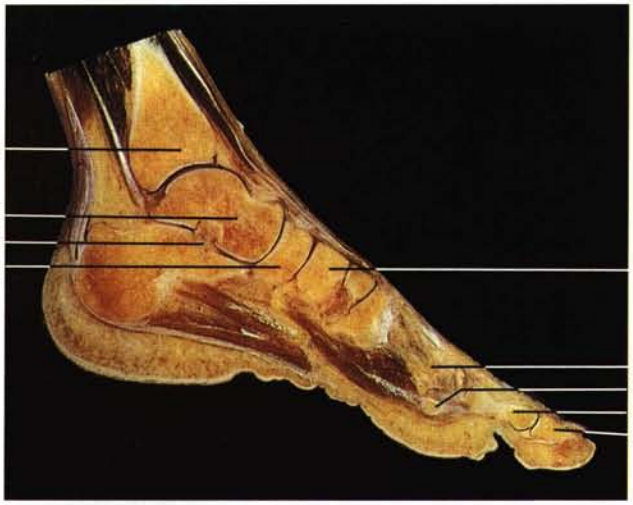

Sagittal section through the lower limb and the foot.

1 Femur
2 Anterior cruciate ligament
3 Lateral meniscus
4 Patella
5 Posterior cruciate ligament
6 Posterior meniscofemoral ligament
7 Medial meniscus
8 Fibular collateral ligament
9 Fibula
10 Tibia
11 Tibial collateral ligament
12 Articular cartilage
13 Articular capsule
14 Tendon of long head of biceps brachii muscle
15 Head of humerus
16 Glenoid labrum
17 Articular cartilage of glenoid fossa

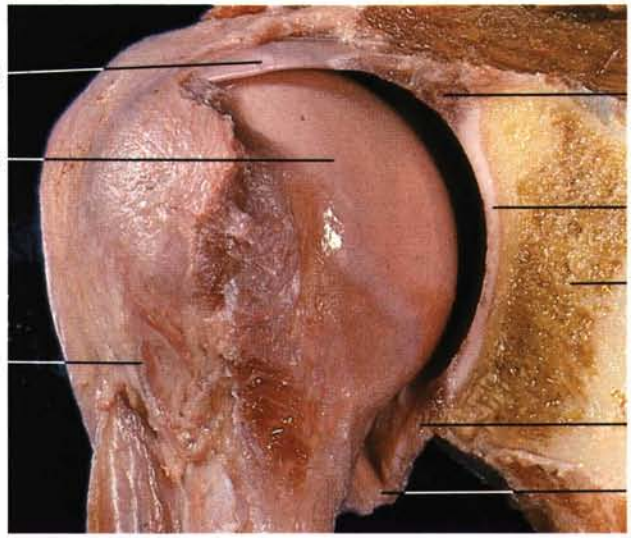

Shoulder joint (anterior view). The anterior part of the articular capsule has been removed.

18 Scapula
19 Talus
20 Interosseous talocalcaneal ligament
21 Navicular bone
22 Medial cuneiform bone
23 First metatarsal bone
24 Proximal phalanx of the hallux (great toe)
25 Distal phalanx of the hallux
26 Sesamoid bone

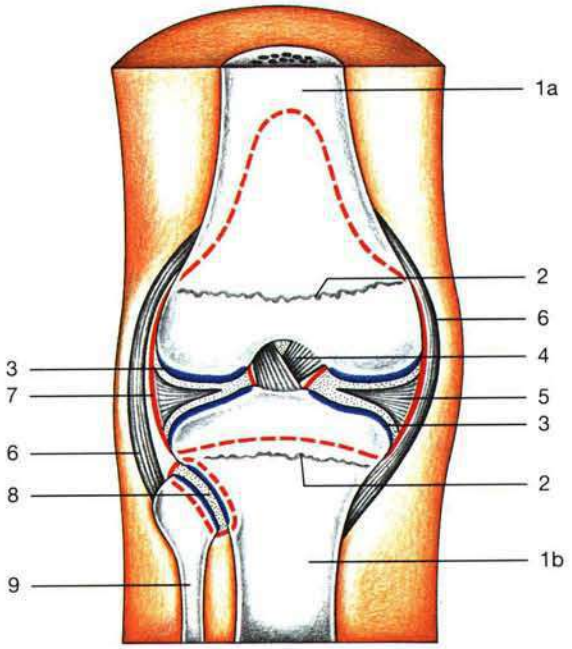

General architecture of a synovial joint with 2 articulating bones and a synovial cavity (right side, anterior view).
Coronal section throught the knee joint.
Red line = Articular capsule with synovial membrane.
Dotted red line = extension of articular capsule (suprapatellar bursa).

1 Articulating bones: a) Femur, b) Tibia
2 Epiphysial line
3 **Articular cartilage**
4 Intraarticular ligaments (e.g. cruciate ligaments)
5 Fibrocartilaginous disk (e.g. meniscus)
6 Collateral ligaments
7 Articular capsule with synovial membrane
8 **Tibiofibular articulation** (example of gliding synovial joint)
9 Fibula

	Movement	Examples
A Fibrous joints		
1 Sutures	No movements	Sutures of the skull
2 Syndesmoses	No movements	Distal tibiofibular joint
3 Gomphosis	No movements	Roots of teeth in alveolar process
B Cartilaginous joints		
1 Synchondroses	No movements	Epiphyseal plates
2 Symphyses	Slight movement	Symphysis pubis intervertebral discs
C Synovial joints		
1 Gliding	Monaxial	Intercarpal joint Intertarsal joint Sacroiliacal joint
2 Hinge	Monaxial	Interphalangeal joint Humeroulnar joint Talocrural joint
3 Pivot	Monaxial	Atlantoaxial joint Radioulnar joint
4 Ellipsoidal	Biaxial	Radiocarpal joint
5 Saddle	Biaxial	Carpometacarpal joint of the thumb
6 Ball-and-socket	Multiaxial	Shoulder and hip joint

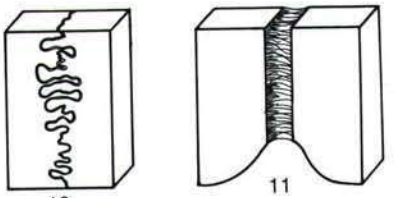

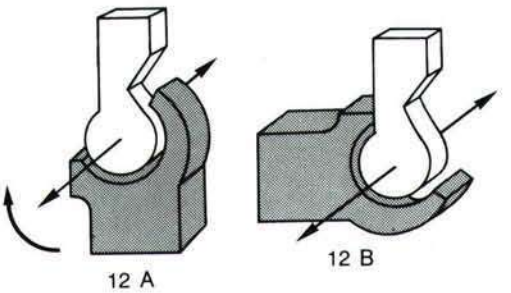

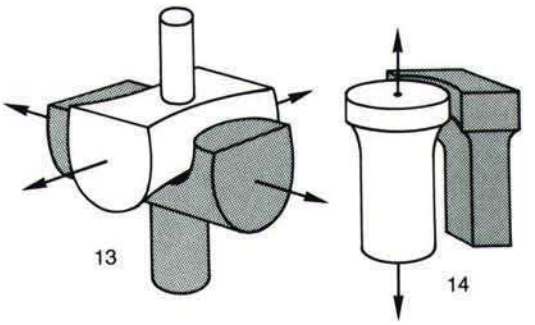

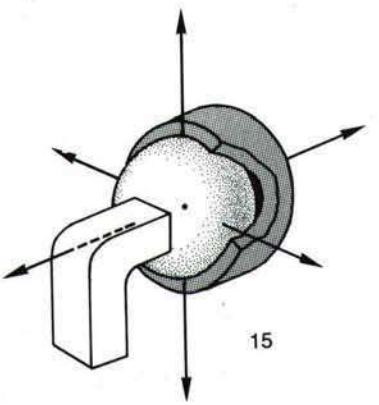

Main types of joints. Arrows: axes of movement.

Fibrous joints (synarthroses)
10 Serrate suture
11 Syndesmosis

Synovial joints (diarthroses)
12 Hinge joints (monaxial) ginglymus
A Extension
B Flexion
13 Saddle joint (biaxial)
14 Pivot joint (monaxial, rotation)
15 Ball-and-socket joint (multiaxial)

An articulation or joint is the functional connection between two or more bones. Joints can be divided into two categories depending upon whether the articulating surfaces of the bones are separated by a real cavity (joint cavity) so that they are movable against each other **(synovial joints)** or whether the bones are firmly connected by fibrous or cartilaginous tissue and practically immovable (**fibrous joints, cartilaginous joints,** symphysis, etc.). Synovial joints always possess a joint capsule (with a vascularized synovial membrane), articular cartilages, and a joint cavity. They are grouped according to the degree of movement they permit. A hinge joint (ginglymus) permits movement in only one plane about a single axis **(uniaxial** or **monaxial),** an ellipsoidal joint permits movements in two planes **(biaxial),** and ball-and-socket joints permit a range of movements around several axes **(multiaxial).** The following survey gives a few examples of these types of articulation.

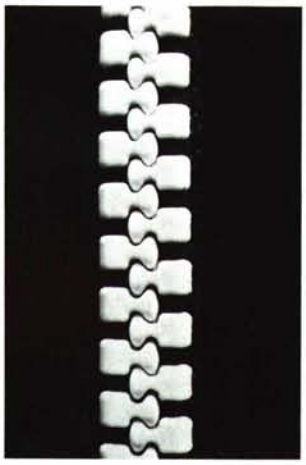

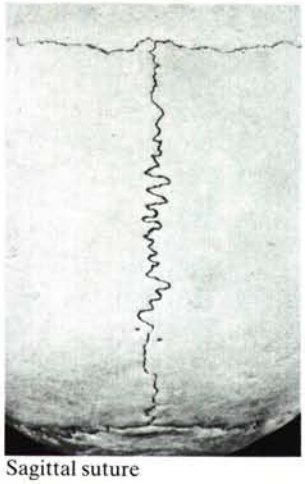

Sagittal suture

Separated

1. Fibrous joints

A Sutures

a) Serrate or dentate suture

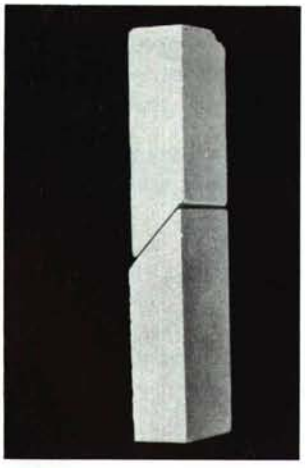

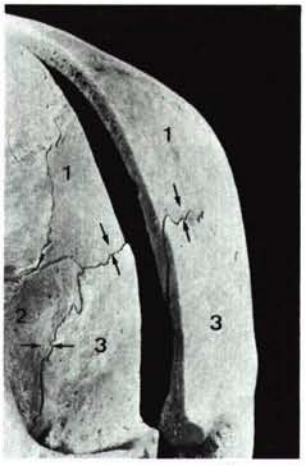

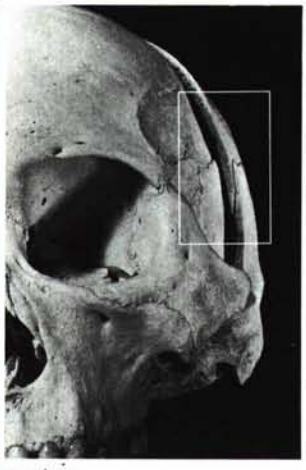

Squamous suture

b) Squamous suture (arrows)

1 Parietal bone
2 Sphenoid bone
3 Squama of temporal bone

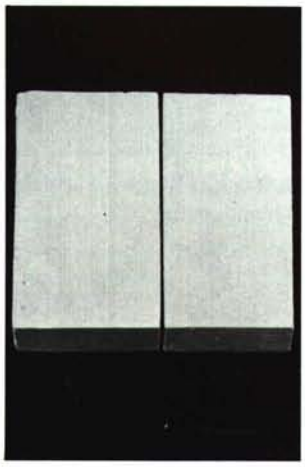

Nasal bones

c) Plane suture (arrow)

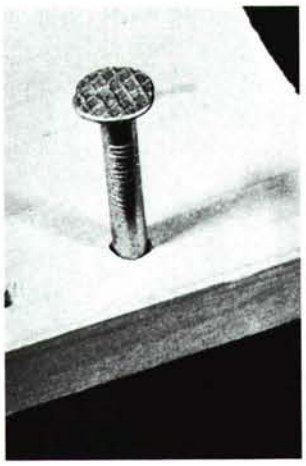

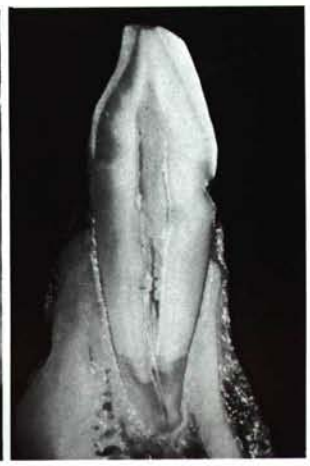

B Peg suture
(Gomphosis)

Teeth and alveolar process

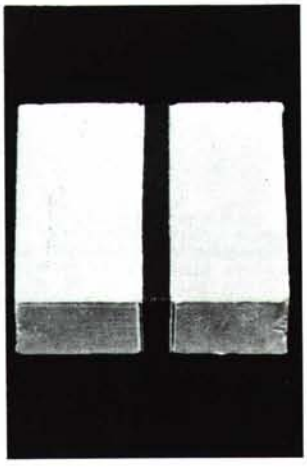

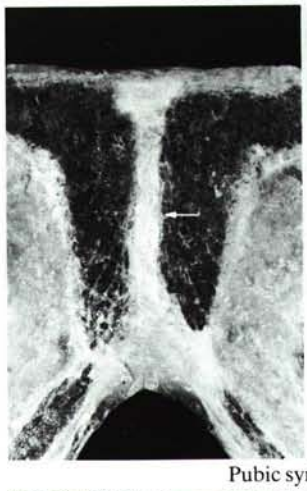

2. Cartilaginous joints

a) Symphysis
(fibrocartilage)

Pubic symphysis

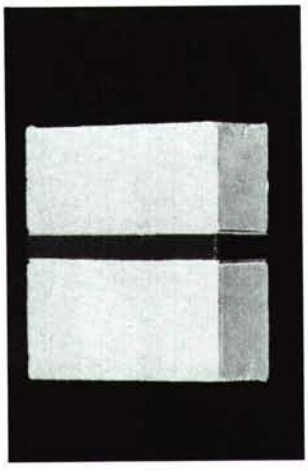

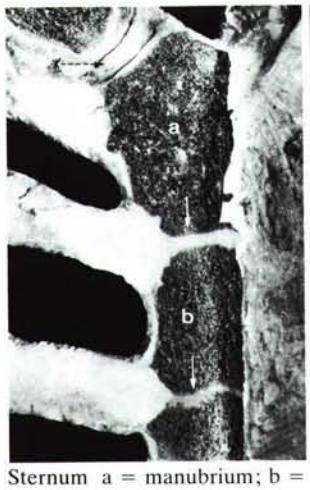

b) Synchondrosis
(hyaline cartilage)

* Articular disc
(sternoclavicular joint)

Sternum a = manubrium; b = body

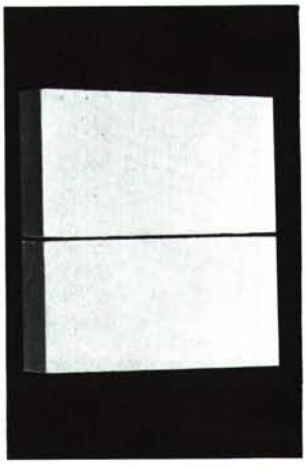

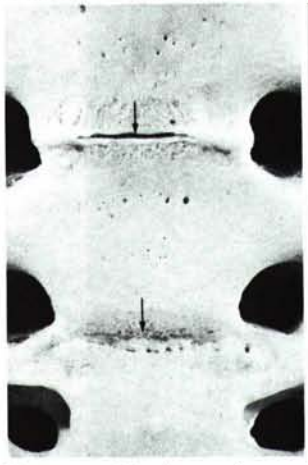

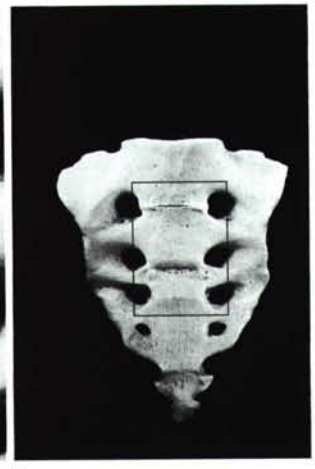

3. Osseous joints
(synostosis)

Transverse ridge (arrows) Sacrum

1. Ball-and-socket joint

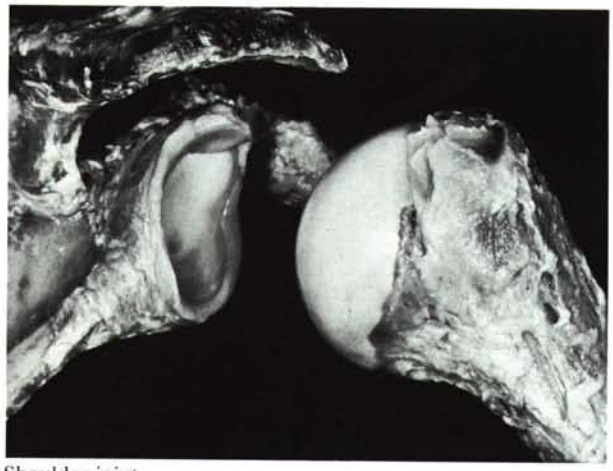

Shoulder joint

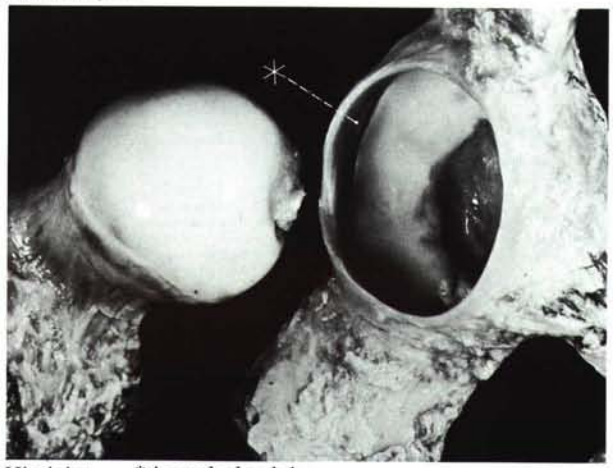

Hip joint *Acetabular labrum

2. Ellipsoid joint

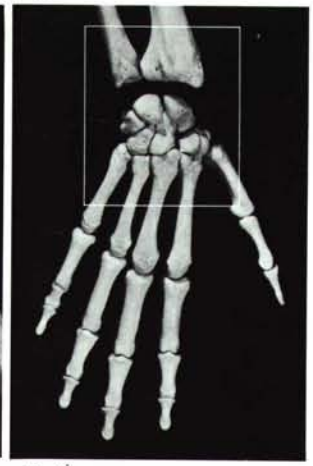

Radiocarpal joint

3. Hinge joint

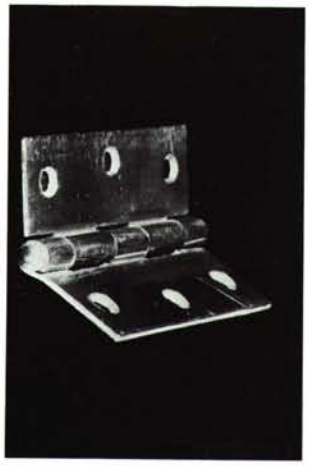

Interphalangeal joint

Elbow joint (humeroulnar joint)

4. Pivot joint

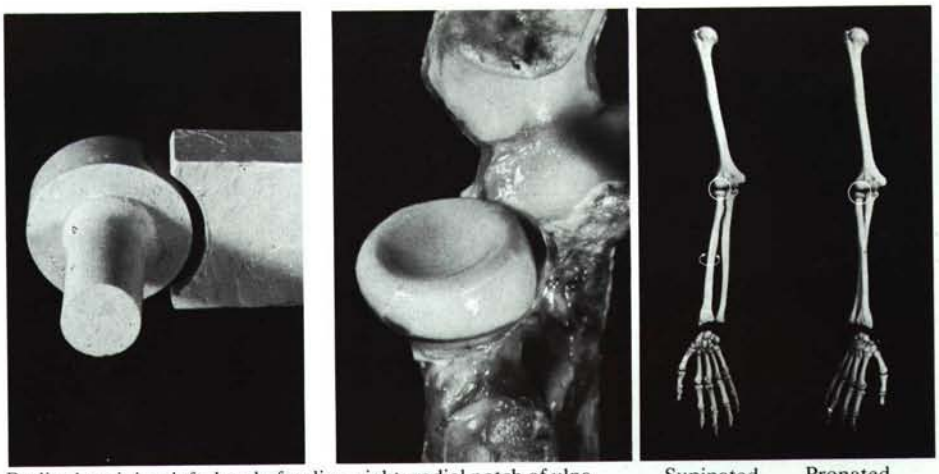

Radioulnar joint; left: head of radius, right: radial notch of ulna

Supinated Pronated

5. Hinge joint

(but a gradually tilting axis produces a slight spiral motion)

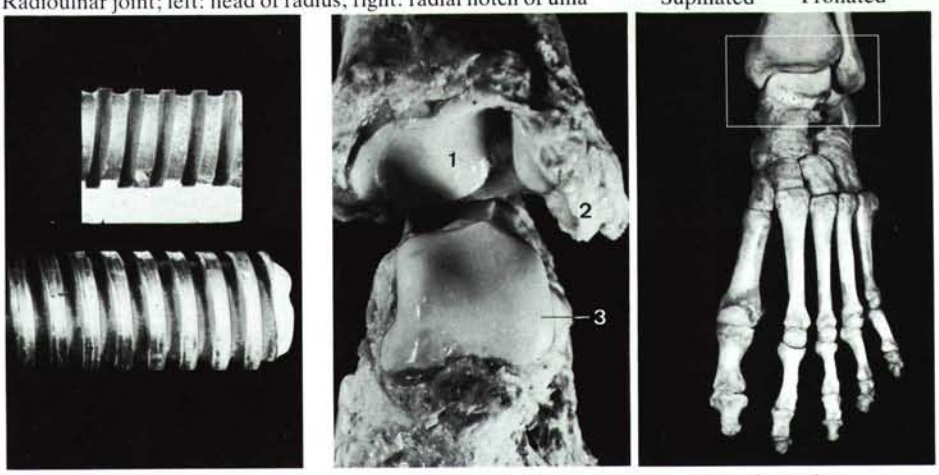

Ankle joint: 1 = tibia; 2 = fibula; 3 = talus

6. Saddle joint

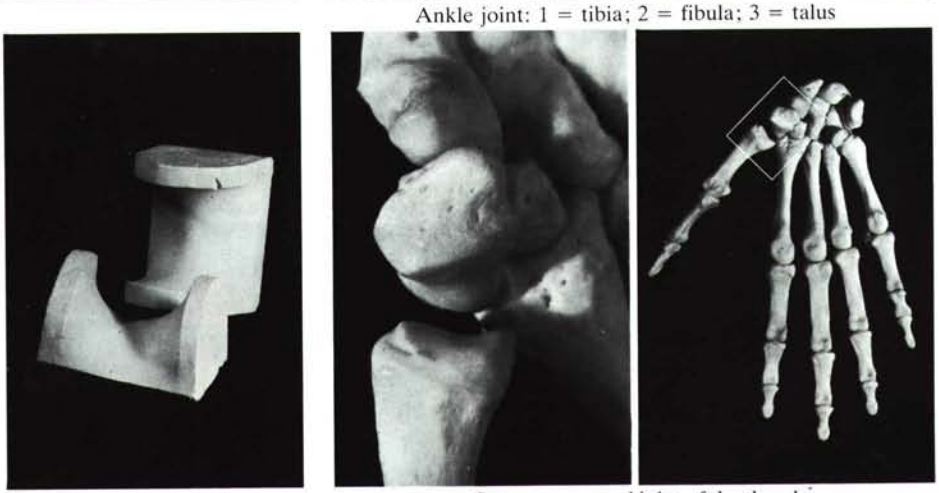

Carpometacarpal joint of the thumb

7. Plane joint

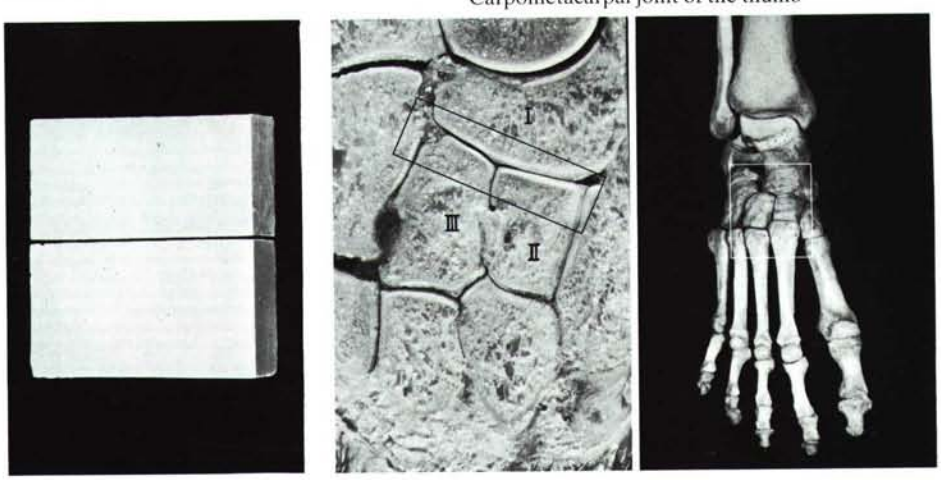

I = navicular; **II** = intermediate cuneiform; **III** = lateral cuneiform

Fusiform
(palmaris longus muscle)

Bicipital
(biceps brachii muscle)

Tricipital
(triceps surae muscle)

Quadricipital
(quadriceps femoris muscle)

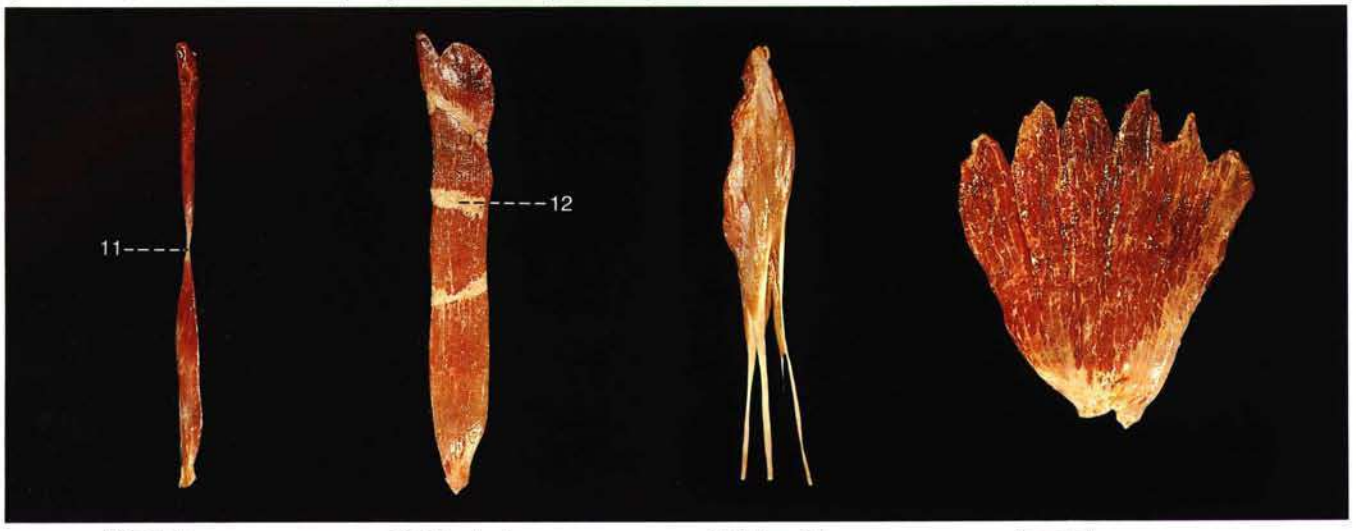

Digastric
(omohyoid muscle)

Multiventral
(rectus abdominis muscle)

Multicaudal
(flexor prof. muscle)

Serrated
(serratus anterior muscle)

Bipennate
(tibialis anterior muscle)

Unipennate
(semimembranous muscle)

Semitendinous
(semitendinous muscle)

Broad, flat muscle
(latissimus dorsi muscle)

Ring-like
(external anal sphincter muscle)

1 Long head
2 Short head
3 Gastrocnemius muscle (medial head, lateral head)
4 Soleus muscle
5 Achilles tendon
6 Vastus intermedius muscle
7 Vastus lateralis muscle
8 Patella
9 Rectus femoris muscle
10 Vastus medialis muscle
11 Intermediate tendon
12 Tendinous intersection
13 Aponeurosis
14 Tendinous intersection

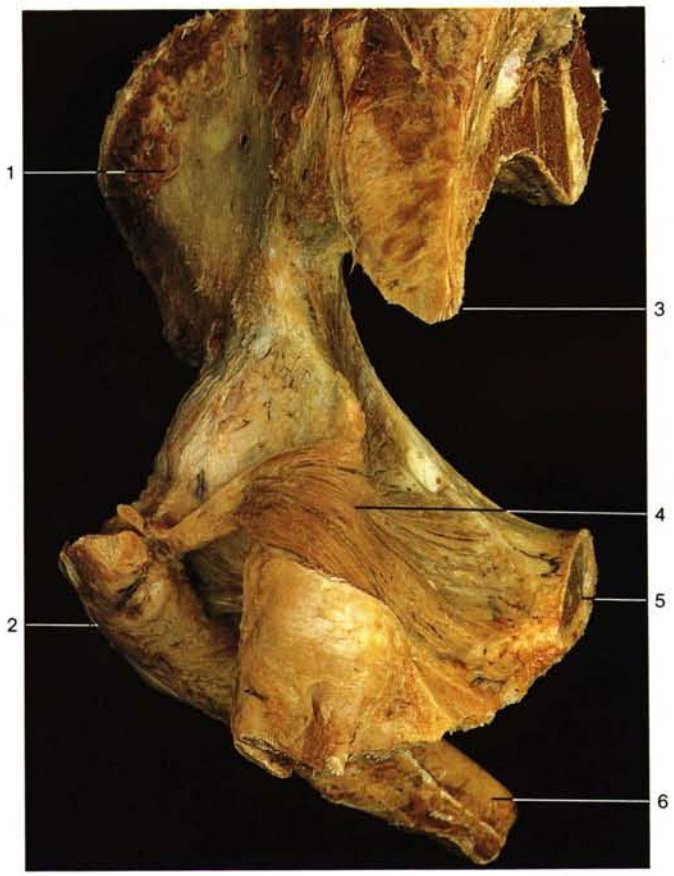

Left half of the pelvis (posterior aspect).
Obturator internus muscle as an example of a muscle, the tendon of which does not act in the direction of the main muscle fibers. Its fibers originate at the internal aspect of the obturator foramen, turn around the posterior rim of the ischium and insert at the greater trochanter of the femur. The ischium thereby serves as a pulley.

1 Ilium
2 Greater trochanter
3 Coccyx
4 Obturator internus muscle
5 Pubis
6 Femur

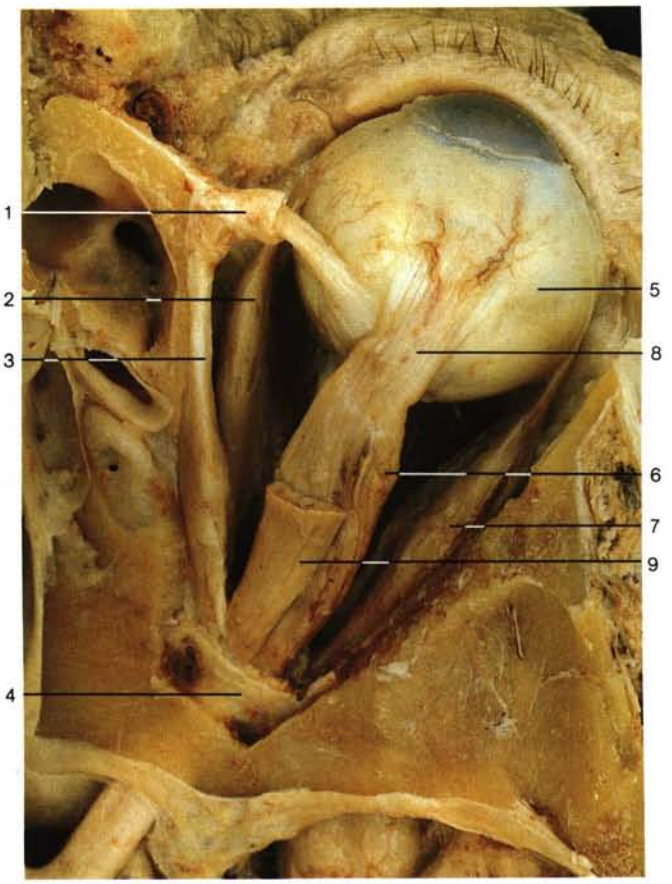

Superior oblique muscle of the eyeball, right eye (superior aspect). The tendon of this muscle bends over the trochlea changing its direction so that it becomes attached to the posterior lateral quadrant of the eyeball.

1 Trochlea
2 Medial rectus muscle
3 Superior oblique muscle
4 Common annular tendon
5 Eyeball
6 Superior rectus muscle
7 Lateral rectus muscle
8 Superior rectus muscle (tendon)
9 Levator palpebrae superioris muscle (divided)

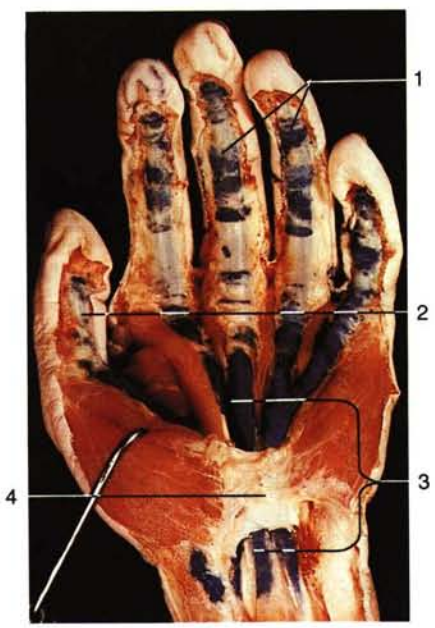

1 Digital synovial sheaths of the tendons of flexor digitorum superficialis and profundus muscles
2 Digital synovial sheaths of the tendon of long flexor pollicis longus muscle
3 Common flexor synovial sheaths of flexor digitorum superficialis and profundus muscles
4 Flexor retinaculum

The synovial sheaths of the tendons on the palmar aspect of the left wrist (colored fluid has been injected).

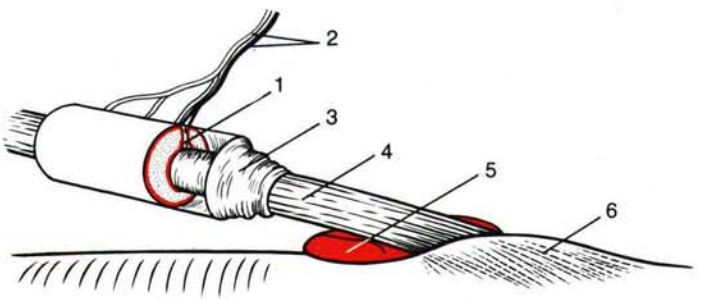

1 Mesotendon
2 Blood vessels
3 Synovial sheath
4 Tendon
5 Synovial bursa
6 Bone (tuberosity)

Structure of a tendon sheath. The synovial membrane which also forms the mesotendon is indicated in red. (Schematic drawing.)

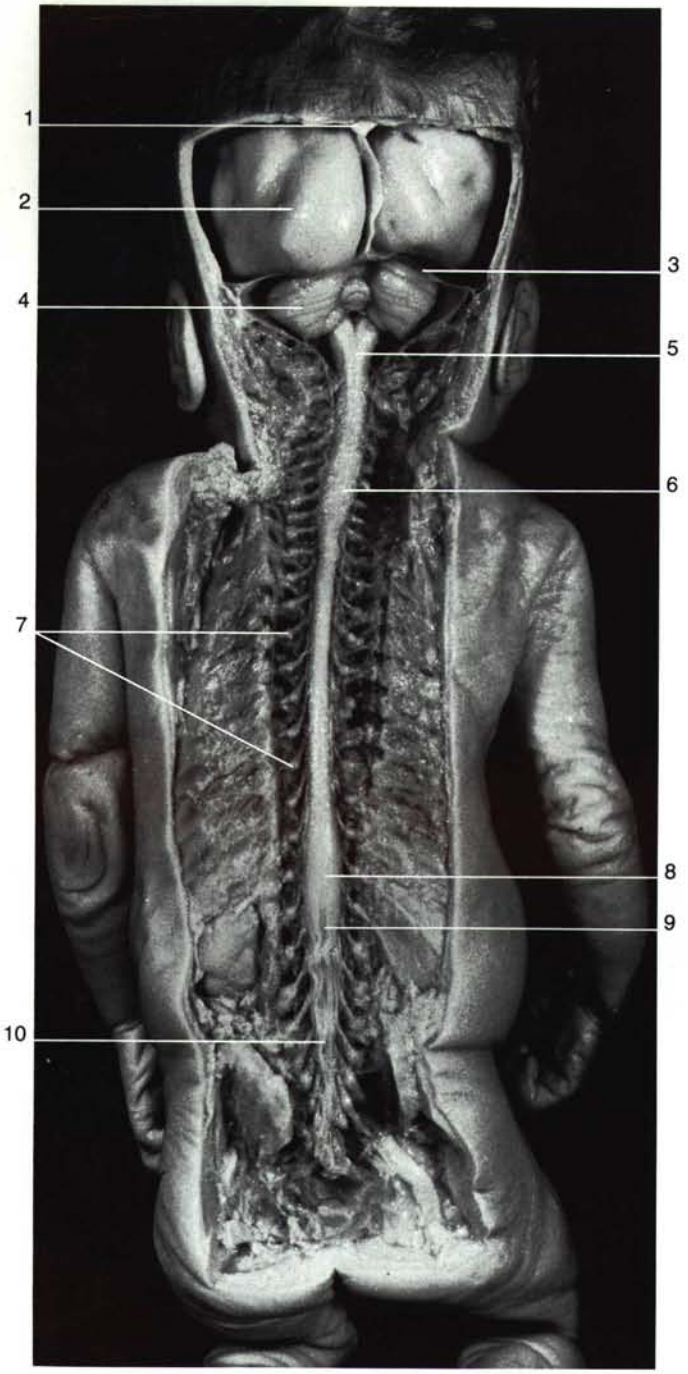

Brain, the **spinal cord** and the **spinal nerves** in the fetus (posterior aspect).

1 Falx cerebri
2 Cerebral hemispheres
3 Tentorium cerebelli
4 Cerebellum
5 Medulla oblongata
6 Spinal cord, cervical enlargement
7 Spinal ganglia
8 Spinal cord, lumbar enlargement
9 Conus medullaris
10 Cauda equina
11 Cervical plexus (formed from ventral rami of C_1–C_4)
12 Brachial plexus (formed from ventral rami of C_5–T_1)
13 Lumbosacral plexus (formed from ventral rami of L_1–S_4)
14 Sympathetic trunk

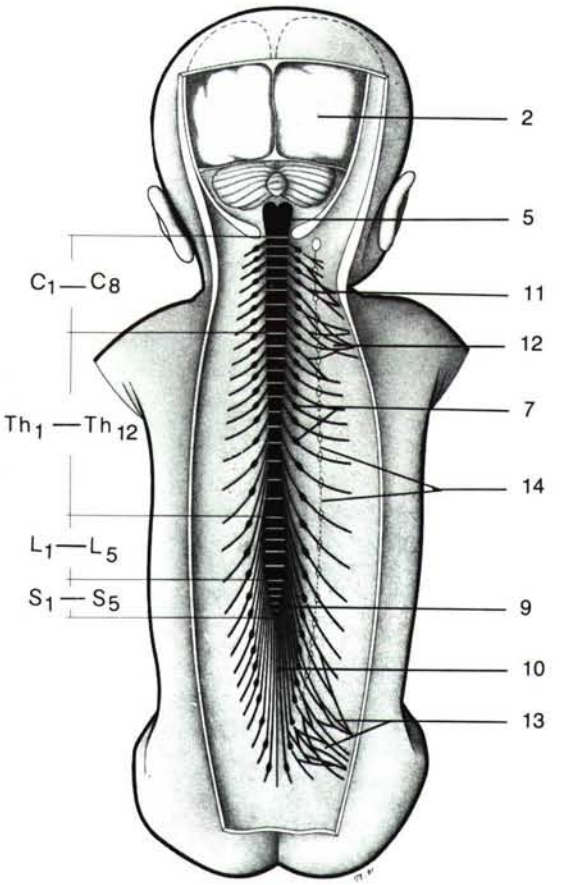

Schematic drawing to illustrate the three main parts of the nervous system in general.

The **nervous system** can be divided into three, functionally distinct parts: 1. the cranial part which comprises the great sensory organs and the brain, 2. the spinal cord which shows a segmental structure and serves predominantly as a reflex-organ, and 3. the autonomic nervous system which controls the unvoluntary functions (subconscious control) of organs and tissues. The autonomic part of the nervous system forms many delicate plexus within the organs. At certain places these plexus contain aggregations of nerve cells (prevertebral and intramural ganglia). The spinal nerves leave the spinal cord at regular intervals, forming the 8 cervical, 12 thoracic, 5 lumbar, 5 sacral and a varying number of coccygeal segments. The ventral rami of the first four cervical spinal nerves (C_1–C_4) form the cervical plexus (for innervation of the anterior neck), the ventral rami of the lower cervical spinal nerves (C_5–T_1) form the brachial plexus which innervates the upper leg, and the ventral rami of the lumbar and sacral spinal nerves form the lumbosacral plexus (L_1–S_4) which innervates the pelvic and genital organs and the lower extremity.

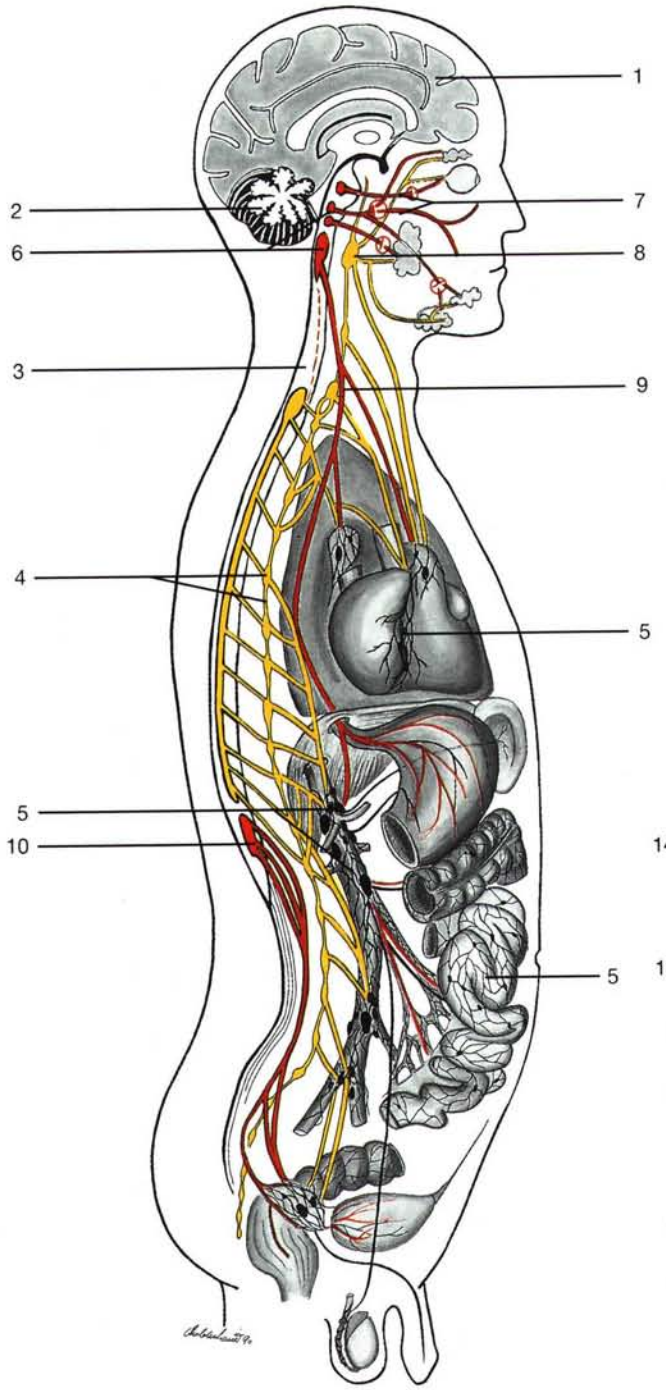

Diagram illustrating the localization of the **three functional portions** of the **nervous system** (brain, spinal cord and autonomic nervous system). Yellow = sympathetic system; red = parasympathetic system.

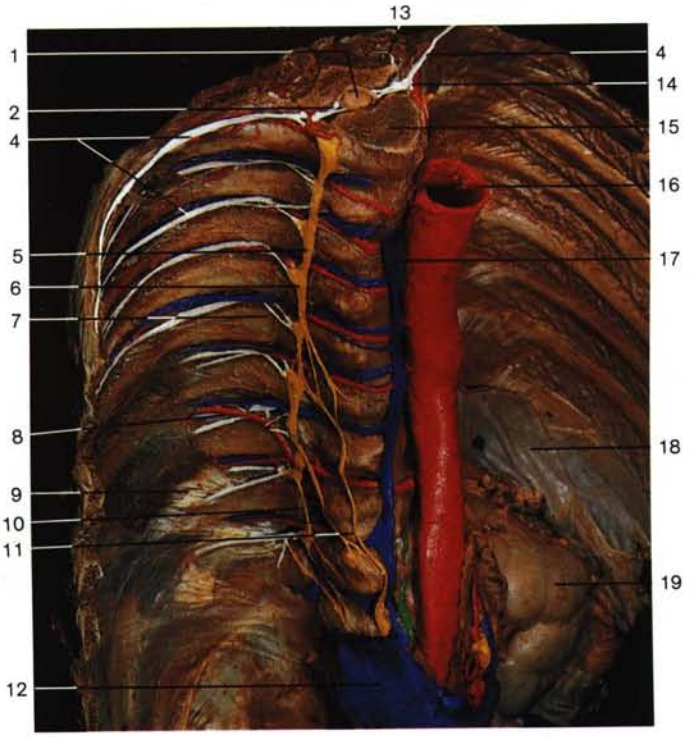

Posterior part of the thorax. Cross-section at the level of the 5th thoracic segment. Spinal nerves and their connections to the sympathetic trunk.

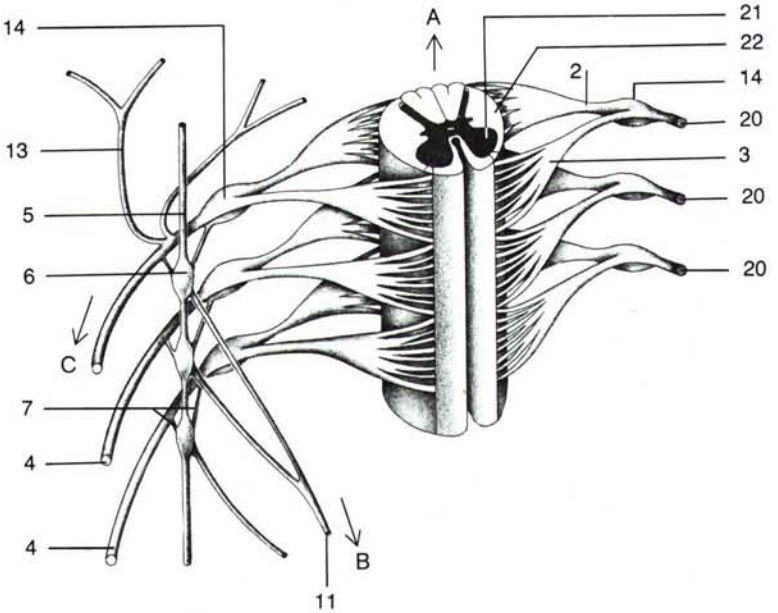

Organization of the spinal cord in structurally equal segments which form the paired spinal nerves. A = connections to the brain; B = connections to the autonomic nervous system; C = connections to the trunk and extremities (intercostal nerves and plexus). (Schematic drawing.)

1 Cerebrum
2 Cerebellum
3 Spinal cord
4 Sympathetic trunk and ganglion
5 Plexus and ganglia of the autonomic nervous system
6 Cranial autonomic system
7 Cranial nerves (n. III and n. VII)
8 Superior cervical ganglion
9 Vagus nerve (n. X)
10 Sacral autonomic system

1 Spinal cord
2 **Dorsal root**
3 **Ventral root**
4 Intercostal nerves
5 Sympathetic trunk
6 Ganglia of the sympathetic trunk
7 Rami communicantes
8 Intercostal artery and vein
9 Subcostalis muscle
10 Lesser splanchnic nerve
11 Greater splanchnic nerve
12 Inferior vena cava
13 Dorsal ramus of spinal nerve
14 **Spinal (dorsal root) ganglion**
15 Body of the vertebra
16 Aorta
17 Azygos vein
18 Diaphragm
19 Left kidney
20 **Spinal nerve**
21 Gray matter of the spinal cord
22 White matter of the spinal cord

Superficial nerves and vessels of the lower leg, illustrating the structural differences between veins and nerves.

1 Crural fascia (fascia cruris)
2 Cutaneous nerves
3 Superficial cutaneous veins
4 Perforating vein

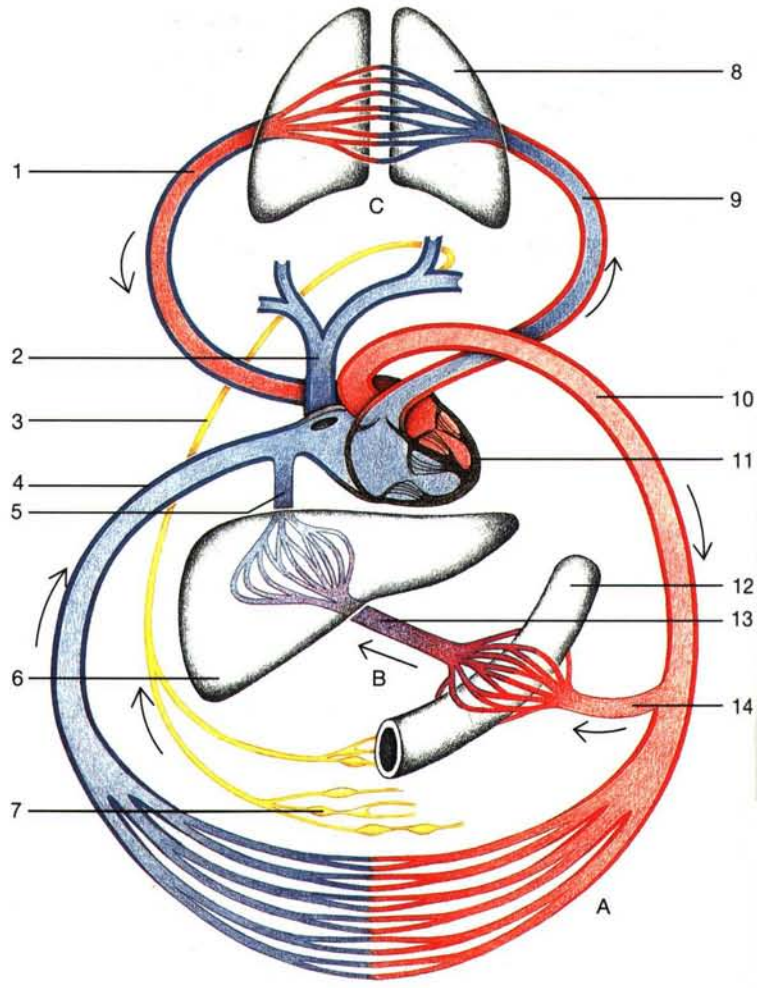

Organization of the circulatory system. Arrows: direction of the blood flow.

Vessel wall in red = Arteries
Vessel wall in blue = Veins
Yellow = Lymphatic vessels

A Systemic circulation
B Hepatic portal circulation
C Pulmonary circulation

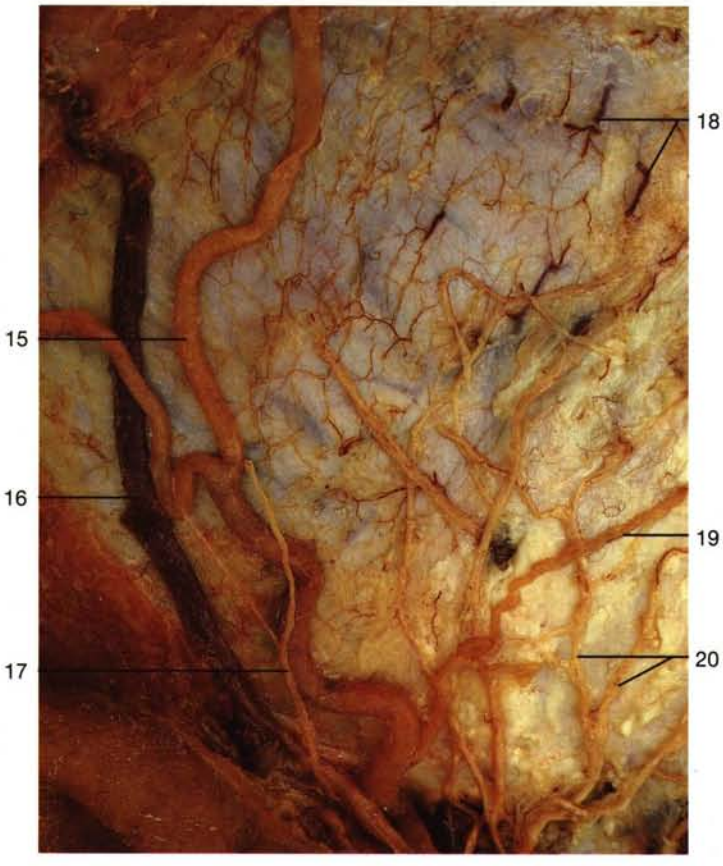

Superficial nerves and vessels. Temporal region. Note the differences between arteries, veins and nerves.

1 Pulmonary vein
2 Superior vena cava
3 Thoracic duct
4 Inferior vena cava
5 Hepatic vein
6 Liver
7 Lymph nodes and lymphatic vessels
8 Lung
9 Pulmonary artery
10 Aorta
11 Heart
12 Small intestine with capillary network
13 Portal vein
14 Mesenteric artery
15 Superficial temporal artery
16 Superficial temporal vein
17 Auriculotemporal nerve
18 Perforating veins for subcutaneous fatty tissue
19 Small artery
20 Small nerves (branches of facial nerve)

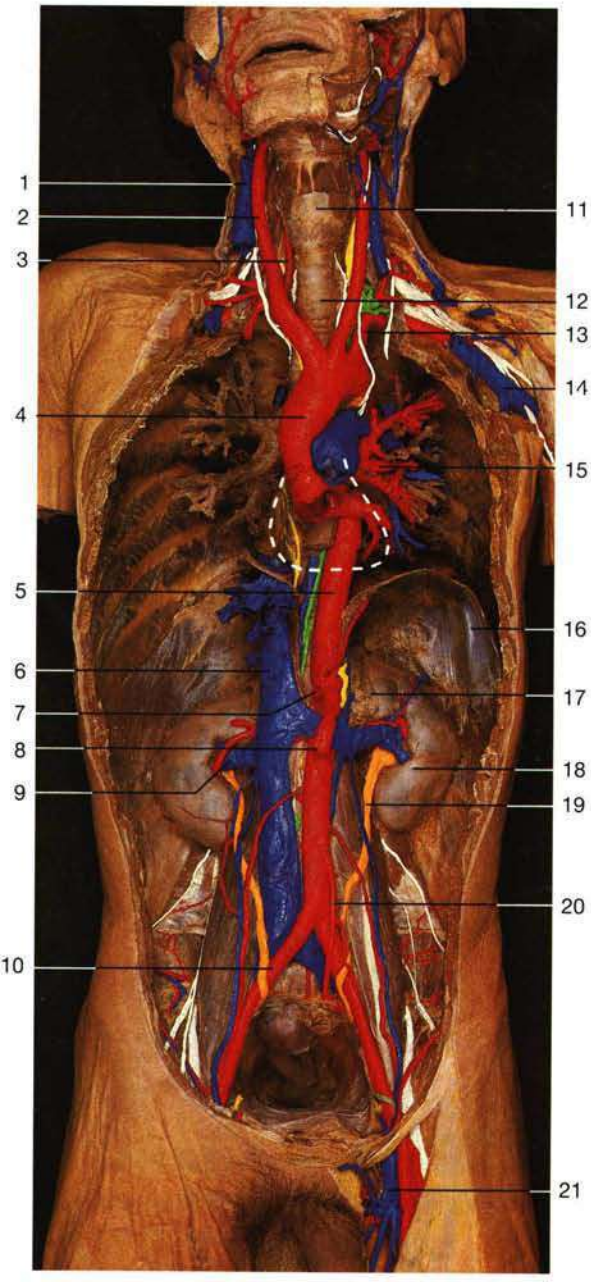

Major vessels of the trunk. The position of the heart is indicated by the dotted line.

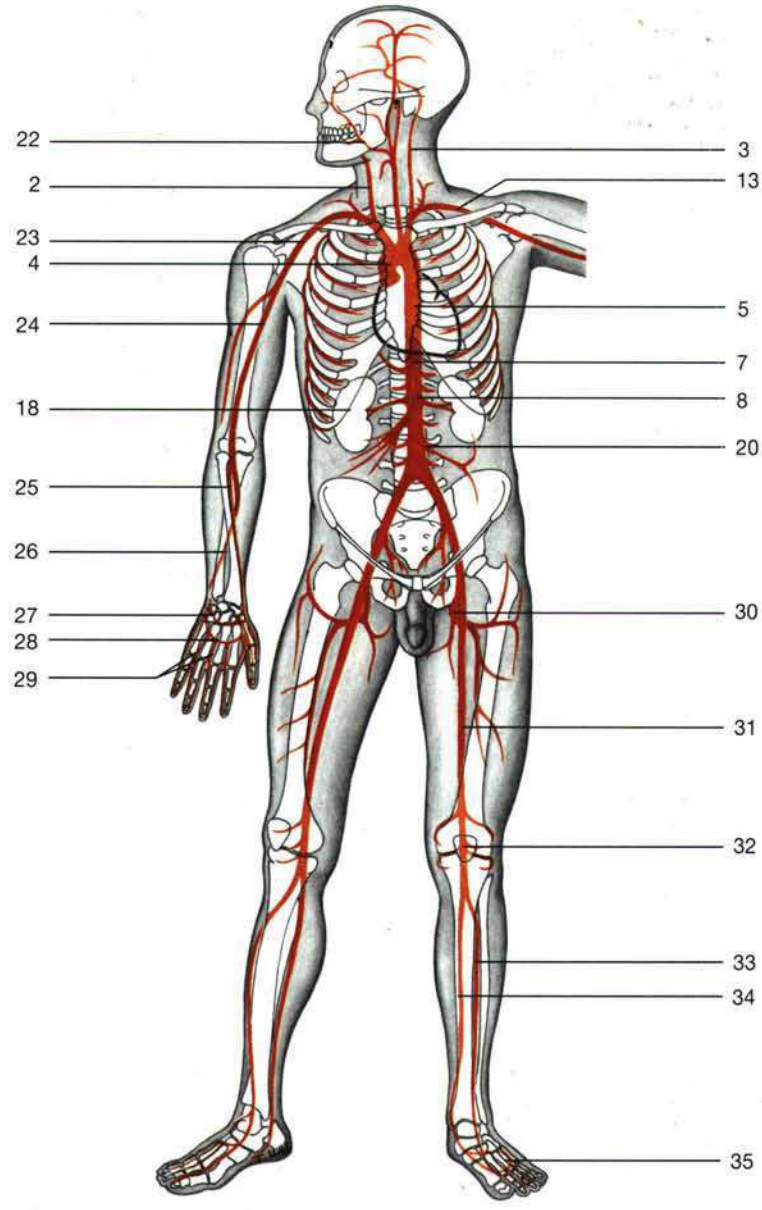

Major arteries of the human body. (Schematic drawing.)

1 Internal jugular vein
2 **Common carotid artery**
3 Vertebral artery
4 **Ascending aorta**
5 **Descending aorta**
6 **Inferior vena cava**
7 Celiac trunk
8 Superior mesenteric artery
9 Renal vein
10 **Common iliac artery**
11 Larynx
12 Trachea
13 Left subclavian artery
14 Left axillary vein
15 **Pulmonary veins**
16 Diaphragm
17 Suprarenal gland
18 Kidney
19 Ureter
20 Inferior mesenteric artery
21 **Femoral vein**
22 Facial artery
23 Axillary artery
24 Brachial artery
25 Radial artery
26 Ulnar artery
27 Deep palmar arch
28 Superficial palmar arch
29 Common palmar digital arteries
30 Profunda femoris artery
31 **Femoral artery**
32 Popliteal artery
33 Anterior tibial artery
34 Posterior tibial artery
35 Plantar arch

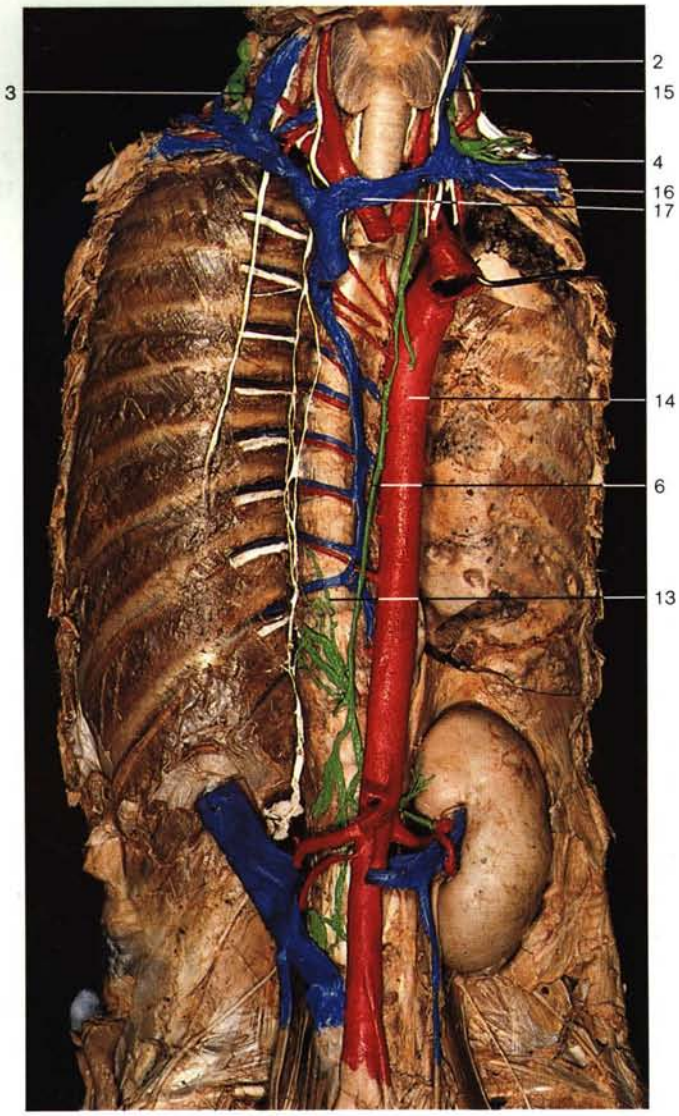

Major lymph vessels of the trunk.

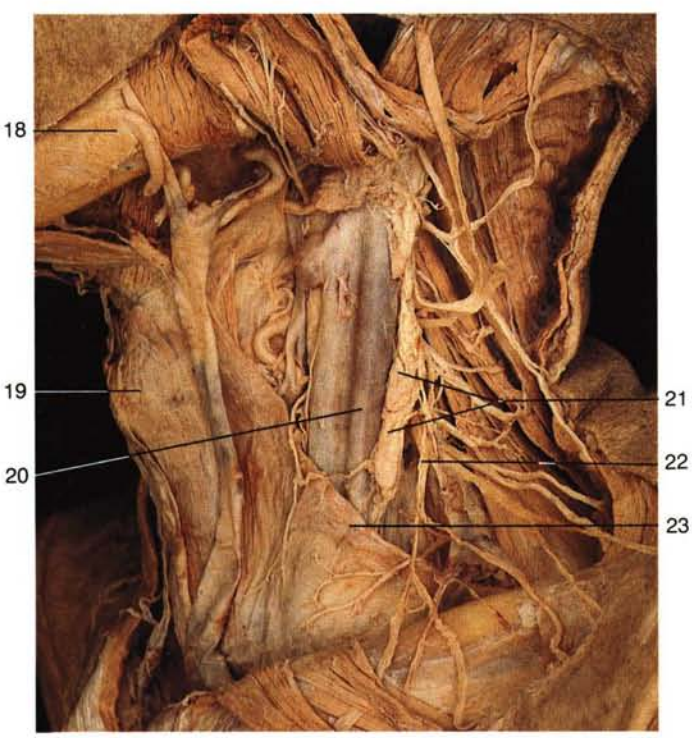

Deep cervical nodes, adjacent to the internal jugular vein.

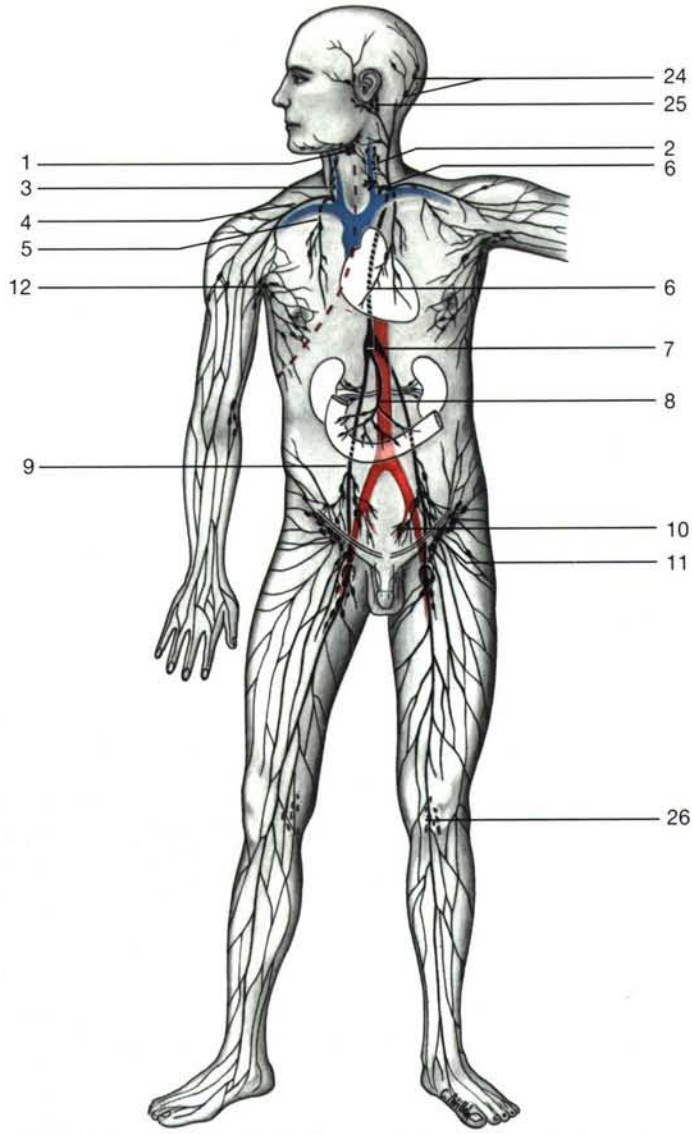

Lymphatic system. Course of the main lymphatic vessels and lymph nodes in the body. Dotted line = border between lymphatic vessels draining towards the right venous angle and towards the left.

1 Submandibular nodes
2 Deep cervical nodes
3 Right jugular trunk
4 Subclavian trunk
5 Right bronchomediastinal trunk
6 Thoracic duct
7 **Cisterna chyli**
8 **Intestinal trunk**
9 Right **lumbar trunk**
10 Internal iliac nodes
11 **Inguinal nodes**
12 **Axillary nodes**
13 Descending trunk
14 Descending aorta
15 Internal jugular vein
16 Subclavian vein
17 Left brachiocephalic vein
18 Mandible
19 Larynx
20 Internal jugular vein
21 **Deep cervical nodes**
22 Cervical nerve plexus
23 Superficial layer of deep fascia
24 Occipital nodes
25 Parotid nodes
26 Popliteal nodes

Lymphatic vessels originate as blind-ending tubes in the tissue spaces (lymph capillaries) and unite to form larger vessels (lymphatics). These resemble veins but have a much thinner wall, more valves and are interrupted by lymph nodes at various intervals. Large groups of lymph nodes are located in the inguinal and axillary regions, deep to the mandible and sternocleidomastoid muscle and within the root of the mesentery of the intestine.

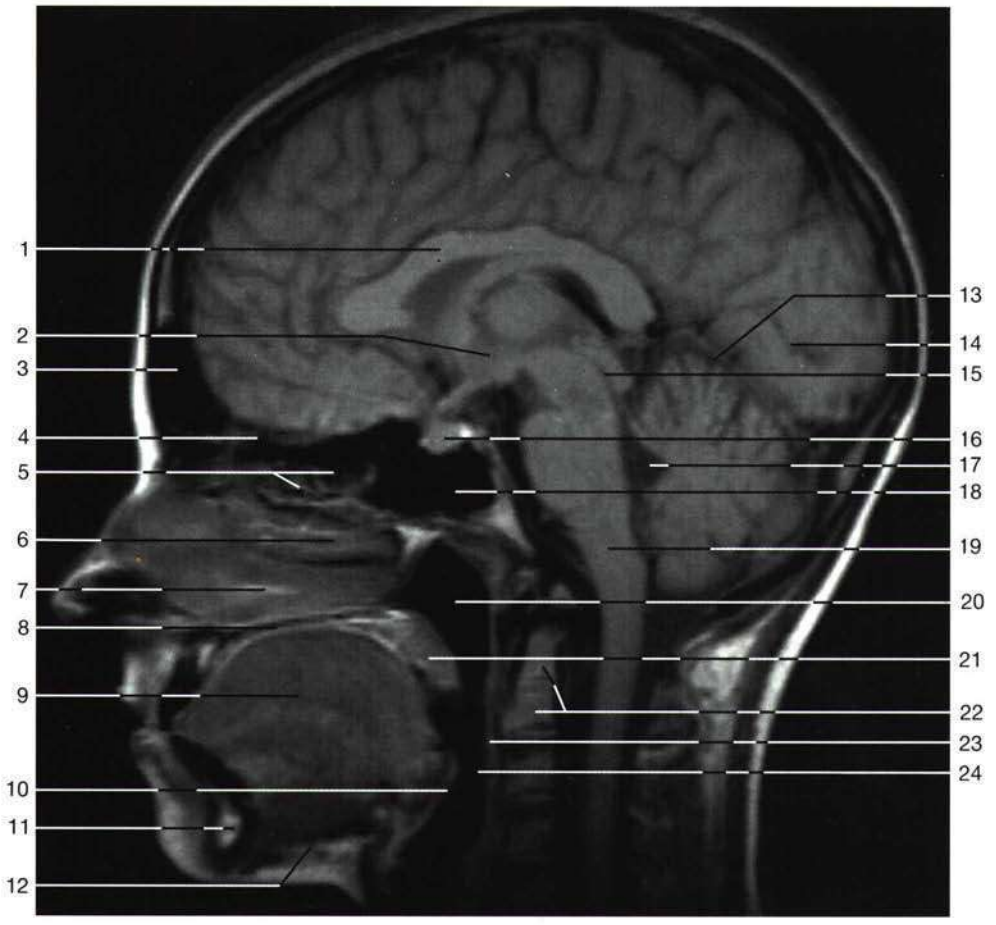

1 Corpus callosum
2 **Hypothalamus**
3 Frontal sinus
4 Cribriform plate
5 Ethmoidal air cells
6 Middle nasal concha
7 Inferior nasal concha
8 Hard palate
9 **Tongue**
10 Epiglottis
11 Mandible
12 Mylohyoid muscle
13 Tentorium of cerebellum
14 Calcarine fissure
15 Cerebral aqueduct
16 **Pituitary gland**
17 Fourth ventricle
18 **Sphenoidal sinus**
19 Medulla oblongata
20 Nasopharynx
21 Uvula
22 Dens of axis
23 Constrictor muscle of pharynx
24 Oral part of pharynx
25 **Cerebrum** (right hemisphere)
26 Calvaria
27 **Cerebellum**

Sagittal section through head and neck (MR-Scan. 23-year-old female, courtesy of PD Dr. A. Heuck, Munich).

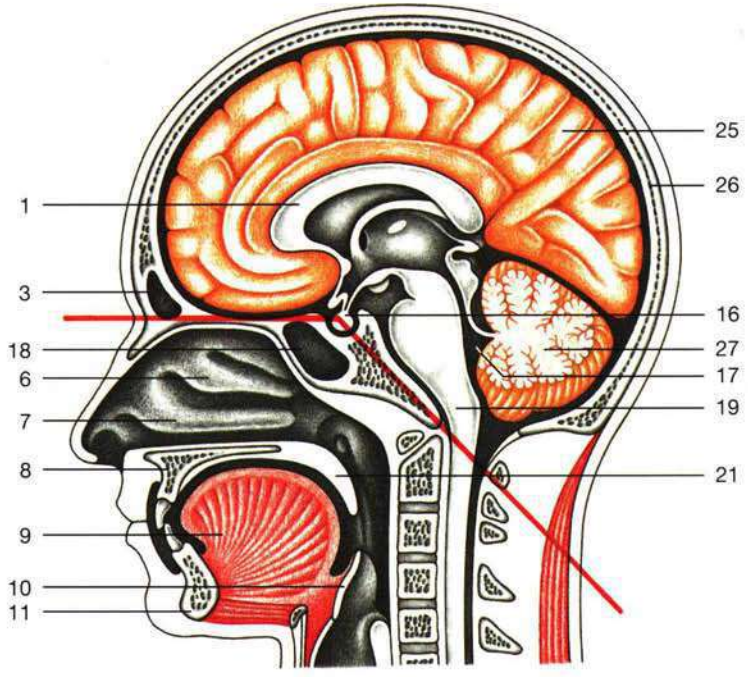

Sagittal section through the head (schematic drawing). The red line represents the border between the neurocranium and viscerocranium forming the clivus angle. The neural cavity contains the brain; the viscerocranium comprises the orbit, the nasal cavity and the oral cavity arranged one beneath the other.

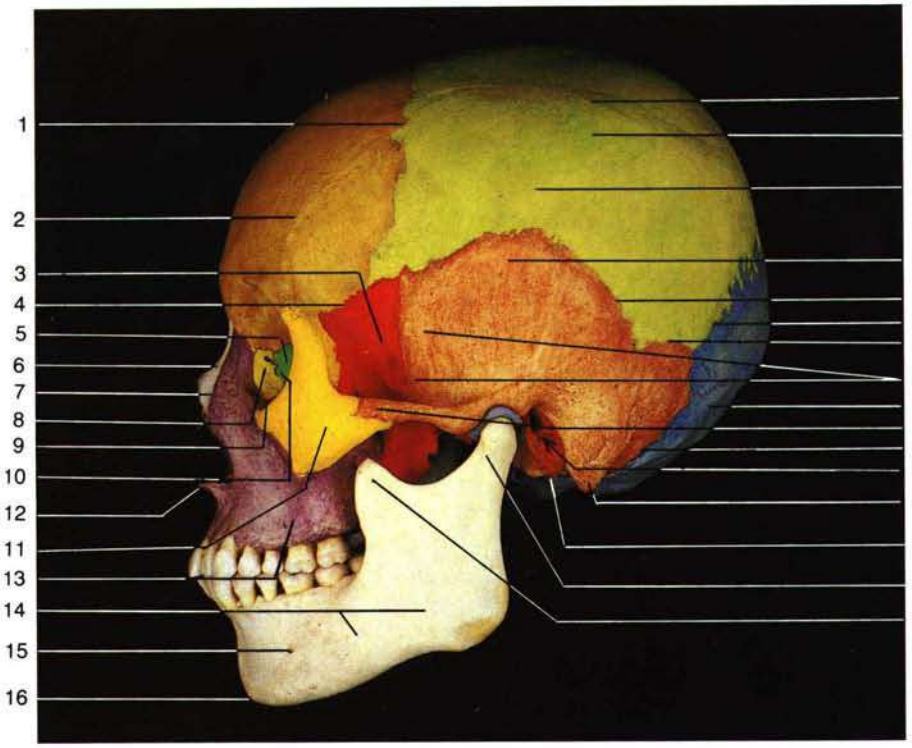

General architecture of the skull (lateral aspect). The different bones are indicated in color (numbers cf. table).

1 Coronal suture
2 **Frontal bone**
3 **Sphenoid bone**
4 Sphenofrontal suture
5 **Ethmoid bone**
6 **Nasal bone**
7 Nasomaxillary suture
8 **Lacrimal bone**
9 Lacrimomaxillary suture
10 Lacrimoethmoid suture
11 **Zygomatic bone**
12 Anterior nasal spine
13 **Maxilla**
14 **Mandible**
15 Mental foramen
16 Mental protuberance
17 Superior temporal line
18 Inferior temporal line
19 **Parietal bone**
20 **Temporal bone**
21 Squamous suture
22 Lambdoid suture
23 Temporal fossa
24 Parietomastoid suture
25 **Occipital bone**
26 Zygomatic arch
27 Occipitomastoid suture
28 External acoustic meatus
29 Mastoid process
30 Tympanic portion of temporal bone
31 Condylar process of mandible
32 Coronoid process of mandible

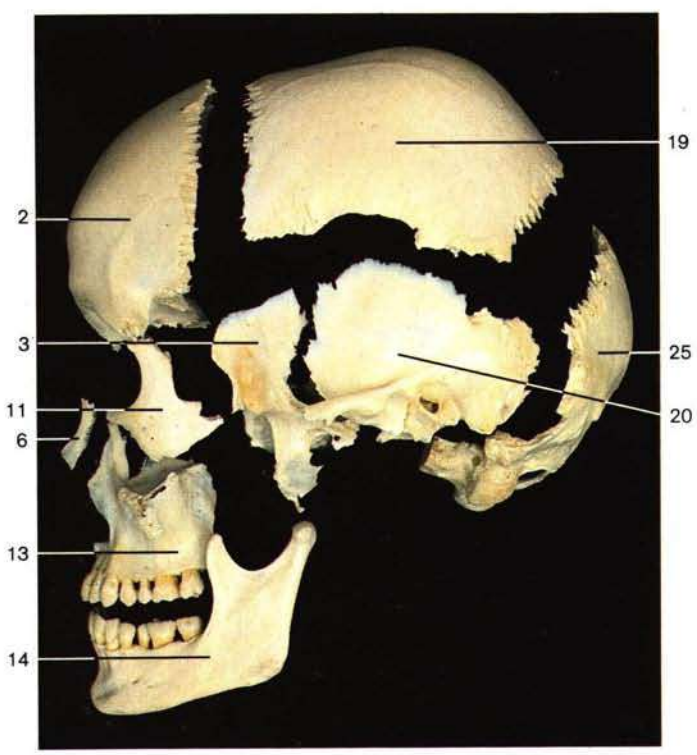

Lateral aspect of the disarticulated skull (palatine bone, lacrimal bone, ethmoid bone and vomer are not depicted).

2 Frontal bone (orange) 19 Parietal bone (light green) 3 Greater wing of sphenoid bone (red) 25 Squama of occipital bone (blue) 20 Squama of temporal bone (brown)	Cranial bones
5 Ethmoid bone (dark green) 3 Sphenoid bone (red) Temporal bone excluding squama (brown) 30 Tympanic portion of temporal bone (dark brown) Occipital bone excluding squama (blue)	Base of skull
6 Nasal bone (white) 8 Lacrimal bone (yellow) Inferior nasal concha Vomer 11 Zygomatic bone (light yellow) Palatine bone 13 Maxilla (violet) 14 Mandible (white)	Facial bones
Malleus, Incus, Stapes } within petrous portion of temporal bone	Auditory ossicles
Hyoid	

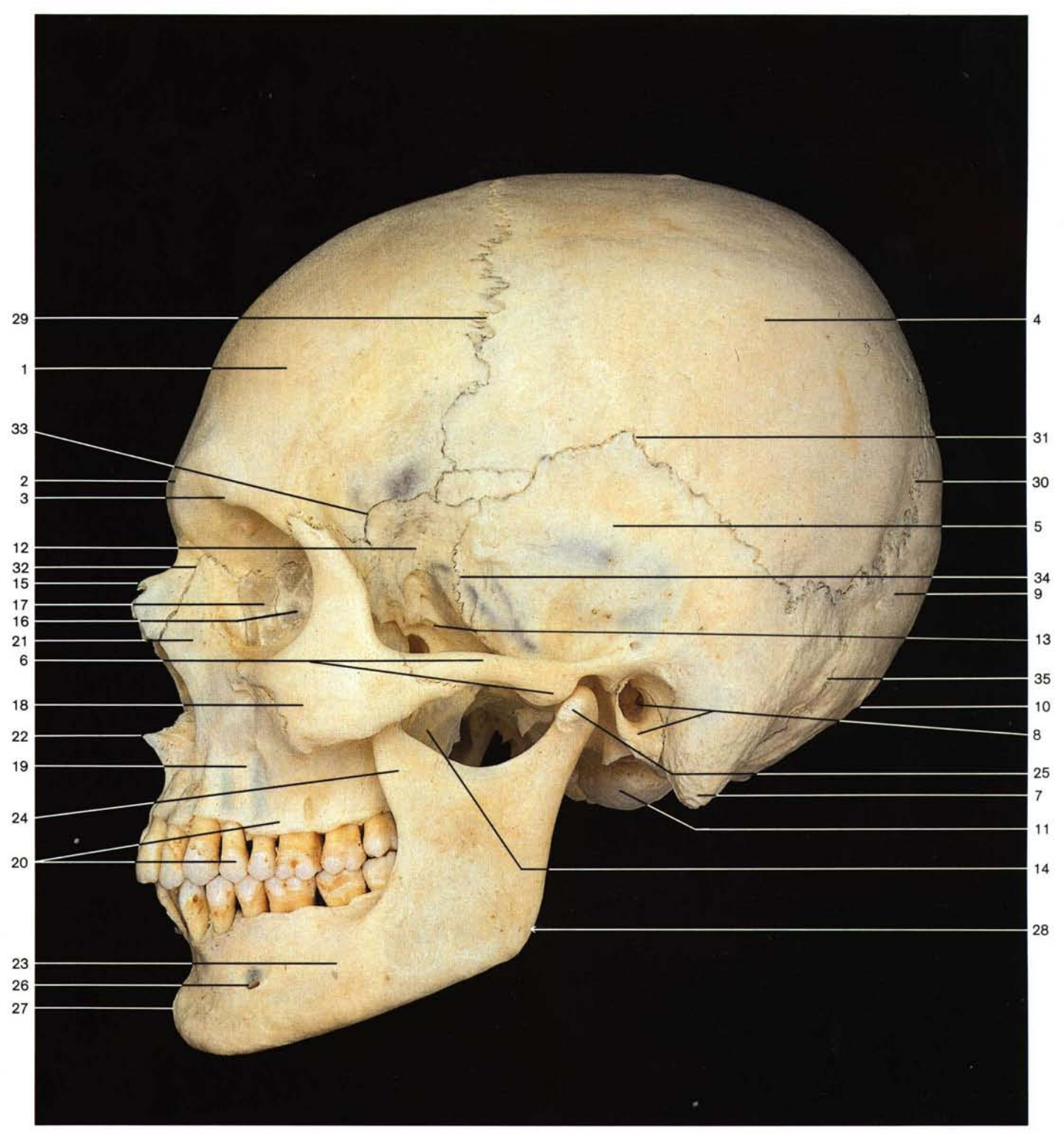

Lateral aspect of the skull.

1 **Frontal bone**
2 Glabella
3 Supraorbital margin
4 **Parietal bone**
5 **Temporal bone** (squamous part)
6 Zygomatic process (articular tubercle)
7 Mastoid process
8 Tympanic part (tympanic plate) and external acoustic meatus
9 **Occipital bone** (squamous part)
10 External occipital protuberance
11 Occipital condyle
12 **Sphenoid bone** (greater wing)
13 Infratemporal crest of sphenoid
14 Pterygoid process (lateral pterygoid plate)
15 **Nasal bone**
16 **Ethmoid bone** (orbital part)
17 **Lacrimal bone**
18 **Zygomatic bone**
19 **Maxilla** (body)
20 Alveolar process and teeth
21 Frontal process
22 Anterior nasal spine
23 **Mandible** (body)
24 Coronoid process
25 Condylar process
26 Mental foramen
27 Mental protuberance
28 Angle of the mandible

Sutures

29 Coronal suture
30 Lambdoid suture
31 Squamous suture
32 Nasomaxillary suture
33 Frontosphenoid suture
34 Sphenosquamosal suture
35 Occipitomastoid suture

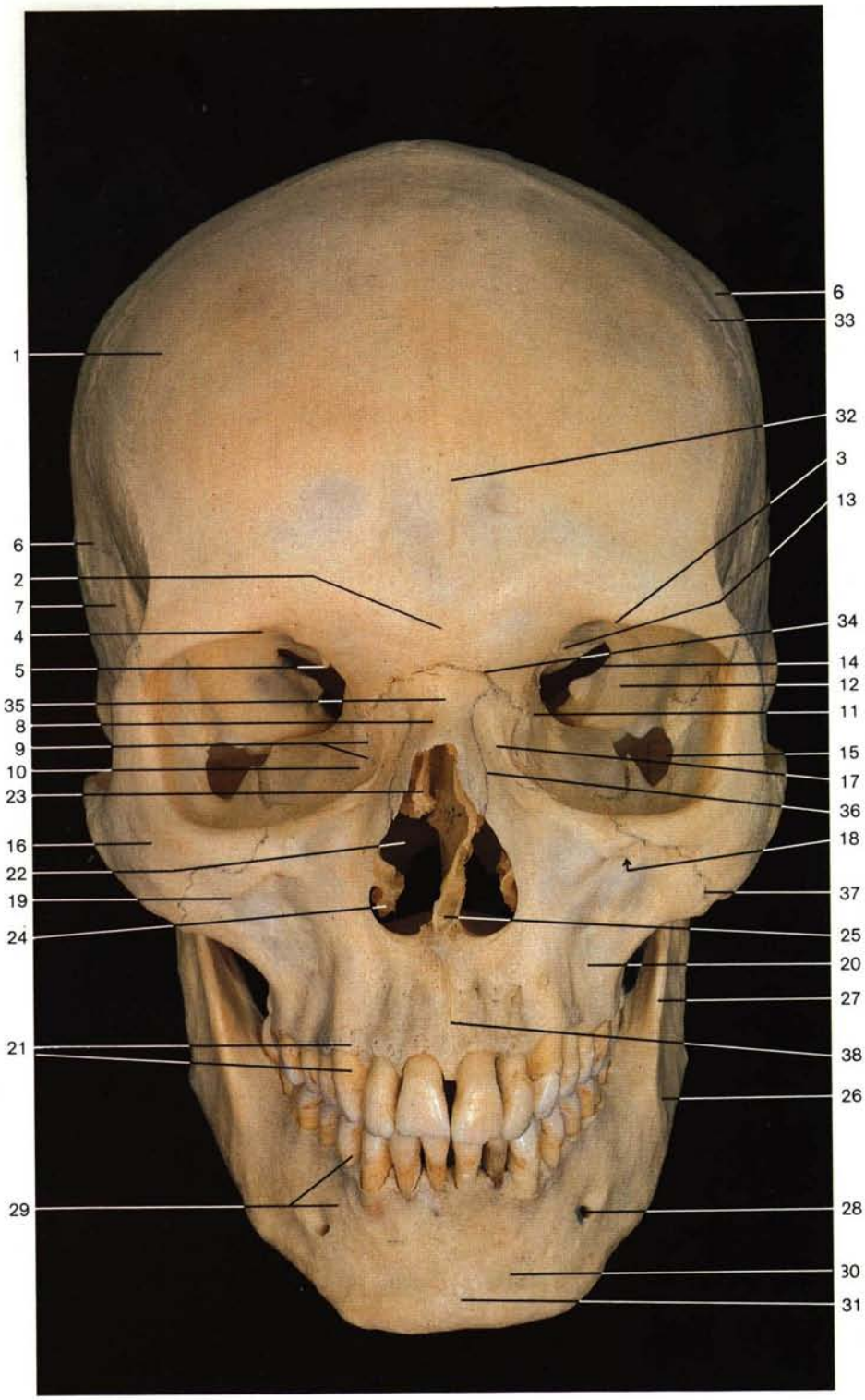

1 **Frontal bone**
2 Glabella
3 Supraorbital margin
4 Supraorbital notch
5 Trochlear spine
6 **Parietal bone**
7 **Temporal bone**
8 **Nasal bone**

Orbit
9 Lacrimal bone
10 Posterior lacrimal crest
11 Ethmoid bone

Sphenoid bone
12 Greater wing of sphenoid bone
13 Lesser wing of sphenoid bone
14 Superior orbital fissure
15 Inferior orbital fissure
16 **Zygomatic bone**

Maxilla
17 Frontal process
18 Infraorbital foramen
19 Zygomatic process
20 Body of maxilla
21 Alveolar process with teeth

Nasal cavity
22 Anterior nasal aperture
23 Middle nasal concha
24 Inferior nasal concha
25 Nasal septum, vomer

Mandible
26 Body of mandible
27 Ramus of mandible
28 Mental foramen
29 Alveolar part with teeth
30 Base of mandible
31 Mental protuberance

Sutures
32 Frontal suture
33 Coronal suture
34 Frontonasal suture
35 Internasal suture
36 Nasomaxillary suture
37 Zygomaticomaxillary suture
38 Intermaxillary suture

Anterior aspect of the skull.

The skull comprises a mosaic of numerous complicated bones which form the cranial cavity protecting the brain **(neurocranium)** and several cavities such as nasal and oral cavities in the facial region. The neurocranium consists of large bony plates which develop directly from the surrounding sheets of connective tissue **(desmocranium).** The bones of the skull base are formed out of cartilaginous tissue **(chondrocranium)** which ossifies secondarily. The **visceral skeleton** which, in fish, gives rise to the gills, has in higher vertebrates been transformed into the bones of the masticatory and auditory apparatus (maxilla, mandible, auditory ossicles and hyoid bone).

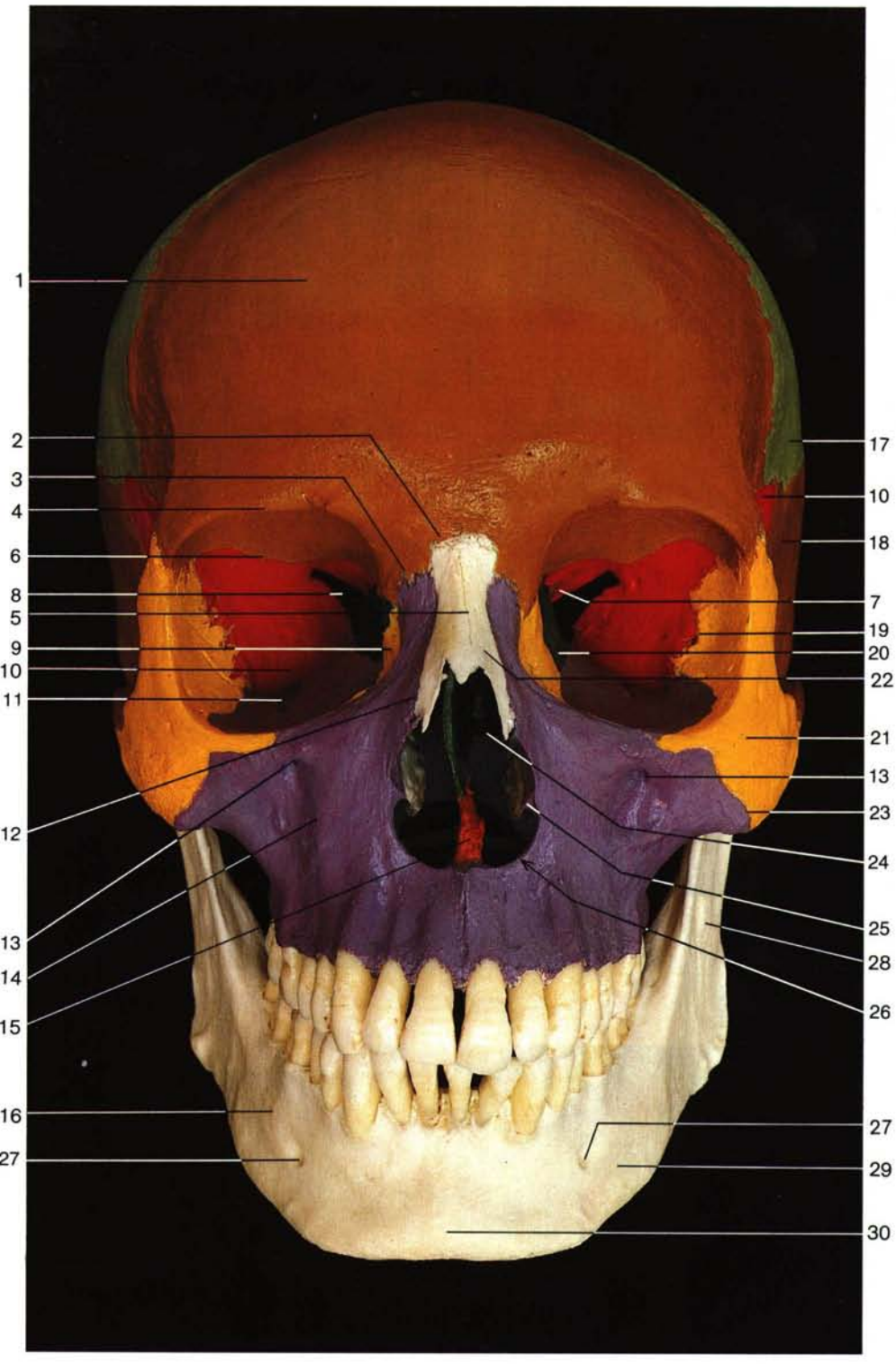

1 **Frontal bone**
2 Frontonasal suture
3 Frontomaxillary suture
4 Supraorbital margin
5 Internasal suture
6 Sphenofrontal suture
7 Optic canal in lesser wing of sphenoid bone
8 Superior orbital fissure
9 **Lacrimal bone**
10 **Sphenoid bone** (greater wing)
11 Inferior orbital fissure
12 Nasomaxillary suture
13 Infraorbital foramen
14 **Maxilla**
15 **Vomer**
16 Body of mandible
17 **Parietal bone**
18 **Temporal bone**
19 Sphenozygomatic suture
20 **Ethmoid bone**
21 **Zygomatic bone**
22 **Nasal bone**
23 Zygomaticomaxillary suture
24 Middle nasal concha
25 Inferior nasal concha
26 Anterior nasal aperture
27 Mental foramen
28 Ramus of mandible
29 Base of mandible
30 Mental protuberance

Bones
Frontal bone (brown)
Parietal bone (light green)
Temporal bone (dark brown)
Sphenoidal bone (red)
Zygomatic bone (yellow)
Ethmoid bone (dark green)
Lacrimal bone (yellow)
Vomer (orange)
Maxilla (violet)
Nasal bone (white)
Mandible (white)

Anterior aspect of the skull (individual bones indicated by color).

The following series of figures are arranged so that the mosaic-like pattern of the skull becomes understandable. It starts with the bones of the **skull base** (sphenoid and occipital) to which the other bones are added step by step. The facial skeleton is built up by the ethmoid bone to which the palatine bone and maxilla are attached laterally; the small nasal and lacrimal bones fill the remaining spaces. Cartilages remain only in the external part of the nose.

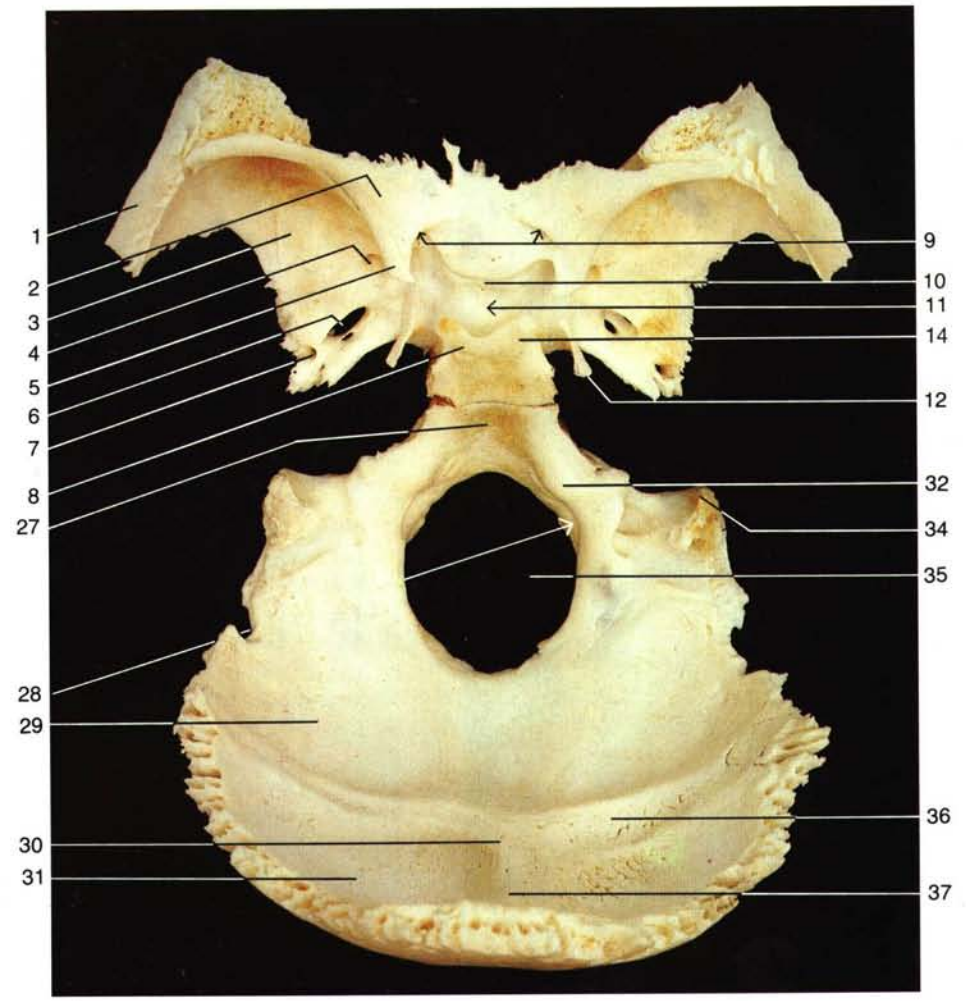

Sphenoid and occipital bone (from above).

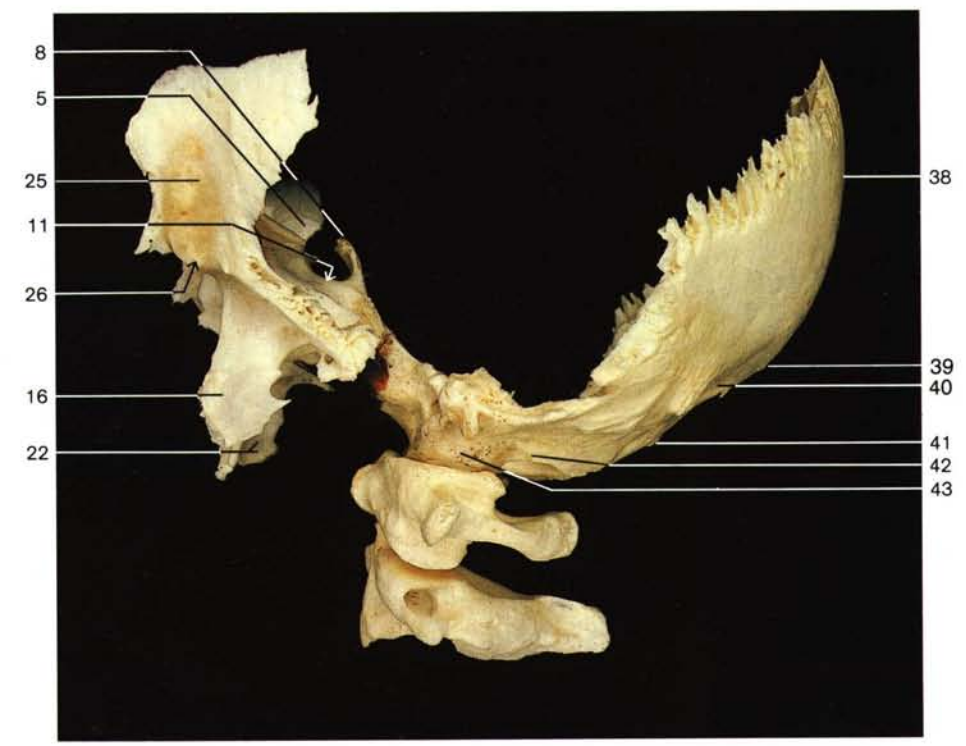

Sphenoid and occipital bone in connection with the atlas and axis (1st and 2nd cervical vertebrae) (left lateral view).

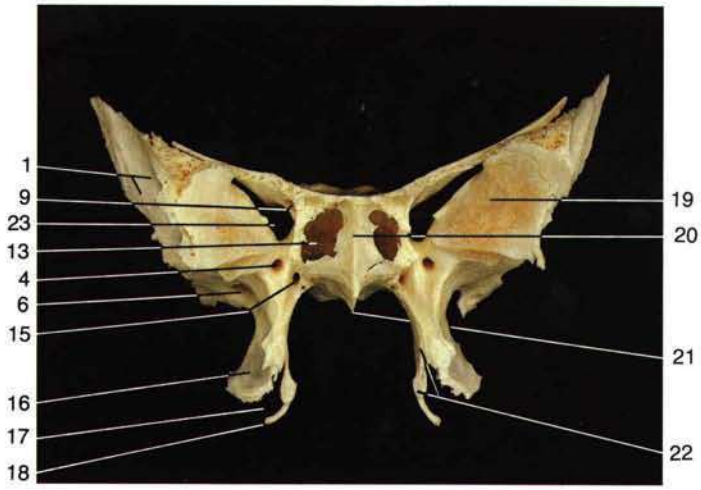

Sphenoid bone (anterior aspect).

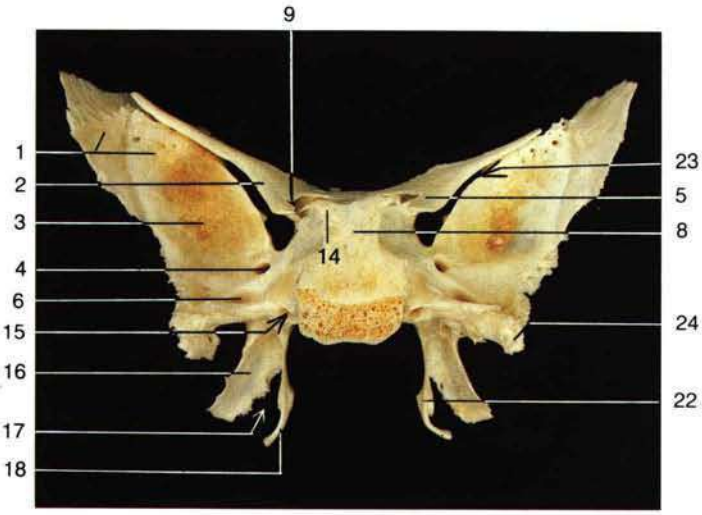

Sphenoid bone (posterior aspect).

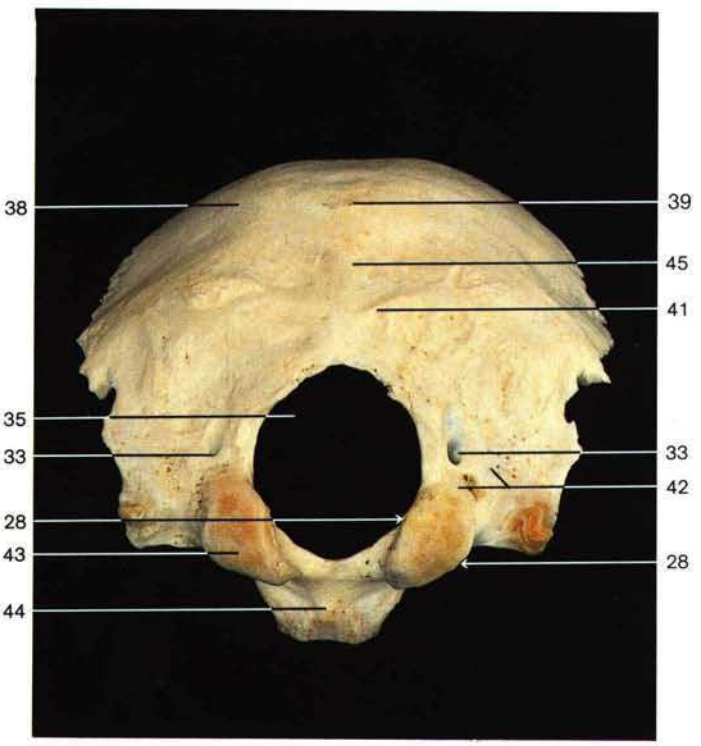

Occipital bone (from below).

Sphenoid bone

1 **Greater wing**
2 **Lesser wing**
3 Cerebral or superior surface of greater wing
4 **Foramen rotundum**
5 Anterior clinoid process
6 Foramen ovale
7 Foramen spinosum
8 Dorsum sellae
9 Optic canal
10 Chiasmatic groove (sulcus chiasmatis)
11 **Hypophysial fossa (sella turcica)**
12 Lingula
13 Opening of sphenoidal sinus
14 Posterior clinoid process
15 **Pterygoid canal**
16 Lateral pterygoid plate of **pterygoid process**
17 Pterygoid notch
18 Pterygoid hamulus
19 Orbital surface of greater wing
20 Sphenoid crest
21 Sphenoid rostrum
22 Medial pterygoid plate
23 Superior orbital fissure
24 Spine of sphenoid
25 Temporal surface of greater wing
26 Infratemporal crest

Occipital bone

27 **Clivus** with basilar part of occipital bone
28 **Hypoglossal canal**
29 Fossa for cerebellar hemisphere
30 Internal occipital protuberance
31 Fossa for cerebral hemisphere
32 Jugular tubercle
33 Condylar canal
34 Jugular process
35 **Foramen magnum**
36 Groove for transverse sinus
37 Groove for superior sagittal sinus
38 **Squamous part** of the occipital bone
39 External occipital protuberance
40 Superior nuchal line
41 Inferior nuchal line
42 Condylar fossa
43 **Condyle**
44 Pharyngeal tubercle
45 External occipital crest

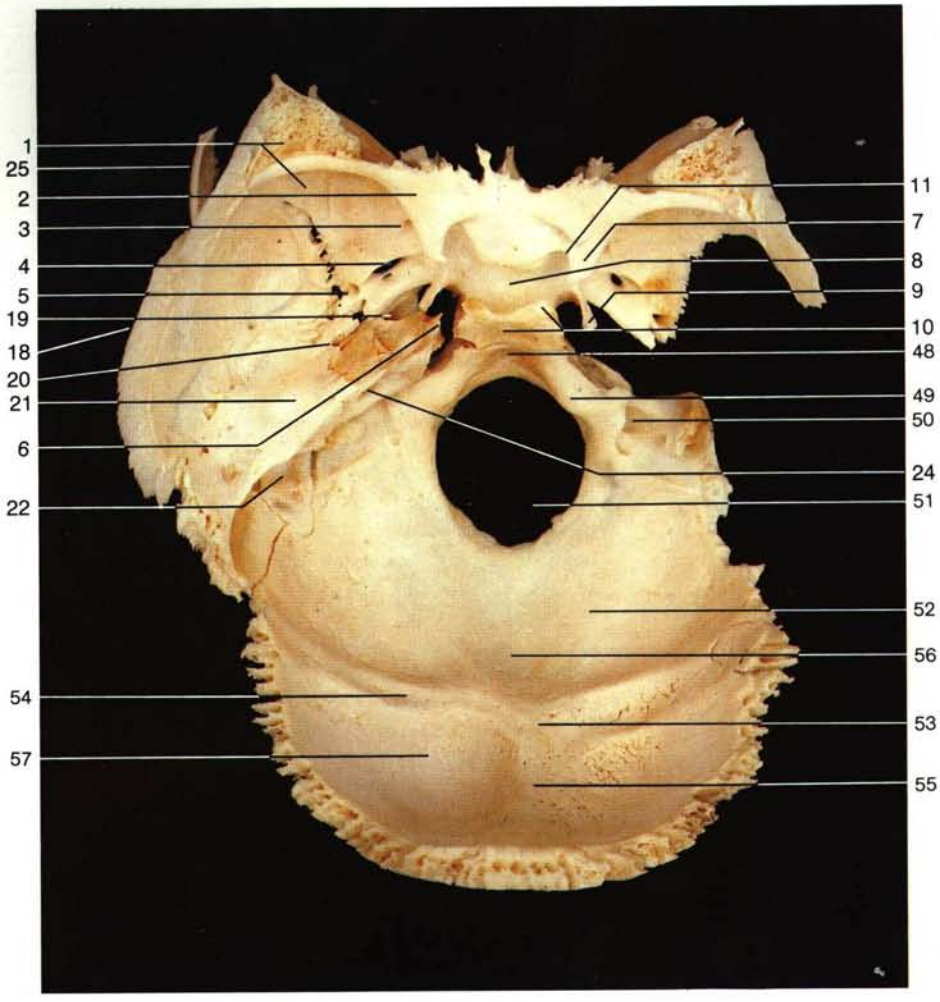

Sphenoid, occipital and left temporal bone (from above). Internal aspect of the base of the skull. The left temporal bone has been added to the preceeding figure.

Sphenoid bone

1 **Greater wing**
2 **Lesser wing**
3 Foramen rotundum
4 **Foramen ovale**
5 Foramen spinosum
6 Foramen lacerum
7 Anterior clinoid process
8 Hypophysial fossa (**sella turcica**)
9 Lingula
10 Dorsum sellae and posterior clinoid process
11 **Optic canal**
12 Sphenoid rostrum
13 Medial pterygoid plate
14 Lateral pterygoid plate
15 Pterygoid hamulus
16 Infratemporal crest
17 Body of the sphenoid

Temporal bone

18 **Squamous part**
19 **Carotid canal**
20 Hiatus of facial canal (for the greater petrosal nerve)
21 Arcuate eminence
22 Groove for the sigmoid sinus
23 Mastoid foramen
24 **Internal acoustic meatus**
25 Zygomatic process
26 Mandibular fossa
27 Petrotympanic fissure
28 Canalis musculotubarius (bony part of auditory tube)
29 **External acoustic meatus**
30 Styloid process (remnant only)
31 Stylomastoid foramen
32 Mastoid canaliculus
33 Jugular fossa
34 **Mastoid process**
35 Mastoid notch

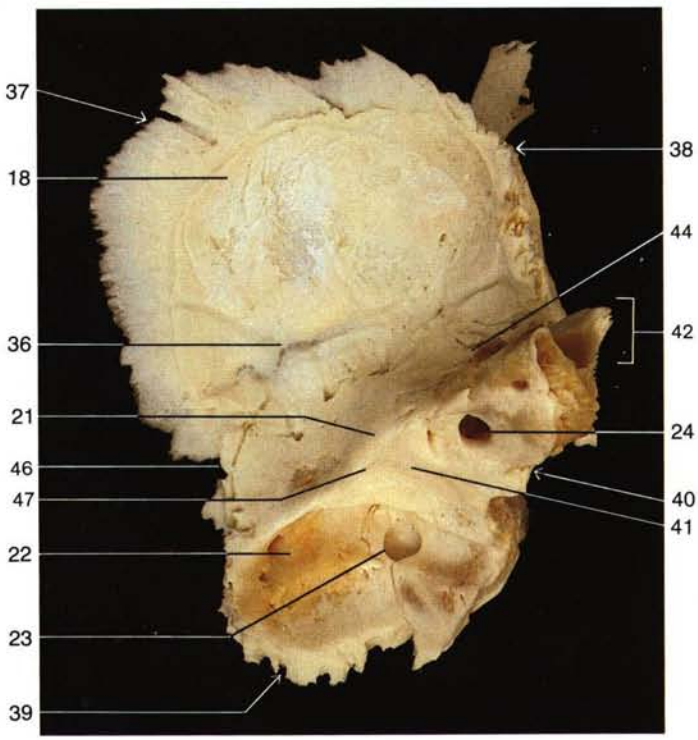

Left temporal bone (medial aspect).

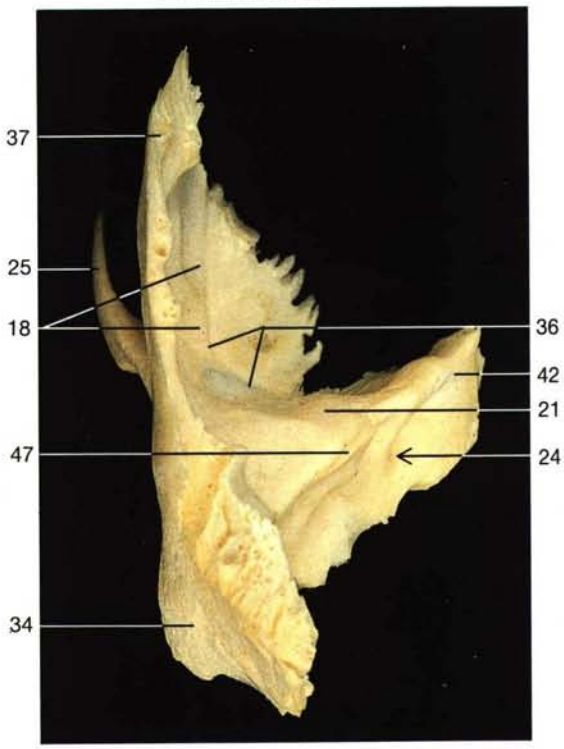

Left temporal bone (from above).

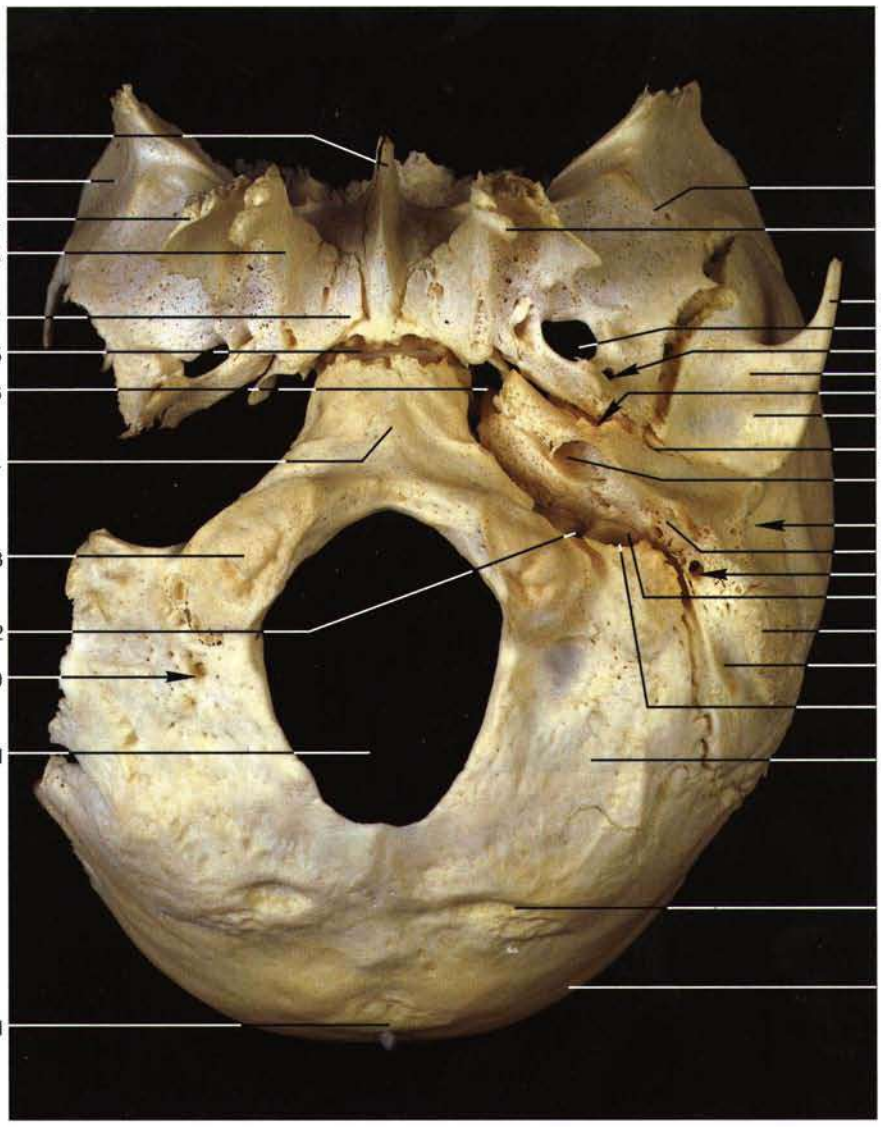

Sphenoid, occipital and left temporal bone.
Base of the skull (external aspect).

36 Groove for middle meningeal vessels
37 Parietal margin
38 Sphenoid margin
39 Occipital margin
40 Cochlear canaliculus
41 Aqueduct of the vestibule
42 Apex of the petrous part
43 **Tympanic part**
44 Trigeminal impression
45 **Articular tubercle**
46 Parietal notch
47 Groove for the superior petrosal sinus

Occipital bone
48 **Clivus**
49 Jugular tubercle
50 Condylar canal
51 **Foramen magnum**
52 Lower part of squamous occipital bone (cerebellar fossa)
53 Internal occipital protuberance
54 Groove for the transverse sinus
55 Groove for the superior sagittal sinus
56 Internal occipital crest
57 Upper part of squamous occipital bone (cerebral fossa)
58 **Condyle**
59 Nuchal plane
60 Superior nuchal line
61 **External occipital protuberance**
62 **Jugular foramen**
63 Inferior nuchal line
64 Pharyngeal tubercle
65 **Sphenooccipital synchondrosis**

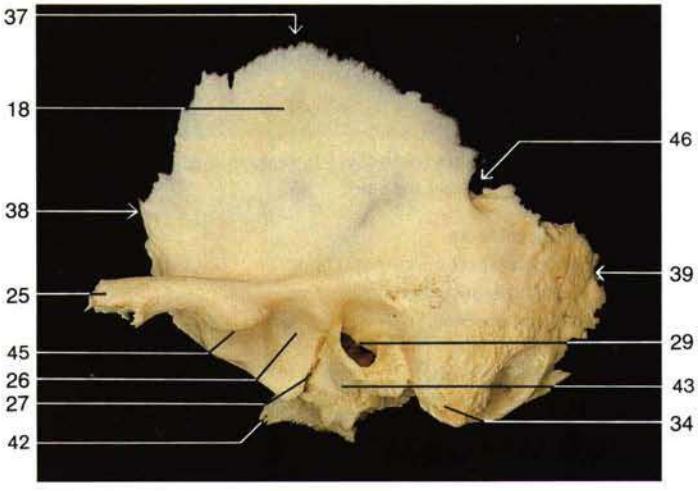

Left temporal bone (lateral aspect).

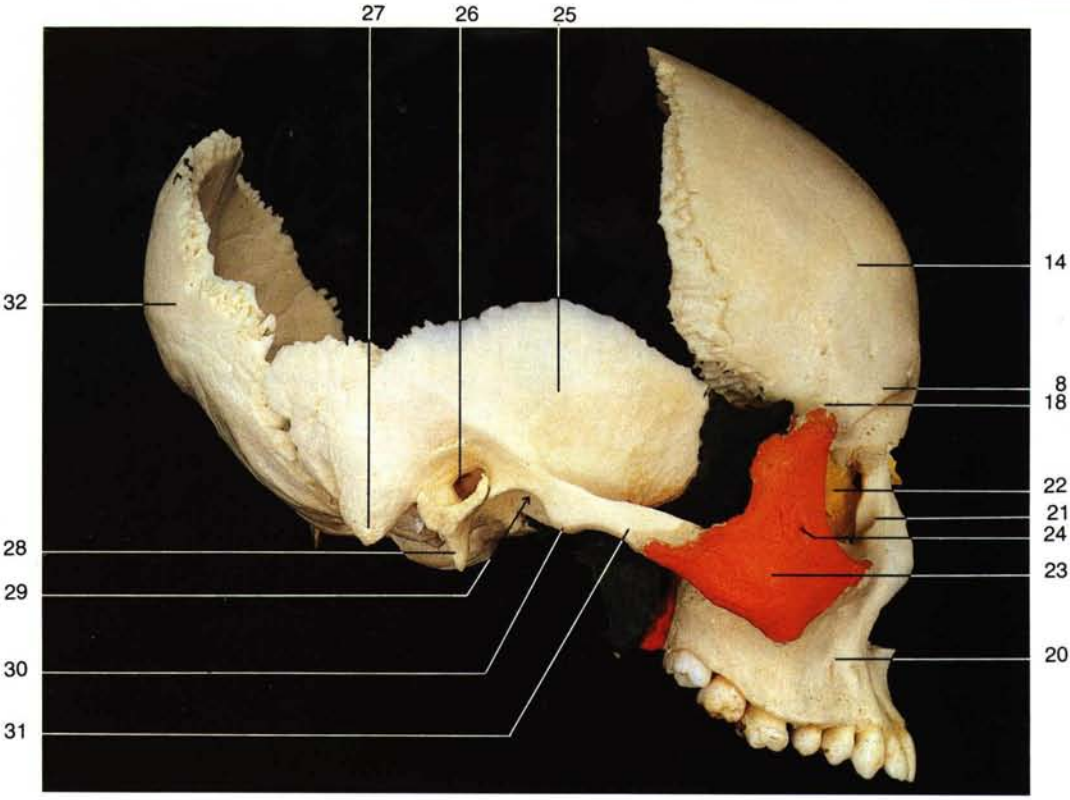

Part of a disarticulated skull (right lateral aspect). The **frontal bone** and the maxilla are connected with the temporal bone by the zygomatic bone (orange). Sphenoid bone (green), palatine bone (red), lacrimal bone (yellow).

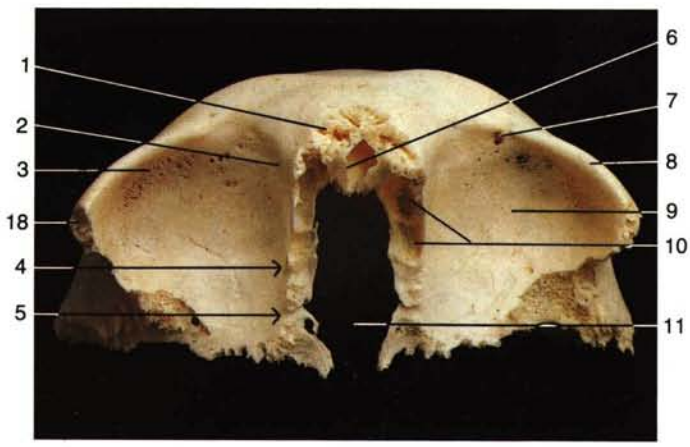

Frontal bone (inferior aspect). The ethmoidal foveolae cover the ethmoidal cavities of the ethmoid bone.

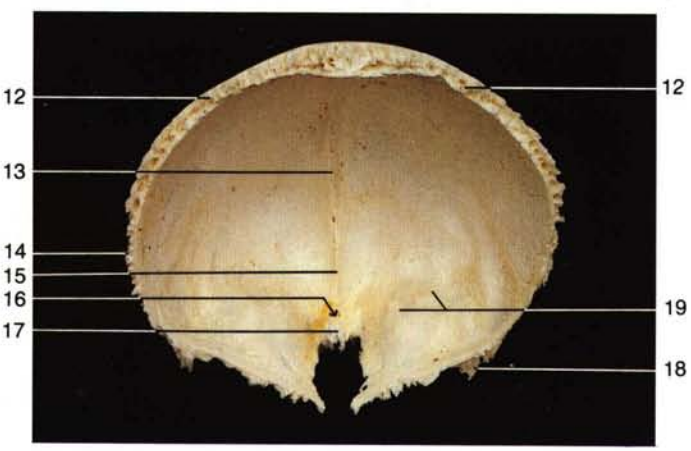

Frontal bone (posterior aspect).

Frontal bone

1 Nasal margin
2 Trochlear fossa
3 Fossa for lacrimal gland
4 Anterior ethmoidal foramen
5 Posterior ethmoidal foramen
6 Nasal spine
7 Supraorbital notch
8 Supraorbital margin
9 **Orbital plate**
10 Roofs of the ethmoidal air cells
11 Ethmoidal notch
12 Parietal margin
13 Groove for superior sagittal sinus
14 **Squamous part** of frontal bone
15 Frontal crest
16 Foramen cecum
17 Nasal spine
18 **Zygomatic process** of frontal bone
19 Juga cerebralia

20 **Maxilla**
21 Frontal process of maxilla
22 **Lacrimal bone** (yellow)
23 **Zygomatic bone** (orange)
24 Zygomaticofacial foramen

Temporal bone

25 Squamous part of temporal bone
26 External acoustic meatus
27 **Mastoid process**
28 **Styloid process**
29 **Mandibular fossa**
30 Articular tubercle
31 Zygomatic process

Occipital bone

32 Squamous part of occipital bone

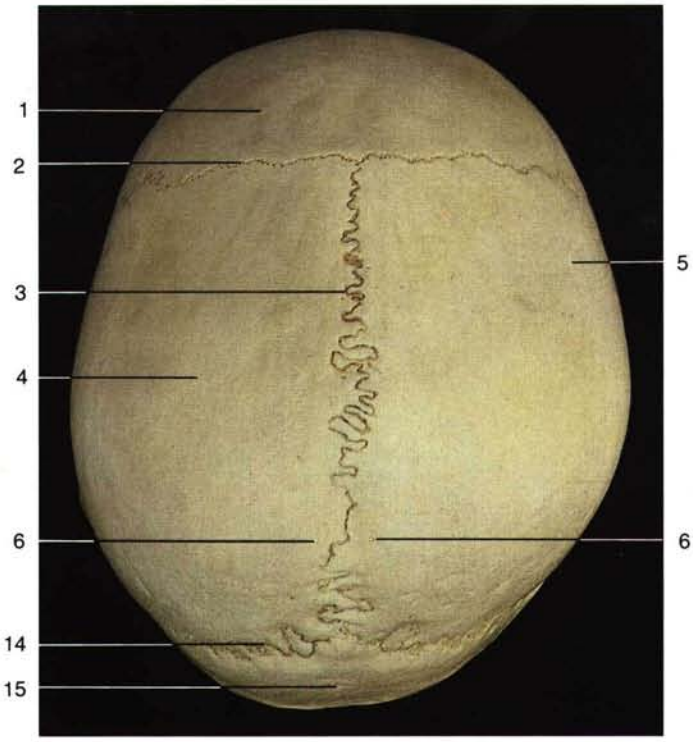

Calvaria (superior aspect).

Calvaria (posterior aspect).

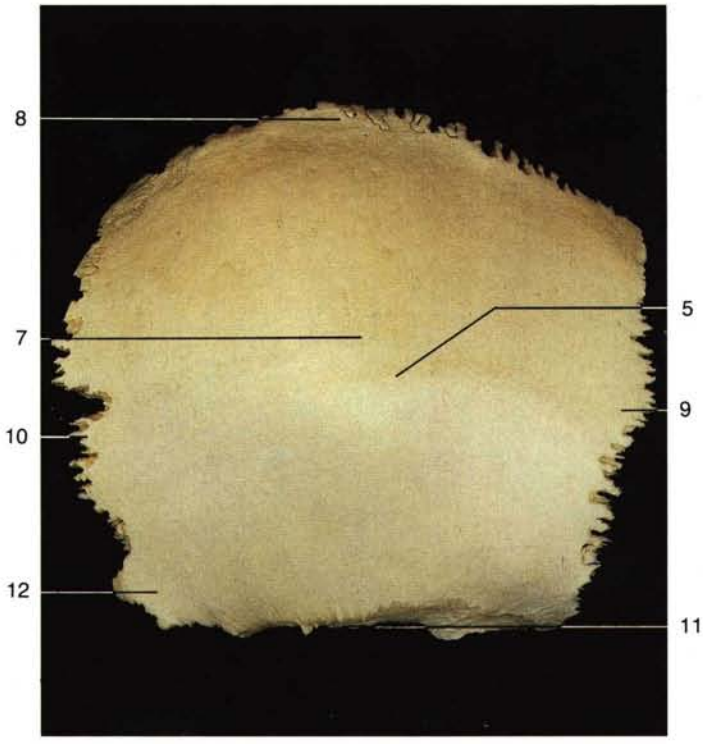

Left parietal bone (external aspect).

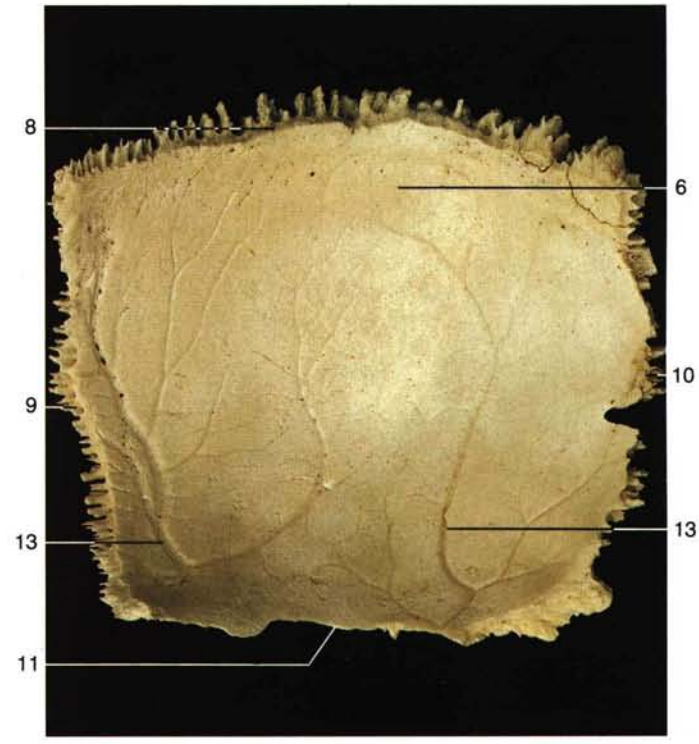

Left parietal bone (internal aspect).

1 **Frontal bone**
2 Coronal suture
3 Sagittal suture
4 **Parietal bone**
5 Superior temporal line
6 Parietal foramen
7 Parietal tuber or eminence
8 Sagittal margin
9 Occipital margin
10 Frontal margin
11 Squamous margin
12 Sphenoidal angle
13 Groove for middle meningeal artery
14 Lambdoid suture
15 **Occipital bone**
16 External occipital protuberance
17 Inferior nuchal line
18 Occipitomastoid suture
19 **Temporal bone**
20 Mastoid process
21 Mastoid notch

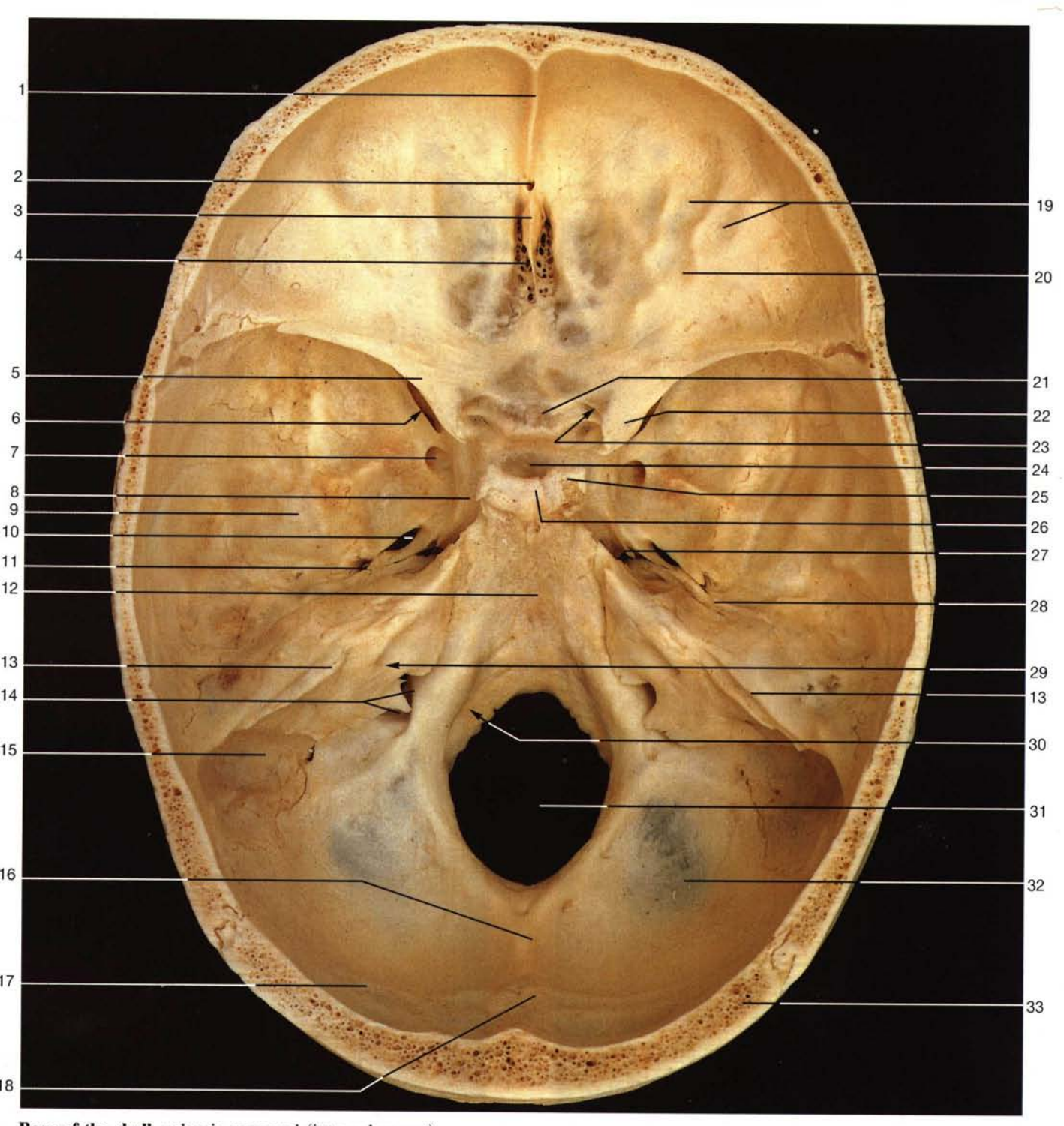

Base of the skull, calvaria removed (internal aspect).

1 Frontal crest
2 Foramen cecum
3 **Crista galli**
4 **Cribriform plate** of ethmoid bone
5 Lesser wing of sphenoid bone
6 Superior orbital fissure
7 **Foramen rotundum**
8 **Carotid sulcus**
9 Middle cranial fossa
10 **Foramen ovale**
11 Foramen spinosum
12 **Clivus**
13 Groove for superior petrosal sinus
14 **Jugular foramen**
15 Groove for sigmoid sinus
16 Internal occipital crest
17 Groove for transverse sinus
18 Internal occipital protuberance
19 Marking of brain convolutions
20 Anterior cranial fossa
21 Chiasmatic sulcus
22 Anterior clinoid process
23 **Optic canal**
24 **Sella turcica** (hypophysial fossa)
25 Posterior clinoid process
26 Dorsum sellae
27 **Foramen lacerum**
28 Groove for greater petrosal nerve
29 Internal acoustic meatus
30 **Hypoglossal canal**
31 **Foramen magnum**
32 Posterior cranial fossa
33 Diploe

Base of the skull (internal aspect, oblique lateral view from left side).

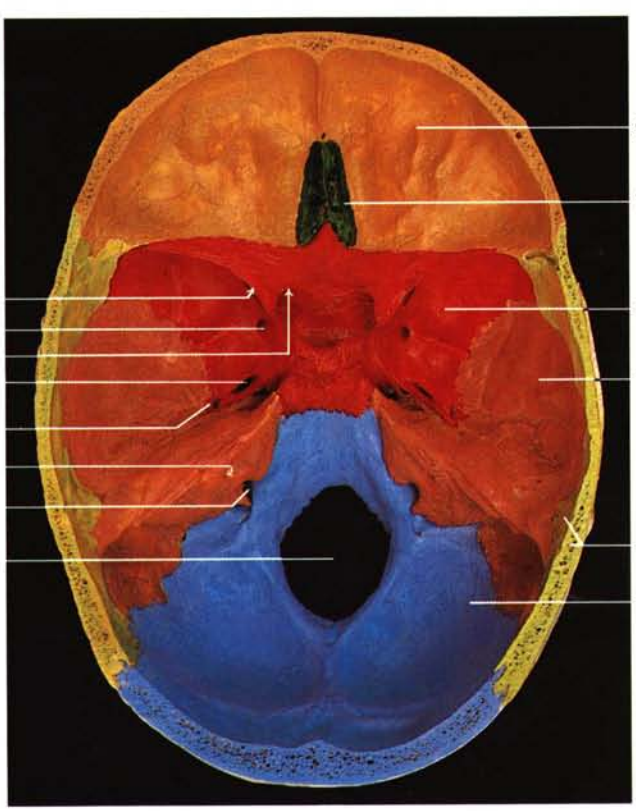

Base of the skull (internal aspect, superior view). Individual bones indicated by color.

Canals, fissures and foramina of the base of the skull

1 Superior orbital fissure
2 Foramen rotundum
3 Optic canal
4 Foramen ovale
5 Foramen spinosum
6 Internal acoustic meatus
7 Jugular foramen
8 Foramen magnum

Bones

9 Frontal bone (orange)
10 Ethmoid bone (dark green)
11 Sphenoid bone (red)
12 Temporal bone (brown)
13 Parietal bone (yellow-green)
14 Occipital bone (blue)

Details of bones

15 Crista galli
16 Cribriform plate
17 Marking of brain convolutions (frontal bone)
18 Lesser wing of sphenoid bone
19 Foramen lacerum
20 Hypophysial fossa (sella turcica)
21 Anterior clinoid process
22 Depression for trigeminal ganglion
23 Petrous part of temporal bone
24 Groove for sigmoid sinus
25 Dorsum sellae (posterior clinoid process)
26 Greater wing of sphenoid bone, groove for middle meningeal artery
27 Hypoglossal canal

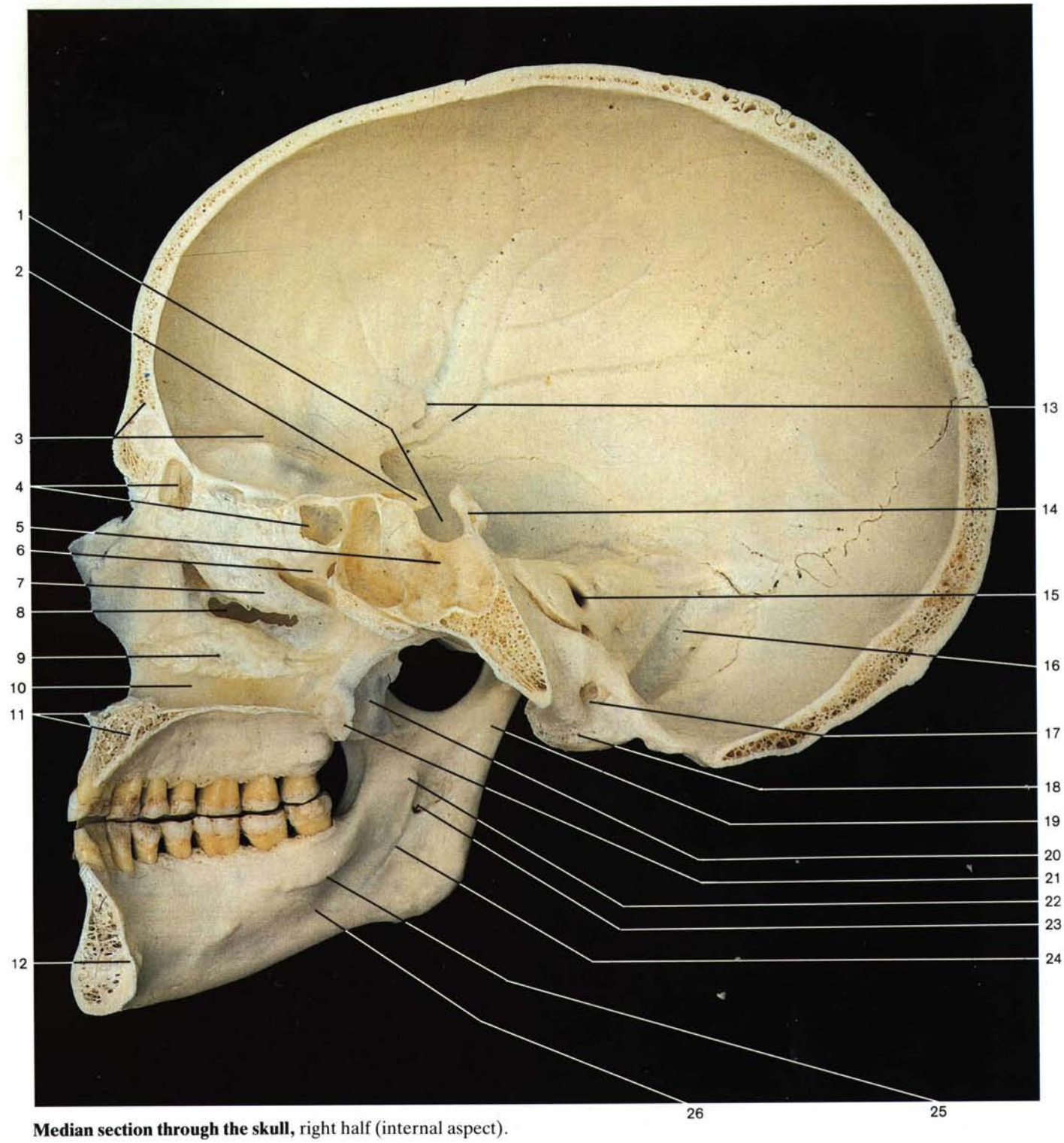

Median section through the skull, right half (internal aspect).

1 **Hypophysial fossa (sella turcica)**
2 Anterior clinoid process
3 **Frontal bone**
4 **Ethmoidal air cells**
5 **Sphenoidal sinus**
6 Superior concha
7 Middle concha
8 **Maxillary hiatus**
9 **Inferior concha**
10 Inferior meatus
11 Anterior nasal spine and maxilla
12 Mental spine or genial tubercle
13 Groove for middle meningeal artery
14 Dorsum sellae
15 Internal acoustic meatus
16 Groove for sigmoid sinus
17 **Hypoglossal canal**
18 **Occipital condyle**
19 Condylar process
20 Lateral pterygoid plate } of **pterygoid process**
21 Medial pterygoid plate } of **pterygoid process**
22 Lingula of mandible
23 **Mandibular foramen**
24 Mylohyoid groove
25 Mylohyoid line
26 Submandibular fovea

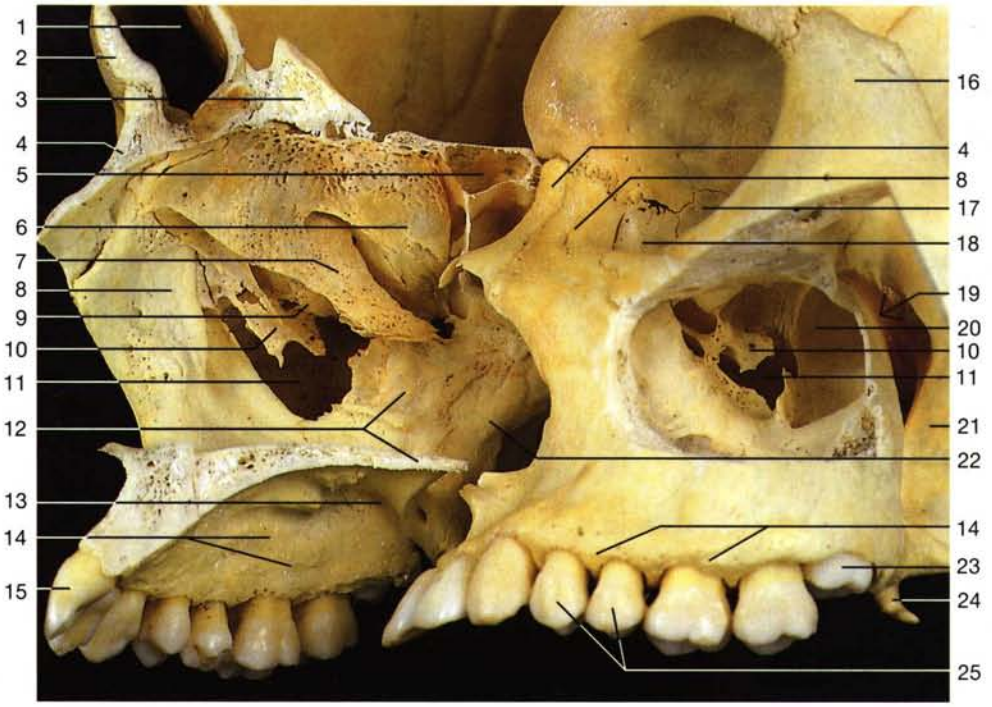

1 Frontal sinus
2 **Frontal bone**
3 Crista galli
4 **Nasal bone**
5 Sphenoidal sinus
6 **Superior concha** } of **ethmoid**
7 **Middle concha** } **bone**
8 Frontal process of maxilla
9 Ethmoidal bulla
10 **Uncinate process**
11 Maxillary hiatus
12 **Palatine bone**
13 Greater palatine foramen
14 **Alveolar process of maxilla**
15 Central incisor
16 Zygomatic bone
17 Ethmoid bone
18 **Lacrimal bone**
19 Pterygopalatine fossa
20 **Maxillary sinus**
21 Lateral pterygoid plate
22 Medial pterygoid plate
23 Third molar tooth
24 Pterygoid hamulus
25 Two premolar teeth

Facial part of the skull (viscerocranium), divided in two halves (lateral and medial aspect). Right inferior concha has been removed to show the maxillary hiatus. Left maxillary sinus opened.

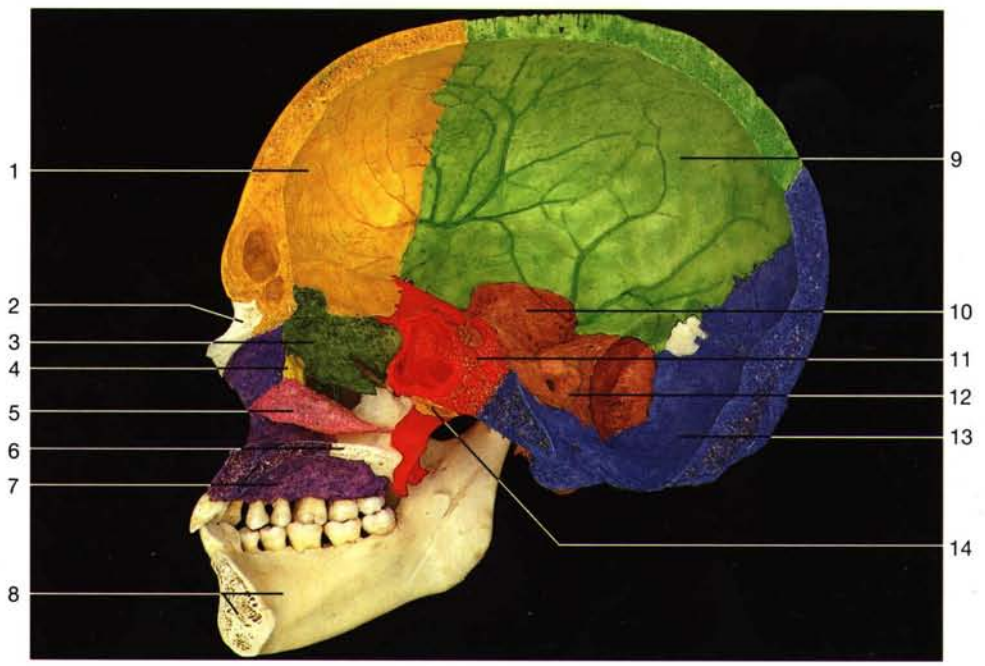

Bones (indicated by colors)
1 Frontal bone (yellow)
2 Nasal bone (white)
3 Ethmoid bone (dark green)
4 Lacrimal bone (yellow)
5 Inferior nasal concha (pink)
6 Palatine bone (white)
7 Maxilla (violet)
8 Mandible (white)
9 Parietal bone (light green)
10 Temporal bone (brown)
11 Sphenoid bone (red)
12 Petrous part of temporal bone (brown)
13 Occipital bone (blue)
14 Ala of vomer (light brown)

Median section through the skull. The nasal septum has been removed. Bones indicated by colors.

Because of the upright posture which the man developed in the course of evolution, the cranial cavity greatly increased in size whereas the facial skeleton decreased. As a result, the base of the skull developed an angulation of about 120° between the clivus and the cribriform plate. The hypophysial fossa containing the pituitary gland lies at the angle formed between these two planes.

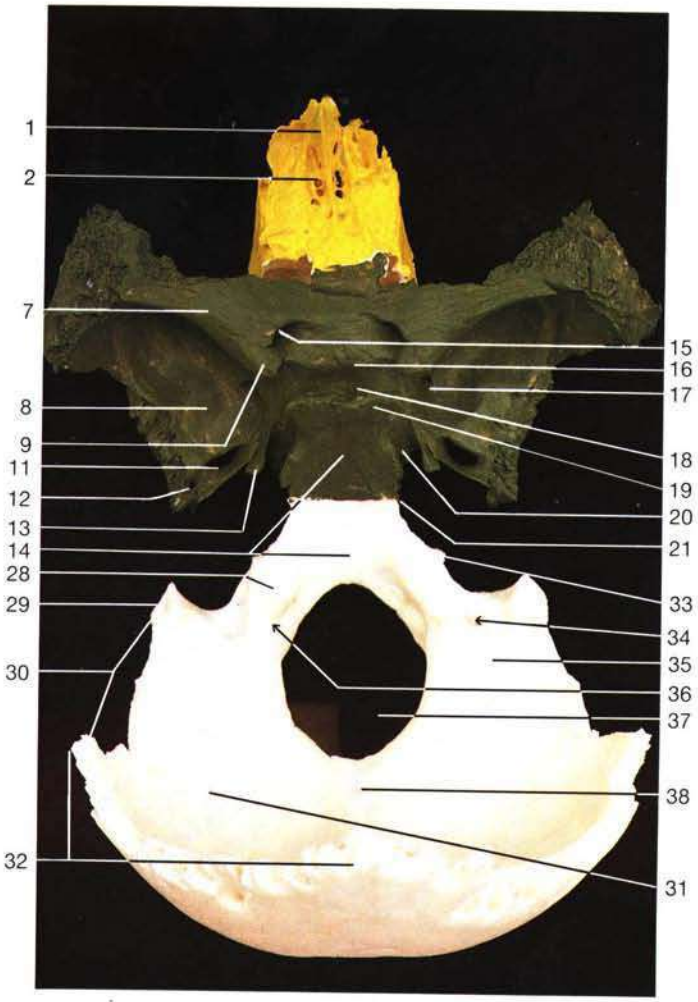

Part of the disarticulated base of the skull.
Ethmoid, sphenoid and occipital bones (from above).
Green = sphenoid bone; yellow = ethmoid bone.

Ethmoid bone

1 Crista galli
2 Cribriform plate
3 Ethmoidal air cells
4 Middle concha
5 Perpendicular plate (part of nasal septum)
6 Orbital plate

Sphenoid bone

7 Lesser wing
8 Greater wing
9 Anterior clinoid process
10 Posterior clinoid process
11 Foramen ovale
12 Foramen spinosum
13 Lingula of the sphenoid
14 Clivus
15 Optic canal
16 Tuberculum sellae
17 Foramen rotundum (right side)
18 Hypophysial fossa (sella turcica)
19 Dorsum sellae
20 Carotid sulcus
21 Sphenooccipital synchondrosis
22 Lateral pterygoid plate
23 Greater wing of sphenoid bone (orbital surface)
24 Greater wing of sphenoid bone (maxillary surface)
25 Foramen rotundum (left side)
26 Superior orbital fissure
27 Infratemporal crest of the greater wing

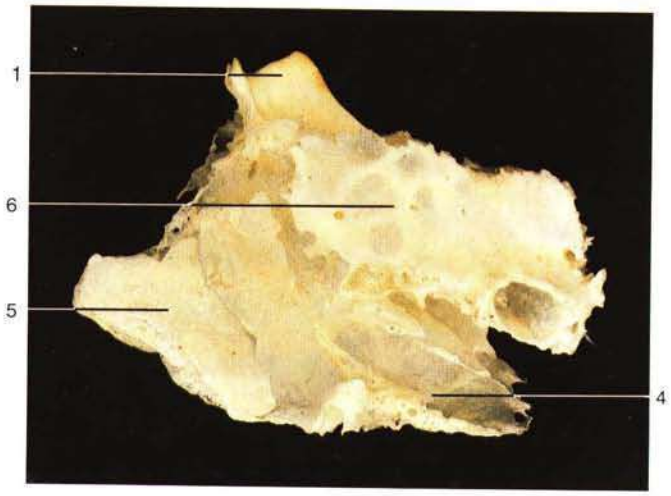

Ethmoid bone (lateral aspect), posterior portion to the right.

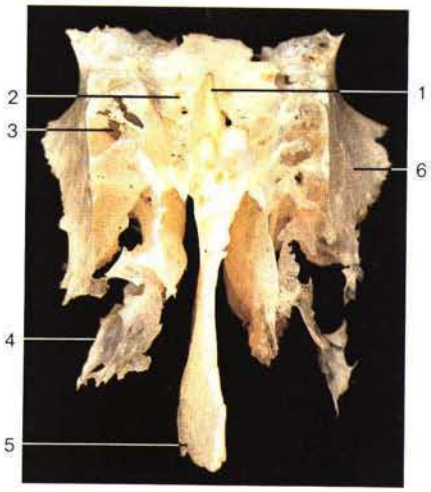

Ethmoid bone (anterior aspect).

Disarticulated base of the skull (anterior aspect).
Green = sphenoid bone; yellow = ethmoid bone; red = palatine bone.

Occipital bone

28 Jugular tubercle
29 Jugular process
30 Mastoid margin
31 Posterior cranial fossa
32 Lambdoid margin
33 Intrajugular process
34 Condylar canal
35 Lateral part of occipital bone
36 Hypoglossal canal
37 Foramen magnum
38 Internal occipital crest
39 Squamous part of occipital bone
40 Internal occipital protuberance

Maxilla

41 Orbital surface
42 Infraorbital groove
43 Maxillary tuberosity with foramina
44 Frontal process
45 Nasolacrimal groove
46 Infraorbital margin
47 Anterior nasal spine
48 Zygomatic process
49 Alveolar process

Palatine bone

50 Orbital process
51 Sphenopalatine notch
52 Sphenoidal process
53 Perpendicular plate
54 Horizontal plate
55 Pyramidal process

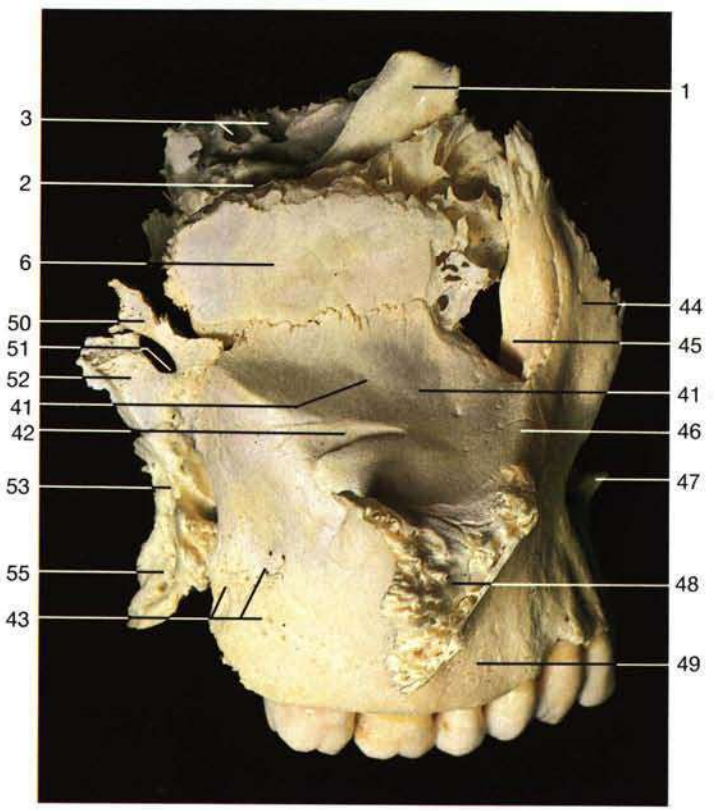

Right maxilla, ethmoid and palatine bone (lateral aspect).

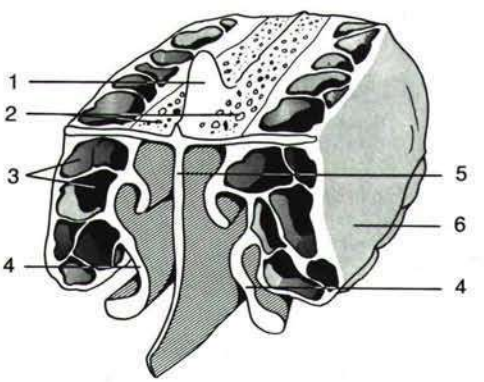

Ethmoid bone (oblique anterior aspect).
(Schematic drawing.)

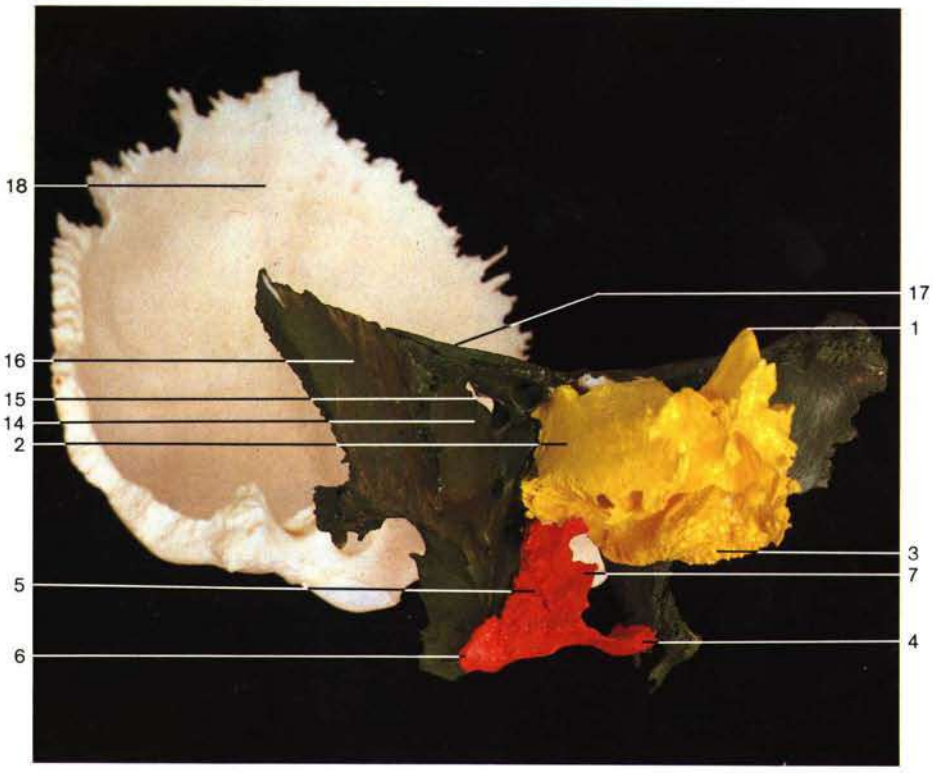

Part of a disarticulated skull base, similar to the foregoing figures, but with palatine bone. Green = sphenoid bone; yellow = ethmoid bone; red = palatine bone.

Ethmoid bone

1 Crista galli
2 Orbital plate
3 Middle concha

Palatine bone

4 Horizontal plate of palatine bone
5 Greater palatine canal
6 Pyramidal process
7 Maxillary process
8 Orbital process
9 Sphenopalatine notch
10 Perpendicular plate of palatine bone
11 Conchal crest
12 Nasal crest
13 Sphenoidal process

Sphenoid bone

14 Greater wing
15 Superior orbital fissure
16 Greater wing (orbital surface)
17 Lesser wing

Occipital bone

18 Squamous part of occipital bone

Maxilla

19 Maxillary tuberosity
20 Frontal process
21 Orbital surface
22 Infraorbital margin
23 Infraorbital groove
24 Zygomatic process
25 Alveolar process

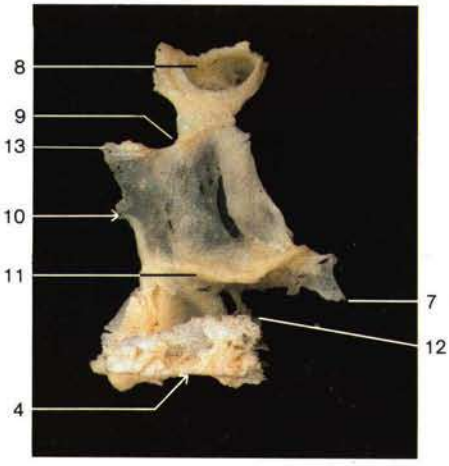

Left palatine bone (medial aspect, posterior aspect to the left).

Left palatine bone (anterior aspect).

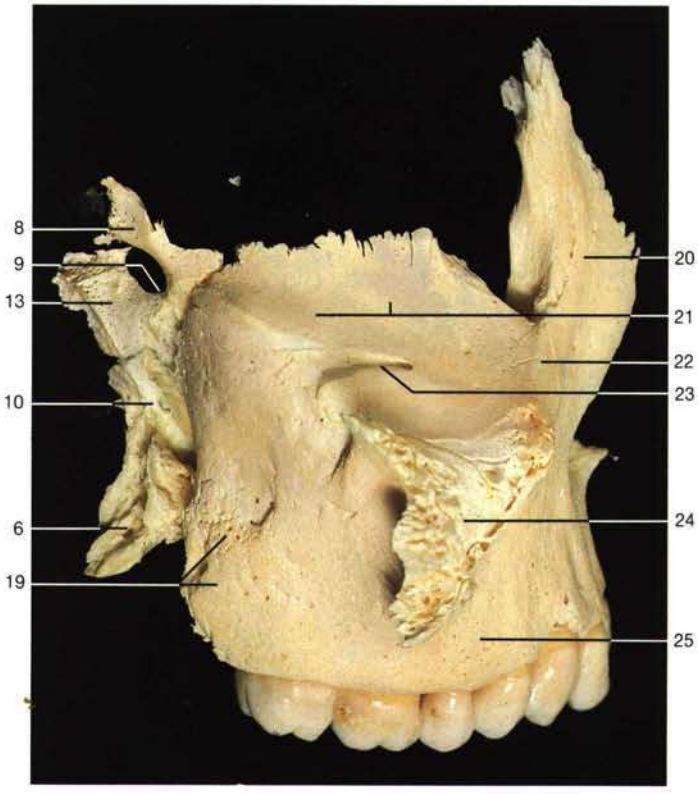

Right maxilla and right palatine bone (lateral aspect).

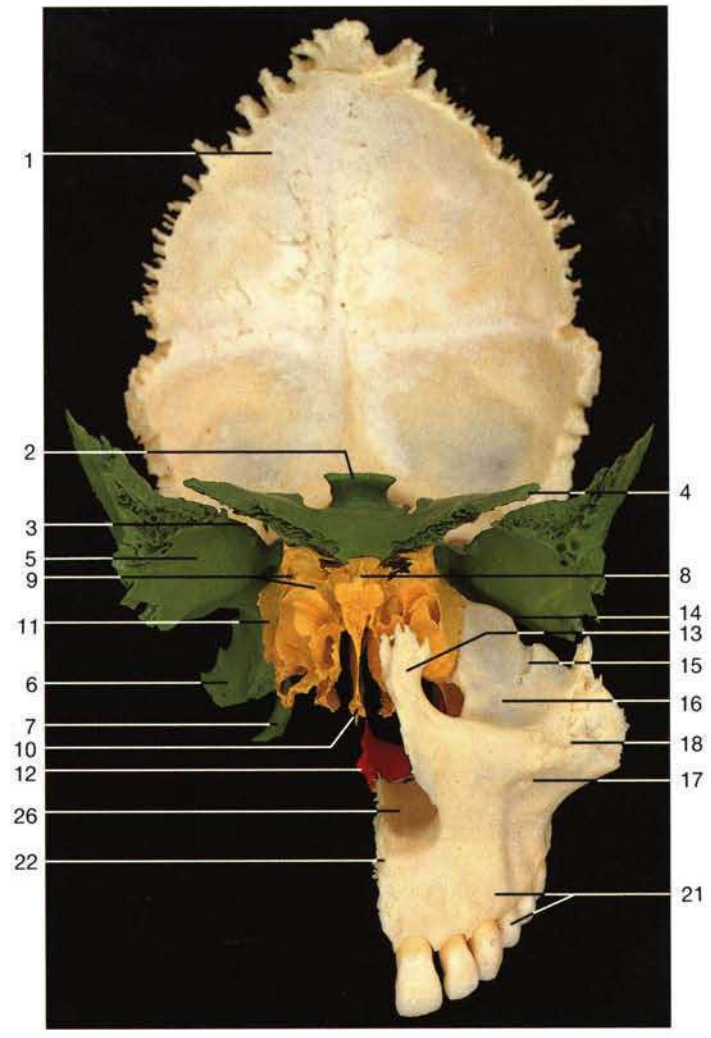

Occipital bone
1 Squamous part

Sphenoid bone
2 Dorsum sellae
3 Superior orbital fissure
4 Lesser wing
5 Greater wing (orbital surface)
6 Lateral pterygoid plate
7 Medial pterygoid plate

Ethmoid bone
8 Crista galli
9 Ethmoidal air cells
10 Perpendicular plate
11 Orbital plate

Palatine bone
12 Horizontal plate (nasal crest)

Maxilla
13 Frontal process
14 Inferior orbital fissure
15 Infraorbital groove
16 Orbital surface
17 Infraorbital foramen
18 Zygomatic process
19 Anterior lacrimal crest
20 Canine fossa
21 Alveolar process with teeth
22 Anterior nasal spine
23 Juga alveolaria (elevations formed by roots of teeth)
24 Lacrimal groove
25 Maxillary tuberosity with alveolar foramina
26 Palatine process of maxilla

Part of a disarticulated skull.
The left **maxilla** is added to the preceding specimen.

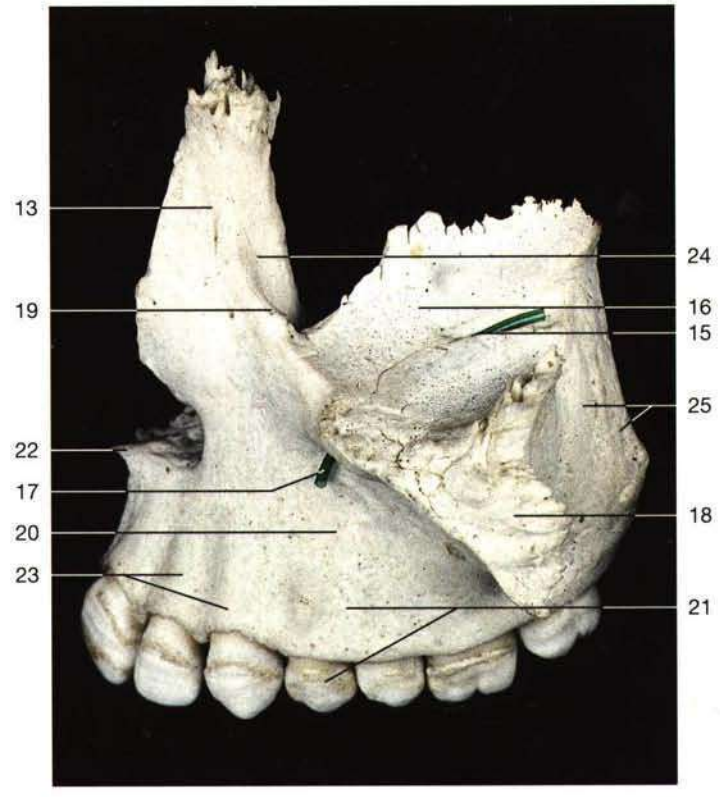

Left maxilla (lateral aspect). Probe = infraorbital canal.

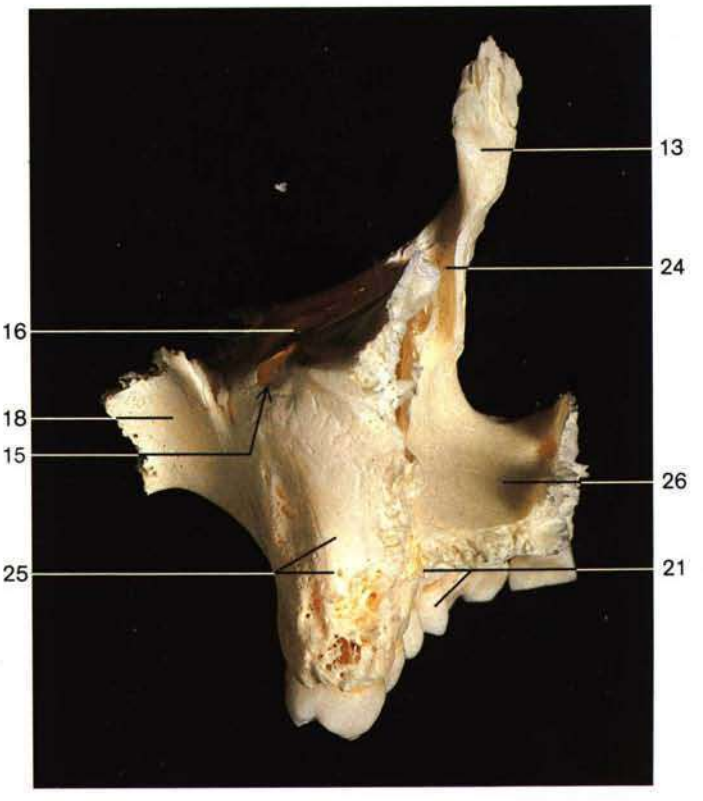

Left maxilla (posterior aspect).

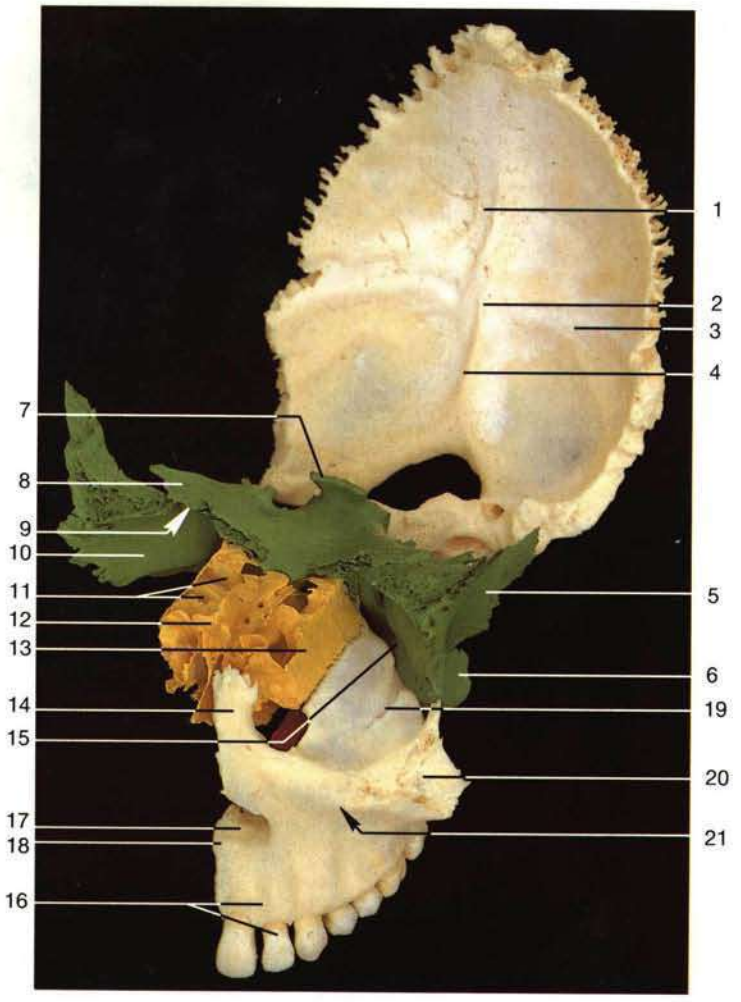

Part of a disarticulated base of skull. The mosaic of the facial bones [sphenoid bone (green), ethmoid bone (yellow), and palatine bone (red)] is seen from the anterior-lateral aspect.

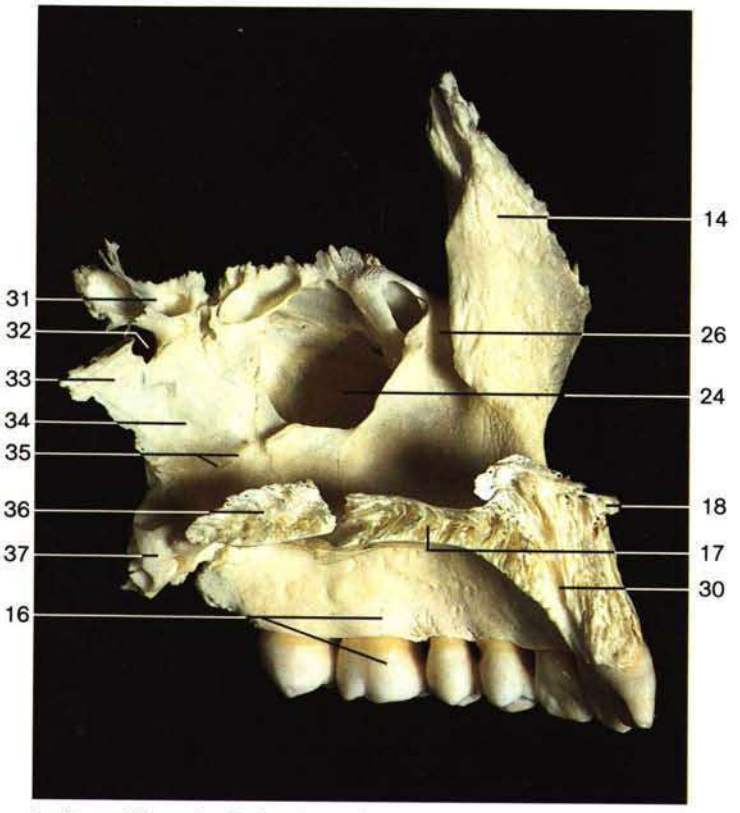

Left maxilla and palatine bone (medial aspect).

Occipital bone

1 Groove for superior sagittal sinus
2 Internal occipital protuberance
3 Groove for transverse sinus
4 Internal occipital crest

Sphenoid bone

5 Greater wing (temporal surface)
6 Lateral pterygoid plate
7 Dorsum sellae
8 Lesser wing
9 Superior orbital fissure
10 Greater wing (orbital surface)

Ethmoid bone

11 Ethmoidal air cells
12 Crista galli
13 Orbital plate

Maxilla

14 Frontal process
15 Inferior orbital fissure
16 Alveolar process with teeth
17 Palatine process
18 Anterior nasal spine
19 Infraorbital groove
20 Zygomatic process
21 Location of infraorbital foramen
22 Middle nasal meatus
23 Inferior nasal meatus
24 Maxillary hiatus (leading to maxillary sinus)
25 Third molar
26 Lacrimal groove
27 Conchal crest
28 Body of maxilla (nasal surface)
29 Nasal crest
30 Incisive canal

Palatine bone

31 Orbital process
32 Sphenopalatine notch
33 Sphenoidal process
34 Perpendicular plate
35 Conchal crest
36 Horizontal plate
37 Pyramidal process

Frontal bone

38 Squamous part
39 Supraorbital foramen
40 Frontal notch
41 Frontal spine

Inferior nasal concha

42 Inferior nasal concha with maxillary process

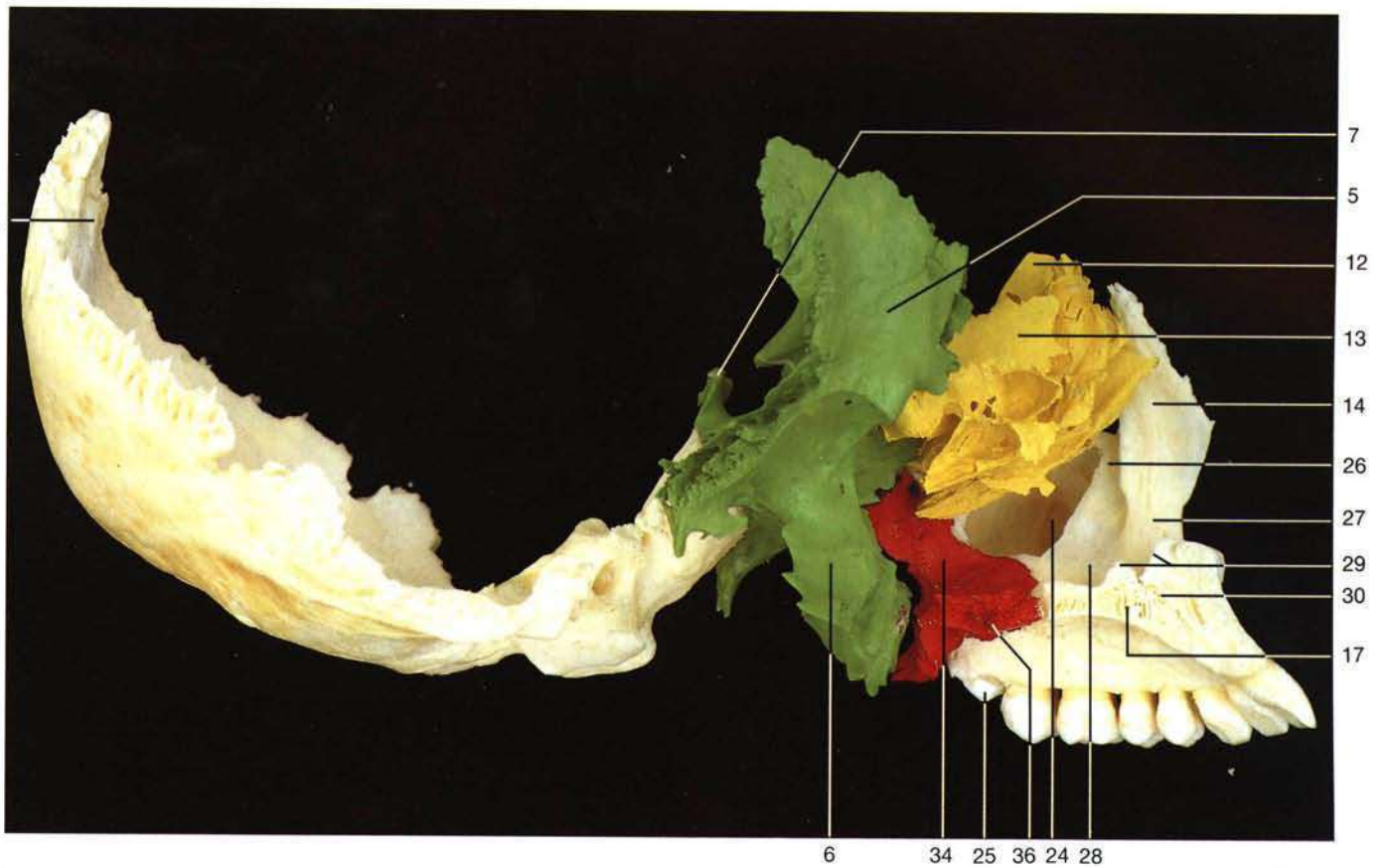

Part of a disarticulated base of skull (medial aspect). Green = sphenoid bone; yellow = ethmoid bone; red = palatine bone; natural colored = left maxilla.

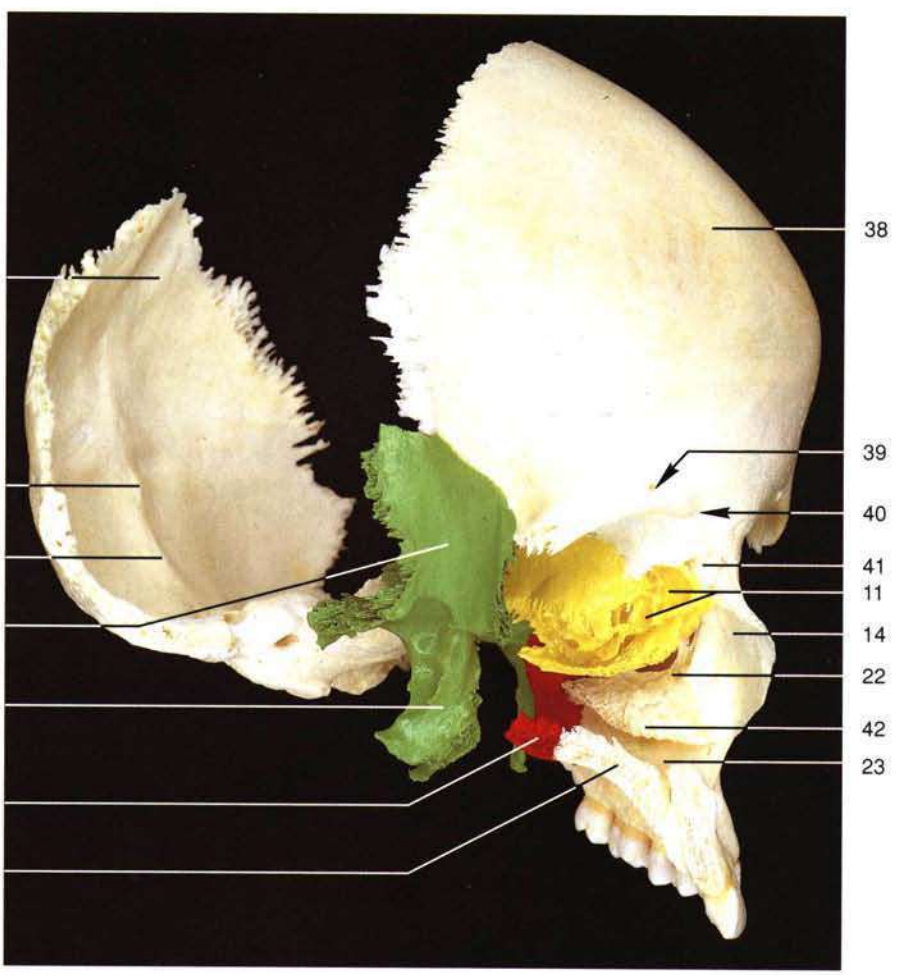

Part of a disarticulated base of skull. The same specimen as shown above but with frontal bone (oblique-lateral aspect).

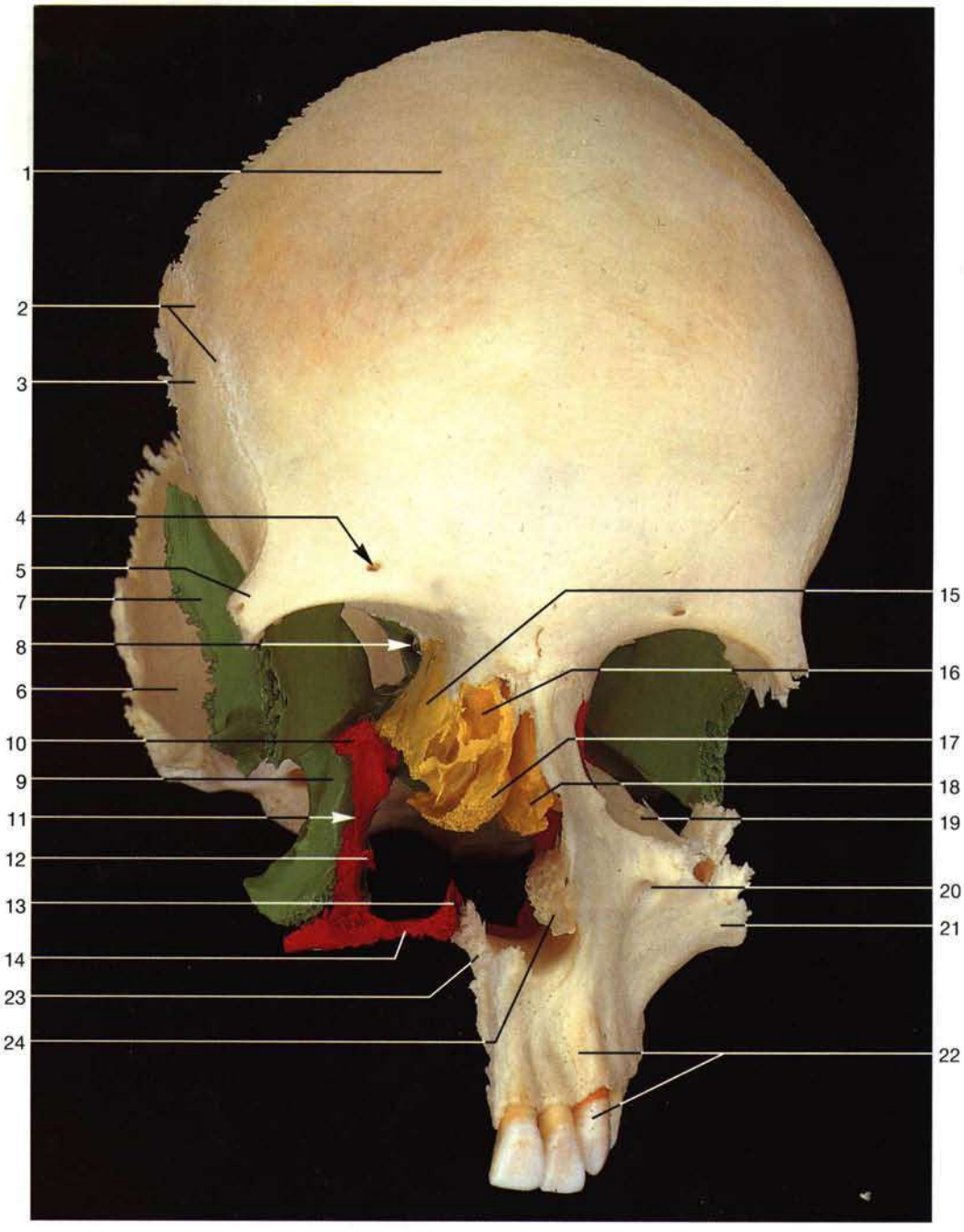

Part of a disarticulated skull, showing the connection of the palatine bone (red) and the maxilla with ethmoid bone (yellow) and sphenoid bone (light green) (anterior aspect).

Frontal bone
1 Squamous part
2 Inferior temporal line
3 Temporal surface
4 Supraorbital foramen
5 Zygomatic process

Occipital bone
6 Squamous part

Sphenoid bone
7 Greater wing (temporal surface)
8 Optic canal within the lesser wing
9 Lateral pterygoid plate

Palatine bone
10 Orbital process
11 Perpendicular plate
12 Conchal crest
13 Nasal crest
14 Horizontal plate

Ethmoid bone
15 Orbital plate
16 Ethmoidal air cell
17 Middle concha
18 Perpendicular plate (part of bony nasal septum)

Maxilla
19 Infraorbital groove
20 Infraorbital foramen
21 Zygomatic process
22 Alveolar process with teeth
23 Palatine process

Left inferior nasal concha
24 Anterior part of inferior concha

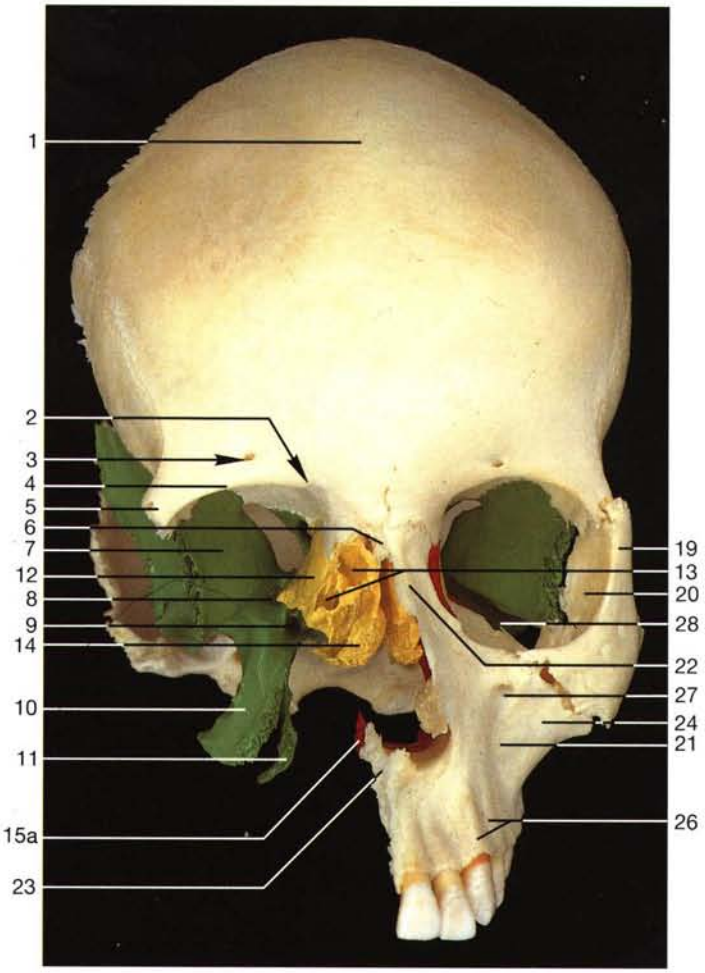

Anterior view of a disarticulated skull, showing the connection of the maxilla with the frontal and zygomatic bones. Yellow = ethmoid bone; red = palatine bone; green = sphenoid bone.

Frontal bone

1 Squamous part
2 Frontal notch
3 Supraorbital foramen
4 Supraorbital margin
5 Zygomatic process
6 Frontal spine

Sphenoid bone

7 Greater wing (orbital surface)
8 Foramen rotundum
9 Pterygoid or Vidian canal
10 Lateral pterygoid plate
11 Medial pterygoid plate

Ethmoid bone

12 Orbital plate
13 Ethmoidal air cells
14 Middle concha

Palatine bone

15 Horizontal plate
15a Nasal crest
16 Pyramidal process
17 Lesser palatine foramen
18 Greater palatine foramen

Zygomatic bone

19 Frontal process
20 Orbital surface

Maxilla

21 Canine fossa
22 Frontal process
23 Palatine process
24 Zygomatic process
25 Alveolar process and teeth
26 Juga alveolaria
27 Infraorbital foramen
28 Infraorbital groove
29 Anterior nasal aperture
30 Anterior nasal spine

Incisive bone

31 Central incisor and incisive bone or premaxilla
32 Incisive fossa

Vomer

33 Ala of the vomer

Sutures and choanae

34 Median palatine suture
35 Transverse palatine suture
36 Choanae

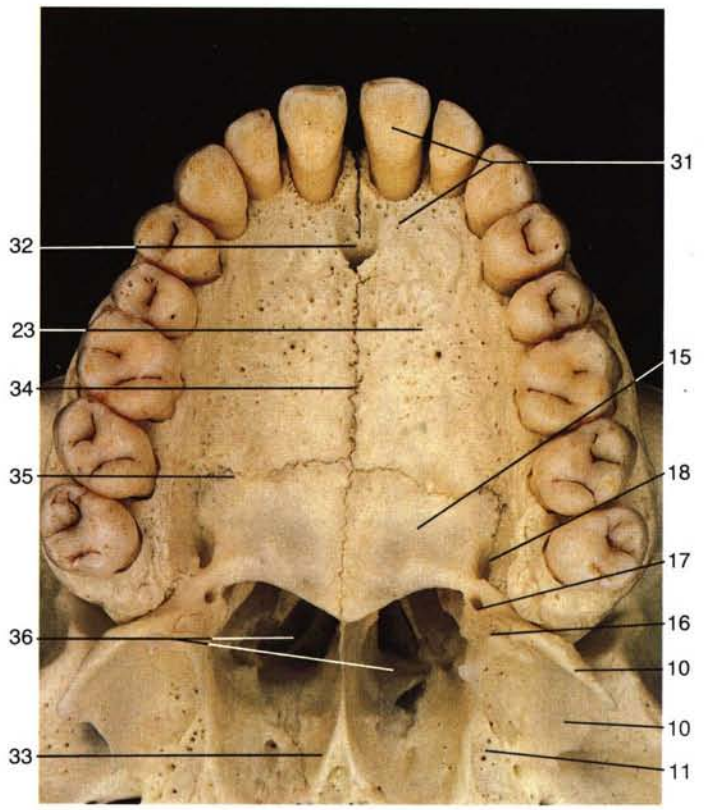

Bony palate and teeth of the maxillae (from below).

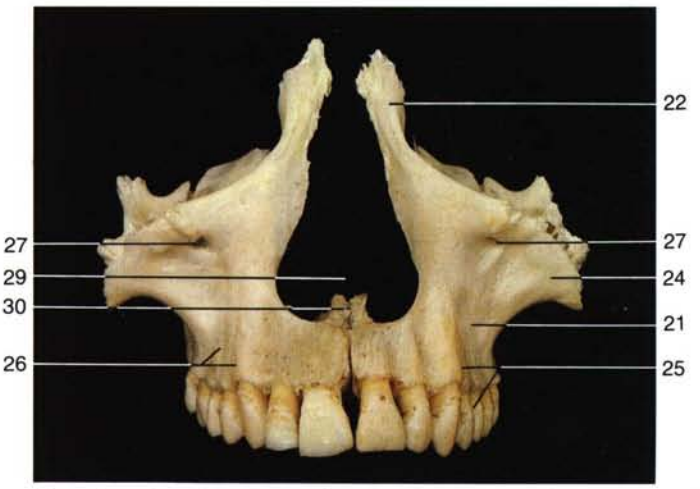

Anterior view of both maxillae, forming the anterior bony aperture of the nose.

Paramedian section through the skull, right side (lateral aspect). Frontal and maxillary sinus are opened.

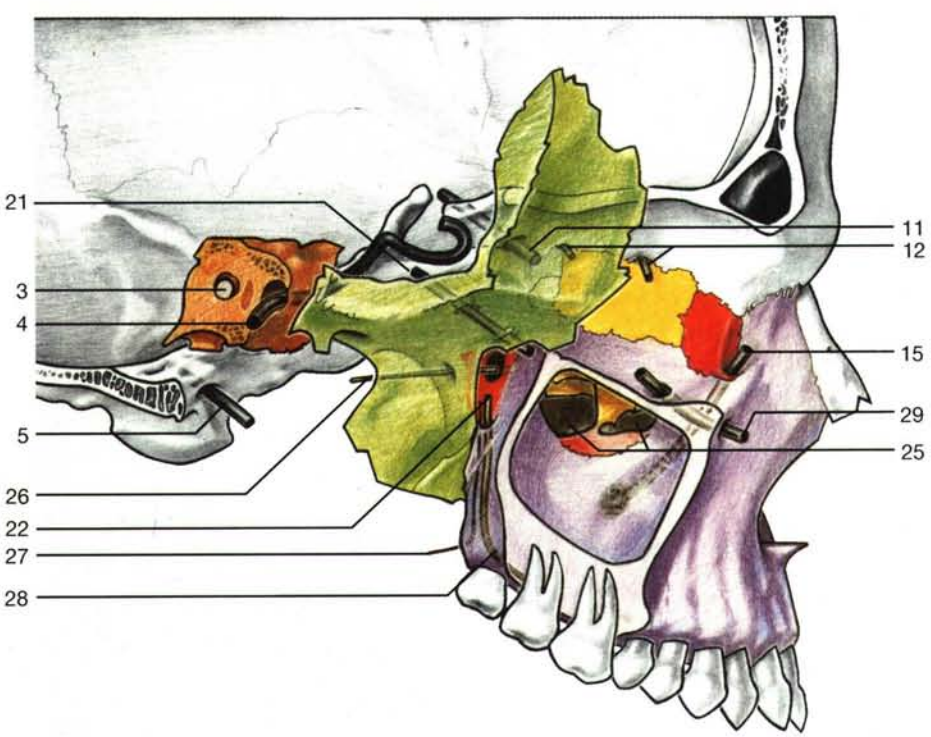

Illustration of canals and foramina connected with the right orbit and pterygopalatine fossa (compare the above figure). The greater wing of sphenoid bone (green) is shown as being transparent. Brown = temporal bone; yellow = ethmoid bone; red = laerimal bone; light red = inferior nasal concha; violet = maxilla; orange = palatine bone.

1 Occipital bone
2 Temporal bone (petrous part)
3 **Internal acoustic meatus**
4 **Carotid canal**
5 **Hypoglossal canal**
6 Occipital condyle
7 Lateral plate of pterygoid process
8 Dorsum of sella turcica
9 Sella turcica
10 Frontal sinus
11 **Optic canal**
12 **Posterior** and **anterior ethmoidal foramen**
13 Orbital plate of ethmoidal bone
14 Nasal bone
15 **Nasolacrimal canal**
16 Uncinate process
17 Inferior nasal concha (maxillary process)
18 Maxillary sinus
19 Anterior nasal spine
20 Alveolar process of maxilla
21 **Foramen rotundum**
22 **Pterygopalatine fossa**
23 Tuberosity of maxilla with **alveolar foramina**
24 **Sphenopalatine foramen**
25 Maxillary hiatus
26 **Pterygoid** or Vidian **canal**
27 **Lesser palatine canal**
28 **Greater palatine canal**
29 **Infraorbital canal**

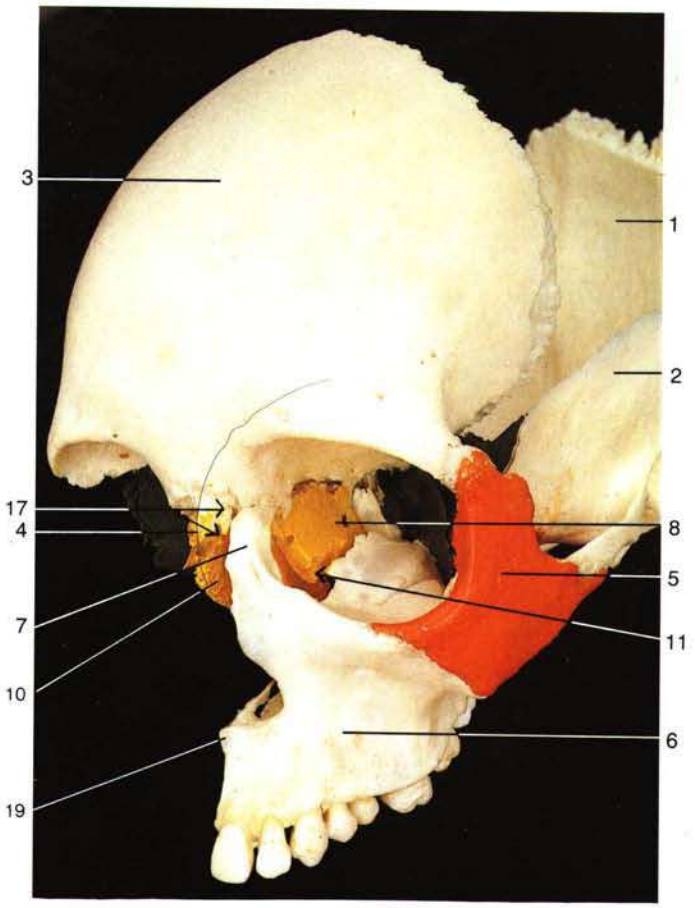

Anterior part of a disarticulated skull.
Orange = zygomatic bone; yellow = ethmoid bone; green = sphenoidal bone. The arrows indicate the locations of the lacrimal bone (11) and the nasal bone (17).

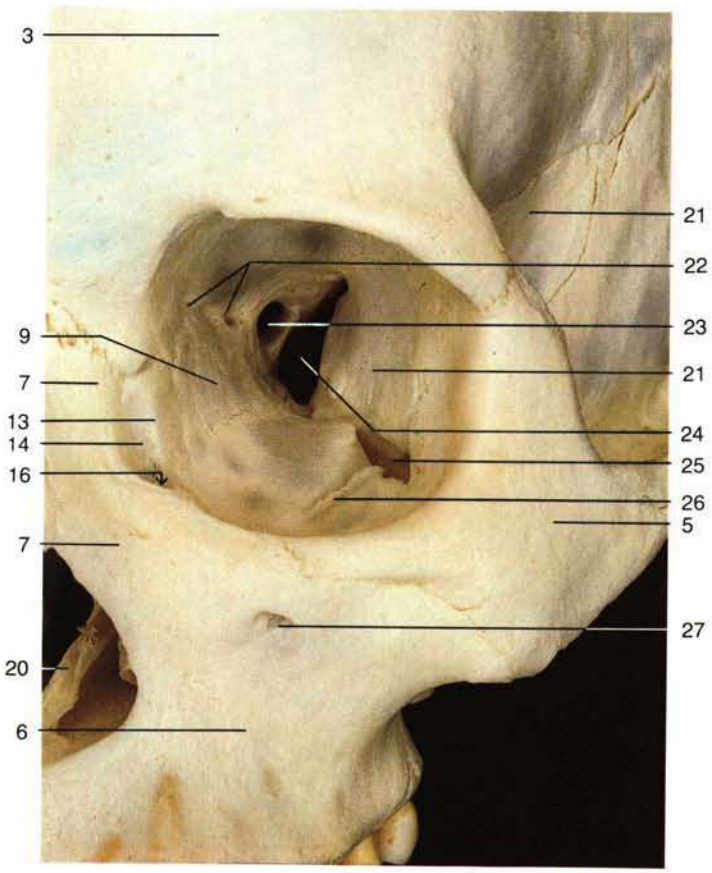

Left orbit (anterior aspect).

1 **Occipital bone**
2 **Temporal bone**
3 **Frontal bone**
4 Nasal spine of frontal bone
5 **Zygomatic bone**
6 **Maxilla**
7 Frontal process of maxilla
8 **Ethmoid bone**
9 Orbital plate of ethmoid bone
10 Perpendicular plate of ethmoid bone
11 Site of **lacrimal bone**
12 Lacrimal groove of lacrimal bone
13 Posterior lacrimal crest
14 Fossa for lacrimal sac
15 Lacrimal hamulus
16 Nasolacrimal canal
17 Site of **nasal bone**
18 Nasal foramina of nasal bone
19 Anterior nasal spine of maxilla
20 **Vomer**
21 Greater wing of sphenoid bone
22 Anterior and posterior ethmoidal foramina
23 Optic canal
24 Superior orbital fissure
25 Inferior orbital fissure
26 Infraorbital groove
27 Infraorbital foramen

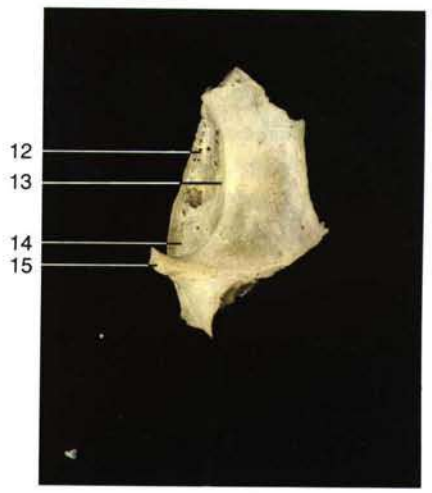

Left lacrimal bone (anterior aspect).

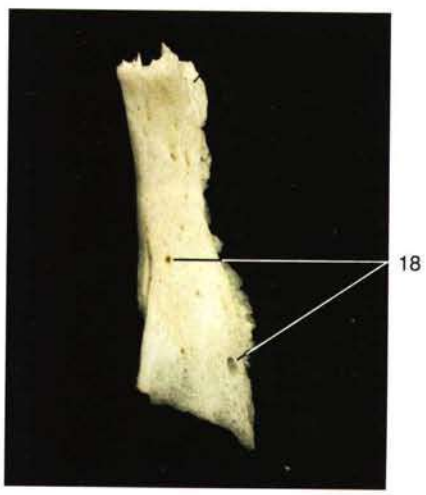

Left nasal bone (anterior aspect).

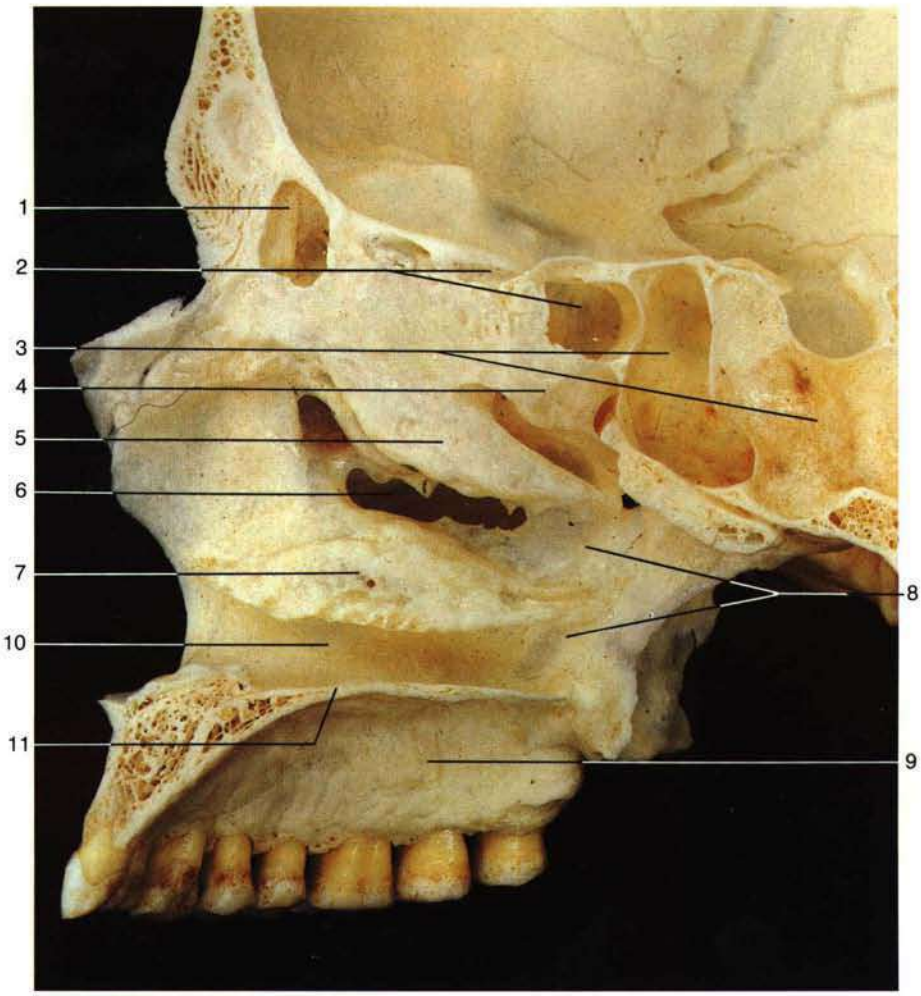

1 Frontal sinus
2 Ethmoidal air cells
3 Sphenoid sinus
4 Superior nasal concha
5 Middle nasal concha
6 Maxillary hiatus
7 Inferior nasal concha
8 Palatine bone
9 Maxilla
10 Inferior meatus
11 Palatine process of the maxilla

Lateral wall of the nasal cavity. Median section through the skull.

▷
To page 49:

Blue	=	Occipital bone
Light green	=	Parietal bone
Light brown	=	Frontal bone
Dark brown	=	Temporal bone
Red	=	Sphenoid bone
Dark green	=	Ethmoid bone
Light blue	=	Nasal bone
Pink	=	Inferior concha
Orange	=	Vomer
Violet	=	Maxilla
White	=	Palatine bone
White	=	Mandible

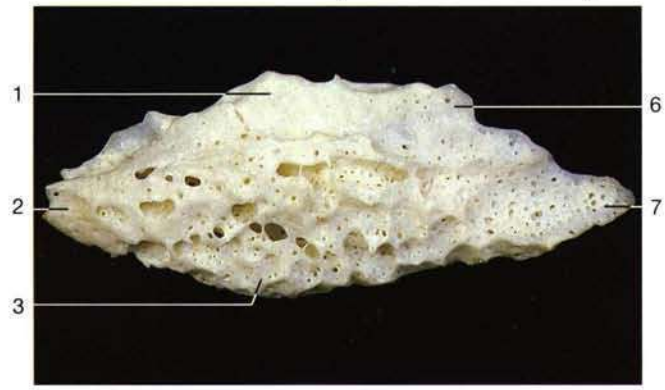

Right inferior nasal concha (medial aspect). Anterior part to the left.

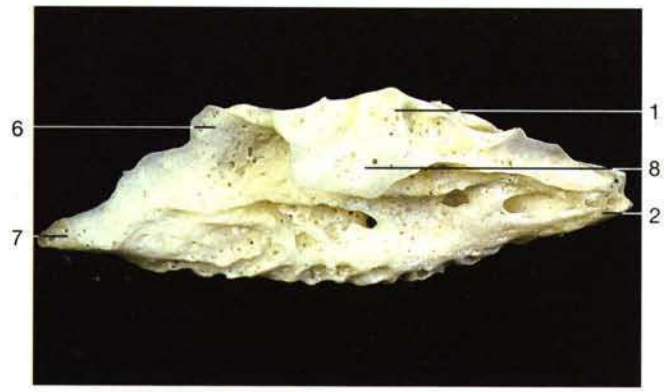

Right inferior nasal concha (lateral aspect). Anterior part to the right.

Inferior Concha and Vomer

1 Ethmoidal process
2 Anterior part of concha
3 Inferior border
4 Ala of vomer
5 Posterior border of nasal septum
6 Lacrimal process
7 Posterior part of concha
8 Maxillary process

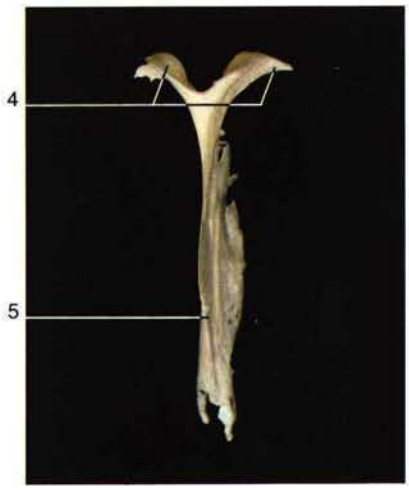

Vomer (posterior aspect).

1 Crista galli
2 Cribriform plate of ethmoid bone
3 Perpendicular plate of ethmoid bone
4 Vomer
5 Ala of the vomer
6 Palatine bone (perpendicular process)
7 Palatine bone (horizontal plate)
8 Mandible
9 Nasal bone
10 Sphenoidal sinus
11 Hypophysial fossa (sella turcica)
12 Grooves for the middle meningeal artery

Cartilages of the nose

13 Lateral nasal cartilages
14 Greater alar cartilage
15 Lesser alar cartilages
16 Septal cartilage

17 Location of nasal bone

Paramedian sagittal section through the skull including the nasal septum.

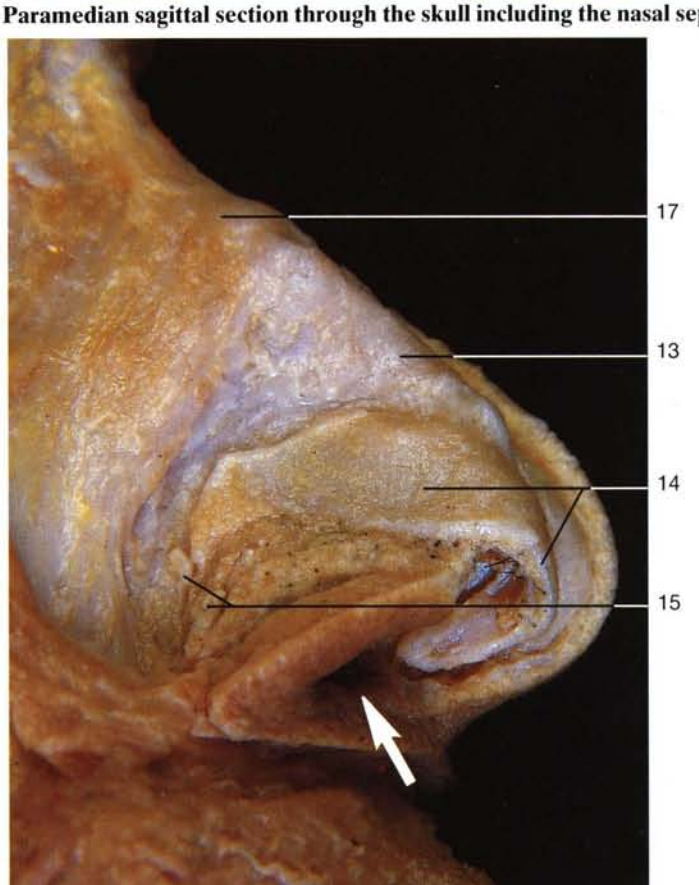

Cartilages of the nose (right anterior aspect). Arrow = nostril, framed by nasal wing.

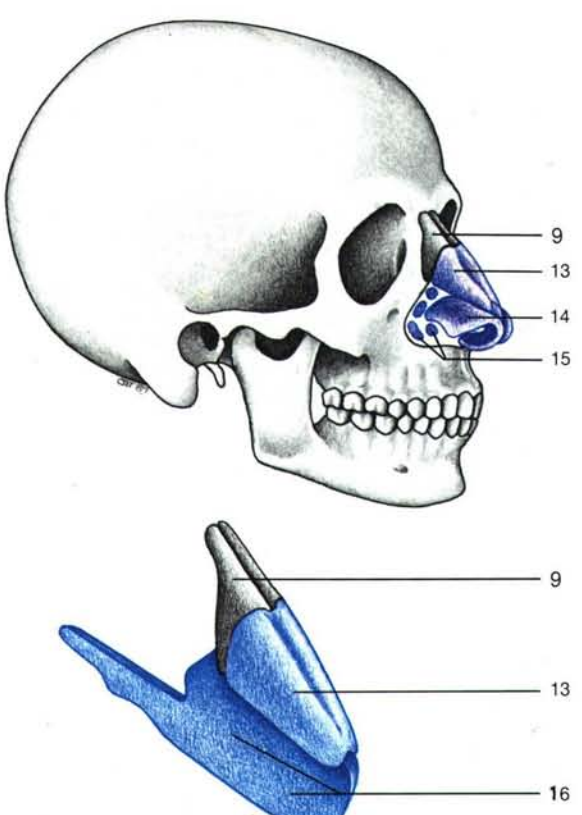

Cartilages of the nose.
Schematic diagram of the external nose.

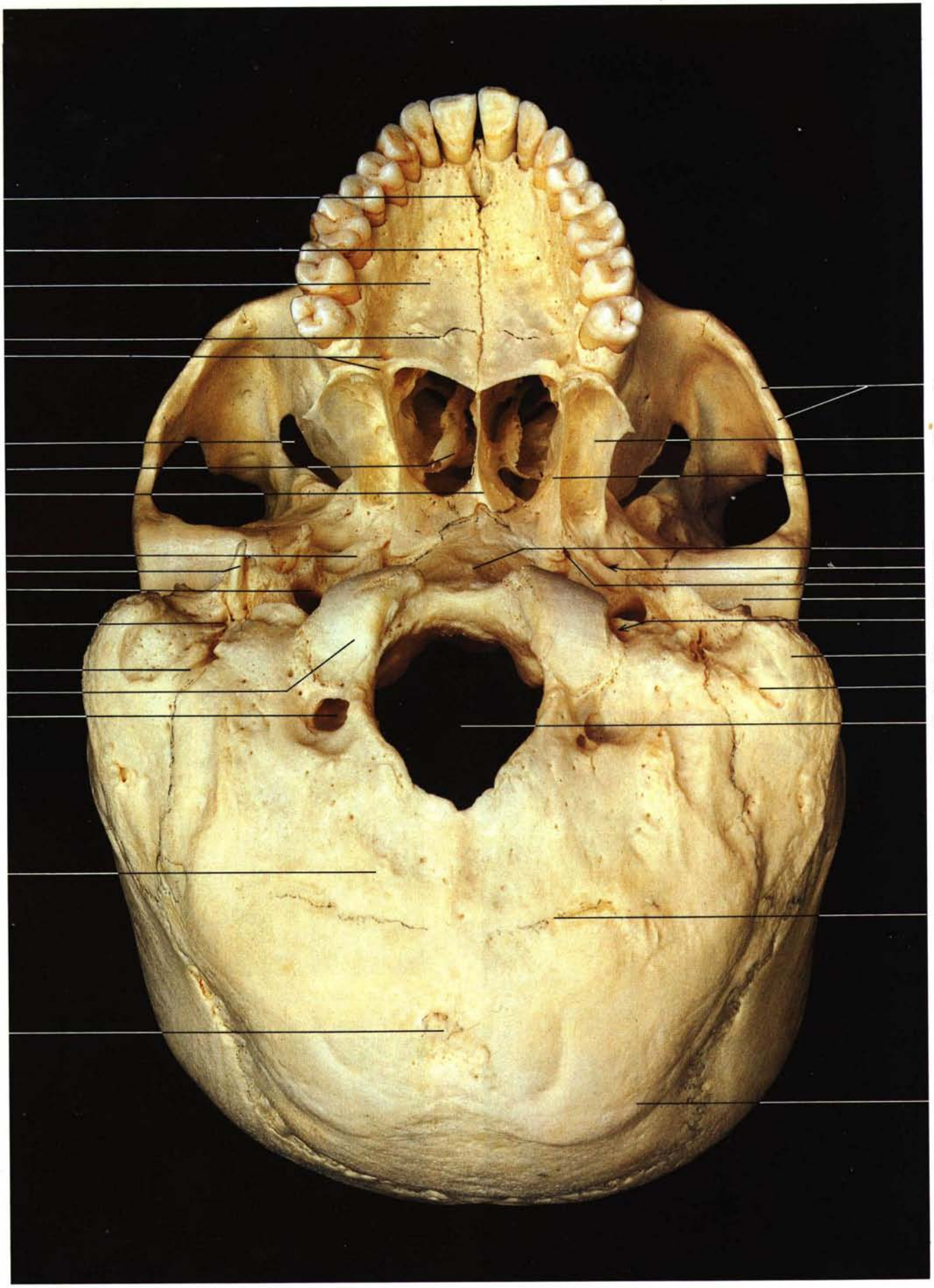

Base of the skull (inferior aspect).

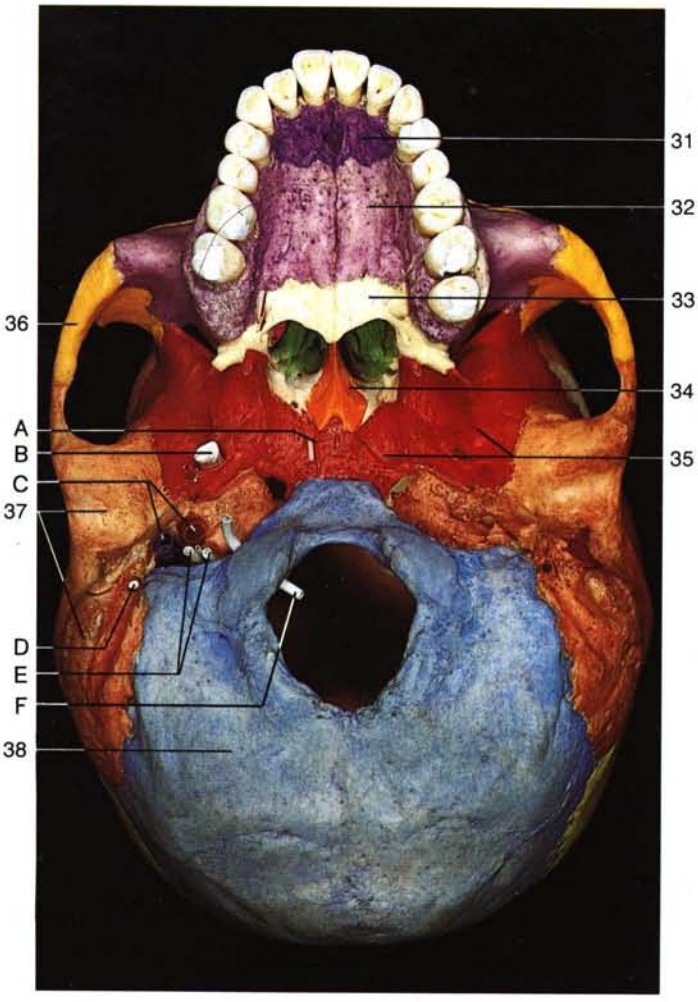

Base of the skull (from below). The individual bones are indicated by different colours.

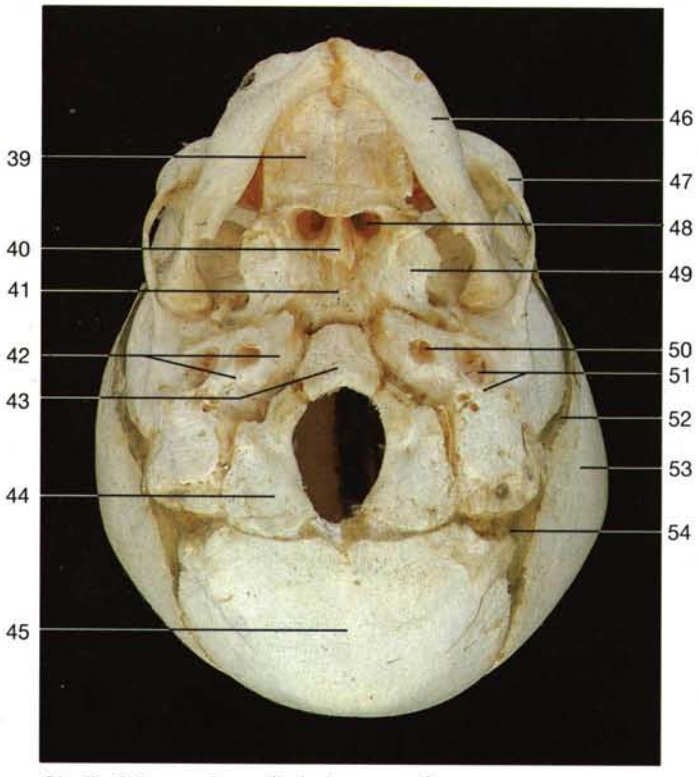

Skull of the newborn (inferior aspect).

A **Pterygoid canal**
B **Foramen ovale**
C Internal carotid artery within **carotid canal** and internal jugular vein within the venous part of jugular foramen
D **Stylomastoid foramen** (facial nerve)
E **Jugular foramen** (glossopharyngeal, vagus and accessory nerves)
F **Hypoglossal canal** (hypoglossal nerve)

1 Incisive canal
2 Median palatine suture
3 Palatine process of maxilla
4 Palatomaxillary suture
5 Greater and lesser palatine foramina
6 Inferior orbital fissure
7 Middle concha (process of ethmoid bone)
8 Vomer
9 Foramen ovale
10 Groove for auditory tube
11 Pterygoid canal
12 Styloid process
13 Carotid canal
14 Stylomastoid foramen
15 Jugular foramen
16 Groove for occipital artery
17 Occipital condyle
18 Condylar canal
19 Nuchal plane
20 External occipital protuberance
21 Zygomatic arch
22 Lateral pterygoid plate
23 Medial pterygoid plate
24 Mandibular fossa
25 Pharyngeal tubercle
26 Superior nuchal line
27 Mastoid process
28 Inferior nuchal line
29 Mastoid notch
30 Foramen magnum
31 **Incisive bone** or **premaxilla** (dark violet)
32 **Maxilla** (violet)
33 **Palatine bone** (white)
34 **Vomer** (orange)
35 **Sphenoid bone** (red)
36 **Zygomatic bone** (yellow)
37 **Temporal bone** (brown)
38 **Occipital bone** (blue)
39 Palatine process of maxilla
40 Vomer
41 Sphenoid bone
42 Petrous part of temporal bone
43 Basilar part } of occipital bone
44 Lateral part } of occipital bone
45 Squamous part } of occipital bone
46 Mandible
47 Zygomatic arch
48 Choana
49 Pterygoid process of sphenoid bone
50 Carotid canal
51 External acoustic meatus (tympanic annulus)
52 Sphenoidal fontanelle
53 Parietal bone
54 Mastoid fontanelle

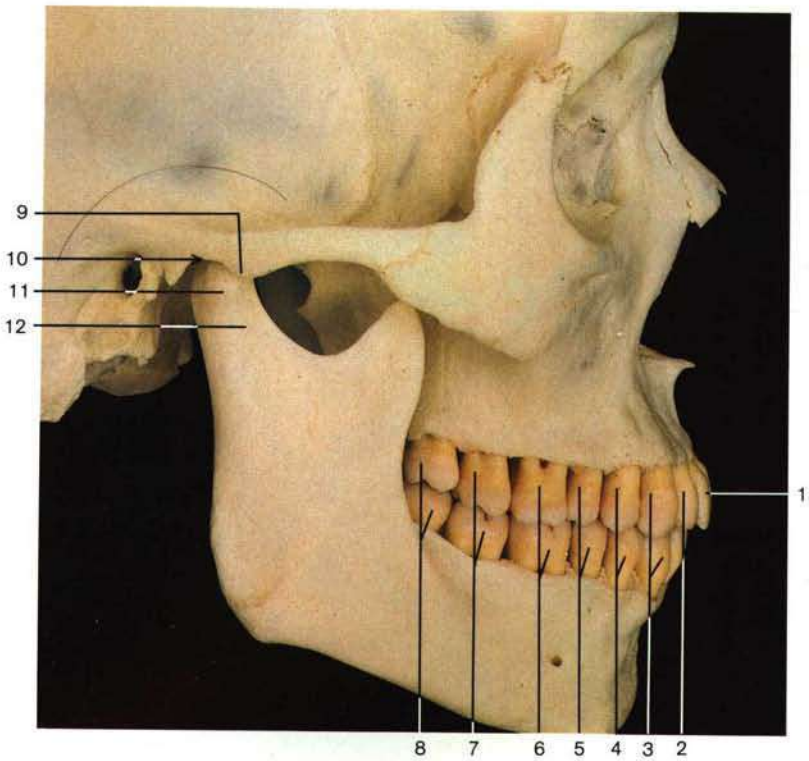

1 Central incisor
2 Lateral incisor
3 Canines
4 First premolars or bicuspids
5 Second premolars or bicuspids
6 First molars
7 Second molars
8 Third molars
9 Articular tubercle
10 Mandibular fossa
11 Head of mandible
12 Condylar process

Normal position of teeth. Dentition in centric occlusion (lateral view).

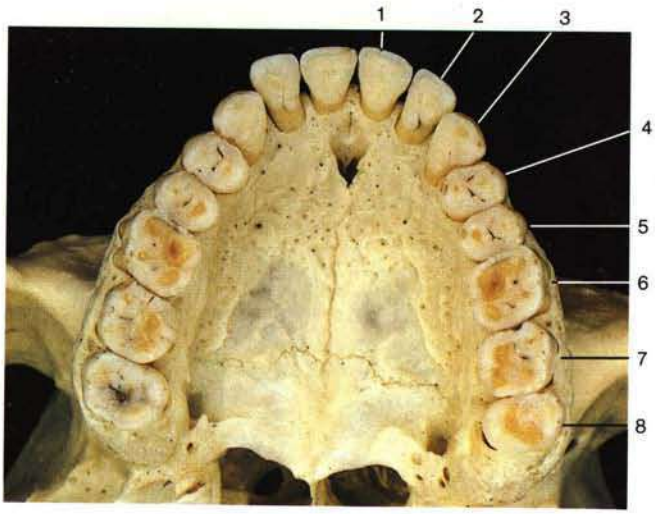

Upper teeth of the adult (inferior aspect).

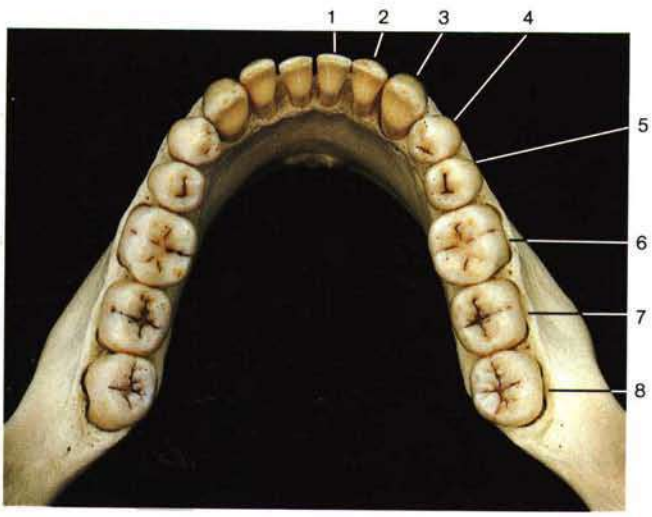

Lower teeth of the adult (superior aspect).

Table of dentition. Eruption of deciduous and permanent teeth (after C. Röse according to A. Kröncke).

Primary dentition (Deciduous teeth)	Maxilla months post partum		Mandible months	
1. Central incisor	10.3		8.6	
2. Lateral incisor	12.2		14.4	
3. Cuspid incisor	19.5		20.1	
4. First molar	15.5		16.5	
5. Second molar	24.8		24.5	
Secondary dentition (Permanent teeth)	years and months ♂	♀	years and months ♂	♀
1. Central incisor	7/8	7/5	6/10	6/7
2. Lateral incisor	8/11	8/6	7/11	7/7
3. Cuspid incisor	12/2	11/7	11/12	10/3
4. First premolar	10/5	10/1	11/3	10/8
5. Second premolar	11/4	11/1	12/0	11/7
6. First molar	6/7	6/6	6/5	6/3
7. Second molar	12/9	12/5	12/3	11/9

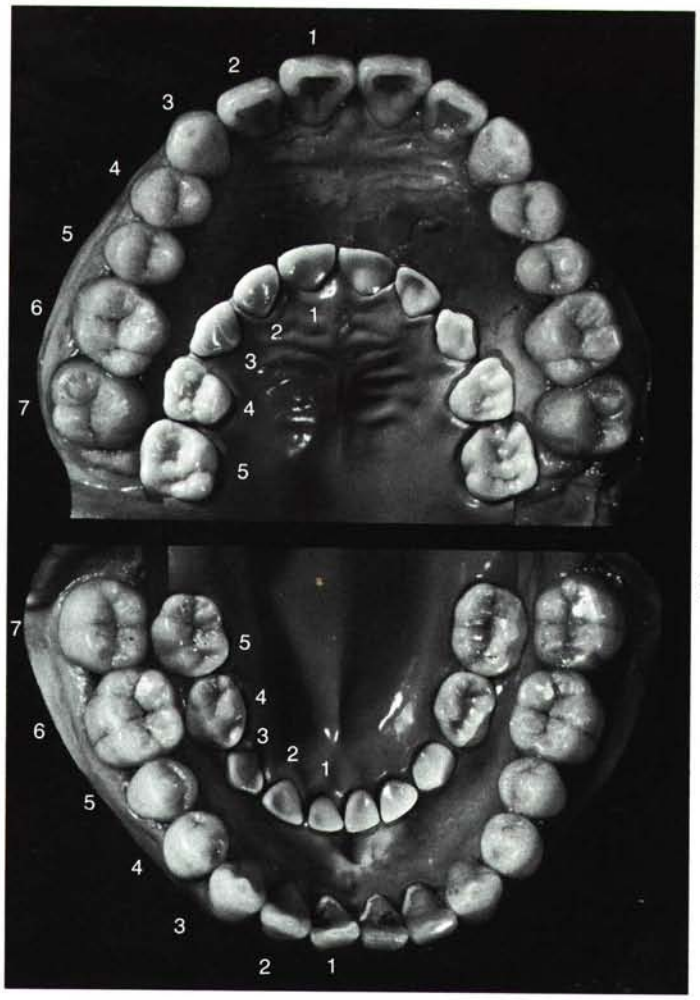

Comparison of the deciduous and permanent teeth. Notice that the breadth of the alveolar arch of the child's mandible and maxilla holding the deciduous teeth is nearly the same as the comparable portion in the jaws of the adult. Note the unerupted third molars.

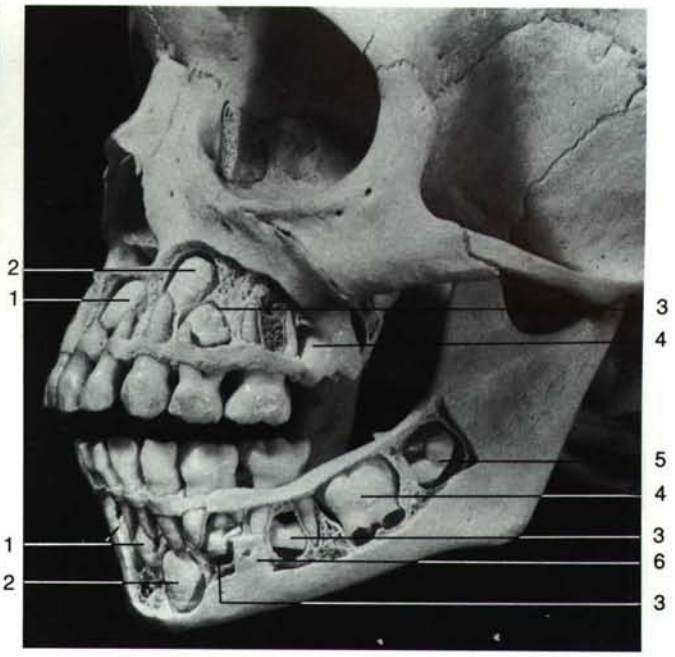

Deciduous teeth in child's skull. The developing crowns of the permanent teeth are displayed in their sockets in the maxilla and mandible.

△

1 Permanent incisors
2 Permanent cuspid (canine)
3 Premolars
4 First permanent molar
5 Second permanent molar
6 Mental foramen

Isolated teeth of the alveolar part of the maxilla (top row) and the mandible (lower row), labial surface of the teeth.

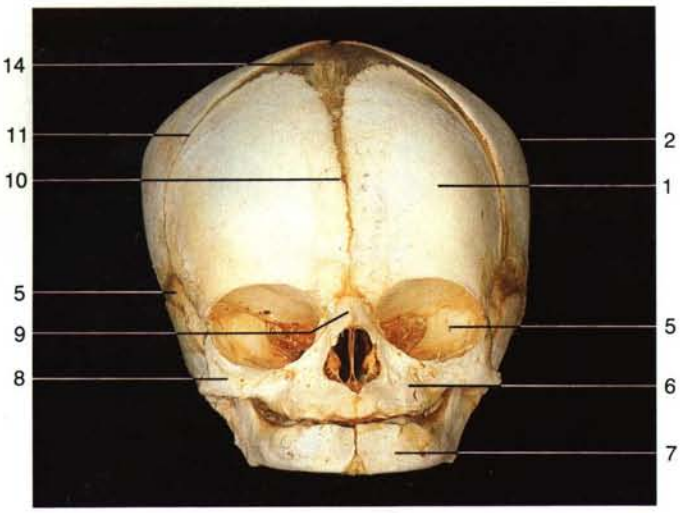

Skull of the newborn (anterior aspect).

Skull of the newborn (superior aspect). Calvaria.

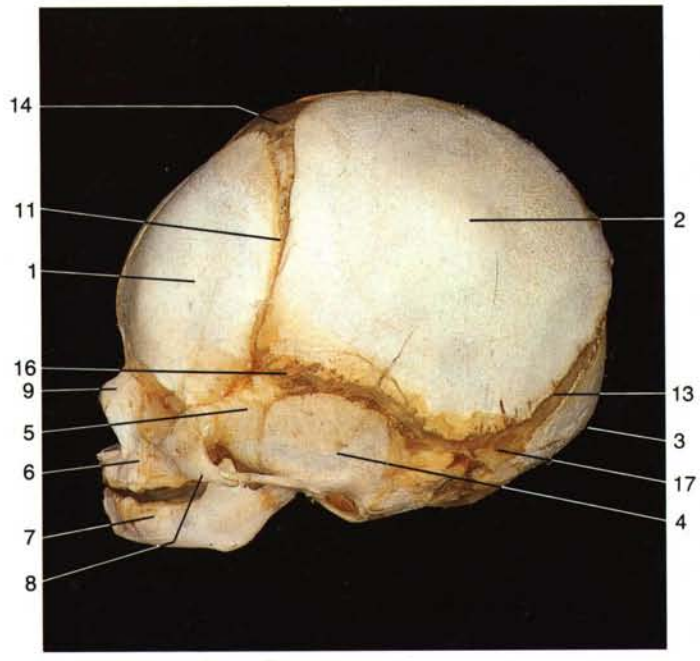

Skull of the newborn (lateral aspect).

Cranial skeleton

1 Frontal tuber or eminence
2 Parietal tuber or eminence
3 Occipital tuber or eminence
4 Squamous part of temporal bone
5 Greater wing of sphenoid bone

Facial skeleton

6 Maxilla
7 Mandible
8 Zygomatic bone
9 Nasal bone

Sutures and fontanelles

10 Frontal suture
11 Coronal suture
12 Sagittal suture
13 Lambdoid suture
14 Anterior fontanelle
15 Posterior fontanelle
16 Sphenoidal fontanelle
17 Mastoid fontanelle

Base of the skull

18 Frontal bone
19 Ethmoid bone
20 Sphenoid bone
21 Hypophysial fossa (sella turcica)
22 Dorsum sellae
23 Temporal bone
24 Mastoid fontanelle
25 Occipital bone

In the newborn the facial skeleton, in contrast to the cranial skeleton, appears relatively small. There are no teeth presenting. The bones of the cranium are separated by wide fontanelles.

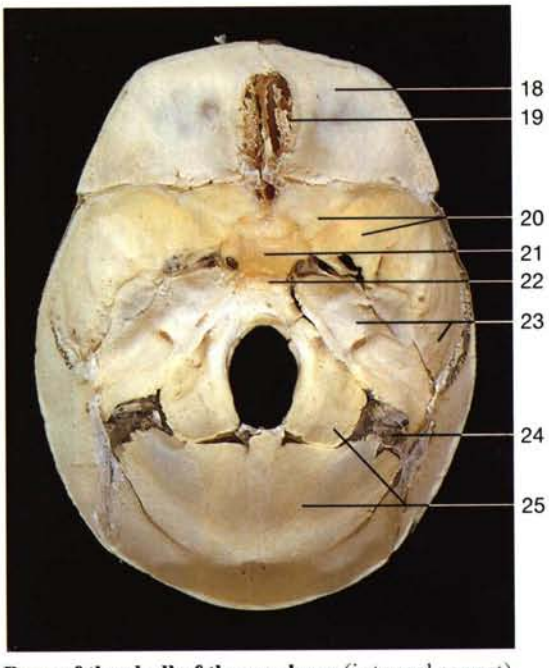

Base of the skull of the newborn (internal aspect).

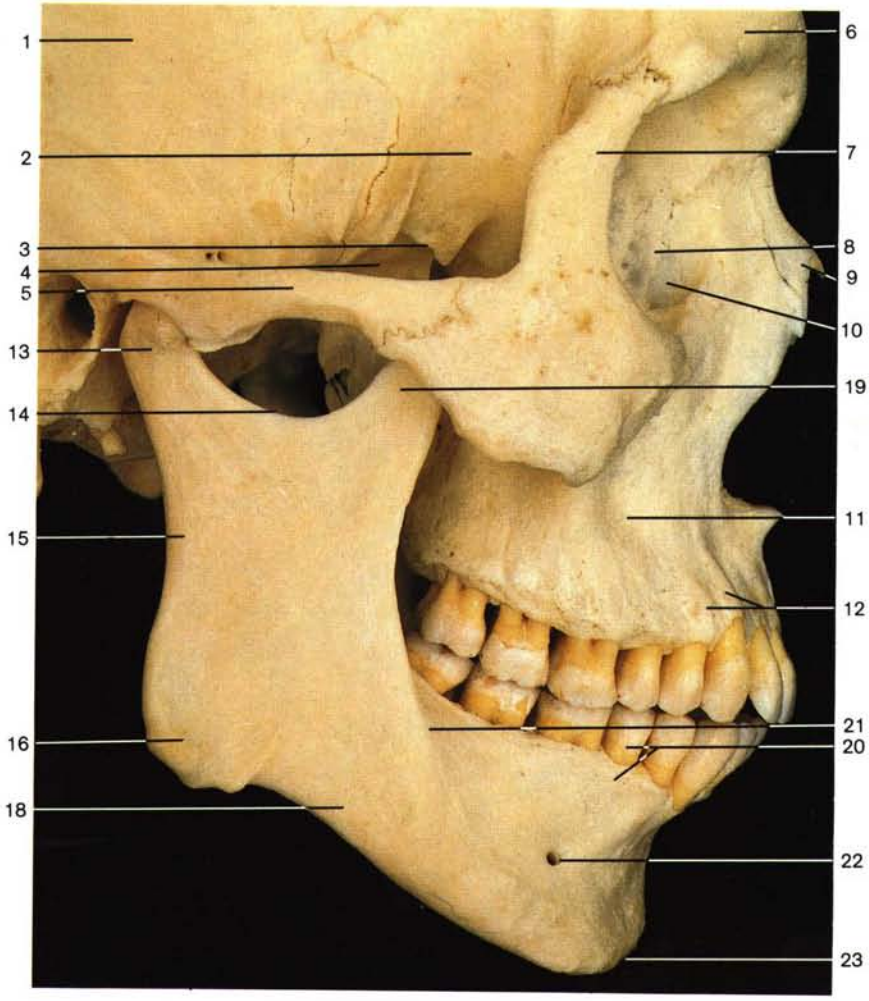

Lateral aspect of the facial bones. Mandible and teeth in the position of occlusion. Upper and lower jaw occluded.

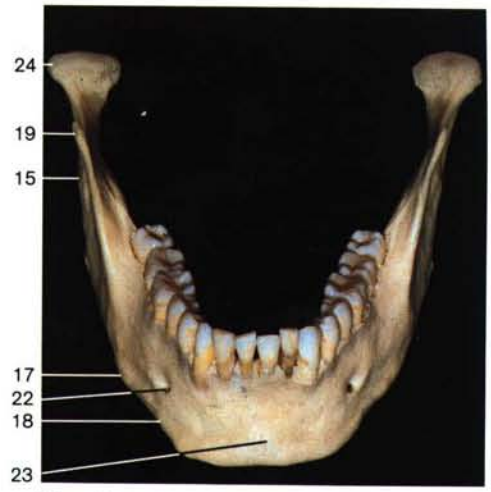

Mandible of the adult (anterior aspect).

Right half of mandible (medial aspect).

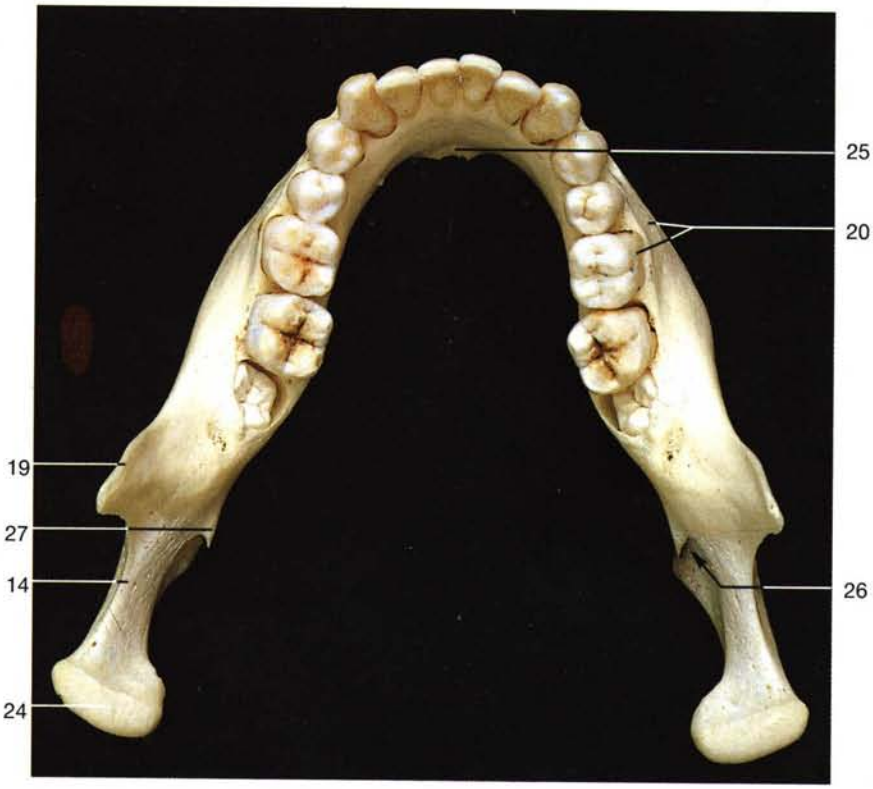

Mandible of the adult (superior aspect).

1 **Temporal bone**
2 Temporal fossa (greater wing of sphenoid bone)
3 Infratemporal crest
4 Infratemporal fossa
5 Zygomatic arch
6 **Frontal bone**

Mandible
13 Condylar process
14 Mandibular notch
15 Ramus of the mandible
16 Masseteric tuberosity
17 Angle of the mandible
18 Body of the mandible
19 Coronoid process
20 Alveolar process including teeth
21 Oblique line
22 Mental foramen
23 Mental protuberance
24 Head of the mandible
25 Genial tubercle or mental spine
26 Mandibular foramen (entrance to mandibular canal)
27 Lingula
28 Mylohyoid sulcus
29 Mylohyoid line
30 Submandibular fossa
31 Sublingual fossa

7 **Zygomatic bone** (frontal process)
8 **Lacrimal bone**
9 **Nasal bone**
10 Lacrimal groove
11 **Maxilla** (canine fossa)
12 Alveolar process of maxilla

Temporomandibular joint with ligaments.

1 Zygomatic arch
2 Articular capsule
3 External acoustic meatus
4 **Lateral ligament**
5 Mandibular notch
6 **Stylomandibular ligament**
7 Ramus of the mandible
8 Zygomatic bone
9 Coronoid process
10 Maxilla
11 Articular cartilage of condylar process
12 Styloid process
13 Mandibular fossa
14 **Articular disc**
15 Articular tubercle
16 Lateral pterygoid muscle
17 Condylar process of mandible

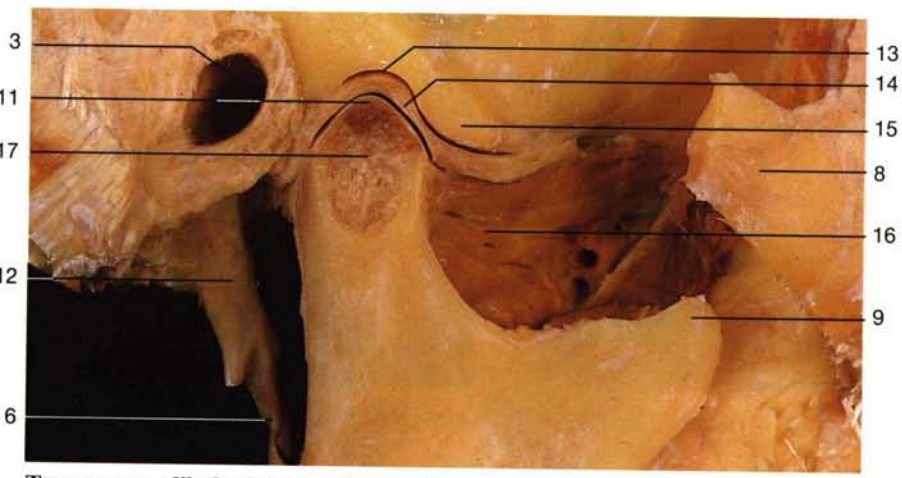

Temporomandibular joint, sagittal section.

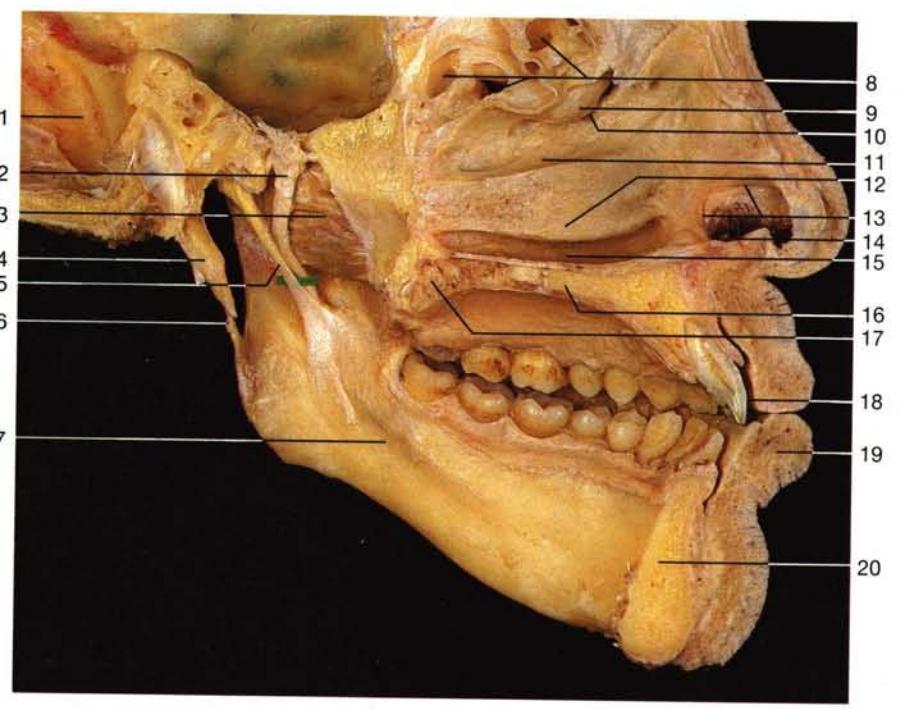

Ligaments of temporomandibular joint. Left half of the head (medial aspect).

1 Groove for sigmoid sinus
2 Mandibular nerve
3 Lateral pterygoid muscle
4 Styloid process
5 **Sphenomandibular ligament**
6 **Stylomandibular ligament**
7 Mylohyoid groove
8 **Ethmoidal air cells**
9 Ethmoidal bulla
10 **Hiatus semilunaris**
11 Middle meatus
12 **Inferior nasal concha**
13 Limen nasi
14 Vestibule with hairs
15 Inferior meatus
16 Hard palate
17 Soft palate
18 Vestibule of oral cavity
19 Lower lip
20 Mandible

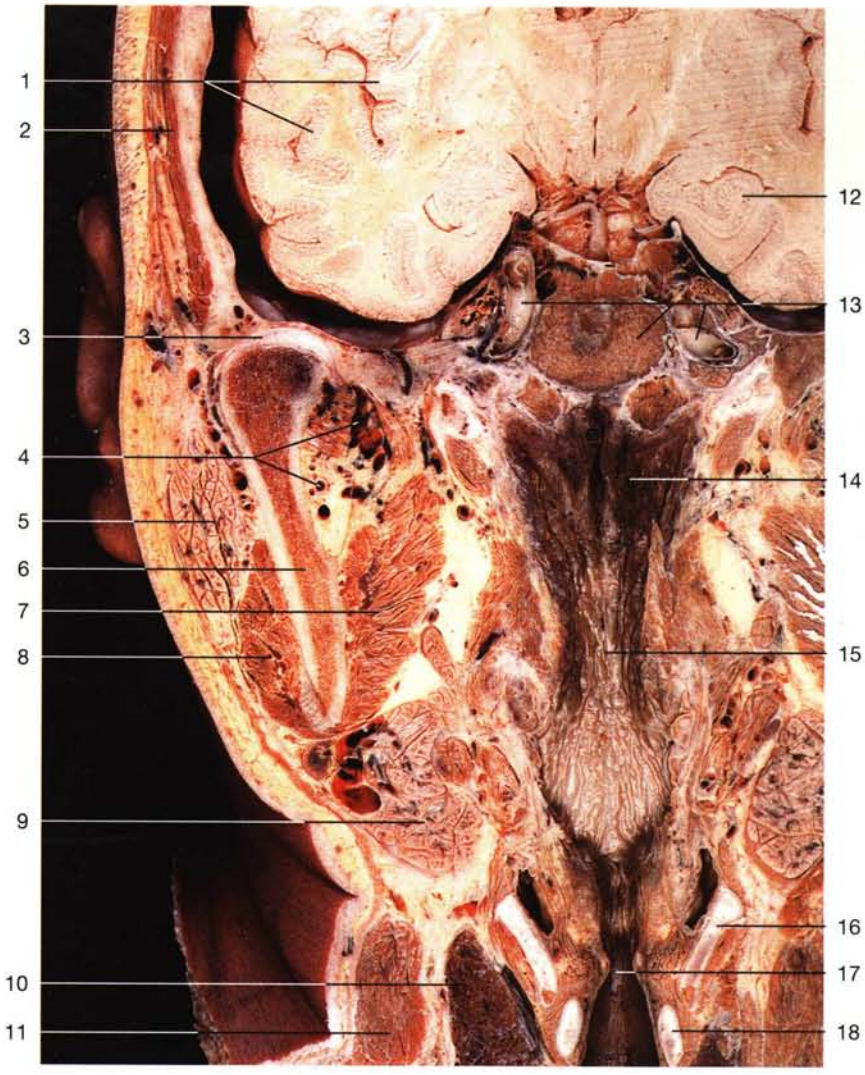

Coronal section through the head at the level of the temporomandibular joint (right side, anterior aspect).

1 Insular lobe and temporal lobe
2 Temporalis muscle
3 **Articular disc of temporomandibular joint**
4 Maxillary artery and pterygoid venous plexus
5 Parotid gland
6 **Mandible**
7 **Medial pterygoid muscle**
8 **Masseter muscle**
9 Submandibular gland
10 Thyroid gland
11 Sternocleidomastoid muscle
12 Hippocampus
13 Internal carotid artery and sphenoid bone
14 Pharyngeal tonsil
15 Pharynx
16 Thyroid cartilage
17 Rima glottidis
18 Cricoid cartilage
19 **Hard palate** and palatine glands
20 **Oral cavity**
21 Upper molar
22 **Oral vestibule**
23 Lower molar
24 Platysma muscle
25 Maxillary sinus
26 Superior longitudinal muscle of tongue
27 **Transverse muscle of tongue**
28 Buccinator muscle
29 Inferior longitudinal muscle of tongue
30 Sublingual gland
31 **Genioglossus muscle**
32 Mastoid process
33 Styloid process
34 **Stylomandibular ligament**
35 Articular capsule
36 **Lateral ligament**
37 Zygomatic arch
38 **Sphenomandibular ligament**
39 Mandibular foramen

Coronal section through the oral cavity

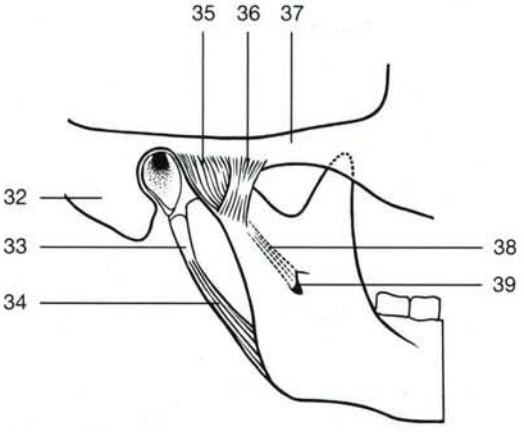

Ligaments related to the **temporomandibular joint**

Temporalis and masseter muscles. The temporal fascia has been removed, the temporomandibular joint severed and the zygomatic arch displayed.

1 Galea aponeurotica
2 **Temporalis muscle**
3 Occipital belly of occipitofrontalis muscle
4 **Temporomandibular joint**
5 External acoustic meatus
6 Deep layer of **masseter muscle**
7 Superficial layer of masseter muscle
8 Stylohyoid muscle
9 Posterior belly of **digastric muscle**
10 Internal jugular vein and external carotid artery
11 **Sternocleidomastoid muscle**
12 Frontal belly of occipitofrontalis muscle
13 Depressor supercilii muscle
14 **Orbicularis oculi muscle**
15 Transverse part of nasalis muscle
16 Levator labii superioris alaeque nasi muscle
17 Levator labii superioris muscle
18 Levator anguli oris muscle
19 **Zygomaticus major muscle**
20 **Orbicularis oris muscle**
21 **Buccinator muscle**
22 Depressor labii inferioris muscle
23 Depressor anguli oris muscle
24 **Submandibular gland**
25 Lateral nasal cartilage
26 Greater alar cartilage (lateral part)
27 Lesser alar cartilages
28 Greater alar cartilage (medial part)
29 Infraorbital nerve
30 Anterior belly of digastric muscle
31 Hypoglossal nerve and hyoglossus muscle
32 Superior thyroid artery
33 **Zygomatic arch**
34 Internal carotid artery
35 Common carotid artery

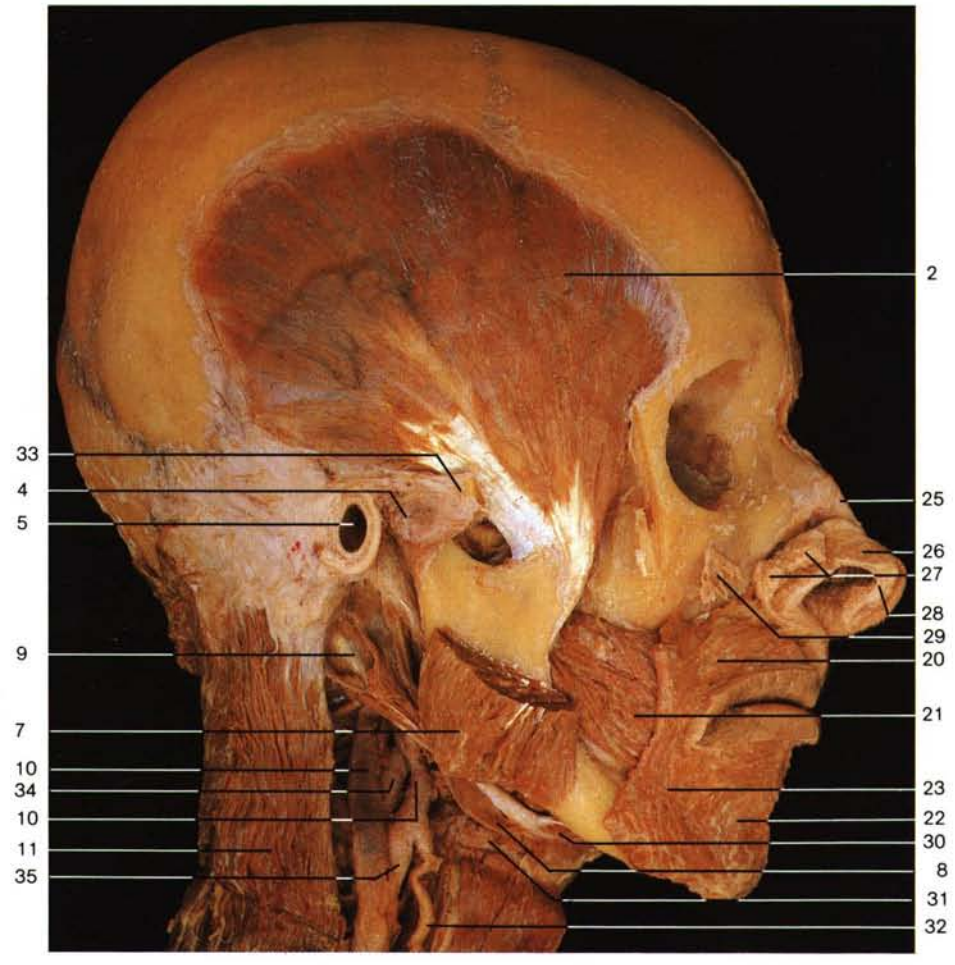

Temporalis muscle and temporomandibular joint. The zygomatic arch and the masseter muscle have been partially severed to display the insertion of the temporalis muscle.

1 Periosteum
2 **Temporalis muscle**
3 Zygomatic arch
4 Articular capsule of temporomandibular joint
5 **Lateral pterygoid muscle** (upper and lower head)
6 **Medial pterygoid muscle**
7 Styloglossus muscle
8 Stylohyoid muscle
9 Posterior belly of digastric muscle
10 **Masseter muscle** (severed)
11 Mandible
12 Sternocleidomastoid muscle
13 Orbicularis oculi muscle
14 Orbicularis oris muscle
15 Buccinator muscle
16 Depressor anguli oris muscle
17 Depressor labii inferioris muscle
18 Platysma muscle
19 **Articular disc of temporomandibular joint**
20 Head of mandible
21 Anterior belly of digastric muscle
22 Mylohyoid muscle
23 Hyoid bone

Medial and lateral pterygoid muscles. A portion of the mandible and the zygomatic arch has been removed revealing the pterygoid region or infratemporal fossa.

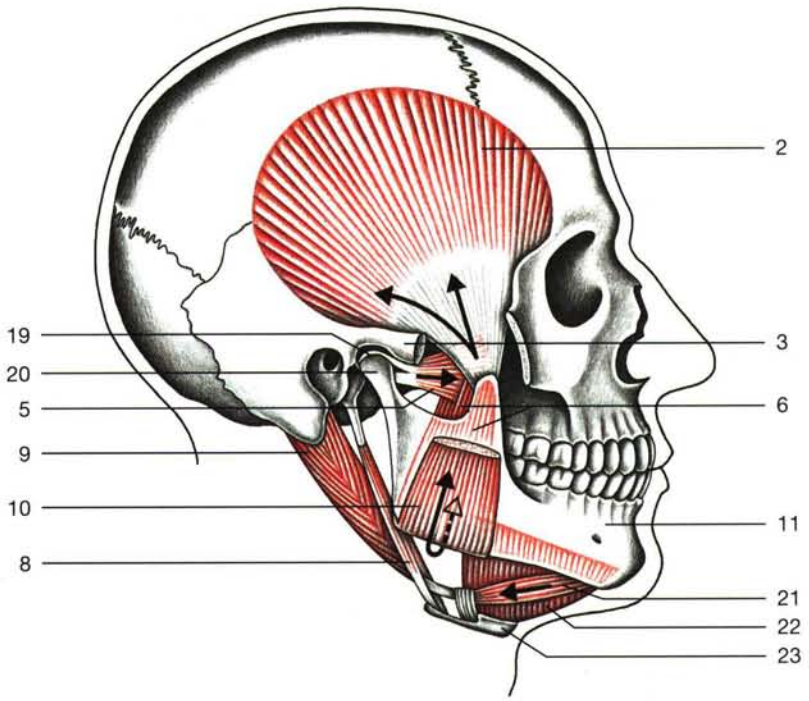

Effect of the muscles of mastication on the temporomandibular joint (arrows).

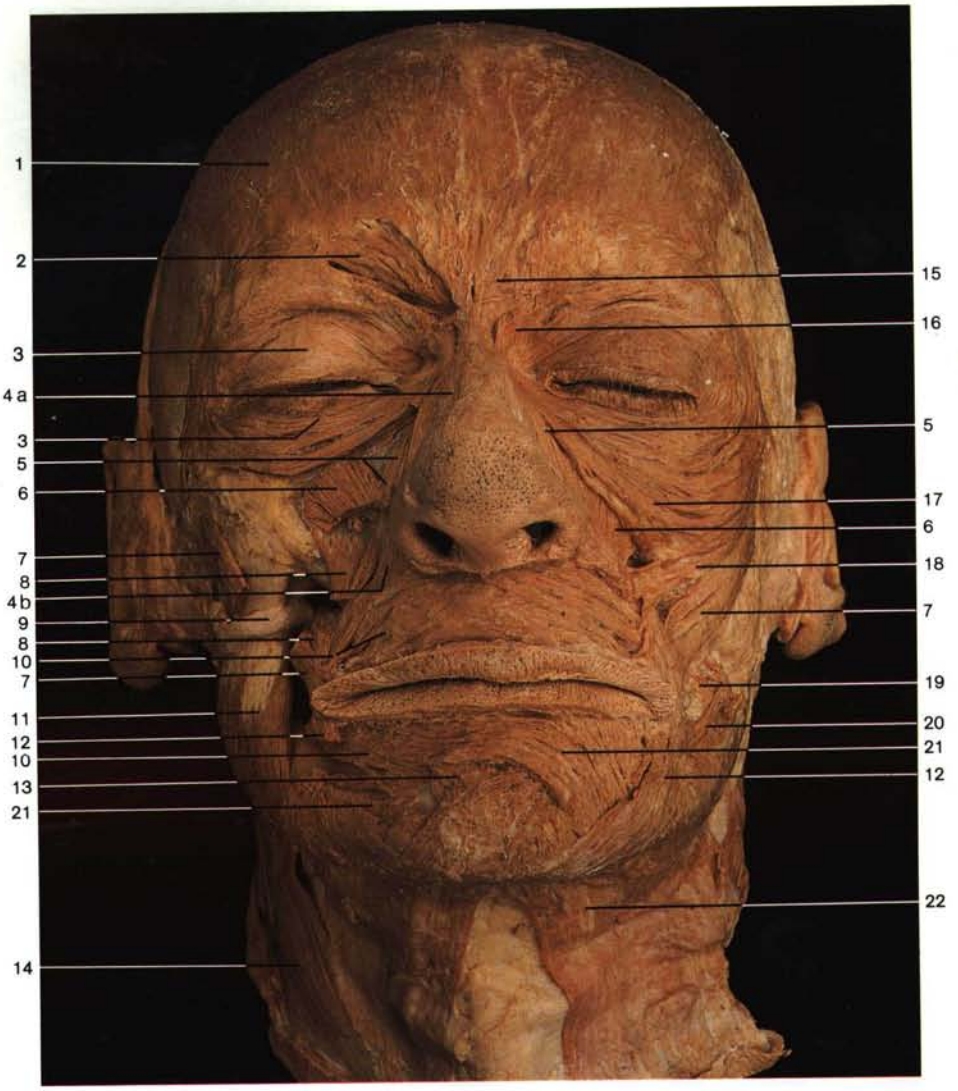

1 Frontal belly of **occipitofrontalis muscle**
2 **Corrugator supercilii muscle**
3 Palpebral part of orbicularis oculi muscle
4a Transverse part of nasalis muscle
4b Alar part of **nasalis muscle**
5 Levator labii superioris alaeque nasi muscle
6 **Levator labii superioris muscle**
7 **Zygomaticus major muscle**
8 Levator anguli oris muscle
9 Parotid duct
10 **Orbicularis oris muscle**
11 Masseter muscle
12 **Depressor anguli oris muscle**
13 **Mentalis muscle**
14 Sternocleidomastoid muscle
15 **Procerus muscle**
16 **Depressor supercilii muscle**
17 Orbital part of **orbicularis oculi muscle**
18 Zygomaticus minor muscle
19 **Buccinator muscle**
20 Risorius muscle
21 **Depressor labii inferioris muscle**
22 **Platysma muscle**
23 Galea aponeurotica
24 **Temporoparietalis muscle**
25 Occipital belly of occipitofrontalis muscle
26 Parotid gland with fascia
27 Temporal fascia
28 **Orbicularis oculi muscle**
29 Parotid duct, masseter muscle

Facial muscles (anterior aspect). Left side: superficial layer; right side: deeper layer.

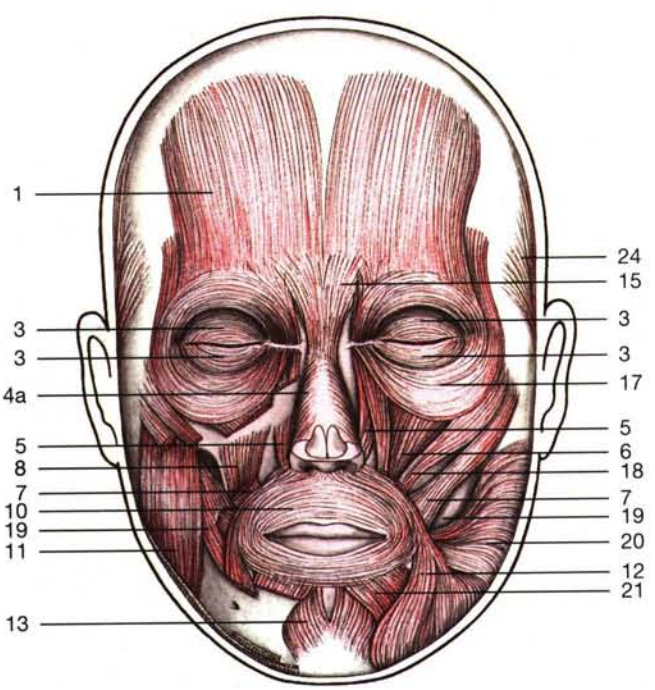

Facial muscles (schematic drawing). Left side: superficial layer; right side: deeper layer.

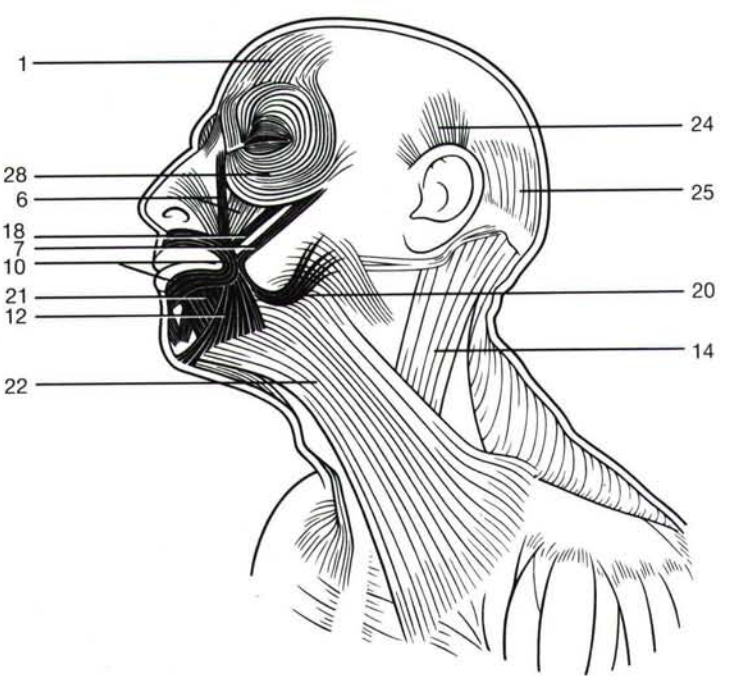

Facial muscles. Sphincter-like muscles surround the orifices of the head. Radially arranged muscles work as their antagonists.

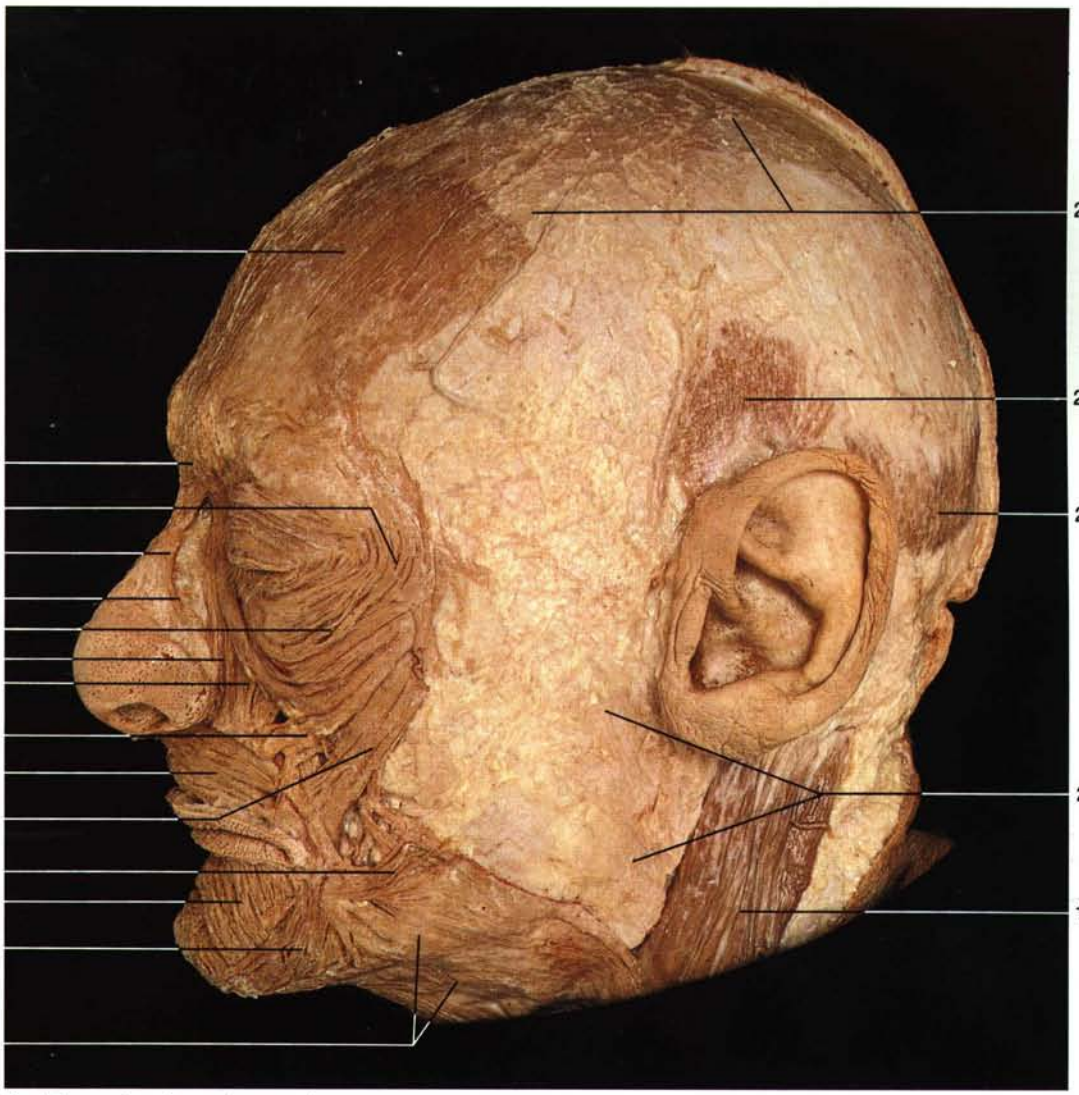

Facial muscles (lateral aspect).

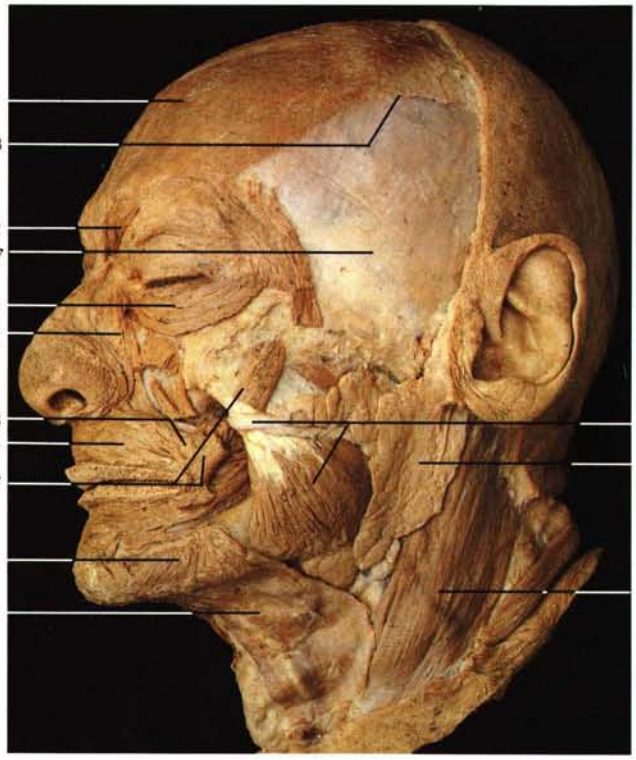

Facial muscles and parotid gland (lateral aspect).

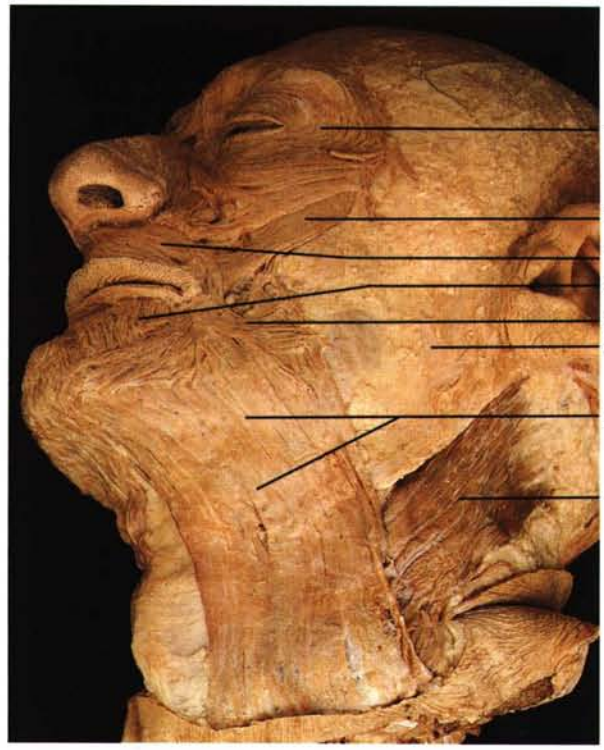

Platysma muscle (oblique lateral aspect). Superficial lamina of cervical fascia partly removed.

Supra- and infrahyoid muscles, pharynx I (lateral aspect). Ramus of mandible, pterygoid muscles and insertion of temporalis muscle removed.

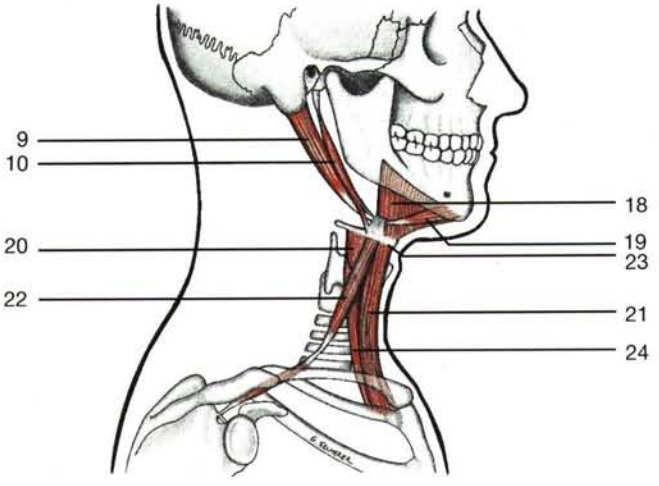

Supra- and infrahyoid muscles. (Schematic diagram.)

1 Galea aponeurotica
2 **Temporal fascia**
3 Tendon of temporalis muscle
4 Zygomatic arch
5 Lateral pterygoid plate
6 Tensor veli palatini muscle (styloid process)
7 Superior constrictor muscle of pharynx
8 **Styloglossus muscle**
9 Posterior belly of digastric muscle
10 **Stylohyoid muscle**
11 Longus capitis muscle
12 Sternocleidomastoid muscle (reflected)
13 Inferior constrictor of pharynx
14 Frontal belly of occipitofrontalis muscle
15 Orbital part of orbicularis oculi muscle
16 **Buccinator muscle**
17 Depressor anguli oris muscle
18 **Mylohyoid muscle**
19 Anterior belly of digastric muscle
20 **Thyrohyoid muscle**
21 **Sternohyoid muscle**
22 **Omohyoid muscle**
23 Hyoid bone
24 Sternothyroid muscle

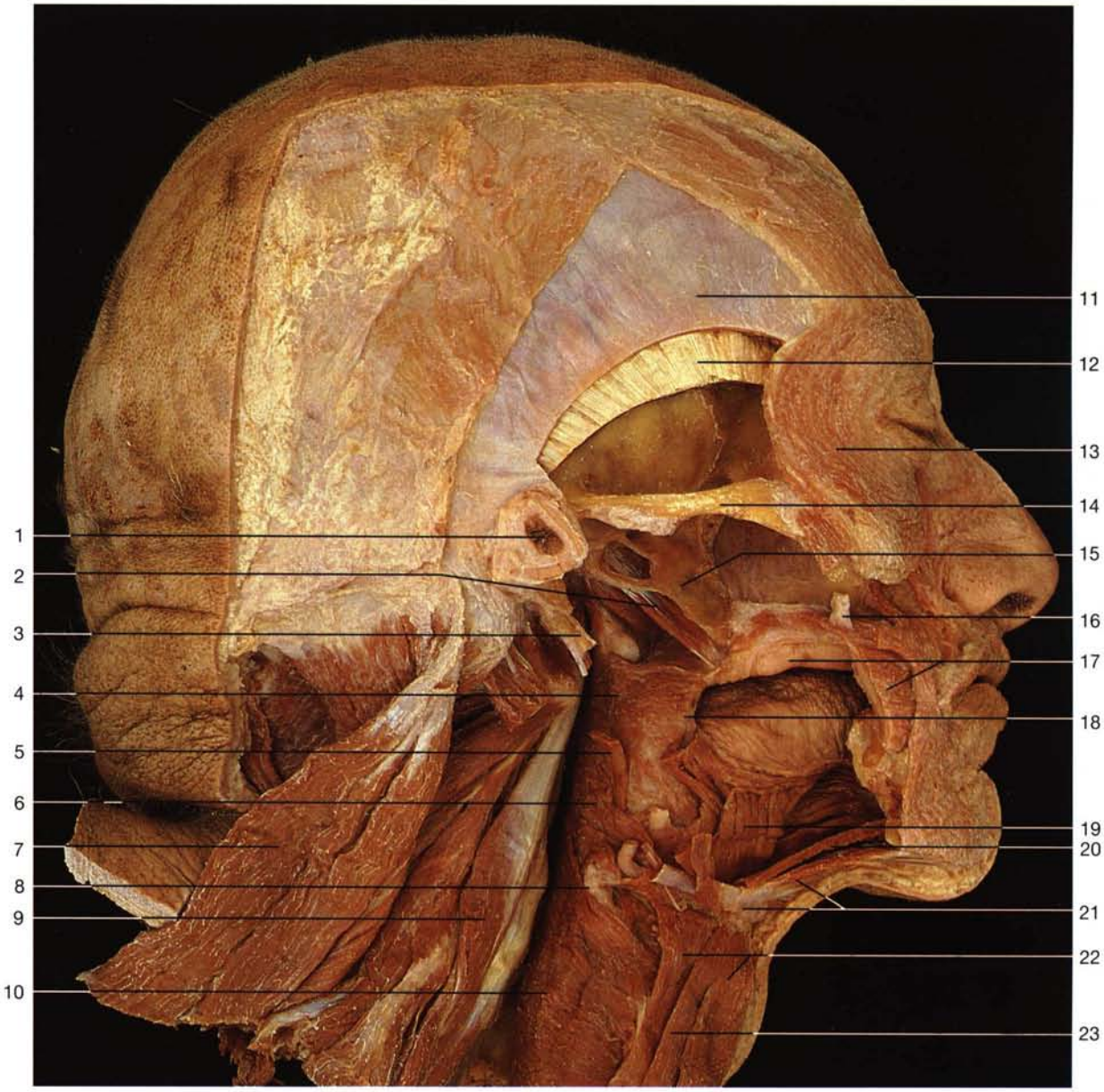

Supra- and infrahyoid muscles, pharynx II. Buccinator muscle removed; oral cavity opened.

1 External acoustic meatus
2 Tensor veli palatini muscle
3 Styloid process
4 **Superior constrictor muscle of pharynx**
5 Stylopharyngeus muscle (divided)
6 **Middle constrictor muscle of pharynx**
7 Sternocleidomastoid muscle
8 Greater horn of hyoid bone
9 Longus capitis
10 **Inferior constrictor muscle of pharynx**
11 Temporal fascia
12 Tendon of temporalis muscle
13 Orbicularis oculi muscle
14 Zygomatic arch
15 Lateral pterygoid plate
16 Parotid duct
17 Gingiva of upper jaw (without teeth), buccinator muscle (divided)
18 Pterygomandibular raphe
19 Hyoglossus muscle
20 Mylohyoid muscle
21 Anterior belly of digastric muscle (hyoid bone)
22 Sternohyoid and thyrohyoid muscles
23 Omohyoid muscle

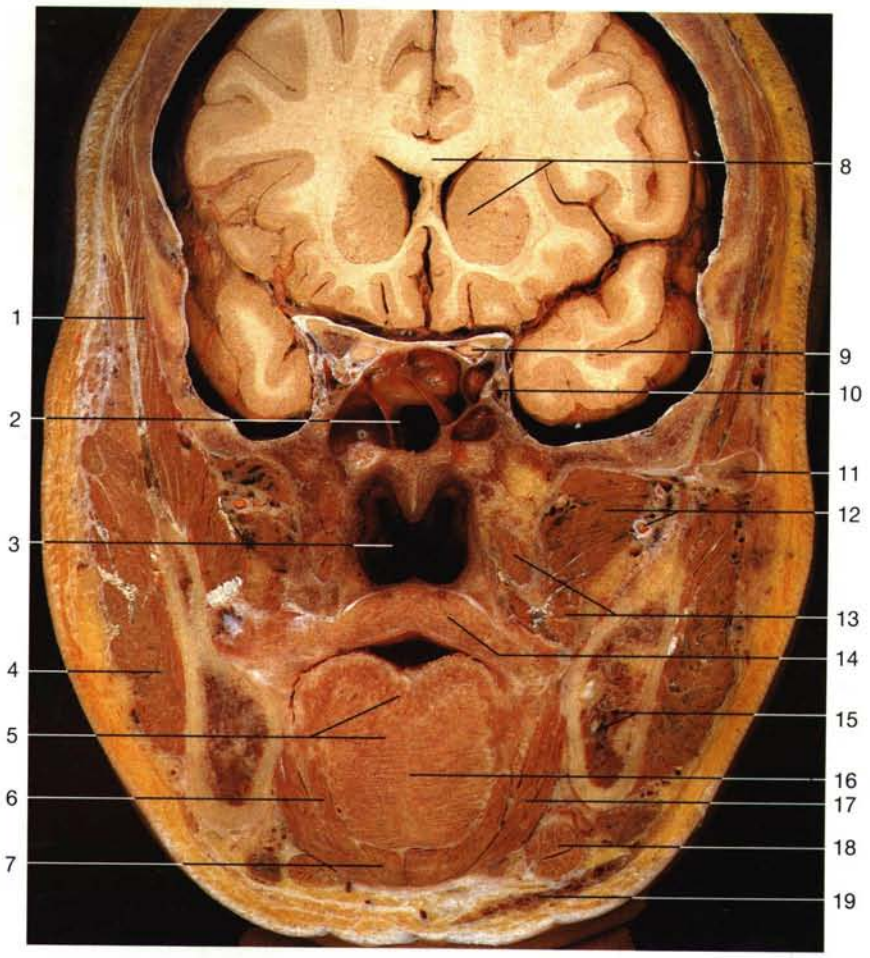

Coronal section through cranial, nasal and oral cavity at the level of sphenoidal sinus.

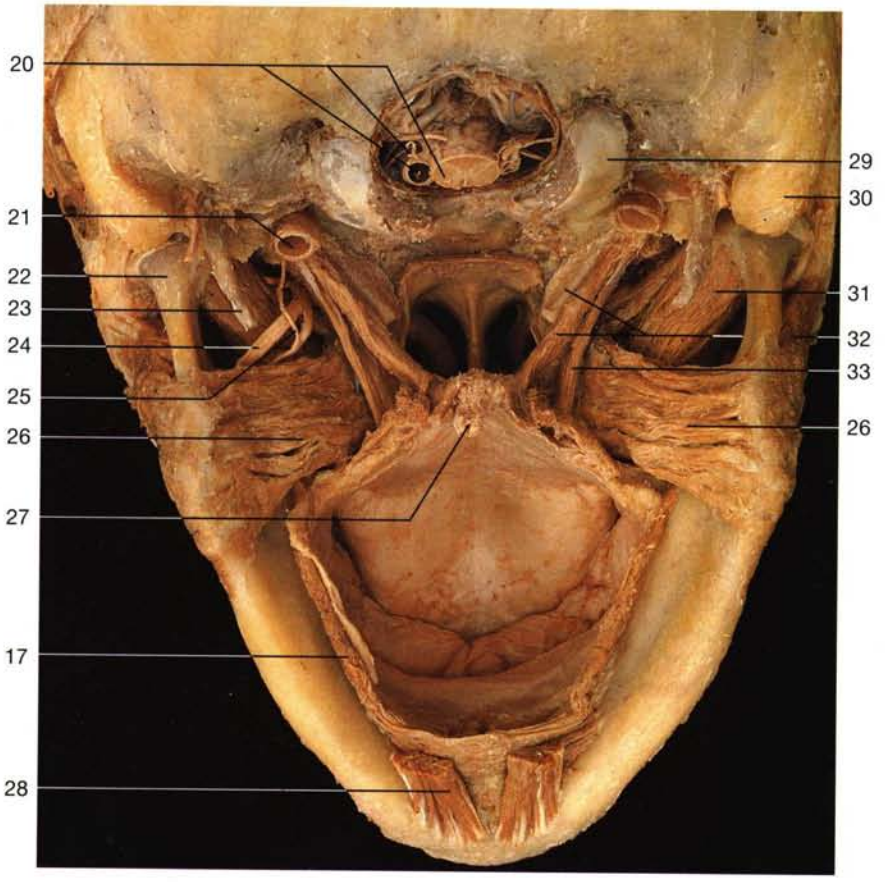

Pterygoid and palatine muscles (posterior aspect).

1 Temporalis muscle
2 **Sphenoidal sinus**
3 **Nasopharynx**
4 Masseter muscle
5 Superior longitudinal, transverse and vertical muscles of tongue
6 Hyoglossus muscle
7 Geniohyoid muscle
8 Corpus callosum (caudate nucleus)
9 Optic nerve
10 Cavernous sinus
11 Zygomatic arch
12 Cross section of lateral pterygoid muscle and maxillary artery
13 Section of medial pterygoid muscle
14 **Soft palate**
15 Mandible and inferior alveolar nerve
16 Septum of the tongue
17 **Mylohyoid muscle**
18 Submandibular gland
19 Platysma muscle
20 Foramen magnum, vertebral artery and spinal cord
21 Internal carotid artery
22 **Head of mandible**
23 Styloid process
24 Inferior alveolar nerve
25 Lingual nerve and chorda tympani nerve
26 **Medial pterygoid muscle**
27 Uvula
28 Anterior belly of digastric muscle (cut)
29 Condyle of occipital bone
30 Mastoid process
31 **Lateral pterygoid muscle**
32 Auditory tube and **levator veli palatini muscle**
33 **Tensor veli palatini muscle**

1 Superficial temporal artery and auriculotemporal nerve
2 Middle meningeal artery
3 **Maxillary artery**
4 Facial nerve (divided and reflected)
5 **External carotid artery**
6 Stylohyoid muscle
7 Digastric muscle (posterior belly)
8 Internal carotid artery and carotid sinus branch of glossopharyngeal nerve
9 Sternocleidomastoid muscle
10 Common carotid artery
11 Temporalis muscle
12 Pterygopalatine fossa
13 Posterior superior alveolar artery
14 Deep temporal artery
15 Infraorbital artery
16 Buccal nerve
17 Inferior alveolar artery and nerve
18 **Facial artery**
19 Submental artery
20 Hyoid bone
21 Superior thyroid artery (divided)

Dissection of maxillary artery. Mandible and lateral pterygoid muscle partly removed.

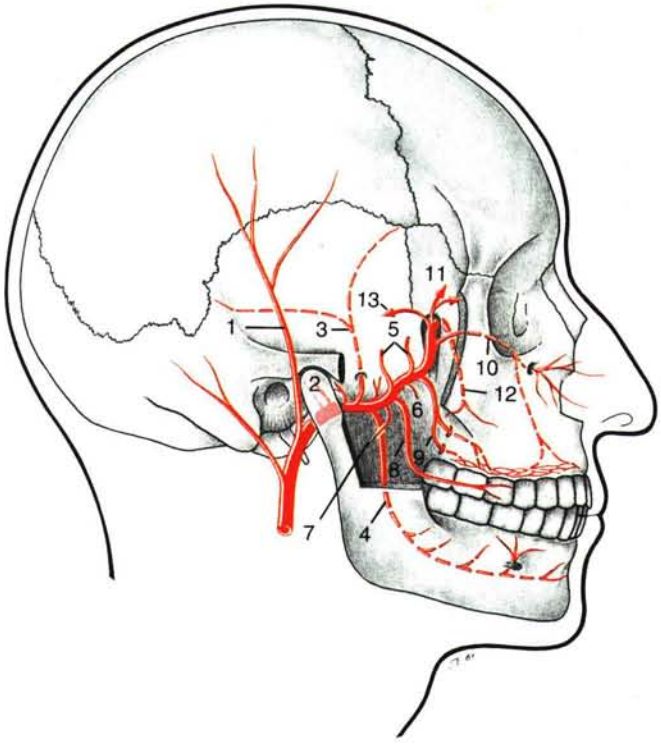

1 Superficial temporal artery

Branches of the first part
2 Deep auricular artery and anterior tympanic artery
3 Middle meningeal artery
4 Inferior alveolar artery

Branches of the second part
5 Deep temporal branches
6 Pterygoid branches
7 Masseteric artery
8 Buccal artery

Branches of the third part
9 Posterior superior alveolar artery
10 Infraorbital artery
11 Sphenopalatine artery and branches to the nasal cavity
12 Descending palatine artery
13 Artery of the pterygoid canal

Main branches of maxillary artery. (Schematic drawing.)

Base of the skull with cranial nerves (internal aspect). Both cerebral hemispheres and upper part of the brain stem removed. Incision on the right tentorium cerebelli to display the cranial nerves of the infratentorial space.

1 Superior sagittal sinus with falx cerebri
2 Olfactory bulb
3 **Olfactory tract**
4 **Optic nerve** and internal carotid artery
5 Anterior clinoid process and anterior attachment of tentorium cerebelli
6 **Oculomotor nerve** (n. III)
7 **Abducens nerve** (n. VI)
8 Tentorial notch (incisura tentorii)
9 **Trochlear nerve** (n. IV)
10 Tentorium cerebelli
11 Falx cerebri and confluence of sinuses
12 Hypophysial fossa, infundibulum, and diaphragma sellae
13 Dorsum sellae
14 Midbrain (divided)
15 **Trigeminal nerve** (n. V)
16 **Facial nerve** (n. VII), nervus intermedius, and **vestibulocochlear nerve** (n. VIII)
17 Cerebral aqueduct
18 Right hemisphere of cerebellum
19 Vermis of cerebellum
20 Straight sinus

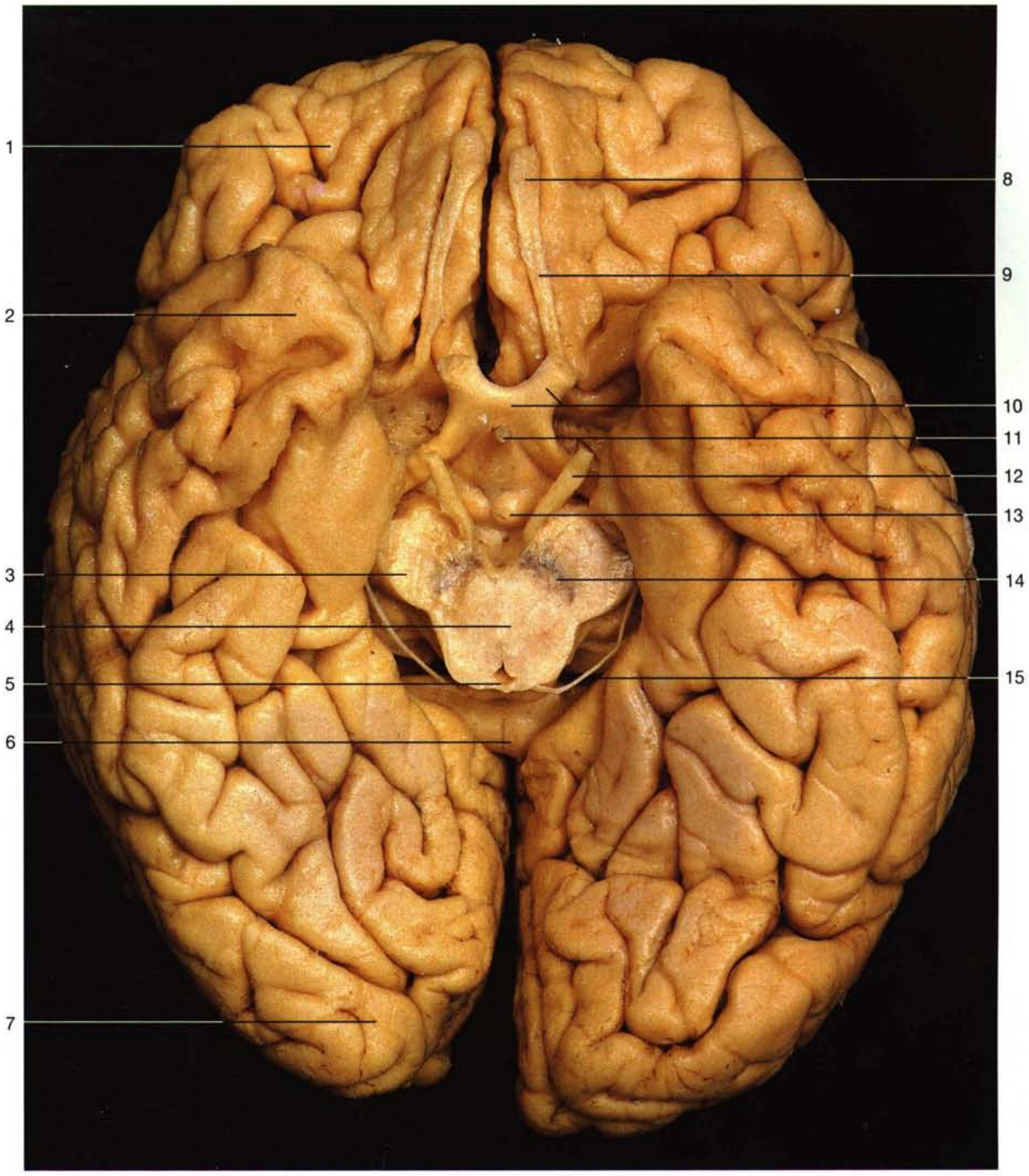

Inferior aspect of the brain with cranial nerves. Midbrain divided.

1 **Frontal lobe**
2 **Temporal lobe**
3 Pedunculus cerebri
4 **Midbrain** (divided)
5 Cerebral aqueduct
6 Splenium of corpus callosum
7 **Occipital lobe**
8 Olfactory bulb
9 Olfactory tract
10 **Optic nerve and optic chiasma**
11 Infundibulum
12 **Oculomotor nerve** (n. III)
13 Mamillary body
14 Substantia nigra
15 **Trochlear nerve** (n. IV)

Cranial nerves	
I = Olfactory nerves	VII = Facial nerve
II = Optic nerve	VIII = Vestibulocochlear nerve
III = Oculomotor nerve	IX = Glossopharyngeal nerve
IV = Trochlear nerve	X = Vagus nerve
V = Trigeminal nerve	XI = Accessory nerve
VI = Abducens nerve	XII = Hypoglossal nerve

Cranial nerves. Brain (inferior aspect).

1 Olfactory sulcus (termination)
2 Orbital gyri
3 Temporal lobe
4 Straight gyrus
5 Olfactory trigone and inferior temporal sulcus
6 Medial occipitotemporal gyrus
7 Parahippocampal gyrus, mamillary body, and interpeduncular fossa
8 Pons and cerebral peduncle
9 **Abducens nerve** (n. VI)
10 Pyramid
11 Lower part of olive
12 Cervical spinal nerves
13 Cerebellum
14 Tonsil of cerebellum
15 Occipital lobe (posterior pole)
16 Olfactory bulb
17 Orbital sulci of frontal lobe
18 **Olfactory tract**
19 **Optic nerve** (n. II) and anterior perforated substance
20 Optic chiasma
21 Optic tract
22 **Oculomotor nerve** (n. III)
23 **Trochlear nerve** (n. IV)
24 **Trigeminal nerve** (n. V)
25 **Facial nerve** (n. VII)
26 **Vestibulocochlear nerve** (n. VIII)
27 Flocculus of cerebellum
28 **Glossopharyngeal** (n. IX) and **vagus nerve** (n. X)
29 **Hypoglossal nerve** (n. XII)
30 **Accessory nerve** (n. XI)
31 Vermis of cerebellum
32 Longitudinal fissure

Brain stem and pharynx with cranial nerves (posterior aspect). Cranial cavity opened and cerebellum removed.

1 Falx cerebri
2 Occipital lobe
3 Straight sinus
4 Tentorium cerebelli
5 Transverse sinus
6 Inferior colliculus of midbrain
7 **Rhomboid fossa**
8 Medulla oblongata
9 Posterior belly of digastric muscle
10 Internal carotid artery
11 Pharynx (middle constrictor muscle)
12 Hyoid bone (greater horn)
13 **Trochlear nerve** (n. IV)
14 **Facial nerve** (n. VII), **vestibulocochlear nerve** (n. VIII)
15 **Glossopharyngeal nerve** (n. IX) and vagus nerve (n. X)
16 **Accessory nerve** (intracranial portion) (n. XI)
17 **Hypoglossal nerve** (intracranial portion) (n. XII)
18 **Accessory nerve** (n. XI)
19 **Hypoglossal nerve** (n. XII)
20 **Vagus nerve** (n. X) and internal carotid artery
21 External carotid artery
22 **Sympathetic trunk** and superior cervical ganglion
23 Ansa cervicalis (superior root of hypoglossal nerve)
24 **Glossopharyngeal nerve** (n. IX) and stylopharyngeus muscle

Cranial nerves of the orbit and pterygopalatine fossa. Left orbit (lateral aspect). Note the zygomaticolacrimal anastomosis (arrow).

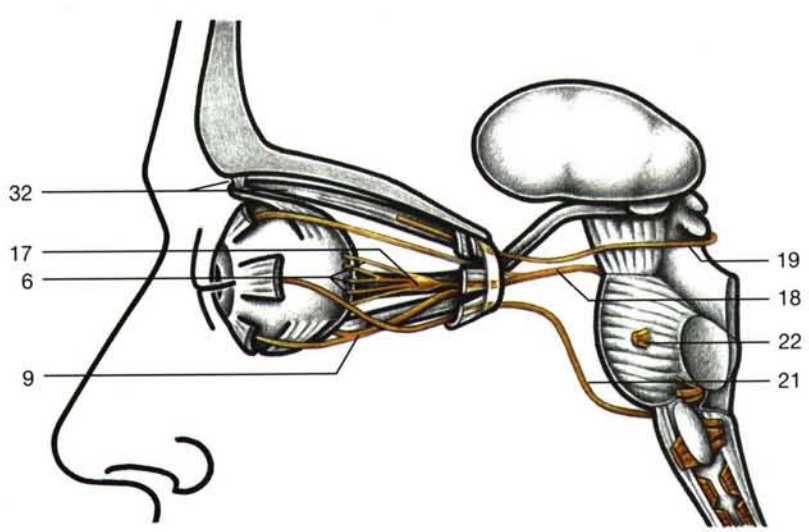

Cranial nerves innervating extraocular muscles (lateral aspect). (Schematic drawing.)

1 Frontal lobe
2 Supraorbital nerve
3 Lacrimal gland
4 Lacrimal nerve
5 Lateral rectus muscle (divided)
6 **Optic nerve** and short ciliary nerves
7 Inferior oblique muscle
8 Zygomatic nerve
9 Inferior branch of oculomotor nerve and inferior rectus muscle
10 Infraorbital nerve
11 Posterior superior alveolar nerves
12 Branches of superior alveolar plexus adjacent to mucous membrane of maxillary sinus
13 Central sulcus of insula
14 Superior rectus muscle
15 Periorbita (roof of orbit)
16 Nasociliary nerve
17 Ciliary ganglion
18 **Oculomotor nerve** (n. III)
19 **Trochlear nerve** (n. IV)
20 **Ophthalmic nerve** (n. V_1)
21 **Abducens nerve** (n. VI) (divided)
22 **Trigeminal nerve** (n. V)
23 Trigeminal ganglion
24 **Maxillary nerve** (n. V_2) and foramen rotundum
25 **Mandibular nerve** (n. V_3)
26 External acoustic meatus
27 Pterygopalatine nerves
28 Deep temporal nerves
29 Buccal nerve
30 Masseteric nerve
31 Auriculotemporal nerve
32 Trochlea and superior oblique muscle

Cranial nerves of the orbit (superior aspect). Right side: superficial layer; left side: middle layer of the orbit (superior rectus muscle and frontal nerve divided and reflected). Tentorium and dura mater partly removed.

Cranial nerves at the skull base. The brain stem was divided and the tentorium fenestrated. Both hemispheres were removed.

1 Frontal sinus (enlarged)
2 Frontal nerve (divided and reflected)
3 Superior rectus muscle (divided) and eyeball
4 Superior oblique muscle
5 Short ciliary nerves and optic nerve (n. II)
6 Nasociliary nerve
7 **Abducens nerve** (n. VI) and lateral rectus muscle
8 Ciliary ganglion and superior rectus muscle (reflected)
9 **Oculomotor nerve** (n. III)
10 **Trochlear nerve** (n. IV)
11 Crus cerebri and midbrain
12 Inferior wall of the third ventricle connected with cerebral aqueduct
13 Lateral and medial branch of supraorbital nerve
14 Supratrochlear nerve
15 Superior rectus muscle
16 Lacrimal nerve
17 Frontal nerve
18 **Ophthalmic nerve** (n. V1)
19 Optic chiasma and internal carotid artery
20 Trigeminal ganglion
21 **Trigeminal nerve** (n. V)
22 Tentorial notch
23 Falx cerebri
24 Cerebellum
25 Infundibulum
26 Olfactory tract

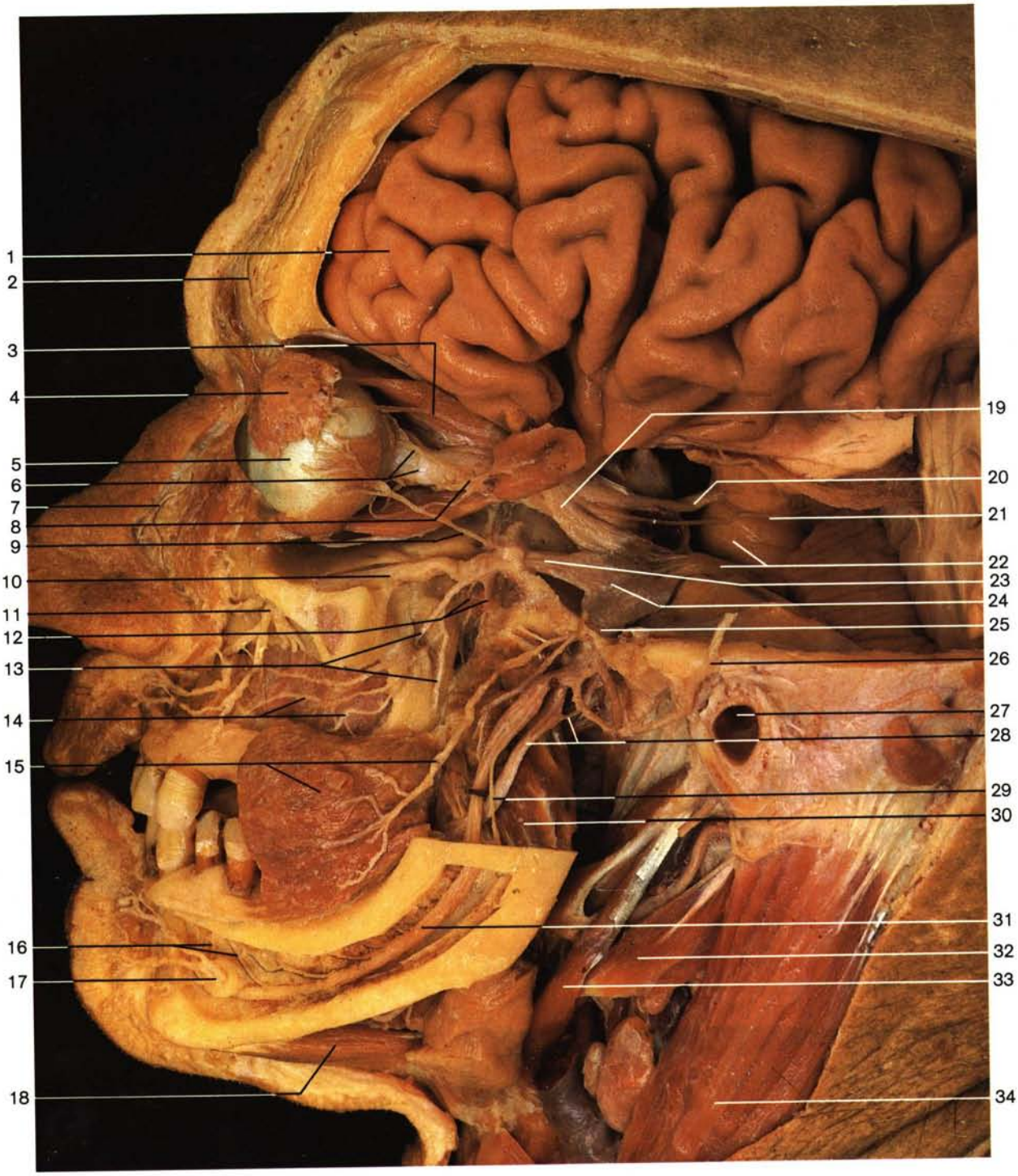

Dissection of the trigeminal nerve in its entirety. Lateral wall of cranial cavity, lateral wall of orbit, zygomatic arch and ramus of the mandible have been removed and the mandibular canal opened.

1 Frontal lobe of cerebrum
2 Supraorbital nerve
3 Lacrimal nerve
4 Lacrimal gland
5 Eyeball
6 **Optic nerve** and short ciliary nerves
7 External nasal branch of anterior ethmoidal nerve
8 Ciliary ganglion
9 Zygomatic nerve
10 Infraorbital nerve
11 Infraorbital foramen and terminal branches of infraorbital nerve
12 Pterygopalatine ganglion and pterygopalatine nerves
13 Posterior superior alveolar nerves
14 Superior dental plexus
15 Buccinator muscle and buccal nerve
16 Inferior dental plexus
17 Mental foramen and mental nerve
18 Anterior belly of digastric muscle
19 **Ophthalmic nerve** (n. V_1)
20 **Oculomotor nerve** (n. III)
21 **Trochlear nerve** (n. IV)
22 Trigeminal nerve and pons
23 **Maxillary nerve** (n. V_2)
24 Trigeminal ganglion
25 **Mandibular nerve** (n. V_3)
26 Auriculotemporal nerve
27 External acoustic meatus (divided)
28 Lingual nerve and chorda tympani
29 Mylohyoid nerve
30 Medial pterygoid muscle
31 Inferior alveolar nerve
32 Posterior belly of digastric muscle
33 Stylohyoid muscle
34 Sternocleidomastoid muscle

Cranial nerves in connection with the brain stem. Left side (lateral superior aspect). Left half of brain and head partly removed. Notice the location of trigeminal ganglion.

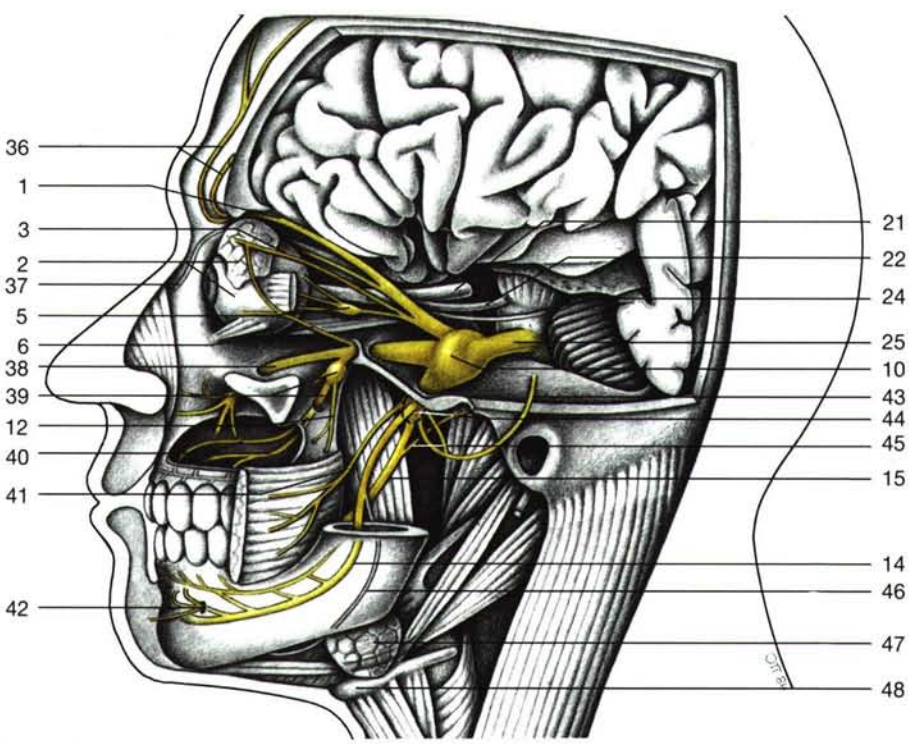

Main branches of trigeminal nerve. (Schematic drawing of figure on opposite page.)

1 Frontal nerve
2 Lacrimal gland and eyeball
3 Lacrimal nerve
4 Lateral rectus muscle
5 **Ciliary ganglion** lateral to optic nerve
6 Zygomatic nerve
7 Inferior branch of oculomotor nerve
8 **Ophthalmic nerve** (n. V_1)
9 **Maxillary nerve** (n. V_2)
10 **Trigeminal ganglion**
11 **Mandibular nerve** (n. V_3)
12 Posterior superior alveolar nerves
13 Tympanic cavity, external acoustic meatus, and tympanic membrane
14 Inferior alveolar nerve
15 Lingual nerve
16 **Facial nerve** (n. VII)
17 **Vagus nerve** (n. X)
18 **Hypoglossal nerve** (n. XII) and superior root of ansa cervicalis
19 External carotid artery
20 **Olfactory tract** (n. I)
21 **Optic nerve** (n. II) (intracranial part)
22 **Oculomotor nerve** (n. III)
23 **Abducens nerve** (n. VI)
24 **Trochlear nerve** (n. IV)
25 **Trigeminal nerve** (n. V)
26 **Vestibulocochlear nerve** (n. VIII) and **facial nerve** (VII)
27 **Glossopharyngeal nerve** (n. IX) (leaving brain stem)
28 Rhomboid fossa
29 **Vagus nerve** (n. X) (leaving brain stem)
30 **Hypoglossal nerve** (n. XII) (leaving medulla oblongata)
31 **Accessory nerve** (n. XI) (ascending from foramen magnum)
32 Vertebral artery
33 Spinal ganglion and dura mater of spinal cord
34 Accessory nerve (n. XI)
35 Internal carotid artery
36 Lateral and medial branch of supraorbital nerve
37 Infratrochlear nerve
38 Infraorbital nerve
39 **Pterygopalatine ganglion** and middle superior alveolar nerve
40 Middle superior alveolar nerves (entering superior dental plexus)
41 Buccal nerve
42 Mental nerve and mental foramen
43 Auriculotemporal nerve
44 Otic ganglion (dotted line)
45 Chorda tympani
46 Mylohyoid nerve
47 Submandibular gland
48 Hyoid bone

Dissection of facial nerve in its entirety. Cranial cavity fenestrated; temporal lobe partly removed. Facial canal and tympanic cavity opened, posterior wall of external acoustic meatus removed. **Branches of facial nerve:** a = temporal branch; b = zygomatic branches; c = buccal branches; d = marginal mandibular branch.

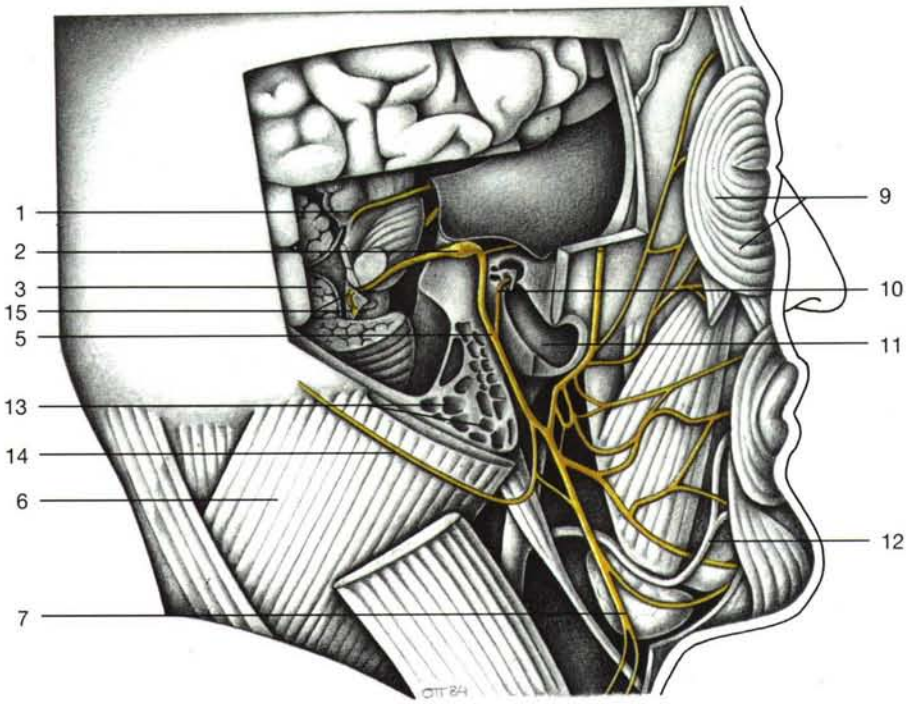

Facial nerve. (Schematic drawing of the dissection above.)

1 Trochlear nerve
2 Facial nerve with geniculate ganglion
3 Cerebellum (right hemisphere)
4 Occipital belly of occipitofrontalis and greater occipital nerve
5 Facial nerve at stylomastoid foramen
6 Splenius capitis muscle
7 Cervical branch of facial nerve
8 Sternocleidomastoid muscle and retromandibular vein
9 Orbicularis oculi muscle
10 Chorda tympani
11 External acoustic meatus
12 Facial artery
13 Mastoid air cells
14 Posterior auricular nerve
15 Nucleus and genu of facial nerve

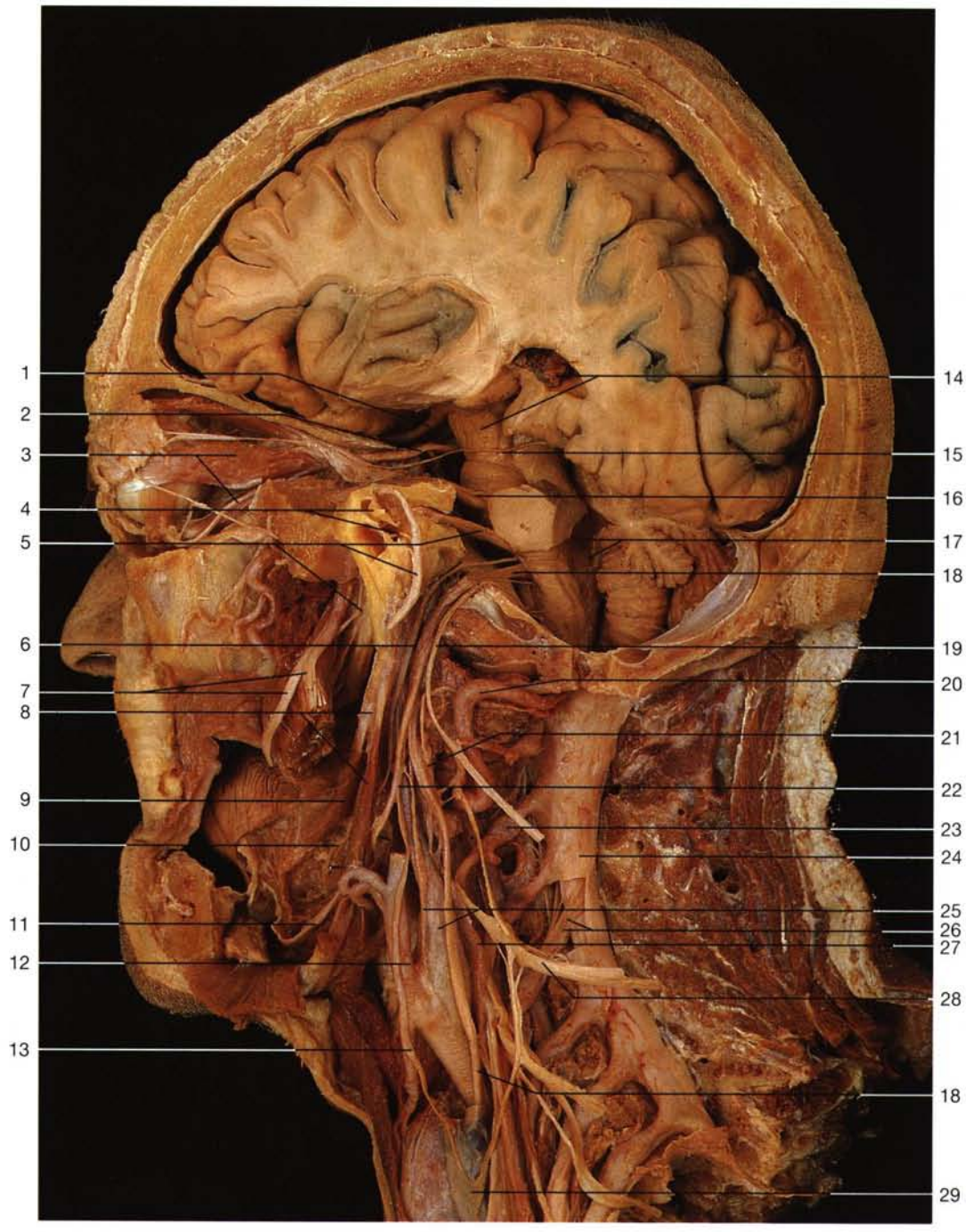

Cranial nerves in connection with the brain stem (oblique-lateral aspect). Lateral portion of the skull, brain, neck and facial structures, lateral wall of orbit and oral cavity have been removed. The tympanic cavity has been opened. The mandible has been divided and the muscles of mastication have been removed.

1 Optic tract
2 **Oculomotor nerve** (n. III)
3 Lateral rectus muscle and inferior branch of oculomotor nerve
4 Malleus and chorda tympani
5 Chorda tympani, **facial nerve** (n. VII), and **vestibulocochlear nerve** (n. VIII)
6 **Glossopharyngeal nerve** (n. XI)
7 Lingual nerve and inferior alveolar nerve
8 Styloid process and stylohyoid muscle
9 Styloglossus muscle
10 Lingual branches of glossopharyngeal nerve
11 Lingual branch of hypoglossal nerve
12 External carotid artery
13 Superior root of ansa cervicalis (branch of hypoglossal nerve)
14 Lateral ventricle with choroid plexus and cerebral peduncle
15 **Trochlear nerve** (n. IV)
16 **Trigeminal nerve** (n. V)
17 Fourth ventricle and rhomboid fossa
18 **Vagus nerve** (n. X)
19 **Accessory nerve** (n. XI)
20 Vertebral artery
21 Superior cervical ganglion
22 **Hypoglossal nerve** (n. XII)
23 Spinal ganglion with dural sheath
24 Dura mater of spinal cord
25 Internal carotid artery and carotid sinus branch of glossopharyngeal nerve
26 Dorsal roots of spinal nerve
27 Sympathetic trunk
28 Branch of cervical plexus (ventral primary ramus of third cervical spinal nerve)
29 Ansa cervicalis

Lateral superficial aspect of the face. Peripheral distribution of facial nerve. a–d = branches of facial nerve: a = temporal branch; b = zygomatic branches; c = buccal branches; d = marginal mandibular branch.

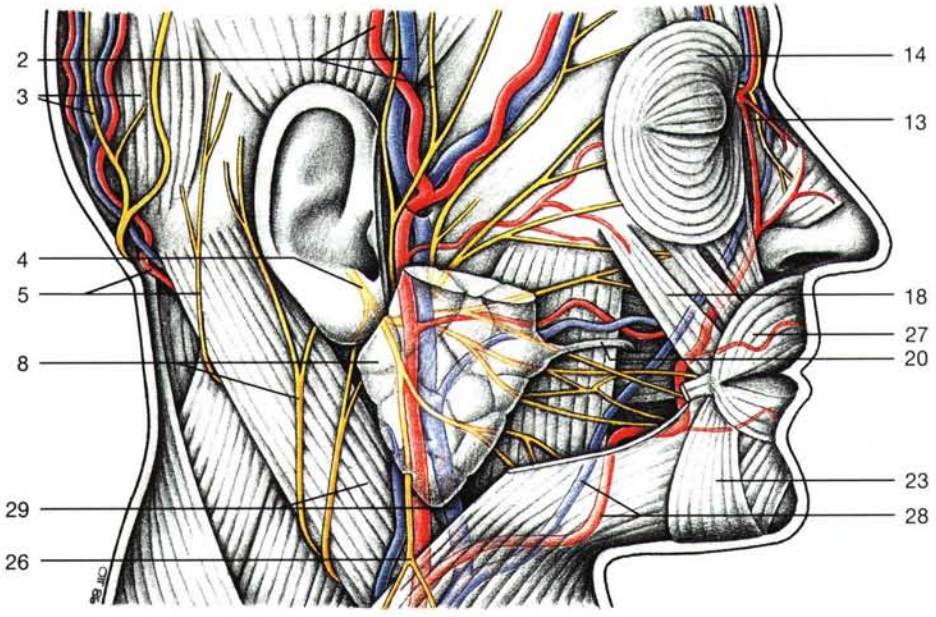

Superficial region of the face. Note the facial plexus within the parotid gland. (Semischematic drawing.)

1 Temporoparietalis muscle
2 Parietal branch of **superficial temporal artery** and **vein** and **auriculotemporal nerve**
3 Occipital belly of occipitofrontalis muscle and greater occipital nerve
4 **Facial nerve**
5 Lesser occipital nerve and **occipital artery**
6 Transverse facial artery
7 Masseter muscle
8 **Parotid gland** and great auricular nerve
9 Sternocleidomastoid muscle and external jugular vein
10 Splenius capitis muscle
11 Branches of cervical plexus
12 Trapezius muscle
13 Angular artery (terminal portion of facial artery)
14 Orbicularis oculi muscle
15 Levator labii superioris alaeque nasi muscle
16 **Facial artery** and zygomaticus minor muscle
17 Levator anguli oris muscle
18 Zygomaticus major muscle
19 Orbicularis oris muscle and superior labial artery
20 Parotid duct
21 Risorius muscle and inferior labial artery

1 Medial branch of supraorbital nerve
2 Nasalis muscle
3 Levator labii superioris alaeque nasi muscle
4 **Orbicularis oculi muscle**
5 Levator labii superioris muscle
6 **Facial artery and vein**
7 Zygomaticus minor and major muscle
8 Transverse facial artery
9 **Orbicularis oris muscle**
10 Buccal nerves, depressor labii inferioris muscle, **facial artery** and **vein**
11 **Parotid gland** and duct and masseter muscle
12 Depressor anguli oris muscle
13 Transverse cervical nerve
14 External jugular vein
15 **Platysma muscle**
16 Supraclavicular nerve
17 Galea aponeurotica
18 Lateral branches of the supraorbital nerve
19 Frontal belly of occipitofrontalis muscle and branches of superficial temporal vein and artery
20 Superficial temporal artery and vein
21 **Auriculotemporal nerve**
22 Zygomatico-orbital artery and temporoparietalis muscle
23 Lesser occipital nerve
24 Occipital belly of occipitofrontalis muscle
25 Occipital vein and occipital artery
26 **Great auricular nerve** and sternocleidomastoid muscle
27 Trapezius muscle
28 Greater petrosal nerve
29 Geniculate ganglion
30 Chorda tympani
31 Posterior auricular nerve
32 Stylomastoid foramen

Superficial layer of the head and neck. Parotid fascia removed. Branches of **facial nerve:** a = temporal branch; b = zygomatic branches; c = buccal branches; d = marginal mandibular branch.

◁ **to page 76**
22 Depressor labii inferioris muscle
23 Depressor anguli oris muscle
24 Platysma muscle
25 Terminal branches of transverse cervical nerve
26 Cervical branch of facial nerve
27 Orbicularis oris muscle
28 **Facial artery** and **vein**
29 Sternocleidomastoid muscle and retromandibular vein

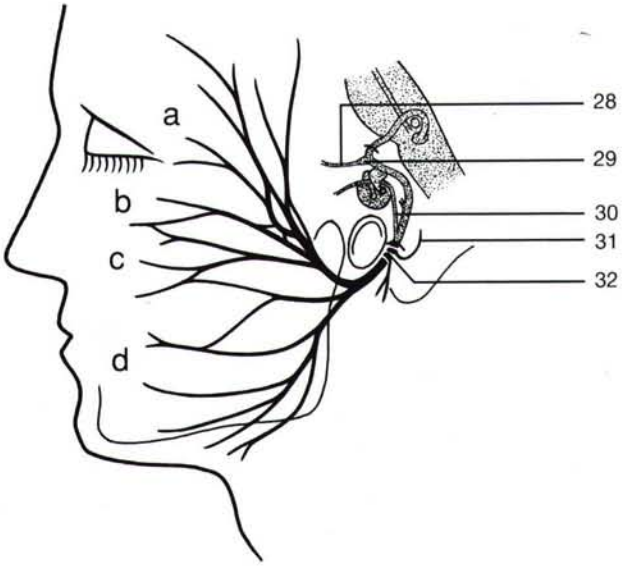

Facial nerve. Main branches. (Schematic diagram.)

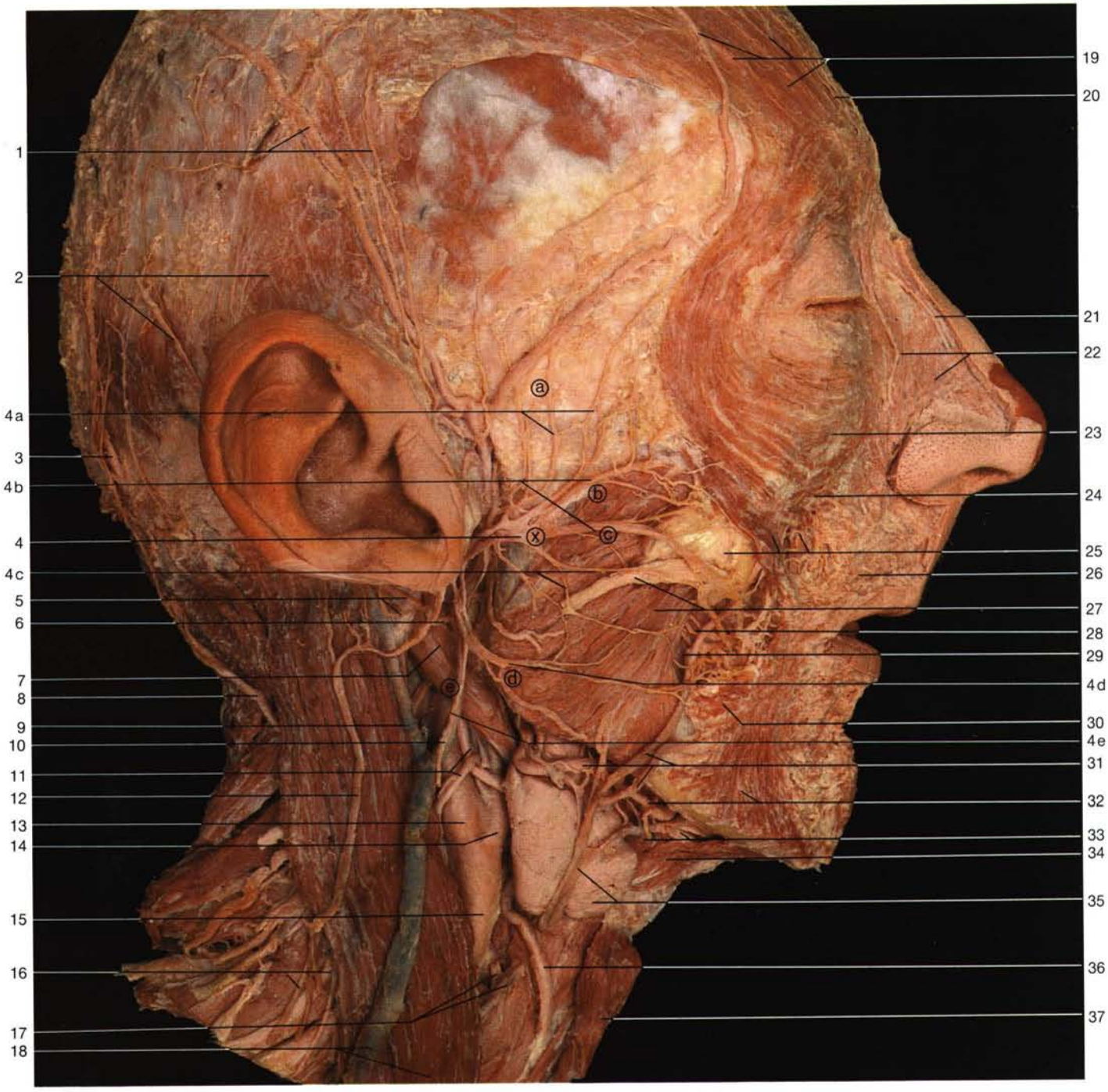

Lateral superficial aspect of the face. The parotid gland has been removed to display the parotid plexus of the facial nerve. a–e = branches of facial nerve: a = temporal branch; b = zygomatic branches; c = buccal branches; d = marginal mandibular branch; e = cervical branch.

1 Superficial temporal artery and auriculotemporal nerve
2 Posterior auricular artery and nerve and temporoparietalis muscle
3 Occipital artery
4 **Facial nerve** (n. VII) (**parotid plexus** [×])
  a Temporal branches
  b Zygomatic branches
  c Buccal branches
  d Marginal mandibular branch
  e Cervical branch
5 Posterior auricular nerve
6 Posterior auricular artery
7 Digastric muscle (posterior belly)
8 **Lesser occipital nerve**
9 Posterior auricular vein
10 Retromandibular vein
11 **Hypoglossal nerve** and sternocleidomastoid artery
12 **Great auricular nerve**
13 Internal carotid artery
14 External carotid artery
15 Common carotid artery
16 Branches of cervical plexus
17 Superior laryngeal artery and vein
18 External jugular vein and sternocleidomastoid muscle
19 Frontal branch of superficial temporal artery, lateral branch of supraorbital nerve, and frontal belly of occipitofrontalis muscle
20 Medial branch of supraorbital nerve
21 Dorsal nasal artery
22 Angular artery and nasalis muscle
23 **Orbicularis oculi muscle**
24 Zygomaticus minor muscle
25 Buccal fat pad, zygomaticus major muscle, and infraorbital nerve
26 **Orbicularis oris muscle**
27 **Parotid duct** and **masseter muscle**
28 Buccal artery and nerve
29 Buccinator muscle
30 Risorius muscle
31 **Facial artery** and **vein**
32 Submental artery and depressor anguli oris muscle
33 Mylohyoid nerve and mylohyoid muscle
34 Digastric muscle (anterior belly)
35 Facial vein and submandibular gland
36 **Superior thyroid artery**
37 Sternohyoid muscle

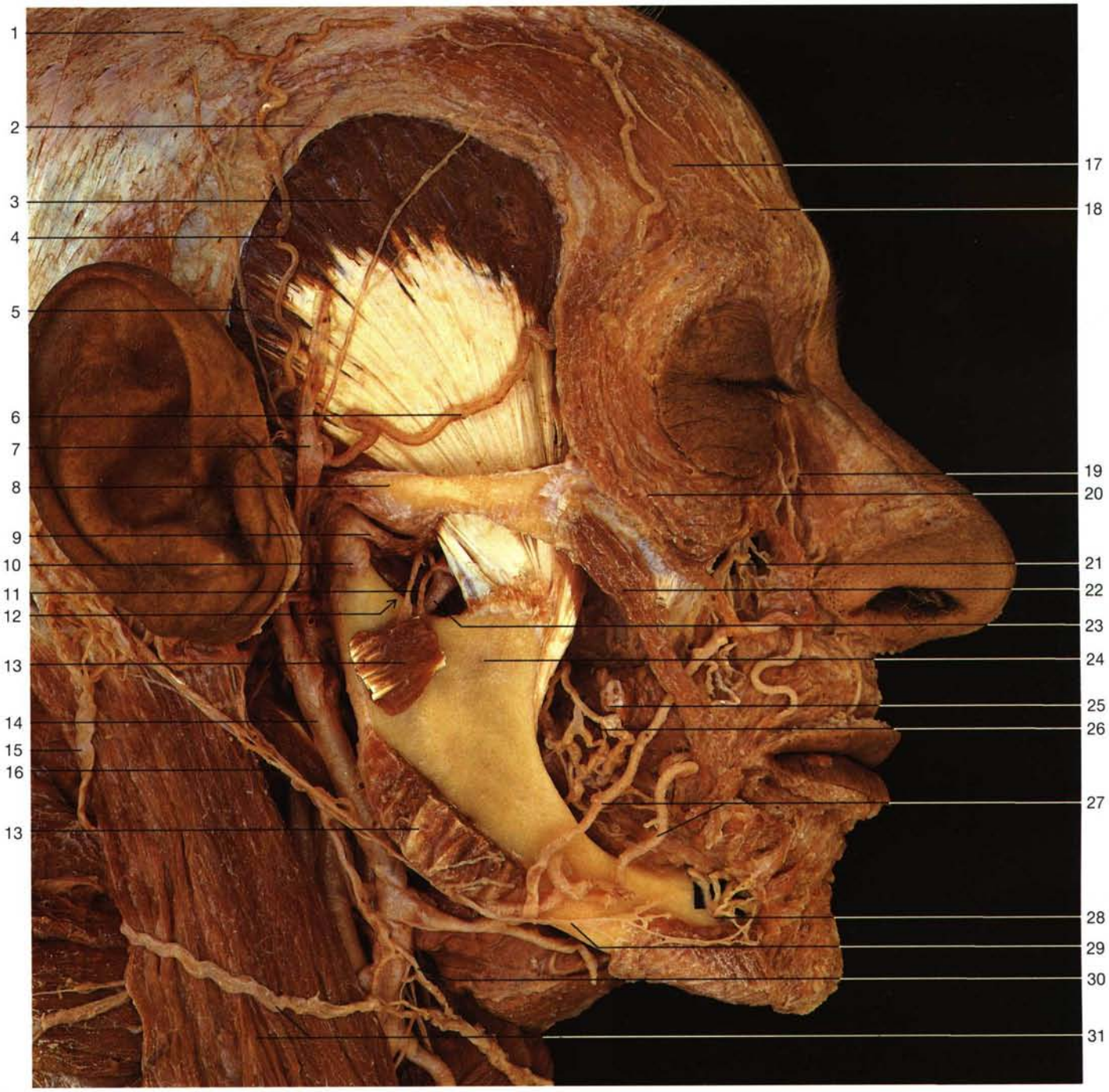

Lateral superficial aspect of the face. Masseter muscle and temporal fascia have been partly removed to display the masseteric artery and nerve.

1 Galea aponeurotica
2 Temporal fascia
3 **Temporalis muscle**
4 Parietal branch of **superficial temporal artery**
5 **Auriculotemporal nerve**
6 Frontal branch of superficial temporal artery
7 Superficial temporal vein
8 **Zygomatic arch**
9 **Articular disc of temporomandibular joint**
10 **Head of mandible**
11 Masseteric artery and nerve
12 Mandibular notch
13 Masseter muscle (divided)
14 **External carotid artery**
15 Great auricular nerve
16 Facial nerve (reflected)
17 Frontal belly of occipitofrontalis muscle
18 Medial branch of supraorbital nerve
19 **Angular artery**
20 Orbicularis oculi muscle
21 **Infraorbital nerve**
22 Zygomaticus major muscle
23 **Maxillar artery**
24 **Coronoid process**
25 Parotid duct (divided)
26 Buccal nerve
27 **Facial artery** and **vein**
28 **Mental nerve**
29 Mandibular branch of facial nerve
30 Cervical branch of facial nerve
31 Transverse cervical nerve (communicating branch with facial nerve) and sternocleidomastoid muscle

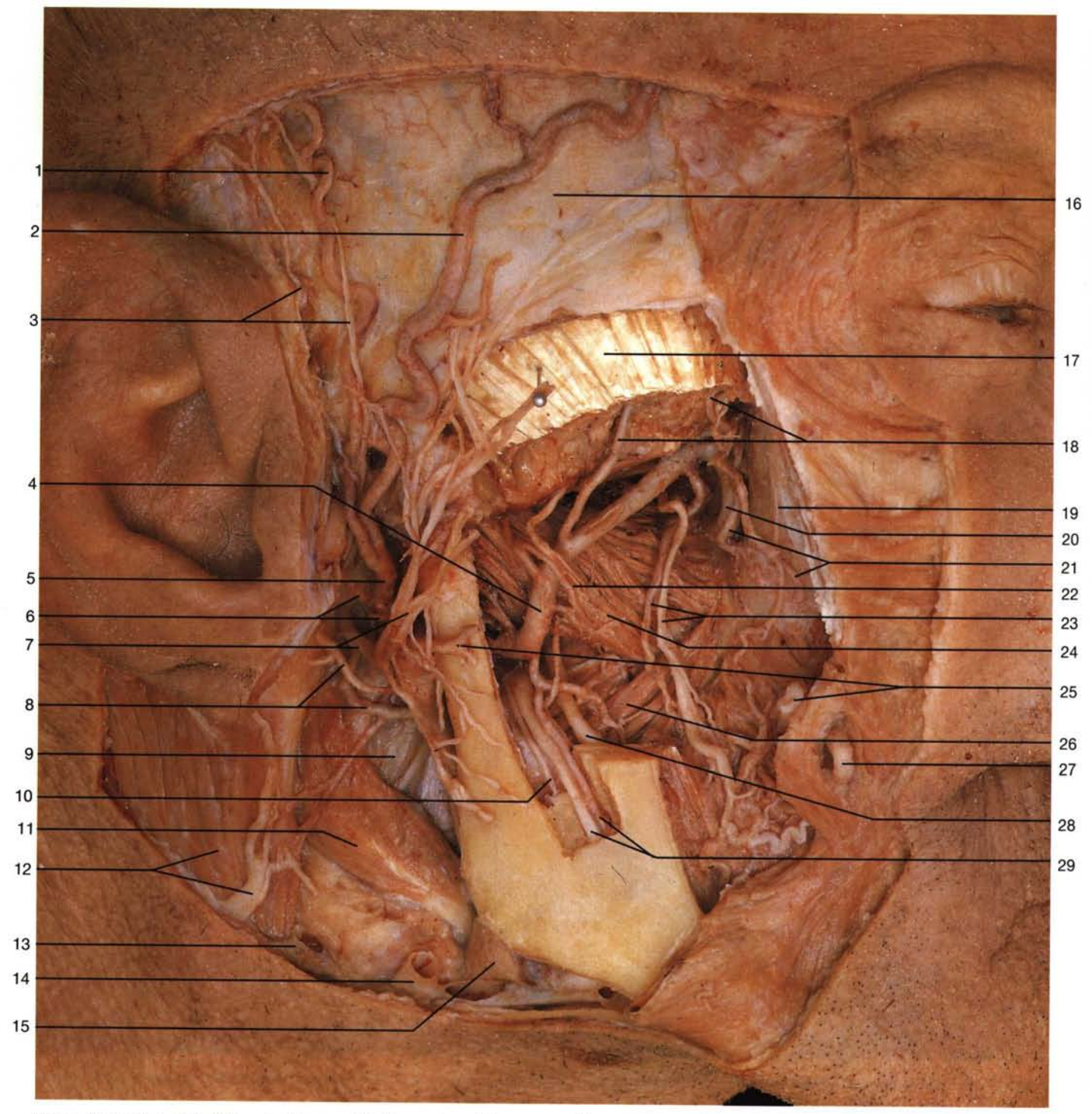

Deep dissection of facial and retromandibular regions. The coronoid process together with the insertions of temporalis muscle have been removed to display the maxillary artery. The upper part of the mandibular canal has been opened.

1 Parietal branch of the superficial temporal artery
2 Frontal branch of the superficial temporal artery
3 Auriculotemporal nerve
4 **Maxillary artery**
5 **Superficial temporal artery**
6 Communicating branches between facial and auriculotemporal nerves
7 **Facial nerve**
8 Posterior auricular artery and anterior auricular branch of superficial temporal artery
9 Internal jugular vein
10 Mylohyoid nerve
11 Posterior belly of digastric muscle
12 Great auricular nerve and sternocleidomastoid
13 External jugular vein
14 Retromandibular vein
15 Submandibular gland
16 Temporal fascia
17 **Temporalis tendon**
18 Deep temporal arteries
19 Posterior superior alveolar nerve
20 Sphenopalatine artery
21 Posterior superior alveolar arteries
22 Masseteric artery and nerve
23 **Buccal nerve and artery**
24 **Lateral pterygoid**
25 Transverse facial artery and parotid duct (divided)
26 Medial pterygoid
27 **Facial artery**
28 **Lingual nerve**
29 **Inferior alveolar artery and nerve** (mandibular canal opened)

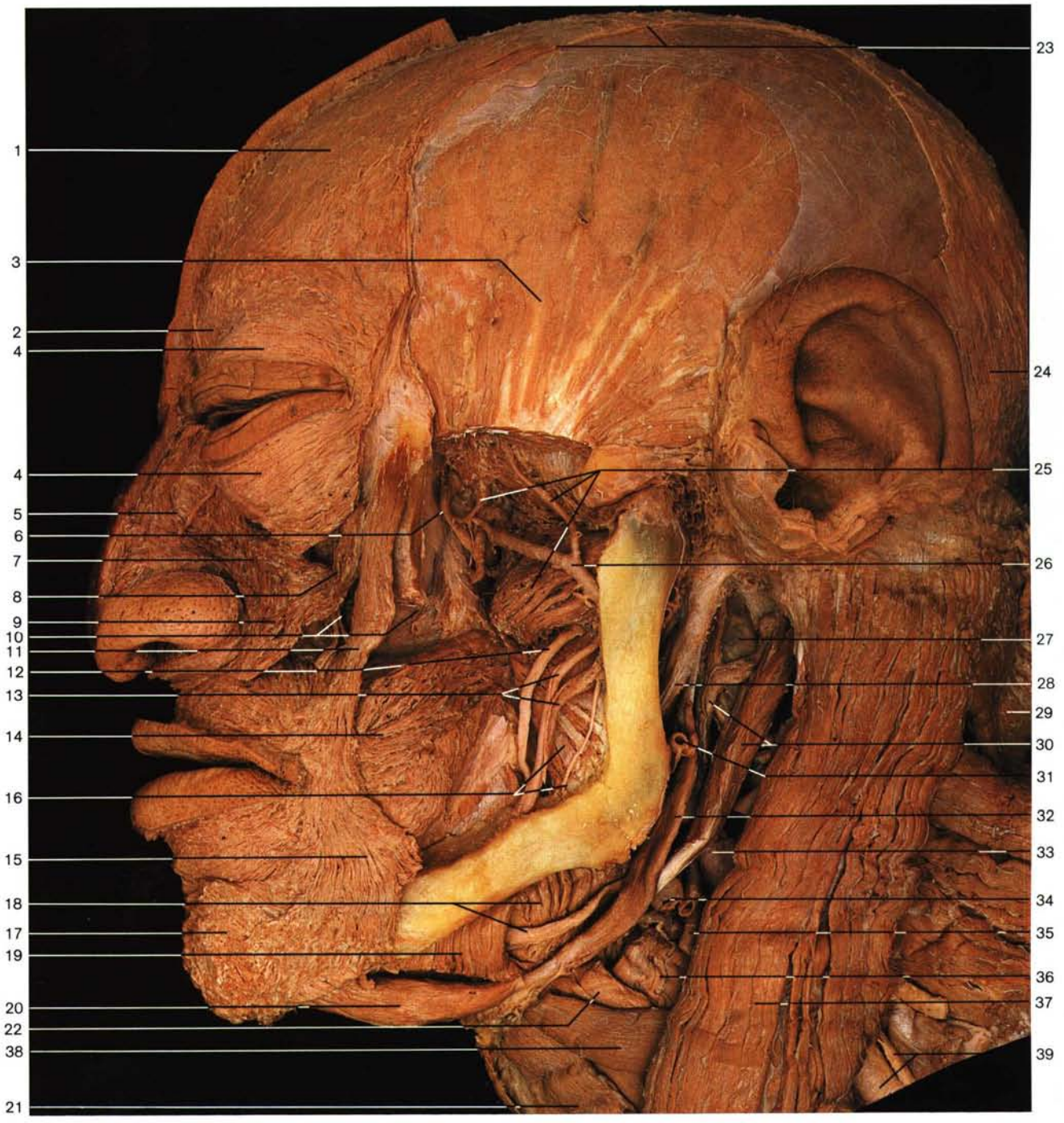

Retromandibular region with maxillary artery and branches of trigeminal nerve (n. V).

1 Frontal belly of occipitofrontalis muscle
2 Depressor supercilii muscle
3 Temporalis muscle
4 **Orbicularis oculi muscle**
5 Nasalis muscle
6 Infraorbital artery
7 Levator labii superioris alaeque nasi muscle
8 Zygomaticus minor muscle
9 Levator labii superioris muscle
10 Infraorbital artery and nerve and posterior superior alveolar artery
11 **Zygomaticus major muscle**
12 **Lingual nerve**
13 **Inferior alveolar artery and nerve**
14 Buccinator muscle
15 Depressor anguli oris muscle
16 **Medial pterygoid muscle** and **mylohyoid nerve**
17 Depressor labii inferioris muscle
18 **Hypoglossal nerve** and **hyoglossus muscle**
19 Mylohyoid muscle
20 Anterior belly of diagastric muscle
21 Sternohyoid muscle
22 Thyrohyoid muscle
23 Galea aponeurotica
24 Occipital belly of occipitofrontalis muscle
25 **Lateral pterygoid muscle** and deep temporal artery
26 **Maxillary artery**
27 Internal jugular vein
28 Styloglossus muscle
29 Splenius capitis muscle
30 Posterior belly of diagastric muscle and occipital artery
31 Superficial temporal artery
32 **Stylohyoid muscle**
33 External carotid artery
34 Retromandibular vein
35 Superior thyroid artery
36 Inferior constrictor muscle of the pharynx
37 Sternocleidomastoid muscle
38 Omohyoid muscle
39 Common carotid artery and vagus nerve

Dissection of deep facial and retromandibular regions after removal of mandible. Pterygoid muscles removed, temporalis muscle fenestrated.

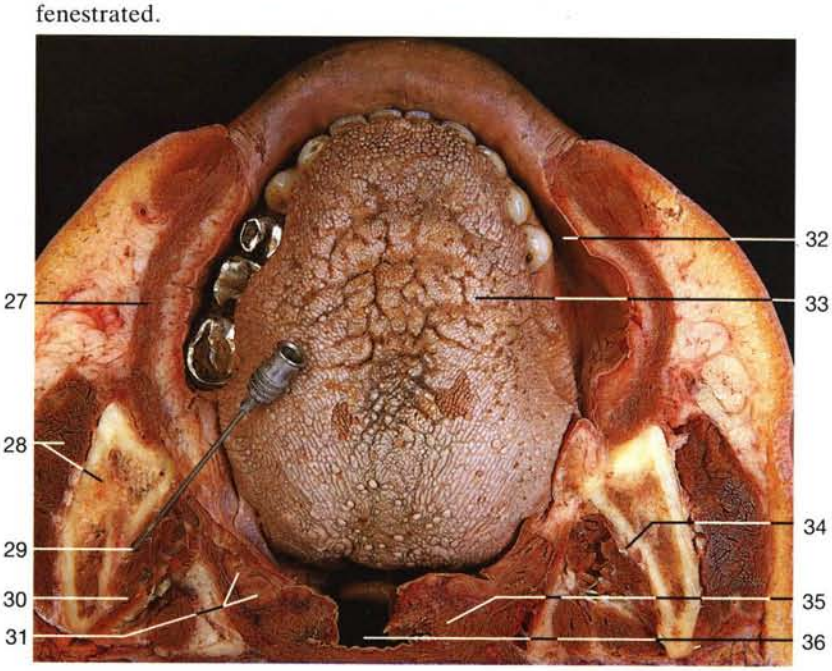

Transverse section through oral cavity and pharynx. The location of inferior alveolar nerve and artery is indicated by a needle.

1 Superficial temporal artery and vein and auriculotemporal nerve
2 Temporalis tendon, deep temporal nerves and artery
3 **Maxillary artery**
4 Middle meningeal artery
5 Occipital artery
6 Inferior alveolar artery and nerve
7 Posterior belly of digastric muscle
8 Great auricular nerve and sternocleidomastoid muscle
9 Hypoglossal nerve and superior root of ansa cervicalis
10 **External carotid artery**
11 Supratrochlear nerve and medial branch of supraorbital artery
12 Angular artery
13 Posterior superior alveolar artery
14 Infraorbital nerve
15 **Facial artery**
16 Parotid duct (divided) and buccinator muscle
17 Buccal artery and nerve

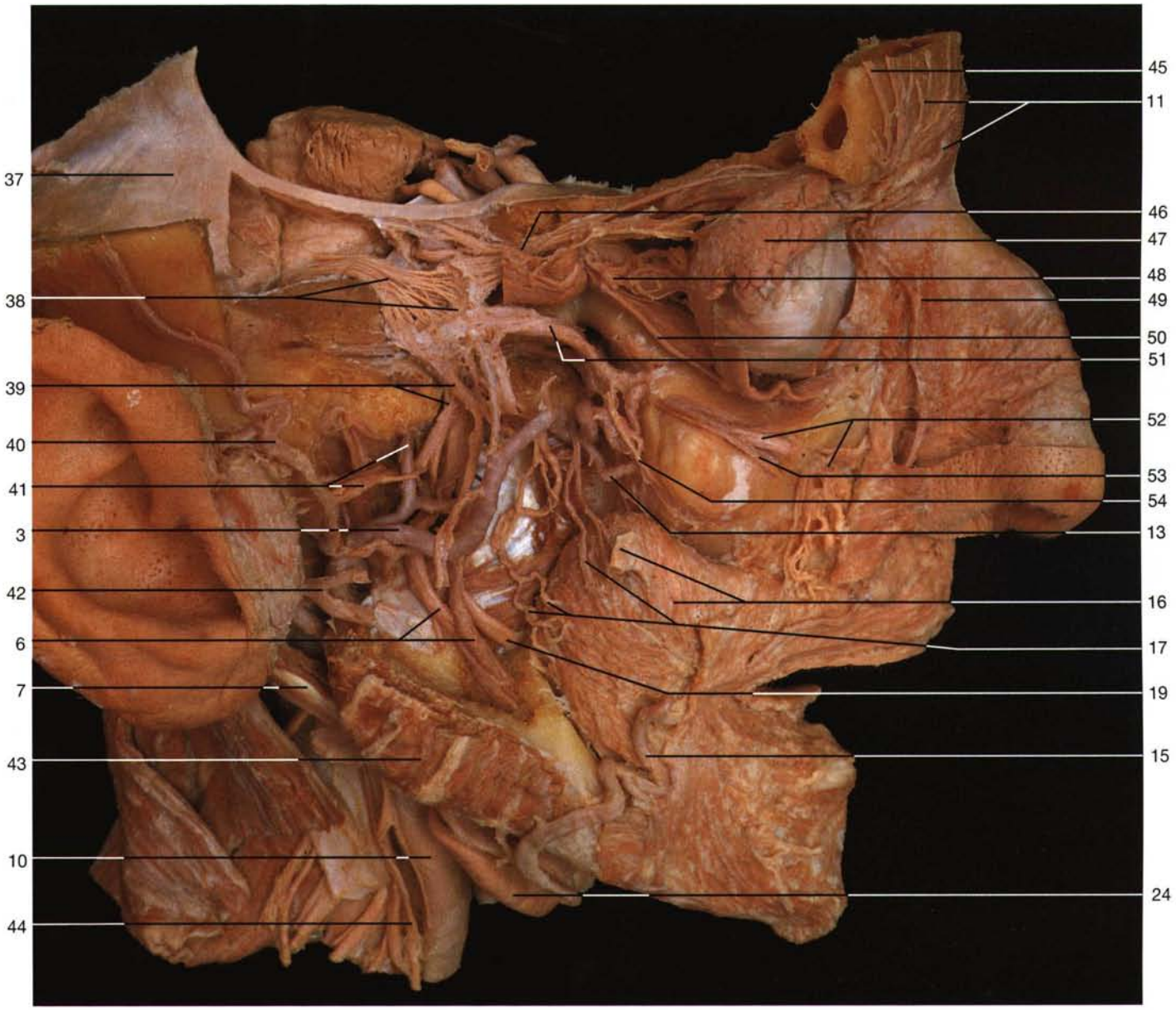

Para- and retropharyngeal regions. The mandible and the lateral wall of the orbit have been removed. The main branches of the trigeminal nerve and its ganglion are displayed.

18 Mylohyoid nerve
19 **Lingual nerve** and **submandibular ganglion**
20 Mental nerve and mental foramen
21 Inferior alveolar nerve
22 Mylohyoid muscle (divided) and hypoglossal nerve
23 Submental artery and vein
24 Submandibular gland
25 Superior thyroid artery
26 Common carotid artery
27 Buccinator muscle
28 Masseter muscle and mandible
29 Entrance of mandibular canal
30 Medial pterygoid muscle
31 Palatine tonsil
32 Oral vestibule
33 Tongue
34 Inferior alveolar nerve, artery and vein
35 Paryngeal constrictor muscle
36 Pharynx
37 Tentorium of cerebellum
38 **Trigeminal nerve** and **ganglion**
39 **Mandibular nerve**
40 Superficial temporal artery
41 Auriculotemporal nerve and middle meningeal artery
42 Facial nerve (divided)
43 Masseter muscle
44 Superior root of ansa cervicalis
45 Lateral branch of supraorbital nerve
46 **Ophthalmic nerve**
47 Lacrimal gland
48 Ciliary ganglion and short ciliary nerves
49 Angular artery
50 Inferior branch of oculomotor nerve
51 **Maxillary nerve**
52 Infraorbital nerve
53 Anterior superior alveolar nerve
54 Posterior superior alveolar nerve

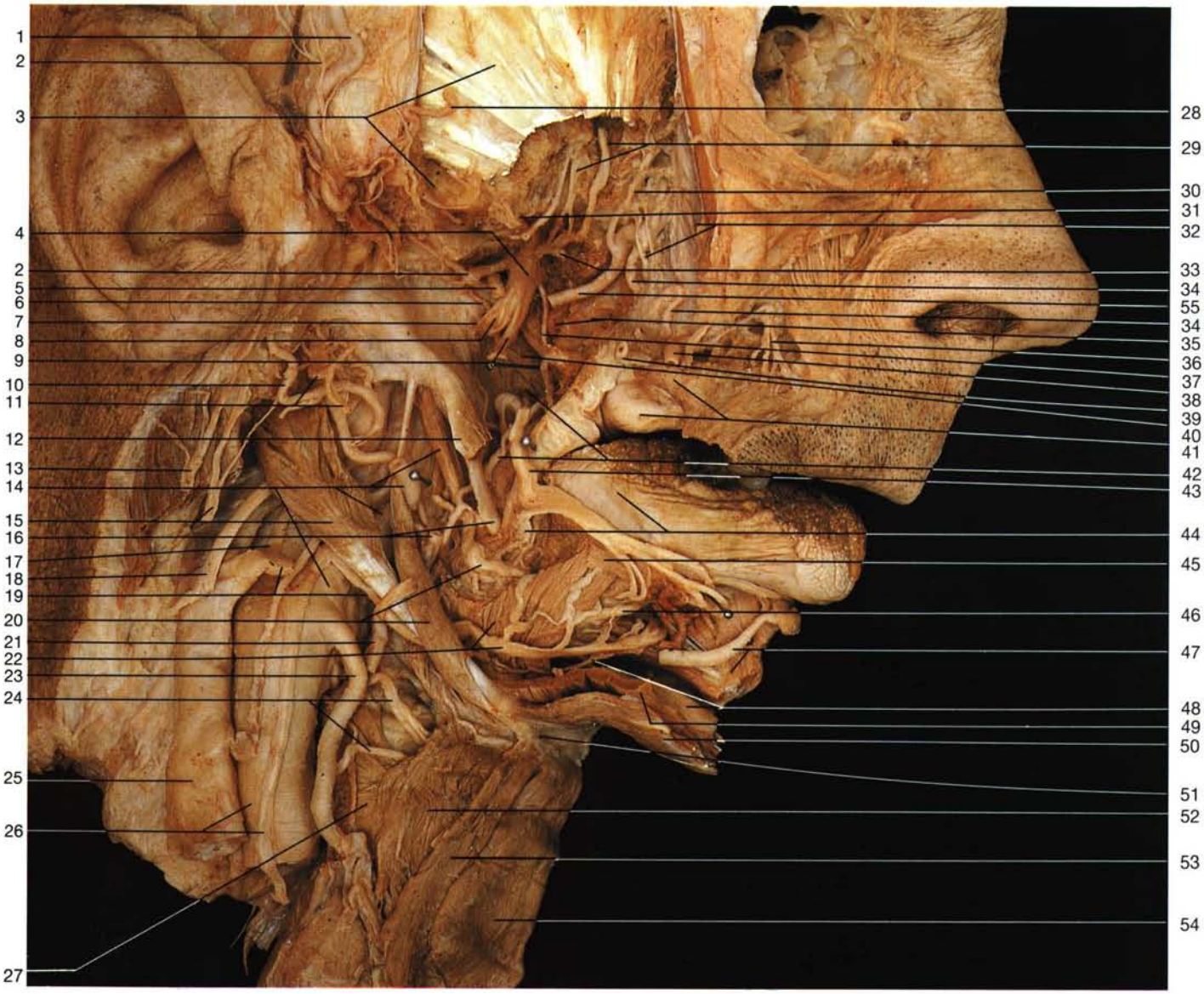

Retropharyngeal and sublingual regions. The mandible has been completely removed.

1 Parietal branch of superficial temporal artery
2 Auriculotemporal nerve
3 Temporalis muscle tendon and zygomatic arch (cut)
4 Mandibular nerve (n. V_3)
5 Middle meningeal artery
6 Chorda tympani
7 Mylohyoid nerve
8 Inferior alveolar nerve
9 **Lingual nerve**
10 Posterior auricular artery
11 Facial nerve
12 Styloglossus muscle (cut)
13 Great auricular nerve
14 Maxillary artery, external carotid artery, and stylopharyngeus muscle
15 Posterior belly of digastric muscle
16 Styloid process and facial artery
17 Vagus nerve (n. X)
18 **Accessory nerve** (n. XI)
19 Hypoglossal nerve (n. XII) and retromandibular vein (cut)
20 Stylohyoid muscle and glossopharyngeal nerve (n. IX)
21 Facial vein
22 Hypoglossal nerve (n. XII) and hyoglossus muscle
23 Superior thyroid artery
24 Superior laryngeal artery and internal laryngeal nerve
25 Internal jugular vein
26 Common carotid artery and superior root of ansa cervicalis
27 Inferior pharyngeal constrictor muscle
28 Middle temporal artery
29 Deep temporal nerves and posterior deep temporal artery
30 Anterior deep temporal artery
31 Masseteric nerve
32 Posterior superior alveolar branches of maxillary artery and nerve
33 Pterygoid branches of mandibular nerve
34 Posterior superior alveolar artery
35 Lateral pterygoid plate and medial pterygoid muscle
36 Infraorbital nerve and artery
37 Buccal nerve
38 Facial artery
39 Parotid duct

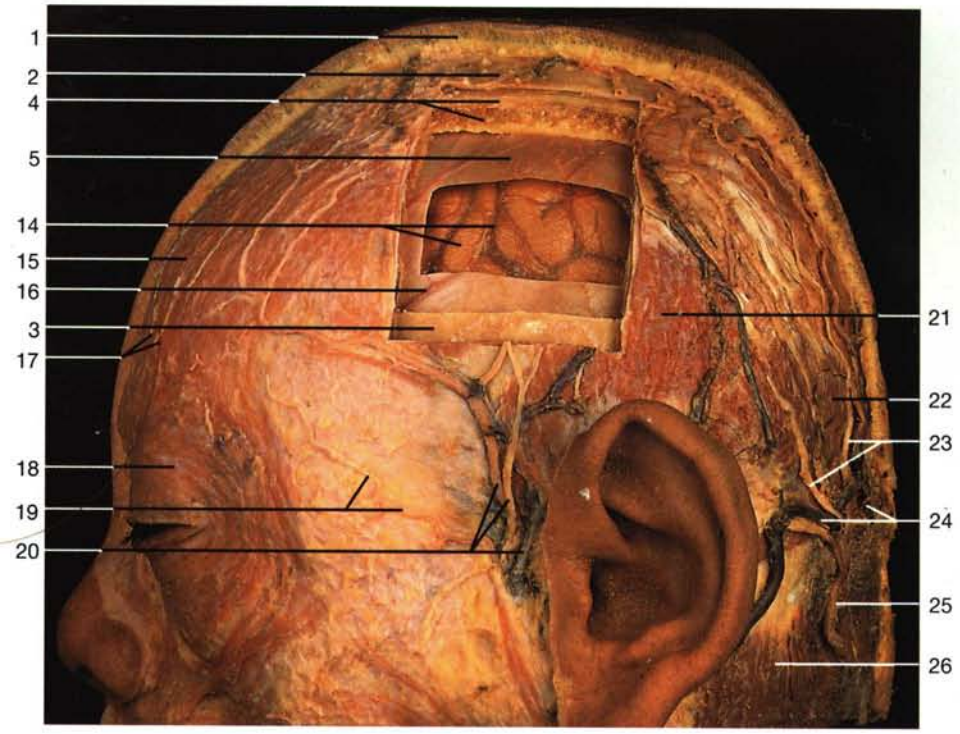

Nerves and blood vessels of the scalp. Scalp and meninges are demonstrated by a series of window-like openings.

1 Skin
2 **Galea aponeurotica**
3 Pericranium (periosteum)
4 Skull with diploe
5 **Dura mater**
6 Subdural space
7 **Arachnoid mater**
8 **Subarachnoid space**
9 Arachnoid granulations
10 Superior sagittal sinus
11 **Pia mater** with cerebral vessels
12 Falx cerebri
13 Cerebral cortex
14 Arachnoid and pia mater with cortical vessels
15 Frontal belly of occipitofrontal muscle
16 Branch of middle meningeal artery
17 Lateral and medial branch of supraorbital nerve
18 Orbicularis oculi muscle
19 Zygomatico-orbital artery
20 **Auriculotemporal nerve, superficial temporal artery** and **vein**
21 Superior auricular muscle
22 Occipital belly of occipitofrontalis muscle
23 Branches of **greater occipital nerve**
24 **Occipital artery** and **vein**
25 Greater occipital nerve
26 Sternocleidomastoid muscle

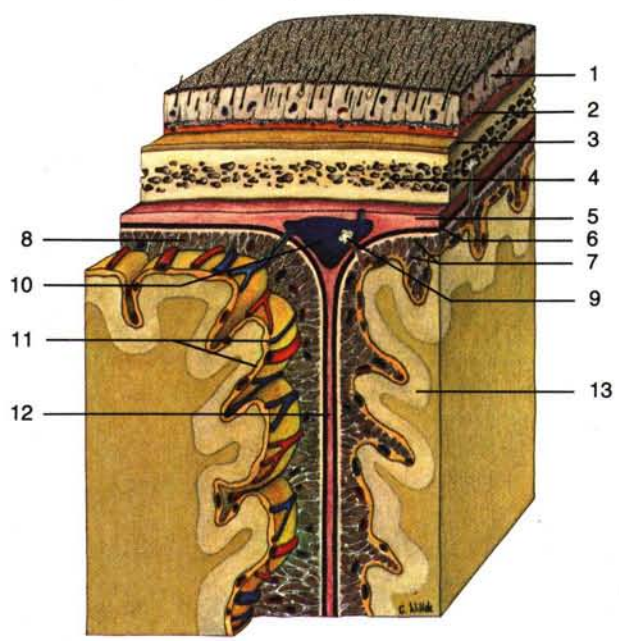

A coronal **section through the vertex of the skull** showing the arrangement of the meninges and vessels of the brain. Together the arachnoid mater and the pia mater form the leptomeninx.

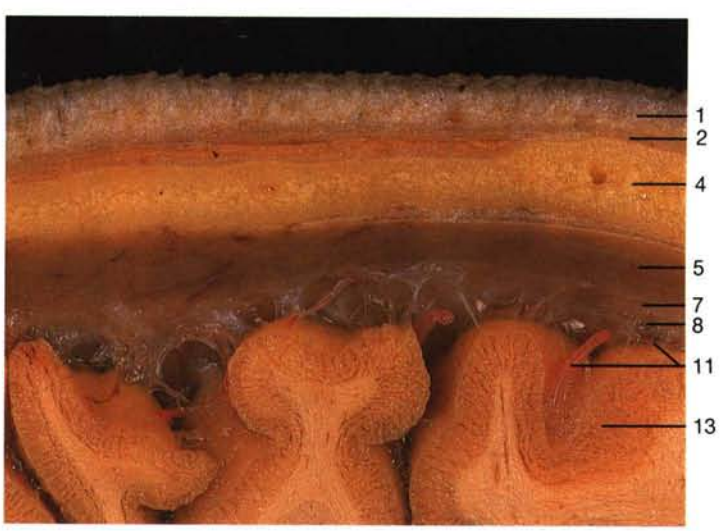

Cross-section of the scalp and the meninges. The subarachnoid space is shown.

40 Levator veli palatini muscle
41 Gingiva and buccinator muscle
42 Ascending pharyngeal artery and superior pharyngeal constrictor muscle (pterygopharyngeal part)
43 Lingual nerve
44 **Submandibular ganglion** and tongue
45 Palatoglossus muscle
46 Deep lingual artery
47 Submandibular duct and genioglossus muscle
48 Geniohyoid nerve and muscle
49 Mylohyoid muscle
50 Anterior belly of digastric muscle
51 Hyoid bone
52 Thyrohyoid muscle
53 Omohyoid muscle (superior belly)
54 Sternohyoid muscle
55 **Maxillary artery**

Median sagittal section through the head and neck.

1 Falx cerebri
2 **Corpus callosum** and **septum pellucidum**
3 Interventricular foramen and fornix
4 Choroid plexus of third ventricle and internal cerebral vein
5 **Third ventricle** and interthalamic adhesion
6 **Pineal body** and **colliculi of the midbrain**
7 Cerebral aqueduct
8 Mamillary body and basilar artery
9 Straight sinus
10 **Fourth ventricle** and **cerebellum**
11 Pons and falx cerebelli
12 **Medulla oblongata**
13 Central canal
14 **Cerebellomedullary cistern**
15 Dens of the axis (odontoid process)
16 Spinal cord
17 **Superior sagittal sinus**
18 **Anterior cerebral artery**
19 Anterior commissure
20 **Frontal sinus**
21 Crista galli
22 Optic chiasma
23 Pituitary gland (hypophysis)
24 Superior nasal concha
25 **Middle nasal concha** and **sphenoid sinus**
26 **Inferior nasal concha**
27 **Pharyngeal opening of auditory tube**
28 Superior longitudinal muscle of tongue
29 Vertical muscle of the tongue
30 **Uvula**
31 Genioglossus muscle
32 Pharynx
33 **Epiglottis**
34 Geniohyoid muscle
35 Mylohyoid muscle
36 Hyoid bone
37 Vocal fold and sinus of larynx
38 Esophagus

Dura mater and venous sinuses of the dura mater. The brain has been removed (oblique lateral aspect).

1 Falx cerebri
2 Position of middle meningeal artery and vein
3 Internal carotid artery
4 **Optic nerves** (n. II)
5 Frontal sinus
6 **Oculomotor nerve** (n. III)
7 Diploe
8 **Dura mater**
9 Superior sagittal sinus
10 Straight sinus
11 **Trigeminal nerve** (n. V)
12 Facial and vestibulocochlear nerve (n. VII and n. VIII)
13 **Tentorium cerebelli**
14 **Pituitary gland (hypophysis)**
15 Inferior sagittal sinus
16 Sigmoid sinus
17 Confluence of sinuses
18 Inferior petrosal sinus
19 Transverse sinus
20 Superior petrosal sinus
21 **Cavernous** and **intercavernous sinuses**

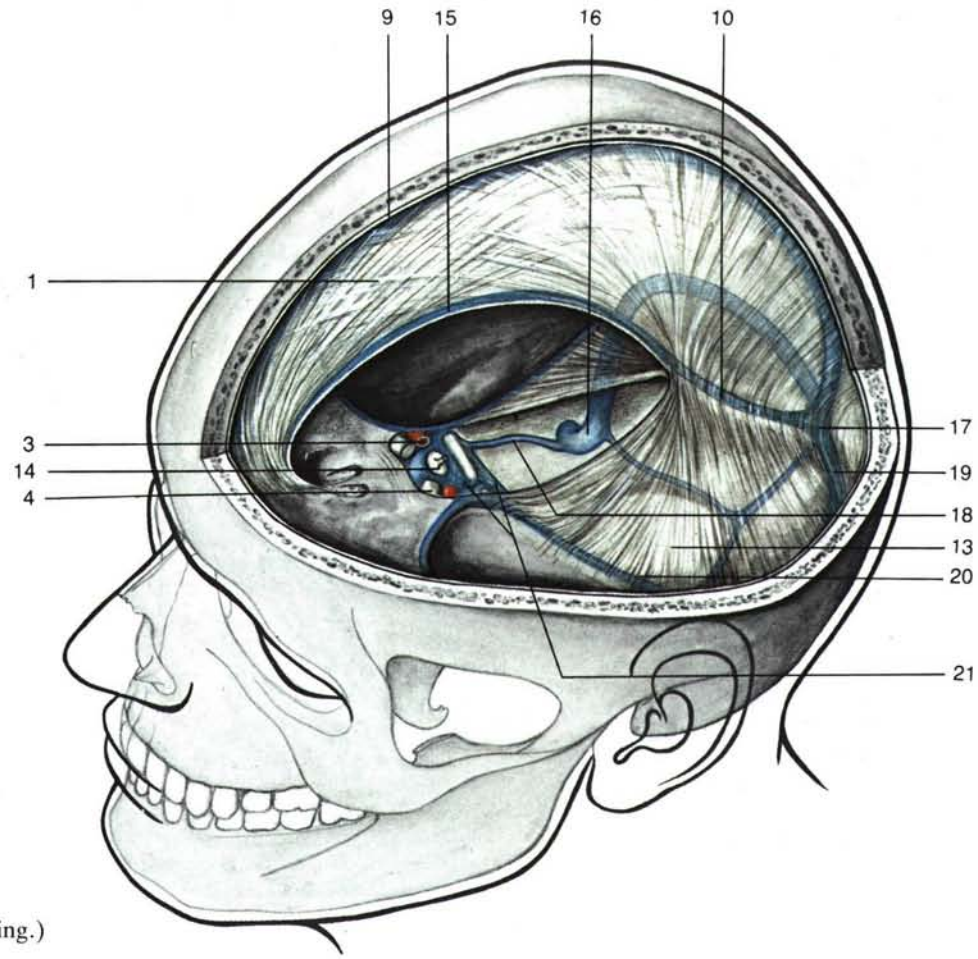

Dura mater and **venous sinuses** (left lateral aspect). (Schematic drawing.)

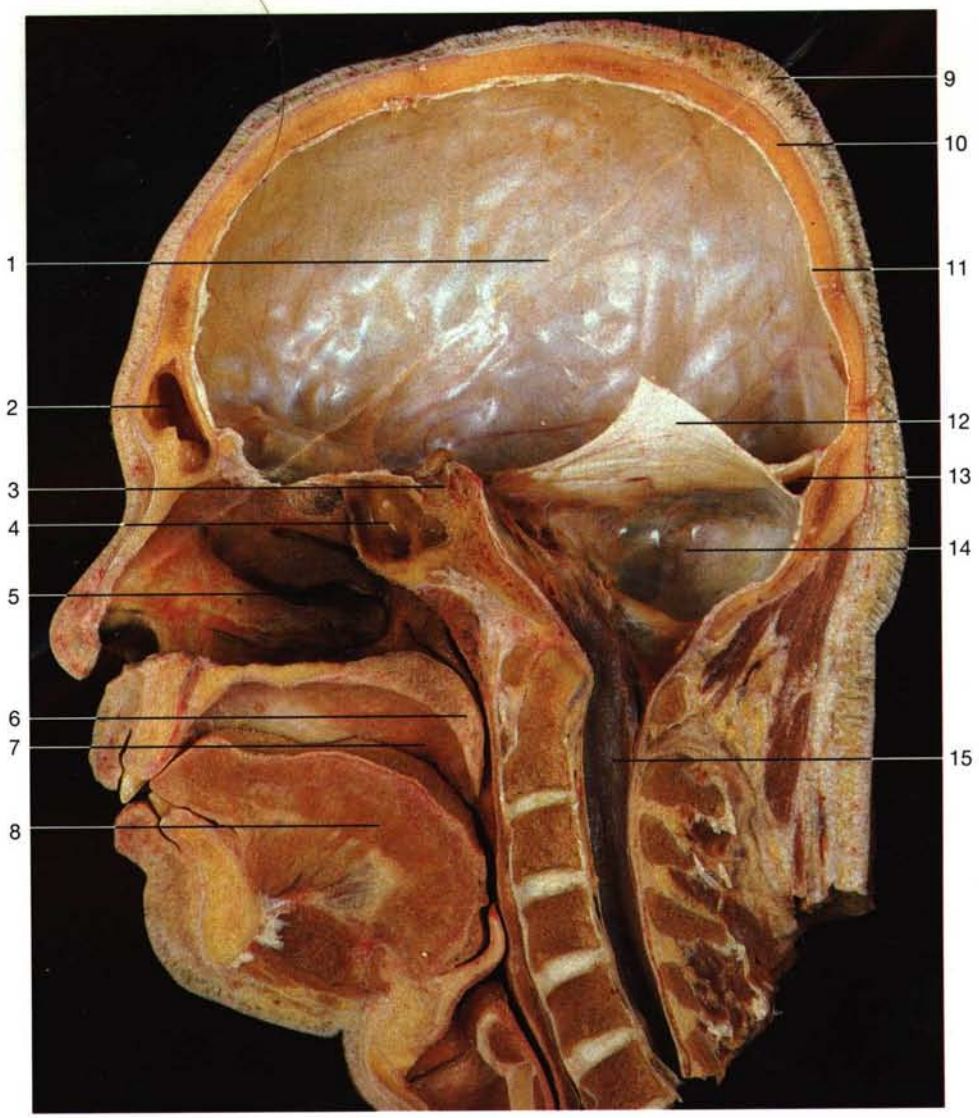

1 **Cranial cavity** with dura mater (the right cerebral hemisphere has been removed)
2 Frontal sinus
3 Hypophysial fossa with pituitary gland
4 Sphenoidal sinus
5 **Nasal cavity**
6 Soft palate (uvula)
7 **Oral cavity**
8 Tongue
9 Skin
10 Calvaria
11 Dura mater
12 Tentorium cerebelli
13 Confluence of sinuses
14 **Infratentorial space** (cerebellum and part of the brain stem have been removed)
15 Vertebral canal
16 Frontal branch of middle meningeal artery and veins
17 Middle meningeal artery
18 **Diploe**
19 Parietal branch of middle meningeal artery and vein
20 Occipital pole of left hemisphere covered with dura mater

Median section through the head. Demonstration of dura mater covering the cranial cavity. Brain and spinal cord are removed (right half of the head, as seen from medial).

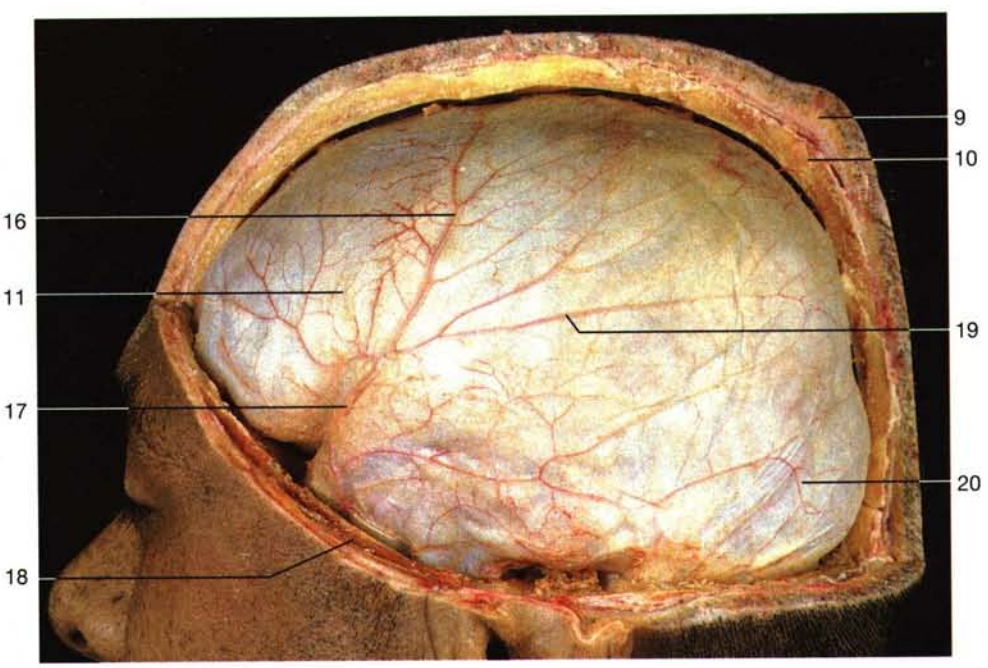

Dissection of dura mater and meningeal vessels. Left half of calvaria removed.

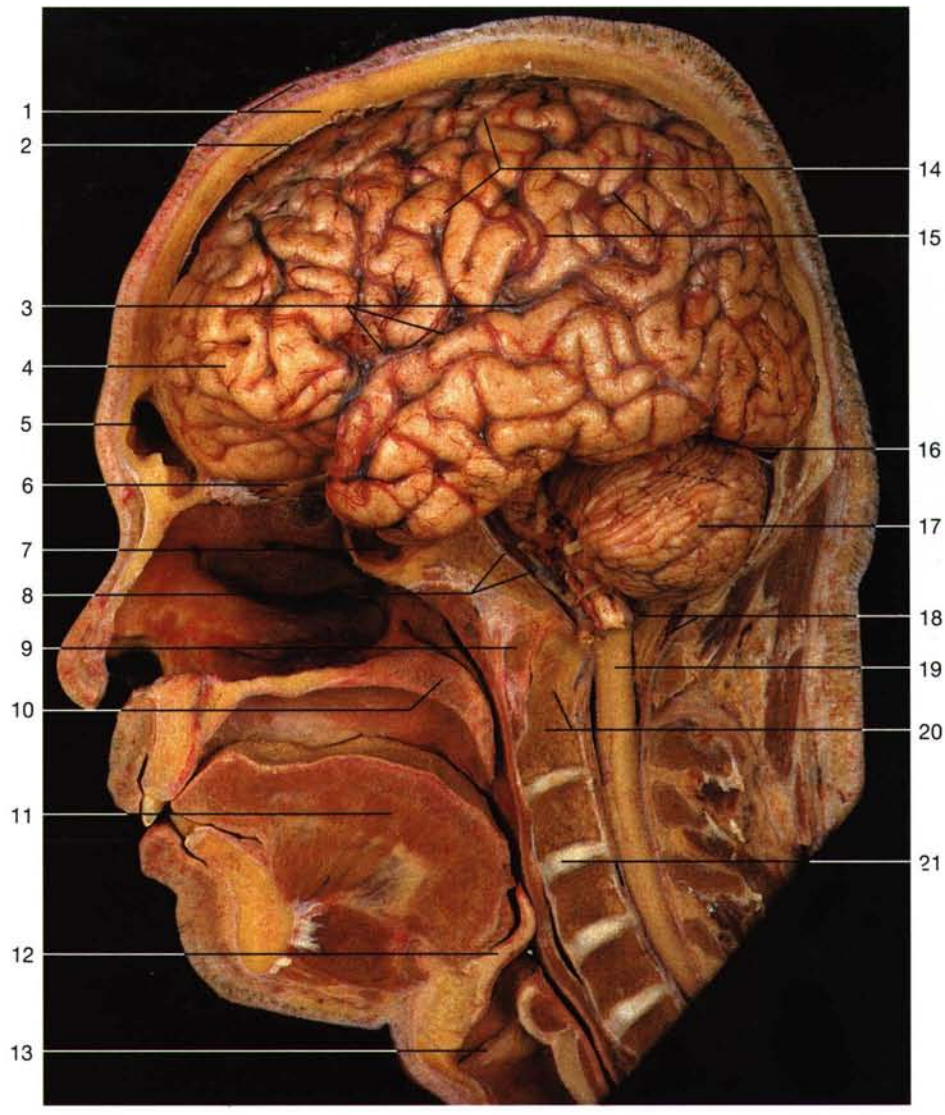

1 Calvaria and skin of the scalp
2 Dura mater (divided)
3 Position of lateral sulcus
4 **Frontal lobe** covered by **arachnoid** and **pia mater**
5 Frontal sinus
6 Olfactory bulb
7 **Sphenoidal sinus**
8 Dura mater on clivus and basilar artery
9 Atlas (anterior arch, divided)
10 **Soft palate**
11 Tongue
12 Epiglottis
13 Vocal fold
14 Position of central sulcus
15 Superior cerebral veins
16 Tentorium (divided)
17 **Cerebellum**
18 Cerebellomedullary cistern
19 Position of foramen magnum and **spinal cord**
20 Dens of axis
21 Intervertebral disc

Dissection of the brain with pia mater and arachnoid in situ. The head is cut in half except for the brain, which is shown in its entirety.

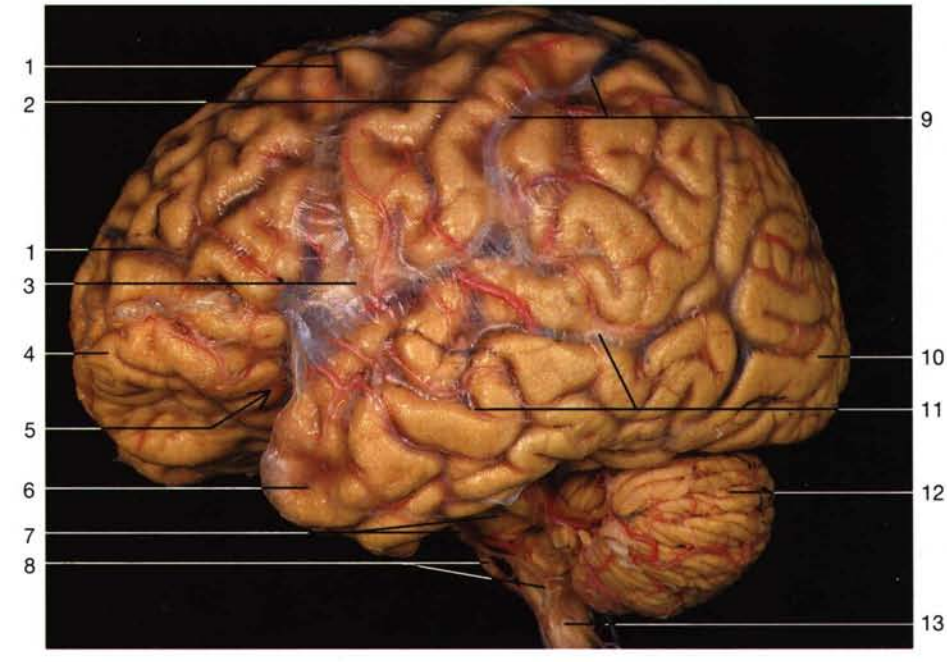

1 Superior cerebral veins
2 Position of central sulcus
3 Position of lateral sulcus and **cistern of lateral cerebral fossa**
4 **Frontal pole**
5 Lateral sulcus (arrow)
6 **Temporal pole**
7 Pons and basilar artery
8 Vertebral arteries
9 Superior anastomotic vein
10 Occipital pole
11 Inferior cerebral veins
12 Hemisphere of **cerebellum**
13 Medulla oblongata

Brain with pia mater and arachnoid. Frontal pole to the left (lateral aspect).

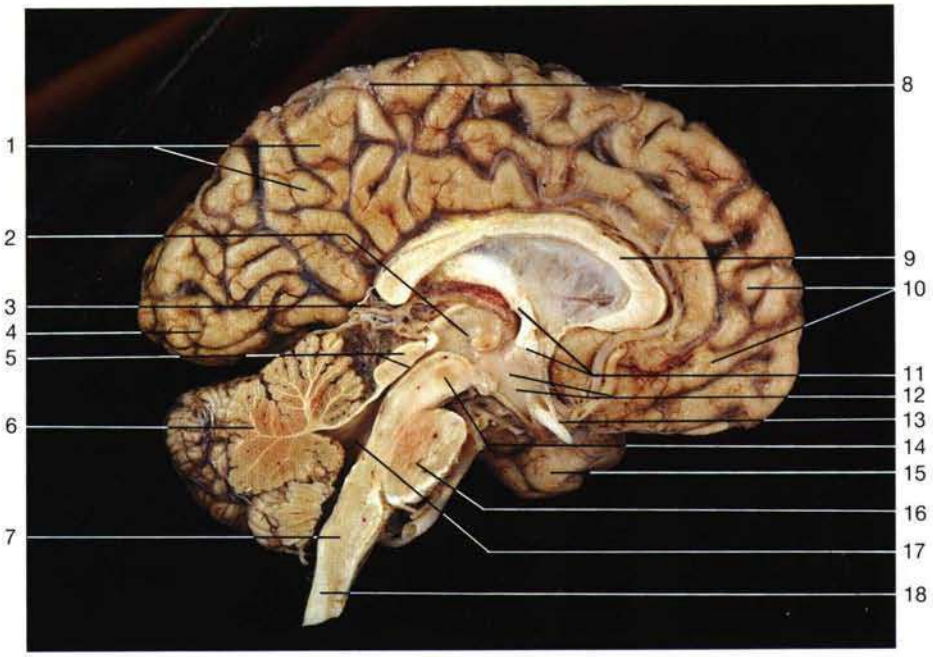

Brain and brain stem, median section. Frontal pole to the right.

1 Parietal lobe
2 Thalamus, **third ventricle** and intermediate mass
3 Great cerebral vein
4 **Occipital lobe**
5 Colliculi of the **midbrain** and **cerebral aqueduct**
6 Cerebellum
7 **Medulla oblongata**
8 Central sulcus
9 **Corpus callosum**
10 **Frontal lobe**
11 Fornix and anterior commissure
12 **Hypothalamus**
13 Optic chiasma
14 Midbrain
15 Temporal lobe
16 **Pons**
17 **Fourth ventricle**
18 Spinal cord
19 Inferior concha and nasal cavity
20 Alveolar process of maxilla
21 Tongue
22 Dens of axis
23 Oral part of pharynx
24 Alveolar process of mandible
25 Epiglottis

Median section through the head. (MR-Scan, cf. section on opposite page.)

1 Frontal lobe of cerebrum
2 Occipital lobe of cerebrum
3 **Corpus callosum**
4 Anterior commissure
5 **Lamina terminalis**
6 Optic chiasma
7 **Hypothalamus**
8 Thalamus and third ventricle
9 **Colliculi of the midbrain**
10 Midbrain (inferior portion)
11 **Cerebellum**
12 Pons
13 **Fourth ventricle**
14 Medulla oblongata
15 Central canal
16 **Spinal cord**

Median section through the head. Regions of the brain. Falx cerebri removed.

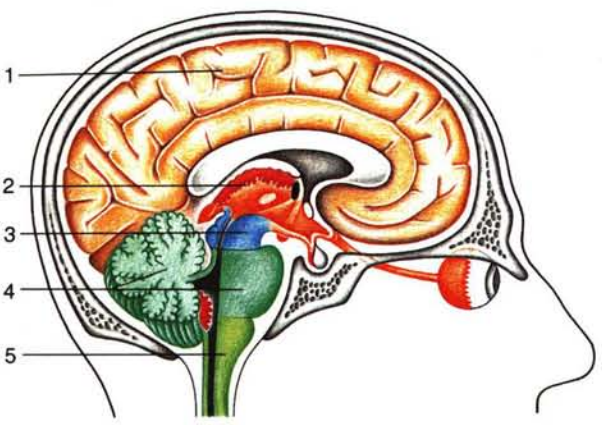

I. Prosencephalon (forebrain)	1. Telencephalon (cerebral hemispheres, striatum, etc.) 2. Diencephalon (thalamus, metathalamus, hypothalamus, etc.)
II. Mesencephalon (midbrain)	3. Mesencephalon (colliculi, cerebral peduncles, tegmentum)
III. Rhombencephalon (hindbrain)	4. Metencephalon (pons, cerebellum) 5. Myelencephalon (medulla oblongata)

Main divisions of the brain
I-III = primary brain vesicles; 1-5 = secondary brain vesicles

Scheme of brain divisions (cf. table).
Red = choroidal plexus. (Schematic drawing.)
1 Telencephalon (yellow) with lateral ventricles
2 Diencephalon (orange) with III ventricle, optic nerve and retina
3 Mesencephalon (blue) with cerebral aqueduct
4 Metencephalon (green) with IV ventricle
5 Myelencephalon (yellow-green)

Midbrain, pons and medulla oblongata are collectively termed the **brain stem.**

1 Superior cerebral veins and parietal lobe
2 Frontal lobe
3 Superficial middle cerebral vein and **cistern of lateral cerebral fossa**
4 Temporal lobe
5 Occipital lobe
6 Inferior cerebral veins and transverse occipital sulcus
7 Inferior anastomotic vein
8 Cerebellum
9 Medulla oblongata

Brain with pia mater. Cerebral veins (bluish). In the lateral sulcus the cistern of the lateral fossa is recognizable. Frontal lobe to the left.

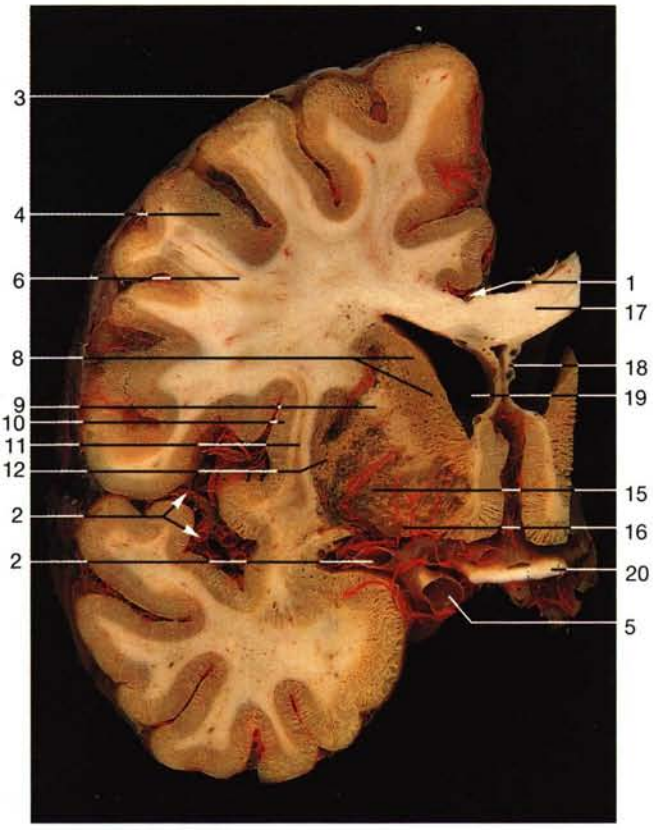

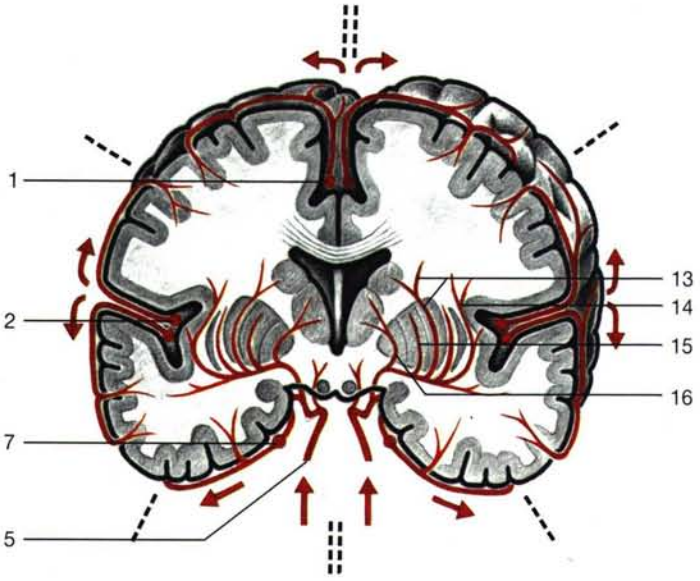

Arteries of the brain. Coronal section. Areas supplied by cortical and central arteries. Dotted lines indicate boundaries of arterial supply areas; arrows: direction of blood flow.

◁

Coronal section through the right hemisphere showing arachnoid, pia mater, and the arterial blood supply (anterior aspect).

1 **Anterior cerebral artery**
2 **Middle cerebral arteries**
3 Arachnoid
4 Cortex
5 **Internal carotid artery**
6 Frontal lobe (white matter)
7 **Posterior cerebral artery**
8 Caudate nucleus
9 Internal capsule
10 Insular lobe
11 Claustrum
12 Putamen
13 Posterior striate branch
14 Insular artery
15 Pallidostriate artery
16 Thalamic artery
17 Corpus callosum
18 Pellucid septum
19 Lateral ventricle
20 Optic chiasm

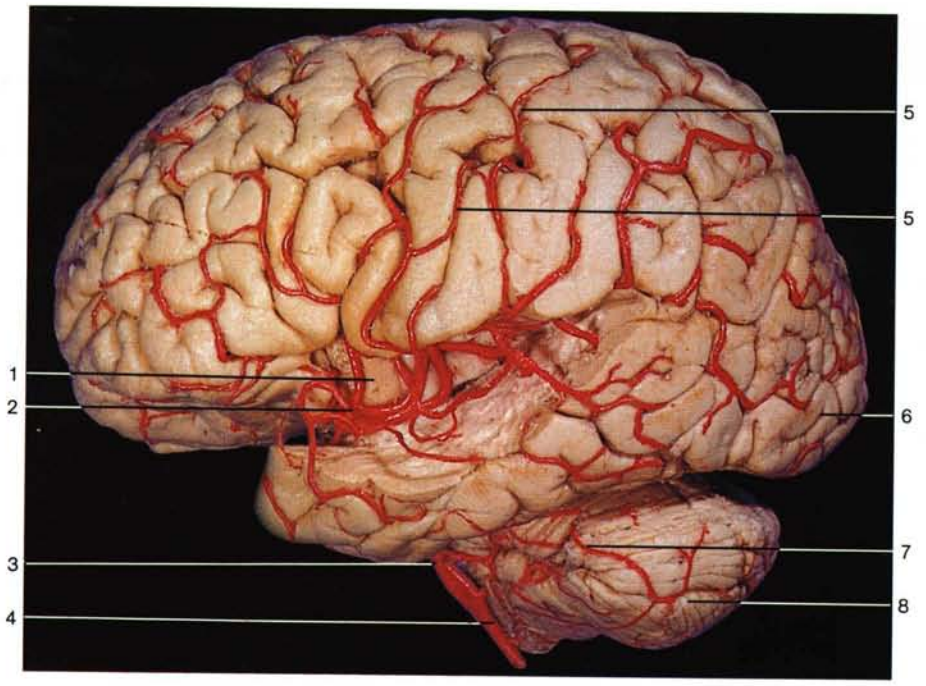

1 Insula
2 **Middle cerebral artery** [2 branches:
a) Parietal branches,
b) Temporal branches]
3 Basilar artery
4 **Vertebral artery**
5 Central sulcus
6 Occipital lobe
7 Superior cerebellar artery
8 Cerebellum
9 **Anterior cerebral artery**
10 Ethmoidal arteries
11 Ophthalmic artery
12 **Internal carotid artery**
13 Posterior communicating artery
14 **Posterior cerebral artery**
15 Anterior inferior cerebellar artery
16 Posterior inferior cerebellar artery

Cerebral arteries. Lateral aspect of the left hemisphere. The upper part of the temporal lobe has been removed to display the insula and cerebral arteries.

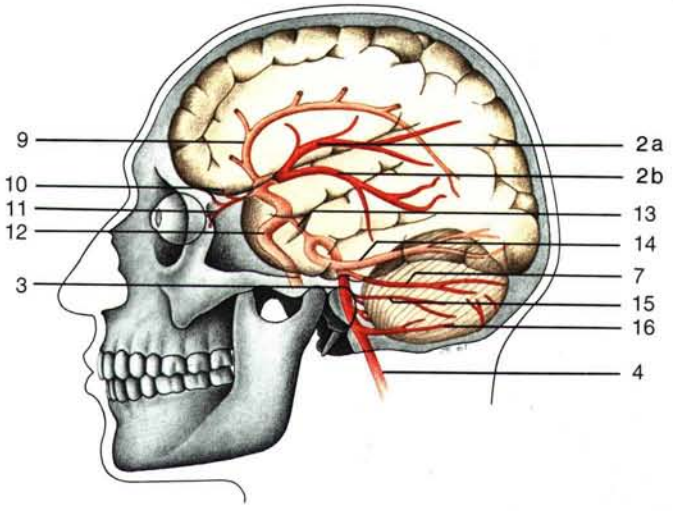

Arteries of the brain.

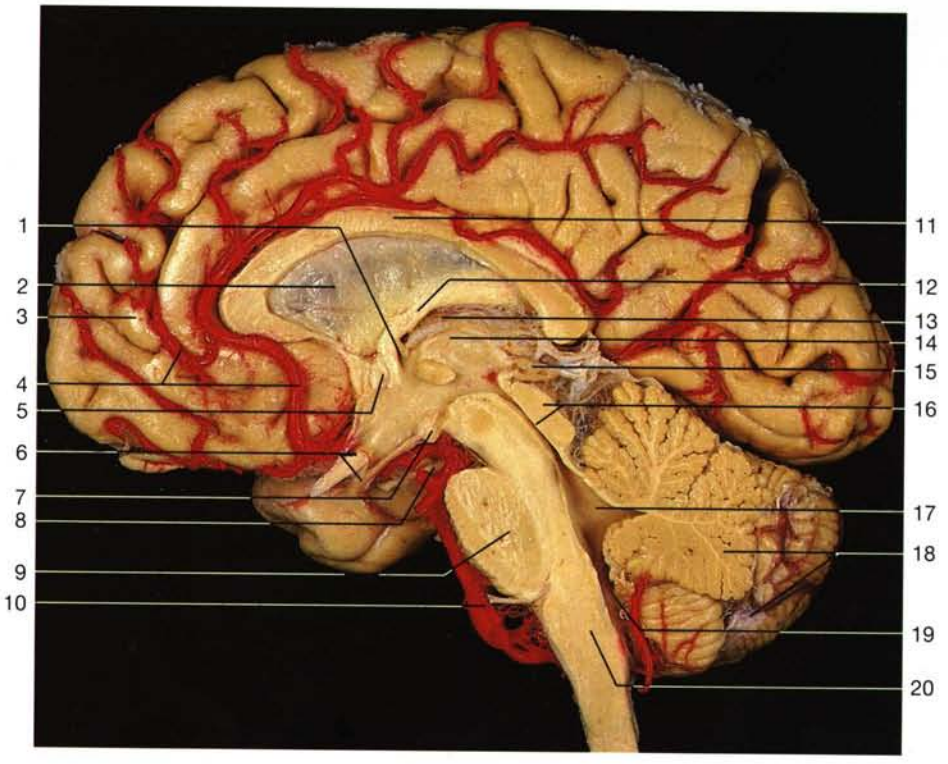

1 Interventricular foramen
2 Septum pellucidum
3 Frontal lobe
4 **Anterior cerebral artery**
5 Anterior commissure
6 Optic chiasma and infundibulum
7 Mamillary body
8 Oculomotor nerve (n. III)
9 Pons
10 **Basilar artery**
11 Corpus callosum
12 **Fornix**
13 **Choroid plexus**
14 **Third ventricle**
15 Pineal body
16 **Tectum** and **cerebral aqueduct**
17 **Fourth ventricle**
18 Cerebellum (arbor vitae, vermis)
19 Median aperture of Magendie
20 Medulla oblongata

Median section through the brain and brain stem. Cerebral arteries injected with red resin.

1 Anterior communicating artery
2 Left **anterior cerebral artery**
3 **Internal carotid artery**
4 Pons and left **superior cerebellar artery**
5 **Anterior inferior cerebellar artery**
6 **Posterior inferior cerebellar artery**
7 Medulla oblongata
8 **Right anterior cerebral artery**
9 Olfactory tract
10 Optic nerve
11 **Middle cerebral artery**
12 Infundibulum
13 Oculomotor nerve and **posterior communicating artery**
14 **Posterior cerebral artery**
15 **Basilar artery** and abducens nerve (n. VI)
16 Anterior spinal artery
17 **Vertebral artery**
18 Cerebellum
19 Labyrinthine arteries
20 Posterior spinal artery
21 Right **superior cerebellar artery**
22 Olfactory bulb

Arteries of the brain (inferior aspect). Frontal pole above; right temporal lobe and cerebellum partly removed.

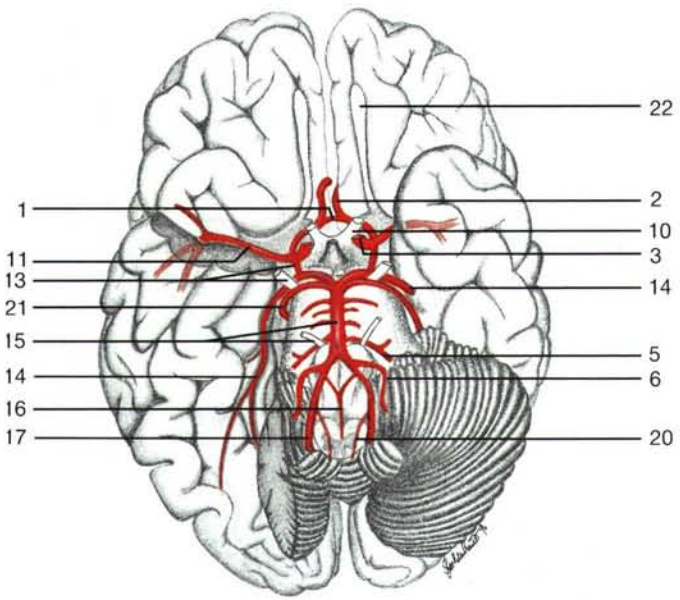

Arteries of the base of the brain, arterial circle of Willis. (Schematic drawing.)

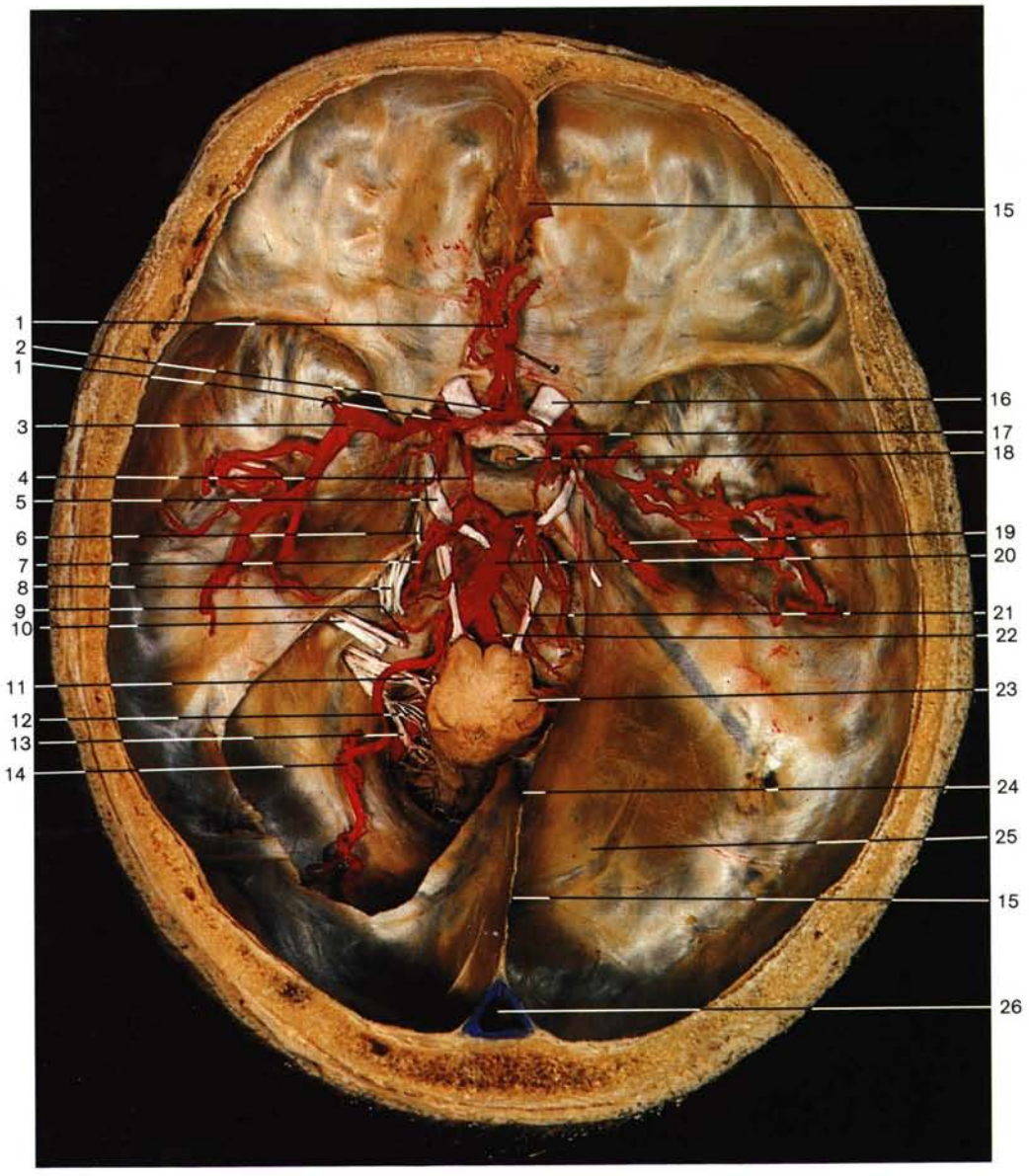

Base of the cranial cavity (internal aspect). Left tentorium has been cut.

1 Anterior cerebral artery
2 Anterior communicating artery
3 **Middle cerebral artery**
4 Posterior communicating artery
5 **Oculomotor nerve**
6 **Trochlear nerve**
7 Posterior cerebral artery
8 **Trigeminal nerve**
9 Internal auditory artery
10 **Facial nerve** and **vestibulocochlear nerve**
11 **Glossopharyngeal nerve** and **vagus nerve**
12 **Hypoglossal nerve**
13 **Accessory nerve**
14 Anterior inferior cerebellar artery
15 Falx cerebri
16 **Optic nerve**
17 Optic chiasma
18 Infundibulum and pituitary gland
19 Anterior choroidal artery and choroid plexus
20 **Basilar artery**
21 **Abducens nerve**
22 Vertebral arteries
23 Medulla oblongata
24 Inferior sagittal sinus
25 Tentorium cerebelli
26 Superior sagittal sinus and confluence of sinuses
27 Anterior spinal artery
28 Internal carotid artery
29 Superior cerebellar artery
30 Anterior inferior cerebellar artery
31 Posterior inferior cerebellar artery
32 Posterior spinal artery
33 Ophthalmic artery

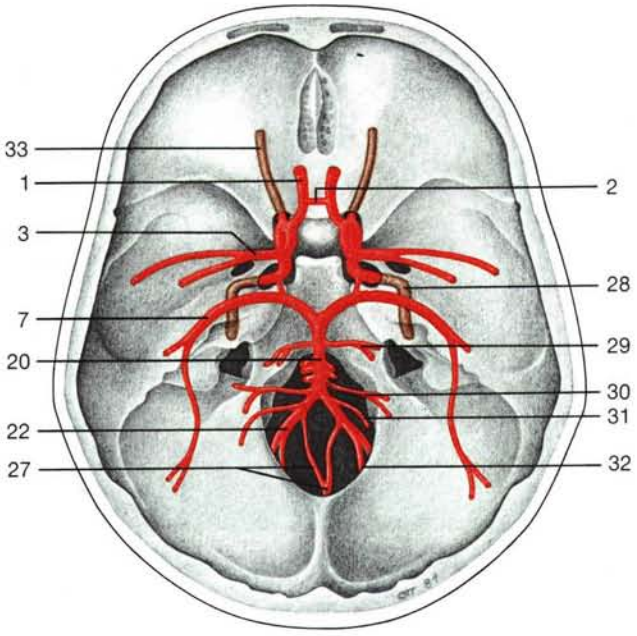

Arterial circle of Willis (superior aspect). (Schematic drawing.)

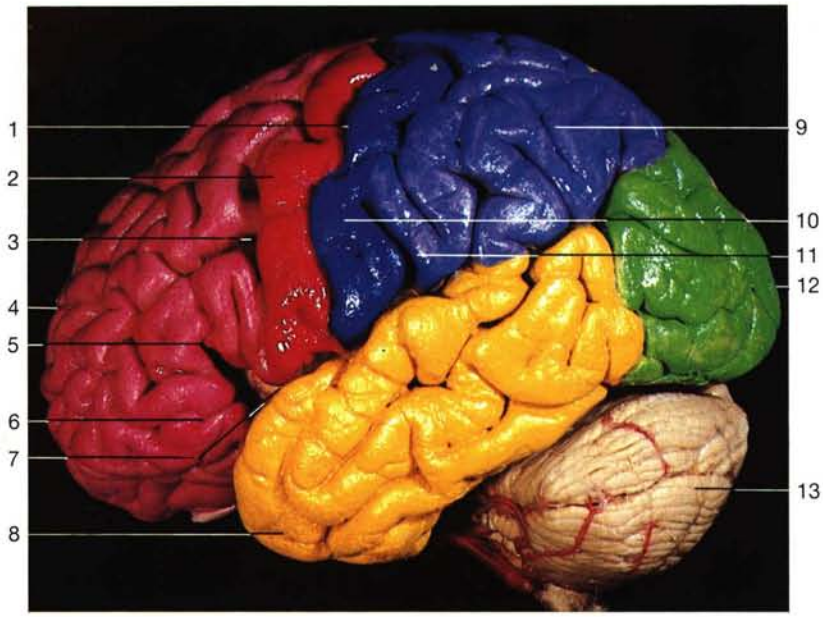

1 Central sulcus
2 Precentral gyrus
3 Precentral sulcus
4 **Frontal lobe**
5 Anterior ascending ramus of lateral sulcus
6 Anterior horizontal ramus of lateral sulcus
7 Lateral sulcus
8 **Temporal lobe**
9 **Parietal lobe**
10 Postcentral gyrus
11 Postcentral sulcus
12 **Occipital lobe**
13 Cerebellum
14 Superior frontal sulcus
15 Middle frontal gyrus
16 Lunate sulcus
17 Longitudinal fissure
18 Arachnoid granulations

Brain, left hemisphere (lateral aspect). Frontal pole to the left.

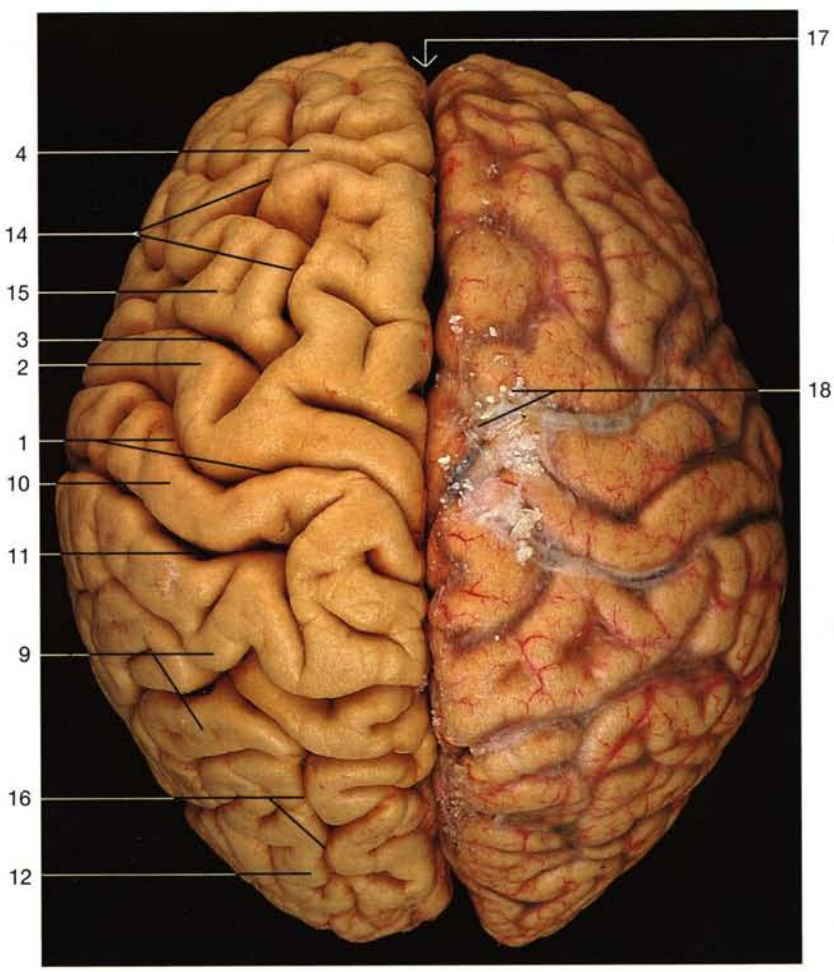

Brain (superior aspect). Right hemisphere with arachnoid and pia mater.

Pink = Frontal lobe
Blue = Parietal lobe
Green = Occipital lobe
Yellow = Temporal lobe
Dark red = Precentral gyrus
Dark blue = Postcentral gyrus

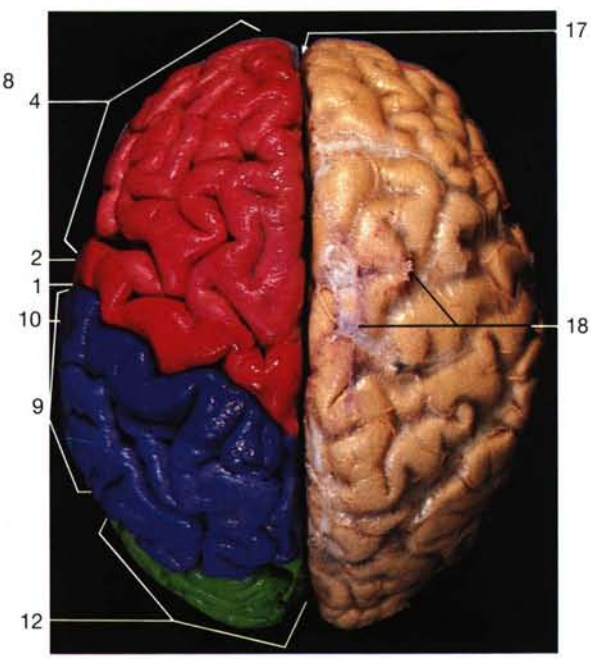

Brain (superior aspect). Lobes of the left hemisphere indicated by color; right hemisphere is covered with arachnoid and pia mater.

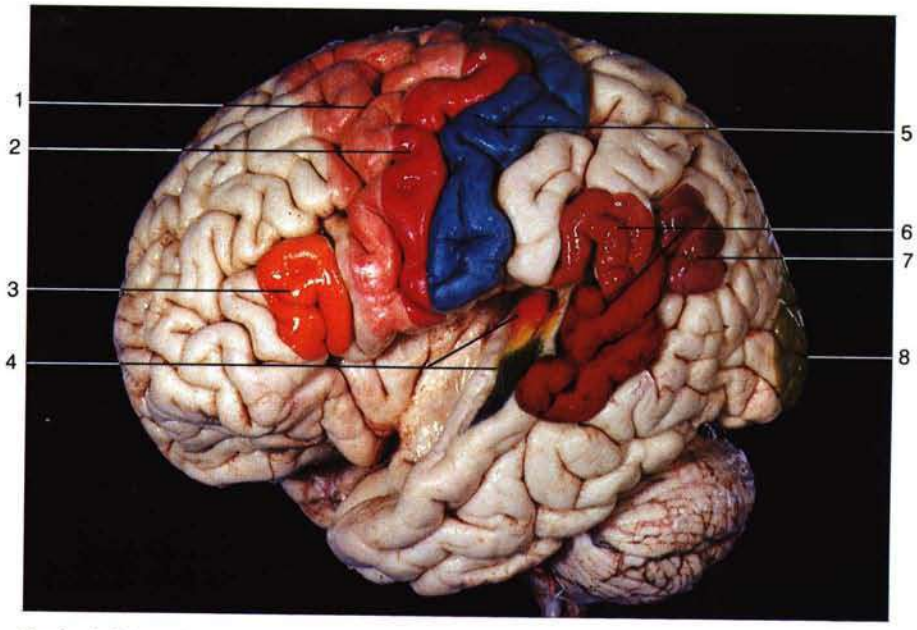

1 Premotor area
2 Somatomotor area
3 Motor speech area of Broca
4 Acoustic area
(red: high tone; blue: low tone)
5 Somatosensory area
6 Sensory speech area of Wernicke
7 Reading comprehension area
8 Visuosensory area

Brain, left hemisphere (lateral aspect). **Main cortical areas** are colored. The lateral sulcus has been opened to display the insula and the inner surface of the temporal lobe.

1 Precentral gyrus
2 Precentral sulcus
3 Superior frontal gyrus
4 **Central sulcus**
5 Middle frontal gyrus
6 Inferior frontal gyrus
7 Ascending ramus } of lateral sulcus
8 Horizontal ramus } of lateral sulcus
9 Posterior ramus } of lateral sulcus
10 Superior temporal gyrus
11 Middle temporal gyrus
12 Inferior temporal gyrus
13 Parietal lobule
14 Postcentral sulcus
15 Postcentral gyrus
16 Supramarginal gyrus
17 Angular gyrus
18 **Occipital lobe**
19 **Cerebellum**
20 Horizontal fissure of cerebellum
21 Medulla oblongata

Brain, left hemisphere (lateral aspect). Frontal pole to the left.

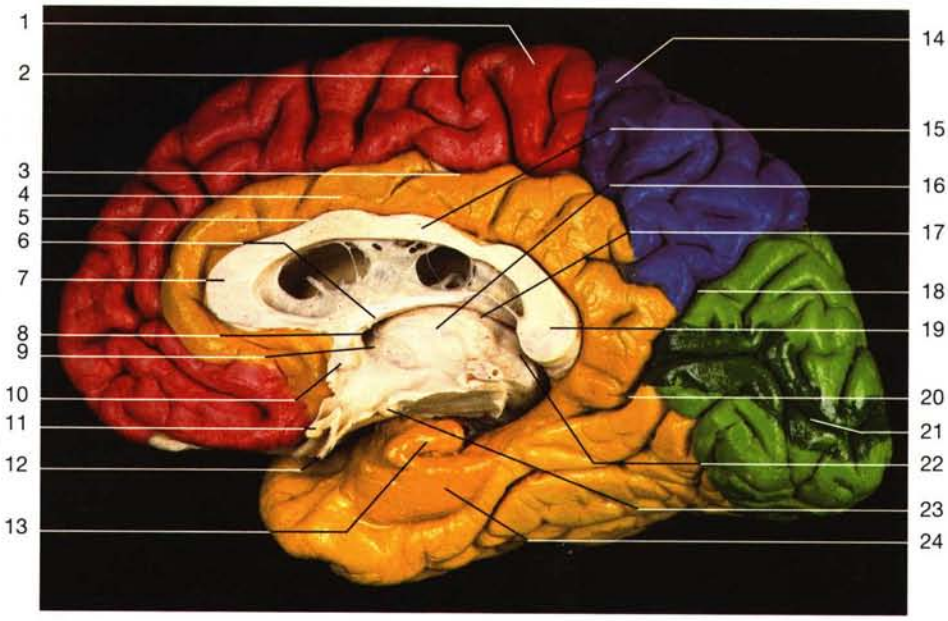

Brain, right hemisphere (medial aspect). Frontal pole to the left (midbrain divided, cerebellum and inferior part of brainstem removed).

1 **Precentral gyrus**
2 Precentral sulcus
3 Cingulate sulcus
4 **Cingulate gyrus**
5 Sulcus of corpus callosum
6 **Fornix**
7 Genu of **corpus callosum**
8 Interventricular foramen
9 Intermediate mass
10 Anterior commissure
11 **Optic chiasma**
12 **Infundibulum**
13 Uncus hippocampi
14 **Postcentral gyrus**
15 Body of corpus callosum
16 Third ventricle and thalamus
17 Stria medullaris
18 **Parietooccipital sulcus**
19 Splenium of corpus callosum
20 Communication of calcarine and parietooccipital sulcus
21 Calcarine sulcus
22 **Pineal body**
23 **Mamillary body**
24 **Parahippocampal gyrus**
25 **Olfactory bulb**
26 Olfactory tract
27 Gyrus rectus
28 **Optic nerve**
29 Infundibulum and optic chiasma
30 **Optic tract**
31 Oculomotor nerve
32 Pedunculus cerebri
33 **Red nucleus**
34 **Cerebral aqueduct**
35 Corpus callosum
36 Longitudinal fissure
37 Orbital gyri
38 Lateral root of olfactory tract
39 Medial root of olfactory tract
40 Olfactory tubercle and anterior perforated substance
41 Tuber cinereum
42 Interpeduncular fossa
43 **Substantia nigra**
44 Colliculi of the **midbrain**
45 Lateral occipitotemporal gyrus
46 Medial occipitotemporal gyrus

Brain (inferior aspect). Midbrain divided. Cerebellum and inferior part of brain stem removed. Frontal pole at the top.

Pink	=	Frontal lobe
Blue	=	Parietal lobe
Green	=	Occipital lobe
Yellow	=	Temporal lobe
Dark red	=	Precentral lobe
Dark blue	=	Postcentral lobe
Orange	=	Limbic cortex (cingulate and parahippocampal gyri)

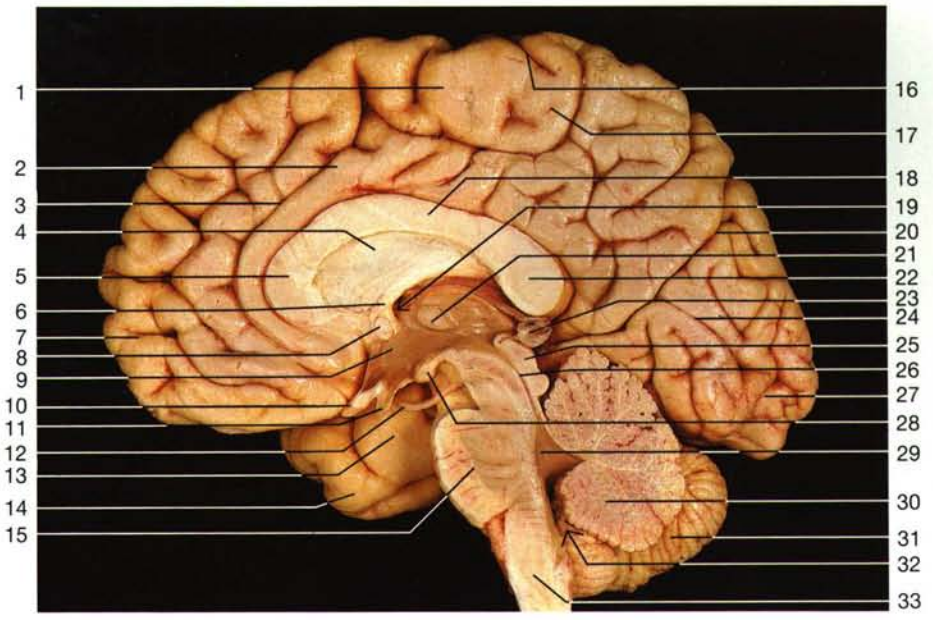

Brain (sagittal section). Frontal pole to the left.

Brain, with pia mater and blood vessels (inferior aspect).

1 Precentral gyrus
2 Cingulate gyrus
3 Cingulate sulcus
4 Septum pellucidum
5 Genu of corpus callosum
6 Fornix
7 **Frontal lobe**
8 Anterior commissure
9 **Hypothalamus**
10 Optic chiasma
11 Infundibulum
12 Oculomotor nerve
13 Uncus
14 **Temporal lobe**
15 Pons
16 Central sulcus
17 Postcentral gyrus
18 Body of corpus callosum
19 Interventricular foramen (arrow)
20 Parietooccipital sulcus
21 Intermediate mass
22 Splenium of corpus callosum
23 Pineal body
24 Calcarine sulcus
25 Colliculi of midbrain
26 Cerebral aqueduct
27 **Occipital lobe**
28 Mamillary body
29 Fourth ventricle
30 Vermis of cerebellum
31 Right hemisphere of cerebellum
32 Median aperture of Magendie (arrow)
33 Medulla oblongata
34 Olfactory tract
35 Optic nerve
36 **Internal carotid artery**
37 Interpeduncular cistern
38 Superior cerebellar artery
39 Anterior inferior cerebellar artery
40 **Vertebral artery**
41 Posterior inferior cerebellar artery
42 **Basilar artery**
43 Trigeminal nerve (n. V)
44 Facial nerve (n. VII)
45 Accessory nerve (n. XI), hypoglossal nerve (n. XII)
46 Cerebellum

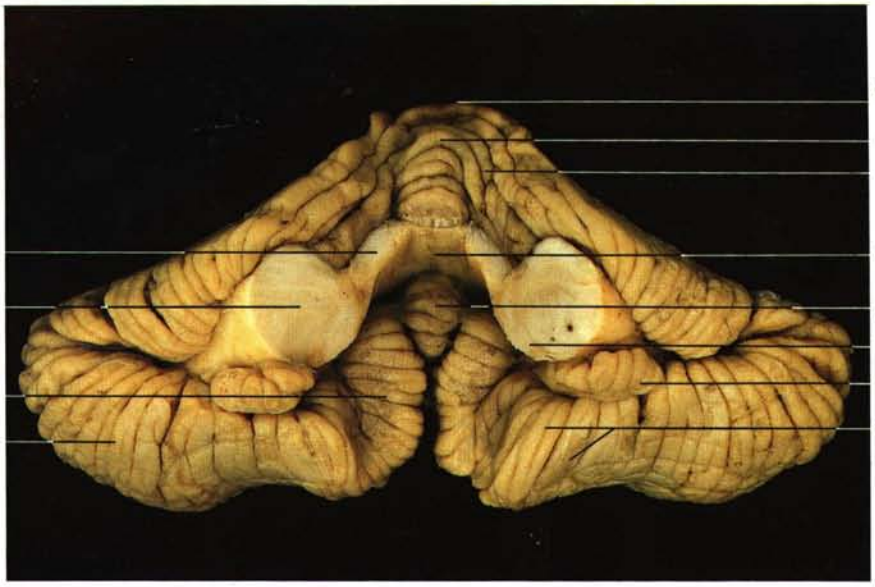

Cerebellum (inferior anterior aspect). The cerebellar peduncles have been severed.

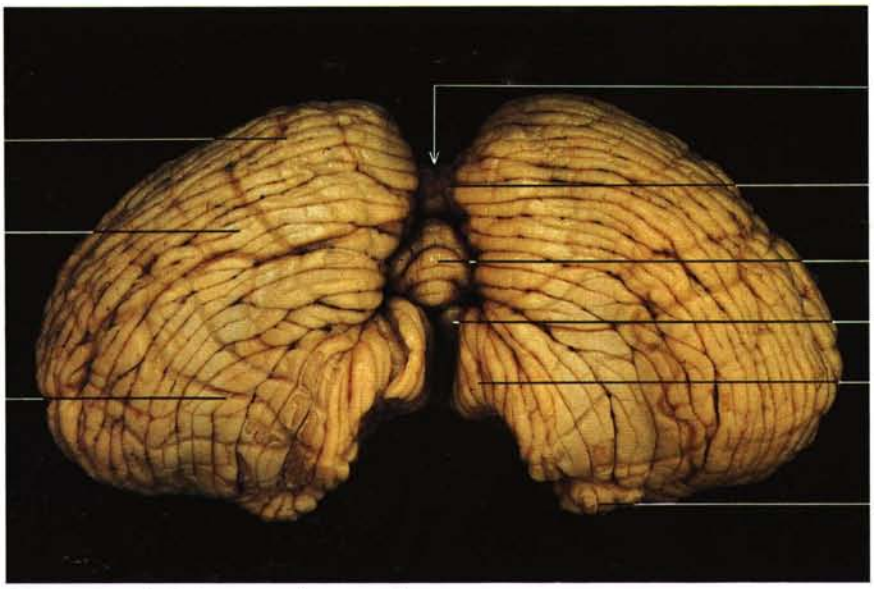

Cerebellum (inferior posterior aspect).

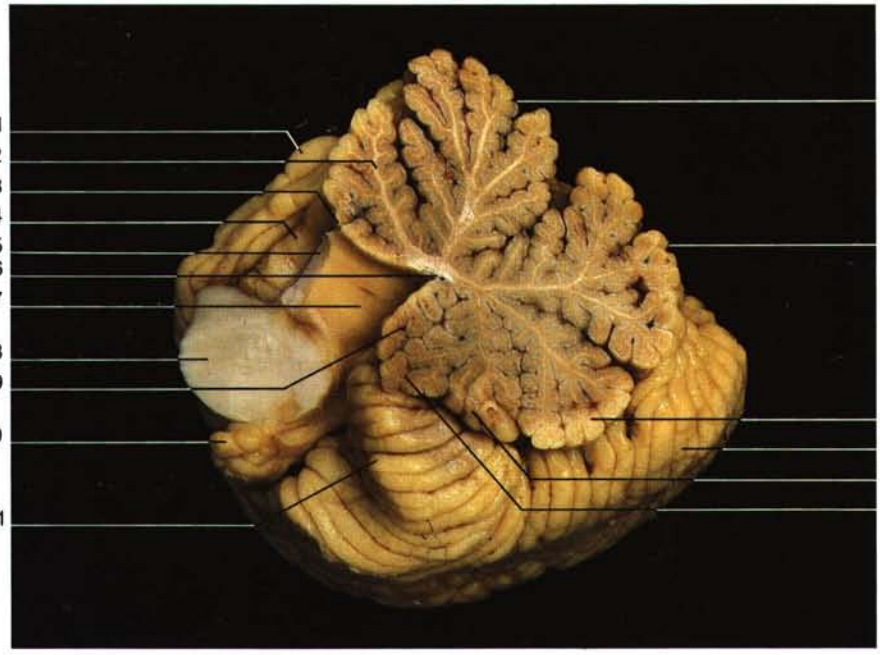

Median section through the cerebellum. Right cerebellar hemisphere and right half of vermis.

1 Superior cerebellar peduncle
2 Middle cerebellar peduncle
3 Cerebellar tonsil
4 Inferior semilunar lobule
5 **Vermis**
6 Central lobule of vermis
7 Inferior cerebellar peduncle
8 Superior medullary velum
9 **Nodule of vermis**
10 **Flocculus**
11 Biventral lobule
12 Left cerebellar hemisphere
13 Inferior semilunar lobule
14 Biventral lobule
15 Vermis of cerebellum
16 Tuber of vermis
17 Pyramid of vermis
18 Uvula of vermis
19 Tonsil of cerebellum
20 Floccule of cerebellum
21 Right cerebellar hemisphere
22 Vermis (central lobule)
23 Cerebellar lingula
24 Ala of central lobule
25 **Superior cerebellar peduncle**
26 Fastigium
27 **Fourth ventricle**
28 **Middle cerebellar peduncle**
29 Nodule of vermis
30 Flocculus of cerebellum
31 Cerebellar tonsil
32 Culmen of vermis
33 Declive of vermis
34 Tuber of vermis
35 Inferior semilunar lobule
36 Pyramis of vermis
37 Uvula of vermis

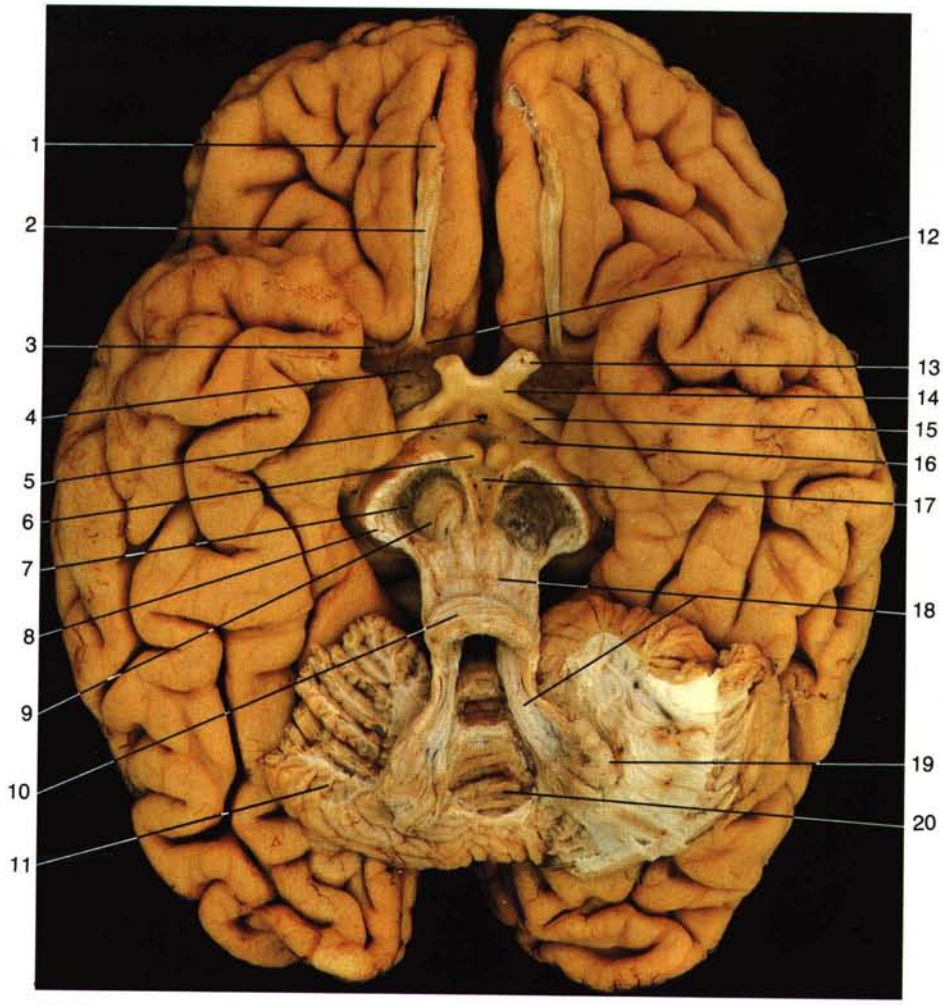

Brain and cerebellum (inferior aspect). Parts of the cerebellum have been removed to display the dentate nucleus and the main pathway to the midbrain (cerebellorubral tract).

Dissection of the cerebellar peduncles and their connection with midbrain and diencephalon. A small part of pulvinar thalami (*) has been cut to show inferior brachium.

1 Olfactory bulb
2 **Olfactory tract**
3 Lateral olfactory stria
4 Anterior perforated substance
5 Infundibulum (divided)
6 **Mamillary body**
7 Substantia nigra
8 Pedunculus cerebri
9 **Red nucleus**
10 Decussation of superior cerebellar peduncle
11 **Cerebellar hemisphere**
12 Medial olfactory stria
13 **Optic nerve**
14 Optic chiasma
15 Optic tract
16 Posterior perforated substance
17 Interpeduncular fossa
18 **Superior cerebellar peduncle** and cerebellorubral tract
19 **Dentate nucleus**
20 **Vermis of cerebellum**
21 Cingulate gyrus
22 **Corpus callosum**
23 Stria terminalis
24 Septum pellucidum
25 **Columna fornicis**
26 Cerebral peduncle at midbrain level
27 Pons
28 Inferior olive
29 Medulla oblongata with lateral pyramidal tract
30 Occipital lobe
31 Calcarine sulcus
32 **Thalamus**
33 Inferior colliculus with brachium
34 **Medial lemniscus**
35 **Superior cerebellar peduncle**
36 **Inferior cerebellar peduncle**
37 **Middle cerebellar peduncle**
38 Cerebellar hemisphere

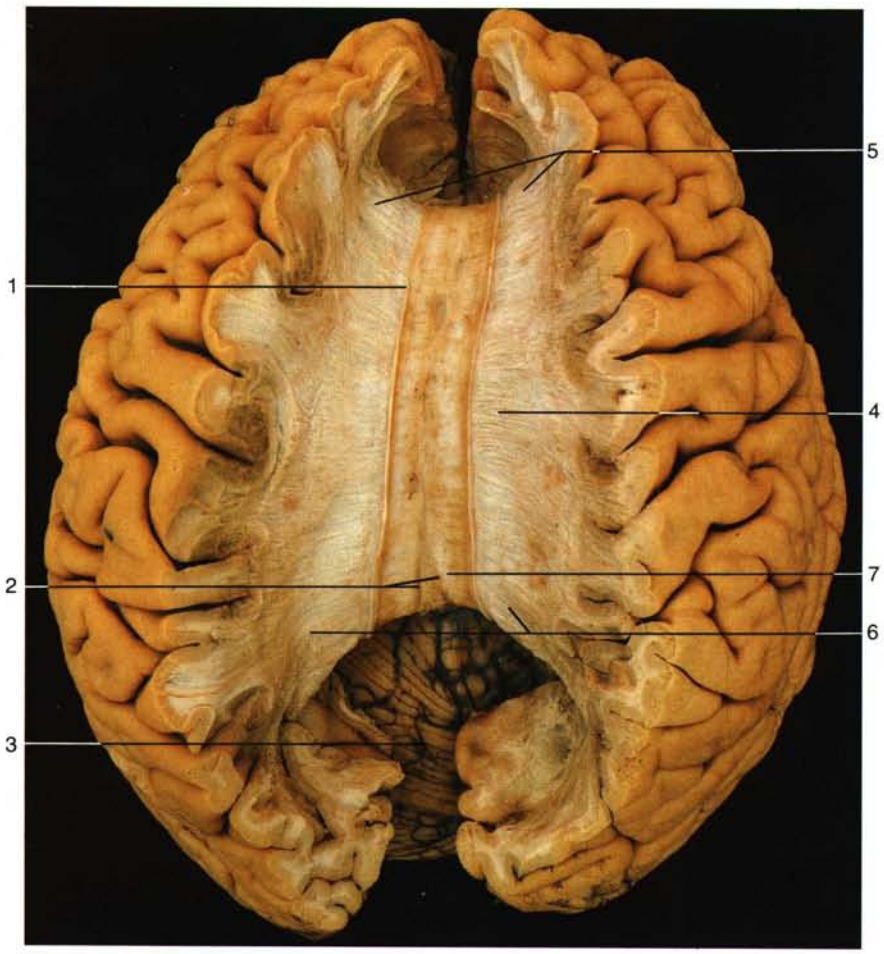

1 Lateral longitudinal stria of indusium griseum
2 Medial longitudinal stria of indusium griseum
3 Cerebellum
4 Radiating fibers of the corpus callosum
5 Forceps minor of corpus callosum
6 Forceps major of corpus callosum
7 Splenium of corpus callosum

Dissection of the brain I. The fiber system of the corpus callosum has been displayed by removing the cortex lying above it. Frontal pole at top.

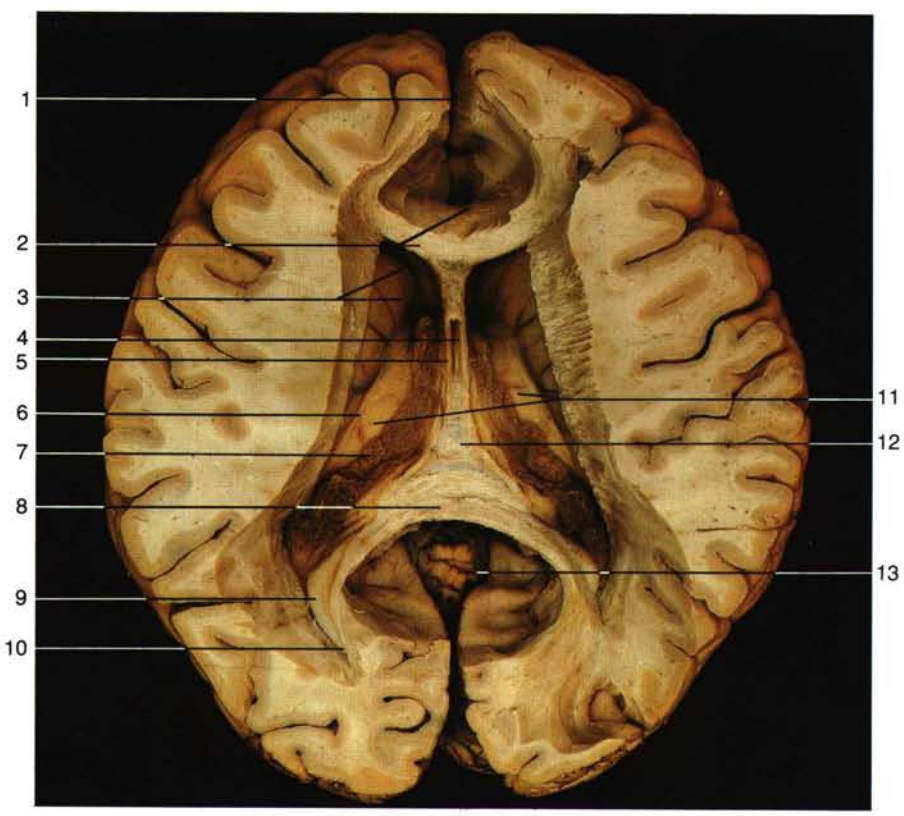

1 Longitudinal cerebral fissure
2 Genu of corpus callosum
3 Head of **caudate nucleus** and anterior horn of lateral ventricle
4 Cavum of septum pellucidum
5 **Septum pellucidum**
6 Stria terminalis
7 **Choroid plexus** of **lateral ventricle**
8 Splenium of **corpus callosum**
9 Calcar avis
10 Posterior horn of lateral ventricle
11 **Thalamus** (lamina affixa)
12 Commissure of fornix
13 Vermis of cerebellum

Dissection of the brain II. The lateral ventricles and subcortical nuclei of the brain are dissected. The corpus callosum has been partly removed. Frontal pole at the top.

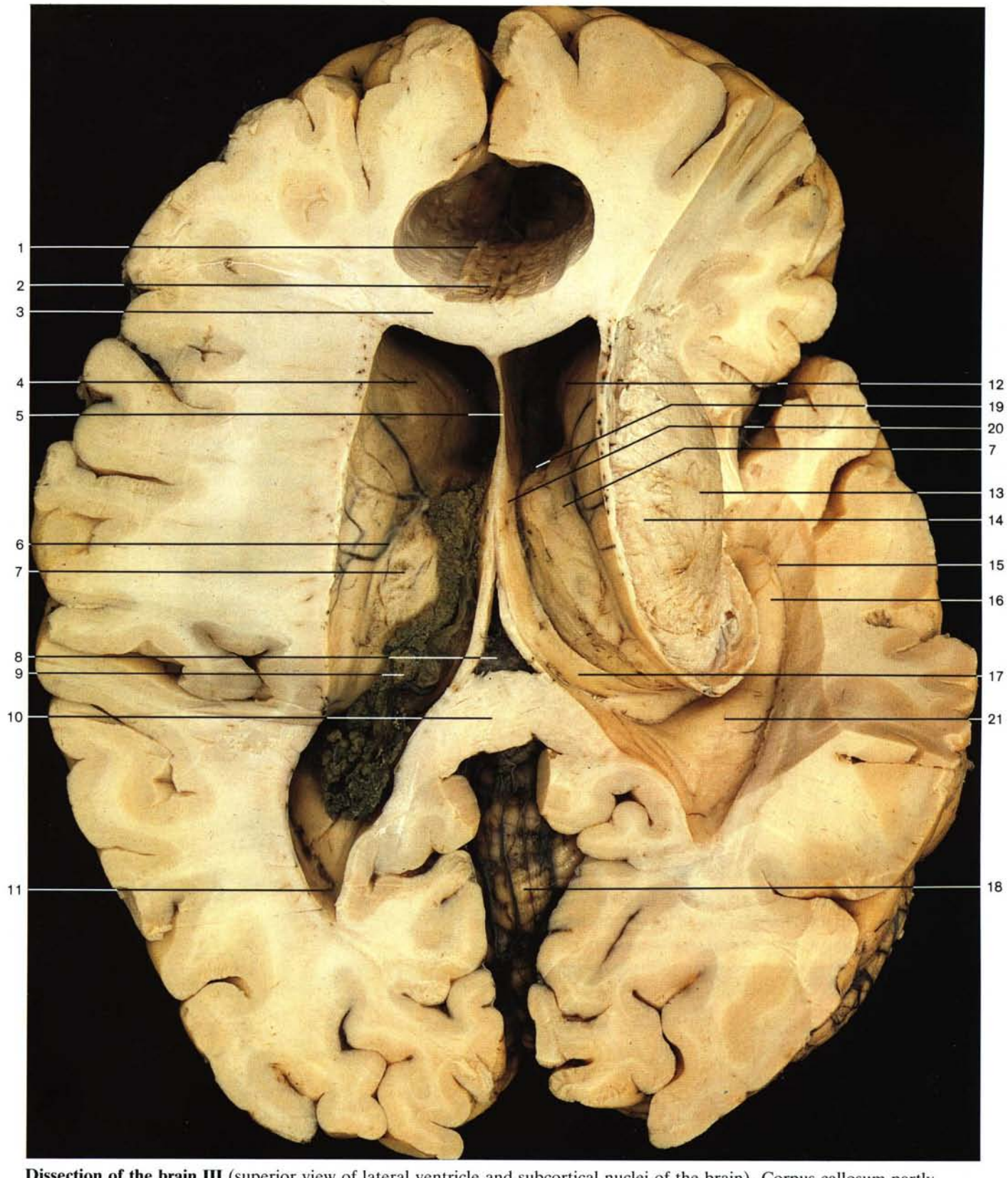

Dissection of the brain III (superior view of lateral ventricle and subcortical nuclei of the brain). Corpus callosum partly removed. At right, the entire lateral ventricle has been opened, the insula with claustrum, extreme and external capsules have been removed, exposing the lentiform nucleus and the internal capsule.

1 Lateral longitudinal stria
2 Medial longitudinal stria
3 Genu of corpus callosum
4 Head of **caudate nucleus**
5 **Septum pellucidum**
6 Stria terminalis
7 **Thalamus** (lamina affixa)
8 Choroid plexus of third ventricle
9 **Choroid plexus** of **lateral ventricle**
10 Splenium of corpus callosum
11 Posterior horn of lateral ventricle
12 Anterior horn of lateral ventricle (head of caudate nucleus)
13 **Putamen** of **lentiform nucleus**
14 **Internal capsule**
15 Inferior horn of lateral ventricle
16 **Pes hippocampi**
17 **Crus of fornix**
18 Vermis of cerebellum with arachnoid and pia mater
19 Interventricular foramen
20 Right column of fornix
21 Collateral eminence

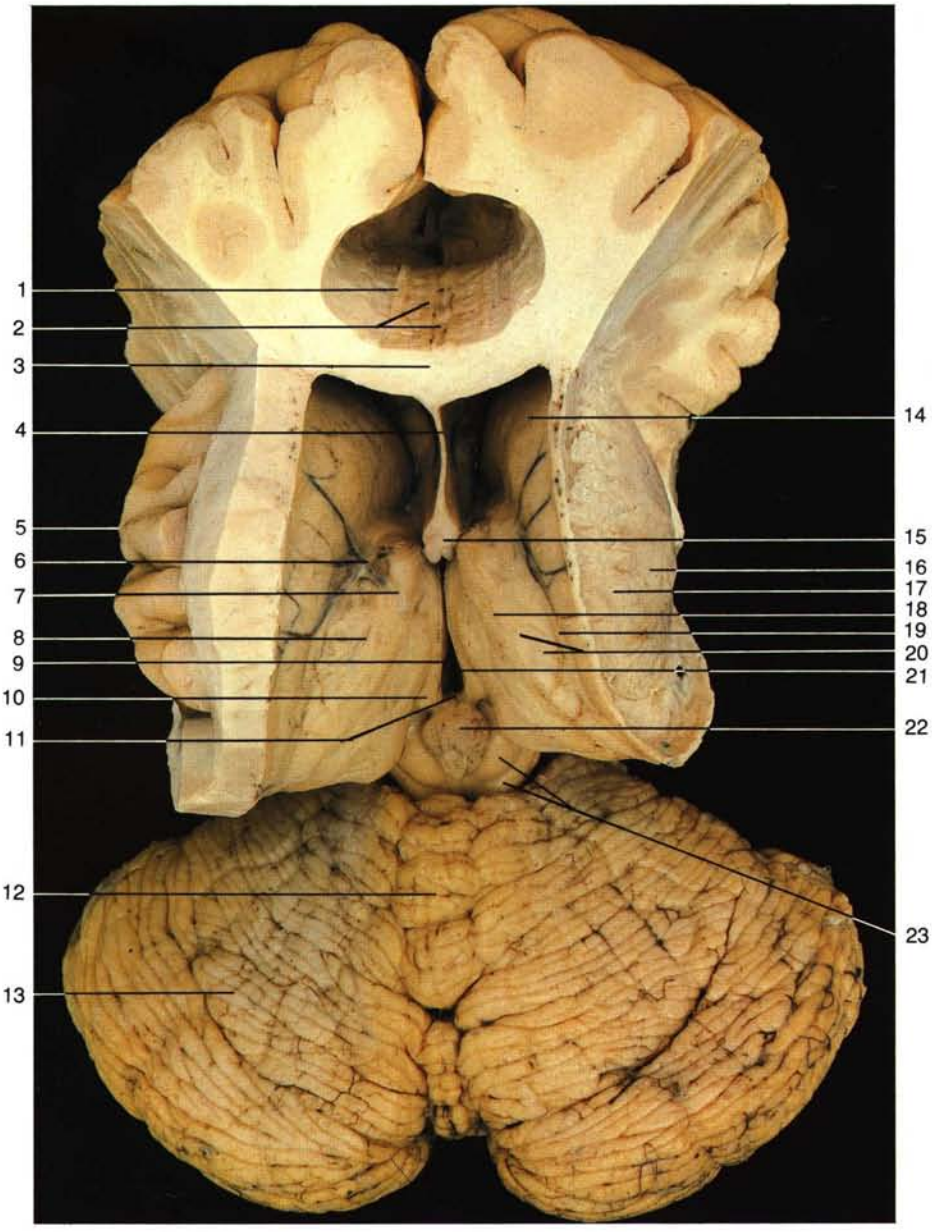

1 Lateral longitudinal stria
2 Medial longitudinal stria
3 **Corpus callosum**
4 Septum pellucidum
5 Insular gyri
6 Thalamostriate vein
7 Anterior tubercle of thalamus
8 **Thalamus**
9 Medullary stria of thalamus
10 Habenular trigone
11 Habenular commissure
12 **Vermis of the cerebellum**
13 Left **hemisphere of cerebellum**
14 **Head of caudate nucleus**
15 **Columns of fornix**
16 Putamen of lentiform nucleus
17 **Internal capsule**
18 Taenia of choroid plexus
19 Stria terminalis and thalamostriate vein
20 Lamina affixa
21 **Third ventricle**
22 **Pineal body**
23 Superior and inferior colliculus of midbrain

Dissection of the brain IVa. Temporal lobe, fornix and the posterior corpus callosum have been removed (this part of the specimen is depicted below). Frontal pole at top (superior aspect).

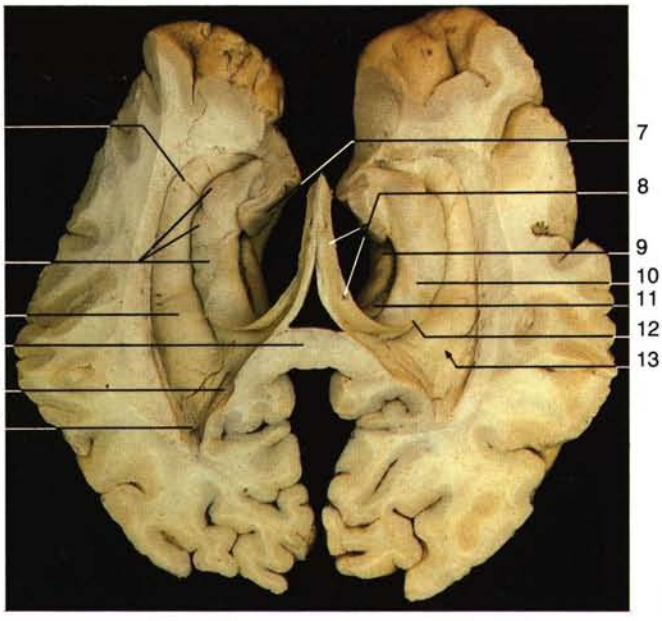

1 Inferior horn of lateral ventricle
2 **Hippocampal digitations**
3 Collateral eminence
4 Splenium of corpus callosum
5 Calcar avis
6 Posterior horn of lateral ventricle
7 **Uncus** of parahippocampal gyrus
8 Body and crus of **fornix**
9 **Parahippocampal gyrus**
10 **Pes hippocampi**
11 Dentate gyrus
12 Hippocampal fimbria
13 Lateral ventricle

Dissection of the brain IVb. Depicted is the portion of the brain removed from the specimen above. **Temporal lobe and limbic system** (superior aspect). Columns of fornix are served.

Dissection of the limbic system. Left side, lateral aspect. Corpus callosum has been cut in the median plane. The left thalamus and the left hemisphere have been partly removed.

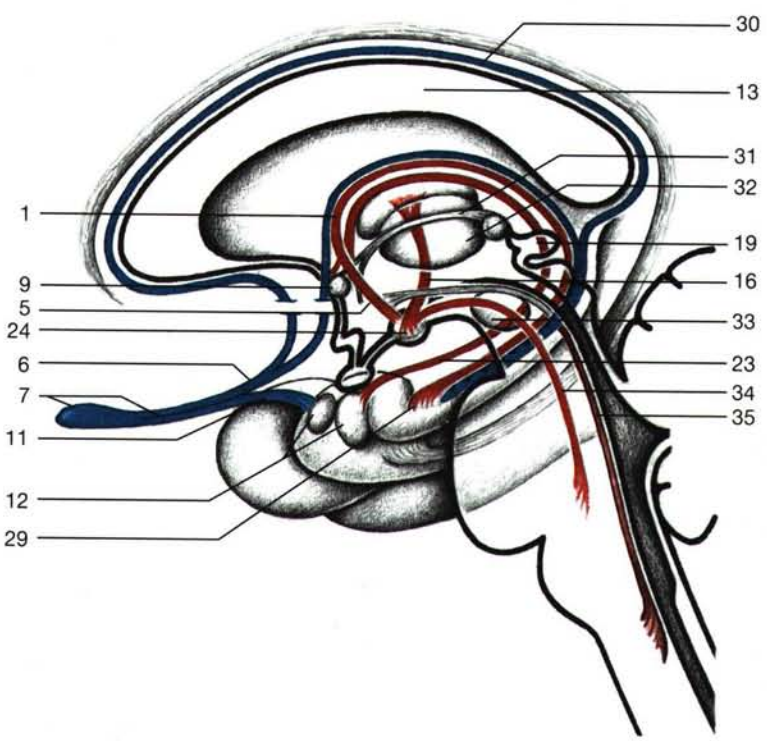

Main pathways of **limbic and olfactory system.** (Schematic drawing.) Blue = afferent pathways; red = efferent pathways.

1 **Body of fornix**
2 Septum pellucidum
3 **Lateral longitudinal stria**
4 Genu of corpus callosum
5 **Column of fornix**
6 Medial olfactory stria
7 **Olfactory bulb** and **olfactory tract**
8 Optic nerve
9 Anterior commissure (left half)
10 Right temporal lobe
11 **Lateral olfactory stria**
12 **Amygdala**
13 Body of **corpus callosum**
14 Interthalamic adhesion
15 Third ventricle and right thalamus
16 **Mamillothalamic tract**
17 Part of the thalamus
18 Habenular commissure
19 Pineal body
20 Splenium of corpus callosum
21 Colliculi of midbrain
22 Vermis of cerebellum
23 **Stria terminalis**
24 **Mamillary body**
25 Fimbria of hippocampus and pes hippocampi
26 Left optic tract and lateral geniculate body
27 Lateral ventricle and **parahippocampal gyrus**
28 Collateral eminence
29 Hippocampal digitations
30 **Supracallosal gyrus** (longitudinal stria)
31 Stria medullaris thalami
32 Thalamus
33 Red nucleus
34 Mamillotegmental tract
35 Dorsal longitudinal fasciculus (Schütz)

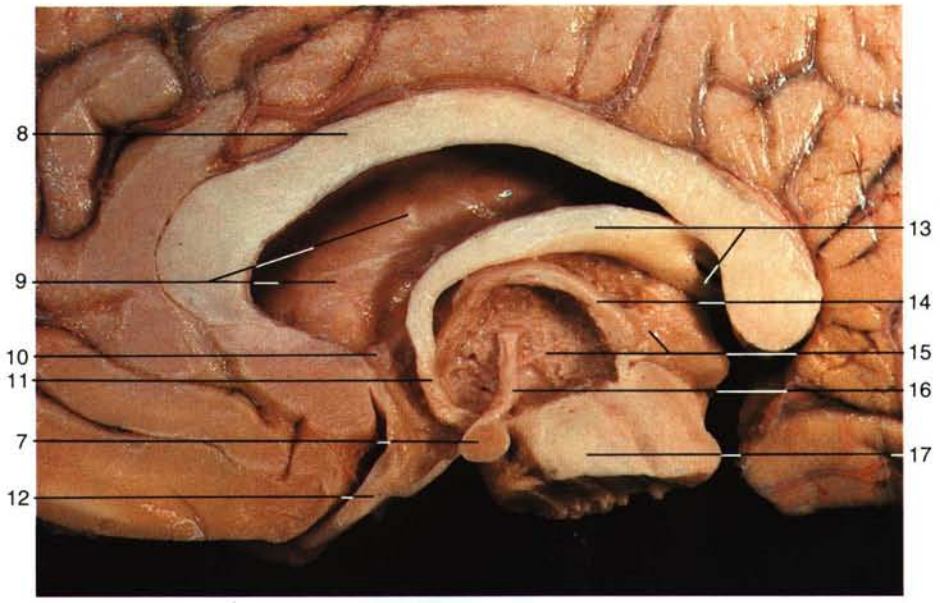

Median section through the diencephalon. Medial part of the thalamus and septum pellucidum have been removed to show the fornix and mamillothalamic tract.

1 Paraventricular nucleus
2 Preoptic nucleus
3 Ventromedial nucleus
4 Supraoptic nucleus
5 Posterior nucleus
6 Dorsomedial nucleus
} **Hypothalamic nuclei**

7 **Mamillary body**
8 **Corpus callosum**
9 Lateral ventricle (showing caudate nucleus)
10 Anterior commissure
11 **Column of fornix**
12 Optic chiasma
13 **Crus of fornix**
14 Medullary stria of thalamus
15 Thalamus and interthalamic adhesion
16 **Mamillothalamic tract** of Vicq d'Azyr
17 Cerebral peduncle
18 **Pineal body**
19 Tectum of midbrain
20 Lamina terminalis

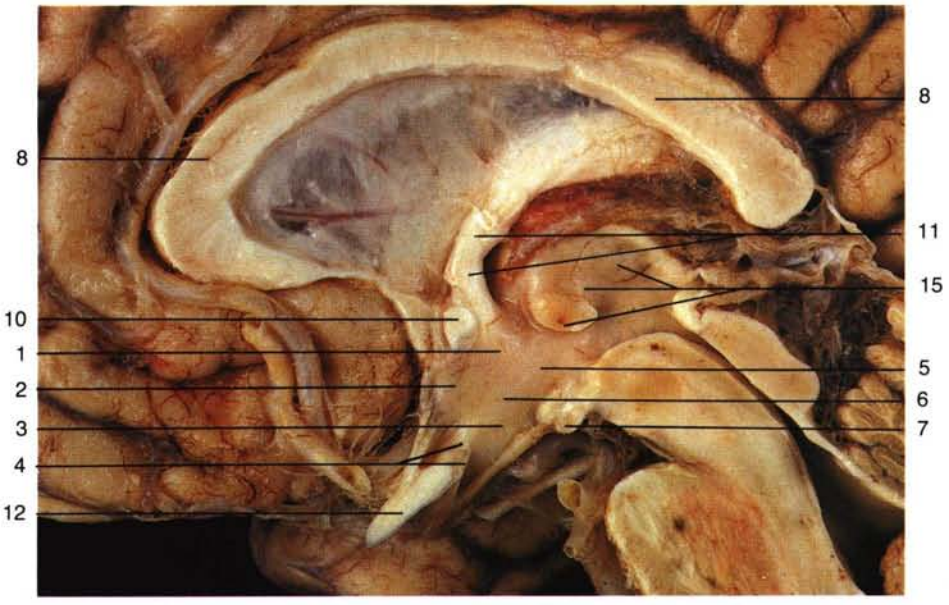

Median section through the diencephalon and midbrain; location of hypothalamic nuclei.

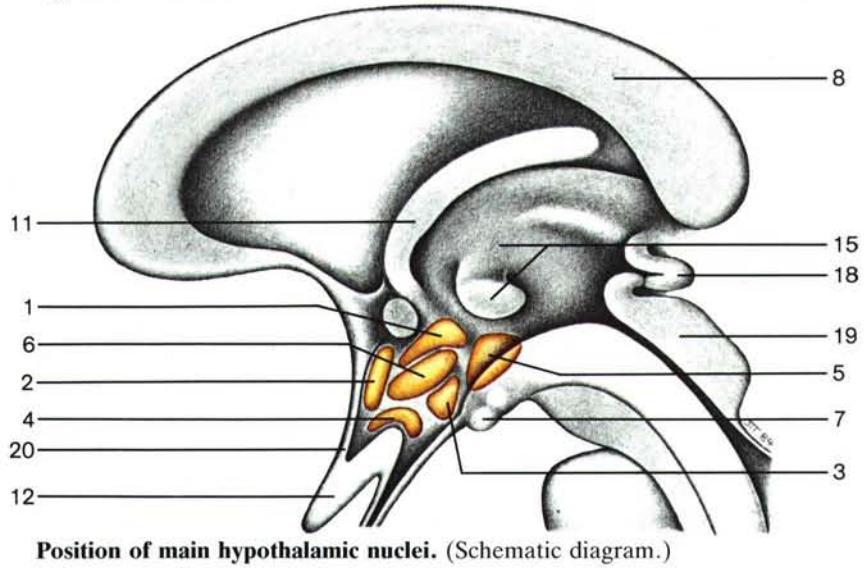

Position of main hypothalamic nuclei. (Schematic diagram.)

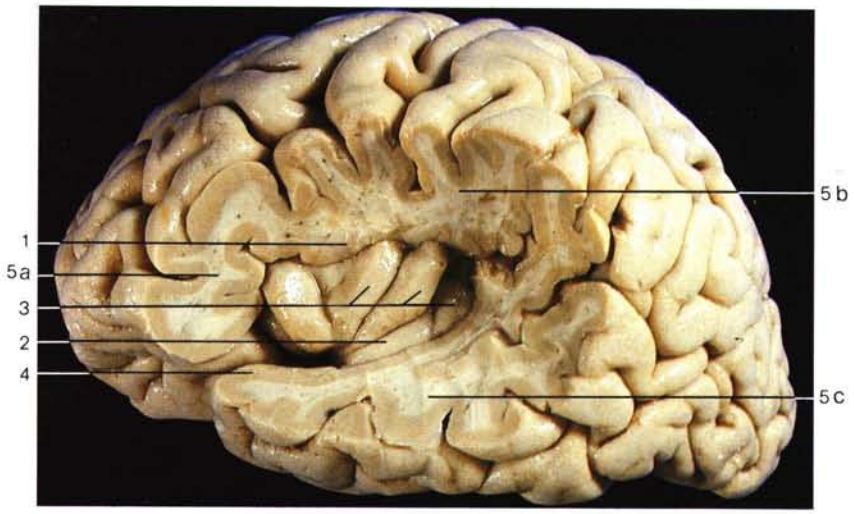

Insula (Reili). The opercula of the frontal, parietal and temporal lobes have been removed to display the insular gyri. Left hemisphere.

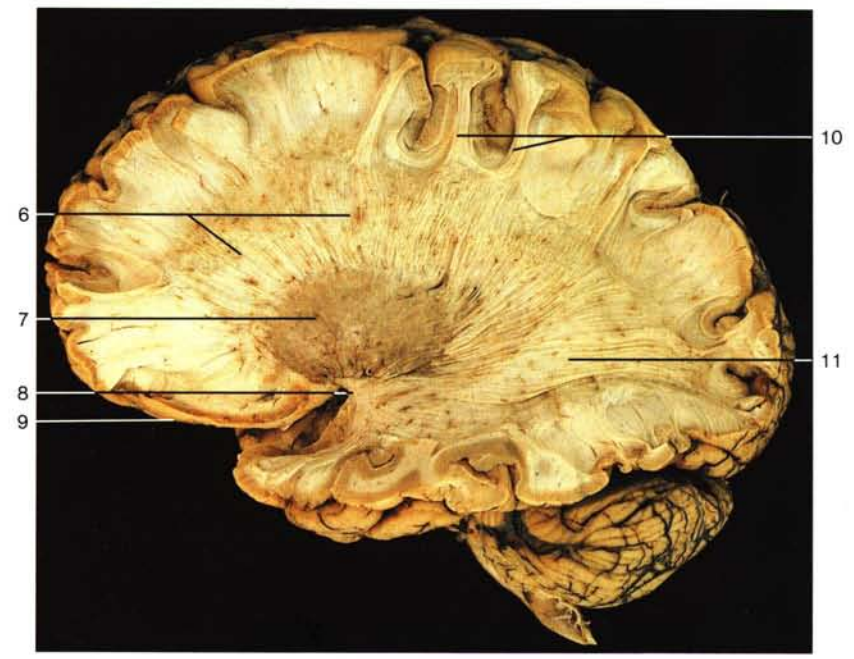

Dissection of the corona radiata, left hemisphere. Frontal pole on the left.

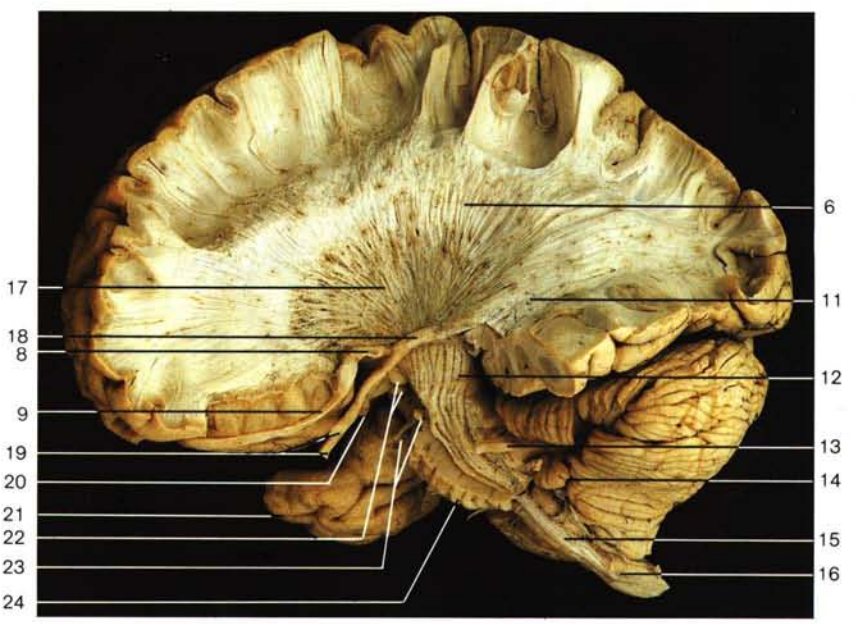

1 Circular sulcus of insula
2 Long gyrus of insula
3 Short gyri of insula
4 Limen insulae
5 Opercula (cut)
a Frontal operculum
b Frontoparietal operculum
c Temporal operculum
6 **Corona radiata**
7 **Lentiform nucleus**
8 **Anterior commissure**
9 Olfactory tract
10 Cerebral arcuate fibers
11 Optic radiation
12 Cerebral peduncle
13 Trigeminal nerve (n. V)
14 Flocculus of cerebellum
15 **Pyramidal tract**
16 Decussation of pyramidal tract
17 Internal capsule
18 Optic tract
19 Optic nerve (n. II)
20 Infundibulum
21 Temporal lobe (right side)
22 Mamillary bodies
23 Oculomotor nerve (n. III)
24 Transverse fibers of pons

◁ **Corona radiata and internal capsule,** left hemisphere. Lentiform nucleus removed (frontal pole to the left).

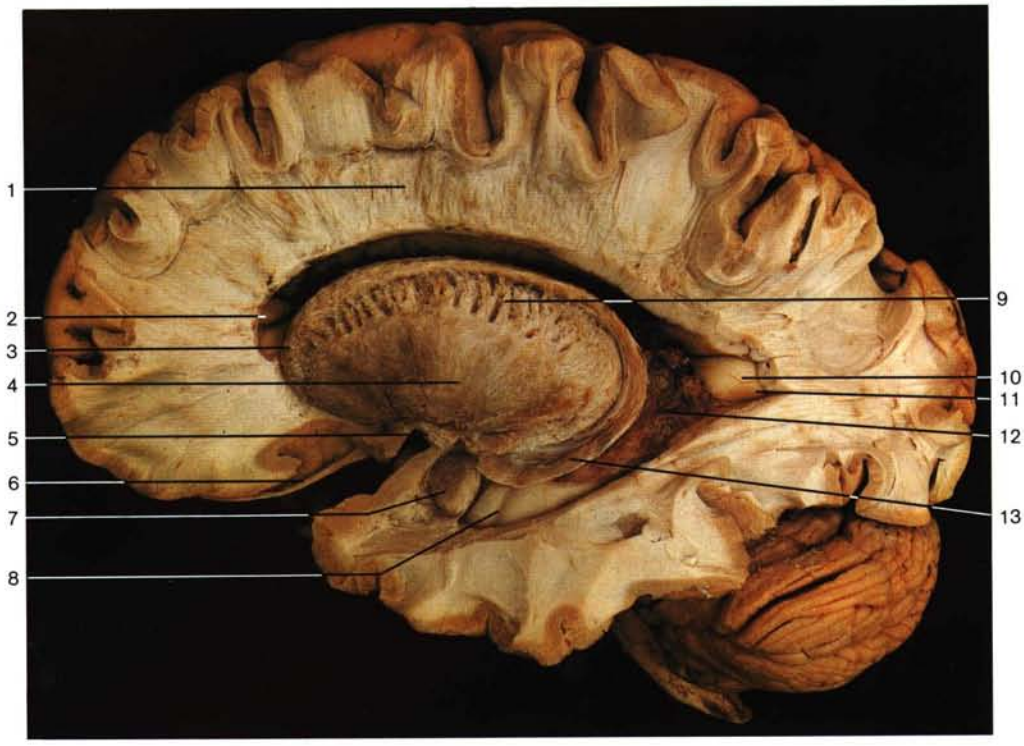

1 Corona radiata
2 Anterior horn of lateral ventricle
3 Head of **caudate nucleus**
4 **Putamen**
5 Anterior commissure
6 Olfactory tract
7 **Amygdala**
8 Hippocampal digitations
9 **Internal capsule**
10 Calcar avis
11 Posterior horn of lateral ventricle
12 Choroid plexus of lateral ventricle
13 Caudal extremity of caudate nucleus
14 **Thalamus**
15 Cerebral arcuate fibers
16 Globus pallidus (remnants)

Dissection of the subcortical nuclei and internal capsule, left hemisphere (lateral aspect). Frontal pole to the left. The lateral ventricle has been opened, and the insular gyri and claustrum have been removed, revealing the lentiform nucleus and the internal capsule.

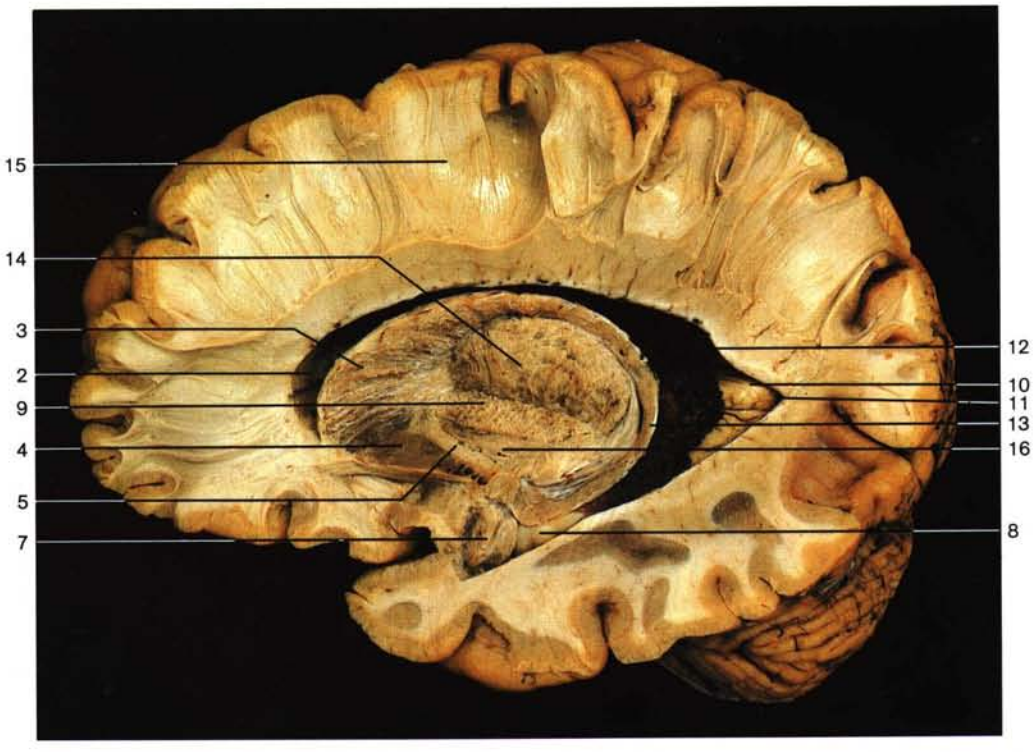

Dissection of the subcortical nuclei (lateral aspect). Lentiform nucleus removed, frontal pole to the left.

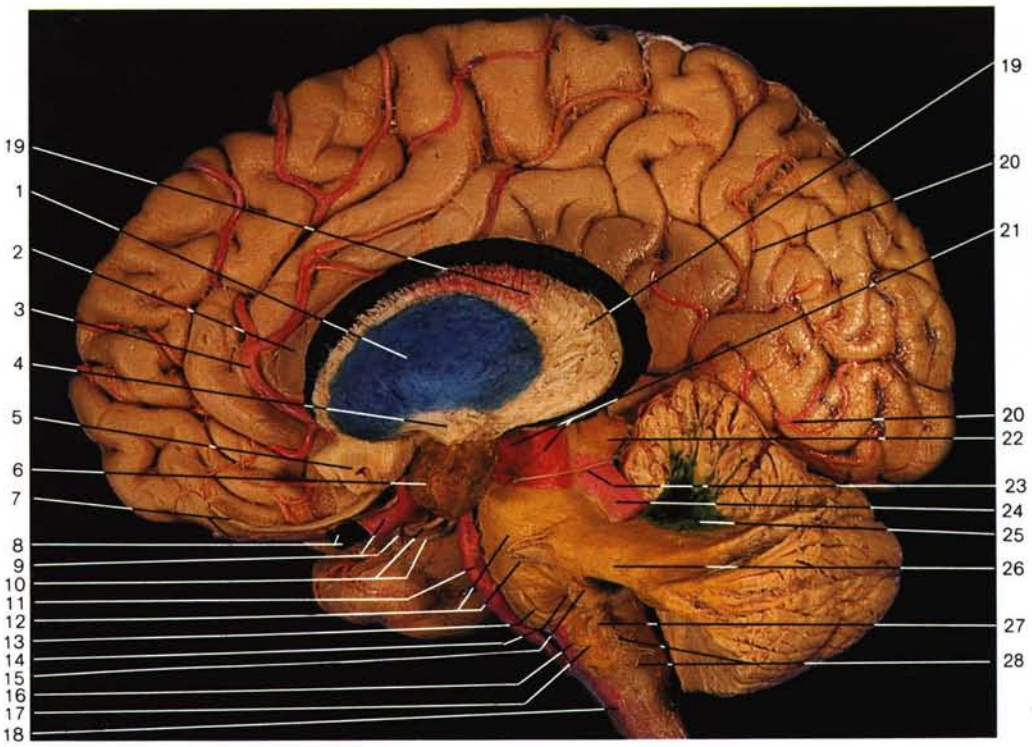

Brain stem and the connections with the cerebellum. Internal capsule (lateral aspect). Red = pyramidal tract; yellow = middle cerebellar peduncle; green = inferior cerebellar peduncle; pink = superior cerebellar peduncle.

1 **Putamen**
2 Genu of corpus callosum
3 Anterior cerebral artery
4 Anterior commissure
5 Subcallosal area
6 Amygdala
7 Olfactory tract
8 Optic nerve (n. II)
9 Internal carotid artery and infundibulum
10 Oculomotor nerve (right and left nerve, n. III)
11 **Basilar artery**
12 Pons and trigeminal nerve (n. V)
13 Abducens nerve (n. VI)
14 Facial nerve (n. VII)
15 Vestibulocochlear nerve (n. VIII)
16 Hypoglossal nerve (n. XII)
17 Olive
18 Pyramidal tract
19 **Internal capsule**
20 Posterior cerebral artery
21 Cerebral peduncle
22 **Colliculi of midbrain**
23 Trochlear nerve (n. IV)
24 Superior cerebellar peduncle
25 Inferior cerebellar peduncle
26 **Middle cerebellar peduncle**
27 Glossopharyngeal nerve (n. IX)
28 Vagus (n. X) and accessory nerves (n. XI)
29 **Corpus callosum**
30 Lateral ventricle (anterior horn)
31 Caudate nucleus
32 Internal capsule (anterior limb)
33 Insula
34 **Claustrum**
35 **Thalamus**
36 Superior and inferior colliculus of midbrain
37 Cerebellum
38 Middle peduncle of cerebellum (efferent tracts)
39 Medulla oblongata (afferent tracts)
40 Putamen of lentiform nucleus
41 **Globus pallidus** of lentiform nucleus
42 Genu and posterior limb of internal capsule

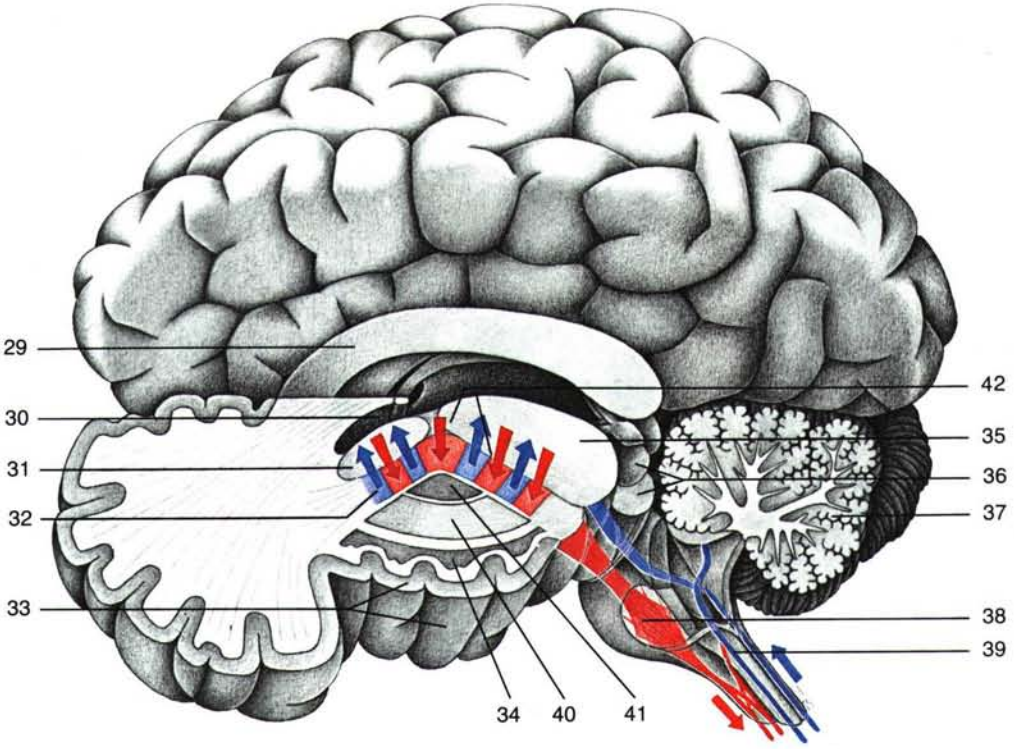

Internal capsule and subcortical nuclei. Left brain, frontal pole to the left, horizontal section. (Semischematic drawing.) Blue = afferent tracts; red = efferent tracts.

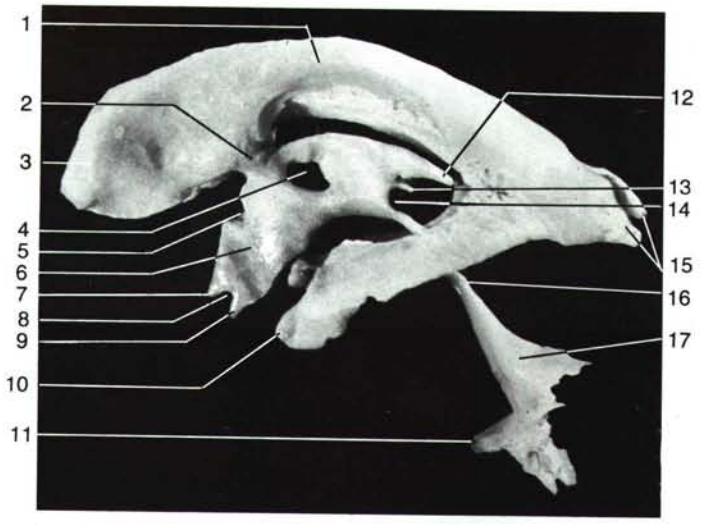

Cast of ventricular cavities of the brain (lateral aspect), frontal pole to the left.

1 Central part of the **lateral ventricle**
2 **Interventricular foramen** of Monro
3 Anterior horn of the lateral ventricle
4 Site of interthalamic adhesion
5 Notch for anterior commissure
6 **Third ventricle**
7 Optic recess
8 Notch for optic chiasma
9 Infundibular recess
10 Inferior horn of lateral ventricle with indentation of amygdaloid body
11 Lateral recess and lateral aperture of Luschka
12 Suprapineal recess
13 Pineal recess
14 Notch for posterior commissure
15 Posterior horn of lateral ventricle
16 **Cerebral aqueduct**
17 **Fourth ventricle**
18 Median aperture of Magendie
19 **Cerebellomedullary cistern**
20 Superior sagittal sinus
21 Inferior sagittal sinus
22 Intervaginal space of optic nerve
23 Arachnoid granulations of Pacchioni
24 Straight sinus

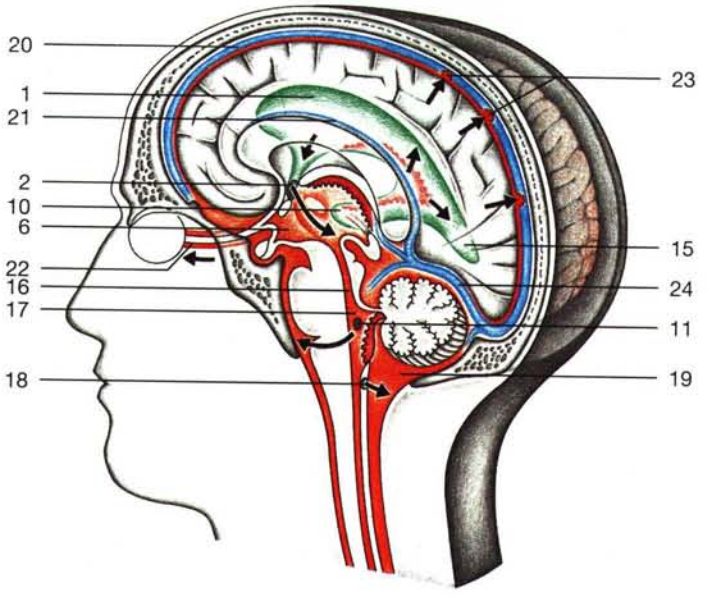

◁ **Position of ventricular cavities.** (Schematic drawing.) The direction of flow of cerebrospinal fluid is indicated by arrows.
Green = Right lateral ventricle
Red = Choroidal plexus

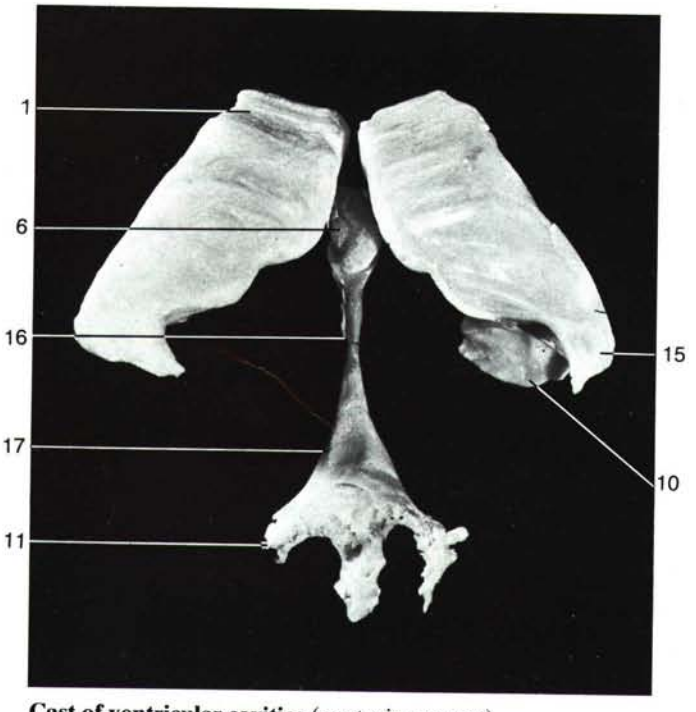

Cast of ventricular cavities (posterior aspect).

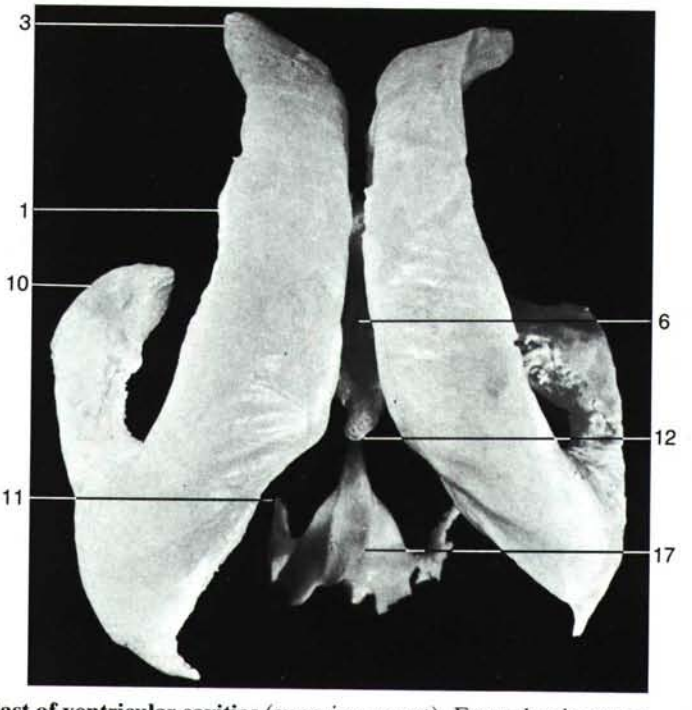

Cast of ventricular cavities (superior aspect). Frontal pole at top.

Brain stem (left lateral aspect). Cerebellar peduncles have been severed, cerebellum and cerebral cortex have been removed.

Brain stem (dorsal aspect). Cerebellum removed.

1 **Internal capsule**
2 Head of the caudate nucleus
3 Olfactory trigone
4 Olfactory tracts
5 Optic nerves
6 Infundibulum
7 Oculomotor nerve
8 **Amygdaloid body**
9 **Pons**
10 Trigeminal nerve
11 Facial and vestibulocochlear nerves
12 Hypoglossal nerve
13 Glossopharyngeal and vagus nerves
14 **Olive**
15 Medulla oblongata
16 **Lentiform nucleus**
17 **Anterior commissure**
18 Tail of caudate nucleus
19 Superior colliculus
20 Inferior colliculus
21 Trochlear nerve
22 Superior cerebellar peduncle
23 Inferior cerebellar peduncle
24 **Middle cerebellar peduncle**
25 Accessory nerve (n. XI)
26 Columns of fornix (divided)
27 Lamina affixa
28 **Third ventricle**
29 Pulvinar of thalamus
30 Inferior brachium
31 Frenulum veli
32 Superior medullary velum
33 Facial colliculus
34 Striae medullares and rhomboid fossa
35 Hypoglossal triangle
36 Stria terminalis and thalamostriate vein
37 Habenular trigone
38 Choroid plexus of lateral ventricle
39 **Pineal body**
40 Medial geniculate body
41 Cerebral peduncle
42 Choroid plexus of fourth ventricle
43 Clava
44 Dorsal root of cervical nerve
45 Cuneate tubercle
46 Lateral geniculate body

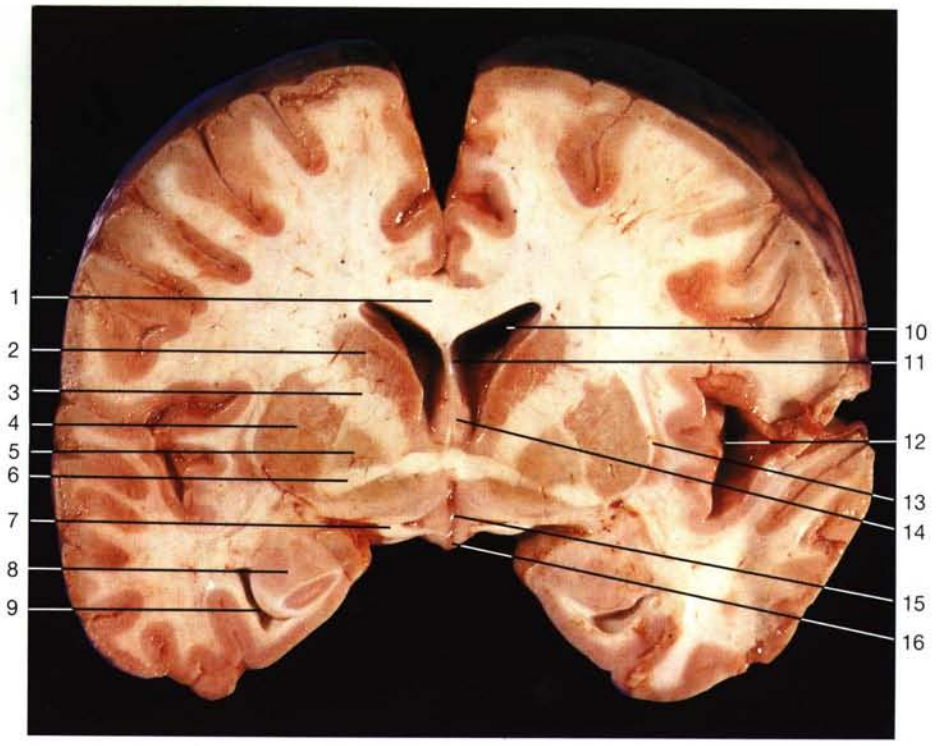

Coronal section through the brain at the level of the anterior commissure. Section 1.

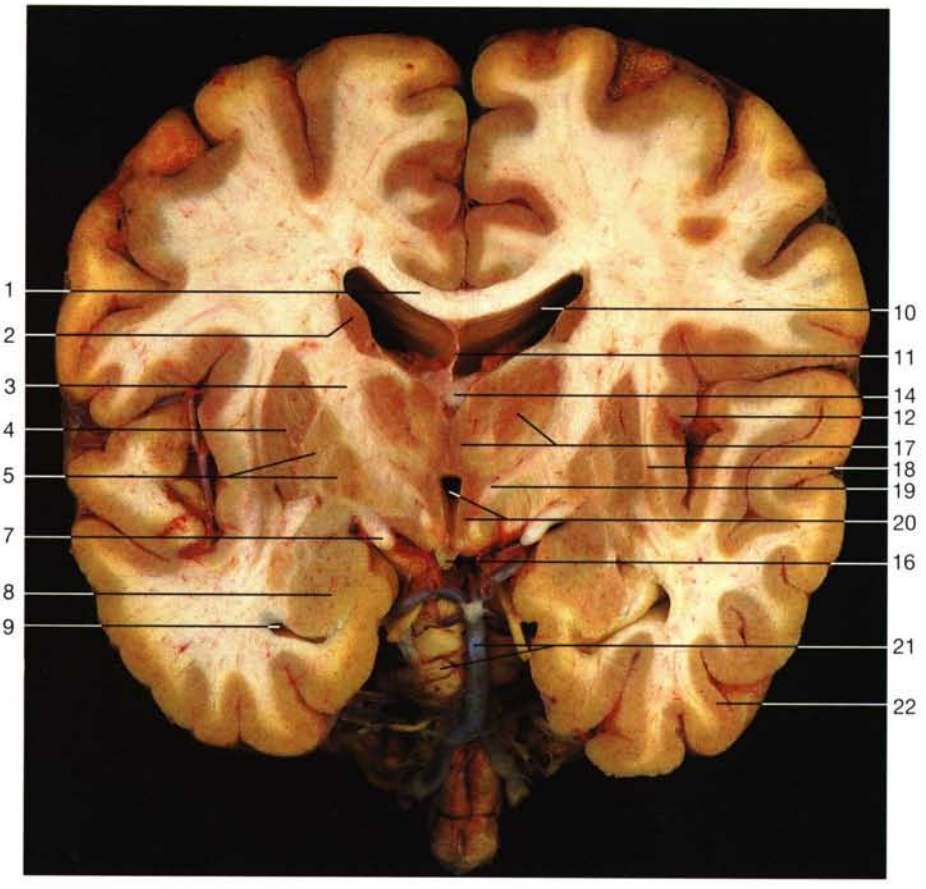

Coronal section through the brain at the level of the third ventricle and the interthalamic adhesion. Section 2.

1 Corpus callosum
2 Head of **caudate nucleus**
3 **Internal capsule**
4 **Putamen**
5 **Globus pallidus**
6 Anterior commissure
7 Optic tract
8 Amygdaloid body
9 Inferior horn of lateral ventricle
10 **Lateral ventricle**
11 Septum pellucidum
12 Lobus insularis (insula)
13 External capsule
14 Column of fornix
15 Optic recess
16 Infundibulum
17 **Thalamus**
18 Claustrum
19 Lenticular ansa
20 **Third ventricle** and **hypothalamus**
21 Basilar artery and pons
22 Cortex of temporal lobe
23 Inferior colliculus
24 Superior colliculus
25 Cerebral aqueduct
26 Red nucleus
27 Substantia nigra
28 Cerebral peduncle
29 Trochlear nerve (n. IV)
30 Gray matter
31 Nucleus of oculomotor nerve (nucleus of Edinger-Westphal)
32 Fibers of oculomotor nerve (n. III)
33 Vermis of cerebellum
34 Fourth ventricle
35 Reticular formation
36 Pons and transverse pontine fibers
37 Emboliform nucleus
38 Dentate nucleus
39 Middle cerebellar peduncle
40 Choroid plexus
41 Hypoglossal nucleus at rhomboid fossa
42 Medial longitudinal fasciculus
43 Trigeminal nerve (n. V.)
44 Inferior olivary nucleus
45 Corticospinal fibers and arcuate fibers
46 Fourth ventricle with choroid plexus
47 Vestibular nuclei
48 Nucleus and tractus solitarius
49 Inferior cerebellar peduncle (restiform body)
50 Reticular formation
51 Medial lemniscus
52 Cuneate nucleus of Burdach
53 Central canal
54 **Pyramidal tract**
55 Flocculus of cerebellum
56 Cerebellar hemisphere with pia mater
57 "Arbor vitae" of cerebellum
58 Nucleus gracilis of Goll
59 Lateral recess of choroid plexus of IVth ventricle
60 Posterior inferior cerebellar artery
61 Choroid plexus of lateral ventricle

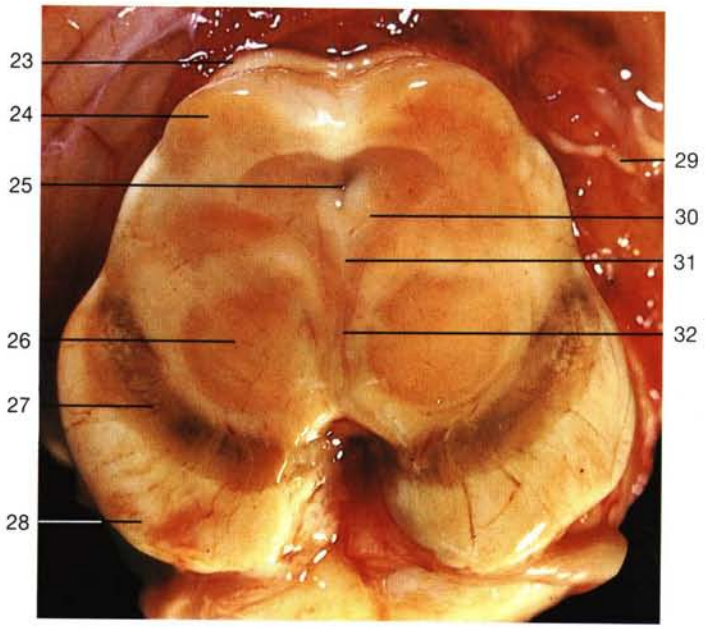

Cross-section of the midbrain (mesencephalon) at the level of the superior colliculus (superior aspect). Section 4.

Cross-section of the rhombencephalon at the level of the olive (inferior aspect). Section 6.

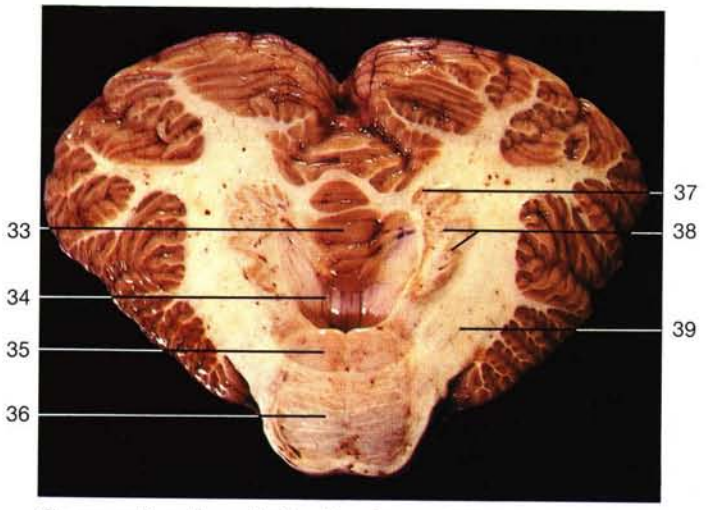

Cross-section through the rhombencephalon at the level of pons (inferior aspect). Section 5.

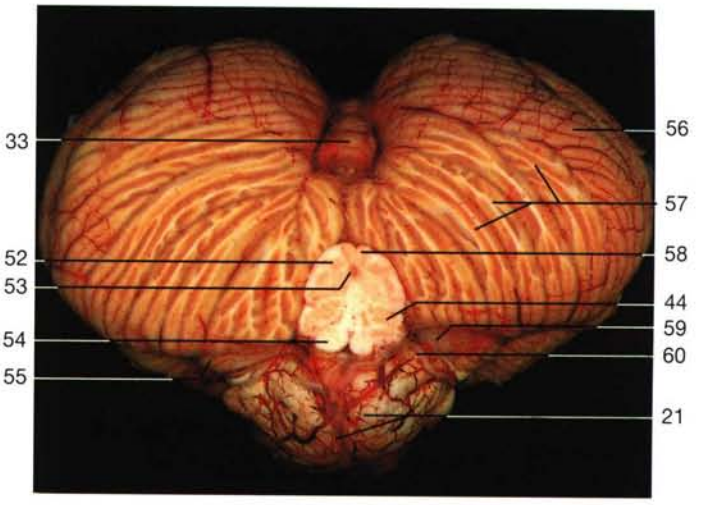

Cross-section through medulla oblongata and cerebellum (inferior aspect). Section 7.

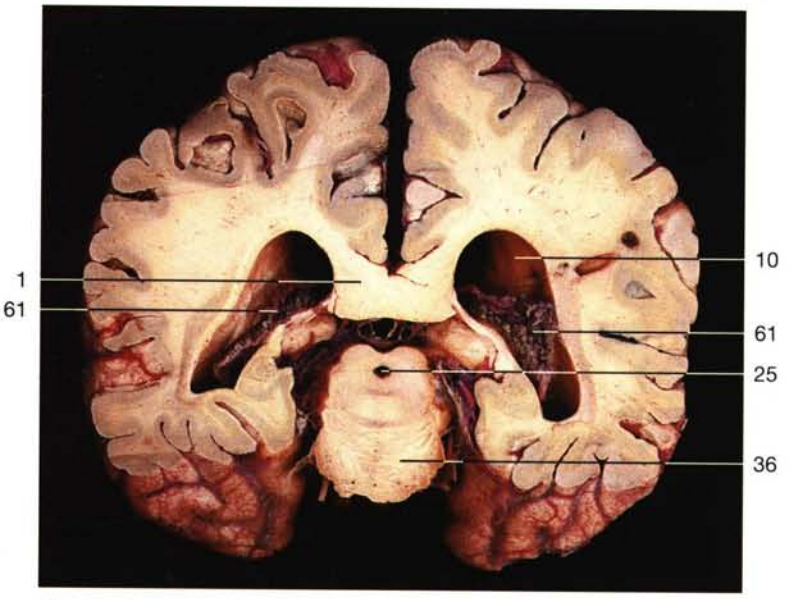

Coronal section at the level of **inferior colliculus** (posterior aspect). Section 3.

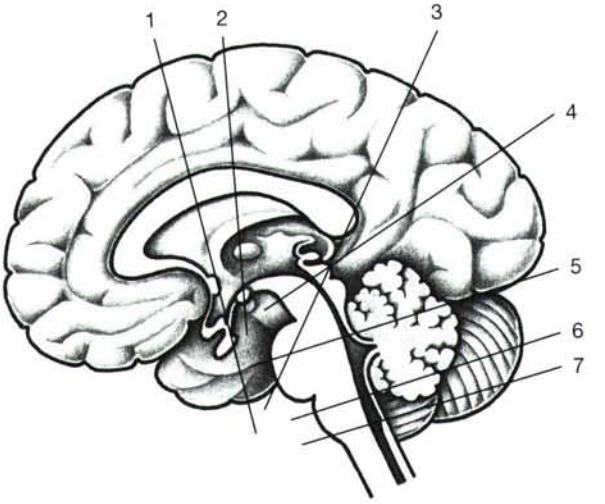

Right half of the brain. Levels of the sections are indicated.

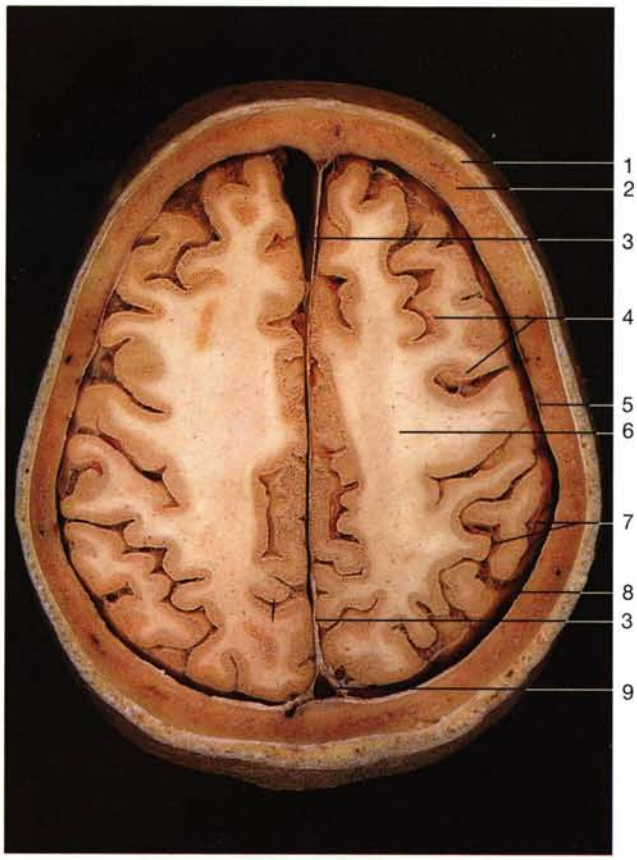

Horizontal section through the head.
Section 1.

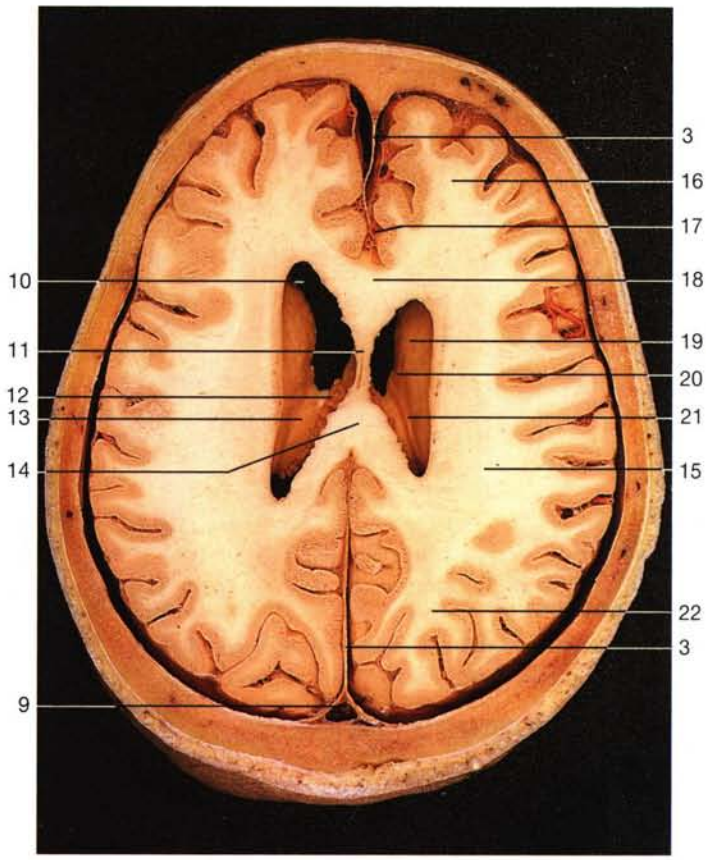

Horizontal section through the head.
Section 2.

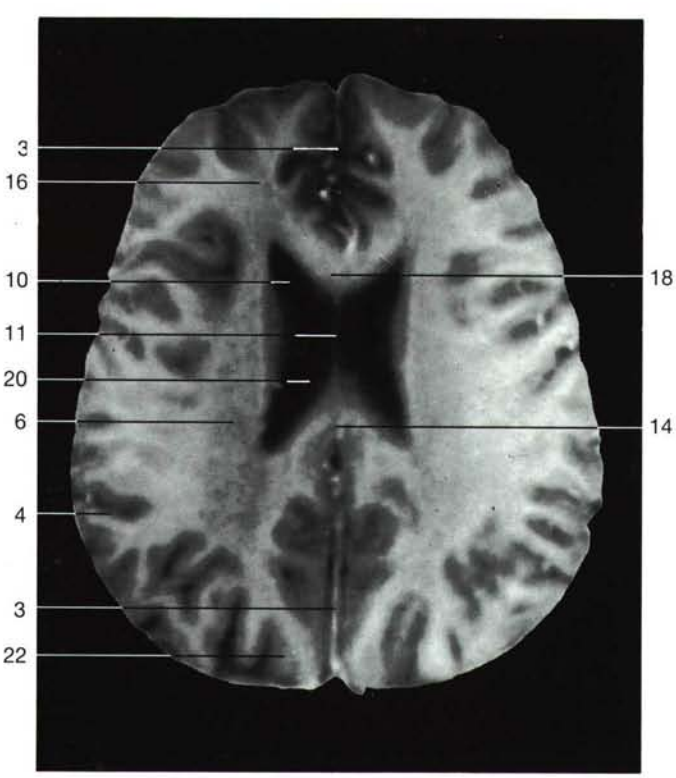

MR-Scan of the human head at the level of section 2.

1 Skin of scalp
2 Calvaria (diploe of the skull)
3 **Falx cerebri**
4 Gray matter of brain (cortex)
5 Dura mater
6 **White matter of brain**
7 Arachnoid and pia mater with vessels
8 Subdural space (slightly expanded due to shrinkage of the brain)
9 Superior sagittal sinus
10 Anterior horn of lateral ventricle
11 **Septum pellucidum**
12 Choroid plexus
13 Thalamus
14 Splenium of corpus callosum
15 Parietal lobe
16 Frontal lobe
17 Anterior cerebral artery
18 Genu of corpus callosum
19 Caudate nucleus
20 **Central part of lateral ventricle**
21 Stria terminalis
22 Occipital lobe

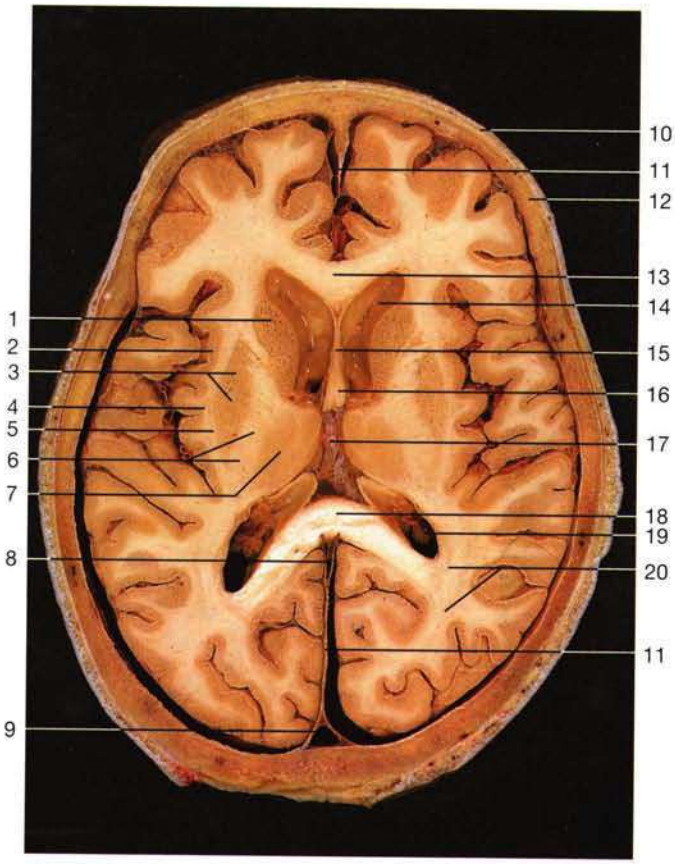

Horizontal section through the head at the level of third ventricle of internal capsule and neighbouring nuclei. Section 3.

1 **Caudate nucleus**
2 Lobus insularis (insula)
3 **Lentiform nucleus**
4 Claustrum
5 External capsule
6 Internal capsule
7 **Thalamus**
8 Straight sinus (sinus rectus)
9 Superior sagittal sinus
10 Skin of scalp
11 Falx cerebri
12 Calvaria (diploe of skull)
13 Genu of corpus callosum
14 Anterior horn of lateral ventricle
15 **Septum pellucidum**
16 Column of fornix
17 Choroid plexus of third ventricle
18 Splenium of corpus callosum
19 Entrance to inferior horn of lateral ventricle with choroid plexus
20 Optic radiation
21 **Third ventricle**

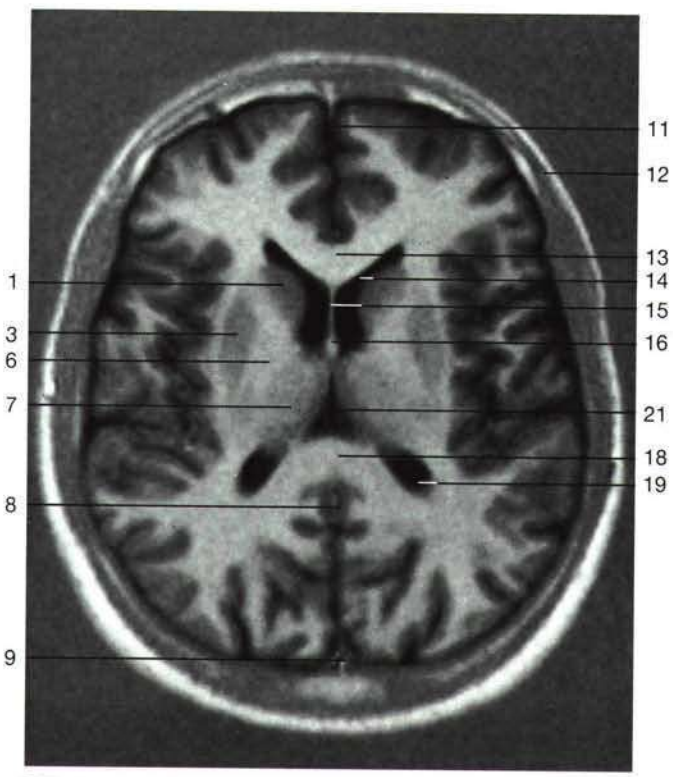

MR-Scan at the corresponding level to the above figure. Section 3.

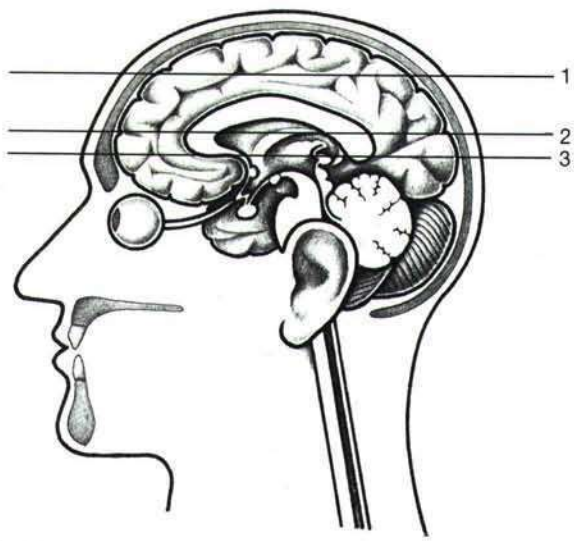

Sagittal section through the head. Levels of the horizontal sections are indicated.

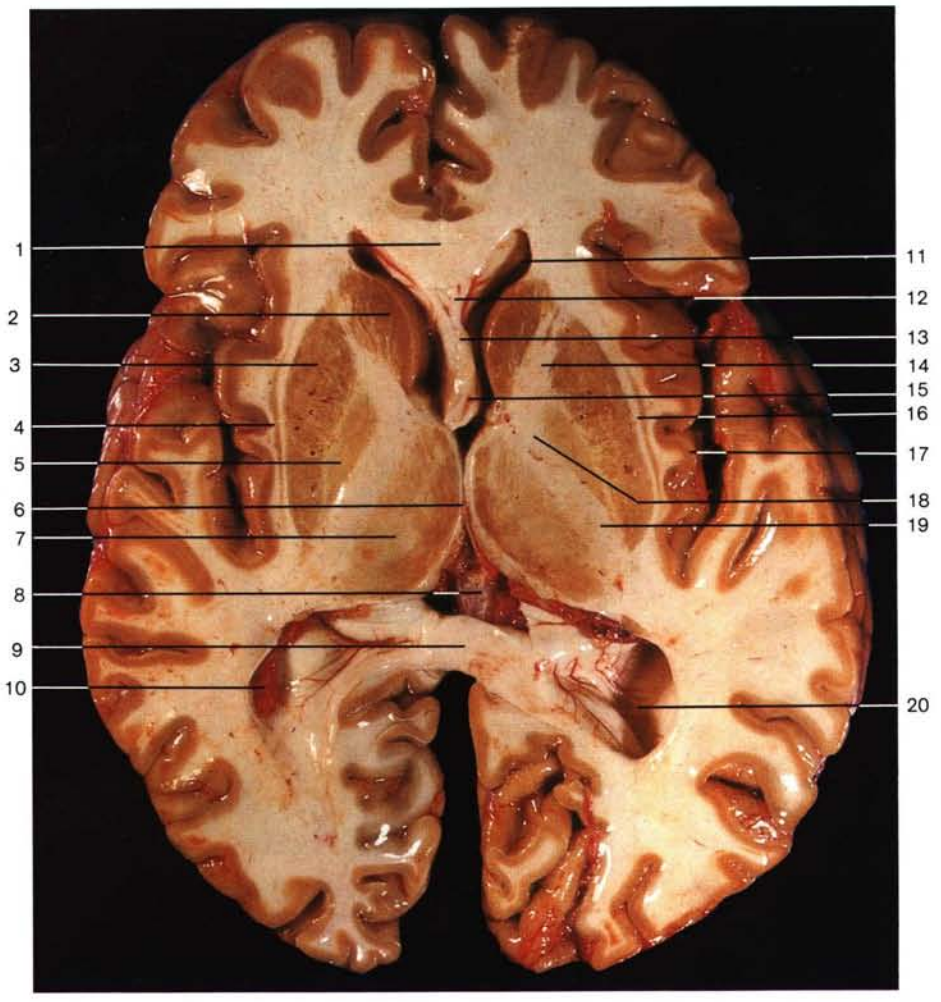

Horizontal section through the brain, showing the subcortical nuclei and internal capsule. Section 1.

1 Genu of corpus callosum
2 **Head of caudate nucleus**
3 **Putamen**
4 **Claustrum**
5 **Globus pallidus**
6 Third ventricle
7 Thalamus
8 **Pineal body**
9 Splenium of corpus callosum
10 Choroid plexus of the lateral ventricle
11 Anterior horn of lateral ventricle
12 Cavity of septum pellucidum
13 **Septum pellucidum**
14 Anterior limb of internal capsule
15 Column of fornix
16 External capsule
17 **Lobus insularis** (insula)
18 Genu of **internal capsule**
19 Posterior limb of internal capsule
20 Posterior horn of lateral ventricle
21 Anterior commissure
22 Optic radiation
23 Falx cerebri
24 **Maxillary sinus**
25 Position of auditory tube
26 Tympanic cavity
27 External acoustic meatus
28 **Medulla oblongata**
29 Fourth ventricle
30 Cerebellum (left hemisphere)
31 Temporomandibular joint
32 Tympanic membrane
33 Base of cochlea
34 Mastoid air cells
35 Sigmoid sinus
36 Vermis of cerebellum
37 Intermediate mass

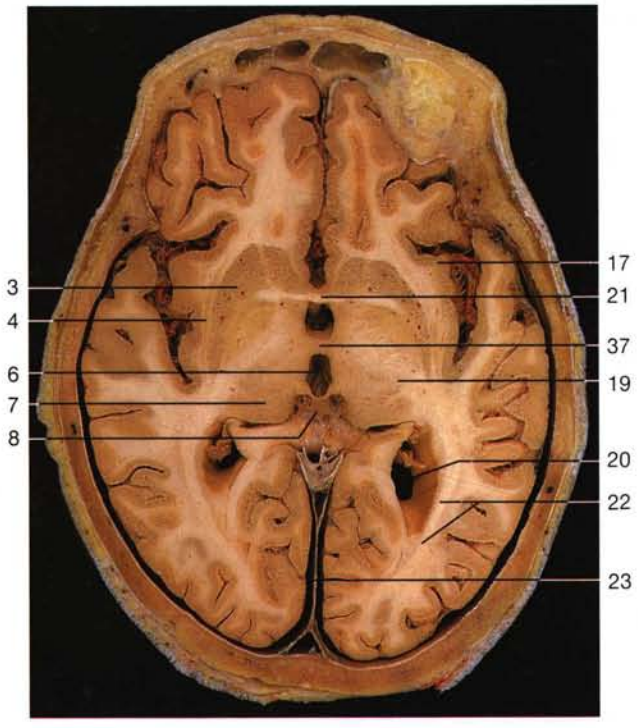

Horizontal section through the head. Section 2.

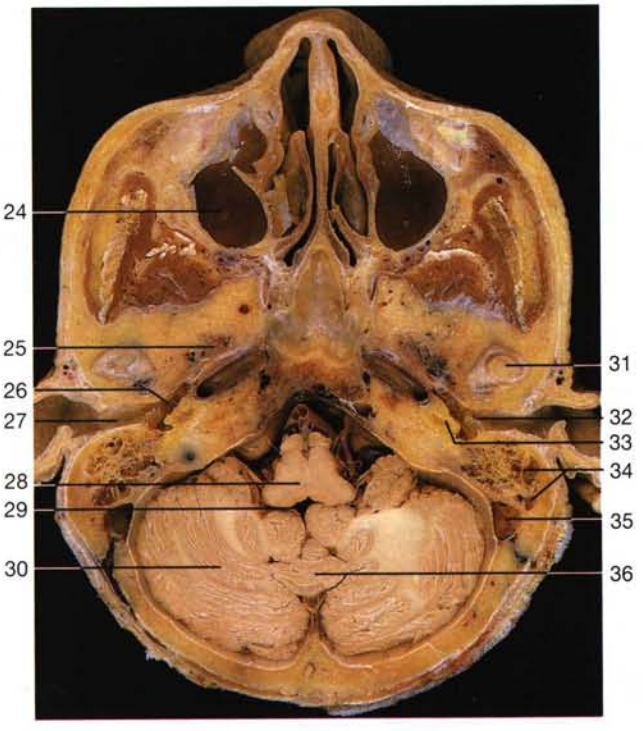

Horizontal section through the head. Section 4.

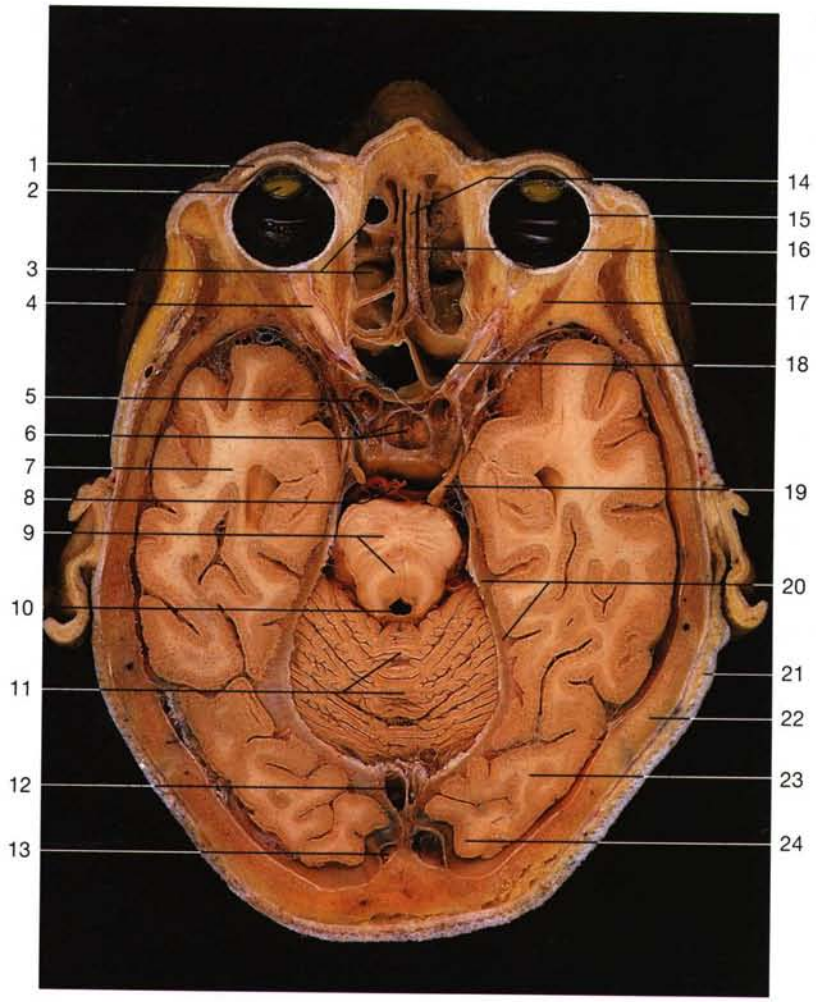

Horizontal section through the head. Section 3.

1 Upper lid (tarsal plate)
2 Lens
3 Ethmoidal sinus
4 **Optic nerve** (n. II)
5 Internal carotid artery
6 Infundibulum and **pituitary gland**
7 Temporal lobe
8 Basilar artery
9 Pons (cross section of brain stem)
10 **Cerebral aqueduct** (beginning of fourth ventricle)
11 Vermis of cerebellum
12 Straight sinus
13 Transverse sinus
14 Nasal septum
15 **Eyeball** (sclera)
16 Nasal cavity
17 Lateral rectus muscle
18 Sphenoidal sinus
19 Oculomotor nerve (n. III)
20 **Tentorium of cerebellum**
21 Skin of scalp
22 Calvaria
23 Occipital lobe
24 Striate cortex (visual cortex)

Horizontal section through the head. (CT-Scan.) Section 3.

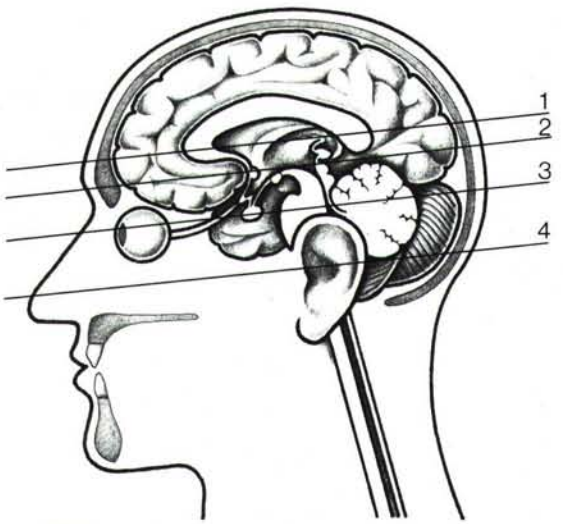

Sagittal section through the head.
Levels of the horizontal sections are indicated.

Longitudinal section through the right temporal bone I. The outer and middle ear and auditory ossicles and tube are shown (anterior aspect).

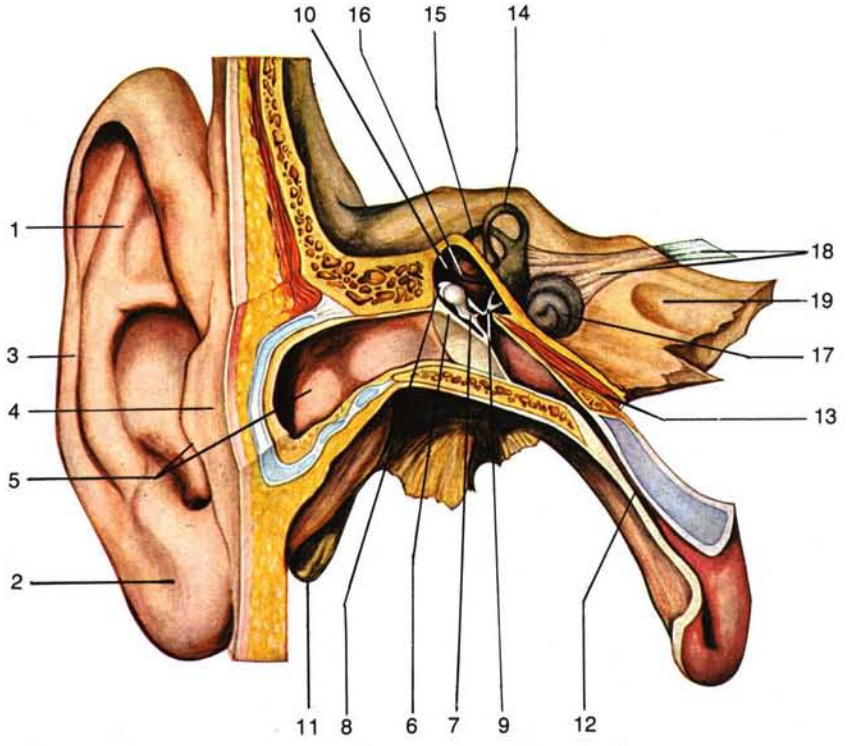

Right **auditory and vestibular apparatus** (anterior aspect). (Schematic drawing.)

Outer ear

1 Auricle
2 Lobule of auricle
3 Helix
4 Tragus
5 External acoustic meatus

Middle ear

6 Tympanic membrane
7 Malleus
8 Incus
9 Stapes
10 Tympanic cavity
11 Mastoid process
12 Auditory tube
13 Tensor tympani muscle

Inner ear

14 Anterior semicircular duct
15 Posterior semicircular duct
16 Lateral semicircular duct
17 Cochlea
18 Vestibulocochlear nerve
19 Petrous part of the temporal bone

Additional structures

20 Superior ligament of malleus
21 Arcuate eminence
22 Internal carotid artery
23 Anterior surface of pyramid with dura mater
24 Stapes
25 Levator veli palatini muscle

Longitudinal section through the right outer, middle and inner ear II. The cochlea and semicircular canals have been further dissected (anterior aspect).

Internal acoustic meatus, left side. The bone was partly removed to show the bottom of the meatus.

1 Roof of tympanic cavity
2 **Lateral osseous semicircular canal**
3 **Facial nerve**
4 Incus
5 **Malleus**
6 External acoustic meatus
7 **Tympanic cavity** and **tympanic membrane**
8 Vestibulochochlear nerve
9 Anterior osseous semicircular canal
10 Geniculate ganglion and greater petrosal nerve
11 **Cochlea**
12 **Stapes**
13 Tensor tympani muscle
14 **Auditory tube**
15 Levator veli palatini muscle
16 Area of facial nerve
17 Superior vestibular area
18 Transverse crest
19 Foramen singulare
20 Foraminous spiral tract (outlet of cochlear part of vestibulocochlear nerve)
21 Base of cochlea

Longitudinal section through the outer, middle and inner ear III. Deeper dissection to display facial nerve and lesser and greater petrosal nerves (anterior aspect).

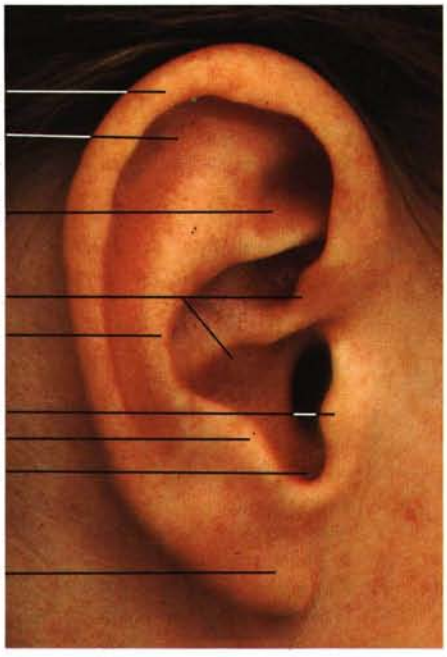

Right auricle (lateral aspect).

1 Helix
2 Scaphoid fossa
3 Triangular fossa
4 Concha
5 Antihelix
6 Tragus
7 Antitragus
8 Intertragic notch
9 Lobule

1 **Anterior osseous semicircular canal** (opened)
2 **Posterior osseous semicircular canal**
3 **Lateral osseous semicircular canal** (opened)
4 Facial nerve and chorda tympani
5 External acoustic meatus
6 Auricle
7 **Facial nerve**
8 Trigeminal nerve
9 Bony base of internal acoustic meatus
10 Internal carotid artery within cavernous sinus
11 **Cochlea**
12 **Facial nerve** with **geniculate ganglion**
13 Greater petrosal nerve
14 Lesser petrosal nerve
15 **Tympanic cavity**
16 Auditory tube
17 Levator veli palatini muscle
18 Internal carotid artery, internal jugular vein
19 Styloid process

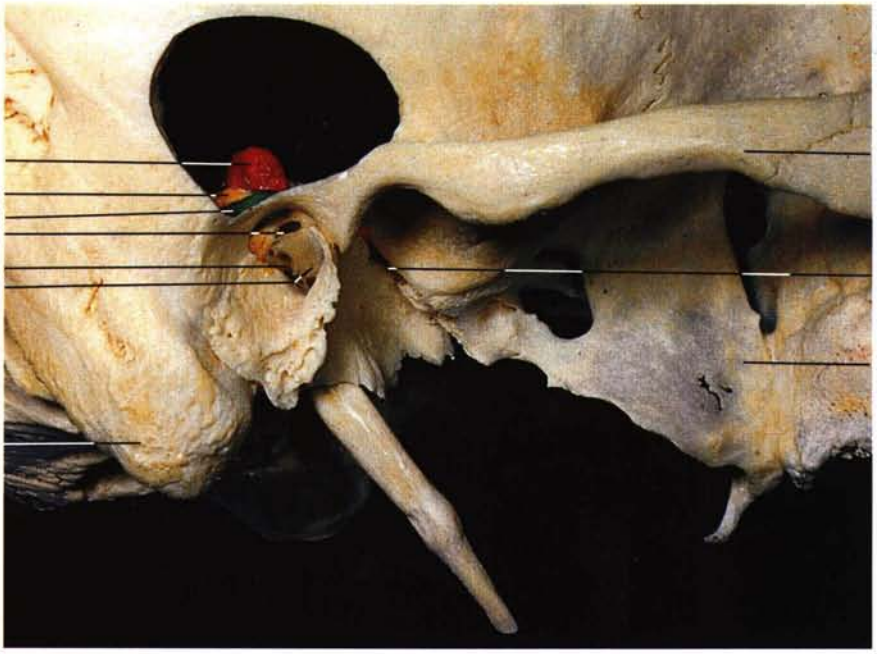

Right temporal bone (lateral aspect). Petrosquamous portion has been partly removed to display the semicircular canals.

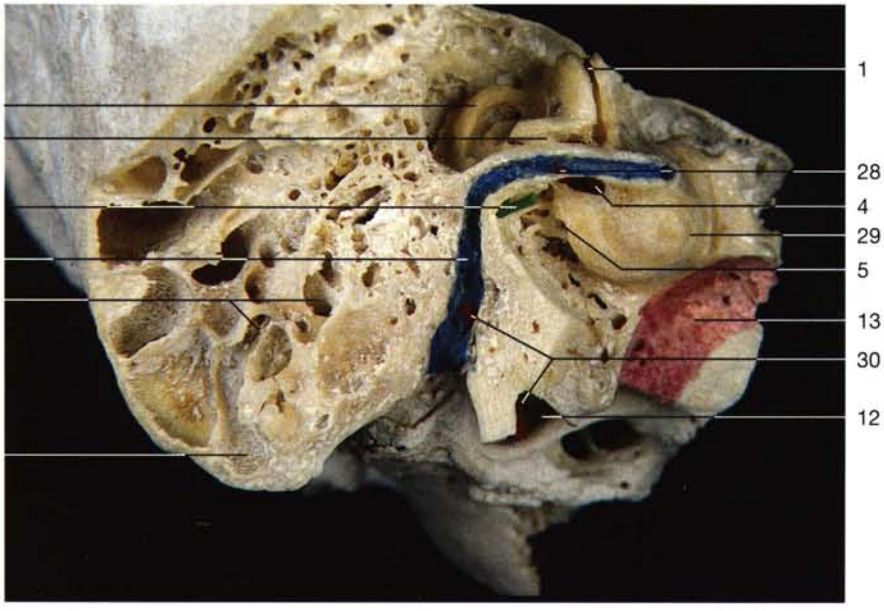

Right temporal bone (lateral aspect). Mastoid air cells and facial canal had been opened. The 3 semicircular canals were dissected.

1 Anterior semicircular canal (red)
2 Posterior semicircular canal (yellow)
3 Lateral or horizontal semicircular canal (green)
4 **Fenestra vestibuli**
5 **Fenestra cochleae**
6 Tympanic cavity
7 Mastoid process
8 **Petrotympanic fissure** (red probe: **chorda tympani**)
9 Lateral pterygoid plate
10 Mastoid air cells
11 **Facial canal** (blue)
12 Foramen ovale
13 **Carotid canal** (red)
14 **Tympanic ring**
15 Petromastoid part of temporal bone
16 Squamous part of temporal bone
17 Squamomastoid suture
18 Zygomatic process of temporal bone
19 Incisure of tympanic ring
20 **Promontory**
21 Apex of cochlea (cupula)
22 Spiral canal of cochlea at base of cochlea
23 Epitympanic recess
24 Auditory ossicles and tympanic cavity
25 Hypotympanic recess
26 Canaliculus chordae tympani (green probe)
27 Mastoid process
28 Canaliculus for stapedius nerve (red)
29 **Cochlea**
30 Canaliculus mastoideus (red probe)

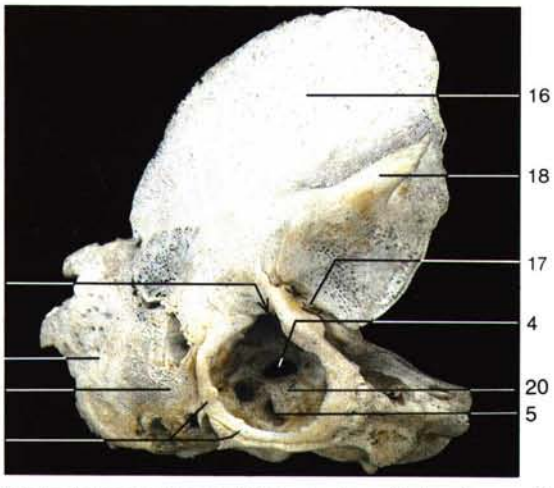

Right temporal bone of the newborn (lateral aspect).

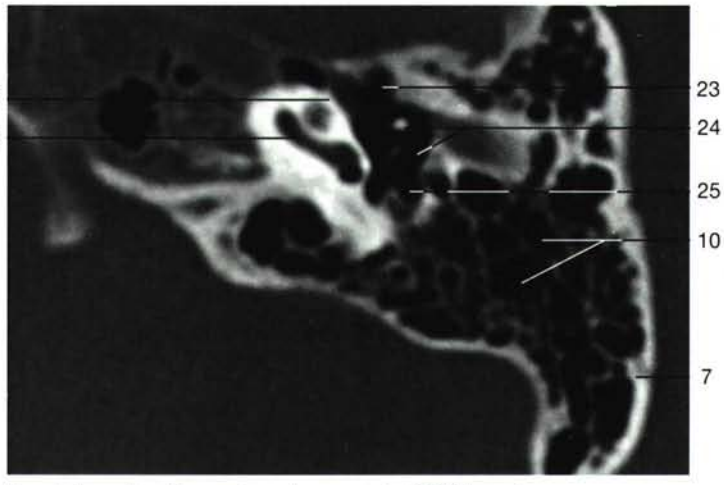

Frontal section through petrous part. (CT-Scan.)

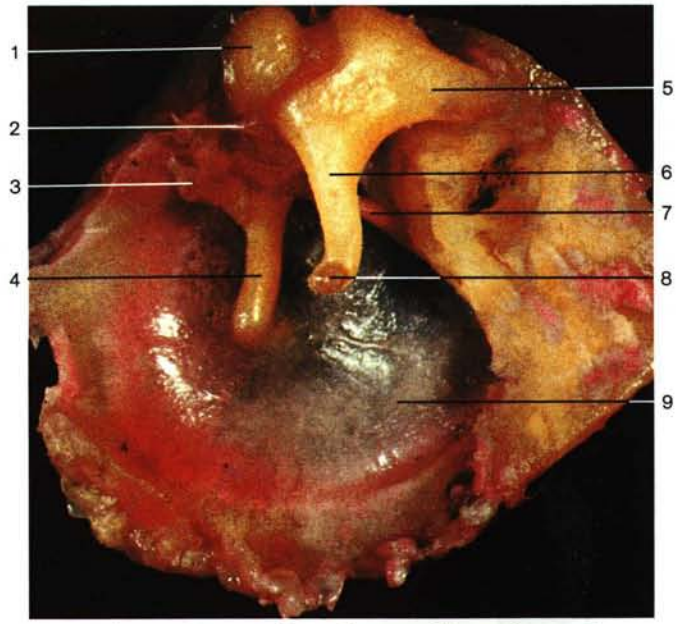

1 Head of malleus
2 Anterior ligament of malleus
3 Tendon of tensor tympani muscle
4 Handle of **malleus**
5 Short crus of incus
6 Long crus of **incus**
7 Chorda tympani
8 Lenticular process
9 **Tympanic membrane**

Tympanic membrane with malleus and incus (internal aspect; right side).

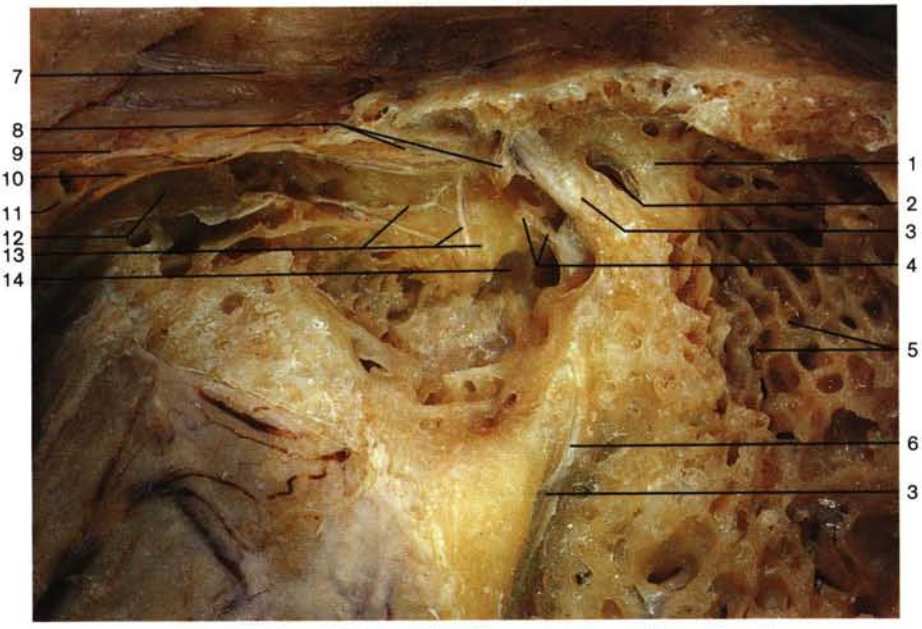

1 Tympanic antrum
2 Lateral semicircular canal (opened)
3 **Facial canal**
4 Stapes with tendon of stapedius
5 Mastoid air cells
6 Chorda tympani (intracranial part)
7 Greater petrosal nerve
8 Tensor tympani muscle (processus cochleariformis)
9 Lesser petrosal nerve
10 Anterior tympanic artery
11 Middle meningeal artery
12 **Auditory tube**
13 **Promontory with tympanic plexus**
14 Fenestra cochleae

Tympanic cavity, medial wall. External auditory meatus and lateral wall of tympanic cavity together with incus. Malleus and tympanic membrane have been removed; mastoid air cells are opened (left side).

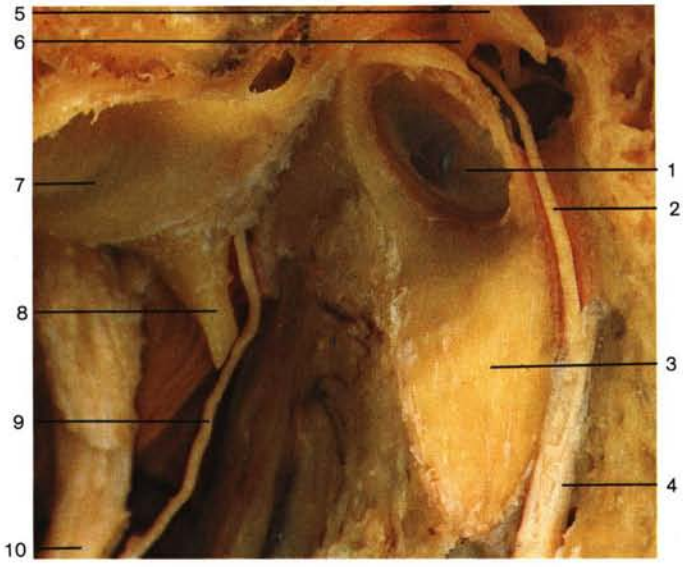

1 **Tympanic membrane**
2 Chorda tympani (intracranial part)
3 Floor of the external acoustic meatus
4 **Facial nerve** and facial canal
5 Incus
6 Head of malleus
7 Mandibular fossa
8 Spine of sphenoid
9 **Chorda tympani** (extracranial part)
10 Styloid process

Tympanic membrane (lateral aspect). External acoustic meatus and facial canal have been opened to expose the chorda tympani (magn. ~1.5×) (left side).

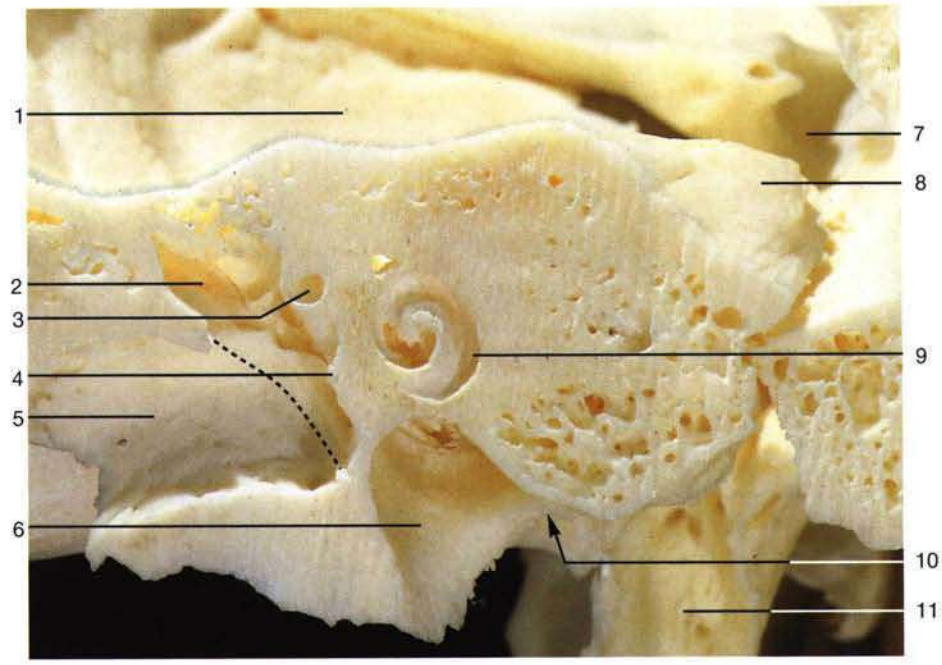

Frontal section through the petrous part of the left temporal bone at the level of the cochlea (posterior aspect). Position of tympanic membrane indicated by dotted line.

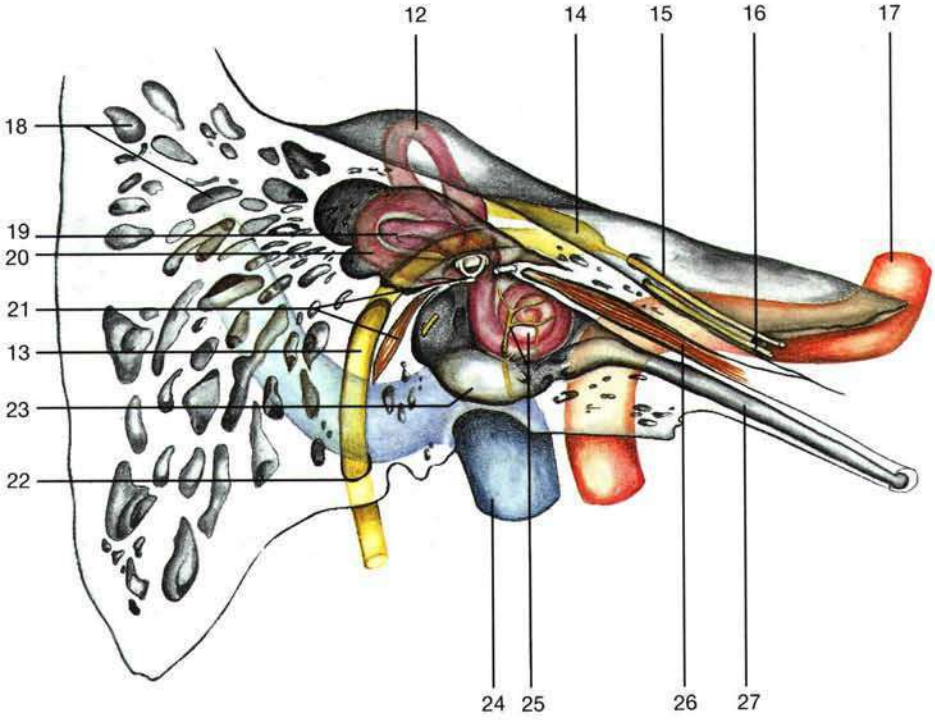

Medial wall of tympanic cavity and its relation to neighboring structures of the inner ear, facial nerve, and blood vessels. (Schematic drawing.) Frontal section through the right temporal bone (anterior aspect).

1 Anterior surface of the pyramid
2 Mastoid antrum
3 Lateral semicircular canal
4 Cochleariform process
5 External acoustic meatus
6 **Jugular fossa**
7 Foramen lacerum
8 Apex of petrous part
9 Position of **cochlea** (modiolus with crista spiralis ossea)
10 Carotid canal
11 Pterygoid process
12 Anterior semicircular duct
13 **Facial nerve**
14 Geniculate ganglion
15 Greater petrosal nerve
16 Lesser petrosal nerve
17 **Internal carotid artery**
18 Mastoid air cells
19 Lateral semicircular duct
20 Posterior semicircular duct
21 Stapes with stapedius muscle
22 Stylomastoid foramen
23 Inferior recess of tympanic cavity (hypotympanon)
24 **Internal jugular vein**
25 Promontory with tympanic plexus (position of cochlea)
26 Tensor muscle of tympanum
27 **Auditory tube**

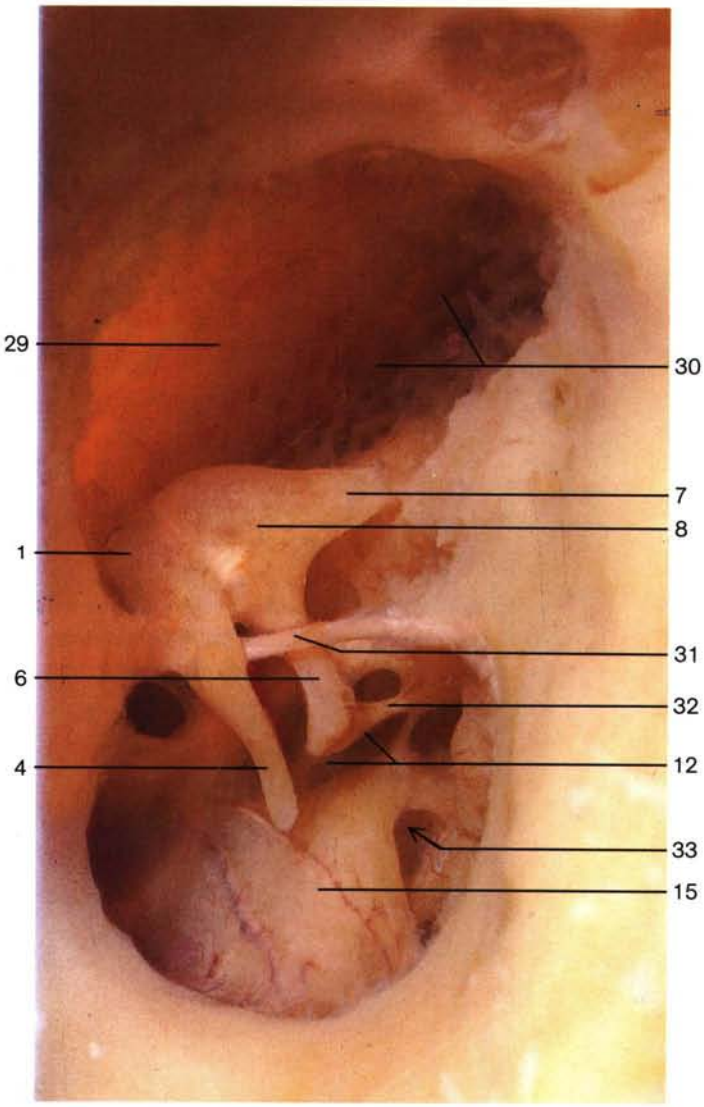

Tympanic cavity with malleus, incus and stapes, left side (lateral aspect). Tympanic membrane removed, mastoid antrum opened.

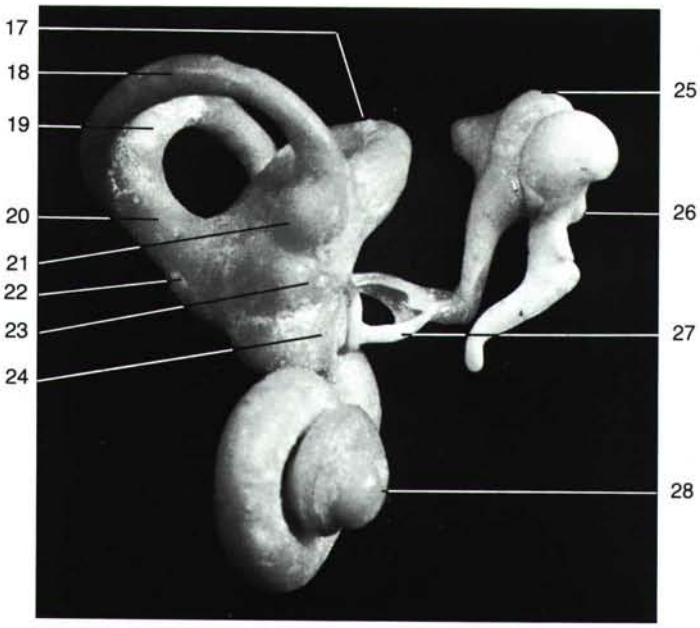

Chain of auditory ossicles in connection with the inner ear, left side (anterior-lateral aspect).

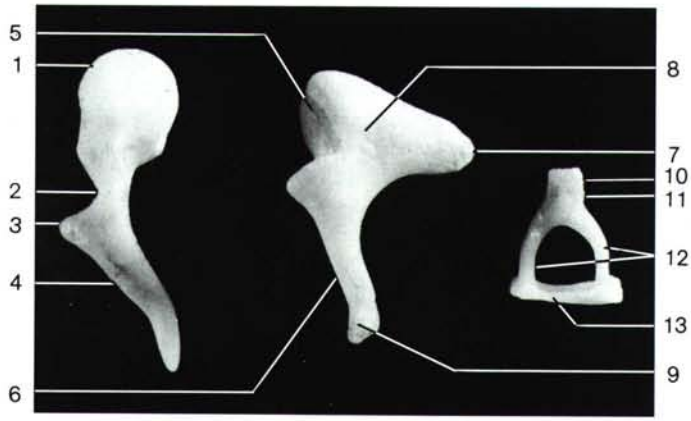

Auditory ossicles (isolated).

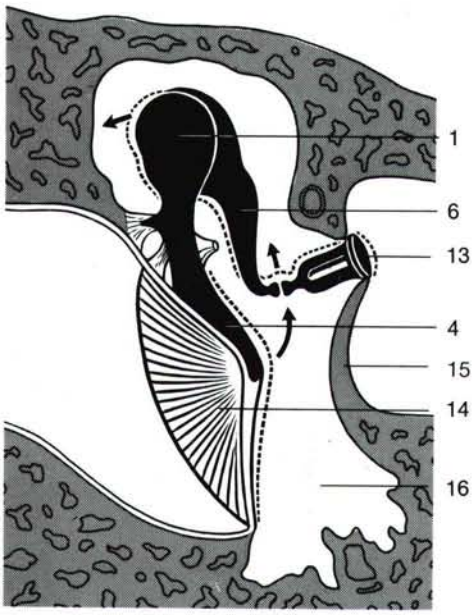

Position and movements of the auditory ossicles. (Schematic diagram.)

Malleus

1 Head
2 Neck
3 Lateral process
4 Handle

Incus

5 Articular facet for malleus
6 Long crus
7 Short crus
8 Body
9 Lenticular process

Stapes

10 Head
11 Neck
12 Anterior and posterior crura
13 Base

Walls of tympanic cavity

14 Tympanic membrane
15 Promontory
16 Hypotympanic recess of tympanic cavity

Internal ear (labyrinth)

17 Lateral semicircular duct
18 Anterior semicircular duct
19 Posterior semicircular duct
20 Common crus
21 Ampulla
22 Beginning of endolymphatic duct
23 Utricular prominence
24 Saccular prominence
25 Incus
26 Malleus
27 Stapes
28 Cochlea

Tympanic cavity

29 Epitympanic recess
30 Mastoid antrum
31 Chorda tympani
32 Tendon of stapedius muscle
33 Round window (fenestra cochleae)

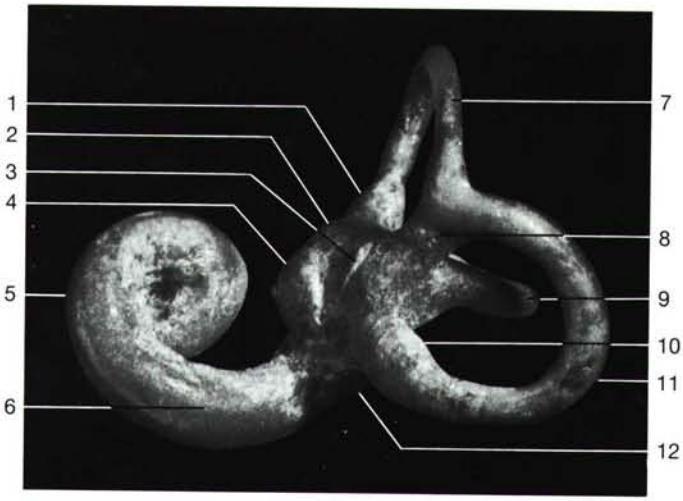

Cast of the right labyrinth (postero-medial aspect).

1 Ampulla (anterior semicircular canal)
2 Elliptical recess
3 Aqueduct of the vestibule
4 Spherical recess
5 **Cochlea**
6 Base of cochlea
7 **Anterior semicircular canal**
8 Crus commune or common limb
9 **Lateral semicircular canal**
10 Posterior bony ampulla
11 **Posterior semicircular canal** (posterior canal)
12 **Fenestra cochleae**
13 Bony ampulla
14 **Fenestra vestibuli**
15 Cupula of cochlea
16 External acoustic meatus
17 Mastoid air cells
18 Tympanic cavity and fenestra cochleae (probe)
19 External acoustic meatus
20 Facial canal
21 Base of cochlea and canalis musculotubarius
22 Malleus and incus
23 Stapes
24 Tympanic membrane
25 Tympanic cavity
26 Aqueduct of cochlea
27 Saccus endolymphaticus
28 Ductus endolymphaticus
29 Macula of utricle
30 Macula of saccule

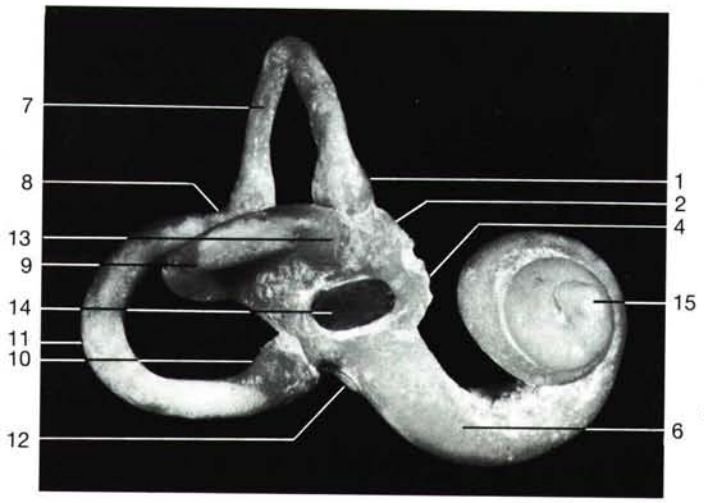

Cast of the right labyrinth (lateral aspect).

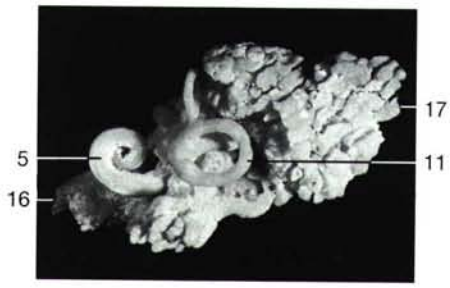

Cast of the labyrinth and mastoid cells. Life size (posterior aspect).

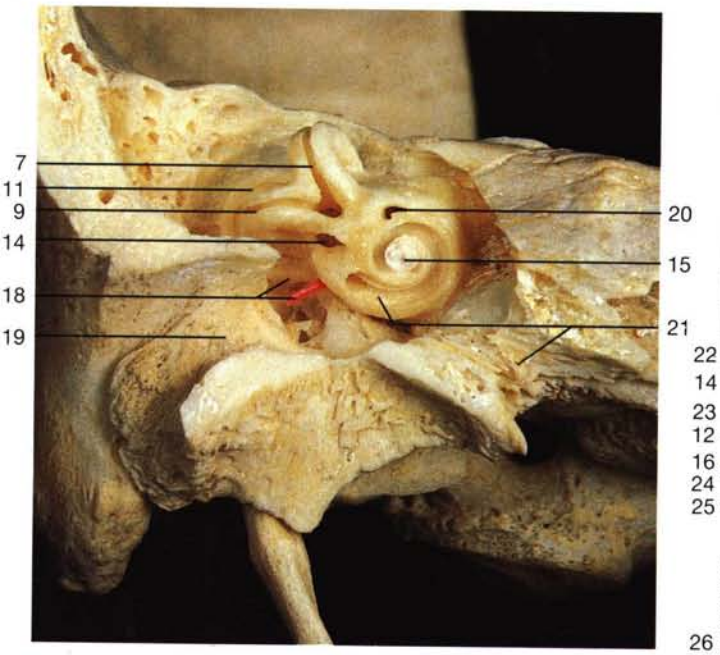

Dissection of bony labyrinth in situ. Semicircular canals and cochlear duct opened.

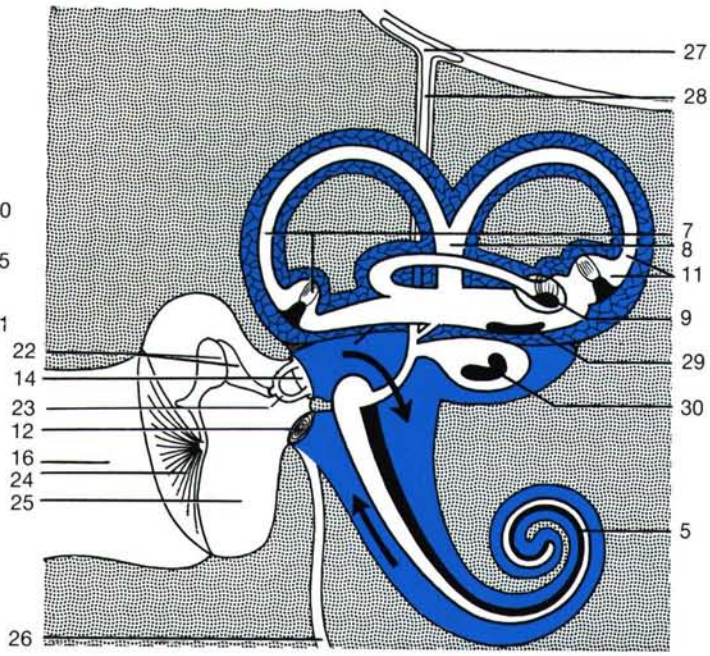

Auditory and vestibular apparatus. Arrows: direction of sound waves; blue = perilymphatic ducts. (Schematic diagram.)

Bony labyrinth, petrous part of the temporal bone (from above). At left: semicircular canals opened, at right closed. Arrows: internal acoustic meatus.

1 Facial canal and semicanal of auditory tube
2 Superior vestibular area
3 Foramen ovale
4 Foramen lacerum
5 Cochlea
6 Vestibule
7 Anterior semicircular canal
8 Lateral semicircular canal
9 Posterior semicircular canal
10 Groove for sigmoid sinus
11 Sigmoid sinus
12 Tympanic cavity
13 Auditory tube
14 Mastoid air cells
15 Facial and vestibulocochlear nerves
16 Temporal fossa
17 Fenestra vestibuli
18 Promontory
19 Zygomatic process
20 Fenestra cochleae
21 Mastoid process

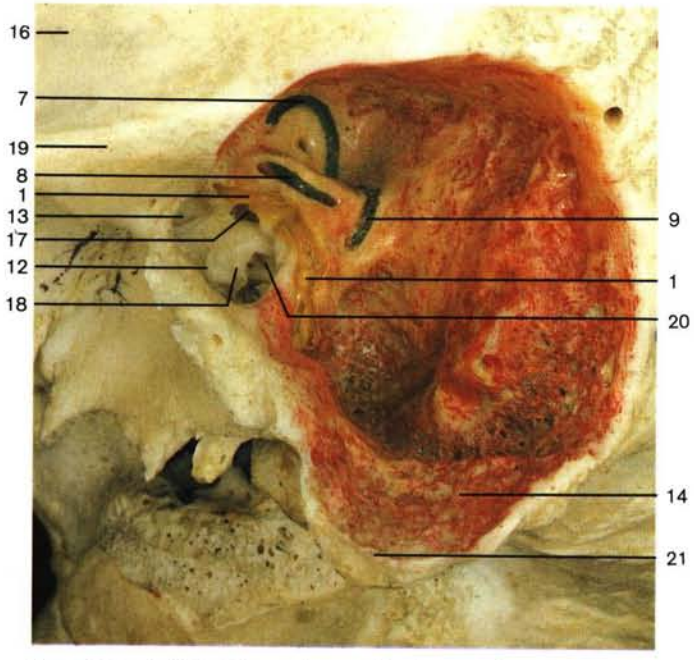

Bony labyrinth (left lateral aspect). Temporal and tympanic bone partly removed, semicircular canals opened.

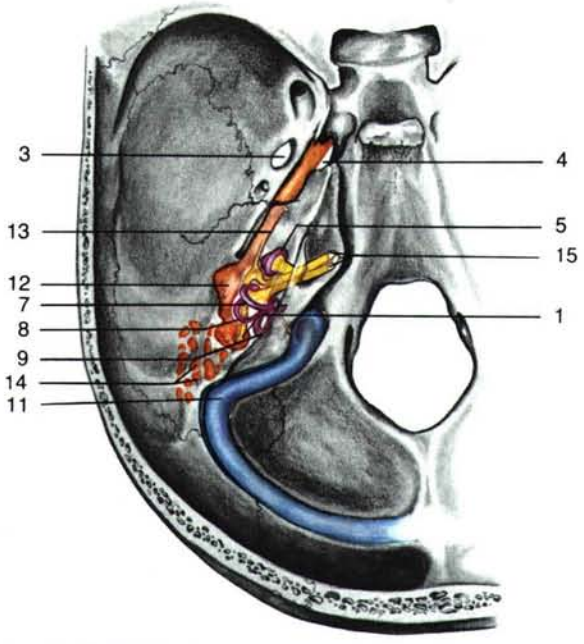

Internal ear. Diagram showing the position of the membranous labyrinth and the tympanic cavity.

Dissection of the brain stem showing the acoustic pathway. Cerebellum and posterior part of the two hemispheres have been removed (dorsal aspect).

1 Left lateral ventricle and corpus callosum
2 Thalamus
3 Pineal gland (epiphysis)
4 Superior colliculus
5 Anterior cerebellar velum and superior cerebellar peduncle
6 Rhomboid fossa
7 **Vestibulocochlear nerve** (n. VIII)
8 **Dorsal acoustic striae** and inferior cerebellar peduncle
9 Insular lobe
10 Caudate nucleus and thalamus
11 Temporal lobe (superior temporal gyrus) (area of **acoustic centers**)
12 Transverse temporal gyri of Heschl (area of **primary acoustic centers**)
13 Acoustic radiation of internal capsule
14 Lateral geniculate body and optic radiation (cut)
15 **Medial geniculate body** and brachium of inferior colliculus
16 **Inferior colliculus**
17 Cerebral peduncle
18 **Lateral lemniscus**
19 Middle cerebellar peduncle
20 Dorsal (posterior) cochlear nucleus
21 Ventral (anterior) cochlear nucleus
22 Inferior olive with tractus olivo-cochlearis of Rasmussen (red)
23 Ganglion spirale
24 Obex
25 Frontal lobe
26 Temporal lobe
27 Middle temporal gyrus (area of **tertiary acoustic centers)**
28 Trapezoid body

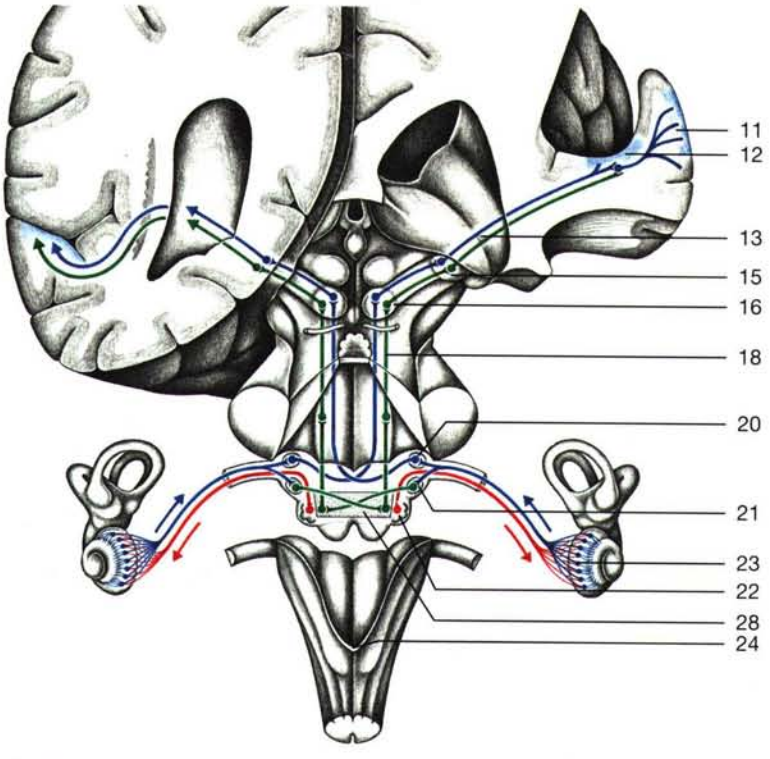

Acoustic pathway (schematic drawing, compare with figure above).

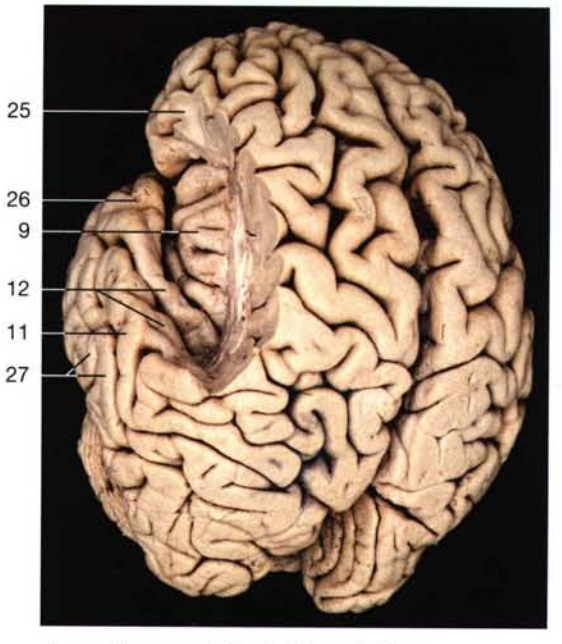

Acoustic areas in the left hemisphere (superior lateral aspect). Parts of the frontal and parietal lobe have been removed.

Bones of the left orbit (indicated by different colors).

Frontal section through the posterior part of the orbit.

Sagittal section through orbit and eyeball.

1 Frontal bone
2 Nasal bone
3 Lacrimal bone
4 Maxilla (frontal process)
5 Ethmoidal foramina
6 Lesser wing of sphenoid bone and optic canal
7 Superior orbital fissure
8 Greater wing of sphenoid bone
9 Orbital process of palatine bone
10 Orbital plate of ethmoid bone
11 Inferior orbital fissure
12 Infraorbital sulcus
13 Nasolacrimal canal
14 Zygomatic bone
15 Frontal sinus
16 **Superior rectus muscle**
17 Orbital fatty tissue
18 **Optic nerve**
19 **Sclera**
20 **Inferior rectus muscle**
21 Periorbita and maxilla
22 Maxillary sinus
23 Levator palpebrae superioris muscle
24 **Superior conjunctival fornix**
25 Superior tarsal plate
26 Inferior tarsal plate
27 **Inferior conjunctival fornix**
28 **Inferior oblique muscle**
29 **Lateral rectus muscle**
30 **Medial rectus muscle**
31 **Superior oblique muscle**
32 Nasal septum
33 Middle nasal concha
34 Inferior nasal concha
35 Tenon's space
36 Ophthalmic artery
37 **Cornea**
38 **Lens**

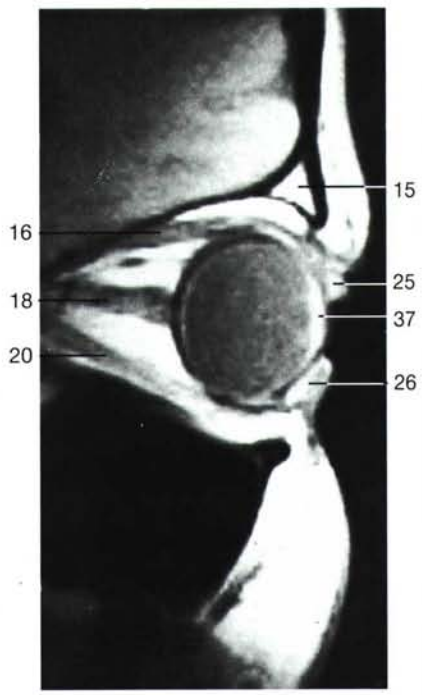

Sagittal section through orbit and eyeball. (MR-Scan.)

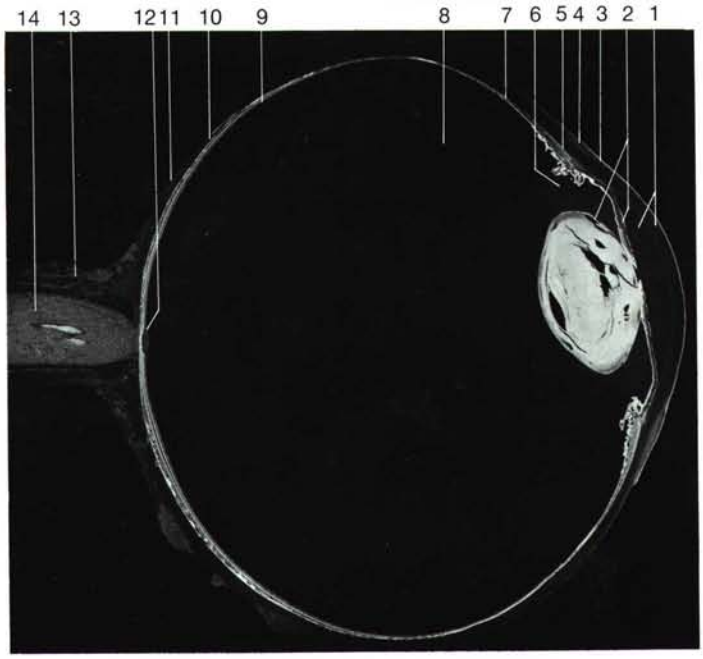

Horizontal section through the human eye (2×).

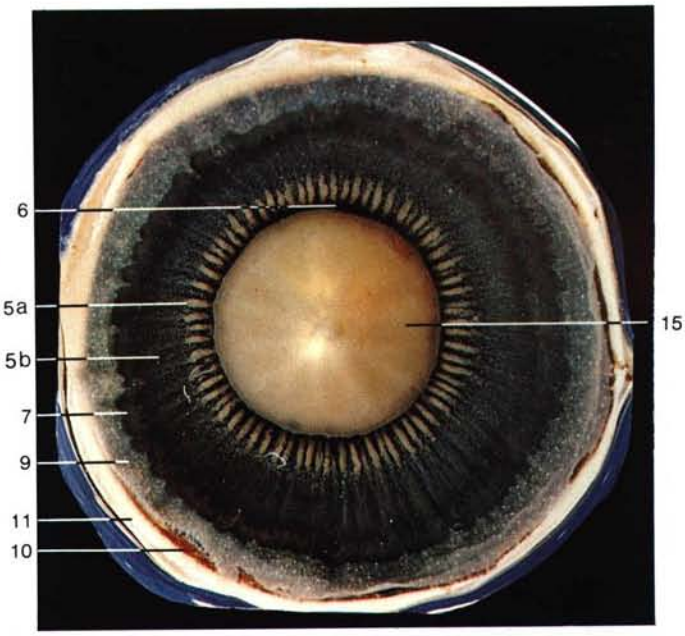

Anterior segment of the eyeball (posterior aspect). The opacity of the lens is an artifact.

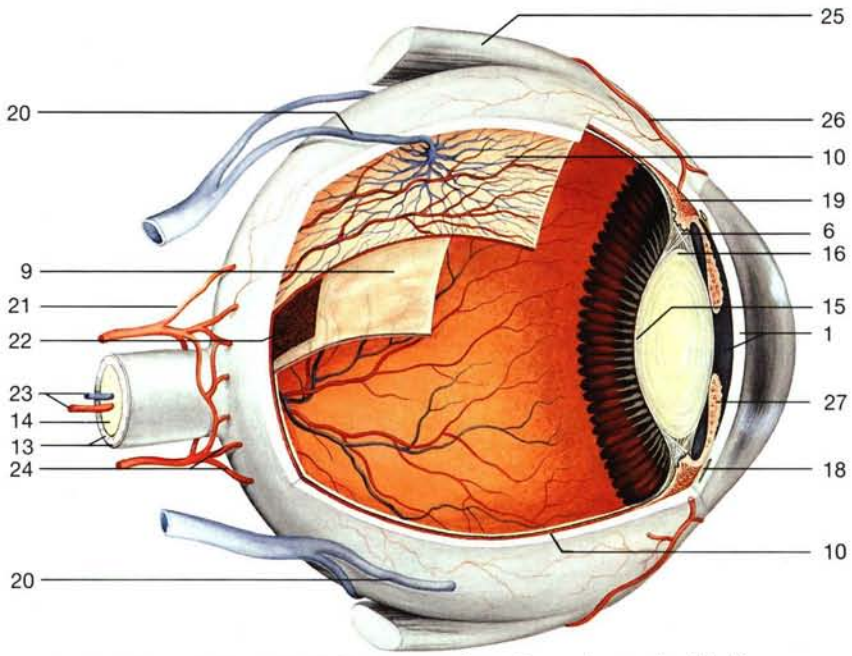

Organization of the eyeball. Demonstration of vascular tunic of bulb. (Schematic drawing.)

Lens (equatorial aspect), anterior pole to the right.

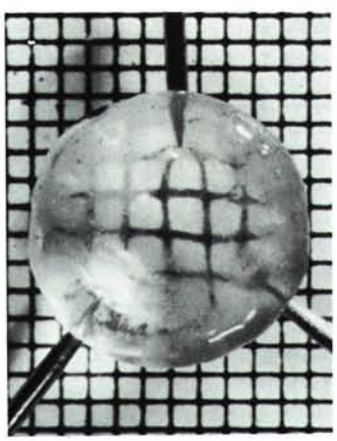

Lens (frontal aspect). Note the magnification effect.

1 **Cornea** and anterior chamber
2 Iris and lens
3 Transitional zone between corneal and conjunctival epithelium
4 Conjunctiva of the eyeball
5 **Ciliary body**
a Ciliary processes (pars plicata)
b Ciliary ring (pars plana)
6 Zonular fibers
7 **Ora serrata**
8 Vitreous body
9 **Retina**
10 **Choroid**
11 **Sclera**
12 Optic disc
13 Dura mater and subarachnoid space
14 **Optic nerve** (n. II)
15 Lens (posterior pole)
16 Equator of lens
17 **Lens** (anterior pole)
18 Canal of Schlemm
19 Ciliary muscle
20 Vena vorticosa
21 Long posterior ciliary artery
22 Retinal pigmented epithelium
23 Central retinal artery and vein
24 Short posterior ciliary arteries
25 External ocular muscle
26 Anterior ciliary artery
27 **Iris**

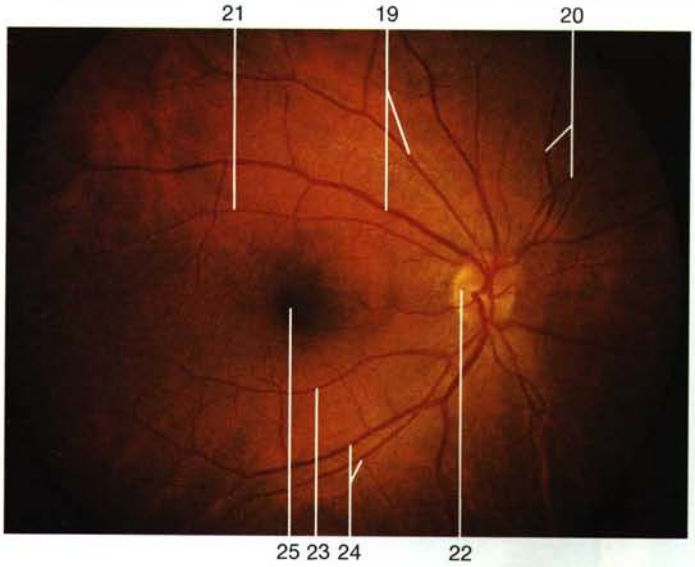

Fundus of a normal right eye (courtesy of Prof. Dr. R. Okamura, Univ. Eye Dept., Kumamoto/Japan). Notice, the arteries are smaller and lighter than the veins.

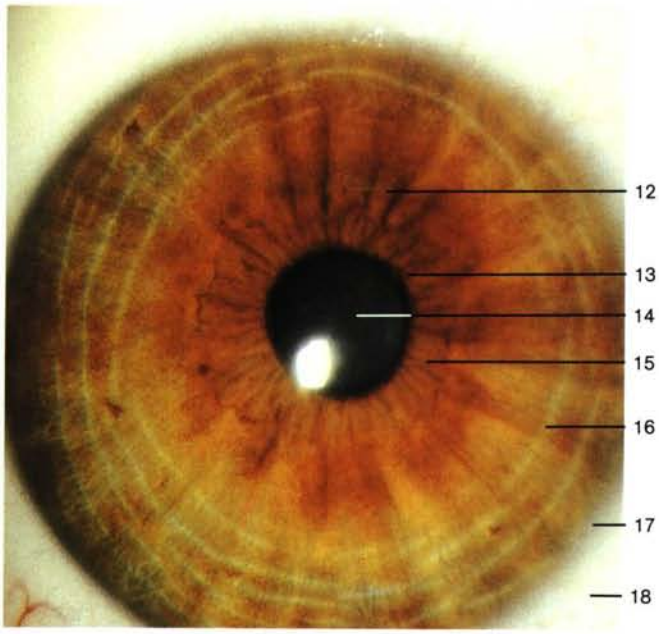

Anterior segment of the human eye (courtesy of Prof. Dr. G. O. H. Naumann, Eye Dept., University of Erlangen, FRG). Note the colored iris and the anterior lens surface.

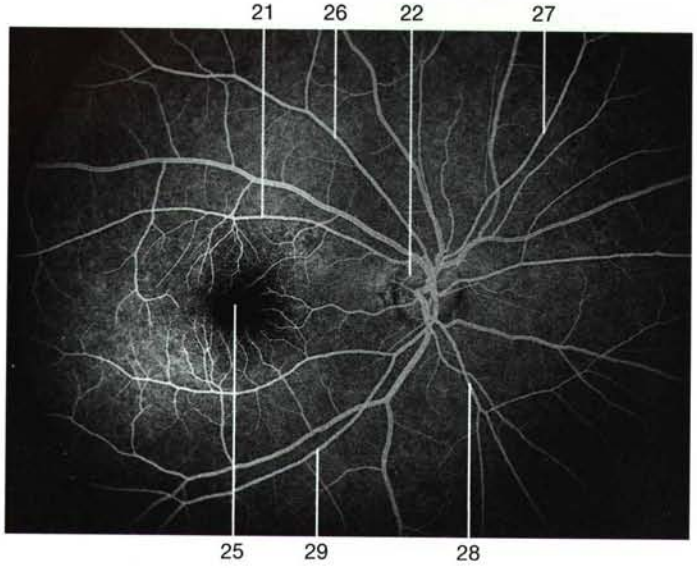

Fluorescent angiography of the right eye; retinal vessels. The same eye as above. (Courtesy of Prof. Dr. R. Okamura, Univ. Eye Dept., Kumamoto/Japan.)

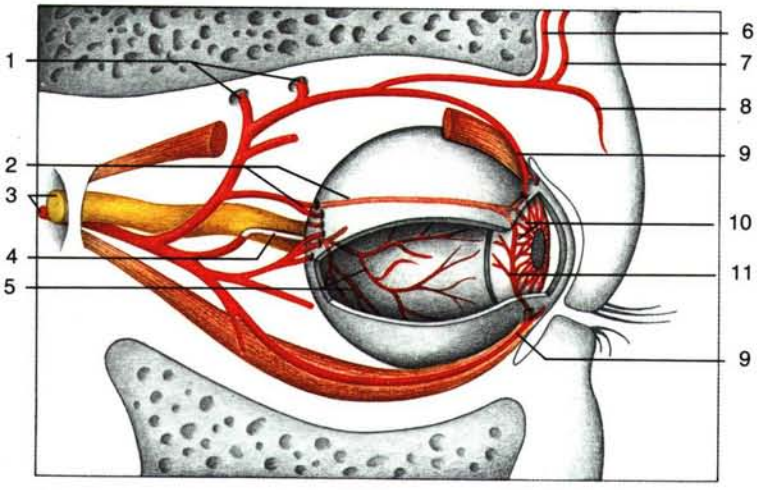

Diagram of the ophthalmic artery and its branches.

1 Posterior and anterior ethmoidal arteries
2 Long and short **posterior ciliary arteries**
3 Optic nerve and **ophthalmic artery**
4 Central retinal artery
5 Retinal arteries
6 Supratrochlear artery
7 Supraorbital artery
8 Dorsal nasal artery
9 **Anterior ciliary artery**
10 Iridial arteries
11 Circulus arteriosus major of iris
12 Iridial fold
13 Pupillary margin of iris
14 Anterior pole of lens
15 Lesser circle of **iris**
16 Greater circle of iris
17 Margin of cornea or limbus
18 Sclera
19 Superior temporal artery and vein of retina
20 Superior nasal artery and vein of retina
21 Superior macular artery
22 **Optic disc**
23 Inferior macular artery
24 Inferior temporal artery and vein
25 **Fovea centralis** and **macula lutea**
26 Superior temporal artery
27 Superior nasal artery
28 Inferior nasal artery
29 Inferior temporal artery
} of retina

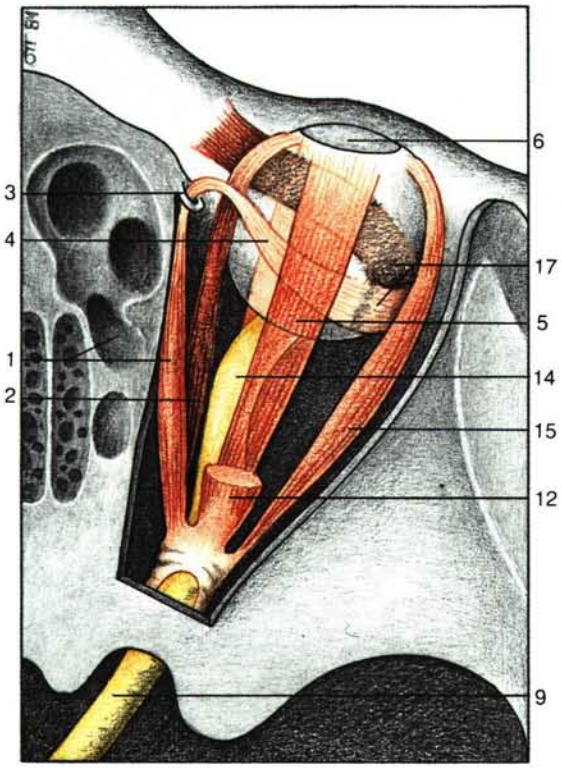

Schematic diagram of the extraocular muscles. Right orbit (from above). Levator palpebrae superioris muscle has been severed.

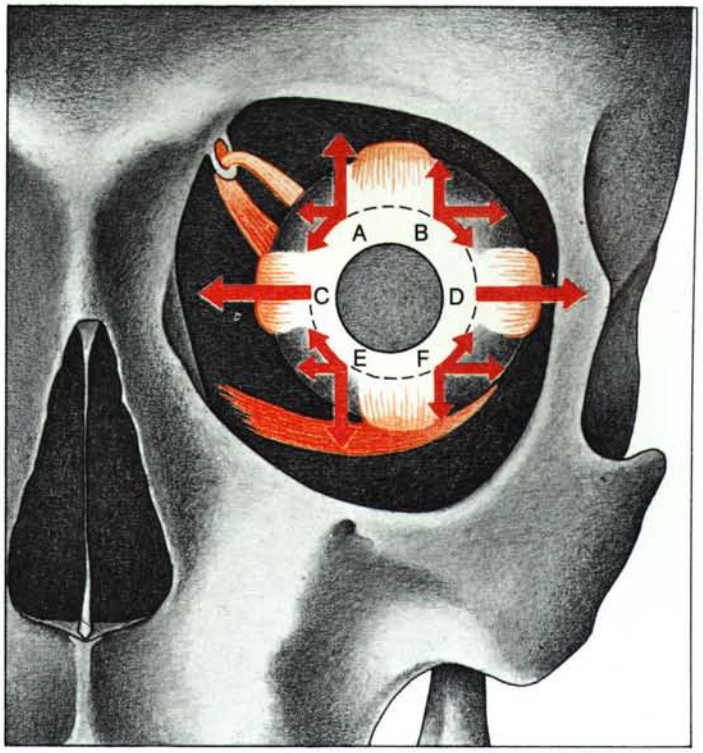

The action of the extraocular muscles. Left orbit (anterior aspect).

A Superior rectus muscle
B Inferior oblique muscle
C Medical rectus muscle
D Lateral rectus muscle
E Inferior rectus muscle
F Superior oblique muscle

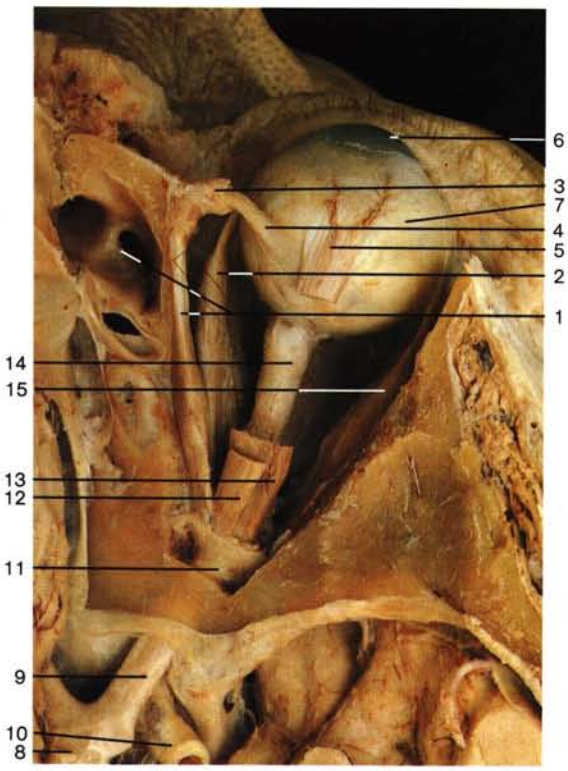

Right orbit with eyeball and extraocular muscles (from above). The roof of the orbit has been removed, the superior rectus muscle and the levator palpebrae superioris muscle have been severed.

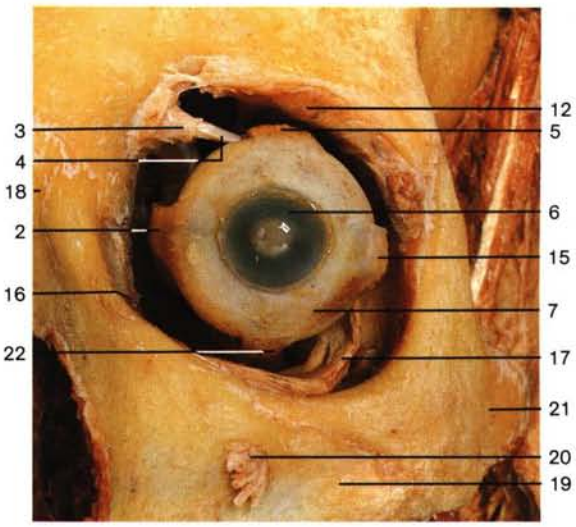

Left orbit with eyeball and extraocular muscles (anterior aspect). Lids, conjunctiva and lacrimal apparatus have been removed.

1 **Superior oblique muscle** and ethmoid air cells
2 **Medial rectus muscle**
3 Trochlea
4 Tendon of superior oblique muscle
5 **Superior rectus muscle**
6 Cornea
7 Eyeball
8 Optic chiasma
9 Optic nerve (intracranial part)
10 Internal carotid artery
11 **Common annular tendon**
12 Levator palpebrae superioris muscle
13 **Superior rectus muscle**
14 Optic nerve (extracranial part)
15 **Lateral rectus muscle**
16 Nasolacrimal duct
17 **Inferior oblique muscle**
18 Nasal bone
19 Maxilla
20 Infraorbital foramen and nerves
21 Zygomatic bone
22 **Inferior rectus muscle**

Extraocular muscles and their nerves (lateral aspect of left eye). Lateral rectus divided and reflected.

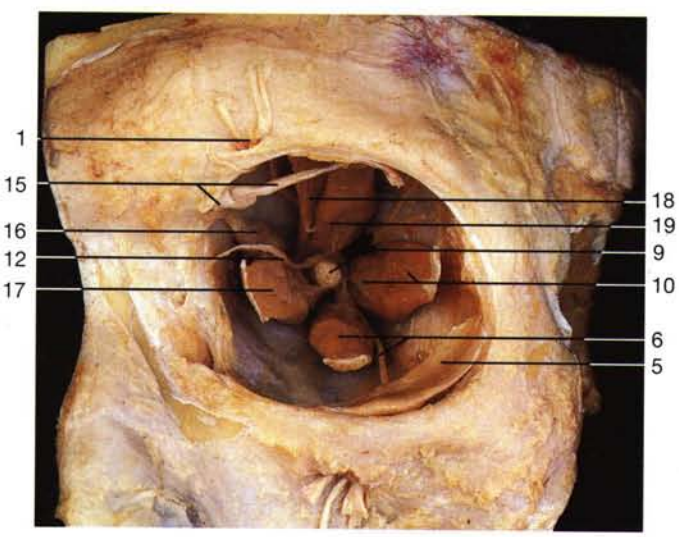

Left orbit with extraocular muscles (anterior aspect). Eyeball removed.

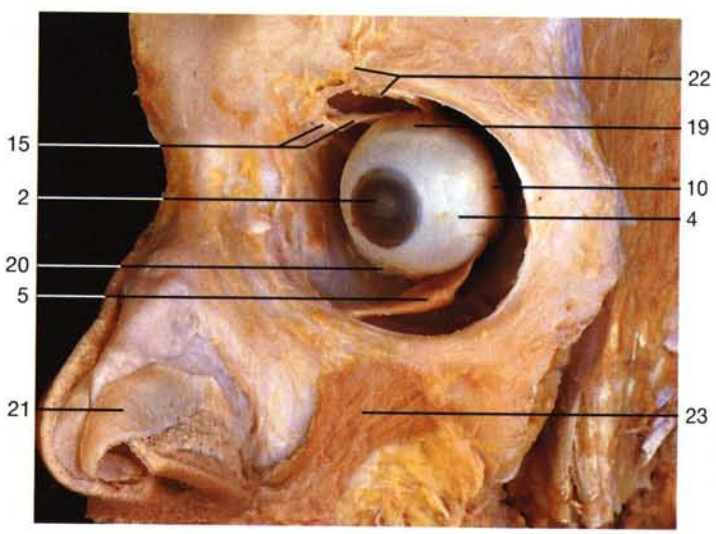

Extraocular eye muscles (anterior-lateral aspect).

1 Supraorbital nerve
2 Cornea
3 Insertion of lateral rectus muscle
4 Eyeball (sclera)
5 **Inferior oblique muscle**
6 **Inferior rectus muscle** and inferior branch of oculomotor nerve
7 Infraorbital nerve
8 Superior rectus muscle and lacrimal nerve
9 Optic nerve
10 **Lateral rectus muscle**
11 Ciliary ganglion and abducens nerve (n. VI)
12 Oculomotor nerve (n. III)
13 Trochlear nerve (n. IV)
14 Ophthalmic nerve (n. V_1) and maxillary nerve (n. V_2)
15 Trochlea and tendon of superior oblique muscle
16 **Superior oblique muscle**
17 **Medial rectus muscle**
18 Levator palpebrae superioris muscle
19 **Superior rectus muscle**
20 **Inferior rectus muscle**
21 Greater alar cartilage
22 Supraorbital nerve and levator palpebrae superioris muscle
23 Levator labii superioris muscle

Dissection of the visual pathway (inferior aspect). Frontal pole at top, midbrain divided.

1 Medial olfactory stria
2 Olfactory trigone
3 Lateral olfactory stria
4 Anterior perforated substance
5 Oculomotor nerve (n. III)
6 Mamillary body
7 Cerebral peduncle
8 **Lateral geniculate body**
9 Medial geniculate body
10 **Pulvinar of thalamus**
11 Optic radiation
12 Splenium of the corpus callosum (commissural fibers)
13 Cuneus
14 Olfactory bulb
15 Olfactory tract
16 **Optic nerve** (n. II)
17 Infundibulum
18 Anterior commissure
19 **Genu of optic radiation**
20 **Optic tract**
21 Interpeduncular fossa and posterior perforated substance
22 Trochlear nerve (n. IV)
23 Substantia nigra
24 Cerebral aqueduct
25 **Visual cortex**
26 Line of Gennari
27 Gyrus of striate cortex
28 **Calcarine sulcus**

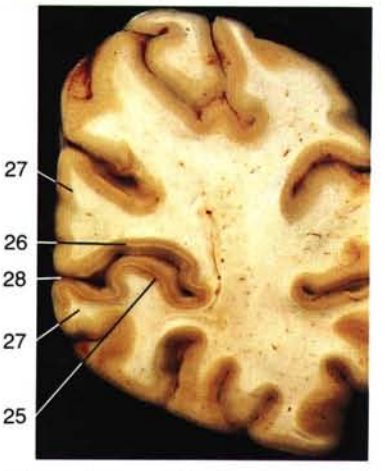

Frontal section of the striate cortex at the level of the striate area in the occipital lobe.

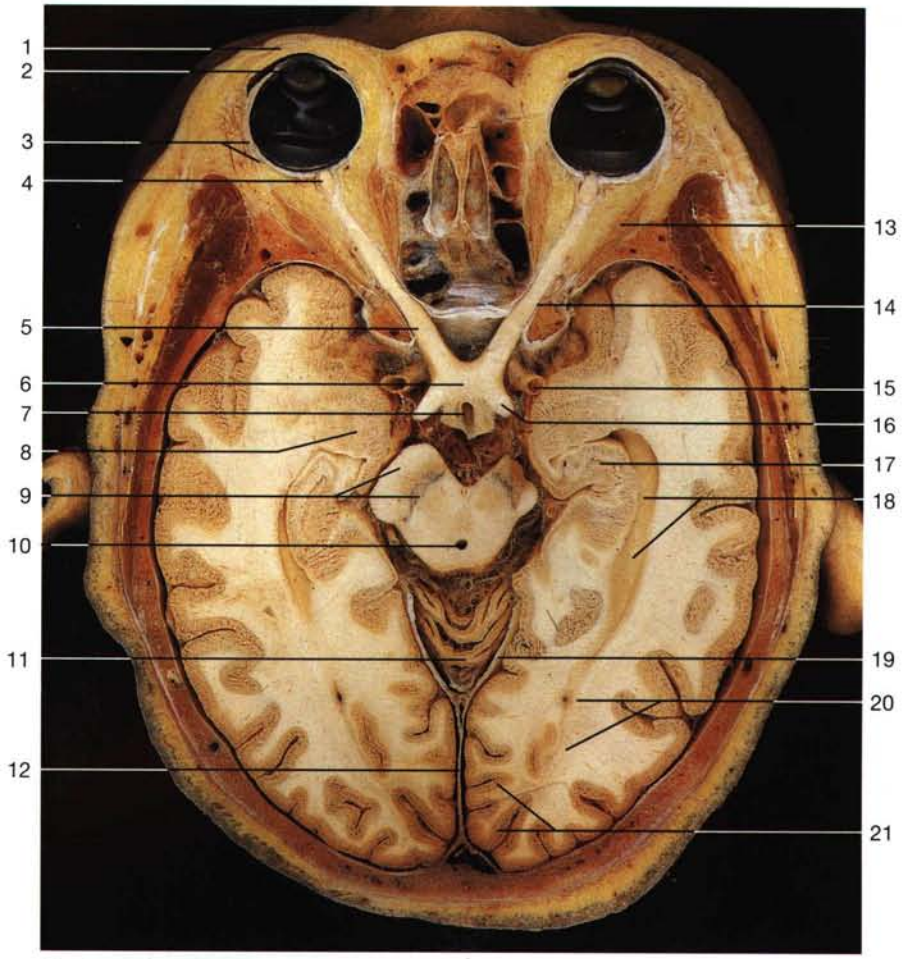

1 Upper lid
2 Cornea
3 **Eyeball** (sclera, retina)
4 Head of optic nerve
5 Optic nerve
6 **Optic chiasma**
7 Infundibular recess of hypothalamus
8 Amygdaloid body
9 Substantia nigra and crus cerebri
10 Cerebral aqueduct
11 Vermis of cerebellum
12 Falx cerebri
13 Lateral rectus muscle
14 Optic canal
15 Internal carotid artery
16 **Optic tract**
17 Hippocampus
18 Inferior horn of lateral ventricle
19 Tentorium cerebelli
20 **Optic radiation** of Gratiolet
21 **Visual cortex (area calcarina, striate cortex)**
22 Lens
23 **Eyeball**
24 Ethmoidal cells
25 Optic nerve with dura sheath
26 Cerebral peduncle
27 Aqueduct of mesencephalon
28 Vermis of cerebellum

Horizontal section through the head at the level of optic chiasma and striate cortex (superior aspect). Note the relationship of hypothalamic infundibulum to optic chiasma.

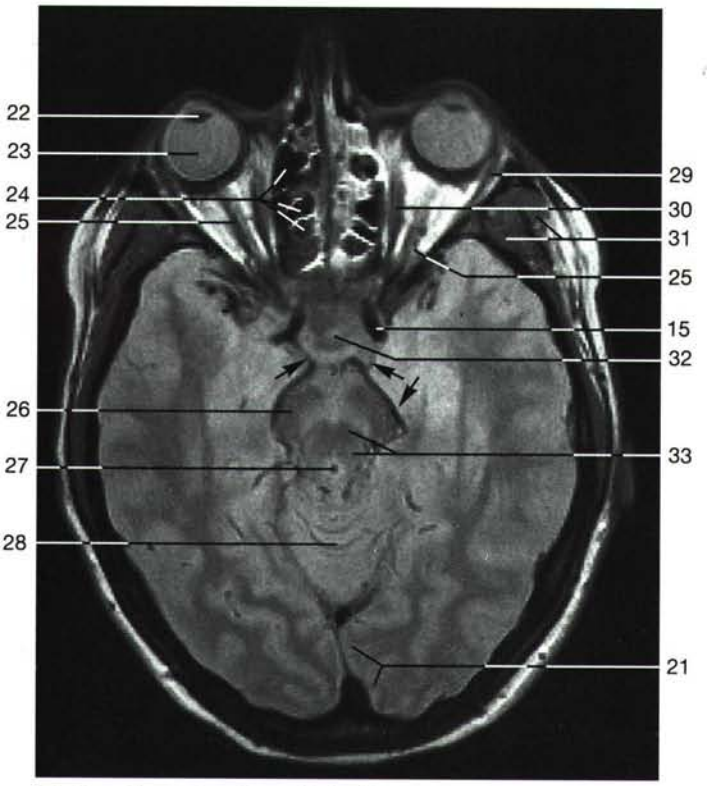

Horizontal section through the human head (MR-Scan, courtesy of Prof. W. J. Huk, Erlangen, FRG). Arrows = branches of arterial circle of Willis.

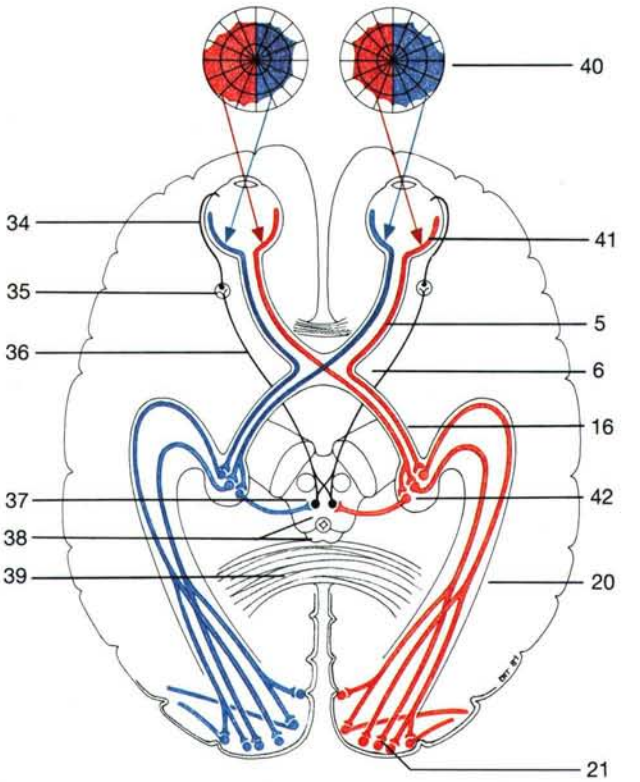

Diagram of the visual pathway and path of the light reflex.

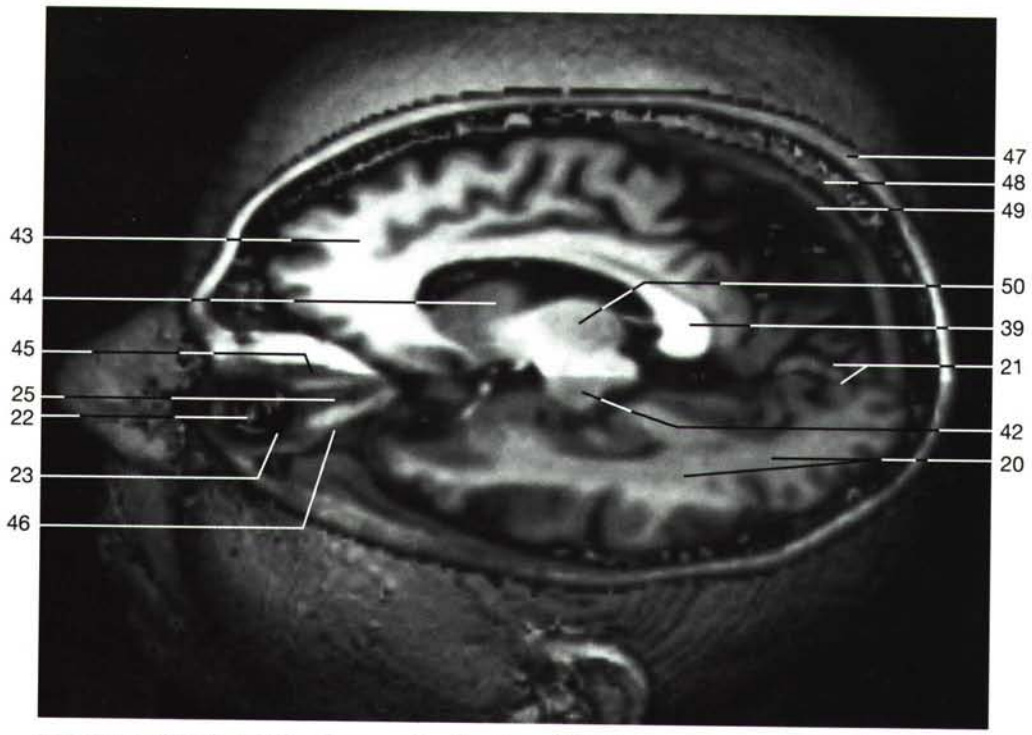

3-D-Reconstruction of the human visual system (MR-Image flash 40°, courtesy of Prof. W. J. Huk, University of Erlangen, FRG).

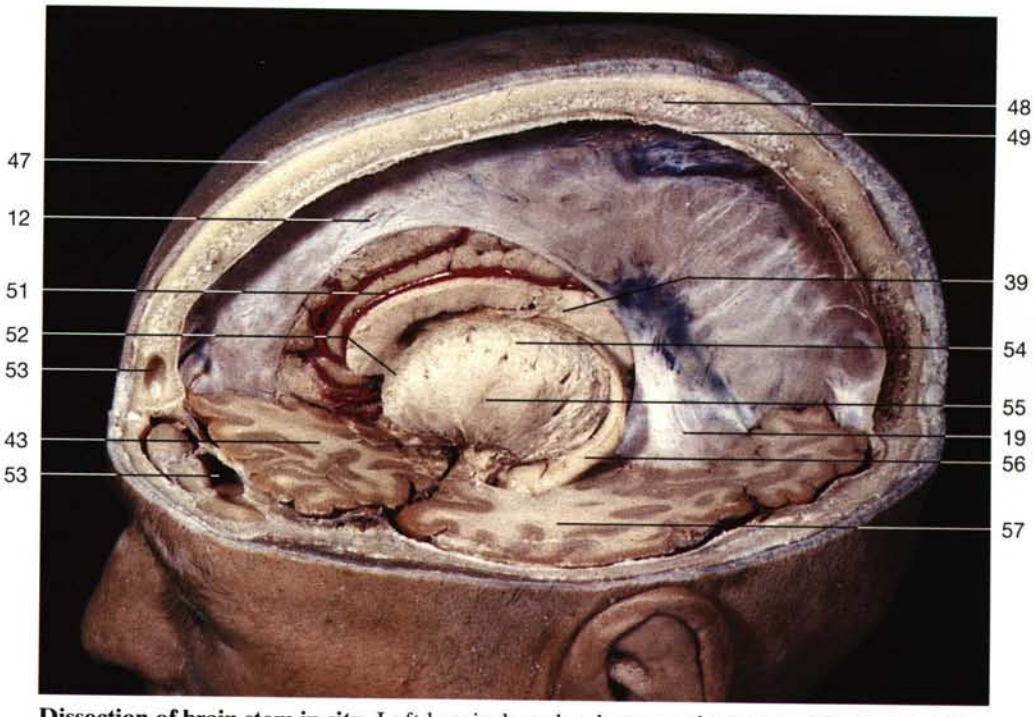

Dissection of brain stem in situ. Left hemisphere has been partly removed (compare with MR-Image above).

29 Lateral rectus muscle
30 Medial rectus muscle
31 Temporalis muscle
32 Hypophysis (pituitary gland)
33 Midbrain
34 Ciliary nerves (long and short)
35 Ciliary ganglion
36 Oculomotor nerve
37 Accessory oculomotor nucleus
38 Colliculi of midbrain
39 Corpus callosum
40 Visual field
41 Retina
42 **Lateral geniculate body**
43 Frontal lobe
44 **Caudate nucleus**
45 Medial rectus muscle
46 Lateral rectus muscle
47 Skin
48 Diploe (skull)
49 Dura mater
50 **Thalamus**
51 **Anterior cerebral artery**
52 Nucleus caudatus
53 Frontal sinus
54 **Internal capsule**
55 Lentiform nucleus (putamen)
56 **Hippocampus**
57 Temporal lobe of left hemisphere

In **binocular vision** the visual field (40) is projected upon portions of both retinae (blue and red in the drawing). In the chiasma the fibers from the two retinal portions are combined to form the left optic tract. The fibers of the two eyes remain separated from each other throughout the entire visual pathway up to their final termination in the calcarine cortex (21). **Injuries** on the optic pathway produce visual defects whose nature depends on the location of the injury. Destruction of one optic nerve produces blindness in the corresponding eye with loss of pupillary light reflex. If lesions of the **chiasma** destroy the crossing fibers of the nasal portions of the retina, both temporal fields of vision are lost (bitemporal hemianopsia). If both lateral angles of the chiasma are compressed, the nondecussating fibers from the temporal retinae are affected resulting in loss of nasal visual fields (binasal hemianopsia). Lesions posterior to the chiasma (i. e., optic tract, lateral geniculate body, optic radiation or visual cortex) result in a loss of the entire opposite field of vision (homonymous hemianopsia).

Superficial layer of the left orbit (superior aspect). The roof of the orbit and a portion of the left tentorium have been removed.

Middle layer of the left orbit (superior aspect). The roof of the orbit has been removed and the superior extraocular muscles have been divided and reflected.

1 Lateral branch of frontal nerve
2 Lacrimal gland
3 Lacrimal vein
4 Lacrimal nerve
5 **Frontal nerve**
6 Superior rectus
7 Middle cranial fossa
8 **Abducens nerve** (n. VI)
9 Trigeminal nerve (n. V)
10 **Trochlear nerve** (intracranial part) (n. IV)
11 Frontal sinus
12 Levator palpebrae superioris muscle
13 Branches of supratrochlear nerve
14 Olfactory bulb
15 Superior oblique muscle
16 **Trochlear nerve** (intraorbital part) (n. IV)
17 **Optic nerve** (intracranial part)
18 Pituitary gland and infundibulum
19 Dorsum sellae
20 **Oculomotor nerve** (n. III)
21 Midbrain
22 Tendon of superior oblique muscle
23 Eyeball
24 Vena vorticosa
25 Short ciliary nerves
26 Optic nerve (extracranial part)
27 **Trigeminal ganglion**

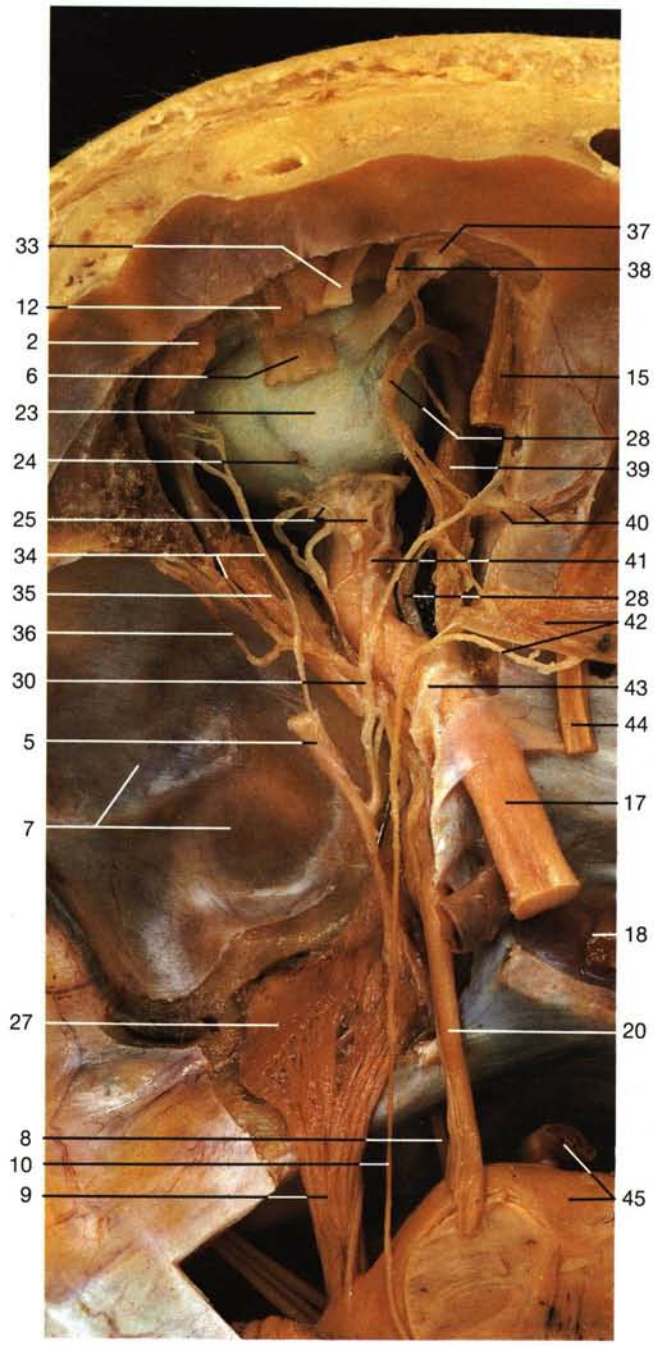

Middle layer of the left orbit (superior aspect). The roof of the orbit and the superior extraocular muscles have been removed.

Deeper layer of the left orbit (superior aspect). The optic nerve has now been removed.

28 **Ophthalmic artery**
29 Superior ophthalmic vein
30 **Nasociliary nerve**
31 Levator palpebrae superioris muscle (reflected)
32 Superior rectus muscle (reflected)
33 Lateral branch of supraorbital nerve
34 Lacrimal nerve and artery
35 Lateral rectus muscle
36 Meningolacrimal artery (anastomosing with middle meningeal artery)
37 Trochlea
38 Medial branch of supraorbital nerve
39 Medial rectus muscle
40 Anterior ethmoidal artery and nerve
41 Long ciliary nerve
42 Superior oblique muscle and trochlear nerve
43 Common tendinous ring
44 Olfactory tract
45 Basilar artery and pons
46 Optic nerve (external sheath of optic nerve, divided)
47 **Ciliary ganglion**
48 Ophthalmic nerve (divided, reflected)
49 Inferior branch of oculomotor nerve and inferior rectus muscle
50 Superior branch of oculomotor nerve
51 Internal carotid artery

Eyelids (superficial layer, right side).

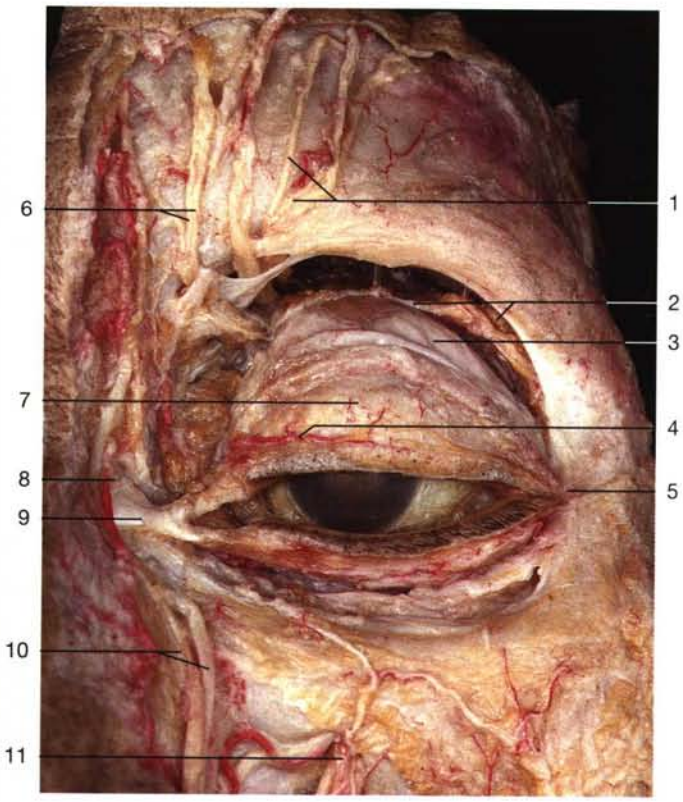

Eyelids (deeper layer, left side).

1 Supraorbital artery, supraorbital nerve, lateral branch
2 **Lacrimal gland**
3 Aponeurosis of levator palpebrae superioris muscle
4 Arterial arch of upper eyelid
5 Lateral palpebral ligament
6 Supratrochlear artery, supratrochlear nerve
7 Upper eyelid, tarsal plate
8 Lacrimal sac
9 **Medial palpebral ligament**
10 Angular artery and vein
11 Infraorbital artery, vein and nerve
12 Orbicularis oculi muscle
13 Infratrochlear nerve
14 Levator labii superioris alaeque nasi muscle
15 Facial artery and vein
16 **Superior lacrimal canaliculus**
17 **Inferior lacrimal canaliculus**
18 Lacrimal papilla and punctum
19 Lacrimal bone
20 **Lacrimal duct**
21 Mucous membrane of nasal cavity
22 Palpebral conjunctiva of lower eyelid
23 Lacrimal sac and superior lacrimal canaliculus
24 Lateral fixation of levator aponeurosis
25 Infraorbital foramen

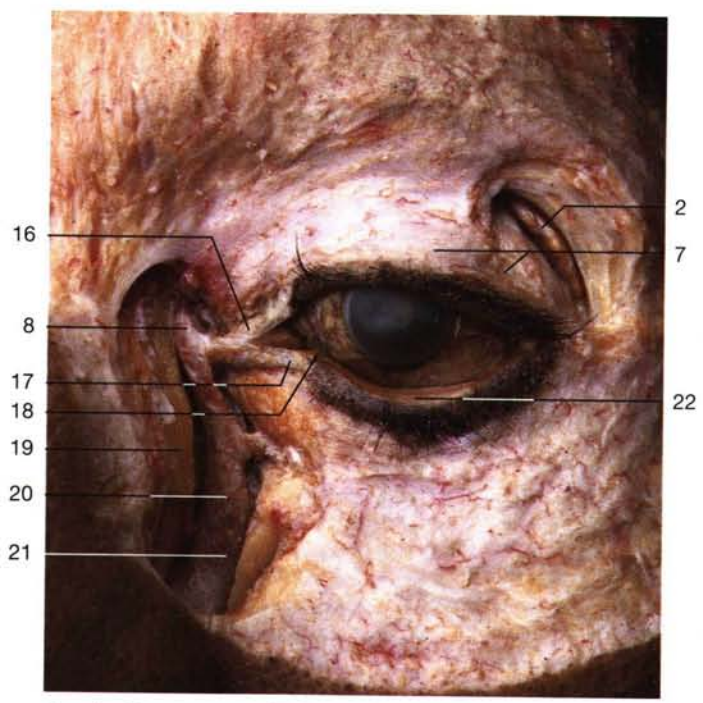

Lacrimal apparatus of left eye (anterior aspect).

Lacrimal apparatus of left eye (anterior aspect).

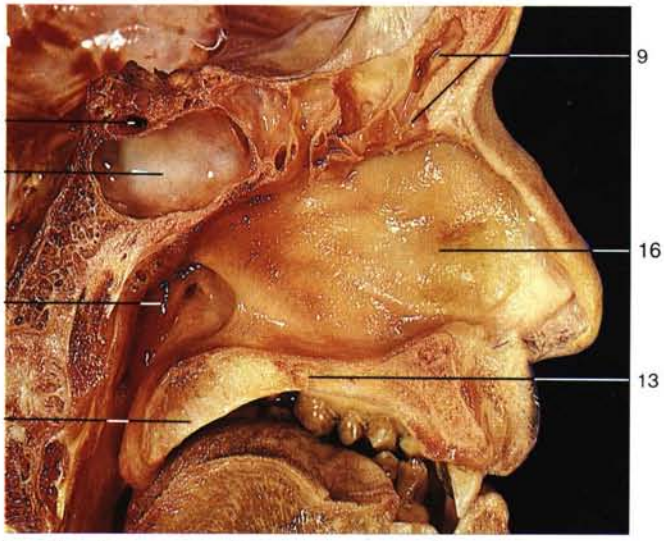

Nasal septum, covered by a mucous membrane.

Nasal septum. Mucous membrane removed.

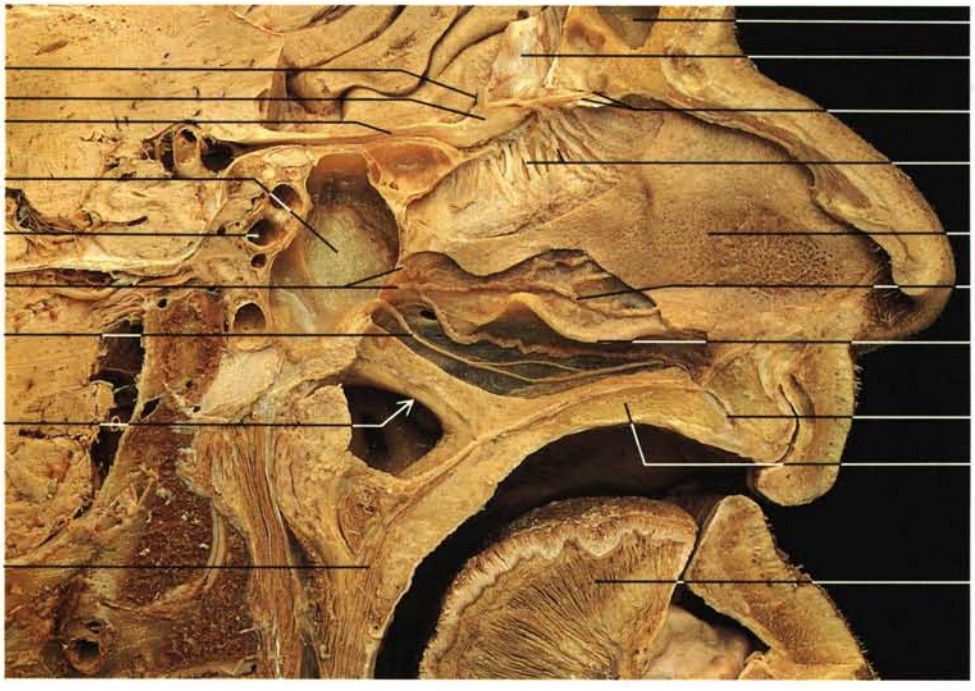

Nasal septum. Dissection of nerves and vessels.

1 Crista galli
2 Pituitary gland and sella turcica
3 **Sphenoidal sinus** (relatively large)
4 Tubal elevation
5 Pharyngeal opening of the **auditory tube**
6 Pharyngeal recess
7 Atlas (anterior arch)
8 Soft palate
9 Frontal sinus
10 Perpendicular plate of ethmoid
11 **Cartilage of nasal septum**
12 Vomer
13 Hard palate
14 Nasal branch of anterior ethmoidal artery and anterior ethmoidal nerve
15 Nasopharynx
16 Nasal septum
17 **Olfactory nerves**
18 Septal artery
19 Crest of nasal septum
20 Incisive canal
21 Anterior ethmoidal artery
22 Olfactory bulb
23 Olfactory tract
24 Internal carotid artery
25 Posterior nasal and septal arteries
26 Nasopalatine nerve
27 **Choana** (arrow)
28 Tongue

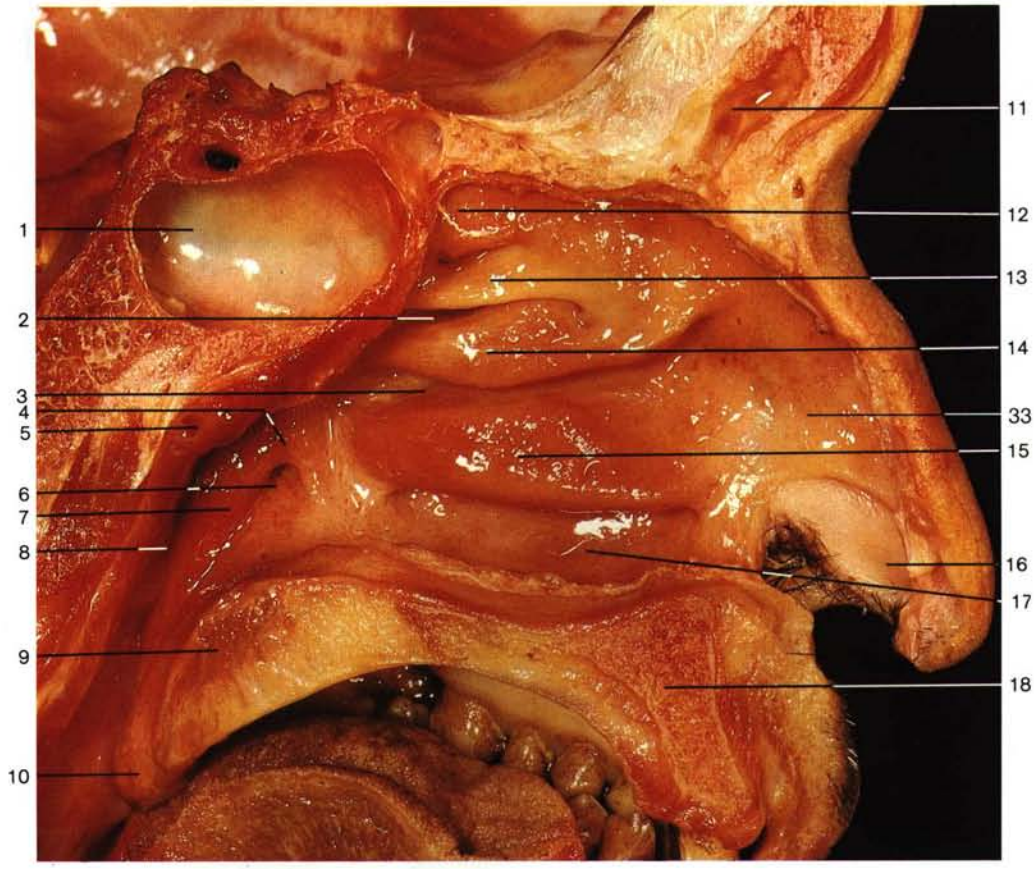

Lateral wall of the nasal cavity, septum removed.

1 Sphenoidal sinus
2 Superior meatus
3 Middle meatus
4 Tubal elevation
5 Pharyngeal tonsil
6 Pharyngeal orifice of **auditory tube**
7 Salpingopharyngeal fold
8 Pharyngeal recess
9 **Soft palate**
10 **Uvula**
11 Frontal sinus
12 Sphenoethmoidal recess
13 **Superior nasal concha**
14 **Middle nasal concha**
15 **Inferior nasal concha**
16 Vestibule
17 Inferior meatus
18 **Hard palate**
19 Grooves for the middle meningeal artery and parietal bone (light green)
20 **Maxillary hiatus**
21 Perpendicular process of palatine bone
22 Openings of ethmoidal air cells
23 Opening of frontal sinus
24 Medial pterygoid plate (red)
25 Horizontal plate of palatine process
26 **Ethmoidal air cells**
27 **Maxillary sinus**
28 Nasal septum
29 Pterygoid hamulus
30 Nasal bone (white)
31 Frontal process of maxilla (violet)
32 Palatine process of maxilla (violet)
33 Nasal atrium

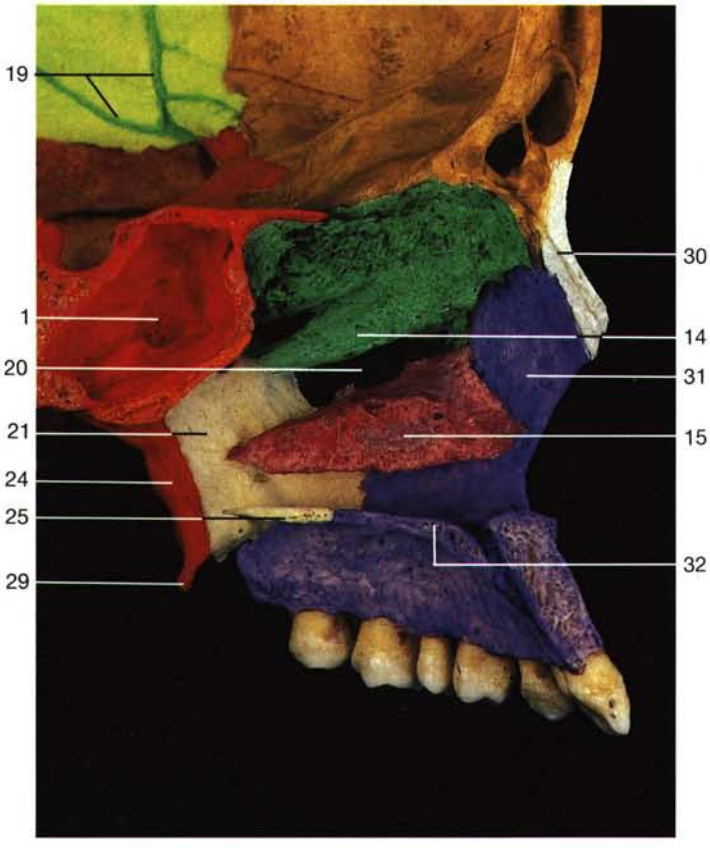

Bones of left nasal cavity, medial aspect.

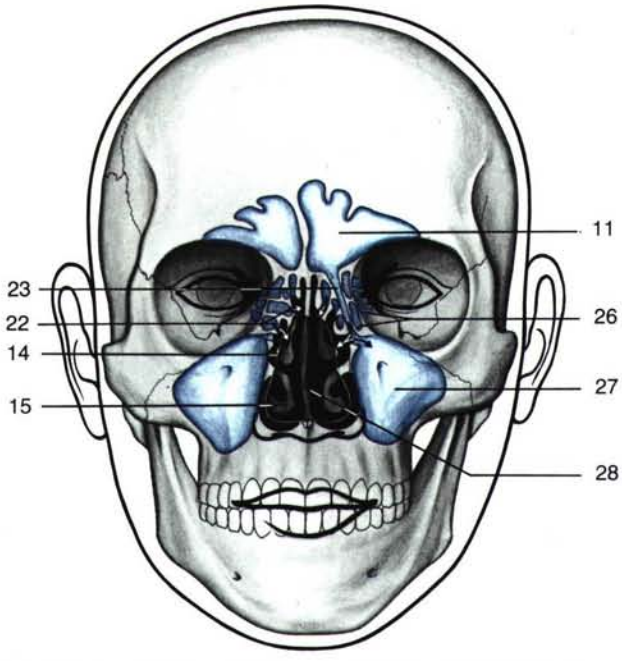

Schematic diagram showing the position of paranasal sinuses, openings indicated by arrows.

Median section through the head with nasal and oral cavity. The middle and inferior nasal conchae have been partly removed to show the openings of paranasal sinuses.

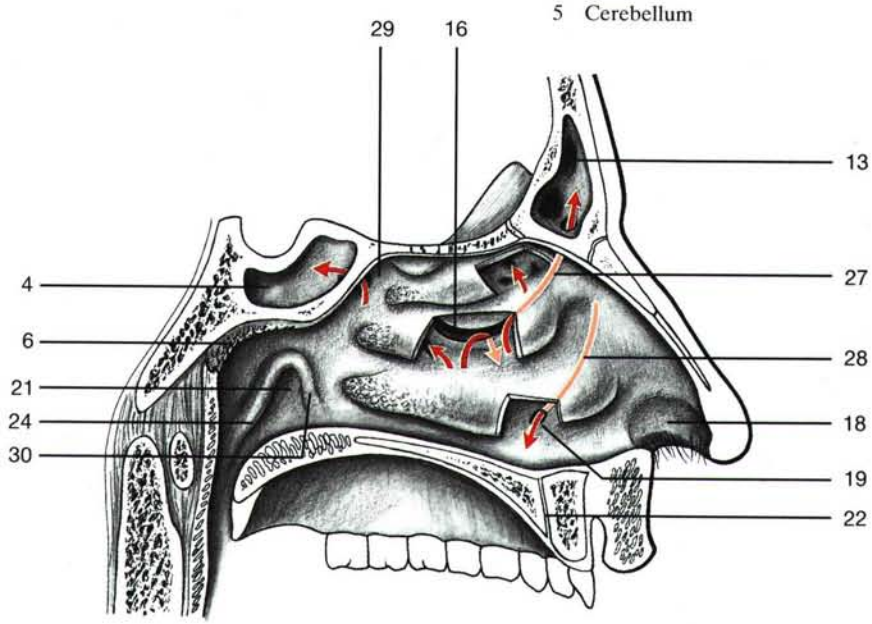

Lateral wall of nasal cavity. Openings indicated by red arrows. (Schematic drawing.)

1 Great cerebral vein (Galen's vein)
2 **Tectum of midbrain**
3 Straight sinus
4 **Sphenoidal sinus**
5 Cerebellum
6 **Pharyngeal tonsil**
7 **Cerebellomedullary cistern**
8 Median atlantoaxial joint
9 Spinal cord
10 Oral part of pharynx
11 Falx cerebri
12 Corpus callosum and anterior cerebral artery
13 **Frontal sinus**
14 Optic chiasm and pituitary gland
15 Superior nasal concha and ethmoidal bulla
16 **Semilunar hiatus**
17 Accessory **openings to maxillary sinus** and cut edge of middle nasal concha
18 Vestibule
19 **Opening of nasolacrimal duct**
20 Inferior nasal concha (cut)
21 **Opening of auditory tube**
22 **Incisive canal**
23 Levator veli palatini muscle
24 Salpingopharyngeal fold
25 Lingual nerve and submandibular ganglion
26 Submandibular duct
27 **Nasofrontal duct**
28 **Nasolacrimal duct**
29 Sphenoethmoidal recess (of Rosenmüller)
30 Salpingopalatine fold

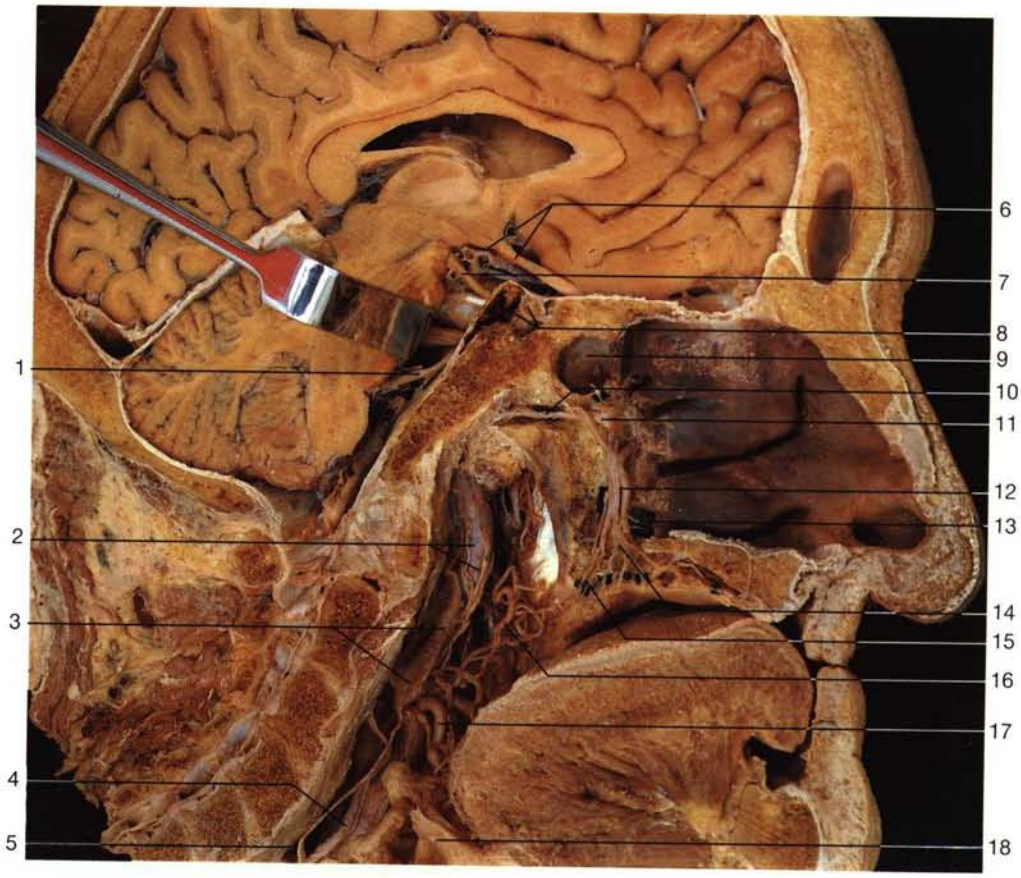

Nerves of the lateral wall of nasal cavity I. Sagittal section through the head. Mucous membranes partly removed, pterygoid canal opened.

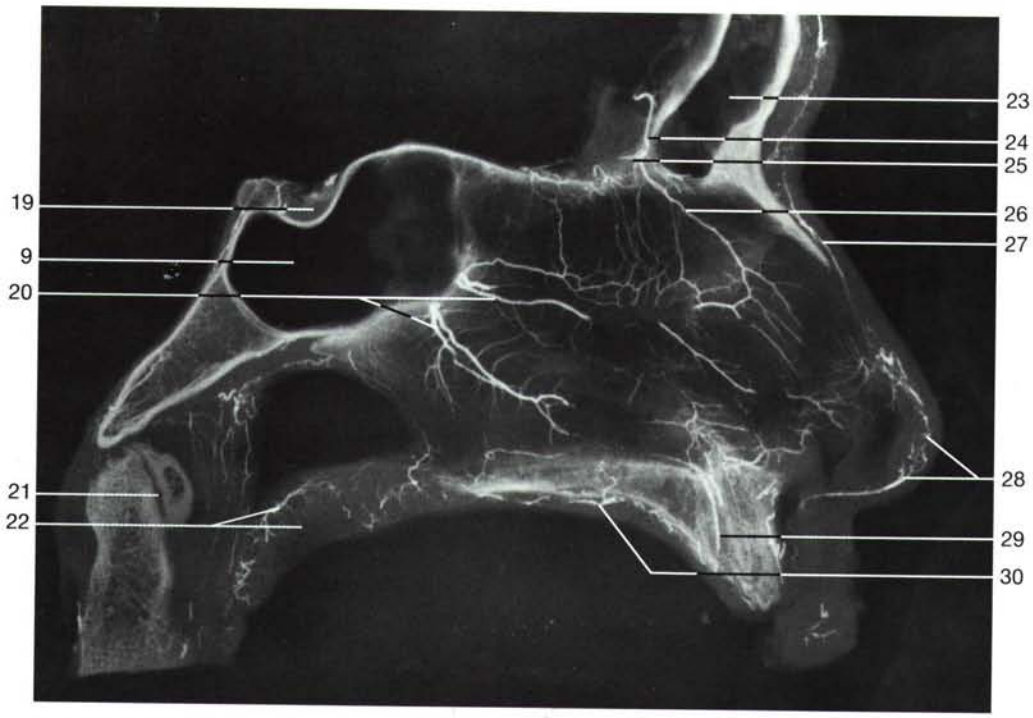

Arteriogram of the nasal septum, left side (lateral aspect).

1 Facial nerve
2 **Internal carotid artery** and **internal carotid plexus**
3 **Superior cervical ganglion**
4 Vagus nerve
5 Sympathetic trunk
6 Optic nerve and ophthalmic artery
7 Oculomotor nerve
8 Internal carotid artery and cavernous sinus
9 Sphenoidal sinus
10 Nerve of the pterygoid canal
11 **Pterygopalatine ganglion**
12 **Descending palatine artery**
13 Lateral inferior posterior nasal branches and lateral posterior nasal and septal arteries
14 Greater palatine nerves and artery
15 Lesser palatine nerves and arteries
16 Branches of ascending pharyngeal artery
17 Lingual artery
18 Epiglottis
19 **Sella turcica**
20 **Posterior lateral nasal and septal arteries** (branches of sphenopalatine artery)
21 Median atlanto-axial joint
22 Soft palate and lesser palatine arteries (branches of descending palatine artery)
23 Frontal sinus
24 Anterior meningeal artery
25 Anterior ethmoidal artery (branch of ophthalmic artery)
26 **Septal branch of anterior ethmoidal artery**
27 Dorsal nasal artery
28 Nasal branches of anterior ethmoidal artery
29 Incisive canal with nasopalatine artery
30 Hard palate and greater palatine artery (branch of descending palatine artery)
31 Tentorium cerebelli
32 Trochlear nerve
33 **Trigeminal nerve** with motor root
34 **Internal carotid plexus**
35 **Lingual nerve with chorda tympani**
36 Medial pterygoid muscle and medial pterygoid plate
37 **Inferior alveolar nerve**
38 **Sympathetic trunk**
39 Oculomotor nerve
40 **Palatine nerves**
41 Tongue
42 **Tringeminal ganglion**
43 Trigeminal nerve (n. V)
44 **Facial nerve** (n. VII)
45 Geniculate ganglion
46 Stylomastoid foramen
47 Medial pterygoid muscle

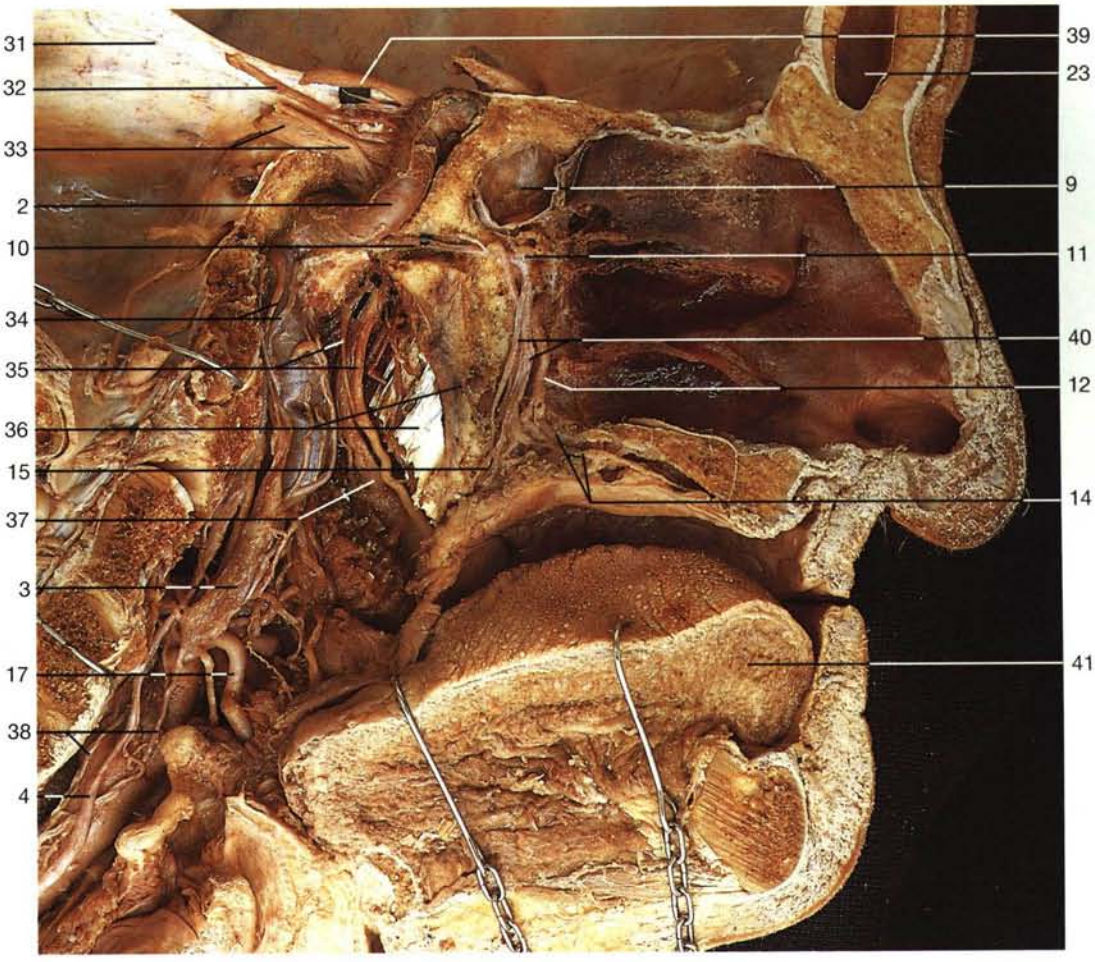

48 **Greater petrosal nerve**
49 Maxillary nerve
50 Olfactory bulb
51 **Olfactory nerves**
52 Internal nasal branches of anterior ethmoidal nerve
53 Lateral superior posterior nasal branches
54 Lateral inferior posterior nasal branches
55 Incisive canal with nasopalatine nerve
56 **Greater palatine nerve**
57 **Deep petrosal nerve**
58 Mandibular nerve
59 Nasal cavity and inferior nasal concha
60 Opening of auditory tube
61 Tensor veli palatini muscle
62 Levator veli palatini muscle
63 Pharyngeal recess in the nasopharynx
64 Uvula
65 Palatoglossal arch
66 Tonsillar branch of ascending palatine artery
67 **Palatine tonsil**
68 Palatopharyngeal arch

Nerves of the lateral wall of nasal cavity II. Carotid canal opened, mucous membranes of pharynx and nasal cavity partly removed.

Dissection of palatine tonsil located in the lateral wall of the nasopharynx (left side). Root of tongue reflected.

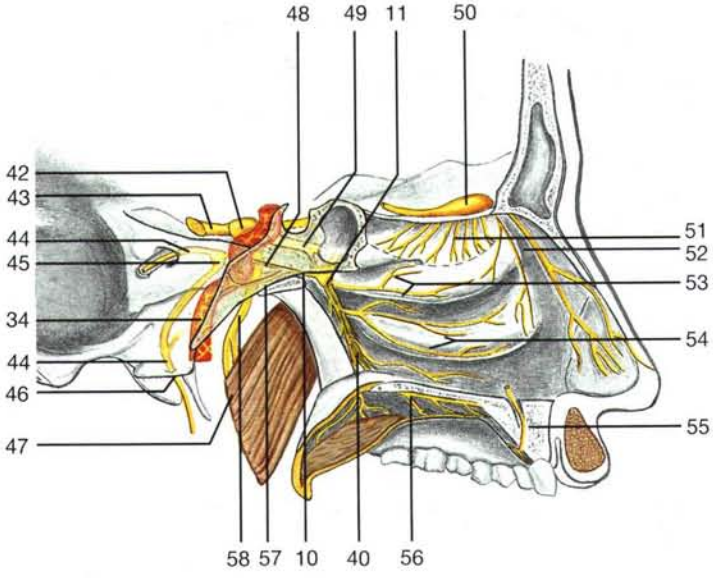

Nerves of the lateral wall of nassal cavity. Body of sphenoid bone appears transparent (schematic drawing).

Horizontal section through the nasal cavity, the orbits and temporal lobes of the brain at the level of pituitary gland.

Horizontal section through the head. CT-Scan. Bar = 2 cm.
Arrow: fracture.

1 Cornea
2 Lens
3 Vitreous body (**eyeball**)
4 Head of optic nerve
5 Medial rectus muscle
6 Lateral rectus muscle
7 **Optic nerve** with dural sheath
8 Internal carotid artery
9 **Pituitary gland** and **infundibulum**
10 **Oculomotor nerve**
11 Superior tarsal plate of eye lid
12 Fornix of conjunctiva
13 **Nasal cavity**
14 Sclera
15 Ethmoidal sinus
16 **Nasal septum**
17 **Sphenoidal sinus**
18 Temporal lobe
19 Clivus
20 Middle cranial fossa
21 External acoustic meatus
22 Superior sagittal sinus
23 Falx cerebri
24 Superior rectus and levator palpebrae superioris muscles
25 Eyeball and lacrimal gland
26 Inferior rectus muscle
27 Zygomatic bone
28 **Maxillary sinus**
29 Inferior nasal concha
30 Hard palate
31 Superior longitudinal muscle of tongue
32 Median fibrosus septum of tongue

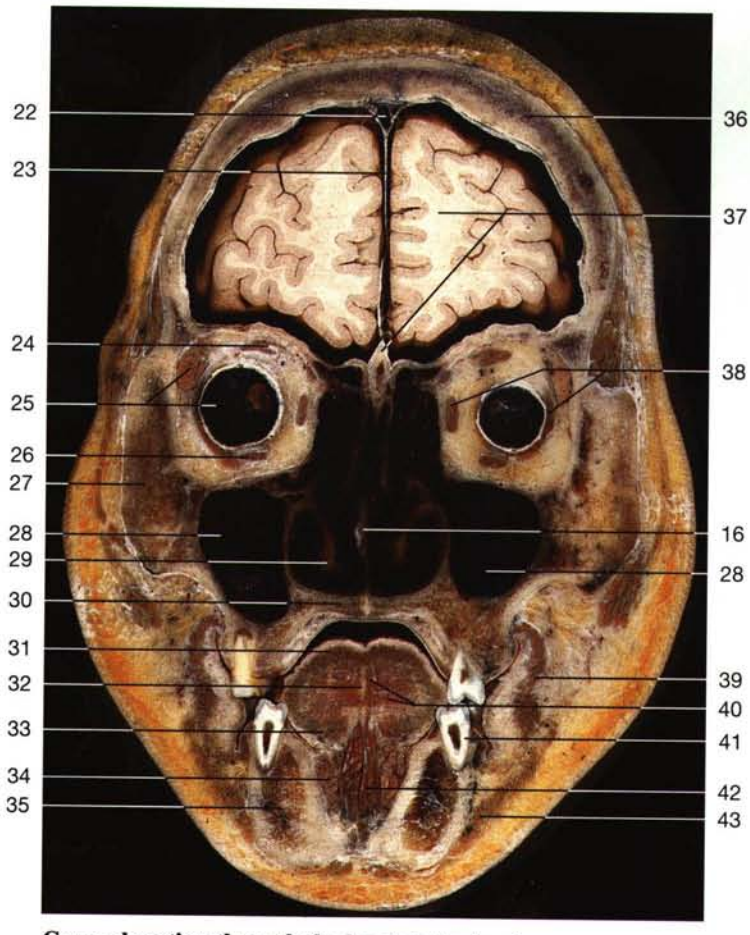

Coronal section through the head at the level of the second premolar of the mandible.

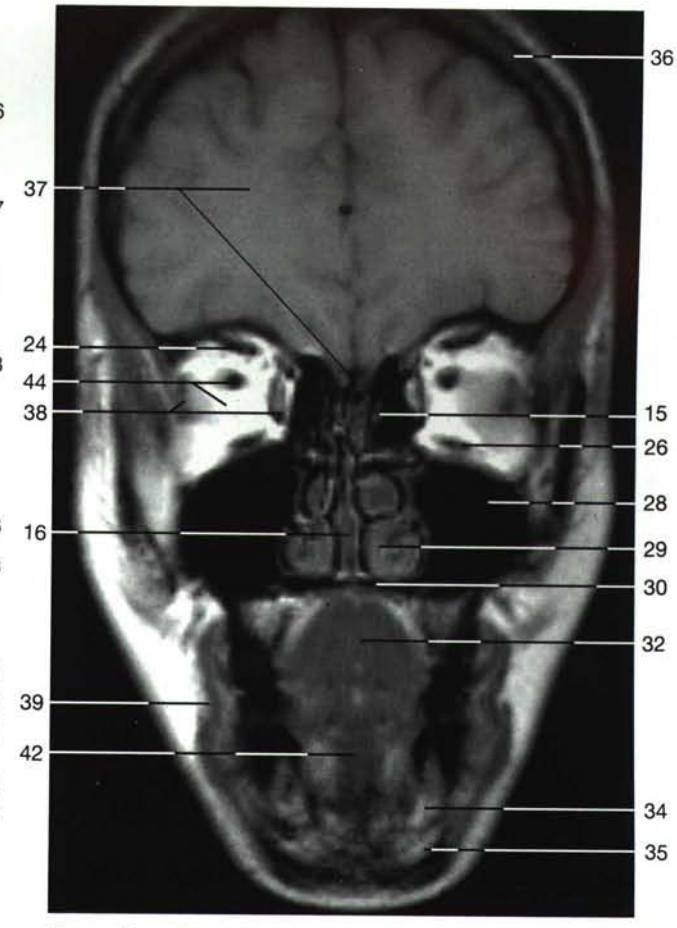

Coronal section through the head (MR-Scan, courtesy of Dr. A. Heuck, Munich). Note the situation of the head cavities.

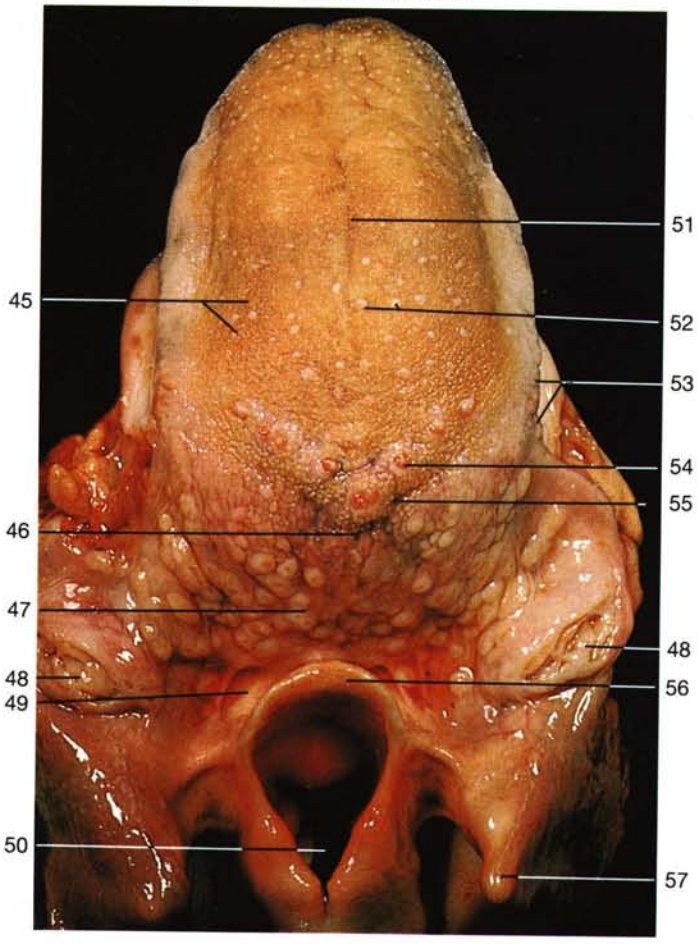

33 Inferior longitudinal muscle of tongue
34 **Sublingual gland**
35 Mandible
36 Calvaria
37 Frontal lobe of brain and crista galli
38 Lateral and medial eye rectus muscles
39 Buccinator muscle
40 Vertical and transverse muscles of tongue
41 Second premolar of mandible
42 Genioglossus muscle
43 Platysma muscle
44 **Orbit** and **optic nerve**
45 Filiform papillae
46 **Foramen cecum**
47 **Root of tongue** (lingual tonsil)
48 **Palatine tonsil**
49 Vallecula of epiglottis
50 Vestibule of larynx
51 Median sulcus of tongue
52 Fungiform papillae
53 Foliate papillae
54 **Circumvallate papilla**
55 Sulcus terminalis
56 Epiglottis
57 Greater cornu of hyoid bone

Dorsal surface of the tongue and laryngeal inlet.

Median sagittal section through the oral cavity and pharynx.

1 Nasal cavity
2 **Hard palate**
3 Upper lip and orbicularis oris muscle
4 Vestibule of oral cavity
5 First incisor
6 Lower lip and orbicularis oris muscle
7 Mandible
8 Genioglossus muscle
9 Geniohyoid muscle
10 Anterior belly of diagastric muscle
11 Mylohyoid muscle
12 **Hyoid bone**
13 **Nasopharynx**
14 **Soft palate** and **uvula**
15 **Oropharynx**
16 Root of tongue and lingual tonsil
17 **Laryngopharynx**
18 Epiglottis
19 Aryepiglottic fold
20 Laryngopharynx continuous with esophagus
21 Larynx

Hyoid bone (oblique lateral aspect).

Hyoid bone (anterior aspect).

1 Greater cornu
2 Lesser cornu
} of hyoid bone
3 Body

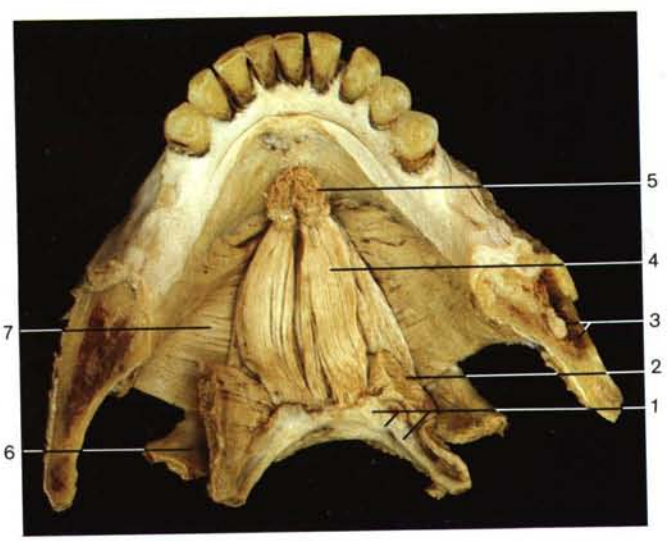

Muscles of the floor of the oral cavity (superior aspect).

1 Lesser cornu and body of hyoid bone
2 Hyoglossus muscle (divided)
3 Ramus of mandible and inferior alveolar nerve
4 **Geniohyoid muscle**
5 Genioglossus muscle (divided)
6 Stylohyoid muscle (divided)

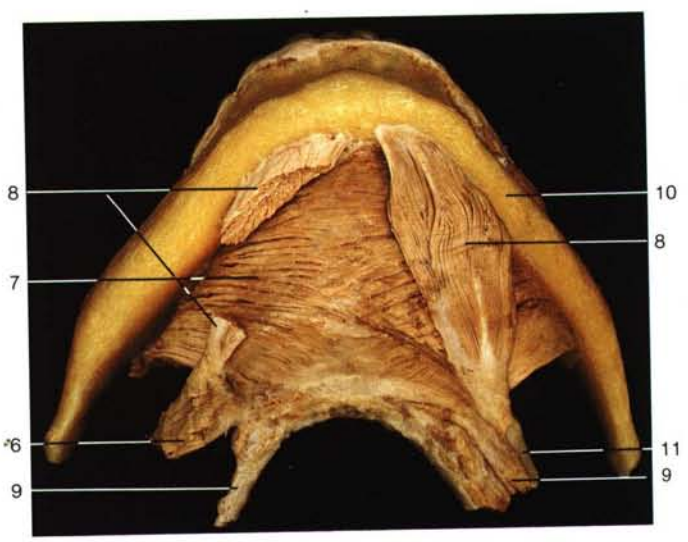

Oral diaphragm, muscles (inferior aspect). Cut on the base.

7 **Mylohyoid muscle**
8 Anterior belly of **digastric muscle**
9 Hyoid bone
10 Mandible
11 Intermediate tendon of digastric muscle

Parapharyngeal and sublingual regions. Innervation of the tongue. Lateral part of face and mandible removed, oral cavity opened. Arrow: submandibular duct.

1 Styloid process
2 Styloglossus muscle
3 Digastric muscle (posterior belly)
4 Vagus nerve (n. X)
5 **Lingual nerve** (n. V_3)
6 **Glossopharyngeal nerve** (n. IX)
7 Submandibular ganglion
8 Hyoglossus muscle
9 **Hypoglossal nerve** (n. XII)
10 Stylohyoid muscle
11 Internal branch of superior laryngeal nerve (branch of vagus nerve, not visible)
12 Middle constrictor muscle of pharynx
13 Omohyoid muscle (divided)
14 Thyrohyoid muscle
15 Sternothyroid muscle
16 Esophagus
17 Parotid duct (divided)
18 Buccinator
19 Superior constrictor muscle of pharynx
20 Tongue
21 Terminal branches of lingual nerve
22 Mandible (divided)
23 Genioglossus and geniohyoid muscles
24 Mylohyoid muscle (divided and reflected)
25 Sternohyoid muscle (divided)
26 Thyroid cartilage
27 Anterior belly of digastric muscle
28 Hyoid bone

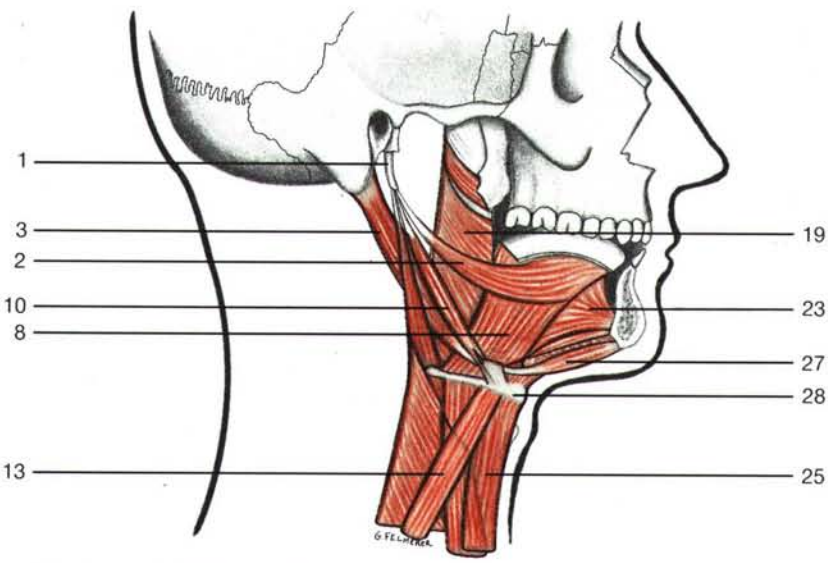

Supra- and infrahyoid muscles and pharynx (schematic drawing).

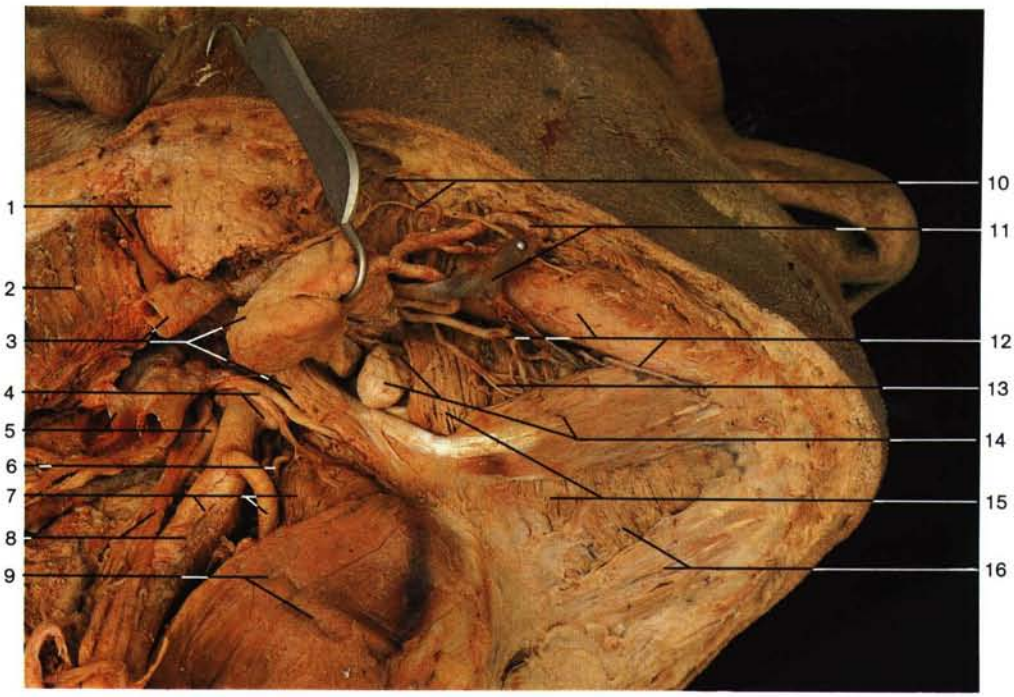

Submandibular triangle, superficial dissection. Right side (inferior aspect). Submandibular gland has been reflected.

Submandibular triangle, deep dissection. Right side. Mylohyoid muscle has been severed and reflected to display the lingual and hypoglossal nerves.

1 Parotid gland and retromandibular vein
2 Sternocleidomastoid
3 Retromandibular vein, submandibular gland and stylohyoid muscle
4 Hypoglossal nerve and lingual artery
5 Vagus nerve and internal jugular vein
6 Superior laryngeal artery
7 **External carotid artery,** thyrohyoid muscle, and superior thyroid artery
8 Common carotid artery and superior root of ansa cervicalis
9 Omohyoid and sternohyoid
10 Masseter and marginal mandibular branch of facial nerve
11 **Facial artery and vein**
12 Mandible and submental artery and vein
13 Mylohyoid nerve
14 Submandibular duct, sublingual gland, and anterior belly of digastric muscle
15 Mylohyoid (right side)
16 Left mylohyoid and anterior belly of left digastric muscle
17 Hyoglossus muscle and lingual artery
18 **Lingual nerve**
19 **Hypoglossal nerve**
20 Geniohyoid muscle
21 Anterior belly of right digastric muscle
22 Submandibular gland and duct

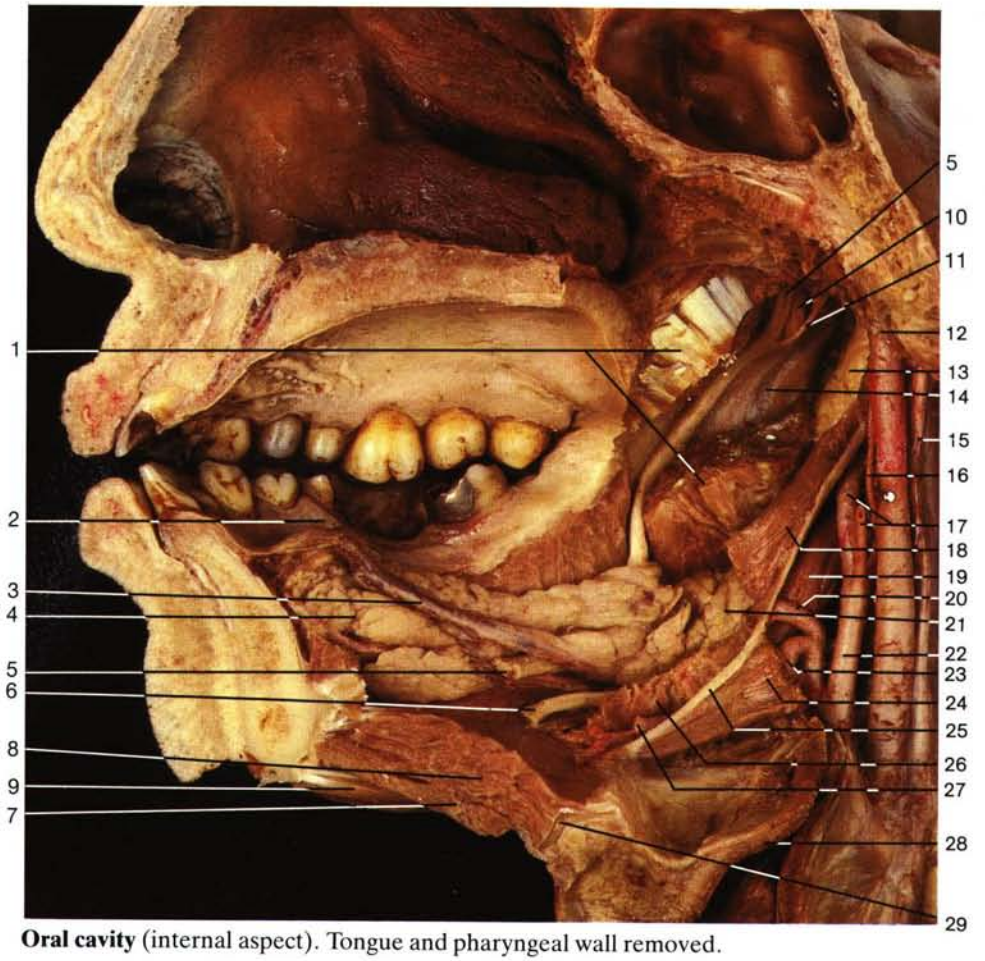

Oral cavity (internal aspect). Tongue and pharyngeal wall removed.

1 Medial pterygoid muscle
2 Sublingual papilla
3 **Submandibular duct**
4 **Sublingual gland**
5 **Lingual nerve**
6 Hypoglossal nerve
7 Mylohyoid muscle
8 Geniohyoid muscle
9 Anterior belly of digastric muscle
10 Inferior alveolar nerve
11 **Chorda tympani**
12 Internal carotid artery
13 Parotid gland
14 Sphenomandibular ligament
15 Vagus nerve
16 Glossopharyngeal nerve
17 Superficial temporal artery and ascending pharyngeal artery
18 Styloglossus muscle
19 Posterior belly of digastric muscle
20 **Facial artery**
21 Submandibular gland
22 **External carotid artery**
23 Lingual artery
24 Middle pharyngeal constrictor muscle
25 **Stylohyoid ligament**
26 Hyoglossus muscle
27 Deep lingual artery
28 Epiglottis
29 Hyoid bone
30 Buccinator muscle
31 Tongue
32 Mandible (divided)
33 **Parotid duct**
34 Masseter muscle
35 Right and left sublingual papilla

Dissection of major salivary glands. Left mandible and buccinator muscle partly removed to view the oral cavity (inferior lateral aspect).

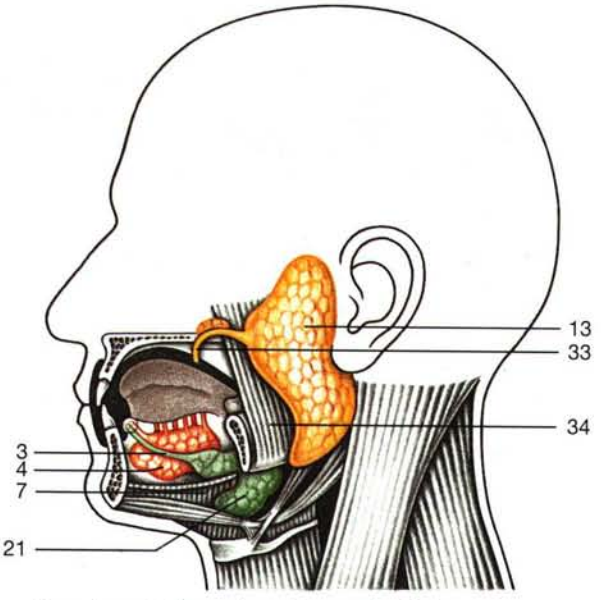

Location of the major salivary glands in relation to the oral cavity.

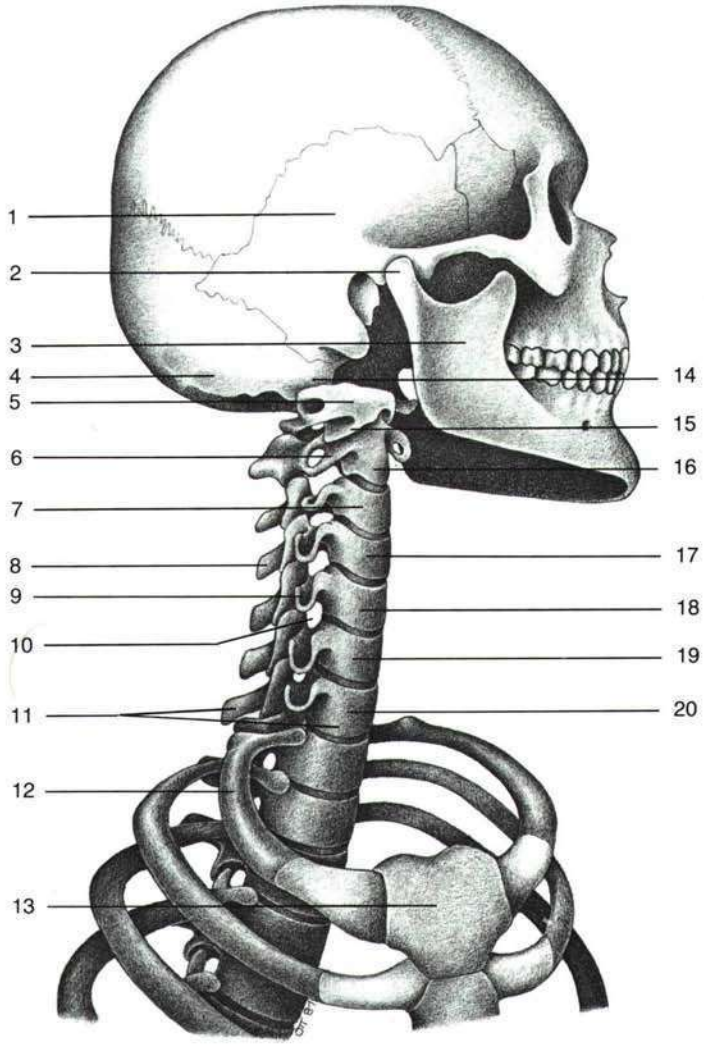

Cervical spine (oblique lateral aspect).

1 Temporal bone
2 Head of mandible
3 Mandible
4 Occipital bone
5 Atlas
6 Axis
7 Third cervical vertebra
8 Spinous process of fourth cervical vertrebra
9 Transverse process of fifth cervical vertebra with groove for spinal nerve
10 Intervertebral foramen
11 Vertebra prominens (C_7)
12 First rib
13 Sternum (manubrium sterni)
14 Atlantooccipital joint
15 Lateral atlantoaxial joint
16 Body of axis
17 Body of fourth cervical vertebra
18 Body of fifth cervical vertebra
19 Body of sixth cervical vertebra
20 Body of seventh cervical vertebra

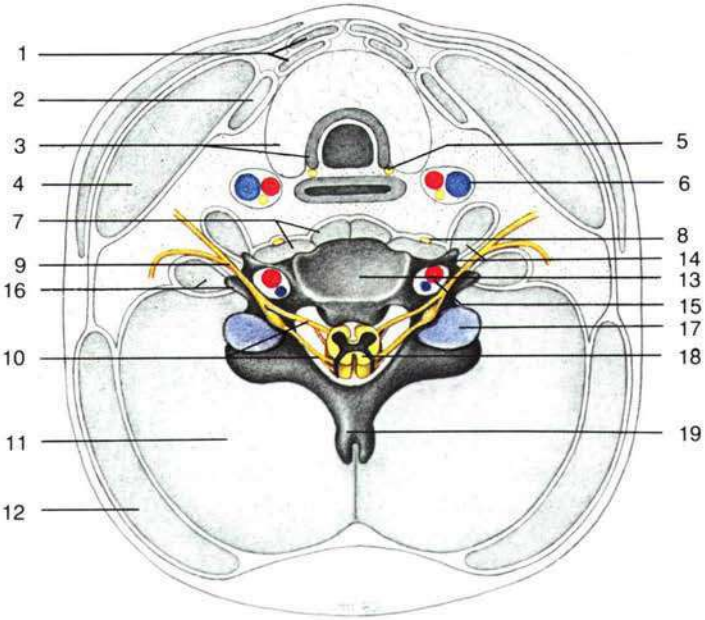

Cervical vertebra and the **organization of the neck.** (Schematic drawing.)

1 Sternohyoid and sternothyroid muscles
2 Omohyoid muscle
3 Thyroid gland and trachea
4 Sternocleidomastoid muscle
5 Recurrent laryngeal nerve
6 Internal jugular vein, common carotid artery and vagus
7 Longus colli and longus capitis muscles
8 Sympathetic trunk
9 Spinal nerve
10 Ventral and dorsal root of spinal nerve
11 True muscles of the neck
12 Trapezius muscle
13 Body of cervical vertebra
14 Anterior tubercle of transverse process and origin of scalenus anterior muscle
15 Vertebral artery and foramen transversarium
16 Posterior tubercle of transverse process and origin of scalenus medius and posterior muscles
17 Superior facet of articular process
18 Spinal cord
19 Spinous process

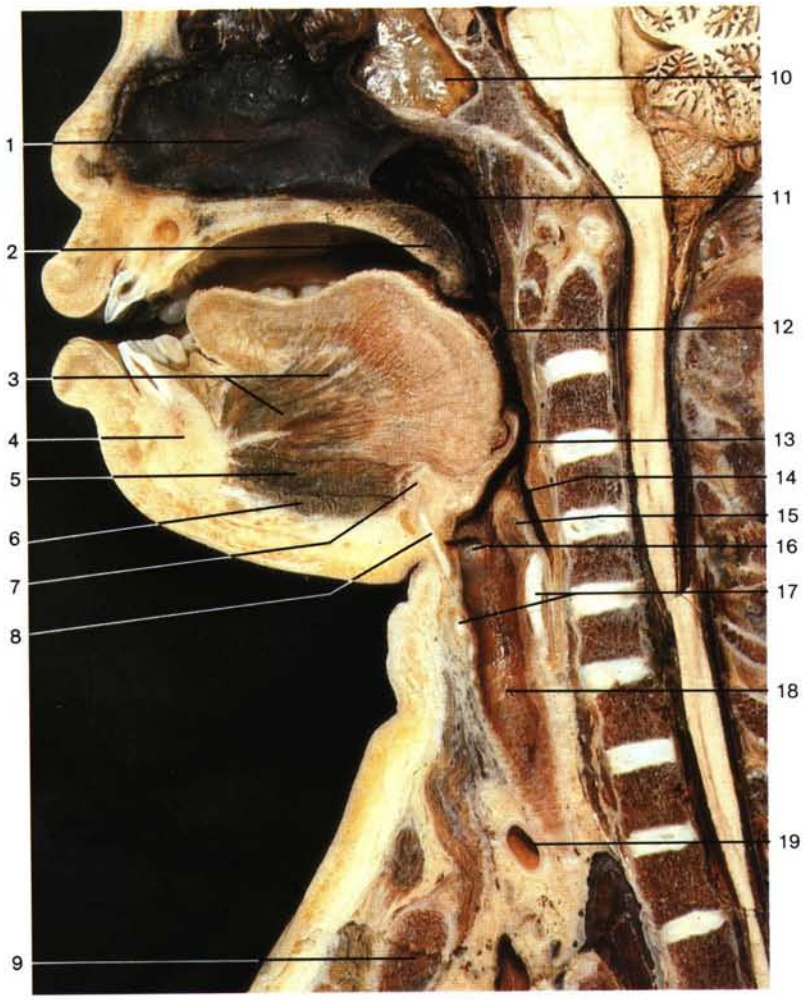

Median section through adult head and neck. Note the low position of the adult larynx when compared with that of the neonate (cf. with figure below).

1 Nasal septum
2 Uvula
3 Genioglossus muscle
4 Mandible
5 Geniohyoid muscle
6 Mylohyoid muscle
7 **Hyoid bone**
8 Thyroid cartilage
9 Manubrium sterni
10 Sphenoidal sinus
11 **Nasopharynx**
12 **Oropharynx**
13 **Epiglottis**
14 **Laryngopharynx**
15 Arytenoid muscle
16 **Vocal fold**
17 Cricoid cartilage
18 **Trachea**
19 Left brachiocephalic vein
20 Thymus
21 Esophagus
22 Occipital lobe
23 Cerebellum and 4th ventricle
24 Medulla oblongata
25 Dens of axis
26 Intervertebral discs of cervical vertebral column

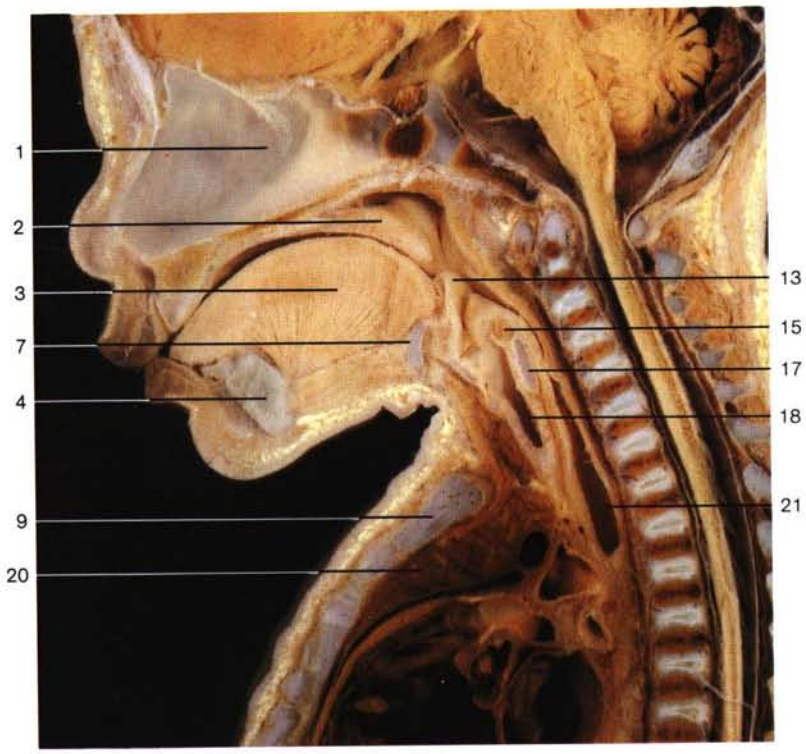

Median section through neonate head and neck. Note the high position of the larynx permitting the epiglottis nearly to reach the uvula (cf. with the figure above).

Sagittal section through the head. (MR-Scan.)

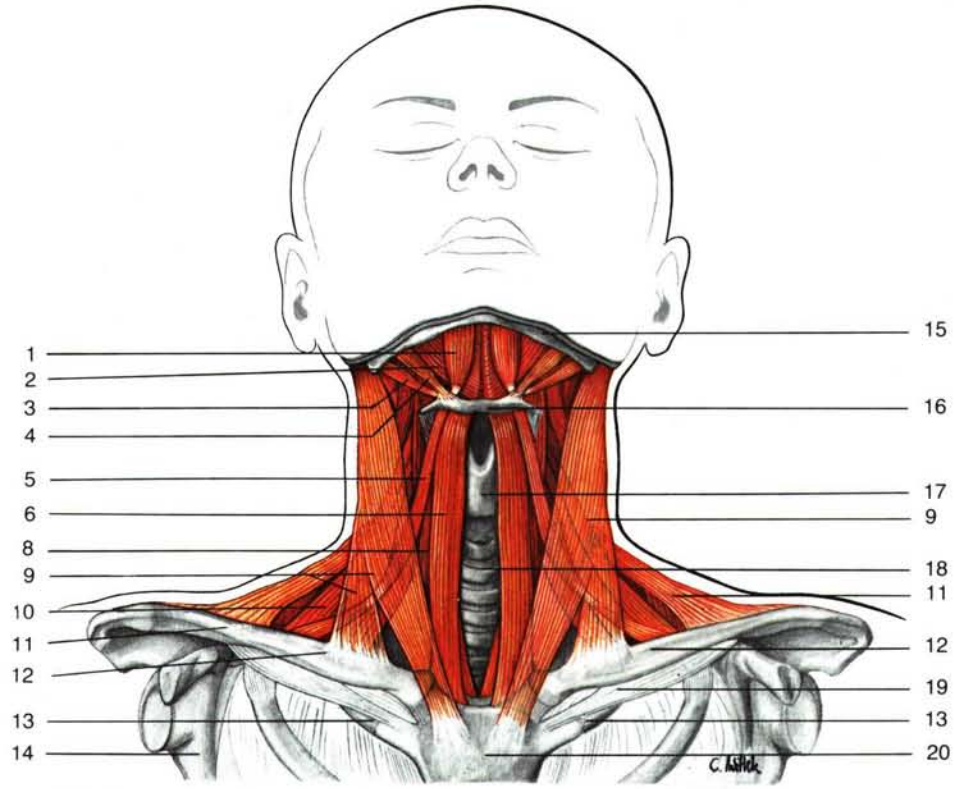

Muscles of the neck (anterior aspect).

Suprahyoid muscles

1 Anterior belly of digastric muscle
2 Mylohyoid muscle
3 Posterior belly of digastric muscle
4 Stylohyoid muscle

Infrahyoid muscles

5 Omohyoid muscle
6 Sternohyoid muscle
7 Thyrohyoid muscle
8 Sternothyroid muscle

Other structures

9 Sternocleidomastoid muscle
10 Scalenus muscles
11 Trapezius muscle
12 Clavicle
13 First rib
14 Scapula
15 Mandible
16 Hyoid bone
17 Larynx (thyroid cartilage)
18 Trachea
19 Subclavius muscle
20 Manubrium sterni
21 Mucous membrane of larynx (conus elasticus)
22 Cricoid cartilage
23 Inferior horn of thyroid cartilage
24 Esophagus
25 Body of cervical vertebra
26 Posterior root ganglion
27 Spinal cord
28 Spinous process
29 Internal jugular vein, common carotid artery, and vagus nerve
30 True muscles of the neck (semispinalis cervicis and capitis muscles)

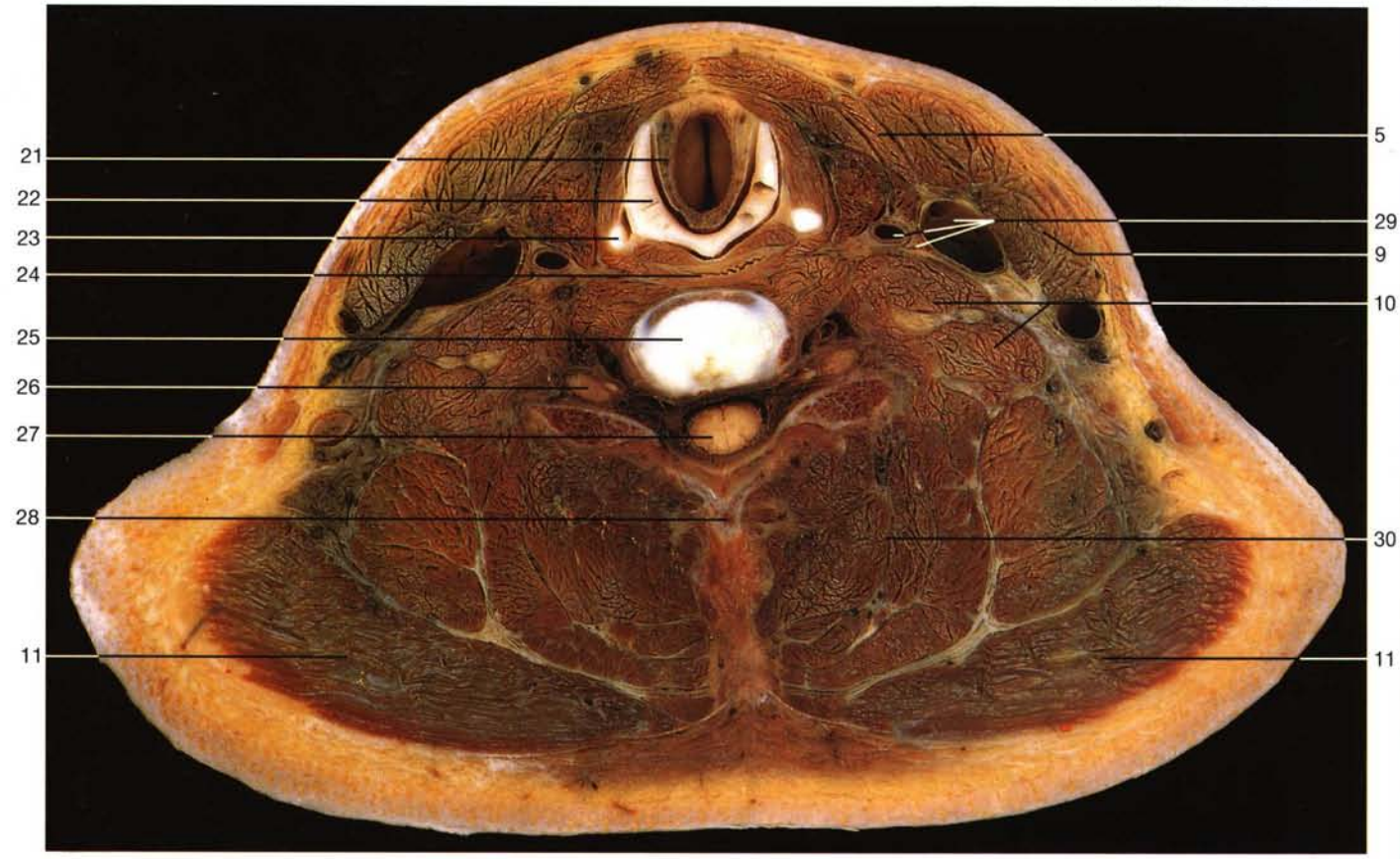

Cross-section of the neck at the level of the intervertebral disc between the 5th and 6th cervical vertebra (inferior aspect).

1 Mandible
2 Masseter muscle and facial artery
3 **Hyoid bone**
4 Median thyrohyoid ligament
5 **Thyrohyoid muscle**
6 **Sternothyroid muscle**
7 Thyroid gland (pyramidal lobe)
8 Pectoralis major muscle
9 Second rib
10 Parotid gland
11 Anterior belly of **digastric muscle**
12 Submandibular gland (divided)
13 **Mylohyoid muscle** and mylohyoid raphe
14 External carotid artery and vagus nerve
15 **Omohyoid muscle**
16 Thyroid cartilage
17 Sternocleidomastoid muscle
18 **Sternohyoid muscle**
19 Clavicle
20 Subclavius
21 Jugular fossa or suprasternal notch

Muscles of the neck (anterior aspect). Sternocleidomastoid and sternohyoid muscles on the right have been divided and reflected.

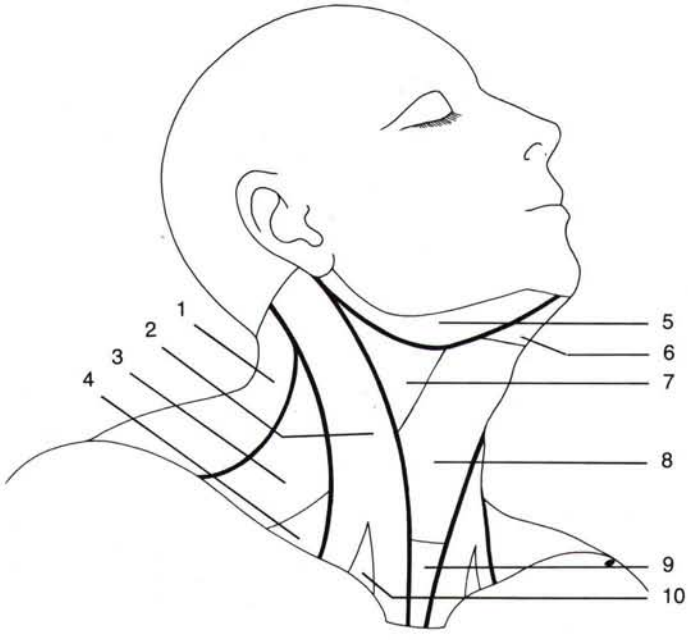

Regions and triangles of the neck.

1 Trapezius
2 Sternocleidomastoid
3 Occipital triangle } **Posterior triangle**
4 Supraclavicular triangle } **Posterior triangle**
5 Submandibular triangle } **Anterior triangle**
6 Submental triangle } **Anterior triangle**
7 Carotid triangle } **Anterior triangle**
8 Muscular triangle
9 Jugular fossa
10 Lesser supraclavicular fossa

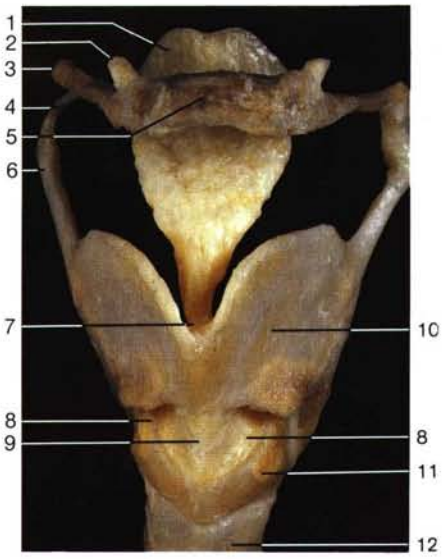

Cartilages of the larynx and the hyoid bone (anterior aspect).

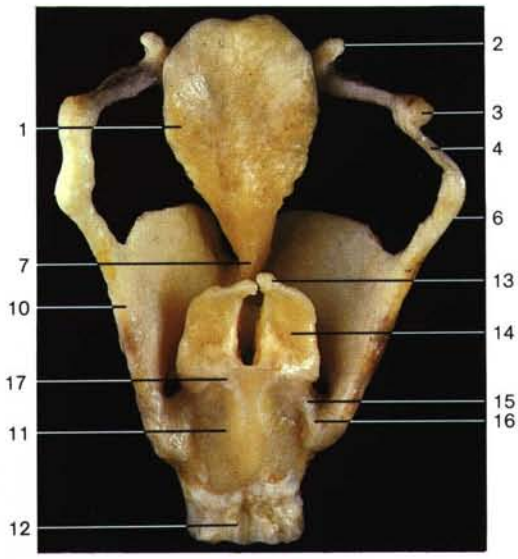

Cartilages of the larynx and the hyoid bone (posterior aspect).

1 **Epiglottis**
2 Lesser cornu of hyoid bone
3 Greater cornu of hyoid bone
4 Lateral thyrohyoid ligament
5 Body of **hyoid bone**
6 Superior cornu of thyroid cartilage
7 Thyroepiglottic ligament
8 Conus elasticus
9 Cricothyroid ligament
10 **Thyroid cartilage**
11 **Cricoid cartilage**
12 Trachea
13 Corniculate cartilage
14 **Arytenoid cartilage**
15 Posterior cricoarytenoid ligament
16 **Cricothyroid joint**
17 **Cricoarytenoid joint**

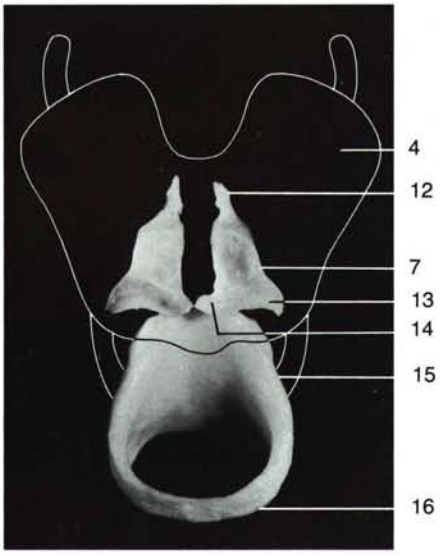

Cartilages of the larynx (anterior aspect). Thyroid cartilage is indicated by the outline.

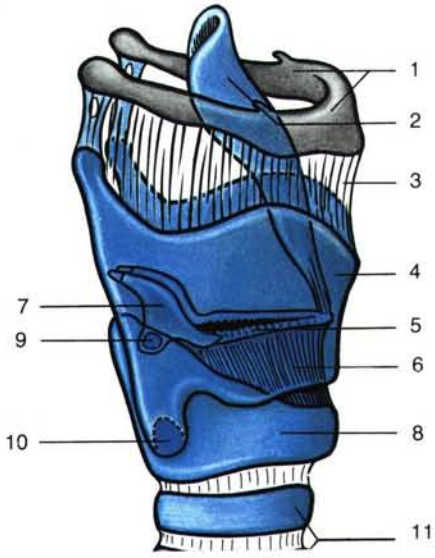

Cartilages and ligaments of the larynx (lateral aspect). (Schematic drawing.)

1 Hyoid bone
2 Epiglottis
3 Thyrohyoid membrane
4 **Thyroid cartilage**
5 **Vocal ligament**
6 Conus elasticus
7 Arytenoid cartilage
8 **Cricoid cartilage**
9 **Cricoarytenoid joint**
10 **Cricothyroid joint**
11 Tracheal cartilages
12 Corniculate cartilage
13 Muscular process of arytenoid cartilage
14 Vocal process of arytenoid cartilage
15 Lamina of cricoid cartilage
16 Arch of cricoid cartilage

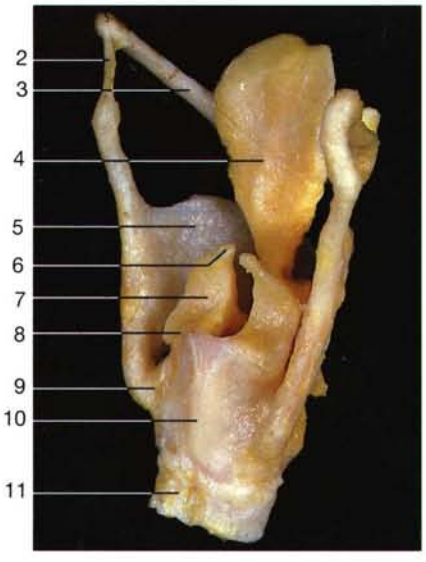

Cartilages of the larynx (oblique-posterior aspect).

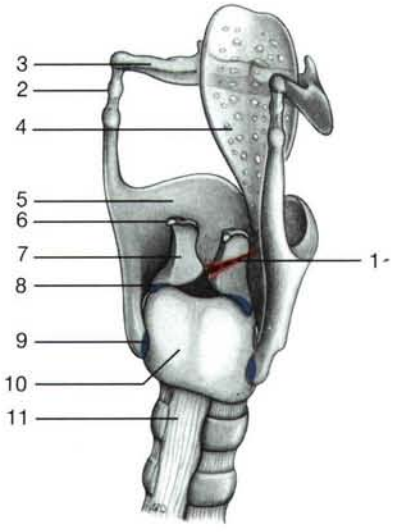

Cartilages of the larynx (oblique-posterior aspect).

1 **Vocal ligament** (red)
2 Lateral thyrohyoid ligament
3 Greater cornu of hyoid bone
4 **Epiglottis**
5 **Thyroid cartilage**
6 Corniculate cartilage
7 **Arytenoid cartilage**
8 **Cricoarytenoid joint**
9 **Cricothyroid joint**
10 **Cricoid cartilage**
11 Trachea

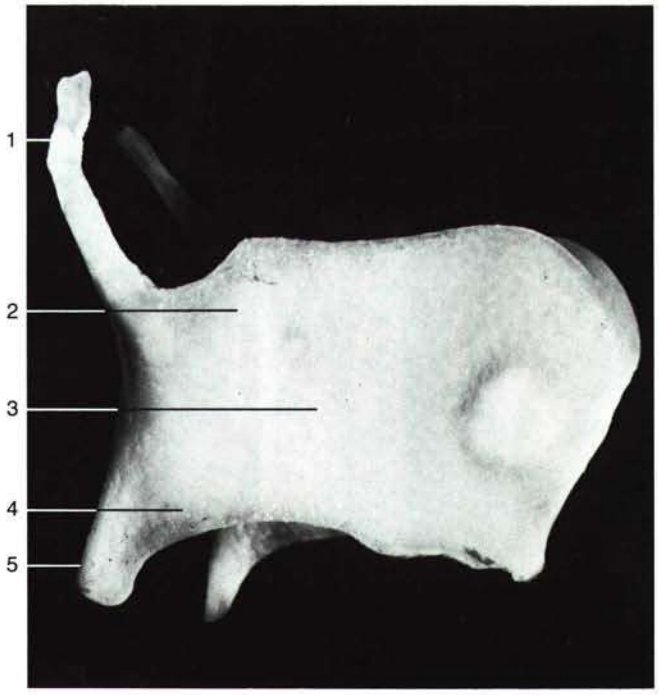

Thyroid cartilage (lateral aspect).

1 Superior cornu
2 Superior thyroid tubercle
3 Lamina of thyroid cartilage

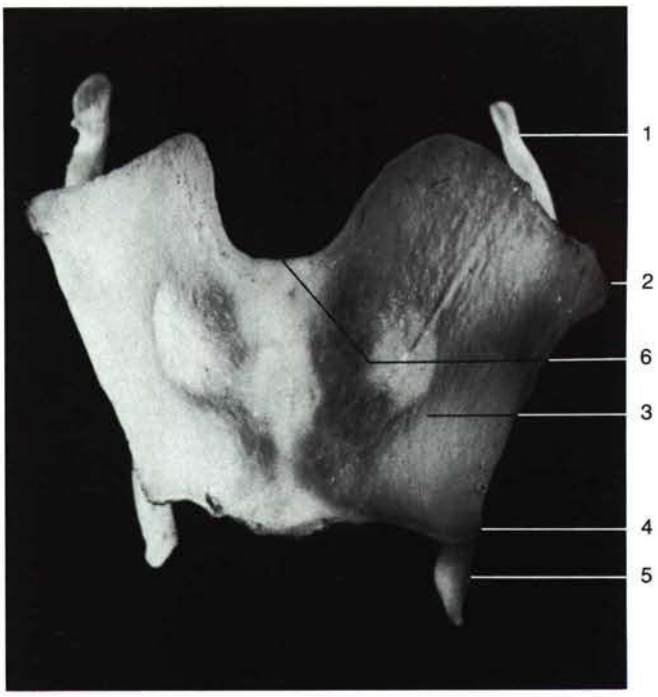

Thyroid cartilage (anterior aspect).

4 Inferior thyroid tubercle
5 Inferior cornu
6 Superior thyroid notch

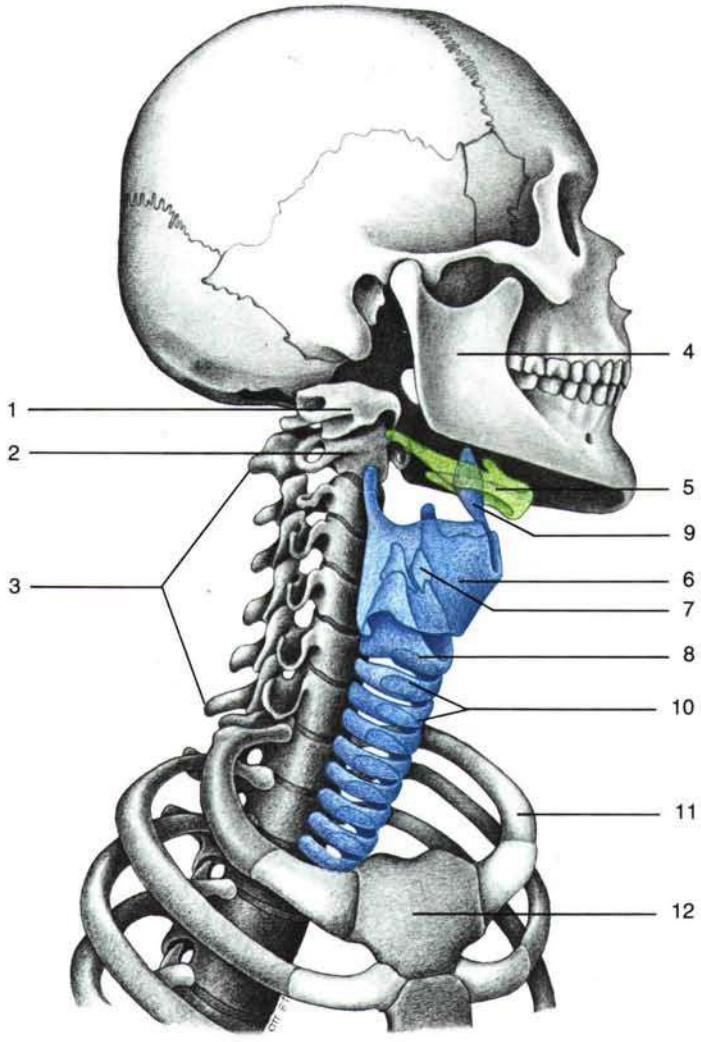

1 Atlas
2 Axis
3 Cervical vertebrae (C_2–C_7)
4 Mandible
5 **Hyoid bone**
6 **Thyroid cartilage**
7 Arytenoid cartilage
8 **Cricoid cartilage**
9 Epiglottis
10 **Tracheal cartilages**
11 First rib
12 Manubrium sterni

Position of the larynx in the neck (oblique lateral aspect). (Schematic drawing.)

Laryngeal muscles I (lateral aspect). Thyroid cartilage and thyroarytenoid muscle have been partly removed.

Laryngeal muscles II (lateral aspect). Half of the right side of the thyroid cartilage has been removed.

1 Epiglottis
2 Greater horn of hyoid bone
3 Corniculate cartilage
4 Transverse arytenoid muscle
5 Oblique arytenoid muscle
6 Lamina of cricoid cartilage
7 **Posterior cricoarytenoid muscle**
8 Aryepiglottic fold and muscle
9 Lamina of thyroid cartilage
10 Thyroepiglottic muscle
11 Thyroarytenoid muscle
12 **Lateral cricoarytenoid muscle**
13 Articular facet of thyroid cartilage
14 Cartilages of trachea
15 Membranous part of trachea
16 Lateral thyrohyoid ligament
17 Thyrohyoid membrane
18 Position of piriform recess
19 Arytenoid cartilage
20 Stem of epiglottis
21 Vestibular fold
22 Vocal folds and fissure of glottis
23 **Vocalis muscle**
24 Cricoarytenoid articulation
25 Arch of cricoid cartilage
26 Cricothyroid muscle
27 Cricoid cartilage
28 Piriform recess
29 Superior horn of thyroid cartilage
30 Vocal ligament

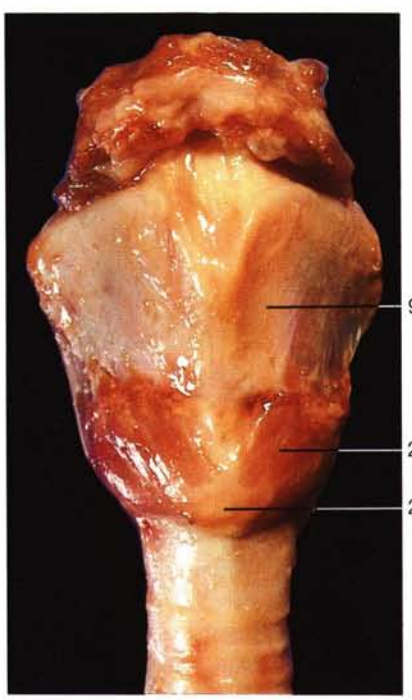

Laryngeal muscles, larynx (anterior aspect).

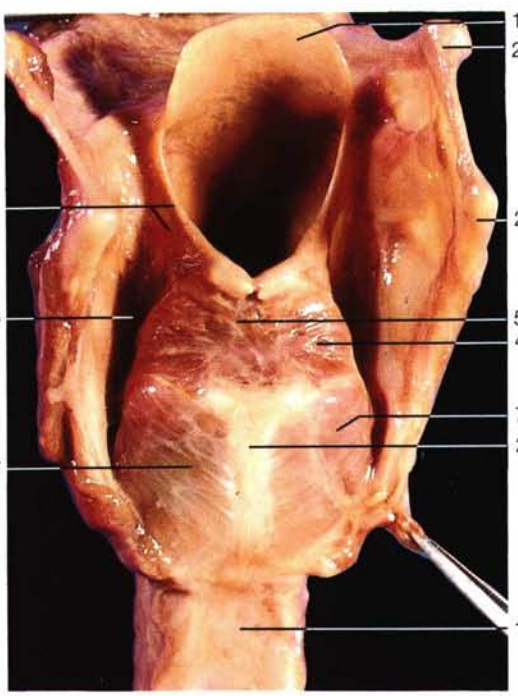

Laryngeal muscles, larynx (posterior aspect).

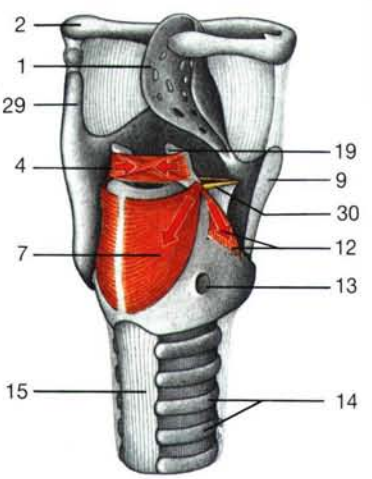

Action of internal muscles of the larynx. (Schematic drawing.)

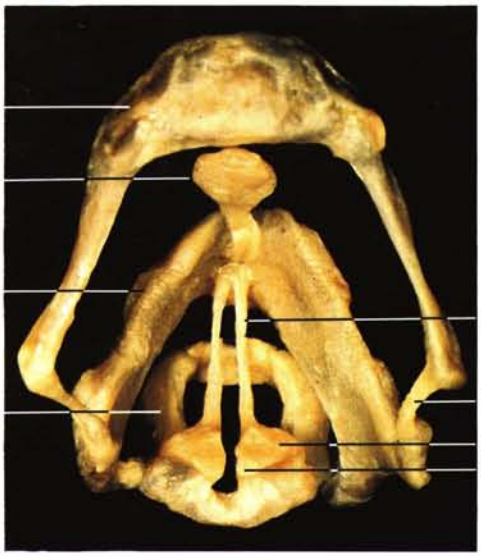

Laryngeal cartilages (superior aspect).

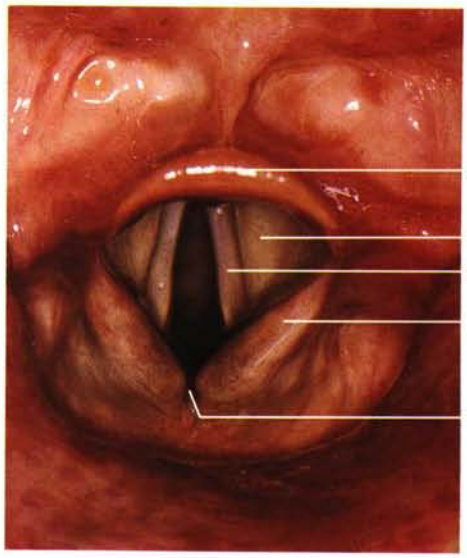

Glottis in vivo (superior aspect).

Horizontal section through the larynx at the level of the vocal folds (superior aspect).

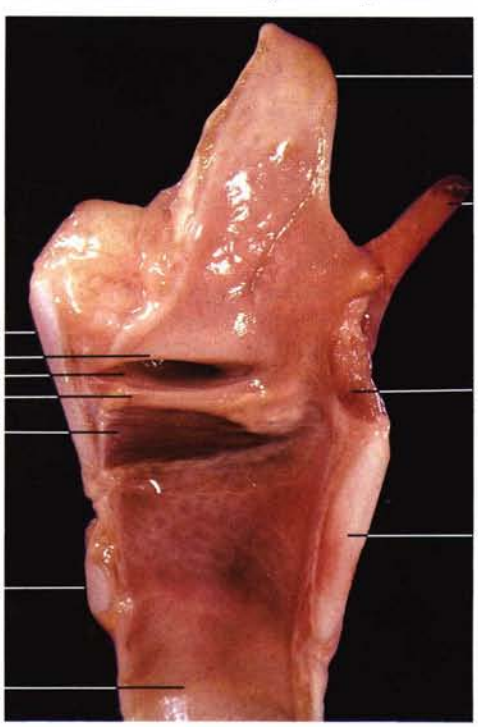

Sagittal section through the larynx.

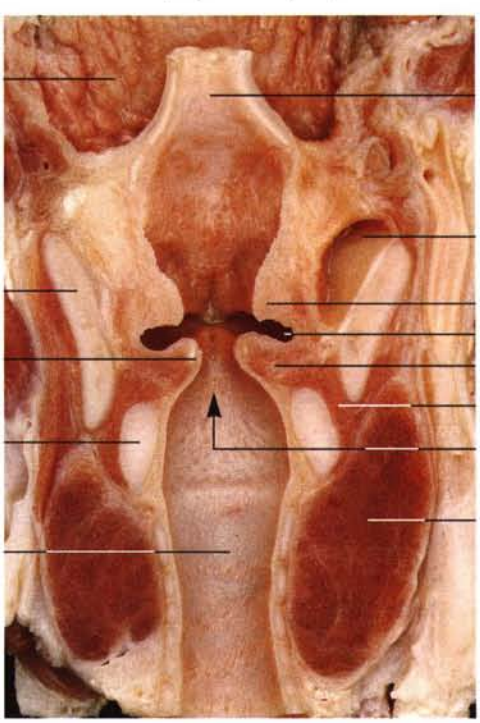

Coronal section through larynx and trachea.

1 **Hyoid bone**
2 Epiglottis
3 **Thyroid cartilage**
4 **Cricoid cartilage**
5 **Vocal ligament**
6 Thyrohyoid ligament
7 **Arytenoid cartilage**
8 Corniculate cartilage
9 **Vocal fold**
10 **Vestibular fold**
11 Aryepiglottic fold
12 Interarytenoid notch
13 Mandible
14 Anterior belly of digastric muscle
15 Mylohyoid muscle
16 Pyramidal process of thyroid gland
17 Sternohyoid and sternothyroid muscles
18 Common carotid artery
19 Internal jugular vein
20 **Rima glottidis**
21 Sternocleidomastoid muscle
22 Transverse arytenoid muscle
23 Pharynx and inferior constrictor muscle
24 Ventricle of larynx
25 **Vocalis muscle**
26 Trachea
27 Superior cornu of thyroid cartilage
28 Root of tongue (lingual tonsil)
29 Piriform recess
30 Vocalis muscle
31 Lateral cricoarytenoid muscle
32 **Thyroid gland**

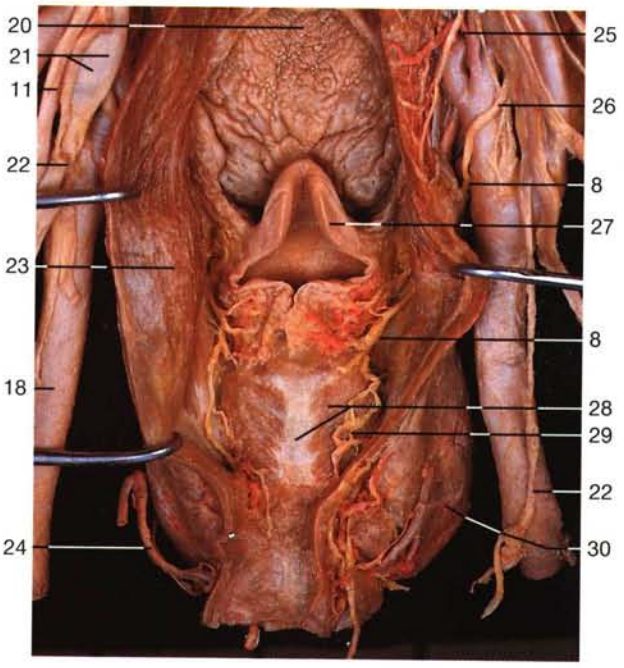

Larynx and its innervation (posterior aspect). Dissection of superior and inferior laryngeal nerves. Pharynx has been opened.

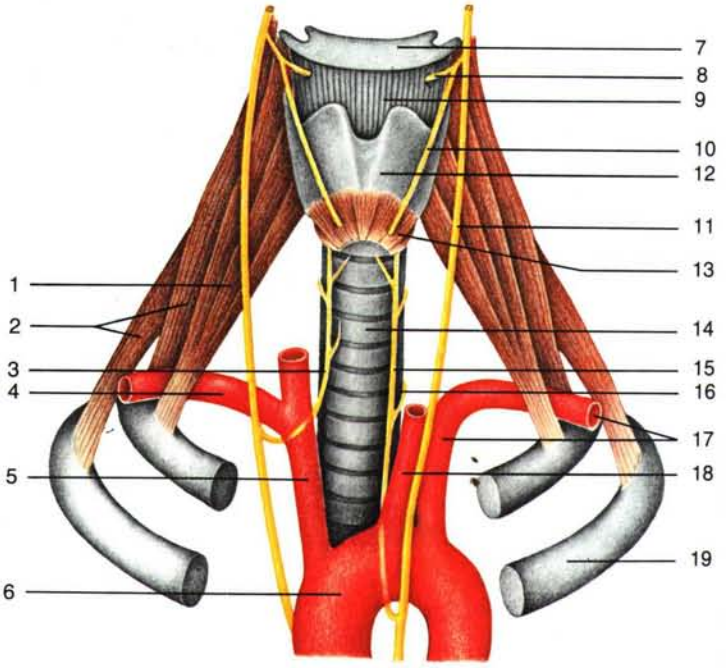

Innervation of the larynx. (Schematic diagram.)

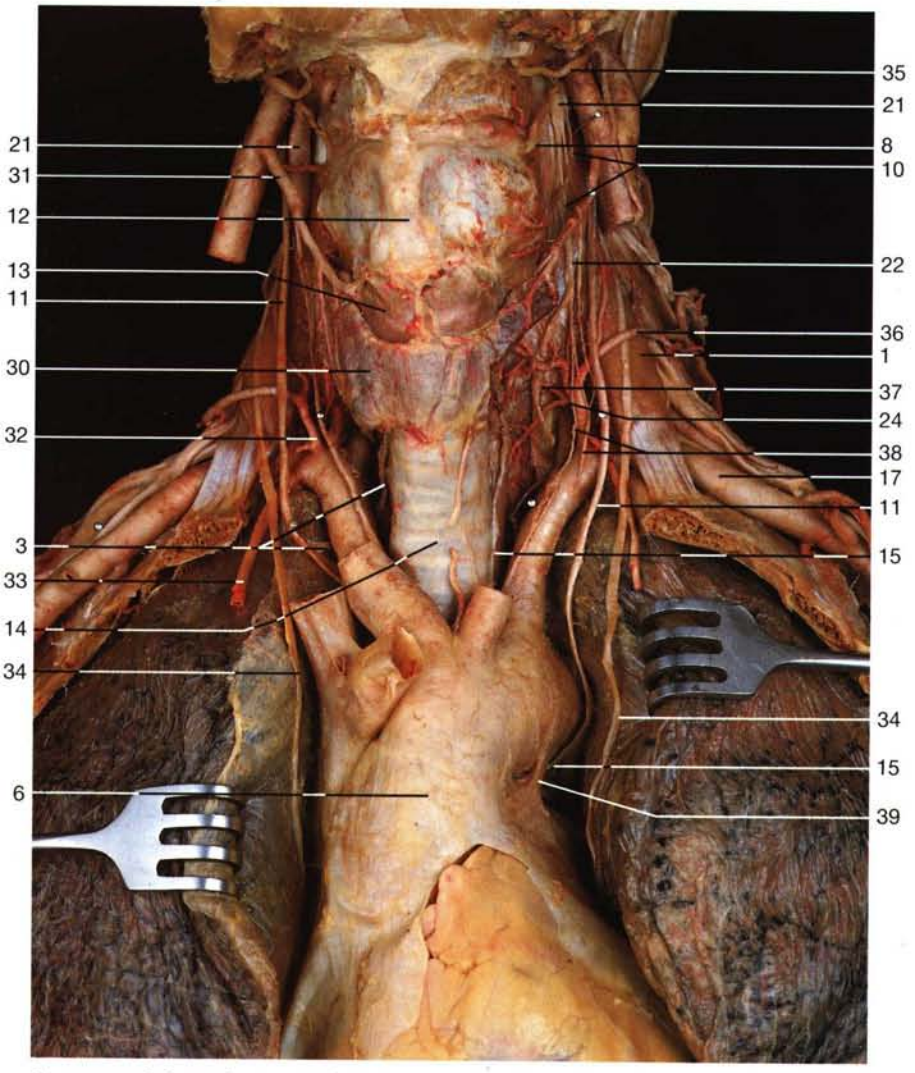

Larynx and thoracic organs (anterior aspect). Dissection of vagus and recurrent laryngeal nerves.

1 Scalenus anterior muscle
2 Scalenus medius and posterior muscles
3 Right **recurrent laryngeal nerve**
4 Right subclavian artery
5 Brachiocephalic trunk
6 Aortic arch
7 Hyoid bone
8 **Internal branch of superior laryngeal nerve**
9 Thyrohyoid membrane
10 **External branch of superior laryngeal nerve**
11 **Vagus nerve**
12 Thyroid cartilage
13 Cricothyroid muscle
14 Trachea
15 **Left recurrent larnygeal nerve**
16 Esophagus
17 Left subclavian artery
18 Left common carotid artery
19 Second rib
20 Tongue
21 Superior cervical ganglion
22 Sympathetic trunk
23 Inferior constrictor muscle of pharynx
24 **Inferior thyroid artery**
25 Glossopharyngeal nerve
26 Superior laryngeal nerve
27 Epiglottis
28 Posterior cricoarytenoid muscle and cricoid cartilage
29 Inferior laryngeal branch of recurrent laryngeal nerve
30 **Thyroid gland**
31 Superior thyroid artery
32 **Thyrocervical trunk**
33 Internal thoracic artery
34 **Phrenic nerve**
35 Hypoglossal nerve
36 Transverse cervical artery
37 Middle cervical ganglion
38 Middle cervical cardiac nerves (branches of sympathetic trunk)
39 Ligamentum arteriosum

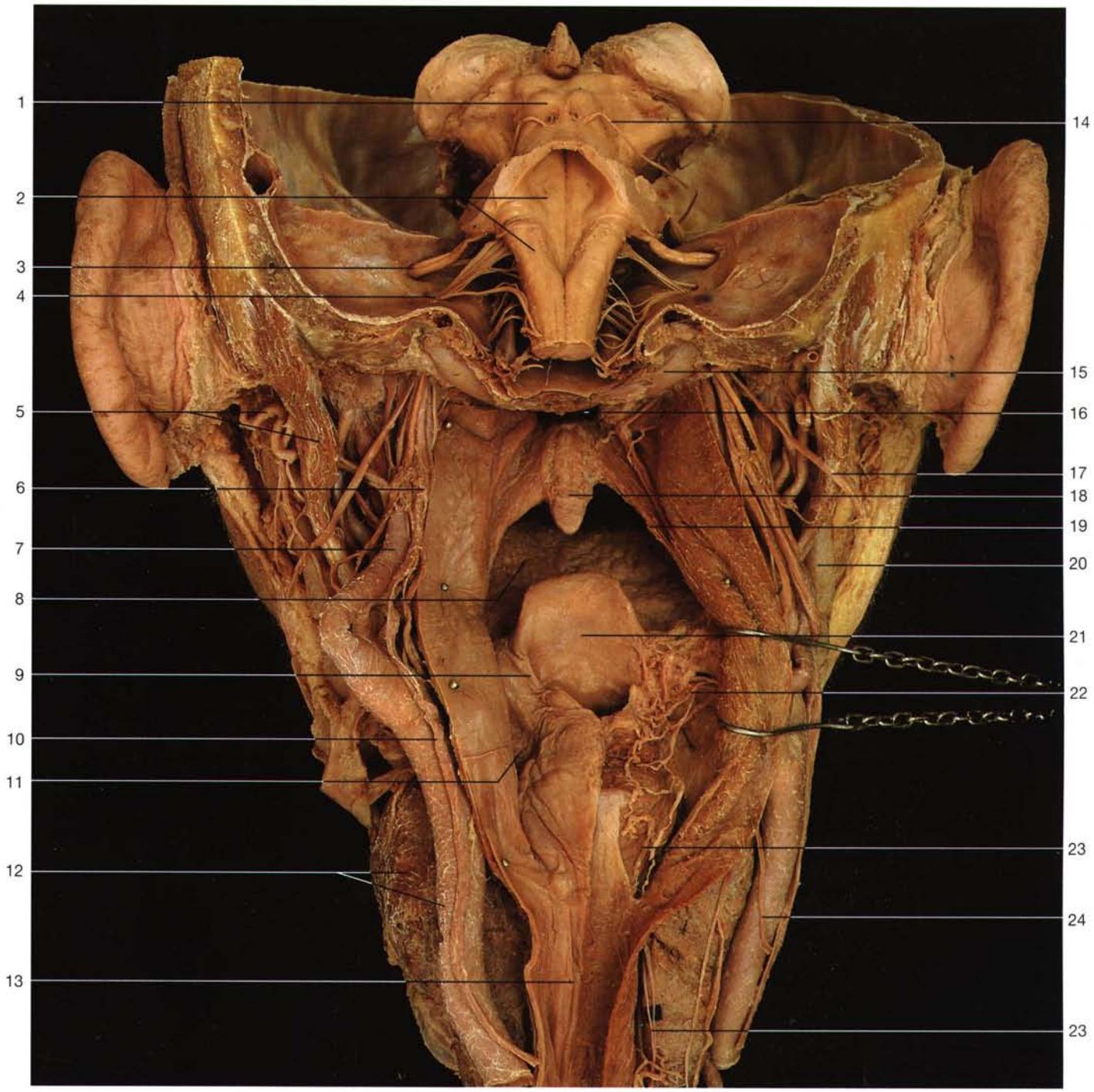

Larynx and oral cavity (posterior aspect). Mucous membrane on the right half of pharynx has been removed.

1 Midbrain (inferior colliculus)
2 Rhomboid fossa and medulla oblongata
3 Vestibulocochlear and facial nerve
4 Glossopharyngeal, vagus and accessory nerve
5 Occipital artery and posterior belly of digastric muscle
6 Superior cervical ganglion
7 **Internal carotid artery**
8 **Oral cavity** (tongue)
9 Aryepiglottic fold
10 Vagus nerve
11 **Piriform recess**
12 Thyroid gland and common carotid artery
13 Esophagus
14 Trochlear nerve
15 Occipital condyle
16 **Nasal cavity** (choana)
17 Accessory nerve
18 **Uvula** and soft palate
19 Palatopharyngeus muscle
20 External carotid artery
21 Epiglottis
22 **Internal branch of superior laryngeal nerve**
23 **Inferior laryngeal nerve**
24 Ansa cervicalis

Pharynx and parapharyngeal nerves in connection with brain stem (posterior aspect).

1 Inferior colliculus of midbrain
2 Facial colliculus in **floor of rhomboid fossa**
3 Vestibulocochlear and facial nerves
4 Glossopharyngeal nerve
5 Vagus nerve
6 Accessory nerve
7 Hypoglossal nerve
8 Pharyngobasilar fascia
9 Superior constrictor muscle of pharynx
10 **Sympathetic trunk** and **superior cervical ganglion** (medially displaced)
11 Middle constrictor muscle of pharynx
12 Greater cornu of hyoid bone
13 Inferior constrictor muscle of pharynx
14 Trochlear nerve
15 **Internal acoustic meatus** with facial and vestibulocochlear nerves
16 **Jugular foramen** with glossopharyngeal, vagus and assessory nerves
17 Occipital condyle
18 Occipital artery
19 Posterior belly of digastric muscle
20 **Accessory nerve** (extracranial part)
21 **Hypoglossal nerve** (extracranial part)
22 External carotid artery
23 Carotid sinus nerve
24 **Internal carotid artery**
25 **Carotid sinus and carotid body**
26 Vagus nerve
27 Thyroid gland
28 Esophagus
29 Choanae
30 Medial pterygoid plate
31 Foramen lacerum
32 **Pharyngeal tubercle**
33 Hard palate
34 Greater and lesser palatine foramen
35 Pterygoid hamulus
36 Lateral pterygoid plate
37 Pterygoid canal
38 Foramen ovale
39 Mandibular fossa
40 Carotid canal
41 Styloid process and stylomastoid foramen

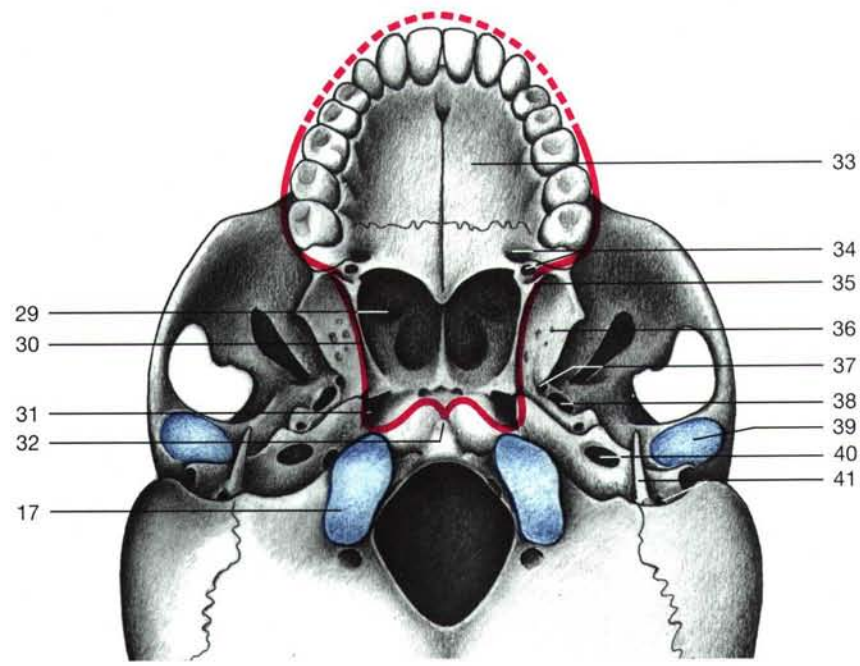

Inferior aspect of the skull. Red line = outline of superior constrictor muscle in continuation with buccinator muscle and orbicularis oris muscle. (Semischematic drawing.)

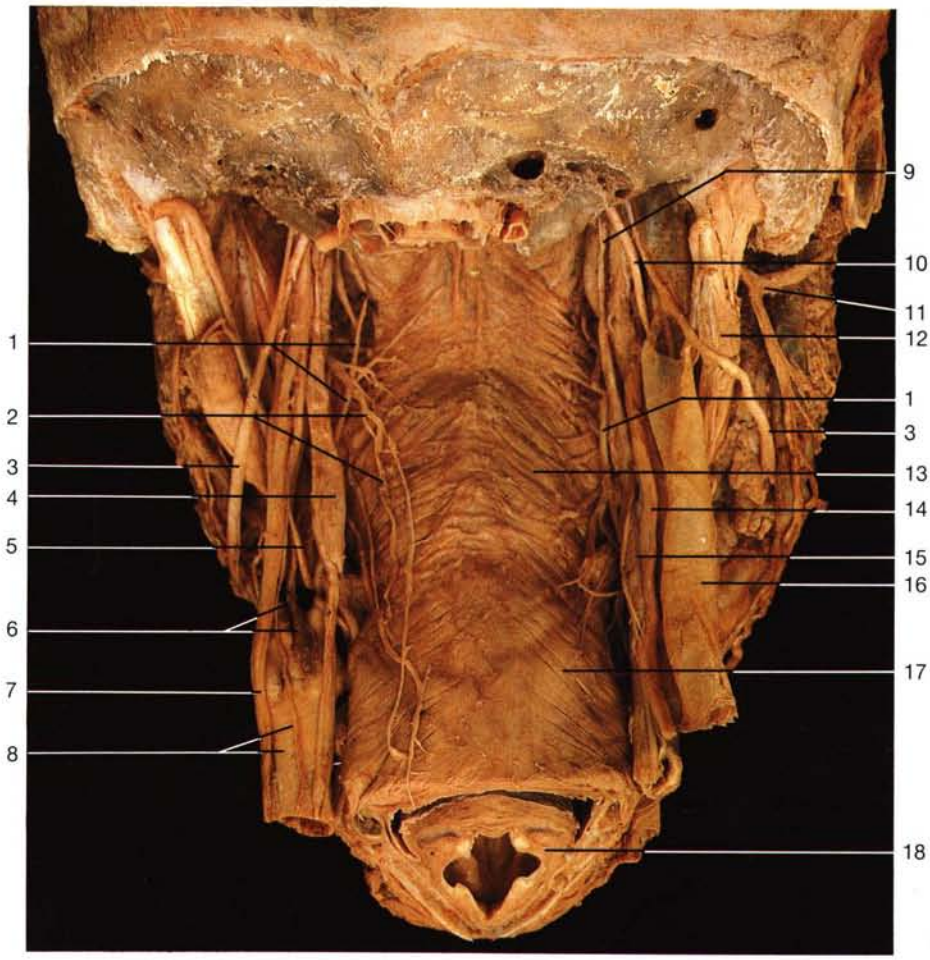

Parapharyngeal nerves and vessels. Dorsal aspect of the pharynx.

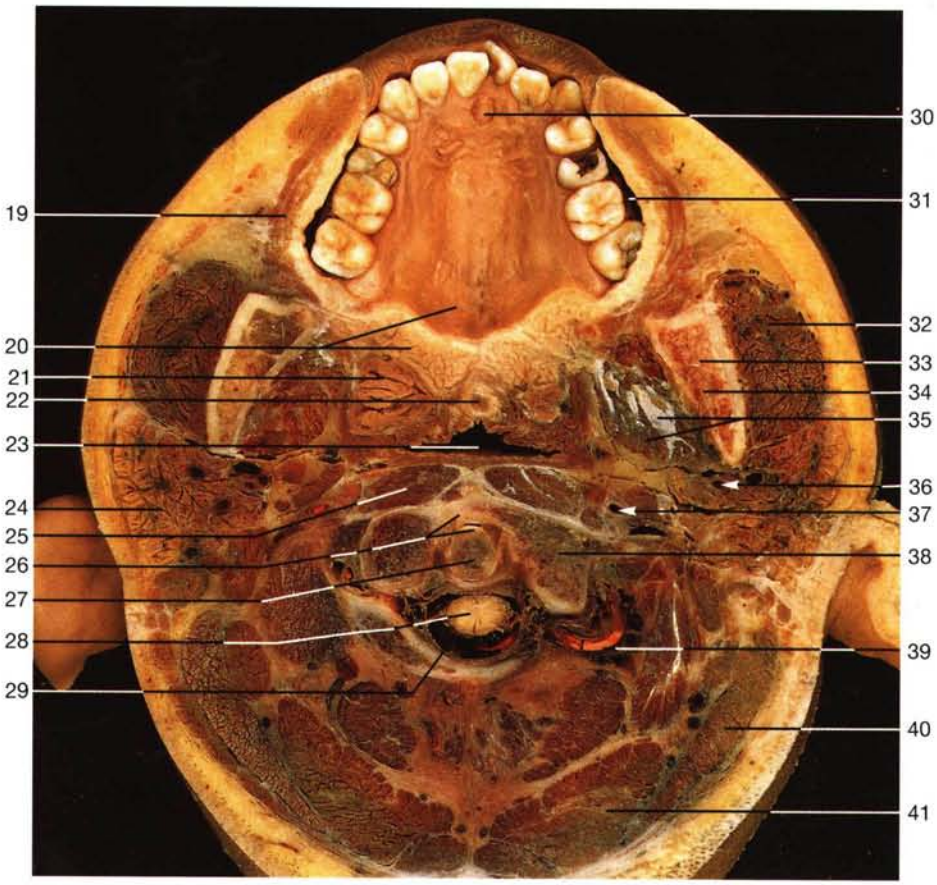

Cross-section of head and neck at the level of the atlas (inferior aspect).

1 Ascending pharyngeal artery
2 Pharyngeal plexus
3 **Accessory nerve**
4 Superior cervical ganglion of sympathetic trunk
5 Superior laryngeal nerve
6 Carotid body and carotid sinus nerve
7 Left **vagus nerve**
8 Common carotid artery and cardiac branch of vagus nerve
9 **Glossopharyngeal nerve**
10 **Hypoglossal nerve**
11 **Facial nerve**
12 Posterior belly of digastric muscle
13 Middle constrictor muscle of pharynx
14 Right vagus nerve
15 **Sympathetic trunk**
16 **Internal jugular vein**
17 Inferior constrictor muscle of pharynx
18 Larynx
19 **Buccinator muscle**
20 Soft palate and palatine glands
21 Palatine tonsil
22 Uvula of palate
23 **Pharynx** (oral part)
24 Parotid gland
25 Longus capitis muscle
26 Medioatlantoaxial joint and anterior arch of atlas
27 **Dens of axis**
28 **Spinal cord**
29 Dura mater
30 Incisive papilla
31 **Oral vestibule**
32 Masseter muscle
33 Mandible
34 Mandibular canal with vessels and nerve
35 Medial pterygoid muscle
36 External carotid artery
37 Internal carotid artery
38 **Atlas**
39 **Vertebral artery**
40 Splenius capitis muscle
41 Semispinalis capitis muscle

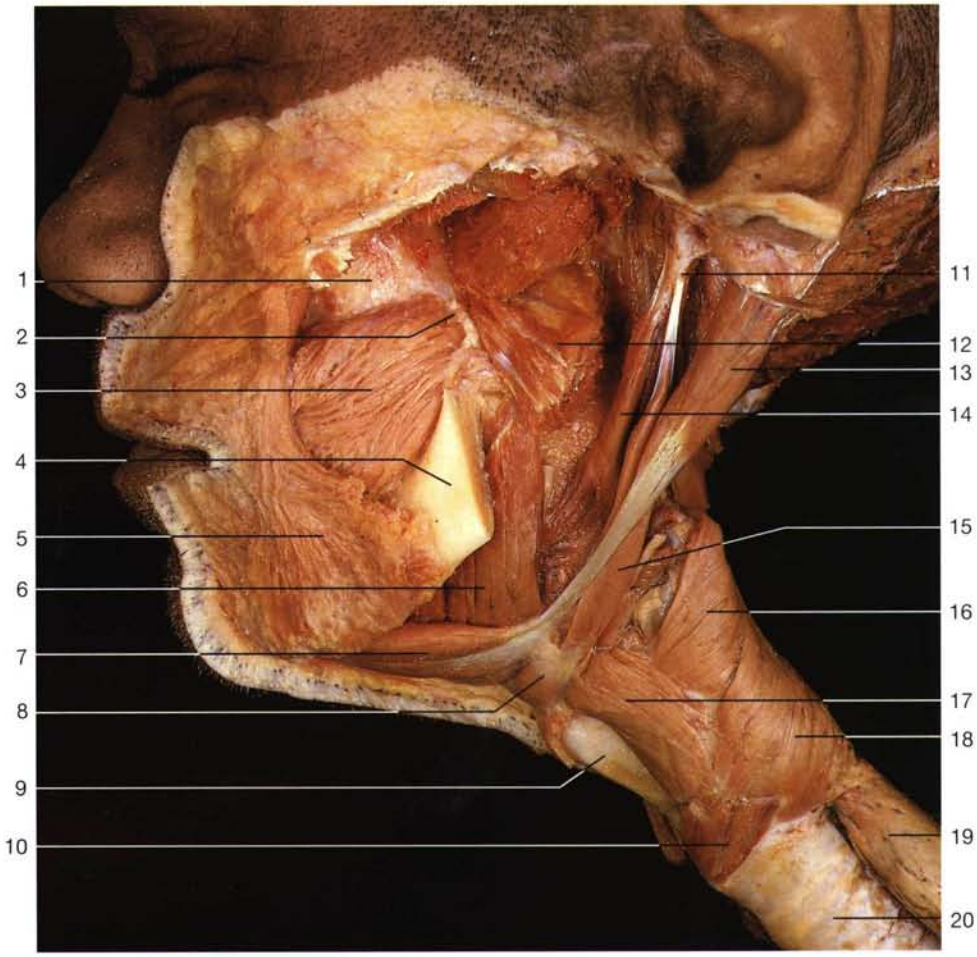

Dissection of pharynx, supra- and infrahyoid muscles I. Mandible partly removed (lateral aspect).

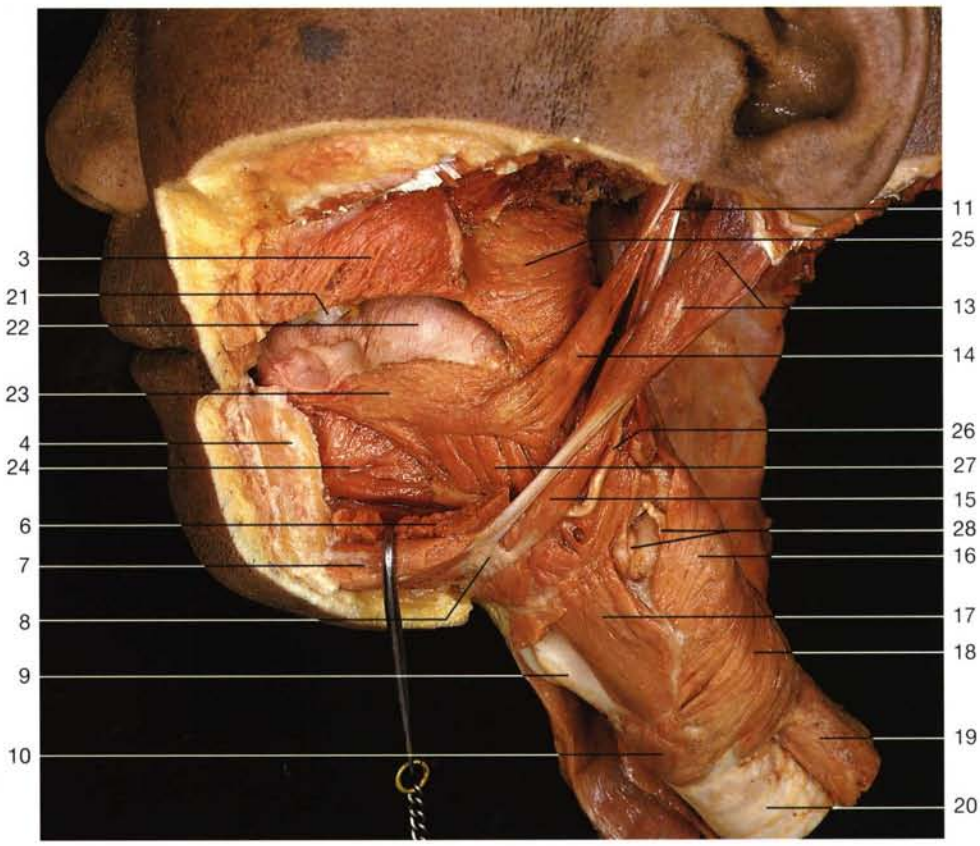

Dissection of pharynx, supra- and infrahyoid muscles II. Oral cavity opened (lateral aspect).

1 Maxilla
2 Pterygomandibular raphe
3 Buccinator muscle
4 Mandible (divided)
5 Depressor anguli oris muscle
6 Mylohyoid muscle
7 Anterior belly of digastric muscle
8 Hyoid bone
9 Thyroid cartilage
10 Cricothyroid muscle
11 Styloid process
12 Medial pterygoid muscle (divided)
13 Posterior belly of digastric muscle
14 Styloglossus muscle
15 Stylohyoid muscle
16 **Thyropharyngeal part of inferior constrictor muscle of pharynx**
17 Thyrohyoid muscle
18 **Cricopharyngeal part of inferior constrictor muscle of pharynx**
19 Esophagus
20 Trachea
21 First molar of maxilla
22 Tongue
23 Inferior longitudinal muscle of tongue
24 Genioglossus muscle
25 **Superior constrictor muscle of pharynx**
26 Hypoglossal nerve
27 Hyoglossus muscle
28 Superior laryngeal nerve and superior laryngeal artery

Muscles of the pharynx (posterior aspect).

1 Sella turcica
2 Internal acoustic meatus and petrous part of temporal bone
3 Pharyngobasilar fascia
4 Fibrous raphe of pharynx
5 **Stylopharyngeus muscle**
6 **Superior constrictor muscle of pharynx**
7 Posterior belly of digastric muscle
8 Stylohyoid muscle
9 **Middle constrictor muscle of pharynx**
10 **Inferior constrictor muscle of pharynx**
11 Muscle-free area (Killian's triangle)
12 Esophagus
13 Trachea
14 Thyroid and parathyroid glands
15 Medial pterygoid muscle
16 Greater horn of hyoid bone
17 Internal jugular vein
18 Parotid gland
19 Accessory nerve
20 **Superior cervical ganglion** of sympathetic trunk
21 Vagus nerve
22 Laimer's triangle (area prone to developing diverticula)
23 Orbicularis oculi muscle
24 Nasalis muscle
25 Levator labii superioris and levator labii alaeque nasi muscles
26 Levator anguli oris muscle
27 Orbicularis oris muscle
28 **Buccinator muscle**
29 Depressor labii inferioris muscle
30 **Hyoglossus muscle**
31 Thyrohyoid muscle
32 Thyroid cartilage
33 Cricothyroid muscle
34 Pterygomandibular raphe
35 Tensor veli palatini muscle
36 Levator veli palatini muscle
37 Depressor anguli oris muscle
38 Mentalis muscle
39 Styloglossus muscle

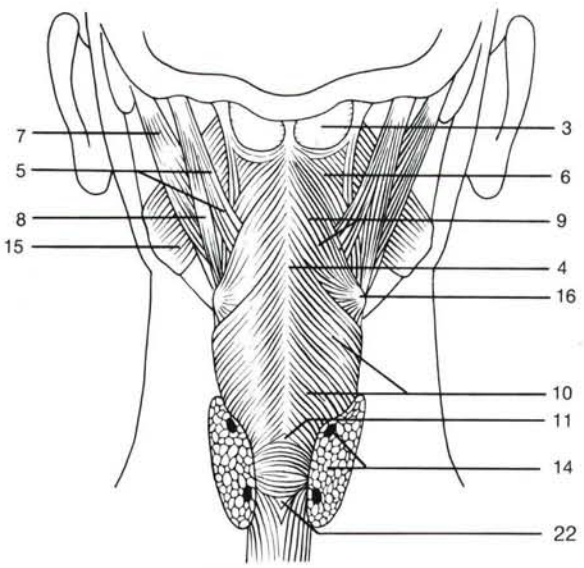

Muscles of the pharynx. (Schematic drawing.)

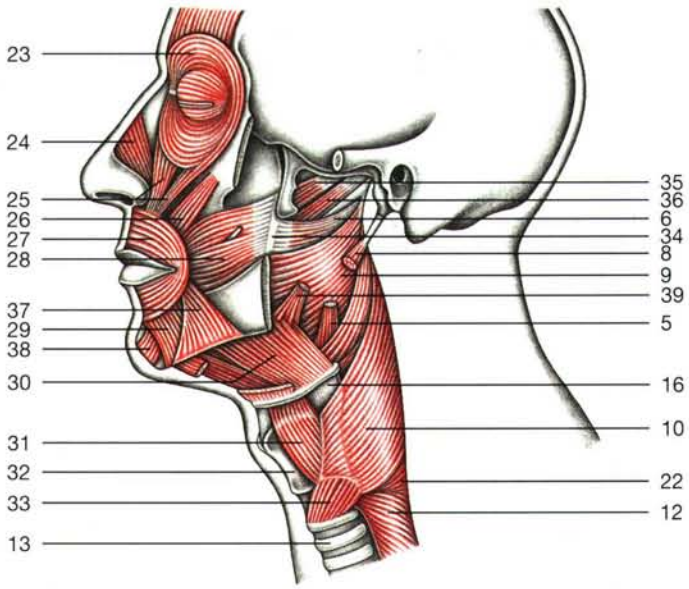

Muscles of pharynx (lateral aspect). (Schematic drawing.)

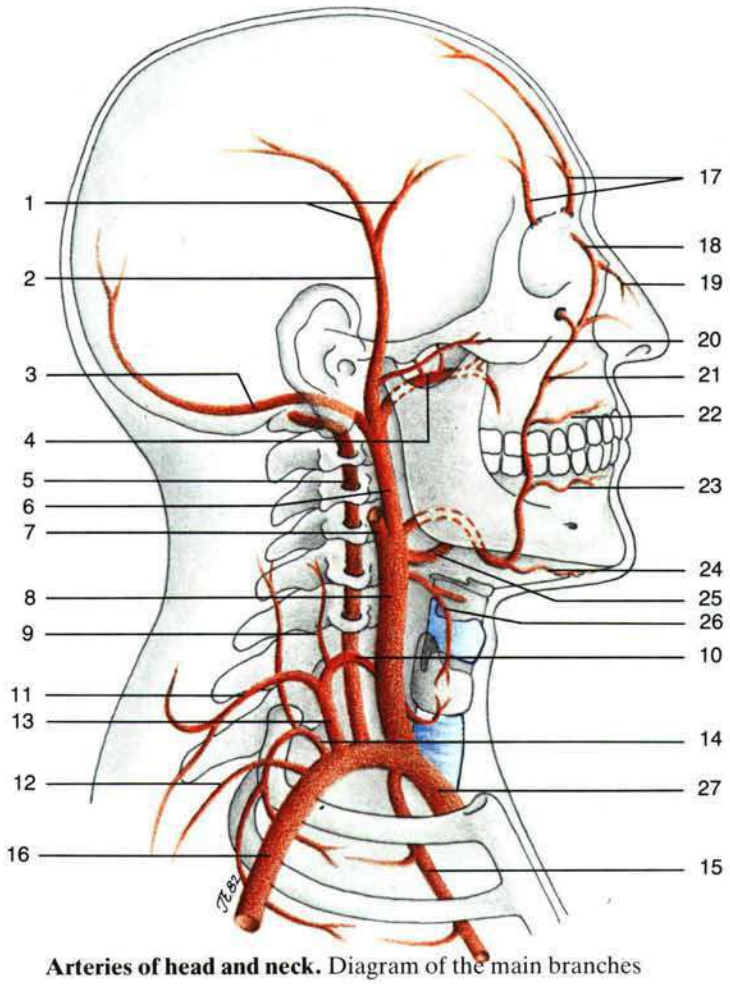

Arteries of head and neck. Diagram of the main branches of external carotid and subclavian artery.

1 Frontal and parietal branches of superficial temporal artery
2 Superficial temporal artery
3 **Occipital artery**
4 **Maxillary artery**
5 **Vertebral artery**
6 **External carotid artery**
7 Internal carotid artery
8 Common carotid artery (divided)
9 Ascending cervical artery
10 Inferior thyroid artery
11 Transverse cervical artery with 2 branches (superficial cervical artery and descending scapular artery)
12 Suprascapular artery
13 **Thyrocervical trunk**
14 Costocervical trunk with 2 branches (deep cervical artery and superior intercostal artery)
15 Internal thoracic artery
16 **Axillary artery**
17 Supraorbital and supratrochlear arteries
18 Angular artery
19 Dorsal nasal artery
20 Transverse facial artery
21 **Facial artery**
22 Superior labial artery
23 Inferior labial artery
24 Submental artery
25 Lingual artery
26 Superior thyroid artery
27 Brachiocephalic trunk

to page 165 ▷

1 Galea aponeurotica
2 Frontal branch } of superficial
3 Parietal branch } temporal artery
4 Superior auricular muscle
5 **Superficial temporal artery and vein**
6 Middle temporal artery
7 Auriculotemporal nerve
8 Branches of facial nerve
9 Facial nerve
10 **External carotid artery** within the retromandibular fossa
11 Posterior belly of digastric muscle
12 Sternocleidomastoid artery
13 Sympathetic trunk and superior cervical ganglion
14 Sternocleidomastoid muscle (divided and reflected)
15 Clavicle (divided)
16 Transverse cervical artery
17 Ascending cervical artery and phrenic nerve
18 Scalenus anterior muscle
19 Suprascapular artery
20 Dorsal scapular artery
21 Brachial plexus and **axillary artery**
22 Thoracoacromial artery
23 Lateral thoracic artery
24 Median nerve (displaced) and pectoralis minor muscle (reflected)
25 Frontal belly of occipitofrontalis muscle
26 Orbital part of orbicularis oculi muscle
27 **Angular artery** and vein
28 **Facial artery**
29 Superior labial artery
30 Zygomaticus major muscle
31 Inferior labial artery
32 Parotid duct
33 Buccal fat pad
34 **Maxillary artery**
35 Masseter muscle
36 Facial artery and mandible
37 Submental artery
38 Anterior belly of digastric muscle

Main branches of head and neck arteries (lateral aspect). Anterior thoracic wall and clavicle partly removed; pectoralis muscles have been reflected to display the subclavian and axillary arteries.

39 Hyoid bone
40 **Internal carotid artery**
41 **External carotid artery**
42 Superior laryngeal artery
43 Superior thyroid artery
44 **Common carotid artery**
45 Thyroid ansa of sympathetic trunk and **inferior thyroid artery**
46 Thyroid gland (right lobe)
47 **Vertebral artery**
48 **Thyrocervical trunk**
49 Vagus nerve
50 Ansa subclavia of sympathetic trunk
51 **Brachiocephalic trunk**
52 Superior vena cava (divided)
53 Aortic arch

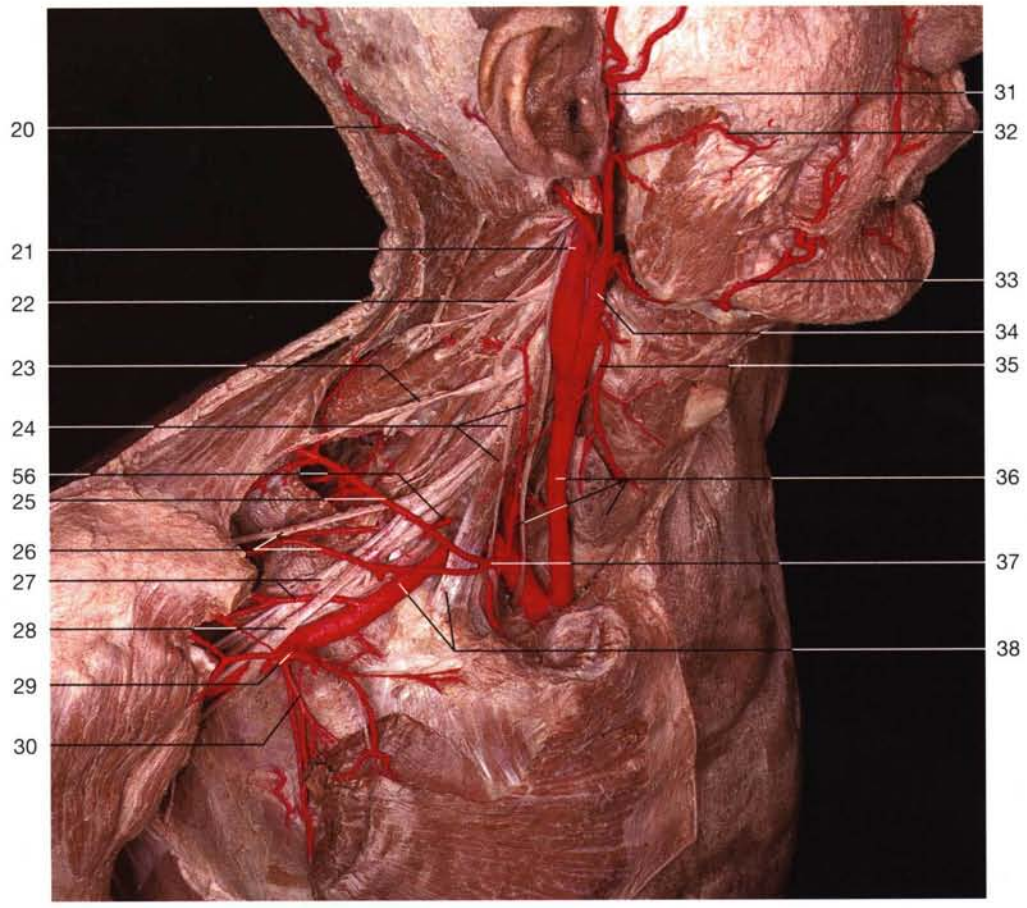

Arteries of head and neck (anterior-lateral aspect). Clavicle, sternocleidomastoid and veins have been partly removed, the arteries were colored.

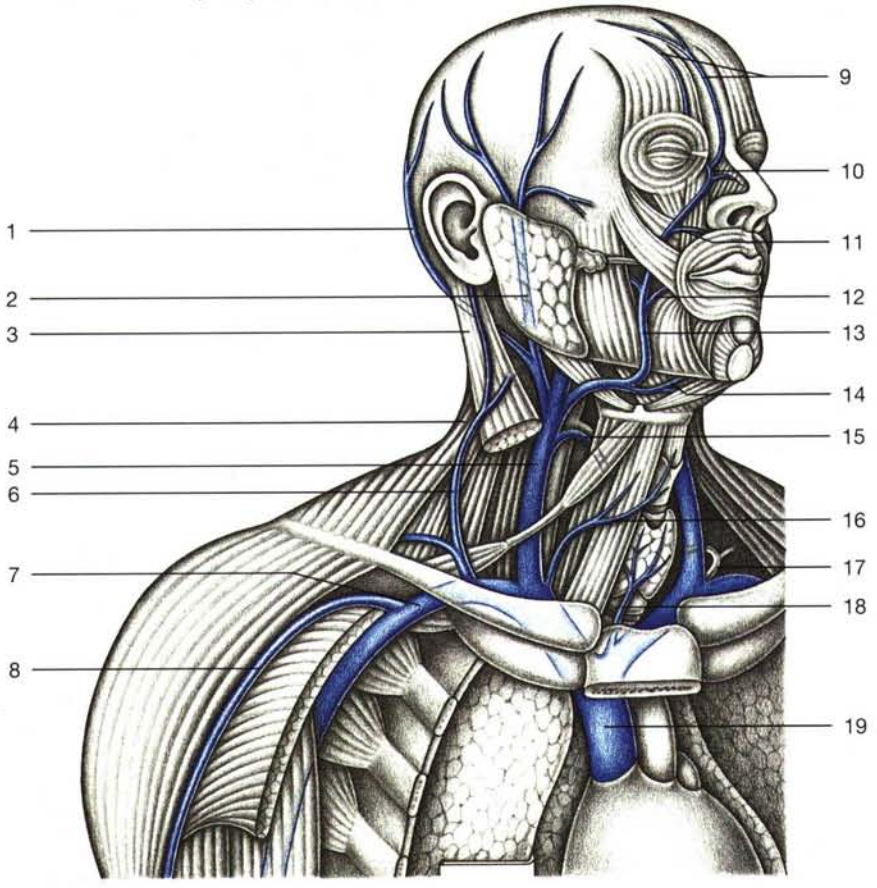

Veins of head and neck. Sternocleidomastoid muscle and anterior thoracic wall partly removed. Note the venous connection with the superior vena cava.

1 Occipital vein
2 Superficial temporal vein
3 Sternocleidomastoid muscle
4 Trapezius muscle
5 **Internal jugular vein**
6 External jugular vein
7 **Subclavian vein**
8 Cephalic vein
9 Supraorbital veins
10 Angular vein
11 Superior labial vein
12 Inferior labial vein
13 **Facial vein**
14 Submental vein
15 Superior thyroid vein
16 Anterior jugular vein
17 Thoracic duct
18 Inferior thyroid vein
19 Superior vena cava
20 Occipital artery
21 **Internal carotid artery**
22 Cervical plexus
23 Supraclavicular nerve
24 Phrenic nerve and ascending cervical artery on scalenus anterior muscle
25 Superficial cervical artery
26 Suprascapular artery and nerve
27 Brachial plexus and anterior circumflex humeral artery
28 Lateral cord of brachial plexus
29 **Thoracoacromial artery**
30 Lateral thoracic artery
31 Superficial temporal artery
32 Transverse facial artery
33 Facial artery
34 External carotid artery
35 Superior thyroid artery
36 **Common carotid artery, vagus nerve** and thyroid gland
37 **Thyrocervical trunk**
38 **Subclavian artery** and scalenus anterior muscle
39 Parotid gland and facial nerve
40 Great auricular nerve
41 External jugular vein
42 Brachial plexus
43 Cephalic vein in deltopectoral groove
44 Axillary vein and artery
45 Right brachiocephalic vein
46 **Superior vena cava**
47 Right lung (reflected)
48 Superficial temporal artery and vein
49 Facial artery and vein
50 Cervical branch of facial nerve and submandibular gland
51 **Internal jugular vein,** common carotid artery and omohyoid muscle
52 Anterior jugular vein and thyroid gland
53 Jugular venous arch
54 Left brachiocephalic vein
55 Pericardium of heart (location of right atrium)
56 Transverse cervical artery

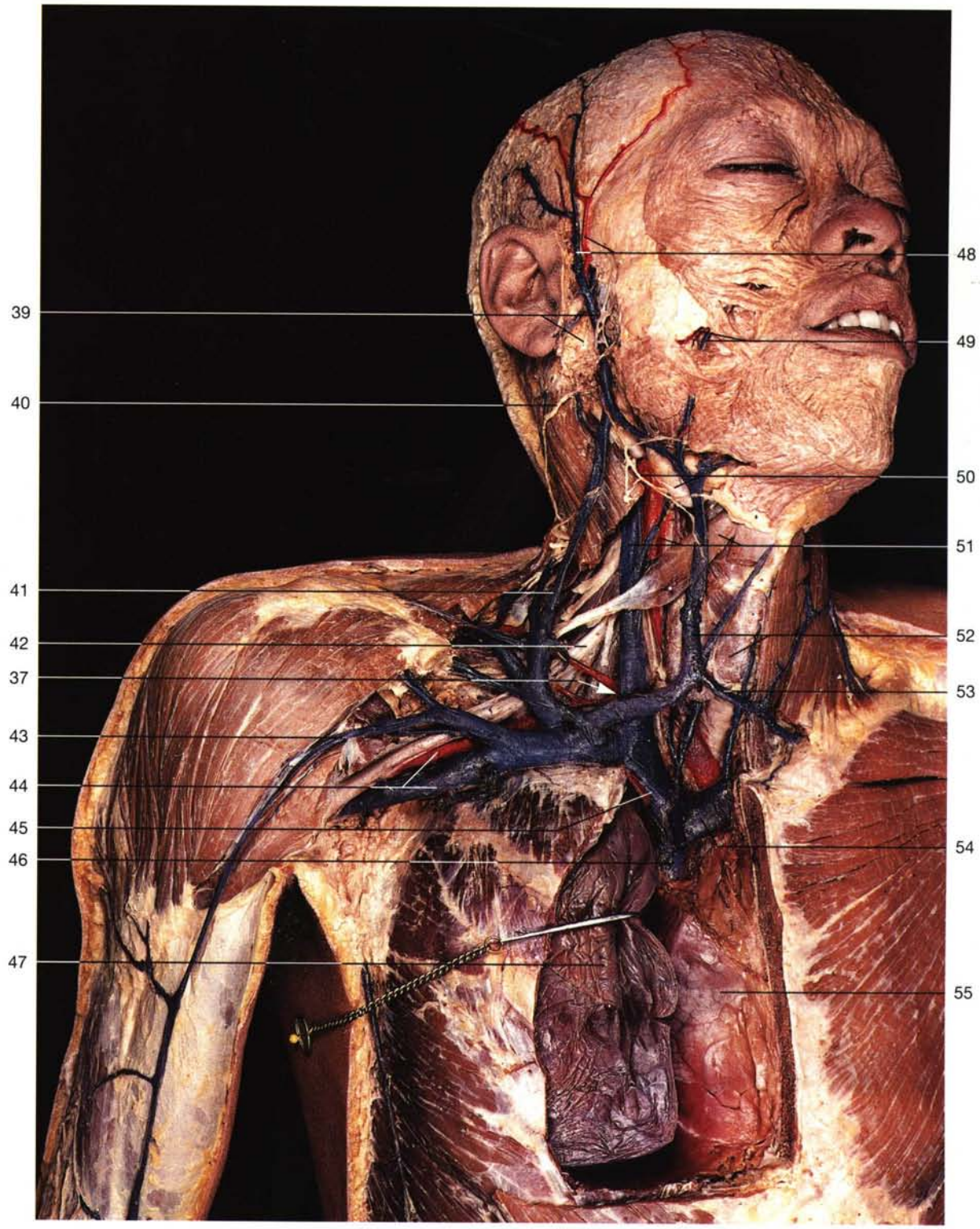

Veins of head and neck (anterior aspect). Part of the thoracic wall, clavicle and sternocleidomastoid muscle have been removed. Veins were colored blue, arteries red.

The **internal jugular vein** is the continuation of the sigmoid sinus which drains most of the venous blood from the brain together with the external cerebrospinal fluid. By joining the subclavian vein it forms the right brachiocephalic vein which continues on the right side directly into the superior vena cava. The common way to introduce the lead from a pacemaker device into the heart is by way of the cephalic vein. On the left side the thoracic duct joins the internal jugular vein at that point where the subclavian vein and the internal jugular vein form the left brachiocephalic vein. Note that the subclavian vein lies in front of the scalenus anterior muscle whereas the subclavian artery together with the plexus brachialis lies posterior to that muscle. The cephalic vein joins the axillary vein by passing into the deltopectoral triangle. The subclavian vein is strongly fixed to the first rib so that it can be punctured with a needle at that point (underneath the sternal end of clavicle) to introduce a catheter (subclavian line).

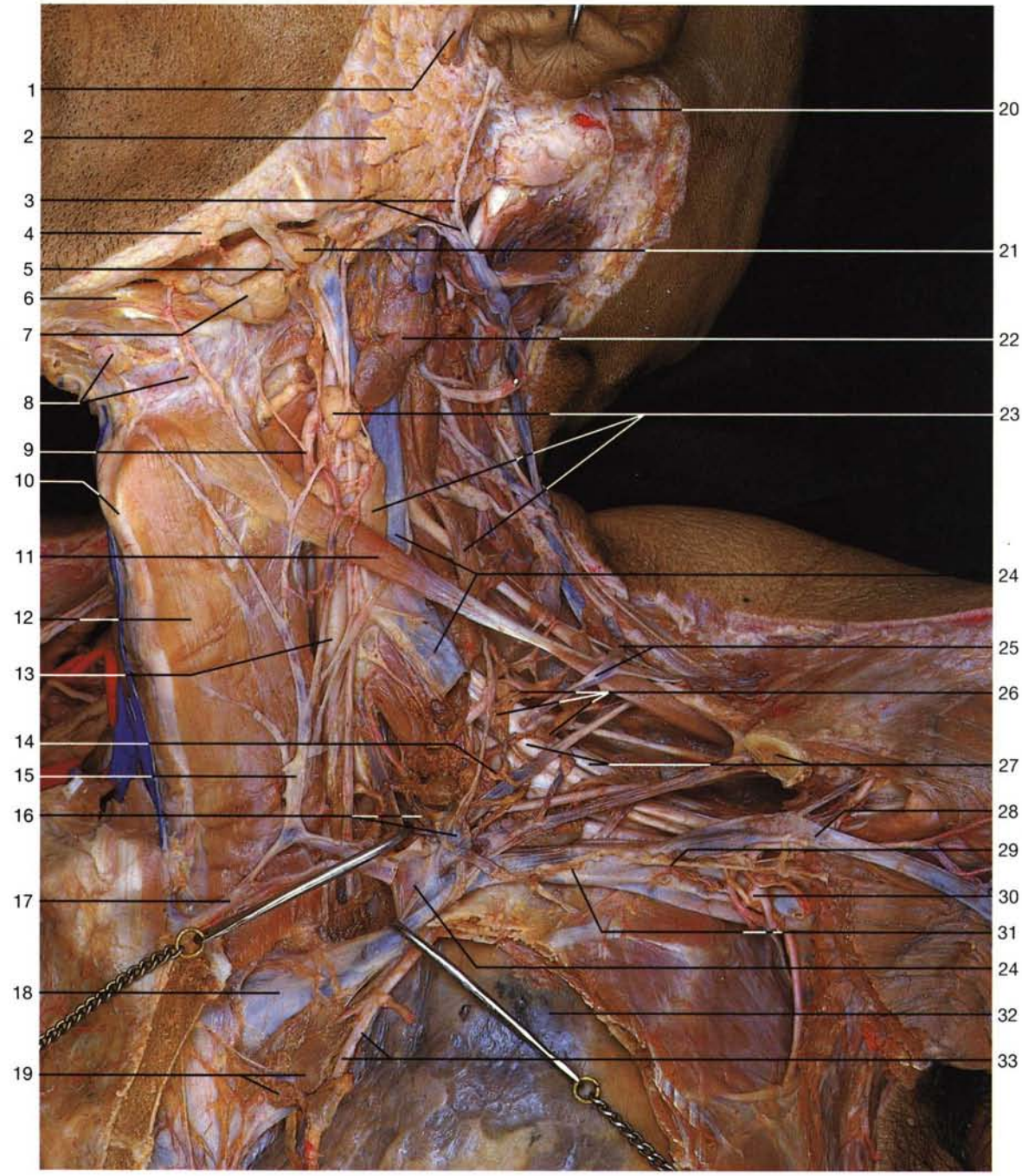

Lymph nodes and lymph vessels of the neck, left side oblique (lateral aspect). The sternocleidomastoid muscle and the left half of the thoracic wall have been removed. Lower part of the internal jugular vein has been cut and laterally displaced to show the thoracic duct.

1 **Superficial parotid lymph node**
2 Parotid gland
3 Great auricular nerve
4 Mandible
5 Facial vein
6 Anterior belly of digastric muscle
7 Submandibular gland
8 **Submental lymph nodes**
9 Superior thyroid artery
10 Thyroid cartilage
11 Omohyoid muscle
12 Sternohyoid muscle
13 Common carotid artery
14 **Supraclavicular lymph nodes**
15 Anterior jugular vein
16 Thoracic duct
17 Jugular venous arch
18 Left brachiocephalic vein
19 **Superior mediastinal lymph nodes**
20 Retroauricular lymph nodes
21 **Submandibular nodes**
22 **Superficial cervical lymph nodes**
23 **Jugulodigastric lymph nodes and jugular trunk**
24 **Internal jugular vein**
25 External jugular vein
26 **Jugulo-omohyoid lymph nodes**
27 Brachial plexus
28 Cephalic vein
29 Subclavian trunk
30 **Infraclavicular lymph nodes**
31 Subclavian vein
32 Lung
33 Internal thoracic artery and vein

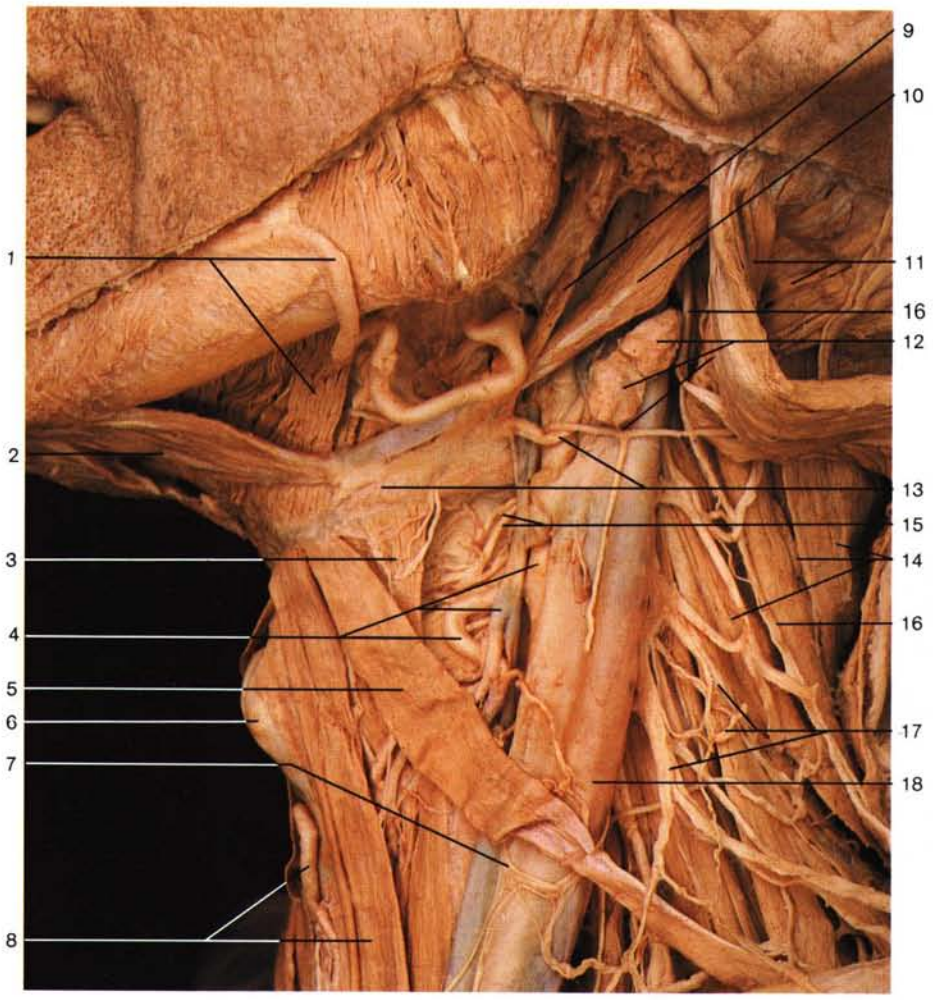

Carotid triangle, left side (lateral aspect). Sternocleidomastoid muscle reflected.

1 Mylohyoid muscle and facial artery
2 Anterior belly of digastric muscle
3 Thyrohyoid
4 **External carotid artery,** superior thyroid artery and vein
5 Omohyoid muscle
6 Thyroid cartilage
7 **Ansa cervicalis**
8 Sternohyoid muscle and superior thyroid artery
9 Stylohyoid muscle
10 Posterior belly of digastric muscle
11 Sternocleidomastoid muscle (reflected)
12 **Superior cervical lymph nodes** and sternocleidomastoid artery
13 Hyoid bone and hypoglossal nerve (n. XII)
14 Splenius capitis and levator scapulae muscles
15 Superior laryngeal artery and internal branch of superior laryngeal nerve
16 **Accessory nerve**
17 Cervical plexus
18 **Internal jugular vein**
19 Facial vein
20 Submental nodes
21 **Thoracic duct**
22 Retroauricular nodes
23 Parotid nodes
24 Occipital nodes
25 **Submandibular nodes**
26 Jugulodigastric nodes } **deep cervical nodes**
27 Jugulo-omohyoid nodes } **deep cervical nodes**
28 Jugular trunk
29 Subclavian trunk
30 Infraclavicular nodes
31 External jugular vein

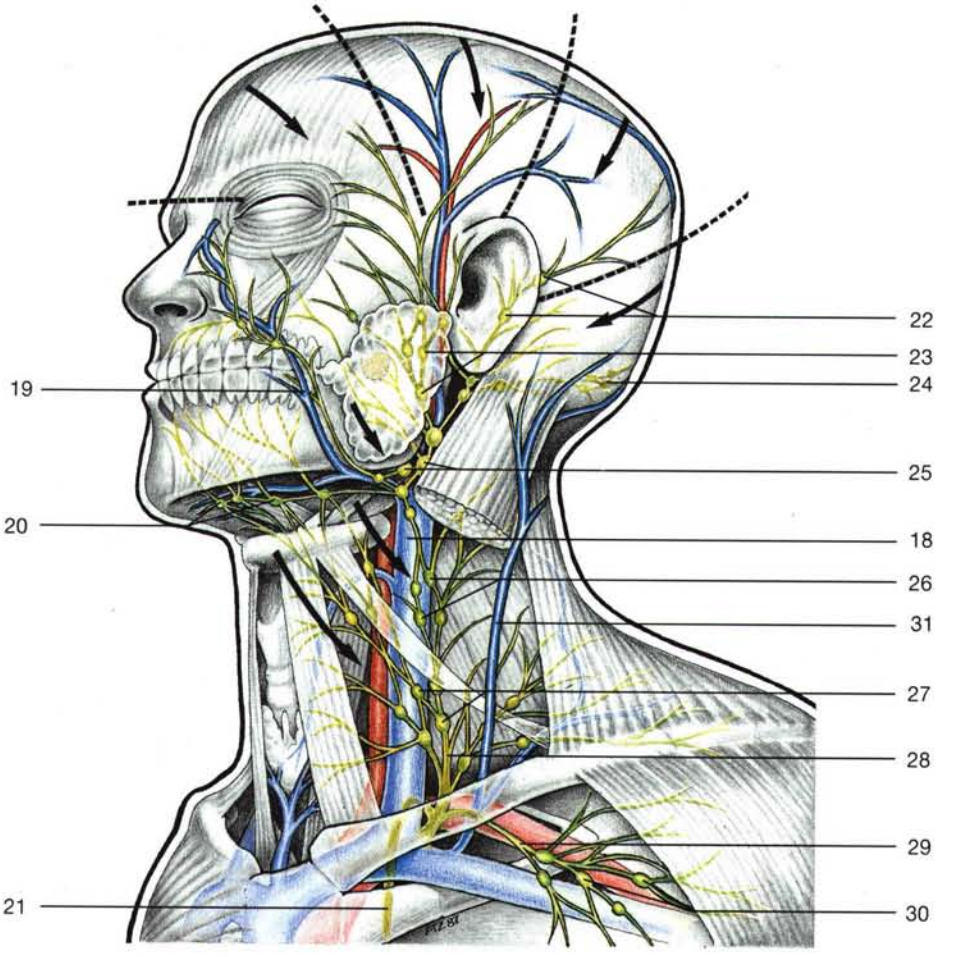

Lymph nodes and veins of head and neck. Dotted lines = border between irrigation areas; arrows: direction of lymph flow.

Neck (anterior aspect).
The superficial fascia has been removed.

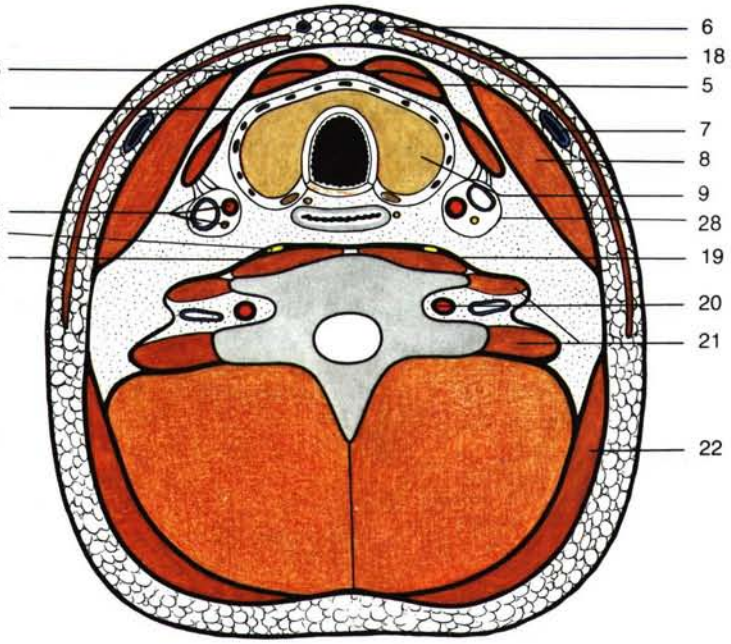

Cross-section of the neck at the level of the thyroid gland.
Notice the postion of the three laminae of cervical fascia (23, 24, 25).

1 Mandible
2 **Facial artery and vein**
3 Anterior belly of digastric muscle
4 Mylohyoid muscle
5 Infrahyoid muscles (sternohyoid, sternothyroid and omohyoid)
6 **Anterior jugular veins**
7 **External jugular vein**
8 Sternocleidomastoid muscle
9 **Thyroid gland**
10 **Submandibular gland**
11 Cervical branch of facial nerve
12 Great auricular nerve
13 Transverse cervical nerves
14 Lateral supraclavicular nerves
15 Middle supraclavicular nerves
16 Medial supraclavicular nerves
} Cutaneous branches of cervical plexus
17 Clavicle
18 Platysma muscle
19 Prevertebral lamina of cervical fascia, covering longus colli muscle
20 Vertebral artery and vein
21 Scalenus muscles
22 Trapezius muscle
23 **Superficial lamina of cervical fascia**
24 **Pretracheal lamina of cervical fascia**
25 **Prevertebral lamina of cervical fascia** with sympathetic trunk
26 **Carotid sheath** with common carotid artery, internal jugular vein and vagus nerve
27 Cervical part of sympathetic trunk
28 Carotid sheath

Anterior triangle (anterior aspect). The pretracheal lamina of cervical fascia and left sternocleidomastoid muscle have been removed.

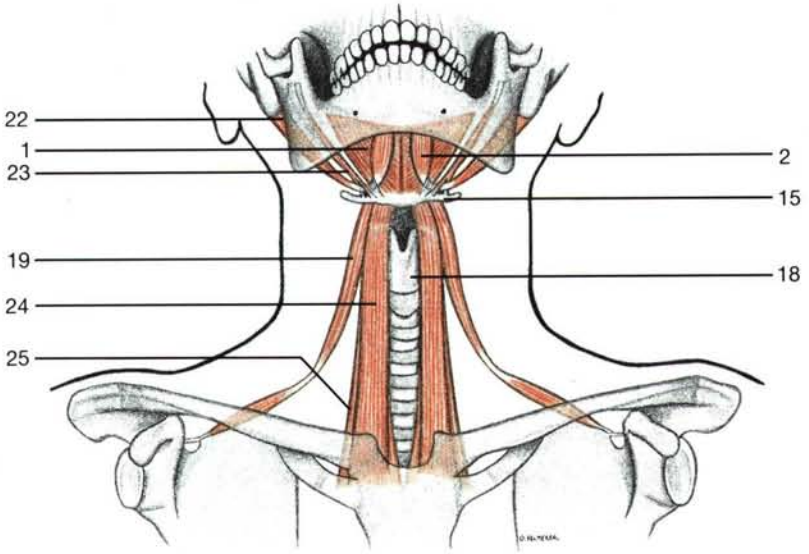

Supra- and infrahyoid muscles. (Schematic drawing.)

1 Mylohyoid muscle
2 Anterior belly of digastric muscle
3 **Facial artery**
4 **Submandibular gland**
5 Great auricular nerve
6 Internal jugular vein and common carotid artery
7 Transverse cervical nerve and omohyoid muscle
8 **Sternohyoid muscle** and **superior thyroid artery**
9 Sternocleidomastoid muscle (sternal head)
10 Left sternocleidomastoid muscle (reflected)
11 Sternocleidomastoid muscle (clavicular head) and lateral supraclavicular nerves
12 Middle **supraclavicular nerves**
13 Medial supraclavicular nerves
14 Mandible
15 **Hyoid bone**
16 Superficial cervical lymph nodes
17 Left superior thyroid artery and external carotid artery
18 **Thyroid cartilage**
19 **Omohyoid muscle** (superior belly)
20 **Internal jugular vein** and branches of ansa cervicalis
21 Thyroid gland and unpaired inferior thyroid vein
22 Posterior belly of digastric muscle
23 **Stylohyoid muscle**
24 **Sternohyoid muscle**
25 **Sternothyroid muscle**

Posterior and carotid triangles (lateral aspect). Superficial dissection.

1 **Parotid gland** and **great auricular nerve**
2 **Lesser occipital nerve**
3 Internal and external jugular veins
4 Retromandibular vein and external carotid artery
5 **Transverse cervical nerve** with communicating branch to cervical branch of facial nerve
6 Trapezius muscle and superficial lamina of cervical fascia
7 Lateral supraclavicular nerves
8 Middle **supraclavicular nerves**
9 Pectoralis major muscle
10 Buccal branch of facial nerve and masseter
11 **Facial artery and vein** and mandibular branch of facial nerve
12 Cervical branch of facial nerve and submandibular gland
13 Thyroid cartilage
14 Omohyoid muscle
15 Sternohyoid muscle
16 Sternocleidomastoid muscle
17 Medial supraclavicular nerves
18 Mandibular branch of facial nerve
19 Cervical branch of facial nerve with communicating branch to transverse cervical nerve

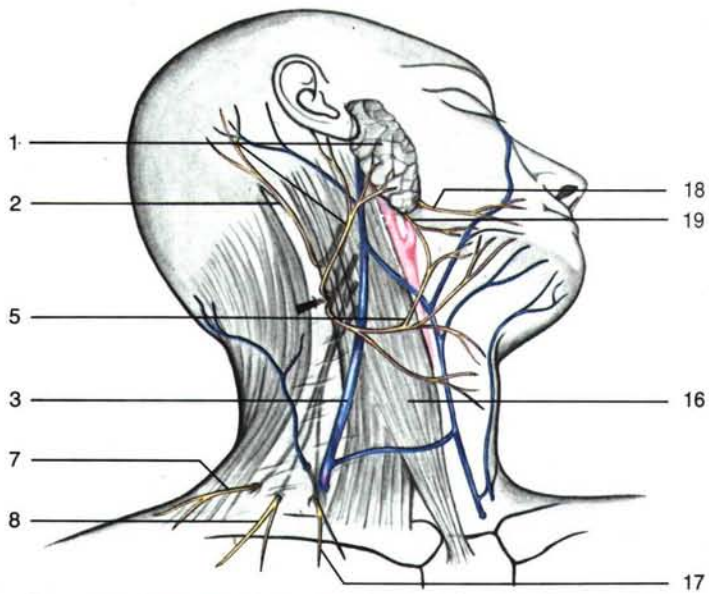

Cutaneous branches of cervical plexus. Erb's point is indicated by an arrowhead. (Schematic diagram.)

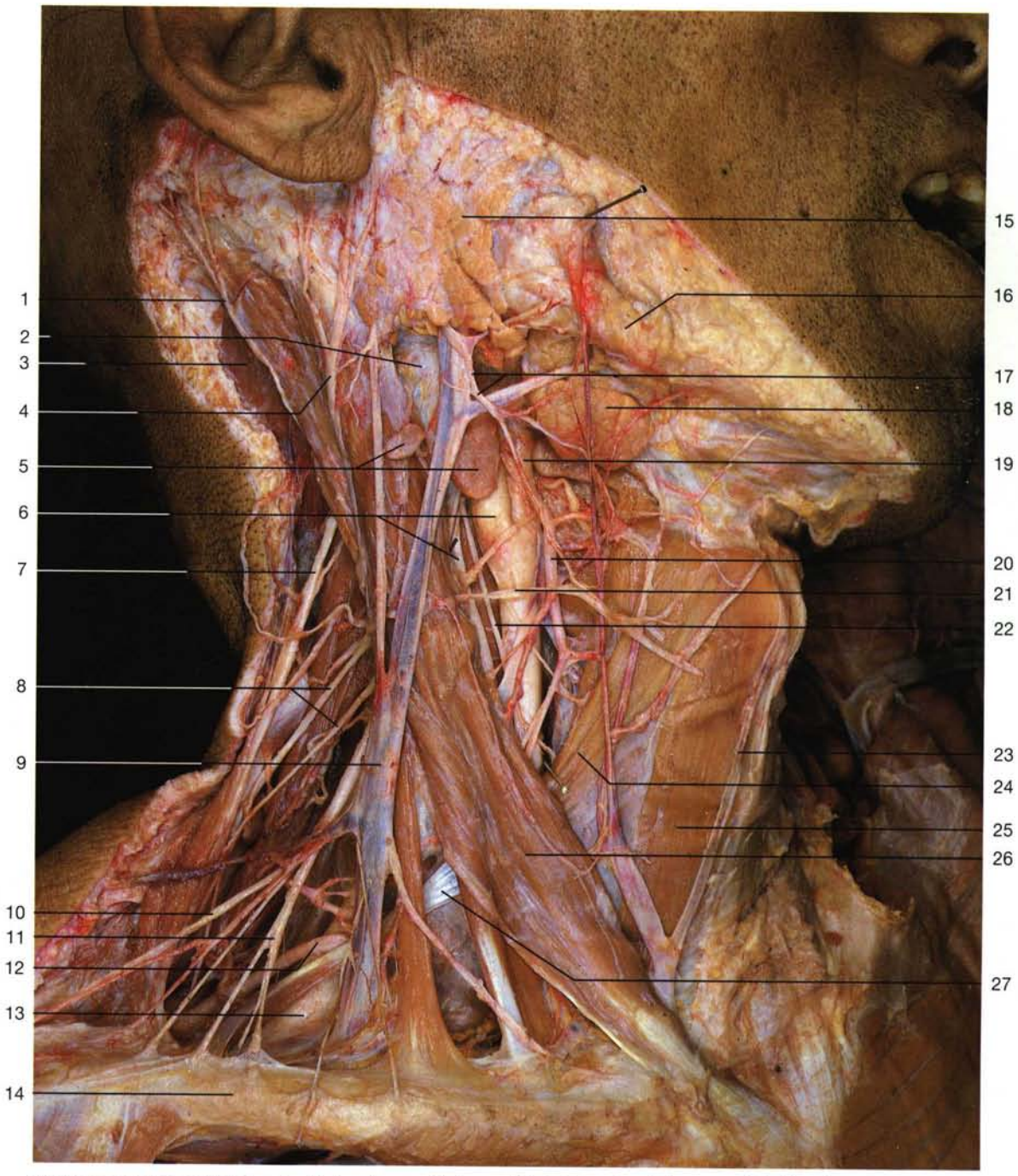

Posterior and carotid triangles (lateral aspect). Superficial dissection. The superficial lamina of cervical fascia has been removed to display the cutaneous branches of the cervical plexus and subcutaneous veins.

1 **Lesser occipital nerve**
2 **Internal jugular vein**
3 Splenius capitis muscle
4 **Great auricular nerve**
5 Submandibular nodes
6 Internal carotid artery and vagus nerve
7 **Accessory nerve**
8 Muscular branches of cervical plexus
9 External jugular vein
10 **Posterior supraclavicular nerves**
11 **Middle supraclavicular nerves**
12 Suprascapular artery
13 Pretracheal lamina of fascia of neck
14 Clavicle
15 Parotid gland
16 Mandible
17 Cervical branch of facial nerve
18 Submandibular gland
19 External carotid artery
20 Superior thyroid artery
21 Transverse cervical nerve
22 Superior root of ansa cervicalis
23 Anterior jugular vein
24 Omohyoid muscle
25 Sternohyoid muscle
26 Sternocleidomastoid muscle
27 Intermediate tendon of omohyoid muscle

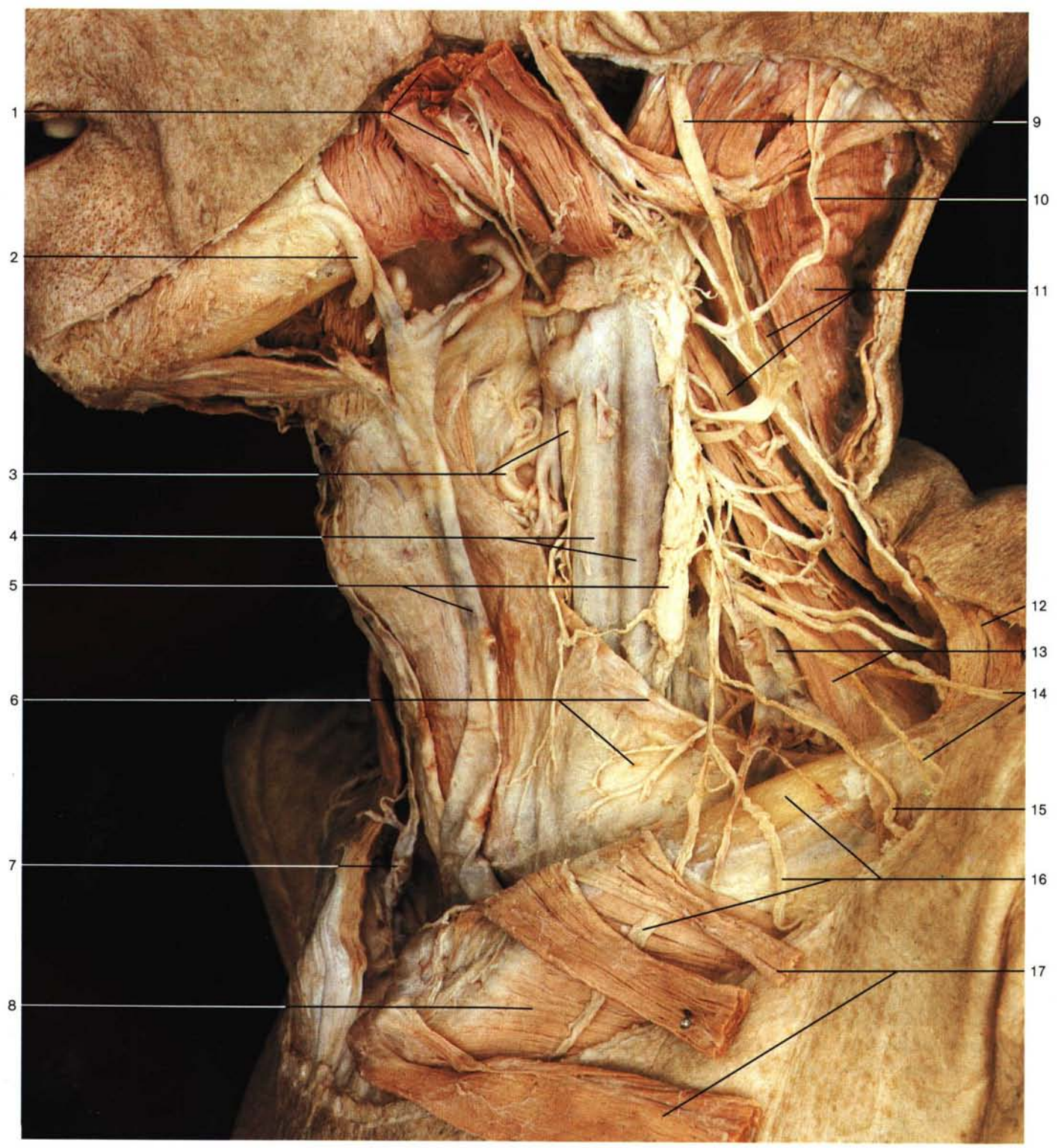

Neck, superficial dissection (lateral aspect). Sternocleidomastoid muscle has been cut and reflected to display the pretracheal lamina of the cervical fascia.

1 Sternocleidomastoid muscle (reflected) and branch of accessory nerve
2 Facial artery
3 External carotid artery and superior thyroid artery
4 **Internal jugular vein**
5 **Deep cervical lymph nodes** and external jugular vein
6 Omohyoid muscle and **pretracheal lamina of cervical fascia**
7 Anterior jugular vein
8 Pectoralis major muscle
9 **Great auricular nerve**
10 **Lesser occipital nerve**
11 Splenius capitis and levator scapulae muscles
12 Trapezius muscle
13 Scalenus medius muscle and brachial plexus
14 Posterior **supraclavicular nerves**
15 Middle supraclavicular nerve
16 Clavicle and anterior supraclavicular nerves
17 Sternocleidomastoid muscle (reflected)

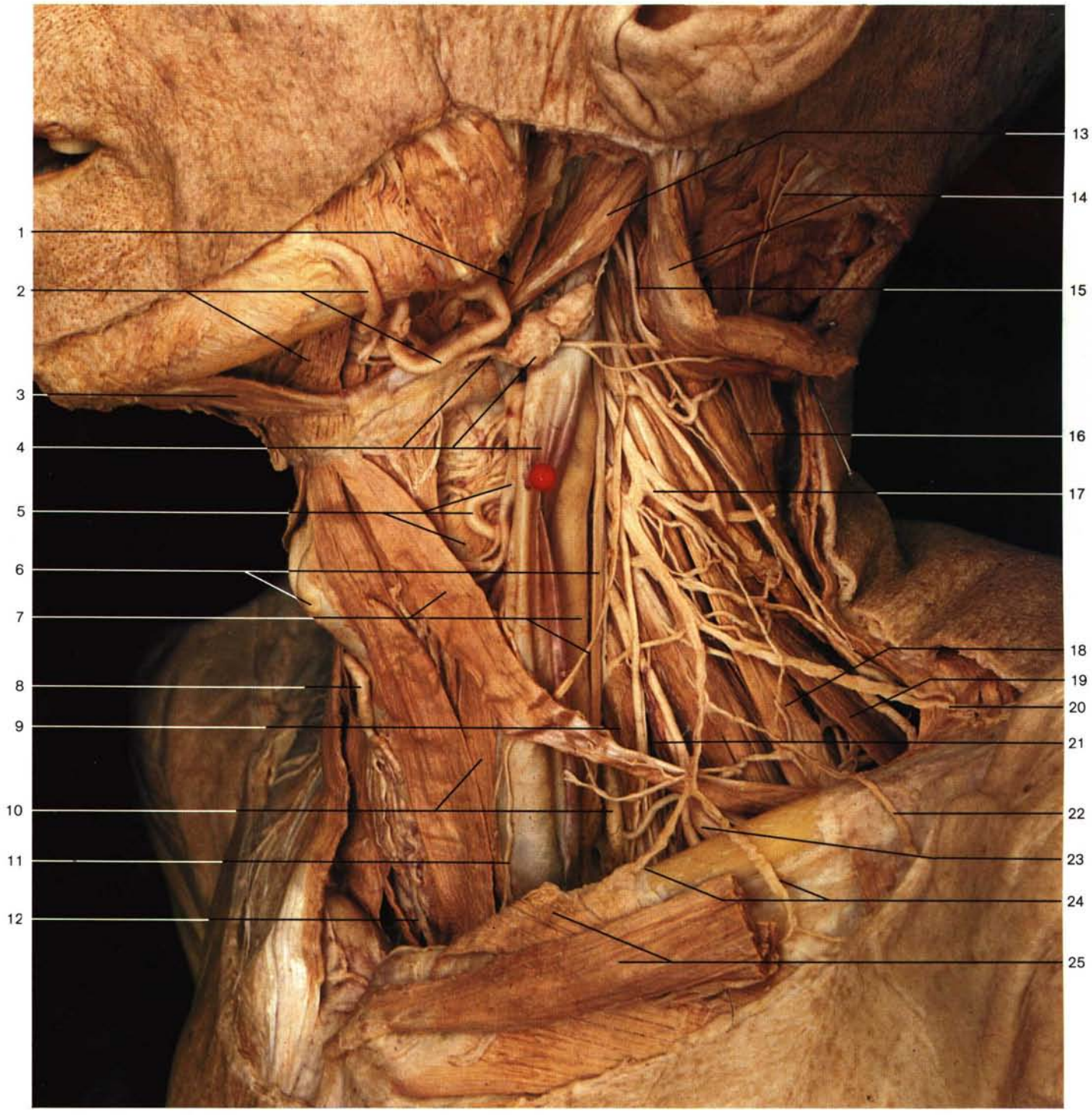

Neck, deep dissection (lateral aspect). The internal jugular vein has been reflected to expose the carotid artery and vagus nerve.

1 Stylohyoid muscle
2 Facial artery and mylohyoid muscle
3 Anterior belly of digastric muscle
4 **Internal jugular vein,** hypoglossal nerve, and superficial cervical lymph nodes
5 **Superior thyroid artery and vein** and inferior pharyngeal constrictor muscle
6 Thyroid cartilage and vagus nerve
7 **Ansa cervicalis,** omohyoid muscle, and common carotid artery
8 Right superior thyroid artery
9 Scalenus anterior muscle
10 Sternothyroid muscle and inferior thyroid artery
11 Muscular branches of ansa cervicalis to the infrahyoid muscles
12 Inferior thyroid vein
13 Posterior belly of digastric muscle
14 Sternocleidomastoid muscle and lesser occipital nerve
15 **Accessory nerve**
16 Splenius capitis muscle
17 **Cervical plexus**
18 Scalenus posterior muscle
19 Levator scapulac muscle
20 Posterior **supraclavicular nerves**
21 Phrenic nerve
22 Middle supraclavicular nerve
23 **Brachial plexus**
24 Anterior supraclavicular nerves
25 Sternocleidomastoid muscle

Neck, deeper dissection (lateral aspect). **Ansa cervicalis.** The cervical fascia and the clavicle are partly removed. Ansa cervicalis and infrahyoid muscles are displayed.

1 Masseter muscle
2 Mylohyoid muscle and facial artery
3 **External carotid artery** and anterior belly of digastric muscle
4 **Hypoglossal nerve**
5 Thyrohyoid muscle
6 Superior thyroid artery and vein and inferior pharyngeal constrictor muscle
7 Omohyoid muscle (superior belly)
8 **Ansa cervicalis,** thyroid gland and **internal jugular vein**
9 Sternothyroid muscle
10 Sternohyoid muscle
11 **Thoracic duct**
12 Pectoralis minor muscle
13 Pectoralis major muscle
14 Posterior belly of digastric muscle
15 Sternocleidomastoid muscle and lesser occipital nerve
16 Splenius capitis muscle
17 Superficial cervical lymph nodes and accessory nerve
18 **Cervical plexus**
19 Scalenus medius muscle
20 Levator scapulae muscle
21 Scalenus posterior muscle
22 Brachial plexus
23 Transverse cervical artery and clavicle
24 Subclavius muscle
25 **Subclavian artery and vein**
26 Thoracoacromial artery
27 Cephalic vein

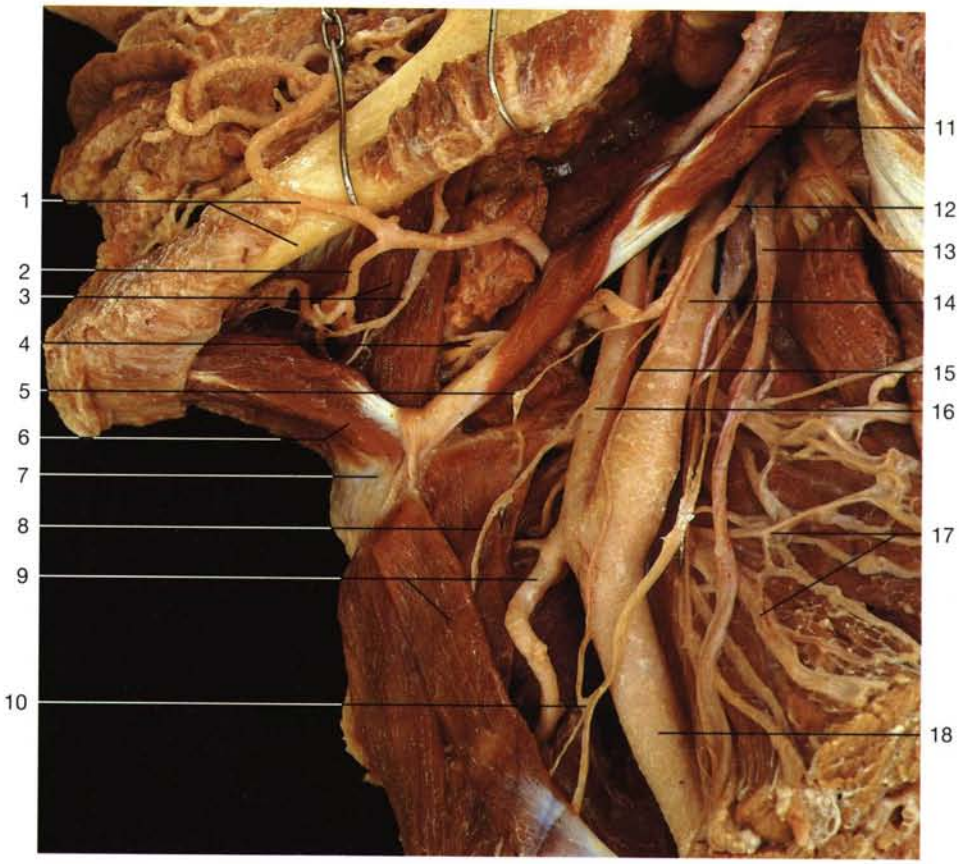

1 **Facial artery** and mandible
2 Submental artery
3 Mylohyoid muscle and nerve
4 **Hypoglossal nerve** (lingual branches)
5 Thyrohyoid branch of hypoglossal nerve (n. XII)
6 Anterior belly of digastric muscle
7 Hyoid bone
8 Omohyoid branch of hypoglossal nerve (n. XII)
9 Omohyoid muscle and superior thyroid artery
10 **Ansa cervicalis**
11 Posterior belly of digastric muscle
12 **Hypoglossal nerve** (n. XII)
13 **Vagus nerve** (n. X)
14 Internal carotid artery
15 Superior root of ansa cervicalis
16 **External carotid artery**
17 Cervical plexus
18 Common carotid artery

Neck, submandibular region (lateral aspect). **Hypoglossal nerve** (n. XII). Mandible slightly elevated.

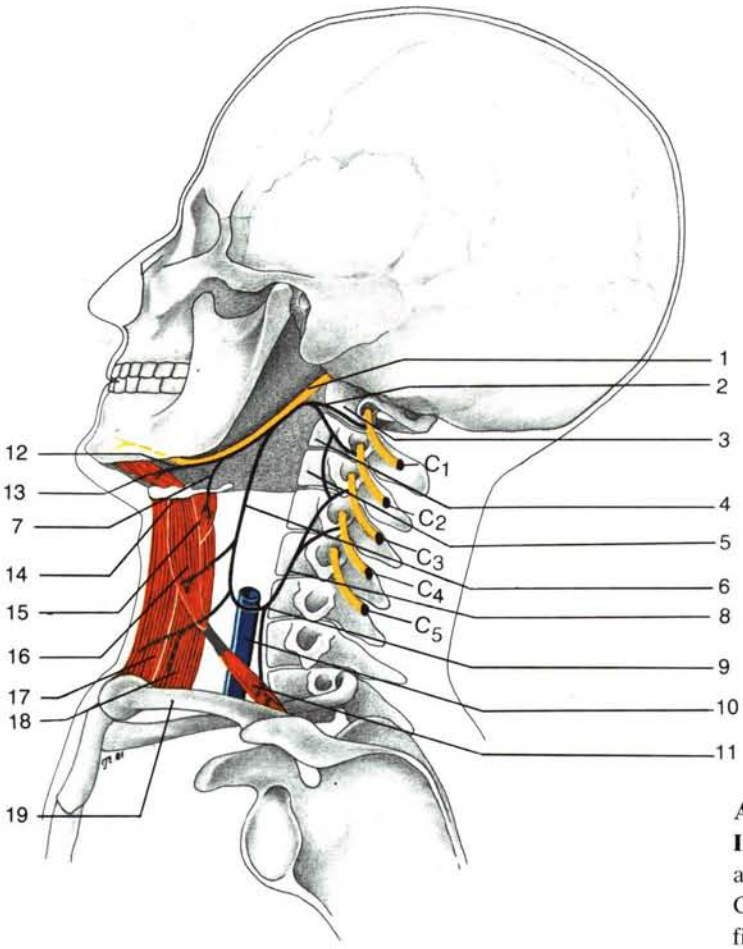

1 **Hypoglossal nerve** (n. XII)
2 Communication from the ventral ramus of the first cervical spinal nerve
3 Atlas
4 Axis
5 Third cervical vertebra
6 Superior root of ansa cervicalis
7 Thyrohyoid branch of hypoglossal nerve
8 Inferior root of ansa cervicalis
9 **Ansa cervicalis**
10 Internal jugular vein
11 Inferior belly of omohyoid muscle
12 Geniohyoid branch of hypoglossal nerve
13 Geniohyoid muscle
14 Hyoid bone
15 Thyrohyoid muscle
16 Superior belly of omohyoid muscle
17 Sternohyoid muscle
18 Sternothyroid muscle
19 Clavicle

Ansa cervicalis.
Innervation of infrahyoid muscles. Cervical plexus and its communication with the hypoglossal nerve. C_1–C_4 = ventral rami of cervical spinal nerves of the first four segments.

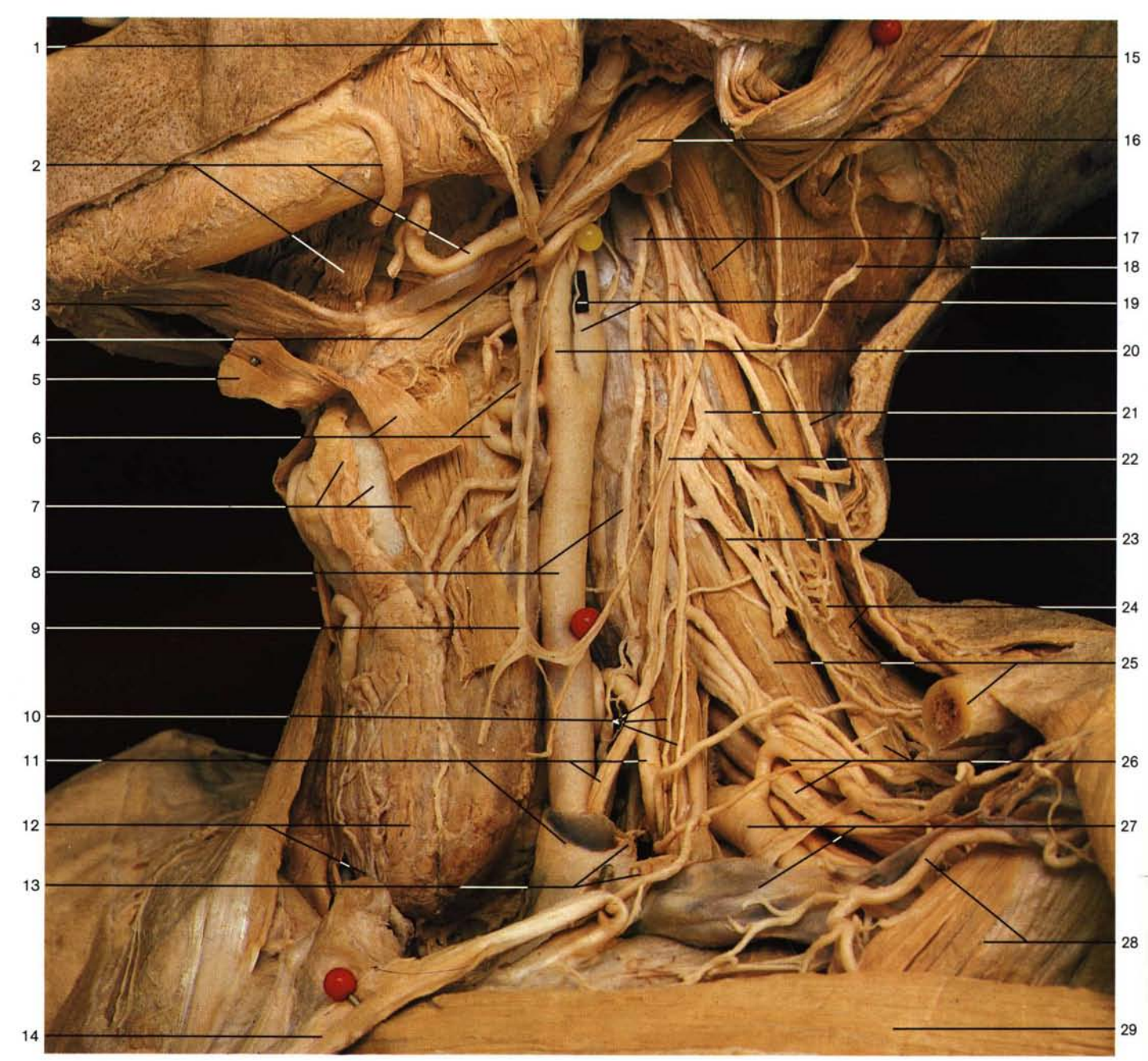

Neck, deep dissection (lateral aspect). Clavicle partly removed to show the slit between the scalenus muscles. Internal jugular vein removed.

1 Masseter muscle
2 Mylohyoid muscle and **facial artery**
3 Anterior belly of digastric muscle
4 **Hypoglossal nerve**
5 Sternohyoid muscle
6 Omohyoid muscle, superior thyroid artery and vein
7 Sternothyroid muscle, thyroid cartilage, and pyramidal lobe of thyroid gland
8 Common carotid artery and sympathetic trunk
9 **Ansa cervicalis**
10 **Phrenic nerve,** ascending cervical artery and anterior scalenus muscle
11 Inferior thyroid artery, vagus nerve and internal jugular vein (cut)
12 **Thyroid gland** and unpaired inferior thyroid venous plexus
13 Thoracic duct and left subclavian trunk
14 Subclavius muscle (reflected)
15 Sternocleidomastoid muscle (reflected)
16 Posterior belly of digastric muscle
17 **Superior cervical ganglion** and splenius muscle
18 **Lesser occipital nerve**
19 Internal carotid artery and branch of the glossopharyngeal nerve to the carotid body
20 **External carotid artery**
21 **Cervical plexus** and **accessory nerve**
22 Inferior root of ansa cervicalis
23 Supraclavicular nerve
24 Levator scapulae muscle
25 Scalenus medius muscle and clavicle
26 Transverse cervical artery, brachial plexus, and scalenus posterior muscle
27 **Subclavian artery and vein**
28 Thoracoacromial artery and pectoralis minor muscle
29 Pectoralis major muscle

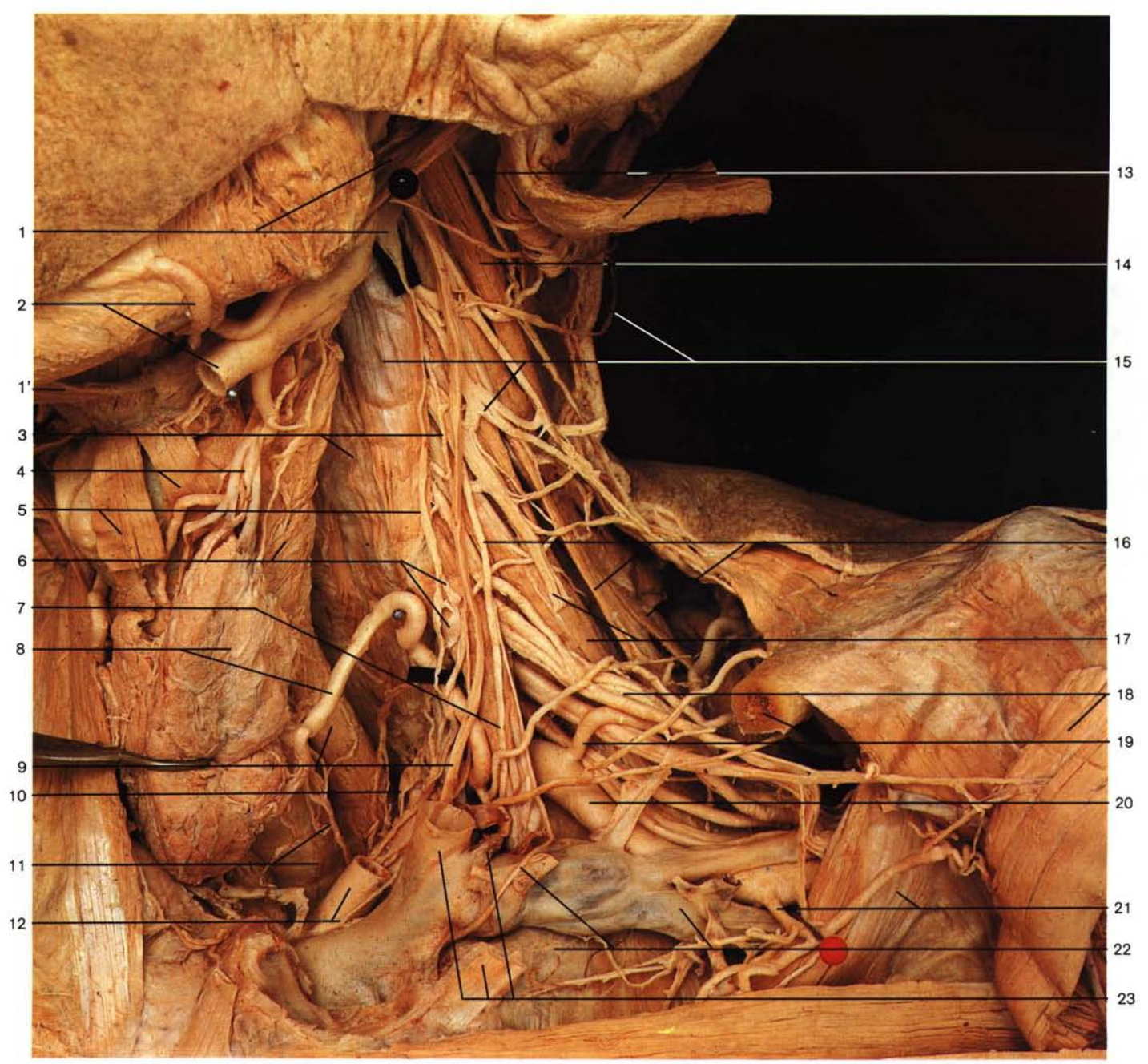

Neck, deepest dissection (anterolateral aspect). Thyroid gland reflected to expose the esophagus and the recurrent laryngeal nerve.

1 **Superior cervical ganglion** of sympathetic trunk and posterior belly of digastric muscle
1' Anterior belly of digastric muscle
2 **Facial artery** and common carotid artery (reflected anteriorly)
3 Ascending cervical artery and longus colli muscle
4 Omohyoid muscle and superior thyroid artery
5 **Sympathetic trunk** and sternohyoid muscle
6 Middle cervical ganglion and inferior pharyngeal constrictor muscle
7 Scalenus anterior muscle and phrenic nerve
8 **Thyroid gland** and **inferior thyroid artery**
9 Vagus nerve and esophagus
10 **Stellate ganglion**
11 **Recurrent laryngeal nerve** and trachea
12 Common carotid artery and cervical cardiac branch of vagus nerve
13 Sternocleidomastoid muscle and accessory nerve
14 Splenius capitis muscle
15 Lesser occipital nerve, longus capitis muscle and cervical plexus
16 **Phrenic nerve,** scalenus posterior muscle and levator scapulae muscle
17 Supraclavicular nerves and scalenus medius muscle
18 Brachial plexus and pectoralis major muscle (clavicular head)
19 Transverse cervical artery and clavicle
20 **Subclavian artery**
21 **Thoracoacromial artery** and pectoralis minor muscle
22 First rib, accessory phrenic nerve and subclavian vein
23 Internal jugular vein, thoracic duct and subclavius muscle

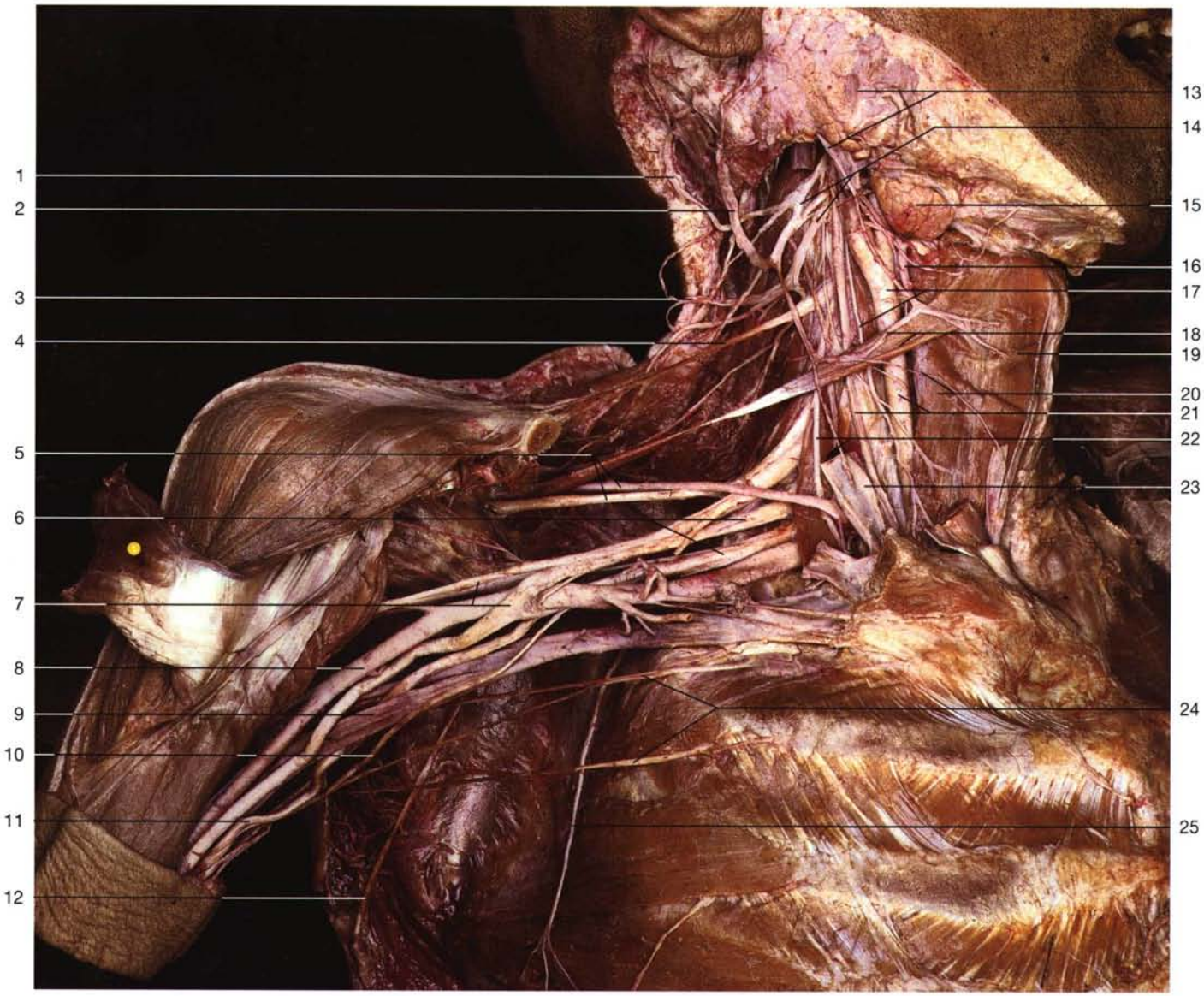

Neck and arm, deepest dissection (anterior-lateral aspect). Cervical and brachial plexus and their relation to the blood vessels are shown. Note the location and content of scalene triangle. Sternocleidomastoid muscle and clavicle have been removed; the internal jugular vein was divided to display the roots of cervical and brachial plexus.

1 Lesser occipital nerve
2 Great auricular nerve
3 Cutaneous branches of cervical plexus
4 Supraclavicular nerve
5 Suprascapular nerve and artery
6 **Brachial plexus**
7 **Median nerve** (with two roots) and musculocutaneous nerve
8 Axillary artery
9 Axillary vein
10 Medial brachial cutaneous nerve
11 **Ulnar nerve**
12 Thoracodorsal nerve
13 Parotid gland and facial nerve (cervical branch)
14 **Cervical plexus**
15 Submandibular gland
16 Superior thyroid artery
17 Common carotid artery dividing in internal and external carotid artery and superior root of ansa cervicalis
18 Omohyoid muscle and cervical branch of facial nerve joining the transverse cervical nerve (C_2, C_3)
19 Sternohyoid muscle
20 Transverse cervical nerve and sternothyroid muscle
21 **Common carotid artery** and **vagus nerve**
22 **Phrenic nerve** and scalenus anterior muscle
23 Internal jugular vein
24 Intercostobrachial nerves
25 Long thoracic nerve

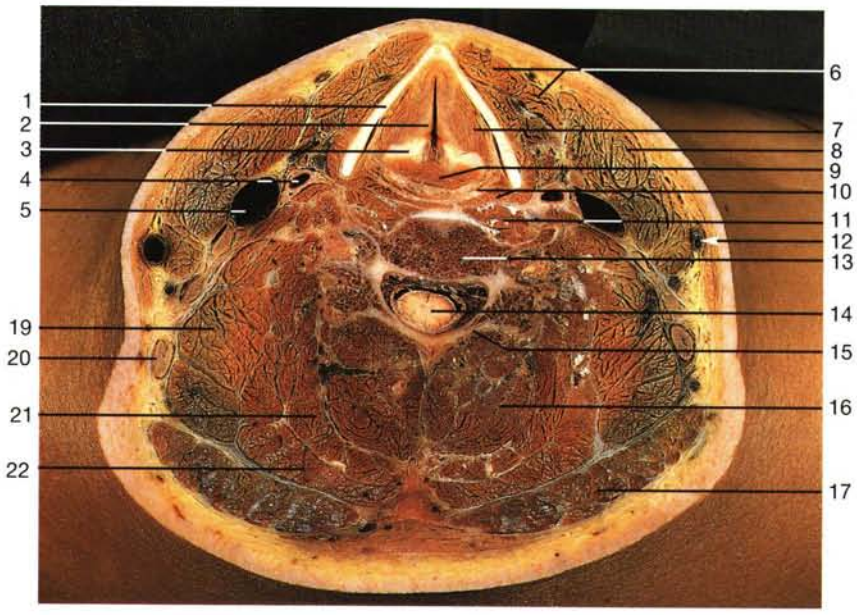

◁ **Horizontal section through the neck** at the level of the fissure of glottis, viewed from above.

1 **Thyroid cartilage**
2 **Vocal fold** and **glottis** (rima glottidis)
3 Arytenoid cartilage
4 Common carotid artery
5 **Internal jugular vein**
6 Infrahyoid muscles
7 Lateral thyroarytenoid muscle
8 Sternocleidomastoid muscle
9 Transverse arytenoid muscle
10 Laryngopharynx and inferior constricor muscle of pharynx
11 Longus colli muscle
12 External jugular vein
13 Body of cervical vertebra (C_5)
14 **Spinal cord**
15 Vertebra arch
16 Deep muscles of neck (semispinalis cervicis muscle)
17 Trapezius muscle
18 Rima glottidis
19 Levator scapulae muscle
20 Lymph node
21 Semispinalis capitis muscle
22 Splenius capitis muscle

◁ **Section through the neck at the level of larynx.** (MR-Scan.)

1 Hyoid bone
2 Thyroid cartilage
3 **Cervical plexus** (C_1-C_4)
4 **Phrenic nerve** (C_4)
5 Sealenus anterior muscle
6 **Brachial plexus** (C_5-T_1)
7 Scalenus medius and posterior muscles
8 Subclavian artery
9 Subclavian vein
10 Superior vena cava
11 Cricoid cartilage
12 Thyroid gland
13 **Internal jugular vein**
14 **Common carotid artery**
15 Inferior thyroid vein
16 Ascending aorta
17 Descending aorta
18 Second rib

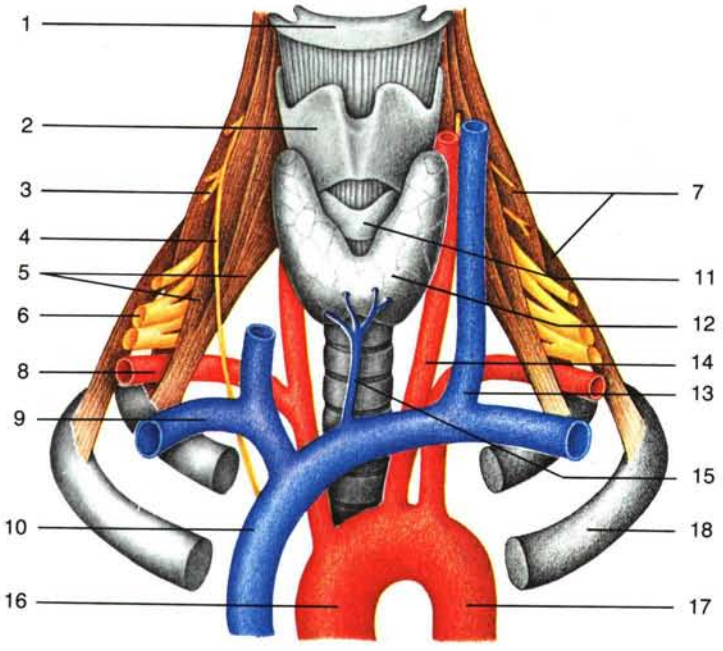

◁ **Scalene triangle, arrangements of blood vessels, and brachial plexus at the lower part of the neck.** (Schematic diagram.)

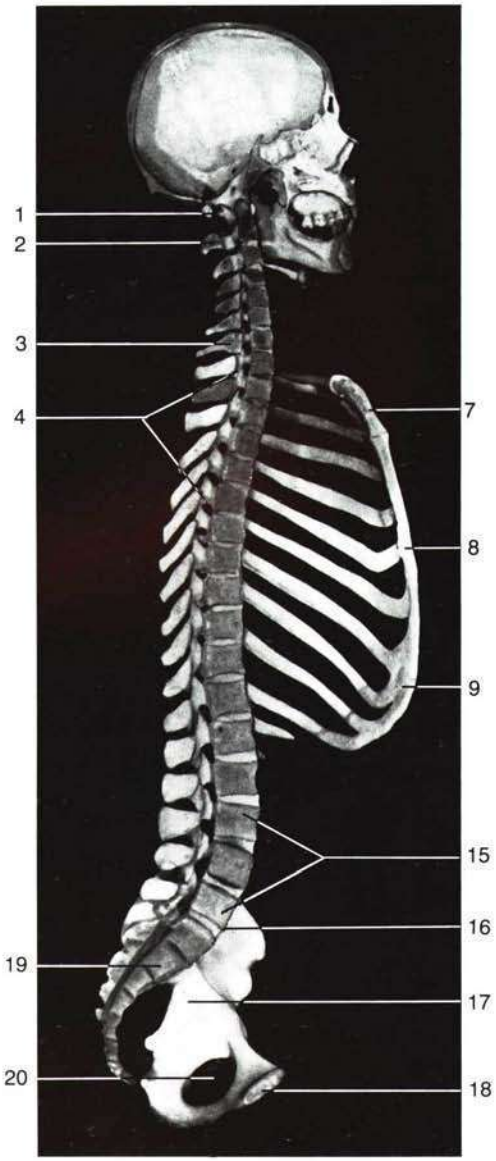

Median sagittal section through the vertebral column, head, and thorax of the adult.

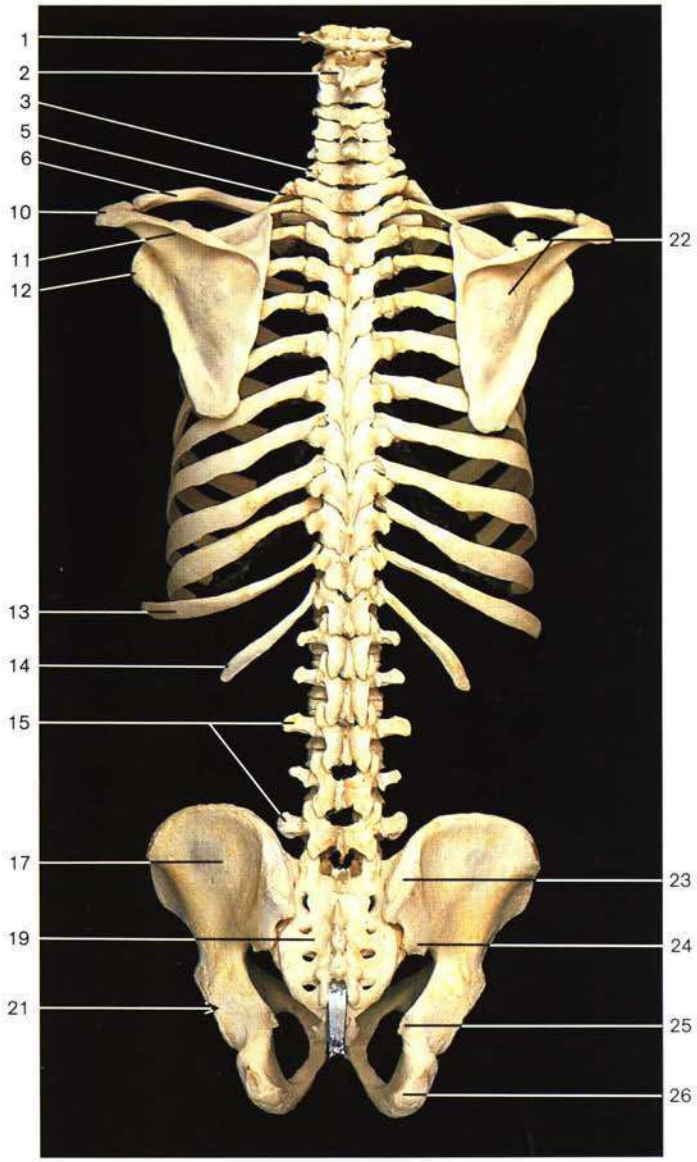

Skeleton of the trunk, vertebral column, thorax, and pelvis (posterior aspect).

1 Atlas
2 Axis
3 Seventh cervical vertebra (vertebra prominens)
4 Vertebral canal
5 First rib
6 Clavicle
7 Manubrium sterni
8 Body of sternum
9 Costal arch
10 Acromion
11 Spine of scapula
12 Glenoid cavity (lateral angle of scapula)
13 Eleventh rib
14 Twelfth rib
15 Lumbar vertebrae
16 Sacral promontory
17 Hip bone
18 Symphysis pubis
19 Sacrum
20 Obturator foramen
21 Acetabulum
22 Scapula with coracoid process
23 Posterior superior iliac spine
24 Posterior inferior iliac spine
25 Ischial spine
26 Ischial tuberosity

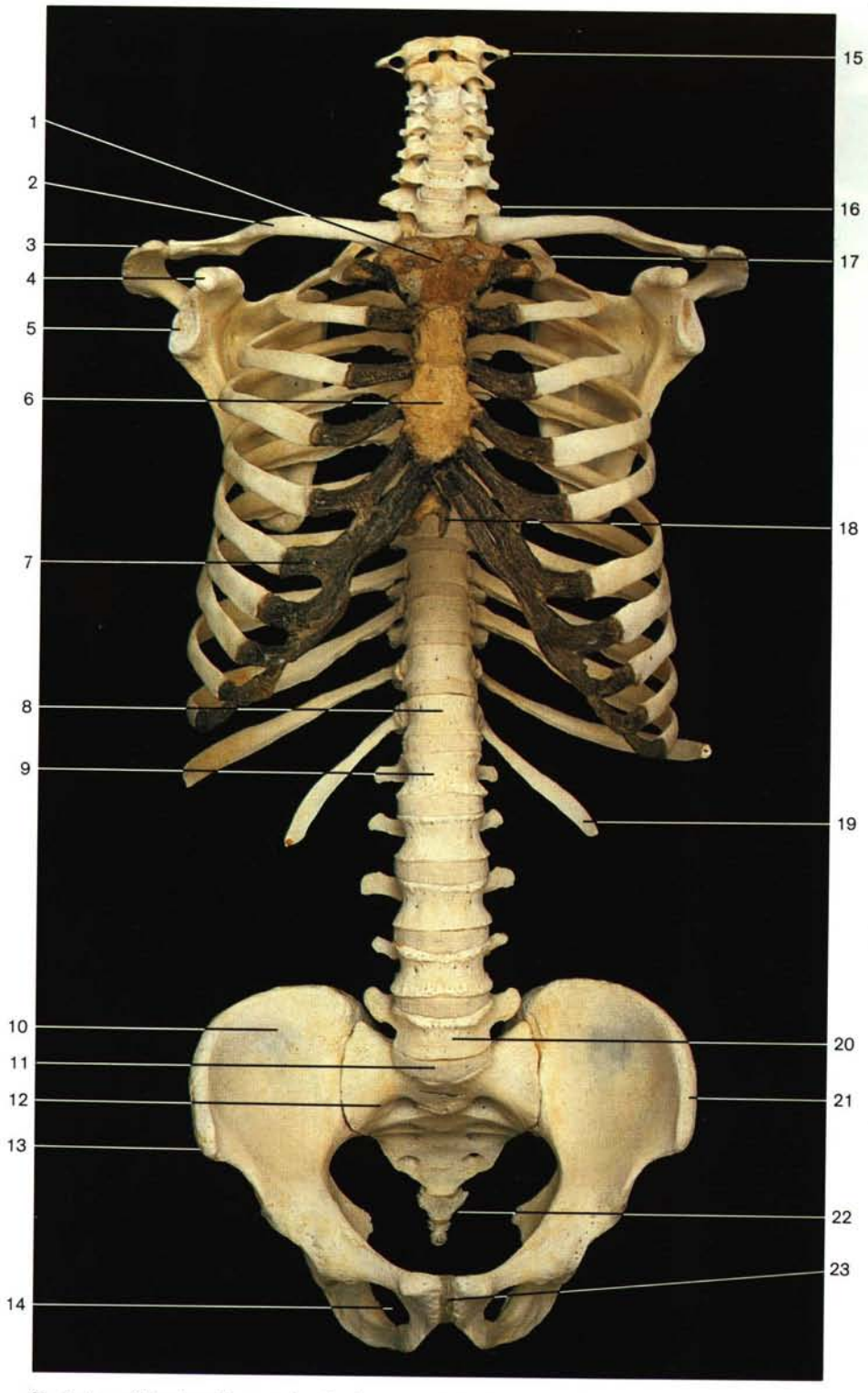

1 Manubrium sterni
2 **Clavicle**
3 Acromion
4 Coracoid process
5 Glenoid cavity
6 Body of **sternum**
7 Costal cartilage
8 Body of the twelfth **thoracic vertebra**
9 Body of the first **lumbar vertebra**
10 Hip bone
11 Sacral promontory
12 **Sacrum**
13 Anterior superior iliac spine
14 Obturator foramen
15 Atlas
16 Seventh **cervical vertebra**
17 First rib
18 Xiphoid process
19 Twelfth rib
20 Body of the fifth lumbar vertebra
21 Iliac crest
22 **Coccyx**
23 Symphysis pubis

Skeleton of the trunk, vertebral column, pelvis, thorax, and shoulder girdle (anterior aspect).

The trunk is divided into segments best visible in the thoracic region, where each segment consists of a pair of ribs connected anteriorly by the sternum and posteriorly by a thoracic vertebra. In the lumbar part of the vertebral column only vestiges of ribs are present which form what appear to be the transverse processes. In cervical vertebrae, remnants of ribs are part of the transverse processes. Each segment also comprises muscles (e.g. intercostal muscles), nerves, and vessels. However, in the cervical and lumbar region the muscular segments fuse with each other forming large muscle plates, for example, the oblique muscles of the abdomen, while vessels and nerves still retain their segmental pattern.

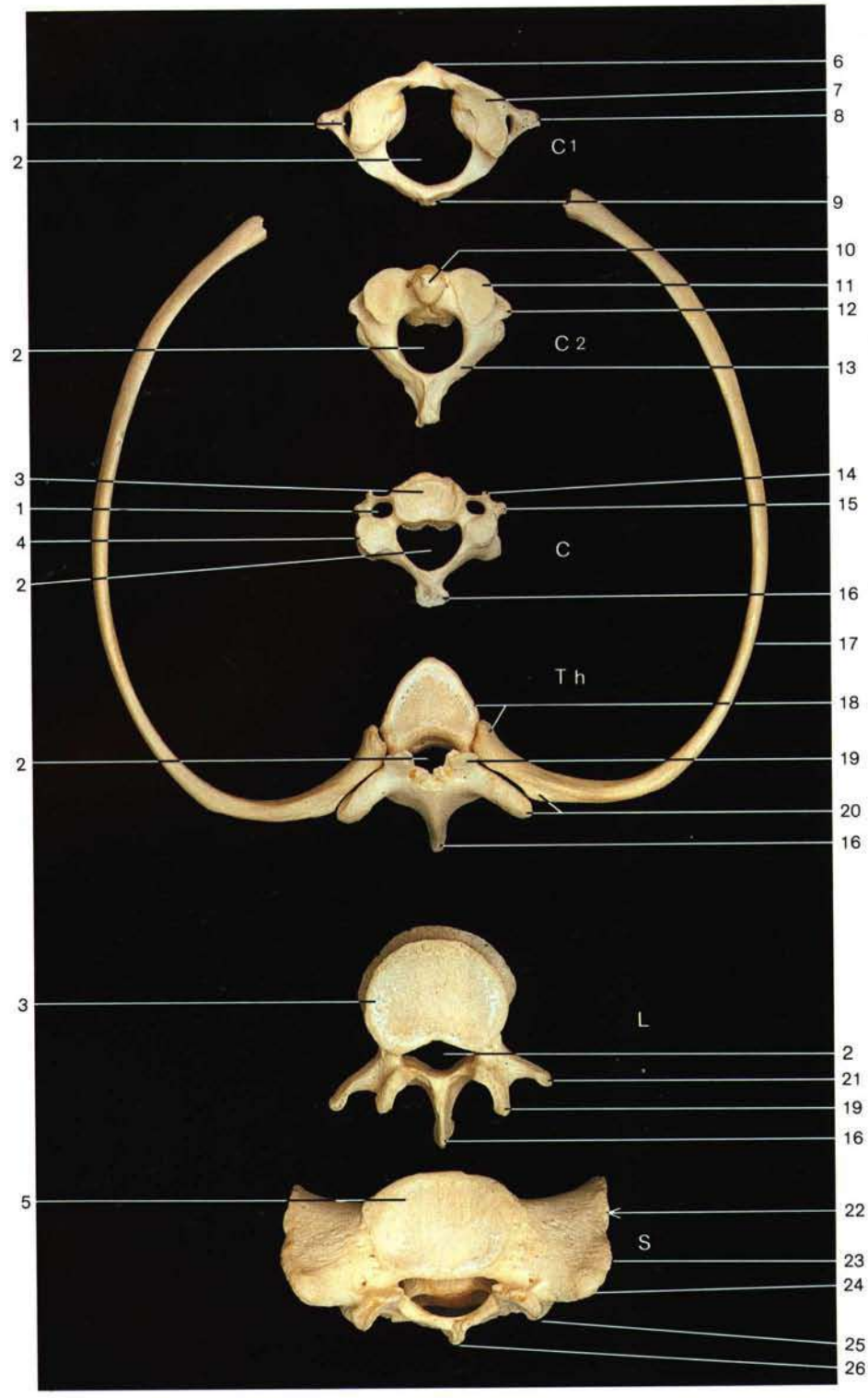

Representative vertebrae from each region of the vertebral column (superior aspect). From top to bottom: atlas (C_1), axis (C_2), cervical vertebra (C), thoracic vertebra (Th), lumbar vertebra (L), and sacrum (S).

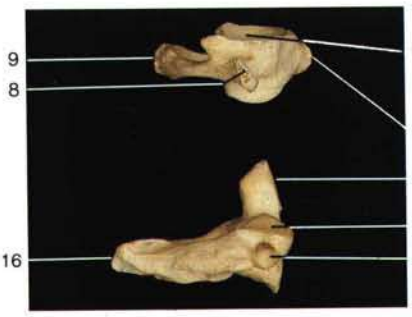

Atlas and axis

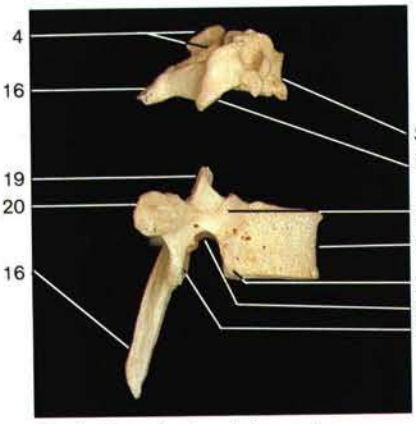

Typical cervical and thoracic vertebrae

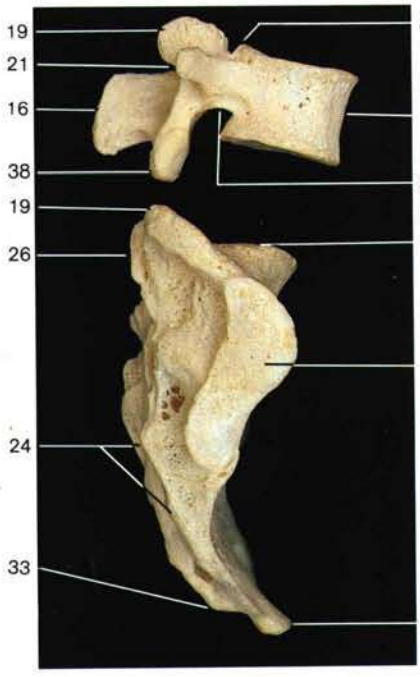

Typical lumbar vertebra and sacrum

Representative vertebrae from each region of the vertebral column (lateral aspect, ventral surface on the right).

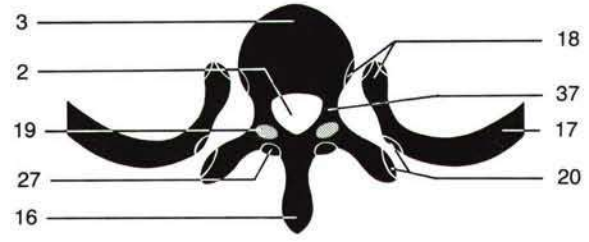

General organization of ribs and vertebrae. (Schematic diagram.)

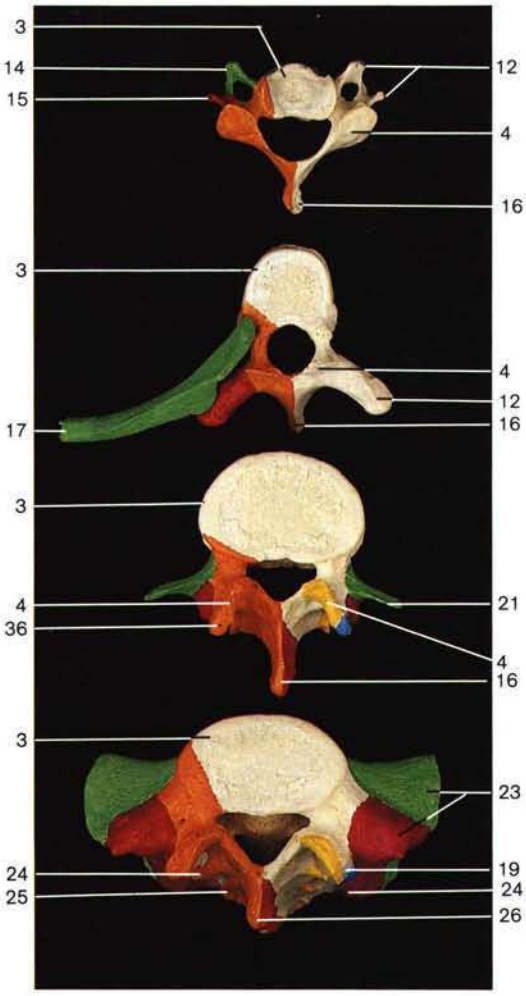

General characteristics of the vertebrae. Typical cervical, thoracic, lumbar vertebrae and sacrum.

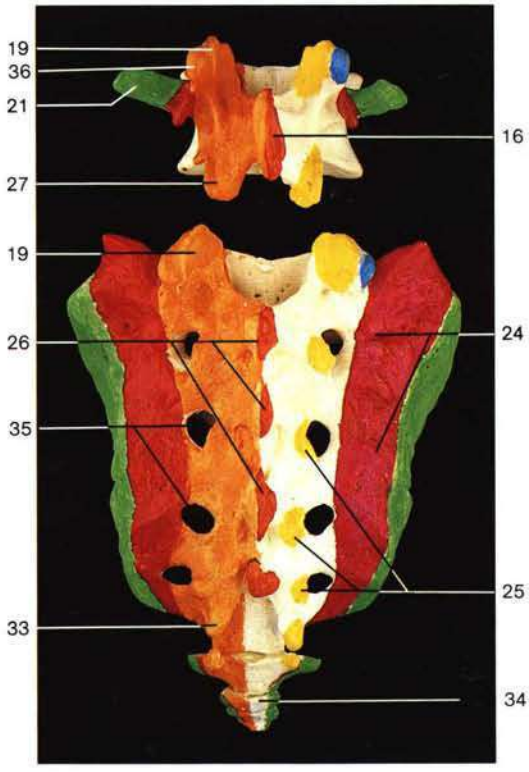

General characteristics of lumbar vertebrae and sacrum (posterior aspect).

1 Foramen transversarium
2 Vertebral foramen
3 Body of vertebra
4 Superior articular facet
5 Base of sacrum
6 Anterior tubercle of atlas
7 Superior articular facet of atlas
8 Transverse process
9 Posterior tubercle of atlas
10 Dens of axis
11 Superior articular surface
12 Transverse process
13 Arch of vertebra
14 Anterior tubercle of transverse process
15 Posterior tubercle of transverse process
16 Spinous process
17 Shaft of rib
18 Body of vertebra and head of rib articulating with each other (costovertebral joint)
19 Superior articular process
20 Transverse process and tubercle of rib articulating with each other (costotransverse joint)
21 Transverse process
22 Auricular surface
23 Lateral part of sacrum
24 Lateral sacral crest
25 Intermediate sacral crest
26 Median sacral crest
27 Inferior articular facet
28 Superior demifacet for head of rib
29 Inferior demifacet for head of rib
30 Inferior vertebral notch
31 Superior vertebral notch
32 Apex of the sacrum
33 Sacral cornu
34 Coccyx
35 Dorsal sacral foramina
36 Mamillary process
37 Pedicle
38 Inferior articular process

Green = Ribs or homologous processes
Red = Muscular processes (transverse and spinous processes)
Orange = Laminae and articular processes
Yellow = Articular facets

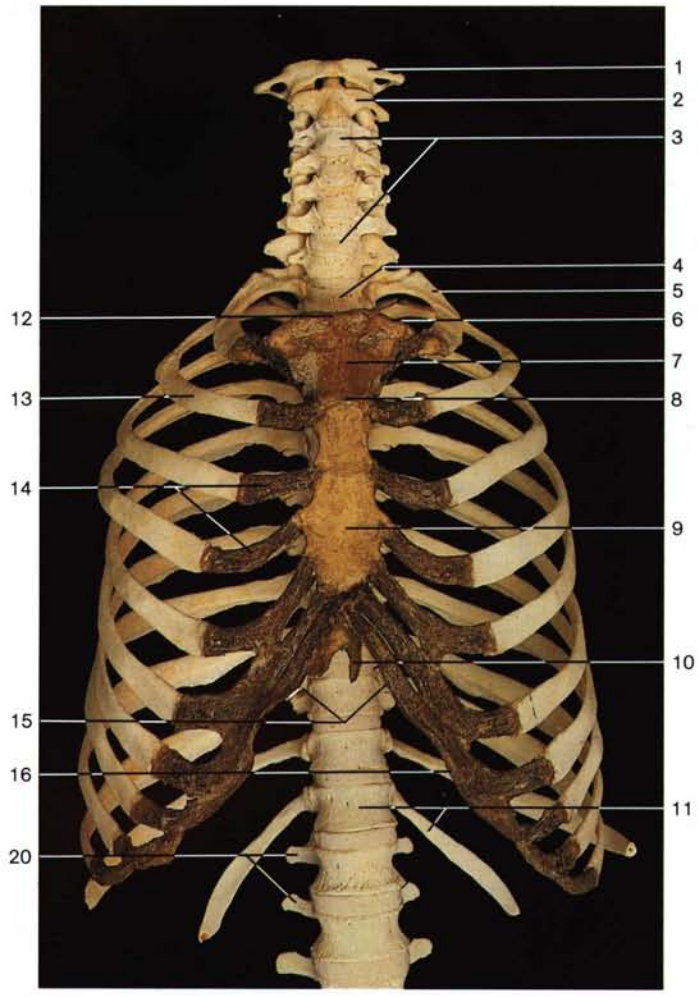

Skeleton of the thorax (anterior aspect).

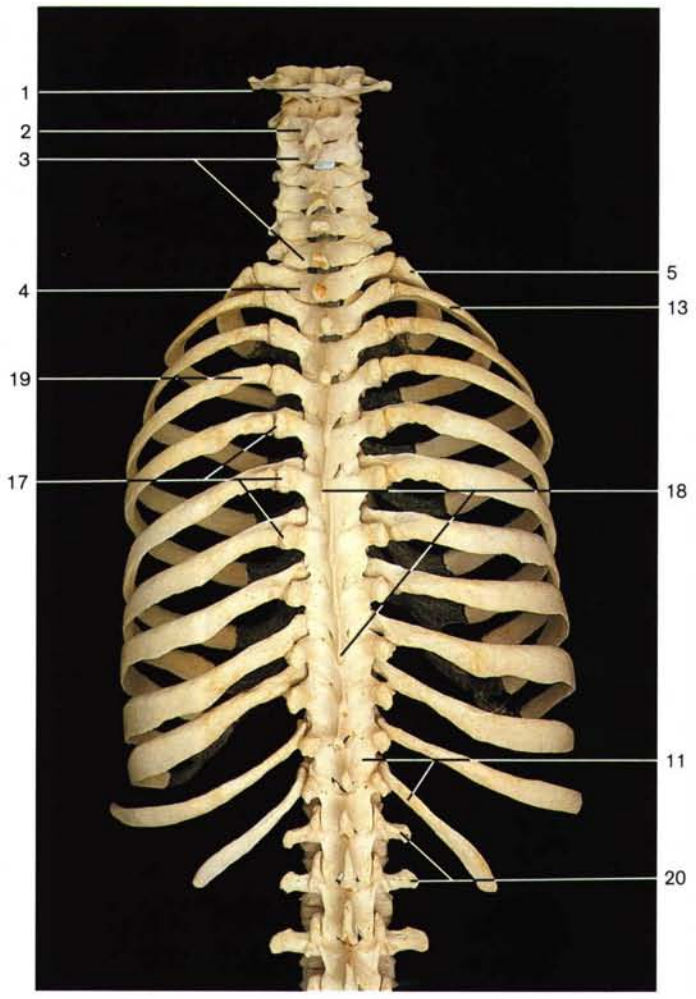

Skeleton of the thorax (posterior aspect).

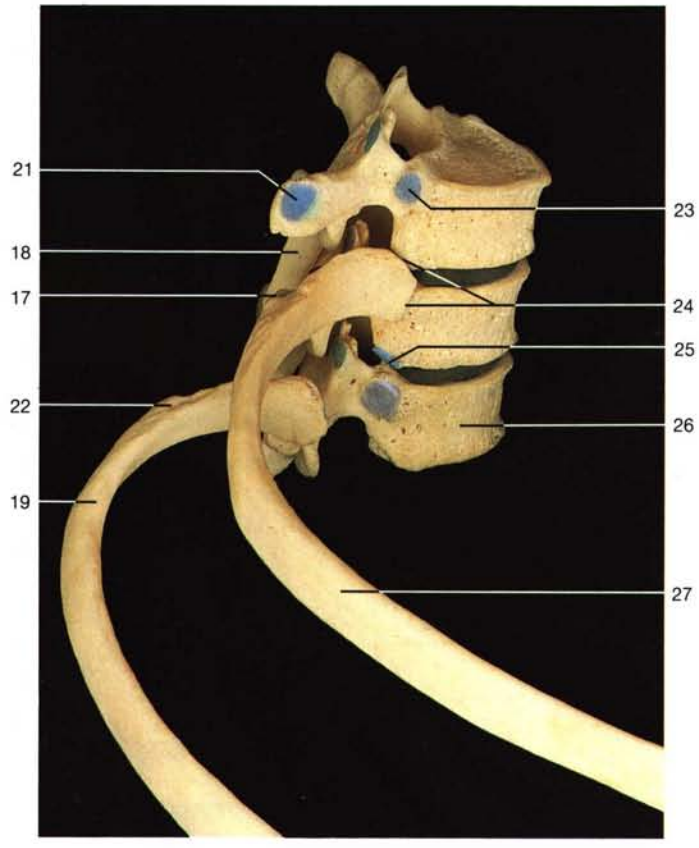

Costovertebral articulation (right lateral aspect).

1 Atlas
2 Axis
3 **Cervical vertebrae**
4 First thoracic vertebra
5 First rib
6 Facet for clavicle and clavicular notch
7 **Manubrium sterni**
8 Sternal angle
9 Body of sternum
10 Xiphoid process
11 Twelfth thoracic vertebra and rib
12 Jugular notch
13 Second rib
14 **Costal cartilages**
15 Infrasternal angle
16 **Costal arch**
17 Costotransverse joints between the transverse processes of thoracic vertebra and the tubercles of the ribs
18 Spinous processes
19 Costal angle
20 Transverse processes of lumbar vertebrae
21 Facet for articulation with rib
22 Tubercle of rib
23 Superior facet for articulation with head of rib
24 Articulation of head of rib with two vertebrae
25 Inferior facet for articulation with head of rib
26 Body of thoracic vertebra
27 Body or shaft of rib

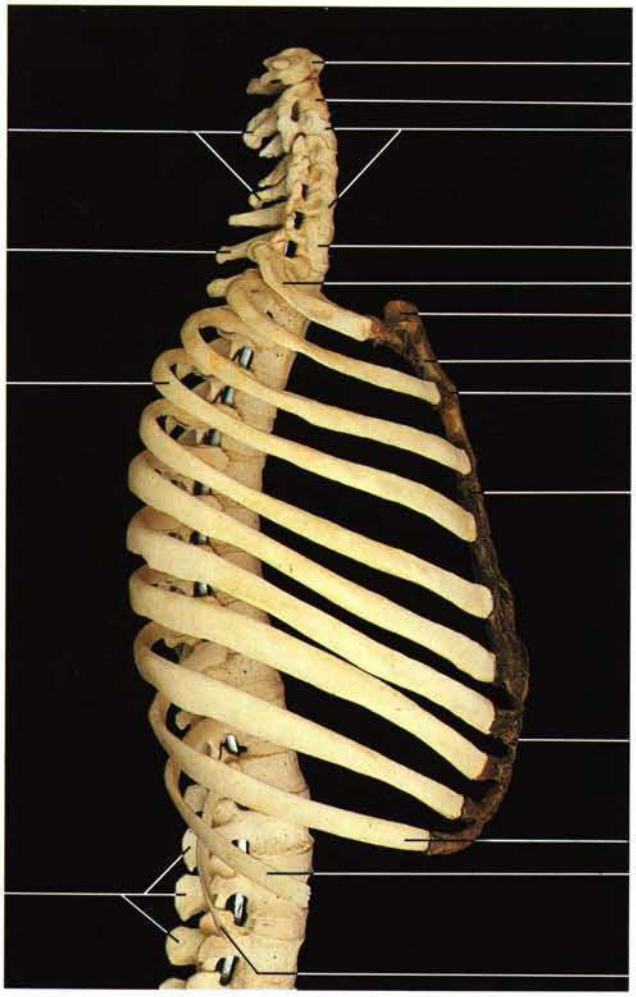

Skeleton of the thorax (right lateral aspect).

1 Atlas
2 Axis
3 **Cervical vertebrae**
4 Seventh cervical vertebra (vertebra prominens)
5 First rib
6 Facet for clavicle
7 **Manubrium sterni**
8 Sternal angle
9 Body of sternum
10 Costal arch
11 Tenth rib
12 Eleventh rib
13 Twelfth rib
14 Spinous processes of cervical vertebrae
15 Spinous processes of thoracic vertebrae
16 Spinous processes of lumbar vertebrae
17 Costal angle
18 Intervertebral foramina
19 Intervertebral discs
20 **Cervical curvature**
21 **Thoracic curvature**
22 **Lumbar curvature**
23 Sacrum
24 Coccyx

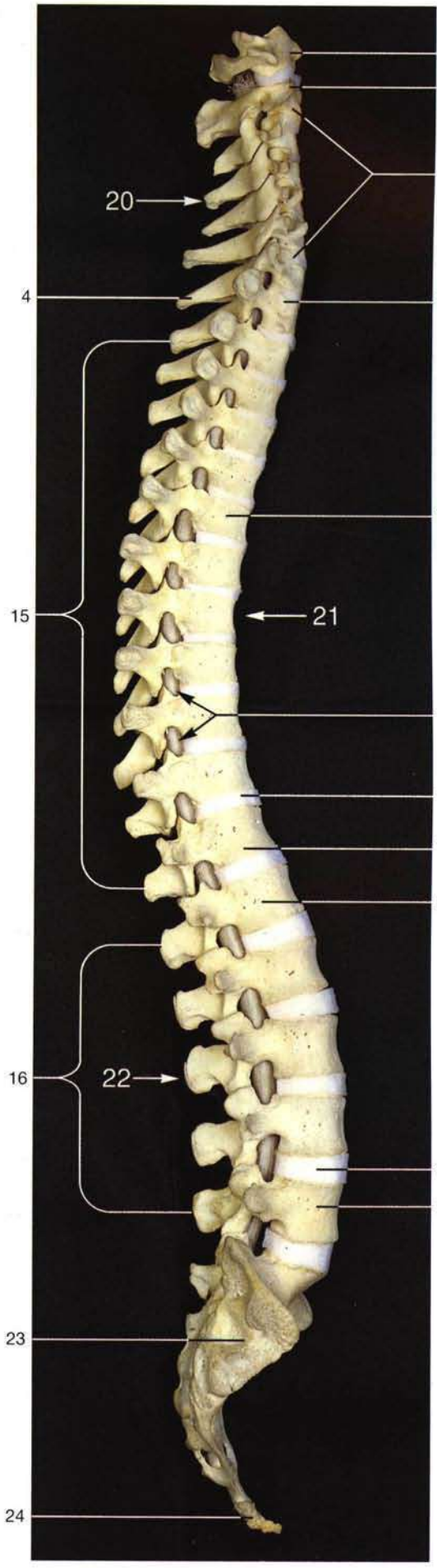

Vertebral column (right lateral aspect).

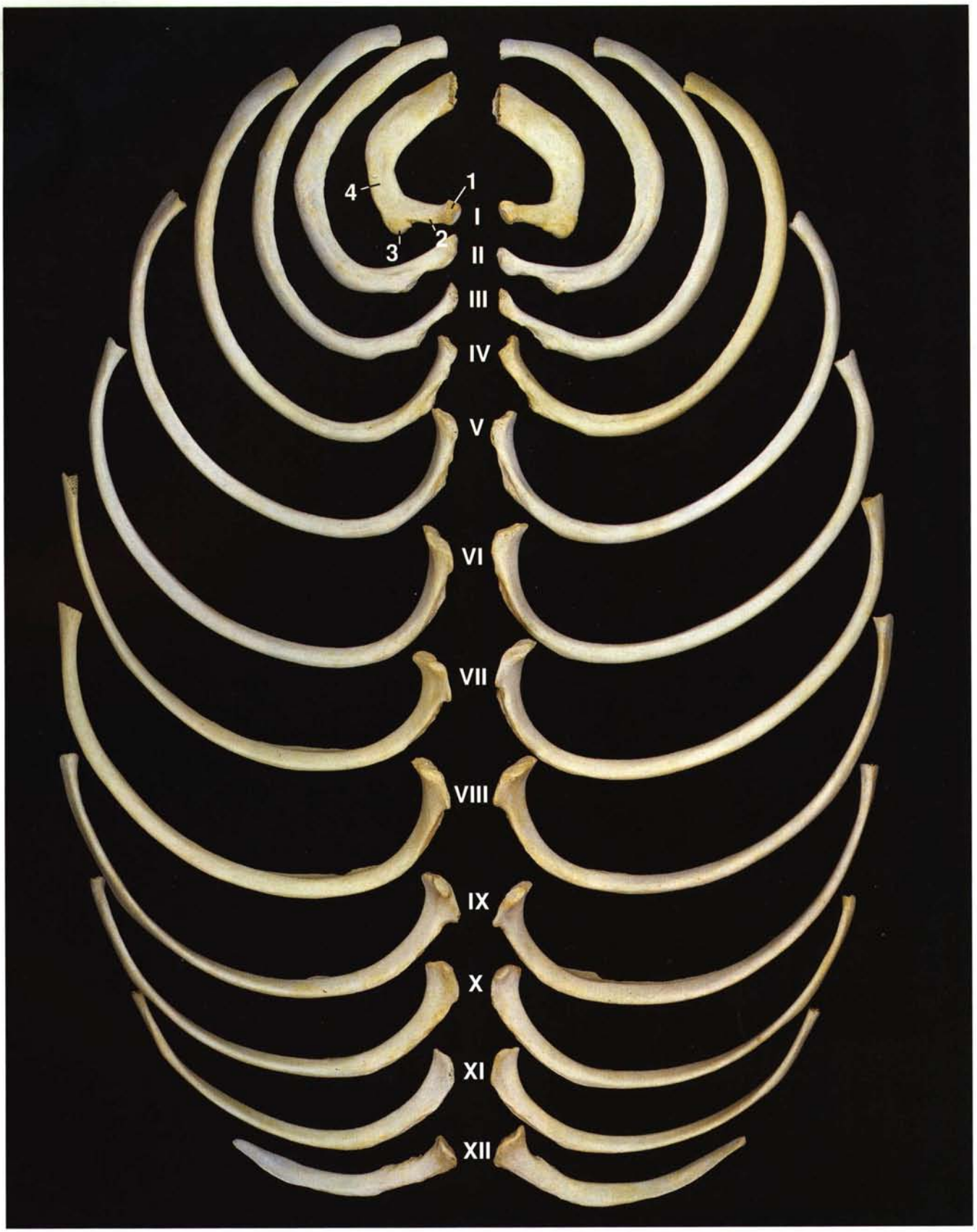

Disarticulated ribs. The twelve ribs arranged in a craniocaudal direction, first rib at top. I–XII = pairs of ribs; 1 = head of rib; 2 = neck of rib; 3 = tubercle of rib; 4 = body of rib.

The head of a typical rib articulates with the bodies of adjacent vertebrae and the intervening intervertebral disc. The first seven ribs are directly connected through the costal cartilages (not depicted) to the sternum (true ribs) while the remaining five ribs are only indirectly connected to the sternum (false ribs, ribs VIII to X) or end freely in the lateral wall of the trunk (floating ribs, ribs XI and XII).

Two **thoracic vertebrae** (left lateral aspect).

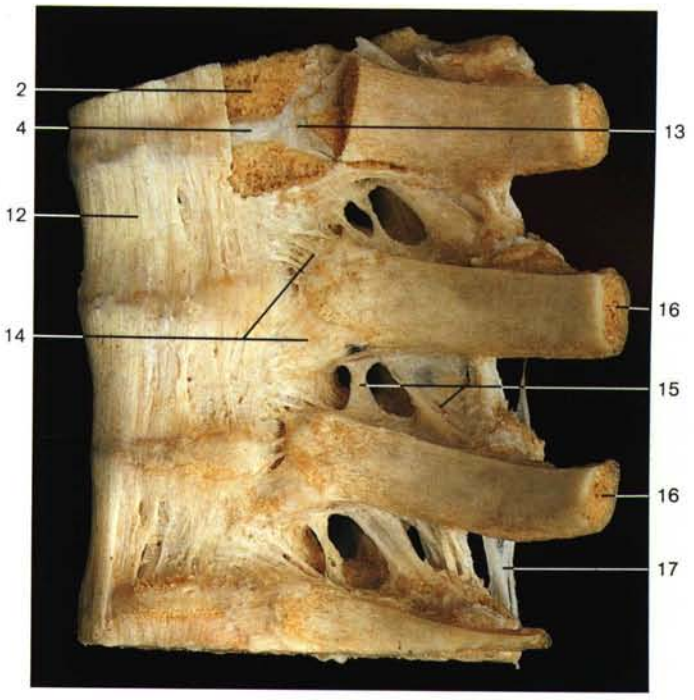

Ligaments of thoracic vertebrae and costovertebral joints (left anterolateral aspect).
In the upper joint, most of the radiate ligament and the anterior part of the head of the rib have been removed to expose the two joint cavities and the interposed intra-articular ligament.

1 Superior demifacet for head of rib
2 Body of vertebra
3 Inferior demifacet for head of rib
4 Intervertebral disc
5 Inferior vertebral notch
6 Superior articular facet and superior articular process
7 Pedicle
8 Transverse process and facet for tubercle of rib
9 Inferior articular process
10 Intervertebral foramen
11 Spinous process
12 Anterior longitudinal ligament
13 Intraarticular ligament
14 Radiate ligament
15 Superior costotransverse ligament
16 Body of rib
17 Intertransverse ligament

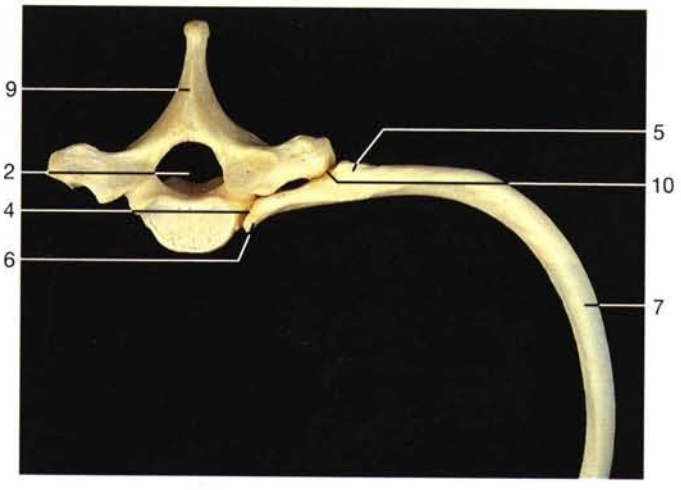

Location of costovertebral joints (superior aspect).

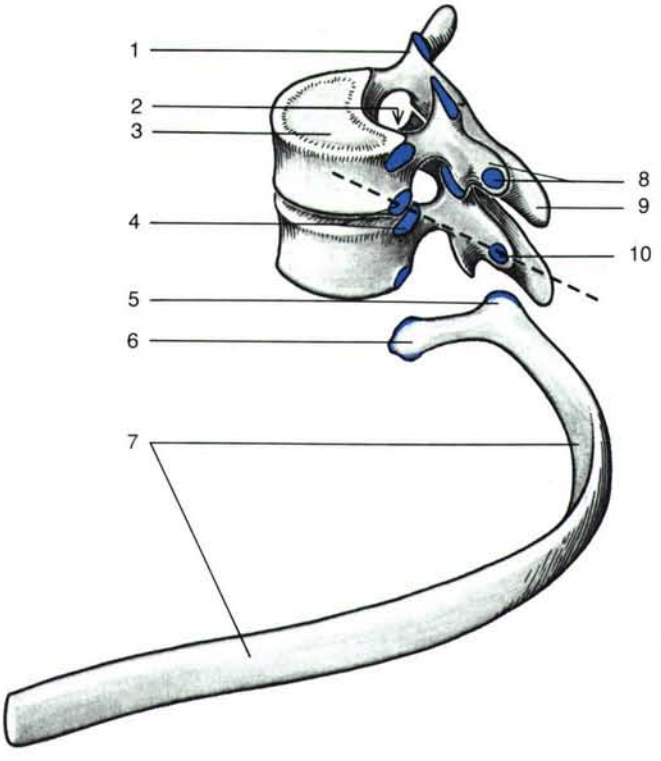

Costovertebral joints. Two thoracic vertebrae with an articulating rib (separated). Axis of movement indicated by dotted line. Blue = articular facets. (Schematic diagram.)

1 Superior articular process
2 Vertebral canal
3 Body of thoracic vertebra
4 Costovertebral joint (articular facets)
5 Tubercle of rib
6 Head of rib
7 Shaft or body of rib
8 Transverse process with articular facet
9 Spinous process
10 Costotransverse joint (articular facets)

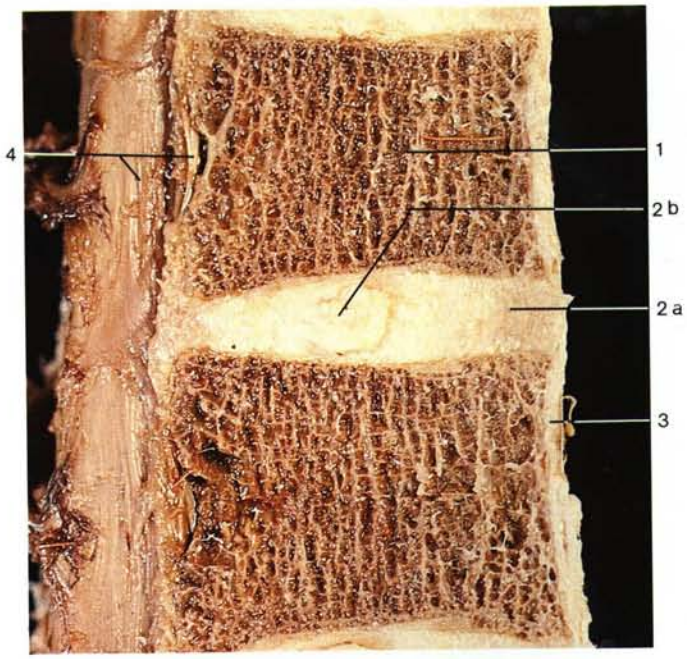

Median-sagittal section of the bodies of the vertebrae showing the **intervertebral discs,** each of which consists of an outer laminated portion and an inner core.

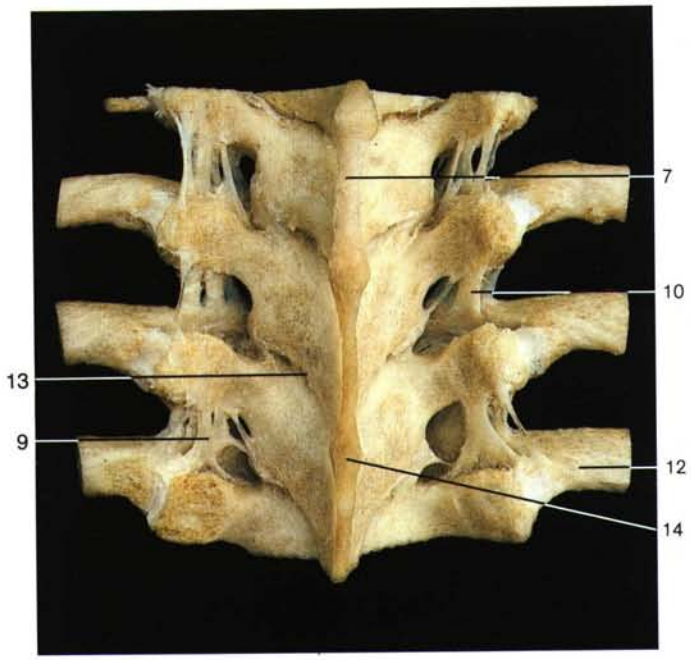

Ligaments of the vertebral column (dorsal aspect).

1 Body of vertebra
2 **Intervertebral disc**
  a Outer portion (anulus fibrosus)
  b Inner core **(nucleus pulposus)**
3 Anterior longitudinal ligament
4 Posterior longitudinal ligament and spinal dura mater
5 Transverse process of lumbar vertebra
6 Sacrum
7 **Supraspinous ligament**
8 Interspinous ligament
9 **Intertransverse ligament**
10 Superior costotransverse ligament
11 Transverse process of thoracic vertebra
12 Rib
13 **Ligamentum flavum**
14 Spinous process
15 Intervertebral foramen

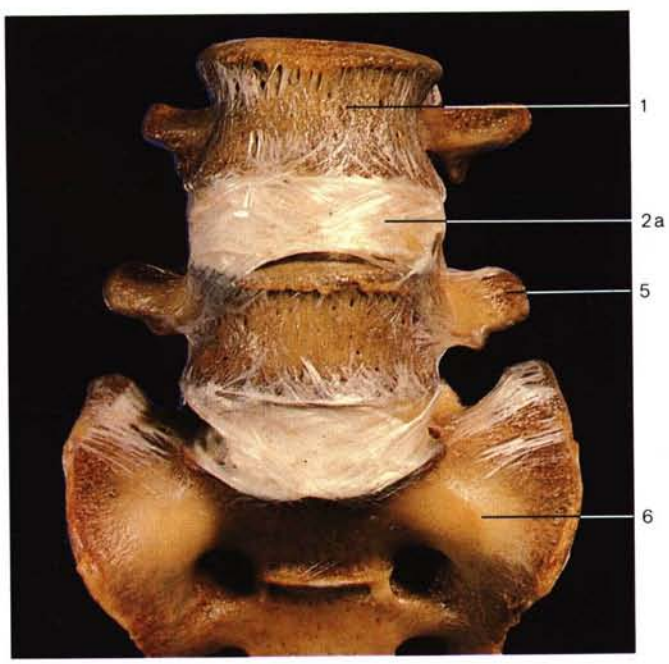

The two caudal lumbar vertebrae and the sacrum with their intervertebral discs (anterior aspect). Anterior longitudinal ligament removed.

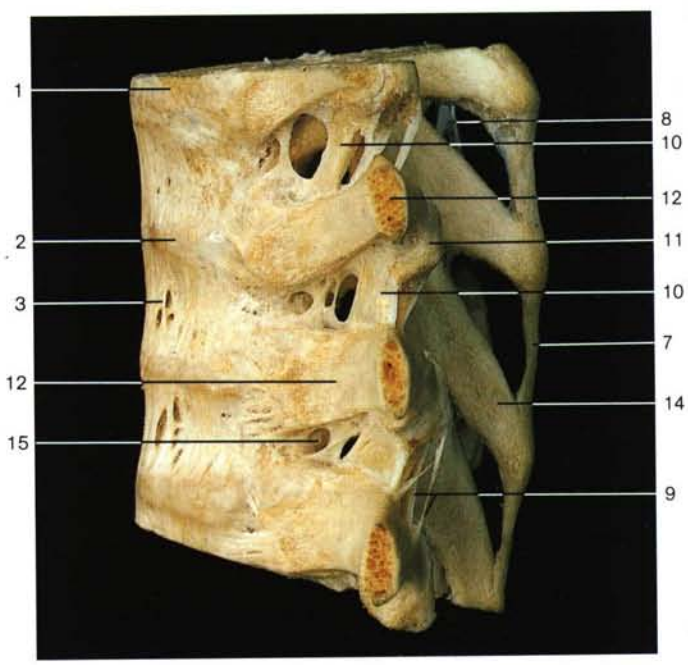

Ligaments of the vertebral column, thoracic part (left lateral aspect).

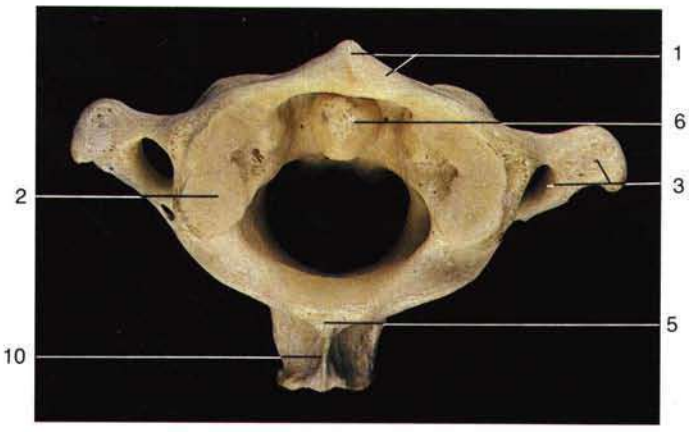

Atlas and axis (from above).

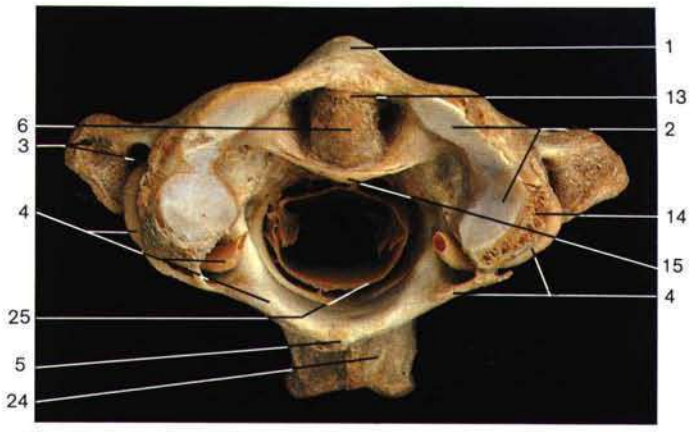

Median atlantoaxial joint and transverse ligament of atlas (from above). Dens of axis partly severed.

1 Anterior arch of **atlas** with anterior tubercle
2 Superior articular facet of atlas
3 **Foramen transversarium** and transverse process
4 Posterior arch of atlas and vertebral artery
5 Posterior tubercle of atlas
6 **Dens of axis**
7 Superior articular surface of axis
8 Body of axis
9 Pedicle and lamina of axis
10 Spinous process
11 Inferior articular process
12 Transverse process and vertebrarterial foramen (foramen transversarium)
13 **Median atlantoaxial joint** (anterior part)
14 Articular capsule of **atlantooccipital joint**
15 Transverse ligament of atlas
16 Occipital bone
17 **Atlantooccipital joint**
18 **Lateral atlantoaxial joint**
19 Third cervical vertebra
20 Superior longitudinal band of cruciform ligament
21 Alar ligaments
22 Transverse ligament of atlas
23 Inferior longitudinal band of cruciform ligament
24 Spinous process of axis
25 Dura mater

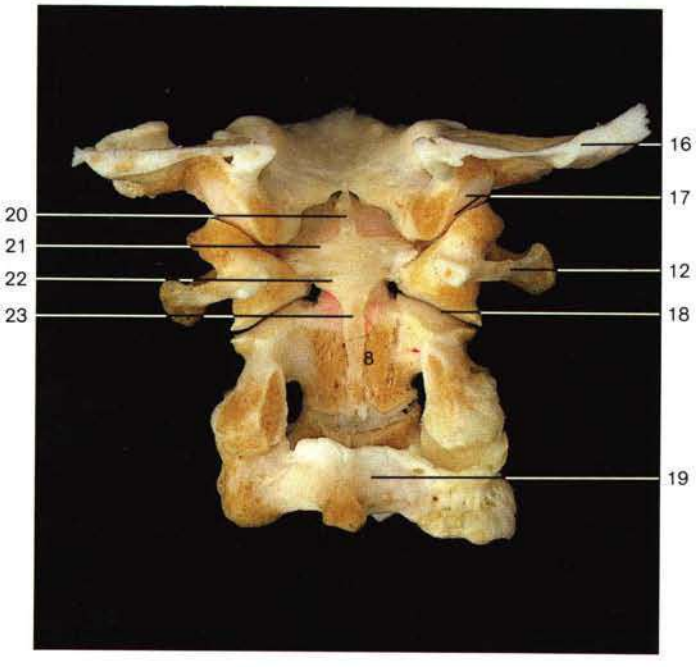

Atlantooccipital and atlantoaxial joints (posterior aspect). Posterior part of occipital bone, posterior arch of atlas and axis have been removed to show the cruciform ligament.

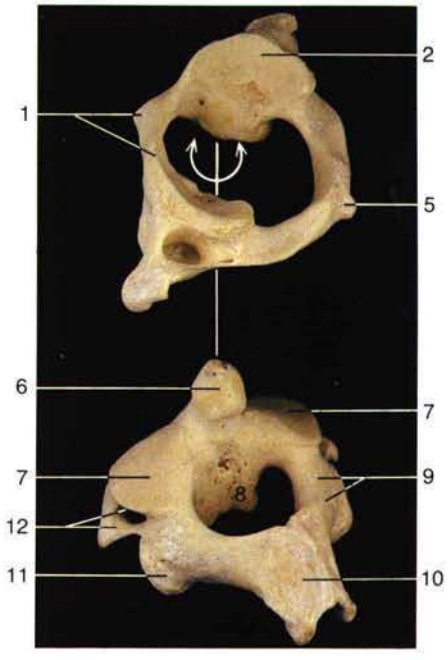

Atlas and axis. Left oblique posterolateral aspect, demonstrating the articulation of the dens of axis with atlas (cf. arrows).

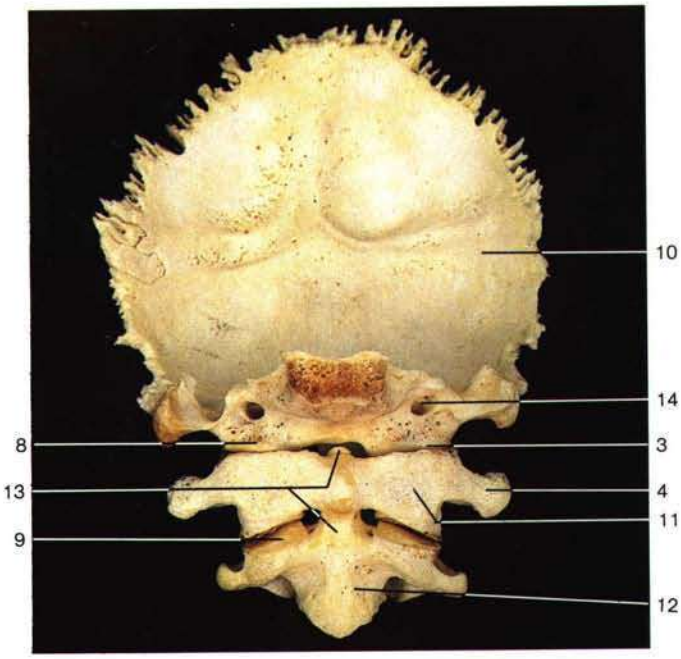

Occipital bone, atlas and axis (anterior aspect).

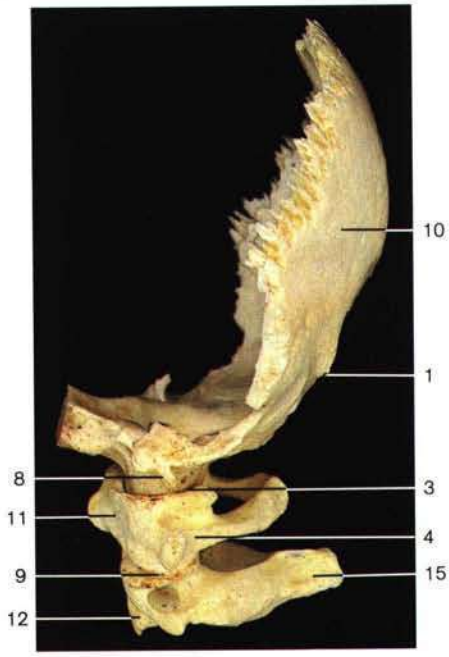

Occipital bone, atlas and axis (left lateral aspect).

1 External occipital protuberance
2 Foramen magnum
3 Atlantooccipital joint
4 Transverse process of atlas
5 Membrana tectoria
6 Posterior longitudinal ligament
7 Spinous process of third cervical vertebra
8 Occipital condyle
9 Lateral atlantoaxial joint
10 Occipital bone
11 Atlas
12 Axis
13 Dens of axis
14 Hypoglossal canal
15 Spinous processes of axis
16 Anterior longitudinal ligament

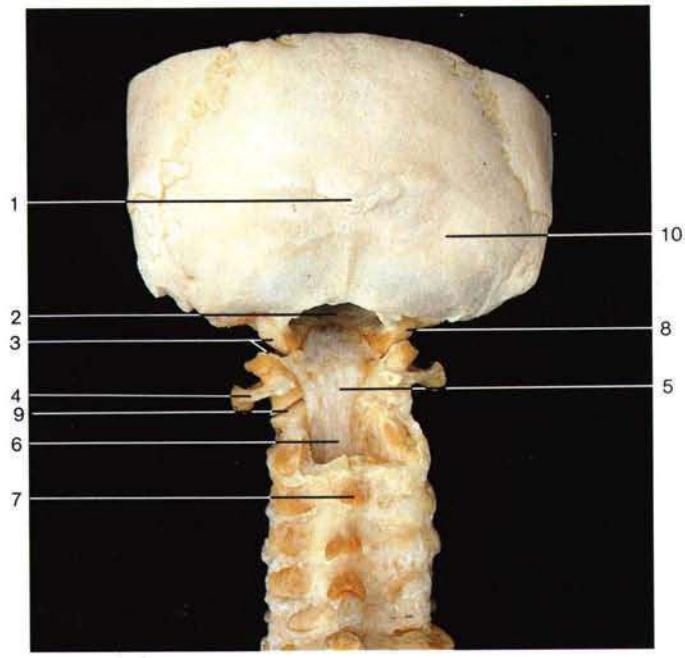

Cervical vertebral column and skull with ligaments (posterior aspect). Posterior arch of atlas and axis removed to show the membrana tectoria.

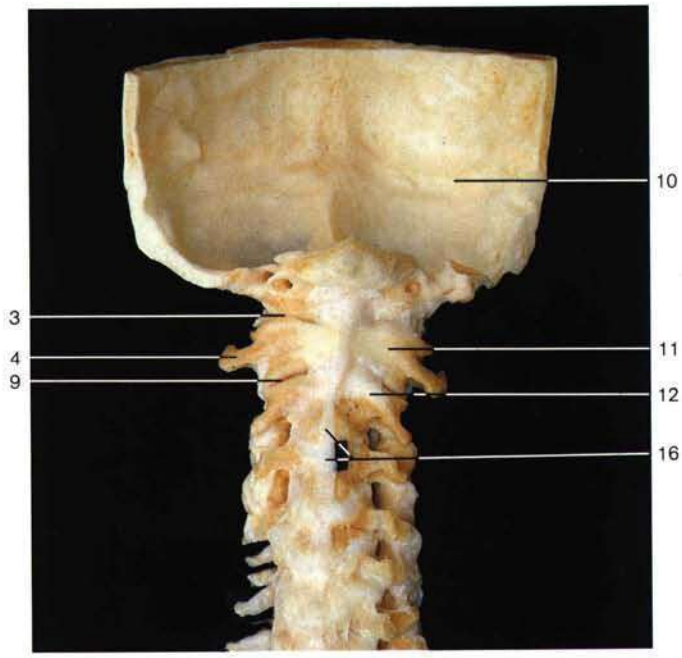

Cervical vertebral column and skull with ligaments (anterior aspect). Anterior part of occipital bone removed.

1 Axillary vein
2 Intercostobrachial nerves
3 Subscapularis muscle and thoracodorsal nerve
4 Long thoracic nerve, lateral thoracic artery and vein
5 Latissimus dorsi muscle
6 **External intercostal** muscles
7 **Serratus anterior** muscle
8 Lateral cutaneous branches of intercostal nerves
9 External abdominal oblique muscle
10 Clavicle (divided)
11 Second rib (costo-chondral junction)
12 **Internal intercostal muscles**
13 External intercostal membrane
14 Position of xiphoid process
15 Costal arch or margin
16 Anterior layer of rectus sheath

Muscles of the thorax, superficial layer (lateral aspect). Upper limb elevated. Pectoralis major and minor muscles have been removed.

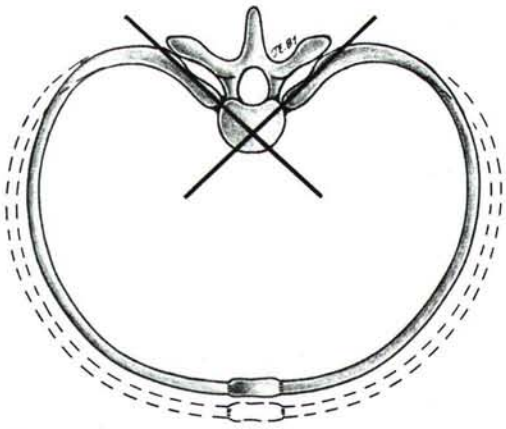

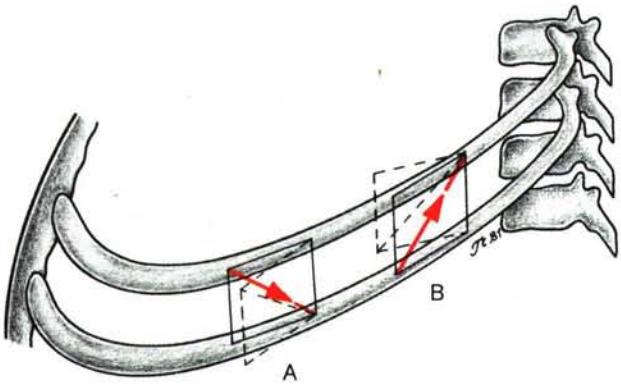

Effect of intercostal muscles on the costovertebral and costotransverse joints. Axes of movement indicated by lines; direction of movements indicated by arrows.

A Action of internal intercostal muscles (expiration)
B Action of external intercostal muscles (inspiration)

Anterior thoracic wall (posterior aspect). Diaphragm partly removed, posterior layer of rectus sheath fenestrated on both sides.

1 Sternocleidomastoid muscle (divided)
2 Clavicle
3 Sternothyroid muscle
4 **Internal intercostal muscle**
5 **Transversus thoracis muscle**
6 Intercostal arteries and nerves
7 Musculophrenic artery
8 Superior epigastric artery and vein
9 Diaphragm (divided)
10 Rectus abdominis muscle
11 Subclavian artery and brachial plexus
12 First rib
13 Internal thoracic artery and vein
14 Sternum
15 Innermost intercostal muscle
16 Intercostal artery and vein
17 Xiphoid process
18 Linea alba and posterior layer of rectus sheath
19 **Transversus abdominis muscle**

Superficial muscles of the anterior thoracic and abdominal wall. The fascia of pectoralis major muscle and the abdominal wall have been removed; the anterior layer of the sheath of the rectus abdominis muscle is displayed.

1 Sternohyoid muscle
1 Sternocleidomastoid muscle
3 Supraclavicular nerves (branches of cervical plexus)
4 Deltoid muscle
5 **Pectoralis major muscle**
6 **Anterior cutaneous branches of intercostal nerves**
7 External abdominal oblique muscle
8 **Lateral cutaneous branches of intercostal nerves**
9 Umbilicus and umbilical ring
10 Clavicle
11 Cephalic vein
12 Serratus anterior muscle
13 Linea alba
14 **Sheath of rectus abdominis** muscle (anterior layer)
15 Inguinal ligament

Thoracic wall I (anterior aspect). Left pectoralis major muscle has been divided and reflected. Note the connection of the cephalic vein with the subclavian vein. Arrow: medial pectoral nerve.

1 Mandible
2 Facial artery
3 Submandibular gland
4 Hyoid bone
5 Thyroid cartilage and sternohyoid muscle
6 Clavicle
7 Subclavius muscle
8 Second rib
9 Anterior cutaneous branches of intercostal nerves
10 External intercostal membrane
11 Parotid gland
12 External carotid artery
13 Sternocleidomastoid muscle and cutaneous branches of cervical plexus
14 Supraclavicular nerves
15 Pectoralis major muscle and lateral pectoral nerves
16 Thoracoacromial artery and subclavian vein
17 **Pectoralis minor muscle**
18 **Median and ulnar nerve**
19 Thoracoepigastric vein
20 Cephalic vein and long head of biceps brachii muscle
21 **Lateral thoracic artery and long thoracic nerve**
22 Lateral cutaneous branches of intercostal nerve
23 Latissimus dorsi muscle
24 Median nerve
25 Axillary artery
26 Intercostobrachial nerves
27 **Thoracodorsal nerve**
28 Long thoracic nerve
29 Latissimus dorsi muscle
30 **Serratus anterior muscle**
31 Thoracoacromial artery
32 Clavicle
33 External intercostal muscle
34 Third rib
35 **Internal intercostal muscle**
36 Anterior intercostal artery and vein
37 Costal arch or margin

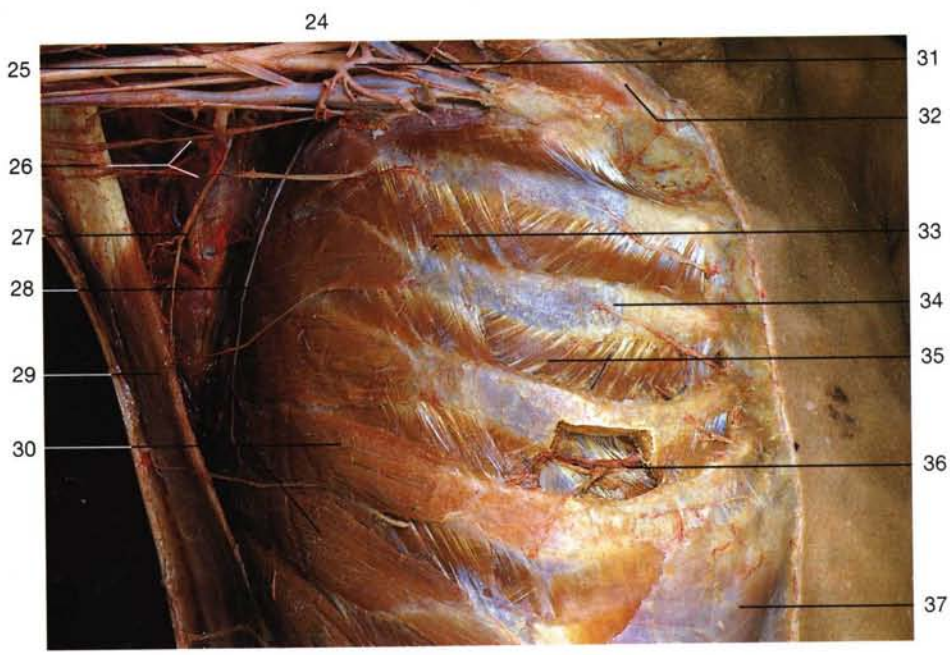

Thoracic wall (lateral aspect). Pectoralis major and minor muscles have been removed. A section of the 4th rib has been cut and removed to display the intercostal vessels and nerve.

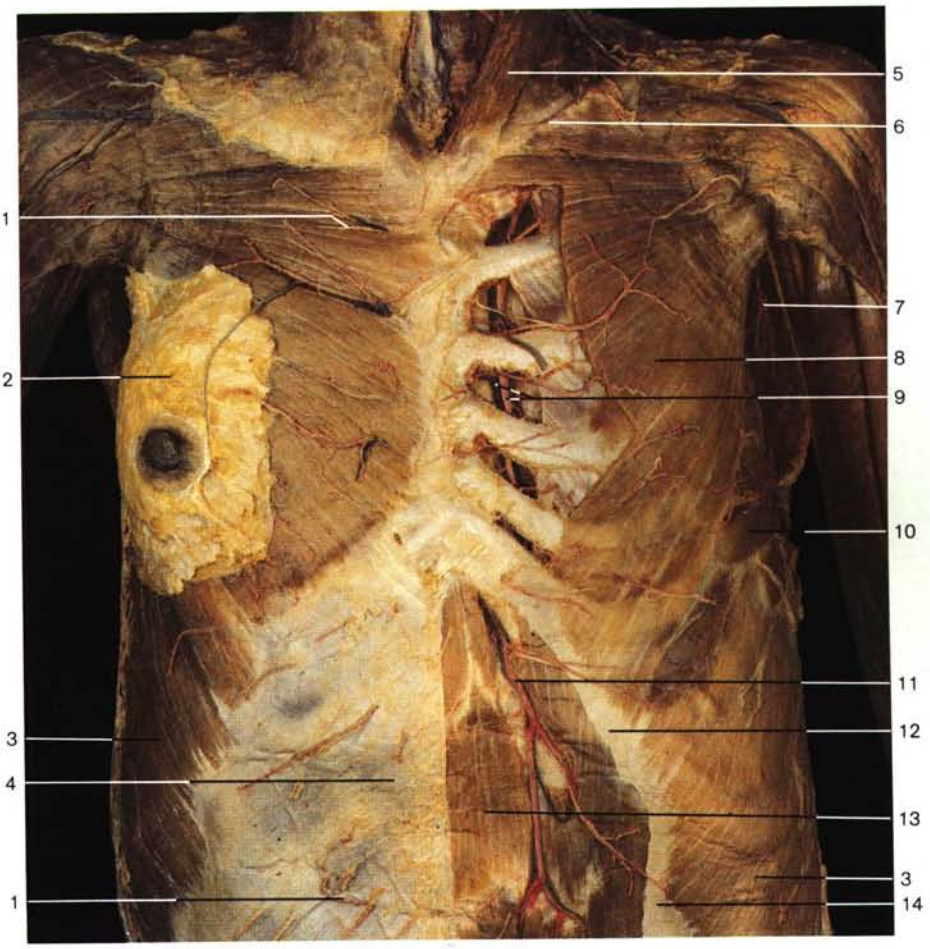

1 Anterior perforating branches of intercostal nerve
2 Mammary gland
3 External abdominal oblique muscle
4 Rectus sheath (anterior layer)
5 Sternocleidomastoid muscle
6 Clavicle
7 Lateral thoracic artery and vein
8 Pectoralis major muscle
9 **Internal thoracic artery and vein**
10 Serratus anterior muscle
11 **Superior epigastric artery and vein**
12 Costal margin
13 Rectus abdominis muscle
14 Cut edge of the anterior layer of the rectus sheath
15 Subclavian artery
16 Highest intercostal artery
17 Internal thoracic artery
18 Musculophrenic artery
19 Superficial epigastric artery
20 Deep circumflex iliac artery
21 Superior epigastric artery
22 Inferior epigastric artery
23 Superficial circumflex iliac artery

Thoracic wall II (anterior aspect). Dissection of the **internal thoracic artery and vein.** Left pectoralis major muscle partly removed. Anterior lamina of the rectus sheath on the left side has been removed.

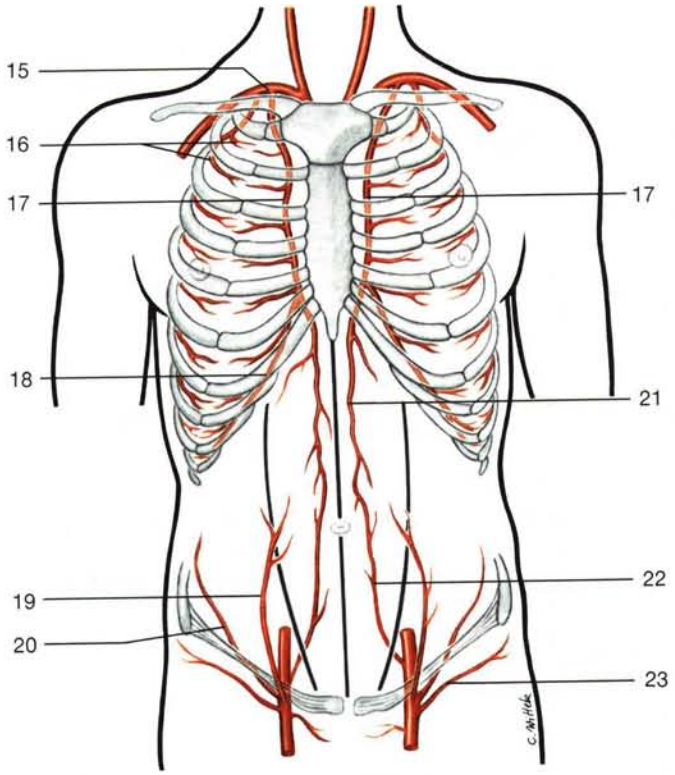

Main arteries of thoracic and abdominal wall.

1 Deltoid muscle
2 Cephalic vein
3 Pectoralis major muscle (divided)
4 **Internal intercostal muscle**
5 **Intercostal artery and vein** (intercostal space, fenestrated)
6 Serratus anterior muscle
7 **External abdominal oblique muscle**
8 **Anterior layer of rectus sheath**
9 Iliac crest
10 Superficial epigastric vein
11 Superficial circumflex iliac vein
12 Saphenous opening
13 Superficial inguinal lymph nodes
14 Superficial external pudendal veins
15 Great saphenous vein
16 Nipple
17 Costal margin
18 Subcutaneous fatty tissue
19 Umbilicus
20 Anterior layer of rectus sheath
21 Rectus abdominis muscle
22 Posterior layer of rectus sheath
23 **Internal abdominal oblique muscle**
24 **External abdominal oblique muscle**
25 **Transversus abdominis muscle**
26 Transversal fascia and peritoneum
27 Psoas major muscle
28 Body of lumbar vertebra (L_4)
29 Quadratus lumborum muscle
30 Medial tract of erector spinae muscle
31 Lateral tract of erector spinae muscle (longissimus and iliocostalis muscles)
32 Small intestine
33 Left ureter
34 **Abdominal aorta**
35 **Inferior vena cava**
36 Descending colon
37 Spinous process

Thoracic and abdominal wall I. Right pectoralis major and minor muscles are divided. Muscles of thoracic and abdominal wall on right side are displayed.

Horizontal section of the trunk at the level of the umbilicus, superior to arcuate line (inferior aspect).

1 Deltoid muscle
2 Pectoralis major muscle (divided)
3 **Internal intercostal muscle**
4 Intercostal artery and vein
5 **Rectus abdominis muscle**
6 Tendinous intersections
7 **External abdominal oblique muscle**
8 Anterior superior iliac spine
9 Superficial circumflex iliac vein
10 Superficial epigastric vein
11 Great saphenous vein
12 Cephalic vein
13 **Pectoralis major muscle**
14 Anterior cutaneous branches of intercostal nerves
15 Nipple
16 Linea alba
17 **Anterior layer of rectus sheath**
18 Umbilicus
19 **Inguinal ligament**
20 Pyramidalis muscle
21 Superficial inguinal ring and spermatic cord
22 Suspensory ligament of penis
23 Longissimus and iliocostalis muscles
24 Multifidus muscle
25 Quadratus lumborum muscle
26 Latissimus dorsi muscle
27 Psoas major muscle
28 Spinous process
29 Body of 1st lumbar vertebra
30 **Transversus abdominis muscle**
31 **Internal abdominal oblique muscle**

Thoracic and abdominal wall II. Right pectoralis major and minor muscles and anterior layer of rectus sheath have been removed on the right side.

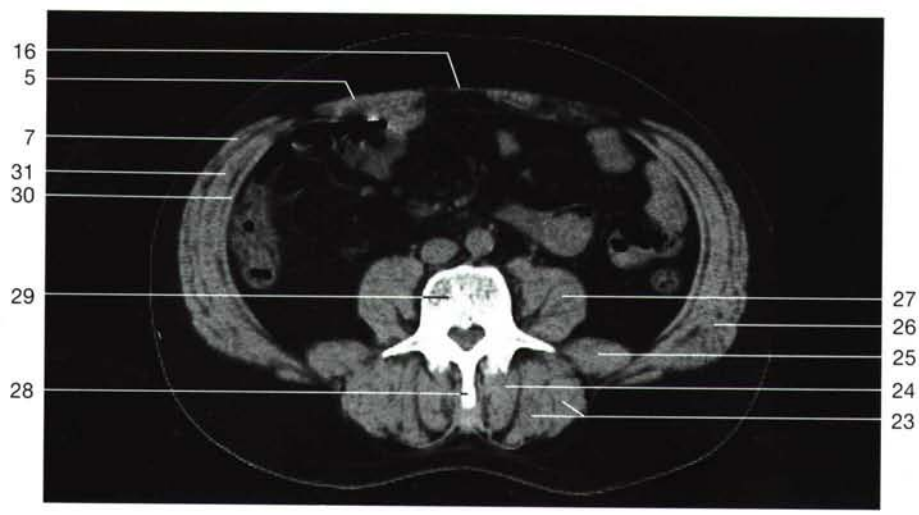

Horizontal section through the body at the level of 4th lumbar vertebra; seen from below. (CT-Scan.)

1 Costal margin
2 **Rectus abdominis muscle**
3 External abdominal oblique muscle (reflected)
4 Thoracoabdominal (intercostal) nerves with accompanying vessels
5 Internal abdominal oblique muscle
6 **Arcuate line** (arrow)
7 **Inferior epigastric artery and vein**
8 Ilioinguinal nerve
9 Position of deep inguinal ring
10 Superficial inguinal lymph nodes
11 Great saphenous vein
12 Linea alba
13 Iliohypogastric nerve
14 **Pyramidalis** muscle
15 Spermatic cord
16 Fundiform ligament of penis

Thoracic and abdominal wall III. External abdominal oblique muscle has been divided and reflected on both sides. The right rectus muscle has been reflected medially to display the posterior layer of rectus sheath. Arrow: location of arcuate line.

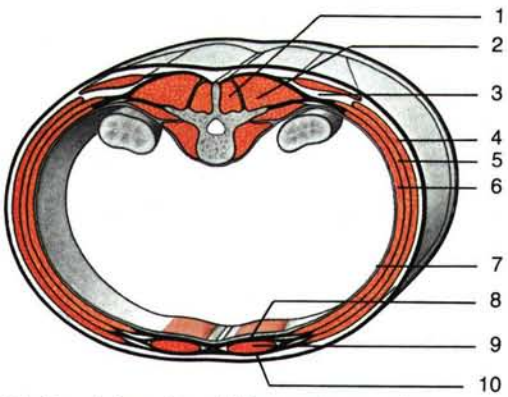

Horizontal-section of the trunk superior to arcuate line. (Schematic drawing.)

1 Medial column of intrinsic muscles of the back
2 Lateral column of erector spinae muscle
3 Thoracolumbar fascia with superficial and deep layer
4 External oblique muscle
5 Internal oblique muscle
6 Transversus abdominis muscle
7 Transversalis fascia
8 Posterior layer of rectus sheath
9 Rectus abdominis muscle
10 Anterior layer of rectus sheath

1 **Rectus abdominis muscle** (reflected)
2 External abdominal oblique muscle (divided)
3 Posterior layer of rectus sheath
4 Umbilical ring
5 **Internal abdominal oblique muscle**
6 Arcuate line (arrow)
7 Inguinal ligament
8 **Inferior epigastric artery and vein** and rectus abdominis muscle (divided and reflected)
9 Costal margin
10 Linea alba
11 Tendinous intersection
12 Iliohypogastric nerve
13 **Ilioinguinal nerve**
14 **Pyramidalis** muscle
15 Spermatic cord

Thoracic and abdominal wall IV. External abdominal oblique muscle has been divided and reflected on both sides. The right rectus muscle has been cut and reflected to display the posterior layer of rectus sheath. Arrow: location of arcuate line.

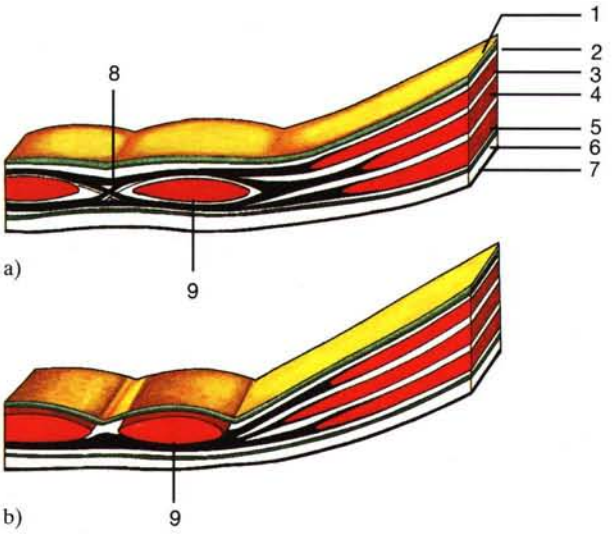

1 Peritoneum
2 Transversalis fascia (green)
3 Transversus abdominis muscle
4 Internal abdominal oblique muscle
5 External abdominal oblique muscle
6 Fascia of external abdominal oblique muscle (green)
7 Skin
8 Linea alba
9 Rectus abdominis muscle

Transverse sections through the abdominal wall superior (a) and inferior (b) to arcuate line.

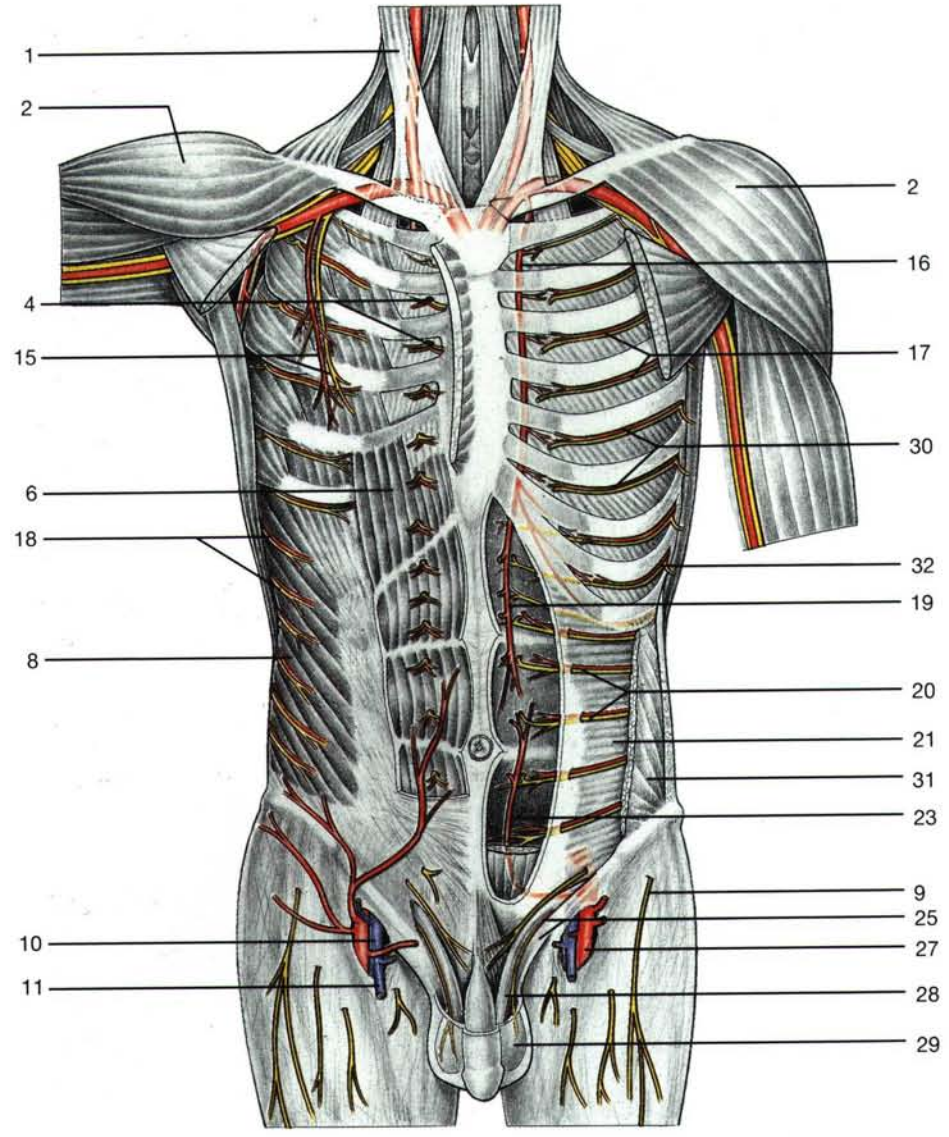

Thoracic and abdominal wall (schematic drawing). Note the segmental organization of the blood vessels and nerves. Right side: superficial layers; left side: deeper layers.

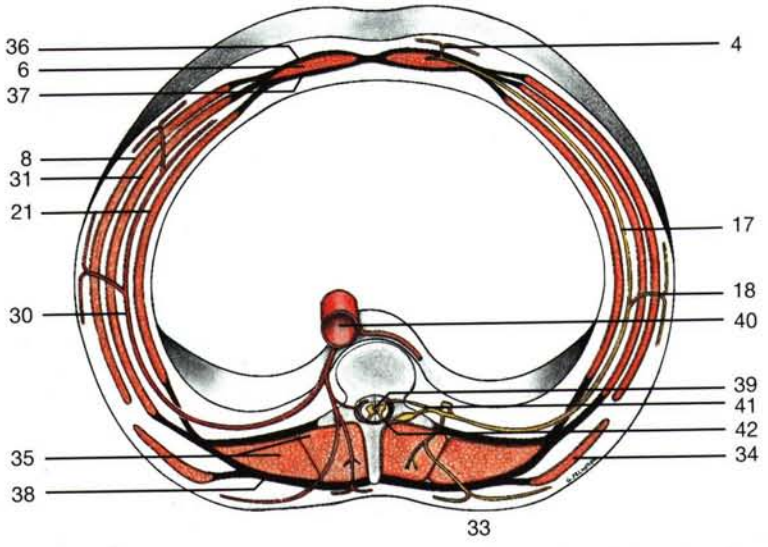

Horizontal section of the abdominal wall (from above) showing the location of the intercostal arteries (left side) and nerves (right side).

1 Sternocleidomastoid muscle
2 Deltoid muscle
3 Pectoralis major muscle
4 **Anterior cutaneous branches of intercostal nerves**
5 Cut edge of anterior layer of rectus sheath
6 **Rectus abdominis muscle**
7 Tendinous intersection
8 External abdominal oblique muscle
9 Lateral femoral cutaneous nerve
10 Femoral vein
11 Great saphenous vein
12 Medial supraclavicular nerves
13 Pectoralis minor muscle (reflected) and medial pectoral nerves
14 Axillary vein
15 Long thoracic nerve and lateral thoracic artery
16 Internal thoracic artery
17 **Intercostal nerves**
18 **Lateral cutaneous branches of intercostal nerves**
19 Superior epigastric artery
20 Thoracoabdominal (intercostal) nerves
21 Transversus abdominis muscle
22 Posterior layer of rectus sheath
23 **Inferior epigastric artery**
24 Lateral femoral cutaneous nerve
25 Inguinal ligament and ilioinguinal nerve
26 Femoral nerve
27 Femoral artery
28 Spermatic cord
29 Testis
30 **Posterior intercostal arteries**
31 Internal abdominal oblique muscle
32 Lateral cutaneous branch of intercostal nerve
33 Dorsal branch of spinal nerve
34 Latissimus dorsi muscle
35 Deep muscles of the back (medial and lateral tract)
36 Anterior layer of rectus sheath
37 Posterior layer of rectus sheath
38 Thoracolumbar fascia
39 Spinal cord
40 Aorta
41 Ventral root } of spinal
42 Dorsal root } nerve

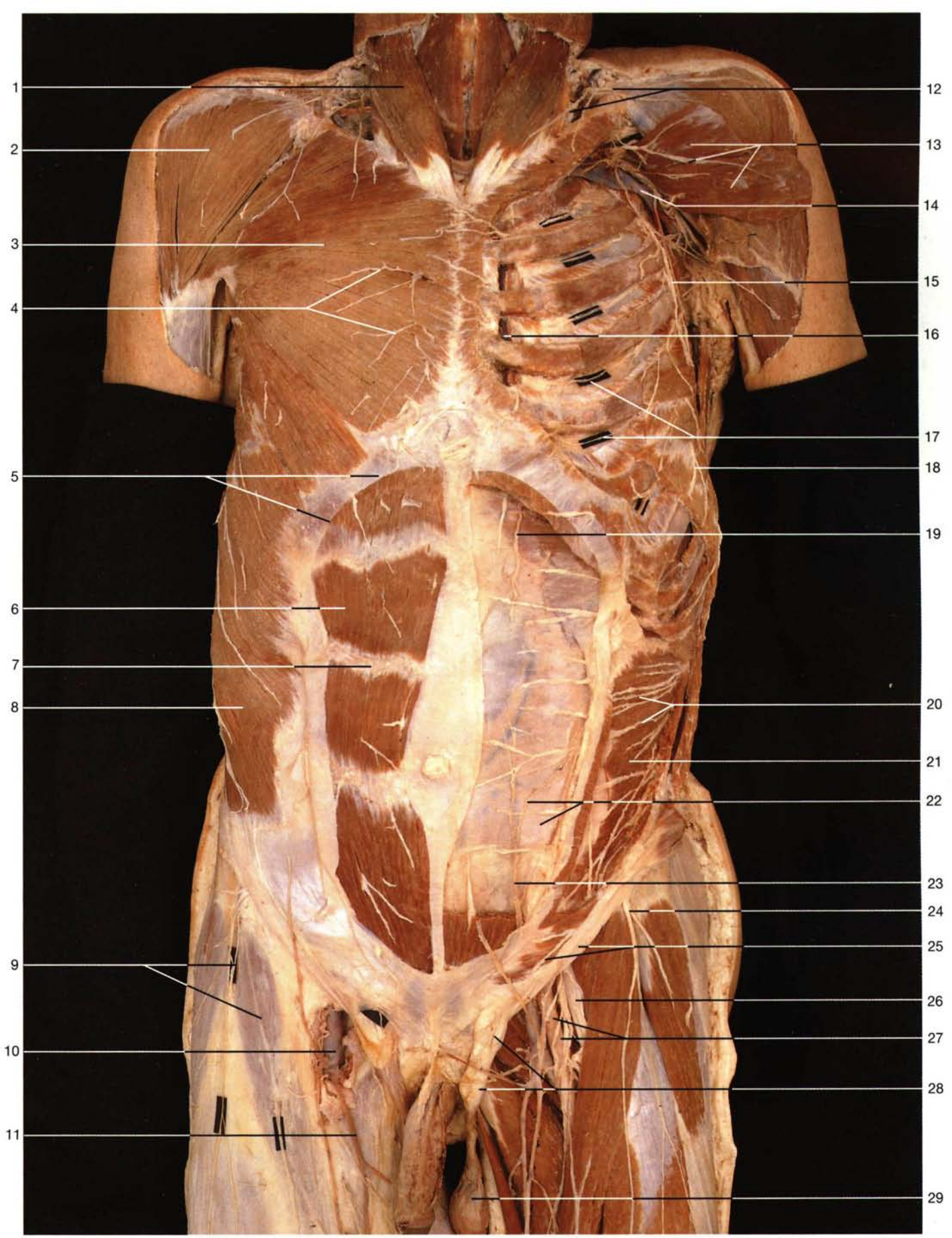

Thoracic wall and abdominal wall V. Right side: superficial layers; left side: deeper layers (anterior aspect). Pectoralis major and minor muscles, the external and internal intercostal muscles on the left side have been removed to display the intercostal nerves. The anterior layer of rectus sheath, the left rectus abdominis muscle and the external and internal abdominal oblique muscle have been removed to show the thoracoabdominal nerves within the abdominal wall.

1 **Rectus abdominis muscle**
2 Tendinous intersection
3 **Internal abdominal oblique muscle**
4 External abdominal oblique muscle (reflected)
5 Anterior superior iliac spine
6 Ilioinguinal nerve
7 Spermatic cord
8 Costal margin
9 Superior epigastric artery
10 **Thoracoabdominal** (intercostal) **nerves**
11 Posterior layer of rectus sheath
12 **Transversus abdominis muscle**
13 Semilunar line
14 Arcuate line
15 **Inferior epigastric artery**
16 Inguinal ligament

Abdominal wall with vessels and nerves. The left rectus abdominis muscle has been divided and reflected to display the inferior epigastric vessels. The left internal abdominal oblique muscle has been removed to show the thoracoabdominal nerves.

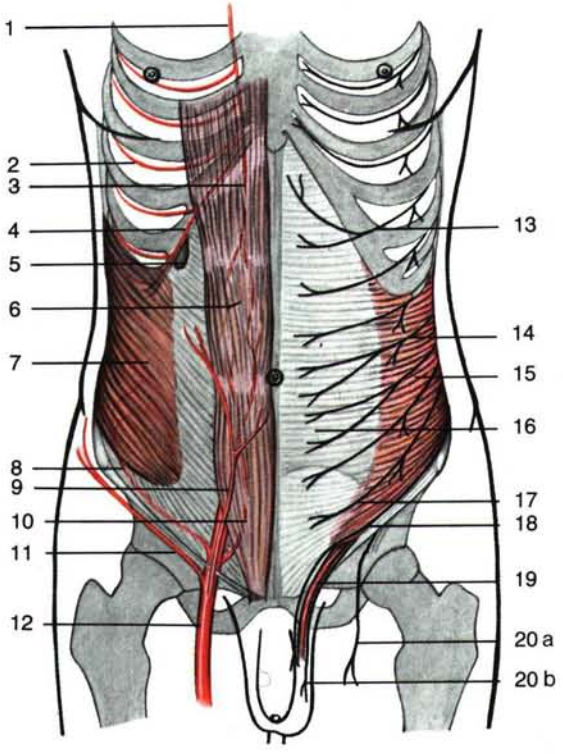

Arteries and nerves which supply the thoracic and abdominal wall. Note their segmental arrangement. (Schematic drawing.)

1 Internal thoracic artery
2 Intercostal artery
3 Superior epigastric artery
4 Musculophrenic artery
5 Gallbladder
6 Rectus abdominis muscle
7 External abdominal oblique muscle
8 Deep circumflex iliac artery
9 Superficial epigastric artery
10 Inferior epigastric artery
11 Superficial circumflex iliac artery
12 Femoral artery
13 Intercostal nerve
14 Thoracoabdominal nerve (T_{10})
15 Transversus abdominis muscle
16 Posterior layer of the rectus sheath
17 Iliohypogastric nerve (L_1)
18 Ilioinguinal nerve (L_1)
19 Spermatic cord
20 Genitofemoral nerve (L_1, L_2)
a Femoral branch
b Genital branch

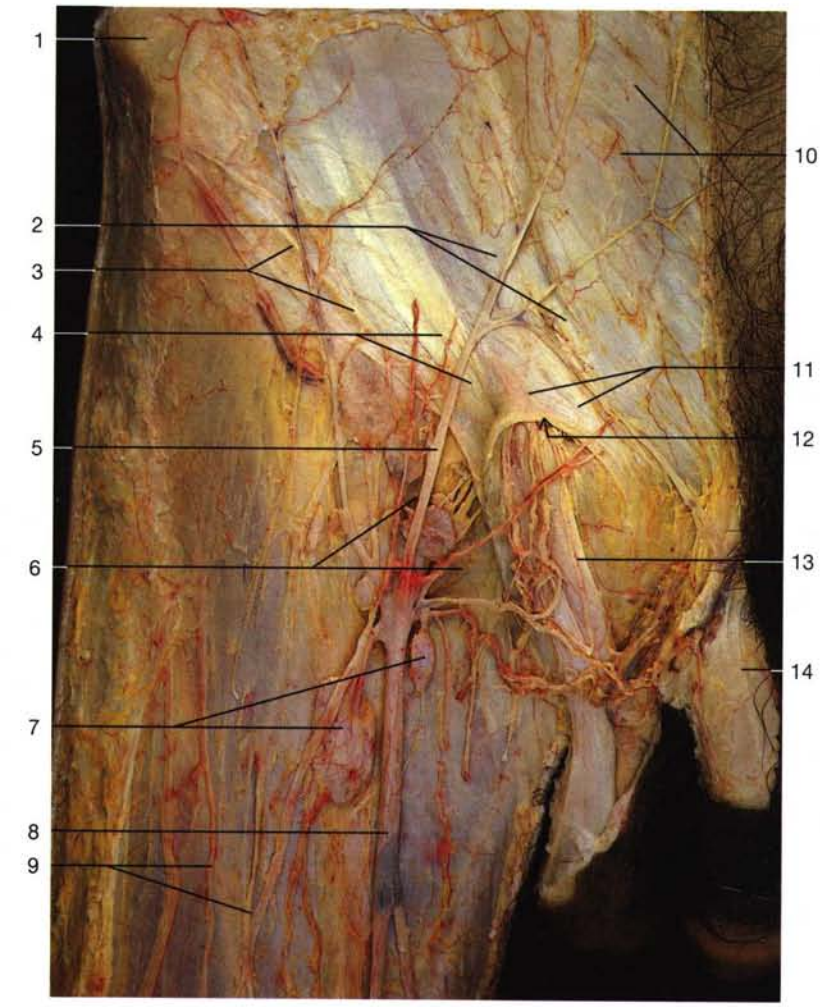

Inguinal canal in the male I. Dissection of superficial layer, right side (anterior aspect).

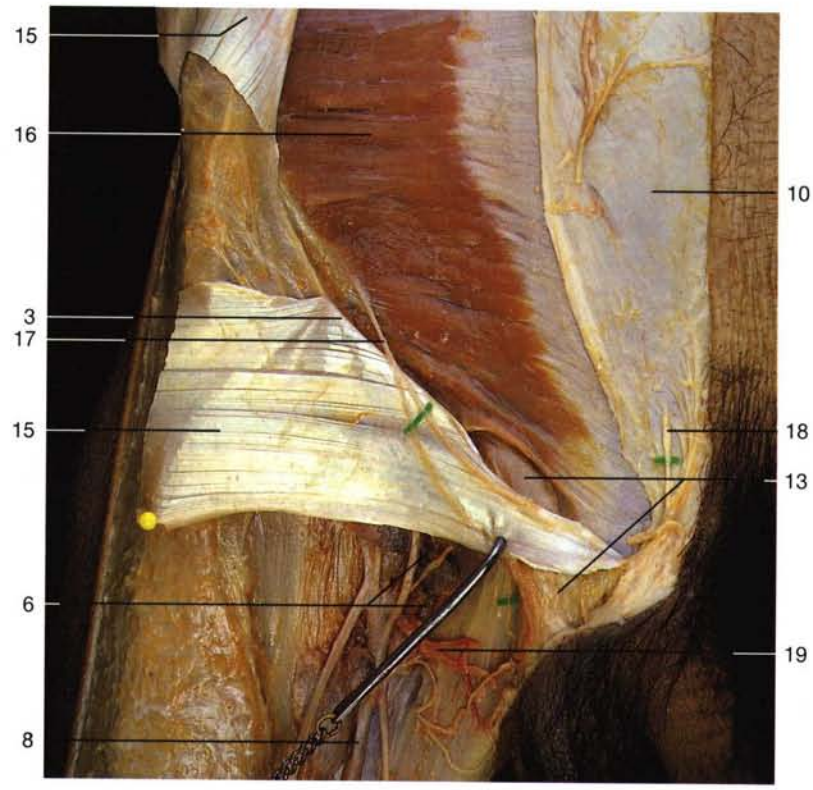

Inguinal canal in the male II, right side (anterior aspect).
The external abdominal oblique muscle has been divided to display the inguinal canal.

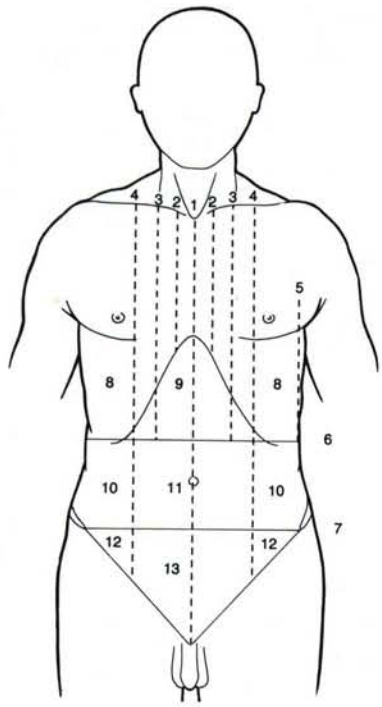

Regions and reference lines
for delineating surface projections.

Reference lines and regions

1 Median line
2 Lateral sternal line
3 Parasternal line
4 Left lateral or midclavicular line
5 Midaxillary line
6 Transpyloric plane
7 Transtubercular plane
8 Hypochondriac region
9 Epigastric region
10 Lumbar region
11 Umbilical region
12 Iliac region
13 Hypogastric region

1 Anterior superior iliac spine
2 Medial crus of inguinal ring
3 Inguinal ligament
4 Lateral crus of inguinal ring
5 Superficial epigastric vein
6 Saphenous opening
7 **Superficial inguinal lymph nodes**
8 **Great saphenous vein**
9 Anterior cutaneous branches of femoral nerve
10 **Anterior layer of rectus sheath**
11 Intercrural fibers
12 **Superficial inguinal ring**
13 Spermatic cord and genital branch of genitofemoral nerve
14 Penis
15 Aponeurosis of external abdominal oblique muscle (divided and reflected)
16 **Internal abdominal oblique muscle**
17 Ilioinguinal nerve
18 Anterior cutaneous branches of iliohypogastric nerve
19 Superficial external pudendal veins

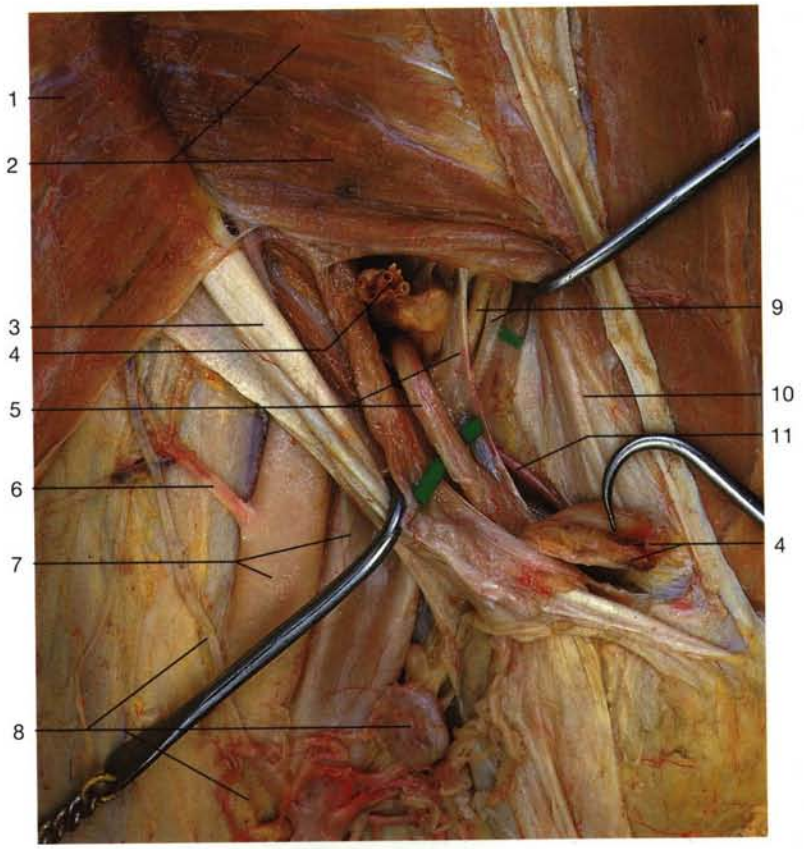

Inguinal canal in the male III. Deep dissection (anterior aspect, right side). Spermatic cord with exception of ductus deferens (probe) has been divided and reflected.

1 Internal abdominal oblique muscle (reflected)
2 Transversus abdominis muscle
3 Inguinal ligament
4 **Spermatic cord** with the exception of the ductus deferens (divided and reflected)
5 **Ductus deferens** and interfoveolar ligament
6 Superficial circumflex iliac artery
7 **Femoral artery and vein**
8 Superficial inguinal lymph nodes and inguinal lymph vessel
9 **Inferior epigastric artery and vein**
10 Falx inguinalis or conjoint tendon (cut)
11 Pubic branch of inferior epigastric artery
12 Superficial inguinal ring
13 Penis
14 External abdominal oblique muscle
15 Anterior superior iliac spine
16 Intercrural fibers
17 Fascia lata and sartorius muscle
18 Saphenous opening and great saphenous vein
19 Deep inguinal ring
20 Skin of scrotum and dartos muscle
21 Cremaster muscle
22 Internal spermatic fascia
23 Ductus deferens
24 Epididymis
25 Peritoneum (blue)
26 Remnant of processus vaginalis
27 Tunica vaginalis testis
28 Rectus abdominis muscle
29 Spermatic cord with ductus deferens covered by external spermatic fascia
30 Anterior layer of rectus sheath
31 Suspensory ligament of penis
32 **Testis and epididymis**
33 Ductus deferens
34 Pampiniform venous plexus and testicular artery
35 Inferior epigastric artery
36 Lateral femoral cutaneous nerve
37 Ilioinguinal nerve
38 **Femoral nerve**
39 Sartorius muscle
40 Deep dorsal vein of penis

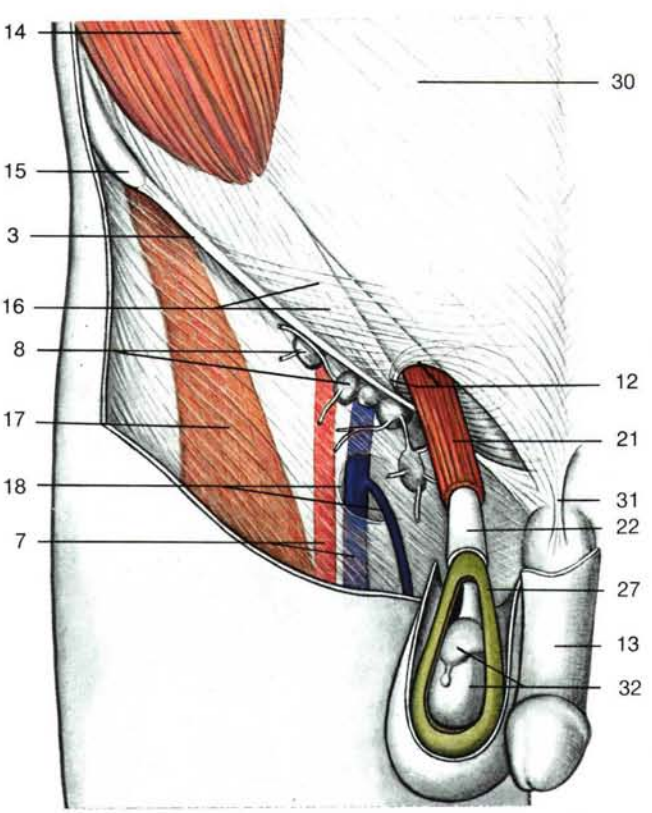

General characteristics of lower part of anterior abdominal wall and inguinal canal. (Schematic drawing.)

Inguinal hernias may either pass through the inguinal canal lateral to the inferior epigastric artery (indirect or lateral inguinal hernias, A and C) or directly penetrate the abdominal wall through the inguinal triangle located medial to the inferior epigastric artery (direct or medial inguinal hernias, B). The lateral hernias can be congenital if the vaginal process remains open (C) or acquired (A) if the hernia develops independently of a patent processus vaginalis.

Femoral hernias generally protrude through the femoral ring below the inguinal ligament. Proper assessment of the site of herniation requires the identification of both the inguinal ligament and the epigastric artery.

Inguinal and femoral region in the male (anterior aspect). On the right, the spermatic cord was dissected to display the ductus deferens and the accompanying vessels and nerves. The fascia lata on the left side has been removed.

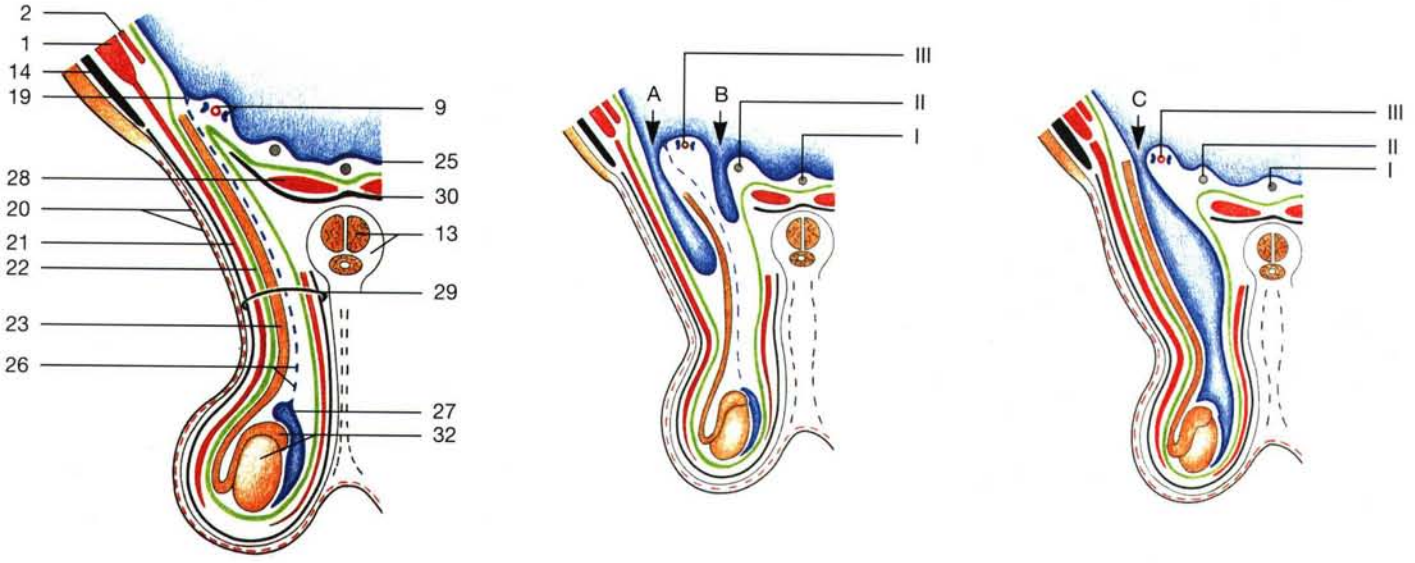

Layers of spermatic cord and types of hernias. Left: Normal situation; middle: Location of acquired inguinal hernias; A = indirect; B = direct inguinal hernia. Right: congenital indirect inguinal hernia (C); the vaginal process remained open.

I = Median umbilical fold, containing urachus chord.

II = Medial umbilical fold with remnants of umbilical artery and vein.

III = Lateral umbilical fold with inferior epigastric artery and vein.

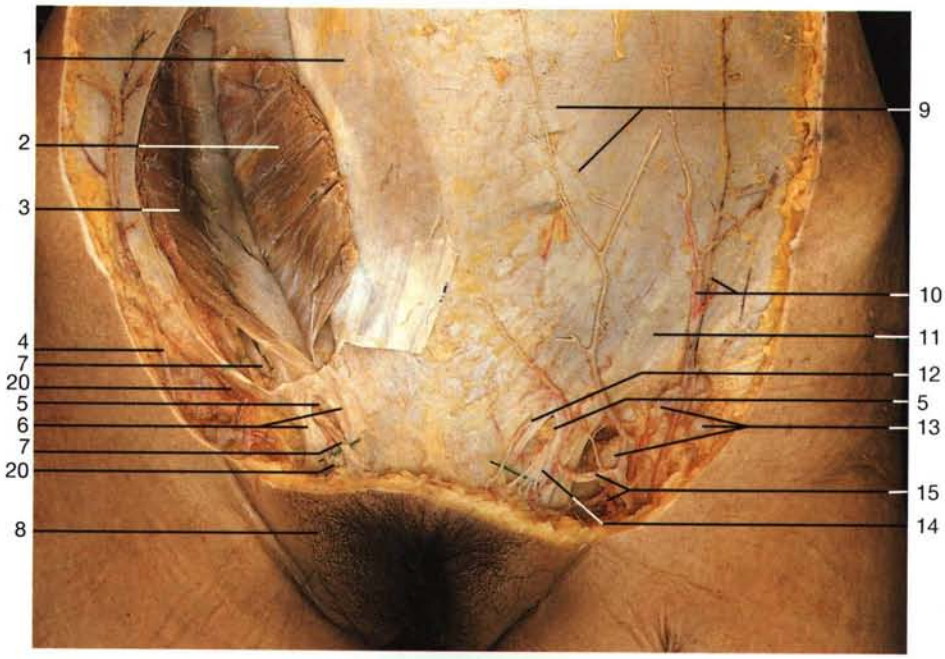

Inguinal region in the female (anterior aspect). Left side: superficial layer; right side: external and internal abdominal oblique muscle divided and reflected.

1 Aponeurosis of external abdominal oblique muscle
2 Internal abdominal oblique muscle (divided and reflected)
3 Transversus abdominis muscle
4 Superficial circumflex iliac artery and vein
5 Superficial inguinal ring with fat pad
6 Medial and lateral crural fibers
7 **Round ligament** (ligamentum teres uteri)
8 Labium majus pudendi
9 Anterior layer of rectus sheath
10 Superficial epigastric artery and vein
11 **Inguinal ligament**
12 Cutaneous branch of ilioinguinal nerve
13 Superficial inguinal lymph nodes
14 Entrance of round ligament into the labium majus
15 External pudendal artery and vein
16 Position of deep inguinal ring
17 **Ilioinguinal nerve**
18 Internal abdominal oblique muscle
19 Pubic branch of inferior epigastric artery
20 Genital branch of genitofemoral nerve
21 Fat pad of inguinal canal
22 Ilioinguinal nerve
23 Sheath of round ligament (inguinal canal)
24 Transversalis fascia

Inguinal canal of the female I (anterior aspect, right side). The external abdominal oblique muscle has been divided and reflected, to display the ilioinguinal nerve and the round ligament.

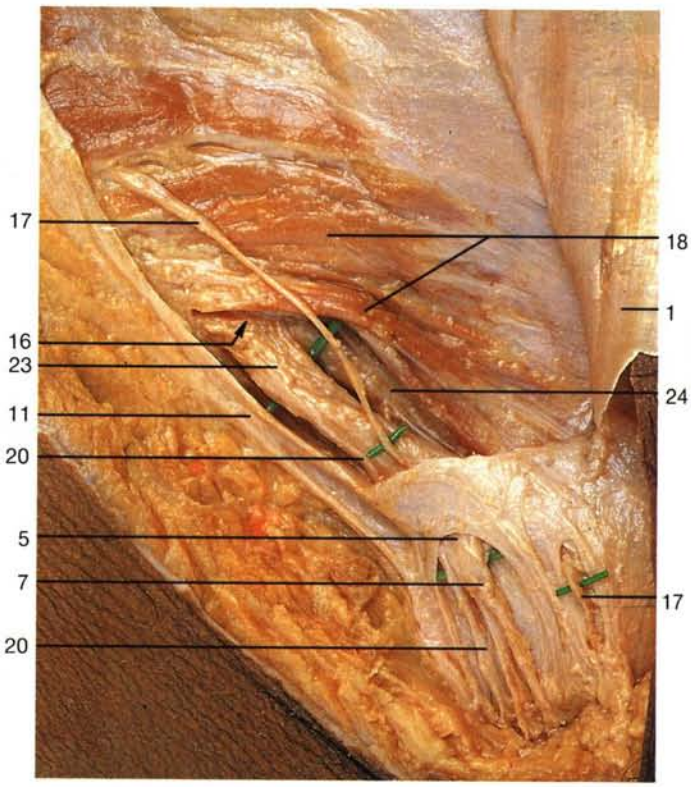

Inguinal canal of the female II (anterior aspect, right side). The external and internal abdominal oblique muscle have been divided and reflected to show the content of the inguinal canal.

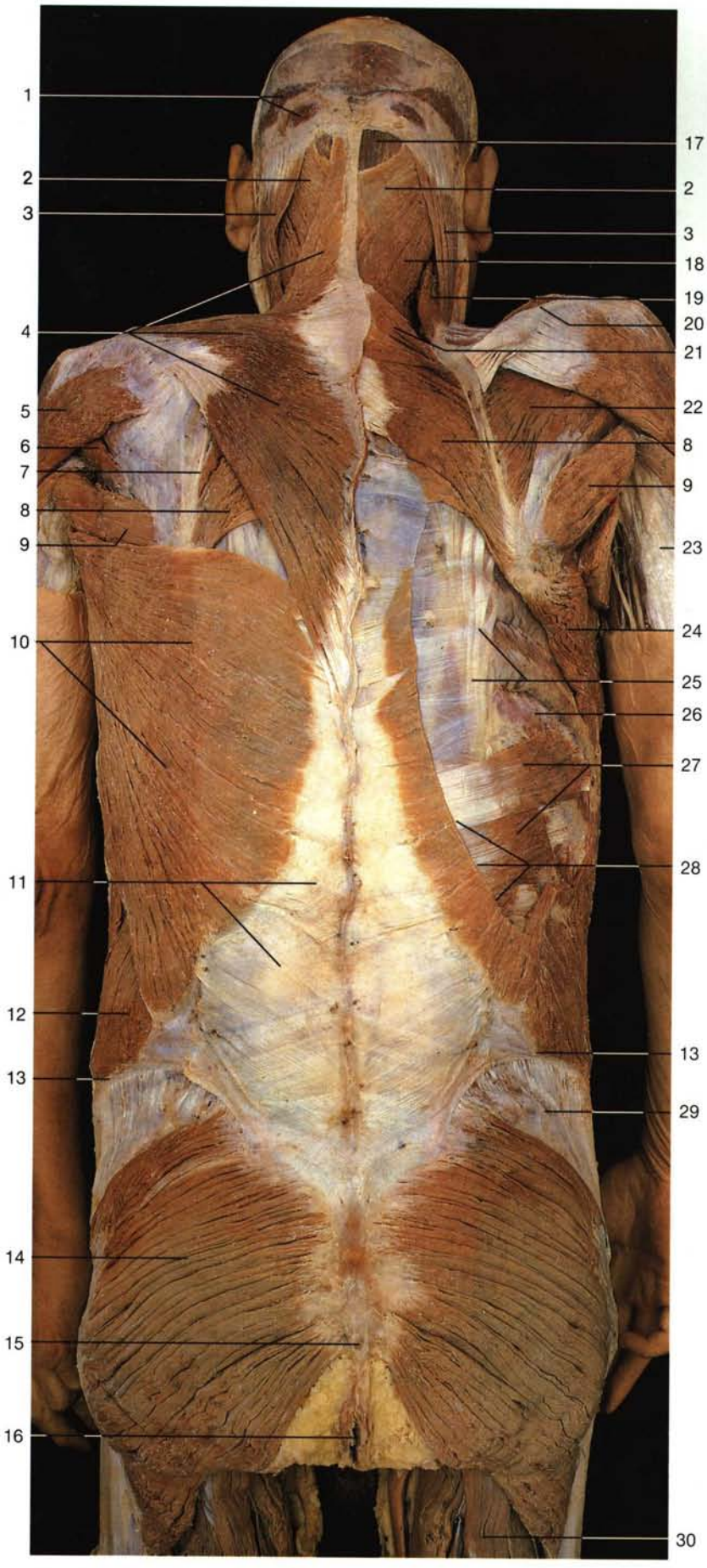

1 Occipital belly of occipitofrontalis muscle
2 Splenius capitis muscle
3 Sternocleidomastoid muscle
4 **Trapezius muscle**
5 Deltoid muscle
6 Teres minor muscle
7 Medial margin of scapula
8 **Rhomboid major muscle**
9 Teres major muscle
10 **Latissimus dorsi muscle**
11 Thoracolumbar fascia
12 External abdominal oblique muscle
13 Iliac crest
14 **Gluteus maximus muscle**
15 Position of last coccygeal vertebra
16 Anus
17 Semispinalis capitis muscle
18 Splenius cervicis muscle
19 Levator scapulae muscle
20 Spine of scapula
21 Rhomboideus minor muscle
22 Infraspinatus muscle
23 Triceps brachii muscle
24 Serratus anterior muscle
25 **Iliocostalis muscle**
26 External intercostal muscle
27 **Serratus posterior inferior muscle**
28 Latissimus dorsi muscle (cut edge)
29 Fascia over gluteus medius muscle
30 Long head of biceps femoris muscle

Muscles of the back I. Superficial layer on the left, deeper layer on the right. Right latissimus dorsi and trapezius muscles are removed.

Muscles of the back II. Dissection of deep muscles. Erector spinae muscle.

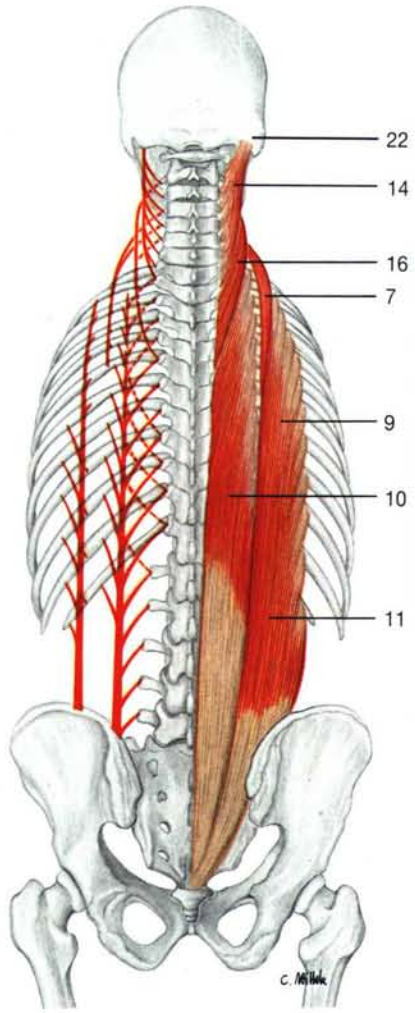

Origin and insertion of iliocostalis and longissimus muscles. (Schematic drawing.)

1 Rectus capitis posterior minor muscle
2 Rectus capitis posterior major muscle
3 Obliquus capitis inferior muscle
4 Spinous process of axis
5 Semispinalis cervicis muscle
6 Spinous process of seventh vertebra
7 **Iliocostalis cervicis muscle**
8 External intercostal muscles
9 **Iliocostalis thoracis muscle**
10 **Longissimus thoracis muscle**
11 **Iliocostalis lumborum muscle**
12 Internal abdominal oblique muscle
13 Semispinalis capitis muscle (divided)
14 **Longissimus capitis muscle**
15 Levator scapulae muscle
16 **Longissimus cervicis muscle**
17 Rhomboid major muscle
18 **Spinalis thoracis muscle**
19 Serratus posterior inferior muscle (reflected)
20 Spinous process of second lumbar vertebra
21 Iliac crest
22 Mastoid process

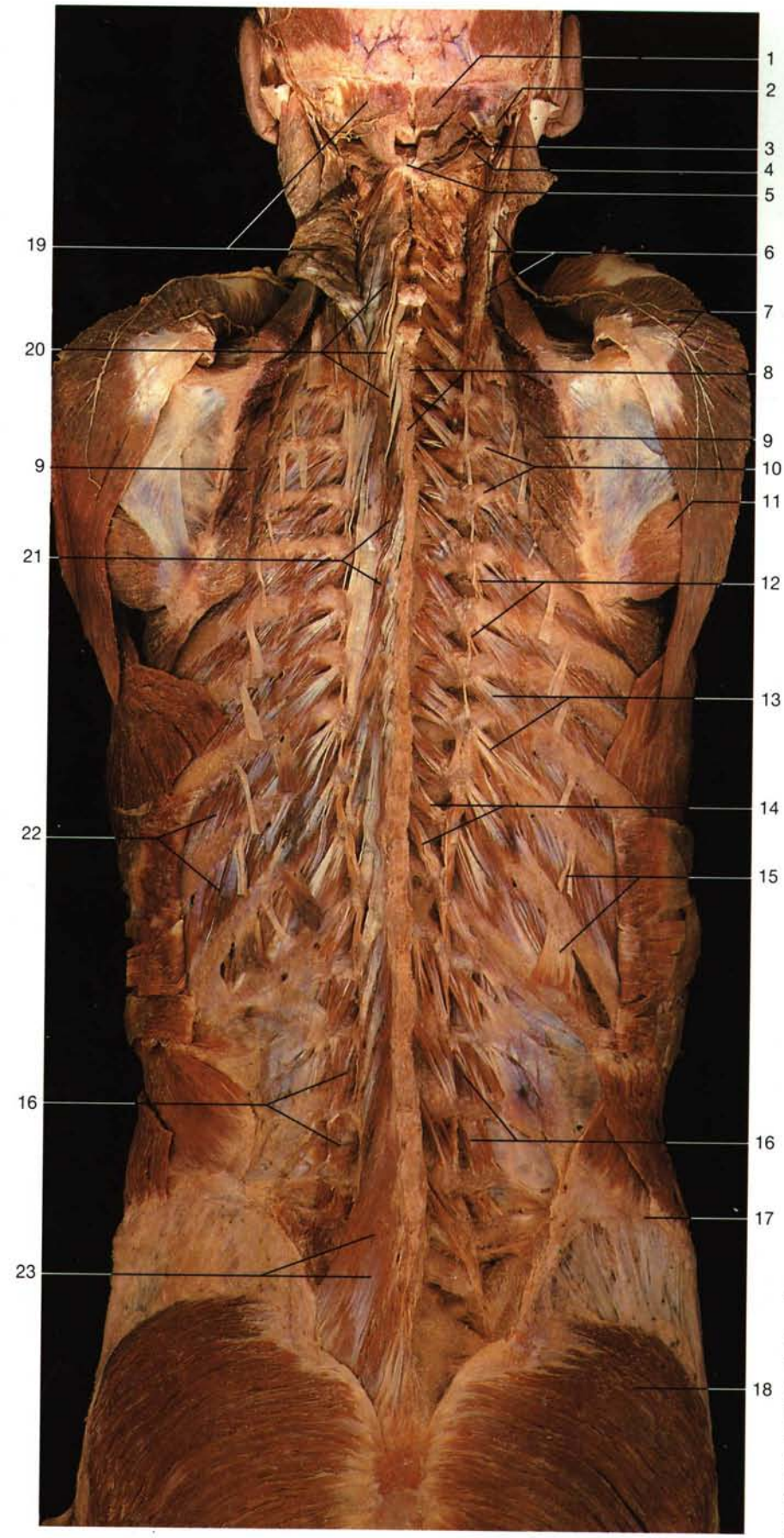

Muscles of the back III. Transversospinal muscles, deepest layer on the right, where all parts of semispinalis and multifidus muscles have been removed.

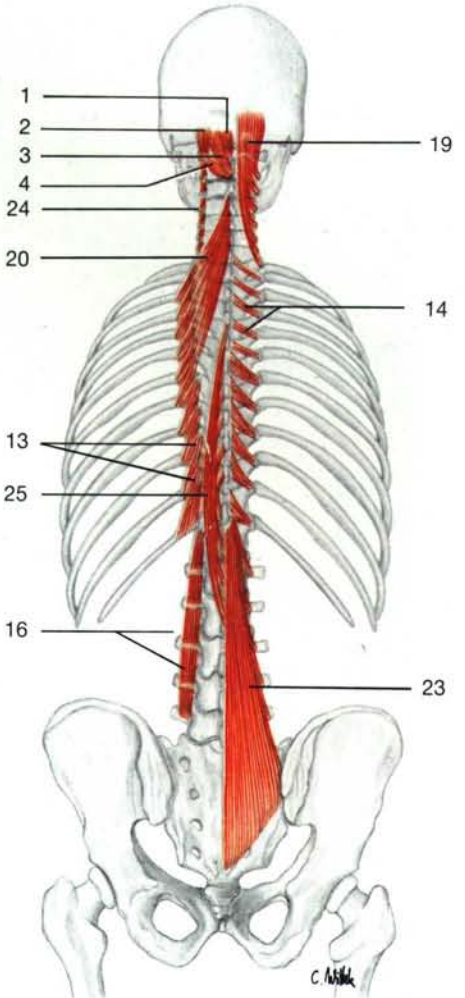

Medial column of intrinsic muscles of the back. Transversospinal and intertransversal system. (Schematic drawing.)

1 Rectus capitis posterior minor muscle
2 Obliquus capitis superior muscle
3 Rectus capitis posterior major muscle
4 Obliquus capitis inferior muscle
5 Spinous process of axis
6 **Longissimus capitis muscle**
7 Trapezius muscle (reflected) and accessory nerve (n. XI)
8 Spinous processes
9 Rhomboid major muscle
10 Transverse processes of thoracic vertebrae
11 Teres major muscle
12 Intertransverse ligaments
13 **Levatores costarum muscles**
14 **Rotatores muscles**
15 Tendons of iliocostalis muscle
16 Intertransversarii lumborum muscle (lateral)
17 Iliac crest
18 Gluteus maximus muscle
19 Semispinalis capitis muscle
20 **Semispinalis cervicis muscle**
21 Semispinalis thoracis muscle
22 **External intercostal muscles**
23 **Multifidus muscle**
24 Intertransversarii posterior cervicis muscle
25 Spinalis thoracis muscle

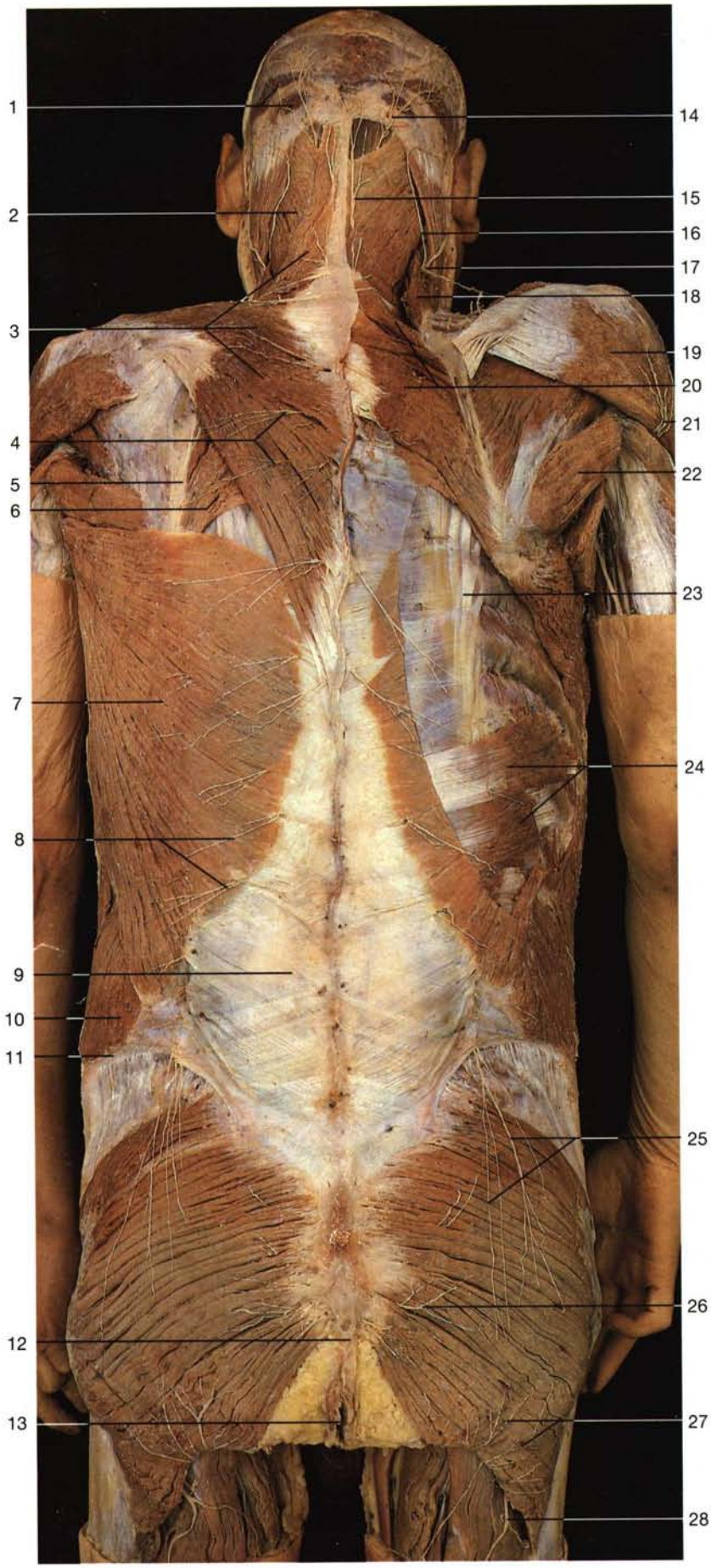

Innervation of the back I. Superficial (left) and deeper (right) layers. Right trapezius and latissimus dorsi muscles removed.

1 Occipital belly of occipitofrontalis muscle
2 Splenius capitis muscle
3 **Trapezius muscle**
4 **Medial cutaneous branches of dorsal rami of spinal nerves**
5 Medial margin of scapula
6 Rhomboid major muscle
7 **Latissimus dorsi muscle**
8 **Lateral cutaneous branches of dorsal rami of spinal nerves**
9 Thoracolumbar fascia
10 External abdominal oblique muscle
11 Iliac crest
12 Last coccygeal vertebra
13 Anus
14 Greater occipital nerve
15 Third occipital nerve
16 Lesser occipital nerve
17 Cutaneous branches of cervical plexus
18 Levator scapulae muscle
19 Deltoid muscle
20 **Rhomboid major and minor muscles**
21 Upper lateral cutaneous nerve of arm (branch of axillary nerve)
22 **Teres major muscle**
23 Iliocostalis thoracis muscle
24 Serratus posterior inferior muscle
25 Superior cluneal nerves
26 Middle cluneal nerves
27 Inferior cluneal nerves
28 Posterior femoral cutaneous nerve

▷
To page 213:

1 **Trapezius muscle**
2 Infraspinatus muscle
3 Left **latissimus dorsi muscle**
4 Thoracolumbar fascia
5 Splenius cervicis muscle
6 Serratus posterior superior muscle
7 **Medial branches of dorsal rami of thoracic spinal nerves**
8 **Lateral branches of dorsal rami of thoracic spinal nerves**
9 Iliocostalis muscle
10 Serratus posterior inferior muscle
11 Latissimus dorsi muscle (reflected)

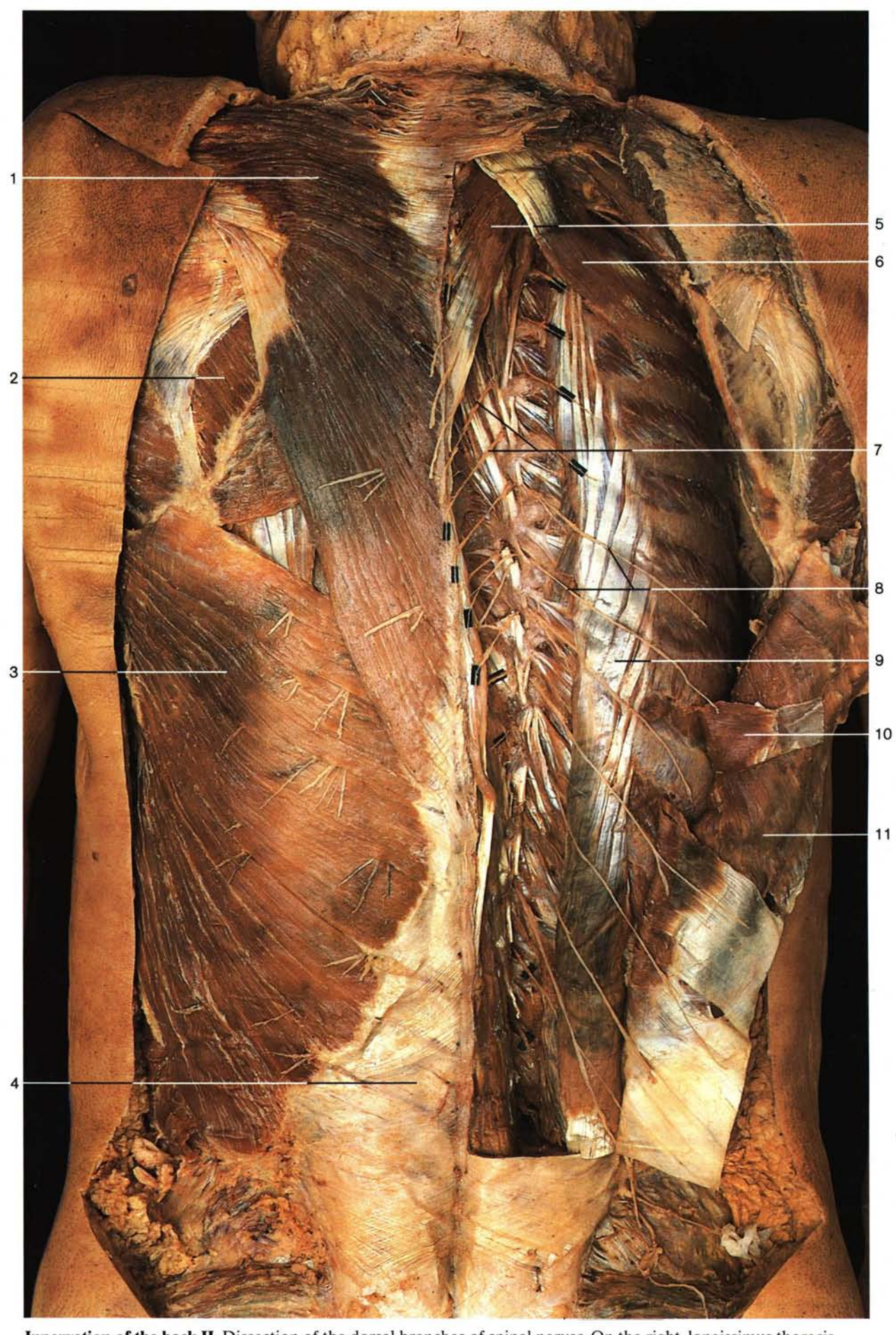

Innervation of the back II. Dissection of the dorsal branches of spinal nerves. On the right, longissimus thoracis muscle has been removed and iliocostalis muscle laterally reflected.

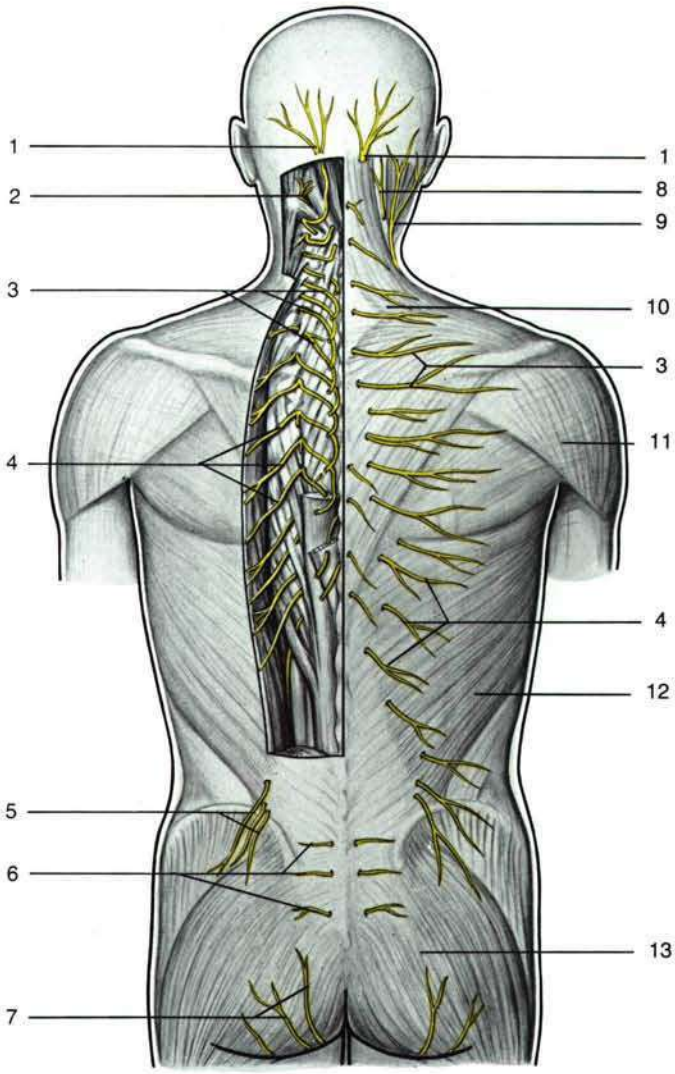

1 Greater occipital nerve (C_2)
2 Suboccipital nerve (C_1)
3 **Medial branches of dorsal rami of spinal nerves**
4 **Lateral branches of dorsal rami of spinal nerves**
5 Superior cluneal nerves (L_1–L_3)
6 Middle cluneal nerves (S_1–S_3)
7 Inferior cluneal nerves (derived from branches of the sacral plexus, ventral rami)
8 Lesser occipital nerve
9 Great auricular nerve
10 Trapezius muscle
11 Deltoid muscle
12 Latissimus dorsi muscle
13 Gluteus maximus muscle
14 External intercostal muscle
15 Internal intercostal muscle
16 Innermost intercostal muscle
17 Dorsal ramus of spinal nerve
18 **Spinal nerve** and **spinal ganglion**
19 Sympathetic trunk with ganglion
20 Intercostal nerve
21 Lateral cutaneous branch } of **intercostal nerve**
22 Anterior cutaneous branch } of **intercostal nerve**

General characteristics of the innervation of the back. Distribution of dorsal branches of spinal nerves. Note the segmental arrangement of the innervation of the dorsal part of the trunk. (Schematic drawing.)

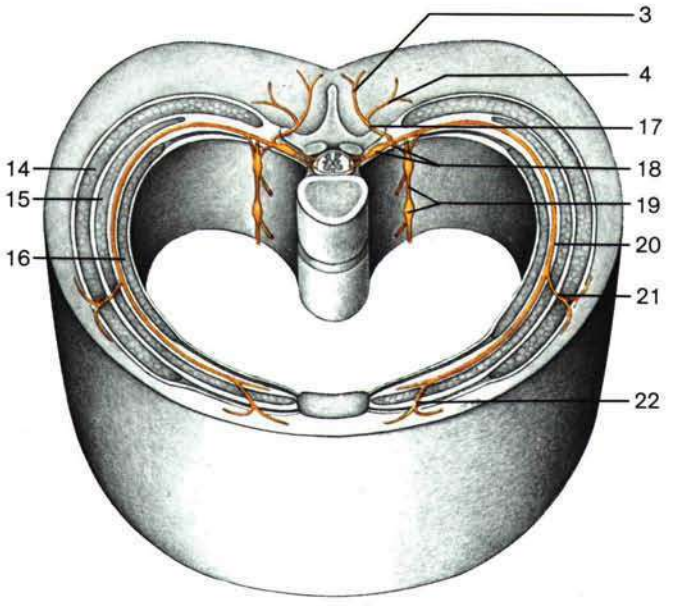

Position and branches of spinal nerves in one segment of thoracic wall. (Schematic drawing.)

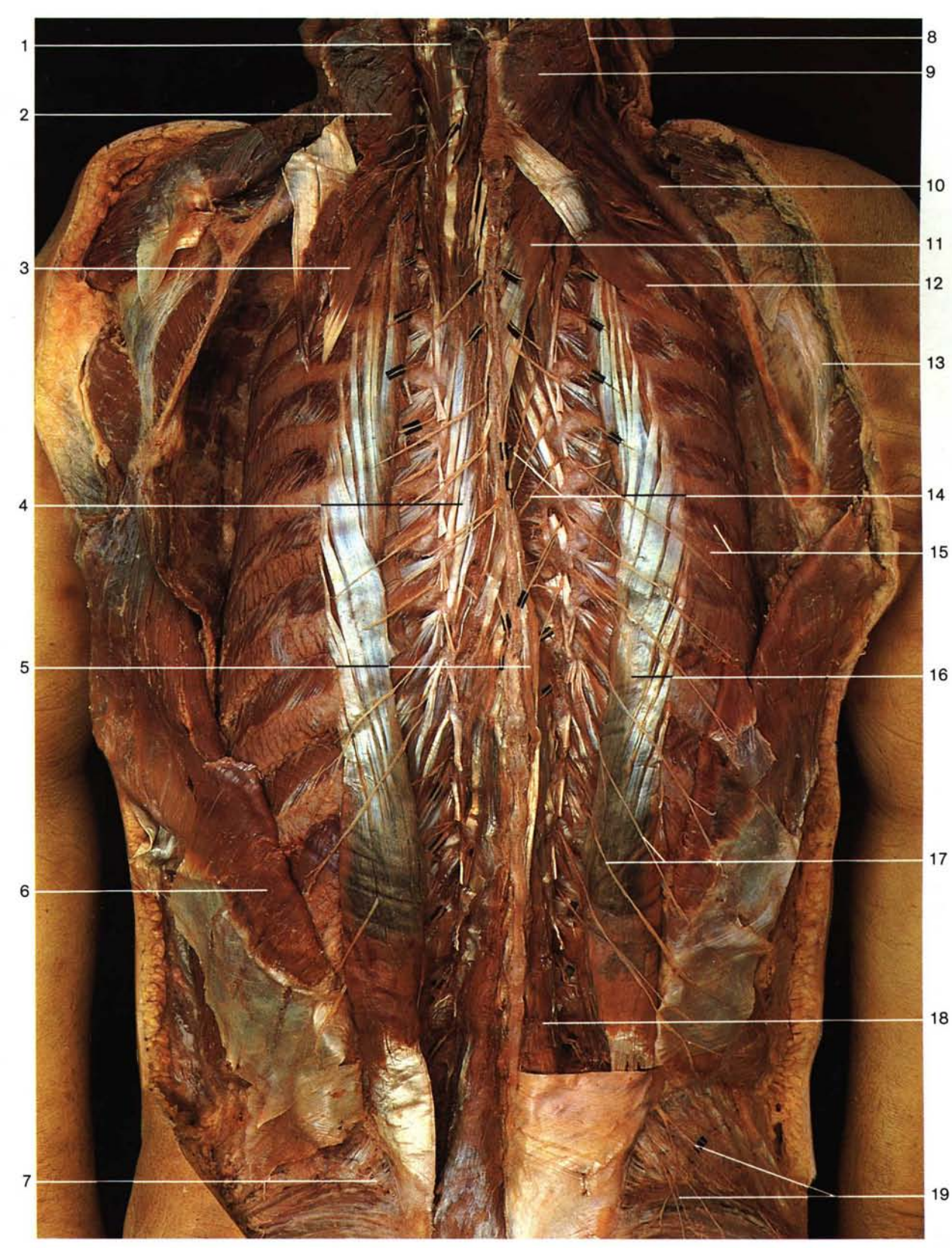

Innervation of the back III. Deeper layer (dorsal aspect).

1 Semispinalis capitis muscle
2 Left splenius capitis muscle (cut and reflected)
3 Left splenius cervicis muscle (cut and reflected)
4 Semispinalis thoracis muscle
5 **Spinalis thoracis muscle**
6 Latissimus dorsi muscle (reflected)
7 Iliac crest
8 Lesser occipital nerve
9 Splenius capitis muscle
10 Levator scapulae muscle
11 **Splenius cervicis muscle**
12 Serratus posterior superior muscle
13 Scapula
14 **Medial branches of dorsal rami of spinal nerves**
15 Rib and external intercostal muscle
16 Iliocostalis thoracis muscle
17 **Lateral branches of dorsal rami of spinal nerves**
18 Multifidus muscle
19 Superior cluneal nerves

Lumbar portion of spinal cord. Note the relation between the nervous and muscular segments.

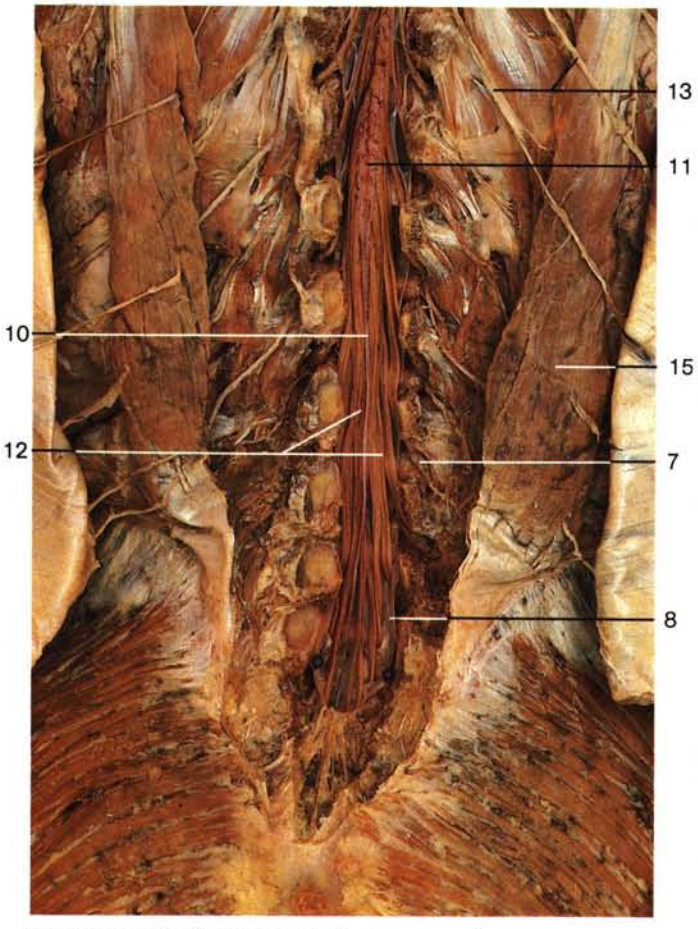

Terminal part of spinal cord. Dura removed.

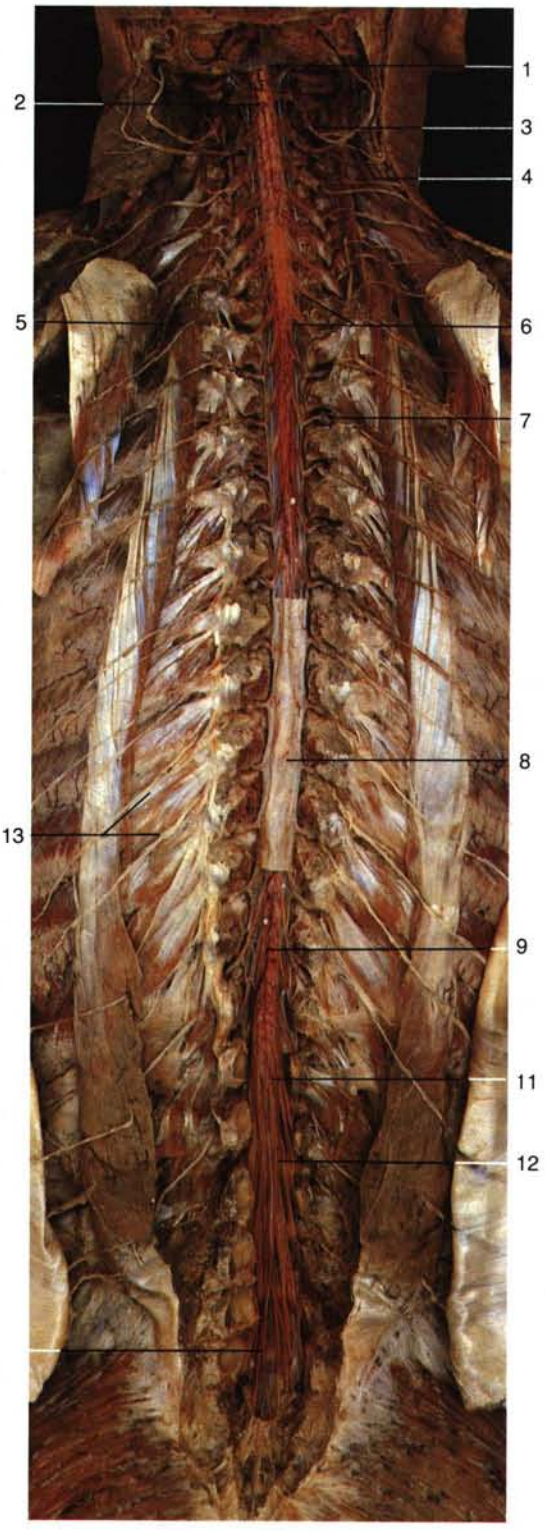

Innervation of the back IV. Spinal cord in the vertebral canal (opened). Longissimus dorsi muscle has been removed and iliocostal muscle reflected.

1 Cerebellomedullary cistern and cerebellum
2 Medulla oblongata
3 Greater occipital nerve (C_2)
4 Third cervical nerve (C_3)
5 Dorsal primary ramus
6 Dorsal roots
7 Spinal ganglion
8 Spinal dura mater
9 Spinal arachnoid mater
10 Filum terminale
11 Conus medullaris
12 Cauda equina
13 Lateral branches of dorsal rami of spinal nerves
14 Ventral ramus of spinal nerve (intercostal nerve)
15 Iliocostalis muscle

Thoracic portion of spinal cord (dorsal aspect). Vertebral canal and dura mater opened.

1 Arch of vertebra (divided)
2 Spinal nerve with meningeal coverings
3 Dorsal roots of thoracic spinal nerves
4 Spinal cord (thoracic portion)
5 Spinal ganglia with meningeal coverings
6 Pia mater with blood vessels
7 Dura mater (opened)
8 Denticulate ligament
9 Lateral branch of dorsal ramus
10 Dorsal ramus of spinal nerve (dividing into a medial and lateral branch)
11 Medial branch of dorsal ramus of spinal nerve
12 Spinal dura mater
13 Spinal nerves of sacral segments
14 Filum terminale

Terminal part of spinal cord with dura mater (dorsal aspect). Dorsal part of sacrum removed.

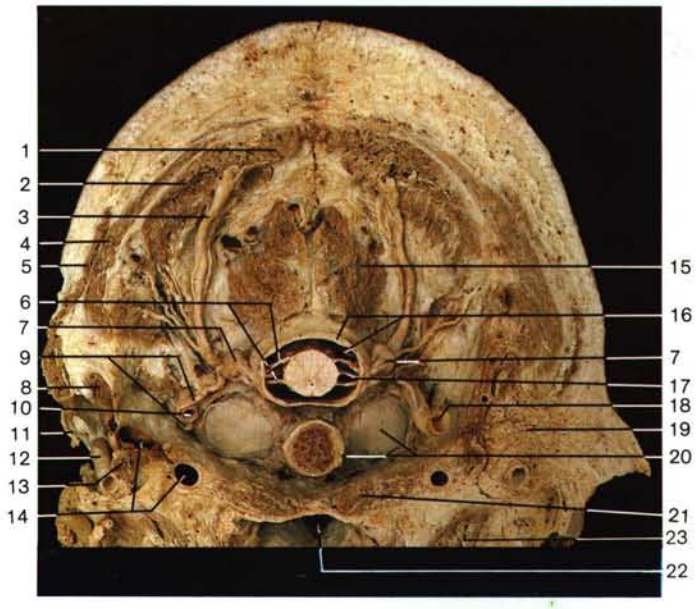

Horizontal section of the neck. Dissection of the second cervical spinal nerve. Posterior surface at top of figure.

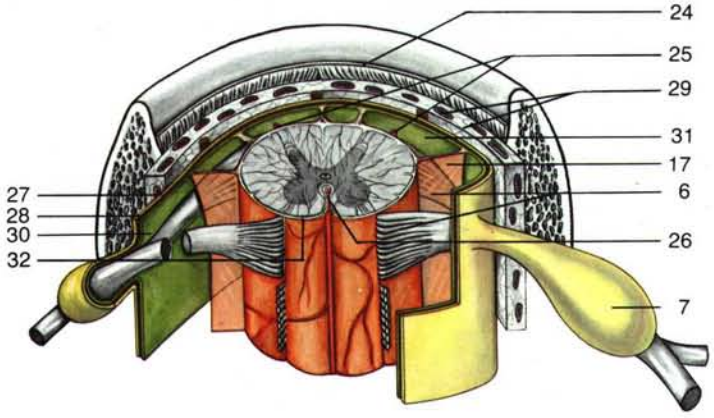

Meningeal coverings of the spinal cord (anterior aspect). (Schematic drawing.)

1 Trapezius muscle
2 Semispinalis capitis muscle
3 **Dorsal ramus of spinal nerve**
4 Sternocleidomastoid muscle
5 Platysma muscle
6 **Dorsal and ventral roots of spinal nerves**
7 **Spinal ganglion**
8 Posterior belly of digastric muscle
9 **Ventral ramus of spinal nerve**
10 Vertebral artery
11 Great auricular nerve
12 Superficial temporal artery
13 Styloid process
14 Internal jugular vein and internal carotid artery
15 Rectus capitis posterior major muscle
16 Dura mater and subarachnoid space
17 Denticulate ligament
18 Vertebral artery
19 Parotid gland
20 Dens of axis (divided) and inferior articular facet of atlas
21 Longus capitis muscle
22 Pharyngeal cavity
23 Medial pterygoid muscle
24 Periosteum of vertebral canal
25 Posterior spinal arteries
26 Anterior spinal artery

Meningeal coverings
27 Dura mater
28 Subdural space
29 Extradural or epidural space with venous plexus and fatty tissue
30 Arachnoid (green)
31 Subarachnoid space
32 Pia mater (pink)
33 Nucleus pulposus
34 Crus of diaphragm
35 Intervertebral disc
36 Body of first lumbar vertebra
37 Spinal cord
38 Conus medullaris
39 Cauda equina
40 Filum terminale
41 Spinous process

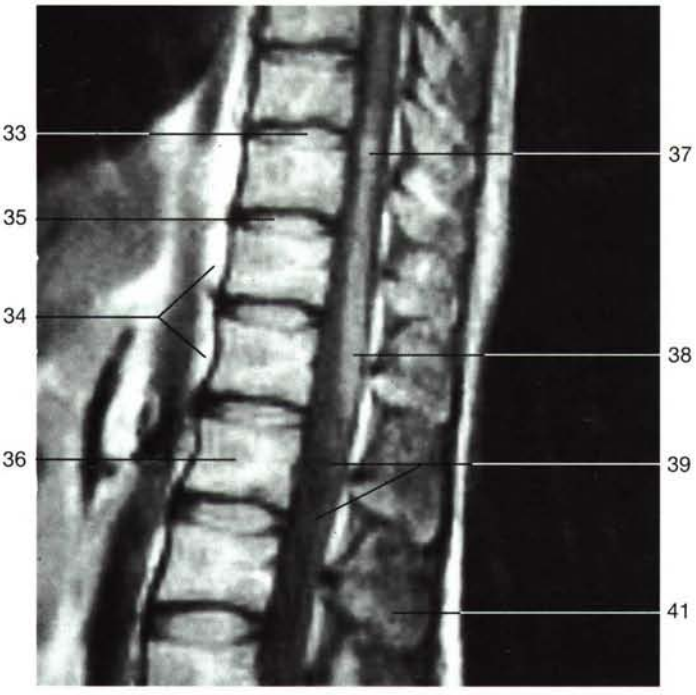

Sagittal section through the vertebral canal, T_9–L_2. (MR-Scan.)

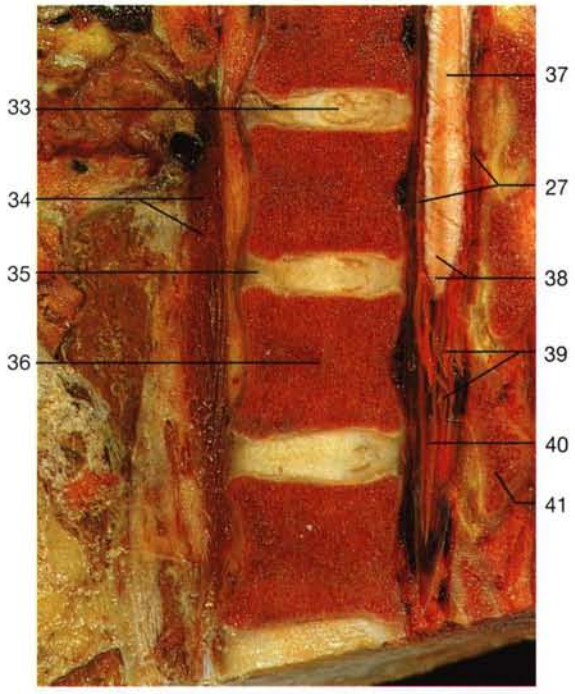

Sagittal section through vertebral canal, T_{12}–L_2. Notice red bone marrow (unfixed).

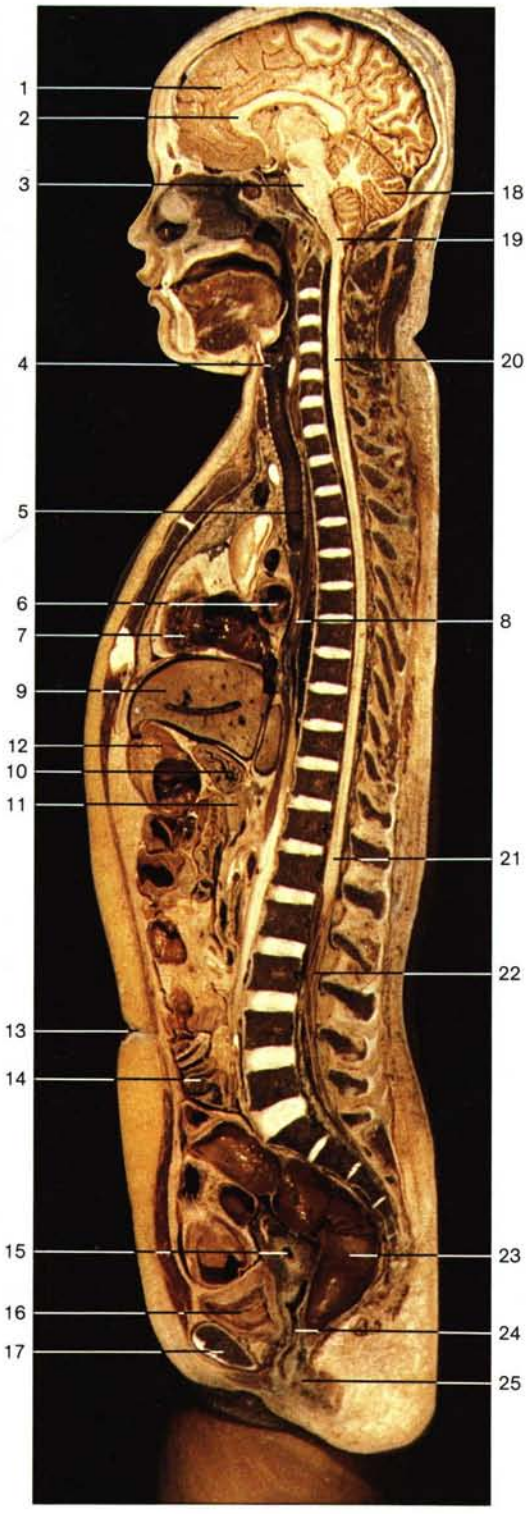

Median section of the head and trunk in the adult (female). The conus medullaris of the spinal cord is located at the level of L_1.

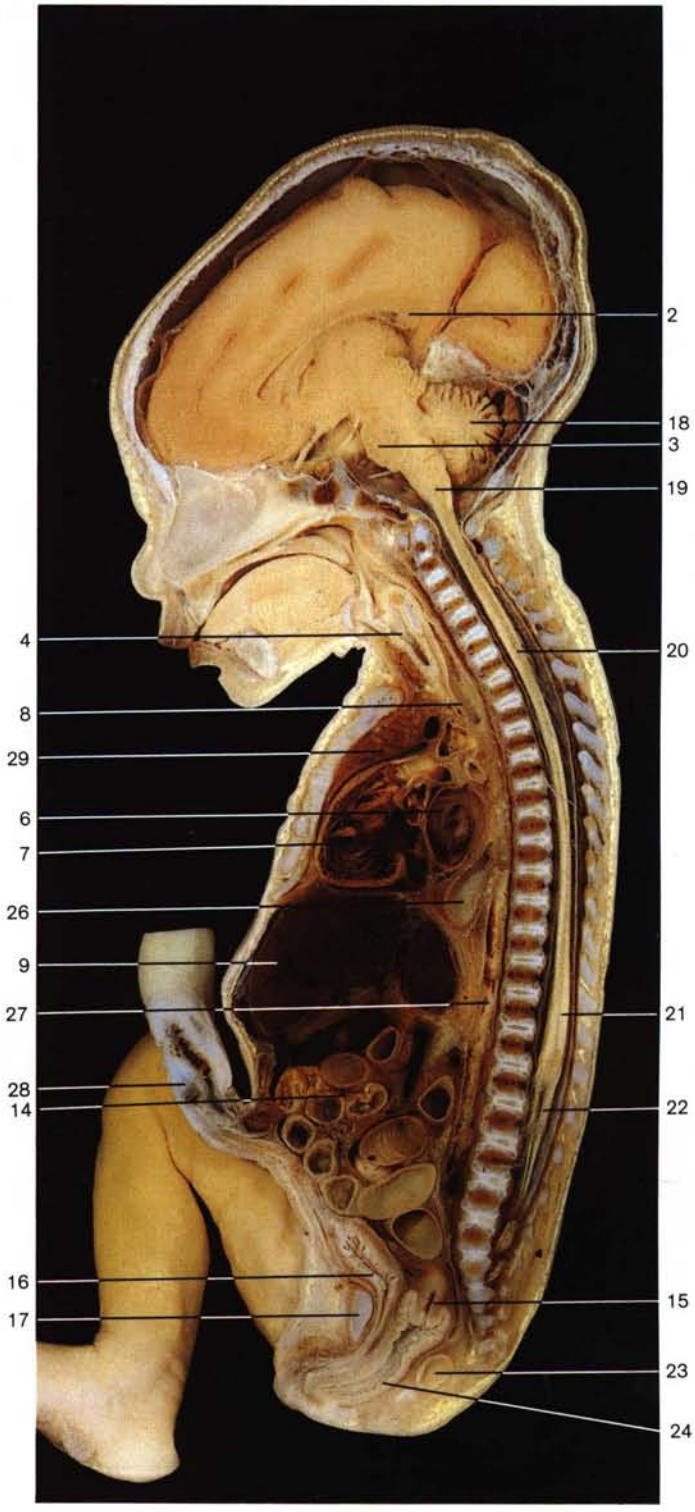

Median section of the head and trunk in the neonate. Note that in the neonate the conus medullaris of the spinal cord extends far more caudally than in the adult.

1 Cerebrum
2 Corpus callosum
3 Pons
4 Larynx
5 Trachea
6 Left atrium
7 **Right ventricle**
8 Esophagus
9 **Liver**
10 Stomach
11 Pancreas
12 Transverse colon
13 Umbilicus
14 Small intestine
15 Uterus
16 Urinary bladder
17 Pubic symphysis
18 Cerebellum
19 Medulla oblongata
20 **Spinal cord**
21 **Conus medullaris**
22 **Cauda equina**
23 Rectum
24 Vagina
25 Anus
26 Inferior vena cava
27 Aorta
28 Umbilical cord
29 Thymus

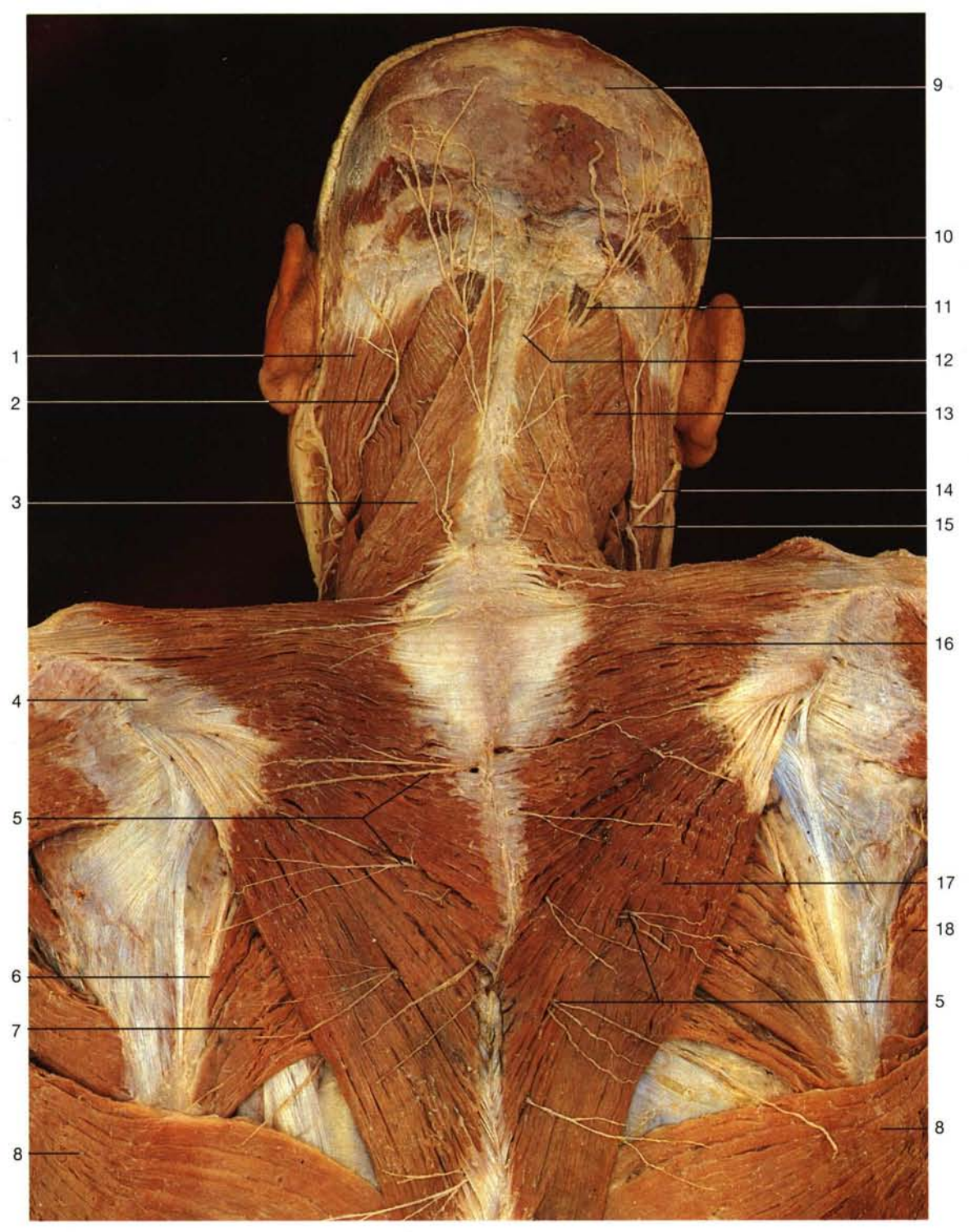

Dorsal aspect of the neck I. Superficial layer. Nuchal region and shoulder.

1 Sternocleidomastoid muscle
2 **Lesser occipital nerve**
3 Descending fibers of trapezius muscle
4 Spine of scapula
5 **Medial cutaneous branches of dorsal rami of spinal nerves**
6 Medial margin of scapula
7 Rhomboid major muscle
8 Latissimus dorsi muscle
9 Galea aponeurotica
10 Occipital belly of occipitofrontalis muscle
11 **Greater occipital nerve**
12 Third occipital nerve
13 Splenius capitis muscle
14 **Great auricular nerve**
15 Cutaneous nerves of cervical plexus
16 Transverse fibers of trapezius muscle
17 Ascending fibers of trapezius muscle
18 Teres major muscle

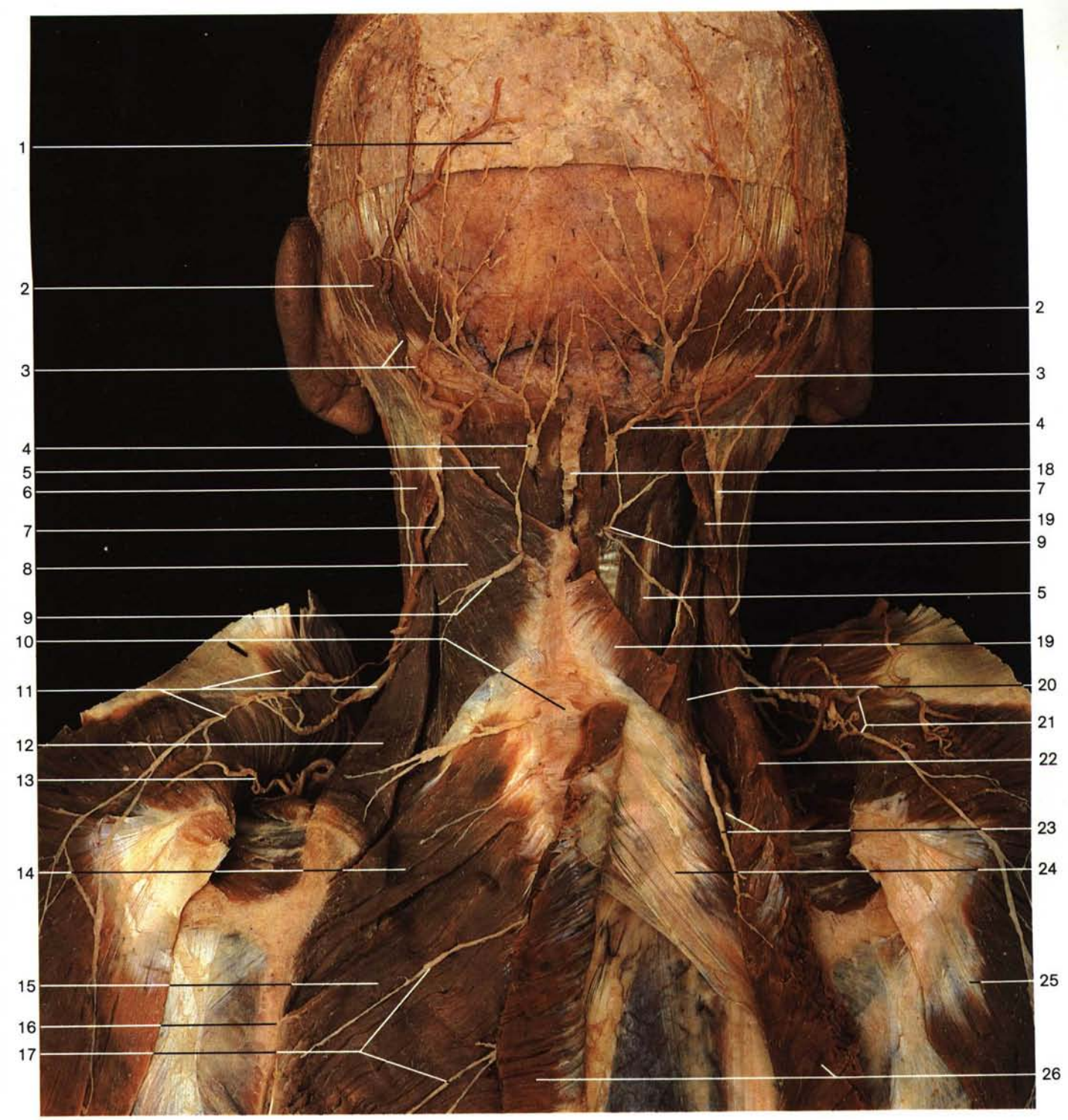

Dorsal aspect of neck II. Deeper layer. The left trapezius muscle has been divided and reflected. On the right, trapezius, rhomboid and splenius muscles have been divided. Right levator scapulae muscle has been slightly reflected.

1 Galea aponeurotica
2 Occipital belly of occipitofrontalis muscle
3 Occipital artery
4 **Greater occipital nerve** (C_2)
5 Semispinalis capitis muscle
6 Sternocleidomastoid muscle
7 Lesser occipital nerve
8 Left splenius capitis muscle
9 Third occipital nerve (C_3)
10 Spinous process of vertebra prominens (C_7)
11 Left trapezius muscle and **accessory nerve**
12 Levator scapulae muscle
13 Superficial branch of transverse cervical artery
14 Thomboid minor muscle
15 Rhomboid major muscle
16 Medial margin of scapula
17 **Medial branches of dorsal rami of spinal nerves**
18 Ligamentum nuchae
19 Splenius capitis muscle (divided)
20 Splenius cervicis muscle
21 Right **accessory nerve** and superficial branch of transverse cervical artery
22 Right levator scapulae muscle
23 Dorsal scapular nerve and deep branch of transverse cervical artery
24 Serratus posterior superior muscle
25 Right trapezius muscle (divided and reflected)
26 Right rhomboid major muscle (divided and reflected)

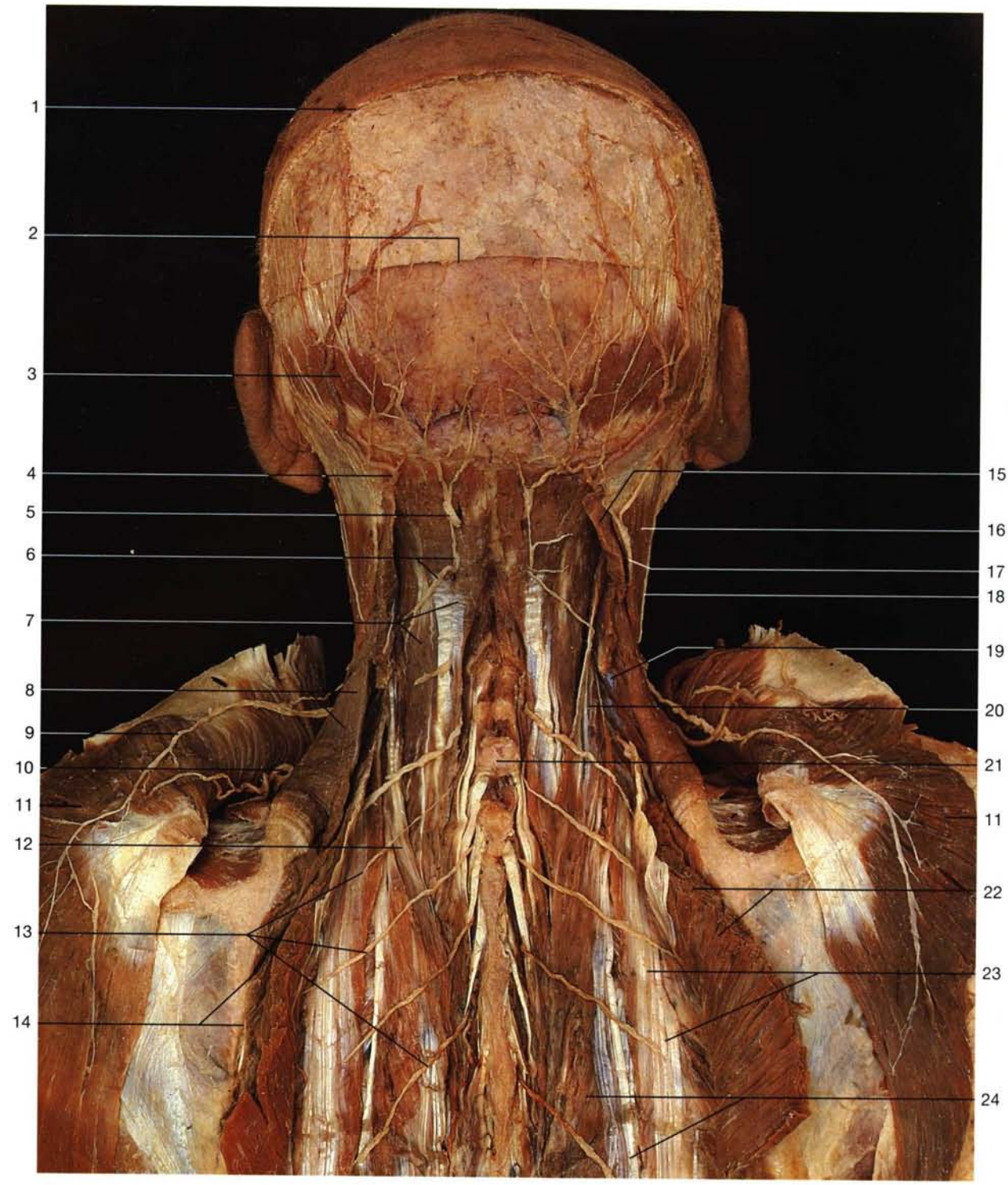

Dorsal aspect of neck III. Deepest layer. Nuchal region. Trapezius muscle and splenius capitis and cervicis muscles have been divided and partly removed or reflected.

1 Skin of scalp
2 Galea aponeurotica
3 Occipital belly of occipitofrontalis muscle
4 **Occipital artery**
5 **Greater occipital nerve**
6 Third occipital nerve
7 Semispinalis capitis muscle
8 Levator scapulae muscle
9 **Accessory nerve** (n. XI)
10 Superficial cervical artery
11 Trapezius muscle (reflected)
12 Longissimus cervicis muscle
13 **Medial cutaneous branches of dorsal rami of spinal nerves**
14 Medial margin of scapula
15 Splenius capitis muscle (divided)
16 Sternocleidomastoid muscle
17 **Lesser occipital nerve**
18 Great auricular nerve
19 Splenius cervicis muscle
20 Longissimus cervicis muscle
21 Spinous process of seventh cervical vertebra (vertebra prominens)
22 Rhomboid muscles (divided)
23 Iliocostalis thoracis muscle
24 Longissimus thoracis muscle

1 Semispinalis capitis muscle (divided)
2 External occipital protuberance
3 Obliquus capitis superior muscle
4 Rectus capitis posterior minor muscle
5 Rectus capitis posterior major muscle
6 Vertebral artery
7 Obliquus capitis inferior muscle
8 **Spinous process of axis**
9 Third cervical vertebra
10 Occipital belly of occipitofrontalis muscle
11 **Greater occipital nerve**
12 Suboccipital nerve (C_1)
13 Lesser occipital nerve
14 Third occipital nerve (C_3)
15 Mastoid process and splenius capitis muscle
16 Atlas
17 Axis
18 Spinous process of third cervical vertebra
19 Right semispinalis cervicis muscle
20 Deep cervical artery
21 Left splenius capitis muscle (divided)
22 Left sternocleidomastoid muscle
23 Great auricular nerve
24 Left **semispinalis capitis muscle**
25 Left longissimus cervicis muscle
26 Levator scapulae muscle
27 Muscular branch of vertebral artery
28 Left semispinalis cervicis muscle (divided)
29 **Medial branches of dorsal rami of spinal nerves**
30 **Occipital artery**
31 Dorsal scapular nerve

Dorsal aspect of neck IV. Deepest layer, suboccipital triangle; right semispinalis capitis muscle divited and reflected.

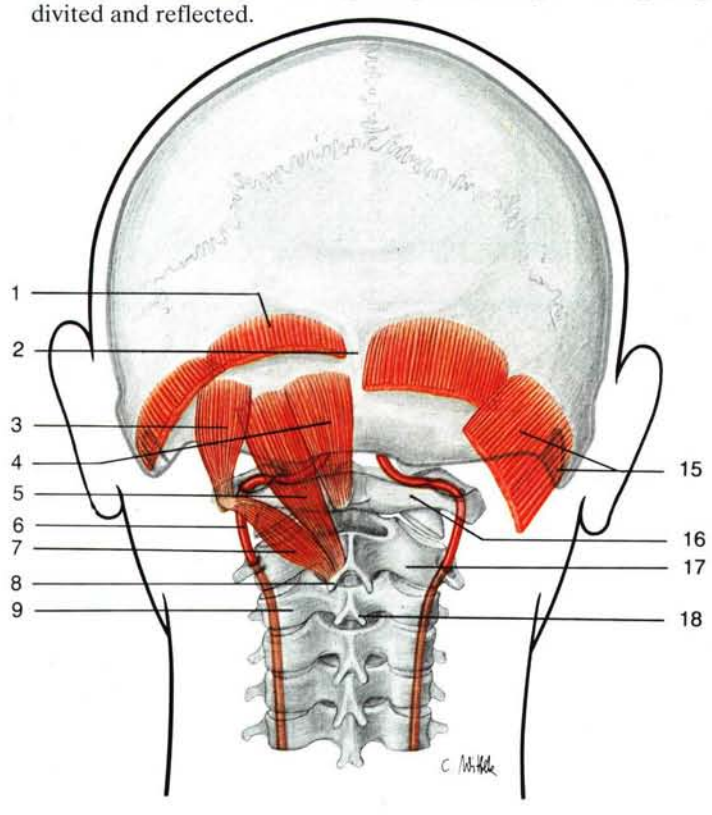

Suboccipital triangle and position of the vertebral artery. (Schematic drawing.)

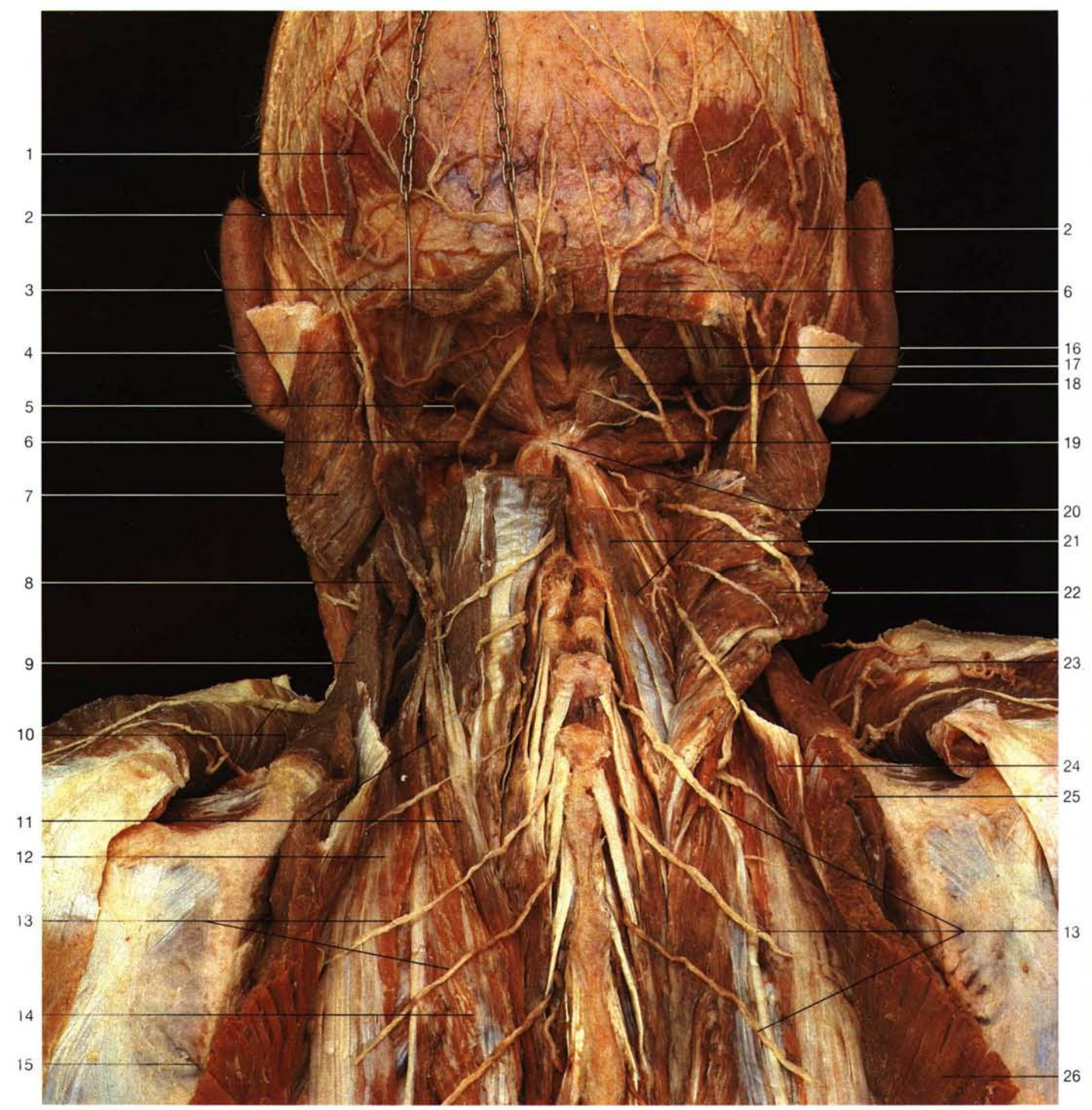

Dorsal aspect of neck V. Nuchal region. Deepest layer. Dissection of suboccipital triangle on both sides.

1 Occipital belly of occipitofrontalis muscle
2 **Occipital artery**
3 Insertion of semispinalis capitis muscle (divided)
4 Lesser occipital nerve (from cervical plexus)
5 Suboccipital nerve (C_1)
6 **Greater occipital nerve** (C_2)
7 Splenius capitis muscle (reflected)
8 Splenius cervicis muscle
9 Levator scapulae muscle
10 Accessory nerve (n. XI), trapezius muscle
11 Longissimus cervicis muscle
12 Iliocostalis cervicis muscle
13 **Medial cutaneus branches of dorsal rami of spinal nerves** (C_7, C_8)
14 Longissimus thoracis muscle
15 Medial margin of scapula
16 **Rectus capitis posterior minor muscle**
17 **Obliquus capitis superior muscle**
18 **Rectus capitis posterior major muscle**
19 **Obliquus capitis inferior muscle**
20 Spinous process of axis
21 Semispinalis cervicis muscle
22 Semispinalis capitis muscle (divided and reflected)
23 Transverse cervical artery (superficial branch)
24 Serratus posterior superior muscle (divided and reflected)
25 Rhomboid minor muscle (divided and reflected)
26 Rhomboid major muscle (divided and reflected)

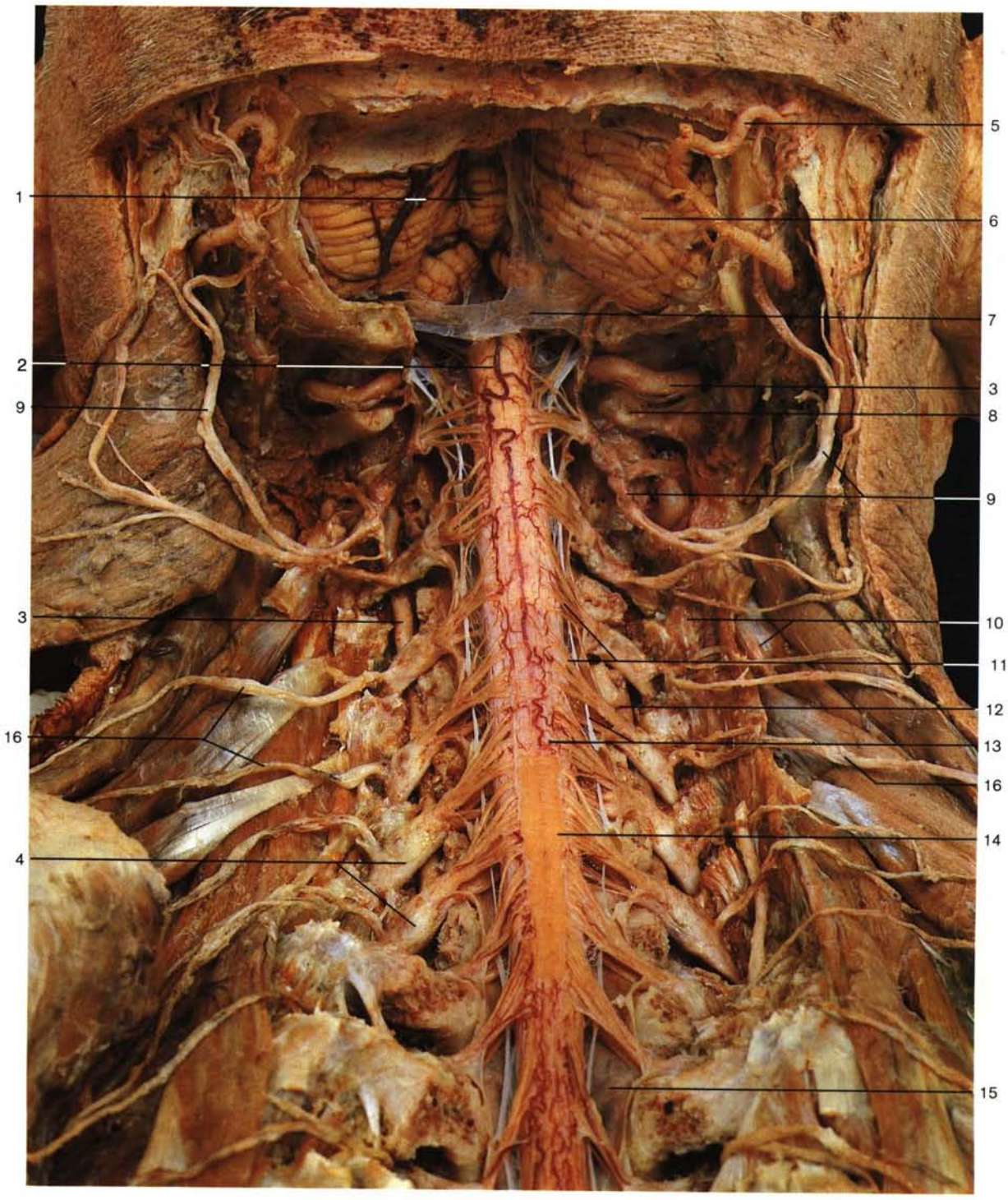

Neck, deepest layer. Spinal cord and medulla oblongata (dorsal aspect). Cranial cavity opened.

1 Vermis of the cerebellum
2 Medulla oblongata and posterior spinal artery
3 Vertebral artery
4 **Spinal ganglion**
5 Occipital artery
6 Cerebellum
7 **Cerebellomedullary cistern**
8 Atlas
9 Greater occipital nerve (C_2)
10 Levator scapulae muscle and intertransverse ligament
11 **Dorsal roots of spinal nerves**
12 Vertebral arch
13 **Denticulate ligament and arachnoid mater**
14 Area where pia mater has been removed
15 Dura mater
16 **Dorsal rami of spinal nerves**

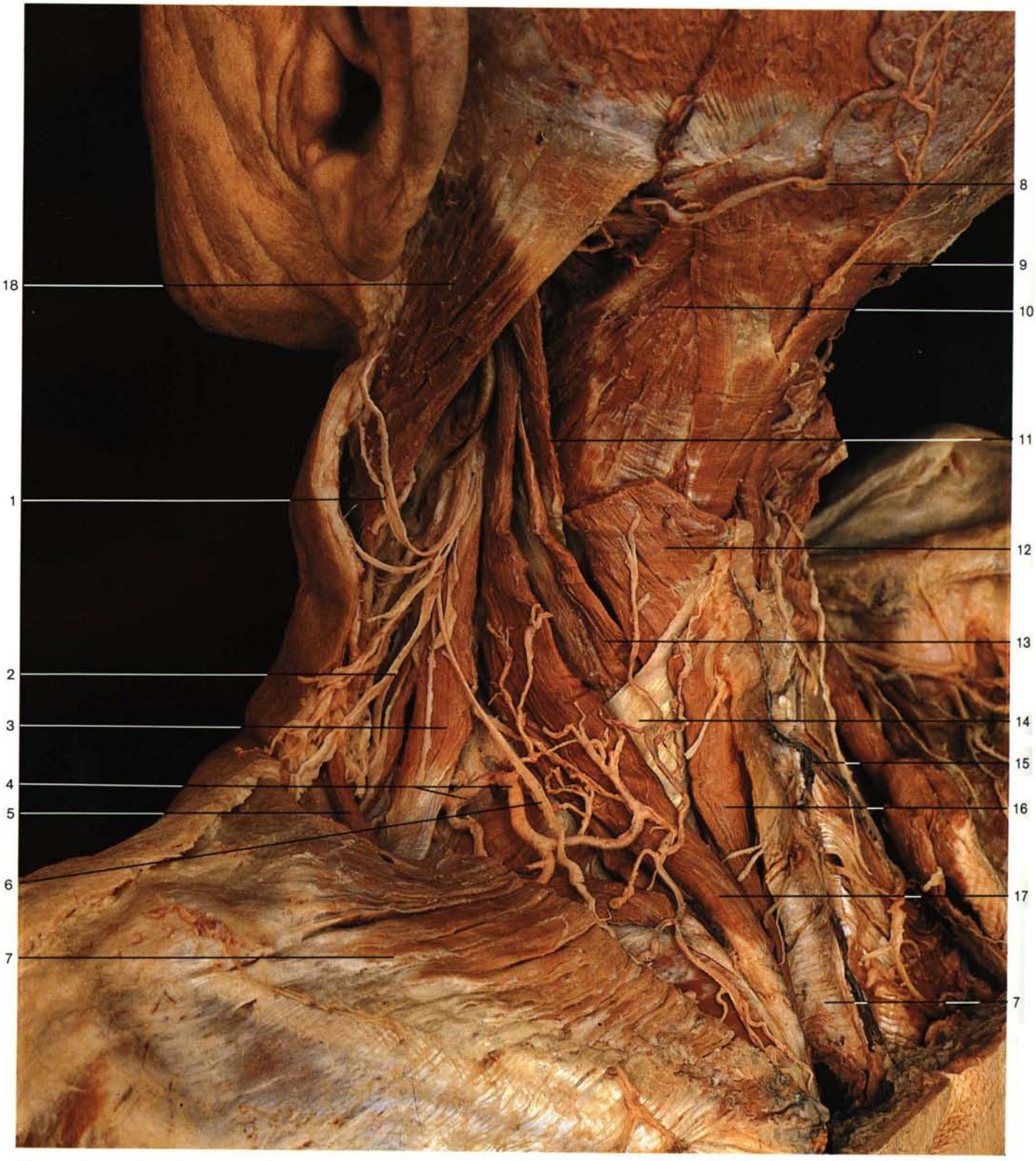

Deeper layer of the dorsal neck (oblique lateral aspect). Trapezius and semispinalis capitis muscles have been divided and partly removed.

1 Great auricular nerve
2 Accessory nerve (n. XI)
3 Scalenus posterior muscle
4 Transverse cervical artery
5 Omohyoid muscle
6 Muscular branch of cervical plexus
7 Trapezius muscle
8 Occipital artery
9 Greater occipital nerve
10 Semispinalis capitis muscle
11 Longissimus capitis muscle
12 Splenius capitis muscle
13 Splenius cervicis muscle
14 Serratus posterior superior muscle
15 Deep cervical vein (variant)
16 Rhomboid minor muscle
17 Levator scapulae muscle
18 Sternocleidomastoid muscle

1 Thyroid gland
2 Internal jugular vein
3 Right common carotid artery
4 Right axillary vein
5 Right brachiocephalic vein
6 Superior lobe of right lung
7 Right atrium
8 Right coronary artery
9 Middle lobe of right lung
10 **Right ventricle**
11 Diaphragm
12 Liver (left lobe) and falciform ligament
13 Left internal jugular vein
14 **Trachea**
15 Left brachiocephalic vein
16 Superior lobe of left lung
17 Cut edge of pericardium
18 **Ascending aorta**
19 **Pulmonary trunk**
20 Inferior lobe of left lung
21 **Left ventricle**
22 Costal margin
23 Right pulmonary artery
24 Esophagus
25 Descending aorta
26 Pericardium
27 Aortic valve
28 Thymus
29 Diaphragm

Thoracic organs, heart and lungs in situ (ventral aspect). Anterior thoracic wall, parietal pleura and pericardium had been removed.

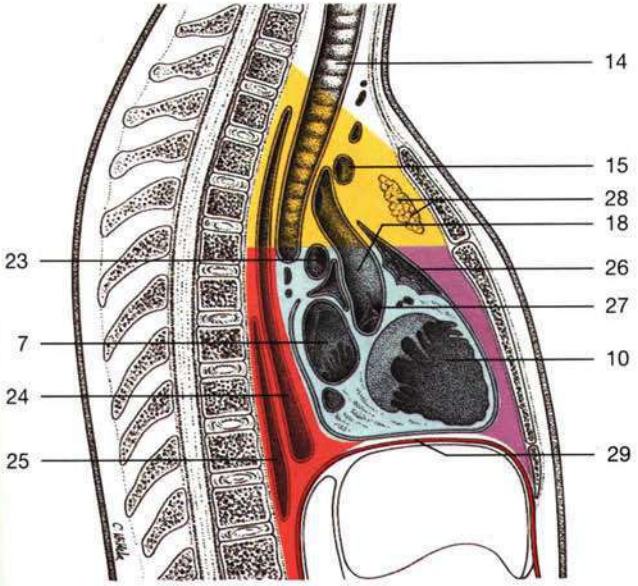

Sagittal section through thoracic cavity. The divisions of the mediastinum are indicated by colors.

Divisions of mediastinum	Main content
Superior mediastinum (yellow)	Trachea, brachiocephalic veins, thymus, aortic arch, esophagus, thoracic duct
Middle mediastinum (light blue)	Heart, ascending aorta, pulmonary trunk, pulmonary veins, phrenic nerves
Posterior mediastinum (red)	Esophagus with vagus nerves, descending aorta, thoracic duct, sympathetic trunks
Anterior mediastinum (pink)	Small vessels, connective and fatty tissue, and thymus in a child

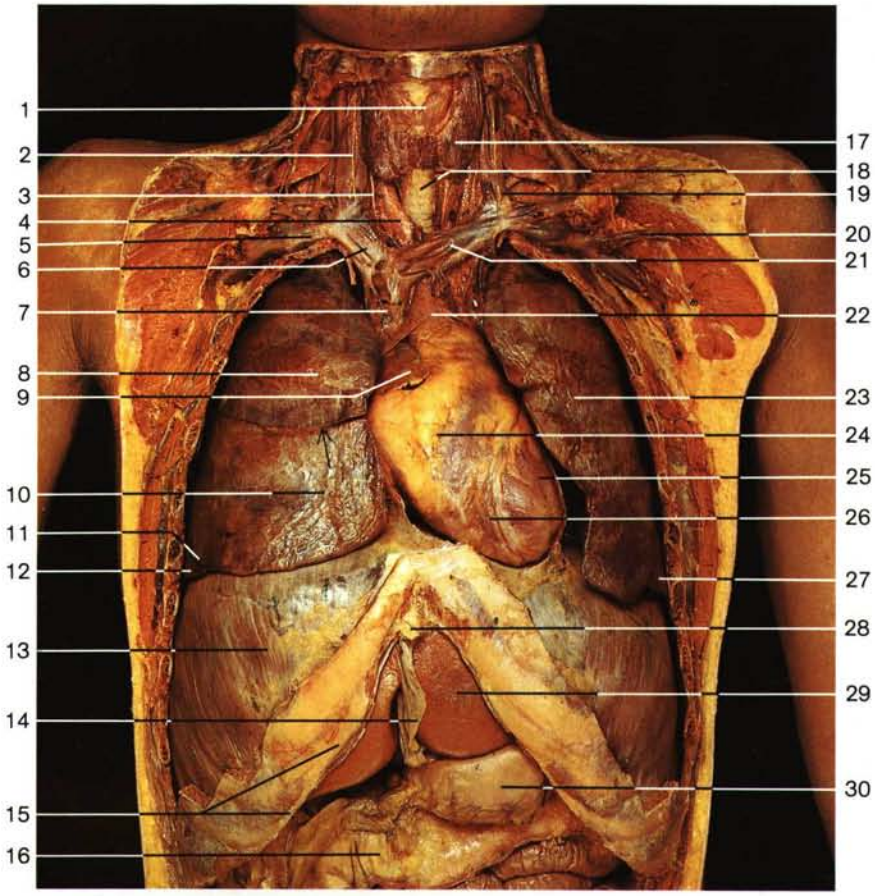

1 Cricothyroid muscle
2 Right internal jugular vein
3 Vagus nerve
4 Right common carotid artery
5 Right subclavian vein
6 Right brachiocephalic vein
7 Superior vena cava
8 Upper lobe of right lung
9 Right auricle
10 Middle lobe of right **lung**
11 Oblique fissure of right lung
12 Lower lobe of right lung
13 **Diaphragm**
14 Falciform ligament
15 Costal margin
16 Transverse colon
17 Thyroid gland
18 Trachea
19 Left internal jugular vein
20 Left cephalic vein
21 Left brachiocephalic vein
22 Pericardium (cut edge)
23 Upper lobe of left **lung**
24 **Right ventricle**
25 **Left ventricle**
26 Anterior interventricular sulcus
27 Lower lobe of left lung
28 Xiphoid process
29 **Liver**
30 **Stomach**
31 Pectoralis major muscle

Positions of thoracic organs. The anterior thoracic wall has been removed. Arrow: horizontal fissure of right lung.

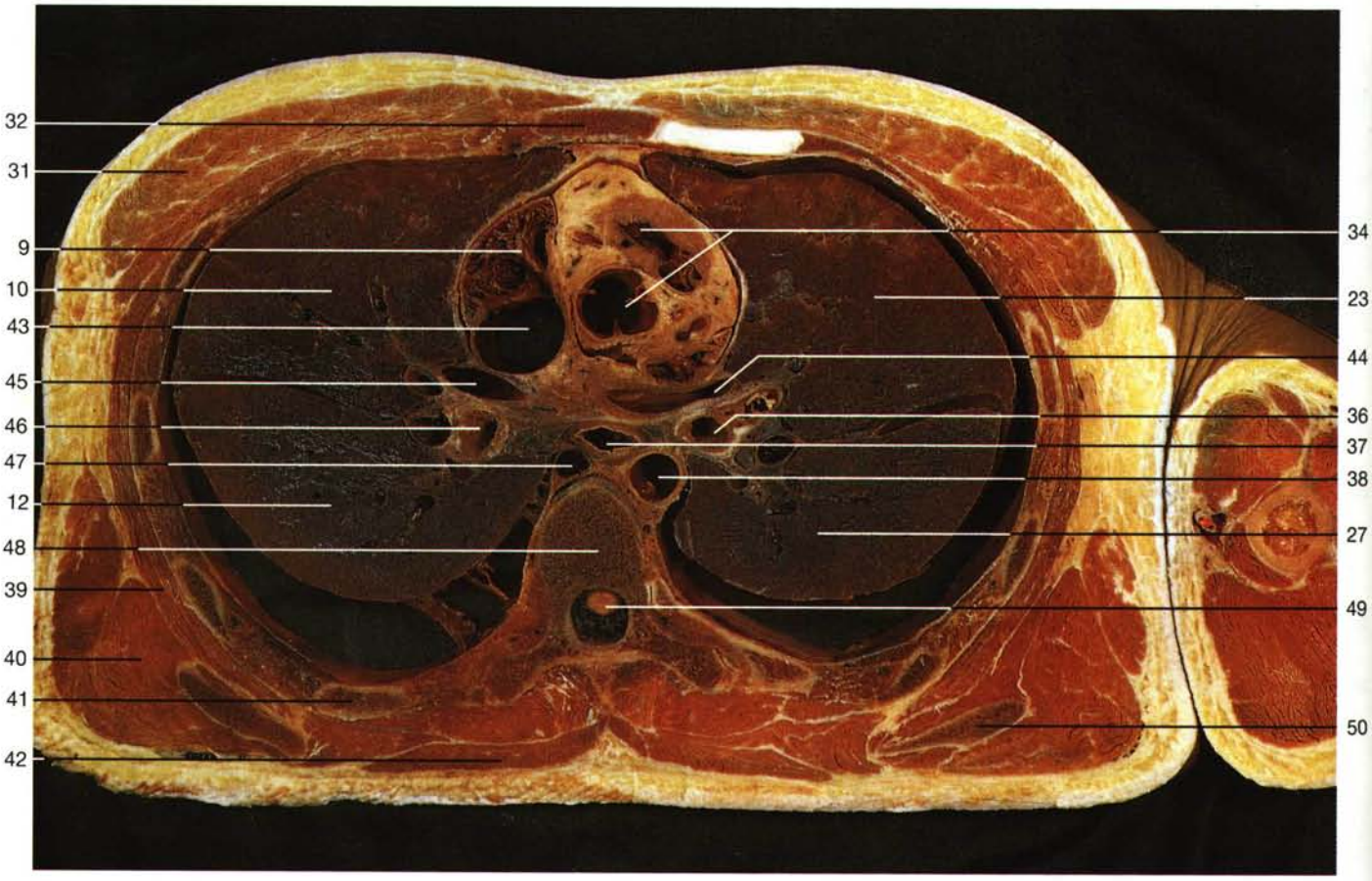

Horizontal section through the thorax at the level of the 7th thoracic vertebra (from below).

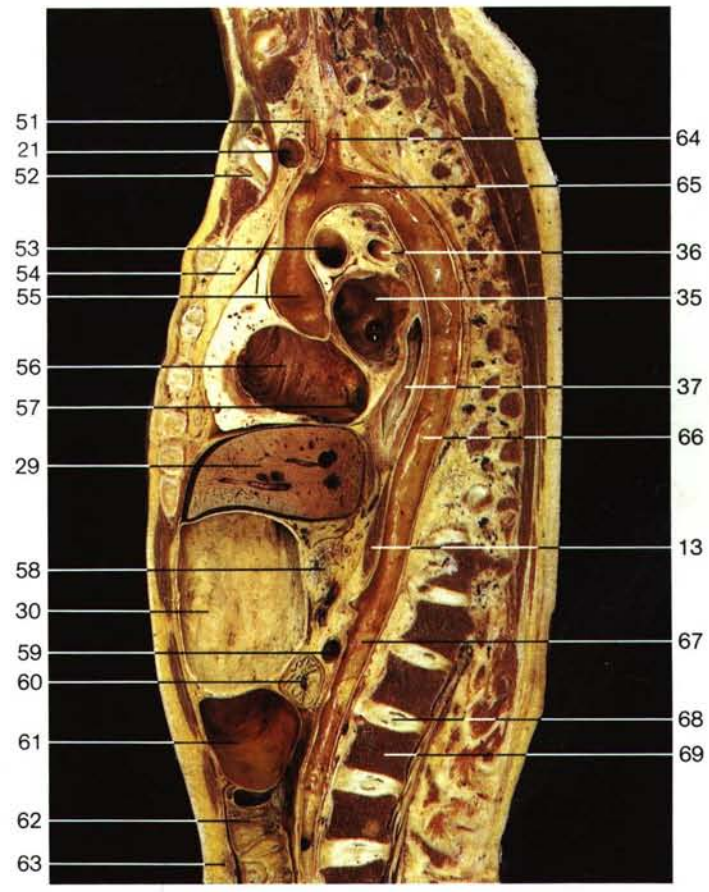

Sagittal section through the left thorax, 2 cm lateral to the median plane.

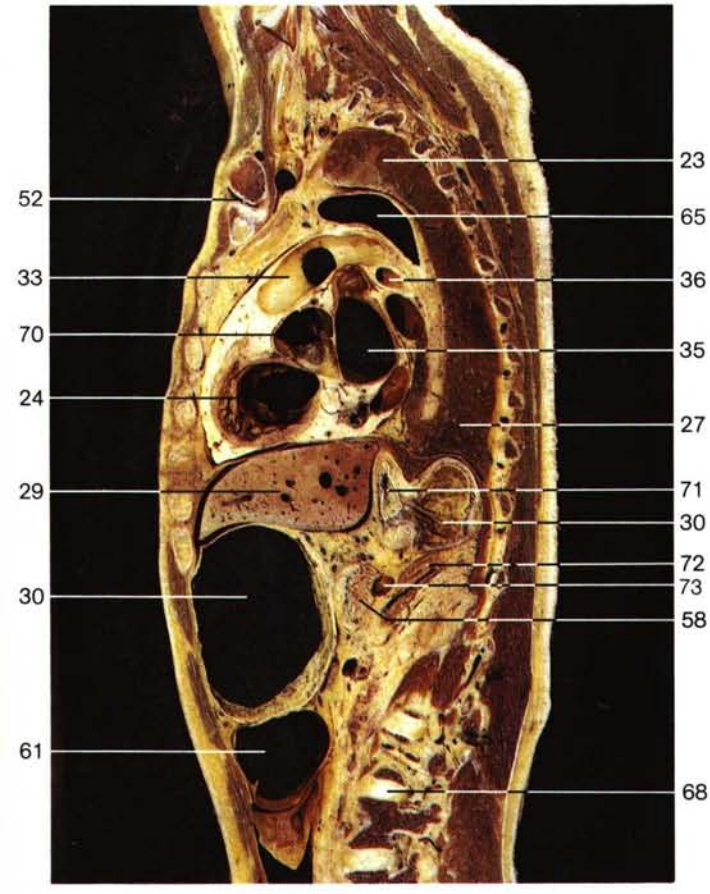

Sagittal section through the left thorax, 3.5 cm lateral to the median plane.

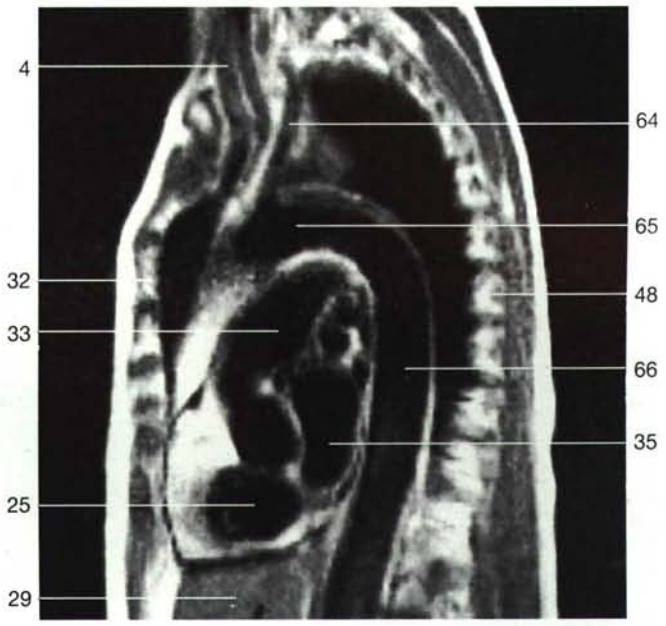

Sagittal section through the thorax. MR-Scan.

32 Sternum
33 Pulmonary trunk
34 Left ventricle and bulb of aorta
35 Left atrium
36 Left main bronchus
37 **Esophagus**
38 Descending aorta
39 Serratus anterior muscle
40 Teres major muscle
41 Rib
42 Trapezius muscle
43 Right atrium
44 Left pulmonary vein
45 Right pulmonary vein
46 Right main bronchus
47 Azygos vein
48 Body of vertebra
49 Spinal cord
50 Scapula
51 Left common carotid artery
52 Sternoclavicular articulation with disc
53 Right pulmonary artery
54 Remnants of thymus
55 Aortic bulb
56 Right atrium
57 Entrance of inferior vena cava in right atrium
58 **Pancreas**
59 Left renal vein
60 **Duodenum**
61 Transverse colon (dilated)
62 Small intestine
63 Umbilicus
64 Left subclavian artery
65 Aortic arch
66 **Thoracic aorta**
67 Abdominal aorta
68 Intervertebral disc
69 Body of lumbar vertebra
70 Aortic valve
71 Cardia of stomach
72 Suprarenal gland
73 Splenic vein

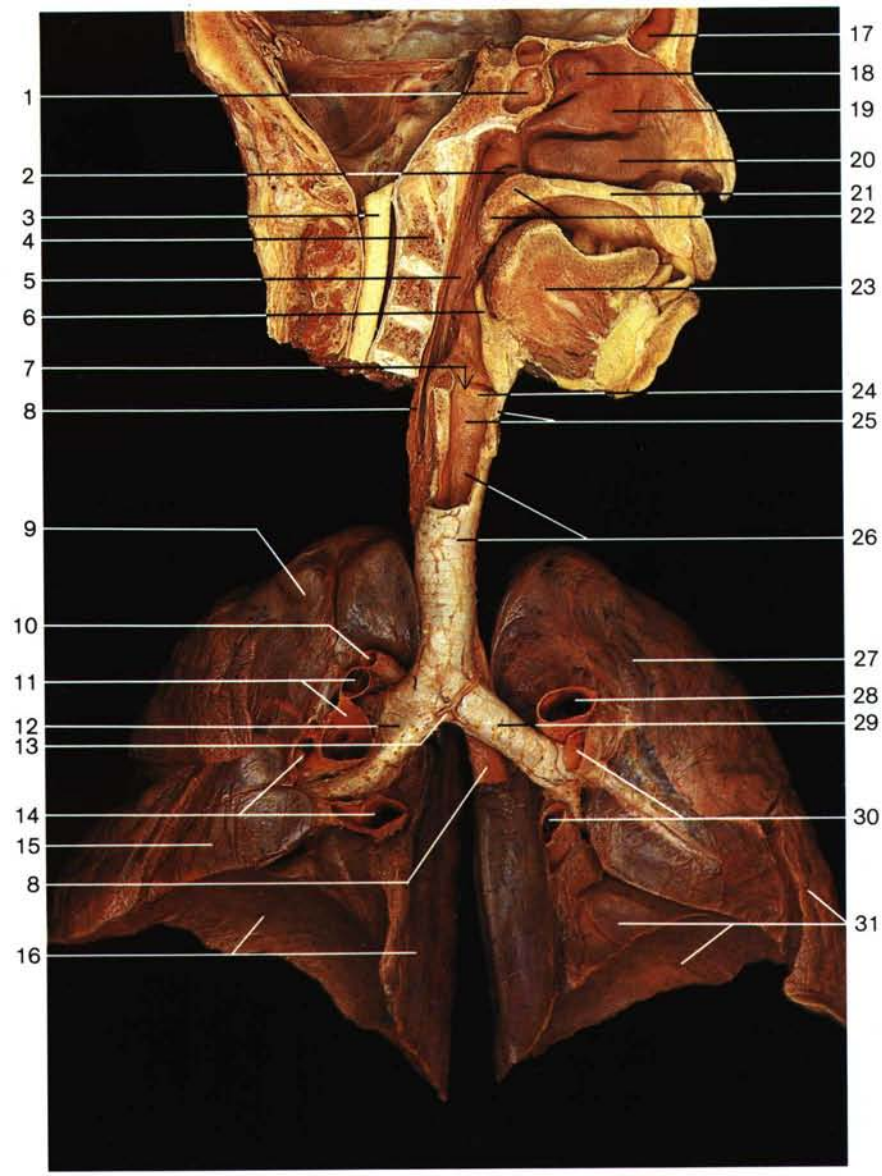

1 Sphenoid sinus
2 Pharyngeal opening of auditory tube
3 Spinal cord
4 Dens of axis
5 **Oropharynx** (oropharyngeal isthmus)
6 Epiglottis
7 Entrance of larynx
8 Esophagus
9 Upper lobe of right lung
10 Azygos vein
11 Branches of pulmonary artery
12 **Right main bronchus**
13 **Bifurcation of trachea**
14 Tributaries of right pulmonary veins
15 Middle lobe of right lung
16 Lower lobe of right lung
17 Frontal sinus
18 Superior nasal concha
19 Middle nasal concha
20 Inferior nasal concha
21 Hard palate
22 Soft palate with uvula
23 Tongue
24 Vocal fold
25 Larynx
26 Trachea
27 Upper lobe of left lung
28 Left pulmonary artery
29 **Left main bronchus**
30 Left pulmonary veins
31 Lower lobe of left lung

Respiratory system. The lungs have been fixed in expiration and turned laterally. Head bisected and turned laterally.

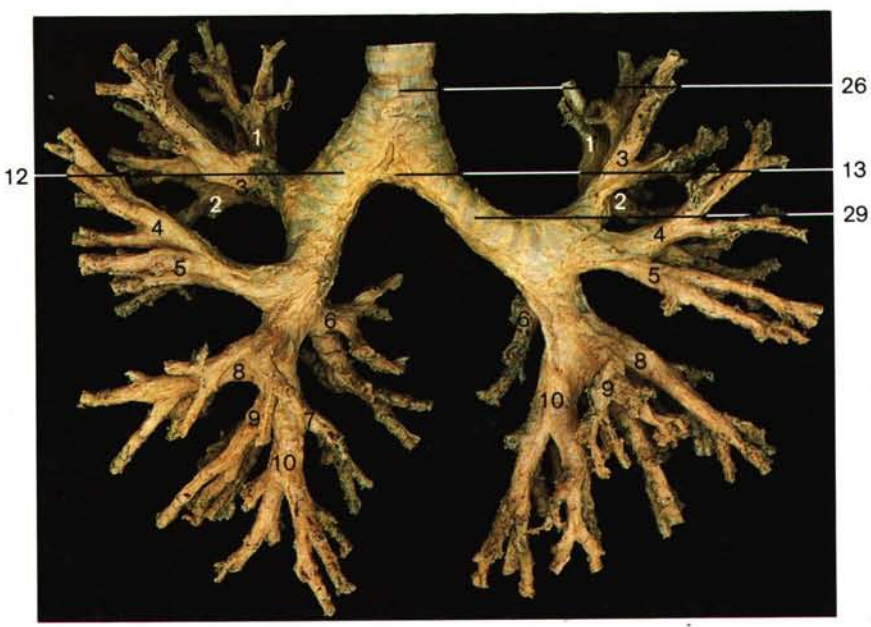

Bronchial tree (ventral aspect). The lung tissue has been removed. The bronchopulmonary segments are numbered 1–10.

▷
To page 231:

1 Nasal cavity
2 Pharynx
3 Larynx (thyroid cartilage)
4 Trachea
5 Upper lobe of right lung
6 Bifurcation of trachea
7 Right main bronchus
8 Horizontal fissure of right lung
9 Middle lobe of right lung
10 Oblique fissures of lungs
11 Lower lobe of right lung
12 Clavicle
13 Upper lobe of left lung
14 Left main bronchus
15 Bronchi supplying bronchopulmonary segments
16 Lower lobe of left lung
17 Costal margin
18 Hyoid bone
19 Right superior lobe bronchus
20 Right middle lobe bronchus
21 Right inferior lobe bronchus
22 Left superior lobe bronchus
23 Left inferior lobe bronchus
24 Segmental bronchi
25 Branches of pulmonary arteries
26 Branches of pulmonary veins

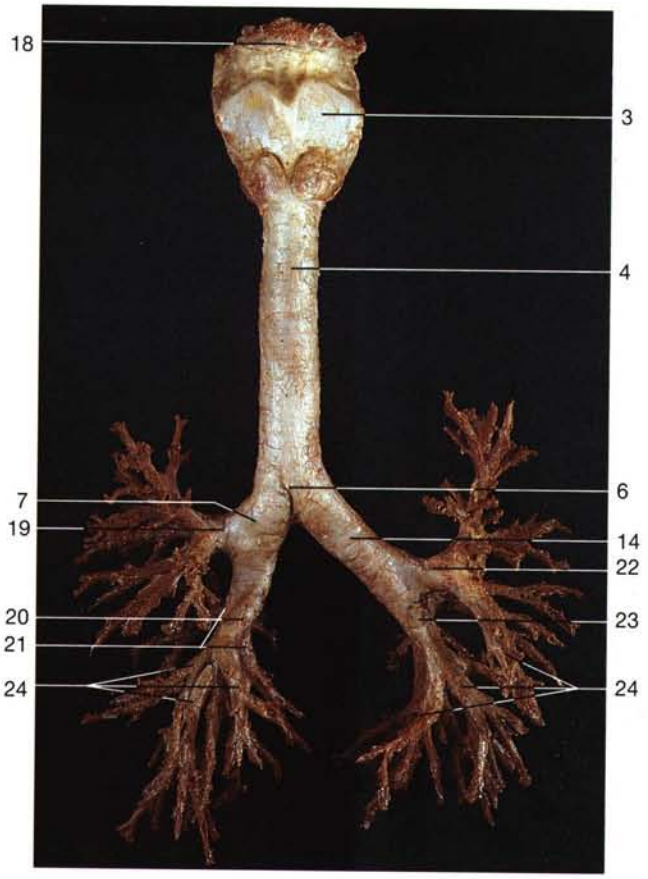

Larynx, trachea and bronchial tree (anterior aspect).

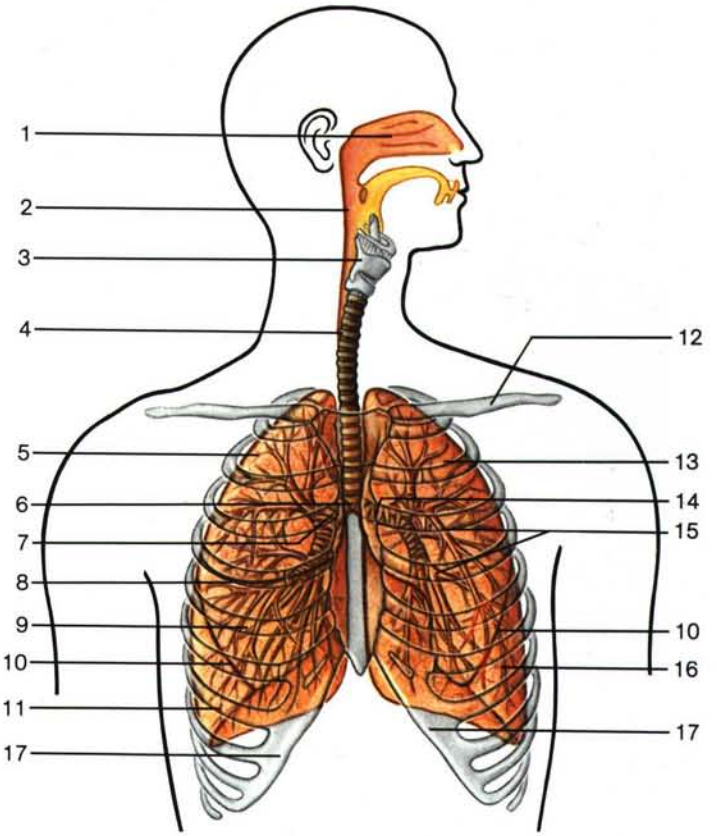

Organization and positions of respiratory organs. (Schematic drawing.)

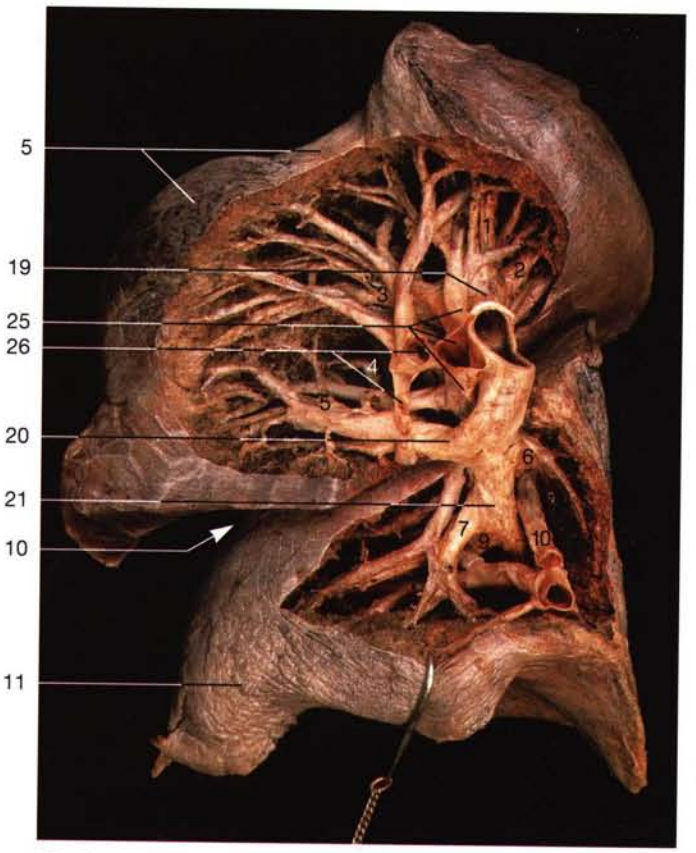

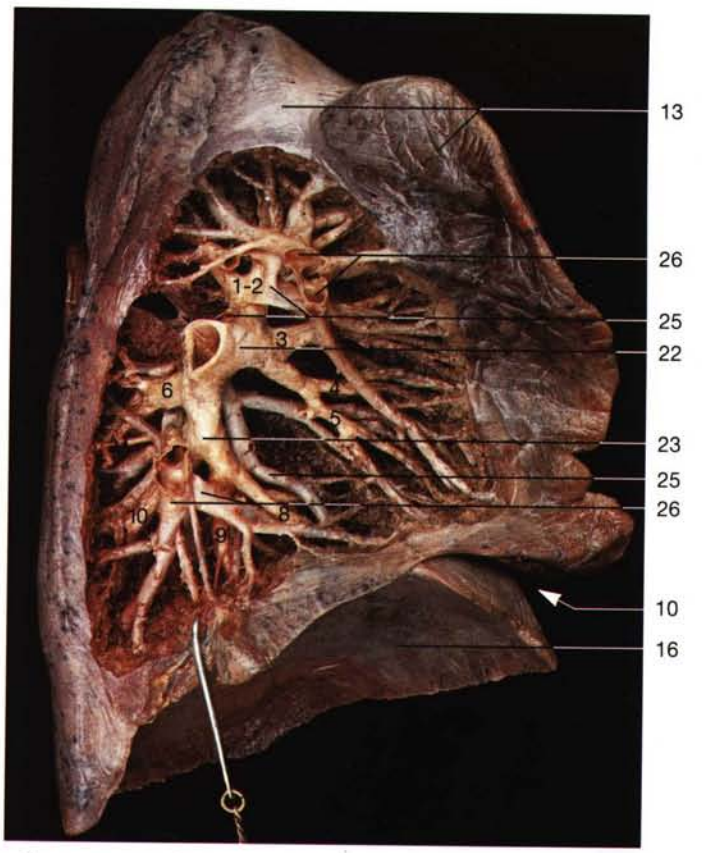

Mediastinal dissection of the bronchial tree, pulmonary veins, and pulmonary arteries of right lung (left) and left lung (right) (medial aspect). Segmental bronchi are numbered 1–10.

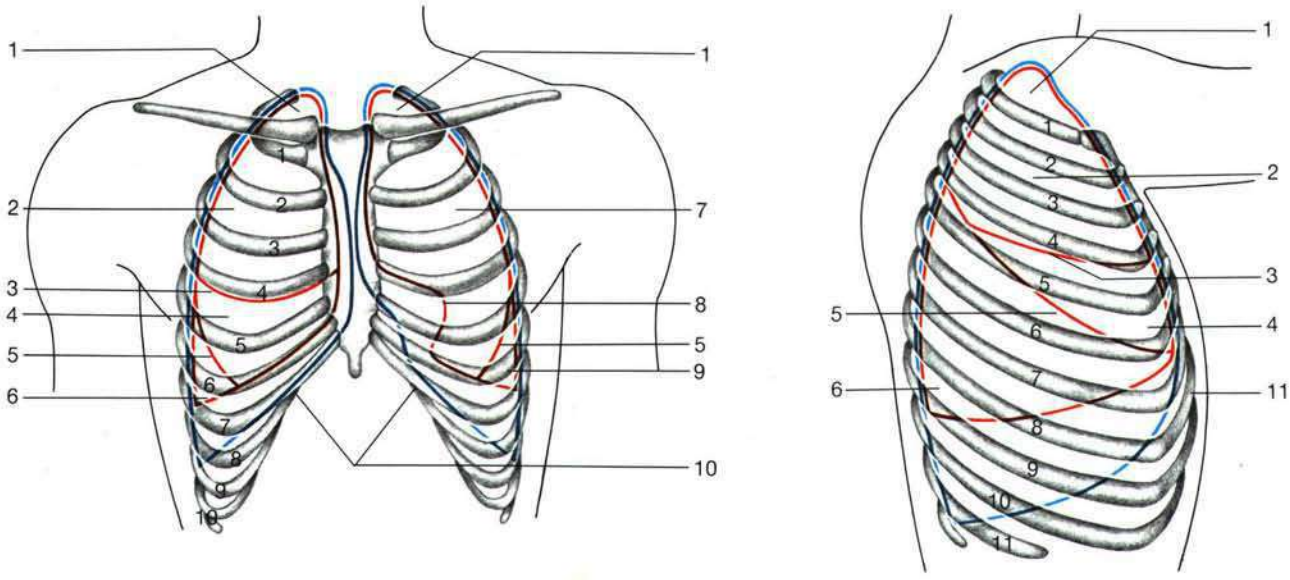

Surface projections of lungs and pleura on the thoracic wall. Left: anterior aspect; right: right-lateral aspect. Red = margins of the lung; blue = margins of pleura. The numbers indicate ribs.

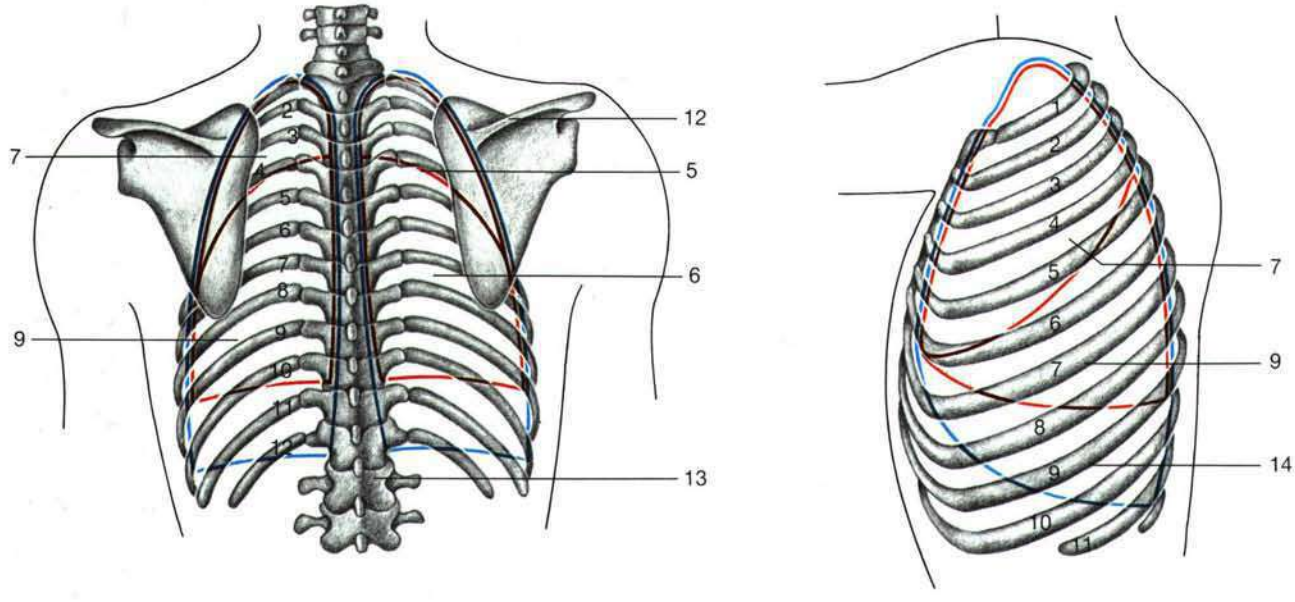

Surface projections of lungs and pleura on thoracic wall. Left: posterior aspect; right: left-lateral aspect. Red = margins of lung; blue = margins of pleura. The number indicate ribs.

1 Apex of lung
2 Upper lobe of right lung
3 Horizontal fissure of right lung
4 Middle lobe of right lung
5 Oblique fissures of lungs
6 Lower lobe of right lung
7 Upper lobe of left lung
8 Cardiac notch of left lung
9 Lower lobe of left lung
10 Infrasternal angle
11 Costal margin
12 Spine of scapula
13 1st lumbar vertebra
14 Space between border of lung and pleura (costodiaphragmatic recess)

Right lung (lateral aspect).

Left lung (lateral aspect).

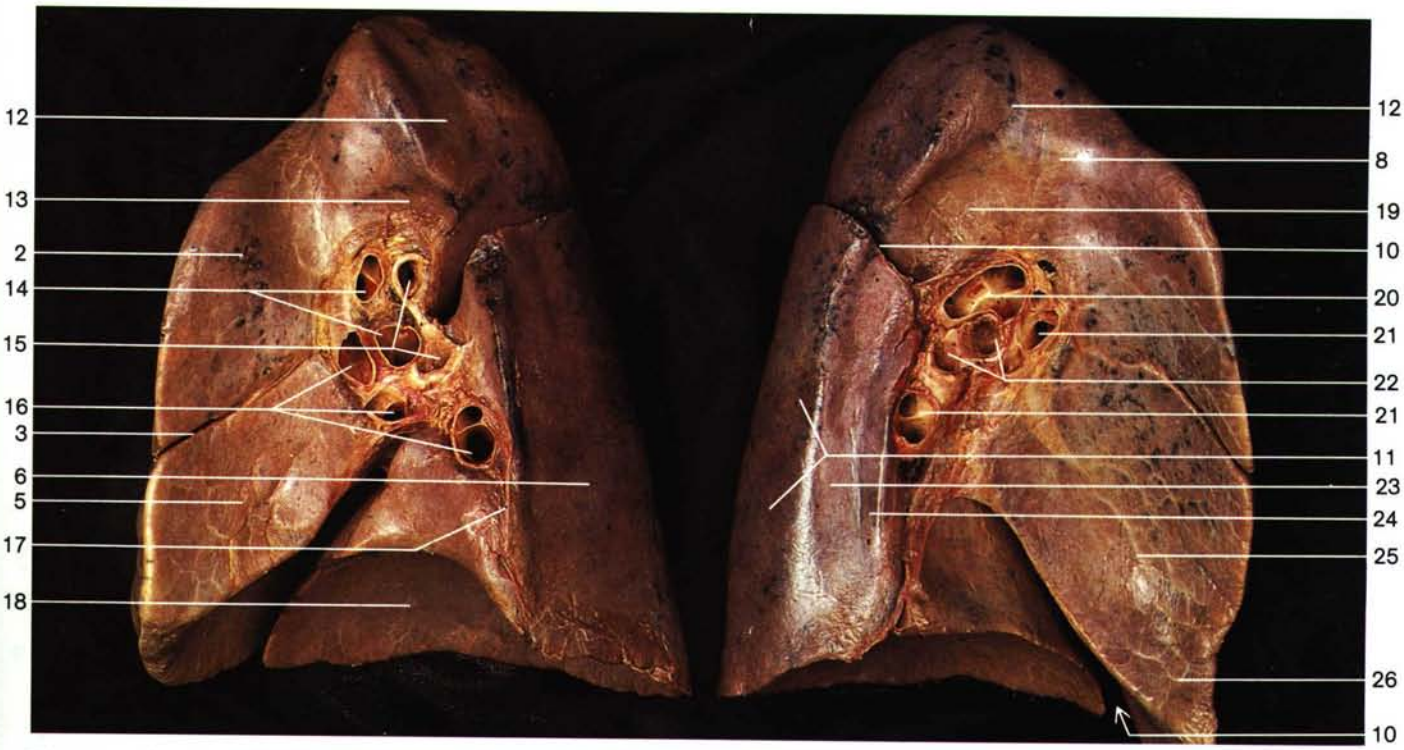

Right lung (medial aspect).

Left lung (medial aspect).

1 Apex of lung
2 Upper lobe of right lung
3 Horizontal fissure of right lung
4 Oblique fissure of right lung
5 Middle lobe of right lung
6 Lower lobe of right lung
7 Inferior border
8 Upper lobe of left lung
9 Impressions of ribs
10 Oblique fissure of left lung
11 Lower lobe of left lung
12 Groove of subclavian artery
13 Groove of azygos arch
14 Branches of right pulmonary artery
15 Bronchi
16 Right pulmonary veins
17 Pulmonary ligament
18 Diaphragmatic surface
19 Groove of aortic arch
20 Left pulmonary artery
21 Branches of left pulmonary veins
22 Left secondary bronchi
23 Groove of thoracic aorta
24 Groove of esophagus
25 Cardiac impression
26 Lingula

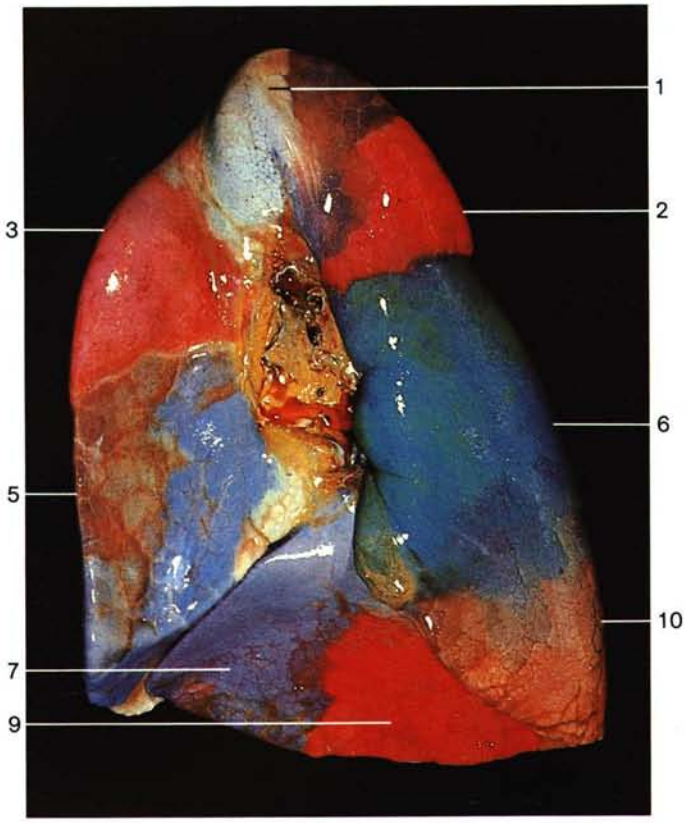

Right lung (medial aspect).

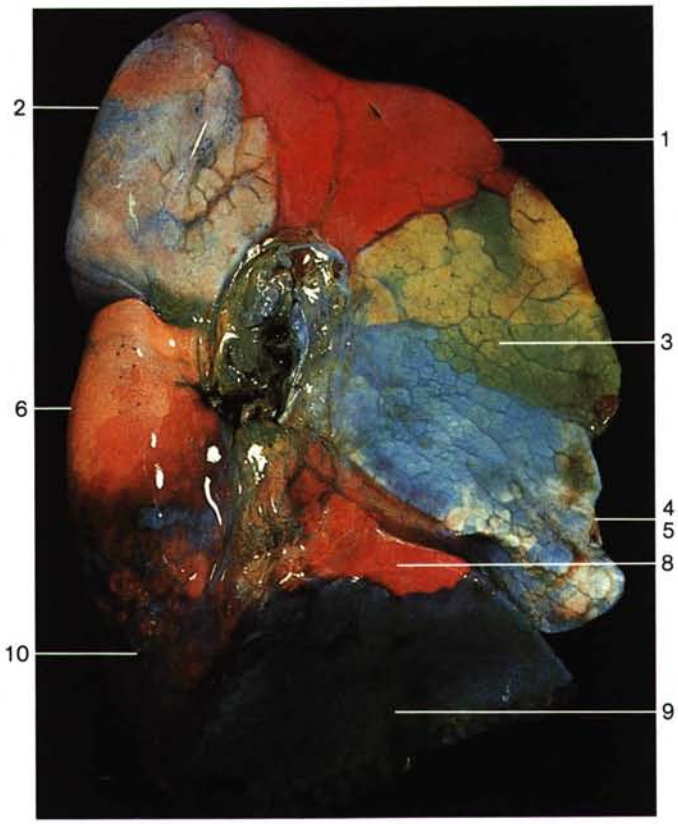

Left lung (medial aspect).

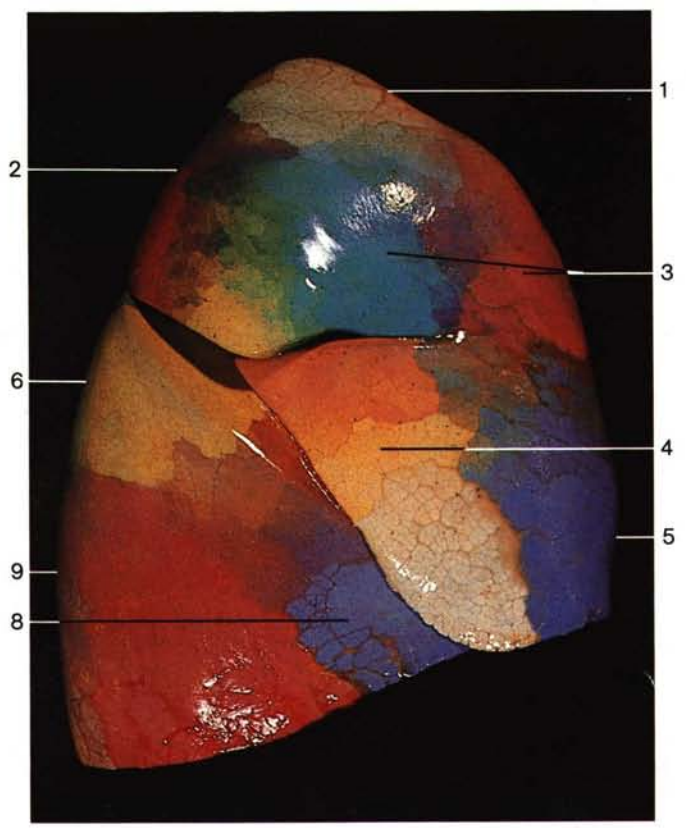

Right lung (lateral aspect).

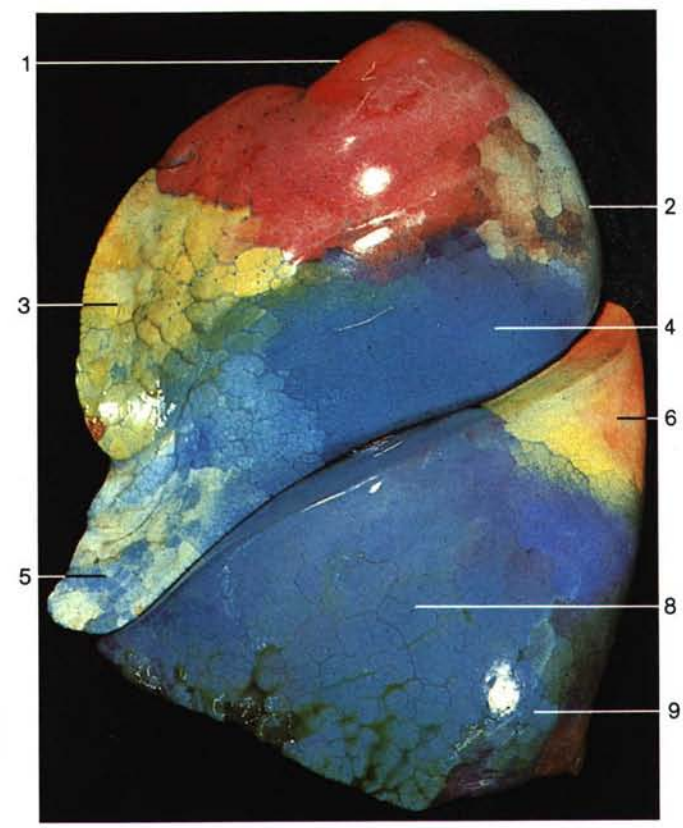

Left lung (lateral aspect).

The bronchopulmonary segments of the lungs are differentiated by the various colors. Notice that there is no segment in the left lung that corresponds to the 7th segment of the right lung. Compare with the schematic drawing on the facing page.

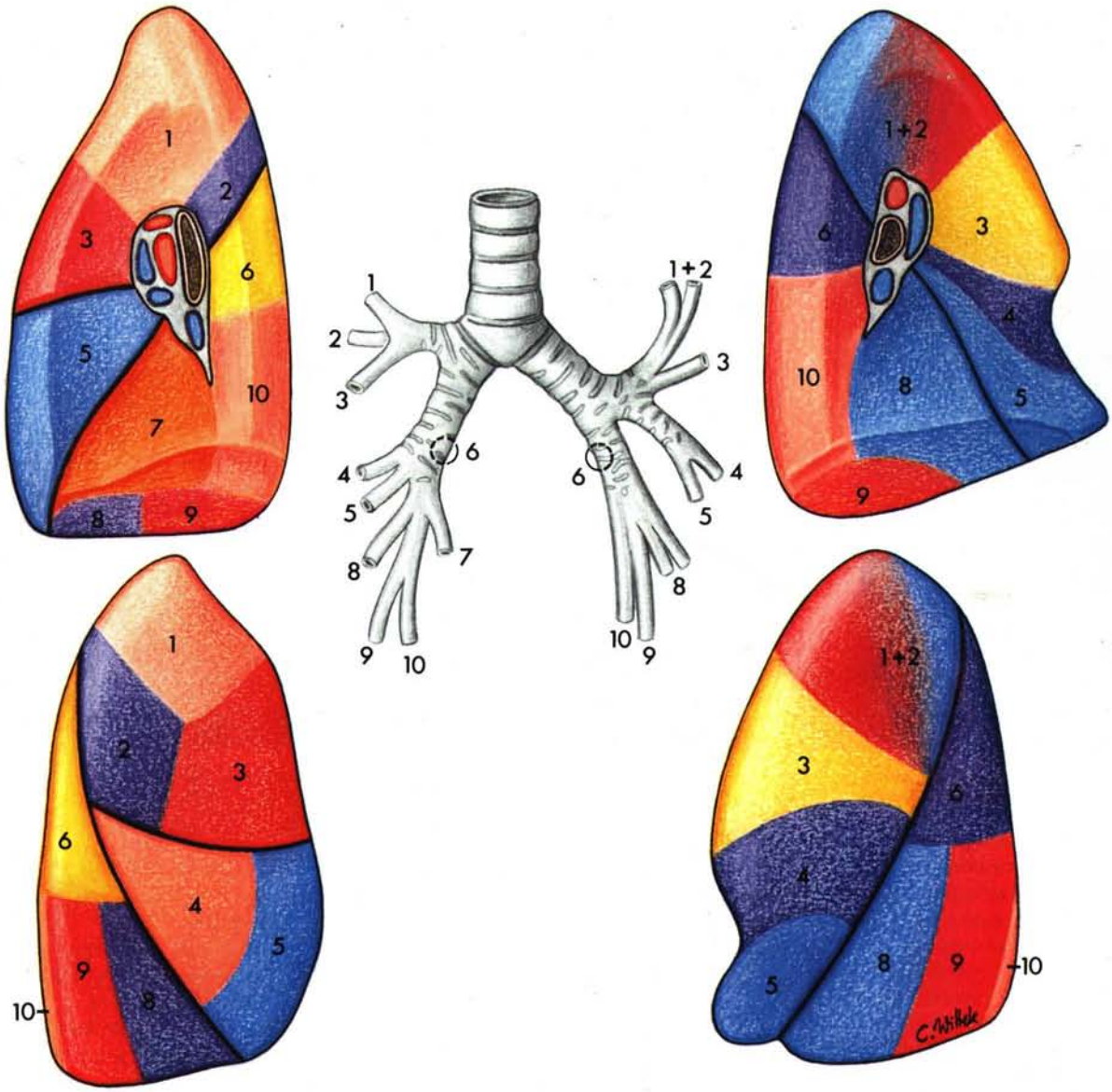

Distribution of bronchopulmonary segments of the lungs and their relation to the bronchial tree (after J. F. Huber).
The bronchopulmonary segments are morphologically and functionally separate independent respiratory units of the lung tissue. Each segment is surrounded by connective tissue which is continuous with the visceral pleura. The segmental bronchi in a segment are central, closely accompanied by branches of the pulmonary arteries whereas the tributaries of the pulmonary veins run **between** the segments. Thus, the veins serve two adjacent segments which drain for the most part into more than one vein. A bronchopulmonary segment is therefore not a complete vascular unit, but segmentation is the result of a specific architecture of the lung vasculature.

Right lung	**Left lung**
1 Apical segment 2 Posterior segment 3 Anterior segment } Upper lobe bronchus 4 Lateral segment 5 Medial segment } Middle lobe bronchus	1+2 Apico-posterior segment 3 Anterior segment } Superior division 4 Superior lingular segment 5 Inferior lingular segment } Inferior division } Upper lobe bronchus
6 Superior (apical) segment 7 Medial basal segment 8 Anterior basal segment 9 Lateral basal segment 10 Posterior basal segment } Lower lobe bronchus	6 Superior (apical) segment 7 Absent 8 Anteromedial basal segment 9 Lateral basal segment 10 Posterior basal segment } Lower lobe bronchus

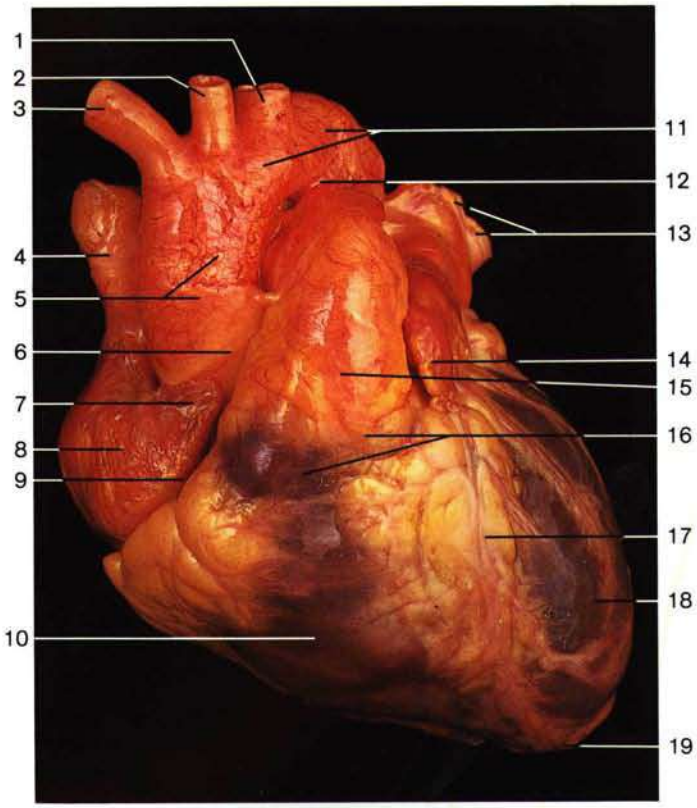

Heart of 30-year-old woman (anterior aspect).

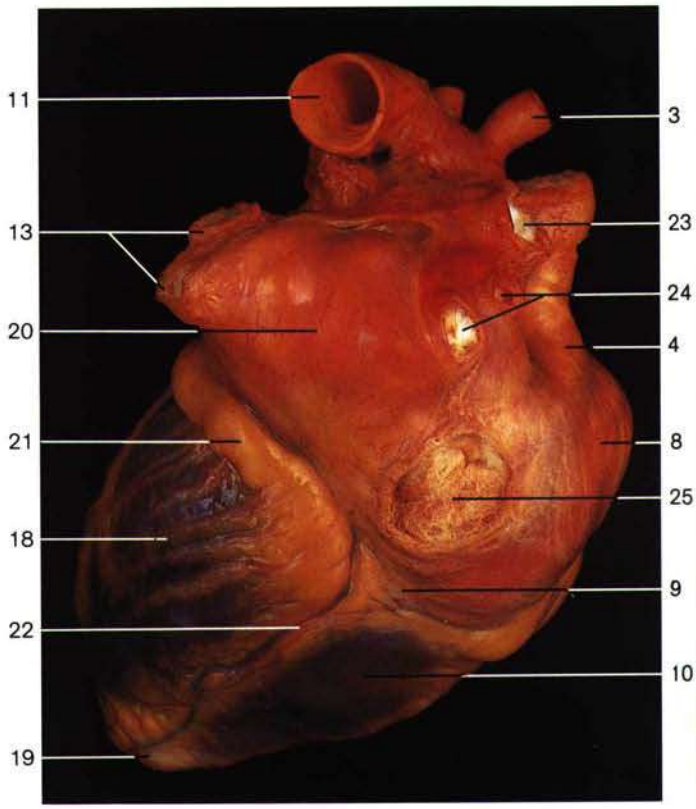

Heart of 30-year-old woman (oblique posterior view).

1 Left subclavian artery
2 Left common carotid artery
3 Brachiocephalic trunk
4 Superior vena cava
5 Ascending aorta
6 Bulb of the aorta
7 Right auricle
8 **Right atrium**
9 Coronary sulcus
10 **Right ventricle**
11 Aortic arch
12 Ligamentum arteriosum
13 Left pulmonary veins
14 Left auricle
15 **Pulmonary trunk**
16 Sinus of pulmonary trunk
17 Anterior interventricular sulcus
18 **Left ventricle**
19 Apex of the heart
20 **Left atrium**
21 Epicardial fat overlying coronary sinus
22 Posterior interventricular sulcus
23 Right pulmonary artery
24 Right pulmonary veins
25 Inferior vena cava

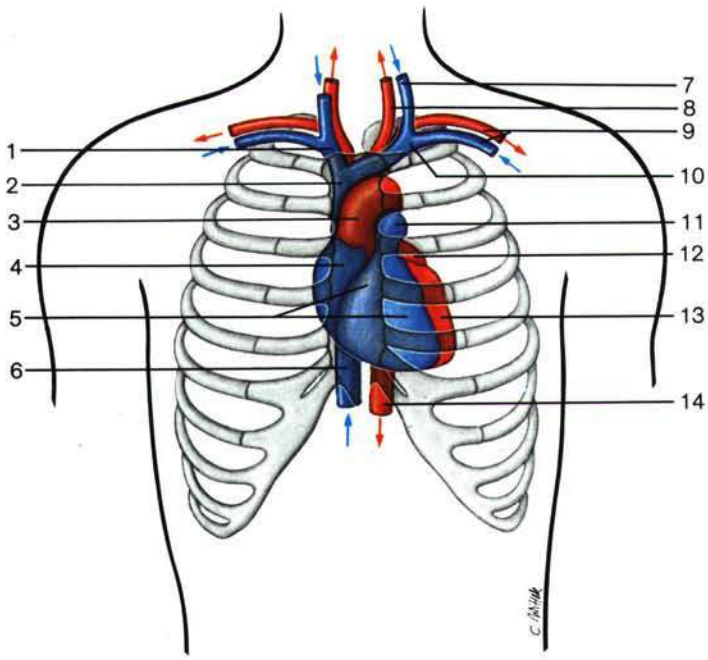

Position of heart and its vessels within the thorax. (Schematic drawing.)

1 Right brachiocephalic vein
2 Superior vena cava
3 Ascending aorta
4 Right atrium
5 Right ventricle
6 Inferior vena cava
7 Left internal jugular vein
8 Left common carotid artery
9 Left axillary artery and vein
10 Left brachiocephalic vein
11 Pulmonary trunk
12 Left auricle
13 Left ventricle
14 Descending aorta

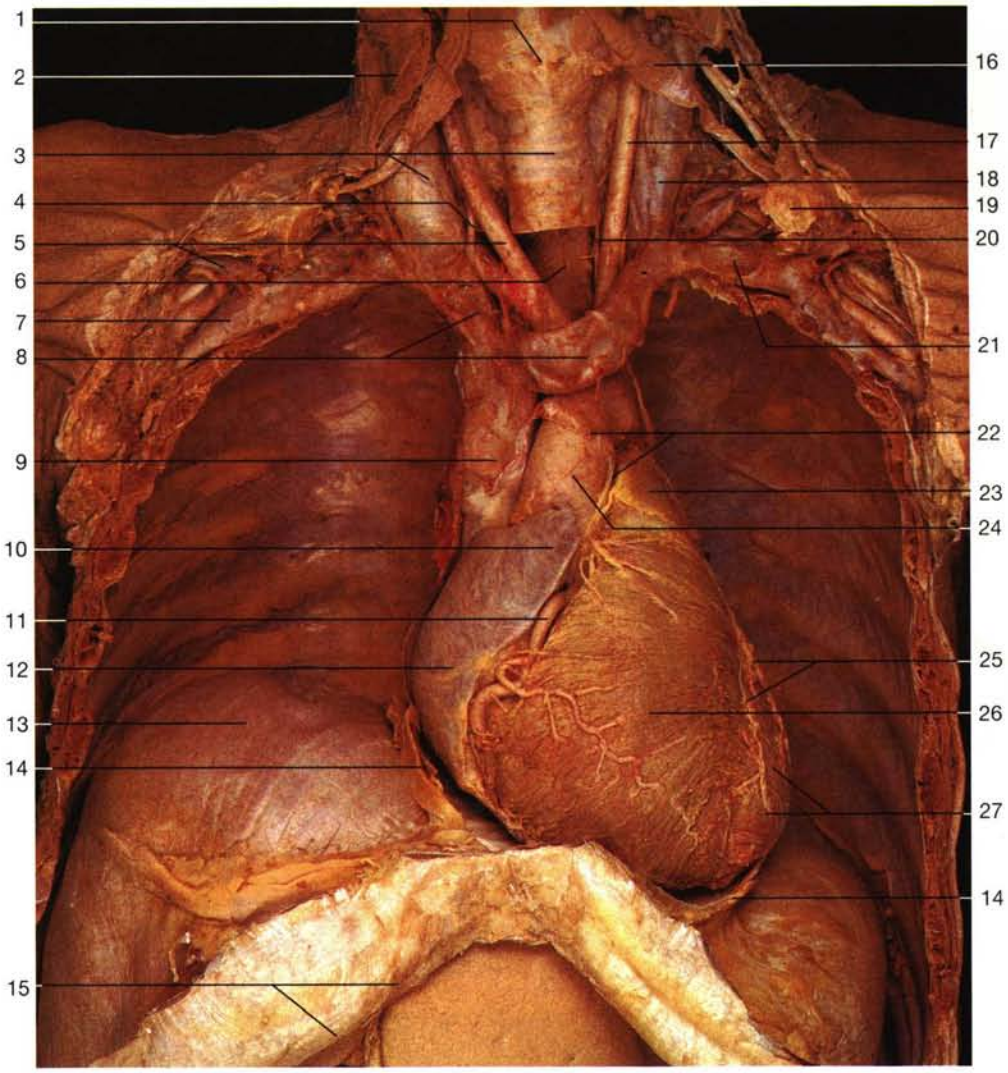

Heart and related vessels in situ (anterior aspect). Anterior thoracic wall, pericardium and epicardium have been removed; trachea divided.

1 Larynx (thyroid cartilage)
2 Sternocleidomastoid muscle (divided)
3 Trachea (divided) and right internal jugular vein
4 Vagus nerve
5 Right common carotid artery and cephalic vein
6 Esophagus
7 Right axillary vein
8 Right and left brachiocephalic veins
9 **Superior vena cava**
10 Right auricle
11 Right **coronary artery**
12 **Right atrium**
13 Diaphragm
14 Pericardium (cut edges)
15 Costal margin
16 Omohyoid muscle
17 Left **common carotid artery**
18 Left internal jugular vein
19 Clavicle (divided)
20 Left recurrent laryngeal nerve
21 **Subclavian vein**
22 Pericardial reflection
23 **Pulmonary trunk**
24 **Ascending aorta**
25 Anterior interventricular sulcus and anterior interventricular branch of left coronary artery
26 **Right ventricle**
27 **Left ventricle**
28 Aortic valve
29 Tricuspid or right atrioventricular valve
30 Inferior vena cava
31 Pulmonary veins
32 Pulmonary valve
33 Left atrioventricular (bicuspid or mitral) valve

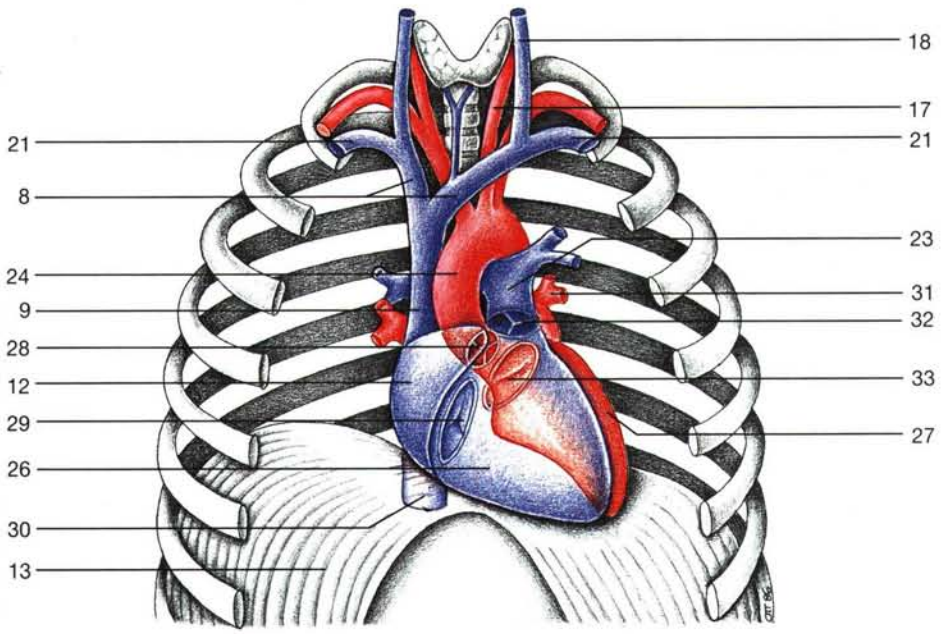

Heart in situ. Position of valves (anterior aspect). (Schematic drawing.)

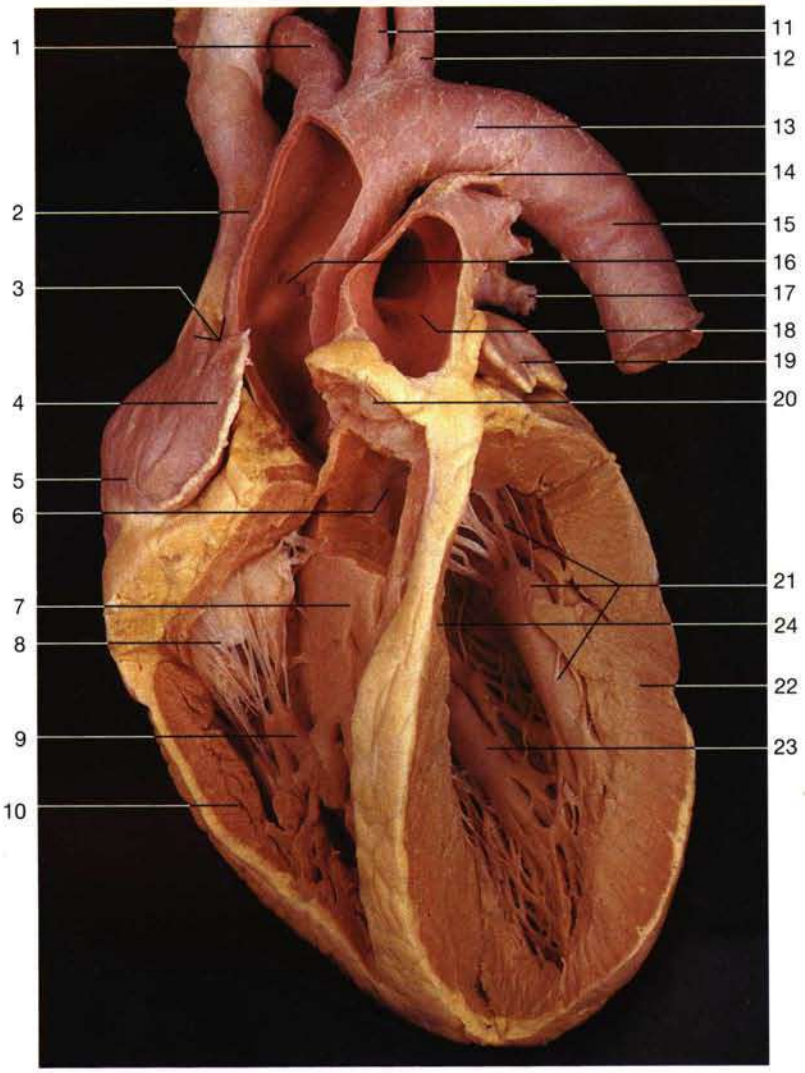

Anterior aspect of the heart. The anterior walls of the ventricles and of the aorta and pulmonary trunk have been fenestrated to show the aortic valve.

1 Brachiocephalic trunk
2 Superior vena cava
3 Sulcus terminalis
4 Right auricle
5 Right atrium
6 **Aortic valve**
7 Conus arteriosus (interventricular septum)
8 **Right atrioventricular (tricuspid) valve**
9 **Anterior papillary muscle**
10 Myocardium of right ventricle
11 Left common carotid artery
12 Left subclavian artery
13 **Aortic arch**
14 Ligamentum arteriosum (remnant of ductus arteriosus)
15 Thoracic aorta (descending aorta)
16 **Ascending aorta**
17 Left pulmonary vein
18 **Pulmonary trunk**
19 Left auricle
20 Pulmonic valve
21 **Anterior papillary muscle** with chordae tendineae
22 Myocardium of left ventricle
23 Posterior papillary muscle
24 Interventricular septum
25 Right and left brachiocephalic veins
26 **Chordae tendineae**
27 Papillary muscles of right ventricle
28 Left atrium
29 Infundibulum
30 Anterior papillary muscle of left ventricle
31 Left atrioventricular (bicuspid or mitral) valve and chordae tendineae
32 Apex of heart

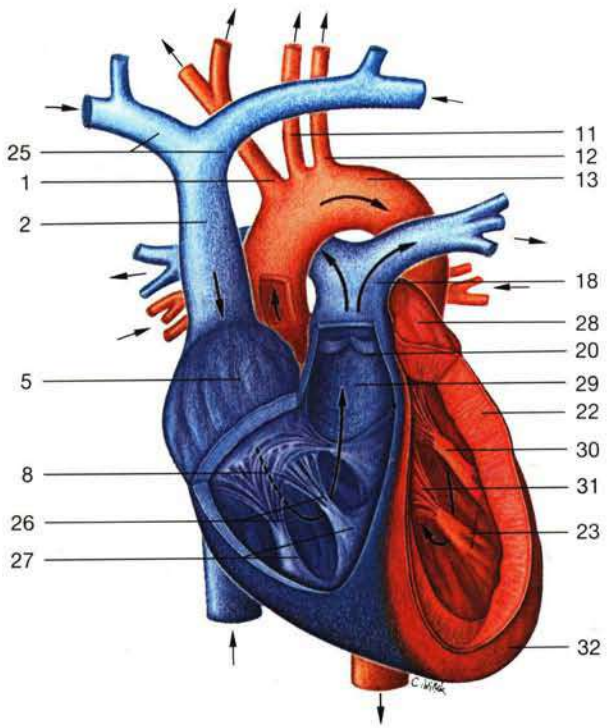

Circulation within the heart. (Schematic drawing.) The arrows indicate the direction of the blood flow through the heart. Blue = right heart; red = left heart.

Heart in situ. Myocardium and coronary arteries (anterior aspect).

1 Internal jugular vein
2 Common carotid artery
3 Brachiocephalic trunk
4 **Ascending aorta**
5 Right lung
6 Right auricle
7 **Right coronary artery**
8 **Myocardium of right ventricle**
9 Diaphragm
10 Costal margin
11 Thyroid gland and internal jugular vein
12 Trachea and left common carotid artery
13 Left brachiocephalic vein
14 Left lung
15 **Pericardium** (cut edge)
16 Pulmonary trunk
17 Anterior interventricular artery
18 **Myocardium of left ventricle**
19 **Muscular vortex** (right ventricle)
20 Posterior interventricular sulcus
21 Anterior interventricular sulcus
22 **Muscular vortex** (left ventricle)
23 Aortic arch
24 Left atrium
25 Coronary sinus
26 Superior vena cava
27 Right pulmonary vein
28 Right atrium
29 Inferior vena cava
30 Coronary sulcus
31 Myocardium of left ventricle
32 Left pulmonary artery
33 Left pulmonary vein
34 Apex of heart

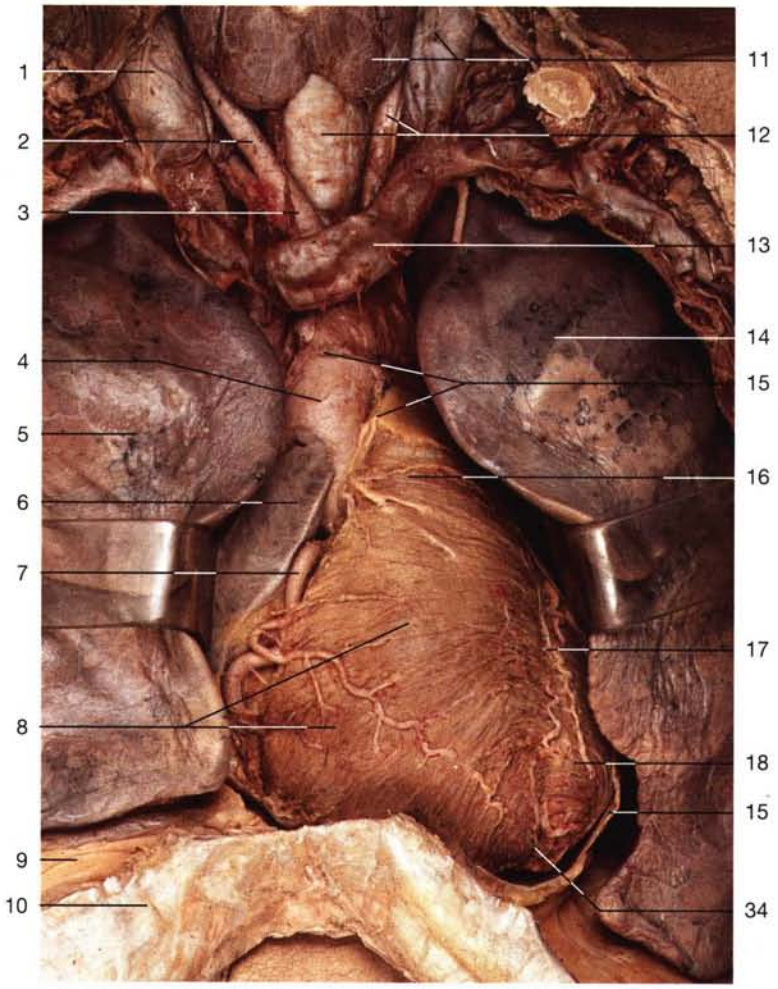

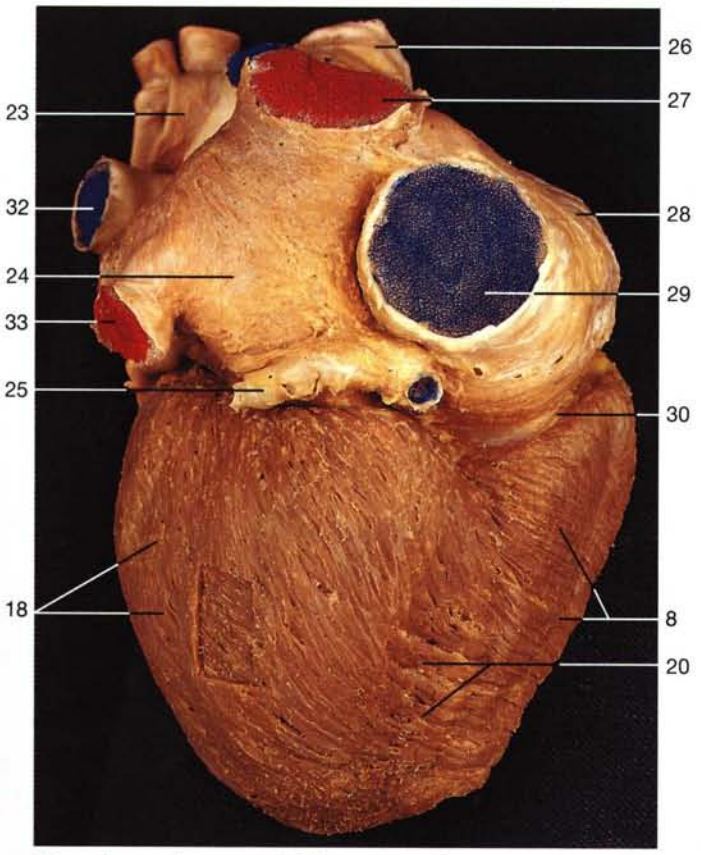

Heart (posterior aspect). The myocardium of the left ventricle has been fenestrated to show the muscle fiber bundles of the deeper layer with their more circular course.

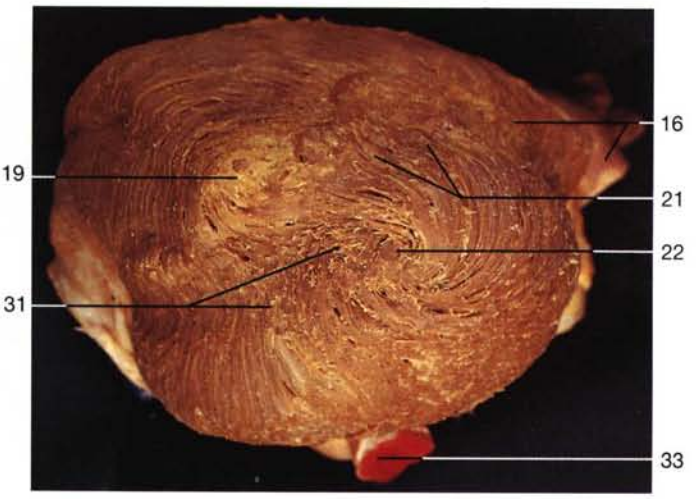

Vortex of cardiac muscle fibers (from below).

1 Superior vena cava
2 Crista terminalis
3 Fossa ovalis
4 Opening of inferior vena cava
5 Opening of coronary sinus
6 Right auricle
7 Right coronary artery and coronary sulcus
8 **Anterior cusp** of **tricuspid valve**
9 Chordae tendineae
10 **Anterior papillary muscle**
11 Myocardium
12 Pulmonary trunk
13 Ascending aorta
14 **Pulmonic valve**
15 Conus arteriosus (interventricular septum)
16 Septal papillary muscles
17 Septomarginal or moderator band
18 Apex of heart
19 Left auricle
20 **Aortic valve**
21 **Left ventricle**
22 Pulmonary veins
23 Position of fossa ovalis
24 **Left atrium**
25 **Left atrioventricular** (bicuspid or **mitral**) **valve**
26 Coronary sinus
27 Left coronary artery
28 Posterior papillary muscle
29 Left subclavian artery
30 Descending aorta
31 Left pulmonary artery

Right heart (anterior aspect). Anterior wall of right atrium and ventricle removed.

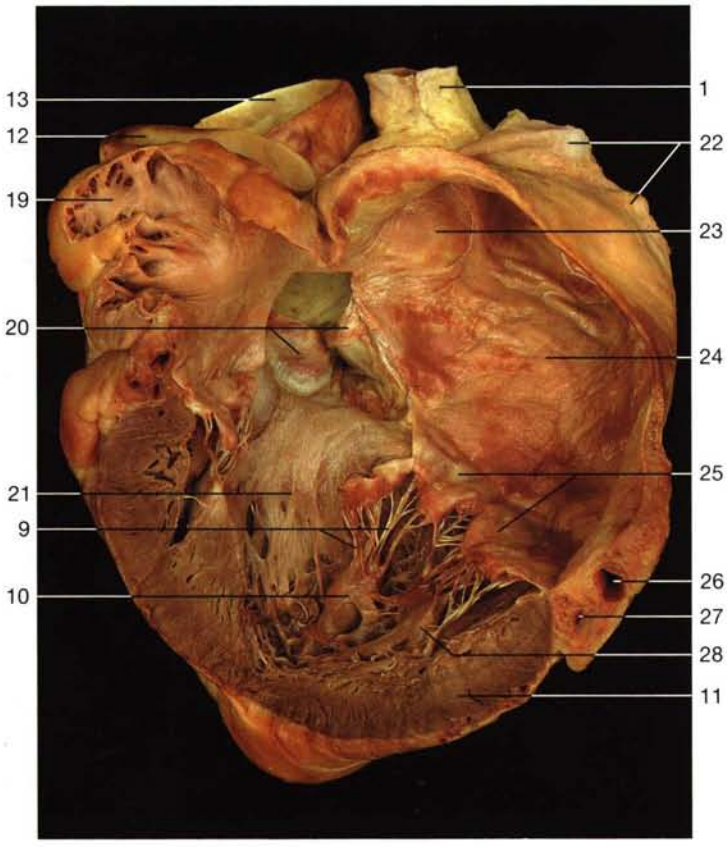

Left heart. Left atrium and ventricle opened. Aortic and mitral valve cut open.

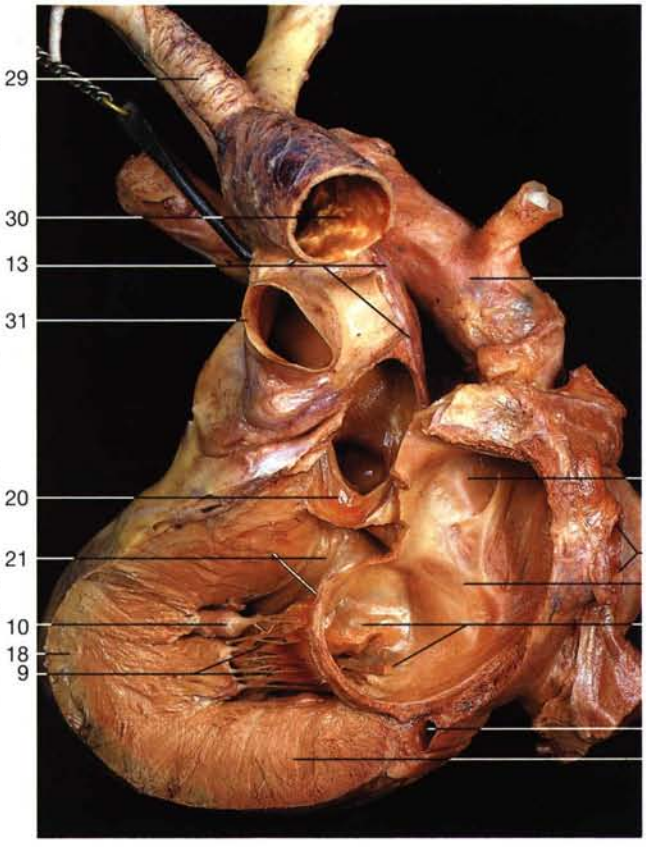

Left heart. Aortic valve cut open. Left atrium and ventricle opened. The atrial wall was cut near the anulus fibrosus of the mitral valve.

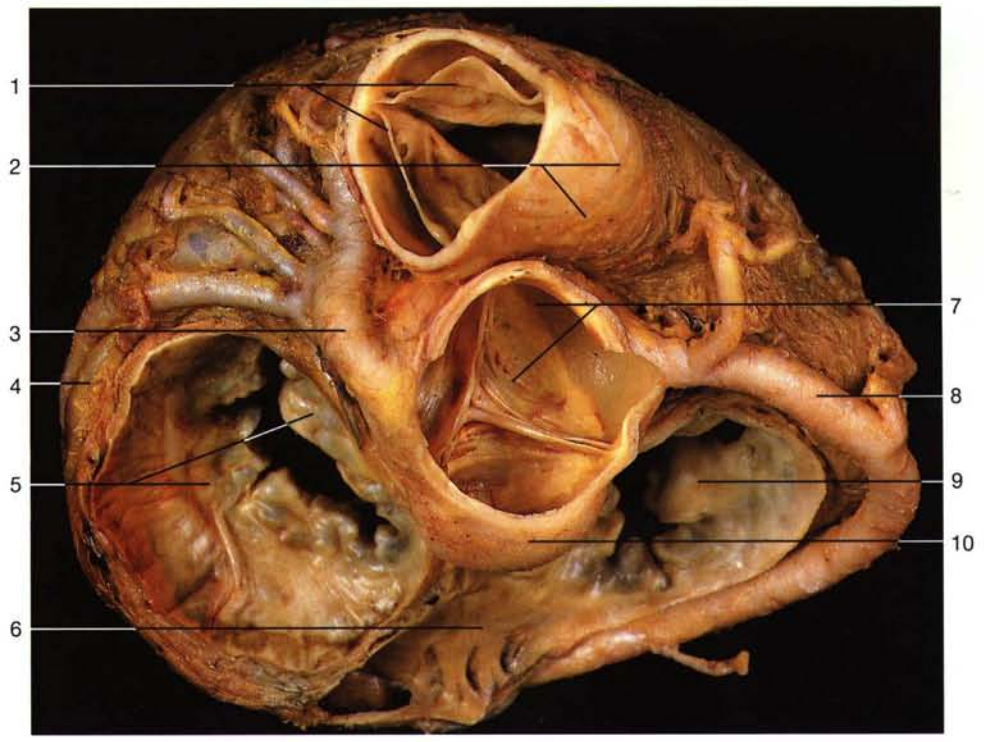

Valves of heart (superior aspect). Left and right atrium removed. Dissection of coronary arteries. Above: anterior wall of the heart.

1 **Pulmonic valve**
2 Sinus of pulmonary trunk
3 **Left coronary artery**
4 Great cardiac vein
5 **Left atrioventricular (mitral) valve**
6 **Coronary sinus**
7 **Aortic valve**
8 **Right coronary artery**
9 **Right atrioventricular (tricuspid) valve**
10 Bulb of aorta
11 Anterior semilunar cusp of pulmonic valve
12 Left semilunar cusp of pulmonic valve
13 Right semilunar cusp of pulmonic valve
14 Left semilunar cusp of aortic valve
15 Right semilunar cusp of aortic valve
16 Posterior semilunar cusp of aortic valve
17 Right atrium
18 Anterior cusp of tricuspid valve
19 **Chordae tendineae**
20 Trabeculae carneae
21 Interventricular septum
22 Septal cusp of tricuspid valve
23 **Anterior papillary muscle**
24 Myocardium of right ventricle

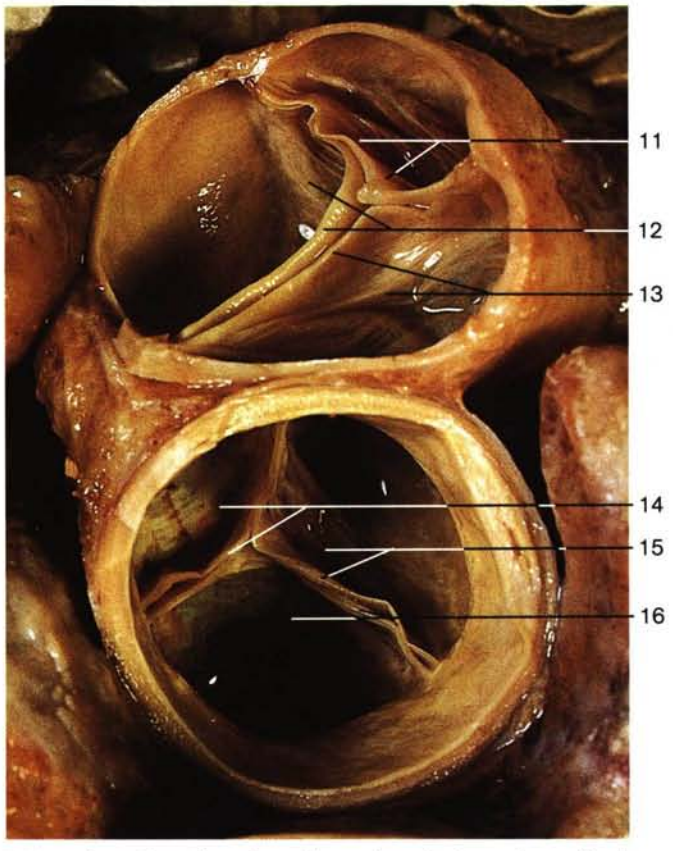

Pulmonic and aortic valves (from above). Anterior wall of the heart at the top. Both valves are closed.

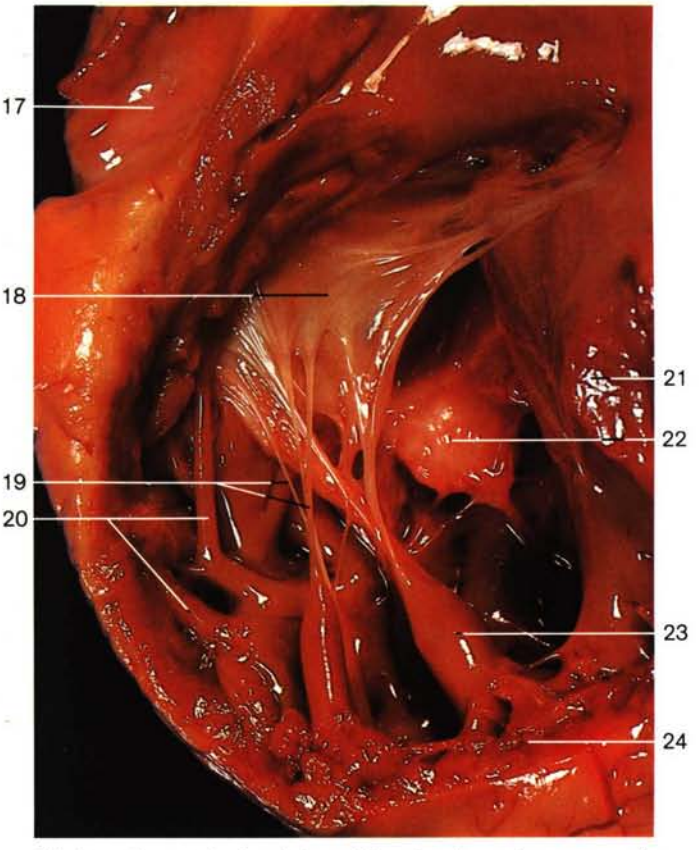

Right atrioventricular (tricuspid) valve (anterior aspect after removal of the anterior wall of the right ventricle).

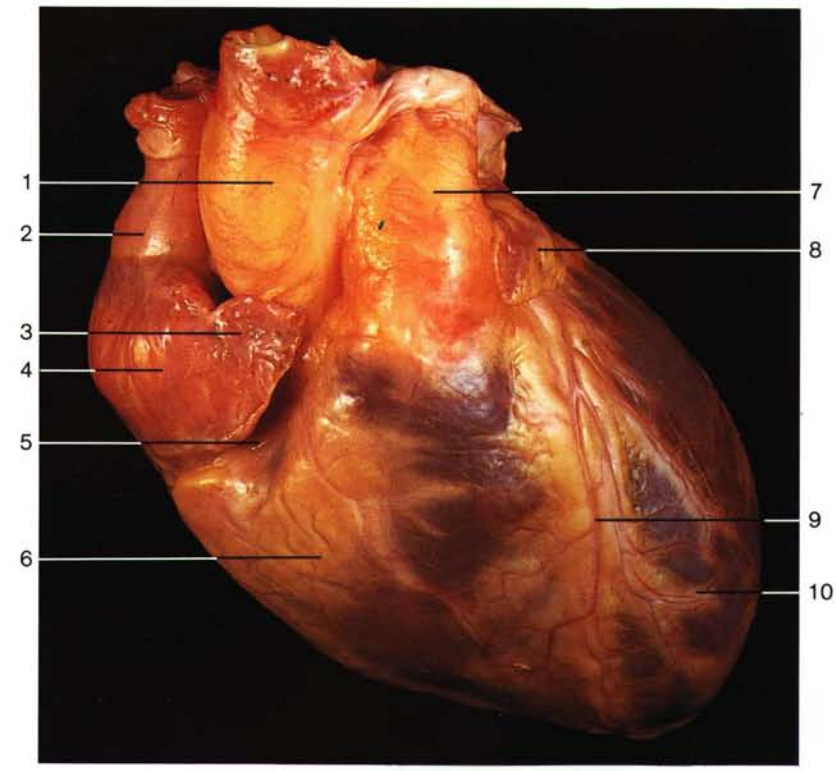

Heart, fixed in **diastole** (anterior aspect). The ventricles are relaxed, atria contracted.

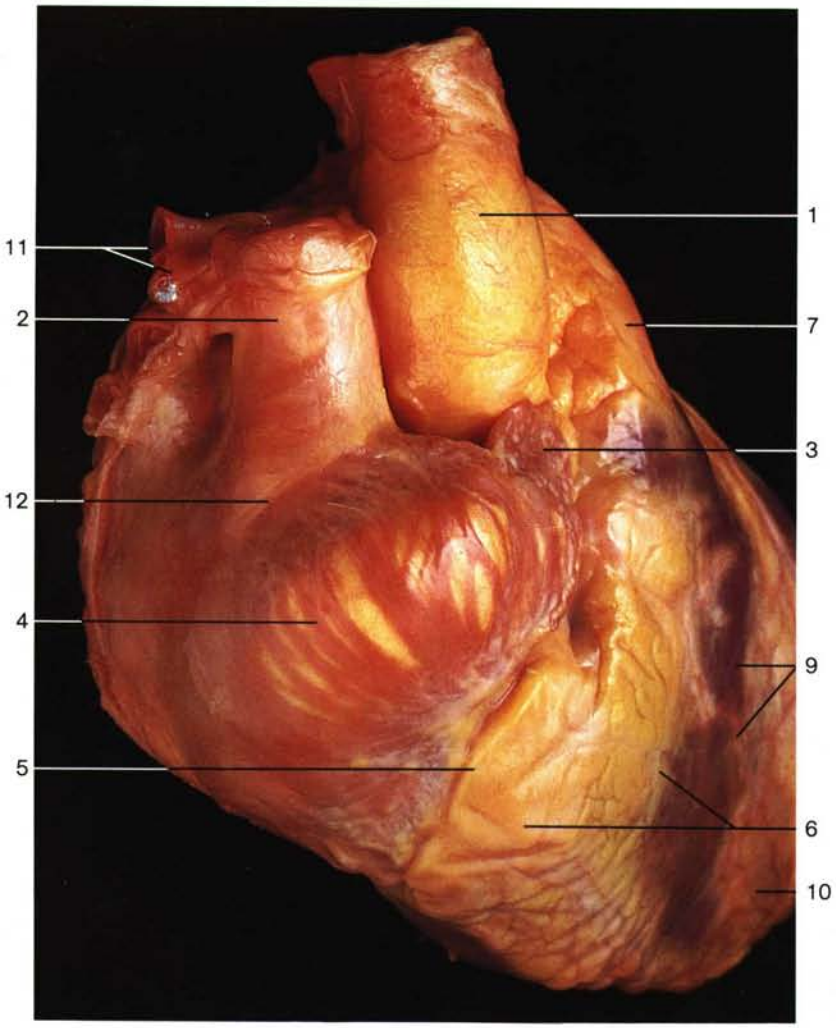

Heart, fixed in **systole** (anterolateral aspect). The ventricles are contracted, atria dilated.

1 Ascending aorta
2 Superior vena cava
3 Right auricle
4 **Right atrium**
5 Coronary sulcus
6 **Right ventricle**
7 Pulmonary trunk
8 Left auricle
9 Anterior interventricular sulcus
10 **Left ventricle**
11 Right pulmonary artery
12 Sulcus terminalis with sinuatrial node
13 Line indicating plane of position of valves
14 Myocardium of right atrium
15 Inferior vena cava
16 Valve of pulmonary trunk
17 Tricuspid valve
18 Myocardium of right ventricle

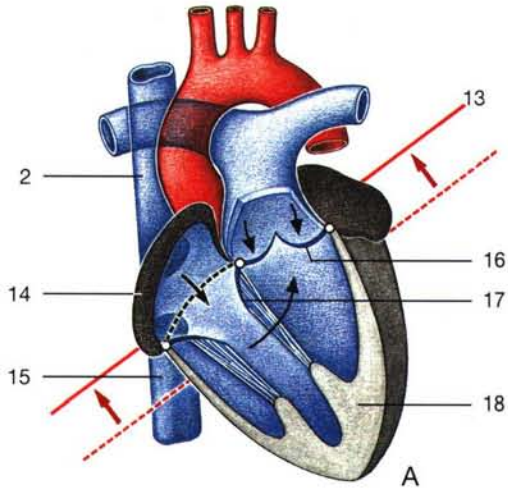

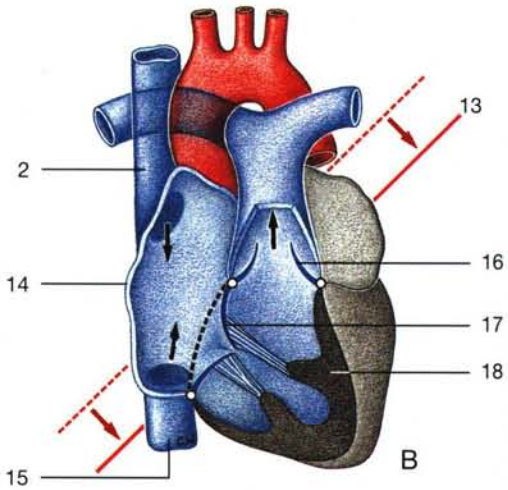

Morphological changes during heart movements. Note the changes in positon of the valves (arrows). Contracted portions of heart are indicated in red.

A. **Diastole,** muscles of the ventricles relaxed; atrioventricular valves open, semilunar valves closed.

B. **Systole,** muscles of ventricles contracted; atrioventricular valves closed, semilunar valves open.

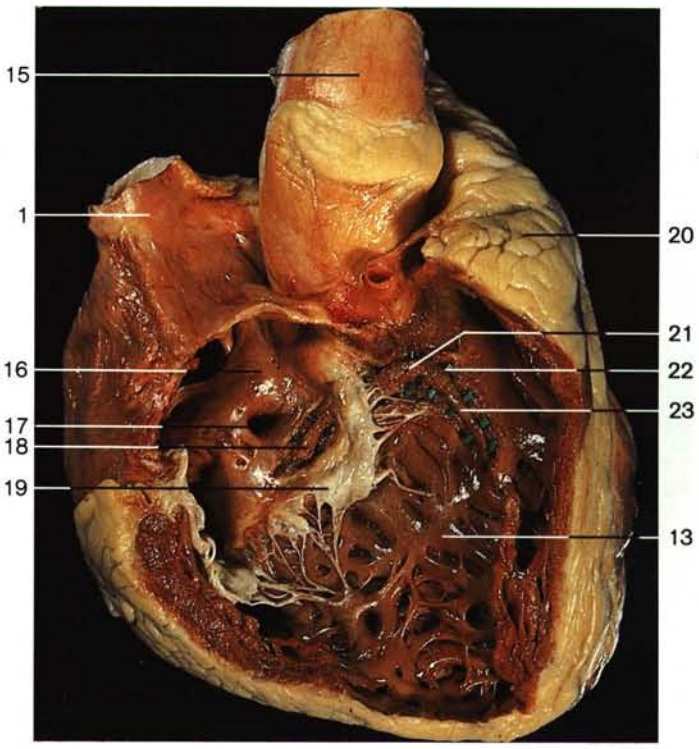

Right ventricle, dissection of **atrioventricular node, atrioventricular bundle (bundle of His)** and right limb or bundle branch (probes).

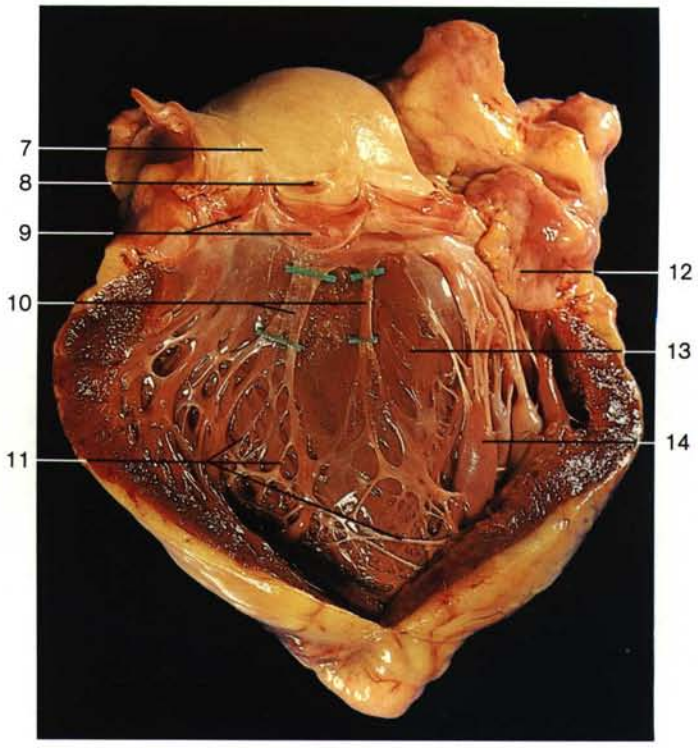

Left ventricle, dissection of the left limb or bundle branch of conducting system (probes).

1 Superior vena cava
2 Sulcus terminalis
3 Bulb of aorta
4 Sinuatrial node (arrows)
5 Muscle fiber bundles of right atrium
6 Coronary sulcus (with right coronary artery)
7 Aortic sinus
8 Entrance to left coronary artery
9 Aortic valve
10 Branches of **left bundle branch**
11 Purkinje fibers
12 Left auricle
13 Interventricular septum
14 Papillary muscles
15 Ascending aorta
16 Right atrium
17 Opening of coronary sinus
18 **Atrioventricular node**
19 Septal cusp of tricuspid valve
20 Pulmonary trunk
21 **Atrioventricular bundle** (bundle of His)
22 Bifurcation of atrioventricular bundle
23 **Right bundle branch**
24 Inferior vena cava
25 Left atrium
26 Left bundle branch
27 Papillary muscles with Purkinje fibers

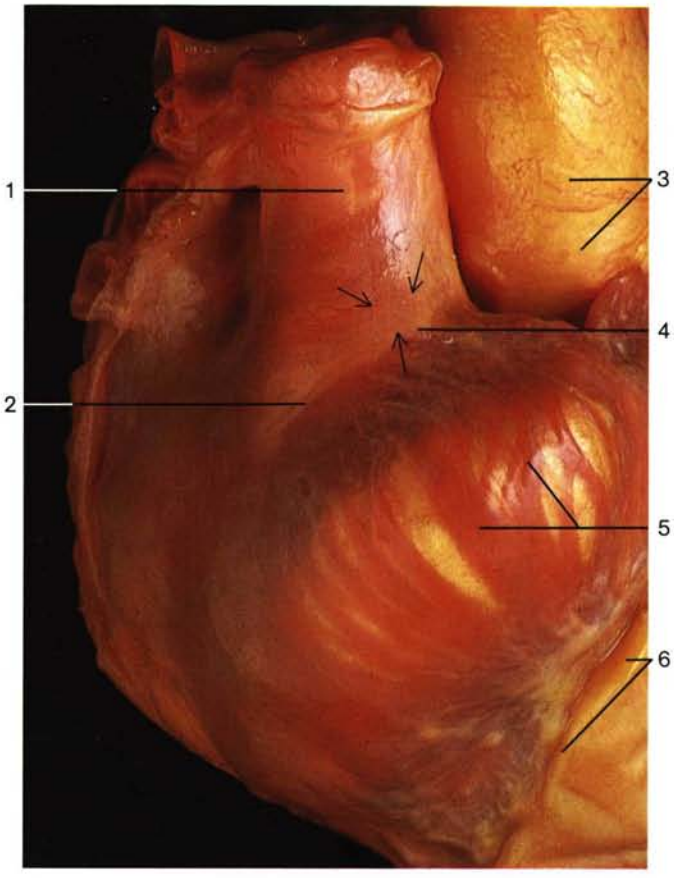

Right atrium, anterior wall, showing the location of the **sinuatrial node** (arrows).

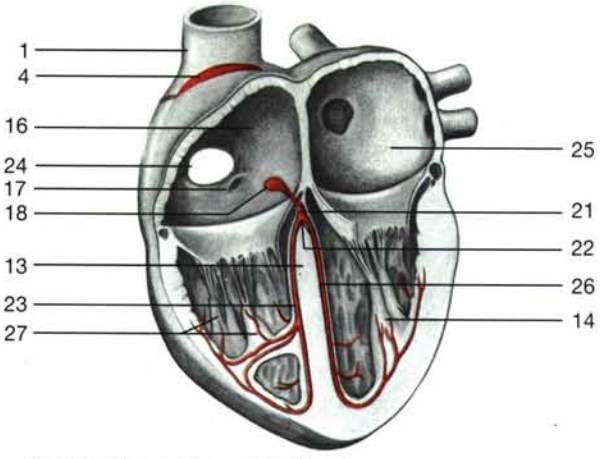

Conducting system of the heart. (Schematic drawing.)

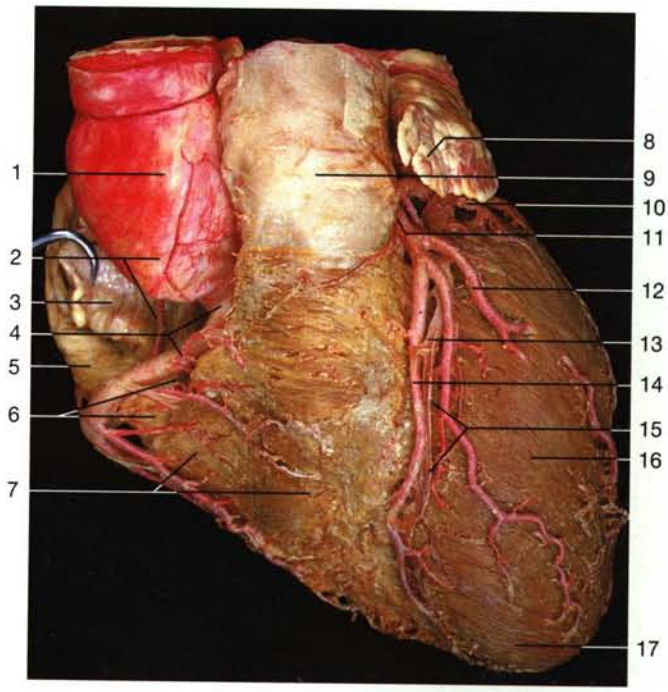

Coronary arteries (anterior aspect). The epicardium and subepicardial fatty tissue have been removed. The arteries have been injected with red resin from the aorta.

1 Ascending aorta
2 Aortic bulb and (in the above specimen) sinuatrial branch of right coronary artery
3 Right auricle
4 **Right coronary artery**
5 Right atrium
6 Coronary sulcus
7 Right ventricle
8 Left auricle
9 Pulmonary trunk

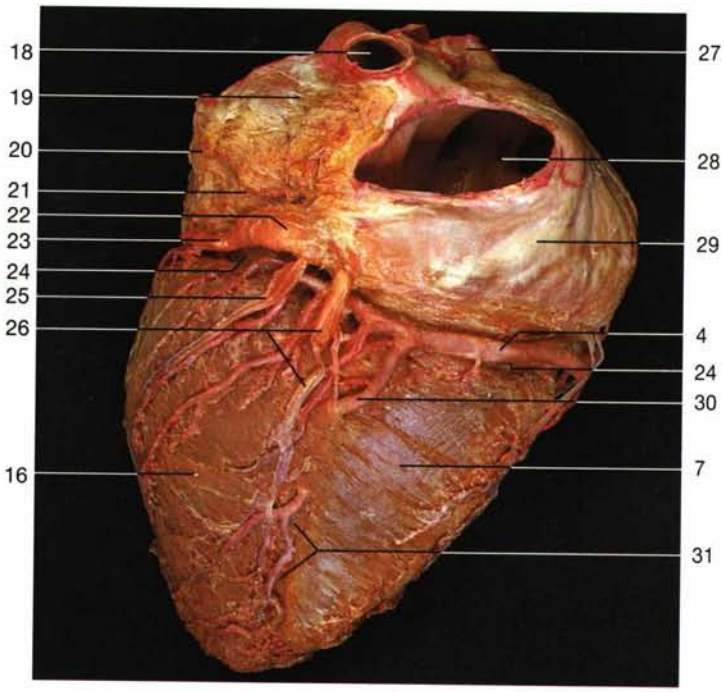

Right coronary artery and veins of the heart (dorsal aspect). The epicardium and subepicardial fatty tissue have been removed.

10 Circumflex branch of left coronary artery
11 **Left coronary artery**
12 Diagonal branch of left artery
13 Great cardiac vein
14 Anterior interventricular artery
15 Anterior interventricular sulcus
16 Left ventricle
17 Apex of heart
18 Right pulmonary vein
19 Left atrium
20 Left pulmonary veins
21 Oblique vein of left atrium (Marshall's vein)
22 **Coronary sinus**
23 **Great cardiac vein**
24 Coronary sulcus (posterior portion)
25 Posterior vein of left ventricle
26 **Middle cardiac vein**
27 Left pulmonary artery
28 Inferior vena cava
29 Right atrium
30 Posterior interventricular branch of right coronary artery
31 Posterior interventricular sulcus
32 Superior vena cava
33 Right marginal branch
34 Branch of sinuatrial node
35 Minimal cardiac veins
36 **Small cardiac vein**

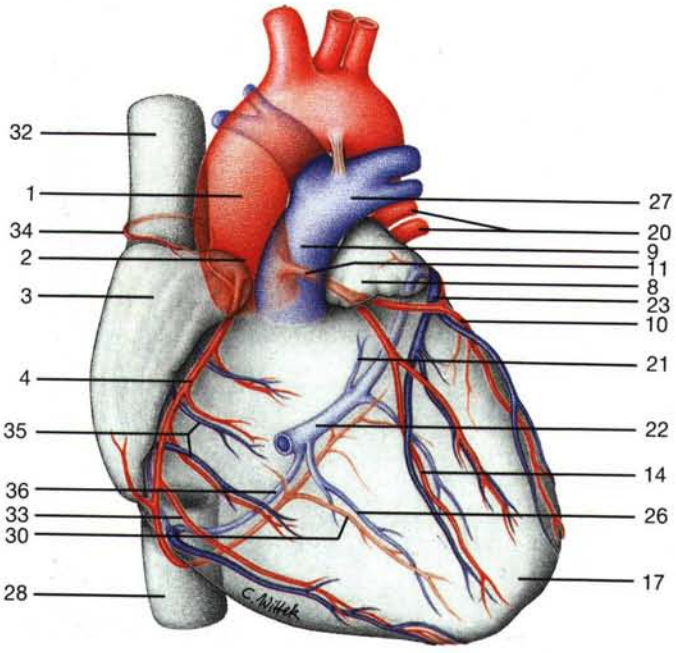

Vessels of the heart. Coronary arteries (red) and veins (blue) of the heart (anterior aspect).

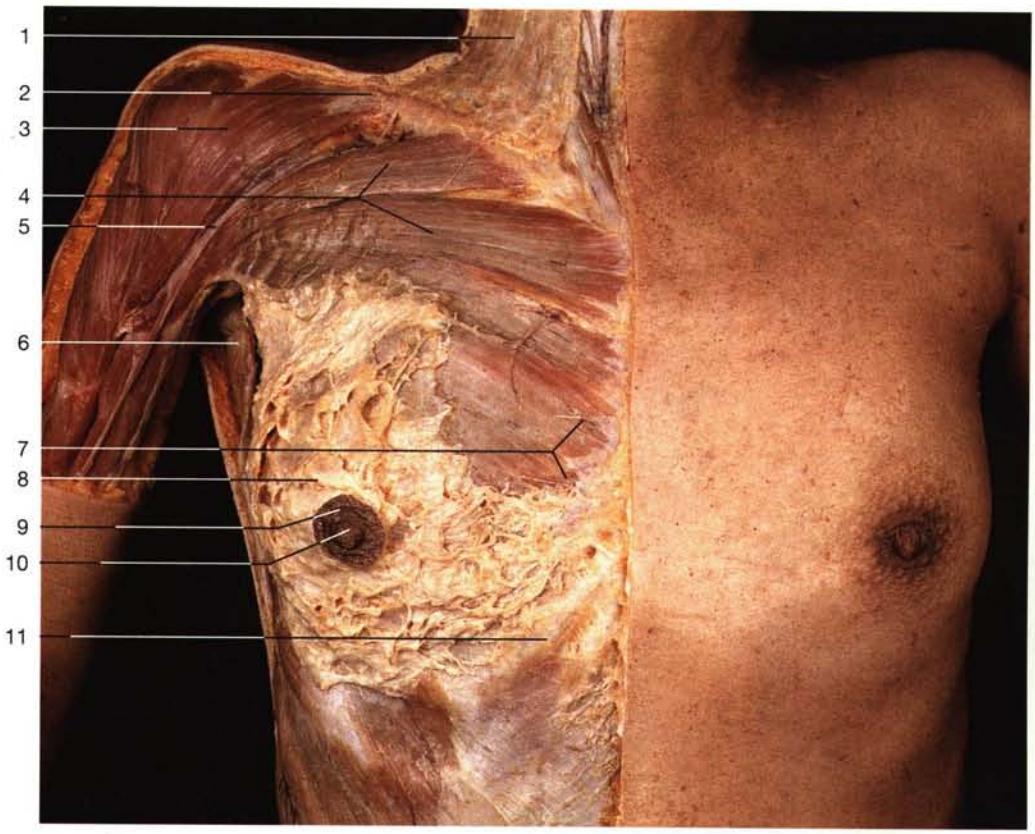

1 Platysma muscle
2 Clavicle
3 Deltoid muscle
4 Pectoralis major muscle
5 Deltopectoral groove and cephalic vein
6 Latissimus dorsi muscle
7 Medial mamarian branches of intercostal nerves
8 Breast tissue
9 Areola
10 Nipple (papilla)
11 Costal margin
12 Pectoral fascia
13 Mammary gland
14 Serratus anterior muscle
15 Lactiferous sinus
16 Lactiferous ducts

Dissection of mammary gland (anterior aspect).

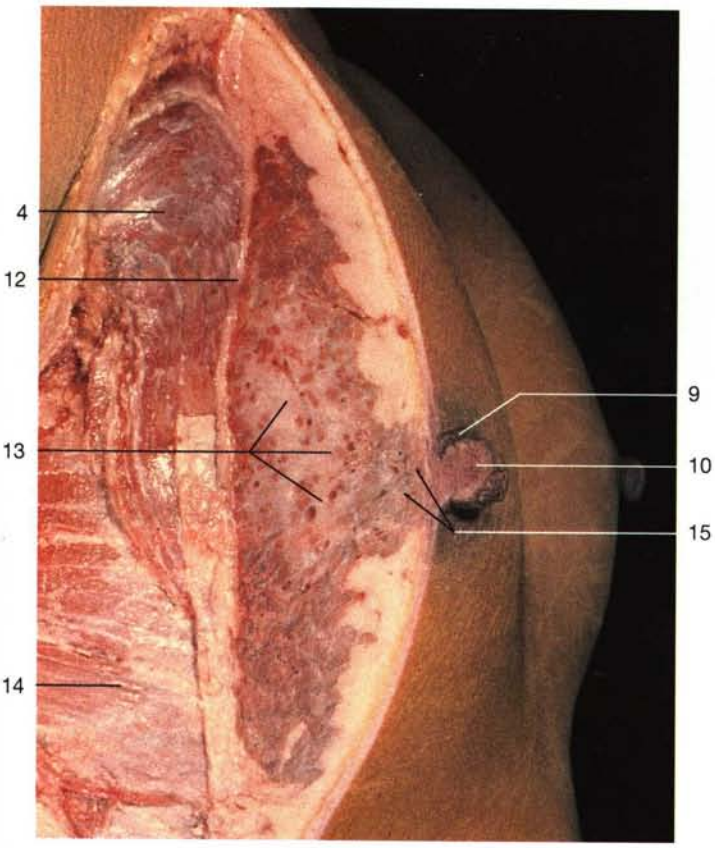

Mammary gland (sagittal section; pregnant female).

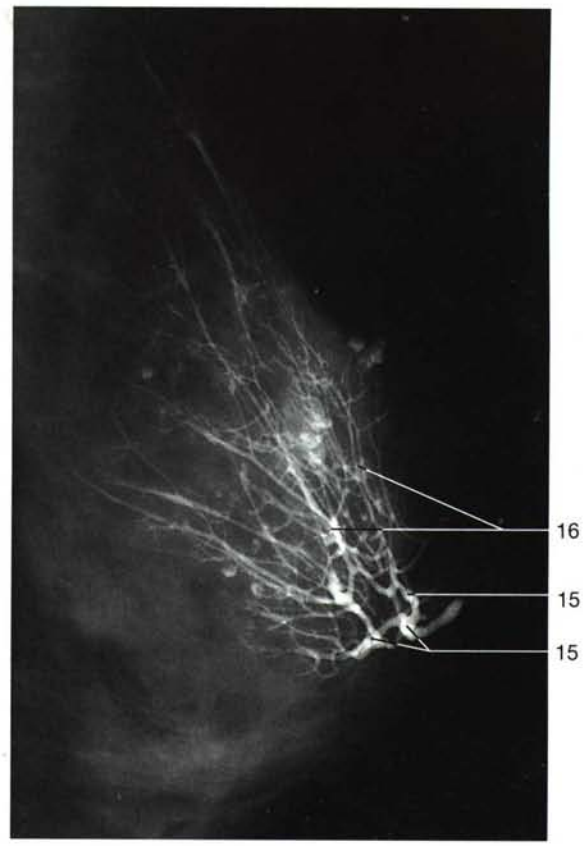

Galactogram of mammary gland (adult female). Notice multiple duct cysts.

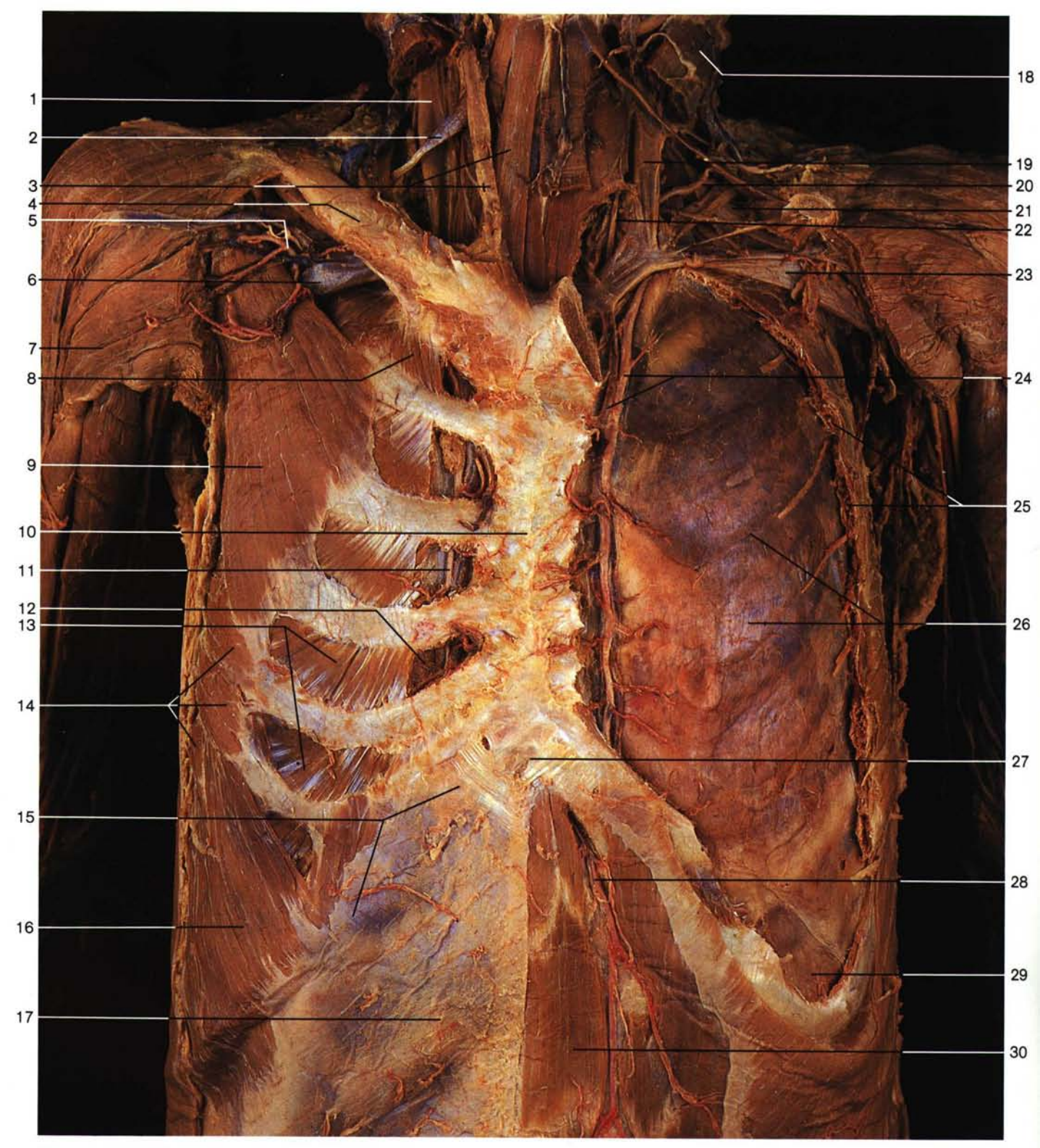

Thoracic wall and organs (ventral aspect). The left clavicle and ribs have been partially removed, and the right intercostal spaces have been opened to show the internal thoracic vein and artery.

1 Right internal jugular vein
2 Omohyoid muscle
3 Sternohyoid muscle and external jugular vein
4 Clavicle
5 Thoracoacromial artery
6 Right **subclavian vein**
7 Pectoralis major muscle
8 External intercostal muscle
9 Pectoralis minor muscle
10 Body of sternum
11 Right internal thoracic artery and vein
12 Fascicles of transversus thoracis muscle
13 **Internal intercostal muscles**
14 Serratus anterior muscle
15 Costal margin
16 External abdominal oblique muscle
17 Anterior sheath of rectus abdominis muscle
18 Sternocleidomastoid muscle
19 Left internal jugular vein
20 Transverse cervical artery
21 Brachial plexus
22 Vagus nerve
23 Left axillary vein
24 Left **internal thoracic artery and vein**
25 Ribs and thoracic wall (cut)
26 Costal pleura
27 Xiphoid process
28 **Superior epigastric artery**
29 Diaphragm
30 Rectus abdominis muscle

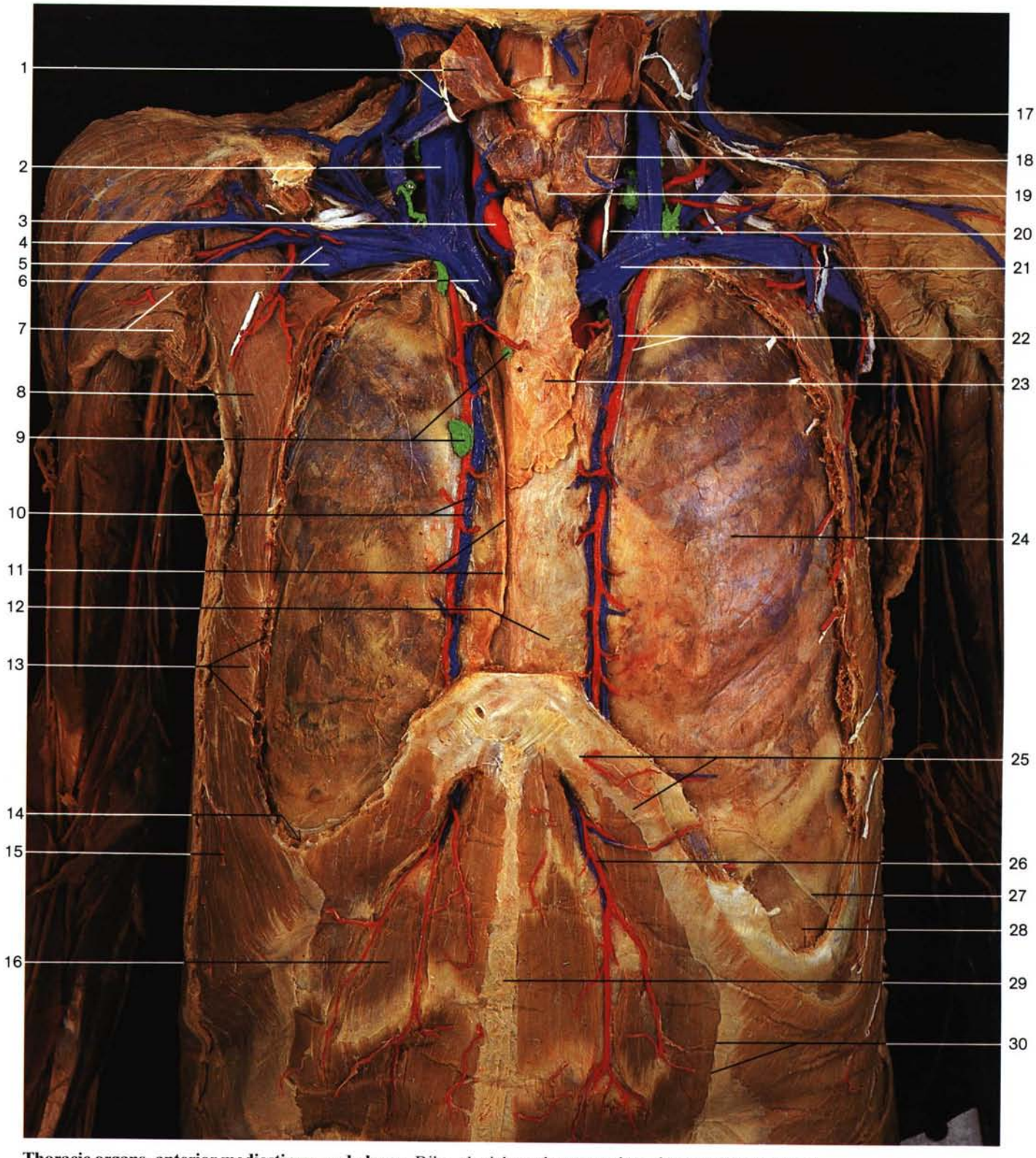

Thoracic organs, anterior mediastinum and pleura. Ribs, clavicle and sternum have been partly removed. Red = arteries; blue = veins; green = lymph vessels and nodes.

1 Sternothyroid muscle and its nerve (a branch of the ansa cervicalis)
2 Right **internal jugular vein**
3 Right common carotid artery
4 Cephalic vein
5 Right **subclavian vein**
6 **Right brachiocephalic vein**
7 Pectoralis major muscle (divided)
8 Pectoralis minor muscle (divided)
9 Parasternal lymph nodes
10 Internal thoracic artery and vein
11 Anterior margin of costal pleura
12 **Pericardium**
13 5th and 6th ribs (divided) and serratus anterior muscle
14 **Costodiaphragmatic recess**
15 External abdominal oblique muscle
16 Rectus abdominis muscle
17 Larynx (thyroid cartilage)
18 Thyroid gland
19 Trachea
20 Left vagus nerve
21 Left brachiocephalic vein
22 Left **internal thoracic artery and vein**
23 Thymus
24 **Costal pleura**
25 Costal margin
26 **Superior epigastric artery**
27 Margin of costal pleura
28 **Diaphragm**
29 **Linea alba**
30 Cut edge of anterior sheath of rectus abdominis muscle

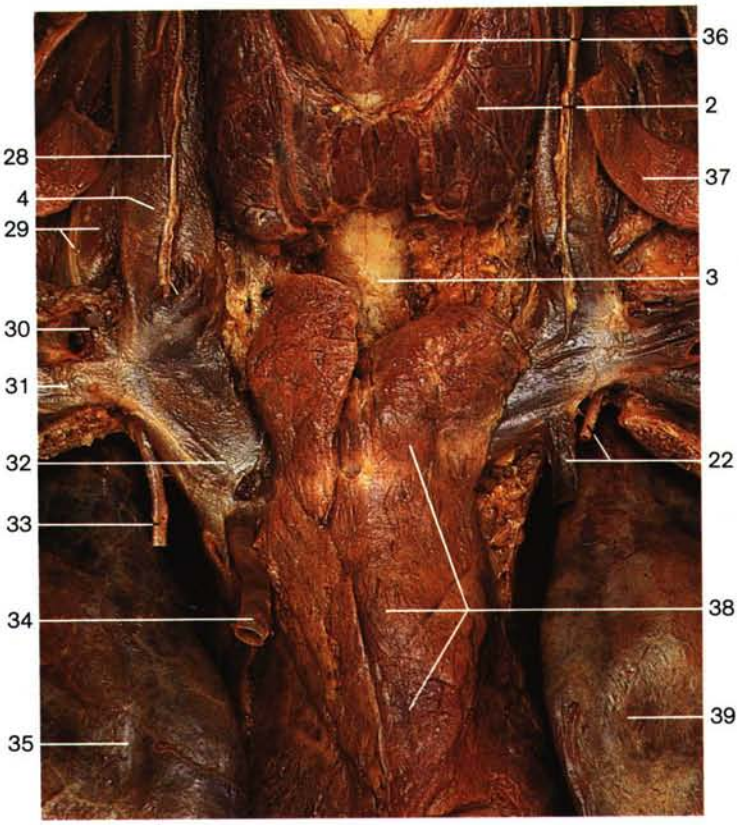

The **thymus** above the heart, showing its position and size.

1 Larynx (thyroid cartilage)
2 Thyroid gland
3 **Trachea**
4 Internal jugular vein
5 Brachial plexus
6 Right brachiocephalic vein and common carotid artery
7 Right phrenic nerve
8 **Ascending aorta**
9 Pectoralis minor muscle (divided)
10 Pulmonary trunk (covered by pericardium)
11 **Costal pleura**
12 Pericardium and **heart**
13 Serratus anterior muscle
14 Xiphoid process
15 Costal margin
16 External abdominal oblique muslce
17 Sternothyroid muscle (divided and reflected)
18 Vagus nerve
19 Left common carotid artery
20 Left sympathetic trunk
21 Left recurrent laryngeal nerve
22 Left internal thoracic artery and vein (divided)
23 Margin of costal pleura
24 Intercostal nerves and vessels
25 Superior epigastric artery
26 Rectus abdominis muscle
27 **Diaphragm**
28 Ansa cervicalis
29 Phrenic nerve and scalenus anterior muscle
30 External jugular vein (divided)
31 Right subclavian vein
32 Right brachiocephalic vein
33 Internal thoracic artery (divided)
34 Internal thoracic vein (divided)
35 Right lung
36 Cricothyroid muscle
37 Omohyoid muscle
38 **Thymus**
39 Left lung

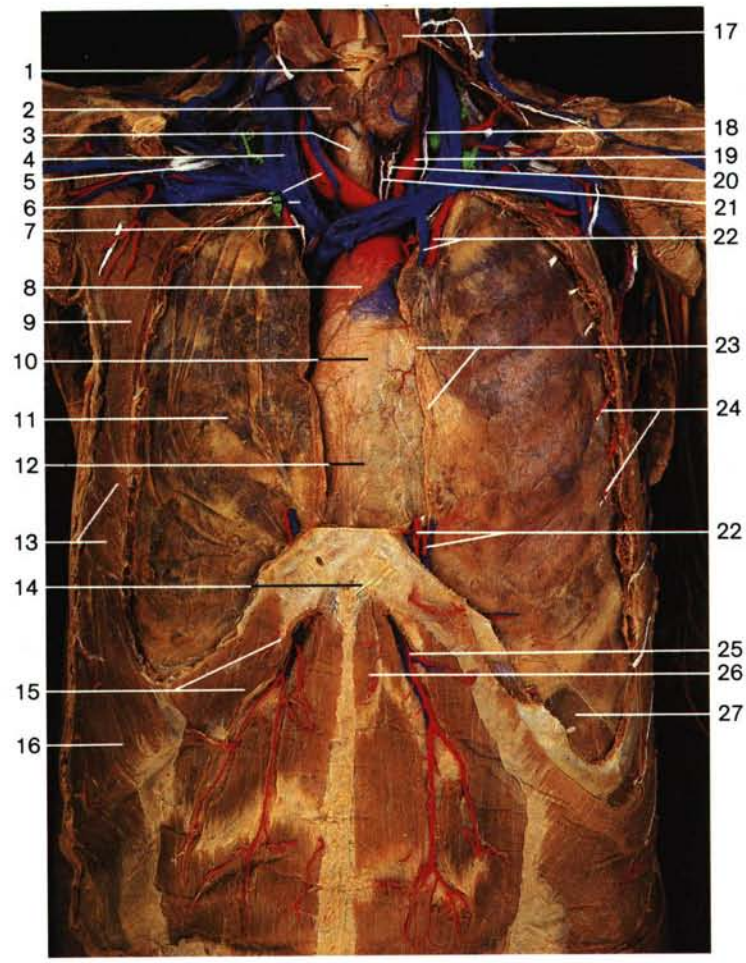

Thoracic organs (ventral aspect). The internal thoracic vessels have been removed, and the anterior margins of the pleura and lungs have been slightly reflected to display the **anterior** and **middle mediastinum,** including the **heart** and **great vessels.**

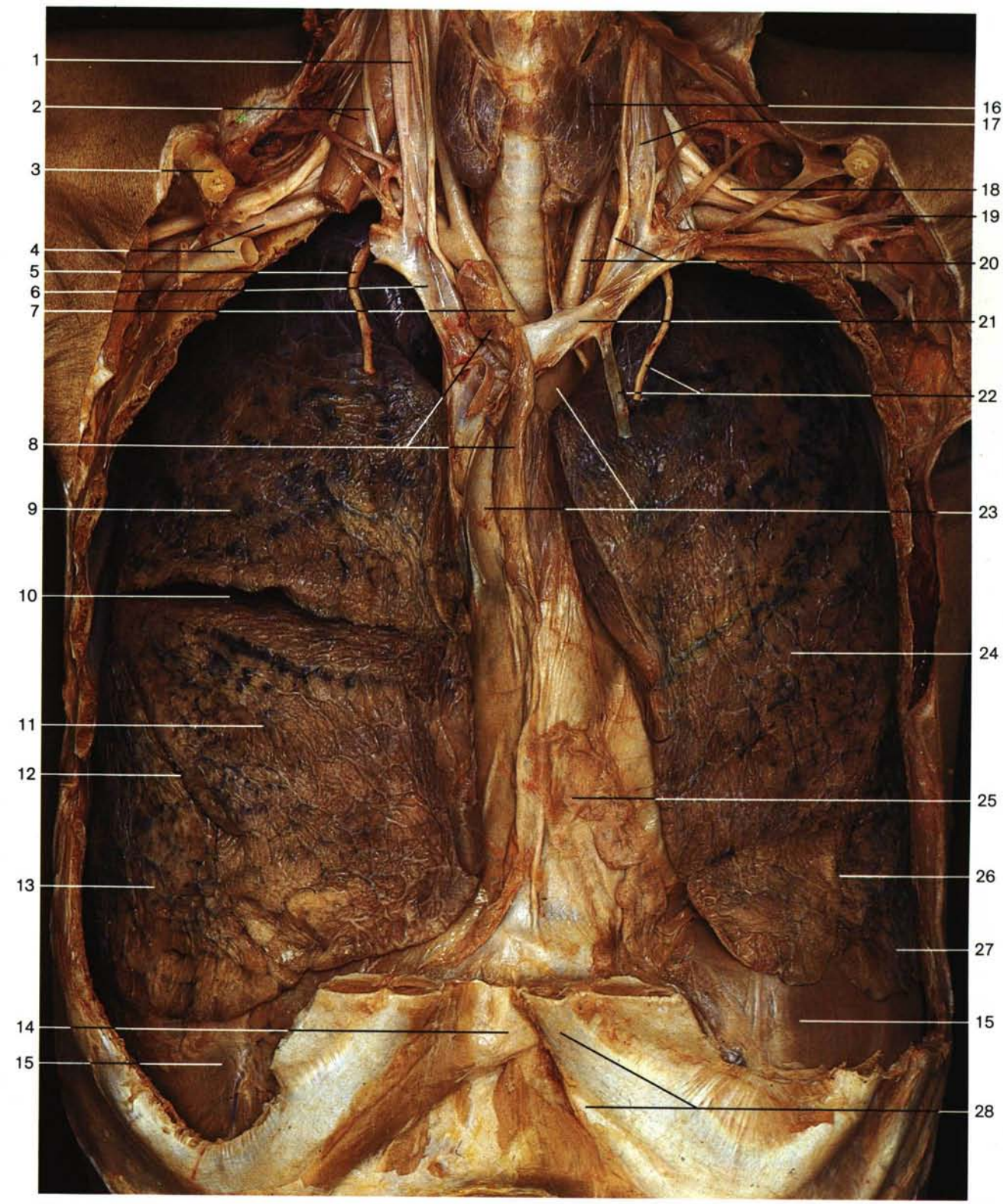

Thoracic organs (ventral aspect). The pleura has been opened and the lungs exposed. Remnants of the thymus and pericardium are seen.

1 Right internal jugular vein
2 Phrenic nerve and scalenus anterior muscle
3 Clavicle (divided)
4 Right subclavian artery and vein
5 Internal thoracic artery
6 **Right brachiocephalic vein**
7 Brachiocephalic trunk
8 Thymus (atrophic)
9 Upper lobe of right **lung**
10 Horizontal fissure of right lung (incomplete)
11 Middle lobe of right lung
12 Oblique fissure of right **lung**
13 Lower lobe of right lung
14 Xiphoid process
15 **Diaphragm**
16 **Thyroid gland**
17 Left internal jugular vein
18 Brachial plexus
19 Left cephalic vein
20 Left common carotid artery and vagus nerve
21 **Left brachiocephalic vein**
22 Internal thoracic artery and vein (divided)
23 **Ascending aorta** and aortic arch
24 Upper lobe of left **lung**
25 **Pericardium**
26 Oblique fissure of left lung
27 Lower lobe of left lung
28 Costal margin

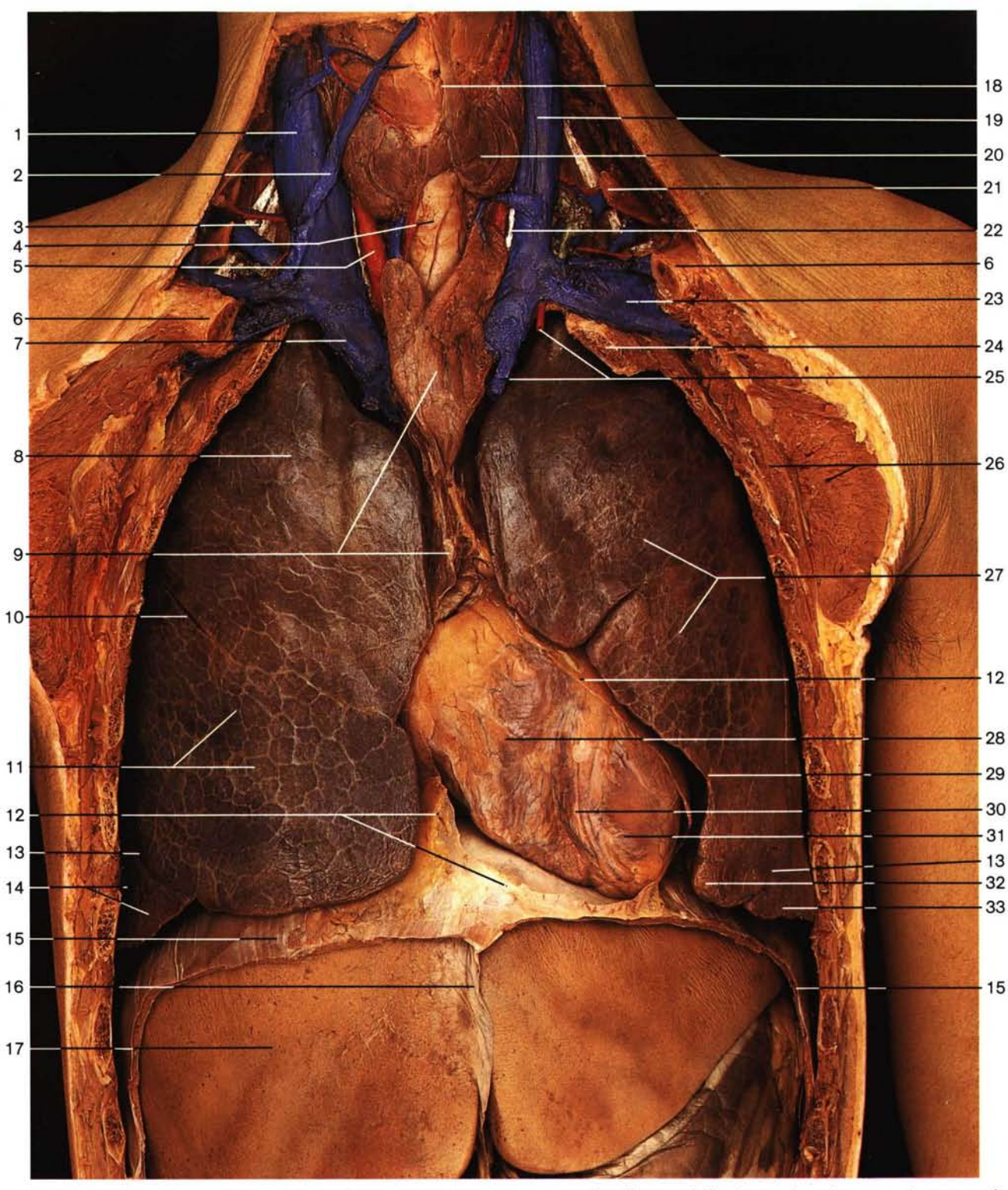

Thoracic organs (ventral aspect). The thoracic wall, costal pleura, pericardium and diaphragm have been partly removed.

1 Internal jugular vein
2 External jugular vein (displaced medially)
3 Brachial plexus
4 Trachea
5 Right common carotid artery
6 Clavicle (divided)
7 Right brachiocephalic vein
8 Upper lobe of right lung
9 **Thymus** (atrophic)
10 Horizontal fissure of right lung
11 Middle lobe of right lung
12 Pericardium (cut edges)
13 Oblique fissure of lung
14 Lower lobe of right lung
15 **Diaphragm**
16 Falciform ligament
17 **Liver**
18 Location of larynx
19 Left internal jugular vein
20 Thyroid gland
21 Omohyoid muscle (divided)
22 Vagus nerve
23 Left subclavian vein
24 1st rib (divided)
25 Internal thoracic artery and vein
26 Pectoralis major and pectoralis minor muscles (cut edges)
27 **Upper lobe of left lung**
28 **Right ventricle**
29 Cardiac notch of left lung
30 Interventricular sulcus of heart
31 **Left ventricle**
32 Lingula
33 Lower lobe of left lung

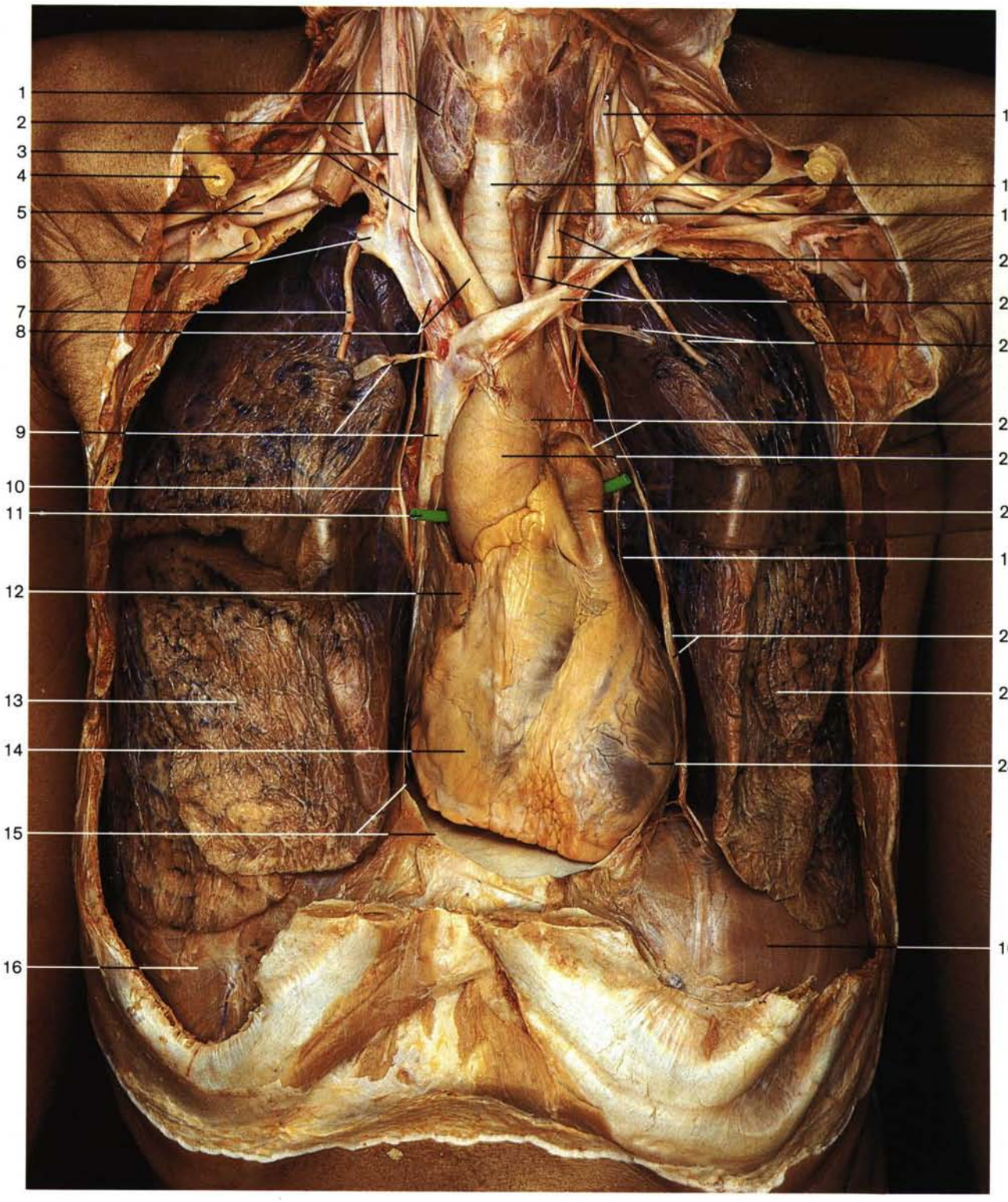

Thoracic organs, position of the heart, middle mediastinum (ventral aspect). The anterior wall of the thorax, the costal pleura and the pericardium have been removed and the lungs slightly reflected.

1 Thyroid gland
2 Phrenic nerve and scalenus anterior muscle
3 Vagus nerve and internal jugular vein
4 Clavicle (divided)
5 Brachial plexus and subclavian artery
6 Subclavian vein
7 Internal thoracic artery
8 Brachiocephalic trunk and right brachiocephalic vein
9 **Superior vena cava** and thymic vein
10 **Right phrenic nerve**
11 Transverse pericardial sinus (probe)
12 Right auricle
13 Middle lobe of right lung
14 **Right ventricle**
15 Cut edge of pericardium
16 Diaphragm
17 Internal jugular vein
18 Trachea
19 Left recurrent laryngeal nerve
20 Left common carotid artery and vagus nerve
21 Left brachiocephalic vein and inferior thyroid vein
22 Left internal thoracic artery and vein (divided)
23 Upper margin of pericardial sac
24 **Ascending aorta**
25 Pulmonary trunk
26 **Left phrenic nerve** and left pericardiacophrenic artery and vein
27 Upper lobe of left lung
28 Left ventricle

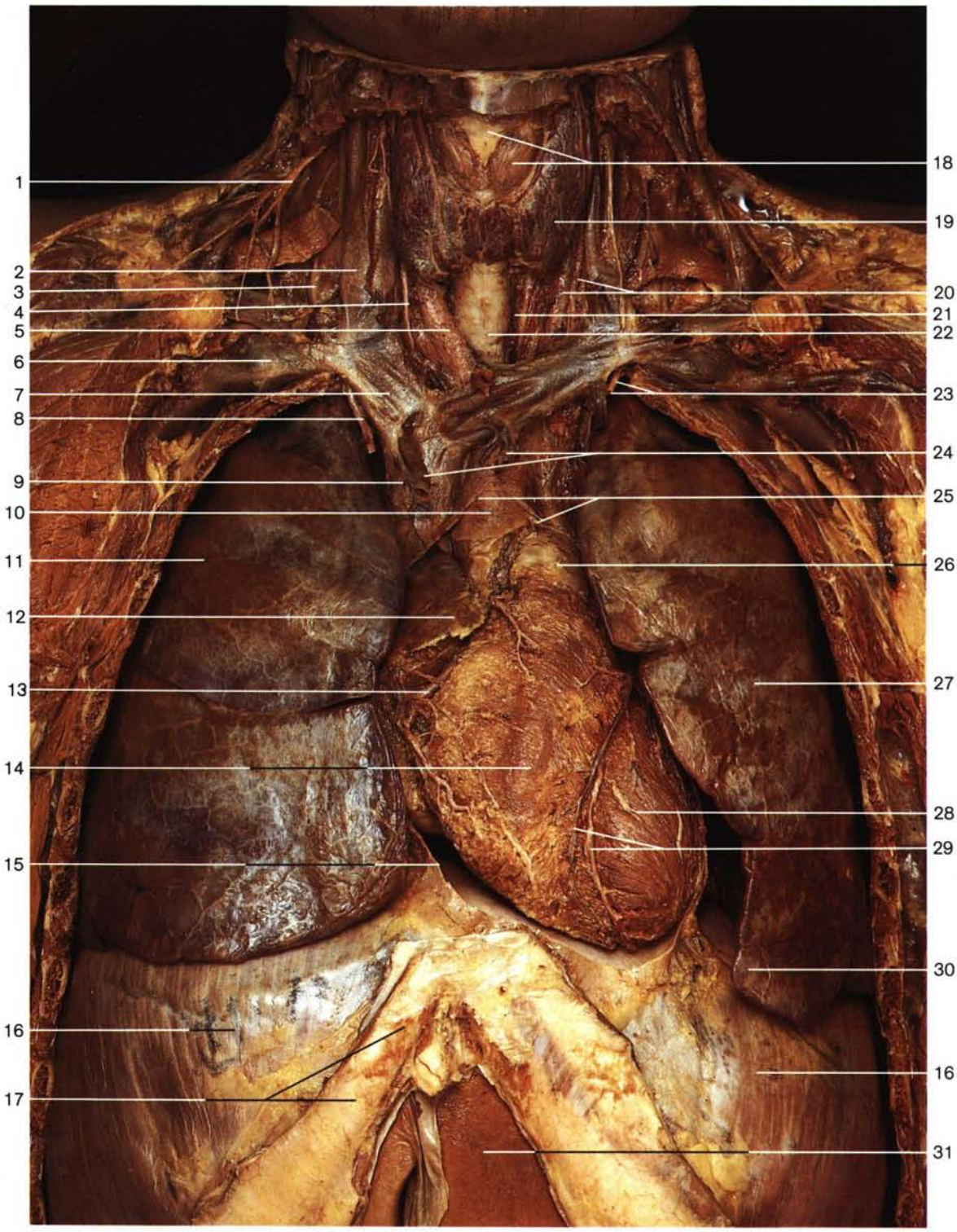

Thoracic organs, position of heart, dissection of coronary vessels in situ (ventral aspect). The anterior wall of thorax, costal pleura and pericardium have been removed.

1 Intermediate supraclavicular nerve
2 Internal jugular vein
3 Right phrenic nerve
4 Right vagus nerve
5 Right common carotid artery
6 Right subclavian vein
7 Right brachiocephalic vein
8 Right internal thoracic artery
9 **Superior vena cava**
10 Ascending aorta
11 Right lung
12 Right atrium
13 **Right coronary artery** and small cardiac vein
14 **Right ventricle**
15 Cut edge of pericardium
17 Costal margin
18 Larynx (cricothyroid muscle and thyroid cartilage)
19 Thyroid gland
20 Left common carotid artery and left vagus nerve
21 Left recurrent laryngeal nerve
22 Trachea
23 Left internal thoracic artery and vein (divided)
24 Thymic veins
25 Margin of pericardial sac
26 Pulmonary trunk
27 Left lung
28 Left ventricle
29 Anterior interventricular artery and vein
30 Lingula
31 Liver

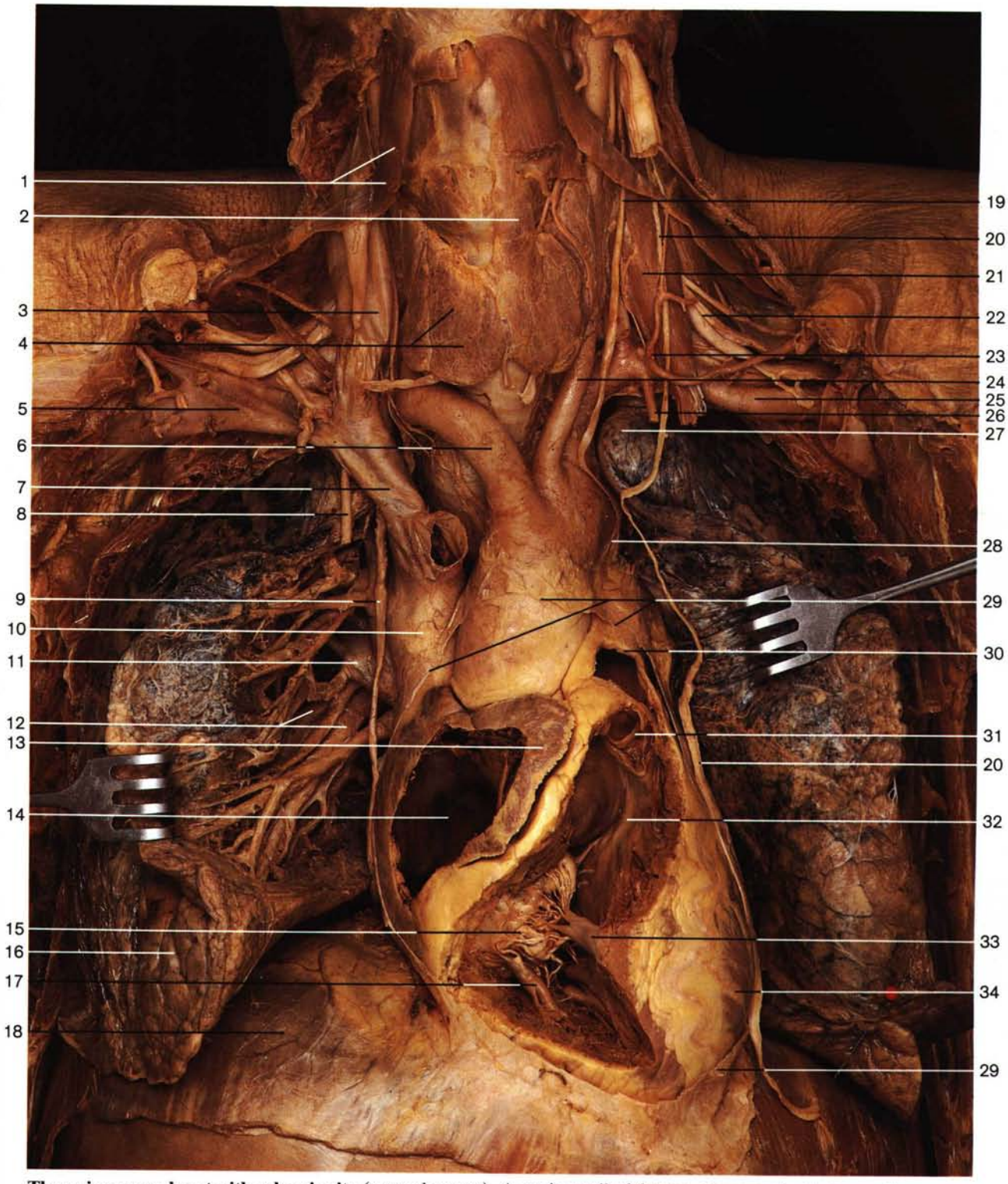

Thoracic organs, heart with valves in situ (ventral aspect). Anterior wall of thorax, pleura and anterior portion of pericardium have been removed. The right atrium and ventricle have been opened to show the right atrioventricular and pulmonary valves.

1 Omohyoid muscle
2 Pyramidal lobe of thyroid gland
3 Internal jugular vein
4 Thyroid gland
5 Right subclavian vein
6 Brachiocephalic trunk
7 **Right brachiocephalic vein**
8 Right internal thoracic artery
9 Right phrenic nerve
10 **Superior vena cava**
11 Pulmonary vein
12 Branches of pulmonary artery
13 Right auricle
14 **Right atrium**
15 **Right atrioventricular (tricuspid) valve**
16 Right lung
17 Posterior papillary muscle
18 Diaphragm
19 Left **vagus nerve**
20 Left **phrenic nerve**
21 Scalenus anterior muscle
22 Brachial plexus
23 Thyrocervical trunk
24 Left common carotid artery
25 Left subclavian artery
26 Left internal thoracic artery
27 Apex of left lung
28 Left recurrent laryngeal nerve
29 Cut edge of **pericardium**
30 Pulmonary trunk (fenestrated)
31 Pulmonic valve
32 Supraventricular crest
33 **Anterior papillary muscle**
34 **Left ventricle**

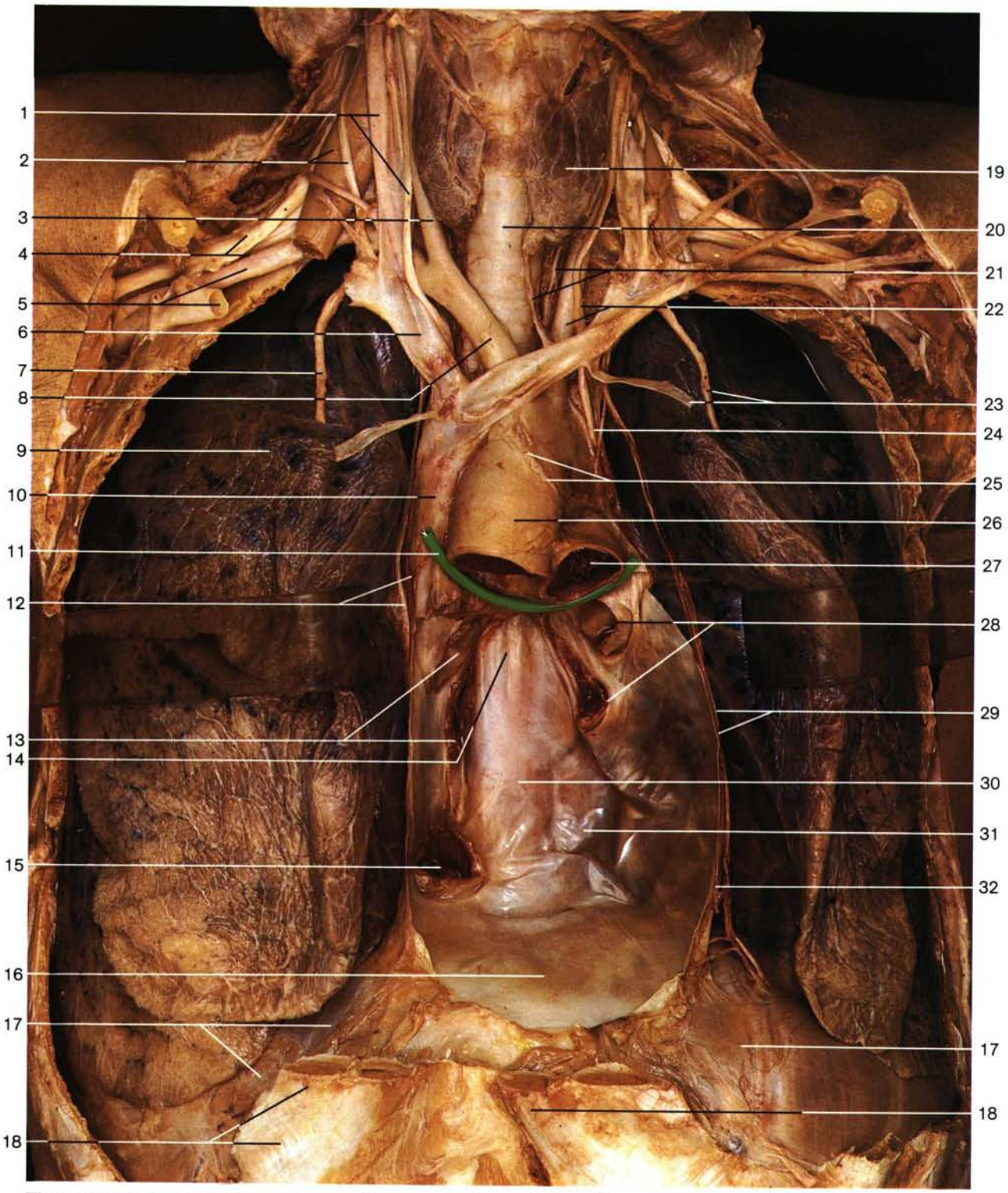

Thoracic organs, pericardium and **mediastinum** (ventral aspect). Anterior wall of thorax and heart have been removed and the lungs slightly reflected. Note probe within transverse pericardial sinus.

1 Right internal jugular vein and right vagus nerve
2 Right phrenic nerve and scalenus anterior muscle
3 Right common carotid artery
4 Brachial plexus
5 Right subclavian artery and vein
6 Right brachiocephalic vein
7 Right internal thoracic artery (divided)
8 **Brachiocephalic trunk**
9 Upper lobe of right lung
10 **Superior vena cava**
11 Transverse pericardial sinus (probe)
12 Right **phrenic nerve** and right pericardiacophrenic artery and vein
13 Right pulmonary veins
14 Oblique sinus of pericardium
15 **Inferior vena cava**
16 Diaphragmatic part of pericardium
17 Diaphragm
18 Costal margin
19 Thyroid gland
20 **Trachea**
21 Left recurrent laryngeal nerve and inferior thyroid vein
22 Left common carotid artery and left vagus nerve
23 Left internal thoracic artery and vein (divided)
24 **Vagus nerve** at aortic arch
25 Cut edge of pericardium
26 **Ascending aorta**
27 Pulmonary trunk (divided)
28 Left pulmonary veins
29 Left phrenic nerve and left pericardiacophrenic artery and vein
30 Contour of esophagus beneath pericardium
31 Contour of aorta beneath pericardium
32 **Pericardium** (cut edge)

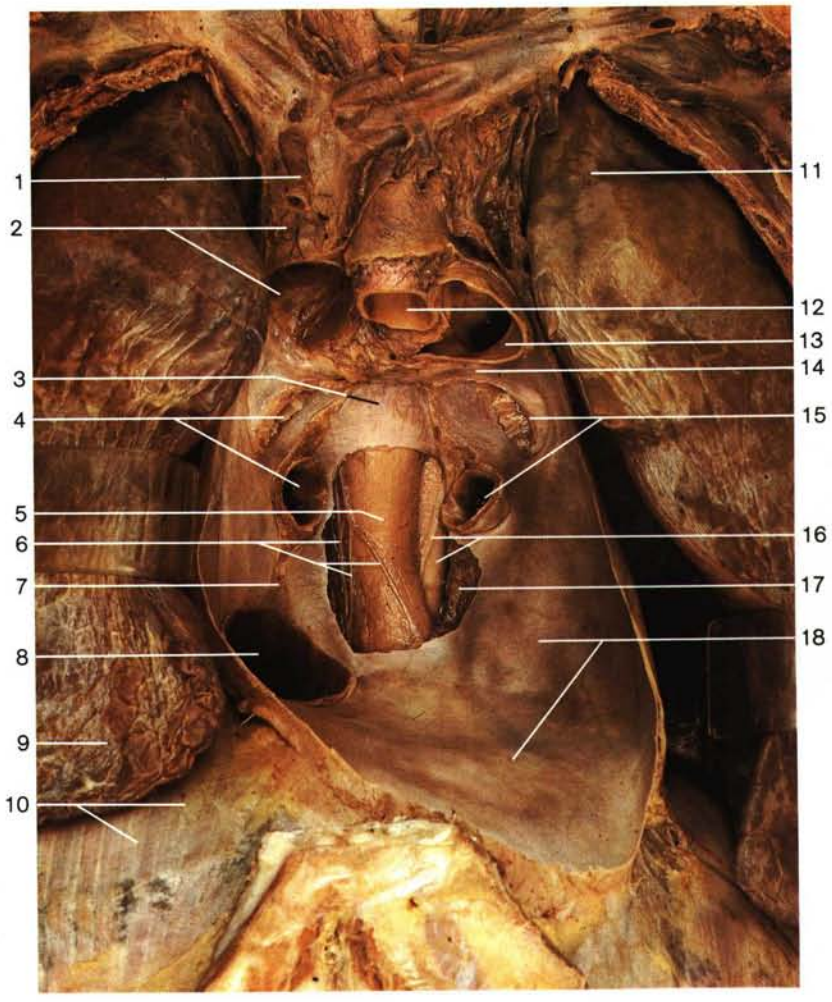

1 Internal thoracic vein
2 Superior vena cava
3 Oblique sinus of pericardium
4 Right pulmonary veins
5 Esophagus
6 Branches of right vagus nerve
7 Mesocardium
8 Inferior vena cava
9 Middle lobe of right lung
10 Diaphragm
11 Upper lobe of left lung
12 Ascending aorta
13 **Pulmonary trunk**
14 Transverse pericardial sinus
15 Left pulmonary veins
16 Descending aorta and left vagus nerve
17 Left lung (adjacent to pericardium)
18 **Pericardium**
19 Left subclavian artery
20 Vagus nerve
21 Left recurrent laryngeal nerve
22 Descending aorta
23 Pulmonary artery
24 **Left atrium**
25 **Left ventricle**
26 Coronary sinus
27 Left common carotid artery
28 Brachiocephalic trunk
29 Azygos arch
30 **Right atrium**
31 **Right ventricle**
32 Aortic arch

Pericardial sac (ventral aspect). The heart has been removed, and the posterior wall of the pericardium has been opened to show the adjacent esophagus and aorta.

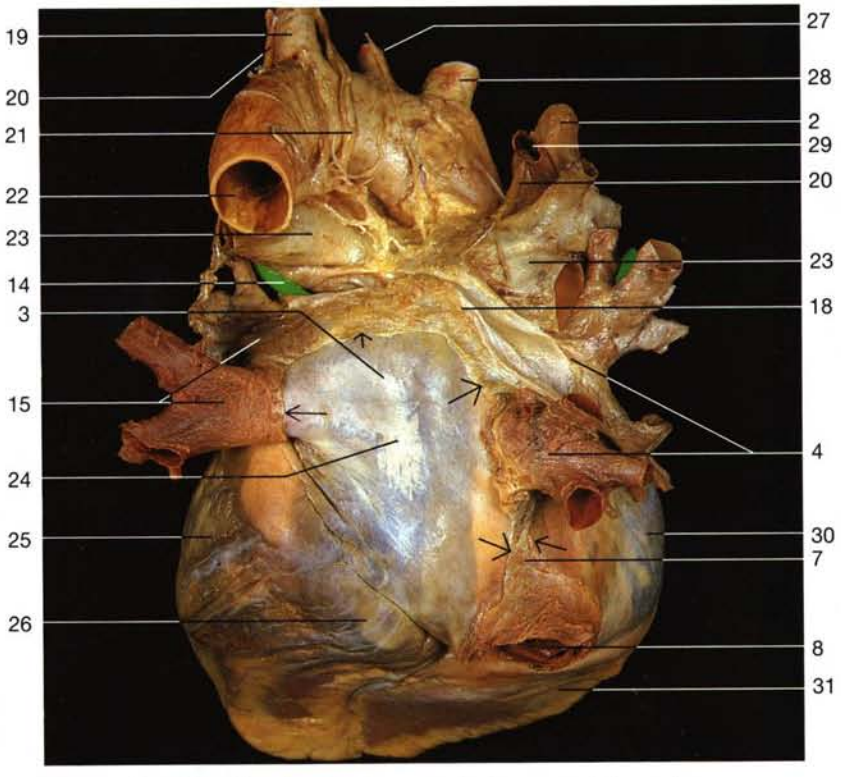

Heart with epicardium (posterior aspect). Arrows: oblique sinus.

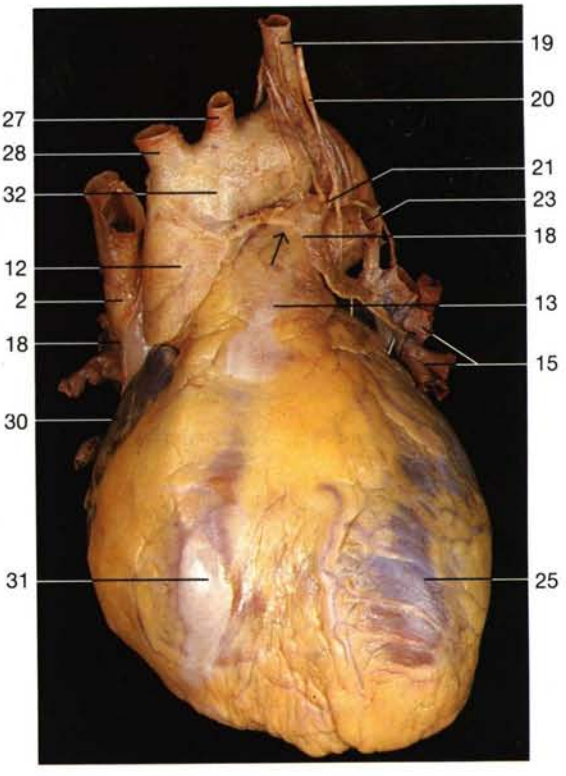

Heart with epicardium (anterior aspect). Arrow: pericardial reflection.

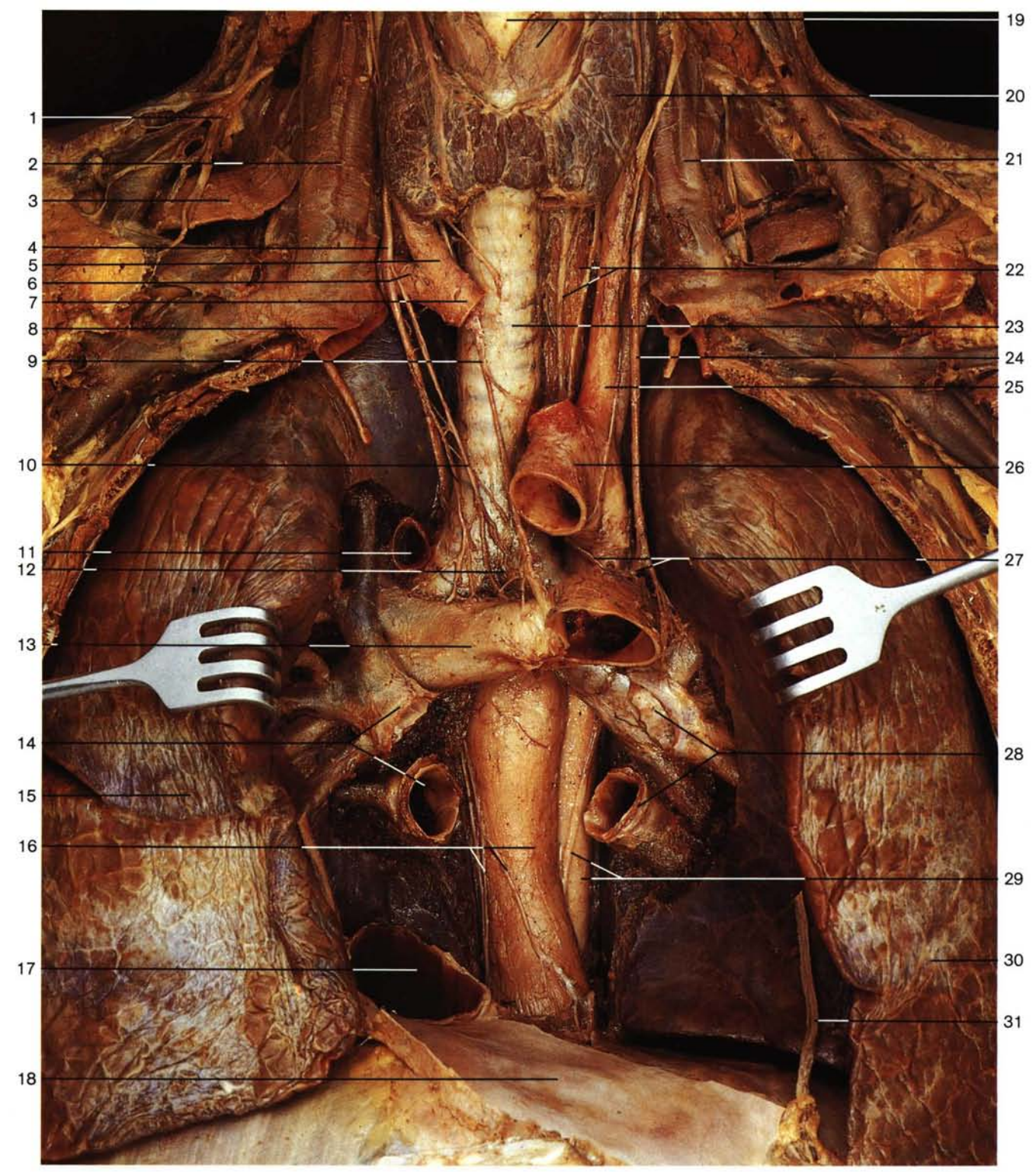

Mediastinal organs after removal of heart and pericardium (ventral aspect). Both lungs have been slightly reflected.

1 Supraclavicular nerves
2 Internal jugular vein
3 Omohyoid muscle
4 **Right vagus nerve**
5 Right common carotid artery
6 Right subclavian artery
7 Brachiocephalic trunk
8 Right brachiocephalic vein
9 Superior cervical cardiac branch of vagus nerve
10 Inferior cervical cardiac branches of vagus nerve
11 Azygos arch (divided)
12 Bifurcation of trachea
13 Right **pulmonary artery**
14 Right **pumonary veins**
15 Right lung
16 **Esophagus** and branches of right vagus nerve
17 Inferior vena cava
18 Pericardium
19 Larynx (thyroid cartilage, cricothyroid muscle)
20 Thyroid gland
21 Internal jugular vein
22 Esophagus and left recurrent laryngeal nerve
23 **Trachea**
24 **Left vagus nerve**
25 Left common carotid artery
26 Aortic arch
27 Left recurrent laryngeal nerve branching off from vagus nerve
28 Left pulmonary veins
29 **Thoracic aorta** and left vagus nerve
30 Left lung
31 Left phrenic nerve (divided)

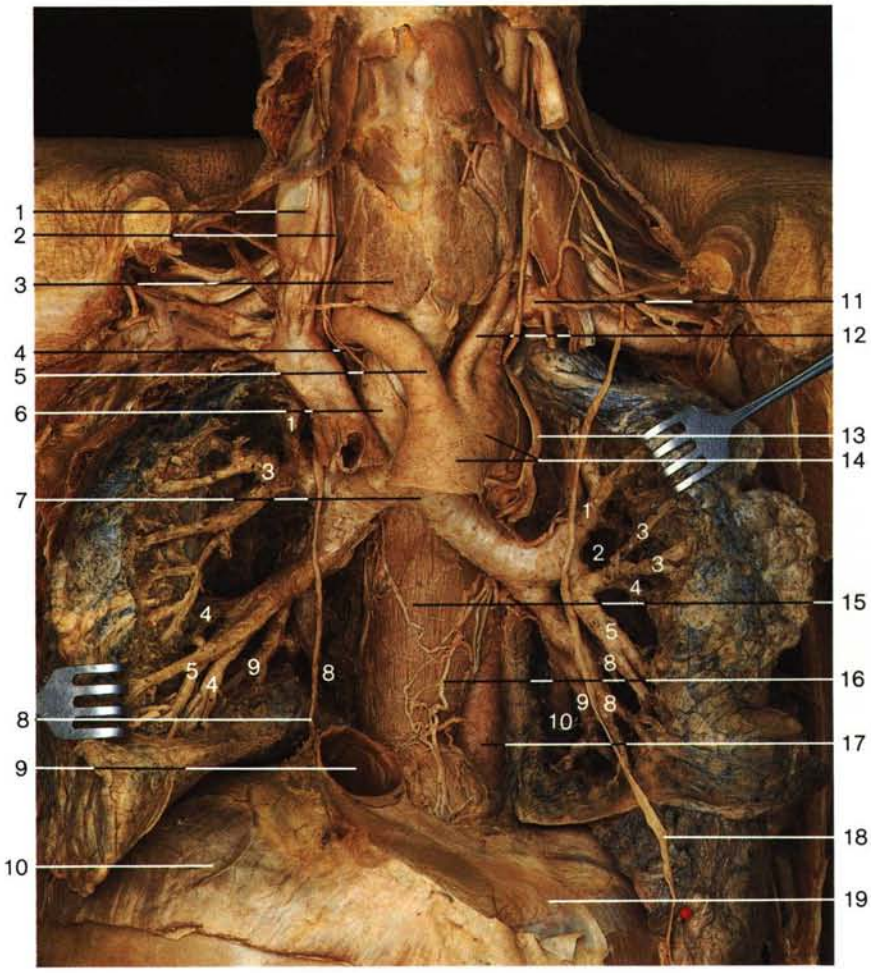

1 Internal jugular vein
2 Right vagus nerve
3 **Thyroid gland**
4 Right recurrent laryngeal nerve
5 Brachiocephalic trunk
6 Trachea
7 **Bifurcation of trachea**
8 **Right phrenic nerve**
9 Inferior vena cava
10 Diaphragm
11 Left subclavian artery
12 Left common carotid artery
13 Left **vagus nerve**
14 Aortic arch
15 **Esophagus**
16 Esophageal plexus
17 Thoracic aorta
18 **Left phrenic nerve**
19 Pericardium at the central tendon of diaphragm
20 Right pulmonary artery
21 Left pulmonary artery
22 Tracheal lymph nodes
23 Superior tracheobronchial lymph nodes
24 Bronchopulmonary lymph nodes

Bronchial tree in situ (ventral aspect). Heart and pericardium have been removed; the bronchi of the bronchopulmonary segments are dissected. 1–10 = numbers of segments.

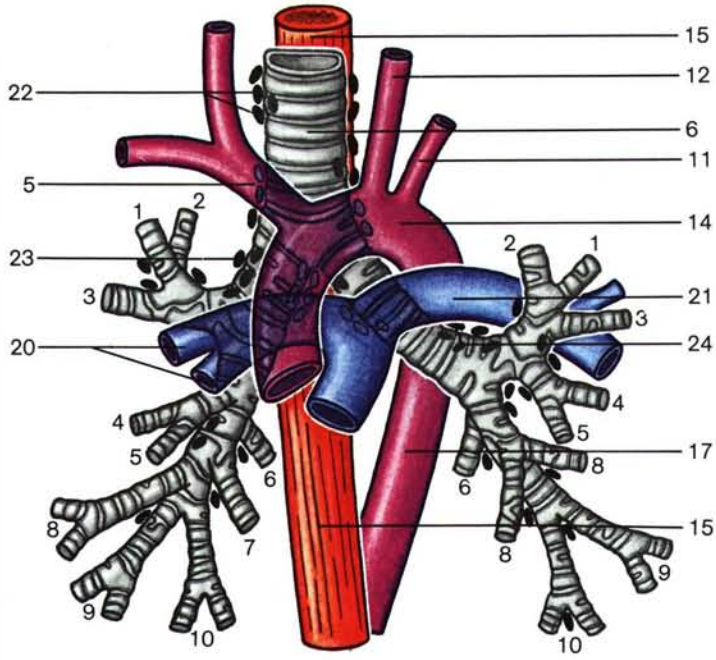

Relation of aorta, pulmonary trunk and esophagus to trachea and bronchial tree. (Schematic drawing.)
1–10 = number of segments (ef. p. 230 and 235).

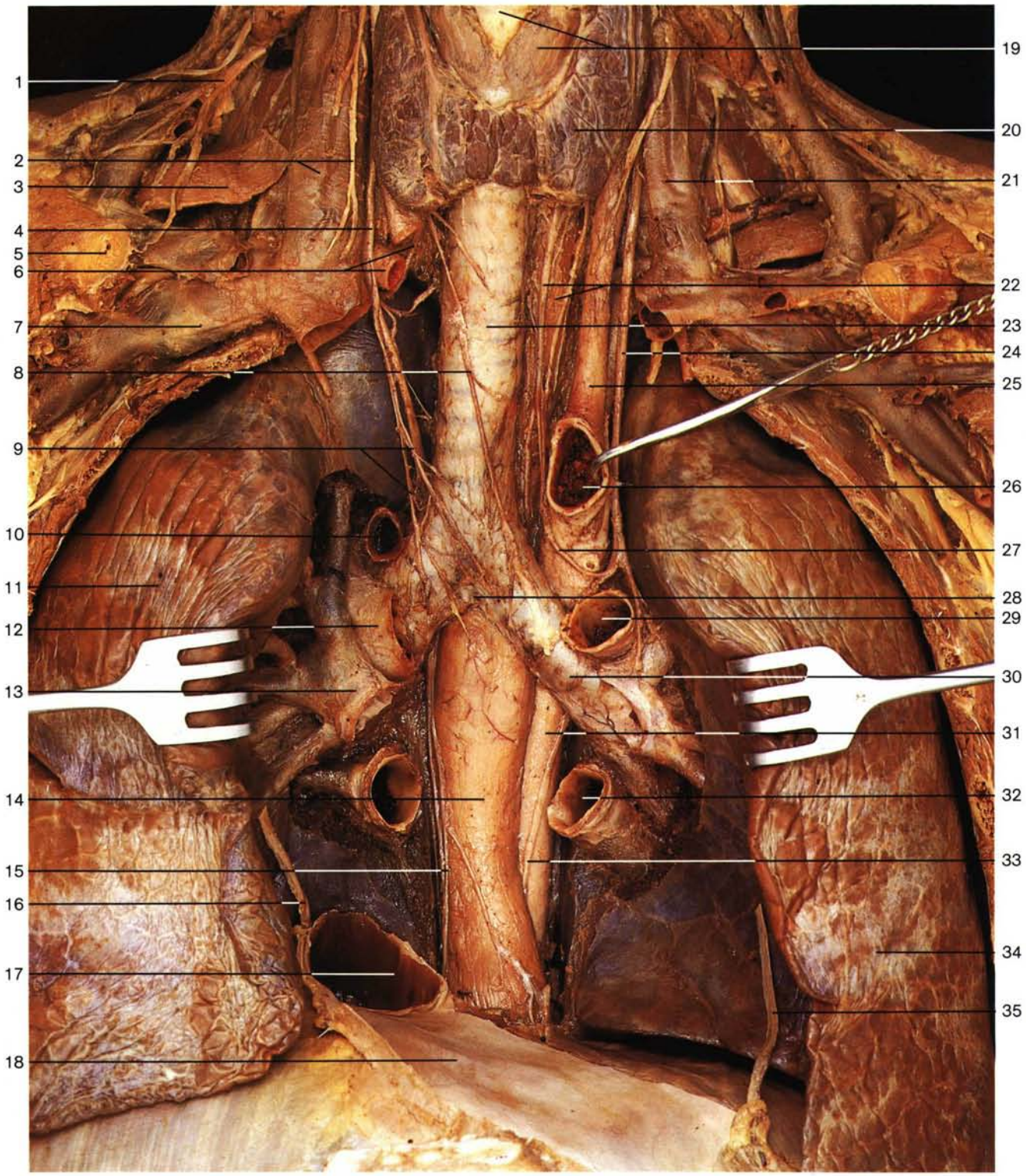

Organs of posterior mediastinum (ventral aspect). The heart with the pericardium has been removed, and the lungs and aortic arch have been slightly reflected to show the vagus nerves and their branches.

1 Supraclavicular nerves
2 Right internal jugular vein with ansa cervicalis
3 Omohyoid muscle
4 **Right vagus nerve**
5 Clavicle
6 Right subclavian artery and recurrent laryngeal nerve
7 Right subclavian vein
8 Superior cervical cardiac branch of vagus nerve
9 Inferior cervical cardiac branch of vagus nerve
10 Azygos arch (divided)
11 **Right lung**
12 **Right pulmonary artery**
13 **Right pulmonary veins**
14 **Esophagus**
15 Esophageal plexus
16 Right phrenic nerve (divided)
17 Inferior vena cava
18 Pericardium covering the diaphragm
19 Larynx (thyroid cartilage and cricothyroid muscle)
20 Thyroid gland
21 Left internal jugular vein
22 Esophagus and left recurrent laryngeal nerve
23 Trachea
24 **Left vagus nerve**
25 Left common carotid artery
26 Aortic arch
27 Left recurrent laryngeal nerve
28 **Bifurcation of trachea**
29 **Left pulmonary artery**
30 Left primary bronchus
31 Descending aorta
32 Left pulmonary veins
33 Branch of left vagus nerve
34 **Left lung**
35 Left phrenic nerve (divided)

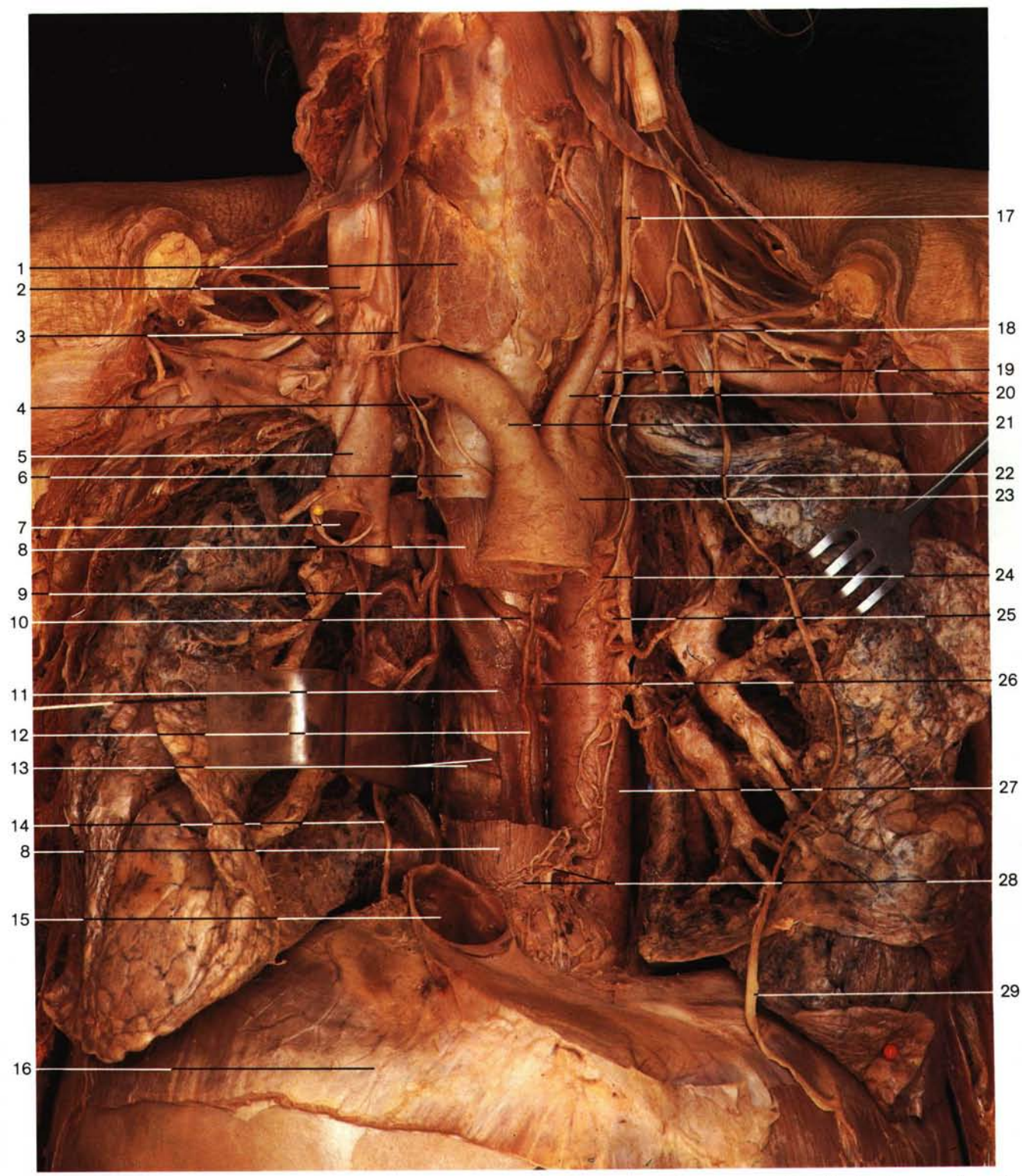

Mediastinal organs (ventral aspect). Heart and distal part of esophagus have been removed to display the vessels and nerves of the posterior mediastinum.

1 Thyroid gland
2 Right internal jugular vein
3 Right vagus nerve
4 Point, where right recurrent laryngeal nerve is branching off the vagus nerve
5 Right brachiocephalic vein
6 Trachea
7 Left brachiocephalic vein (reflected)
8 Esophagus
9 Right bronchial artery
10 Posterior intercostal artery
11 **Azygos vein**
12 **Thoracic duct**
13 Posterior intercostal artery and vein (in front of the vertebral column)
14 Right phrenic nerve
15 Inferior vena cava
16 Diaphragm
17 **Left vagus nerve**
18 Thyrocervical trunk
19 Left subclavian artery
20 Left common carotid artery
21 Brachiocephalic trunk
22 Left vagus nerve
23 **Aortic arch**
24 Left recurrent laryngeal nerve
25 Left bronchial artery
26 Lymph node
27 Thoracic aorta
28 Esophageal plexus
29 Left phrenic nerve

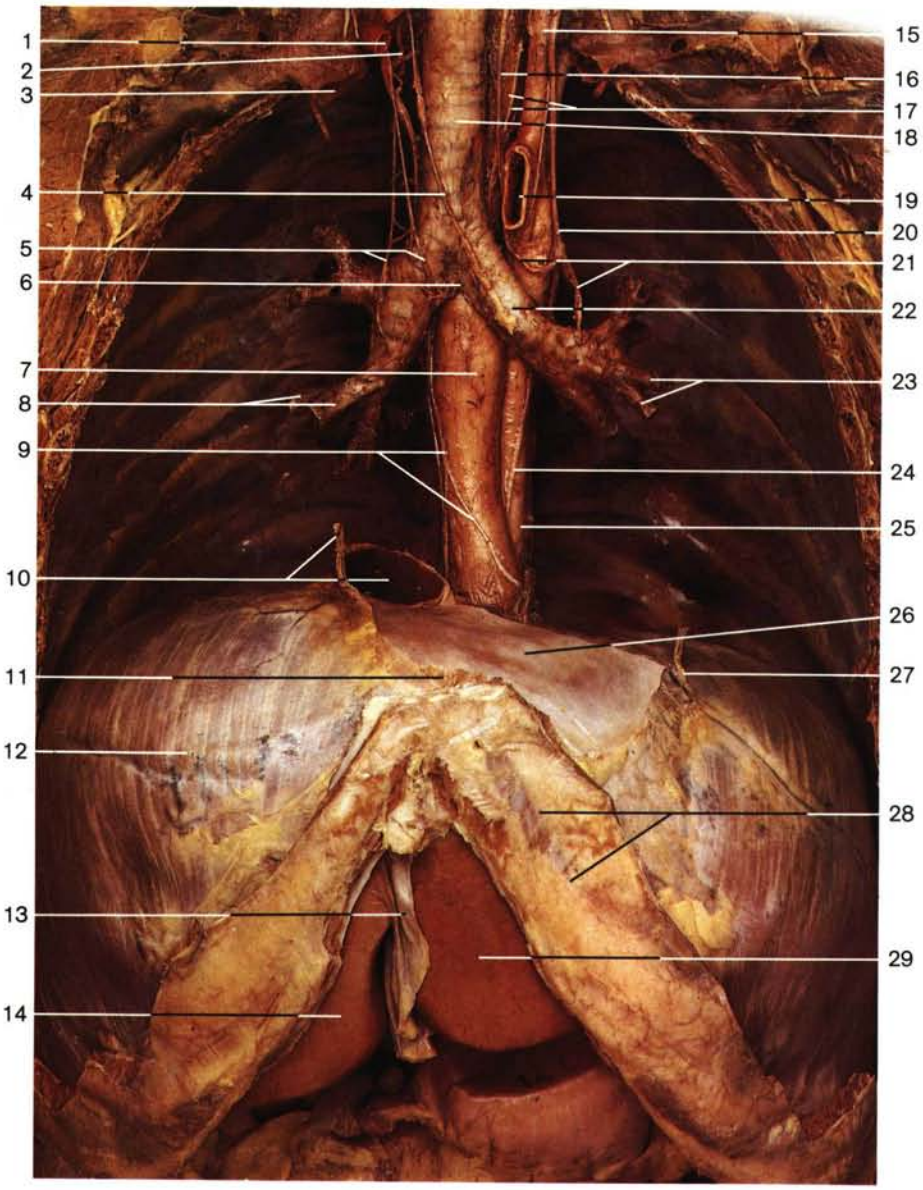

Diaphragm and organs of mediastinum (anterior aspect). Heart and lungs have been removed; the costal margin remains in place. Note the different courses of left and right vagus.

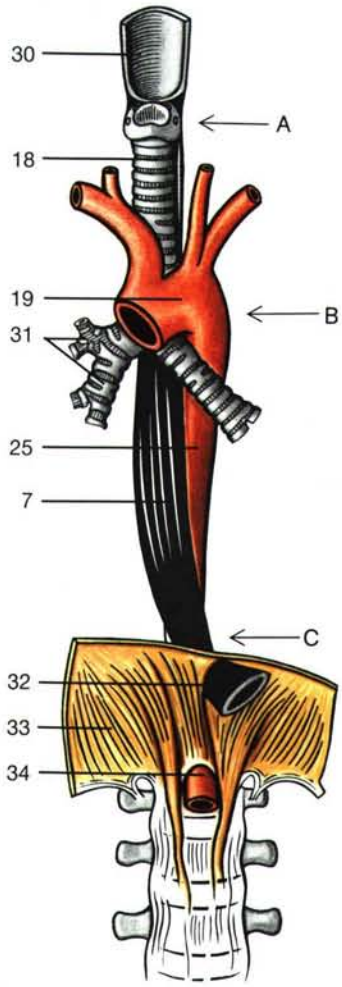

Organs of posterior mediastinum (ventral aspect). (Schematic drawing.) Three regions in which the esophagus is narrowed are shown:

A at the level of the cricoid cartilage;

B at the level of the aortic arch;

C at the level of the diaphragm.

1 Right subclavian artery
2 Right recurrent laryngeal nerve
3 Right brachiocephalic vein
4 Superior cervical cardiac nerve
5 Inferior cervical cardiac nerves and pulmonary branches
6 **Bifurcation of trachea**
7 **Esophagus** (thoracic part)
8 Bronchi of lateral and medial segments of middle lobe
9 Esophageal plexus and branches of right vagus nerve
10 Inferior vena cava and right phrenic nerve (cut)
11 Sternal part of diaphragm
12 Costal part of **diaphragm**
13 Falciform ligament of liver
14 Liver (quadrate lobe)
15 Left common carotid artery
16 Left recurrent laryngeal nerve
17 Esophageal branches of left vagus nerve and esophagus
18 **Trachea**
19 Aortic arch
20 Left vagus nerve
21 Left recurrent laryngeal nerve with inferior cardiac nerve
22 Left primary bronchus
23 Superior and inferior lingular bronchi
24 Esophageal plexus of left vagus nerve
25 **Descending aorta**
26 Central tendon of diaphragm covered with pericardium
27 Left phrenic nerve (divided)
28 Costal margin
29 **Liver,** left lobe
30 Pharynx
31 Secondary bronchi
32 Esophagus (abdominal part)
33 Diaphragm
34 Abdominal aorta

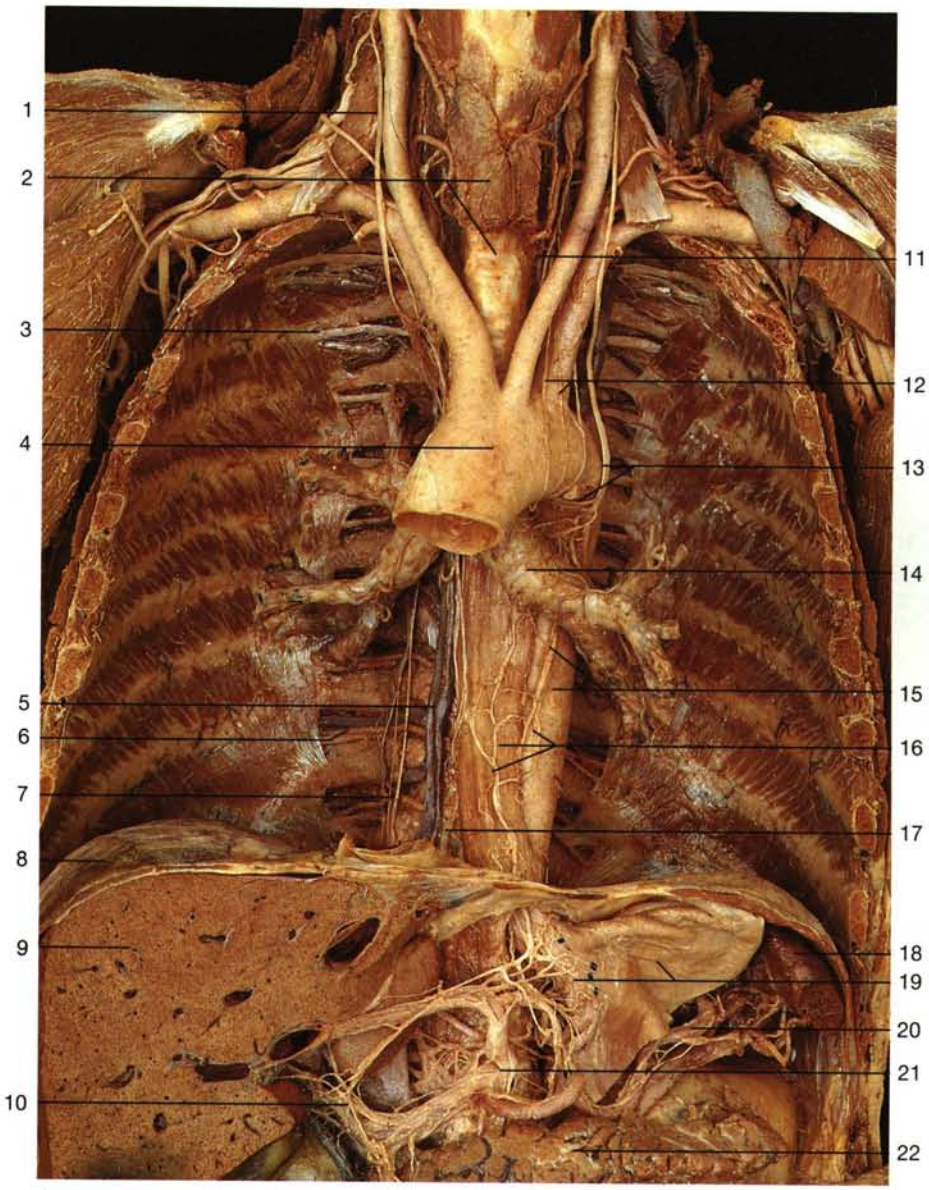

Organs of posterior mediastinum (anterior aspect).

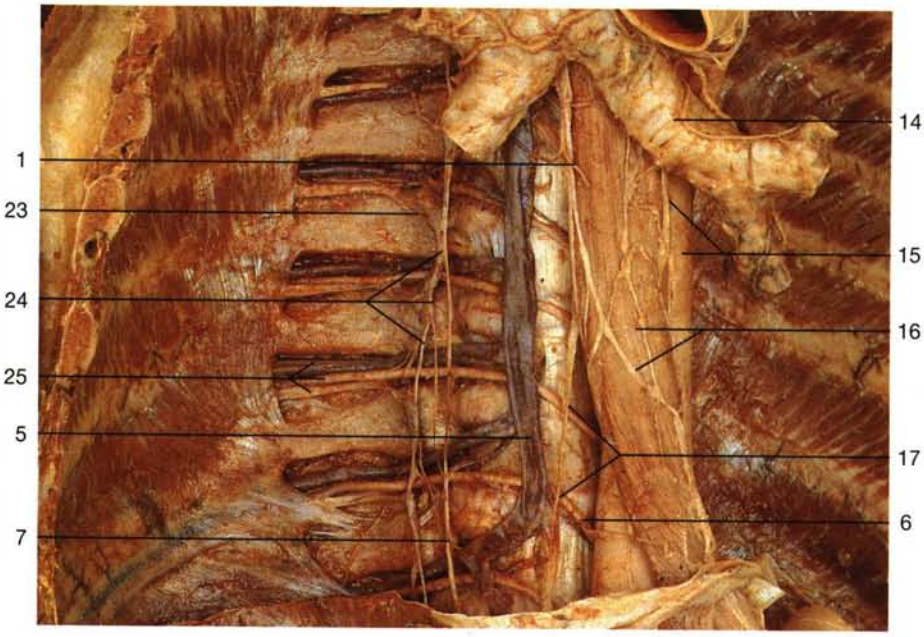

Inferior segment of posterior mediastinum (anterior aspect).

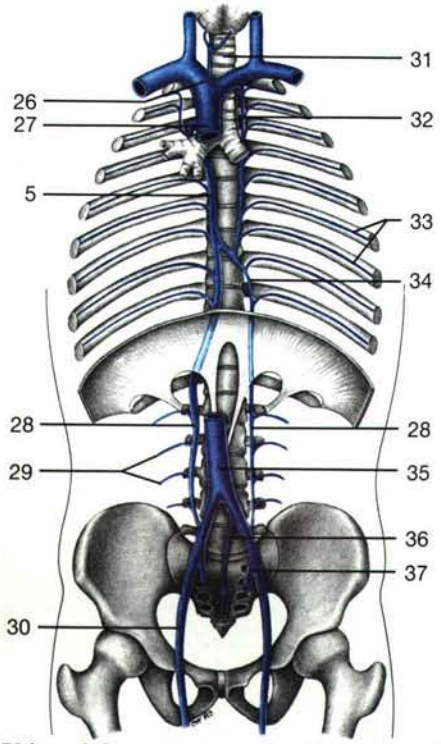

Veins of the posterior wall of thoracic and abdominal cavity. (Schematic drawing.)

1 Right vagus nerve
2 Thyroid gland and trachea
3 Intercostal nerve
4 Aortic arch
5 **Azygos vein**
6 Posterior intercostal artery
7 **Greater splanchnic nerve**
8 Diaphragm
9 Liver
10 Proper hepatic artery and hepatic plexus
11 Left recurrent laryngeal nerve
12 Inferior cervical cardiac nerves
13 Left vagus nerve and left recurrent laryngeal nerve
14 Left primary bronchus
15 Thoracic aorta and left vagus nerve
16 Esophagus and esophageal plexus
17 **Thoracic duct**
18 Spleen
19 Anterior gastric plexus and stomach (divided)
20 Splenic artery and splenic plexus
21 **Celiac trunk** and **celiac plexus**
22 Pancreas
23 Ramus communicans
24 **Sympathetic trunk** and **sympathetic ganglion**
25 **Posterior intercostal vein** and **artery** and **intercostal nerve**
26 Right brachiocephalic vein
27 Superior vena cava
28 Ascending lumbar vein
29 Lumbar veins
30 Right external iliac vein
31 Trachea
32 Accessory hemiazygos vein
33 Posterior intercostal veins
34 Hemiazygos vein
35 Inferior vena cava
36 Median sacral vein
37 Internal iliac vein

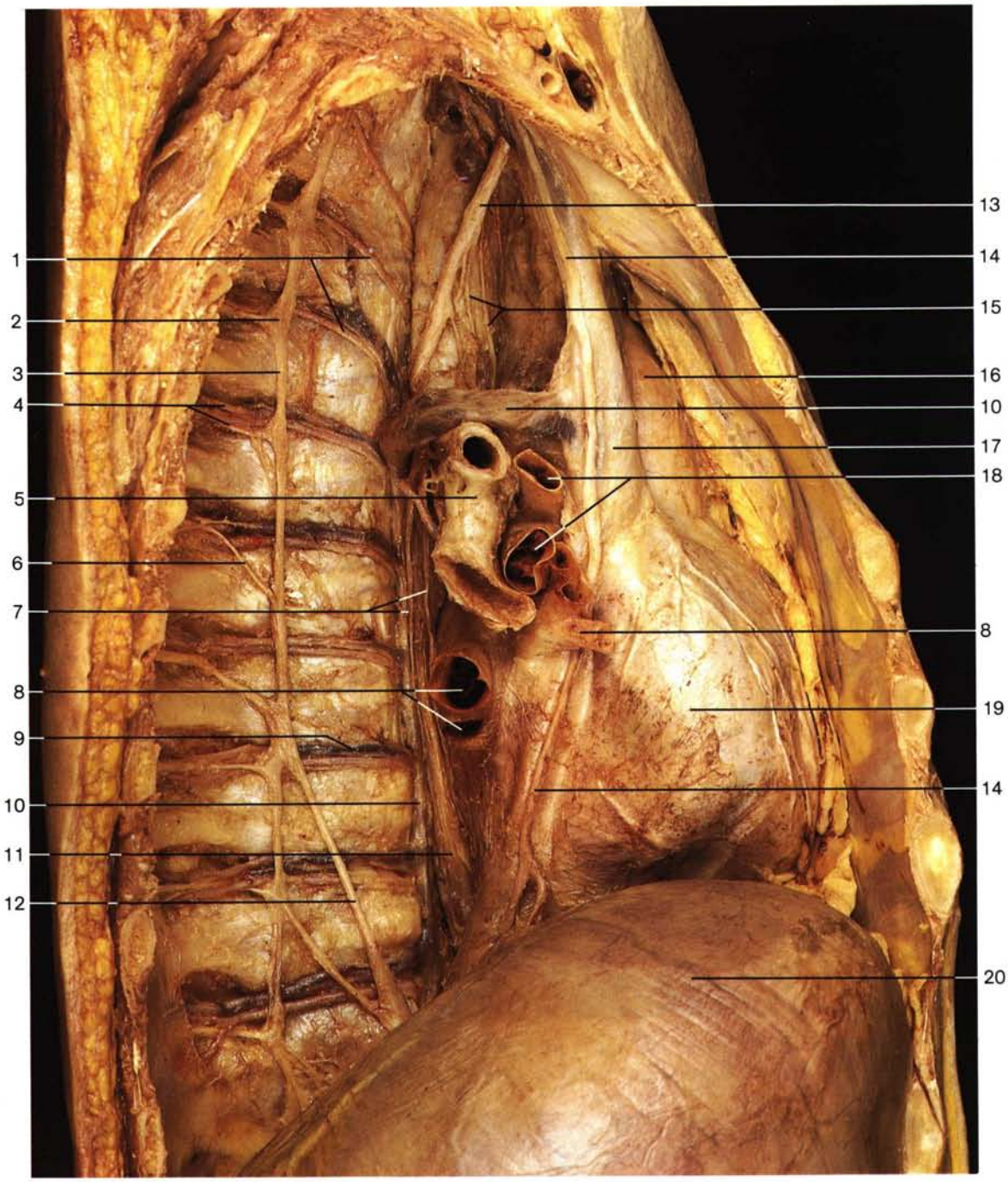

Mediastinal organs (right lateral aspect). Right lung and pleura of right half of the thorax have been removed.

1 Posterior intercostal arteries
2 Ganglion of sympathetic trunk
3 **Sympathetic trunk**
4 Vessels and nerves of the intercostal space (from above: Posterior intercostal vein and artery and intercostal nerve)
5 Right primary bronchus
6 Ramus communicans of sympathetic trunk
7 Esophageal plexus (branches of right vagus nerve)
8 Pulmonary veins
9 Posterior intercostal vein
10 **Azygos vein**
11 Esophagus
12 Greater splanchnic nerve
13 **Right vagus nerve**
14 **Right phrenic nerve**
15 Inferior cervical cardiac branches of vagus nerve
16 **Aortic arch**
17 Superior vena cava
18 Right pulmonary artery
19 **Heart with pericardium**
20 Diaphragm

Organs of posterior and superior mediastinum (left lateral aspect).

1 Subclavian artery
2 Subclavian vein
3 Clavicle (divided)
4 **Left vagus nerve**
5 First rib (divided)
6 Left superior intercostal vein
7 **Left atrium with pericardium**
8 **Left phrenic nerve** and pericardiacophrenic artery and vein
9 Esophageal plexus (branches derived from left vagus nerve)
10 **Apex of heart with pericardium**
11 Brachial plexus
12 Scapula (divided)
13 Posterior intercostal arteries
14 Ramus communicans of sympathetic trunk
15 **Sympathetic trunk**
16 Aortic arch
17 Left vagus nerve and left recurrent laryngeal nerve
18 Left pulmonary artery
19 Left primary bronchus
20 **Thoracic aorta**
21 Pulmonary vein
22 Esophagus (thoracic part)
23 **Diaphragm**
24 Posterior intercostal artery and vein and intercostal nerve

1 Superior vena cava
2 Right atrium
3 Right ventricle
4 **Costal part of diaphragm**
5 Costal margin
6 Position of costodiaphragmatic recess
7 **Lateral arcuate ligament**
8 **Medial arcuate ligament**
9 Right crus of lumbar part of diaphragm
10 Quadratus lumborum muscle
11 Ascending aorta
12 Pulmonary trunk
13 Left ventricle
14 Pericardium, diaphragm
15 Esophageal hiatus and abdominal part of esophagus (cut)
16 **Lumbar part of diaphragm**
17 Aortic hiatus
18 Psoas major muscle
19 Lumbar vertebra

Diaphragm in situ (anterior aspect). Anterior walls of thoracic and abdominal cavities have been removed. Natural position of the heart above the central tendon on the diaphragm is shown.

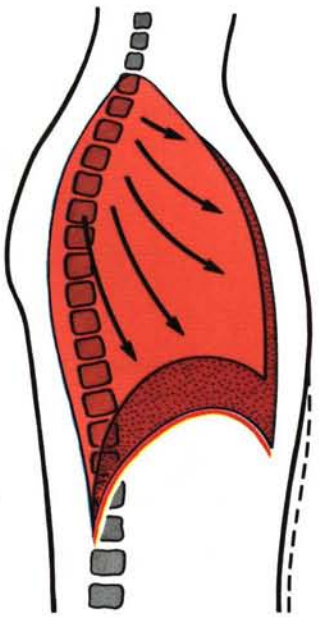

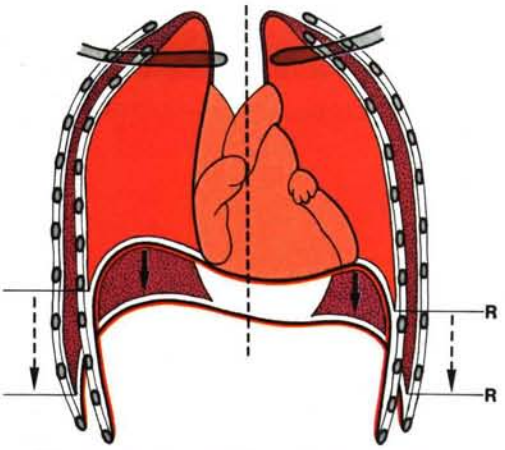

Changes in the position of the diaphragm and thoracic cage during respiration. Left: lateral aspect; right: anterior aspect. During inspiration the diaphragm moves downwards and the lower part of the thoracic cage expands forward and laterally, causing the costodiaphragmatic recess (R) to enlarge (cf. dotted arrows).

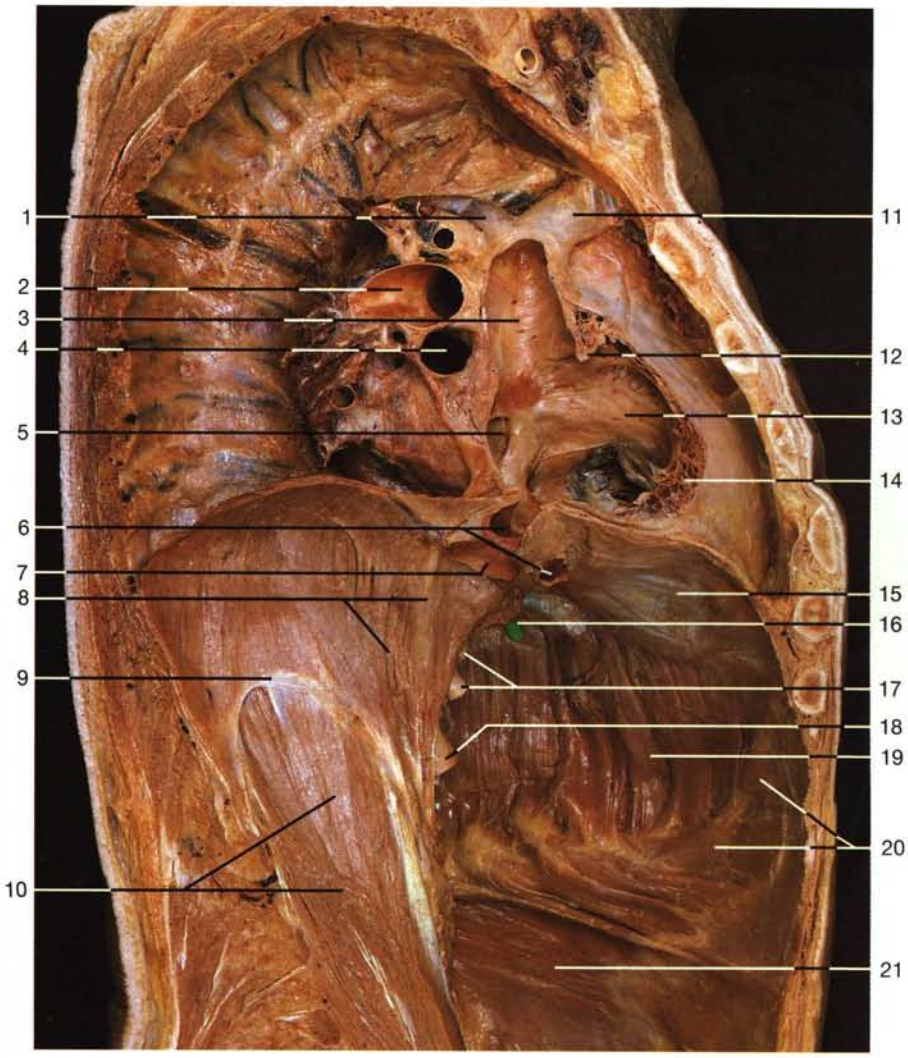

1 Azygos venous arch
2 Right pulmonary artery
3 **Superior vena cava**
4 Right pulmonary vein
5 Fossa ovalis
6 Hepatic veins
7 Inferior vena cava
8 Right crus of lumbar part of **diaphragm**
9 Medial arcuate ligament
10 Psoas major muscle
11 Left brachiocephalic vein
12 Terminal crista
13 **Right atrium**
14 Right auricle
15 Central tendon of diaphragm
16 Esophagus
17 Celiac trunk and superior mesenteric artery
18 Aorta
19 Costal part of diaphragm
20 Costal margin
21 Transversus abdominis muscle

Diaphragm. Paramedian section to the right of the median plane through thoracic and upper abdominal cavity. The plane passes through the superior and inferior vena cava just to the right of the vertebral bodies. Most of the heart remains in situ to the left of this plane (specimen is viewed from the right side).

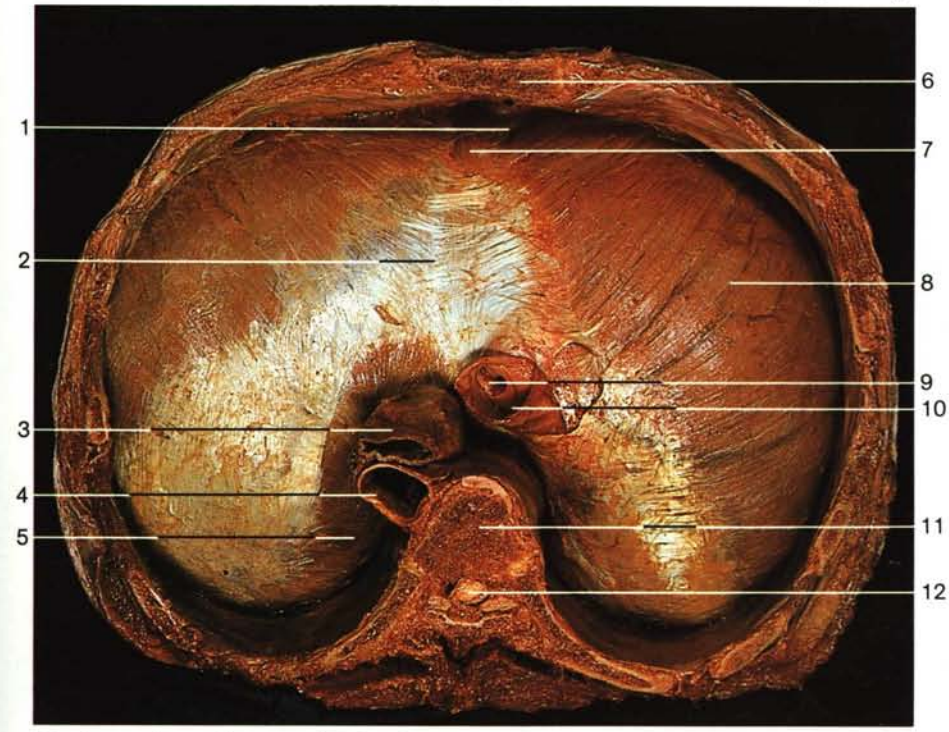

1 Sternocostal triangle
2 Central tendon (from above)
3 Esophagus
4 Aorta
5 Lumbar part of diaphragm
6 Sternum
7 Sternal part of diaphragm
8 Costal part of diaphragm
9 Entrance of hepatic veins
10 Inferior vena cava
11 Body of 9th thoracic vertebra
12 Spinal cord

Diaphragm (superior aspect). The pleura and thoracic wall have been removed.

Coronal section through the thorax at the level of ascending aorta (anterior aspect).

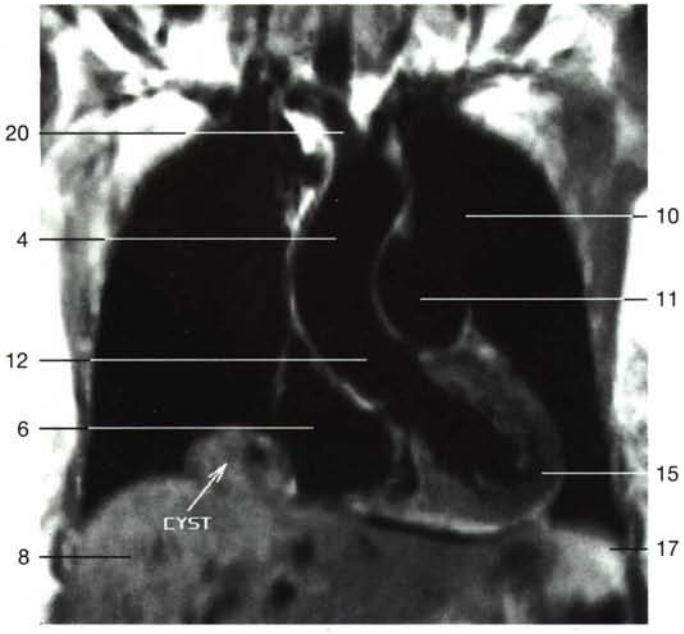

Coronal section through the thorax at the level of ascending aorta. (MR-Scan.)

1 Clavicle
2 Left brachiocephalic vein
3 Superior lobe of right lung
4 Aortic arch
5 Superior vena cava
6 Right atrium (entrance of inferior vena cava)
7 Coronary sinus
8 Liver
9 Second rib
10 Superior lobe of left lung
11 Pulmonary trunk
12 Ascending aorta and left coronary artery
13 Aortic valve
14 Pericardium
15 Myocardium of left ventricle
16 Lower lobe of left lung
17 Diaphragm
18 Colic flexures
19 Stomach
20 Brachiocephalic trunk

Coronal section through the thorax at the level of superior and inferior vena cava (anterior aspect).

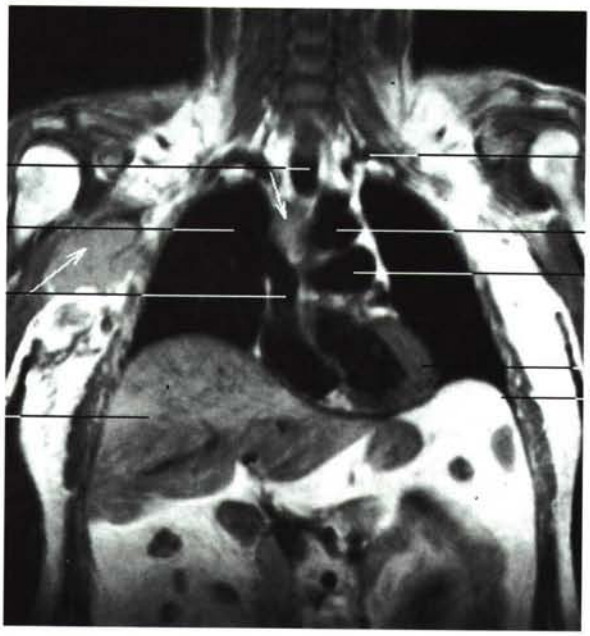

Coronal section through the thorax at the level of superior vena cava. (MR-Scan.)

1 Trachea
2 Upper lobe of right lung
3 **Superior vena cava**
4 Right pulmonary veins
5 Inferior vena cava and right atrium
6 Liver
7 Left common carotid artery
8 Left subclavian vein
9 Upper lobe of left lung
10 Aortic arch
11 Left pulmonary artery
12 Left auricle
13 **Left atrium** with orifices of pulmonary veins
14 Left ventricle (myocardium)
15 Pericardium
16 Diaphragm
17 Left colic flexure
18 Stomach
19 Left subclavian artery

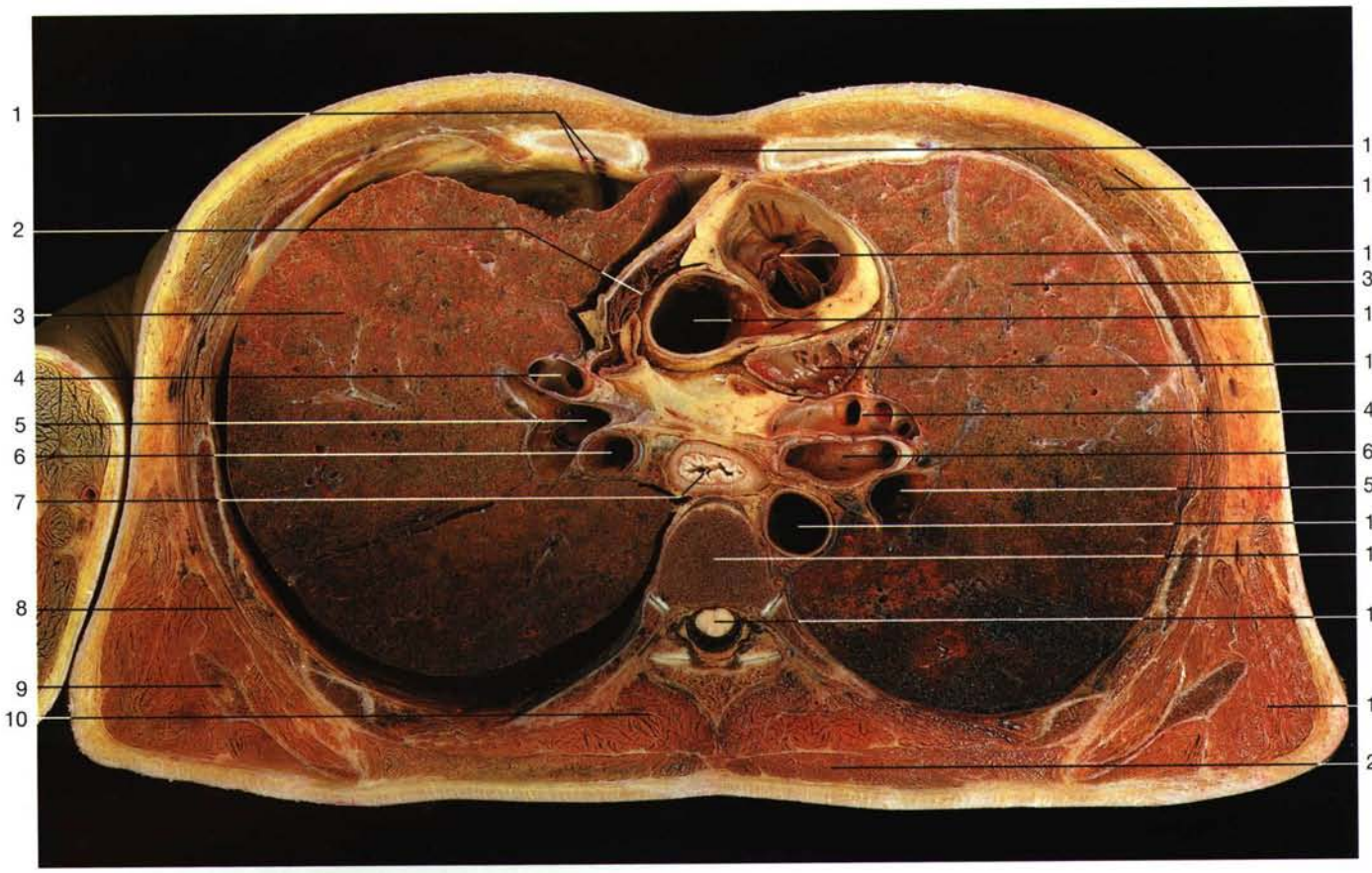

Horizontal section through the thorax at level 1 (from below).

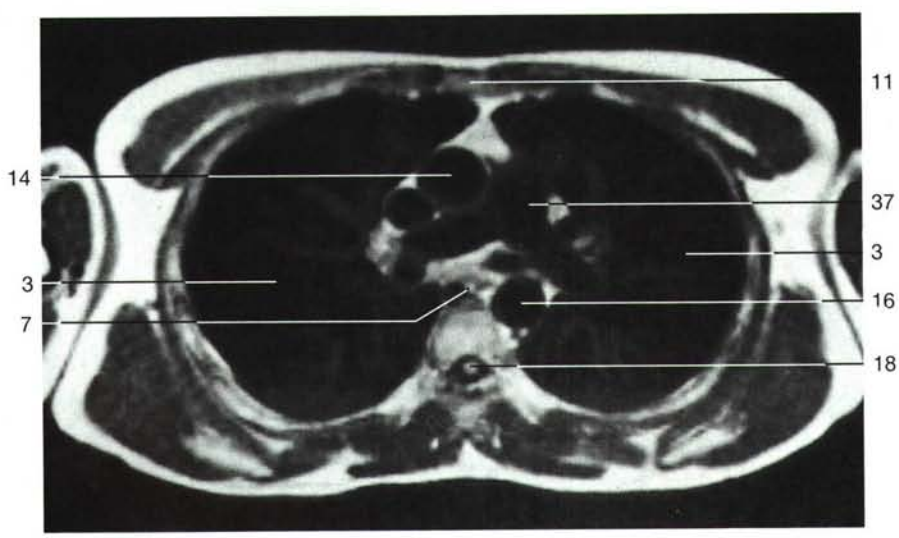

Horizontal section through the thorax at level 1 (from below). (MR-Scan.)

1 Internal thoracic artery and vein
2 Right atrium
3 Lung
4 Pulmonary artery
5 Pulmonary vein
6 Primary bronchus
7 Esophagus
8 Serratus anterior muscle
9 Scapula
10 Longissimus thoracis muscle
11 Sternum
12 Pectoralis major and minor muscles
13 Conus arteriosus (right ventricle), pulmonic valve
14 Ascending aorta and left coronary artery (only in upper figure)
15 **Left atrium**
16 **Descending aorta**
17 Thoracic vertebra
18 Spinal cord
19 Latissimus dorsi muscle
20 Trapezius muscle

Horizontal section through the thorax at level 2 (from below).

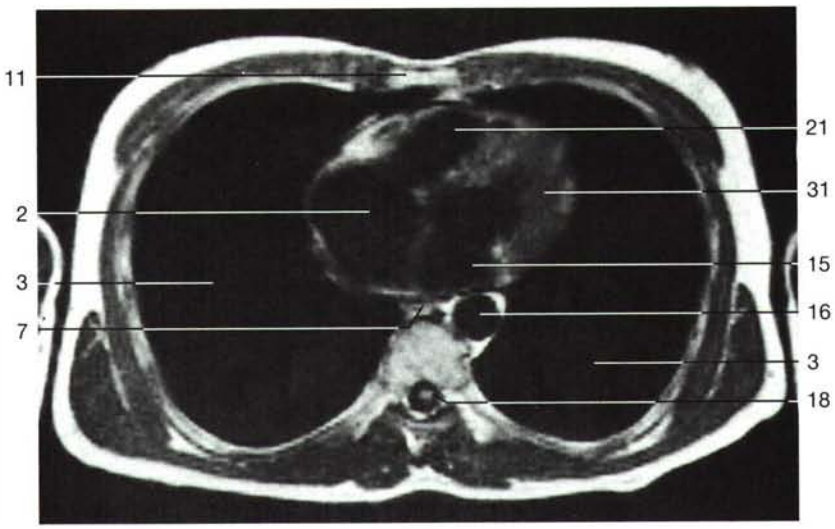

Horizontal section through the thorax at level 2 (from below). (MR-Scan.)

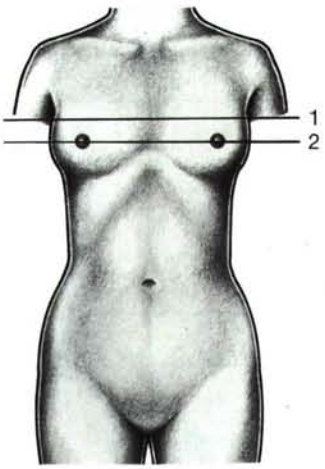

Levels of sections.

21 **Right ventricle**
22 Right coronary artery
23 Right atrioventricular valve
24 Lung (upper lobe)
25 **Left atrium**
26 Pulmonary veins
27 Lung (lower lobe)
28 Erector muscle of spine
29 Third costal cartilage
30 Nipple
31 **Left ventricle**
32 Pericardium
33 Left atrioventricular valve
34 Left coronary artery and coronary sinus
35 Accessory hemiazygos vein
36 Serratus anterior muscle
37 Pulmonary trunk

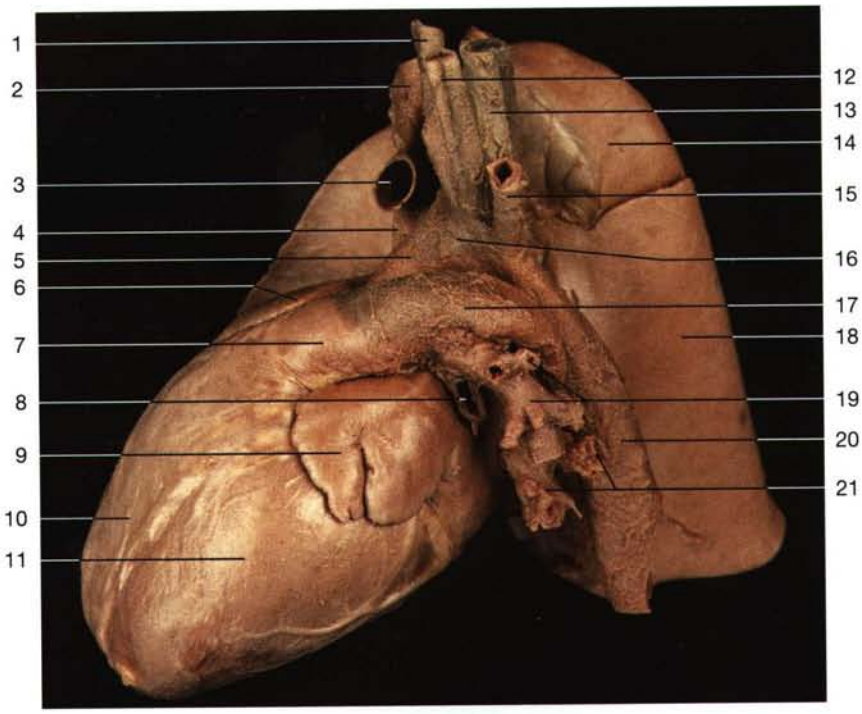

Heart and right lung of the fetus (viewed from left side). The left lung has been removed. Note the ductus arteriosus (Botalli).

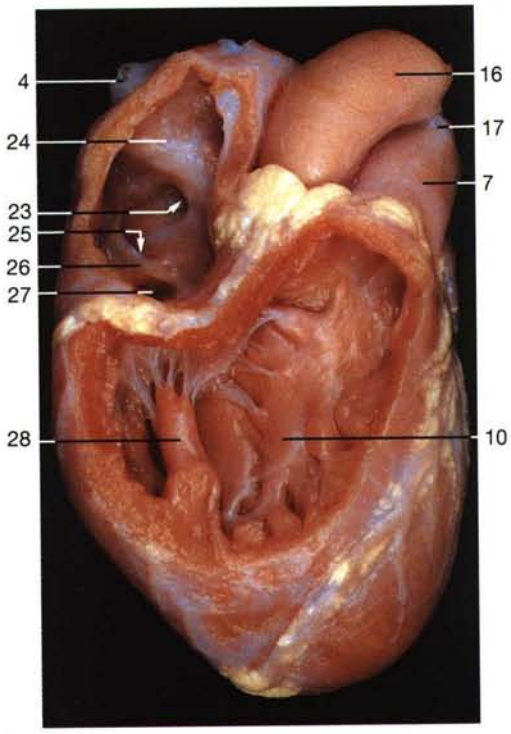

Heart of the fetus (anterior aspect). Right atrium and ventricle opened.

Shunts in the fetal circulation system		
1. Ductus venosus (of Arantius)	between umbilical vein and inferior vena cava	bypass of liver circulation
2. Foramen ovale	between right and left atrium	bypass of pulmonary circulation
3. Ductus arteriosus (Botalli)	between pulmonary trunk and aorta	

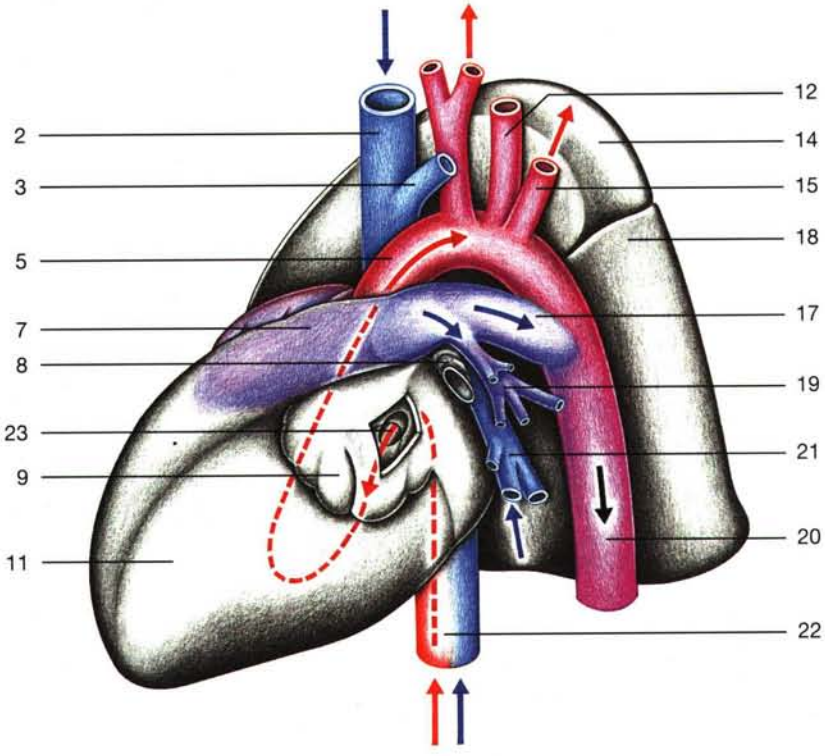

1 Right common carotid artery
2 Right brachiocephalic vein
3 Left brachiocephalic vein
4 Superior vena cava
5 **Ascending aorta**
6 Right auricle
7 **Pulmonary trunk**
8 Left primary bronchus
9 Left auricle
10 Right ventricle
11 Left ventricle
12 Left common carotid artery
13 Trachea
14 Superior lobe of right lung
15 Left subclavian artery
16 Aortic arch
17 **Ductus arteriosus (Botalli)**
18 Inferior lobe of right lung
19 Left pulmonary artery with branches to the left lung
20 **Descending aorta**
21 Left pulmonary veins
22 Inferior vena cava
23 **Foramen ovale**
24 Right atrium
25 Opening of inferior vena cava
26 Valve of inferior vena cava (Eustachian valve)
27 Opening of coronary sinus
28 Anterior papillary muscle of right ventricle

◁ **Heart of the fetus** (schematic drawing). Direction of blood flow indicated by arrows. Note the change in oxygenation of blood after ductus arteriosus entry into aorta.

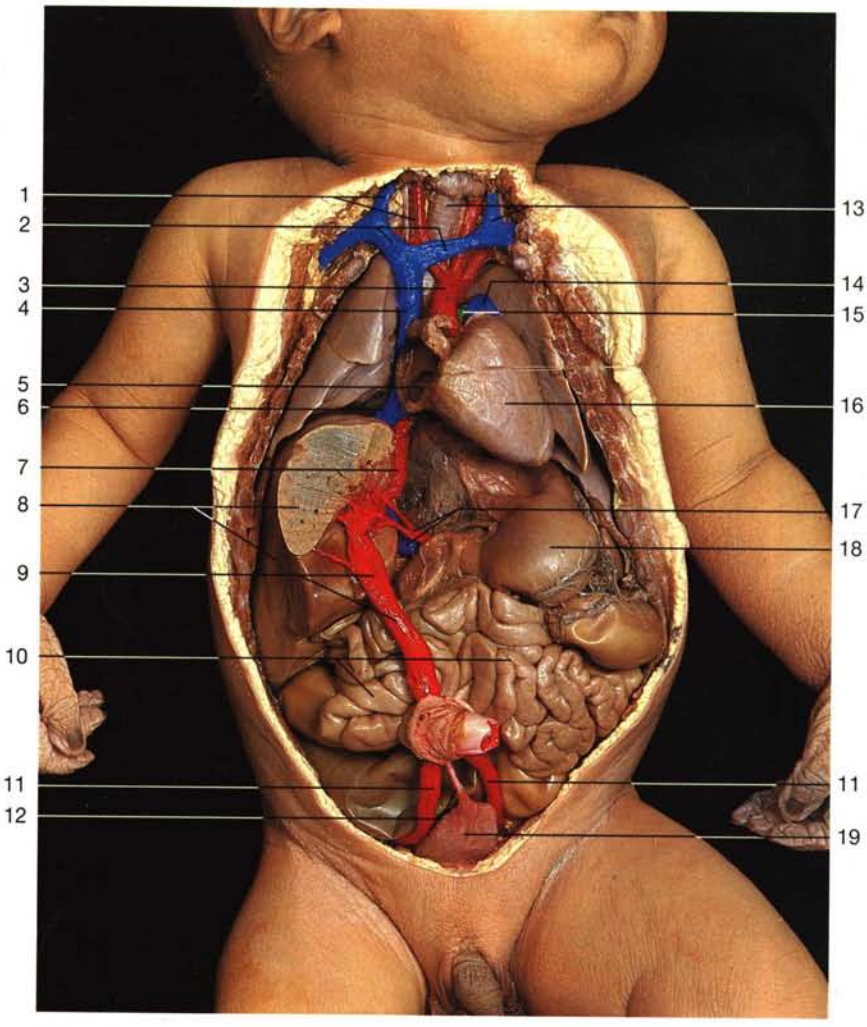

Thoracic and abdominal organs in the newborn (anterior aspect). The right atrium has been opened to show the foramen ovale. The left lobe of the liver has been removed.

1 Internal jugular vein and right common carotid artery
2 Right and left brachiocephalic vein
3 Aortic arch
4 Superior vena cava
5 **Foramen ovale**
6 Inferior vena cava
7 **Ductus venosus**
8 Liver
9 **Umbilical vein**
10 Small intestine
11 **Umbilical artery**
12 Urachus
13 Trachea and left internal jugular vein
14 Left pulmonary artery
15 **Ductus arteriosus** (Botalli)
16 Right ventricle
17 Hepatic arteries (red) and portal vein (blue)
18 Stomach
19 Urinary bladder
20 Portal vein
21 Pulmonary veins
22 Descending aorta
23 Placenta

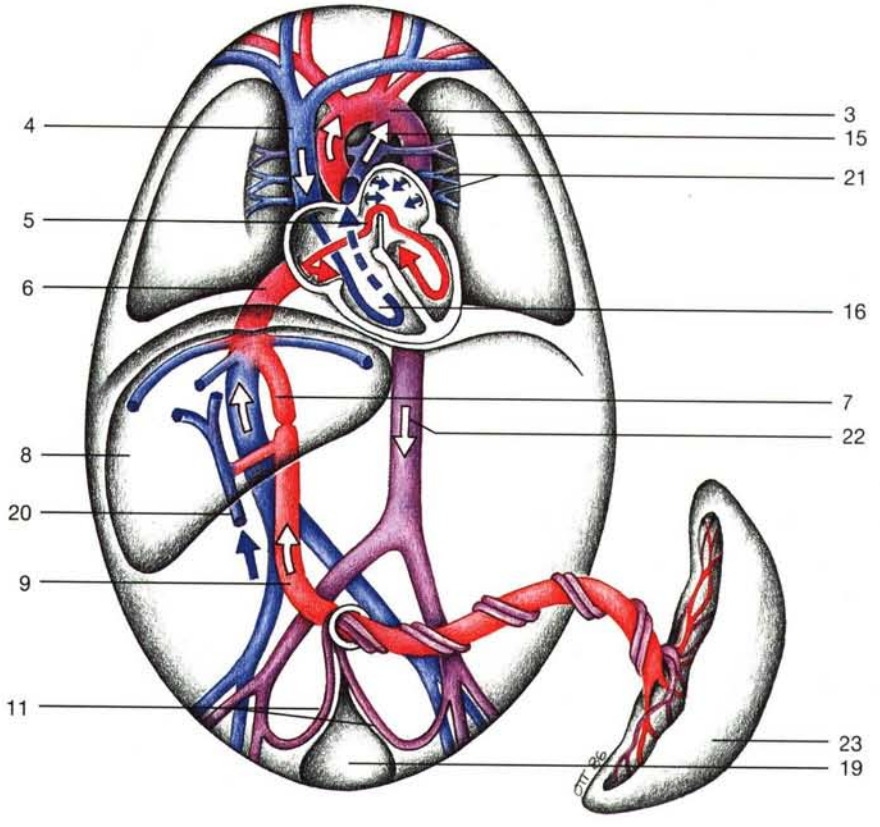

◁ **Fetal circulatory system.** (Schematic drawing.) The oxygen gradient is indicated by color.

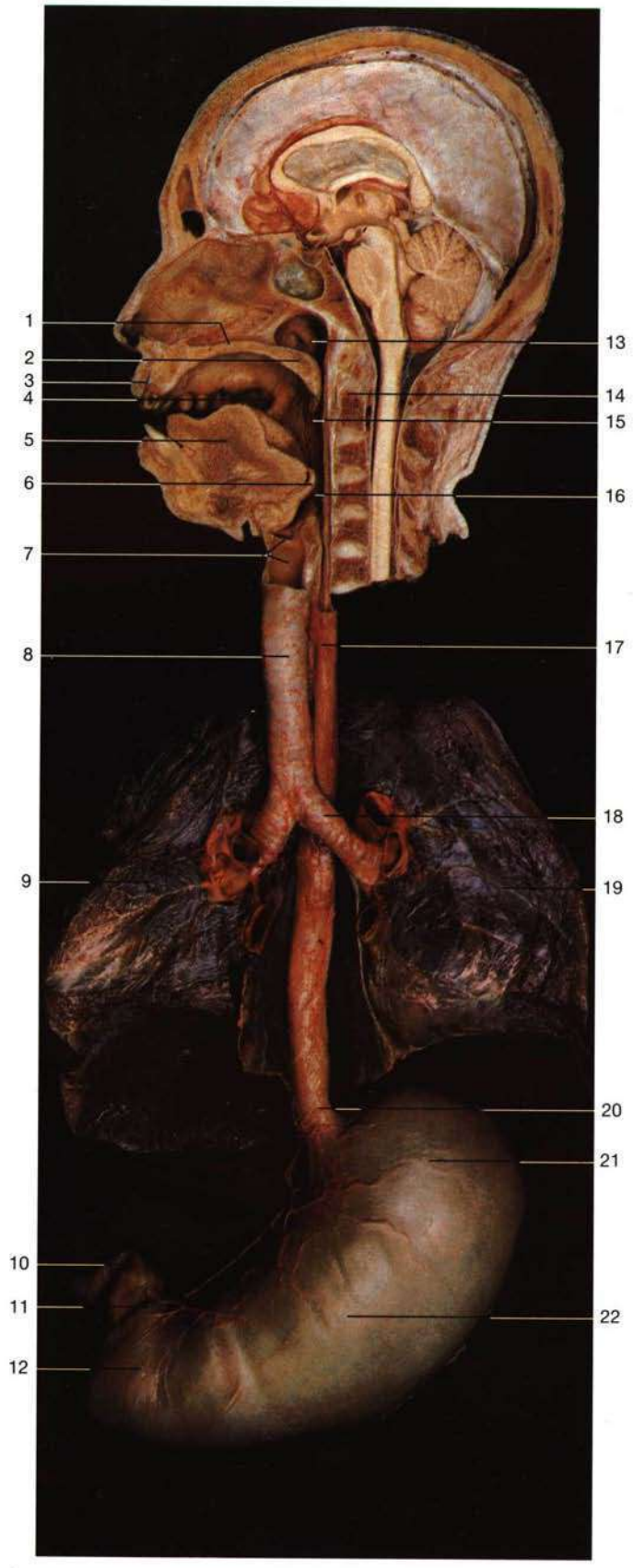

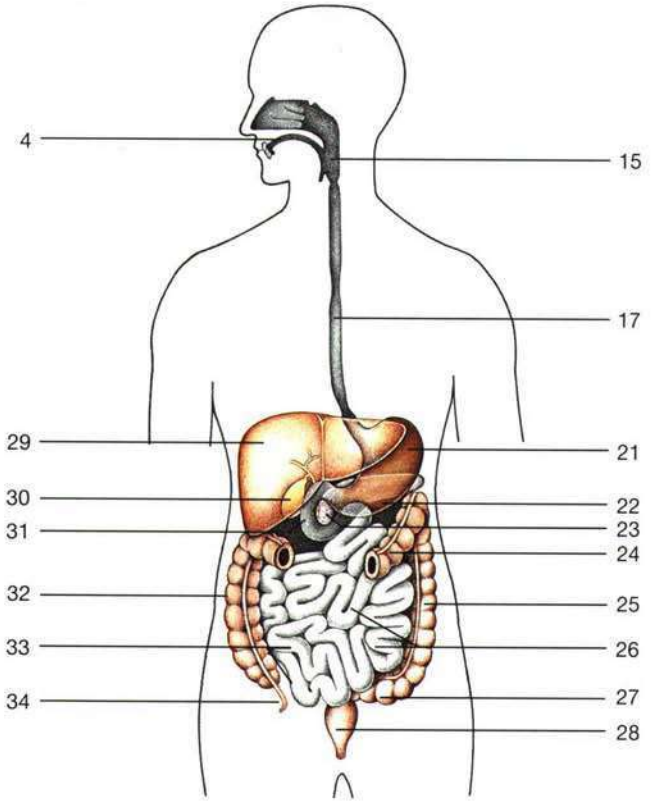

Organization of digestive system. Position of abdominal organs.

1 Hard palate
2 Soft palate with uvula
3 Vestibule of the mouth
4 Oral cavity proper
5 Tongue
6 Epiglottis
7 Vocal ligament and larynx
8 Trachea
9 Right lung
10 Superior part of duodenum
11 Pylorus
12 Pyloric antrum
13 Nasopharynx
14 Dens of axis
15 Oropharynx
16 Laryngopharynx
17 Esophagus (thoracic part)
18 Left primary bronchus
19 Left lung
20 Abdominal part of esophagus and cardia
21 Fundus of stomach
22 Body of stomach
23 Pancreas
24 Transverse colon (divided)
25 Descending colon
26 Jejunum
27 Sigmoid colon
28 Rectum
29 Liver
30 Gall bladder
31 Duodenum
32 Ascending colon
33 Ileum
34 Vermiform appendix

Survey of upper portion of digestive system.
Oral cavity, pharynx, esophagus, and stomach.

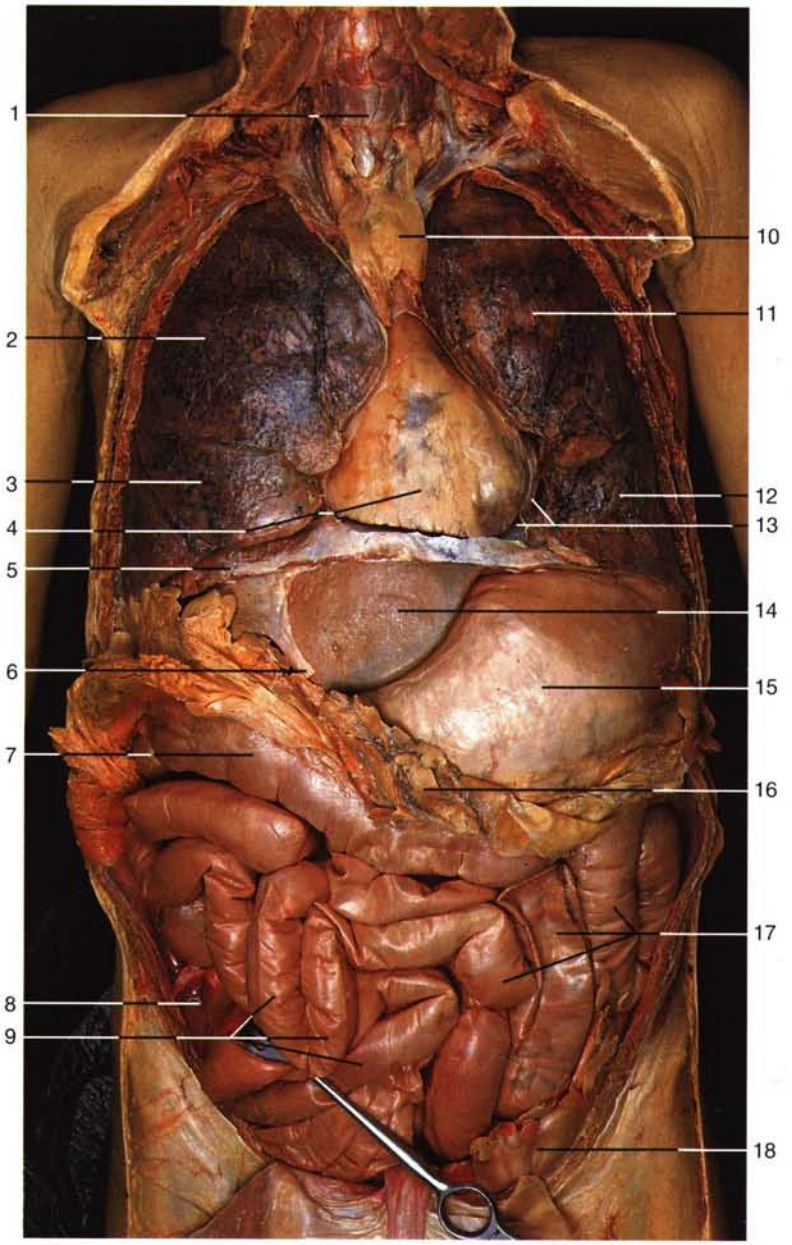

Abdominal organs in situ. The greater omentum has been partly removed or reflected.

1 Thyroid gland
2 Upper lobe of right lung
3 Middle lobe of right lung
4 Heart
5 Diaphragm
6 Round ligament of liver (ligamentum teres)
7 **Transverse colon**
8 Cecum
9 Small intestine (ileum)
10 Thymus
11 Upper lobe of left lung
12 Lower lobe of left lung
13 Pericardium (cut edge)
14 **Liver** (left lobe)
15 **Stomach**
16 Greater omentum
17 **Small intestine** (jejunum)
18 Sigmoid colon
19 Rectus abdominis muscle
20 Small intestine (section)
21 Rib
22 Common bile duct, duodenum and pancreas
23 Inferior vena cava
24 Liver
25 Body of second lumbar vertebra
26 Right kidney
27 Cauda equina and dura mater
28 Linea alba
29 Stomach and pylorus
30 Superior mesenteric artery and vein
31 Abdominal aorta
32 Left renal artery and vein
33 Left kidney
34 Psoas major muscle
35 Deep muscles of the back
36 Pancreas adjacent to lesser sac (bursa omentalis)
37 Falciform ligament with ligamentum teres

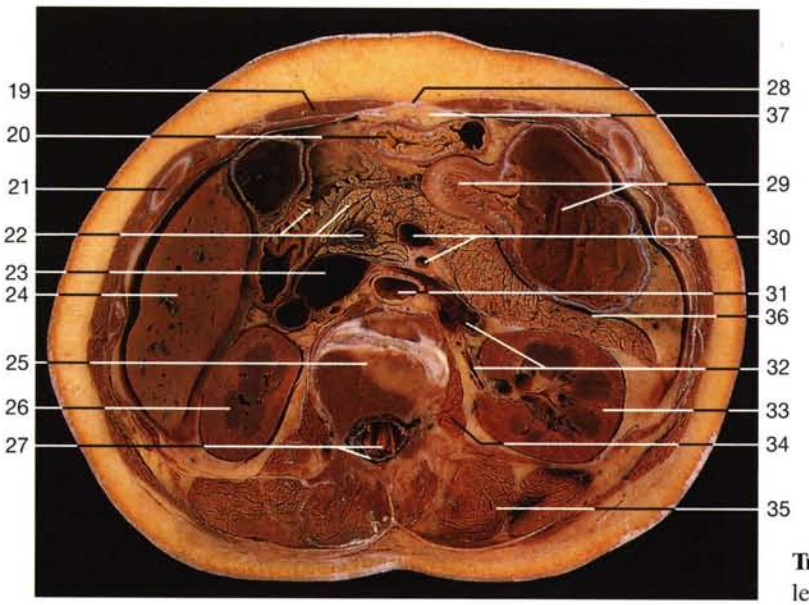

Transverse section through the abdominal cavity at the level of the second lumbar vertebra (from below).

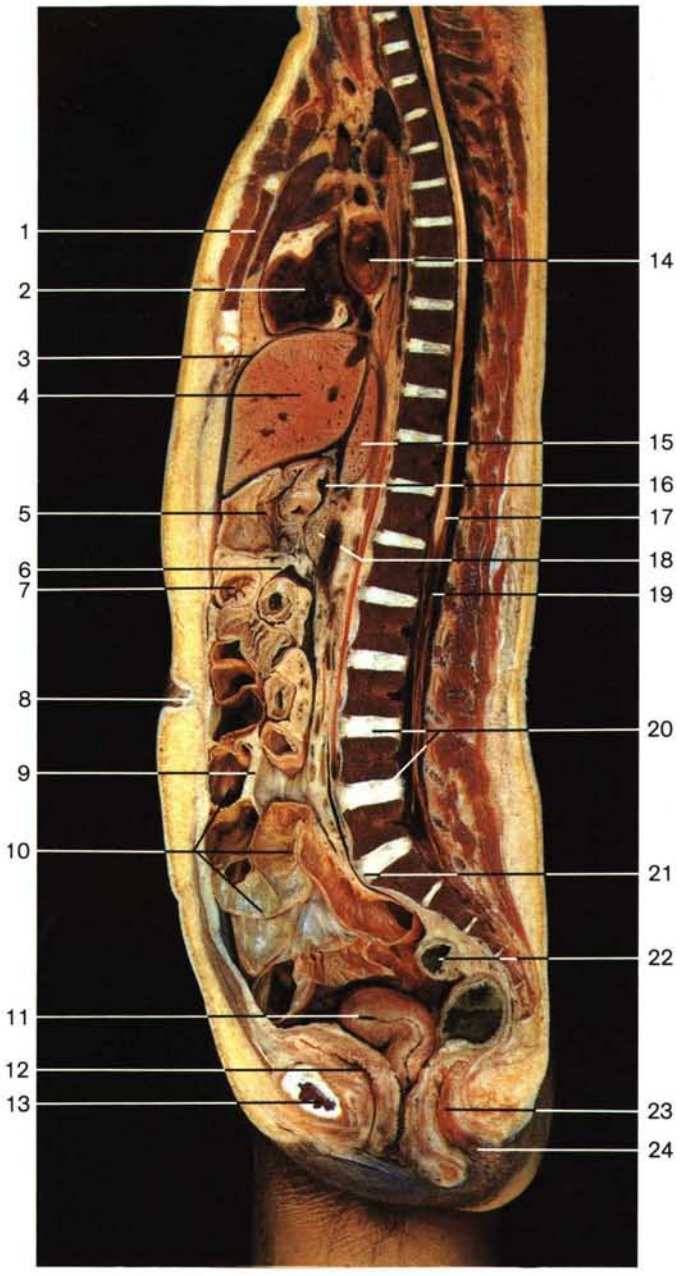

Midsagittal section through the trunk (female).

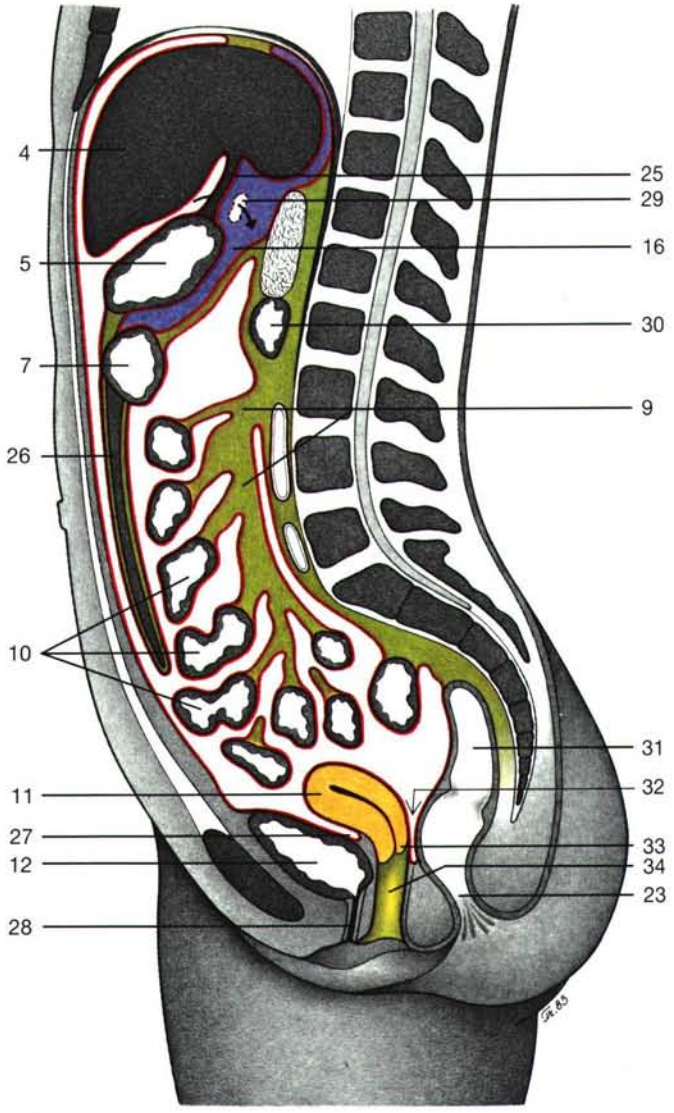

Midsagittal section through the tunk (female).
(Schematic drawing.)
Blue = omental bursa; red = peritoneum.

1 Sternum
2 Right ventricle of heart
3 Diaphragm
4 Liver
5 Stomach
6 Transverse mesocolon
7 Transverse colon
8 Umbilicus
9 Mesentery
10 Small intestine
11 Uterus
12 Urinary bladder
13 Pubic symphysis
14 Left atrium of heart
15 Caudate lobe of liver
16 Omental bursa or lesser sac
17 Conus medullaris
18 Pancreas
19 Cauda equina
20 Intervertebral discs (lumbar vertebral column)
21 Sacral promontory
22 Sigmoid colon
23 Anal canal
24 Anus
25 Lesser omentum
26 Greater omentum
27 Vesicouterine pouch
28 Urethra
29 Epiploic (omental) foramen
30 Duodenum
31 Rectum
32 Rectouterine pouch
33 Vaginal part of cervix of uterus
34 Vagina

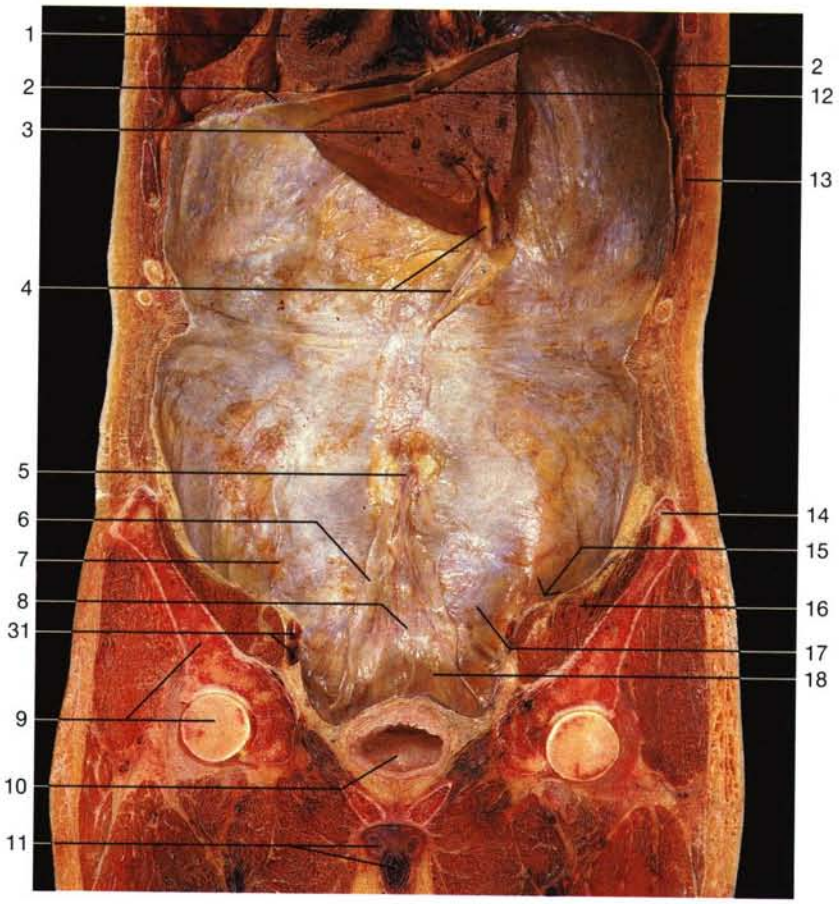

Anterior abdominal wall with pelvic cavity and thigh (frontal section, male) (internal aspect).

1 Left ventricle with pericardium
2 Diaphragm
3 Remnant of liver
4 Ligamentum teres (free margin of falciform ligament)
5 Site of umbilicus
6 **Medial umbilical fold** (containing the obliterated umbilical artery)
7 **Lateral umbilical fold** (containing inferior epigastric artery and vein)
8 **Median umbilical fold** (containing remnant of urachus)
9 Head of femur and pelvic bone
10 Urinary bladder
11 Root of penis
12 Falciform ligament of liver
13 Rib (divided)
14 Iliac crest (divided)
15 Site of **deep inguinal ring** and lateral inguinal fossa
16 Iliopsoas muscle (divided)
17 Medial inguinal fossa
18 Supravesical fossa
19 Posterior layer of rectus sheath
20 Transversus abdominis muscle
21 Umbilicus and arcuate line
22 **Inferior epigastric artery**
23 Femoral nerve
24 Iliopsoas muscle
25 Remnant of umbilical artery
26 Femoral artery and vein
27 Tendinous intersection of rectus abdominis muscle
28 Rectus abdominis muscle
29 Interfoveolar ligament
30 Pubic symphysis (divided)
31 External iliac artery and vein

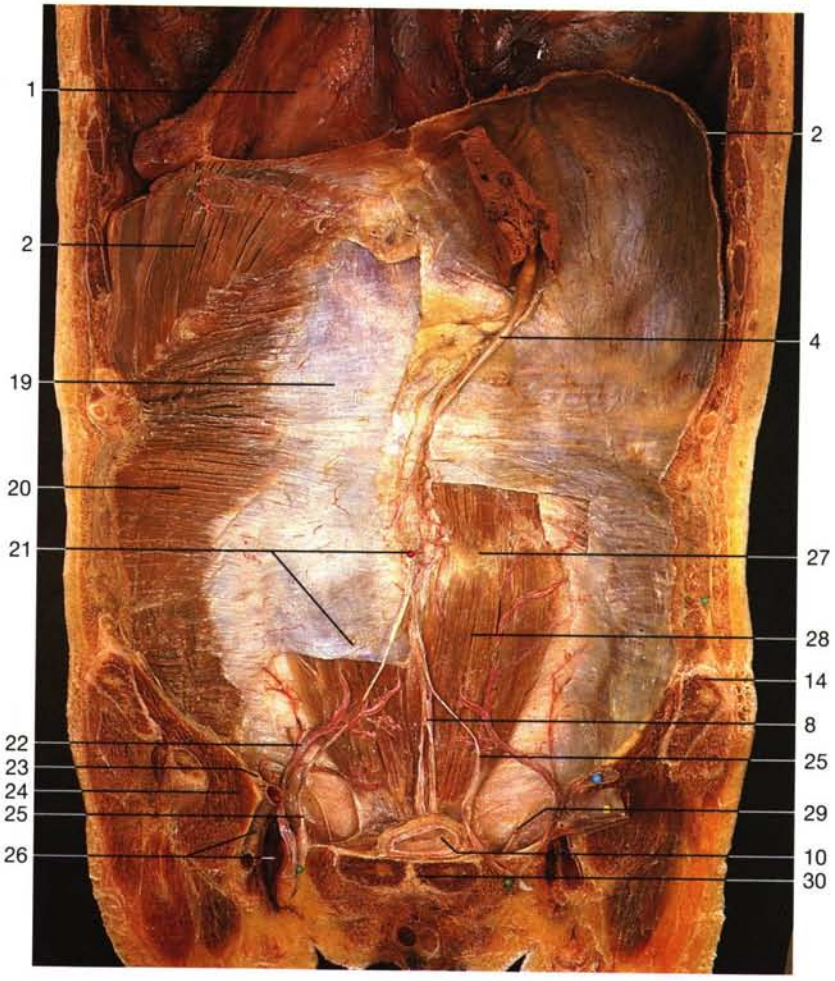

Anterior abdominal wall (male) (internal aspect). The peritoneum and parts of the posterior layer of rectus sheath have been removed. Dissection of inferior epigastric arteries and veins.

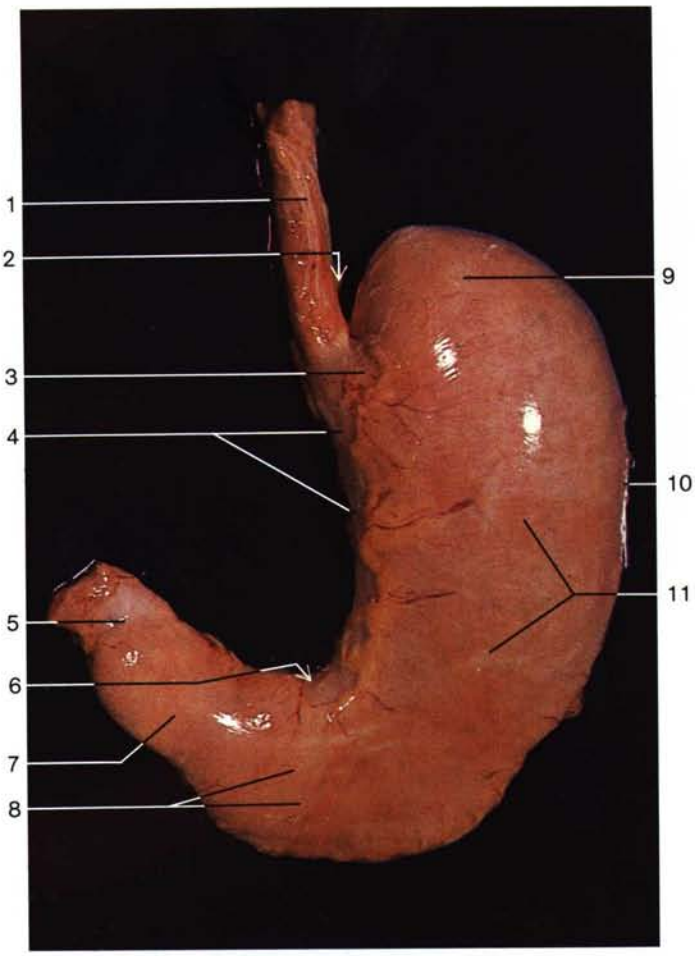

Stomach (ventral aspect).

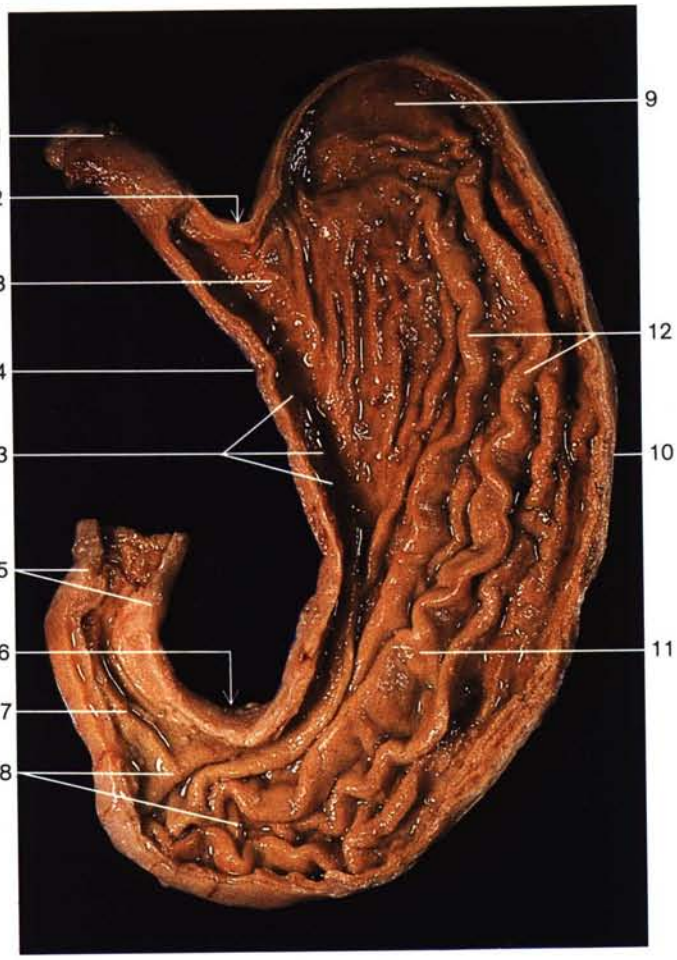

Mucosa of posterior wall of stomach (ventral aspect).

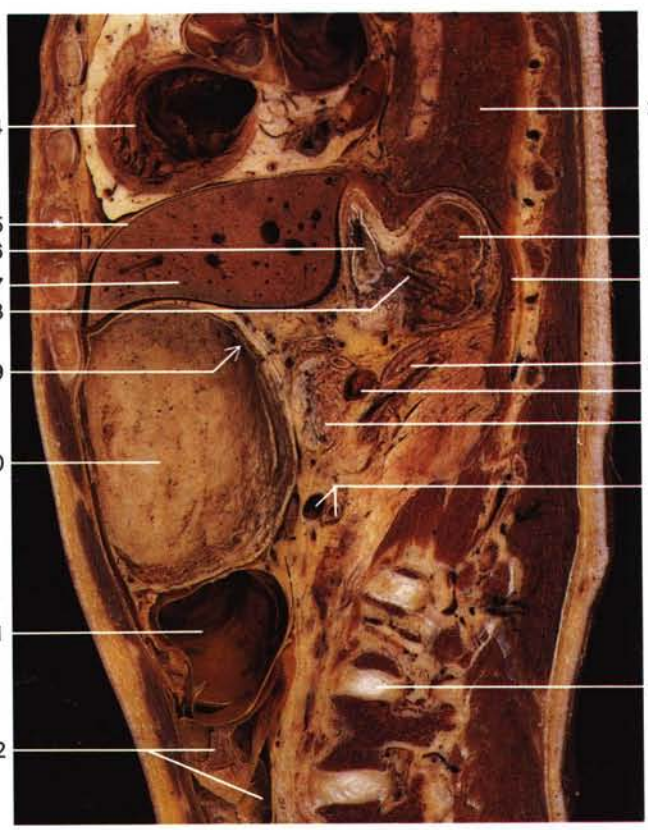

Parasagittal section through upper part of left abdominal cavity 3.5 cm lateral to median plane.

1 Esophagus
2 Cardiac notch
3 **Cardiac part of stomach**
4 Lesser curvature of stomach
5 **Pyloric sphincter**
6 Angular notch (incisura angularis)
7 **Pyloric canal**
8 Pyloric antrum
9 Fundus of stomach
10 Greater curvature of stomach
11 Body of stomach
12 Folds of mucous membrane (gastric rugae)
13 **Gastric canal**
14 Right ventricle of heart
15 Diaphragm (cut edge)
16 Abdominal portion of esophagus
17 **Liver**
18 Cardiac part of stomach (cut edge)
19 Position of pyloric canal
20 Body of stomach
21 Transverse colon
22 Small intestine
23 Lung (cut edge)
24 Fundus of stomach (section)
25 Lumbar portion of diaphragm (cut edge)
26 Suprarenal gland
27 Splenic vein
28 Pancreas
29 Superior mesenteric artery and vein
30 Intervertebral disc

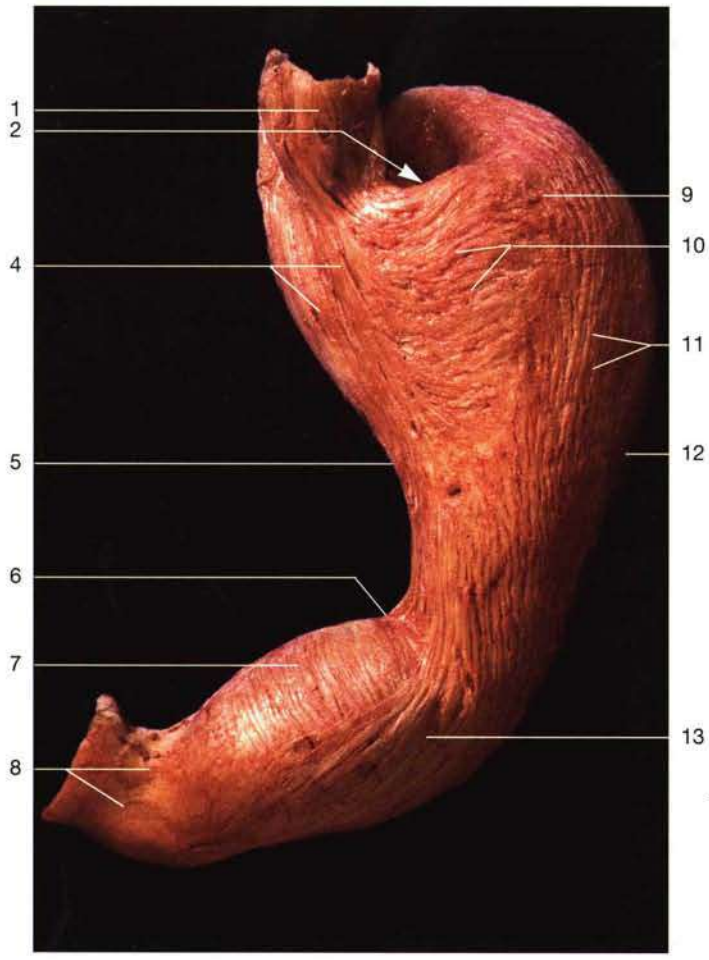

Muscular coat of stomach, outer layer (ventral aspect).

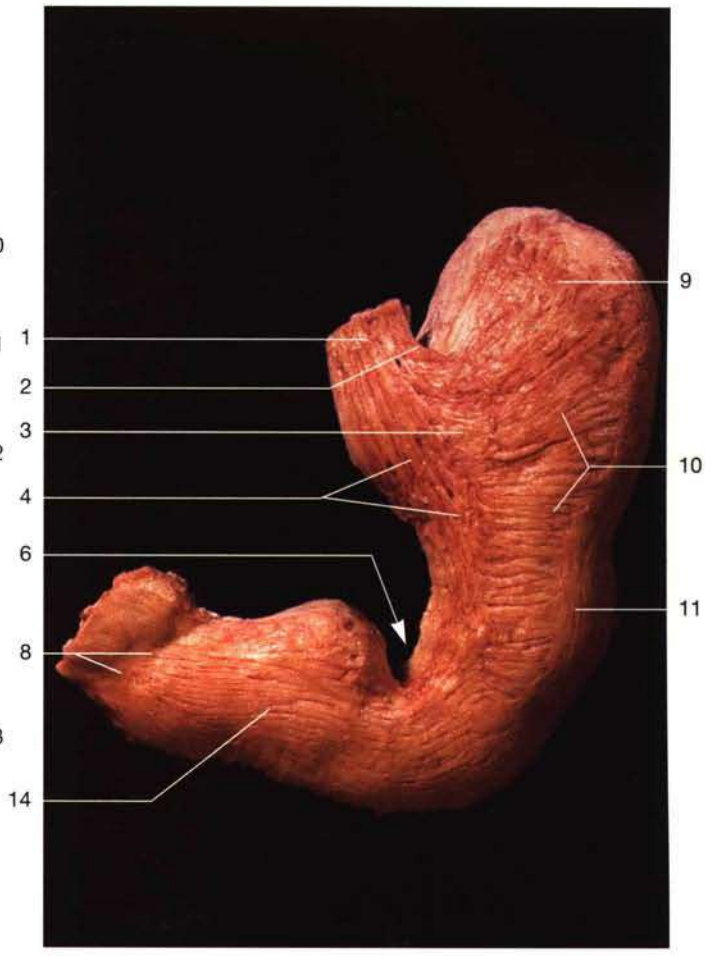

Muscular coat of stomach, middle layer (ventral aspect).

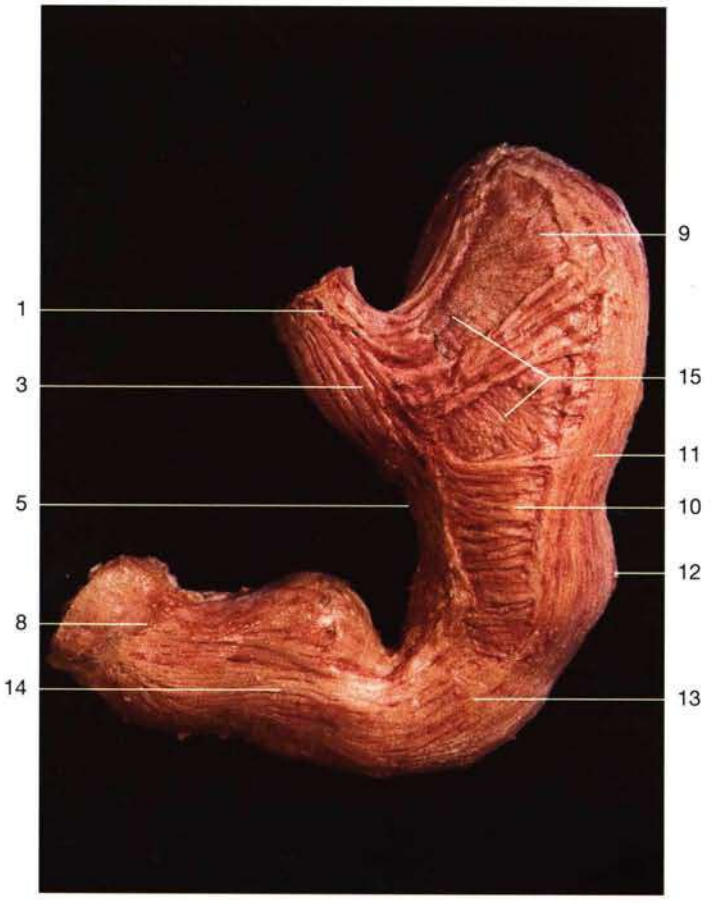

Muscular coat of stomach, inner layer (ventral aspect).

1 Esophagus (abdominal part)
2 Cardiac notch
3 Cardiac part of stomach
4 Longitudinal muscle layer at lesser curvature of stomach
5 Lesser curvature
6 Incisura angularis
7 Circular muscle layer of pyloric part of stomach
8 Pyloric sphincter muscle
9 Fundus of stomach
10 Circular muscle layer of fundus of stomach
11 Longitudinal muscle layer of greater curvature of stomach
12 Greater curvature of stomach
13 Longitudinal muscle layer (transition from body to pyloric part of stomach)
14 Pyloric part of stomach
15 Oblique muscle fibers

Pancreas with adjacent duodenum, spleen, liver and related arteries. The pancreatic ducts have been partly dissected. The duodenal papillae are indicated by probes. Left half of liver has been removed.

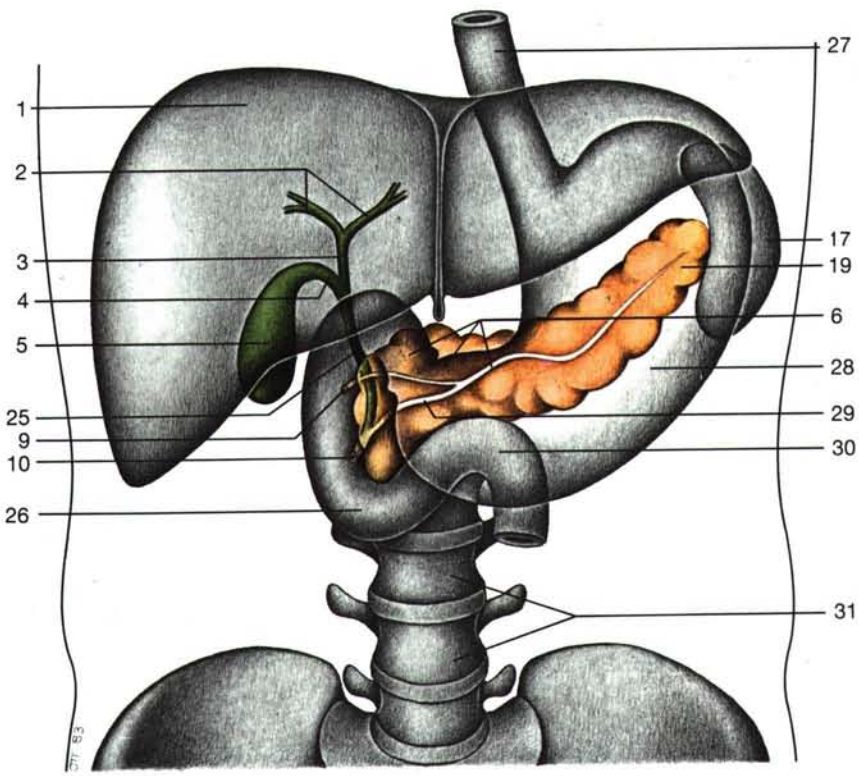

Pancreas, duodenum and extrahepatic bile ducts. (Schematic drawing.)

1 Liver
2 Left and right hepatic ducts
3 Common hepatic duct
4 **Cystic duct**
5 **Gall bladder**
6 **Pancreatic duct** and **head of pancreas**
7 Pylorus (cut)
8 Right gastro-omental (gastroepiploic) artery
9 Minor duodenal papilla (probe)
10 Major duodenal papilla (probe)
11 **Superior mesenteric artery and vein**
12 Inferior vena cava and iliocolic artery
13 **Portal vein,** lig. teres of liver and lig. venosum
14 Aorta
15 Common hepatic artery, left branch of hepatic artery proper, and right gastric artery
16 **Gastroduodenal artery** and right branch of hepatic artery proper
17 Spleen
18 Splenic artery and vein
19 **Tail of pancreas**
20 Left gastro-omental artery
21 Middle colic artery
22 Jejunum
23 Jejunal arteries
24 Aorta and inferior mesenteric artery
25 **Common bile duct**
26 Horizontal part of duodenum
27 Esophagus
28 Stomach
29 **Pancreatic duct**
30 Duodenojejunal flexure
31 Lumbal vertebrae

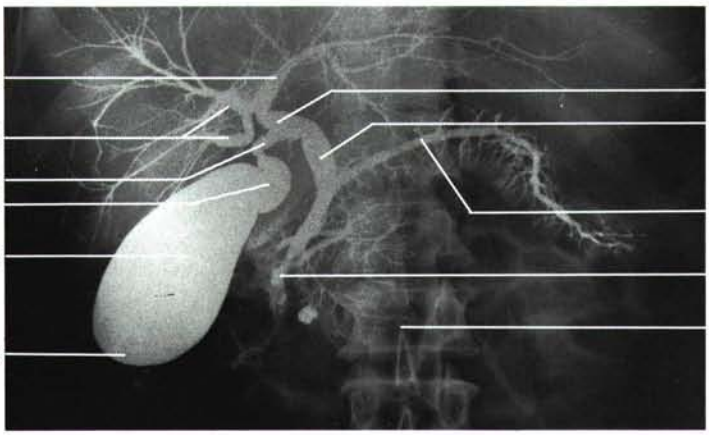

Radiograph of biliary ducts, gallbladder and pancreatic duct (anterior-posterior view).

Isolated gallbladder and cystic duct (anterior aspect).
The gallbladder has been opened to display the mucous membrane.

1 Left hepatic duct
2 Right hepatic duct
3 Cystic duct
4 Neck of gallbladder
5 Body of gallbladder
6 Fundus of gallbladder
7 Common hepatic duct
8 **Common bile duct**
9 **Pancreatic duct**
10 **Greater duodenal papilla**
11 Second lumbar vertebra
12 Folds of mucous membrane of gallbladder
13 Muscular coat of gallbladder
14 Neck of gallbladder (opened)
15 Cystic duct with spiral fold
16 **Lesser duodenal papilla**
17 **Accessory pancreatic duct**
18 Uncinate process
19 Plica circularis of duodenum (Kerckrings's fold)
20 Head of pancreas
21 Body of pancreas
22 Tail of pancreas
23 Descending part of duodenum
24 Incisure of pancreas

Pancreas with descending part of duodenum (posterior aspect). The duodenum was opened to display the duodenal papillae. Pancreatic duct has been dissected, the common bile duct has been divided. The sphincter of Oddi is shown.

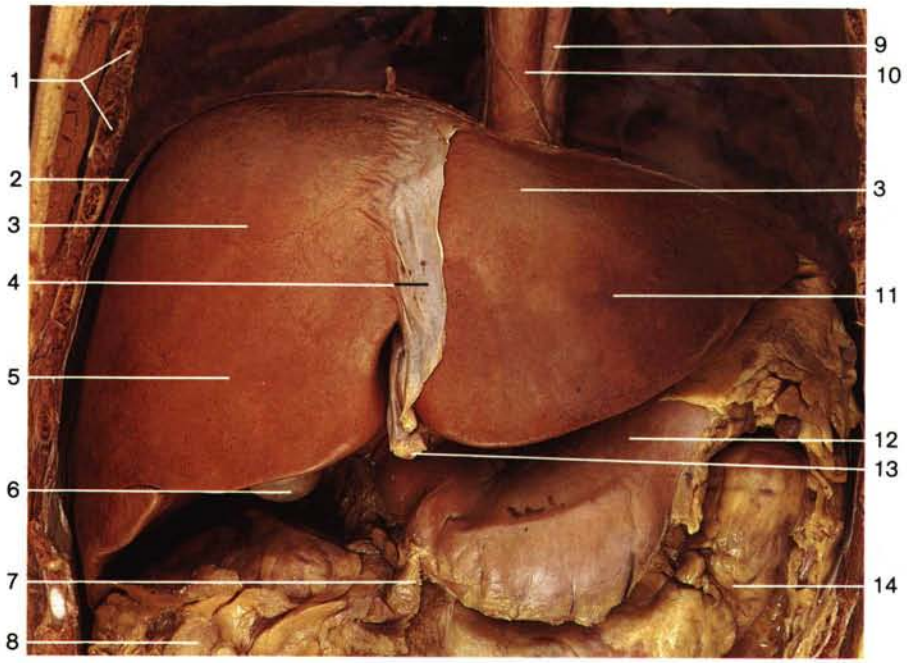

Liver in situ (ventral aspect). Part of the diaphragm has been removed.

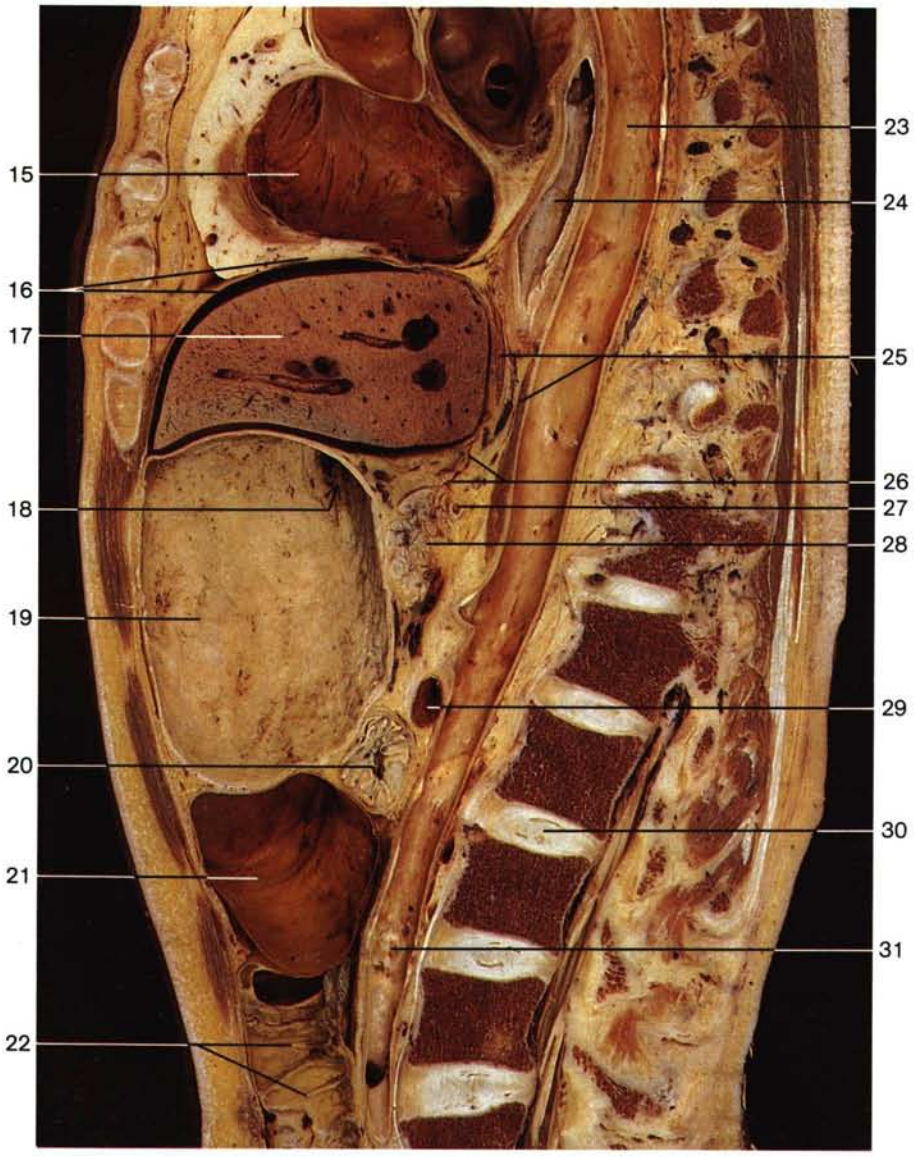

Liver in situ. Parasagittal section through the left side of the abdomen 2 cm lateral to median plane.

1 Ribs (cut edges)
2 Diaphragm
3 Diaphragmatic surface of liver
4 **Falciform ligament of liver**
5 Right lobe of liver
6 Fundus of gallbladder
7 Gastrocolic ligament
8 Greater omentum
9 Aorta
10 Esophagus
11 Left lobe of liver
12 Stomach
13 **Ligamentum teres**
14 Transverse colon
15 Right atrium of heart
16 Central tendon and sternal portion of diaphragm
17 Liver (cut edge)
18 Entrance to duodenum (pylorus)
19 Stomach
20 Duodenum
21 Transverse colon (divided, dilated)
22 Small intestine
23 Thoracic aorta (longitudinally divided)
24 Esophagus (longitudinally divided)
25 Esophageal hiatus of diaphragm
26 **Omental bursa** (lesser sac)
27 Splenic artery
28 Pancreas
29 Left renal vein
30 Intervertebral disc
31 Abdominal aorta (longitudinally divided)

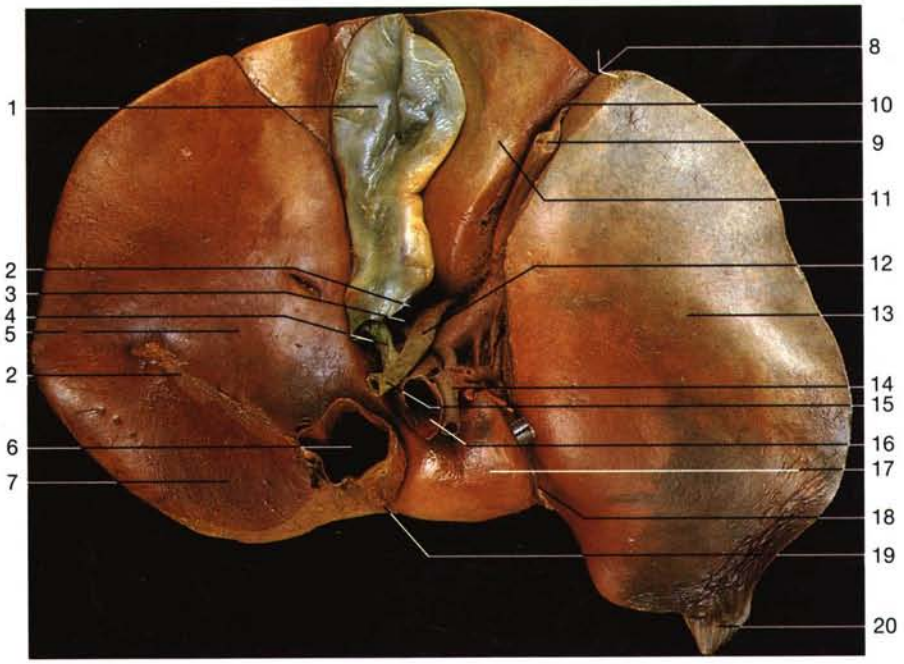

Liver (inferior aspect). Dissection of porta hepatis. Gallbladder partly collapsed. Ventral margin of liver above.

1 Fundus of **gallbladder**
2 Peritoneum (cut edges)
3 Cystic artery
4 **Cystic duct**
5 **Right lobe of liver**
6 Inferior vena cava
7 Bare area of liver
8 Notch for ligamentum teres and falciform ligament
9 Ligamentum teres
10 Falciform ligament of liver
11 **Quadrate lobe of liver**
12 Common hepatic duct
13 **Left lobe of liver**
14 Hepatic artery proper
15 **Common bile duct**
16 **Portal vein**
} Portal triad
17 **Caudate lobe of liver**
18 Ligamentum venosum
19 Ligament of inferior vena cava
20 Appendix fibrosa (left triangular ligament)
21 Coronary ligament of liver
22 Hepatic veins
23 Porta hepatis

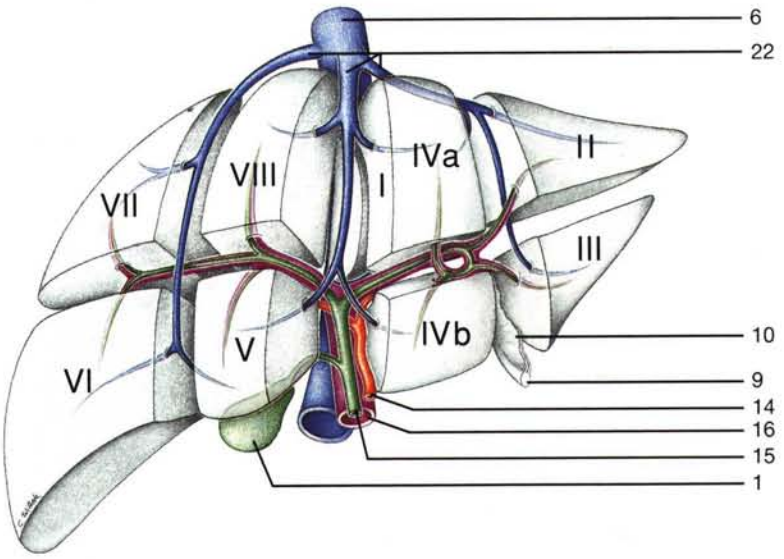

Segmentation of the liver (anterior aspect). Liver segments indicated by roman numerals.

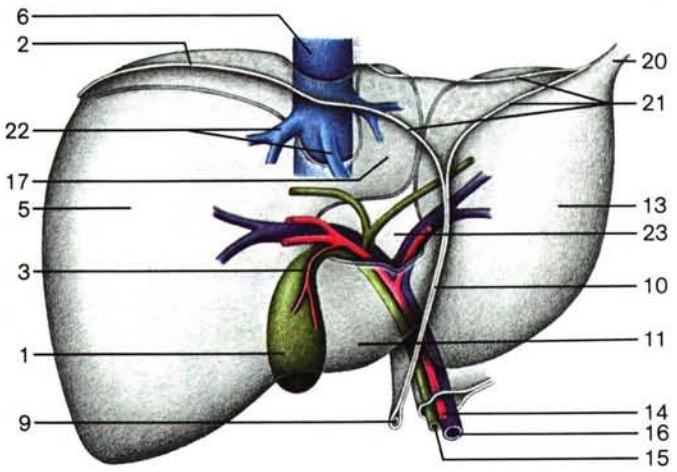

Liver (ventral aspect) (transparent drawing illustrating margins of peritoneal folds).

It should be noted that the anatomical left and right lobes of the liver do not reflect the internal distribution of the hepatic artery, portal vein, and biliary ducts. Using these structures as criteria the left lobe includes both the caudate and quadrate lobes and thus the line dividing the liver into left and right functional lobes passes through the gallbladder and inferior vena cava. The three main hepatic veins drain segments of the liver which have no visible external markings.

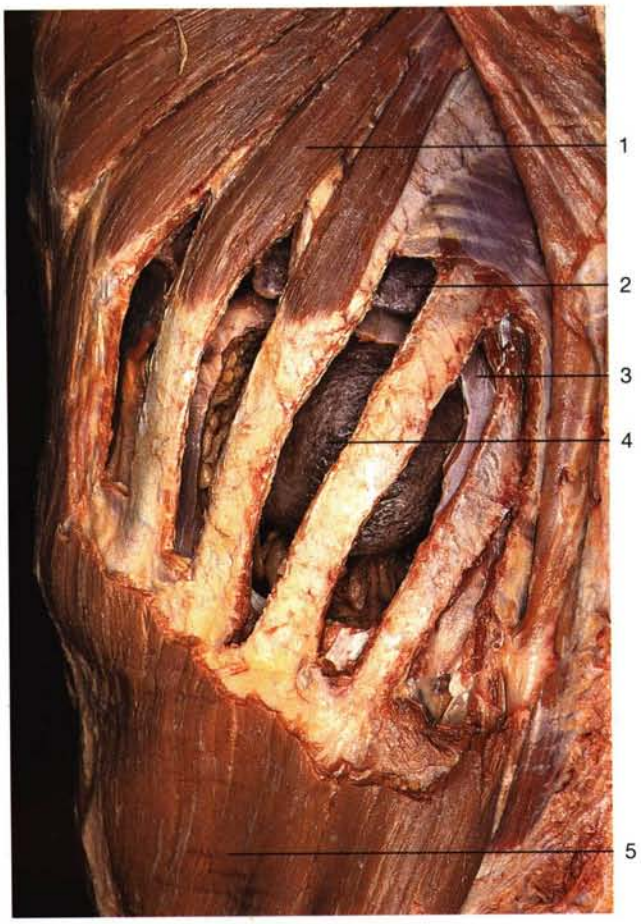

Location of the spleen in situ (left-lateral aspect). Intercostal spaces and diaphragm have been fenestrated.

1 Serratus anterior muscle
2 Left lung
3 Diaphragm
4 Spleen
5 External abdominal oblique muscle
6 Gastrosplenic ligament
7 **Splenic artery**
8 Pancreas tail
9 Superior margin of spleen
10 Anterior border of spleen
11 Inferior vena cava
12 **Hepatic veins**
13 Liver
14 **Portal vein**
15 Paraumbilical veins within falciform ligament
16 **Superior mesenteric vein**
17 Middle colic vein
18 Right colic vein
19 Ascending colon
20 Ileocolic vein
21 Vermiform appendix and appendicular vein
22 Spleen
23 Gastric and esophageal veins
24 **Splenic vein**
25 **Inferior mesenteric vein**
26 Right gastroepiploic vein
27 Ileal veins
28 Descending colon
29 Ileum
30 Sigmoid veins
31 Superior rectal veins

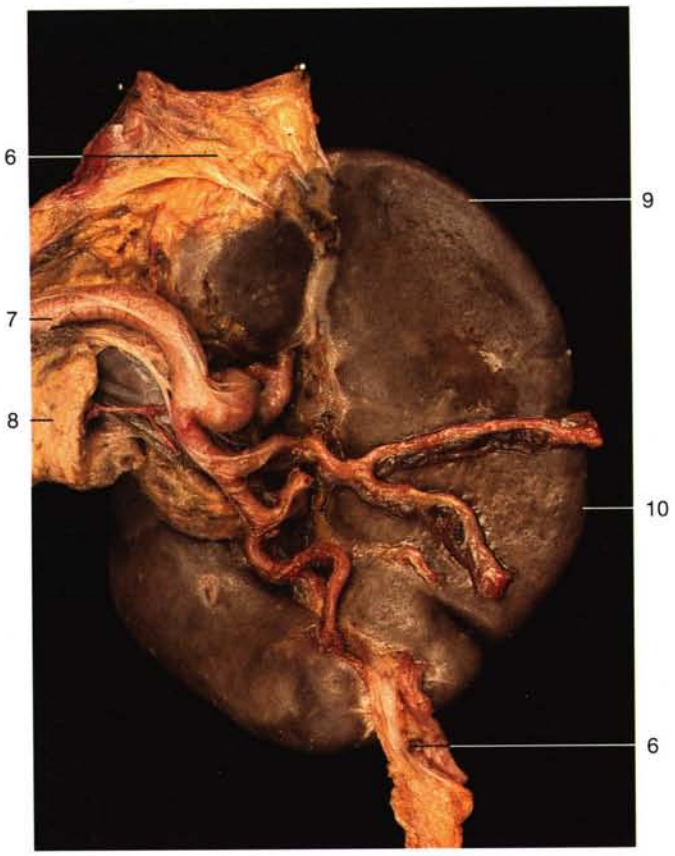

Spleen (visceral surface), hilum of spleen with vessels, nerves and ligaments.

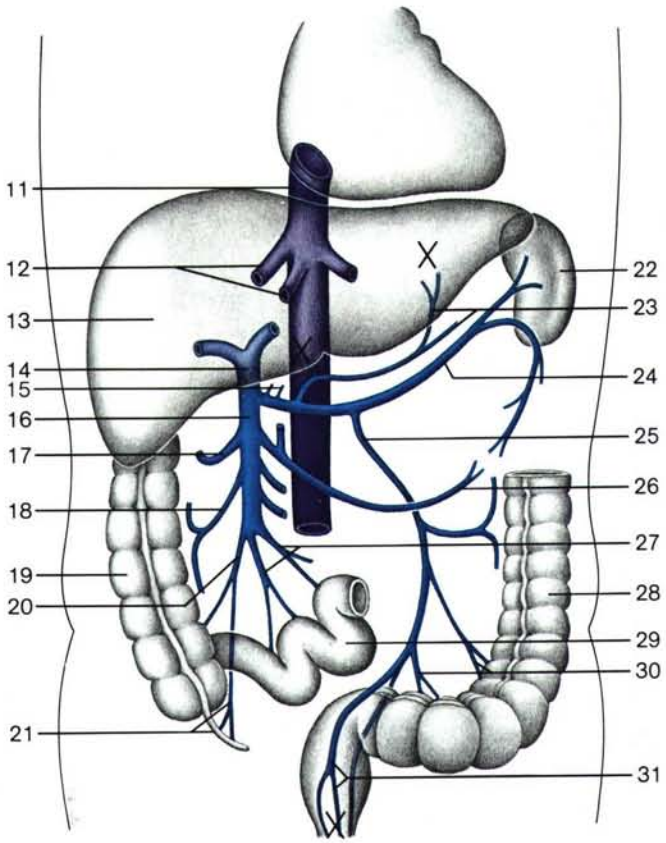

Main tributaries of portal vein. (Schematic drawing of portal circulation.) X = sites of portocaval anastomoses.

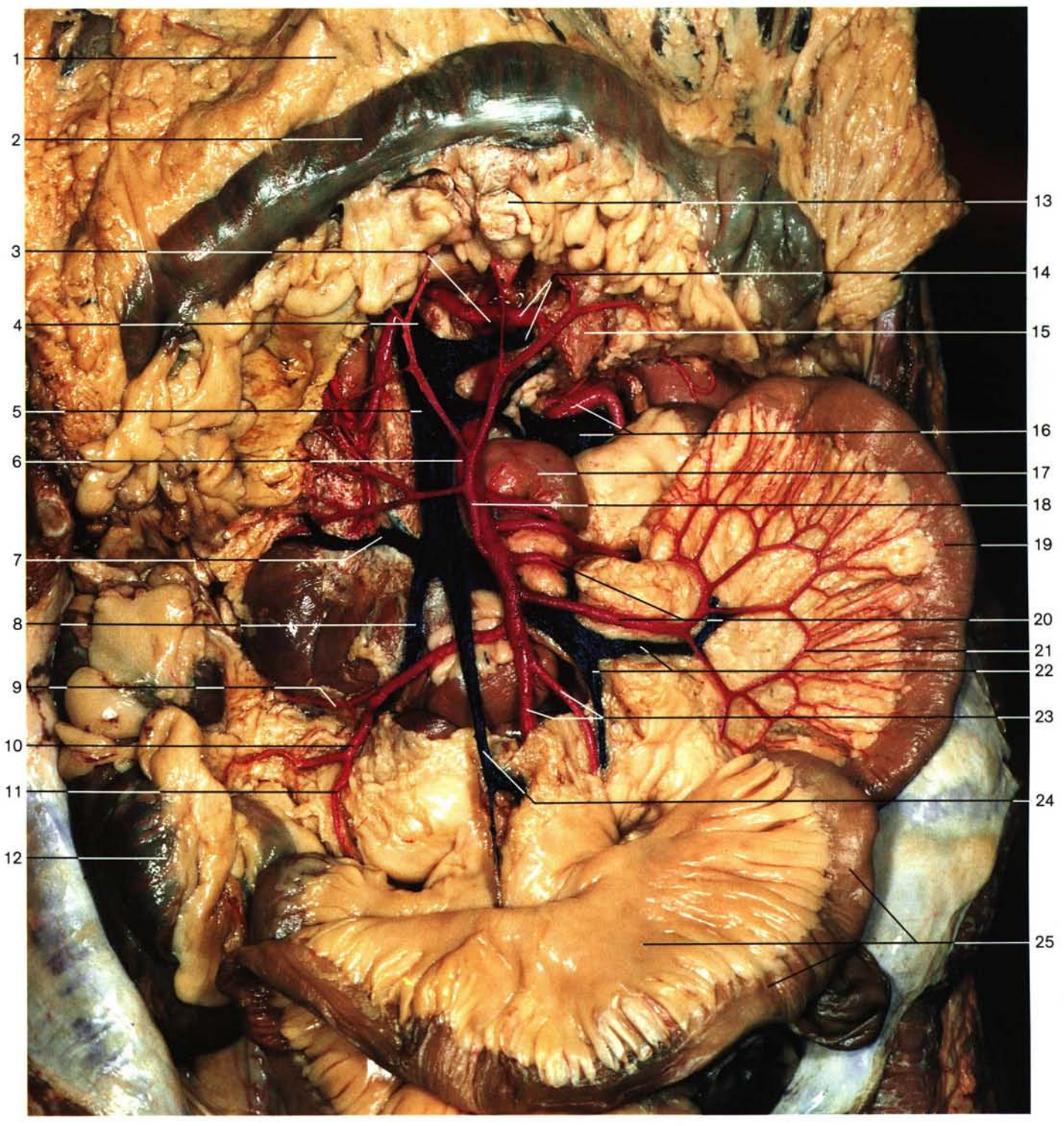

Tributaries of portal vein and branches of superior mesenteric artery (injected with colored solutions). Blue = veins; red = arteries. One layer of peritoneum has been removed to display the vascular arcades of the intestine. Part of the head of the pancreas and the mesocolon have also been removed to show the deeper vessels.

1 Greater omentum (reflected)
2 Transverse colon (raised cranially)
3 Celiac trunk
4 **Portal vein**
5 Superior mesenteric vein
6 **Superior mesenteric artery**
7 Right colic vein
8 Iliocolic vein
9 Right colic artery
10 Iliocolic artery
11 Appendicular artery
12 Cecum
13 Transverse mesocolon
14 **Splenic artery and vein**
15 Pancreas (divided)
16 Renal artery and vein
17 Duodenojejunal flexure
18 Middle colic artery
19 Jejunum
20 Jejunal arteries
21 Arterial arcades to intestine
22 Jejunal veins
23 Ileal arteries
24 Ileal vein
25 Ileum with mesentery

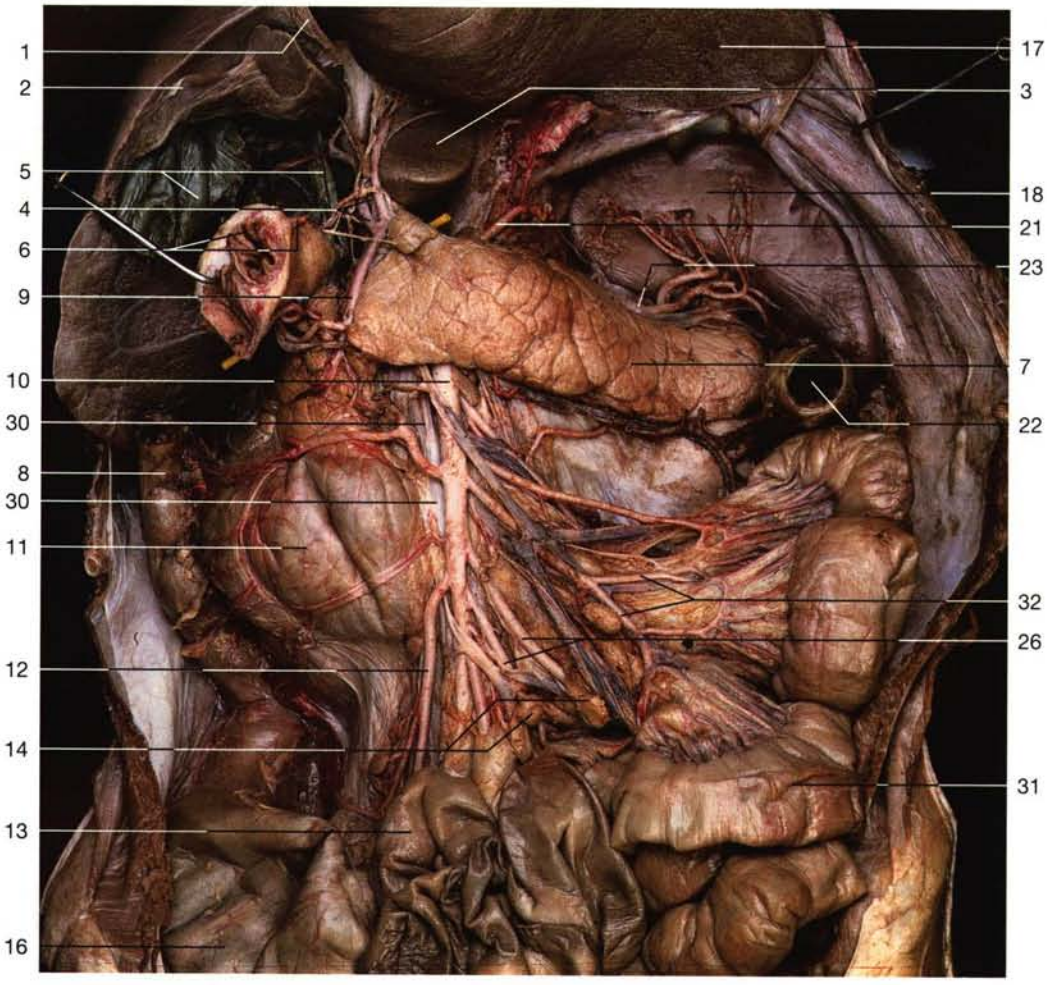

Superior mesenteric artery in relation to pancreas and duodenum. Stomach and transverse colon has been removed and the liver elevated. Note the location of the spleen. A yellow probe is inserted through the omental foramen.

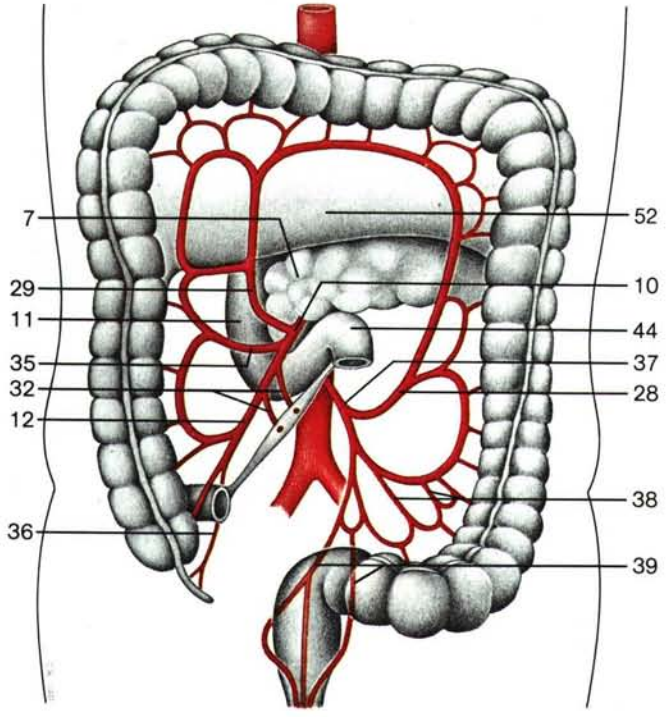

Main branches of superior and inferior mesenteric arteries. (Schematic drawing.)

1 Ligamentum teres
2 Liver
3 Caudate lobe of liver
4 Proper hepatic artery and portal vein
5 Gallbladder and common bile duct
6 Right gastric artery and pylorus
7 **Pancreas**
8 Ascending colon
9 Gastroduodenal artery
10 **Superior mesenteric artery**
11 Duodenum
12 **Iliocolic artery**
13 Terminal part of ileum
14 Mesenteric lymph nodes
15 Right common iliac artery
16 Cecum
17 Left lobe of liver
18 **Spleen**
19 Left gastro-omental (gastroepiploic) artery
20 Stomach
21 Left gastric artery
22 Left colic flexure (cut)
23 **Splenic artery**
24 Right gastro-omental (gastroepiploic) artery
25 Renal artery
26 **Ileal arteries**
27 Left kidney
28 Left colic artery
29 Middle colic artery
30 **Superior mesenteric vein**
31 Jejunum
32 **Jejunal arteries**
33 Inferior vena cava
34 Sigmoid colon
35 Right colic artery
36 Appendicular artery
37 **Inferior mesenteric artery**
38 Sigmoid arteries
39 Superior rectal artery
40 Fundus of gallbladder
41 Common bile duct
42 Portal vein
43 Beginning of jejunum
44 Duodenojejunal flexure
45 Ureter
46 Splenic artery
47 **Celiac trunk**
48 Renal vein
49 Abdominal aorta
50 Descending colon
51 Left common iliac artery
52 Transverse mesocolon

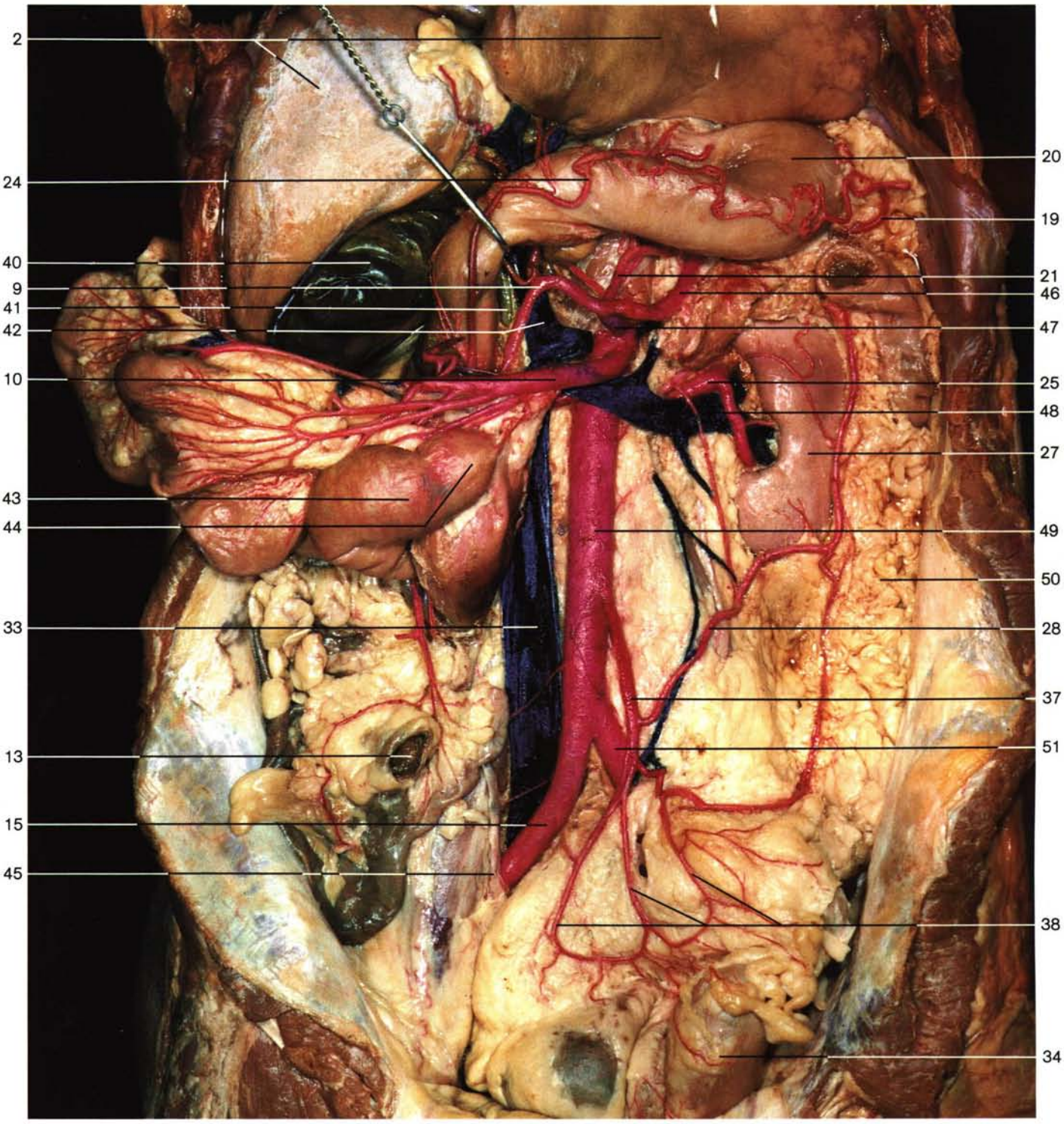

Vessels of abdominal organs; dissection of inferior mesenteric artery and celiac trunk (injected with colored solutions). Blue = veins; red = arteries. The small intestine including the duodenojejunal flexure has been reflected laterally and the stomach and liver have been raised. The peritoneum of the posterior abdominal wall has been removed to display the inferior mesenteric artery and its branches to the colon.

The **superior mesenteric artery** arises at the level of L_1 vertebra branching of the abdominal aorta. It supplies jejunum and ileum as the ascending and transverse colon. At the left colic flexure branches of the middle colic artery (from the superior mesenteric artery) and of the left colic artery (from the inferior mesenteric artery) anastomose (so-called **Riolani's anastomosis**). The arteries unite to form loops or arches called arterial arcades from which straight vessels (vasa recta) arise. The vasa recta do not anastomose within the mesentery. Thus occlusion of these vessels may lead to necrosis of the segment of bowel concerned and ileus. Within the wall of the intestine there are many anastomoses of blood vessels so that blockage of a single vessel is usually not dangerous. The arteries are paralleled by veins which convey blood to the portal vein. Closely related to the venous drainage are the lymph vessels which pass the mesenteric lymph nodes and finally enter the cisterna chyli.

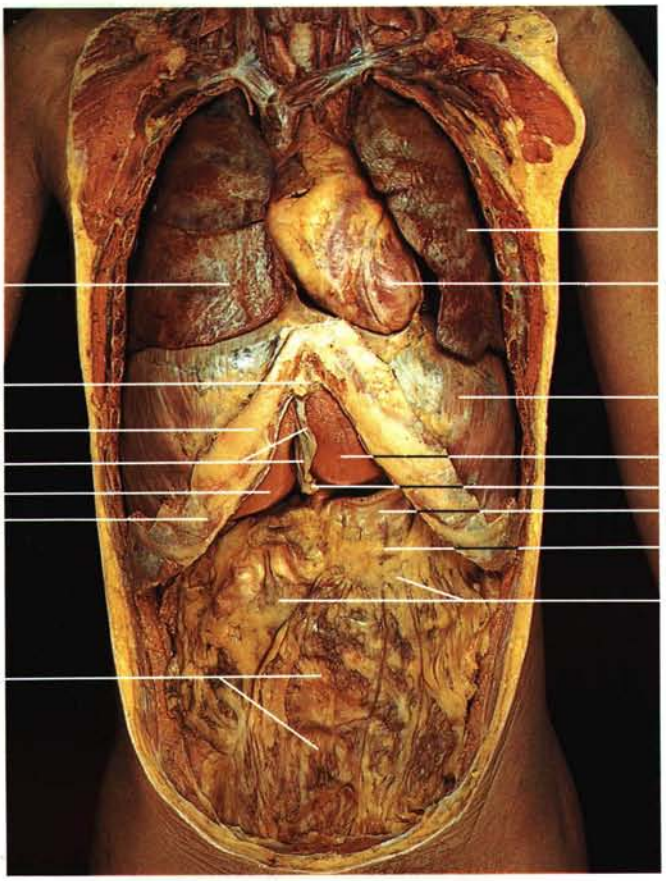

1 Middle lobe of right lung
2 Xiphoid process
3 Costal margin
4 Falciform ligament of liver
5 Quadrate lobe of liver
6 Greater omentum
7 Upper lobe of left lung
8 Heart
9 Diaphragm
10 Left lobe of liver
11 Ligamentum teres
12 Stomach
13 Gastrocolic ligament
14 Transverse colon
15 Taenia coli
16 Appendices epiploicae
17 Cecum
18 Taenia coli
19 Ileum
20 Transverse mesocolon
21 Jejunum
22 Sigmoid colon
23 Position of root of mesentery
24 Vermiform appendix
25 Duodenojejunal flexure
26 Mesentery

Abdominal organs. The anterior thoracic and abdominal walls have been removed.

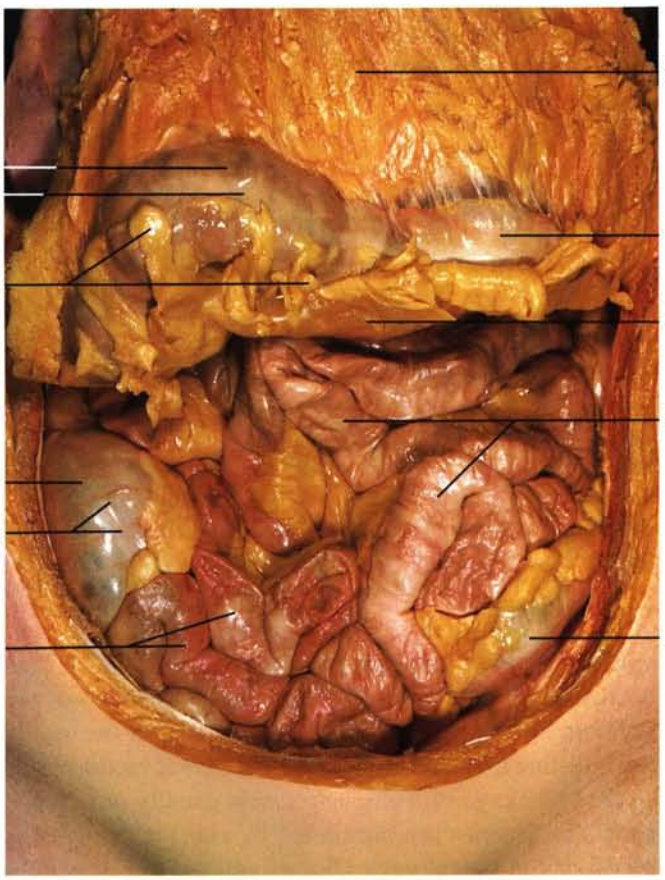

Abdominal organs. The greater omentum which is fixed to the transverse colon has been raised.

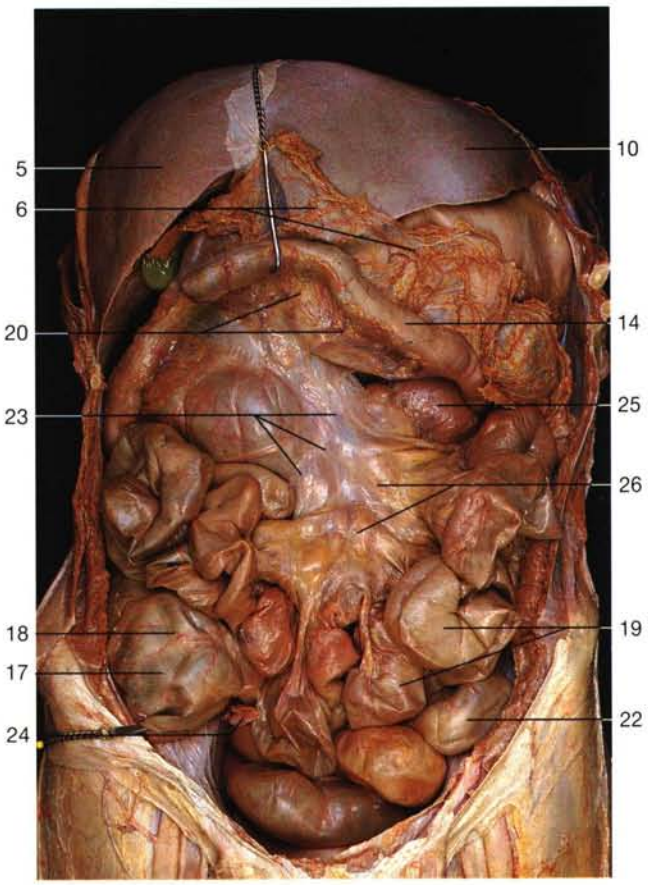

Abdominal organs (anterior aspect). The transverse colon has been reflected.

1 Diaphragm
2 Costal margin
3 Transverse colon
4 Ascending colon with haustra
5 Free taenia of cecum
6 **Ileum**
7 **Cecum**
8 **Falciform ligament of liver**
9 Liver
10 Stomach
11 Gastrocolic ligament
12 Jejunum
13 Sigmoid colon
14 **Vermiform appendix**
15 Terminal ileum
16 **Mesoappendix**
17 Mesentery

Abdominal organs in situ. The greater omentum has been removed.

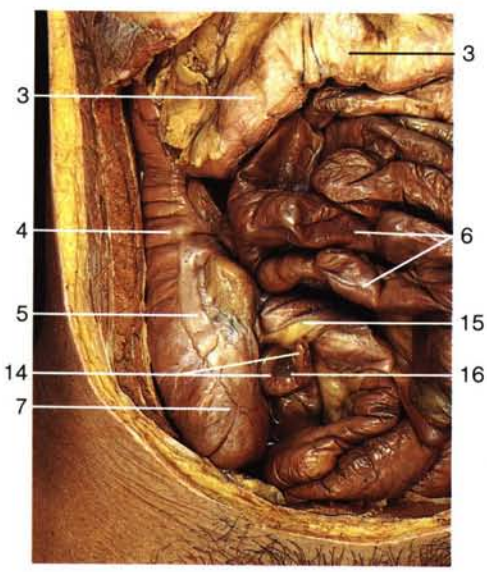

Ascending colon, cecum and vermiform appendix (detail of the preceding figure).

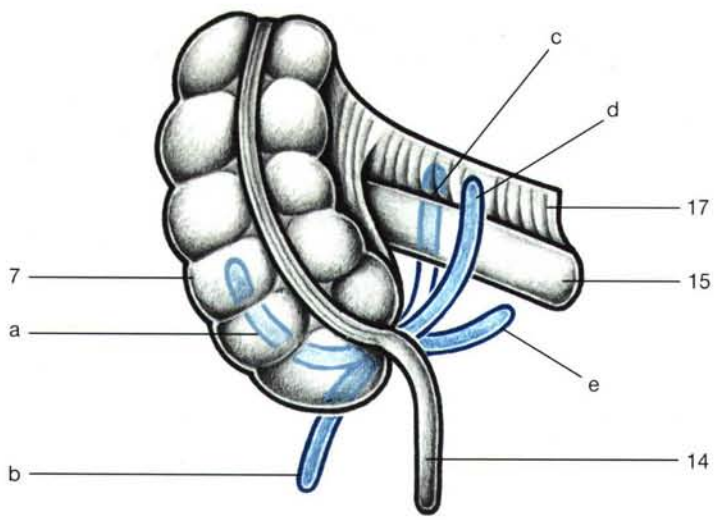

Variations in the position of the vermiform appendix. a = retrocecal; b = paracolic; c = retroileal; d = preileal; e = subcecal.

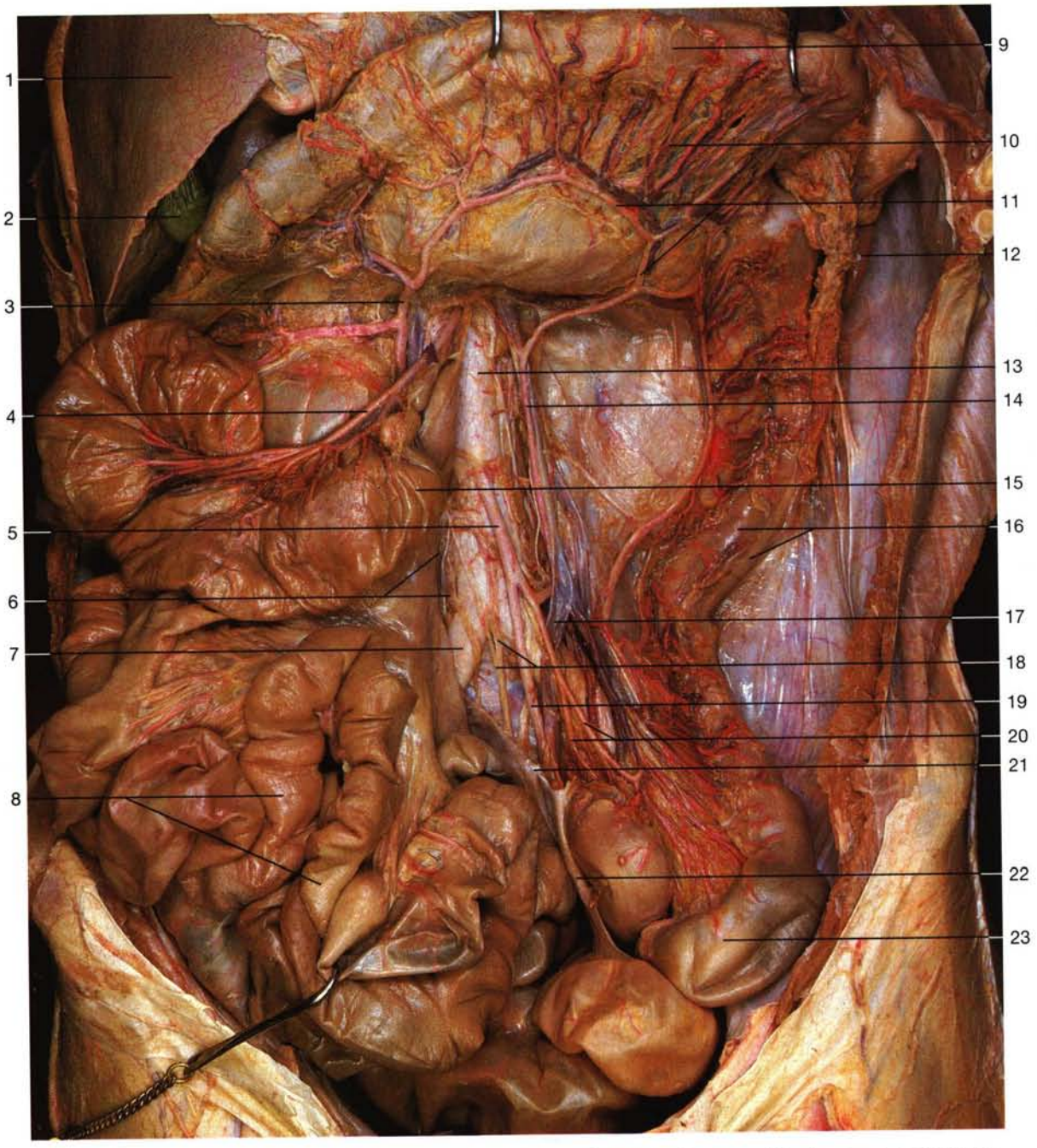

Abdominal organs. Dissection of inferior mesenteric artery and autonomic plexus. The transverse colon with mesocolon has been raised and the small intestine reflected.

1 Liver
2 Gallbladder
3 **Middle colic artery**
4 Jejunal artery
5 **Inferior mesenteric artery**
6 Sympathetic nerves and ganglia
7 Right common iliac artery
8 Small intestine (ileum)
9 Transverse colon (reflected)
10 Transverse mesocolon
11 Anastomosis between middle and left colic artery
12 Spleen
13 **Abdominal aorta**
14 **Left colic artery**
15 Duodenojejunal flexure
16 Descending colon (free taenia of colon)
17 Inferior mesenteric vein
18 Superior hypogastric plexus
19 **Superior rectal artery**
20 Sigmoid arteries
21 Peritoneum (cut edge)
22 Sigmoid mesocolon
23 Sigmoid colon

1 Liver
2 Middle colic artery
3 Horizontal part of duodenum (extended)
4 **Superior mesenteric artery and vein**
5 **Right colic artery**
6 **Iliocolic artery**
7 Ascending colon
8 Cecum
9 Greater omentum (reflected)
10 Transverse colon
11 Transverse mesocolon
12 Duodenojejunal flexure
13 **Jejunal arteries**
14 Jejunum
15 **Ileal arteries**
16 **Mesenteric lymph nodes and lymph vessels**
17 Ileum
18 Abdominal aorta
19 Inferior vena cava
20 Stomach
21 Spleen
22 Splenic artery
23 Head of pancreas
24 Superior mesenteric artery

Abdominal organs. Superior mesenteric artery. Mesenteric lymph nodes. Transverse colon reflected.

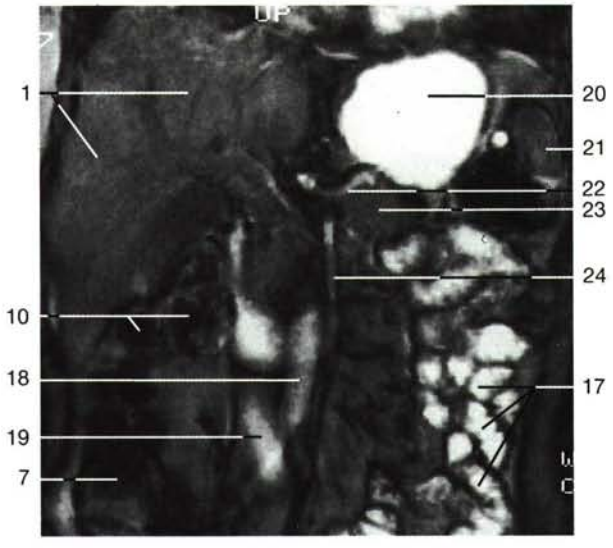

Frontal section through the abdominal cavity. (MR-Scan: the intestinal tract and vessels are filled with a paramagnetic substance [Gadolinium]; courtesy of Dr. W. Rödl, Erlangen.)

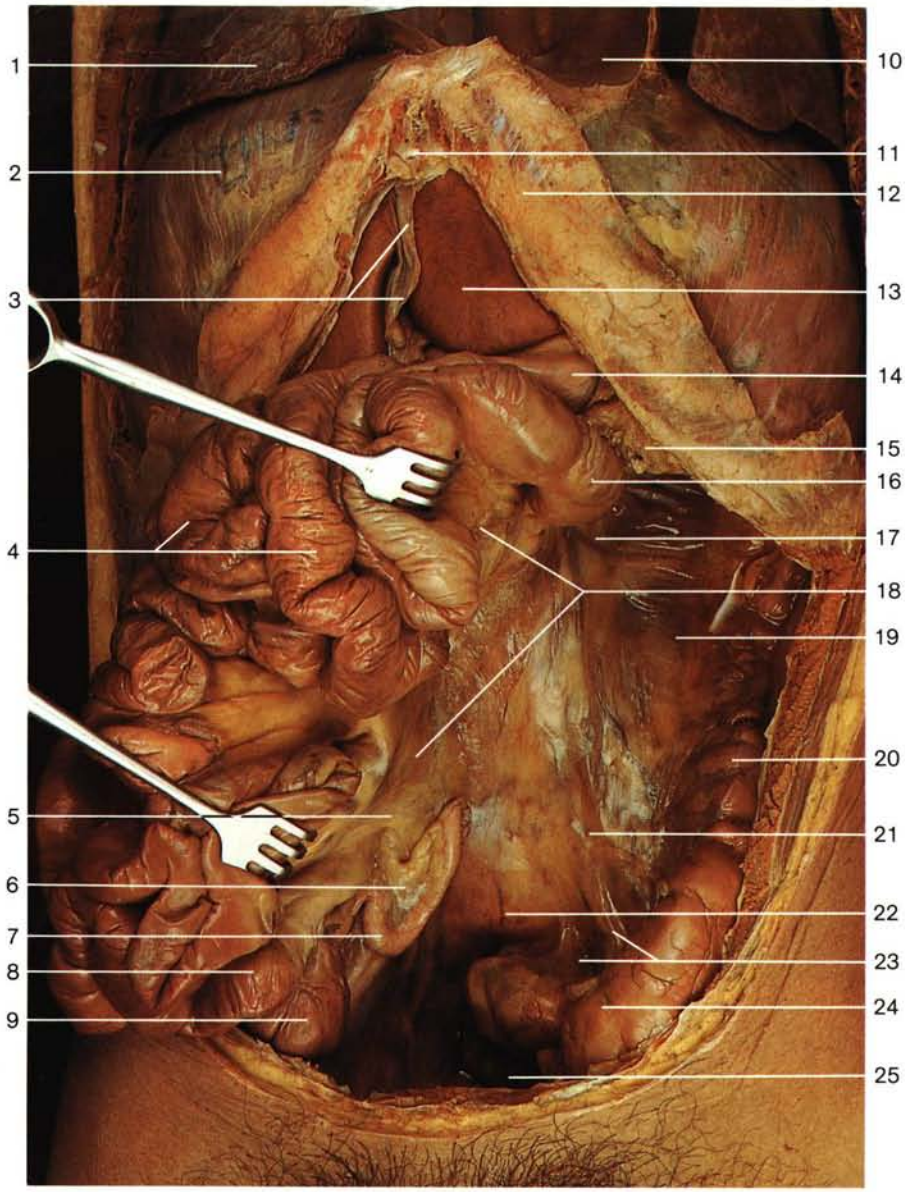

1 Lung
2 Diaphragm
3 Falciform ligament of liver
4 Jejunum
5 Ileocecal fold
6 Mesoappendix
7 Vermiform appendix
8 **Ileocecal junction**
9 **Cecum**
10 Pericardial sac
11 Xiphoid process
12 Costal margin
13 Liver
14 Stomach
15 Transverse colon
16 **Duodenojejunal flexure**
17 Inferior duodenal fold
18 Mesentery
19 Position of left kidney
20 Descending colon
21 Position of left common iliac artery
22 Sacral promontory
23 Sigmoid mesocolon
24 Sigmoid colon
25 Rectum
26 Beginning of jejunum
27 Peritoneum of posterior abdominal wall
28 Transverse mesocolon
29 Superior duodenal fold
30 Superior duodenal recess
31 Retroduodenal recess
32 Free taenia of ascending colon
33 **Ileocecal valve**
34 Frenulum of ileocecal valve
35 Orifice of vermiform appendix (probe)
36 **Ileocolic artery**
37 Vermiform appendix with appendicular artery
38 Ascending colon

Abdominal cavity. Mesenteries. The small intestine has been reflected laterally to demonstrate the mesentery.

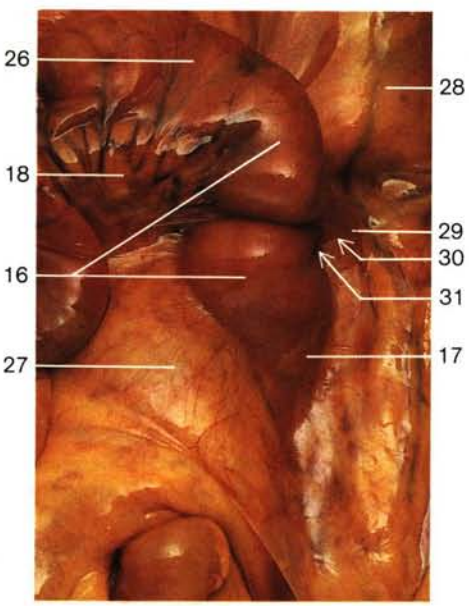

Duodenojejunal flexure (enlargement of preceding figure).

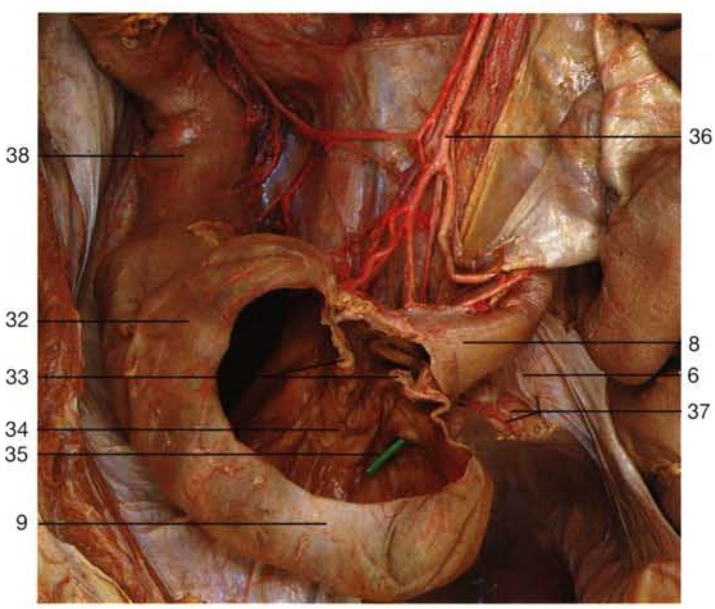

Ileocecal valve (ventral aspect). The cecum and terminal part of the ileum have been opened.

Upper abdominal organs. Thorax and anterior part of diaphragm have been removed and the liver raised to display the lesser omentum. A probe has been inserted into the epiploic foramen and lesser sac.

1 Falciform ligament and ligamentum teres
2 Liver
3 Gallbladder (fundus)
4 Hepatoduodenal ligament
5 **Epiploic foramen** (probe)
6 Pylorus
7 Descending part of duodenum
8 Right colic flexure
9 **Gastrocolic ligament**
10 Caudate lobe of liver (behind lesser omentum)
11 **Lesser omentum**
12 Stomach
13 Lesser curvature of stomach
14 Superior part of duodenum
15 Diaphragm
16 Greater curvature of stomach with gastroepiploic vessels
17 Twelfth thoracic vertebra
18 Right kidney
19 Right suprarenal gland
20 Inferior vena cava
21 Falciform ligament of liver
22 Abdominal aorta
23 Spleen
24 Lienorenal ligament
25 Gastrosplenic ligament
26 Pancreas
27 Lesser sac

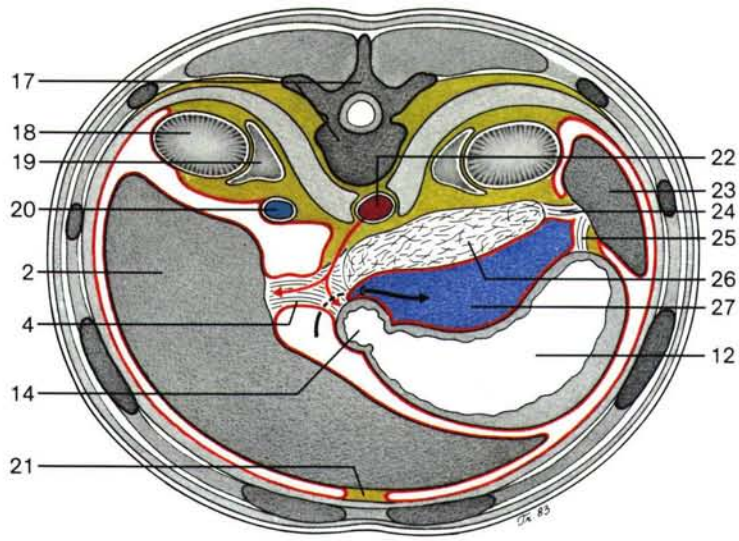

Horizontal section through omental bursa above the level of epiploic foramen (black arrow). Viewed from above. Red arrows: routes of the arterial branches of celiac trunk to liver, stomach, duodenum, and pancreas. (Schematic drawing.)

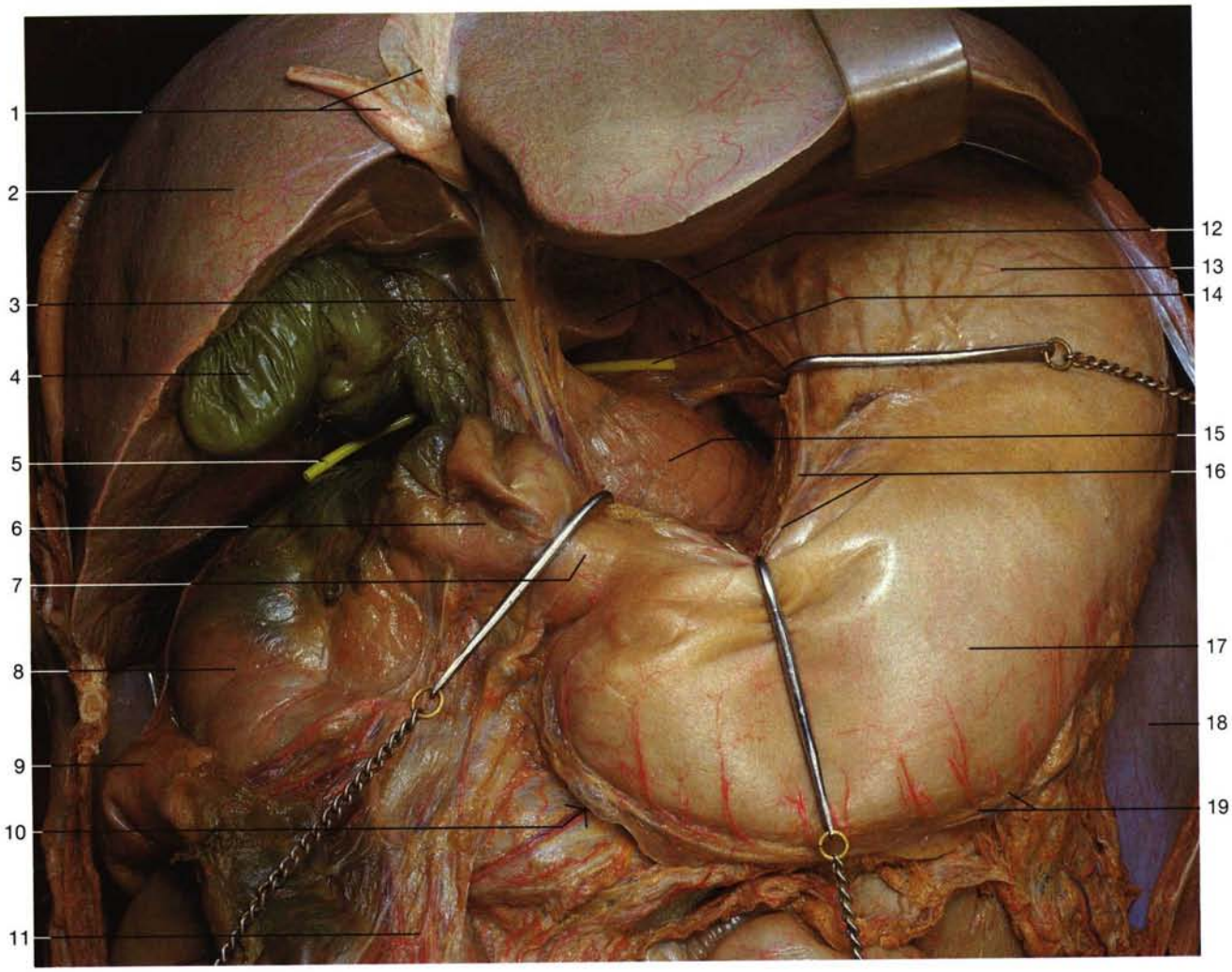

Upper abdominal organs, lesser sac, omental bursa (anterior aspect). Lesser omentum partly removed, liver and stomach slightly reflected.

1 Falciform ligament and ligamentum teres
2 Liver
3 **Hepatoduodenal ligament**
4 Gallbladder
5 Probe within the epiploic foramen
6 Superior part of duodenum
7 Pylorus
8 Descending part of duodenum
9 Right colic flexure
10 Gastrocolic ligament
11 **Greater omentum**
12 Caudate lobe of liver
13 Fundus of stomach
14 Probe at the level of the vestibule of lesser sac
15 **Head of pancreas**
16 Lesser curvature of stomach
17 **Body of stomach**
18 Diaphragm
19 Greater curvature with gastroepiploic vessels
20 Head of pancreas and gastropancreatic fold
21 **Spleen**
22 **Tail of pancreas**
23 Left colic flexure
24 Root of transverse mesocolon
25 Transverse mesocolon
26 Gastrocolic ligament (cut edge)
27 **Transverse colon**
28 Umbilicus
29 Small intestine
30 Lesser omentum
31 Lesser sac (omental bursa)
32 Duodenum
33 Mesentery
34 Sigmoid colon

Upper abdominal organs, lesser sac, omental bursa (anterior aspect). The gastrocolic ligament has been divided and the whole stomach raised to display the posterior wall of the lesser sac.

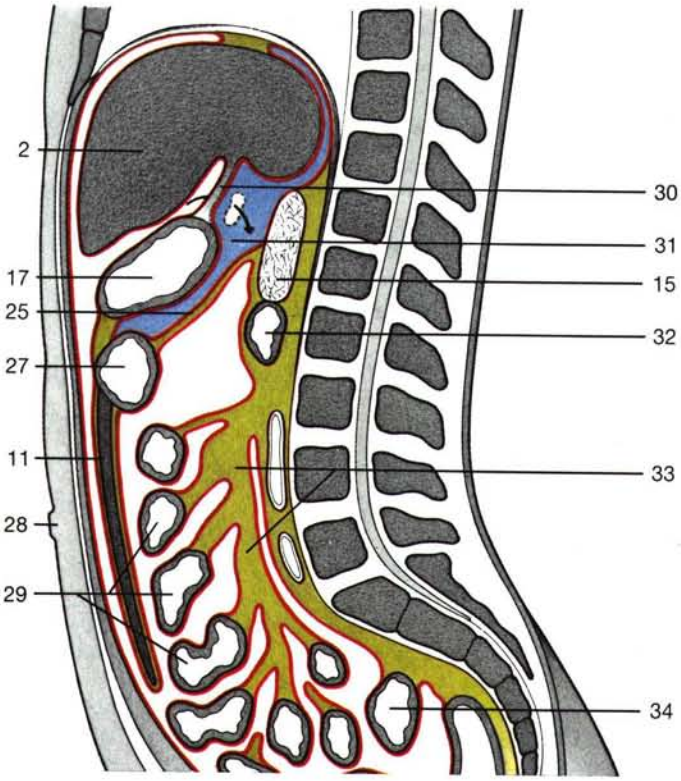

Midsagittal section through abdominal cavity demonstrating the site of lesser sac (blue). (Schematic drawing.) The epiploic foramen, entrance to the lesser sac, is indicated by an arrow. Red = peritoneum.

Arteries of upper abdominal organs; dissection of celiac trunk. The lesser omentum has been removed and the lesser curvature of the stomach reflected to display the branches of the celiac trunk.

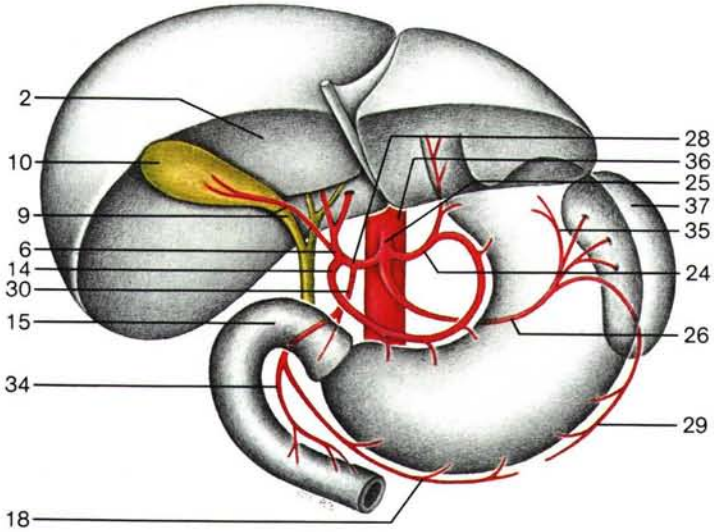

Branches of celiac trunk.
(Schematic drawing.)

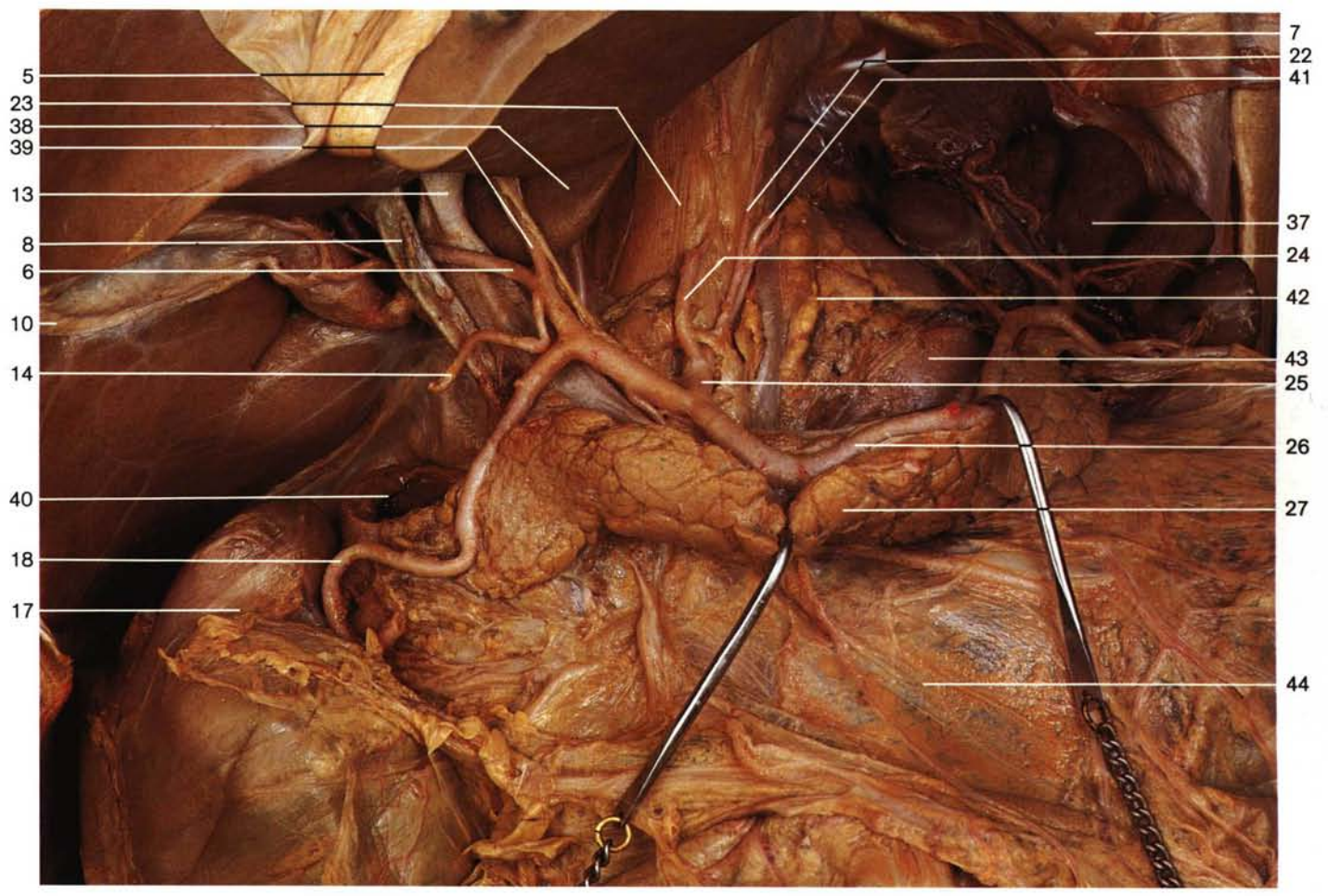

Branches of celiac trunk; blood supply of liver, pancreas and spleen. The stomach, superior part of duodenum and celiac ganglion have been removed to reveal the anterior aspect of the posterior wall of the lesser sac (omental bursa) and the vessels and ducts of the hepatoduodenal ligament. The pancreas has been slightly reflected anteriorly.

1 Lung
2 Liver (visceral surface)
3 Lymph node
4 Inferior vena cava
5 Ligamentum teres (reflected)
6 Right branch of hepatic artery proper
7 Diaphragm
8 **Common hepatic duct** (dilated)
9 Cystic duct and artery
10 Gallbladder
11 Probe in epiploic foramen
12 Right lobe of liver
13 **Portal vein**
14 **Right gastric artery**
15 Duodenum
16 Pylorus
17 Right colic flexure
18 **Right gastro-omental** (gastroepiploic) **artery**
19 Transverse colon
20 Abdominal part of esophagus (cardiac part of stomach)
21 Fundus of stomach
22 Esophageal branches of left gastric artery
23 Lumbar part of diaphragm
24 Left gastric artery
25 **Celiac trunk**
26 **Splenic artery**
27 Pancreas
28 **Common hepatic artery**
29 Left gastro-omental (gastroepiploic) artery
30 **Gastroduodenal artery**
31 Pyloric part of stomach
32 Greater curvature of stomach
33 Gastrocolic ligament
34 Superior pancreaticoduodenal artery
35 Short gastric arteries
36 Aorta
37 Spleen
38 Caudate lobe of liver
39 Left branch of hepatic artery proper
40 Descending part of duodenum (cut)
41 Left inferior phrenic artery
42 Suprarenal gland
43 Kidney
44 Transverse mesocolon

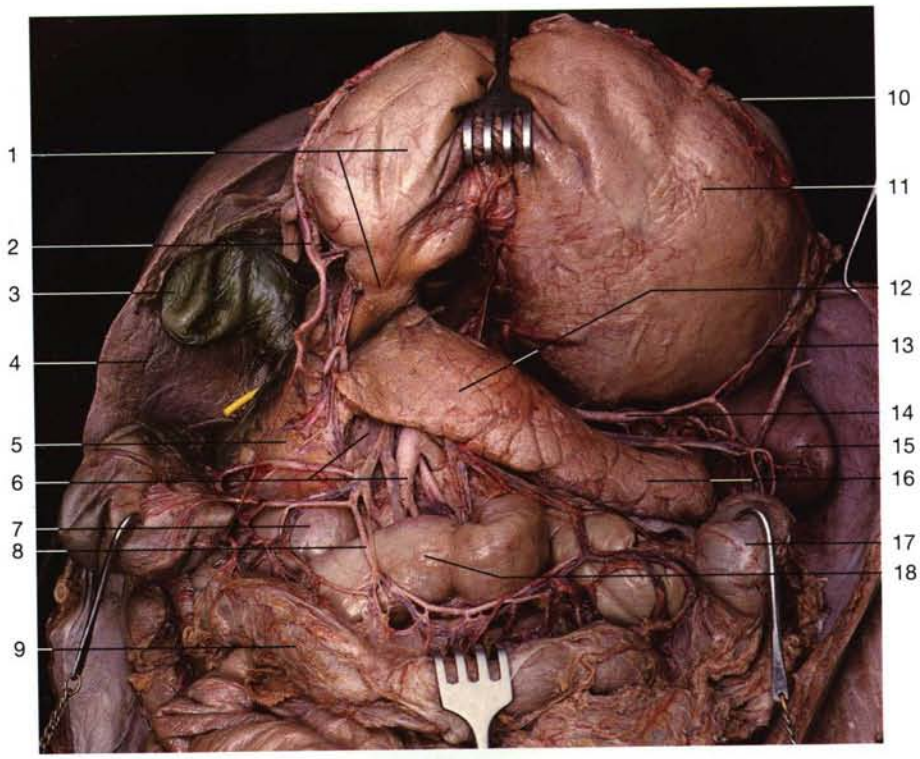

Pancreas and extrahepatic bile ducts in situ (anterior aspect). The gastrocolic ligament has been divided, the transverse colon and the stomach were replaced to display pancreas and superior mesenteric vessels.

1 Stomach (pyloric part) and pylorus
2 Right gastro-omental (gastroepiploic) artery
3 **Fundus of gall bladder**
4 Liver, right lobe
5 Head of pancreas
6 **Superior mesenteric artery** and **vein**
7 Duodenum
8 Middle colic artery
9 Transverse colon
10 Greater curvature of stomach (remnants of gastrocolic ligament)
11 **Body of stomach**
12 **Body of pancreas**
13 Left gastro-omental (gastroepiploic) artery
14 **Splenic artery**
15 **Spleen**
16 Tail of pancreas
17 Left colic flexure
18 Jejunum
19 Lung (inferior lobe of right lung)
20 **Common hepatic duct** and caudate lobe of liver
21 Portal vein and hepatic artery proper
22 Cystic duct and right gastric artery
23 **Common bile duct** and gastroduodenal artery
24 Pylorus (cut)
25 Major and minor duodenal papillas (probes)
26 Duodenum (inferior part)
27 Right kidney, ureter and ovarian artery
28 Iliocolic artery and inferior vena cava
29 Right and left vagal trunk
30 **Esophagus**
31 Diaphragm
32 Cut edge of esophagus, left gastric artery with esophageal branch
33 Celiac trunk
34 Common hepatic artery, main pancreatic duct and **pancreas**
35 Renal vein
36 Duodenojejunal flexure
37 Left kidney
38 Jejunal arteries
39 Abdominal aorta and left ureter
40 Cystic artery
41 Hepatic proper artery
42 **Celiac trunk**
43 Right gastric artery
44 **Common hepatic artery**
45 Gastroduodenal artery
46 **Superior mesenteric artery**
47 Superior posterior pancreaticoduodenal artery
48 Superior anterior pancreaticoduodenal artery
49 Short gastric arteries
50 Left gastric artery
51 Posterior pancreatic branch of splenic artery
52 Inferior pancreaticoduodenal artery

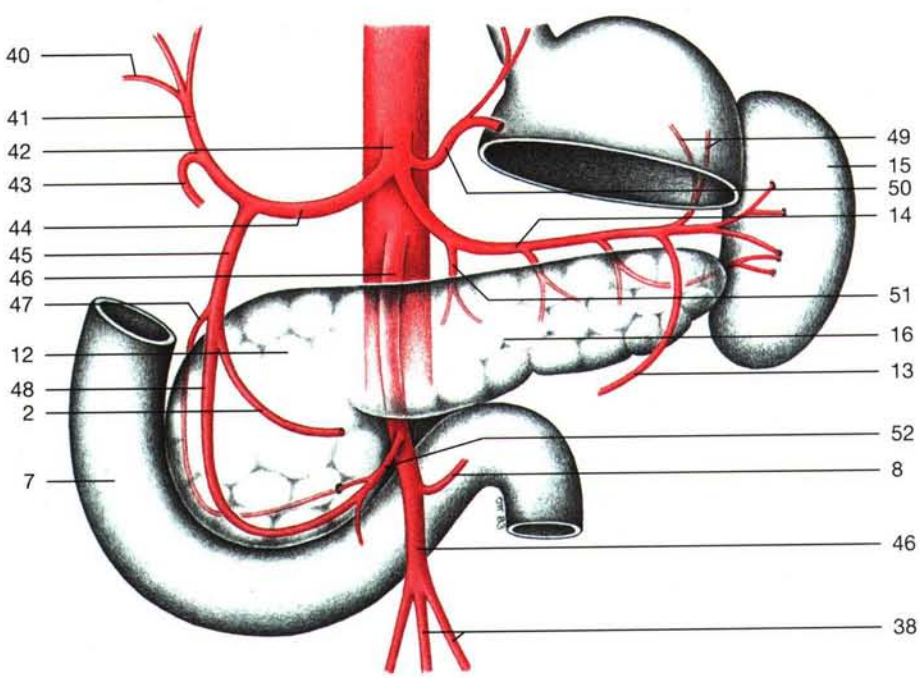

Blood supply of upper abdominal organs (branches of the celiac trunk and superior mesenteric artery). (Schematic drawing.)

Pancreas, extrahepatic bile ducts, spleen and liver with their vessels in situ. The stomach has been removed, the liver elevated, and the descending part of duodenum opened to display the openings of pancreatic ducts. The pancreatic ducts were dissected. Note the location of superior mesenteric artery and vein between duodenum and pancreas. Arrow: probe within the epiploic foramen.

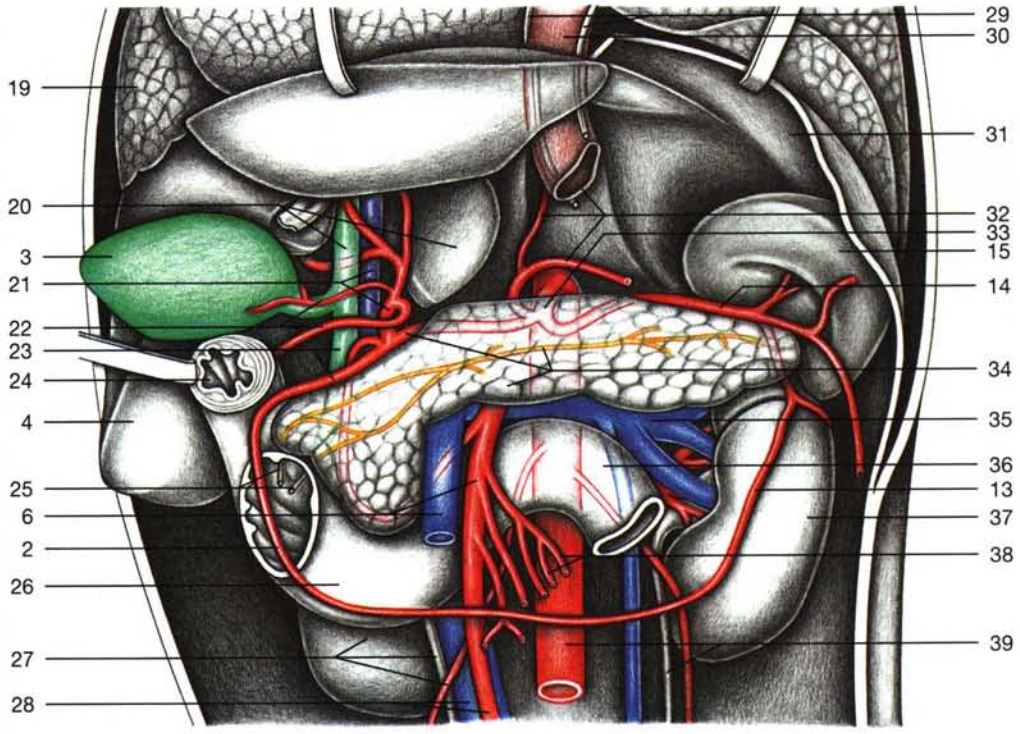

Schematic drawing of the specimen depicted above. Note the branching of the celiac trunk (partially covered by the body of the pancreas), and the main arterial supply of liver, spleen, pancreas, stomach, and duodenum.

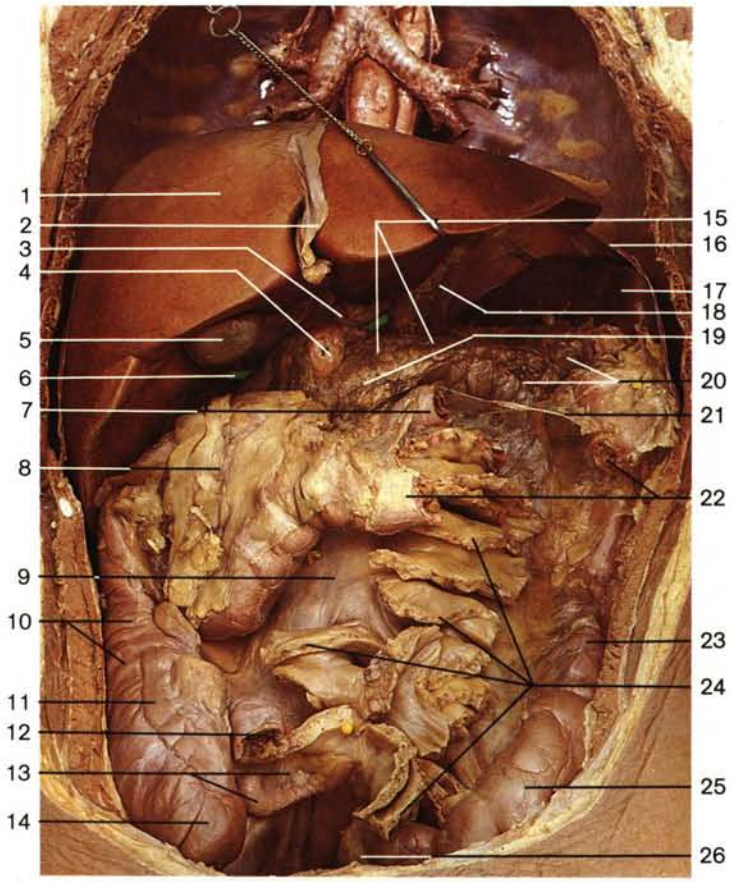

Abdominal cavity after removal of stomach, jejunum, ileum and part of the transverse colon. Liver has been slightly raised.

1 Liver
2 Falciform ligament
3 Hepatoduodenal ligament
4 Pylorus (divided)
5 Gallbladder
6 Probe within the epiploic foramen
7 Duodenojejunal flexure (divided)
8 **Greater omentum**
9 Root of mesentery
10 Ascending colon
11 Free colic taenia
12 End of ileum (divided)
13 Vermiform appendix with mesoappendix
14 Cecum
15 Pancreas and site of lesser sac
16 Diaphragm
17 Spleen
18 Cardia (part of stomach, divided)
19 Head of pancreas
20 Body and tail of pancreas
21 **Transverse mesocolon**
22 Transverse colon (divided)
23 Descending colon
24 Cut edge of mesentery
25 Sigmoid colon
26 Rectum
27 Attachment of bare area of liver
28 Inferior vena cava
29 Kidney
30 Attachment of right colic flexure
31 Root of transverse mesocolon
32 Junction between descending and horizontal parts of duodenum
33 Bare surface for ascending colon
34 **Ileocecal recess**
35 **Retrocecal recess**
36 Root of mesoappendix
37 Superior recess } of **lesser sac (omental bursa)**
38 Isthmus (opening) } of **lesser sac (omental bursa)**
39 Splenic recess } of **lesser sac (omental bursa)**
40 **Superior duodenal recess**
41 **Inferior duodenal recess**
42 Bare surface for descending colon
43 **Paracolic recesses**
44 **Root of mesentery**
45 Root of mesosigmoid
46 **Intersigmoid recess**
47 Hepatic veins
48 Duodenojejunal flexure
49 Attachment of left colic flexure
50 Esophagus
51 Entrance to lesser sac through the epiploic foramen

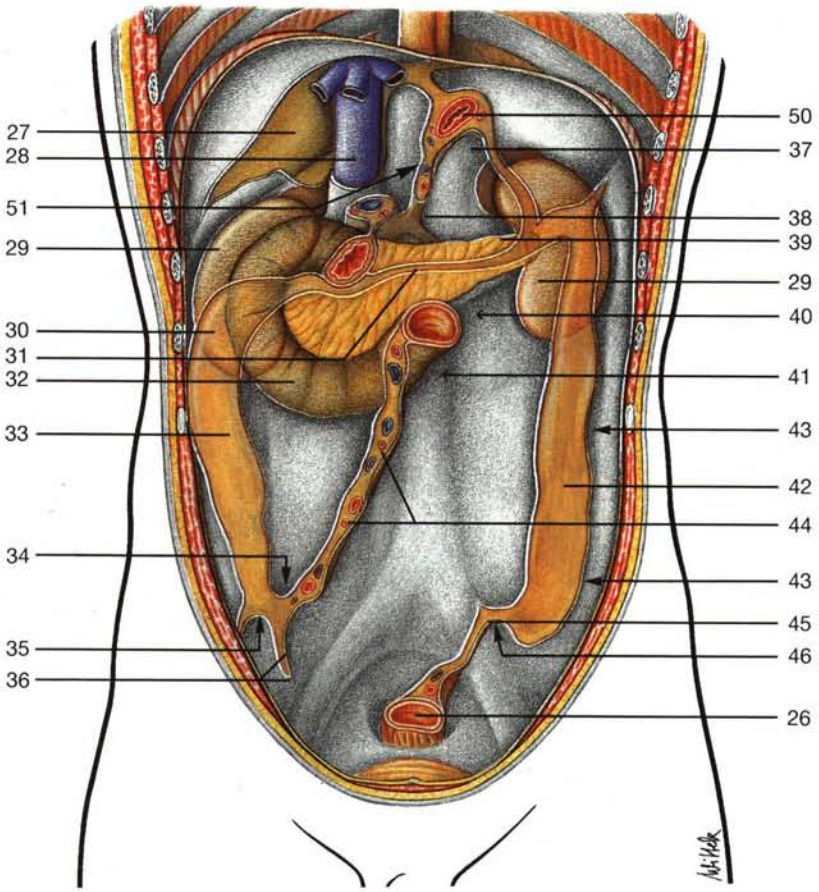

Peritoneal reflections from organs and the position of root of mesentry and peritoneal recesses on the posterior abdominal wall. (Schematic drawing.)

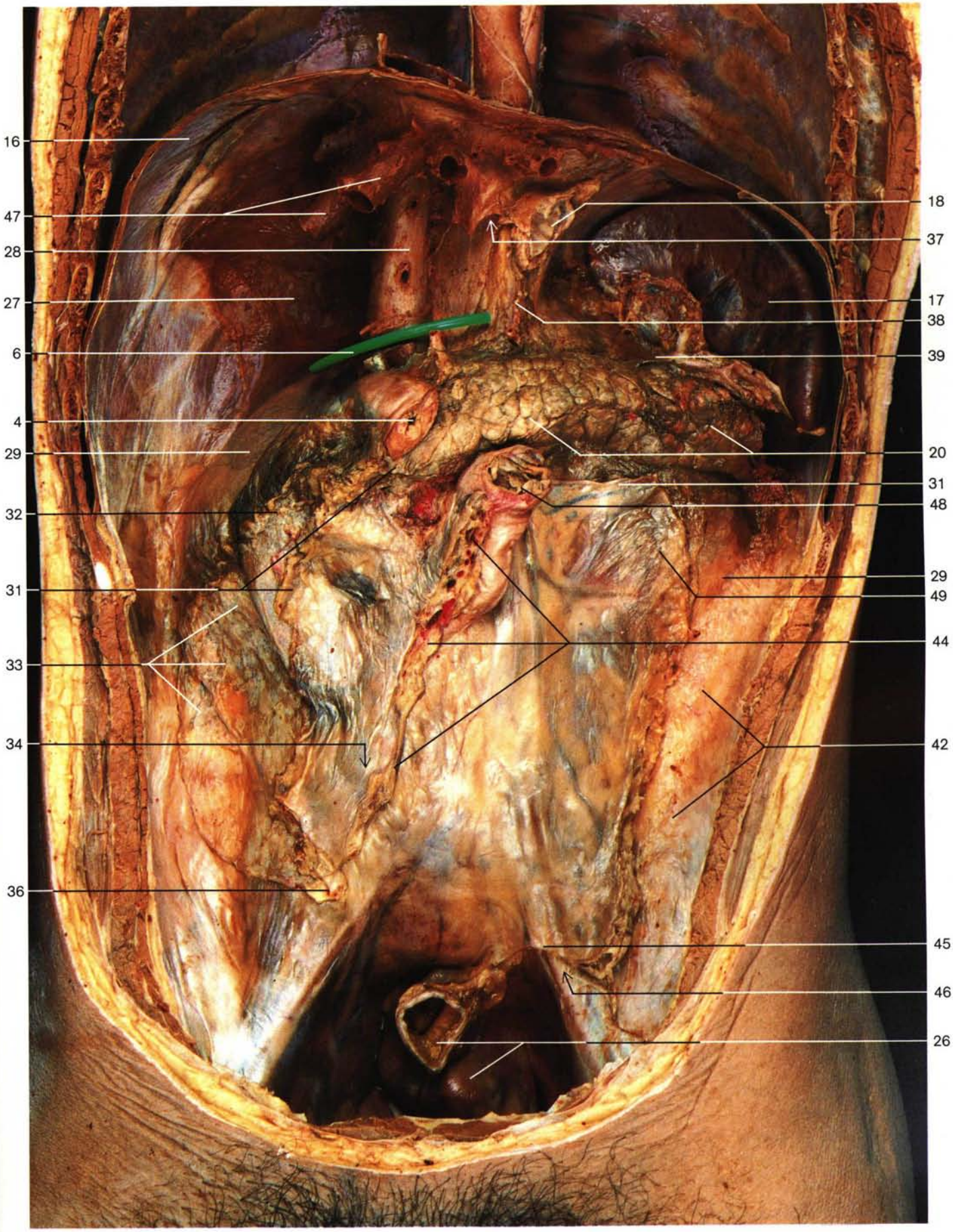

Peritoneal recesses on the posterior abdominal wall. The liver, stomach, jejunum, ileum, and colon have been removed. The duodenum, pancreas, and spleen have been left in place.

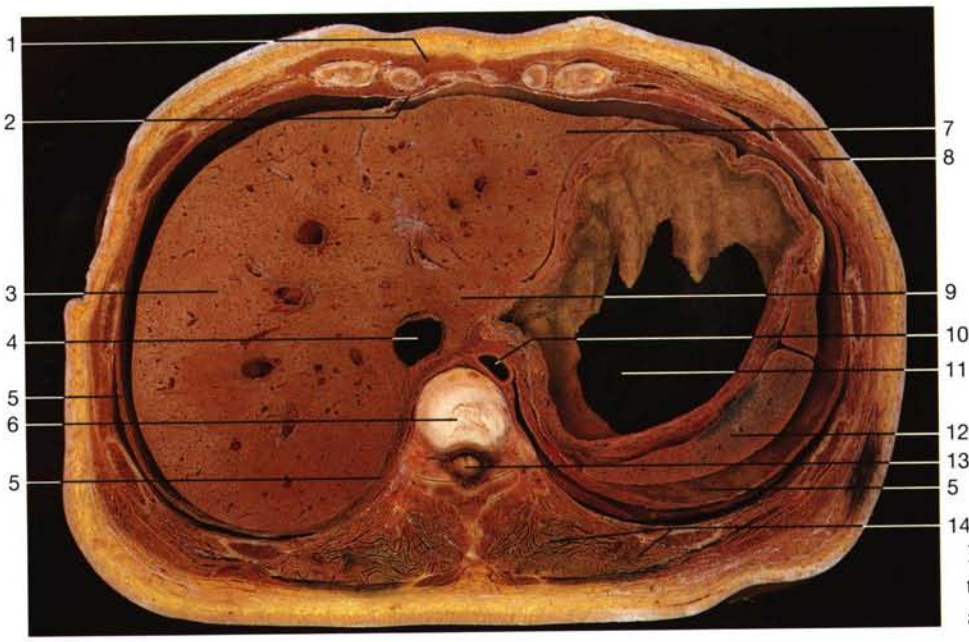

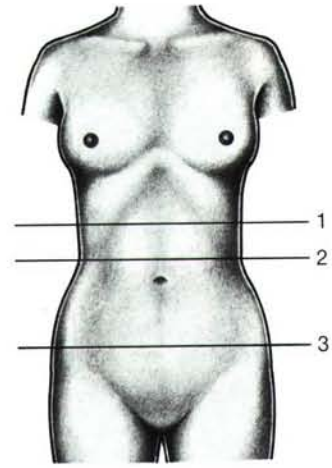

Horizontal section through the abdominal cavity at level 1 (from below).

Horizontal section through the body. MR-Scan, corresponding to level 1. Arrow: stomach.

1 Rectus abdominis muscle
2 Falciform ligament
3 **Liver** (right lobe)
4 Inferior vena cava
5 Diaphragm
6 Intervertebral disc
7 Liver (left lobe)
8 Rib
9 Liver (caudate lobe)
10 Abdominal (descending) aorta
11 **Stomach**
12 **Spleen**
13 Spinal cord
14 Longissimus and iliocostalis muscles
15 Body of vertebra
16 Abdominal rectus muscle
17 External abdominal oblique muscle
18 Transverse colon
19 **Head of pancreas**
20 **Greater duodenal papilla**
21 Duodenum
22 Suprarenal gland and ureter
23 **Kidney**
24 Round ligament of liver
25 Superior mesenteric artery and vein
26 Psoas major muscle
27 Descending colon
28 Quadratus lumborum muscle
29 Cauda equina
30 Right renal vein
31 **Small intestine**
32 Iliacus muscle
33 Ilium
34 **Ileocecal valve**
35 Cecum
36 Common iliac artery and vein
37 Gluteus medius muscle
38 Vertebral canal and dura mater

Horizontal section through the abdominal cavity at the level of greater duodenal papilla (from below).

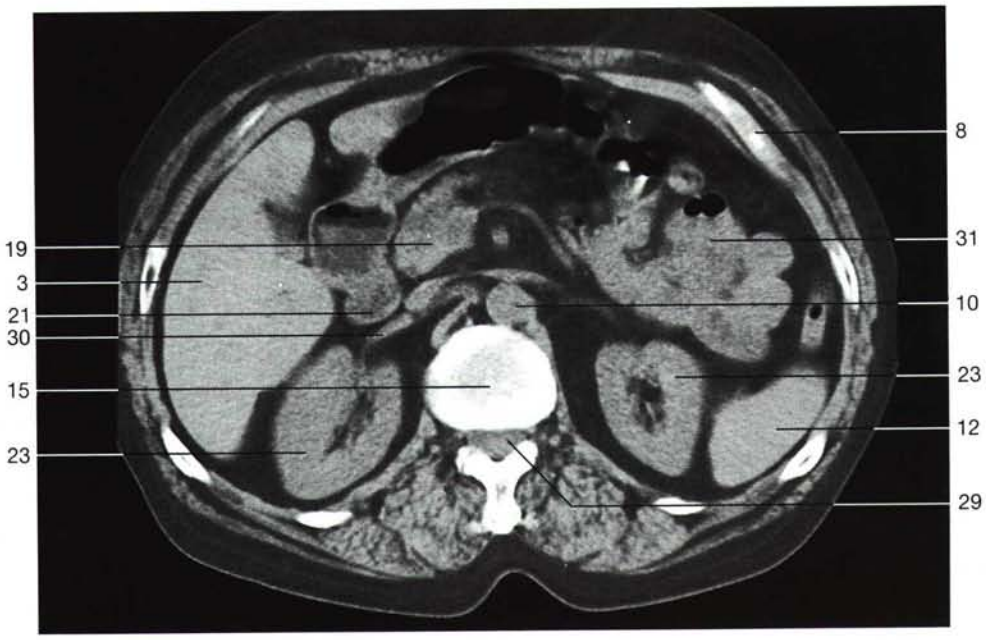

Horizontal section through the body. CT-Scan, corresponding to level 2.

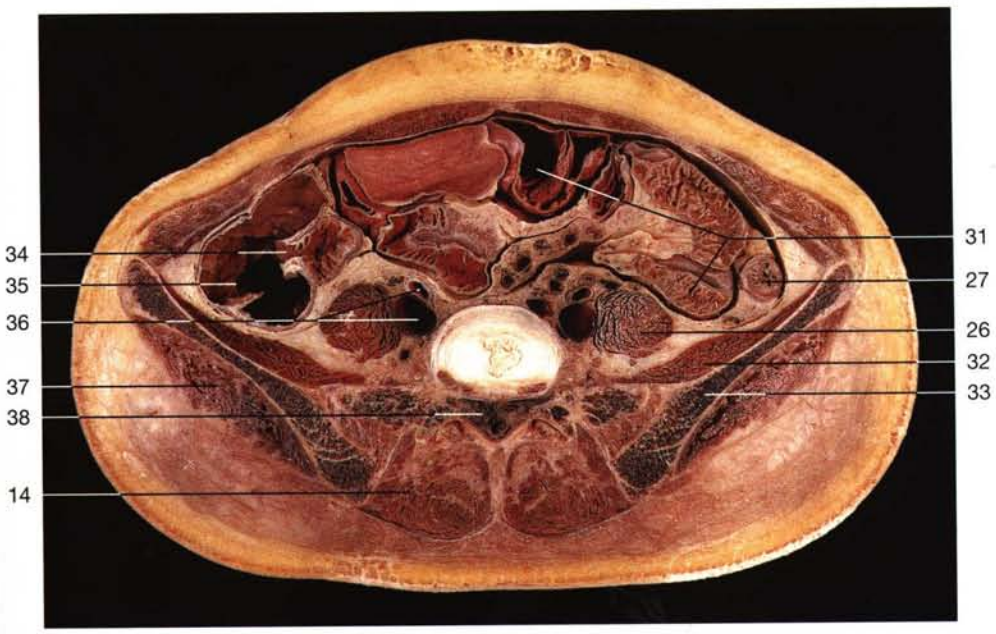

Horizontal section through the abdominal cavity at level 3 (from below).

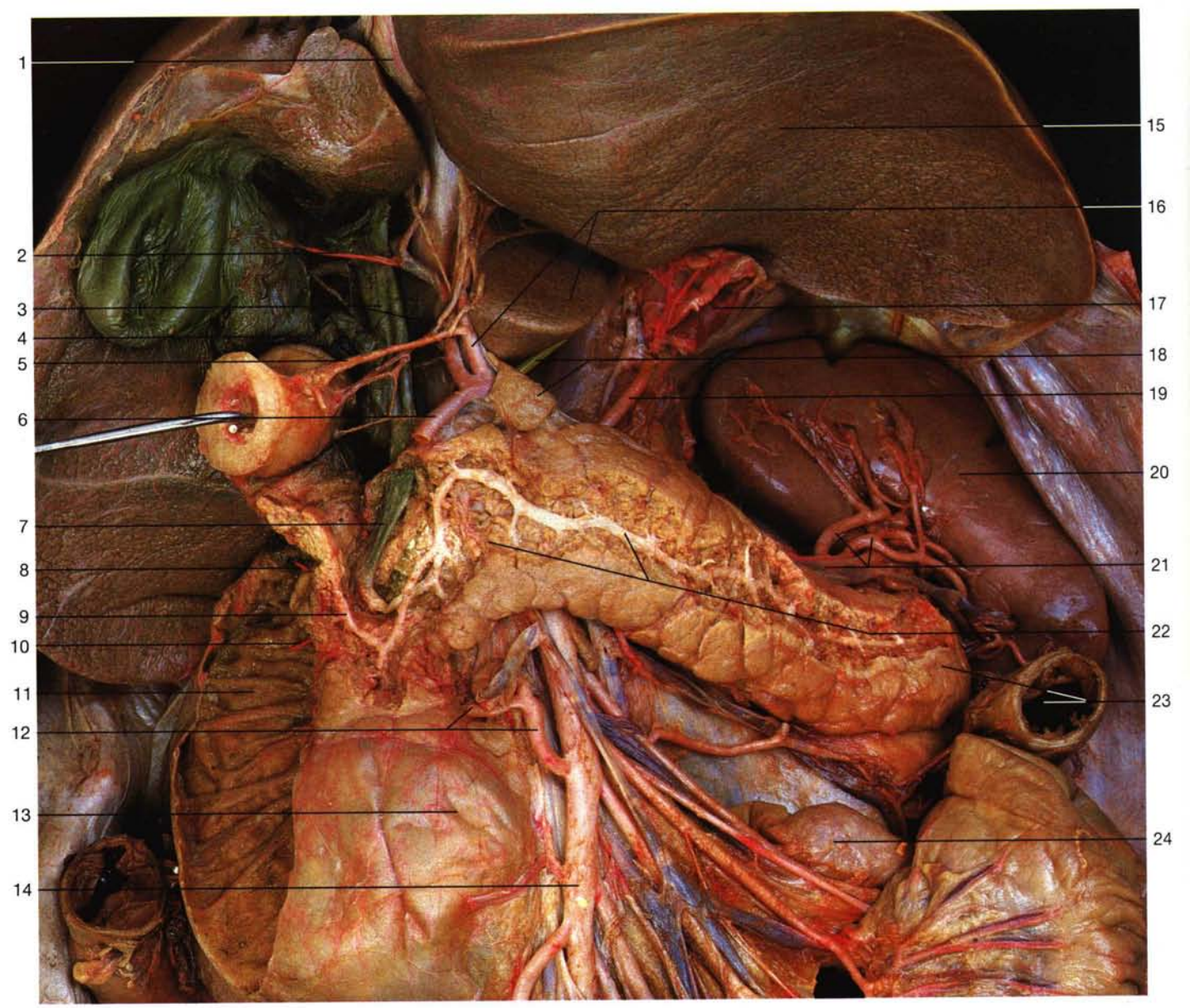

Posterior abdominal wall with duodenum, pancreas and spleen (anterior aspect). Dissection of pancreatic and common bile duct. The stomach has been removed, the liver raised and the duodenum anteriorly opened.

1 Ligamentum teres
2 Gallbladder and cystic artery
3 Common hepatic duct and portal vein
4 Cystic duct
5 Right gastric artery (pylorus with superior part of duodenum, cut and reflected)
6 Gastroduodenal artery
7 **Common bile duct**
8 Probe within the minor duodenal papilla
9 **Accessory pancreatic duct**
10 Probe within the major duodenal papilla
11 Descending part of duodenum (opened)
12 Middle colic artery and inferior pancreaticoduodenal artery
13 Horizontal part of duodenum (distended)
14 **Superior mesenteric artery**
15 Liver (left lobe)
16 Caudate lobe of liver and hepatic artery proper
17 Abdominal part of esophagus (cut)
18 Probe in epiploic foramen and lymph node
19 Left gastric artery
20 **Spleen**
21 Splenic vein and branches of splenic artery
22 **Pancreatic duct and head of pancreas**
23 Left colic flexure and tail of pancreas
24 Duodenojejunal flexure

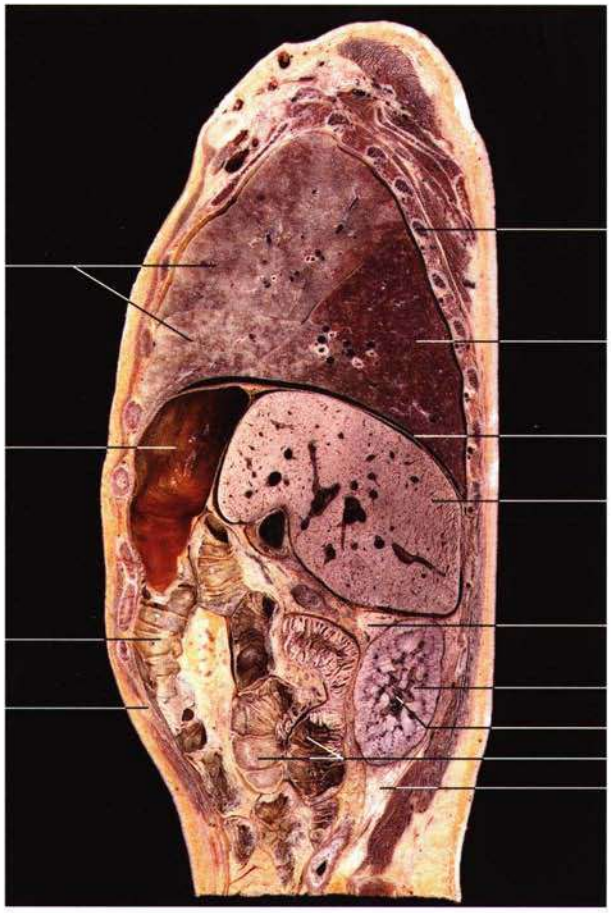

Parasagittal section through the thoracic and abdominal cavities (right side, medial aspect).

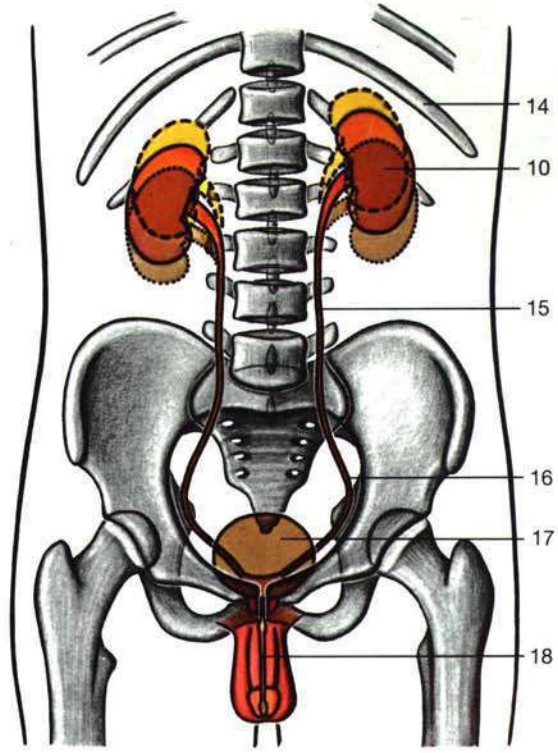

Position of kidneys and urinary system (anterior view). The excursions of the kidneys with the respiratory movements of the diaphragm are indicated. (Schematic drawing.)

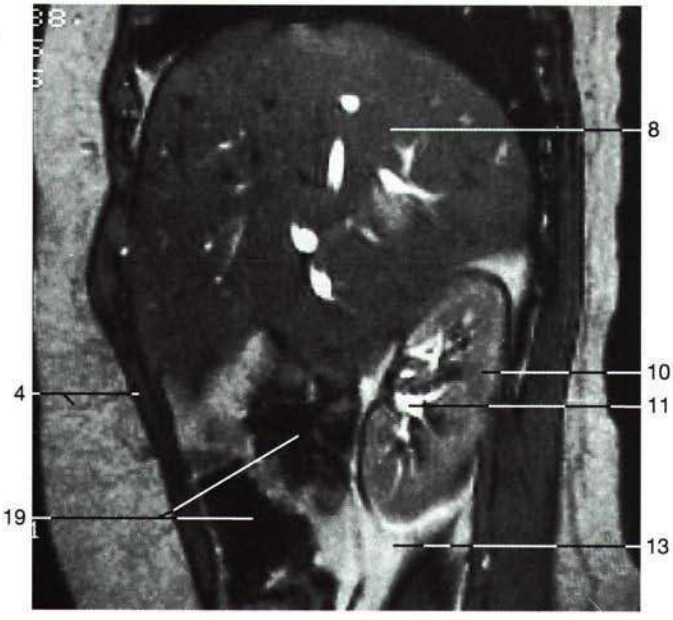

Parasagittal section through the abdominal cavity. (MR-Scan, courtesy of Prof. W. Rödl, Erlangen).

1 Right lung (superior and middle lobes)
2 Transverse colon
3 Jejunum
4 Abdominal wall
5 Fourth rib
6 Right lung (inferior lobe)
7 Diaphragm
8 **Liver**
9 Suprarenal gland
10 **Kidney**
11 Renal pelvis
12 Small intestine
13 Perirenal fat pad
14 Eleventh rib
15 Ureter (abdominal part)
16 Ureter (pelvic part)
17 **Urinary bladder**
18 Urethra
19 Colon

Horizontal section through the abdominal cavity at the level of the first lumbar vertebra (from below).

1 Pyloric antrum
2 Gastroduodenal artery
3 Descending part of duodenum
4 Vestibule of lesser sac
5 Inferior vena cava and liver
6 Body of first lumbar vertebra
7 Cauda equina
8 **Right kidney**
9 Latissimus dorsi muscle
10 Iliocostalis muscle
11 Rectus abdominis muscle
12 Stomach
13 Lesser sac
14 Splenic vein
15 Superior mesenteric artery
16 Pancreas
17 Aorta and left renal artery
18 Transverse colon
19 Renal artery and vein
20 Spleen
21 **Left kidney**
22 Psoas major muscle
23 Multifidus muscle
24 Margin of lung
25 Margin of pleura
26 Renal pelvis
27 **Left ureter**
28 Descending colon
29 Rectum
30 Right suprarenal gland
31 12th rib
32 Ascending colon
33 **Right ureter**
34 Cecum
35 Vermiform appendix
36 Urinary bladder
37 Liver
38 Anterior layer of renal fascia
39 Duodenum
40 Perirenal fatty tissue
41 Posterior layer of renal fascia
42 Abdominal cavity

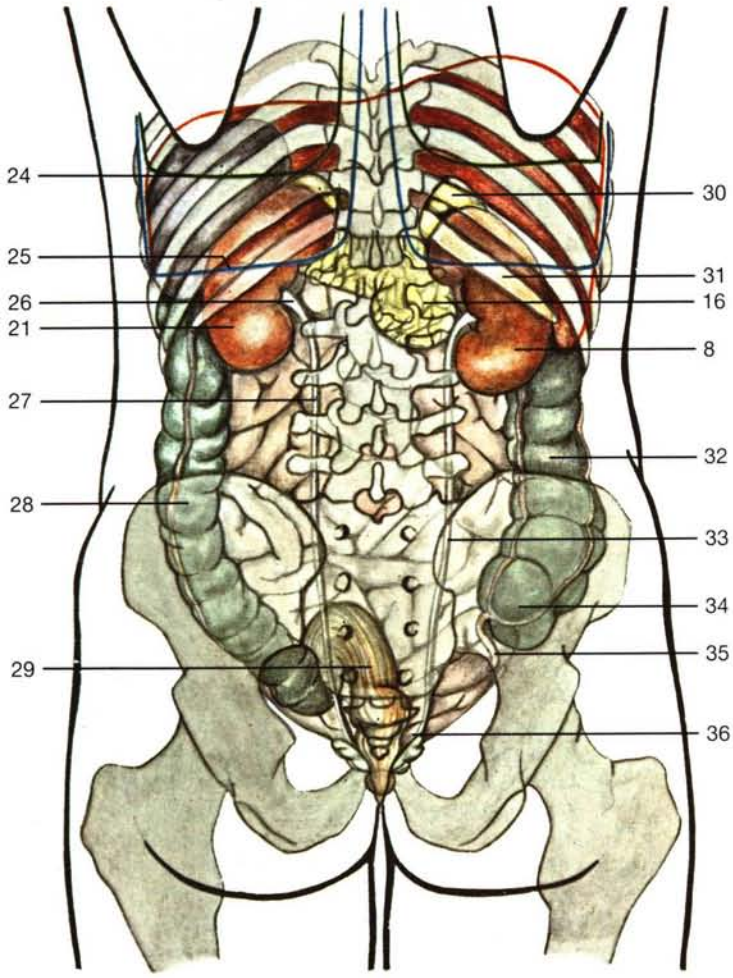

Positions of urinary organs (posterior view). Notice that the upper part of the kidney reaches the level of the margin of pleura and lung.

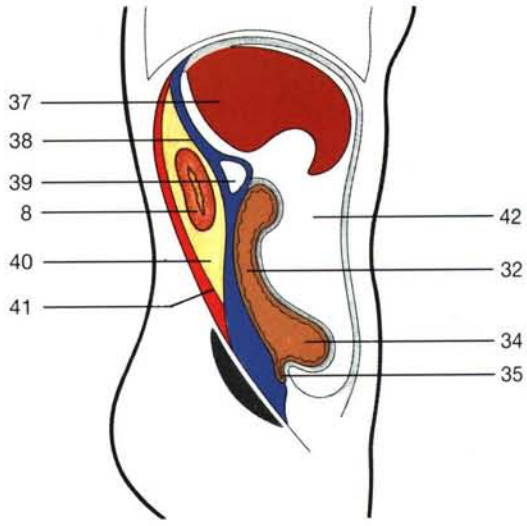

Retroperitoneal tissue, location of the right kidney. (Schematic drawing.) Yellow = adipose capsule of kidney.

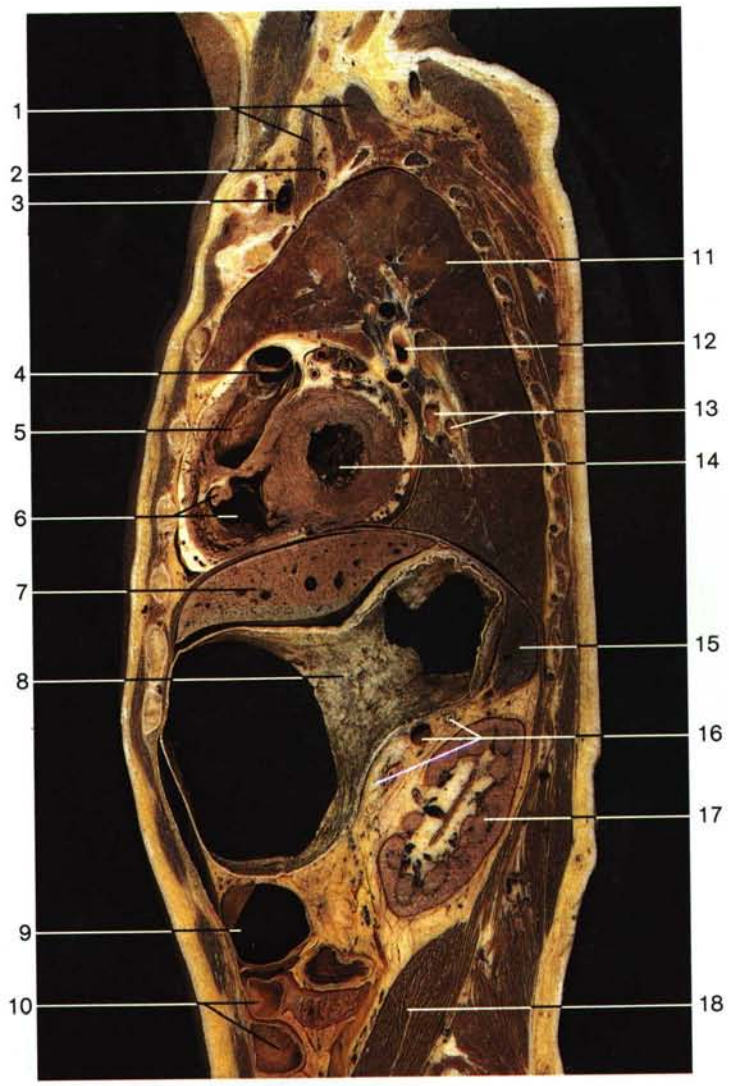

1 Scalenus anterior, medius and posterior muscles
2 Left subclavian artery
3 Left subclavian vein
4 Pulmonic valve
5 Arterial cone
6 Right ventricle of heart
7 Liver
8 Stomach
9 Transverse colon
10 Small intestine
11 Left lung
12 Left main bronchus
13 Branches of pulmonary vein
14 Left ventricle of heart
15 Spleen
16 Splenic artery and vein and pancreas
17 **Left kidney**
18 Psoas major muscle
19 Inferior vena cava
20 **Renal vein**
21 Body of 12th thoracic vertebra and vertebral canal
22 **Right kidney**
23 Superior mesenteric artery
24 Superior mesenteric vein
25 Pancreas
26 Abdominal aorta
27 Left psoas major and quadratus lumborum muscles
28 Anterior layer of renal fascia } of Gerota
29 Posterior layer of renal fascia } of Gerota
30 Perirenal fatty tissue
31 Abdominal cavity
32 Descending and sigmoid colon

Parasagittal section through the thoracic and abdominal cavities at the level of the left kidney
(5.5 cm left of median plane).

Horizontal section through the retroperitoneal region
at level of 12th thoracic vertebra. (CT-Scan, from below.)

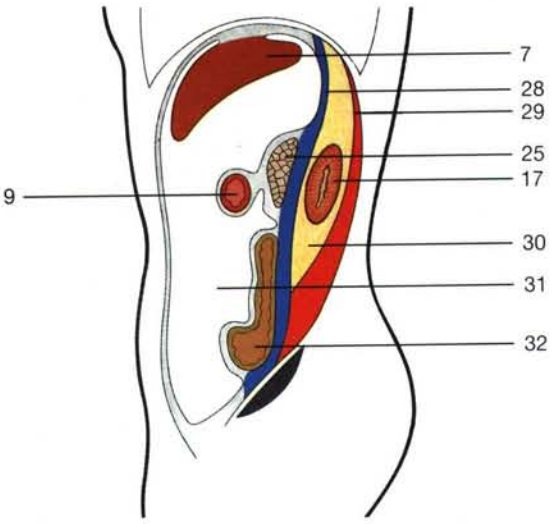

Retroperitoneal tissue, position of left kidney.
(Schematic drawing.)

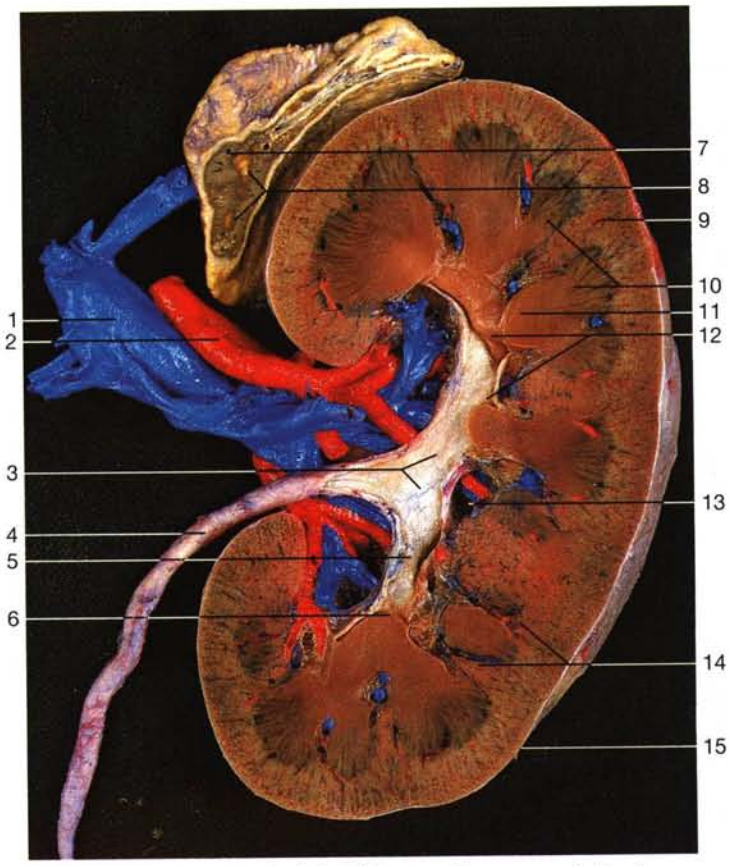

Coronal section through right kidney and suprarenal gland (posterior view). The renal pelvis has been opened and the fatty tissue removed to display the renal vessels.

1 Renal vein
2 Renal artery
3 **Renal pelvis**
4 Abdominal part of ureter
5 **Major renal calyx**
6 Cribriform area of renal papilla
7 Cortex of suprarenal gland
8 Medulla of suprarenal gland
9 **Cortex of kidney**
10 **Medulla of kidney**
11 Renal papilla
12 **Minor renal calyx**
13 Renal sinus
14 Renal columns
15 **Fibrous capsule** of kidney

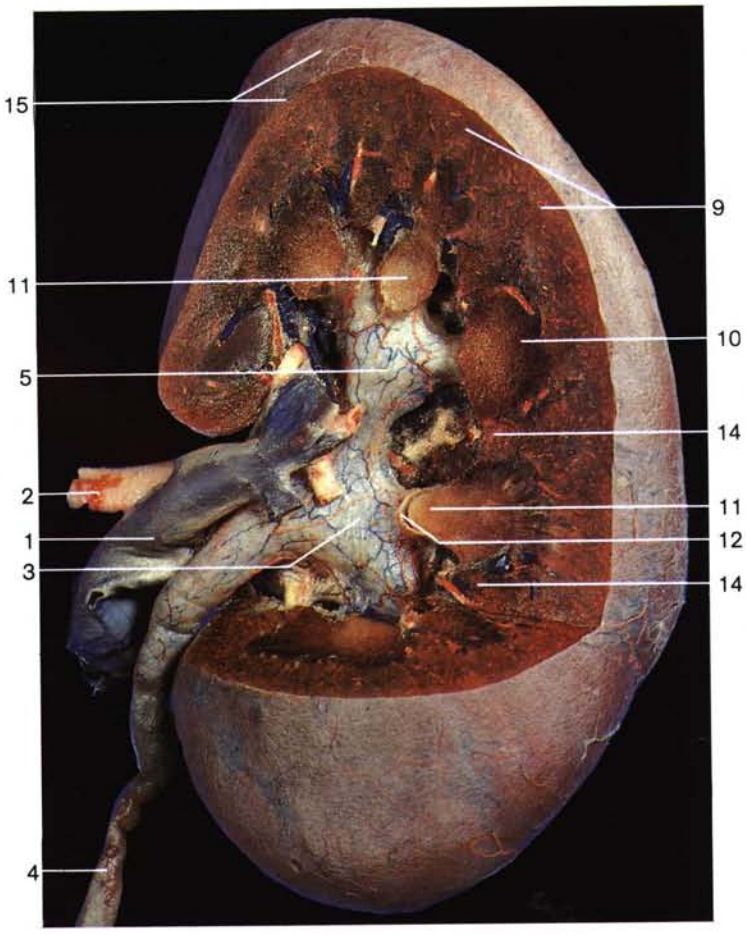

Right kidney (posterior view). Partial coronal section to expose internal aspect of the kidney.

Each kidney can be divided in five segments supplied by individual interlobar arteries considered as end arteries. Thus, obstruction leads to infarcts marking the trace of segment borders. The anterior kidney surface reveals four segments, the posterior only three (No. 1, 4 and 5).

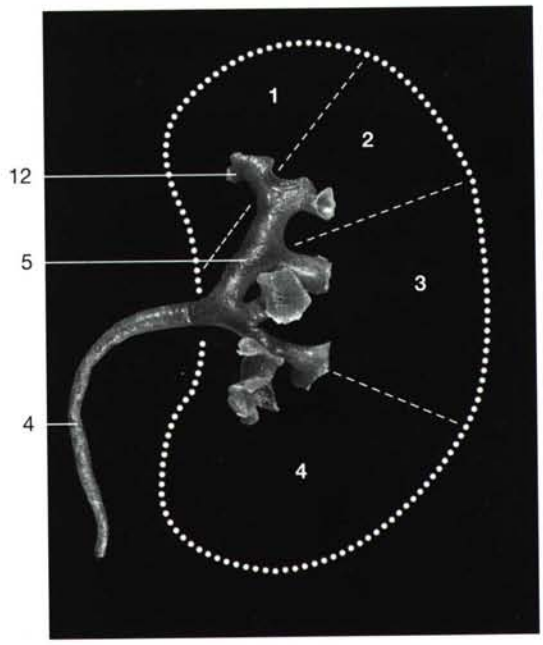

Cast of renal pelvis and calices.
1–4 = Renal segments on anterior surface.

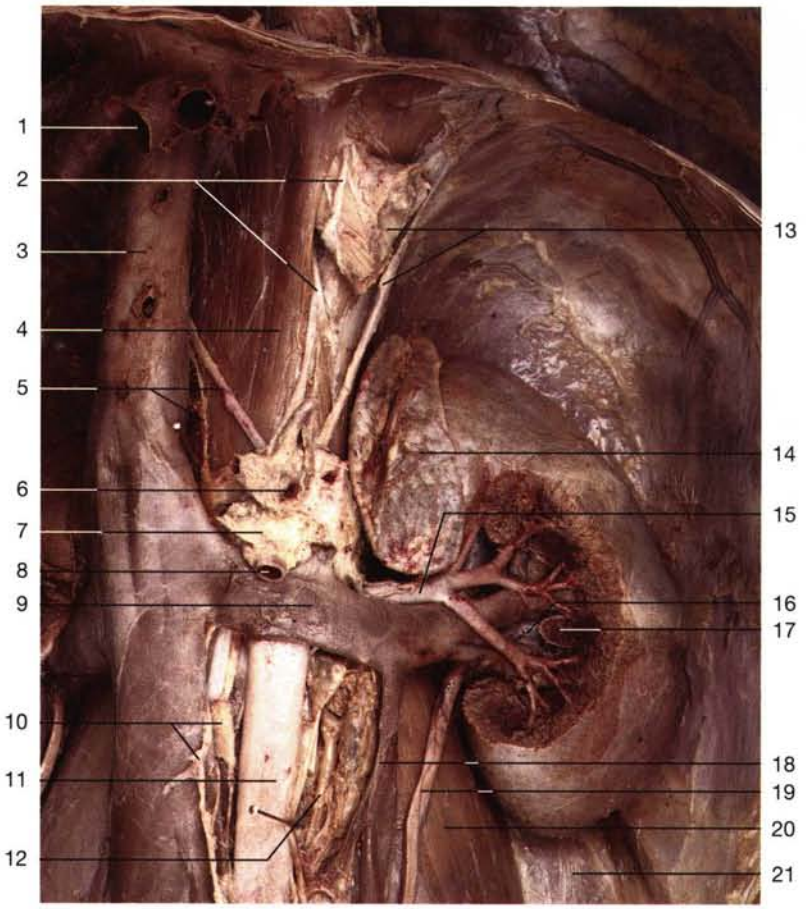

Left kidney and suprarenal gland in situ. The anterior cortical layer of the kidney has been removed to display the renal pelvis and papillae.

1 Hepatic vein
2 Anterior and posterior vagal trunk
3 Inferior vena cava
4 Lumbar part of diaphragm
5 Right greater and lesser splanchnic nerves
6 Celiac trunk
7 Celiac ganglion and plexus
8 Superior mesenteric artery
9 Left renal vein
10 Right sympathetic trunk and ganglion
11 Abdominal aorta
12 Left sympathetic trunk
13 Esophagus (cut), left greater splanchnic nerve
14 **Left suprarenal gland**
15 Left renal artery
16 Renal pelvis
17 Renal papilla with minor calyx
18 Left testicular vein
19 Left ureter
20 Psoas major muscle
21 Quadratus lumborum muscle
22 **Glomerulus**
23 Afferent arteriole of glomerulus
24 **Glomeruli**
25 Radiating cortical artery
26 Subcortical or arcuate artey
27 Subcortical or arcuate vein
28 Interlobular vein
29 Interlobular artery
30 **Interlobar artery and vein**
31 Vessels of renal capsule
32 Efferent arteriole of glomerulus
33 **Vasa recta** of renal medulla
34 Spiral arteries of renal pelvis

Glomeruli (210×). Scanning electron micrograph showing glomeruli and associated arteries.

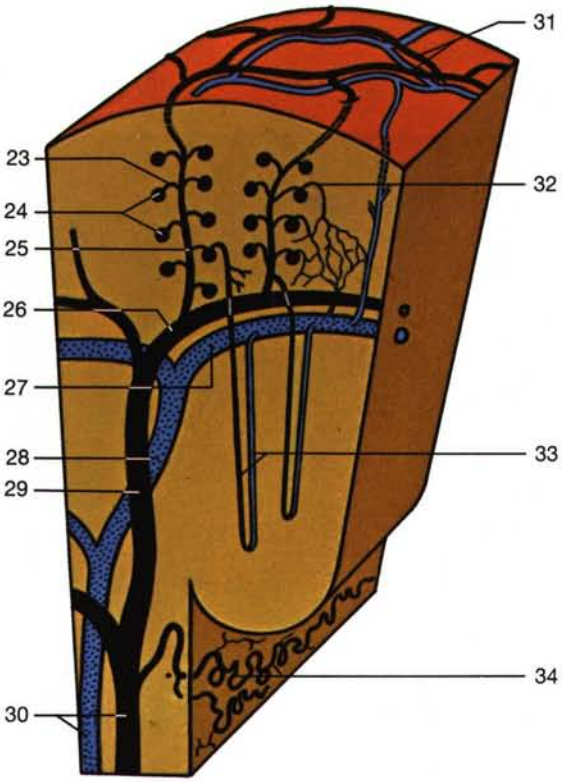

Architecture of vascular system of kidney. (Schematic drawing.)

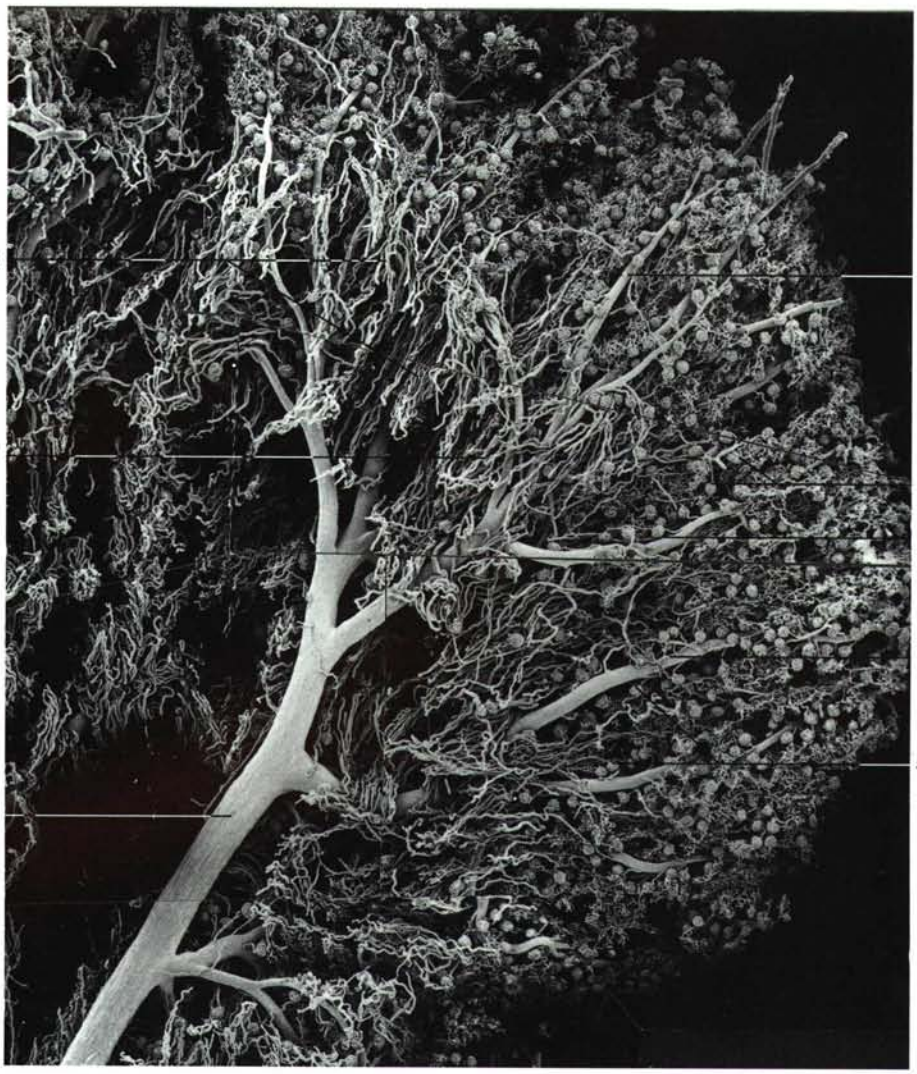

1 Arteriolae rectae of renal medulla
2 Interlobar artery
3 Interlobular arteries
4 Cortical glomeruli
5 Juxtamedullary glomeruli
6 Body of first lumbar vertebra
7 Left renal artery
8 Abdominal aorta with catheter
9 Upper pole of kidney
10 Anterior branch } of renal artery
11 Posterior branch } of renal artery
12 Anterior inferior segmental artery
13 Lower pole of kidney
14 Celiac trunk
15 Superior mesenteric artery
16 Middle colic artery
17 Splenic artery

Resin cast of kidney arteries. (Scanning electron micrograph.)

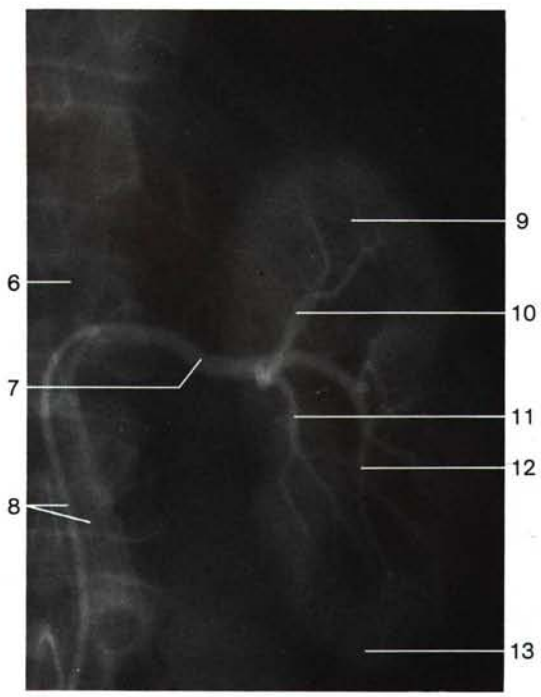

Left kidney. (Arteriogram.)

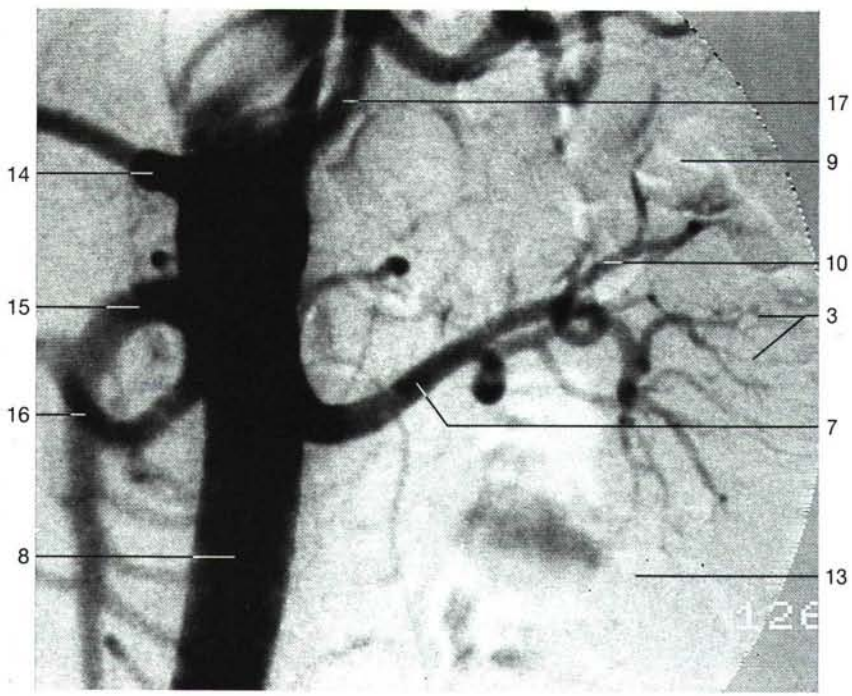

Abdominal aorta. (Subtraction angiograph.)

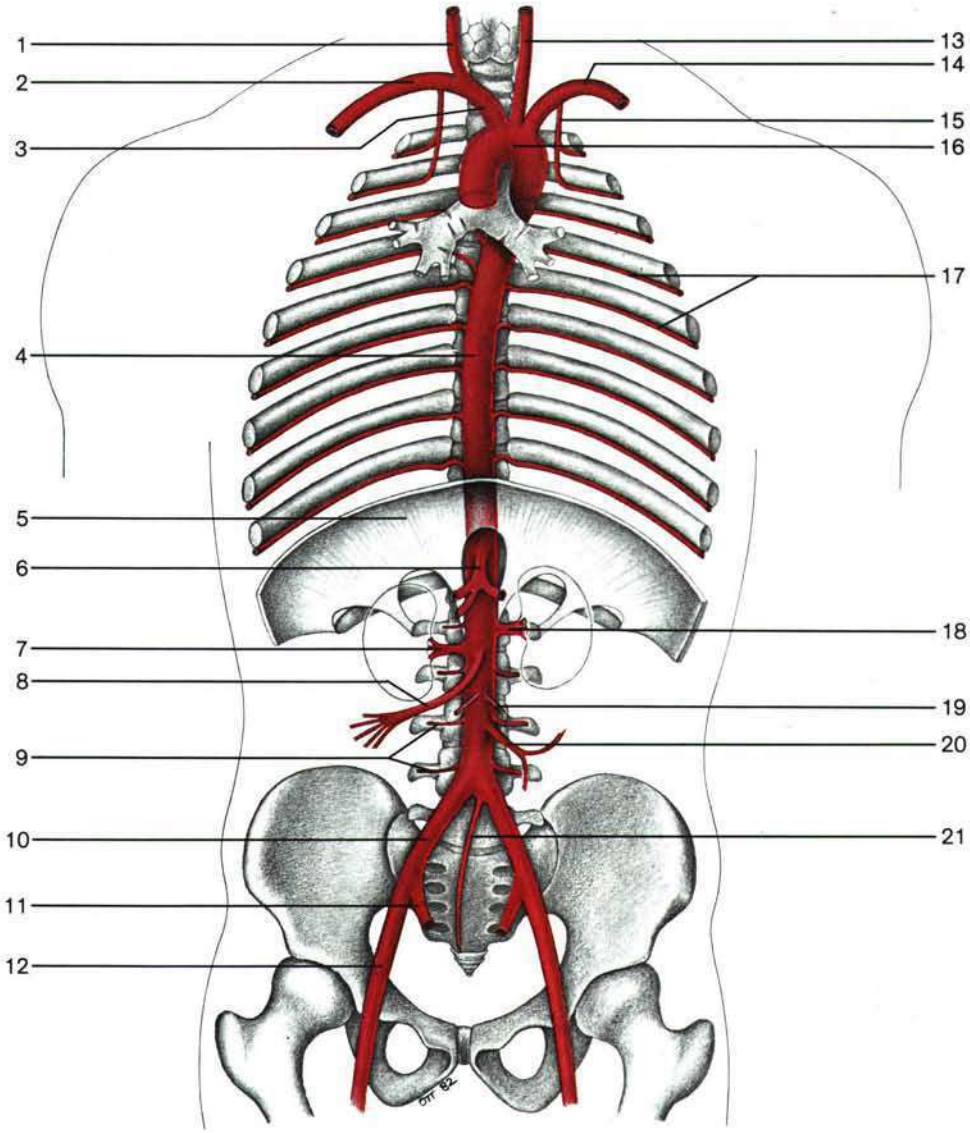

1 Right common carotid artery
2 Right subclavian artery
3 **Brachiocephalic trunk**
4 Thoracic aorta
5 Diaphragm
6 **Celiac trunk**
7 Right **renal artery**
8 **Superior mesenteric artery**
9 Lumbar arteries
10 Right **common iliac artery**
11 Internal iliac artery
12 **External iliac artery**
13 Left common carotid artery
14 Left subclavian artery
15 Highest intercostal artery
16 Aortic arch
17 **Posterior intercostal arteries**
18 Left renal artery
19 Left testicular (or ovarian) artery
20 Inferior mesenteric artery
21 Median sacral artery
22 **Superior suprarenal artery**
23 Upper capsular artery
24 Anterior branch of renal artery
25 Perforating artery
26 Lower capsular artery
27 Ureter
28 Right inferior phrenic artery
29 Left inferior phrenic artery
30 **Middle suprarenal artery**
31 **Inferior suprarenal artery**
32 Posterior branch of renal artery

Main branches of descending aorta. (Schematic drawing.)

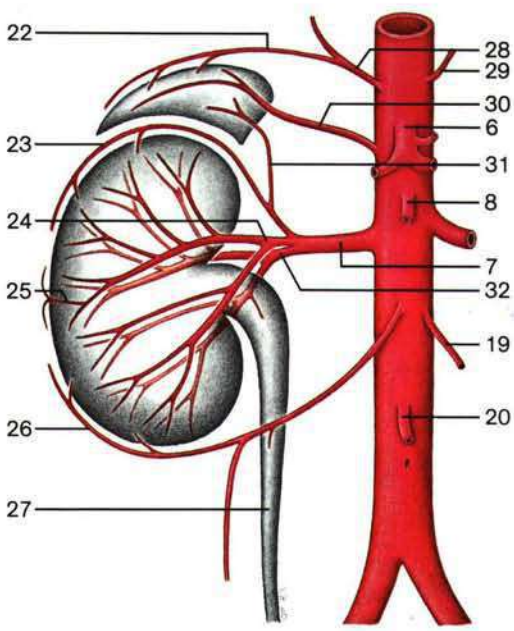

Arteries of kidney and suprarenal gland. (Schematic drawing.)

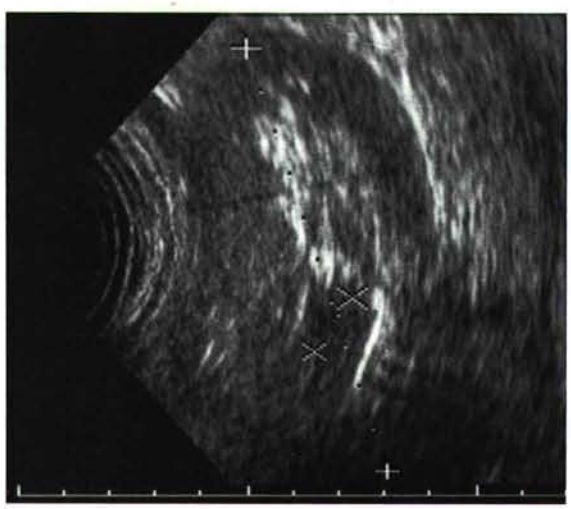

Ultrasound image of the right kidney (upper and lower border of the kidney marked by crosses. × = small cortical cyst).

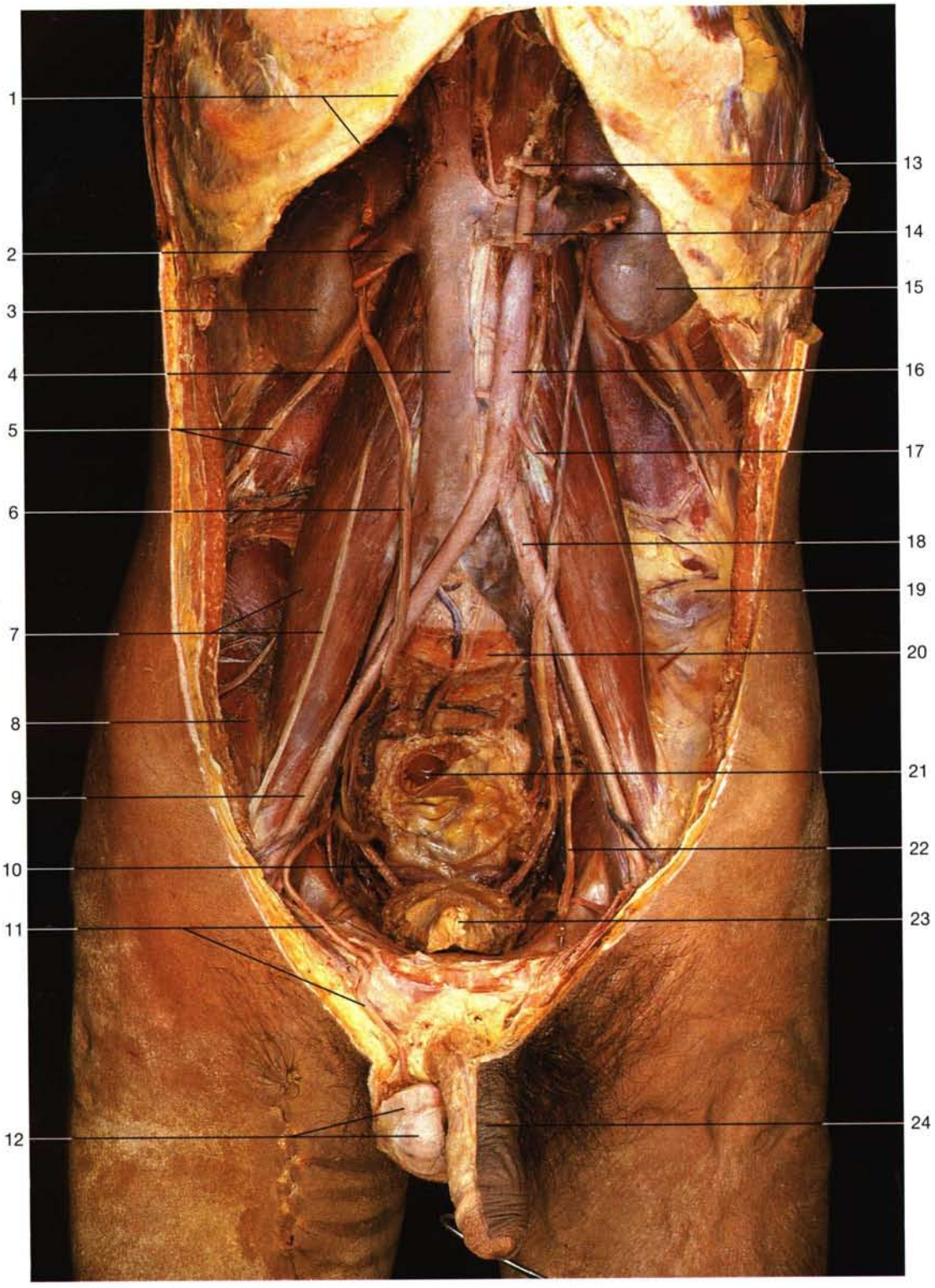

Retroperitoneal organs, urinary system in the male (anterior view). The peritoneum has been removed.

1 Costal arch
2 Right renal vein
3 **Right kidney**
4 Inferior vena cava
5 Iliohypogastric nerve and quadratus lumborum
6 **Ureter** (abdominal part)
7 Psoas major muscle and genitofemoral nerve
8 Iliacus muscle
9 External iliac artery
10 **Ureter** (pelvic part)
11 Ductus deferens
12 Testis and epididymis
13 Celiac trunk
14 Superior mesenteric artery
15 **Left kidney**
16 Abdominal aorta
17 Inferior mesenteric artery
18 Common iliac artery
19 Iliac crest
20 Sacral promontory
21 Rectum (cut)
22 Medial umbilical ligament
23 **Urinary bladder**
24 Penis

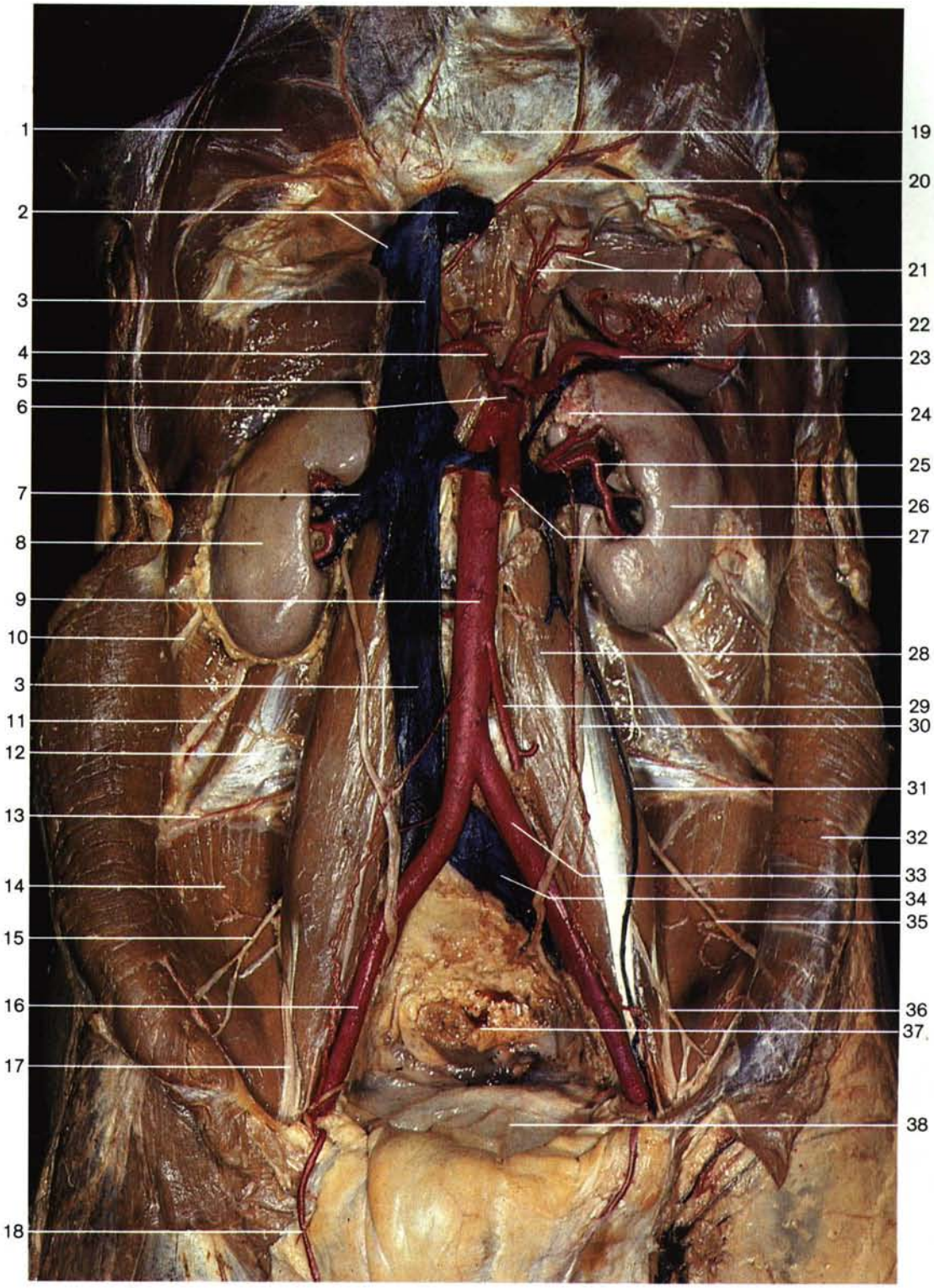

Retroperitoneal organs, urinary system in situ (anterior view). The peritoneum has been removed. Red = arteries; blue = veins.

1 Diaphragm
2 Hepatic veins
3 **Inferior vena cava**
4 Common hepatic artery
5 Right suprarenal gland
6 Celiac trunk
7 **Right renal vein**
8 Right kidney
9 **Abdominal aorta**
10 Subcostal nerve
11 Iliohypogastric nerve
12 Quadratus lumborum muscle
13 Iliac crest
14 Iliacus muscle
15 Right lateral femoral cutaneous nerve
16 **External iliac artery**
17 Femoral nerve
18 Right inferior epigastric artery
19 Central tendon of diaphragm
20 Inferior phrenic artery
21 Cardiac part of stomach and esophageal branches of left gastric artery
22 Spleen
23 Splenic artery
24 Left suprarenal gland
25 **Left renal artery**
26 Left **kidney**
27 Superior mesenteric artery
28 Psoas major muscle
29 Inferior mesenteric artery
30 **Ureter**
31 Testicular artery and vein
32 Transversus abdominis muscle
33 Left common iliac artery
34 Left common iliac vein
35 Lateral femoral cutaneous nerve
36 Genitofemoral nerve
37 Rectum (cut)
38 Urinary bladder

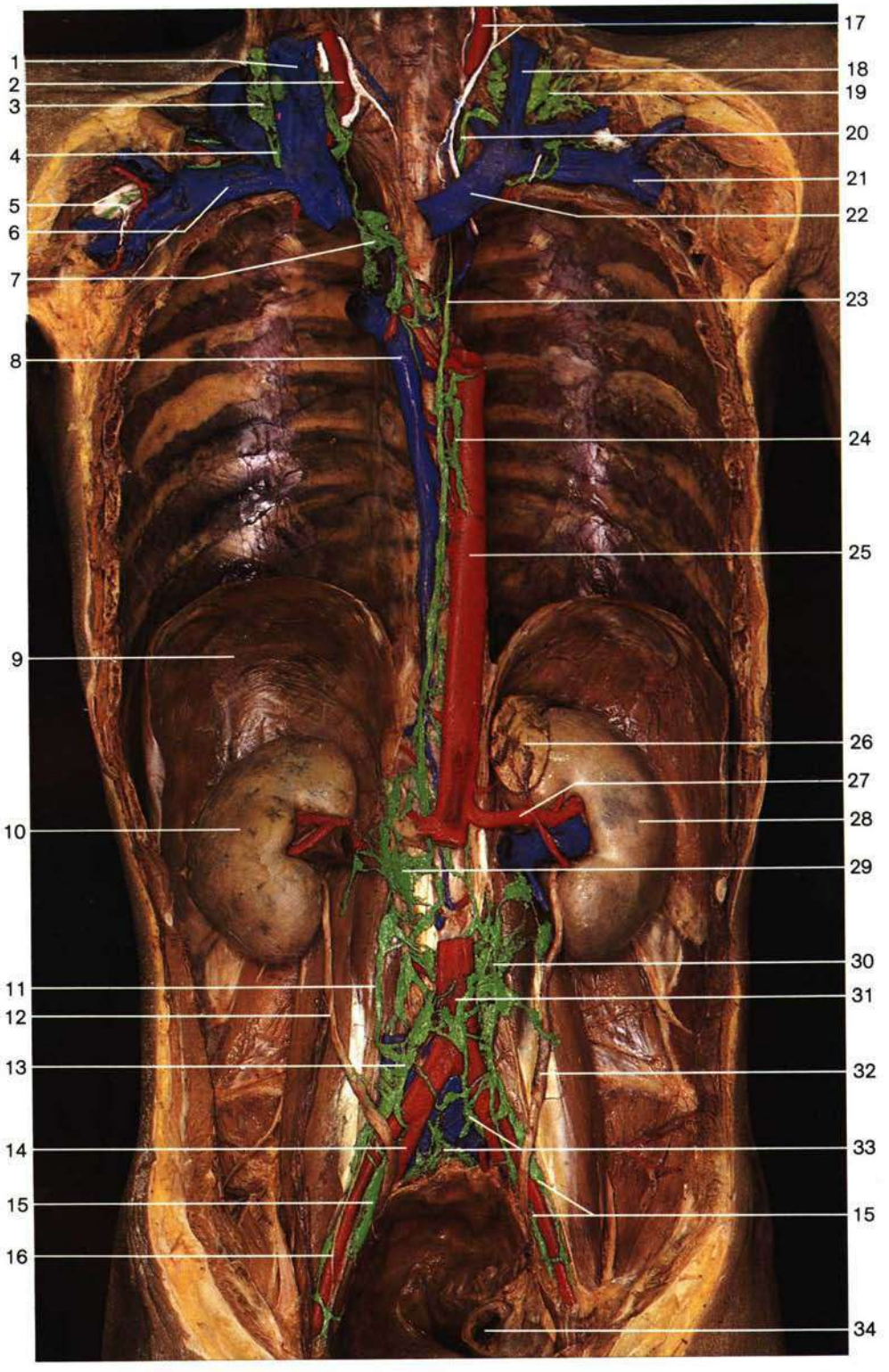

1 Internal jugular vein
2 Right common carotid artery and right vagus nerve
3 Jugulo-omohyoid lymph node
4 Right lymphatic duct
5 Subclavian trunk
6 Right subclavian vein
7 Bronchomediastinal trunk
8 Azygos vein
9 Diaphragm
10 Right kidney
11 Right lumbar trunk
12 Right ureter
13 Common iliac lymph nodes
14 Right internal iliac artery
15 External iliac lymph nodes
16 Right external iliac artery
17 Left common carotid artery and left vagus nerve
18 Internal jugular vein
19 Deep cervical lymph nodes
20 Thoracic duct entering left jugular angle
21 Left subclavian vein
22 Left brachiocephalic vein
23 **Thoracic duct**
24 Mediastinal lymph nodes
25 Thoracic aorta
26 Left suprarenal gland
27 Left renal artery
28 Left kidney
29 Cisterna chyli
30 Lumbar lymph nodes
31 Abdominal aorta
32 Left ureter
33 Sacral lymph nodes
34 Rectum (cut edge)

Lymph vessels and lymph nodes of the posterior wall of thoracic and abdominal cavities (anterior aspect). Green = lymph vessels and nodes; blue = veins; red = arteries; white = nerves.

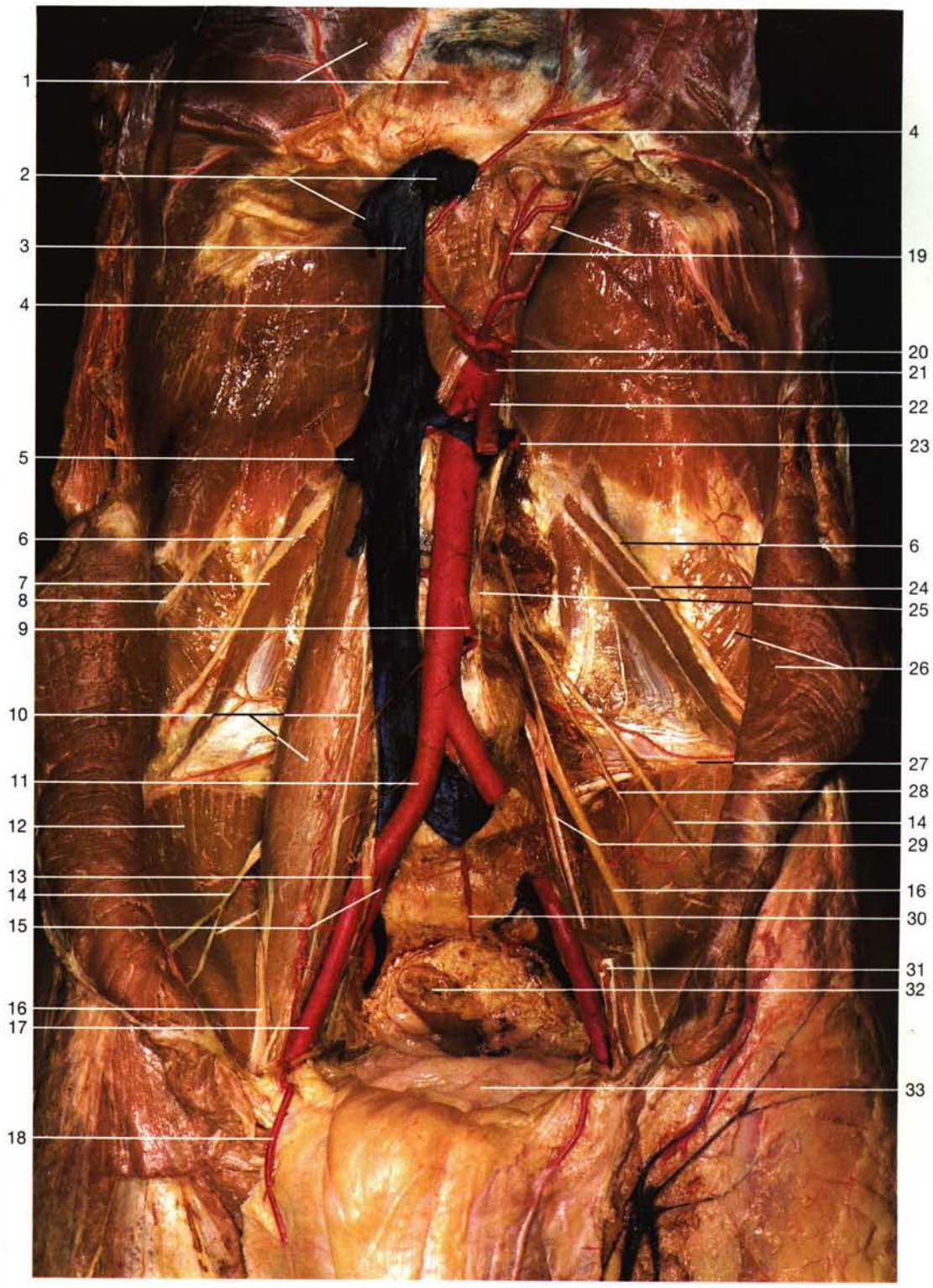

Vessels and nerves of posterior abdominal wall (anterior aspect). Part of the left psoas major muscle has been removed to display the lumbar plexus. Red = arteries, blue = veins.

1 Diaphragm
2 Hepatic veins
3 **Inferior vena cava**
4 Inferior phrenic artery
5 Right renal vein
6 **Iliohypogastric nerve**
7 Quadratus lumborum muscle
8 Subcostal nerve
9 Inferior mesenteric artery
10 Right genitofemoral nerve and psoas major muscle
11 Common iliac artery
12 Iliacus muscle
13 Right ureter (divided)
14 Lateral femoral cutaneous nerve
15 Internal iliac artery
16 **Femoral nerve**
17 **External iliac artery**
18 Inferior epigastric artery
19 Cardiac part of stomach and oesophageal branches of left gastric artery
20 Splenic artery
21 Celiac trunk
22 **Superior mesenteric artery**
23 Left renal artery
24 **Ilioinguinal nerve**
25 Sympathetic trunk
26 Transversus abdominis muscle
27 Iliac crest
28 Left **genitofemoral nerve**
29 Left **obturator nerve**
30 Median sacral artery
31 Psoas major muscle (divided) with supplying artery
32 Rectum (cut)
33 Urinary bladder

Posterior wall of thoracic and abdominal cavities with sympathetic trunk, vagus nerve and autonomic ganglia (anterior view). Thoracic and abdominal organs removed, except esophagus and aorta.

Organization of autonomic nervous system (after Mattuschka). (Schematic drawing.) Yellow = parasympathetic nerves; green = sympathetic nerves.

Ganglia and plexus of the autonomic nervous system within the retroperitoneal space (anterior view). The kidneys and the inferior vena cava with its tributaries have been removed (compare pp. 264 and 265).

1 **Right vagus nerve**
2 Right subclavian artery
3 Esophagus
4 Aortic arch
5 **Sympathetic trunk**
6 Greater splanchnic nerve
7 Intercostal nerve
8 Abdominal part of esophagus and vagal trunk
9 **Celiac trunk with celiac ganglion**
10 **Superior mesenteric artery and ganglion**
11 Psoas major muscle and genitofemoral nerve
12 Common iliac artery
13 **Superior hypogastric plexus and ganglion**
14 **Left vagus nerve**
15 Brachial plexus
16 Left subclavian artery
17 Left recurrent laryngeal nerve
18 Inferior cervical cardiac nerve
19 Thoracic aorta
20 Esophageal plexus
21 Azygos vein
22 Diaphragm
23 Splenic artery
24 Left renal artery and plexus
25 **Inferior mesenteric ganglion and artery**
26 Left external iliac artery
27 **Superior cervical ganglion** of sympathetic trunk
28 Superior cardiac branch of sympathetic trunk
29 Middle cervical ganglion of sympathetic trunk
30 **Inferior cervical ganglion** of sympathetic trunk
31 Right recurrent laryngeal nerve
32 Lesser splanchnic nerve
33 Lumbar splanchnic nerves
34 Sacral splanchnic nerves
35 **Inferior hypogastric ganglion and plexus**
36 Left recurrent laryngeal nerve
37 **Aorticorenal plexus** and renal artery
38 Ganglion impar
39 Esophagus with branches of vagus nerve
40 Hepatic veins
41 Right crus of diaphragm
42 Inferior phrenic artery
43 Right vagus nerve entering the celiac ganglion
44 Right lumbar lymph trunk
45 Lumbar part of right sympathetic trunk
46 Lumbar artery and vein
47 Psoas major muscle
48 Iliac crest
49 Inferior vena cava
50 Iliacus muscle
51 Ureter
52 Left vagus nerve forming the esophageal plexus
53 Left vagus nerve forming the gastric plexus
54 Esophagus continuing into the cardiac part of stomach
55 Lumbocostal triangle
56 Position of 12th rib
57 Left lumbar lymph trunk
58 Ganglion of sympathetic trunk
59 Quadratus lumborum muscle
60 Lumbar part of left sympathetic trunk
61 Iliac lymph vessels

Male genital organs isolated (right lateral aspect).

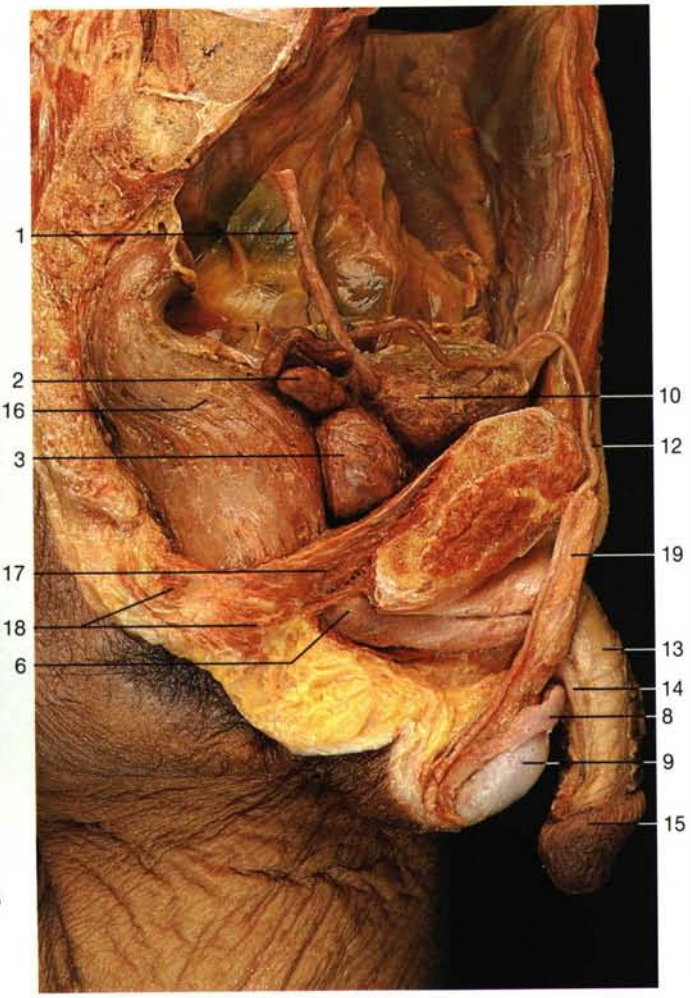

Male genital organs in situ (right lateral aspect).

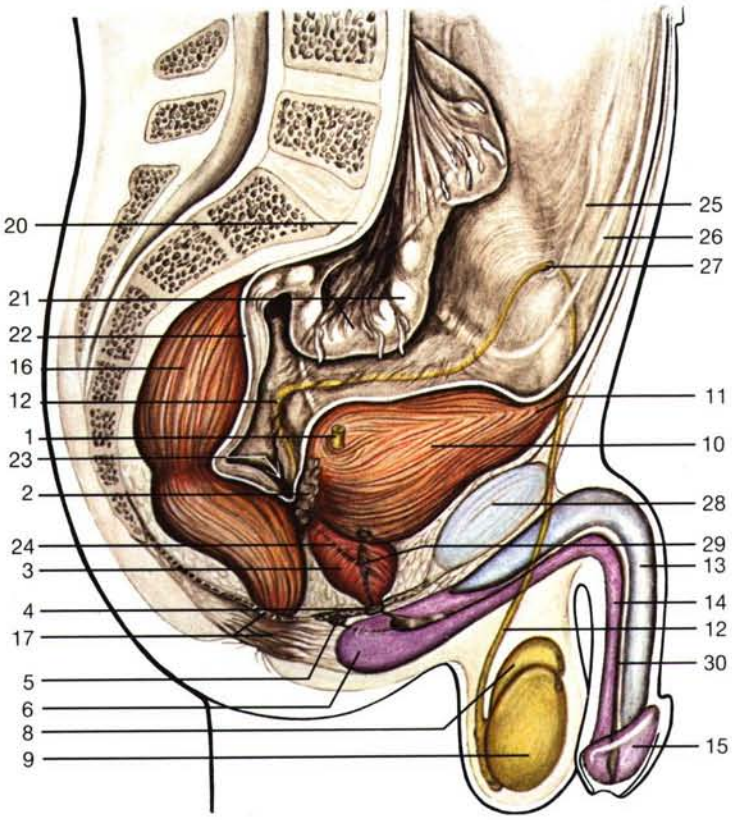

Positions of male genital organs (right lateral aspect). (Schematic drawing.)

1 Ureter
2 **Seminal vesicle**
3 **Prostate gland**
4 Urogenital diaphragm and membranous part of urethra
5 Bulbourethral or Cowper's gland
6 Bulb of penis
7 Left and right crus penis
8 **Epididymis**
9 **Testis**
10 **Urinary bladder**
11 Apex of urinary bladder
12 **Ductus deferens**
13 **Corpus cavernosum of penis**
14 **Corpus spongiosum of penis**
15 Glans penis
16 Ampulla of rectum
17 Levator ani muscle
18 Anal canal and external anal sphincter muscle
19 Spermatic cord (cut)
20 Sacral promontory
21 Sigmoid colon
22 Peritoneum (cut edge)
23 **Rectovesical pouch**
24 Ejaculatory duct
25 Lateral umbilical fold
26 Medial umbilical fold
27 Deep inguinal ring and ductus deferens
28 Pubic symphysis
29 Prostatic part of urethra
30 Spongy urethra

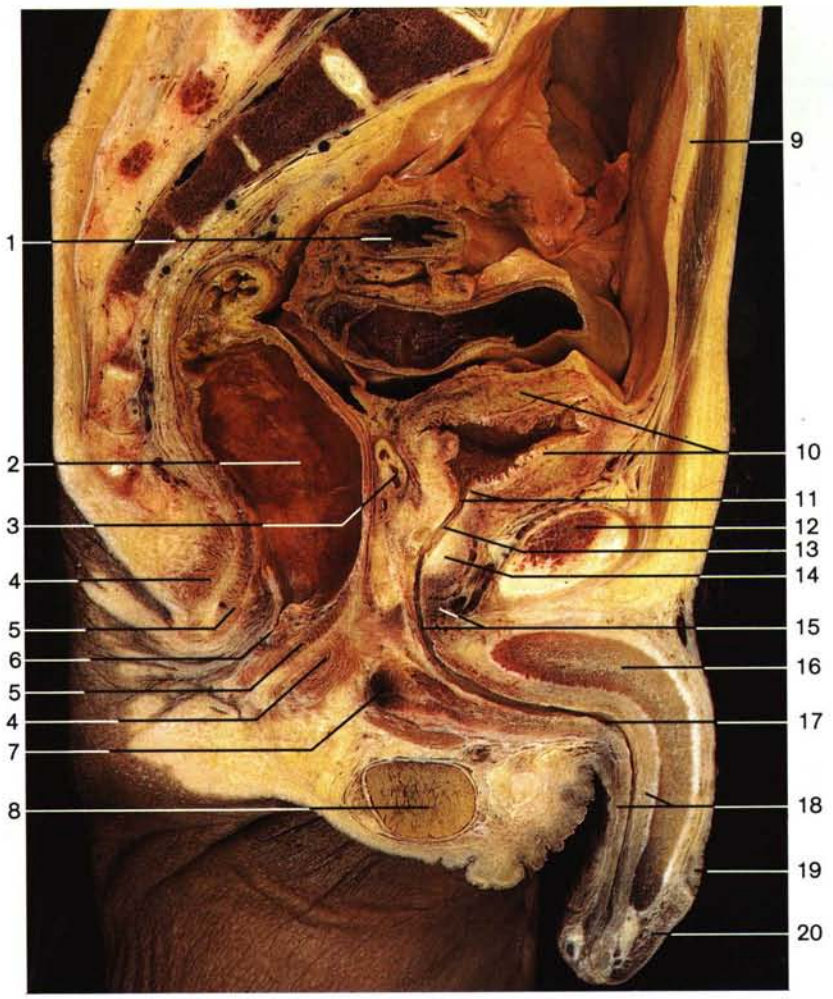

Midsagittal section through the male pelvis.

1 Sigmoid colon
2 **Ampulla of rectum**
3 Ampulla of ductus deferens
4 External anal sphincter muscle
5 Internal anal sphincter muscle
6 **Anal canal**
7 Bulb of penis
8 **Testis** (cut surface)
9 Median umbilical ligament
10 **Urinary bladder**
11 Internal urethral orifice and sphincter muscle (sphincter vesicae)
12 Pubic symphysis
13 Prostatic part of urethra
14 **Prostate gland**
15 Membranous urethra and external urethral sphincter muscle
16 **Corpus cavernosum of penis**
17 **Spongy urethra**
18 Corpus spongiosum of penis
19 Foreskin or prepuce
20 Glans penis
21 Kidney
22 Renal pelvis
23 Abdominal part of ureter
24 Pelvic part of ureter
25 Seminal vesicle
26 Ejaculatory duct
27 Bulbourethral or Cowper's gland
28 **Ductus deferens**
29 **Epididymis**
30 Umbilicus
31 Trigone of bladder and ureteric orifice
32 Navicular fossa of urethra
33 External urethral orifice
34 **Testis**

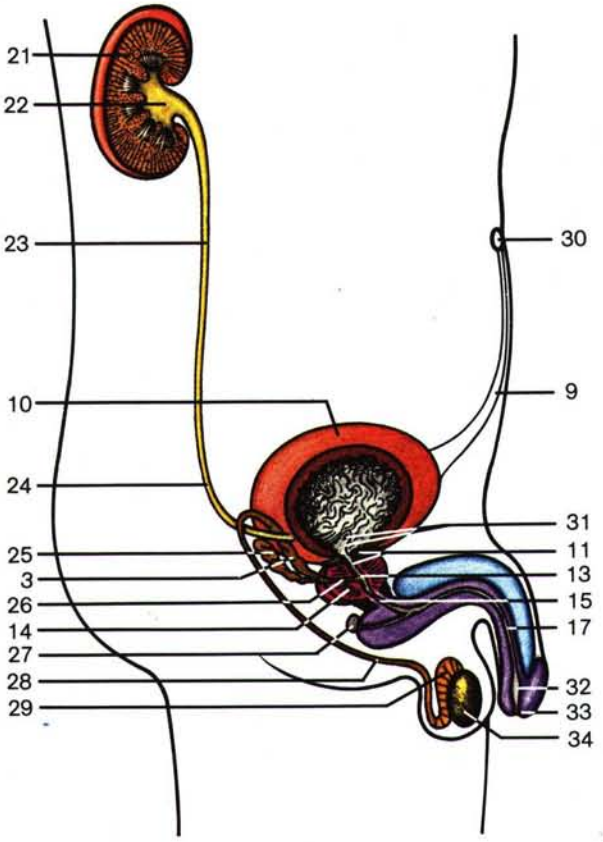

Male urogenital system. (Schematic drawing.)

The **prostate** is located between bladder and urogenital diaphragm. The penis includes the **urethra** and thus serves for both ejaculation and micturition. The internal (involuntary) and external (voluntary) urethral sphincters are widely separated. The **ureter** having crossed the ductus deferens enters the urinary bladder at its base. The peritoneum is reflected off the posterior surface of the bladder onto the rectum thus forming the rectovesical pouch.

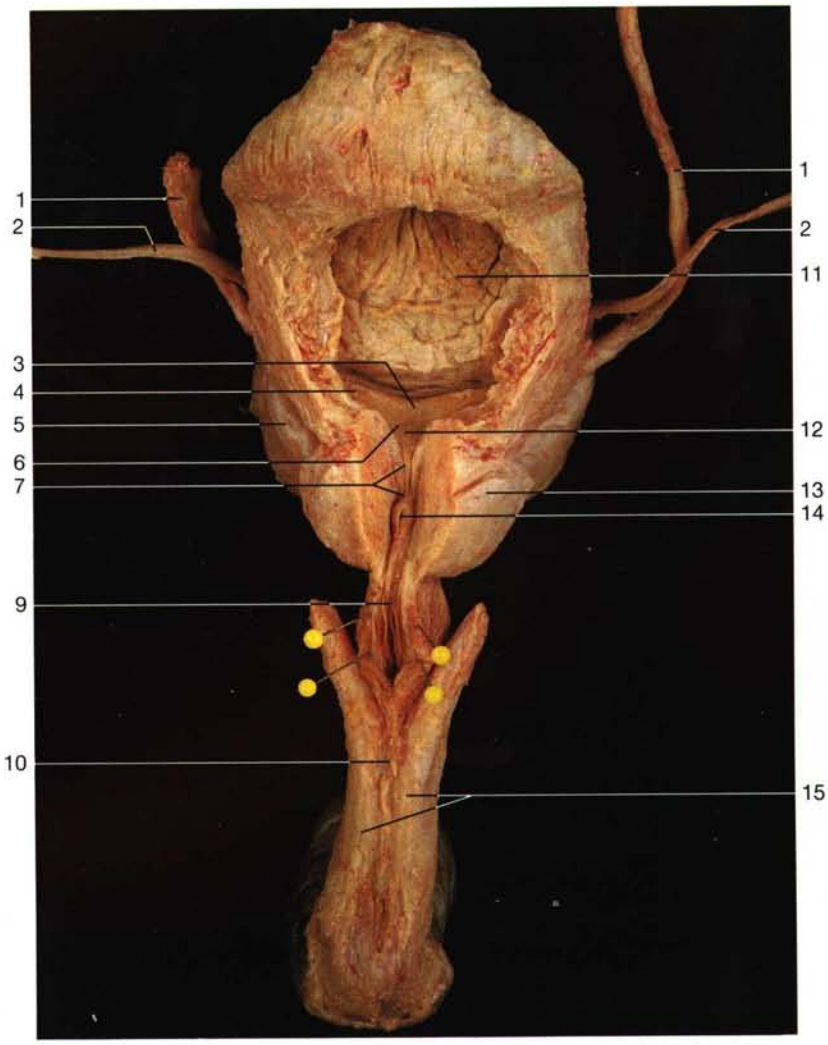

Male urogenital organs; isolated (anterior view). Urinary bladder, prostate and urethra have been opened. The urinary bladder is contracted.

1 Ureter
2 Ductus deferens
3 Interureteric fold
4 Ureteric orifice
5 Seminal vesicle
6 Trigone of bladder
7 Prostatic urethra with seminal colliculus and urethral crest
8 Deep transverse perineal muscle
9 Membranous urethra
10 **Spongy urethra**
11 Mucous membrane of **urinary bladder**
12 Internal urethral orifice and uvula of bladder
13 **Prostate**
14 Prostatic utricle
15 Right and left **corpus cavernosum of penis**
16 Ejaculatory duct
17 **Sphincter urethrae muscle**
18 **Median umbilical fold** with remnant of urachus
19 Medial umbilical fold with remnant of umbilical artery
20 **Urinary bladder**
21 Rectovesical pouch
22 Rectum
23 Sacrum
24 Deep iliac circumflex artery
25 Deep inguinal ring and ductus deferens
26 External iliac artery and vein
27 **Femoral nerve**
28 Obturator nerve
29 Ilium and sacrum
30 Inferior epigastric artery
31 Iliopsoas muscle

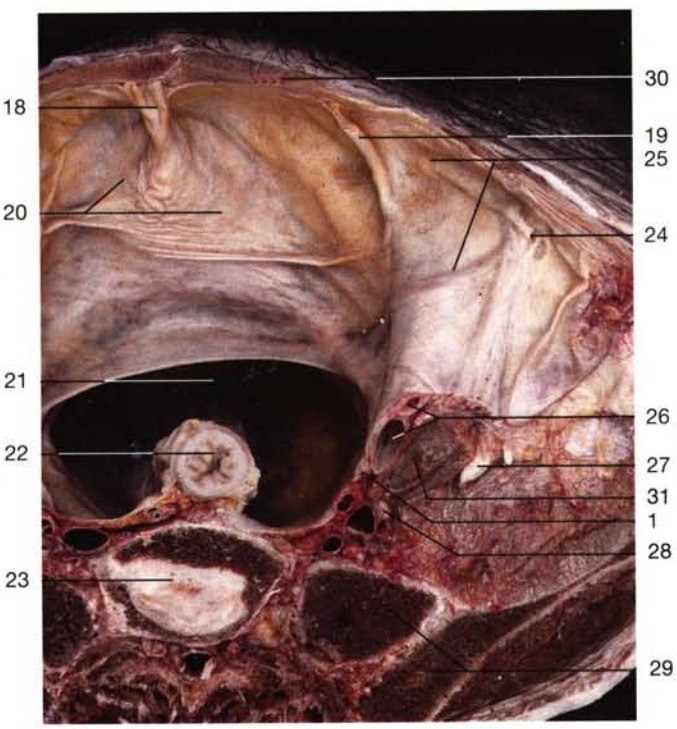

Pelvic cavity in the male (viewed from above).

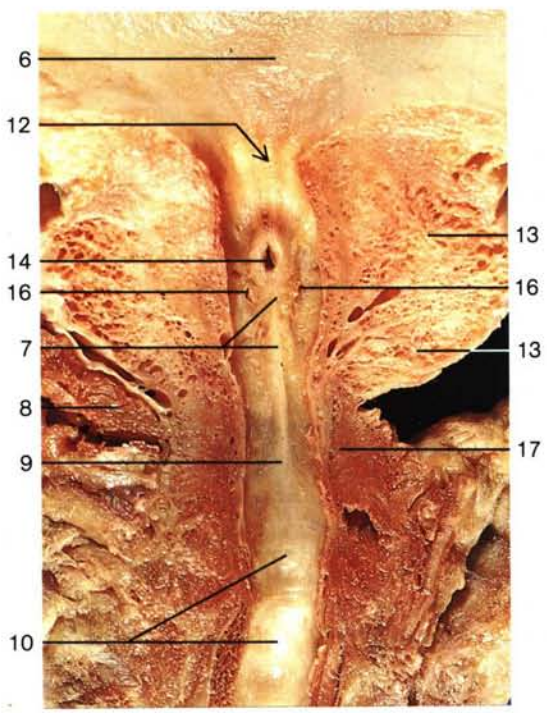

Posterior half of male urethra and prostate in continuity with neck of bladder (anterior view).

1 Apex of urinary bladder with urachus
2 **Urinary bladder**
3 Ureter
4 **Ductus deferens**
5 Ampulla of ductus deferens
6 **Seminal vesicle**
7 **Prostate**
8 **Bulbourethral or Cowper's gland**
9 **Bulb of penis**
10 Crus penis
11 **Corpus spongiosum of penis**
12 **Corpus cavernosum of penis**
13 Testis and epididymis with coverings
14 Glans penis
15 Fundus of bladder
16 **Head of epididymis**
17 **Testis**
18 Mucous membrane of bladder
19 Trigone of bladder
20 Ureteric orifice
21 Internal urethral orifice
22 Seminal colliculus
23 Prostate
24 Prostatic urethra
25 Membranous urethra
26 Spongy (penile) urethra
27 Skin of penis
28 Deep dorsal vein of penis (unpaired)
29 Dorsal artery of penis (paired)
30 **Tunica albuginea of corpora cavernosa**
31 Septum of penis
32 Deep artery of penis
33 Tunica albuginea of corpus spongiosum
34 Deep fascia of penis

Male genital organs, isolated (posterior view).

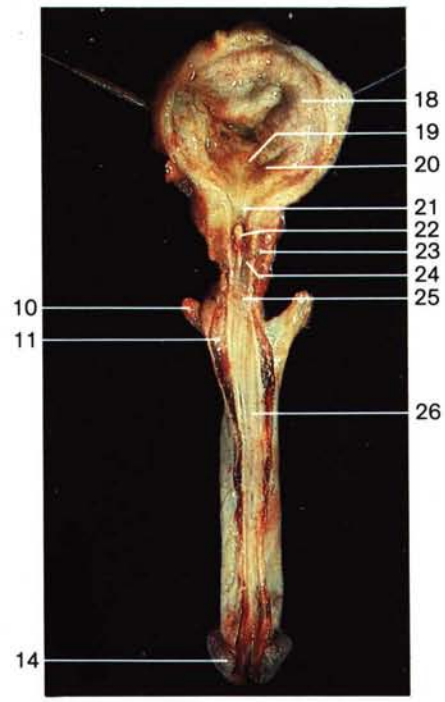

Urinary bladder, urethra and penis (anterior view, open longitudinally).

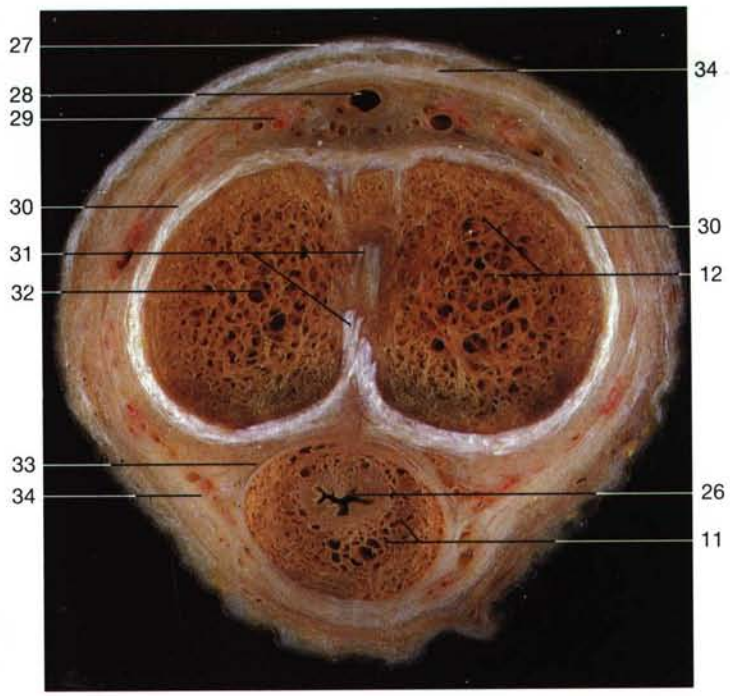

Cross-section of penis (anterior view).

1 **Corpus cavernosum of penis**
2 **Corpus spongiosum of penis**
3 Corona of glans penis
4 **Glans penis**
5 Suspensory ligament of penis
6 Inferior pubic ramus
7 Crus penis
8 **Bulb of penis**
9 Deep dorsal vein of penis
10 Septum pectiniforme
11 Dorsal artery of penis
12 Bulbourethral or Cowper's gland
13 Urinary bladder
14 **Seminal vesicle**
15 Ampulla of ductus deferens
16 Ductus deferens
17 Membranous urethra
18 **Prostate**
19 Ureter

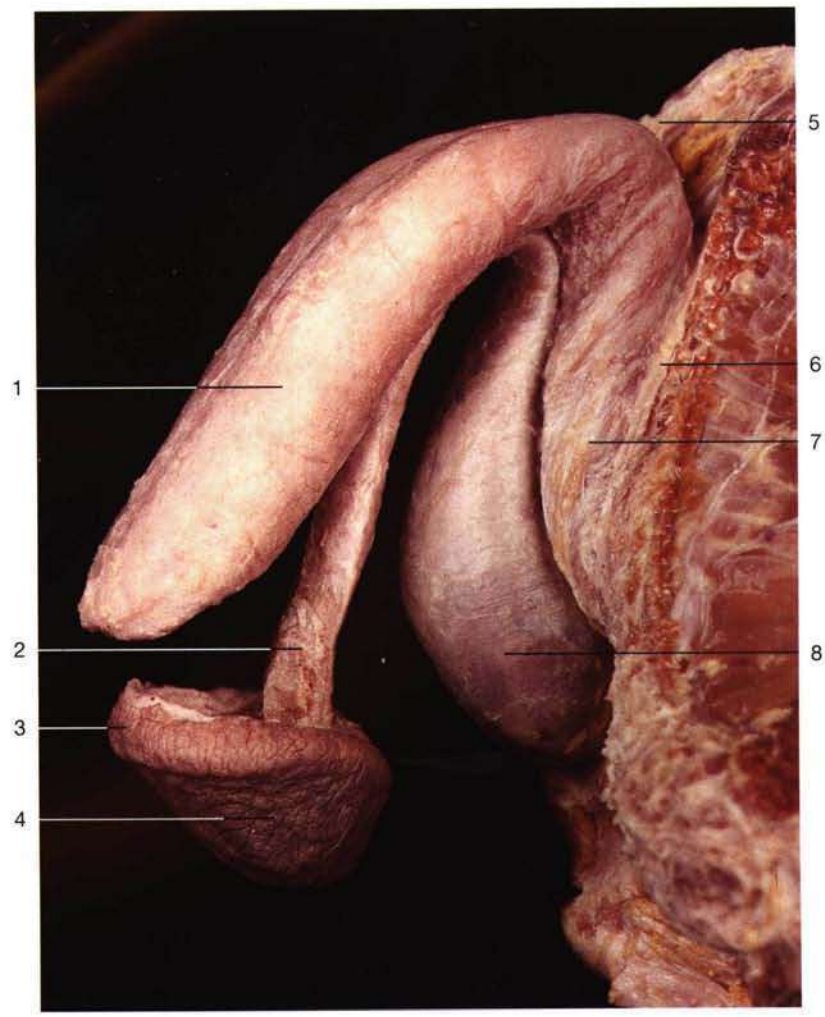

Male external genital organs (lateral view). The corpus spongiosum of the penis with the glans penis has been isolated and reflected.

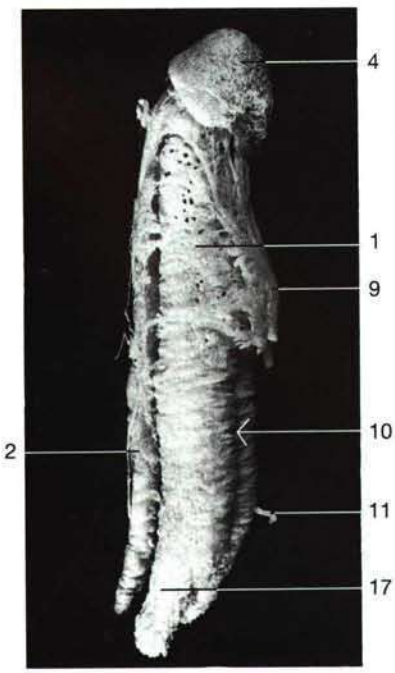

Resin cast of erected penis.

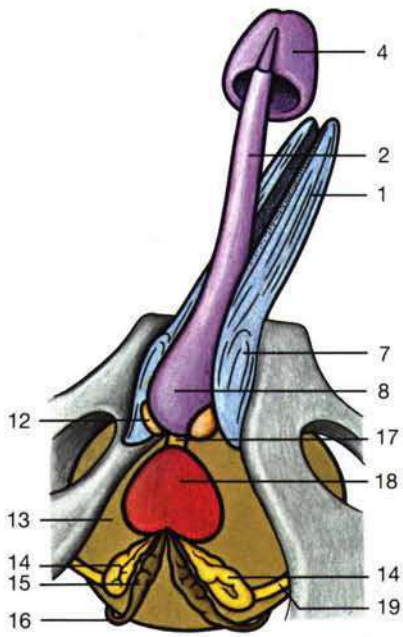

Male external genital organs and accessory glands. (Schematic drawing.)

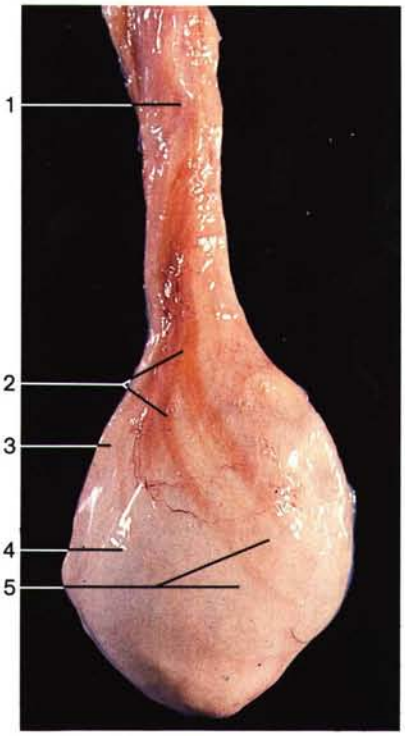

Testis and epididymis with investing layers (lateral view).

1 Spermatic cord covered with cremasteric fascia
2 **Cremaster muscle**
3 Position of epididymis
4 Internal spermatic fascia
5 Position of testis
6 Internal spermatic fascia with adjacent investing layers of testis (cut surface)

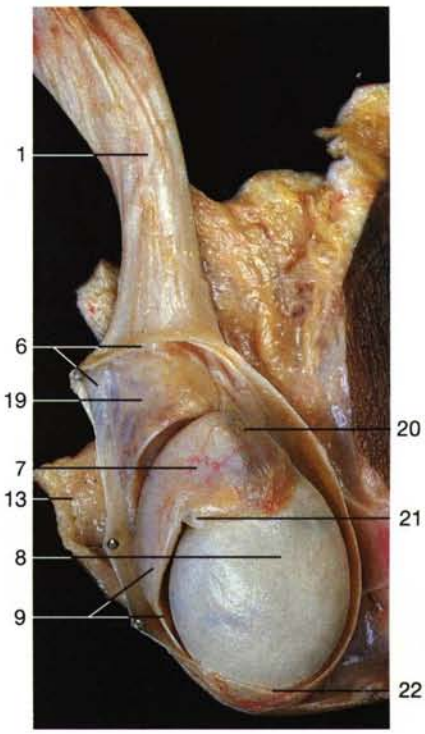

Testis and epididymis (lateral view). The tunica vaginalis has been opened.

7 **Head of epididymis**
8 **Testis with tunica vaginalis** (visceral layer)
9 Body of epididymis
10 Pampiniform venous plexus (anterior veins)
11 **Testicular artery**

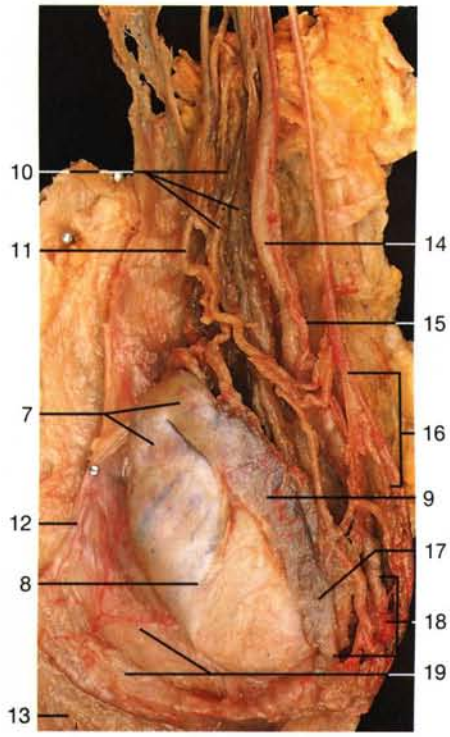

Testis, epididymis and spermatic cord. Dissection of spermatic cord and ductus deferens (left side, posterolateral aspect).

12 Tunica vaginalis (parietal layer, cut edge)
13 Skin and dartos muscle (reflected)
14 **Ductus deferens**
15 Artery of ductus deferens
16 Posterior veins of pampiniform plexus
17 Tail of epididymis
18 Transition of epididymal duct to ductus deferens and venous plexus
19 Parietal layer of tunica vaginalis
20 Appendix of epididymis
21 Appendix of testis
22 Gubernaculum testis

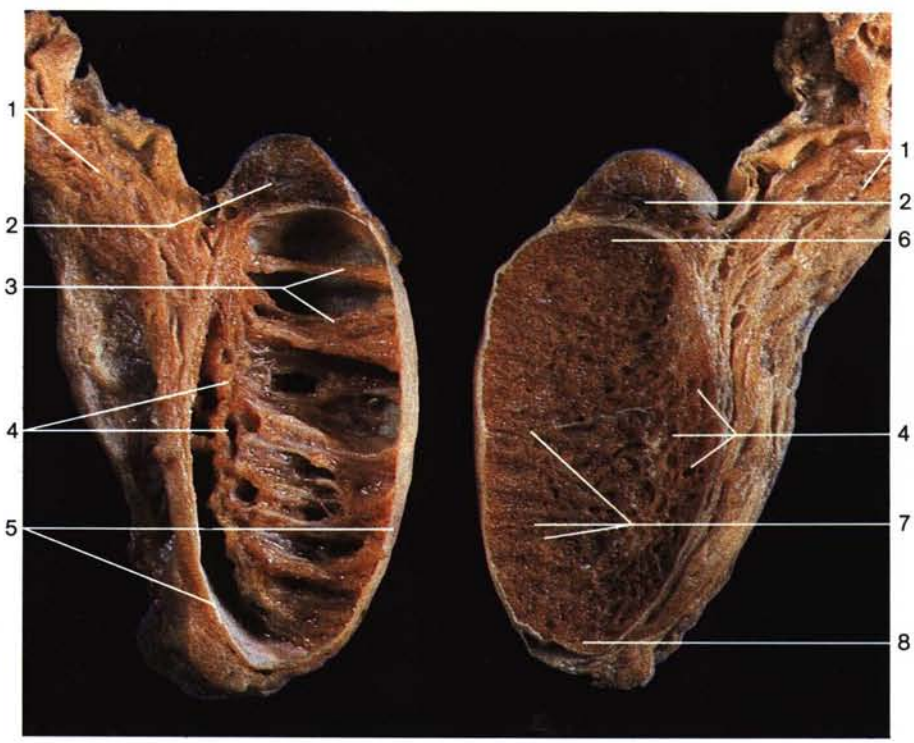

Longitudinal section through testis and epididymis. The left figure shows the testicular septa after removal of the seminiferous tubules.

1 Spermatic cord (cut surface)
2 Head of epididymis (cut surface)
3 Septa of testis
4 Mediastinum testis
5 Tunica albuginea
6 Superior pole of testis
7 Convoluted seminiferous tubules
8 Inferior pole of testis

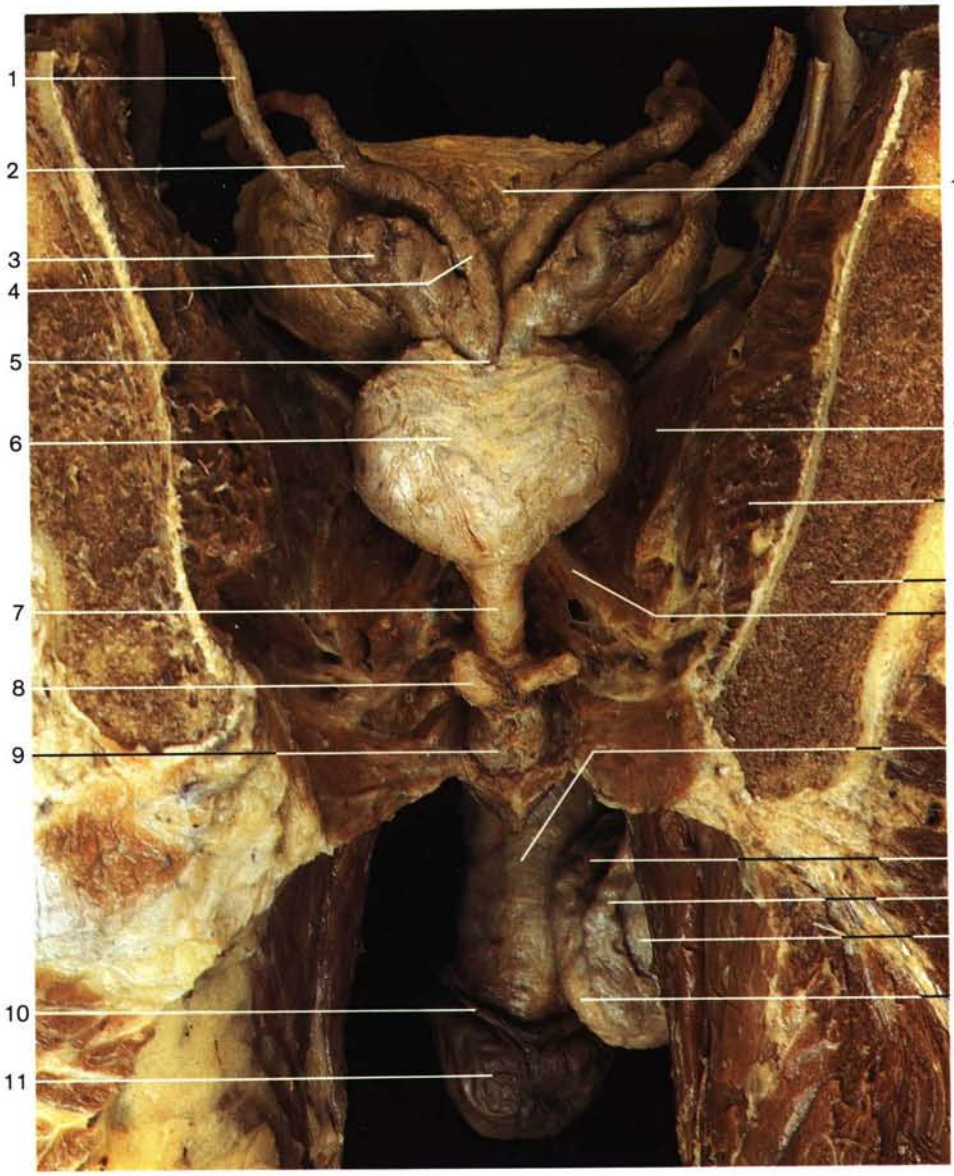

1 Ureter
2 **Ductus deferens**
3 **Seminal vesicle**
4 **Ampulla of ductus deferens**
5 Ejaculatory duct (proximal portion)
6 **Prostate**
7 Membranous urethra
8 **Bulbourethral or Cowper's gland**
9 Bulb of penis
10 Penis
11 Glans penis
12 **Urinary bladder**
13 Levator ani muscle
14 Obturator internus muscle
15 Pelvic bone (cut edge)
16 Puboprostatic ligament
17 **Corpus spongiosum of penis**
18 Head of epididymis
19 Beginning of ductus deferens
20 Testis
21 Tail of epididymis
22 Corpus cavernosum of penis
23 Spermatic cord
24 Pectineus and adductor muscles
25 Pubic bone
26 **Prostatic part of urethra** (seminal colliculus)
27 Rectum
28 Sciatic nerve
29 Great saphenous vein
30 Sartorius muscle
31 Femoral artery and vein
32 Rectus femoris muscle
33 Tensor fasciae latae muscle
34 Pectineus muscle
35 Iliopsoas muscle
36 Vastus lateralis muscle
37 Obturator externus muscle
38 Femur
39 Ischial tuberosity
40 Gluteus maximus muscle

Accessory glands of male genital organs in situ. Coronal section through the pelvic cavity. Posterior aspect of urinary bladder, prostate and seminal vesicles.

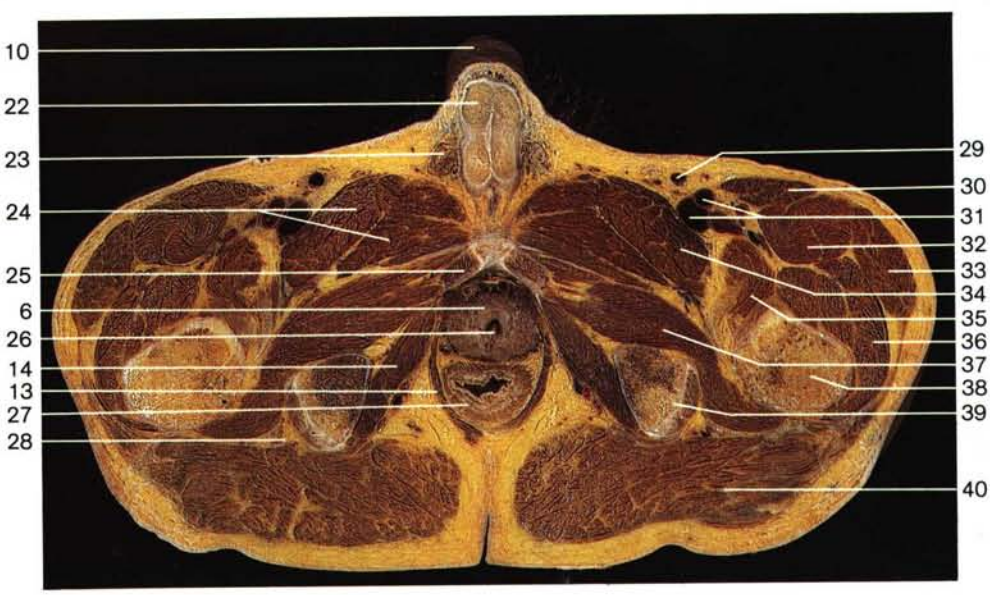

Horizontal section through pelvic cavity at level of prostate.

Coronal section through pelvic cavity at the level of prostate and hip joint (anterior aspect).

1 Acetabulum of hip joint
2 **Urinary bladder**
3 Head of femur
4 Internal urethral orifice
5 **Prostate**
6 **Seminal colliculus**
7 Obturator internus muscle
8 Ischiorectal fossa
9 Membranous urethra
10 Deep transverse perineus muscle
11 **Crus penis** and **ischiocavernosus muscle**
12 Prostatic part of urethra
13 Prostatic plexus
14 Levator ani muscle
15 Obturator externus muscle
16 **Bulb of penis**
17 Ampulla of rectum
18 **Anal canal**
19 Seminal vesicle
20 Internal anal sphincter muscle
21 External anal sphincter muscle
22 Anus
23 Psoas major muscle
24 Intervertebral disc
25 Ilium
26 Ligament of the head of the femur
27 Sacral promontory

Coronal section through pelvic cavity. (MR-Scan.)

Coronal section through anal canal.

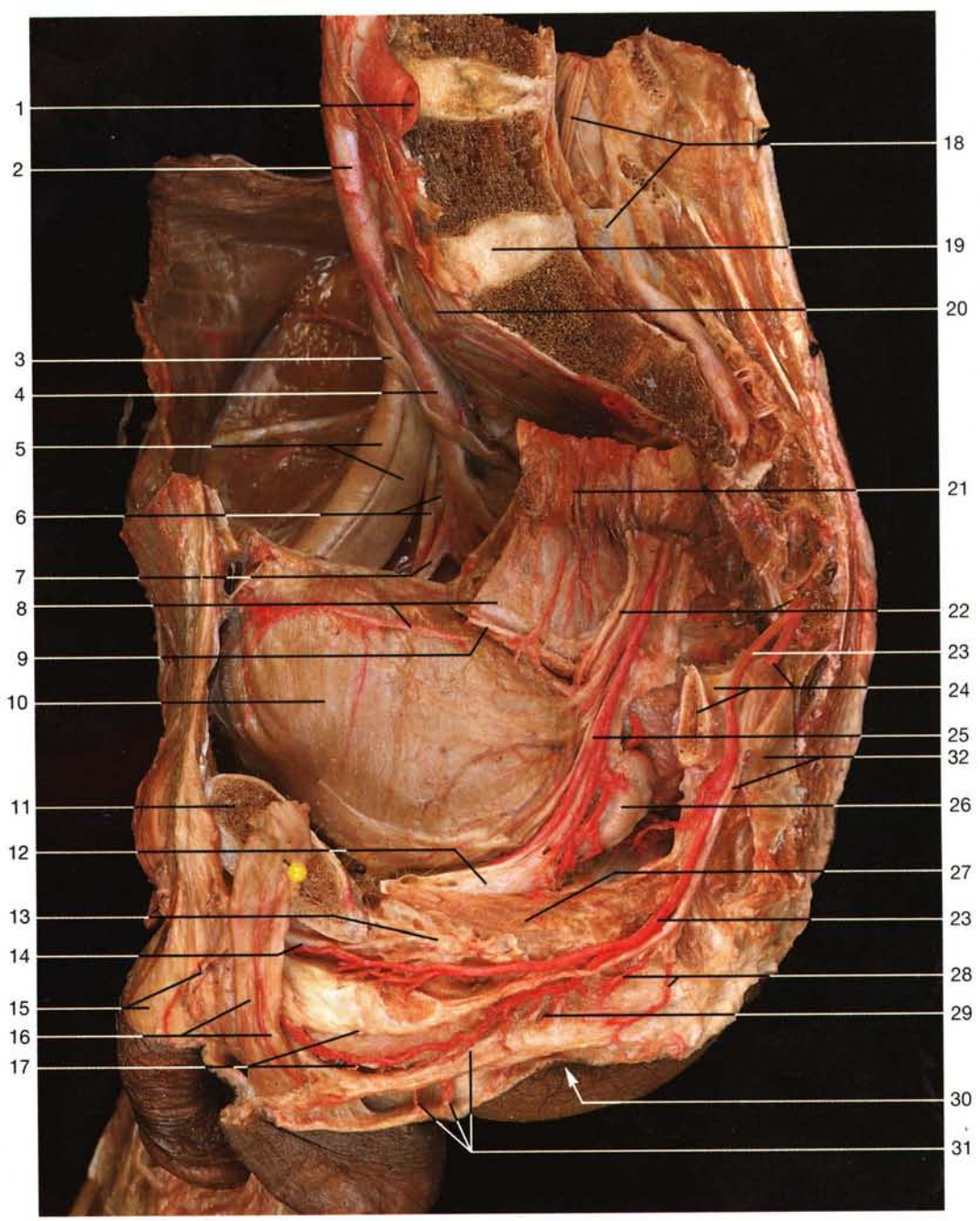

Pelvic cavity in the male (right half of parasagittal section). The arteries have been injected with red resin. The parietal layer of peritoneum has been removed. The urinary bladder is filled to a great extent.

1 Left common iliac artery
2 Right common iliac artery
3 Right **ureter**
4 Right **internal iliac artery**
5 Right external iliac artery and vein
6 Right **obturator artery and nerve**
7 Umbilical artery
8 Sigmoid and superior vesical artery
9 Left ductus deferens
10 **Urinary bladder**
11 Pubic bone (cut)
12 **Prostate**
13 Vesicoprostatic venous plexus
14 Deep dorsal vein of penis and dorsal artery of penis
15 Penis and superficial dorsal vein
16 **Spermatic cord and testicular artery**
17 **Bulb of penis** and deep artery of penis
18 Cauda equina and dura mater (divided)
19 Intervertebral disc between fifth lumbar vertebra and sacrum
20 **Sacral promontory**
21 Mesosigmoid
22 Left ureter
23 Left internal pudendal artery
24 Ischial spine (cut), sacrospinal ligament, inferior gluteal artery
25 Left **inferior vesical artery**
26 **Seminal vesicle**
27 Levator ani muscle
28 Branches of inferior rectal artery
29 Perineal artery
30 Anus
31 Posterior scrotal branches
32 **Pudendal nerve** and sacrotuberal ligament

Vessels of pelvic cavity in the male (midsagittal section, right side, medial aspect). Urinary bladder and rectum have been partly removed; the hypogastric plexus is displayed.

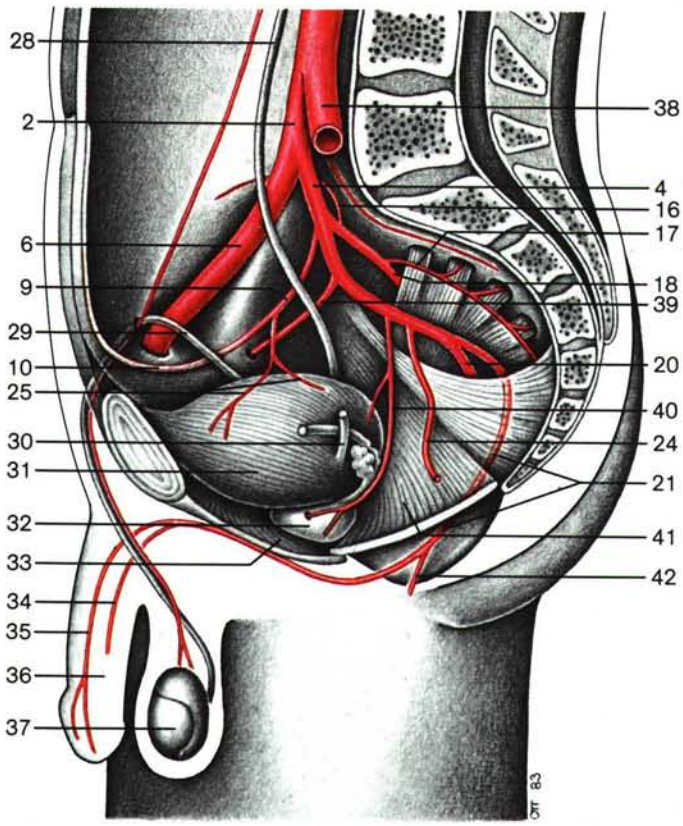

Main branches of internal iliac artery in the male. (Schematic drawing.)

1 Intervertebral disc between 5th lumbar vertebra and sacrum
2 Right common iliac artery
3 Sacral promontory
4 Right **internal iliac artery**
5 Femoral nerve
6 Right external iliac artery
7 Psoas major muscle
8 Right external iliac vein
9 **Umbilical artery**
10 Medial umbilical ligament with obliterated umbilical artery
11 Obturator nerve
12 Superior hypogastric plexus
13 Sacral spinal nerve (of sacral plexus)
14 **Inferior hypogastric plexus**
15 Rectum
16 Iliolumbar artery
17 Lateral sacral artery
18 Superior gluteal artery and sacral spinal nerve
19 Pudendal and coccygeal plexus
20 Inferior gluteal artery
21 **Internal pudendal artery**
22 Coccygeus muscle overlying sacrospinous ligament
23 **Anal canal**
24 Middle rectal artery
25 **Superior vesical artery** and branch to the ductus deferens
26 Pubic bone (cut)
27 Bulbospongiosus muscle
28 Right ureter
29 Right ductus deferens
30 Left ureter
31 **Urinary bladder**
32 **Prostate**
33 Urogenital diaphragm
34 Deep artery of penis
35 Dorsal artery of penis
36 Penis
37 Testis
38 Left common iliac artery
39 **Obturator artery**
40 **Inferior vesical artery**
41 Levator ani muscle
42 Inferior rectal artery

Male external genital organs with penis, testis, and spermatic cord, superficial layers (anterior view).

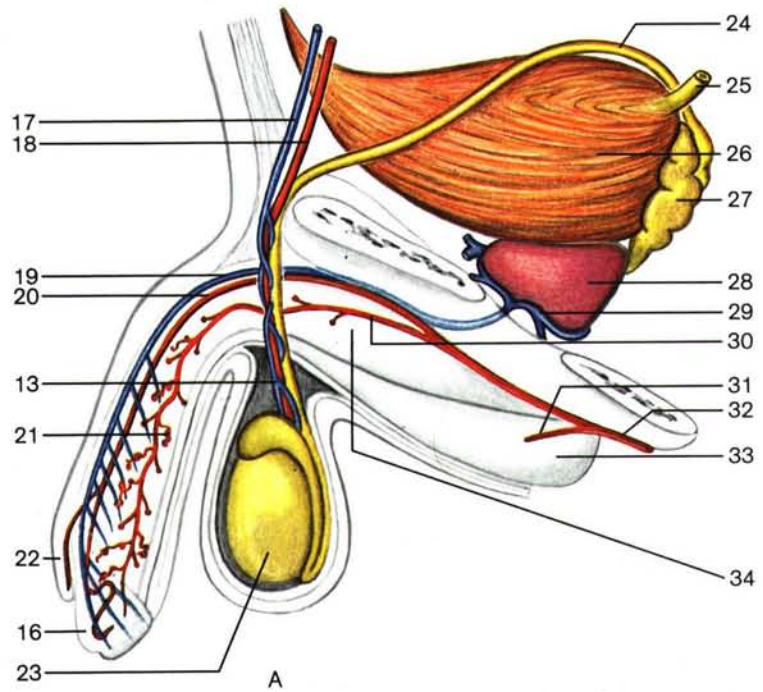

Vessels of male genital organs. (Schematic drawing.)
A = lateral aspect; B = cross-section of penis.

1 Femoral nerve
2 Femoral artery and vein
3 Femoral branch of genitofemoral nerve
4 **Spermatic cord** with genital branch of genitofemoral nerve
5 **Penis** with deep fascia
6 Great saphenous vein
7 Cremaster muscle
8 **Testis with cremaster muscle**
9 Superficial inguinal ring
10 Internal spermatic fascia (cut edge)
11 Ilioinguinal nerve
12 Left spermatic cord
13 Pampiniform venous plexus
14 External spermatic fascia
15 Superficial dorsal vein of penis
16 **Glans penis**

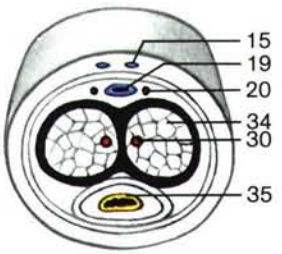

Male external genital organs with penis, testis and spermatic cord, deeper layers (ventral aspect). The deep fascia of the penis has been opened to display the dorsal nerves and vessels.

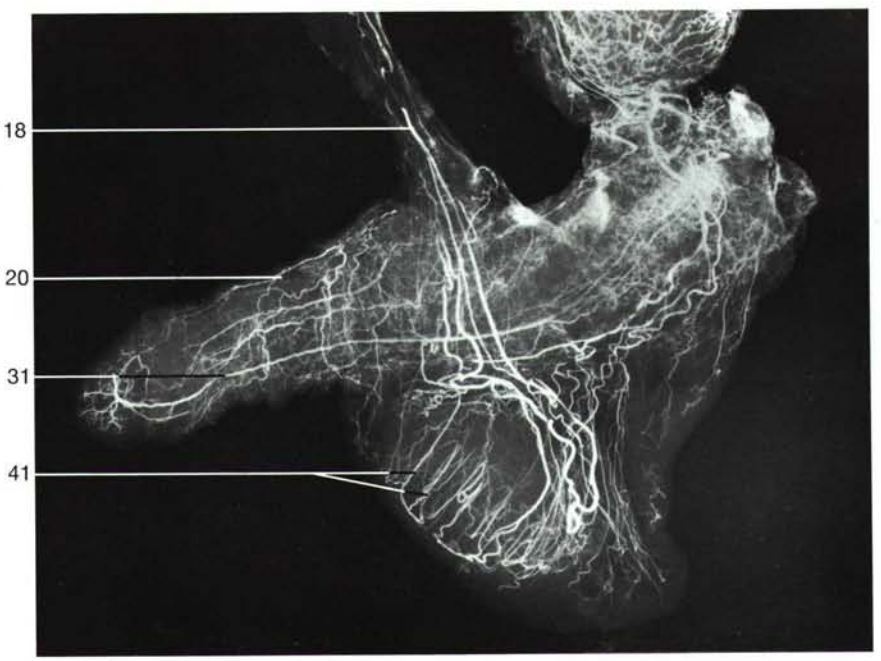

Male genital organs (lateral aspect). (Arteriogram.)

17 Testicular vein
18 Testicular artery
19 Deep dorsal vein of penis
20 **Dorsal artery of penis**
21 **Helicine arteries**
22 Prepuce
23 **Testis** with **tunica albuginea**
24 Ductus deferens
25 Ureter
26 Urinary bladder
27 **Seminal vesicle**
28 **Prostate**
29 Vesicoprostatic venous plexus
30 **Deep artery of penis**
31 Artery of bulb of penis
32 Internal pudendal artery
33 Corpus spongiosum of penis
34 Corpus cavernosum of penis
35 Urethra
36 Cremasteric fascia with cremaster muscle
37 Dorsal nerve of penis
38 **Epididymis**
39 **Tunica vaginalis** (visceral layer)
40 Tunica vaginalis (parietal layer)
41 Testis with vascular loops

1 Glans penis
2 **Corpus spongiosum of penis**
3 **Corpus cavernosum of penis**
4 Gracilis muscle
5 Adductor muscles
6 Ischiocavernosus muscle overlying crus of penis
7 Perineal body
8 Gluteus maximus muscle
9 Coccyx
10 **Bulbospongiosus** muscle
11 Deep transverse perineus muscle covered by inferior fascia of urogenital diaphragm
12 Superficial transverse perineus muscle
13 Anus
14 **External anal sphincter muscle**
15 **Levator ani muscle**
16 Anococcygeal ligament
17 Obturator internus muscle
18 Urethra
19 **Deep transverse perineus muscle**

Muscles of urogenital and pelvic diaphragms in the male (from below).

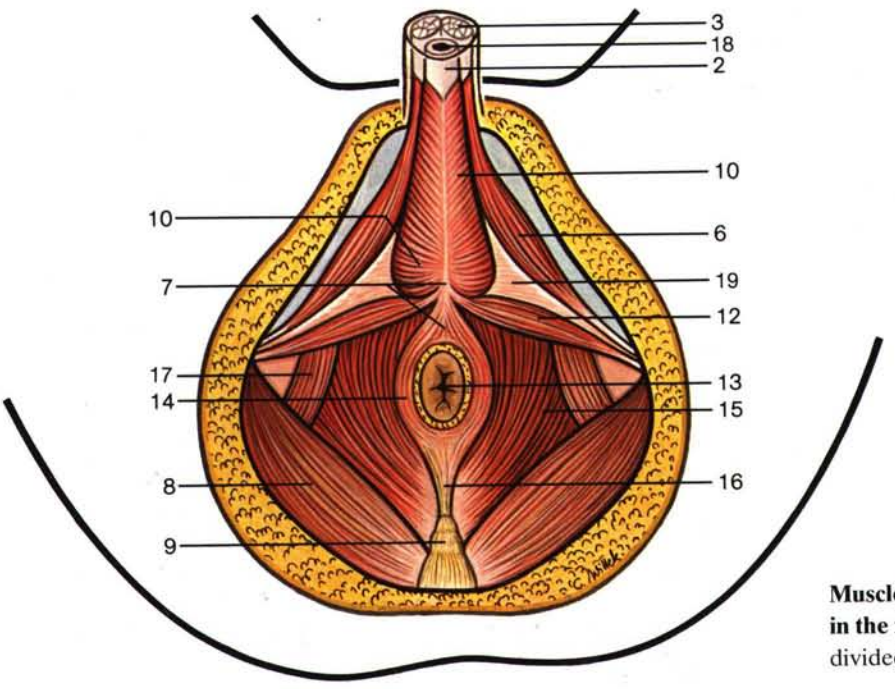

Muscles of urogenital and pelvic diaphragms in the male (from below). The penis has been divided. (Schematic drawing.)

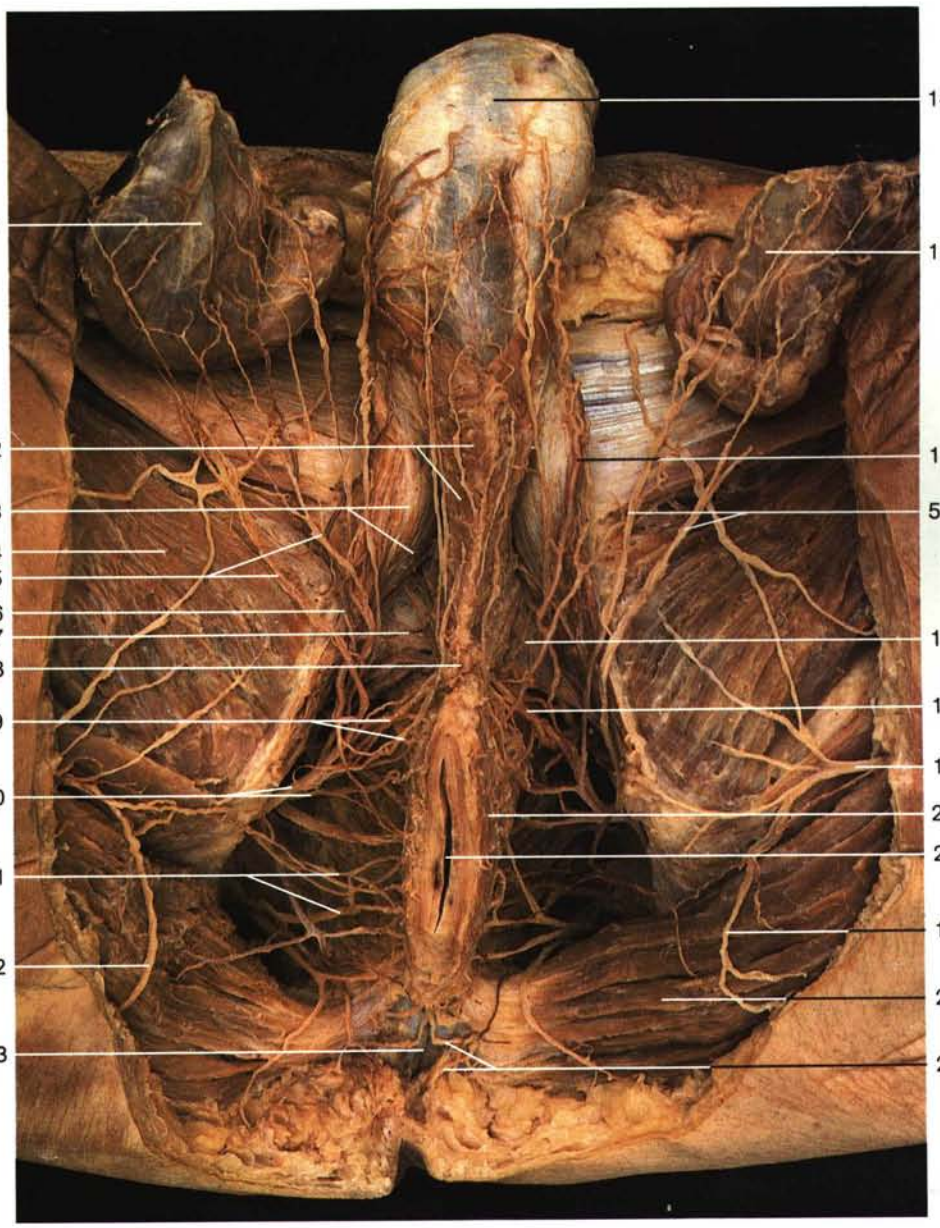

1 Right testis (reflected laterally and upward)
2 Bulbospongiosus muscle
3 Ischiocavernosus muscle
4 Adductor magnus muscle
5 **Posterior scrotal nerves** and superficial perineal arteries
6 Posterior scrotal artery and vein
7 Right artery of bulb of penis
8 Perineal body
9 Perineal branches of pudendal nerve
10 **Pudendal nerve** and **internal pudendal artery**
11 **Inferior rectal arteries** and **nerves**
12 Inferior cluneal nerve
13 Coccyx (location)
14 Penis
15 Left testis (reflected laterally)
16 Left posterior scrotal artery
17 **Deep transverse perineal muscle**
18 Left artery of bulb of penis
19 **Posterior femoral cutaneous nerve**
20 External anal sphincter muscle
21 **Anus**
22 Gluteus maximus muscle
23 Anococcygeal nerves
24 Acetabulum (femur removed)
25 Ligament of femoral head
26 Body of ischium (cut)
27 Sciatic nerve
28 Coccygeus muscle
29 **Levator ani muscle**
a iliococcygeus muscle
b pubococcygeus muscle
c puborectalis muscle
30 Prostatic venous plexus
31 Body of pubis
32 Testis

Urogenital diaphragm and external genital organs in the male with vessels and nerves (from below). The testes have been reflected laterally.

Pelvic diaphragm and external genital organs in the male. The right half of the pelvis including obturator internus muscle and femur had been removed to display the right half of levator ani muscle.

1 Right testis (reflected)
2 Posterior scrotal nerves
3 Left crus penis with ischiocavernosus muscle
4 **Anus**
5 Inferior cluneal nerves
6 **Penis**
7 Left testis (reflected)
8 Dorsal artery and nerve of penis
9 Urethra
10 **Deep transverse perineus muscle**
11 **Perineal branch of pudendal nerve**
12 Artery of bulb of penis (reflected)
13 Branch of posterior femoral cutaneous nerve
14 **Internal pudendal artery** and **pudendal nerve**
15 Inferior rectal arteries and nerves
16 Gluteus maximus muscle
17 Dorsal nerve of penis
18 Posterior femoral cutaneous nerve
19 Perineal branches of pudendal nerve
20 Inferior rectal nerves
21 **Bulbocavernosus muscle** (inside: dorsal artery of penis)
22 Perineal artery
23 **External anal sphincter muscle**
24 Inferior rectal artery and veins

Urogenital diaphragm and external genital organs in the male (from below). The left crus penis has been isolated and reflected laterally together with the bulb of the penis. The urethra has been cut.

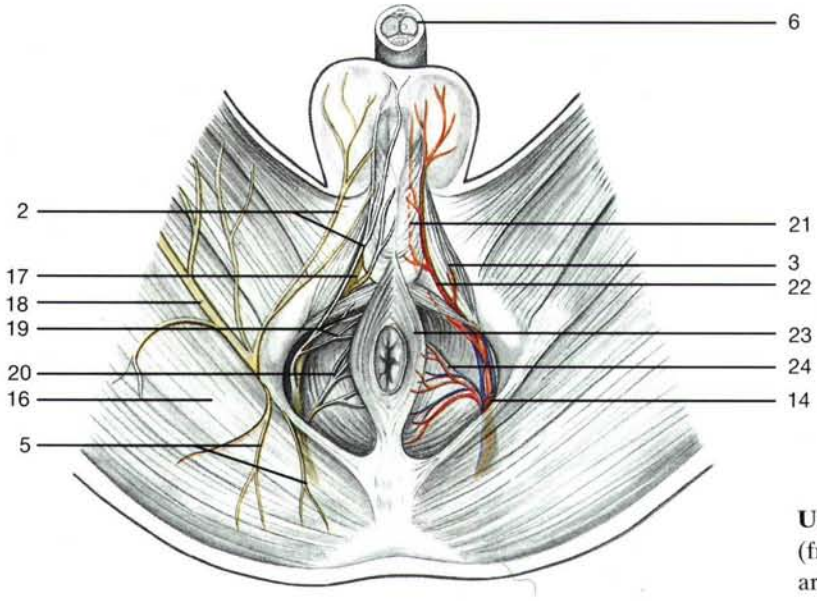

Urogenital and anal region in the male (from below). Right side: nerves; left side: arteries and veins.

1 Right testis (reflected)
2 **Corpus spongiosum of penis**
3 **Corpus cavernosum of penis**
4 Perineal branch of posterior femoral cutaneous nerve
5 Posterior scrotal arteries and nerves
6 Deep artery of penis
7 Deep transverse perineal muscle
8 Right **perineal nerves**
9 **Inferior rectal nerves**
10 Inferior cluneal nerve
11 Anococcygeal nerves
12 Left spermatic cord
13 Left testis (cut surface)
14 Dorsal artery and nerve of penis
15 Deep dorsal vein of penis
16 **Urethra** (cut)
17 Artery of bulb of penis
18 Superficial transverse perineus muscle
19 Left artery of bulb of penis
20 Perineal branch of pudendal nerve
21 **Anus**
22 **External anal sphincter muscle**
23 Gluteus maximus muscle
24 Internal pudendal artery and pudendal nerve
25 Sacrotuberous ligament
26 Coccyx
27 Urogenital diaphragm (deep transverse perineus muscle)
28 Tendinous center of perineum (perineal body)
29 **Levator ani muscle**
30 Anococcygeal ligament
31 Obturator internus muscle
32 Dorsal artery of penis

Anal and urogenital region in the male (from below). The root of penis has been cut. Dissection of the urogenital diaphragm.

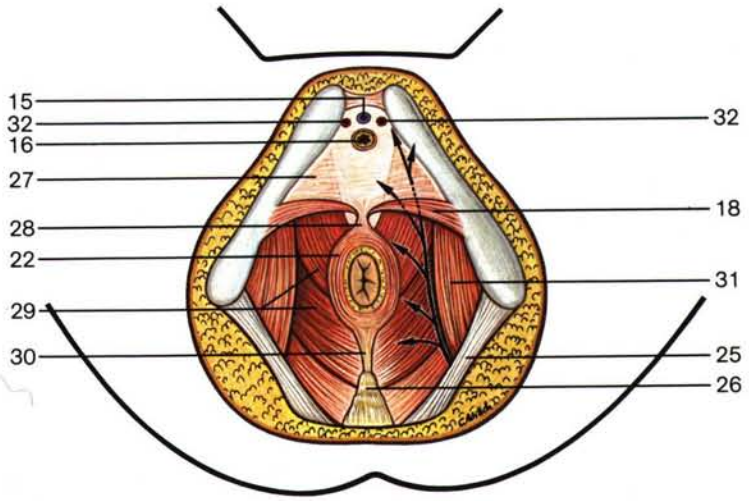

Urogenital and pelvic diaphragms in the male (from below). The penis has been removed. The arrows indicate the course of vessels and nerves. (Schematic drawing.)

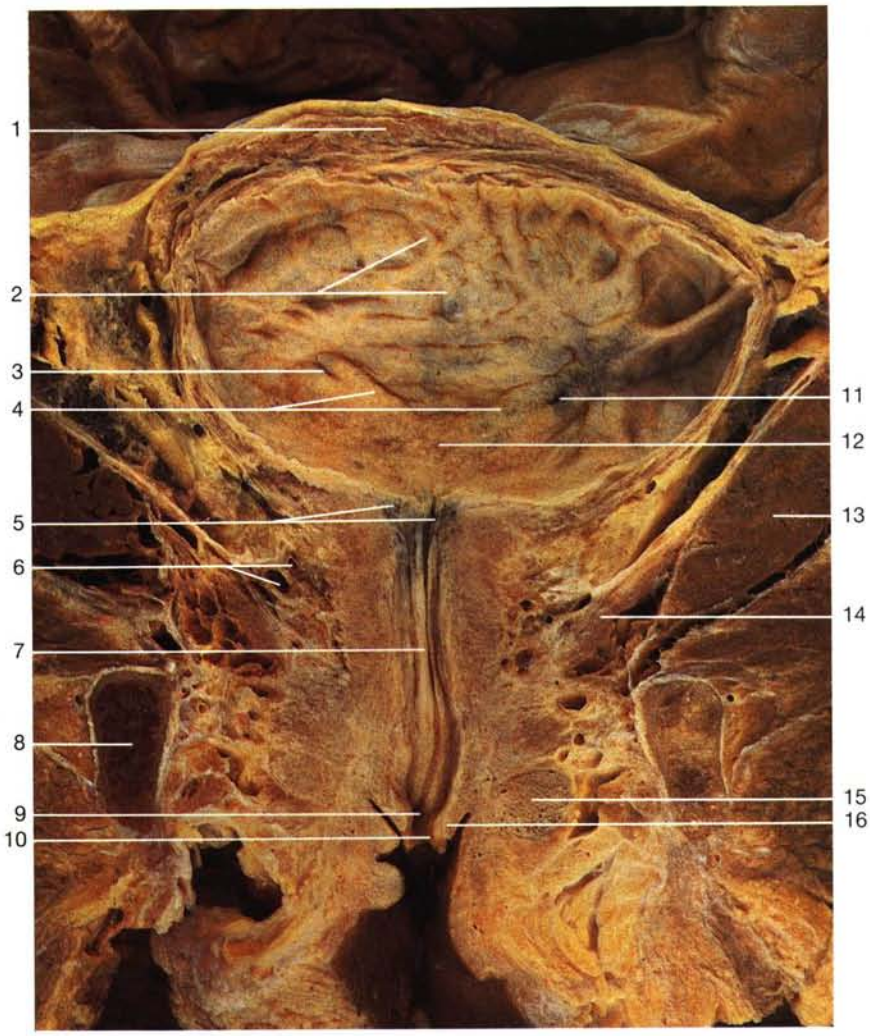

Coronal section through the female urinary bladder and urethra (anterior view).

1 Muscular coat of **urinary bladder**
2 Folds of mucous membrane of urinary bladder
3 Right ureteric orifice
4 Interureteric fold
5 Internal urethral orifice
6 Vesicouterine venous plexus
7 **Urethra**
8 Pubic bone (cut edge)
9 External urethral orifice
10 Vestibule of vagina
11 Left ureteric orifice
12 Trigone of bladder
13 Obturator internus muscle
14 Levator ani muscle
15 Bulb of the vestibule
16 Left labium minus
17 **Uterine tube**
18 Mesosalpinx
19 **Ovary**
20 Sigmoid colon
21 Saphenous opening
22 Urinary bladder
23 Vesicouterine pouch
24 **Fundus of uterus**
25 Rectouterine pouch (of Douglas)
26 Ampulla of rectum
27 Kidney
28 Abdominal part of ureter
29 Pelvic part of ureter
30 Anal canal
31 Perineum (perineal body)
32 Umbilicus
33 Infundibulum of uterine tube
34 Vaginal portion of cervix of uterus
35 Vagina
36 Pubic symphysis
37 Clitoris
38 Deep transverse perineus muscle

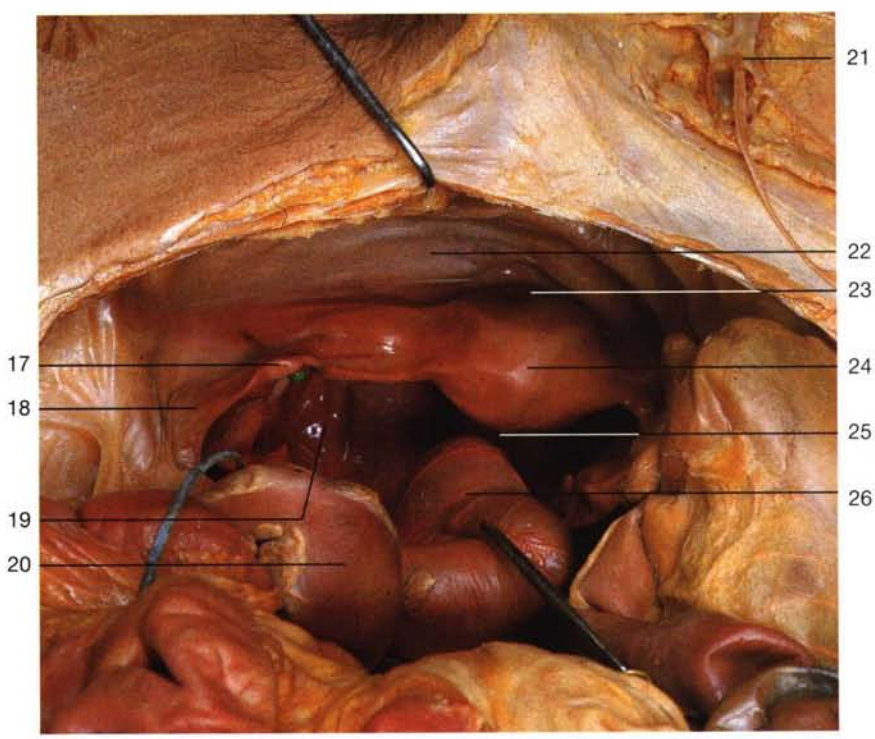

Femal internal genital organs. Pelvic cavity (from above).

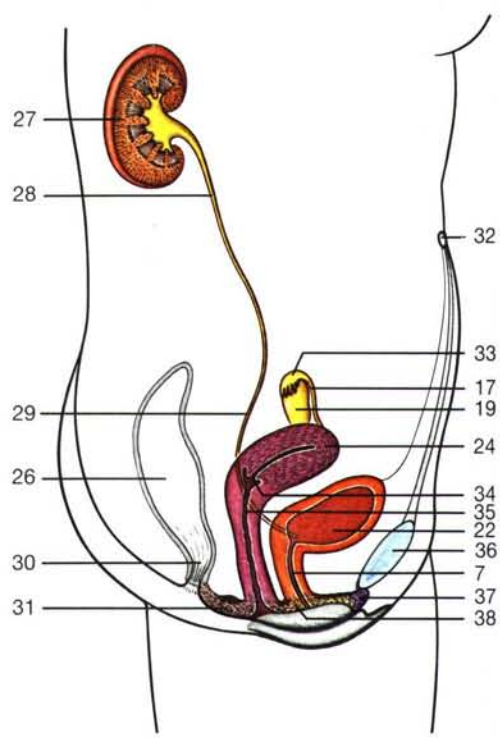

Female urogenital system (midsagittal section). (Schematic drawing.)

Female internal genital organs. Pelvic cavity, seen from above. The uterus has been reflected to the right.

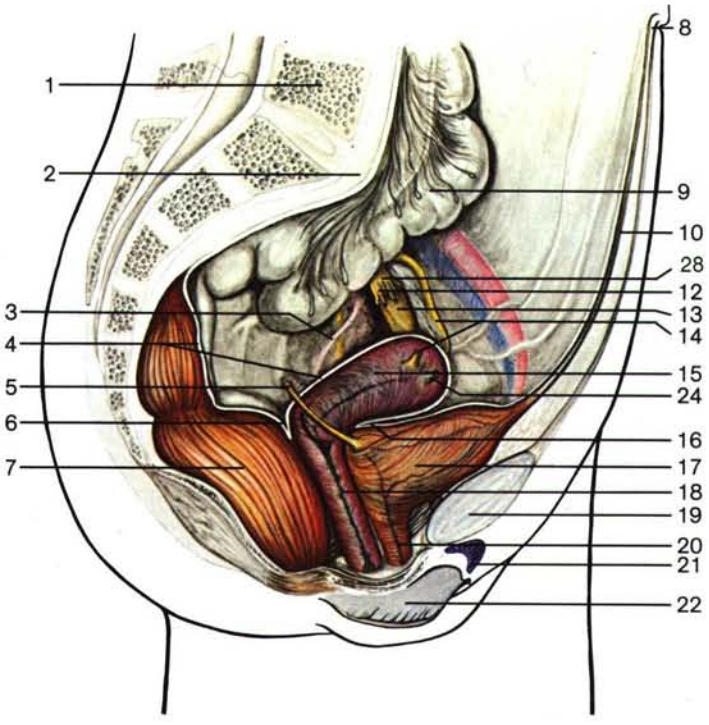

Regional relations of female internal genital organs (medial aspect). (Schematic drawing.)

1 Body of 5th lumbar vertebra
2 Sacral promontory
3 Left ureter
4 Peritoneum (cut edge)
5 Right ureter (divided)
6 **Recto-uterine pouch** (of Douglas)
7 Rectum
8 Umbilicus
9 Sigmoid colon
10 Median umbilical fold with urachus
11 Ampulla of uterine tube
12 Fimbriae of uterine tube
13 **Ovary**
14 **Uterine tube** (isthmus)
15 **Uterus**
16 **Vesico-uterine pouch**
17 Urinary bladder
18 Vagina
19 Pubic symphysis
20 Urethra
21 Clitoris
22 Labium minus
23 Insertion of uterine tube at fundus of uterus
24 Round ligament of uterus
25 Ligament of the ovary
26 Suspensory ligament of ovary
27 Right common iliac artery (covered by peritoneum)
28 Infundibulum of uterine tube

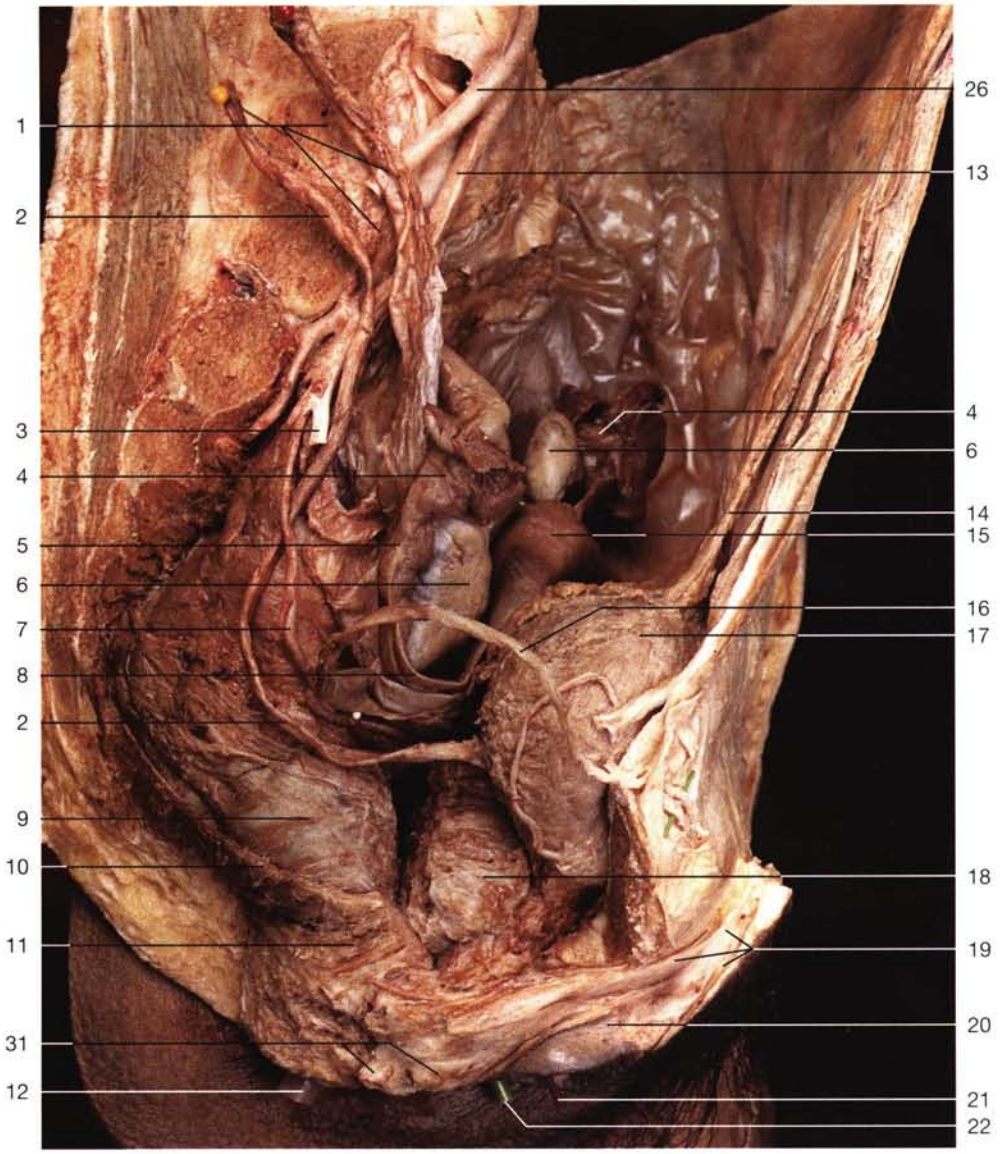

Female internal genital organs in situ. Right half of the pelvis and sacrum have been removed.

1 Body of 5th lumbar vertebra, **suspensory ligament of ovary** and sacral promontory
2 Ureter
3 Medial umbilical ligament (remnant of umbilical artery) (cut)
4 Infundibulum of uterine tube
5 Ampulla of uterine tube
6 **Ovary**
7 **Uterine artery**
8 **Uterine tube**
9 Rectum
10 Levator ani muscle (pelvic diaphragm – cut edge)
11 External anal sphincter muscle
12 Anus (probe)
13 Internal iliac artery
14 Remnant of urachus (median umbilical ligament)
15 **Uterus**
16 Round ligament of uterus
17 **Urinary bladder**
18 **Vagina**
19 **Clitoris**
20 Labium minus
21 External orifice of urethra (red probe)
22 Vaginal orifice (green probe)
23 Lateral umbilical ligament
24 Inferior epigastric artery
25 Obturator artery, vein and nerve
26 External iliac artery
27 **Recto-uterine pouch** (of Douglas)
28 Recto-uterine fold
29 Vesico-uterine pouch
30 Suspensory ligament of ovary
31 Greater vestibular gland and bulb of the vestibule

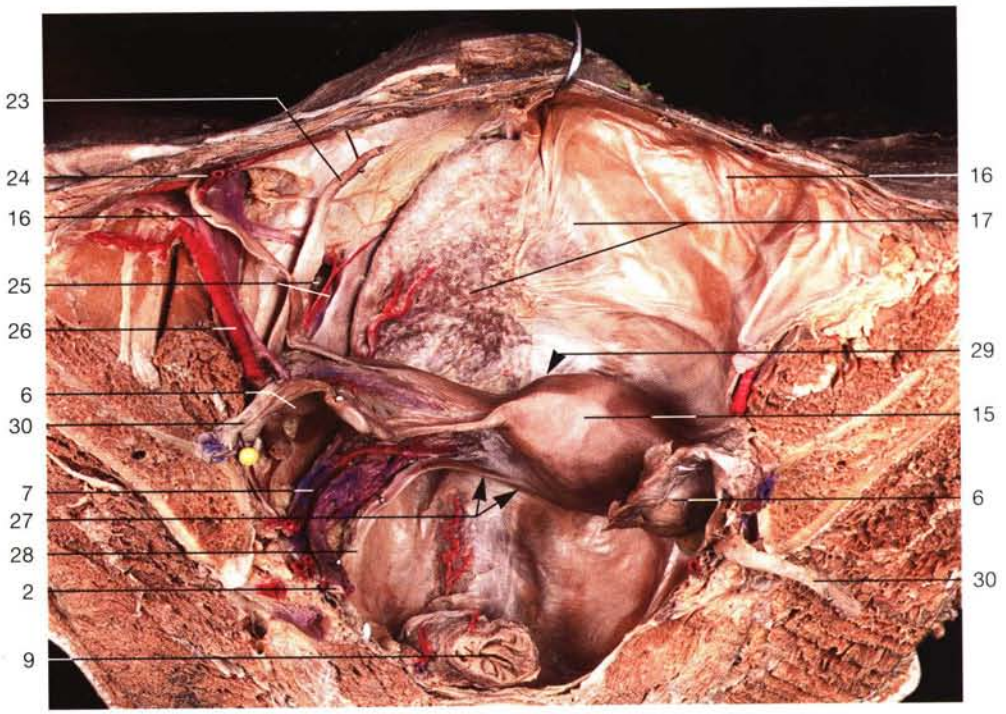

Female internal genital organs in situ (seen from above).
The peritoneum at the left half of pelvic cavity has been removed to display uterine tube, vessels, and nerves.

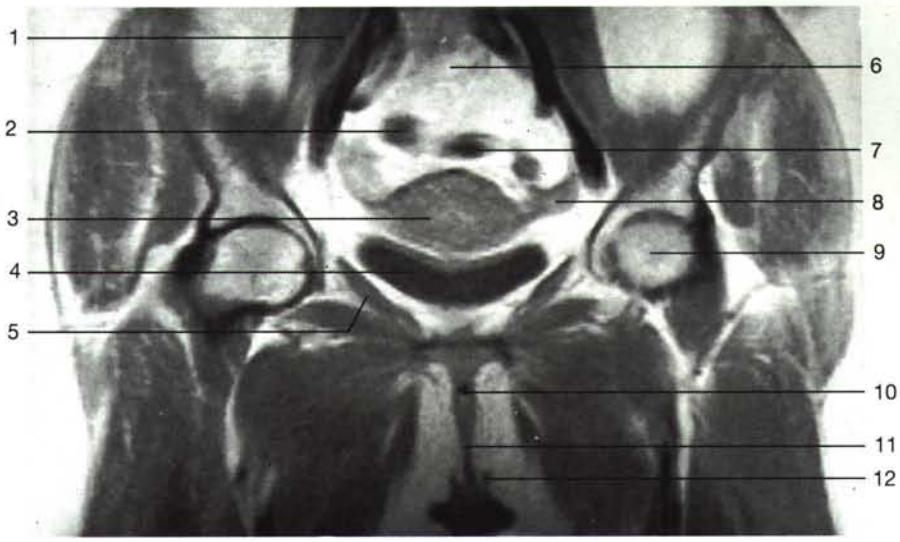

Coronal section through the pelvic cavity of the female. (MR-Scan.)

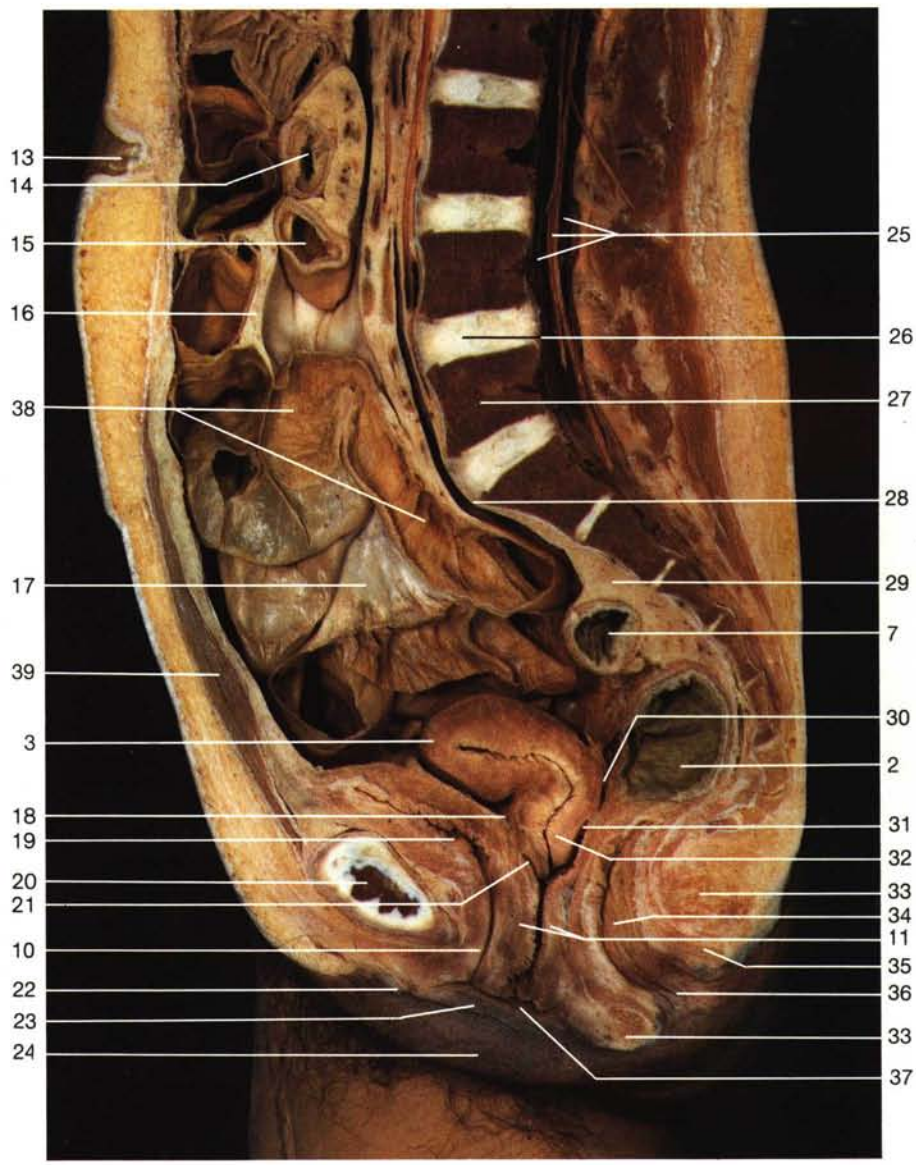

Midsagittal section through the female trunk. The urinary bladder is empty, the position and shape of the uterus are normal.

1 Psoas major muscle
2 Ampulla of rectum
3 **Uterus**
4 **Urinary bladder**
5 Obturator internus muscle
6 Promontory
7 Sigmoid colon
8 **Uterine tube**
9 Head of femur
10 Urethra
11 Vagina
12 Labium minus
13 Umbilicus
14 Duodenum
15 Ascending part of duodenum
16 Root of mesentery
17 Mesentery
18 Vesico-uterine pouch
19 **Urinary bladder** (collapsed)
20 Pubic symphysis
21 Anterior fornix of vagina
22 Clitoris
23 Labium minus
24 Labium majus
25 Vertebral canal with cauda equina
26 Intervertebral disc
27 Body of 5th lumbar vertebra
28 Sacral promontory
29 Mesosigmoid
30 **Recto-uterine pouch** (of Douglas)
31 Posterior fornix of vagina
32 Cervix of uterus
33 External anal sphincter muscle
34 Anal canal
35 Internal anal sphincter muscle
36 Anus
37 Hymen
38 Small intestine
39 Rectus abdominis muscle

Female genital organs, isolated (anterior view). The anterior wall of the vagina has been opened to display the vaginal portion of the cervix.

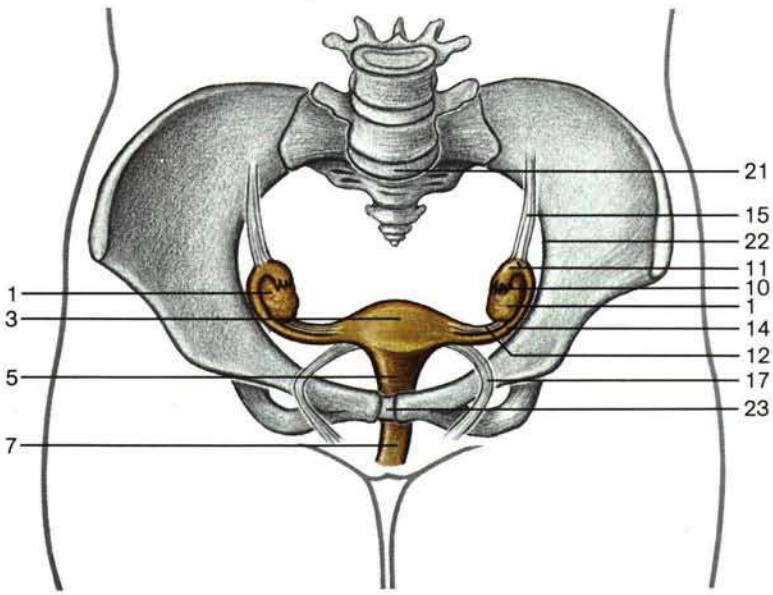

Female internal genital organs. (Schematic drawing.)

1 **Ovary**
2 Mesovarium
3 **Fundus of uterus**
4 Vesico-uterine pouch
5 **Cervix of uterus**
6 Vaginal portion of cervix
7 **Vagina**
8 Crus of clitoris
9 Labium minus
10 Fimbriae of uterine tube
11 Infundibulum of uterine tube
12 Ligament of the ovary
13 Mesosalpinx
14 Uterine tube
15 Suspensory ligament of ovary (caudally displaced)
16 Broad ligament of uterus
17 Round ligament of uterus
18 Corpus cavernosum of clitoris
19 Glans of clitoris
20 **Hymen, vaginal orifice**
21 Promontory
22 Linea terminalis of pelvis
23 Pubic symphysis

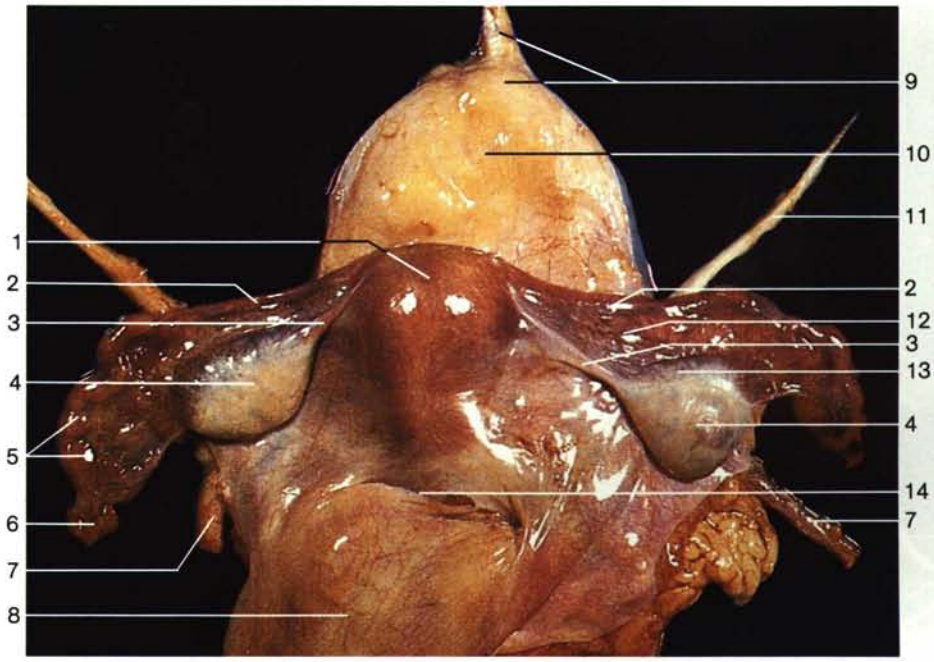

Female internal genital organs, isolated (superior-posterior view).

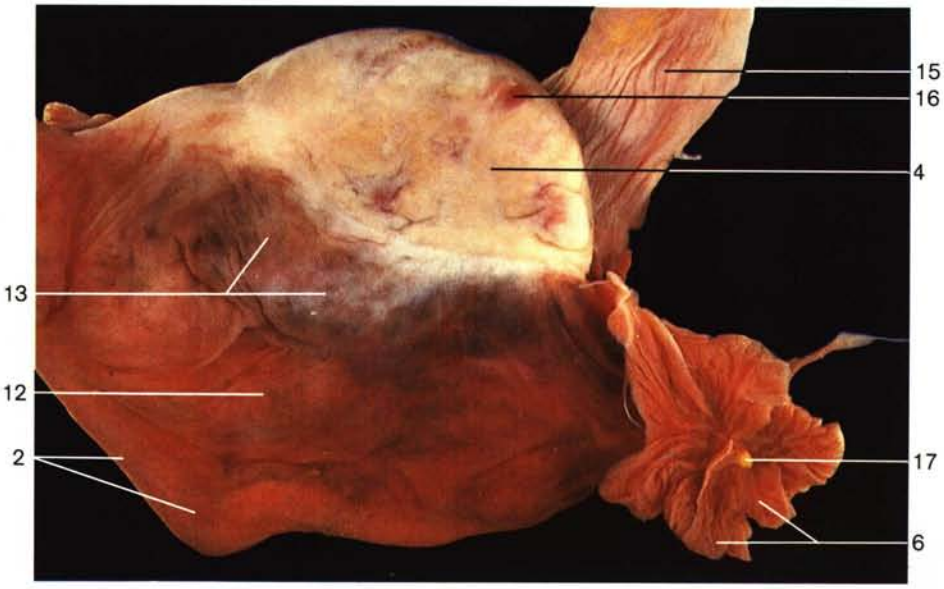

Right ovary and uterine tube, isolated (superior-posterior view). The fimbriae of the uterine tube have been reflected to show the abdominal ostium.

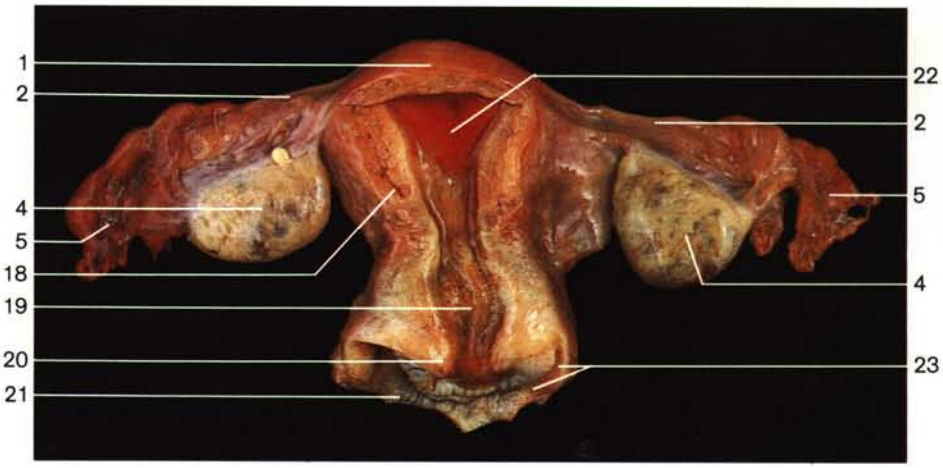

Uterus and related organs (posterior view). The posterior wall of the uterus has been opened.

1 **Fundus of uterus**
2 Uterine tube
3 **Ligament of the ovary**
4 **Ovary**
5 Infundibulum of uterine tube
6 Fimbriae of uterine tube
7 **Ureter**
8 Rectum
9 Apex of urinary bladder and median umbilical ligament
10 Urinary bladder
11 Round ligament of uterus
12 **Mesosalpinx**
13 **Mesovarium**
14 **Rectouterine pouch** (of Douglas)
15 Suspensory ligament of ovary
16 Scarring of ovary (following ovulation)
17 Abdominal opening of uterine tube
18 Body of uterus
19 **Cervical canal**
20 Vaginal portion of cervix of uterus (congestion)
21 Vagina
22 Mucous membrane of uterus
23 Anterior fornix of vagina

Arteriogram of female genital organs (anterior-posterior view). Notice the helicine arteries supplying uterus and ovary.

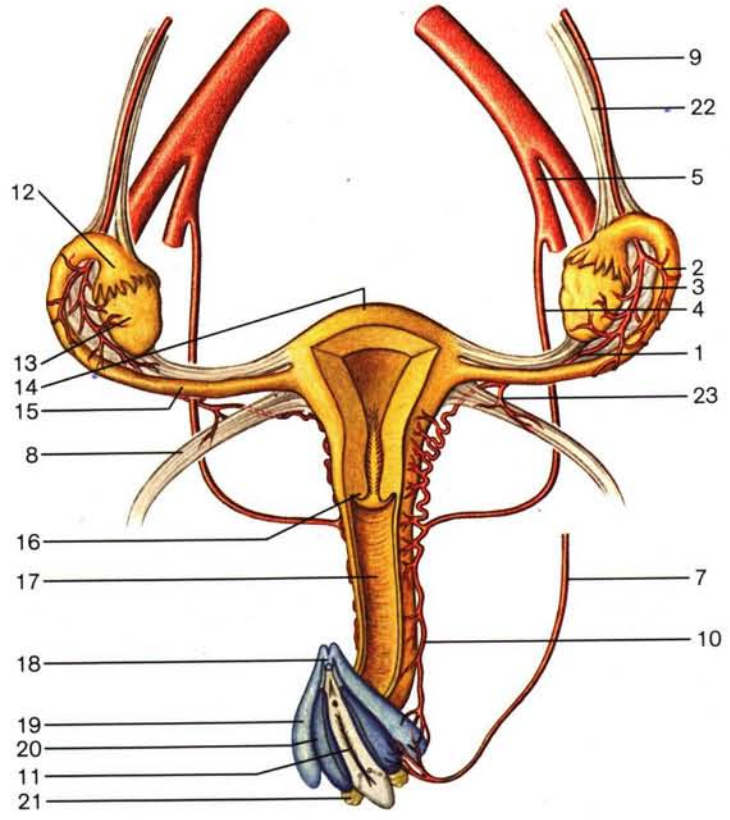

Arteries of female genital organs. (Schematic drawing.)

1 Ovarian branch of uterine artery (anastomoses with ovarian artery)
2 Tubal branch of ovarian artery
3 Ovarian branch of ovarian artery
4 **Uterine artery**
5 Internal iliac artery
6 Inferior gluteal artery
7 **Internal pudendal artery**
8 Round ligament of uterus
9 **Ovarian artery**
10 Vaginal artery
11 Vaginal orifice
12 Infundibulum of uterine tube
13 Ovary
14 Fundus of uterus
15 Uterine tube
16 Vaginal portion of cervix of uterus
17 Vagina
18 Clitoris
19 Corpus cavernosum of clitoris
20 Bulb of vestibule
21 Greater vestibular gland
22 Suspensory ligament of ovary
23 Artery of round ligament

Full term uterus with placenta (anterior view). The anterior wall of the uterus has been removed to show the location of the placenta.

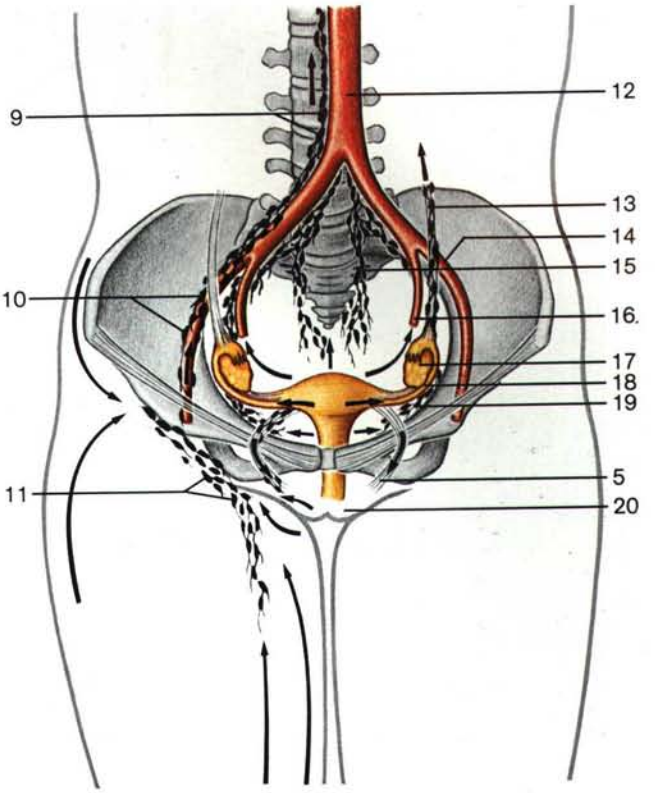

Main drainage routes of lymph vessels of uterus and its adnexa (indicated by arrows). (Schematic drawing.)
Red = arteries; black = lymph vessels and nodes; yellow = internal genital organs.

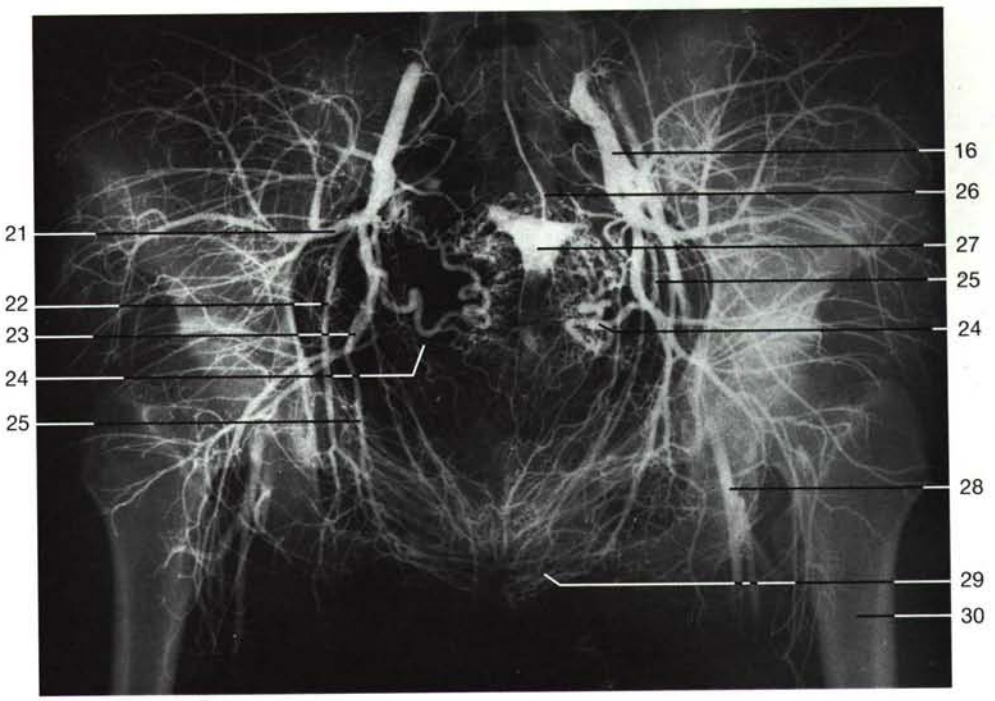

Pelvic vessels in the female (anterior-posterior view). (Arteriogram.)

1 Placenta
2 Amnion and chorion
3 Adnexa of uterus (uterine tube and ovaries)
4 Myometrium
5 Round ligament of uterus
6 Internal orifice of uterus
7 Cervix of uterus
8 **Umbilical cord**
9 Lumbar lymph nodes
10 External iliac lymph nodes
11 Inguinal lymph nodes
12 Abdominal aorta
13 Suspensory ligament of ovary
14 External iliac artery
15 Sacral lymph nodes
16 Internal iliac artery
17 Ovary
18 Uterine tube
19 Internal iliac lymph nodes
20 External genital organs
21 Superior gluteal artery
22 **Obturator artery**
23 Inferior gluteal artery
24 **Uterine artery**
25 **Internal pudendal artery**
26 Middle sacral artery
27 **Uterine cavity**
28 Femoral artery
29 Vessels of labium majus
30 Femur

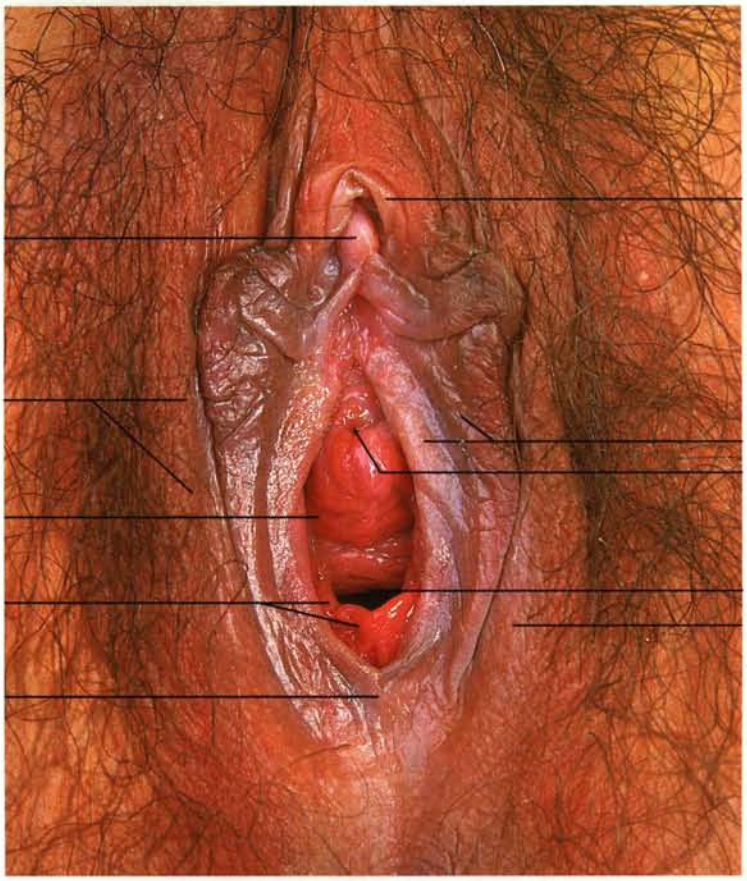

External genital organs in the female (anterior aspect). Labia reflected.

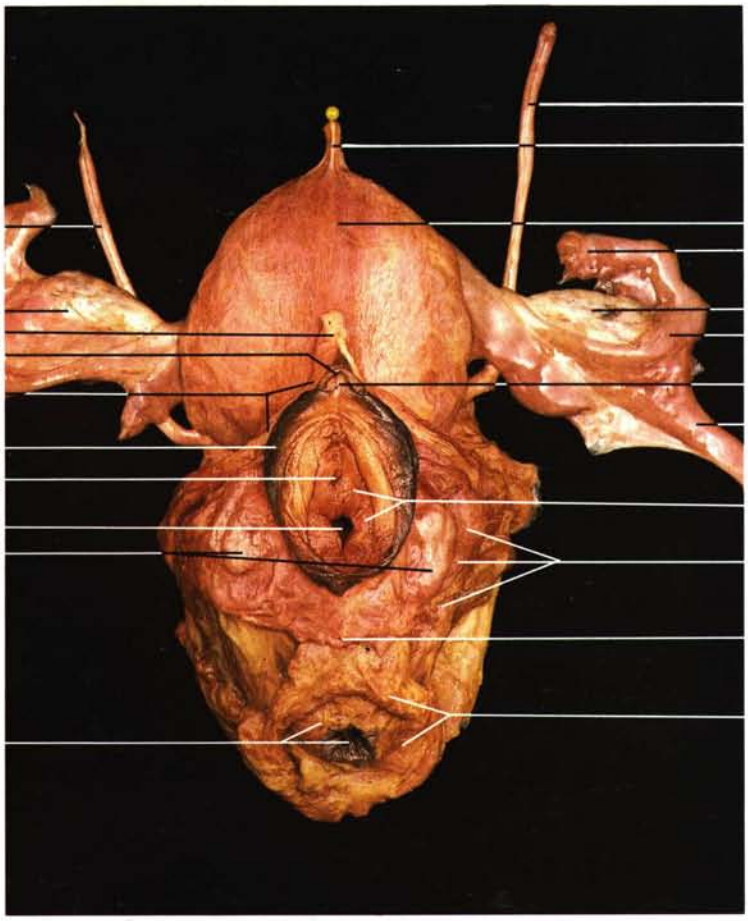

Female external genital organs in relation to internal genital organs and urinary system (isolated, anterior aspect).

1 **Glans clitoris**
2 Labium majus
3 Vestibule of vagina
4 **Hymen**
5 Posterior labial commissure
6 **Body of clitoris**
7 Labium minus
8 **External orifice of urethra**
9 **Vaginal orifice**
10 Ureter
11 Adnexa of uterus
12 Prepuce of clitoris
13 Crus of clitoris
14 Greater vestibular glands
15 Anus and internal anal sphincter muscle
16 Median umbilical ligament containing urachus
17 **Urinary bladder**
18 Infundibulum of uterine tube
19 Ovary
20 Ampulla of uterine tube
21 Suspensory ligament of the ovary
22 Bulbospongiosus muscle and bulb of vestibule
23 Central tendon of perineum (perineal body)
24 External anal sphincter muscle

Cavernous tissue of female external genital organs, isolated (anterior aspect).

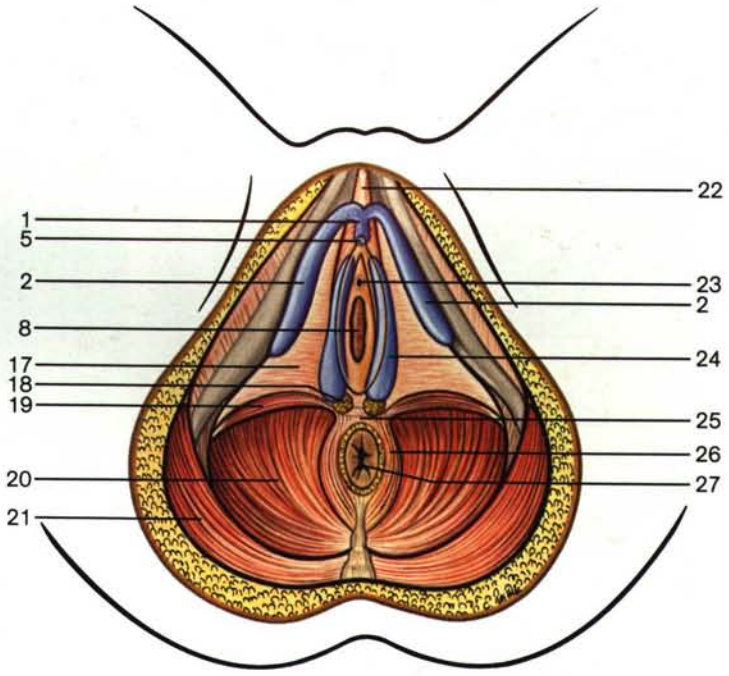

Urogenital and pelvic diaphragms (anterior aspect). (Schematic drawing.)
Blue = cavernous tissue of clitoris and bulb of vestibule.

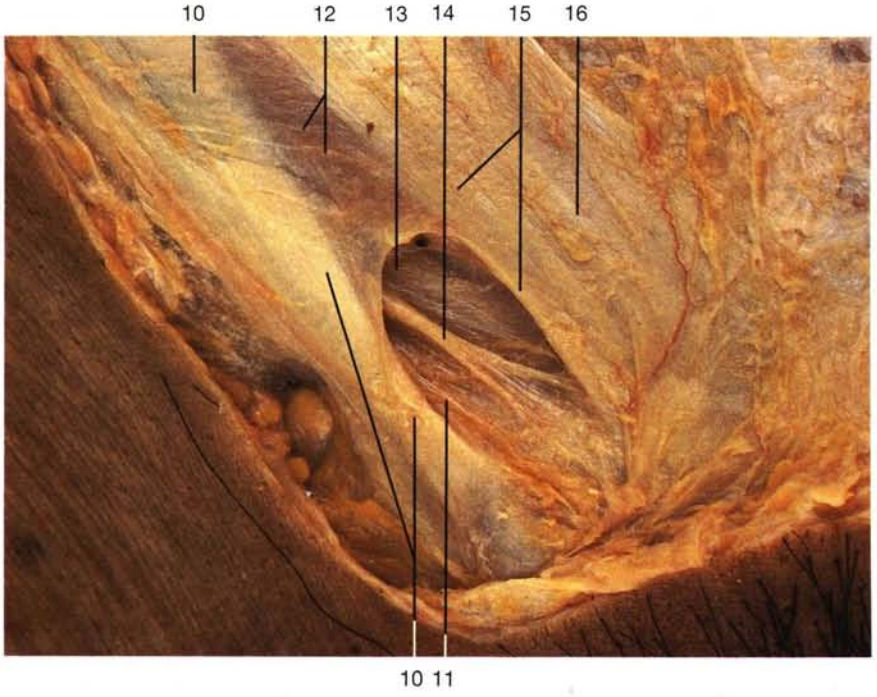

Inguinal canal and round ligament of uterus in situ (right side, ventral aspect).

1 Body of clitoris
2 Crus of clitoris
3 Bulb of vestibule
4 Prepuce of clitoris
5 **Glans of clitoris**
6 Frenulum of clitoris
7 **Labium minus**
8 **Vaginal orifice**
9 **Greater vestibular gland**
10 Lateral crus of superficial inguinal ring
11 Ilioinguinal nerve
12 Intercrural fibers
13 Superficial inguinal ring
14 **Round ligament of uterus**
15 Medial crus of superficial inguinal ring
16 Aponeurosis of external abdominal oblique muscle
17 Deep transverse perineus muscle with fascia
18 Greater vestibular gland
19 Superficial transverse perineus muscle
20 Levator ani muscle
21 Gluteus maximus muscle
22 Suspensory ligament of clitoris
23 External orifice of urethra
24 Bulb of vestibule
25 Perineal body
26 External anal sphincter muscle
27 Anus

Female urogenital diaphragm and external genital organs, superficial layer (from below).

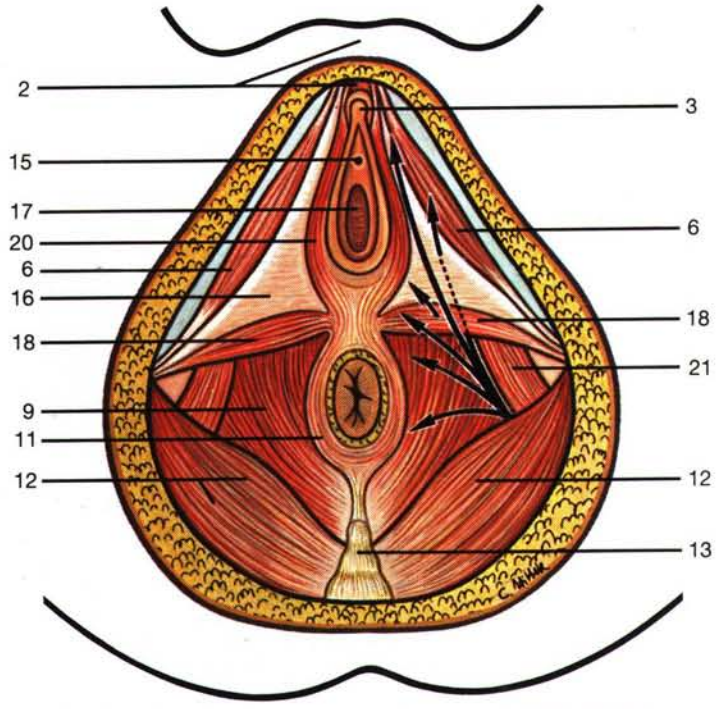

Muscles of pelvic and urogenital diaphragms (from below). (Schematic drawing.)

1 Fatty tissue encasing round ligament
2 Position of pubic symphysis
3 Clitoris
4 Labium minus
5 Bulb of vestibule
6 Ischiocavernosus muscle
7 **Greater vestibular gland**
8 Perineal branches of pudendal nerve
9 Levator ani muscle
10 Inferior rectal nerves
11 External anal sphincter muscle
12 Gluteus maximus muscle
13 Coccyx
14 Fatty tissue of mons pubis
15 **External orifice of urethra**
16 Urogenital diaphragm with fascia of deep transverse perineus muscle
17 **Vaginal orifice**
18 Superficial transverse perineal muscle
19 **Anus**
20 Bulbospongiosus muscle
21 Obturator internus muscle

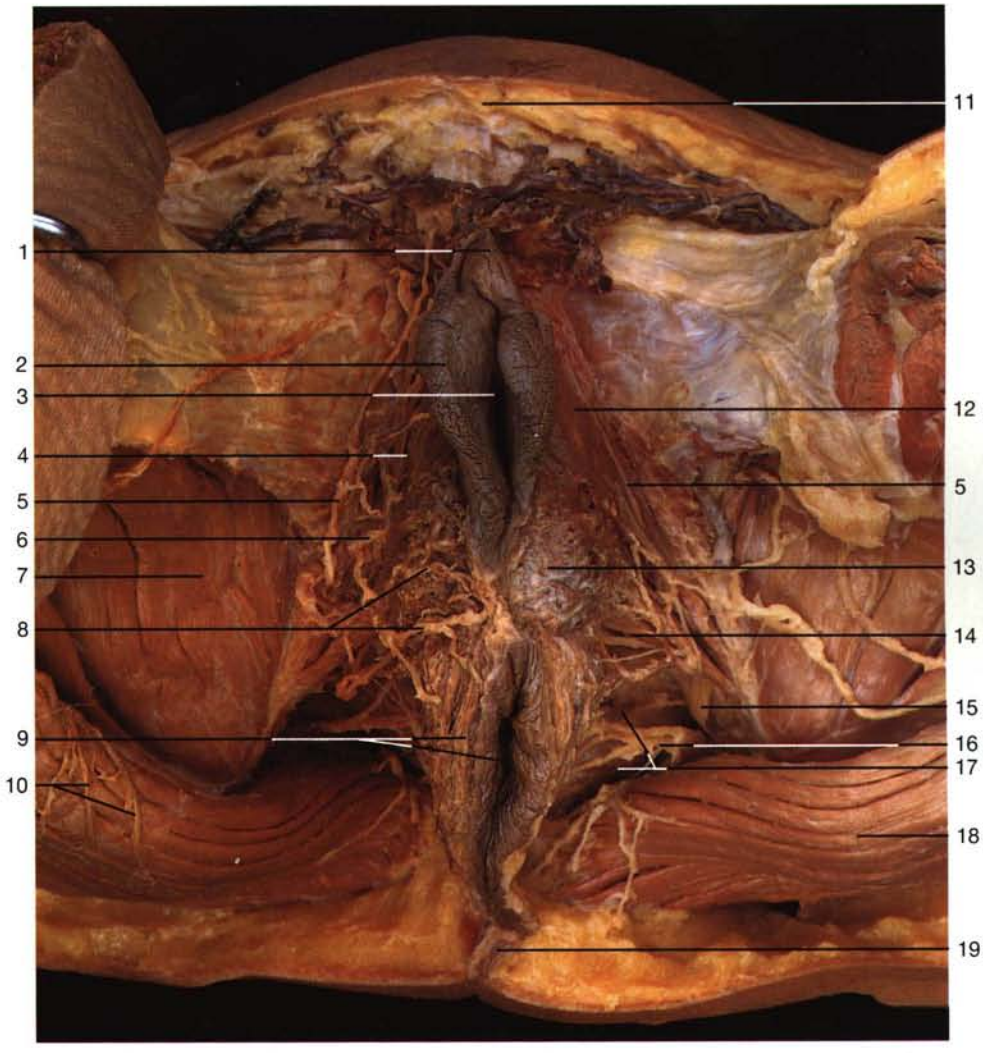

Urogenital diaphragm and external genital organs in the female, superficial layer (from below). On the right side the bulb of vestibule has been removed.

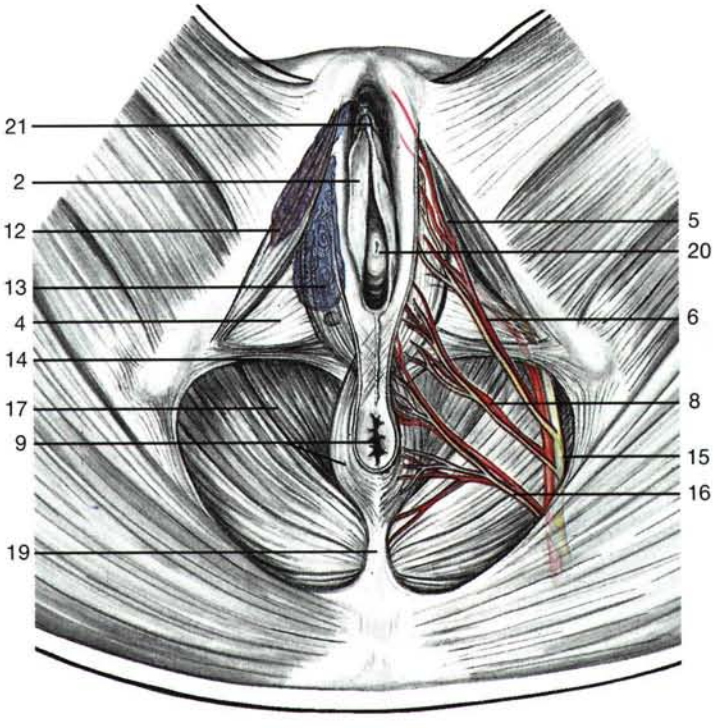

External female genital organs. Position of arteries and nerves; bulb of vestibule in blue. (Schematic drawing.)

1 **Prepuce of clitoris**
2 Labium minus
3 **Vaginal orifice**
4 Deep transverse perineus muscle
5 Dorsal nerve of clitoris
6 Posterior labial nerves
7 Great adductor muscle
8 Perineal branches of pudendal nerve
9 **Anus** and **external anal sphincter muscle**
10 Inferior cluneal nerves
11 Mons pubis
12 Crus of clitoris with ischiocavernosus muscle
13 Bulb of vestibule
14 Superficial transverse perineus muscle
15 **Pudendal nerve** and **internal pudendal artery**
16 Inferior rectal nerves
17 Levator ani muscle
18 Gluteus maximus muscle
19 Anococcygeal ligament
20 External urethral orifice
21 Glans of clitoris

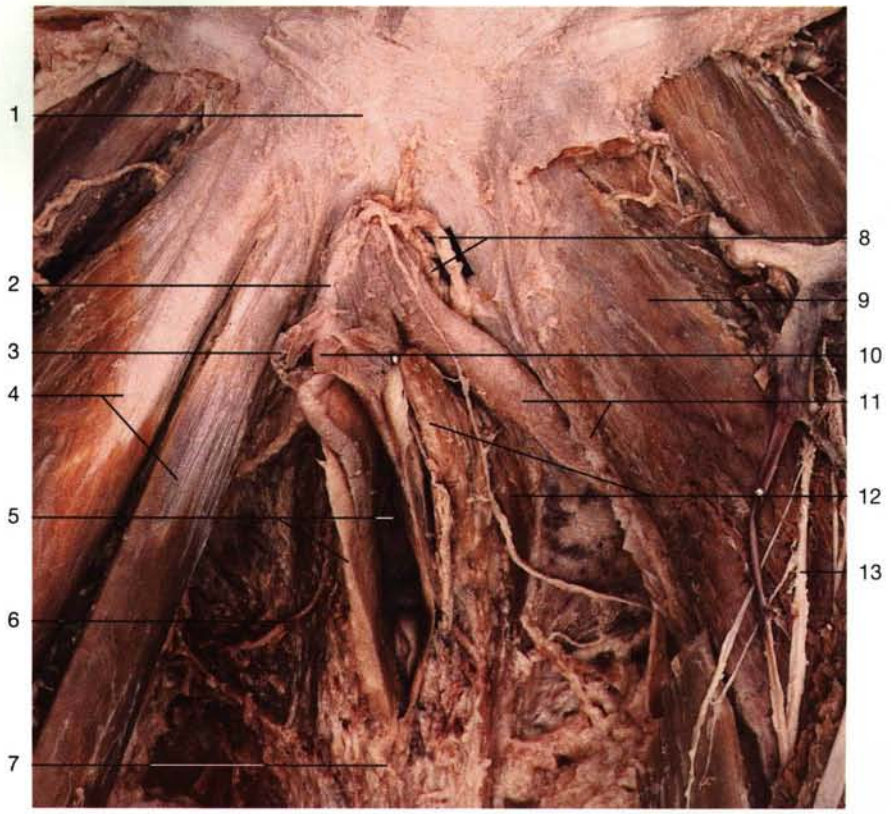

External female genital organs (inferior aspect). The clitoris has been dissected and slightly reflected to the right. The prepuce of clitoris has been divided to display the glans.

1 Position of pubic symphysis
2 Body of clitoris
3 Prepuce of clitoris
4 Adductor longus and gracilis muscles
5 External orifice of vagina and labium minus
6 Posterior labial nerve
7 Perineal body
8 Deep artery of clitoris and dorsal nerve of clitoris
9 Adductor brevis muscle
10 **Glans of clitoris**
11 Crus of clitoris and ischiocavernosus muscle
12 **Bulb of vestibule** and bulbospongiosus muscle
13 Anterior branch of obturator nerve
14 Labium minus
15 **Vaginal orifice**
16 Posterior labial nerves
17 Branches of pudendal nerve
18 External sphincter of anus
19 **Anus**
20 Bulb of vestibule (divided)
21 Dorsal artery of clitoris
22 Superficial transverse perineus muscle
23 Perineal branch of posterior femoral cutaneous nerve
24 Levator ani muscle
25 **Pudendal nerve** and **internal pudendal artery**
26 Inferior rectal nerves
27 Gluteus maximus muscle
28 Anococcygeal ligament

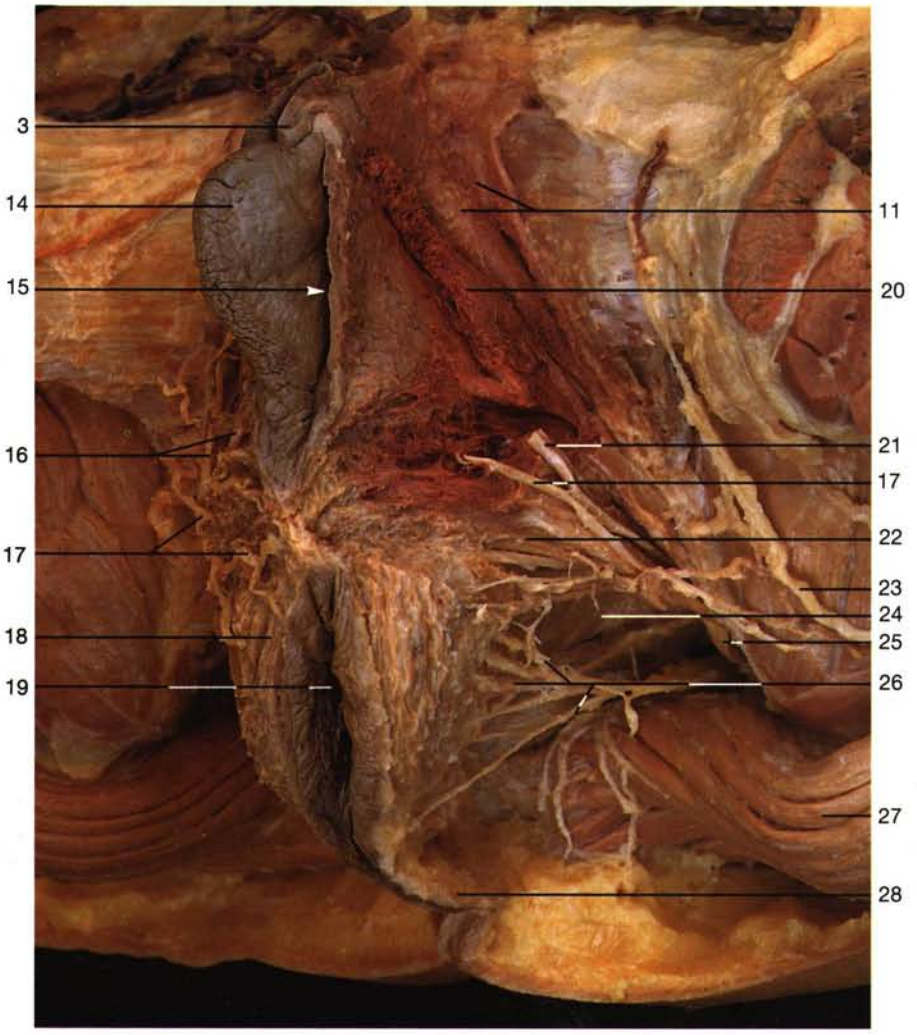

Urogenital diaphragm and external genital organs in the female (lateral inferior view). The bulb of vestibule has partly been removed, the left labium minus was cut away.

Coronal section through the pelvic cavity of the female (cf. MR-Scan on page 335).

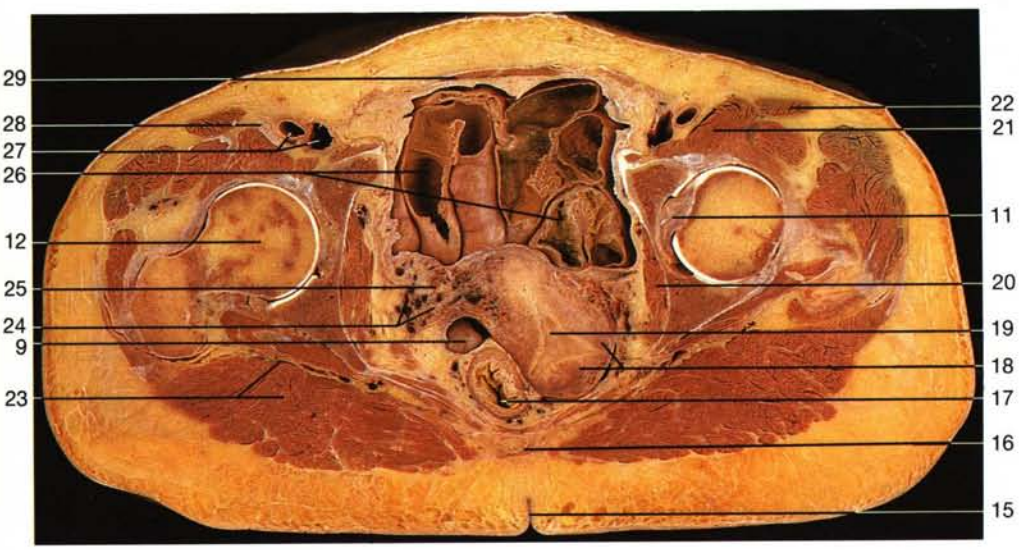

Horizontal section through pelvic cavity at level of uterus (from below). The uterus is retroverted to the left.

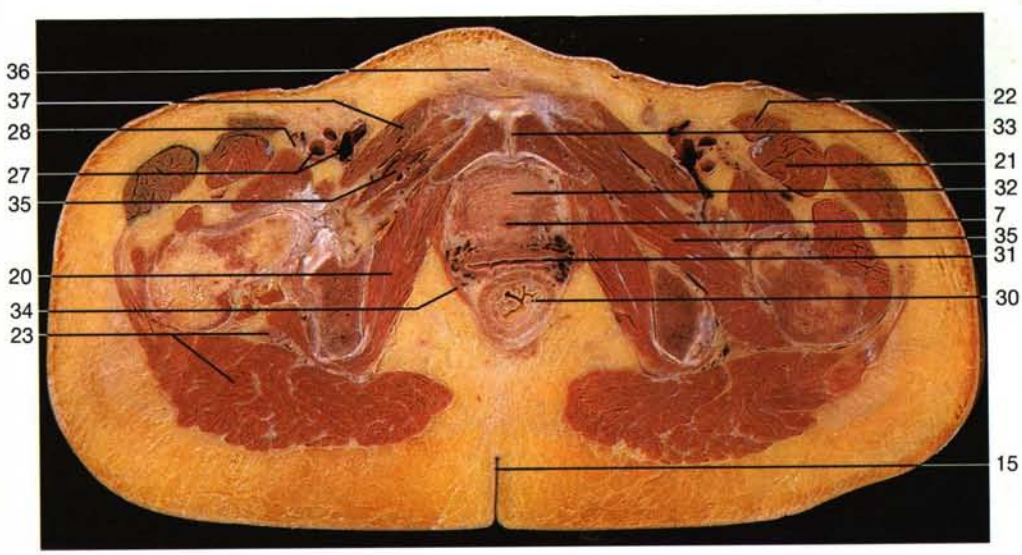

Horizontal section through the pelvic cavity at level of the urethral sphincter and vagina (from below).

1 Ilium
2 Rectum
3 Rectouterine fold
4 **Ovary**
5 **Uterine tube**
6 **Urinary bladder**
7 **Urethra**
8 Labium minus
9 Rectouterine pouch of Douglas
10 **Uterus** (uterovesical pouch)
11 Ligament of the head of the femur
12 Head of femur
13 Vestibule of vagina
14 Labium majus
15 Anal cleft
16 Coccyx
17 Rectum
18 Myometrium of uterus
19 Uterine cavity
20 Obturator internus muscle
21 Iliopsoas muscle
22 Sartorius muscle
23 Sciatic nerve and gluteus maximus muscle
24 Uterine venous plexus
25 Broad ligament
26 Small intestine
27 Femoral artery and vein
28 Femoral nerve
29 Pyramidalis muscle
30 Rectum (anal canal)
31 Vagina
32 Urethral sphincter muscle (base of urinary bladder)
33 Pubic symphysis
34 Levator ani muscle
35 Obturator externus muscle
36 Mons pubis
37 Pectineus muscle

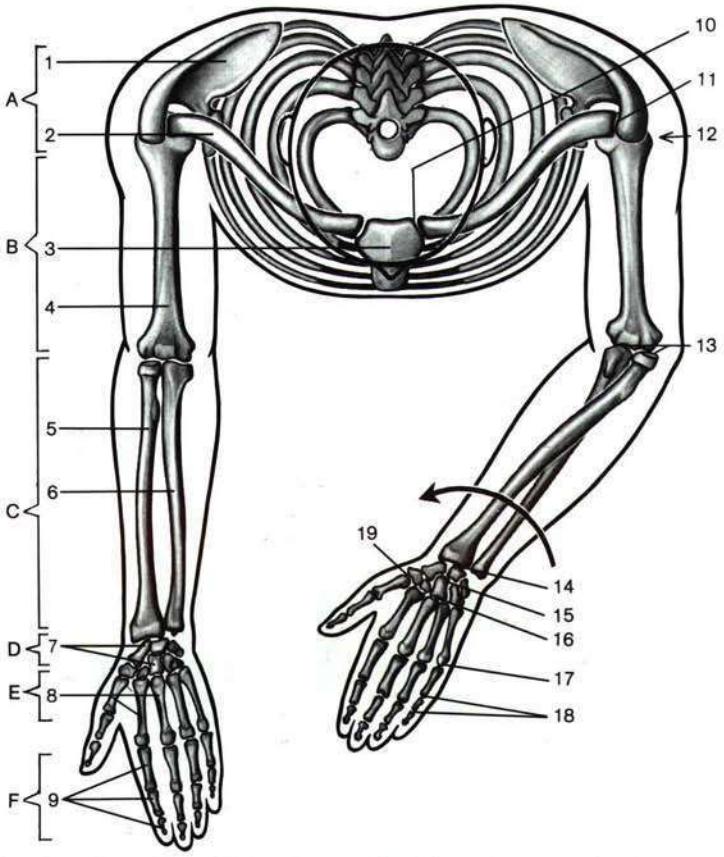

A Pectoral girdle
B Arm
C Forearm
D Wrist
E Palm of hand
F Finger

Bones

1 Scapula
2 Clavicle
3 Sternum
4 Humerus
5 Radius
6 Ulna
7 Carpal bones
8 Metacarpal bones
9 Phalanges

Joints

10 Sternoclavicular joint
11 Acromioclavicular joint
12 Shoulder joint
13 Elbow joint
14 Radiocarpal joint
15 Midcarpal joint
16 Carpometacarpal joint
17 Metacarpophalangeal joint
18 Interphalangeal joints of fingers
19 Carpometacarpal joint of thumb

Organization of shoulder girdle and upper limb (superior aspect). The two positions of the forearm essential to manual skills in the human, supination (right arm) and pronation (left arm) are shown.

Bones of shoulder girdle articulated with the thorax (superior aspect).

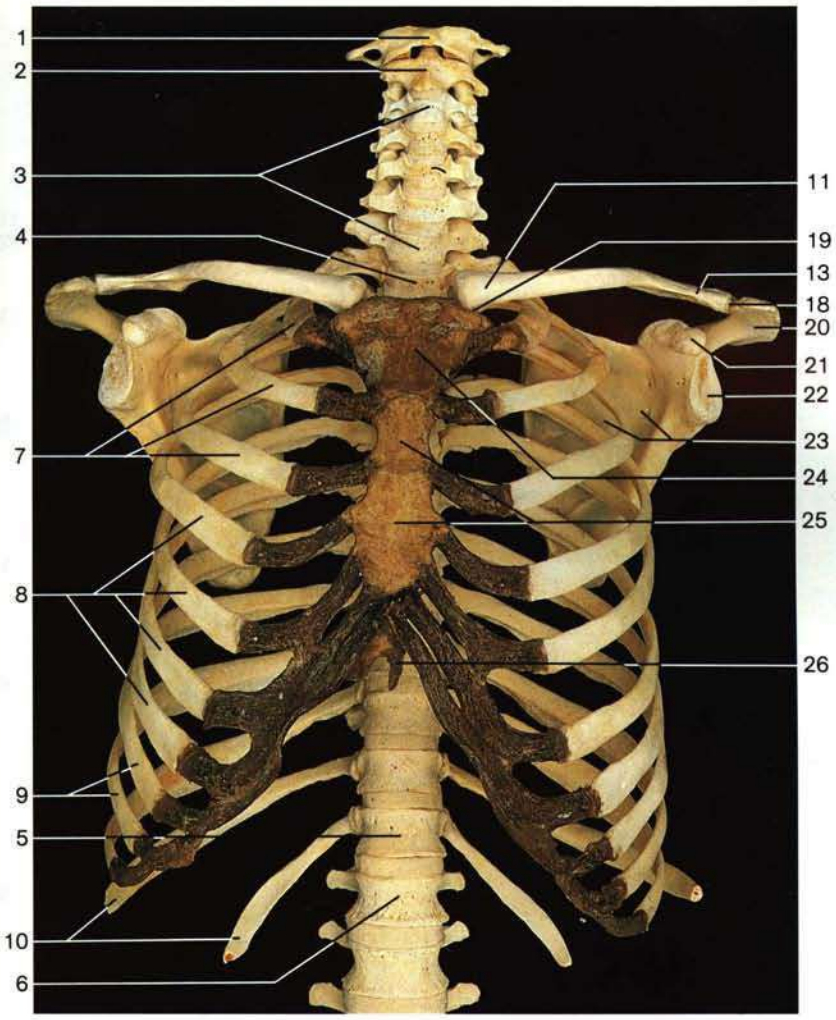

Skeleton of shoulder girdle and thorax (anterior aspect). The cartilaginous parts of the ribs appear dark brown.

Vertebral column

1 Atlas
2 Axis
3 3rd-7th cervical vertebrae
4 1st thoracic vertebra
5 12th thoracic vertebra
6 1st lumbar vertebra

Ribs

7 1st–3rd ribs } True ribs
8 4th–7th ribs
9 8th–10th ribs } False ribs
10 11th and 12th ribs (floating ribs)

Clavicle

11 Sternal end
12 Articular facet for sternum
13 Acromial end
14 Articular facet for acromion
15 Impression for costoclavicular ligament
16 Conoid tubercle
17 Trapezoid line
18 Site of acromioclavicular joint
19 Site of sternoclavicular joint

Scapula

20 Acromion
21 Coracoid process
22 Glenoid cavity
23 Costal surface

Sternum

24 Manubrium
25 Body
26 Xiphoid process

Right clavicle (superior aspect).

Right clavicle (inferior aspect).

Because of his upright posture, man's upper limb has developed a high degree of mobility. The shoulder girdle is to a great extent movable in the thorax and is connected with the trunk only by the sternoclavicular joint. A further characteristic of man's forearm is the capacity for rotation (i.e., pronation and supination).

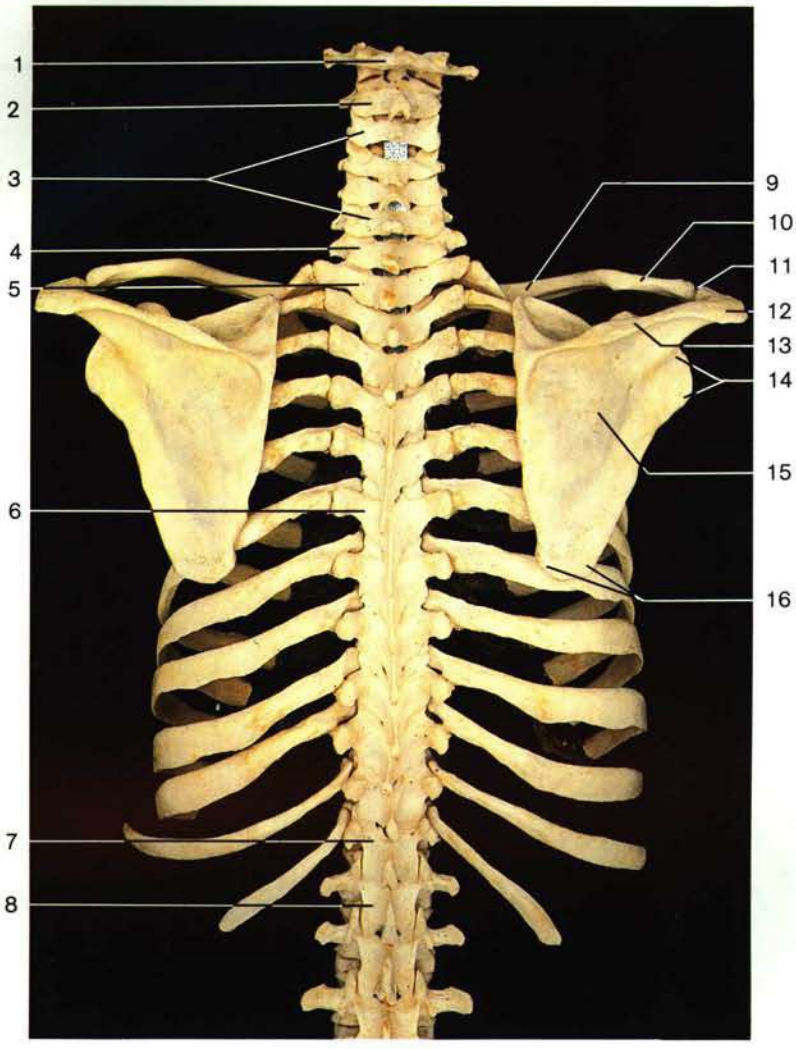

Skeleton of shoulder girdle and thorax (posterior view).

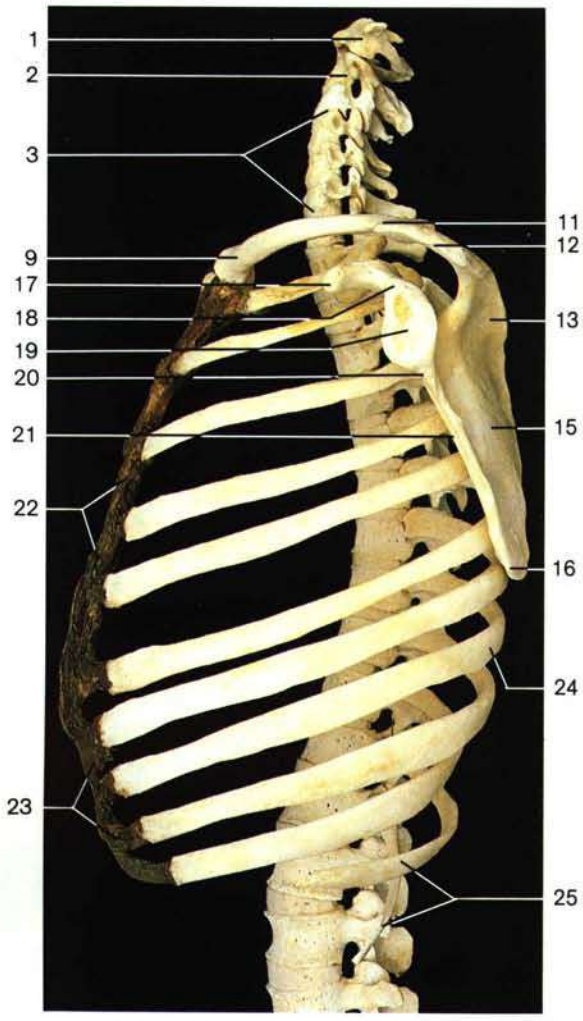

Skeleton of shoulder girdle and thorax (lateral view).

Vertebral column

1 Atlas
2 Axis
3 3rd–6th cervical vertebrae
4 7th vertebra (vertebra prominens)
5 1st thoracic vertebra
6 6th thoracic vertebra
7 12th thoracic vertebra
8 1st lumbar vertebra

Clavicle

9 Sternal end
10 Acromial end
11 Site of acromioclavicular joint

Scapula

12 Acromion
13 Spine of scapula
14 Lateral angle
15 Posterior surface
16 Inferior angle
17 Coracoid process
18 Supraglenoid tubercle
19 Glenoid cavity
20 Infraglenoid tubercle
21 Lateral margin

Thorax

22 Body of sternum
23 Costal arch
24 Angle of ribs
25 Free ribs

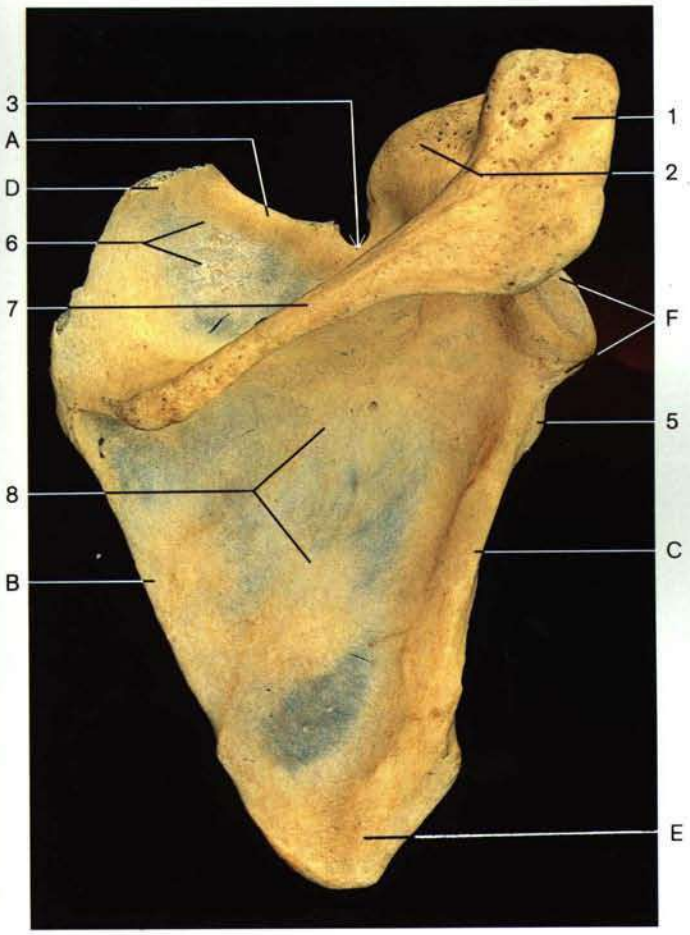

Right scapula (posterior aspect).

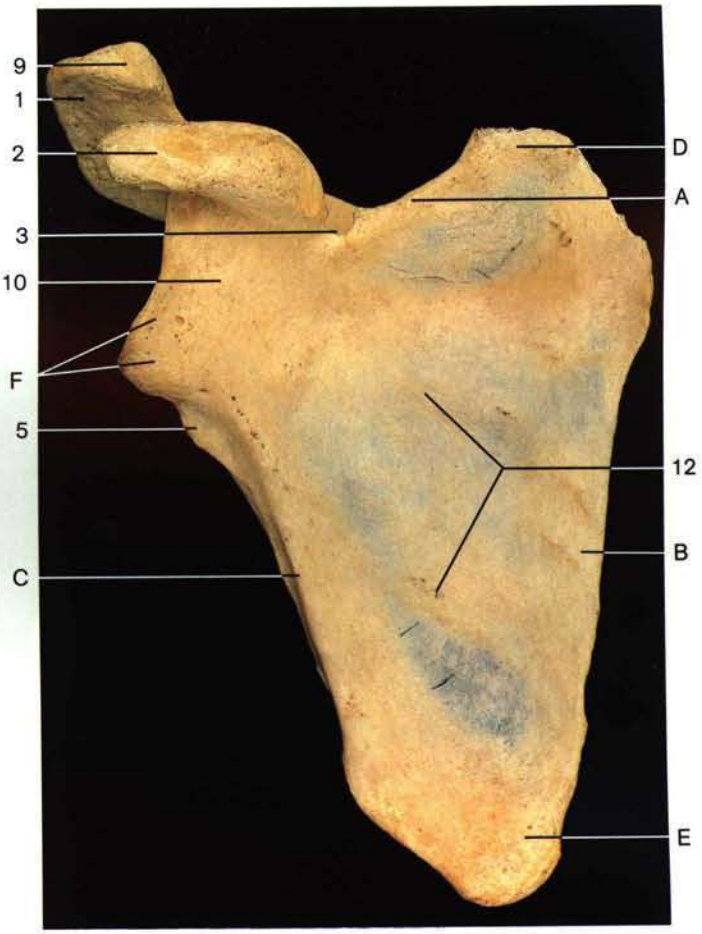

Right scapula (anterior aspect costal surface).

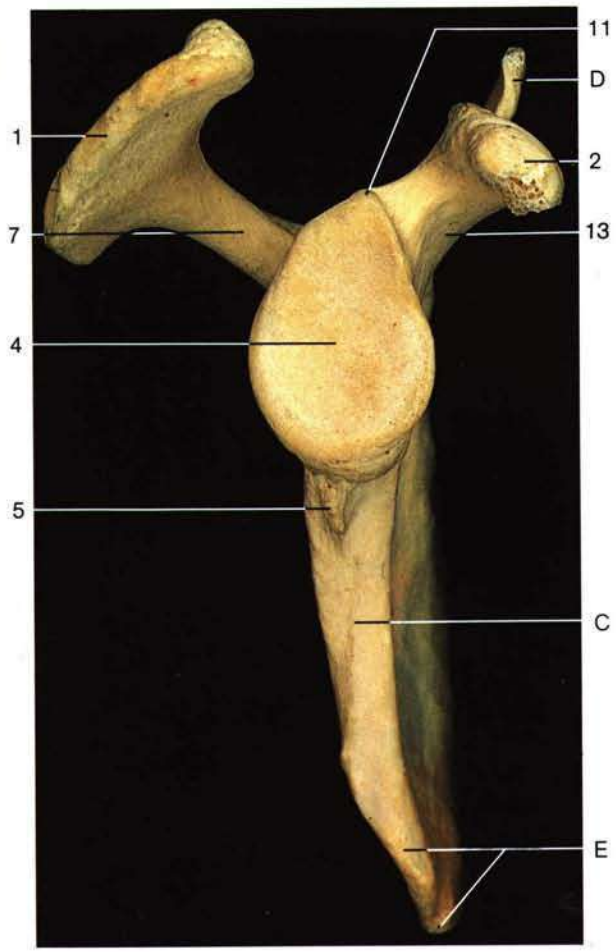

Right scapula (lateral view).

Scapula

A Superior border
B Medial border
C Lateral border
D Superior angle
E Inferior angle
F Lateral angle

1 Acromion
2 Coracoid process
3 Scapular notch
4 Glenoid cavity
5 Infraglenoid tubercle
6 Supraspinous fossa
7 Spine
8 Infraspinous fossa
9 Articular facet for acromion
10 Neck
11 Supraglenoid tubercle
12 Costal (anterior) surface
13 Base of coracoid process

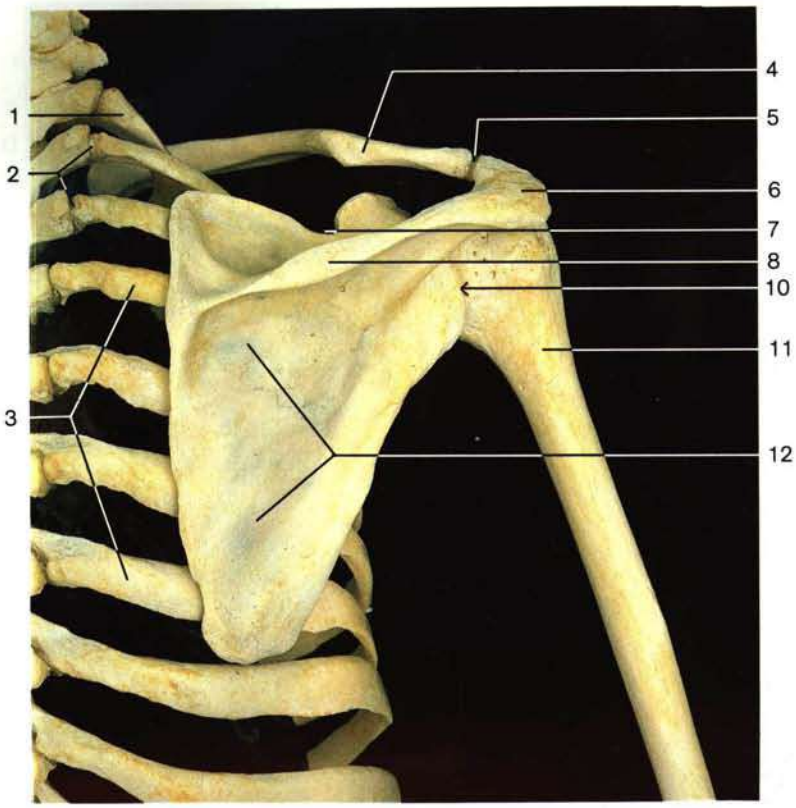

1 1st rib
2 Position of costotransverse joints
3 4th–7th ribs
4 **Clavicle**
5 Position of acromioclavicular joint
6 Acromion
7 Scapular notch
8 Spine of scapula
9 Head of humerus
10 Glenoid cavity
11 Surgical neck of humerus
12 Posterior surface of **scapula**
13 Coracoid process
14 Infraglenoid tubercle
15 Greater tubercle of **humerus**
16 Anatomical neck of humerus

Bones of shoulder joint (posterior aspect).

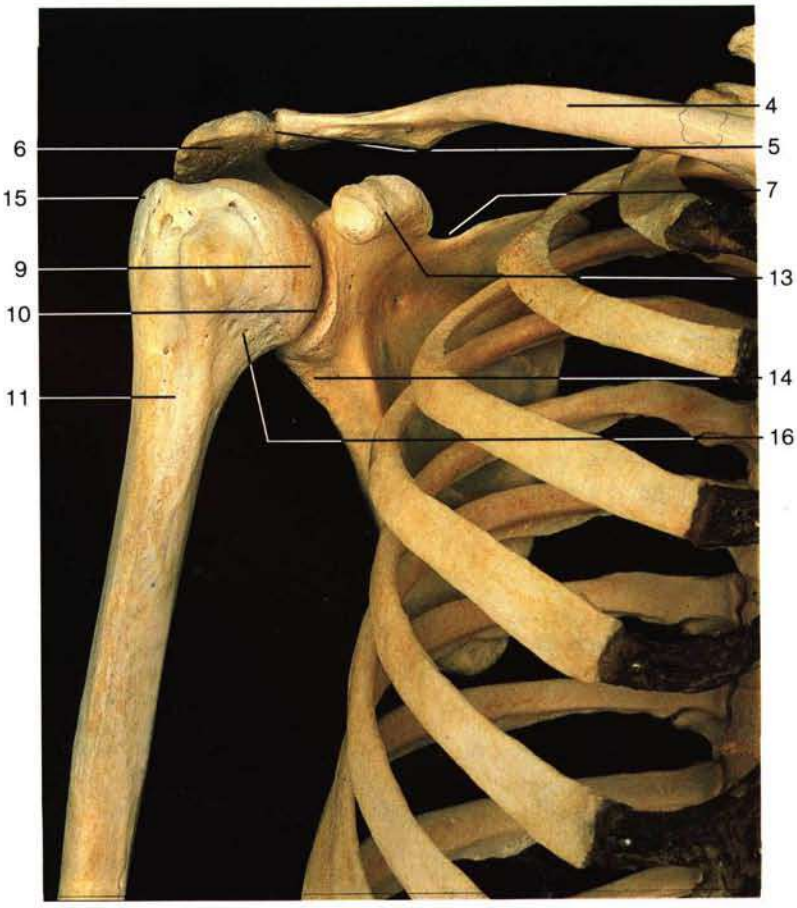

Bones of shoulder joint (anterior aspect).

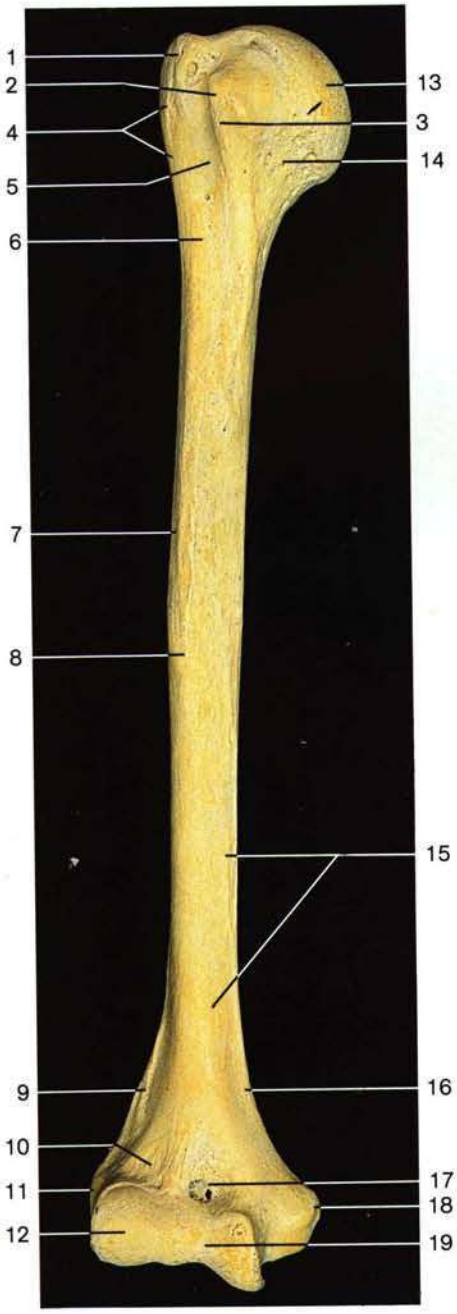

Right humerus (anterior aspect).

Right humerus (medial aspect).

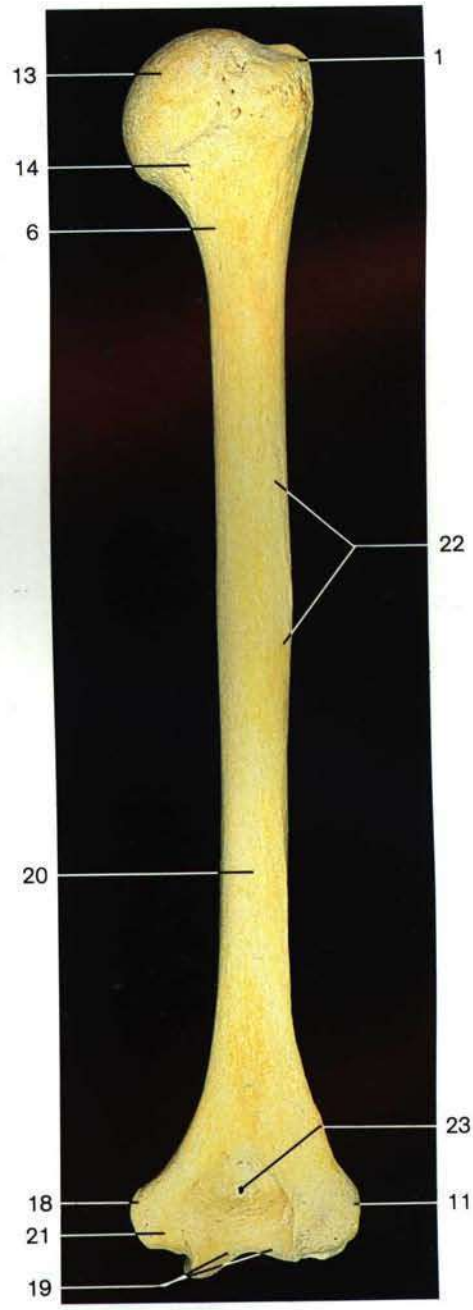

Right humerus (posterior aspect).

Humerus

1 Greater tubercle
2 Lesser tubercle
3 Crest of lesser tubercle
4 Crest of greater tubercle
5 Intertubercular sulcus
6 Surgical neck
7 Deltoid tuberosity
8 Anterolateral surface
9 Lateral supracondylar ridge
10 Radial fossa
11 Lateral epicondyle
12 Capitulum
13 Head
14 Anatomical neck
15 Anteromedial surface
16 Medial supracondylar ridge
17 Coronoid fossa
18 Medial epicondyle
19 Trochlea
20 Posterior surface
21 Groove for ulnar nerve
22 Groove for radial nerve
23 Olecranon fossa

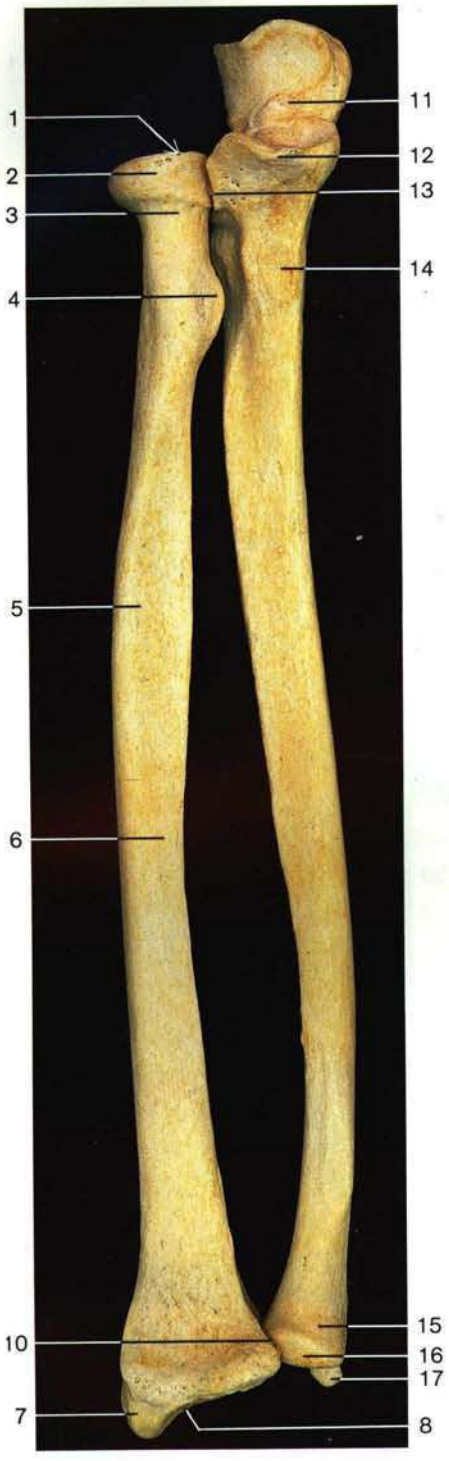

Bones of right forearm, radius and ulna (anterior aspect).

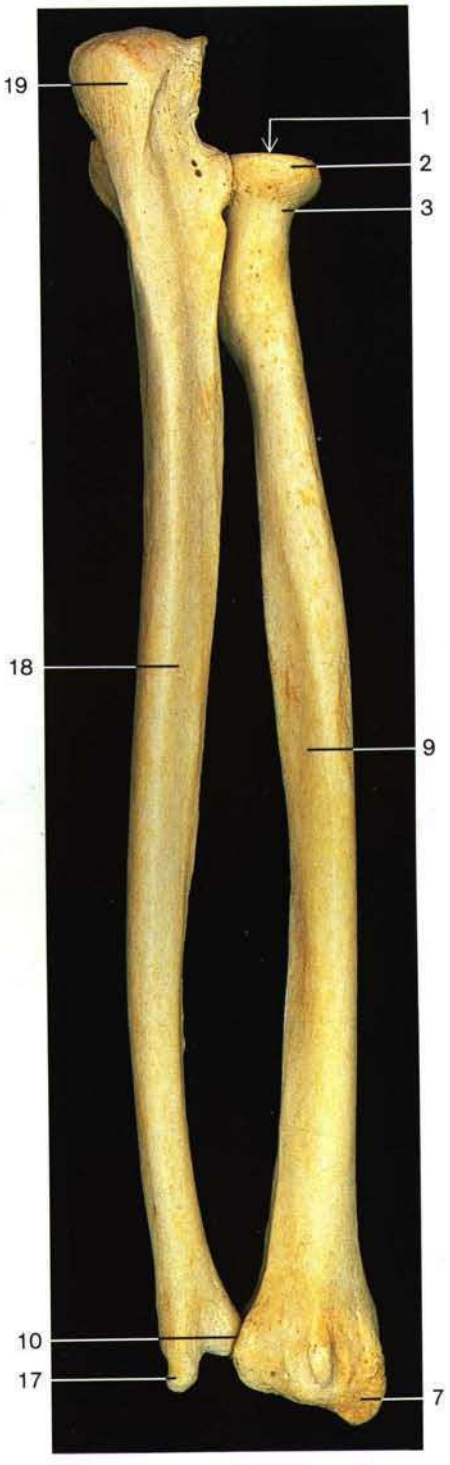

Bones of right forearm, radius and ulna (posterior aspect).

Radius

1 Head
2 Articular circumference
3 Neck
4 Radial tuberosity
5 Shaft
6 Anterior surface
7 Styloid process
8 Articular surface
9 Posterior surface
10 Ulnar notch

Ulna

11 Trochlear notch
12 Coronoid process
13 Radial notch
14 Ulnar tuberosity
15 Head
16 Articular circumference
17 Styloid process
18 Posterior surface
19 Olecranon

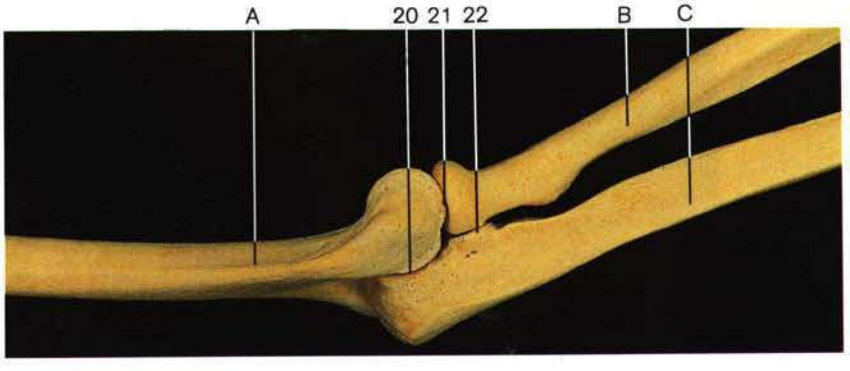

Bones of right elbow joint (lateral aspect).

Articulations at the right elbow

20 Site of humeroulnar joint
21 Site of humeroradial joint
22 Site of proximal radioulnar joint

A Humerus
B Radius
C Ulna

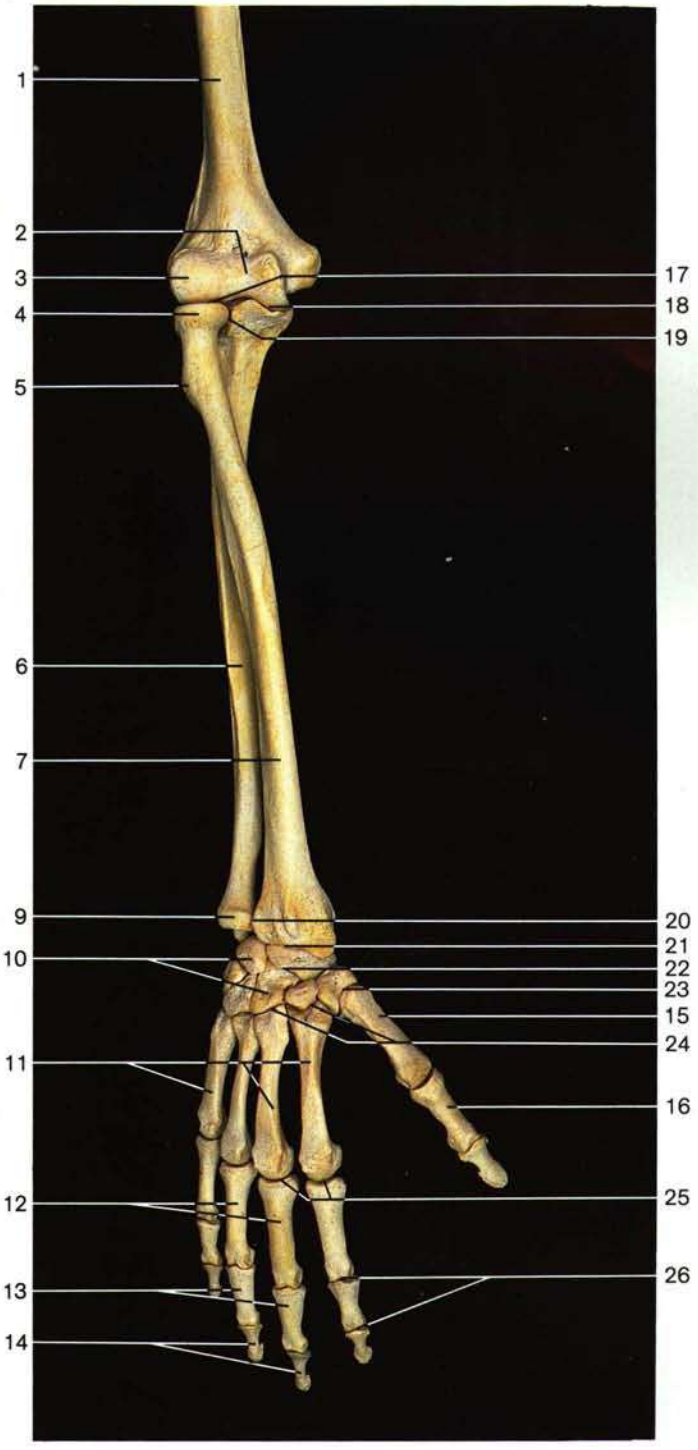

Skeleton of right forearm and hand in pronation.

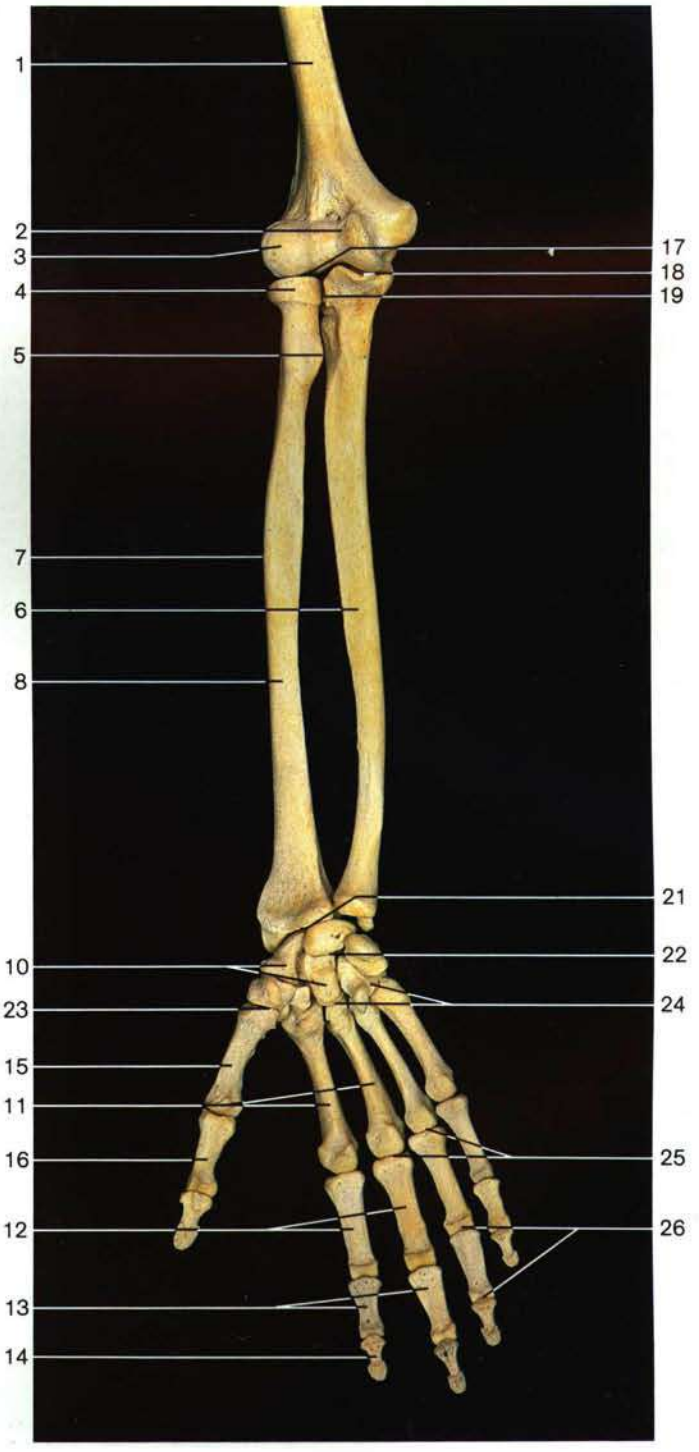

Skeleton of right forearm and hand in supination.

1 Humerus
2 Trochlea of humerus
3 Capitulum of humerus
4 Articular circumference of radius
5 Radial tuberosity
6 Anterior surface of **ulna**
7 Posterior surface of **radius**
8 Anterior surface of radius
9 Articular circumference of ulna
10 **Carpal bones**
11 **Metacarpal bones**
12 Proximal phalanges
13 Middle phalanges
14 Distal phalanges
15 Metacarpal bone of thumb
16 Proximal phalanx of thumb

Sites of joints

17 Humeroradial joint
18 Humeroulnar joint
19 Proximal radioulnar joint
20 Distal radioulnar joint
21 Radiocarpal joint
22 Midcarpal joint
23 Carpometacarpal joint of thumb
24 Carpometacarpal joints
25 Metacarpophalangeal joints
26 Interphalangeal joints of fingers

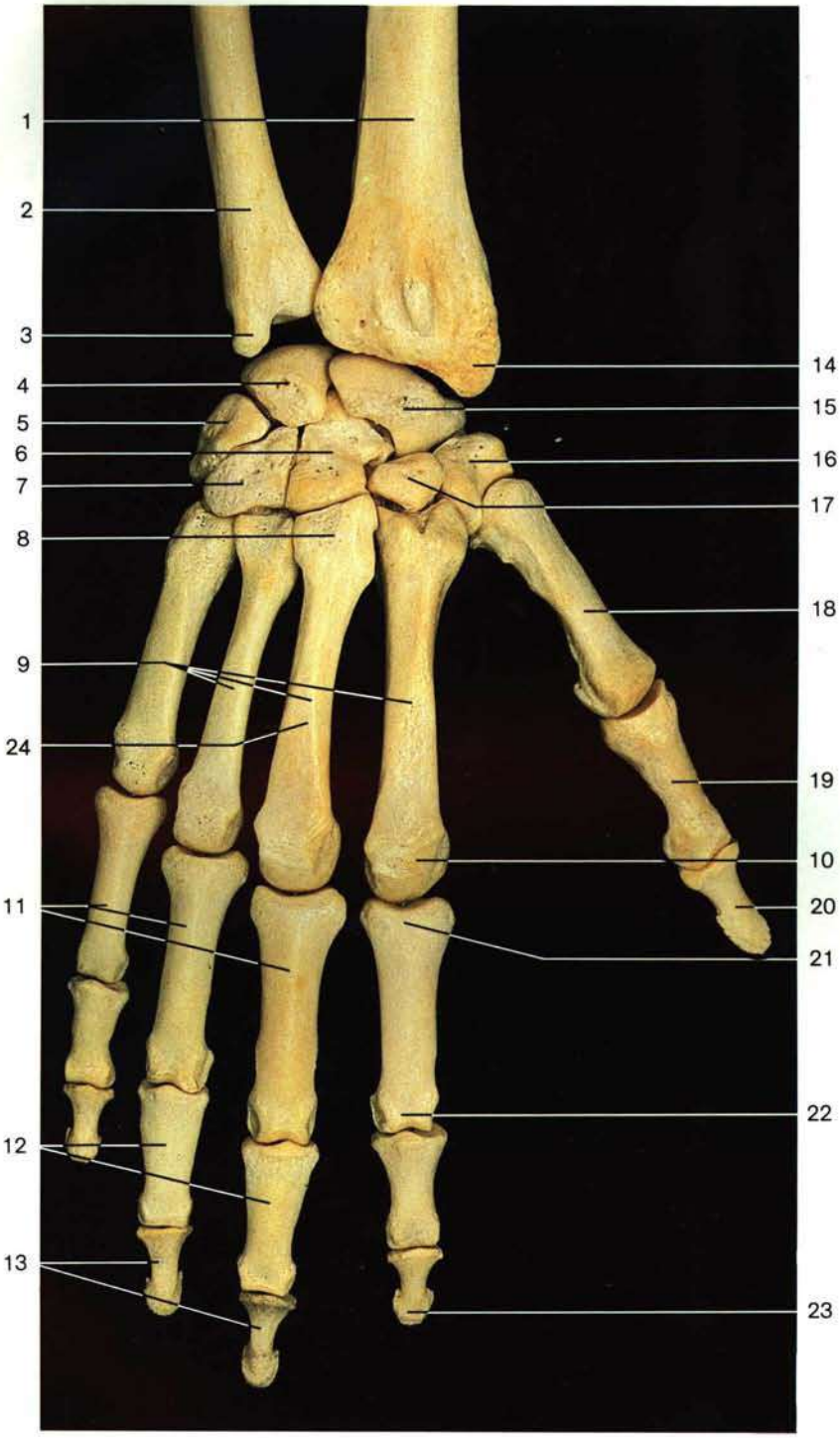

Skeleton of right wrist and hand (dorsal aspect).

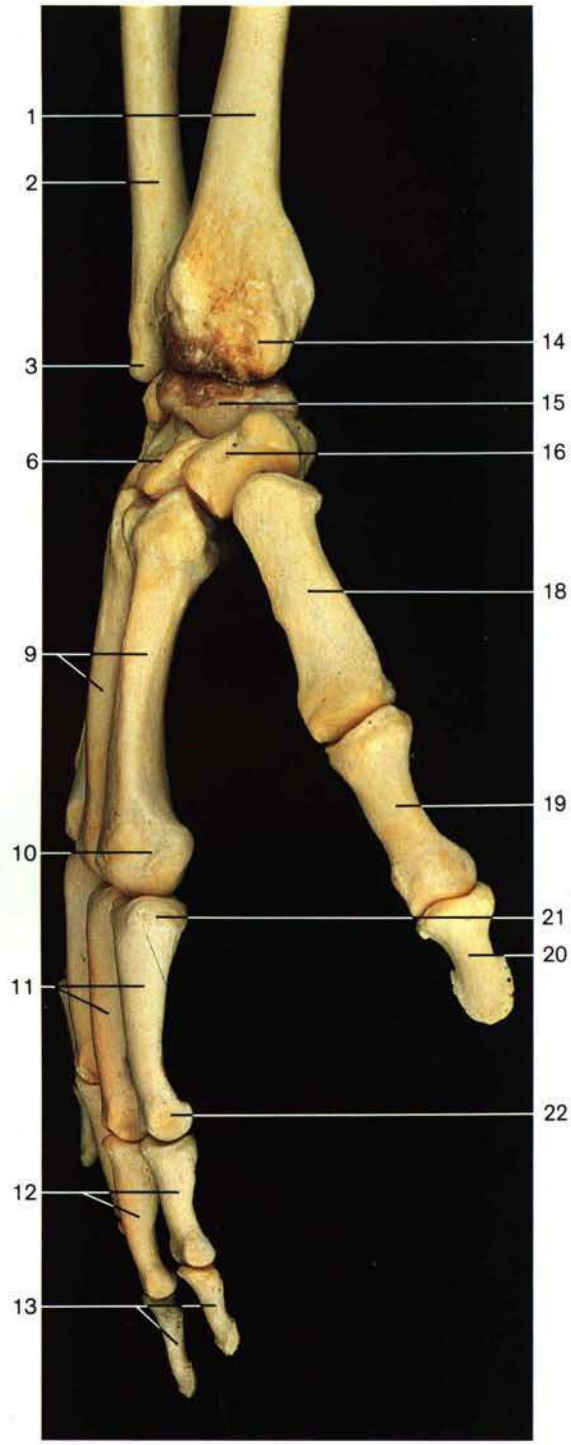

Skeleton of right wrist and hand (medial aspect).

1 **Radius**
2 **Ulna**
3 Styloid process of ulna
4 Lunate bone
5 Triquetral bone
6 Capitate bone
7 Hamate bone
} **Carpal bones**
8 Base of third metacarpal bone
9 **Metacarpal bones**
10 Head of metacarpal bone
11 Proximal phalanges of hand
12 Middle phalanges of hand
13 Distal phalanges of hand
14 Styloid process of radius
15 Scaphoid bone
16 The trapezium
17 Trapezoid bone
} **Carpal bones**
18 Metacarpal bone of thumb
19 Proximal phalanx of thumb
20 Distal phalanx of thumb
21 Base of second proximal phalanx
22 Head of second proximal phalanx
23 Tuberosity of distal phalanx
24 Body of third metacarpal bone

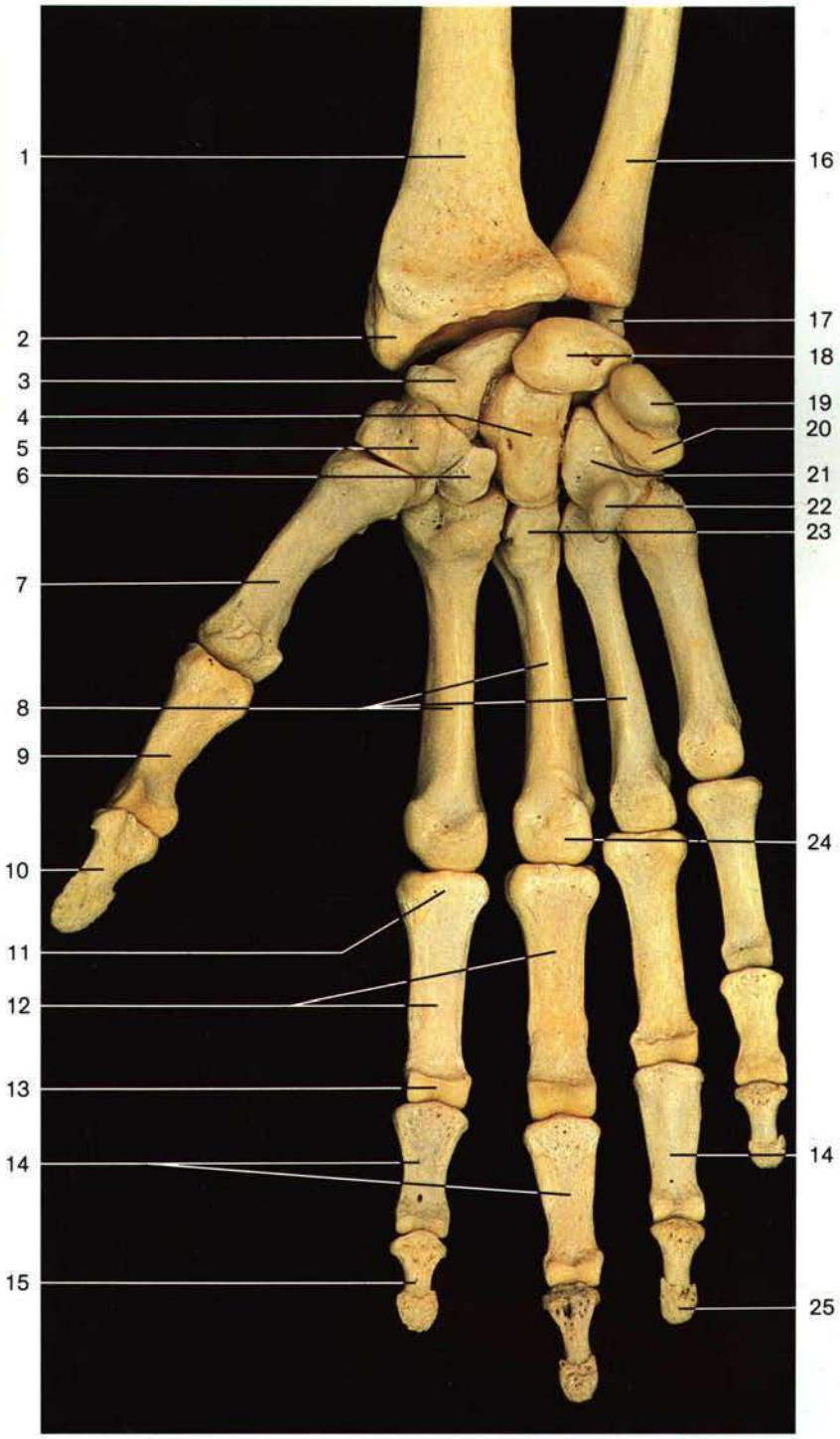

1 Radius
2 Styloid process of radius
3 Scaphoid bone
4 Capitate bone
5 Trapezium
6 Trapezoid bone
} **Carpal bones**
7 First metacarpal bone
8 Second to fourth metacarpal bones
9 Proximal phalanx of thumb
10 Distal phalanx of thumb
11 Base of second proximal phalanx
12 Proximal phalanges
13 Head of second proximal phalanx
14 Middle phalanges
15 Distal phalanx
16 Ulna
17 Styloid process of ulna
18 Lunate bone
19 Pisiform bone
20 Triquetral bone
21 Hamate bone
22 Hamulus or hook of hamate bone
} **Carpal bones**
23 Base of third metacarpal bone
24 Head of metacarpal bone
25 Tuberosity of distal phalanx

Skeleton of right wrist and hand (palmar aspect).

The human hand is one of the most admirable structures of the human body. The carpometacarpal joint of the thumb, a saddle joint, enjoys wide mobility, so that the thumb can get in contact with all other fingers, thus enabling the hand to become an instrument for grasping and psychologic expression. During evolution these newly developed functions appeared after the erect posture of the human body was achieved. An inevitable prerequisite for the development of human cultures is not only the differentiation of the brain but also the development of an organ capable of realizing its ideas: the human hand.

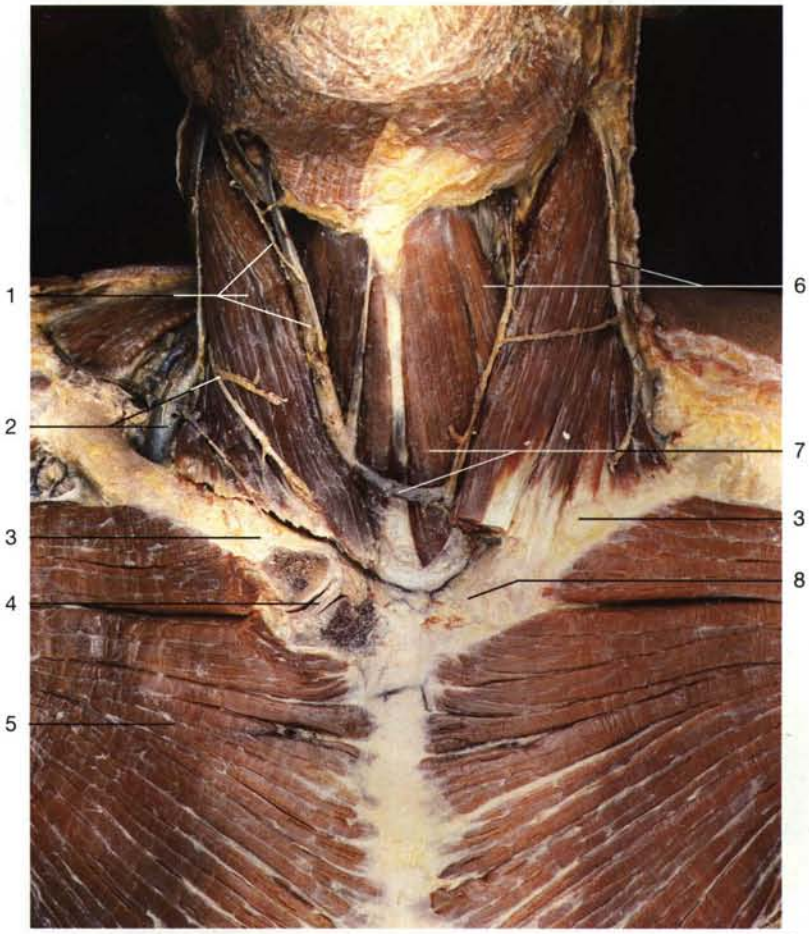

Right sternoclavicular joint (anterior aspect). On the right side the joint has been opened by a coronal section. Note the articular disk.

1 Sternocleidomastoid muscle, cervical branch of facial nerve and anterior jugular vein
2 External jugular vein and transverse cervical nerve (inferior branch)
3 Clavicle
4 **Sternoclavicular joint** (opened) with articular disk
5 Pectoralis major muscle
6 Omohyoid muscle and external jugular vein
7 Jugular venous arch and sternohyoid muscle
8 Sternoclavicular joint (not opened)
9 Acromial end of clavicle
10 **Acromioclavicular joint**
11 Acromion
12 Tendon of supraspinatus muscle (attached to the articular capsule)
13 Coraco-acromial ligament
14 Tendon of long head of biceps brachii muscle
15 Tendon of subscapularis muscle (attached to the articular capsule)
16 Intertubercular sulcus
17 Articular capsule of shoulder joint
18 Humerus
19 Trapezoid ligament
20 Coracoid process
21 Glenoid labrum
22 **Shoulder joint** (joint cavity)
23 Scapula
24 Supraspinatus muscle
25 Cartilage of glenoid cavity
26 Tendon of long head of triceps brachii muscle
27 Head of humerus (articular cartilage)
28 Epiphyseal line

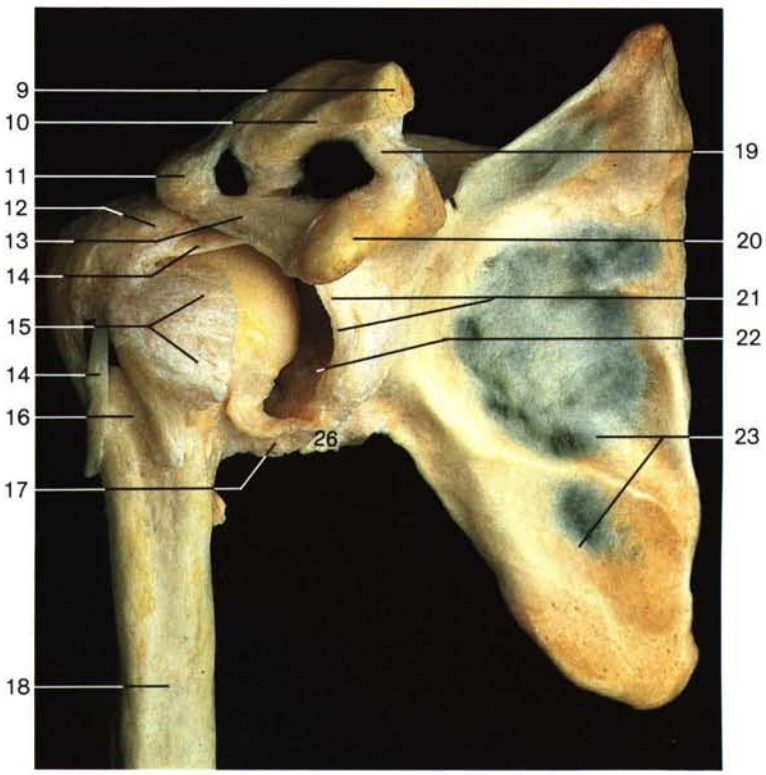

Right shoulder joint. The anterior part of the articular capsule has been removed and the head of the humerus has been slightly rotated outward to show the cavity of the joint.

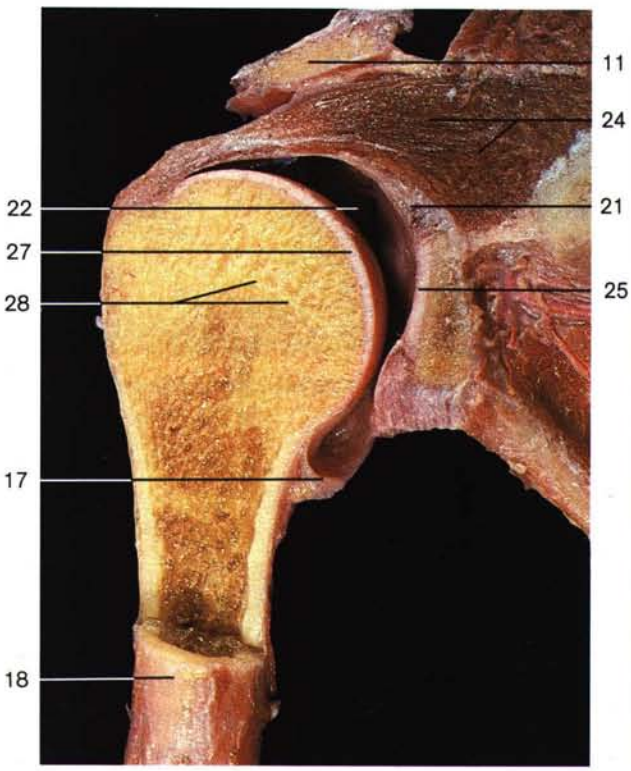

Coronal section of the right shoulder joint (anterior aspect).

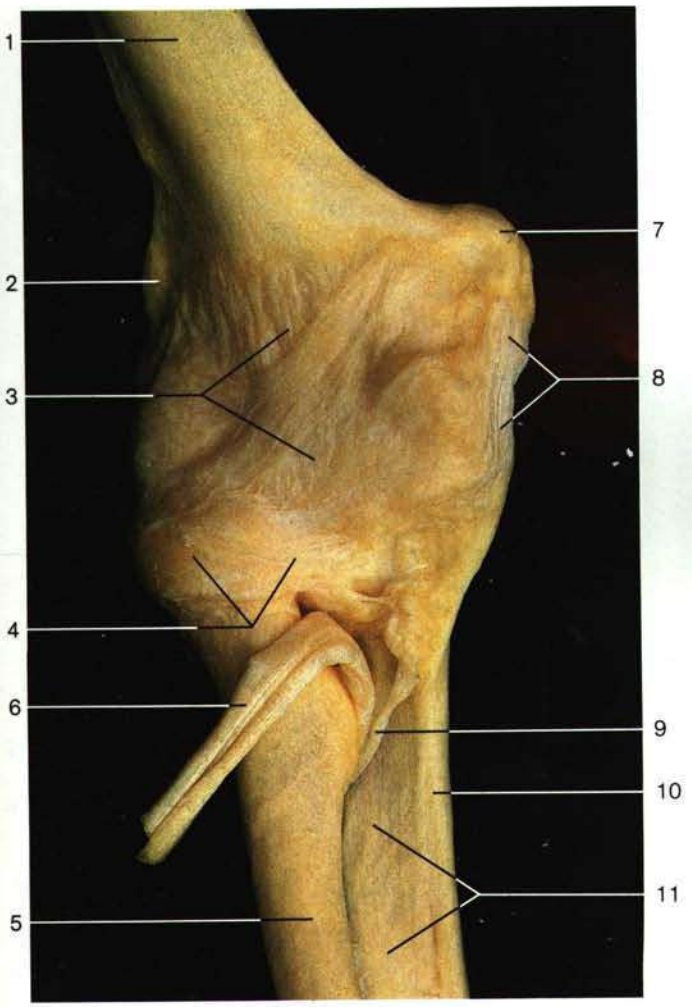

1 Humerus
2 Lateral epicondyle of humerus
3 Articular capsule
4 Annular ligament of proximal radioulnar joint
5 Radius
6 Tendon of biceps brachii muscle
7 Medial epicondyle of humerus
8 Ulnar collateral ligament
9 Oblique cord
10 Ulna
11 Interosseous membrane
12 Radial fossa
13 Capitulum of humerus
14 Head of radius
15 Radial collateral ligament
16 Coronoid fossa
17 Trochlea of humerus
18 Coronoid process of ulna
19 Olecranon
20 Radial tuberosity

Ligaments of elbow joint (anterior aspect).

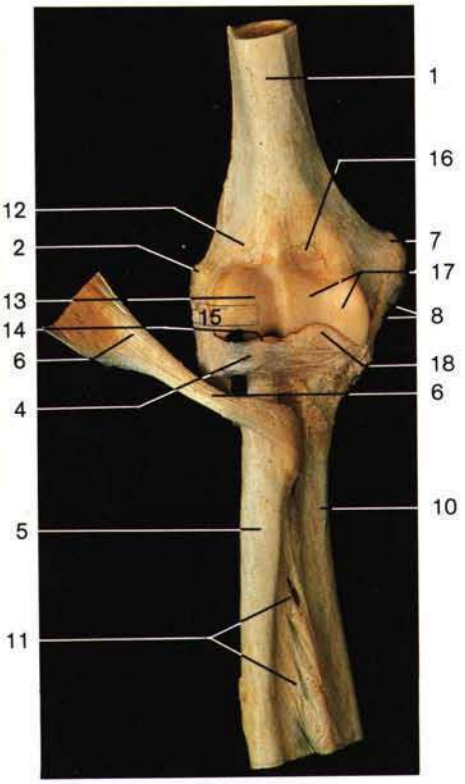

Elbow joint with ligaments (anterior aspect). Articular capsule has been removed to show the annular ligament.

Elbow joint with collateral ligaments (medial aspect).

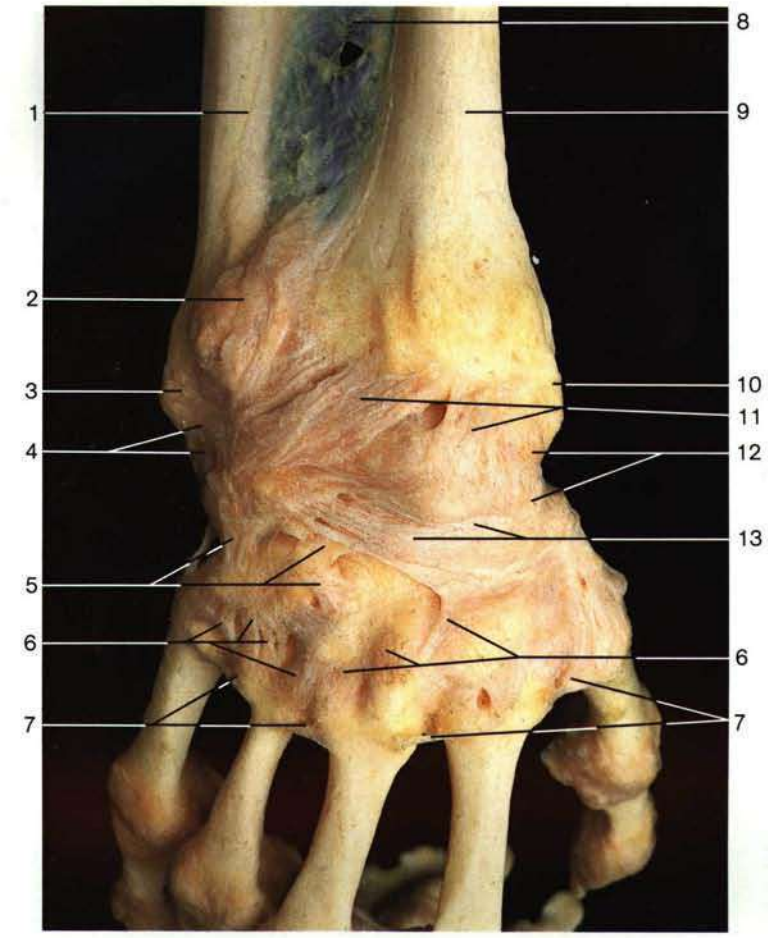

Ligaments of hand and wrist (dorsal aspect).

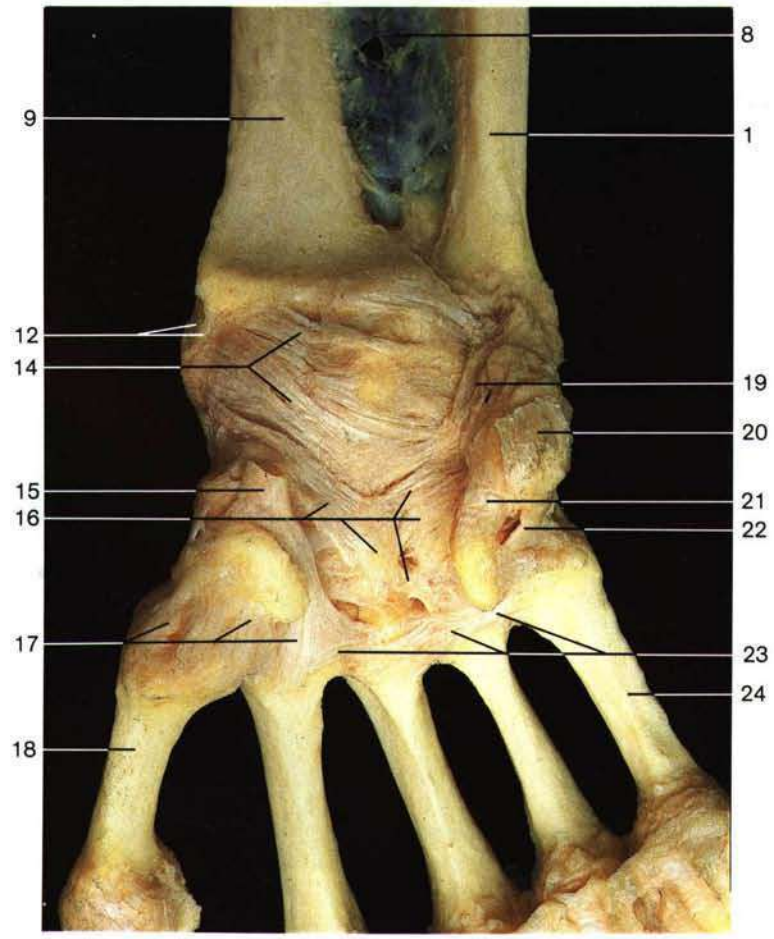

Ligaments of hand and wrist (palmar aspect).

1 Ulna
2 Exostosis (pathological)
3 Head of ulna
4 **Ulnar carpal collateral ligament**
5 Deep intercarpal ligaments
6 Dorsal carpometacarpal ligaments
7 Dorsal metacarpal ligaments
8 Interosseous membrane
9 Radius
10 Styloid process of radius
11 **Dorsal radiocarpal ligament**
12 Radial collateral ligament
13 Articular capsule and dorsal intercarpal ligaments
14 **Palmar radiocarpal ligament**
15 Tendon of flexor carpi radialis muscle (cut)
16 Radiating carpal ligament
17 Palmar carpometacarpal ligaments
18 1st metacarpal bone
19 Palmar ulnocarpal ligament
20 Tendon of flexor carpi ulnaris muscle (cut)
21 **Pisohamate ligament**
22 **Pisometacarpal ligament**
23 Palmar metacarpal ligaments
24 5th metacarpal bone

1 Radius
2 Styloid process of radius
3 Palmar radiocarpal ligament
4 Tendon of flexor carpi radialis muscle (cut)
5 Radiating carpal ligament
6 Articular capsule of **carpometacarpal joint of thumb**
7 Articular capsule of **metacarpophalangeal joint of thumb**
8 Palmar ligaments and articular capsule of metacarpophalangeal joints
9 Palmar ligaments and articular capsule of interphalangeal joints
10 Articular capsule
11 Interosseous membrane
12 Ulna
13 **Distal radioulnar joint**
14 Styloid process of ulna
15 Palmar ulnocarpal ligament
16 Pisiform bone with tendon of flexor carpi ulnaris muscle
17 Pisometacarpal ligament
18 Pisohamate ligament
19 Metacarpal bone
20 Deep transverse metacarpal ligament
21 Tendons of extensor muscles and articular capsule
22 Collateral ligament of **interphalangeal joint**
23 Collateral ligaments of **metacarpophalangeal joints**
24 Second metacarpal bone

Ligaments of right forearm, hand and fingers (palmar aspect). The arrow indicates the location of the carpal tunnel.

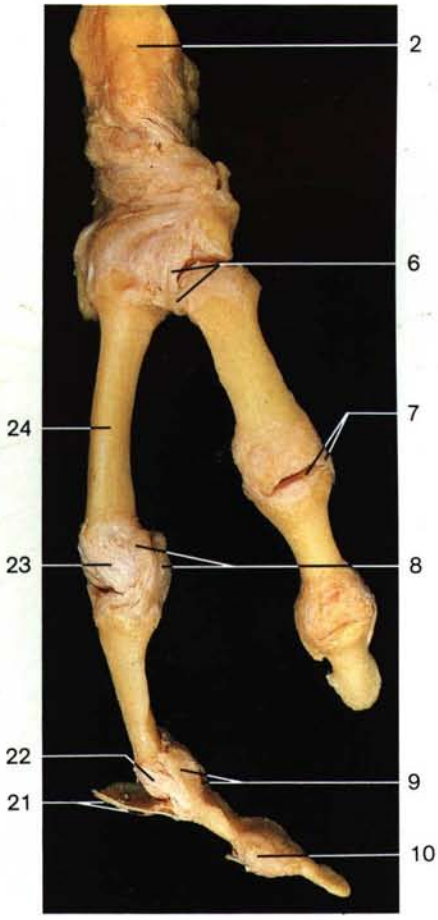

Ligaments of fingers (lateral aspect).

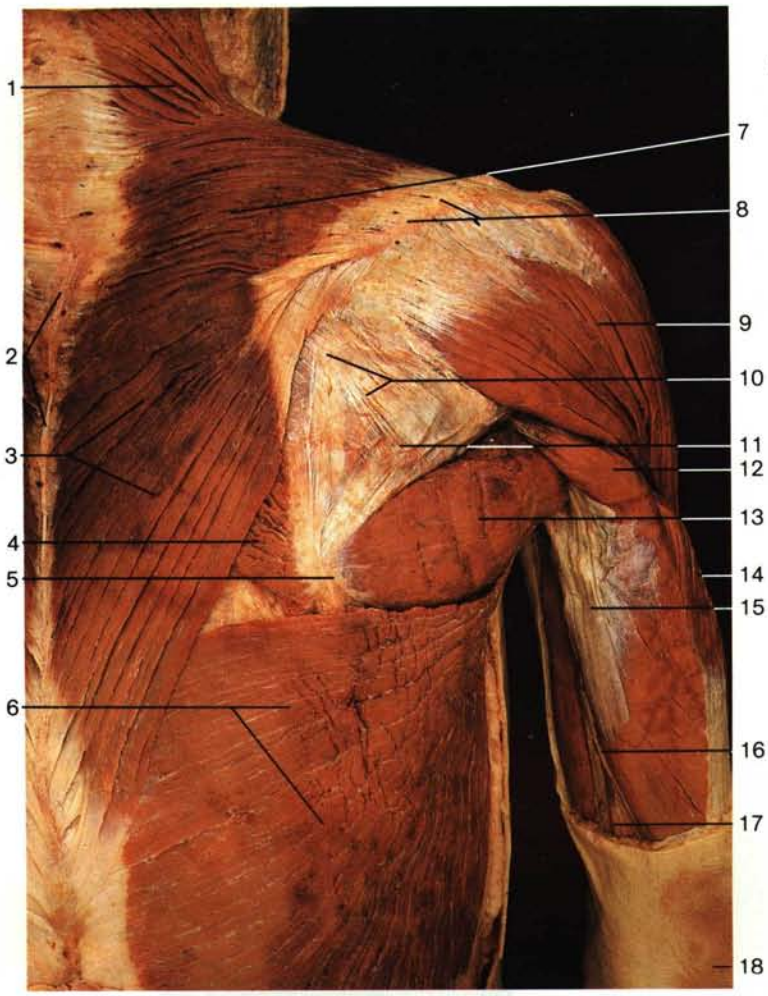

1 Descending fibers of **trapezius muscle**
2 Spinous processes of thoracic vertebrae
3 Ascending fibers of trapezius muscle
4 Rhomboid major muscle
5 Inferior angle of scapula
6 **Latissimus dorsi muscle**
7 Transverse fibers of trapezius muscle
8 Spine of scapula
9 Posterior fibers of **deltoid muscle**
10 Infraspinatus muscle and infraspinous fascia
11 Teres minor muscle and fascia
12 Long head of triceps brachii muscle
13 **Teres major muscle**
14 Lateral head of triceps brachii muscle
15 Medial head of triceps brachii muscle
16 Medial intermuscular septum
17 Ulnar nerve
18 Olecranon

Muscles of shoulder and arm, superficial layer (dorsal aspect).

1 Trapezius muscle (reflected)
2 Levator scapulae muscle
3 Supraspinatus muscle
4 Rhomboid minor muscle
5 Medial border of scapula
6 **Rhomboid major muscle**
7 **Infraspinatus muscle**
8 **Teres major muscle**
9 Inferior angle of scapula
10 Cut edge of trapezius muscle
11 Intrinsic muscles of back with fascia
12 **Latissimus dorsi muscle**
13 Acromion
14 Spine of scapula
15 **Deltoid muscle**
16 Teres minor
17 Long head of triceps brachii muscle
18 Lateral head of triceps brachii muscle
19 Medial head of triceps brachii muscle
20 Medial intermuscular septum
21 Tendon of triceps brachii muscle

Muscles of shoulder and arm, deeper layer (right side, dorsal aspect). The trapezius has been cut near its origin at the vertebral column and reflected upward.

1 **Splenius capitis muscle**
2 Sternocleidomastoideus muscle
3 Trapezius muscle (reflected)
4 **Lateral supraclavicular nerves**
5 Clavicle
6 Levator scapulae muscle
7 Supraspinatus muscle
8 Spine of scapula
9 Deltoid muscle (reflected)
10 Rhomboid minor muscle
11 **Rhomboid major muscle**
12 **Axillary nerve and posterior circumflex humeral artery**
13 **Infraspinatus muscle**
14 **Teres minor muscle**
15 Long head of triceps brachii muscle
16 **Teres major muscle**
17 Inferior angle of scapula
18 Triceps brachii muscle
19 **Latissimus dorsi muscle**

Muscles of shoulder and arm, deeper layer, right side (dorsal aspect). The trapezius and deltoid muscles have been divided and reflected.

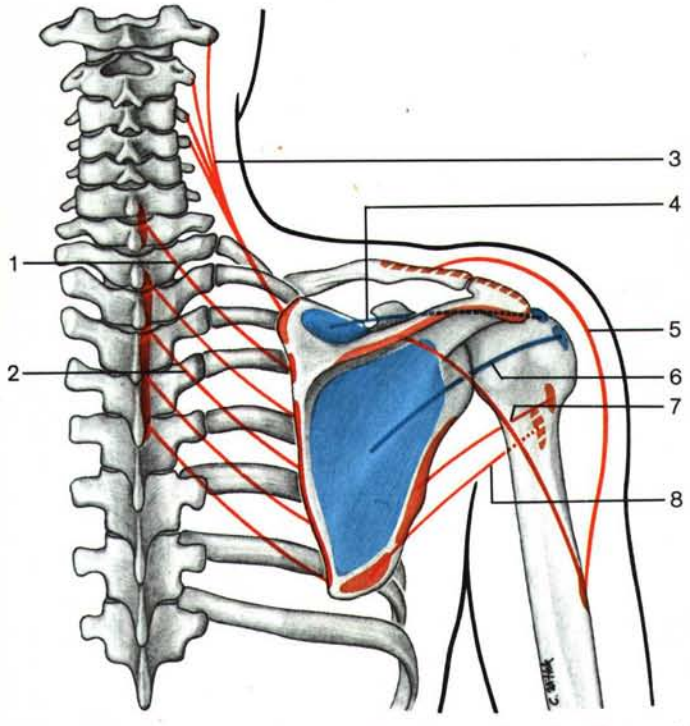

Shoulder muscles, schematic diagram illustrating the course of the main muscles of the dorsal aspect of the shoulder.

1 Rhomboid minor muscle (red)
2 Rhomboid major muscle (red)
3 Levator scapulae muscle (red)
4 Supraspinatus muscle (blue)
5 Deltoid muscle (red)
6 Infraspinatus muscle (blue)
7 Teres minor muscle (red)
8 Teres major muscle (red)

Muscles of shoulder and arm, superficial layer (ventral aspect).

1 Trapezius muscle
2 Acromion
3 Deltopectoral triangle
4 Clavicular part of deltoid muscle (anterior fibers)
5 Acromial part of deltoid muscle (central fibers)
6 Clavicular part of pectoralis major muscle
7 Sternocostal part of pectoralis major muscle
8 Short head of biceps brachii muscle
9 Long head of biceps brachii muscle
10 Abdominal part of pectoralis major muscle
11 Brachialis muscle
12 Serratus anterior muscle
13 External abdominal oblique muscle
14 Sternocleidomastoid muscle
15 Infrahyoid muscles
16 Clavicle
17 Manubrium sterni
18 Body of sternum
19 Xiphoid process
20 Anterior layer of sheath of rectus abdominis muscle

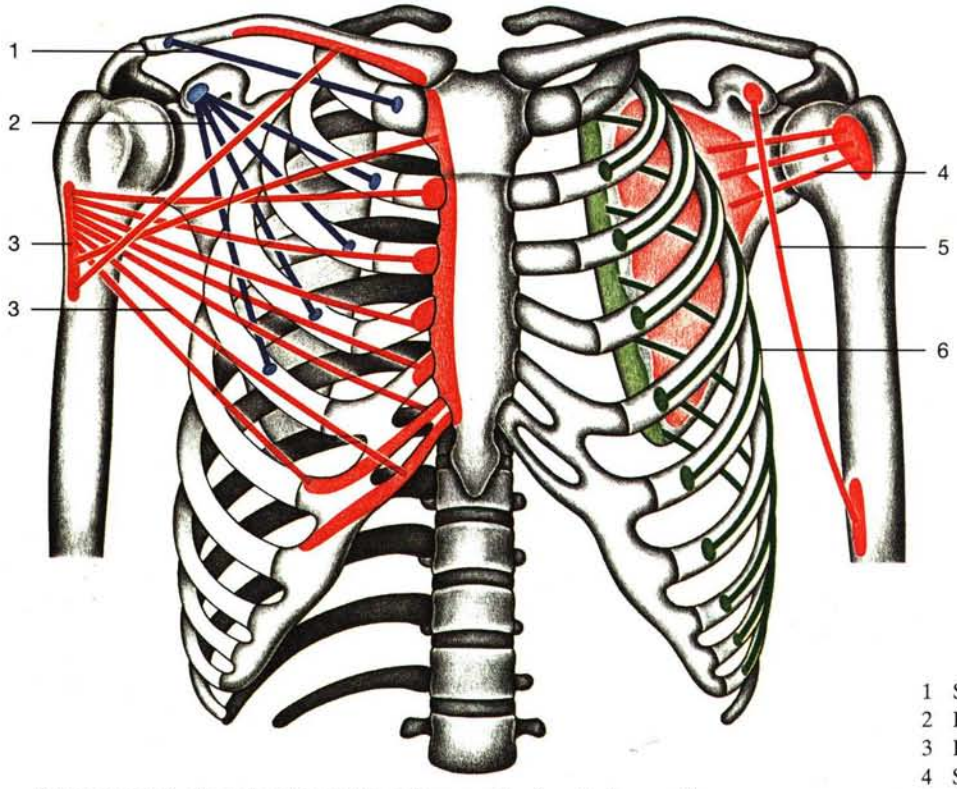

Arrangement of pectoral and shoulder muscles (ventral aspect). (Schematic drawing.)

1 Subclavius muscle (blue)
2 Pectoralis minor muscle (blue)
3 Pectoralis major muscle (red)
4 Subscapularis muscle (red)
5 Coracobrachialis muscle (light-blue)
6 Serratus anterior muscle (green)

Muscles of shoulder and arm, deep layer (ventral aspect).

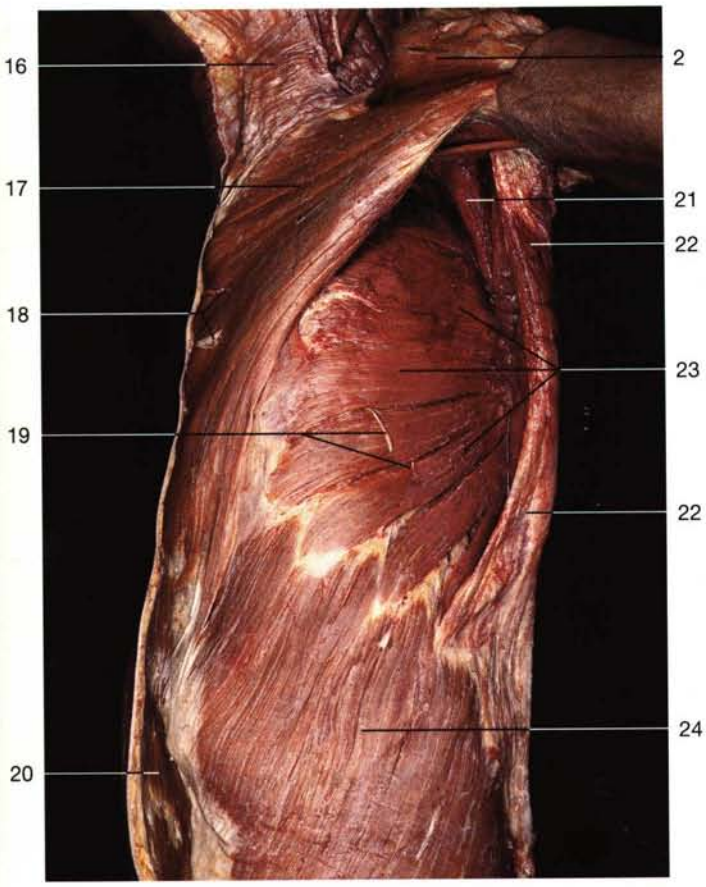

1 Acromion
2 Clavicular part of deltoid muscle
3 Pectoralis major muscle (reflected)
4 Coracobrachialis muscle
5 Short head of biceps brachii muscle
6 Deltoid muscle (insertion on humerus)
7 Long head of biceps brachii muscle
8 Brachialis muscle
9 Sternocleidomastoid muscle
10 Clavicle
11 Subclavius muscle
12 Pectoralis minor muscle
13 Sternum
14 3rd rib
15 Pectoralis major muscle
16 Platysma muscle
17 Pectoralis major muscle forming the anterior axillary fold
18 Anterior cutaneous branches of intercostal nerves
19 Lateral cutaneous branches of intercostal nerves
20 Rectus abdominis muscle
21 Subscapularis muscle
22 **Latissimus dorsi muscle** forming the posterior axillary fold
23 **Serratus anterior muscle** forming the medial wall of the axilla
24 External abdominal oblique muscle

Axillary fossa and serratus anterior muscle (left side, lateral aspect).

Muscles of right arm (lateral aspect).

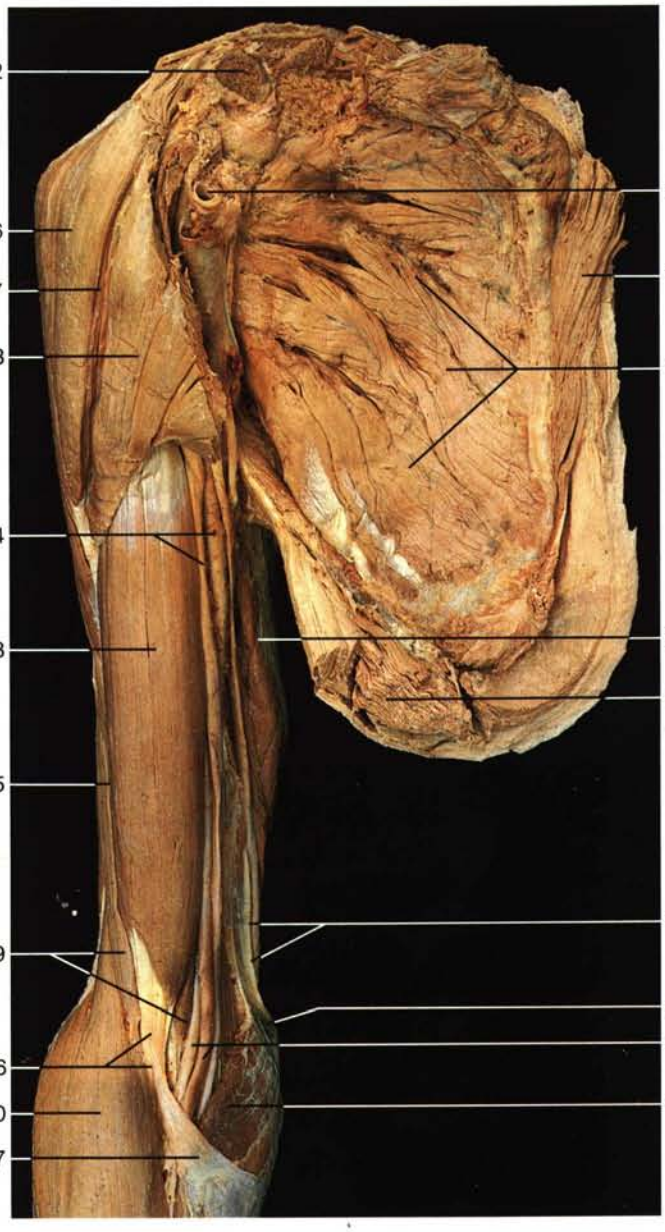

Muscles of right arm (ventral aspect). The arm with the scapula and attached muscles has been removed from the trunk.

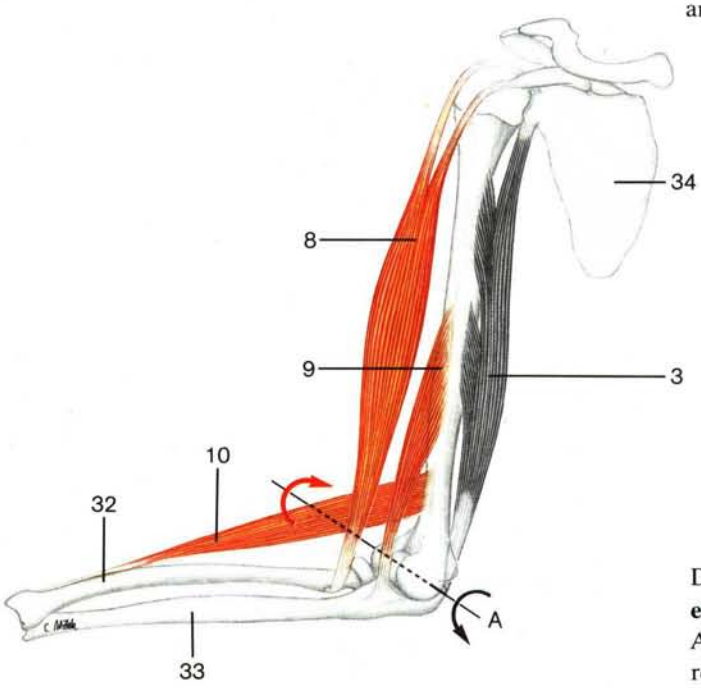

Diagram illustrating the position of the **flexor and extensor muscles of the arm** and their effect on the elbow joint. A = axis; arrows = direction of movements; red = flexion; black = extension.

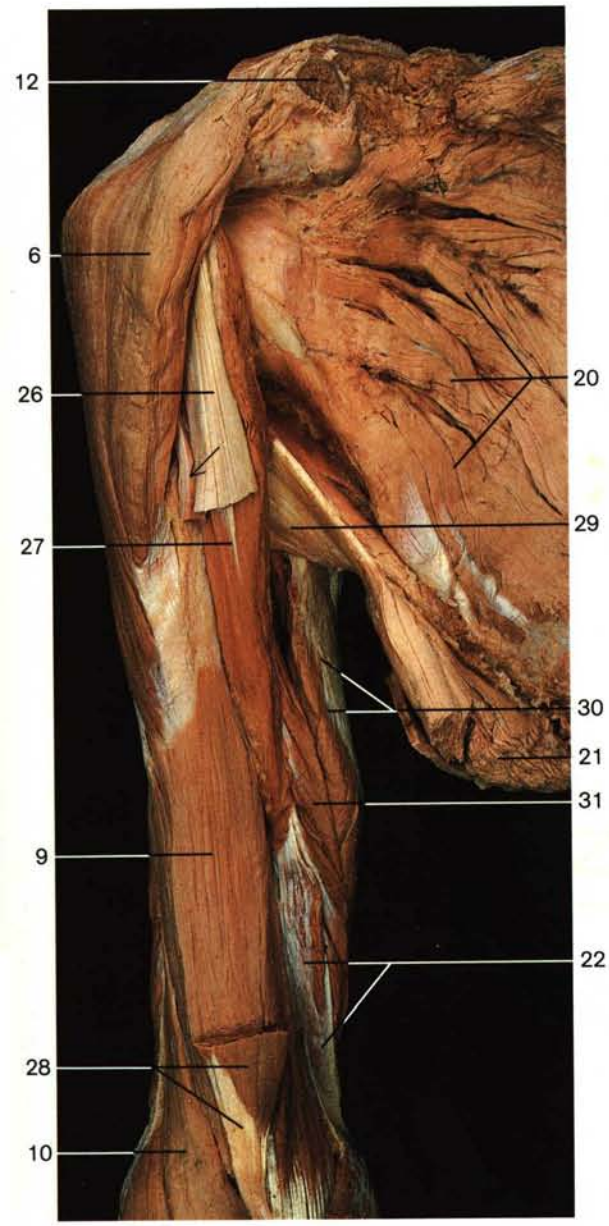

Flexor muscles of right arm (ventral aspect). Part of the biceps brachii muscle has been removed. Arrow: tendon of long head of biceps brachii muscle.

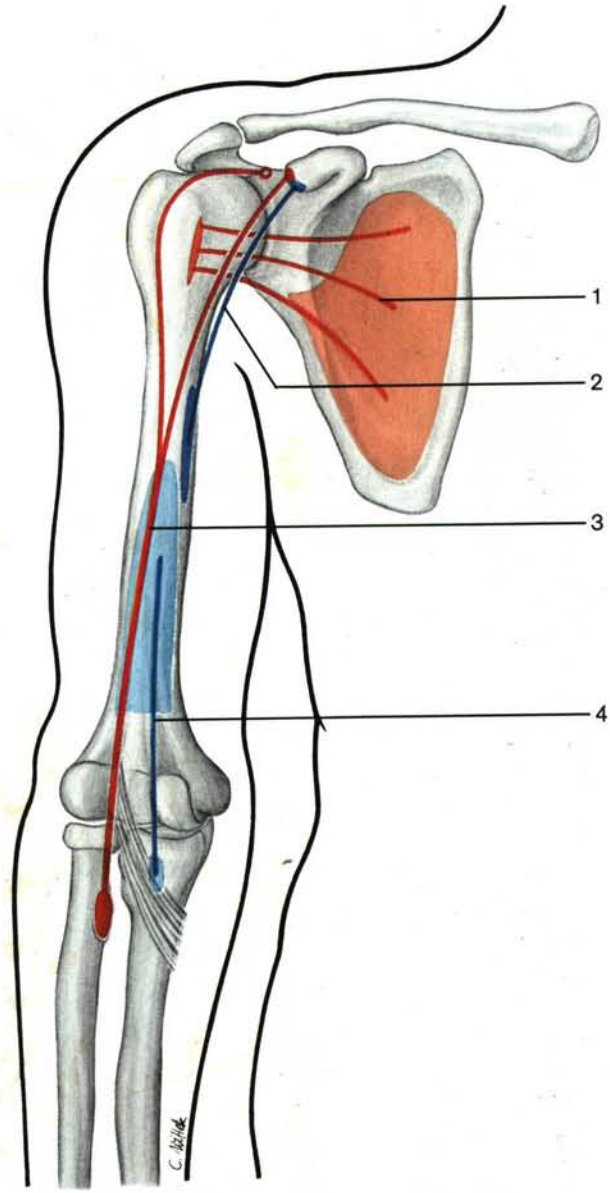

Position and course of flexors of arm. (Schematic drawing.)

1 Subscapularis muscle (red)
2 Coracobrachialis muscle (blue)
3 Biceps brachii muscle (red)
4 Brachialis muscle (blue)

1 Acromial part of deltoid muscle (central fibers)
2 Scapular part of deltoid muscle (posterior fibers)
3 Triceps brachii muscle
4 Tendon of triceps brachii muscle
5 Olecranon
6 Clavicular part of deltoid muscle (anterior fibers)
7 Deltopectoral groove
8 **Biceps brachii muscle**
9 **Brachialis muscle**
10 **Brachioradialis muscle**
11 Extensor carpi radialis longus muscle
12 Clavicle (divided)
13 Pectoralis major muscle
14 Medial intermuscular septum with vessels and nerves
15 Lateral intermuscular septum
16 Tendon of biceps brachii muscle
17 Bicipital aponeurosis
18 Axillary artery
19 Rhomboid major muscle
20 **Subscapularis muscle**
21 Latissimus dorsi muscle (divided)
22 Medial intermuscular septum
23 Medial epicondyle of humerus
24 Brachial artery and median nerve
25 Pronator teres muscle
26 Tendon of short head of biceps brachii muscle
27 Coracobrachialis muscle
28 Distal part of biceps brachii muscle
29 Teres major muscle
30 Long head of triceps brachii muscle
31 Medial head of triceps brachii muscle
32 Radius
33 Ulna
34 Scapula

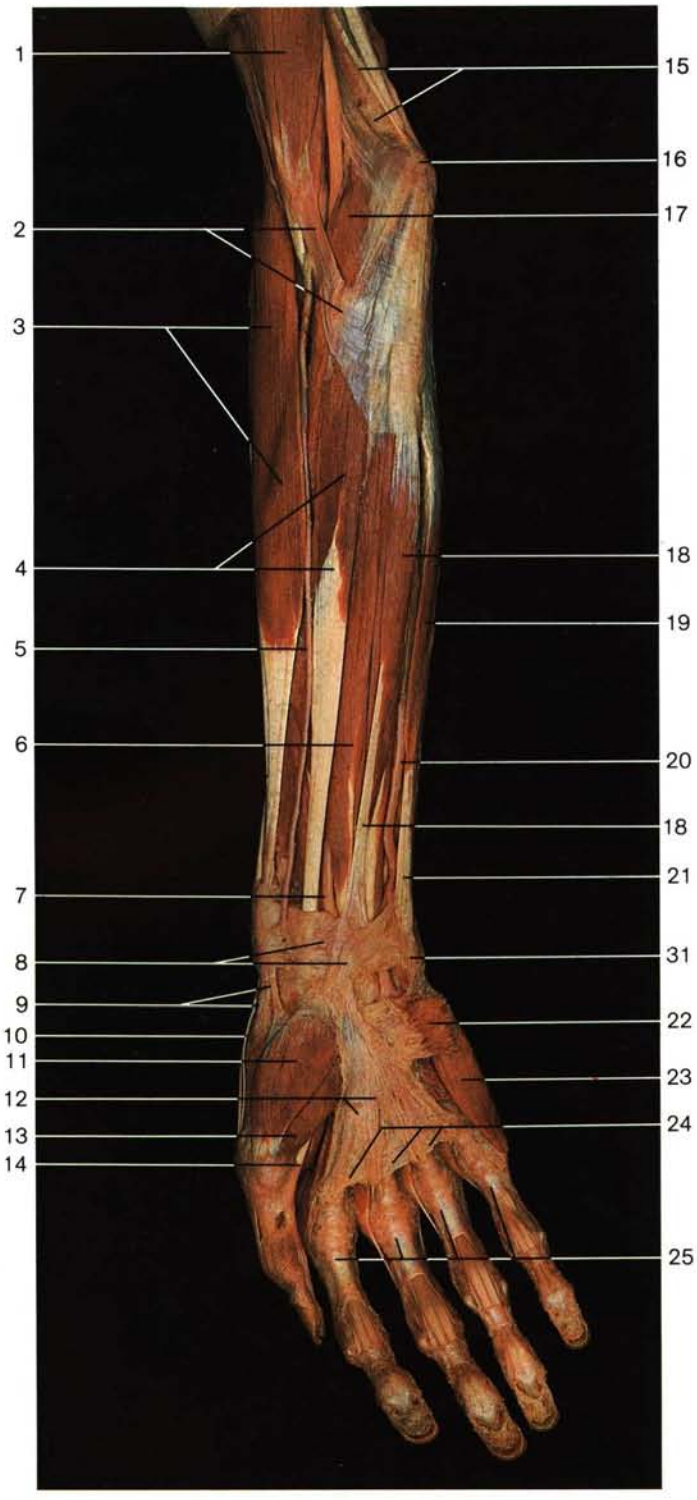

Forearm muscles, superficial layer (ventral aspect).

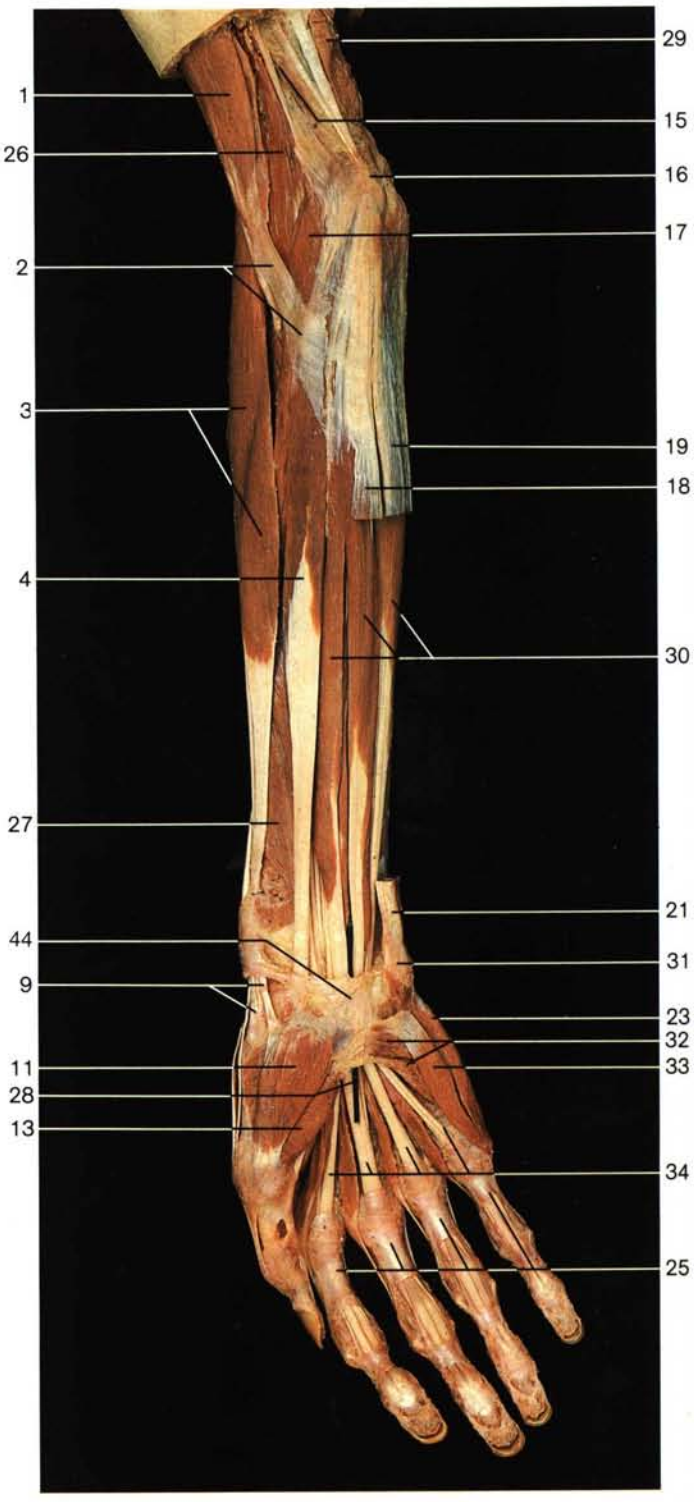

Forearm muscles, superficial layer (ventral aspect). The palmaris longus and flexor carpi ulnaris muscles have been removed.

1 Biceps brachii muscle
2 Bicipital aponeurosis
3 **Brachioradialis muscle**
4 **Flexor carpi radialis muscle**
5 Radial artery
6 **Flexor digitorum superficialis muscle**
7 Median nerve
8 Antebrachial fascia and tendon of palmaris longus muscle
9 Tendon of abductor pollicis longus muscle
10 Tendon of extensor pollicis brevis muscle
11 Abductor pollicis brevis muscle
12 **Palmar aponeurosis**
13 Superficial head of flexor pollicis brevis muscle
14 Tendon of flexor pollicis longus muscle
15 Medial intermuscular septum
16 Medial epicondyle of humerus
17 Humeral head of pronator teres muscle
18 **Palmaris longus muscle**
19 **Flexor carpi ulnaris muscle**
20 **Ulnar artery**
21 Tendon of flexor carpi ulnaris muscle
22 Palmaris brevis muscle

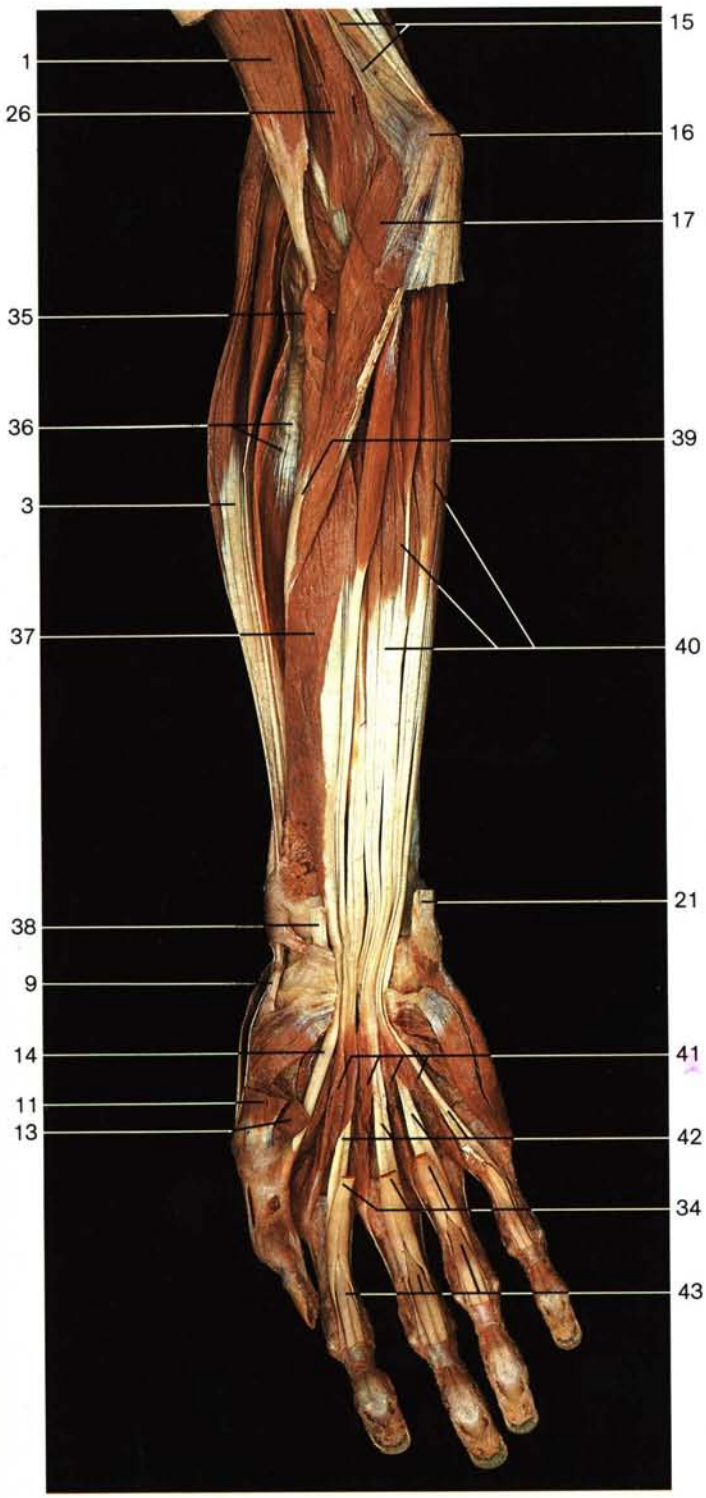

Forearm muscles, middle layer (ventral aspect). The palmaris longus, flexor carpi radialis and ulnaris muscles have been partly removed. The flexor retinaculum has been divided.

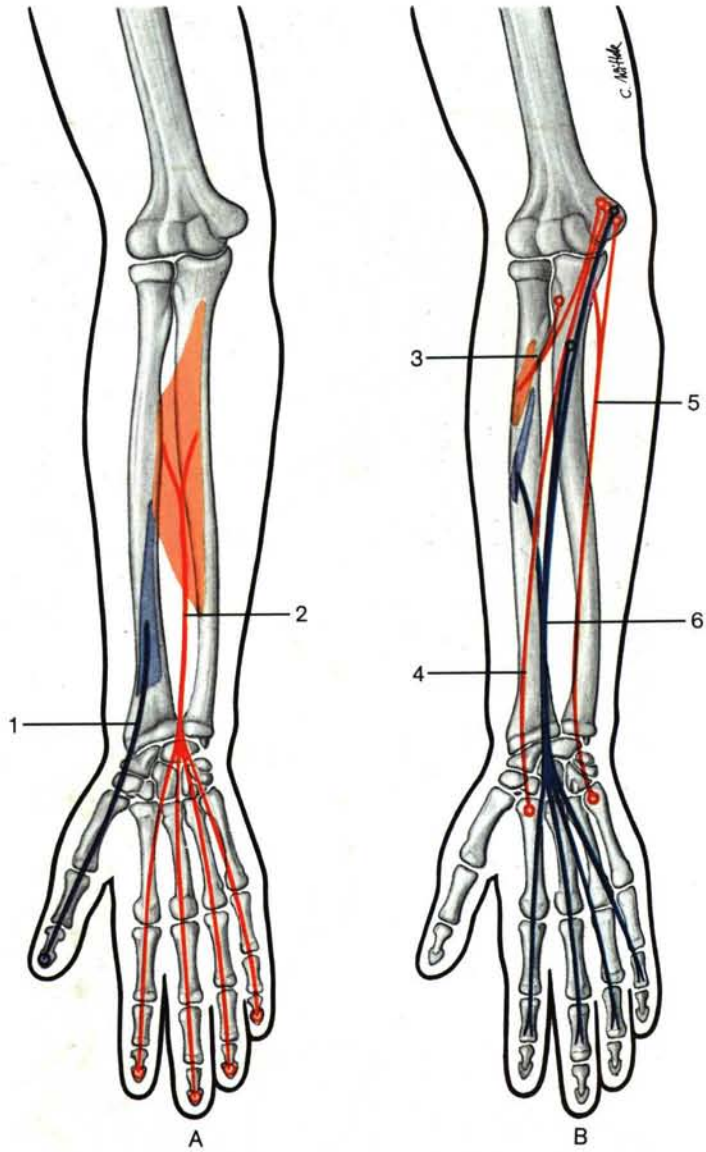

Position of flexors of fingers and hand. (Schematic drawing.)

A Deep layer
B Superficial layer
△
1 Flexor pollicis longus muscle (blue)
2 Flexor digitorum profundus muscle (red)
3 Pronator teres muscle (red)
4 Flexor carpi radialis muscle (red)
5 Flexor carpi ulnaris muscle (red)
6 Flexor digitorum superficialis muscle (blue)

23 Abductor digiti minimi muscle
24 Transverse fasciculi of palmar aponeurosis
25 Digital fibrous sheaths of tendons of flexor digitorum muscle
26 Brachialis muscle
27 Flexor pollicis longus muscle
28 Carpal tunnel (canalis carpi, probe)
29 Triceps brachii muscle
30 Flexor digitorum superficialis muscle
31 Pisiform bone
32 Opponens digiti minimi muscle
33 Flexor digiti minimi brevis muscle
34 Tendons of flexor digitorum superficialis muscle
35 Supinator muscle
36 Extensor carpi radialis brevis muscle
37 **Flexor pollicis longus muscle**
38 Tendon of flexor carpi radialis muscle
39 **Pronator teres muscle** (insertion of radius)
40 **Flexor digitorum profundus muscle**
41 **Lumbrical muscles**
42 Tendons of flexor digitorum profundus muscle
43 Tendons of flexor digitorum profundus muscle having passed through the divided tendons of the flexor digitorum superficialis muscle
44 Flexor retinaculum

1 Biceps brachii muscle
2 Brachialis muscle
3 **Pronator teres muscle**
4 Brachioradialis muscle
5 **Radius**
6 Tendon of flexor carpi radialis muscle
7 Tendon of abductor pollicis longus muscle
8 Opponens pollicis muscle
9 Adductor pollicis muscle
10 Tendon of flexor pollicis longus muscle
11 Triceps brachii muscle
12 Medial intermuscular septum
13 Medial epicondyle of humerus
14 Common flexor mass (divided)
15 **Ulna**
16 **Interosseous membrane**
17 **Pronator quadratus muscle**
18 Tendon of flexor carpi ulnaris muscle
19 Pisiform bone
20 Abductor digiti minimi muscle
21 Flexor digiti minimi brevis muscle
22 Tendons of flexor digitorum profundus muscle
23 Tendons of flexor digitorum superficialis muscle
24 Flexor retinaculum
25 Hypothenar muscles
26 Thenar muscles
27 Common synovial sheath of flexor tendons
28 Synovial sheath of tendon of flexor pollicis longus muscle
29 Digital synovial sheaths of flexor tendons

Muscles of the forearm, deep layer (ventral aspect). All flexors have been removed to display the pronator quadratus and pronator teres muscles together with the interosseous membrane. Forearm in supination.

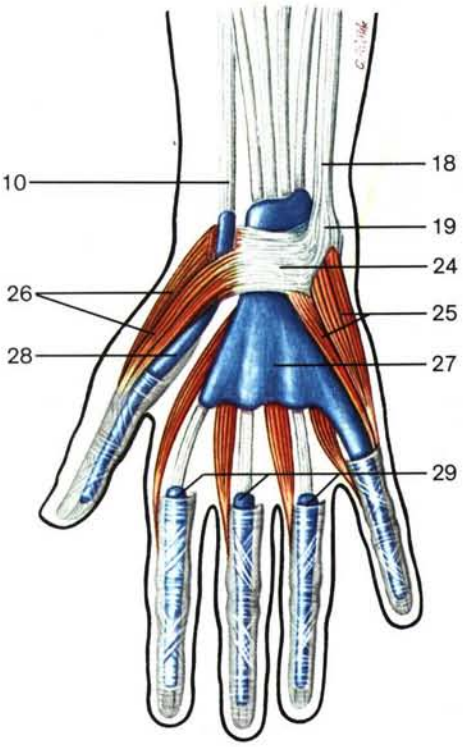

Synovial sheaths of flexor tendons (palmar aspect of right hand). (Semischematic drawing.)

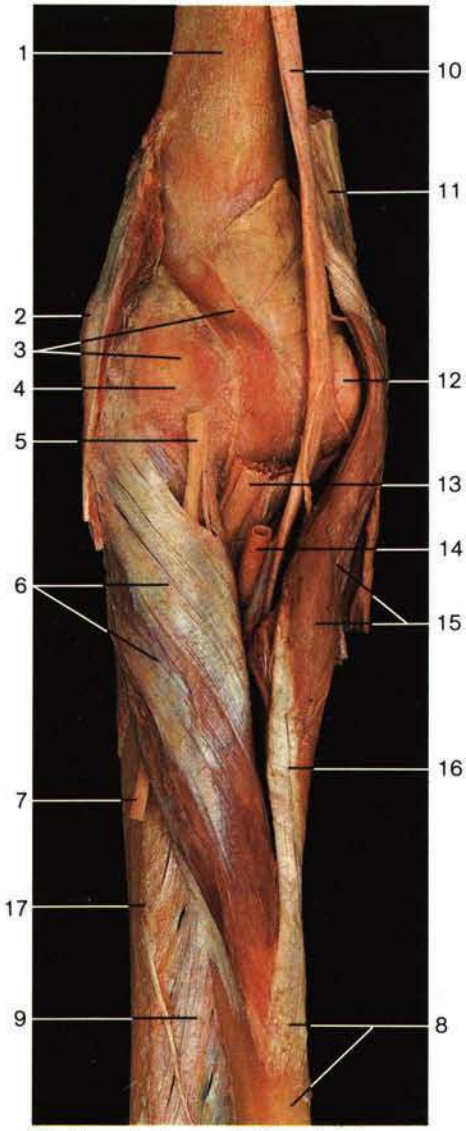

Right supinator and elbow joint
(anterior aspect). Forearm in pronation.

1 Humerus
2 Lateral epicondyle of humerus
3 Articular capsule
4 Position of capitulum of humerus
5 Deep branch of radial nerve
6 **Supinator muscle**
7 Entrance of deep branch of radial nerve to extensor muscles
8 Radius and insertion of pronator teres muscle
9 Interosseous membrane
10 Median nerve
11 Triceps brachii muscle
12 Trochlea of humerus
13 Tendon of biceps brachii muscle
14 Brachial artery
15 **Pronator teres muscle**
16 Tendon of pronator teres muscle
17 Ulna
18 **Pronator quadratus muscle**
19 Tendon of flexor carpi radialis muscle
20 Thenar muscles
21 Synovial sheath of tendon of flexor pollicis longus muscle
22 Fibrous sheath of flexor tendons
23 **Digital synovial sheath of flexor tendons**
24 Flexor digitorum superficialis muscle
25 Tendon of flexor carpi ulnaris muscle
26 **Common synovial sheath of flexor tendons**
27 Position of pisiform bone
28 Flexor retinaculum
29 Hypothenar muscles

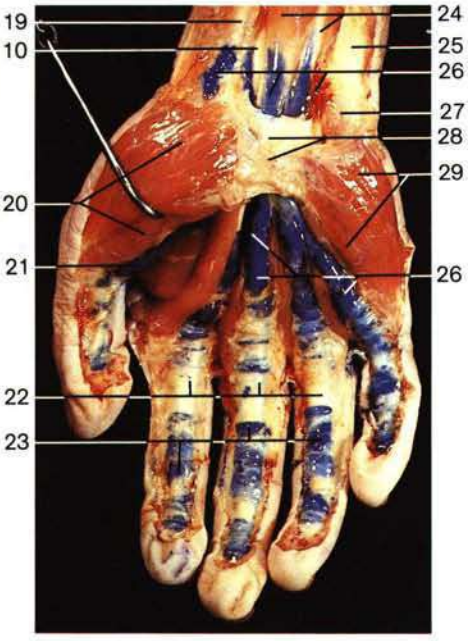

Synovial sheaths of flexor muscles
(palmar aspect of right hand).
Blue PVA-solution has been injected into the sheaths.

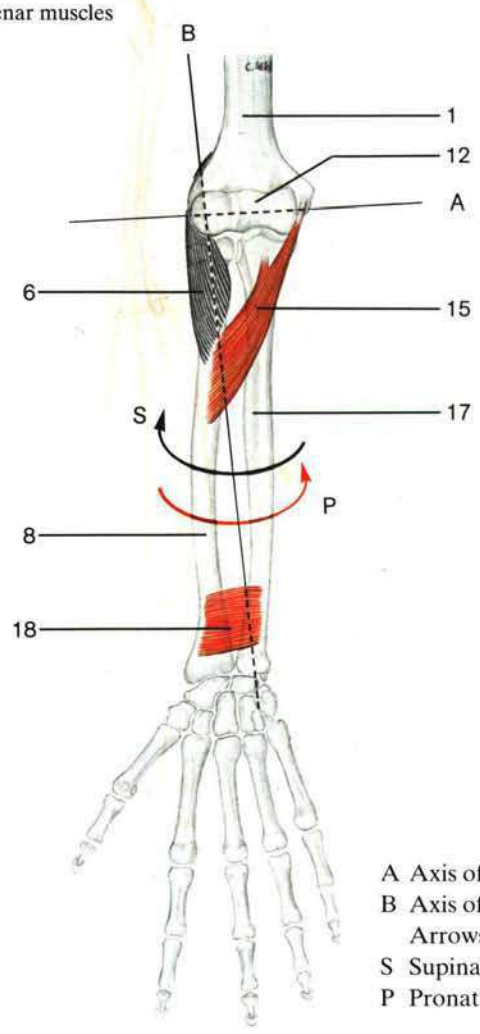

A Axis of flexion and extension
B Axis of rotation
Arrows:
S Supination
P Pronation

Diagram illustrating the two **axes of the elbow joint.**

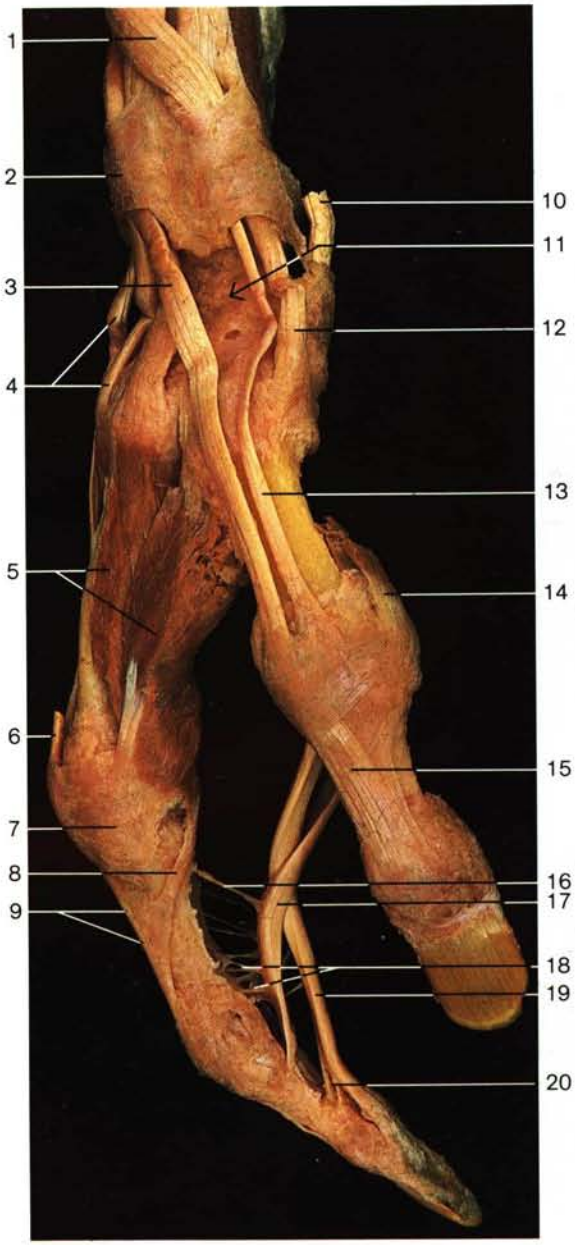

Muscles of thumb and index finger (medial aspect). The tendons of the extensor muscles of the thumb and the insertion of the flexor tendons of the index finger are displayed.

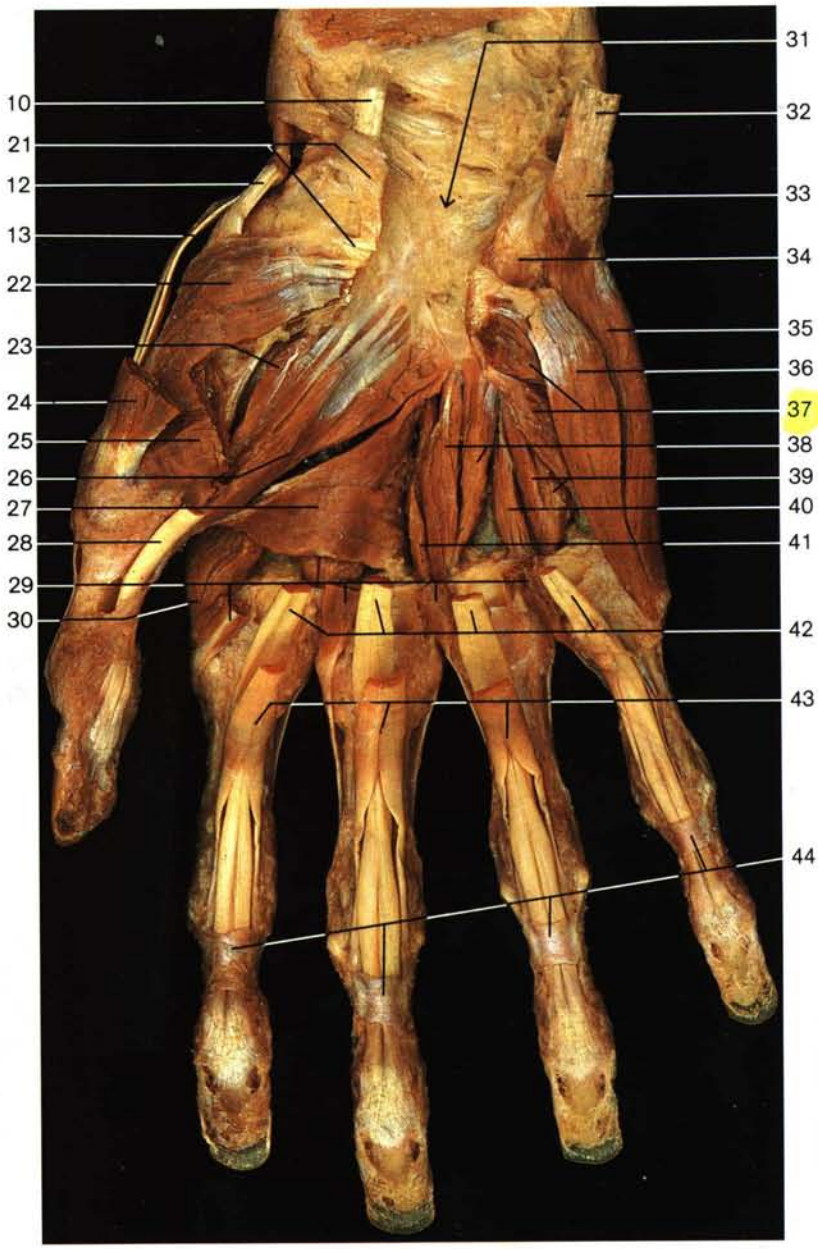

Muscles of hand (palmar aspect). The tendons of the flexor muscles and parts of the thumb muscles have been removed. The carpal tunnel has been opened.

1 Tendons of extensor pollicis brevis and abductor pollicis longus muscle
2 Extensor retinaculum
3 Tendon of extensor pollicis longus muscle
4 Tendons of extensor carpi radialis longus and brevis muscles
5 First dorsal interosseous muscle
6 Tendon of extensor digitorum muscle for index finger
7 Location of metacarpophalangeal joint
8 Tendon of lumbrical muscle
9 Extensor expansion of index finger
10 Tendon of flexor carpi radialis muscle (cut)
11 **Anatomical snuffbox**
12 Tendon of abductor pollicis longus muscle
13 Tendon of extensor pollicis brevis muscle
14 Tendon of abductor pollicis brevis muscle
15 Extensor expansion of extensor of thumb
16 Vinculum longum
17 Tendons of flexor digitorum superficialis muscle dividing to allow passage of deep tendons
18 Vincula of flexor tendons
19 Tendon of flexor digitorum profundus muscle
20 Vinculum breve
21 Radial carpal eminence (cut edge of flexor retinaculum)
22 Opponens pollicis muscle
23 Deep head of flexor pollicis brevis muscle
24 Abductor pollicis brevis muscle (cut)
25 Superficial head of flexor pollicis brevis muscle (cut)
26 Oblique head of adductor pollicis muscle
27 Transverse head of adductor pollicis muscle
28 Tendon of flexor pollicis longus muscle (cut)
29 Lumbrical muscles (cut)
30 First dorsal interosseous muscle
31 **Position of carpal tunnel**
32 Tendon of flexor carpi ulnaris muscle
33 Location of pisiform bone
34 Hook of hamate bone
35 Abductor digiti minimi muscle
36 Flexor digiti minimi brevis muscle
37 Opponens digiti minimi muscle
38 2nd palmar interosseous muscle
39 3rd palmar interosseous muscle
40 4th dorsal interosseous muscle
41 3rd dorsal interosseous muscle
42 Tendon of flexor digitorum profundus muscle (cut)
43 Tendons of flexor digitorum superficialis muscle (cut)
44 Fibrous flexor sheaths

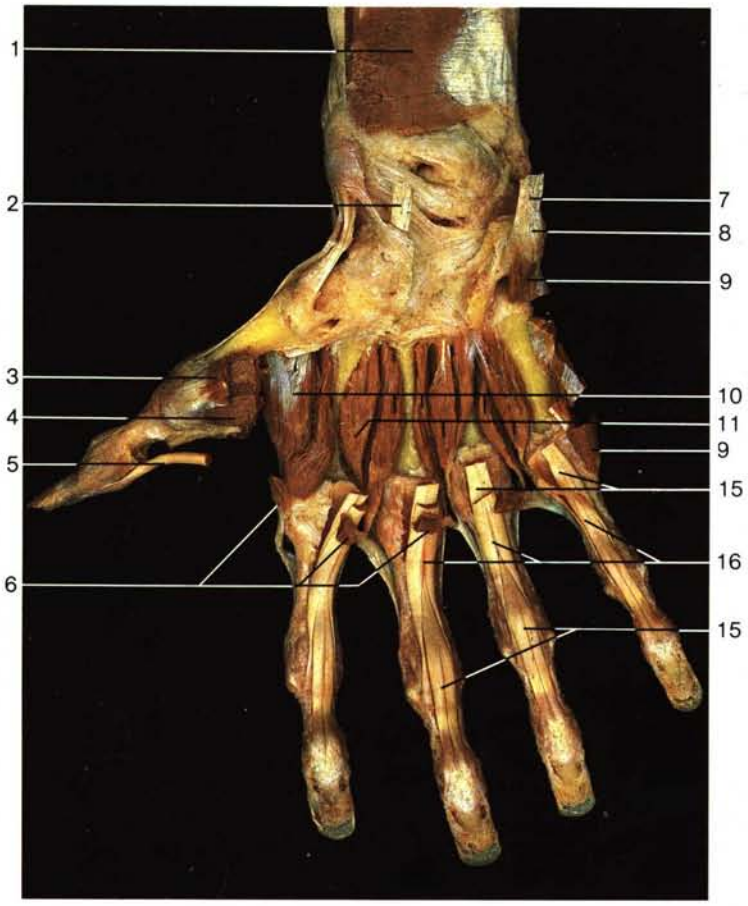

Muscles of right hand, deep layer (palmar aspect). The thenar and hypothenar muscles have been removed to display the interosseous muscles.

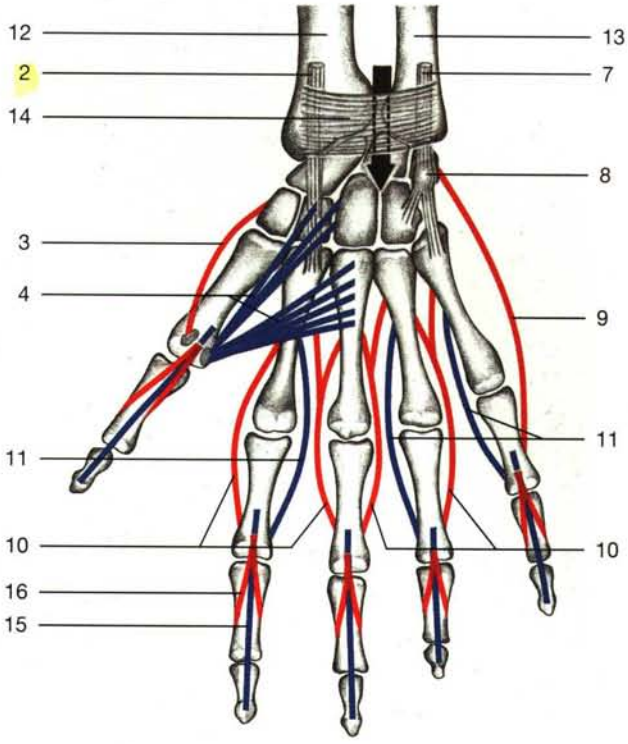

Actions of interosseous muscles in abduction and adduction of fingers (palmar aspect, schematic drawing). Arrow: carpal tunnel.
Red = **abduction**
(dorsal interosseous muscles, abductor digiti minimi and abductor pollicis brevis muscles).
Black = **adduction**
(palmar interosseous muscles, adductor pollicis muscle).

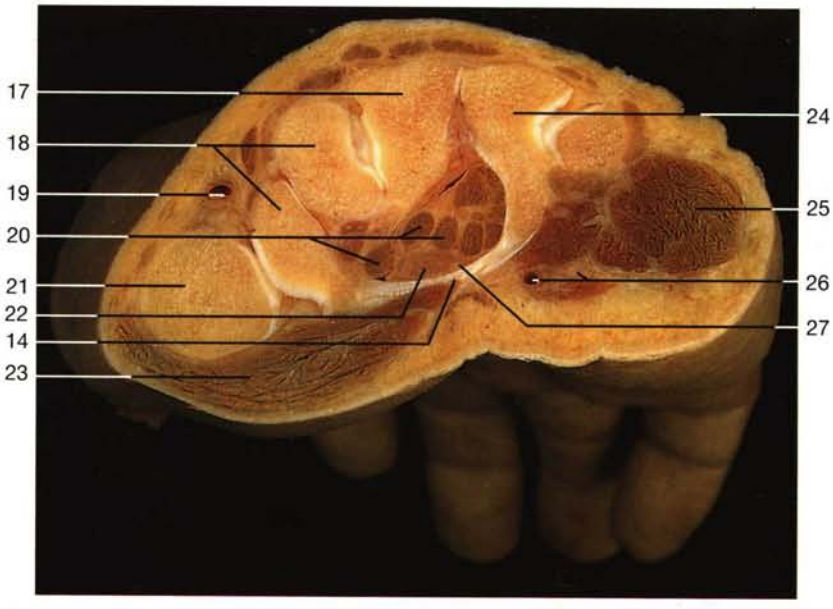

Transverse section through the right hand, showing the carpal tunnel (canalis carpi).

1 Pronator quadratus muscle
2 Tendon of flexor carpi radialis muscle
3 Abductor pollicis brevis muscle (divided)
4 Adductor pollicis muscle (divided)
5 Tendon of flexor pollicis longus muscle
6 Lumbrical muscles (cut)
7 Tendon of flexor carpi ulnaris muscle
8 Pisiform bone
9 Abductor digiti minimi muscle (divided)
10 **Dorsal interosseous muscles**
11 **Palmar interosseous muscles**
12 Radius
13 Ulna
14 **Flexor retinaculum**
15 Tendons of flexor digitorum profundus muscle
16 Tendons of flexor digitorum superficialis muscle
17 Capitate bone
18 Trapezium bone and trapezoid bone
19 Radial artery
20 Tendon of flexor muscles
21 First metacarpal bone
22 **Median nerve**
23 Thenar muscles
24 Hamate bone
25 Hypothenar muscles
26 Ulnar artery and nerve
27 **Carpal tunnel** (canalis carpi)

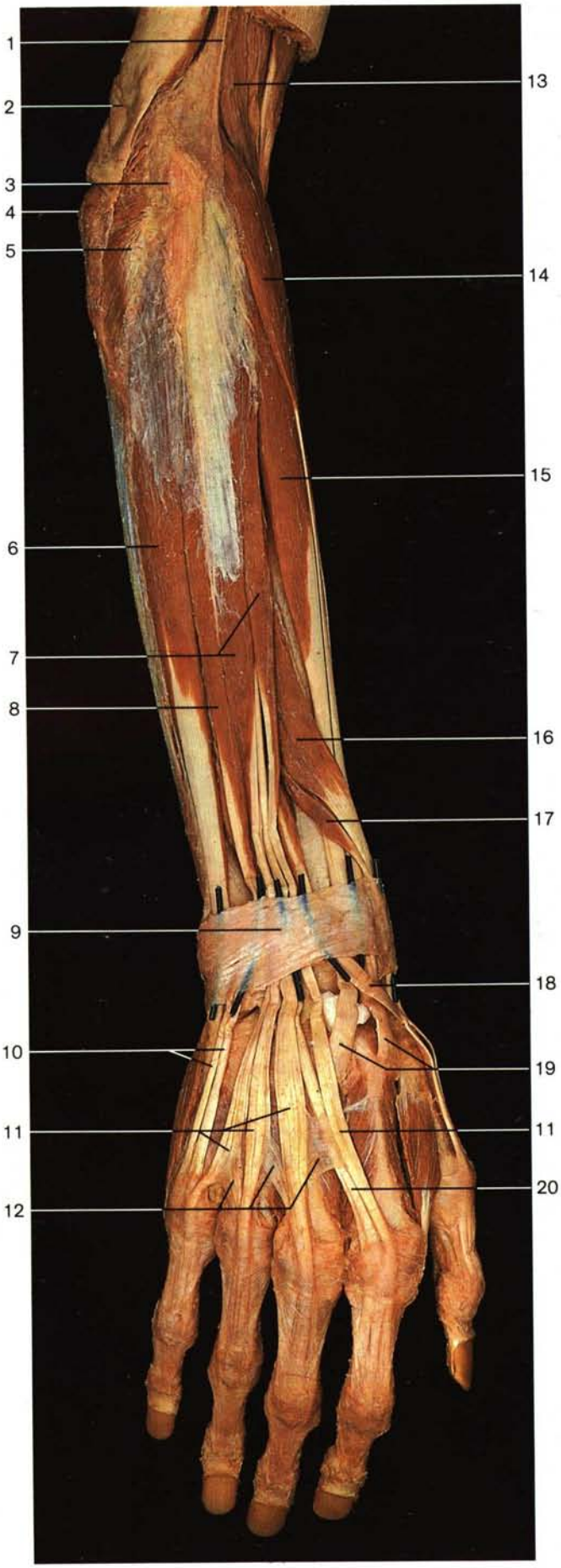

Extensor muscles of forearm and hand, superficial layer (dorsal aspect). Tunnels for extensor tendons indicated by probes.

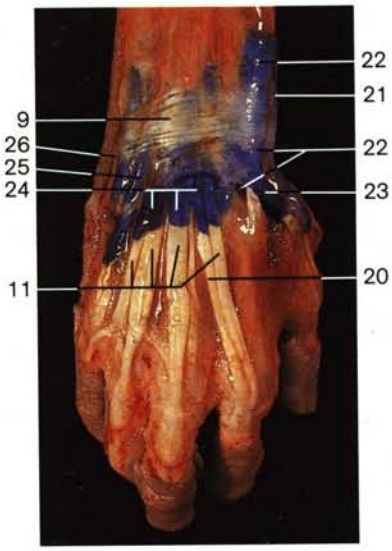

Synovial sheaths of extensor tendons. The sheaths have been injected with blue gelatin.

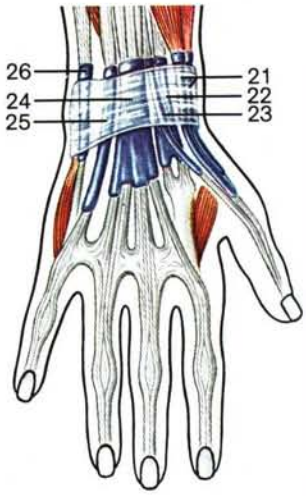

Synovial sheaths of extensor tendons on the back of right wrist (indicated in blue). Notice the six tunnels for the passage of the extensor tendons beneath the extensor retinaculum. (Semischematic drawing.)

1 Lateral intermuscular septum
2 Tendon of triceps brachii muscle
3 Lateral epicondyle of humerus
4 Olecranon
5 Anconeus
6 Extensor carpi ulnaris muscle
7 Extensor digitorum muscle
8 Extensor digiti minimi muscle
9 Extensor retinaculum
10 Tendons of extensor digiti minimi muscle
11 Tendons of extensor digitorum muscle
12 Intertendinous connections
13 Brachioradialis muscle
14 Extensor carpi radialis longus muscle
15 Extensor carpi radialis brevis muscle
16 Abductor pollicis longus muscle
17 Extensor pollicis brevis muscle
18 Tendon of extensor pollicis longus muscle
19 Tendons of both extensor carpi radialis longus and extensor carpi radialis brevis muscles
20 Tendon of extensor indicis muscle
21 1st tunnel: Abductor pollicis longus muscle
Extensor pollicis brevis muscle
22 2nd tunnel: Extensor carpi radialis longus and brevis muscles
23 3rd tunnel: Extensor pollicis longus muscle
24 4th tunnel: Extensor digitorum muscle
Extensor indicis muscle
25 5th tunnel: Extensor digiti minimi muscle
26 6th tunnel: Extensor carpi ulnaris muscle

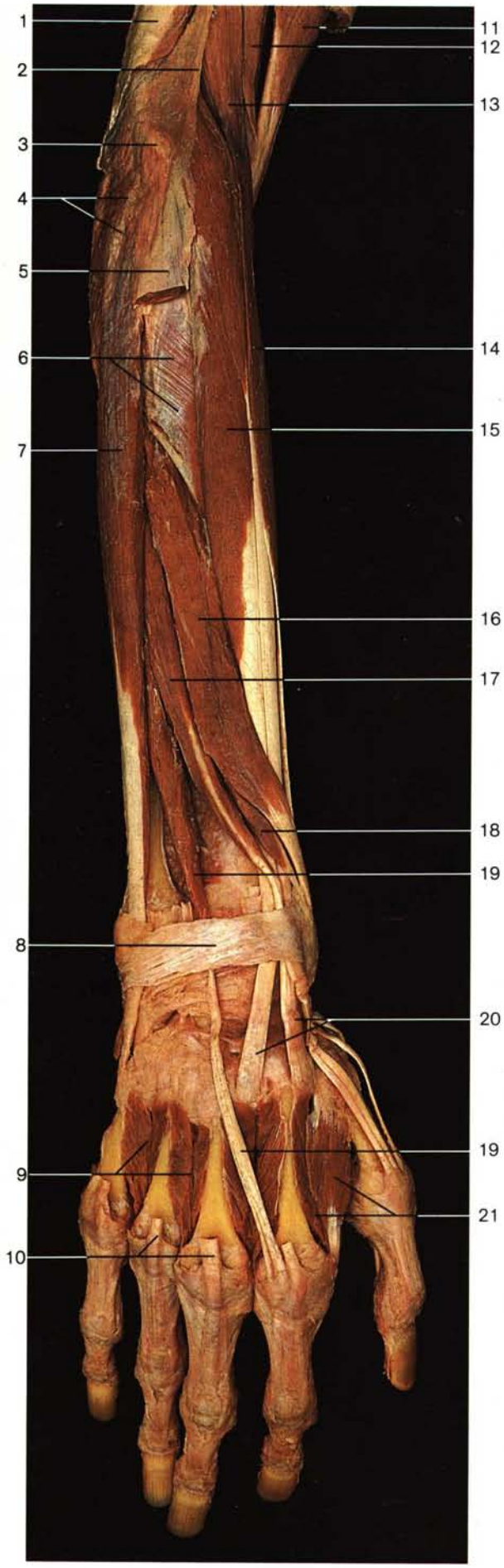

Extensor muscles of forearm and hand, deep layer (dorsal aspect).

1 Triceps brachii muscle
2 Lateral intermuscular septum
3 Lateral epicondyle of humerus
4 Anconeus muscle
5 Extensor digitorum and extensor digiti minimi muscles (cut)
6 Supinator muscle
7 Extensor carpi ulnaris muscle
8 Extensor retinaculum
9 3rd and 4th dorsal interosseous muscles
10 Tendons of extensor digitorum muscle (cut)
11 Biceps brachii muscle
12 Brachialis muscle
13 Brachioradialis muscle
14 Extensor carpi radialis longus muscle
15 Extensor carpi radialis brevis muscle
16 Abductor pollicis longus muscle
17 Extensor pollicis longus muscle
18 Extensor pollicis brevis muscle
19 Extensor indicis muscle
20 Tendons of the extensor carpi radialis longus and extensor carpi radialis brevis muscles
21 1st dorsal interosseous muscle

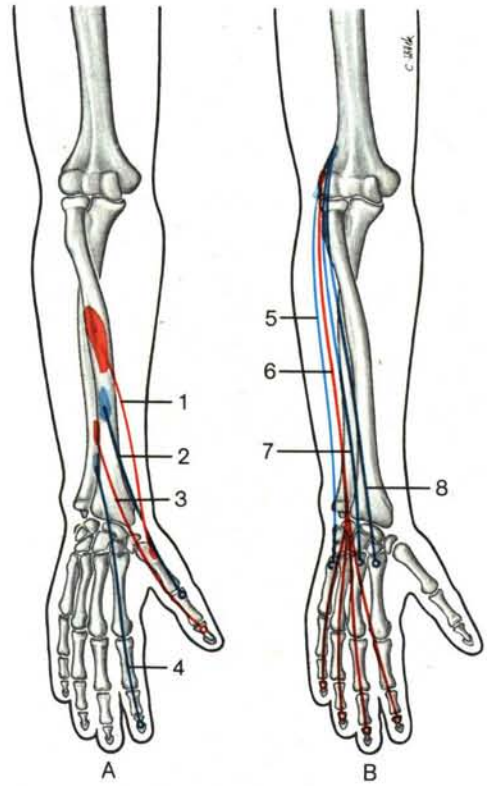

Position of extensor muscles of forearm and hand. (Semischematic drawing.)

A Extensors of thumb

1 Abductor pollicis longus muscle (red)
2 Extensor pollicis brevis muscle (blue)
3 Extensor pollicis longus muscle (red)
4 Extensor indicis muscle (blue)

B Extensors of fingers and hand

5 Extensor carpi ulnaris muscle (blue)
6 Extensor digitorum muscle (red)
7 Extensor carpi radialis brevis muscle (blue)
8 Extensor carpi radialis longus muscle (blue)

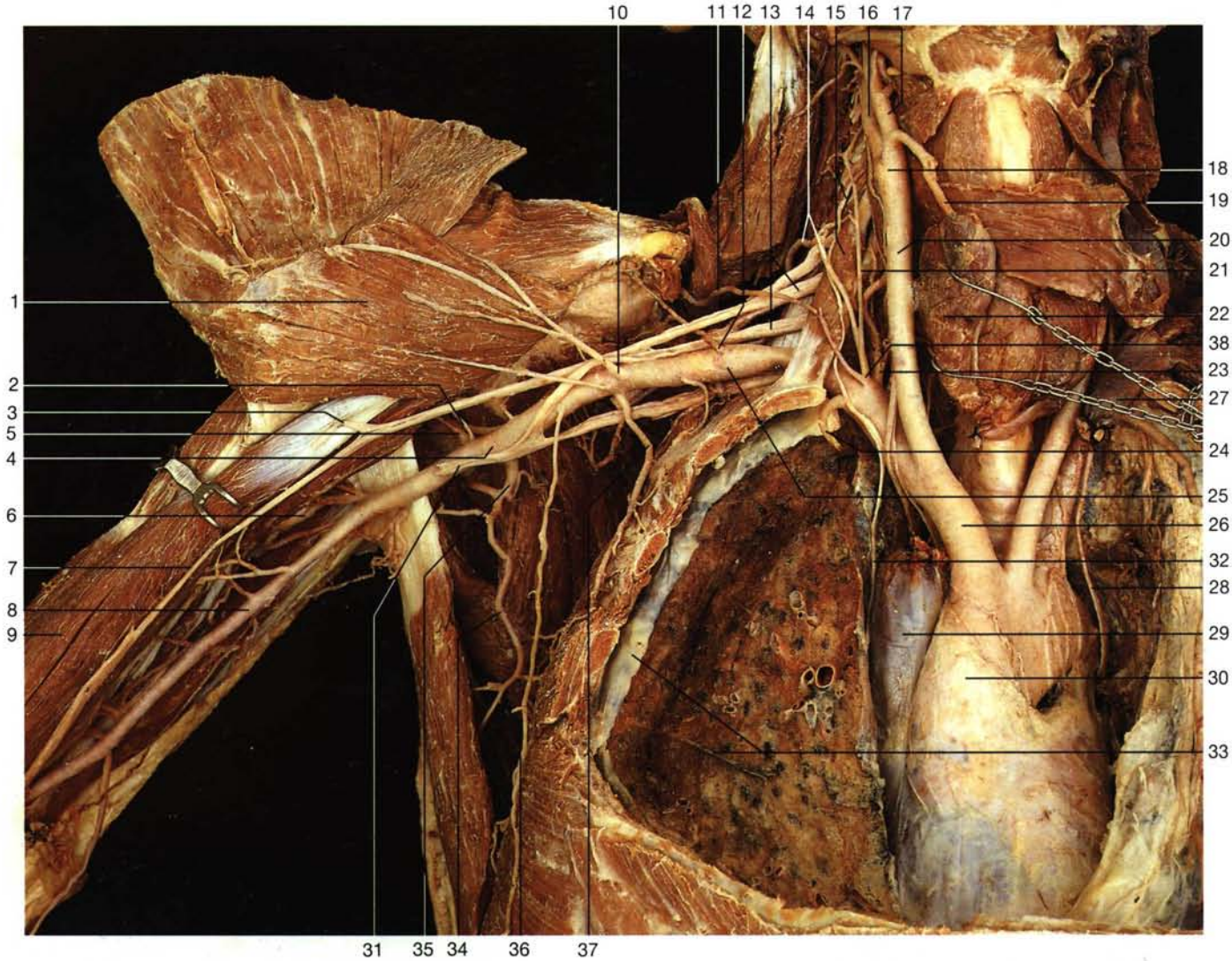

Main branches of right subclavian and axillary arteries (anterior aspect). Pectoralis muscles have been reflected, clavicle and anterior wall of thorax removed and right lung divided. Left lung with pleura and thyroid gland have been reflected laterally to display aortic arch and common carotid artery with their branches.

1 Pectoralis minor muscle (reflected)
2 Anterior circumflex humeral artery
3 Musculocutaneous nerve (divided)
4 **Axillary artery**
5 Posterior circumflex humeral artery
6 **Profunda brachii artery**
7 Median nerve (var.)
8 **Brachial artery**
9 Biceps brachii muscle
10 Thoracoacromial artery
11 Suprascapular artery
12 Descending scapular artery
13 Brachial plexus (middle trunk)
14 Transverse cervical artery
15 Scalenus anterior muscle and phrenic nerve
16 Right internal carotid artery
17 Right external carotid artery
18 Carotid sinus
19 Superior thyroid artery
20 Right common carotid artery
21 Ascending cervical artery
22 Thyroid gland
23 Inferior thyroid artery
24 Internal thoracic artery
25 Right subclavian artery
26 Brachiocephalic trunk
27 Left brachiocephalic vein (divided)
28 Left vagus nerve
29 Superior vena cava (divided)
30 **Ascending aorta**
31 Median nerve (divided)
32 Phrenic nerve
33 Right lung (divided) and pulmonary pleura
34 Thoracodorsal artery
35 Subscapular artery
36 Lateral mammary branches (variant)
37 Lateral thoracic artery
38 **Thyrocervical trunk**
39 Superior thoracic artery
40 Superior ulnar collateral artery
41 Inferior ulnar collateral artery
42 Middle collateral artery
43 Radial collateral artery
44 Radial recurrent artery
45 **Radial artery**
46 Anterior and posterior interosseous arteries
47 Princeps pollicis artery
48 **Deep palmar arch**
49 Common palmar digital arteries
50 Ulnar recurrent artery
51 Recurrent interosseous artery
52 Common interosseous artery
53 **Ulnar artery**
54 **Superficial palmar arch**
55 Median nerve and brachial artery
56 Biceps brachii muscle
57 Ulnar nerve
58 Flexor pollicis longus muscle
59 Palmar digital arteries
60 Anterior interosseous artery
61 Flexor carpi ulnaris muscle
62 Superficial palmar branch of radial artery

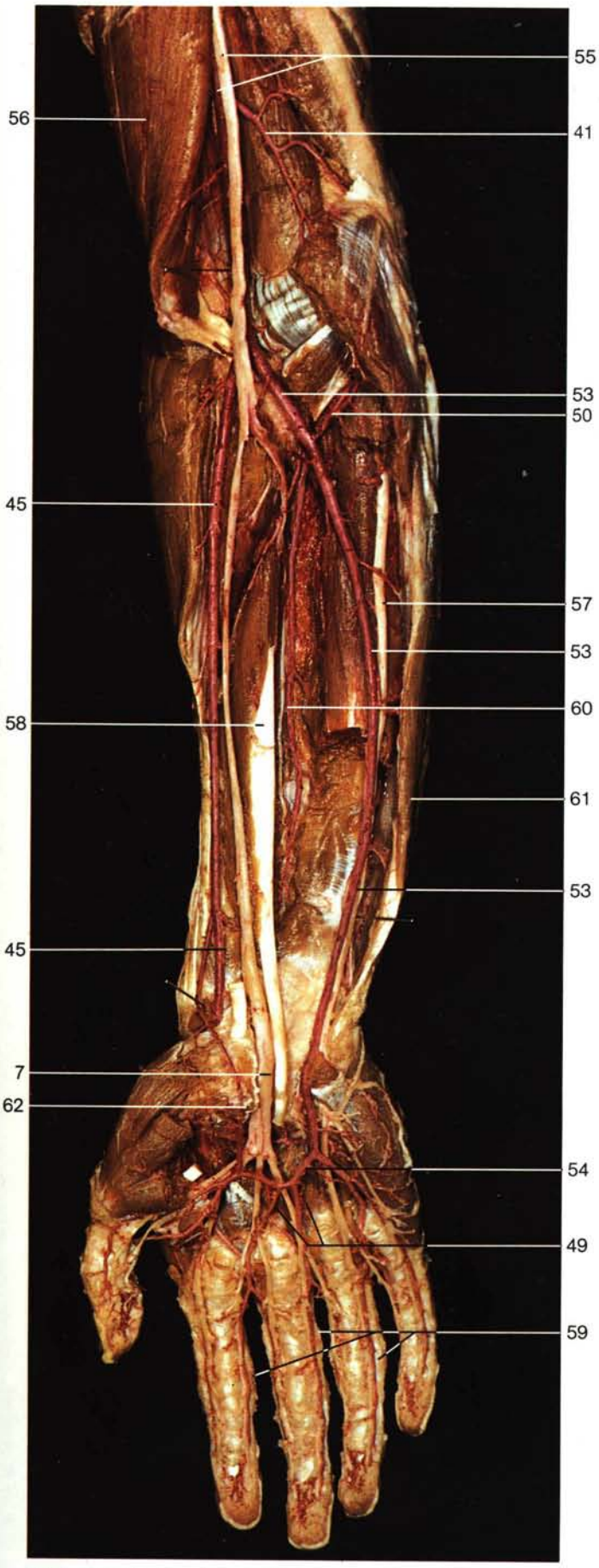

Dissection of the arteries of forearm and hand. The superficial flexor muscles have been removed, the carpal tunnel opened and the flexor retinaculum cut. The arteries have been filled with colored resin.

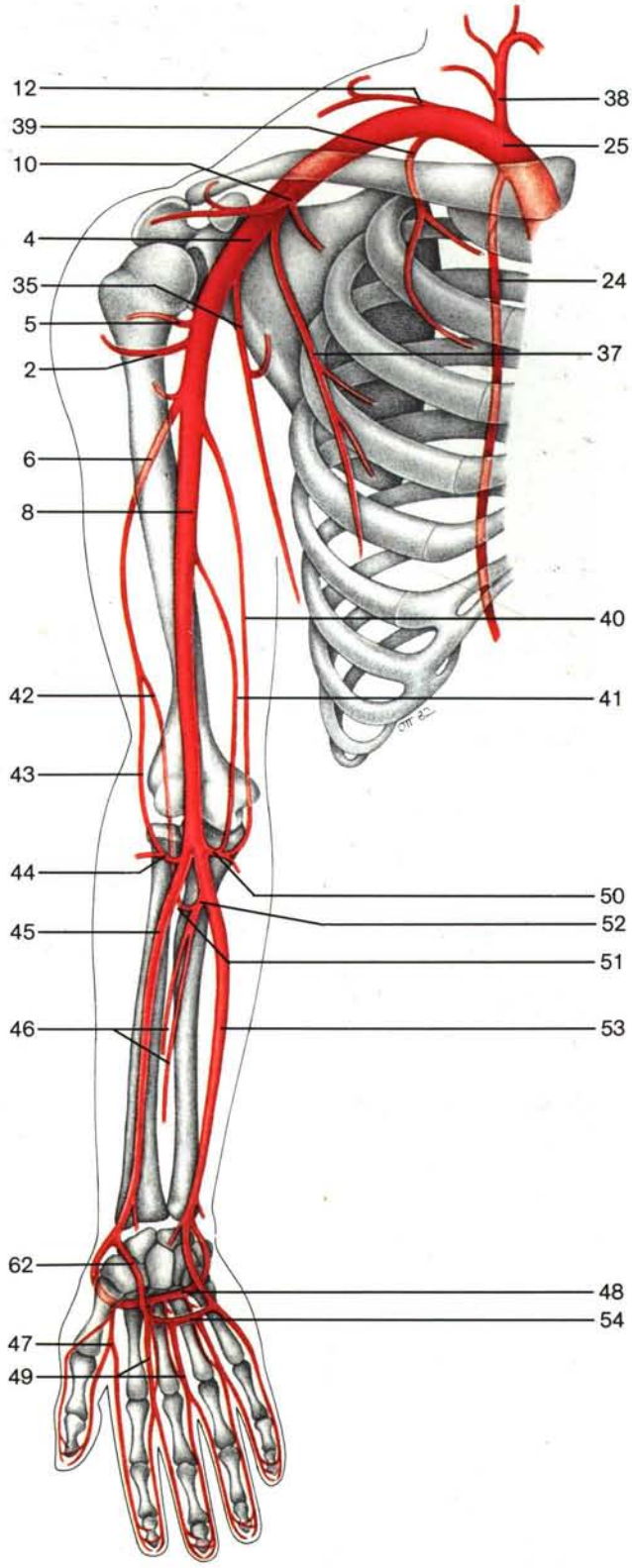

Arteries of the upper limb. (Schematic drawing.)

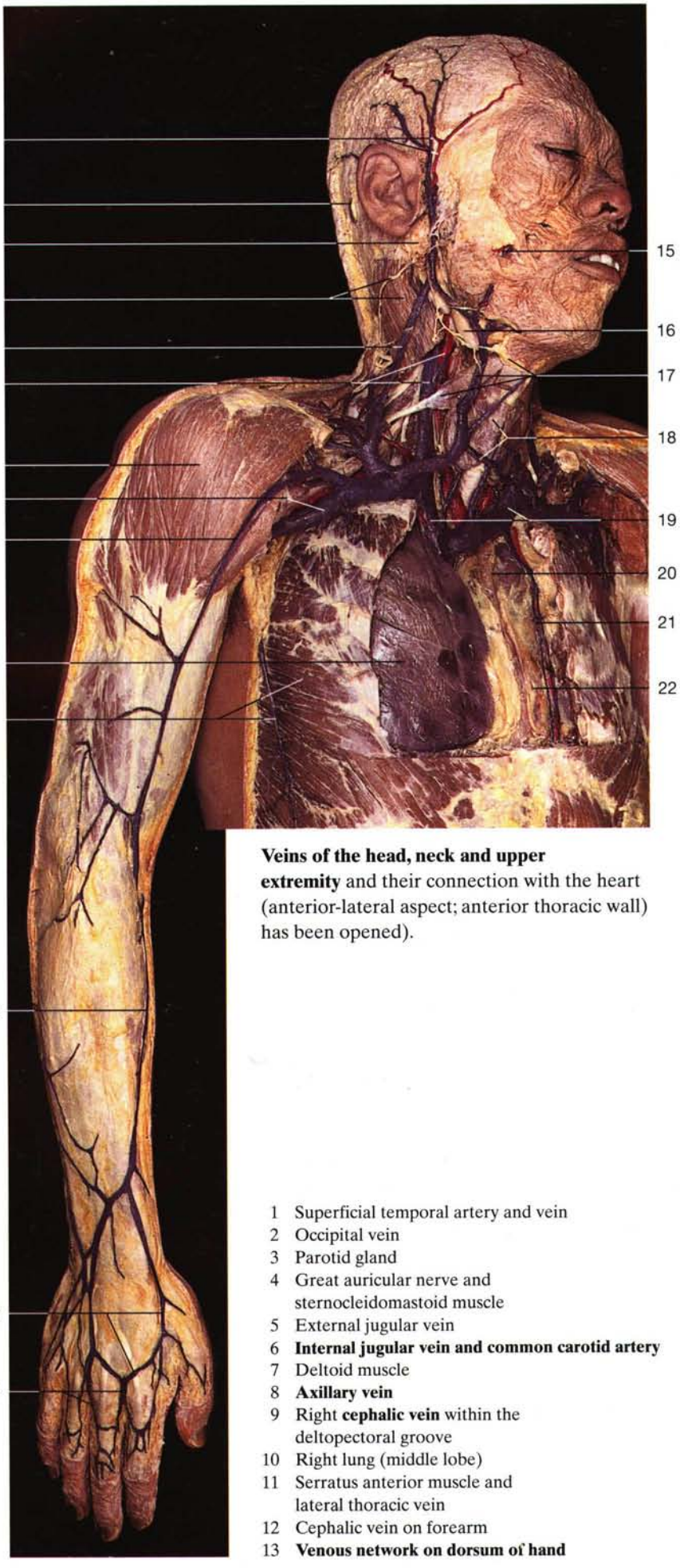

Veins of the head, neck and upper extremity and their connection with the heart (anterior-lateral aspect; anterior thoracic wall) has been opened).

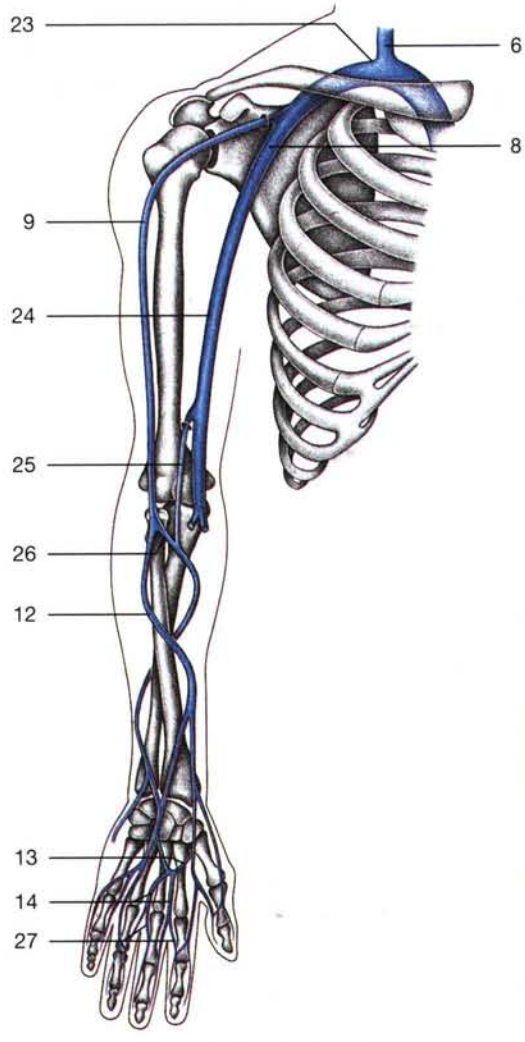

Superficial veins of upper limb. (Semischematic drawing.)

1 Superficial temporal artery and vein
2 Occipital vein
3 Parotid gland
4 Great auricular nerve and sternocleidomastoid muscle
5 External jugular vein
6 **Internal jugular vein and common carotid artery**
7 Deltoid muscle
8 **Axillary vein**
9 Right **cephalic vein** within the deltopectoral groove
10 Right lung (middle lobe)
11 Serratus anterior muscle and lateral thoracic vein
12 Cephalic vein on forearm
13 **Venous network on dorsum of hand**
14 Dorsal metacarpal veins
15 **Facial artery and vein**
16 Submandibular gland
17 Anterior jugular vein, hyoid bone and omohyoid muscle
18 Jugular venous arch and thyroid gland
19 Right and left brachiocephalic veins
20 Retrosternal body (remnant of thymus gland)
21 Internal thoracic artery and vein
22 Heart with pericardium
23 Right venous angle
24 **Brachial vein**
25 Basilic vein
26 Median cubital vein
27 Digital veins

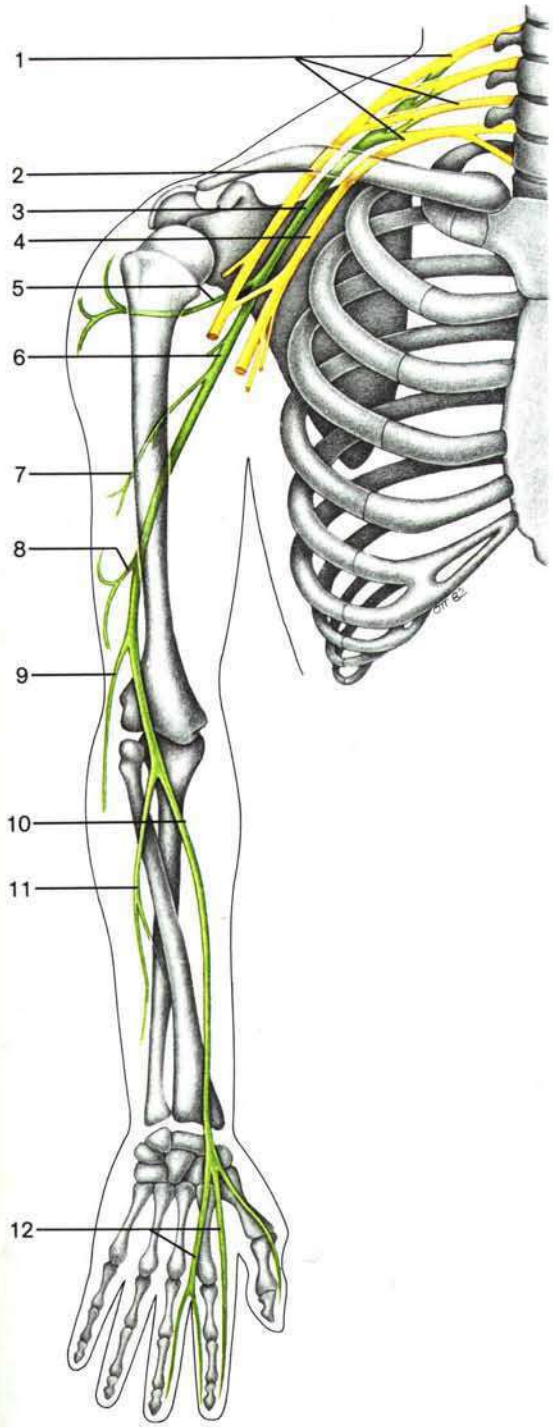

Main branches of radial nerve. (Schematic drawing.) Posterior divisions of trunks and posterior cord and its branches are indicated in green.

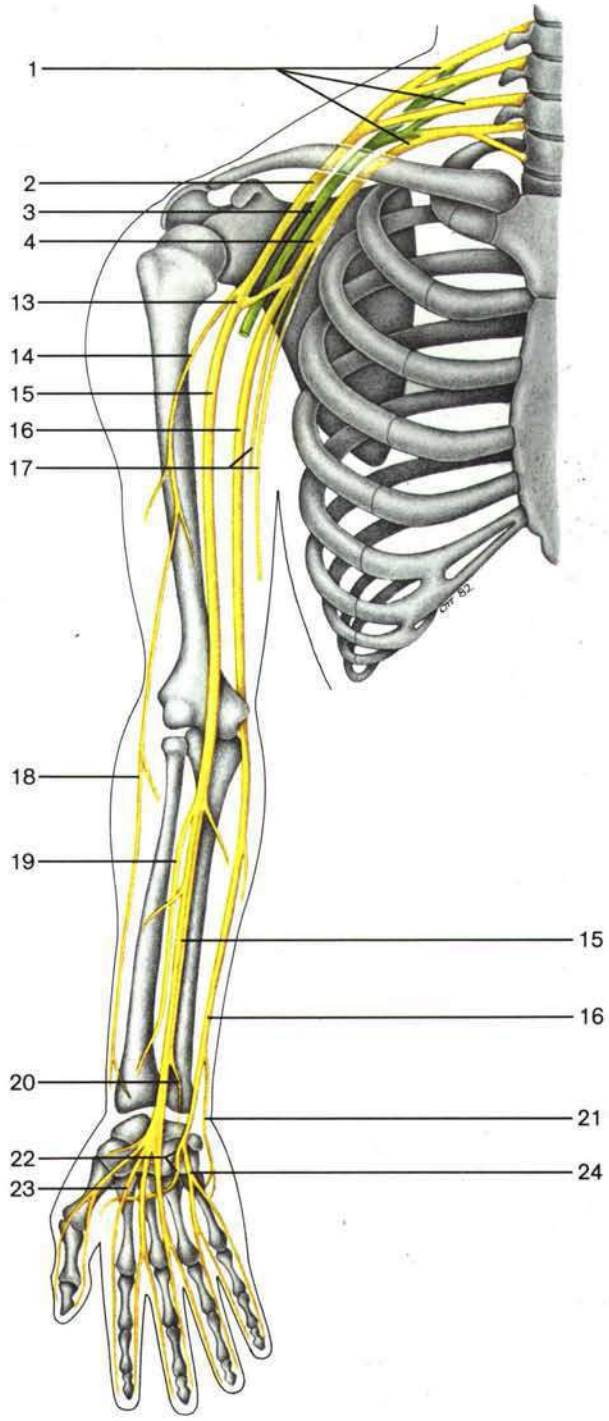

Main branches of musculocutaneous, median and ulnar nerves. (Schematic drawing.)
Anterior divisions of the trunks and all the components arising from them are indicated in yellow.

1 Brachial plexus
2 Lateral cord of brachial plexus
3 Posterior cord of brachial plexus
4 Medial cord of brachial plexus
5 **Axillary nerve**
6 **Radial nerve**
7 Posterior cutaneous nerve of arm
8 Lower lateral cutaneous nerve of arm
9 Posterior cutaneous nerve of forearm
10 Superficial branch of radial nerve
11 Deep branch of radial nerve
12 Dorsal digital nerves

13 Roots of median nerve
14 Musculocutaneous nerve
15 **Median nerve**
16 **Ulnar nerve**
17 Medial cutaneous nerves of arm and forearm
18 Lateral cutaneous nerve of forearm
19 Anterior interosseous nerve
20 Palmar branch of median nerve
21 Dorsal branch of ulnar nerve
22 Deep branch of ulnar nerve
23 Common palmar digital nerves of median nerve
24 Superficial branch of ulnar nerve

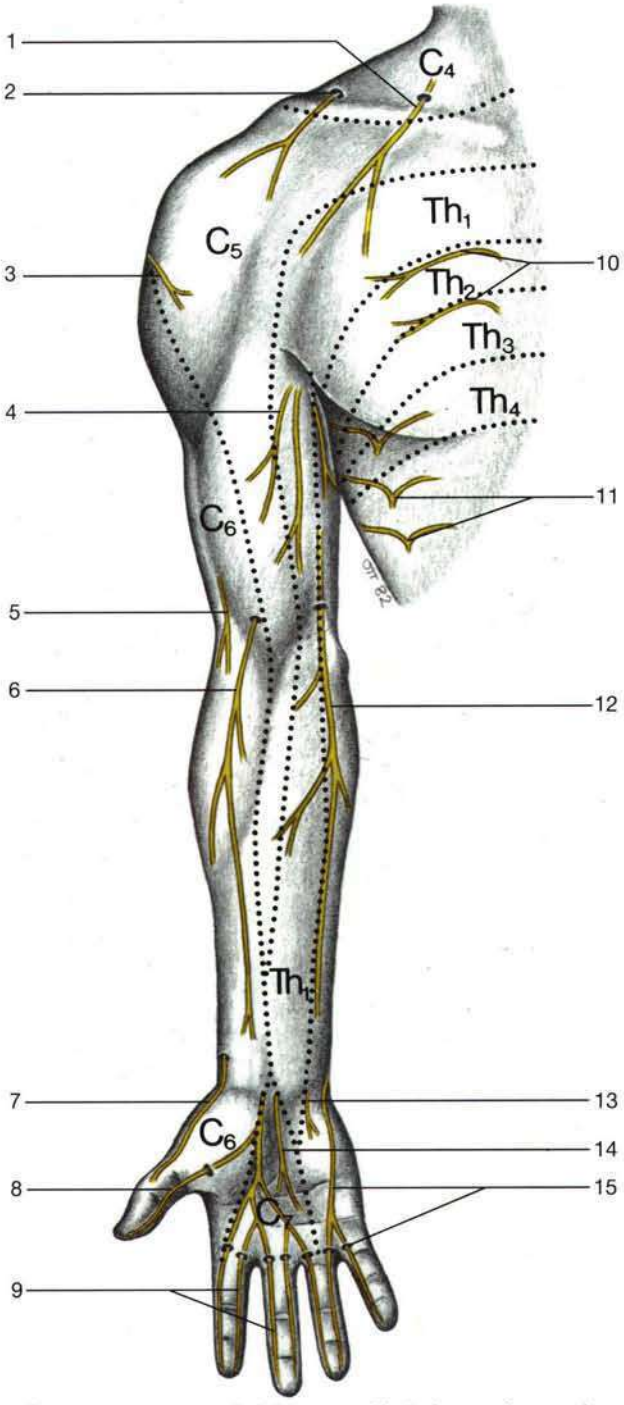

Cutaneous nerves of right upper limb (ventral aspect). (Semischematic drawing.)

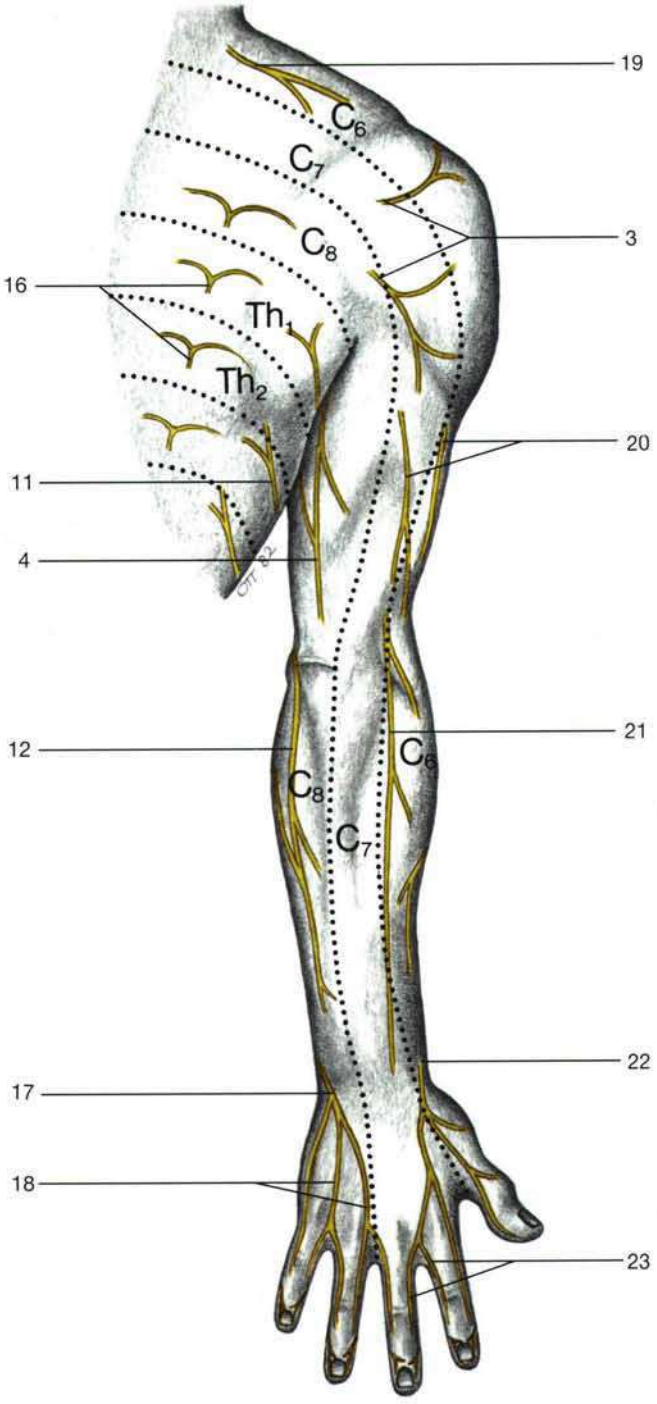

Cutaneous nerves of the upper limb (dorsal aspect). (Semischematic drawing.)

1 Medial supraclavicular nerve
2 Intermediate supraclavicular nerve
3 Upper lateral cutaneous nerve of arm
4 Terminal branches of intercostobrachial nerves
5 Lower lateral cutaneous nerve of arm
6 Lateral cutaneous nerve of forearm
7 Terminal branch of superficial branch of radial nerve
8 Palmar digital nerve of thumb (branch of median nerve)
9 Palmar digital branches of median nerve
10 Anterior cutaneous branches of intercostal nerves
11 Lateral cutaneous branches of intercostal nerves
12 Medial cutaneous nerve of forearm
13 Palmar cutaneous branch of ulnar nerve
14 Palmar branch of median nerve
15 Palmar digital branches of ulnar nerve
16 Cutaneous branches of dorsal rami of spinal nerves
17 Dorsal branch of ulnar nerve
18 Dorsal digital nerves
19 Posterior supraclavicular nerve
20 Posterior cutaneous nerve of arm
21 Posterior cutaneous nerve of forearm
22 Superficial branch
23 Dorsal digital branches

} from radial nerve

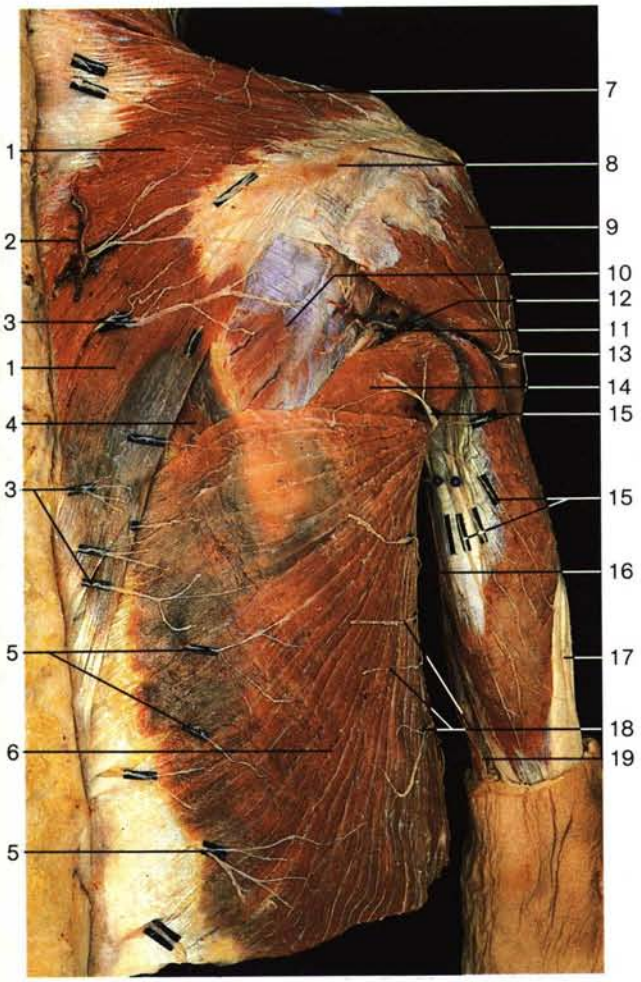

Dorsal region of shoulder, superficial layer. Note the segmental arrangement of the cutaneous nerves of the back.

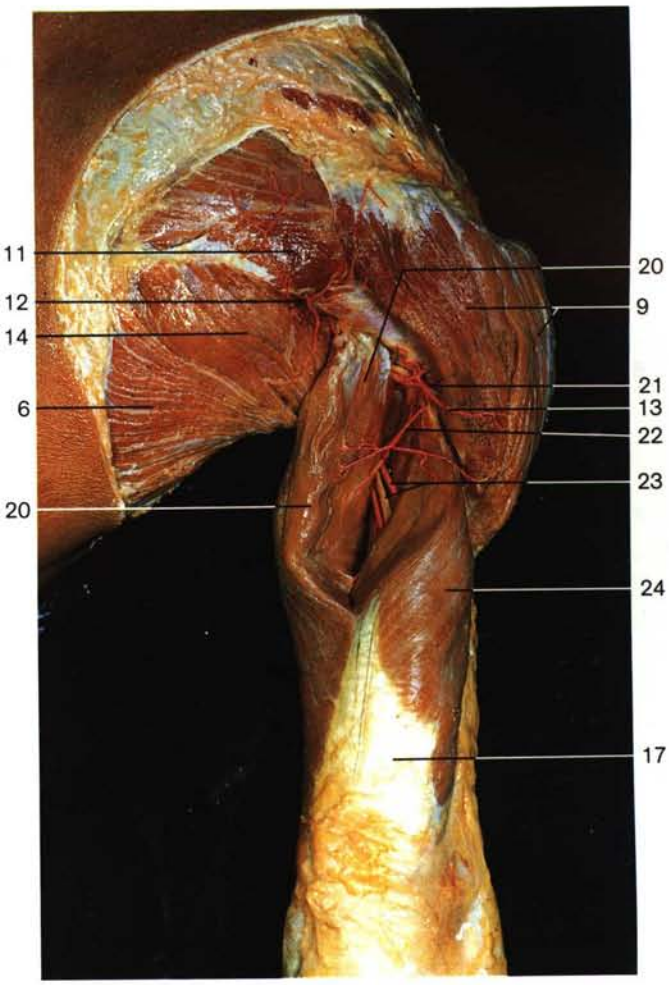

Shoulder and arm (dorsal aspect). Dissection of the quadrangular and triangular spaces of the axillary region.

1 Trapezius muscle
2 Dorsal branches of posterior intercostal artery and vein (medial cutaneous branches)
3 Medial branches of dorsal rami of spinal nerves
4 Rhomboid major muscle
5 Lateral branches of dorsal rami of spinal nerves
6 Latissimus dorsi muscle
7 Posterior supraclavicular nerves
8 Spine of scapula
9 Deltoid muscle
10 Infraspinatus muscle
11 Teres minor muscle
12 **Triangular space** with circumflex scapular artery and vein
13 Upper lateral cutaneous nerve of arm with artery
14 Teres major muscle
15 Terminal branches of intercostobrachial nerve
16 Medial cutaneous nerve of arm
17 Tendon of triceps brachii muscle
18 Lateral cutaneous branches of intercostal nerves
19 Medial cutaneous nerve of forearm
20 Long head of triceps brachii muscle
21 **Quadrangular space** with axillary nerve and posterior humeral circumflex artery
22 Anastomosis between profunda brachii artery and posterior humeral circumflex artery
23 Course of radial nerve and profunda brachii artery
24 Lateral head of triceps brachii muscle
25 Course of descending scapular artery and dorsal scapular nerve
26 Course of suprascapular nerve, artery and vein

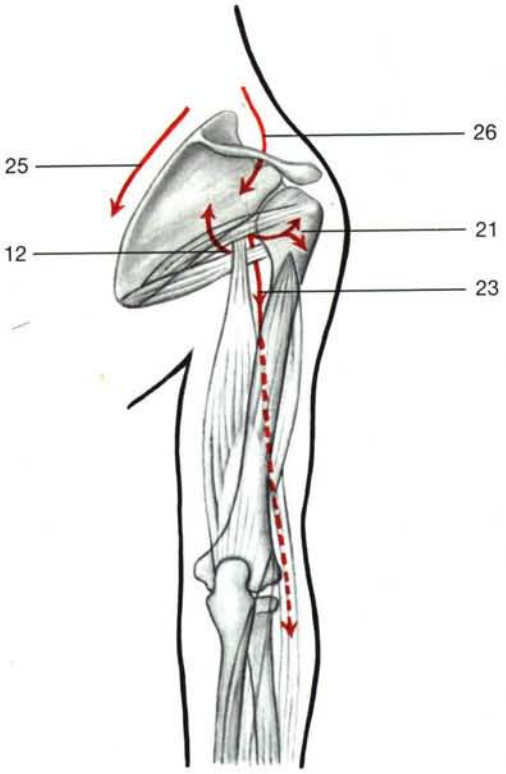

Course of vessels and nerves to shoulder and upper limb. (Schematic drawing.)

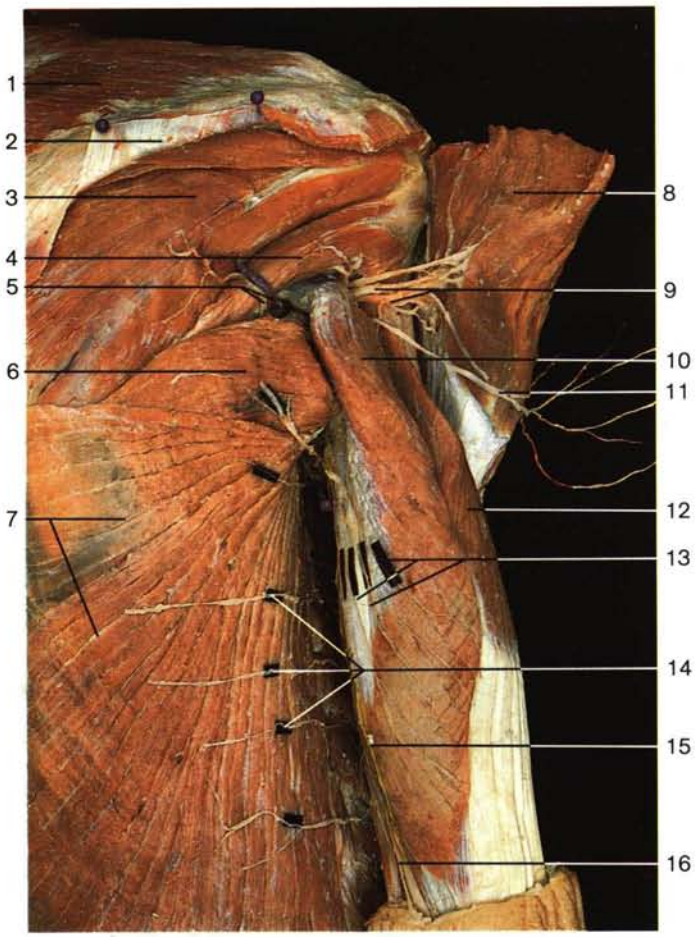

Scapular region, arm and shoulder, deep layer (dorsal aspect). Part of deltoid muscle has been cut and reflected to display the quadrangular and triangular spaces of the axillary region.

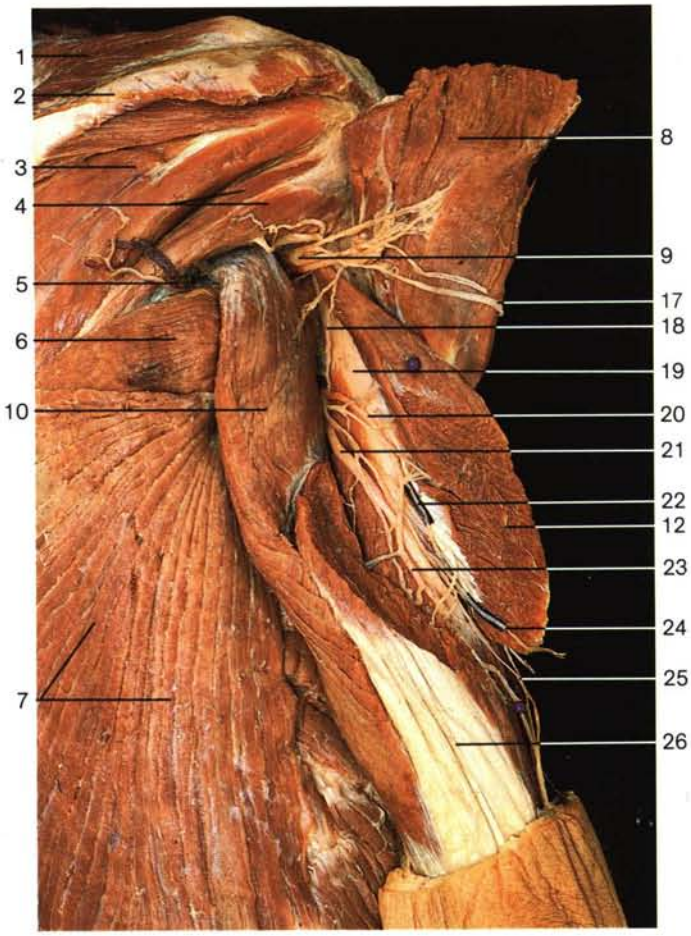

Scapular region and posterior brachial region; arm and shoulder, deep layer (dorsal aspect). The lateral head of the triceps brachii muscle has been cut to display the radial nerve and accompanying vessels.

1 Trapezius muscle
2 Spine of scapula
3 Infraspinatus muscle
4 Teres minor muscle
5 **Triangular space** containing circumflex scapular artery and vein
6 Teres major muscle
7 Latissimus dorsi muscle
8 Deltoid muscle (cut and reflected)
9 **Quadrangular space** containing axillary nerve and posterior circumflex humeral artery and vein
10 Long head of triceps brachii muscle
11 Cutaneous branch of axillary nerve
12 Lateral head of triceps brachii muscle
13 Terminal branches of intercostobrachial nerve
14 Lateral cutaneous branches of intercostal nerves
15 Medial cutaneous nerve of arm
16 Medial cutaneous nerve of forearm
17 Upper lateral cutaneous nerve of arm
18 Anastomosis between profunda brachii artery and posterior humeral circumflex artery
19 Humerus
20 **Profunda brachii artery**
21 **Radial nerve**
22 Radial collateral artery
23 Middle collateral artery
24 Lower lateral cutaneous nerve of arm
25 Posterior cutaneous nerve of forearm
26 Tendon of triceps brachii muscle

Dorsal regions of neck and shoulder (dorsal aspect). Left side: superficial layer. Right side: trapezius and latissimus dorsi muscles have been removed. Dissection of dorsal branches of spinal nerves.

1 **Greater occipital nerve**
2 Ligamentum nuchae
3 Splenius capitis muscle
4 Sternocleidomastoid muscle
5 Lesser occipital nerve
6 Splenius cervicis muscle
7 Descending and transverse fibers of trapezius muscle
8 **Medial cutaneous branches of dorsal rami of spinal nerves**
9 Ascending fibers of trapezius muscle
10 Latissimus dorsi muscle
11 Cutaneous branch of third occipital nerve
12 Great auricular nerve
13 Accessory nerve (n. XI)
14 Posterior supraclavicular nerve and levator scapulae muscle
15 Branches of suprascapular artery
16 Deltoid muscle
17 Rhomboid major muscle
18 Infraspinatus muscle
19 Teres minor muscle
20 Upper lateral cutaneous nerve of arm (branch of axillary nerve)
21 Teres major muscle
22 Medial margin of scapula
23 Long head of triceps muscle
24 Posterior cutaneous nerve of arm (branch of radial nerve)
25 Latissimus dorsi muscle (divided)
26 Ulnar nerve and brachial artery
27 **Lateral cutaneous branches of dorsal rami of spinal nerves** and iliocostalis thoracis muscle
28 External intercostal muscle and seventh rib
29 Serratus posterior inferior muscle

1 Clavicle
2 Deltoid muscle
3 **Suprascapular artery**
4 Suprascapular nerve
5 Superior transverse scapular ligament
6 Teres minor muscle
7 Axillary nerve and posterior circumflex humeral artery
8 Long head of triceps muscle
9 **Circumflex scapular artery**
10 Teres major muscle
11 Greater occipital nerve
12 Lesser occipital nerve
13 Great auricular nerve
14 Splenius capitis muscle
15 Accessory nerve (n. XI)
16 Third occipital nerve and levator scapulae muscle
17 Serratus posterior superior muscle
18 Spine of scapula
19 Descending scapular artery and dorsal scapular nerve
20 Rhomboid major muscle
21 Infraspinatus muscle and medial margin of scapula
22 Radial nerve and profunda brachii artery
23 Thoracodorsal artery
24 Thyrocervical trunk
25 Roots of brachial plexus

Dorsal region of shoulder; deepest layer. Rhomboid and scapular muscles fenestrated; posterior part of deltoid muscle reflected.

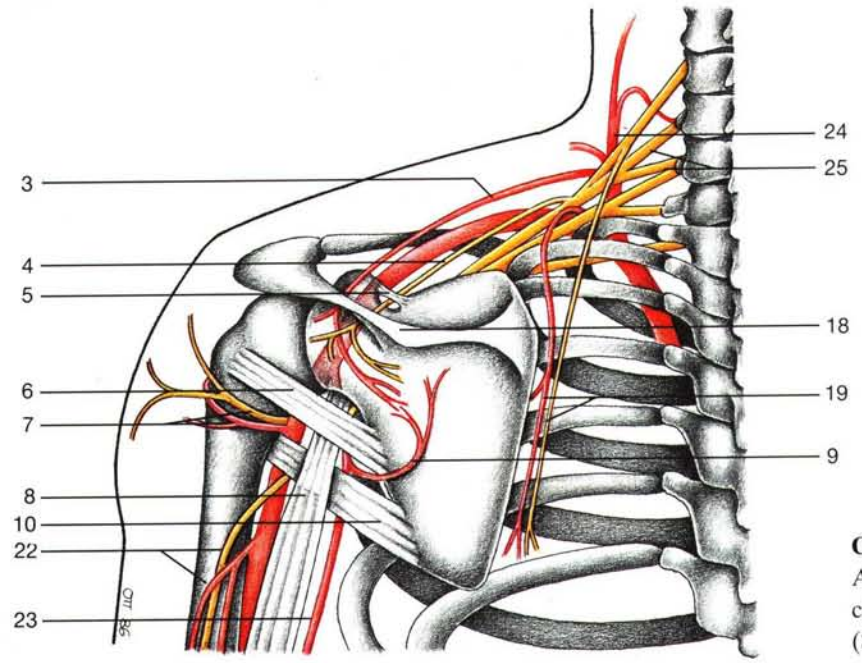

Collateral circulation of shoulder. Anastomosis of suprascapular and circumflex scapular arteries. (Semischematic drawing.)

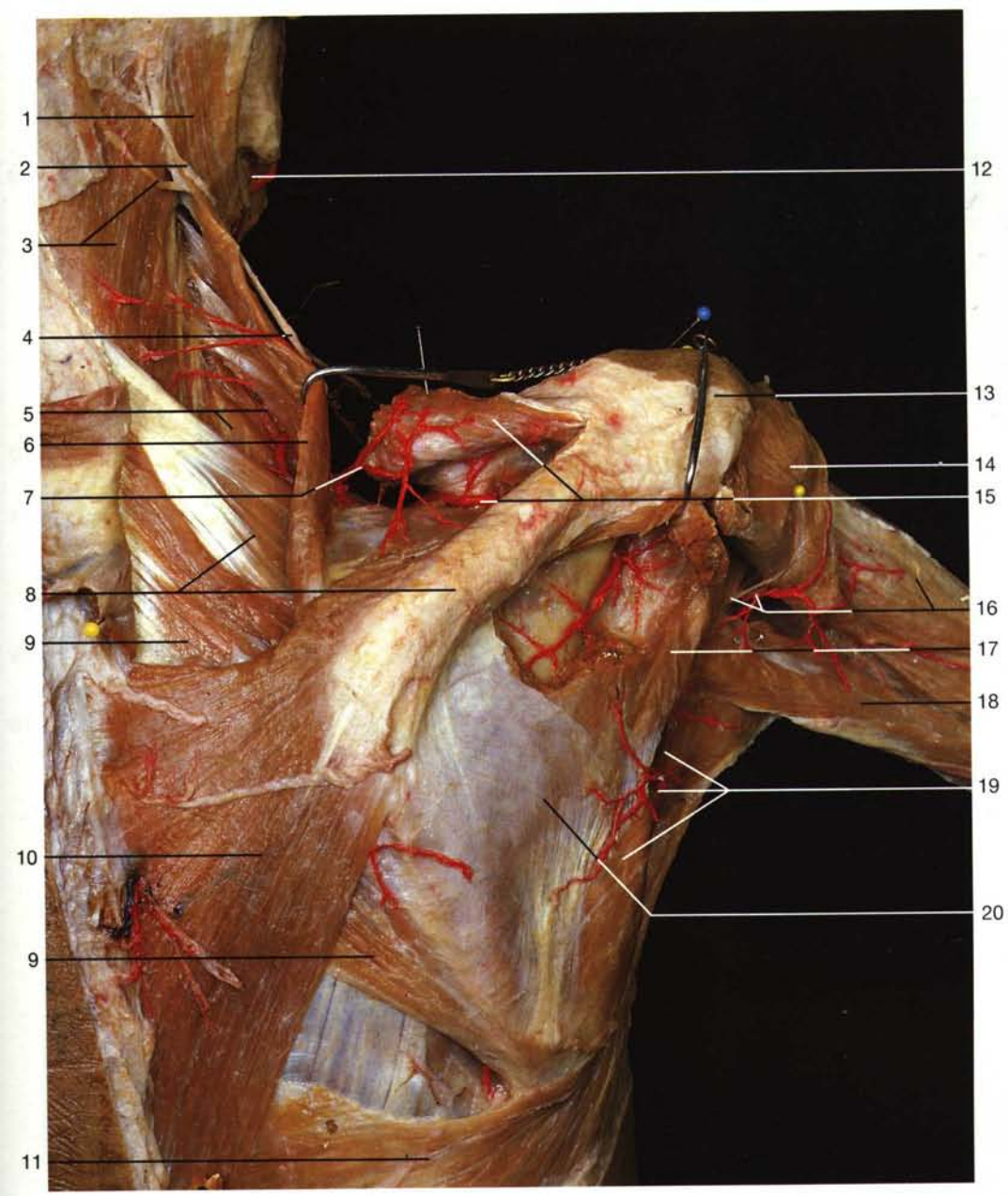

Scapular region, arm and shoulder (dorsal aspect). Arteries of scapular region are injected. Trapezius, deltoid and infraspinatus muscles are partially removed or reflected.

1 Sternocleidomastoid muscle
2 Lesser occipital nerve
3 Splenius capitis muscle and third occipital nerve
4 **Accessory nerve** (N. XI)
5 Splenius cervicis muscle and transverse cervical artery (deep branch)
6 Levator of scapula muscle
7 **Transverse cervical artery** (superficial branch)
8 Spine of scapula and serratus posterior superior muscle
9 Rhomboid major muscle
10 Trapezius muscle
11 Latissimus dorsi muscle
12 Facial artery
13 Acromion
14 Deltoid muscle
15 **Suprascapular artery** and supraspinatus muscle (reflected)
16 **Axillary nerve,** posterior circumflex humeral artery and lateral head of triceps, brachii muscle
17 Teres minor muscle
18 Long head of triceps brachii muscle
19 **Circumflex scapular artery** and teres major
20 Infraspinatus muscle

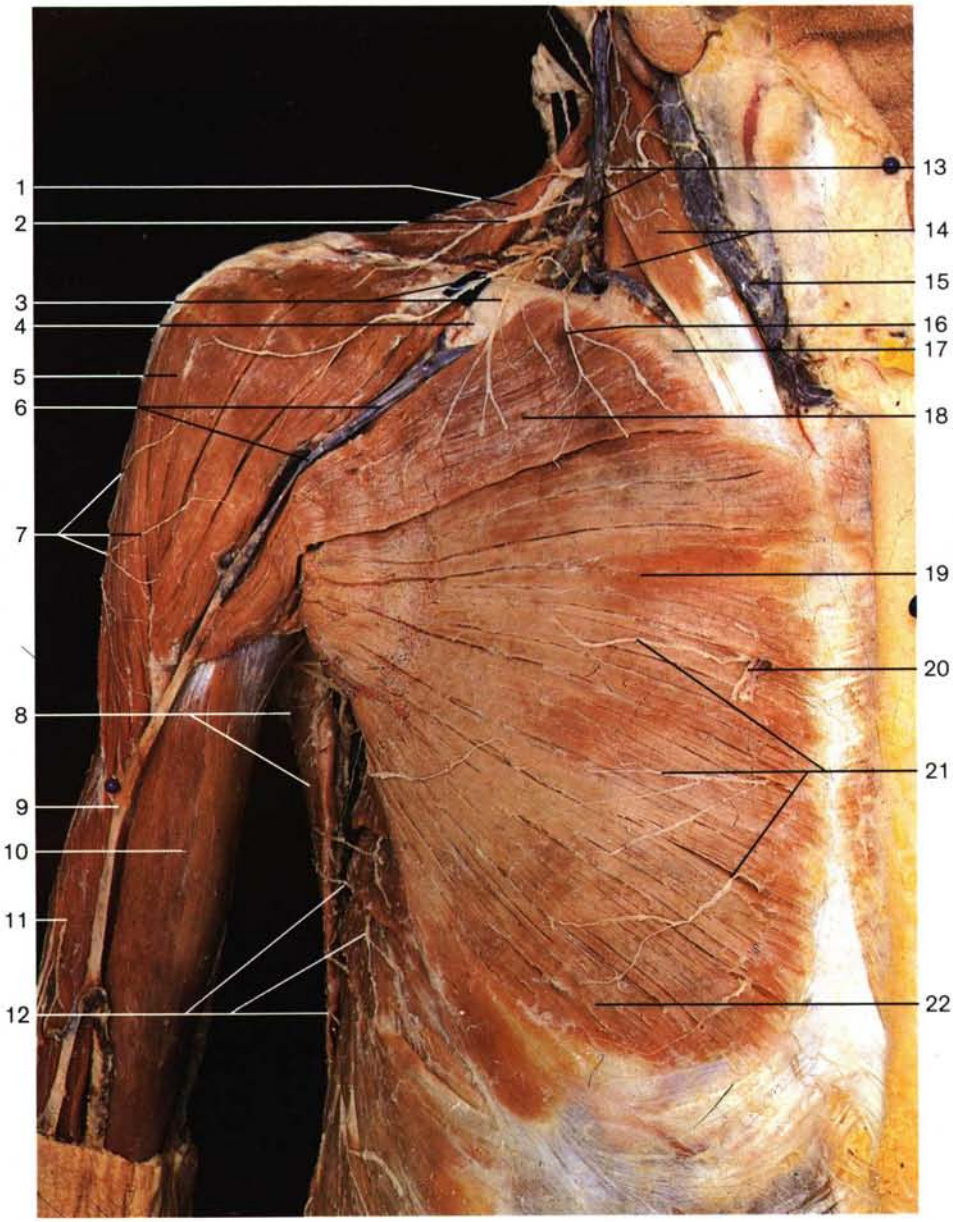

Right shoulder and thoracic wall, superficial layer (anterior aspect). Dissection of the cutaneous nerves and veins.

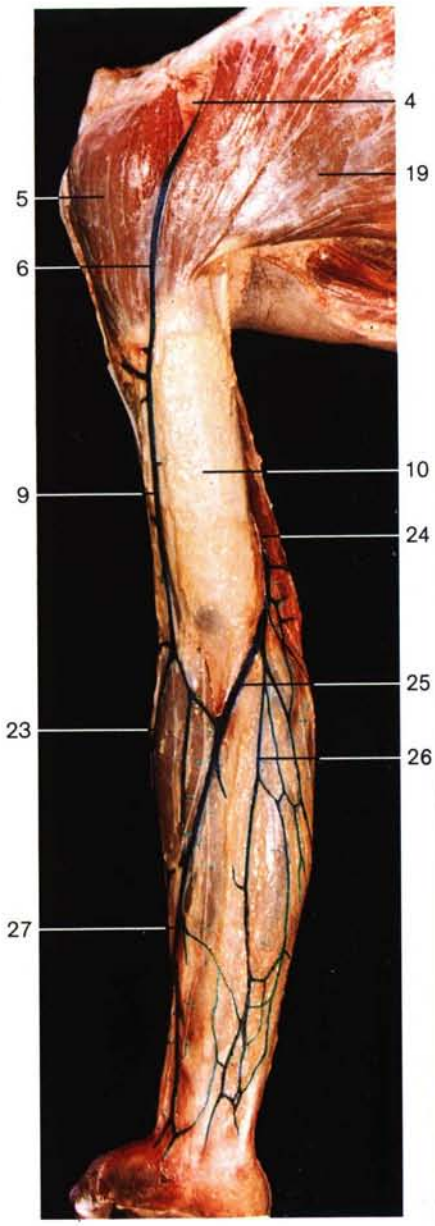

Superficial veins of right arm have been injected with blue gelatin.

1 Trapezius muscle
2 Posterior supraclavicular nerve
3 Middle supraclavicular nerve
4 Deltopectoral triangle
5 Deltoid muscle
6 Cephalic vein within the deltopectoral groove
7 Upper lateral cutaneous nerve of arm (branch of axillary nerve)
8 Latissimus dorsi muscle
9 **Cephalic vein**
10 Biceps brachii muscle
11 Triceps brachii muscle
12 Lateral cutaneous branches of intercostal nerves
13 Transverse cervical nerve and external jugular vein
14 Sternocleidomastoid muscle
15 **Anterior jugular vein**
16 Anterior supraclavicular nerve
17 Clavicle
18 Clavicular part of pectoralis major muscle
19 Sternocostal part of pectoralis major muscle
20 Perforating branch of internal thoracic artery
21 **Anterior cutaneous branches of intercostal nerves**
22 Abdominal part of pectoralis major muscle
23 Accessory cephalic vein
24 **Basilic vein**
25 **Median cubital vein**
26 Median vein of forearm
27 Cephalic vein in the forearm

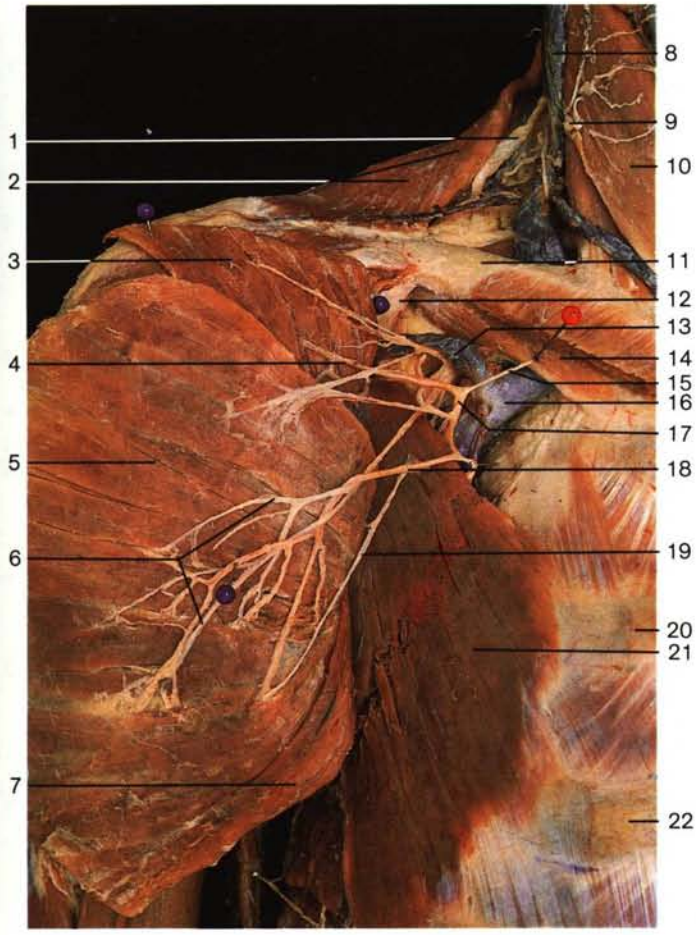

Right deltopectoral triangle, infraclavicular region (anterior aspect). The pectoralis major muscle has been cut and reflected.

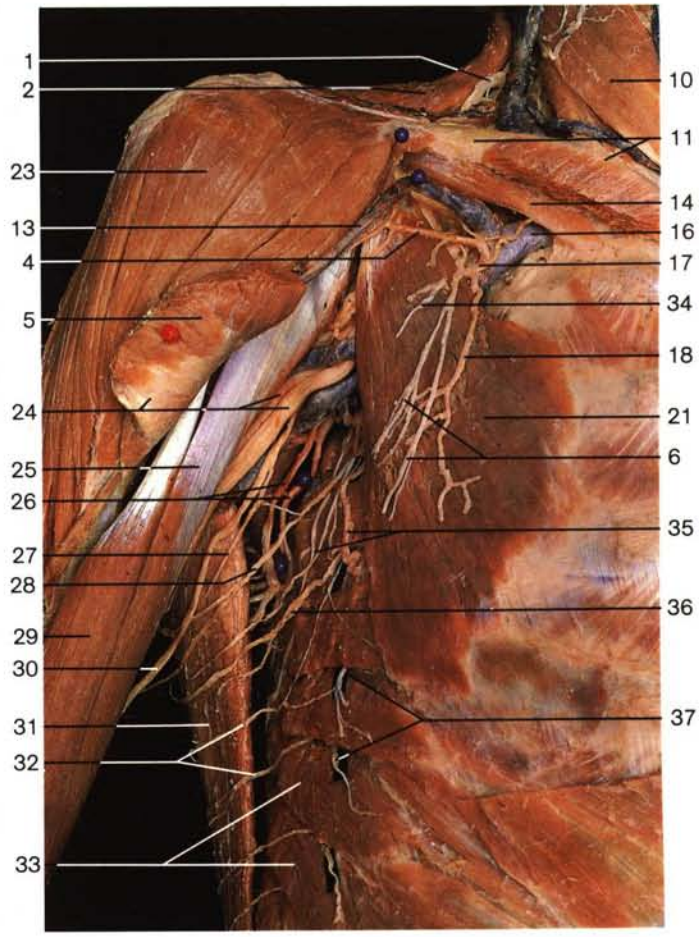

Thoracic wall and shoulder, deep layer. **Right axillary region** (anterior aspect). The pectoralis major muscle has been cut and partly removed.

1 Accessory nerve
2 Trapezius muscle
3 Pectoralis major muscle (clavicular part)
4 Acromial branch of thoracoacromial artery
5 Pectoralis major muscle
6 **Lateral pectoral nerves**
7 Abdominal part of pectoralis major muscle
8 External jugular vein
9 Cutaneous branches of cervical plexus
10 Sternocleidomastoid muscle
11 Clavicle
12 Clavipectoral fascia
13 **Cephalic vein**
14 Subclavius muscle
15 Clavicular branch of thoracoacromial artery
16 **Subclavian vein**
17 **Thoracoacromial artery**
18 Pectoral branch of thoracoacromial artery
19 Medial pectoral nerve
20 2nd rib
21 Pectoralis minor muscle
22 3rd rib
23 Deltoid muscle
24 Pectoralis major muscle (reflected), **brachial artery and median nerve**
25 Short head of biceps brachii muscle
26 Thoracodorsal artery and nerve
27 Medial cutaneous nerve of arm
28 Intercostobrachial nerve (T_2)
29 Long head of biceps brachii muscle
30 Medial cutaneous nerve of forearm
31 Latissimus dorsi muscle
32 Lateral cutaneous branches of intercostal nerves (posterior branches)
33 Serratus anterior muscle
34 Medial pectoral nerve
35 **Long thoracic nerve and lateral thoracic artery**
36 Intercostobrachial nerve (T_3)
37 **Lateral cutaneous branches of intercostal nerves** (anterior branches)

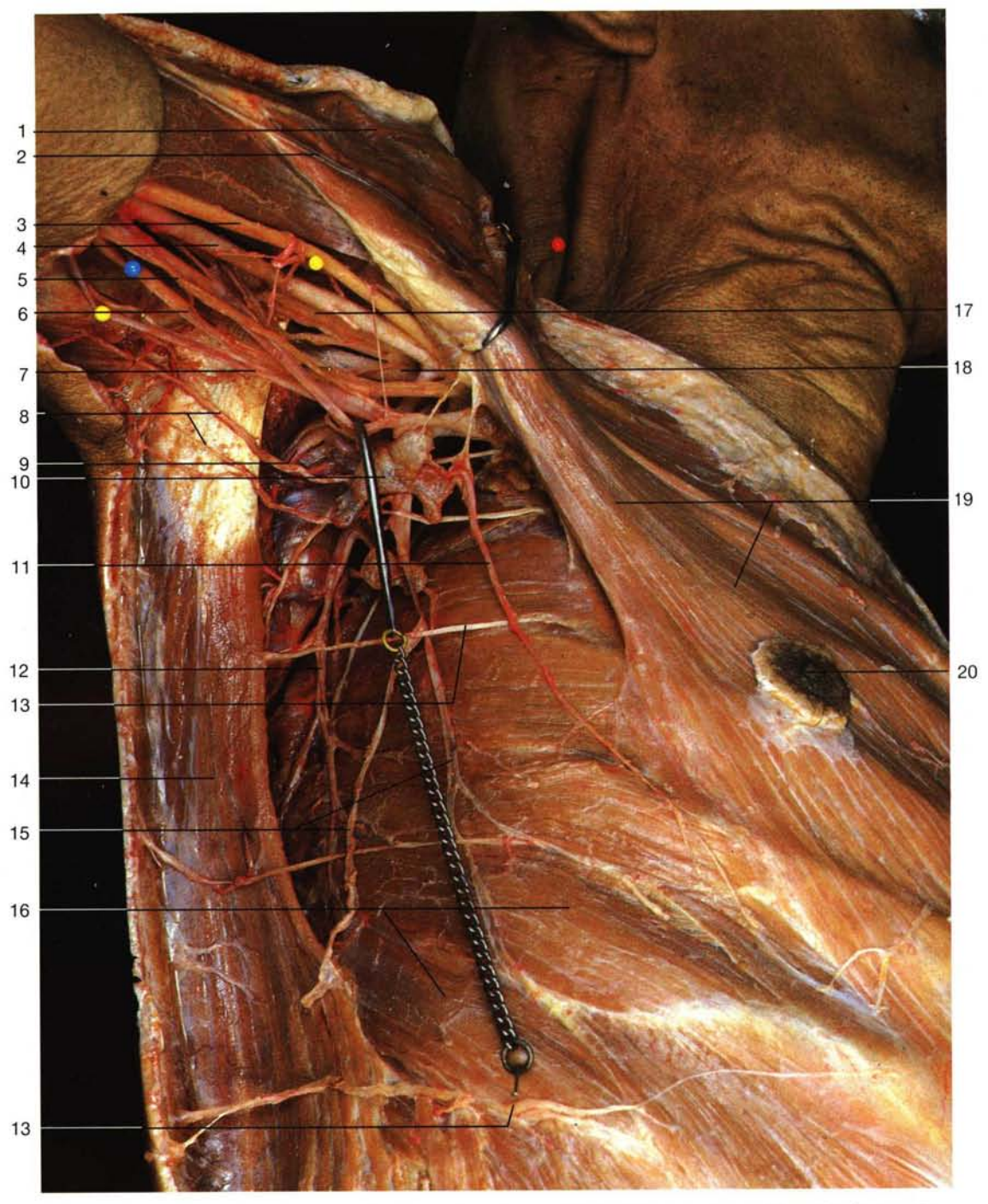

Right axillary region (inferior aspect). **Dissection of superficial axillary nodes and lymphatic vessels.** The pectoralis major muscle has been slightly elevated.

1 Deltoid muscle
2 Cephalic vein
3 Median nerve
4 Brachial artery
5 Medial cutaneous nerves of arm and forearm
6 Ulnar nerve
7 Basilic Vein
8 Intercostobrachial nerves
9 Circumflex scapular artery
10 **Superficial axillary nodes**
11 Lateral thoracic artery
12 Thoracodorsal artery
13 Lateral cutaneous branch of intercostal nerve
14 Latissimus dorsi muscle
15 Thoracoepigastric vein
16 Serratus anterior muscle
17 Musculocutaneous nerve
18 Radial nerve
19 Pectoralis major muscle
20 Nipple

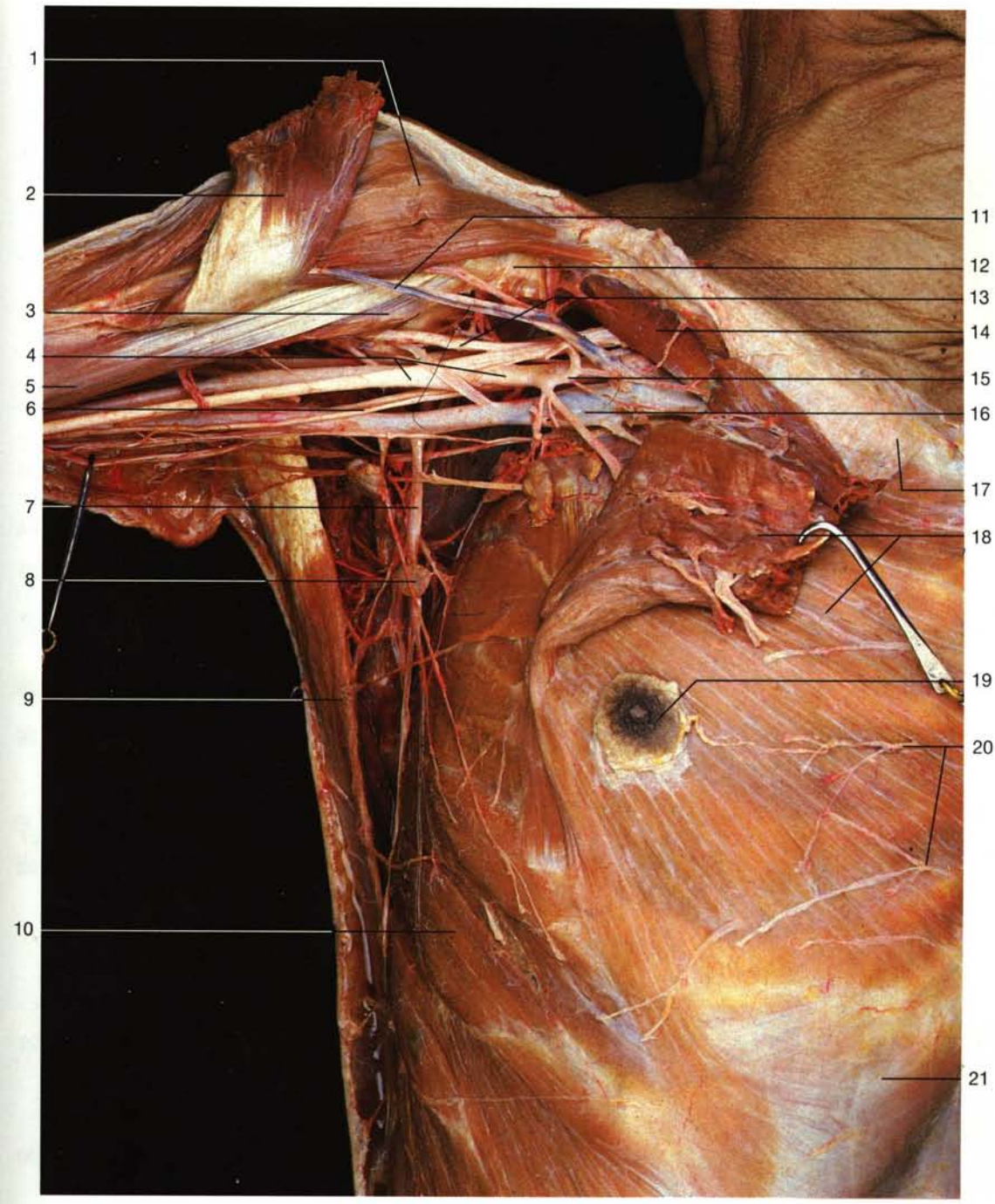

Axillary region (anterior aspect). Dissection of deep axillary nodes. Pectoralis major and minor divided and reflected. Shoulder girdle and arm elevated and reflected.

1 Deltoid muscle
2 Insertion of pectoralis major muscle
3 Coracobrachialis muscle
4 Roots of **median nerve, axillary artery**
5 Short head of biceps brachii muscle
6 **Ulnar nerve** and medial cutaneous nerve of forearm
7 Thoracoepigastric vein
8 Deep axillary node
9 Latissimus dorsi muscle
10 Serratus anterior muscle
11 Cephalic vein
12 Insertion of pectoralis minor muscle (coracoid process)
13 **Musculocutaneous nerve**
14 Subclavius muscle
15 Thoracoacromial artery
16 **Axillary vein**
17 Clavicle
18 Pectoralis major and minor muscles (reflected)
19 Nipple
20 Anterior cutaneous branches of intercostal nerves
21 Anterior layer of rectus sheaths

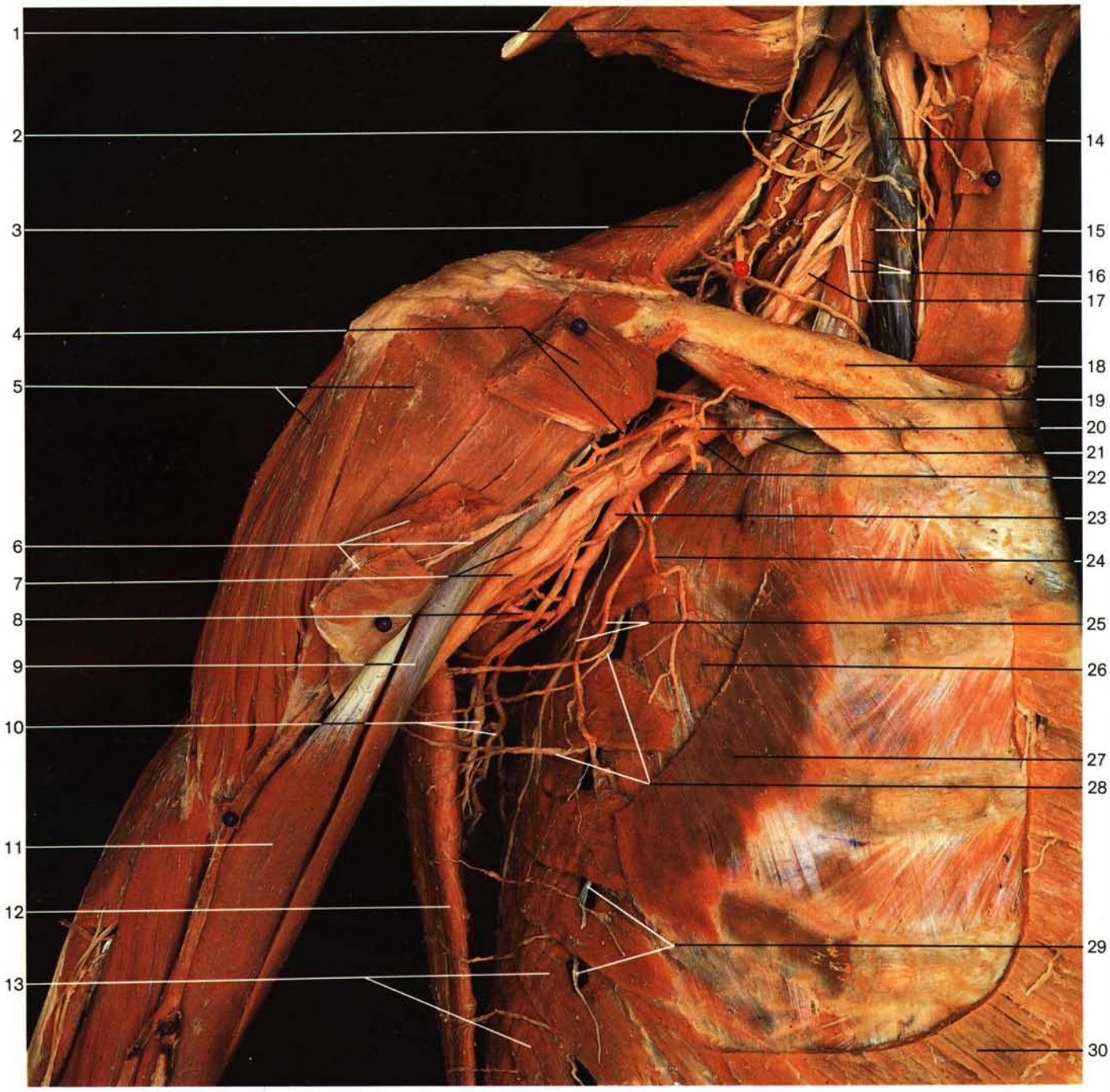

Right axillary region (anterior aspect). The pectoralis major and minor muscles have been cut and reflected to display the vessels and nerves of the axilla.

1 Sternocleidomastoid muscle (cut and reflected)
2 Cervical plexus
3 Trapezius muscle
4 Pectoralis minor muscle and medial pectoral nerve
5 Deltoid muscle
6 Pectoralis major muscle and lateral pectoral nerve
7 **Median nerve and brachial artery**
8 Circumflex scapular artery
9 Short head of biceps brachii muscle
10 **Thoracodorsal artery and nerve**
11 Long head of biceps brachii muscle
12 Latissimus dorsi muscle
13 Serratus anterior muscle
14 Internal jugular vein
15 **Scalenus anterior muscle**
16 Phrenic nerve and ascending cervical artery
17 Brachial plexus (at the levels of the trunks)
18 Clavicle
19 Subclavius muscle
20 **Thoracoacromial artery**
21 Subclavian vein (cut)
22 **Axillary artery**
23 Subscapular artery
24 Superior thoracic artery
25 **Lateral thoracic artery and long thoracic nerve**
26 External intercostal muscle
27 Insertion of pectoralis minor muscle
28 Intercostobrachial nerves
29 Lateral cutaneous branches of intercostal nerves
30 Insertion of pectoralis major muscle

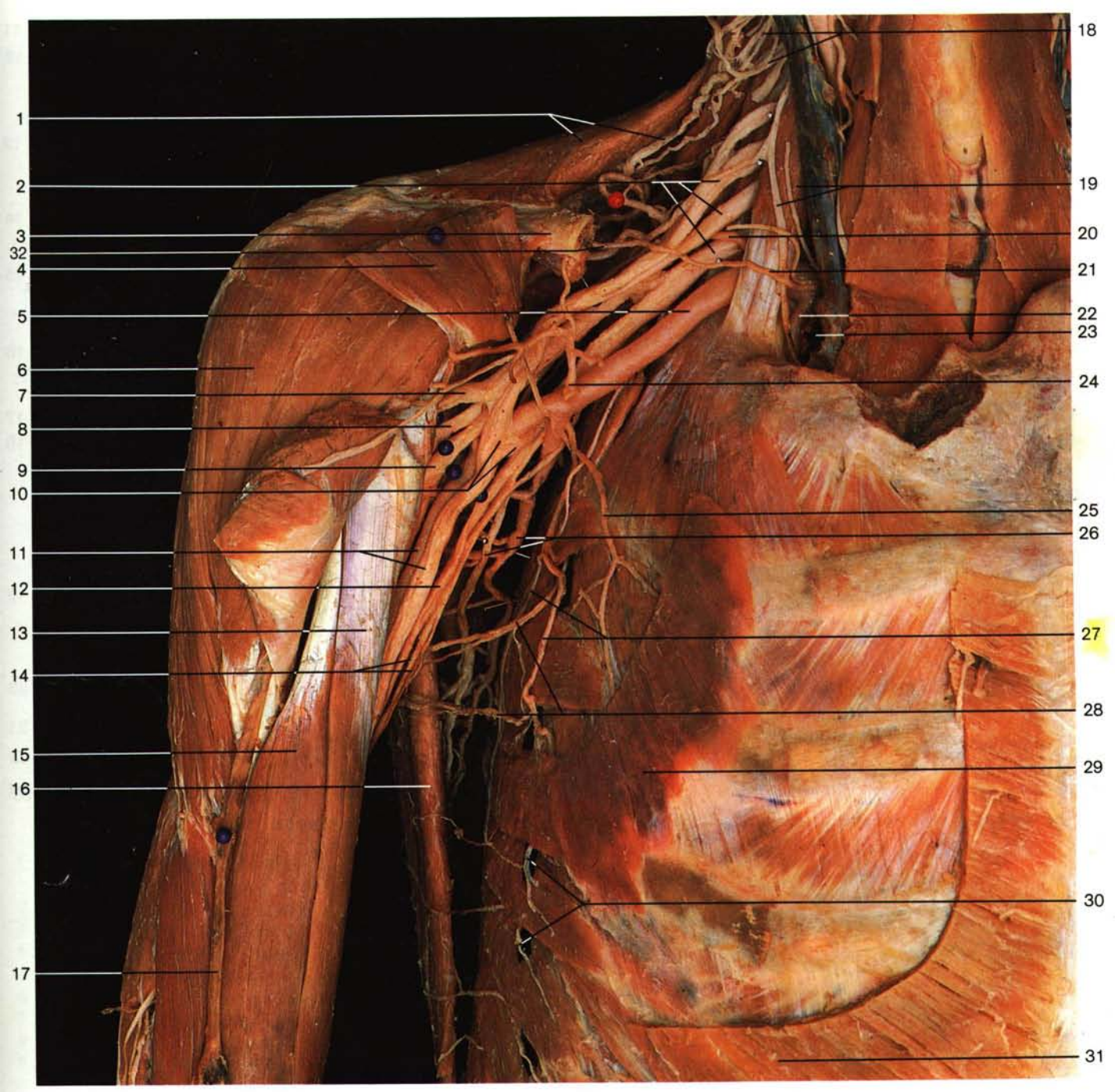

Shoulder and arm. Axillary region with brachial plexus (anterior aspect). The pectoralis major and minor muscles have been partly removed. The clavicle has been cut.

1 Trapezius muscle and accessory nerve
2 Brachial plexus
3 Clavicle (cut)
4 Pectoralis minor muscle
5 **Axillary (subclavian) artery**
6 Deltoid muscle
7 **Musculocutaneous nerve**
8 **Axillary nerve**
9 **Radial nerve**
10 Medial and lateral roots of the median nerve
11 **Median nerve** and **brachial artery**
12 **Ulnar nerve**
13 Short head of biceps brachii muscle
14 Medial cutaneous nerves of arm and forearm
15 Long head of biceps brachii muscle
16 Latissimus dorsi muscle
17 Cephalic vein
18 Cervical plexus
19 **Phrenic nerve and scalenus anterior muscle**
20 Descending scapular artery
21 Suprascapular artery
22 Internal thoracic artery
23 Subclavian vein
24 Thoracoacromial artery
25 Superior thoracic artery
26 **Thoracodorsal artery and nerve**
27 **Lateral thoracic artery and long thoracic nerve**
28 Intercostobrachial nerves
29 Insertion of the pectoralis minor muscle into the 2nd–5th ribs
30 Lateral cutaneous branches of intercostal nerves
31 Abdominal part of pectoralis major muscle (cut)
32 Suprascapular nerve and artery

Brachial plexus (anterior aspect). Clavicle and the two pectoralis muscles have been partly removed.

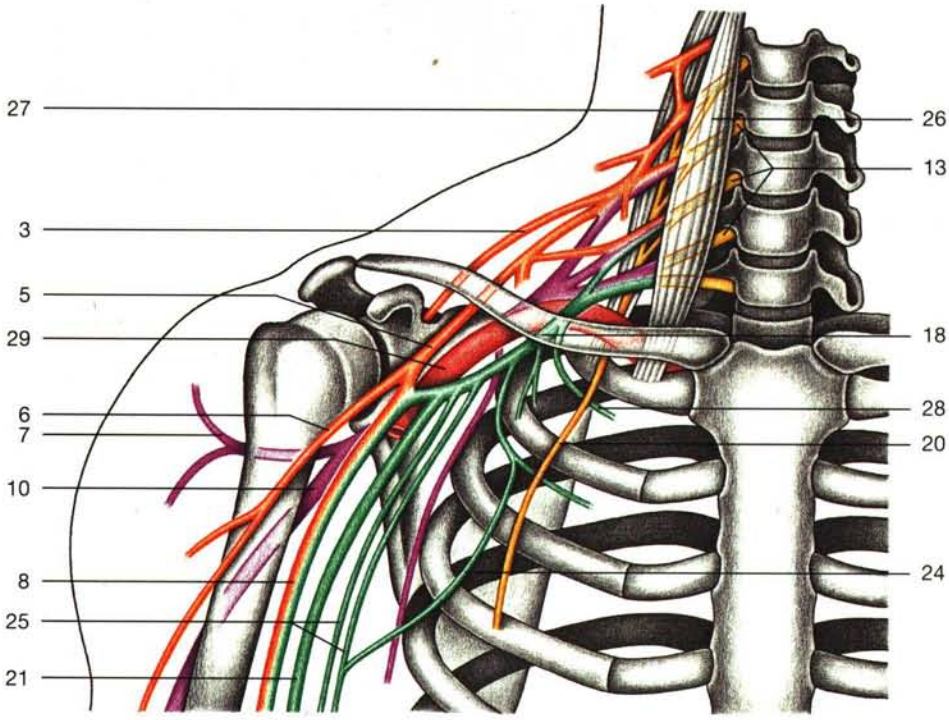

Main branches of brachial plexus. Posterior cord in red, lateral cord in yellow and medial cord in purple. (Schematic drawing.)

1 Accessory nerve
2 Dorsal scapular artery
3 Suprascapular nerve
4 Clavicle and pectoralis minor muscle
5 **Lateral cord** of brachial plexus
6 Musculocutaneous nerve
7 Axillary nerve
8 Median nerve
9 Brachial artery
10 Radial nerve
11 Cervical plexus
12 Common carotid artery
13 Roots of brachial plexus (C_5–T_1)
14 Phrenic nerve
15 Dorsal scapular artery
16 Subclavian artery
17 **Posterior cord** of brachial plexus
18 **Medial cord** of brachial plexus
19 Subscapular artery
20 Long thoracic nerve
21 Ulnar nerve
22 Medial cutaneous nerve of forearm
23 Thoracodorsal nerve
24 Intercostobrachial nerve
25 Medial cutaneous nerves of arm and forearm
26 Scalenus anterior muscle
27 Scalenus medius muscle
28 Intercostal nerve (T_1)
29 Axillary artery

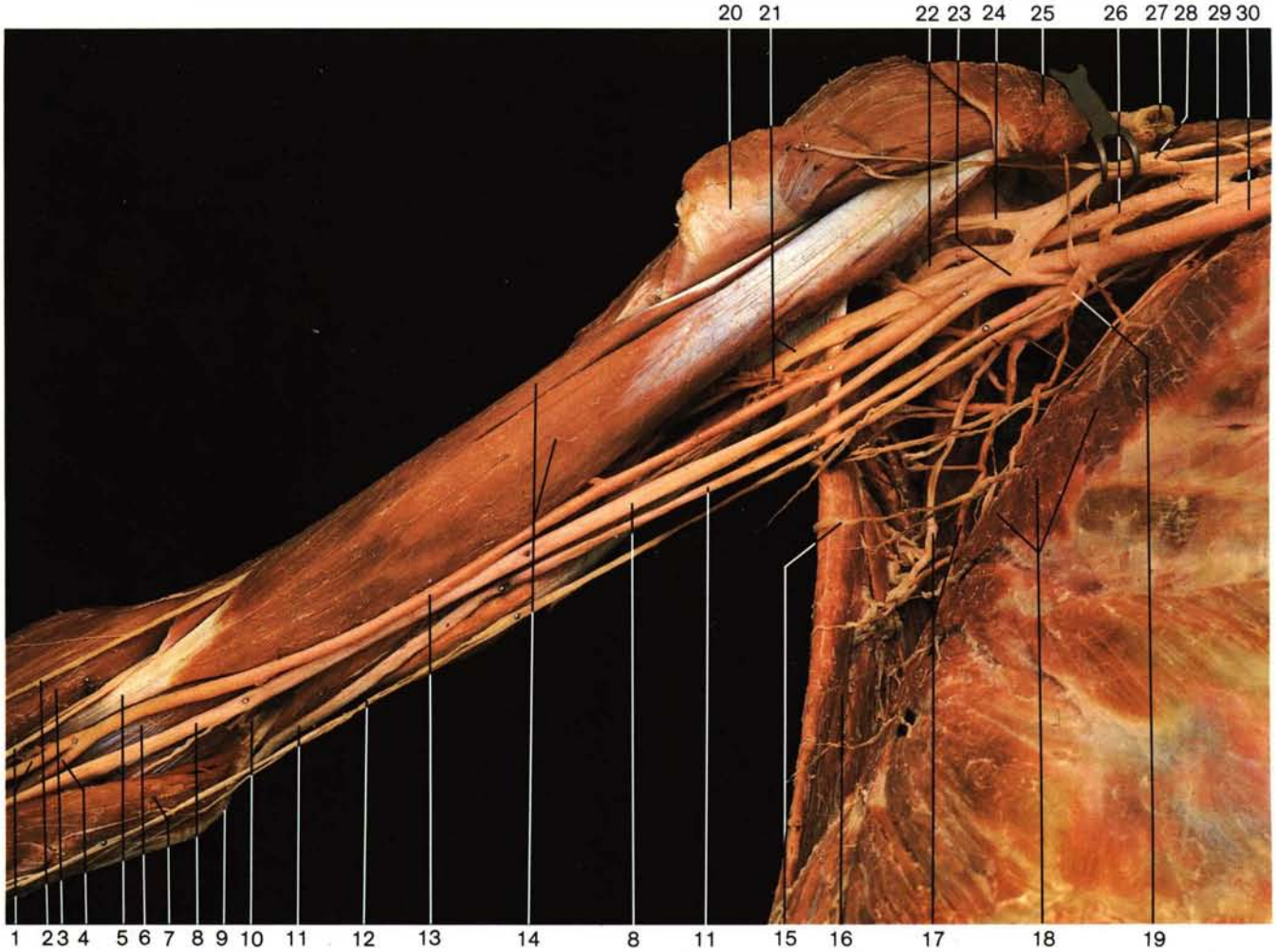

Right arm. Dissection of vessels and nerves (medial aspect). Shoulder girdle has been reflected slightly.

1 Radial artery and superficial branch of radial nerve
2 Lateral cutaneous nerve of forearm
3 Brachioradialis muscle
4 Ulnar artery
5 Tendon of biceps brachii muscle
6 Brachialis muscle
7 Pronator teres muscle
8 **Median nerve**
9 Medial epicondyle of humerus
10 Inferior ulnar collateral artery
11 **Ulnar nerve**
12 Medial cutaneous nerve of forearm
13 **Brachial artery**
14 Biceps brachii muscle
15 Intercostobrachial nerve (T_3)
16 Latissimus dorsi muscle
17 Thoracodorsal nerve and artery
18 Serratus anterior muscle
19 Subscapular artery
20 Pectoralis major muscle (reflected) and lateral pectoral nerve
21 Radial nerve and profunda brachii artery
22 Axillary nerve
23 Roots of the median nerve with axillary artery
24 **Musculocutaneous nerve**
25 Pectoralis minor muscle (reflected) and medial pectoral nerve
26 Posterior cord of brachial plexus
27 Clavicle (cut)
28 Lateral cord of brachial plexus
29 Medial cord of brachial plexus
30 Subclavian artery
31 Brachial vein

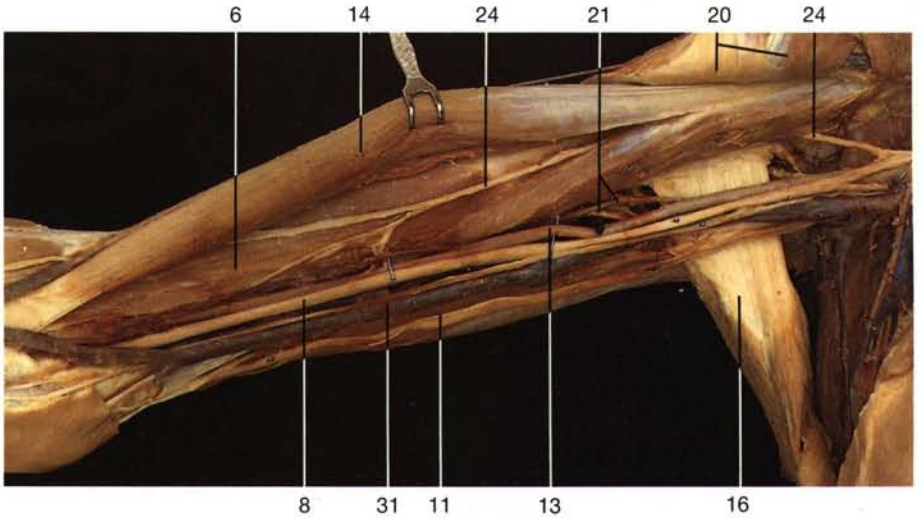

Right arm. Dissection of vessels and nerves, deeper layer. Biceps muscle has been reflected.

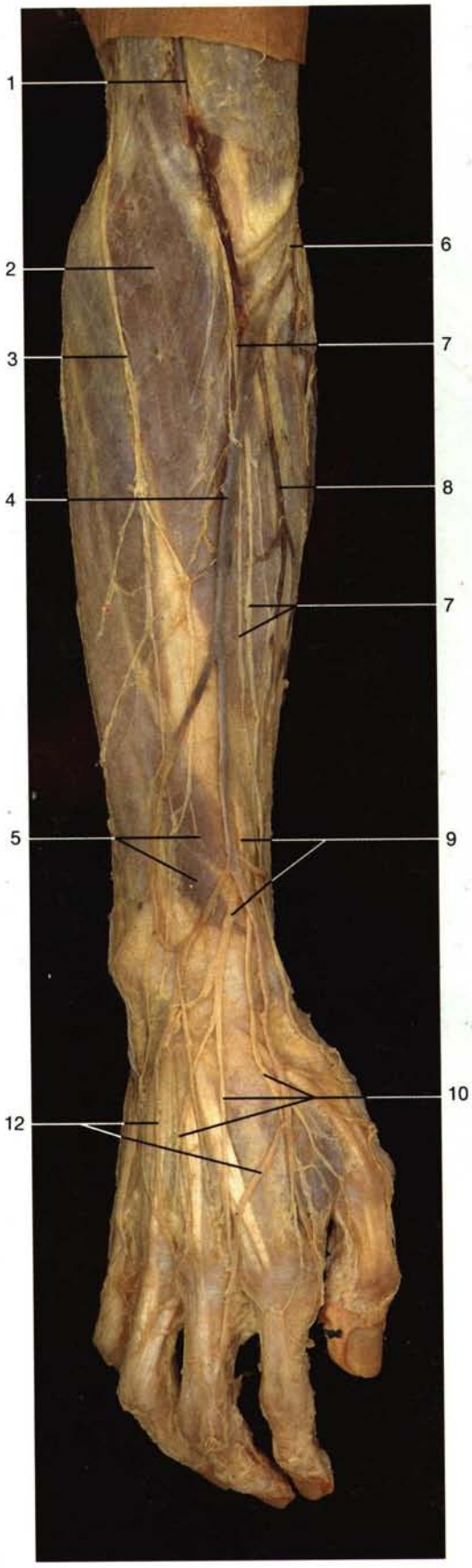

Superficial veins and cutaneous nerves of forearm and hand (anteromedial aspect).

Sagittal section through the forearm at the level of the humeroulnar articulation (MR-Scan, after A. Heuck, G. Luttke, and J. W. Rohen, 1994).

1 **Cephalic vein**
2 Brachioradialis muscle covered by its fascia
3 **Posterior cutaneous nerve of forearm** (branch of radialis nerve)
4 Cephalic vein of forearm
5 Extensor pollicis longus and brevis muscles covered by their fascia
6 **Median cubital vein**
7 Lateral cutaneous nerves of forearm (branch of musculocutaneous nerve)
8 Intermedian vein of forearm
9 Superficial branch of radial nerve
10 **Dorsal digital branches of radial nerve**
11 Triceps brachii muscle
12 Venous network on the dorsum of the hand
13 Olecranon
14 **Humeroulnar articulation**
15 Ulna
16 Radiocarpal articulation
17 Biceps brachii muscle
18 Trochlea of humerus
19 Flexor muscles of forearm
20 Carpal bones

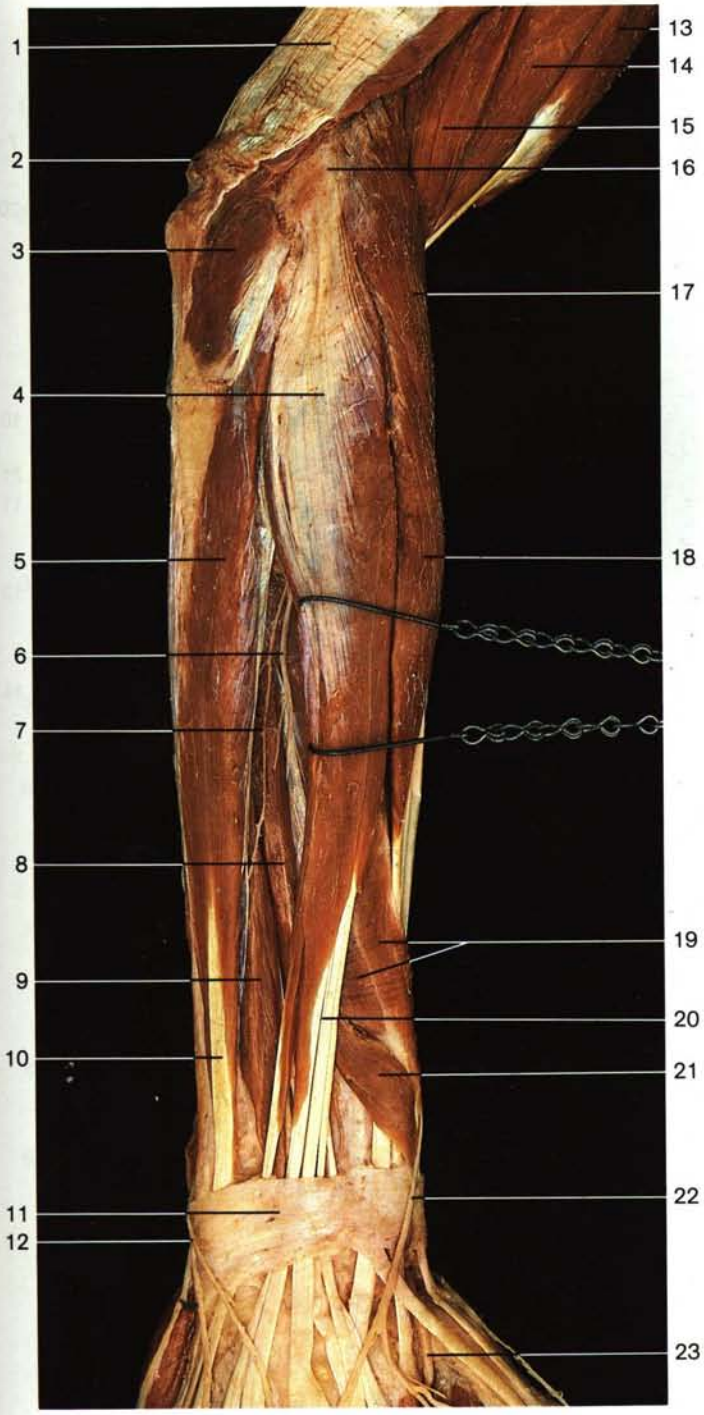

Vessels and nerves of right forearm, superficial layer (dorsal aspect).

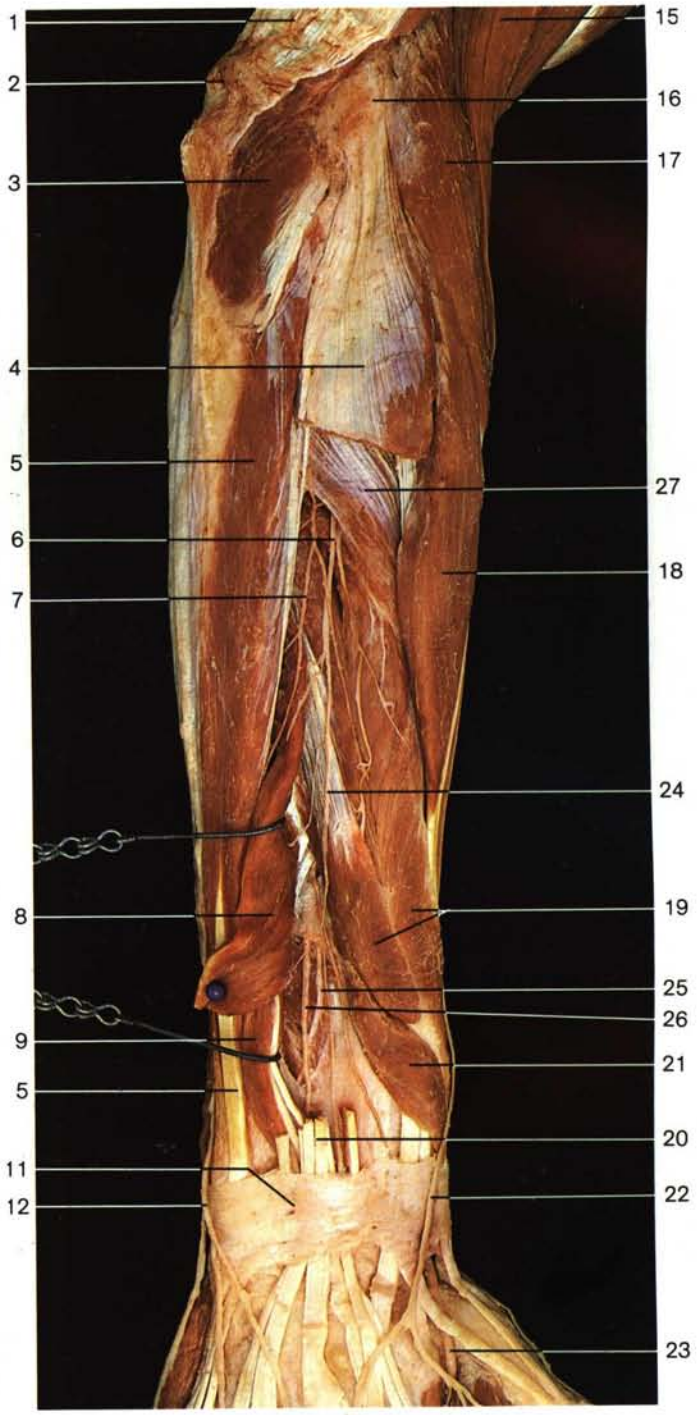

Vessels and nerves of right forearm, deep layer (dorsal aspect).

1 Tendon of triceps brachii muscle
2 Olecranon
3 Anconeus muscle
4 Extensor digitorum muscle
5 Extensor carpi ulnaris muscle
6 Deep branch of radial nerve
7 Posterior interosseous artery
8 Extensor pollicis longus muscle
9 Extensor indicis muscle
10 Tendon of extensor carpi ulnaris muscle
11 Extensor retinaculum
12 Dorsal branch of ulnar nerve
13 Biceps brachii muscle
14 Brachialis muscle
15 Brachioradialis muscle
16 Lateral epicondyle of humerus
17 Extensor carpi radialis longus muscle
18 Extensor carpi radialis brevis muscle
19 Abductor pollicis longus muscle
20 Tendons of extensor digitorum muscle
21 Extensor pollicis brevis muscle
22 Superficial branch of radial nerve
23 Radial artery
24 Posterior interosseous nerve
25 Posterior interosseous branch of radial nerve
26 Posterior branch of anterior interosseous artery
27 Supinator muscle

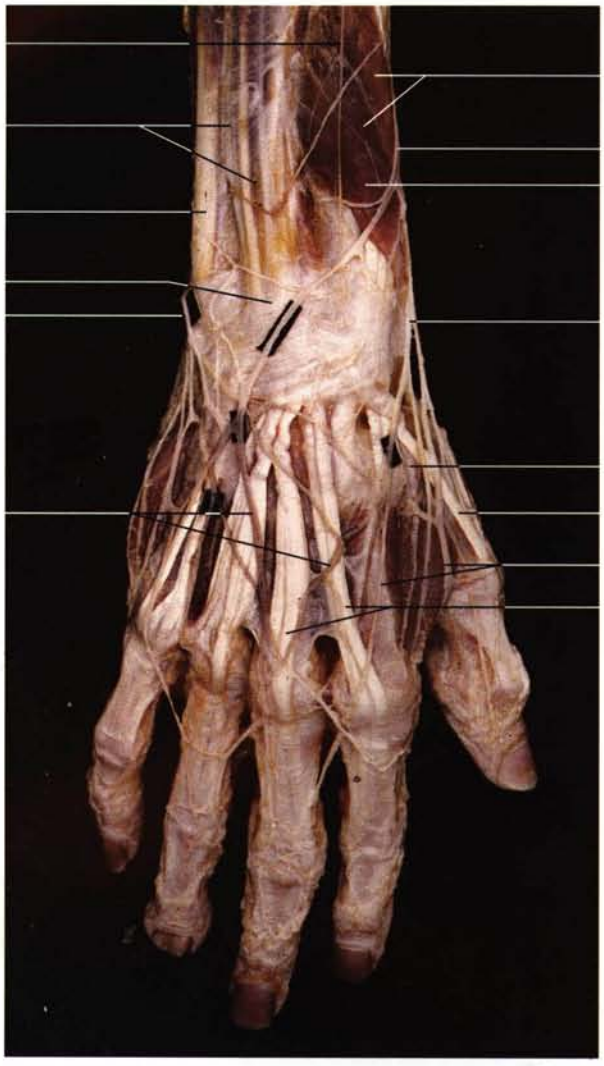

Cutaneous nerves and veins of forearm and hand (superficial layer, dorsal aspect).

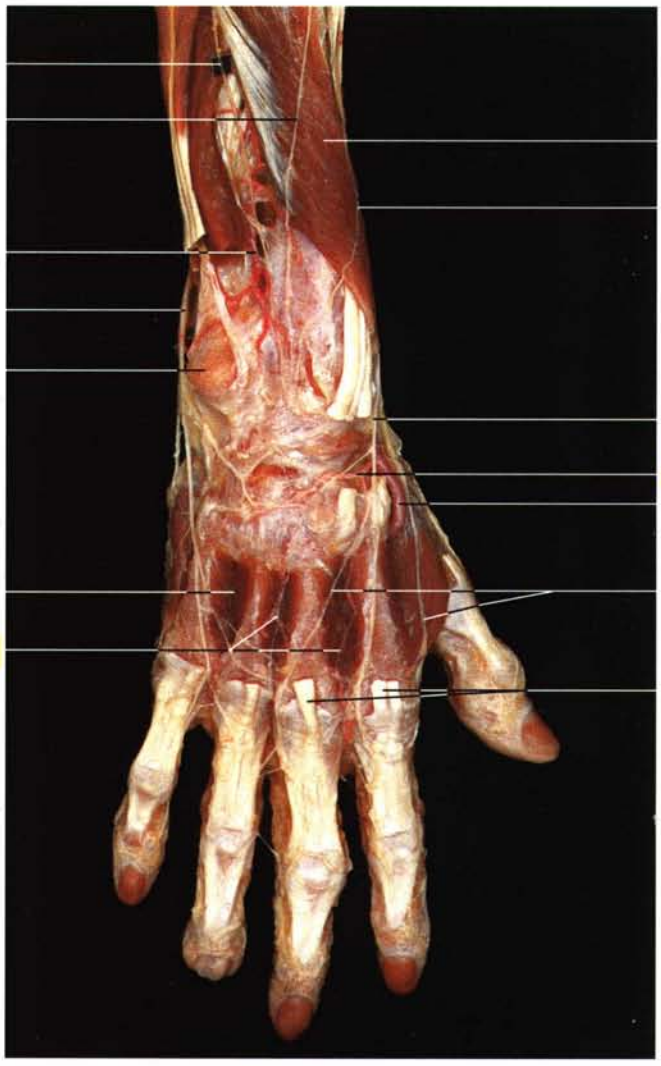

Dorsal region of forearm and hand (deeper layer). Extensor digitorum muscle has been partly removed.

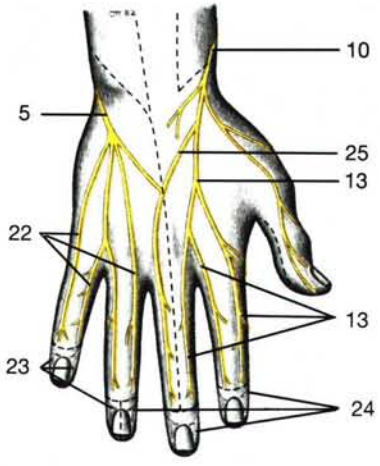

Innervation pattern of dorsal surfaces of hand
$2^1/_2$ digits by radial nerve, $2^1/_2$ digits by ulnar nerve. Note that the terminal branches to the dorsal surfaces of the distal phalanges are derived from the palmar digital nerves. The cutaneous distribution varies; often $3^1/_2$ digits are innervated by the radial and $1^1/_2$ digits by the ulnar nerve.

1 **Posterior cutaneous nerve of forearm** (branch of radial nerve)
2 Extensor digitorum muscle
3 Tendon of extensor carpi ulnaris muscle
4 Extensor retinaculum
5 **Ulnar nerve**
6 **Venous network of dorsum of hand**
7 Abductor pollicis longus muscle
8 Cephalic vein
9 Extensor pollicis brevis muscle
10 **Radial nerve,** superficial branch
11 **Radial artery**
12 Tendon of extensor pollicis longus muscle
13 Dorsal digital branches of radial nerve
14 Tendons of extensor digitorum muscle with intertendinous connections
15 Posterior interosseus nerve (branch of the deep radial nerve)
16 Posterior interosseous artery
17 Styloid process of ulna
18 Dorsal interosseus muscle IV
19 Dorsal carpal branch of radial artery
20 **Lateral cutaneous nerve of forearm** (branch of musculocutaneous nerve)
21 Dorsal metacarpal artery
22 Proper dorsal digital branches of ulnar nerve
23 Regions supplied by palmar digital nerves (ulnar nerve)
24 Regions supplied by palmar digital nerves (median nerve)
25 Communicating branch with ulnar nerve

Cubital region (anterior aspect), dissection of cutaneous nerves and veins.

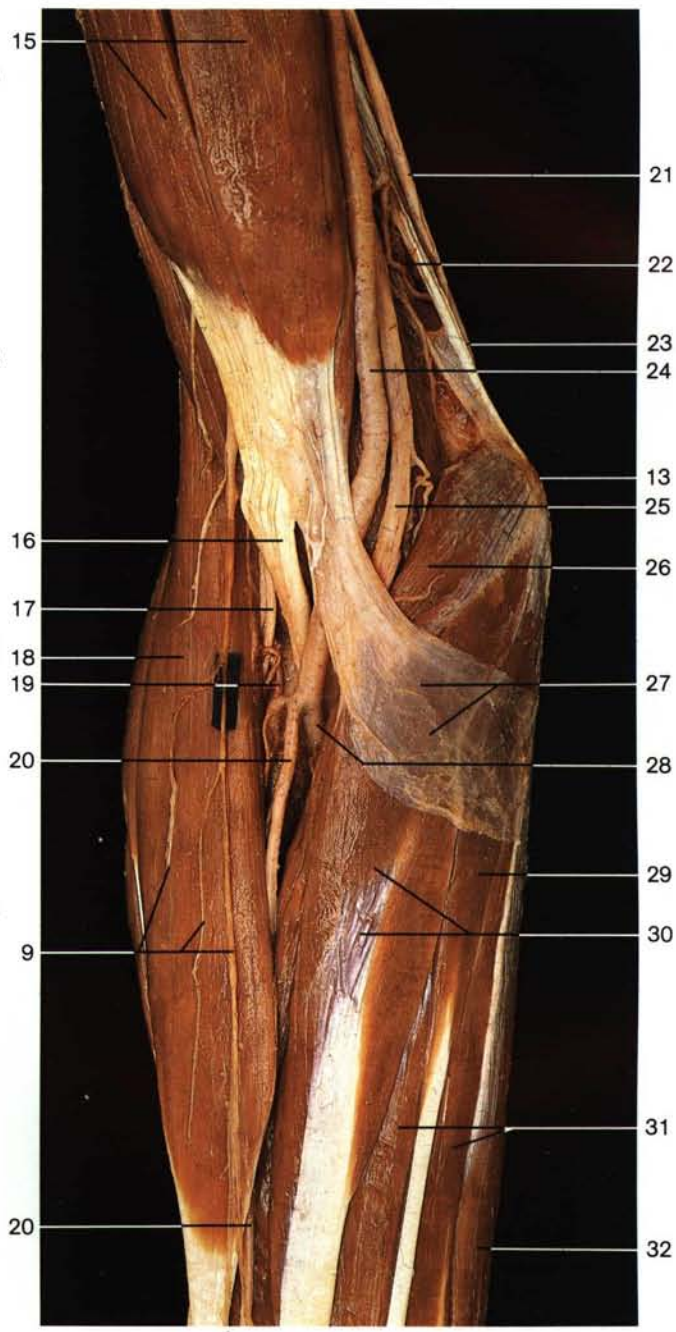

Cubital region, superficial layer (anterior aspect). The fasciae of the muscles have been removed.

1 Biceps brachii muscle with fascia
2 **Cephalic vein**
3 **Median cubital vein**
4 Lateral cutaneous nerve of forearm
5 Tendon and aponeurosis of biceps brachii muscle (covered by the antebrachial fascia)
6 Brachioradialis muscle with fascia
7 Accessory cephalic vein
8 Median vein of forearm
9 Branches of lateral cutaneous nerve of forearm
10 Terminal branches of medial cutaneous nerve of arm
11 **Medial cutaneous nerve of forearm**
12 **Basilic vein**
13 Medial epicondyle of humerus
14 Terminal branches of medial cutaneous nerve of forearm
15 Biceps brachii muscle
16 Tendon of biceps brachii muscle
17 Radial nerve
18 Brachioradiaiis muscle
19 Radial recurrent artery
20 **Radial artery**
21 **Ulnar nerve**
22 Superior ulnar collateral artery
23 Medial intermuscular septum
24 **Brachial artery**
25 **Median nerve**
26 Pronator teres muscle
27 Bicipital aponeurosis
28 Ulnar artery
29 Palmaris longus muscle
30 Flexor carpi radialis muscle
31 Flexor digitorum superficialis muscle
32 Flexor carpi ulnaris muscle

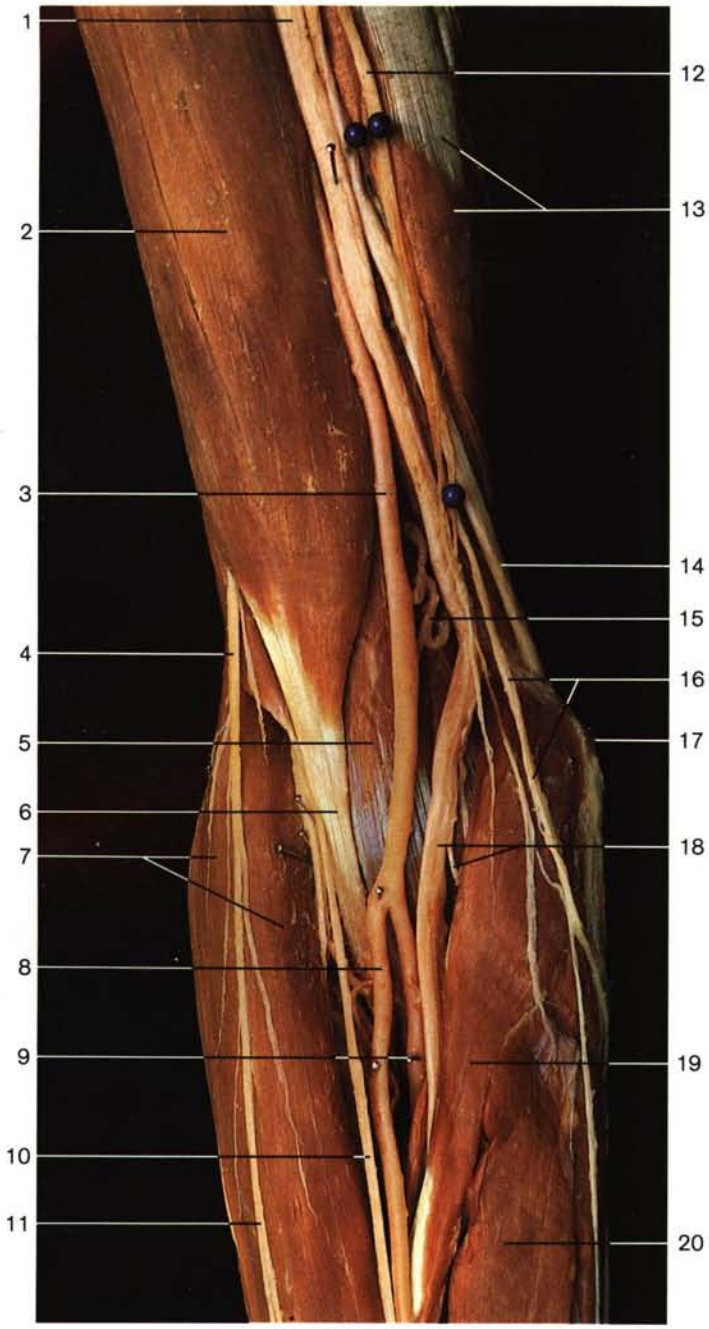

Cubital region, middle layer (anterior aspect). The bicipital aponeurosis has been removed.

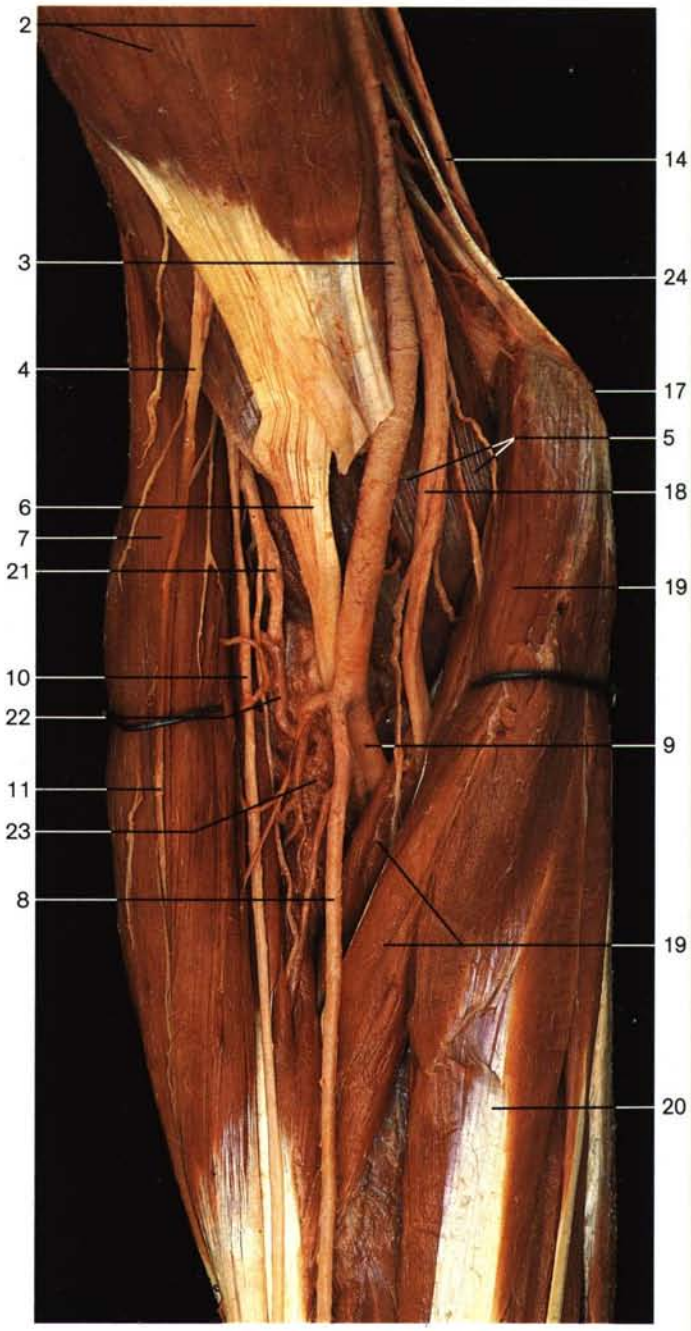

Cubital region, middle layer (anterior aspect). The pronator teres and brachioradialis muscles have been slightly reflected.

1 Median nerve
2 Biceps brachii muscle
3 Brachial artery
4 Lateral cutaneous nerve of forearm (terminal branch of **musculocutaneous nerve**)
5 Brachialis muscle
6 Tendon of biceps brachii muscle
7 Brachioradialis muscle
8 **Radial artery**
9 **Ulnar artery**
10 Superficial branch of radial nerve
11 Lateral cutaneous nerve of forearm
12 Medial cutaneous nerve of forearm
13 Triceps brachii muscle
14 **Ulnar nerve**
15 Inferior ulnar collateral artery
16 Anterior branch of medial cutaneous nerve of forearm
17 Medial epicondyle of humerus
18 **Median nerve** with branches to pronator teres muscle
19 Pronator teres muscle
20 Flexor carpi radialis muscle
21 Deep branch of radial nerve
22 Radial recurrent artery
23 Supinator muscle
24 Medial intermuscular septum of arm

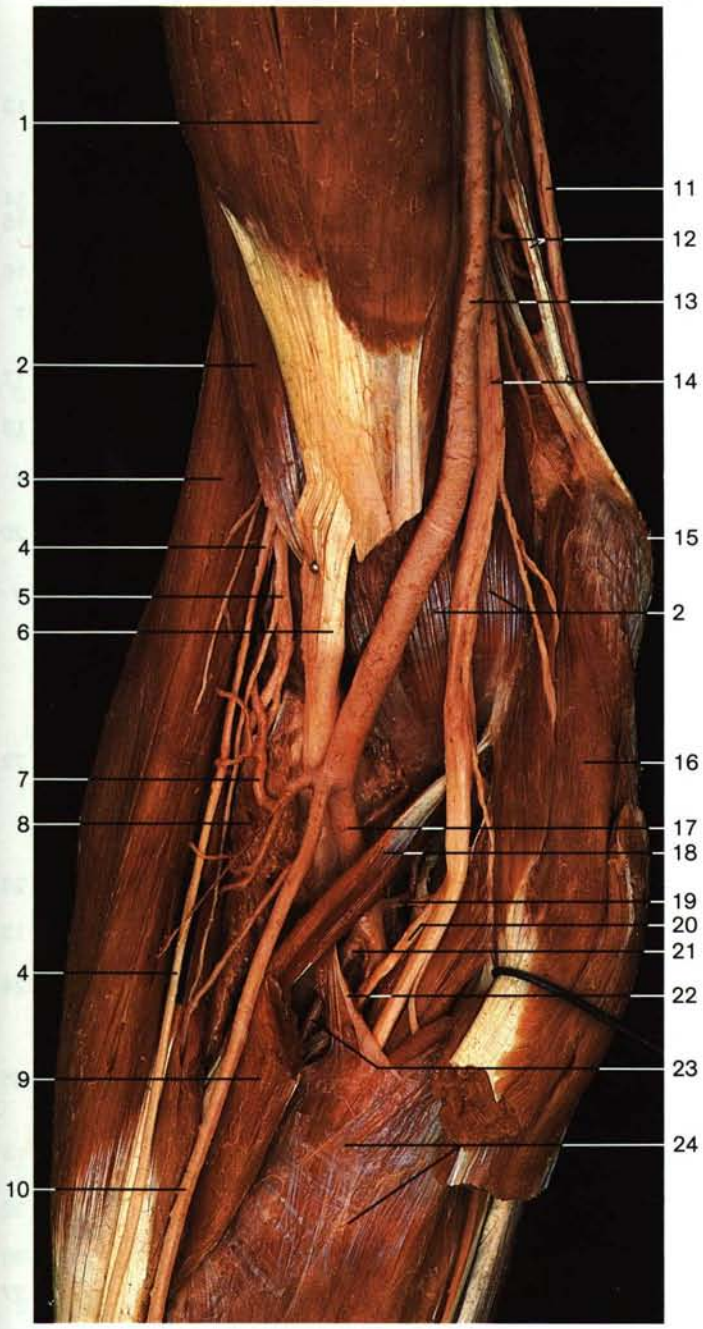

Cubital region, deep layer (anterior aspect). The pronator teres and flexor carpi ulnaris muscles have been cut and reflected.

Cubital region, deepest layer (anterior aspect). The flexor digitorum superficialis and the ulnar head of the pronator teres have been cut and reflected.

1 Biceps brachii muscle
2 Brachialis muscle
3 Brachioradialis muscle
4 Superficial branch of radial nerve
5 Deep branch of radial nerve
6 Tendon of biceps brachii muscle
7 **Radial recurrent artery**
8 Supinator muscle
9 Insertion of pronator teres muscle
10 **Radial artery**
11 **Ulnar nerve**
12 Medial intermuscular septum of arm and superior ulnar collateral artery
13 **Brachial artery**
14 **Median nerve**
15 Medial epicondyle of humerus
16 Humeral head of pronator teres muscle
17 **Ulnar artery**
18 Ulnar head of pronator teres muscle
19 **Ulnar recurrent artery**
20 Anterior interosseous nerve
21 Common interosseous artery
22 Tendinous arch of flexor digitorum superficialis muscle
23 Anterior interosseous artery
24 Flexor digitorum superficialis muscle
25 Flexor digitorum profundus muscle
26 Flexor pollicis longus muscle

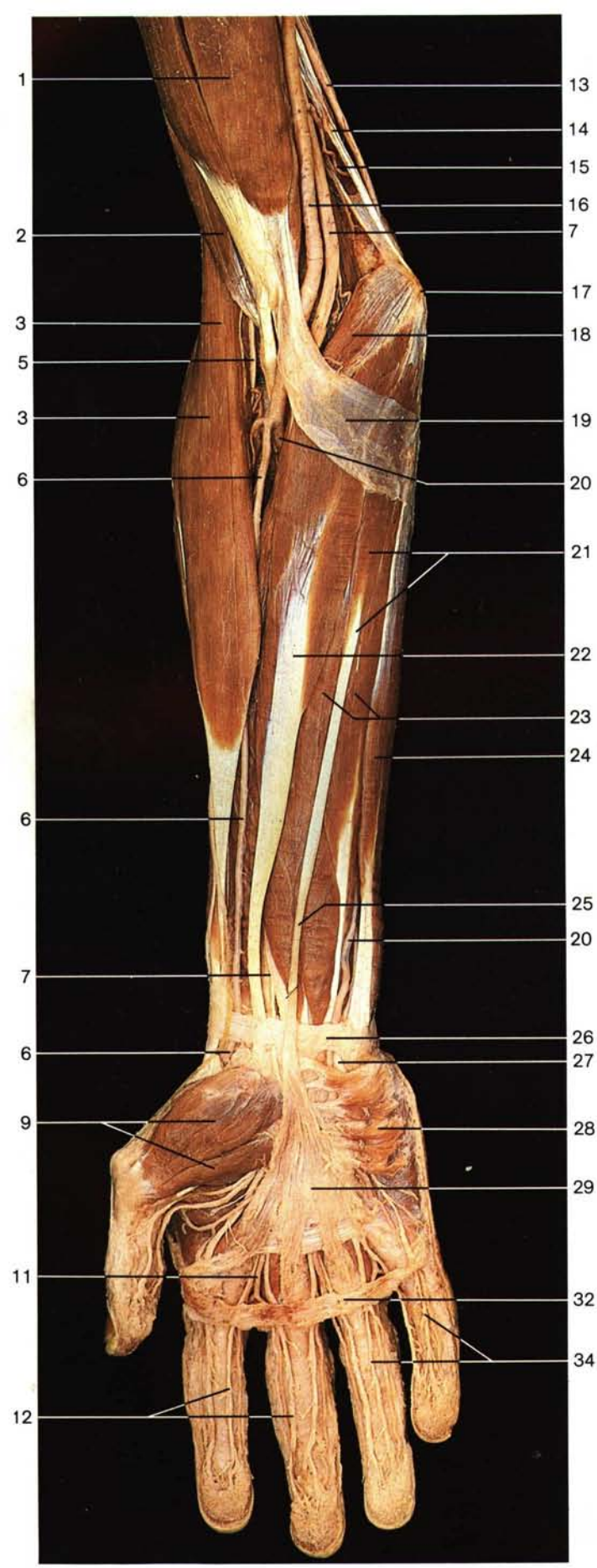

Vessels and nerves of right forearm and hand, superficial layer (palmar aspect).

Vessels and nerves of right forearm and hand, superficial layer (palmar aspect). The palmar aponeurosis of the hand and the bicipital aponeurosis have been removed.

1 Biceps brachii muscle
2 Brachialis muscle
3 Brachioradialis muscle
4 **Deep branch of radial nerve**
5 **Superficial branch of radial nerve**
6 **Radial artery**
7 **Median nerve**
8 Flexor retinaculum
9 Thenar muscles
10 Common palmar digital branches of median nerve
11 Common palmar digital arteries
12 Proper palmar digital nerves (median nerve)
13 **Ulnar nerve**
14 Medial intermuscular septum of arm
15 Superior ulnar collateral artery
16 Brachial artery
17 Medial epicondyle of humerus
18 Pronator teres muscle
19 Bicipital aponeurosis
20 **Ulnar artery**
21 Palmaris longus muscle
22 Flexor carpi radialis muscle
23 Flexor digitorum superficialis muscle
24 Flexor carpi ulnaris muscle
25 Tendon of palmaris longus muscle
26 Remnant of antebrachial fascia
27 Superficial branch of ulnar nerve
28 Palmaris brevis muscle
29 Palmar aponeurosis
30 Hypothenar muscles
31 **Superficial palmar arch**
32 Superficial transverse metacarpal ligament
33 Common palmar digital branch of ulnar nerve
34 Proper palmar digital branches of ulnar nerve
35 Anterior interosseous artery and nerve
36 Flexor digitorum profundus muscle
37 Common palmar digital arteries
38 Palmar branch of median nerve
39 Flexor pollicis longus muscle
40 Palmar branch of ulnar nerve

Vessels and nerves of forearm and hand, deep layer (anterior aspect). The superficial layer of the flexor muscles has been removed.

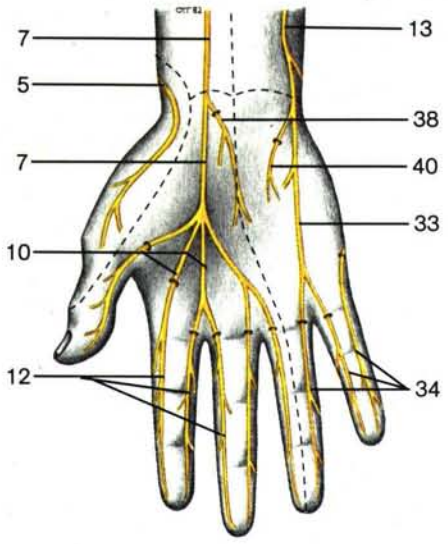

Cutaneous innervation of hand (palmar aspect). (Schematic drawing.)
Cutaneous innervation of palmar surface:
3½ digits by median nerve,
1½ digits by ulnar nerve.

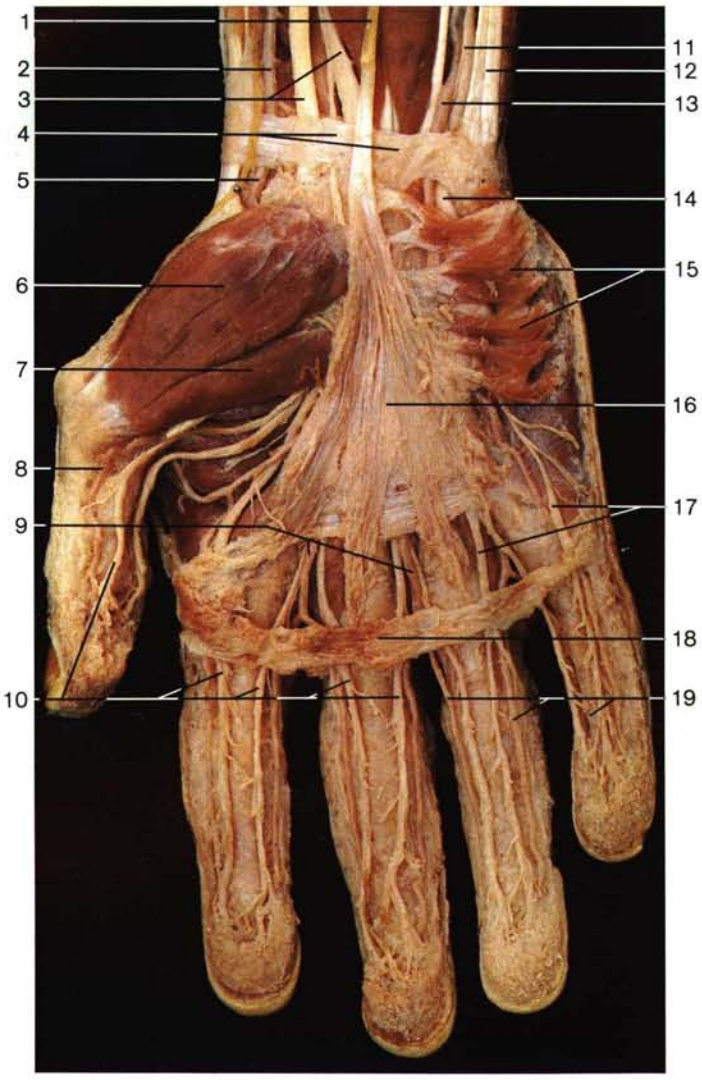

Right hand, superficial layer, dissection of vessels and nerves (palmar aspect).

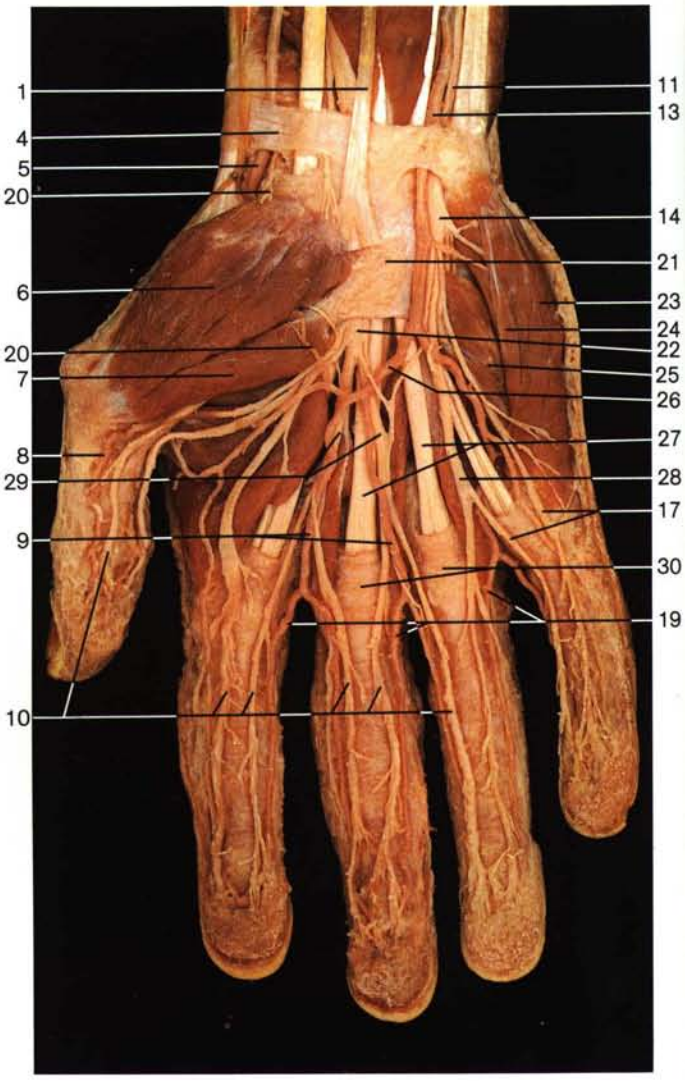

Right hand, superficial layer, dissection of vessels and nerves (palmar aspect). The palmar aponeurosis has been removed to display the superficial palmar arch.

1 Tendon of palmaris longus muscle
2 **Radial artery**
3 Tendon of flexor carpi radialis muscle and median nerve
4 Distal part of antebrachial fascia
5 Radial artery passing into the anatomical snuffbox
6 Abductor pollicis brevis muscle
7 Superficial head of flexor pollicis brevis muscle
8 Palmar digital artery of thumb
9 Common palmar digital arteries
10 Proper palmar digital nerves (median nerve)
11 **Ulmar nerve**
12 Tendon of flexor carpi ulnaris muscle
13 **Ulnar artery**
14 Superficial branch of ulnar nerve
15 Palmaris brevis muscle
16 Palmar aponeurosis
17 Palmar digital nerves (ulnar nerve)
18 Superficial transverse metacarpal ligament
19 Proper palmar digital arteries
20 Superficial palmar branch of radial artery (contributing to the superficial palmar arch)
21 Flexor retinaculum
22 **Median nerve**
23 Abductor digiti minimi muscle
24 Flexor digiti minimi brevis muscle
25 Opponens digiti minimi muscle
26 **Superficial palmar arch**
27 Tendons of flexor digitorum superficialis muscle
28 Common palmar digital branch of ulnar nerve
29 Common palmar digital branch of median nerve
30 Fibrous sheath of flexor tendons

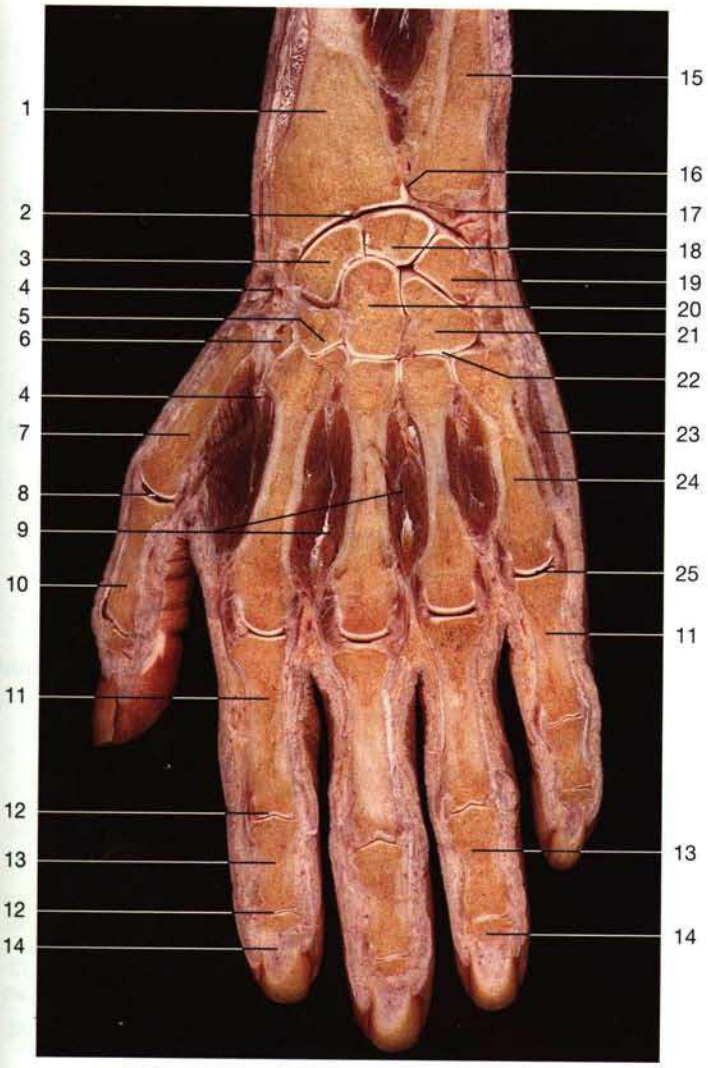

Coronal section through the right hand (palmar aspect).

Coronal section through the right hand (palmar aspect) (MR-Scan, courtesy of Dr. A. Heuck, Munich).

1 Radius
2 **Radiocarpal articulation**
3 **Scaphoid** (navicular) **bone**
4 **Radial artery**
5 Trapezoid bone
6 **Trapezium bone**
7 First metacarpal bone
8 Metacarpophalangeal articulation of thumb
9 Interosseous muscles
10 Proximal phalanx of thumb
11 Proximal phalanx of fingers
12 Interphalangeal articulations
13 Middle phalanx
14 Distal phalanx
15 **Ulna**
16 Distal radioulnar articulation
17 Articular disk
18 **Lunate bone**
19 **Triangular bone**
20 **Capitate bone**
21 **Hamate bone**
22 Carpometacarpal articulations
23 Abductor digiti minimi muscle
24 Fifth metacarpal bone
25 Metacarpophalangeal articulation
26 Adductor pollicis muscle
27 Proper palmar digital arteries

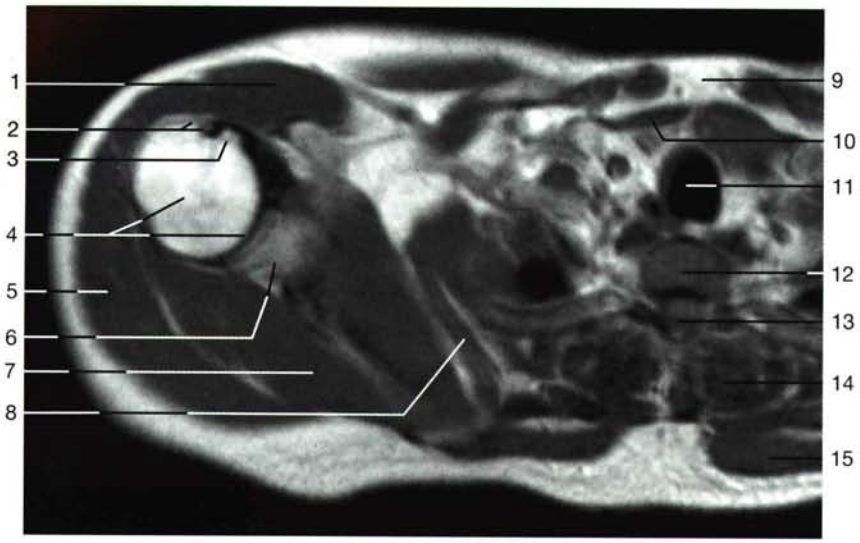

Horizontal section through shoulder joint (section 1; MR-Image; inferior aspect).

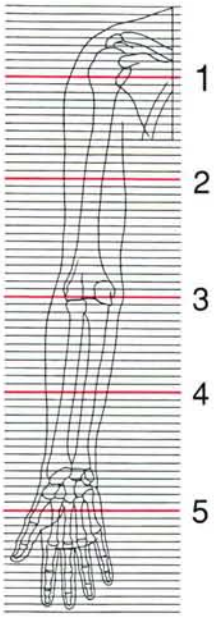

Upper extremity, location of sections 1–5 (MR-Scans, courtesy of Dr. A. Heuck, Munich).

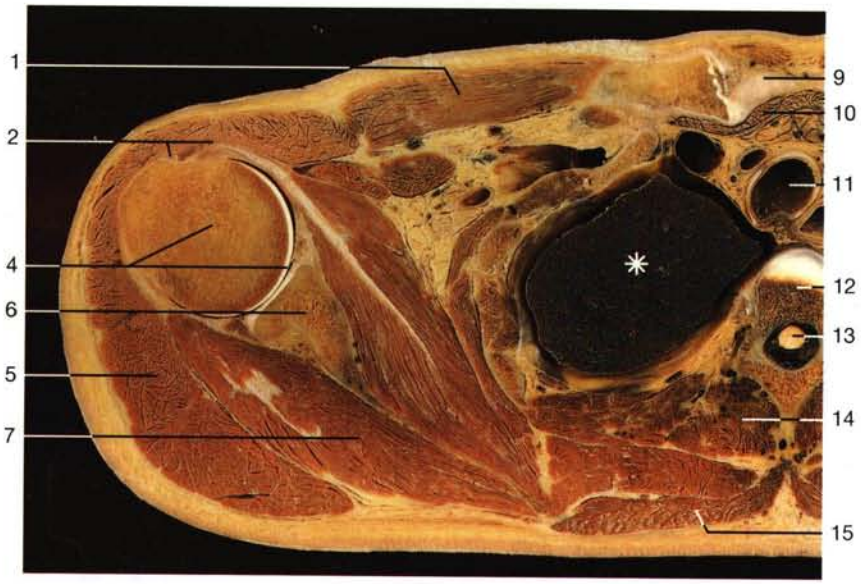

Horizontal section through the right shoulder at the level of T_1 (section 1; inferior aspect). * = Upper lobe of lung.

1 Pectoralis major muscle
2 Greater tubercle and tendon of biceps muscle
3 Lesser tubercle
4 Head of humerus and articular cavity of shoulder joint
5 Deltoid muscle
6 Scapula
7 Infraspinatus muscle
8 Serratus anterior muscle
9 Sternum
10 Infrahyoid muscles
11 Trachea
12 Body of thoracic vertebra
13 Vertebral canal and spinal cord
14 Deep muscles of the back
15 Trapezius muscle
16 Brachialis muscle
17 Radial nerve and profunda brachii vessels

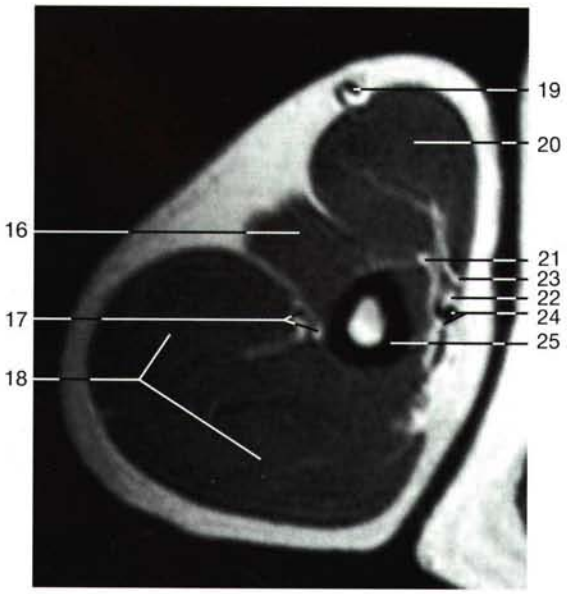

Axial section through the middle of the right arm (section 2; MR-Image, inferior aspect).

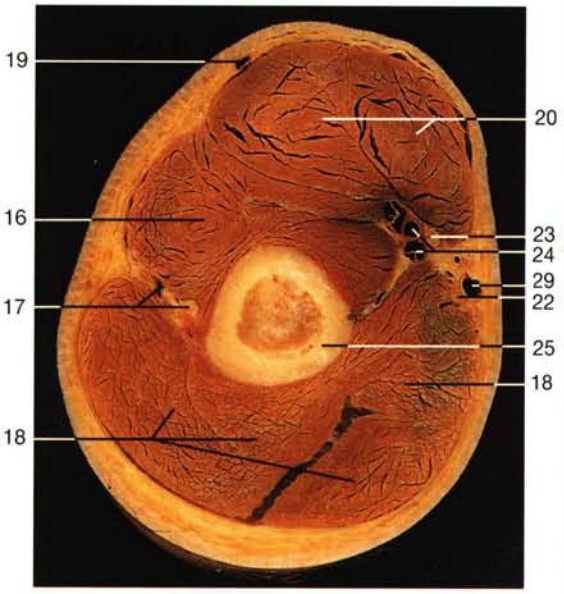

Axial section through the middle of the right arm (section 2; inferior aspect).

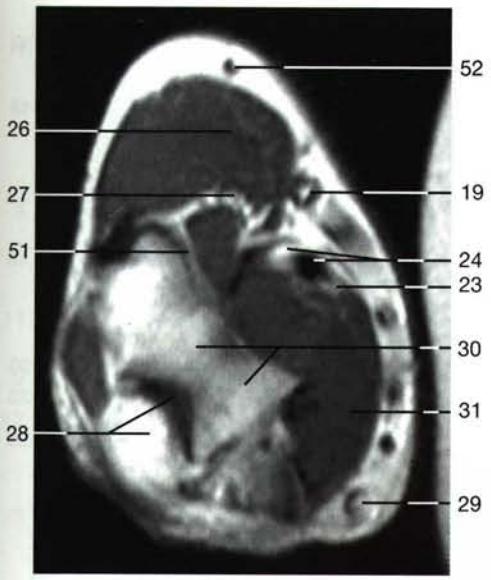

Axial section through the elbow joint (section 3; MR-Image; inferior aspect).

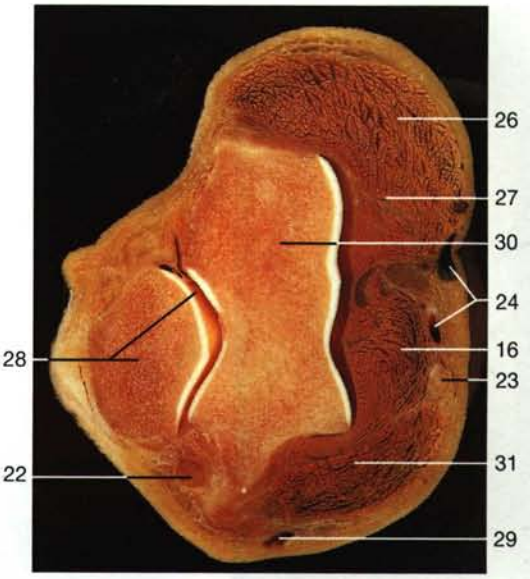

Axial section through the right elbow joint (section 3; inferior aspect).

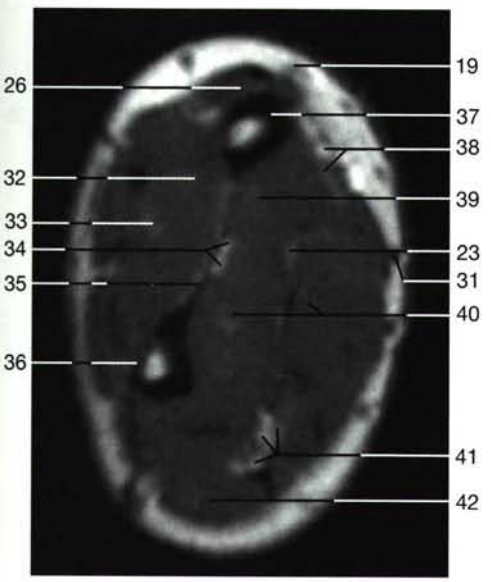

Axial section through the middle of the forearm (section 4; MR-Image; inferior aspect).

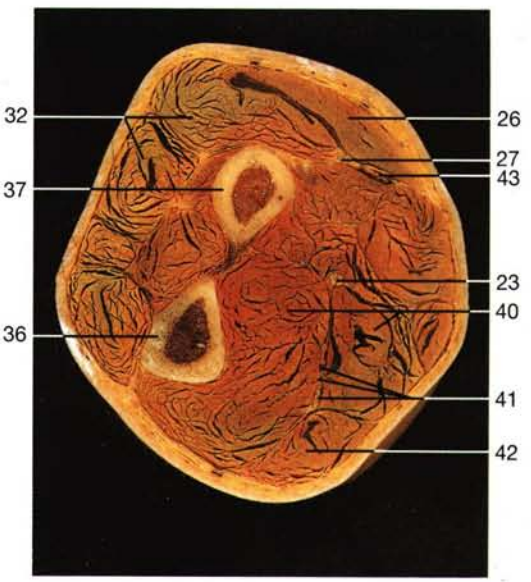

Axial section through the right forearm (section 4; inferior aspect).

18 Triceps brachii muscle
19 Cephalic vein
20 Biceps brachii muscle
21 Musculocutaneous nerve
22 **Ulnar nerve**
23 **Medianus nerve**
24 Brachial artery and vein
25 Shaft of humerus
26 Brachioradialis muscle
27 **Radial nerve**
28 Olecranon and articular cavity of elbow joint
29 Basilic vein
30 Humerus
31 Pronator teres muscle
32 Extensor muscles of forearm
33 Ramus profundus of radialis nerve
34 Anterior interosseus vessels and nerve
35 Interosseous membrane
36 Ulna
37 Radius
38 Radial artery and superficial branch of radial nerve
39 Flexor pollicis longus muscle
40 Flexor digitorum superficialis and profundus muscles
41 Ulnar nerve, ulnar artery and vein
42 Flexor carpi ulnaris muscle
43 Radial artery
44 Metacarpal bones III and IV
45 **Carpal canal** with tendons of flexor digitorum muscles
46 Hypothenar muscle
47 Median nerve
48 Interosseus muscles
49 First metacarpal bone
50 Thenar muscles
51 Articular cavity of humeroradial joint
52 Median cubital vein

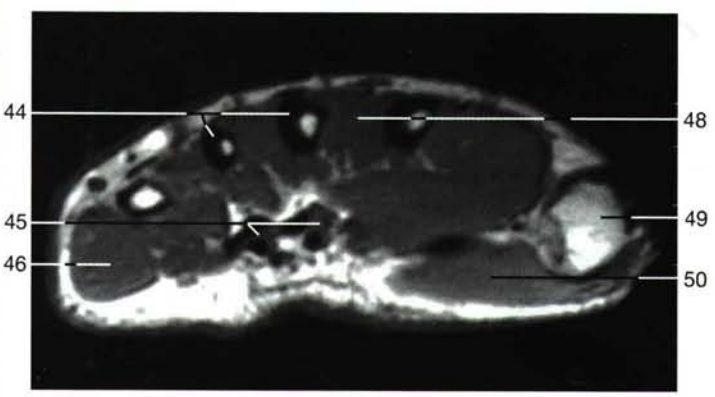

Axial section through the right hand (metacarpus; MR-Image; section 5; inferior aspect).

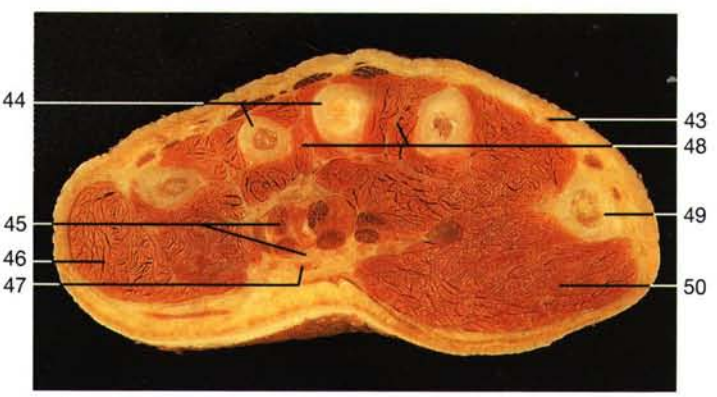

Axial section through the right hand at the level of the metacarpus (section 5; inferior aspect).

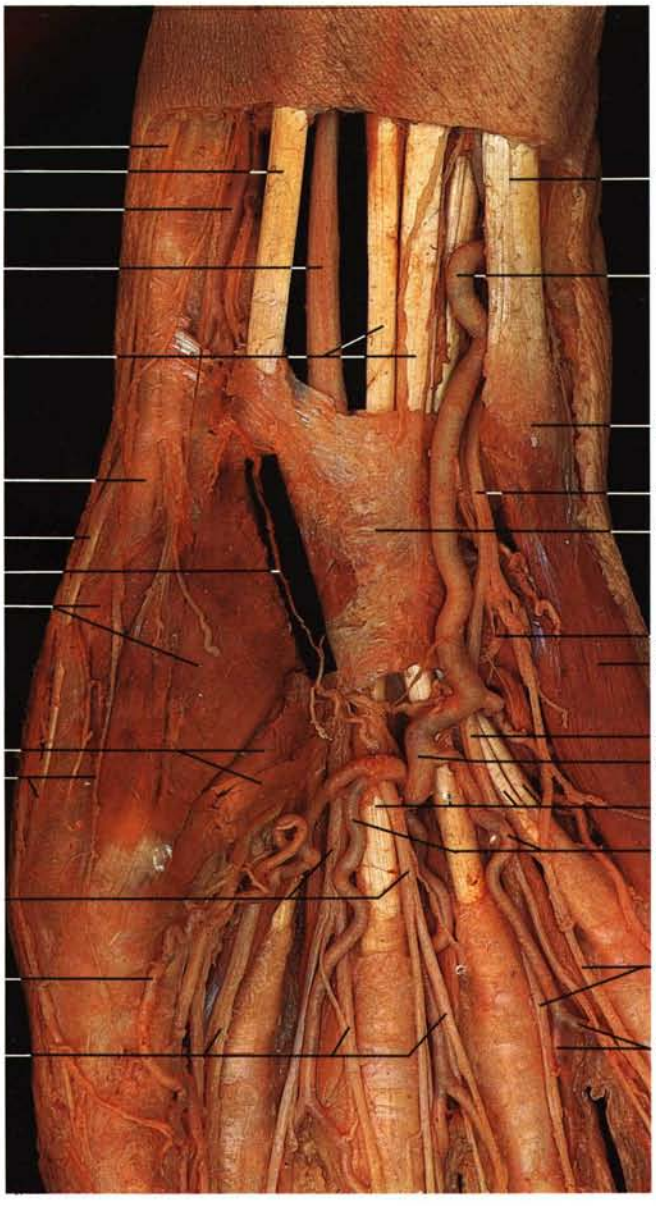

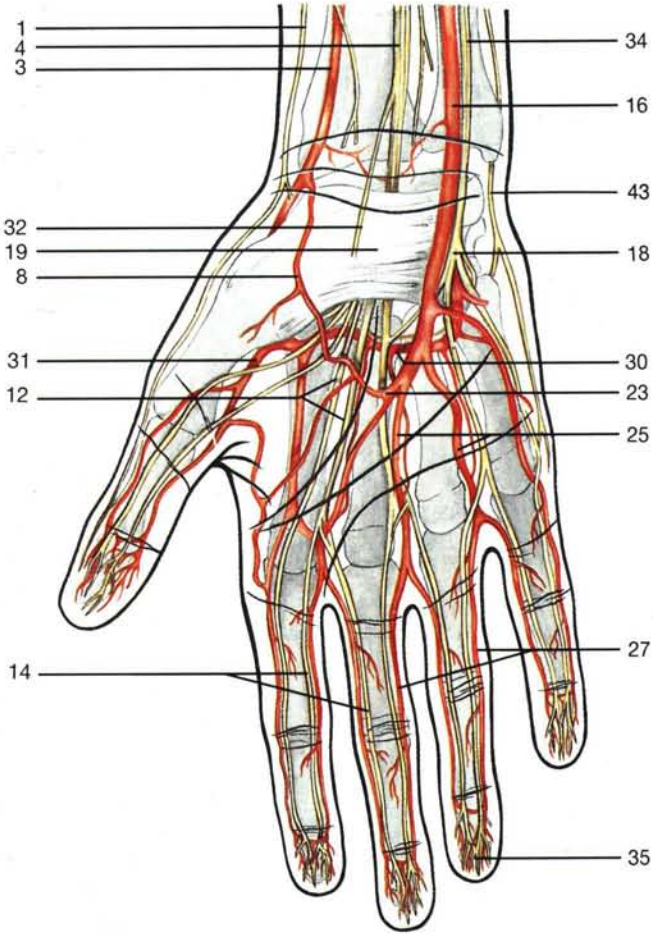

Arteries and nerves of the right hand (palmar aspect). (Schematic drawing.)

◁
Right hand, superficial layer (palmar aspect). Dissection of the superficial palmar arch.

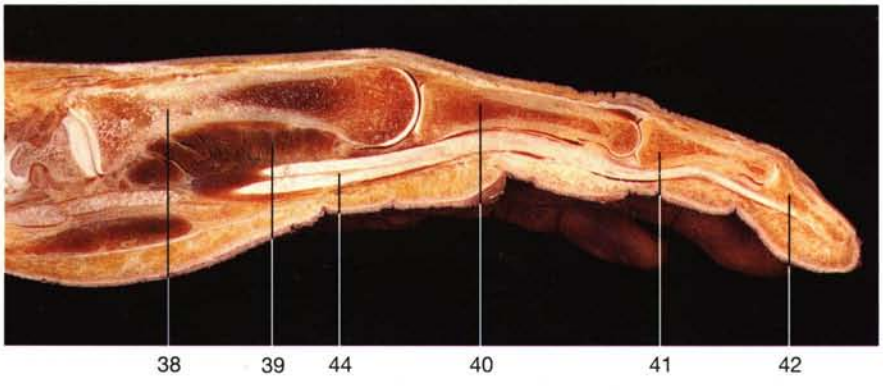

Longitudinal section through the hand at the level of the third finger

Longitudinal section through the hand at the level of the third finger (MR-Scan, courtesy of Dr. A. Heuck, Munich).

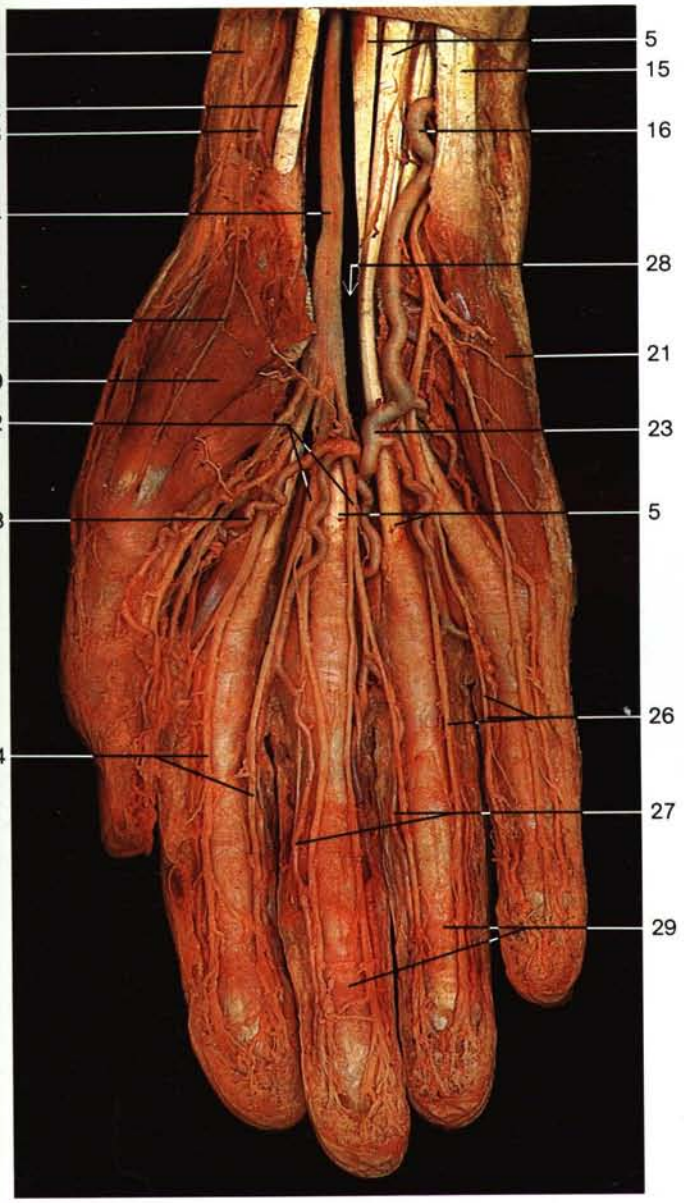

Right hand, middle layer (palmar aspect). The flexor retinaculum has been removed.

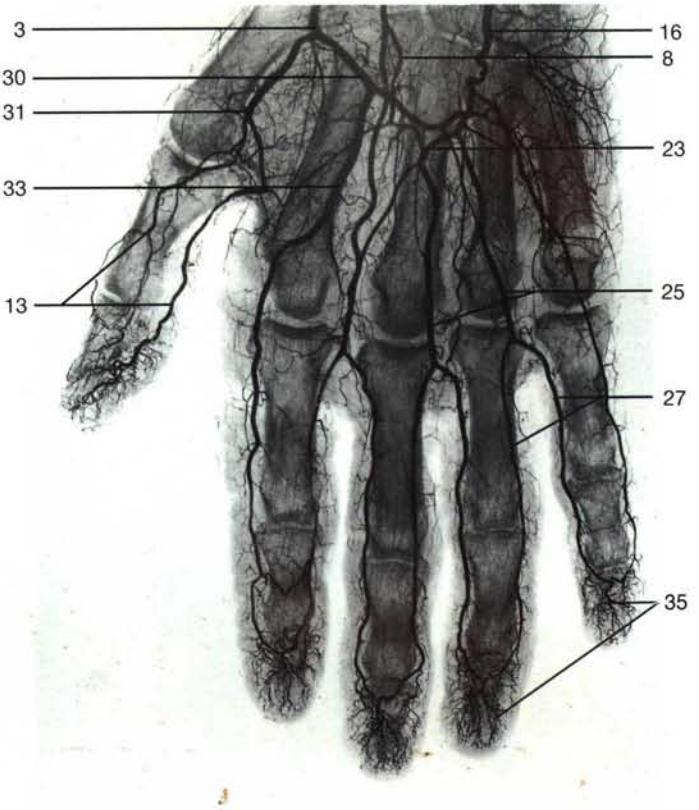

Arteriogram of the right hand (palmar aspect).

1 Superficial branch of radial nerve
2 Tendon of flexor carpi radialis muscle
3 **Radial artery**
4 **Median nerve**
5 Tendon of flexor digitorum superficialis muscle
6 Tendon of abductor pollicis longus muscle
7 Tendon of extensor pollicis brevis muscle
8 Superficial palmar branch of radial artery
9 Abductor pollicis brevis muscle
10 Superficial head of flexor pollicis brevis muscle
11 Terminal branches of superficial branch of radial nerve
12 Common palmar digital nerves (median nerve)
13 Proper palmar digital arteries of thumb
14 Proper palmar digital nerves (median nerve)
15 Tendon of flexor carpi ulnaris muscle
16 **Ulnar artery**
17 Position of pisiform bone
18 Superficial branch of ulnar nerve
19 Flexor retinaculum
20 Deep branch of ulnar nerve
21 Abductor digiti minimi muscle
22 Common palmar digital nerves (ulnar nerve)
23 **Superficial palmar arch**
24 Tendons of flexor digitorum muscles
25 Common palmar digital arteries
26 **Palmar digital nerves (ulnar nerve)**
27 **Proper palmar digital arteries**
28 **Carpal tunnel**
29 Fibrous sheaths for the tendons of flexor digitorum muscles
30 **Deep palmar arch**
31 Princeps pollicis artery
32 Palmar branch of median nerve
33 Common digital palmar artery
34 Ulnar nerve
35 Capillary network of finger
36 Radius
37 Carpal bones
38 Metacarpal bone
39 Interosseus muscles
40 Proximal phalanx
41 Middle phalanx
42 Distal phalanx
43 Dorsal branch of ulnar nerve
44 Tendons of flexor digitorum profundus (upper) and superficialis (lower) muscles

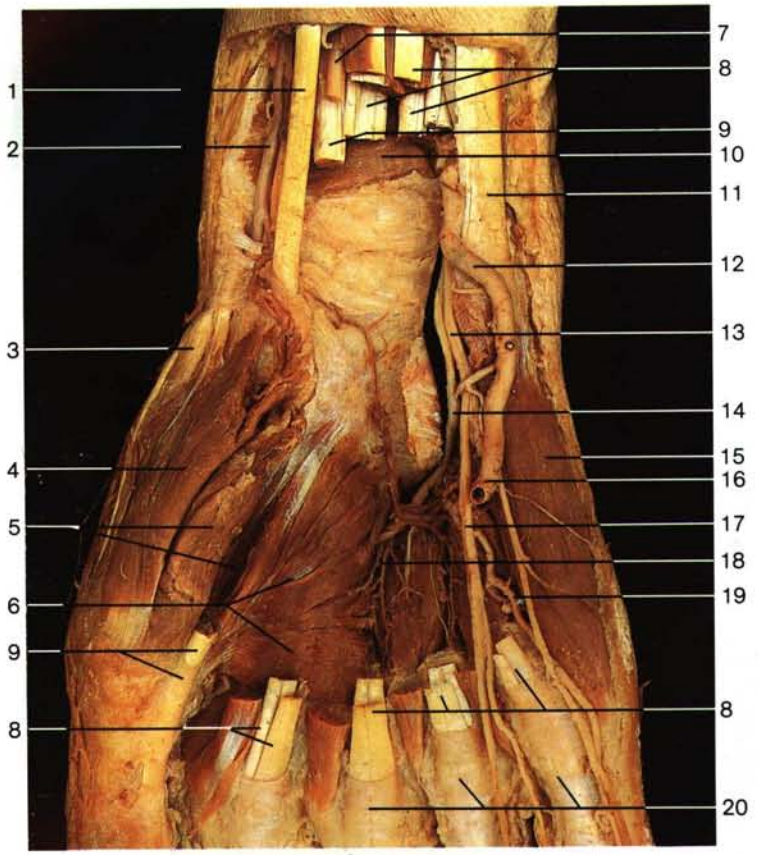

Right hand, deep layer (palmar aspect). The carpal tunnel has been opened, the tendons of the flexor muscles have been removed and the superficial palmar arch has been cut.

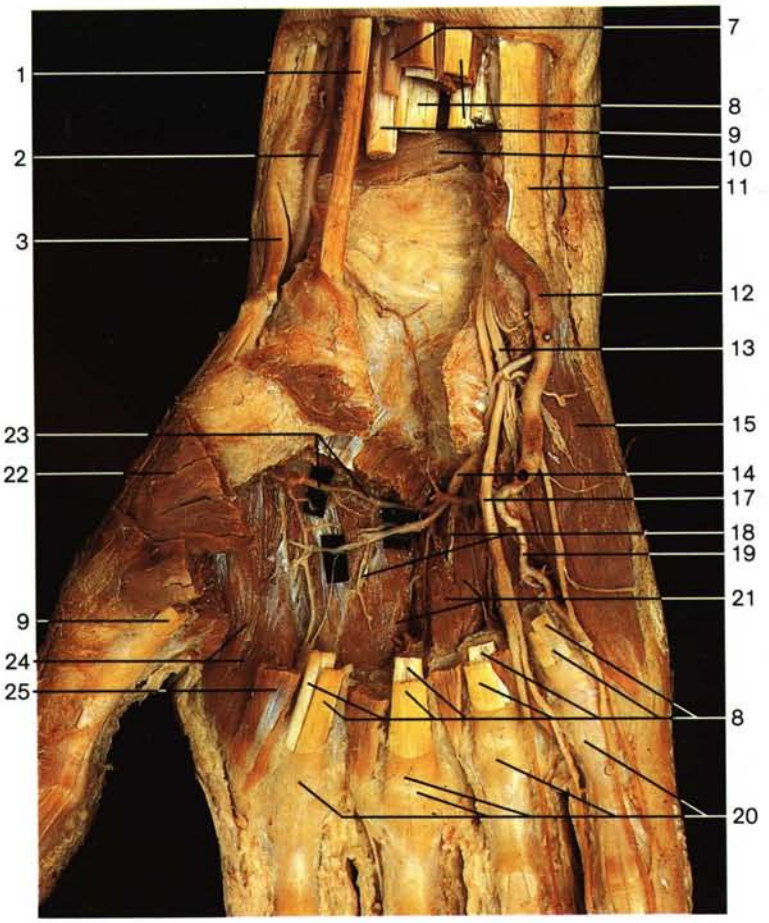

Right hand, deep layer (palmar aspect). Dissection of the deep palmar arch.

1 Tendon of flexor carpi radialis muscle
2 Radial artery
3 Tendon of abductor pollicis longus muscle
4 Abductor pollicis brevis muscle
5 Superficial and deep heads of flexor pollicis brevis muscle
6 Oblique and transverse heads of adductor pollicis muscle
7 Median nerve
8 Tendons of flexor digitorum superficialis and profundus muscles
9 Tendon of flexor pollicis longus muscle
10 Pronator quadratus muscle
11 Tendon of flexor carpi ulnaris muscle
12 Ulnar artery
13 Superficial branch of ulnar nerve
14 Deep branch of ulnar nerve
15 Abductor digiti minimi muscle
16 **Superficial palmar arch** (cut end)
17 Common palmar digital nerves (ulnar nerve)
18 Palmar metacarpal arteries of deep palmar arch
19 Palmar digital artery of the 5th finger
20 Fibrous sheaths of tendons of flexor muscles
21 Palmar interosseous muscles
22 Opponens pollicis muscle (cut)
23 **Deep palmar arch**
24 1st dorsal interosseous muscle
25 1st lumbrical muscle

Coronal section through pelvis and thighs (MR-Scan of the male, courtesy of Dr. A. Heuck, Munich).

Coronal section through the legs (MR-Scan, courtesy of Dr. A. Heuck, Munich).

1 Sacral promontory
2 Gluteus medius muscle
3 Small intestine and urinary bladder
4 Acetabulum
5 **Head of femur**
6 Greater trochanter of femur
7 Vastus lateralis muscle
8 **Femur**
9 Adductor muscles
10 **Knee joint** with menisci
11 **Tibia**
12 Soleus muscle
13 Tibialis anterior muscle
14 Distal tibiofibular articulation
15 **Talocrural articulation**
16 Fibula (lateral malleolus)

The lower limb (extremity) is specialized for support of the upright posture, locomotion and maintaining balance. In contrast to the upper limb, the lower limb is more restricted in its movements, and the joints are more tight and fixed by strong ligaments. The hip joint is a ball and socket type of synovial joint between head of femur and acetabulum. The knee joint is a hinge type of synovial joint that permits only limited rotation. The talocrural joint is a hinge joint between talus, fibula, and tibia only allowing movements of flexion and extension.

The long axis of the foot is at right angle to that of the leg, thus, forming an effective arch for the upright stance of man.

A Pelvic girdle
B Thigh
C Leg
D Foot
1 Right hip bone
2 Sacrum
3 Femur
4 Patella
5 Fibula
6 Tibia
7 Tarsal bones
8 Metatarsal bones
9 Phalanges
10 Sacroiliac joint
11 Pubic symphysis
12 Hip joint
13 Knee joint
14 Proximal tibiofibular joint
15 Distal tibiofibular joint
16 Talocrural joint
17 Talocalcaneonavicular joint
18 Tarsometatarsal joints
19 Metatarsophalangeal joints
20 Interphalangeal joints

The pelvic girdle is firmly connected to the vertebral column at the sacroiliac joint. Therefore the body can be kept upright more easily even if only one limb is used for support (as in walking). The mobility of the lower limb is more limited than that of the upper limb.

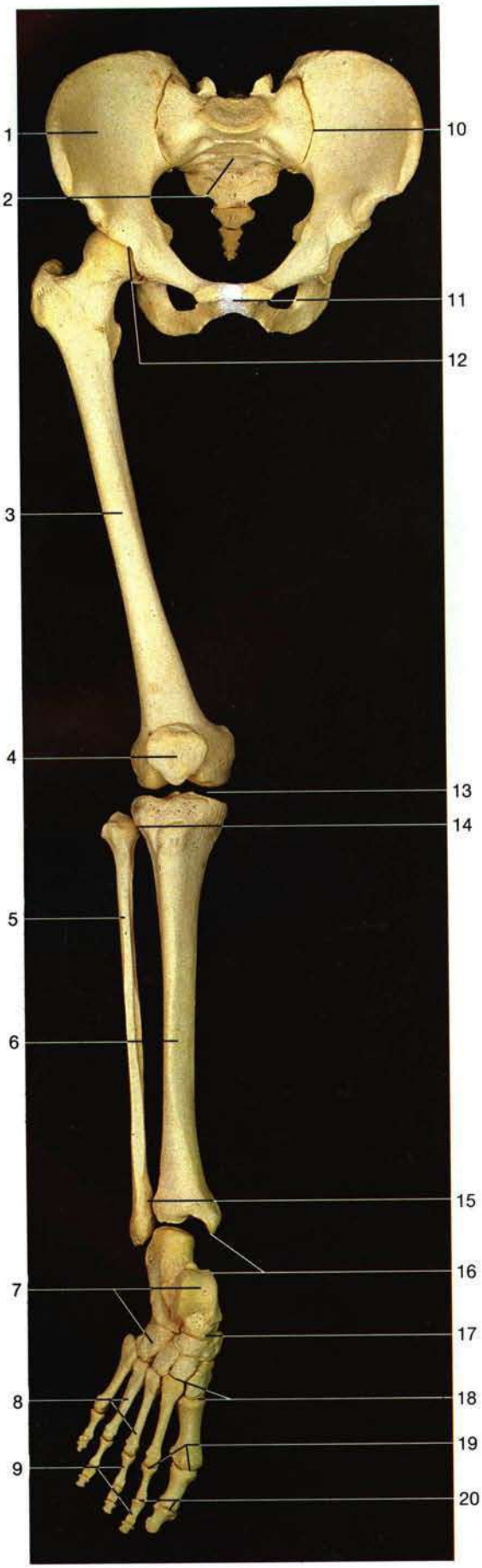

Skeleton of pelvic girdle and lower limb (anterior aspect). The talocrural joint has been dislocated.

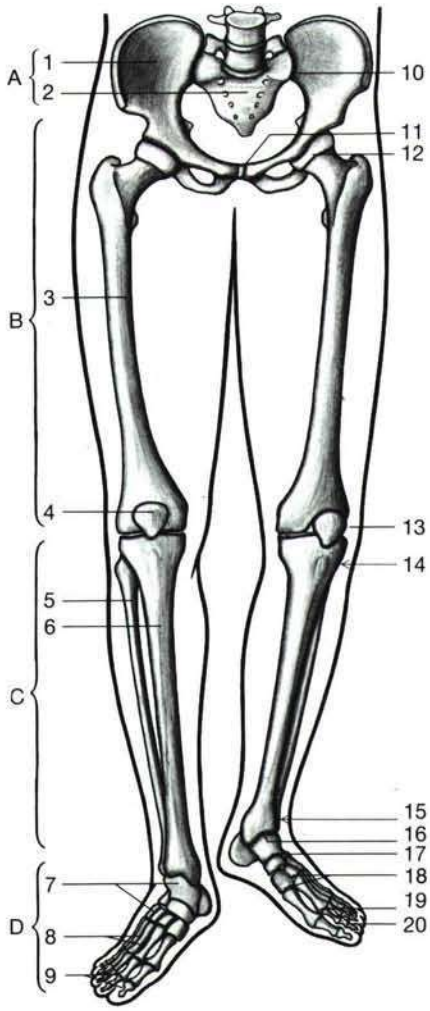

Organization of pelvic girdle and lower limb.

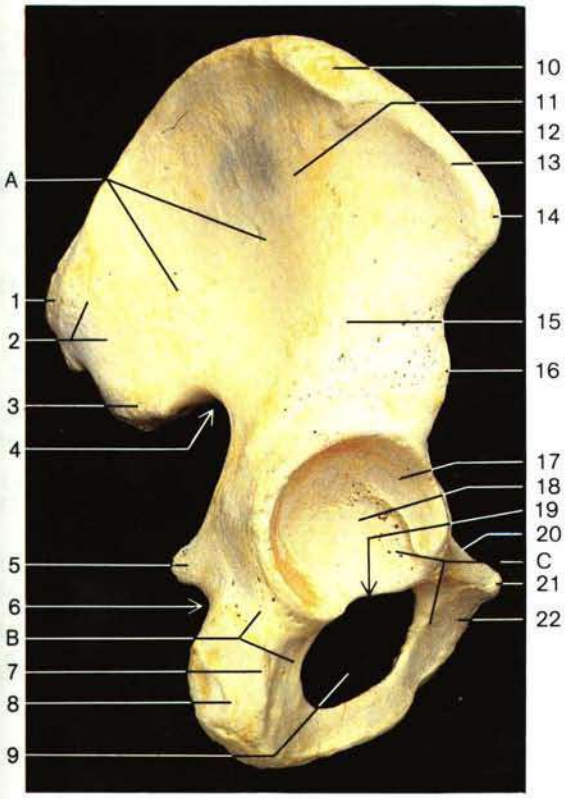

Right hip bone (lateral aspect).

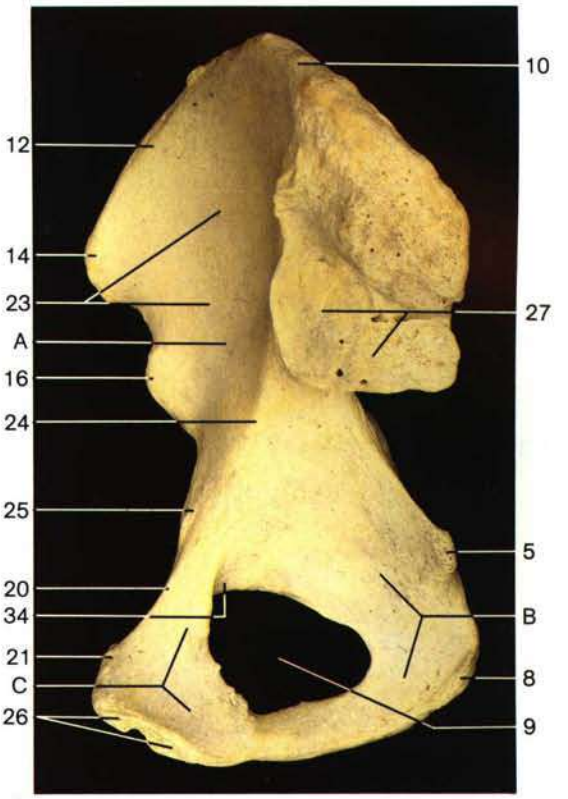

Right hip bone (medial aspect).

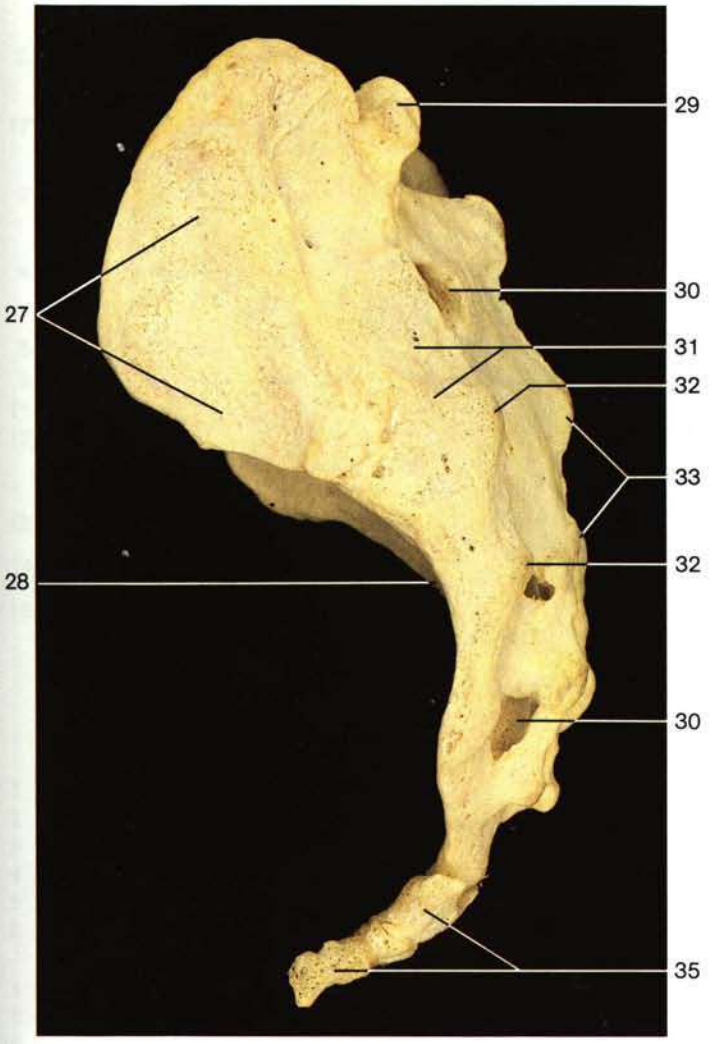

Sacrum and coccyx (lateral aspect).

A Ilium
B Ischium
C Pubis

1 Posterior superior iliac spine
2 Posterior gluteal line
3 Posterior inferior iliac spine
4 Greater sciatic notch
5 Ischial spine
6 Lesser sciatic notch
7 Body of ischium
8 Ischial tuberosity
9 Obturator foramen
10 Iliac crest
11 Anterior gluteal line
12 Internal lip of iliac crest
13 External lip of iliac crest
14 Anterior superior iliac spine
15 Inferior gluteal line
16 Anterior inferior iliac spine
17 Lunate surface of acetabulum
18 Acetabular fossa
19 Acetabular notch
20 Pecten pubis
21 Pubic tubercle
22 Body of pubis
23 Iliac fossa
24 Arcuate line
25 Iliopubic eminence
26 Articular surface of pubis
27 Auricular surface of sacrum
28 Pelvic surface of sacrum
29 Superior articular process of sacrum
30 Dorsal sacral foramina
31 Sacral tuberosity
32 Lateral sacral crest
33 Median sacral crest
34 Obturator groove
35 Coccyx

Sacrum (posterior aspect).

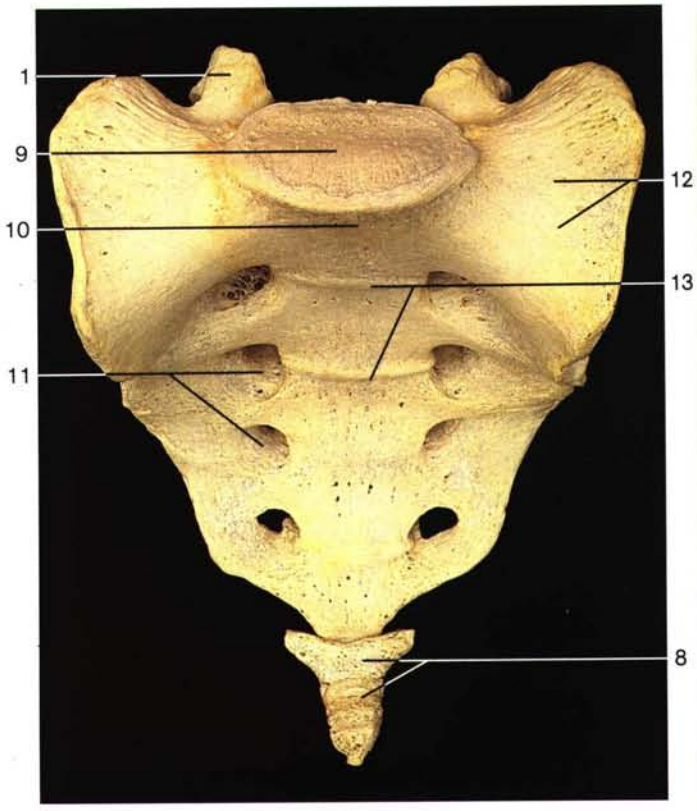

Sacrum (anterior aspect).

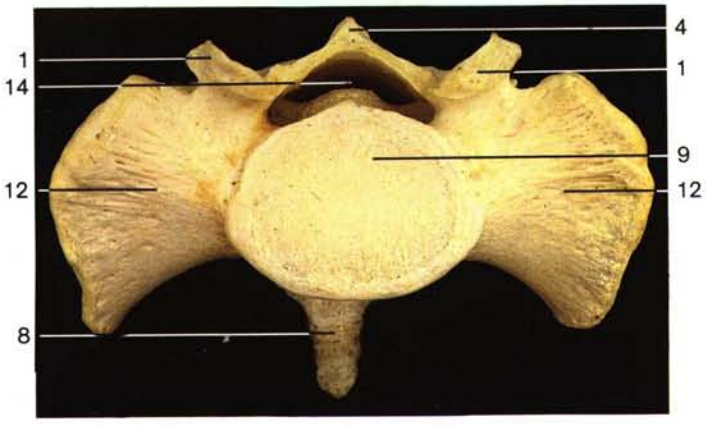

Sacrum (superior aspect).

1 Superior articular process of sacrum
2 Dorsal sacral foramina
3 Sacral hiatus
4 Median sacral crest
5 Lateral sacral crest
6 Sacral tuberosity
7 Intermediate sacral crest
8 Coccyx
9 Base of sacrum
10 Sacral promontory
11 Anterior sacral foramina
12 Lateral part of sacrum (ala)
13 Transverse line of sacrum
14 Sacral canal
15 Linea terminalis
16 True conjugate
17 Diagonal conjugate
18 Transverse diameter
19 Oblique diameter
20 Inferior pelvic aperture or outlet

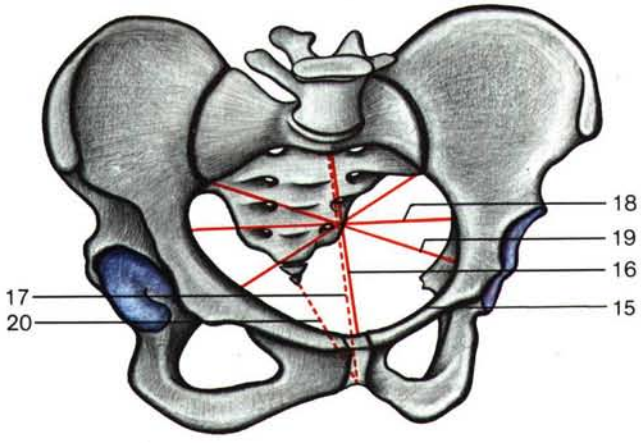

Diameters of pelvis (oblique superior aspect). (Schematic drawing.)

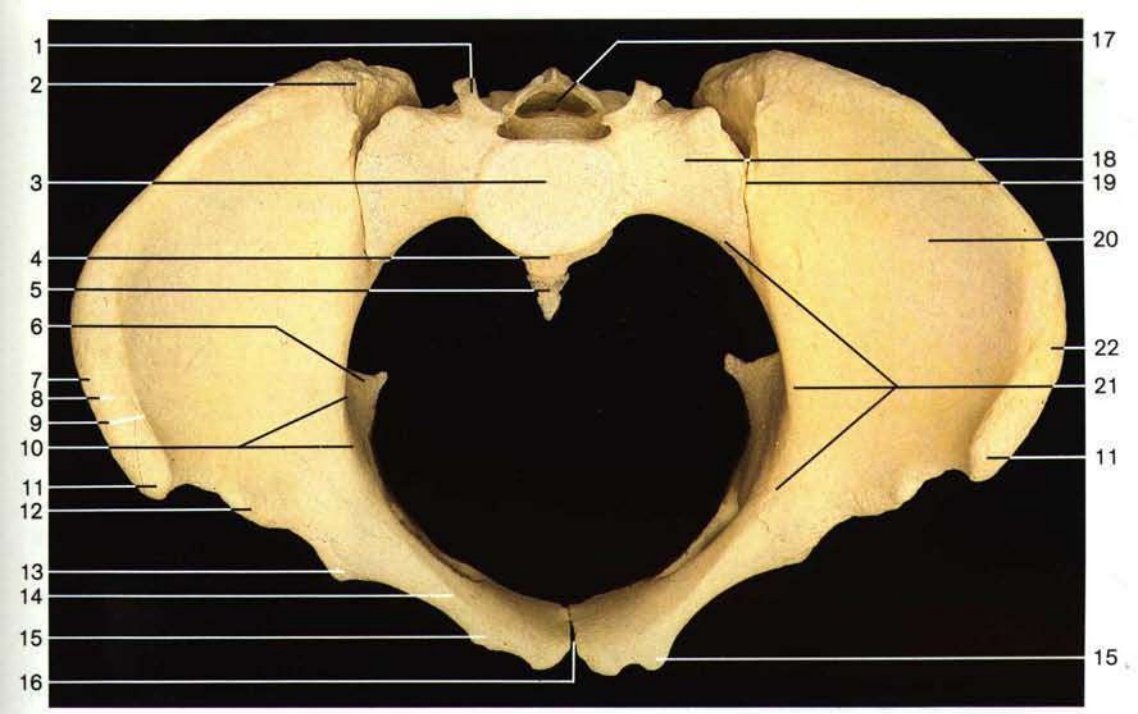

Female pelvis (superior aspect). Note the differences between the male and the female pelvis, predominantly in the form and dimensions of the sacrum, the superior and inferior apertures and the alae of the ilium.

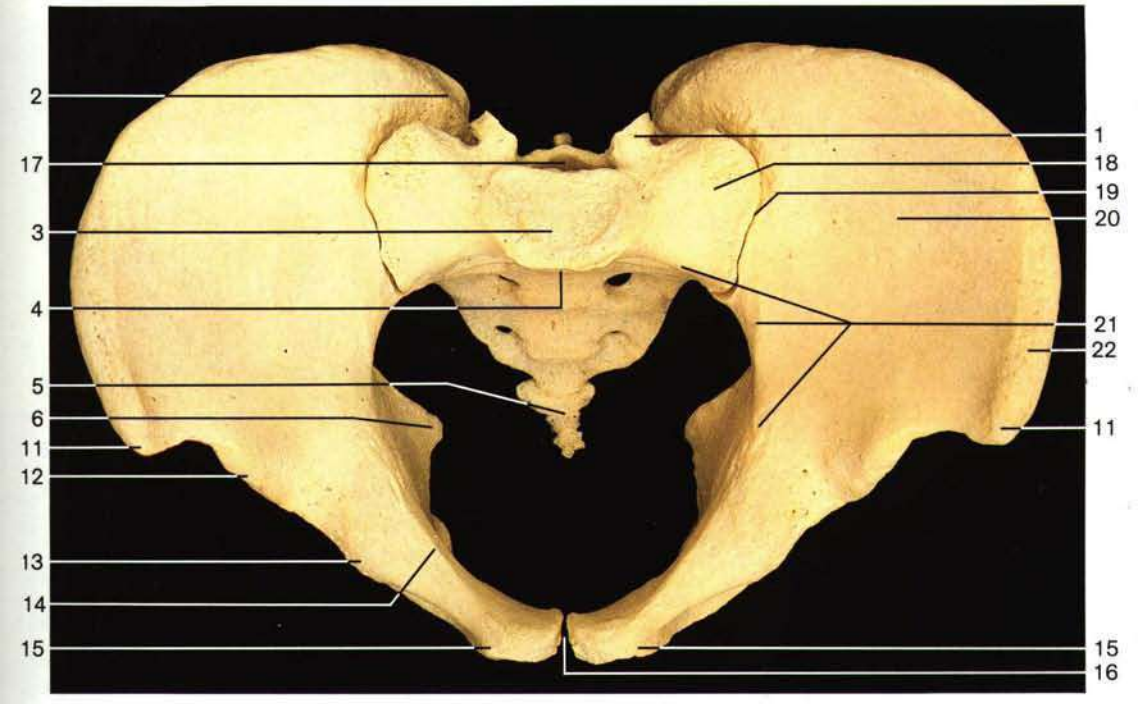

Male pelvis (superior aspect). Compare with the female pelvis (depicted above).

1 Superior articular process of sacrum
2 Posterior superior iliac spine
3 Base of sacrum
4 Sacral promontory
5 Coccyx
6 Ischial spine
7 External lip } of iliac crest
8 Intermediate line } of iliac crest
9 Internal lip } of iliac crest
10 Arcuate line
11 Anterior superior iliac spine
12 Anterior inferior iliac spine
13 Iliopubic eminence
14 Pecten pubis
15 Pubic tubercle
16 Pubic symphysis
17 Sacral canal
18 Ala of sacrum
19 Position of sacroiliac joint
20 Iliac fossa
21 Linea terminalis
22 Iliac crest

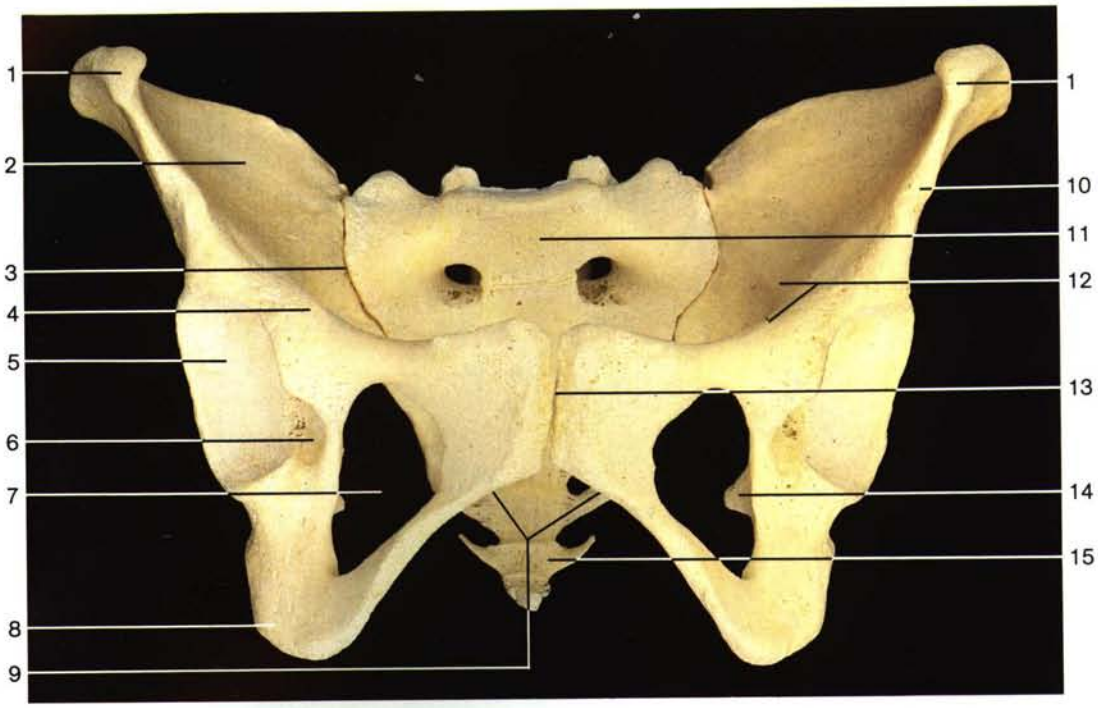

Female pelvis (anterior aspect). Note the differences between the form and dimensions of the male and the female pelvis. The female pubic arch is wider than the male. The obturator foramen in the female pelvis is triangular, while that in the male pelvis is ovoid.

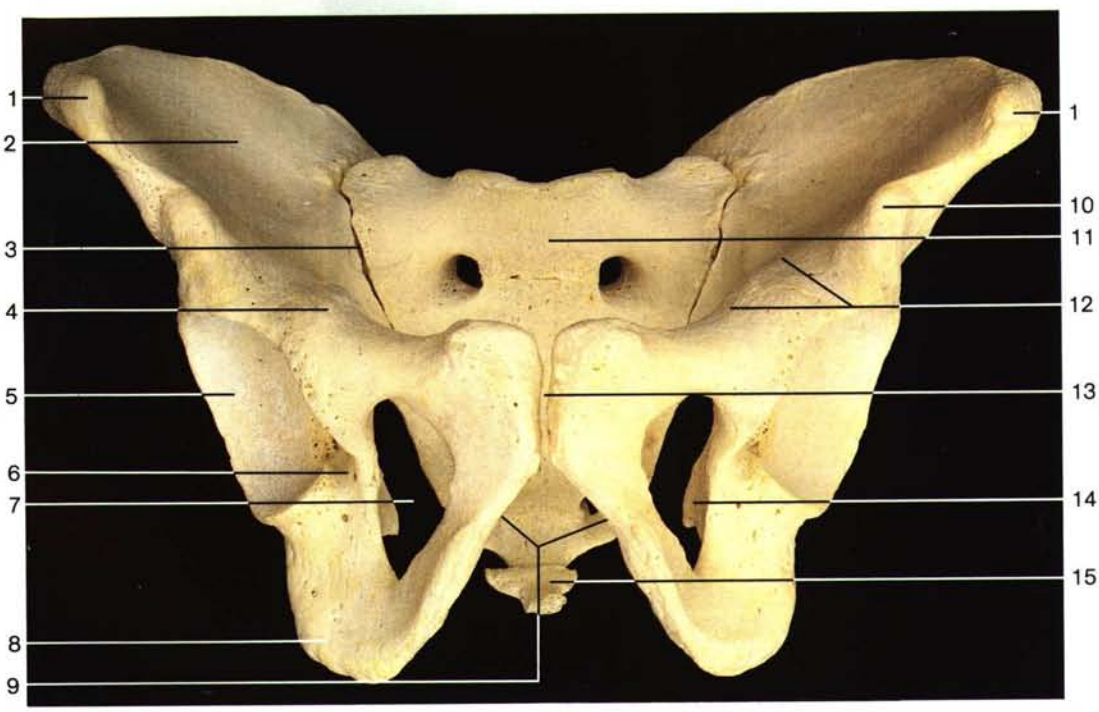

Male pelvis (anterior aspect). Compare with foregoing figure.

1 Anterior superior iliac spine
2 Iliac fossa
3 Position of sacroiliac joint
4 Iliopubic eminence
5 Lunate surface of acetabulum
6 Acetabular notch
7 Obturator foramen
8 Ischial tuberosity
9 Pubic arch
10 Anterior inferior iliac spine
11 Sacrum
12 Linea terminalis (at margin of superior aperture)
13 Pubic symphysis
14 Ischial spine
15 Coccyx

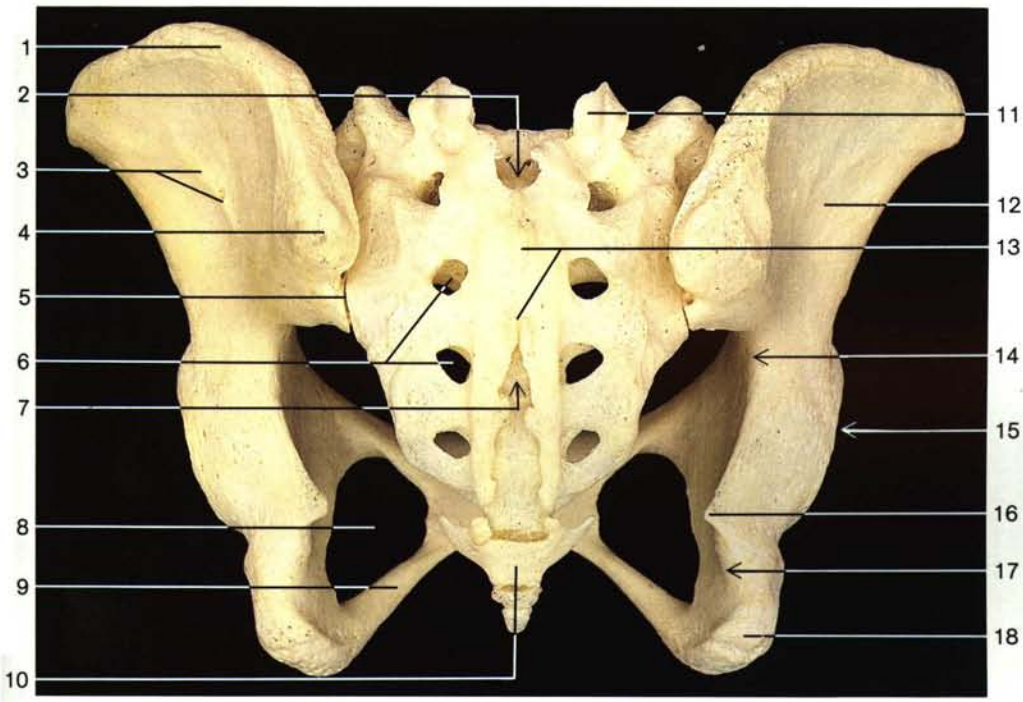

Female pelvis (posterior inferior aspect). Note the differences between the female and male pelvis, especially with respect to the inferior aperture, the shape of the sacrum, the two sciatic notches and the pubic arch.

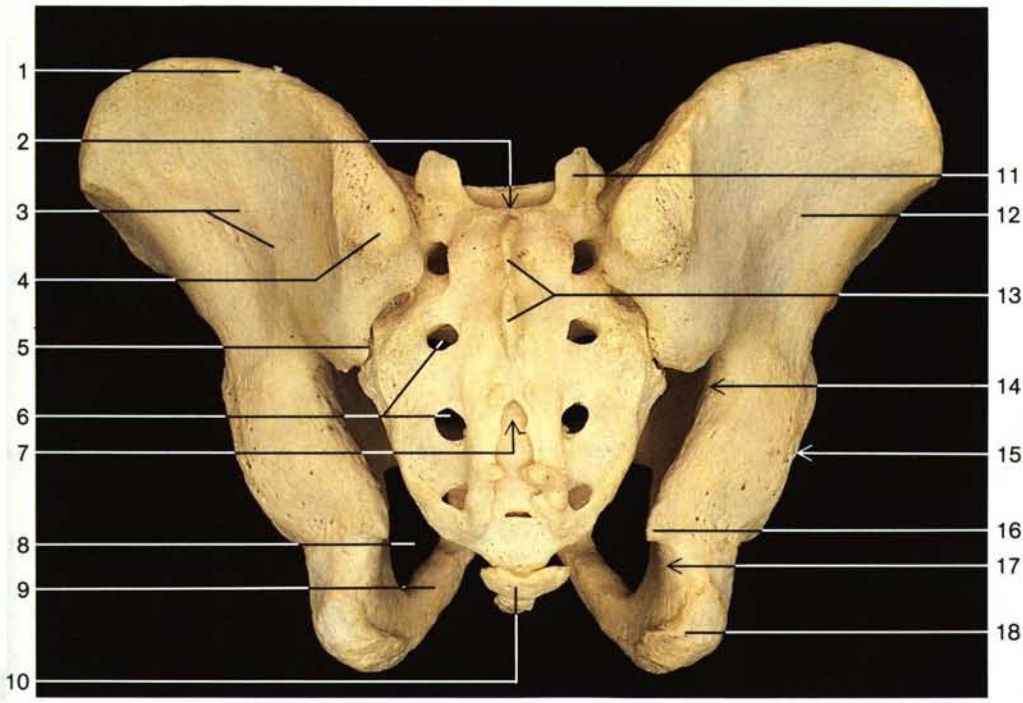

Male pelvis (posterior inferior aspect). Compare with the female pelvis (depicted above).

1 Iliac crest
2 Sacral canal
3 Posterior gluteal line
4 Posterior superior iliac spine
5 Position of sacroiliac joint
6 Dorsal sacral foramina
7 Sacral hiatus
8 Obturator foramen
9 Ramus of ischium
10 Coccyx
11 Superior articular process of sacrum
12 Gluteal surface of ilium
13 Median sacral crest
14 Greater sciatic notch
15 Position of acetabulum
16 Ischial spine
17 Lesser sciatic notch
18 Ischial tuberosity

Bones of right hip joint (anterior aspect).

1 Iliac crest
2 Lateral part of sacrum (ala)
3 Position of sacroiliac joint
4 Anterior superior iliac spine
5 Linea terminalis
6 Iliopubic eminence
7 Bony margin of acetabulum
8 Head of femur
9 Greater trochanter
10 Neck of femur
11 Intertrochanteric line
12 Shaft of femur
13 5th lumbar vertebra
14 Imitation intervertebral disc between 5th lumbar vertebra and sacrum
15 Sacral promontory
16 Anterior sacral foramina
17 Pubic tubercle
18 Obturator foramen
19 Ramus of ischium
20 Lesser trochanter
21 Dorsal sacral foramina
22 Greater sciatic notch
23 Ischial spine
24 Pubic symphysis
25 Pubis
26 Ischial tuberosity
27 Intertrochanteric crest
28 Symphysial surface

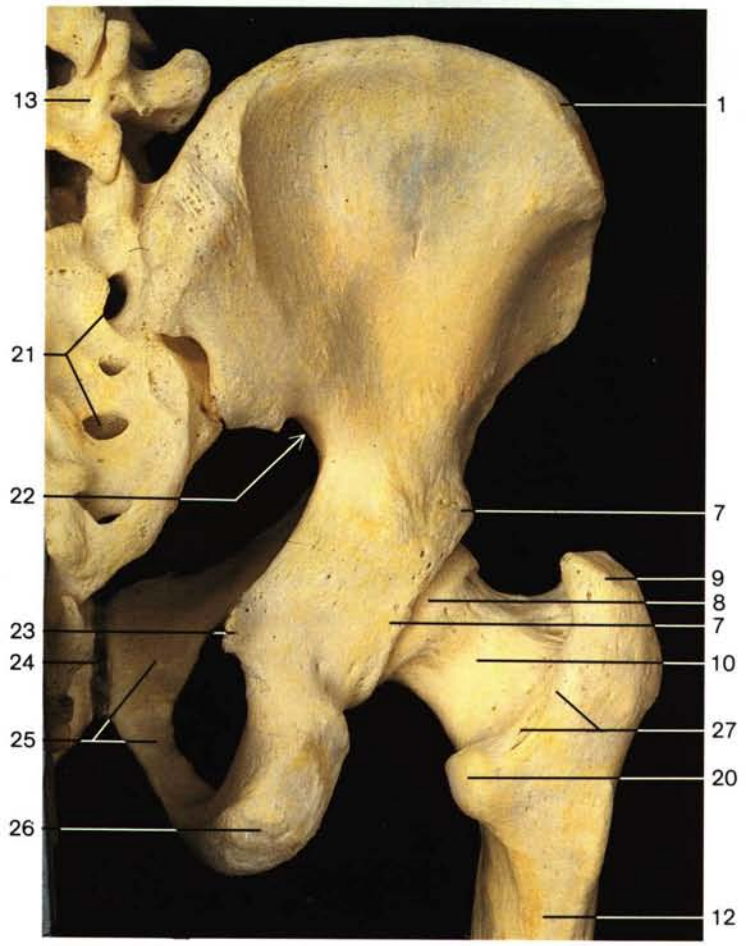

Bones of right hip joint (posterior aspect).

Diameters of the pelvis

A True conjugate (11–11.5 cm) (Conjugata vera)
B Diagonal conjugate (12.5–13 cm)
C Largest diameter of pelvis
D Inferior pelvic aperture
E Pelvic inclination (60°)

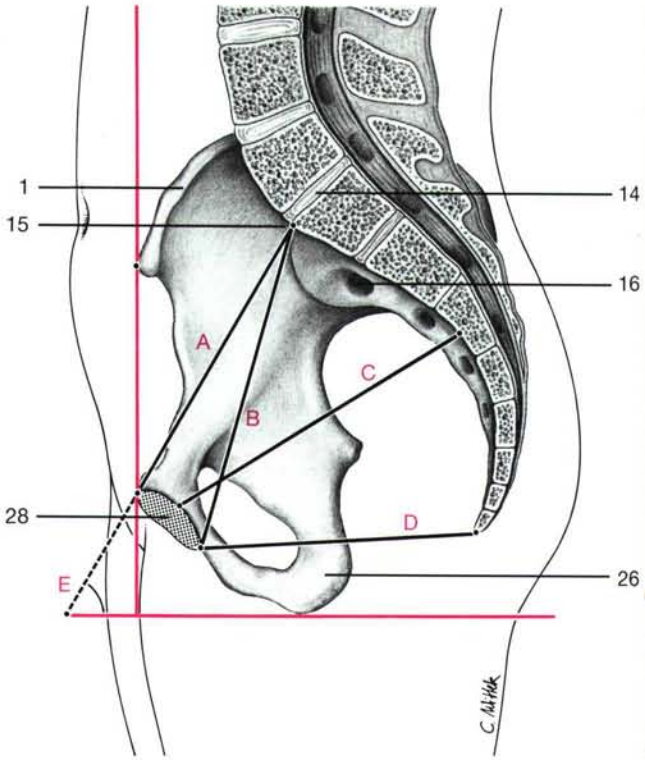

Inclination and diameters of the female pelvis, right half (medial aspect).

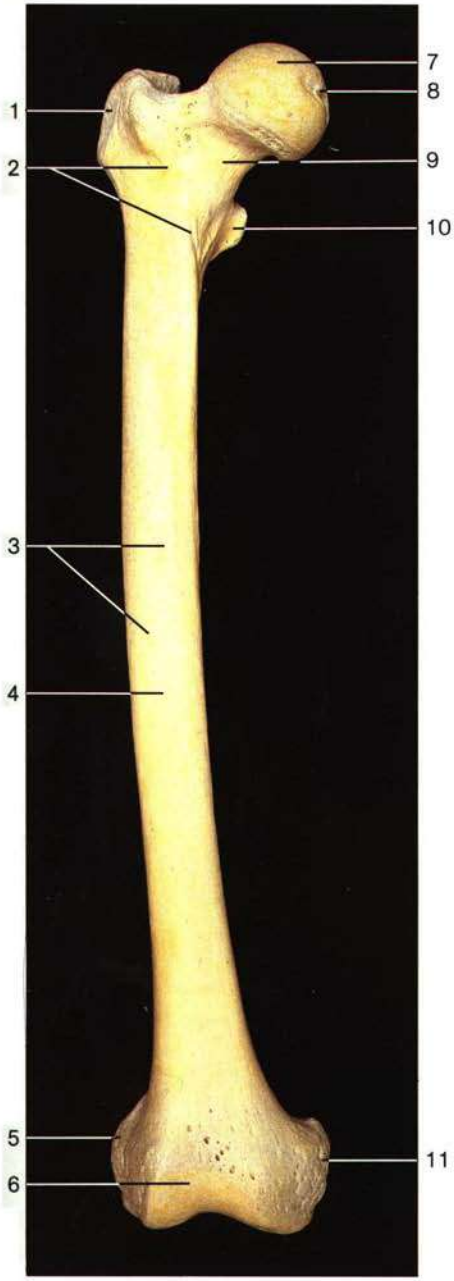

Right femur (anterior aspect).

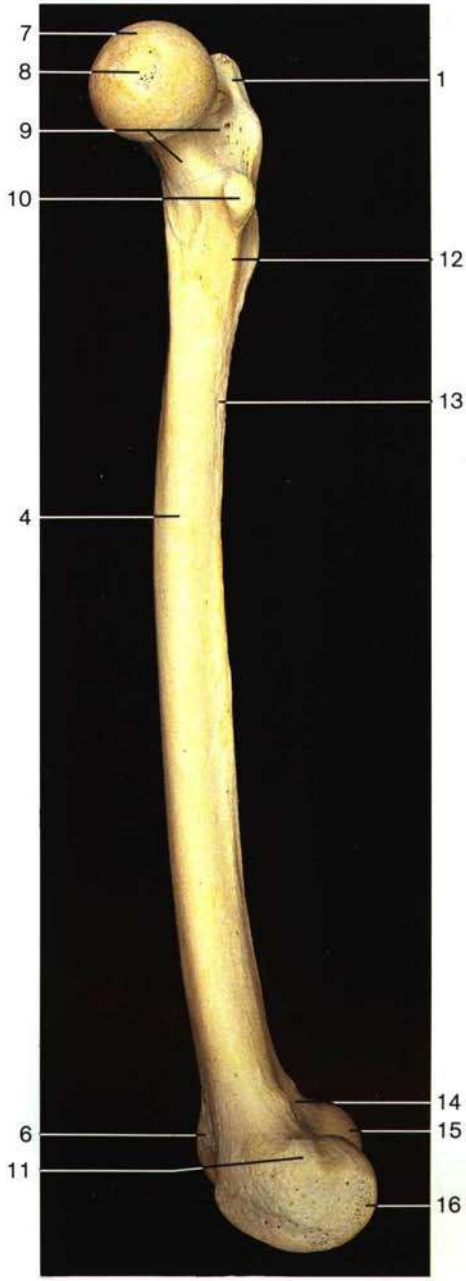

Right femur (medial aspect).

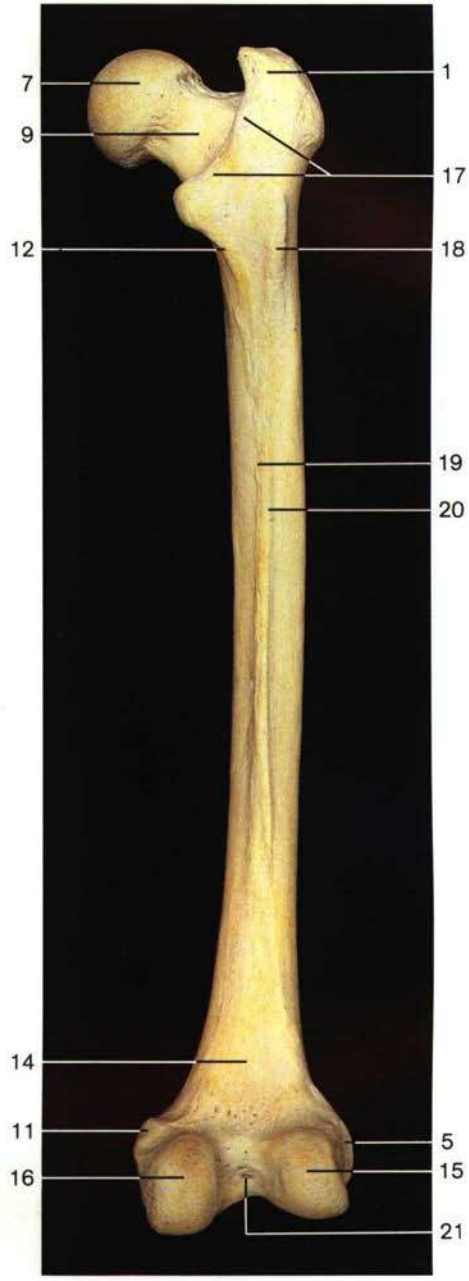

Right femur (posterior aspect).

1 Greater trochanter
2 Intertrochanteric line
3 Nutrient foramina
4 Shaft of femur (diaphysis)
5 Lateral epicondyle
6 Patellar surface
7 Head
8 Fovea of head
9 Neck
10 Lesser trochanter
11 Medial epicondyle
12 Pectineal line
13 Linea aspera
14 Popliteal surface
15 Lateral condyle
16 Medial condyle
17 Intertrochanteric crest
18 Third trochanter
19 Medial lip of linea aspera
20 Lateral lip of linea aspera
21 Intercondylar fossa

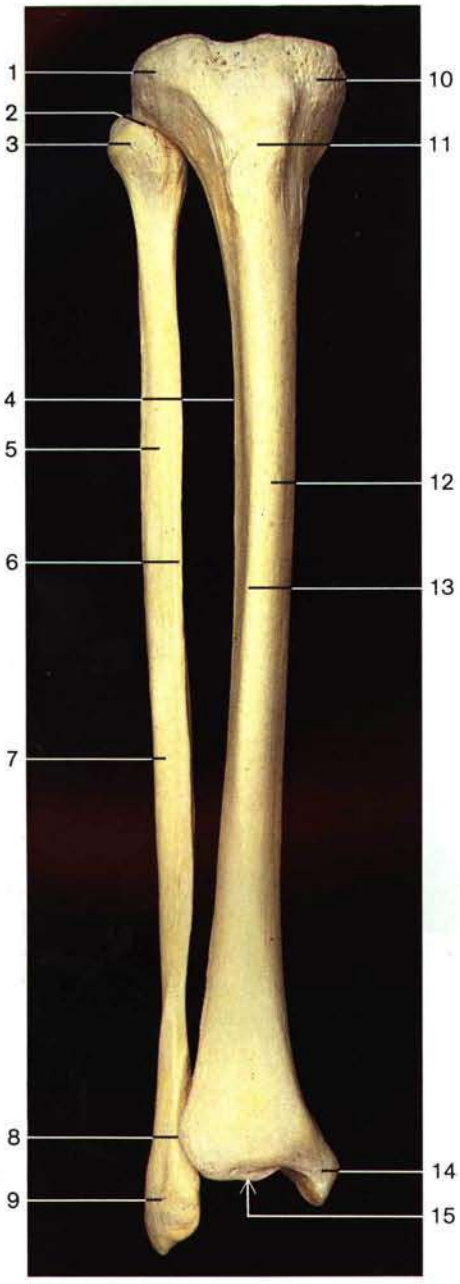

Bones of leg, right tibia and fibula (anterior aspect).

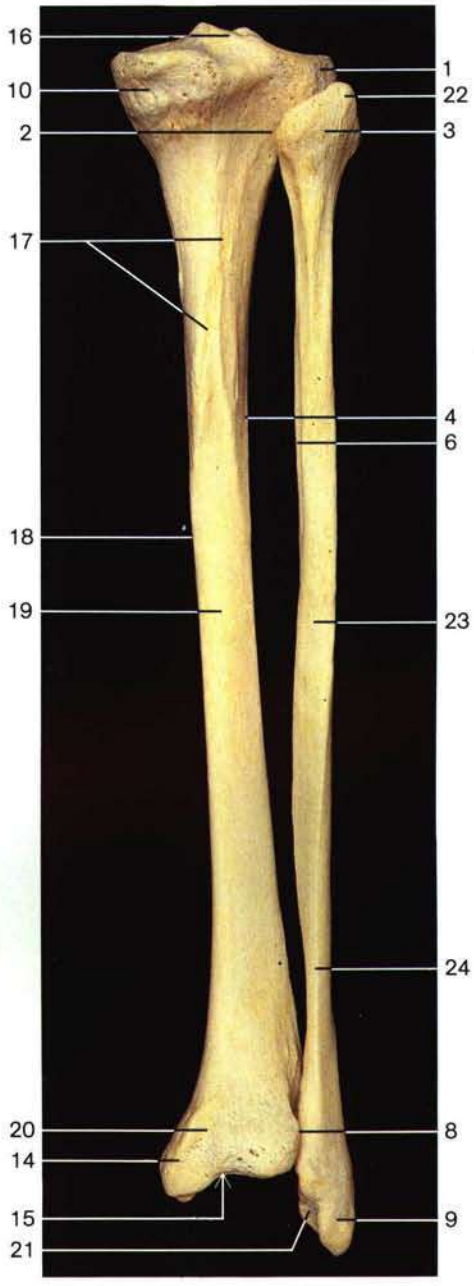

Bones of leg, right tibia and fibula (posterior aspect).

1 Lateral condyle of tibia
2 Position of tibiofibular joint
3 Head of fibula
4 Interosseous border of tibia
5 Shaft of fibula
6 Interosseous border of fibula
7 Lateral surface of fibula
8 Position of tibiofibular joint
9 Lateral malleolus
10 Medial condyle of tibia
11 Tuberosity of tibia
12 Shaft of tibia (diaphysis)
13 Anterior margin of tibia
14 Medial malleolus
15 Inferior articular surface of tibia
16 Intercondylar eminence
17 Soleal line
18 Medial border of tibia
19 Posterior surface of tibia
20 Malleolar sulcus of tibia
21 Malleolar articular surface of fibula
22 Apex of head of fibula
23 Posterior surface of fibula
24 Posterior border of fibula
25 Medial intercondylar tubercle
26 Posterior intercondylar area
27 Anterior intercondylar area
28 Lateral intercondylar tubercle

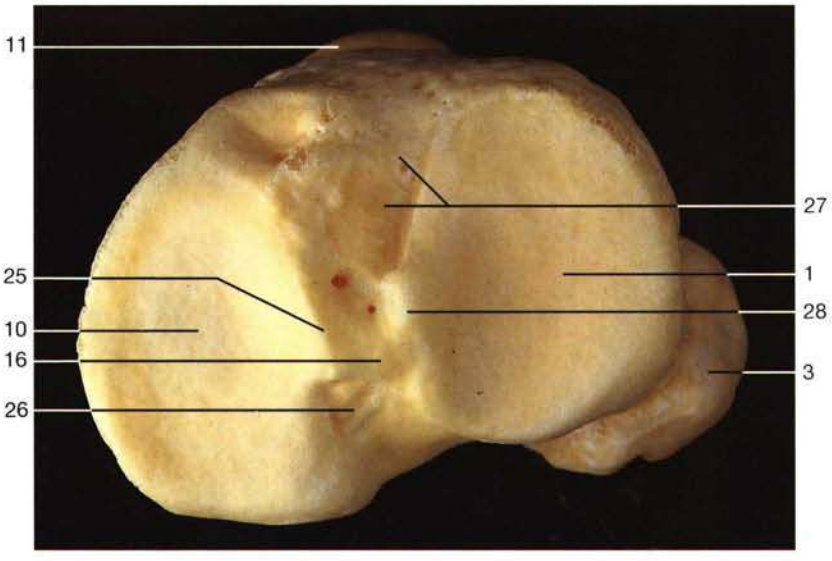

Upper end of right tibia with fibula (from above), anterior margin of tibia above. Superior articular surface of tibia.

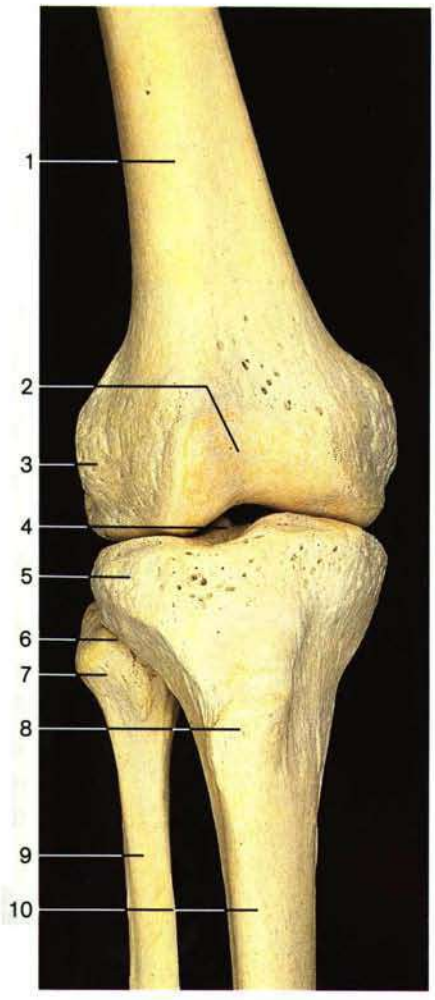

Bones of right knee joint (anterior aspect).

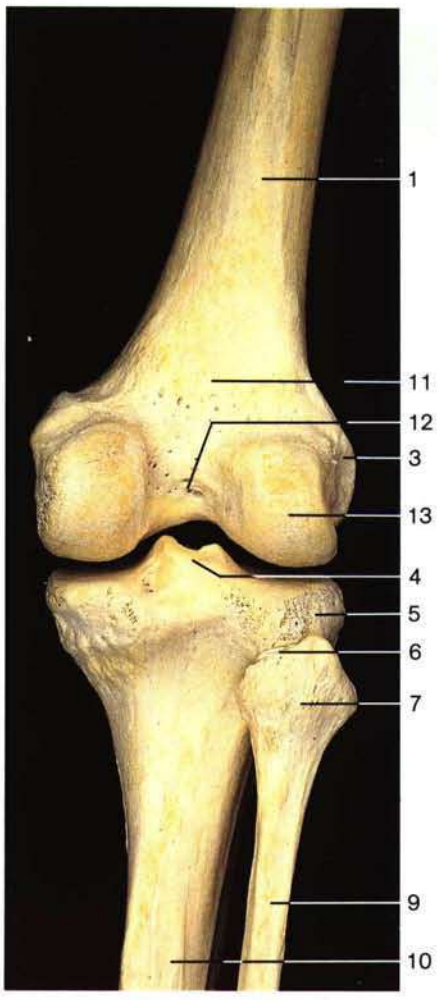

Bones of right knee joint (posterior aspect).

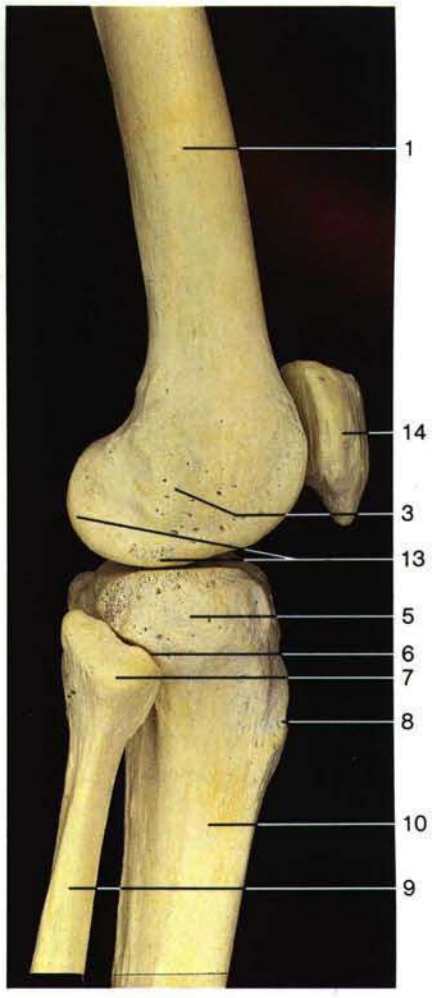

Bones of right knee joint (lateral aspect).

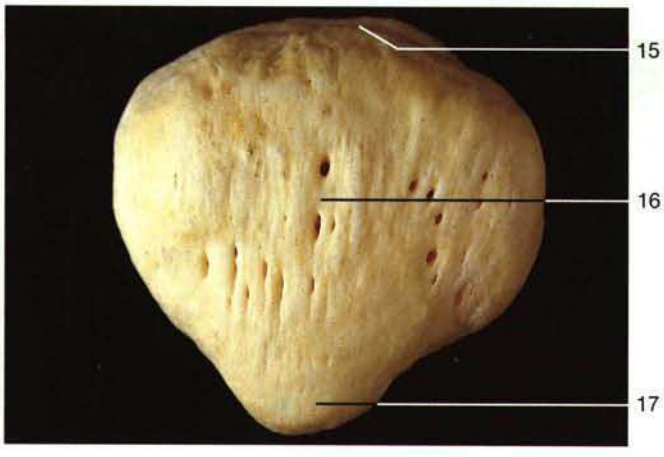

Right patella (anterior aspect).

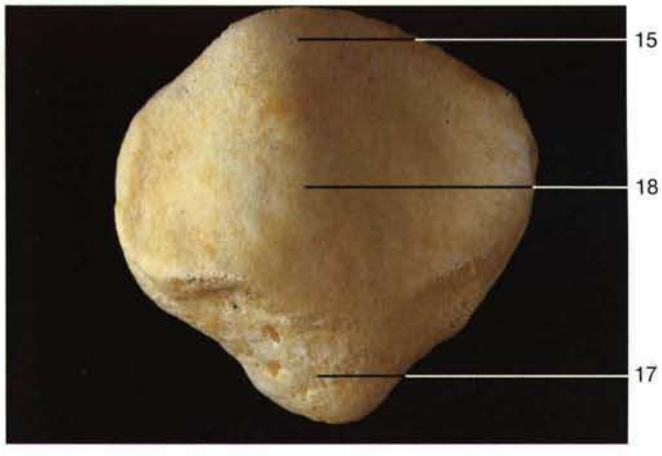

Right patella (posterior aspect).

1 **Femur**
2 Patellar surface of femur
3 Lateral epicondyle of femur
4 Intercondylar eminence of tibia
5 Lateral condyle of tibia
6 Position of tibiofibular joint
7 Head of fibula
8 Tuberosity of **tibia**
9 **Fibula**
10 Shaft of tibia
11 Popliteal surface of femur
12 Intercondylar fossa of femur
13 Lateral condyle of femur
14 **Patella**
15 Base of patella
16 Anterior surface of patella
17 Apex of patella
18 Articular surface of patella

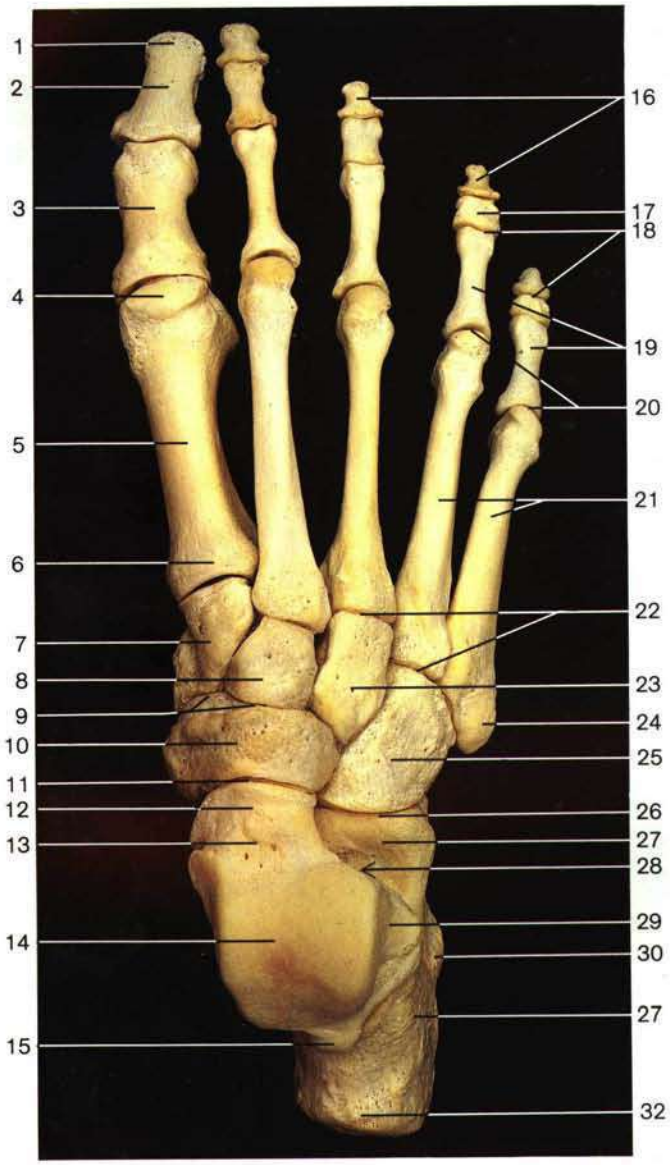

Bones of right foot (dorsal aspect).

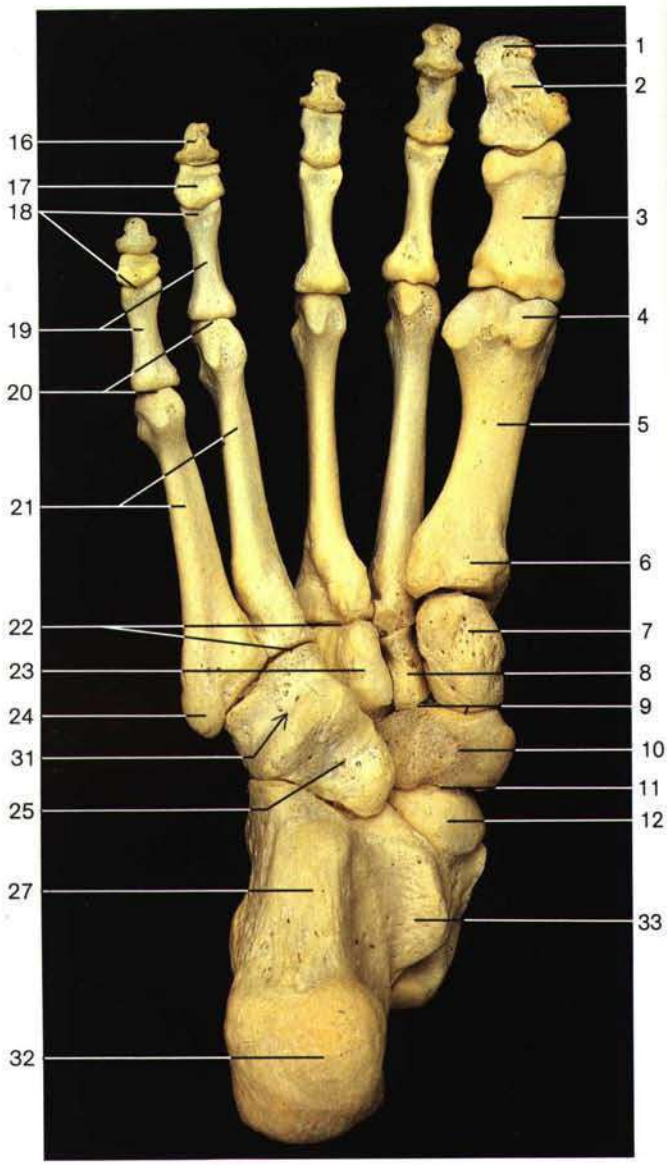

Bones of right foot (plantar aspect).

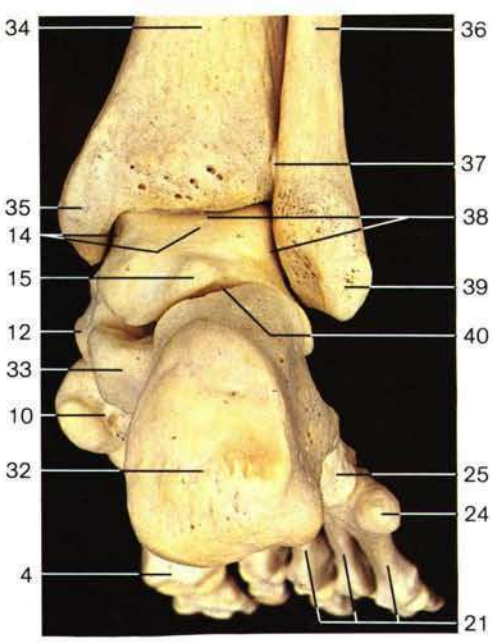

Bones of right foot together with tibia and fibula (posterior aspect).

1 Tuberosity of distal phalanx of great toe
2 Distal phalanx of great toe
3 Proximal phalanx of great toe
4 Head of first metatarsal bone
5 First metatarsal bone
6 Base of first metatarsal bone
7 Medial cuneiform bone
8 Intermediate cuneiform bone
9 Position of **cuneonavicular joint**
10 Navicular bone

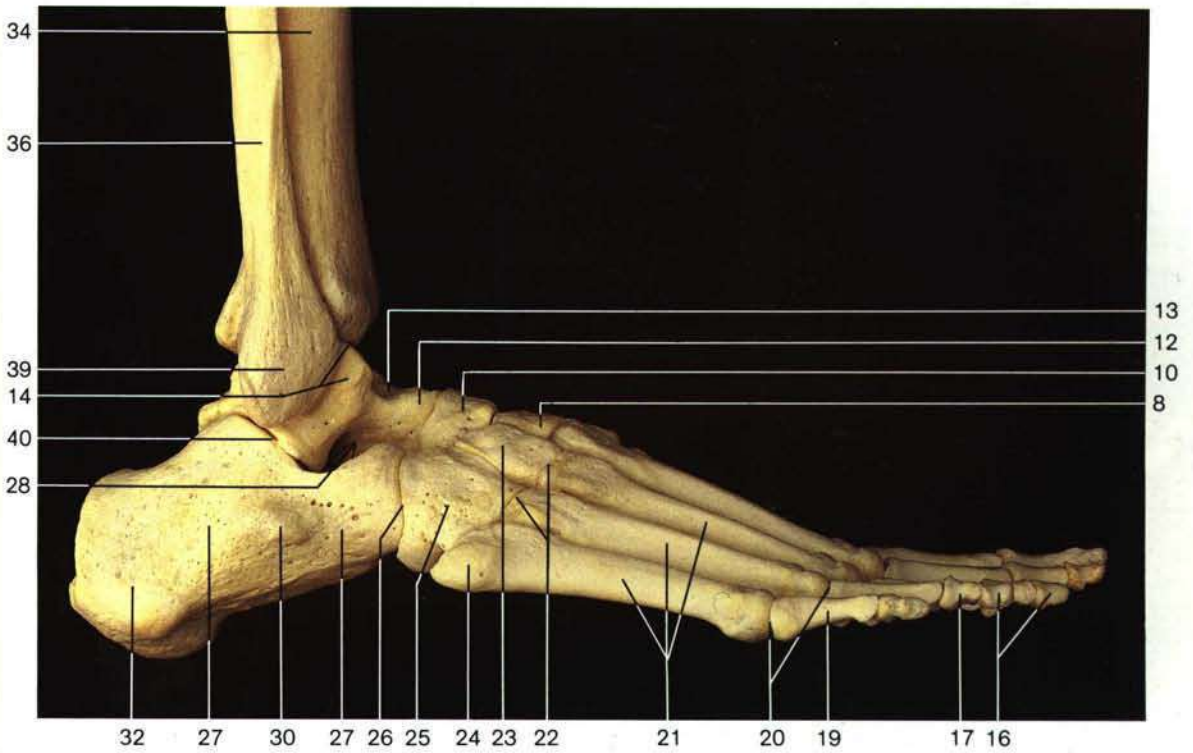

Bones of right foot, tibia and fibula (lateral aspect).

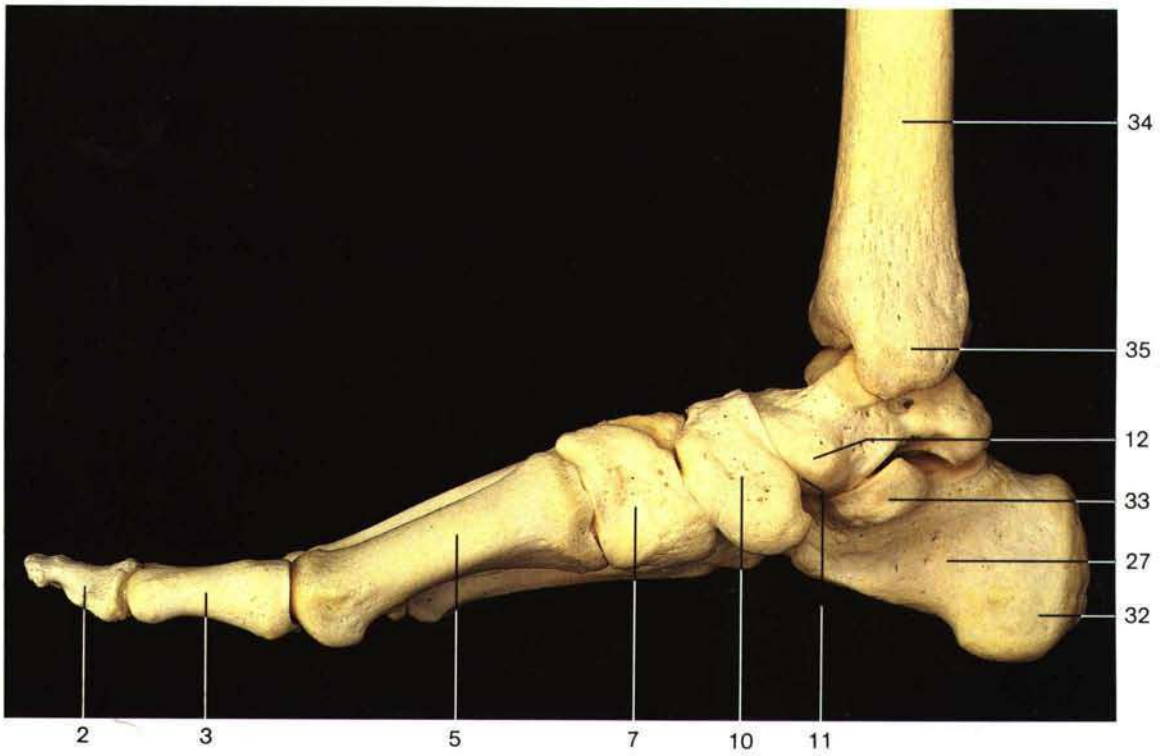

Bones of right foot, tibia and fibula (medial aspect).

11 Position of **talocalcaneonavicular joint**
12 Head of talus
13 Neck of talus
14 Trochlea of talus
15 Posterior talar process
16 Distal phalanges
17 Middle phalanges
18 Position of **interphalangeal joints**
19 Proximal phalanges
20 Position of **metatarsophalangeal joints**
21 Metatarsal bones
22 Position of **tarsometatarsal joints**
23 Lateral cuneiform bone
24 Tuberosity of 5th metatarsal bone
25 Cuboid bone
26 Position of **calcaneocuboid joint**
27 Calcaneus
28 Tarsal sinus
29 Lateral malleolar surface of talus
30 Peroneal trochlea of calcaneus
31 Groove for tendon of peroneus longus
32 Calcaneal tuberosity
33 Sustentaculum tali
34 Tibia
35 Medial malleolus
36 Fibula
37 Position of tibiofibular syndesmosis
38 Position of **talocrural joint**
39 Lateral malleolus
40 Position of **subtalar joint**

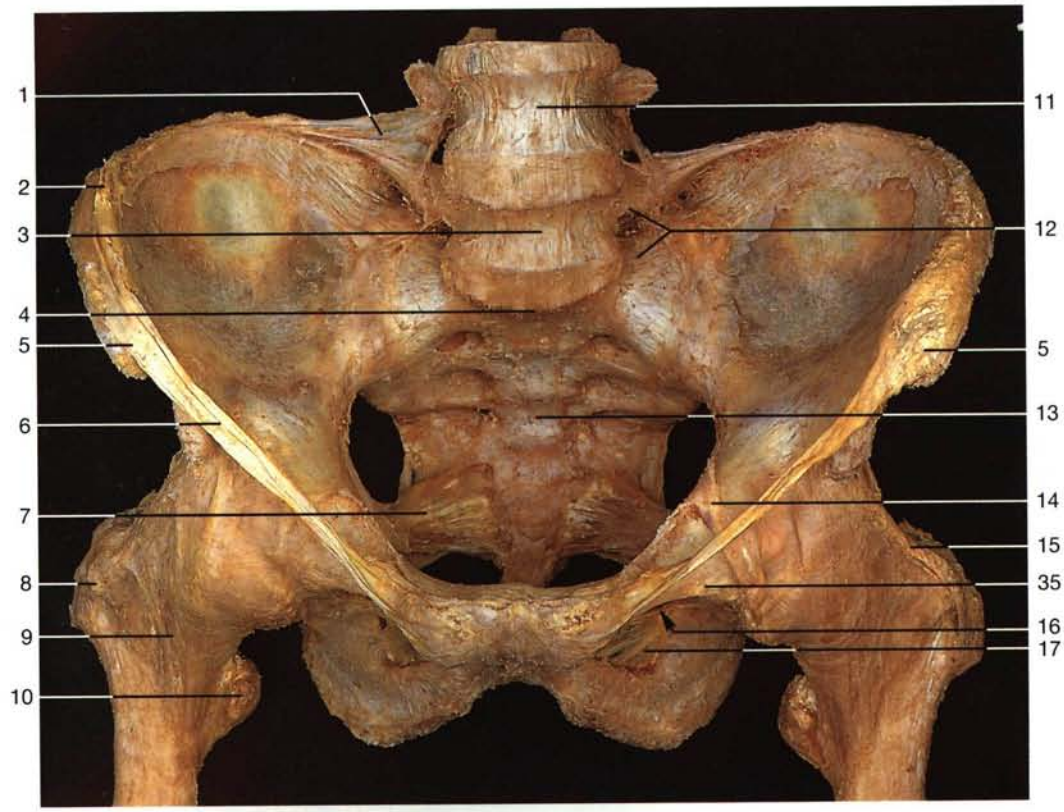

Ligaments of pelvis and hip joint (anterior aspect).

1 **Iliolumbar ligament**
2 Iliac crest
3 Fifth lumbar vertebra
4 Sacral promontory
5 Anterior superior iliac spine
6 Inguinal ligament
7 Sacrospinous ligament
8 Greater trochanter
9 **Iliofemoral ligament** (vertical band)
10 Lesser trochanter
11 Forth lumbar vertebra
12 Iliolumbar and ventral sacroiliac ligaments
13 Sacrum
14 Iliopectineal arch
15 **Iliofemoral ligament** (horizontal band)
16 Obturator canal
17 Obturator membrane
18 **Greater sciatic foramen**
19 Sacrospinous ligament
20 Sacrotuberous ligament
21 **Lesser sciatic foramen**
22 Ischial tuberosity
23 **Ischiofemoral ligament**
24 Intertrochanteric crest
25 Femur
26 Articular capsule of hip joint
27 Dorsal sacroiliac ligaments
28 Coccyx with superficial dorsal sacrococcygeal ligament
29 Head of femur
30 Articular cartilage of head of femur
31 Articular cavity of hip joint
32 Acetabular lip
33 Spongy bone
34 Ligament of head of femur
35 Pubofemoral ligament
36 Zona orbicularis

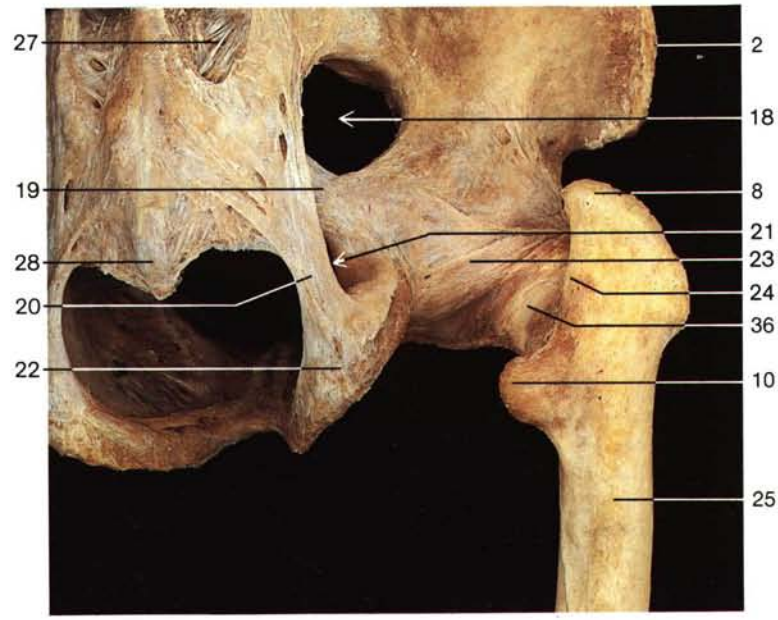

Ligaments of pelvis and hip joint (left anterior aspect).

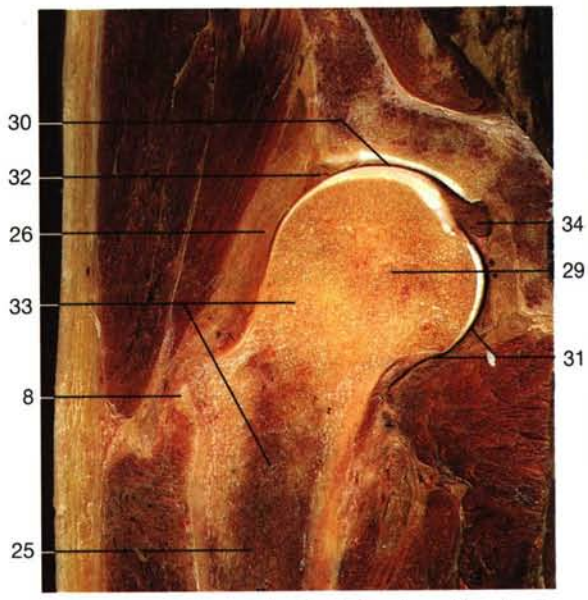

Coronal section of right hip joint (anterior view).

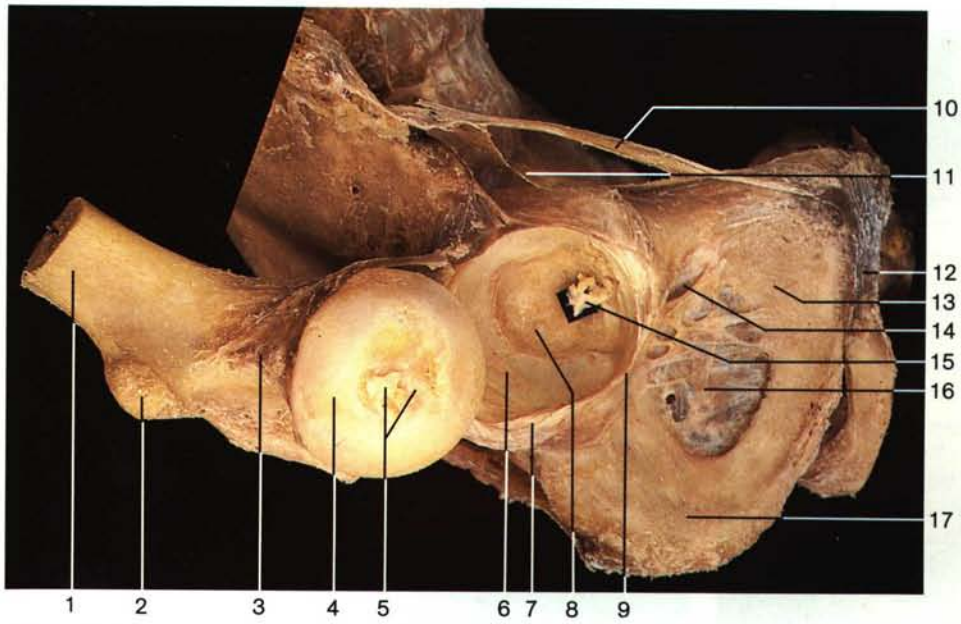

Right hip joint, opened (lateral anterior aspect). The ligament of the head of the femur has been divided and the femur has been posteriorly reflected.

1 Femur
2 Lesser trochanter
3 Neck of femur
4 Head of femur
5 Fovea of head with remnants of ligament of head
6 Lunate surface of acetabulum
7 Acetabular lip
8 Acetabular fossa
9 Transverse acetabular ligament
10 Inguinal ligament
11 Iliopectineal arch
12 Pubic symphysis
13 Pubic bone
14 Obturator canal
15 Ligament of head of femur
16 Obturator membrane
17 Ischium
18 Anterior longitudinal ligament (level of fifth lumbar vertebra)
19 Sacral promontory
20 Iliolumbar ligament
21 Iliac crest
22 Anterior superior iliac spine
23 Iliofemoral ligament (horizontal band)
24 Iliofemoral ligament (vertical band)
25 Greater trochanter
26 Pubofemoral ligament
27 Anterior inferior iliac spine
28 Ventral sacroiliac ligaments
29 **Sacrospinous ligament**
30 **Sacrotuberous ligament**
31 Intertrochanteric line
32 **Ischiofemoral ligament**
33 Zona orbicularis

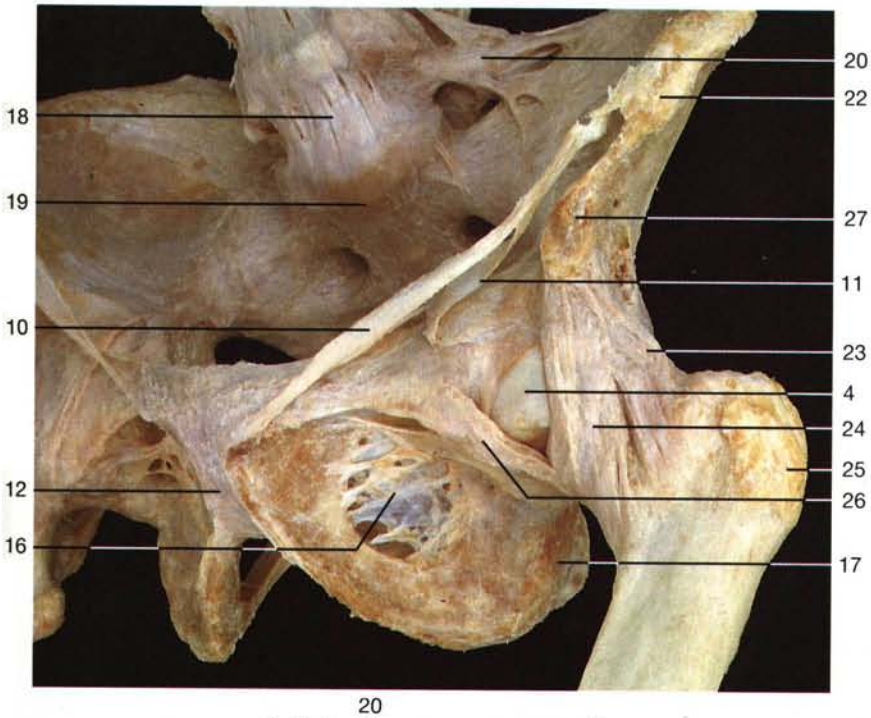

Ligaments of the pelvis and hip joint (antero-lateral aspect).

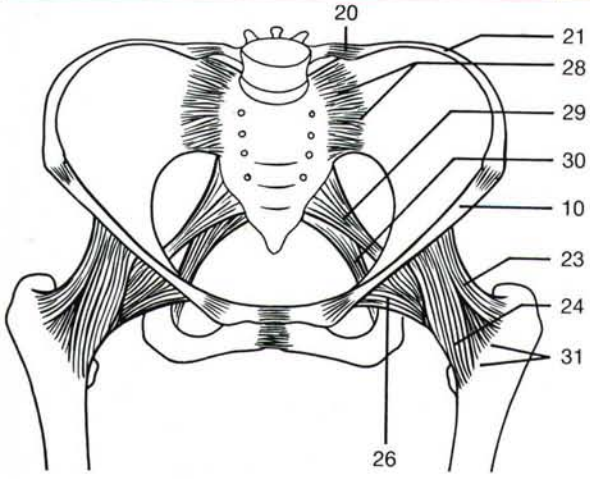

Ligaments of hip joint (anterior aspect). (Schematic drawing.)

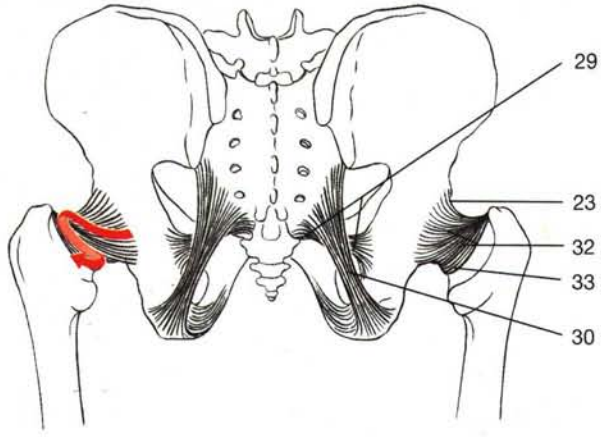

Ligaments of hip joints (posterior aspect). (Schematic drawing.)

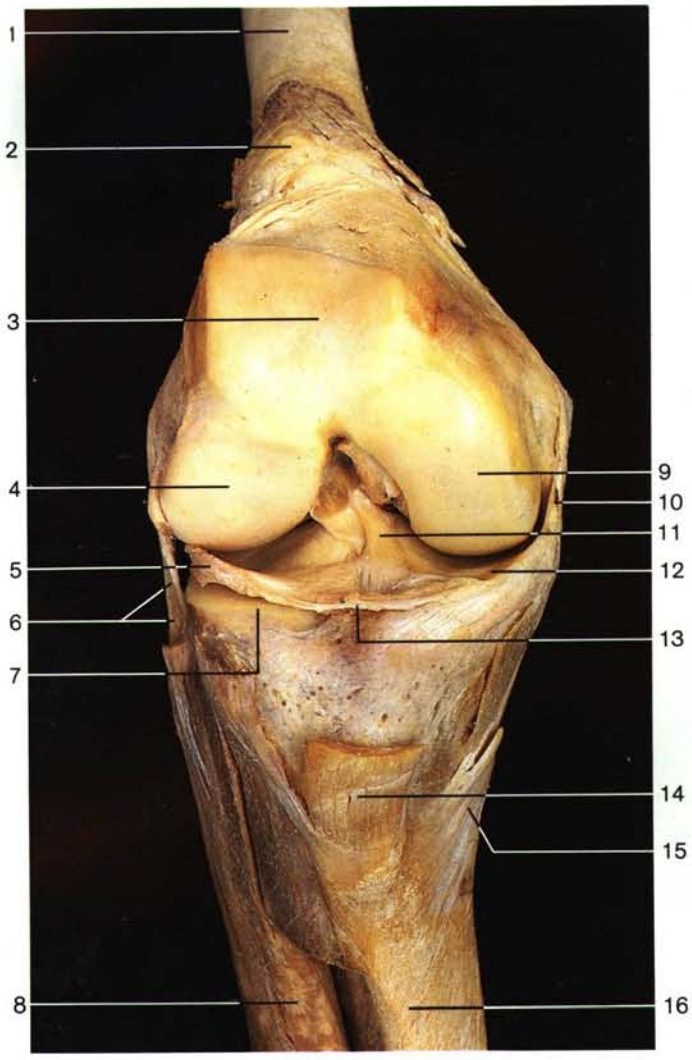

Right knee joint (opened) **with ligaments** (anterior aspect). The patella and articular capsule have been removed and the femur slightly flexed.

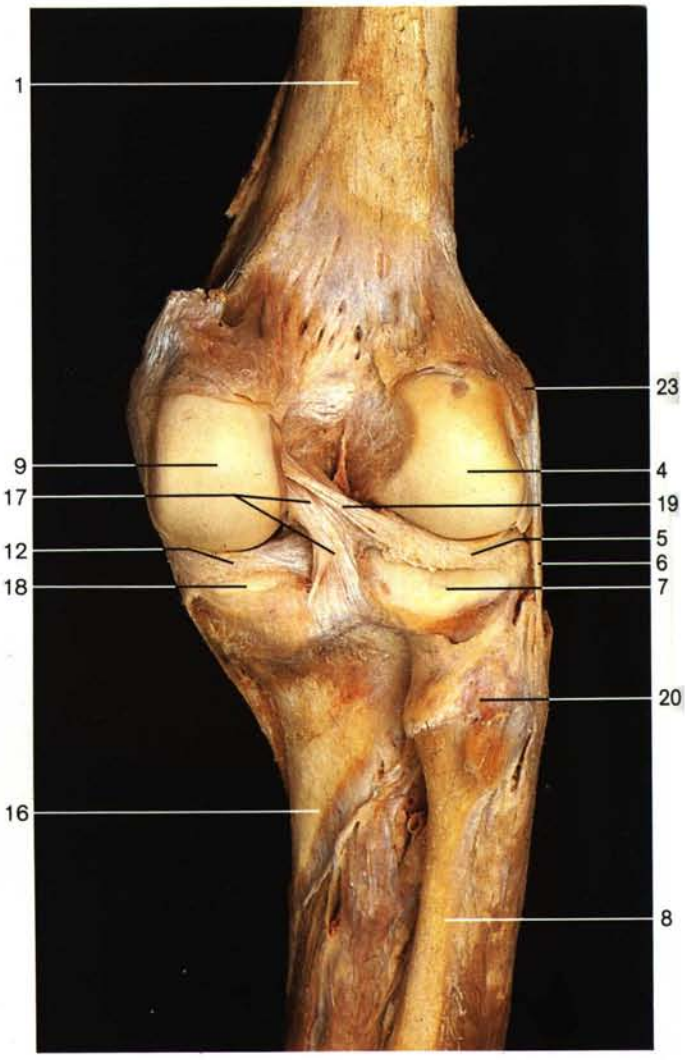

Right knee joint with ligaments (posterior aspect). The joint is extended and the articular capsule has been removed.

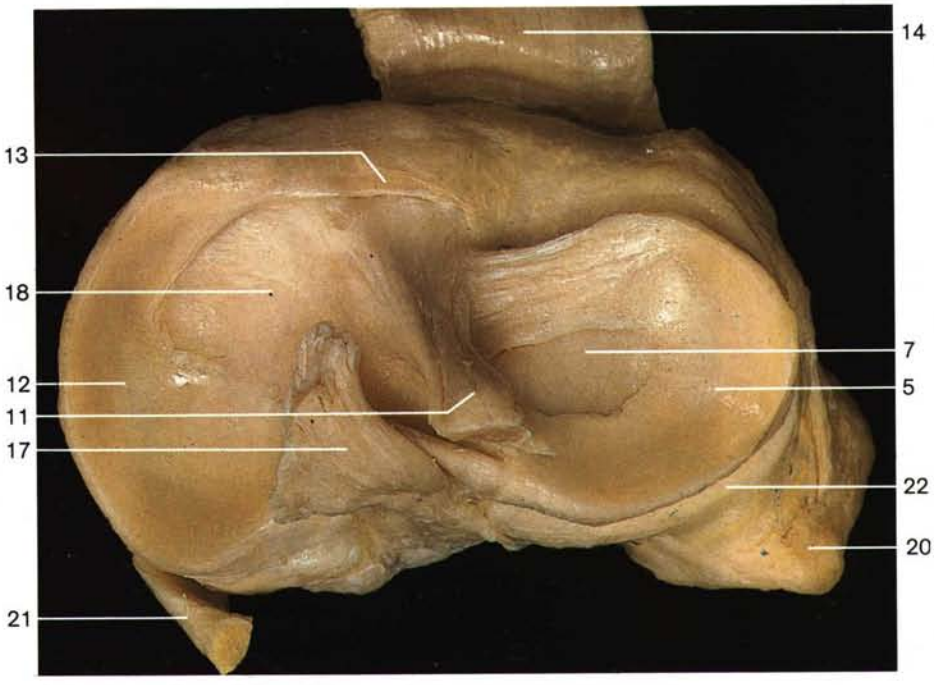

Articular surface of right tibia, menisci, and cruciate ligaments (superior aspect). Anterior margin of tibia above.

1 Femur
2 Articular capsule with **suprapatellar bursa**
3 Patellar surface
4 Lateral condyle of femur
5 **Lateral meniscus of knee joint**
6 Fibular collateral ligament
7 Lateral condyle of tibia (superior articular surface)
8 Fibula
9 Medial condyle of femur
10 Tibial collateral ligament
11 **Anterior cruciate ligament**
12 **Medial meniscus of knee joint**
13 Transverse ligament of knee
14 Patellar ligament
15 Common tendon of sartorius, semitendinosus and gracilis muscles
16 Tibia
17 **Posterior cruciate ligament**
18 Medial condyle of tibia (superior articular surface)
19 Posterior meniscofemoral ligament
20 Head of fibula
21 Tendon of semimembranosus muscle
22 Posterior attachment of articular capsule of knee joint
23 Lateral epicondyle of femur

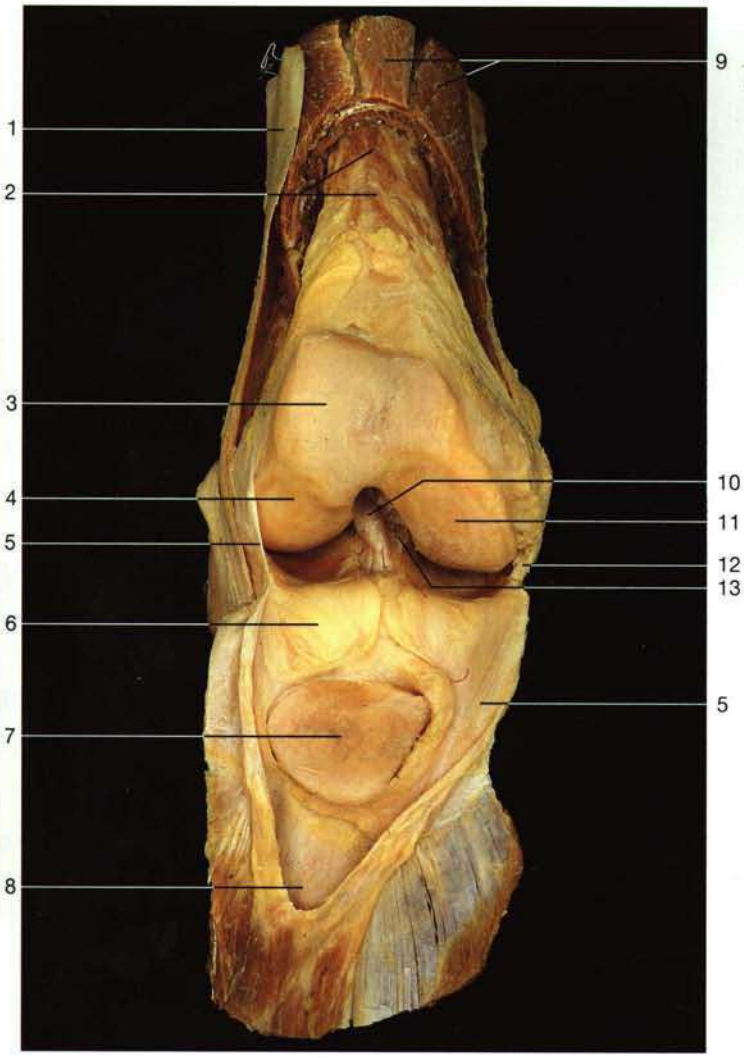

Right knee joint, opened (anterior aspect). Patellar ligament with patella reflected.

1 Iliotibial tract
2 Articular muscle of knee
3 Patellar surface
4 Lateral condyle of femur
5 Articular capsule
6 Infrapatellar fat pad
7 **Patella** (articular surface)
8 **Suprapatellar bursa**
9 Quadriceps muscle of thigh (divided)
10 **Anterior cruciate ligament**
11 Medial condyle of femur
12 Fibular collateral ligament
13 **Posterior cruciate ligament**
14 Medial epicondyle of femur
15 Intercondylar fossa of femur
16 Tibial collateral ligament
17 **Medial meniscus** of knee joint
18 Medial intercondylar tubercle
19 Femur
20 Lateral epicondyle of femur
21 **Lateral meniscus** of knee joint
22 Epiphyseal line of tibia
23 Tibia

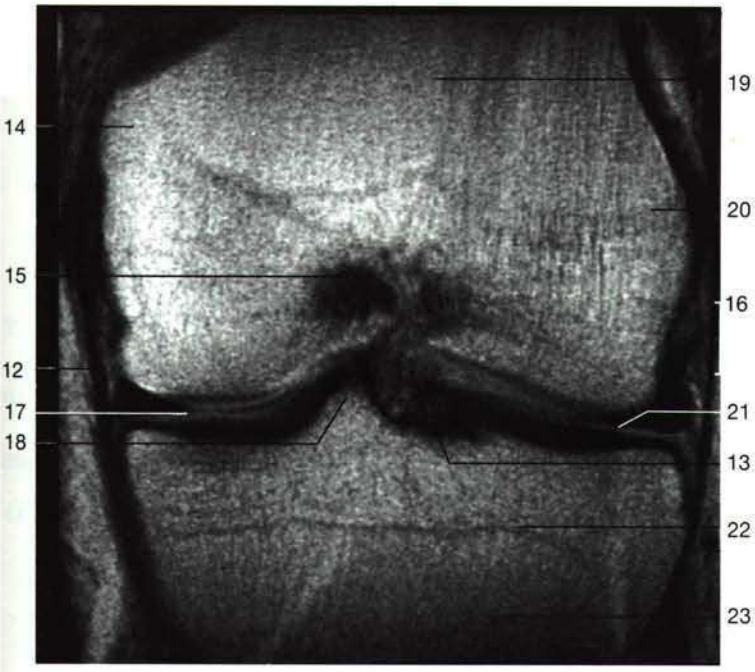

Right knee joint. Frontal section through the central part of the joint (posterior aspect, MR-Scan). (See also p. 10.)

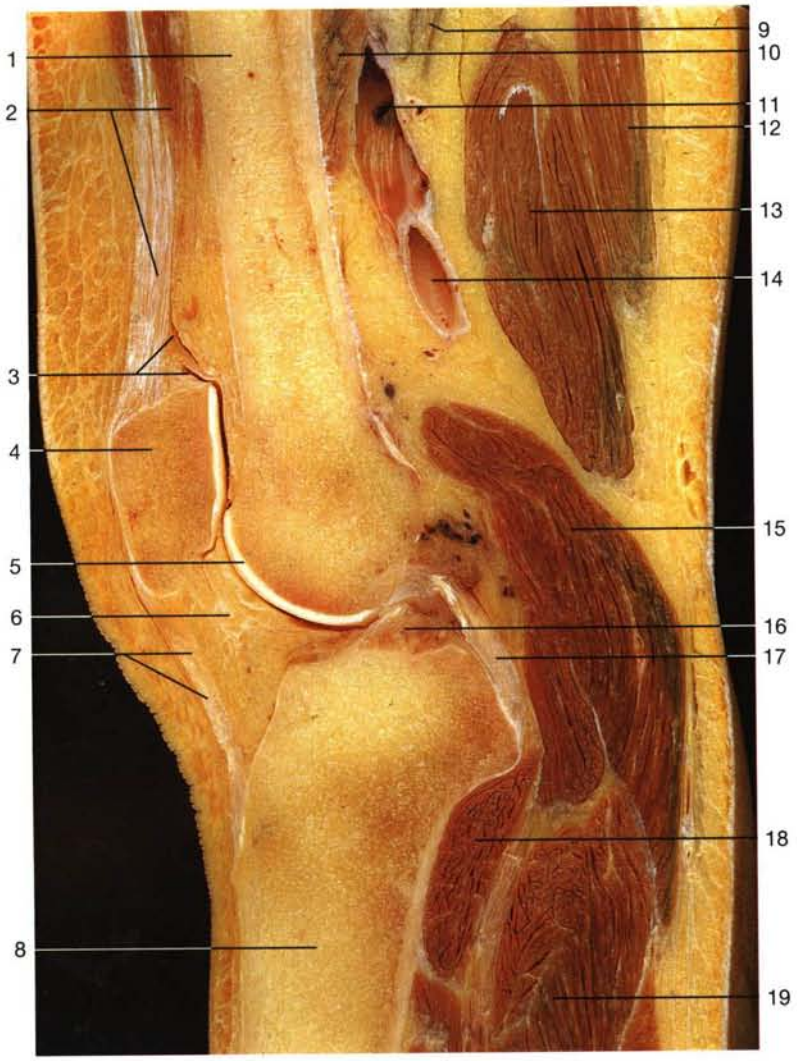

Sagittal section through the knee joint. Anterior surface to the left.

1 **Femur**
2 Quadriceps femoris muscle
3 Suprapatellar bursa and articular cavity
4 **Patella**
5 Patellar surface (articular cartilage)
6 Infrapatellar fat pad
7 Patellar ligament
8 **Tibia**
9 Tibial nerve
10 Adductor magnus muscle
11 Popliteal vein
12 Semitendinosus muscle
13 Semimembranosus muscle
14 **Popliteal artery**
15 Gastrocnemius muscle
16 Anterior cruciate ligament
17 Posterior cruciate ligament
18 Popliteus muscle
19 Soleus muscle
20 Deep flexor muscles of leg
21 Calcaneal tendon
22 Epiphyseal line of tibia
23 Calcaneus
24 **Talocrural joint**
25 Talus
26 Lateral meniscus
27 Epiphysial line

Talocrural joint. Sagittal section; anterior part to the left. (MR-Scan.)

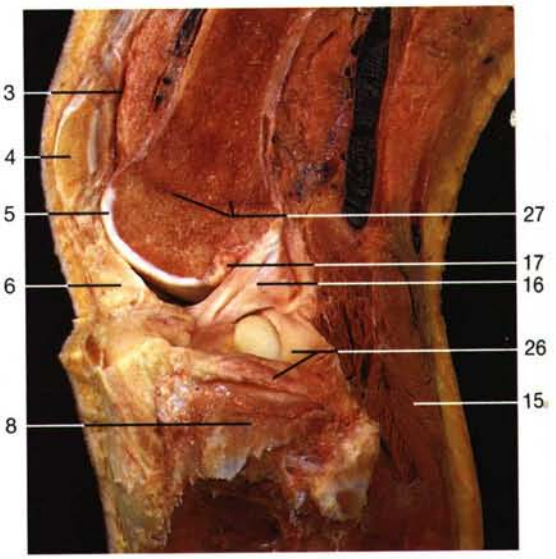

Left knee joint. Anterior cruciate ligament (lateral aspect).

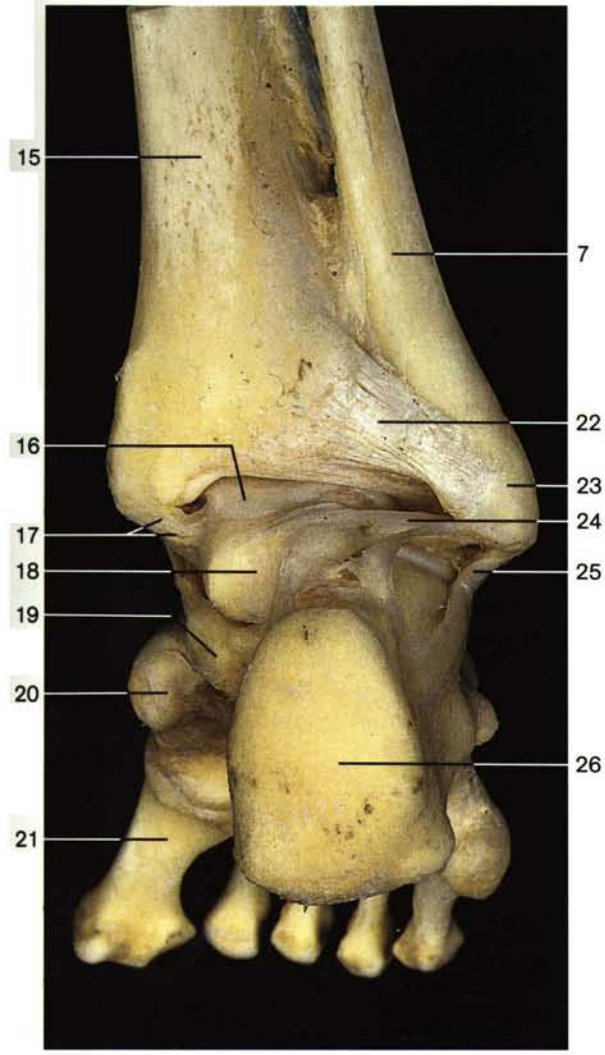

Ligaments of talocrural joint, right leg (posterior aspect).

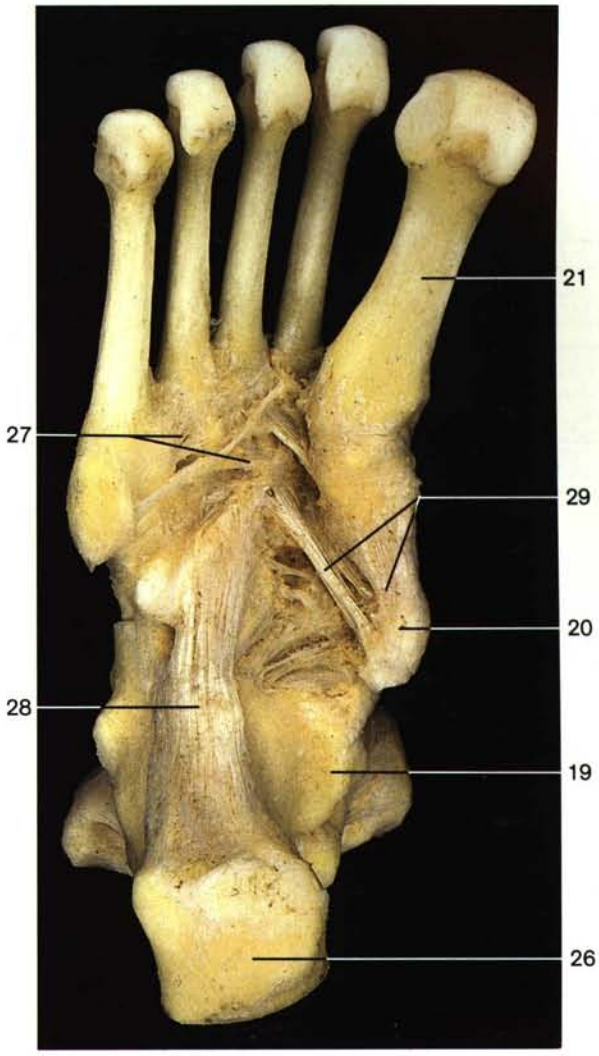

Deep ligaments of the foot, right foot (plantar aspect). The toes have been removed.

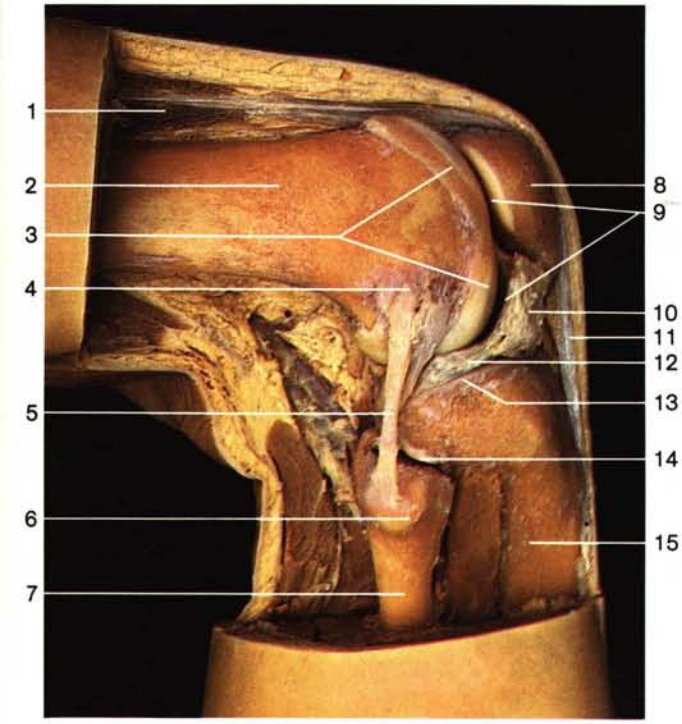

Right knee joint and tibiofibular joint with ligaments. Note the position of the lateral meniscus.

1 Quadriceps femoris muscle
2 Femur
3 Patellar surface
4 Lateral epicondyle of femur
5 **Fibular collateral ligament**
6 Head of fibula
7 **Fibula**
8 **Patella**
9 Articular cavity of knee joint
10 Infrapatellar fat pad
11 **Patellar ligament**
12 **Lateral meniscus of knee joint**
13 Lateral condyle of tibia (superior articular facet)
14 **Tibiofibular joint**
15 Tibia
16 Trochlea of talus (superior surface)
17 Deltoid ligament of ankle (posterior tibiotalar part)
18 Talus
19 **Sustentaculum tali**
20 Navicular bone
21 1st metatarsal bone
22 **Posterior tibiofibular ligament**
23 Lateral malleolus
24 Posterior talofibular ligament
25 Calcaneofibular ligament
26 Calcaneal tuberosity
27 Plantar tarsometatarsal ligaments
28 **Long plantar ligament**
29 Plantar cuneonavicular ligaments

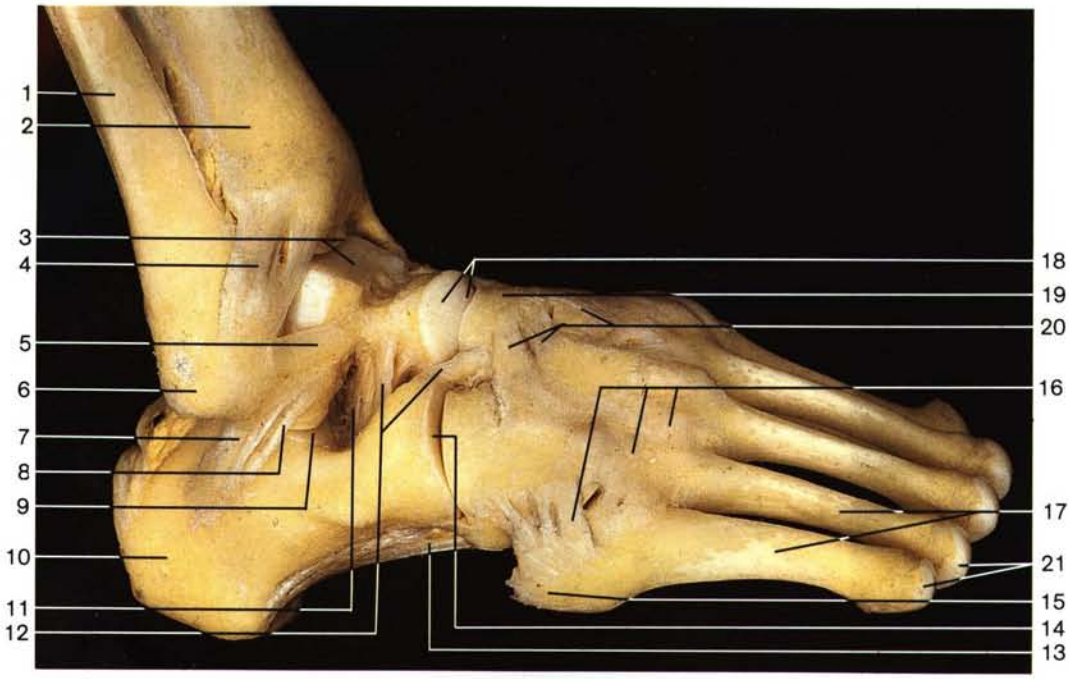

Ligaments of right foot (lateral aspect).

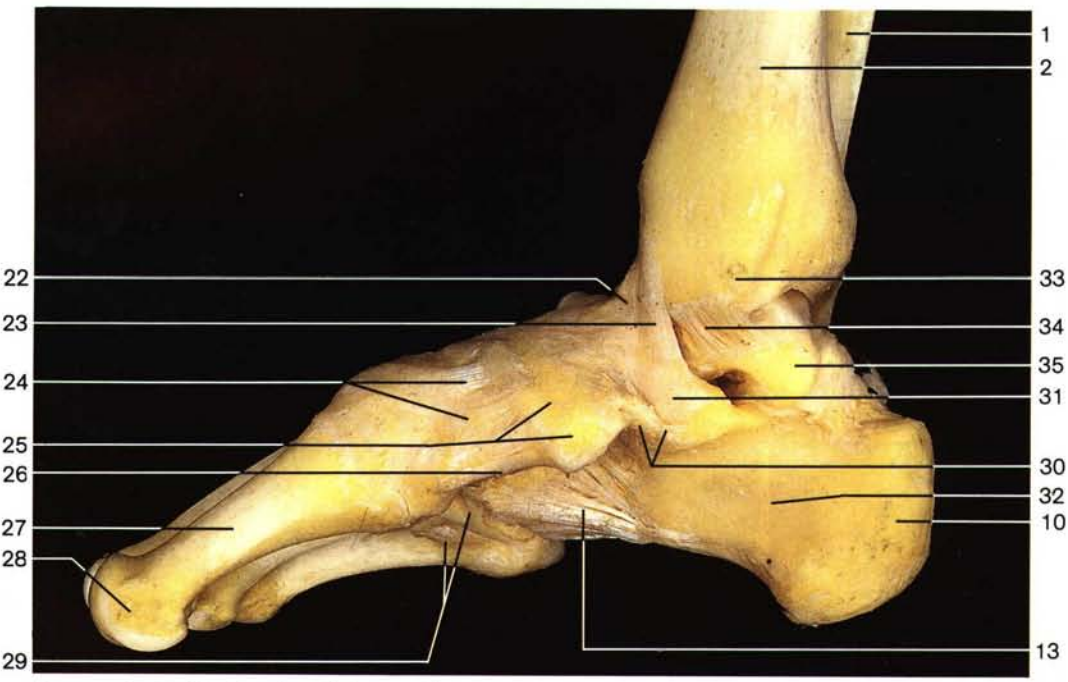

Ligaments of right foot (medial aspect).

1 Fibula
2 Tibia
3 Trochlea of talus and **talocrural joint**
4 Anterior tibiofibular ligament
5 Anterior talofibular ligament
6 Lateral malleolus
7 Calcaneofibular ligament
8 Lateral talocalcaneal ligament
9 **Subtalar joint**
10 Tuber calcanei
11 Interosseous talocalcaneal ligament
12 Bifurcate ligament
13 Long plantar ligament
14 **Calcaneocuboid joint**
15 Tuberosity of 5th metatarsal bone
16 Dorsal tarsometatarsal ligaments
17 Metatarsal bones
18 Head of talus and **talocalcaneonavicular joint**
19 Navicular bone
20 Dorsal cuneonavicular ligaments
21 Heads of metatarsal bones
22 Medial or deltoid ligament of ankle (tibionavicular part)
23 Medial or deltoid ligament of ankle (tibiocalcaneal part)
24 Dorsal cuneonavicular ligaments
25 Navicular bone
26 Plantar cuneonavicular ligament
27 1st metatarsal bone
28 Head of 1st metatarsal bone
29 Plantar tarsometatarsal ligaments
30 Plantar calcaneonavicular ligament
31 Sustentaculum tali
32 Calcaneus
33 Medial malleolus
34 Medial or deltoid ligament of ankle (posterior part)
35 Talus

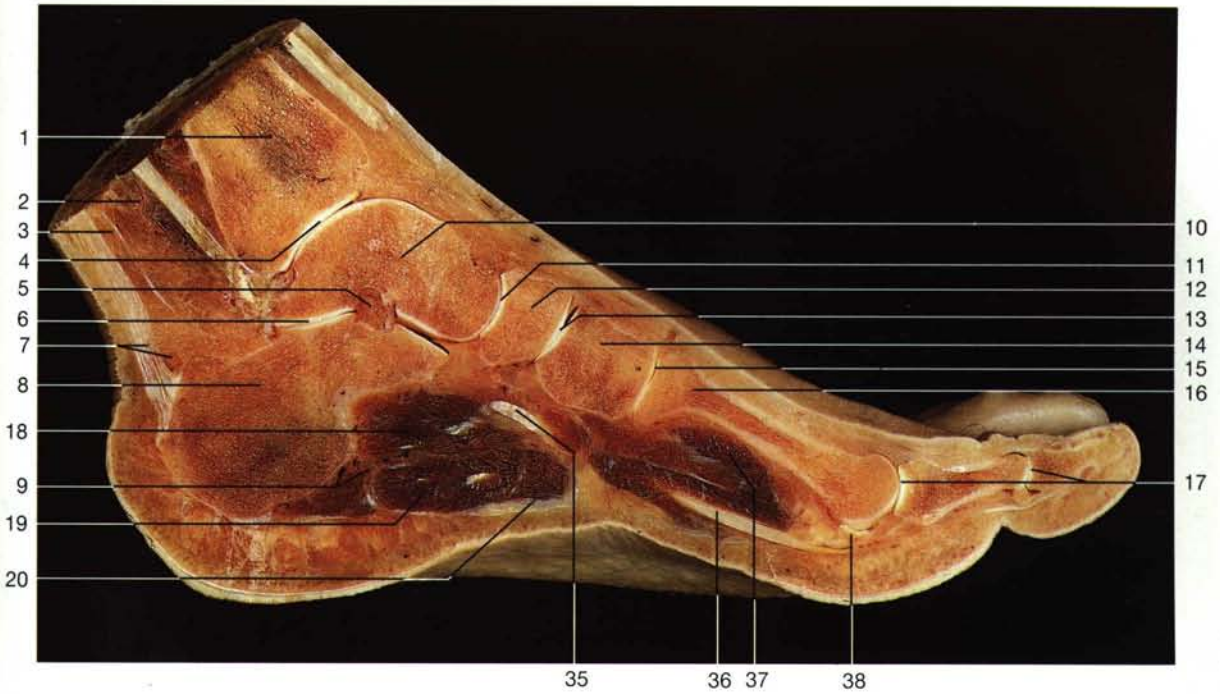

Longitudinal section through the foot at the level of first phalanx.

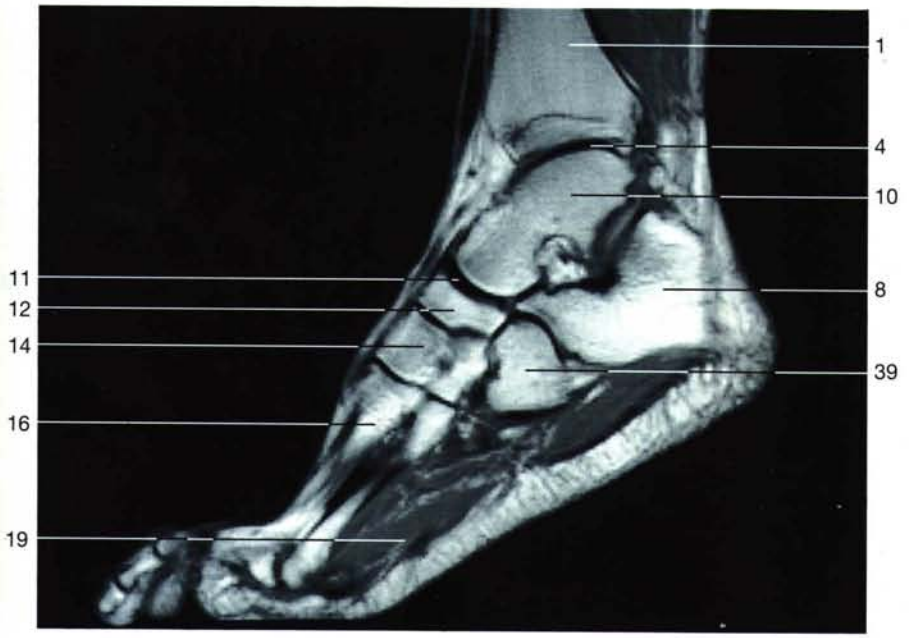

Sagittal section through the foot
(MR-Scan after A. Heuck, G. Luttke and J. W. Rohen, 1994).

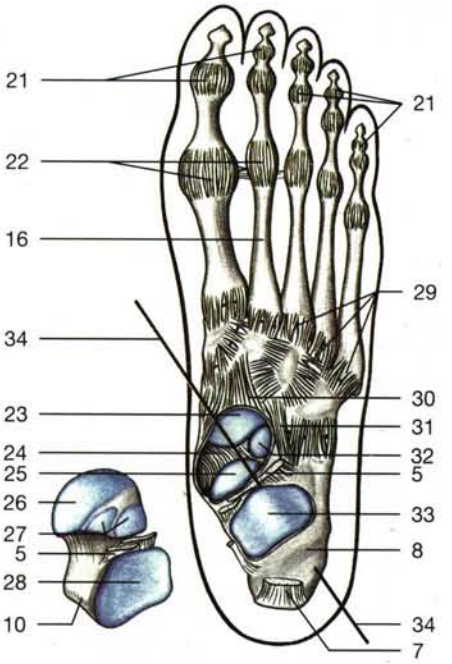

Talocalcaneonavicular joint.
The talus has been rotated to show the articular surfaces of the joint.

1 Tibia
2 Deep flexor muscles
3 Superficial flexor muscles
4 **Talocrural joint**
5 Interosseous talocalcaneal ligament
6 **Subtalar joint**
7 Calcaneal or Achilles tendon and bursa
8 Calcaneus
9 Vessels and nerves of foot
10 Talus
11 **Talocalcaneonavicular joint**
12 Navicular bone
13 **Cuneonavicular joint**
14 Intermediate cuneiform bone
15 **Tarsometatarsal joints**
16 Metatasal bones
17 **Metatarsophalangeal** and **interphalangeal joints**
18 Quadratus plantae muscle with flexor tendons
19 Flexor digitorum brevis muslce
20 Plantar aponeurosis
21 Articular capsules of interphalangeal joints
22 Articular capsules of metatarsophalangeal joints
23 Articular surface of navicular bone
24 Plantar calcaneonavicular ligament
25 Middle talar articular surface of calcaneus
26 Navicular articular surface of talus
27 Anterior and middle calcaneal surfaces of talus
28 Posterior calcaneal surface of talus
29 Dorsal tarsometatarsal ligaments
30 Talonavicular ligament
31 Bifurcate ligament
32 Anterior talar articular surface of calcaneus
33 Posterior talar articular surface of calcaneus
34 Axis for inversion and eversion
35 Tendon of tibialis posterior muscle
36 Tendon of flexor hallucis longus muscle
37 Flexor hallucis brevis muscle
38 Sesamoid bone
39 Cuboid bone

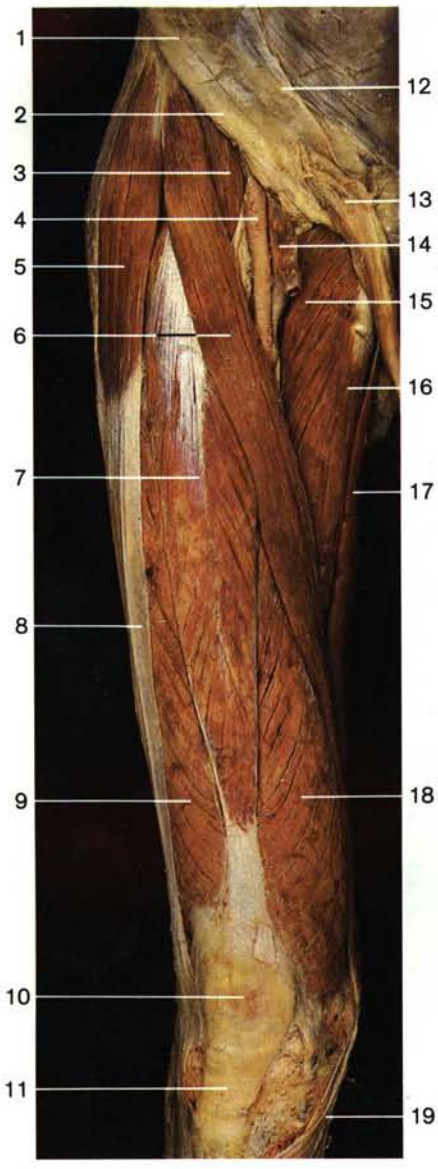

Extensor and adductor muscles of thigh, right thigh (anterior aspect).

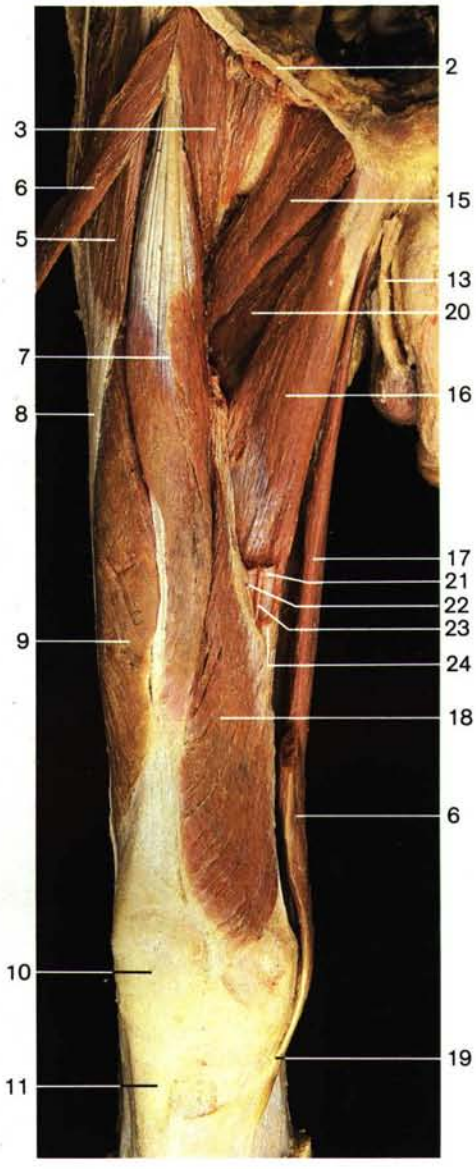

Quadriceps muscle and superficial layer of adductor muscles, right thigh (anterior aspect). The sartorius muscle has been divided.

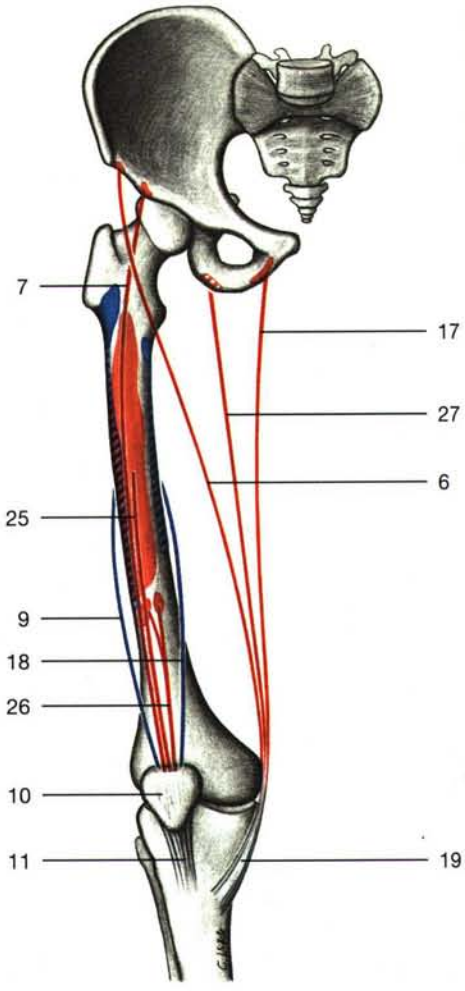

Course of extensor muscles of thigh and muscles inserting with common tendon on tibia.

1 Anterior superior iliac spine
2 Inguinal ligament
3 Iliopsoas muscle
4 Femoral artery
5 Tensor fasciae latae muscle
6 Sartorius muscle
7 **Rectus femoris muscle**
8 Iliotibial tract
9 Vastus lateralis muscle
10 Patella
11 Patellar ligament
12 Aponeurosis of external abdominal oblique muscle
13 Spermatic cord
14 Femoral vein
15 **Pectineus muscle**
16 **Adductor longus muscle**
17 Gracilis muscle
18 Vastus medialis muscle
19 Common tendon of sartorius, gracilis, and semitendinosus muscles (pes anserinus)
20 **Adductor brevis muscle**
21 Femoral artery } entering the adductor canal
22 Femoral vein } entering the adductor canal
23 Saphenous nerve } entering the adductor canal
24 Fascia of adductor canal
25 Vastus intermedius muscle
26 Articularis genus muscle
27 Semitendinosus muscle

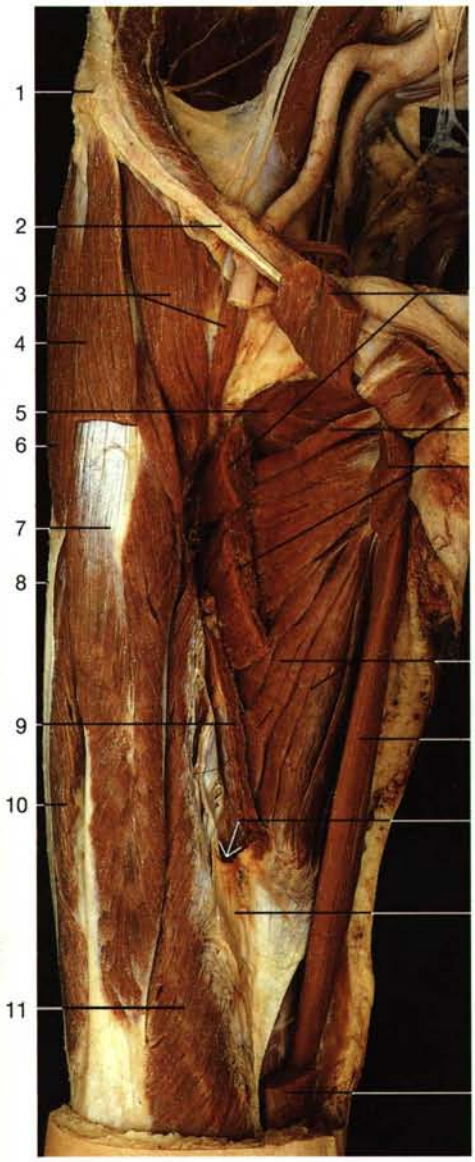

Deep layer of adductor muscles. Adductor magnus muscle (anterior aspect). Pectineus, adductor longus and brevis muscles have been divided.

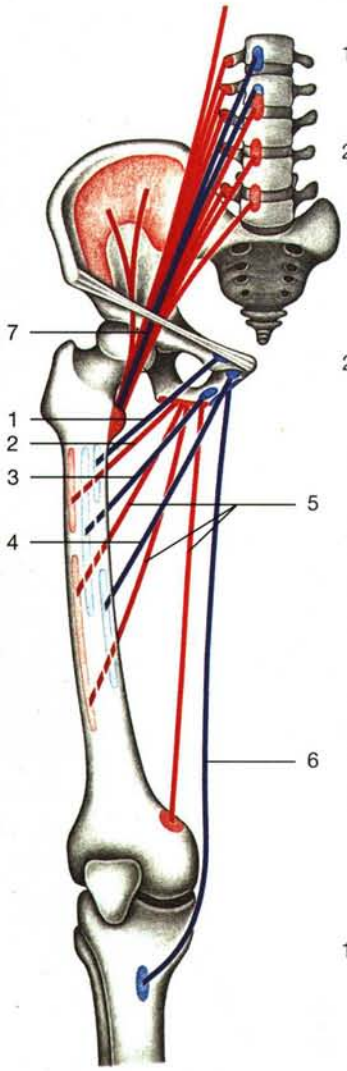

Course of adductor muscles. (Schematic drawing.)

1 Pectineus muscle (blue)
2 Adductor minimus muscle (red)
3 Adductor brevis muscle (blue)
4 Adductor longus muscle (blue)
5 Adductor magnus muscle (red)
6 Gracilis muscle (blue)
7 Iliopsoas muscle (red/blue)

Iliopsoas and adductor muscles; deepest layer (anterior aspect). Pectineus, adductor longus and brevis, rectus femoris muscles have been divided.

1 Anterior superior iliac spine
2 Inguinal ligament
3 **Iliopsoas muscle**
4 Sartorius muscle
5 Obturator externus muscle
6 Tensor fasciae latae muscle
7 Rectus femoris muscle
8 Iliotibial tract
9 Adductor longus muscle (divided)
10 Vastus lateralis muscle
11 Vastus medialis muscle
12 Pectineus muscle (divided)
13 Adductor minimus muscle
14 Adductor brevis muscle (cut)
15 **Adductor magnus**
16 Gracilis muscle
17 Adductor hiatus
18 Vastoadductor membrane
19 Diaphragm
20 Quadratus lumborum muscle
21 **Iliacus muscle**
22 Vastus intermedius muscle
23 Aorta in aortic hiatus
24 Twelfth rib
25 **Psoas minor muscle**
26 **Psoas major muscle**
27 Iliopectineal arch

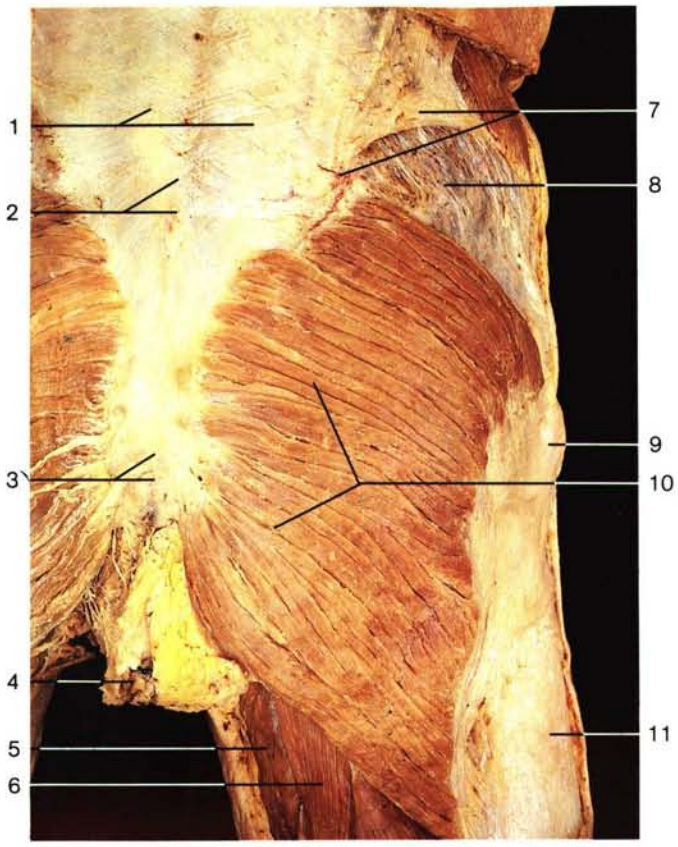

Gluteal muscles, superficial layer (posterior aspect).

Gluteal muscles, deeper layer (posterior aspect).

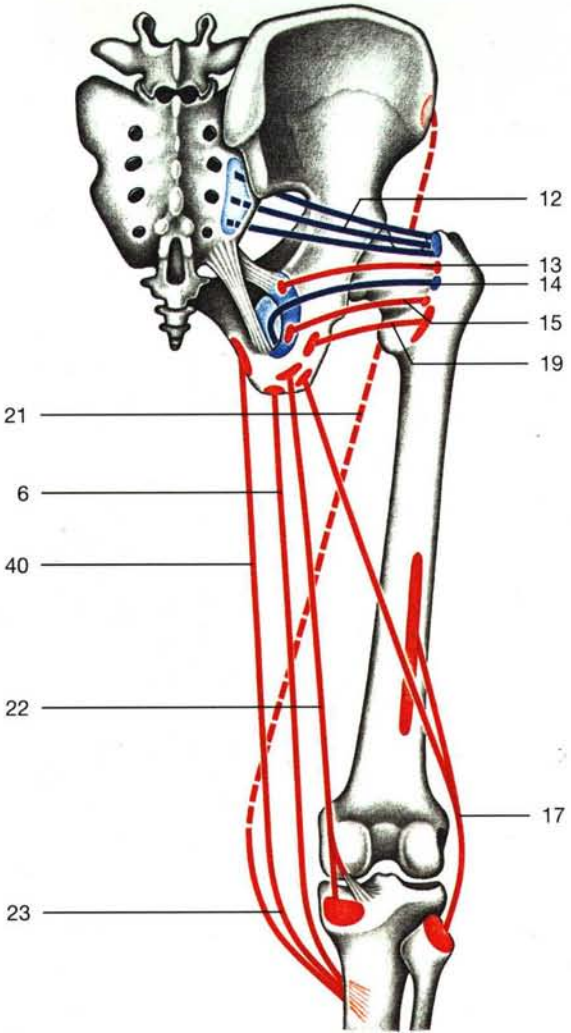

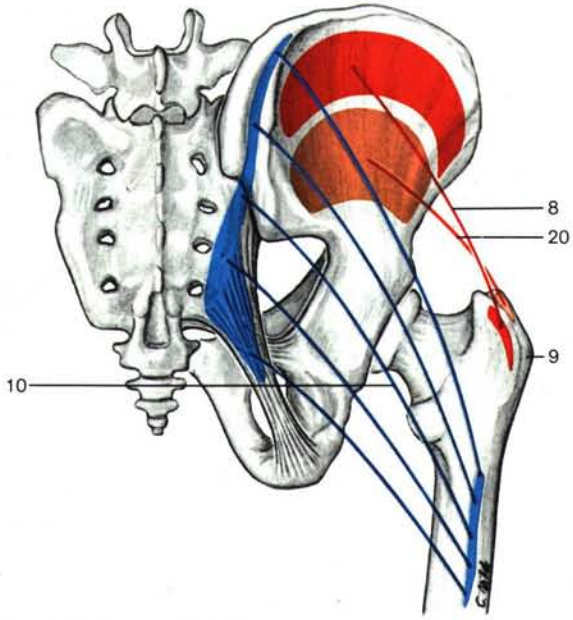

Course of gluteal muscles
(posterior aspect; schematic drawing).

◁
Course of gluteal muscles (deeper layer) **and of ischiocrural muscles** (posterior aspect).
Sartorius muscle is indicated by dotted line (schematic drawing).

1 Thoracolumbar fascia
2 Spinous processes of lumbar vertebrae
3 Coccyx
4 Anus
5 Adductor magnus muscle
6 Semitendinosus muscle
7 Iliac crest
8 **Gluteus medius muscle**
9 Greater trochanter
10 **Gluteus maximus muscle**
11 Iliotibial tract
12 Piriformis muscle
13 Superior gemellus muscle
14 Obturator internus muscle
15 Inferior gemellus muscle
16 Ischial tuberosity
17 **Biceps femoris muscle**
18 Tensor fasciae latae muscle
19 Quadratus femoris muscle
20 **Gluteus minimus muscle**
21 Sartorius muscle
22 Semimembranosus muscle
23 Tendon of gracilis muscle
24 Tibial nerve
25 Medial head of gastrocnemius muscle
26 Common peroneal nerve
27 Tendon of biceps femoris muscle
28 Lateral head of gastrocnemius muscle
29 Rectus femoris muscle
30 Vastus medialis muscle
31 Vastus intermedius muscle
32 Vastus lateralis muscle
33 Sciatic nerve
34 Gluteus maximus muscle (insertion)
35 Great saphenous vein
36 Femoral artery
37 Femoral vein
38 Adductor longus muscle
39 Femur
40 Gracilis muscle
41 Septum between semitendinosus and semimembranosus muscles

Flexors of the right thigh, superficial layer (dorsal aspect).

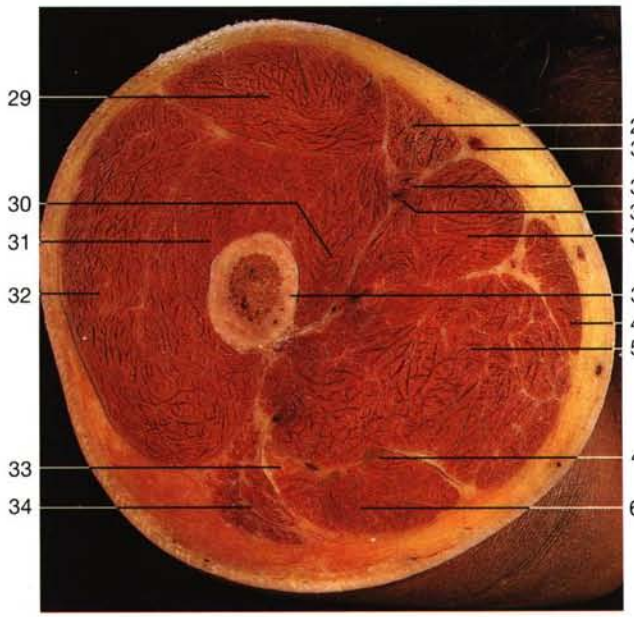

Cross section of right thigh (inferior aspect). Anterior side on top.

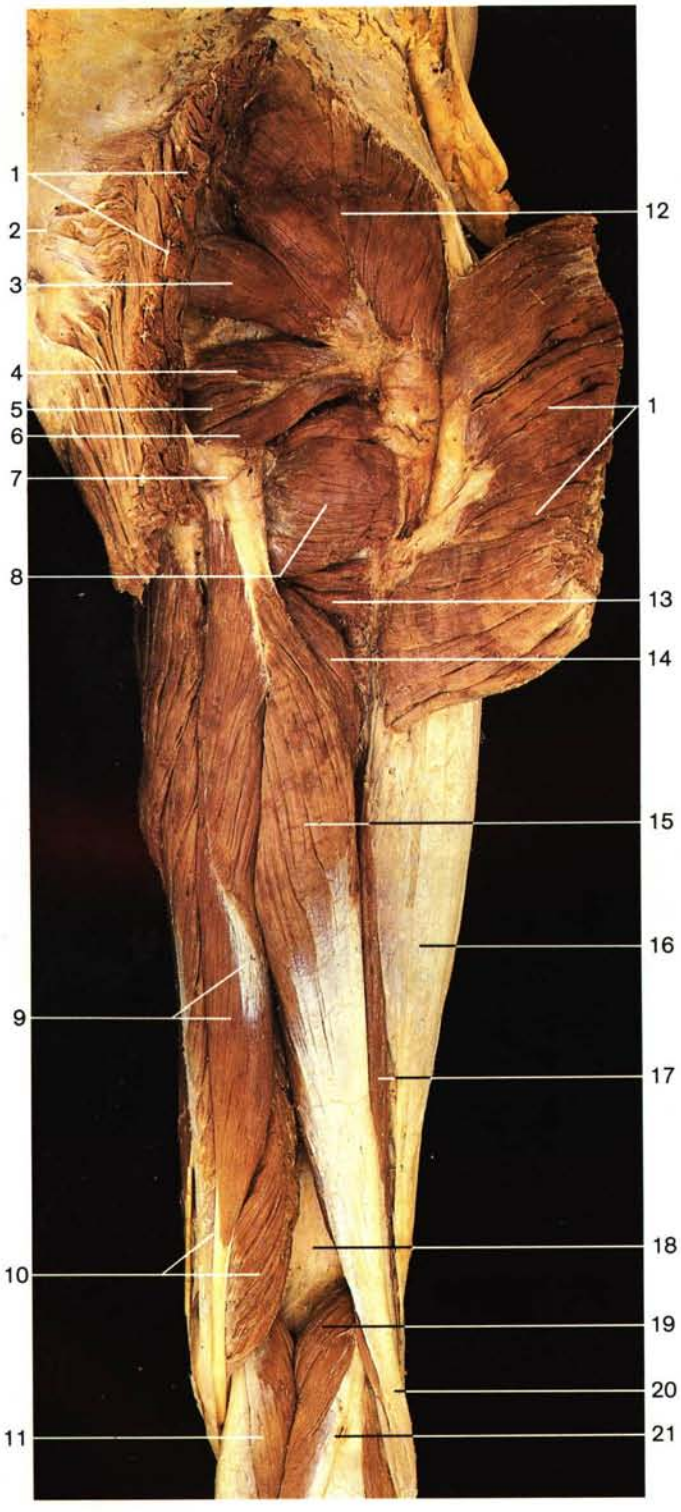

Dorsal muscles of right thigh (posterior aspect).
The gluteus maximus muscle has been cut and reflected.

Dorsal muscles of right thigh (posterior aspect).
The gluteus maximus muscle and the long head of biceps femoris muscle have been divided and displaced.

1 Gluteus maximus muscle (divided)
2 Position of coccyx
3 Piriformis muscle
4 Superior gemellus muscle
5 Obturator internus muscle
6 Inferior gemellus muscle
7 Ischial tuberosity
8 Quadratus femoris muscle
9 Semitendinosus muscle with intermediate tendon
10 Semimembranosus muscle
11 Medial head of gastrocnemius muscle
12 Gluteus medius muscle
13 Adductor minimus muscle
14 Adductor magnus muscle
15 Long head of biceps femoris muscle
16 Iliotibial tract
17 Short head of biceps femoris muscle
18 Popliteal surface of femur
19 Plantaris muscle
20 Tendon of biceps femoris muscle
21 Lateral head of gastrocnemius muscle
22 Membranous part of semimembranosus muscle

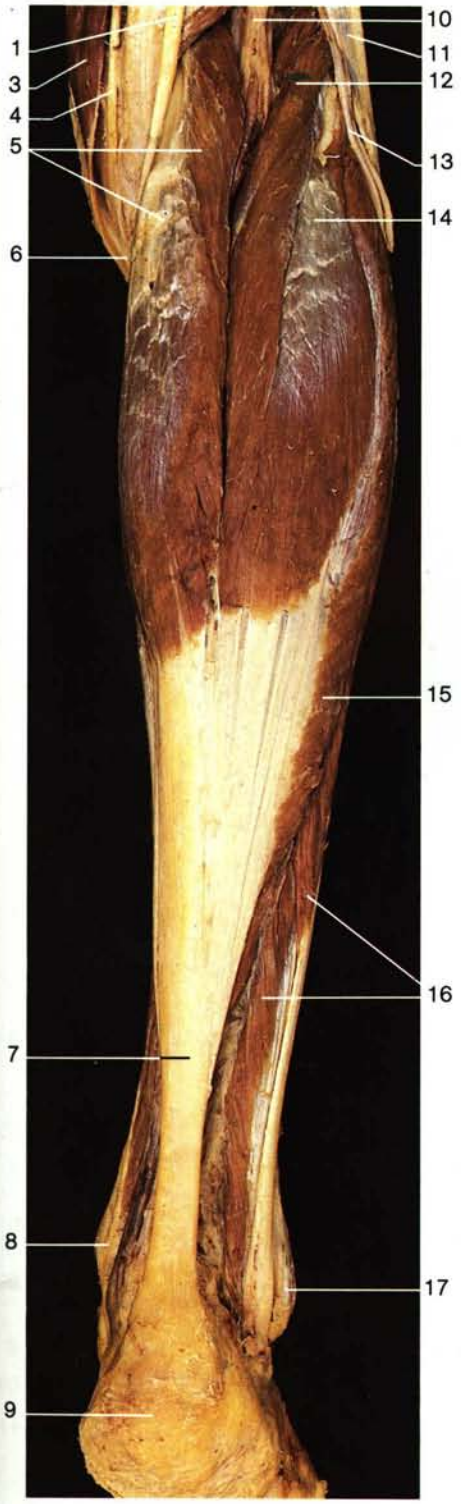

Flexor muscles of right leg (posterior aspect).

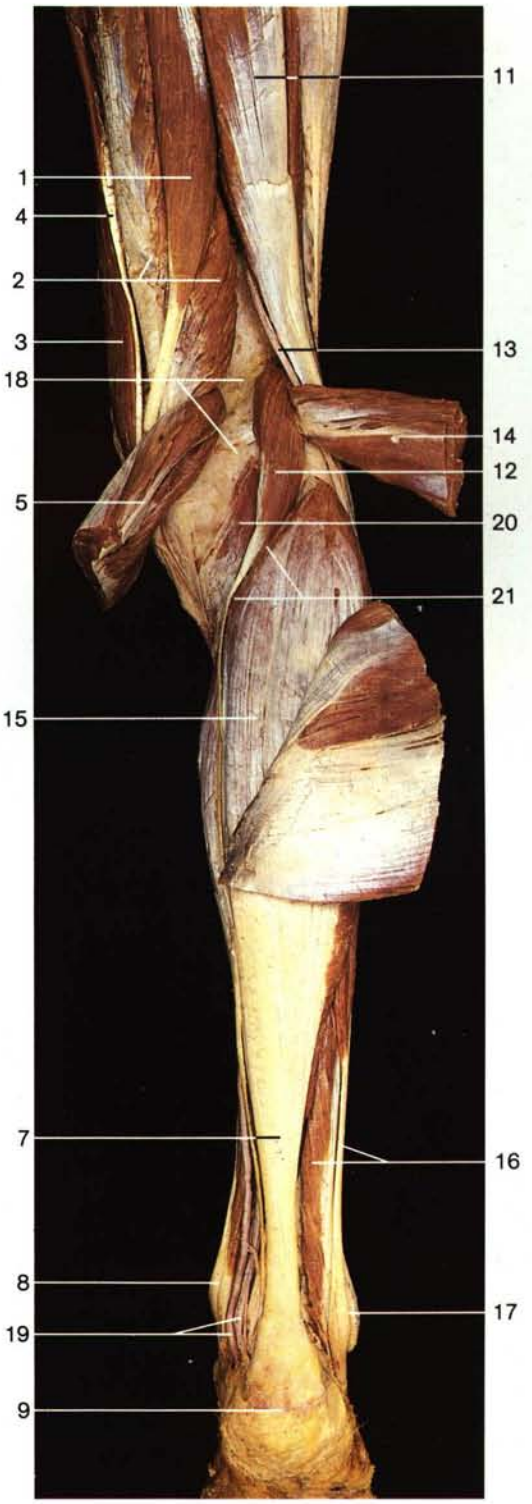

Flexor muscles of right leg (posterior aspect). Both heads of the gastrocnemius muscle have been cut and reflected.

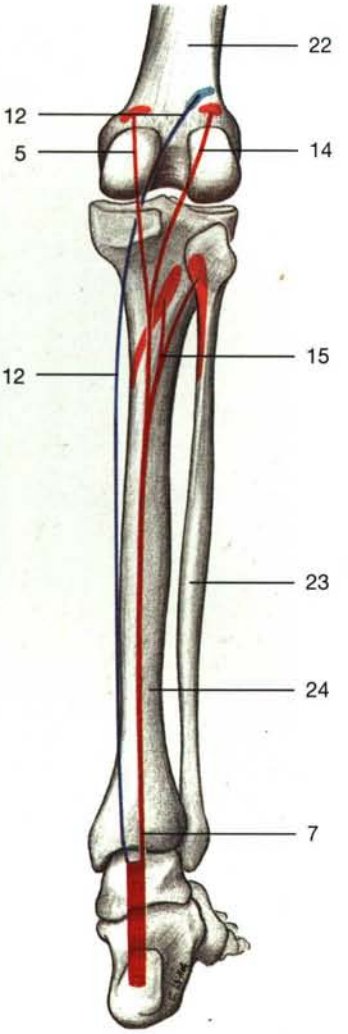

Flexor muscles of the leg (right side).

1 Semitendinosus muscle
2 Semimembranosus muscle
3 Sartorius muscle
4 Tendon of gracilis muscle
5 **Medial head of gastrocnemius muscle**
6 Common tendon of gracilis, sartorius, and semitendinosus muscles
7 Calcaneal or Achilles tendon
8 Medial malleolus
9 Calcaneal tuberosity
10 Tibial nerve
11 Biceps femoris muscle
12 Plantaris muscle
13 Common peroneal nerve
14 **Lateral head of gastrocnemius muscle**
15 **Soleus muscle**
16 Peroneus longus and brevis muscles
17 Lateral malleolus
18 Popliteal fossa
19 Tibial nerve and posterior tibial artery
20 Popliteus muscle
21 Tendinous arch of soleus muscle
22 Femur
23 Fibula
24 Tibia

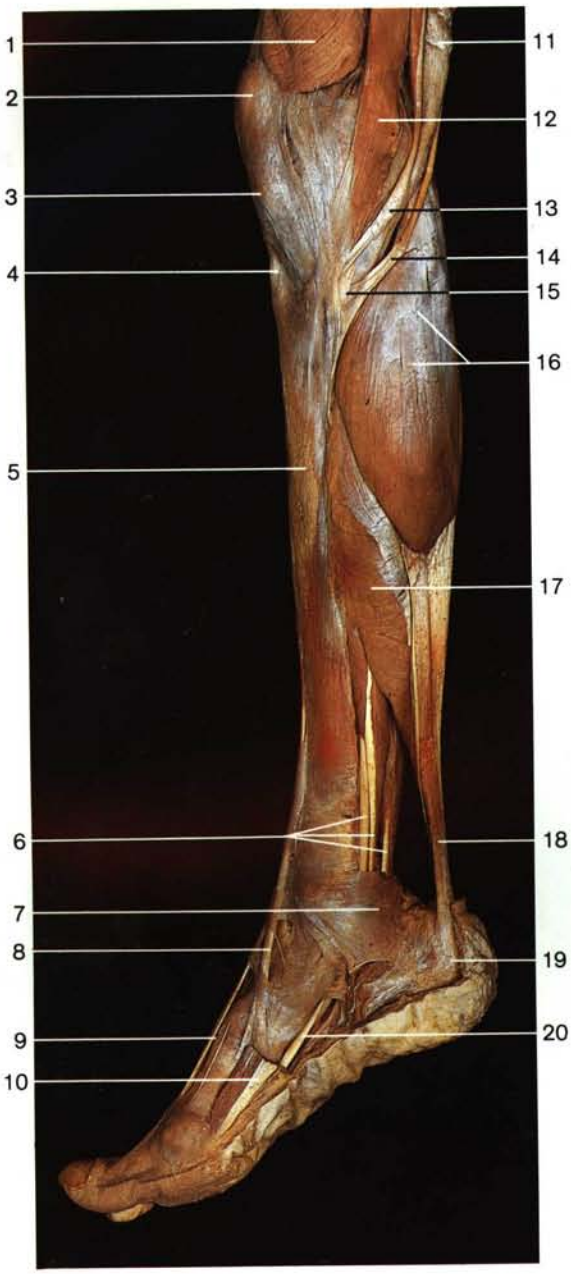

Muscles of right leg and foot (medial aspect).

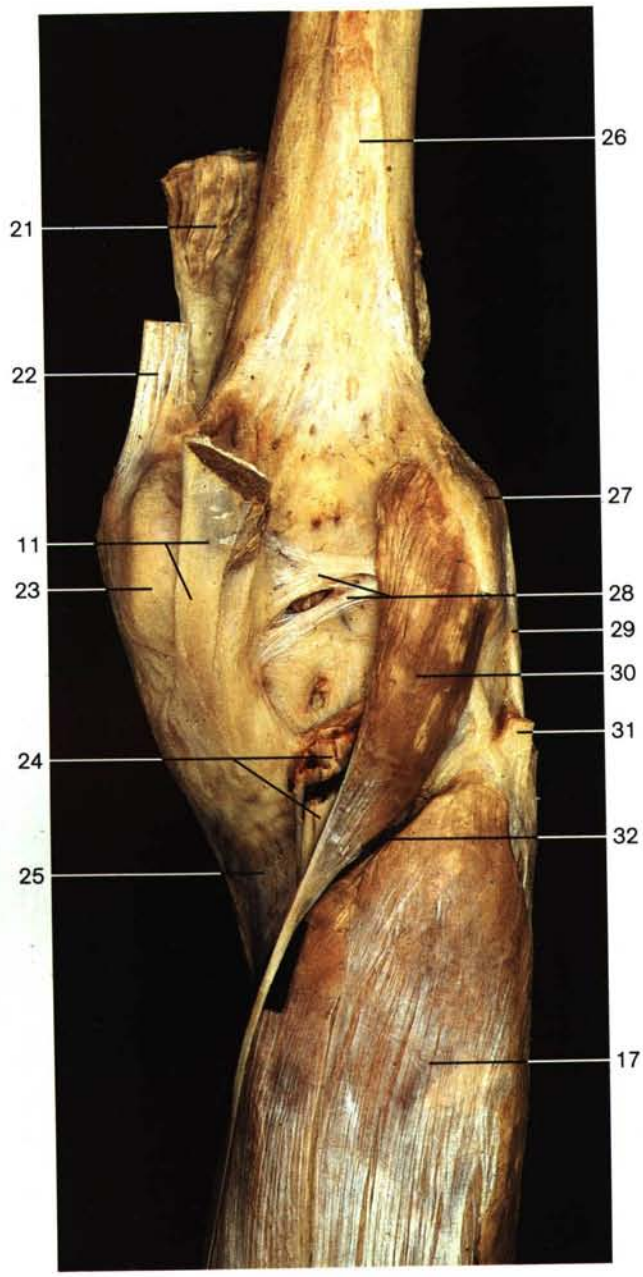

Popliteal region with plantaris and soleus, right side (dorsal aspect). Notice the insertion of the tendon of semimembranosus.

1 Vastus medialis muscle
2 Patella
3 Patellar ligament
4 Tibial tuberosity
5 Tibia
6 Tendons of deep flexor muscles (from anterior to posterior: 1. tibialis posterior; 2. flexor digitorum longus; 3. flexor hallucis longus muscles)
7 Flexor retinaculum
8 Tendon of tibialis anterior muscle
9 Tendon of extensor hallucis longus muscle
10 Abductor hallucis muscle
11 Semimembranosus muscle
12 Sartorius muscle
13 Tendon gracilis muscle
14 Tendon of semitendinosus muscle
15 Common tendon of gracilis, semitendinosus and sartorius muscles
16 **Medial head of gastrocnemius muscle**
17 **Soleus muscle**
18 Calcaneal or Achilles tendon
19 Calcaneus muscle
20 Tendon of flexor hallucis longus muscle
21 Quadriceps femoris muscle (divided)
22 Tendon of adductor magnus muscle (divided)
23 Medial condyle of femur
24 Popliteal artery and vein, tibial nerve
25 Tibia
26 Femur
27 Lateral epicondyle of femur
28 Oblique popliteal ligament
29 Lateral (fibular) collateral ligament
30 **Plantaris muscle**
31 Tendon of biceps femoris muscle (divided)
32 Tendinous arch of soleus muscle

Muscles of right leg and foot (lateral aspect).

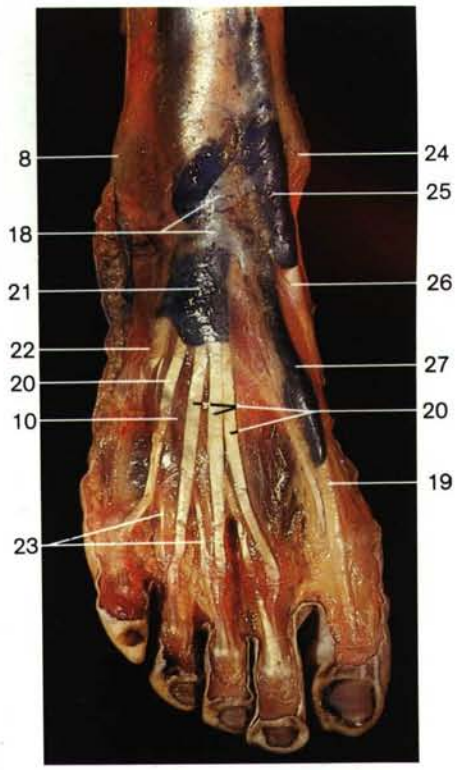

Right foot with synovial sheaths of extensor muscles (dorsal aspect). The synovial sheaths have been injected with blue solution.

1 Common peroneal nerve
2 Head of fibula
3 Lateral head of gastrocnemius muscle
4 Soleus muscle
5 Peroneus longus muscle
6 Peroneus brevis muscle
7 Calcaneal or Achilles tendon
8 Lateral malleolus muscle
9 Tendon of peroneus longus muscle
10 Extensor digitorum brevis muscle
11 Tendon of peroneus brevis muscle
12 Patella
13 Patellar ligament
14 Tuberosity of tibia
15 Tibialis anterior muscle
16 Extensor digitorum longus muscle
17 Superior extensor retinaculum
18 Inferior extensor retinaculum
19 Tendon of extensor hallucis longus muscle
20 Tendons of extensor digitorum longus muscle
21 Common synovial sheath of extensor digitorum longus muscle
22 Tendon of peroneus tertius muscle to the lateral margin of foot
23 Tendons of extensor digitorum brevis muscle
24 Medial malleolus
25 Synovial sheath of tendon of tibialis anterior muscle
26 Tendon of tibialis anterior muscle
27 Synovial sheath of tendon of extensor hallucis longus muscle

1 Medial condyle of femur
2 Popliteus muscle
3 Flexor digitorum longus muscle
4 Crosssing of tendons in leg
5 Tendon of tibialis posterior muscle
6 Tendon of flexor digitorum longus muscle
7 Medial malleolus
8 Lateral condyle of femur
9 Head of fibula
10 Tibialis posterior muscle
11 Flexor hallucis longus muscle
12 Peroneus longus muscle
13 Peroneus brevis muscle
14 Tendon of flexor hallucis longus muscle
15 Calcaneal tendon (divided)
16 Lateral malleolus

Deep flexor muscles of right leg (posterior aspect).

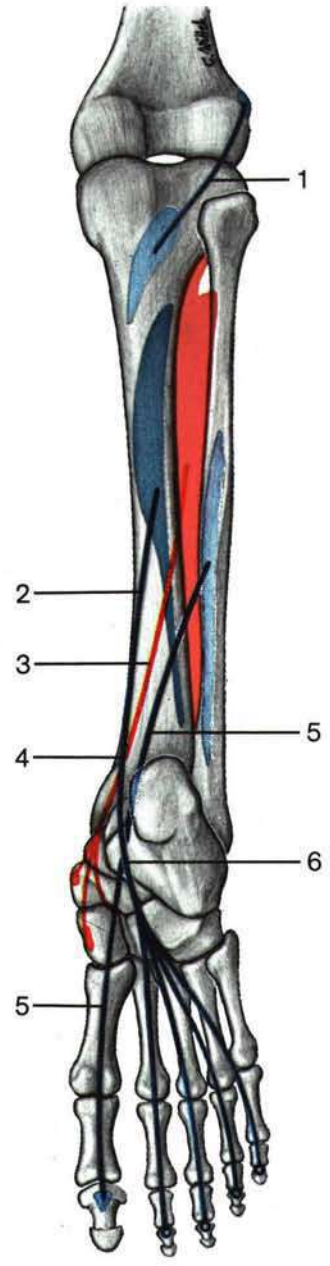

1 Popliteus muscle (blue)
2 Flexor digitorum longus muscle (blue)
3 Tibialis posterior muscle (red)
4 Crossing of tendons in leg
5 Flexor hallucis longus muscle (blue)
6 Crossing of tendons in sole

Course of deep flexor muscle of leg. (Schematic drawing.)

1 Medial condyle of femur
2 Tibia
3 Flexor digitorum longus muscle
4 Crossing of tendons in leg
5 Tendon of tibialis posterior muscle
6 Abductor hallucis muscle
7 Tendon of flexor hallucis longus muscle
8 Lateral condyle of femur
9 Head of fibula
10 Tibialis posterior muscle
11 Tendon of flexor digitorum longus muscle
12 Flexor retinaculum
13 Calcaneal tendon
14 Calcaneal tuberosity
15 Crossing of tendons in sole
16 Quadratus plantae muscle
17 Tendons of flexor digitorum longus muscle
18 Tendon of tibialis anterior muscle
19 Area of insertion of tibialis posterior muscle
20 Lumbrical muscles
21 Flexor hallucis longus muscle

Deep flexor muscles of right leg and foot (posterior oblique medial aspect). Felxor digitorum brevis and flexor hallucis longus muscles have been removed.

Sole of foot; tendons of long flexor muscles (oblique medial and inferior aspect).

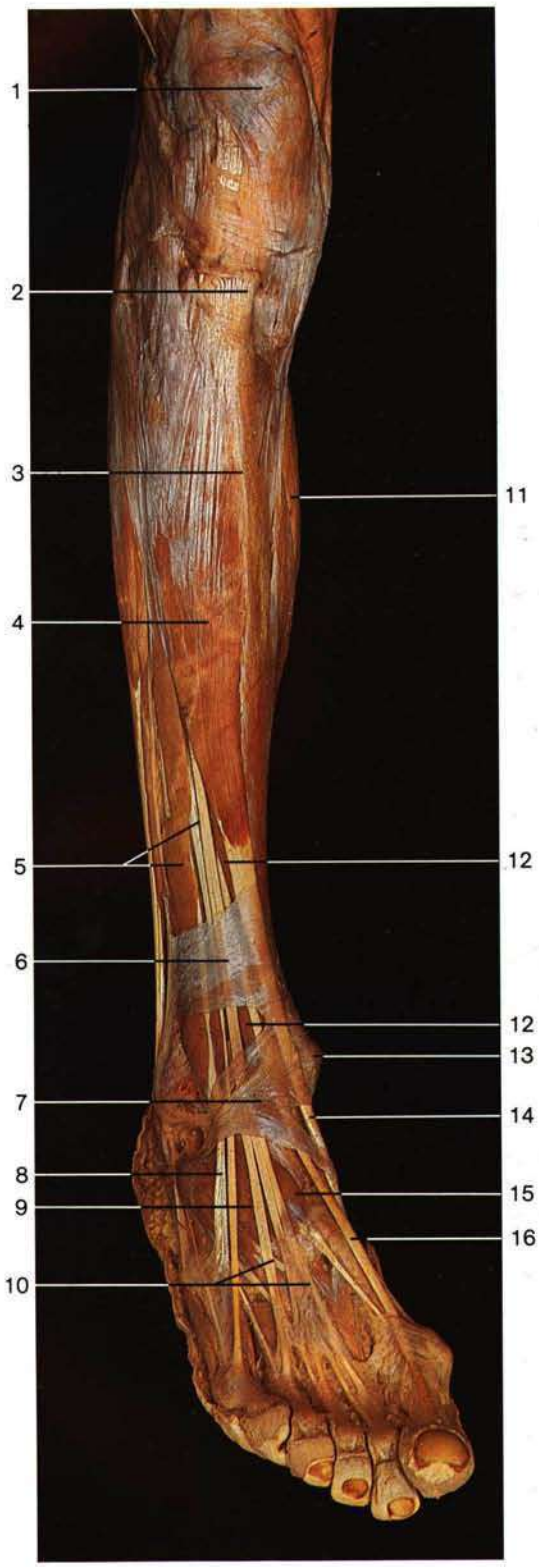

Extensor muscles of right leg and foot (oblique anterolateral aspect).

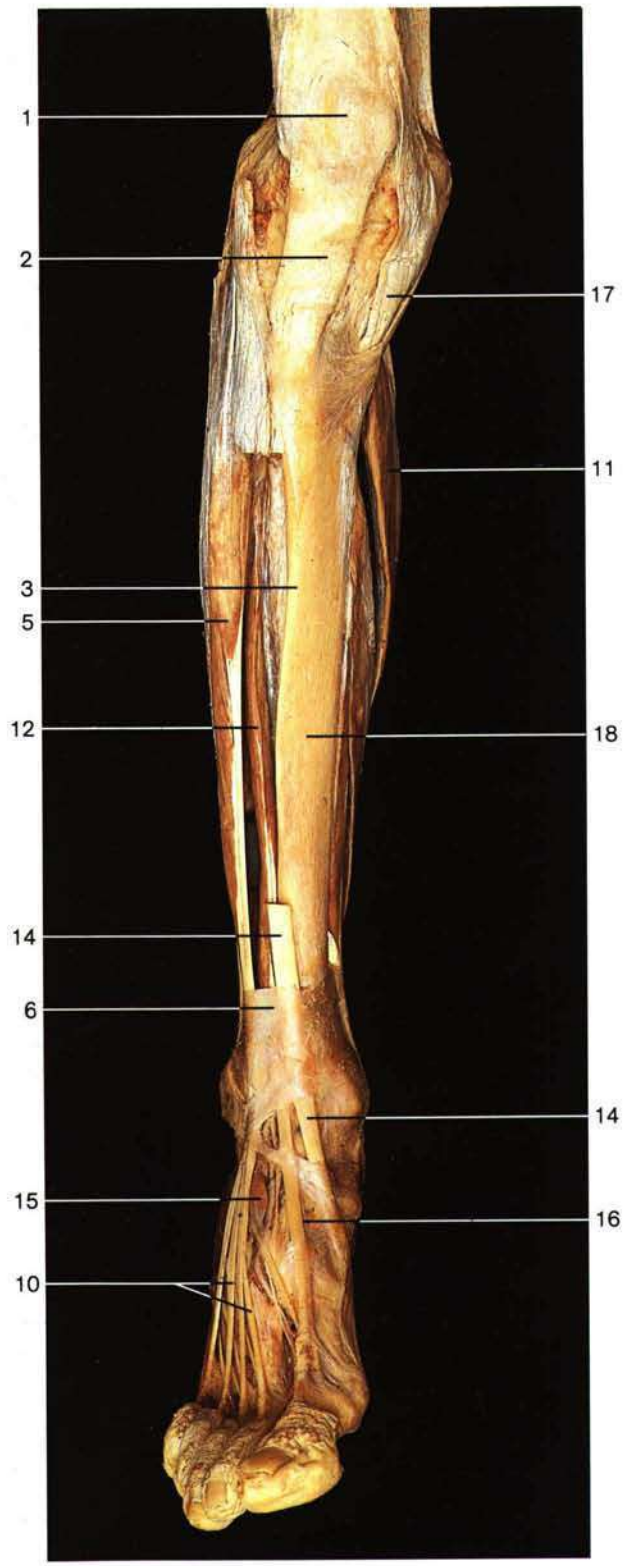

Extensors of the right leg and foot (anterior aspect). Part of the tibialis anterior muscle has been removed.

1 Patella
2 Patellar ligament
3 Anterior margin of tibia
4 Tibialis anterior muscle
5 Extensor digitorum longus muscle
6 Superior extensor retinaculum
7 Inferior extensor retinaculum
8 Tendon of peroneus tertius muscle
9 Extensor digitorum brevis muscle
10 Tendons of extensor digitorum longus muscle
11 Soleus
12 Extensor hallucis longus muscle
13 Medial malleolus
14 Tendon of tibialis anterior muscle
15 Extensor hallucis brevis muscle
16 Tendon of extensor hallucis longus muscle
17 Common tendon of gracilis, semitendinosus and sartoris muscles
18 Tibia

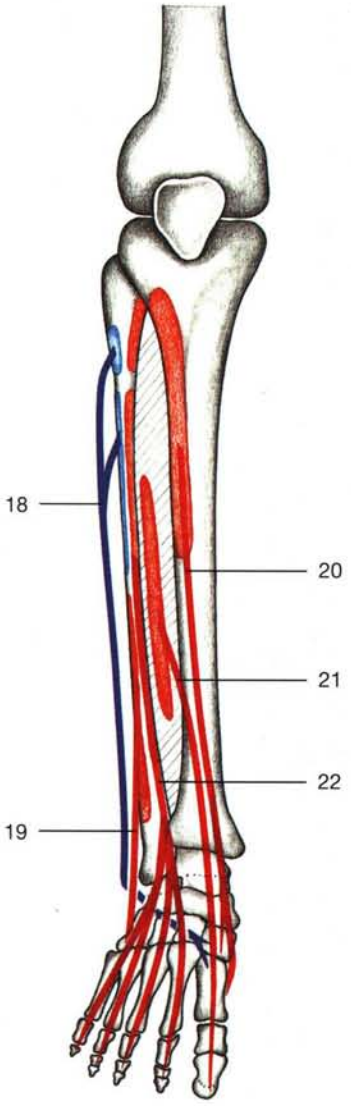

Extensor muscles of the leg (right side).

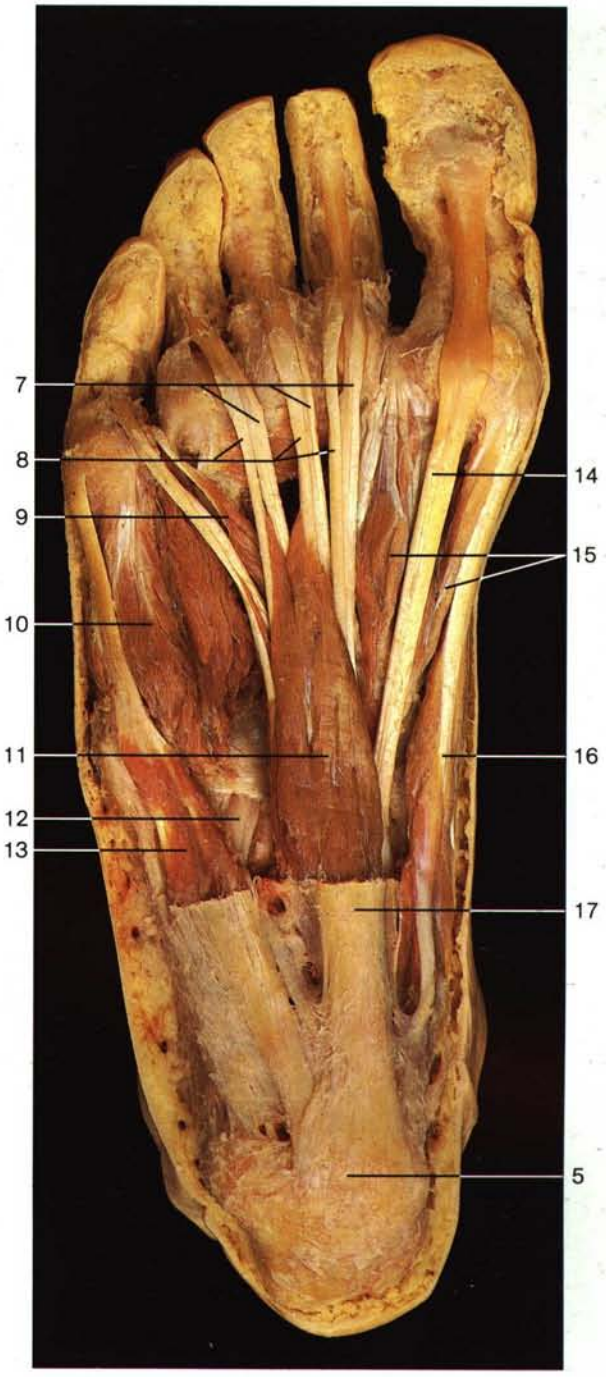

Sole of foot, first layer of muscles (from below). The plantar aponeurosis and the fasciae of the superficial muscles have been removed.

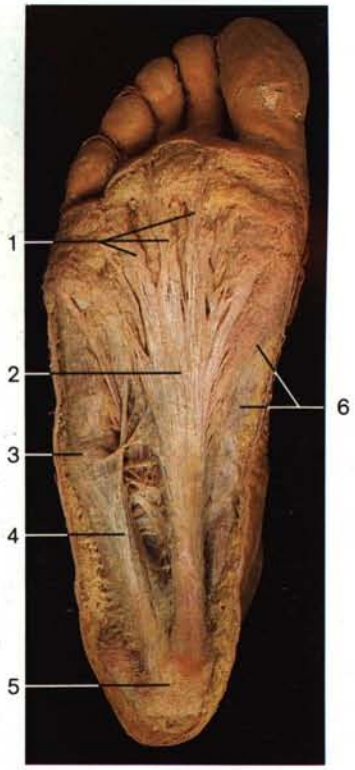

Sole of foot, plantar aponeurosis (from below).

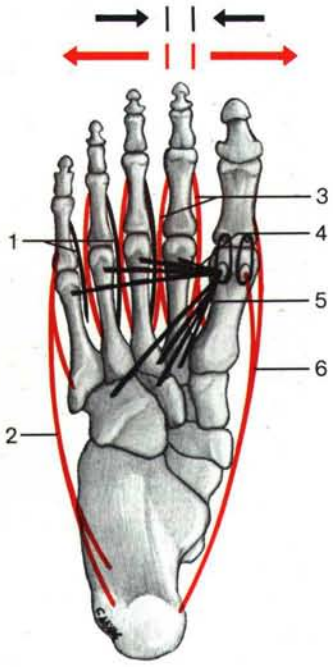

Course of abductor and adductor muscles of foot. (Schematic drawing.)
Red arrows = abduction.
Black arrows = adduction.

1 Longitudinal bands of plantar aponeurosis
2 Plantar aponeurosis
3 Position of tuberosity of 5th metatarsal bone
4 Muscles of 5th toe with fascia
5 Calcaneal tuberosity
6 Muscles of great toe with fascia
7 Tendons of flexor digitorum longus muscle
8 Tendons of flexor digitorum brevis muscle
9 Lumbrical muscle
10 Flexor digiti minimi brevis muscle
11 Flexor digitorum brevis muscle
12 Tendon of peroneus longus muscle
13 Abductor digiti minimi muscle
14 Tendon of flexor hallucis longus muscle
15 Flexor hallucis brevis muscle
16 Abductor hallucis muscle
17 Plantar aponeurosis (cut)
18 Peroneus longus muscle
19 Peroneus brevis muscle
20 Tibialis anterior muscle
21 Extensor hallucis longus muscle
22 Extensor digitorum longus muscle

1 Plantar interossei muscles (black)
2 Abductor digiti minimi muscles (red)
3 Dorsal interossei muscle (red)
4 Transverse head of adductor muscle (black)
5 Oblique head of adductor muscle (black)
6 Abductor hallucis muscle (red)

Muscles of sole of foot, second layer (from below). The flexor digitorum brevis muscle has been divided.

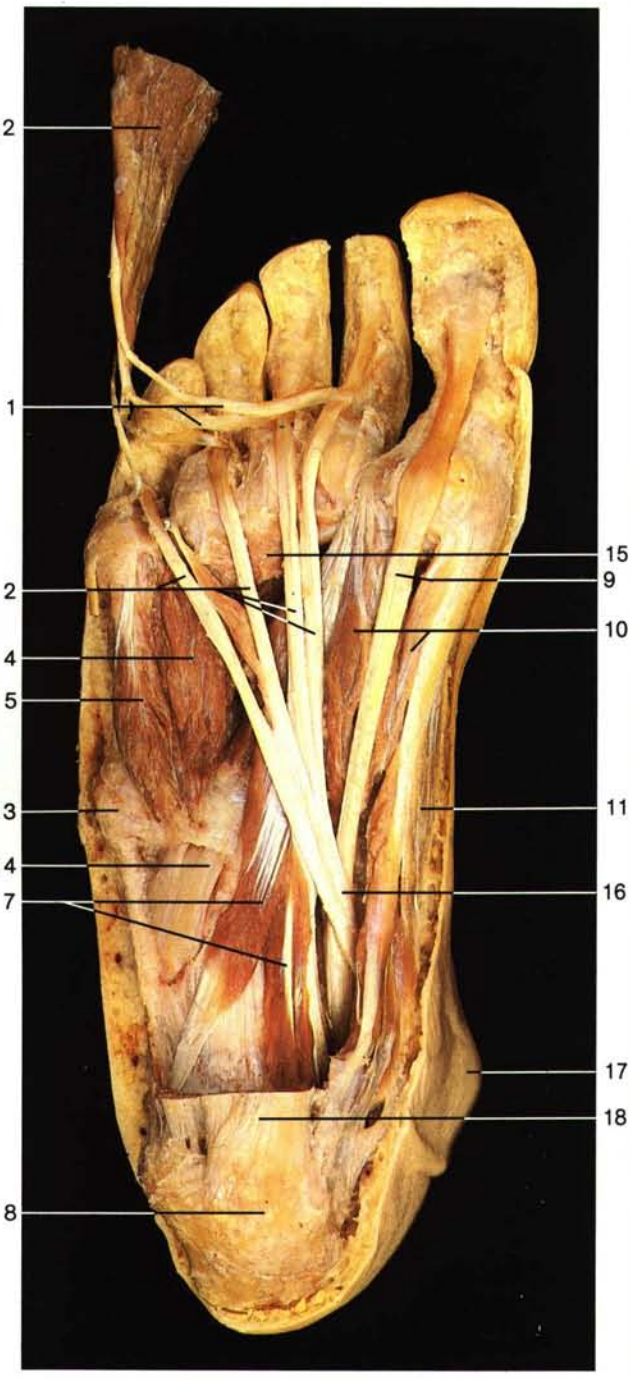

Muscles of sole of foot, second layer (from below). The tendons of the flexor muscles and the crossing of tendons are displayed. The flexor digitorum brevis muscle has been divided and reflected.

1 Tendons of flexor digitorum brevis muscle
2 Tendons of flexor digitorum longus muscle
3 Lumbrical muscles
4 Interossei muscles
5 Flexor digiti minimi brevis muscle
6 Abductor digiti minimi muscle
7 Quadratus plantae muscle
8 Calcaneal tuberosity
9 Tendon of flexor hallucis longus muscle
10 Flexor hallucis brevis muscle
11 Abductor hallucis muscle
12 Flexor digitorum brevis muscle (divided)
13 Tuberosity of 5th metatarsal bone
14 Tendon of peroneus longus muscle
15 Transverse head of adductor hallucis muscle
16 Crossing of tendons in sole of foot
17 Medial malleolus
18 Plantar aponeurosis (divided)

Muscles of sole of foot, third layer (from below). The flexor digitorum brevis muscle has been removed, and the quadratus plantae muscle and the abductor hallucis and digiti minimi muscles have been divided.

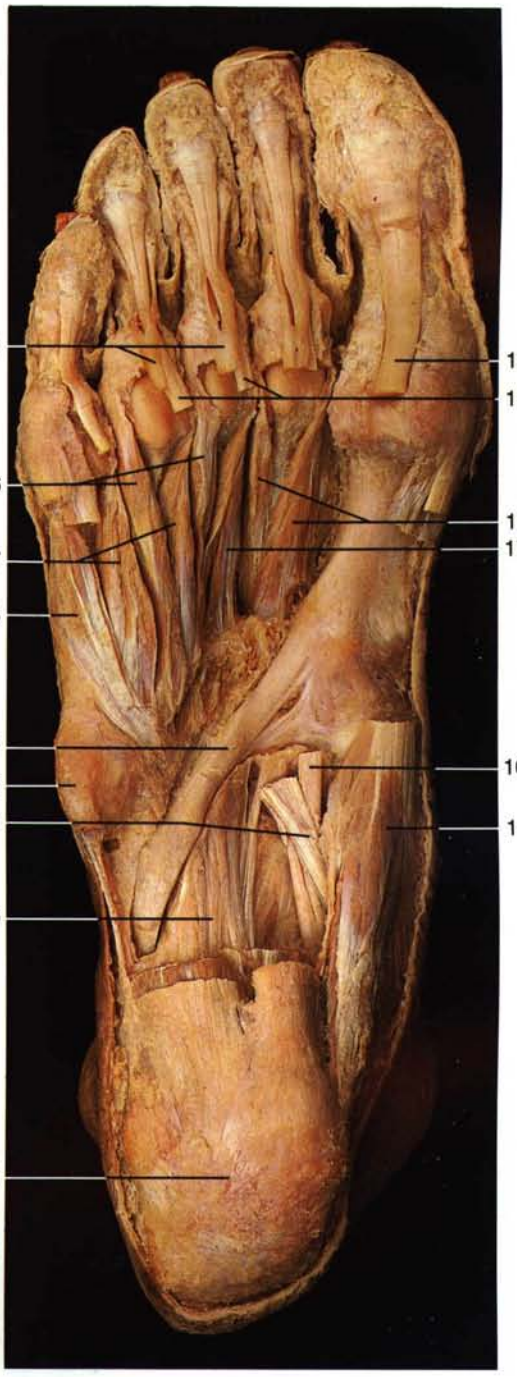

Muscles of sole of foot, fourth layer (from below). The interosseous muscles and the canal for the tendon of peroneus longus muscle are shown.

1 Tendons of flexor digitorum brevis muscle
2 Transverse head of adductor hallucis muscle
3 Abductor digiti minimi muscle
4 Interosseous muscles
5 Flexor digiti minimi brevis muscle
6 Opponens digiti minimi muscle
7 Tendon of peroneus longus muscle
8 Quadratus plantae muscle with tendon of flexor digitorum longus muscle
9 Calcaneal tuberosity
10 Tendons of flexor hallucis longus muscle (divided)
11 Tendon of flexor digitorum longus muscle
12 Flexor hallucis brevis muscle
13 Oblique head of adductor hallucis muscle
14 Abductor hallucis muscle (cut)
15 Tendon of tibialis posterior muscle
16 Dorsal interossei muscles
17 Plantar interossei muscles
18 Tuberosity of 5th metatarsal bone
19 Tendon of flexor digitorum longus muscle (crossing of plantar tendons)
20 Long plantar ligament

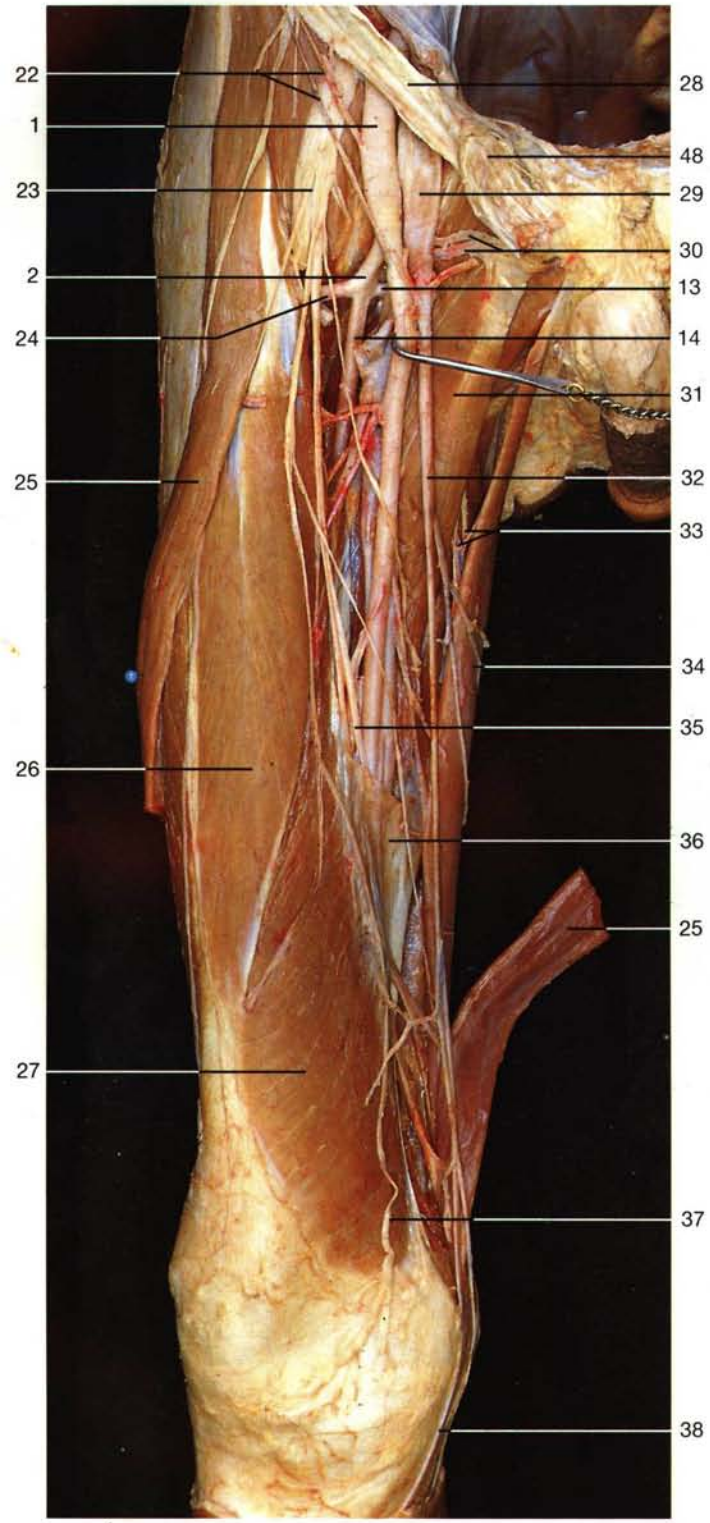

Main arteries and nerves of right thigh (anterior aspect). Sartorius muscle has been divided and reflected. The femoral vein has been partly removed to show the deep femoral artery. Notice: the vessels enter the adductor canal to reach the popliteal fossa.

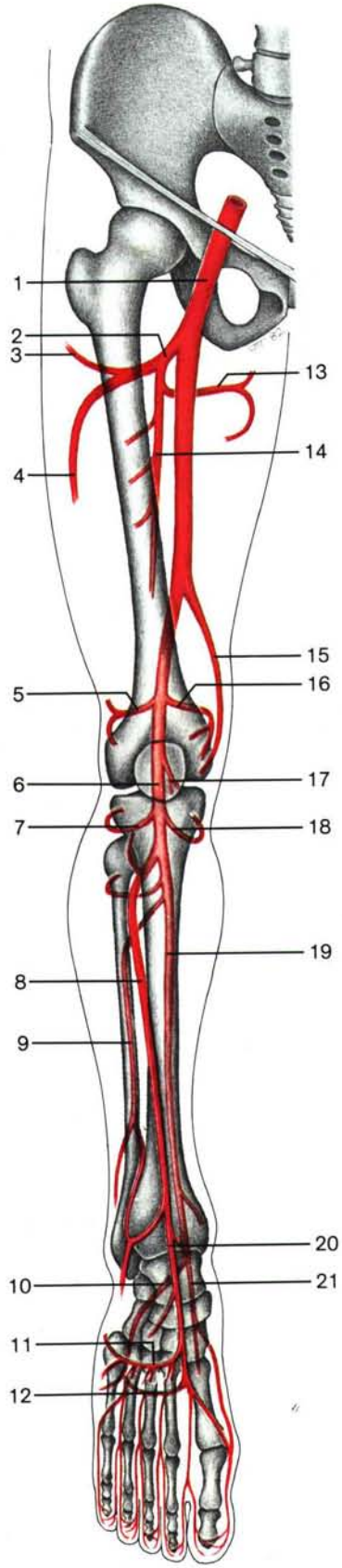

Main arteries of lower extremity, right side (ventral aspect). (Schematic drawing.)

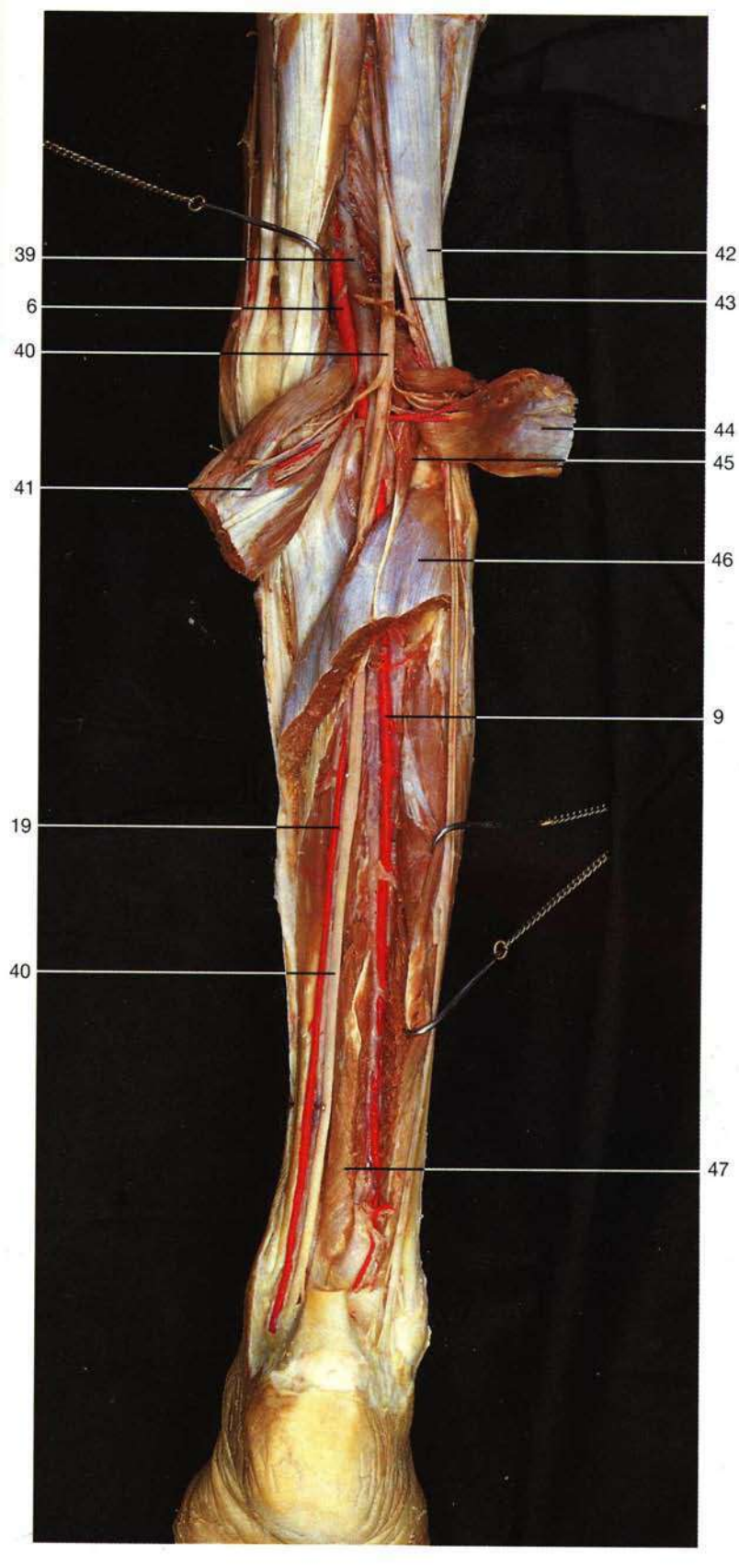

1 **Femoral artery**
2 **Profunda femoris artery**
3 Ascending branch of lateral circumflex femoral artery
4 Descending branch of lateral circumflex femoral artery
5 Lateral superior genicular artery
6 Popliteal artery
7 Lateral inferior genicular artery
8 **Anterior tibial artery**
9 **Peroneal artery**
10 **Lateral plantar artery**
11 Arcuate artery with dorsal metatarsal arteries
12 **Plantar arch** with plantar metatarsal arteries
13 Medial circumflex femoral artery
14 Profunda femoris artery with perforating arteries
15 Descending genicular artery
16 Medial superior genicular artery
17 Middle genicular artery
18 Medial inferior genicular artery
19 **Posterior tibial artery**
20 **Dorsalis pedis artery**
21 **Medial plantar artery**
22 Superficial and deep circumflex iliac arteries
23 **Femoral nerve**
24 Lateral circumflex femoral artery
25 Sartorius muscle (cut and reflected)
26 Rectus femoris muscle
27 Vastus medialis muscle
28 Inguinal ligament
29 **Femoral vein** (cut)
30 External pudendal artery and vein
31 Adductor longus muscle
32 Great saphenous vein
33 Obturator artery and nerve
34 Gracilis muscle
35 Saphenous nerve
36 Tendinous wall of adductor canal
37 Anterior cutaneous branch of femoral nerve
38 Infrapatellar branch of saphenous nerve
39 Popliteal vein
40 **Tibial nerve**
41 Medial head of gastrocnemius muscle
42 Biceps femoris muscle
43 Common peroneal nerve
44 Lateral head of gastrocnemius muscle
45 Plantaris muscle
43 Soleus muscle
47 Flexor hallucis longus muscle
48 Spermatic cord

Arteries of the right leg (posterior aspect).

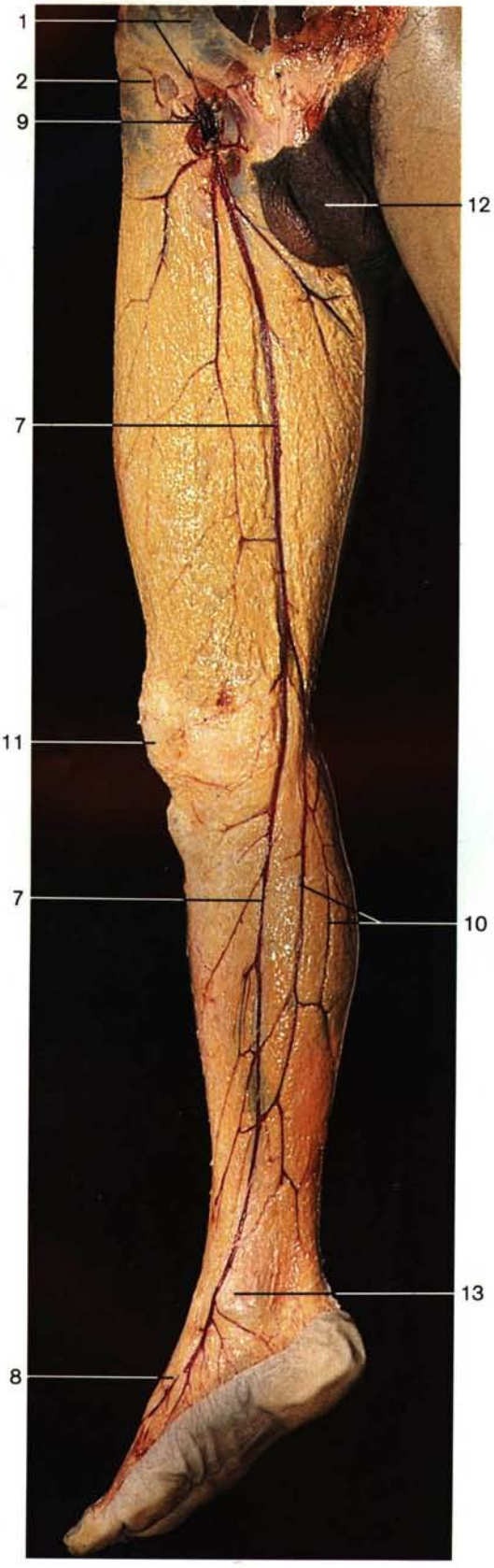

Superficial veins of lower limb, right side (medial anterior aspect). The veins have been injected with red solution.

Main veins of lower limb, right side (anterior aspect). (Schematic drawing.)

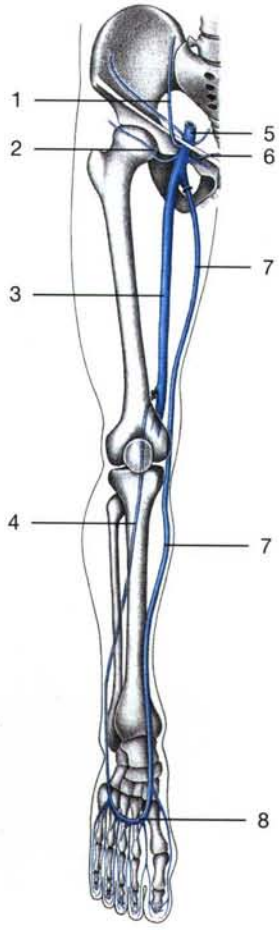

1 Superficial epigastric vein
2 Superficial circumflex iliac vein
3 Femoral vein
4 **Small saphenous vein**
5 External iliac vein
6 External pudendal vein
7 **Great saphenous vein**
8 Dorsal venous arch
9 Saphenous opening with femoral vein
10 Venous anastomoses of small saphenous vein with great saphenous vein
11 Patella
12 Penis
13 Medial malleolus
14 Popliteal fossa
15 Perforating veins
16 Lateral malleolus
17 Dorsal digital veins of foot
18 Dorsal venous arch of foot
19 Dorsal metatarsal veins of foot
20 Anterior tibial artery and veins
21 Tibia
22 **Posterior tibial artery and veins**
23 Fibula
24 Peroneal artery and vein
25 Deep layer of crural fascia
26 Superficial layer of crural fascia
27 Perforating veins I-III (of Cockett)
28 **Tibial nerve**
29 **Arcuate vein**
30 **Saphenous nerve**
31 Medial dorsal cutaneous nerve (branch of superficial peroneal nerve)
32 Posterior tibial vein

▷

Medial malleolar region. Dissection of tibial nerve, posterior tibial vessels, and great saphenous vein (veins injected with blue resin).

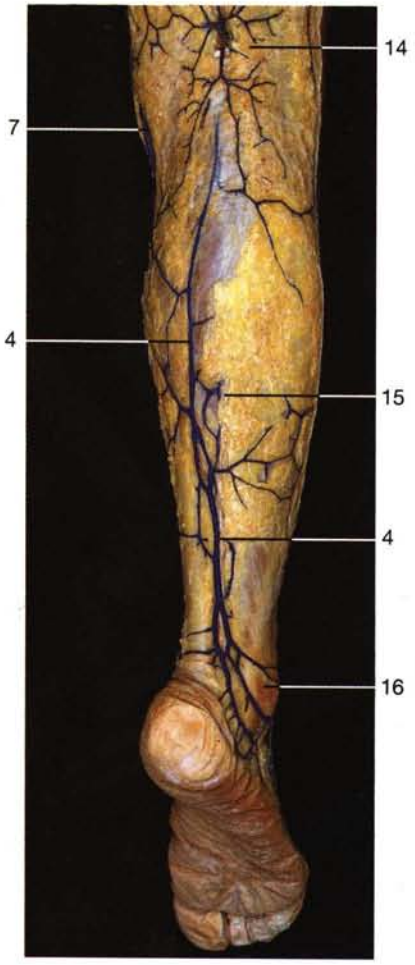

Superficial veins of leg (posterior aspect; injected with blue resin).

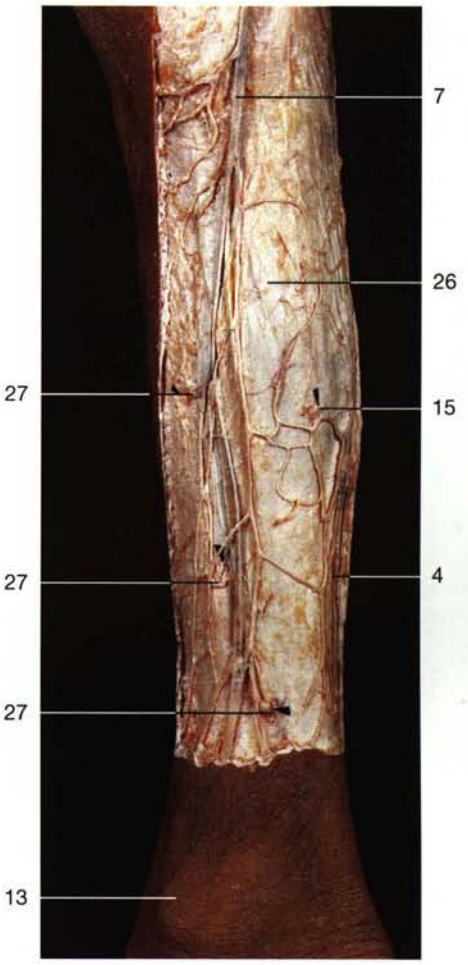

Superficial veins of leg. The perforating veins of Cockett have been dissected.

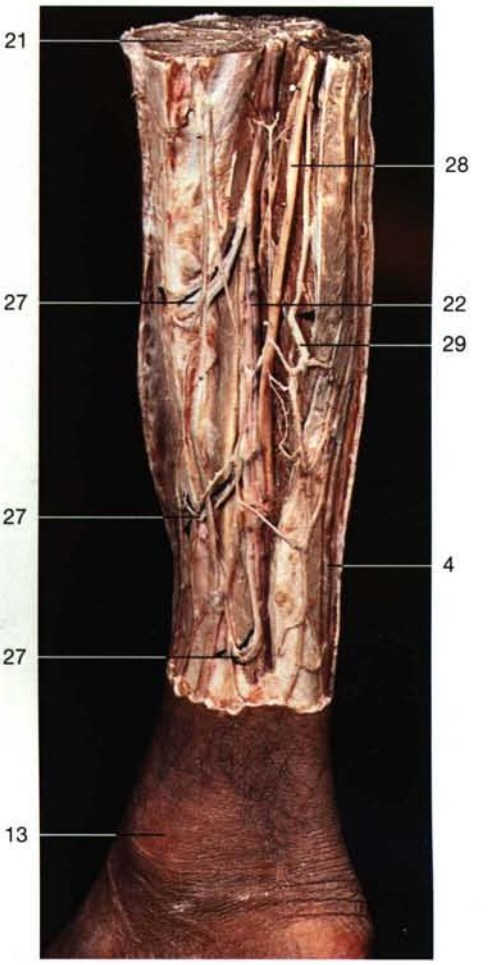

Veins of leg. The anastomoses between superficial and deeper veins are dissected.

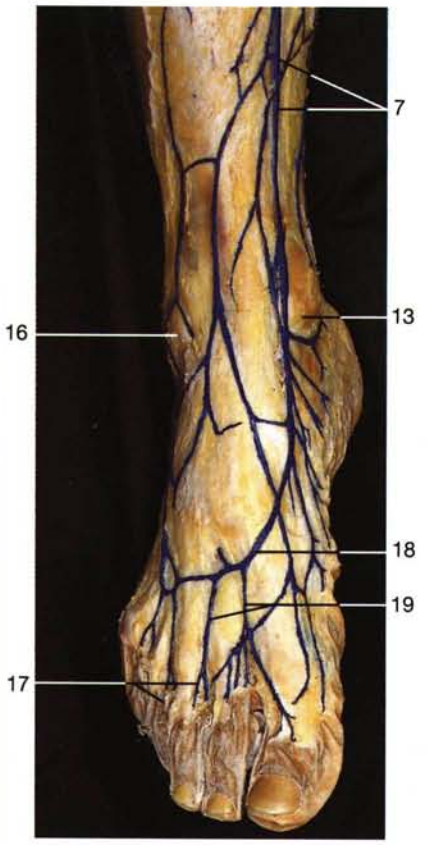

Anastomoses between superficial and deep veins of the leg (after Aigner). (Schematic drawing.) Arrows: directions of blood flow. ▷

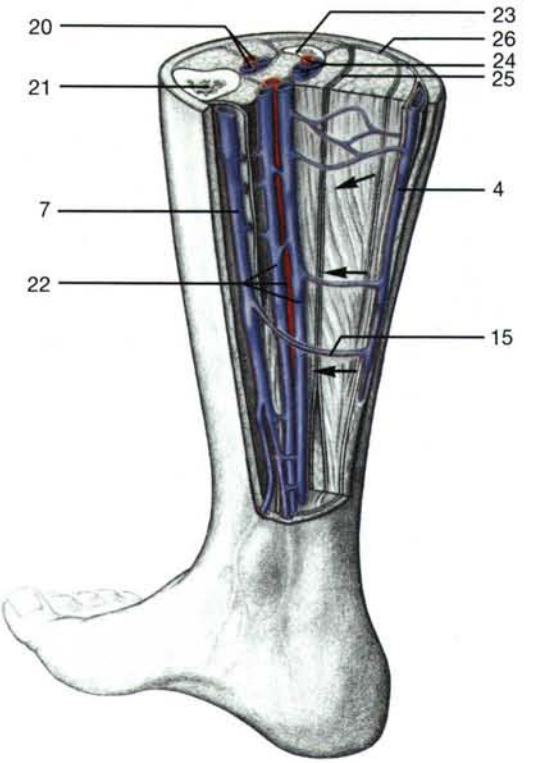

◁

Superficial veins on dorsum of foot (injected with blue resin).

1 Transversus abdominis muscle
2 Iliohypogastric nerve
3 Ilioinguinal nerve
4 **Femoral nerve**
5 Lateral femoral cutaneous nerve
6 Obturator nerve
7 Obturator internus muscle
8 Pubic bone (cut edge)
9 Levator ani muscle (remnant)
10 Dorsal nerve of penis
11 Posterior scrotal nerves
12 Adductor longus muscle
13 Gracilis muscle
14 Body of 4th lumbar vertebra
15 Cauda equina
16 Intervertebral disc
17 Sacral promontory
18 Sympathetic trunk
19 Sacrum
20 **Lumbosacral trunk**
21 **Sciatic plexus**
22 Coccyx
23 Sacrospinous ligament
24 **Pudendal nerve**
25 Inferior rectal nerves
26 Perineal nerves
27 Subcutaneous fat tissue of gluteal region

Lumbosacral plexus in situ, right side (medial aspect).
Pelvic organs with peritoneum and part of the levator ani muscle have been removed.

1 Subcostal nerve
2 Iliohypogastric nerve
3 Ilioinguinal nerve
4 Lateral femoral cutaneous nerve
5 Genitofemoral nerve
6 **Pudendal nerve**
7 **Femoral nerve**
8 Obturator nerve
9 **Sciatic nerve**
10 Lumbar plexus (L_1–L_4) } lumbosacral plexus
11 Sacral plexus (L_4–S_4) } lumbosacral plexus
12 "Pudendal" plexus (S_2–S_4) } lumbosacral plexus
13 Inferior cluneal nerve
14 Posterior femoral cutaneous nerve
15 Common peroneal nerve
16 **Tibial nerve**
17 Lateral sural cutaneous nerve
18 Medial and lateral plantar nerves
19 Saphenous nerve
20 Infrapatellar branch of saphenous nerve
21 **Deep peroneal nerve**
22 **Superficial peroneal nerve**

Nerves of lower limb, right side (lateral aspect). (Schematic drawing.)

Main branches of lumbosacral plexus (ventral aspect). (Schematic drawing.)

Spinal cord with intercostal nerves. Inferior thoracic region (anterior aspect). Anterior portion of thoracic vertebrae removed, dural sheath opened and spinal cord slightly reflected to the right to display the dorsal and ventral roots.

1 Dura mater
2 **Spinal cord**
3 Costotransverse ligament
4 Innermost intercostal muscle
5 Vertebral arches (cut surfaces)
6 Eleventh rib
7 **Intercostal nerve**
8 Collateral branch of intercostal nerve
9 Intercostal nerve (entering the intermuscular interval)
10 Anterior root filaments
11 Spinal (dorsal root) ganglion
12 Posterior root filaments
13 Arachnoid mater and denticulate ligament
14 Anterior spinal artery

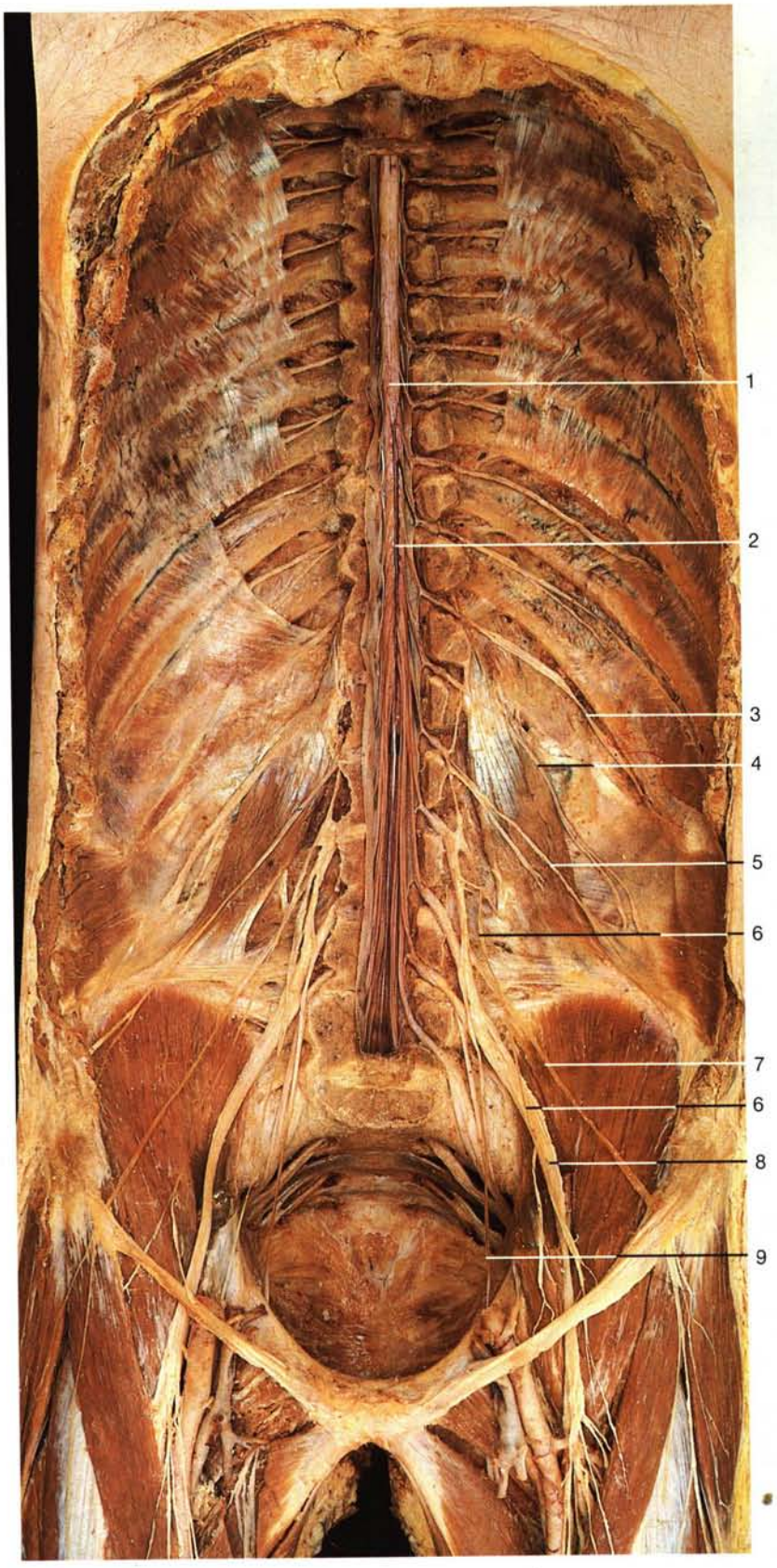

Spinal cord and lumbar plexus in situ (anterior aspect).

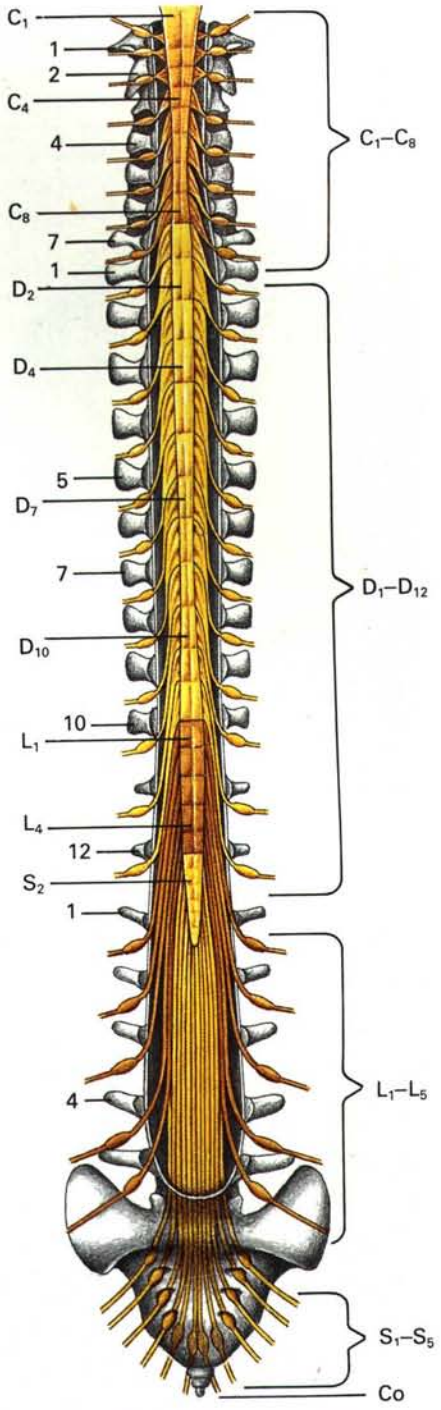

Organization of spinal cord segments in relation to the vertebral column (anterior aspect). C = cervical; D = thoracic; L = lumbar; S = sacral segments; Co = coccygeal bone.

1 Conus medullaris
2 Filum terminale
3 Subcostal nerve
4 Iliohypogastric nerve
5 Ilioinguinal nerve
6 Genitofemoral nerve
7 Lateral femoral cutaneous nerve
8 Femoral nerve
9 Obturator nerve

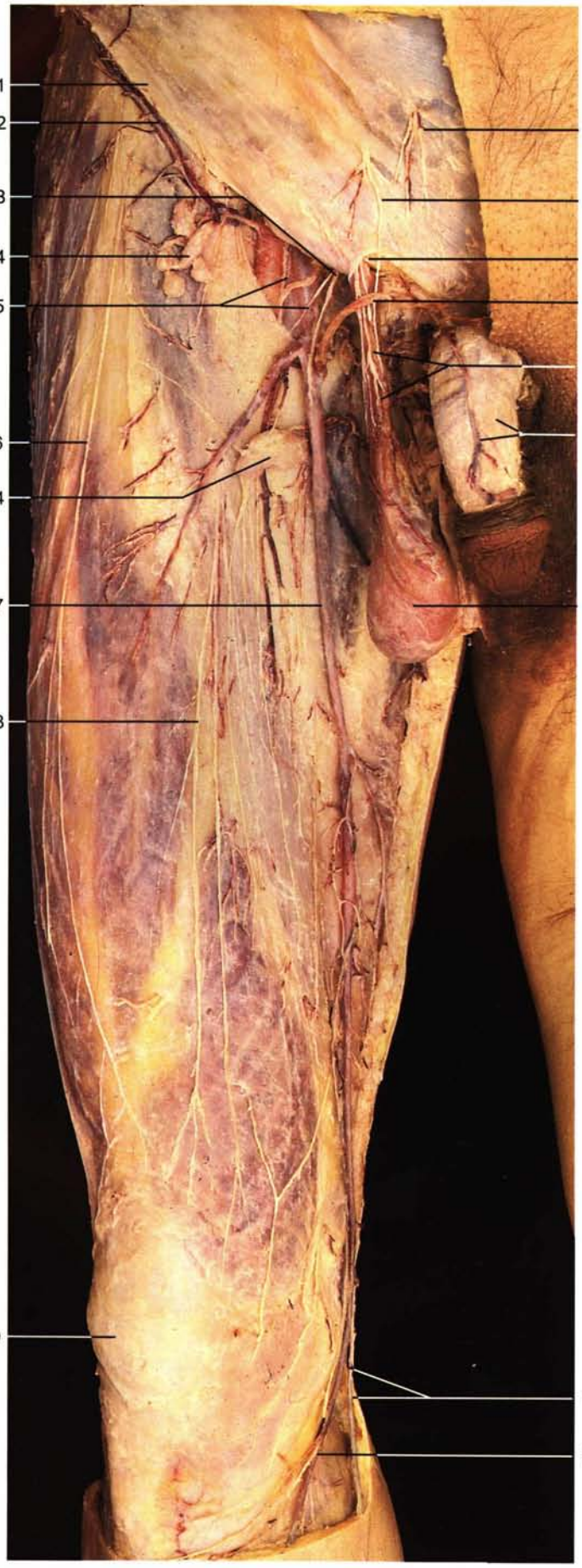

Cutaneous nerves and veins of thigh (anterior aspect).

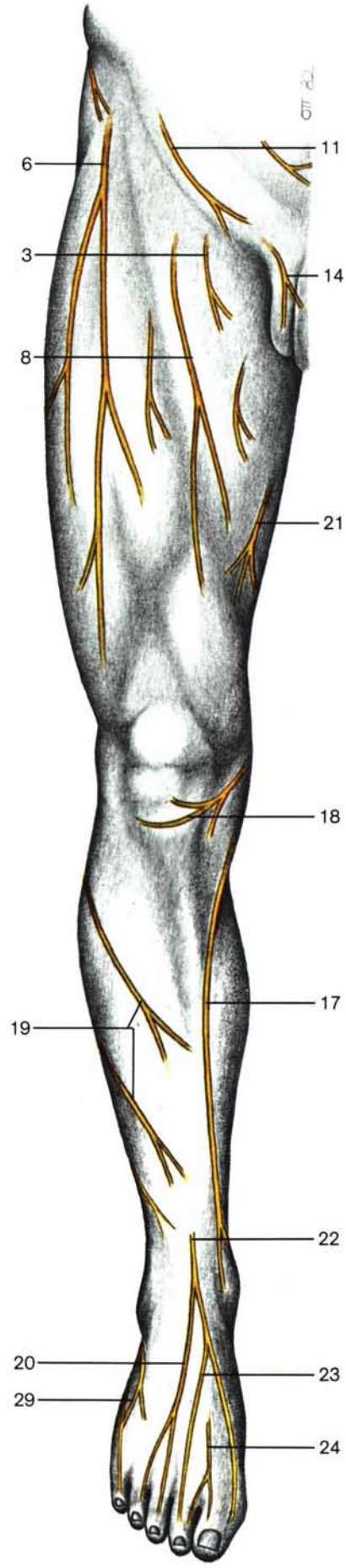

Cutaneous nerves of lower limb (anterior aspect). (Schematic drawing.)

1 Inguinal ligament
2 Superficial circumflex iliac vein
3 Femoral branch of genitofemoral nerve
4 Superficial inguinal lymph nodes
5 Saphenous opening with femoral artery and vein
6 Lateral femoral cutaneous nerve
7 Great saphenous vein
8 Anterior cutaneous branches of femoral nerve
9 Patella
10 Terminal branches of subcostal nerve
11 Terminal branches of iliohypogastric nerve
12 Superficial inguinal ring
13 External pudendal vein
14 Spermatic cord with genital branch of genitofemoral nerve
15 Penis with superficial dorsal vein of penis
16 Testis and its coverings
17 Saphenous nerve
18 Infrapatellar branch of saphenous nerve
19 Lateral sural cutaneous nerves
20 Intermediate dorsal cutaneous branch of superficial peroneal nerve
21 Cutaneous branch of obturator nerve
22 Superficial peroneal nerve
23 Medial dorsal cutaneous branch of superficial peroneal nerve
24 Deep peroneal nerve
25 **Femoral nerve**
26 **Femoral artery**
27 Superficial epigastric vein
28 **Femoral vein**
29 Lateral dorsal cutaneous branch of sural nerve
30 **Inguinal nodes** (enlarged)
31 Lympathic vessels
32 Sartorius muscle

Cutaneous nerves and veins of thigh (anterior aspect). The fascia lata and fasciae of the thigh muscles have been removed.

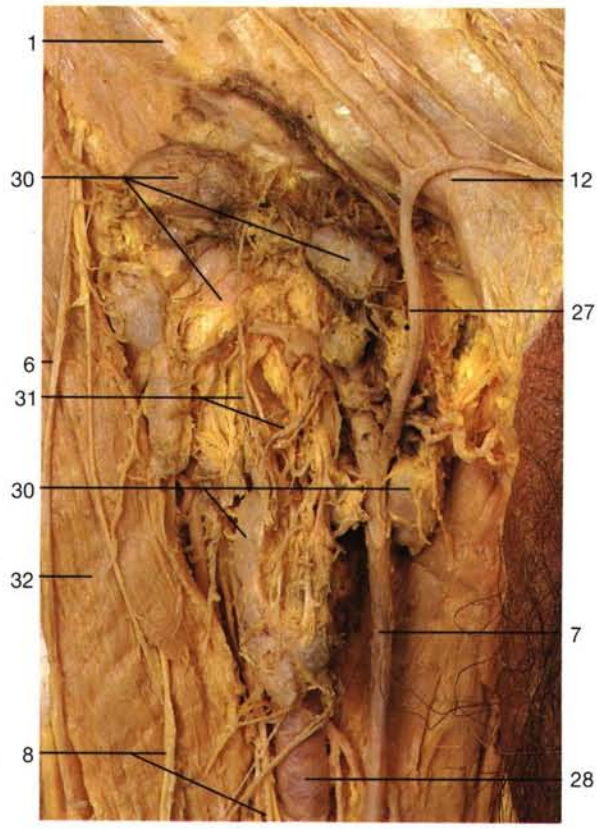

Inguinal nodes with lymphatic vessels (anterior aspect).

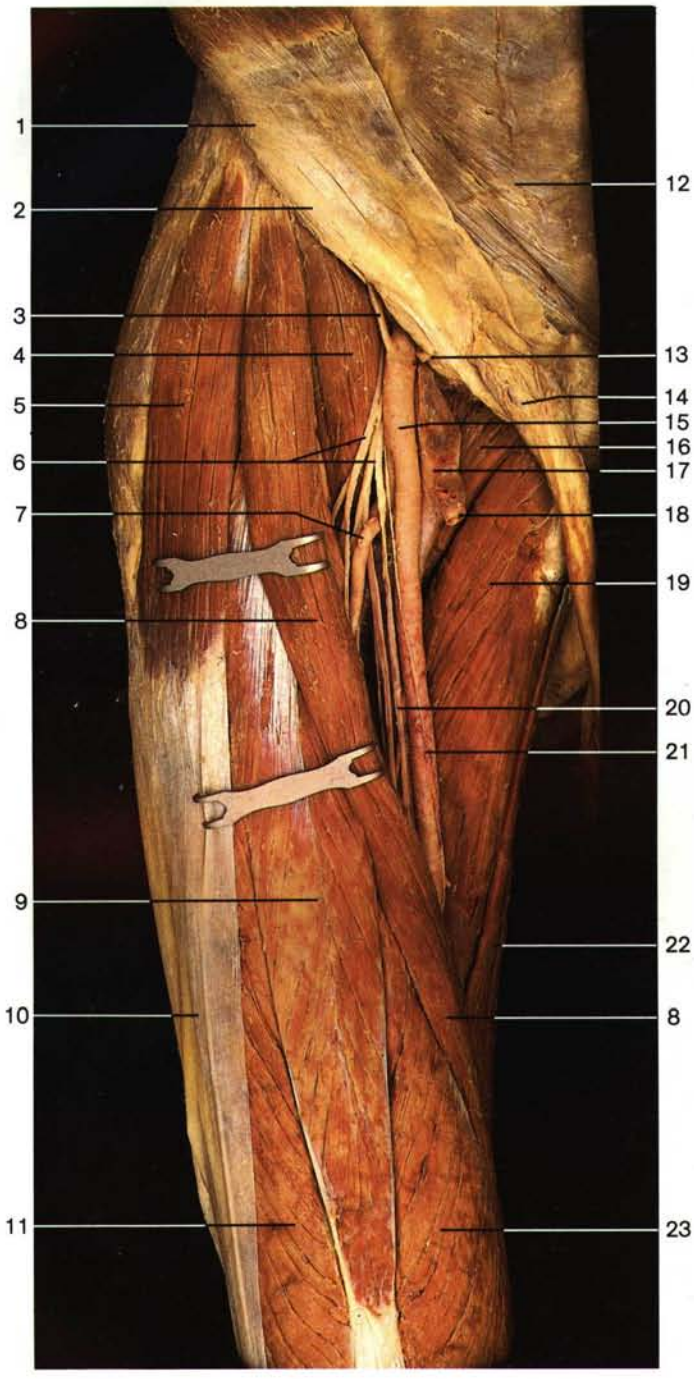

Anterior region of right thigh (anterior aspect).
The fascia lata has been removed, and the sartorius muscle has been slightly reflected.

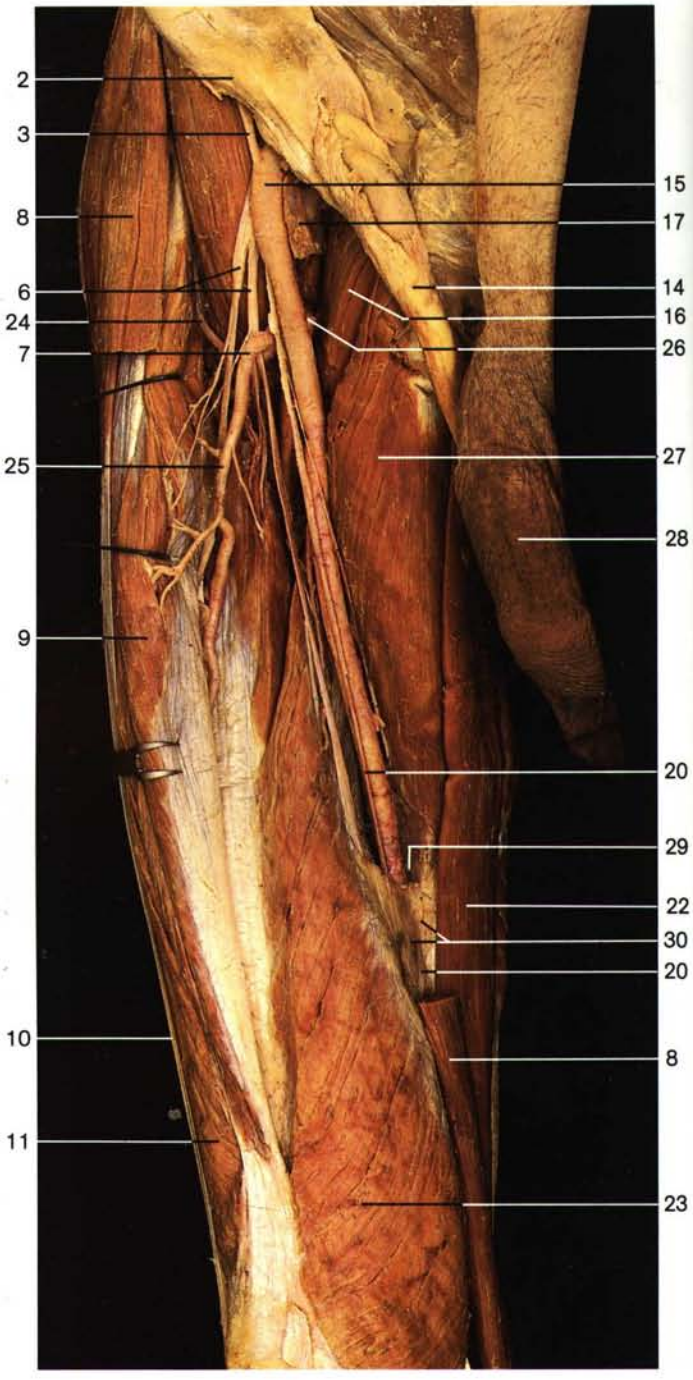

Anterior region of right thigh (anterior aspect).
The fascia lata has been removed, and the sartorius muscle has been divided.

1 Anterior superior iliac spine
2 Inguinal ligament
3 **Deep circumflex iliac artery**
4 Iliopsoas muscle
5 Tensor fasciae latae muscle
6 **Femoral nerve**
7 **Lateral circumflex femoral artery**
8 Sartorius muscle
9 Rectus femoris muscle
10 Iliotibial tract
11 Vastus lateralis muscle
12 Anterior sheath of rectus abdominis muscle
13 Inferior epigastric artery
14 Spermatic cord
15 **Femoral artery**
16 Pectineus muscle
17 **Femoral vein**
18 Great saphenous vein (divided)
19 Adductor longus muscle
20 **Saphenous nerve**
21 Muscular branch of femoral nerve
22 Gracilis muscle
23 Vastus medialis muscle
24 Ascending branch of lateral circumflex femoral artery
25 Descending branch of lateral circumflex femoral artery
26 Medial circumflex femoral artery
27 Adductor longus muscle
28 Penis
29 Entrance to adductor canal
30 Vastoadductory lamina of fascia beneath sartorius muscle

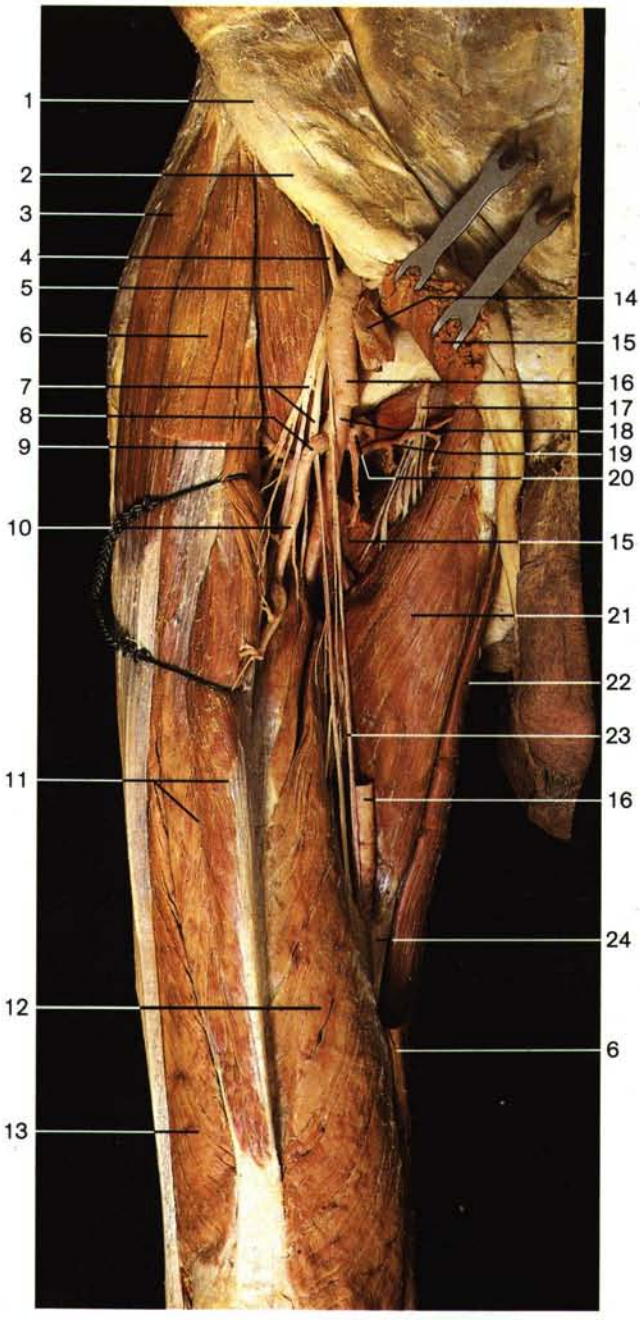

Anterior region of right thigh (anterior aspect). The fascia lata has been removed. Sartorius muscle, pectineus muscle and femoral artery have been cut to display the deep femoral artery with its branches. The rectus femoris muscle has been slightly reflected.

1 Anterior superior iliac spine
2 Inguinal ligament
3 Tensor fasciae latae muscle
4 Deep circumflex iliac artery
5 Iliopsoas muscle
6 Sartorius muscle (cut)
7 **Femoral nerve**
8 **Lateral circumflex femoral artery**
9 Ascending branch of lateral circumflex femoral artery
10 Descending branch of lateral circumflex femoral artery
11 Rectus femoris muscle
12 Vastus medialis muscle
13 Vastus lateralis muscle
14 Femoral vein
15 Pectineus muscle (cut)
16 Femoral artery (cut)

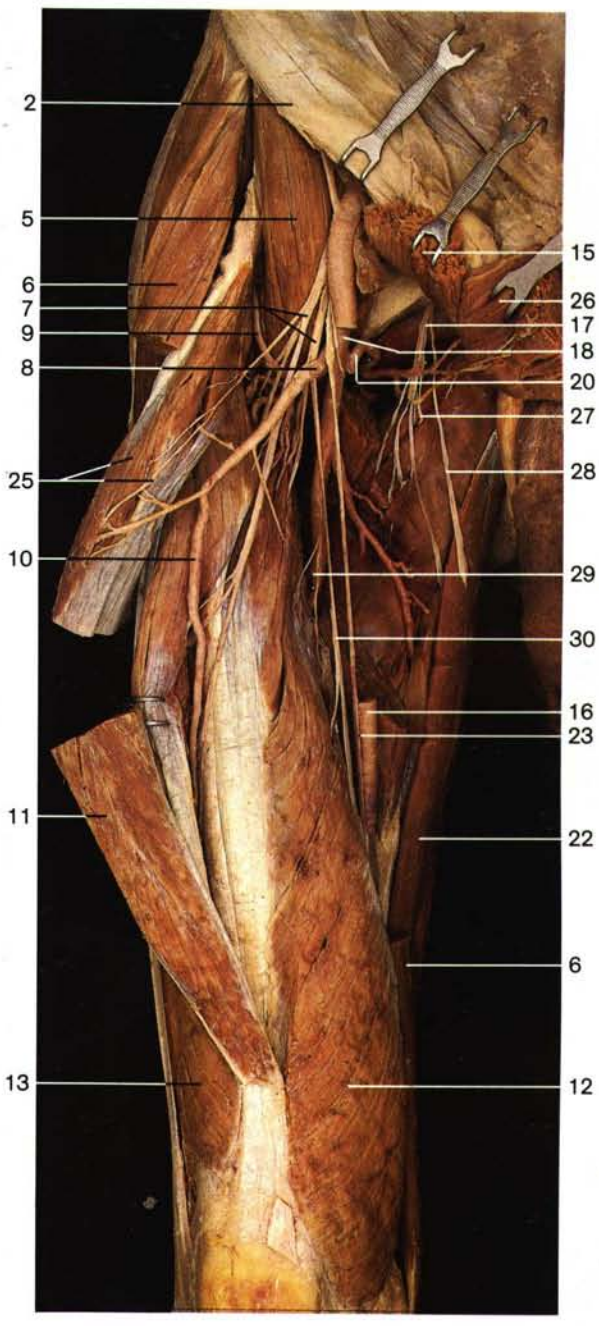

Anterior region of right thigh (anterior aspect). The sartorius, pectineus, adductor longus and rectus femoris muscles have been divided and reflected. The greater part of the femoral artery has been removed.

17 **Obturator nerve**
18 **Profunda femoris artery**
19 Ascending branch of medial circumflex femoral artery
20 **Medial circumflex femoral artery**
21 Adductor longus muscle
22 Gracilis muscle
23 Saphenous nerve
24 Distal part of vastoadductory lamina
25 Rectus femoris muscle with muscular branch of femoral nerve
26 Adductor longus muscle (divided)
27 Posterior branch of obturator nerve
28 Anterior branch of obturator nerve
29 Point at which perforating artery branches off from profunda femoris artery
30 Muscular branch to vastus medialis muscle

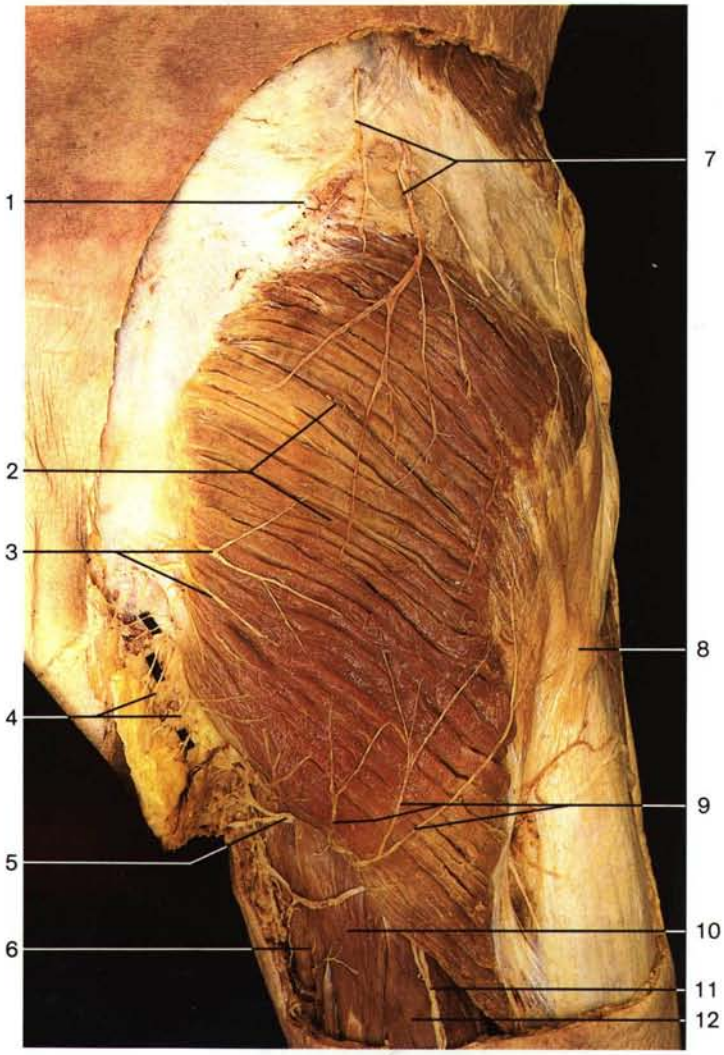

Gluteal region, right side (posterior aspect).

1 Iliac crest
2 Gluteus maximus muscle
3 **Middle cluneal nerves**
4 Anococcygeal nerves
5 Perineal branch of posterior femoral cutaneous nerve
6 Adductor magnus muscle
7 **Superior cluneal nerves**
8 Position of greater trochanter
9 **Inferior cluneal nerves**
10 Semitendinosus muscle
11 Posterior femoral cutaneous nerve
12 Long head of biceps femoris muscle

A **Suprapiriform foramen**
of greater sciatic foramen

Superior gluteal artery, vein and nerve

B **Infrapiriform foramen**
of greater sciatic foramen

Sciatic nerve
Inferior gluteal artery, vein and nerve
Posterior femoral cutaneous nerve
Internal pudendal artery and vein
Pudendal nerve

C **Lesser sciatic foramen**

Pudendal nerve
Internal pudendal artery and vein

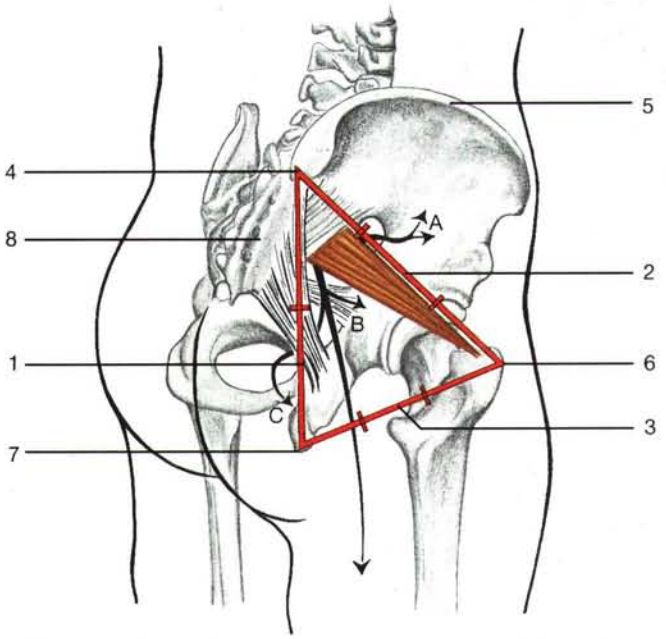

Gluteal region, right side (posterolateral aspect). Location of sciatic foramina in relation to the bones.
(Schematic drawing.)

Red lines

1 **Spine-tuber line.**
In the middle of this line the infrapiriform foramen is situated
2 **Spine-trochanter line.**
In the upper third the suprapiriform foramen is located
3 **Tuber-trochanter line.**
Between the middle and posterior third, the ischiadic nerve can be found

Other structures

4 Posterior superior iliac spine
5 Iliac crest
6 Greater trochanter
7 Ischial tuberosity
8 Sacrum

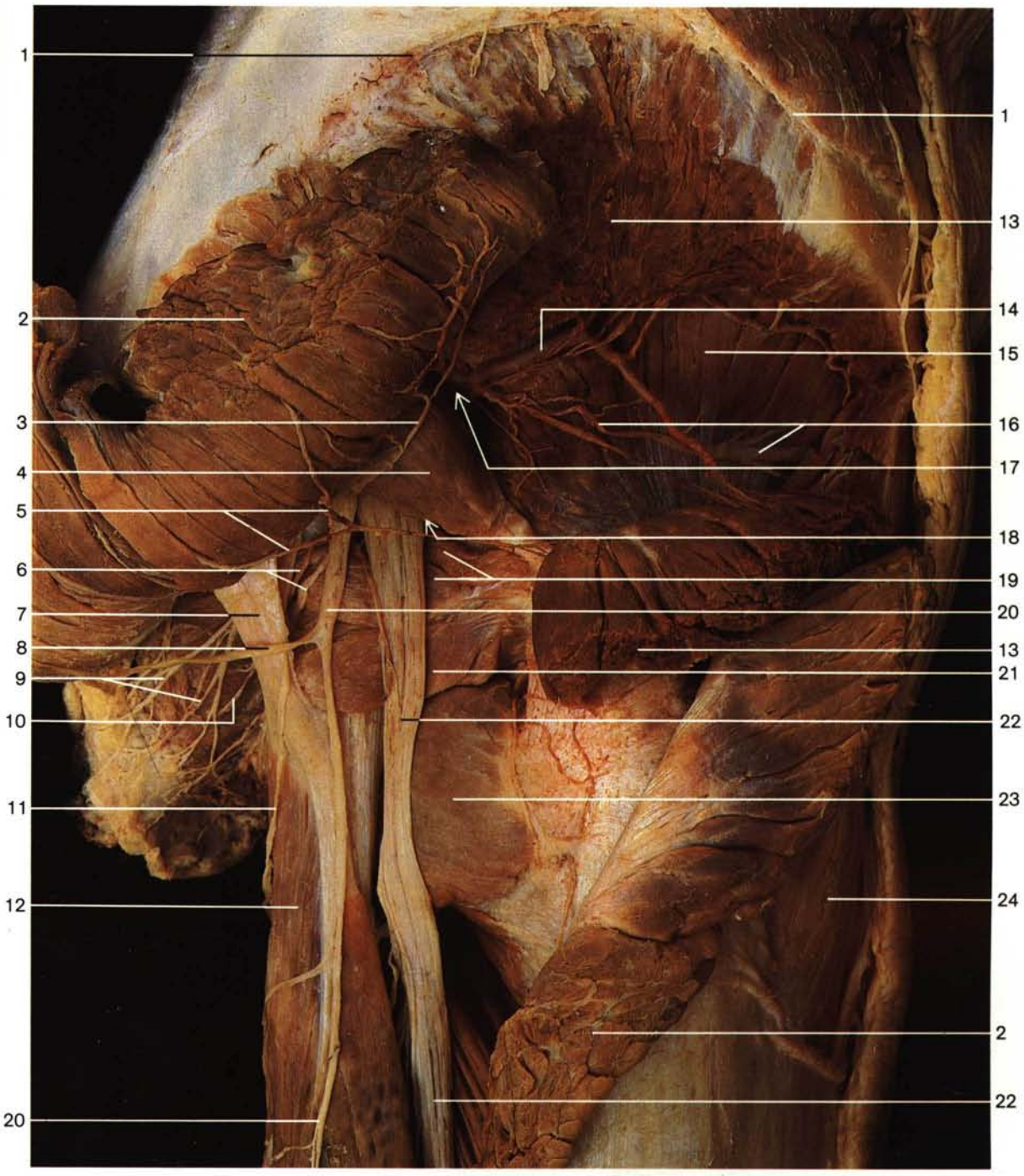

Gluteal region, right side (dorsal aspect). The gluteus maximus and gluteus medius muscles have been divided and reflected. Notice the position of the foramina above and below piriformis muscles and the lesser sciatic foramen.

1 Iliac crest
2 Gluteus maximus muscle (cut)
3 Inferior gluteal nerve
4 Piriformis muscle
5 Muscular branches of inferior gluteal artery
6 Pudendal nerve and internal pudendal artery within the **lesser sciatic foramen** (entrance to the pudendal canal)
7 Sacrotuberous ligament
8 Inferior cluneal nerve
9 Inferior rectal nerves
10 Inferior rectal arteries
11 Perforating cutaneous nerve
12 Long head of biceps femoris muscle
13 Gluteus medius muscle (cut)
14 Deep branch of superior gluteal artery
15 Gluteus minimus muscle
16 Superior gluteal nerve
17 **Suprapiriform foramen** } greater sciatic foramen
18 **Infrapiriform foramen** } greater sciatic foramen
19 Tendon of obturator internus and superior gemellus muscles
20 Posterior femoral cutaneous nerve
21 Inferior gemellus muscle
22 **Sciatic nerve**
23 Quadratus femoris muscle
24 Tensor fasciae latae muscle

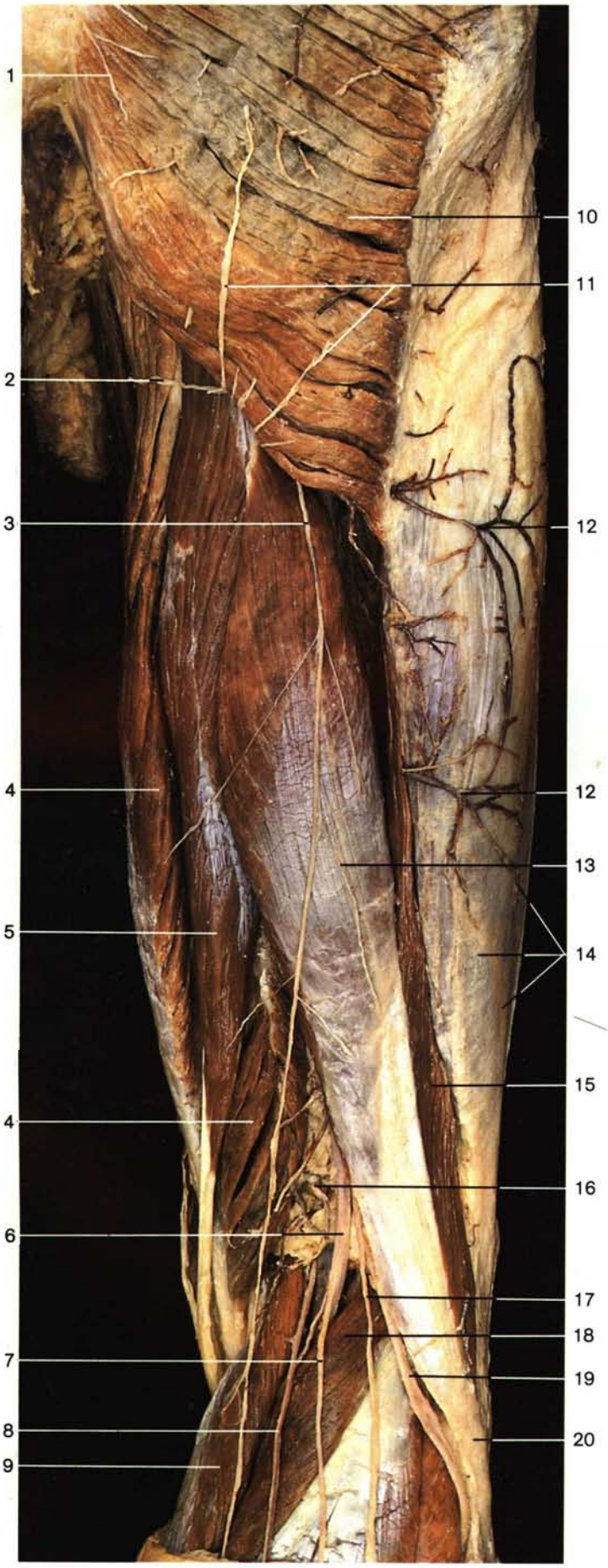

1 Middle cluneal nerves
2 Perineal branch of posterior femoral cutaneous nerve
3 **Posterior femoral cutaneous nerve**
4 Semimembranosus muscle
5 Semitendinosus muscle
6 **Tibial nerve**
7 Medial sural cutaneous nerve
8 **Small saphenous vein**
9 Medial head of gastrocnemius muscle
10 Gluteus maximus muscle
11 **Inferior cluneal nerve**
12 Cutaneous veins
13 Long head of biceps femoris muscle
14 Iliotibial tract
15 Short head of biceps femoris muscle
16 Popliteal fossa
17 Lateral sural cutaneous nerve
18 Lateral head of gastrocnemius muscle
19 **Common peroneal nerve**
20 Tendon of biceps femoris muscle
21 Inferior gluteal nerve
22 Sacrotuberous ligament
23 Inferior rectal branches of pudendal nerve
24 Anus
25 Gluteus medius muscle
26 Piriformis muscle
27 **Sciatic nerve**
28 Inferior gluteal artery
29 Gluteus maximus muscle (cut)
30 Quadratus femoris muscle
31 Sciatic nerve dividing into its two branches: the common peroneal nerve and the tibial nerve
32 Muscular branches of sciatic nerve to hamstring muscles
33 **Popliteal artery**
34 **Popliteal vein**
35 Small saphenous vein (cut)
36 Long head of biceps femoris muscle (cut)
37 **Superficial peroneal nerve**

Cutaneous nerves of thigh (posterior aspect).
The fascia lata and the fasciae of muscles have been removed.

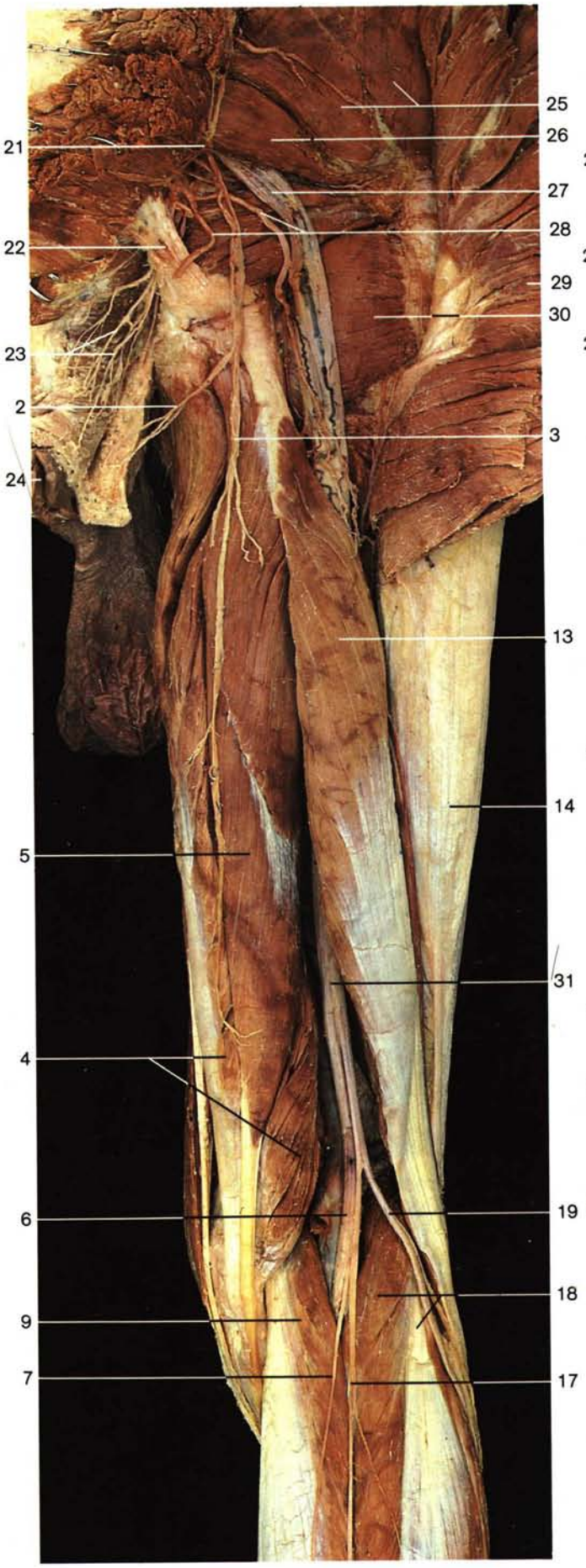

Posterior femoral region and gluteal region, right side (posterior aspect). The gluteus maximus muscle has been divided and reflected.

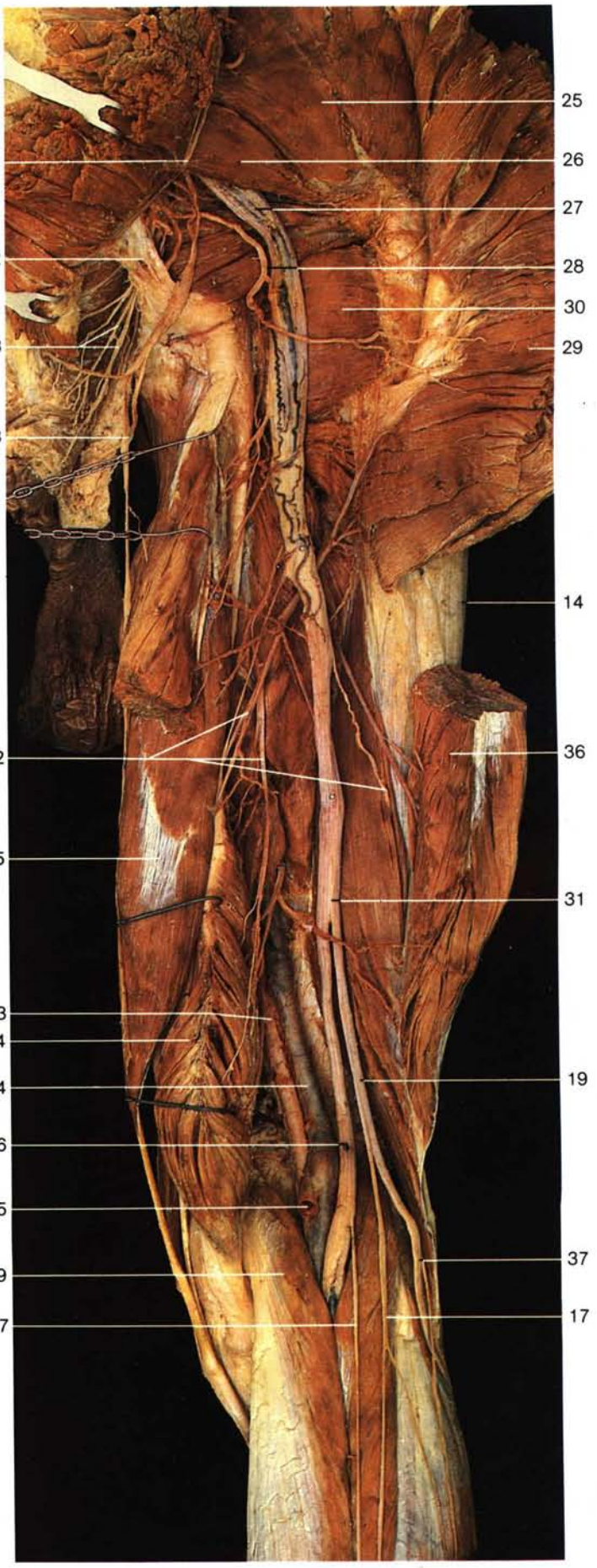

Posterior femoral region and gluteal region, right side (posterior aspect). The gluteus maximus muscle and the long head of biceps femoris muscle have been divided and reflected.

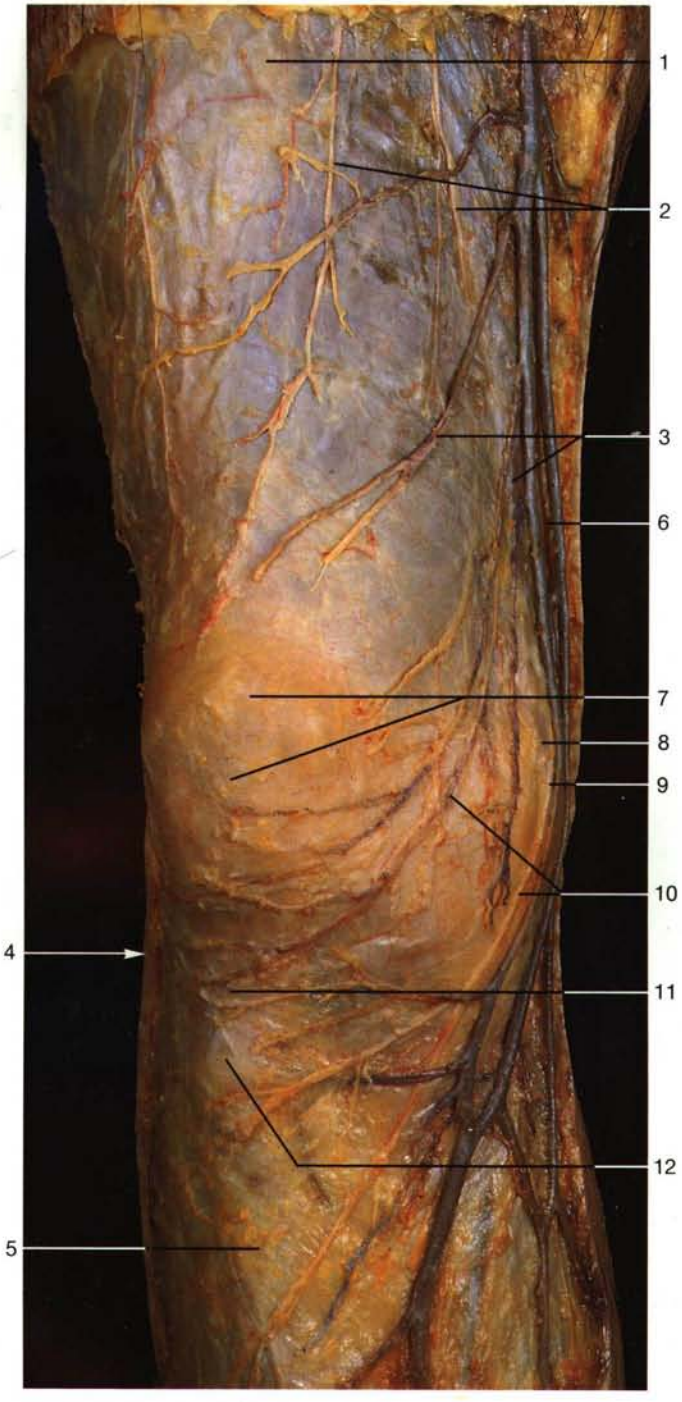

Anterior region of right knee; cutaneous nerves and veins (anterior aspect).

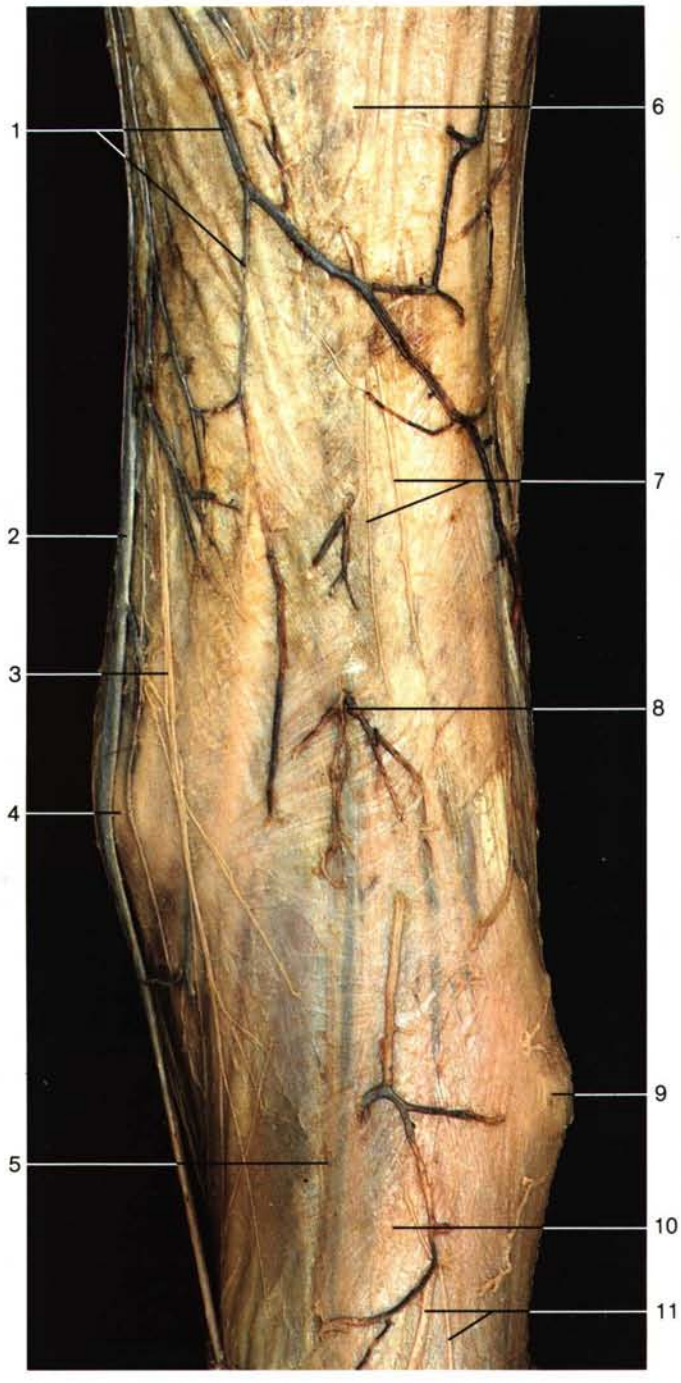

Posterior region of right kneee; cutaneous nerves and veins (posterior aspect).

1 Fascia lata
2 Terminal branches of anterior cutaneous branches of femoral nerve
3 Venous network around knee
4 Position of head of fibula
5 Superficial crural fascia
6 **Great saphenous vein**
7 Patella
8 Position of medial epicondyle of femur
9 Saphenous nerve
10 **Infrapatellar branches of saphenous nerve**
11 Patellar ligament
12 Position of tuberosity of tibia

1 Cutaneous veins (tributaries of great saphenous vein)
2 **Great saphenous vein**
3 Cutaneous branch of femoral nerve
4 Position of medial epicondyle of femur
5 Position of small saphenous vein
6 Fascia lata
7 Terminal branches of **posterior femoral cutaneous nerve**
8 Cutaneous veins of popliteal fossa
9 Position of head of fibula
10 Superficial layer of fascia cruris
11 Lateral sural cutaneous nerve

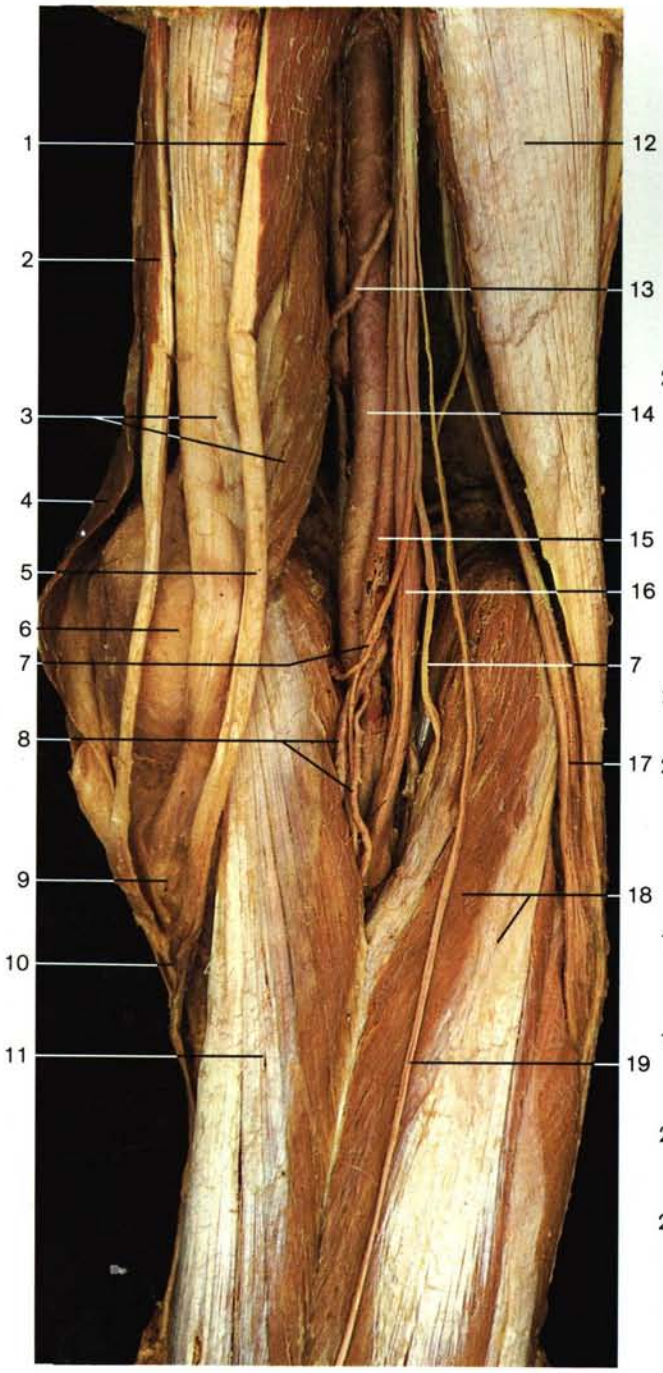

Right leg, posterior crural region (posterior aspect). The gastrocnemius muscle has been divided and reflected.

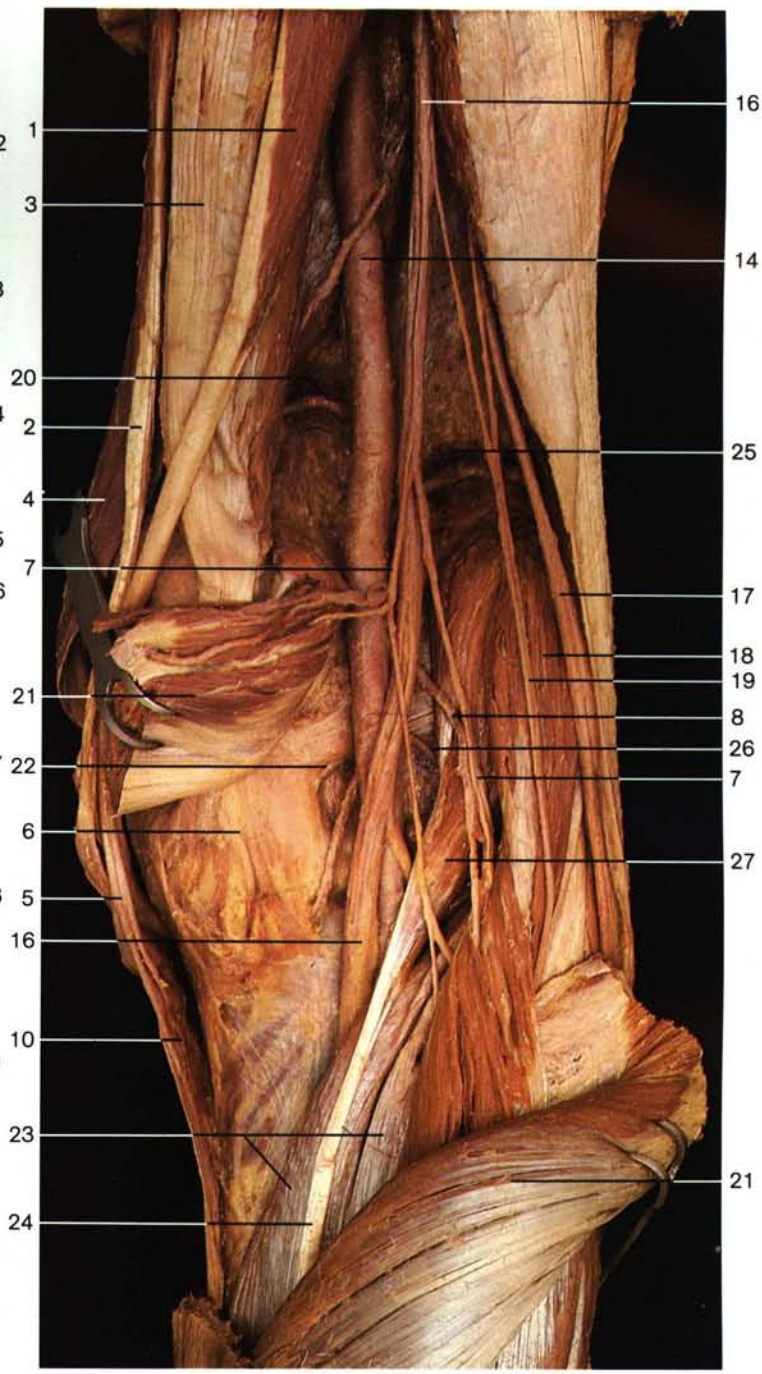

Right leg, posterior crural region, deep layer (posterior aspect). The gastrocnemius and the soleus muscles have been divided and reflected.

1 Semitendinosus muscle
2 Gracilis muscle
3 Semimembranosus muscle
4 Sartorius muscle
5 Tendon of semitendinosus muscle
6 Position of medial condyle of femur
7 Muscular branches of tibial nerve
8 Sural arteries and veins
9 Tendon of semimembranosus muscle
10 Common tendon of gracilis, semitendinosus and sartorius muscles
11 Medial head of gastrocnemius muscle
12 Biceps femoris muscle
13 Muscular branch of popliteal artery
14 **Popliteal artery**
15 **Popliteal vein**
16 **Tibial nerve**
17 **Common peroneal nerve**
18 Lateral head of gastrocnemius muscle
19 Medial sural cutaneous nerve
20 Medial superior genicular artery
21 Medial head of gastrocnemius muscle (cut and reflected)
22 Medial inferior genicular artery
23 Soleus muscle
24 Tendon of plantaris muscle
25 Lateral superior genicular artery
26 Lateral inferior genicular artery
27 Plantaris muscle

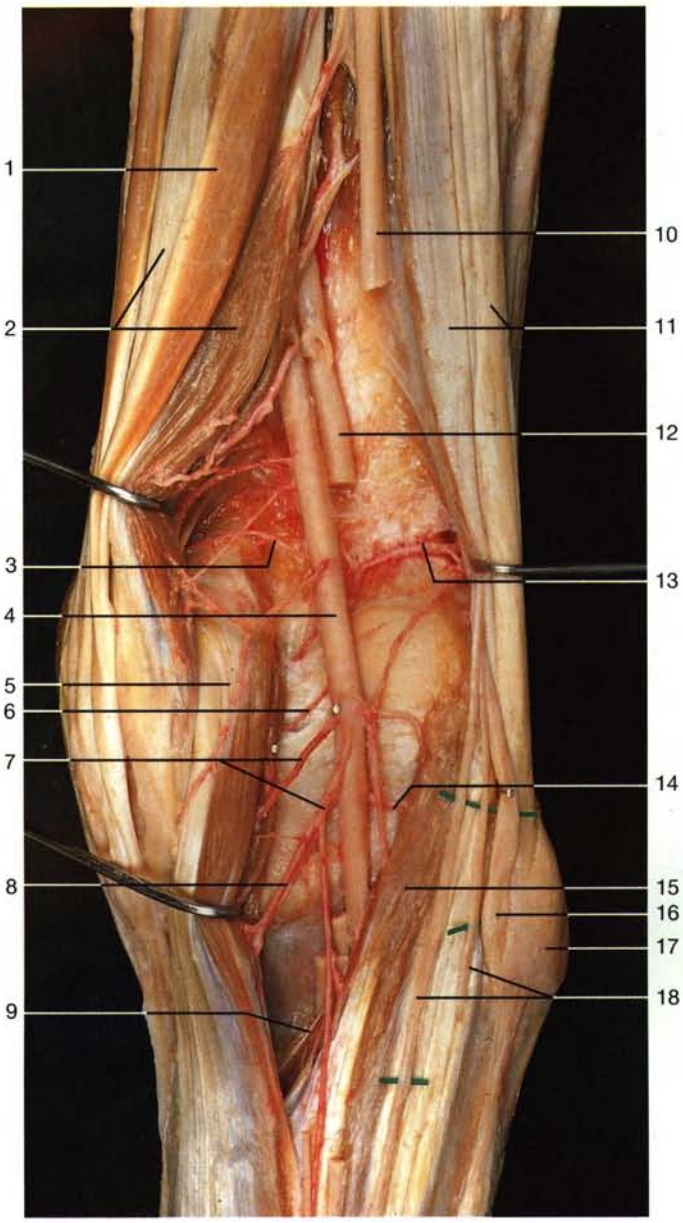

Right leg, popliteal fossa, deep layer (posterior aspect). The muscles have been reflected to display the genicular arteries.

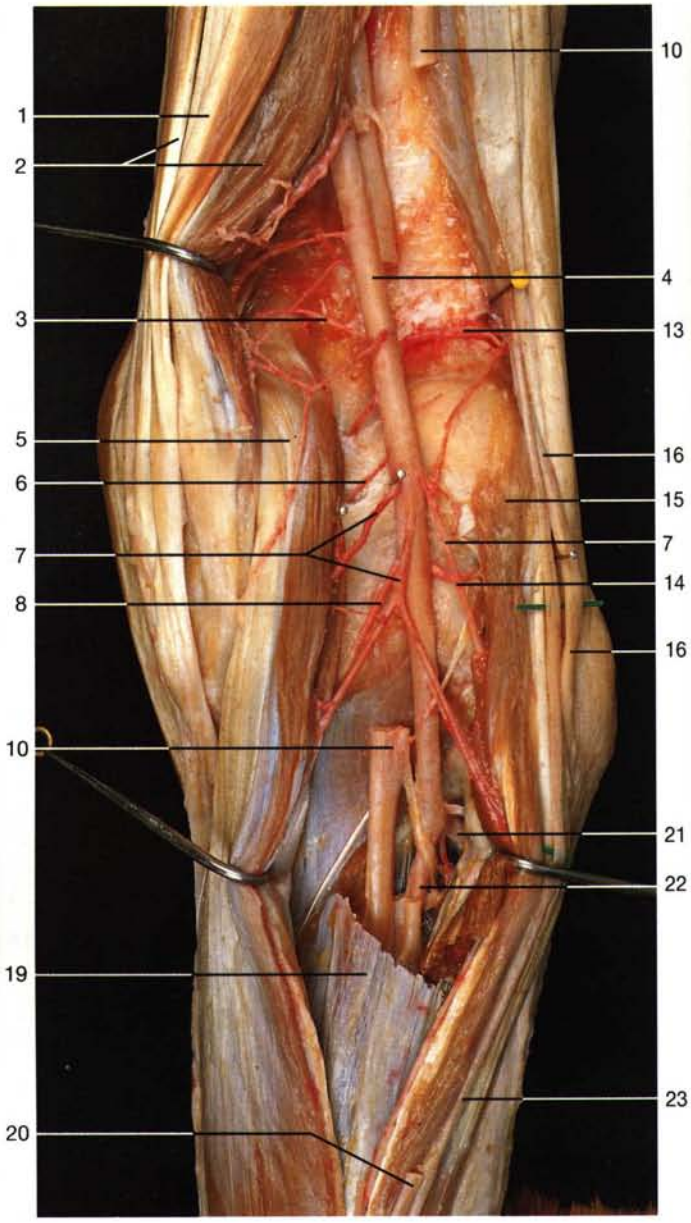

Right leg, popliteal fossa, deepest layer (posterior aspect). Tibial nerve and popliteal vein have been partly removed and a portion of the soleus muscle was cut away to display the anterior tibial artery.

1 Semitendinosus muscle
2 Semimembranosus muscle
3 Medial superior genicular artery
4 **Popliteal artery**
5 Medial head of gastrocnemius muscle
6 Middle genicular artery
7 Muscular branches
8 Medial inferior genicular artery
9 Tendon of plantaris muscle
10 **Tibial nerve** (cut)
11 Biceps femoris muscle
12 Popliteal vein (cut)
13 Lateral superior genicular artery
14 Lateral inferior genicular artery
15 Lateral head of gastrocnemius muscle
16 **Common peroneal nerve**
17 Head of fibula
18 Lateral sural cutaneous nerves
19 Soleus muscle
20 Medial sural cutaneous nerve
21 **Anterior tibial artery**
22 **Posterior tibial artery**
23 Lateral sural cutaneous nerve

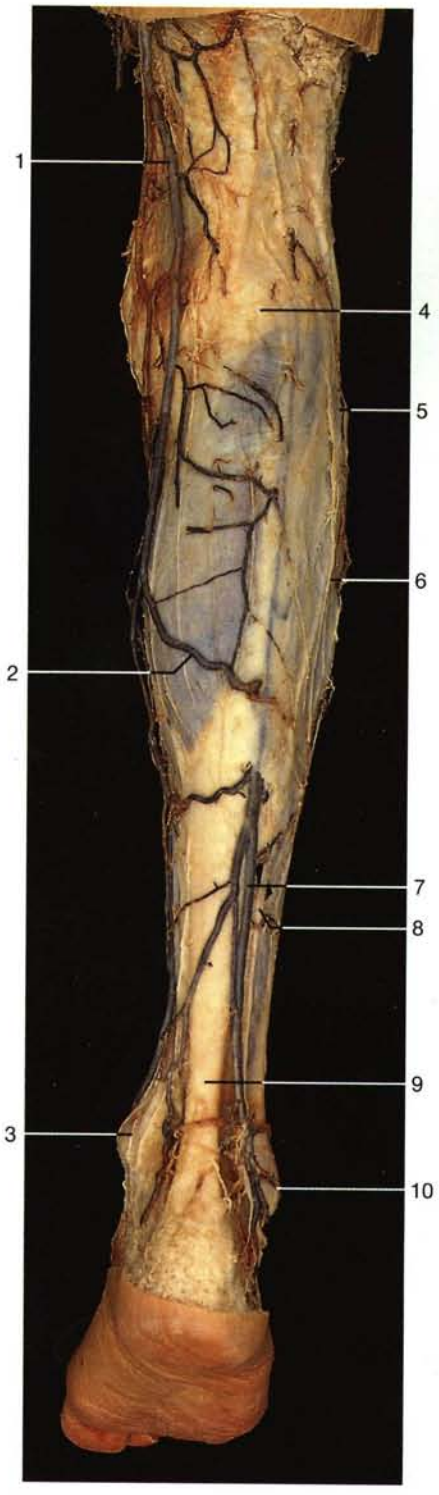

Right leg. Cutaneous veins and nerves (posterior aspect).

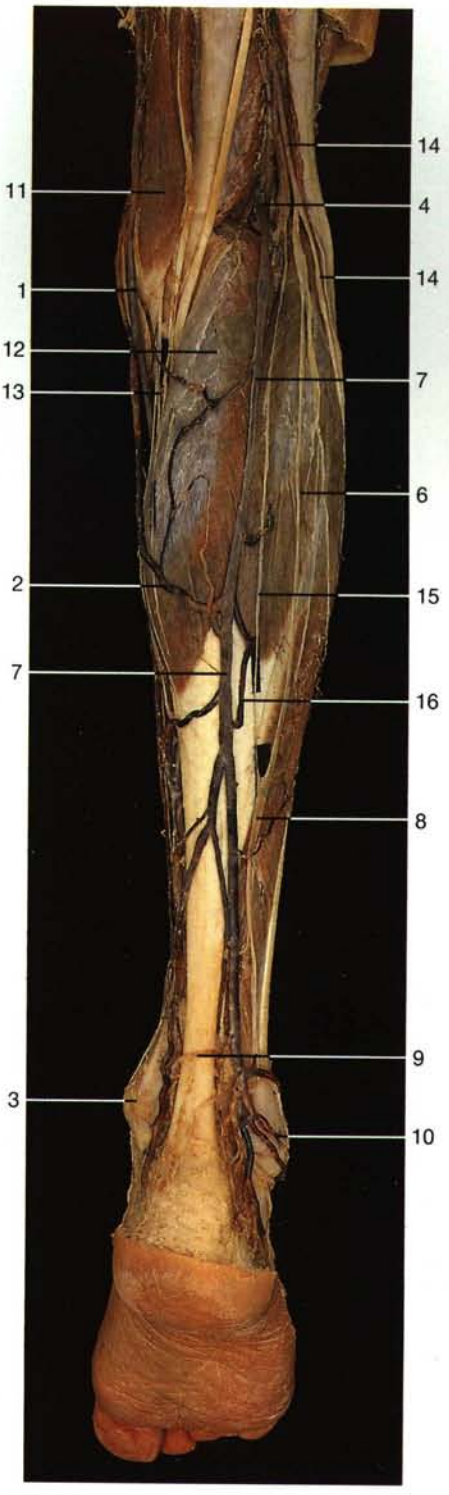

Right leg. Cutaneous nerves and veins (posterior aspect). The superficial layer of the crural fascia has been removed.

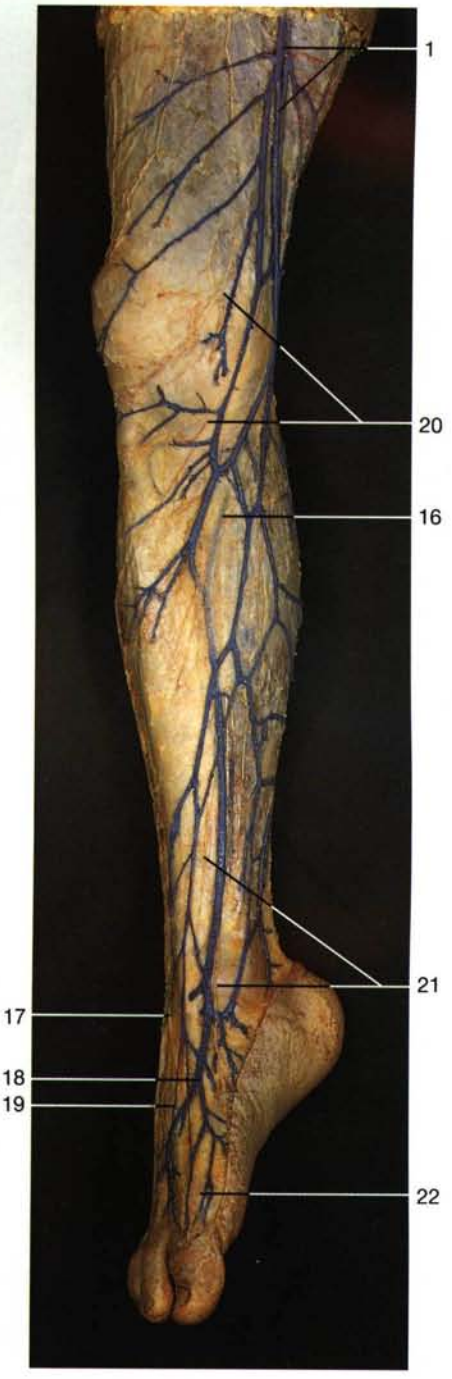

Cutaneous veins and nerves of the right leg (anterior-medial aspect; veins are colored).

1 **Great saphenous vein**
2 Venous anastomosis between small and great saphenous veins
3 Medial malleolus
4 Popliteal fossa
5 Position of head of fibula
6 Lateral sural cutaneous nerve
7 **Small saphenous vein**
8 **Sural nerve**
9 Calcaneal tendon
10 Lateral malleolus
11 Semitendinosus muscle
12 Medial head of gastrocnemius muscle
13 Saphenous nerve
14 **Common peroneal nerve**
15 Medial sural cutaneous nerve
16 Perforating veins
17 Superficial peroneal nerve
18 Dorsal venous arch
19 Intermediate dorsal cutaneous nerve
20 **Infrapatellar branches of saphenous nerve**
21 Terminal branches of saphenous nerve
22 Medial dorsal cutaneous nerve

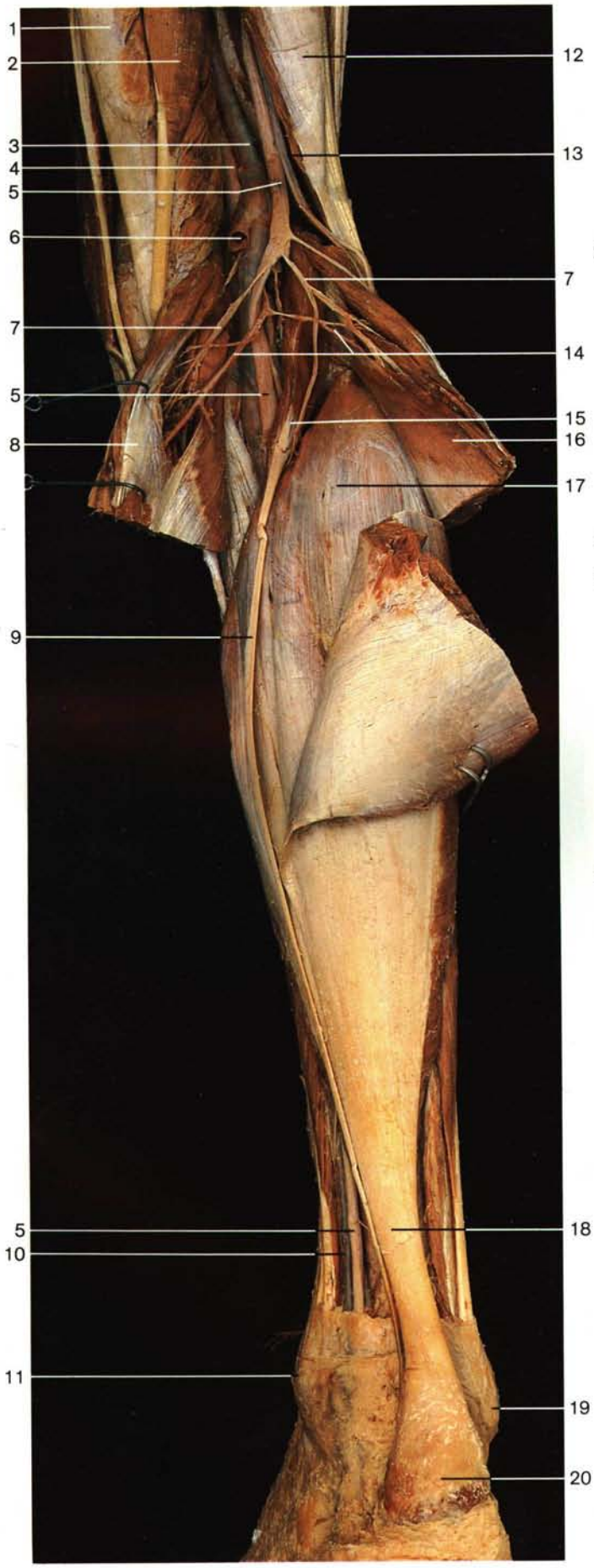

Right popliteal fossa, middle layer (posterior aspect). The cutaneous veins and nerves have been removed.

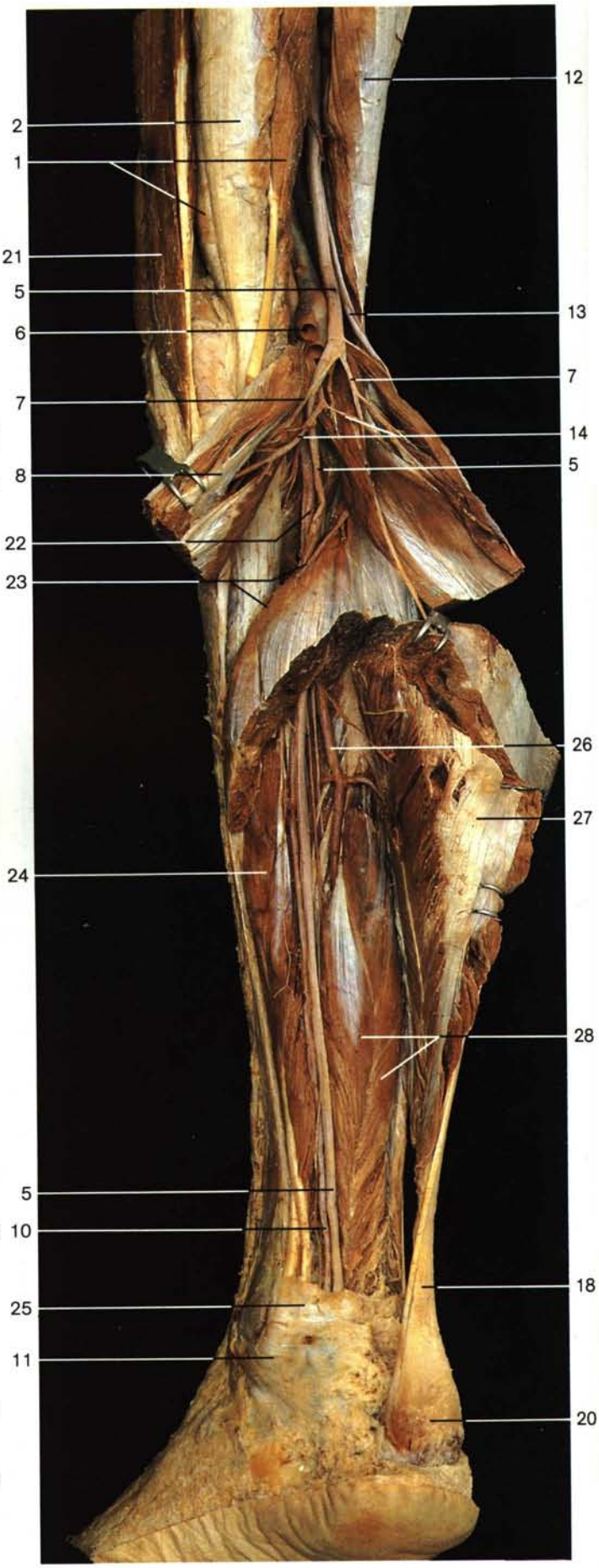

Right popliteal fossa, deep layer (posterior aspect). The medial head of gastrocnemius muscle has been divided and reflected.

1 Semimembranosus muscle
2 Semitendinosus muscle
3 **Popliteal vein**
4 Popliteal artery
5 **Tibial nerve**
6 Small saphenous vein (cut)
7 Muscular branch of tibial nerve
8 Medial head of gastrocnemius muscle
9 Tendon of plantaris muscle
10 **Posterior tibial artery**
11 Medial malleolus
12 Biceps femoris muscle
13 Common peroneal nerve
14 **Sural arteries**
15 Plantaris muscle
16 Lateral head of gastrocnemius muscle
17 Soleus muscle
18 Calcaneal tendon
19 Lateral malleolus
20 Calcaneal tuberosity
21 Sartorius muscle
22 **Popliteal artery**
23 Tendinous arch of soleus muscle
24 Flexor digitorum longus muscle
25 Flexor retinaculum
26 **Peroneal artery**
27 Soleus muscle
28 Flexor hallucis longus muscle
29 **Anterior tibial artery**
30 Muscular branches of tibial nerve
31 Tibialis posterior muscle
32 Communicating branch of peroneal artery
33 Tendon of tibialis anterior muscle
34 Tibia
35 Tendon of extensor hallucis longus muscle
36 Tendons of extensor digitorum longus muscle
37 Anterior tibialis artery
38 Fibula
39 Tendons of peroneus longus and brevis muscles

Right leg, posterior crural region, deepest layer (posterior aspect). Triceps surae (gastrocnemius and soleus) and flexor hallucis longus muscles have been cut and reflected.

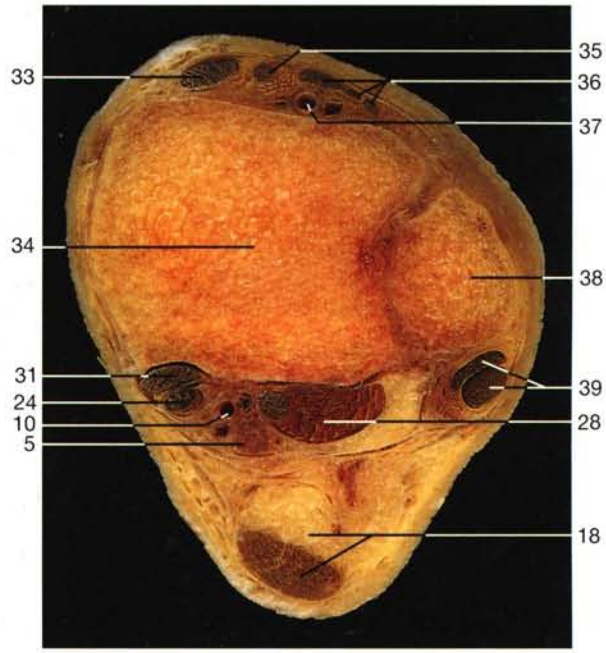

Cross-section of the leg, superior to the malleoli (from below).

Right leg and foot, anterior crural region and dorsum of foot; cutaneous nerves and veins (anterior aspect).

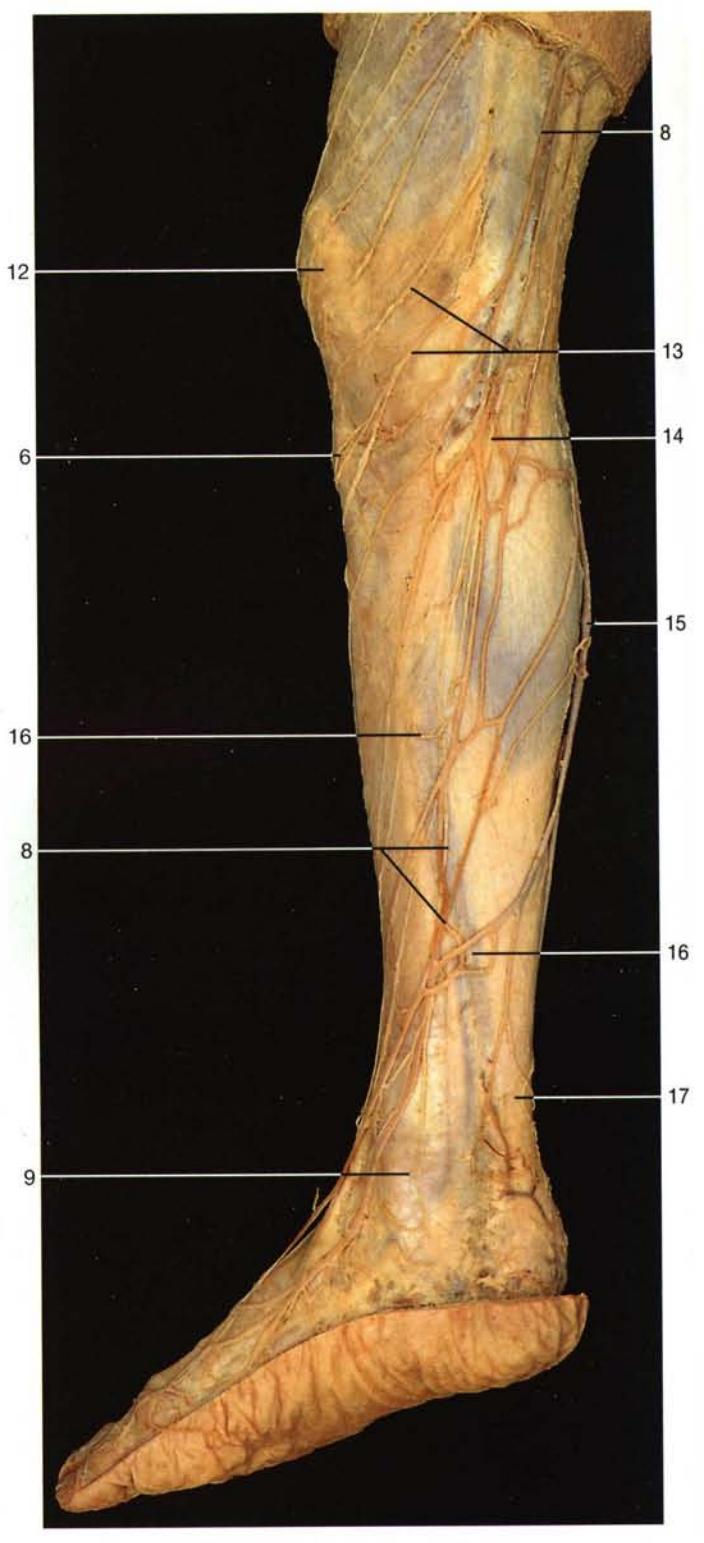

Right leg and foot; cutaneous nerves and veins (medial aspect).

1 Superficial crural fascia
2 Medial cutaneous branch of superficial peroneal nerve
3 Lateral malleolus
4 Lateral cutaneous branch of superficial peroneal nerve
5 Cutaneous branch of sural nerve
6 Position of tuberosity of tibia
7 Anterior margin of tibia
8 **Great saphenous vein**
9 Medial malleolus
10 Deep peroneal nerve
11 **Venous arch of dorsum of foot**
12 Position of patella
13 **Infrapatellar branches of saphenous nerve**
14 Saphenous nerve
15 Small saphenous vein
16 Perforating vein
17 Calcaneal tendon

Right leg and dorsum of foot; cutaneous nerves and veins (lateral aspect).

1 Position of fibula
2 **Sural nerve**
3 **Small saphenous vein**
4 Calcaneal tendon
5 Lateral calcaneal branches of sural nerve
6 Venous network at lateral malleolus
7 Cutaneous branch of sural nerve
8 Tendon of peroneus brevis muscle
9 Tendons of extensor digitorum longus muscle
10 Fascia cruris
11 Superficial peroneal nerve
12 Position of tibia
13 Lateral cutaneous branch } of **superficial peroneal nerve**
14 Medial cutaneous branch } of **superficial peroneal nerve**
15 Lateral malleolus
16 Dorsal digital nerves
17 Dorsal venous arch
18 Deep peroneal nerve

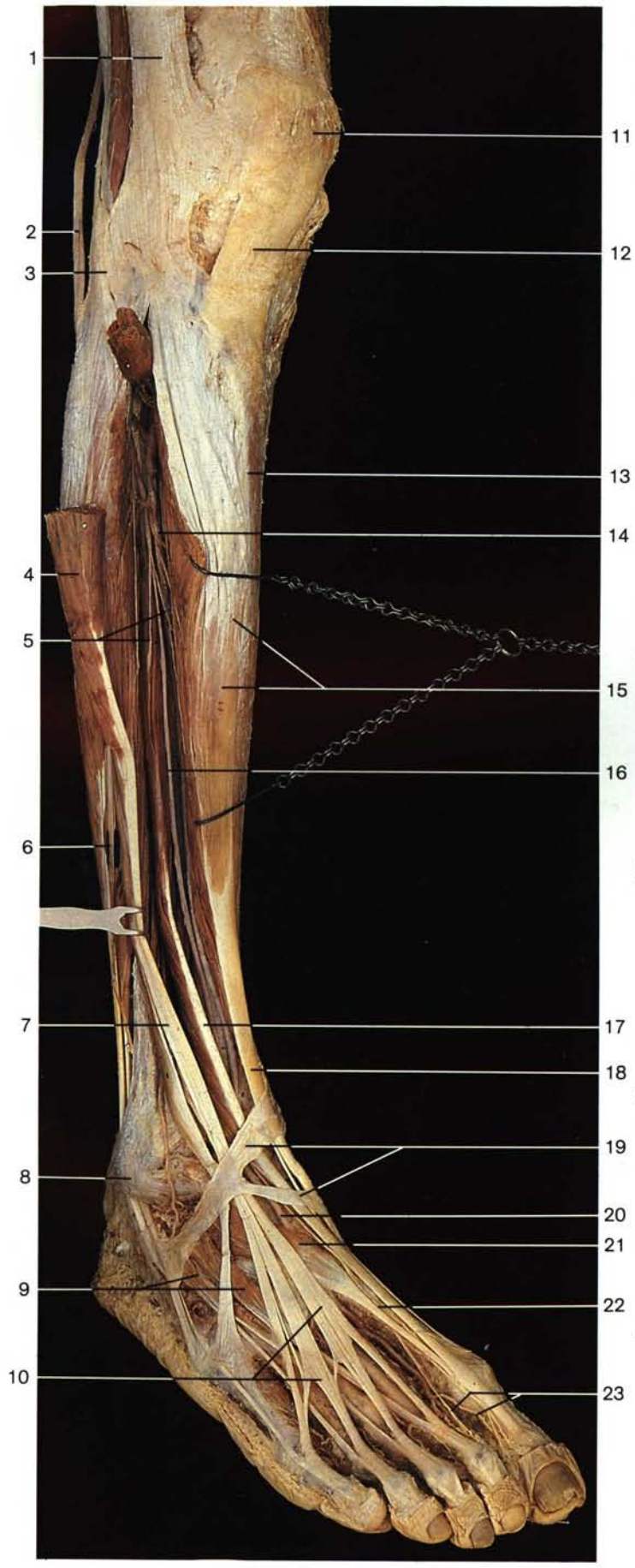

Right leg and dorsum of foot, middle layer (anterior lateral aspect). The extensor digitorum longus muscle has been divided and reflected laterally.

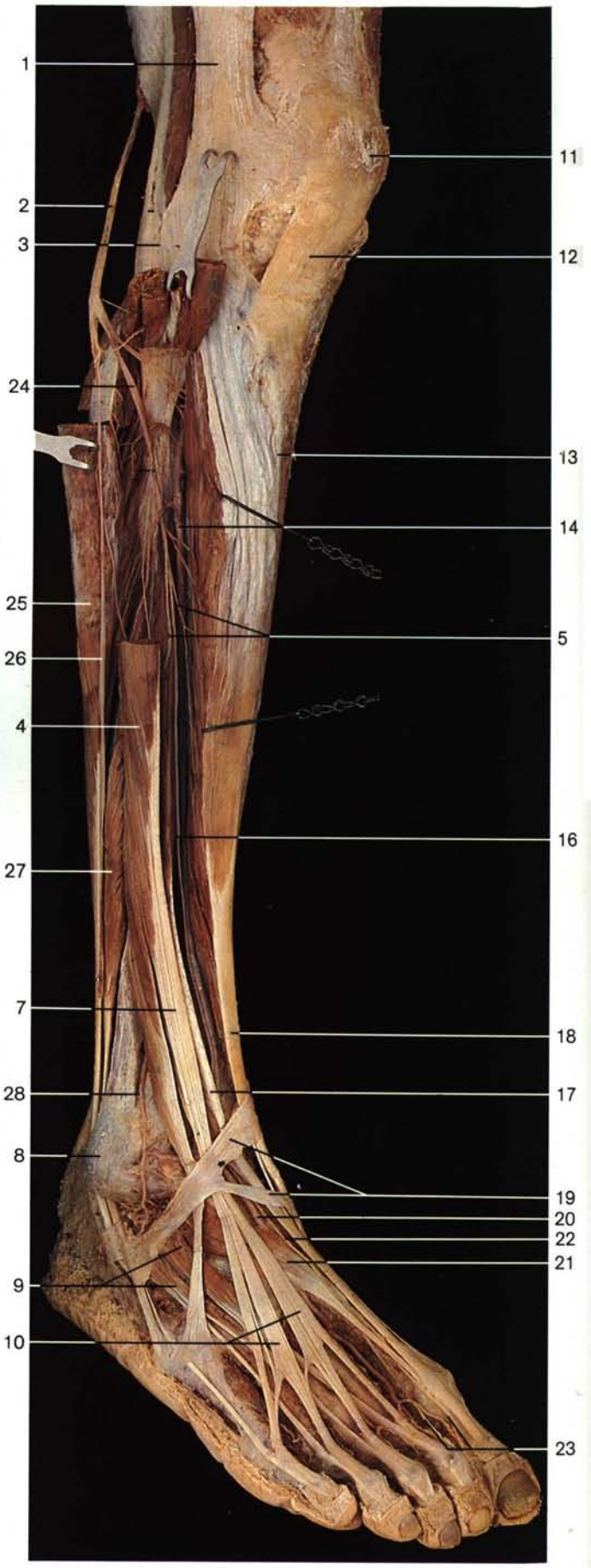

Right leg, deep layer (anterior lateral aspect). The extensor digitorum longus and the peroneus longus muscle have been divided or removed. The common peroneal nerve has been elevated to show its course around the head of fibula.

Coronal section through the foot and talocrural joint (anterior aspect).

Coronal section through the foot (MR-Scan after A. Heuck, G. Luttke and J. W. Rohen, 1994).

1 Iliotibial tract
2 **Common peroneal nerve**
3 Position of head of fibula
4 Extensor digitorum longus muscle
5 Muscular branches of deep peroneal nerve
6 **Superficial peroneal nerve**
7 Tendon of extensor digitorum longus muscle
8 Lateral malleolus
9 Extensor digitorum brevis muscle
10 Tendons of extensor digitorum longus muscle
11 Patella
12 Patellar ligament
13 Anterior margin of tibia
14 **Anterior tibial artery**
15 Tibialis anterior muscle
16 Deep peroneal nerve
17 Extensor hallucis longus muscle
18 Tendon of tibialis anterior muscle
19 Extensor retinaculum
20 Dorsalis pedis artery
21 Extensor hallucis brevis muscle
22 Deep peroneal nerve (on dorsum of foot)
23 Dorsal digital nerves (terminal branches of deep peroneal nerve)
24 **Deep peroneal nerve**
25 Peroneus longus muscle (cut)
26 Superficial peroneal nerve (with peroneal muscles laterally reflected)
27 Peroneus brevis muscle
28 Lateral anterior malleolar artery
29 Fibula
30 Distal tibiofibular joint (syndesmosis)
31 Interosseous talocalcaneal ligament
32 Calcaneus
33 Tendon of peroneus brevis muscle
34 Cuboid bone
35 Lateral cuneiform bone
36 Metatarsal bones
37 Dorsal interosseous muscles
38 Tibia
39 **Talocrural joint**
40 Medial malleolus
41 Talus
42 **Talocalcaneonavicular joint**
43 Navicular bone
44 Medial cuneiform bone
45 Intermediate cuneiform bone
46 First metatarsal bone
47 Metatarsophalangeal joint of great toe
48 Proximal phalanx of great toe
49 Distal phalanx of great toe
50 Heads of metatarsal bones II-IV

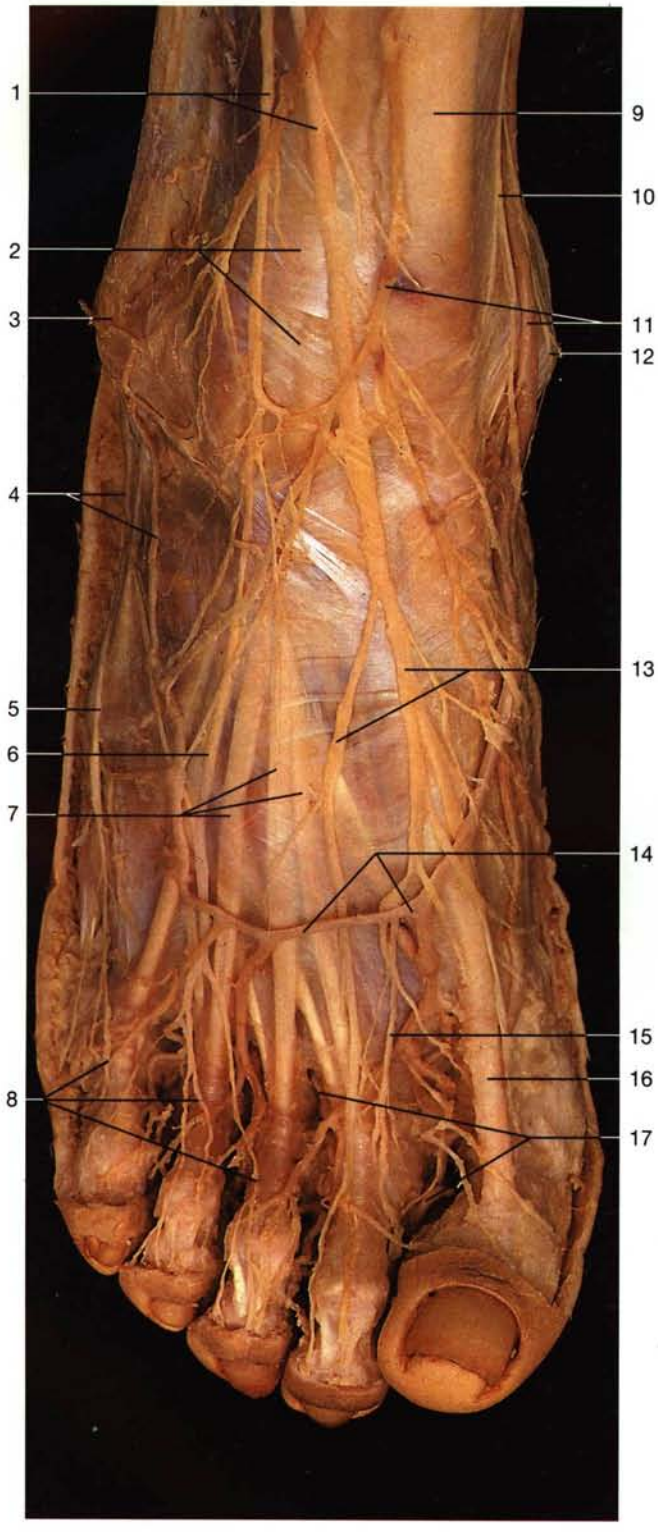

Dorsum of the right foot, superficial layer (anterior aspect).

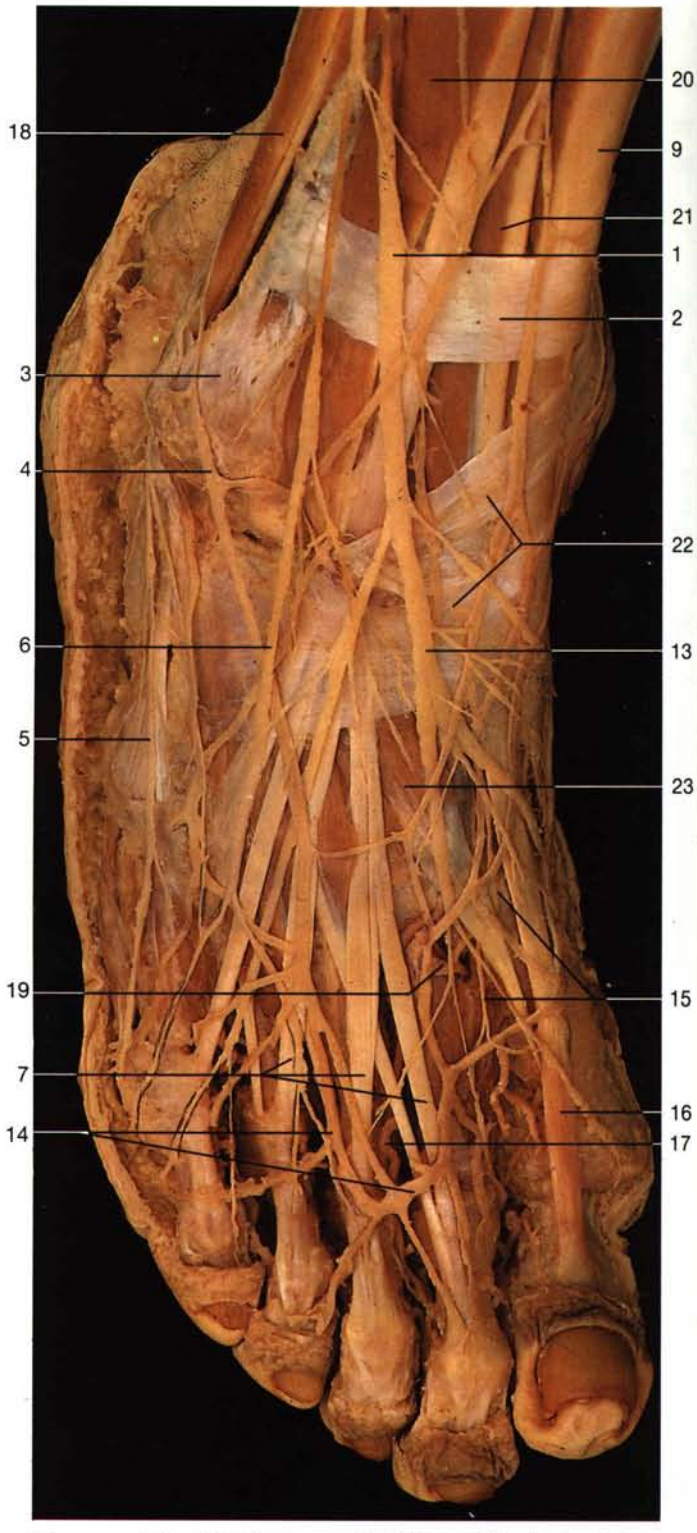

Dorsum of the right foot, superficial layer. The fascia of the dorsum has been removed.

1 **Superficial peroneal nerve**
2 Superior extensor retinaculum
3 Lateral malleolus
4 Venous network of lateral malleolus and tributaries of small saphenous vein
5 Lateral dorsal cutaneous nerve (branch of sural nerve)
6 Intermediate dorsal cutaneous nerve
7 Tendons of extensor digitorum longus muscle
8 Dorsal digital nerves
9 Tendon of tibialis anterior muscle
10 **Saphenous nerve**
11 Venous network of medial malleolus and tributaries of great saphenous vein
12 Medial malleolus
13 Medial dorsal cutaneous nerves
14 **Dorsal venous arch**
15 Dorsal digital nerve (of deep peroneal nerve)
16 Tendon of extensor hallucis longus muscle
17 Dorsal digital arteries
18 Peroneal muscles
19 Deep plantar branch of dorsalis pedis artery anastomosing with plantar arch
20 Extensor digitorum longus muscle
21 Extensor hallucis longus muscle
22 Inferior extensor retinaculum
23 Extensor hallucis brevis muscle

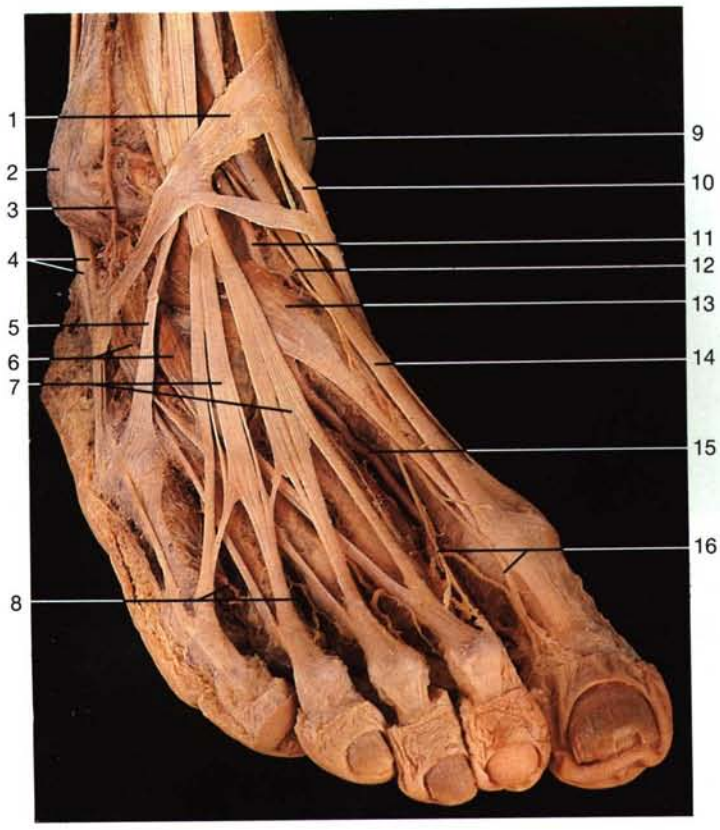

Dorsum of right foot, middle layer (anterior lateral aspect). The cutaneous nerves have been removed.

1 Extensor retinaculum
2 Lateral malleolus
3 Lateral anterior malleolar artery
4 Tendons of peroneal muscles
5 Tendon of peroneus tertius muscle
6 Extensor digitorum brevis muscle
7 Tendons of extensor digitorum longus muscle
8 Dorsal metatarsal arteries
9 Medial malleolus
10 Tendon of tibialis anterior muscle
11 **Dorsalis pedis artery**
12 Deep peroneal nerve (on dorsum of foot)
13 Extensor hallucis brevis muscle
14 Tendon of extensor hallucis longus muscle
15 Dorsalis pedis artery with deep plantar branch to the plantar arch
16 Dorsal digital nerves (terminal branches of deep peroneal nerve)
17 Lateral tarsal artery
18 Extensor digitorum brevis muscle (divided)
19 Arcuate artery
20 Dorsal interosseous muscles
21 **Deep peroneal nerve**
22 Medial cuneiform and 1st metatarsal bone
23 Tendon of peroneus longus muscle
24 Abductor hallucis and flexor hallucis brevis muscles
25 Medial plantar artery, vein and nerve
26 Fourth and fifth metatarsal bone
27 Adductor hallucis (oblique head)
28 Tendons of flexor digitorum longus muscle
29 Lateral plantar artery, vein and nerve
30 Flexor digitorum brevis muscle
31 Plantar aponeurosis

Dorsum of right foot, deep layer (anterior lateral aspect). The extensor digitorum and hallucis breves muscles have been removed.

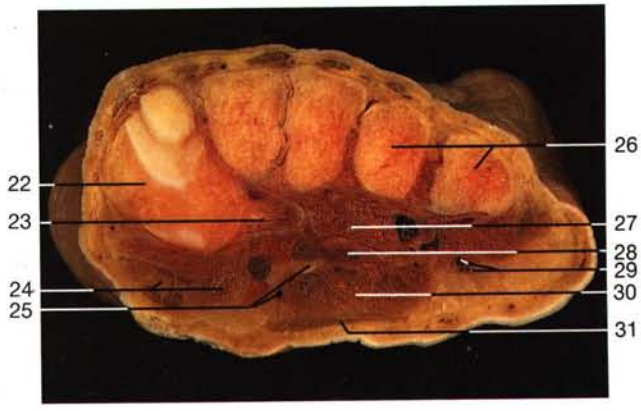

Cross-section of the rigt foot at the level of the metatarsal bones (posterior aspect).

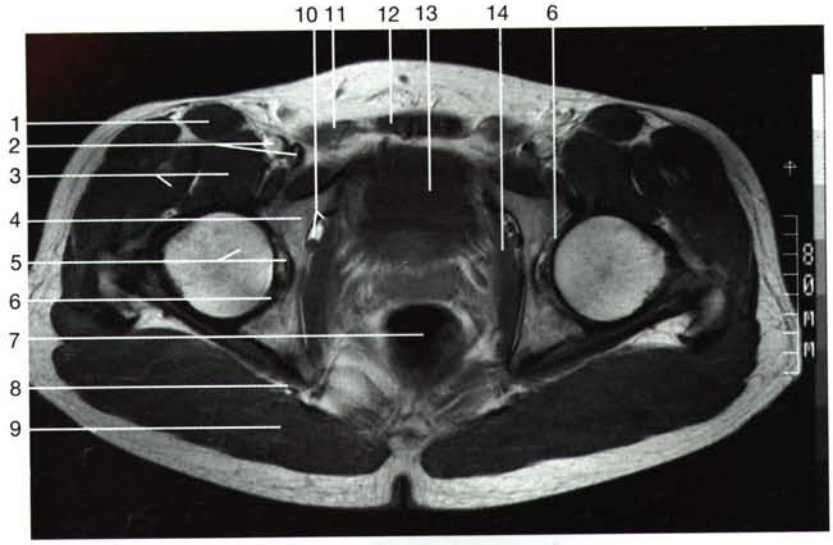

Axial section through the pelvis and the hip joints (section 1; MR-Image; inferior aspect).

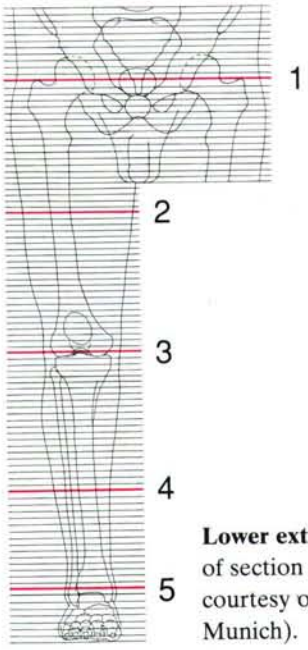

Lower extremity, location of section 1–5. (MR-Images, courtesy of Dr. A. Heuck, Munich).

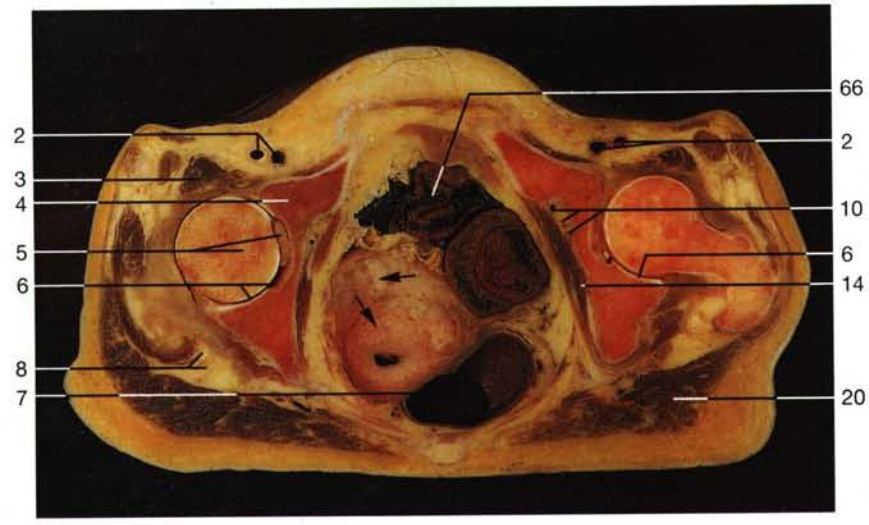

Axial section through the pelvis and hip joints in the female (section 1, inferior aspect). (Arrows: uterus, myometrium with myoma).

1 Sartorius muscle
2 **Femoral artery and vein**
3 Iliopsoas muscle
4 Pelvic bone
5 **Femoral head** with ligament of femoral head
6 Articular cavity
7 Rectum
8 **Sciatic nerve** and accompanying artery
9 Gluteus maximus muscle
10 Obturator vessels and obturator nerve
11 Rectus abdominis muscle
12 Pyramidalis muscle
13 Urinary bladder
14 Obturator internus muscle
15 Rectus femoris muscle
16 Vastus intermedius and vastus lateralis of quadriceps femoris muscle

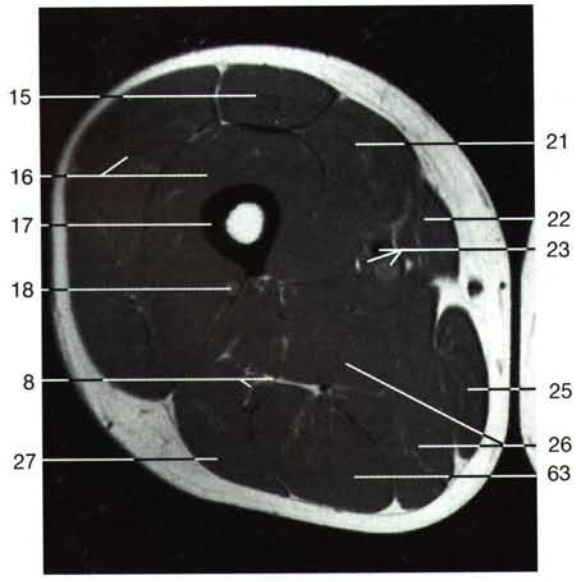

Axial section through the middle of the thigh (section 2; MR-Image, inferior aspect).

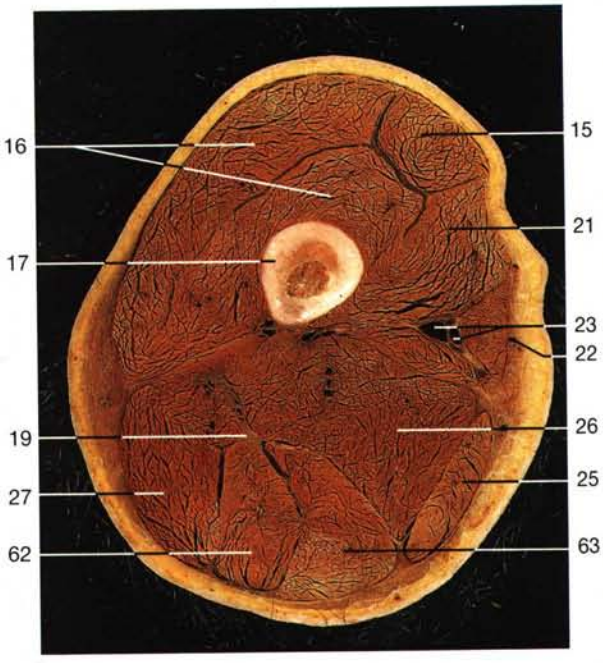

Axial section through the middle of the right thigh (section 2, inferior aspect).

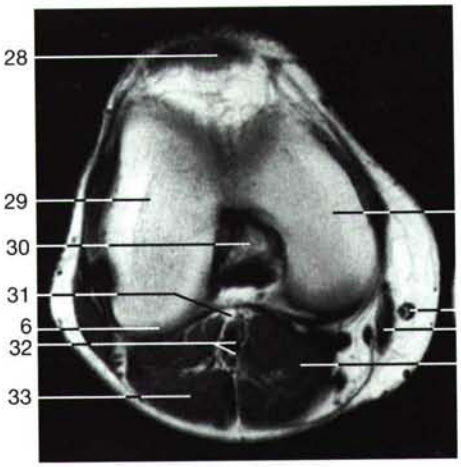

Axial section through the knee joint (section 3; MR-Image; inferior aspect).

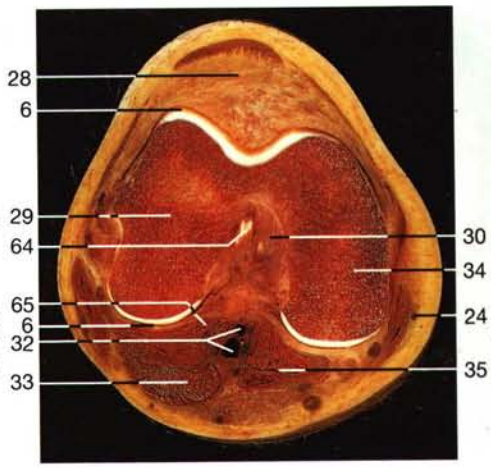

Axial section through the right knee joint (section 3, inferior aspect).

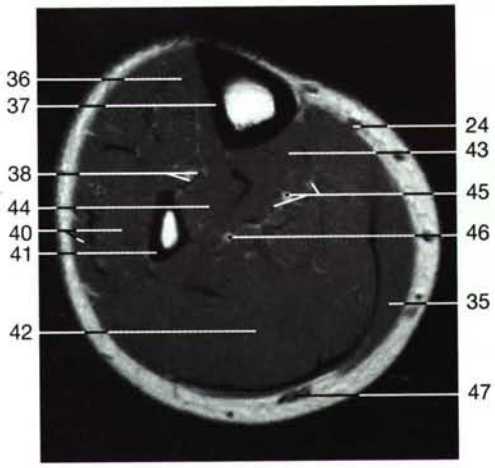

Axial section through the middle of the leg (section 4, MR-Image, inferior aspect).

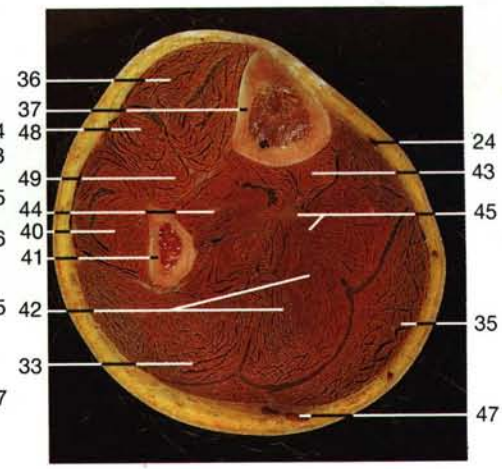

Axial section through the middle of the right leg (section 4, inferior aspect).

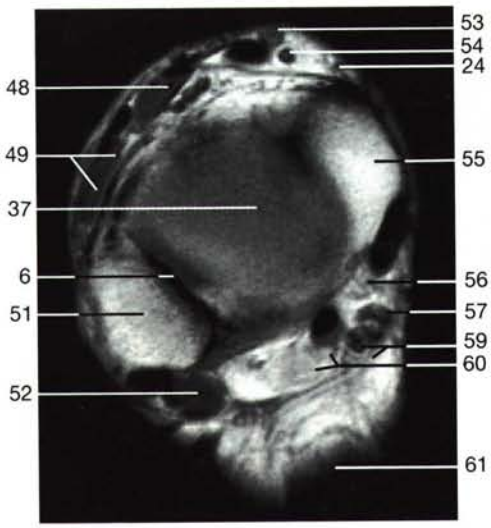

Axial section through the end of the right leg (section 5; MR-Image; inferior aspect).

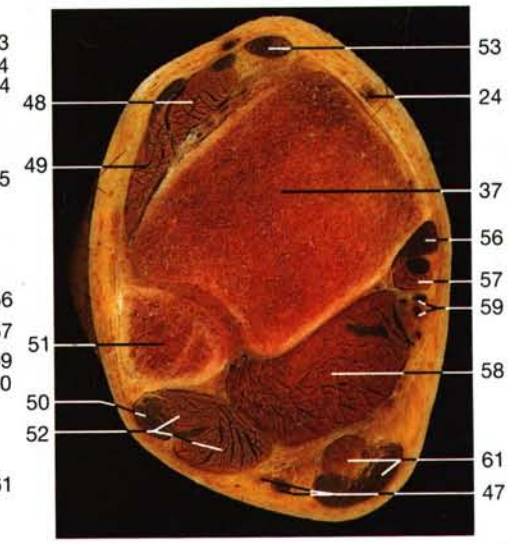

Axial section through the end of the right leg (section 5; inferior aspect).

17 **Femur**
18 Perforating artery
19 Sciatic nerve
20 Gluteus maximus muscle (insertion)
21 Vastus medialis muscle
22 Sartorius muscle
23 Femoral artery and vein
24 Great saphenous vein
25 Gracilis muscle
26 Adductor muscles
27 Biceps femoris muscle
28 Patellar ligament
29 Lateral condyle of femur
30 Posterior cruciate ligament
31 **Tibialis nerve**
32 Popliteal artery and vein
33 Lateral head of gastrocnemius muscle
34 Medial condyle of femur
35 Medial head of gastrocnemius muscle
36 Tibialis anterior muscle
37 **Tibia**
38 Deep peroneal nerve, anterior tibial artery and vein
40 Peroneus longus and brevis muscles
41 **Fibula**
42 Soleus muscle
43 Flexor digitorum longus muscle
44 Tibialis posterior muscle
45 Posterior tibial artery and vein and tibial nerve
46 Peroneal artery
47 Small saphenous vein and sural nerve
48 Extensor hallucis longus muscle
49 Extensor digitorum longus muscle
50 Tendon of peroneus longus muscle
51 Lateral malleolus (fibula)
52 Peroneus brevis muscle
53 Tibialis anterior muscle (tendon)
54 Dorsalis pedis artery
55 Medial malleolus (tibia)
56 Tibialis posterior muscle (tendon)
57 Flexor digitorum longus muscle (tendon with synovial sheath)
58 Flexor hallucis longus muscle
59 Posterior tibial artery and vein
60 Lateral and medial plantar nerves
61 Calcaneal tendon
62 Semitendinosus muscle
63 Semimembranosus muscle
64 Anterior cruciate ligament
65 Plantaris muscle
66 Small intestine

Sole of the right foot, superficial layer (from below): dissection of cutaneous nerves and vessels.

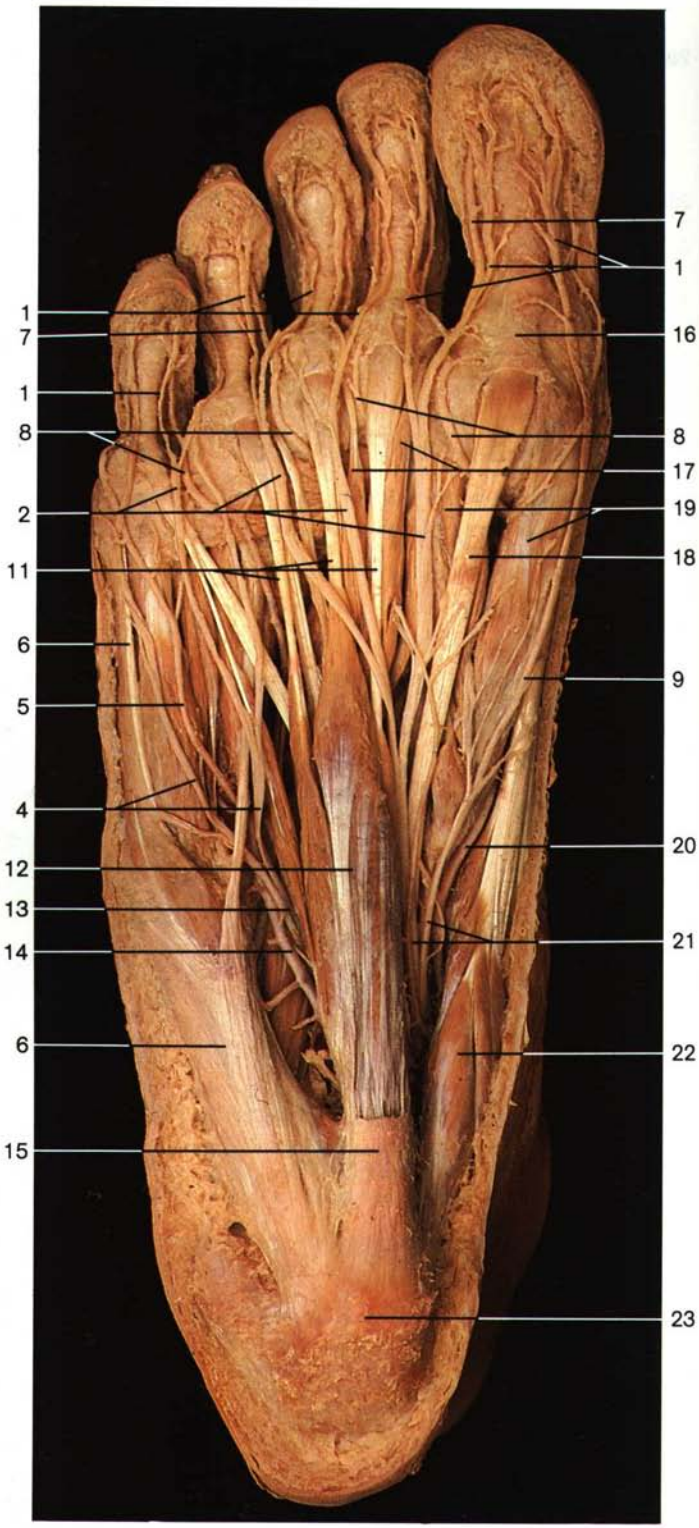

Sole of the right foot, middle layer (from below). The plantar aponeurosis has been removed.

1 Proper plantar digital nerves
2 Common plantar digital nerves
3 Plantar aponeurosis
4 Superficial branch of lateral plantar nerve
5 Superficial branch of lateral plantar artery
6 Abductor digiti minimi
7 Proper plantar digital arteries
8 Common plantar digital arteries
9 Digital branch of medial plantar nerve to great toe
10 Medial calcaneal branches
11 Tendons of flexor digitorum brevis muscle
12 Flexor digitorum brevis muscle
13 **Superficial branch of lateral plantar nerve**
14 **Lateral plantar artery**
15 Plantar aponeurosis (remnant)
16 Digital synovial sheath
17 Lumbrical muscles
18 Tendon of flexor hallucis longus muscle
19 Flexor hallucis brevis muscle
20 **Medial plantar artery**
21 **Medial plantar nerve**
22 Abductor hallucis muscle
23 Calcaneal tuberosity
24 Tendons of flexor digitorum longus muscle
25 Quadratus plantae muscle
26 **Lateral plantar nerve**
27 Flexor digitorum brevis muscle (cut)
28 Synorial sheaths
29 Plantar arch

Sole of the right foot, middle layer (from below); dissection of vessels and nerves. The flexor digitorum brevis muscle has been divided and anteriorly reflected.

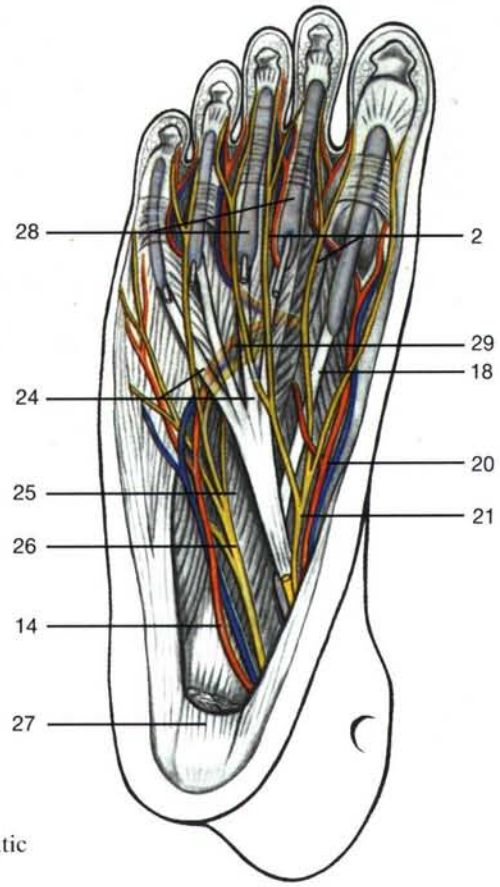

Sole of the right foot. Synovial sheaths of flexor tendons indicated in light blue. (Schematic drawing.)

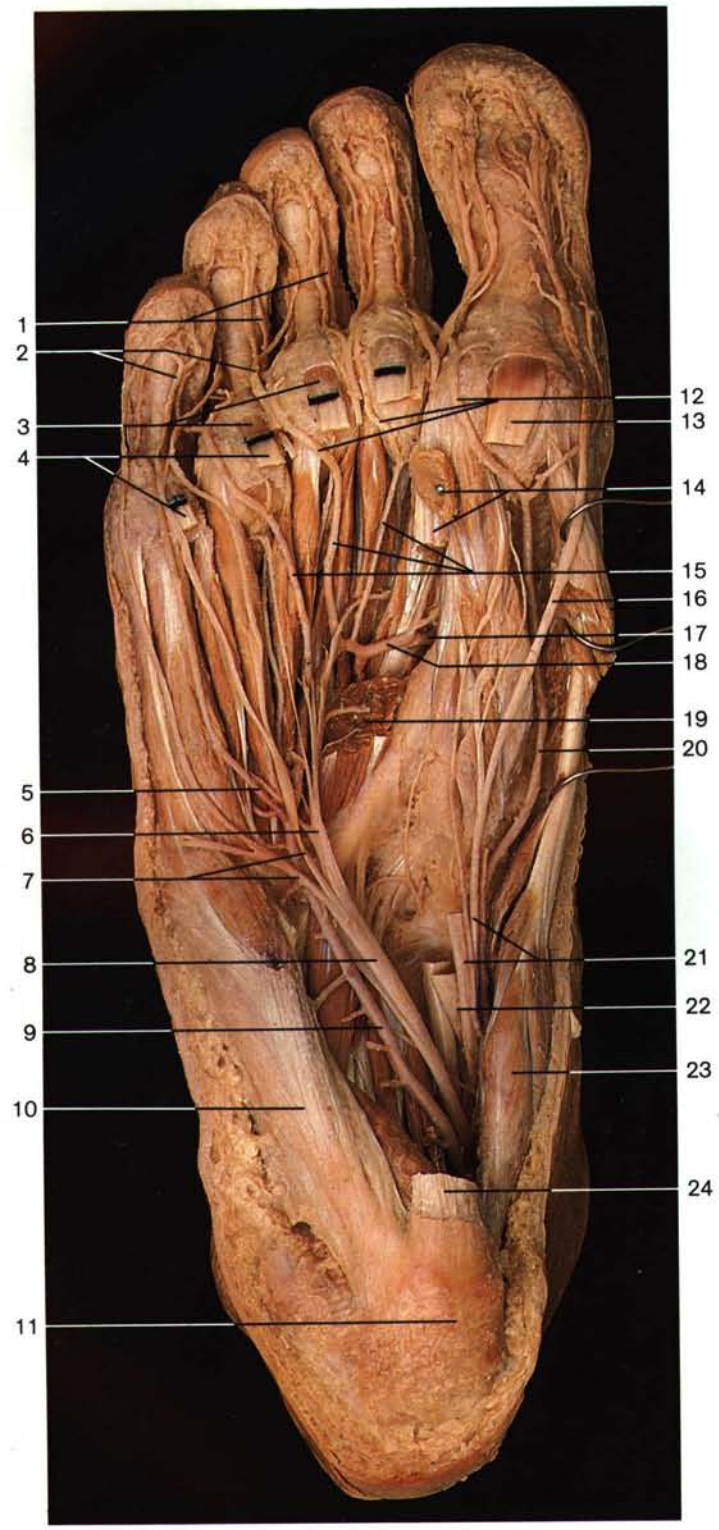

1 Proper plantar digital arteries
2 Proper plantar digital nerves
3 Tendons of flexor digitorum brevis muscle
4 Tendons of flexor digitorum longus muscle
5 Superficial branch of lateral plantar artery
6 Deep branch of lateral plantar nerve
7 Superficial branch of lateral plantar nerve
8 **Lateral plantar nerve**
9 **Lateral plantar artery**
10 Abductor digiti minimi muscle
11 Calcaneal tuberosity
12 Common plantar digital arteries
13 Tendon of flexor hallucis longus muscle
14 Insertion of both heads of adductor hallucis muscle
15 Plantar metatarsal arteries
16 Medial plantar nerve of great toe
17 Deep plantar branch of dorsalis pedis artery (perforating branch)
18 **Plantar arch**
19 Oblique head of adductor hallucis muscle (cut)
20 **Medial plantar artery**
21 **Medial plantar nerve**
22 Crossing of tendons in sole of foot (flexor hallucis longus and flexor digitorum longus muscles)
23 Abductor hallucis muscle
24 Origin of flexor digitorum brevis muscle

Sole of the right foot, deep layer (from below); dissection of vessels and nerves. The flexor digitorum brevis muscle, the quadratus plantae muscle with the tendons of the flexor digitorum longus muscle and some branches of the medial plantar nerve have been removed. The flexor hallucis brevis and adductor hallucis muscles have been cut and portions removed to show the somewhat atypical course of the medial plantar artery and deep muscles of the foot.

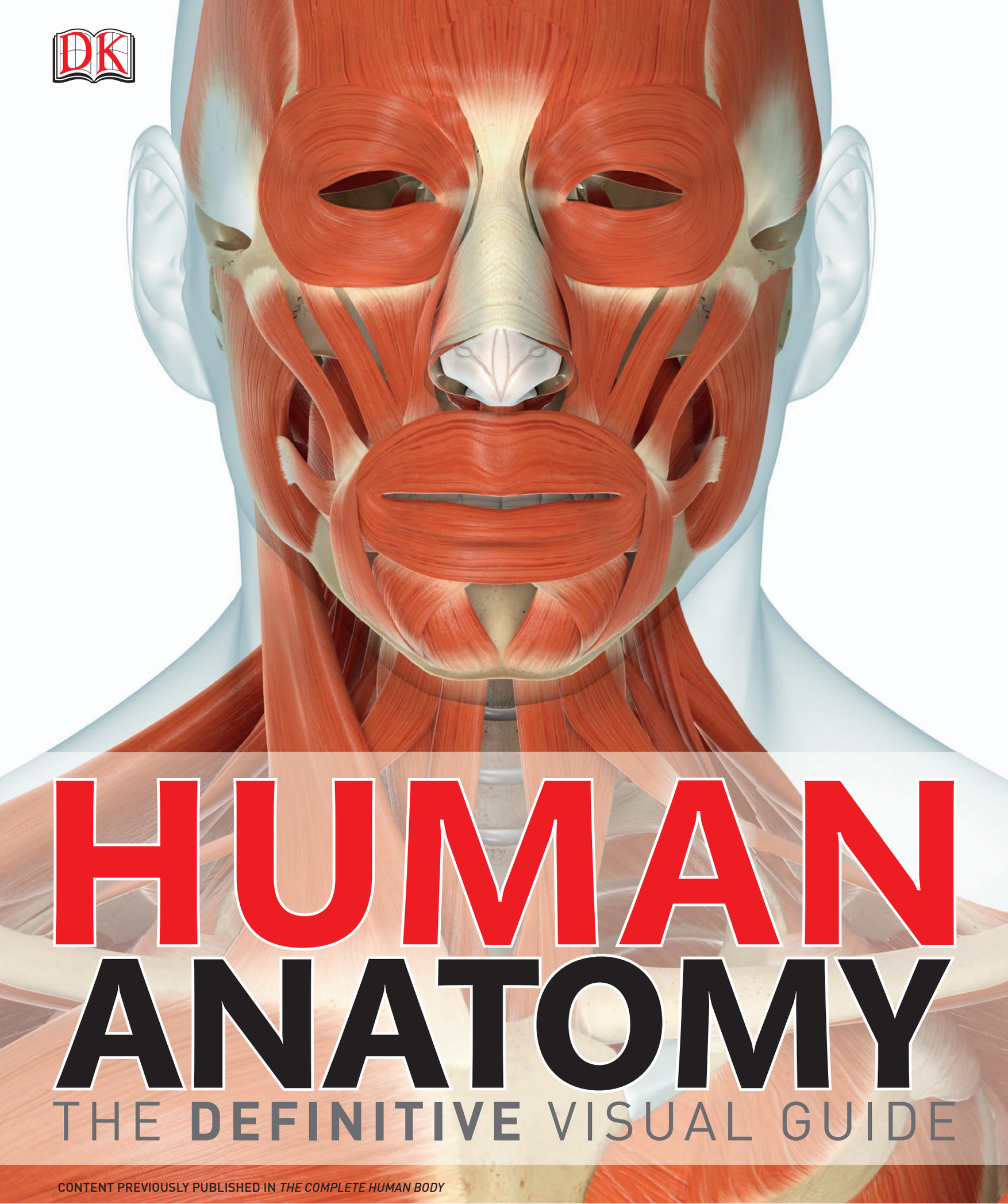

HUMAN ANATOMY

THE DEFINITIVE VISUAL GUIDE

CONTENT PREVIOUSLY PUBLISHED IN *THE COMPLETE HUMAN BODY*

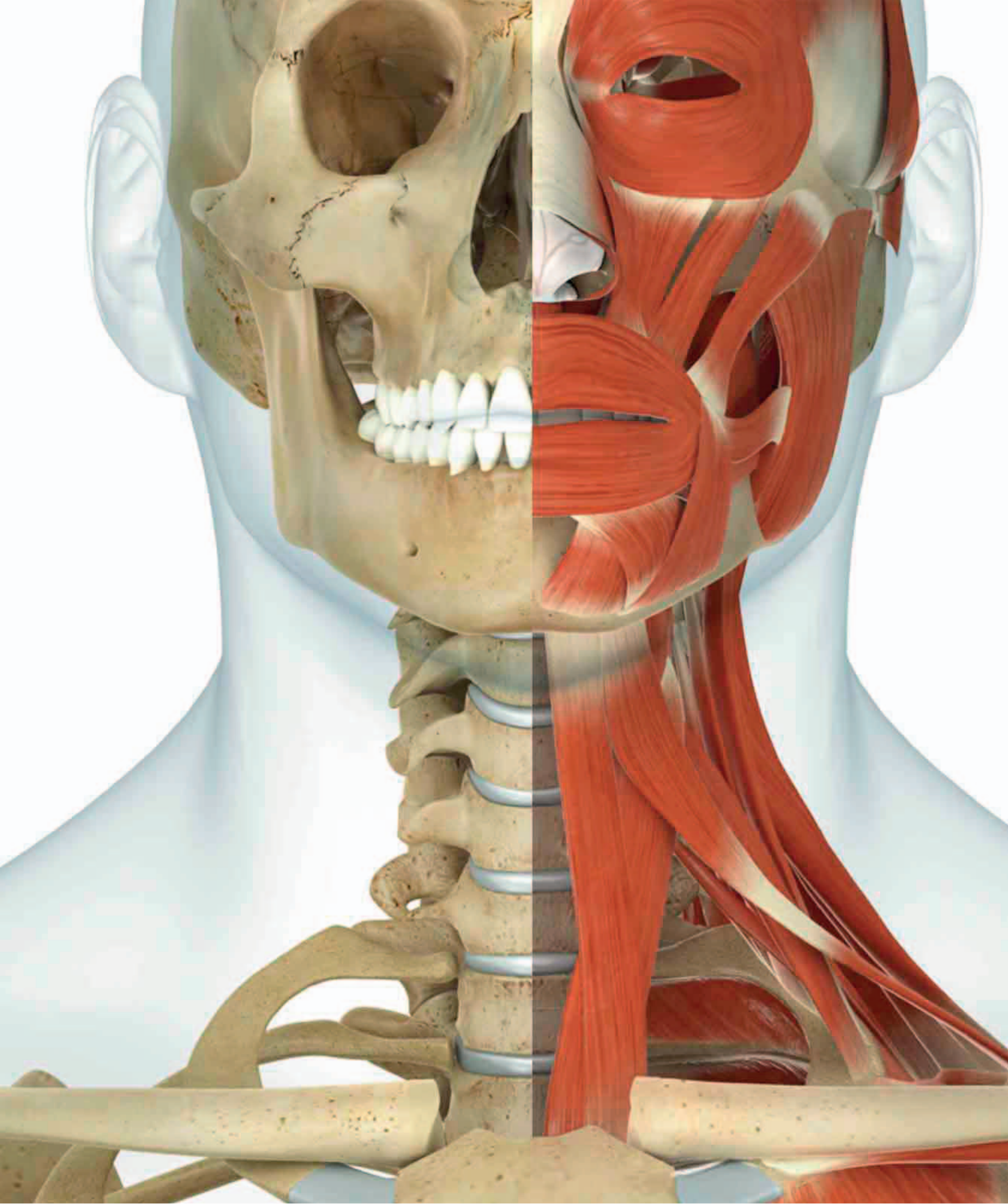

LONDON, NEW YORK, MELBOURNE,
MUNICH, AND DELHI

DK LONDON

Project Art Editor Duncan Turner
Jacket Designer Duncan Turner
Pre-Production Producer Vikki Nousiainen
Managing Art Editor Michelle Baxter
Production Editor Rachel Ng
Art Director Philip Ormerod
Associate Publishing Director Liz Wheeler
Project Editor Ruth O'Rourke-Jones
US Editor Cheryl Ehrlich
Editor Lili Bryant
Jacket Editor Manisha Majithia
Managing Editor Angeles Gavira
Publisher Sarah Larter
Publishing Director Jonathan Metcalf

DK DELHI

Senior Editor Anita Kakkar
Art Editors Suhita Dharamjit, Amit Malhotra
Deputy Managing Art Editor Sudakshina Basu
DTP Designer Vishal Bhatia
Production Manager Pankaj Sharma
Editor Pallavi Singh
Design Assistant Anjali Sachar
Managing Editor Rohan Sinha
Pre-Production Manager Balwant Singh
Picture Researcher Sumedha Chopra

Illustrators

Medi-Mation (Creative Director: Rajeev Doshi)
Antbits Ltd. (Richard Tibbitts)
Dotnamestudios (Andrew Kerr)
Deborah Maizels

Editor-in-Chief Professor Alice Roberts

Authors

THE INTEGRATED BODY
Linda Geddes

BODY SYSTEMS, IMAGING THE BODY
Professor Alice Roberts

Consultants

THE INTEGRATED BODY
Professor Mark Hanson,
University of Southampton, UK

BODY SYSTEMS, IMAGING THE BODY
Professor Harold Ellis, King's College, London
Professor Susan Standring, King's College, London

Content previously published in *The Complete Human Body*
First American Edition, 2010
This American Edition, 2014

Published in the United States by
DK Publishing
4th floor, 345 Hudson Street
New York, New York 10014

14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

256502–05/14

Copyright © 2010, 2014 Dorling Kindersley Limited
Foreword copyright © Alice Roberts
All rights reserved

Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form, or by any means (electronic, mechanical, photocopying, recording, or otherwise), without the prior written permission of both the copyright owner and the above publisher of this book.
Published in Great Britain by Dorling Kindersley Limited.

A catalog record for this book is available from the Library of Congress.

ISBN 978-1-4654-1954-5

DK books are available at special discounts when purchased in bulk for sales promotions, premiums, fund-raising, or educational use. For details, contact: DK Publishing Special Markets, 345 Hudson Street, New York, New York 10014 or SpecialSales@dk.com.

Printed and bound in China by South China

Discover more at
www.dk.com

CONTENTS

006 Foreword

008
THE INTEGRATED BODY

010 Human genetic formula
012 Cell
014 Body composition
016 Body systems
018 Terminology and planes

020
BODY SYSTEMS

022 SKIN, HAIR, AND NAIL

024 SKELETAL SYSTEM OVERVIEW
026 Head and neck
036 Thorax
040 Spine
042 Abdomen and pelvis
046 Pelvis

048 Shoulder and upper arm
054 Lower arm and hand
056 Hand and wrist joints
058 Hip and thigh
062 Hip and knee
064 Lower leg and foot
066 Foot and ankle

068 MUSCULAR SYSTEM OVERVIEW
070 Head and neck
076 Thorax
082 Abdomen and pelvis
086 Shoulder and upper arm
094 Lower arm and hand
098 Hip and thigh
106 Lower leg and foot

110 NERVOUS SYSTEM OVERVIEW
112 Brain
118 Head and neck
120 Brain (transverse and coronal sections)
122 Head and neck (cranial nerves)
124 Eye
126 Ear
128 Neck
130 Thorax
132 Abdomen and pelvis
134 Shoulder and upper arm
138 Lower arm and hand
140 Hip and thigh
144 Lower leg and foot

146 RESPIRATORY SYSTEM OVERVIEW
148 Head and neck
150 Thorax
152 Lungs

154 CARDIOVASCULAR SYSTEM OVERVIEW
156 Head and neck
160 Thorax
162 Heart
166 Abdomen and pelvis
168 Shoulder and upper arm
172 Lower arm and hand
174 Hip and thigh
178 Lower leg and foot

180 LYMPHATIC AND IMMUNE SYSTEM OVERVIEW
182 Head and neck
184 Thorax
186 Abdomen and pelvis
188 Shoulder and upper arm
190 Hip and thigh

192 DIGESTIVE SYSTEM OVERVIEW
194 Head and neck
196 Thorax
198 Abdomen and pelvis
200 Stomach and intestines
202 Liver, pancreas, and gallbladder

204 URINARY SYSTEM OVERVIEW
206 Abdomen and pelvis

208 REPRODUCTIVE SYSTEM OVERVIEW
210 Thorax
212 Abdomen and pelvis

216 ENDOCRINE SYSTEM OVERVIEW
218 Head and neck

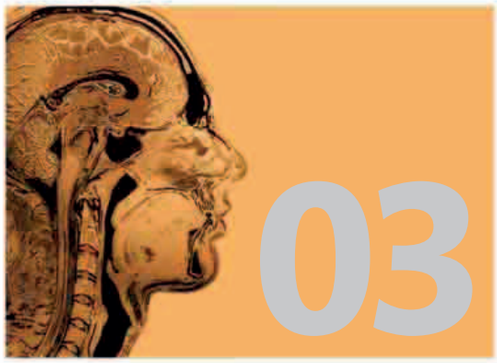

220
IMAGING THE BODY

222 Imaging techniques
224 Head and neck
226 Thorax
228 Abdomen and pelvis
230 Lower arm and hand
232 Lower limb and foot

234 GLOSSARY
241 INDEX
255 ACKNOWLEDGMENTS

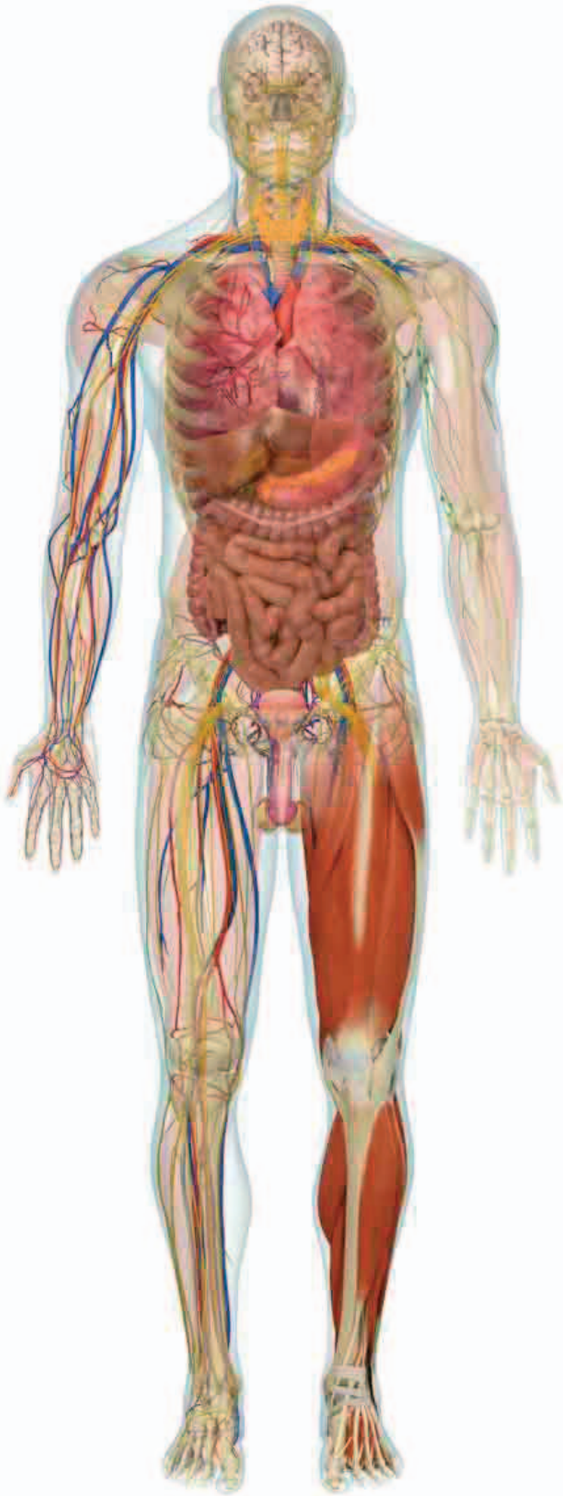

01 The Integrated Body

The human body comprises trillions of cells, each one a complex unit with intricate workings in itself. Cells are the building blocks of tissues, organs, and eventually, the integrated body systems that all interact—allowing us to function and survive.

010 Human genetic formula
012 Cell
014 Body composition
016 Body systems
018 Terminology and planes

HUMAN GENETIC FORMULA

DNA (deoxyribonucleic acid) is the blueprint for all life, from the humblest yeast to human beings. It provides a set of instructions for how to assemble the many thousands of different proteins that make us who we are. It also tightly regulates this assembly, ensuring that the components of the assemby do not run out of control.

THE MOLECULE OF LIFE

Although we all look different, the basic structure of our DNA is identical. It consists of chemical building blocks called bases, or nucleotides. What varies between individuals is the precise order in which these bases connect into pairs. When base pairs are strung together they can form functional units called genes, which "spell out" the instructions for making a protein. Each gene encodes a single protein, although some complex proteins are encoded by more than one gene. Proteins have a wide range of vital functions in the body. They form structures such as skin or hair, carry signals around the body, and fight off infectious agents such as bacteria. Proteins also make up cells, the basic units of the body, and carry out the thousands of basic biochemical processes needed to sustain life. However, only about 1.5 per cent of our DNA encodes genes. The rest consists of regulatory sequences, structural DNA, or has no obvious purpose – so-called "junk DNA".

DNA double helix
In the vast majority of organisms, including humans, long strands of DNA twist around each other to form a right-handed spiral structure called a double helix. The helix consists of a sugar (deoxyribose) and phosphate backbone and complementary base pairs that stick together in the middle. Each twist of the helix contains around ten base pairs.

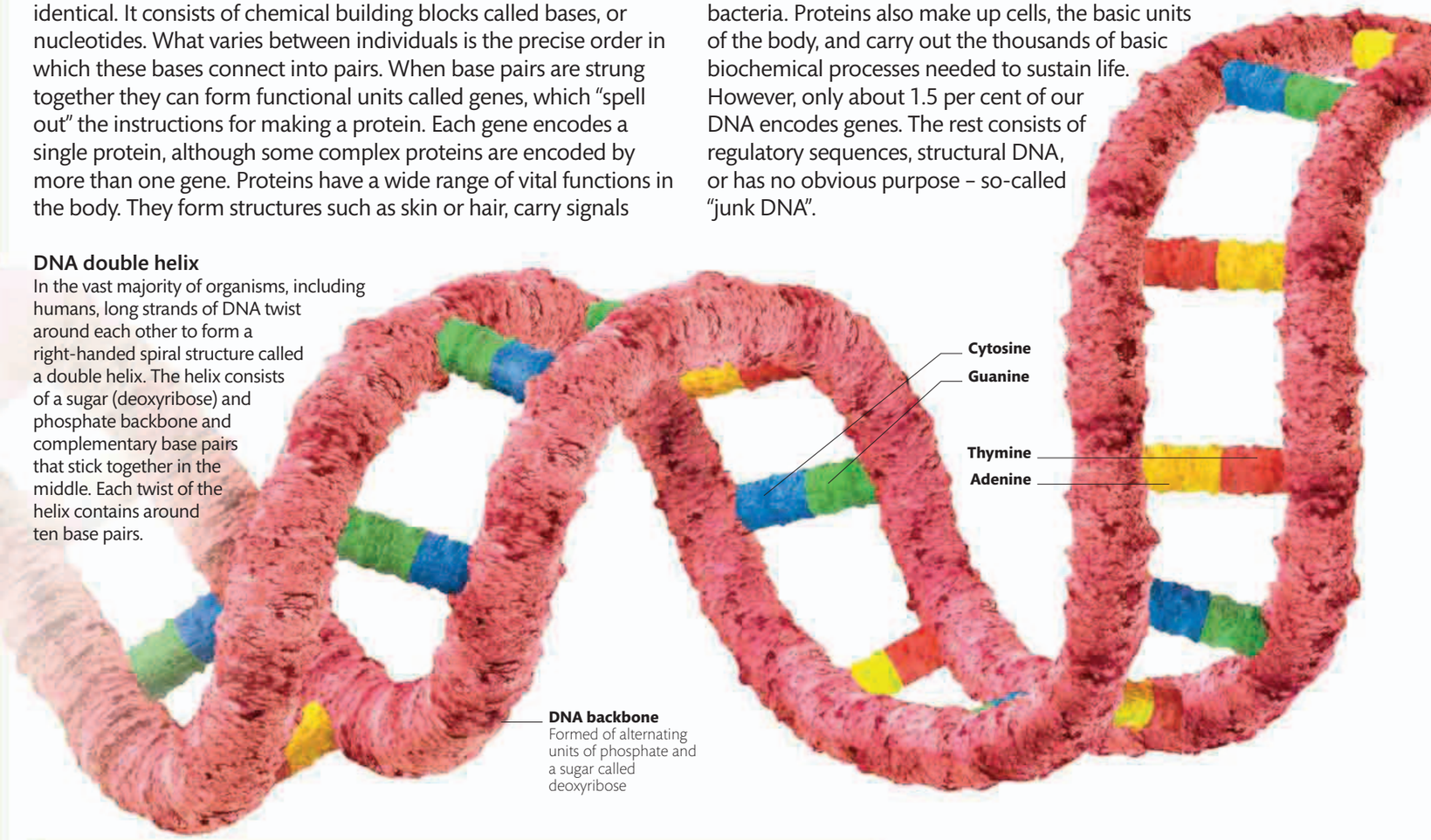

PACKAGING DNA

The human genome is composed of approximately 3 billion bases of DNA—about 6½ft (2m) of DNA in every cell if it were stretched from end to end. So our DNA must be packaged in order to fit inside each cell. DNA is concentrated into dense structures called chromosomes. Each cell has 23 pairs of chromosomes (46 in total)—one set from each parent. To package DNA, the double helix must first be coiled around histone proteins, forming a structure that looks like a string of beads. These histone "beads" wind around and lock together into densely coiled "chromatin", which, when a cell prepares to divide, further winds back on itself into tightly coiled chromosomes.

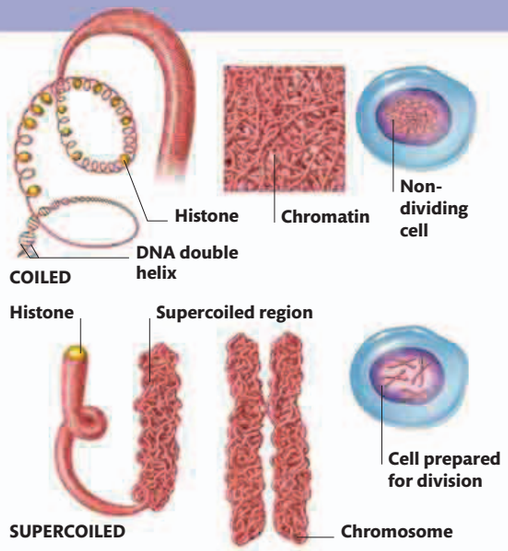

MAKING PROTEINS

Proteins consist of building blocks called amino acids, strung together in chains and folded. Every three base pairs of DNA codes for one amino acid. The body makes 20 different amino acids—others are obtained from the diet. Protein synthesis occurs in two steps: transcription and translation. In transcription, the DNA double helix unwinds, exposing single-stranded DNA. Complementary sequences of a related molecule called RNA (ribonucleic acid) then create a copy of the DNA sequence that locks into the exposed DNA bases to be translated into protein. This "messenger RNA" travels to ribosomes, where it is translated into strings of amino acids. These are then folded into the 3D structure of a particular protein.

BASE PAIRS

DNA consists of building blocks called bases. There are four types: adenine (A), thymine (T), cytosine (C), and guanine (G). Each base is attached to a phosphate group and a deoxyribose sugar ring to form a nucleotide. In humans, bases pair up to form a double-stranded helix in which adenine pairs with thymine, and cytosine with guanine. The two strands are "complementary" to each other. Even if they are unwound and unzipped, they can realign and rejoin.

GENES

A gene is a unit of DNA needed to make a protein. Genes range in size from just a few hundred to millions of base pairs. They control our development, but are also switched on and off in response to environmental factors. For example, when an immune cell encounters a bacterium, genes are switched on that produce antibodies to destroy it. Gene expression is regulated by proteins that bind to regulatory sequences within each gene. Genes contain regions that are translated into protein (exons) and non-coding regions (introns).

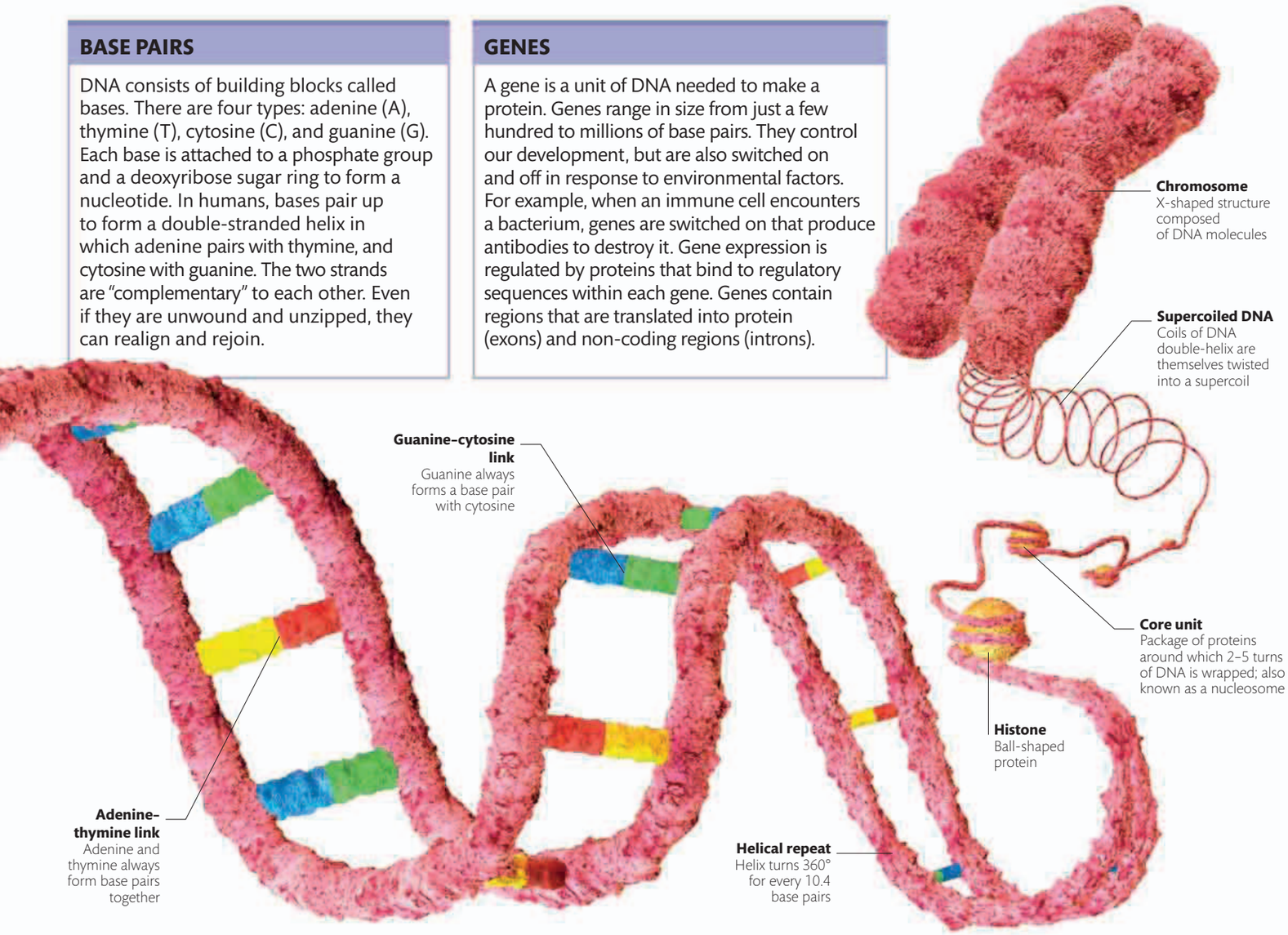

THE HUMAN GENOME

Different organisms contain different genes, but a surprisingly large proportion of genes are shared between organisms. For example, roughly half of the genes found in humans are also found in bananas. However, it would not be possible to substitute the banana version of a gene for a human one because variations in the order of the base pairs within each gene also distinguish us. Humans possess more or less the same genes, but many of the differences between individuals can be explained by subtle variations within each gene. In humans, DNA differs by only about 0.2 per cent, while human DNA differs from chimpanzee DNA by around 5 per cent. Human genes are divided unevenly between 23 pairs of chromosomes, and each chromosome consists of gene-rich and gene-poor sections. When chromosomes are stained, differences in these regions show up as light and dark bands, giving chromosomes a striped appearance. We still don't know the exact number of protein-coding genes in the human genome, but researchers currently estimate between 20,000 and 25,000.

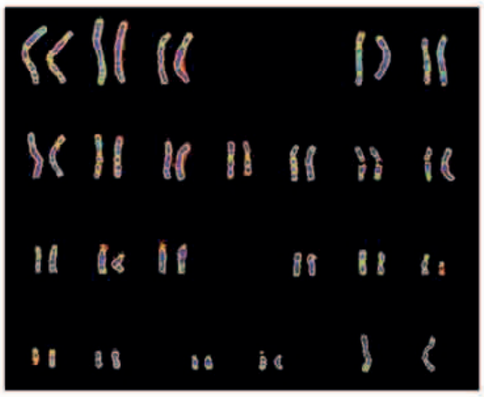

Karyotype
This is an organized profile of the chromosomes in someone's cells, arranged by size. Studying someone's karyotype enables doctors to determine whether any chromosomes are missing or abnormal.

GENETIC ENGINEERING

This form of gene manipulation enables us to substitute a defective gene with a functional one, or introduce new genes. Glow-in-the-dark mice were created by introducing a jellyfish gene that encodes a fluorescent protein into the mouse genome. Finding safe ways of delivering replacement genes to the correct cells in humans could lead to cures for many types of inherited diseases—so-called gene therapy.

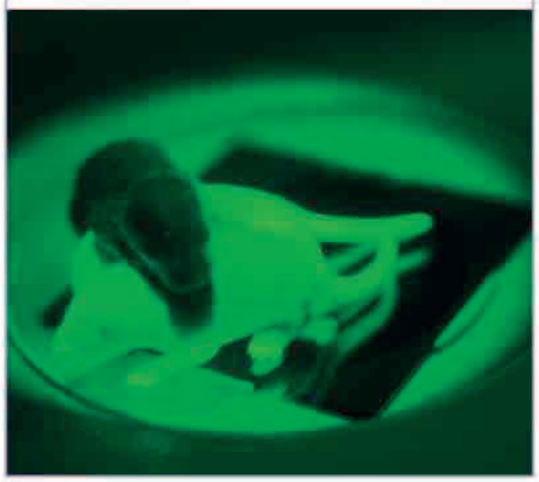

CELL

It is hard to comprehend what 75 trillion cells looks like, but observing yourself in a mirror would be a good start. That is how many cells exist in the average human body – and we replace millions of these cells every single day.

CELL ANATOMY

The cell is the basic functional unit of the human body. Cells are extremely small, typically only about 0.01mm across – even our largest cells are no bigger than the width of a human hair. They are also immensely versatile: some can form sheets like those in your skin or lining your mouth, while others can store or generate energy, such as fat and muscle cells. Despite their amazing diversity, there are certain features that all cells have in common, including an outer membrane, a control center called a nucleus, and tiny powerhouses called mitochondria.

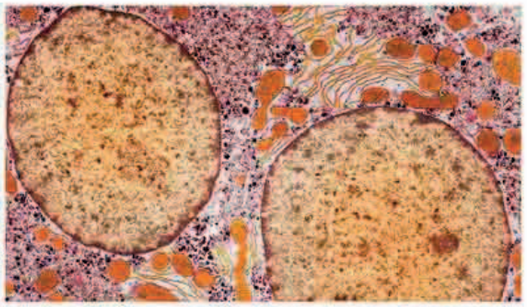

Liver cell
These cells make protein, cholesterol, and bile, and detoxify and modify substances from the blood. This requires lots of energy, so liver cells are packed with mitochondria (orange).

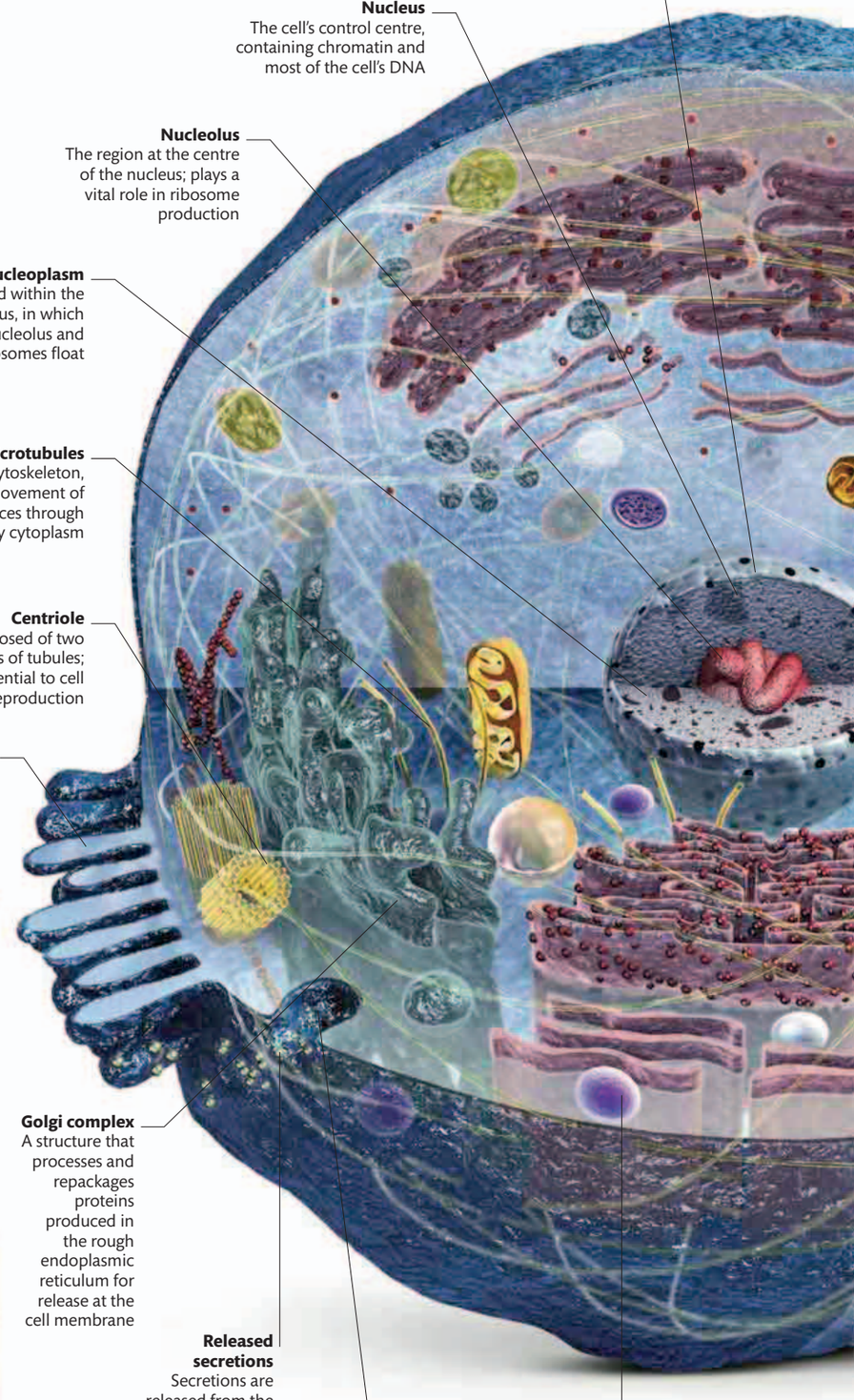

CELL METABOLISM

When a cell breaks down nutrients to generate energy for building new proteins or nucleic acids, it is known as cell metabolism. Cells use a variety of fuels to generate energy, but the most common one is glucose, which is transformed into adenosine triphosphate (ATP). This takes place in structures called mitochondria through a process called cellular respiration: enzymes within the mitochondria react with oxygen and glucose to produce ATP, carbon dioxide, and water. Energy is released when ATP is converted into adenoside diphosphate (ADP) via the loss of a phosphate group.

Mitochondrion
While the number of mitochondria varies between different cells, all have the same basic structure: an outer membrane and a highly folded inner membrane, where the production of energy actually takes place.

Generic cell

At a cell's heart is the nucleus, where the genetic material is stored and the first stages of protein synthesis occur. Cells also contain other structures for assembling proteins, including ribosomes, the endoplasmic reticulum, and the Golgi apparatus. The mitochondria provide the cell with energy.

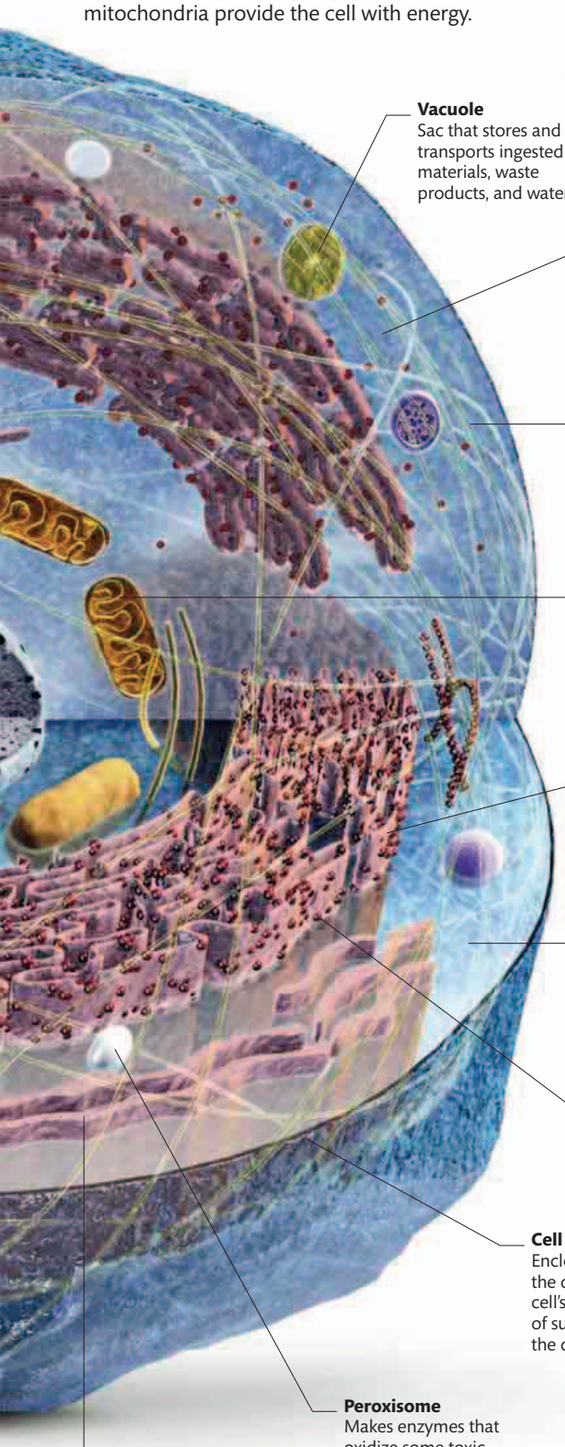

CELL TRANSPORT

Materials are constantly being transported in and out of the cell via the cell membrane. Such materials include fuel for generating energy, or building blocks for protein assembly. Some cells secrete signalling molecules to communicate with the rest of the body. The cell membrane is studded with proteins that help transport, allow cells to communicate, and identify a cell to other cells. The membrane is permeable to some molecules, but others need active transport through special channels in the membrane. Cells have three methods of transport: diffusion, facilitated diffusion, and active transport.

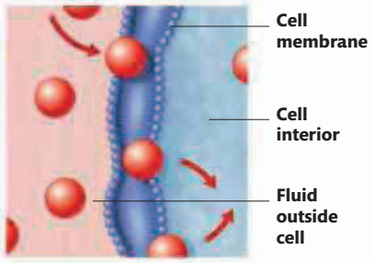

Diffusion
Molecules passively cross the membrane from areas of high to low concentration. Water and oxygen both cross by diffusion.

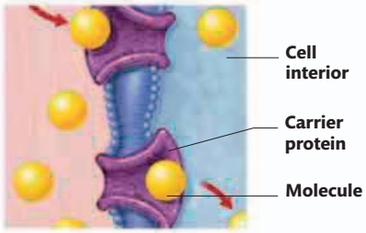

Facilitated diffusion
A carrier protein, or protein pore, binds with a molecule outside the cell, then changes shape and ejects the molecule into the cell.

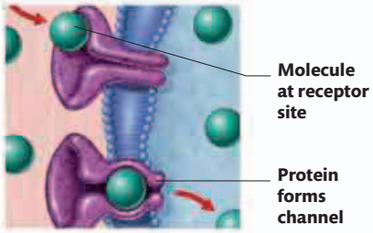

Active transport
Molecules bind to a receptor site on the cell membrane, triggering a protein, which changes into a channel that molecules travel through.

MAKING NEW BODY CELLS

While the cells lining the mouth are replaced every couple of days, some of the nerve cells in the brain have been there since before birth. Stem cells are specialized cells that constantly divide and give rise to new cells, such as blood cells. Cell division requires that a cell's DNA is accurately copied and then shared equally between two "daughter" cells, by a process called mitosis. The chromosomes are first replicated before being pulled to opposite ends of the cell. The cell then divides to produce two daughter cells, with the cytoplasm and organelles being shared between the two cells.

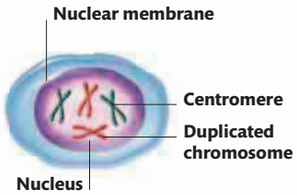

1 Preparation
The cell produces proteins and new organelles, and duplicates its DNA. The DNA condenses into X-shaped chromosomes.

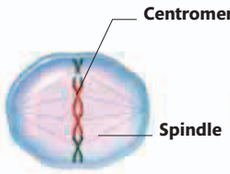

2 Alignment
The chromosomes line up along a network of filaments – spindle – linked to a larger network, called the cytoskeleton.

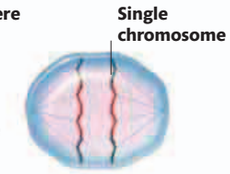

3 Separation
The chromosomes are pulled apart and move to opposite ends of the cell. Each end has an identical set of chromosomes.

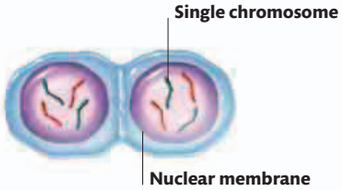

4 Splitting
The cell now splits into two, with the cytoplasm, cell membrane, and remaining organelles being shared roughly equally between the two daughter cells.

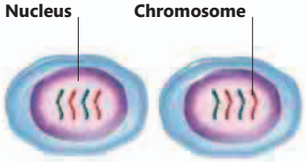

5 Offspring
Each daughter cell contains a complete copy of the DNA from the parent cell; this enables it to continue growing, and eventually divide itself.

BODY COMPOSITION

Cells are building blocks from which the human body is made. Some cells work alone—such as red blood cells, which carry oxygen—but many are organized into tissues. These tissues form organs, which in turn form specific body systems, where cells with various functions join forces to accomplish one or more tasks.

CELL TYPES

There are more than 200 types of cells in the body, each type specially adapted to its own particular function. Every cell contains the same genetic information, but not all of the genes are "switched on" in every cell. It is this pattern of gene expression that dictates the cell's appearance, its behavior, and its role in the body. A cell's fate is largely determined before birth, influenced by its position in the body and the cocktail of chemical messengers that it is exposed to in that environment. Early during development, stem cells begin to differentiate into three layers of specialized cells called the ectoderm, endoderm, and mesoderm. Cells of the ectoderm will form the skin and nails, the epithelial lining of the nose, mouth, and anus, the eyes, and the brain and spinal cord. Cells of the endoderm will become the inner linings of the digestive tract, the respiratory linings, and glandular organs. Mesoderm cells will develop into muscles, and the circulatory and excretory systems.

STEM CELLS

A few days after fertilization, an embryo consists of a ball of "embryonic stem cells" (ESCs). These cells have the potential to develop into any type of cell in the body. Scientists are trying to harness this property to grow replacement body parts. As the embryo grows, the stem cells become increasingly restricted in their potential and most are fully differentiated by the time we are born. Only a small number of stem cells remain in parts of the adult body, including in the bone marrow. Scientists believe that these cells could also be used to help cure diseases.

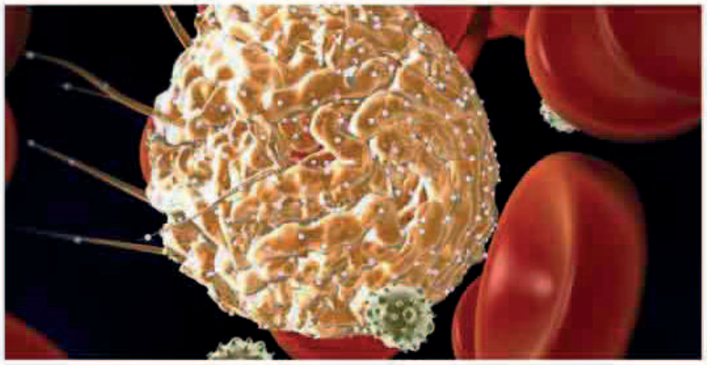

Adult stem cells
Adult stem cells, such as the large white cell in this image, are present in bone marrow, where they multiply and produce millions of blood cells, including red blood cells, also seen here.

Red blood cells
Unlike other cells, red blood cells lack a nucleus and organelles. Instead, they have an oxygen-carrying protein (hemoglobin), which gives blood its red color.

Epithelial cells
The skin cells and the cells lining the lungs and reproductive tracts are among the barrier cells, called epithelial cells, which line the cavities and surfaces of the body.

Adipose (fat) cells
These cells are highly adapted for storing fat – the bulk of their interior is taken up by a large droplet of semi-liquid fat. When we gain weight, they fill up with more fat.

Nerve cells
These electrically excitable cells transmit electrical signals down an extended stem called an axon. Found throughout the body, they enable us to feel sensations.

Photoreceptor cells
Located in the eye, these areof two types—cone and rod (left). Both have a light-sensitive pigment and generate electrical signals when struck by light, helping us see.

Smooth muscle cells
One of three types of muscle cell, smooth muscle cells are spindle-shaped cells found in the arteries and the digestive tract that produce contractions.

Ovum (egg) cells
The largest cells in the female human body, eggs are female reproductive cells. Like sperm, they have just 23 chromosomes.

Sperm cells
Sperm are male reproductive cells, with tails that enable them to swim up the female reproductive tract and fertilize an egg.

LEVELS OF ORGANIZATION

The overall organization of the human body can be visualized as a hierarchy of levels. At its lowest are the body's basic chemical constituents, forming organic molecules, such as DNA, the key to life. As the hierarchy ascends, the number of components in each of its levels —cells, tissues, organs, and systems—decreases, culminating in a single being at its apex. Cells are the smallest living units, with each adapted to carry out a specific role, but not in isolation. Groups of similar cells form tissues, which in turn form organs with a specific role. Organs with a common purpose are linked within a system, such as the cardiovascular system, shown right. These interdependent systems combine to produce a human body (see pp.16–17).

TISSUE TYPES

Cells of the same kind often group together to form tissues that carry out a specific function. However, not all cells within a tissue are necessarily identical. The four main types of tissue in the human body are muscle, connective tissue, nervous tissue, and epithelial tissue. Within these groups, different forms of these tissues can have very different appearances and functions. For example, blood, bone, and cartilage are all types of connective tissue, but so are fat layers, tendons, ligaments, and the fibrous tissue that holds organs and epithelial layers in place. Organs such as the heart and lungs are composed of several different kinds of tissue.

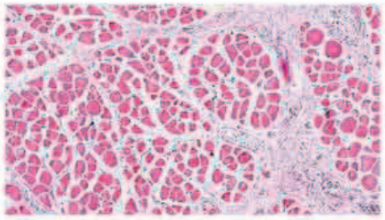

Skeletal muscle
This tissue enables voluntary limb movements. Its cells are arranged into bundles of fibers that connect to bones via tendons. They are packed with filaments that slide over one another to produce contractions.

Smooth muscle
Able to contract involuntarily in long, wavelike motions, smooth muscle is found in sheets on the walls of specific organs. It is vital for maintaining blood pressure and for pushing food through the system.

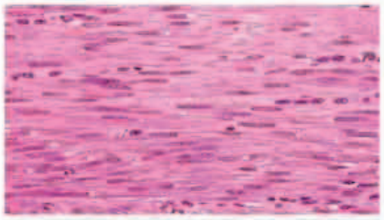

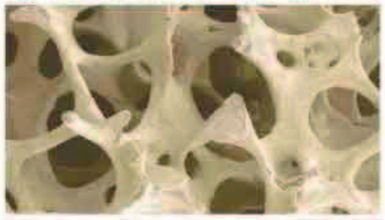

Spongy bone
Spongy bone is found in the center of bones (see p.24) and is softer and weaker than compact bone. The latticelike spaces in spongy bone are filled with bone marrow or connective tissue.

Cartilage
Its high water-content makes this tissue rubbery yet stiff. It is composed of cells, called chondrocytes, set in a matrix of gel-like materials secreted by the cells. Cartilage is found in the bone joints and in the ear and nose.

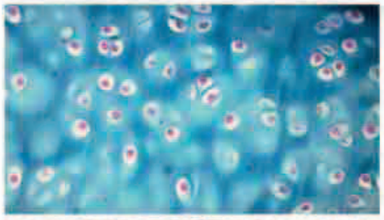

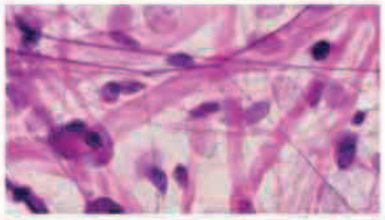

Loose connective tissue
This tissue type also contains cells called fibroblasts, which secrete loosely-organized fibers that make the tissue pliable. Loose connective tissue holds organs in place and provides support.

Dense connective tissue
This contains fibroblast cells, which secrete the fibrous protein called type 1 collagen. The fibers are organized into a regular parallel pattern, making the tissue very strong. This tissue type occurs in the base layer of skin.

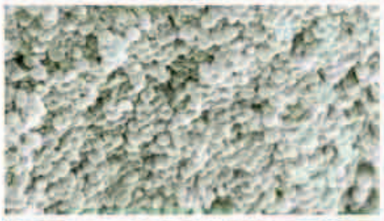

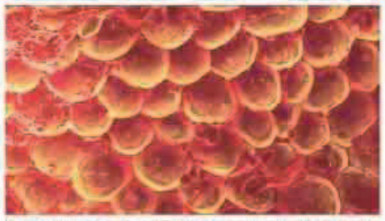

Adipose tissue
A type of connective tissue, adipose tissue is composed of fat cells called adipocytes, as well as some immune cells, fibroblast cells, and blood vessels. Its main task is to store energy, and to protect and insulate the body.

Epithelial tissue
This tissue forms a covering or lining for internal and external body surfaces. Some epithelial tissues can secrete substances such as digestive enzymes; others can absorb substances like food or water.

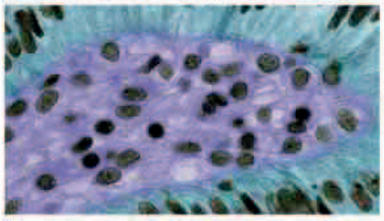

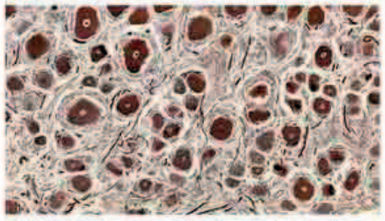

Nervous tissue
This forms the brain, spinal cord, and the nerves that control movement, transmit sensation, and regulate many body functions. It is mainly made up of networks of nerve cells (see opposite).

1. CHEMICALS

Key among the chemicals inside all cells is DNA (see pp.10–11). Its long molecules provide the instructions for making proteins. These, in turn, perform many roles, such as building cells.

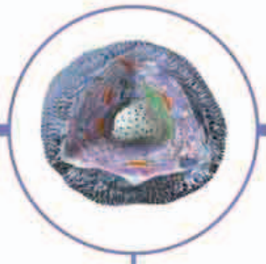

2. CELLS

While cells may differ in size and shape, all have the same basic features: an outer membrane; organelles floating within jellylike cytoplasm; and a nucleus containing DNA (see pp.12–13).

3. HEART TISSUE

One of the three types of muscle tissue, cardiac muscle is found only in the walls of the heart. Its cells contract together, as a network, to make the heart squeeze and pump blood.

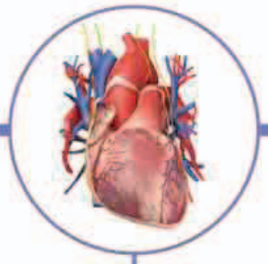

4. HEART

Like other organs, the heart is made of several types of tissue, including cardiac muscle tissue. Among the others are connective and epithelial tissues, found in the chambers and valves.

5. CARDIOVASCULAR SYSTEM

The heart, blood, and blood vessels form the cardiovascular system. Its main tasks are to pump blood, deliver nutrients, and remove waste from the tissue cells.

BODY SYSTEMS

The human body can do many different things. It can digest food, think, move, even reproduce and create new life. Each of these tasks is performed by a different body system—a group of organs and tissues working together to complete that task. However, good health and body efficiency rely on the different body systems working together in harmony.

SYSTEM INTERACTION

Think about what your body is doing right now. You are breathing, your heart is beating, and your blood pressure is under control. You are also conscious and alert. If you were to start running, specialized cells called chemoreceptors would detect a change in your body's metabolic requirements and signal to the brain to release adrenaline. This would in turn signal the heart to beat faster, boosting blood circulation and providing more oxygen to the muscles. After a while, cells in the hypothalamus might detect an increase in temperature and send a signal to the skin to produce sweat, which would evaporate and cool you down.

The individual body systems are linked together by a vast network of positive and negative feedback loops. These use signalling molecules such as hormones and electrical impulses from nerves to maintain equilibrium. Here, the basic components and functions of each system are described, and examples of system interactions are examined.

LYMPHATIC AND IMMUNE SYSTEM

The lymphatic system includes a network of vessels and nodes, which drain tissue fluid and return it to the veins. Its main functions are to maintain fluid balance within the cardiovascular system and to distribute immune cells around the body. Movement of lymphatic fluid relies on the muscles within the muscular system.

ENDOCRINE SYSTEM

The endocrine system communicates with the other systems, enabling them to be monitored and controlled. It uses chemical messengers, called hormones, which are secreted into the blood by specialized glands.

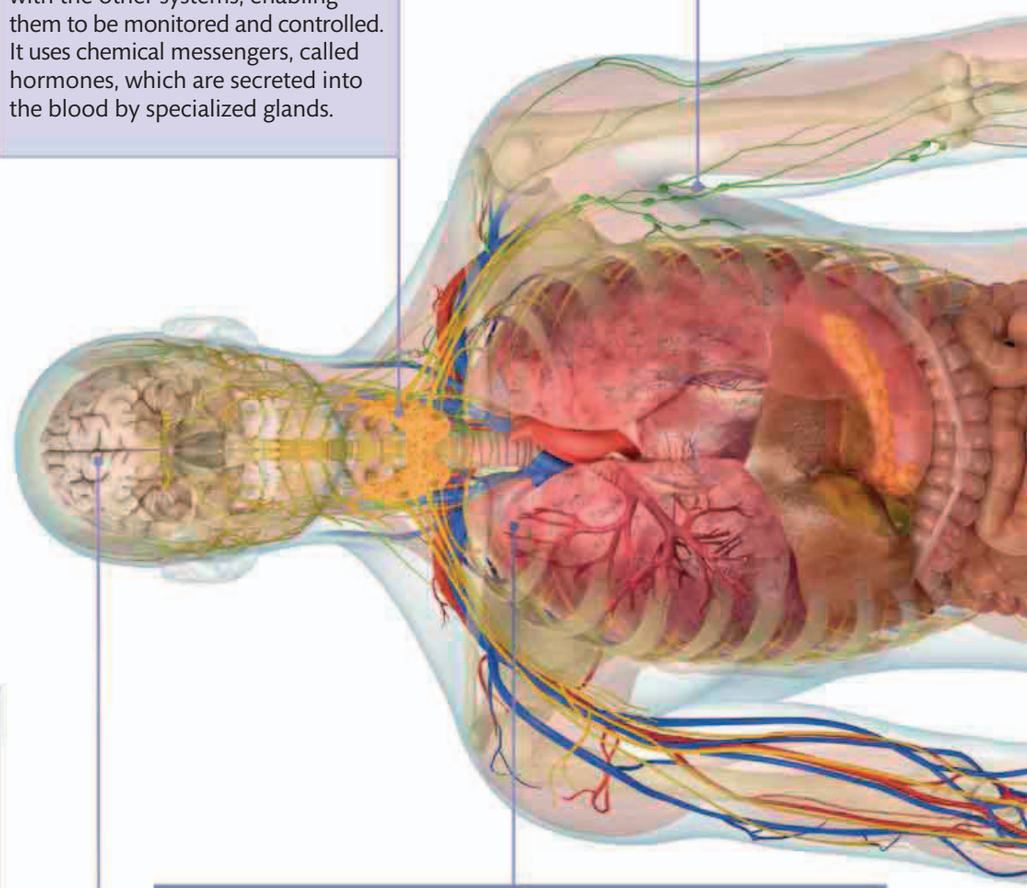

BREATHING IN AND OUT

The mechanics of breathing rely upon an interaction between the respiratory and muscular systems. Together with three accessory muscles, the intercostal muscles and the diaphragm contract to increase the volume of the chest cavity. This draws air down into the lungs. A different set of muscles is used during forced exhalation. These rapidly compress the chest cavity, forcing air out of the lungs.

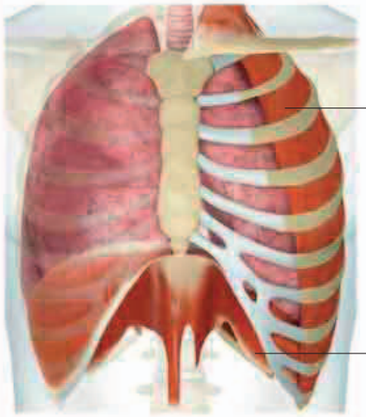

RESPIRATORY SYSTEM

Every cell in the body needs oxygen and must dispel carbon dioxide in order to function. The respiratory system ensures this by breathing air into the lungs, where the exchange of these molecules occurs between the air and blood. The cardiovascular system transports oxygen and carbon dioxide between the cells and the lungs.

NERVOUS SYSTEM

The brain, spinal cord, and nerves collect, process, and disseminate information from the body's internal and external environments. The nervous system communicates through networks of nerve cells, which connect with other systems. The brain controls and monitors all the other systems to ensure they are performing normally.

DIGESTIVE SYSTEM

As well as oxygen, every cell needs energy in order to function. The digestive system processes and breaks down the food we eat so that a variety of nutrients can be absorbed from the intestines into the circulatory system. These are then delivered to the cells of every body system in order to provide them with energy.

MUSCULAR SYSTEM

The muscular system is made up of three types of muscle: skeletal, smooth, and cardiac. It is responsible for generating movement – both in the limbs and within the other body systems. For example, smooth muscle aids the digestive system by helping to propel food down the esophagus and through the stomach, intestines, and rectum. The respiratory system needs the thoracic muscles to contract to fill the lungs with air (see opposite).

SKELETAL SYSTEM

This system uses bones, cartilage, ligaments, and tendons to provide the body with structural support and protection. It encases much of the nervous system within a skull and vertebrae, and the vital organs of the respiratory and circulatory systems within the ribcage. The skeletal system also supports our immune and the circulatory systems by manufacturing red and white blood cells.

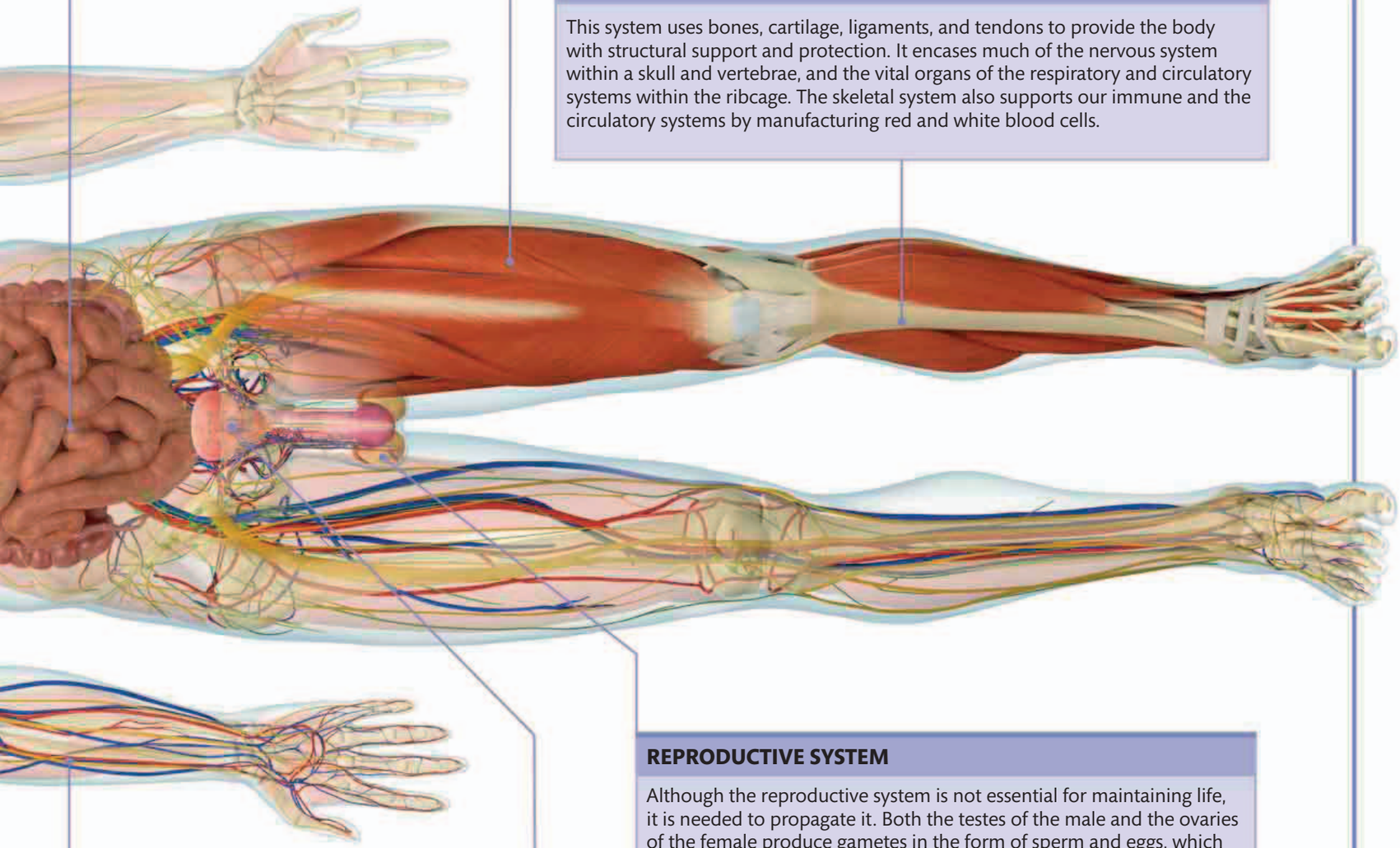

REPRODUCTIVE SYSTEM

Although the reproductive system is not essential for maintaining life, it is needed to propagate it. Both the testes of the male and the ovaries of the female produce gametes in the form of sperm and eggs, which fuse to create an embryo. The testes and ovaries also produce hormones including estrogen and testosterone, thus forming part of the endocrine system.

CARDIOVASCULAR SYSTEM

The cardiovascular system uses blood to carry oxygen from the respiratory system and nutrients from the digestive system to cells of all the body's systems. It also removes waste products from these cells. At the center of the cardiovascular system lies the muscular heart, which pumps blood through the blood vessels.

URINARY SYSTEM

The urinary system filters and removes many of the waste products generated by cells of the body. It does this by filtering blood through the kidneys and producing urine, which is collected in the bladder and then excreted through the urethra. The kidneys also help regulate blood pressure within the cardiovascular system by ensuring that the correct amount of water is reabsorbed by the blood.

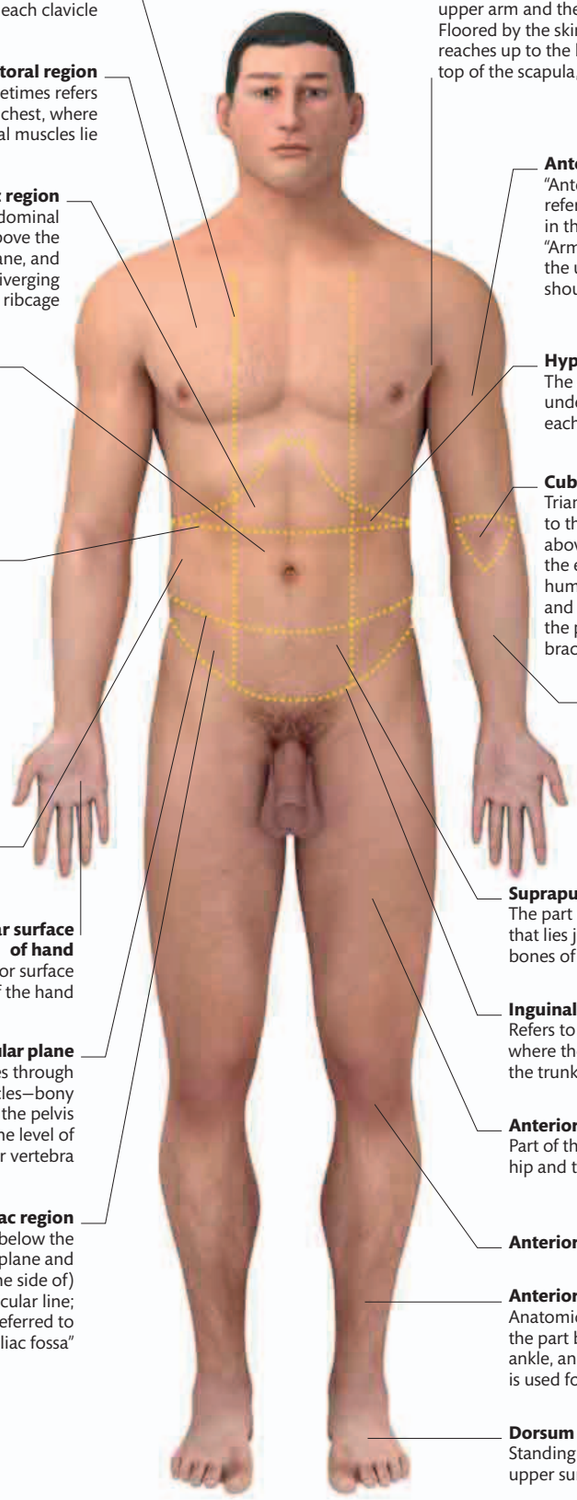

Midclavicular line
A vertical line running down from the midpoint of each clavicle

Pectoral region
The chest; sometimes refers to just the upper chest, where the pectoral muscles lie

Epigastric region
Area of the abdominal wall above the transpyloric plane, and framed by the diverging margins of the ribcage

Umbilical region
Central region of the abdomen, around the umbilicus (navel)

Transpyloric plane
Horizontal plane joining tips of the ninth costal cartilages, at the margins of the ribcage; level with the first lumbar vertebra and pylorus of the stomach

Lumbar region
The side of the abdominal wall

Palmar surface of hand
Anterior surface of the hand

Intertubercular plane
This plane passes through the iliac tubercles—bony landmarks on the pelvis—and lies at the level of the fifth lumbar vertebra

Iliac region
The area below the intertubercular plane and lateral to (to the side of) the midclavicular line; may also be referred to as the "iliac fossa"

Axilla
The armpit; more precisely, the pyramid-shaped part of the body between the upper arm and the side of the thorax. Floored by the skin of the armpit, it reaches up to the level of the clavicle, top of the scapula, and first rib

Anterior surface of arm
"Anterior" means front, and refers to the body when it is in this "anatomical position". "Arm" relates to the part of the upper limb between the shoulder and the elbow

Hypochondrial region
The abdominal region under the ribs on each side

Cubital fossa
Triangular area anterior to the elbow, bounded above by a line between the epicondyles of the humerus on each side, and framed below by the pronator teres and brachioradialis muscles

Anterior surface of forearm
The part of the body between the elbow and the wrist

Suprapubic region
The part of the abdomen that lies just above the pubic bones of the pelvis

Inguinal region
Refers to the groin area, where the thigh meets the trunk

Anterior surface of thigh
Part of the body between the hip and the knee

Anterior surface of knee

Anterior surface of leg
Anatomically, "leg" just refers to the part between the knee and ankle, and the term "lower limb" is used for the whole limb

Dorsum of foot
Standing upright, this is the upper surface of the foot

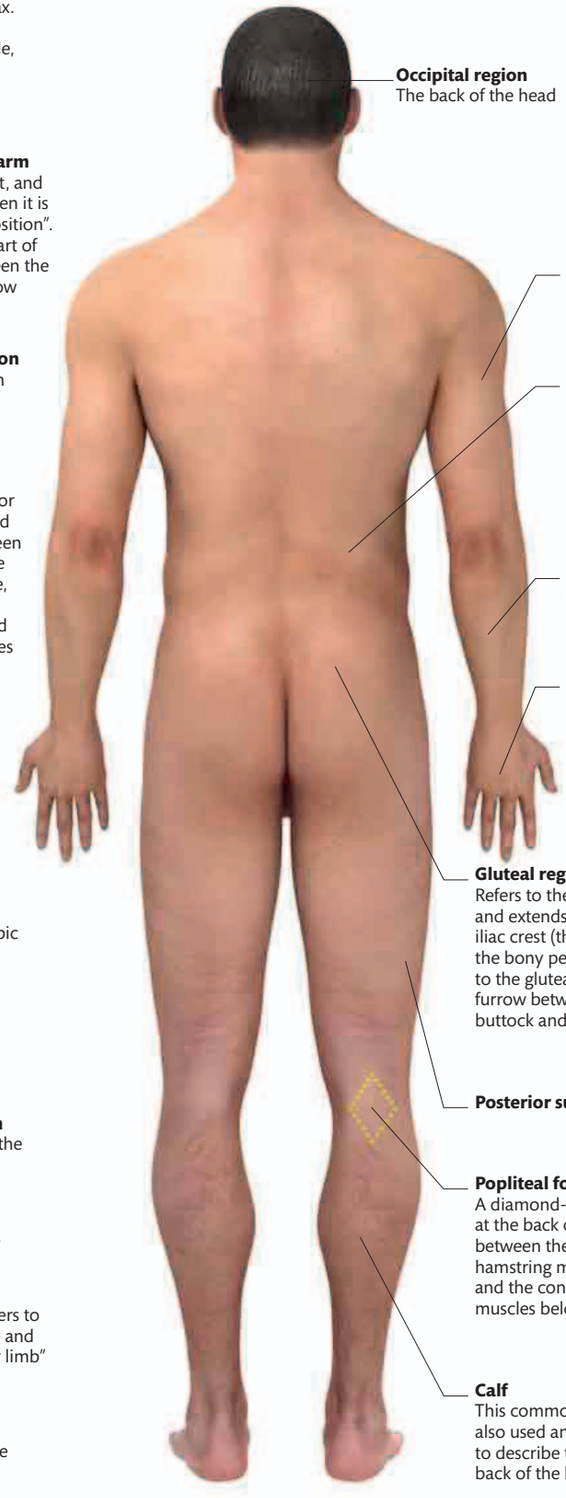

Occipital region
The back of the head

Posterior surface of arm

Lumbar region
On the back of the body it refers to the part between the thorax and the pelvis

Posterior surface of forearm

Dorsum of hand
The back of the hand

Gluteal region
Refers to the buttock, and extends from the iliac crest (the top of the bony pelvis) above, to the gluteal fold (the furrow between the buttock and thigh) below

Posterior surface of thigh

Popliteal fossa
A diamond-shaped cavity at the back of the knee, between the diverging hamstring muscles above and the converging calf muscles below

Calf
This common term is also used anatomically, to describe the fleshy back of the leg

Anterior surface regions

The anterior surface of the body is divided into general anatomical areas by imaginary lines drawn on the body. The location of many of these lines is defined by reference to underlying features such as muscles or bony prominences; for example, the cubital fossa is defined by reference to epicondyles of the humerus, and the pronator teres and brachioradialis muscles. Many of the regions may be divided into smaller areas. For instance, the upper part of the anterior thigh contains the femoral triangle.

Posterior surface regions

As with the anterior surface, the posterior surface can also be divided into anatomical regions. The anterior surface of the abdomen is divided by planes and mapped into nine regions—allowing doctors to describe precisely where areas of tenderness or lumps are felt on abdominal examination. The back is not divided into as many regions. This illustration shows some of the terms used for the broader regions of back of the body.

TERMINOLOGY AND PLANES

Anatomical language allows us to describe the structure of the body accurately and unambiguously. The illustrations here show the main regions of the anterior (front) and posterior (back) surfaces of the body. Sometimes it is easier to understand anatomy by dividing the body into two dimensional slices. The orientation of these planes through the body also have specific anatomical names. There are also terms to describe the relative position of structures within the body.

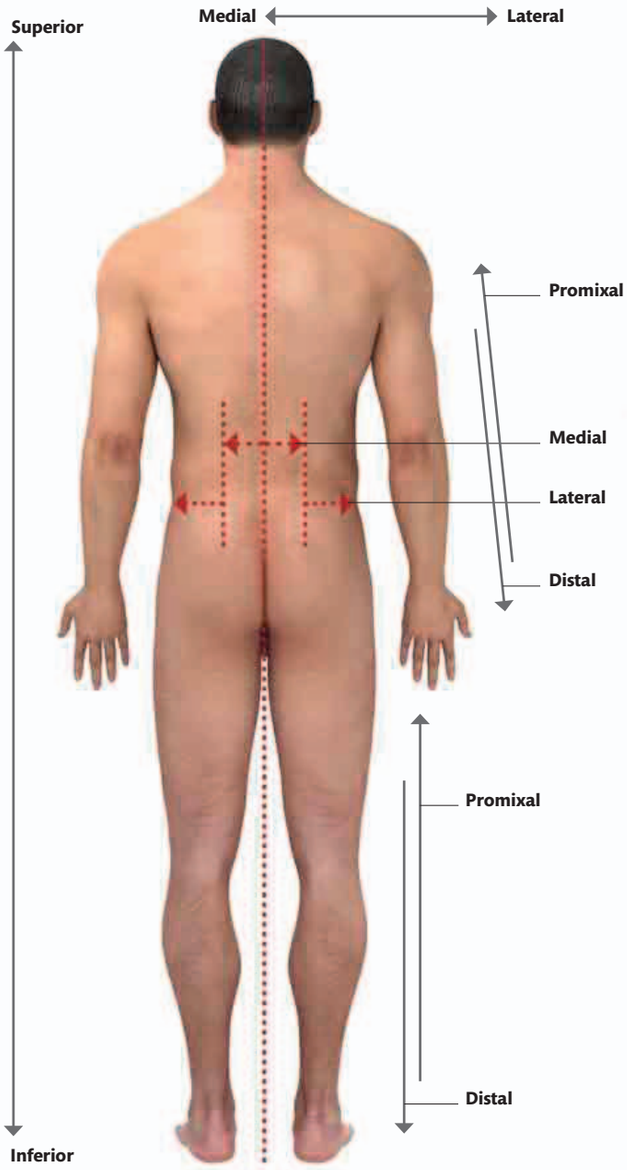

Directions and relative positions

As well as defining parts of the body, anatomical terminology also allows us to precisely and concisely describe the relative positions of various structures. These terms always refer back to relative positions of structures when the body is in the "anatomical position" (shown above). Medial and lateral describe positions of structures toward the midline, or toward the side of the body, respectively. Superior and inferior refer to vertical position—toward the top or bottom of the body. Proximal and distal are useful terms, describing a relative position toward the center or periphery of the body.

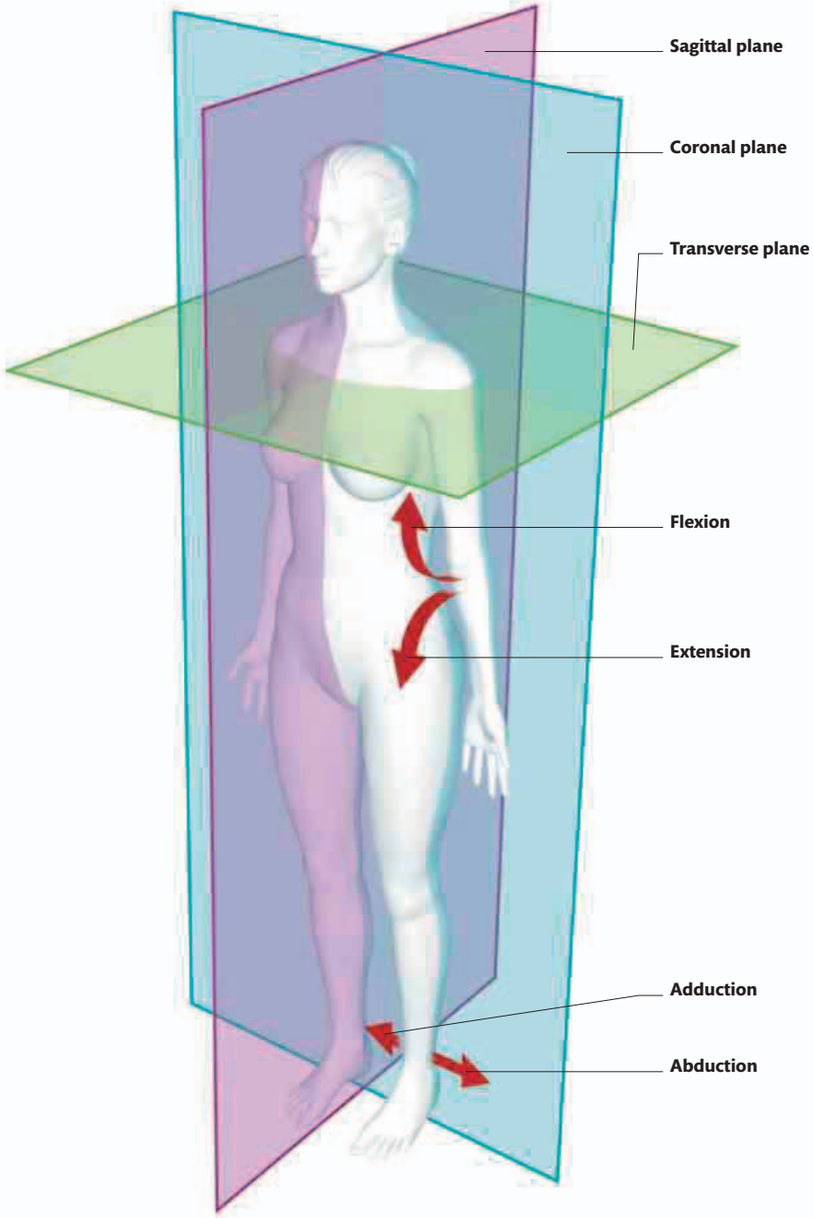

Anatomical terms for planes and movement

The diagram above shows the three planes—sagittal, coronal, and transverse—cutting through a body. It also illustrates some medical terms that are used to describe certain movements of body parts: flexion decreases the angle of a joint, such as the elbow, while extension increases it; adduction draws a limb closer to the sagittal plane, while abduction moves it further away from that plane.

02 Body Systems

The human body is made up of eleven functional systems. No one system works in isolation: for example the endocrine and nervous systems work closely to keep the body regulated, while the respiratory and cardiovascular systems combine to deliver vital oxygen to cells. To build the clearest picture of how the body is put together it is, however, helpful to strip back our anatomy and consider it system by system. This chapter gives an overview of the basic structure of each system before looking at each region in detail.

022 Skin, hair, and nail
024 Skeletal system
068 Muscular system
110 Nervous system
146 Respiratory system
154 Cardiovascular system
180 Lymphatic and immune system
192 Digestive system
204 Urinary system
208 Reproductive system
216 Endocrine system

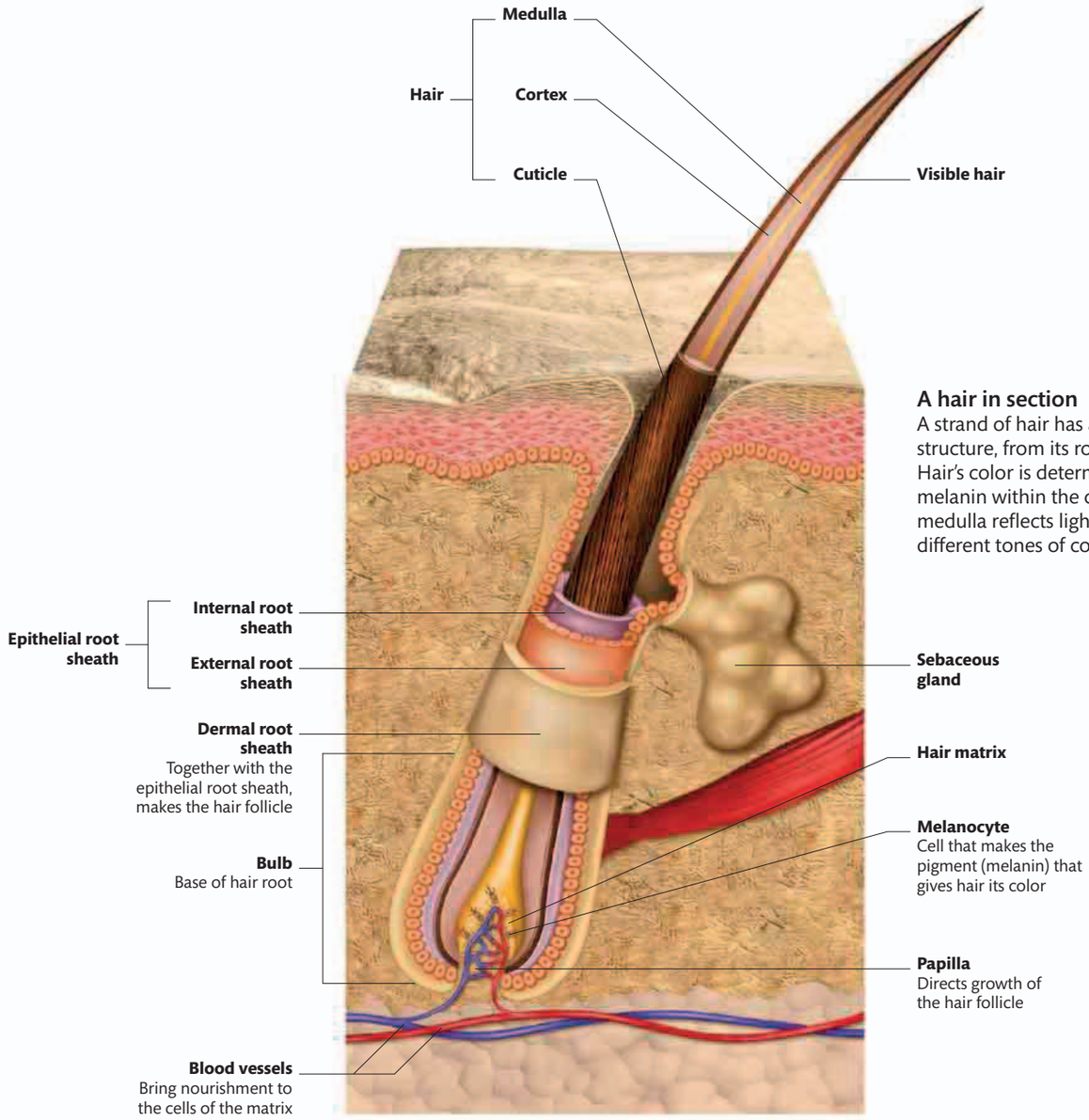

A hair in section
A strand of hair has a multilayered structure, from its root to the tip. Hair's color is determined by melanin within the cortex; the medulla reflects light so the different tones of color are seen.

SECTION THROUGH A HAIR

SKIN, HAIR, AND NAIL

The skin is our largest organ, weighing about 9lb (4kg) and covering an area of about 21 square feet (2 square meters). It forms a tough, waterproof layer, which protects us from the elements. However, it offers much more than protection: the skin lets us appreciate the texture and temperature of our environment; it regulates body temperature; it allows excretion in sweat; communication through blushing; gripping thanks to ridges on our fingertips, and vitamin D production in sunlight. Thick head hairs and fine body hairs help to keep us warm and dry. All visible hair is in fact dead; hairs are only alive at their root. Constantly growing and self-repairing, nails protect fingers and toes but also enhance their sensitivity.

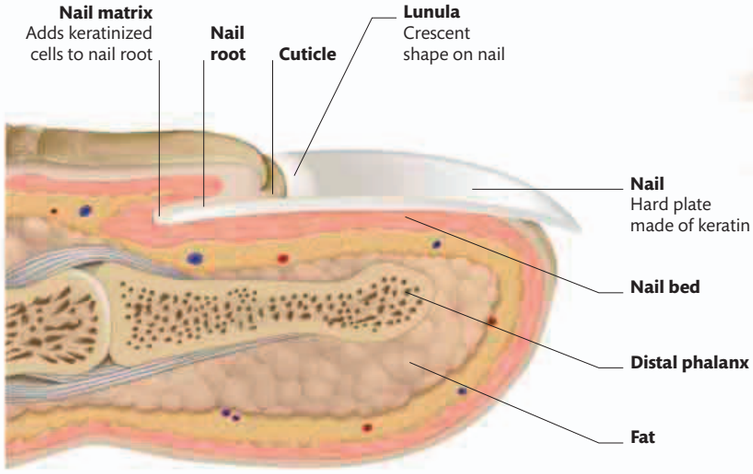

SECTION THROUGH A NAIL

SECTION THROUGH SKIN

Skin in section

In just one square centimeter (⅙in²) of the skin, there are, on average, 21½in (55cm) of nerve fibers, 27½in (70cm) of blood vessels, 15 sebaceous glands, 100 sweat glands, and over 200 sensory receptors.

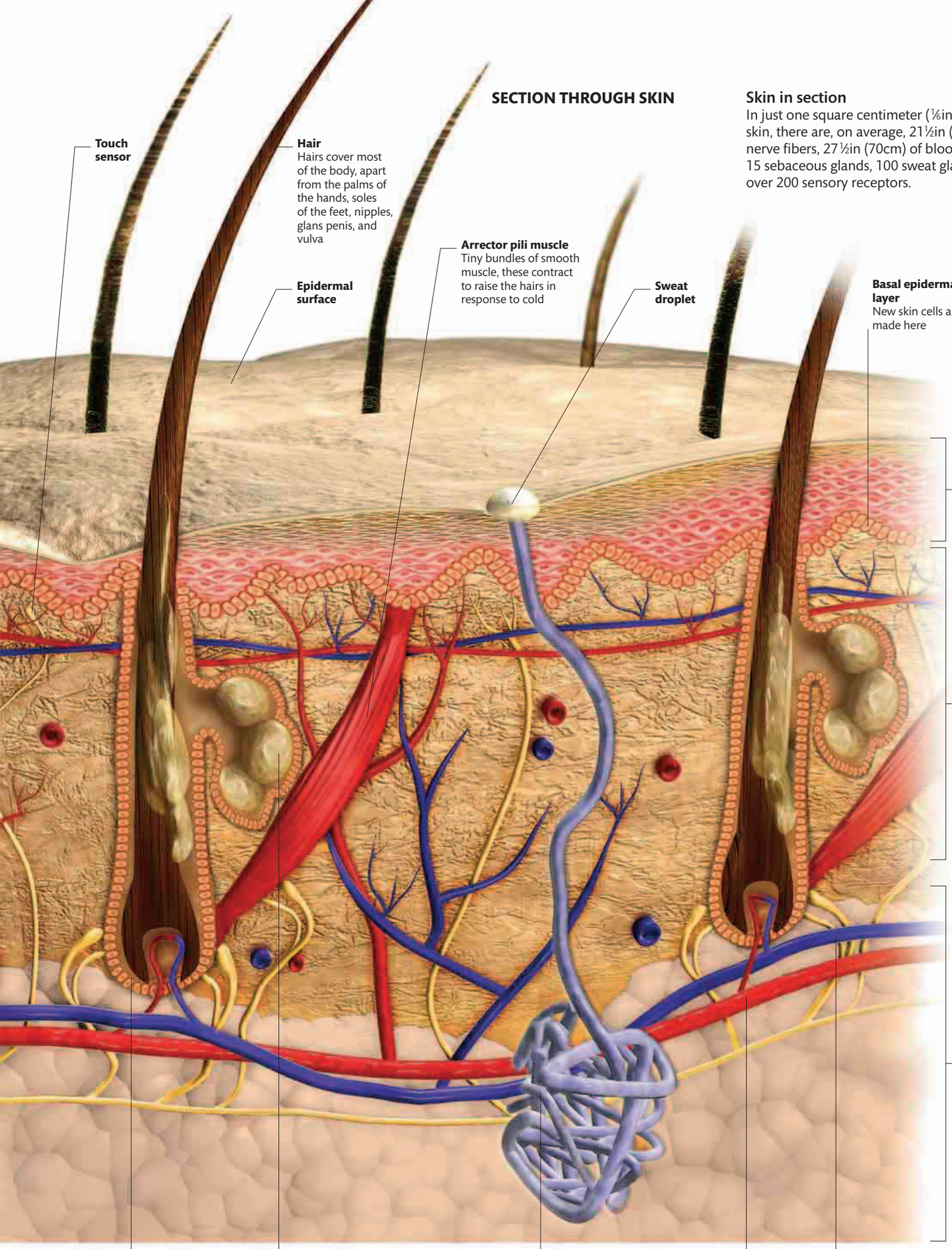

SKELETAL SYSTEM OVERVIEW

The human skeleton gives the body its shape, supports the weight of all our other tissues, provides attachment for muscles, and forms a system of linked levers that the muscles can move. It also protects delicate organs and tissues, such as the brain within the skull, the spinal cord within the arches of the vertebrae, and the heart and lungs within the ribcage. The skeletal system differs between the sexes. This is most obvious in the pelvis, which is usually wider in a woman than in a man. The skull also varies, with men having a larger brow and more prominent areas for muscle attachment on the back of the head. The entire skeleton tends to be larger and more robust in a man.

BONE STRUCTURE

Most of the human skeleton develops first as cartilage, which is later replaced by bone throughout fetal development and childhood. Both bone and cartilage are connective tissues. Bone tissue consists of cells that are embedded in a mineralized matrix, making it extremely hard and strong. Bone is full of blood vessels and repairs easily.

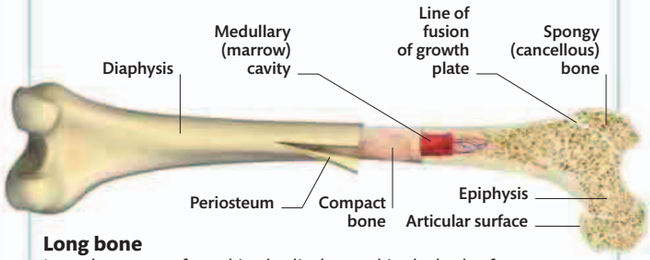

Long bone
Long bones are found in the limbs, and include the femur (shown above), humerus, radius, ulna, tibia, fibula, metatarsals, metacarpals, and phalanges. A long bone has flared ends (epiphyses), which narrow to form a neck (metaphysis), tapering down into a cylindrical shaft (diaphysis). Cartilage growth plates near the ends of bones allow rapid growth in childhood, but disappear by adulthood.

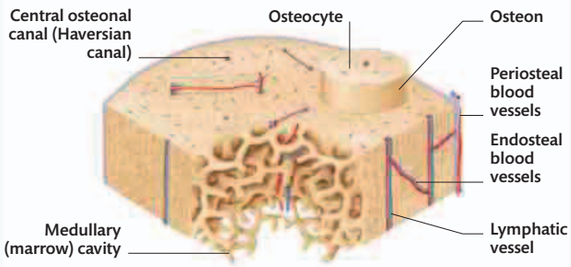

Compact bone
Also called cortical bone, compact bone is made up of osteons: concentric cylinders of bone tissue, each around 0.1–0.4mm in diameter, with a central vascular canal. Bone is full of blood vessels: those in the osteons connect to blood vessels within the medullary cavity of the bone as well as to vessels in the periosteum on the outside.

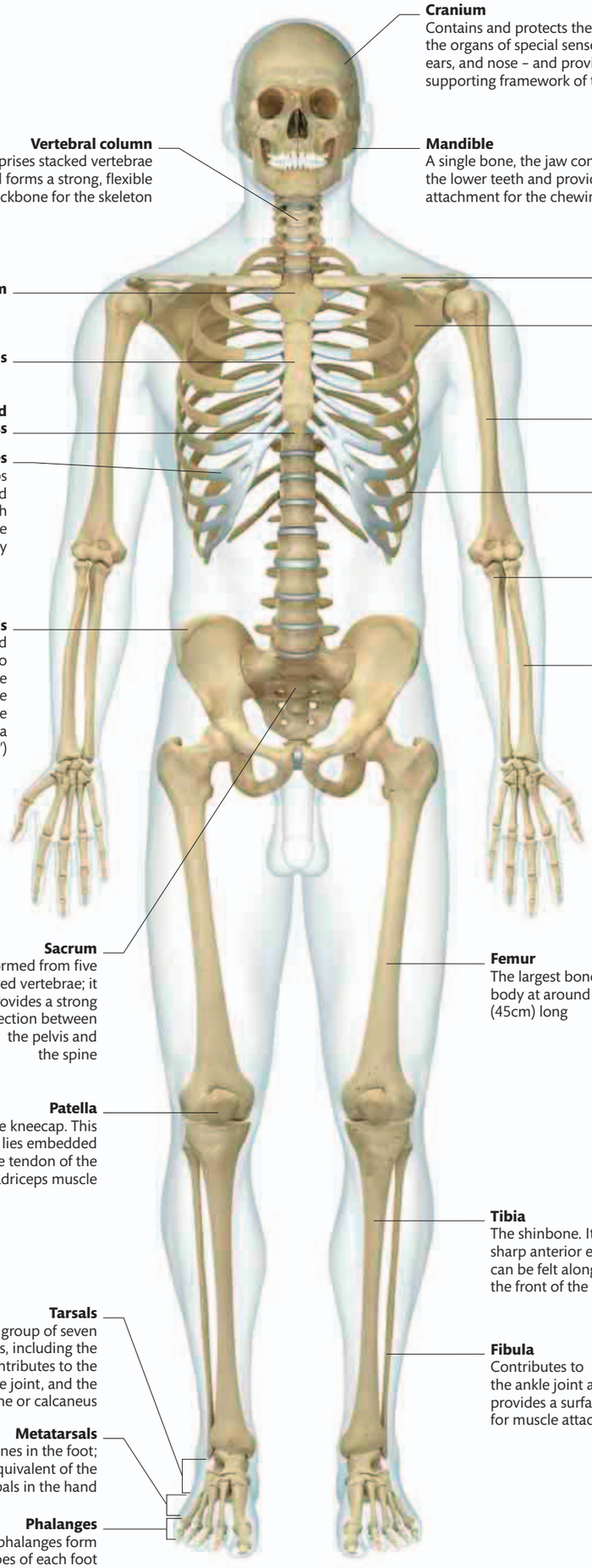

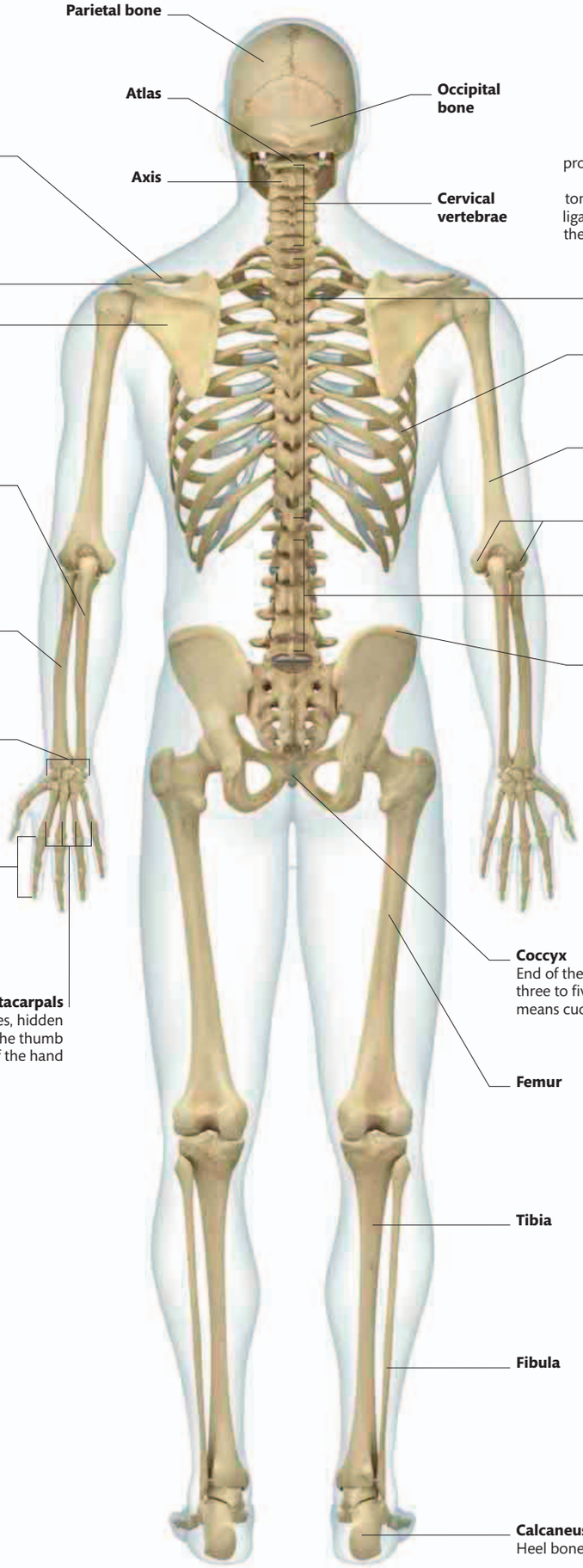

POSTERIOR (BACK)

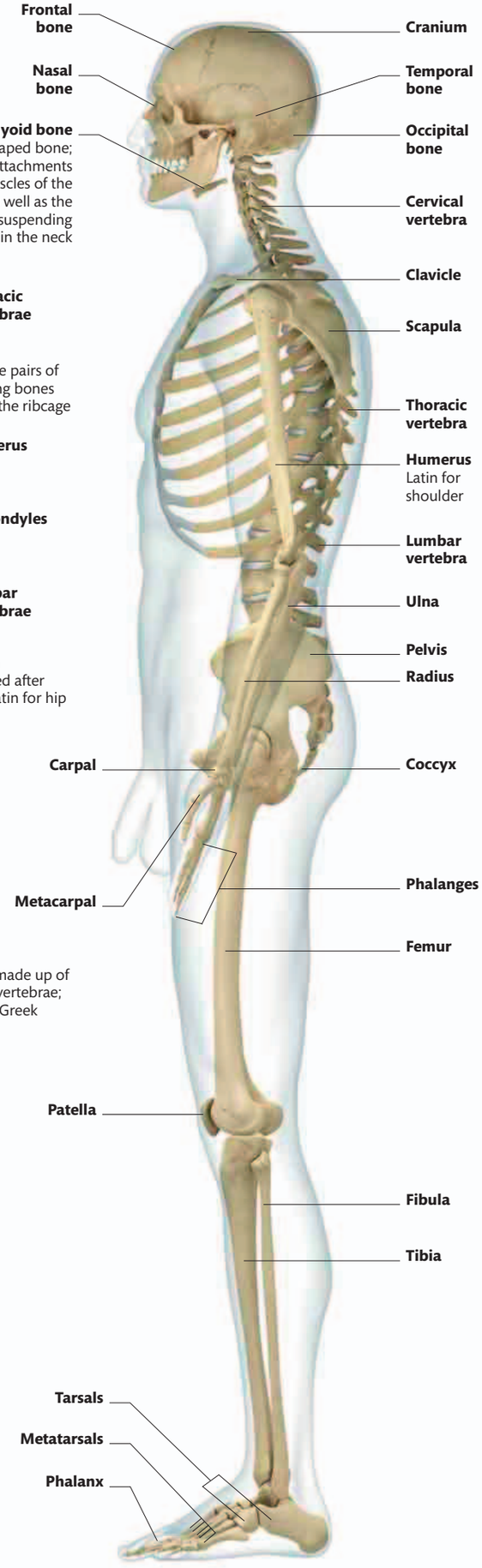

SIDE

HEAD AND NECK

The skull comprises the cranium and mandible. It houses and protects the brain and the eyes, ears, nose, and mouth. It encloses the first parts of the airway and of the alimentary canal, and provides attachment for the muscles of the head and neck. The cranium itself comprises more than 20 bones that meet each other at fibrous joints called sutures. In addition to the main bones labeled on these pages, there are sometimes extra bones along the sutures. In a young adult skull, the sutures are visible as tortuous lines between the cranial bones; they gradually fuse with age. The mandible of a newborn baby is in two halves, with a fibrous joint in the middle. The joint fuses during early infancy, so that the mandible becomes a single bone.

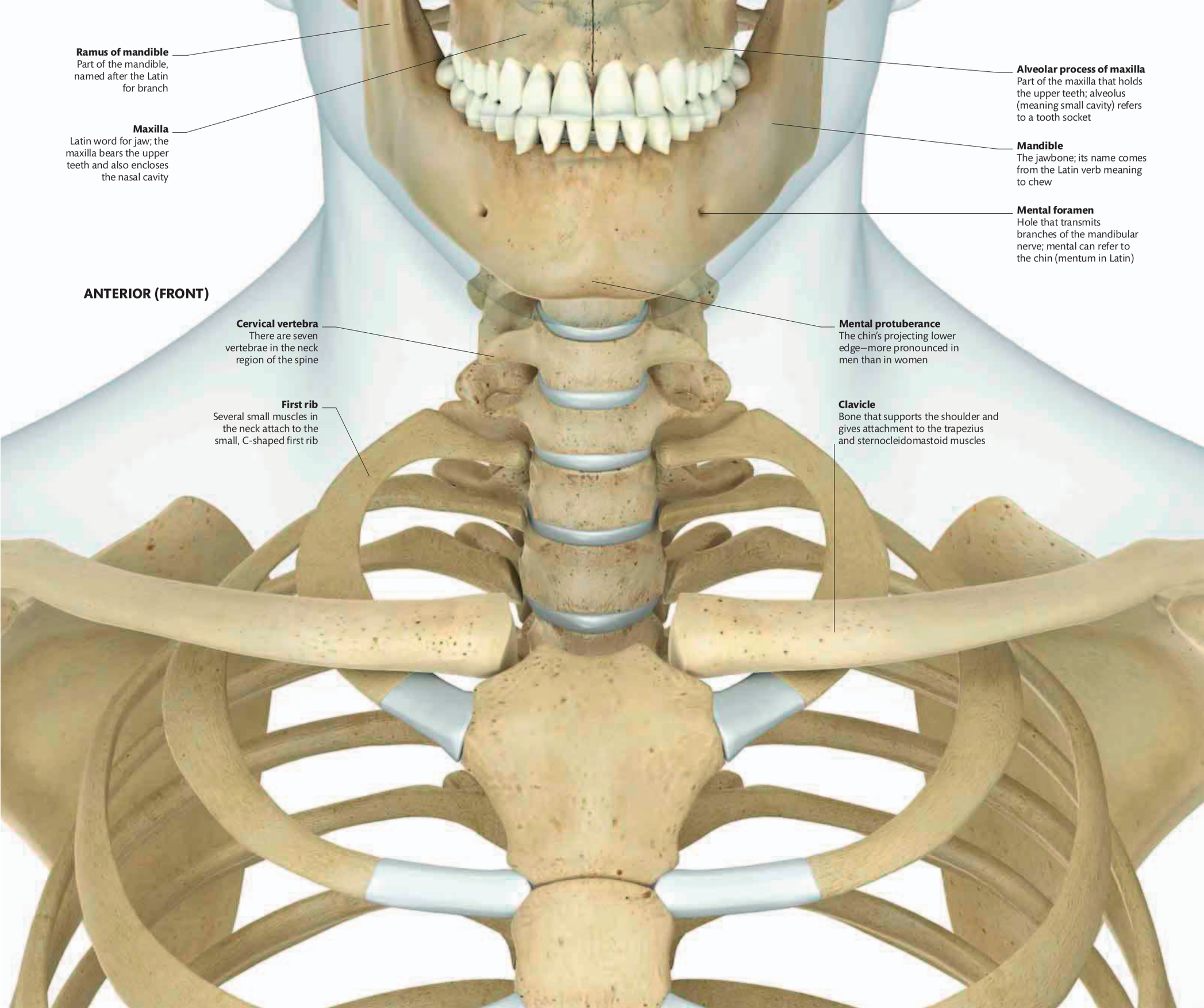
ANTERIOR (FRONT)
Ramus of mandible
Part of the mandible, named after the Latin for branch
Maxilla
Latin word for jaw; the maxilla bears the upper teeth and also encloses the nasal cavity
Alveolar process of maxilla
Part of the maxilla that holds the upper teeth; alveolus (meaning small cavity) refers to a tooth socket
Mandible
The jawbone; its name comes from the Latin verb meaning to chew
Mental foramen
Hole that transmits branches of the mandibular nerve; mental can refer to the chin (mentum in Latin)
Cervical vertebra
There are seven vertebrae in the neck region of the spine
Mental protuberance
The chin's projecting lower edge—more pronounced in men than in women
First rib
Several small muscles in the neck attach to the small, C-shaped first rib
Clavicle
Bone that supports the shoulder and gives attachment to the trapezius and sternocleidomastoid muscles

HEAD AND NECK

The cervical spine includes seven vertebrae, the top two of which have specific names. The first vertebra, which supports the skull, is called the atlas, after the Greek god who carried the sky on his shoulders. Nodding movements of the head occur at the joint between the atlas and the skull. The second cervical vertebra is the axis, from the Greek word for axle, so-called because when you shake your head from side to side, the atlas rotates on the axis. In this side view, we can also see more of the bones that make up the cranium, as well as the temporomandibular (jaw) joint between the mandible and the skull. The hyoid bone is also visible. This small bone is a very important anchor for the muscles that form the tongue and the floor of the mouth, as well as muscles that attach to the larynx and pharynx.

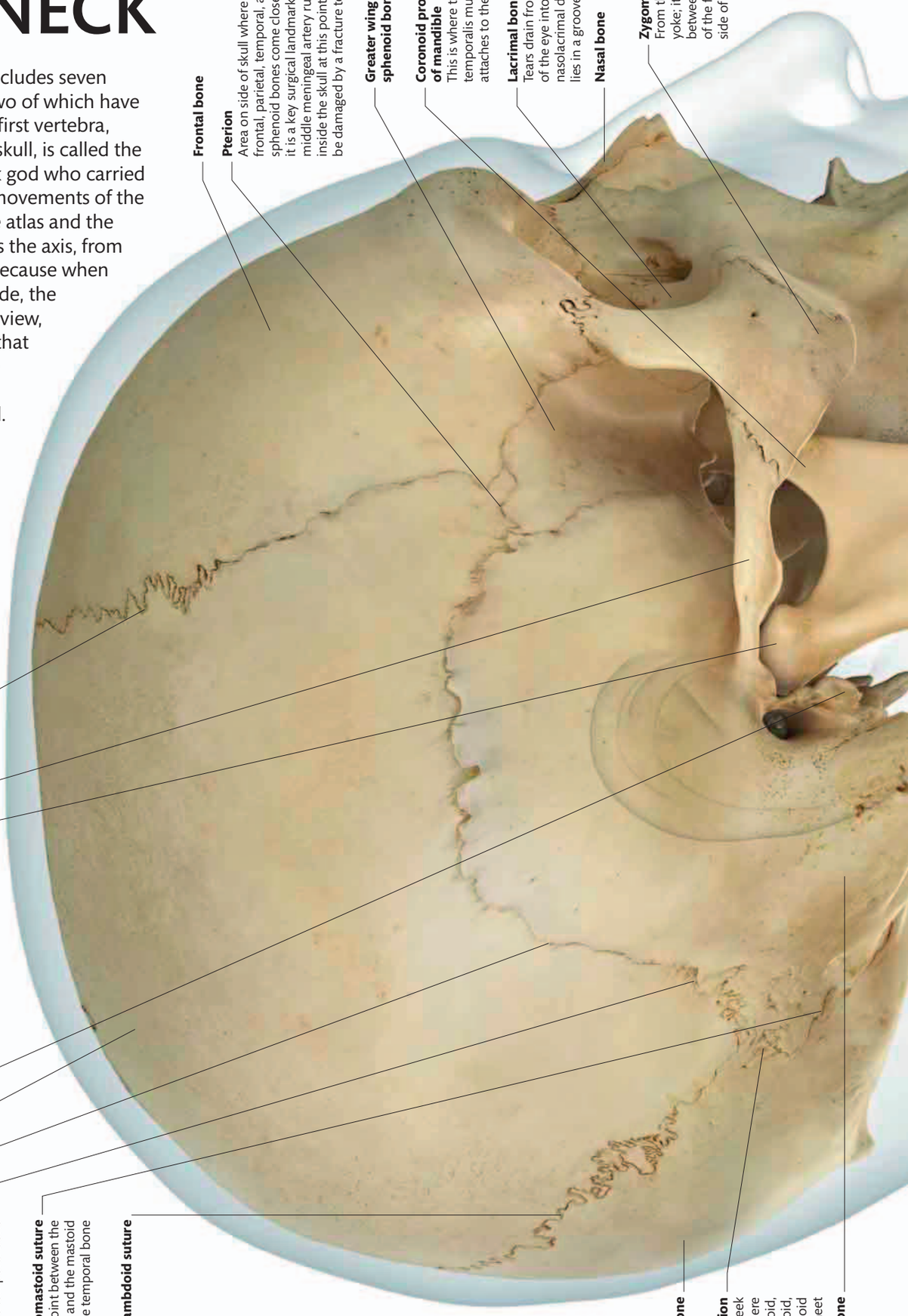

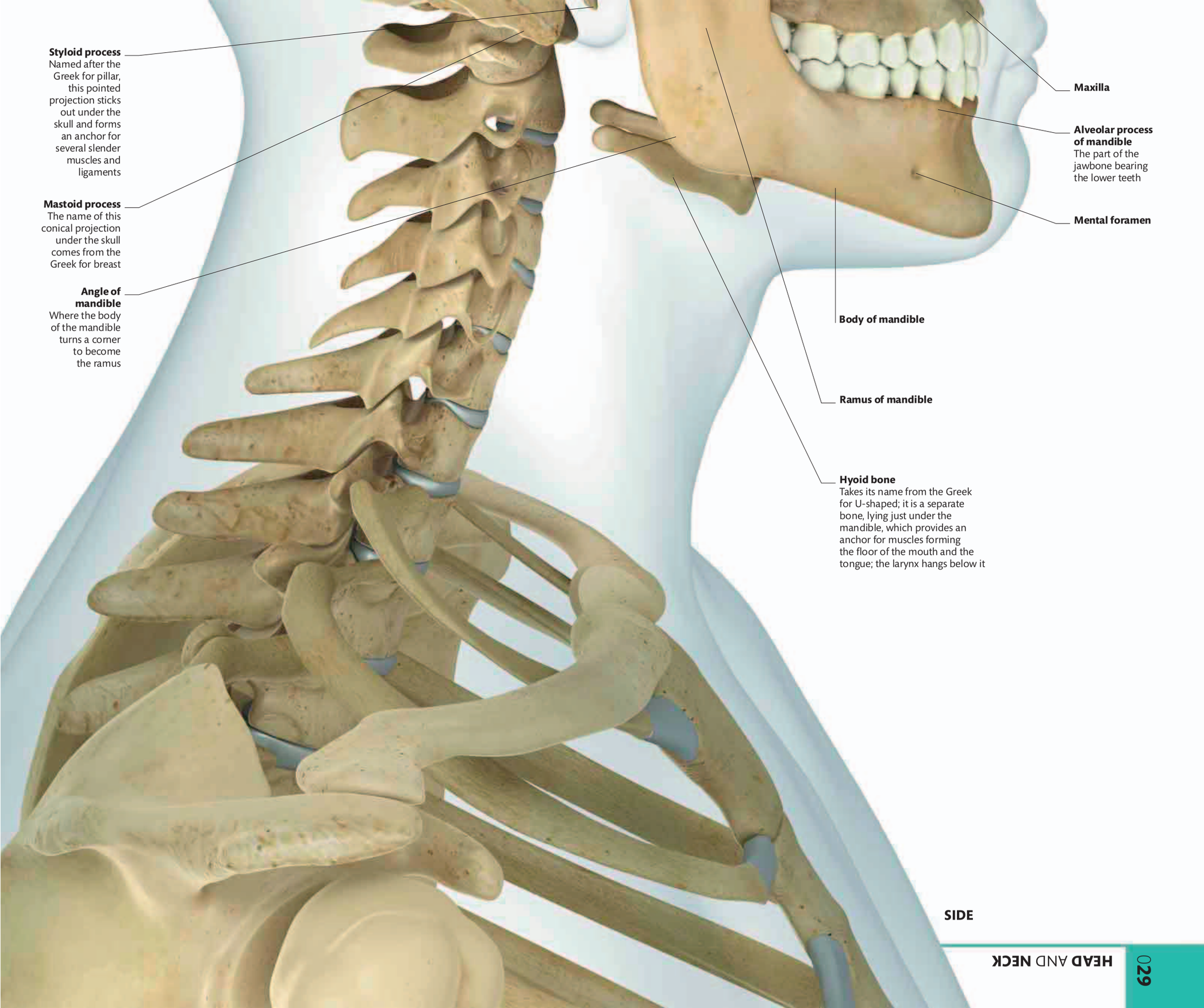
Styloid process
Named after the Greek for pillar, this pointed projection sticks out under the skull and forms an anchor for several slender muscles and ligaments
Mastoid process
The name of this conical projection under the skull comes from the Greek for breast
Angle of mandible
Where the body of the mandible turns a corner to become the ramus
Maxilla
Alveolar process of mandible
The part of the jawbone bearing the lower teeth
Mental foramen
Body of mandible
Ramus of mandible
Hyoid bone
Takes its name from the Greek for U-shaped; it is a separate bone, lying just under the mandible, which provides an anchor for muscles forming the floor of the mouth and the tongue; the larynx hangs below it

SIDE

HEAD AND NECK

The most striking features of the skull viewed from these angles are the holes in it. In the middle, there is one large hole—the foramen magnum—through which the brain stem emerges to become the spinal cord. But there are also many smaller holes, most of them paired. Through these holes, the cranial nerves from the brain escape to supply the muscles, skin, and mucosa, and the glands of the head and neck. Blood vessels also pass through some holes, on their way to and from the brain. At the front, we can also see the upper teeth sitting in their sockets in the maxillae, and the bony, hard palate.

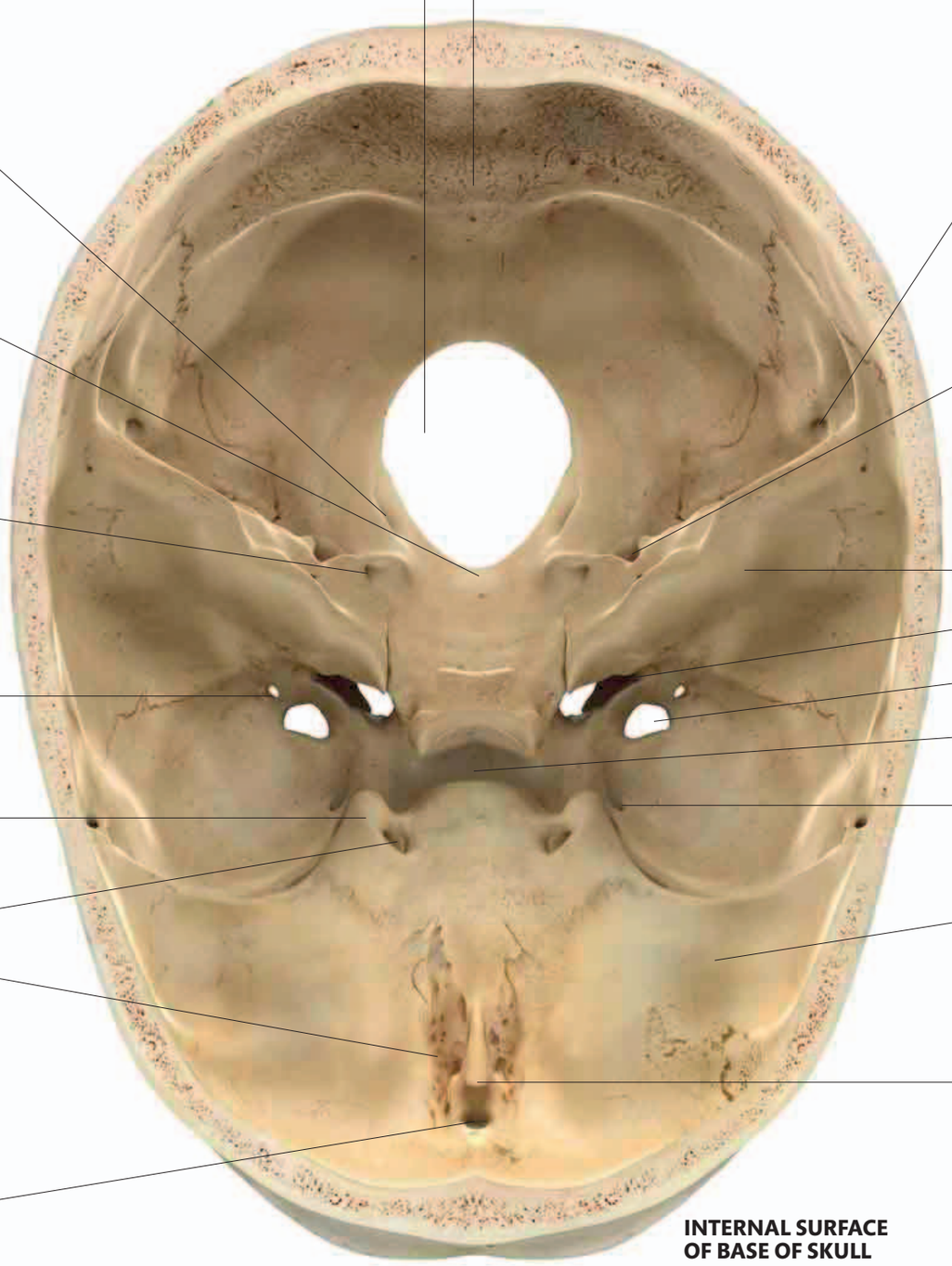

INTERNAL SURFACE OF BASE OF SKULL

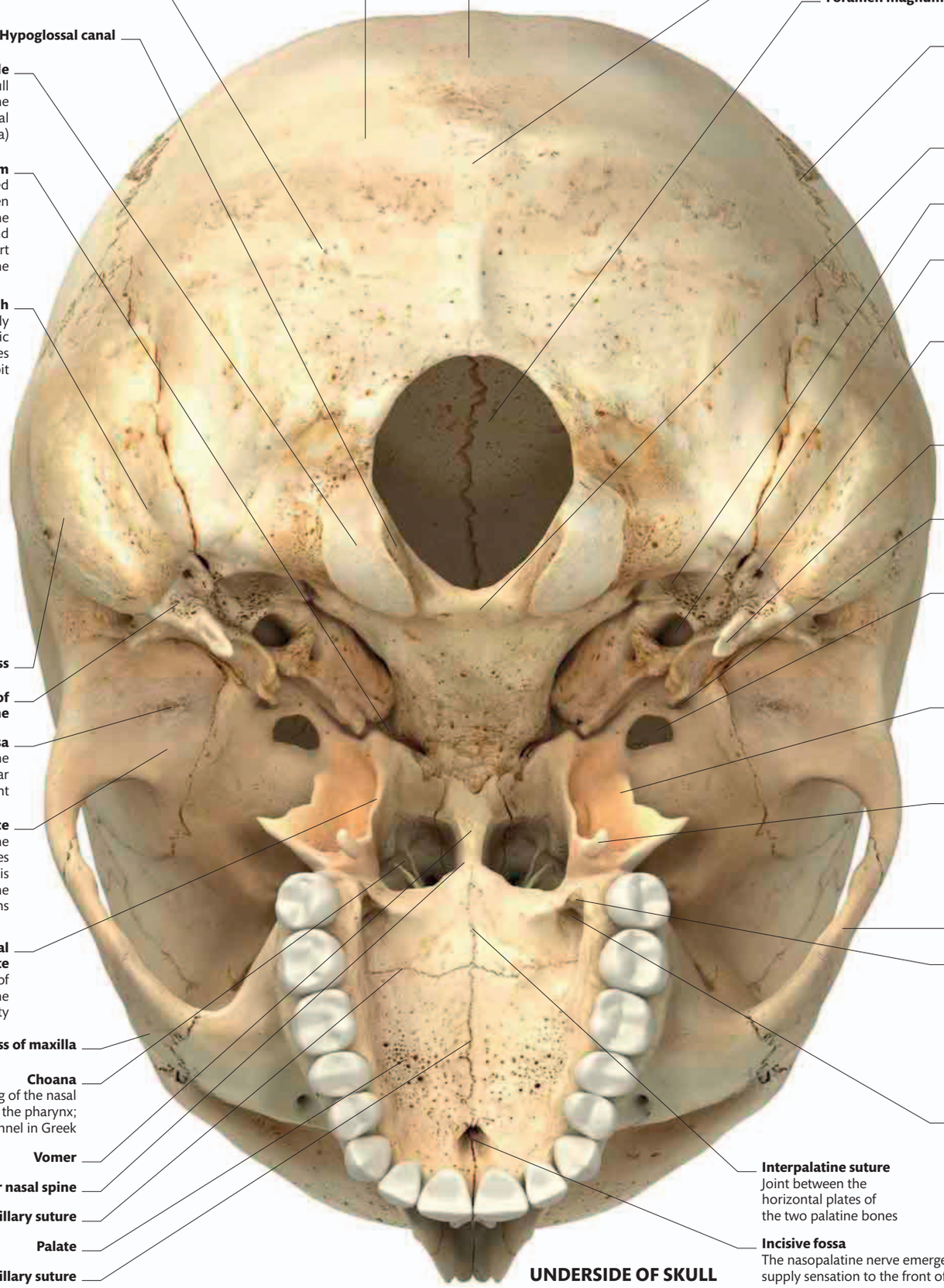

UNDERSIDE OF SKULL

HEAD AND NECK

This section—right through the middle of the skull—lets us in on some secrets. We can clearly appreciate the size of the cranial cavity, which is almost completely filled by the brain, with just a small gap for membranes, fluid, and blood vessels. Some of those blood vessels leave deep grooves on the inner surface of the skull: we can trace the course of the large venous sinuses and the branches of the middle meningeal artery. We can also see that the skull bones are not solid, but contain trabecular bone (or diploe), which itself contains red marrow. Some skull bones also contain air spaces, like the sphenoidal sinus visible here. We can also appreciate the large size of the nasal cavity, hidden away inside the skull.

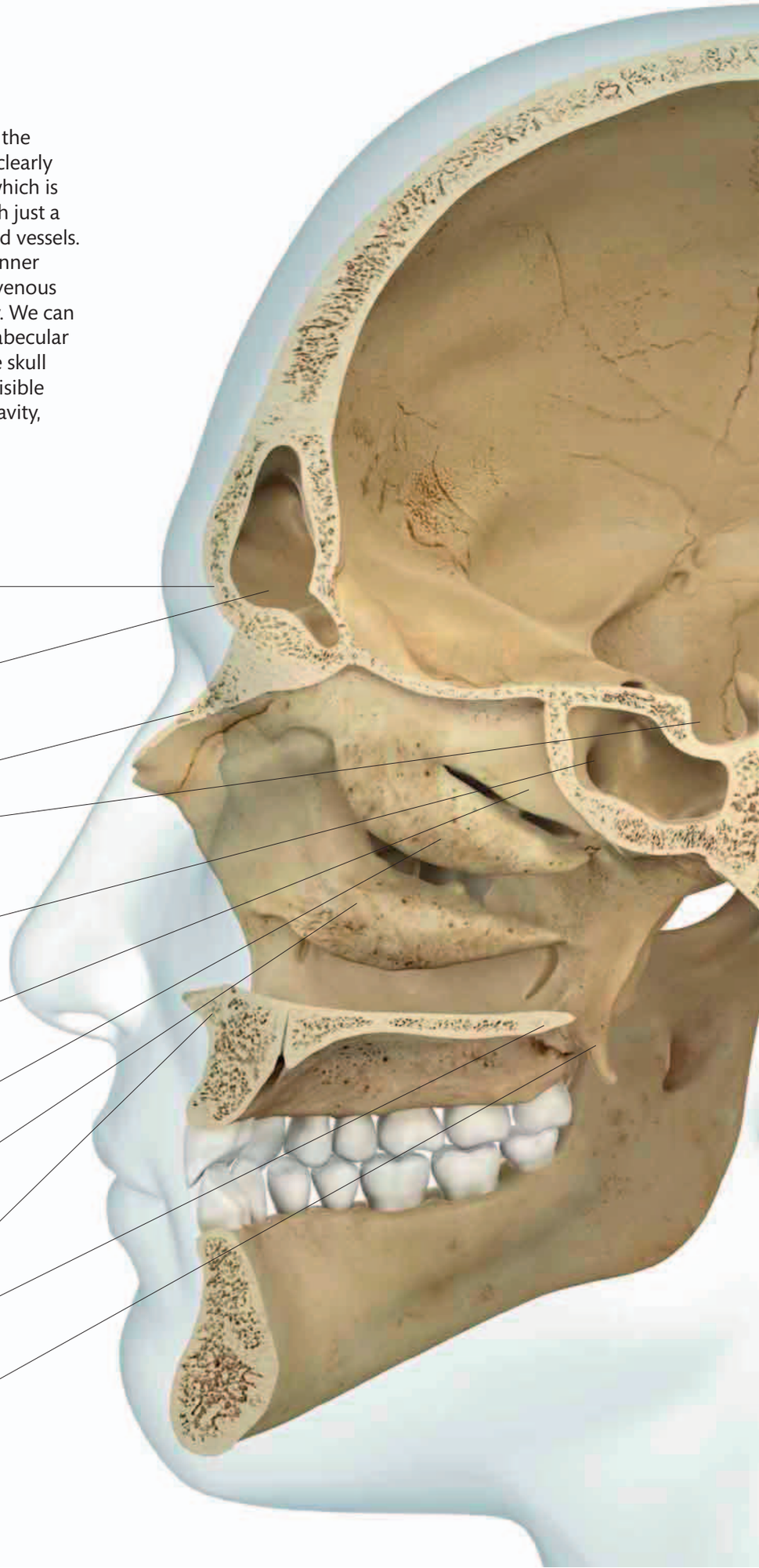

Frontal bone
Forms the anterior cranial fossa, where the frontal lobes of the brain lie, inside the skull

Frontal sinus
One of the paranasal air sinuses that drain into the nasal cavity, this is an air space within the frontal bone

Nasal bone

Pituitary fossa
Fossa is the Latin word for ditch; the pituitary gland occupies this small cavity on the upper surface of the sphenoid bone

Sphenoidal sinus
Another paranasal air sinus; it lies within the body of the sphenoid bone

Superior nasal concha
Part of the ethmoid bone, which forms the roof and upper side walls of the nasal cavity

Middle nasal concha
Like the superior nasal concha, this is also part of the ethmoid bone

Inferior nasal concha
A separate bone, attached to the inner surface of the maxilla; the conchae increase the surface area of the nasal cavity

Anterior nasal crest

Palatine bone
Joins to the maxilla and forms the back of the hard palate

Pterygoid process
Sticking down from the greater wing of the sphenoid bone, this process flanks the back of the nasal cavity and provides attachment for muscles of the palate and jaw

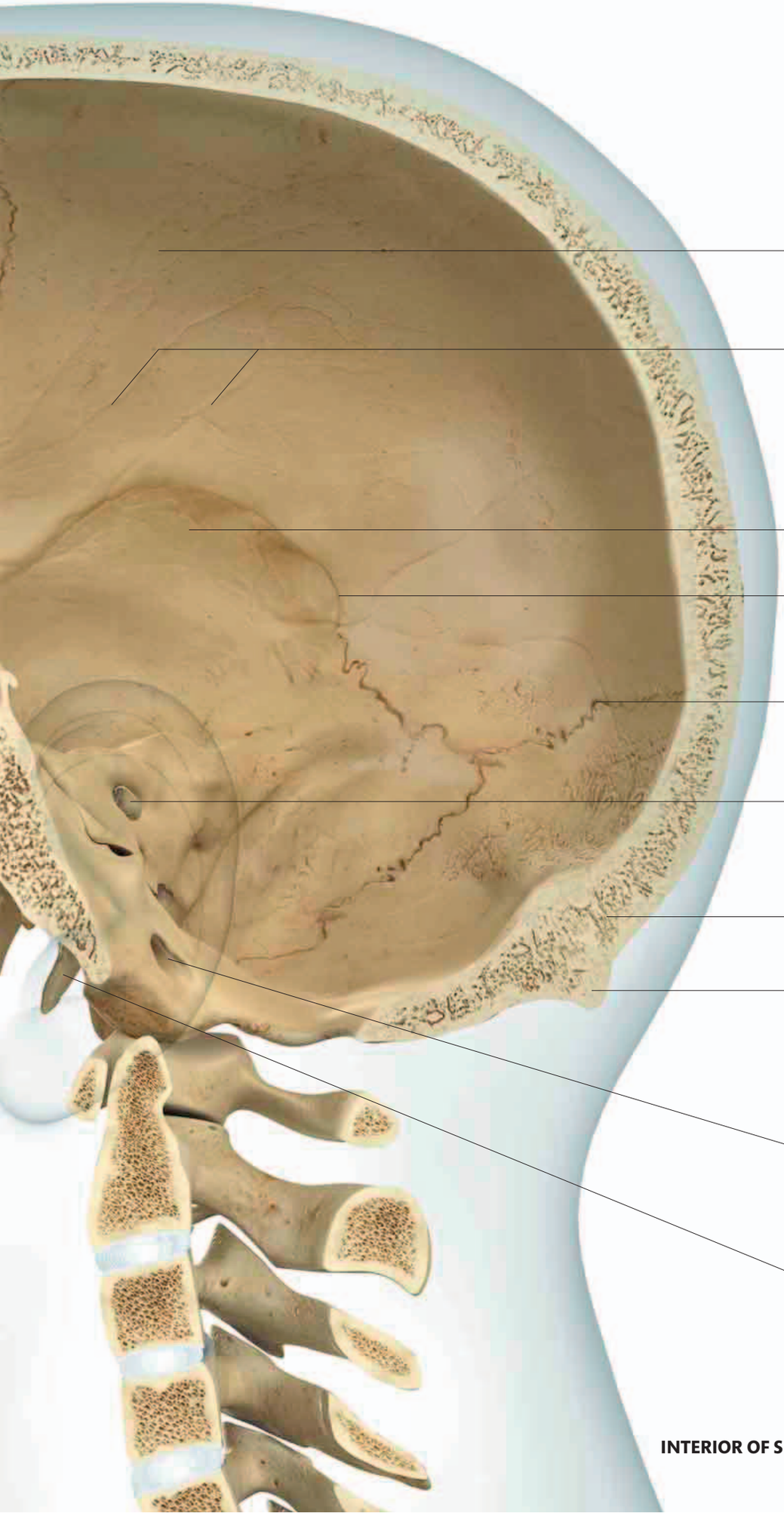

INTERIOR OF SKULL

HEAD AND NECK

In this view of the skull, we can clearly see that it is not one single bone, and we can also see how the various cranial bones fit together to produce the shape we are more familiar with. The butterfly-shaped sphenoid bone is right in the middle of the action—it forms part of the skull base, the orbits, and the side walls of the skull, and it articulates with many of the other bones of the skull. The temporal bones also form part of the skull's base and side walls. The extremely dense petrous parts of the temporal bones contain and protect the delicate workings of the ear, including the tiny ossicles (malleus, incus, and stapes) that transmit vibrations from the eardrum to the inner ear.

FIBROUS JOINTS

In places, the connective tissue between developing bones solidifies to create fibrous joints. Linked by microscopic fibers of collagen, these fixed joints anchor the edges of adjacent bones, or bone and tooth, so that they are locked together. Such joints include the sutures of the skull, the teeth sockets (gomphoses), and the lower joint between the tibia and fibula.

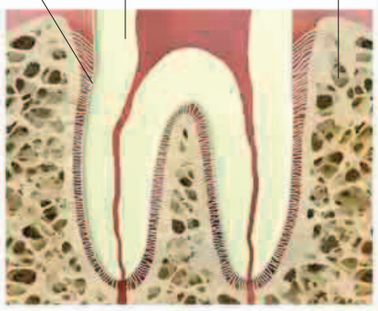

TOOTH

Gomphosis
This name comes from the Greek word for bolted together. The fibrous tissue of the periodontal ligament connects the cement of the tooth to the bone of the socket.

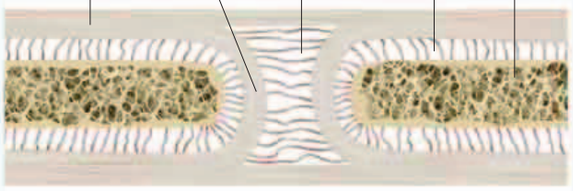

SKULL

Suture
These joints exist between flat bones of the skull. They are flexible in the skull of a newborn baby, and allow growth of the skull throughout childhood. The sutures in the adult skull are interlocking, practically immovable joints, and eventually fuse completely in later adulthood.

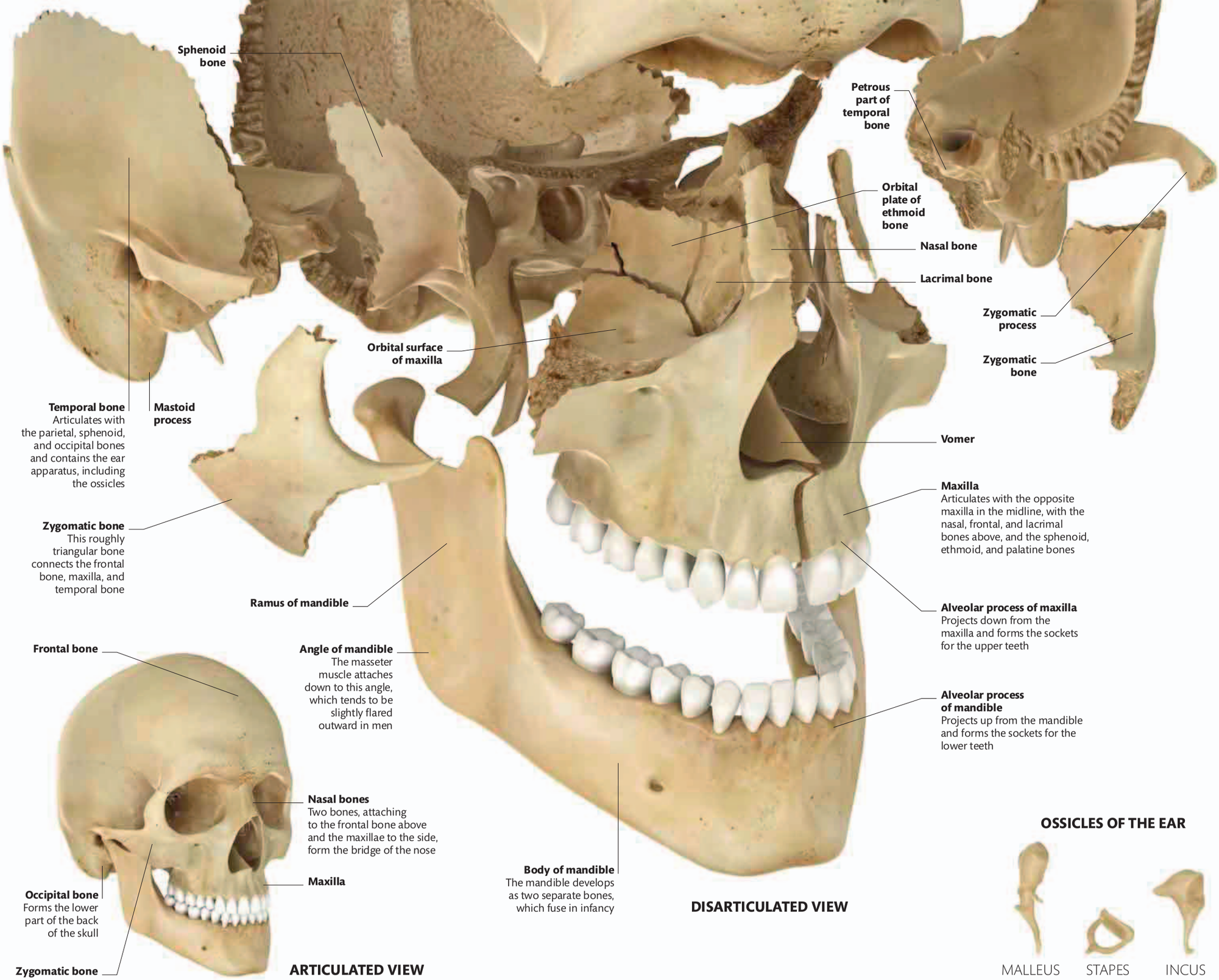
Sphenoid bone
Petrous part of temporal bone
Orbital plate of ethmoid bone
Nasal bone
Lacrimal bone
Zygomatic process
Zygomatic bone
Orbital surface of maxilla
Temporal bone
Articulates with the parietal, sphenoid, and occipital bones and contains the ear apparatus, including the ossicles
Mastoid process
Vomer
Maxilla
Articulates with the opposite maxilla in the midline, with the nasal, frontal, and lacrimal bones above, and the sphenoid, ethmoid, and palatine bones
Zygomatic bone
This roughly triangular bone connects the frontal bone, maxilla, and temporal bone
Ramus of mandible
Alveolar process of maxilla
Projects down from the maxilla and forms the sockets for the upper teeth
Frontal bone
Angle of mandible
The masseter muscle attaches down to this angle, which tends to be slightly flared outward in men
Alveolar process of mandible
Projects up from the mandible and forms the sockets for the lower teeth
Nasal bones
Two bones, attaching to the frontal bone above and the maxillae to the side, form the bridge of the nose
Maxilla
Occipital bone
Forms the lower part of the back of the skull
Zygomatic bone
Body of mandible
The mandible develops as two separate bones, which fuse in infancy
ARTICULATED VIEW
DISARTICULATED VIEW
OSSICLES OF THE EAR
MALLEUS
STAPES
INCUS

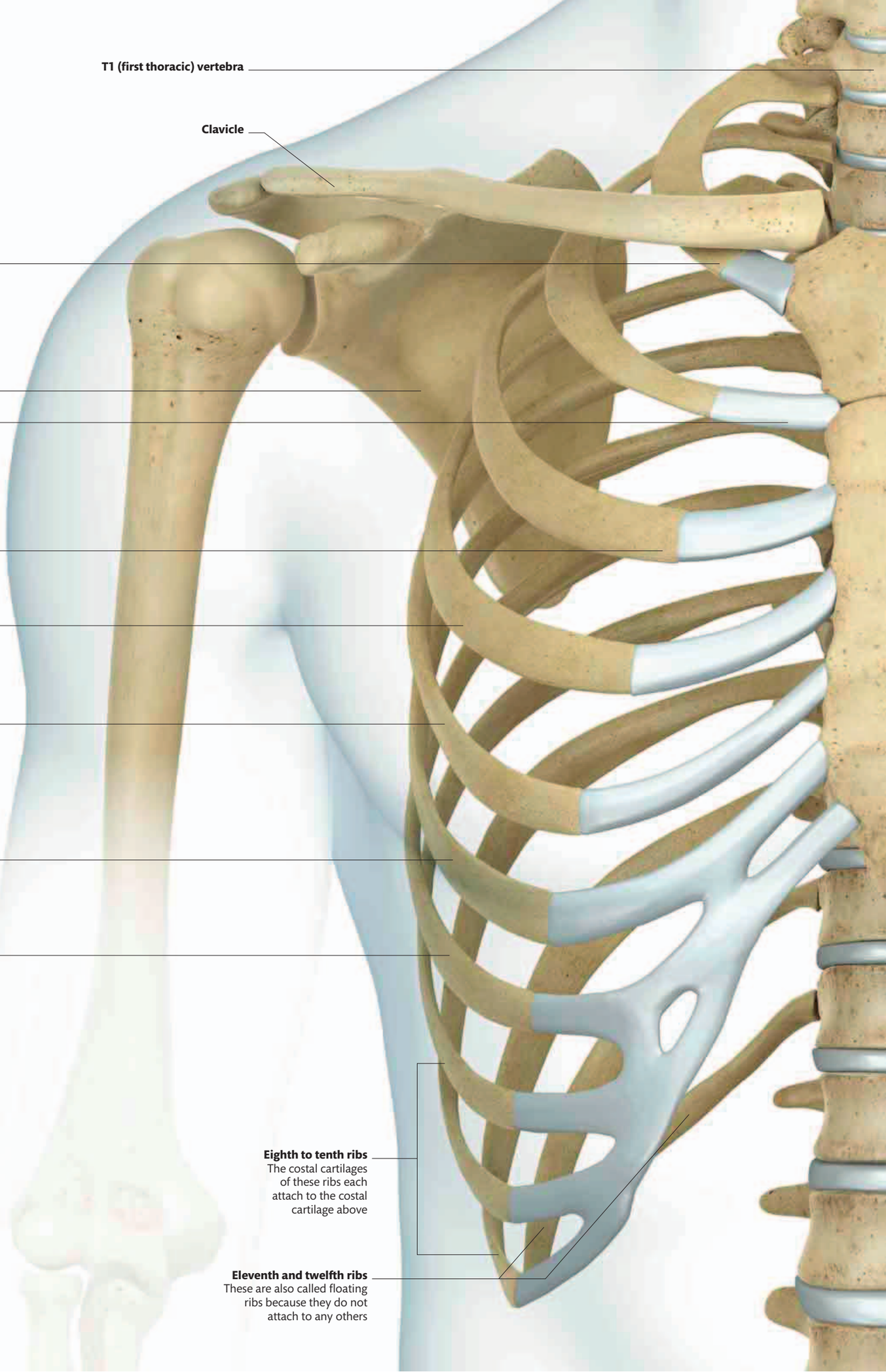
T1 (first thoracic) vertebra
Clavicle
First rib
Smaller and more curved than all the other ribs; the thoracic inlet is formed by the first rib on each side, together with the manubrium sterni and the body of the T1 vertebra
Scapula
Second costal cartilage
The upper seven ribs are true ribs, and all attach directly to the sternum via costal cartilages
Third rib
Fourth rib
Fifth rib
Sixth rib
Seventh rib
Eighth to tenth ribs
The costal cartilages of these ribs each attach to the costal cartilage above
Eleventh and twelfth ribs
These are also called floating ribs because they do not attach to any others

Transverse process of T1
Each rib articulates with the transverse processes of the corresponding thoracic vertebra

Head of first rib
The heads of the ribs articulate with the bodies of vertebrae

Manubrium sterni
The sternum is shaped like a dagger or short sword; manubrium means handle or hilt in Latin

Manubriosternal joint

Body of sternum
Sternum comes from the Greek for breastbone

Xiphisternal joint

Xiphoid process
The tip of the sternum takes its name from the Greek word for sword

THORAX

The skeleton of the thorax plays several extremely important roles. It not only acts as an anchor for muscle attachment, but during breathing the ribs move up and out to increase the volume inside the thoracic cavity and draw air into the lungs. It also forms a protective cage around the precious organs inside: the heart and lungs. The bony thorax includes the 12 thoracic vertebrae, 12 pairs of ribs and costal cartilages, and the breastbone, or sternum. The upper seven ribs all articulate with the sternum via their costal cartilages. The eighth to the tenth costal cartilages each join to the cartilage above, creating the sweeping curve of the ribcage below the sternum on each side. The eleventh and twelfth ribs are short and do not join any other ribs—they are sometimes referred to as free or floating ribs.

ANTERIOR (FRONT)

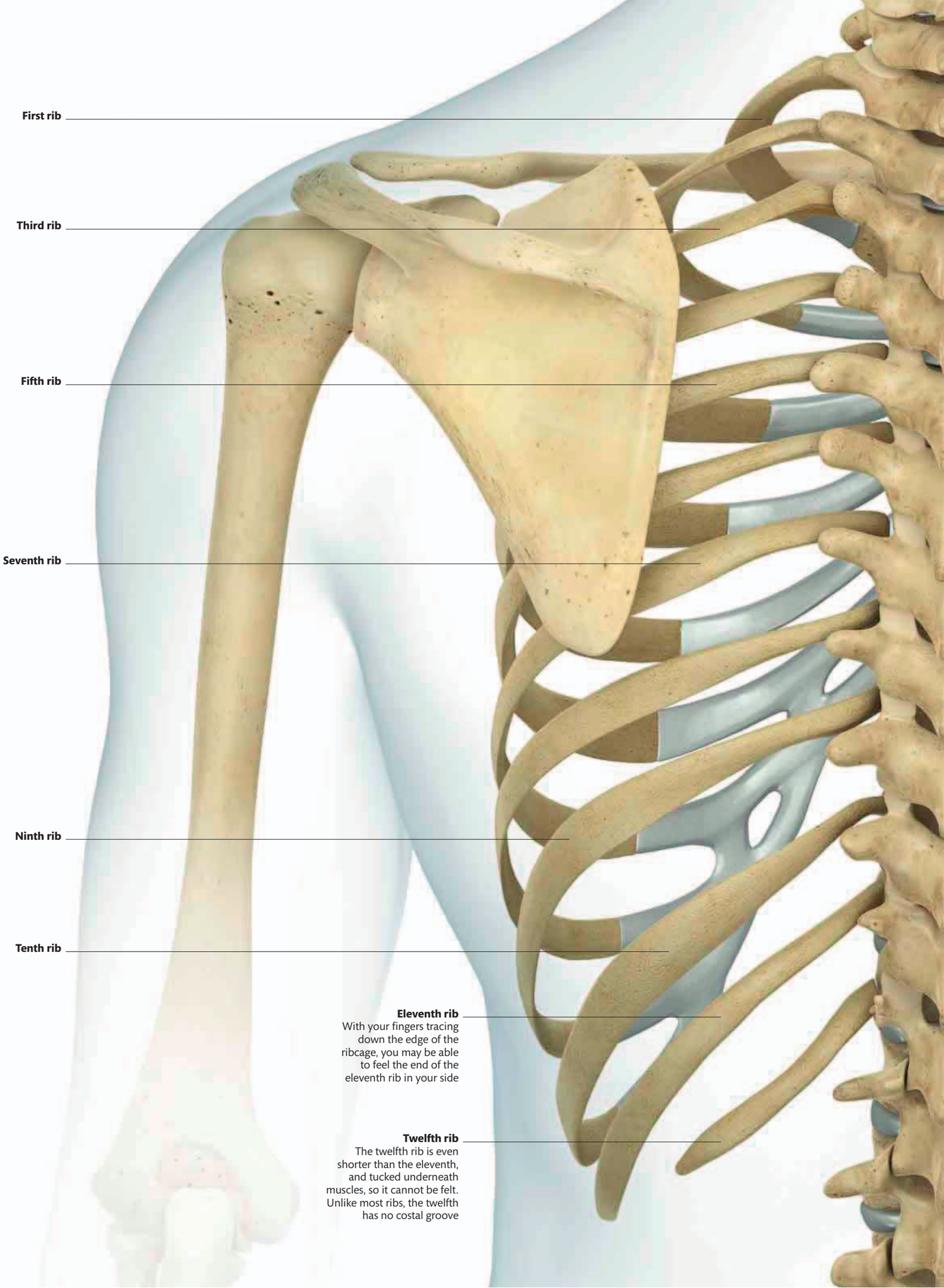
First rib
Third rib
Fifth rib
Seventh rib
Ninth rib
Tenth rib
Eleventh rib
With your fingers tracing down the edge of the ribcage, you may be able to feel the end of the eleventh rib in your side
Twelfth rib
The twelfth rib is even shorter than the eleventh, and tucked underneath muscles, so it cannot be felt. Unlike most ribs, the twelfth has no costal groove

C7 (seventh cervical vertebra)

Transverse process of T1

Costal groove

POSTERIOR (BACK)

THORAX

There are cartilaginous joints between the vertebrae at the back of the thorax, and between the parts of the sternum at the front. The joints between the ribs and the vertebrae at the back are synovial, allowing the ribs to move during breathing. When taking a breath, the anterior (front) ends of the upper ribs, along with the sternum, lift up and forward to increase the chest's front-to-back diameter, while the lower ribs move up and out, increasing the side-to-side diameter. Most ribs have a costal groove marking the lower border, on the inner surface, where nerves and vessels of the thoracic wall lie.

SPINE

The spine, or vertebral column, occupies a central position in the skeleton, and plays several extremely important roles: it supports the trunk, encloses and protects the spinal cord, provides sites for muscle attachment, and contains blood-forming bone marrow. The entire vertebral column is about 28in (70cm) long in men, and 24in (60cm) long in women. About a quarter of this length is made up by the cartilaginous intervertebral discs between the vertebrae. The number of vertebrae varies from 32 to 35, mostly due to variation in the number of small vertebrae that make up the coccyx. Although there is a general pattern for a vertebra —most possess a body, a neural arch, and spinous and transverse processes—there are recognizable features that mark out the vertebrae of each section of the spine.

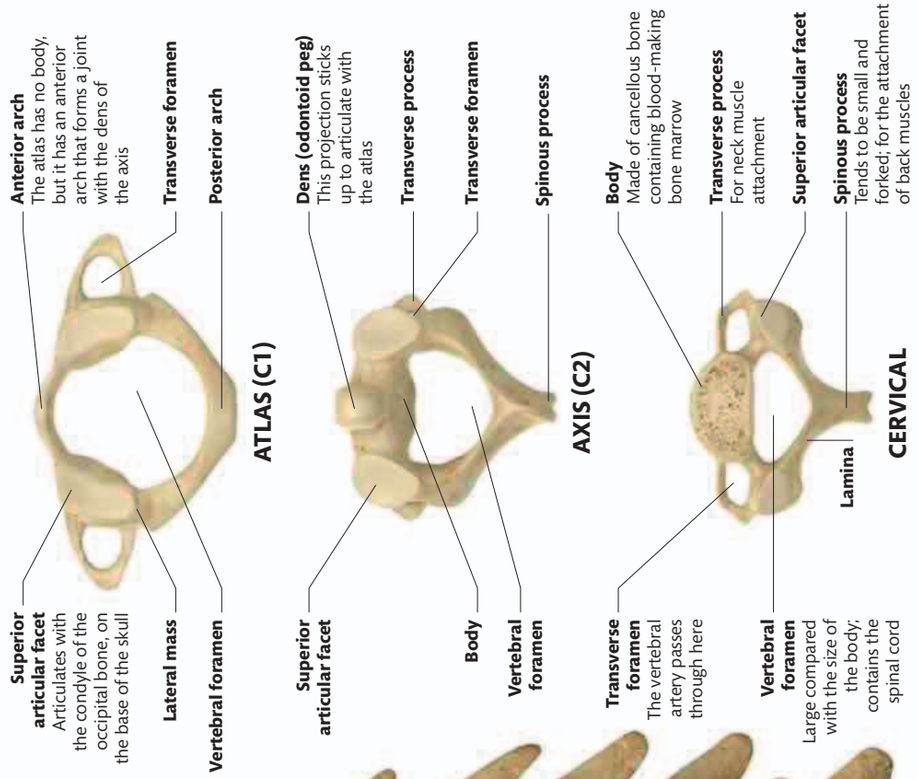

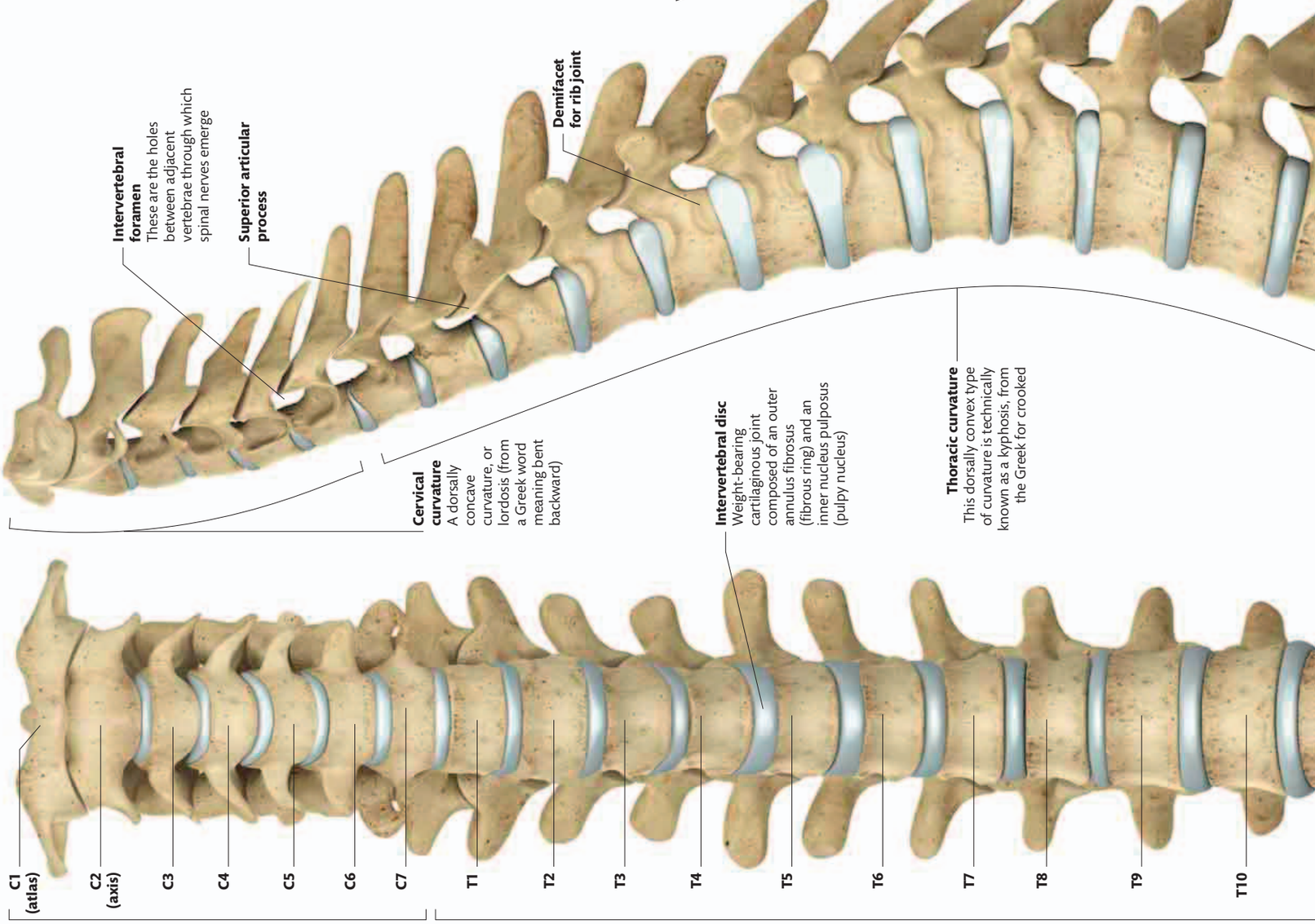

Cervical spine
(Seven vertebrae make up the spine in the neck)

Thoracic spine
(Twelve vertebrae, providing attachment for twelve pairs of ribs)

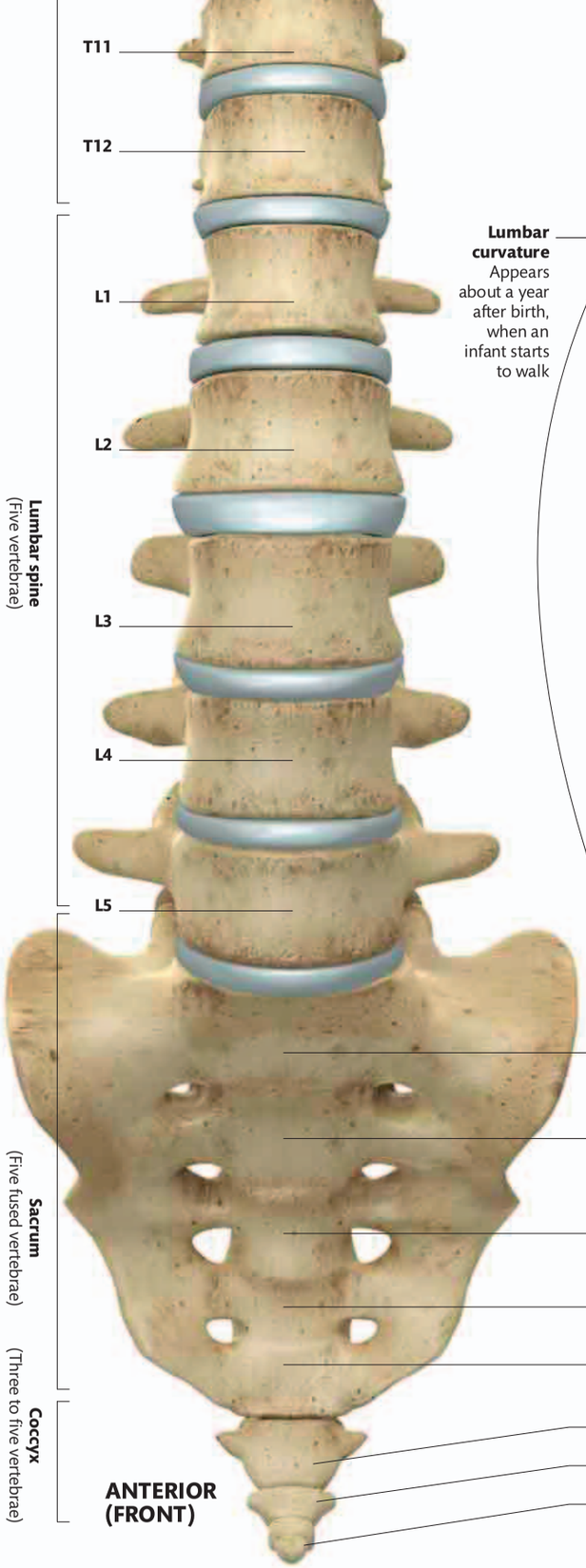
T11
T12
L1
L2
L3
L4
L5
S1
S2
S3
S4
S5
co1
co2
co3
Lumbar spine
(Five vertebrae)
Sacrum
(Five fused vertebrae)
Coccyx
(Three to five vertebrae)
ANTERIOR
(FRONT)

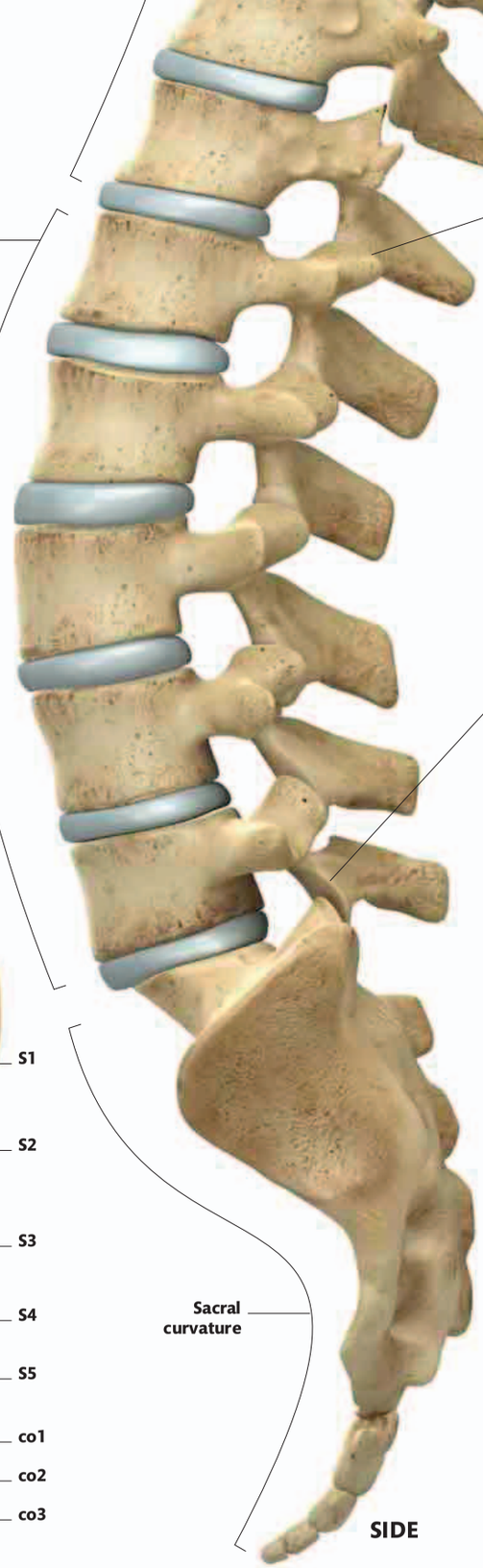
Lumbar curvature
Appears about a year after birth, when an infant starts to walk
Zygapophyseal (facet) joint
Synovial joints between the adjacent articular processes allow variable degrees of movement in different sections of the spine; in disk degeneration, these joints end up bearing more weight and may be a source of back pain
Inferior articular process
Sacral curvature
SIDE

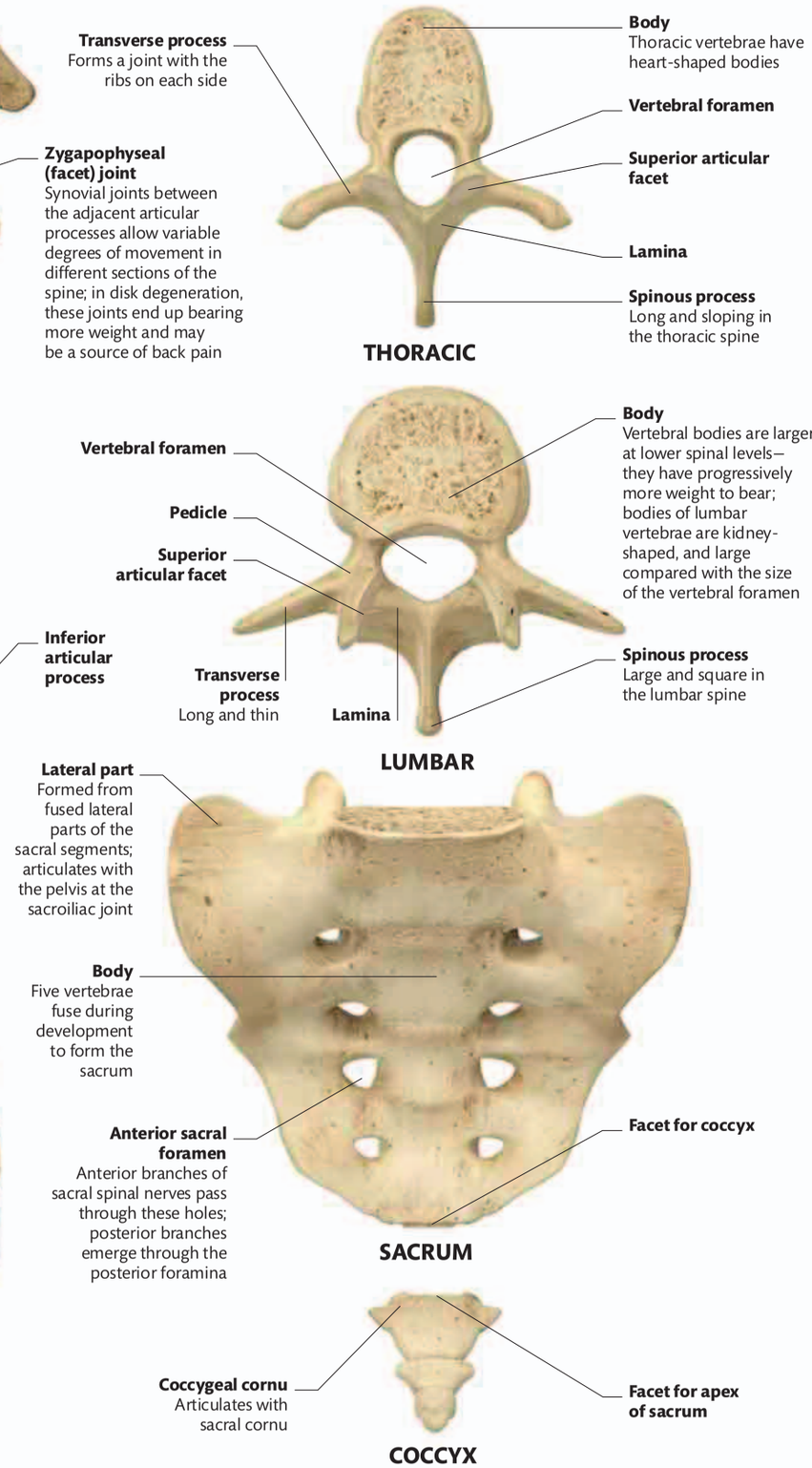
Transverse process
Forms a joint with the ribs on each side
Body
Thoracic vertebrae have heart-shaped bodies
Vertebral foramen
Superior articular facet
Lamina
Spinous process
Long and sloping in the thoracic spine
THORACIC
Vertebral foramen
Pedicle
Superior articular facet
Transverse process
Long and thin
Lamina
Body
Vertebral bodies are larger at lower spinal levels—they have progressively more weight to bear; bodies of lumbar vertebrae are kidney-shaped, and large compared with the size of the vertebral foramen
Spinous process
Large and square in the lumbar spine
LUMBAR
Lateral part
Formed from fused lateral parts of the sacral segments; articulates with the pelvis at the sacroiliac joint
Body
Five vertebrae fuse during development to form the sacrum
Anterior sacral foramen
Anterior branches of sacral spinal nerves pass through these holes; posterior branches emerge through the posterior foramina
Facet for coccyx
SACRUM
Coccygeal cornu
Articulates with sacral cornu
Facet for apex of sacrum
COCCYX

ABDOMEN AND PELVIS

The bony boundaries of the abdomen include the five lumbar vertebrae at the back, the lower margin of the ribs above, and the pubic bones and iliac crest of the pelvic bones below. The abdominal cavity itself extends up under the ribcage, as high as the gap between the fifth and sixth ribs, due to the domed shape of the diaphragm. This means that some abdominal organs, such as the liver, stomach, and spleen are, in fact, largely tucked up under the ribs. The pelvis is a basin shape, and is enclosed by the two pelvic (or innominate) bones, at the front and sides, and by the sacrum at the back. Each pelvic bone is made of three fused bones: the ilium at the rear, the ischium at the lower front, and the pubis above it.

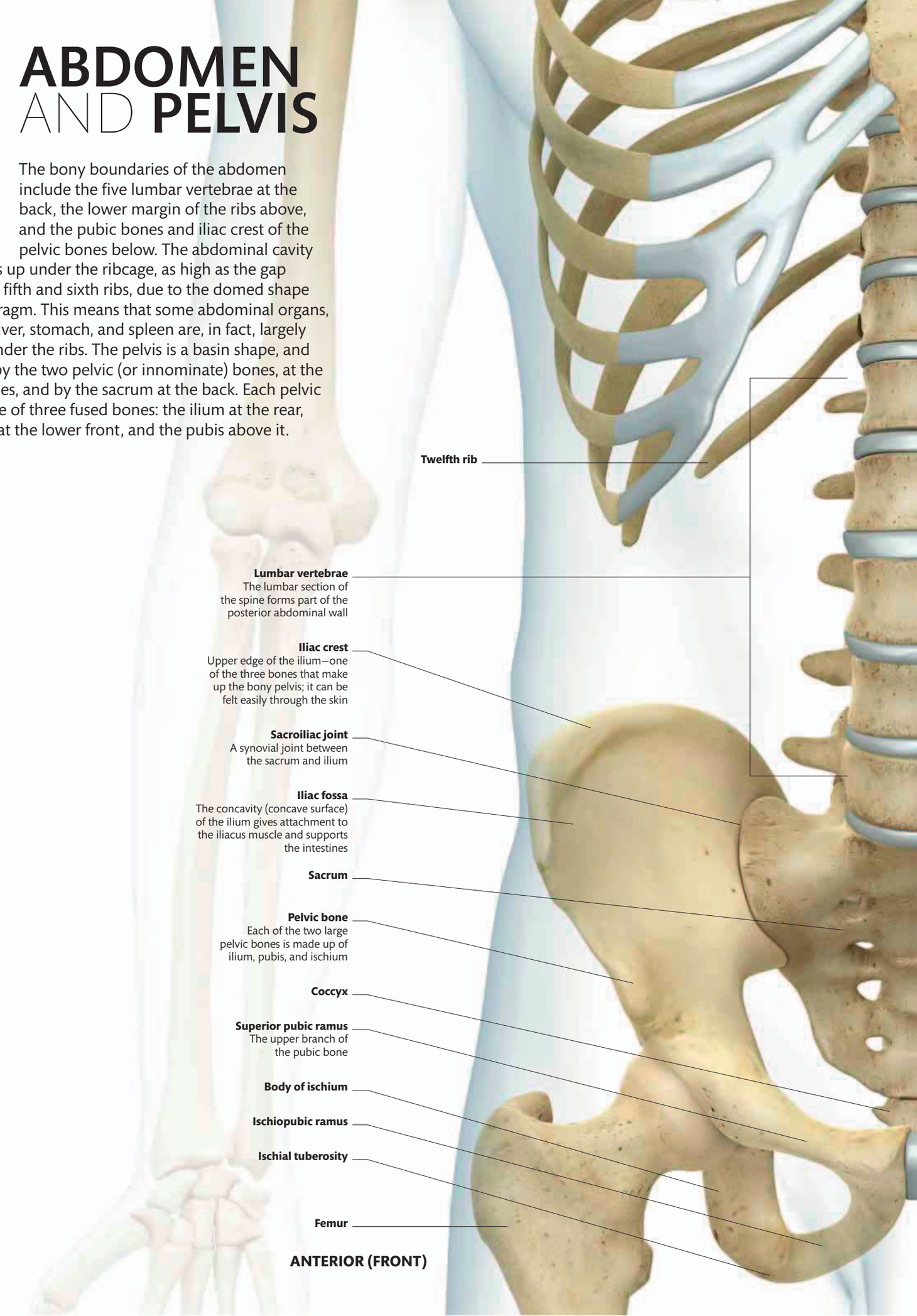

CARTILAGINOUS JOINTS

Semi-movable cartilaginous joints are formed by bones separated by a disc of resilient and compressible fibrocartilage, which allows limited movement. Cartilaginous joints include the junctions between ribs and costal cartilages, joints between the components of the sternum, and the pubic symphysis. The intervertebral discs are also specialized cartilaginous joints.

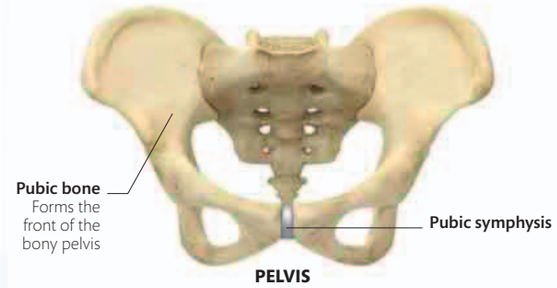

Pubic symphysis
At the front of the bony pelvis, the two pubic bones meet each other. The articular surface of each is covered with hyaline cartilage, with a pad of fibrocartilage joining them in the middle.

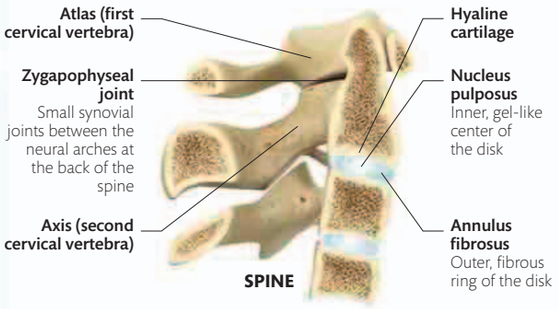

Intervertebral disc
Each fibrocartilage pad or disk between vertebrae is organized into an outer annulus fibrosus and an inner nucleus pulposus.

Anterior superior iliac spine
This is the anterior (front) end of the iliac crest

Ala of sacrum
The bony masses to the sides of the sacrum are called the alae, which means wings in Latin

Anterior sacral foramina
Anterior (frontal) branches of the sacral spinal nerves pass out through these holes

Pubic symphysis
A cartilaginous joint between the two pubic bones

Pubic tubercle
This small bony projection provides an attachment point for the inguinal ligament

Obturator foramen
This hole is largely closed over by a membrane, with muscles attaching on either side; its name comes from the Latin for stopped up

ABDOMEN AND PELVIS

The orientation of the facet joints (the joints between the vertebrae) of the lumbar spine means that rotation of the vertebrae is limited, but flexion and extension can occur freely. There is, however, rotation at the lumbosacral joint, which allows the pelvis to swing during walking. The sacroiliac joints are unusual in that they are synovial joints (which are usually very movable), yet they are particularly limited in their movement. This is because strong sacroiliac ligaments around the joints bind the ilium (part of the pelvic bone) tightly to the sacrum on each side. Lower down, the sacrospinous and sacrotuberous ligaments, stretching from the sacrum and coccyx to the ilium, provide additional support and stability.

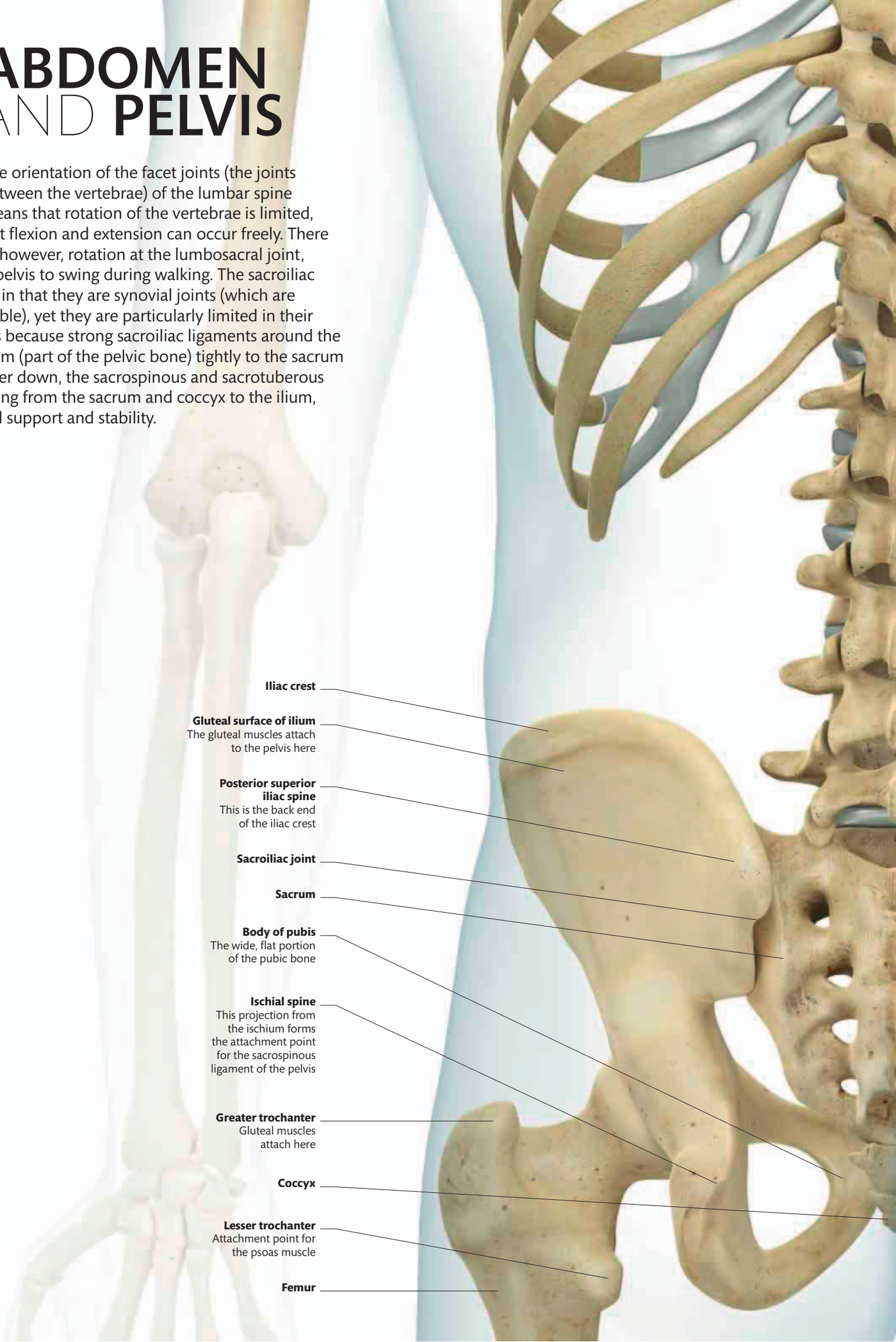

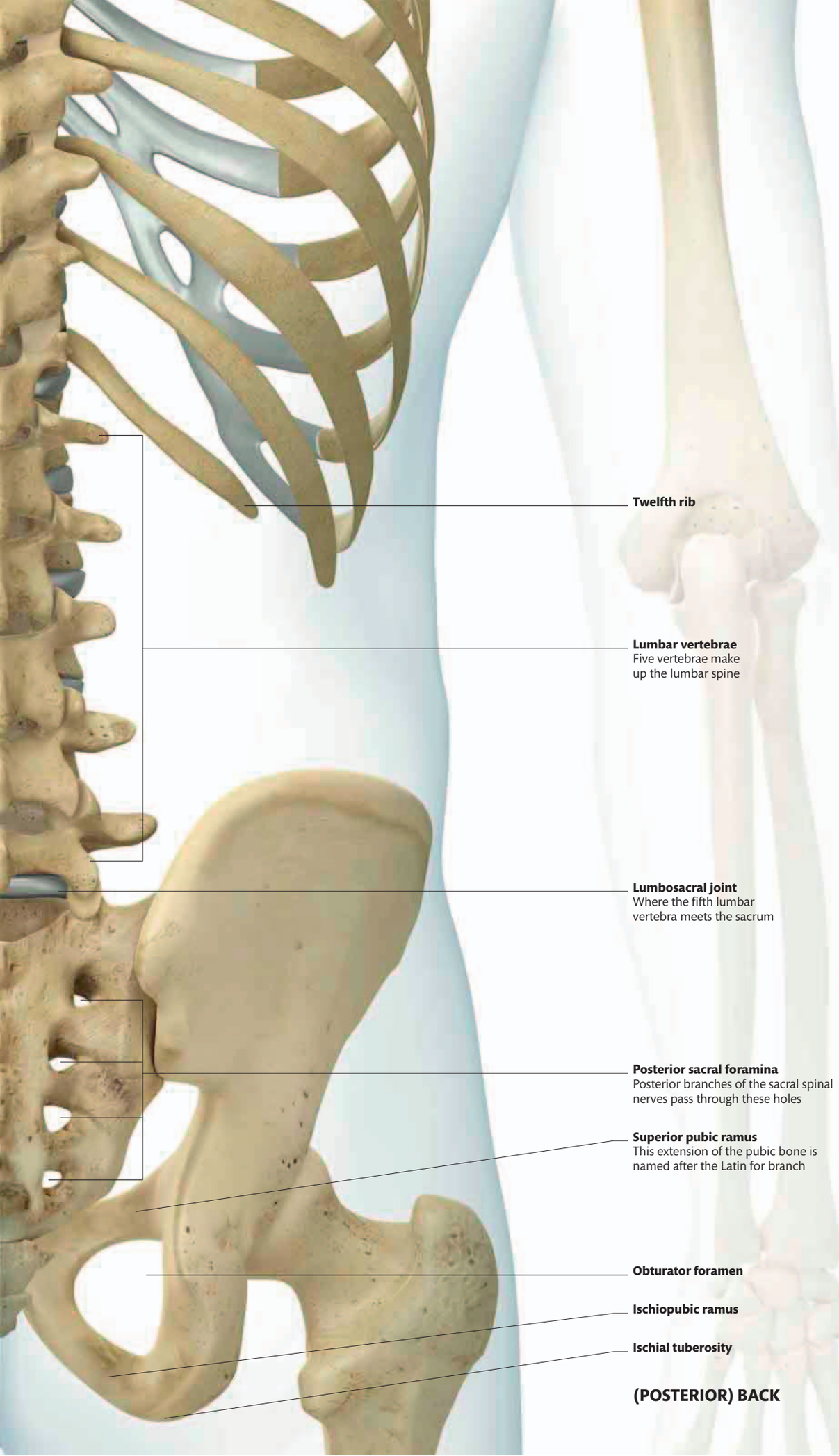

(POSTERIOR) BACK

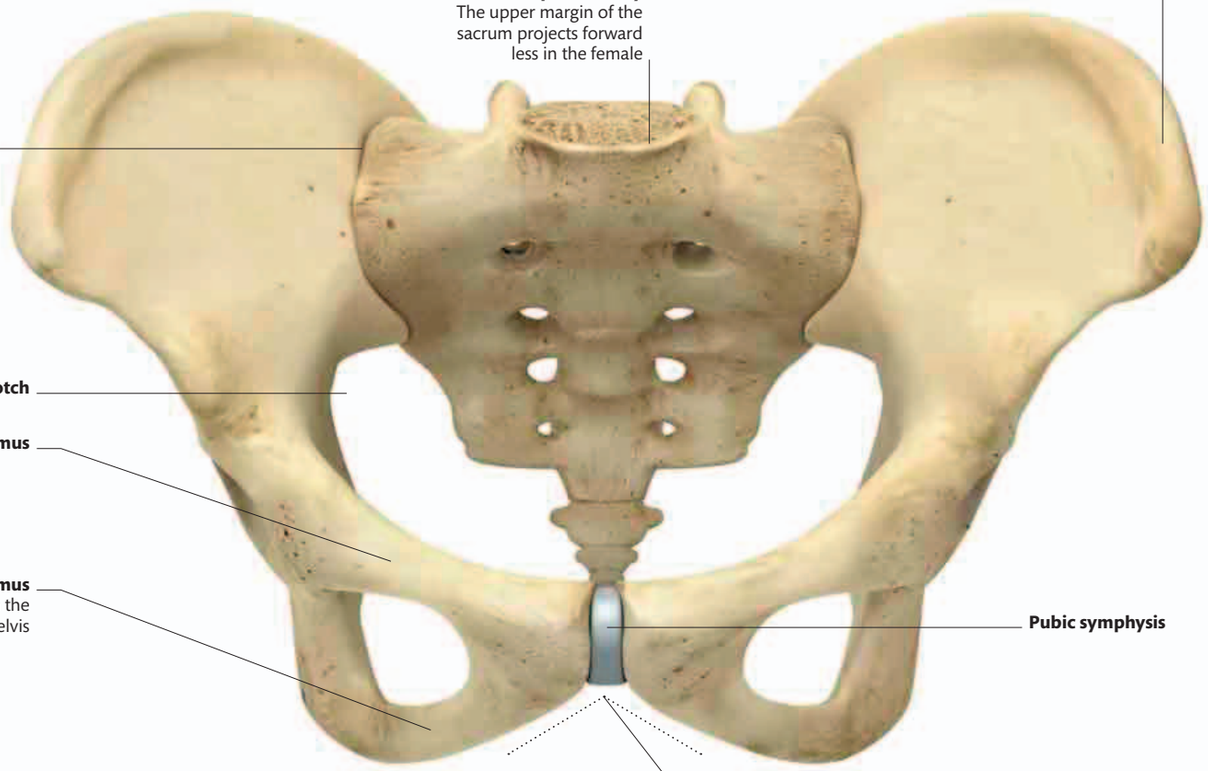

FEMALE PELVIS ANTERIOR (FRONT)

PELVIS

The bony pelvis is the part of the skeleton that is most different between the sexes, because the pelvis in the female has to accommodate the birth canal, unlike the male pelvis. Comparing the pelvic bones of a man and a woman, there are obvious differences between the two. The shape of the ring formed by the sacrum and the two pelvic bones —the pelvic brim—tends to be a wide oval in the woman and much narrower and heart-shaped in a man. The subpubic angle, underneath the joint between the two pubic bones, is much narrower in a man than it is in with a woman. As with the rest of the skeleton, the pelvic bone also tends to be more chunky or robust in a man, with more obvious ridges where muscles attach.

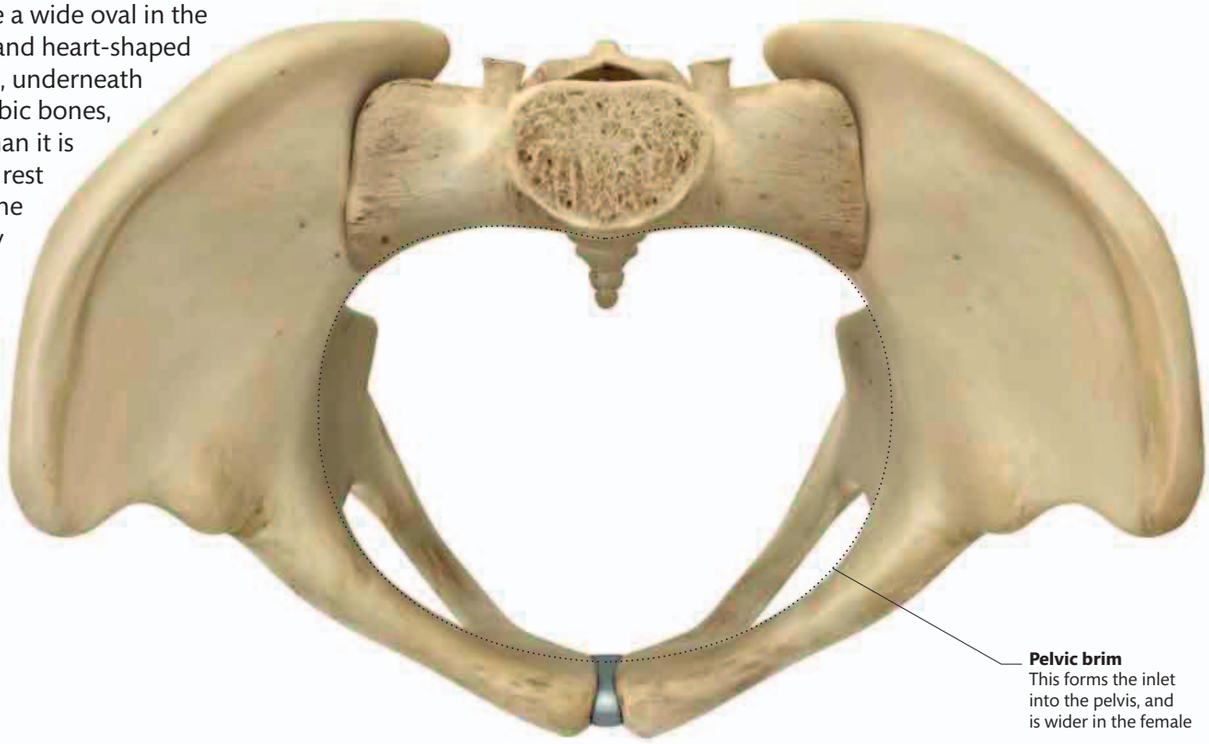

FEMALE PELVIS VIEWED FROM ABOVE

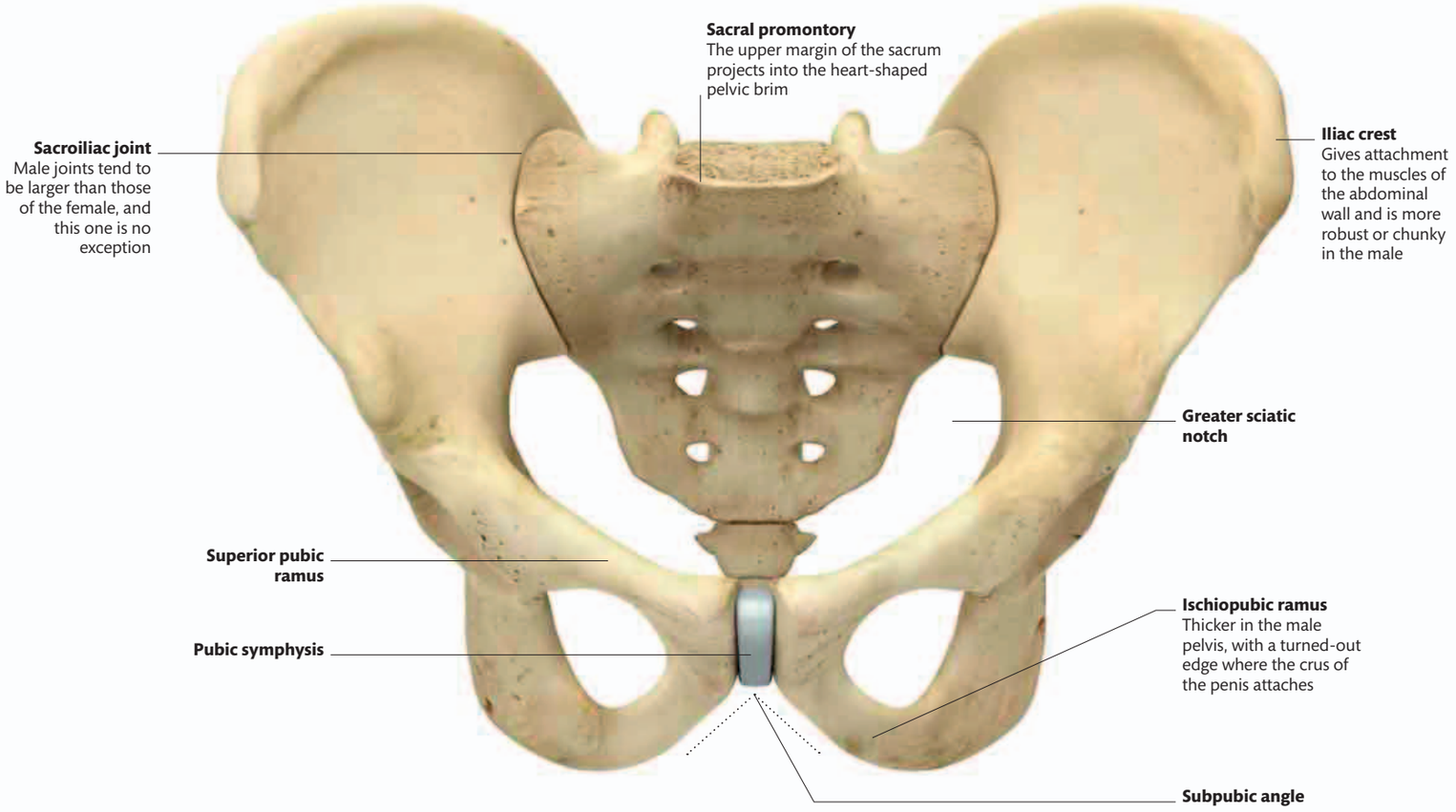

MALE PELVIS ANTERIOR (FRONT)

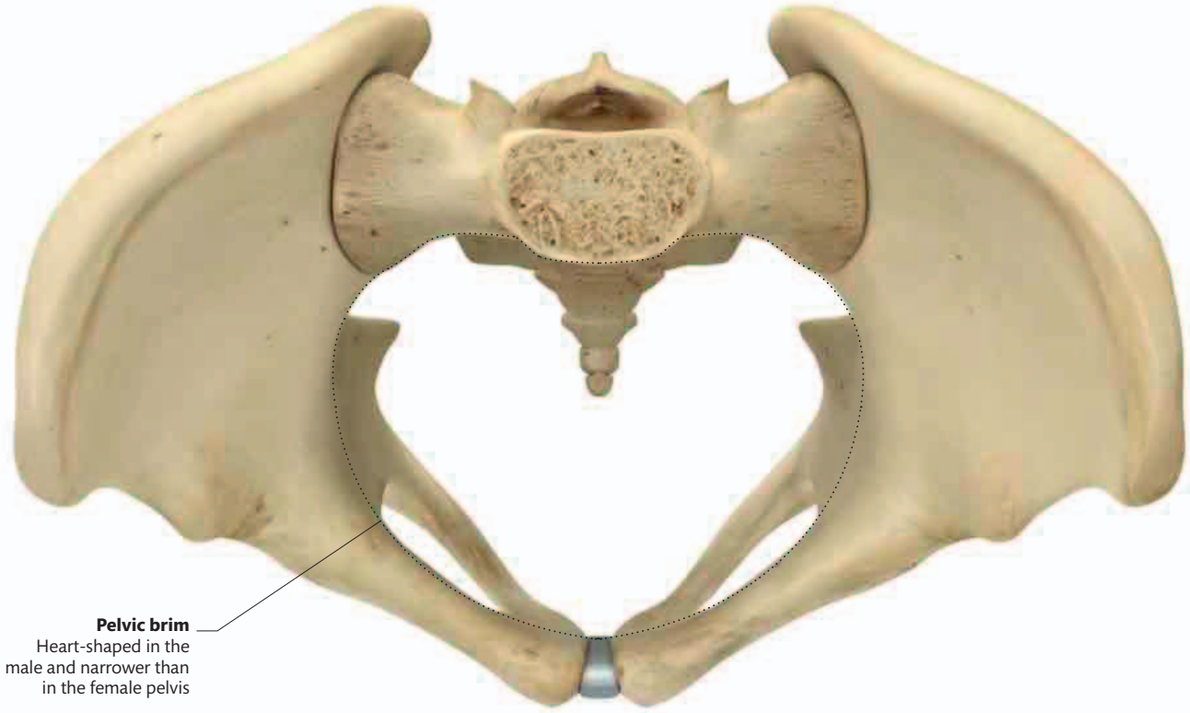

MALE PELVIS VIEWED FROM ABOVE

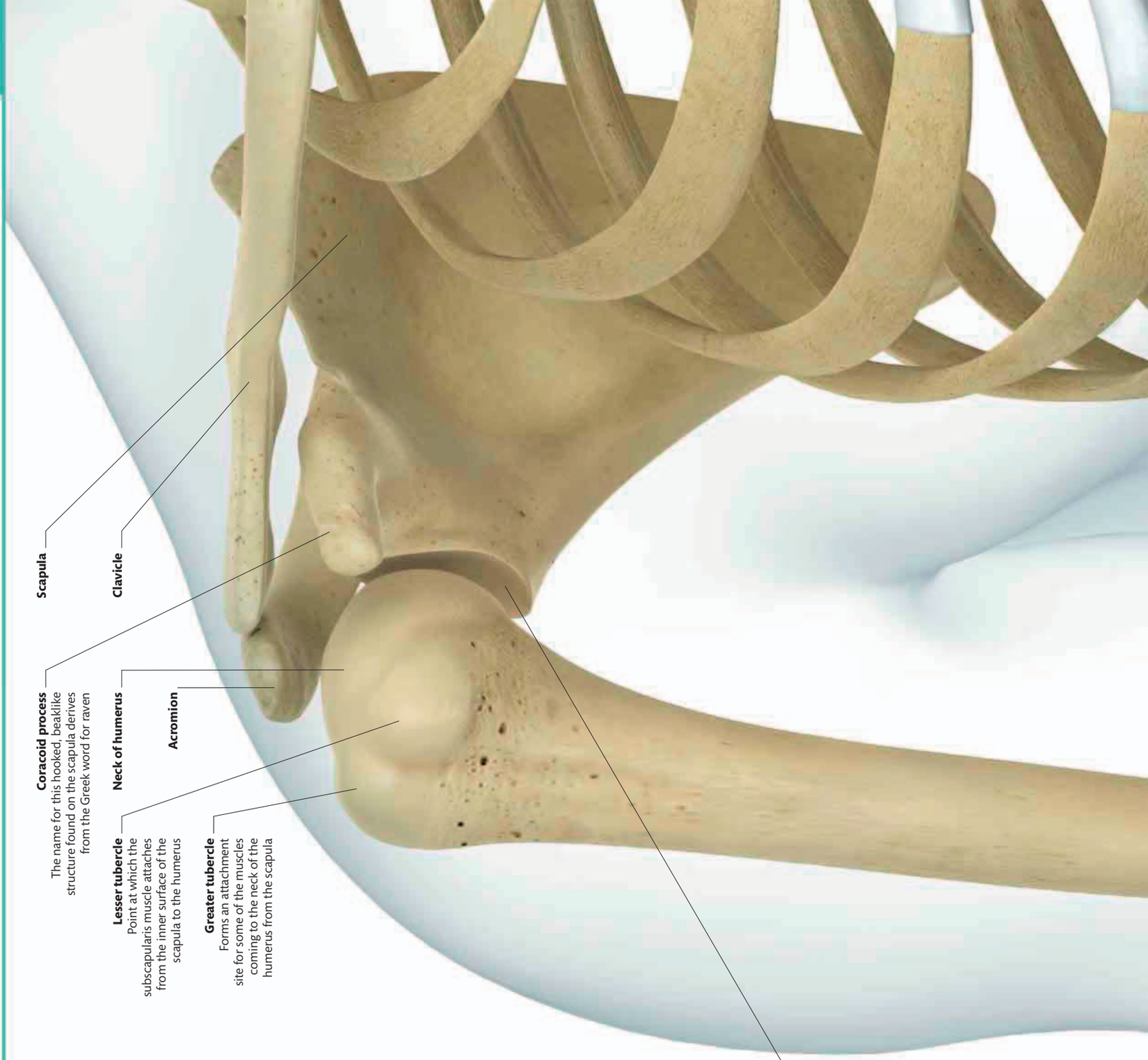

SHOULDER AND UPPER ARM

The scapula and clavicle make up the shoulder girdle, which anchors the arm to the thorax. This is a very mobile attachment—the scapula "floats" on the ribcage, attached to it by muscles only (rather than by a true joint) that pull the scapula around on the underlying ribs, altering the position of the shoulder joint. The clavicle has joints—it articulates with the acromion of the scapula laterally (at the side) and the sternum at the other end—and helps hold the shoulder out to the side while allowing the scapula to move around. The shoulder joint, the most mobile joint in the body, is a ball-and-socket joint, but the socket is small and shallow, allowing the ball-shaped head of the humerus to move freely.

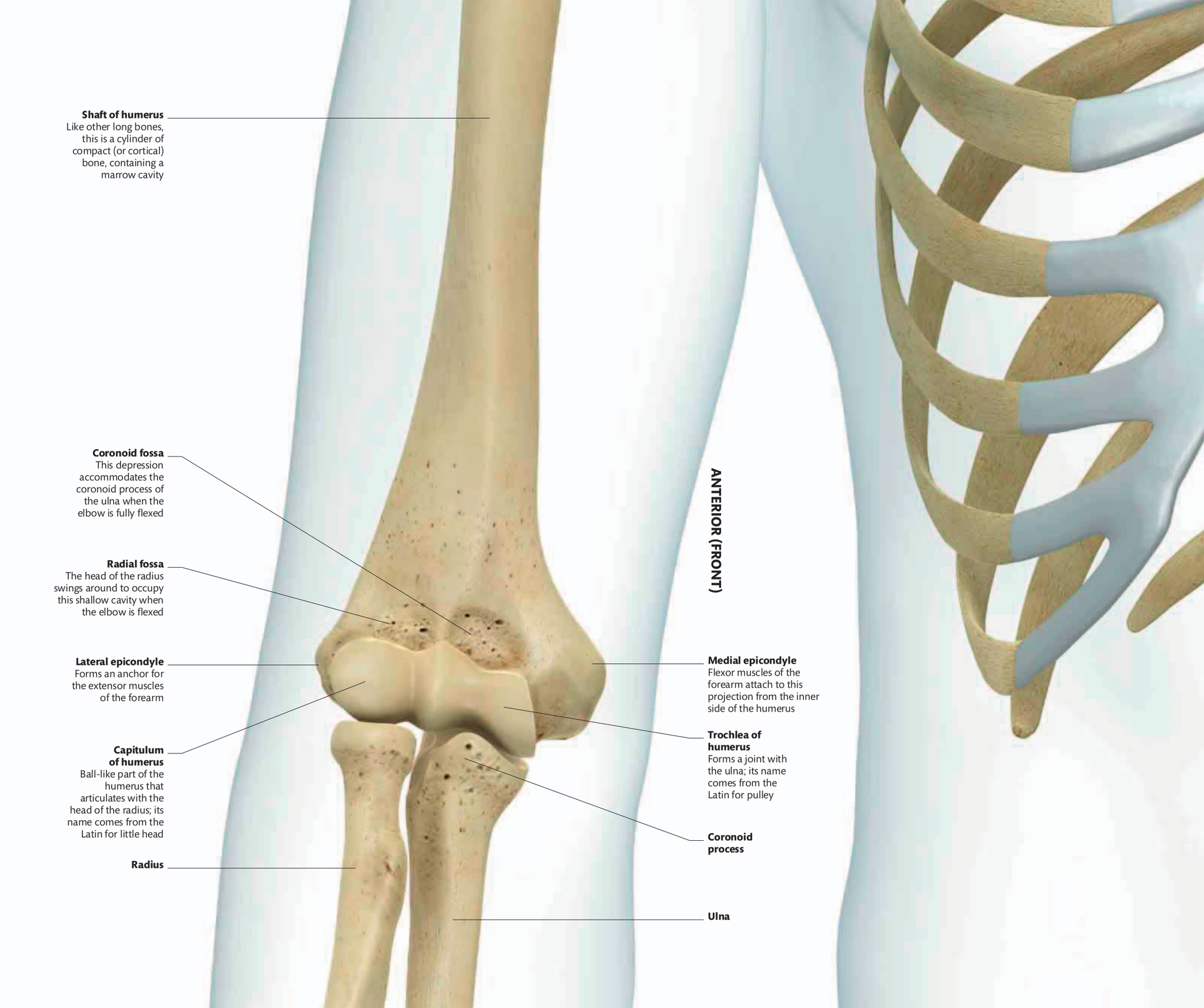
Shaft of humerus
Like other long bones, this is a cylinder of compact (or cortical) bone, containing a marrow cavity
Coronoid fossa
This depression accommodates the coronoid process of the ulna when the elbow is fully flexed
Radial fossa
The head of the radius swings around to occupy this shallow cavity when the elbow is flexed
Lateral epicondyle
Forms an anchor for the extensor muscles of the forearm
Capitulum of humerus
Ball-like part of the humerus that articulates with the head of the radius; its name comes from the Latin for little head
Radius
ANTERIOR (FRONT)
Medial epicondyle
Flexor muscles of the forearm attach to this projection from the inner side of the humerus
Trochlea of humerus
Forms a joint with the ulna; its name comes from the Latin for pulley
Coronoid process
Ulna

SHOULDER AND UPPER ARM

The back of the scapula is divided into two sections by its spine. The muscles that attach above this spine are called supraspinatus; those that attach below are called infraspinatus. They are part of the rotator cuff muscle group, which enables shoulder movements and stabilizes the shoulder joint. The spine of the scapula runs to the side and projects out above the shoulder joint to form the acromion, which can be easily felt on the top of the shoulder. The scapula rests in the position shown here when the arm is hanging at the side of the body. If the arm is abducted (raised to the side), the entire scapula rotates so that the glenoid cavity points upward and the inferior angle moves outward.

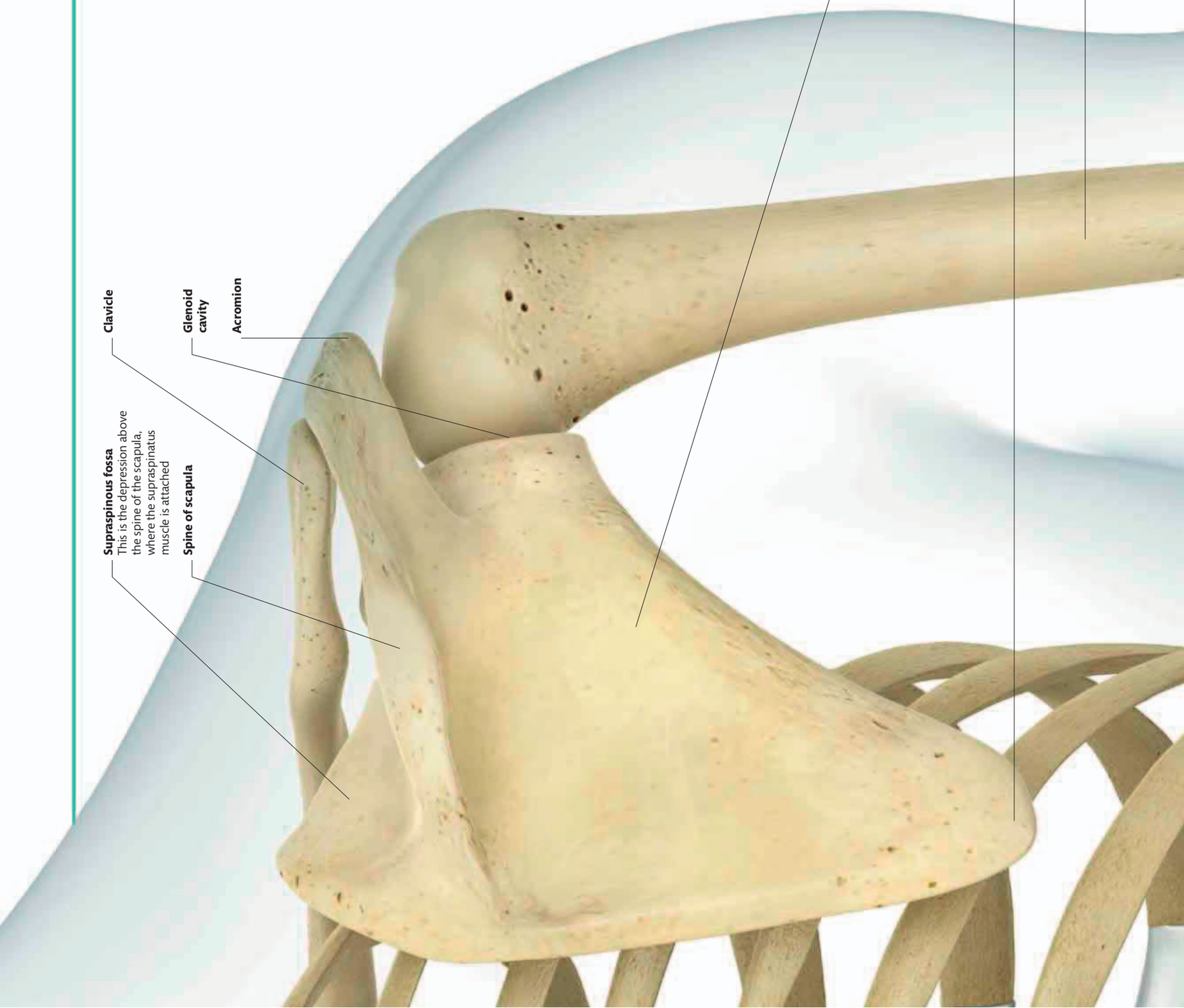

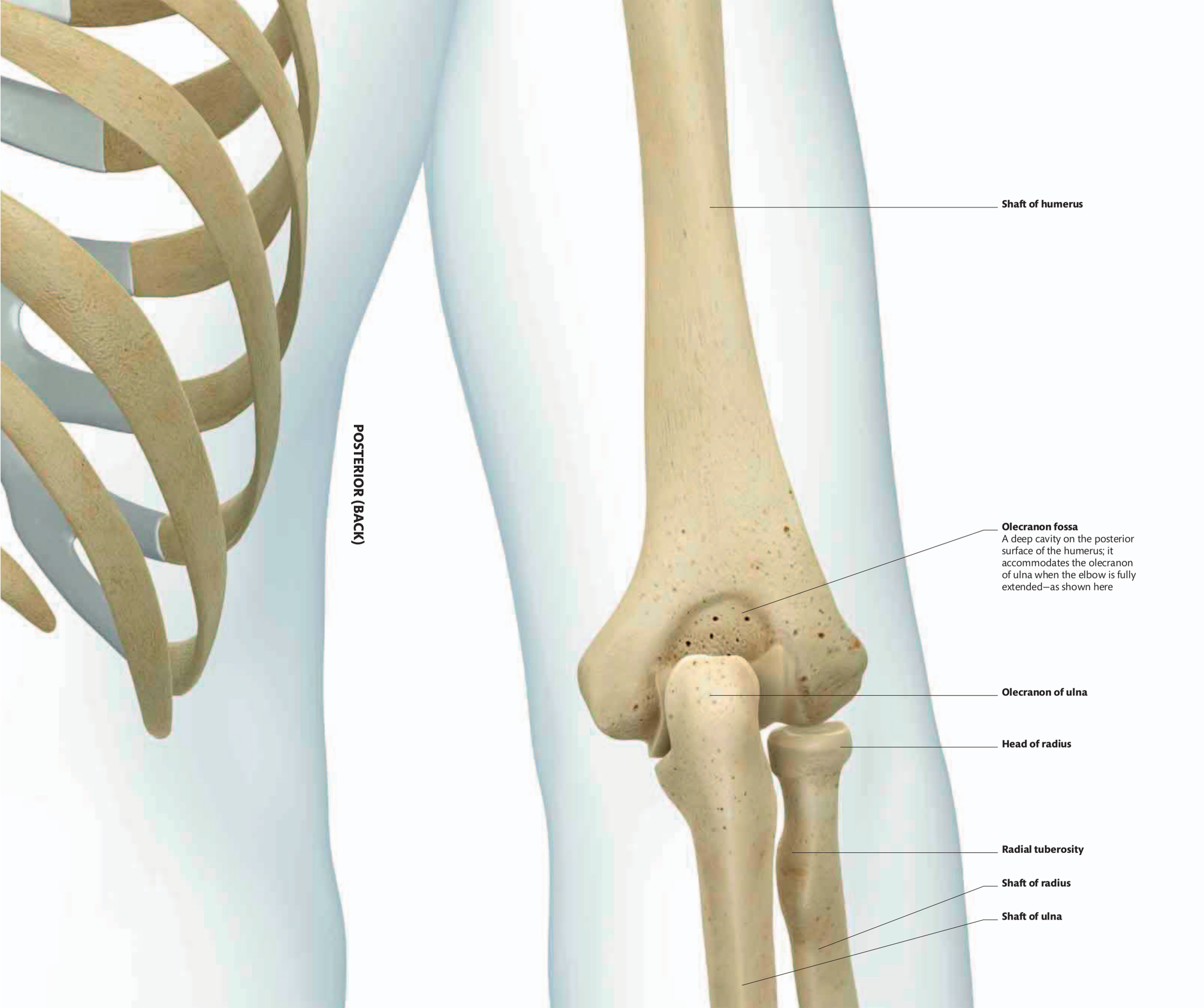
POSTERIOR (BACK)
Shaft of humerus
Olecranon fossa
A deep cavity on the posterior surface of the humerus; it accommodates the olecranon of ulna when the elbow is fully extended—as shown here
Olecranon of ulna
Head of radius
Radial tuberosity
Shaft of radius
Shaft of ulna

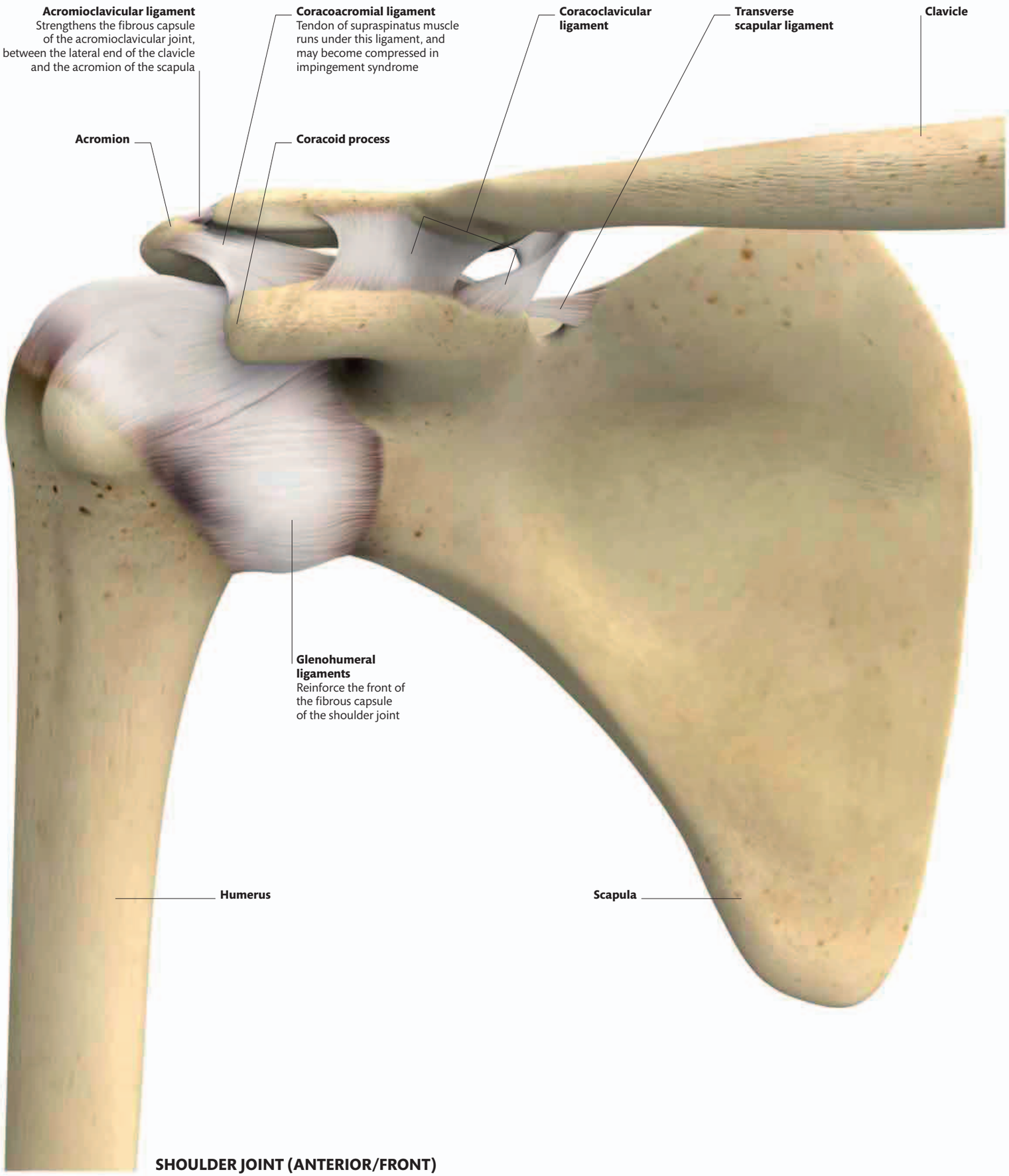

SHOULDER JOINT (ANTERIOR/FRONT)

SHOULDER AND UPPER ARM

In any joint, there is always a play off between mobility and stability. The extremely mobile shoulder joint is therefore naturally unstable, and so it is not surprising that this is the most commonly dislocated joint in the body. The coracoacromial arch, formed by the acromion and coracoid process of the scapula with the strong coracoacromial ligament stretching between them, prevents upward dislocation; when the head of the humerus dislocates, it usually does so in a downward direction. The elbow joint is formed by the articulation of the humerus with the forearm bones: the trochlea articulates with the ulna, and the capitulum with the head of the radius. The elbow is a hinge joint, stabilized by collateral ligaments on each side.

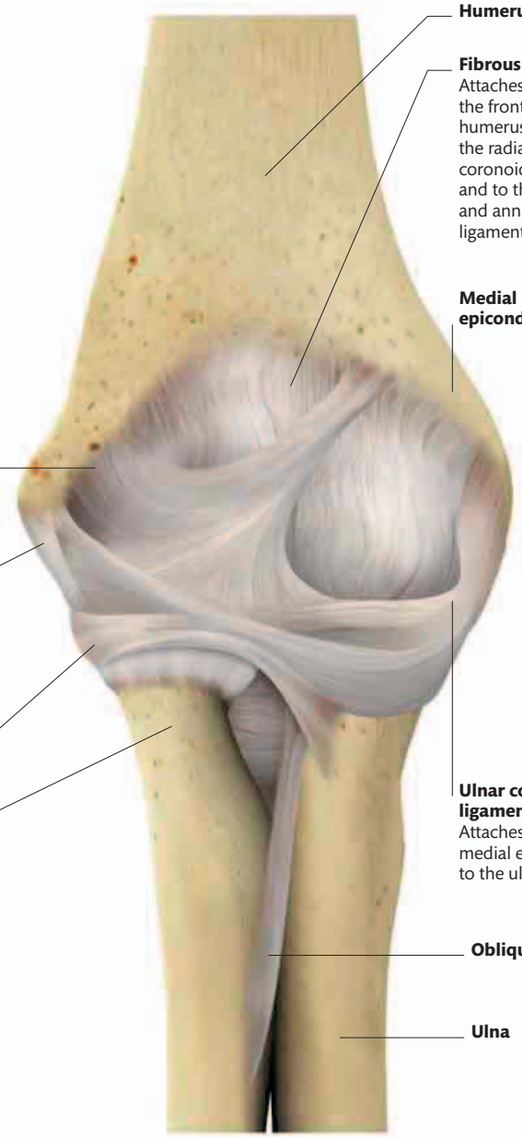

ELBOW (ANTERIOR/FRONT)

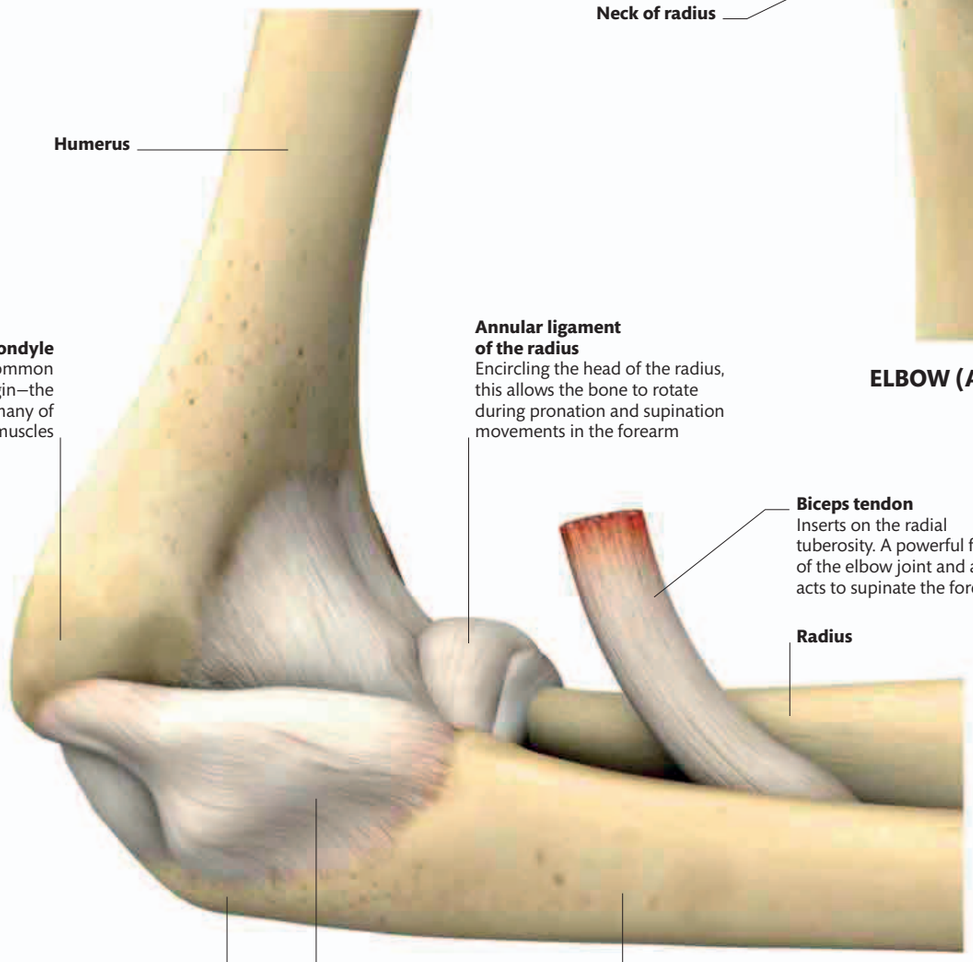

ELBOW (LATERAL/ OUTER SIDE)

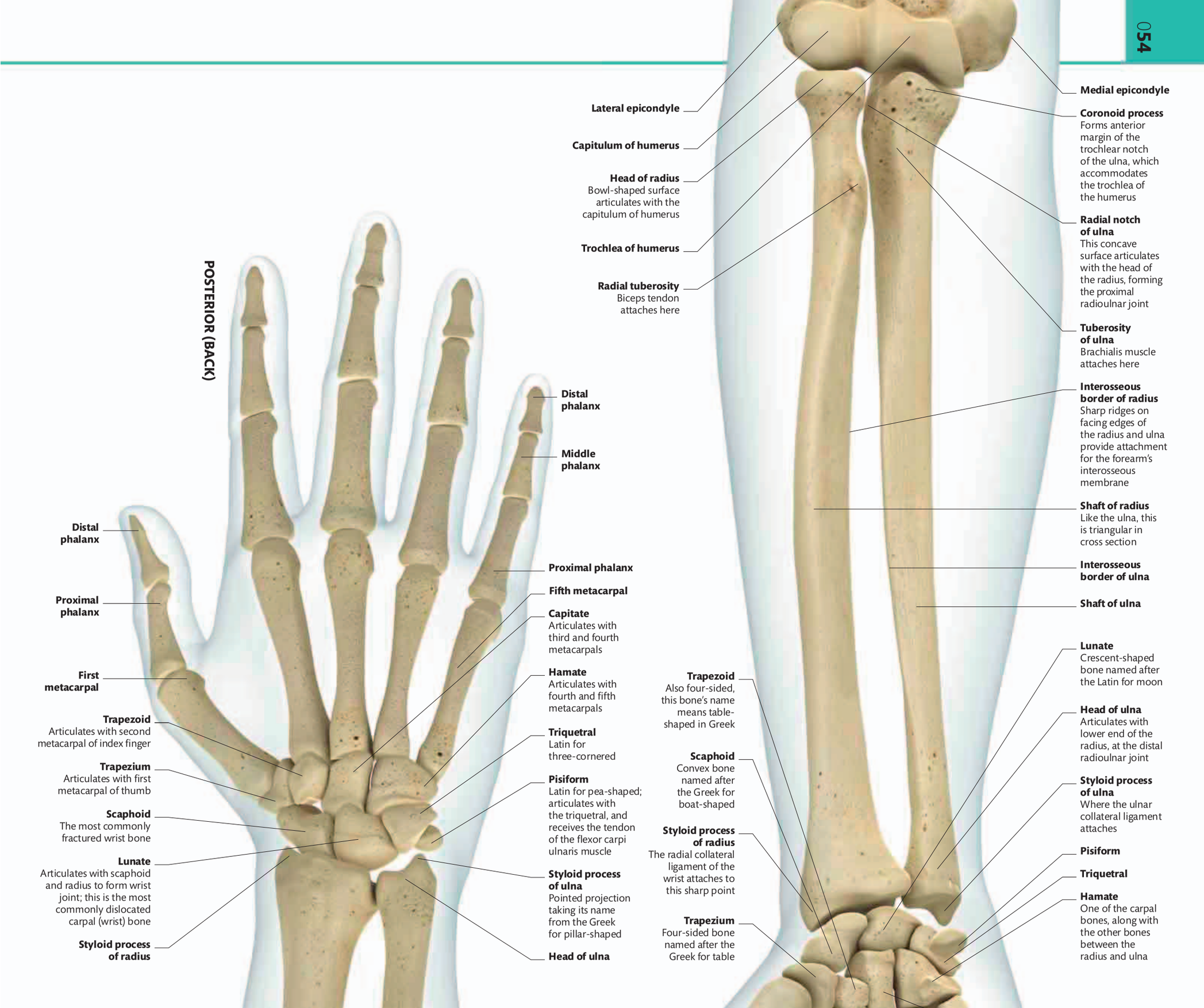
POSTERIOR (BACK)
Distal phalanx
Middle phalanx
Distal phalanx
Proximal phalanx
Proximal phalanx
Fifth metacarpal
Capitate
Articulates with third and fourth metacarpals
First metacarpal
Hamate
Articulates with fourth and fifth metacarpals
Trapezoid
Articulates with second metacarpal of index finger
Triquetral
Latin for three-cornered
Trapezium
Articulates with first metacarpal of thumb
Pisiform
Latin for pea-shaped; articulates with the triquetral, and receives the tendon of the flexor carpi ulnaris muscle
Scaphoid
The most commonly fractured wrist bone
Lunate
Articulates with scaphoid and radius to form wrist joint; this is the most commonly dislocated carpal (wrist) bone
Styloid process of ulna
Pointed projection taking its name from the Greek for pillar-shaped
Styloid process of radius
Head of ulna
Lateral epicondyle
Capitulum of humerus
Head of radius
Bowl-shaped surface articulates with the capitulum of humerus
Trochlea of humerus
Radial tuberosity
Biceps tendon attaches here
Medial epicondyle
Coronoid process
Forms anterior margin of the trochlear notch of the ulna, which accommodates the trochlea of the humerus
Radial notch of ulna
This concave surface articulates with the head of the radius, forming the proximal radioulnar joint
Tuberosity of ulna
Brachialis muscle attaches here
Interosseous border of radius
Sharp ridges on facing edges of the radius and ulna provide attachment for the forearm's interosseous membrane
Shaft of radius
Like the ulna, this is triangular in cross section
Interosseous border of ulna
Shaft of ulna
Trapezoid
Also four-sided, this bone's name means table-shaped in Greek
Scaphoid
Convex bone named after the Greek for boat-shaped
Styloid process of radius
The radial collateral ligament of the wrist attaches to this sharp point
Trapezium
Four-sided bone named after the Greek for table
Lunate
Crescent-shaped bone named after the Latin for moon
Head of ulna
Articulates with lower end of the radius, at the distal radioulnar joint
Styloid process of ulna
Where the ulnar collateral ligament attaches
Pisiform
Triquetral
Hamate
One of the carpal bones, along with the other bones between the radius and ulna

LOWER ARM AND HAND

The two forearm bones, the radius and ulna, are bound together by a flat sheet of ligament called the interosseous membrane, and by synovial joints between the ends of the two bones. Known as radioulnar joints, these joints allow the radius to move around the ulna. Hold your hand out in front of you, palm upward. Now turn your hand so that the palm faces the ground. This movement is called pronation, and is achieved by bringing the radius to cross over the ulna. The movement that returns the palm to an upward-facing position is called supination. Since the forearm bones are bound together by ligaments, joints, and muscles, it is common for both bones to be involved in a serious forearm injury. Often, one bone is fractured and the other dislocated. The skeleton of the hand comprises the eight carpal bones (bones between the radius and ulna), five metacarpals, and fourteen phalanges.

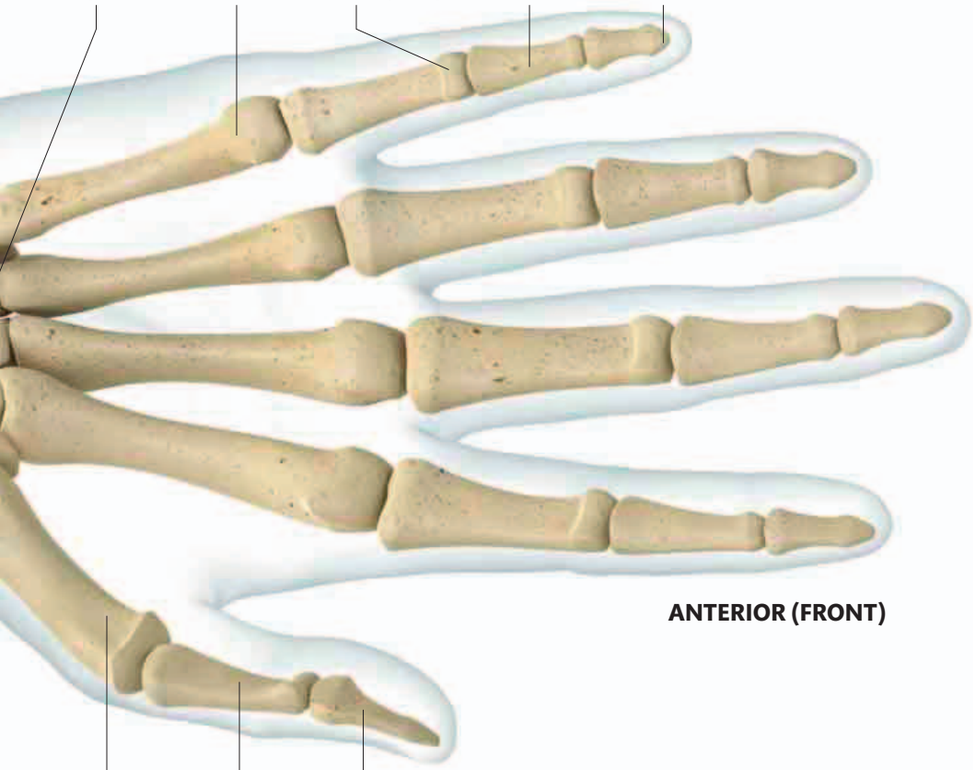

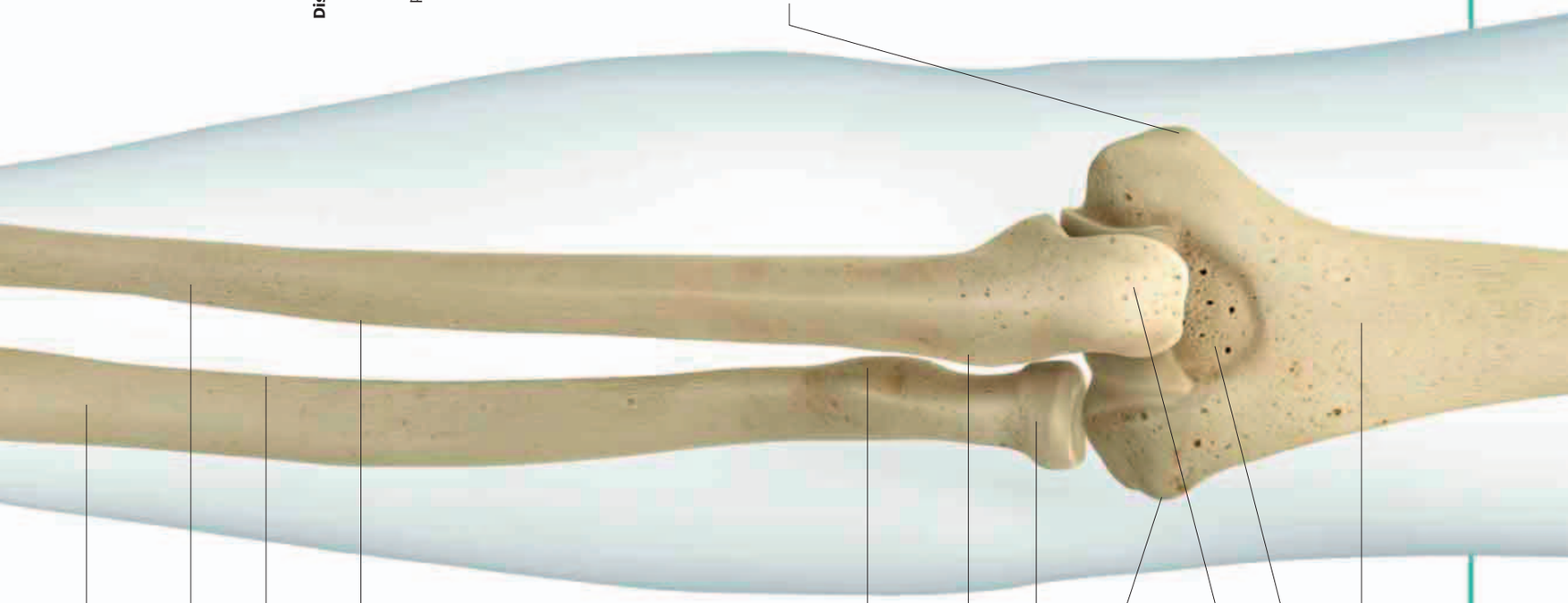

HAND AND WRIST JOINTS

The radius widens out at its distal (lower) end to form the wrist joint with the closest two carpal bones, the lunate and scaphoid. This joint allows flexion, extension, adduction, and abduction (see pp.16–17). There are also synovial joints (see p.60) between the carpal bones in the wrist, which increase the range of motion during wrist flexion and extension. Synovial joints between metacarpals and phalanges allow us to spread or close our fingers, as well as flexing or extending the whole finger. Joints between the individual finger bones, or phalanges, enable fingers to bend and straighten. In common with many other primates, humans have opposable thumbs. The joints at the base of the thumb are shaped differently from those of the fingers. The joint between the metacarpal of the thumb and the wrist bones is especially mobile and allows the thumb to be brought across the palm of the hand so that the tip of the thumb can touch the other fingertips.

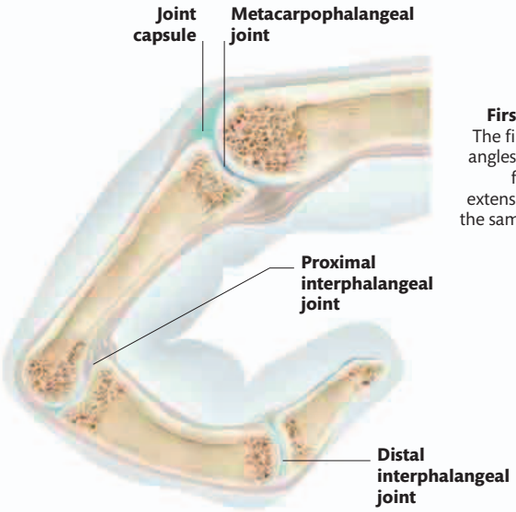

FINGER (SAGITTAL SECTION)

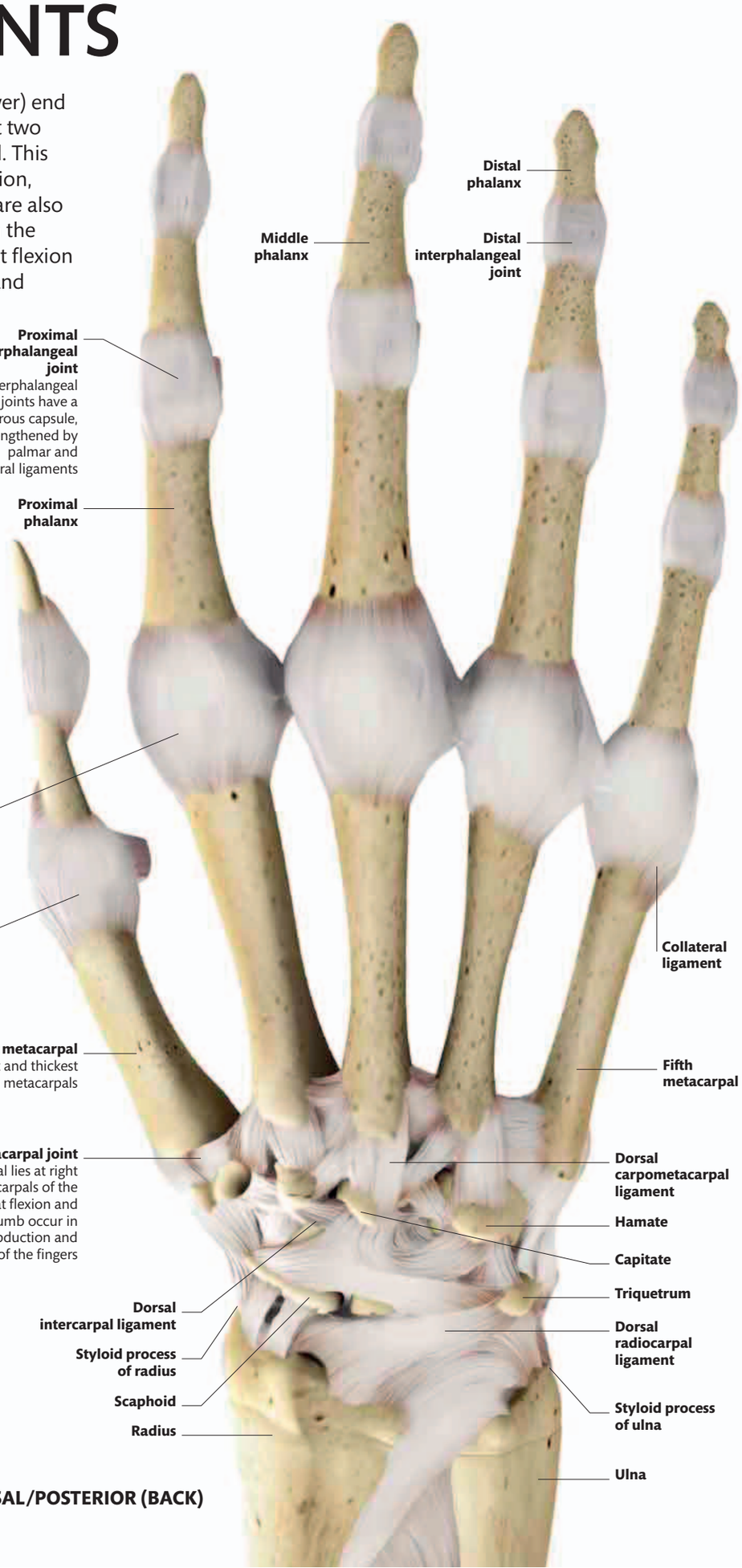

DORSAL/POSTERIOR (BACK)

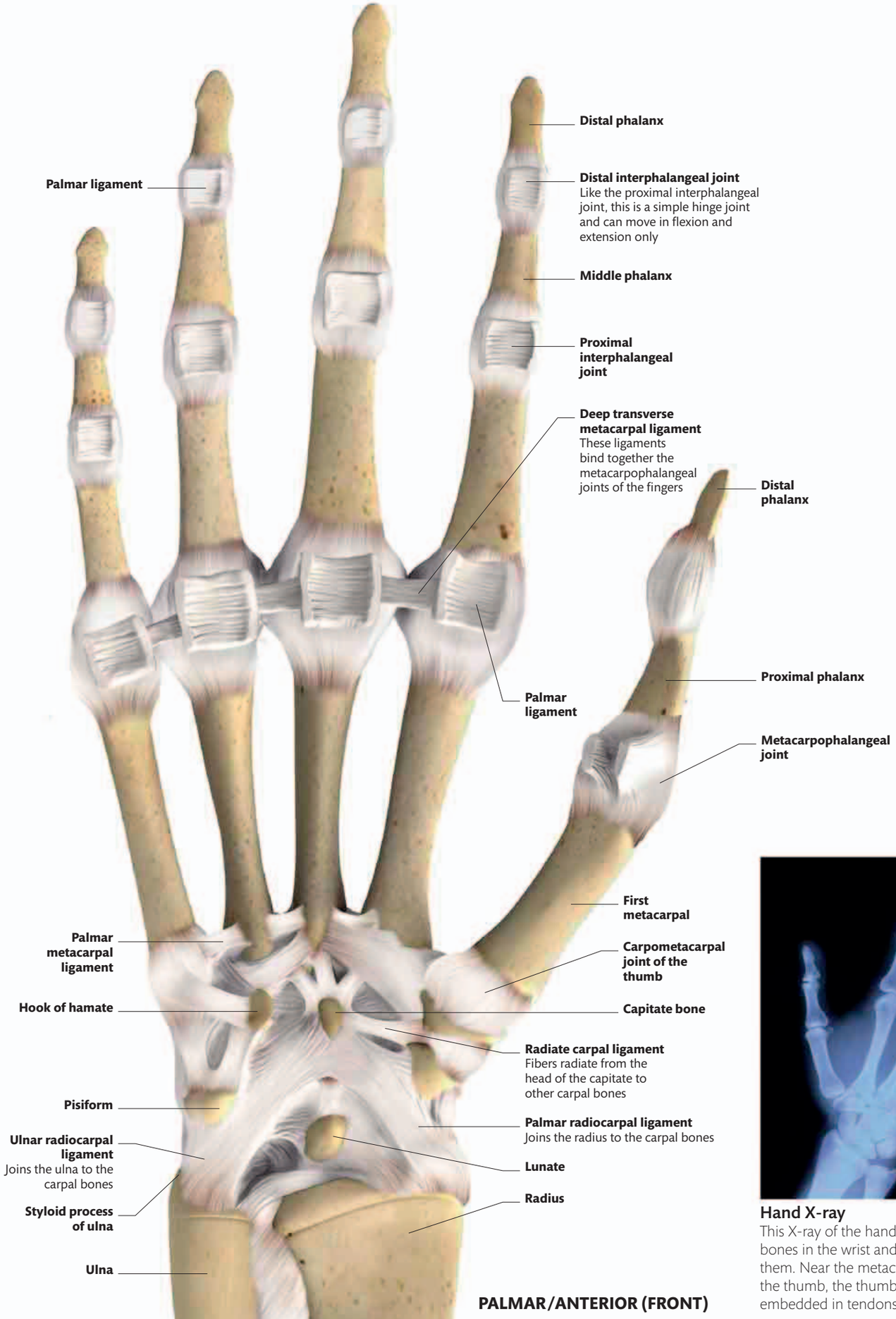

PALMAR/ANTERIOR (FRONT)

Hand X-ray

This X-ray of the hand clearly shows the carpal bones in the wrist and the joints between them. Near the metacarpophalangeal joint of the thumb, the thumb's tiny sesamoid bones, embedded in tendons, are also visible.

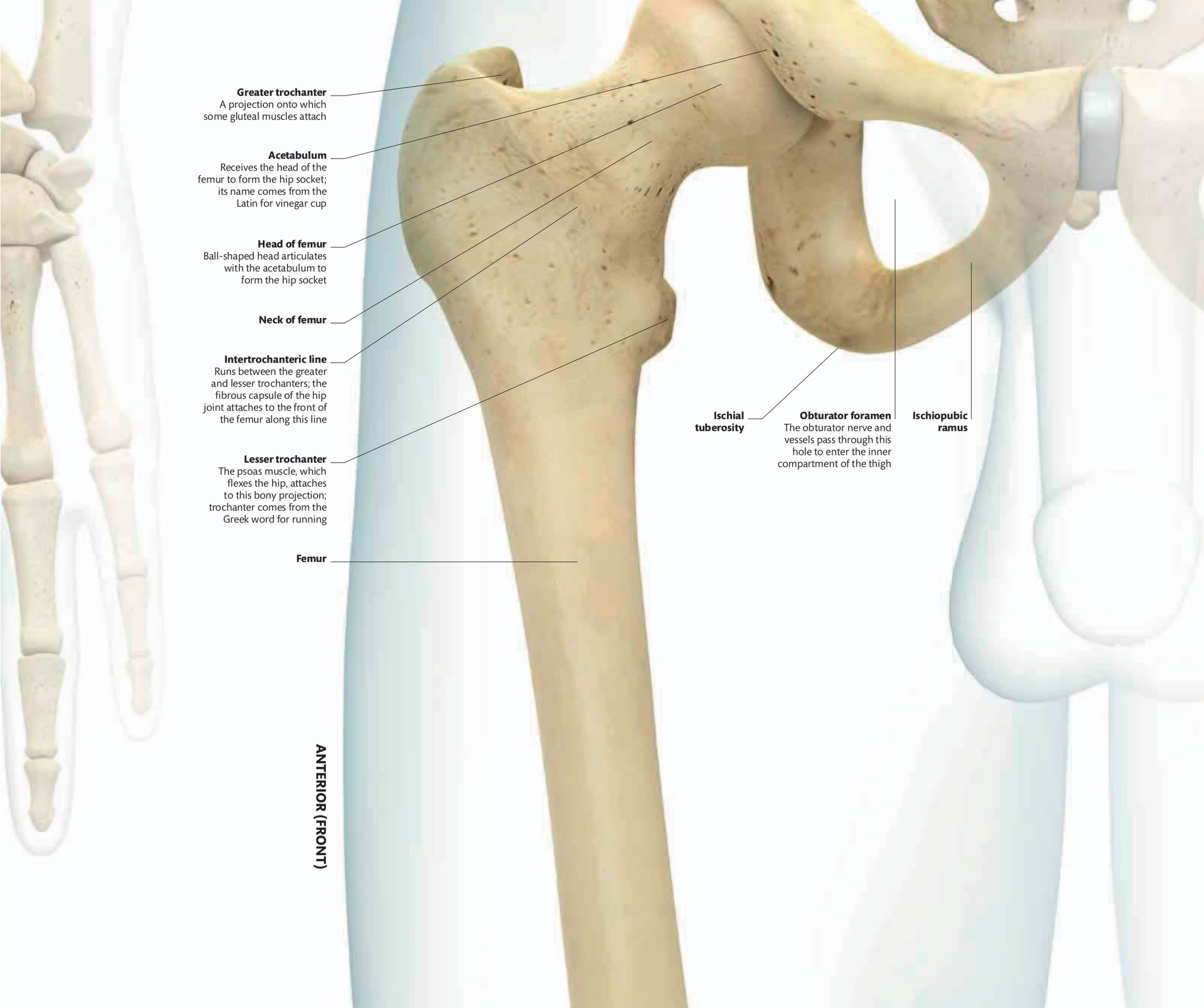
Greater trochanter
A projection onto which some gluteal muscles attach
Acetabulum
Receives the head of the femur to form the hip socket; its name comes from the Latin for vinegar cup
Head of femur
Ball-shaped head articulates with the acetabulum to form the hip socket
Neck of femur
Intertrochanteric line
Runs between the greater and lesser trochanters; the fibrous capsule of the hip joint attaches to the front of the femur along this line
Lesser trochanter
The psoas muscle, which flexes the hip, attaches to this bony projection; trochanter comes from the Greek word for running
Femur
Ischial tuberosity
Obturator foramen
The obturator nerve and vessels pass through this hole to enter the inner compartment of the thigh
Ischiopubic ramus
ANTERIOR (FRONT)

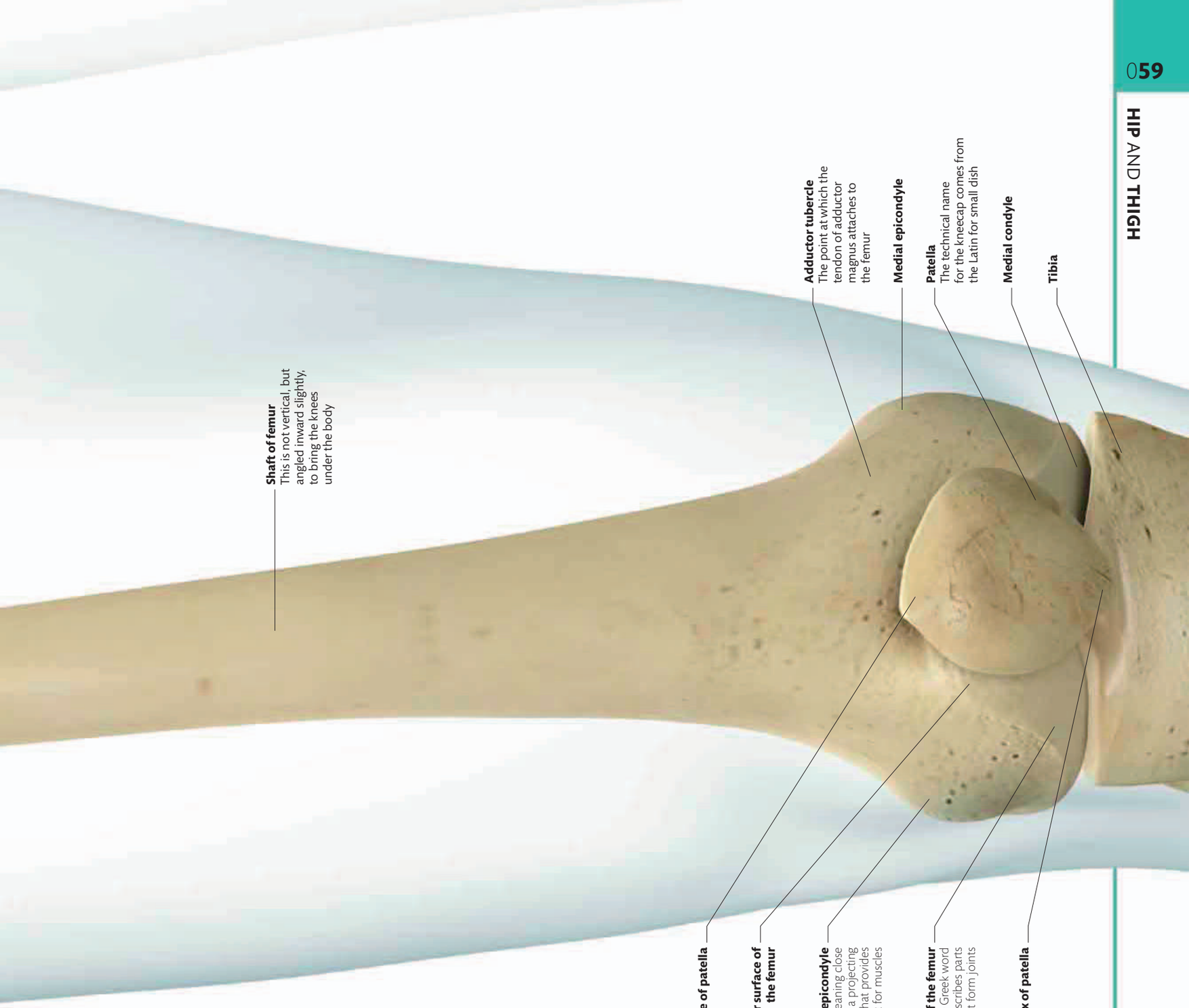

HIP AND THIGH

The leg or, to be anatomically precise, the lower limb, is attached to the spine by the pelvic bones. This is a much more stable arrangement than that of the shoulder girdle, which anchors the arm, because the legs and pelvis must bear our body weight as we stand or move around. The sacroiliac joint provides a strong attachment between the ilium of the pelvis and the sacrum, and the hip joint is a much deeper and more stable ball-and-socket joint than that in the shoulder. The neck of the femur joins the head at an obtuse angle. A slightly raised diagonal line on the front of the neck (the intertrochanteric line) shows where the fibrous capsule of the hip joint attaches to the bone.

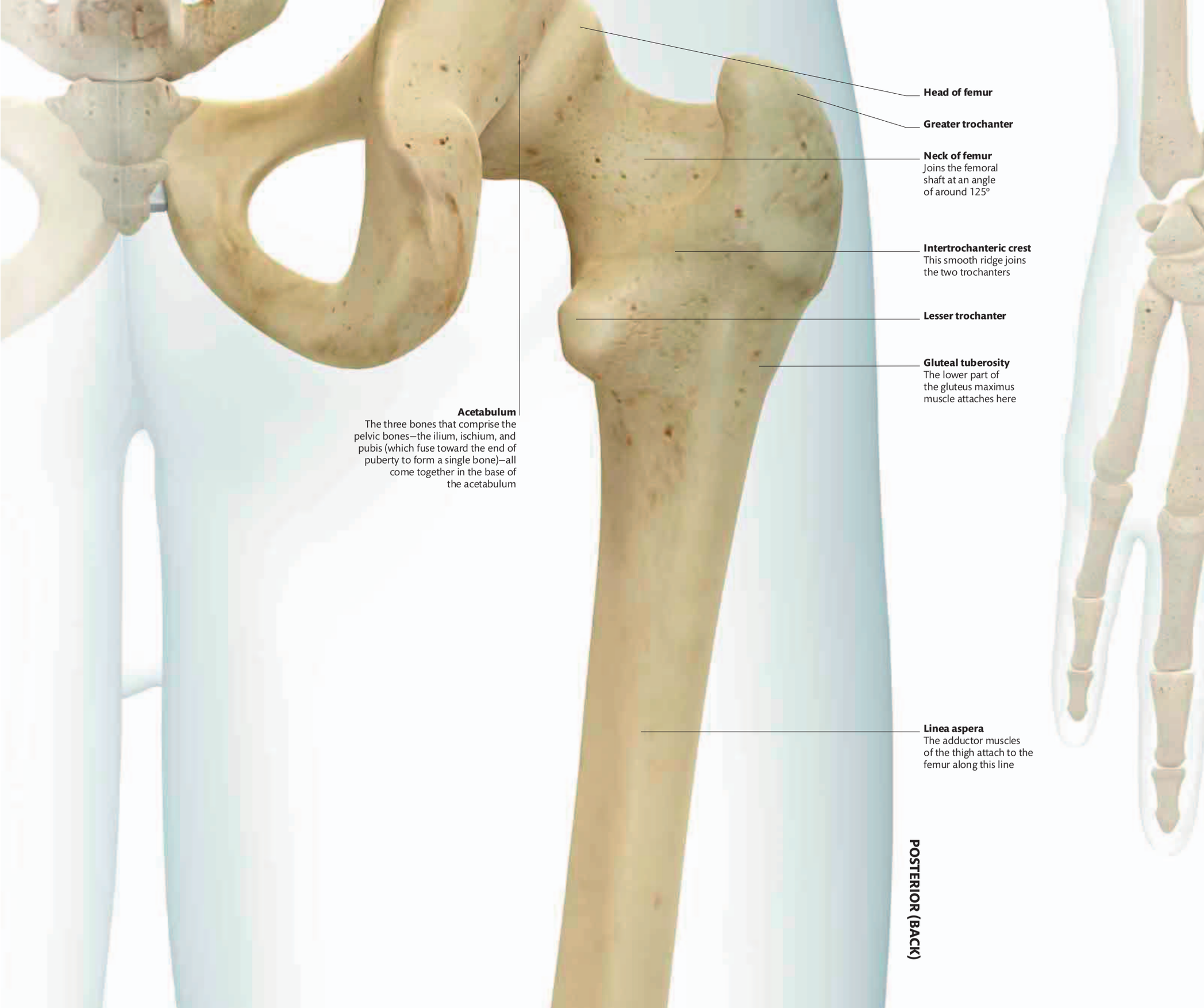
Head of femur
Greater trochanter
Neck of femur
Joins the femoral shaft at an angle of around 125°
Intertrochanteric crest
This smooth ridge joins the two trochanters
Lesser trochanter
Gluteal tuberosity
The lower part of the gluteus maximus muscle attaches here
Acetabulum
The three bones that comprise the pelvic bones—the ilium, ischium, and pubis (which fuse toward the end of puberty to form a single bone)—all come together in the base of the acetabulum
Linea aspera
The adductor muscles of the thigh attach to the femur along this line
POSTERIOR (BACK)

HIP AND THIGH

The shaft of the femur (thighbone) is cylindrical, with a marrow cavity. The linea aspera runs down along the back of the femoral shaft. This line is where the inner thigh's adductor muscles attach to the femur. Parts of the quadriceps muscle also wrap right around the back of the femur to attach to the linea aspera. At the bottom—or distal—end, toward the knee, the femur widens to form the knee joint with the tibia and the patella. From the back, the distal end of the femur has a distinct double-knuckle shape, with two condyles (rounded projections) that articulate with the tibia.

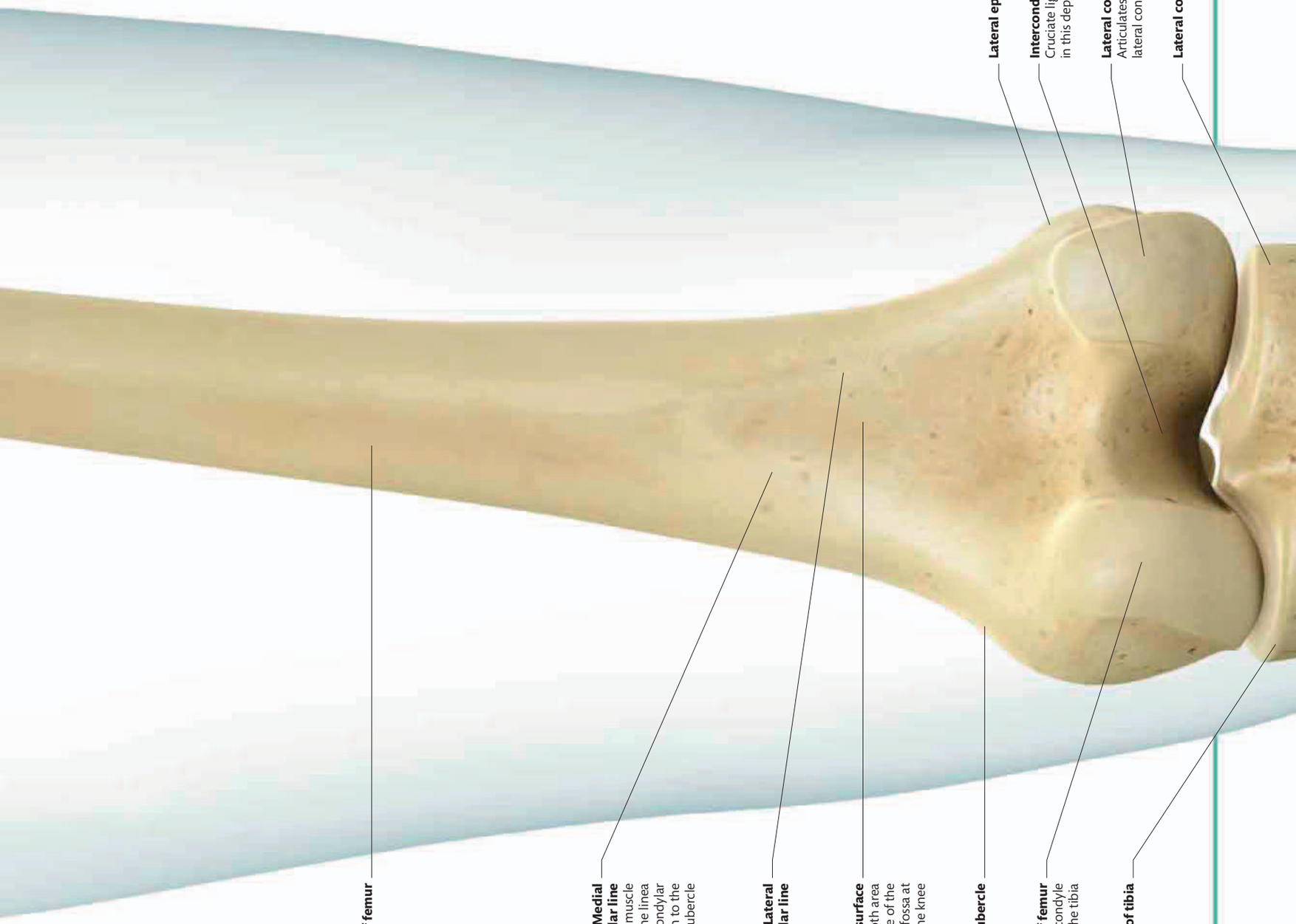

HIP AND KNEE

The hip joint is very stable. Its fibrous capsule is strengthened by ligaments that attach from the neck of the femur to the pelvic bone. These are the iliofemoral and pubofemoral ligaments at the front and the ischiofemoral ligament at the back. Inside the joint capsule, a small ligament attaches from the edge of the acetabulum (hip socket) to the head of the femur. The knee joint is formed by the articulation of the femur with the tibia and patella. Although primarily a hinge joint, the knee also permits some rotation to occur. These complex movements are reflected by the complexity of the joint: there are crescent-shaped articular disks (menisci) inside the joint, powerful collateral ligaments on either side of the joint, as well as crossed-over cruciate ligaments binding the femur to the tibia, and numerous extra pockets of synovial fluid, called bursae, that lubricate tendons around the joint.

Ilium
Ilium, pubis, and ischium of the pelvis all meet in the acetabulum or hip socket

Tendon of rectus femoris
Attaches to the anterior inferior iliac spine

Pubofemoral ligament
Blends with the inner side of the hip capsule

Iliofemoral ligament
This strong ligament strengthens the front of the fibrous capsule of the hip joint

Superior pubic ramus
Forms the upper border of the obturator foramen

Body of pubic bone

Ischiopubic ramus
Forms the lower border of the obturator foramen

Obturator membrane
Covers over the obturator foramen, leaving just a small gap at the top where the obturator nerve and vessels pass out of the pelvis into the thigh

Greater trochanter of femur

Ischium

Ischial tuberosity
Hamstring muscles of the thigh attach here

Intertrochanteric line of the femur
The iliofemoral ligament attaches to the femur along this line

Lesser trochanter of femur

HIP (ANTERIOR/FRONT)

SYNOVIAL JOINTS

The majority of the body's 320 or so joints, including those in the finger, knee, and shoulder, are free-moving synovial joints. The joint surfaces are lined with smooth hyaline cartilage to reduce friction, and contain lubricating synovial fluid.

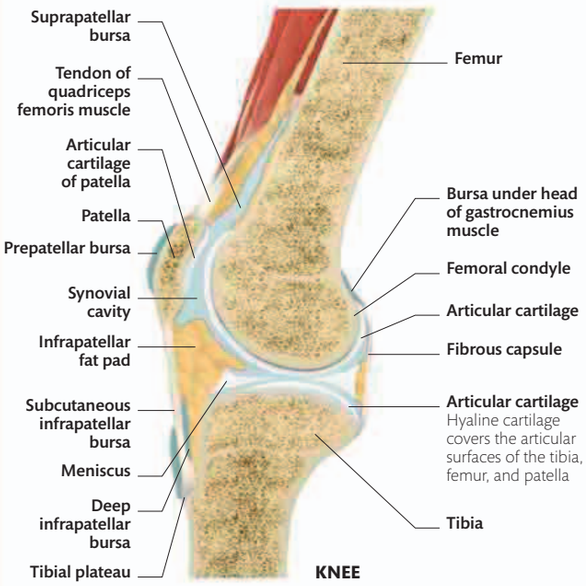

KNEE

Complex joint
A complex synovial joint, such as the knee, has articular discs or menisci inside the synovial cavity. The knee is also a compound hinge joint, as it involves more than two bones. Its complex anatomy allows it to move in flexion and extension, but some sliding and axial rotation of the femur on the tibia also occurs.

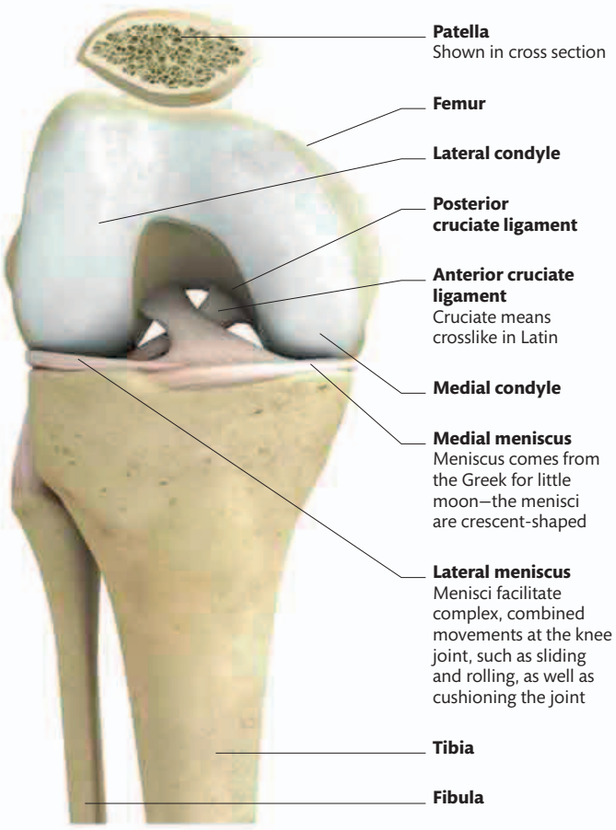

KNEE (FLEXED)

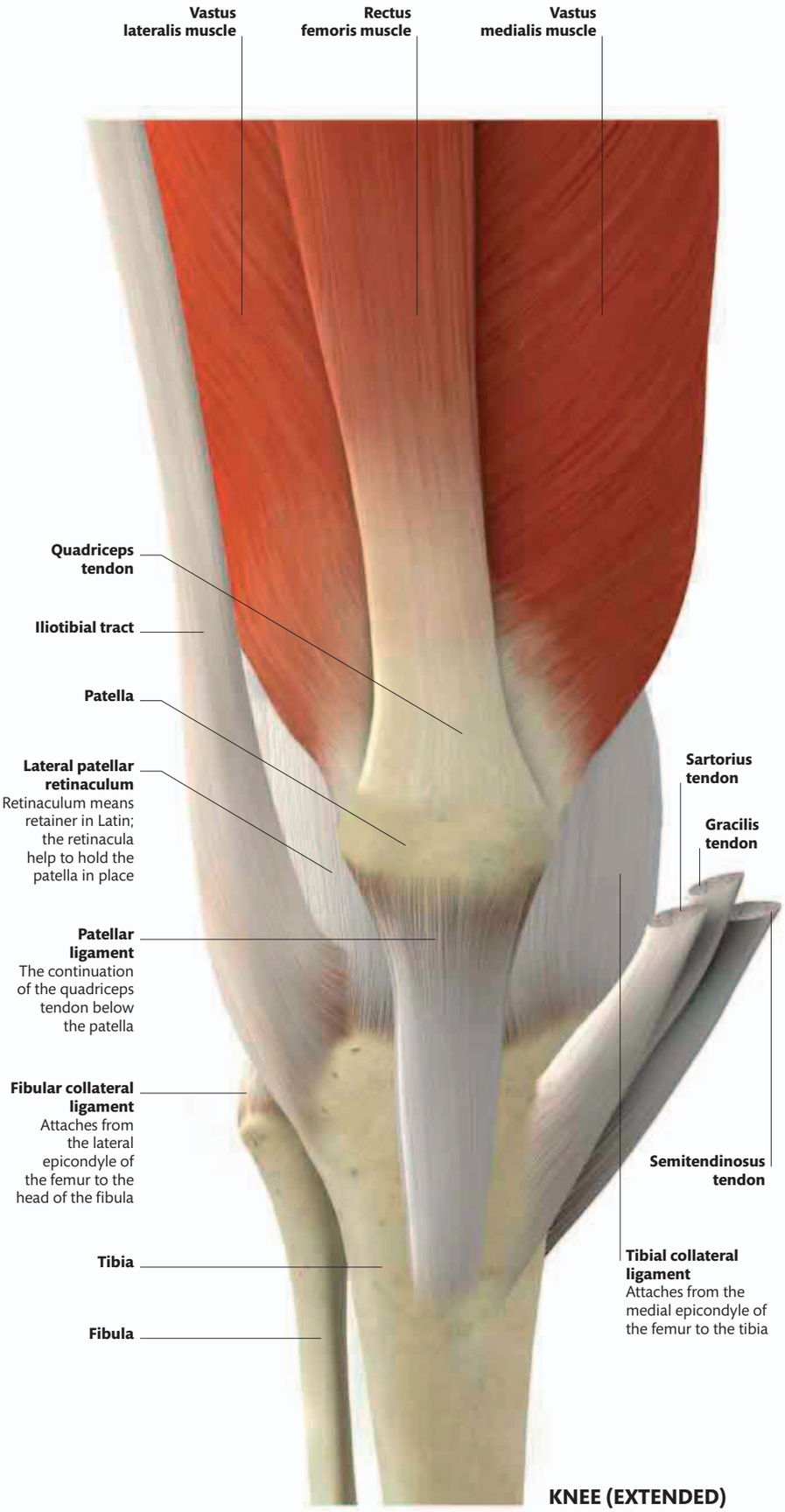

KNEE (EXTENDED)

LOWER LEG AND FOOT

The tibia is the main weight-bearing bone of the lower leg. The fibula, which attaches to the tibia below the knee joint, provides extra areas for the attachment of muscles in the shin and calf and also forms part of the ankle joint. The foot comprises the tarsal bones, metatarsals, and phalanges. The arrangement of these bones is very similar to that of the carpals, metacarpals, and phalanges in the hand. In fact, each limb can be seen to be constructed to a common plan, with a limb girdle providing attachment to the thorax or spine, a single long bone in the first segment, two long bones in the second, a collection of small bones (at the wrist or ankle), and a fan of long, slender bones forming fingers or toes.

LATERAL (OUTSIDE)

Fibula

Tibia

Anterior tibiofibular ligament

Anterior talofibular ligament
One of the ligaments that make up the lateral collateral ligament of the ankle; it connects the lateral malleolus to the neck of the talus

Lateral malleolus

Calcaneofibular ligament
A constituent of the lateral collateral ligament of the ankle

Talus

Navicular

Dorsal tarsometatarsal ligaments

Calcaneus

Dorsal metatarsal ligaments

First metatarsal

Proximal phalanx

Short plantar ligament
Attaches from the calcaneus to the cuboid, and helps to support the lateral arch

Middle phalanx

Calcaneal (Achilles) tendon

Long plantar ligament
Attaching from the calcaneus all the way to the bases of the outer metatarsals, this helps support the arch of the foot on the outer (lateral) side

Cuboid

Fibularis brevis tendon
Attaches to the base of the fifth metacarpal

Distal phalanx

Calcaneonavicular ligament

Calcaneocuboid ligament

Deep transverse metatarsal ligaments

X-ray on tiptoe
This X-ray shows the foot in action. The calf muscles are pulling up on the lever of the calcaneus to flex the ankle down (plantarflex), while the metatarsophalangeal joints are extended.

FOOT AND ANKLE

The ankle joint is a simple hinge joint. The lower ends of the tibia and fibula are firmly bound together by ligaments, forming a strong fibrous joint, and making a wrench shape that neatly sits around the nut of the talus. The joint is stabilized by strong collateral ligaments on either side. The talus forms synovial joints (see p.61) with the calcaneus beneath it, and the navicular bone in front of it. Level with the joint between the talus and the navicular is a joint between the calcaneus and the cuboid. These joints together allow the foot to be angled inward or outward—these movements are called inversion and eversion respectively. The skeleton of the foot is a sprung structure, with the bones forming arches, held together by ligaments and also supported by tendons.

POSTERIOR (BACK)

Top view of the foot bones
This is a dorsal-plantar X-ray of the foot, showing the bones as if you were looking down at your right foot. The two small bones near the head of the first metatarsal are sesamoid bones, embedded in the tendons of the short muscles operating the big toe.

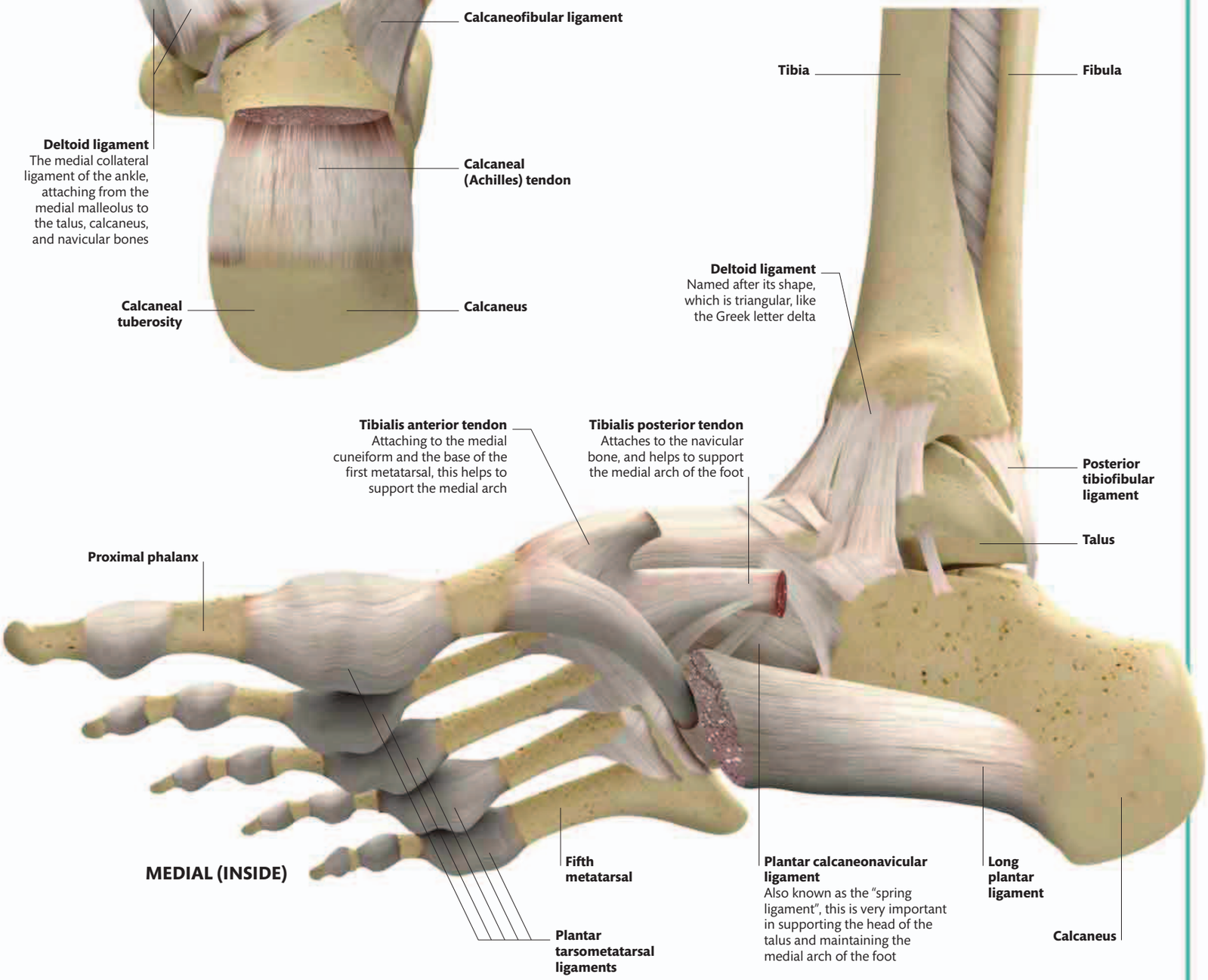

MEDIAL (INSIDE)

MUSCULAR SYSTEM **OVERVIEW**

There are three types of muscle in the body: skeletal, smooth, and cardiac. The main role of skeletal muscles is to generate movement. A muscle's movement, or "action" is produced when it contracts. The force it generates depends on the shape of the muscle. For instance, long, thin muscles contract a lot but exert low forces. Muscles attach to the skeleton by means of tendons, aponeuroses, and connective tissue called fascia. While muscles are well supplied with blood vessels and appear reddish, tendons have a sparse vascular supply and look white. The muscles in our body are located at varied depths. The deep layer sits closest to the bone, while the superficial one lies beneath the skin.

SKELETAL MUSCLE STRUCTURE

Skeletal muscle includes familiar muscles such as biceps or quadriceps. It is built up of parallel bundles of muscle fibers, which are conglomerations of many cells. These muscles are supplied by somatic motor nerves, which are part of the peripheral nervous system and are generally under conscious control.

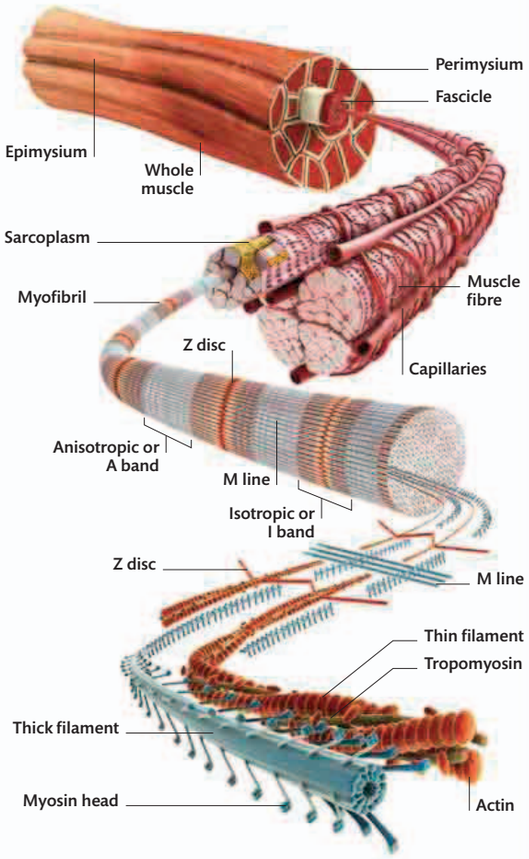

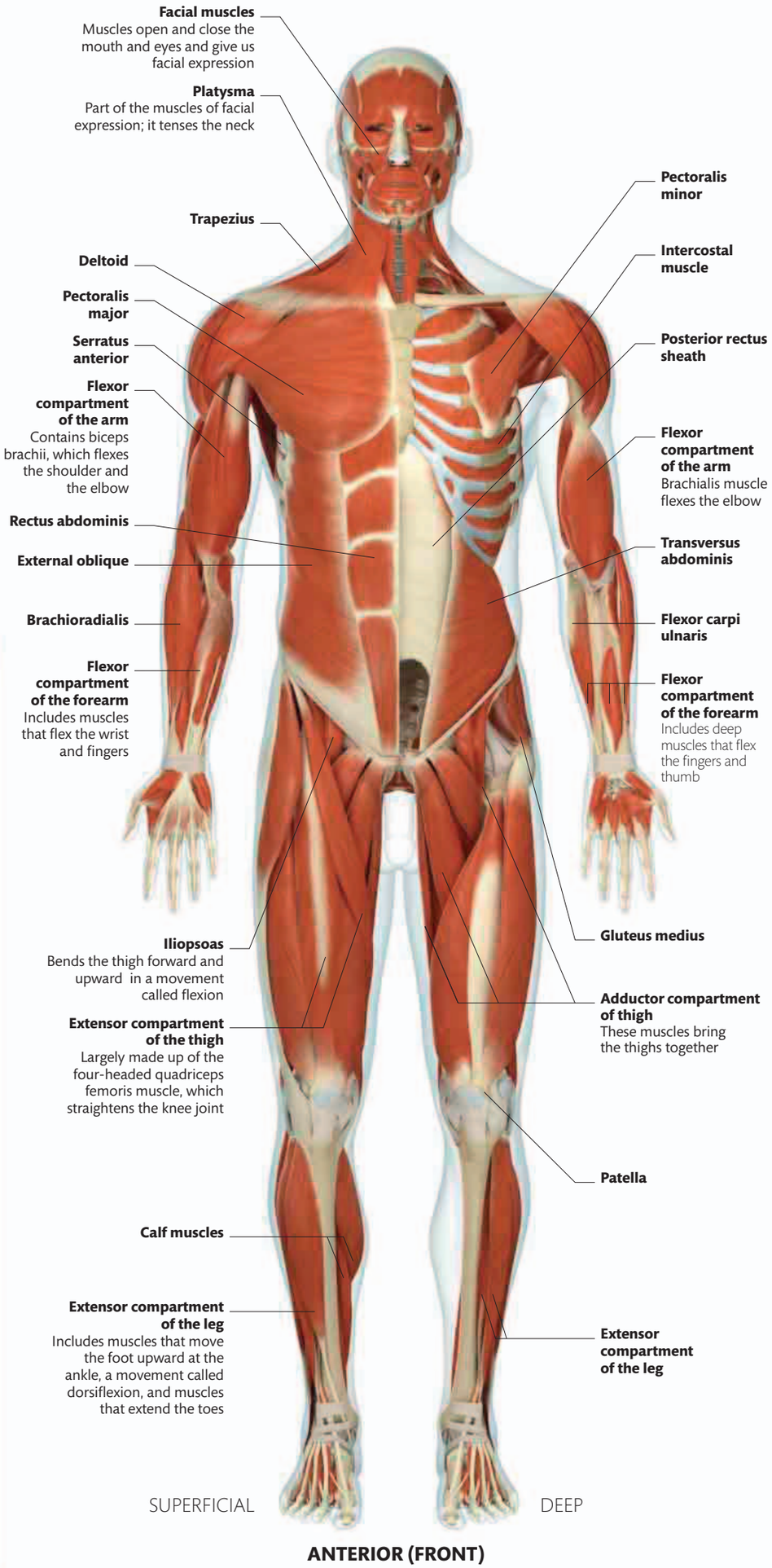

ANTERIOR (FRONT)

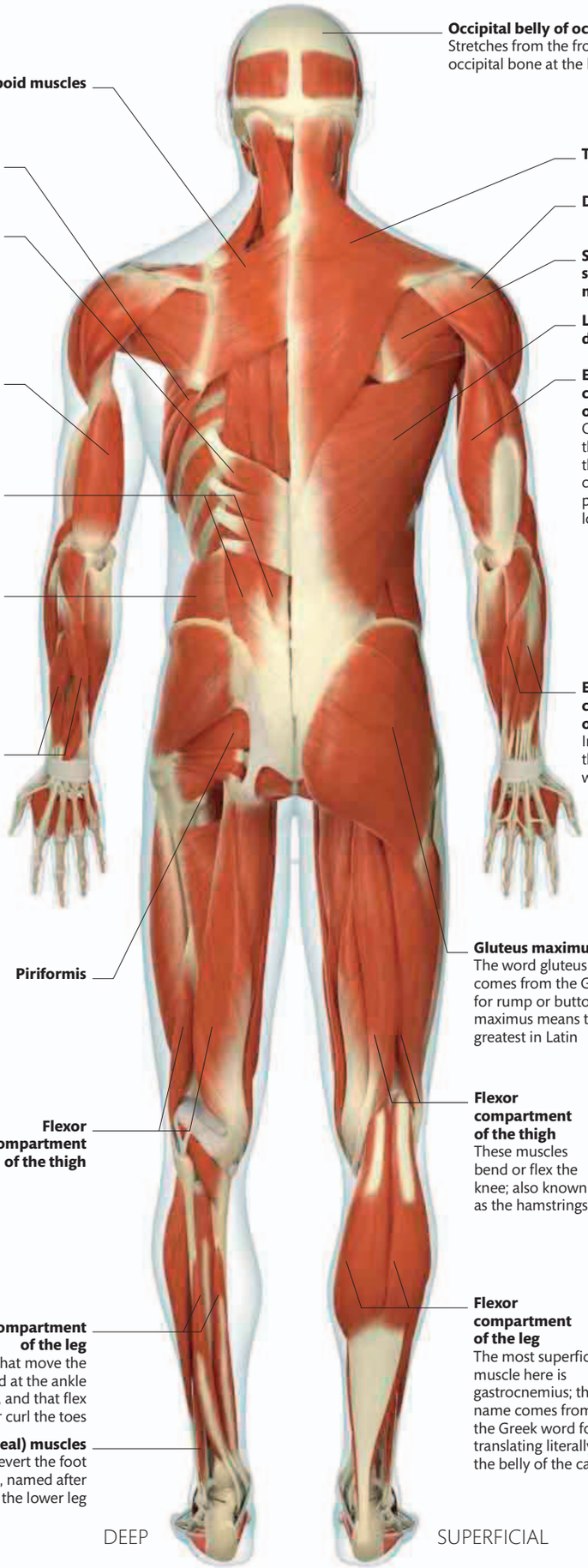

Occipital belly of occipitofrontalis
Stretches from the frontal bone to the occipital bone at the back of the skull

Rhomboid muscles

Serratus anterior

Serratus posterior inferior

Trapezius

Deltoid

Short scapular muscles

Latissimus dorsi

Extensor compartment of the arm

Extensor compartment of the arm
Contains one muscle, the triceps, which means three-headed; here we can see the two superficial parts of the muscle, the long and lateral heads

Erector spinae
This muscle group, as its name suggests, helps to keep the spine erect

Transversus abdominis
The innermost of three sheet-like anterolateral (front-side) abdominal muscles

Extensor compartment of the forearm
Includes deep muscles that extend the thumb or move it out to the side

Extensor compartment of the forearm
Includes muscles that extend the wrist and fingers

Piriformis

Gluteus maximus
The word gluteus comes from the Greek for rump or buttock; maximus means the greatest in Latin

Flexor compartment of the thigh

Flexor compartment of the thigh
These muscles bend or flex the knee; also known as the hamstrings

Flexor compartment of the leg
Deep muscles that move the foot downward at the ankle (plantarflexion), and that flex or curl the toes

Flexor compartment of the leg
The most superficial muscle here is gastrocnemius; the name comes from the Greek word for calf, translating literally as the belly of the calf

Fibular (peroneal) muscles
Two muscles that evert the foot (move it outward), named after the fibula bone in the lower leg

DEEP

SUPERFICIAL

POSTERIOR (BACK)

Temporalis

Deltoid

Brachialis

Short scapular muscles
Problems with these muscles can lead to osteoarthritis at the shoulder joint

Serratus anterior
Anchors the scapula against the chest wall and also helps to move it

Transversus abdominis

Erector spinae

Flexor compartment of the forearm
Some of these muscles originate from the medial epicondyle of the humerus

Gluteus maximus

Quadriceps femoris
Largest muscle in the extensor compartment of the thigh

Flexor compartment of the thigh
Hamstring injuries are common in athletes: the long muscles in this compartment stretch across two joints—the hip and the knee—and are at risk of tearing if overstretched

Flexor compartment of the leg
These muscles combine to form the Achilles tendon, which can be ruptured if overstretched in a sporting injury

Extensor compartment of the leg
The bony attachments of these muscles can become inflamed and painful in the condition "shin splints"

Achilles tendon

SIDE

HEAD AND NECK

The muscles of the face have very important functions. They open and close the apertures in our faces—our eyes, noses, and mouths. But they also play an extremely important role in communication, and this is why these muscles are often known, collectively, as "the muscles of facial expression". These muscles are attached to bone at one end and skin at the other. It is these muscles that allow us to raise our eyebrows in surprise, frown, or knit our brows in concentration, to scrunch up our noses in distaste, to smile gently or to grin widely, and to pout. As we age, and our skin forms creases and wrinkles, these reflect the expressions we have used throughout our lives. The wrinkles and creases lie perpendicular to the direction of the underlying muscle fibers.

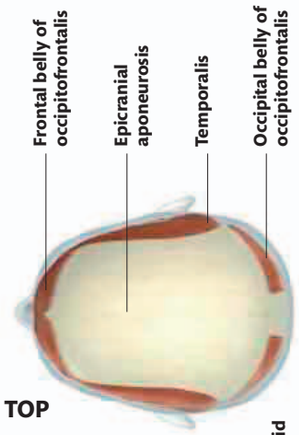

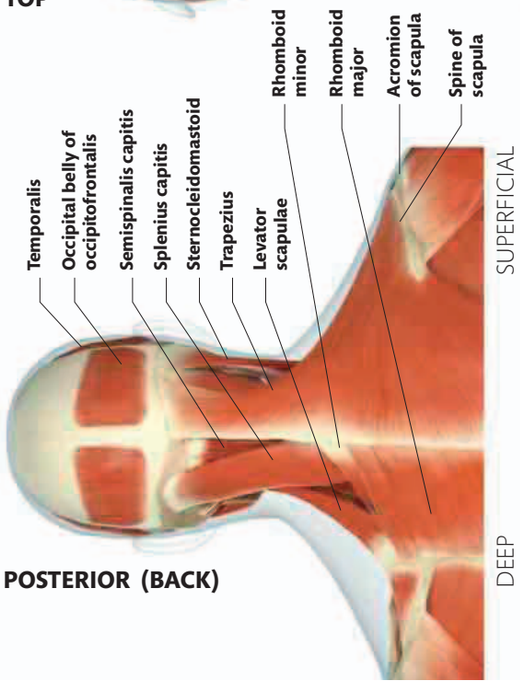

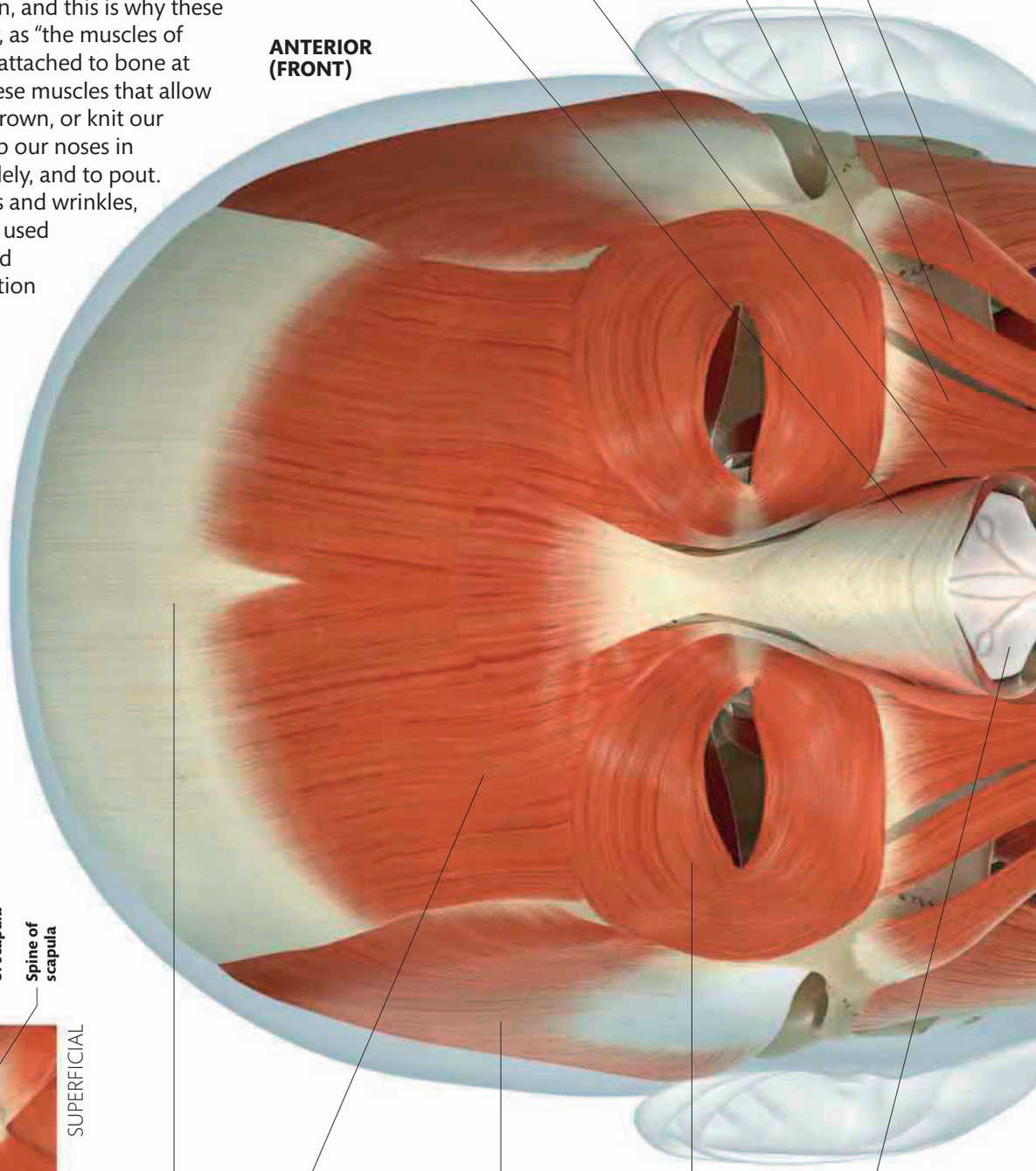

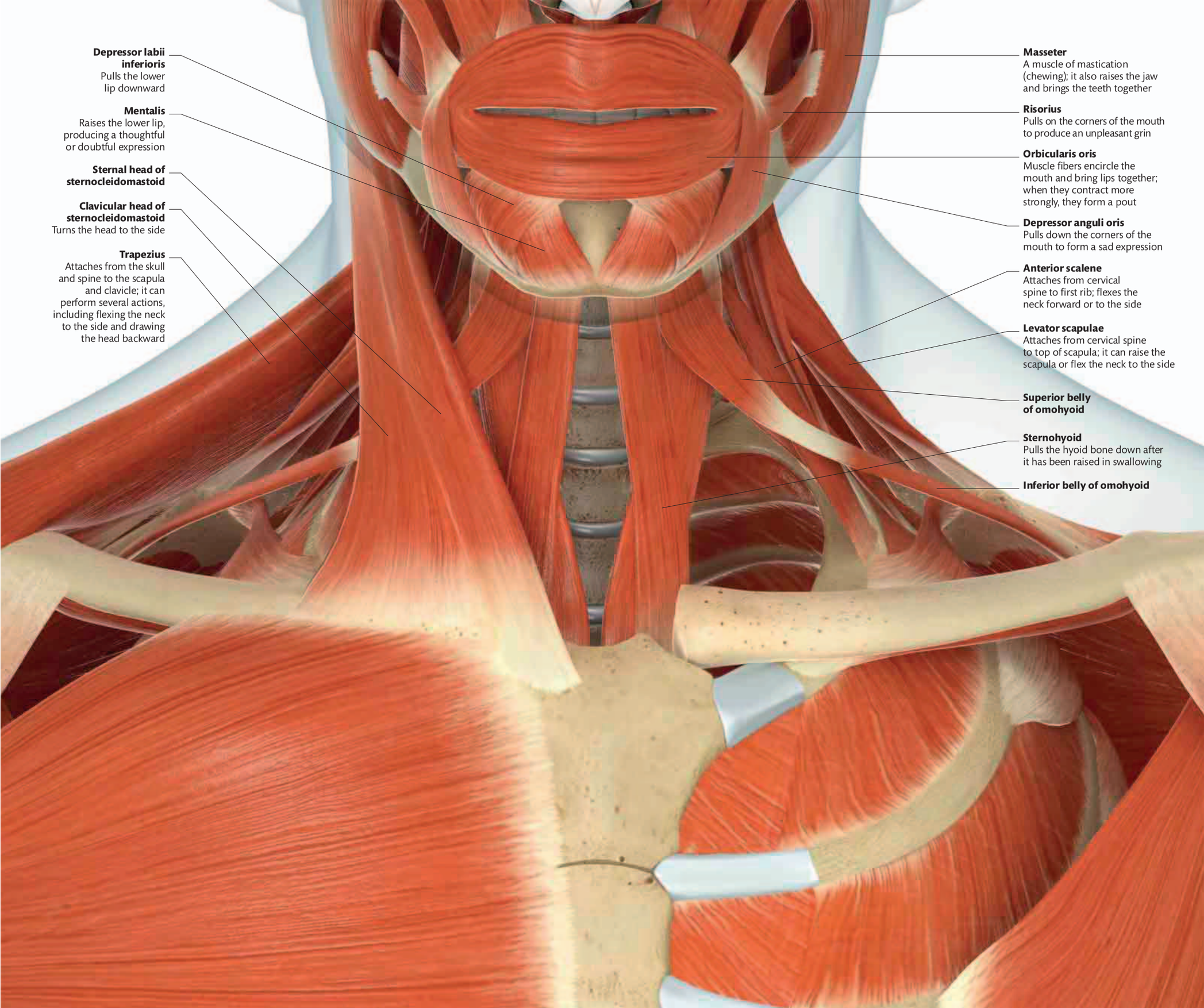
Depressor labii inferioris
Pulls the lower lip downward
Mentalis
Raises the lower lip, producing a thoughtful or doubtful expression
Sternal head of sternocleidomastoid
Clavicular head of sternocleidomastoid
Turns the head to the side
Trapezius
Attaches from the skull and spine to the scapula and clavicle; it can perform several actions, including flexing the neck to the side and drawing the head backward
Masseter
A muscle of mastication (chewing); it also raises the jaw and brings the teeth together
Risorius
Pulls on the corners of the mouth to produce an unpleasant grin
Orbicularis oris
Muscle fibers encircle the mouth and bring lips together; when they contract more strongly, they form a pout
Depressor anguli oris
Pulls down the corners of the mouth to form a sad expression
Anterior scalene
Attaches from cervical spine to first rib; flexes the neck forward or to the side
Levator scapulae
Attaches from cervical spine to top of scapula; it can raise the scapula or flex the neck to the side
Superior belly of omohyoid
Sternohyoid
Pulls the hyoid bone down after it has been raised in swallowing
Inferior belly of omohyoid

HEAD AND NECK

The muscles of mastication (chewing) attach from the skull to the mandible (jawbone), operating to open and shut the mouth, and to grind the teeth together to crush the food we eat. In this side view, we can see the two largest muscles of mastication, the temporalis and masseter muscles. Two smaller muscles attach to the inner surface of the mandible. Human jaws don't just open and close, they also move from side to side, and these four muscles act in concert to produce complex chewing movements. In this view, we can also see how the frontal bellies (fleshy central parts) of the occipitofrontalis muscle are connected to occipital bellies at the back of the head by a thin, flat tendon, or aponeurosis. This makes the entire scalp movable on the skull.

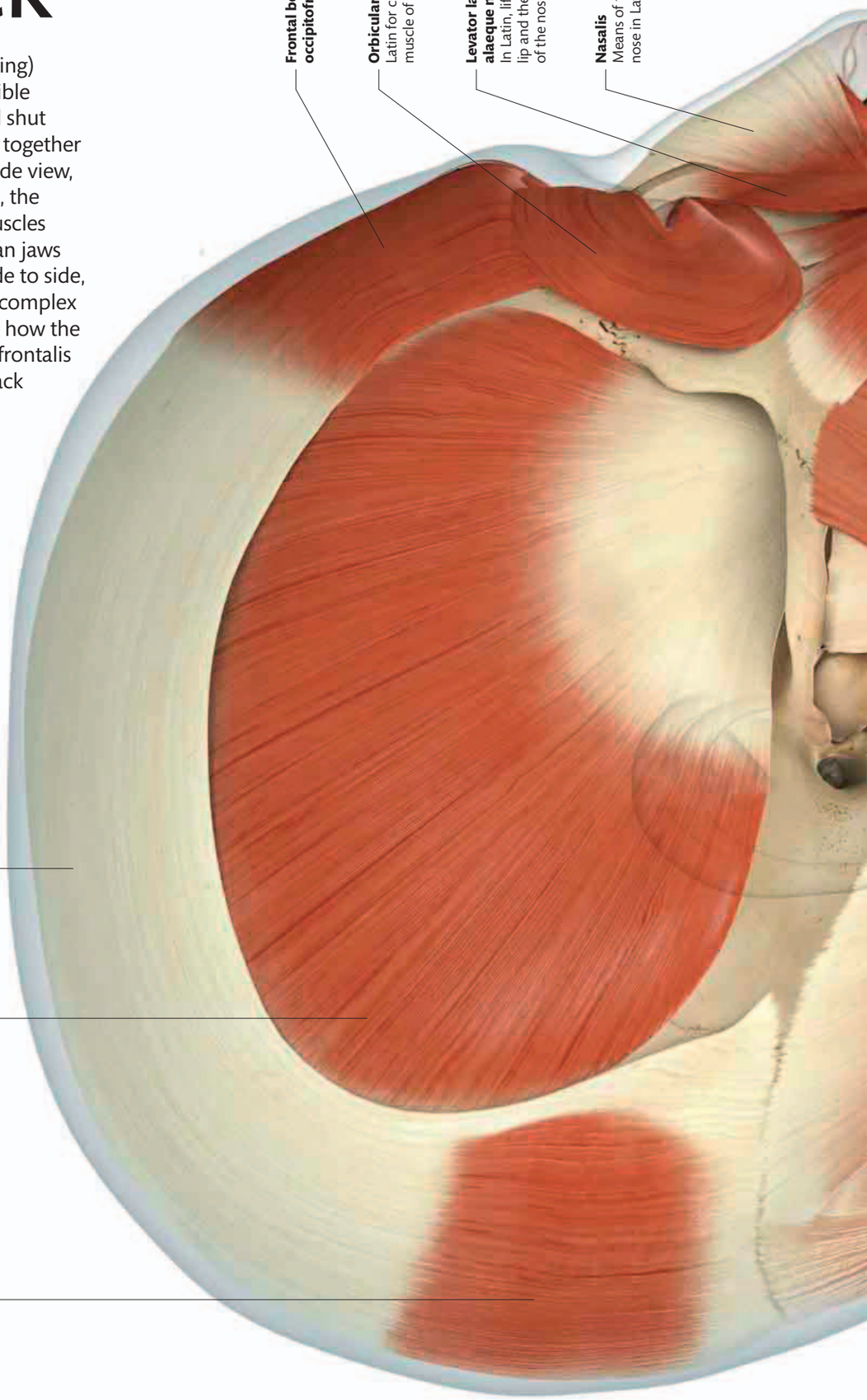

SIDE

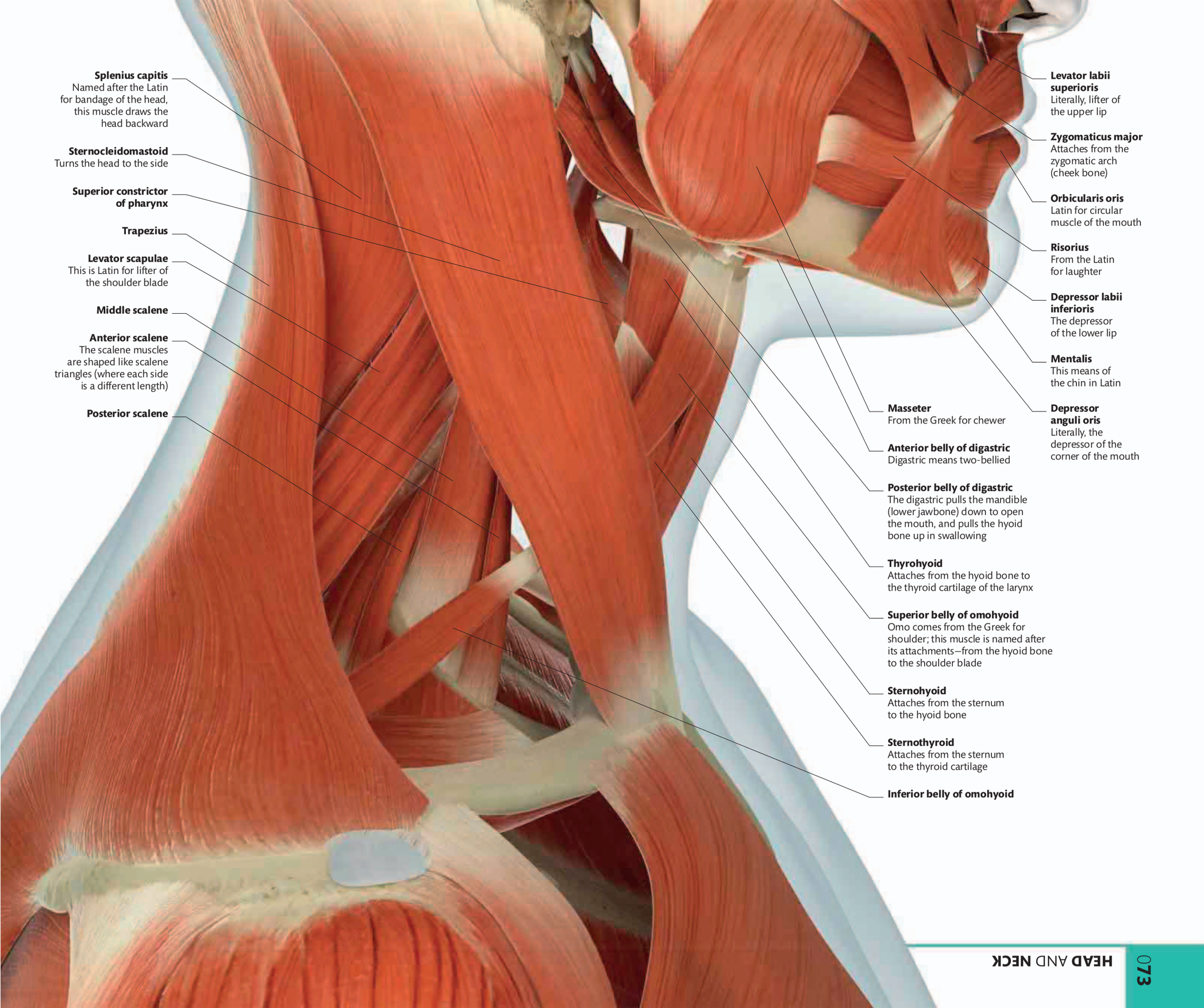
Splenius capitis
Named after the Latin for bandage of the head, this muscle draws the head backward
Sternocleidomastoid
Turns the head to the side
Superior constrictor of pharynx
Trapezius
Levator scapulae
This is Latin for lifter of the shoulder blade
Middle scalene
Anterior scalene
The scalene muscles are shaped like scalene triangles (where each side is a different length)
Posterior scalene
Levator labii superioris
Literally, lifter of the upper lip
Zygomaticus major
Attaches from the zygomatic arch (cheek bone)
Orbicularis oris
Latin for circular muscle of the mouth
Risorius
From the Latin for laughter
Depressor labii inferioris
The depressor of the lower lip
Mentalis
This means of the chin in Latin
Depressor anguli oris
Literally, the depressor of the corner of the mouth
Masseter
From the Greek for chewer
Anterior belly of digastric
Digastric means two-bellied
Posterior belly of digastric
The digastric pulls the mandible (lower jawbone) down to open the mouth, and pulls the hyoid bone up in swallowing
Thyrohyoid
Attaches from the hyoid bone to the thyroid cartilage of the larynx
Superior belly of omohyoid
Omo comes from the Greek for shoulder; this muscle is named after its attachments—from the hyoid bone to the shoulder blade
Sternohyoid
Attaches from the sternum to the hyoid bone
Sternothyroid
Attaches from the sternum to the thyroid cartilage
Inferior belly of omohyoid

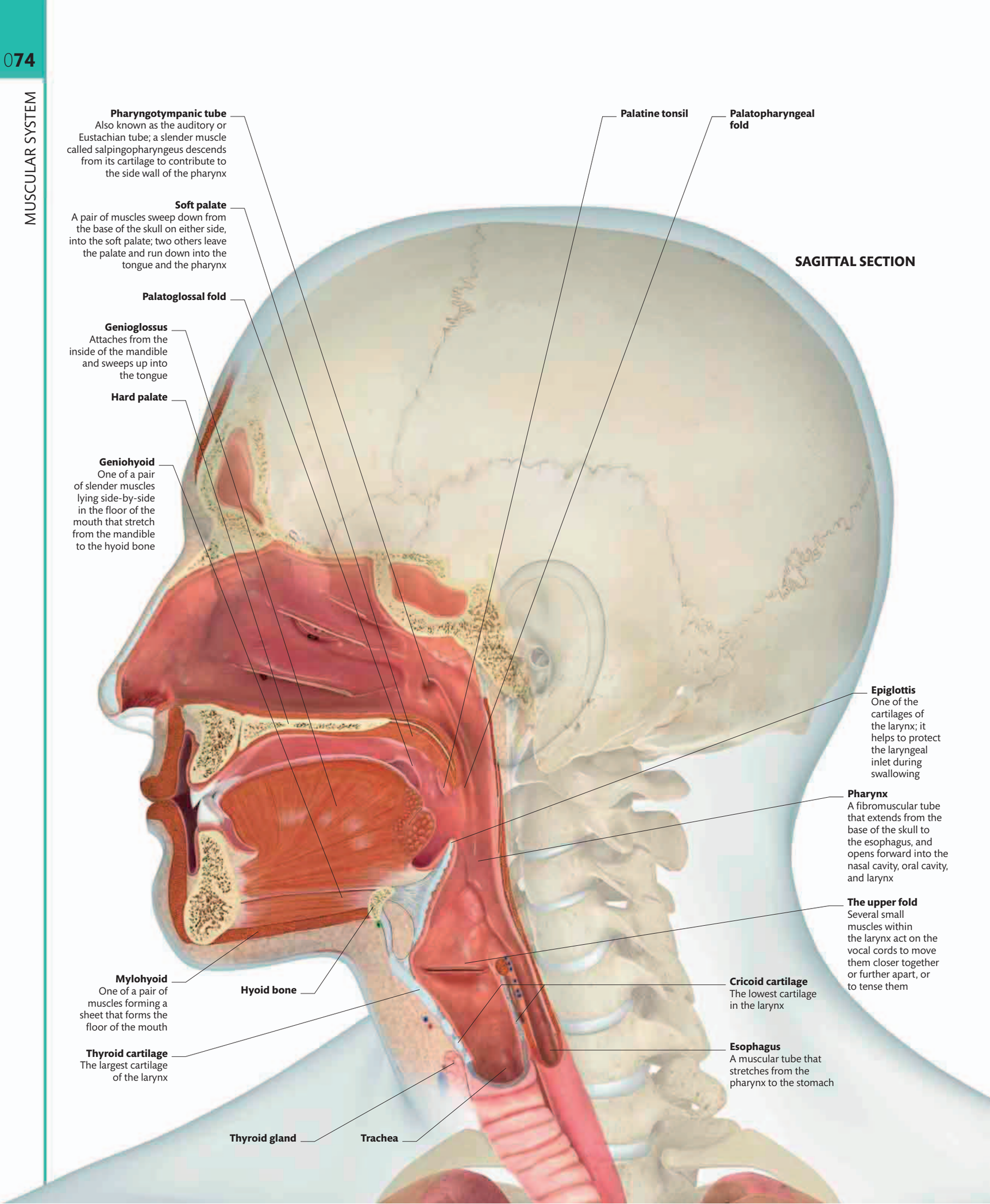
SAGITTAL SECTION
Pharyngotympanic tube
Also known as the auditory or Eustachian tube; a slender muscle called salpingopharyngeus descends from its cartilage to contribute to the side wall of the pharynx
Soft palate
A pair of muscles sweep down from the base of the skull on either side, into the soft palate; two others leave the palate and run down into the tongue and the pharynx
Palatoglossal fold
Genioglossus
Attaches from the inside of the mandible and sweeps up into the tongue
Hard palate
Geniohyoid
One of a pair of slender muscles lying side-by-side in the floor of the mouth that stretch from the mandible to the hyoid bone
Palatine tonsil
Palatopharyngeal fold
Epiglottis
One of the cartilages of the larynx; it helps to protect the laryngeal inlet during swallowing
Pharynx
A fibromuscular tube that extends from the base of the skull to the esophagus, and opens forward into the nasal cavity, oral cavity, and larynx
The upper fold
Several small muscles within the larynx act on the vocal cords to move them closer together or further apart, or to tense them
Mylohyoid
One of a pair of muscles forming a sheet that forms the floor of the mouth
Hyoid bone
Thyroid cartilage
The largest cartilage of the larynx
Cricoid cartilage
The lowest cartilage in the larynx
Esophagus
A muscular tube that stretches from the pharynx to the stomach
Thyroid gland
Trachea

HEAD AND NECK

In the section through the head (opposite), we see the soft palate, tongue, pharynx, and larynx, all of which contain muscles. The soft palate comprises five pairs of muscles. When relaxed, it hangs down at the back of the mouth but, during swallowing, it thickens and is drawn upward to block off the airway. The tongue is a great mass of muscle, covered in mucosa. Some of its muscles arise from the hyoid bone and the mandible, and anchor it to these bones and move it around. Other muscle fibers are entirely within the tongue and change its shape. The pharyngeal muscles are important in swallowing, and the laryngeal muscles control the vocal cords. The muscles that move the eye can be seen on p.122.

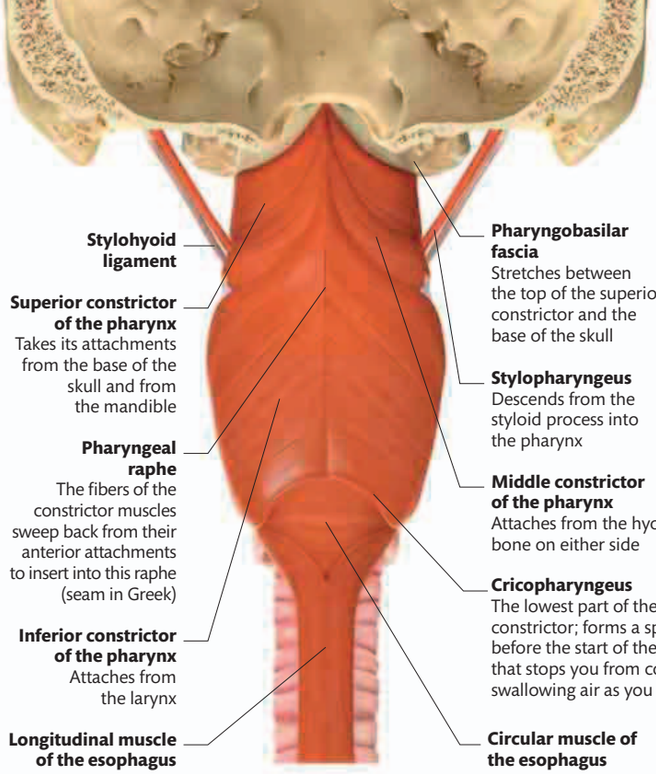

PHARYNX POSTERIOR (BACK)

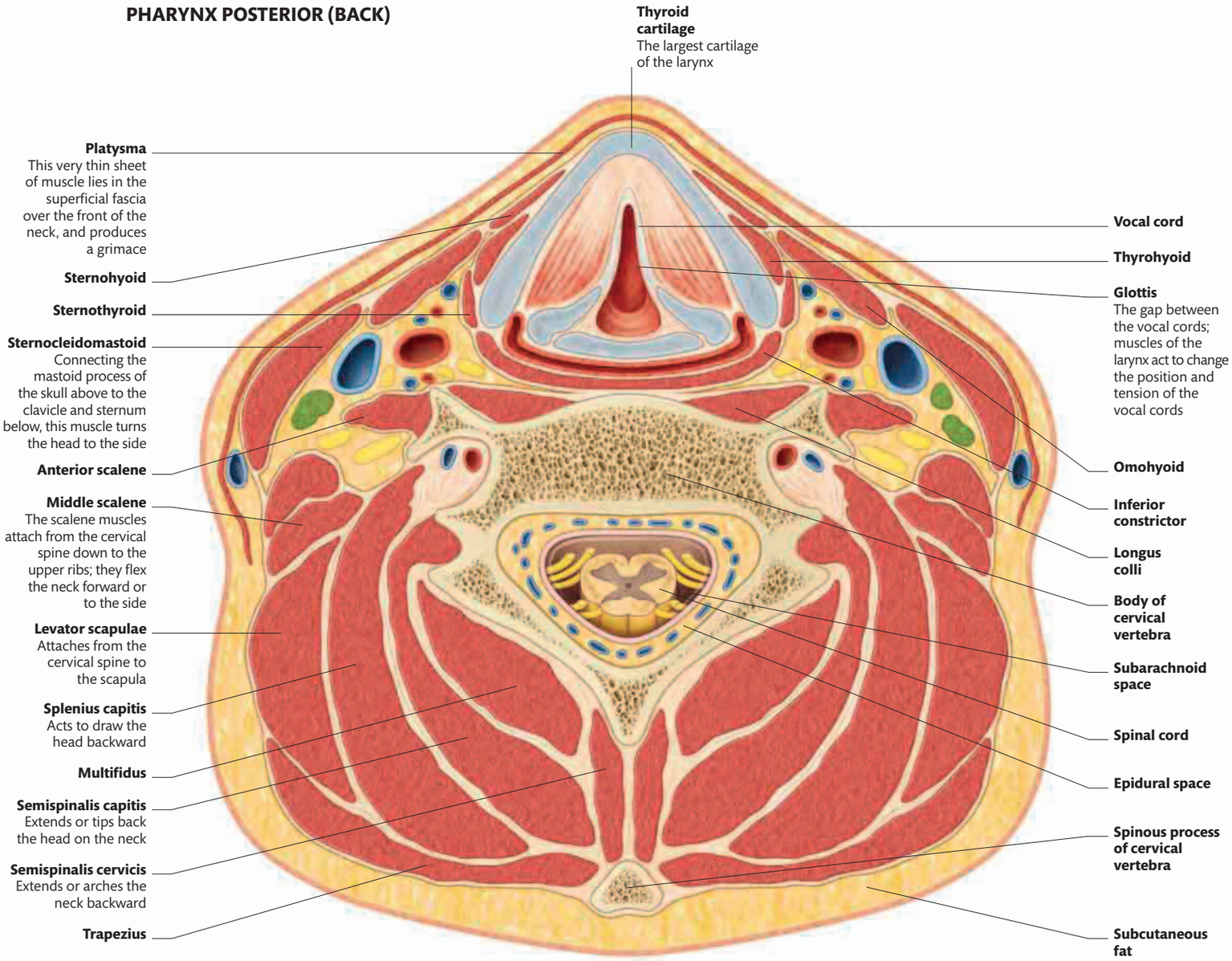

TRANSVERSE SECTION OF THE NECK AT THE VOCAL CORDS

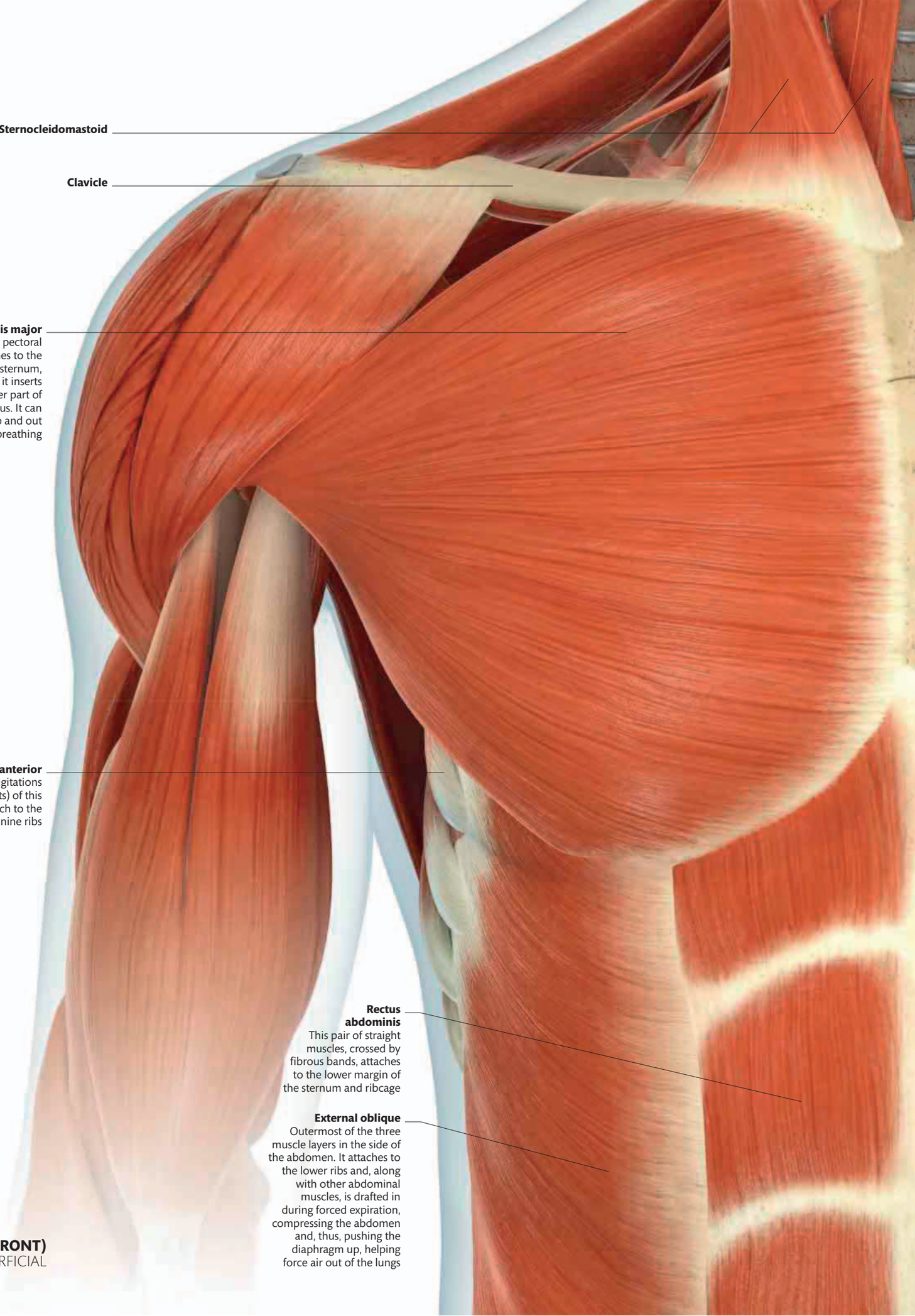

ANTERIOR (FRONT)
SUPERFICIAL

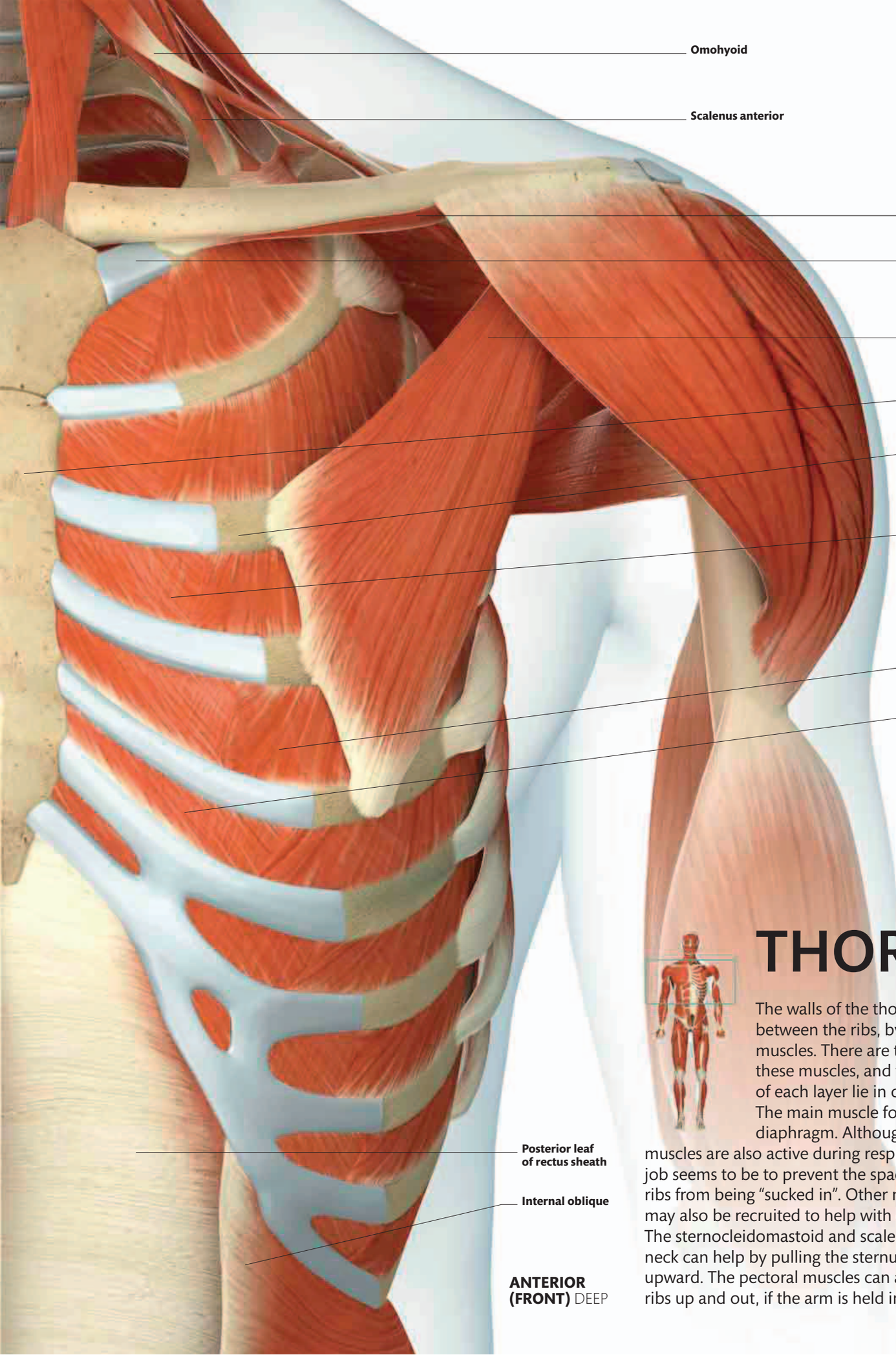

THORAX

The walls of the thorax are filled in, between the ribs, by the intercostal muscles. There are three layers of these muscles, and the muscle fibers of each layer lie in different directions. The main muscle for breathing is the diaphragm. Although the intercostal muscles are also active during respiration, their main job seems to be to prevent the spaces between the ribs from being "sucked in". Other muscles seen here may also be recruited to help with deep breathing. The sternocleidomastoid and scalene muscles in the neck can help by pulling the sternum and upper ribs upward. The pectoral muscles can also pull the ribs up and out, if the arm is held in a fixed position.

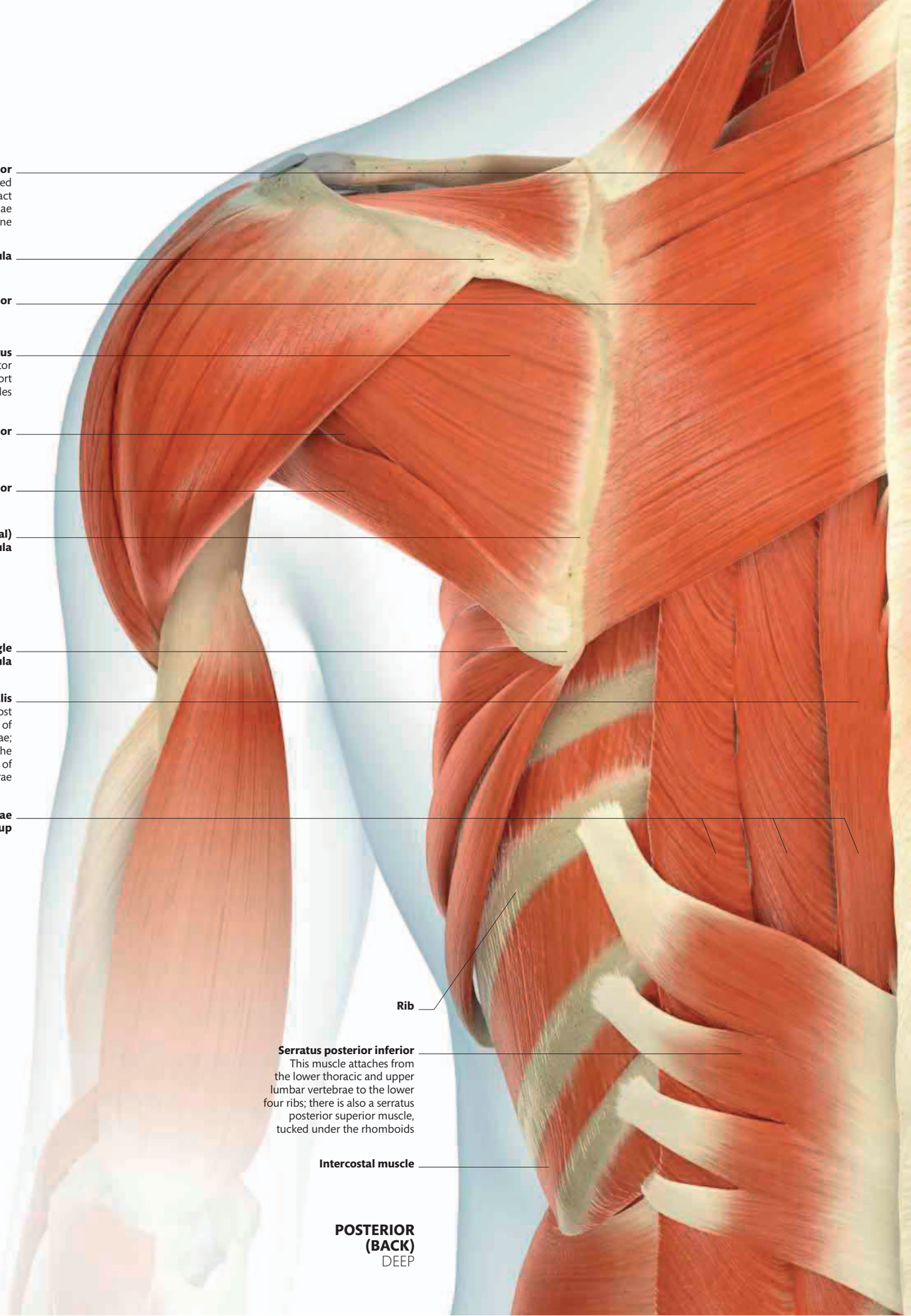

POSTERIOR (BACK)
DEEP

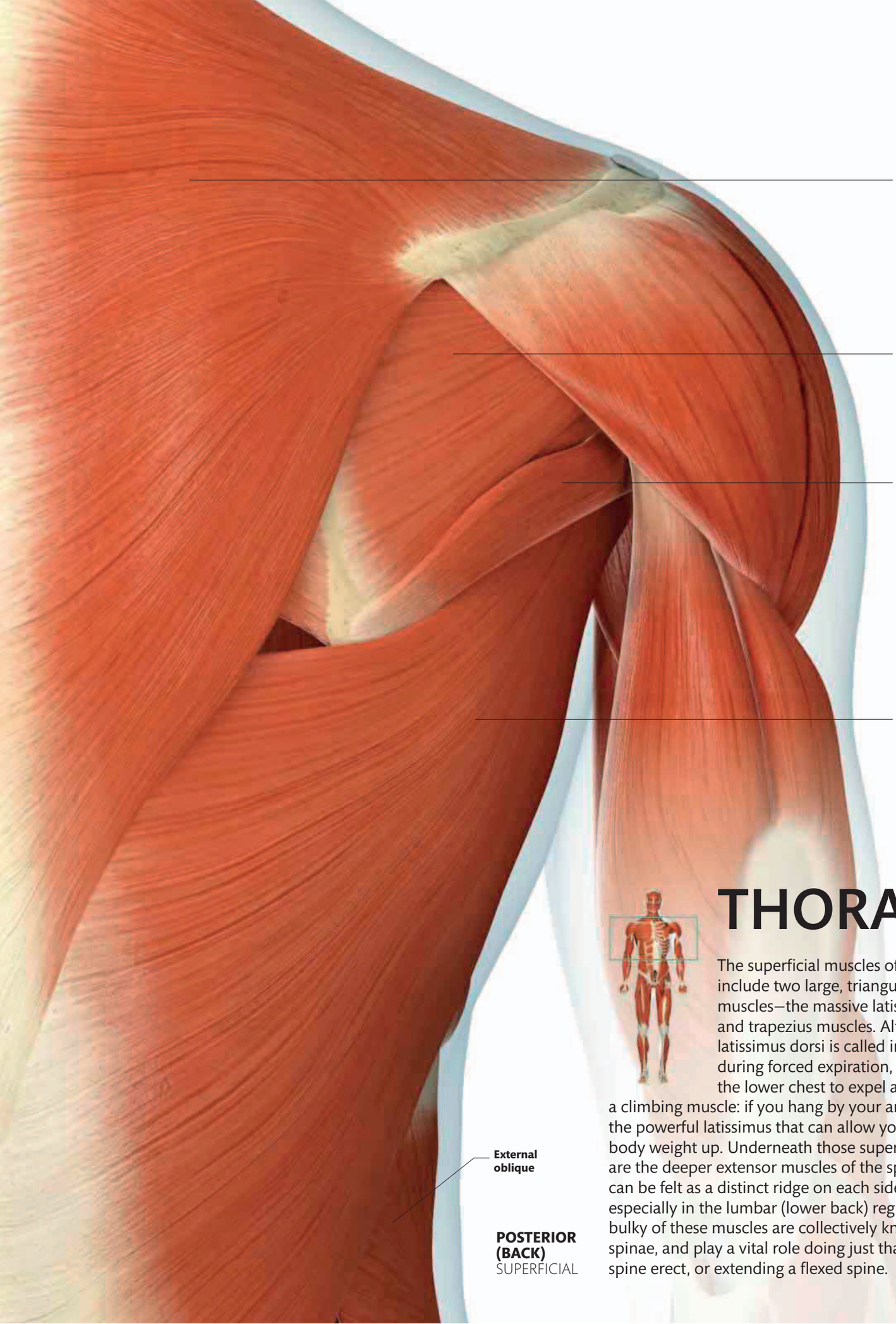

POSTERIOR (BACK)
SUPERFICIAL

THORAX

The superficial muscles of the back include two large, triangular-shaped muscles—the massive latissimus dorsi and trapezius muscles. Although latissimus dorsi is called into action during forced expiration, squeezing the lower chest to expel air, it is really a climbing muscle: if you hang by your arms, it is largely the powerful latissimus that can allow you to pull your body weight up. Underneath those superficial muscles are the deeper extensor muscles of the spine, which can be felt as a distinct ridge on each side of the spine, especially in the lumbar (lower back) region. The most bulky of these muscles are collectively known as erector spinae, and play a vital role doing just that—keeping the spine erect, or extending a flexed spine.

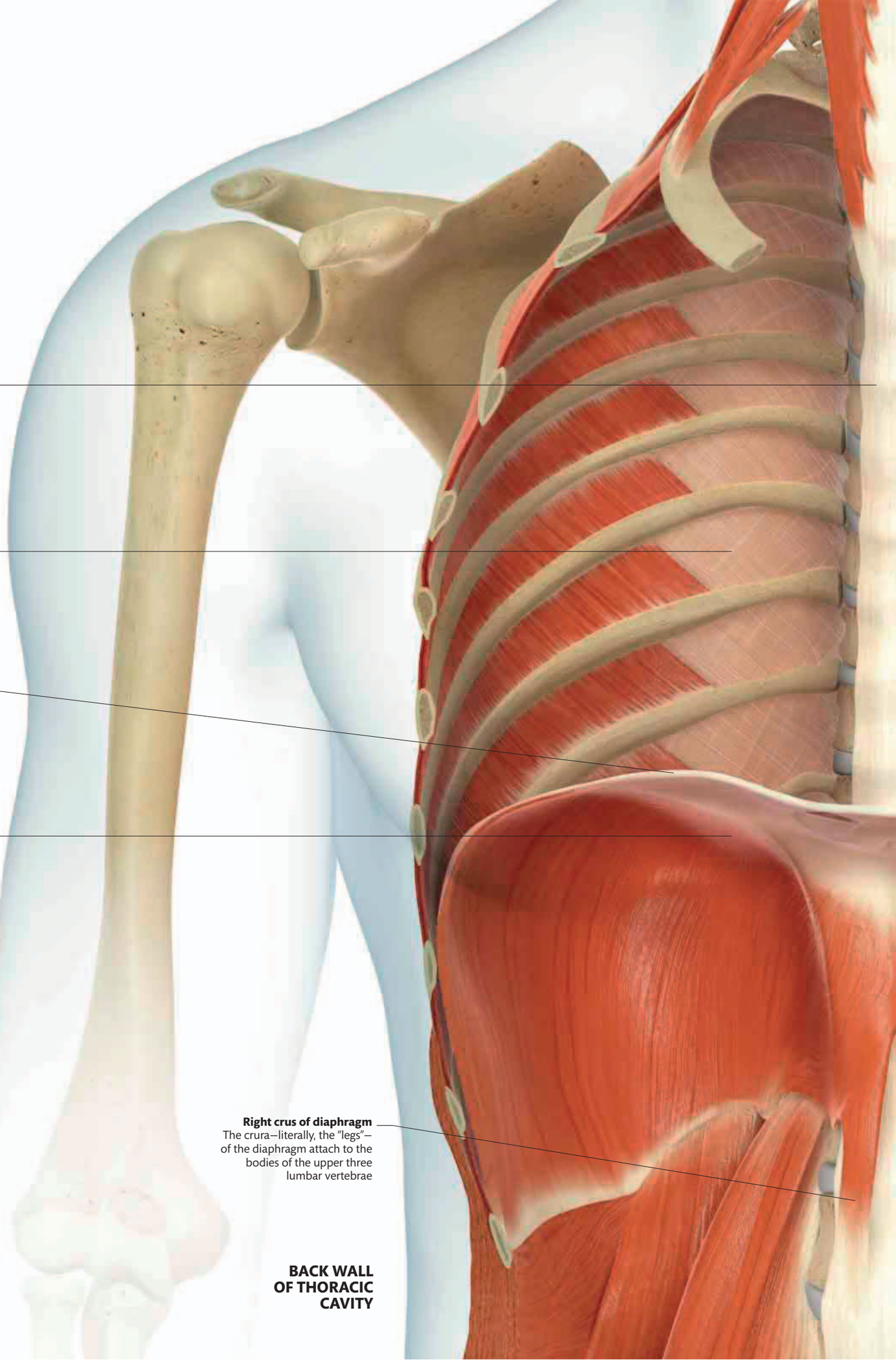

BACK WALL OF THORACIC CAVITY

Middle scalene

Anterior scalene

Longus colli

External intercostal muscle
These muscles are replaced by a membrane around the front of the thorax. (Seen here after removal of internal intercostal membrane)

Internal intercostal muscle
The intercostal muscles are supplied by intercostal nerves

Left crus of diaphragm

THORAX

The diaphragm, which divides the thorax and abdomen, is the main muscle of respiration. It attaches to the spine and to deep muscles in the back, around the margins of the rib cage, and to the sternum at the front. Its muscle fibers radiate out from a central, flat tendon to these attachments. The diaphragm contracts and flattens during inspiration, increasing the volume inside the chest cavity, and pulling air into the lungs; during expiration, it relaxes back into a domed shape. The intercostal muscles and diaphragm are "voluntary" muscle, and you can consciously control your breathing. But most of the time you don't have to think about breathing, since they work to a rhythm set by the brain stem, producing about 12 to 20 breaths per minute in an adult.

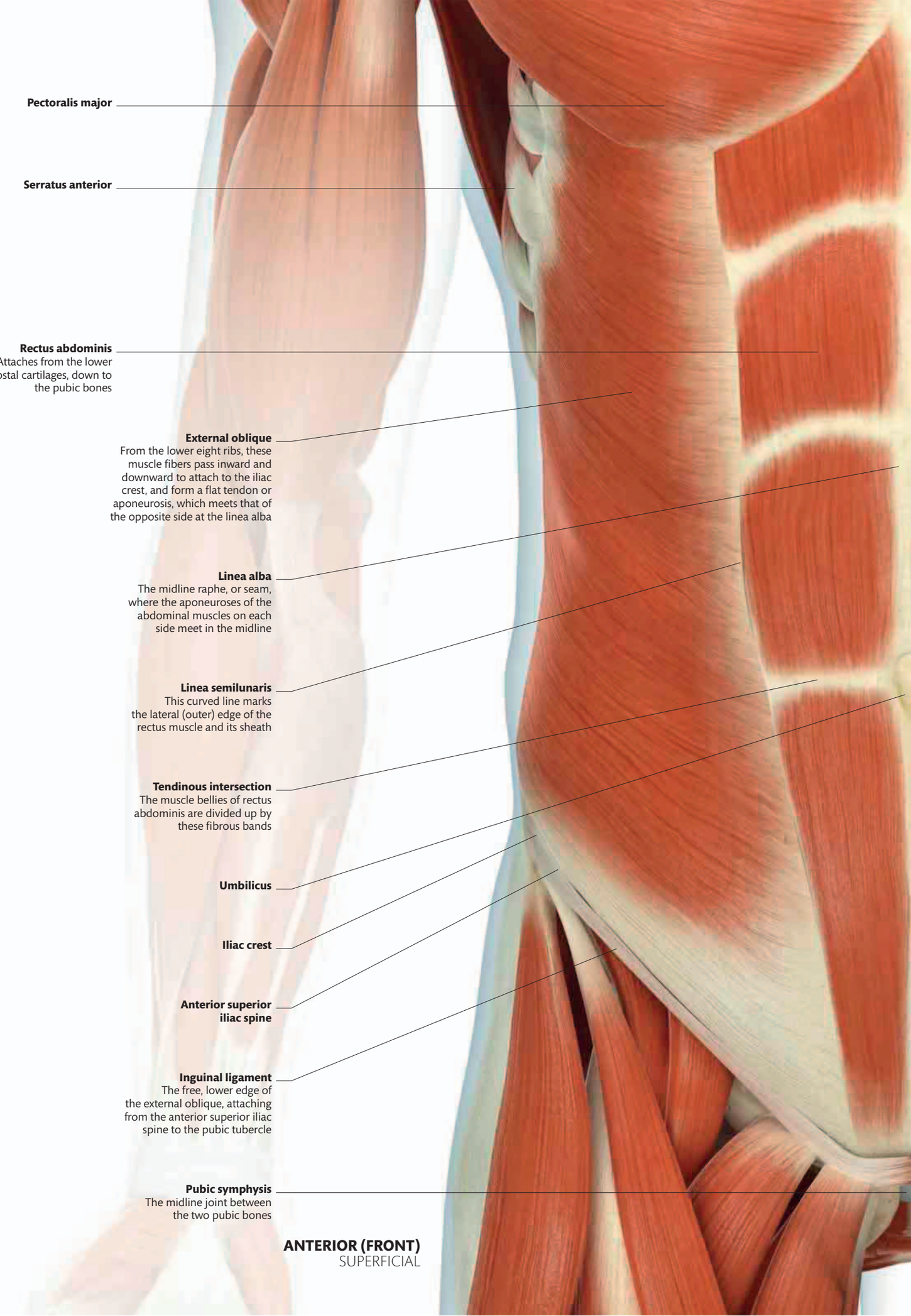

ANTERIOR (FRONT)
SUPERFICIAL

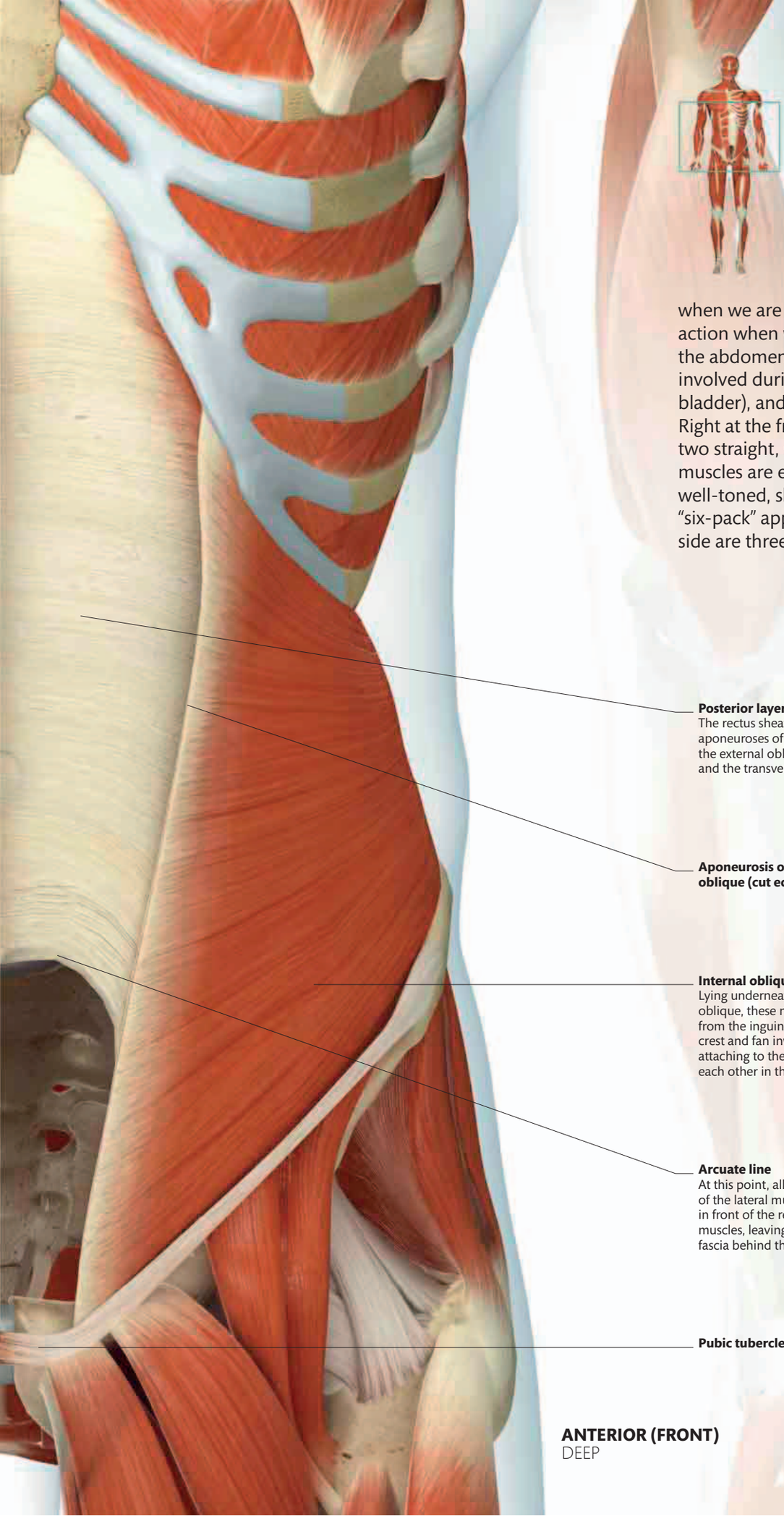

ABDOMEN AND PELVIS

The abdominal muscles can move the trunk —flexing the spine to the front or to the side, or twisting the abdomen from side to side. They are very important muscles in posture, helping support the upright spine when we are standing or sitting, and are also called into action when we lift heavy objects. Because they compress the abdomen and raise the pressure internally, they are involved during defecation, micturition (emptying the bladder), and in forced expiration of air from the lungs. Right at the front, lying either side of the midline, there are two straight, straplike rectus abdominis muscles. These muscles are each broken up by horizontal tendons: in a well-toned, slim person, this creates the much-sought-after "six-pack" appearance. Flanking the recti muscles on each side are three layers of broad, flat muscles.

Posterior layer of rectus sheath
The rectus sheath is formed by the aponeuroses of the muscles to the sides: the external oblique, the internal oblique, and the transversus abdominis

Aponeurosis of internal oblique (cut edge)

Internal oblique
Lying underneath the external oblique, these muscle fibers spring from the inguinal ligament and iliac crest and fan inward and upward, attaching to the lower ribs and to each other in the midline

Arcuate line
At this point, all the aponeuroses of the lateral muscles swap to lie in front of the rectus abdominis muscles, leaving only a layer of fascia behind that muscle

Pubic tubercle

ANTERIOR (FRONT)
DEEP

ABDOMEN AND PELVIS

The most superficial muscle of the lower back is the incredibly broad latissimus dorsi. Underneath this, lying along the spine on each side, there is a large bulk of muscle that forms two ridges in the lumbar region in a well-toned person. This muscle mass is collectively known as the erector spinae, and its name suggests its importance in keeping the spine upright. When the spine is flexed forward, the erector spinae can pull it back into an upright position, and even take it further, into extension. The muscle can be divided up into three main strips on each side: iliocostalis, longissimus, and spinalis. Most of the muscle bulk of the buttock comes down to just one muscle: the fleshy gluteus maximus, which extends the hip joint. Hidden beneath the gluteus maximus are a range of smaller muscles that also move the hip.

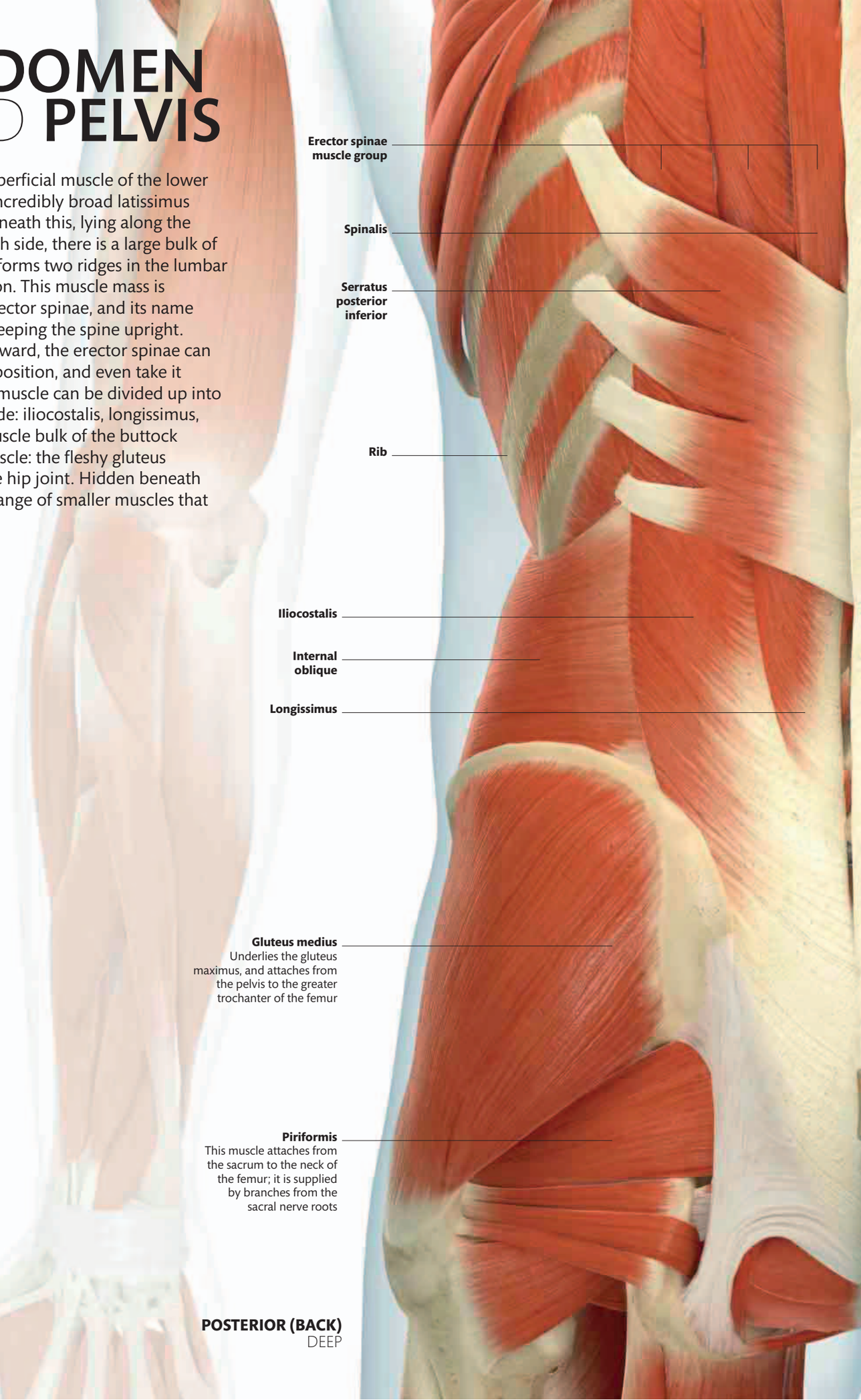

POSTERIOR (BACK)
DEEP

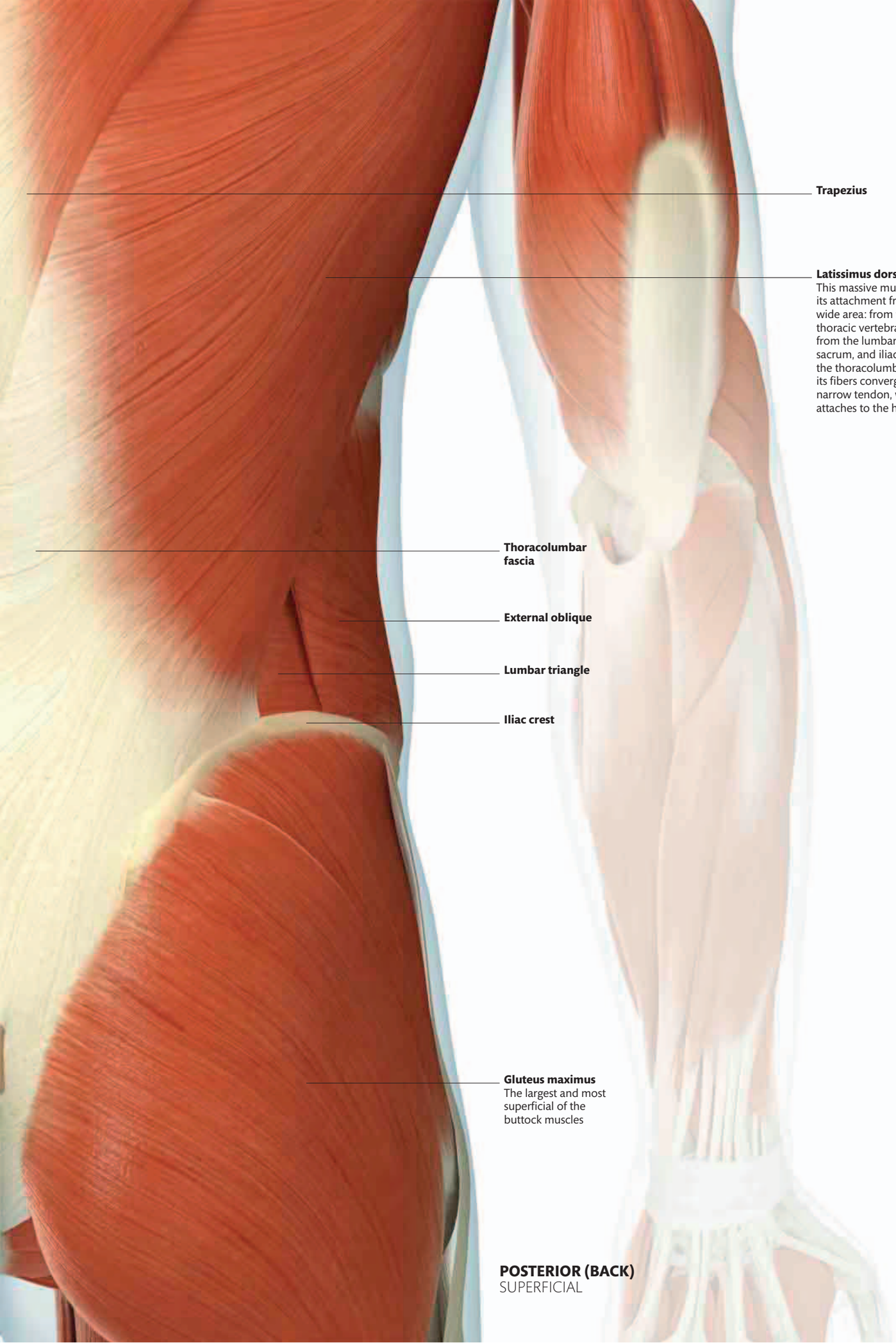

POSTERIOR (BACK)
SUPERFICIAL

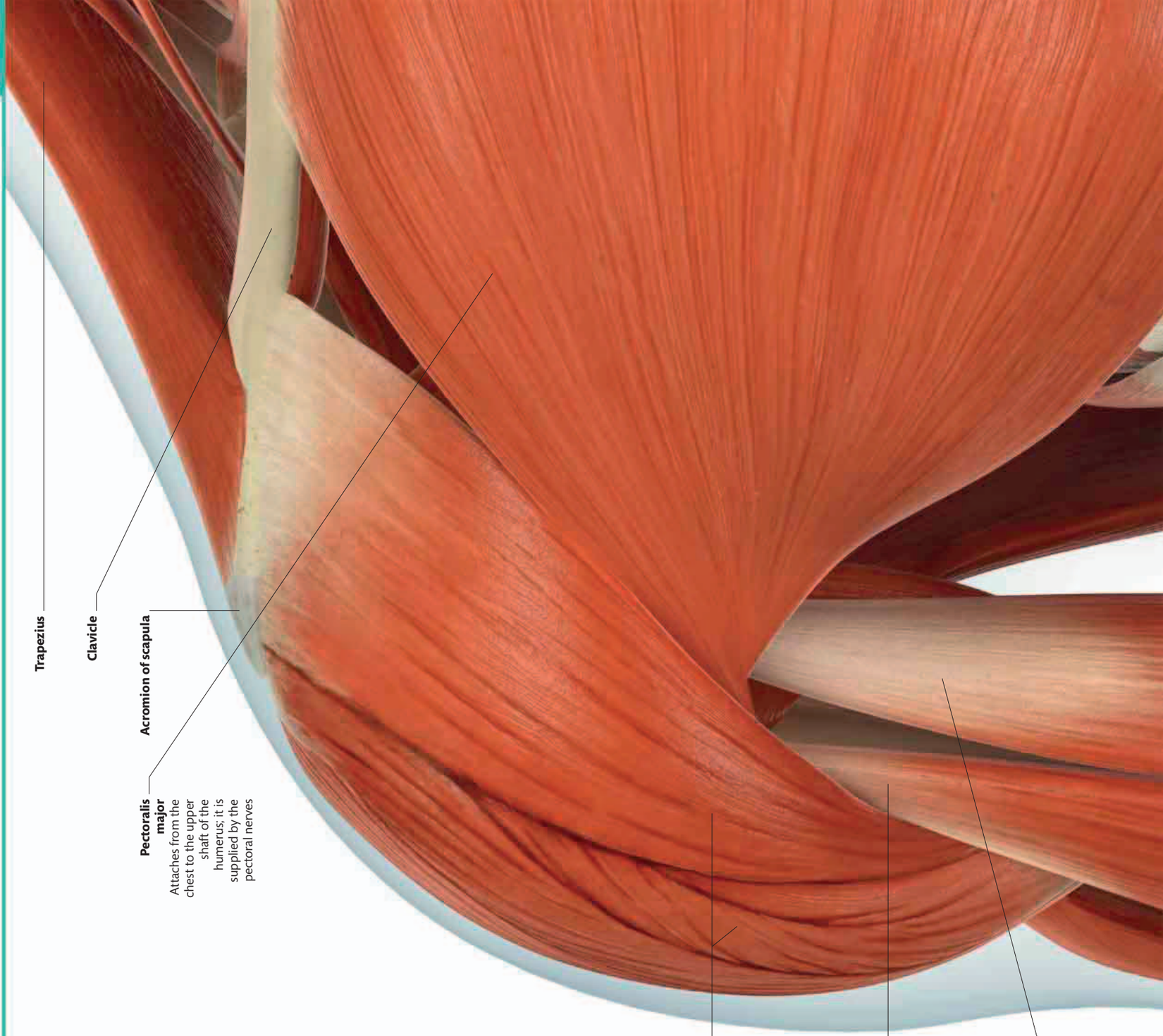

SHOULDER AND UPPER ARM

The triangular deltoid muscle lies over the shoulder. Acting as a whole, this muscle raises the arm to the side (abduction), but the fibers of the deltoid attaching to the front of the clavicle can also move the arm forward. The pectoralis major muscle can also act on the shoulder joint, flexing the arm forward or pulling it in to the side of the chest (adduction). The biceps brachii muscle forms much of the muscle bulk on the front of the arm. The biceps tendon inserts on the radius, and also has an aponeurosis (flat tendon) that fans out over the forearm muscles. The biceps is a powerful flexor of the elbow, and can also rotate the radius to position the lower arm so the palm faces upward (supination).

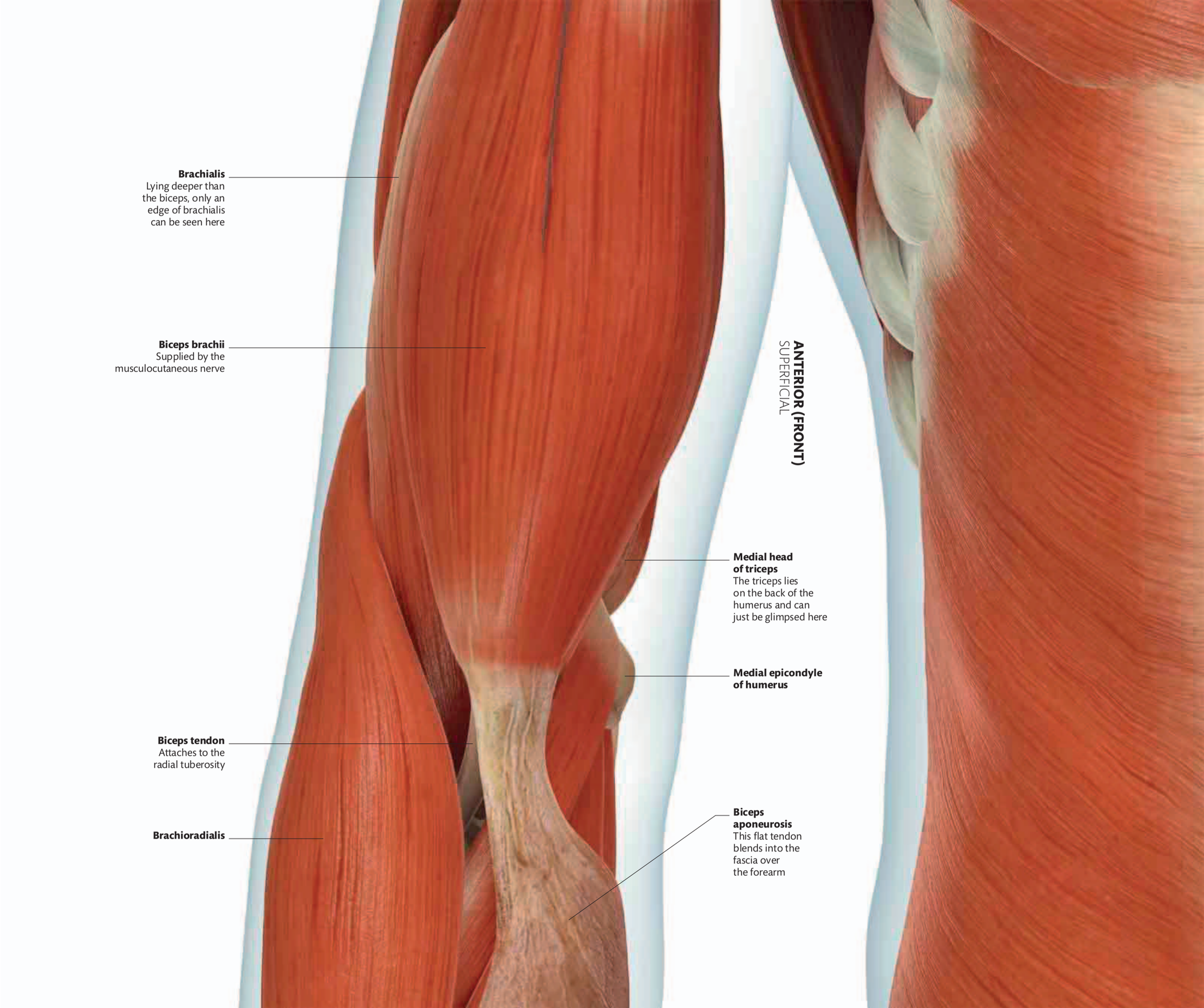
ANTERIOR (FRONT)
SUPERFICIAL
Brachialis
Lying deeper than the biceps, only an edge of brachialis can be seen here
Biceps brachii
Supplied by the musculocutaneous nerve
Biceps tendon
Attaches to the radial tuberosity
Brachioradialis
Medial head of triceps
The triceps lies on the back of the humerus and can just be glimpsed here
Medial epicondyle of humerus
Biceps aponeurosis
This flat tendon blends into the fascia over the forearm

SHOULDER AND UPPER ARM

The posterior fibers of the deltoid attach from the spine of the scapula (shoulder blade) down to the humerus, and this part of the muscle can draw back the arm, or extend it. Latissimus dorsi (a broad muscle attaching from the back of the trunk and ending in a narrow tendon that secures onto the humerus) can also extend the arm. The triceps brachii muscle is the sole extensor of the elbow. In a superficial dissection (represented in this view) only two of the three heads of the triceps can be seen—the long and lateral heads. The triceps tendon attaches to the leverlike olecranon of the ulna, which forms the bony knobble at the back of the elbow.

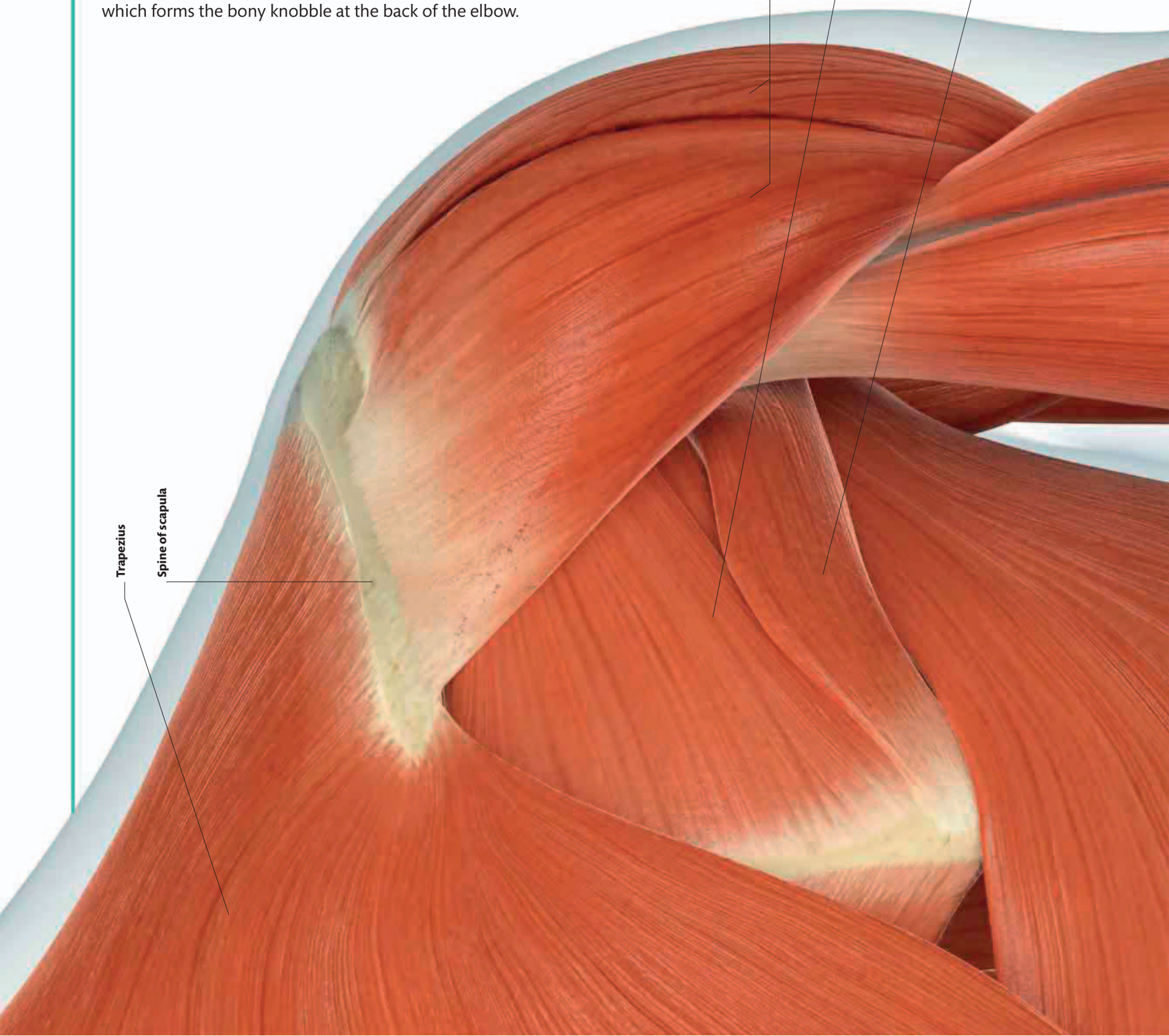

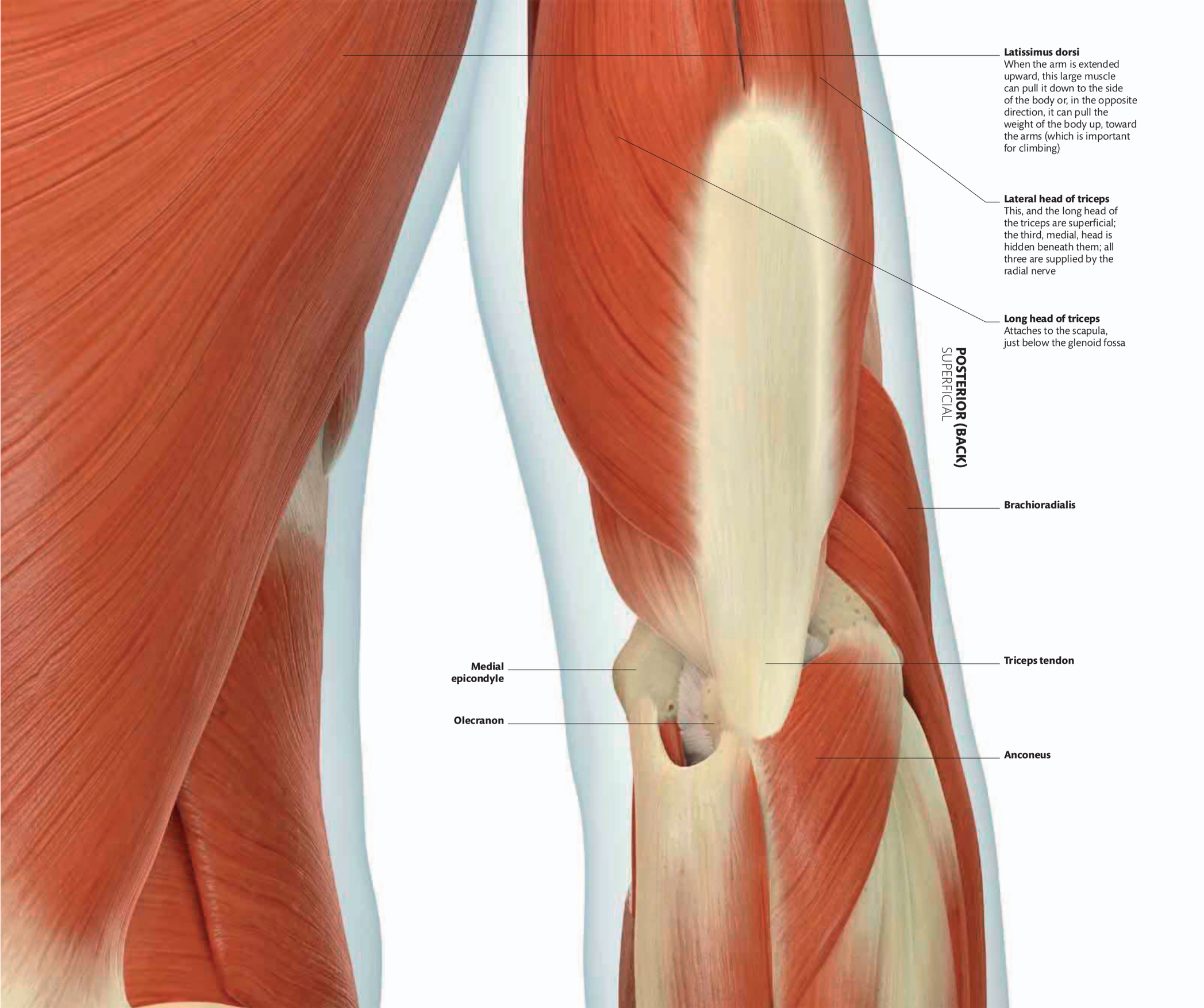
Latissimus dorsi
When the arm is extended upward, this large muscle can pull it down to the side of the body or, in the opposite direction, it can pull the weight of the body up, toward the arms (which is important for climbing)
Lateral head of triceps
This, and the long head of the triceps are superficial; the third, medial, head is hidden beneath them; all three are supplied by the radial nerve
Long head of triceps
Attaches to the scapula, just below the glenoid fossa
POSTERIOR (BACK)
SUPERFICIAL
Brachioradialis
Triceps tendon
Medial epicondyle
Olecranon
Anconeus

SHOULDER AND UPPER ARM

The deep muscles around the shoulder include the so-called rotator cuff group, two of which can be seen here: the subscapularis (which attaches from the deep surface of the scapula) and the supraspinatus (which runs from the scapula, over the shoulder joint, to attach to the humerus). The supraspinatus tendon passes through a narrow gap between the head of the humerus and the acromion of the scapula, and may become compressed and damaged here, in impingement syndrome. On the front of the humerus, the biceps (see p.85) has been removed to reveal brachialis, which runs from the lower humerus down to the ulna. Like the biceps, brachialis is a flexor of the elbow.

ANTERIOR (FRONT)
DEEP
Medial head of triceps
Brachialis
The name of this muscle simply means of the arm in Latin. It attaches from the front of the humerus to the tuberosity on the front of the ulna, and it flexes the elbow. It lies underneath the biceps
Medial epicondyle of humerus
Brachioradialis
Supinator

SHOULDER AND UPPER ARM

More of the rotator cuff muscles—the supraspinatus, infraspinatus, and teres minor—can be seen from the back. In addition to moving the shoulder joint in various directions, including rotation, these muscles are important in helping to stabilize the shoulder joint: they hug the head of the humerus into its socket during movements at the shoulder. On the back of the arm, a deeper view reveals the third, medial head of the triceps, which attaches from the back of the humerus. It joins with the lateral and long heads to form the triceps tendon, attaching to the olecranon. Most of the forearm muscles take their attachment from the epicondyles of the humerus, just above the elbow, but the brachioradialis and extensor carpi radialis longus have higher origins from the side of the humerus, as shown here.

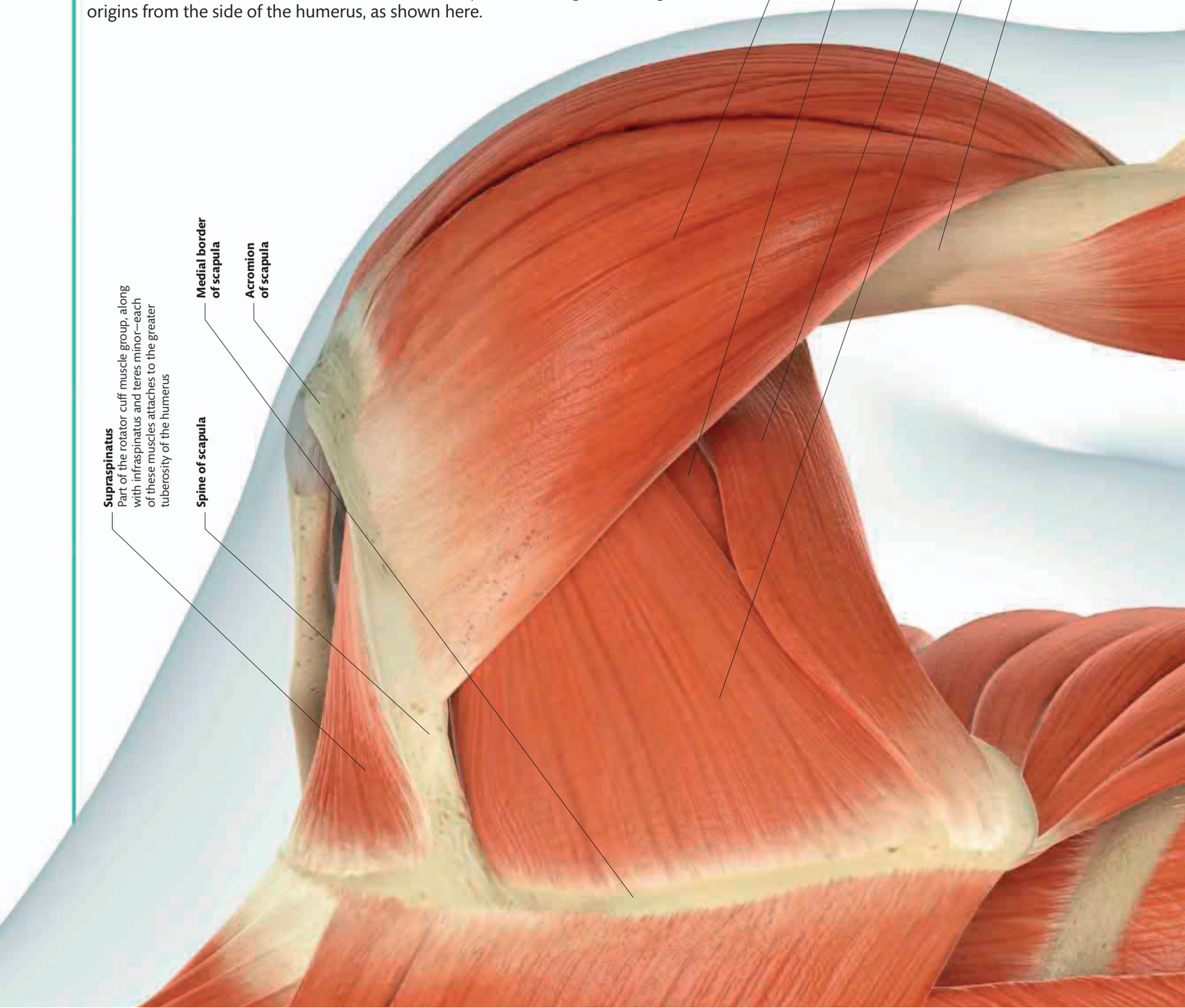

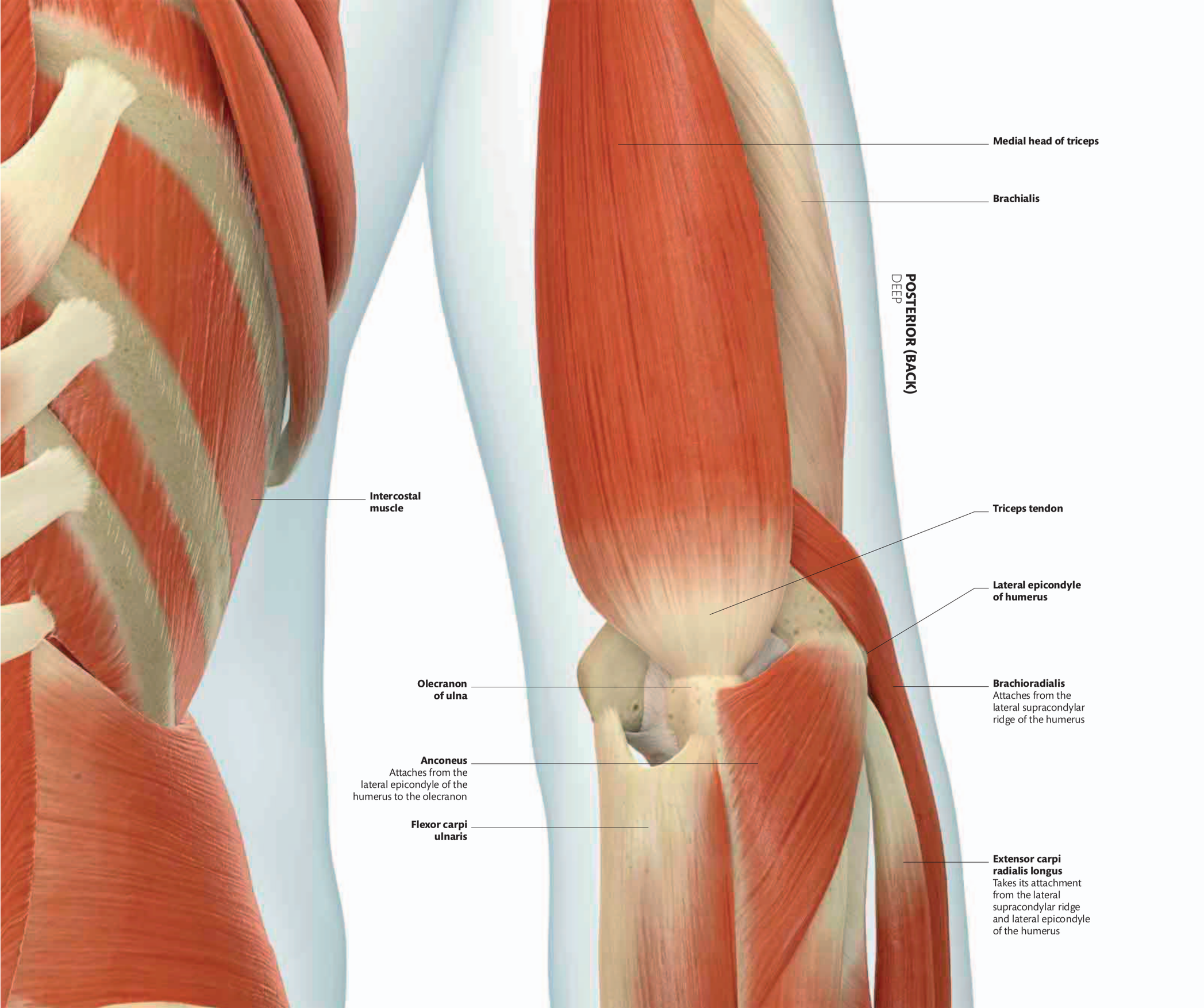
POSTERIOR (BACK)
DEEP
Medial head of triceps
Brachialis
Intercostal muscle
Triceps tendon
Lateral epicondyle of humerus
Olecranon of ulna
Brachioradialis
Attaches from the lateral supracondylar ridge of the humerus
Anconeus
Attaches from the lateral epicondyle of the humerus to the olecranon
Flexor carpi ulnaris
Extensor carpi radialis longus
Takes its attachment from the lateral supracondylar ridge and lateral epicondyle of the humerus

ANTERIOR (FRONT)
SUPERFICIAL

POSTERIOR (BACK)
SUPERFICIAL

LOWER ARM AND HAND

There are five superficial muscles on the front of the forearm, all taking their attachment from the medial epicondyle of the humerus. Pronator teres attaches across to the radius, and can pull this bone into pronation (held with the palm turned downward). The other muscles run farther down the forearm, becoming slender tendons that attach around the wrist, or continue into the hand. Flexor digitorum superficialis splits into four tendons, one for each finger. On the back of the forearm, seven superficial extensor muscles attach to the lateral epicondyle of the humerus. Most of these tendons run down to the wrist or into the hand.

Brachialis
Medial epicondyle of humerus
Also known as the common flexor origin
ANTERIOR (FRONT)
DEEP
Flexor carpi ulnaris
Flexor retinaculum
Supinator
Brachioradialis
Flexor pollicis longus
This long flexor of the thumb arises from the radius and interosseous membrane; its tendon runs into the thumb to attach to the base of the distal phalanx
POSTERIOR (BACK)
DEEP
Dorsal interosseous muscles
These muscles spread the fingers
Extensor retinaculum
Extensor pollicis brevis
Extensor of the index finger; it joins the tendon of extensor digitorum (see pp.92–93) of the index finger
Extensor pollicis brevis
Short extensor of the thumb; it attaches to the proximal phalanx and pulls the thumb out to the side

LOWER ARM AND HAND

Stripping away the superficial muscles on the front of the forearm reveals a deeper layer attaching to the radius and ulna, and to the interosseous membrane between the bones. The long, quill-like flexor of the thumb (flexor pollicis longus) can be seen clearly. Deep muscles on the back of the forearm include the long extensors of the thumb and index finger and the supinator, which pulls on the radius to rotate the pronated arm (held with palm facing downward) into supination (with palm facing up). In the hand, a deep dissection reveals the interosseous muscles that act on the metacarpophalangeal joints in order to either spread or close the fingers.

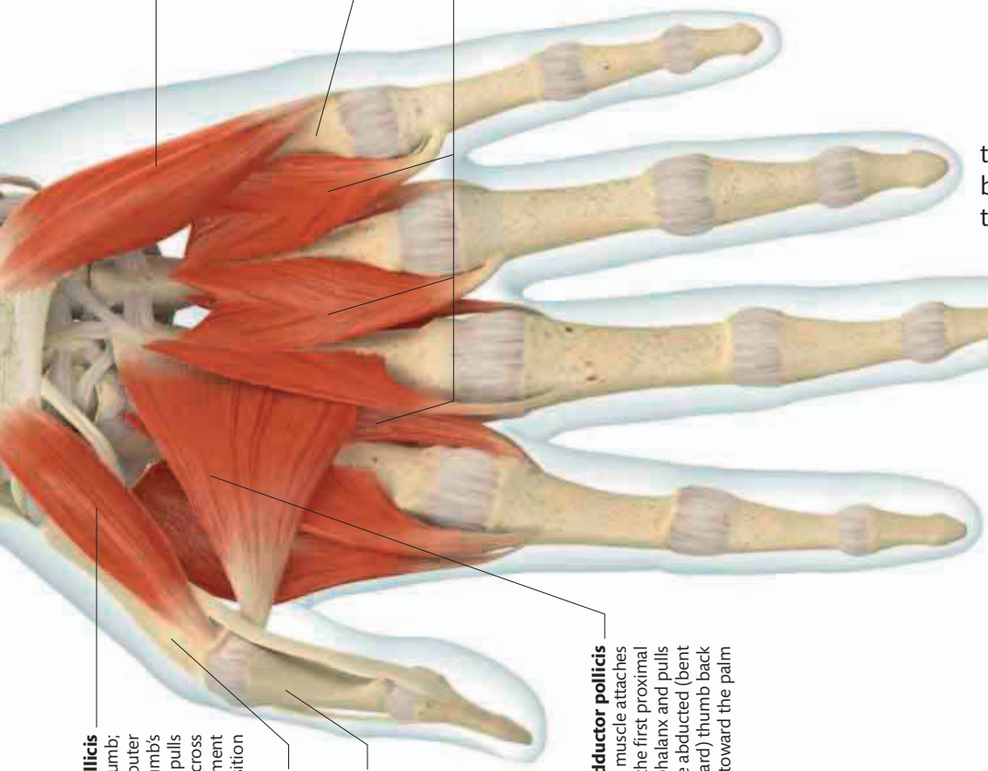

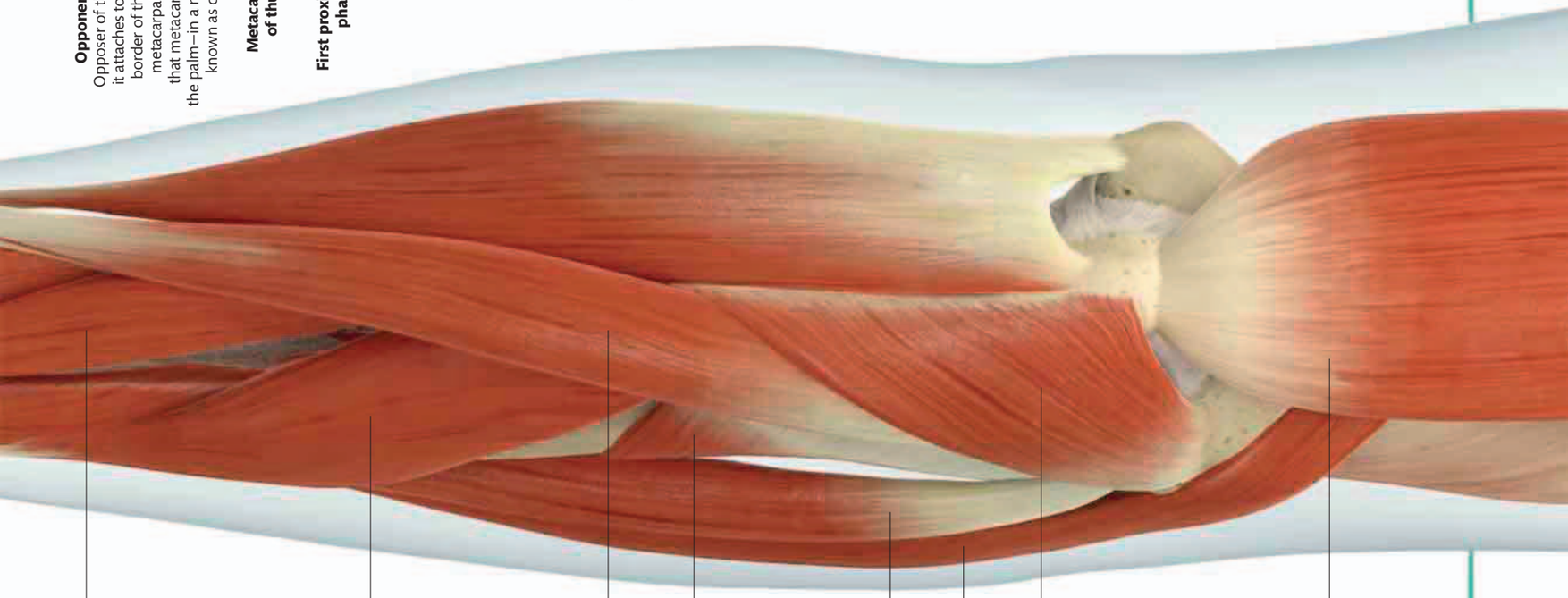

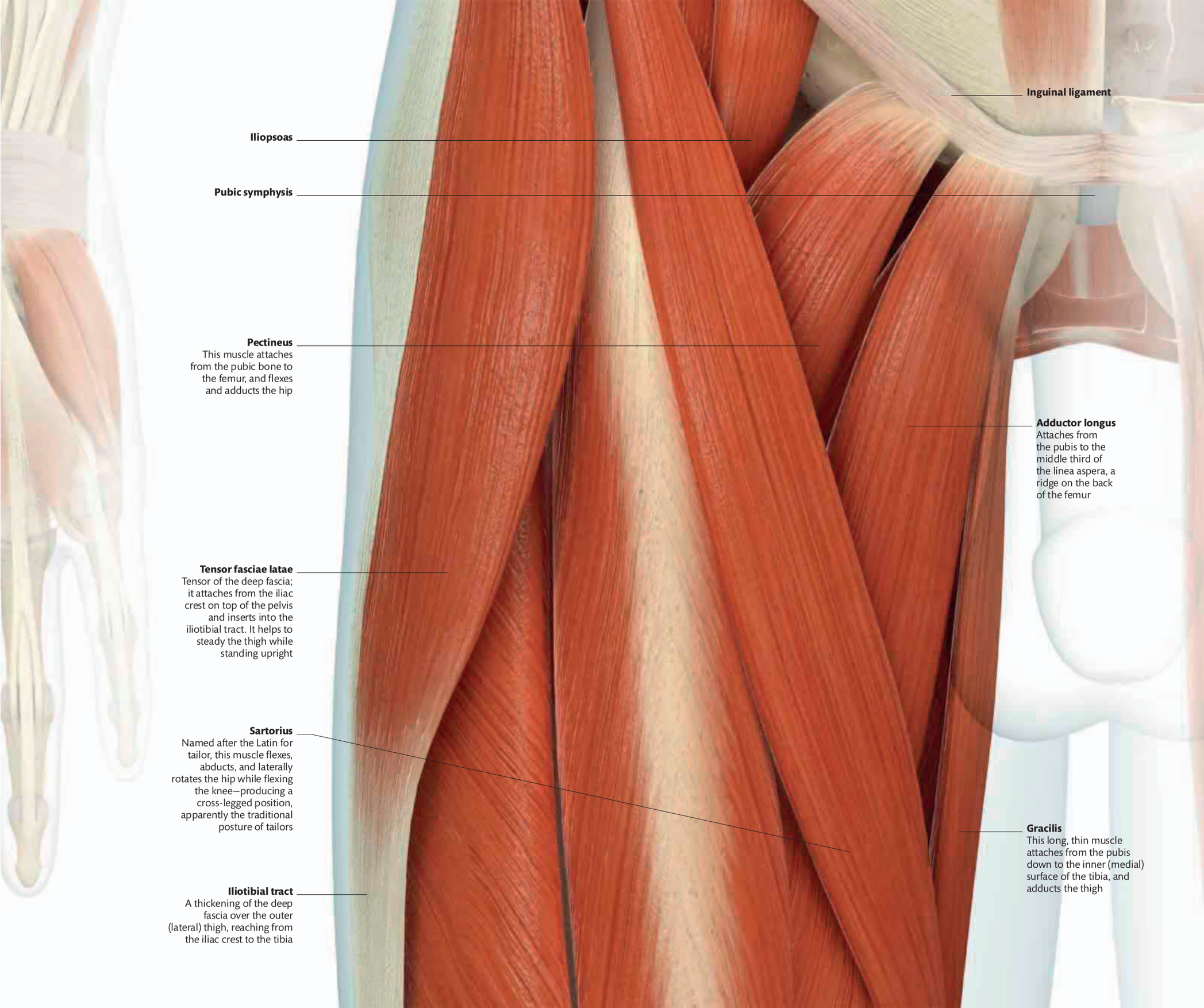
Inguinal ligament
Iliopsoas
Pubic symphysis
Pectineus
This muscle attaches from the pubic bone to the femur, and flexes and adducts the hip
Adductor longus
Attaches from the pubis to the middle third of the linea aspera, a ridge on the back of the femur
Tensor fasciae latae
Tensor of the deep fascia; it attaches from the iliac crest on top of the pelvis and inserts into the iliotibial tract. It helps to steady the thigh while standing upright
Sartorius
Named after the Latin for tailor, this muscle flexes, abducts, and laterally rotates the hip while flexing the knee—producing a cross-legged position, apparently the traditional posture of tailors
Gracilis
This long, thin muscle attaches from the pubis down to the inner (medial) surface of the tibia, and adducts the thigh
Iliotibial tract
A thickening of the deep fascia over the outer (lateral) thigh, reaching from the iliac crest to the tibia

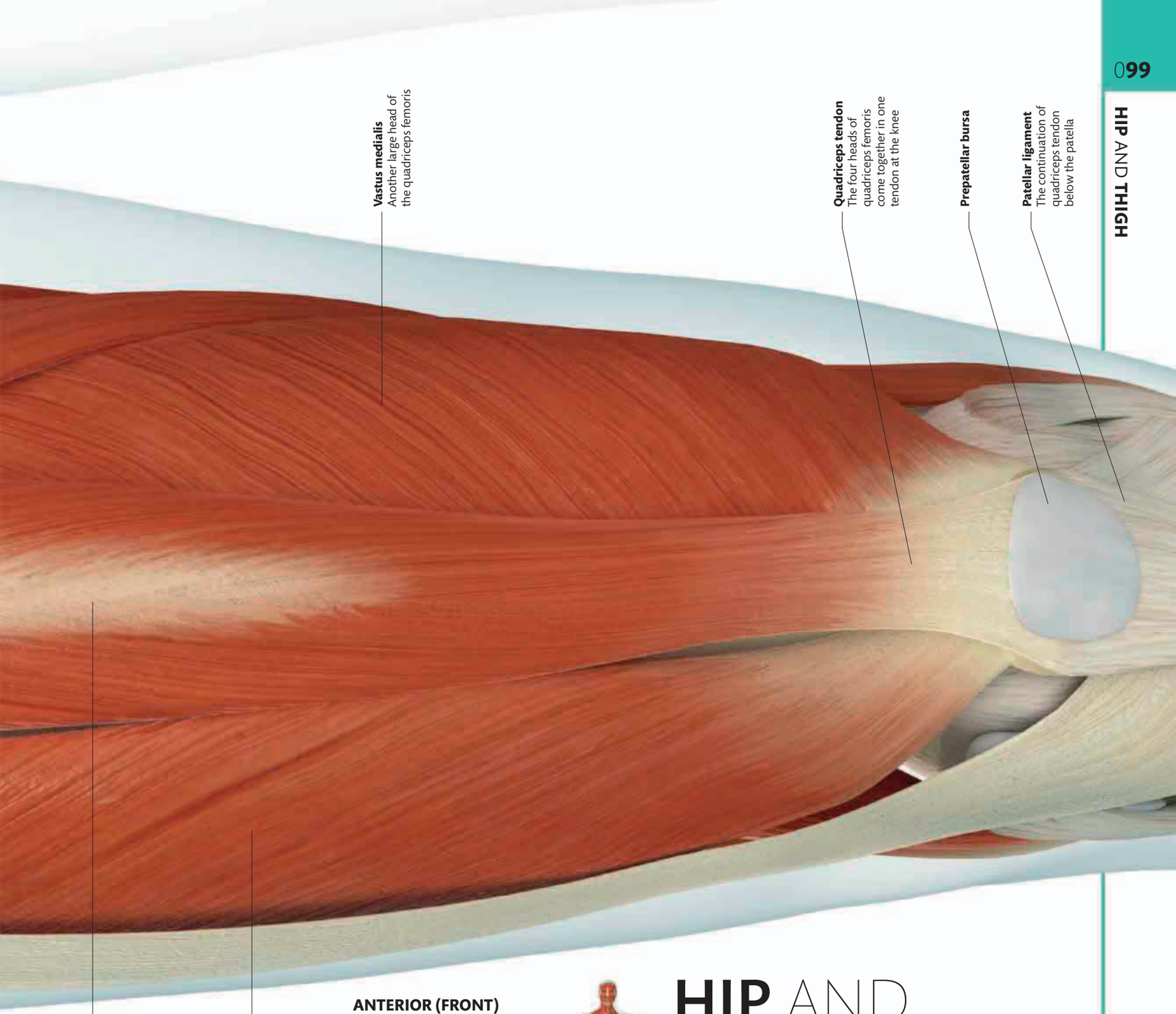

ANTERIOR (FRONT)
SUPERFICIAL

HIP AND THIGH

Most of the muscle bulk on the front of the leg is the four-headed quadriceps femoris. Three of its heads can be seen in a superficial dissection of the thigh: the rectus femoris, vastus lateralis, and vastus medialis. The quadriceps extends the knee, but it can also flex the hip, since the rectus femoris part has an attachment from the pelvis, above the hip joint. The patella is embedded in the quadriceps tendon; this may protect the tendon from wear and tear, but it also helps to give the quadriceps good leverage in extending the knee. The part of the tendon below the patella is usually called the patellar ligament. Tapping this with a tendon hammer produces a reflex contraction in the quadriceps—the "knee jerk".

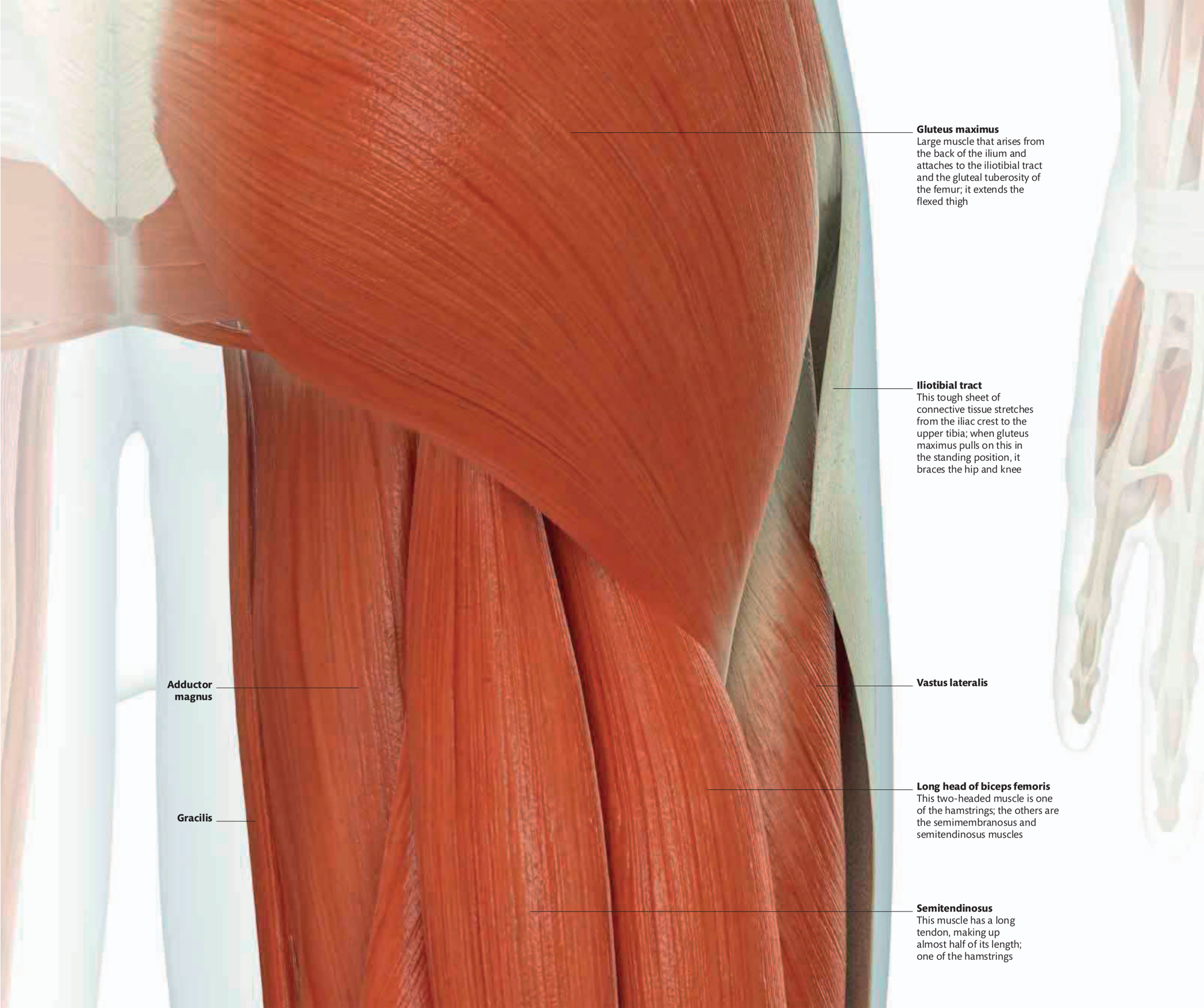
Gluteus maximus
Large muscle that arises from the back of the ilium and attaches to the iliotibial tract and the gluteal tuberosity of the femur; it extends the flexed thigh
Iliotibial tract
This tough sheet of connective tissue stretches from the iliac crest to the upper tibia; when gluteus maximus pulls on this in the standing position, it braces the hip and knee
Adductor magnus
Vastus lateralis
Long head of biceps femoris
This two-headed muscle is one of the hamstrings; the others are the semimembranosus and semitendinosus muscles
Gracilis
Semitendinosus
This muscle has a long tendon, making up almost half of its length; one of the hamstrings

HIP AND THIGH

On the back of the hip and thigh, a superficial dissection reveals the large gluteus maximus, an extensor of the hip joint, and the three hamstrings. The gluteus maximus acts to extend the hip joint, swinging the leg backward. While it doesn't really contribute to gentle walking, it is very important in running, and also when the hip is being extended from a flexed position, such as when getting up from sitting on the floor or when climbing the stairs. The hamstrings—the semimembranosus, semitendinosus, and biceps femoris muscles—attach from the ischial tuberosity of the pelvis and sweep down the back of the thigh to the tibia and fibula. They are the main flexors of the knee.

POSTERIOR (BACK)
SUPERFICIAL

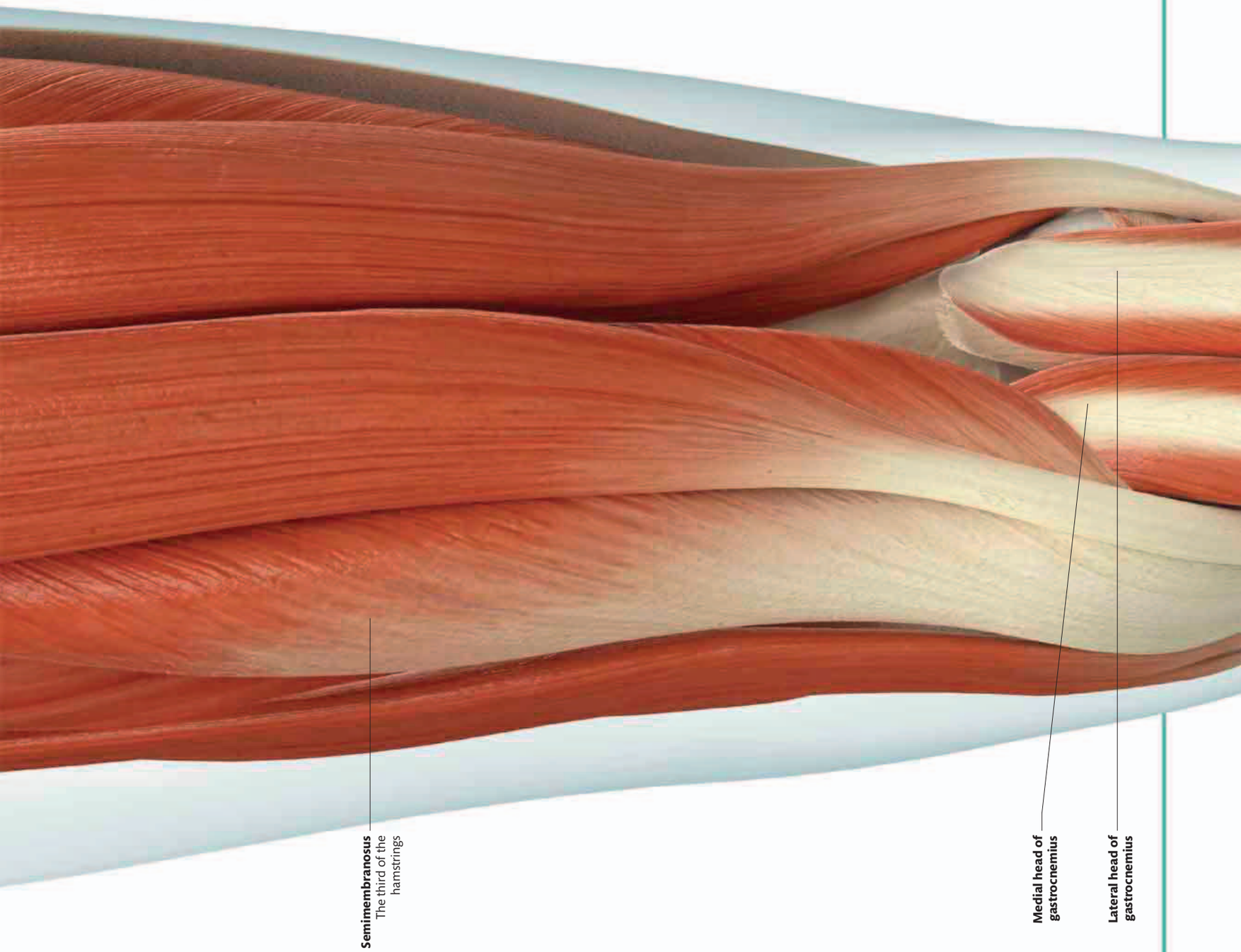

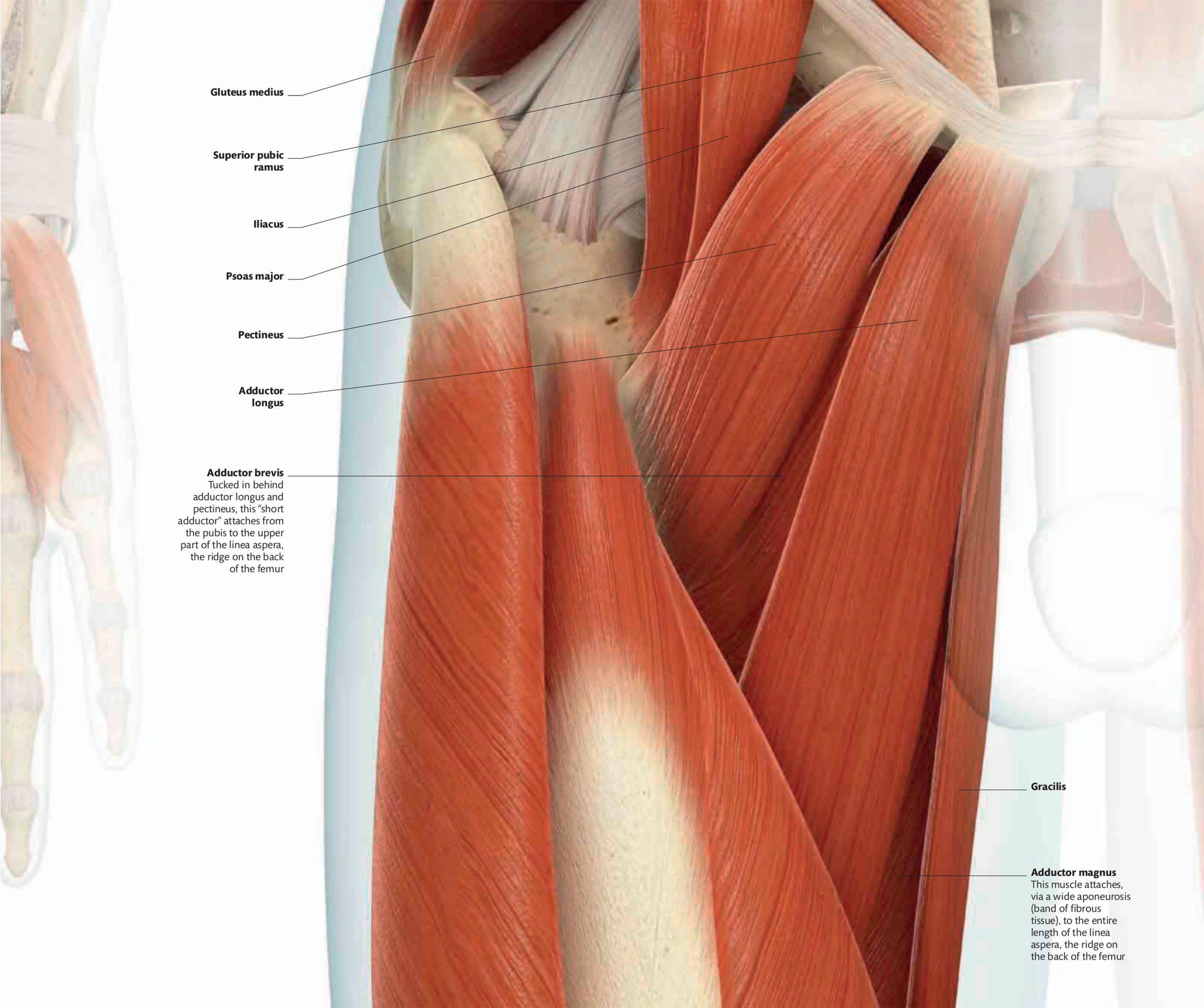
Gluteus medius
Superior pubic ramus
Iliacus
Psoas major
Pectineus
Adductor longus
Adductor brevis
Tucked in behind adductor longus and pectineus, this "short adductor" attaches from the pubis to the upper part of the linea aspera, the ridge on the back of the femur
Gracilis
Adductor magnus
This muscle attaches, via a wide aponeurosis (band of fibrous tissue), to the entire length of the linea aspera, the ridge on the back of the femur

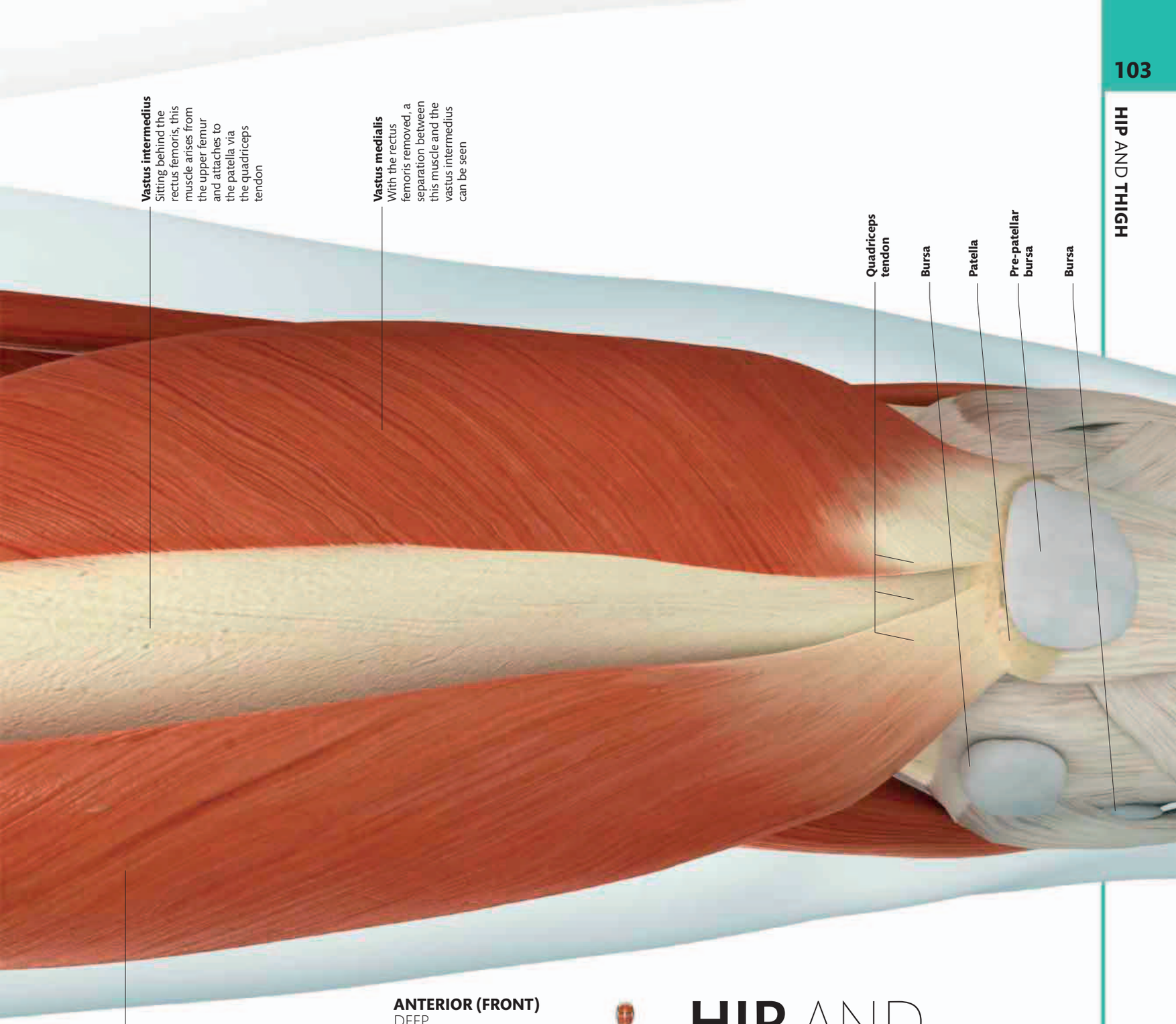

ANTERIOR (FRONT)
DEEP

HIP AND THIGH

With the rectus femoris and sartorius muscles stripped away, we can see the deep, fourth head of the quadriceps, known as vastus intermedius. The adductor muscles that bring the thighs together can also be seen clearly, including the gracilis, which is long and slender, as its name suggests. The largest adductor muscle—the adductor magnus—has a hole in its tendon, through which the main artery of the leg (the femoral artery) passes. The adductor tendons attach from the pubis and ischium of the pelvis, and the sporting injuries referred to as "groin pulls" are often tears in these particular tendons.

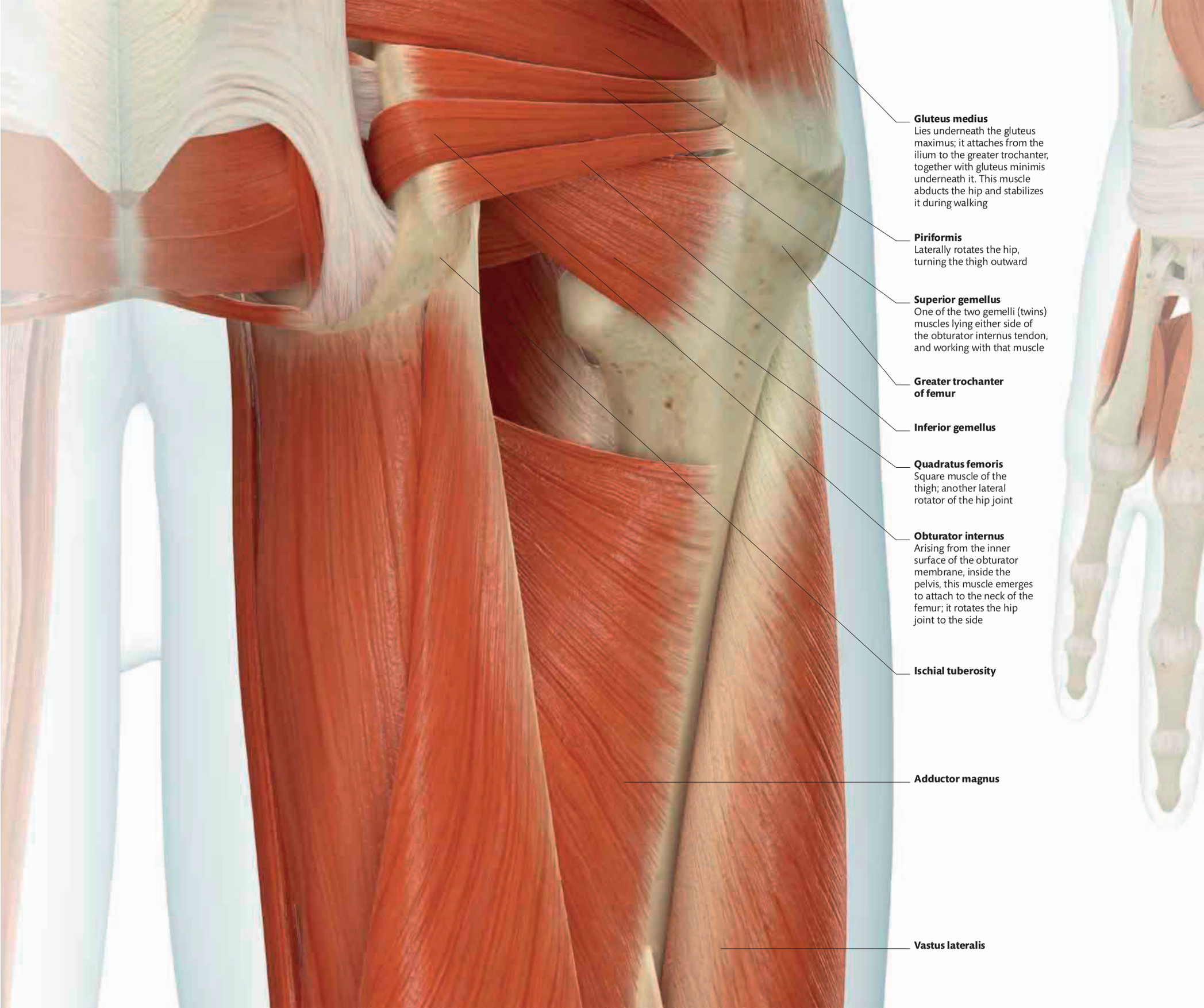
Gluteus medius
Lies underneath the gluteus maximus; it attaches from the ilium to the greater trochanter, together with gluteus minimis underneath it. This muscle abducts the hip and stabilizes it during walking
Piriformis
Laterally rotates the hip, turning the thigh outward
Superior gemellus
One of the two gemelli (twins) muscles lying either side of the obturator internus tendon, and working with that muscle
Greater trochanter of femur
Inferior gemellus
Quadratus femoris
Square muscle of the thigh; another lateral rotator of the hip joint
Obturator internus
Arising from the inner surface of the obturator membrane, inside the pelvis, this muscle emerges to attach to the neck of the femur; it rotates the hip joint to the side
Ischial tuberosity
Adductor magnus
Vastus lateralis

HIP AND THIGH

On the back of the hip, with the gluteus maximus removed, the short muscles that rotate the hip out to the side are clearly revealed. These include the piriformis, obturator internus, and quadratus femoris muscles. With the long head of the biceps femoris removed, we can now see the deeper, short head attaching to the linea aspera on the back of the femur. The semitendinosus muscle has also been cut away to reveal the semimembranosus underneath it, with its flat, membranelike tendon at the top. Popliteus muscle is also visible at the back of the knee joint, as is one of the many fluid filled bursae around the knee.

POSTERIOR (BACK)
DEEP

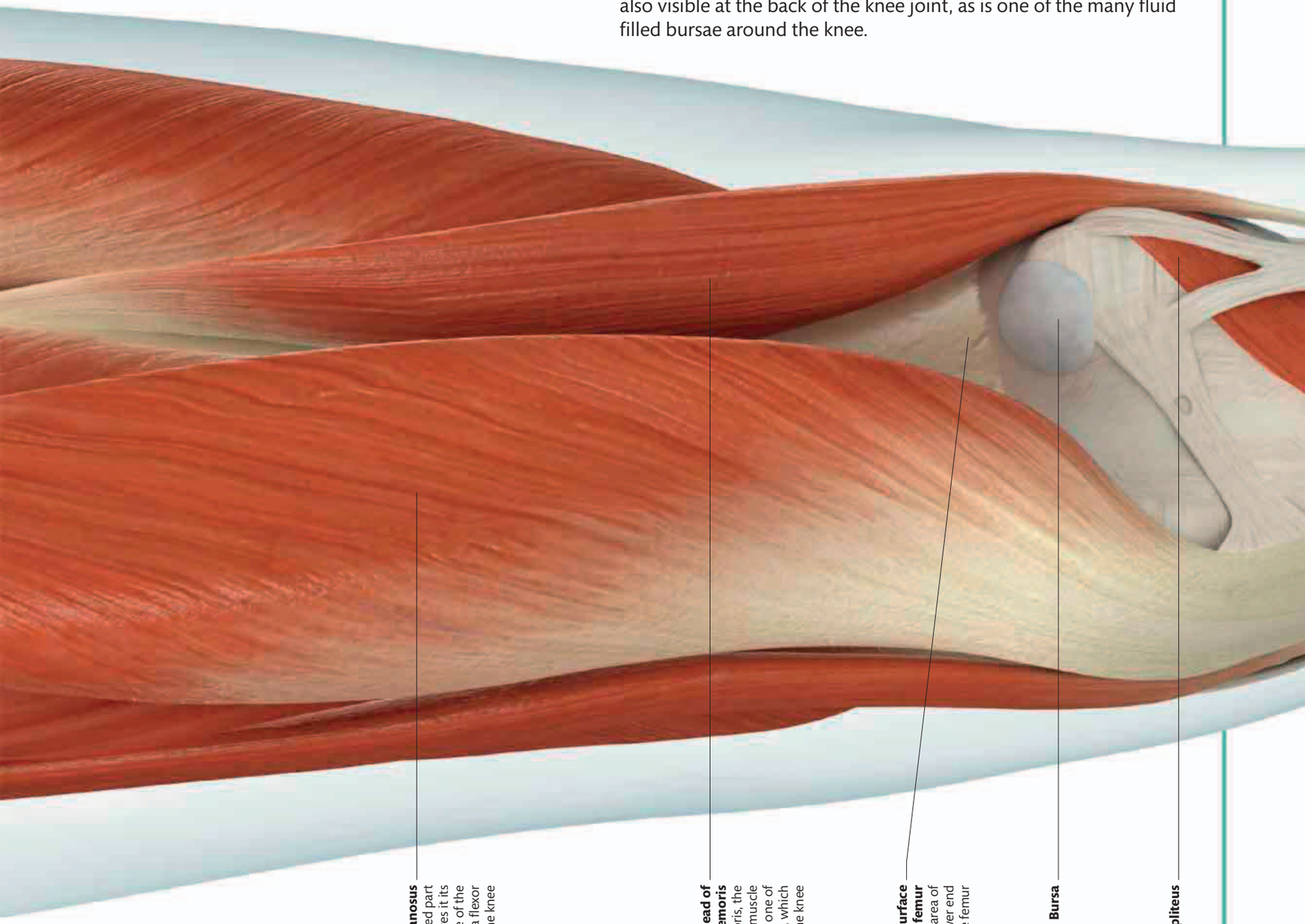

POSTERIOR (BACK)
SUPERFICIAL

ANTERIOR (FRONT)
SUPERFICIAL

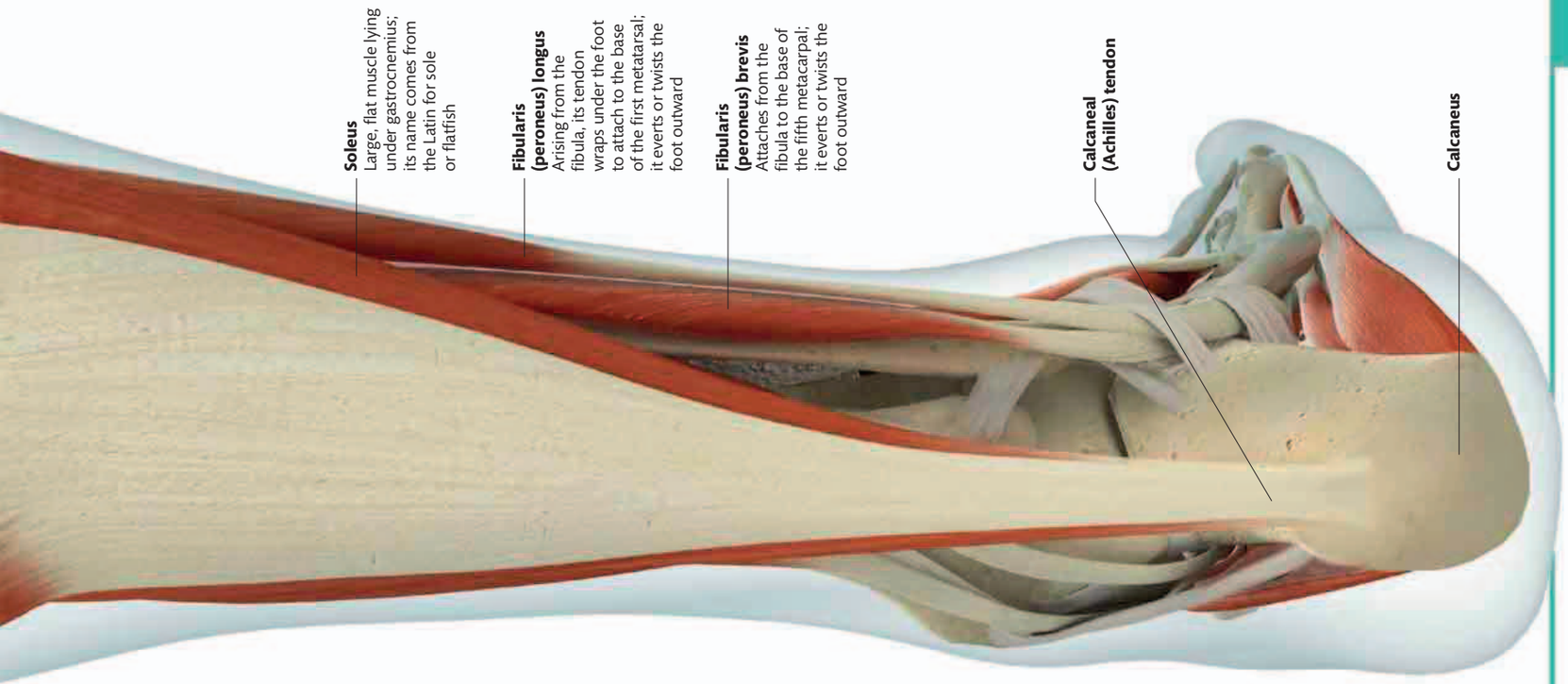

LOWER LEG AND FOOT

You can feel the medial surface of the tibia easily, just under the skin on the front of your lower leg, on the inner side. Move your fingers outward, and you feel the sharp border of the bone, and then a soft wedge of muscles alongside it. These muscles have tendons that run down to the foot. They can pull the foot upward at the ankle, in a movement called dorsiflexion. Some extensor tendons continue all the way to the toes. There are much bulkier muscles on the back of the leg, and these form the calf. The gastrocnemius, and soleus underneath it, are large muscles that join together to form the Achilles tendon. They pull up on the lever of the calcaneus, pushing the ball of the foot down. They are involved as the foot pushes off from the ground during walking and running.

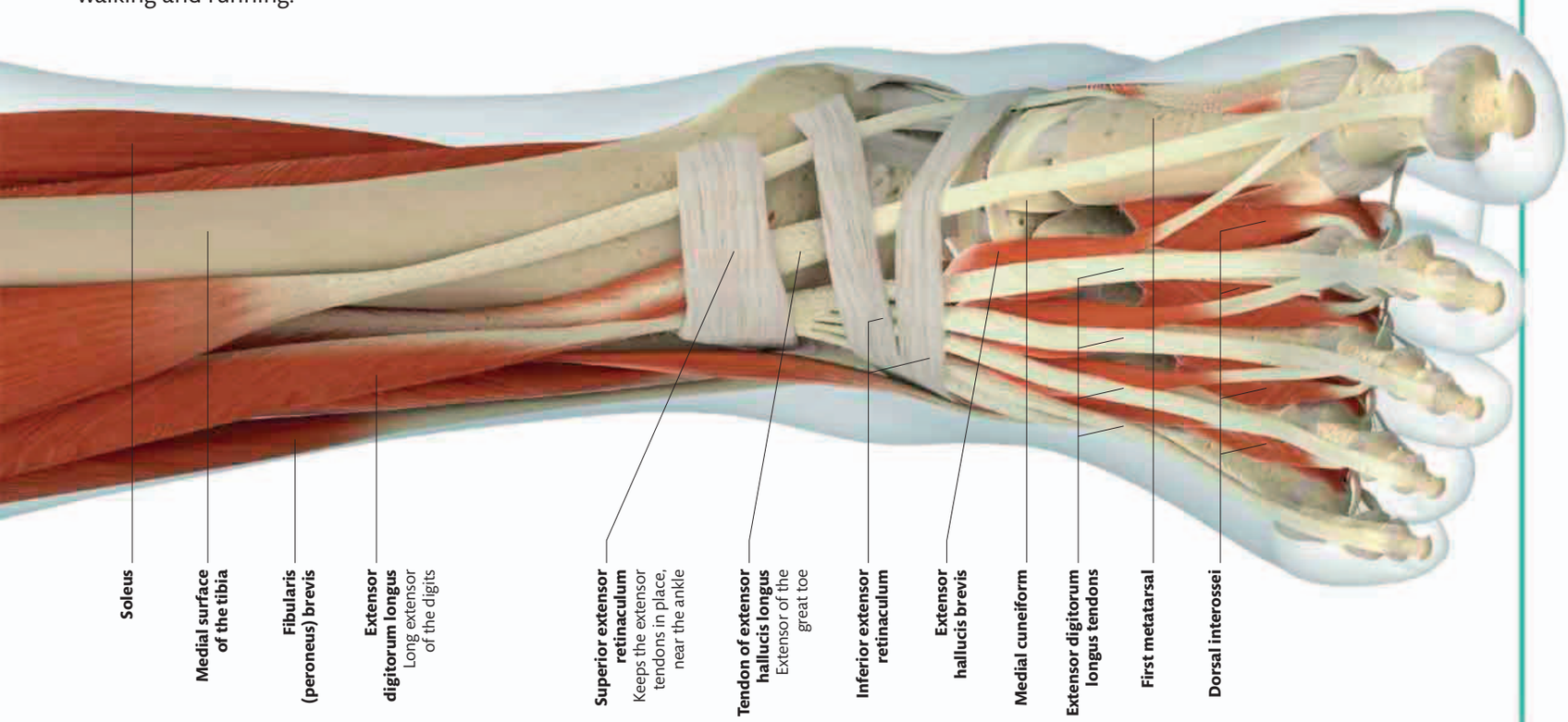

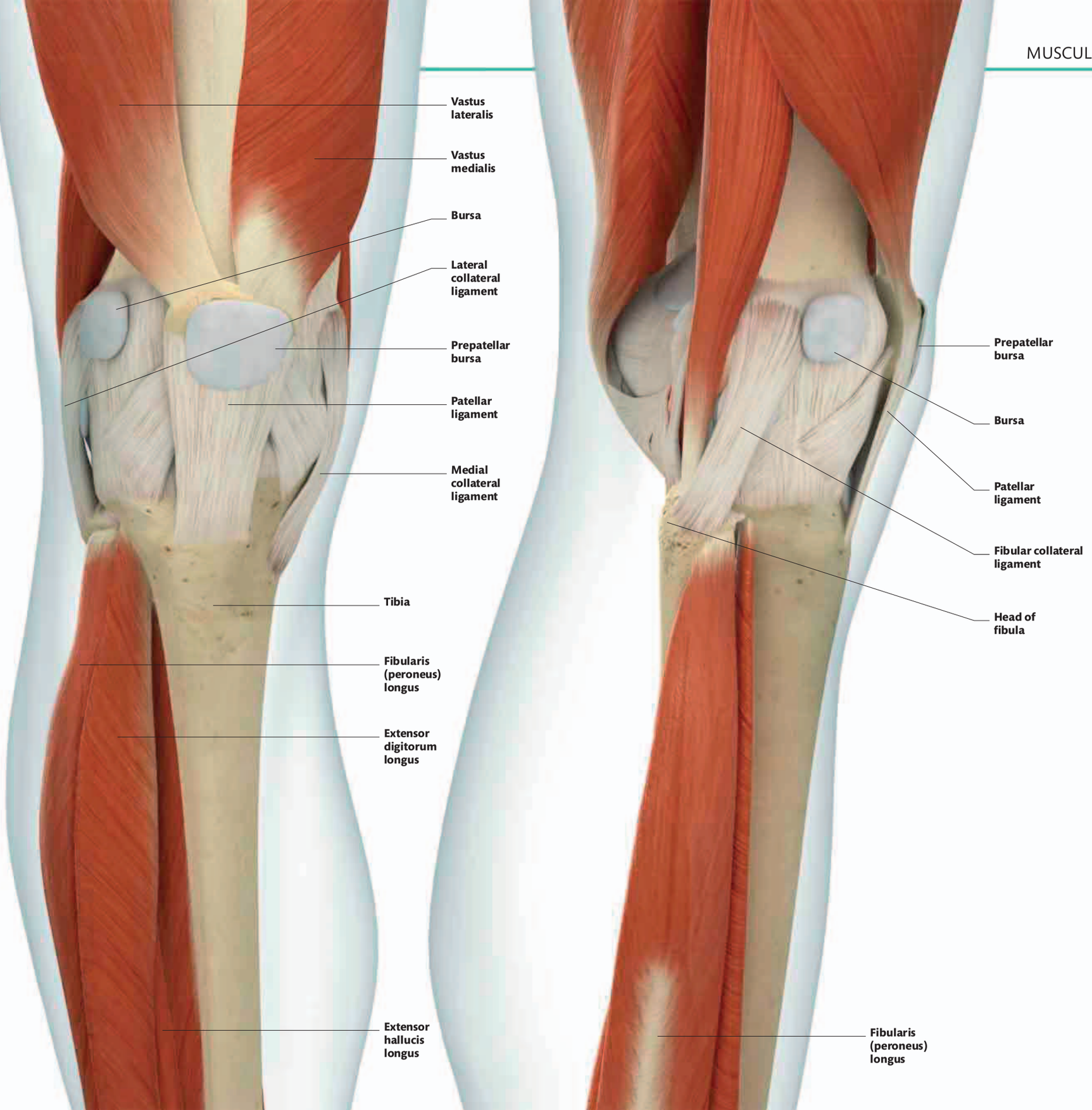
Vastus lateralis
Vastus medialis
Bursa
Lateral collateral ligament
Prepatellar bursa
Patellar ligament
Medial collateral ligament
Tibia
Fibularis (peroneus) longus
Extensor digitorum longus
Extensor hallucis longus
Prepatellar bursa
Bursa
Patellar ligament
Fibular collateral ligament
Head of fibula
Fibularis (peroneus) longus

LOWER LEG AND FOOT

Two muscles run along the outer, or lateral, side of the leg, down into the foot: the fibularis longus and fibularis brevis (see pp.104–05). These muscles pull the outer side of the foot upward, in a movement called eversion. The tendon of fibularis longus runs right underneath the foot, to attach on the inner side, and helps to maintain the transverse arch of the foot. The flexor hallucis longus arises from the fibula and interosseous membrane, and sends its tendon down, behind the medial malleolus and into the sole of the foot, to attach to the distal phalanx of the big toe.

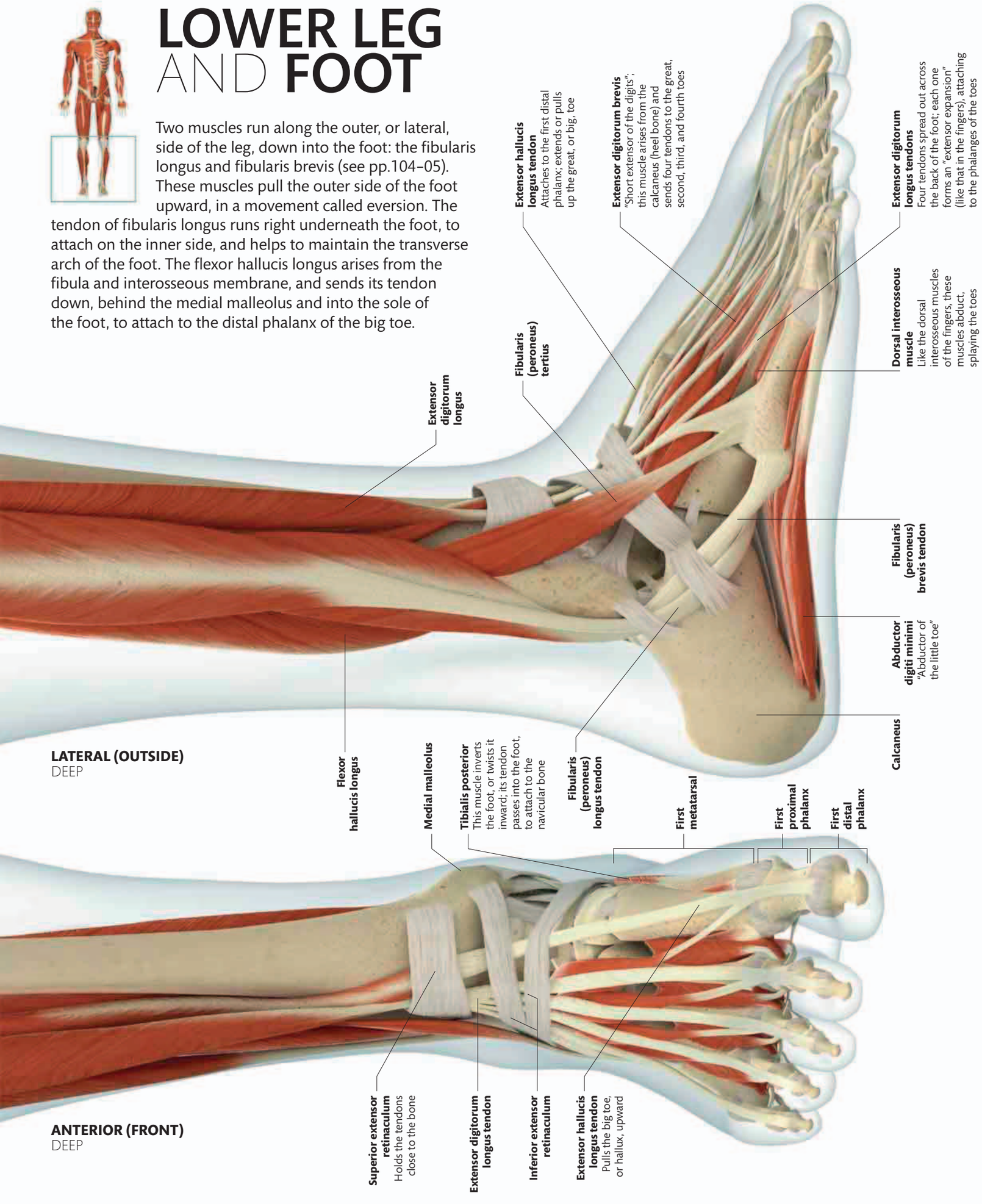

NERVOUS SYSTEM **OVERVIEW**

The nervous system contains billions of intercommunicating nerve cells, or neurons. It can be broadly divided into the central nervous system (brain and spinal cord) and the peripheral nervous system (cranial and spinal nerves and their branches). The brain and spinal cord are protected by the skull and vertebral column respectively. Twelve cranial nerves emerge from the brain and exit through holes in the skull to supply the head and neck; thirty-one pairs of spinal nerves leave via gaps between vertebrae to supply the rest of the body. You can also divide the nervous system by function. The part that deals more with the way we sense and interact with our surroundings is called the somatic nervous system. The part involved with sensing and controlling our internal environments—affecting glands or heart rate, for example—is the autonomic nervous system.

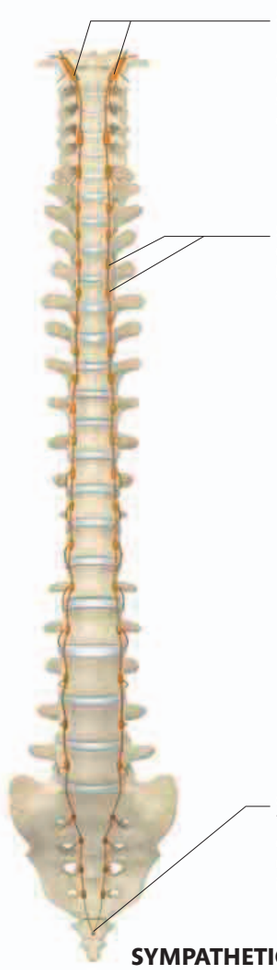

SYMPATHETIC TRUNK

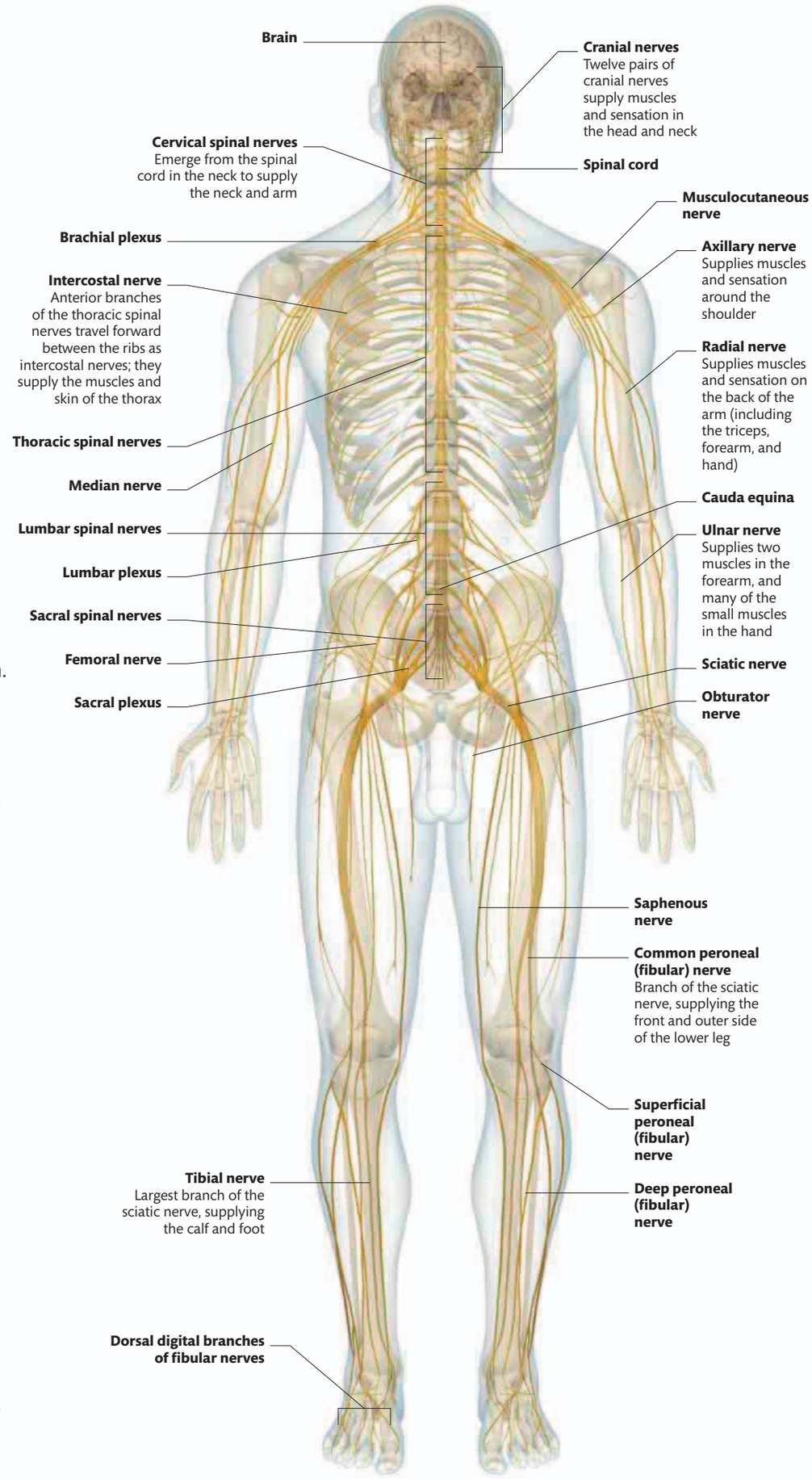

ANTERIOR (FRONT)

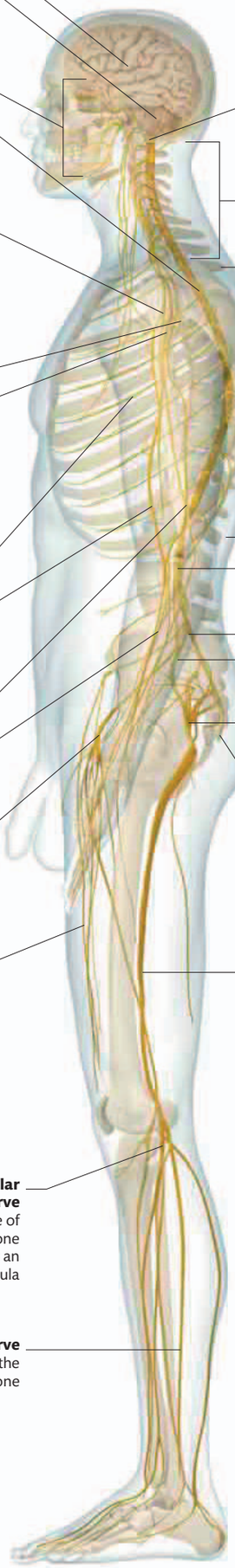

Cerebrum

Cerebellum
Literally little brain in Latin, this part of the brain is involved with balance and coordination of movement

Cranial nerves

Spinal cord
The continuation of the brain stem, lying protected within the vertebral canal of the spine

Musculocutaneous nerve
Supplies the muscles in the front of the upper arm, (including the biceps), as well as sensation to the skin of the outer side of the forearm

Axillary nerve

Brachial plexus
Anterior branches of the lower cervical spinal nerves, together with the first thoracic spinal nerve, form a network, or plexus, from which branches emerge to supply the arm, forearm, and hand

Intercostal nerve

Median nerve
Supplies most of the muscles in the front of the forearm, and also some in the hand

Radial nerve

Ulnar nerve
This nerve lies on the ulnar, or inner, side of the arm and forearm

Femoral nerve
Supplies sensation over the thigh and inner leg, and muscles in the front of the thigh, including the quadriceps

Obturator nerve
Supplies the muscles and skin of the inner thigh

Common fibular (peroneal) nerve
Lies on the outer side of the leg and is named after the bone around which it wraps; perona is an alternative Latin name for fibula

Tibial nerve
Named after the other bone of the lower leg—the tibia, or shinbone

Brainstem
Emerges from the foramen magnum in the base of the skull

Cervical spinal nerves
Cervical means of the neck; cervix is Latin for neck

Thoracic spinal nerves
Thorax is Latin for chest so the term thoracic means of the chest

Lumbar plexus
Anterior branches of the lumbar spinal nerves form a network here, from which nerves emerge to supply the leg

Sacral plexus
Anterior branches of sacral spinal nerves come together here as a network; the network provides nerves to the buttock and leg

Cauda equina
Below the end of the spinal cord, the lumbar and sacral nerve roots continue for some way inside the vertebral canal, before emerging from the spine

Sciatic nerve
Largest nerve in the body, which supplies the hamstrings in the back of the thigh; its branches supply muscles and sensation in the lower leg and foot

SIDE

NEURON STRUCTURE

A single neuron such as the cell shown below from the central nervous system can make contact with hundreds of other neurons, creating an incredibly complex network of connections. Each neuron's cell body has projections or, dendrites. One is usually longer and thinner than the rest, and this is the axon. Some axons within the brain are less than 1/32in (1mm) in length; others, stretching from the spinal cord to limb muscles, can measure over 39in (1m) long.

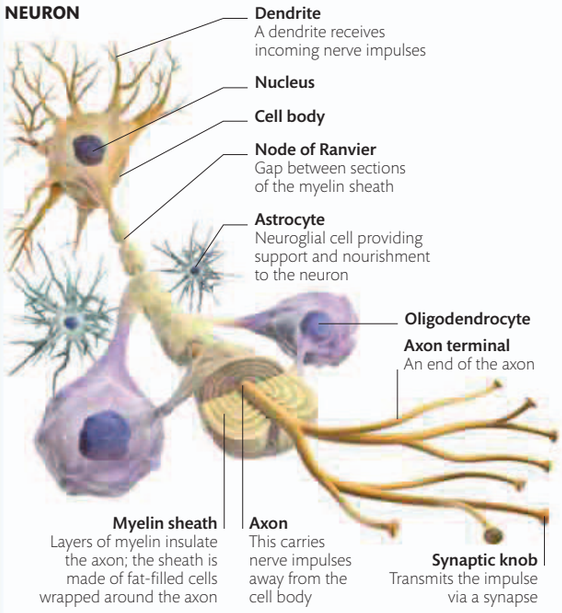

BRAIN

Compared to other animals, humans have massive brains for the size of our bodies. The human brain has grown larger and larger over the course of evolution, and it is now so overblown that the frontal lobes of the brain lie right over the top of the orbits that contain the eyes. Think about any other mammal, perhaps a dog or a cat for easy reference, and you will quickly realize what an odd shape the human head is—and most of that is a result of our huge brains. Looking at a side view of the brain, you can see all the lobes that make up each cerebral hemisphere: the frontal, parietal, temporal, and occipital lobes (individually colored, below). Tucked under the cerebral hemispheres at the back of the brain is the cerebellum (Latin for little brain). The brain stem leads down, through the foramen magnum of the skull, to the spinal cord.

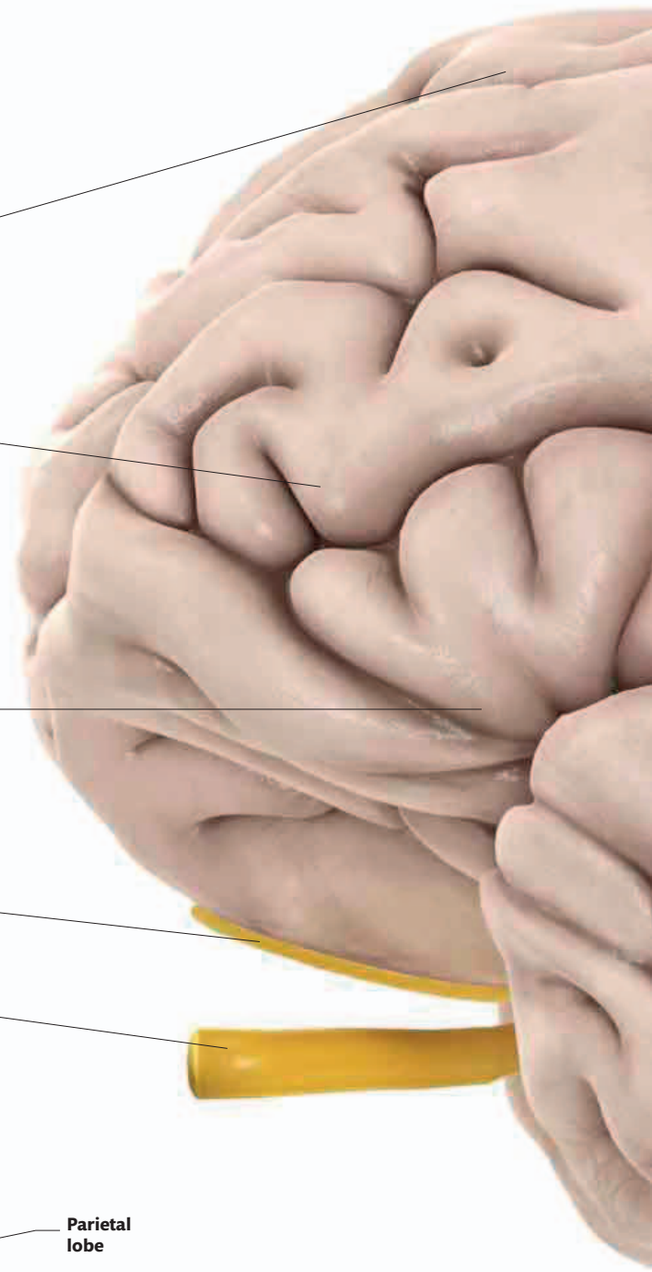

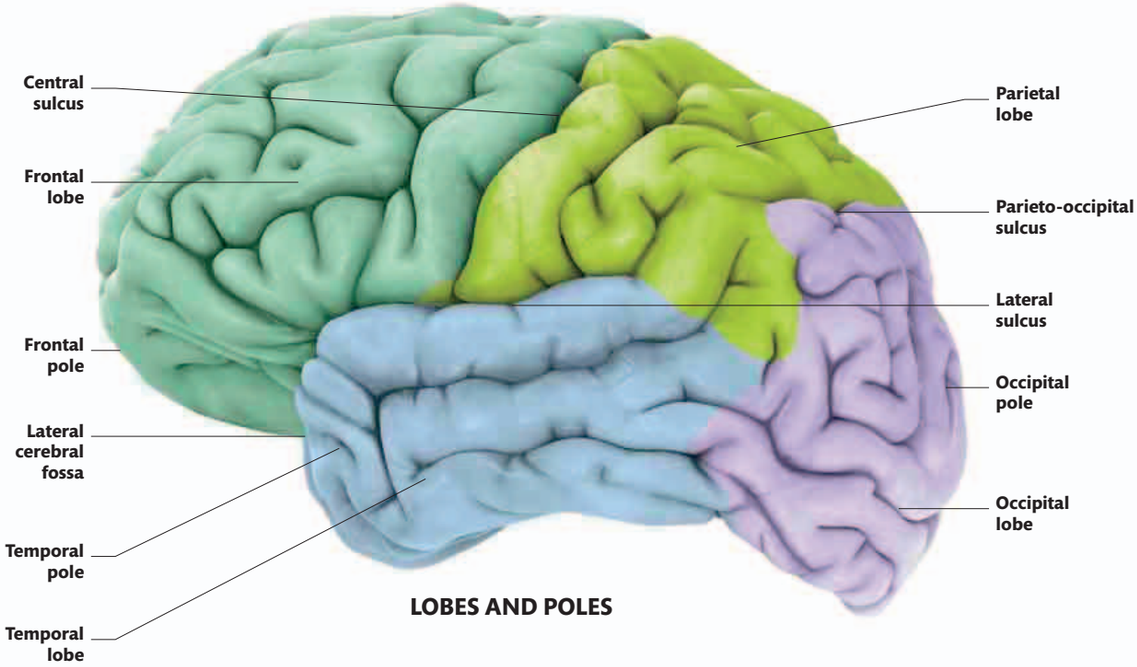

LOBES AND POLES

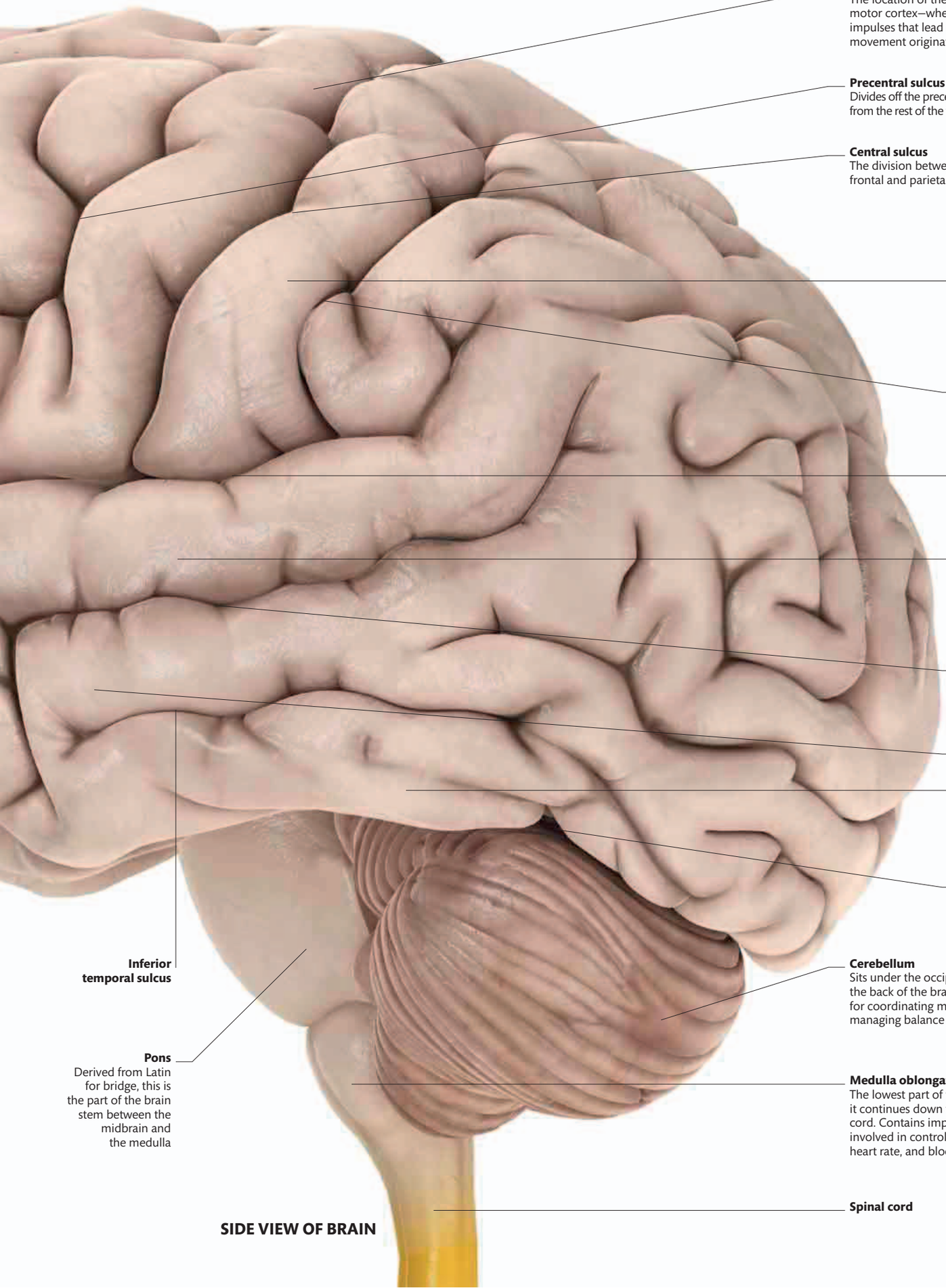

Precentral gyrus
The location of the primary motor cortex—where nerve impulses that lead to muscle movement originate

Precentral sulcus
Divides off the precentral gyrus from the rest of the frontal lobe

Central sulcus
The division between the frontal and parietal lobes

Postcentral gyrus
Lies just behind the central sulcus. The primary somatosensory cortex, which receives sensory information from all over the body

Postcentral sulcus
Separates the postcentral gyrus from the rest of the parietal lobe

Lateral sulcus
A deep cleft dividing the frontal and parietal lobes from the temporal lobe below

Superior temporal gyrus
Includes the primary auditory cortex, where sensory information related to hearing is received

Superior temporal sulcus
Sulcus is a Latin word meaning groove or furrow

Middle temporal gyrus

Inferior temporal gyrus

Preoccipital notch

Inferior temporal sulcus

Cerebellum
Sits under the occipital lobes at the back of the brain; responsible for coordinating movement and managing balance and posture

Pons
Derived from Latin for bridge, this is the part of the brain stem between the midbrain and the medulla

Medulla oblongata
The lowest part of the brain stem; it continues down to form the spinal cord. Contains important centers involved in controlling breathing, heart rate, and blood pressure

Spinal cord

SIDE VIEW OF BRAIN

BRAIN

From an anatomist's point of view, the brain is quite an ugly and unprepossessing organ. It looks rather like a large, pinkish gray, wrinkled walnut—especially when viewed from above. The outer layer of gray matter, called the cortex, is highly folded. Underneath the brain we see some more detail, including some of the cranial nerves that emerge from the brain itself. To the naked eye, there is little to suggest that the brain is the most complicated organ in the human body. Its true complexity is only visible through a microscope, revealing billions of neurons that connect with each other to form the pathways that carry our senses, govern our actions, and create our minds.

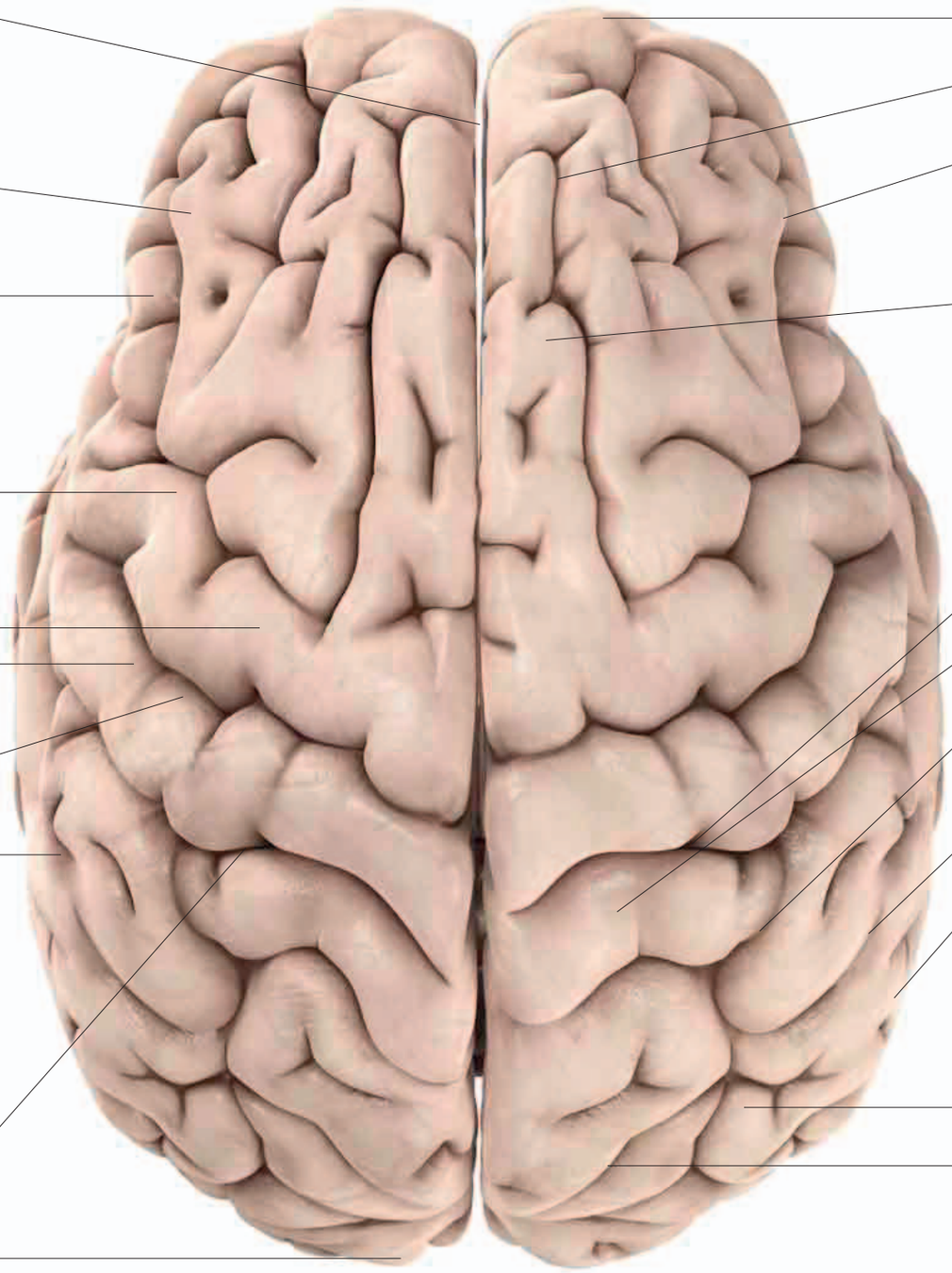

TOP VIEW OF BRAIN

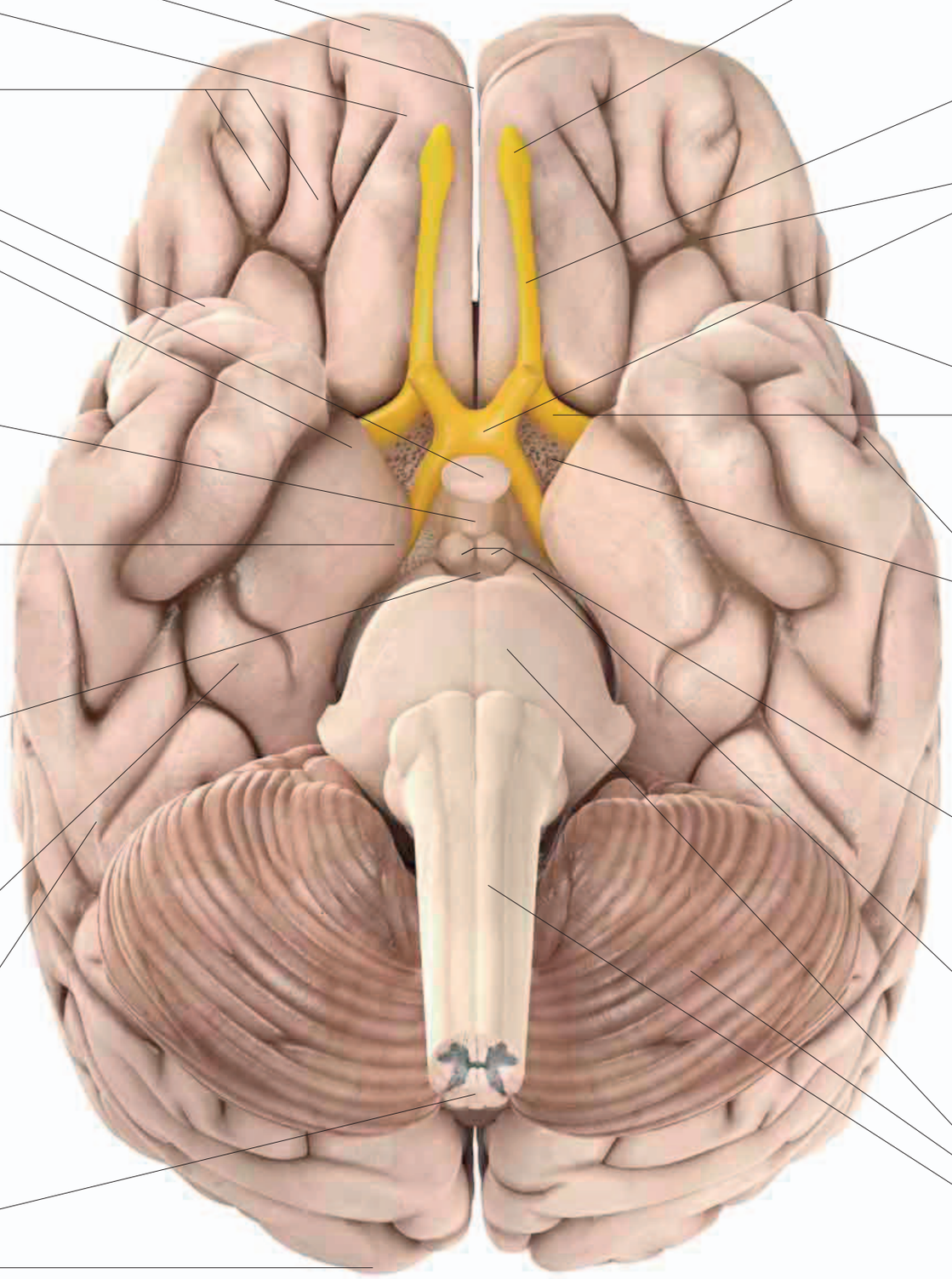

Frontal pole

Longitudinal (cerebral) fissure

Straight gyrus

Orbital gyri
Lie around the edges of the H-shaped orbital sulcus, and seem to play some role in empathy

Temporal pole

Pituitary gland

Parahippocampal gyrus
This part of the cortex, close to the hippocampus, plays an important role in memory and recognition

Tuber cinereum
Small bump of gray matter under the brain; part of the hypothalamus (see p.116)

Uncus
Hooked-under end of the parahippocampal gyrus; contains the primary olfactory cortex, receiving olfactory (smell) information

Interpeduncular fossa
Area enclosed by the cerebral peduncles on each side, by the optic chiasma in front, and the pons of the brain stem behind

Occipitotemporal fusiform gyrus

Inferior temporal gyrus

Spinal cord

Occipital pole

Olfactory bulb
Receives olfactory nerves, which have emerged from the top of the nasal cavity through the cribiform plate of the ethmoid bone, to enter the inside of the skull

Olfactory tract
Carries olfactory (smell) information back to the uncus

Orbital sulcus

Optic chiasma
Where the two optic nerves meet and swap fibers with each other, to form the optic tracts; chiasma means a cross

Lateral cerebral fossa

Olfactory trigone
The olfactory tract splays out into this triangular shape, just in front of the anterior perforated substance

Inferior temporal sulcus

Anterior perforated substance
Area of gray matter between the olfactory trigone, the optic chiasma, and the uncus; pierced by small arteries from the anterior and middle cerebral arteries

Mammillary bodies
Two breast-like bumps that are part of the limbic system, which is involved in memory, emotions, and behavior

Cerebral peduncle
"Stalk" of the brain, containing motor nerve fibers that descend from the cerebral cortex to the brain stem and spinal cord

Pons

Cerebellum

Pyramid
A prominence on the front of medulla containing motor nerve fibers that run from the cortex of the brain to the spinal cord

UNDERSIDE OF BRAIN

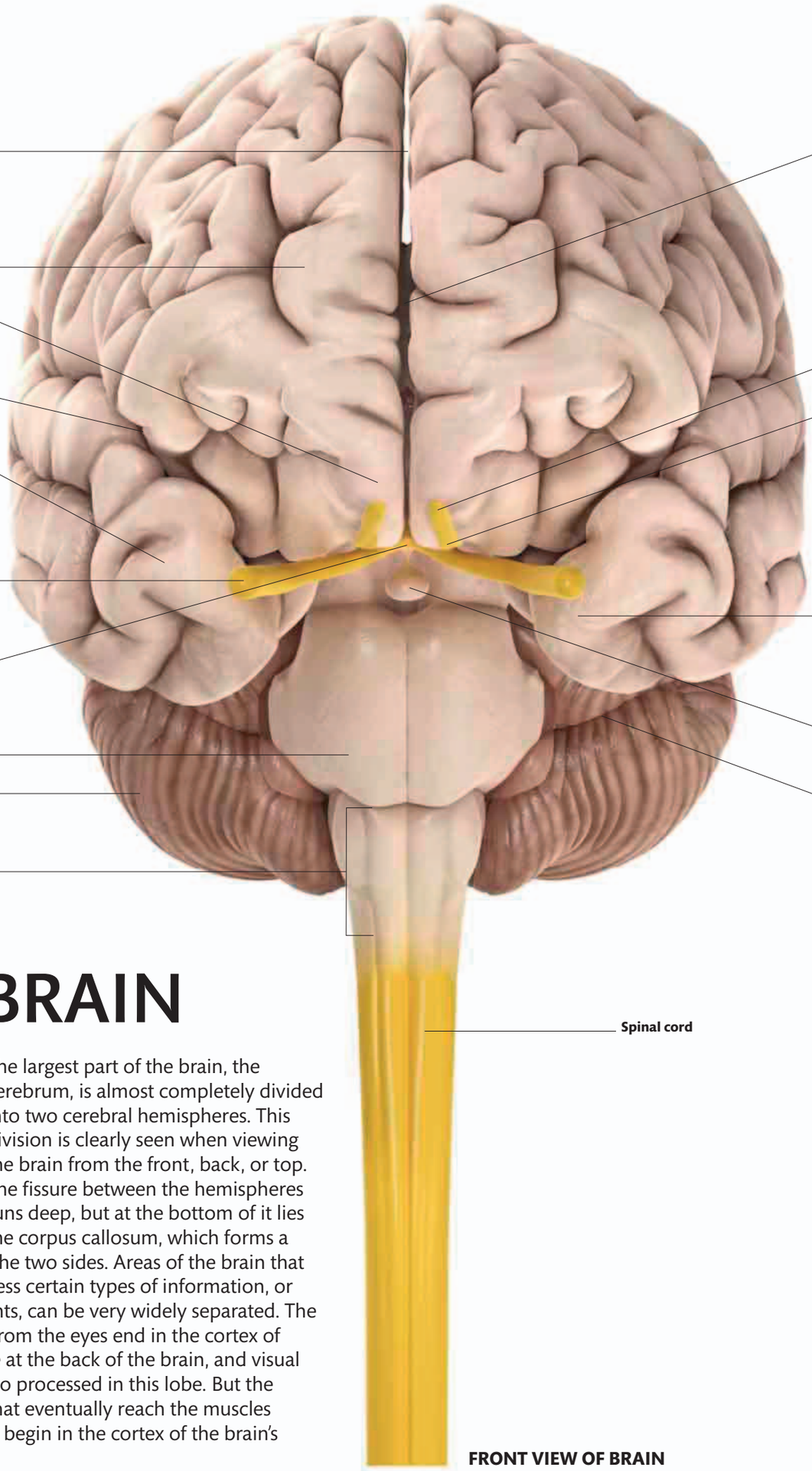

FRONT VIEW OF BRAIN

BRAIN

The largest part of the brain, the cerebrum, is almost completely divided into two cerebral hemispheres. This division is clearly seen when viewing the brain from the front, back, or top. The fissure between the hemispheres runs deep, but at the bottom of it lies the corpus callosum, which forms a bridge between the two sides. Areas of the brain that receive and process certain types of information, or govern movements, can be very widely separated. The visual pathways from the eyes end in the cortex of the occipital lobe at the back of the brain, and visual information is also processed in this lobe. But the nerve impulses that eventually reach the muscles to move the eyes begin in the cortex of the brain's frontal lobe.

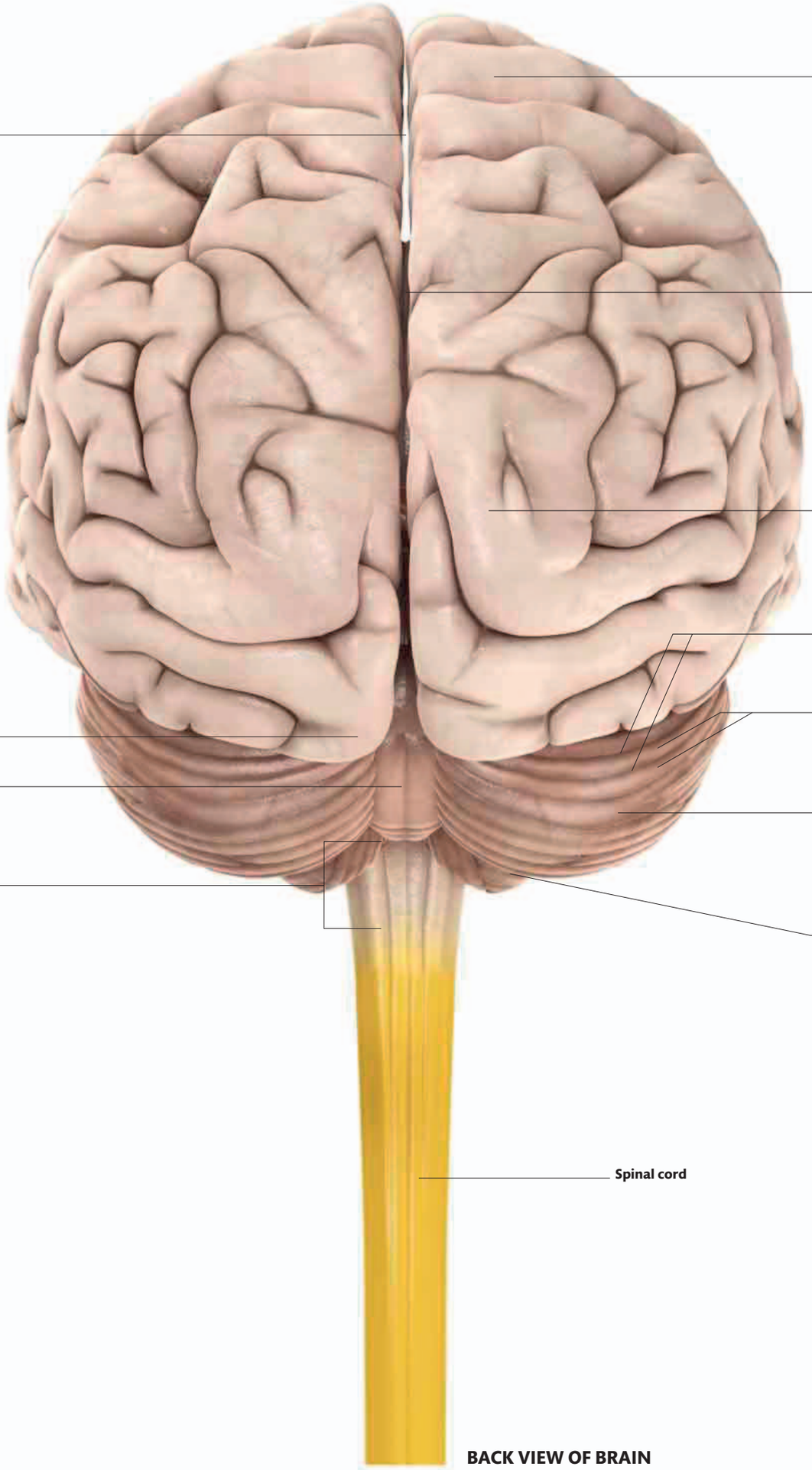

BACK VIEW OF BRAIN

Body of corpus callosum
The largest commissure (or bundle of connecting nerve fibers) between the two hemispheres, this forms the roofs of the lateral ventricles

Superior frontal gyrus

Cingulate gyrus
Cingulum is the Latin for girdle and this gyrus wraps closely around the corpus callosum; it is part of the limbic system, which is involved with emotional responses and behaviors

Septum pellucidum
This translucent partition is a thin dividing wall between the two lateral ventricles

Genu of corpus callosum
The anterior (front) end of the corpus callosum is bent over—genu means knee in Latin

Anterior commissure
A bundle of nerve fibers connecting parts of the two cerebral hemispheres

Optic chiasma
The crossover point where the two optic nerves meet and swap fibers, then part company as the optic tracts, which continue on each side of the brain toward the thalamus

SAGITTAL SECTION THROUGH BRAIN

Hypothalamus
Plays an important role in regulating the internal environment of the body, by keeping a check on body temperature, blood pressure, and blood sugar level, for instance

Pituitary gland
Produces many hormones and forms a link between the brain and endocrine system

Mammillary body
Part of the limbic system of the brain

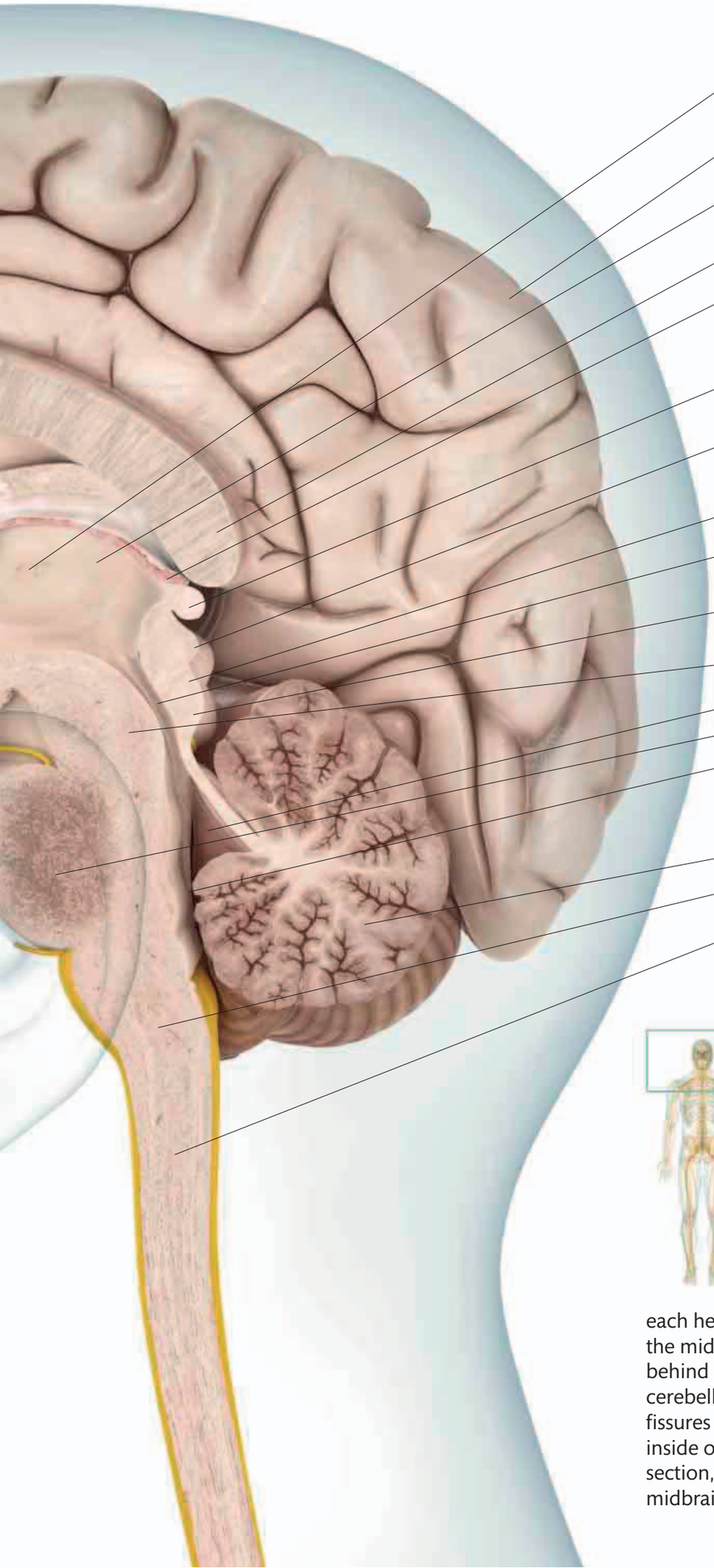

Interthalamic adhesion
Connection between the thalami on each side of the brain

Cerebrum
The largest part of the brain, consisting of the two cerebral hemispheres

Thalamus
Processes and relays sensory and motor information to higher brain centers

Splenium of corpus callosum
The posterior end of the corpus callosum

Choroid plexus of the third ventricle
A choroid plexus is formed where the inner and outer membranes of the brain come together; it is full of capillaries and produces cerebrospinal fluid, which flows into the ventricle

Pineal gland
Produces the hormone melatonin and is involved in the regulation of sleep–wake cycles

Superior colliculus
Involved in visual reflex pathways, including the pupillary light reflex, which makes the pupils constrict when bright light hits the retina

Tectum of the midbrain
The roof of the midbrain

Cerebral aqueduct
A narrow channel connecting the third and fourth ventricles

Inferior colliculus
Involved with auditory pathways, including reflex responses to loud noises

Tegmentum of midbrain

Fourth ventricle

Pons

Median aperture of the fourth ventricle
Cerebrospinal fluid escapes from the fourth ventricle via this opening in the midline, as well as through an opening on each side, into the subarachnoid space around the brain and spinal cord

Cerebellum

Medulla oblongata

Spinal cord

HEAD AND NECK

This median sagittal section—a vertical slice right through the middle of the brain—shows clearly the corpus callosum, which links the two hemispheres. We also see that the brain is not solid: there are cavities within it. Two spaces (or ventricles) lie inside each hemisphere, while the third and fourth ventricles are located on the midline. These spaces are full of cerebrospinal fluid. Beneath and behind the cerebrum sits the cerebellum. The gray cortex of the cerebellum is more finely folded than that of the cerebrum, with fissures separating its leaves (or folia). Sliced through this way, the inside of the cerebellum reveals a beautiful, treelike pattern. In this section, we can also see clearly all the parts of the brain stem—the midbrain, pons, and medulla.

BRAIN

The brain is protected by three membranes called the meninges (which become inflamed in meningitis). The tough dura mater layer is the outermost covering, which surrounds the brain and the spinal cord. Under the dura mater is the cobweblike arachnoid mater layer. The delicate pia mater is a thin membrane on the surface of the brain. Between the pia mater and the arachnoid mater there is a slim gap—the subarachnoid space—which contains cerebrospinal fluid (CSF). Mainly produced by the choroid plexus in the brain's lateral ventricles, CSF flows through the third ventricle into the fourth, where it can escape via small apertures into the subarachnoid space.

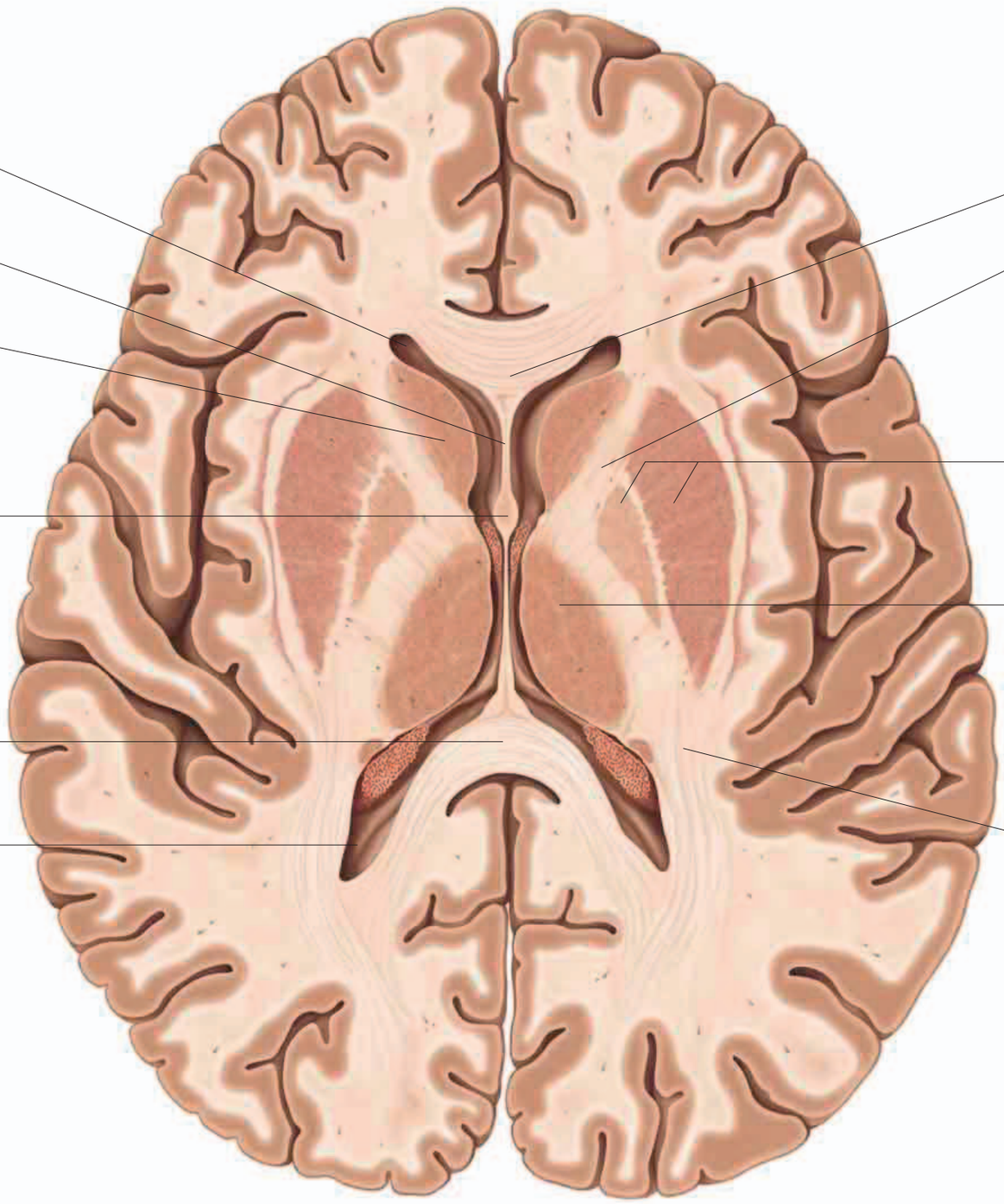

TRANSVERSE SECTION OF BRAIN

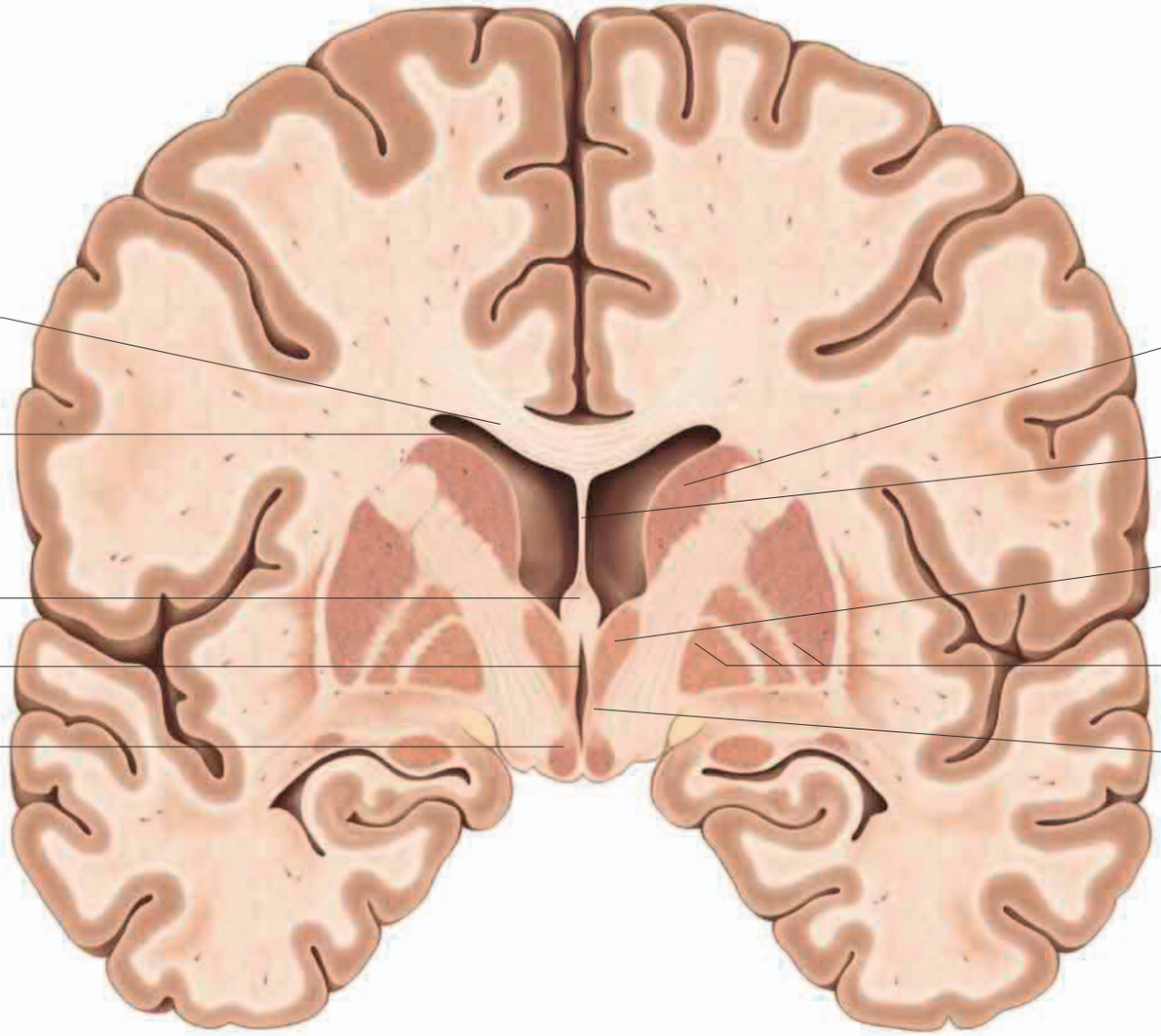

CORONAL SECTION OF BRAIN

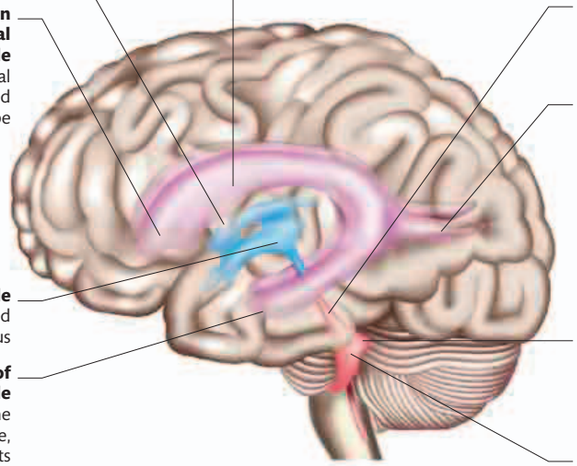

VENTRICLES OF BRAIN

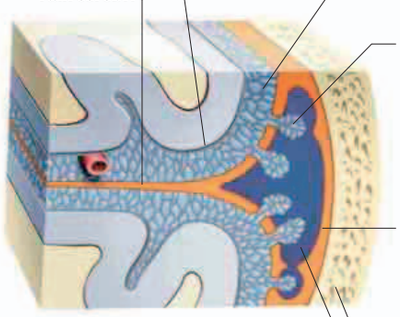

MENINGES SECTION

HEAD AND NECK

The 12 pairs of cranial nerves (the standard abbreviation for which is CN) emerge from the brain and brain stem, leaving through holes, or "foramina", in the base of the skull. Some nerves are purely sensory, some just have motor functions, but most contain a mixture of motor and sensory fibers. A few also contain autonomic nerve fibers. The olfactory nerve and the optic nerve attach to the brain itself. The other 10 pairs of cranial nerves emerge from the brain stem. All the cranial nerves supply parts of the head and neck, except the vagus nerve. This has branches in the neck, but then continues on to supply organs in the thorax and down to the abdomen. Careful testing of cranial nerves, including tests of sight, eye and head movement, taste, and so on, can help doctors to pinpoint neurological problems in the head and neck.

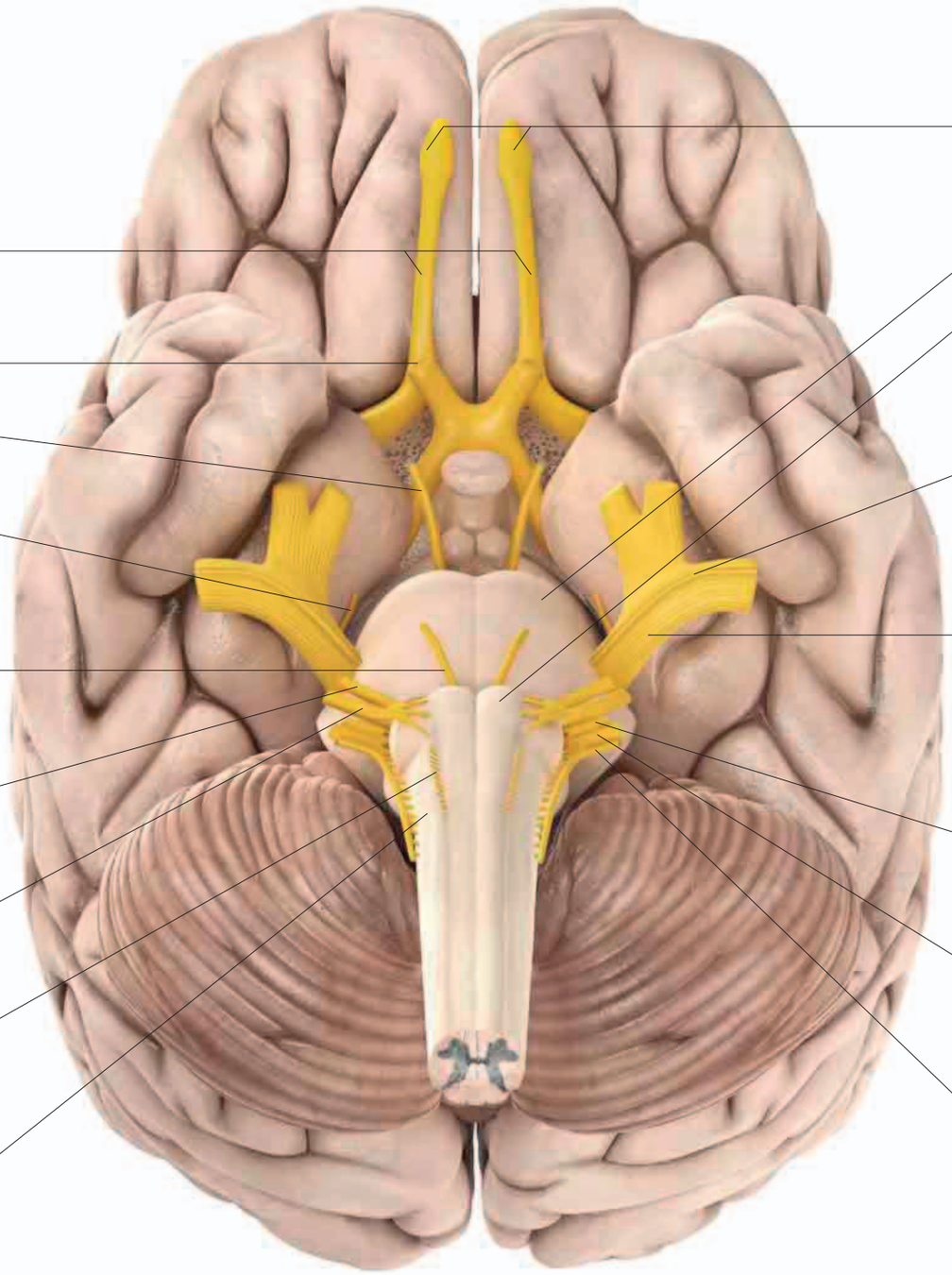

ORIGIN OF CRANIAL NERVES (UNDERSIDE OF BRAIN)

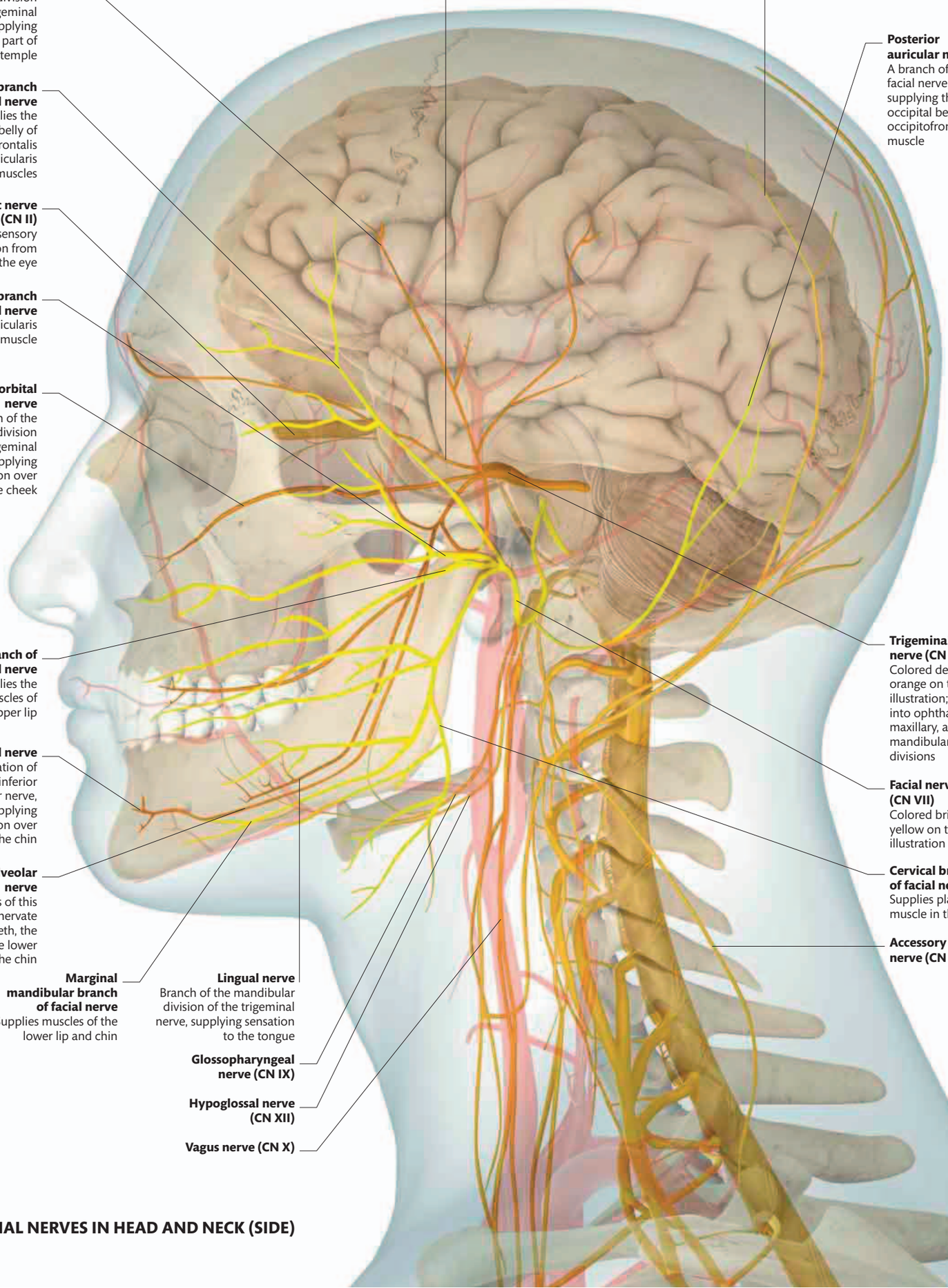

CRANIAL NERVES IN HEAD AND NECK (SIDE)

EYE

The eyes are precious organs. They are well protected inside the eye sockets, or bony orbits, of the skull. They are also protected by the eyelids, and bathed in tears produced by the lacrimal glands. Each eyeball is only 1in (2.5cm) in diameter. The orbit provides an anchor for the muscles that move the eye, and the rest of the space inside the orbit is largely filled up with fat. Holes and fissures at the back of this bony cavern transmit nerves and blood vessels, including the optic nerve, which carries sensory information from the retina to the brain. Other nerves supply the eye muscles and the lacrimal glands, and even continue on to the face to supply sensation to the skin of the eyelids and forehead.

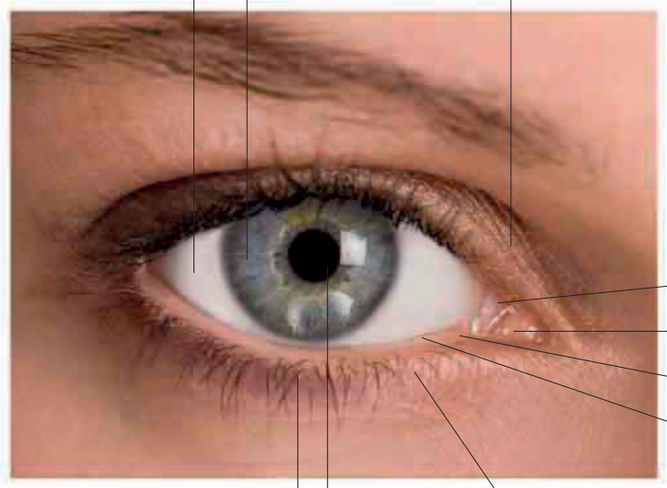

EXTERNAL EYE

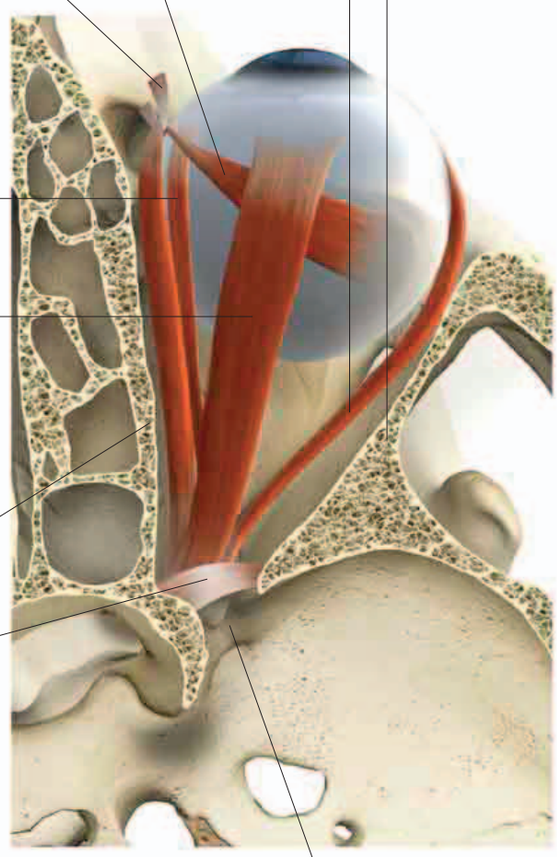

MUSCLES OF THE EYE (FROM ABOVE)

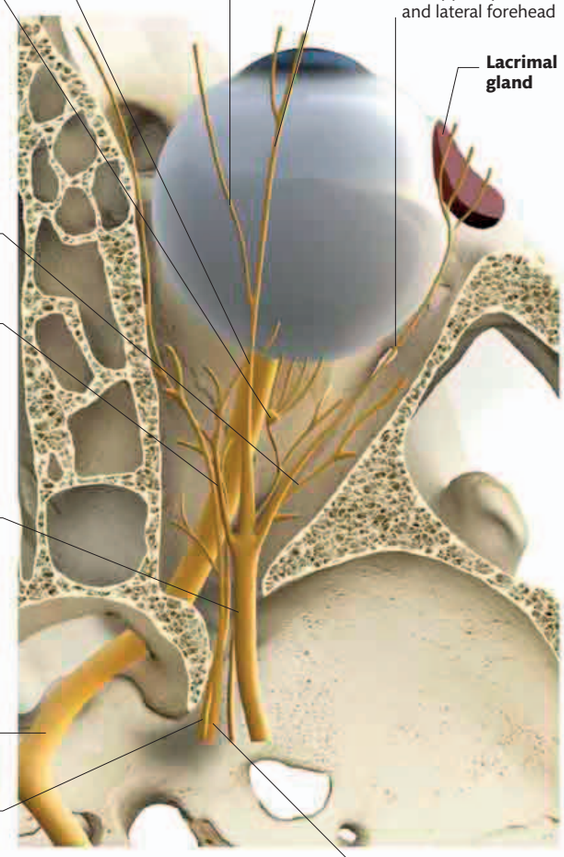

NERVES OF THE ORBIT (FROM ABOVE)

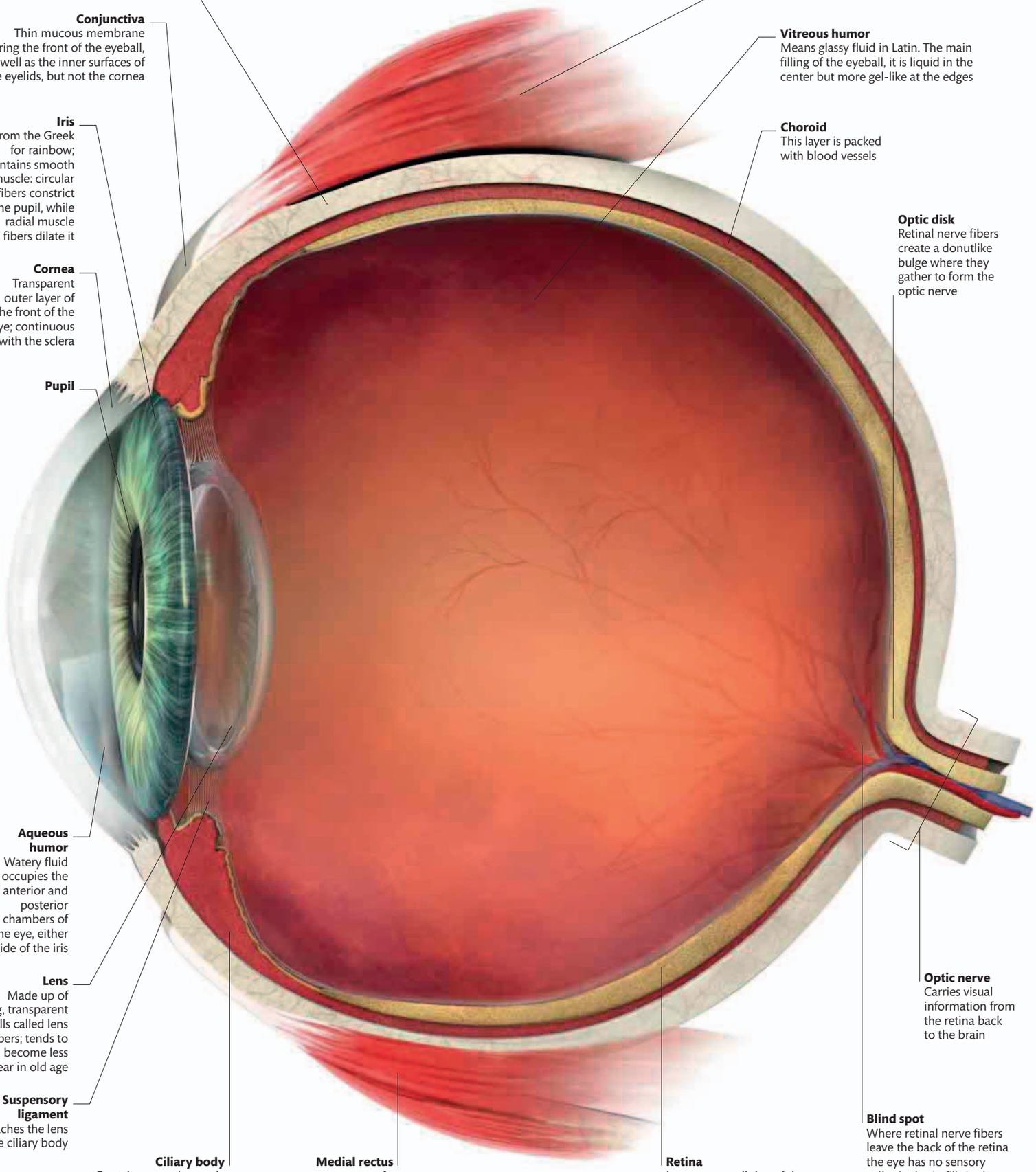

HORIZONTAL SECTION THROUGH THE EYEBALL

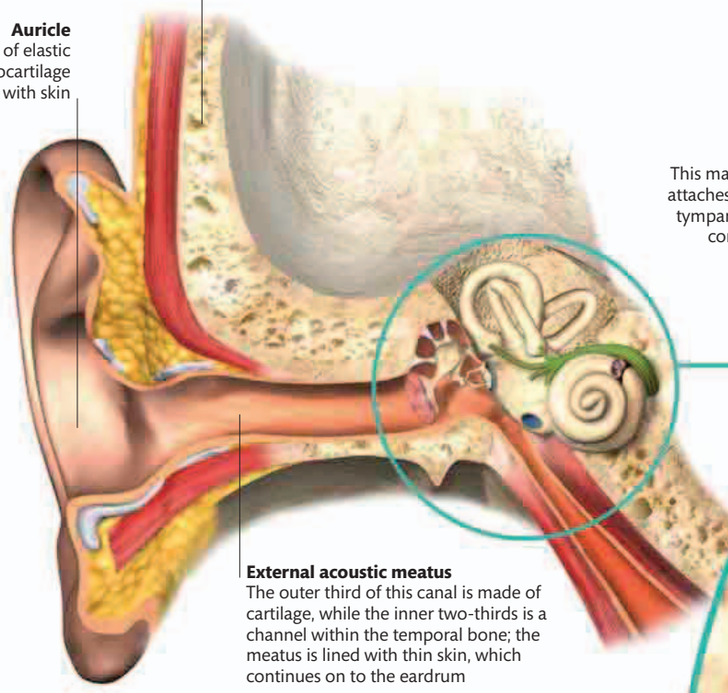

EXTERNAL EAR

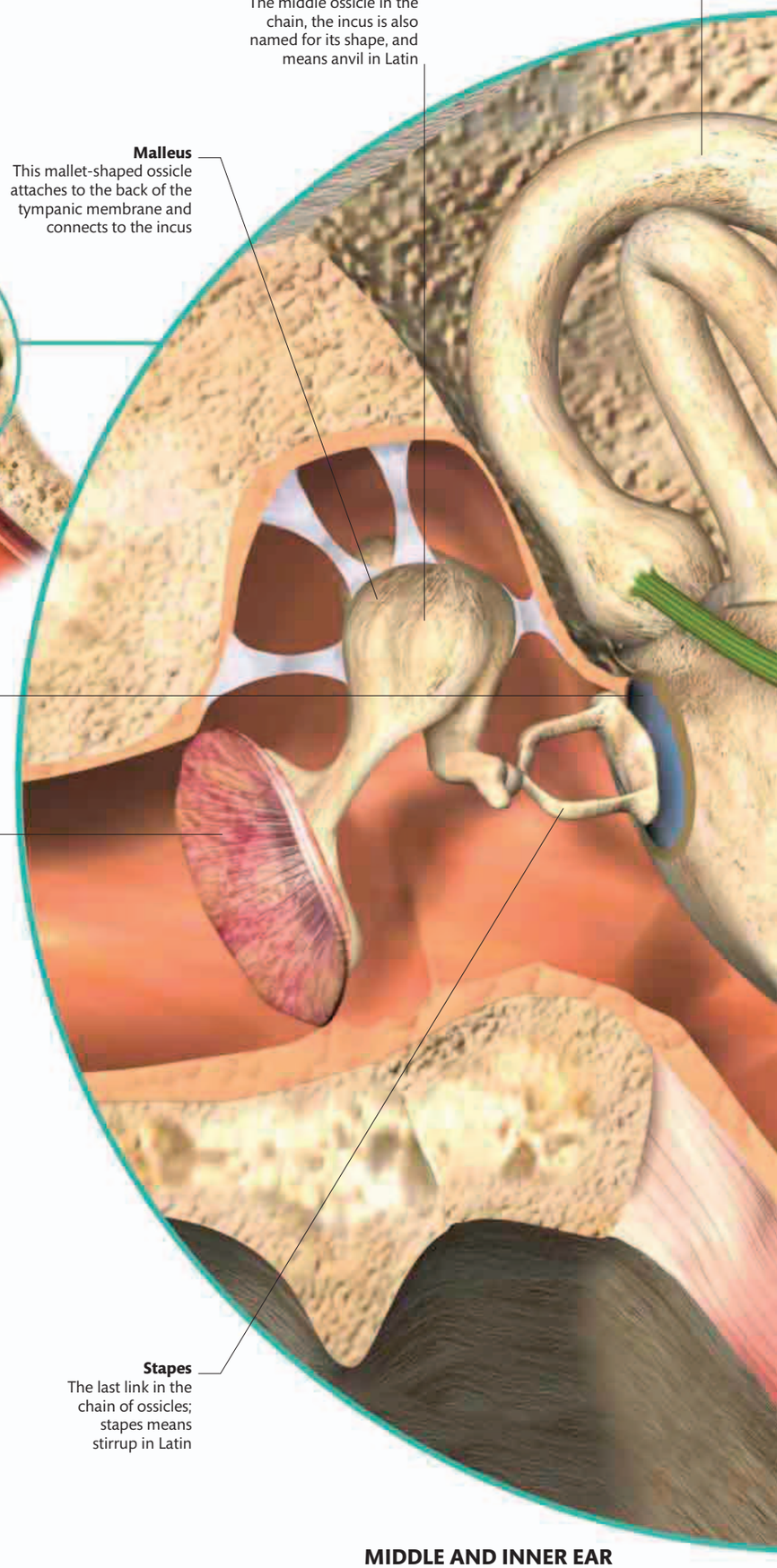

MIDDLE AND INNER EAR

EAR

The ear can be divided up into external, middle, and internal parts. The external ear includes the auricle on the outside of the head, and the external acoustic meatus—the canal that leads to the eardrum, or tympanic membrane. The middle ear is an air space inside the temporal bone. It contains the ossicles (ear bones) and is linked to the pharynx by the pharyngotympanic, or Eustachian, tube. Minute hair cells inside the inner ear convert vibrations in the fluid within the cochlea into an electrical nerve impulse. Similar hair cells in the vestibular apparatus (the semicircular canals, utricle, and saccule) convert mechanical stimuli, produced by motions of the head, into nerve impulses. The sensory nerves leaving the inner ear join to form the vestibulocochlear nerve.

Lateral semicircular canal
This is positioned horizontally

Anterior semicircular canal
Positioned vertically, but at right angles to the plane of the posterior semicircular canal

Vestibular nerve
Carries sensory information from the vestibular apparatus—including the semicircular canals

Cochlear nerve
Conveys sensory information about sound from the cochlea

Section cut from cochlea
From top to bottom shows vestibular canal, cochlear duct, and tympanic canal

Vestibulocochlear nerve
The cochlear nerve conveys sensory information about sound from the cochlea. It joins the vestibular nerve to the vestibulocochlear nerve

Cochlea
Not surprisingly, cochlea means snail in Latin

Vestibule
Contains the utricle and saccule, organs of balance

Round window
Vibrations can travel in the fluid inside the cochlea, all the way up to its apex and back down to the round window

Pharyngotympanic tube
Passage connecting the middle ear to the back of the throat, and allowing air pressure either side of the eardrum to be equalized

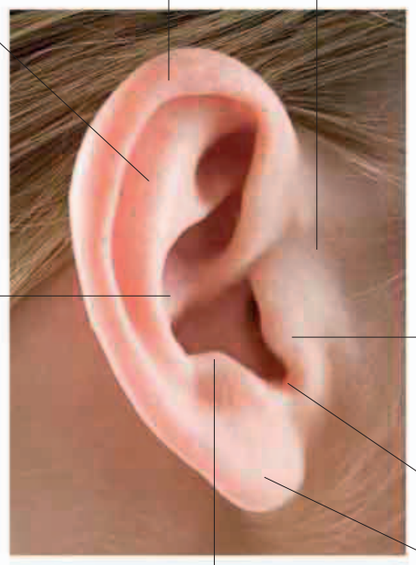

Helix
The outer rim of the auricle

External acoustic meatus

Antihelix
A curved prominence, parallel to the helix

Concha
This hollow is named after the Greek for shell

Tragus
This little flap overlaps the external acoustic meatus

Intertragic notch

Lobule

Antitragus
A small tubercle opposite the tragus

AURICLE

Tympanic membrane
As seen with an otoscope, a healthy eardrum has a pearly, almost translucent appearance

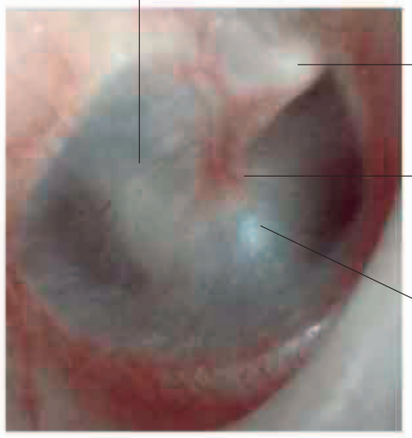

Lateral process of malleus

Handle of malleus

Cone of light
Light is reflected in the front, lower quadrant of the eardrum

EARDRUM

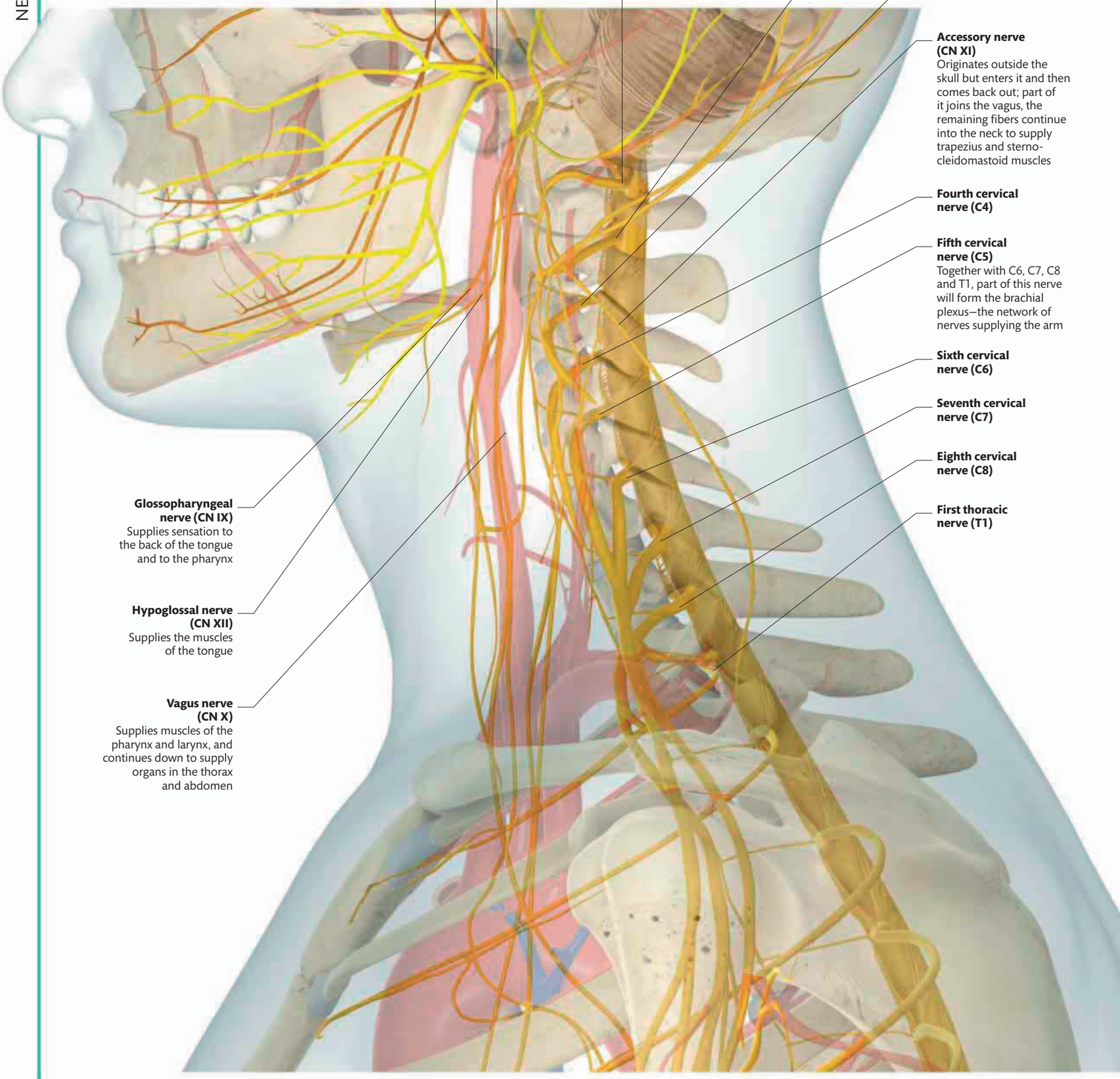

NERVES OF THE NECK (SIDE)

NECK

The last four cranial nerves all appear in the neck. The glossopharyngeal nerve supplies the parotid gland and the back of the tongue, then runs down to the pharynx. The vagus nerve is sandwiched between the common carotid artery and the internal jugular vein, and it gives branches to the pharynx and larynx before continuing down into the thorax. The accessory nerve supplies the sternocleidomastoid and trapezius muscles in the neck, while the last cranial nerve, the hypoglossal, dips down below the mandible, then curves back up to supply the muscles of the tongue. We can also see spinal nerves in the neck. The upper four cervical nerves supply neck muscles and skin, while the lower four contribute to the brachial plexus and are destined for the arm.

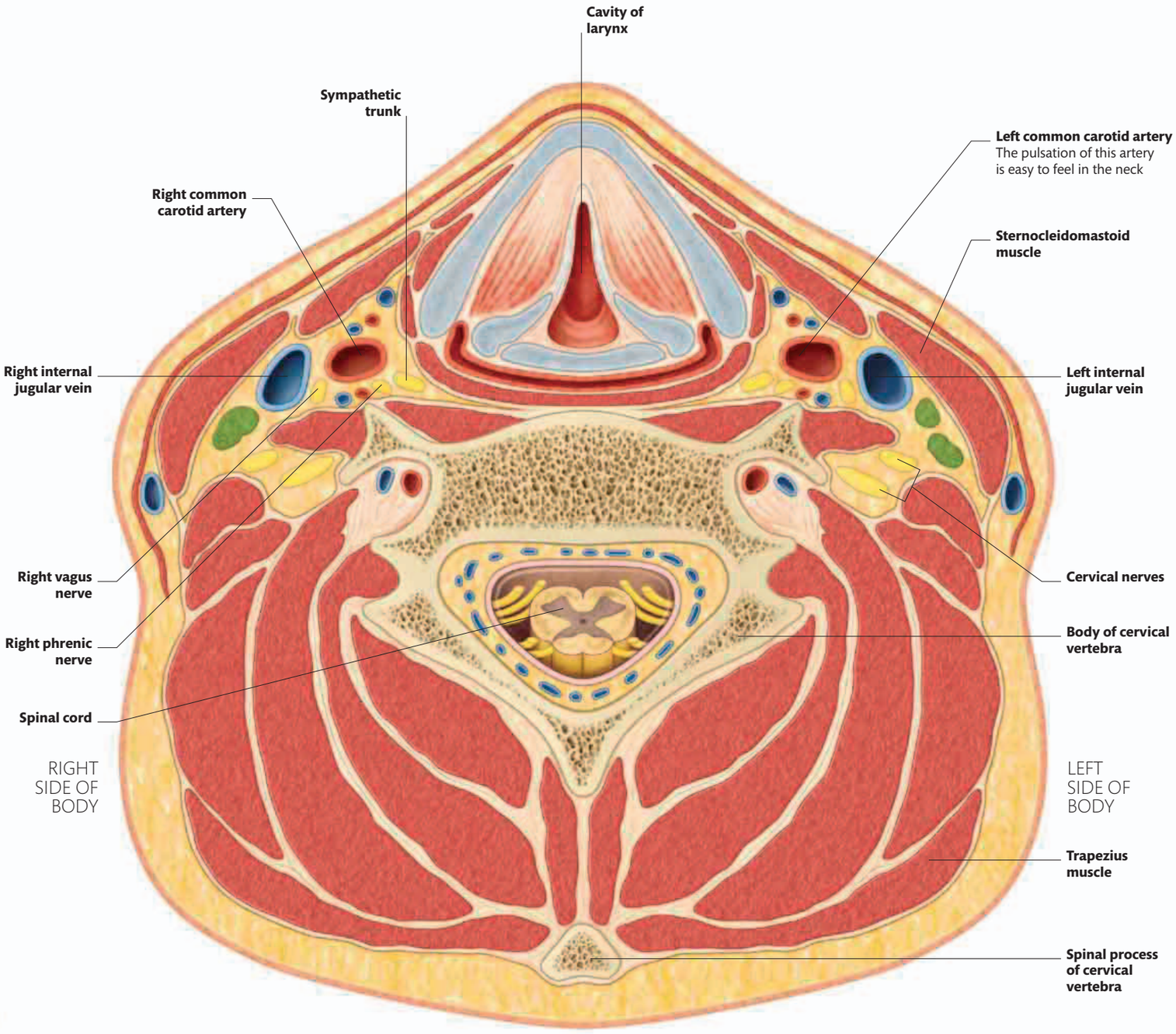

TRANSVERSE SECTION OF THE NECK

Vagus nerve
The tenth cranial nerve strays a long way beyond the neck to supply structures in the thorax and abdomen as well; its name means wandering or straying

First rib

First intercostal nerve
Anterior branch of T1 (first thoracic) spinal nerve

Phrenic nerve
Comes from the third, fourth, and fifth cervical nerves; supplies the muscle of the diaphragm and the membranes lining either side of it—the pleura on the thoracic side and peritoneum on the abdominal side

ANTERIOR (FRONT)

THORAX

Pairs of spinal nerves emerge via the intervertebral foramina (openings) between the vertebrae. Each nerve splits into an anterior and a posterior branch. The posterior branch supplies the muscles and skin of the back. The anterior branches of the upper 11 thoracic spinal nerves run, one under each rib, as intercostal nerves, supplying the intercostal muscles and overlying skin. The anterior branch of the last thoracic spinal nerve runs under the twelfth rib as the subcostal nerve. In addition to motor and sensory fibers, thoracic spinal nerves contain sympathetic nerve fibers that are linked by tiny connecting branches to the sympathetic chain or trunk (see pp.108–109). This allows sympathetic nerves originating from one level of the spinal cord to travel up and down, and spread out to several body segments.

Sixth rib

Eighth rib

Eighth intercostal nerve
Like each intercostal nerve, this supplies the muscles lying in the same intercostal space, and also supplies sensation to a strip of skin around the thorax

Twelfth rib

Eleventh rib

Subcostal nerve
Anterior branch of T12 nerve, in series with the intercostal nerves; named subcostal as it lies under the last rib

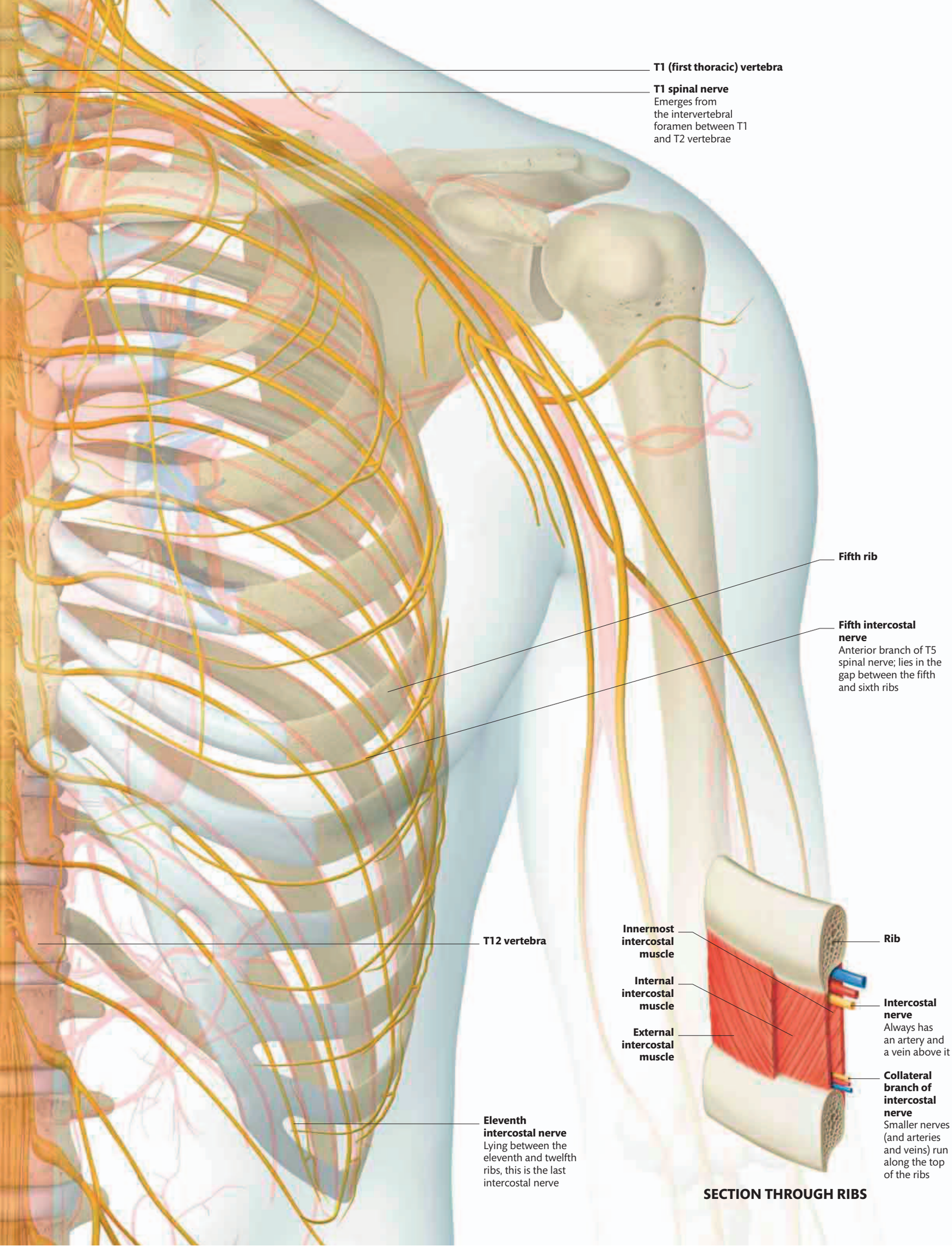

SECTION THROUGH RIBS

ABDOMEN AND PELVIS

The lower intercostal nerves continue past the lower edges of the ribcage at the front to supply the muscles and skin of the abdominal wall. The lower parts of the abdomen are supplied by the subcostal and iliohypogastric nerves. The abdominal portion of the sympathetic trunk receives nerves from the thoracic and first two lumbar spinal nerves, and sends nerves back to all the spinal nerves. The lumbar spinal nerves emerge from the spine and run into the psoas major muscle at the back of the abdomen. Inside the muscle, the nerves join up and swap fibers to form a network, or plexus. Branches of this lumbar plexus emerge around and through the psoas muscle and make their way into the thigh. Lower down, branches of the sacral plexus supply pelvic organs and enter the buttock. One of these branches, the sciatic nerve, is the largest nerve in the entire body. It supplies the back of the thigh, as well as the rest of the leg and foot.

Genitofemoral nerve
Splits into two branches: the genital branch supplies some of the scrotum or labium majus, while the femoral branch supplies a small patch of skin at the top of the thigh

Iliohypogastric nerve
Runs around the side of the lower abdomen to supply the lowest parts of the muscles and skin of the abdominal wall

Ilioinguinal nerve
Travels through the layers of the abdominal wall, then down to supply sensation in the front of the scrotum in the male, or the labia majora in the female

Femoral nerve
Supplies the front of the thigh

Sacral plexus
Nerve roots from the fourth and fifth lumbar nerves join the upper four sacral nerves to form this network. Pelvic splanchnic nerves come from the second to fourth sacral nerve roots, and convey parasympathetic nerve fibers to the pelvic organs, via the pelvic plexus on each side

Lateral cutaneous nerve of the thigh
Supplies the skin of the side of the thigh

Obturator nerve
Travels along the inside of the pelvis, then emerges through the obturator foramen to supply the inner thigh

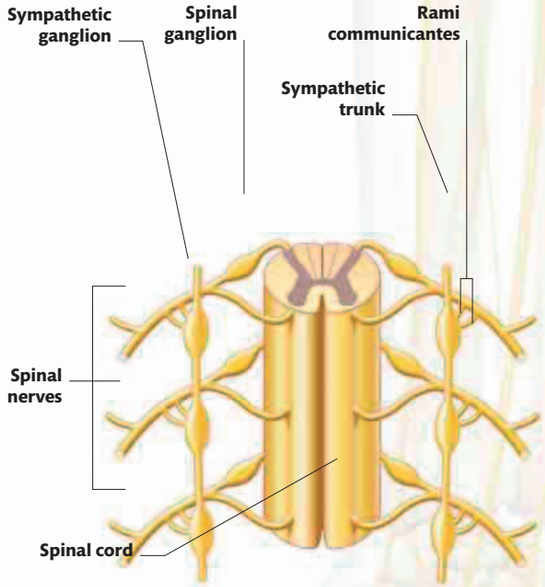

Section of sympathetic trunk and spinal cord
Branches from the sympathetic trunk innervate the organs of the abdomen and pelvis

T12 (twelfth thoracic) vertebra

Twelfth rib

Intercostal nerve

Subcostal nerve

Lumbar plexus

Iliac crest

Lumbosacral trunk
Carries nerve fibers from the fourth and fifth lumbar nerves down to join the sacral plexus

Superior gluteal nerve
Branch of the sacral plexus that supplies muscles and skin in the buttock

Anterior sacral foramen

Sciatic nerve

ANTERIOR (FRONT)

STRUCTURE OF THE SPINAL CORD

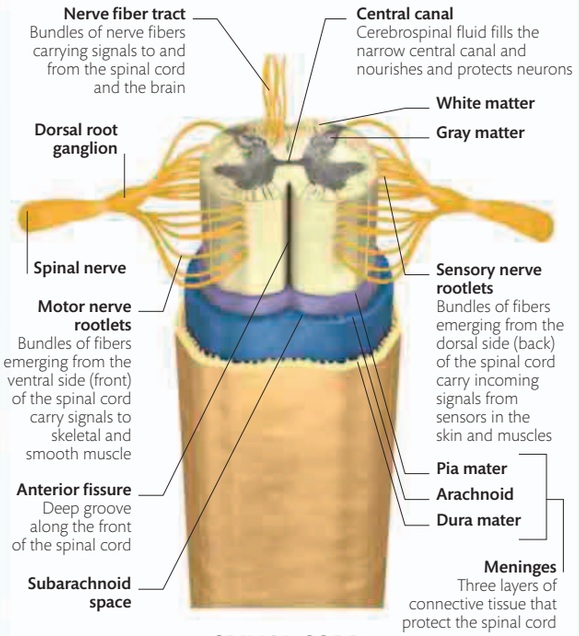

SPINAL CORD

The walls of the abdomen and pelvis are supplied by nerves emerging from the spinal cord. Like the brain, the spinal cord contains gray matter (mostly neuron cell bodies) and white matter (axons), and is covered in the same three layers of meninges: dura mater, arachnoid, and pia mater (see p.119).

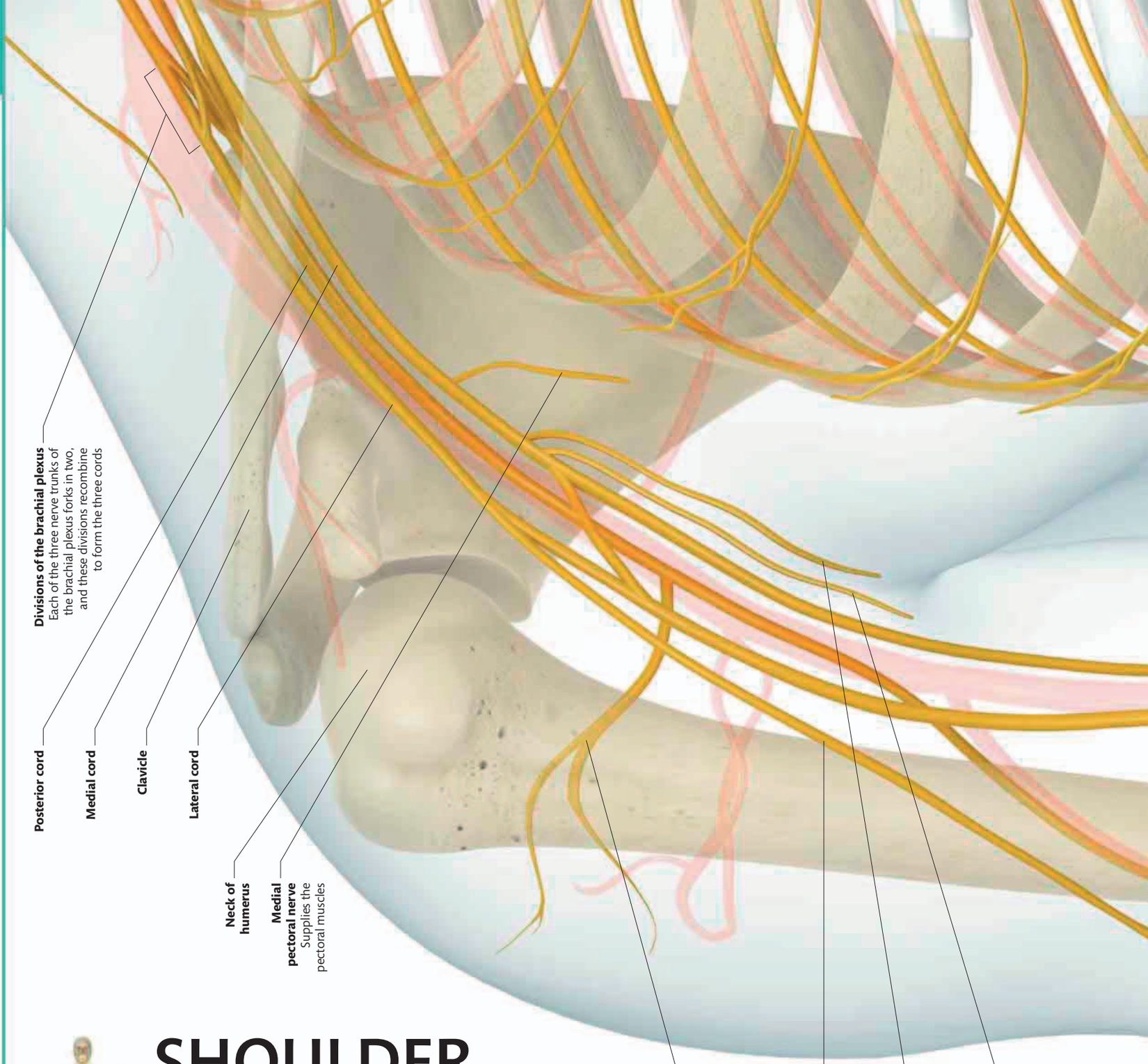

SHOULDER AND UPPER ARM

The upper limb (shoulder to hand) is supplied by five thick nerve roots that branch from the last four cervical and first thoracic spinal nerves. Emerging between the scalene muscles of the neck, they link up to form a complex skein (network) called the brachial plexus that dives under the clavicle to enter the axilla—the space between the upper arm and the chest. At this point, the plexus comprises three cords lying around the axillary artery. The network's five major nerves—musculocutaneous, median, ulnar, axillary, and radial—provide sensation to the upper limb and supply its muscles. The musculocutaneous nerve supplies the muscles in the front of the arm: the biceps, brachialis, and coracobrachialis.

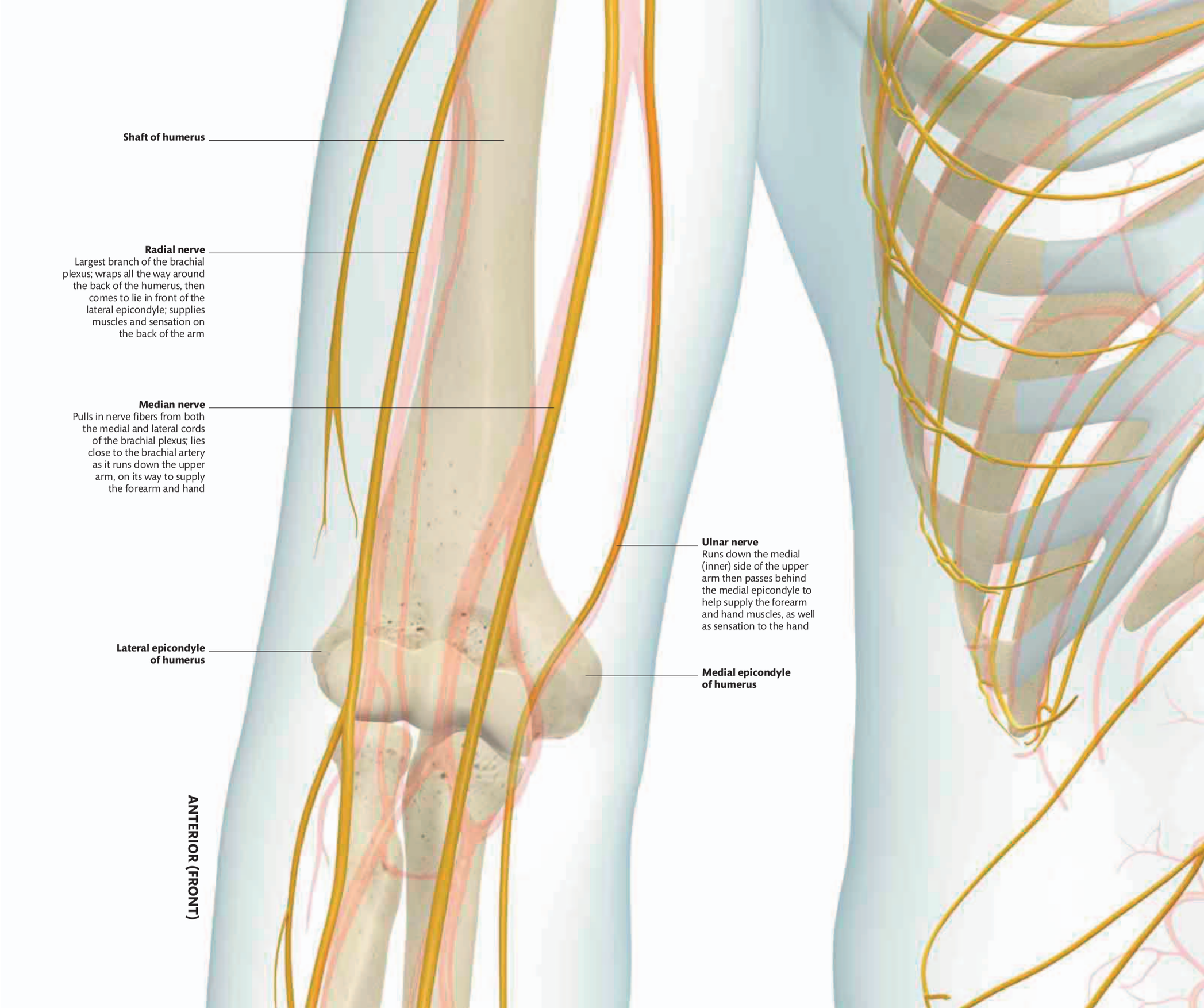
Shaft of humerus
Radial nerve
Largest branch of the brachial plexus; wraps all the way around the back of the humerus, then comes to lie in front of the lateral epicondyle; supplies muscles and sensation on the back of the arm
Median nerve
Pulls in nerve fibers from both the medial and lateral cords of the brachial plexus; lies close to the brachial artery as it runs down the upper arm, on its way to supply the forearm and hand
Ulnar nerve
Runs down the medial (inner) side of the upper arm then passes behind the medial epicondyle to help supply the forearm and hand muscles, as well as sensation to the hand
Lateral epicondyle of humerus
Medial epicondyle of humerus
ANTERIOR (FRONT)

SHOULDER AND UPPER ARM

The axillary and radial nerves emerge from the back of the brachial plexus and run behind the humerus. The axillary nerve wraps around the neck of the humerus, just underneath the shoulder joint, and supplies the deltoid muscle. The radial nerve—the largest branch of the brachial plexus—supplies all the extensor muscles in the upper arm and in the forearm. It spirals around the back of the humerus, lying right against the bone, and sends branches to supply the heads of the triceps. The radial nerve then continues in its spiral, running forward to lie just in front of the medial epicondyle of the humerus at the elbow.

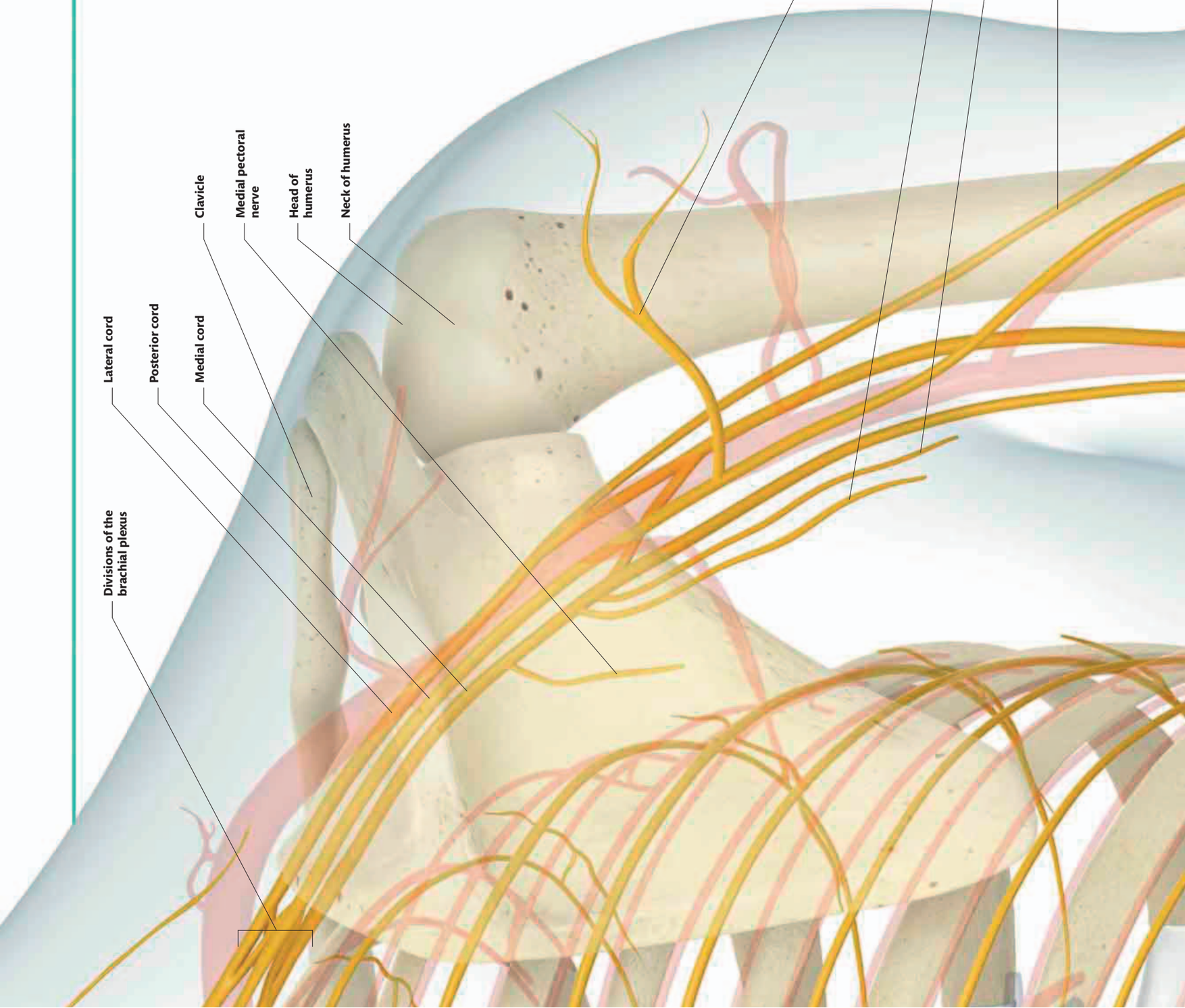

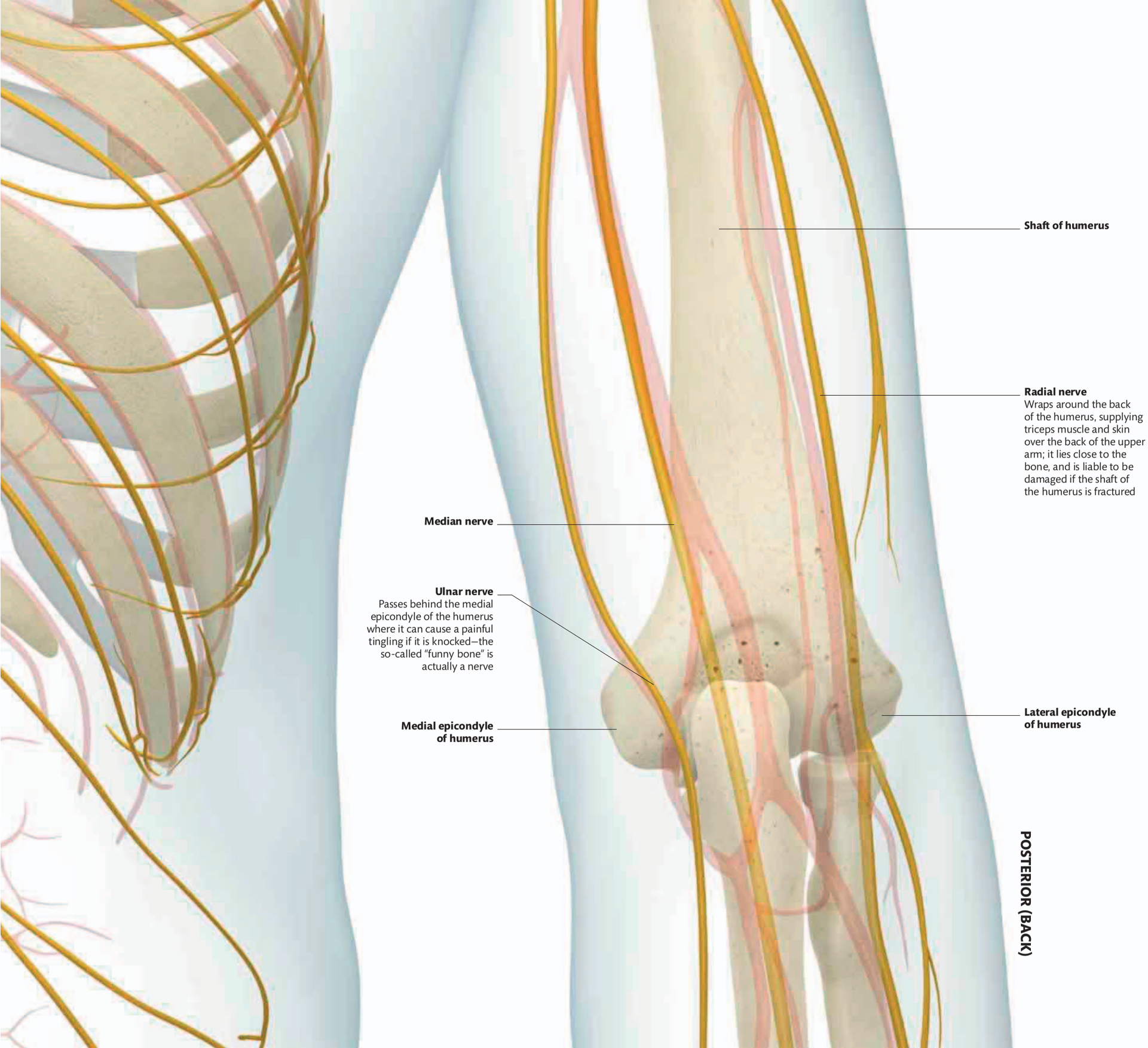
Shaft of humerus
Radial nerve
Wraps around the back of the humerus, supplying triceps muscle and skin over the back of the upper arm; it lies close to the bone, and is liable to be damaged if the shaft of the humerus is fractured
Median nerve
Ulnar nerve
Passes behind the medial epicondyle of the humerus where it can cause a painful tingling if it is knocked—the so-called "funny bone" is actually a nerve
Medial epicondyle of humerus
Lateral epicondyle of humerus
POSTERIOR (BACK)

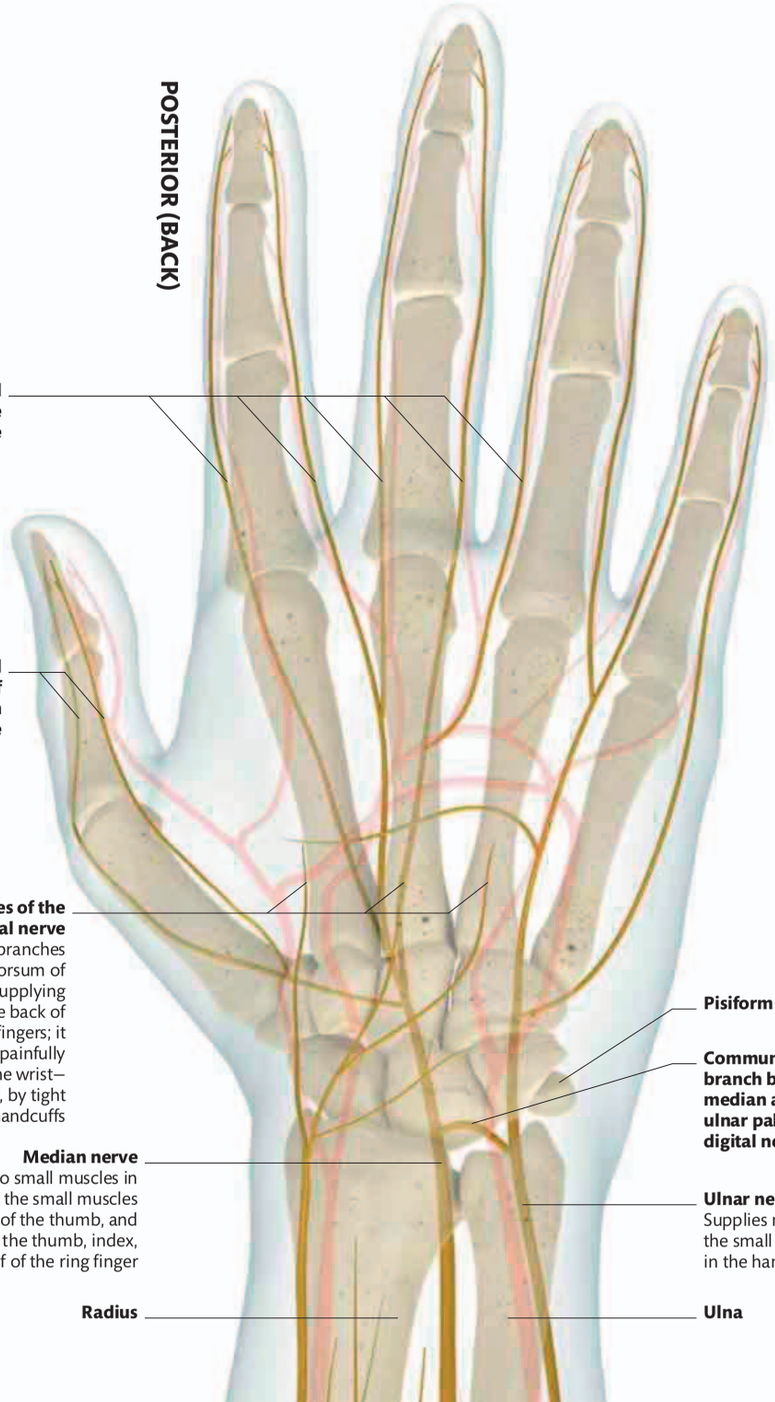
POSTERIOR (BACK)
Digital branches of the median nerve
Digital branches of the median nerve
Branches of the superficial radial nerve
This nerve branches across the dorsum of the hand, supplying sensation to the back of the hand and fingers; it can become painfully compressed at the wrist—for example, by tight bracelets or handcuffs
Median nerve
Supplies two small muscles in the hand, the small muscles in the ball of the thumb, and sensation to the thumb, index, middle, and half of the ring finger
Radius
Pisiform bone
Communicating branch between median and ulnar palmar digital nerves
Ulnar nerve
Supplies most of the small muscles in the hand
Ulna

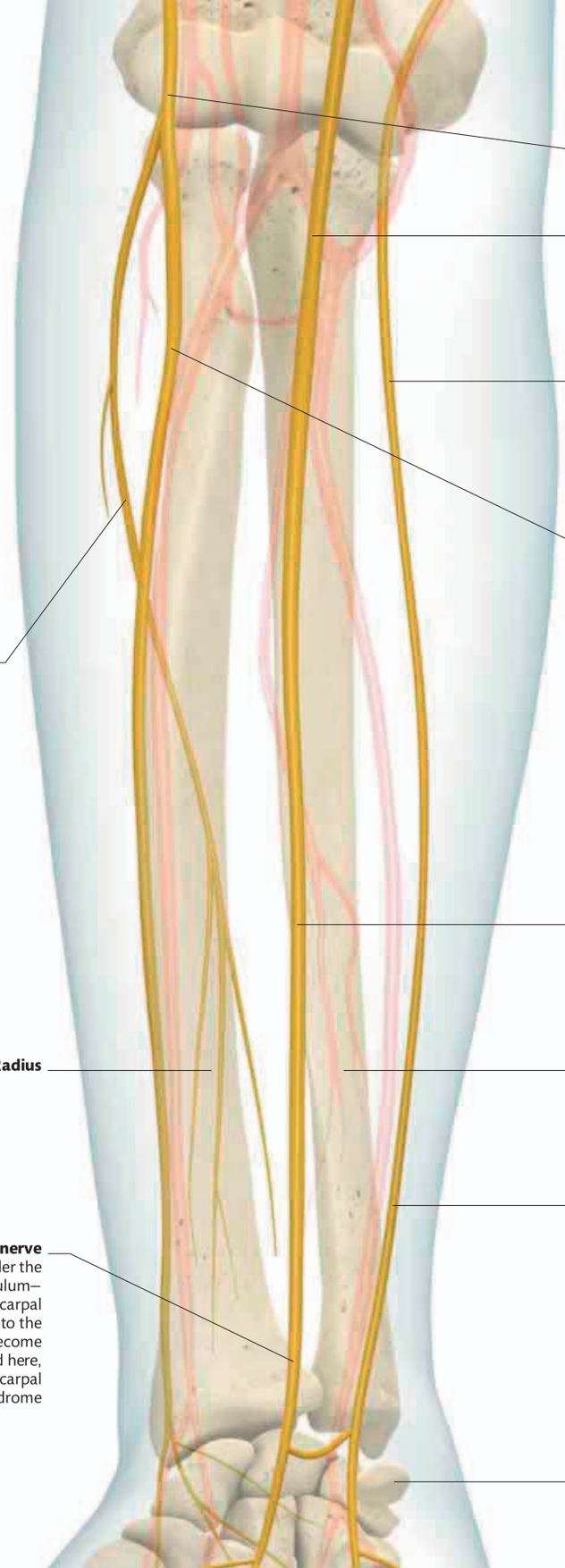
ANTERIOR (FRONT)
Radial nerve
Branches at the elbow
Median nerve
Passes through a triangular area on the inside of the elbow known as the cubital fossa
Ulnar nerve
Supplies flexor carpi ulnaris and some of flexor digitorum profundus muscles in the forearm
Superficial radial nerve
Branch of the radial nerve that runs down to the wrist, just tucked underneath brachioradialis muscle on the outer (radial) side of the forearm
Posterior interosseous nerve
Branch of the radial nerve that passes backward around the radius
Median nerve
Supplies most of the flexor muscles in the front of the forearm
Radius
Ulna
Ulnar nerve
Passes into the hand, close to the pisiform bone, where it can get compressed—for example, when using vibrating machinery or when holding a motorbike handlebar
Median nerve
Passes under the flexor retinaculum—through the carpal tunnel—into the wrist; it may become compressed here, producing carpal tunnel syndrome
Pisiform bone

LOWER ARM AND HAND

The front of the forearm is supplied by the musculocutaneous, median, and ulnar nerves. The musculocutaneous nerve supplies sensation to the lateral side of the forearm. The median nerve runs down the middle of the forearm, supplying most of the flexor muscles. It then travels over the wrist and into the hand to supply some of the thumb muscles, as well as sensation to the palm, thumb, and some fingers. The ulnar nerve courses down the inner side of the forearm, where it supplies just two muscles. It continues on to supply most of the small muscles in the hand and provide sensation to the inner side of the ring finger, and also the little finger. On the back of the forearm, the radial nerve and its branches supply all the extensor muscles. Branches of the radial nerve fan out over the back of the hand, where they provide sensation.

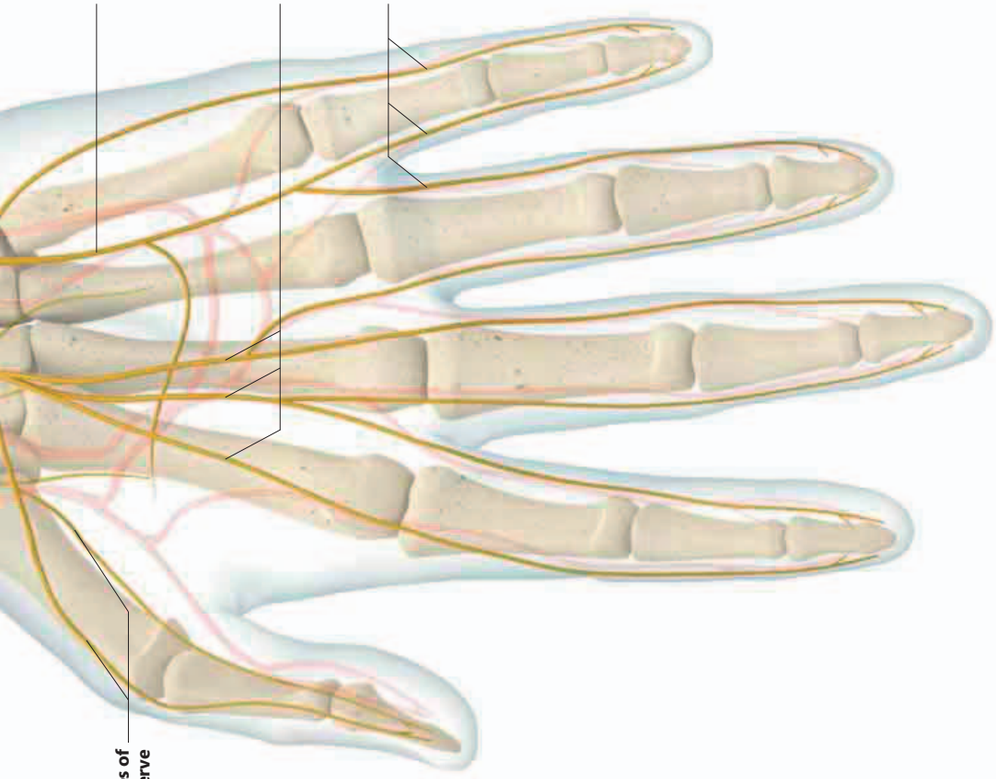

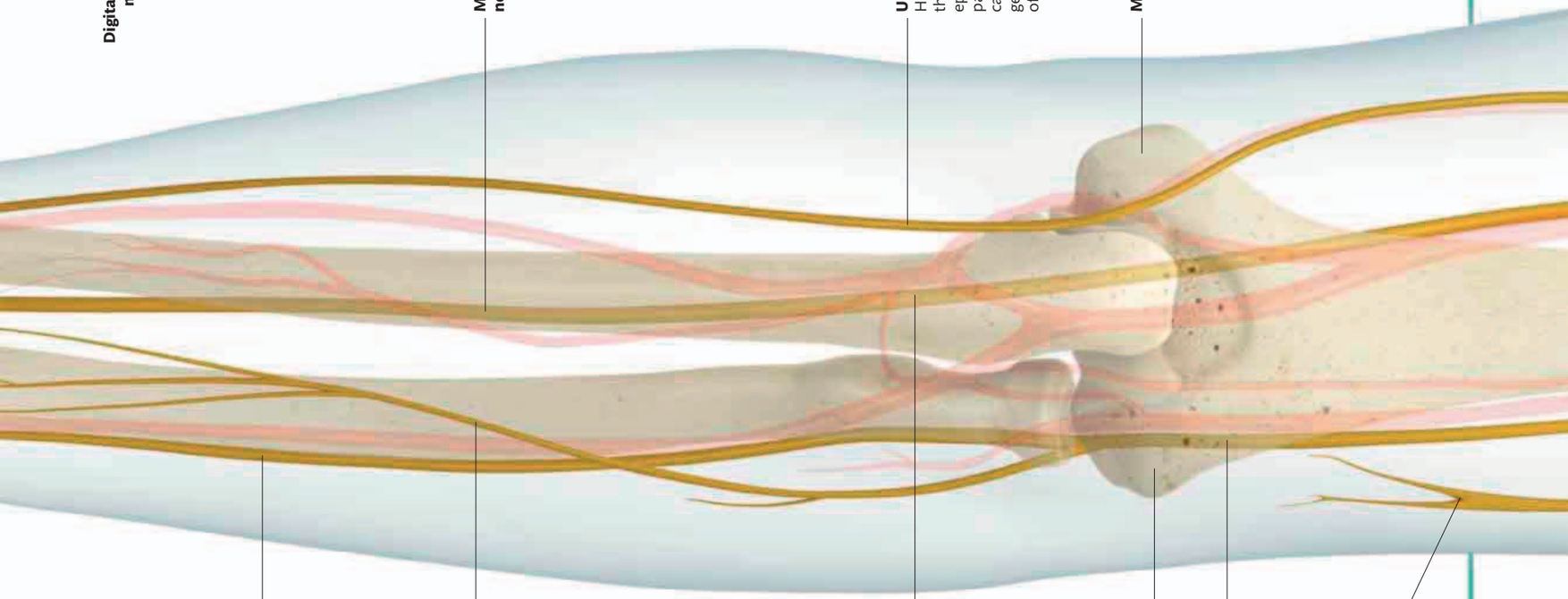

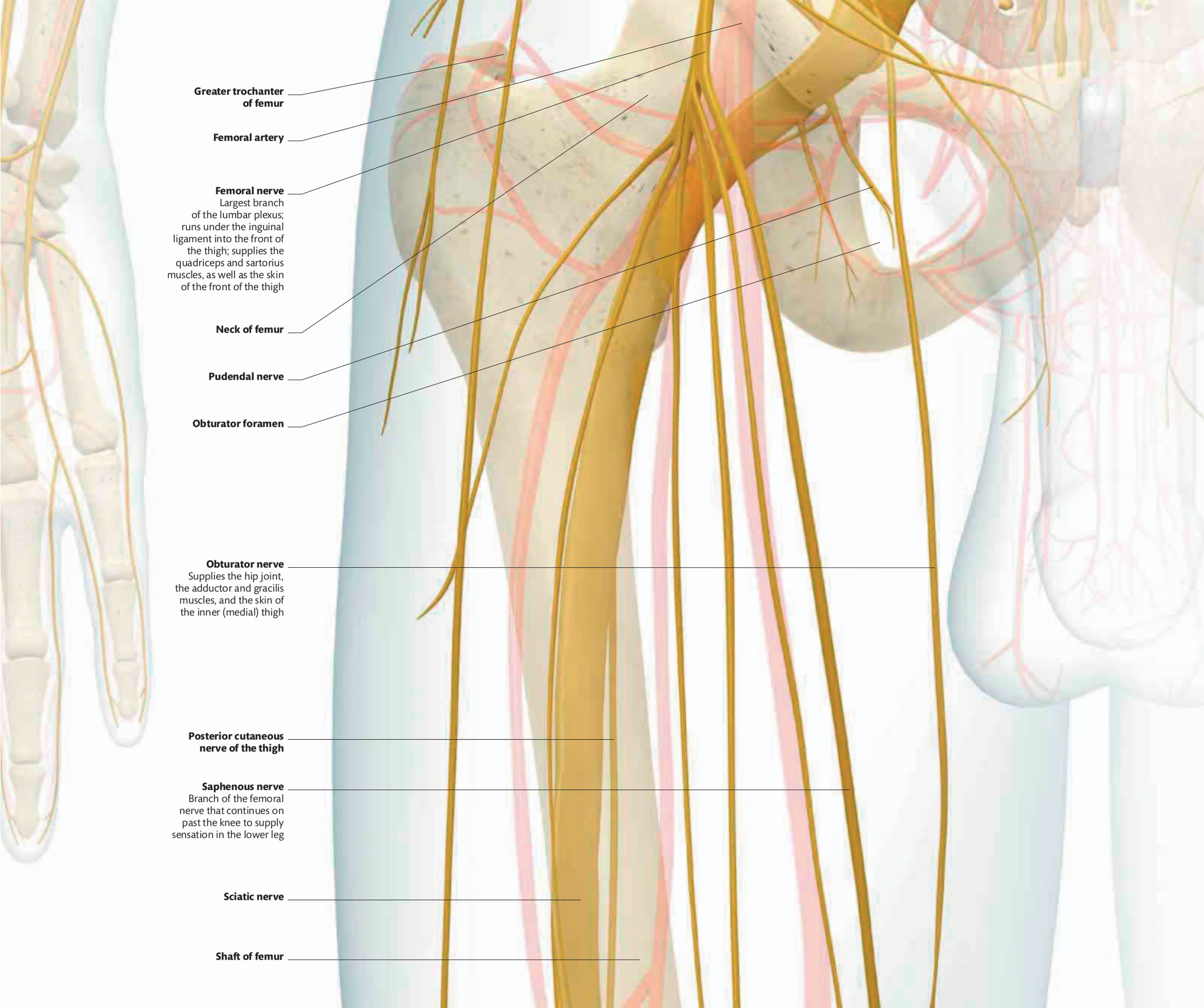
Greater trochanter of femur
Femoral artery
Femoral nerve
Largest branch of the lumbar plexus; runs under the inguinal ligament into the front of the thigh; supplies the quadriceps and sartorius muscles, as well as the skin of the front of the thigh
Neck of femur
Pudendal nerve
Obturator foramen
Obturator nerve
Supplies the hip joint, the adductor and gracilis muscles, and the skin of the inner (medial) thigh
Posterior cutaneous nerve of the thigh
Saphenous nerve
Branch of the femoral nerve that continues on past the knee to supply sensation in the lower leg
Sciatic nerve
Shaft of femur

ANTERIOR (FRONT)

Tibial nerve

Common peroneal (fibular) nerve

Patella

Tibia

Medial femoral cutaneous nerve
A branch of the femoral nerve

Intermediate femoral cutaneous nerve
Also a branch of the femoral nerve

Lateral femoral cutaneous nerve
Emerges under or through the inguinal ligament, to supply the skin of the upper, outer thigh

HIP AND THIGH

The lower limb (hip, thigh, leg, and foot) receives nerves from the lumbar and sacral plexuses. Three main nerves supply the thigh muscles: the femoral, obturator, and sciatic nerves (the last in the back). The femoral nerve runs over the pubic bone to supply the quadriceps and sartorius muscles in the front. The saphenous nerve, a slender branch of the femoral, continues past the knee and supplies skin on the inside of the lower leg and the inner side of the foot. The obturator nerve passes through the obturator foramen in the pelvic bone to supply the adductor muscles of the inner thigh and provide sensation to the skin there. Some smaller nerves just supply skin, such as the femoral cutaneous nerves.

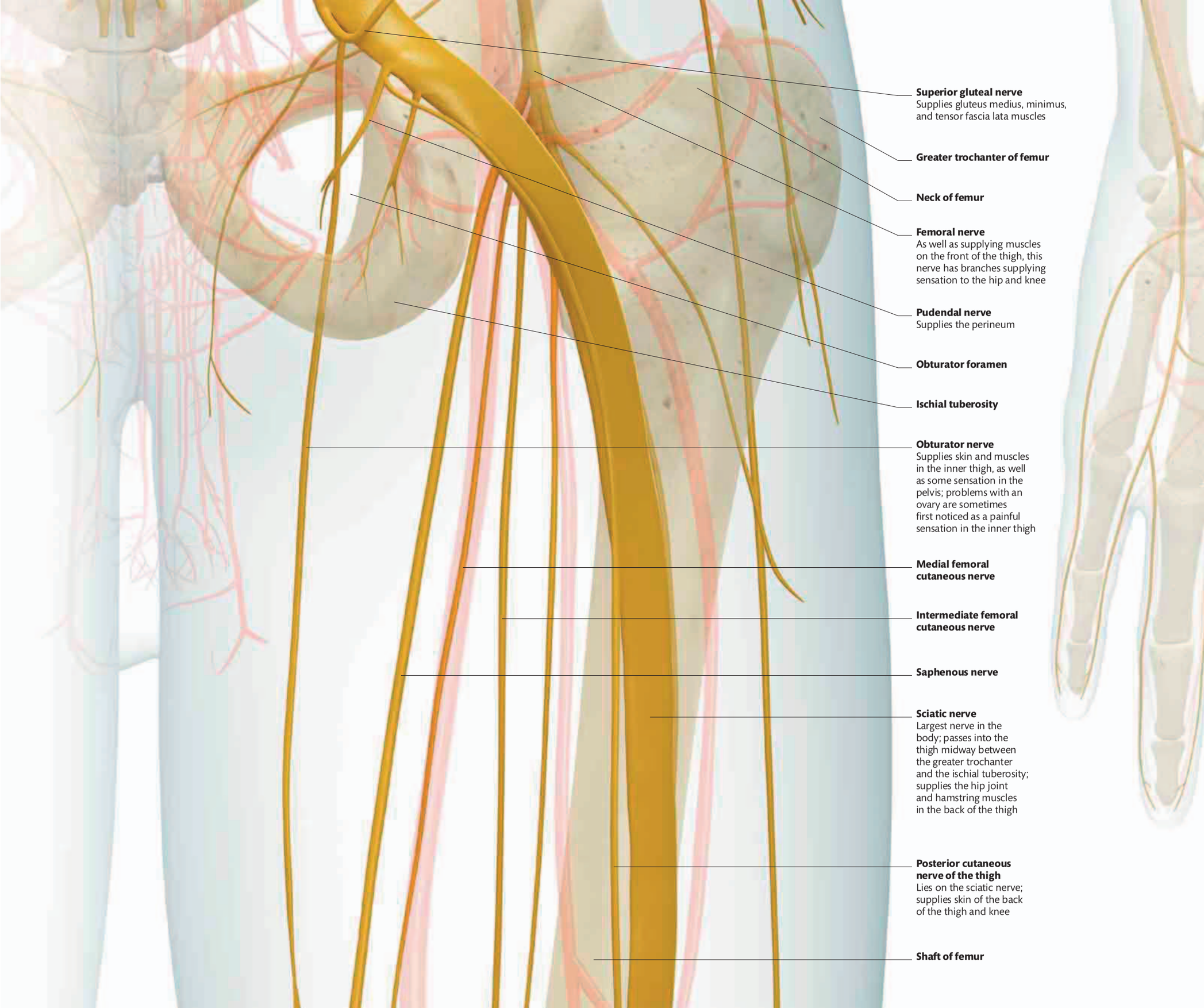
Superior gluteal nerve
Supplies gluteus medius, minimus, and tensor fascia lata muscles
Greater trochanter of femur
Neck of femur
Femoral nerve
As well as supplying muscles on the front of the thigh, this nerve has branches supplying sensation to the hip and knee
Pudendal nerve
Supplies the perineum
Obturator foramen
Ischial tuberosity
Obturator nerve
Supplies skin and muscles in the inner thigh, as well as some sensation in the pelvis; problems with an ovary are sometimes first noticed as a painful sensation in the inner thigh
Medial femoral cutaneous nerve
Intermediate femoral cutaneous nerve
Saphenous nerve
Sciatic nerve
Largest nerve in the body; passes into the thigh midway between the greater trochanter and the ischial tuberosity; supplies the hip joint and hamstring muscles in the back of the thigh
Posterior cutaneous nerve of the thigh
Lies on the sciatic nerve; supplies skin of the back of the thigh and knee
Shaft of femur

HIP AND THIGH

Gluteal nerves from the sacral plexus emerge via the greater sciatic foramen, at the back of the pelvis, to supply the muscles and skin of the buttock. The sciatic nerve also emerges through the greater sciatic foramen into the buttock. The gluteus maximus is a good site for injections into a muscle, but these should always be given in the upper, outer part of the buttock to make sure the needle is well away from the sciatic nerve. The sciatic nerve runs down the back of the thigh, supplying the hamstrings. In most people, the sciatic nerve runs halfway down the thigh then splits into two branches, the tibial and common peroneal nerves. These continue into the popliteal fossa (back of the knee) and on into the lower leg.

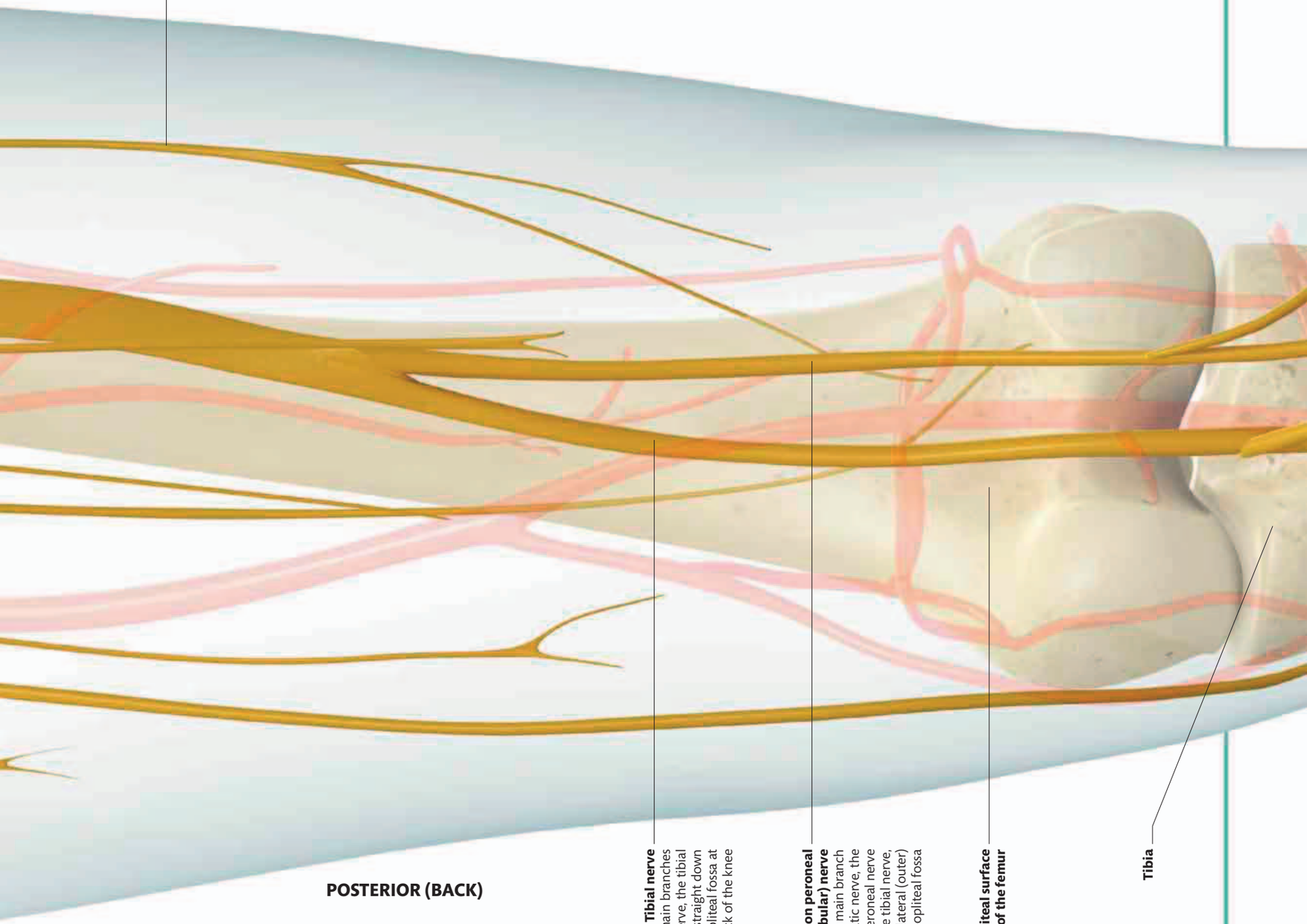

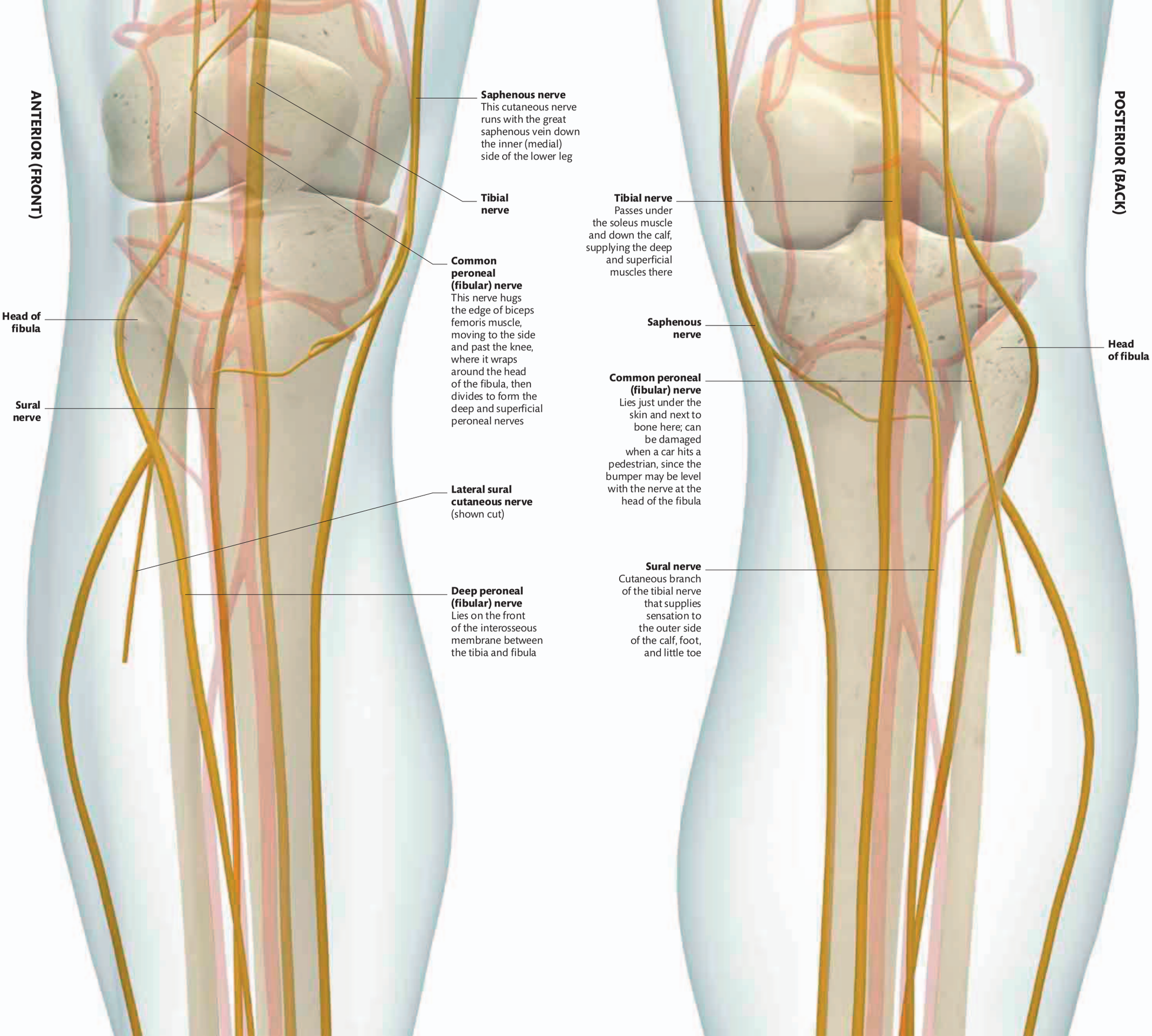
ANTERIOR (FRONT)
Saphenous nerve
This cutaneous nerve runs with the great saphenous vein down the inner (medial) side of the lower leg
Tibial nerve
Common peroneal (fibular) nerve
This nerve hugs the edge of biceps femoris muscle, moving to the side and past the knee, where it wraps around the head of the fibula, then divides to form the deep and superficial peroneal nerves
Head of fibula
Sural nerve
Lateral sural cutaneous nerve
(shown cut)
Deep peroneal (fibular) nerve
Lies on the front of the interosseous membrane between the tibia and fibula
POSTERIOR (BACK)
Tibial nerve
Passes under the soleus muscle and down the calf, supplying the deep and superficial muscles there
Saphenous nerve
Head of fibula
Common peroneal (fibular) nerve
Lies just under the skin and next to bone here; can be damaged when a car hits a pedestrian, since the bumper may be level with the nerve at the head of the fibula
Sural nerve
Cutaneous branch of the tibial nerve that supplies sensation to the outer side of the calf, foot, and little toe

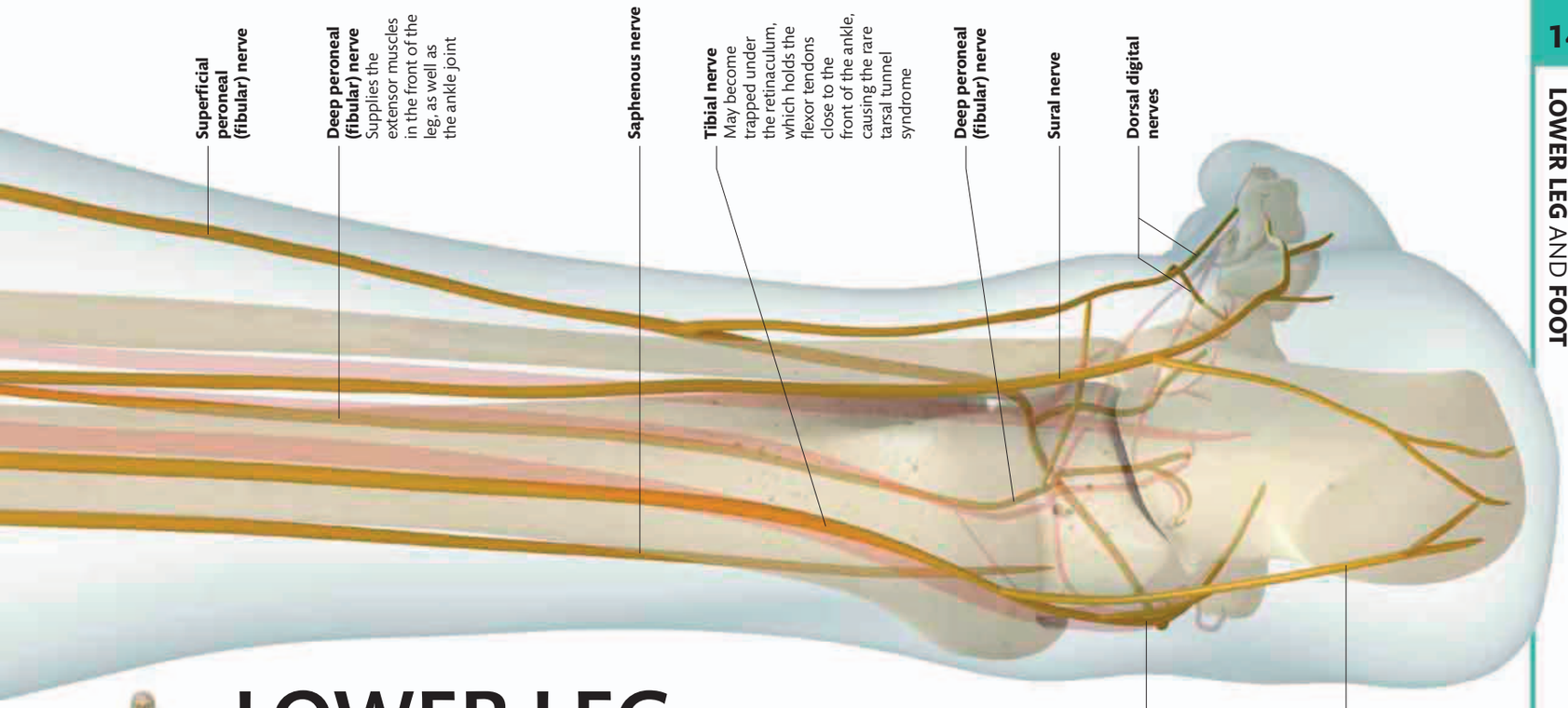

LOWER LEG AND FOOT

The common peroneal nerve runs past the knee and wraps around the neck of the fibula. Then it splits into the deep and superficial peroneal nerves. The deep peroneal nerve supplies the extensor muscles of the shin, then fans out to provide sensation to the skin at the back of the foot. The superficial peroneal nerve stays on the side of the leg and supplies the peroneal muscles. The tibial nerve runs through the popliteal fossa (back of the knee), under the soleus muscle, and between the deep and superficial calf muscles, which it supplies. It continues behind the medial malleolus and under the foot, then splits into two plantar nerves that supply the small muscles of the foot and the skin of the sole.

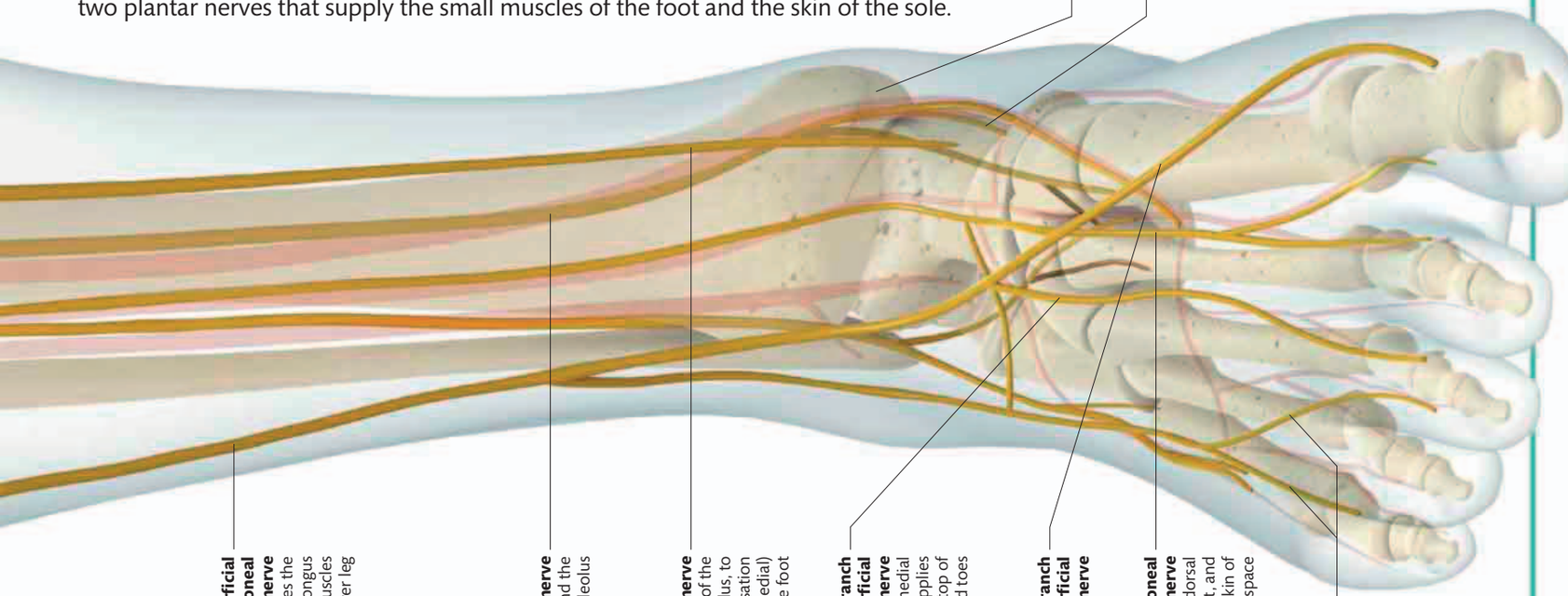

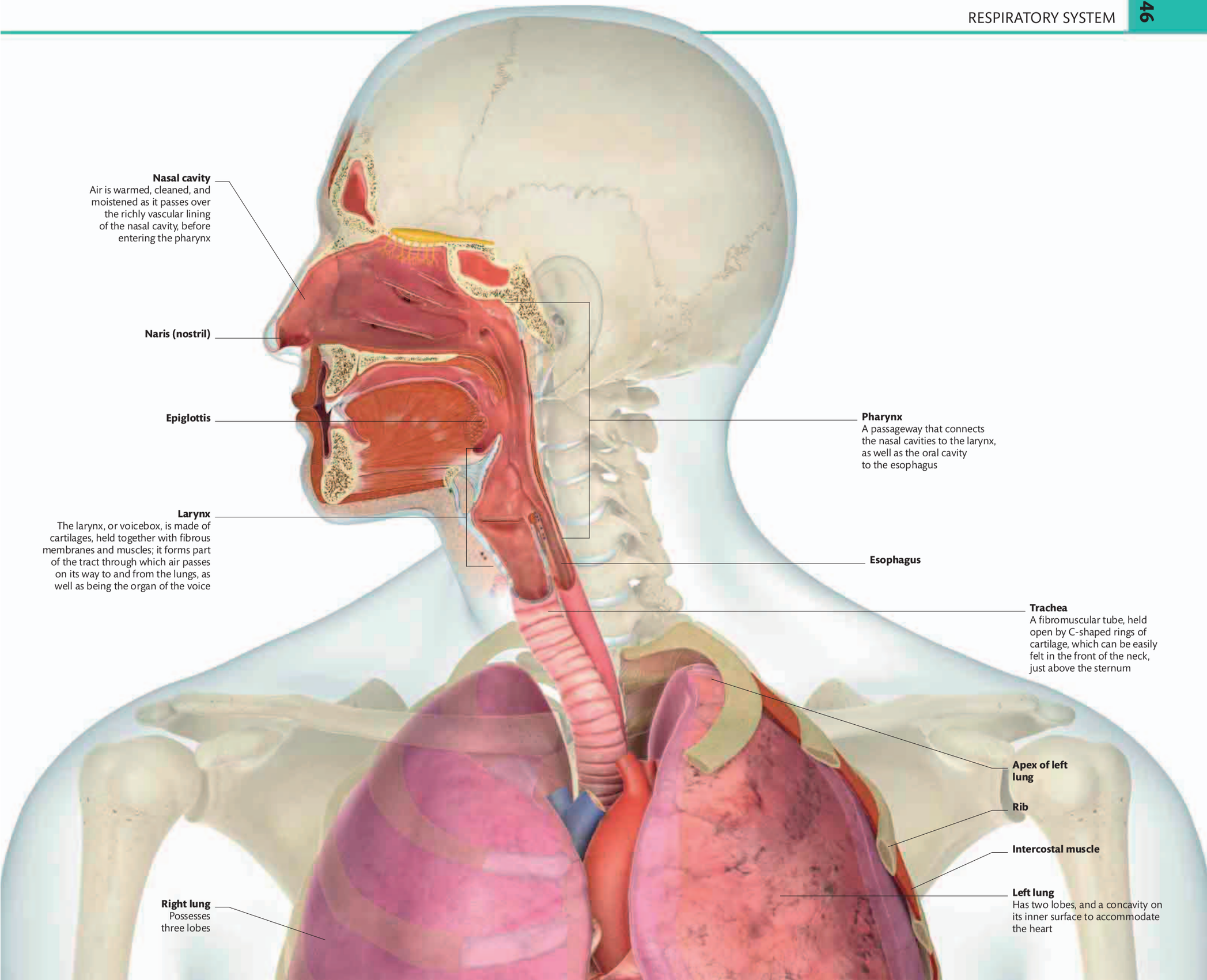
Nasal cavity
Air is warmed, cleaned, and moistened as it passes over the richly vascular lining of the nasal cavity, before entering the pharynx
Naris (nostril)
Epiglottis
Larynx
The larynx, or voicebox, is made of cartilages, held together with fibrous membranes and muscles; it forms part of the tract through which air passes on its way to and from the lungs, as well as being the organ of the voice
Pharynx
A passageway that connects the nasal cavities to the larynx, as well as the oral cavity to the esophagus
Esophagus
Trachea
A fibromuscular tube, held open by C-shaped rings of cartilage, which can be easily felt in the front of the neck, just above the sternum
Apex of left lung
Rib
Intercostal muscle
Left lung
Has two lobes, and a concavity on its inner surface to accommodate the heart
Right lung
Possesses three lobes

RESPIRATORY SYSTEM **OVERVIEW**

Every cell in the human body needs to get oxygen, and to get rid of carbon dioxide. These gases are transported around the body in the blood, but the actual transfer of gases between the air and the blood occurs in the lungs. The lungs have extremely thin membranes that allow the gases to pass across easily. But air also needs to be regularly drawn in and out of the lungs, to expel the building carbon dioxide and to bring in fresh oxygen, and this is brought about by respiration—commonly called breathing. The respiratory system includes the airways on the way to the lungs: the nasal cavities, parts of the pharynx, the larynx, the trachea, and the bronchi (see p.149).

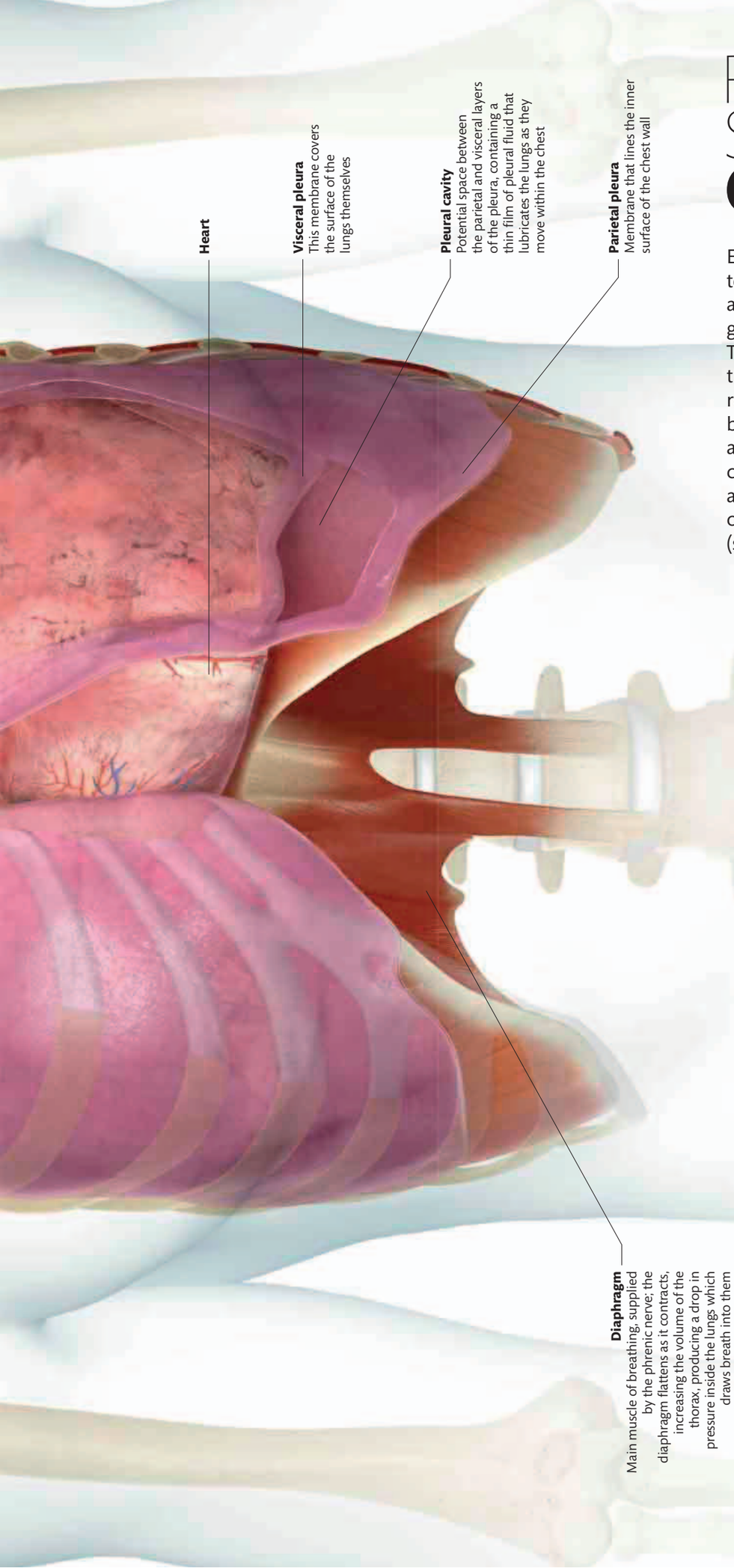

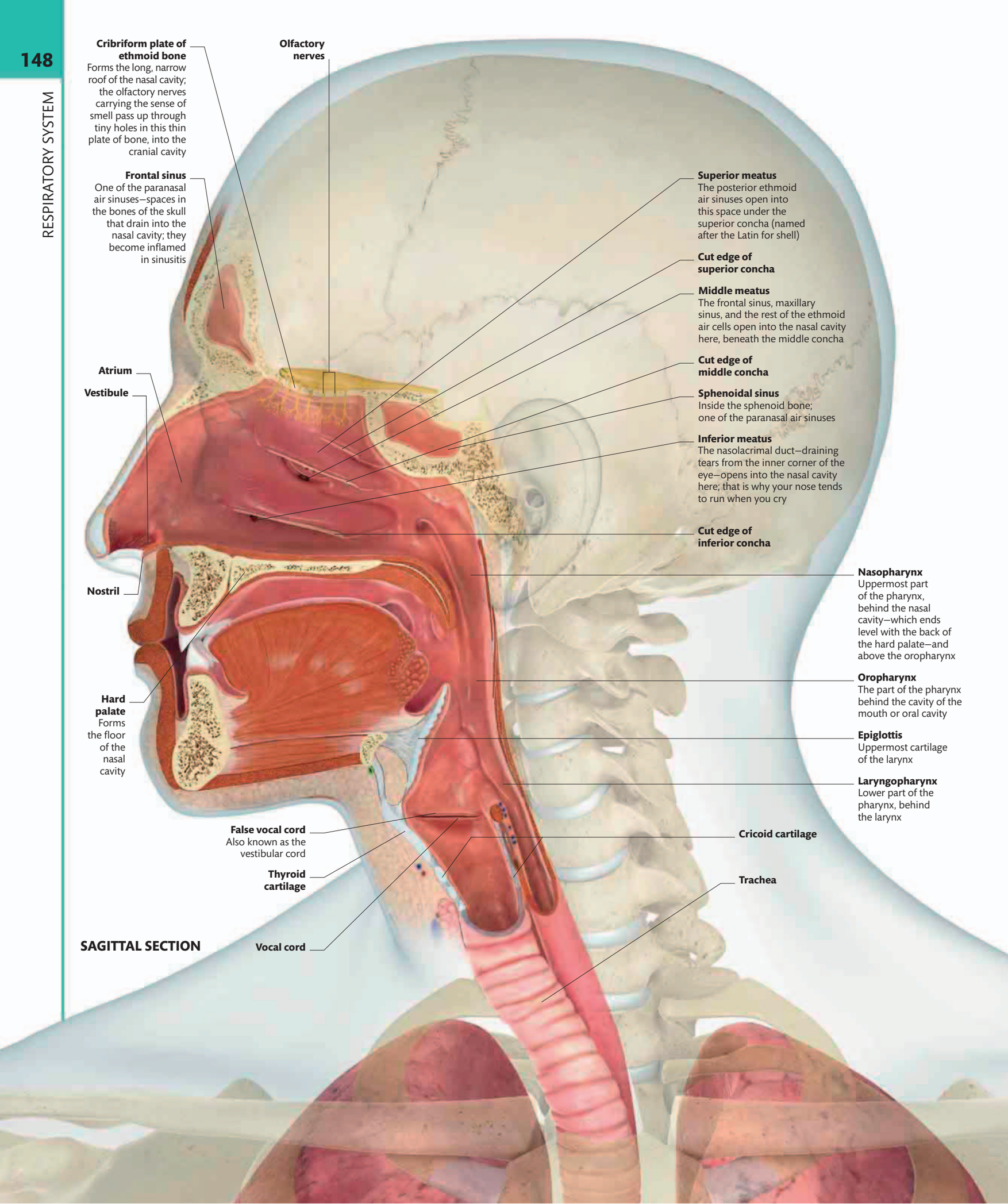

SAGITTAL SECTION

HEAD AND NECK

When we take a breath, air is pulled in through our nostrils, into the nasal cavities. Here the air is cleaned, warmed, and moistened before its onward journey. The nasal cavities are divided by the thin partition of the nasal septum, which is composed of plates of cartilage and bone. The lateral walls of the nasal cavity are more elaborate, with bony curls (conchae) that increase the surface area over which the air flows. The nasal cavity is lined with mucosa, which produces mucus. This often undervalued substance does an important job of trapping particles and moistening the air. The nasal sinuses, also lined with mucosa, open via tiny orifices into the nasal cavity. Below and in front of the pharynx is the larynx—the organ of speech. The way that air passes through this can be modulated to produce sound.

X-RAY OF HEAD SHOWING SINUSES

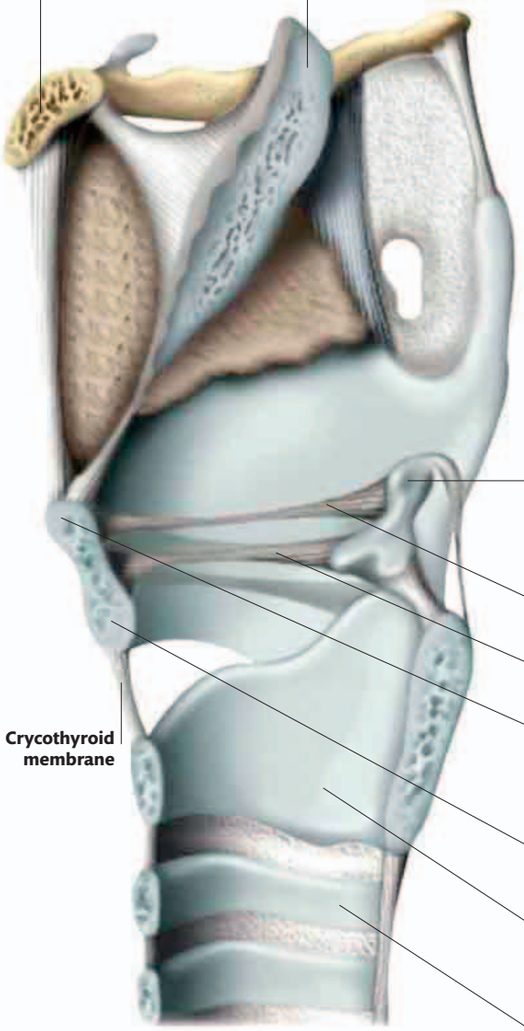

LARYNX

THORAX

The trachea, commonly known as the windpipe, passes from the neck into the thorax, where it divides into two airways called bronchi—each supplying one lung. The trachea is supported and held open by 15–20 C-shaped pieces of cartilage, and there is smooth muscle in its wall that can alter the width of the trachea. Cartilage in the walls of the bronchi prevents them from collapsing when air enters the lungs under low pressure. Inside the lungs, the bronchi branch and branch again, forming smaller airways called bronchioles; the bronchioles are just muscular tubes, completely lacking in cartilage. The smallest bronchioles end in a cluster of alveoli, these are air sacs surrounded by capillaries, where oxygen passes from the air into the blood, and carbon dioxide passes in the opposite direction.

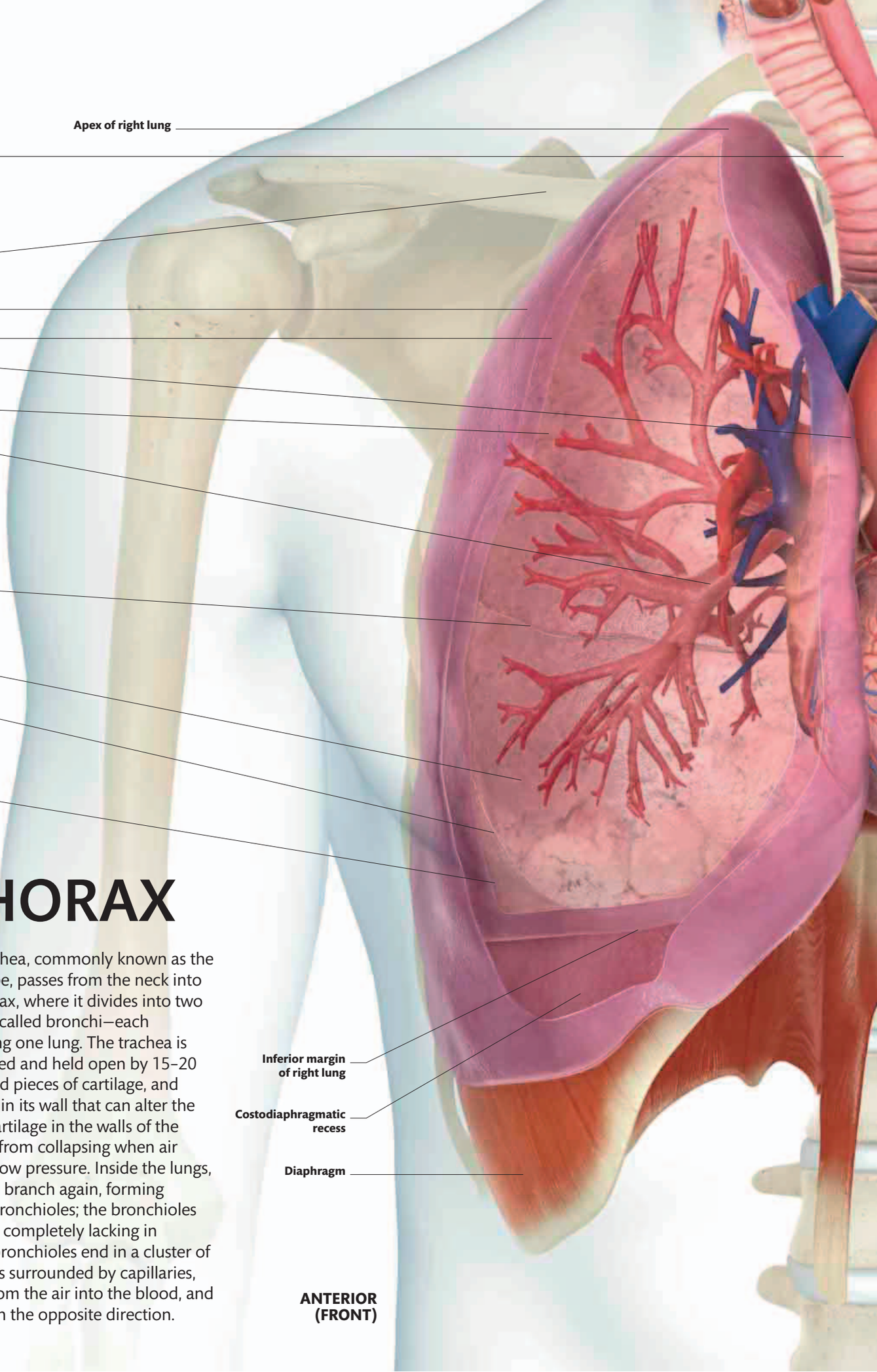

Apex of left lung
The apex, or topmost point, of each lung projects some $^{3}/_{4}$in (2cm) above the clavicle

Left clavicle (cut away to show lung behind)

Bronchus of left lung
Bronchi are lined with epithelium, which produces mucus to trap particles, and carpeted with tiny hairlike projections called cilia that waft mucus up and out of the lungs

Superior lobe of left lung

Anterior margin of left lung

Cardiac notch of left lung
Anterior edge of the left lung that curves inward slightly to accommodate the heart

Oblique fissure of left lung
Divides the superior and inferior lobes of the left lung

Bronchiole

Inferior lobe of left lung

Inferior margin of left lung

Lingula
Slight projection of the front edge of the left lung; name originates from the Latin for little tongue

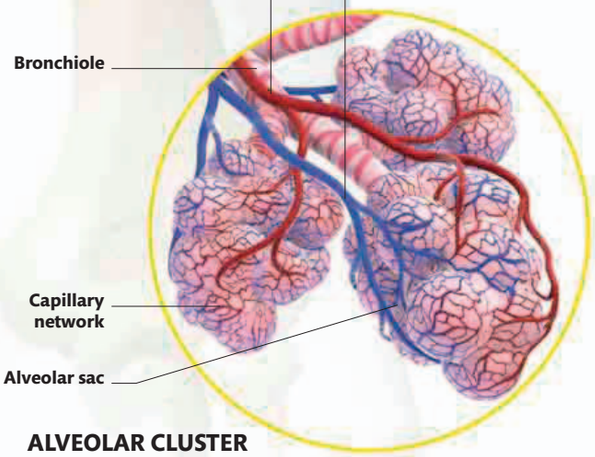

ALVEOLAR CLUSTER

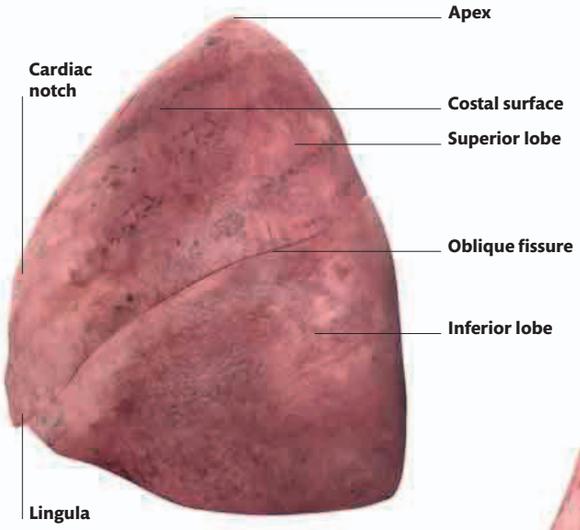

LEFT LUNG (LATERAL VIEW)

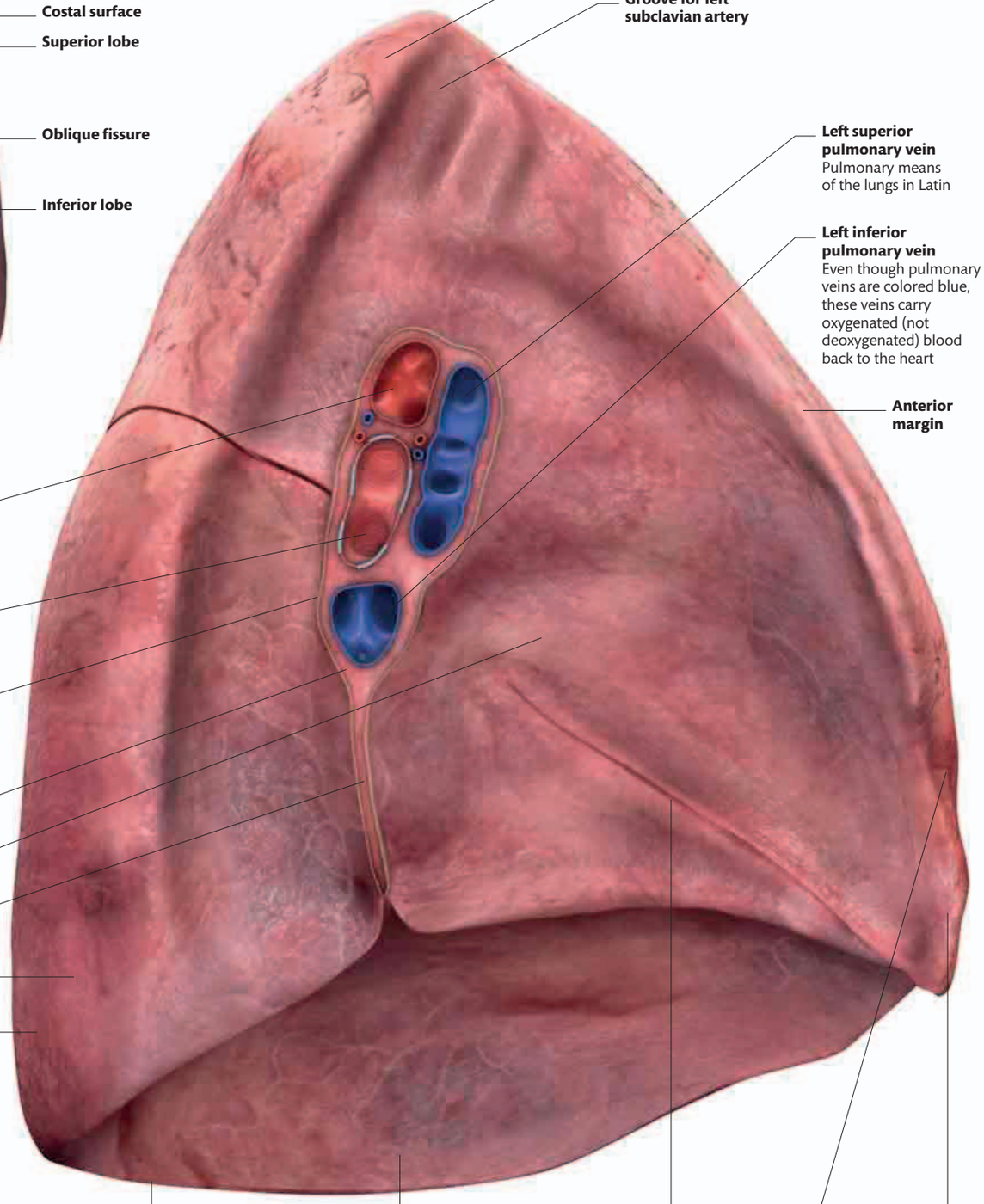

LEFT LUNG (MEDIAL VIEW)

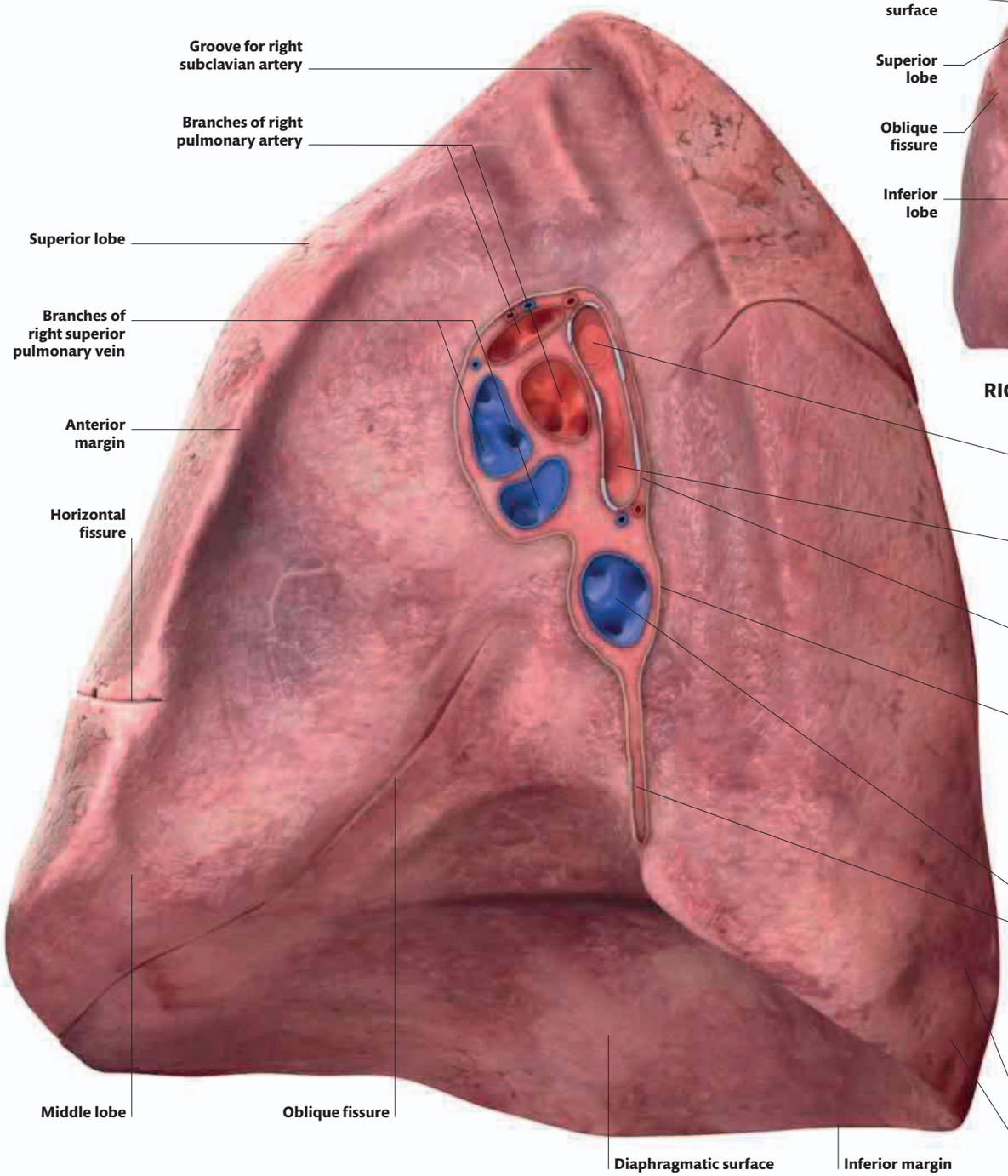

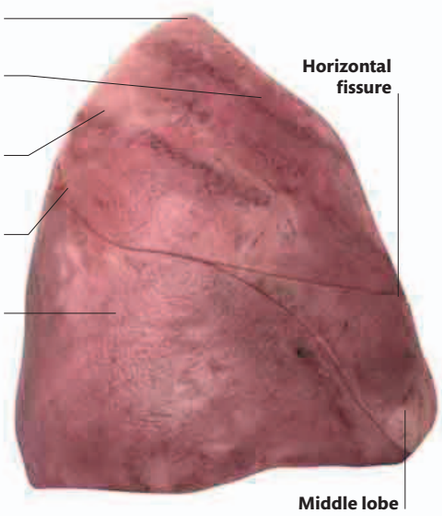

RIGHT LUNG (LATERAL VIEW)

Superior lobar bronchus
The structure of the wall of the bronchus, containing rings of cartilage, can be seen in cross section here

Right main bronchus
Having already given off the superior lobar bronchus, this will branch to form the middle and inferior lobar bronchi

Hilum
Depression where the bronchi and major pulmonary vessels enter and leave the lung

Pleura
Around the hilum, the visceral pleura doubles back on itself to form the parietal pleura, which continues on to the inner surface of the ribcage and the upper surface of the diaphragm; the visceral and parietal pleura thus form a continuous, closed envelope around the lung

Right inferior pulmonary vein

Pulmonary ligament
The line where the visceral pleura doubles back on itself to become the parietal pleura hangs down below the hilum, forming this "ligament", which allows the pulmonary veins to expand freely

Costal surface

Inferior lobe

RIGHT LUNG (MEDIAL VIEW)

LUNGS

Each lung fits snugly inside its half of the thoracic cavity. The surface of each lung is covered with a thin pleural membrane (visceral pleura), and the inside of the chest wall is also lined with pleura (parietal pleura). Between the two pleural layers lies a thin film of lubricating fluid that allows the lungs to slide against the chest wall during breathing movements, but it also creates a fluid seal, effectively sticking the lungs to the ribs and the diaphragm. Because of this seal, when you inhale, the lungs are pulled outward in all directions, and air rushes into them. The bronchi and blood vessels enter each lung at the hilum on its inner or medial surface. Although the two lungs may appear to be similar at first glance, there is some asymmetry. The left lung is concave to fit around the heart and has only two lobes, whereas the right lung has three lobes, marked out by two deep fissures.

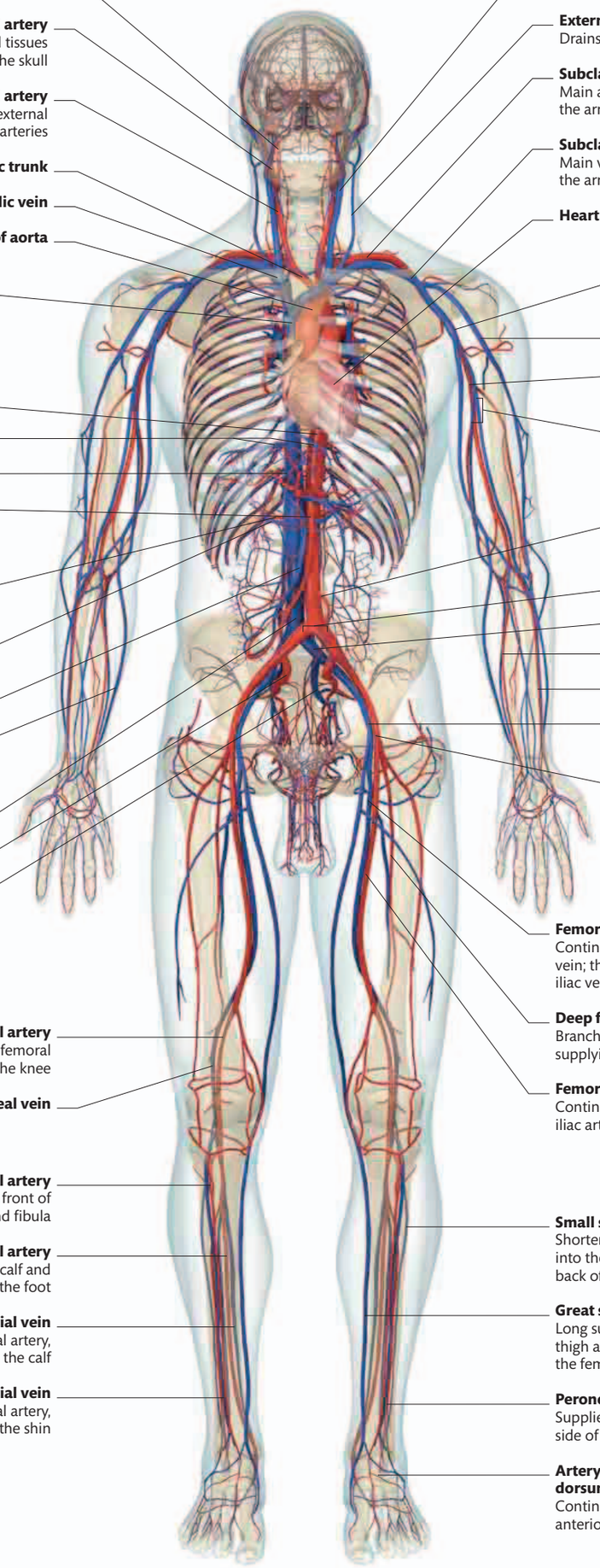

ANTERIOR (FRONT)

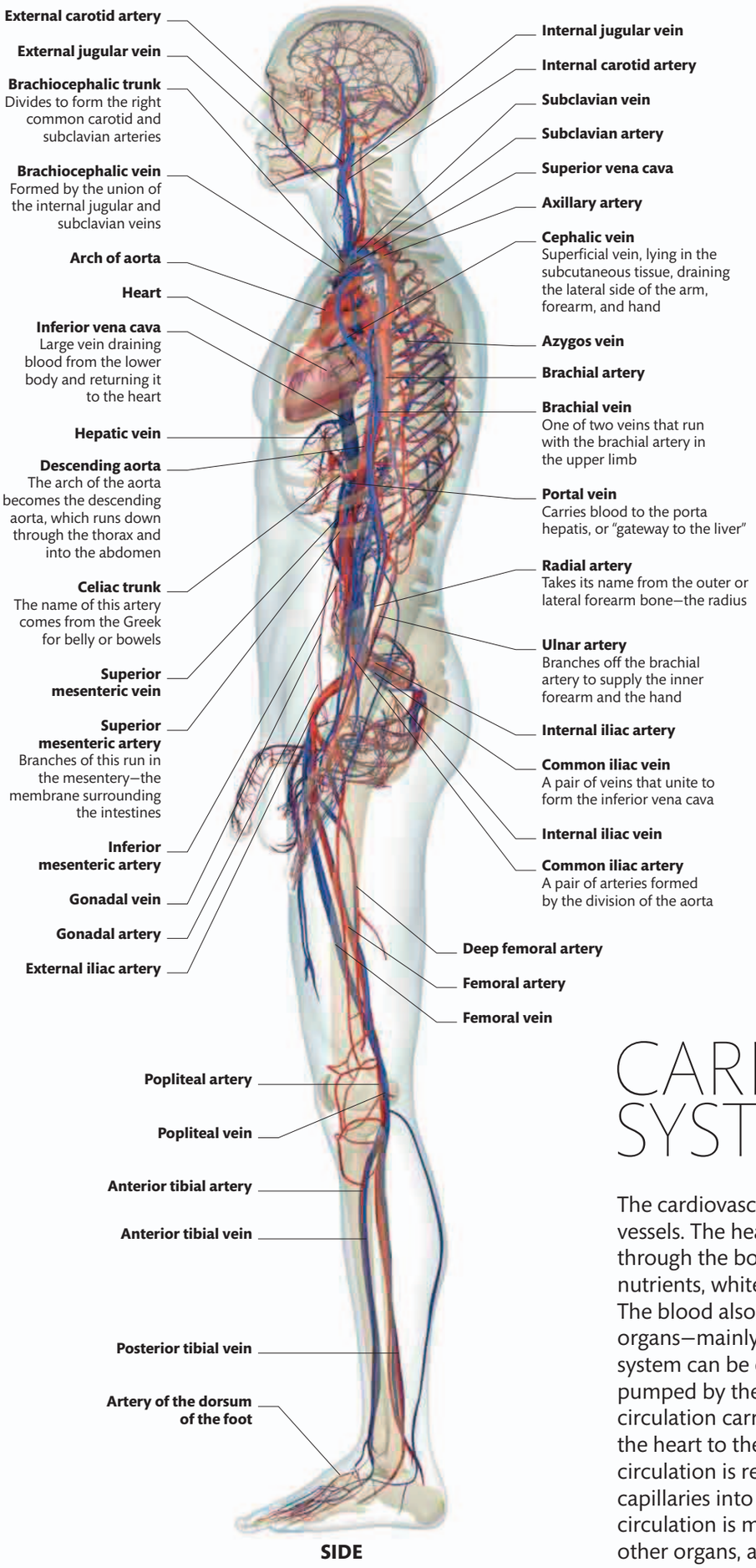

SIDE

BLOOD VESSELS

The heart contracts to keep blood moving through a vast network of blood vessels—arteries, arterioles, capillaries, venules, and veins. A thick elastic wall in arteries helps them to carry high-pressure blood from the heart to organs and tissues; veins contain valves that prevent backflow of blood when carrying it back to the heart. Arteries and smaller vessels branch into capillaries—the smallest blood vessels. The endothelial wall of a capillary is one cell thick.

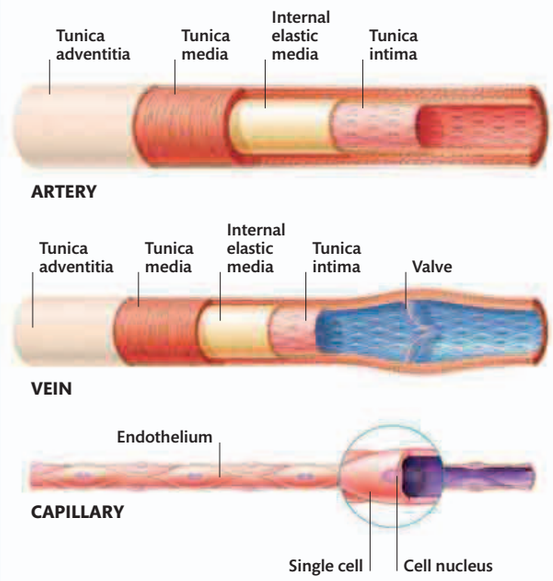

CARDIOVASCULAR SYSTEM **OVERVIEW**

The cardiovascular system consists of the heart, blood, and blood vessels. The heart—a muscular pump—contracts to push blood through the body's network of vessels in order to deliver oxygen, nutrients, white blood cells, and hormones to the tissues of the body. The blood also removes waste products and takes them to other organs—mainly the liver and kidneys—for excretion. The circulatory system can be divided in two: pulmonary circulation carries blood pumped by the right side of the heart to the lungs, and systemic circulation carries blood pumped by the more powerful left side of the heart to the rest of the body. Pressure in the pulmonary circulation is relatively low, to prevent fluid being forced out of capillaries into the alveoli of the lungs. Pressure in the systemic circulation is much higher, to push blood up to the brain, into all other organs, and out into the fingers and toes.

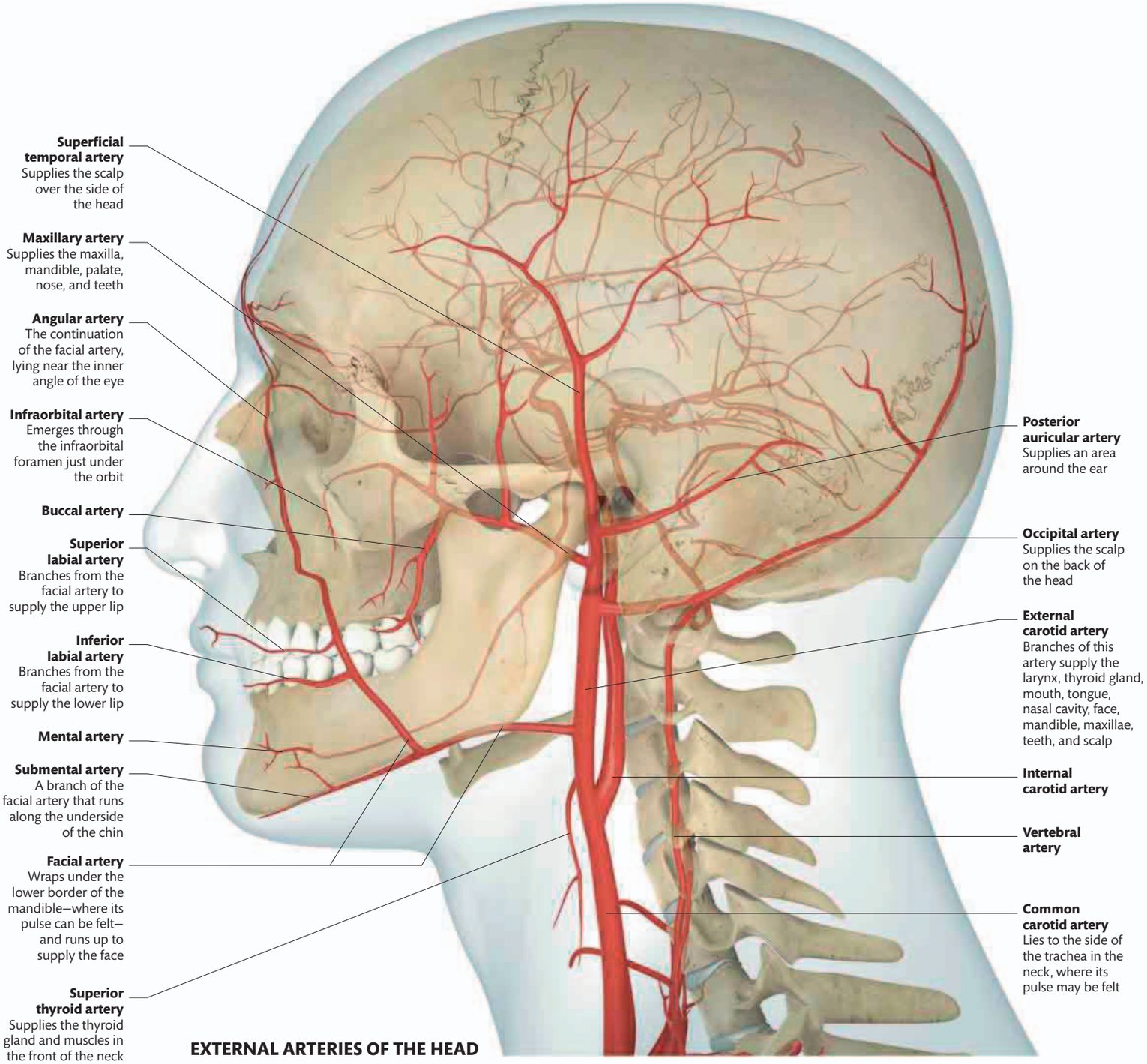

EXTERNAL ARTERIES OF THE HEAD

HEAD AND NECK

The main vessels supplying oxygenated blood to the head and neck are the common carotid and vertebral arteries. The vertebral artery runs up through holes in the cervical vertebrae and eventually enters the skull through the foramen magnum. The common carotid artery runs up the neck and divides in two—the internal carotid artery supplies the brain, and the external carotid artery gives rise to a profusion of branches, some of which supply the thyroid gland, the mouth, tongue, and nasal cavity. Veins of the head and neck come together like river tributaries, draining into the large internal jugular vein, behind the sternocleidomastoid muscle, and into the subclavian vein, low in the neck.

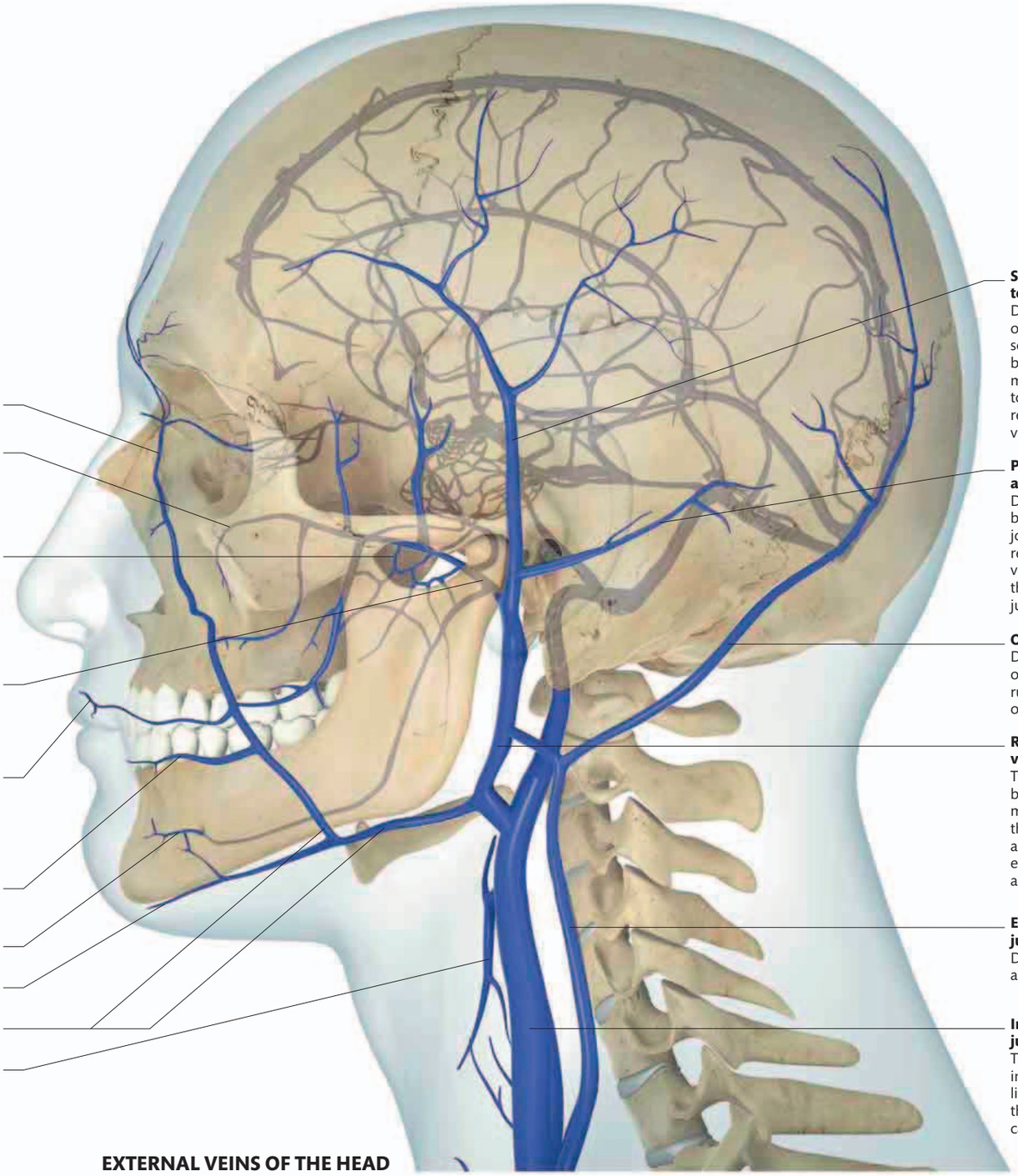

EXTERNAL VEINS OF THE HEAD

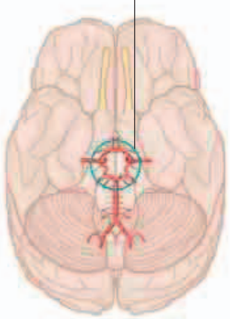

LOCATION OF CIRCLE OF WILLIS

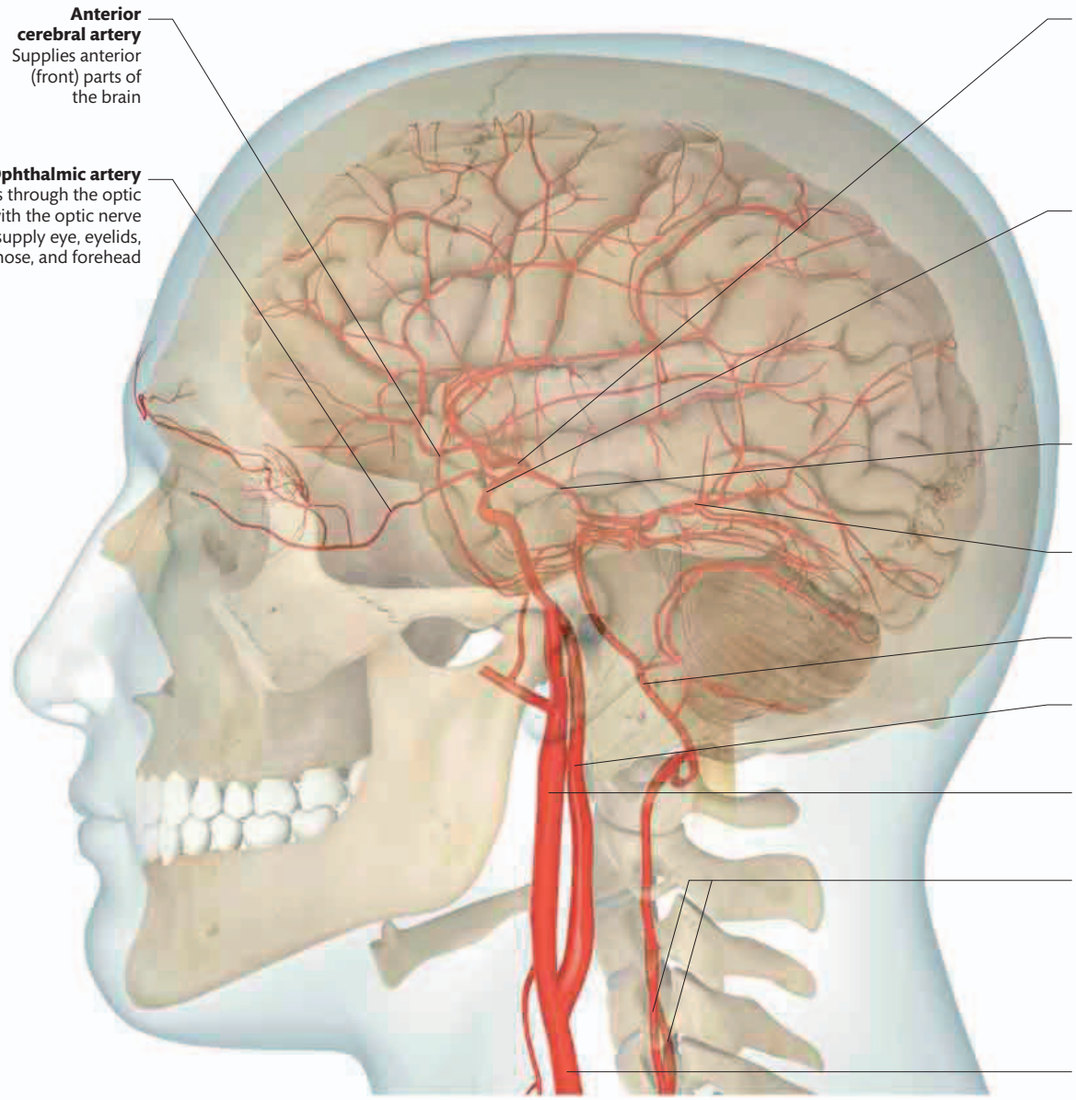

ARTERIES AROUND THE BRAIN

HEAD AND NECK

The brain has a rich blood supply, which arrives via the internal carotid and vertebral arteries. The vertebral arteries join together to form the basilar artery. The internal carotid arteries and basilar artery join up on the undersurface of the brain to form the Circle of Willis. From there, three pairs of cerebral arteries make their way into the brain. The veins of the brain and the skull drain into venous sinuses, which are enclosed within the dura mater (the outermost layer of the meninges) and form grooves on the inner surface of the skull. The sinuses join up and eventually drain out of the base of the skull, into the internal jugular vein.

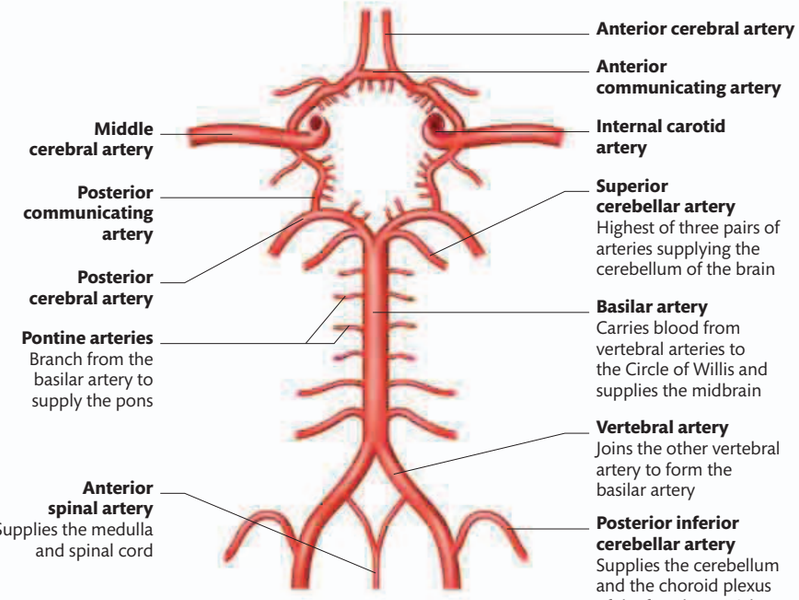

CIRCLE OF WILLIS

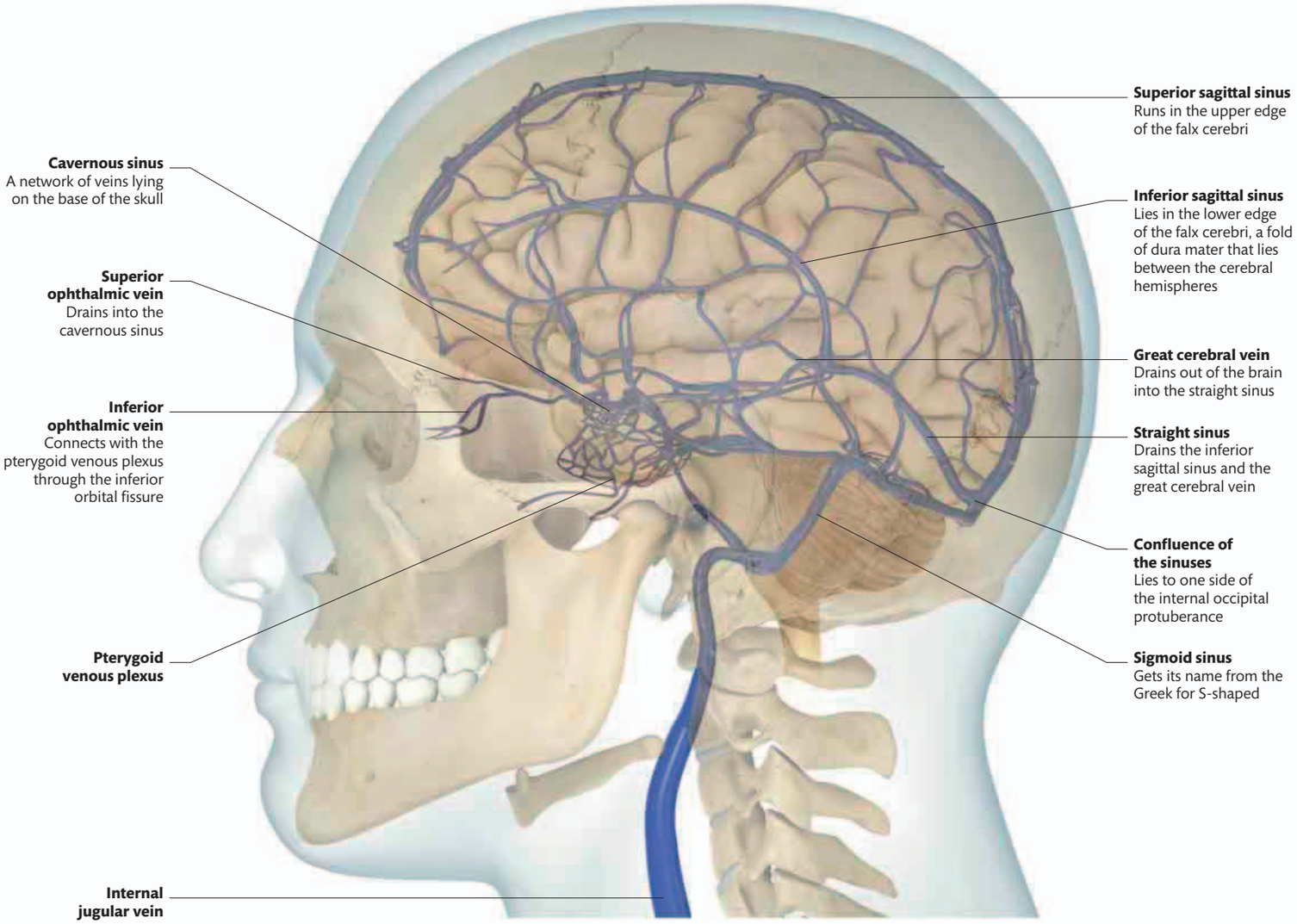

VEINS AROUND THE BRAIN

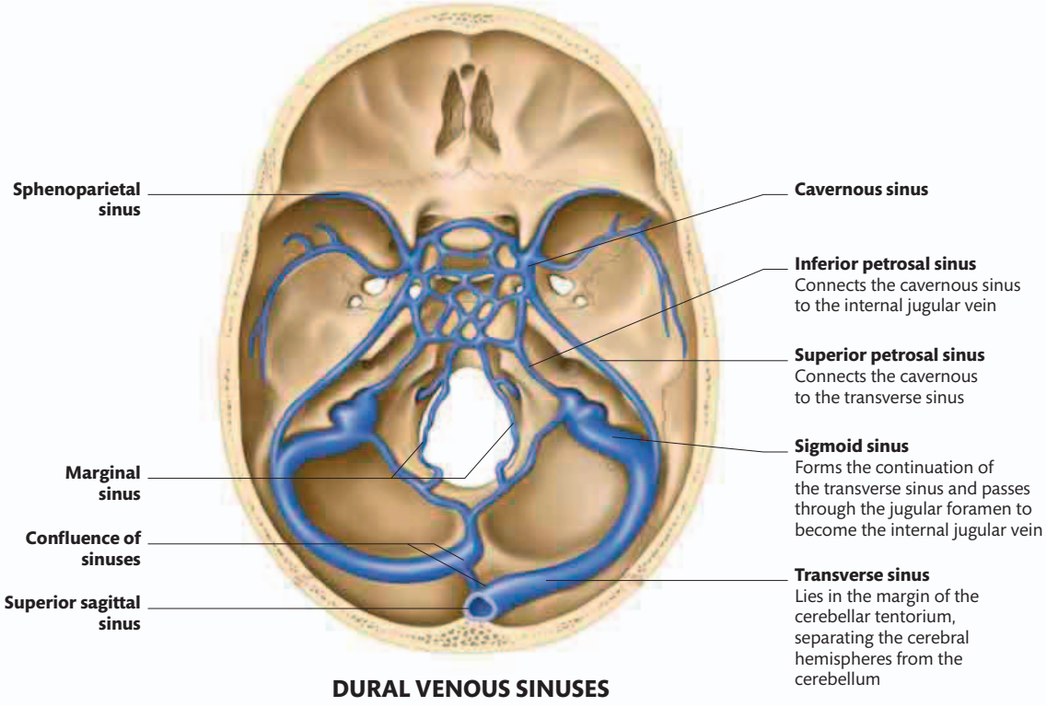

DURAL VENOUS SINUSES

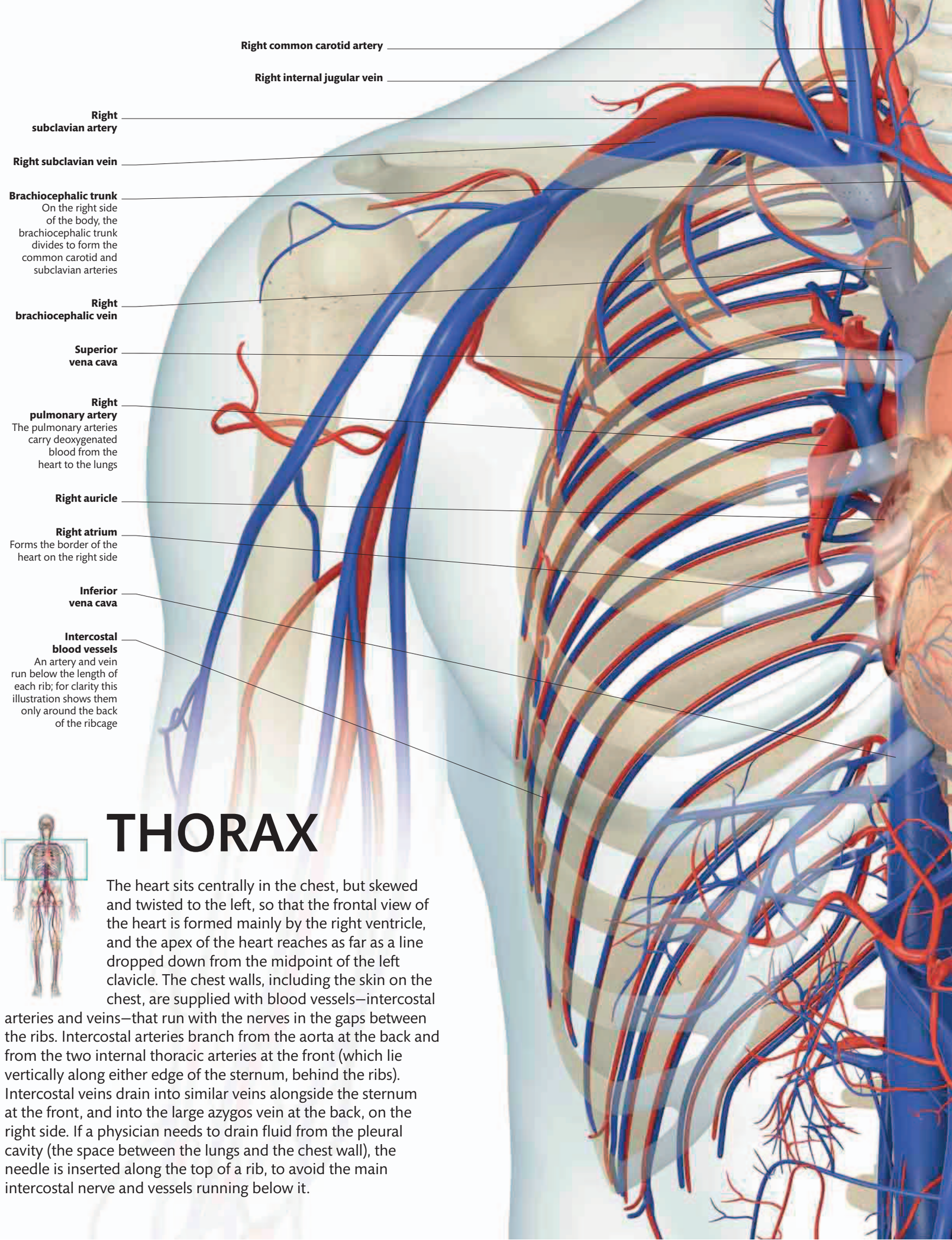

THORAX

The heart sits centrally in the chest, but skewed and twisted to the left, so that the frontal view of the heart is formed mainly by the right ventricle, and the apex of the heart reaches as far as a line dropped down from the midpoint of the left clavicle. The chest walls, including the skin on the chest, are supplied with blood vessels—intercostal arteries and veins—that run with the nerves in the gaps between the ribs. Intercostal arteries branch from the aorta at the back and from the two internal thoracic arteries at the front (which lie vertically along either edge of the sternum, behind the ribs). Intercostal veins drain into similar veins alongside the sternum at the front, and into the large azygos vein at the back, on the right side. If a physician needs to drain fluid from the pleural cavity (the space between the lungs and the chest wall), the needle is inserted along the top of a rib, to avoid the main intercostal nerve and vessels running below it.

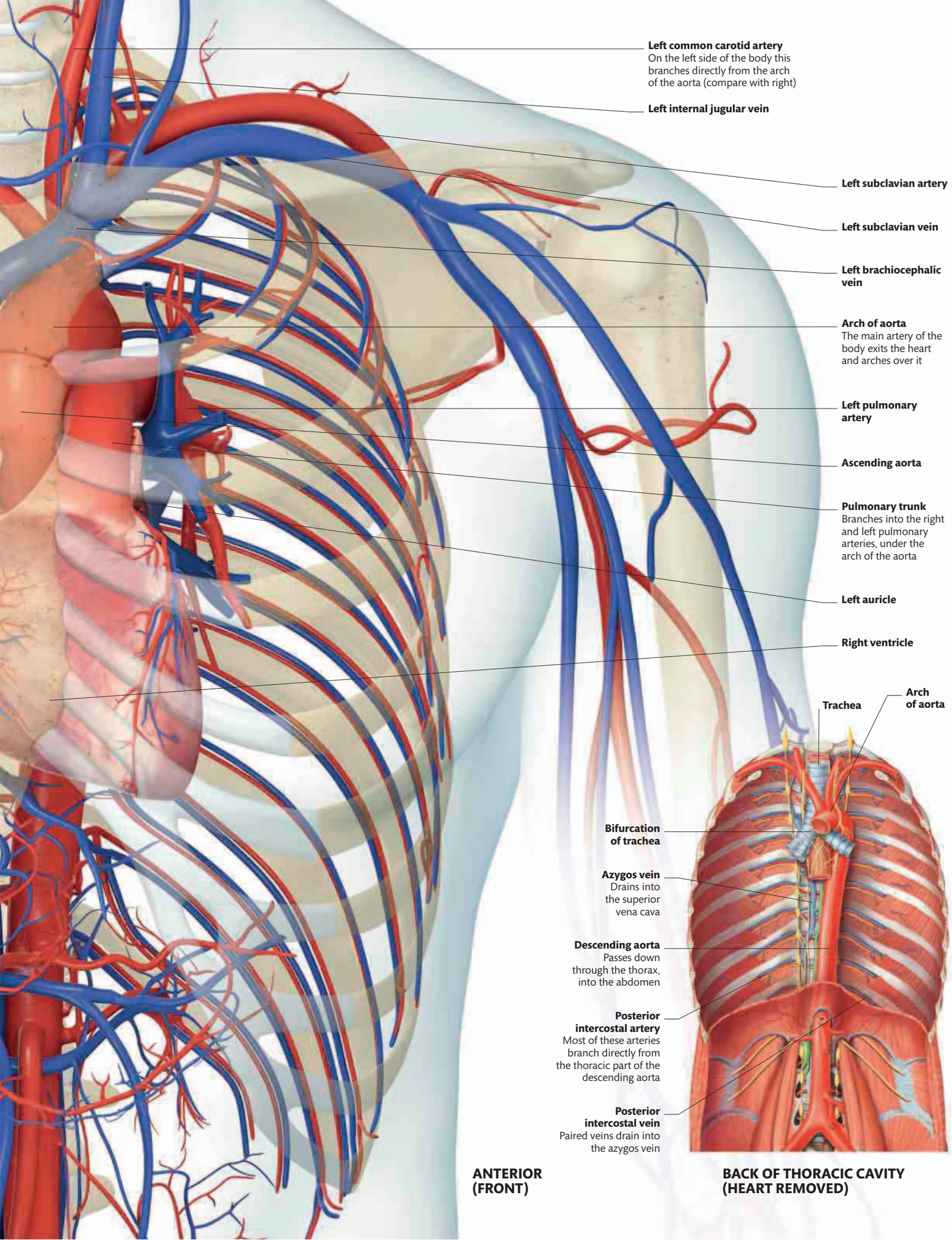
Left common carotid artery
On the left side of the body this branches directly from the arch of the aorta (compare with right)
Left internal jugular vein
Left subclavian artery
Left subclavian vein
Left brachiocephalic vein
Arch of aorta
The main artery of the body exits the heart and arches over it
Left pulmonary artery
Ascending aorta
Pulmonary trunk
Branches into the right and left pulmonary arteries, under the arch of the aorta
Left auricle
Right ventricle
Trachea
Arch of aorta
Bifurcation of trachea
Azygos vein
Drains into the superior vena cava
Descending aorta
Passes down through the thorax, into the abdomen
Posterior intercostal artery
Most of these arteries branch directly from the thoracic part of the descending aorta
Posterior intercostal vein
Paired veins drain into the azygos vein
ANTERIOR (FRONT)
BACK OF THORACIC CAVITY (HEART REMOVED)

Right vagus nerve
The vagus nerves travel through the thorax, giving branches to the heart and lungs, then travel close to the esophagus, behind the heart, as they descend into the abdomen

Right phrenic nerve

Superior vena cava

Right pulmonary artery

Right auricle
A pocketlike projection from the right atrium; its name comes from the Latin for little ear

Small cardiac vein
Drains into the coronary sinus

Right coronary artery
The word coronary comes from the Latin for crown; the right and left coronary arteries encircle the heart

Right ventricle

Marginal artery
A branch of the right coronary artery

Left phrenic nerve
The phrenic nerves are branches from the cervical plexus in the neck; they supply the muscle of the diaphragm

Left vagus nerve

Arch of aorta

Left recurrent laryngeal nerve
This branch of the left vagus nerve loops under the arch of the aorta before travelling back up to the neck to supply the larynx

Cut edge of pericardium

Left pulmonary artery

Pulmonary trunk

Left auricle
Similar to the right auricle, this is a projection from the left atrium

Great cardiac vein
Drains into the coronary sinus

Anterior interventricular artery
A branch of the left coronary artery that runs down between the two ventricles on the front of the heart

Left ventricle

Apex

Pericardium

ANTERIOR (FRONT)

HEART

The heart is encased in the pericardium. This has a tough outer layer that is fused to the diaphragm below and to the connective tissue around the large blood vessels above the heart. Lining the inside of this cylinder (and the outer surface of the heart) is a thin membrane called the serous pericardium. Between these two layers is a thin film of fluid that lubricates the movement of the heart as it beats. Inflammation of this membrane, known as pericarditis, can be extremely painful. Branches of the right and left coronary arteries, which spring from the ascending aorta, supply the heart muscle itself. The heart is drained by cardiac veins, most of which drain into the coronary sinus.

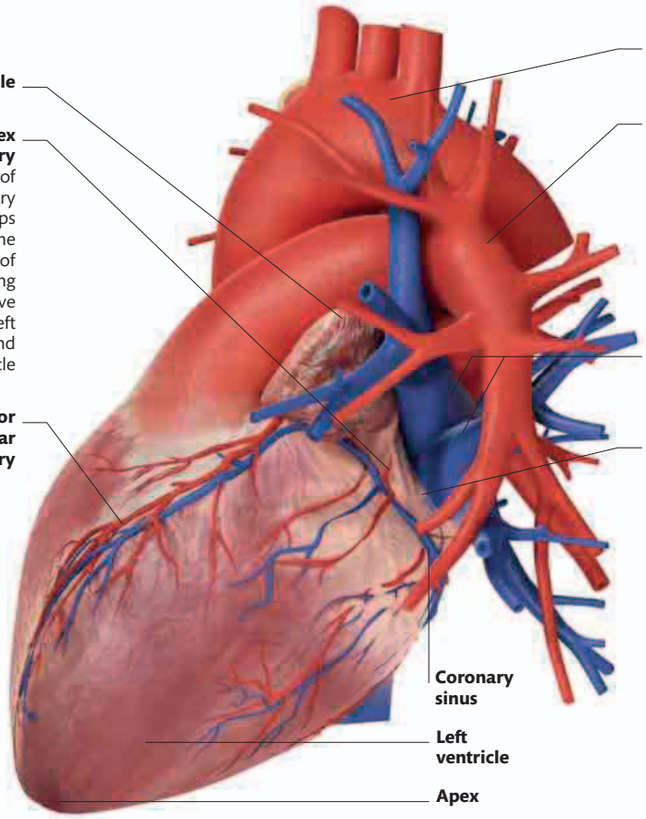

LEFT

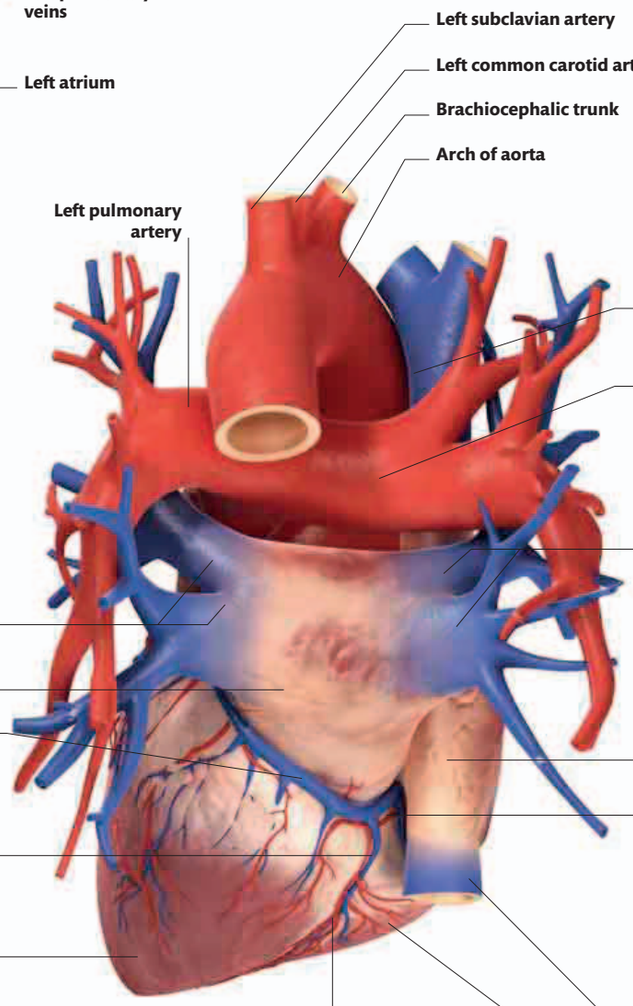

POSTERIOR (BACK)

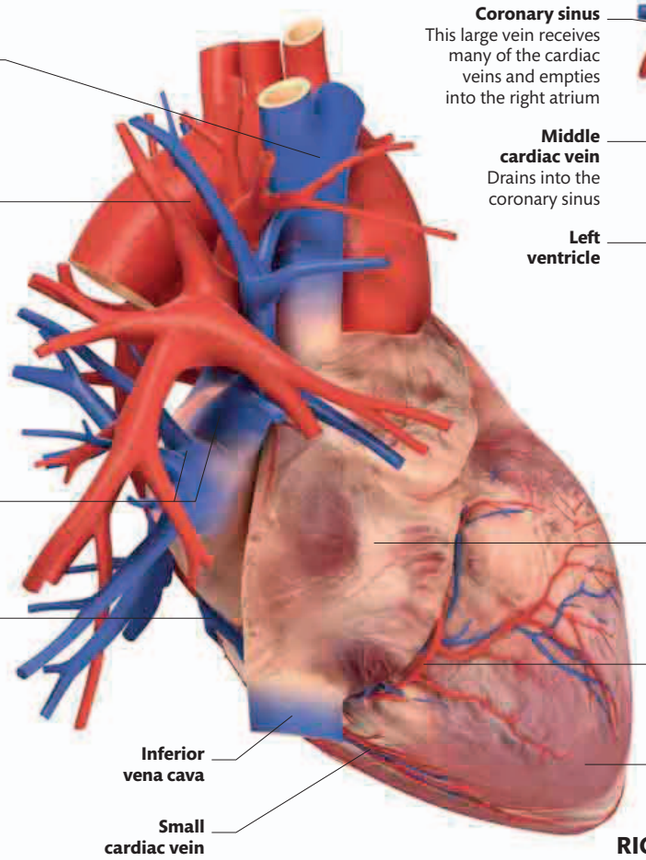

RIGHT

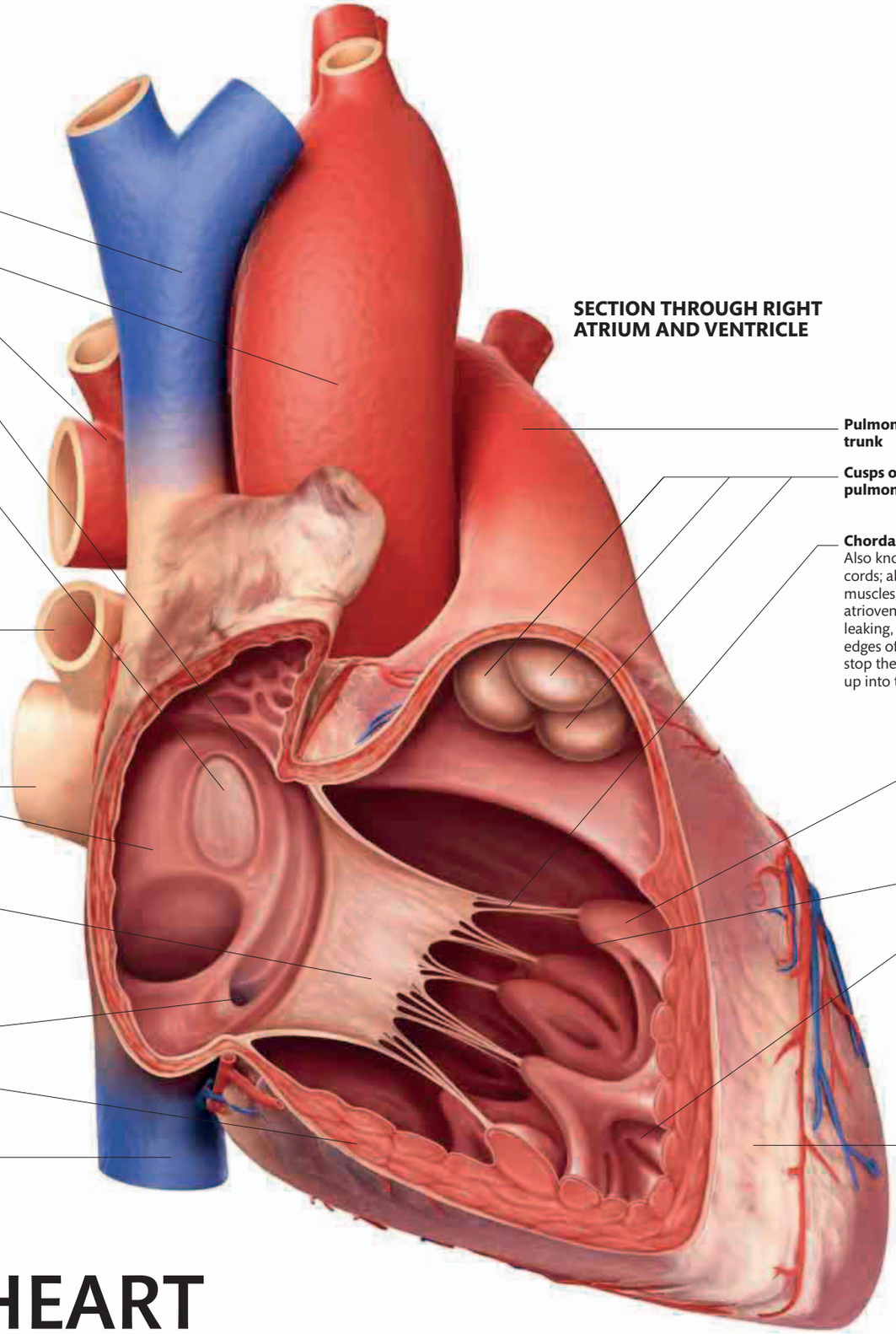

HEART

The heart receives blood from veins and pumps it out through arteries. It has four chambers: two atria and two ventricles. The heart's left and right sides are separate. The right side receives deoxygenated blood from the body via the superior and inferior venae cavae, and pumps it to the lungs through the pulmonary trunk. The left gets oxygenated blood from the lungs via the pulmonary veins, and pumps it into the aorta for distribution. Each atrium opens into its corresponding ventricle via a valve (on the right, the tricuspid valve, and the bicuspid valve on the left), which shuts when the ventricle contracts, to stop blood flowing back into the atrium. The aorta and pulmonary trunk also have valves.

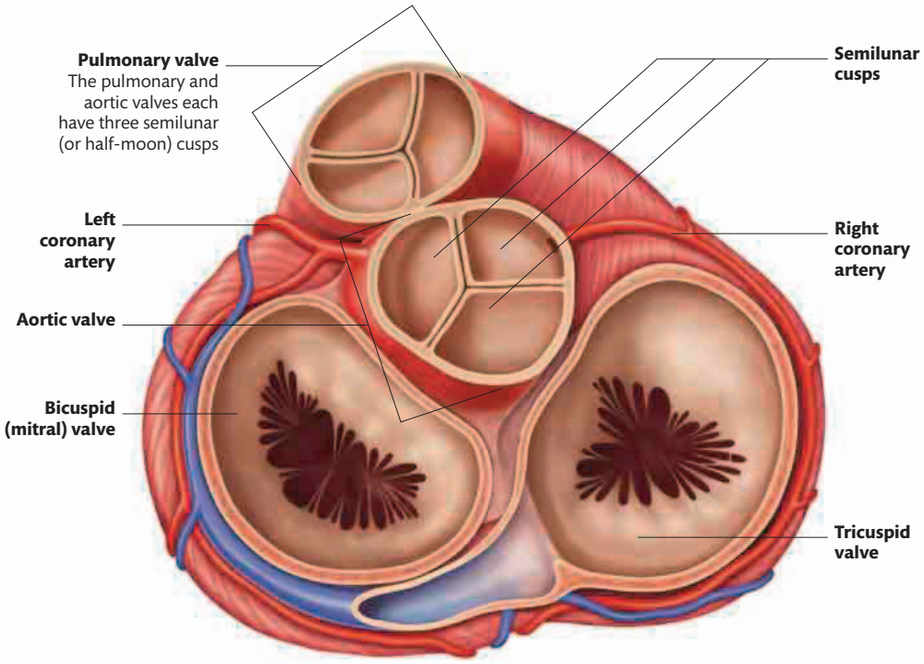

TRANSVERSE SECTION SHOWING VALVES

CARDIAC MUSCLE STRUCTURE

Cardiac muscle (myocardium) is essential to the heart's function as a pump. When the myocardium contracts, blood is squeezed out of the heart. The myocardium is a network of interconnected fibers, which contract rhythmically and spontaneously. Autonomic nerves can adjust the rate of contraction, matching the heart's output to the body's need.

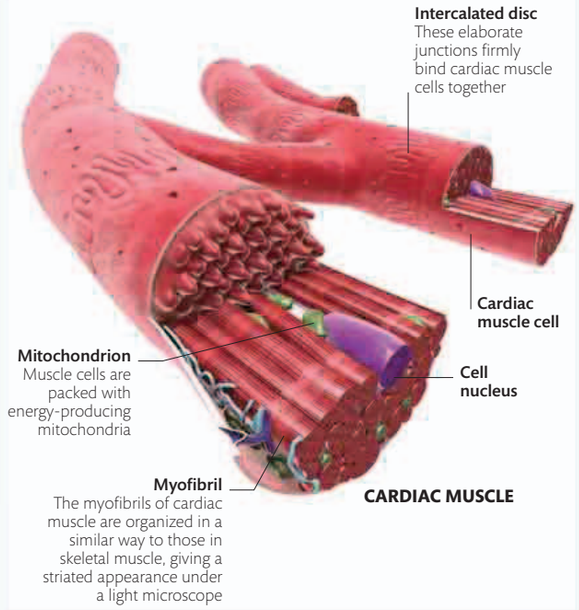

CARDIAC MUSCLE

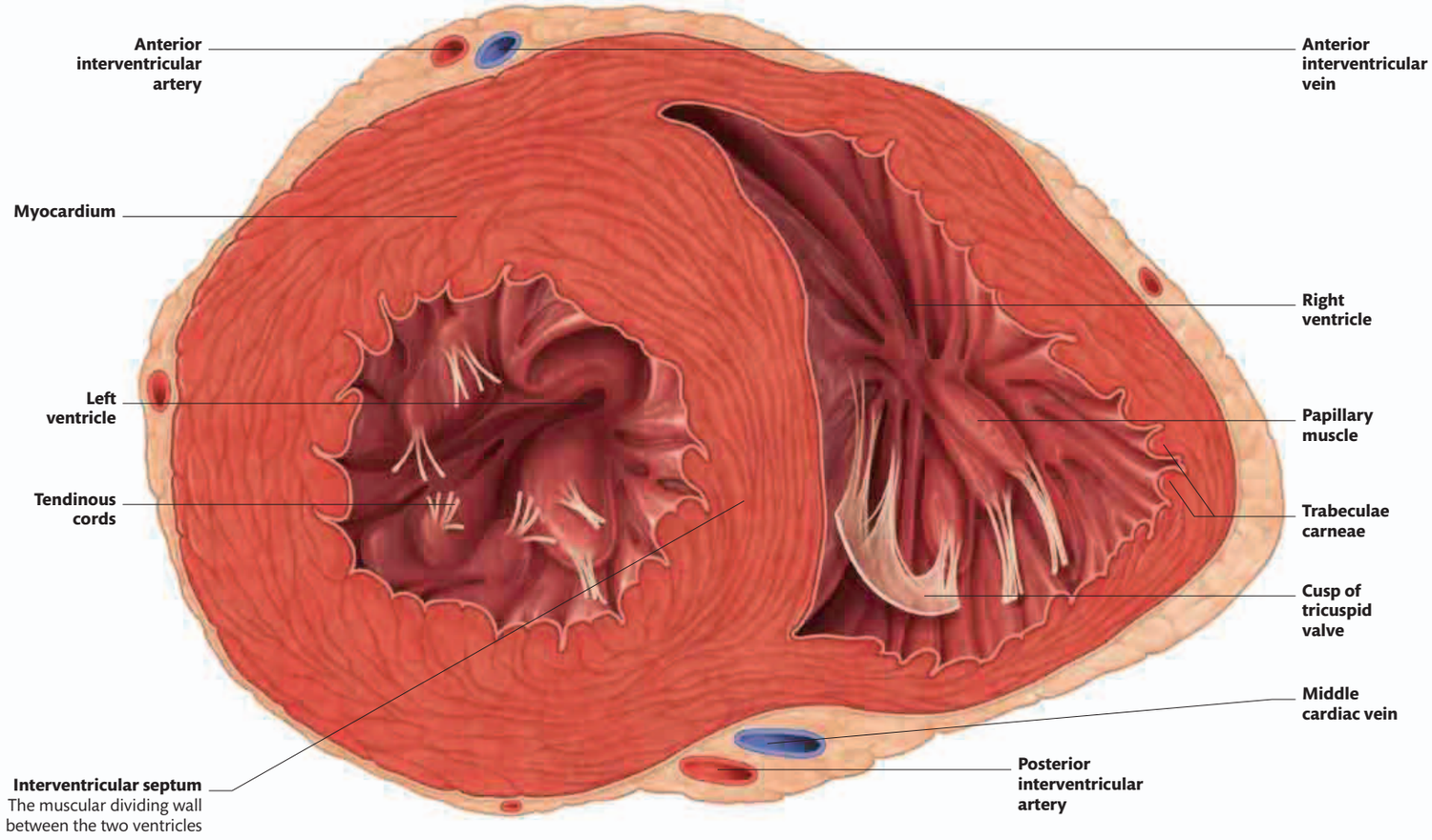

TRANSVERSE SECTION THROUGH VENTRICLES

ABDOMEN AND PELVIS

The aorta passes behind the diaphragm, level with the twelfth thoracic vertebra, and enters the abdomen. Pairs of arteries branch from the sides of the aorta to supply the walls of the abdomen, the kidneys, suprarenal glands, and the testes or ovaries with oxygenated blood. A series of branches emerge from the front of the abdominal aorta to supply the abdominal organs: the celiac trunk gives branches to the liver, stomach, pancreas, and spleen, and the mesenteric arteries provide blood to the gut. The abdominal aorta ends by splitting into two, forming the common iliac arteries. Each of these then divides, in turn, forming an internal iliac artery (which supplies the pelvic organs) and an external iliac artery (which continues into the thigh, becoming the femoral artery). Lying to the right of the aorta is the major vein of the abdomen: the inferior vena cava.

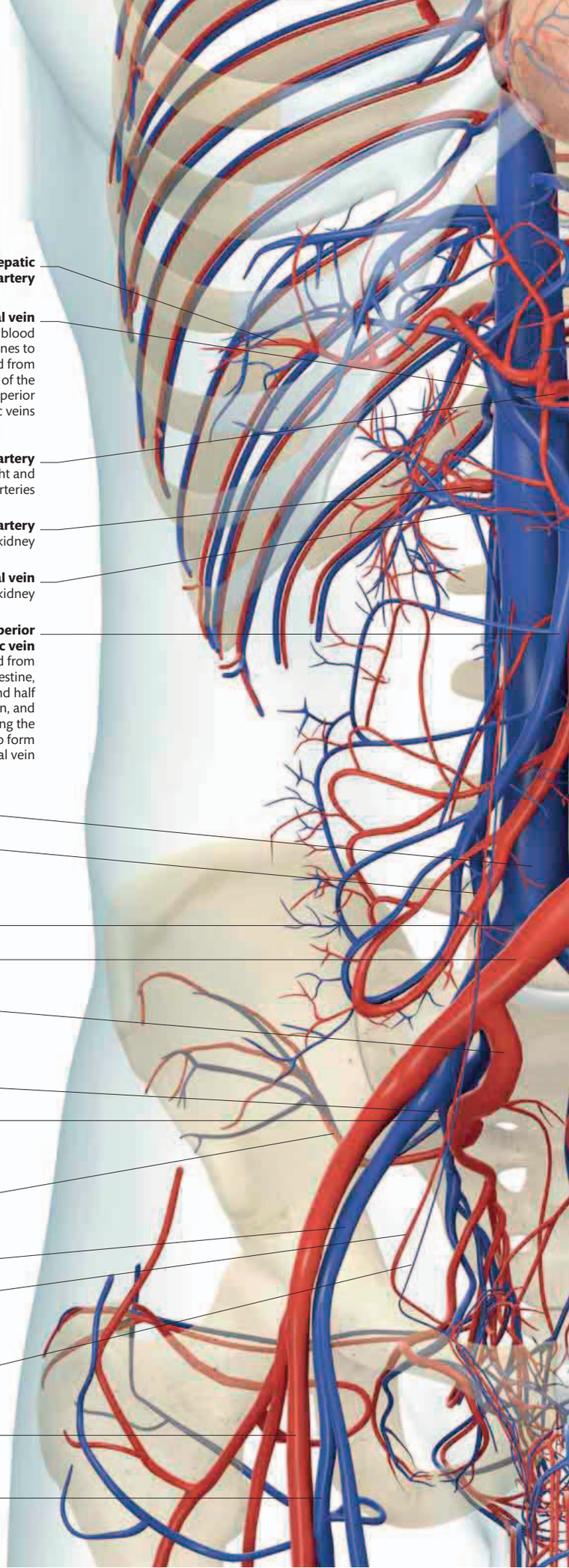

ANTERIOR (FRONT)

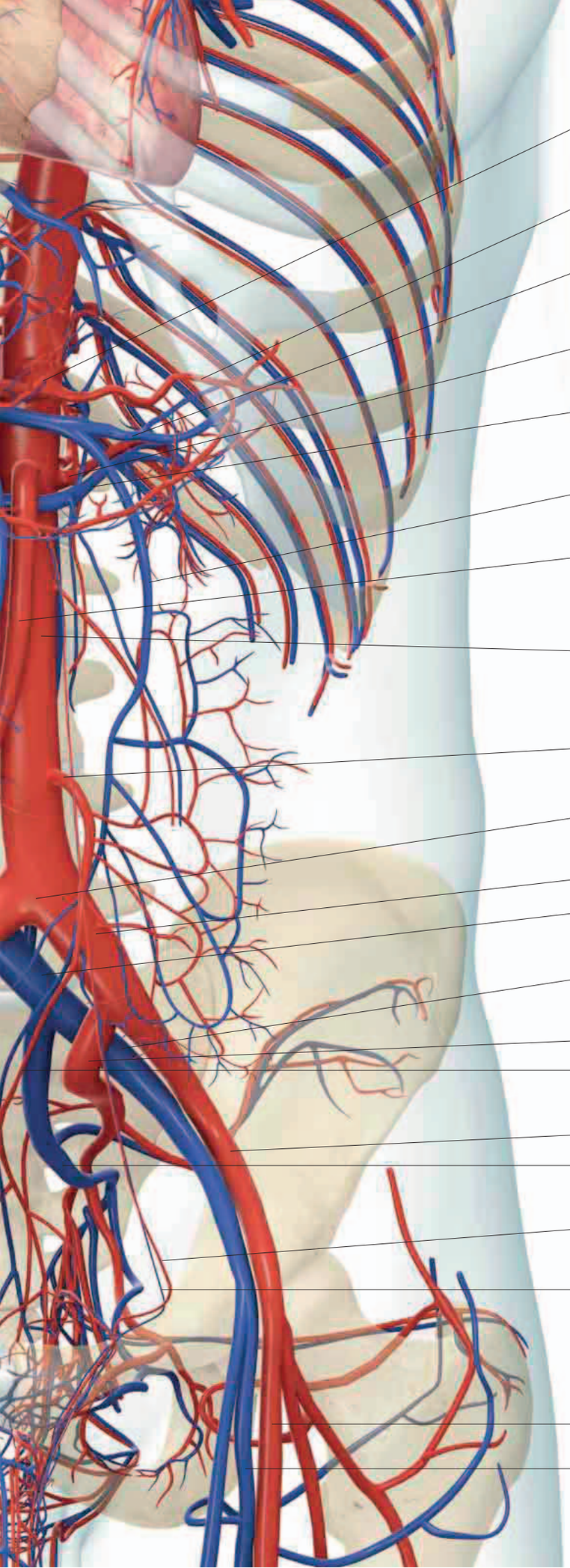

Celiac trunk
Only just over ⅜in (1cm) long, it quickly branches into the left gastric, splenic, and common hepatic arteries

Splenic artery
Supplies the spleen, as well as most of the pancreas, and the upper part of the stomach

Splenic vein
Drains the spleen and receives other veins from the stomach and pancreas, as well as the inferior mesenteric vein

Left renal artery
Shorter than the right renal artery, this supplies the left kidney

Left renal vein
Longer than its counterpart on the right, this drains the left kidney and receives the left gonadal vein

Inferior mesenteric vein
Drains blood from the colon and rectum and ends by emptying into the splenic vein

Superior mesenteric artery
Branches within the mesentery to supply a great length of intestine, including all of the jejunum and ileum and half of the colon

Abdominal aorta
The thoracic aorta becomes the abdominal aorta as it passes behind the diaphragm, level with the twelfth thoracic vertebra

Inferior mesenteric artery
Supplies the last third of the transverse colon, the descending and sigmoid colon, and the rectum

Bifurcation of aorta
The abdominal aorta divides in front of the fourth lumbar vertebra

Left common iliac artery

Left common iliac vein
Formed from the union of the external and internal iliac veins

Left external iliac vein
The continuation of the femoral vein, after it has passed into the pelvis

Left internal iliac artery

Superior rectal artery
The last branch of the inferior mesenteric artery passes down into the pelvis to supply the rectum

Left external iliac artery

Left internal iliac vein
Drains the pelvic organs, perineum, and buttock

Left gonadal artery
Gonadal arteries branch from the aorta just below the renal arteries

Left gonadal vein
Drains the ovary or testis, and empties into the left renal vein

Left femoral artery

Left femoral vein
The main vein from the leg; becomes the external iliac vein

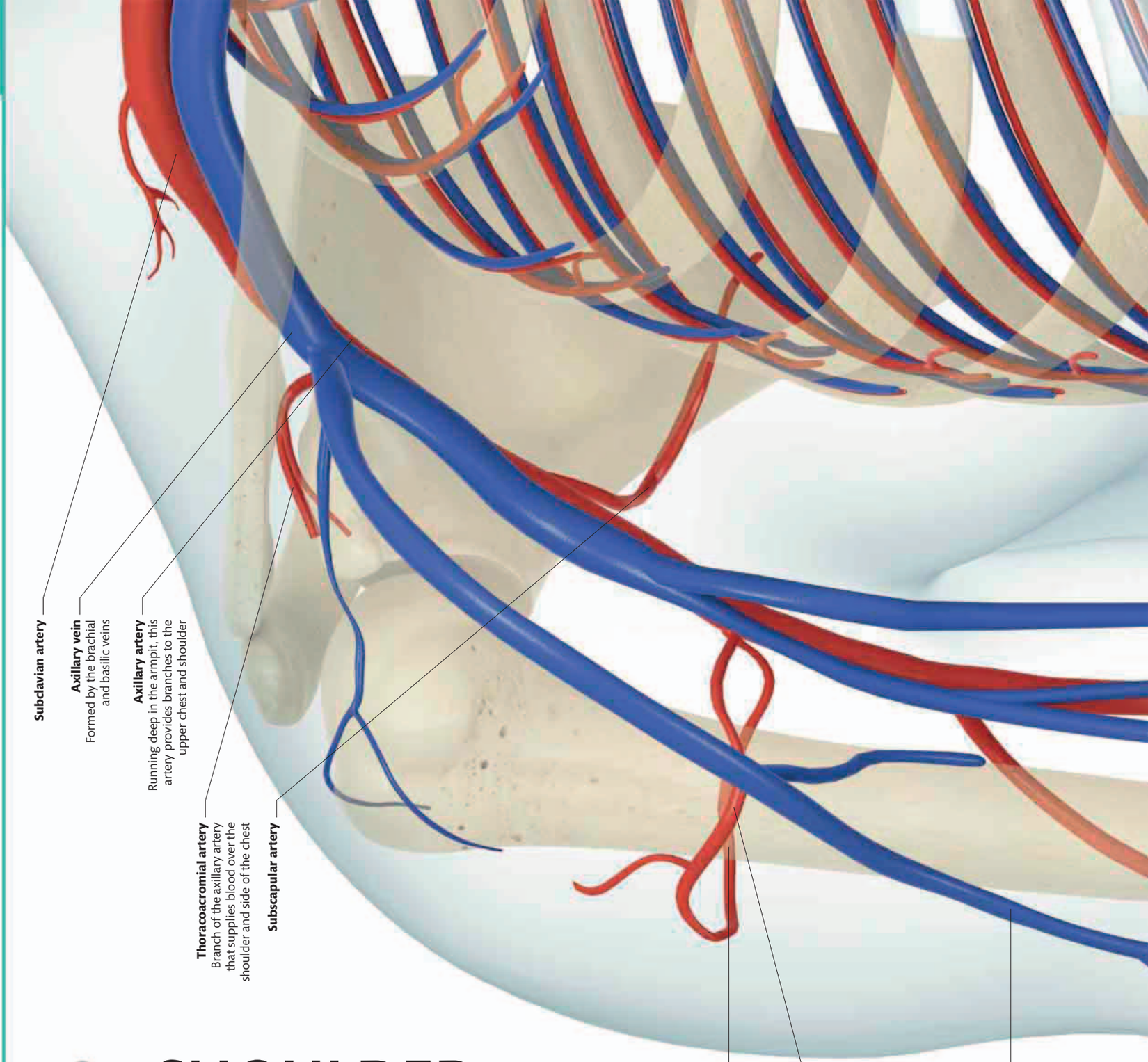

SHOULDER AND UPPER ARM

The subclavian artery is the main arterial supply to the upper limb. When this artery passes under the clavicle and into the axilla (armpit), it becomes the axillary artery. Several branches spring off in this region, running backward toward the scapula, up to the shoulder, and around the humerus. Beyond the armpit, the name of the axillary artery changes to the brachial artery, which runs down the front of the arm, usually accompanied by a pair of companion veins. Two superficial veins that drain blood from the back of the hand end in the arm by draining into deep veins: the basilic vein drains into brachial veins; the cephalic vein runs up to the shoulder, then plunges deeper to join the axillary vein.

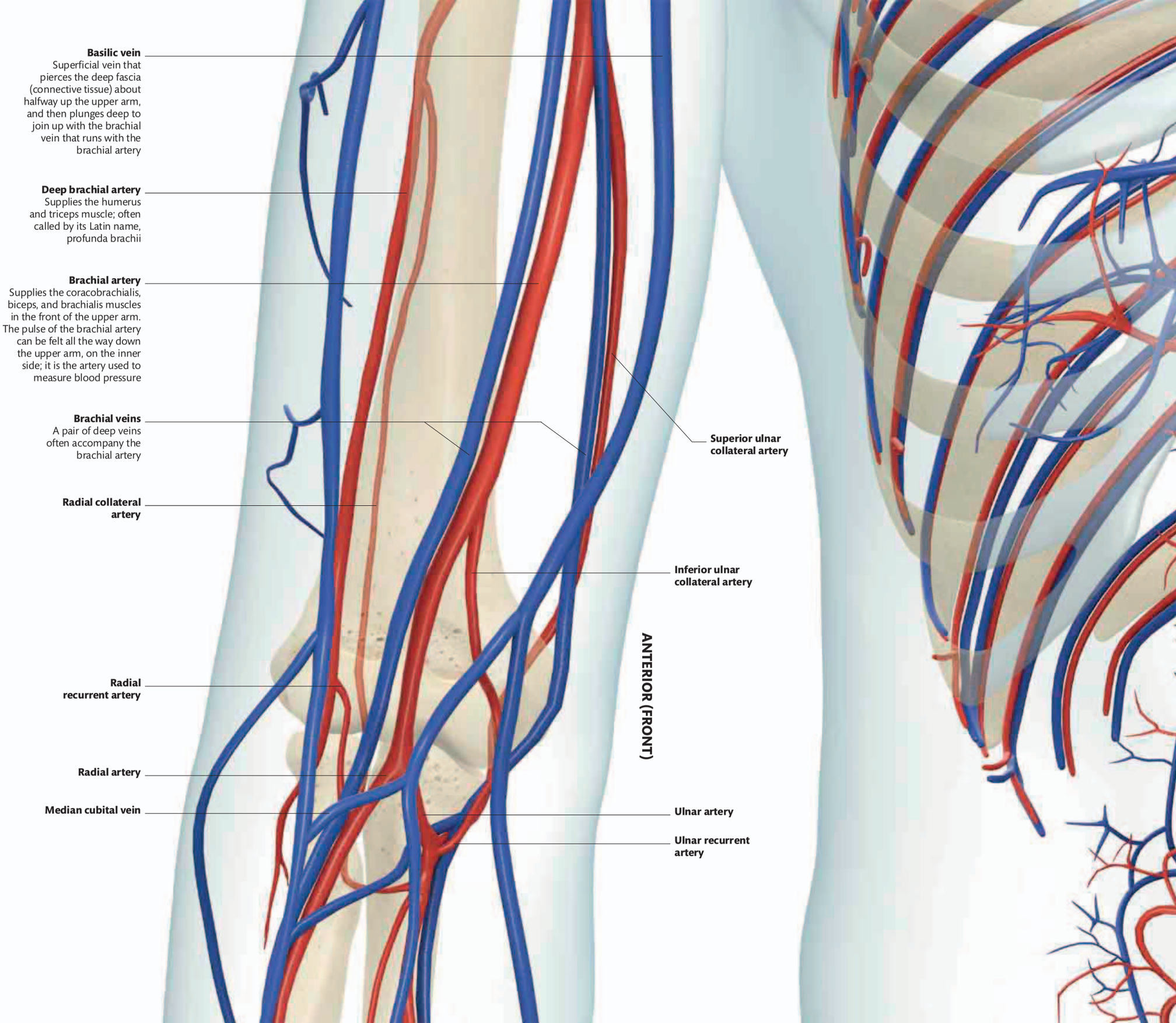
Basilic vein
Superficial vein that pierces the deep fascia (connective tissue) about halfway up the upper arm, and then plunges deep to join up with the brachial vein that runs with the brachial artery
Deep brachial artery
Supplies the humerus and triceps muscle; often called by its Latin name, profunda brachii
Brachial artery
Supplies the coracobrachialis, biceps, and brachialis muscles in the front of the upper arm. The pulse of the brachial artery can be felt all the way down the upper arm, on the inner side; it is the artery used to measure blood pressure
Brachial veins
A pair of deep veins often accompany the brachial artery
Radial collateral artery
Radial recurrent artery
Radial artery
Median cubital vein
Superior ulnar collateral artery
Inferior ulnar collateral artery
ANTERIOR (FRONT)
Ulnar artery
Ulnar recurrent artery

SHOULDER AND UPPER ARM

Various branches from the axillary and brachial arteries supply the back of the shoulder and upper arm. The posterior circumflex humeral artery, which runs with the axillary nerve, curls around the upper end of the humerus. The deep brachial artery runs with the radial nerve, spiraling around the back of the bone. From this artery, and from the brachial artery itself, collateral branches run down the arm and join up, or anastomose, with recurrent branches running back up from the ulnar and radial arteries of the forearm. There are also anastomoses (links) between branches of the subclavian and axillary arteries around the shoulder. Anastomoses like this, where branches from different regions join up, can provide alternative routes through which blood can flow if the main vessel becomes squashed or blocked.

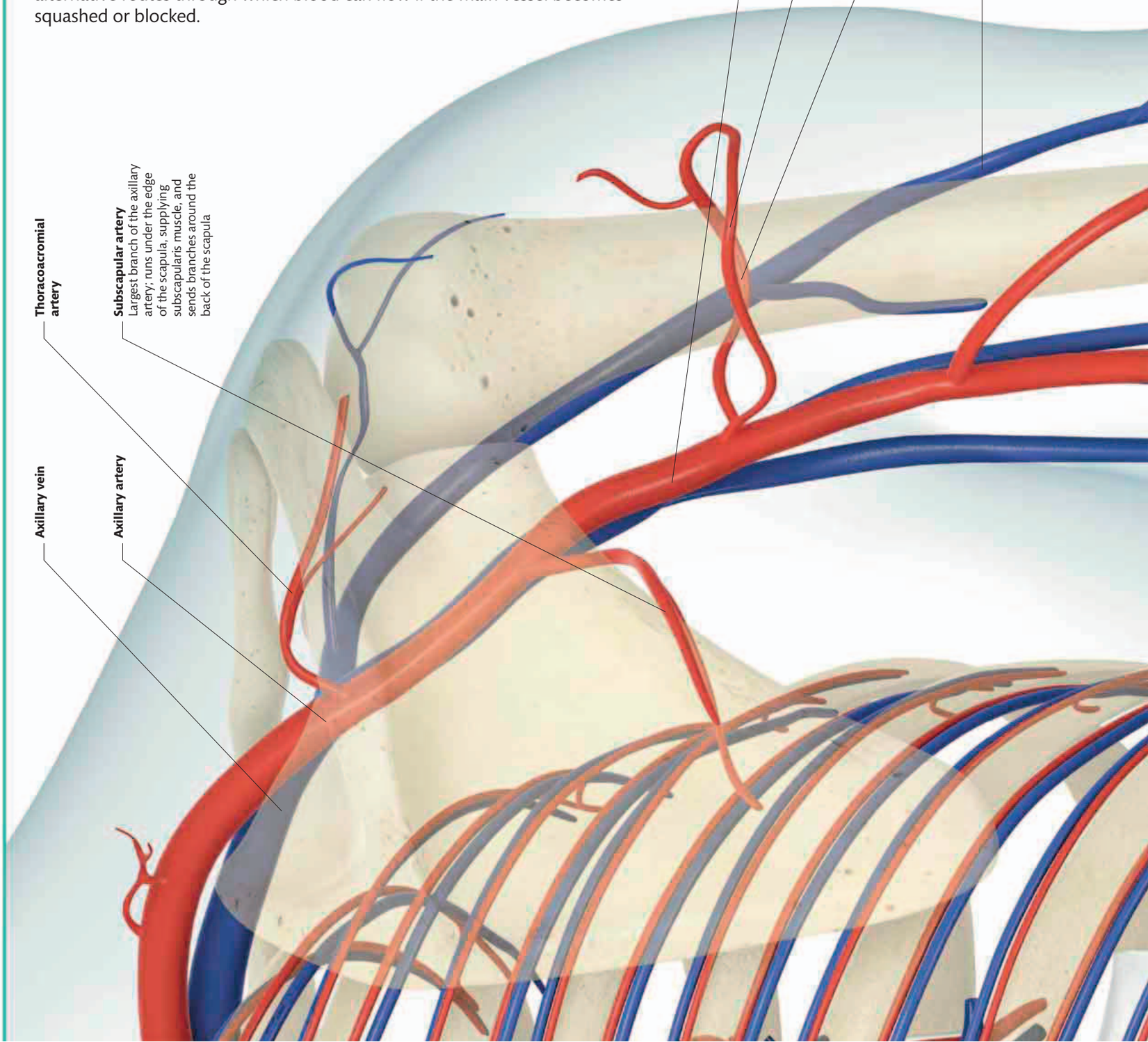

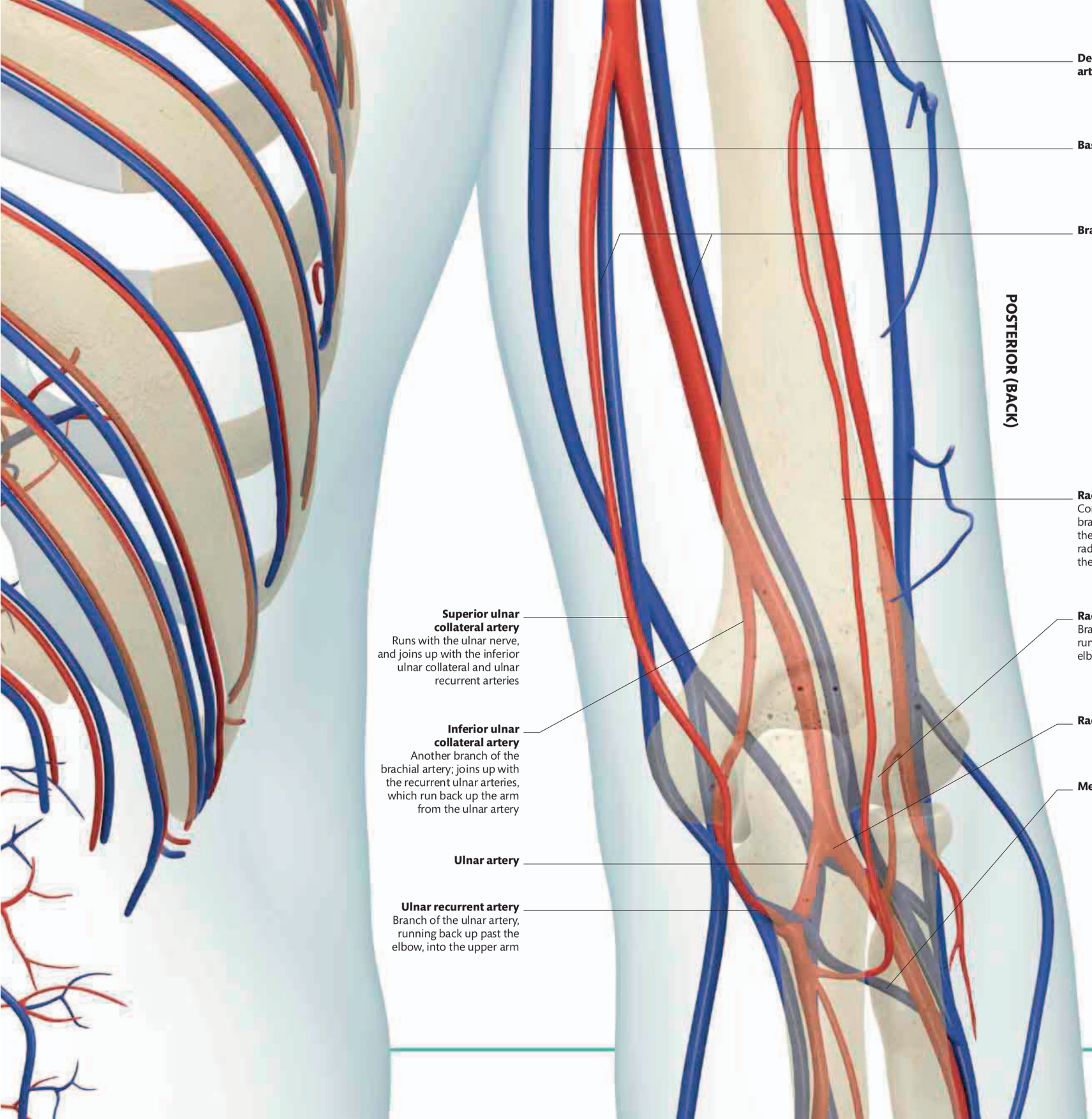
POSTERIOR (BACK)
Deep brachial artery
Basilic vein
Brachial veins
Radial collateral artery
Continuation of the deep brachial artery, running down the side of the arm, with the radial nerve, to join up with the radial recurrent artery
Radial recurrent artery
Branch of the radial artery, running back up past the elbow, into the upper arm
Radial artery
Median cubital vein
Superior ulnar collateral artery
Runs with the ulnar nerve, and joins up with the inferior ulnar collateral and ulnar recurrent arteries
Inferior ulnar collateral artery
Another branch of the brachial artery; joins up with the recurrent ulnar arteries, which run back up the arm from the ulnar artery
Ulnar artery
Ulnar recurrent artery
Branch of the ulnar artery, running back up past the elbow, into the upper arm

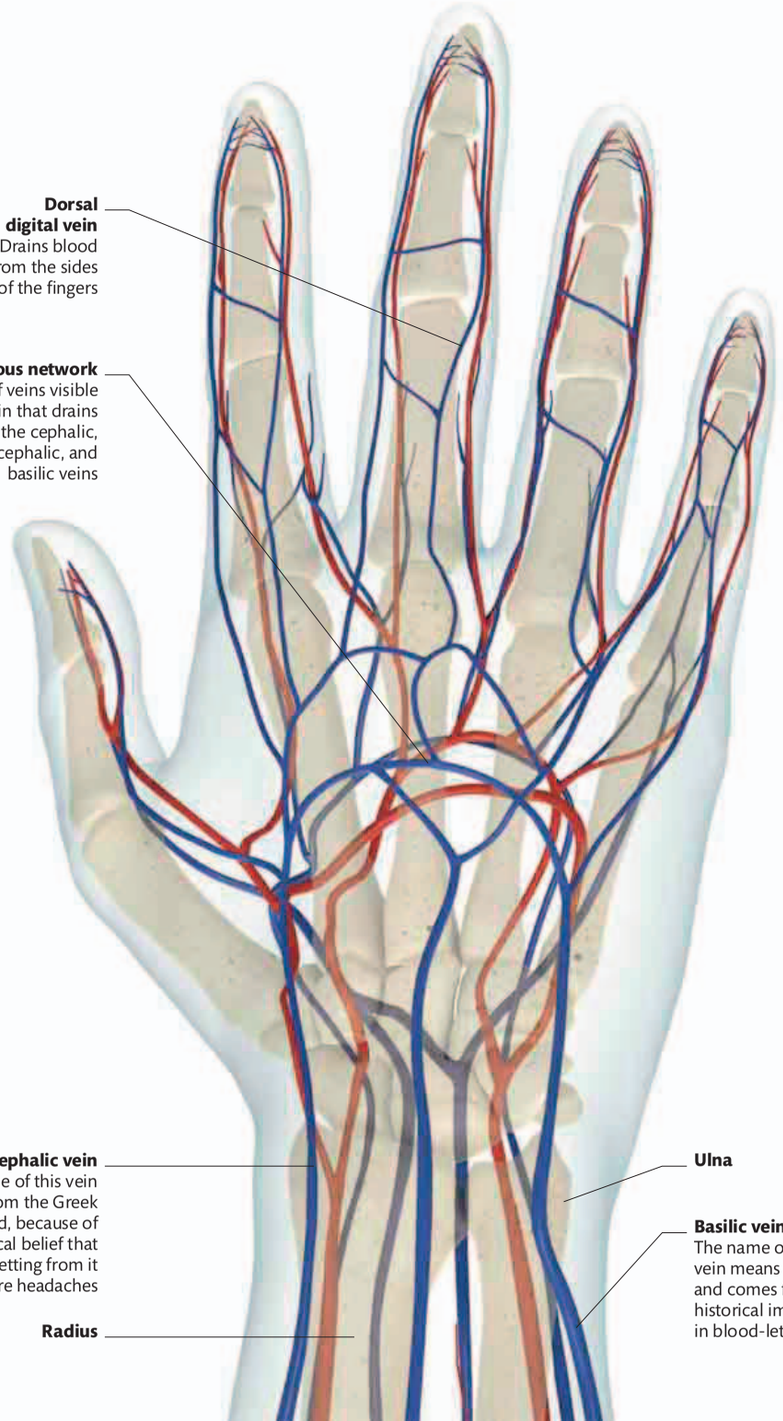

ANTERIOR (FRONT)

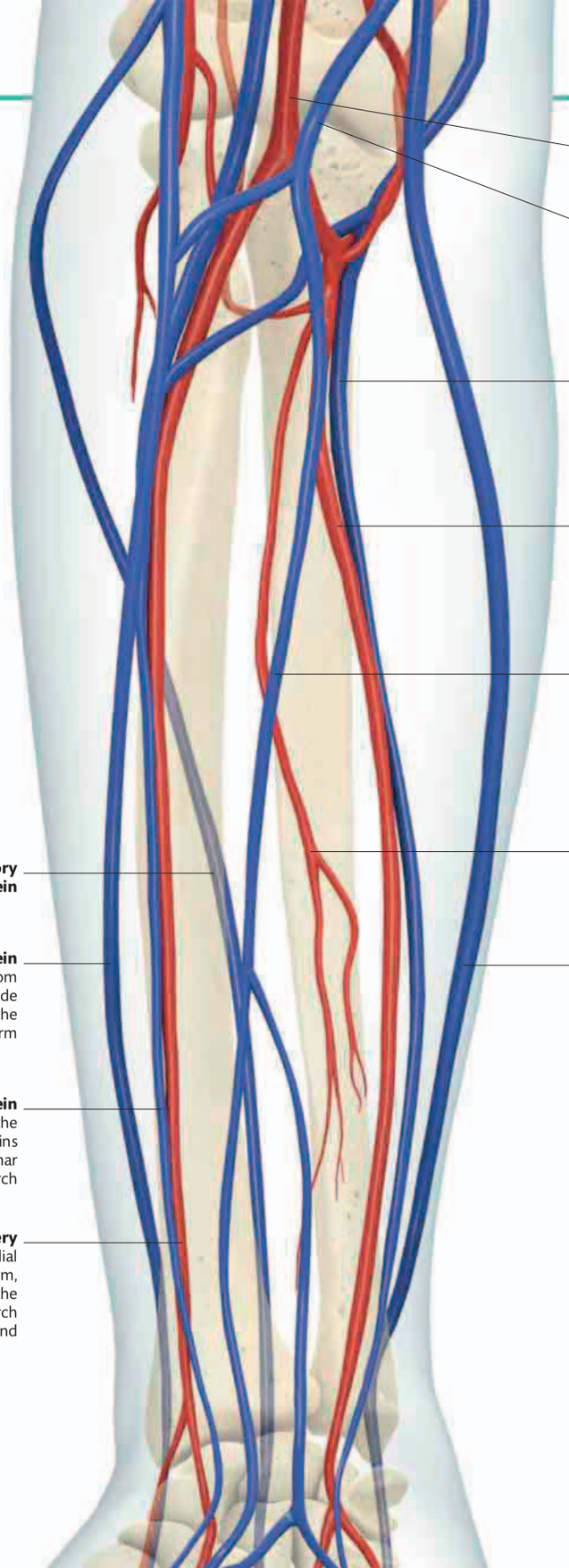

LOWER ARM AND HAND

The brachial artery divides into two arteries, which take their names from the bones of the forearm: the radial and ulnar arteries. The radial artery can be felt at the wrist, and this is the most common place for taking a pulse as the strong pulsations are easy to feel when the artery is pressed against the bone beneath it. Furthermore, taking a pulse here doesn't even require any undressing. The radial and ulnar arteries end by joining up to form arterial arches in the wrist and palm. Digital arteries, destined for the fingers, spring off from the palmar arch. Superficial veins are concentrated on the back of the hand, rather than on the palm—otherwise those thin-walled vessels would be compressed every time a person gripped something. The dorsal venous network of the hand drains into two main vessels: the basilic and cephalic veins.

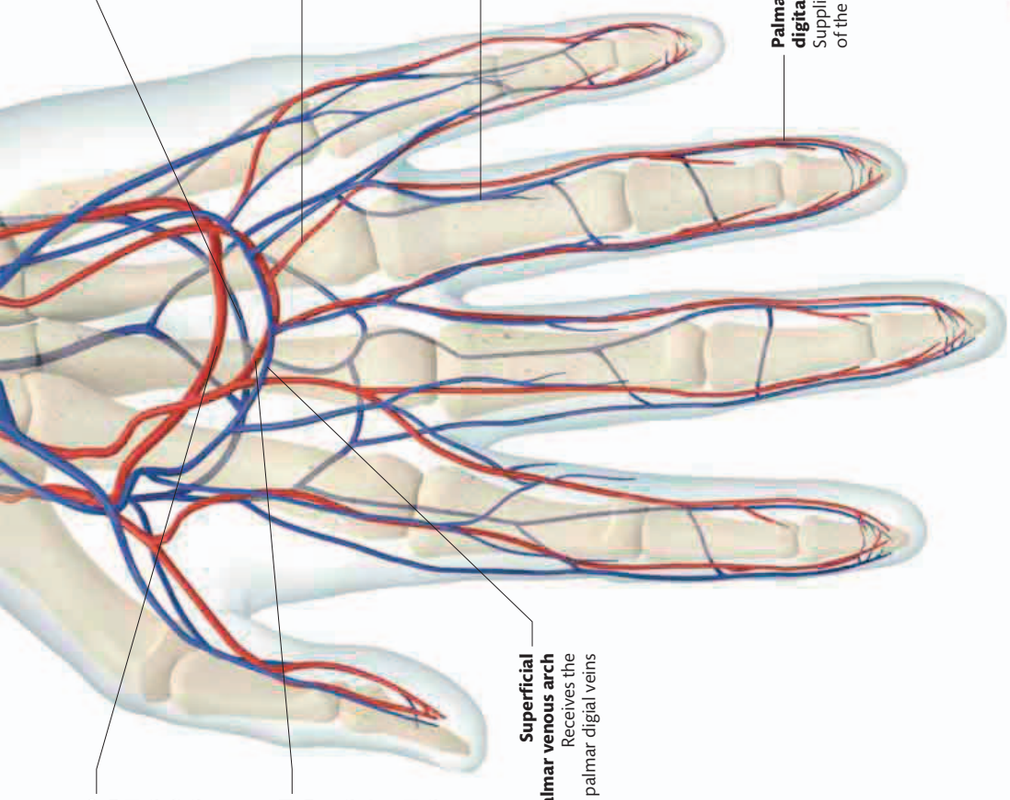

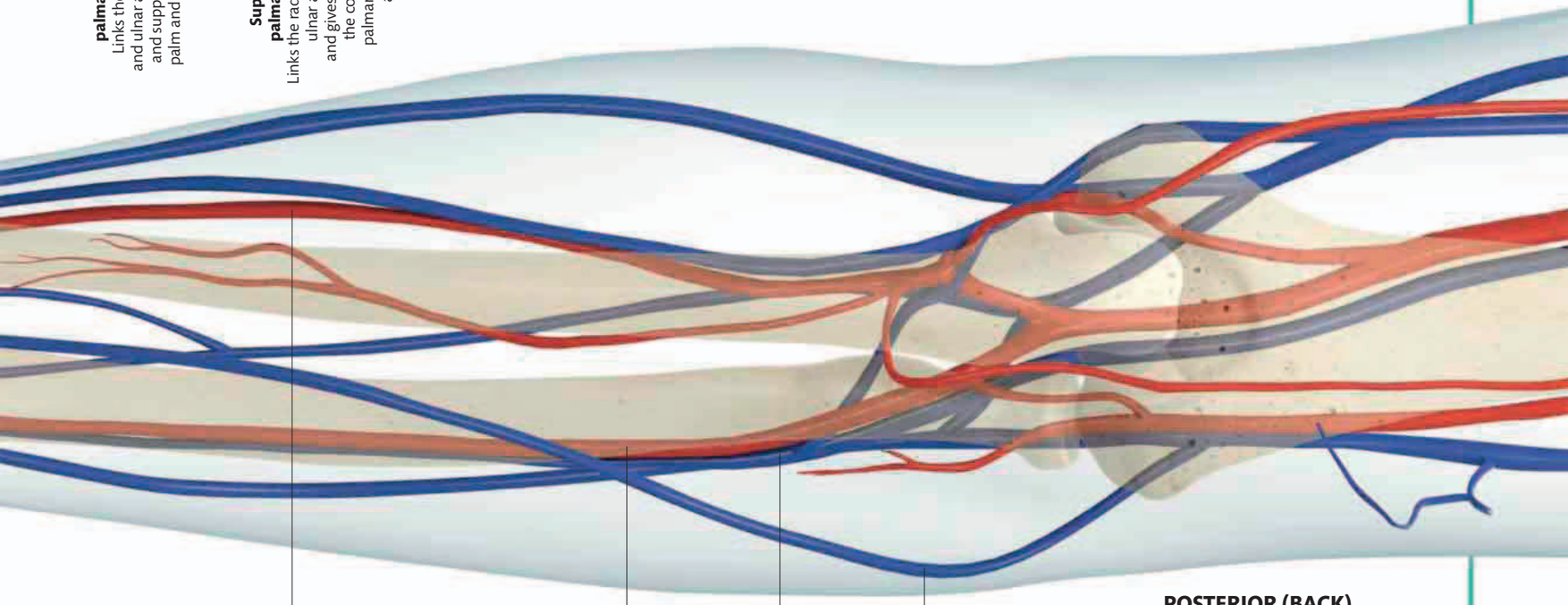

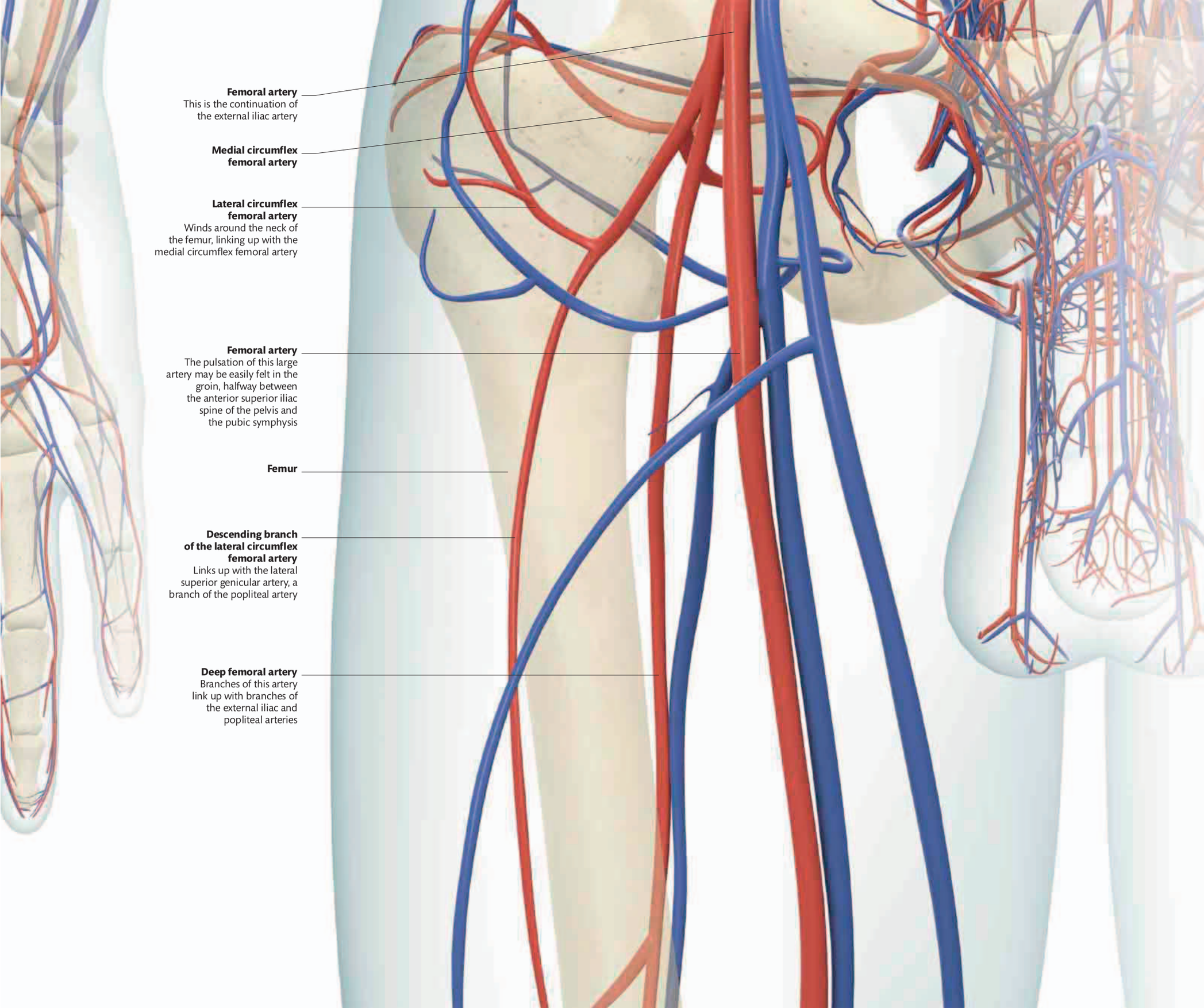
Femoral artery
This is the continuation of the external iliac artery
Medial circumflex femoral artery
Lateral circumflex femoral artery
Winds around the neck of the femur, linking up with the medial circumflex femoral artery
Femoral artery
The pulsation of this large artery may be easily felt in the groin, halfway between the anterior superior iliac spine of the pelvis and the pubic symphysis
Femur
Descending branch of the lateral circumflex femoral artery
Links up with the lateral superior genicular artery, a branch of the popliteal artery
Deep femoral artery
Branches of this artery link up with branches of the external iliac and popliteal arteries

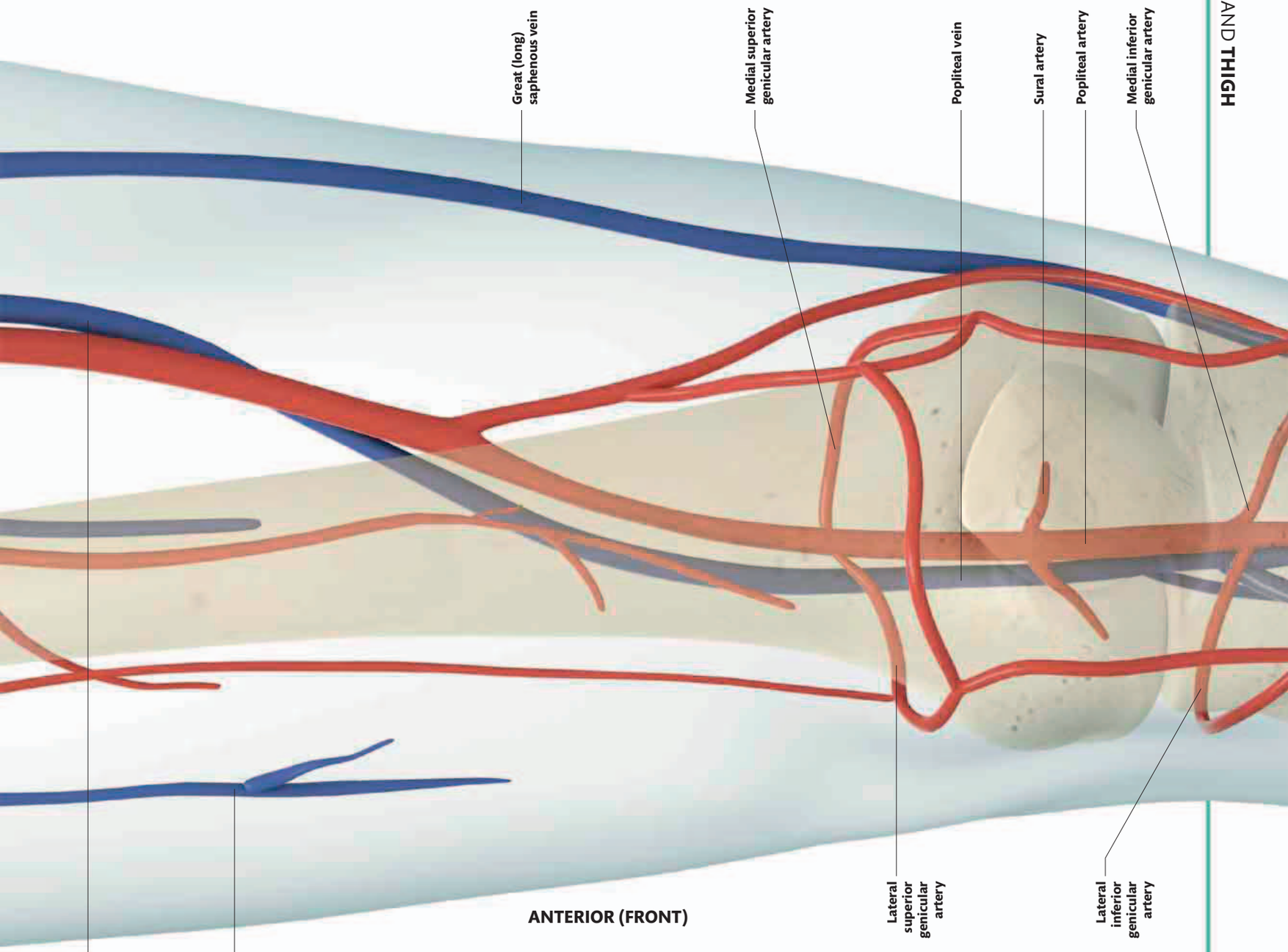

HIP AND THIGH

As the external iliac artery runs over the pubic bone and underneath the inguinal ligament, it changes its name to the femoral artery—the main vessel carrying blood to the lower limb. The femoral artery lies exactly halfway along a line between the anterior superior iliac spine of the pelvis and the pubic symphysis. It has a large branch, the deep femoral artery, that supplies the muscles of the thigh. The femoral artery then runs toward the inner thigh, passing through the hole in the adductor magnus tendon, where its name changes to the popliteal artery. Deep veins run with the arteries, but—just as in the arm—there are also superficial veins. The great (or long) saphenous vein drains up the inner side of the leg and thigh, and ends by joining the femoral vein near the hip.

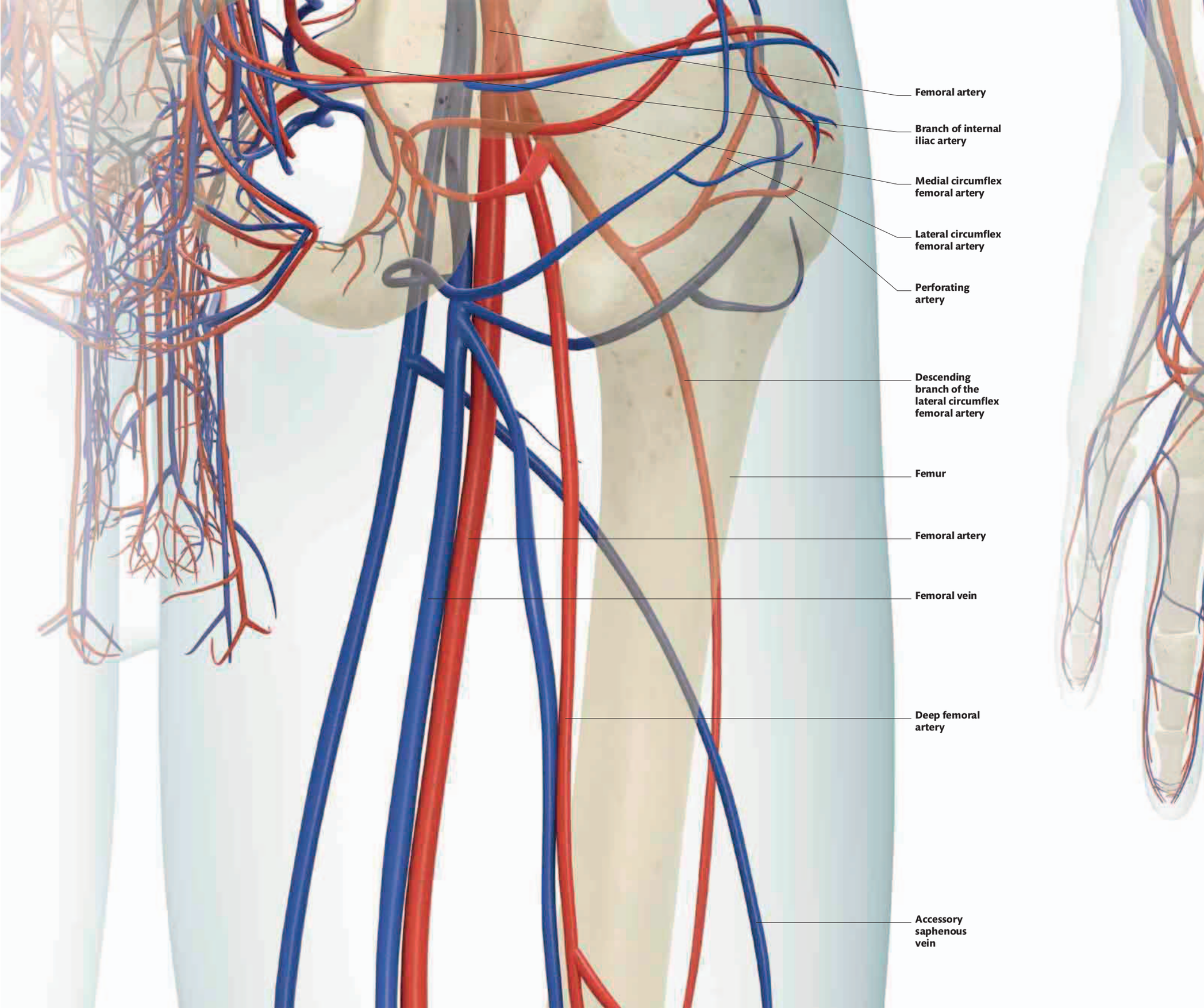
Femoral artery
Branch of internal iliac artery
Medial circumflex femoral artery
Lateral circumflex femoral artery
Perforating artery
Descending branch of the lateral circumflex femoral artery
Femur
Femoral artery
Femoral vein
Deep femoral artery
Accessory saphenous vein

HIP AND THIGH

In this back view, gluteal branches of the internal iliac artery can be clearly seen, emerging through the greater sciatic foramen to supply the buttock. The muscles and skin of the inner part and back of the thigh are supplied by branches of the deep femoral artery. These are known as the perforating arteries because they pierce through the adductor magnus muscle. Higher up, the circumflex femoral arteries encircle the femur. The popliteal artery, formed after the femoral artery passes through the hiatus (gap) in adductor magnus, lies on the back of the femur, deep to the popliteal vein.

POSTERIOR (BACK)

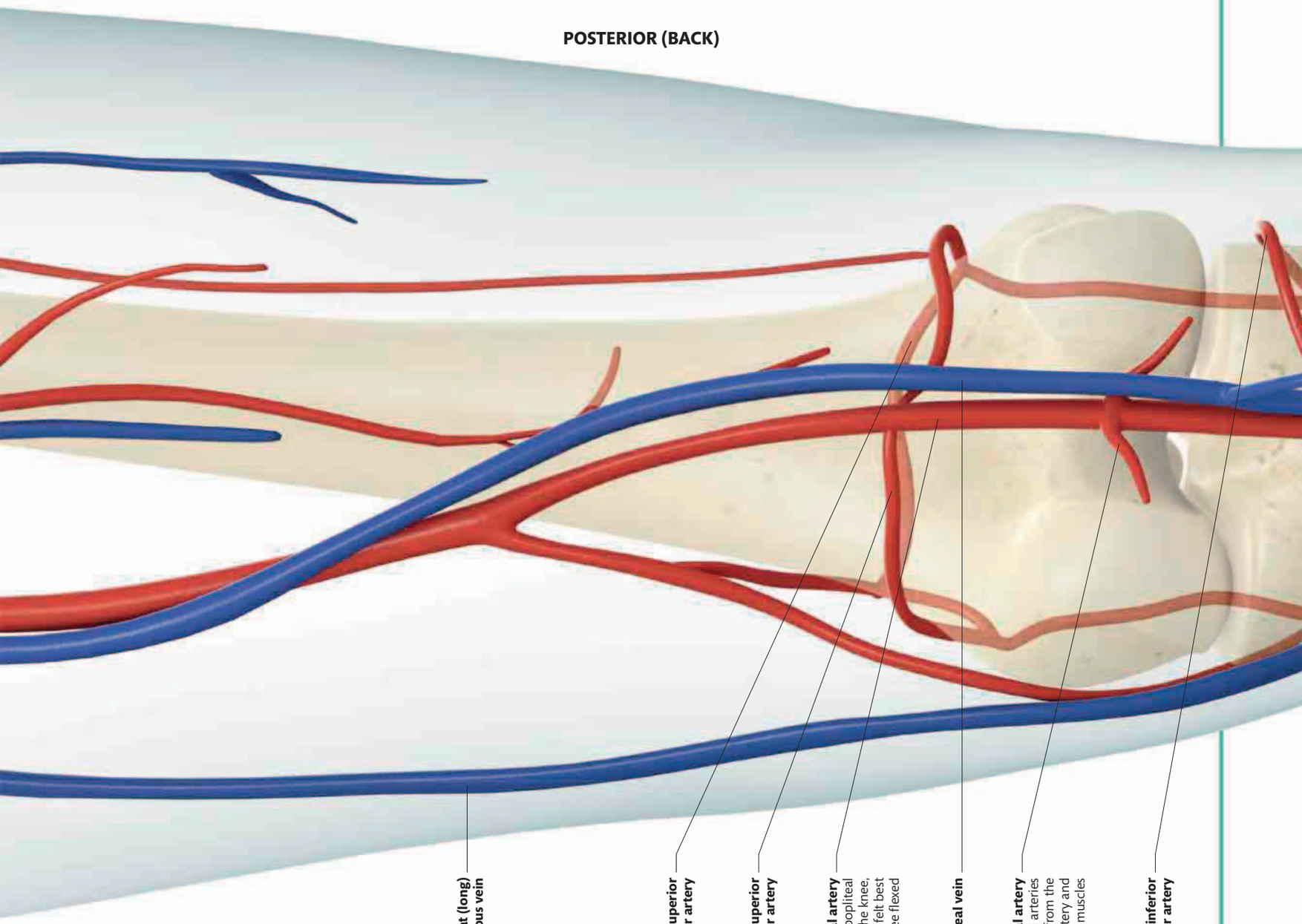

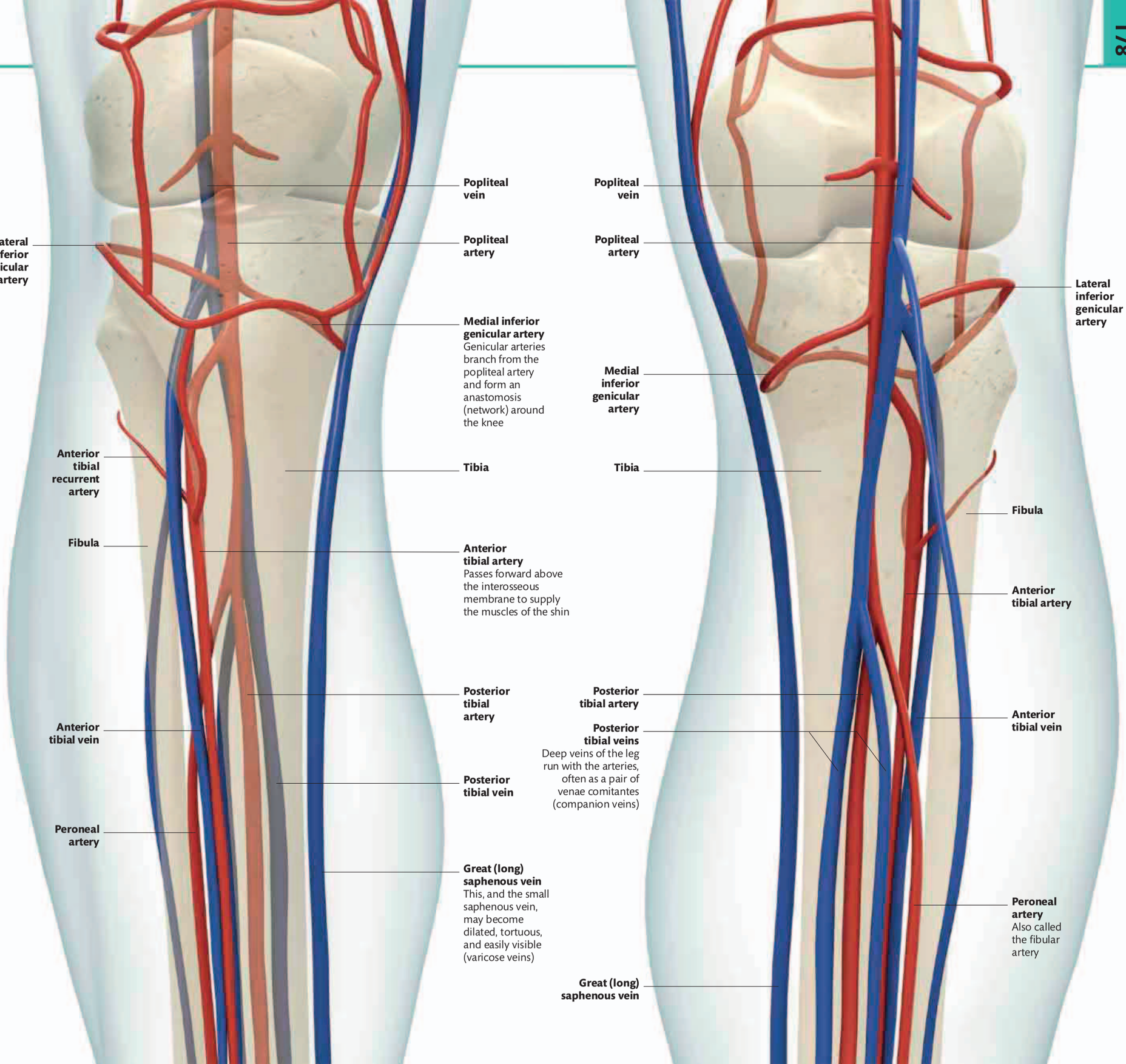
Popliteal vein
Lateral inferior genicular artery
Popliteal artery
Medial inferior genicular artery
Genicular arteries branch from the popliteal artery and form an anastomosis (network) around the knee
Anterior tibial recurrent artery
Tibia
Fibula
Anterior tibial artery
Passes forward above the interosseous membrane to supply the muscles of the shin
Posterior tibial artery
Anterior tibial vein
Posterior tibial vein
Peroneal artery
Great (long) saphenous vein
This, and the small saphenous vein, may become dilated, tortuous, and easily visible (varicose veins)
Popliteal vein
Popliteal artery
Lateral inferior genicular artery
Medial inferior genicular artery
Tibia
Fibula
Anterior tibial artery
Posterior tibial artery
Anterior tibial vein
Posterior tibial veins
Deep veins of the leg run with the arteries, often as a pair of venae comitantes (companion veins)
Peroneal artery
Also called the fibular artery
Great (long) saphenous vein

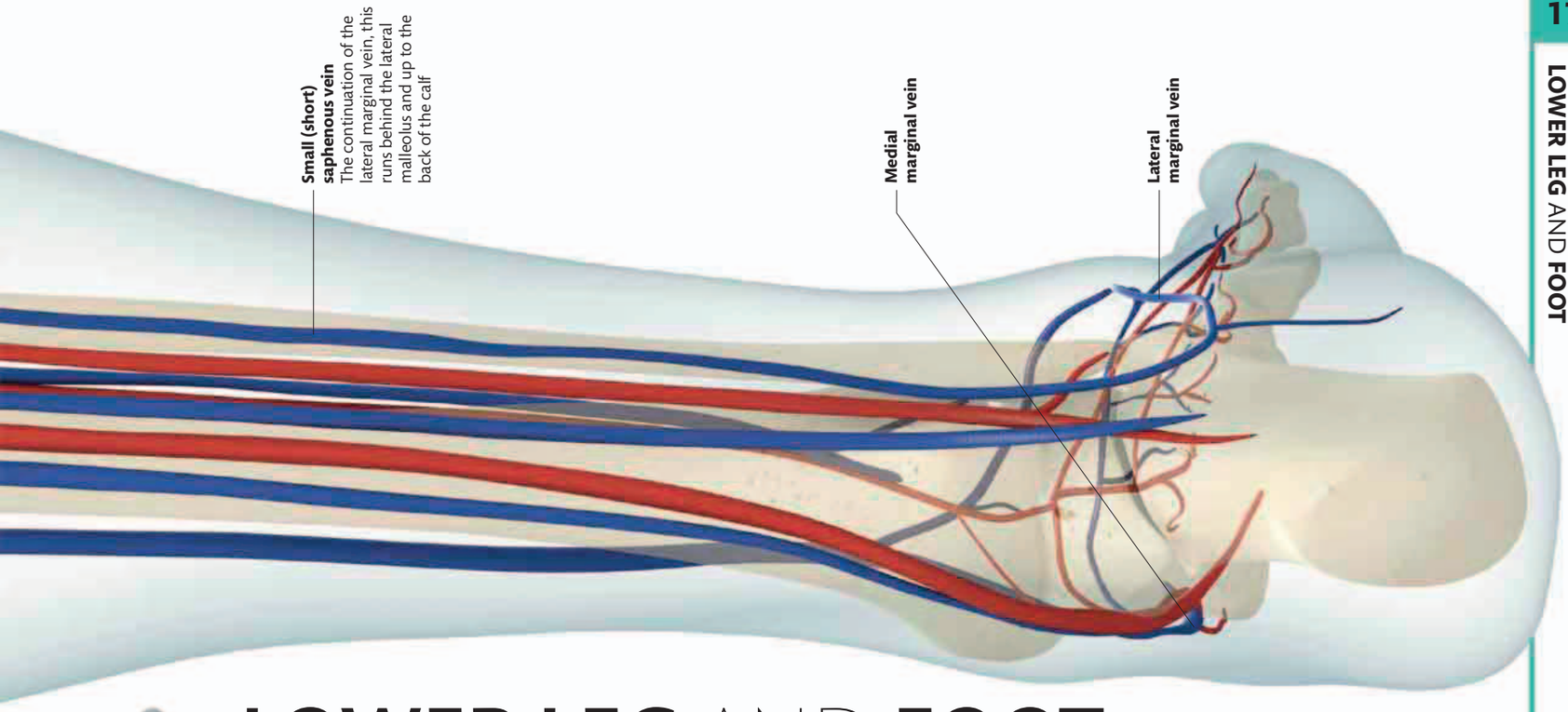

POSTERIOR (BACK)

LOWER LEG AND FOOT

The popliteal artery runs deep across the back of the knee, dividing into two branches: the anterior and posterior tibial arteries. The former runs forward, piercing the interosseous membrane between the tibia and fibula, to supply the extensor muscles of the shin. It runs down past the ankle, onto the top of the foot, as the dorsalis pedis artery. The latter gives off a peroneal branch, supplying the muscles and skin on the leg's outer side. The posterior tibial artery itself continues in the calf, running with the tibial nerve and, like the nerve, divides into plantar branches to supply the sole of the foot. A network of superficial veins on the back of the foot is drained by the saphenous veins.

ANTERIOR (FRONT)

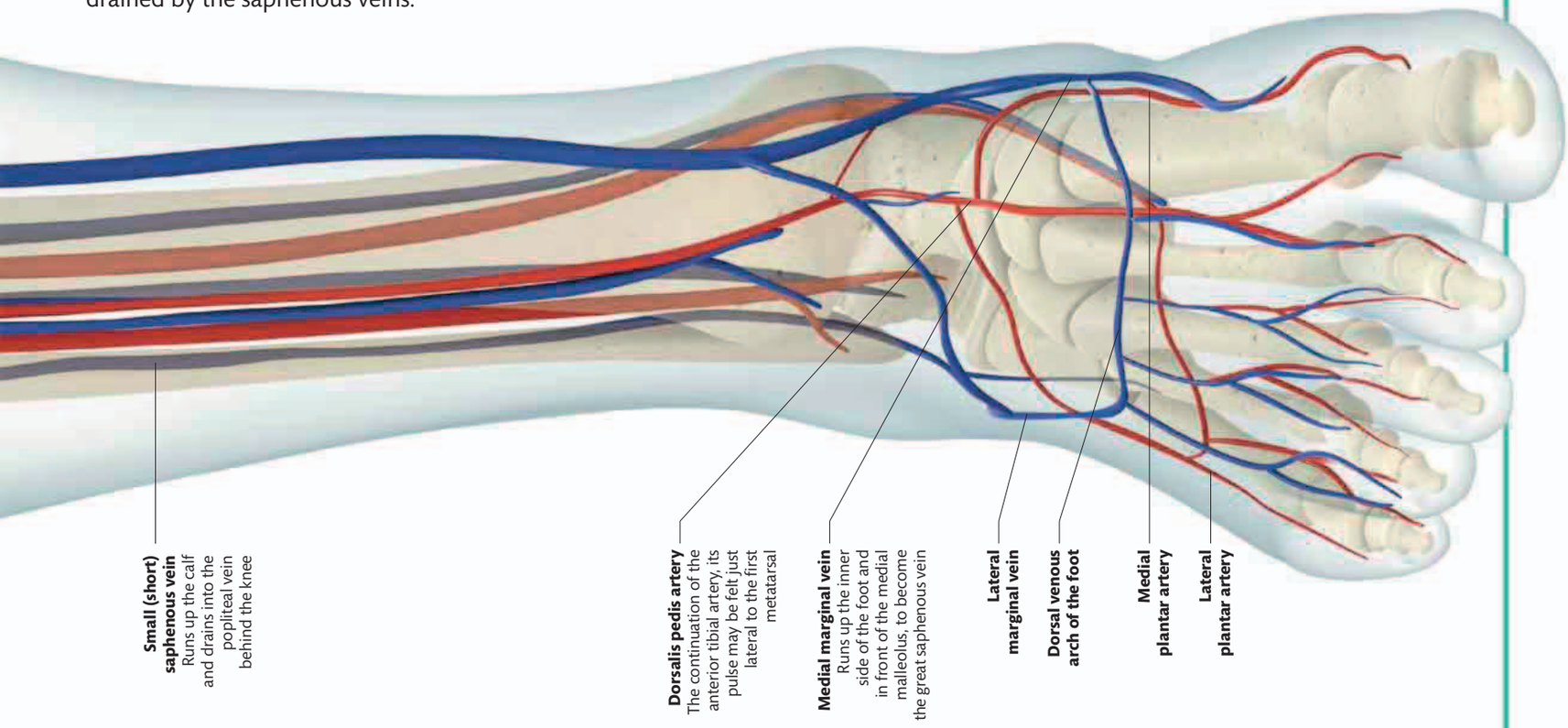

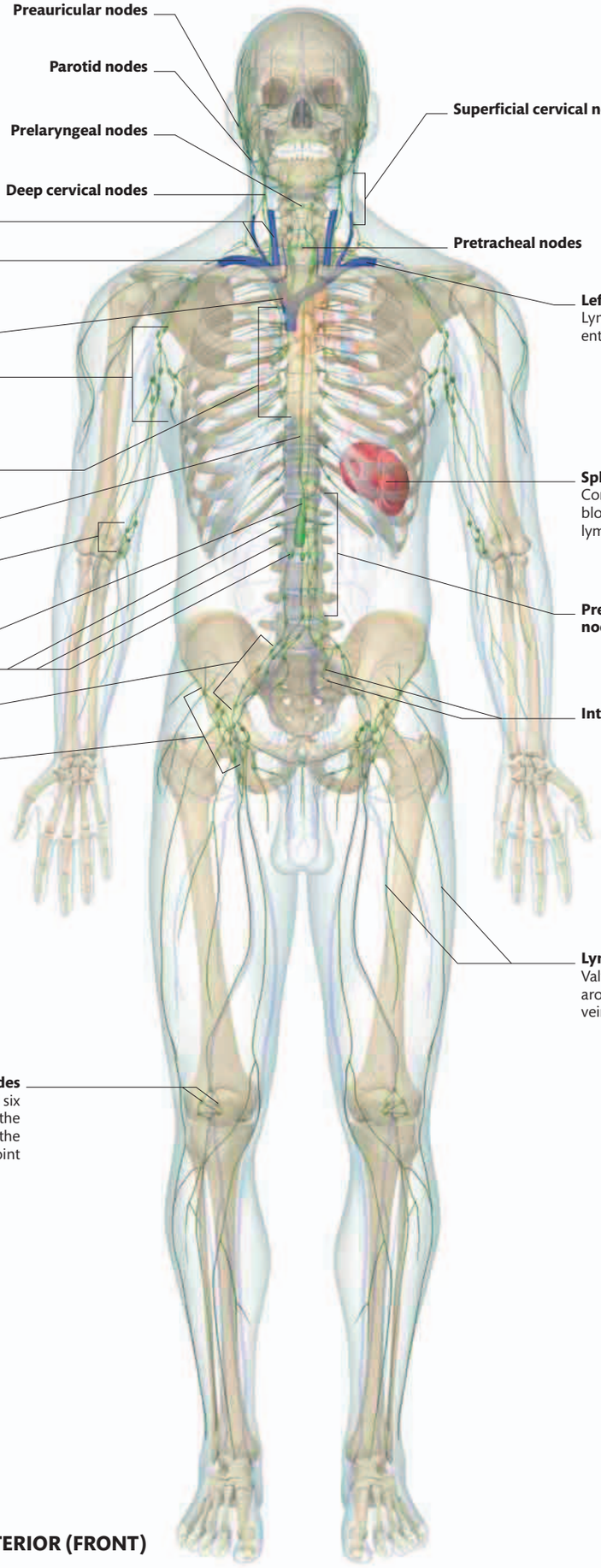
Preauricular nodes
Parotid nodes
Prelaryngeal nodes
Deep cervical nodes
Jugular veins
Right subclavian vein
Lymph from right arm, and right side of head and chest enters bloodstream here
Superior vena cava
Axillary nodes
Drain upper trunk and arm
Parasternal nodes
Thoracic duct
Supratrochlear nodes
Lymph from the hand and forearm drains to nodes at the elbow
Cisterna chyli
Lateral aortic nodes
External iliac nodes
Inguinal nodes
Popliteal nodes
A group of around six nodes sit within the popliteal fossa, at the back of the knee joint
Superficial cervical nodes
Pretracheal nodes
Left subclavian vein
Lymph from the thoracic duct enters bloodstream here
Spleen
Contains lymphocytes and filters blood; the largest organ in the lymphatic system
Pre- and para-aortic nodes
Internal iliac nodes
Lymphatics
Valved vessels transport lymph fluid around the body in a way similar to veins transporting blood
ANTERIOR (FRONT)

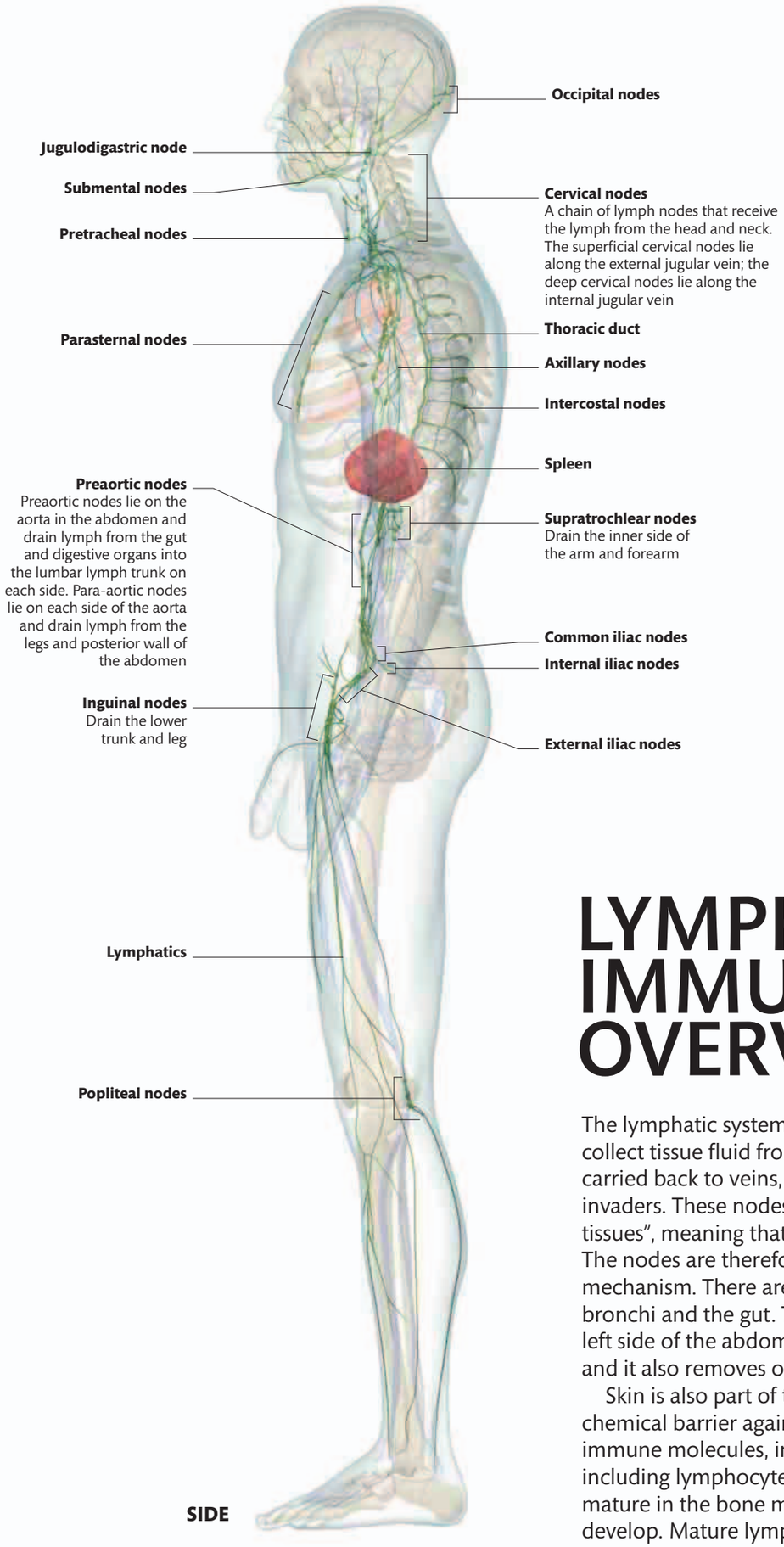

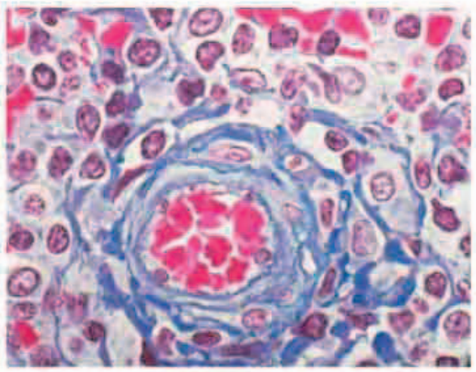

Lymphoid tissue
At a high magnification, individual lymphocytes (purple) can be seen in a section of lymphoid tissue. The blue circle in the image is an arteriole, packed full of blood cells (stained pink).

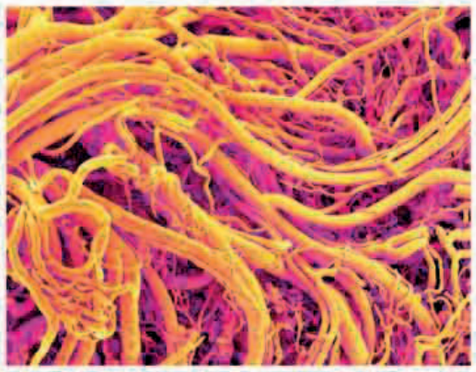

Blood vessels of lymph node
This image, produced using a scanning electron microscope, shows a resin cast of the dense network of tiny blood vessels inside a lymph node.

LYMPHATIC AND IMMUNE SYSTEM OVERVIEW

The lymphatic system consists of a network of lymphatic vessels that collect tissue fluid from the spaces between cells. Before this fluid is carried back to veins, it is delivered to lymph nodes to check for potential invaders. These nodes, like the tonsils, spleen, and thymus, are "lymphoid tissues", meaning that they contain immune cells known as lymphocytes. The nodes are therefore part of the immune system, the body's defense mechanism. There are also patches of lymphoid tissue in the walls of the bronchi and the gut. The spleen, which lies tucked under the ribs on the left side of the abdomen, has two important roles: it is a lymphoid organ, and it also removes old red blood cells from circulation.

Skin is also part of the immune system as it forms a physical and chemical barrier against infections. The formation of some important immune molecules, including antibodies, and a range of immune cells, including lymphocytes, happens in the bone marrow. Some lymphocytes mature in the bone marrow, whereas others move to the thymus to develop. Mature lymphocytes stay in the lymph nodes, where they perform their function.

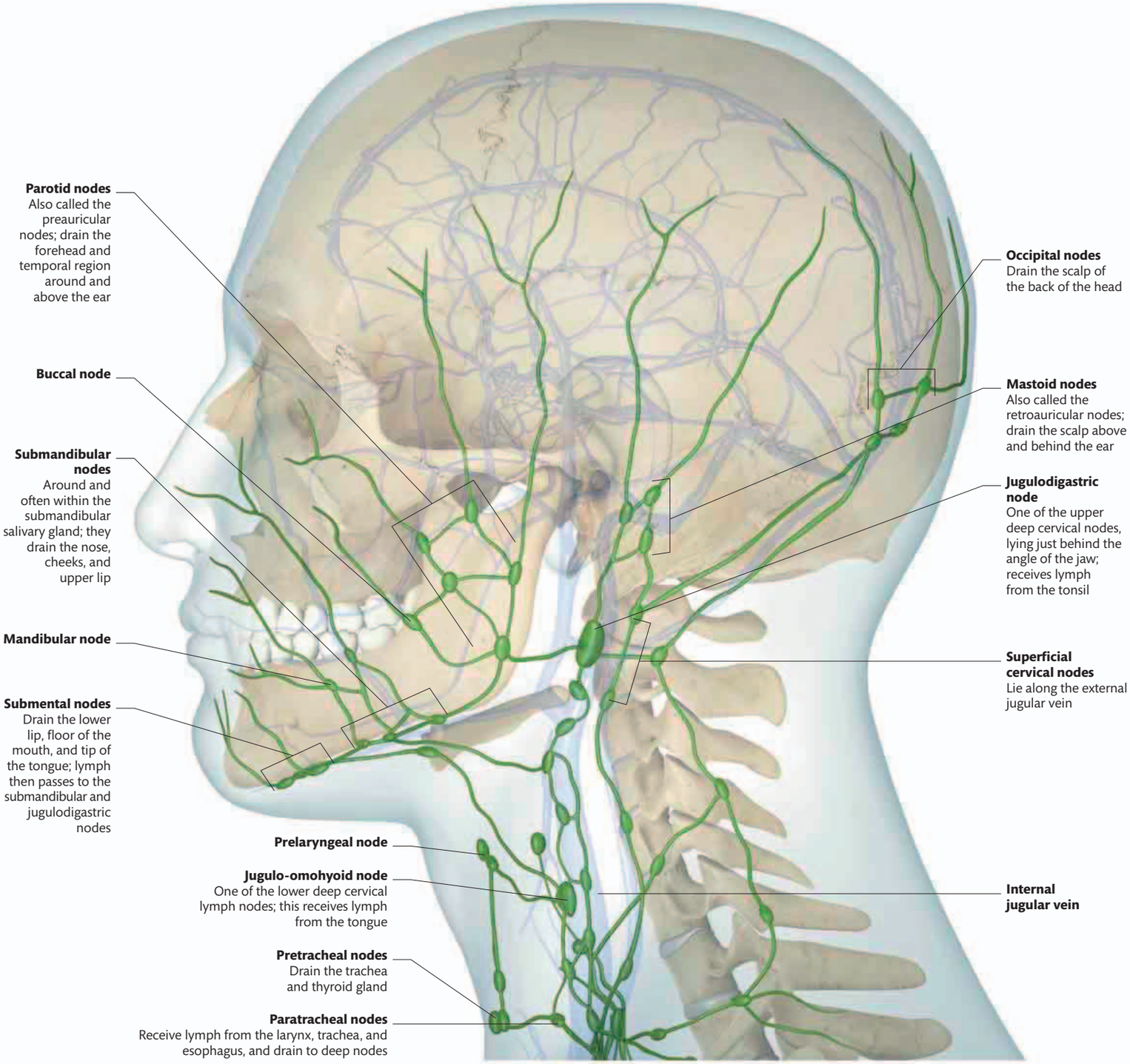

LYMPH NODES OF HEAD

HEAD AND NECK

A ring of lymph nodes lies close to the skin where the head meets the neck, from the occipital nodes (against the skull at the back) to the submandibular and submental nodes (which are tucked under the jaw). Superficial nodes lie along the sides and front of the neck, and deep nodes are clustered around the internal jugular vein, under cover of sternocleidomastoid muscle. Lymph from all other nodes passes to these deep ones, then into the jugular lymphatic trunk before draining back into veins in the base of the neck. Lymphoid tissue, in the form of the palatine, pharyngeal, and lingual tonsils, forms a protective ring around the upper parts of the respiratory and digestive tracts.

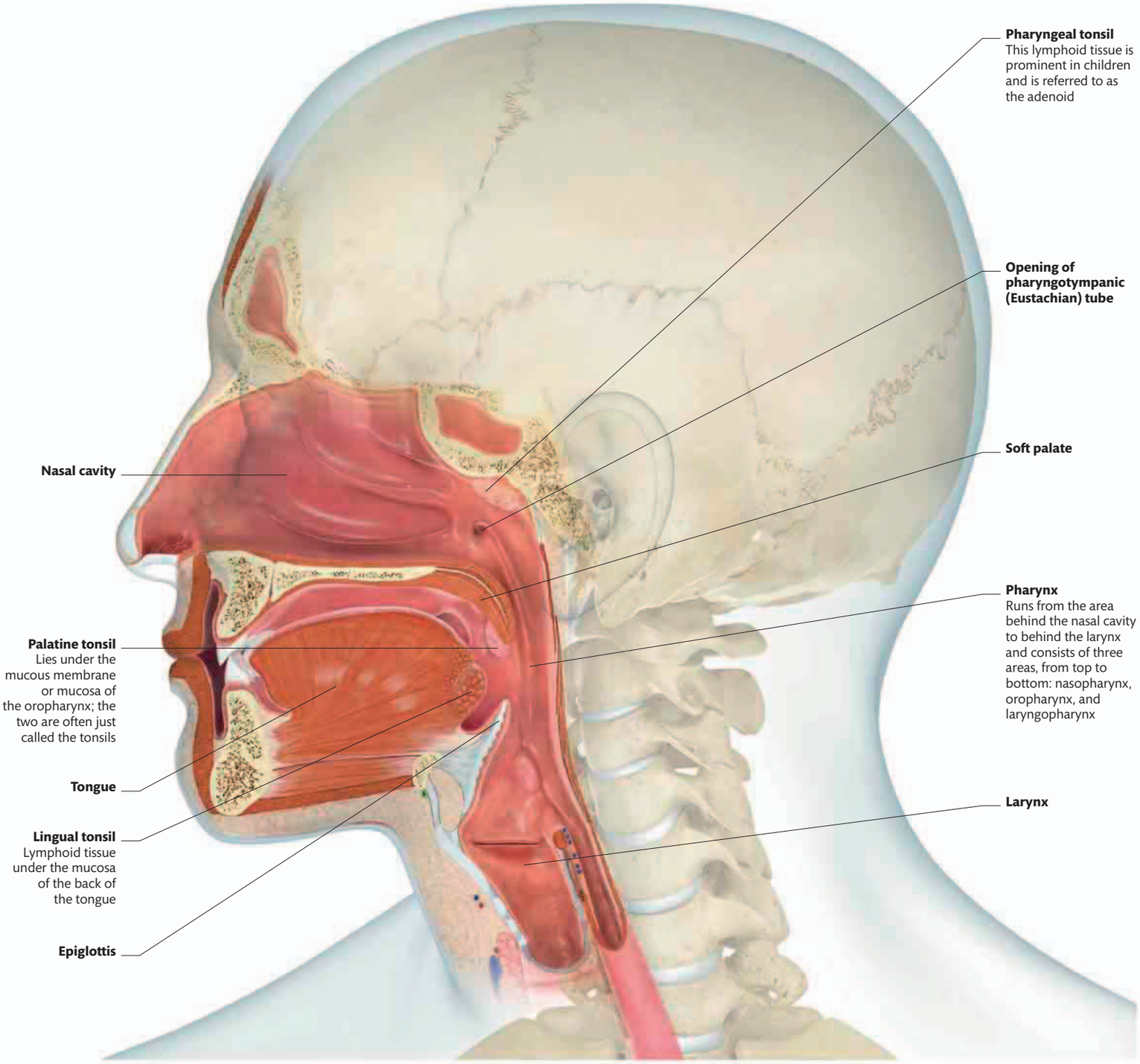

LOCATION OF TONSILS

Right lymphatic duct
Lymph from the right arm and the right side of the neck and thorax drains into the junction of the right internal jugular and subclavian veins

Parasternal nodes
Also called internal thoracic nodes; these lie in the gaps between the ribs, either side of the sternum on the inside of the ribcage; they drain some of the lymph from the front of the thorax – including from the breast in a woman

Axillary nodes
Receive lymph from superficial tissues of the thorax, upper limb, and breast

Intercostal nodes
Sitting in the intercostal spaces between the ribs at the back of the ribcage, these drain lymph from the deeper tissues at the sides and back of the thorax

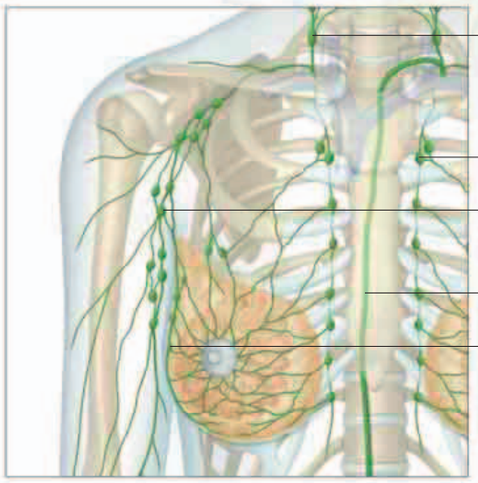

ANTERIOR (FRONT) / FEMALE

ANTERIOR (FRONT) / MALE

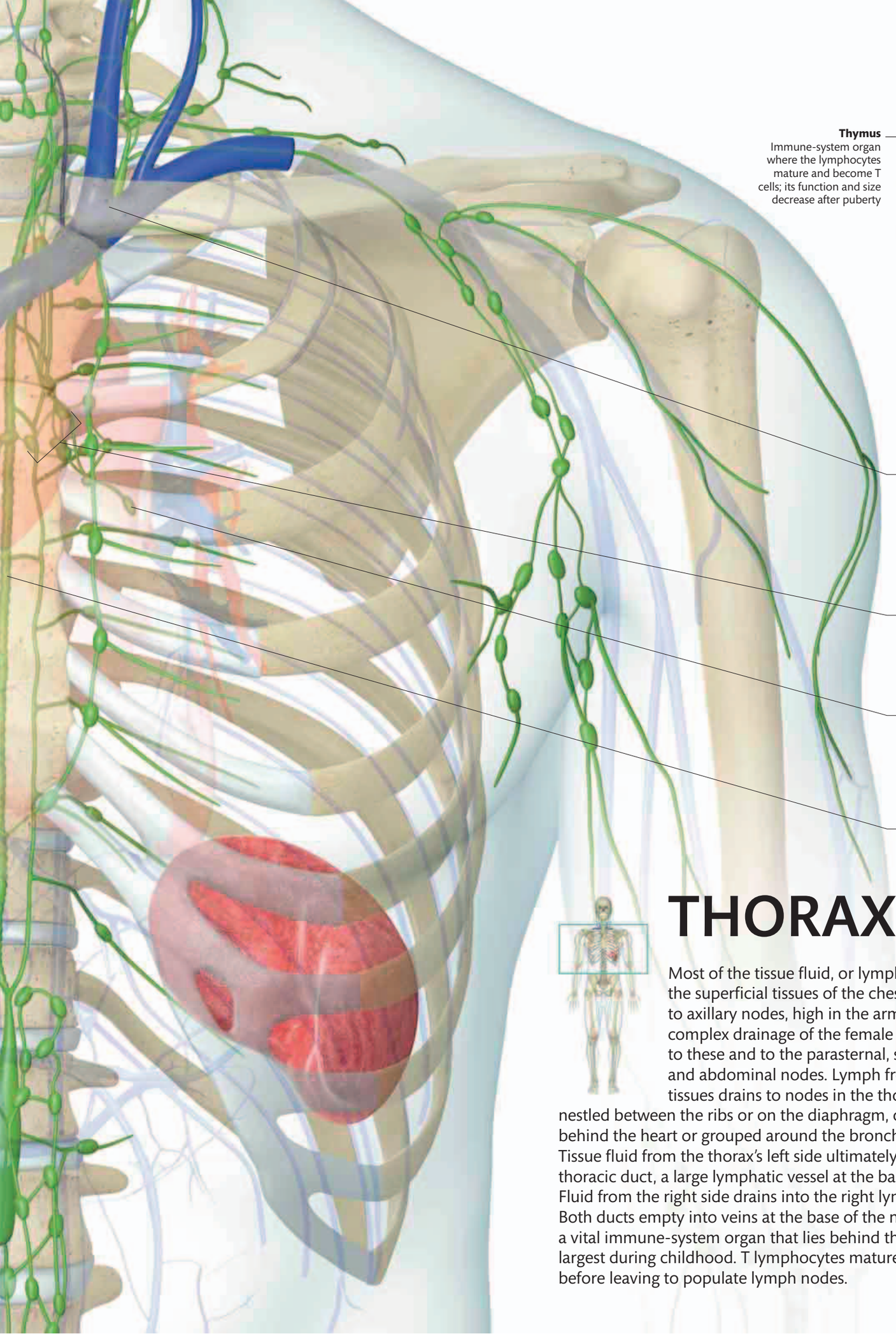

THORAX

Most of the tissue fluid, or lymph, from the superficial tissues of the chest drains to axillary nodes, high in the armpits. The complex drainage of the female breast passes to these and to the parasternal, supraclavicular, and abdominal nodes. Lymph from deeper tissues drains to nodes in the thorax, some nestled between the ribs or on the diaphragm, others tucked behind the heart or grouped around the bronchi and trachea. Tissue fluid from the thorax's left side ultimately drains into the thoracic duct, a large lymphatic vessel at the back of the thorax. Fluid from the right side drains into the right lymphatic duct. Both ducts empty into veins at the base of the neck. The thymus, a vital immune-system organ that lies behind the sternum, is largest during childhood. T lymphocytes mature in the thymus before leaving to populate lymph nodes.

ABDOMEN AND PELVIS

The deep lymph nodes of the abdomen are clustered around arteries. Nodes lying along each side of the aorta receive lymph from paired structures, such as the muscles of the abdominal wall, the kidneys and adrenal glands, and the testes or ovaries. Iliac nodes collect lymph returning from the legs and pelvis. Nodes clustered around the branches on the front of the aorta collect lymph from the gut and abdominal organs. Eventually, all this lymph from the legs, pelvis, and abdomen passes into a swollen lymphatic vessel called the cisterna chyli; this narrows down to become the thoracic duct, which runs up into the chest. Most lymph nodes are small, bean-sized structures, but the abdomen also contains a large and important organ of the immune system—the spleen.

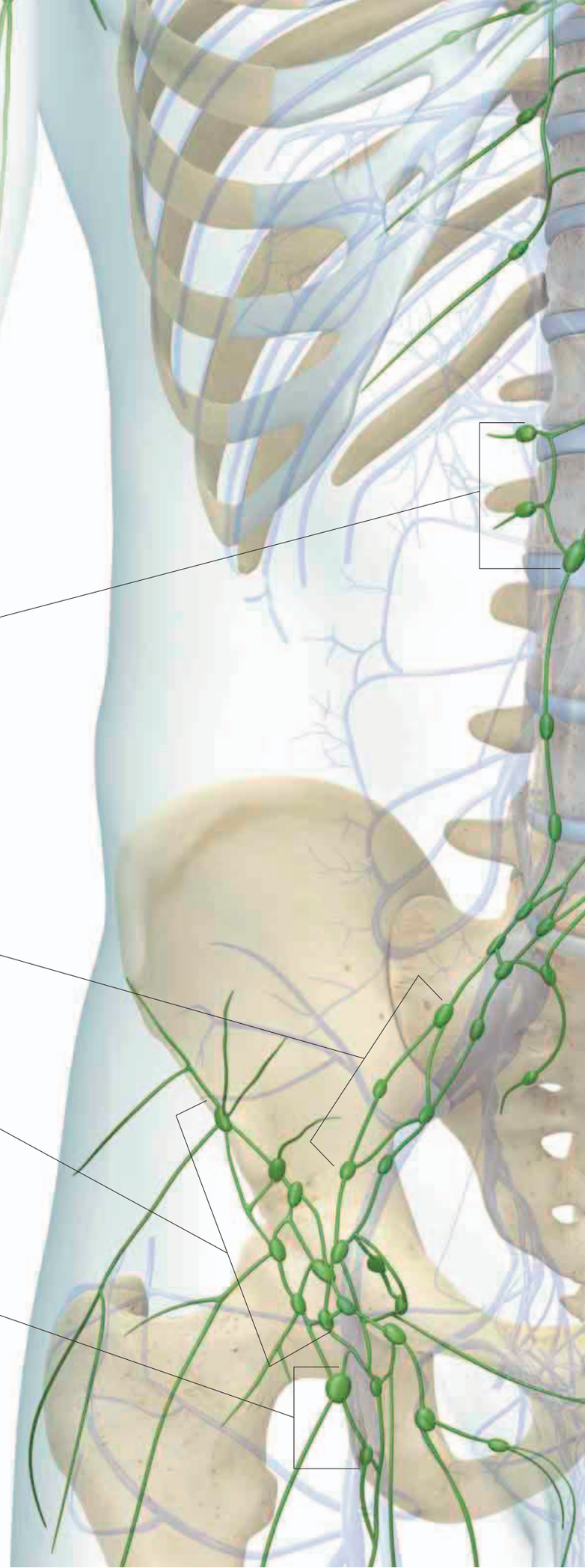

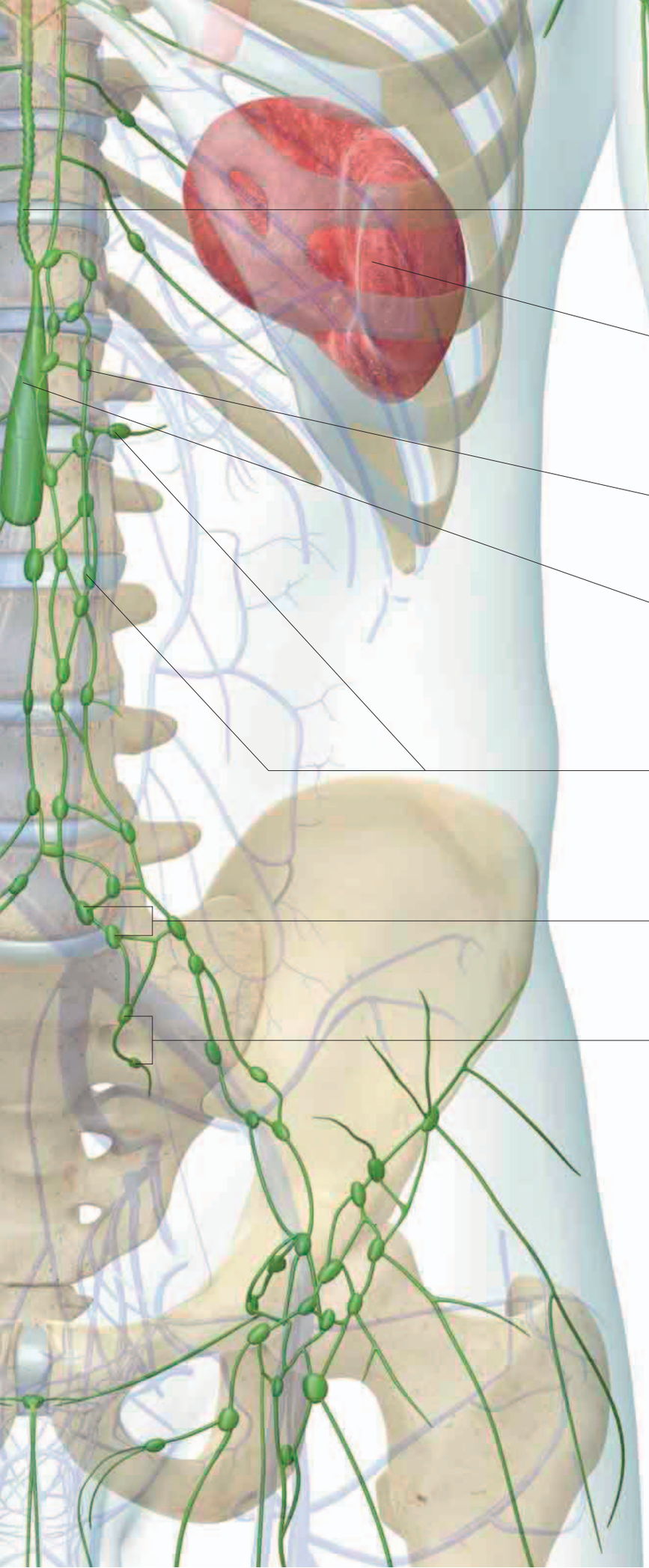
Thoracic duct
Spleen
Contains red pulp, into which old, tired red blood cells are removed from circulation, and also white pulp, which is full of lymphocytes, making it much like a massive lymph node
Celiac nodes
Drain lymph from the organs supplied by the celiac artery, including the liver, pancreas, and stomach
Cisterna chyli
Formed by the confluence of the main lymph trunks—the lumbar and intestinal trunks—in the abdomen. This leads into the thoracic duct. In Greek, the name means juice reservoir
Mesenteric nodes
Nestled around the origins of the superior and inferior mesenteric arteries from the aorta, these drain most of the lymph from the intestines
Common iliac nodes
Receive lymph from the external and internal iliac nodes and drain up to the lateral aortic nodes
Internal iliac nodes
Drain lymph from the organs of the pelvis
ANTERIOR (FRONT)

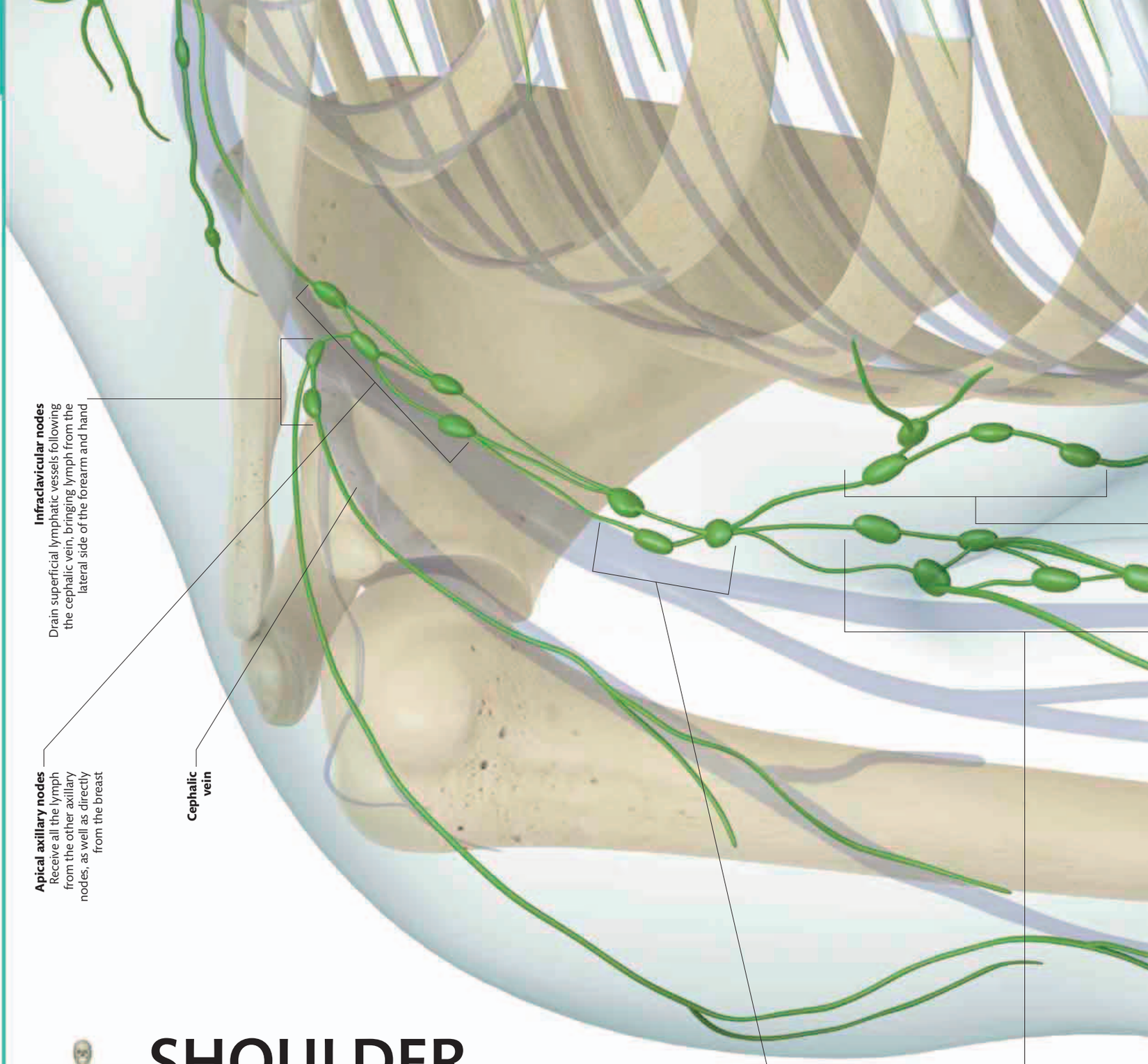

SHOULDER AND UPPER ARM

Ultimately, all the lymph from the hand, forearm, and arm drains to the axillary nodes in the armpit. But there are a few nodes, lower in the arm, that lymph may pass through on its way to the axilla. The supratrochlear nodes lie in the subcutaneous fat on the inner arm, above the elbow. They collect lymph that has drained from the medial side of the hand and forearm. The infraclavicular nodes, lying along the cephalic vein, below the clavicle, receive lymphatics draining from the thumb and the lateral side of the forearm and arm. Axillary nodes drain lymph from the arm and receive it from the chest wall. They may become infiltrated with cancerous cells spreading from a tumor in the breast.

Anterior axillary nodes
Drain lymph from the trunk above the umbilicus, including the front of the chest and breast
Supratrochlear nodes
Drain superficial tissues on the inner (medial) side of the hand and forearm
ANTERIOR (FRONT)

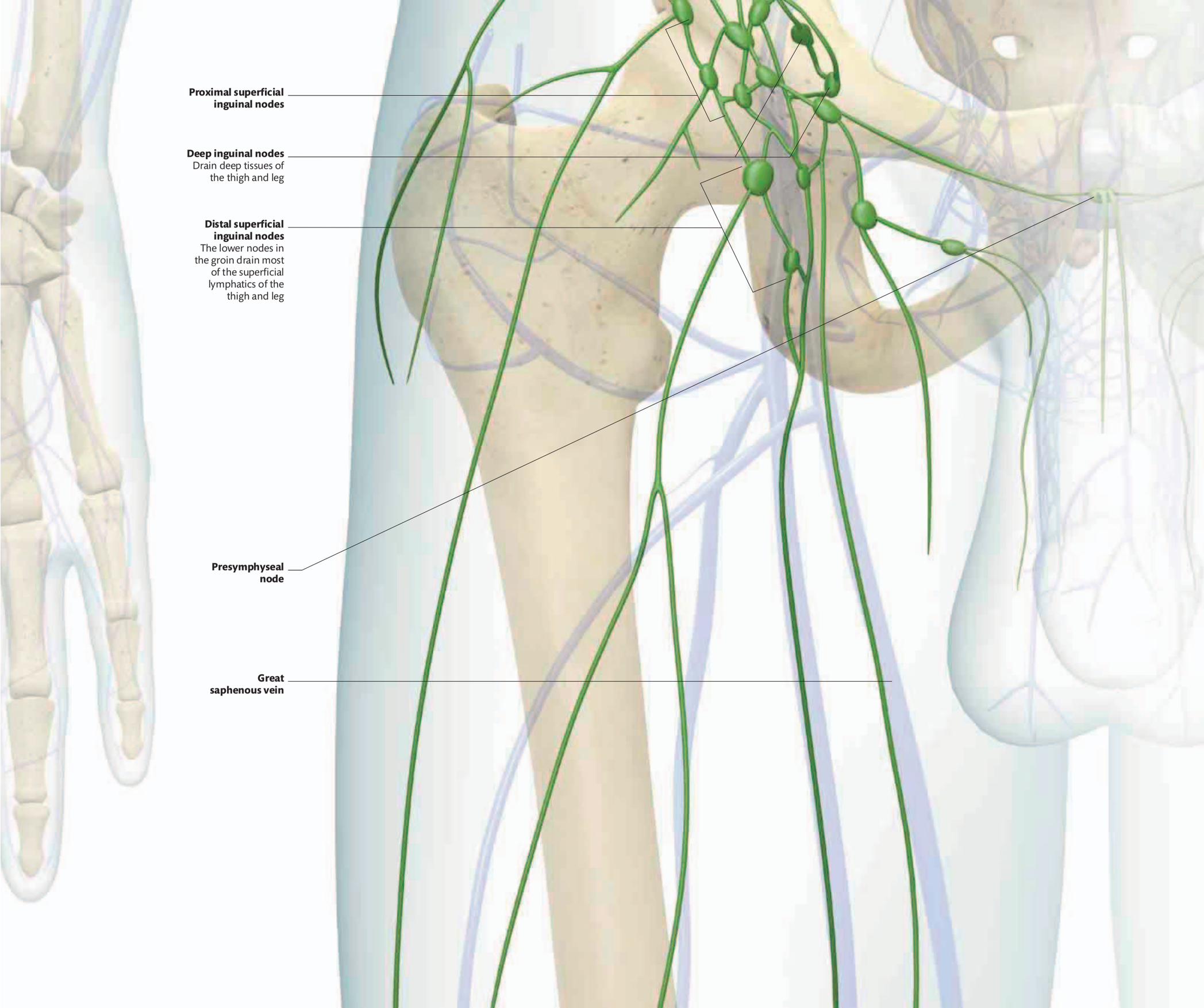
Proximal superficial inguinal nodes
Deep inguinal nodes
Drain deep tissues of the thigh and leg
Distal superficial inguinal nodes
The lower nodes in the groin drain most of the superficial lymphatics of the thigh and leg
Presymphyseal node
Great saphenous vein

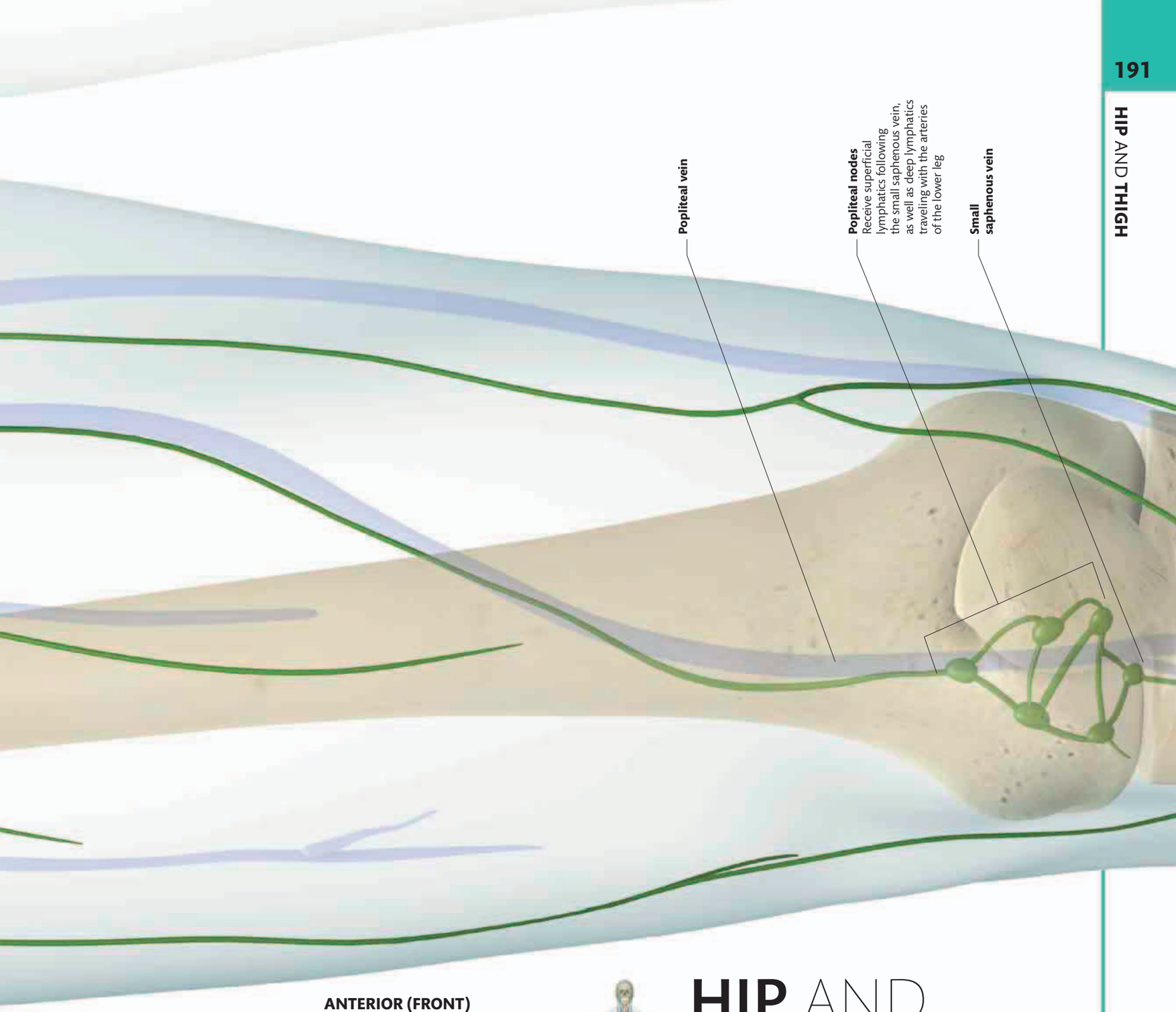

HIP AND THIGH

Most lymph from the thigh, leg, and foot passes through the inguinal group of lymph nodes, which are in the groin. But lymph from the deep tissues of the buttock passes straight to nodes inside the pelvis (see pp.184–85), along the internal and common iliac arteries. Eventually, all the lymph from the leg reaches the lateral aortic nodes, on the back wall of the abdomen. As in the arm, there are groups of nodes clustered around points at which superficial veins drain into deep veins. Popliteal nodes are close to the drainage of the small saphenous vein into the popliteal vein, while the superficial inguinal nodes lie close to the great saphenous vein, just before it empties into the femoral vein.

DIGESTIVE SYSTEM **OVERVIEW**

The digestive system comprises the organs that enable us to take in food, break it down physically and chemically, extract its useful nutrients, and excrete what we don't need. This process begins in the mouth, where the teeth, tongue, and saliva work together to form food into a moist ball that can be swallowed. The mouth, pharynx, stomach, intestines, rectum, and anal canal form a long tube that is referred to as the digestive tract. It usually takes between one and two days for ingested food to travel all the way from the mouth to the anus. Other organs—including the salivary glands, liver, gallbladder, and pancreas—complete the digestive system.

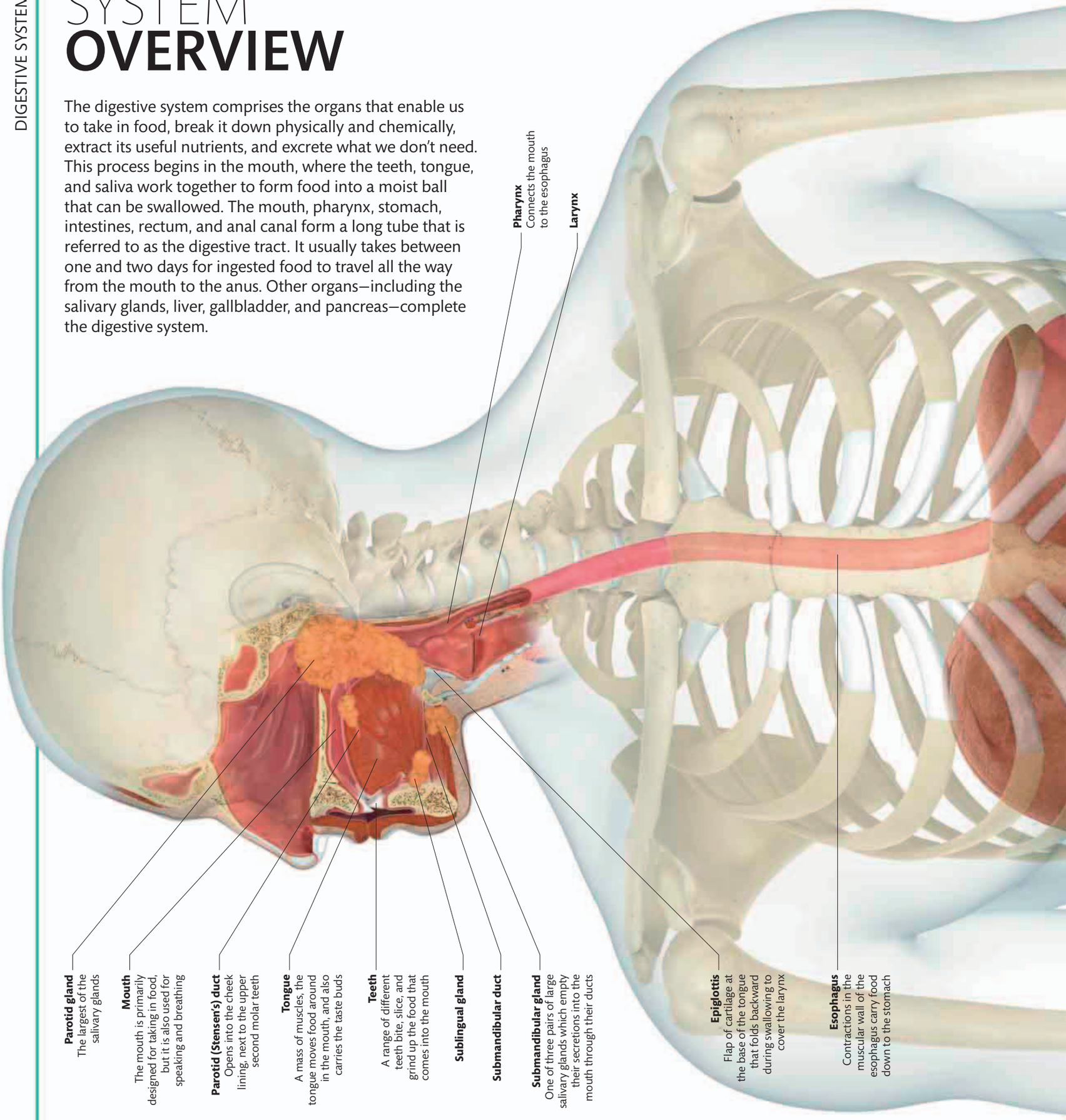

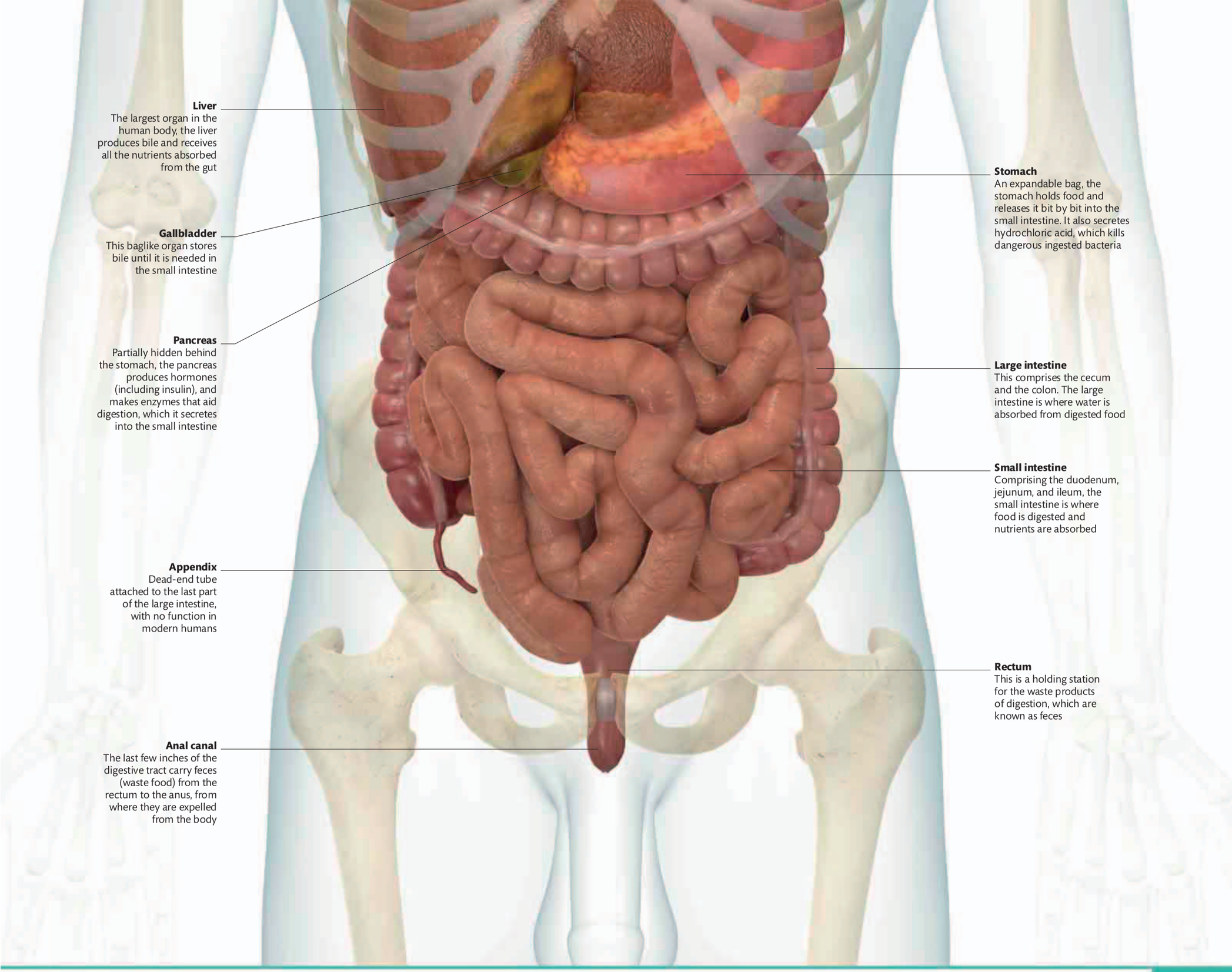
Liver
The largest organ in the human body, the liver produces bile and receives all the nutrients absorbed from the gut
Gallbladder
This baglike organ stores bile until it is needed in the small intestine
Pancreas
Partially hidden behind the stomach, the pancreas produces hormones (including insulin), and makes enzymes that aid digestion, which it secretes into the small intestine
Appendix
Dead-end tube attached to the last part of the large intestine, with no function in modern humans
Anal canal
The last few inches of the digestive tract carry feces (waste food) from the rectum to the anus, from where they are expelled from the body
Stomach
An expandable bag, the stomach holds food and releases it bit by bit into the small intestine. It also secretes hydrochloric acid, which kills dangerous ingested bacteria
Large intestine
This comprises the cecum and the colon. The large intestine is where water is absorbed from digested food
Small intestine
Comprising the duodenum, jejunum, and ileum, the small intestine is where food is digested and nutrients are absorbed
Rectum
This is a holding station for the waste products of digestion, which are known as feces

SAGITTAL SECTION

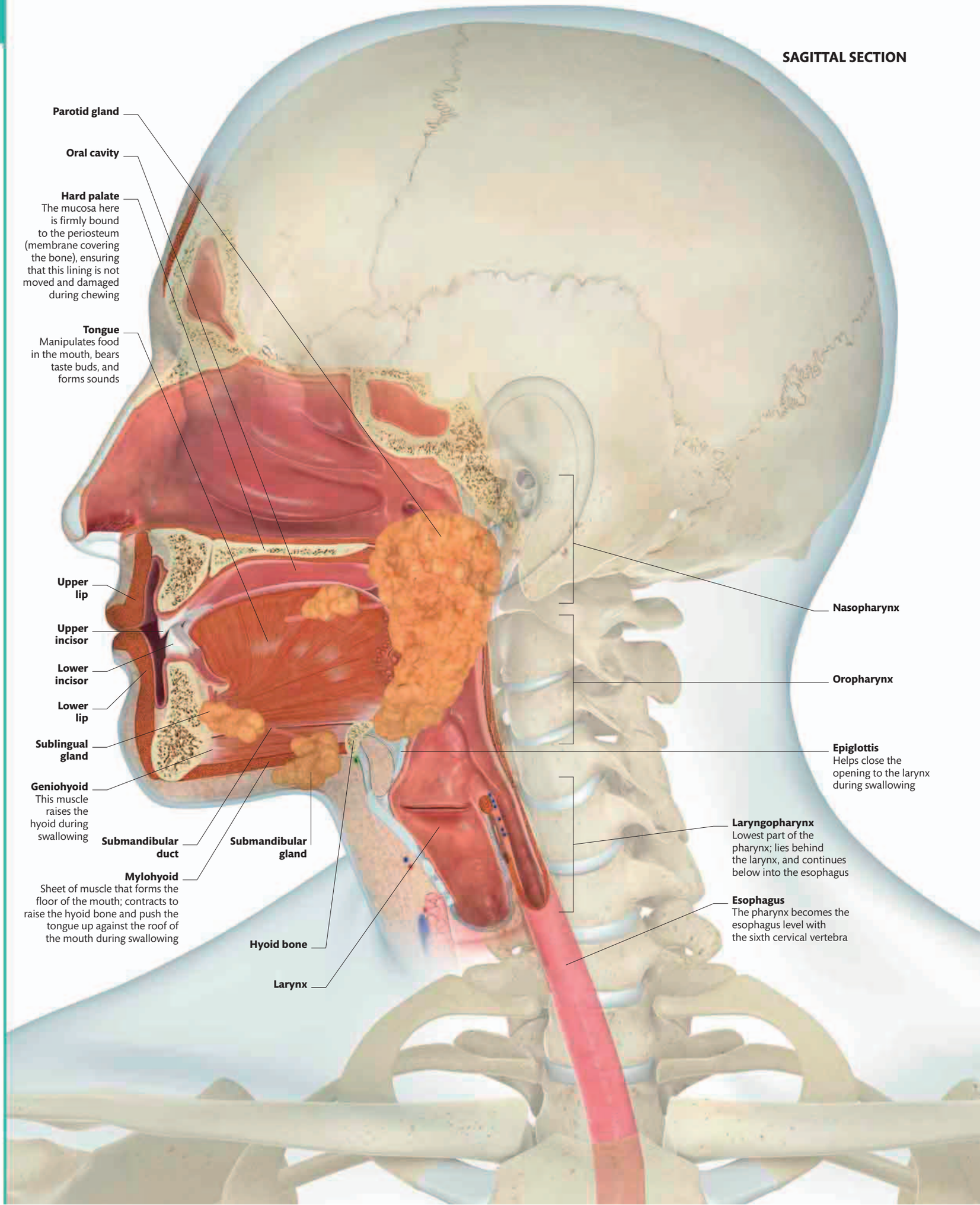

HEAD AND NECK

The mouth is the first part of the digestive tract, and it is here that the processes of mechanical and chemical digestion get underway. Your teeth grind each mouthful, and you have three pairs of major salivary glands—parotid, submandibular, and sublingual—that secrete saliva through ducts into the mouth. Saliva contains digestive enzymes that begin to chemically break down the food in your mouth. The tongue manipulates the food, and also has taste buds that allow you to quickly make the important distinction between delicious food and potentially harmful toxins. As you swallow, the tongue pushes up against the hard palate, the soft palate seals off the airway, and the muscular tube of the pharynx contracts in a wave to push the ball of food down into the esophagus, ready for the next stage of its journey.

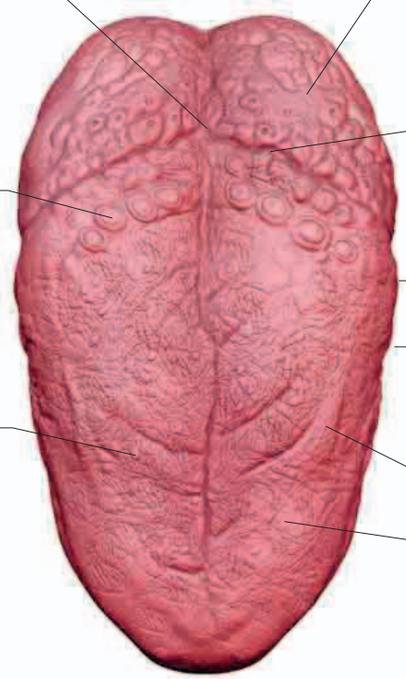

TONGUE

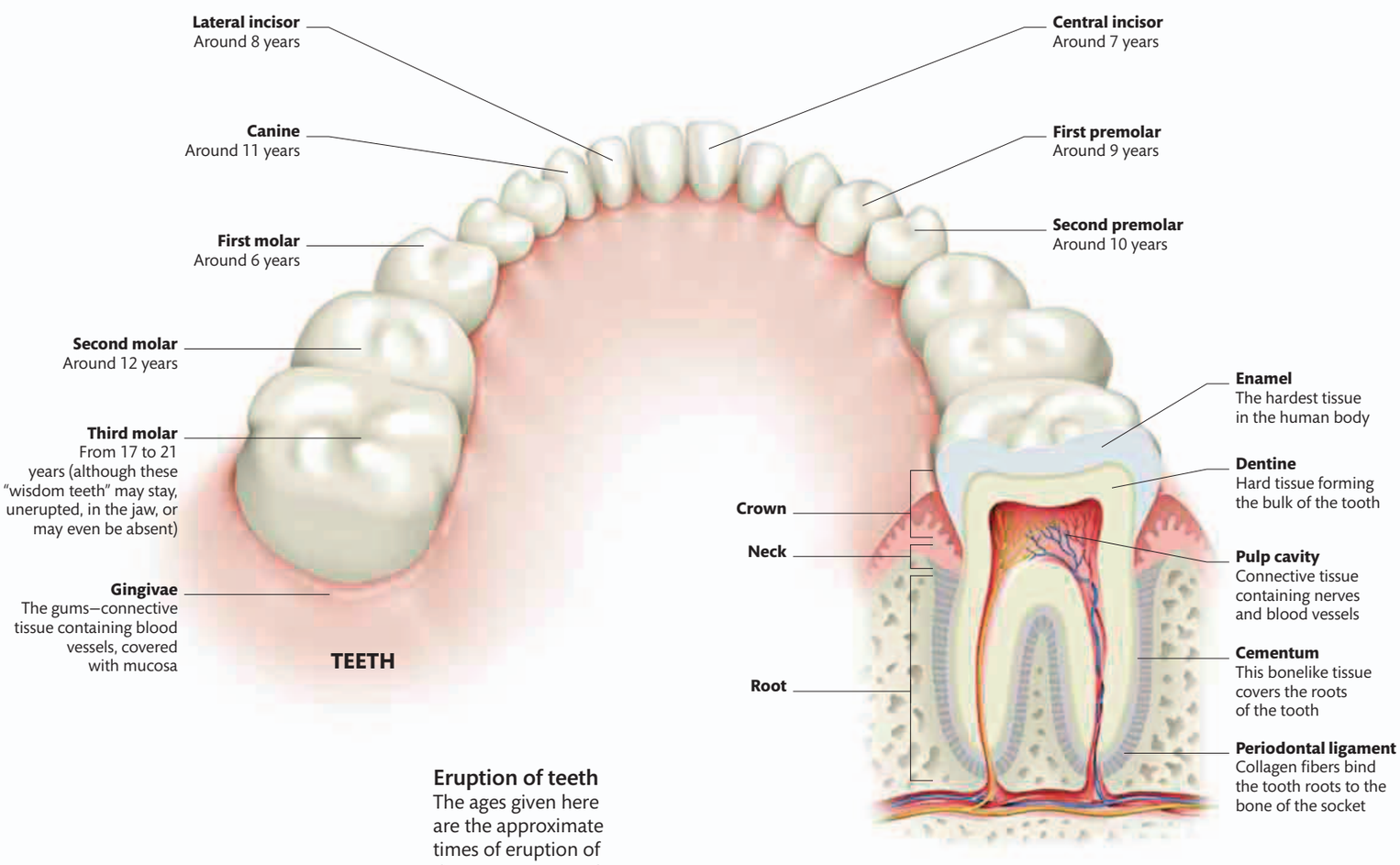

TEETH

Eruption of teeth
The ages given here are the approximate times of eruption of the permanent teeth.

Esophagus
In the neck, the esophagus lies behind the trachea

Thoracic part of the esophagus
The esophagus is slightly constricted here by the left main bronchus, which crosses in front of it

Liver
Lies under the right dome of the diaphragm, and largely under cover of the ribs

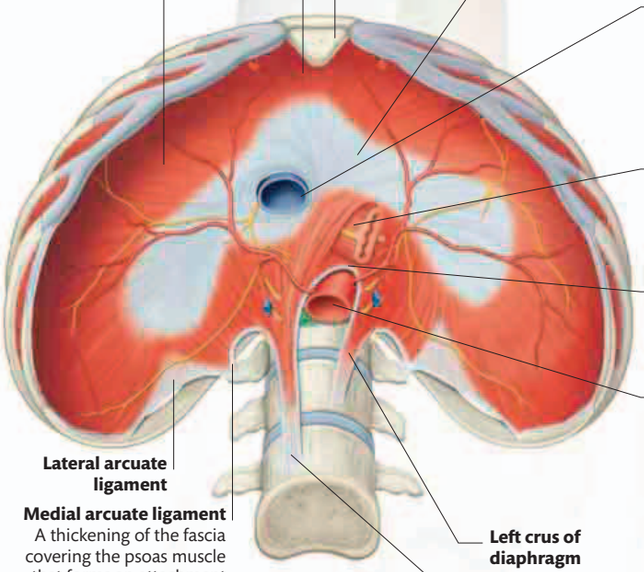

DIAPHRAGM FROM BELOW

ANTERIOR (FRONT)

Fundus of stomach
The upper part of the stomach lies below the left dome of the diaphragm, under the ribs

THORAX

There are several large tubes crammed into the space behind the heart. These include the descending aorta, the azygos vein, and the lymphatic duct, but also a part of the digestive tract—the esophagus. This tube of smooth muscle starts in the neck as a continuation of the pharynx. It runs down through the thorax, slightly to the left of center, and pierces through the diaphragm level with the tenth thoracic vertebra. A couple of centimeters below this, it empties into the stomach and ends. The esophagus, like much of the digestive tract, has an outer layer of longitudinal muscle and an inner layer of circular muscle within its wall. During swallowing, a wave of constriction passes downward to push food or fluid down into the stomach.

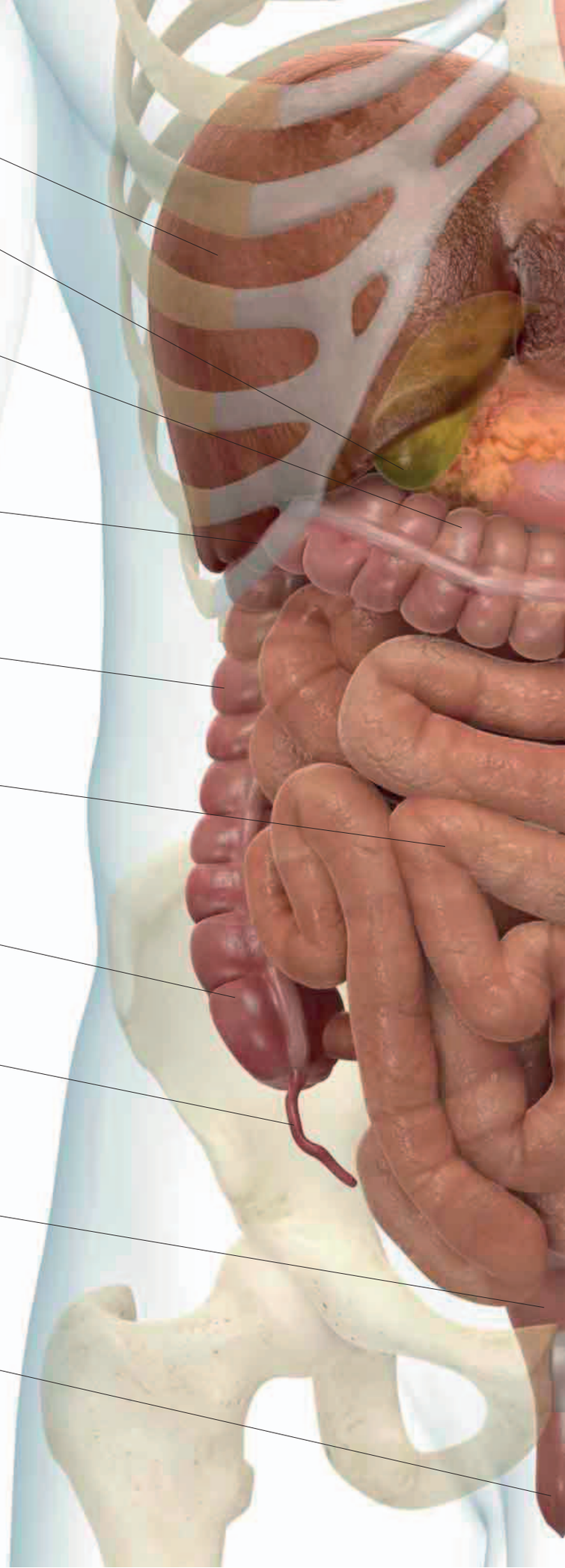

Right lobe of liver

Fundus of gallbladder
Bottom of the baglike gallbladder, which just sticks out under the liver

Transverse colon
Hanging down below the liver and stomach, this part of the colon has a mesentery (fold of the peritoneum that connects the intestines to the dorsal abdominal wall) through which its blood vessels and nerves travel

Hepatic flexure of colon
Junction between the ascending and transverse colon, tucked under the liver

Ascending colon
This part of the large intestine is firmly bound to the back wall of the abdomen

Ileum
Lying mainly in the suprapubic region of the abdomen, this part of the small intestine is about 13ft (4m) long; ileum simply means entrails in Latin

Cecum
First part of the large intestine, lying in the right iliac fossa of the abdomen

Appendix
Properly known as the vermiform (wormlike) appendix; usually a few centimeters long, it is full of lymphoid tissue, and thus forms part of the gut's immune system

Rectum
About 4¾in (12cm) long, this penultimate part of the gut is stretchy; it can expand to store faeces, until a convenient time for emptying presents itself

Anal canal
Muscular sphincters in and around the anal canal keep it closed; the sphincters relax during defecation, as the diaphragm and abdominal wall muscles contract to raise pressure in the abdomen and force the feces out

ANTERIOR (FRONT)

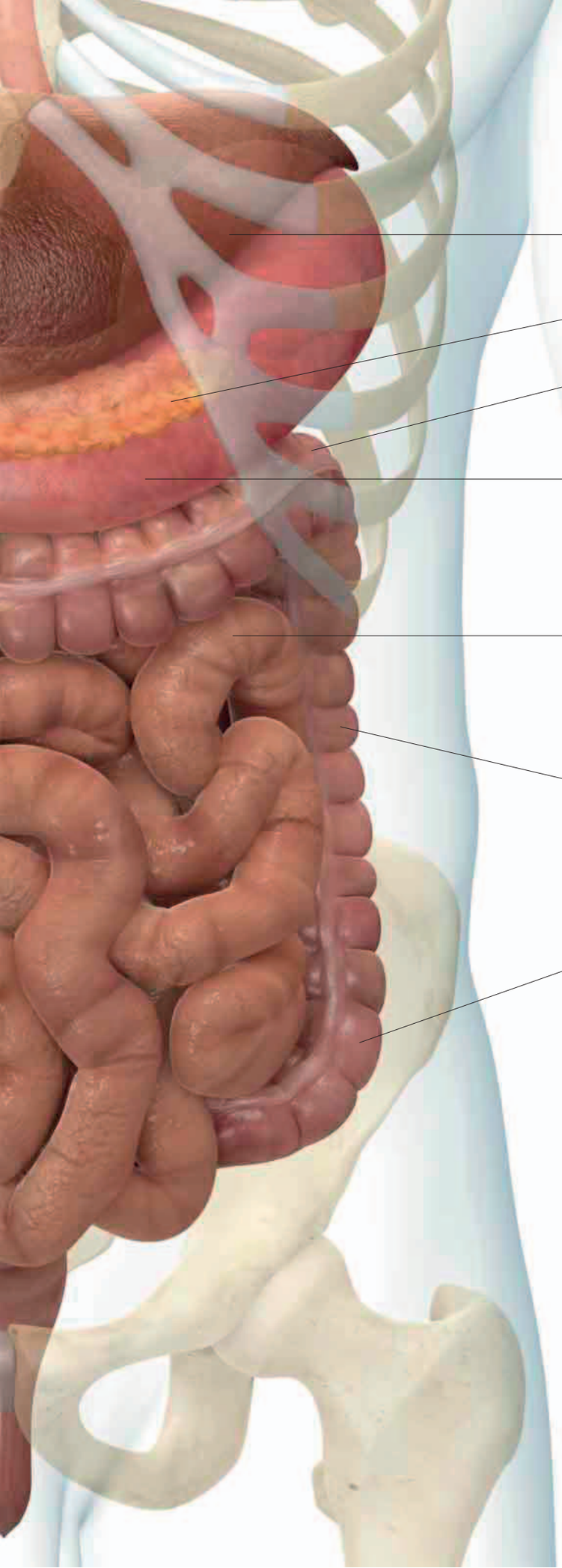

ABDOMEN AND PELVIS

With the organs in situ, it is clear how much the abdominal cavity extends up under the ribs. The upper abdominal organs—the liver, stomach, and spleen—are largely under cover of the ribcage. This gives them some protection, but it also means that they are vulnerable to injury if a lower rib is fractured. The large intestine forms an M shape in the abdomen, starting with the cecum low down on the right, and the ascending colon running up the right flank and tucking under the liver. The transverse colon hangs down below the liver and stomach, and the descending colon runs down the left side of the abdomen. This becomes the S-shaped sigmoid colon, which runs down into the pelvis to become the rectum. The coils of the small intestine occupy the middle of the abdomen.

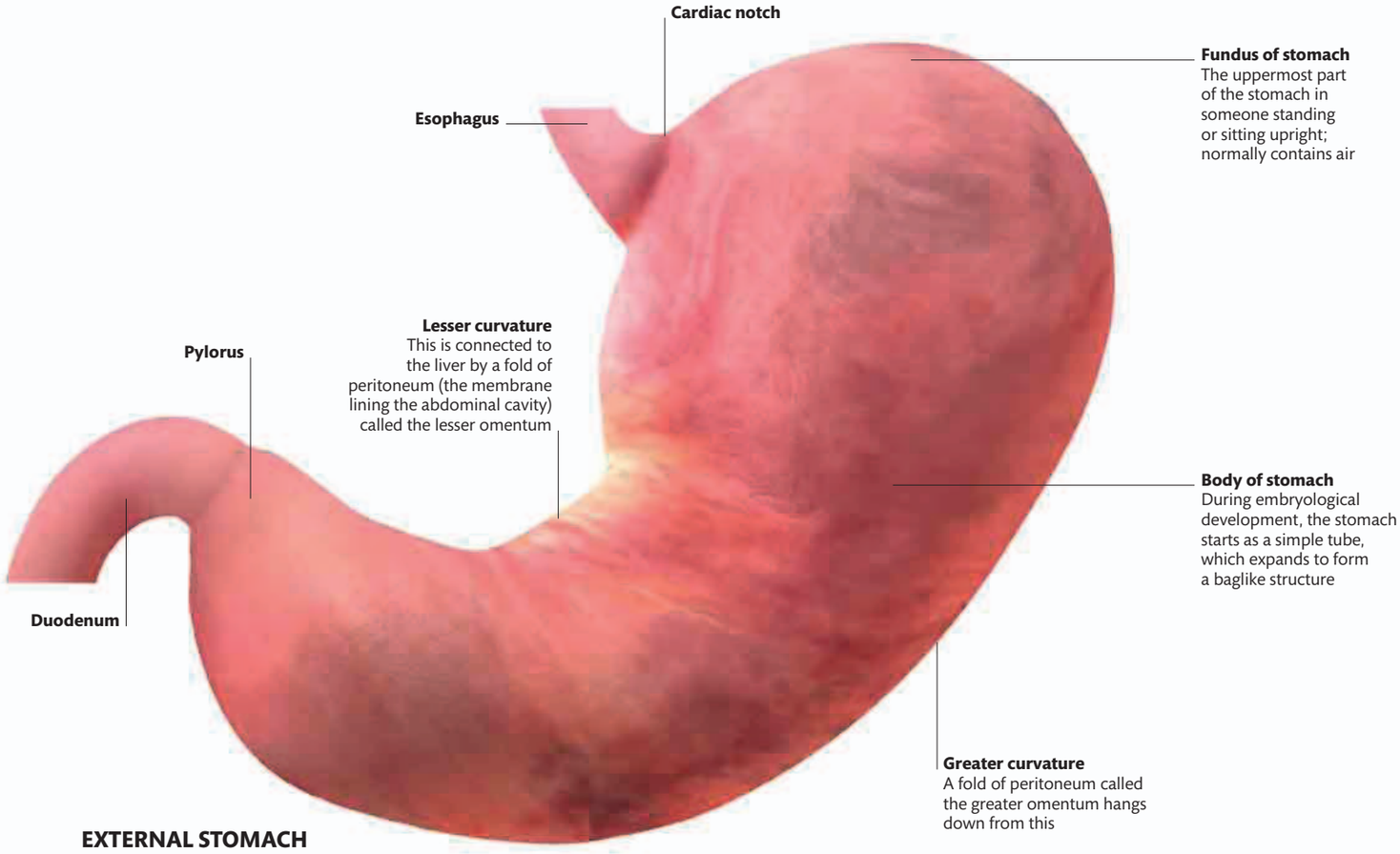

EXTERNAL STOMACH

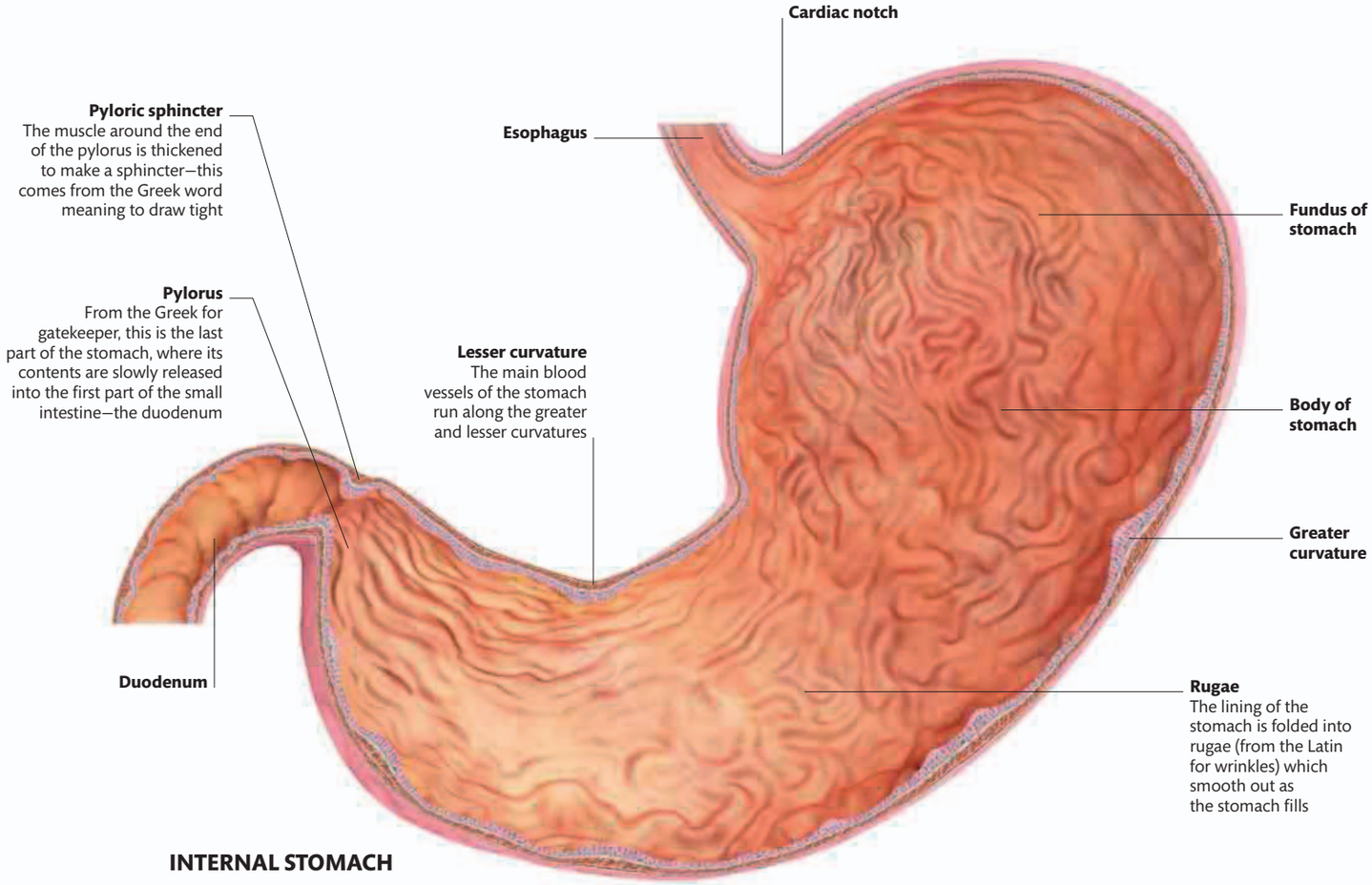

INTERNAL STOMACH

STOMACH AND INTESTINES

The stomach is a muscular bag, where food is held before moving on to the intestines. Inside the stomach, food is exposed to a cocktail of hydrochloric acid, which kills off bacteria, and protein-digesting enzymes. The layered muscle of the stomach wall contracts to churn up its contents. Semidigested food is released from the stomach into the first part of the small intestine, the duodenum, where bile and pancreatic juices are added. Contractions in the intestine wall then push the liquid food into the jejunum and ileum, where digestion continues. What is left passes into the cecum, the beginning of the large intestine. In the colon, the next part of the large intestine, water is absorbed so that the gut contents become more solid. The resulting feces pass into the rectum, where they are stored until excretion.

SMOOTH MUSCLE STRUCTURE

Functions of the gut, blood vessels, and respiratory tract are carried out involuntarily, at a subconscious level, with the help of a special type of muscle called smooth muscle. This is supplied by autonomic motor nerves.

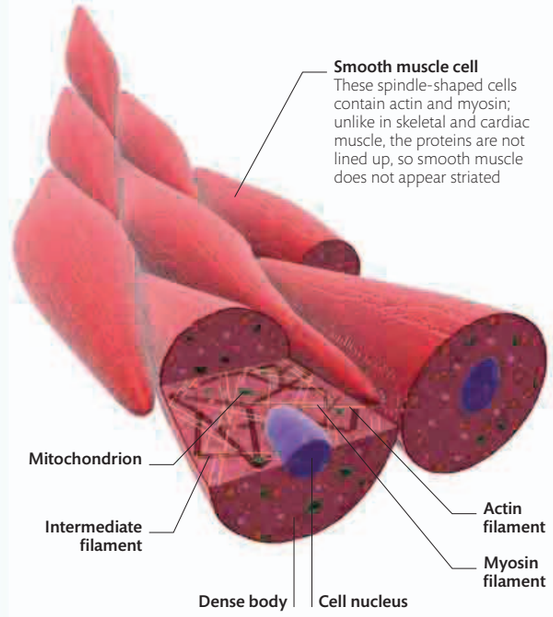

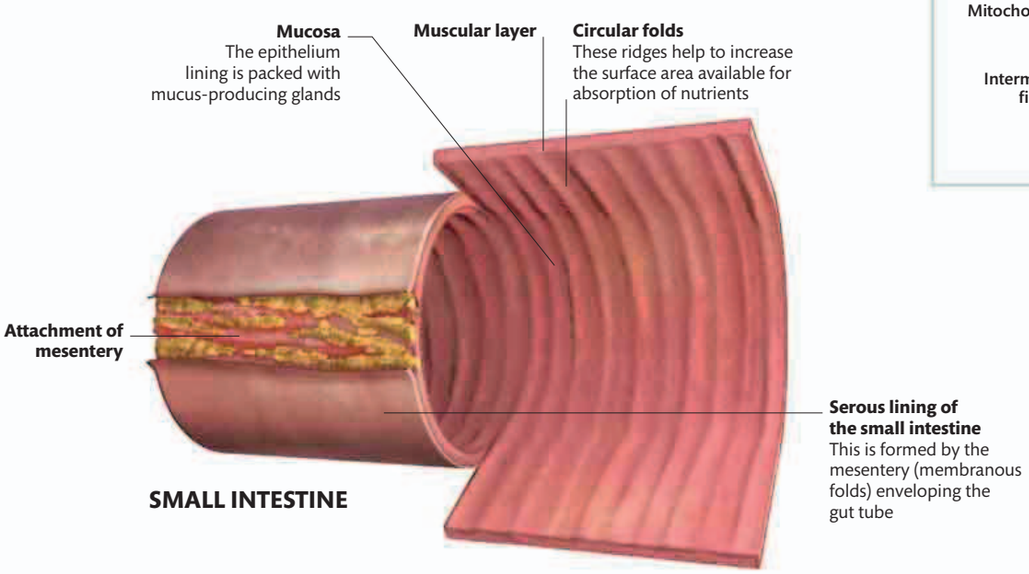

SMALL INTESTINE

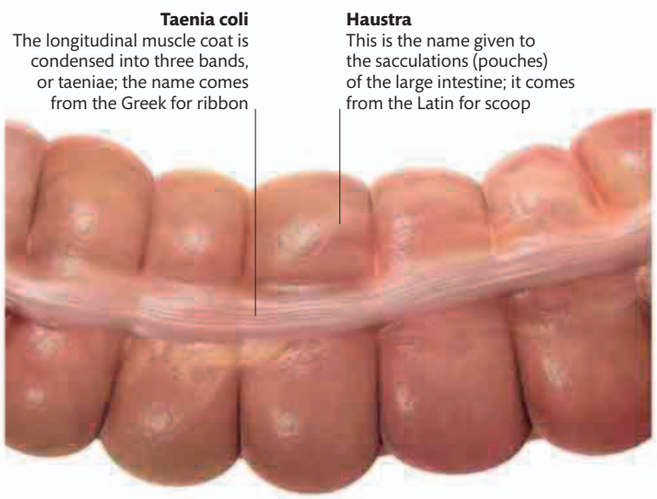

LARGE INTESTINE

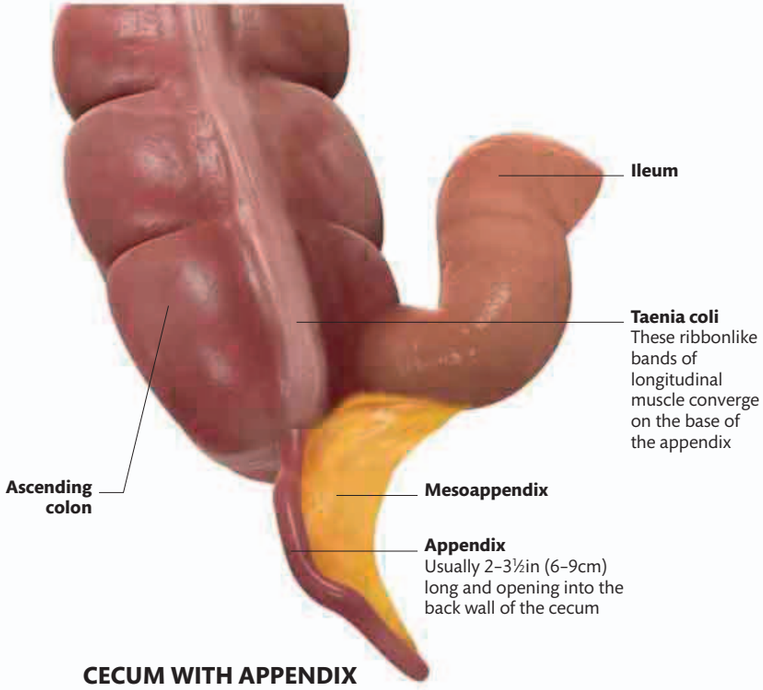

CECUM WITH APPENDIX

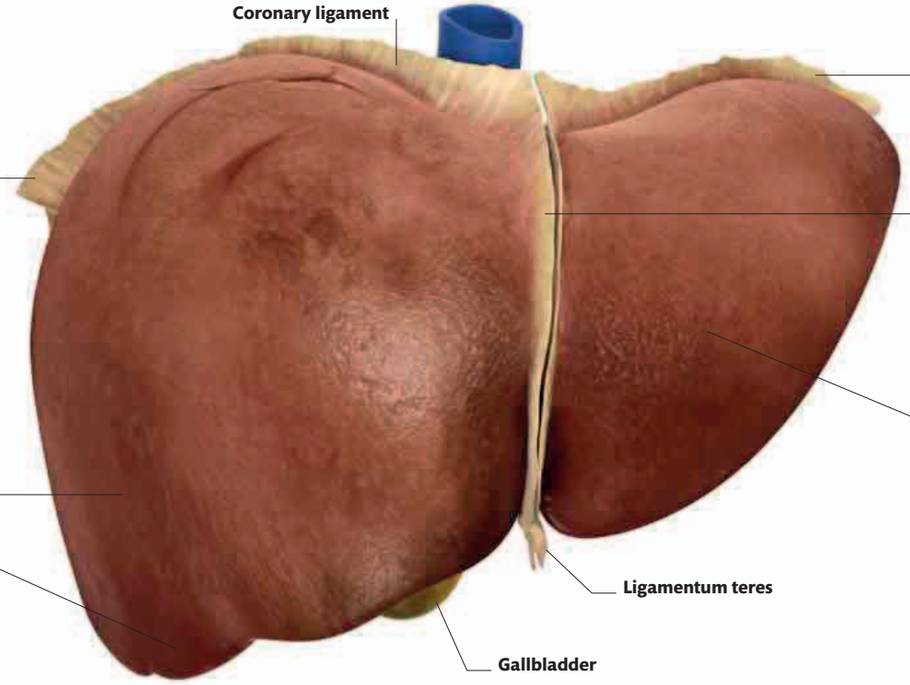

LIVER ANTERIOR (FRONT)

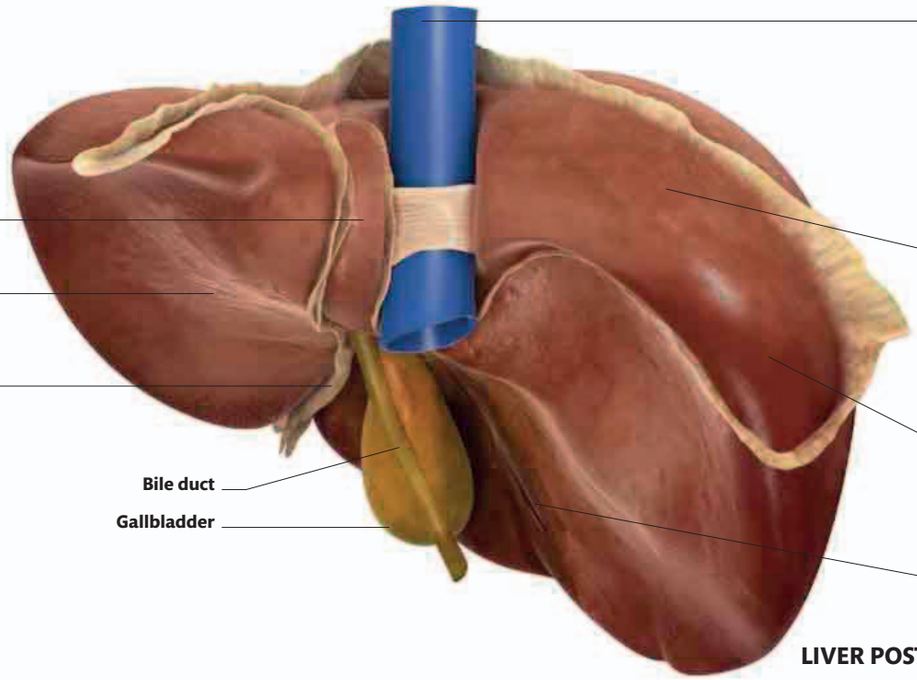

LIVER POSTERIOR (BACK)

LIVER, PANCREAS, AND GALLBLADDER

The liver, the largest internal organ, can weigh up to 6lb (3kg). It does hundreds of jobs simultaneously, many of them related to digestion. It produces bile, which is stored in the gallbladder and helps to digest fats. It also receives nutrients from the gut via the portal vein and processes them. It breaks down or builds up proteins, carbohydrates, and fats according to need; detoxifies or deactivates substances such as alcohol and drugs; and plays a role in the immune system. The pancreas, a long, thin, leaf-shaped gland lying under the liver and behind the stomach, produces hormones that are secreted into the blood, and makes pancreatic juice, full of digestive enzymes, which it empties into the duodenum.

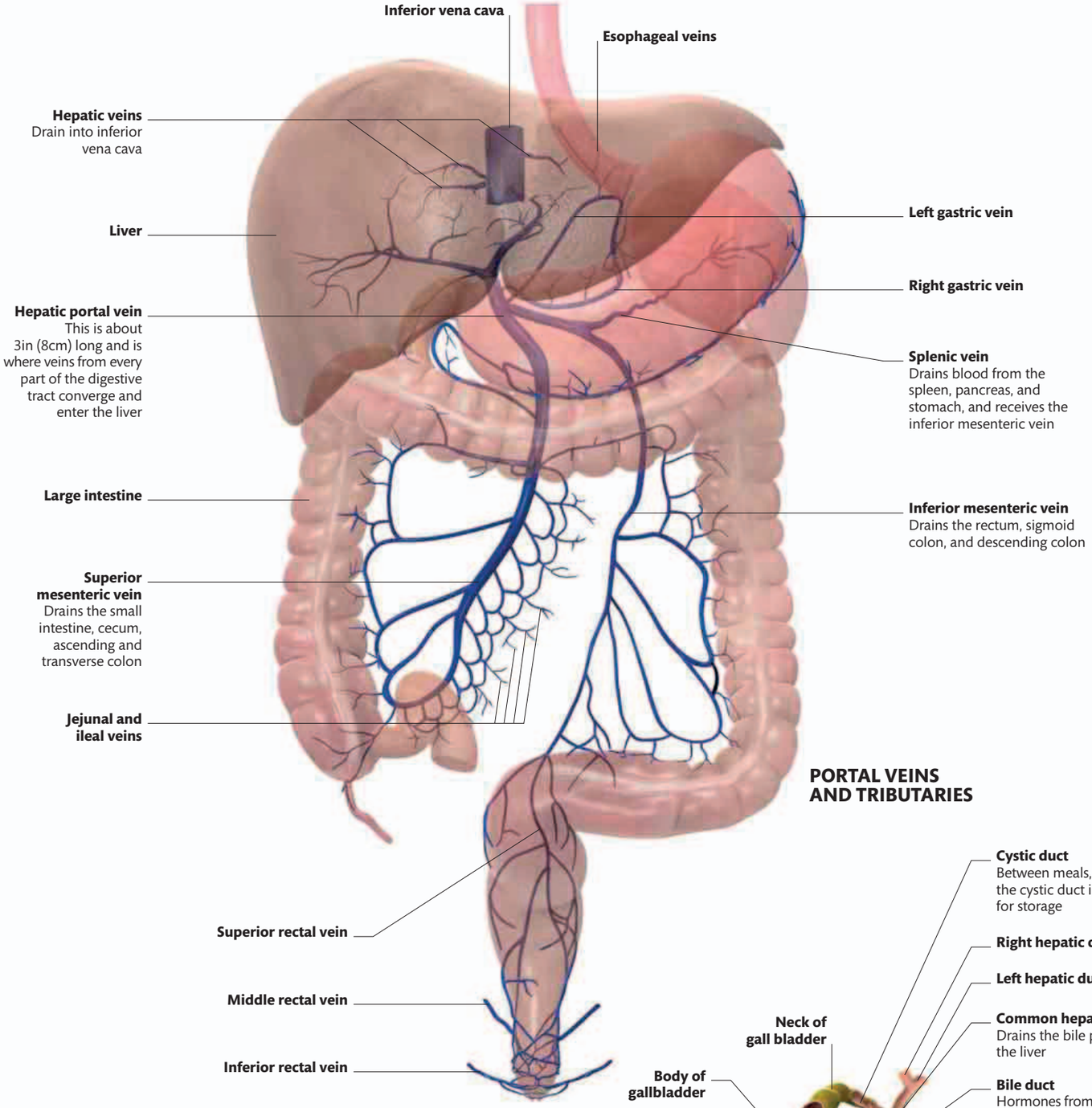

PORTAL VEINS AND TRIBUTARIES

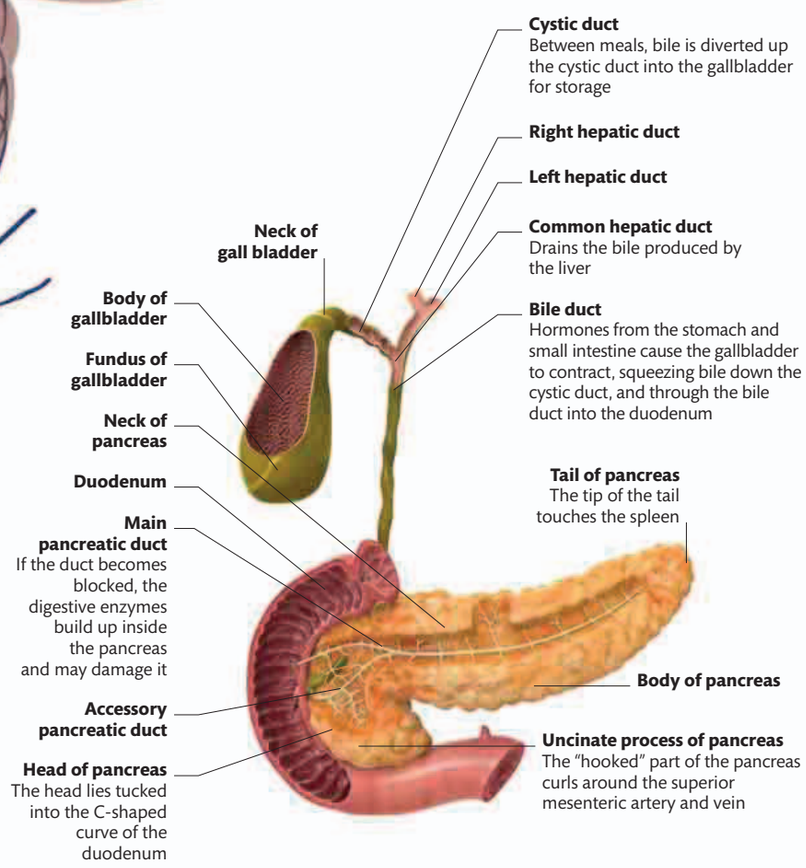

GALLBLADDER AND PANCREAS

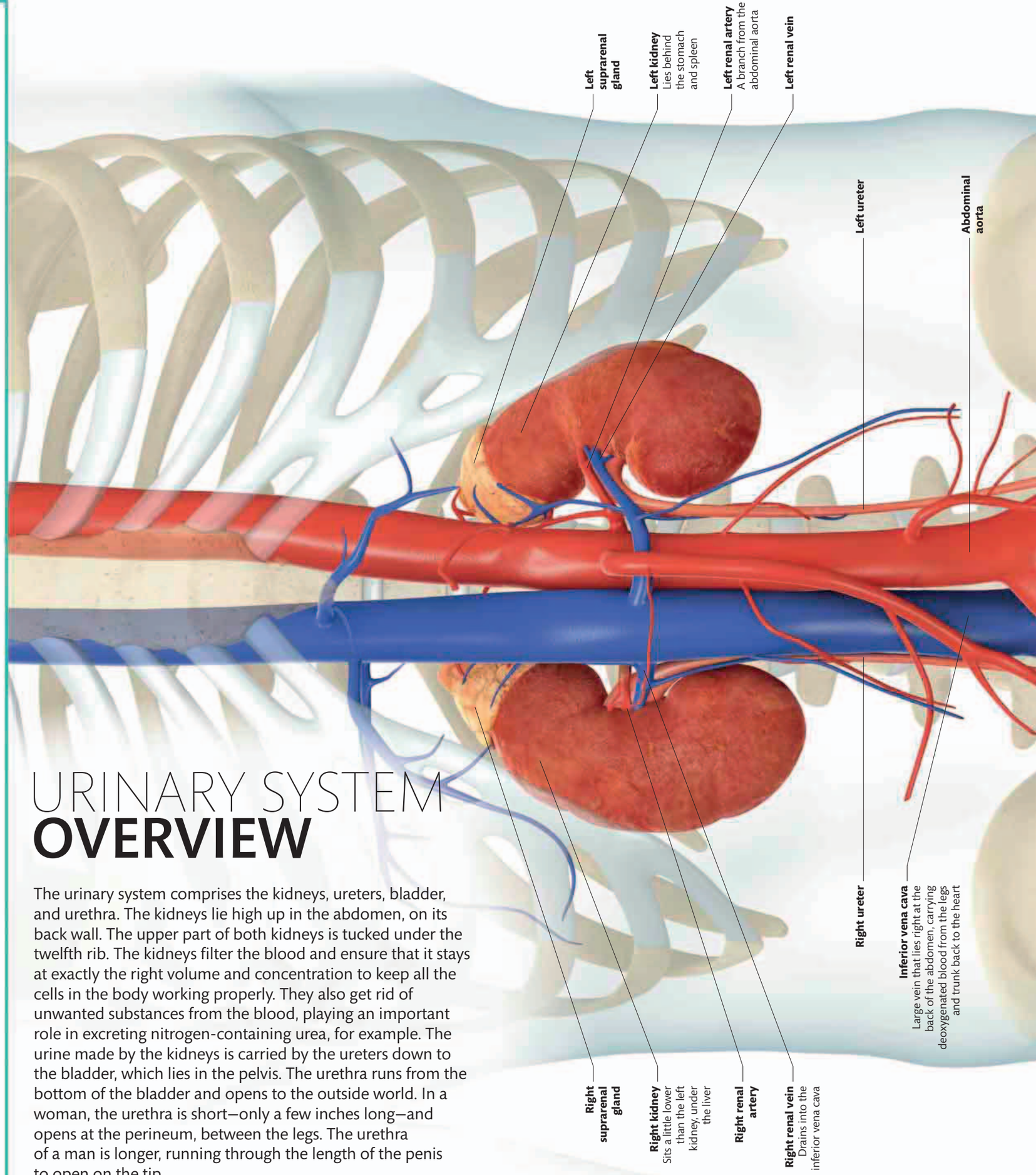

URINARY SYSTEM **OVERVIEW**

The urinary system comprises the kidneys, ureters, bladder, and urethra. The kidneys lie high up in the abdomen, on its back wall. The upper part of both kidneys is tucked under the twelfth rib. The kidneys filter the blood and ensure that it stays at exactly the right volume and concentration to keep all the cells in the body working properly. They also get rid of unwanted substances from the blood, playing an important role in excreting nitrogen-containing urea, for example. The urine made by the kidneys is carried by the ureters down to the bladder, which lies in the pelvis. The urethra runs from the bottom of the bladder and opens to the outside world. In a woman, the urethra is short—only a few inches long—and opens at the perineum, between the legs. The urethra of a man is longer, running through the length of the penis to open on the tip.

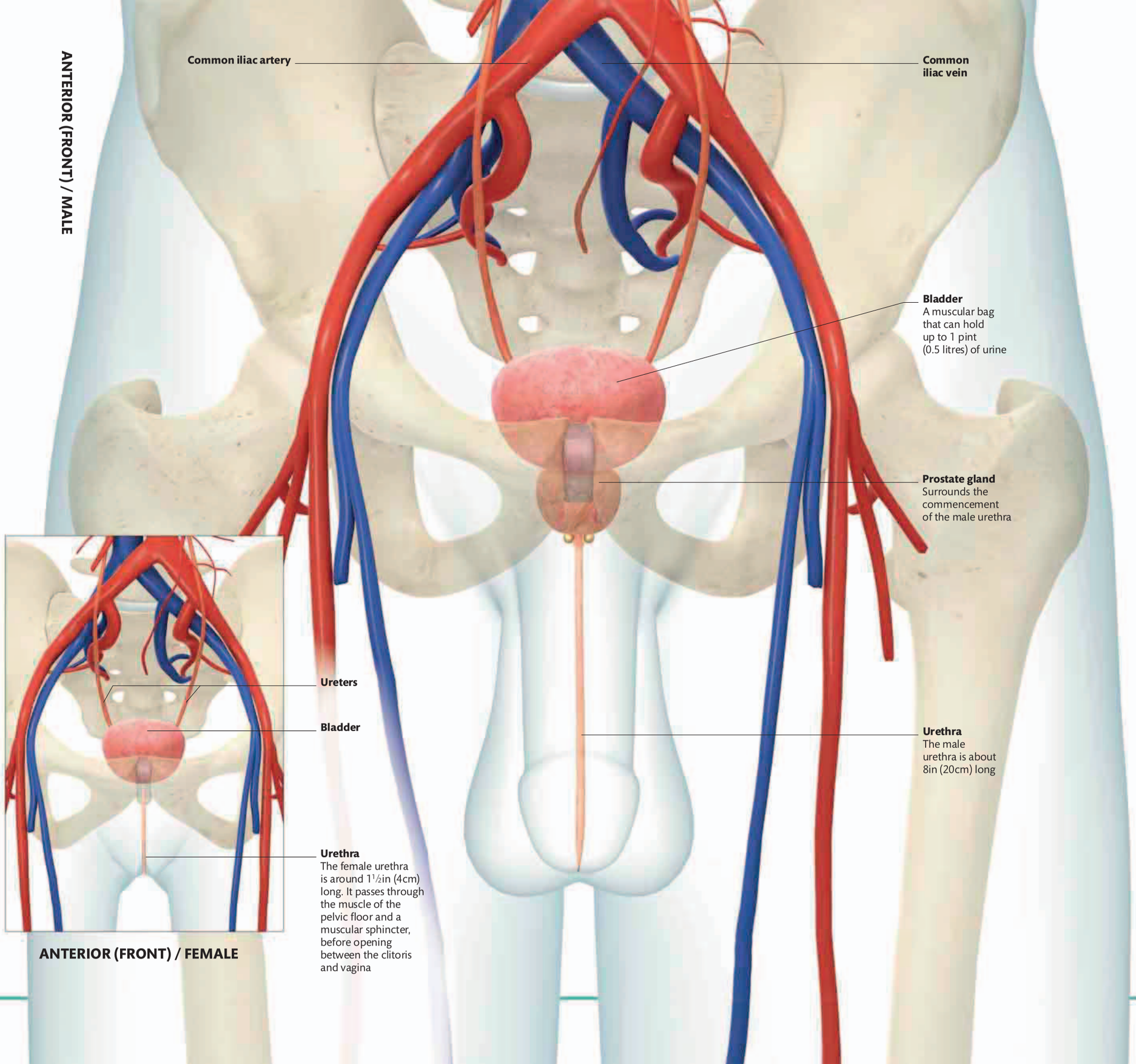

ANTERIOR (FRONT) / FEMALE

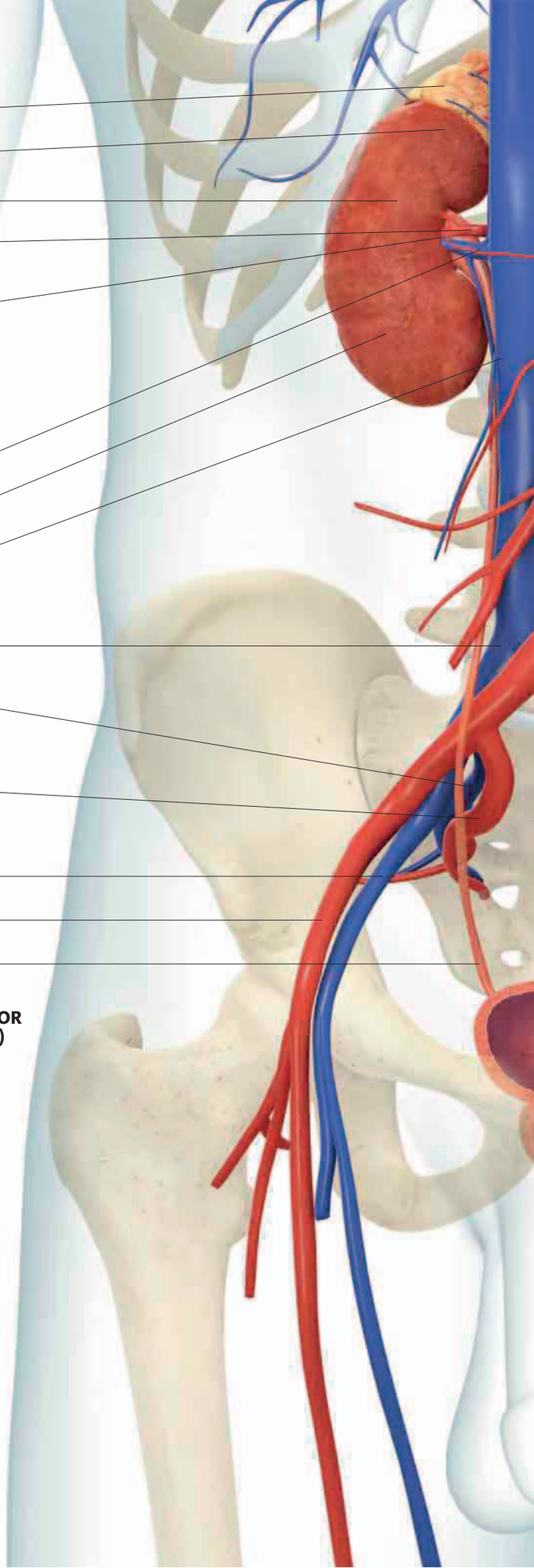

ABDOMEN AND PELVIS

The kidneys lie high up on the back wall of the abdomen, tucked up under the twelfth ribs. A thick layer of perinephric fat surrounds and protects each kidney. The kidneys filter the blood, which is carried to them via the renal arteries. They remove waste from the blood, and keep a tight check on blood volume and concentration. The urine they produce collects first in cup-shaped calyces, which join to form the renal pelvis. The urine then flows out of the kidneys and down narrow, muscular tubes called ureters to the bladder in the pelvis. The bladder is a muscular bag that can expand to hold up to about 1 pint (0.5 litres) of urine, and empties itself when the individual decides it is convenient. Urine travels through the urethra before leaving the body.

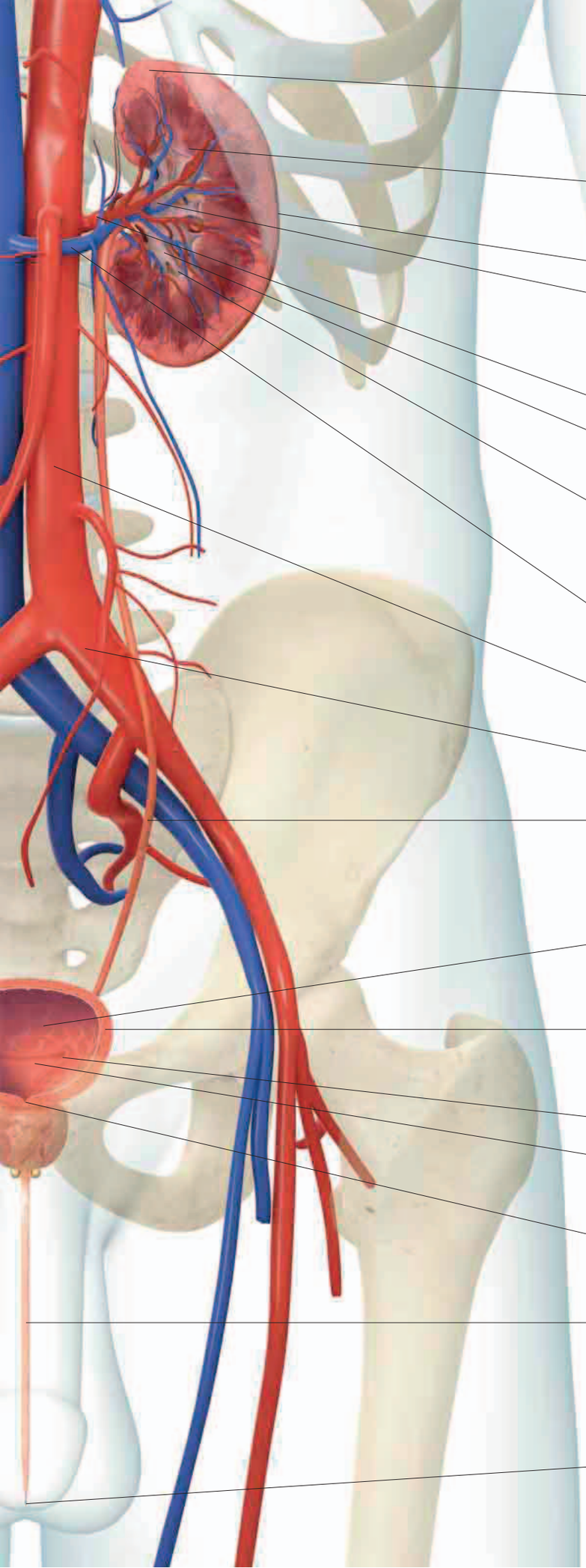

Renal cortex
Cortex means rind or bark; this is the outer tissue of the kidney

Renal medullary pyramid
Medulla means marrow or pith; this core tissue of the kidney is arranged as pyramids, which look triangular in cross section

Left kidney

Renal pelvis
Collects all urine from the kidney, and empties into the ureter; pelvis means basin in Latin, and the renal pelvis should not be confused with the bony pelvis—also shaped like a large basin

Left renal artery

Major calyx
The major calyces collect urine from the minor calyces, then themselves join together to form the renal pelvis

Minor calyx
Calyx originally meant flower covering in Greek, but because it is similar to the Latin word for cup it is used to describe cup-shaped structures in biology; urine from the microscopic collecting tubules of the kidney flows out into the minor calyces

Left renal vein

Abdominal aorta

Left common iliac artery

Left ureter
This name comes from the Greek for to make water; the two ureters carry urine from the kidneys to the bladder

Bladder
The empty bladder lies low down, in the true pelvis, behind the pubic symphysis; as the bladder fills, it expands up into the abdomen

Detrusor muscle
The crisscrossing smooth muscle bundles of the bladder wall give the inner surface of the bladder a netlike appearance

Ureteric orifice

Trigone
The three-cornered region of the back wall of the bladder, between the ureteric orifices and the internal urethral orifice

Internal urethral orifice

Urethra
From the Greek for urinate; this tube carries urine from the bladder to the outside world, a distance of around 1½in (4cm) in women, and about 8in (20cm) in men (as it travels the length of the penis)

External urethral orifice
The male urethra opens at the tip of the glans penis

REPRODUCTIVE SYSTEM **OVERVIEW**

Most organs in the body are similar in men and women. However, when it comes to the reproductive organs, there is a world of difference. In a woman, the ovaries, which produce eggs and female sex hormones, are tucked away, deep inside the pelvis. Also located within the pelvis are the vagina, uterus, and paired oviducts, or fallopian tubes, in which eggs are conveyed from the ovaries to the uterus. The woman's reproductive system also includes the mammary glands, which are important in providing milk for the newborn.

In a man, the testes, which produce sperm and sex hormones, hang well outside the pelvis, in the scrotum. The rest of the male reproductive system consists of a pair of tubes called the vasa deferentia (singular, vas deferens), the accessory sex glands (the seminal vesicles and the prostate), and the urethra.

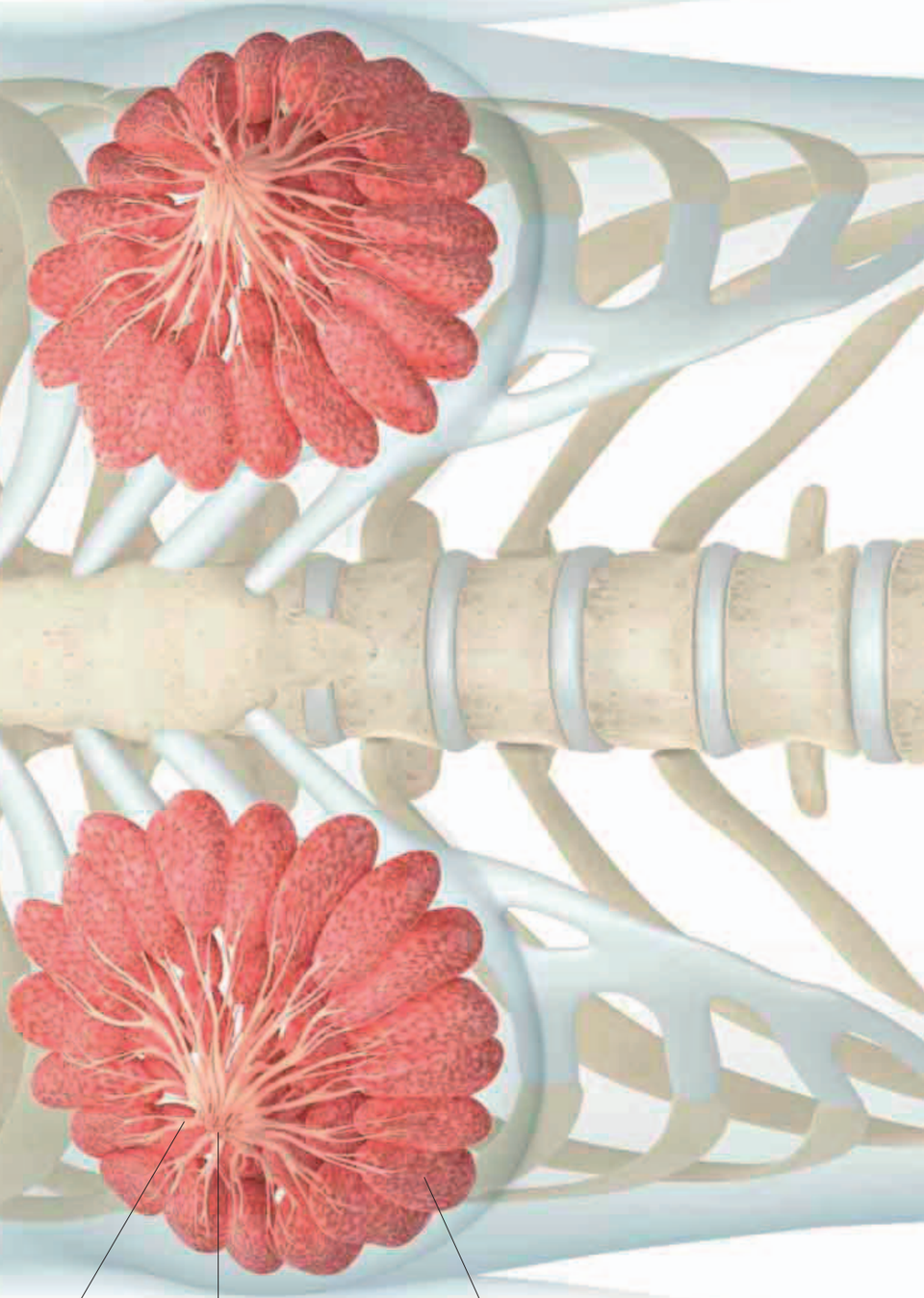

ANTERIOR (FRONT) / FEMALE

Oviduct
Also known as fallopian tubes, oviducts collect eggs produced at ovulation and transport them to the uterus; oviducts are also the place where fertilization normally occurs

Fimbriae
Fingerlike projections that form a feathery end to each oviduct

Ovary
Female gonad; is hidden away, deep within the pelvis

Fundus of uterus
The uterus is angled forward, so the fundus—the farthest point from the opening—lies toward the front

Body of uterus

Cervix of uterus
The cervix, or neck of the uterus, projects down into the vagina

Vagina
Flexible muscular tube that accommodates the male penis during coitus; during childbirth, it expands to allow the fetus to pass through

Vas deferens

Seminal vesicle
Contributes fluid to semen

Prostate gland
Accessory gland located at the base of the bladder; contributes some fluid to semen

Shaft of penis
Formed by masses of erectile tissue, which become engorged with blood during erection

Urethra
Conveys sperm and urine through penis

Epididymis
A tightly-coiled tube on the back of the testis; sperm are stored and mature here

Glans penis

Testis
Male gonad; hangs outside body cavity, in the scrotum

Scrotum
Pouch of skin and muscle that encases testis

ANTERIOR (FRONT) / MALE

Axillary tail
This part of the illustration shows the extent of the breast; breast tissue continues upward and outward on the chest wall—right up into the axilla (armpit)

Nipple

Areola
Means small area in Latin

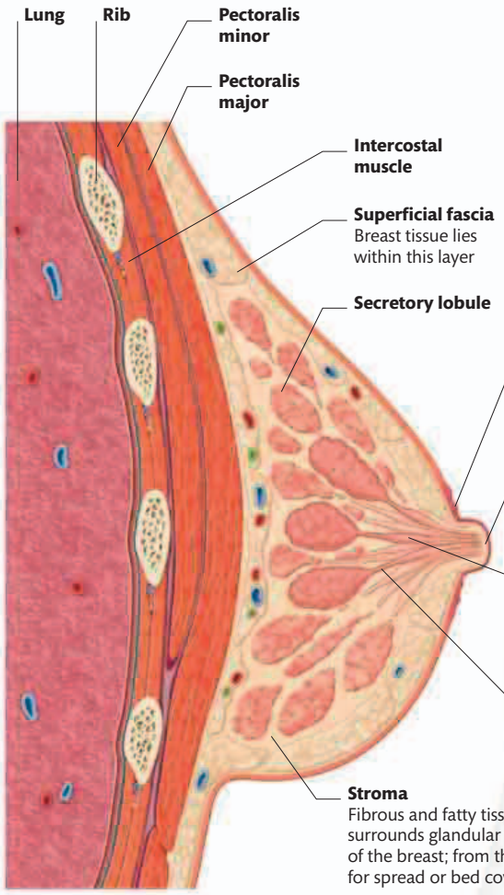

CROSS SECTION OF BREAST

ANTERIOR (FRONT) / FEMALE

Lactiferous duct
From the Latin for milk-carrying duct

Secretory lobule
Lactiferous ducts branch during puberty to form lobules, where milk is produced and secreted

THORAX

The breasts, or mammary glands, are an important part of the reproductive system in women. Like all other mammals, human females have mammary glands to provide the newborn with milk. But while many mammals have multiple mammary glands, humans (and other apes) have just two, on the front of the chest. The breasts develop at puberty, when they grow due to the increased production of glandular tissue and fat. The breasts lie on the pectoralis major muscle on each side. Each breast contains 15 to 20 lobes, which are connected to the nipple by lactiferous ducts. There seems to be a basic plan in the developing embryo, so that male nipples appear, although the breast does not form.

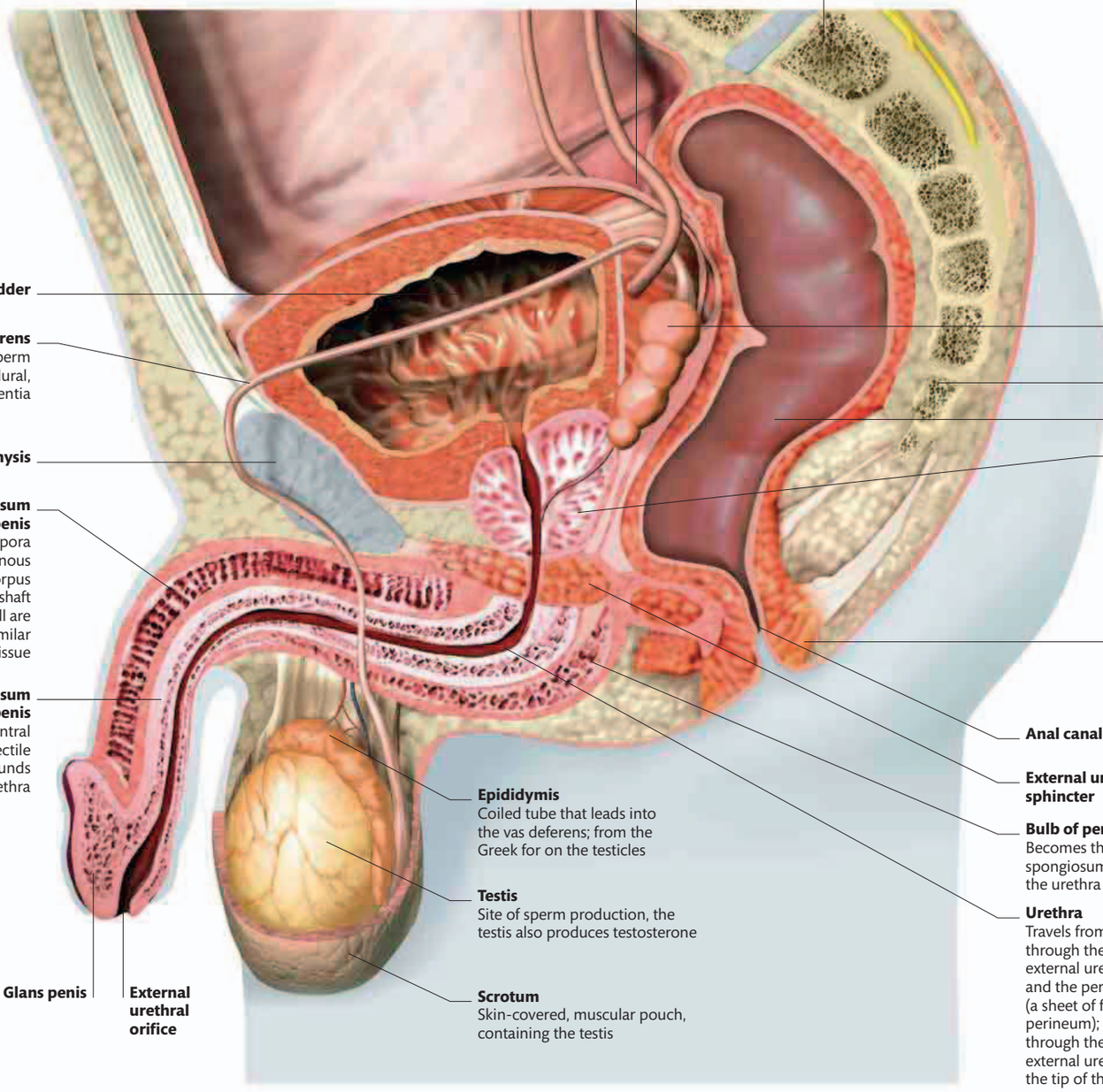

SAGITTAL SECTION / MALE

ABDOMEN AND PELVIS

The male and female reproductive systems are both comprised of a series of internal and external organs, although structurally these are very different. It is true that both sexes possess gonads (ovaries in women and testes in men) and a tract, or set of tubes, but the similarity ends there. When we look in detail at the anatomy of the pelvis in each sex, the differences are obvious. The pelvis of a man contains only part of the reproductive tract, as well as the lower parts of the digestive and urinary tracts, including the rectum and bladder. Beneath the bladder is the prostate gland; this is where the vasa deferentia, which bring sperm from the testis, empty into the urethra. A woman's pelvic cavity contains more of the reproductive tract than a man's. The vagina and uterus are situated between the bladder and rectum in the pelvis.

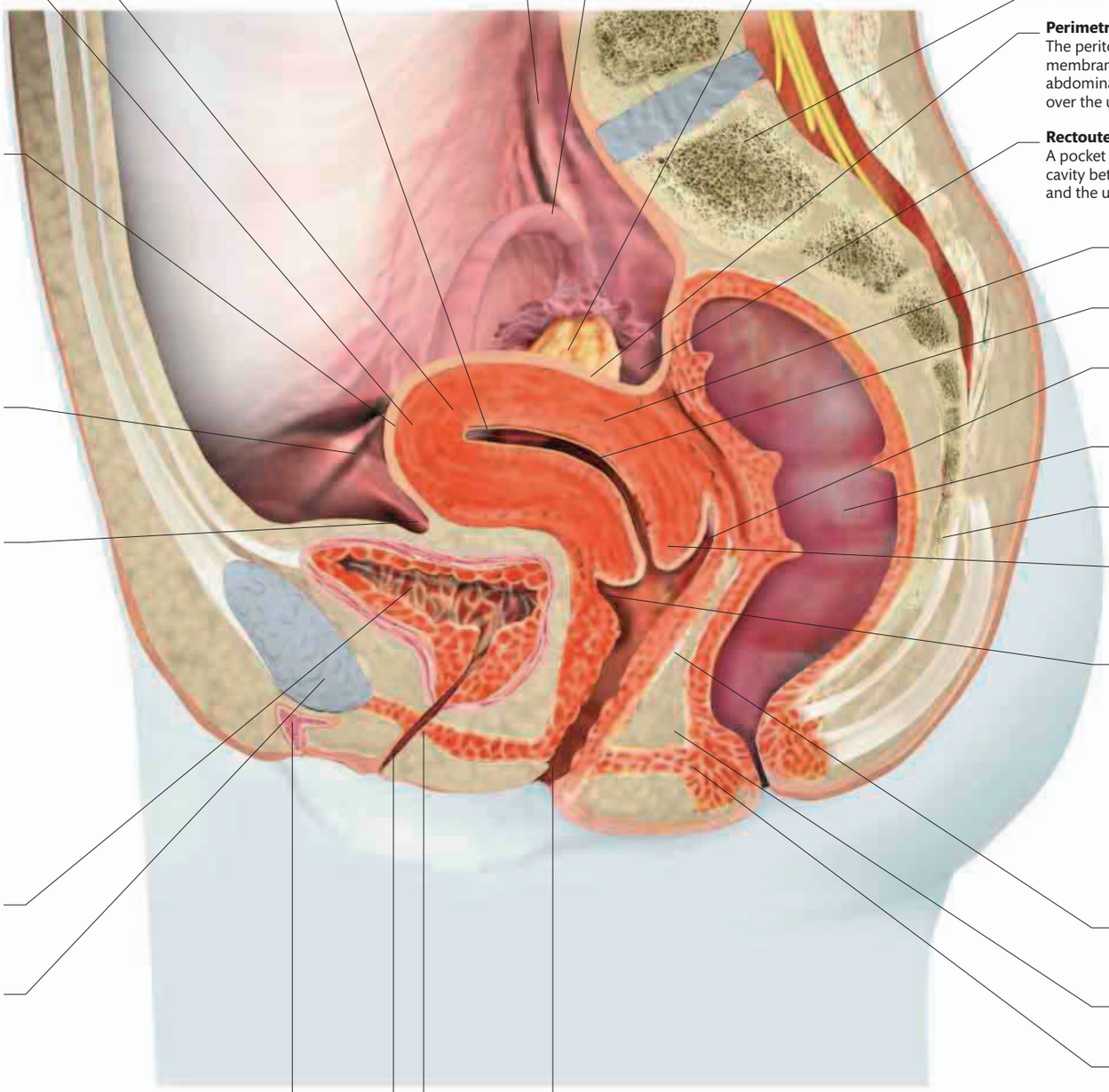

Myometrium
Thick, smooth muscle layer of uterus

Fundus of uterus
The top portion of the uterus, farthest from the cervix

Uterus
The uterus is shaped like a flattened pear, and normally lies in the position shown here—bent forward over an empty bladder

Round ligament of the uterus

Vesicouterine pouch
A pocket of peritoneal cavity between the bladder and the uterus. The peritoneal cavity is a potential space between the peritoneum lining the abdominal walls, and the abdominal and pelvic organs

Bladder

Pubic symphysis
Cartilage joint at front of pelvis; softens during pregnancy then widens slightly during childbirth

Endometrium
Lining of the uterus; innermost layer of endometrium is shed during menstruation. From the Greek for within the womb

Suspensory ligament of ovary
Carries ovarian arteries and veins to and from the ovary

Oviduct
Literally, egg-duct; each one is about 4in (10cm) long

Ovary
Means egg place in Latin; each of the two ovaries lies on the pelvic side wall, in the angle between the internal and external iliac arteries

Sacrum

Perimetrium
The peritoneum (serous membrane lining the abdominal cavity) lies over the uterus

Rectouterine pouch
A pocket of peritoneal cavity between the rectum and the uterus

Body of uterus

Cavity of uterus

Posterior fornix of vagina

Rectum

Coccyx

Cervix of uterus
Literally, the neck of the uterus

Anterior fornix of vagina
Fornices are gutter-like areas that form as the cervix projects down into the vagina, known as anterior, lateral, and posterior fornices; fornix means arch or vault in Latin

Rectovaginal septum

Anal canal

External anal sphincter

SAGITTAL SECTION / FEMALE

Clitoris
Contains spongy, erectile tissue similar to that in the penis

Urethra

External urethral sphincter

Vagina
Tube, around 3½in (9cm) long, with walls of fibrous and muscular tissue

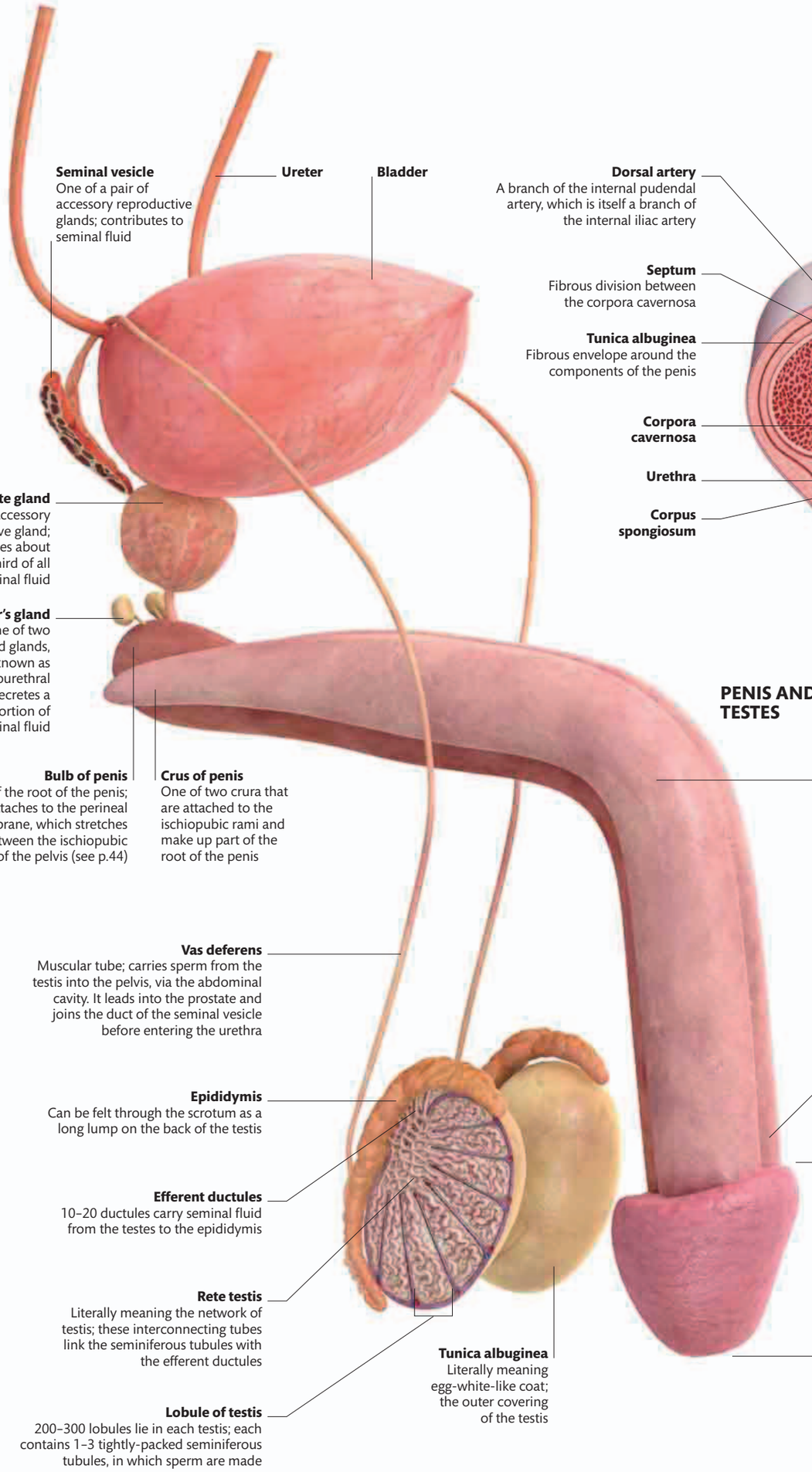
Seminal vesicle
One of a pair of accessory reproductive glands; contributes to seminal fluid
Ureter
Bladder
Prostate gland
Accessory reproductive gland; contributes about a third of all seminal fluid
Cowper's gland
One of two pea-sized glands, also known as bulbourethral glands; secretes a tiny portion of seminal fluid
Bulb of penis
Part of the root of the penis; attaches to the perineal membrane, which stretches between the ischiopubic rami of the pelvis (see p.44)
Crus of penis
One of two crura that are attached to the ischiopubic rami and make up part of the root of the penis
Vas deferens
Muscular tube; carries sperm from the testis into the pelvis, via the abdominal cavity. It leads into the prostate and joins the duct of the seminal vesicle before entering the urethra
Epididymis
Can be felt through the scrotum as a long lump on the back of the testis
Efferent ductules
10–20 ductules carry seminal fluid from the testes to the epididymis
Rete testis
Literally meaning the network of testis; these interconnecting tubes link the seminiferous tubules with the efferent ductules
Lobule of testis
200–300 lobules lie in each testis; each contains 1–3 tightly-packed seminiferous tubules, in which sperm are made
Tunica albuginea
Literally meaning egg-white-like coat; the outer covering of the testis
PENIS AND TESTES
Corporus cavernosa
One of two columns of tissue, formed where the crura join at the shaft of the penis; known together as corpora cavernosa
Corpus spongiosum
The bulb of the penis continues on to form spongy erectile tissue called the corpus spongiosum, or spongy body
Glans penis
Expanded part of the corpus spongiosum
Dorsal artery
A branch of the internal pudendal artery, which is itself a branch of the internal iliac artery
Dorsal veins
These drain up to the venous plexus around the prostate and, ultimately, to the internal iliac veins
Septum
Fibrous division between the corpora cavernosa
Tunica albuginea
Fibrous envelope around the components of the penis
Corpora cavernosa
Urethra
Corpus spongiosum
CROSS SECTION OF PENIS

Ampulla of oviduct
Slightly widened part of oviduct, where fertilization normally occurs; from the Latin for flask

Isthmus of oviduct
Last third of oviduct; narrower than ampulla. Isthmus is from Greek for neck, or narrow passage

Body of uterus

Secondary follicle
Follicle that has started to accumulate fluid

Fimbriae
Fingerlike projections, help pick up the ovulated egg. From the Latin for fringe

Mature follicle
Filled with fluid; this will burst to release the egg at ovulation

Corpus luteum
Remains of follicle after ovulation, literally meaning yellow body

Infundibulum of oviduct
Funnel-shaped end of oviduct, closest to ovary; from the Latin for funnel

Ovary
With a volume of around 4¼in³ (11cm³), the ovary contains follicles in various stages of the ovarian cycle

Primary follicle
Contains a developing egg (or oocyte), surrounded by follicle cells

Corpus albicans
If the ovulated egg is unfertilized, the corpus luteum shrivels up to form this scarlike structure

Cavity of uterus

Cervical canal

Cervix of uterus
About 1in (2.5cm) long

Lateral fornix of vagina
The recesses of the vagina around the cervix are called the fornices; there is a lateral fornix on either side of the cervix

Vagina
The ridged passage (or lumen) through the vagina is folded into an H-shape, enabling it to expand

UTERUS

ABDOMEN AND PELVIS

At a very fundamental level, the reproductive systems of man and woman must work together to allow eggs and sperm to meet. These views of the isolated organs and reproductive tracts show clearly how the anatomy is arranged to achieve this. The ovaries, where eggs (or ova) are produced, are deep inside the female pelvis. The eggs are collected from the ovaries by a pair of tubes, the oviducts, and it is usually here that fertilization takes place. The fertilized egg then moves along the oviduct, dividing into a ball of cells. The embryo eventually reaches the uterus, which is designed to accommodate and support the growing fetus. The vagina provides both a way for sperm to get in, and the route for the baby to get out at birth.

Ischiocavernosus
Muscle covering the crus of the clitoris

Prepuce, overlying body of clitoris

Crus of clitoris
Smaller in size than the crus of the penis; attached to the ischiopubic ramus of the bony pelvis

Glans of clitoris
Erectile organ, equivalent to the penis; the body of the clitoris comprises two corpora cavernosa

External urethral orifice

Vestibule
Area between the labia minora; Latin for entrance court

Bulb of vestibule
One of a pair of structures equivalent to the single bulb of the penis; made of spongy erectile tissue

Vaginal orifice

Labia minora
Folds of skin either side of the vestibule; singular is labium minus

Bulbospongiosus
Muscle covering the bulb of vestibule; helps increase pressure in the underlying spongy tissue

Anus

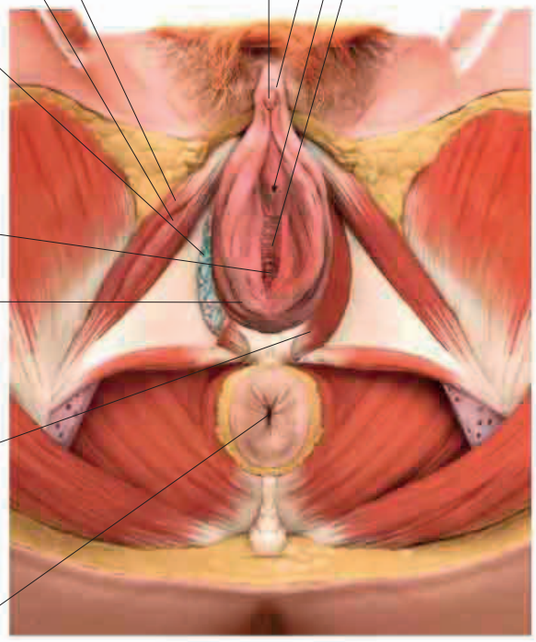

EXTERNAL FEMALE GENITALIA

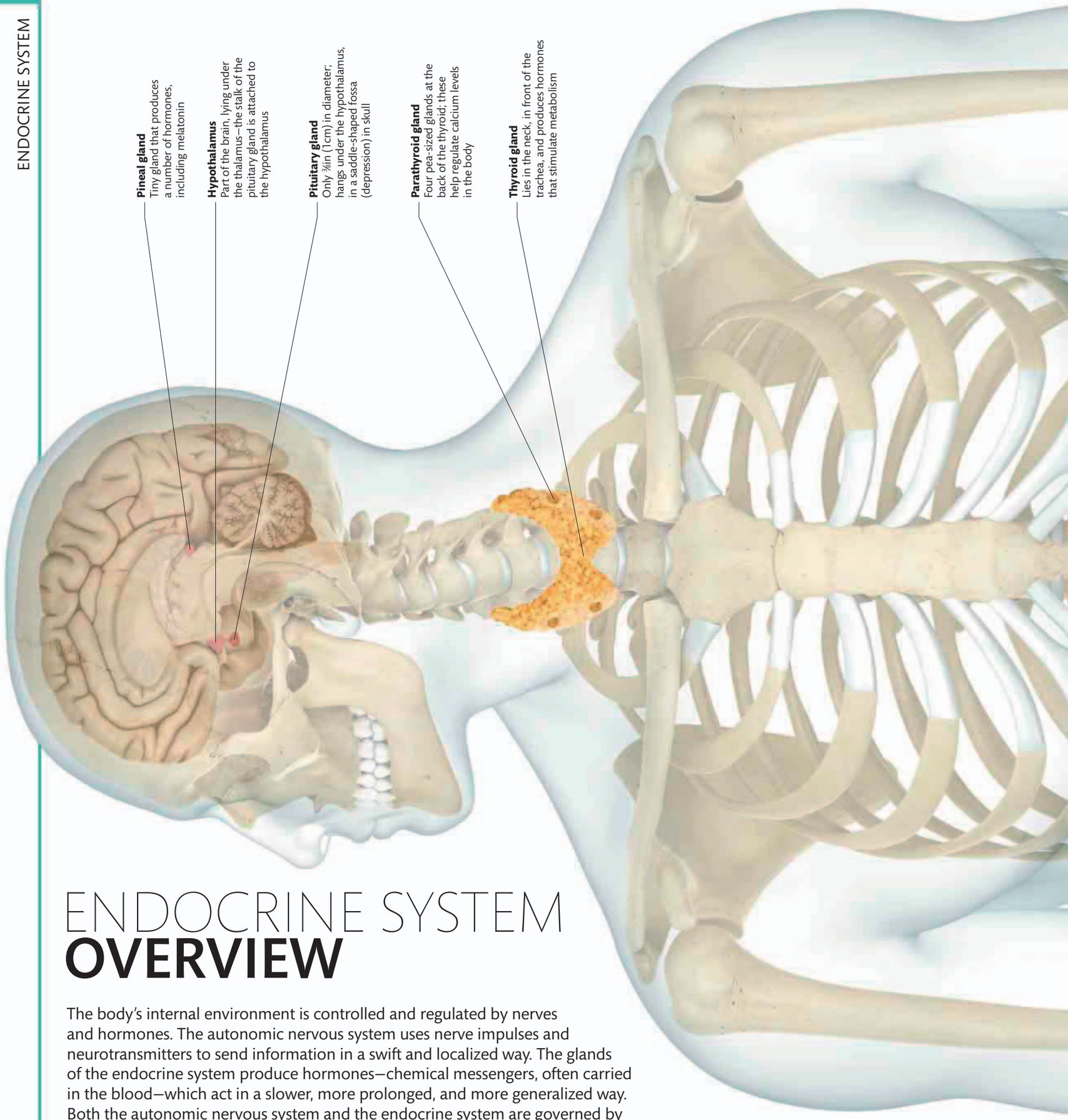

ENDOCRINE SYSTEM **OVERVIEW**

The body's internal environment is controlled and regulated by nerves and hormones. The autonomic nervous system uses nerve impulses and neurotransmitters to send information in a swift and localized way. The glands of the endocrine system produce hormones—chemical messengers, often carried in the blood—which act in a slower, more prolonged, and more generalized way. Both the autonomic nervous system and the endocrine system are governed by the hypothalamus in the brain. The pituitary gland produces hormones that affect other endocrine glands, which sometimes form discrete organs. There are also hormone-producing cells in the tissues of many other organs.

Suprarenal gland
A pair of glands, also known as adrenal glands, that produce epinephrine, also called adrenaline

Pancreas
Has cells that produce hormones controlling glucose metabolism: insulin and glucagon; also produces digestive enzymes

Ovary
Ovaries produce sex hormones as well as gametes (reproductive cells) called ova

FEMALE

Testis
Testes produce sex hormones as well as gametes (reproductive cells) called sperm

HEAD AND NECK

The insides of our bodies are regulated by the autonomic nervous and endocrine systems. There is overlap between these two systems, and their functions are integrated and controlled within the hypothalamus of the brain. The pituitary gland has two lobes; its posterior lobe develops as a direct extension of the hypothalamus. Both lobes of the pituitary gland secrete hormones into the bloodstream, in response to nerve signals or blood-borne releasing factors from the hypothalamus. Many of the pituitary hormones act on other endocrine glands, including the thyroid gland in the neck, the suprarenal glands on top of the kidneys, and the ovaries or testes.

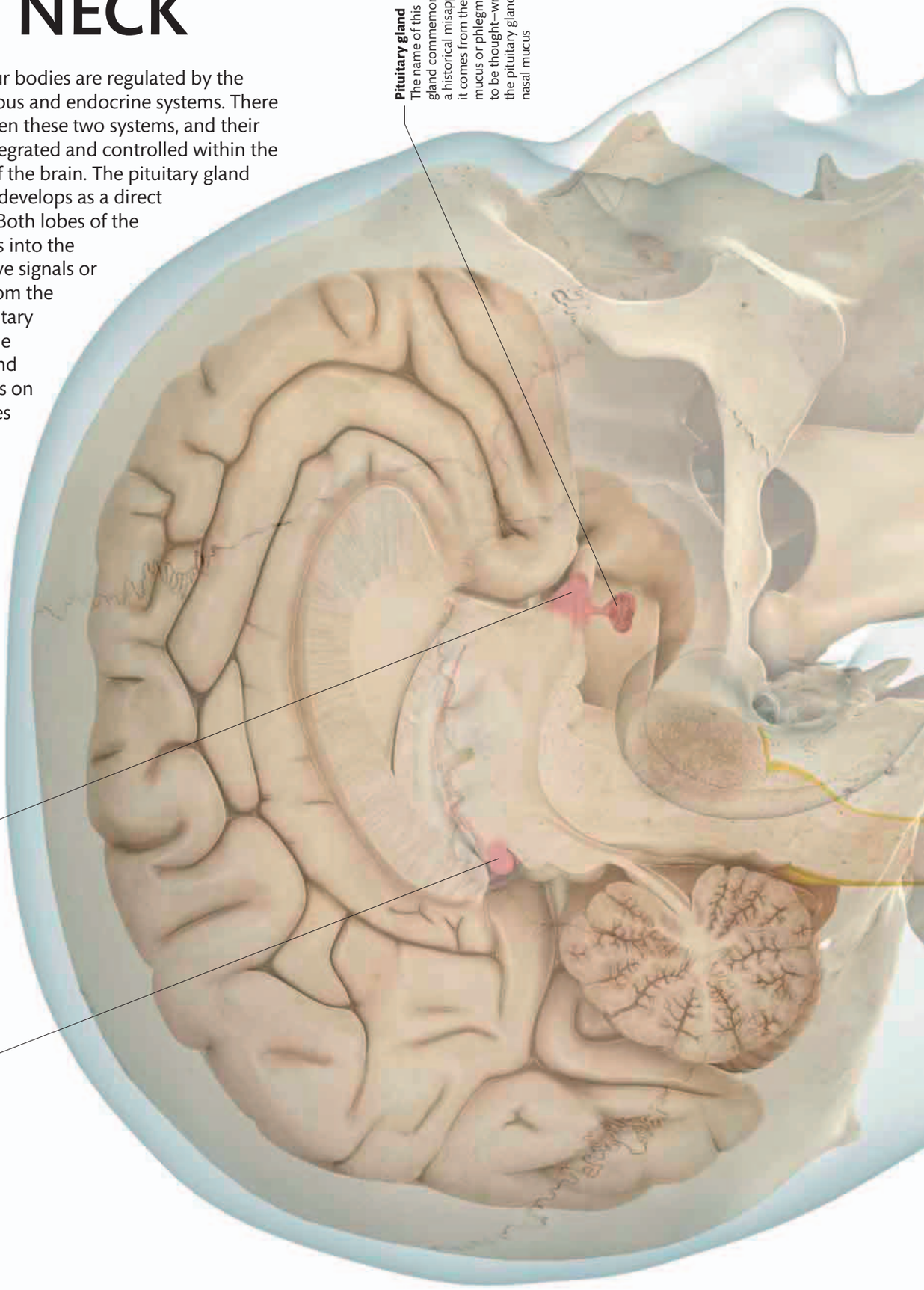

SIDE

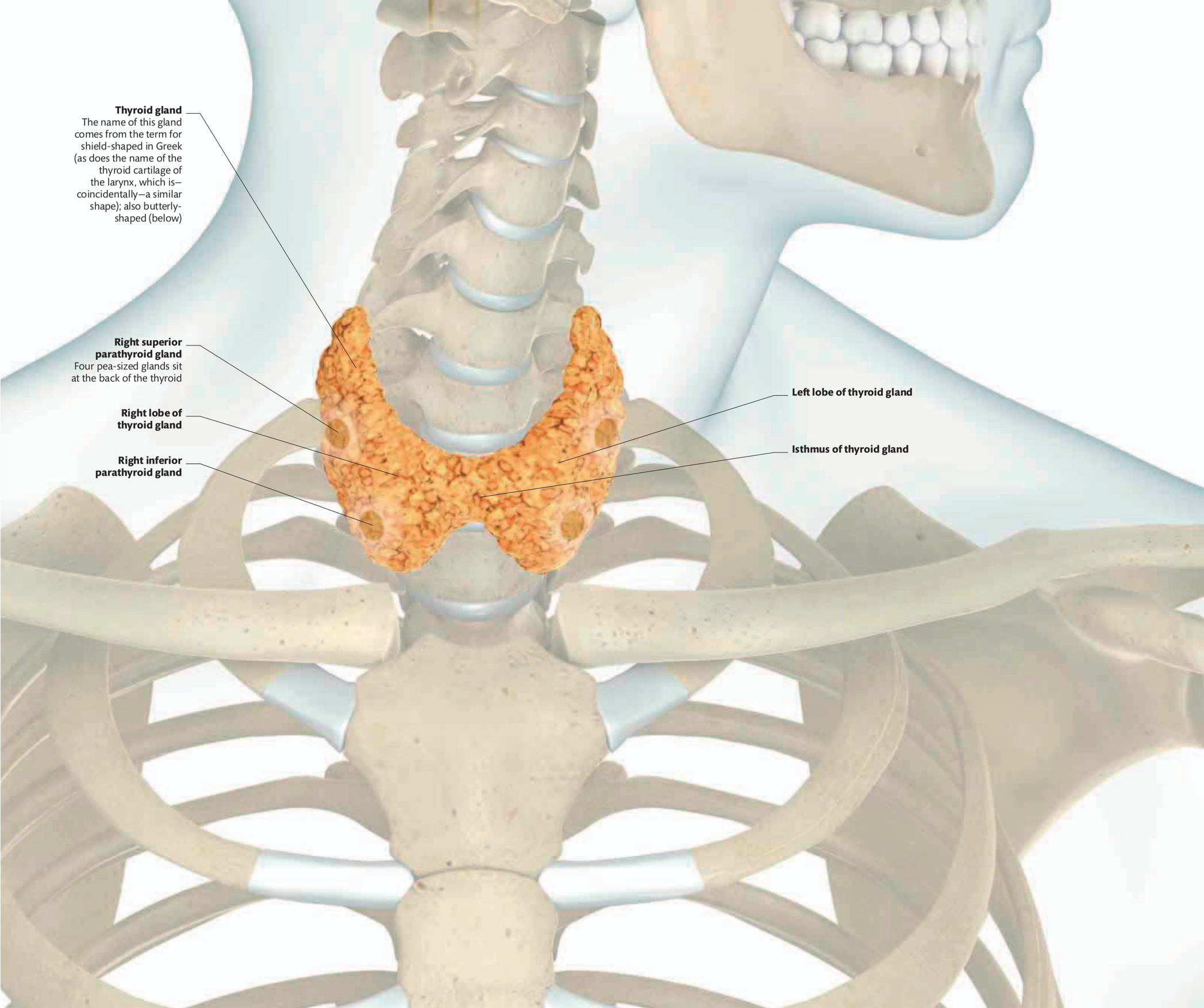
Thyroid gland
The name of this gland comes from the term for shield-shaped in Greek (as does the name of the thyroid cartilage of the larynx, which is—coincidentally—a similar shape); also butterly-shaped (below)
Right superior parathyroid gland
Four pea-sized glands sit at the back of the thyroid
Right lobe of thyroid gland
Right inferior parathyroid gland
Left lobe of thyroid gland
Isthmus of thyroid gland

03 Imaging the Body

The human body is a "living machine" with many complex working parts. To understand how the body functions, and to cure the various ailments that afflict it, it is crucial for medical professionals to examine it in detail. Advances in technology have made it possible to view human anatomy without dissecting the body. Techniques such as magnetic resonance imaging (MRI) reveal the inside of the body with great accuracy and allow us to build up a complete picture of our anatomy from every possible angle.

222 Imaging techniques
224 Head and neck
226 Thorax
228 Abdomen and pelvis
230 Lower arm and hand
232 Lower limb and foot

IMAGING TECHNIQUES

Imaging is vital to diagnose illness, unravel disease processes, and evaluate treatments. Modern techniques provide detailed information with minimum discomfort to the patient and have largely replaced surgery in establishing the presence and extent of disease. Imaging has also helped advance biological research.

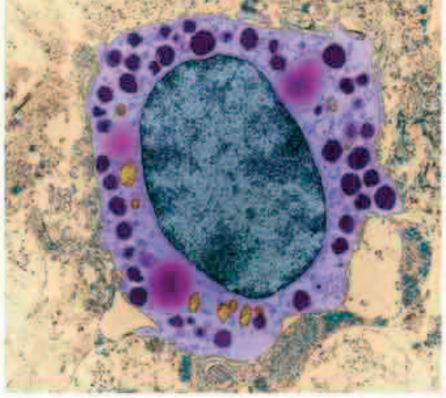

TEM of mast cell
Magnifications of several million times are possible using transmission electron microscopy (TEM). This image shows a mast cell with granules (dark purple) that it releases when it is damaged or is fighting microbes.

The invention of the X-ray in 1895 made the development of noninvasive medicine possible. Without the ability to see inside the body, many internal disorders could only be found after major surgery. Computerized imaging now helps doctors make early diagnoses, which at times greatly increase the likelihood of recovery. Computers process and enhance raw data to aid our visual ability. However, sometimes direct observation is essential. Viewing techniques have also become less invasive with the development of instruments such as the endoscope (see opposite).

MICROSCOPY

Light microscopy (LM) uses magnifying lenses to focus light rays. The light passes through a thin section of material and enlarges it up to 2,000 times. Higher magnifications are achieved with beams of electrons (subatomic particles). In scanning electron microscopy (SEM), the beam runs across a specimen coated with gold film and bounces off the surface to create a three-dimensional image.

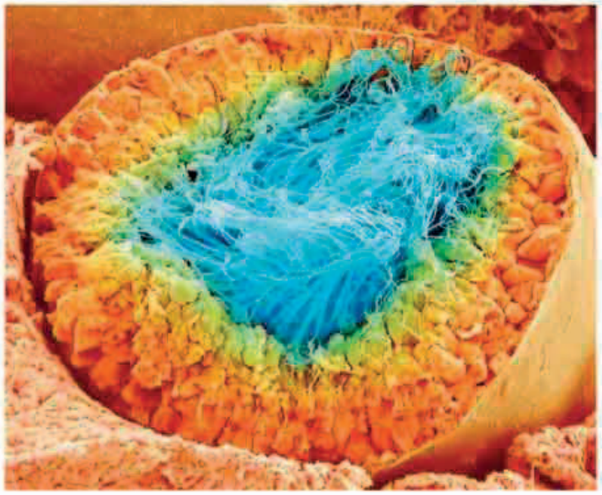

SEM of seminiferous tubule
This freeze-fracture image—in which the specimen is frozen and then cracked open before being scanned—shows sperm heads buried in Sertoli cells (orange) with the tails (blue) projecting into the tubule's lumen.

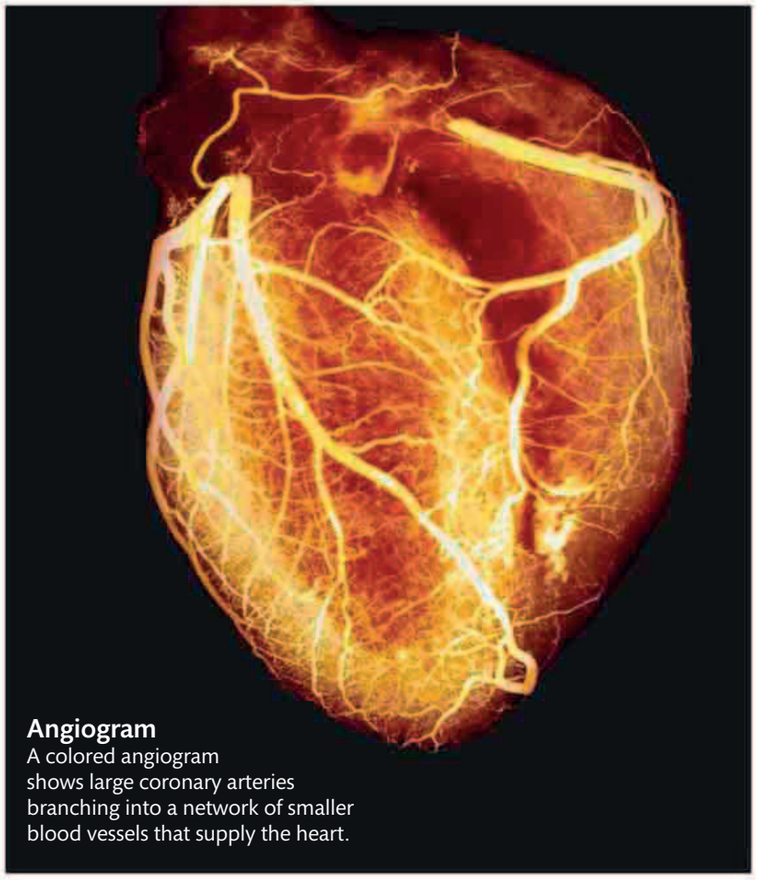

Angiogram
A colored angiogram shows large coronary arteries branching into a network of smaller blood vessels that supply the heart.

X-RAY

Like light rays, X-rays are electromagnetic energy, but of short wavelength. When passed through the body to strike photographic film, they create shadow images (radiographs). Dense structures, such as bone, absorb more X-rays and show up as white, while soft tissues, such as muscle, appear gray. Air-filled spaces, such as the lungs, appear black. The spaces inside the digestive tract, or within blood vessels, may be visualized by filling them with a contrast medium—such as iodine or barium—that absorbs X-rays. A contrast X-ray image of blood vessels is known as an angiogram.

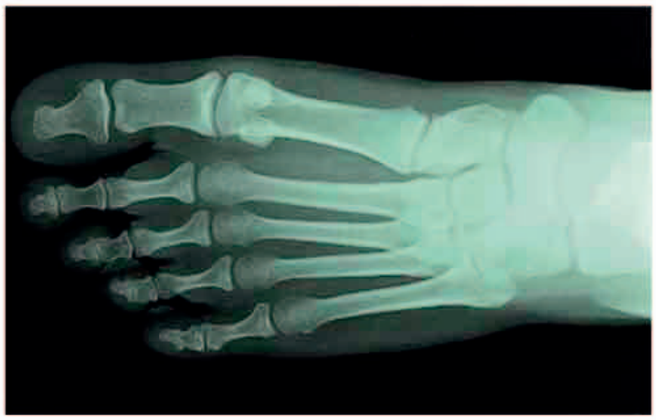

X-ray of foot
This X-ray image shows the foot bones of an adult from above. X-rays are especially useful for viewing dense tissue, such as bone.

RADIONUCLIDE AND PET SCANNING

In radionuclide imaging, a radioactive substance is injected into the body and is absorbed by the area to be imaged. As the substance decays, it emits gamma rays, which a computer forms into an image. Positron emission tomography (PET) is a type of radionuclide scanning where the injected chemical emits radioactive particles called positrons. PET gives data about how the brain functions rather than anatomy.

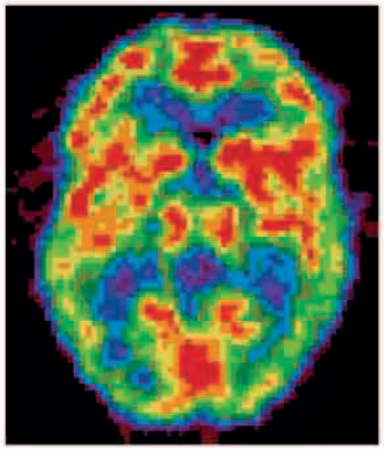

PET brain scan
This scan of the brain shows that the organ is active even in sleep mode. Areas in red, orange, and yellow represent high levels of activity.

ULTRASOUND

In ultrasound, a device called a transducer emits very high-frequency sound waves as it is passed over the body part being examined. The sound waves echo back to the device based on the density of the tissues they encounter. A computer analyses the reflected waves and creates an image.

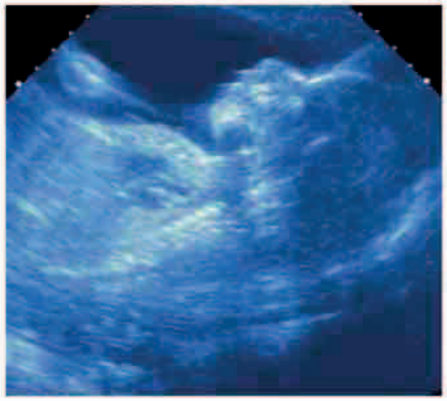

Fetal ultrasound
Low-intensity ultrasound is a safe way to monitor fetal development. In this scan the fetus's head can be seen clearly in profile on the right.

ENDOSCOPY

Telescopelike endoscopes are inserted through natural orifices or incisions to image the body's interior. They can be bent and controlled, and may carry instruments for other purposes as well, such as surgery and biopsy. Endoscopes have been designed to fit various body parts—bronchoscope for airways, gastroscope for the stomach, laparoscope for abdomen, and proctoscope for lower bowel.

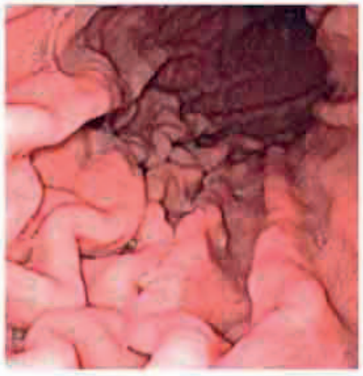

Endoscopic view of stomach
The gastric mucosa (inner lining) of a healthy stomach as seen through an endoscope. This procedure may be performed to investigate upper digestive tract disorders.

MRI AND CT SCANNING

Computerized tomography (CT) and magnetic resonance imaging (MRI) detail various tissue types. In CT, a scanner using X-rays rotates around the patient as a computer records the levels of electromagnetic energy passing through tissues of different densities. A cross section is built from layers of data. In MRI, a person lies in a magnetic chamber that causes hydrogen atoms in the body to align. A pulse of radiowaves is released, throwing the atoms out of alignment. As they realign, the atoms emit radio signals that are used to create an image.

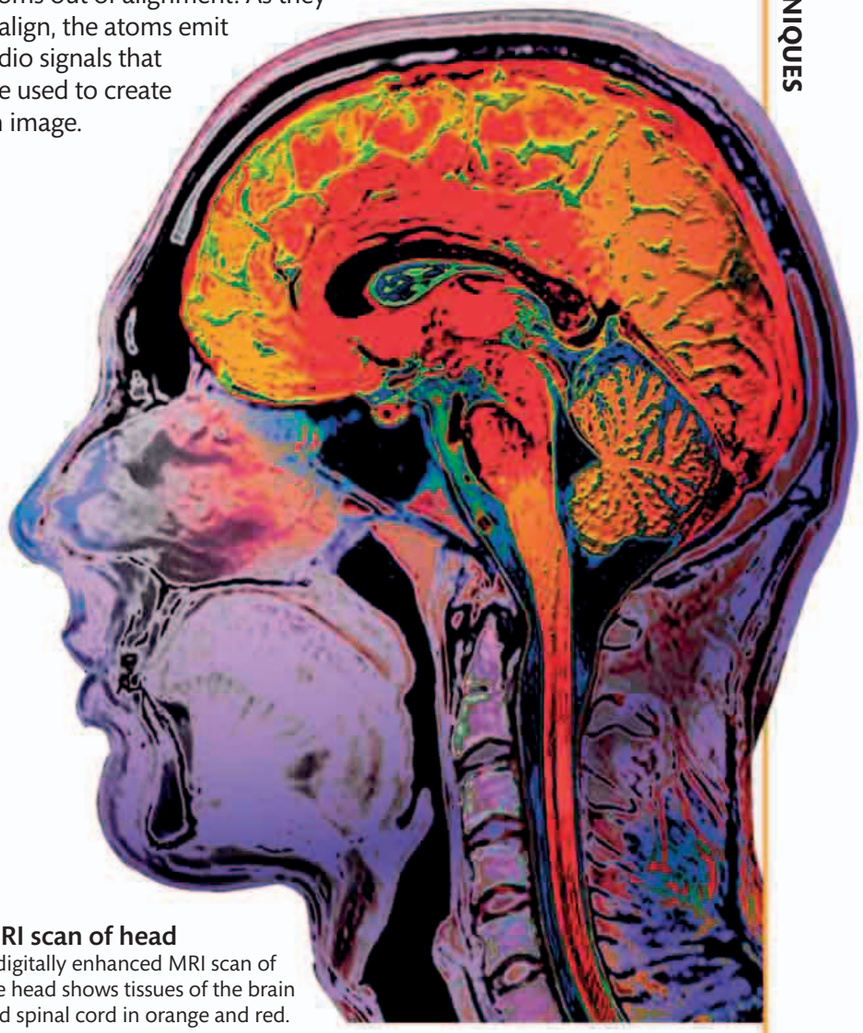

MRI scan of head
A digitally enhanced MRI scan of the head shows tissues of the brain and spinal cord in orange and red.

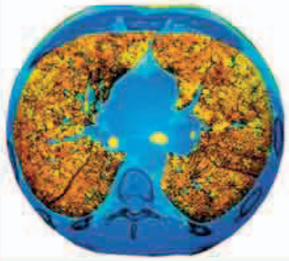

CT scan of lungs
In a horizontal slice through the chest, the spongy tissues and airways of the healthy lungs (oranges and yellows) show up clearly. The heart and major blood vessels between the lungs are mid-blue; the vertebrae, ribs, and sternum are dark blue.

ELECTRICAL ACTIVITY

Monitoring electrical activity in the body can reveal whether it is functioning normally. Signals coming from muscles and nerves are detected by applying sensor pads to the skin. The signals are sent to a computer that coordinates, amplifies, and displays them as a real-time trace—usually a spiked or wavy line. Examples of this technique include electrocardiography (ECG) of the heart, electromyography (EMG) of skeletal muscles, and electroencephalography (EEG) of the brain's nerve activity.

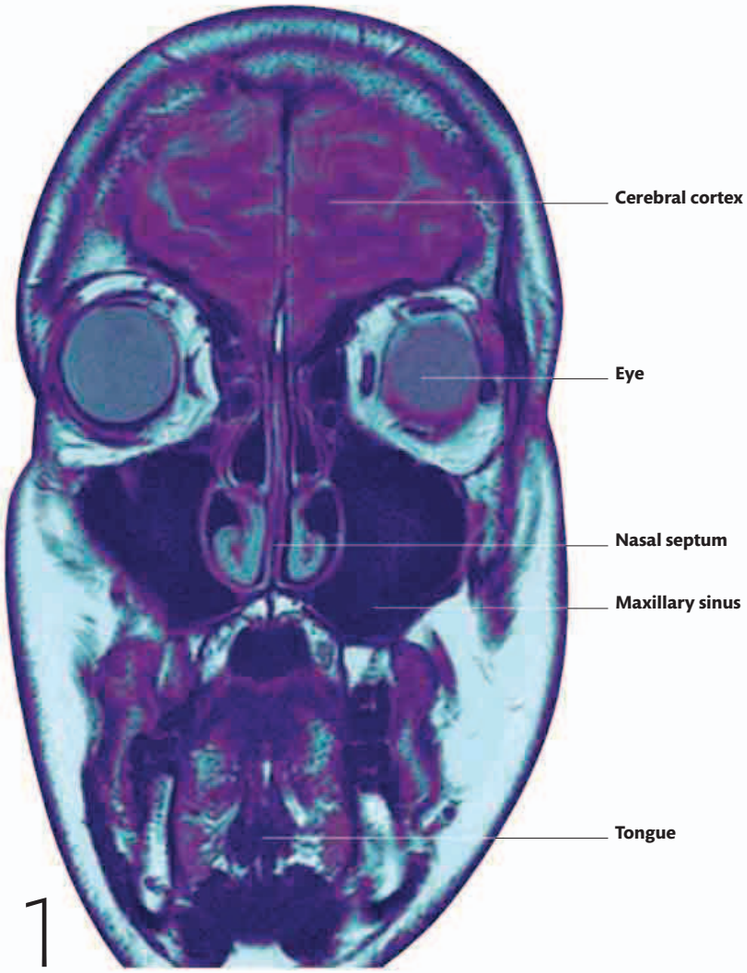

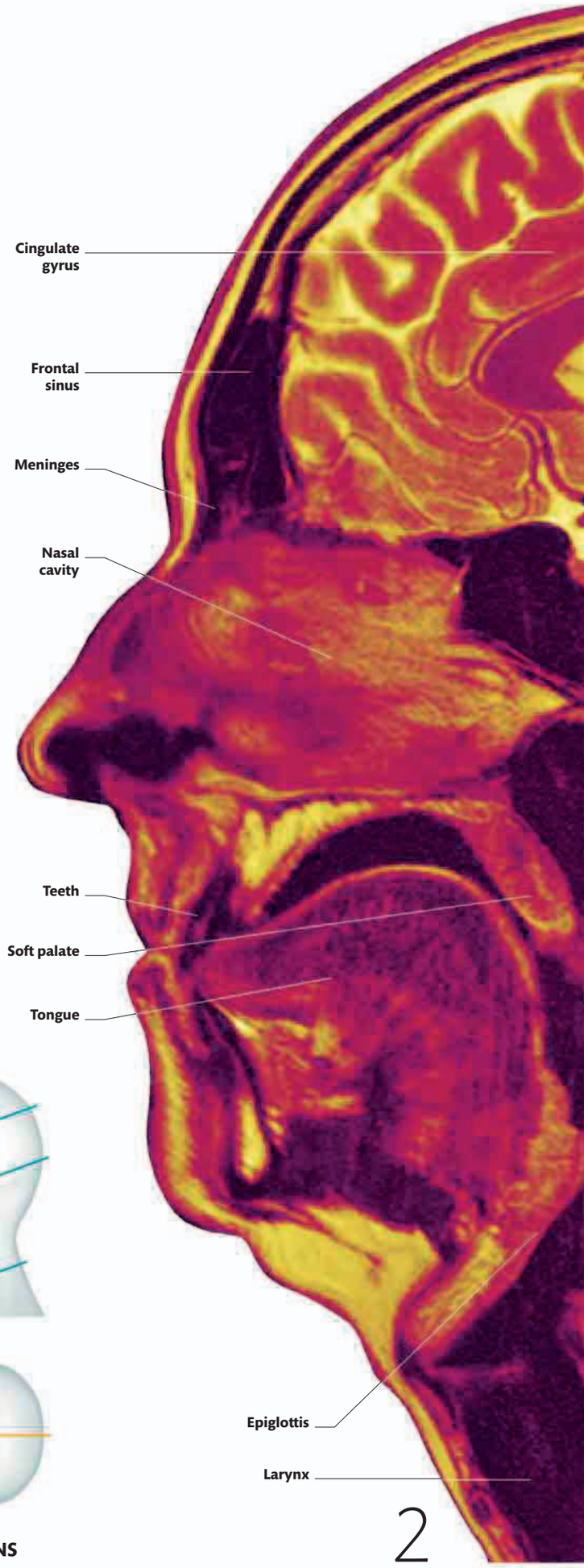

HEAD AND NECK

The discovery of X-rays at the end of the 19th century suddenly created the possibility of looking inside the human body—without having to physically cut it open. Medical imaging is now an important diagnostic tool, as well as being used for the study of normal anatomy and physiology. In computed tomography (CT), X-rays are used to produce virtual sections or slices through the body. Another form of sectional imaging, using magnetic fields rather than X-rays to create images, is magnetic resonance imaging (MRI), as shown here. MRI is very useful for looking in detail at soft tissue, for instance, muscle, tendons, and the brain. Also seen clearly in these sections are the eyes (1 and 3), the tongue (1 and 2), the larynx, vertebrae, and spinal cord (2 and 5).

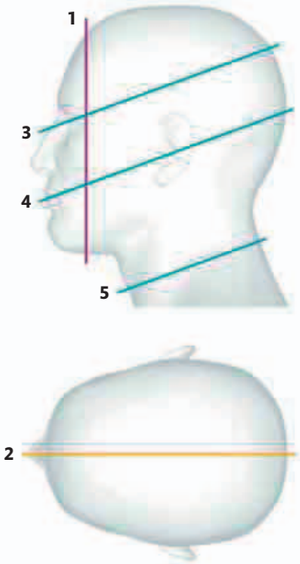

LEVELS OF SCANS

Skull
Corpus callosum
Thalamus
Hypothalamus
Pons
Cerebellum
Medulla oblongata
Spinal cord
Vertebra
Intervertebral disc

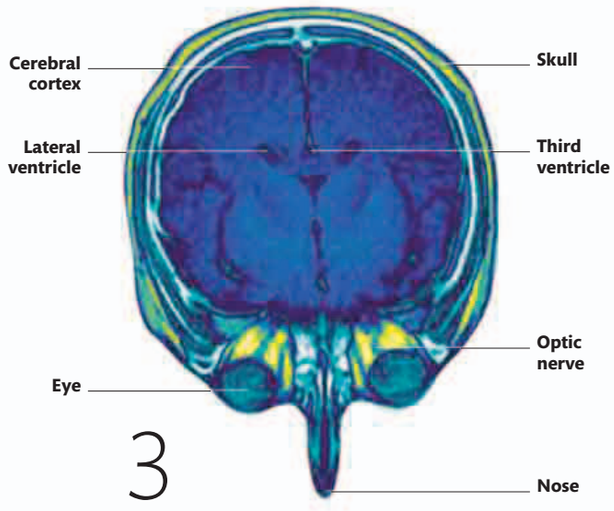

3

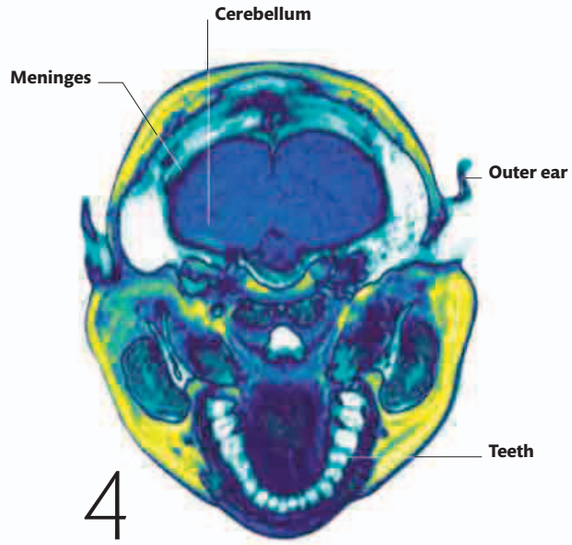

4

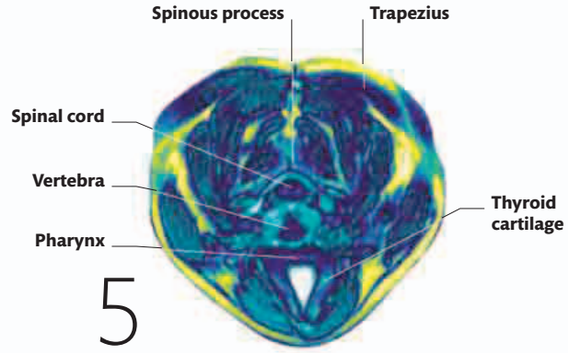

5

1

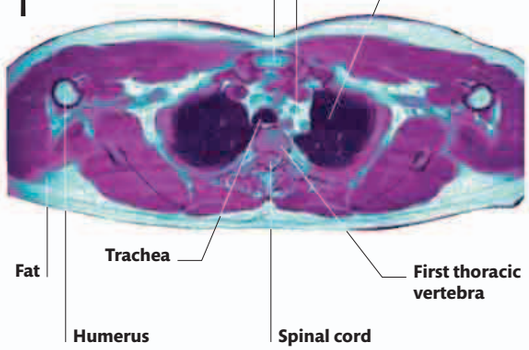
Sternum
Right common carotid artery
Apex of the left lung
Fat
Trachea
First thoracic vertebra
Humerus
Spinal cord

2

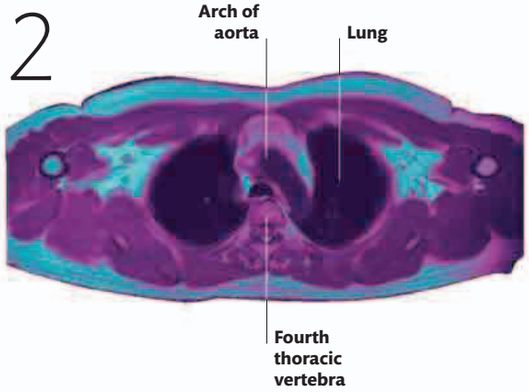
Arch of aorta
Lung
Fourth thoracic vertebra

4

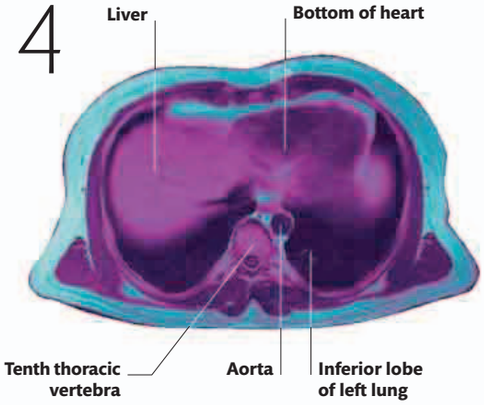
Liver
Bottom of heart
Tenth thoracic vertebra
Aorta
Inferior lobe of left lung

3

Lung
Superior vena cava
Left atrium
Sternum
Inferior lobe of right lung
Right inferior pulmonary artery
Seventh thoracic vertebra
Spinal cord

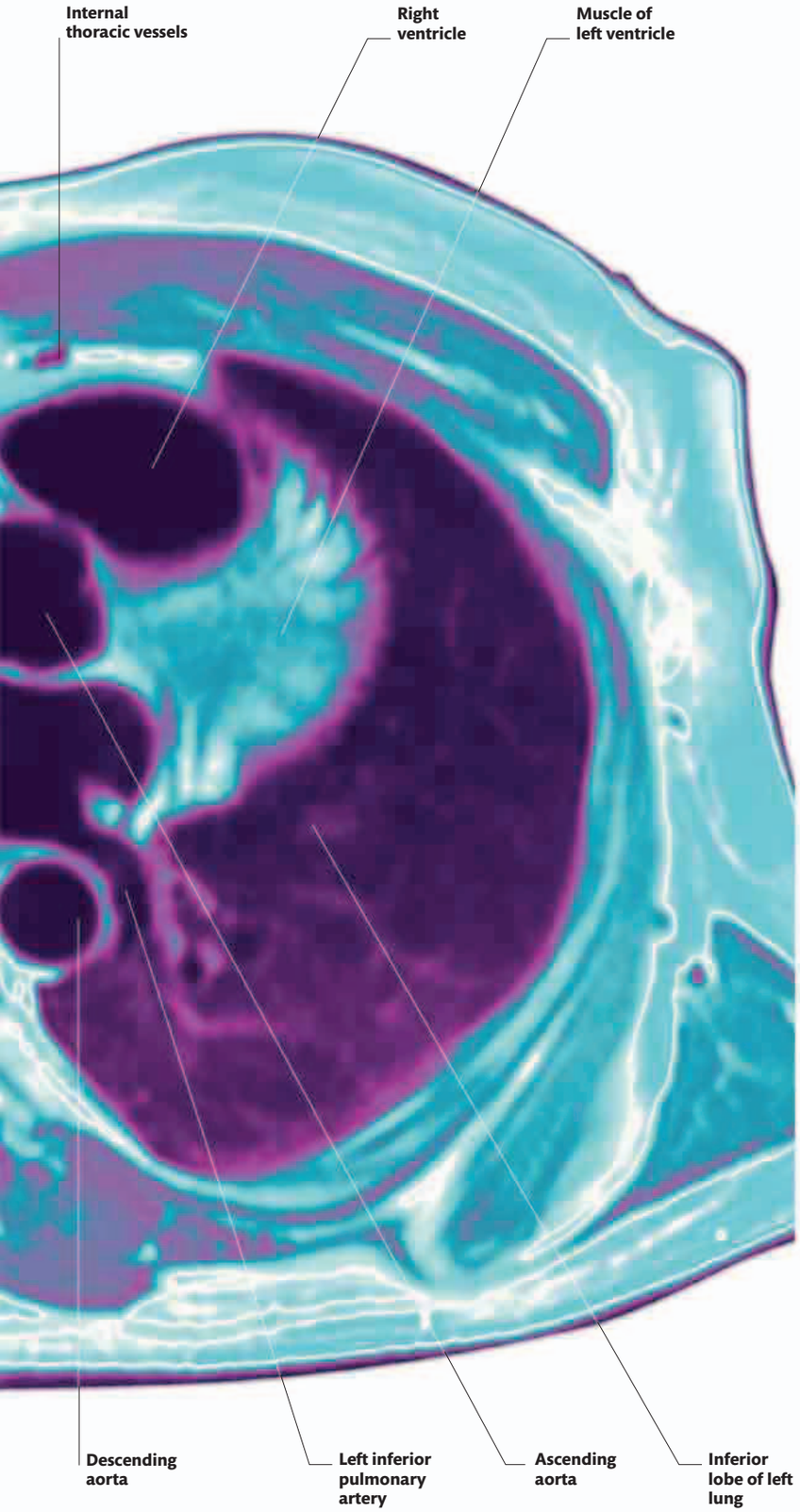

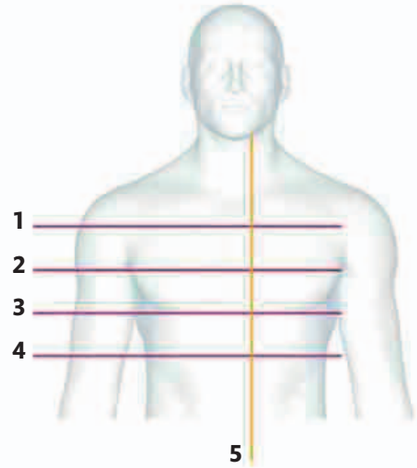

LEVELS OF SCANS

THORAX

The axial, or transverse, sections through the chest (sections 1–4) show the heart and large blood vessels lying centrally within the thorax, flanked by the lungs, and all set within the protective, bony casing of the ribcage. Section 1 shows the clavicles, or collarbones, joining the sternum at the front, the apex (top) of the lungs, and the great vessels passing between the neck and the thorax. Section 2 is lower down in the chest, just above the heart, while section 3 shows the heart with detail of its different chambers. The aorta appears to be to the right of the spine in this image, rather than to the left, but this is the usual way in which scans are viewed. You need to imagine yourself standing at the foot of the bed, looking down at the patient. This means that the left side of the body appears on the right side of the image as you view it. Section 4 shows the very bottom of the heart, and the inferior lobes of the lungs.

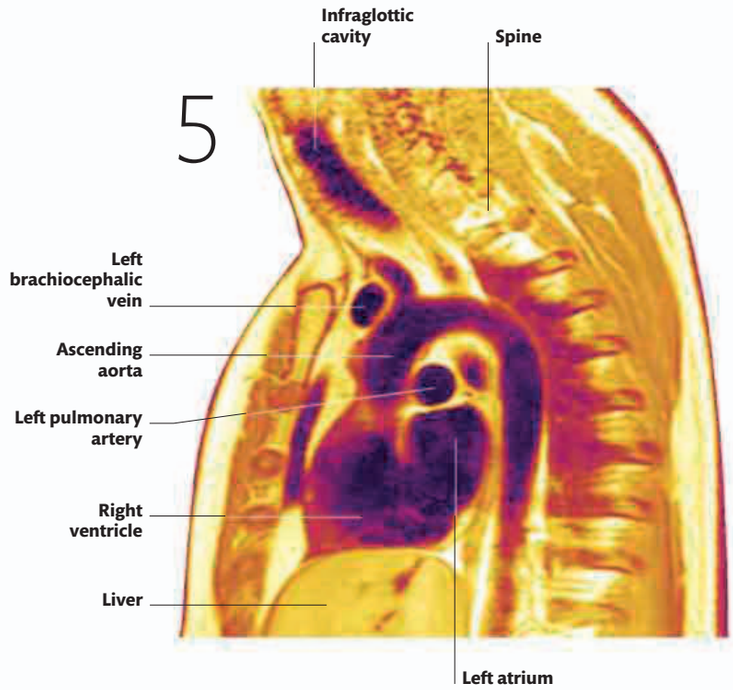

1

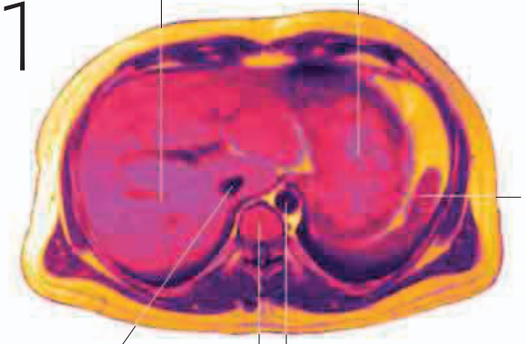
Liver
Stomach
Spleen
Inferior vena cava
Lumbar vertebra
Aorta

3

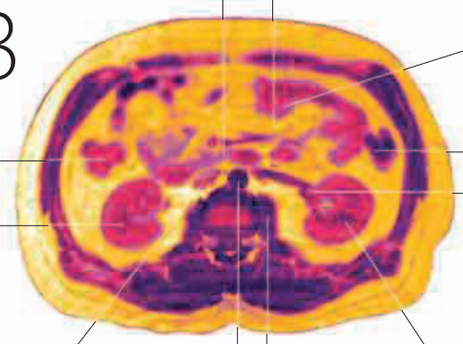
Duodenum
Pancreas
Stomach
Ascending colon
Descending colon
Renal hilum
Right kidney
Inferior vena cava
Aorta
Psoas
Left kidney

2

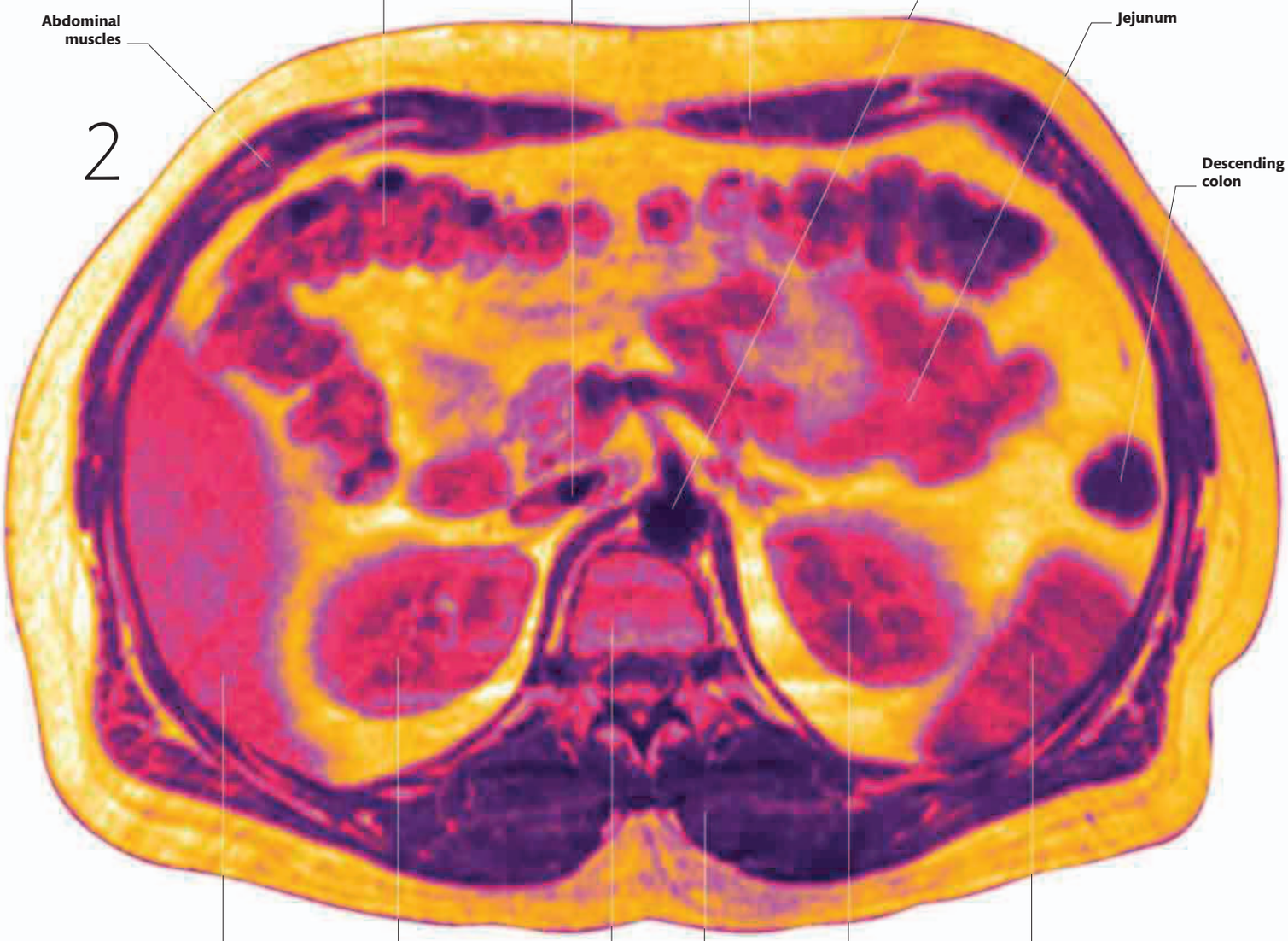
Colon
Inferior vena cava
Rectus abdominus
Aorta
Abdominal muscles
Jejunum
Descending colon
Liver
Right kidney
Lumbar vertebra
Erector spinae
Left kidney
Spleen

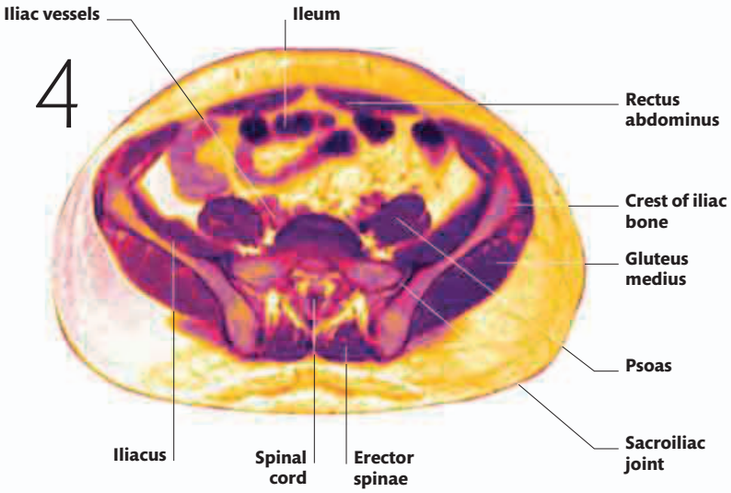

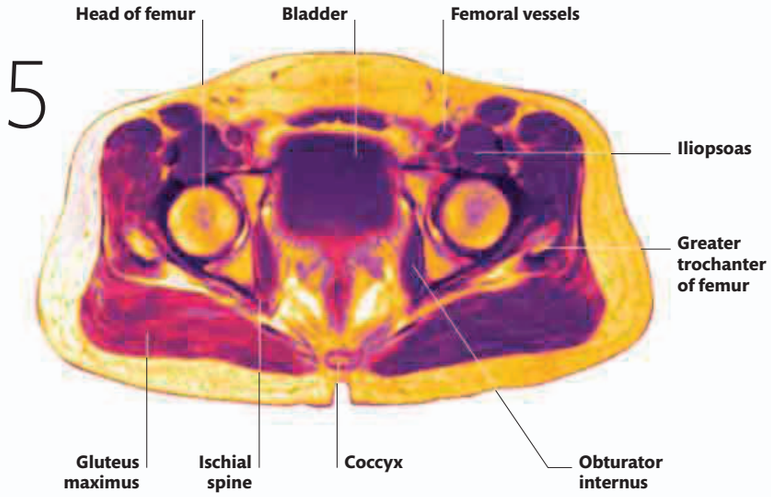

ABDOMEN AND PELVIS

MRI is a useful way of looking at soft tissues—and for visualizing the organs of the abdomen and pelvis, which only appear as subtle shadows on a standard X-ray. In the series of axial or transverse sections through the abdomen and pelvis, we can clearly see the dense liver, and blood vessels branching within it (section 1); the right kidney lying close to the liver, and the left kidney close to the spleen (section 2); the kidneys at the level where the renal arteries enter them (section 3), with the stomach and pancreas lying in front; coils of small intestine, the ileum, resting in the lower part of the abdomen, cradled by the iliac bones (section 4); and the organs of the pelvis at the level of the hip joints (section 5). The sagittal view (section 6) shows how surprisingly shallow the abdominal cavity is, in front of the lumbar spine. In a slim person, it is possible to press down on the lower abdomen and feel the pulsations of the descending aorta—right at the back of the abdomen.

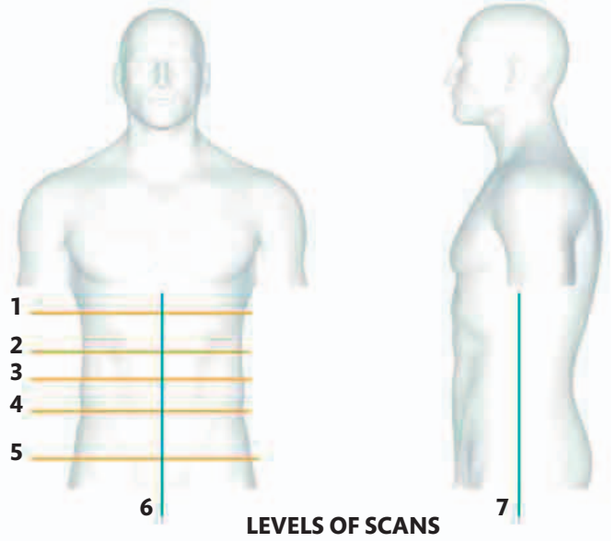

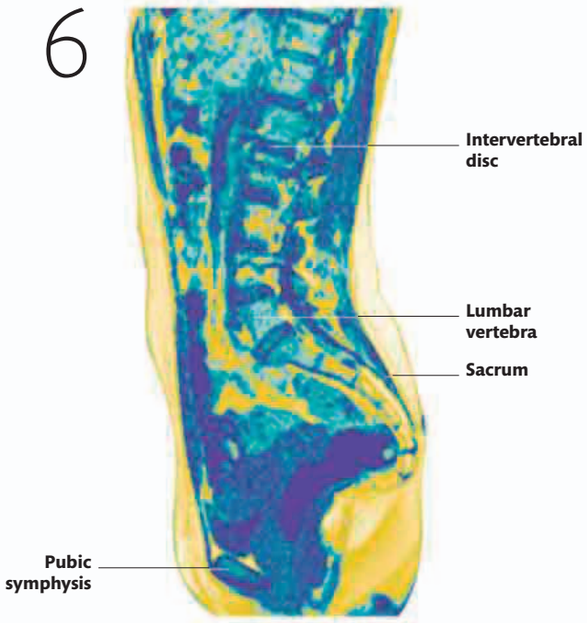

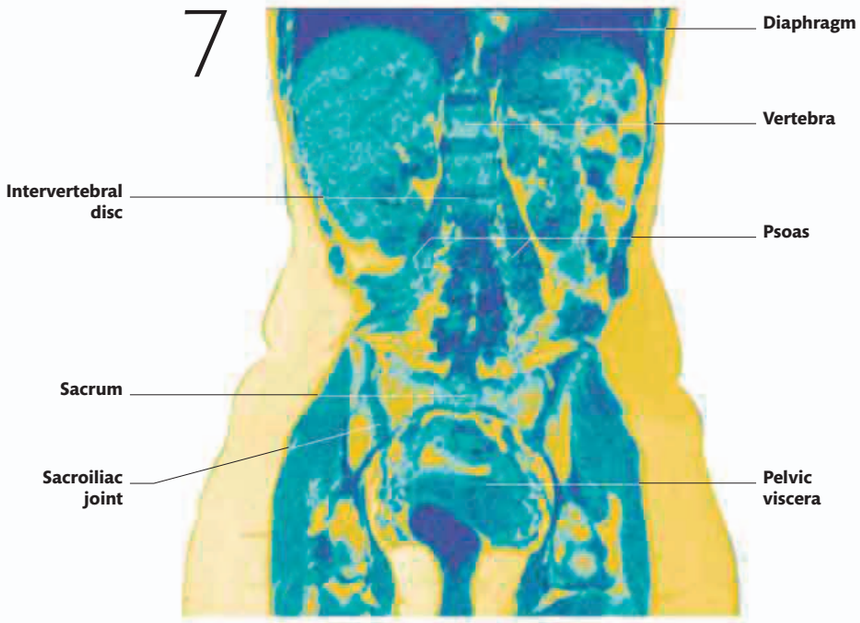

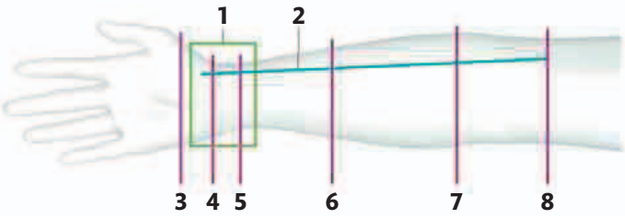

LEVELS OF SCANS

LOWER ARM AND HAND

These scans of the arm, forearm, and hand show how tightly packed the structures are. Section 1 reveals the bones of the wrist—the carpals—interlocking like a jigsaw. The wrist joint itself is the articulation between the radius and the scaphoid and lunate bones. In section 2, part of the elbow joint is visible, with the bowl-shaped head of the radius cupping the rounded end of the humerus. Muscles in the forearm are grouped into two sets, flexors on the front and extensors behind the forearm bones and interosseous membrane. Compare sections 3–8 with sections through the leg (see pp.234–35)—both limbs have a single bone (humerus or femur) in the upper part, two bones in the lower part (radius and ulna in the forearm; tibia and fibula in the lower leg), a set of bones in the wrist and ankle (carpals and tarsals), fanning out to five digits at the end of the limb. Evolutionarily, these elements developed from the rays of a fish fin.

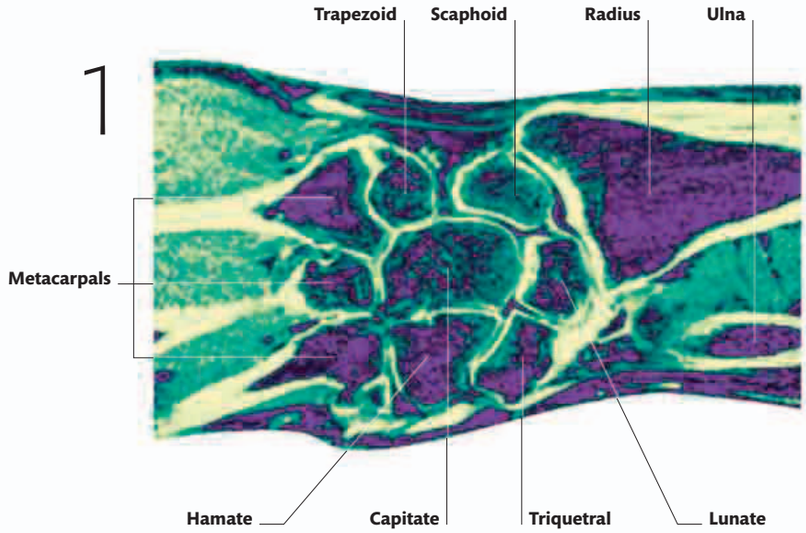

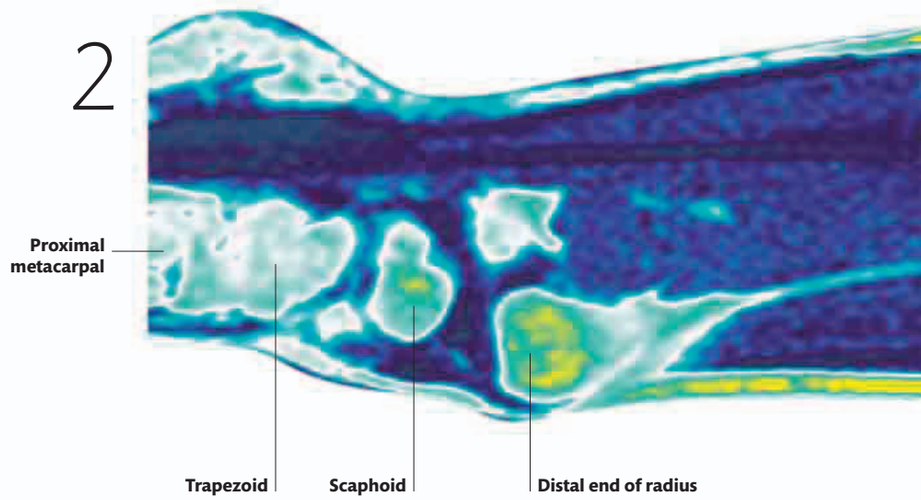

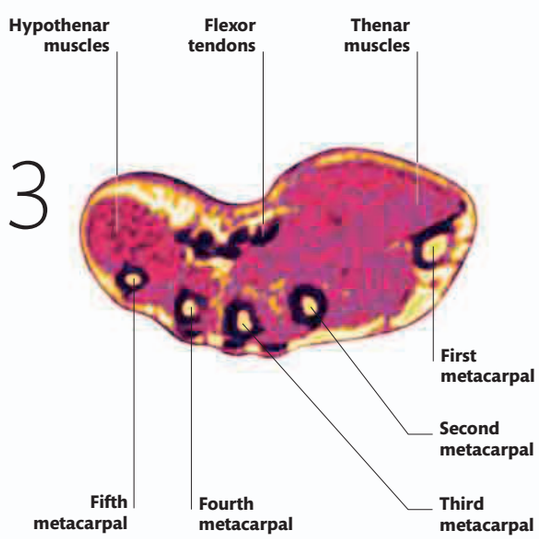

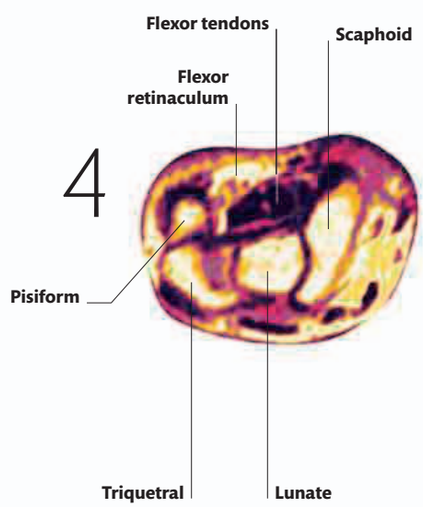

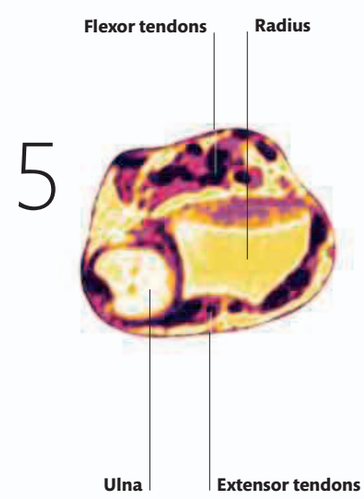

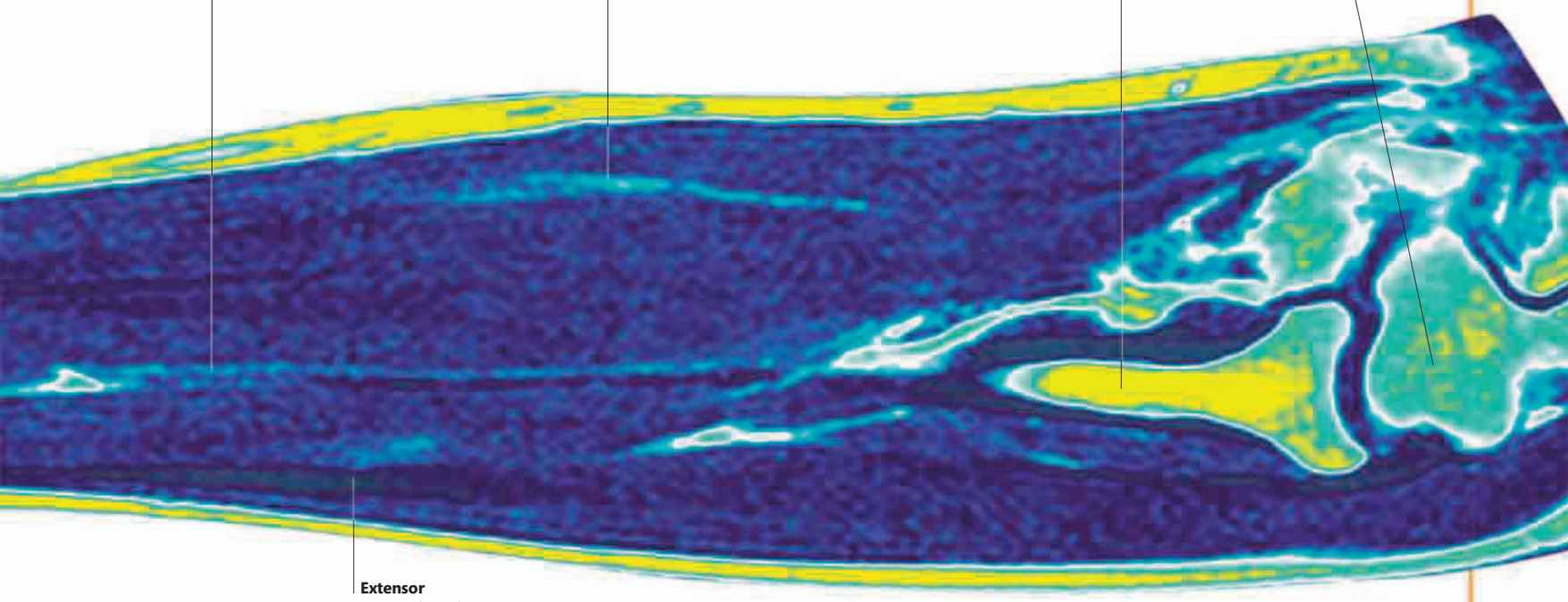
Interosseous membrane
Flexor compartment of forearm
Radius
Humerus
Extensor compartment of forearm

6

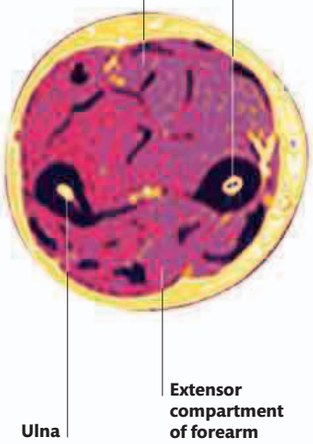
Flexor compartment of forearm
Radius
Ulna
Extensor compartment of forearm

7

Ulna
Radius

8

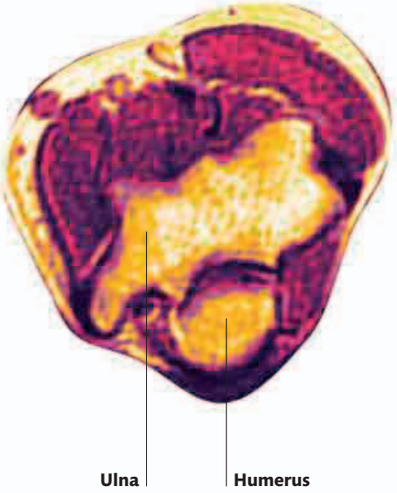
Ulna
Humerus

1

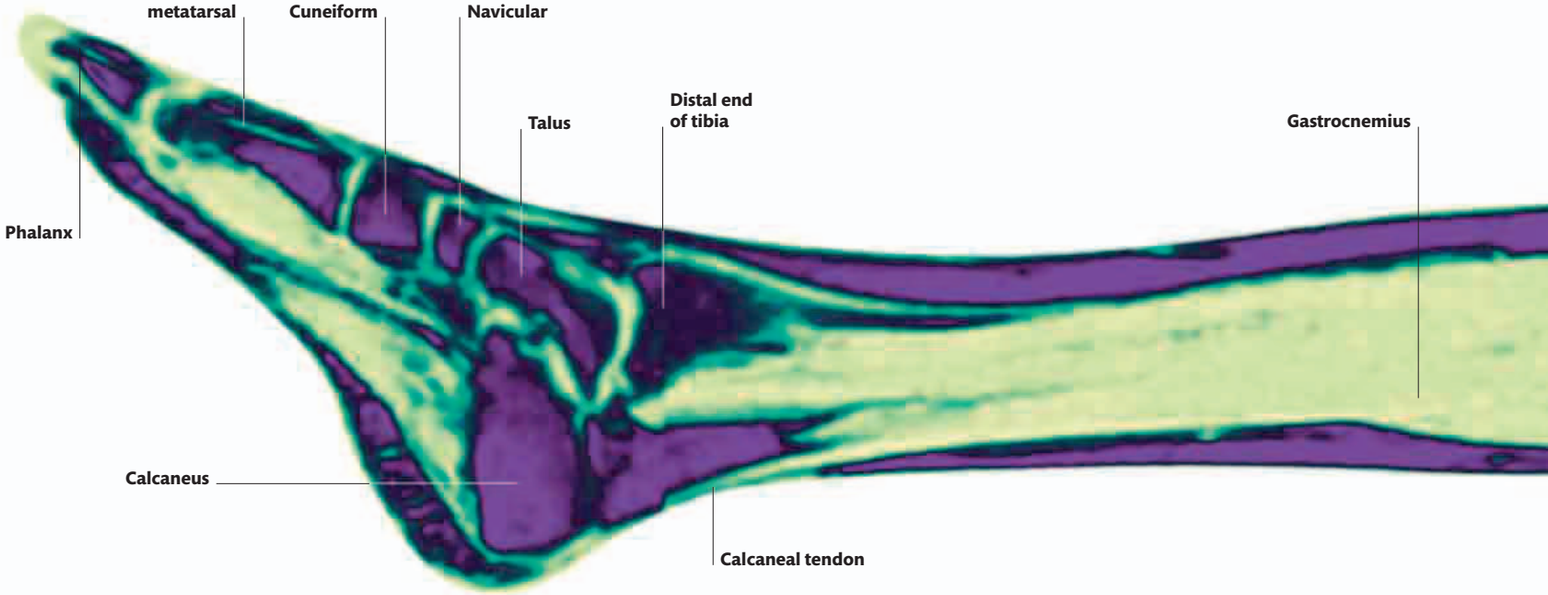

2

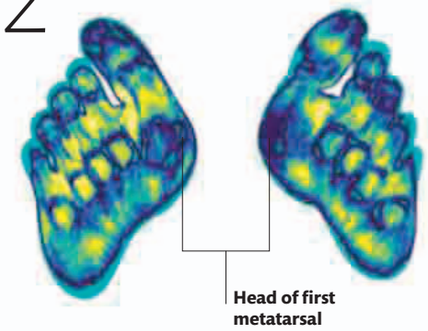

3

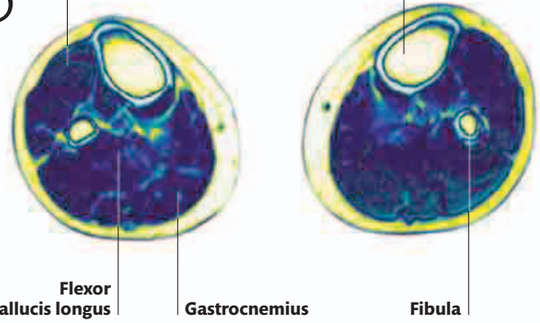

4

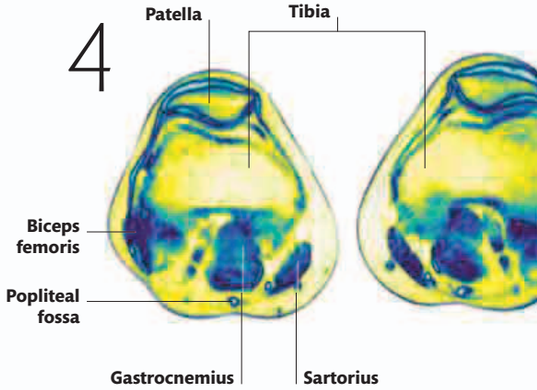

LOWER LIMB AND FOOT

The sequence of axial and transverse sections through the thigh and lower leg show how the muscles are arranged around the bones. Groups of muscles are bound together with fascia—fibrous packing tissue—forming three compartments in the thigh (the flexor, extensor, and adductor muscles), and three in the lower leg (flexor, extensor, and peroneal or fibular muscles). Nerves and deep blood vessels are also packaged together in sheaths of fascia, forming "neurovascular bundles". Section 2 shows the bones of the forefoot, while the tightly packed muscles surrounding the tibia and fibula in the lower leg are visible in section 3. At the knee joint, shown in section 4, the patella can be seen to fit neatly against the reciprocal shape of the femoral condyles. The neurovascular bundle is clearly visible here, at the back of the knee, in a space known as the popliteal fossa—with the hamstring muscles on either side. Sections 5 and 6, through the middle and upper thigh, show the powerful quadriceps and hamstring muscles surrounding the thigh bone, or femur.

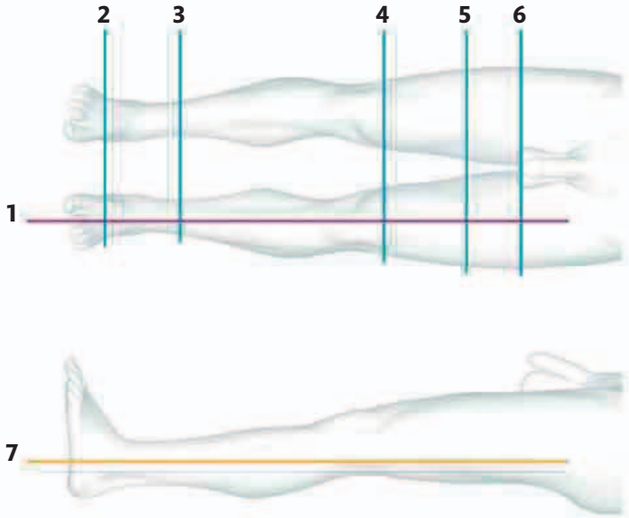

LEVELS OF SCANS

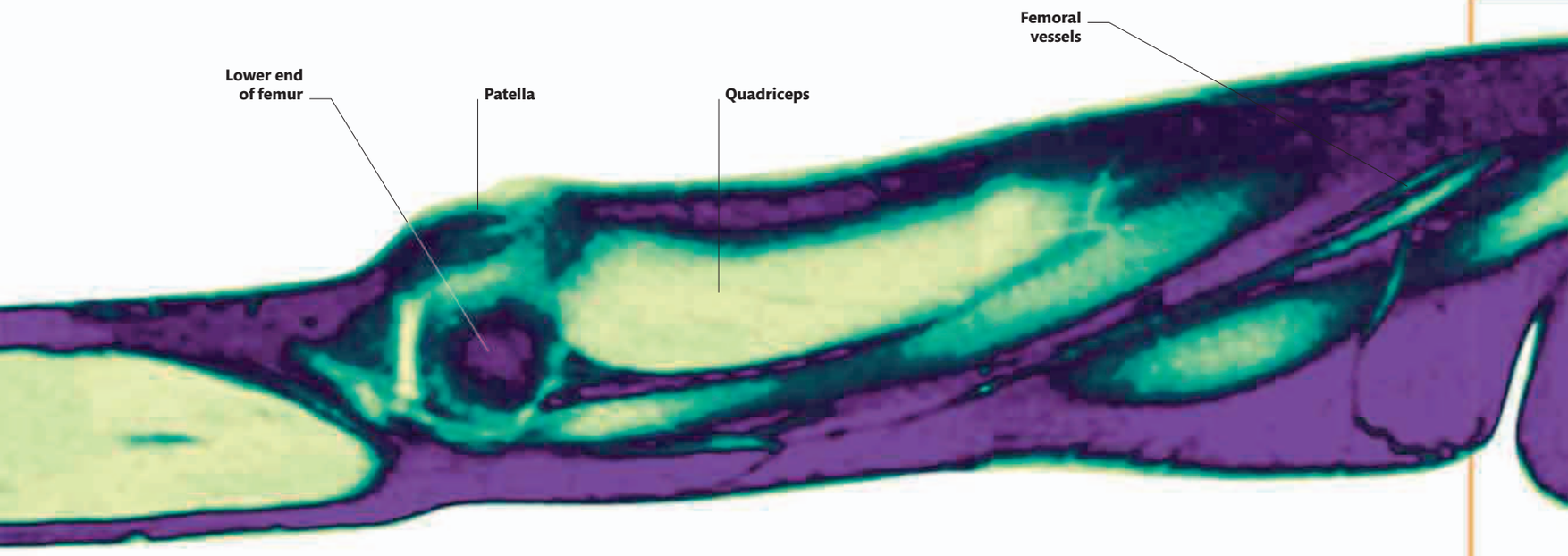
Femoral vessels
Lower end of femur
Patella
Quadriceps

5

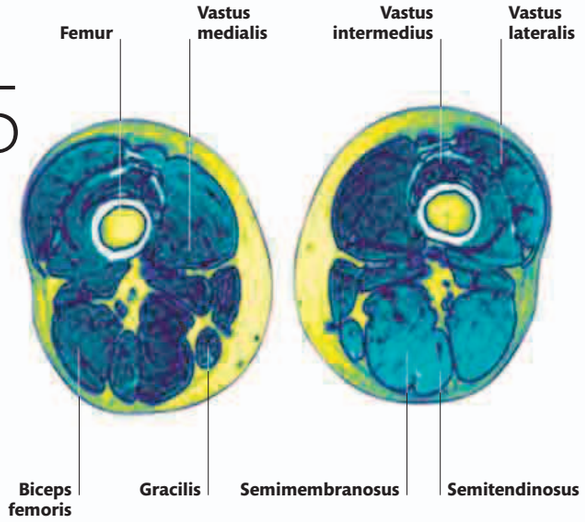
Femur
Vastus medialis
Vastus intermedius
Vastus lateralis
Biceps femoris
Gracilis
Semimembranosus
Semitendinosus

6

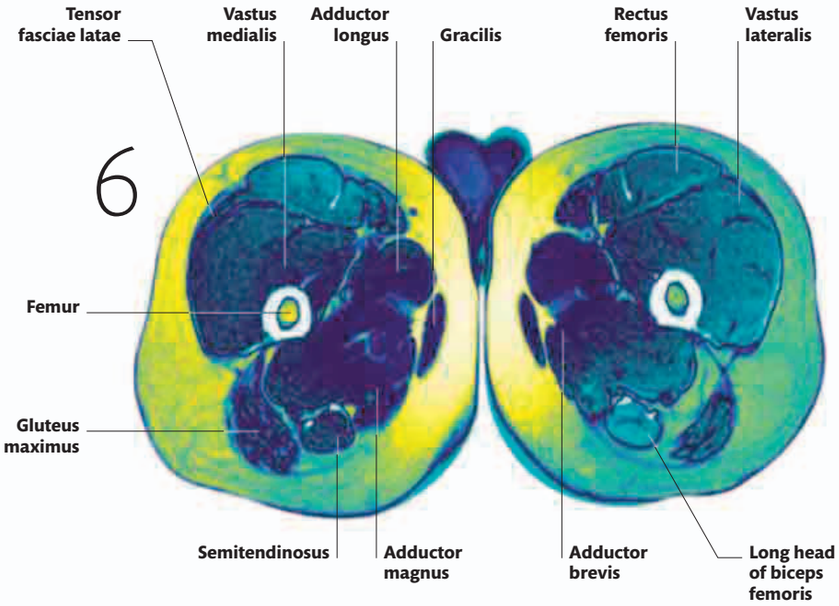
Tensor fasciae latae
Vastus medialis
Adductor longus
Gracilis
Rectus femoris
Vastus lateralis
Femur
Gluteus maximus
Semitendinosus
Adductor magnus
Adductor brevis
Long head of biceps femoris

7

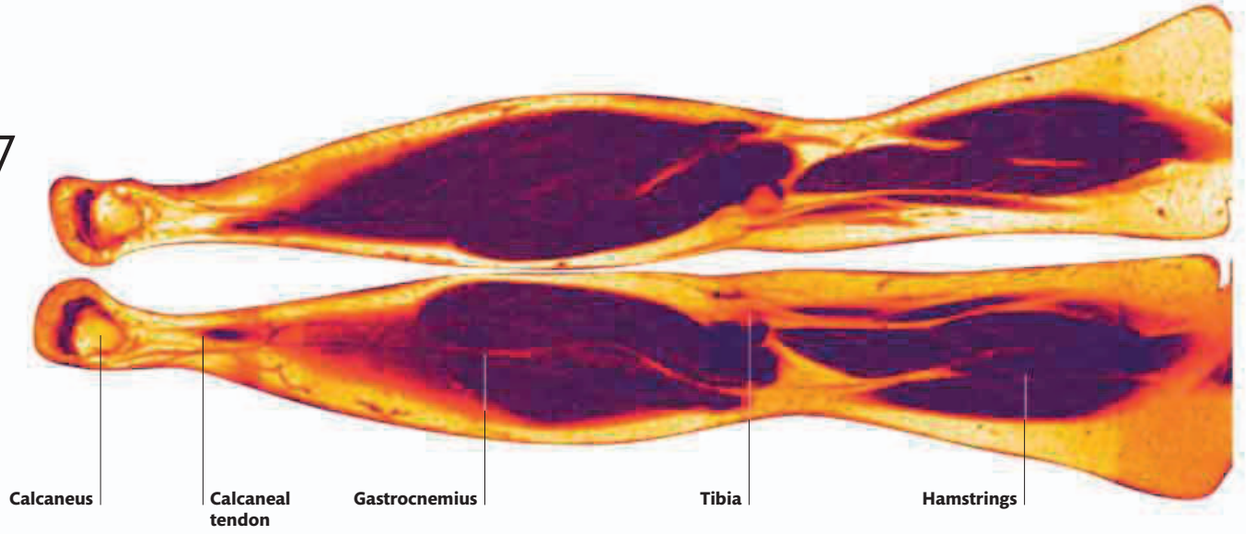
Calcaneus
Calcaneal tendon
Gastrocnemius
Tibia
Hamstrings

Get Full Access and More at ExpertConsult.com

ATLAS OF PAIN INJECTION TECHNIQUES

SECOND EDITION

Therese O'Connor & Stephen Abram

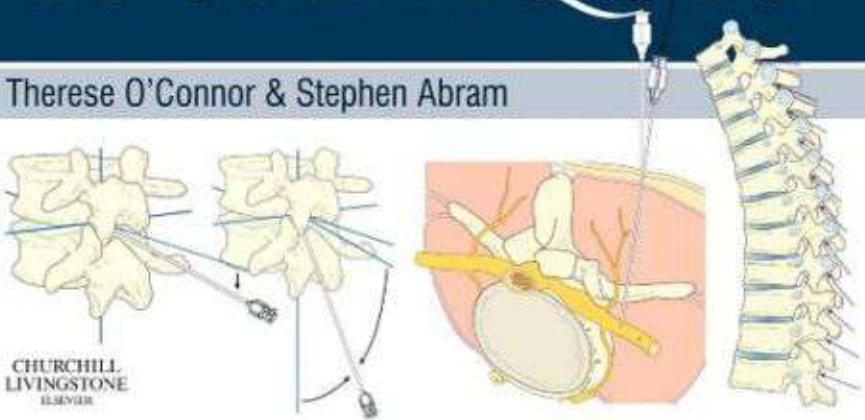

CHURCHILL LIVINGSTONE
ELSEVIER

JOINT INJECTIONS 2

2.1 LUMBAR FACET JOINT INJECTION

Anatomy

The zygopophyseal or facet joints (Fig. 2.1.1) are paired articular surfaces between the posterior aspects of adjacent vertebrae. In the cervical region, rotation and flexion are possible as the joint surfaces lie midway between the coronal and the axial planes. Rotation is prevented in the lumbar region but flexion is possible as the anterior portion of the lumbar facet joints lie in the coronal plane and the posterior portions in the sagittal plane. In the thoracic region, the joints' inferior and superior articular surfaces overlap each other in an almost vertical incline.

The facet joints bear most of the shear forces when the spine is flexed. In addition, when the intervertebral discs are degenerated, the facet joints carry increased load and weight, especially when the spine is extended. Innervation of the facet joints is via the medial branches of the dorsal rami of the spinal nerves. These nerves also innervate the muscles and ligaments surrounding the joints. Each medial branch divides into proximal and distal branches (Fig. 2.1.2). The proximal branch innervates the

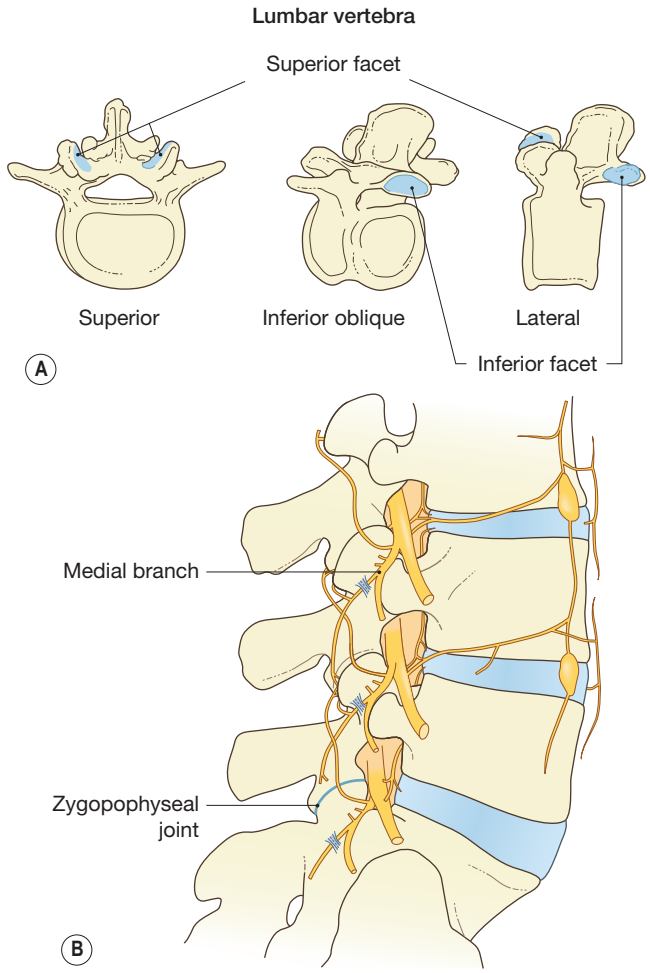

Fig. 2.1.1

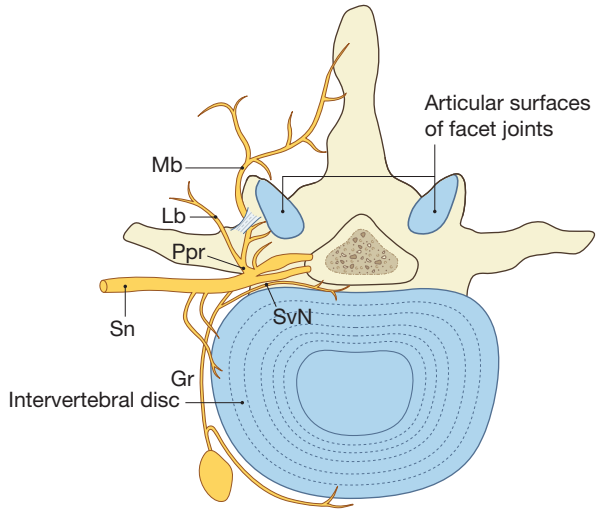

Fig. 2.1.2 Lumbar spine innervation. Innervation of the lumbar spinal structures in the transverse view. Note the posterior primary ramus (Ppr) leaving the spinal nerve (Sn) and splitting into a lateral branch (Lb) and a medial branch (Mb). The medial branch passes under the mamillo-accessory ligament to innervate the facet joint and capsule, the spinous process and the multifidus muscles. Sensory fibers traveling with the gray rami (Gr) form the sinu-vertebral nerve (SvN) and provide sensory function to the disc annulus. (Reproduced with permission from *Cousins and Bridenbaugh's Neural Blockade in Clinical Anesthesia and Management of Pain*, 4th edition, Wolters Kluwer/Lippincott Williams & Wilkins, 2009.)

adjacent facet joint, and the distal branch innervates the next facet joint below. The medial branch also innervates the interspinous ligaments and the multifidus muscles and the lateral branch innervates other adjacent muscles. Thus, pain from irritation of a joint may cause generalized sensitization of the dorsal rami with secondary hyperactivity and spasm of the innervated muscles and may be difficult to localize.

The facet joints contain vascular, highly innervated intra-articular synovial inclusions, which may become trapped and inflamed when the joint is injured, causing pain.

Equipment

- 2 ml and 10 ml syringes
- 25 G needle
- 22 G spinal needle, end-opening
- Non-ionic radio-opaque contrast medium
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- C-arm fluoroscopy or ultrasound

Drugs

- Lidocaine (lignocaine) 1% 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 25 mg (or its equivalent)
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 2.1.3*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The lumbar midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution.
- The spinous processes of the vertebrae are marked.
- **The insertion point of the needle lies 2–3 cm lateral to the cephalic end of the spinous process of the vertebra** (*Fig. 2.1.4 a,b*).
- C-arm fluoroscopy is positioned at an angle of about 30°, tilted towards the side of the joint to be injected. The angle is adjusted until the joint is well visualized. A radio-opaque object, e.g. the tip of a hemostat, is positioned over the joint and the skin is marked.
- Thereby, with the aid of fluoroscopy, the insertion point is identified.
- A skin wheal is raised and the area is infiltrated with lidocaine (lignocaine) 1%.
- A spinal needle is introduced in a **vertical direction to the skin**, until the needle is observed to enter the joint space, preferably near the lowest aspect of the joint (inferior recess). Confirmation of intra-articular placement is made by **observation of the needle tip remaining on the joint line as the fluoroscope is rotated laterally** (*Fig. 2.1.5*).
- After negative aspiration, 0.5 ml of non-ionic radio-opaque contrast medium (that is compatible with nerve tissue) is injected.
- The correct placement is indicated by outlining the joint with non-ionic radio-opaque contrast medium, visible on anteroposterior and oblique views (*Fig. 2.1.6 a,b*).

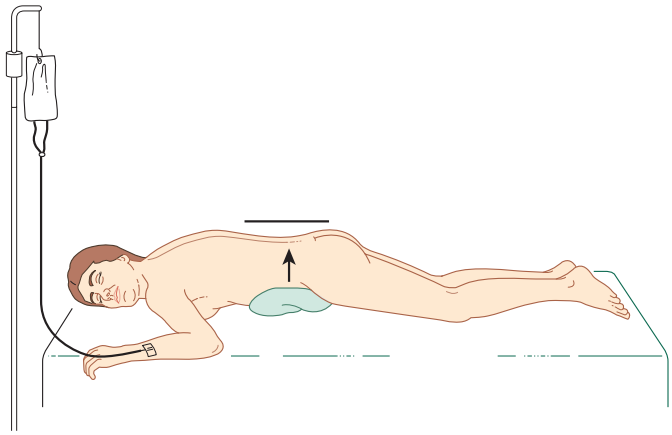

Fig. 2.1.3

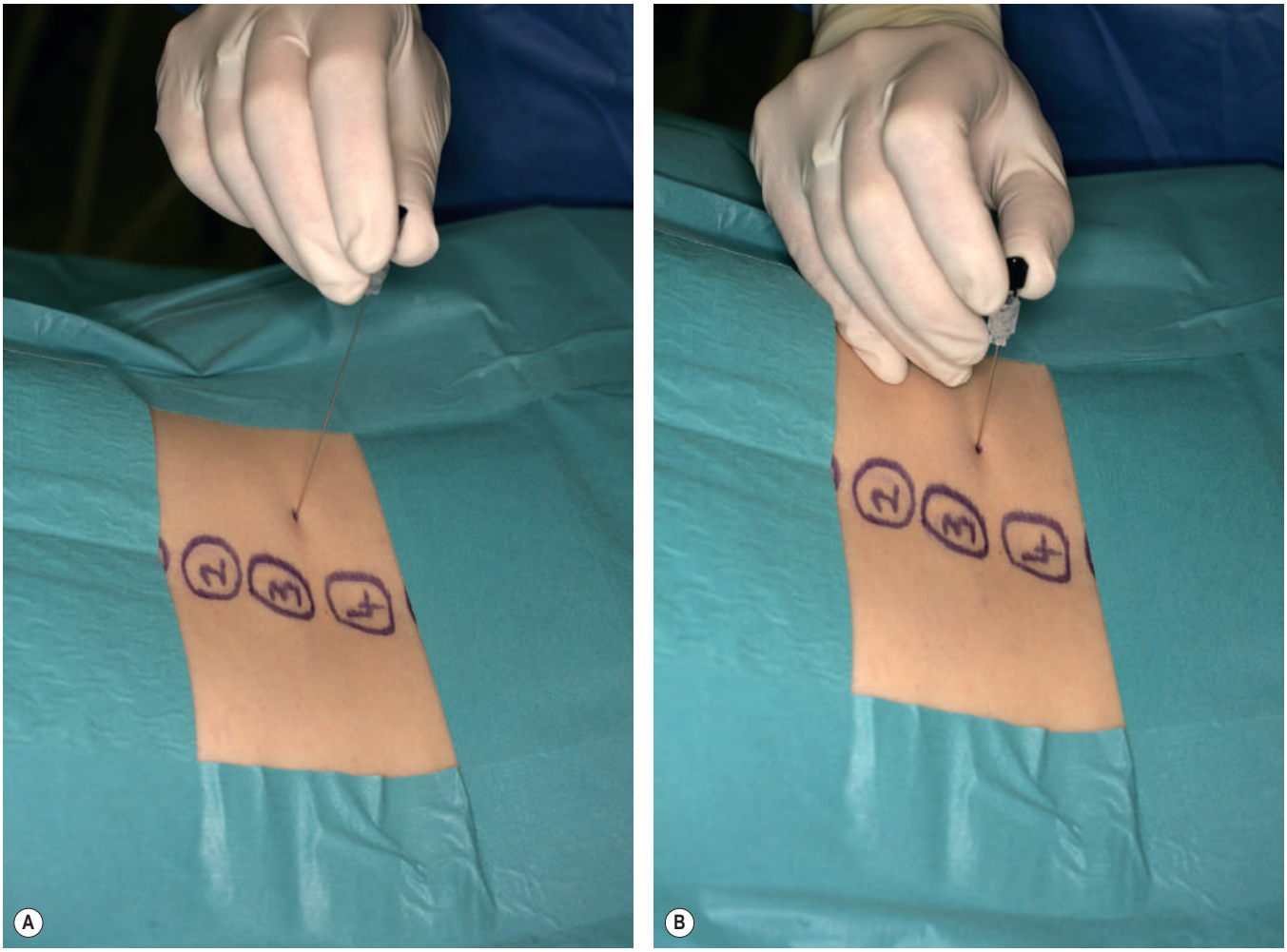

Fig. 2.1.4

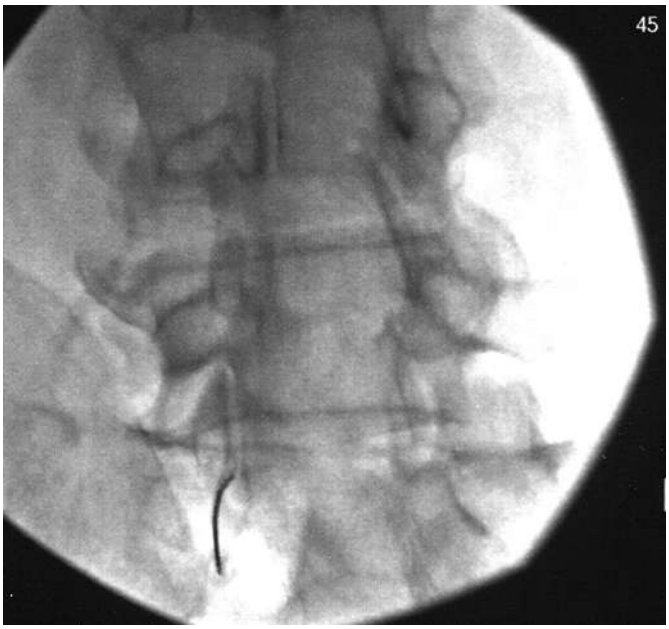

Fig. 2.1.5

Ultrasound may also be helpful in identifying the facet joint (*Fig. 2.1.7*).

- When the correct placement of the needle is confirmed, lidocaine (lignocaine) 1% 0.5 ml plus corticosteroid, e.g. triamcinolone diacetate 25 mg, may be injected and the needle removed while clearing with lidocaine (lignocaine) 1% 1 ml.

Confirmation of a successful injection

- Relief of pain.

Tips

- Care must be taken to inject only a small amount of volume as described above. A total volume of more than 1 ml may damage the joint. If the joint is disrupted anteriorly, drug may spread to the epidural space.

Potential problems

- Solution may spread to the epidural space via the anteromedial capsule.
- Nerve root injection.

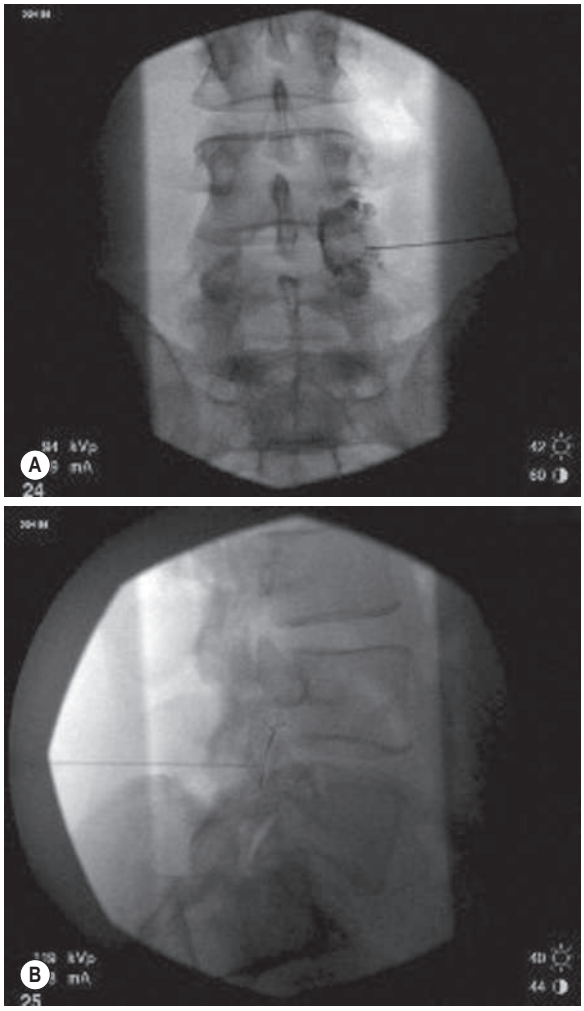

Fig. 2.1.6

Lumbar Facet Nerve Injection

- Facet nerve injection may be carried out by placing a spinal needle at the point where the superior articular and transverse processes join as the median branch passes over the cephalad edge of the transverse process (*Fig. 2.1.8*).
- The direct posterior approach should be avoided as the needle placement may be obstructed by the superior portion of the facet joint.
- Approach to the target site from a **lateral oblique angle 30° to skin** is recommended.
- The needle is advanced towards the target site (the posterior-superior edge of the transverse process) until bone is encountered.
- It is recommended that the transverse process be approached first, to determine depth.
- The needle is then repositioned medially until the lateral edge of the facet joint is reached.
- The needle is then moved superiorly until it just "falls off" the superior edge of the transverse process (*Fig. 2.1.9*).
- The optimum position is obtained by repositioning the needle to the postero-superior edge of the transverse process.
- The patient may now report reproduction of back pain.
- Injection of lidocaine (lignocaine) 1% 0.5 ml plus triamcinolone diacetate 25 mg may be carried out for therapeutic effect. Diagnostic blockade may be unreliable as anesthesia of a facet joint means that both nerves supplying the joint should be blocked. However, this means that the joint above and the joint below will also be partially blocked and therefore diagnosis of pain in a particular joint using nerve block is not feasible.

Potential problems

- The same potential problems may occur as described for lumbar facet joint injection (see above).

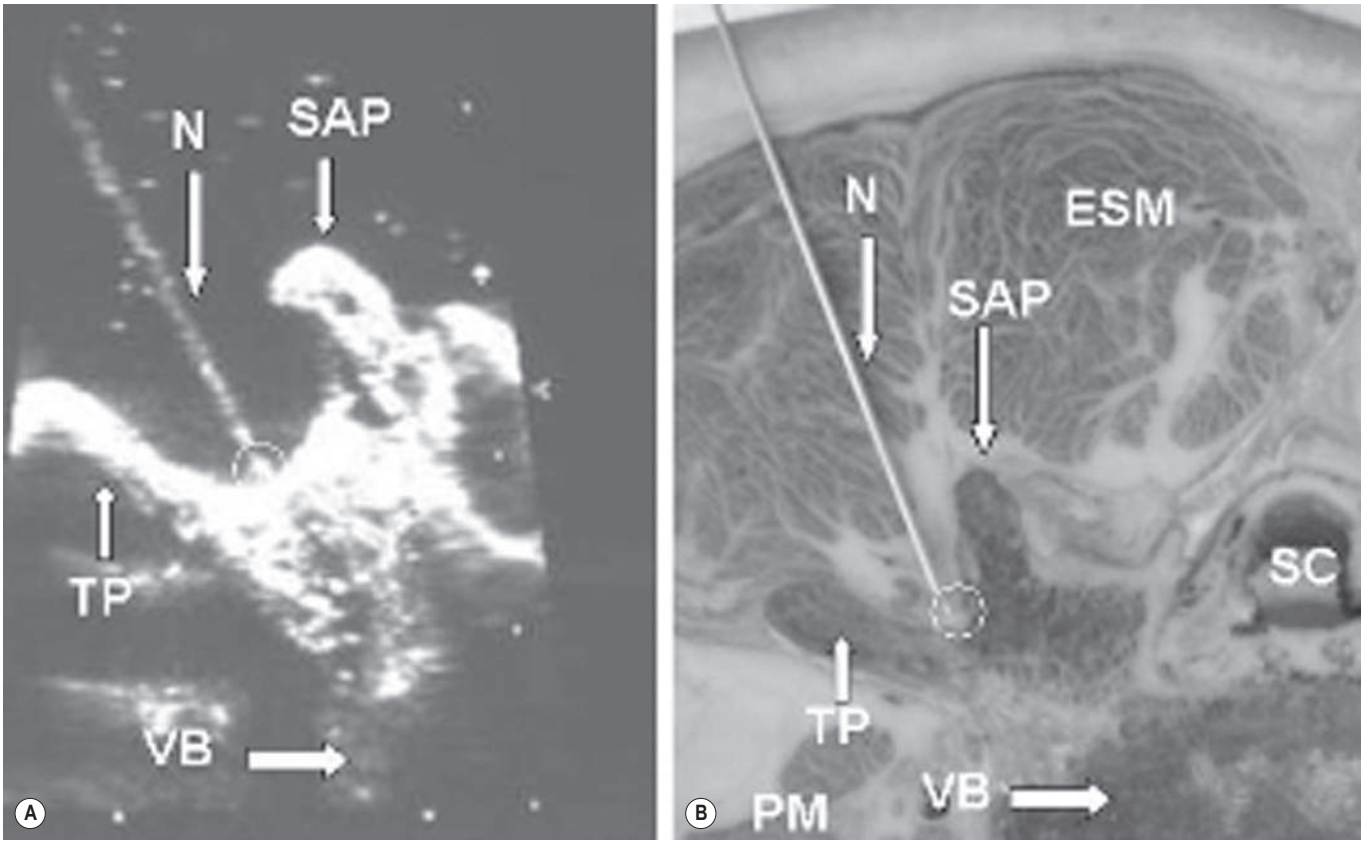

Fig. 2.1.7 **A** High-resolution sonogram (15-MHz linear transducer) of vertebral bone L3 immersed in water in the cross-axis view. **B** Corresponding anatomic cross-sectional cadaver preparation. Circles indicate targets. ESM erector spinae muscle; N needle; PM psoas muscle; SAP superior articular process; SC spinal channel; TP transverse process; VB vertebral body. (From Greher M, Scharbert G, Kamolz LP, et al, Ultrasound-guided lumbar facet nerve block: a sonoanatomic study of a new methodologic approach. Anesthesiology 2004; 100:1242–8 © 2004 American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins, Inc.)

Fig. 2.1.8

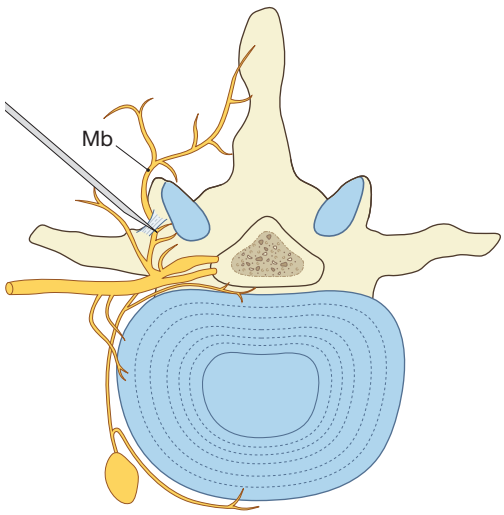

Fig. 2.1.9

2.2 CERVICAL FACET JOINT INJECTION

Anatomy

The anatomy relevant to injection of the cervical facet joints is similar to that relevant to the lumbar facet joints. The cervical facet joints below the C2–3 level are innervated by the medial branches of the cervical posterior primary rami. These divide into lateral and medial branches after leaving the posterior spinal canal and the splenius capitis muscles cover the medial branch posteriorly. The medial branches lie in close proximity to the vertebral artery and the epidural space is in close proximity to the anterior joint capsule (**Fig. 2.2.1**). The C2–3 facet joint is innervated by the medial branch of the third occipital nerve, which travels beneath the tendonous origin of the splenius capitis muscle where it may be accessed for local anesthetic blockade (**Fig. 2.2.2**).

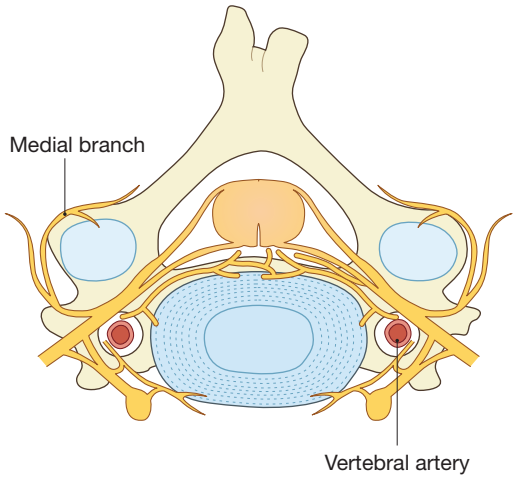

Fig. 2.2.1

Equipment

- 2 ml and 10 ml syringes
- 25 G needle
- 22 G spinal needle, end-opening
- Radio-opaque contrast medium
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- C-arm fluoroscopy or ultrasound

Drugs

- Lidocaine (lignocaine) 1%, 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 25 mg (or its equivalent)
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Neck slightly flexed (*Fig. 2.2.3*).

Needle puncture and technique

Caution: injection of 0.5–1 ml of lidocaine (lignocaine) 1% into the vertebral artery may result in immediate convulsion and/or loss of consciousness with possible cardiovascular system (CVS) collapse.

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The cervical midline and an area of 7 cm × 5 cm laterally is cleaned with antiseptic solution.
- The spinous processes are marked.

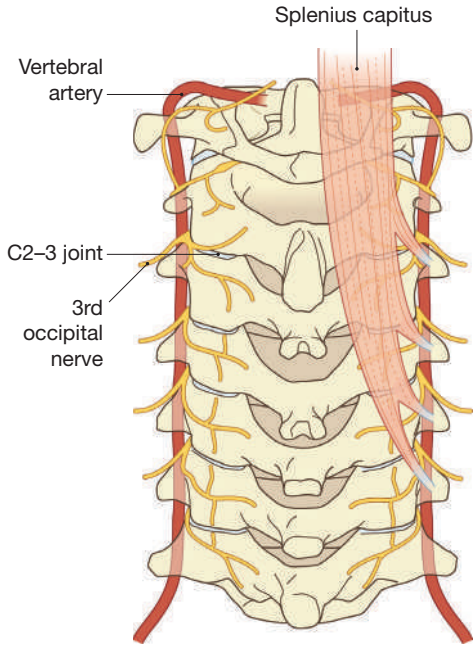

Fig. 2.2.2

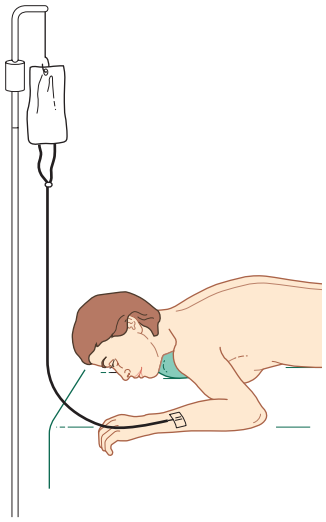

Fig. 2.2.3

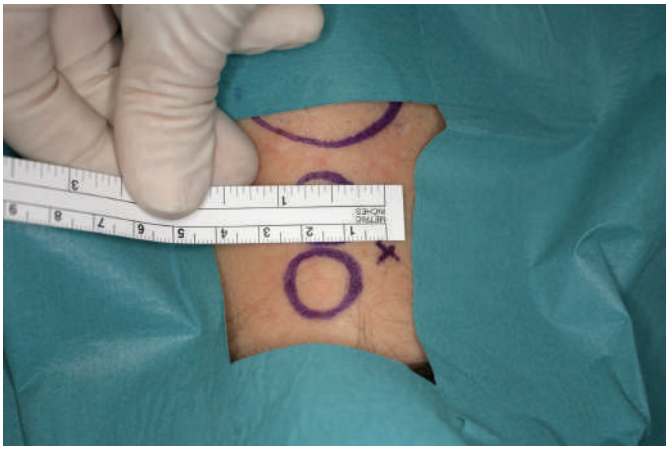

Fig. 2.2.4

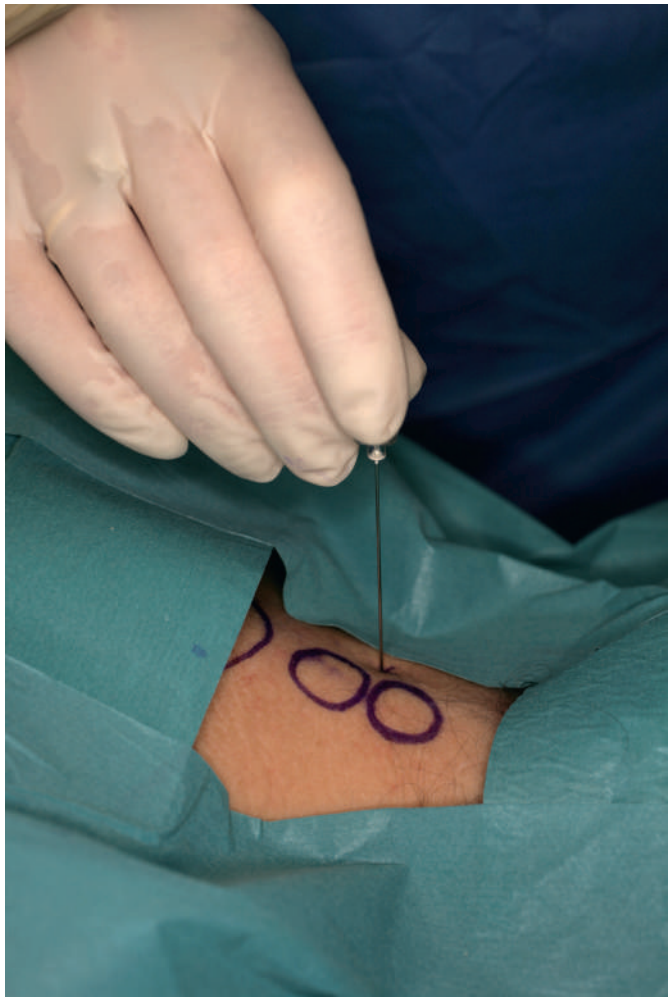

Fig. 2.2.5

- **The insertion point of the needle lies 2–3 cm lateral to the cephalic end of the spinous process of the vertebra** (*Fig. 2.2.4*).
- C-arm fluoroscopy is positioned at an angle of about 30°, tilted towards the side of the joint to be injected. The angle is adjusted until the joint is well visualized. A radio-opaque object, e.g. the tip of a hemostat, is positioned over the joint and the skin is marked.
- Thereby, with the aid of fluoroscopy, the insertion point is identified.
- A skin wheal is raised and the area is infiltrated with lidocaine (lignocaine) 1%.
- A spinal needle is introduced in a vertical direction to the skin, until the needle is observed to enter the joint space (*Fig. 2.2.5*). Confirmation of intra-articular placement is made by observation of the needle tip remaining on the joint line as the fluoroscope is rotated (*Fig. 2.2.6*) or on ultrasound.
- After negative aspiration, 0.5 ml of non-ionic radio-opaque contrast medium (that is compatible with nerve tissue) is injected.
- The correct placement is indicated by outlining the joint with non-ionic radio-opaque contrast medium, visible on anteroposterior and oblique views.
- When the correct placement of the needle is confirmed, lidocaine (lignocaine) 1% 0.5 ml plus corticosteroid, e.g. triamcinolone diacetate 25 mg, may be injected and the needle removed while clearing with lidocaine (lignocaine) 1% 1 ml.

Confirmation of a successful injection

- Relief of pain.

Fig. 2.2.6

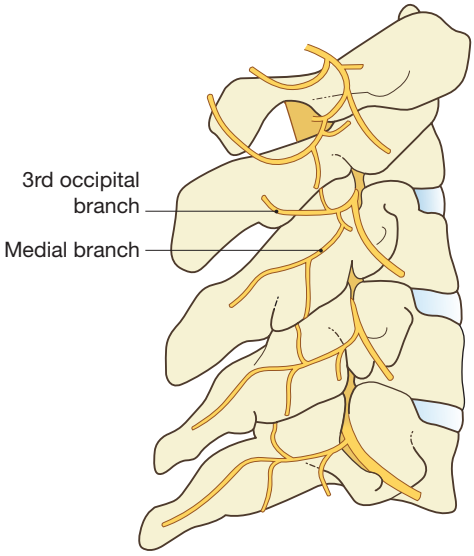

Fig. 2.2.7

Tips

- Care must be taken to inject only a small amount of volume, as described above. A total volume of more than 1 ml may damage the joint. If the joint is disrupted anteriorly, a drug may spread to the epidural space.

Potential problems

- Solution may spread to the epidural space via the anteromedial capsule.
- Nerve root injection.
- Intrathecal injection resulting in spinal anesthesia may occur if local anesthetic is inadvertently injected into the nerve root sleeve. Prompt recognition of this complication is vital during cervical procedures, because the patient's breathing may be arrested and there may be immediate convulsion and/or loss of consciousness with CVS collapse requiring immediate resuscitation. In addition, the patient's head should be immediately elevated after the injection to ensure that the lidocaine (lignocaine) flows inferiorly. Some practitioners elevate the head of the table during all cervical injections to help prevent this complication. Intravenous injection may be harmless, but it results in a suboptimal or false-negative result.
- Intra-arterial injection may result in immediate convulsion and/or loss of consciousness with possible CVS collapse. Intrarterial injection can be dangerous if the agent is injected into the vertebral artery or radicular branches that enter the neural foramina at various levels and, rarely, persisting paraplegia or paraparesis have been reported after cervical facet joint injection or nerve root block. These complications may be due to embolism from intra-arterial injection of particulate corticosteroid. However, even when contrast injection prior to steroid infiltration confirms extravascular needle placement, nerve damage may occur, which suggests an alternative cause for the complication, such as vasospasm or direct arterial injury from the needle-tip. Regardless of the cause, contrast injection is recommended to at least potentially reduce the risk of intravascular injection. Ultrasound will have limitations in this regard.
- Hematoma may occur (avoid performing block on patients who have coagulopathy).

CERVICAL FACET NERVE INJECTION

- Facet nerve injection may be carried out by placing a spinal needle at the point where the superior articular and transverse processes join as the median branch passes over the cephalad edge of the transverse process (*Fig. 2.2.7*).
- The direct posterior approach should be avoided as the needle placement may be obstructed by the superior portion of the facet joint.
- Approach to the target site from a **lateral oblique angle 30° to the skin** is recommended.
- The needle is advanced towards the target site (the posterior–superior edge of the transverse process) until bone is encountered.
- It is recommended that the transverse process be approached first, to determine depth.

- The needle is then repositioned medially until the lateral edge of the facet joint is reached.
- The needle is then moved superiorly until it just "falls off" the superior edge of the transverse process.
- The optimum position is obtained by repositioning the needle to the posterosuperior edge of the transverse process.
- The patient may now report reproduction of back pain.
- Injection of lidocaine (lignocaine) 1% 0.5 ml plus triamcinolone diacetate 25 mg may be carried out for therapeutic effect. Diagnostic blockade may be unreliable as anesthesia of a facet joint means that both nerves supplying the joint should be blocked. However, this means that the joint above and the joint below will also be partially blocked and therefore diagnosis of pain in a particular joint using nerve block is not feasible.

2.3 SACRO-ILIAC JOINT INJECTION

Anatomy

The surfaces of the sacrum and ilium form a synovial joint, the sacro-iliac joint. Ligaments and connective tissue surround the joint, conferring stability and preventing excessive movement of the joint (**Figs 2.3.1**, **2.3.2**). The joint is innervated by L4, L5, S1 (the superior gluteal nerve), S2, and L3. Localization of the pain is therefore difficult due to this wide nerve supply to the joint.

Equipment

- 2 ml and 10 ml syringes
- 25 G needle
- 22 G spinal needle, end-opening
- Radio-opaque contrast medium
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- C-arm fluoroscopy or ultrasound

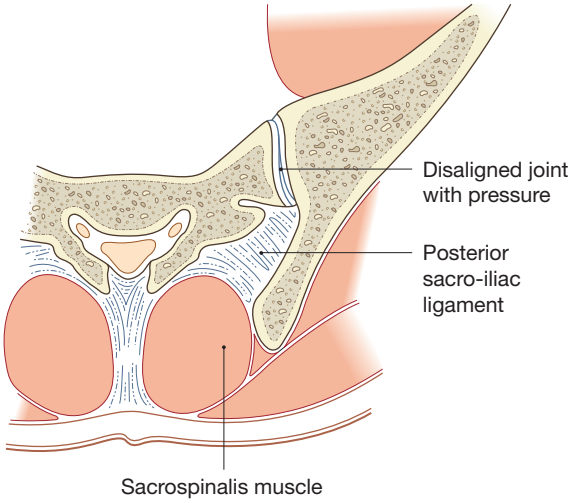

Fig. 2.3.1

Drugs

- Lidocaine (lignocaine) 1%, 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 25 mg (or its equivalent)
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 2.3.3*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The sacral area is prepared antiseptically.
- An AP image is obtained, centered over the joint to be injected.
- Two joint lines are observed. The posterior joint line is located more medial in a direct AP view (*Fig. 2.3.4a*). The image intensifier (positioned above the patient) is rotated toward the opposite side until the two joint lines

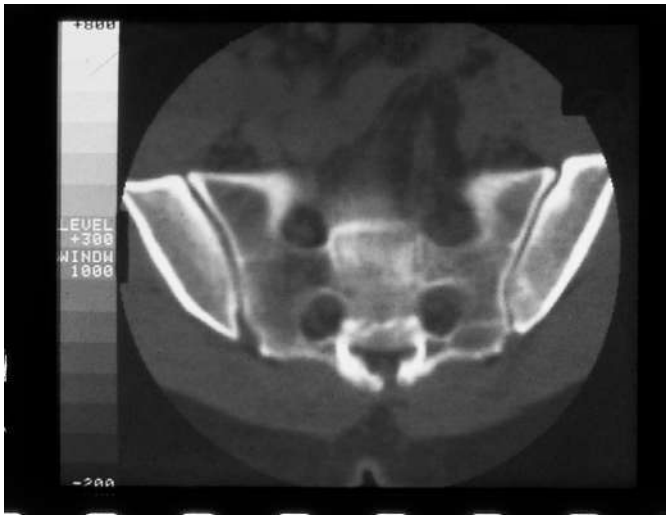

Fig. 2.3.2

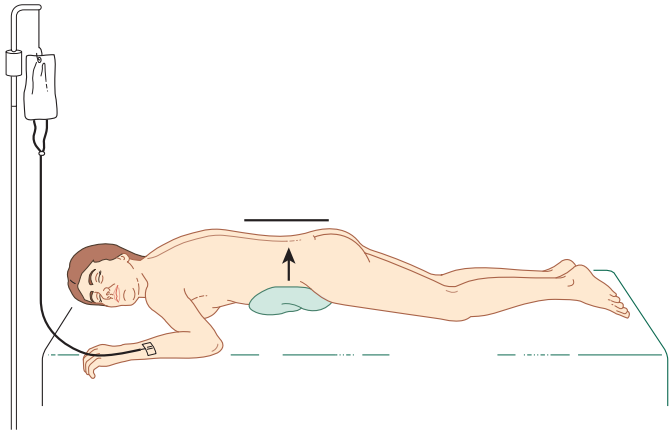

Fig. 2.3.3

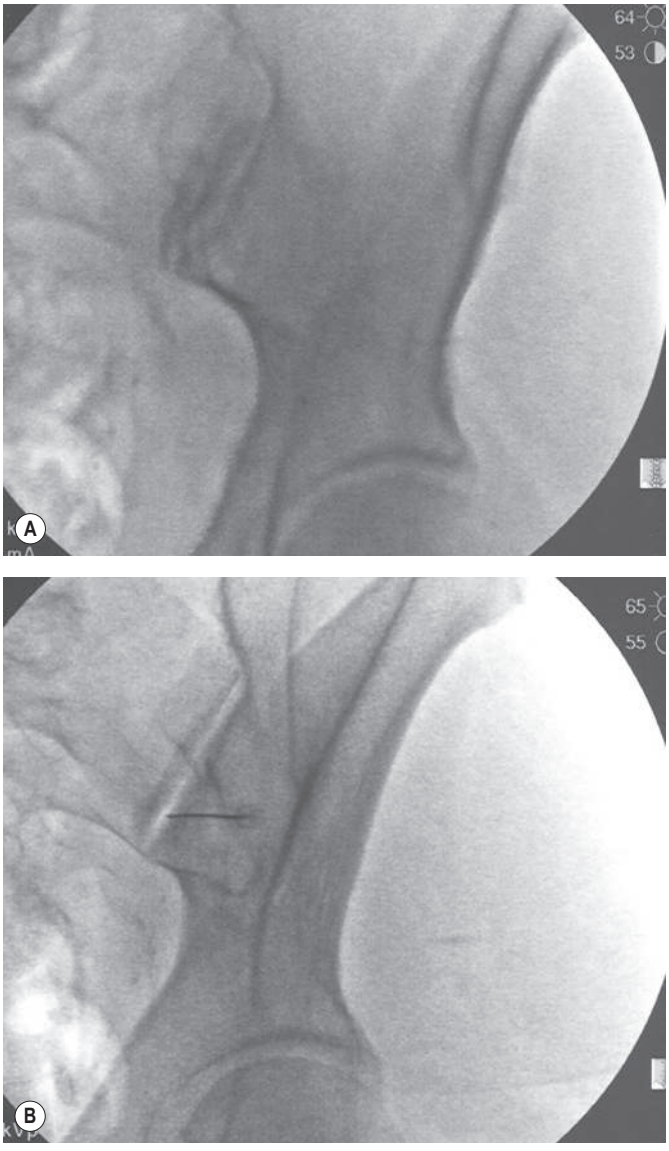

Fig. 2.3.4

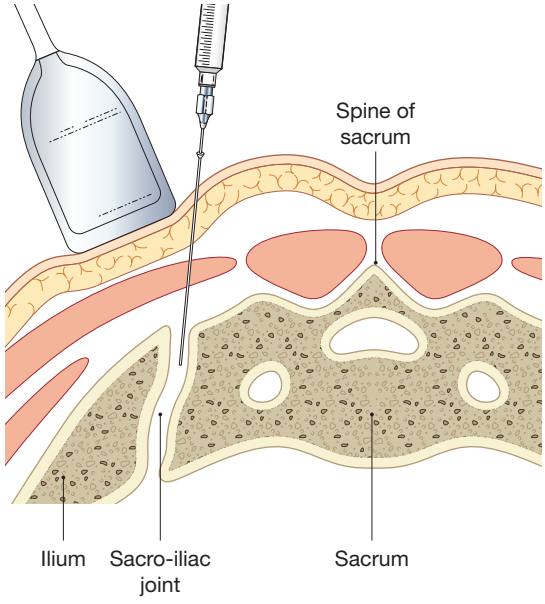

Fig. 2.3.5

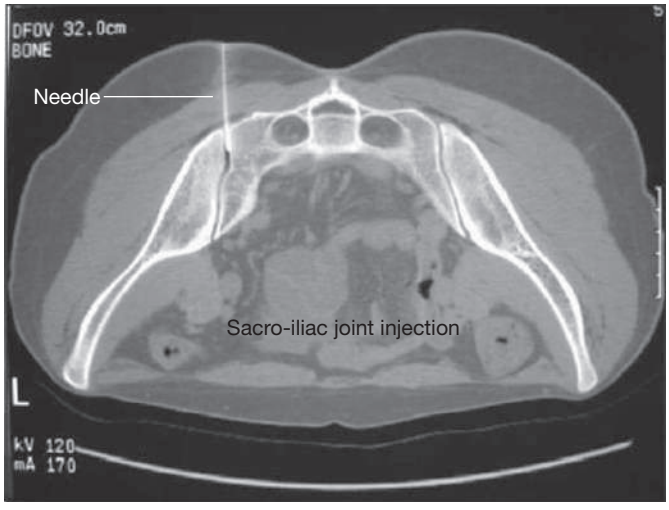

Fig. 2.3.6

are superimposed (usually about 10–20°) (*Fig. 2.3.4b*). The skin is marked and a skin wheal is raised. The area is infiltrated with lidocaine (lignocaine) 1% **over the joint line 1 cm above the most caudal point of the joint**.

- A 22 or 25 G 3½ in spinal needle is advanced no more than 1 cm into the joint. Some resistance is usually felt as the needle contacts the joint.
- A lateral view is then obtained. The needle should traverse no more than half the distance across the sacrum, and should never be advanced beyond the anterior cortex.
- Contrast dye, 0.5 to 1 ml, may be injected to ensure intra-articular spread. Intravascular injection is best detected during "live" fluoroscopy injection. In the AP view, dye should be seen within the joint space. Some extravasation outside the joint is common. If extensive, the needle should be repositioned.
- 1–2 ml lidocaine 1% is injected alone for diagnostic purposes. Reproduction of the patient's pain during needle positioning and injection as well as pain relief following the block will help confirm the sacro-iliac joint as the pain generator (avoid sedation with opioids for diagnostic procedures).
- Corticosteroid, e.g. triamcinolone diacetate 25 mg, plus 1–2 ml 1% lidocaine may be injected for therapeutic effect.
- Ultrasound may also be used to identify the joint (*Fig. 2.3.5*).
- CT scan may also be used to identify the joint but is not usually necessary (*Fig. 2.3.6*).

Confirmation of a successful injection

- Reproduction of pain during injection and relief of pain following injection confirms correct placement.
- Radiologic assessment of the X-ray image after injection of contrast medium may demonstrate tears in the joint capsule.

Tips

- While the joint may be easily entered, injection can be difficult where the joint is heavily invested with connective tissue and ligaments. This is especially true in elderly patients, where the joint is rigid and the joint space cannot expand to accommodate a volume of liquid. In such cases it may be possible to inject only as the needle is being removed from the joint.
- Sometimes, injection into the deep sacro-iliac ligaments around the joint may be helpful for pain relief. Introduction of a spinal needle just above the midline of the upper sacrum and advanced at 45° to the skin, under the rim of the ilium and in the direction of the joint, will access these ligaments. Lidocaine (lignocaine) 1% 4 ml plus triamcinolone diacetate 25 mg may then be injected.

Potential problems

- Discomfort on injection.
- Epidural injection.
- Sacral nerve root blockade.
- Subperiosteal injection (painful in the awake patient).

EPIDURAL INJECTION 3

Intervertebral disc disease may produce inflammation of spinal nerve roots, which may be the cause of radicular pain. The L5 and S1 nerve roots are most commonly affected, probably because they exit the bony canal through a narrow lateral bony recess, therefore increasing the likelihood of nerve compression and irritation. Lumbo-sacral radiculopathy consists of low-back pain that radiates a varying distance into the lower extremity, and which may be associated with motor and sensory loss consistent with damage to the affected nerve root. If bowel and bladder symptoms of dysfunction are present, large midline disc protrusion is suspected and prompt surgical intervention is indicated. Otherwise, if severe pain exists after treatment with immobilization and mild analgesics, epidural steroid injection may be carried out. Similarly, pain of thoracic or cervical disc origin may respond to epidural steroid injection.

Triamcinolone diacetate is the most commonly administered preparation and injection should be carried out as close to the affected nerve root as possible. Injection of a small amount of local anesthetic with the steroid will help to confirm drug placement and provide analgesia. In patients with S1 pathology the drug may not spread to the affected nerve root using the lumbar approach and the caudal approach to the epidural space may be required. Cervical epidural injection accesses the cervical spinal nerve roots, while in the thoracic region a paramedian approach is usually more successful.

3.1 LUMBAR EPIDURAL BLOCK

Anatomy

Structures encountered when inserting an epidural needle include skin, subcutaneous tissue, supraspinous ligament, interspinous ligament, ligamentum flavum (5–6 mm thick in the midline of the lumbar region, 3–5 mm thick in the midline of the thoracic region), prior to reaching the epidural space itself (**Fig. 3.1.1**). Beyond this space lies the dura mater, the arachnoid mater and intrathecal space containing the cerebrospinal fluid. The spinal cord usually ends at the L2 level (**Fig. 3.1.2**).

One should expect a distance of 3.5–6 cm from skin to the epidural space using a midline approach. In the lumbar region the spinous processes are generally perpendicular to the vertebral bodies (**Fig. 3.1.3**). In the thoracic region the spinous processes lie at an angle of 30–45° to the thoracic vertebral body, thus making midline epidural injection a little more difficult, and sometimes necessitating a paravertebral approach. Other relevant anatomy of the vertebral bodies is illustrated in **Fig. 3.1.4**.

Equipment

- 2 ml and 10 ml syringes
- 18 G, 20 G, and 25 G needles
- ECG, BP, and SpO_2 monitors
- 18 G epidural set (*Fig. 3.1.5*)
- Filter aspiration needle
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy (optional)
- Ultrasound (optional)

Drugs

- Lidocaine (lignocaine) 1%, 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 50 mg (or its equivalent)
- Saline (NaCl) 10 ml
- Resuscitation drugs (*see Appendix 3*)

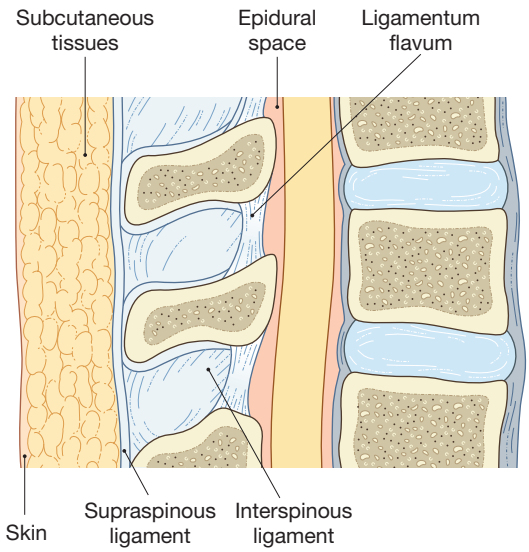

Fig. 3.1.1

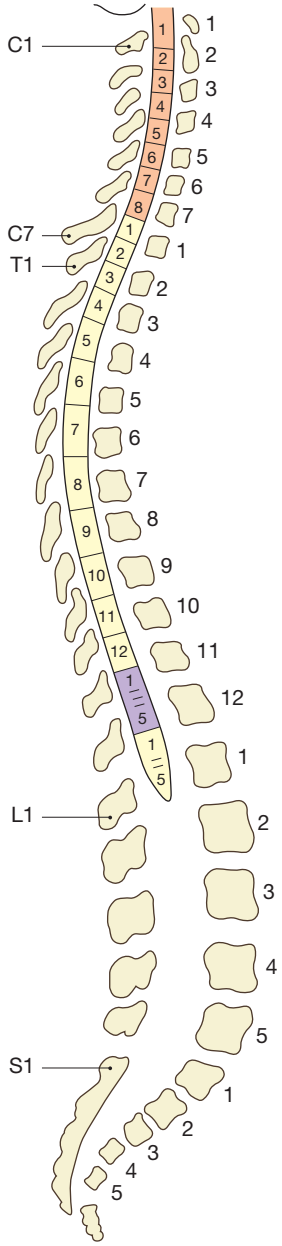

Fig. 3.1.2

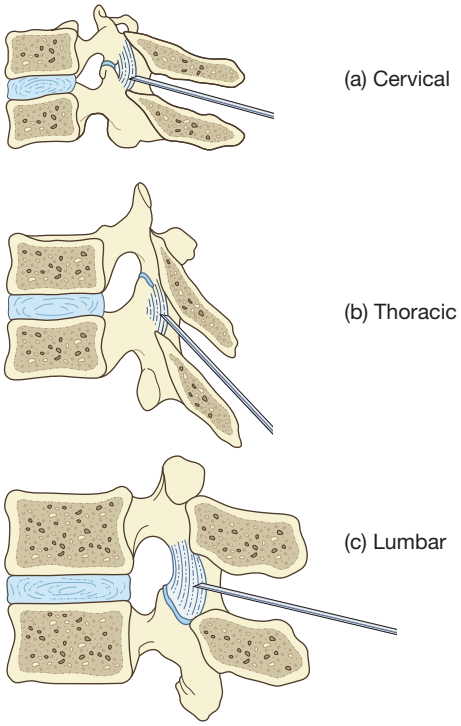

Fig. 3.1.3

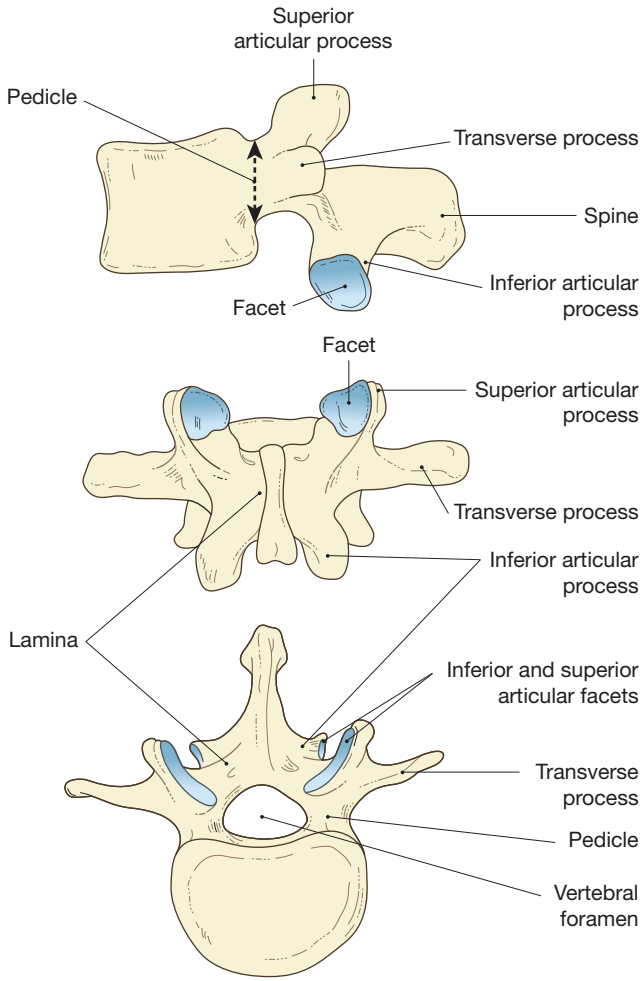

Fig. 3.1.4

Position of patient

- Lateral, usually lying on the side of the radiculopathy.
- Shoulders and buttocks parallel to the edge of the bed, perpendicular to the floor, with spine flexed.

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- Lidocaine (lignocaine) 1% 2 ml is drawn up into three 2 ml syringes.
- Lidocaine (lignocaine) 1% 2 ml, plus corticosteroid, e.g. triamcinolone diacetate 50 mg is drawn up into the 10 ml syringe.
- NaCl 10 ml is drawn up into the 10 ml loss-of-resistance syringe.
- The iliac crest is palpated and the intercrestal line (this corresponds with the inferior aspect of the spinous process of L4 or may lie in the L4–5 interspace) is identified (*Fig. 3.1.6*).
- The spinous processes are palpated, and the level requiring injection is identified.
- This may be confirmed by fluoroscopy or ultrasound.

Fig. 3.1.5

MIDLINE APPROACH FOR THE RIGHT-HANDED OPERATOR

With the left hand

- The fore- and middle fingers are placed each side of the interspace.

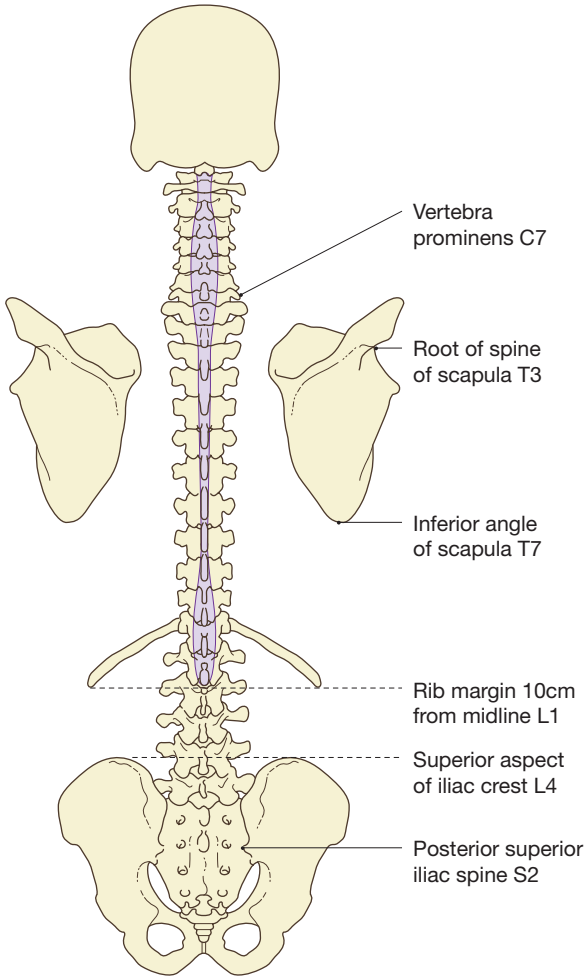

Fig. 3.1.6

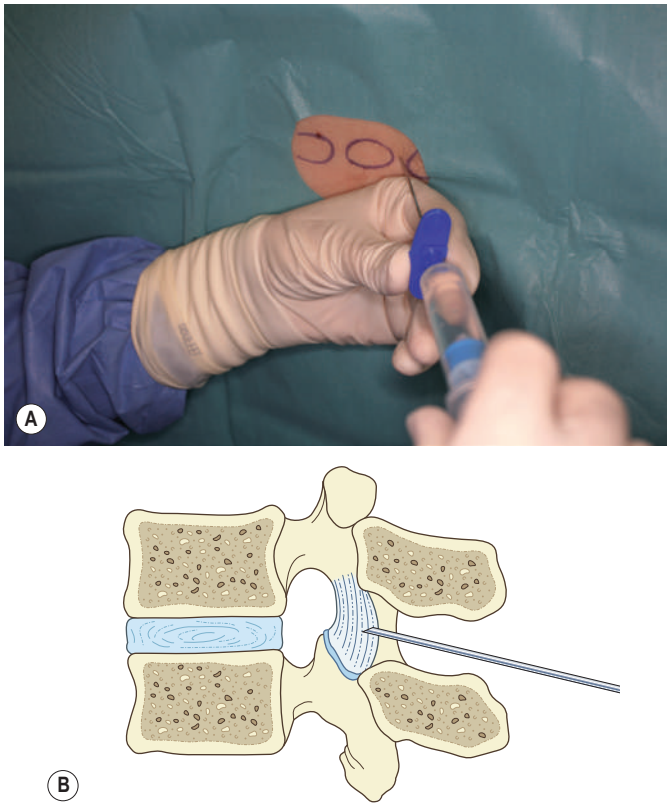

Fig. 3.1.7

- These fingers are kept in place until the epidural needle is gripped by the interspinous ligament.

With the right hand

- The interspinous ligament is infiltrated with lidocaine (lignocaine) 1% 2 ml.
- **The epidural needle is inserted, bevel facing the side of the radiculopathy, between the fore- and middle fingers of the left hand in a direction 60° cephalad, perpendicular to the spine, parallel to the floor, until it is gripped by the interspinous ligament** (*Fig. 3.1.7 a,b*). In the case of thoracic epidural injection (midline approach), the point of entry of the needle should be as close as possible to the caudal end of the interspinous space and the needle directed 30–45° cephalad to enter between the spinous processes.
- The hub of the needle is gripped with the fore- and middle fingers of the left hand and this hand is steadied by leaning the wrist against the patient's back.
- The stylet is removed and the loss-of-resistance syringe is applied.
- The needle is slowly and carefully advanced, while constant pressure is applied to the plunger, the left hand aiding the advance, while at the same time applying a brake if required (*Fig. 3.1.8 a,b*).
- At the point at which the needle enters the ligamentum flavum, absolute resistance to injection is experienced.
- At this point the needle is advanced very slowly until a sudden loss of resistance to the pressure on the plunger is experienced, the point at which the epidural space is entered.
- After negative aspiration for blood or cerebrospinal fluid (CSF), lidocaine (lignocaine) 1% 3 ml is injected.
- After 5 minutes the patient is questioned about any changes in sensation or power, and any changes in heart rate or blood pressure are noted.
- If the injection is for diagnostic purposes only, the needle may be removed at this point.
- If therapeutic effect is required, lidocaine (lignocaine) 1% 2 ml, plus corticosteroid if indicated, e.g. triamcinolone diacetate 50 mg (or its equivalent), may be injected. Alternatively, a catheter may be inserted through the needle if indicated.
- The needle is flushed with NaCl 1 ml and removed.
- The patient is allowed to lie in the lateral position, on the side of the pain.

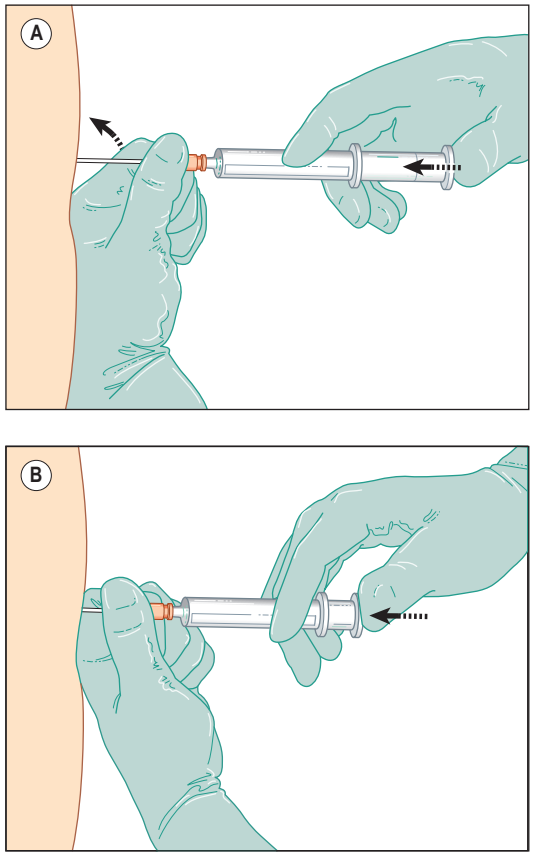

Fig. 3.1.8

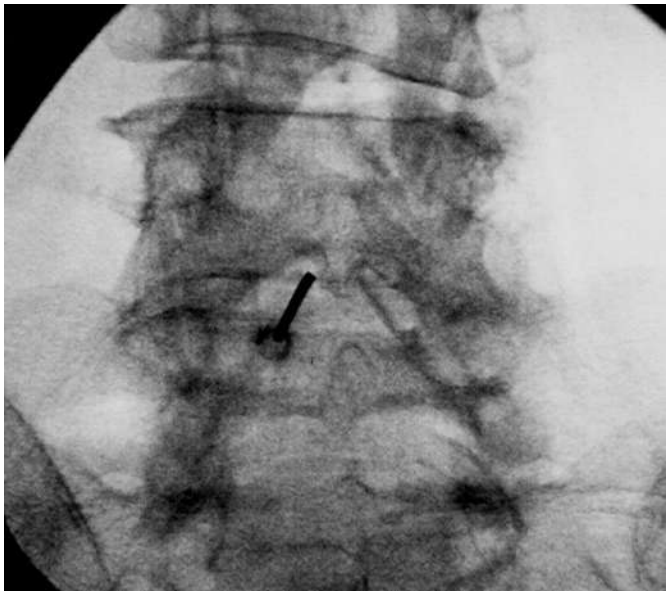

Fig. 3.1.9

Fig. 3.1.10

- Monitors should be left attached and i.v. access left in situ for at least 30 minutes.
- The patient is advised to contact the hospital should the anesthesia remain after several hours.

FLUOROSCOPIC GUIDED LUMBAR EPIDURAL INJECTION

- Position the patient prone with a pillow under the lower abdomen to increase lumbar flexion.
- The interlaminar space for the desired segmental level is identified fluoroscopically using a straight AP view. Angling the fluoroscope slightly cephalad may open the space if it appears very narrow.
- Prepare skin with antiseptic and sterile drape.
- Raise a local anesthetic skin wheal just below the interlaminar space, about 0.5 cm from the midline toward the symptomatic side.
- Advance the Tuohy needle, angling slightly toward the midline until resistance of the ligamentum flavum is encountered, keeping the trajectory just lateral to the midline to avoid contacting the spinous process. Repeat imaging periodically to ensure that the needle is approaching the space, not the lamina or spinous process (see *Fig. 3.1.9*). Advance the needle through the ligamentum flavum using a loss of resistance technique with air or saline. Once loss of resistance is achieved, obtain a lateral view to ensure the needle is barely into the bony spinal canal.
- Attach a low volume extension set to the needle, aspirate to ensure there is no blood return, and inject 0.5–1 ml of non-ionic contrast medium during live fluoroscopy. Dye should be seen spreading within the bony canal (see *Fig. 3.1.10*). Prior to injecting local anesthetic or steroid, obtain an AP view to reconfirm epidural spread.

- Aspirate again, then inject local anesthetic (1–2 ml) and 50 mg triamcinolone diacetate or equivalent.
- Loss of resistance is occasionally encountered with the needle superficial to the epidural space. Dye will be seen dorsal to the epidural space on the lateral view, and spread lateral to the spinal canal will be seen on the AP view. The needle can then be advanced through the ligamentum flavum, again using loss of resistance, followed by dye confirmation.

Confirmation of a successful block

- Relief of pain.
- Anesthesia in the distribution of the blocked nerves.
- For lumbar epidural steroid injection, improvement in straight-leg raising may be evident.

Tips

- Air may be used instead of NaCl to determine loss of resistance. If this technique is used it is advisable to avoid constant pressure on the plunger, as the air is compressible; instead it should be bounced intermittently with the thumb to test for resistance and loss of resistance.
- Advocates claim identification of CSF is easier with this technique.
- Advocates of the use of NaCl point out that absolute resistance to pressure identifies the ligamentum flavum, and that by applying constant pressure to the plunger one can identify loss of resistance earlier, thereby more easily avoiding the possibility of dural tap.
- An epidural catheter may be inserted through the needle and the needle removed, taking care not to withdraw the catheter when removing the needle. However, a test dose of lidocaine (lignocaine) 1% 4 ml with/without epinephrine (adrenaline) 1:200 000 is given after insertion, before any injection through the catheter is carried out.
- Ultrasound may guide the insertion of the needle as spinous proccesses are easily visible on ultrasound (optional) (*Fig. 3.1.11*).

Fig. 3.1.11

- Injection of radio-opaque dye under direct fluoroscopy can confirm epidural placement.
- Insertion of a radio-opaque epidural catheter may be carried out also under fluoroscopy.

Potential problems

IMMEDIATE

- Failure to locate epidural space (sitting position may be successful).
- Intravascular injection (test dose important); addition of epinephrine (adrenaline) to test dose may help identification of intravascular injection.
- Intrathecal injection (test dose important).
- Hypotension due to sympathetic blockade (give i.v. fluid; consider ephedrine).
- Headache (possible dural puncture).
- Allergic reaction.

LATER

- Infection (epidural abscess; bacterial meningitis).
- Aseptic meningitis; usually a result of intrathecal injection (test dose important).
- Cushingoid symptoms; usually as a result of repeated steroid injections.

3.2 THORACIC EPIDURAL BLOCK

Anatomy

In the thoracic region, the spinous processes lie at an angle of 30–45° to the thoracic vertebral body (**Fig. 3.2.1 a,b**) thus making midline epidural injection a little more difficult, and sometimes necessitating a paramedian approach (**Fig. 3.2.2 a–d**).

Equipment

- 2 ml and 10 ml syringes
- 18 G, 20 G, and 25 G needles
- ECG, BP, and SpO_2 monitors
- 18 G epidural set
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy or ultrasound (optional)

Drugs

- Lidocaine (lignocaine) 1%, 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 50 mg (or its equivalent)
- Saline (NaCl) 10 ml
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Lateral, usually lying on the side of the radiculopathy.
- Shoulders and buttocks parallel to the edge of the bed, perpendicular to the floor, with spine flexed.

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- Lidocaine (lignocaine) 1% 2 ml is drawn up into three 2 ml syringes.
- Lidocaine (lignocaine) 1% 2 ml, plus corticosteroid, e.g. triamcinolone diacetate 50 mg, is drawn up into the 10 ml syringe.
- NaCl 10 ml is drawn up into the 10 ml loss-of-resistance syringe.
- The spinous processes are palpated, and the level requiring injection is identified.
- This may be confirmed by fluoroscopy or ultrasound.

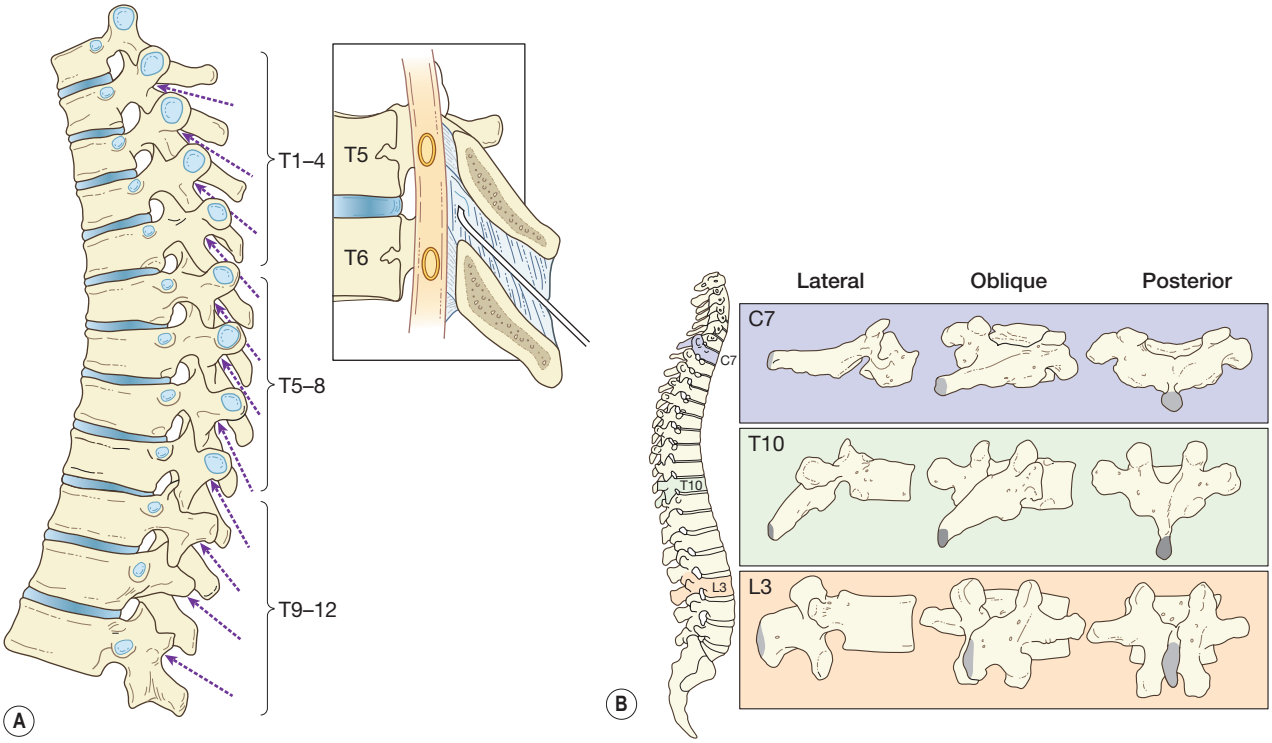

Fig. 3.2.1

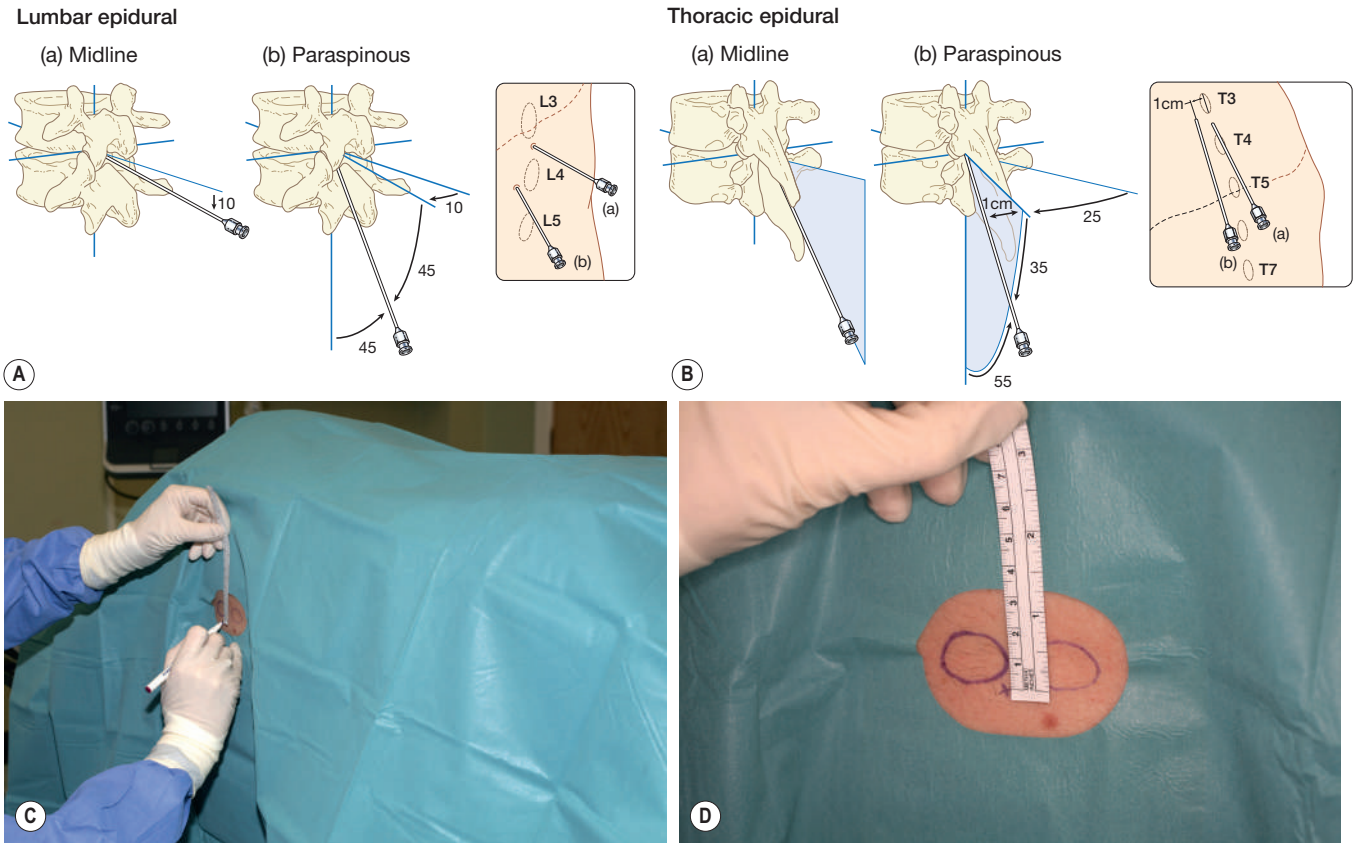

Fig. 3.2.2

Paramedian approach

FOR THE RIGHT-HANDED OPERATOR

With the left hand

- The fore- and middle fingers are placed each side of the interspace.
- These fingers are kept in place until the epidural needle is gripped by the interspinous ligament.

With the right hand

- The interspinous ligament is infiltrated with lidocaine (lignocaine) 1% 2 ml.
- **The insertion point of the epidural needle in the paravertebral approach lies 1 cm lateral to the midline, at the lower border of the spinous process** (*Fig. 3.2.3 a,b*). The epidural needle is inserted, bevel facing the side of radiculopathy, between the fore- and middle fingers of the left hand, perpendicular to the spine, parallel to the floor, until it is gripped by the interspinous ligament.
- **The direction of the needle is 130° cephalad and 15° medial to the midline.** Care must be taken as the ligamentum flavum is not as thick laterally, and may not be identified as easily. Therefore, it is usually easiest to first identify the lamina and **walk the needle off the lamina in a cephalad direction** until the needle enters the ligamentum flavum. At that point the loss-of-resistance technique may be performed.
- The hub of the needle is gripped with the fore- and middle fingers of the left hand and this hand is steadied by leaning the wrist against the patient's back.
- The stylet is removed and the loss-of-resistance syringe is applied.
- The needle is slowly and carefully advanced until the osseous endpoint of the lamina is encountered.
- It is then walked off the lamina in the cephalad direction until it enters the ligamentum flavum.
- At the point at which the needle enters the ligamentum flavum, absolute resistance to injection is experienced.
- It is then carefully advanced further while constant pressure is applied to the plunger, the left hand aiding the advance, while at the same time applying a brake if required (*Fig. 3.2.4*, viewed from above).
- The needle is advanced very slowly until a sudden loss of resistance to the pressure on the plunger is experienced, the point at which the epidural space is entered.
- After negative aspiration for blood or cerebrospinal fluid, lidocaine (lignocaine) 1% 2 ml is injected.

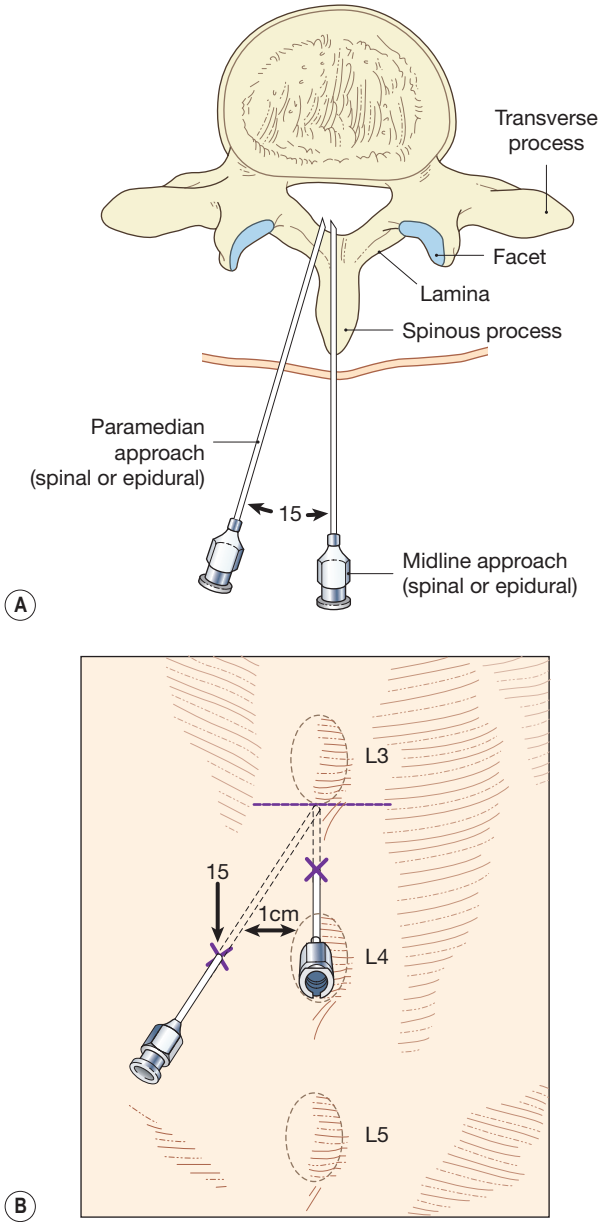

Fig. 3.2.3

- After 5 minutes the patient is questioned about changes in sensation or power, and any changes in heart rate or blood pressure are noted.
- If the injection is for diagnostic purposes only, the needle may be removed at this point.
- If therapeutic effect is required, lidocaine (lignocaine) 1% 2 ml plus corticosteroid, e.g. triamcinolone diacetate 50 mg, may be injected.
- The patient is allowed to lie in the lateral position, on the side of the pain.
- Monitors should be left attached and i.v. access should remain in situ for at least 30 minutes.

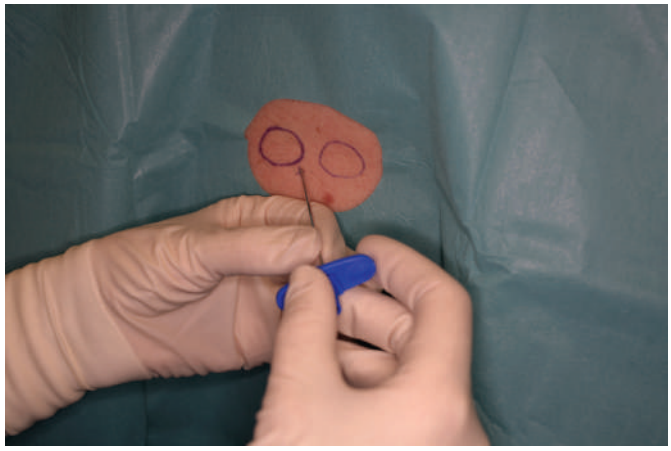

Fig. 3.2.4

- The patient is advised to contact the hospital should anesthesia remain after several hours.

Confirmation of a successful block

- Relief of pain.
- Anesthesia in the distribution of blocked nerves.
- For lumbar epidural steroid injection, improvement in straight-leg raising may be evident.

Tips

- Air may be used instead of NaCl to determine loss of resistance. If this technique is used it is advisable to avoid constant pressure on the plunger, as the air is compressible; instead the plunger should be bounced intermittently with the thumb to test for resistance and loss of resistance.
- Advocates claim identification of CSF is easier with this technique.
- Advocates of the use of NaCl point out that absolute resistance to pressure identifies the ligamentum flavum, and that by applying constant pressure to the plunger one can identify loss of resistance more immediately, thereby avoiding the possibility of dural tap more easily.
- An epidural catheter may be inserted through the needle and the needle removed, taking care not to withdraw the catheter when removing the needle. However, a test dose of lidocaine (lignocaine) 1% 4 ml with epinephrine (adrenaline) 1:200 000 is given after insertion, before any injection through the catheter is carried out.
- Identification of the insertion point may be aided by ultrasound (*Fig. 3.2.5*).
- Injection of radio-opaque dye under direct fluoroscopy can confirm epidural placement.
- Insertion of a radio-opaque epidural catheter may be carried out also under fluoroscopy.

Fig. 3.2.5

Potential problems

IMMEDIATE

- Failure to locate epidural space (sitting position may be successful).
- Intravascular injection (test dose important); addition of epinephrine (adrenaline) to test dose may help identification of intravascular injection.
- Intrathecal injection (test dose important).
- Hypotension due to sympathetic blockade (give i.v. fluid, consider ephedrine).
- Headache (possible dural puncture).
- Allergic reaction.
- Spinal cord injury may occur if the epidural space is not recognized. Deep sedation should be avoided during needle insertion and drug injection.

LATER

- Infection (epidural abscess, bacterial meningitis).
- Aseptic meningitis, usually the result of intrathecal injection (test dose important).
- Cushingoid symptoms (usually as a result of repeated injections).
- Epidural hematoma. This complication should be suspected when sensory or motor function loss occurs minutes to hours after the procedure. Immediate diagnostic imaging (CT or MRI) is essential. Prompt surgical decompression may be required.

3.3 CERVICAL EPIDURAL BLOCK

Anatomy

In the cervical region, the spinous processes are almost perpendicular to the vertebral bodies, especially in the lower part. They also widen and become bifid. As a result, insertion of the needle is often easy. However, it must be remembered that the epidural space is relatively narrow in this area (2–4 mm), and that the spinal cord lies very close to it (**Fig. 3.3.1 a,b**). Most workers prefer to use the "hanging drop" technique when accessing the cervical epidural space, as there exists a significant negative pressure in the cervical region in the sitting position.

Equipment

- 2 ml, 5 ml, and 10 ml syringes
- 18 G, 20 G, and 25 G needles
- ECG, BP, and SpO_2 monitors
- 18 G epidural set
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy or ultrasound (optional)

Drugs

- Lidocaine (lignocaine) 1%, 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 50 mg (or its equivalent)
- Saline (NaCl) 10 ml
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Sitting.
- Head flexed forward.

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- Lidocaine (lignocaine) 1% 2 ml is drawn up into two 2 ml syringes.
- Lidocaine (lignocaine) 1% 1 ml plus corticosteroid, e.g. triamcinolone diacetate 50 mg, is drawn up into a 5 ml syringe.
- NaCl 10 ml is drawn up into a 10 ml syringe.
- The patient is allowed to sit up straight for a moment, and the spinous process of T3, which lies opposite the root of the spine of the scapula, is identified and marked. The prominent spinous process of C7 (vertebra prominens) is identified (*Fig. 3.3.2*) and marked (*Fig. 3.3.3 a,b*). Ultrasound can be used to guide needle placement (*Fig. 3.3.3 c,d*). The interspace to be used for epidural injection is also marked.

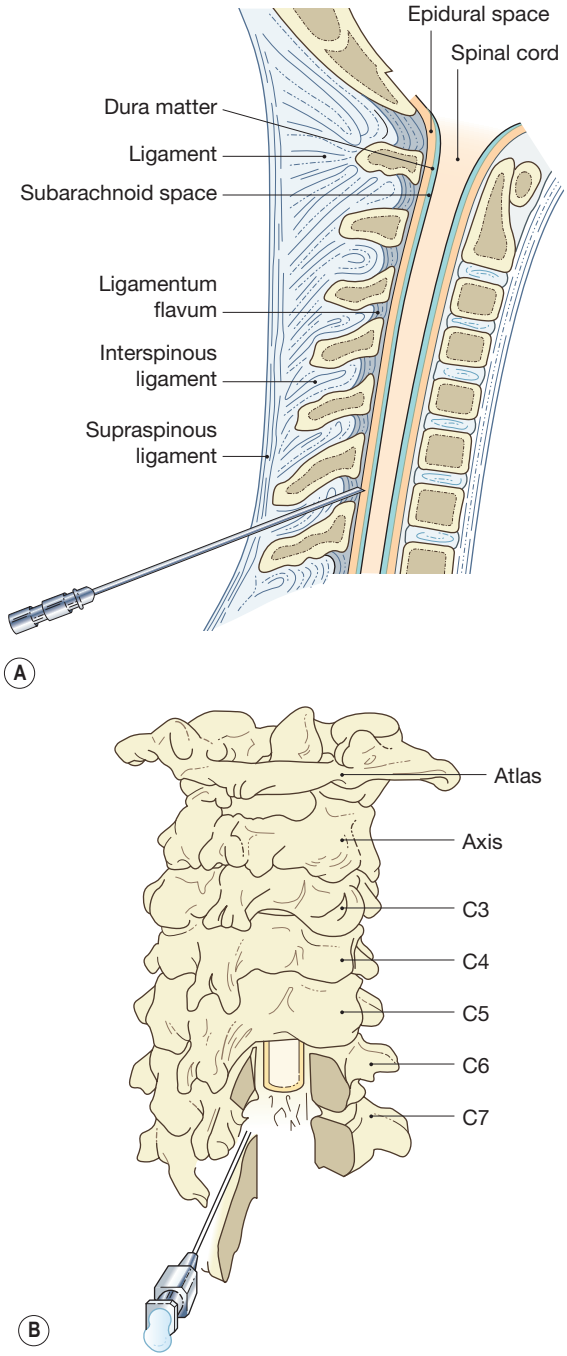

Fig. 3.3.1

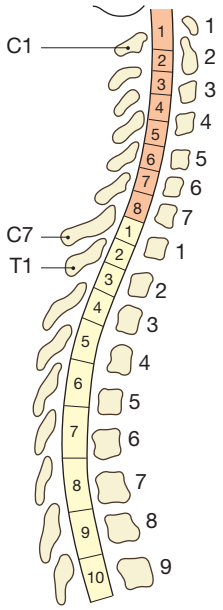

Fig. 3.3.2

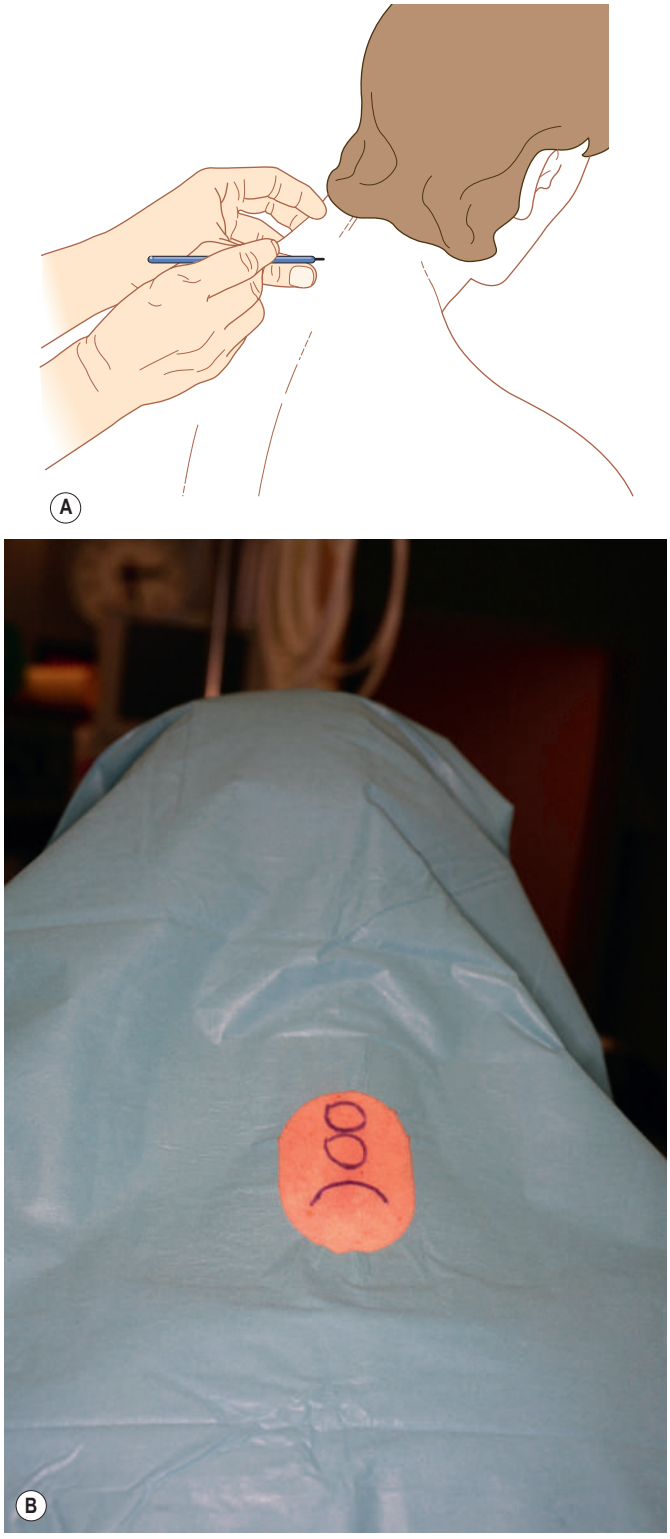

Fig. 3.3.3

FOR THE RIGHT-HANDED OPERATOR

With the left hand

- The fore- and middle fingers are placed each side of the interspace.
- These fingers are kept in place until the epidural needle is gripped by the interspinous ligament.

With the right hand

- The interspinous ligament is infiltrated with lidocaine (lignocaine) 1% 2 ml.
- **The epidural needle is inserted, bevel facing caudad, between the fore- and middle fingers of the left hand in a direction 60° cephalad, until it is gripped firmly by the interspinous ligament.**
- The hub of the needle is gripped with the fore- and middle fingers of the left hand and this hand is steadied by leaning the wrist against the patient's spine.
- The stylet is removed.
- The hub of the epidural needle is filled with saline (*Fig. 3.3.4*) such that a "hanging drop" appears (*Fig. 3.3.5*).

With both hands

- The wings of the epidural needle are gripped with each hand, steadying the hands by resting the wrists against the posterior thoracic wall (*Fig. 3.3.6*).
- The needle is slowly and carefully advanced with both hands.
- It is prudent periodically to confirm high resistance of the needle in the ligament by testing with an air-filled syringe, then replace the stylet to make sure there is no tissue blocking the needle before resuming the "hanging drop" technique.

Fig. 3.3.3, cont'd

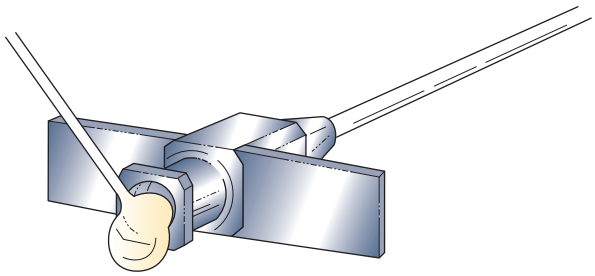

Fig. 3.3.4

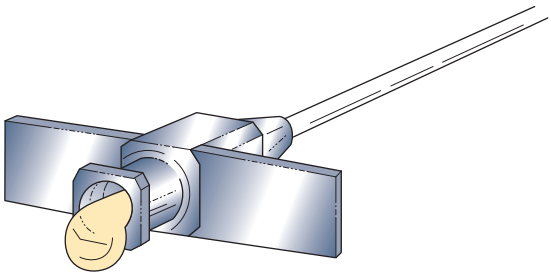

Fig. 3.3.5

- The "hanging drop" at the hub of the needle is watched closely, and the patient is asked periodically about the presence of paresthesia.
- At the point at which the epidural needle enters the epidural space, the drop should appear to be sucked into the needle (*Figs 3.3.7*, *3.3.8*).
- After negative aspiration for blood or CSF, lidocaine (lignocaine) 1% 2 ml is injected.
- After 1–2 minutes the patient is questioned about changes in sensation or power, and any changes in heart rate or blood pressure are noted.
- If the injection is for diagnostic purposes only, the needle may be removed at this point.
- If therapeutic effect is required, lidocaine (lignocaine) 1% 1 ml plus corticosteroid, e.g. triamcinolone diacetate 50 mg, may be injected.
- The patient is allowed to lie in the lateral position, on the side of the pain.
- Monitors should be left attached and i.v. access kept in situ for at least 30 minutes.

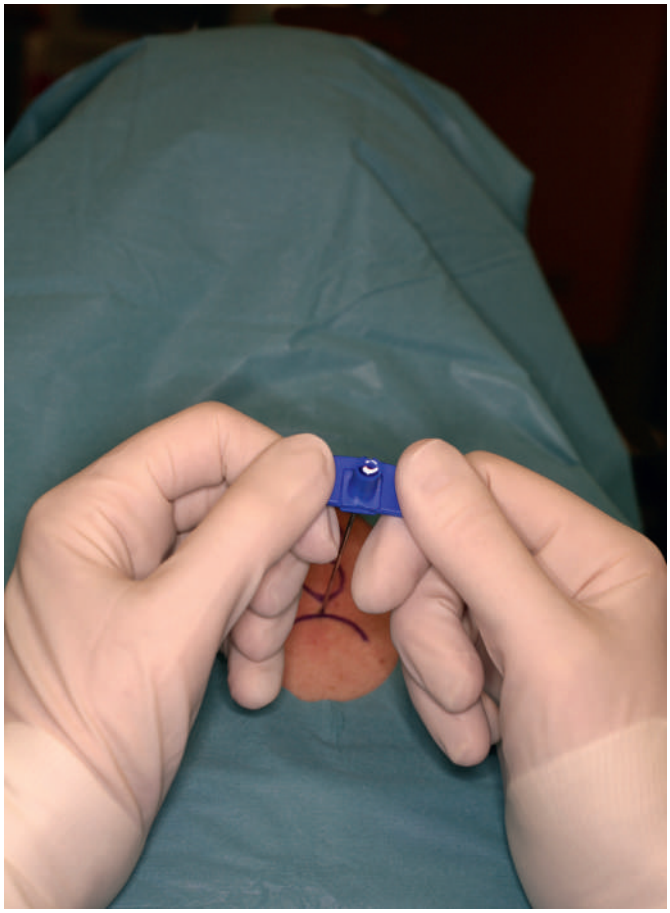

Fig. 3.3.6

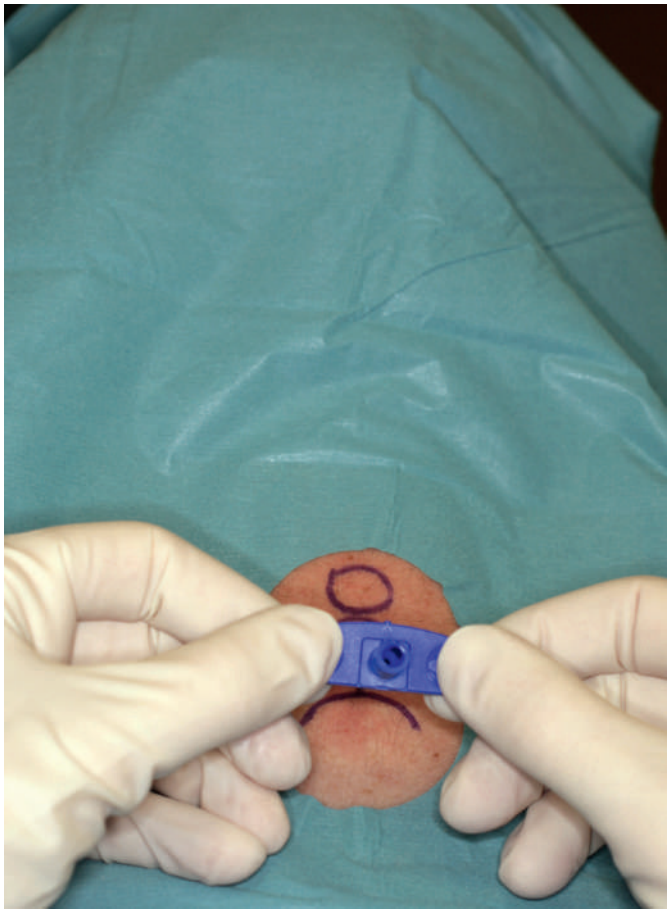

Fig. 3.3.8

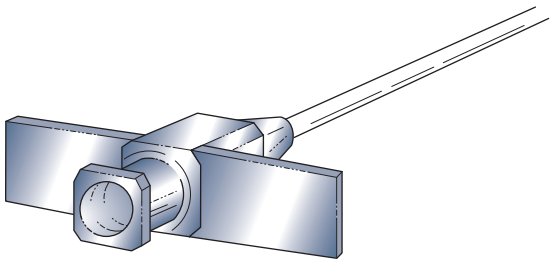

Fig. 3.3.7

- Ultrasound may aid in identifying the interspinous space as spinous processes are easily visible.

FLUOROSCOPIC GUIDED CERVICAL EPIDURAL INJECTION

- Check MRI to ensure that the spinal cord is not displaced posteriorly. If the posterior epidural space is compromised or the spinal cord is shifted posteriorly, it is safer to enter the upper thoracic epidural space and advance a radio-opaque catheter to the low cervical level.
- Position patient prone, with pillow under shoulders, neck flexed, arms at sides, shoulders as far downward as possible.
- Identify targeted interlaminar space (T1–2, C7–T1, or C6–7) in direct AP fluoroscopic view. Adjust angle upward or downward slightly to maximize view of interlaminar space.
- Mark skin over lower border of T1–2, C7–T1 (preferred) or C6–7 interlaminar space just lateral to the midline. Do not perform epidural injection above C6–7 because of absence of midline epidural fat at higher levels.
- Prepare skin with antiseptic and sterile drape.
- Provide only minimal sedation or no sedation. Instruct patient to report any pain or paresthesia during the procedure.
- Infiltrate skin and subcutaneous tissue with lidocaine (lignocaine) 1%.
- Advance the Tuohy needle through the insertion point into the interspinous ligament and check the fluoroscopy image to ensure the needle tip is directed toward the midline (*Fig. 3.3.9*).
- Begin to advance through the ligament using loss of resistance with air or saline. Check lateral view if possible (may be obscured by the shoulders). Proceed with needle advancement. When loss of resistance

Fig. 3.3.9

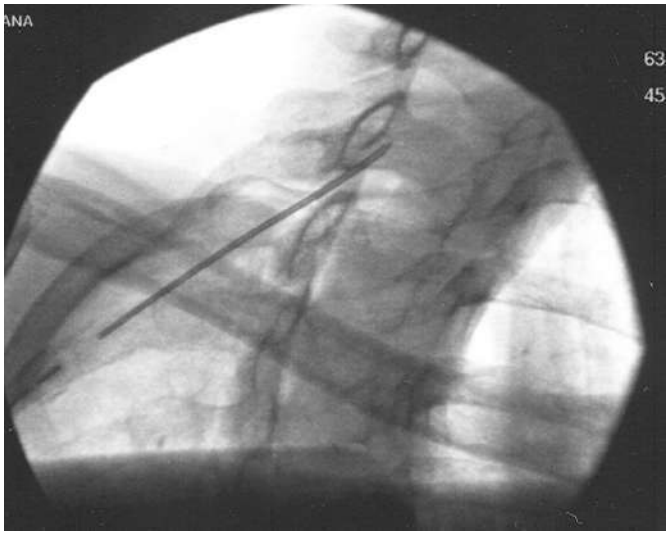

Fig. 3.3.10

occurs, recheck the lateral view. If unable to visualize the spinal canal, obtain slightly oblique view (*Fig. 3.3.10*).

- Attach a low volume extension set to the needle and inject a small volume (1 ml or less) of non-ionic contrast medium under live fluoroscopy, preferably in lateral view (AP or oblique view is used if this is not possible) (*Fig. 3.3.11*), then observe the AP view to confirm epidural dye spread.
- Aspirate to ensure there is no blood return, then inject 1–2 ml lidocaine (lignocaine) 1% followed by 25–50 mg triamcinolone diacetate or equivalent.

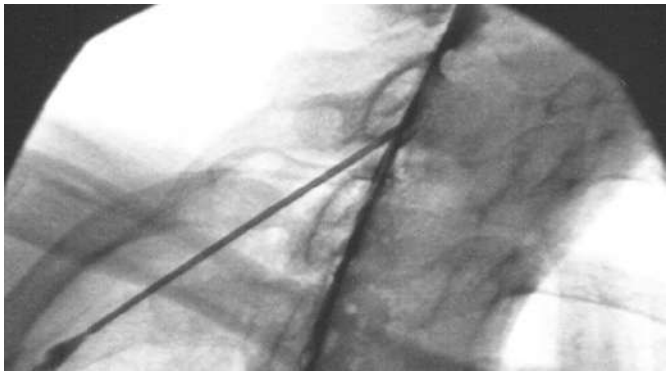

Fig. 3.3.11

Confirmation of a successful block

- Relief of pain.

Tips

- Loss-of-resistance techniques may also be used to access the cervical epidural space.
- The steroid may be given soon after the test dose, as hypotension may be a problem if the patient remains in the sitting position.

Potential problems

IMMEDIATE

- Failure to locate epidural space (lateral position with loss of resistance technique may be successful).
- Pain on injection (caution: close proximity to spinal cord).
- Intrathecal injection (test dose important).
- Hypotension ± bradycardia due to sympathetic blockade (maximum 3 ml local anesthetic administered in this technique).
- Vasovagal syncope is common in young adult patients.
- Intravascular injection; addition of epinephrine (adrenaline) 1:200 000 to the test dose may aid identification of intravascular injection.
- Headache (possible dural puncture).
- Allergic reaction.
- Spinal cord injury may occur if the epidural space is not recognized. Deep sedation should be avoided during needle insertion and drug injection.

LATER

- Infection (epidural abscess, bacterial meningitis).
- Aseptic meningitis, usually the result of intrathecal injection (test dose important).
- Cushingoid symptoms (usually as a result of repeated injections).
- Epidural hematoma. This complication should be suspected when sensory or motor function loss occurs minutes to hours after the procedure. Immediate diagnostic imaging (CT or MRI) is essential. Prompt surgical decompression may be required.

3.4 CAUDAL EPIDURAL BLOCK

Anatomy

Injection of anesthetic through the sacral hiatus allows access to the sacral epidural space or caudal anesthesia. The sacrum is roughly triangular in shape and is made up of five fused sacral vertebrae (**Fig. 3.4.1**). Its dorsal aspect is convex and there is a midline sacral canal that allows passage of the sacral nerves through four pairs of foramen, anteriorly and posteriorly. At the caudal end lies the coccyx, and at the cephalad end lies the fifth lumbar vertebra. The posterior wall of S5, and sometimes S4, is unfused. The thick fibrous sacro-coccygeal membrane or sacral hiatus covers the defect. This may be variable in size as the posterior wall of other sacral vertebrae may also be unfused. Penetration of this membrane allows access to the sacral epidural space.

Equipment

- 2 ml, 5 ml, and 20 ml syringes
- 18 G, 20 G, and 25 G needles
- 22 G, 3–5 cm short-bevel needle, with stylet
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy or ultrasound (optional)

Drugs

- Lidocaine (lignocaine) 1%, 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 50 mg (or its equivalent)
- Saline (NaCl) 10 ml
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under abdomen and/or operating table broken to allow flexion of the lumbo-sacral spine.
- Lower limbs abducted 15°, toes rotated to point towards the opposite foot (*Fig. 3.4.2*).

Needle puncture and technique (adult)

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.

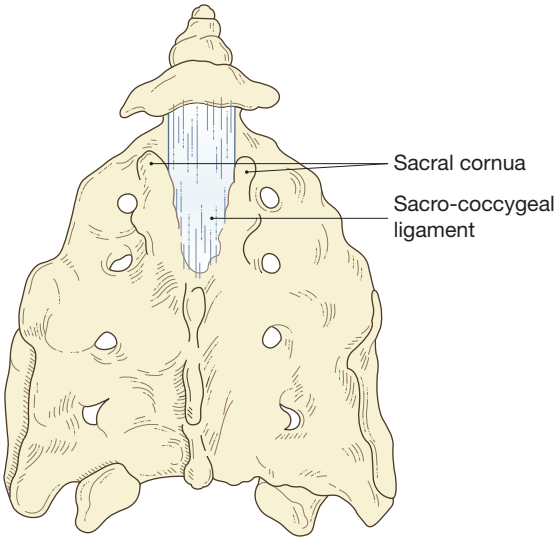

Fig. 3.4.1

Fig. 3.4.2

- Lidocaine (lignocaine) 1% 2 ml is drawn up into three 2 ml syringes.
- Lidocaine (lignocaine) 1% 15 ml plus corticosteroid, e.g. triamcinolone diacetate 50 mg, is drawn up into the 20 ml syringe.
- NaCl 10 ml is drawn up into the 10 ml syringe.

FOR THE RIGHT-HANDED OPERATOR

With the left hand

- The posterior superior iliac spines are identified.
- The sacral cornua (the unfused spinous processes of S5) are also identified and marked.
- Between the cornua lies the base of the sacral hiatus, a roughly triangular fibroelastic structure.
- The index and middle fingers of the left hand are placed on each of the sacral cornua (*Fig. 3.4.3*).
- **The insertion point lies between these two fingers.**

With the right hand

- The insertion point is infiltrated with lidocaine (lignocaine) 1% 2 ml.
- The 22 G short-bevel needle with stylet is inserted at an angle of 45° to the skin (*Fig. 3.4.4. a,b*). Ultrasound can be used to guide needle placement (*Fig. 3.4.4c*).
- As the needle passes through the fibroelastic sacral hiatus, a "pop" may be experienced, although this is not always evident in adults, and bone may be contacted.
- After passing through the sacral hiatus, the needle is withdrawn a little, and redirected to an angle to the skin of 15–20° (*Fig. 3.4.5*). This should allow further advancement of 1–2 cm, as the needle enters the long axis of the caudal epidural space.
- After negative aspiration, lidocaine (lignocaine) 1% 2 ml is injected.
- After 5 minutes the patient is questioned about changes in sensation or power of the lower limbs, and any changes in heart rate or blood pressure are noted.
- Then lidocaine (lignocaine) 1% 5–15 ml plus corticosteroid, e.g. triamcinolone diacetate 50 mg, may be injected in order to promote spread to upper sacral and lower lumbar segments (a volume of at least 10 ml should be used if the symptoms are at the level of S1 nerve root or higher) (*Fig. 3.4.6*). The needle is then

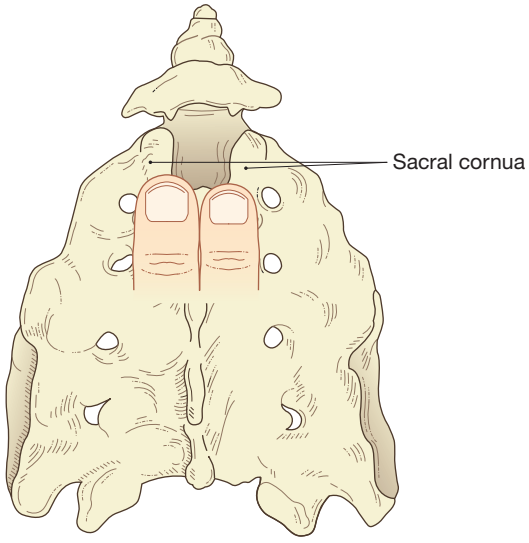

Fig. 3.4.3

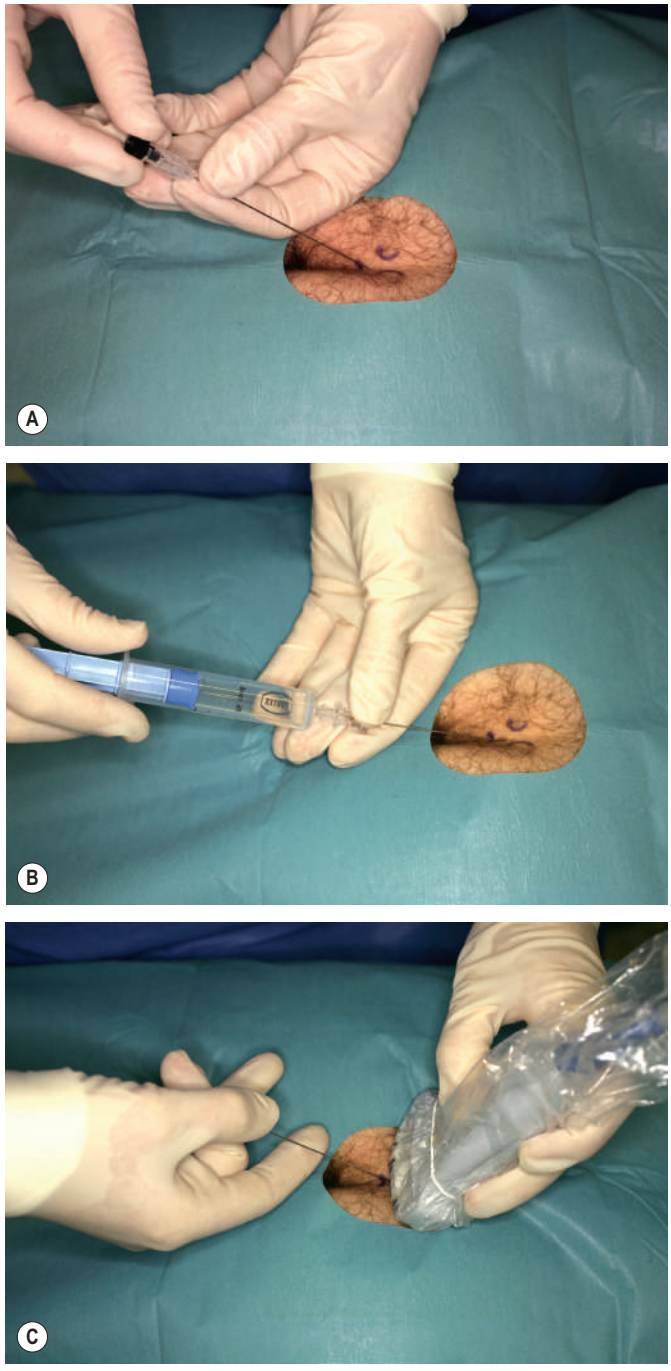

Fig. 3.4.4

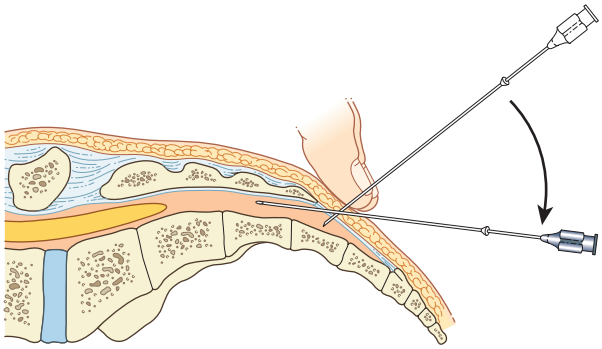

Fig. 3.4.5

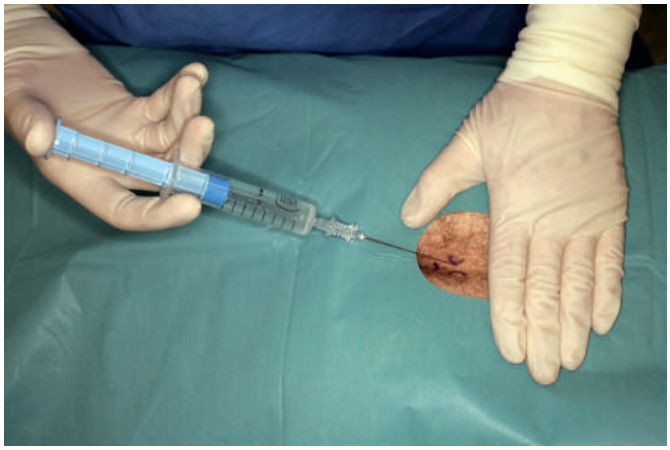

Fig. 3.4.6

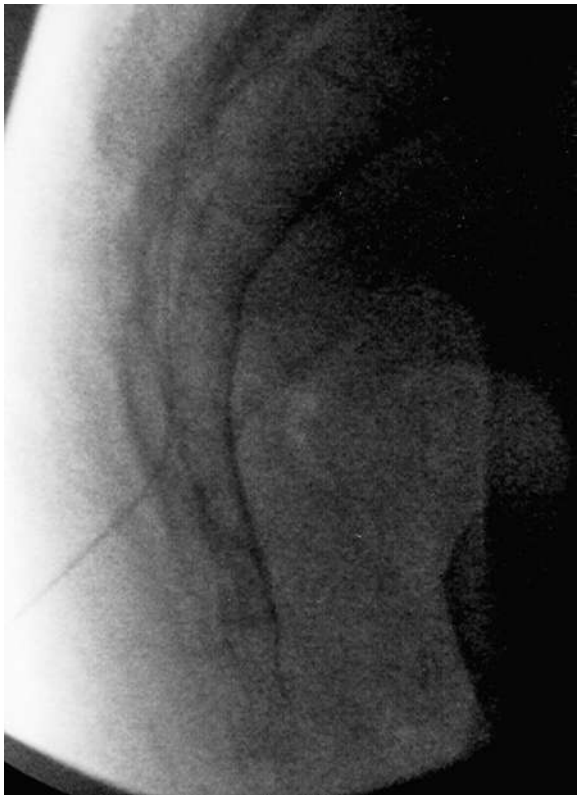

Fig. 3.4.7

removed while clearing it with lidocaine (lignocaine) 1% 2 ml.

- Monitors should be left attached and i.v. access kept in situ for at least 30 minutes.
- It is prudent to warn the patient about possible loss of sensation and or power of one or both lower limbs.

FLUOROSCOPIC GUIDED CAUDAL EPIDURAL INJECTION

- This approach may be used for treating L5 or S1 radiculopathy. It is a reasonable alternative to the interlaminar approach when surgery has disrupted the posterior spinal anatomy.
- Position the patient prone.
- Locate the sacral hiatus using the sacral cornua as landmarks.
- Prepare skin with antiseptic and sterile drape.
- Place the tip of a sterile blunt instrument over the sacral hiatus and obtain a lateral fluoroscopic view of the sacrum.
- Raise a skin wheal with 1% lidocaine (lignocaine) just below the sacral hiatus and infiltrate with lidocaine down to the sacral hiatus with a small gauge needle.
- Advance a Tuohy needle at a 45° angle to the skin through the sacral hiatus, checking a lateral fluoroscopic view to make sure the needle has entered the spinal canal. Lower the needle angle and advance the needle slightly. Recheck lateral fluoroscopy to ensure the needle is in the spinal canal.
- Aspirate to rule out intravascular placement and inject 0.5–1 ml contrast medium under live fluoroscopy. Check a lateral and AP image to ensure epidural spread.
- If no epidural catheter is used, inject a mixture of local anesthetic and steroid. Inject 10 ml 0.5% lidocaine plus 50 mg triamcinolone diacetate. This volume should be sufficient to reach the L5 or S1 nerve roots.
- Alternatively, a radio-opaque catheter can be inserted through the needle and advanced to the desired level. Check the catheter position in both AP and lateral views (*Fig. 3.4.7*). Attach the injection hub to the catheter and inject 0.5–1 ml contrast medium under live fluoroscopy, rechecking dye spread in both AP and lateral views (*Fig. 3.4.8*). Inject 1–2 ml 1% lidocaine plus 50 mg triamcinolone diacetate or equivalent.

Fig. 3.4.8

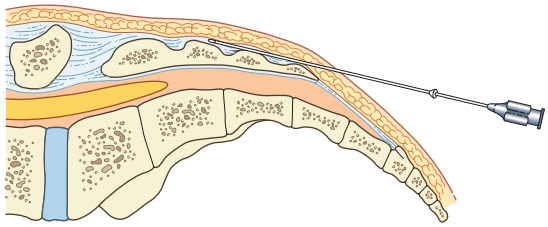

Fig. 3.4.9

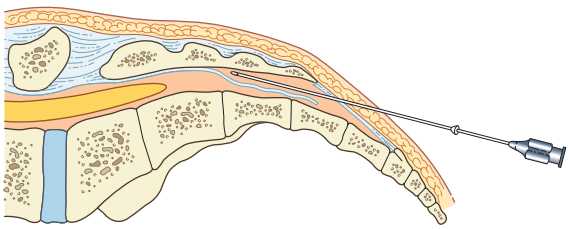

Fig. 3.4.10

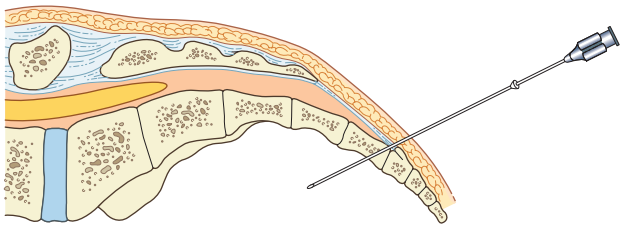

Fig. 3.4.11

Confirmation of a successful block

- Relief of pain.
- Anesthesia or diminished sensation in distribution of blocked nerves.
- Improvement in straight-leg raising (although for sacral nerve root-related pain, this may not have been positive prior to caudal blockade).

Tips

- After location of the caudal epidural space, the left hand may be placed over the sacral hiatus while 10 ml saline is rapidly injected (*Fig. 3.4.6*). Misplacement of the needle in the subcutaneous tissue (*Fig. 3.4.9*) should be evident if the injection is appreciated by the palpating left hand.
- Subperiosteal injection in the awake patient will cause pain (*Fig. 3.4.10*). The needle angle is important as the tip may come to lie anterior to the sacrum (*Fig. 3.4.11*).
- Anatomic variations exist in many patients, making access to the caudal epidural space difficult or impossible. Fluoroscopy using lateral views is useful to confirm the epidural needle position. Ultrasound may also be helpful in identifying the sacro-coccygeal membrane.

Potential problems

IMMEDIATE

- Failure to locate epidural space.
- Intravascular injection (test dose important); addition of epinephrine (adrenaline) to test dose may help identification of intravascular injection.
- Intrathecal injection, rare but possible (test dose important).
- Hypotension due to sympathetic blockade (give i.v. fluid, consider ephedrine).
- Transient exacerbation of radiculopathic pain (caution patient).
- Headache (possible dural puncture, rare).
- Allergic reaction.

LATER

- Infection (epidural abscess, bacterial meningitis).
- Aseptic meningitis, usually the result of intrathecal injection (test dose important).
- Cushingoid symptoms (usually as a result of repeated injections).
- Infection (epidural abscess).
- Epidural hematoma.

3.5 LONG-TERM EPIDURAL CATHETER INSERTION

Anatomy

As described for lumbar epidural block in Section 3.1.

Equipment

- 2 ml and 10 ml syringes
- 18 G, 20 G, and 25 G needles
- ECG, BP, and SpO_2 monitors
- 18 G epidural set
- Epidural catheter passer
- A surgical pack, including small scalpel and suture set
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy or ultrasound (optional)

Drugs

- Lidocaine (lignocaine) 1% (preservative free) 20 ml (or its equivalent)
- Lidocaine (lignocaine) 1% (preservative free) 4 ml plus epinephrine (adrenaline) 1:200 000
- Saline (NaCl) 10 ml

Position of patient

- Lateral, lying on side of radiculopathy (*Fig. 3.5.1*).
- Shoulders and buttocks parallel to the edge of the bed, perpendicular to the floor, with spine flexed.

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- Lidocaine (lignocaine) 1%, 2 ml is drawn up into three 2 ml syringes.
- Lidocaine (lignocaine) 1% 10 ml is drawn up into one 10 ml syringe.
- NaCl 10 ml is drawn up into the 10 ml loss-of-resistance syringe.
- An epidural catheter is inserted through the needle as previously described, 5–6 cm into epidural space (see Section 3.1).
- The needle is withdrawn 1–1.5 cm, but *is not removed*.
- After negative aspiration, a test dose of lidocaine (lignocaine) 1% 4 ml, with epinephrine (adrenaline) 1:200 000, is given. After 5 minutes the patient is questioned about changes in sensation or power, and any changes in heart rate or blood pressure are noted.
- After a negative reaction lidocaine (lignocaine) 1% 10 ml is injected slowly over 10 minutes. Assessment of the level of blockade is carried out after a further 15 minutes.
- Subcutaneous infiltration around the epidural needle with lidocaine (lignocaine) 1% is carried out. Note: a small incision is made to include the epidural needle (*Fig. 3.5.2*).
- A purse string suture is placed around the epidural needle, *but is not tied* (*Fig. 3.5.3*).
- Another small incision in the lateral abdominal wall is made after subcutaneous infiltration with lidocaine (lignocaine) 1% (*Fig. 3.5.4*). A catheter passer is tunneled through the subcutaneous tissue between the two incision sites.
- A catheter is manually bent to a curve and tunneled through the subcutaneous tissue between the two incision sites in the direction from the abdominal site to the epidural needle site (*Fig. 3.5.5*).
- The epidural needle is carefully removed (*Figs 3.5.6, 3.5.7*).
- The catheter is secured to subcutaneous tissue in the midline by tightening the purse string suture (*Fig. 3.5.8*).
- The catheter is further secured using an angle piece and then threaded in a lateral direction through the catheter

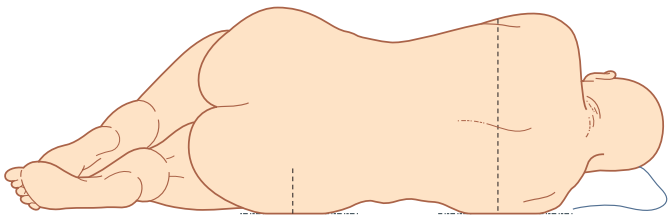

Fig. 3.5.1

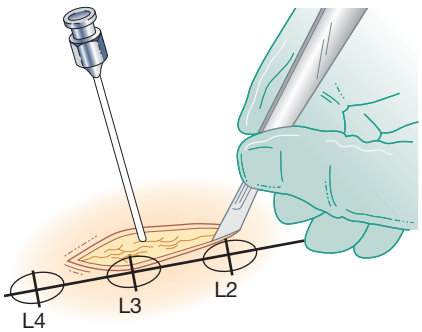

Fig. 3.5.2

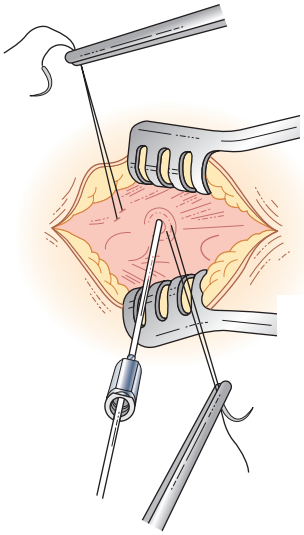

Fig. 3.5.3

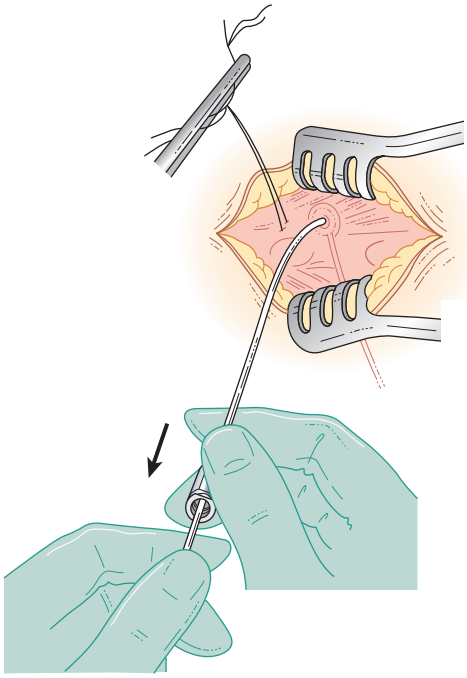

Fig. 3.5.6

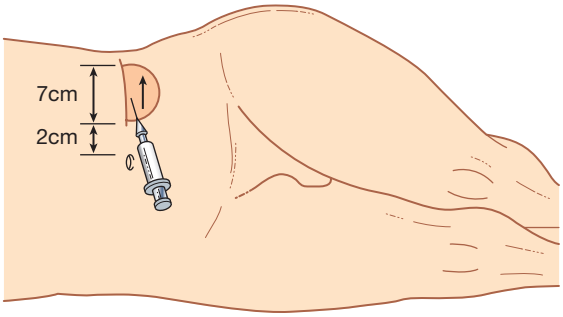

Fig. 3.5.4

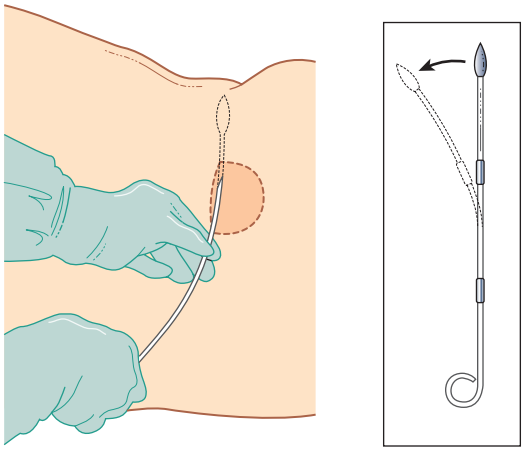

Fig. 3.5.5

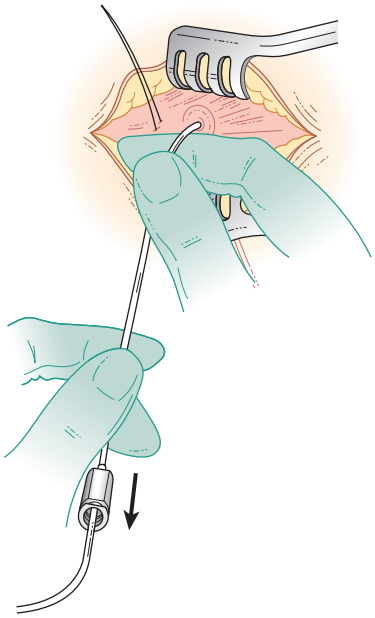

Fig. 3.5.7

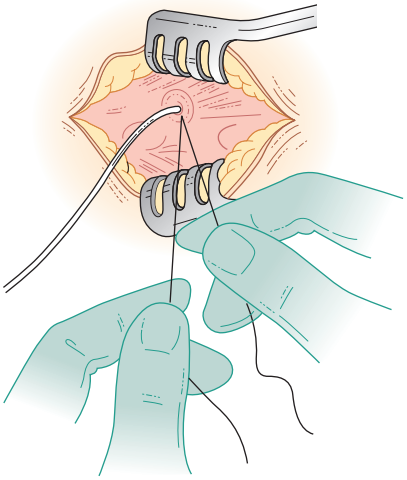

Fig. 3.5.8

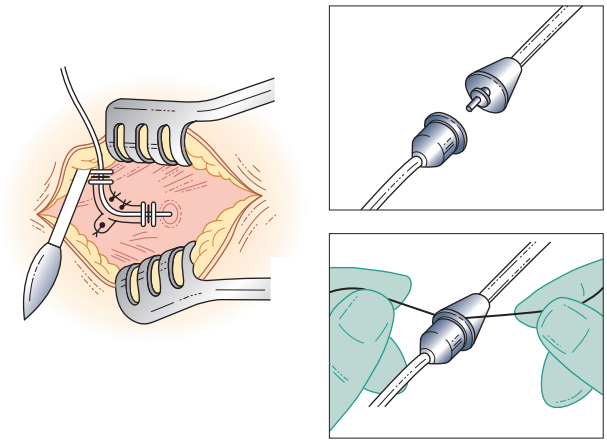

Fig. 3.5.9

passer from the needle site to the abdominal wall site and connections are secured (*Fig. 3.5.9*).

- A pump may be placed in the abdominal wall site and the skin incisions closed.

Confirmation of a successful block

- Relief of pain.
- Anesthesia or diminished sensation in the distribution of affected nerves.

Tips

- Injection of radio-opaque dye under direct fluoroscopy can confirm epidural placement.
- Insertion of a radio-opaque epidural catheter may be carried out also under fluoroscopy.
- Ultrasound may aid insertion

Potential problems

- As described for lumbar epidural block in Section 3.1.
- However, in view of the long-term nature of epidural catheter implantation, any symptoms of infection should be immediately investigated and treated.
- Epidural catheters should not be inserted or removed during anticoagulation. Coagulation and platelet function should be normalized before catheter removal.

SOMATIC NERVE BLOCKADE 4

Mechanical nerve root compression was originally thought to be the cause of pain in discogenic radiculopathy. However, it has been found that many asymptomatic patients demonstrate substantial disc protrusion on magnetic resonance (MR) imaging, myelography and subsequent autopsy examination. In addition, surgical decompression does not result in uniform success in the relief of such pain. Following a period of mechanical nerve-root compression it is likely that an acute inflammatory process ensues, resulting in intraneural accumulation of serum proteins and fluid, raised intraneural pressure, ischemia and axonal degeneration. Degenerating glycoprotein material from the nucleus pulposis may also contribute to the inflammatory process.

There are many situations in which injection of spinal nerve roots with local anesthetic may be helpful in the diagnosis of radicular pain. These include those where investigations including electromyography, computer tomography (CT) or MR imaging are not consistent with the clinical findings, where there are multiple levels of pathology, and after spinal surgery with subsequent scarring in the area of the surgery. In addition, the contribution of the somatic nerve root may be elucidated in pain of uncertain origin, e.g. chest pain or abdominal pain, by specific nerve root local anesthetic injection.

It may be used therefore to determine the level of surgery, if indicated, and the addition of steroid may produce longer-lasting pain relief.

4.1 INTERCOSTAL NERVE BLOCK

Anatomy

The intercostal nerve is made up of several types of nerves: sympathetic white and grey rami communicantes, cutaneous and motor fibers supplied by dorsal rami, sensory fibers to the chest wall, anterior and posterior, via the lateral cutaneous branch, and further sensory fibers to the anterior chest wall via the anterior cutaneous branch. The lateral cutaneous branch exits just distal to the angle of the rib. Just below the inferior edge of the rib, in the intercostal groove, lie the intercostal nerve, artery and vein, the latter lying superior to the nerve. The optimal site to block the intercostal nerve is the most posterior point at which the rib is palpable, usually the angle of the rib (**Fig. 4.1.1 a,b**).

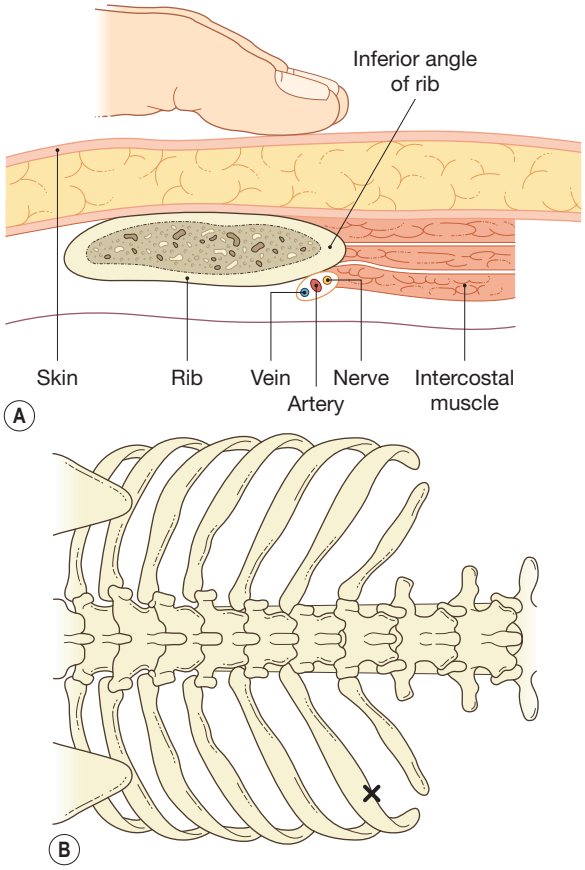

Fig. 4.1.1

Equipment

- 2 ml and 5 ml syringes
- 30 G needle
- 22 G 3–4 cm short-bevel needle
- Extension set (optional)
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- Ultrasound (optional)

Drugs

- Lidocaine (lignocaine) 1% 2 ml for skin infiltration
- Lidocaine (lignocaine) 1% 5 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 50 mg (or its equivalent)
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone (this allows best access, although a lateral or supine position may also be used).
- Pillow under mid-abdomen to widen the intercostal spaces.
- Arms hanging over sides of table to rotate the scapulae laterally.

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- Sedation may be administered if multiple blocks are being performed.
- The midline and an area 10 cm × 10 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- The midline is palpated and marked.
- **The inferior edge of the rib is palpated and marked at the most posterior point at which the rib is palpable; this is the insertion point** (*Fig. 4.1.2*). If multiple blocks are planned these marks will form a line that becomes more medial towards the cephalad end as the scapulae are avoided laterally (*Fig. 4.1.2*).

FOR THE RIGHT-HANDED OPERATOR

With the left hand

- The inferior edge of the rib is palpated with the fore- and middle fingers (*Fig. 4.1.3*).

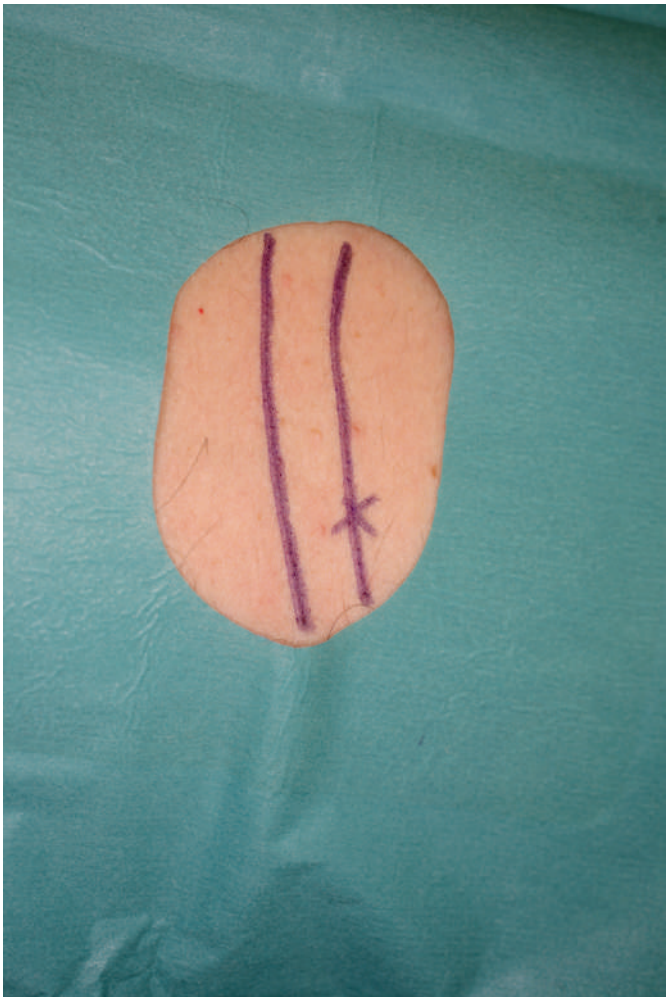

Fig. 4.1.2

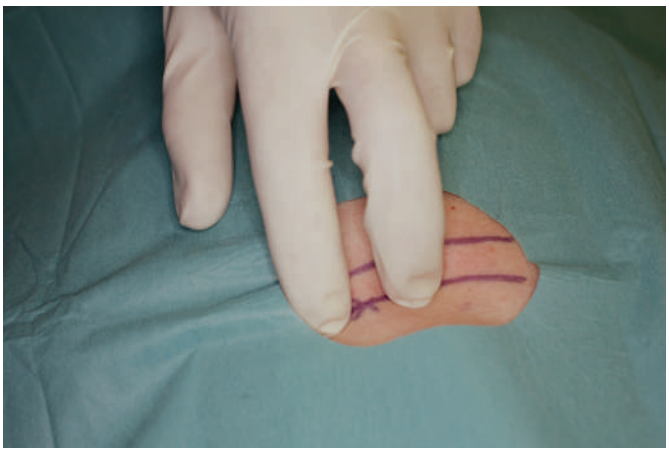

Fig. 4.1.3

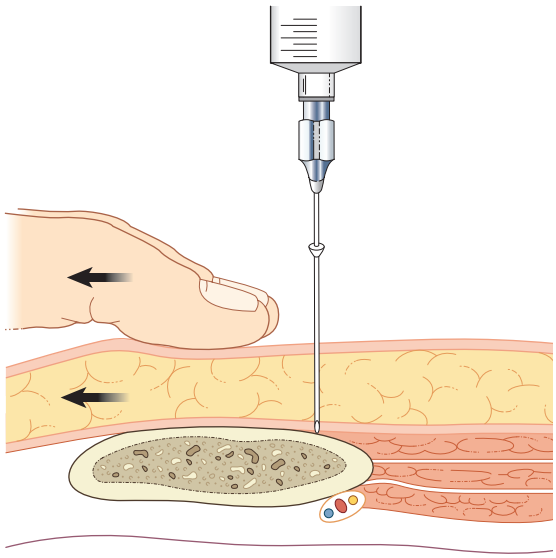

Fig. 4.1.4

- The skin is drawn up over the rib itself.
- The fingers of the left hand will grip the needle-hub for controlled advancement of the needle once contact with the rib is made during injection.

With the right hand

- The insertion point is infiltrated with lidocaine (lignocaine) 1% using a 2 ml syringe and a 30 G needle.
- The 22 G short-bevel needle with syringe attached is inserted between fore- and middle finger of the left hand in a direction 15–20° cephalad, until it makes contact with the rib (*Figs 4.1.4, 4.1.5*).
- The needle-hub is gripped with the fingers of the left hand and this hand is steadied by leaning the wrist against the patient's posterior chest wall.
- With the right and left hands acting as one unit, the needle is walked off the edge of the rib until it enters the intercostal space immediately below the rib (*Fig 4.1.6*). Alternatively, a catheter may be inserted between the needle and syringe and injection may then be carried out by a second operator, while the first maintains the needle steady in the correct position (*Fig. 4.1.7*).
- It is then advanced 2 mm.
- After negative aspiration, 3–4 ml of local anesthetic, plus corticosteroid if indicated, is injected and the needle is withdrawn.
- Monitors are left attached and i.v. access left in situ for at least 30 minutes.
- Chest X-ray is performed if pneumothorax is suspected.

Fig. 4.1.5

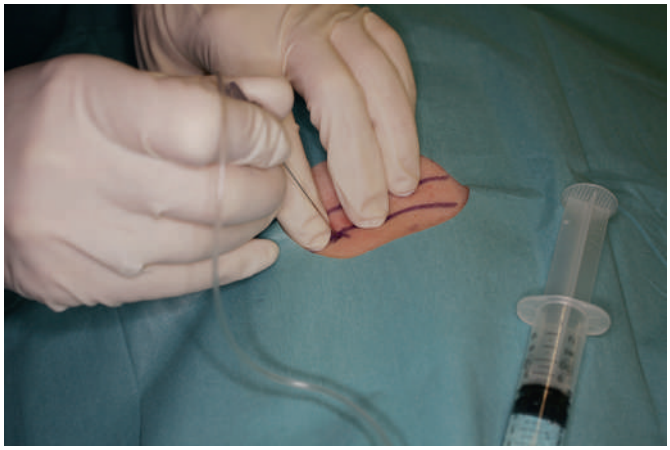

Fig. 4.1.7

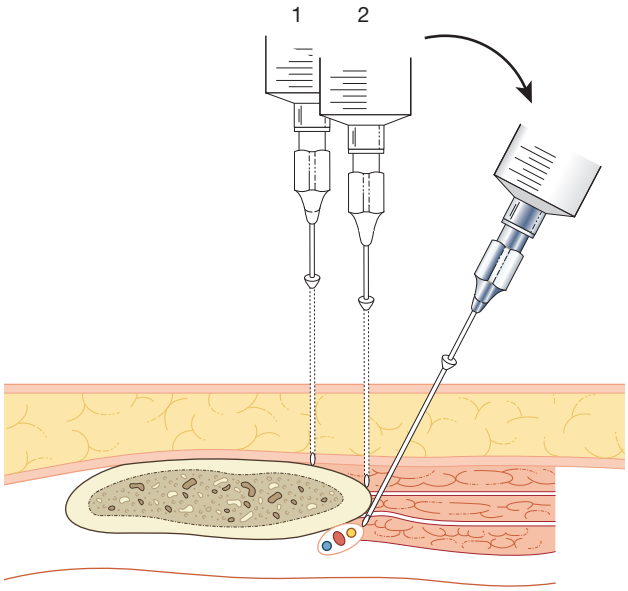

Fig. 4.1.6

Confirmation of a successful block

- Relief of pain.
- Anesthesia in the distribution of the blocked nerve.

Tips

- The approximate depth to the rib may be determined with the left fore- and middle fingers before insertion of the needle.
- Insertion of the needle > 2 mm deeper than the rib when intercostal space is reached is avoided. This will minimize the risk of pneumothorax, as the average distance from the rib to the pleura is 8 mm. If patient coughs on injection, pneumothorax is suspected.
- Neurolytic intercostal nerve block, e.g. with alcohol 50% 3 ml (made up by combining equal parts of alcohol 100% and lidocaine (lignocaine) 1% or its equivalent), or phenol 6%, may be carried out after local anesthetic block confirms accurate placement of the needle as described above. However, it is important to note that injection may result in subarachnoid spread of a neurolytic agent with resultant possible permanent spinal cord damage.
- Ultrasound may aid accurate placement of the needle (*Fig. 4.1.8*). Injection of non-ionic radio-constrast medium may also aid accurate placement of the needle in neurolytic block (*Fig. 4.1.9*).
- Placement of a catheter into the intercostal space can be achieved by threading 3 cm of catheter through an 18 G epidural needle after the intercostal neurovascular bundle has been identified as above.

Fig. 4.1.8

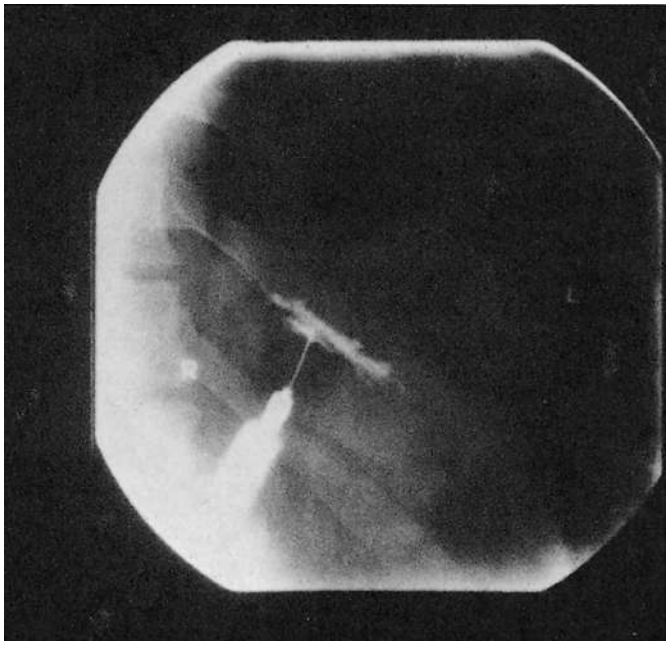

Fig. 4.1.9

Radiofrequency lesioning

- Radiofrequency lesioning of the intercostal nerve is simple and has a low level of side effects. The lesioning is carried out using the same method of placement of the needle as described for intercostal nerve block with local anesthetic.
- However, after placement of the needle and confirmation of accuracy by fluoroscopy (as described above), a trial of stimulation is carried out using 2 V at 50 Hz. If the needle has been placed accurately the patient should experience paresthesiae in the distribution of that intercostal nerve. A pulsed radiofrequency lesion may then be produced at 40–45 °C for 5 minutes or 49–60 °C for 90 seconds.

Potential problems

- Injection within the nerve sheath can result in the spread of anesthetic to the subarachnoid space.
- Intercostal block in patients with severe respiratory problems should be avoided as there is risk of a pneumothorax. Careful observation of a small pneumothorax is usually all that is required but failure to re-expand the lung may require chest tube insertion.
- Because of the vascularity of the intercostal space, there may be rapid absorption of local anesthetic and systemic effects can occur quickly, especially with multiple blocks. However, peak plasma concentration of local anesthetic may occur 15–20 minutes after the block is performed, when systemic toxicity effects may develop. Addition of epinephrine (adrenaline) to the anesthetic solution may decrease the peak plasma concentration of local anesthetic.
- If aspiration of blood occurs, the needle should be removed, keeping the left fore- and middle fingers in place. The needle is cleared, reinserted to contact the rib again, and the block is continued as above.

- Care of the airway must be remembered if sedation is administered to a patient in the prone position.
- Complications of neurolytic intercostal nerve block include pneumothorax, infection (especially in the immunocompromised patient) as well as post-lesioning intercostal nerve neuritis. The frequency of the latter in radiofrequency lesioning increases as higher temperatures are used. Intercostal nerve neuritis usually responds to local injection of lidocaine (lignocaine) 1% 3 ml plus triamcinolone diacetate 20 mg to the lesion site.

4.2 INTERPLEURAL BLOCK

Anatomy

The parietal pleura lines the thoracic wall, the thoracic surface of the diaphragm and the lateral mediastinum. The visceral pleura completely covers the surface of the lung. Both the pleural layers become contiguous at the root of the lung. Between the two pleural layers lies the interpleural space (**Fig. 4.2.1**). Injection of local anesthetic into this space produces an interpleural block by topical contact with free nerve endings within the pleura, and by local diffusion to nerves in the vicinity of the injection site. These include the intercostal nerves, the sympathetic chain, and the inferior part of the brachial plexus. Local anesthetic solution may also track to the epidural and subarachnoid spaces producing blockade.

Equipment

- 2 ml and 10 ml syringes
- 18 G, 20 G, and 25 G needles
- ECG, BP, and SpO_2 monitors
- 18 G epidural set
- Well-lubricated 5 ml glass syringe
- Resuscitation equipment (*see Appendix 3*)

Drugs

- Lidocaine (lignocaine) 1% 10 ml
- Levobupivacaine 0.25%, 20 ml (or its equivalent)
- Saline (NaCl) 10 ml
- Resuscitation drugs (*see Appendix 3*)

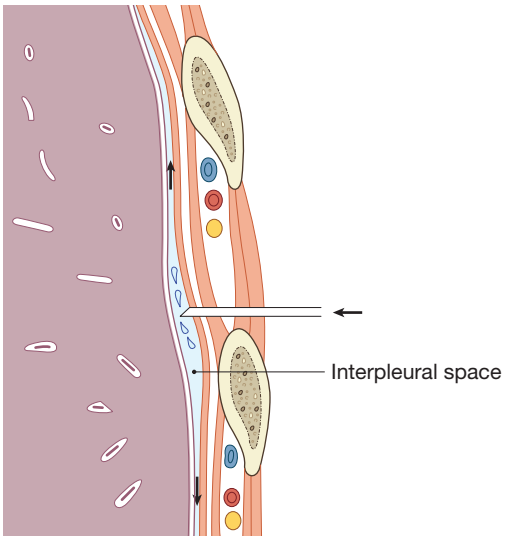

Fig. 4.2.1

Position of patient

- The technique described here relies on negative interpleural pressure to identify the interpleural space. Therefore the patient should be breathing spontaneously for this technique.
- Semi-prone.
- Side to be blocked uppermost, supported by a pillow (*Fig. 4.2.2*).
- The arm should be allowed to fall forwards in front of the body to rotate the scapula anterolaterally.

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The midline and an area 15 cm × 12 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- Lidocaine (lignocaine) 2%, 2 ml is drawn up.
- Levobupivacaine 0.25% 20 ml is drawn up.
- Air is drawn up into a well-lubricated 5 ml syringe.
- The seventh and eighth ribs are palpated and marked.

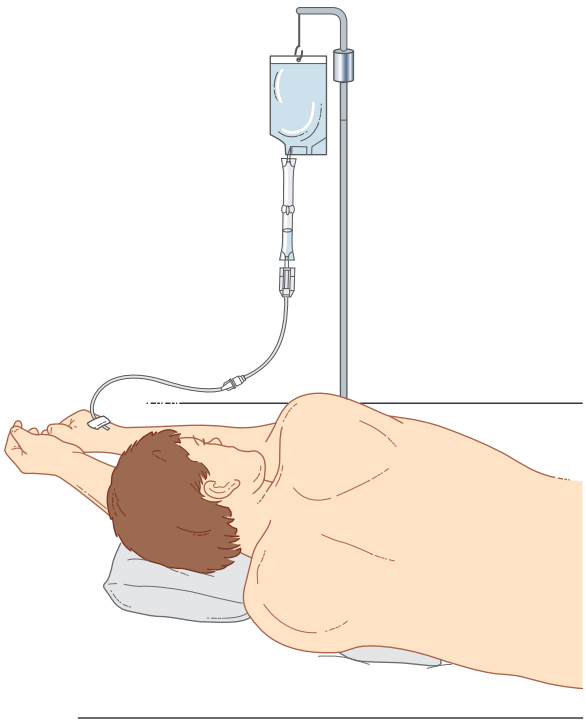

Fig. 4.2.2

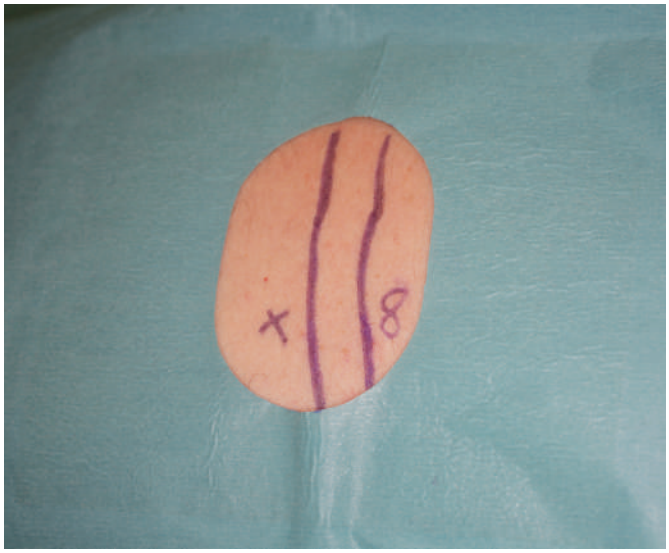

Fig. 4.2.3

- **A point approximately 10 cm from the midline, immediately superior to the eighth rib, is marked; this is the insertion point** (*Fig. 4.2.3*).

FOR THE RIGHT-HANDED OPERATOR

With the left hand

- The fore- and middle fingers are placed each side of the insertion point, palpating the superior aspect of the eighth rib.
- These fingers are kept in place until the epidural needle passes through the subcutaneous tissue.

With the right hand

- A skin wheal is raised at the insertion point.
- The epidural needle is inserted between the fore- and middle fingers of the left hand, taking care that the point of entry of the needle is as close as possible to the superior aspect of the eighth rib. This helps to avoid damage to the neurovascular bundle, which lies immediately inferior to the seventh rib.
- After passage through the subcutaneous tissue, the hub of the needle is gripped with the fore- and middle fingers of the left hand and this hand is steadied by leaning the wrist against the patient's posterior chest wall.
- The stylet is removed and the well-lubricated glass syringe containing 3 ml air is applied (*Fig. 4.2.4*).
- The needle is slowly and carefully advanced, with no pressure applied to the plunger, the left hand aiding the advance, while at the same time applying a brake if required.
- Resistance from the tissues prevents the plunger from advancing.
- At the point at which the needle enters the interpleural space a definite click is experienced, negative pressure draws the air in and the barrel drops (*Fig. 4.2.5*).

Fig. 4.2.4

- Taking care not to allow air entry into the interpleural cavity, a catheter is inserted approximately 10 cm into the interpleural space (*Fig. 4.2.6*). **Once the catheter is in position it is best to place the patient supine, tilted slightly, with the side to be blocked upwards** (*Fig. 4.2.7 a,b*). **(If blockade of the upper thoracic segments is required the patient is tilted head-down).**
- After negative aspiration and a test dose, 10–15 ml levobupivacaine 0.25% (or its equivalent) in divided doses of 5 ml is injected.
- Infusion of local anesthetic may be set up for continuous analgesia.
- Monitors should be left attached and i.v. access left in situ while the catheter is in place.
- Chest radiograph may be performed to rule out pneumothorax.

Confirmation of a successful block

- Relief of pain.
- Anesthesia in the distribution of the blocked nerves.

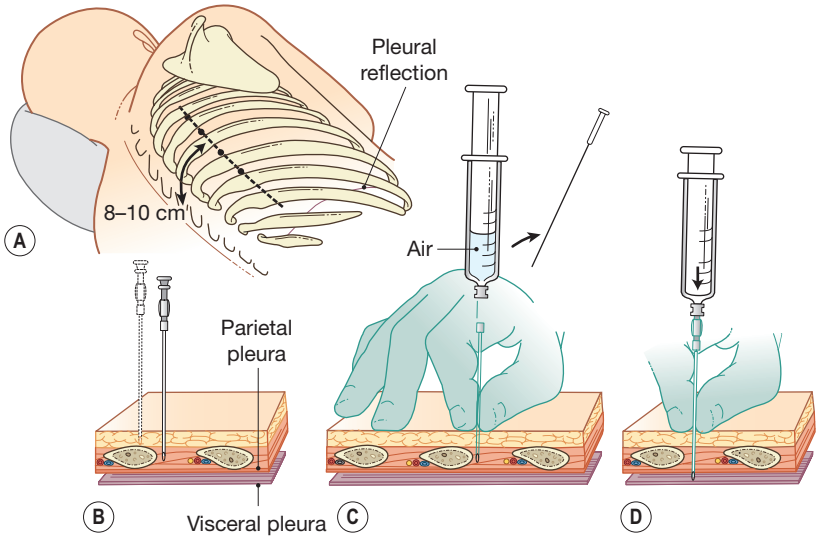

Fig. 4.2.5

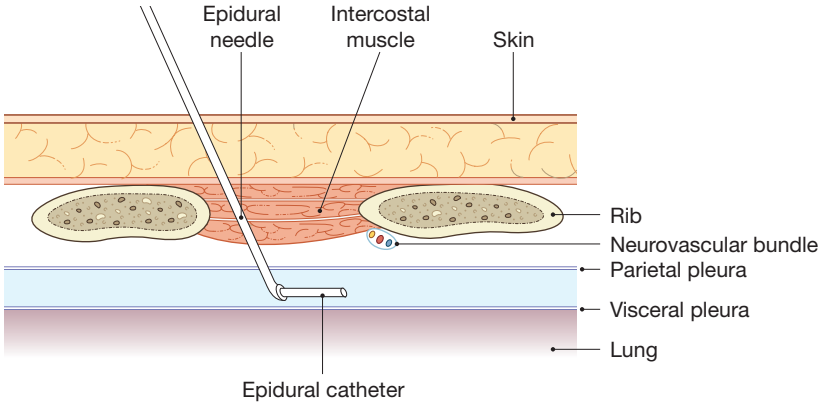

Fig. 4.2.6

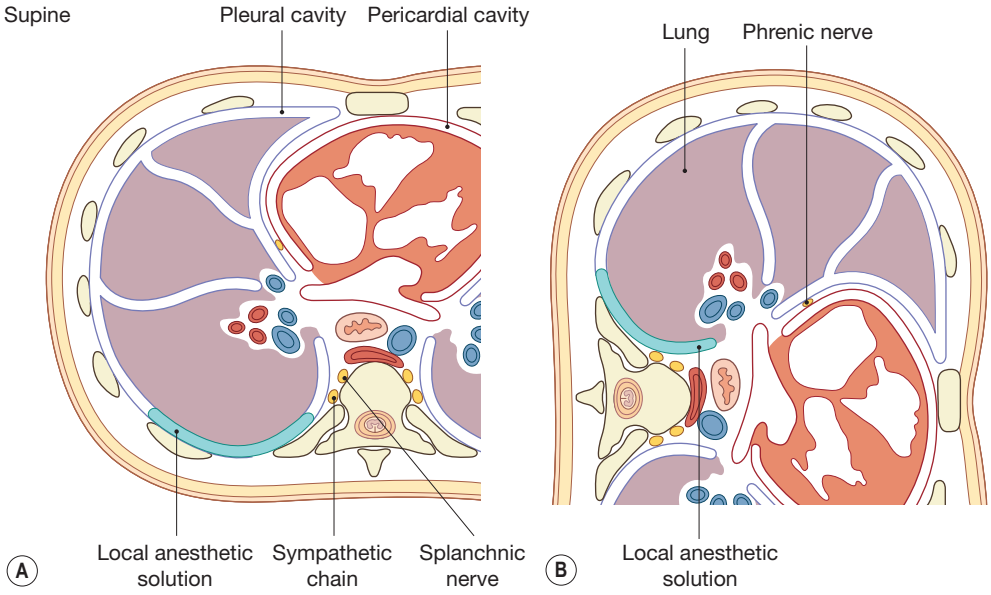

Fig. 4.2.7

Tips

- As an alternative technique, the barrel of the syringe may be removed and the open syringe filled with NaCl or local anesthetic and advanced until the fluid level begins to fall as the solution is sucked into the interpleural space (*Fig. 4.2.8*).
- Also, a bag containing NaCl 500 ml may be attached to the epidural needle via a giving set, and drops observed on entry to the interpleural space (*Fig. 4.2.9*).
- Ultrasound may aid placement of the needle (*Fig. 4.2.10*).

Potential problems

- Pneumothorax.
- Unpredictable analgesia. The mechanism of action of interpleural block is still unproven and spread of the local anesthetic solution may be unpredictable. The duration of the block may be decreased when a thoracotomy drainage tube is present.

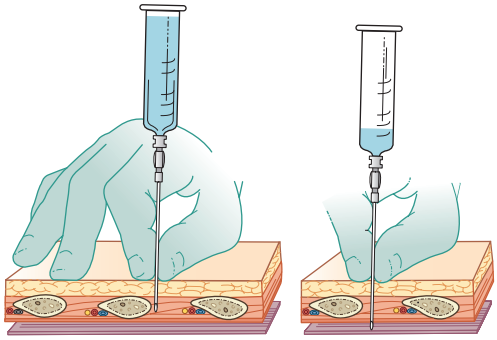

Fig. 4.2.8

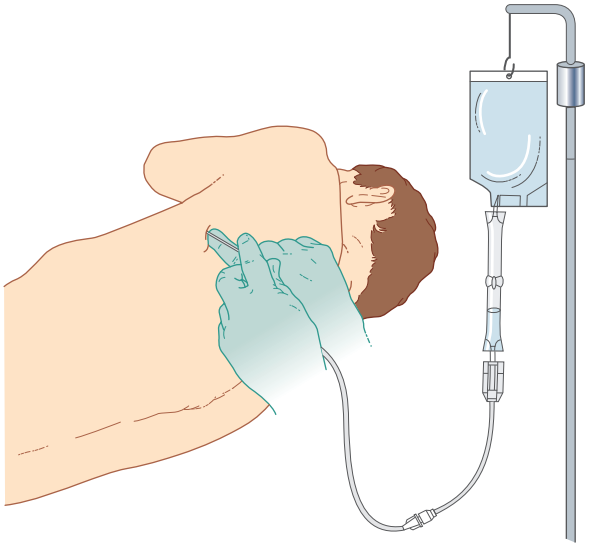

Fig. 4.2.9

Fig. 4.2.10

4.3 LUMBAR NERVE ROOT BLOCK

Anatomy

The lumbar nerves are made up of sensory and motor fibers to the trunk and lower limbs, and sympathetic white and grey rami communicantes. Each lumbar nerve exits via the intervertebral foramen which lies just inferior to the caudad edge of the transverse process of the respective vertebral body, and passes anteriorly over the lateral aspect of the transverse process of the vertebral body below (**Fig. 4.3.1**). It then branches into posterior and anterior branches. The posterior branch supplies the paravertebral muscles and cutaneous fibers to the back. The anterior branch passes through the substance of the psoas muscle and branches further, communicating with the other anterior branches to form the lumbar plexus. As a result there is significant overlap of nerve supply. The fascial layers of the psoas muscle prevent spread of local anesthetic to the sympathetic lumbar chain. It may be helpful to consider blockade of a lumbar nerve root to be similar to intercostal nerve block except that the transverse process is present instead of a rib and the insertion site is therefore more medial.

Equipment

- 2 ml and 10 ml syringes
- 25 G needle

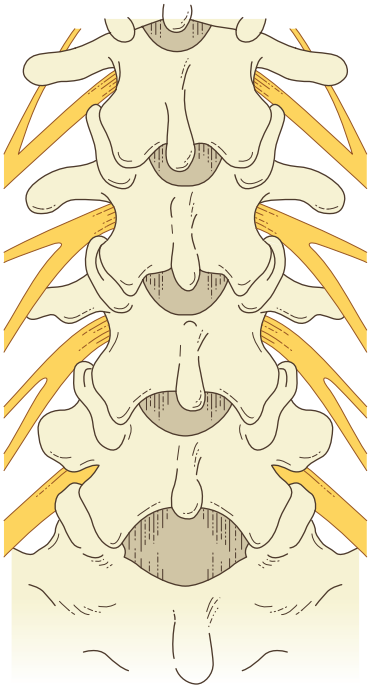

Fig. 4.3.1

- 22 G spinal needle, end-opening
- Radio-opaque contrast medium
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy or ultrasound

Drugs

- Lidocaine (lignocaine) 1% 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 25 mg (or its equivalent)
- Resuscitation drugs

Position of patient

- Prone.
- Pillow under the anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 4.3.2*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The lumbar midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- The iliac crests are palpated and the intercrestal line is identified. This corresponds with the inferior aspect of the spinous process of L4 or may lie in the L4–5 interspace (*Fig. 4.3.3 a*).
- The spinous processes are counted until the level to be blocked is identified and confirmed with fluoroscopy.
- The spinous processes of the vertebrae are marked (*Fig. 4.3.3 b*).
- **The insertion point of the needle lies 2–3 cm lateral to the cephalic end of the spinous process of the vertebra. The nerve corresponding to each vertebra emerges just below the transverse process of that vertebra at this site** (*see Appendix 6*).
- Therefore, with the aid of fluoroscopy, the insertion point is identified.
- A skin wheal is raised and the area is infiltrated with lidocaine (lignocaine) 1%.
- A spinal needle is introduced in a vertical direction to the skin, until the needle contacts bone at an approximate depth of 3–5 cm, the transverse process of that vertebra (*Fig. 4.3.4 a,b*).

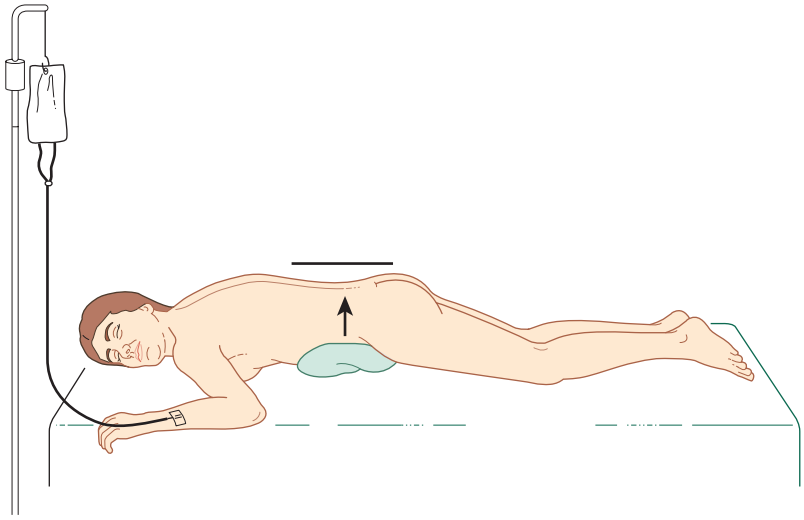

Fig. 4.3.2

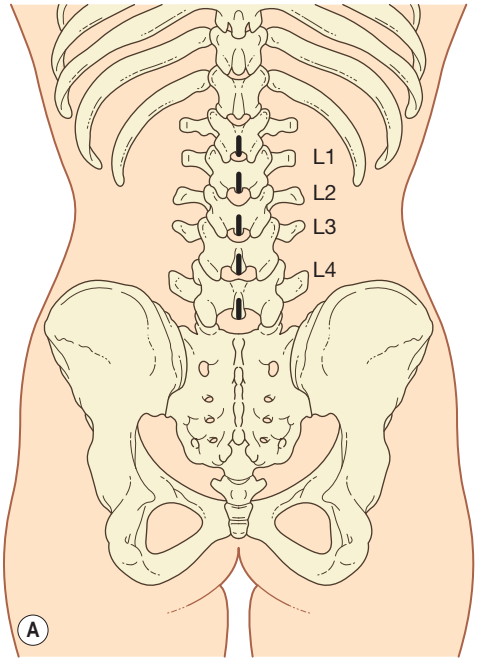

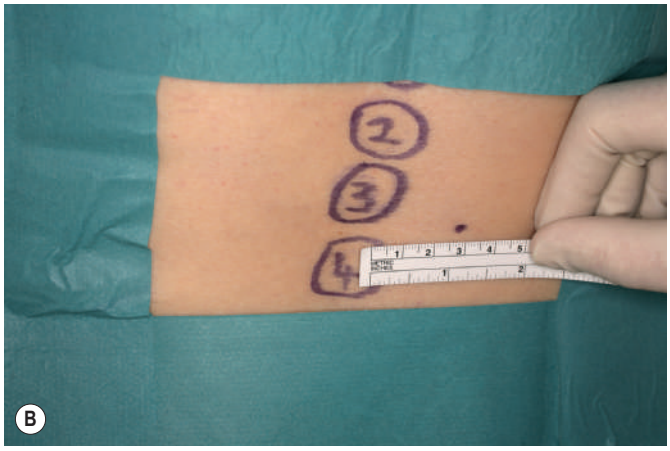

Fig. 4.3.3

- The needle is then walked off the transverse process in the caudad direction and advanced 1.5–2 cm, the site of the emerging nerve root (*Fig. 4.3.5*).
- It is useful to confirm the needle tip over the intervertebral foramen with fluoroscopy.
- Paresthesia in the distribution of the nerve may be experienced.
- After aspiration, non-ionic radio-opaque contrast medium 1 ml is injected.
- The correct placement is indicated by outlining the nerve root with non-ionic radio-opaque contrast medium, visible on anteroposterior and lateral fluoroscopic views (*Fig. 4.3.6*).
- After further aspiration, lidocaine (lignocaine) 1% 0.5–1 ml is injected.
- After 5 minutes the patient is questioned about changes in pain, sensation and power of the lower limb.
- For diagnostic nerve root blockade, the needle may be removed when the level causing pain is identified.
- Ultrasound may also aid placement of the needle (*Fig. 4.3.7*).

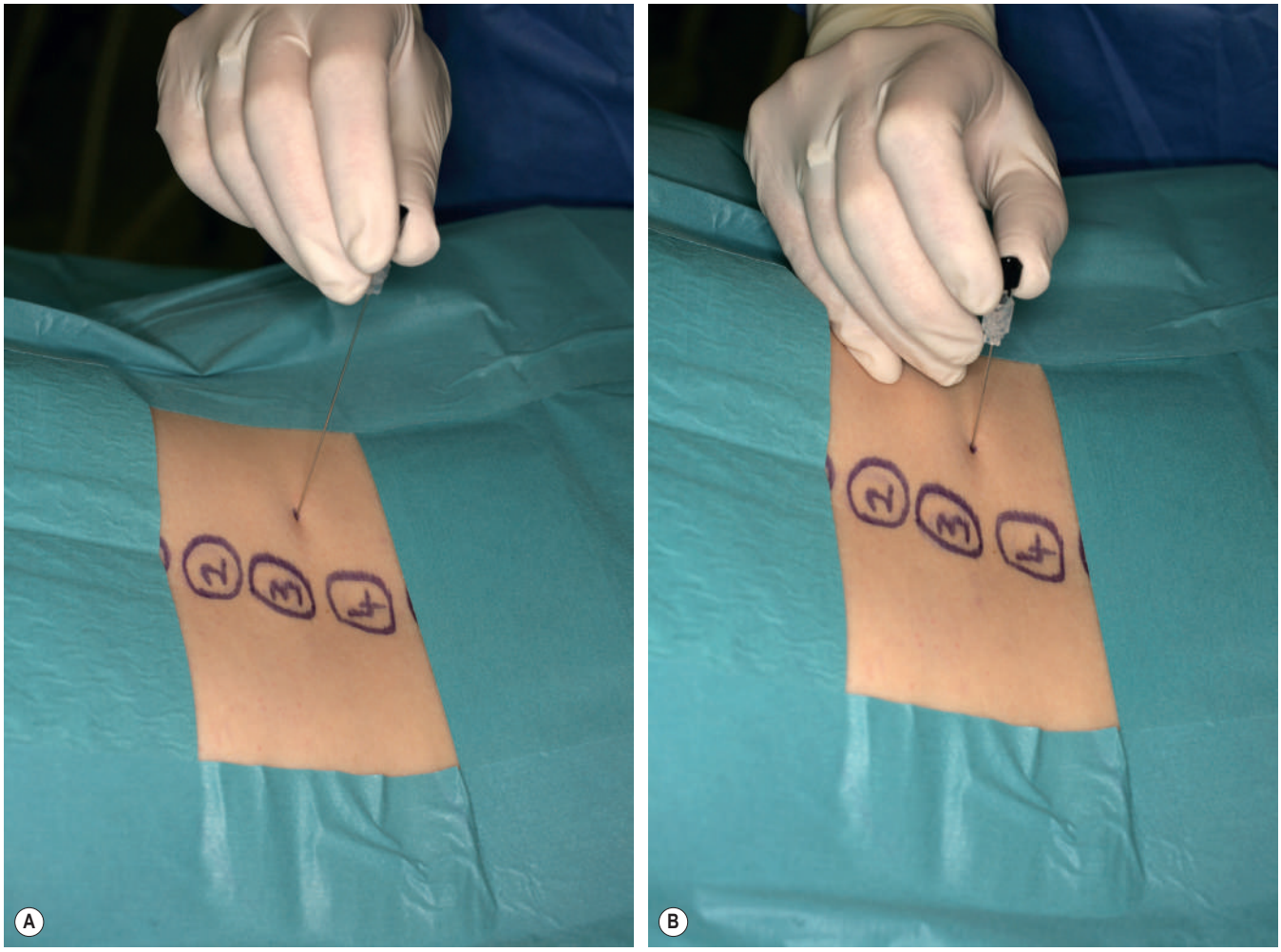

Fig. 4.3.4

Confirmation of a successful block

- Relief of pain and anesthesia in distribution of the blocked nerve.

Tips

- As in the case of thoracic nerve root block, it has also been recommended that the needle is angled 20° medially after entering the paravertebral space. However care must be taken, with the aid of fluoroscopy, not to inject local anesthetic solution into the nerve sheath allowing tracking of the solution centrally to produce intrathecal blockade.

Potential problems

- Intrathecal injection.
- Epidural blockade usually occurs with this block, but this is not a problem once low volumes are used. However, even small volumes of epidural spread may confound the diagnostic value of the block.
- Sympathetic blockade is unlikely, but it may occur and cause hemodynamic changes.
- Intravascular injection. Injection of particulate steroids into a radicular artery can cause spinal cord infarction. Particulate steroids should never be injected near the foramen unless intravascular placement has been ruled out using live fluoroscopy contrast dye injection, preferably with digital subtraction technique.

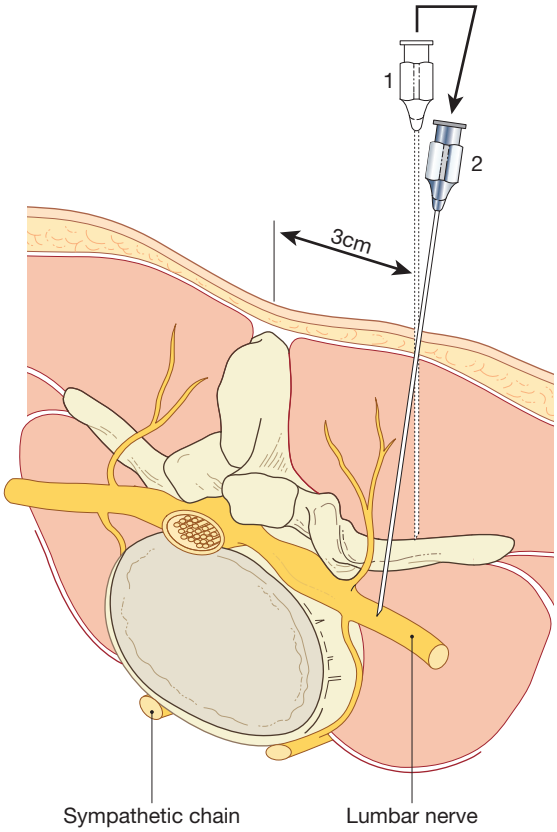

Fig. 4.3.5

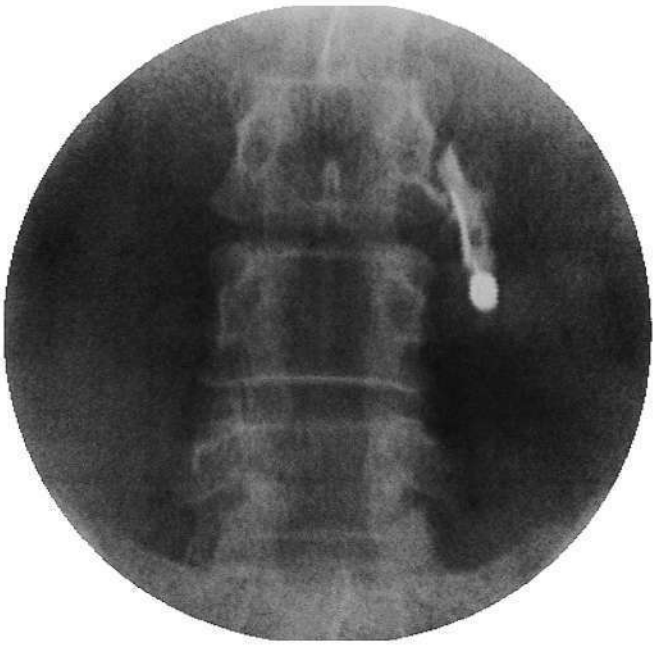

Fig. 4.3.6

Fig. 4.3.7

4.4 THORACIC NERVE ROOT BLOCK

Anatomy

The anatomy relevant to thoracic paravertebral nerve root blockade is very similar to the anatomy relevant to lumbar paravertebral nerve root block, except that:

- ribs are present instead of the rudimentary ribs of the lumbar spine, the transverse processes;
- the lung and pleura are in close proximity to the injection site, therefore the risk of pneumothorax is significant;
- unlike the lumbar region, where the needle passes through the substance of the psoas muscle, the needle passes through connective tissue only in the thoracic region (*Fig. 4.4.1 a,b*);
- the fascial layers of the psoas muscle prevent spread to the sympathetic lumbar chain. However, such layers do not exist in the thoracic region, and paravertebral injection usually results in sympathetic blockade.

Equipment

- 2 ml and 10 ml syringes
- 25 G needle
- 22 G spinal needle, end-opening
- Radio-opaque contrast medium
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment and drugs (*see Appendix 3*)
- Fluoroscopy or ultrasound

Drugs

- Lidocaine (lignocaine) 1% 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 25 mg (or its equivalent)
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone (the block requires localization of the transverse process and can be performed in a lateral, sitting or prone position) (*Fig. 4.4.2*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The thoracic midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution.

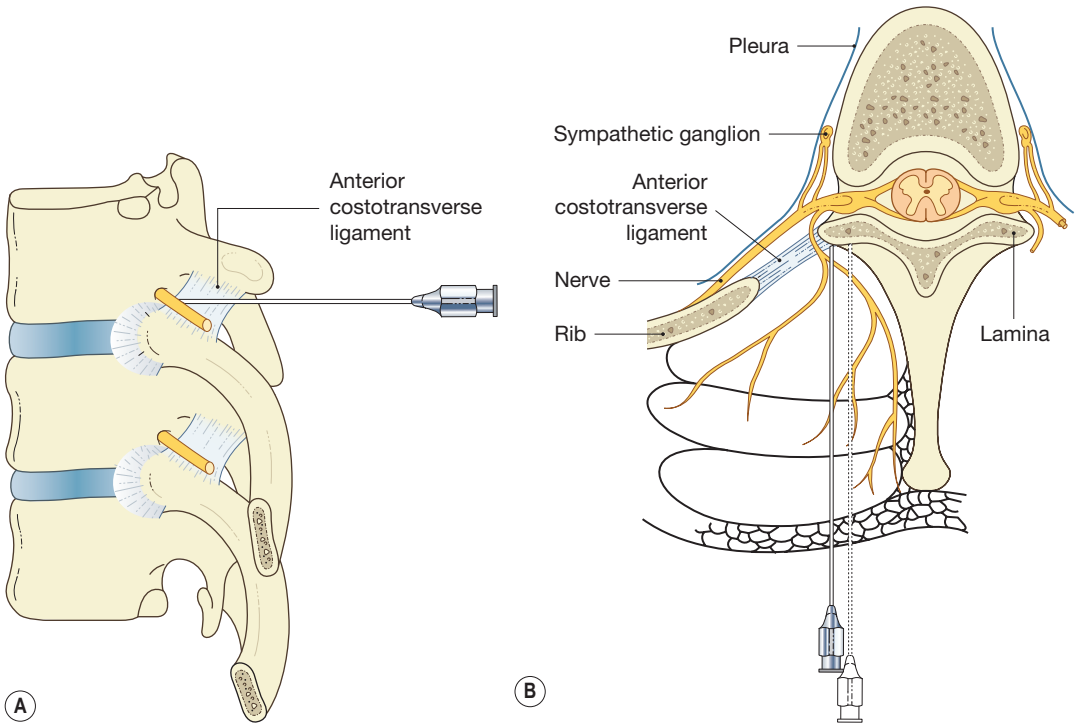

Fig. 4.4.1

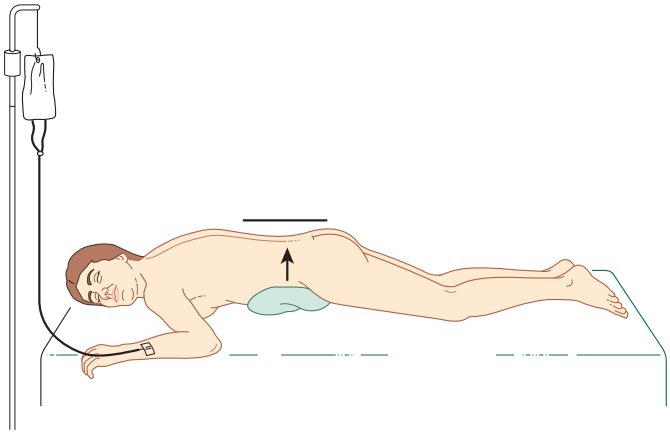

Fig. 4.4.2

- The inferior angle of the scapula is identified; this lies at the level of the spinous process of T7.
- The root of the spine of the scapula is identified; this lies at the level of the spinous process of T3.
- The spinous processes are counted until the level to be blocked is identified, and confirmed with fluoroscopy.
- The spinous processes of vertebrae are then marked.
- **The insertion point of the needle lies 1.5–3 cm lateral to the cephalic end of the spinous process of the vertebra, usually midway between the ribs** (*Fig. 4.4.3 a,b*).
- The nerve corresponding to each vertebra emerges just below the transverse process of that vertebra at this site.
- Therefore, with the aid of fluoroscopy, the insertion point is identified.
- A skin wheal is raised and the area is infiltrated with lidocaine (lignocaine) 1%.
- The transverse process is identified under fluoroscopy and the spinal needle is introduced in a direction perpendicular to the skin until the needle contacts bone.
- A slightly mesiad inclination avoids the pleura and increases the chances of placing the needle tip near the nerve root.
- Care must be taken to avoid the pleura by not advancing the needle any further than is necessary to locate the transverse process, approximately 3 cm. If the needle does not contact the transverse process by 3 cm, it should be withdrawn and re-advanced in a more caudad, and subsequently more cephalic, direction, again taking care not to advance beyond 3 cm.
- The needle is then walked off the caudad edge of the transverse process until it slips off the edge (*Fig. 4.4.4 a,b*). It is then advanced 1 cm, the site of the emerging nerve root at that level.
- It is useful to confirm the needle tip over the intervertebral foramen with fluoroscopy.
- Paresthesia in the distribution of the nerve may be experienced.

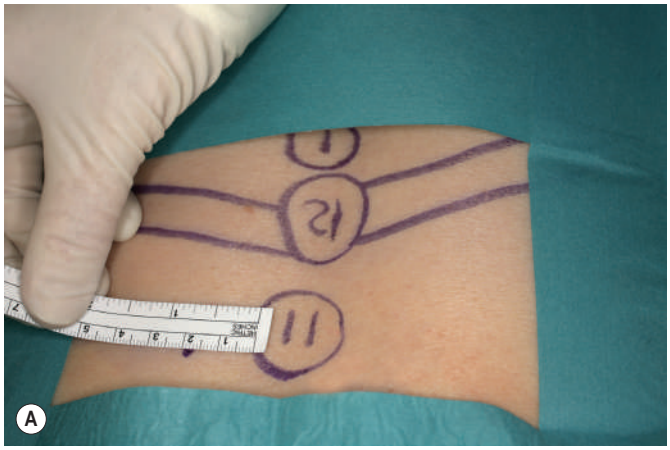

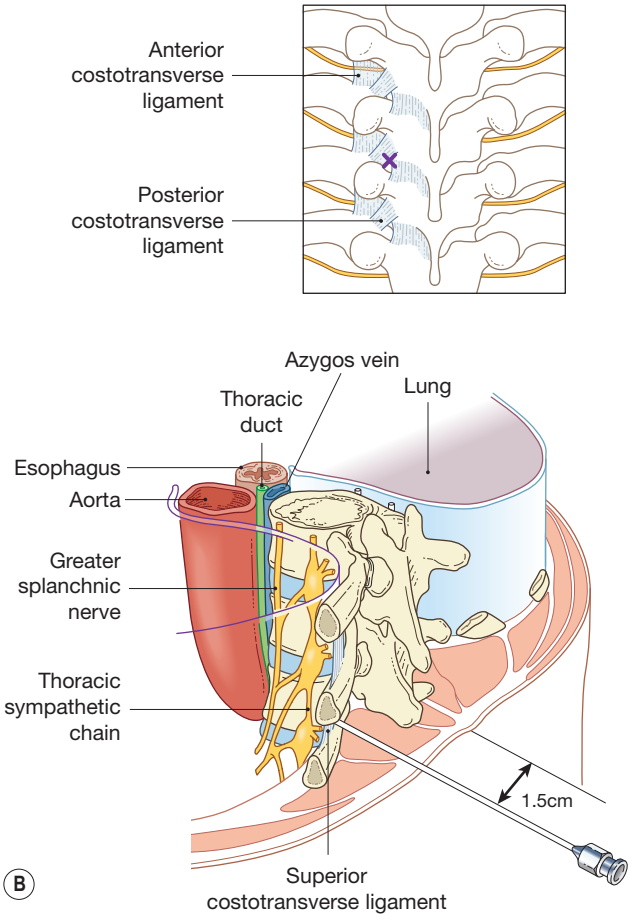

Fig. 4.4.3

- After aspiration, radio-opaque contrast medium 1 ml, is injected.
- The correct placement is indicated by outlining the nerve root with non-ionic radio-opaque contrast medium, visible on anteroposterior and lateral views (as in the description of lumbar somatic nerve injection in Section 4.3).
- Ultrasound may aid placement of needle.

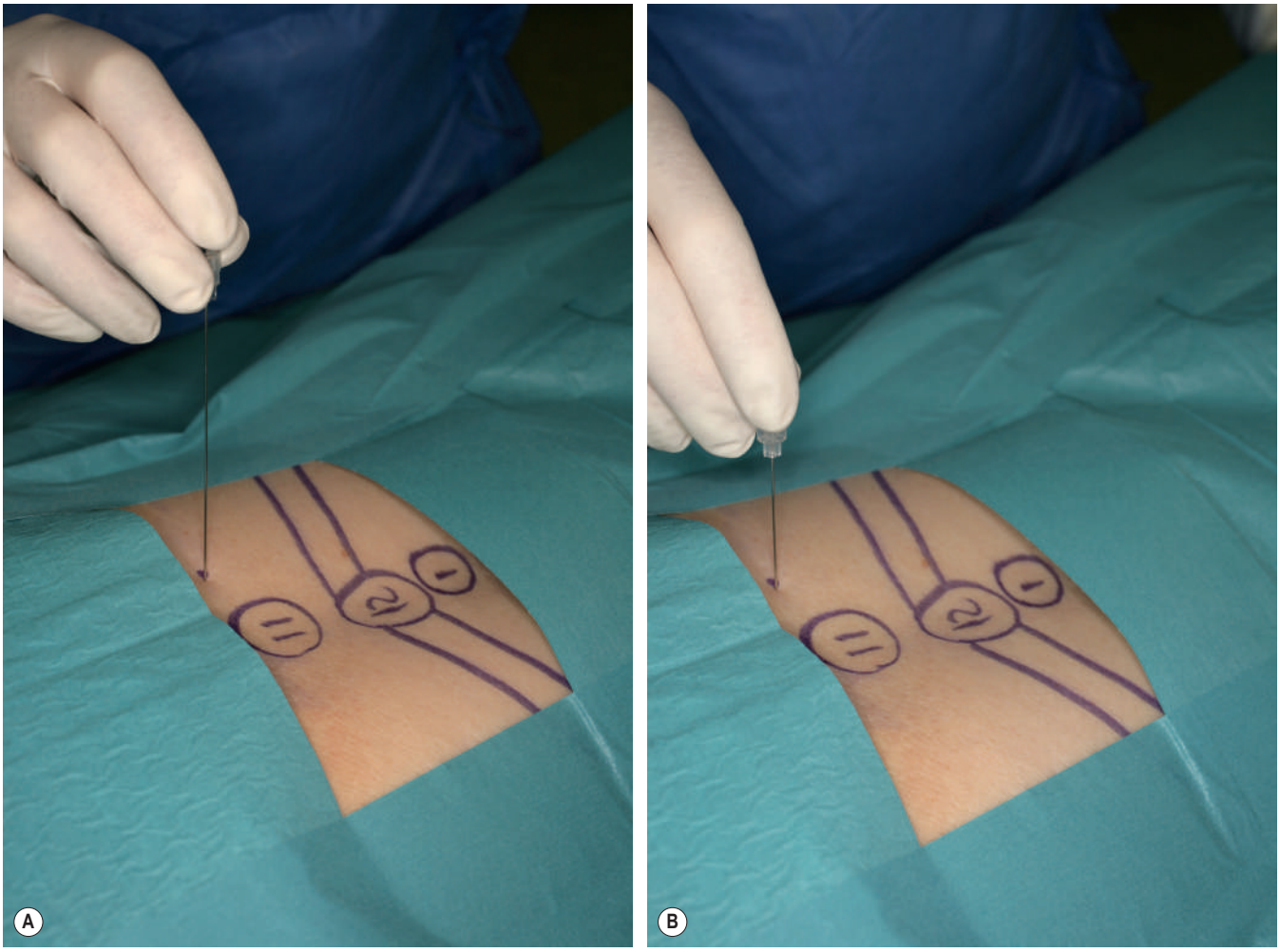

Fig. 4.4.4

- After further aspiration, lidocaine (lignocaine) 1% 0.1–1 ml is injected.
- After 5 minutes the patient is questioned about changes in pain and sensation in the distribution of the nerve root.
- For diagnostic nerve root blockade the needle may be removed when the level causing pain is identified.

Confirmation of a successful block

- Relief of pain and anesthesia in distribution of the blocked nerve.

Tips

- An alternative approach is to advance the needle in a mesiad direction until the lamina is contacted. The needle is inserted more medially, 1.5 cm lateral to the cephalad edge of the spinous process and then walked laterally off the edge of the lamina until it slips into the costovertebral ligament and is advanced 1 cm.
- Some workers advocate applying an air-filled loss-of-resistance syringe to the needle after it has been walked off the transverse process or lamina, and advancing the needle while applying pressure to the plunger. Loss of resistance has been described as the needle pops through the costotransverse ligament to enter the thoracic paravertebral space.
- A catheter may be passed into the paravertebral space via an epidural needle, with the bevel medial, by using this technique.

Potential problems

- Pneumothorax.
- While it has been recommended that the needle be angled 20° medially after entering the paravertebral space, care must be taken with the aid of fluoroscopy that the local anesthetic solution is not injected into the nerve sheath causing the solution to track centrally to produce intrathecal blockade.

- Epidural blockade usually occurs with this block, but this is not a problem once low volumes are used. However, even small volumes of epidural spread may confound the diagnostic value of the block.
- Sympathetic blockade may cause hemodynamic changes.
- Neuritis may occur with catheter placement.
- Intravascular injection. Injection of particulate steroids into a radicular artery can cause spinal cord infarction. Particulate steroids should never be injected near the foramen unless intravascular placement has been ruled out using live fluoroscopy contrast dye injection, preferably with digital subtraction technique.

4.5 SACRAL NERVE ROOT BLOCK

Anatomy

Each of the five sacral nerves is accessible by passing a needle into the sacral foramen via the posterior opening at the level of the nerve. The sacral canal is the caudal extension of the epidural space and nerves of the cauda equina leave via the sacral foramina (**Fig. 4.5.1**). The distal dural sac ends at S2, the level of the posterior superior iliac spines. The epidural space ends at the sacral hiatus (**Fig. 4.5.2**). While variability in the bony anatomy of the sacrum is common, this occurs usually in the midline.

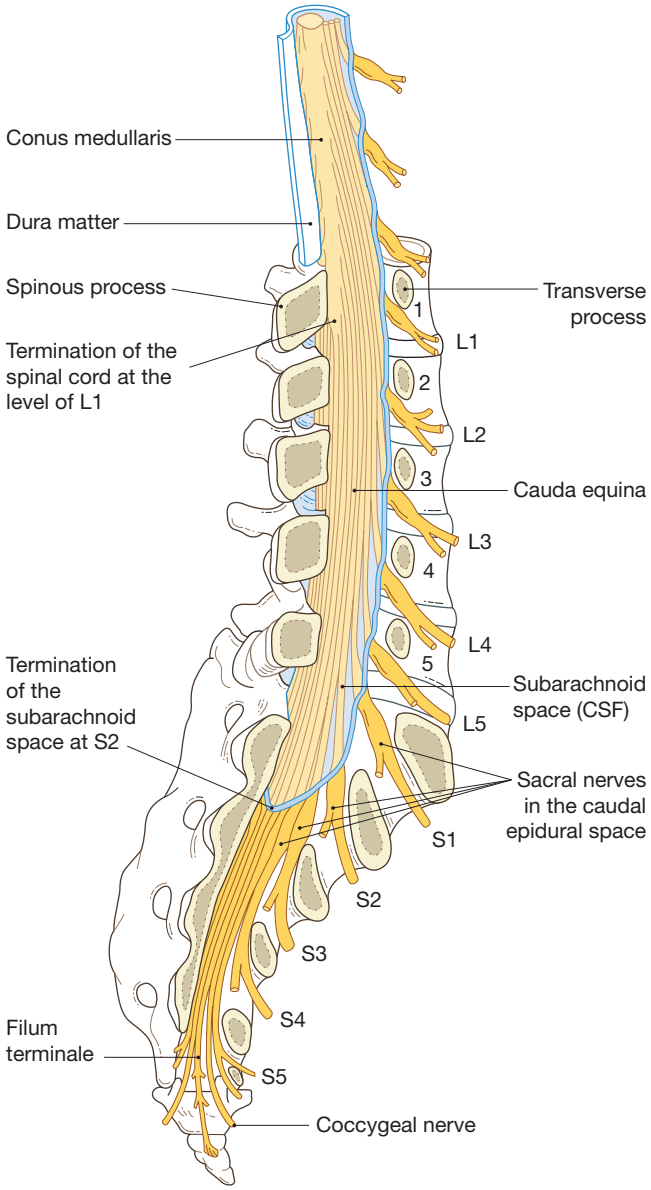

Fig. 4.5.1

Equipment

- 2 ml and 10 ml syringes
- 25 G needle
- 22 G spinal needle, end-opening
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*) and drugs
- Fluoroscopy or ultrasound

Drugs

- Lidocaine (lignocaine) 1% 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 50 mg (or its equivalent)
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 4.5.3*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The sacral midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.

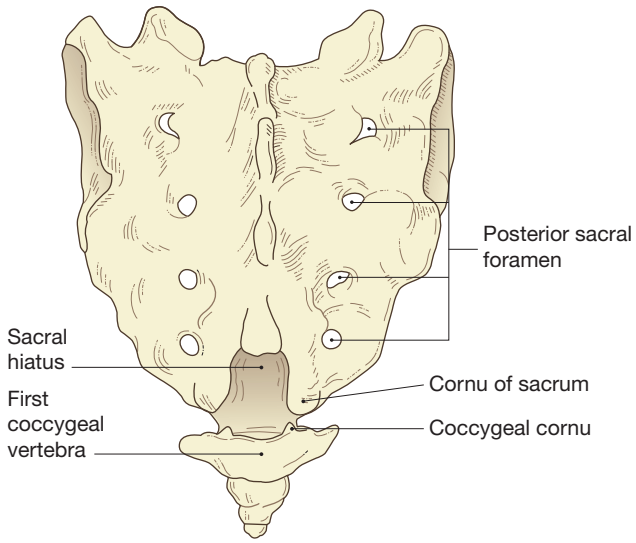

Fig. 4.5.2

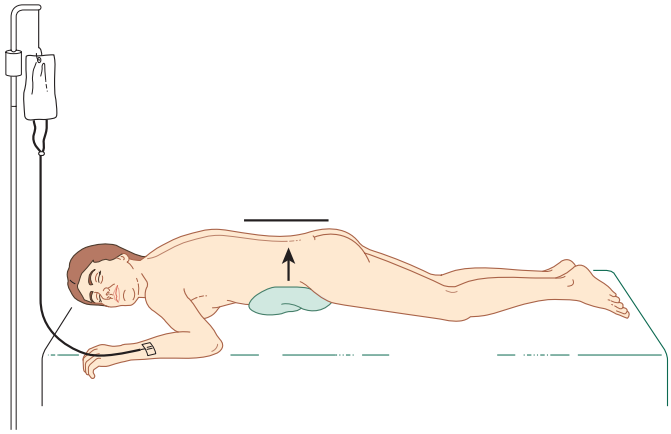

Fig. 4.5.3

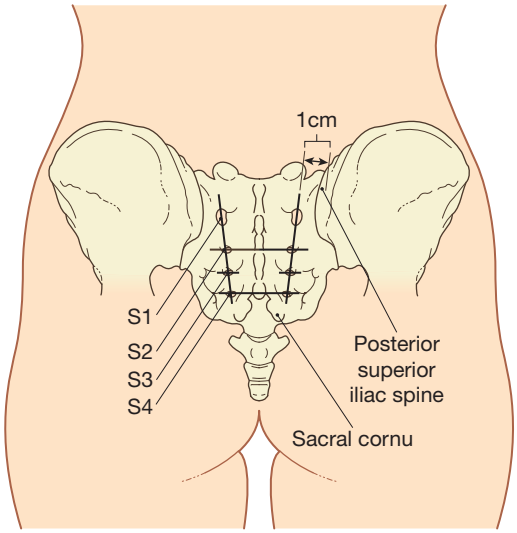

Fig. 4.5.4

- Iliac crests are palpated and an intercrestal line is identified.
- This corresponds with the inferior aspect of the spinous process of L4 or may lie in the L4–5 interspace.
- For this block, it is almost essential that the posterior opening of the sacral foramina is identified with fluoroscopy. The X-ray beam is used to guide the needle tip into the foramen. It is important to note that when the anteroposterior X-ray view is used, it is usually the anterior opening of the foramen that is the most prominent.
- **The insertion point of the needle lies 2–3 cm lateral to the midline (variable) and approximately 1 cm medial to the posterior iliac spine** (*Figs 4.5.4, 4.5.5 a,b*).
- Therefore, with the aid of fluoroscopy, the insertion point is identified.
- A skin wheal is raised and the area is infiltrated with lidocaine (lignocaine) 1%.

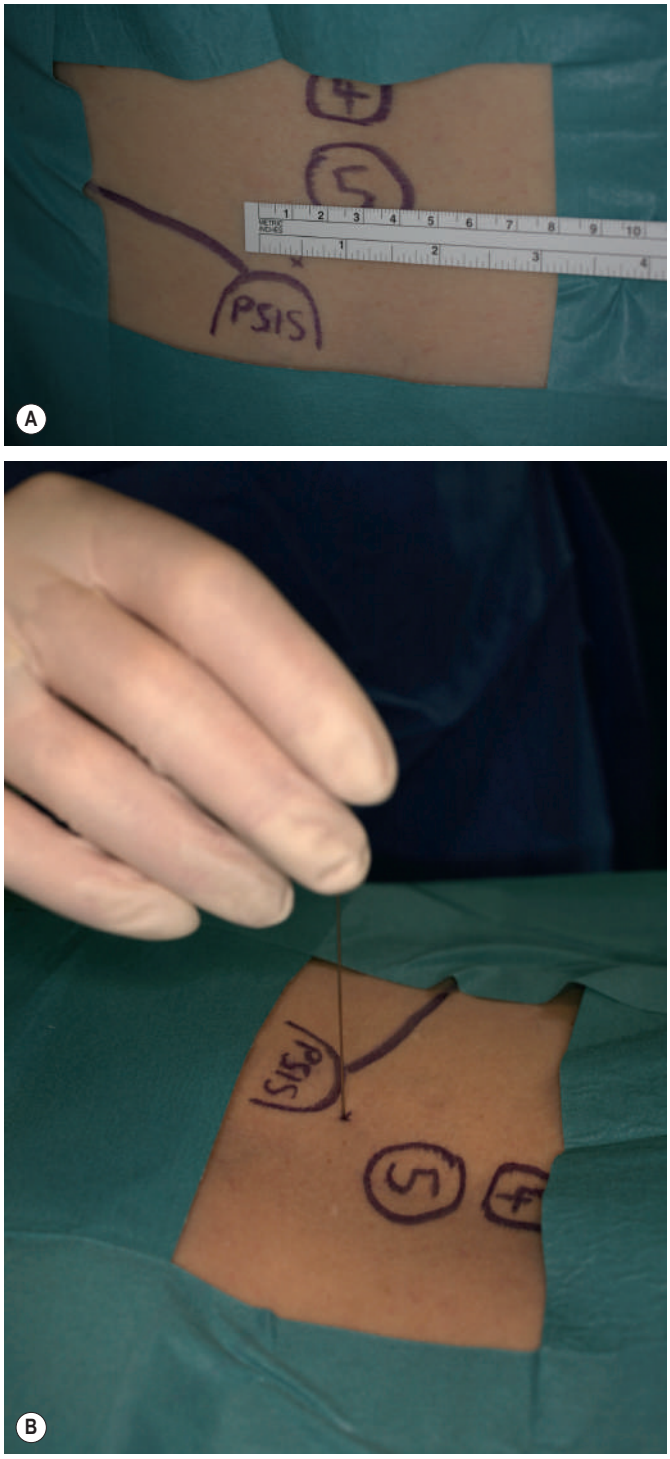

Fig. 4.5.5

- A spinal needle is introduced in a vertical direction to the skin, and aimed slightly cephalad until bone is contacted (*Fig. 4.5.6*).
- The needle is then walked off the sacrum in a caudad direction until it slips into the foramen. After confirmation with the aid of fluoroscopy it is then advanced 1 cm.

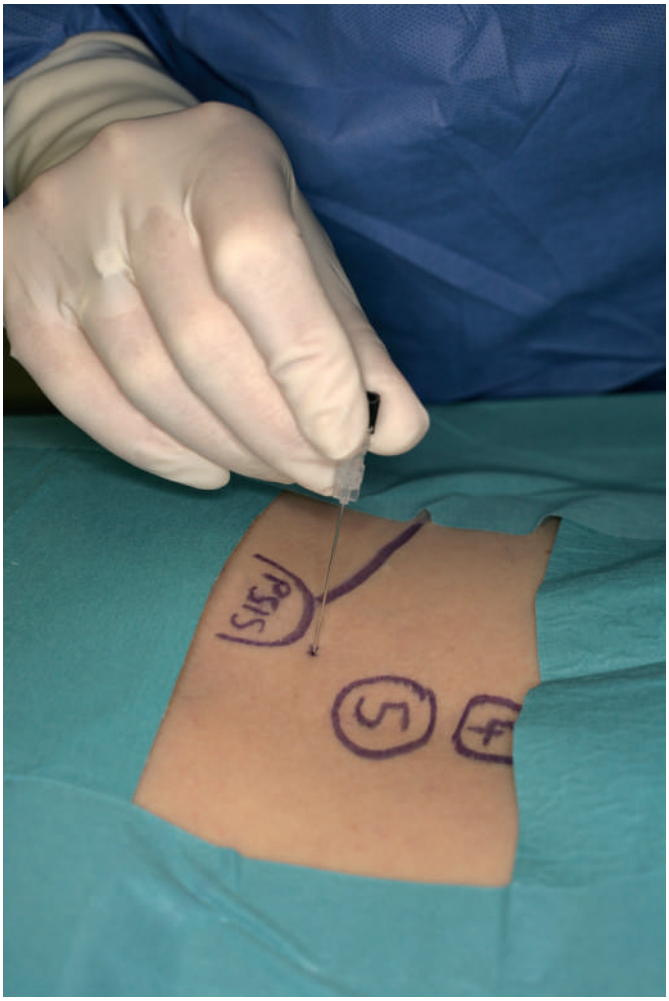

Fig. 4.5.6

- Paresthesia may be produced.
- After aspiration, non-ionic radio-opaque contrast medium 0.5 ml is injected.
- The correct placement is indicated by a needle tip flush with the anterior surface of the spinal canal in the lateral fluoroscopic view.
- Injection of 0.5 ml non-ionic contrast should spread diagonally along the S1 spinal nerve (*Fig. 4.5.7*).
- After further aspiration, lidocaine (lignocaine) 1% 0.5 ml is injected.
- After 5 minutes, the patient is questioned about changes in pain, sensation and power of the lower limb.
- For diagnostic nerve root blockade the needle may be removed when the level causing pain is identified.
- Ultrasound may also aid placement.

Confirmation of a successful block

- Relief of pain.
- Anesthesia in the distribution of the blocked nerve.

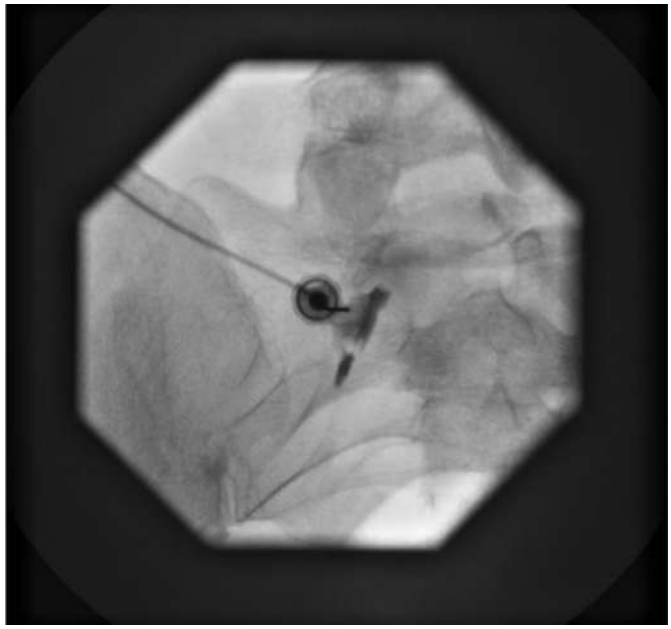

Fig. 4.5.7

Tips

- Some workers advocate drawing a line from a point 2–3 cm medial to the posterior superior iliac spine to a point 1–2 cm lateral to the sacral cornua. The sacral foramina usually lie along this line.
- It is best to angle the X-ray beam caudally, thereby perpendicular to the sacrum, superimposing the anterior and posterior sacral foramina. Consequently, when the needle is introduced in a direction perpendicular (*Fig. 4.5.8*) to the skin, fluoroscopic guidance is easier. Optimally, the needle makes gentle contact with the spinal nerve in the middle of the canal (*Figs 4.5.9*, *4.5.10*) and produces paresthesia in the distribution of the nerve.
- While all sacral nerve roots are accessible, using this technique for blockade of S5 is achieved by walking the needle caudally off the inferior edge of the bony plate of the sacrum and advancing the needle 1 cm.

Potential problems

- Caudal epidural blockade may occur with this block, but this is not a problem once low volumes are used. However, even small volumes of epidural spread may confound the diagnostic value of the block.
- Intravascular injection. Injection of particulate steroids into a radicular artery can cause spinal cord infarction. Particulate steroids should never be injected near the foramen unless intravascular placement has been ruled out using live fluoroscopy contrast dye injection, preferably with digital subtraction technique.

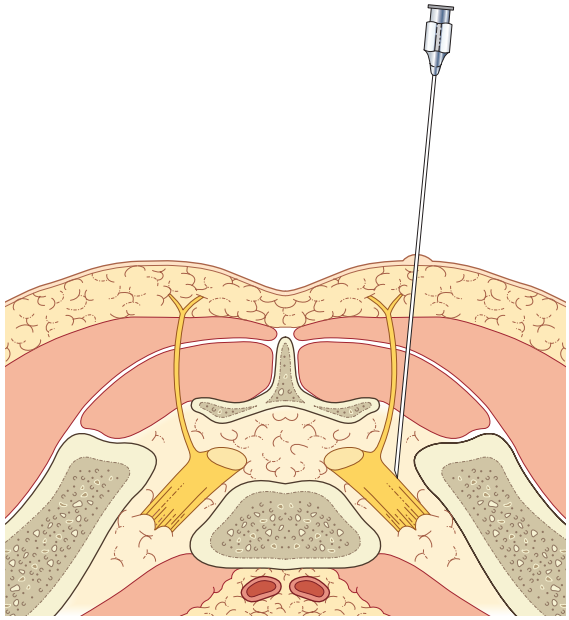

Fig. 4.5.8

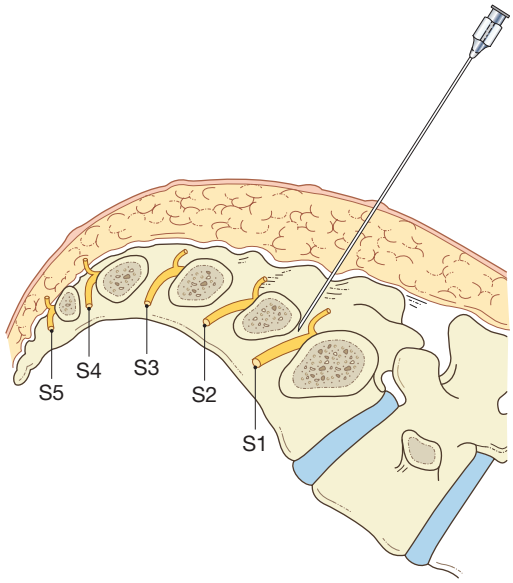

Fig. 4.5.9

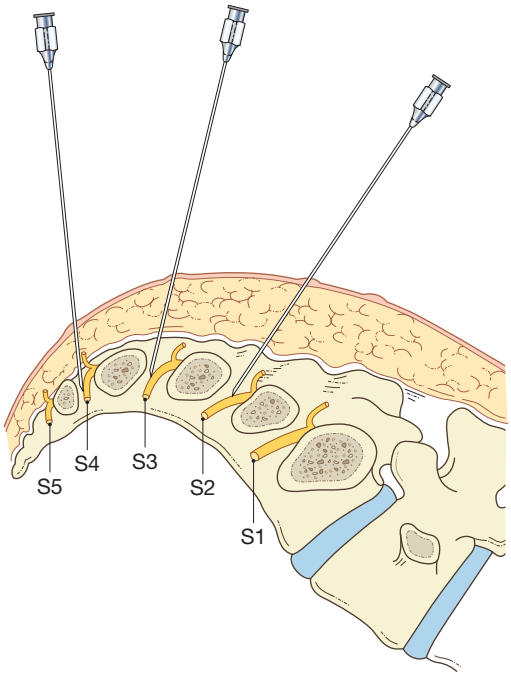

Fig. 4.5.10

4.6 OCCIPITAL NERVE BLOCK

Anatomy

The greater occipital nerve arises from the dorsal rami of the second cervical nerve. From here it passes through the muscles of the neck and becomes subcutaneous at the superior nuchal line, where it emerges with the occipital artery (Fig. 4.6.1). The superior nuchal line extends from the mastoid process to the greater occipital protuberance bilaterally (Fig. 4.6.2).

Equipment

- 10 ml syringe
- 25 G needle

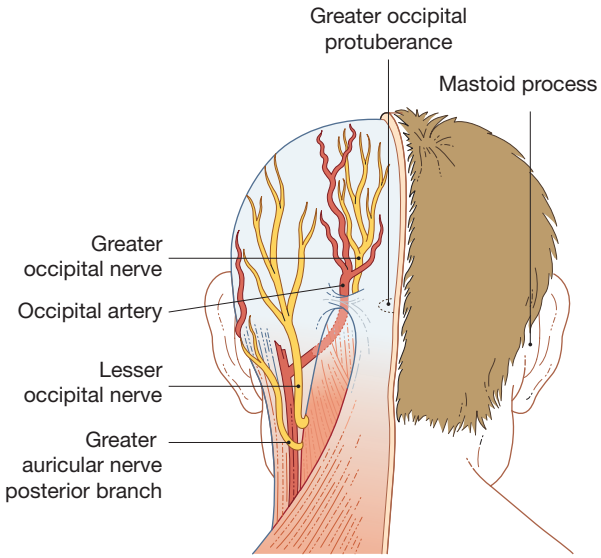

Fig. 4.6.1

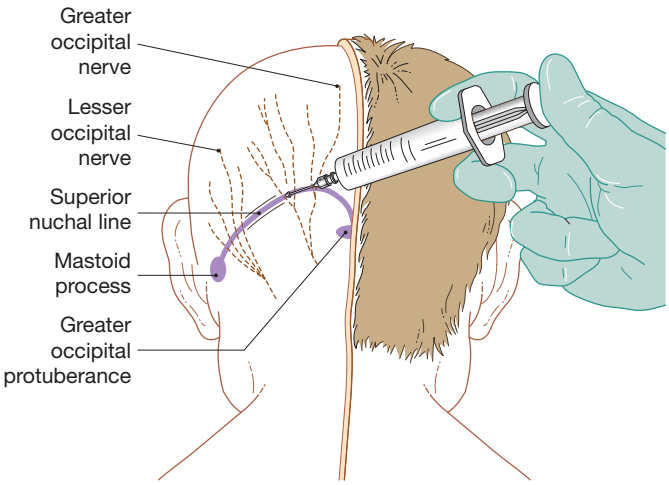

Fig. 4.6.2

- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)

Drugs

- Lidocaine (lignocaine) 1% 10 ml (or its equivalent)
- Corticosteroid if indicated, e.g. triamcinolone diacetate 50 mg (or its equivalent)
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Sitting.
- Neck flexed.

Needle puncture and technique

- The superior nuchal line is cleaned with antiseptic solution (no drape is required) (*Fig. 4.6.3*).

Fig. 4.6.3

- The occipital artery is palpated 2 cm lateral to the greater occipital protuberance on the superior nuchal line (*Fig. 4.6.4*).
- A 25 G needle is inserted subcutaneously at this point.
- After negative aspiration, 3–5 ml of lidocaine (lignocaine) 1% or its equivalent, plus corticosteroid if indicated, is injected to surround the occipital artery.
- It is often difficult to feel an occipital artery pulse. If this is the case, it is best to pick a point midway between the occipital protuberance and the mastoid bone and fan out the injection in both directions, medially and laterally from that site (*Fig. 4.6.5*).

Confirmation of a successful block

- Relief of pain and anesthesia in distribution of nerve.

Tips

- Bone should be contacted at a depth no greater than 1–2 cm.
- The lesser occipital nerve is blocked by redirecting the needle towards the mastoid process along the greater nuchal line and injecting a further 3 ml of solution.

Potential problems

- Injection into the cerebrospinal fluid (CSF) of the cisterna magna is possible and will produce a total spinal block.

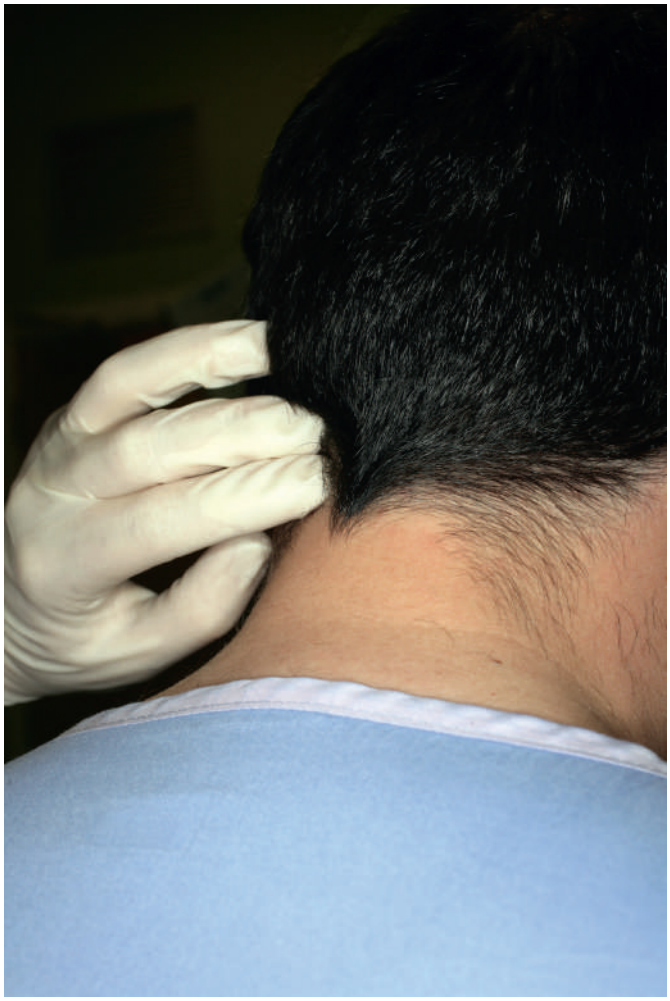

Fig. 4.6.4

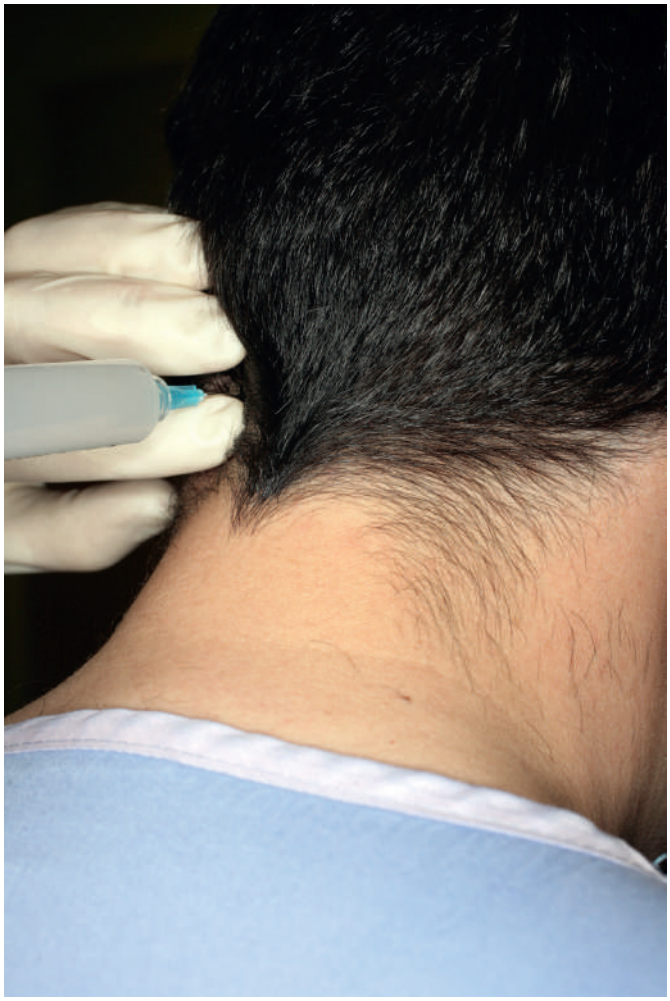

Fig. 4.6.5

4.7 TRIGEMINAL GANGLION (GASSERIAN) BLOCK

Anatomy

The trigeminal ganglion gives rise to the fifth cranial nerve and divides into three branches, the ophthalmic, maxillary, and mandibular nerves (Fig. 4.7.1). These provide the sensory nerve supply to the ipsilateral face and the anterior two-thirds of the head (Fig. 4.7.2). The mandibular nerve also provides motor supply to the muscles of mastication.

The trigeminal ganglion is located at the apex of the petrous temporal bone in a fold of dura mater, "Meckel's cave". This dural invagination covers the posterior two-thirds of the ganglion and contains cerebrospinal fluid (CSF). Posterior to Meckel's cave lies the brainstem, superior to it lies the temporal lobe, and medially lies the cavernous sinus which contains the internal carotid artery and cranial nerves III, IV, and VI. Accordingly, extreme care must be taken when carrying out this block, especially if neurolytic agents are used. Blockade of the ganglion is carried out by passage of a needle through the foramen ovale, which lies immediately below it (Fig. 4.7.3).

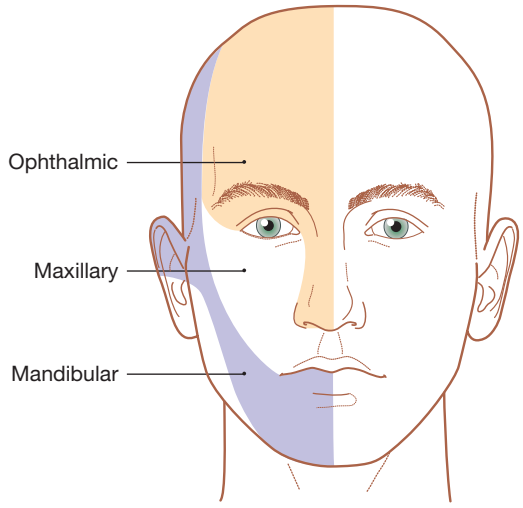

Fig. 4.7.1

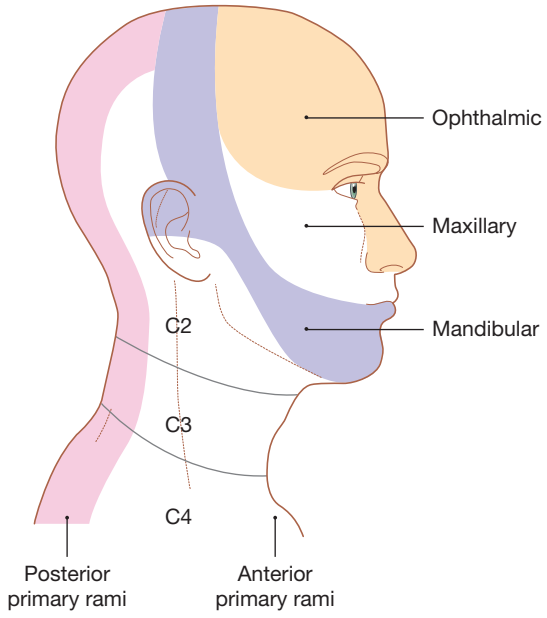

Fig. 4.7.2

Equipment

- 2 ml and 10 ml syringes
- 25 G needle
- 22 G 8–10 cm needle
- Non-ionic radio-opaque contrast medium
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy

Drugs

- Lidocaine (lignocaine) 2% 10 ml
- Lidocaine (lignocaine) 1% 10 ml (or its equivalent)
- Neurolytic agent, e.g. phenol 6% plus glycerol (or its equivalent)
- Sedative, e.g. midazolam, propofol
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Supine.
- Eyes directed straight ahead.

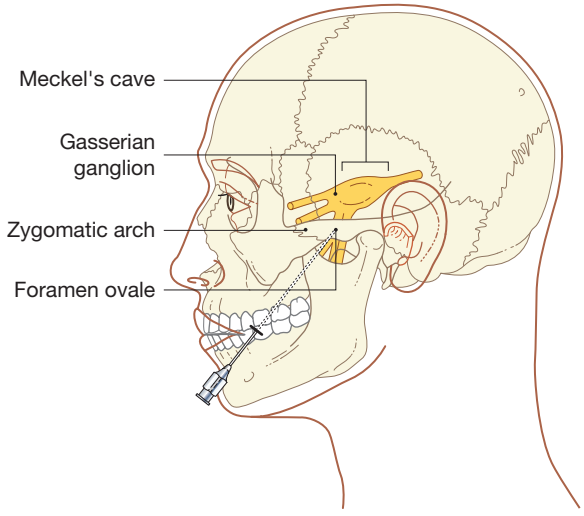

Fig. 4.7.3

Needle puncture and technique

Caution: Injection of 0.25 ml of lidocaine (lignocaine) 1% into the CSF may result in immediate convulsion and/or loss of consciousness with possible cardiovascular system (CVS) collapse.

- The cheek on the side of the block is cleaned with antiseptic solution or saline.
- Mild sedation is induced.
- It is best to stand on the side of the block, just below the shoulder.
- **The insertion point lies 1–3 cm posterior to the lateral margin of the mouth, at the medial edge of the masseter muscle (located by asking the patient to clench the jaw) and is marked.**
- **In edentulous patients the insertion point should be more caudad** as sufficient angle towards the infratemporal surface of the sphenoid bone may not be achieved.
- One finger is placed inside the upper lip to avoid injection into the buccal cavity and possible bacterial contamination, and a skin wheal is raised at this site.
- Viewed from above, a 22 G 8–10 cm needle is advanced towards the ipsilateral pupil (*Figs 4.7.4–4.7.6*), or viewed from the side the needle advances towards the mid-point of the zygomatic arch (see Anatomy above) until bone is contacted; the roof of the infratemporal fossa. This lies just anterior to the foramen ovale and lateral to the base of the pterygoid process. The location of the needle tip is confirmed with fluoroscopy.
- A depth mark is set and the needle is withdrawn to the subcutaneous tissue. With the aid of fluoroscopy the needle is reinserted to walk off the bone and enter the foramen ovale (*Fig. 4.7.7*).
- Paresthesia in the distribution of the mandibular nerve (sometimes the other branches of the trigeminal nerve) or contraction of the muscles of mastication may be experienced at this point.
- The needle is advanced a further 1 cm to bring the tip to lie in the trigeminal ganglion. The correct placement is indicated by a visible outline of Meckel's cave on injection of 0.25 ml non-ionic radio-opaque contrast medium under fluoroscopy (*Figs 4.7.8*, *4.7.9*).
- The patient is allowed to awaken from sedation and is questioned about the presence and distribution of paresthesia and pain.
- A stimulating device may aid placement in patients who are not able to locate the paresthesia with accuracy.
- If necessary, analgesia may be administered, although this may affect accurate assessment of the blockade.
- Adjustment of the needle may be required to place the needle near the appropriate nerve division.

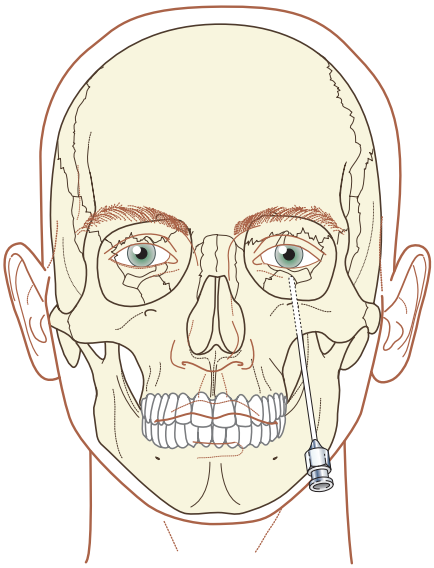

Fig. 4.7.4

Fig. 4.7.5

Fig. 4.7.6

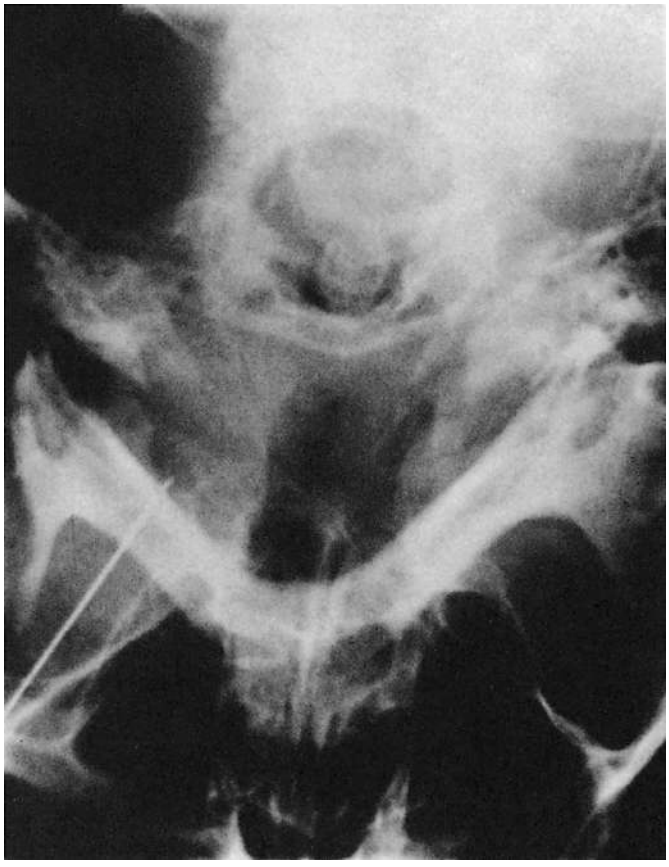

Fig. 4.7.8

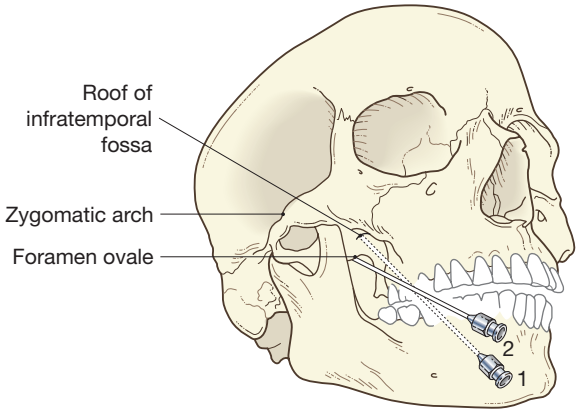

Fig. 4.7.7

Fig. 4.7.9

- After careful negative aspiration for CSF or blood, lidocaine (lignocaine) 1% 0.25 ml is injected (***Caution: injection into CSF may cause loss of consciousness***.) This is followed by further boluses of lidocaine (lignocaine) 1% 0.25 ml until a total of 1 ml is given.
- After 5 minutes the patient is questioned about pain relief and changes in sensation.
- When the desired analgesia has been achieved for diagnostic blockade the needle may be removed.

Confirmation of a successful block

- Relief of pain and anesthesia in the distribution of the trigeminal nerve or its desired branches.

Tips

- Intravenous anesthesia using a short-acting agent, e.g. propofol, may be induced to allow placement of the needle. The patient is then allowed to awaken and a stimulating device may aid accurate placement of the needle.
- The needle is advanced beyond the infratemporal bone by 0.5 cm for location of the mandibular division, 1.0 cm for the maxillary division, and 1.5 cm for the ophthalmic division.
- Gangliolysis using thermocoagulation may be employed for trigeminal-nerve division destruction after location of the ganglion using this technique. Further intravenous anesthesia using a short-acting agent, e.g. propofol, may be induced after placement of the insulated needle to facilitate this painful procedure.
- Injection of glycerol alone may produce pain relief with this injection technique. This involves placement of the needle in the cul-de-sac of CSF, positioning the patient face-down or supine, with the head extended to prevent spill-over into the posterior cranial fossa. After entry into the CSF, and positive aspiration of CSF, 0.1–0.3 ml of glycerol may be injected.

Potential problems

- Injection of 0.25 ml of lidocaine (lignocaine) 1% into the CSF may result in immediate convulsion and/or loss of consciousness with possible CVS collapse.
- Spread of hyperbaric neurolytic solution may immediately affect cranial nerves VI, VIII, IX, X, XI, and XII.
- Spread of hypobaric neurolytic solution may immediately affect the oculomotor and trochlear nerves.
- Neurolytic blocks of the trigeminal ganglion commonly produce corneal and hemifacial anesthesia.

AUTONOMIC BLOCKADE 5

Autonomic blockade is useful in the diagnosis and treatment of pain of autonomic origin. In cases of thoracic, abdominal or pelvic pain, it is often difficult to distinguish between that of visceral origin and that of somatic. Pain of visceral origin, e.g. pancreatic cancer or pelvic cancer, may cause pain that responds to celiac or hypogastric plexus blockade, respectively. A prognostic block may be carried out prior to neurolytic blockade for relief of cancer pain. Pain of the upper abdominal viscera may also be relieved by celiac plexus block, proceeding to neurolytic blockade as appropriate for cancer-related pain. The retrocrural approach to the celiac plexus also targets the splanchnic nerves to produce a splanchnic nerve block if required. Chest pain may be of somatic origin, e.g. intercostal neuralgia and costochondritis, or visceral origin, e.g. pulmonary or cardiac-related pain. Stellate ganglion blockade may be helpful in the diagnosis and management of the latter.

In addition, increased sympathetic activity is thought to contribute to a large number of pain states. These are generally grouped under the term Complex Regional Pain Syndrome Type I and II. In these cases, trophic changes and alterations in blood flow are often obvious but the pathophysiologic origin is not. Blockade of sympathetic innervation may therefore help in diagnosis and management of such pain. This may also indicate other therapies that could be beneficial, e.g. sympathetically active drugs or destructive therapies. Similarly, these therapies would not be indicated if sympathetic blockade failed to relieve the pain. If blockade did succeed in relieving this type of pain, further blocks may effect lasting relief.

There are a number of other conditions in which the diagnosis is clear but there is a possible contribution of sympathetic activity in the pathogenesis of the pain. These conditions include central pain, post-herpetic neuralgia, trigeminal neuralgia, peripheral vascular disease and others. Blockade of sympathetic activity may help to clarify the sympathetic contribution to the pain and therefore help to indicate management options.

5.1 STELLATE GANGLION BLOCK—C6 (CLASSIC) APPROACH

Anatomy

The cervical sympathetic trunk—the superior, middle, and stellate ganglia—supplies the sympathetic innervation of the head, neck, and upper limbs. The stellate ganglion is made up of a combination of the lower cervical and first thoracic ganglia. It lies on the prevertebral fascia of the seventh cervical and first thoracic vertebrae (**Fig. 5.1.1**). However, as the sixth cervical anterior tubercle (Chassaignac's tubercle) is easy to palpate, injection of a large volume of local anesthetic is made at this level and allowed to track caudally along the prevertebral fascia to block the stellate ganglion. The vertebral and carotid arteries, the pleura, and the brachial plexus are in close proximity to the stellate ganglion.

Equipment

- 10 ml syringe
- 22 G short-bevel needle
- Extension set (optional)
- ECG, BP, SpO_2 monitors
- Skin temperature monitor
- Resuscitation equipment (*see Appendix 3*)
- Ultrasound (optional)

Drugs

- Lidocaine (lignocaine) 1%, 10 ml
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Supine.
- Thin pillow under head.
- Roll under neck.
- Eyes directed at ceiling.
- Mouth slightly open.

Needle puncture and technique

Caution: Injection of 0.5–1 ml of lidocaine (lignocaine) 1% into the vertebral artery may result in immediate convulsion and/or loss of consciousness with possible cardiovascular system (CVS) collapse.

- Intravenous access is inserted.
- Monitors are attached, each temperature probe is attached to the palmar aspect of the middle finger of each hand (*Fig. 5.1.2*).
- Resuscitation equipment and drugs are checked and made ready for use.
- The side of the neck is cleaned with antiseptic solution.
- It is best to stand at the same side of the neck as the ganglion to be blocked.

FOR THE RIGHT-HANDED OPERATOR

With the left hand

- The thyroid cartilage is located and marked.
- The cricoid cartilage is identified and marked (*Figs 5.1.3, 5.1.4*).

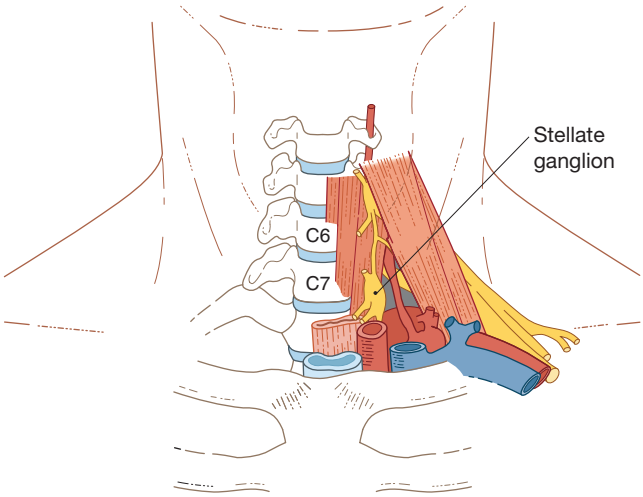

Fig. 5.1.1

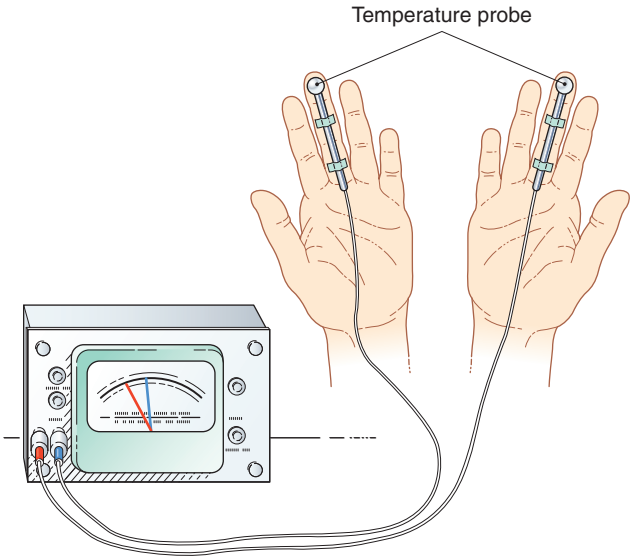

Fig. 5.1.2

Fig. 5.1.3

Fig. 5.1.6

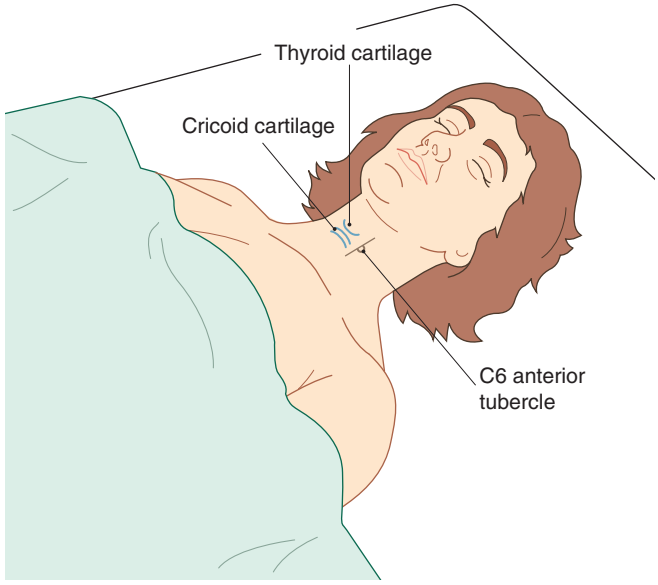

Fig. 5.1.4

Fig. 5.1.7

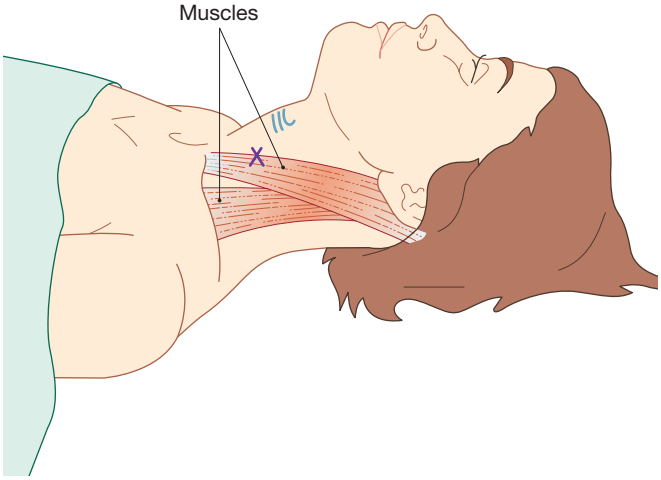

Fig. 5.1.5

- Chassaignac's tubercle is palpated with the middle finger, just lateral to the cricoid cartilage (*Fig. 5.1.5*).
- The sternocleidomastoid (SCM) muscle is gently pulled laterally and the carotid pulse is palpated (*Fig. 5.1.6*).
- Chassaignac's tubercle is palpated again and positioned between the fore- and middle fingers.

With the right hand

- The needle is inserted between the fore- and middle fingers of the other hand, directly perpendicular to the floor, aiming for Chassaignac's tubercle (*Figs 5.1.7, 5.1.8*).
- When contact with the tubercle is reached, the injecting hand is steadied and the needle is withdrawn 2 mm.
- The hub of the needle is held in place with the other hand.

Fig. 5.1.8

Fig. 5.1.9

- After negative aspiration, lidocaine (lignocaine) 1%, 0.5 ml, is injected.
- The patient is questioned about sensation and any change in level of consciousness is noted.
- If negative, the same procedure is repeated as 0.5 ml boluses are given until 10 ml is injected.
- The needle is withdrawn and the patient is immediately put in the sitting position.
- Monitors should be left attached and i.v. access left in situ for at least 30 minutes. The patient is requested not to eat or drink for 2 hours, as the recurrent laryngeal nerve may be blocked.
- Note: if aspiration of blood occurs during the block the needle is removed and cleared, keeping left fore- and middle fingers in place. It is then reinserted and the block is continued as above.
- If hematoma occurs before the solution is injected it may be worth performing the block at the C7 level.
- If there is pain on injection and/or paresthesia, it is likely that the brachial plexus may have been contacted, the needle is withdrawn and the landmarks are rechecked.

Confirmation of a successful block

- Skin temperature, measured over the palmar aspect of the hand or fingers on the blocked side, should begin to rise within 2–3 minutes. Extensive sympathetic blockade is confirmed by a rise in skin temperature to >33 °C.
- Ptosis of eyelid.
- Miosis of pupil.
- Unilateral blockage of nose on side of block.
- Unilateral flushing of conjunctiva of eye on side of block.
- Blockade of the upper sympathetic chain can occur in the absence of sympathetic denervation of the upper extremity, resulting in Horner's syndrome without a rise in skin temperature in the hand.

Tips

- The external jugular vein usually crosses the SCM muscle at the level of C6.
- Skin infiltration prior to block should be avoided if possible, as this will make landmarks more difficult to locate.
- If palpation is painful or difficult it may be helpful to try to bounce the middle finger off the tubercle during identification.
- An extension set may be inserted between the needle and syringe for better stability of needle (*Fig. 5.1.9*), but an assistant is then required to continue the procedure as described above.
- Consideration should be given to performance of the block under fluoroscopy or CT control if the landmarks are difficult to locate.
- Lidocaine (lignocaine) 1% 15 ml may be given if a previous block failed to relieve sympathetically maintained pain in the presence of correctly placed solution. This may improve tracking of the solution caudally to produce more effective blockade of the stellate ganglion.
- Ultrasound may aid placement of needle (*Fig. 5.1.10 a,b*).

Potential problems

- Intra-arterial injection or intrathecal injection may result in immediate convulsion and/or loss of consciousness with possible CVS collapse.

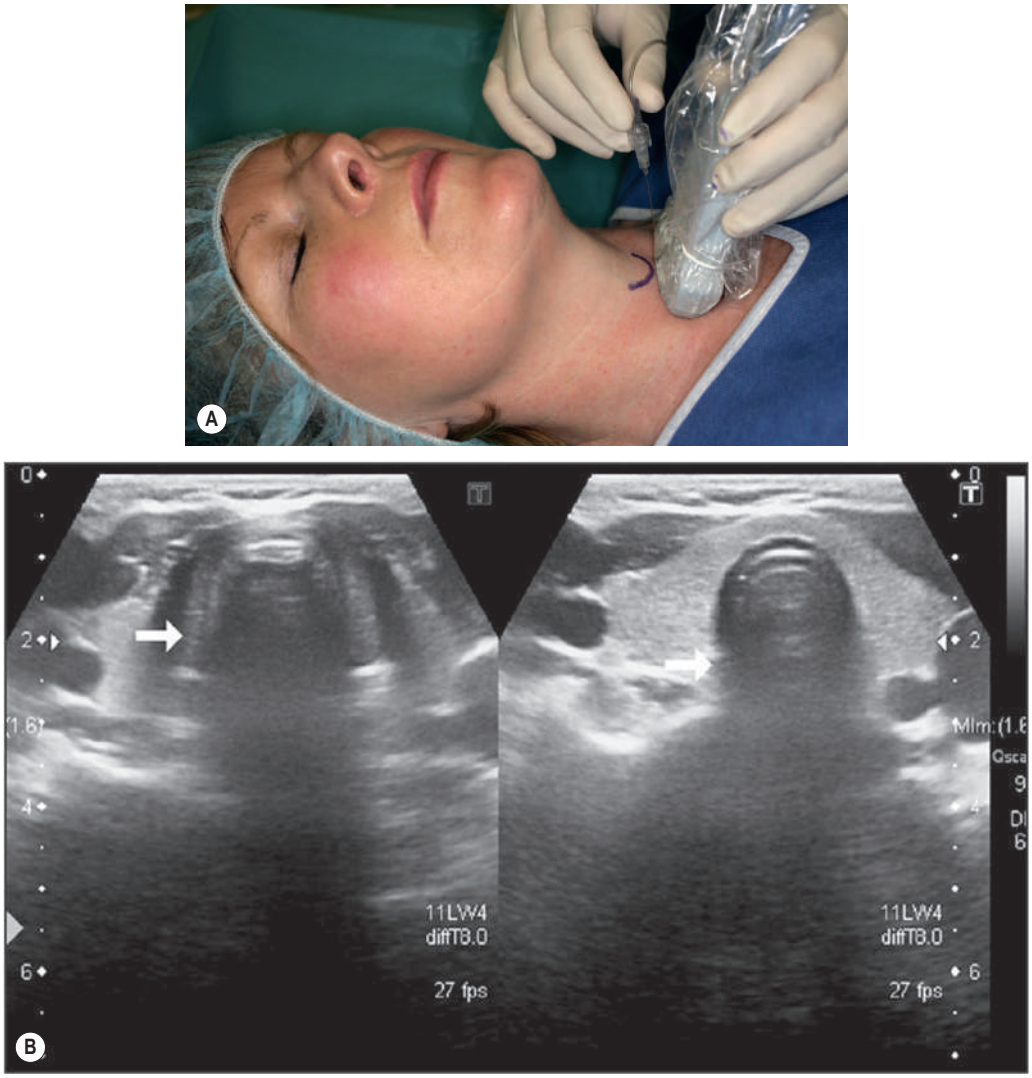

Fig. 5.1.10 From Gupta Prashant K, Gupta Kumkum, Dwivedi Amit Nandan D, Jain Manish. Potential role of ultrasound in anesthesia and intensive care, Anesthesia Essays and Research, 2011 Volume 5, Issue Number 1, Page: 11-19.

- Hematoma may occur (avoid performing block on patients who have coagulopathy).
- Pneumothorax may occur.
- Recurrent laryngeal nerve block may occur and it is prudent to advise the patient about possible hoarseness post-blockade. Bilateral stellate ganglion blockade should be avoided for the same reason.
- Phrenic nerve block may occur and it is prudent to caution the patient about possible shortness of breath post blockade of the stellate ganglion.

5.2 STELLATE GANGLION BLOCK—C7 APPROACH

Anatomy

As for stellate ganglion block (C6 approach) in Section 5.1 (**Figs 5.2.1 a,b**).

Equipment

- 10 ml syringe
- 22 G short-bevel needle
- Extension set (optional)
- ECG, BP, SpO_2 monitors
- Skin temperature monitor
- Fluoroscopy or CT (optional)
- Resuscitation equipment (*see Appendix 3*)
- Ultrasound (optional)

Drugs

- Lidocaine (lignocaine) 1%, 10 ml
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Supine.
- Thin pillow under head.
- Roll under neck.
- Eyes directed at ceiling.
- Mouth slightly open.

Needle puncture and technique

Caution: injection of 0.5–1 ml of lidocaine (lignocaine) 1% into the vertebral artery may result in immediate convulsion and/or loss of consciousness with possible cardiovascular (CVS) collapse. The risk of pneumothorax is greater with this approach.

- Intravenous access is inserted.
- Monitors are attached.
- Each temperature probe is attached to the palmar aspect of the middle finger of each hand. (*Fig. 5.2.2*).
- Resuscitation equipment and drugs are checked and made ready for use.
- The side of the neck is cleaned with antiseptic solution.

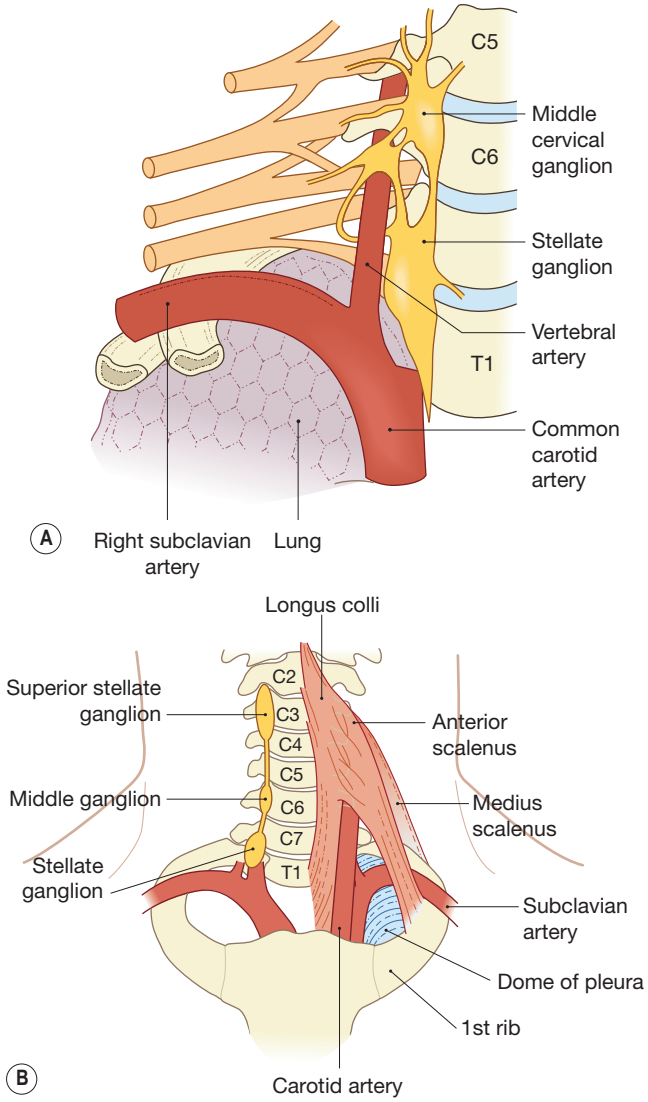

Fig. 5.2.1

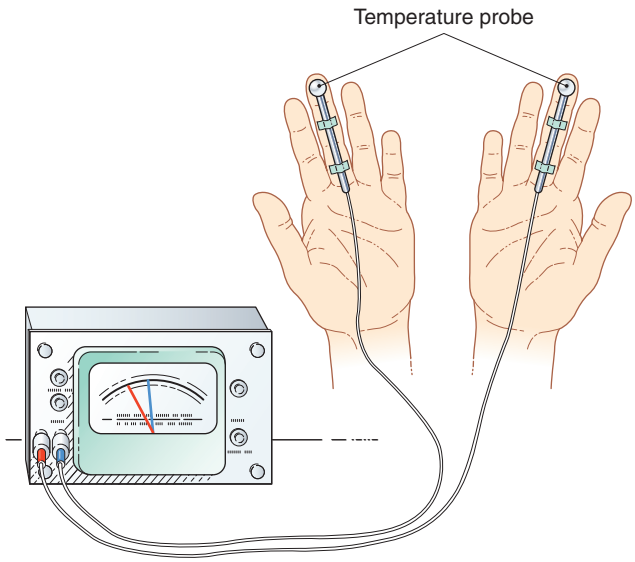

Fig. 5.2.2

- It is best to stand at the same side of the neck as the ganglion to be blocked.

FOR THE RIGHT-HANDED OPERATOR

With the left hand

- The thyroid cartilage is located and marked.
- The cricoid cartilage is identified and marked (*Figs 5.2.3, 5.2.4*).
- The sternoclavicular junction is palpated and marked (*Fig. 5.2.5*).
- The SCM muscle is gently pulled laterally and the carotid pulse is palpated (*Fig. 5.2.6*).
- **The site of insertion of the needle lies 3 cm above the sternoclavicular junction or one to two finger-breadths below the level of the cricoid cartilage.**

Fig. 5.2.3

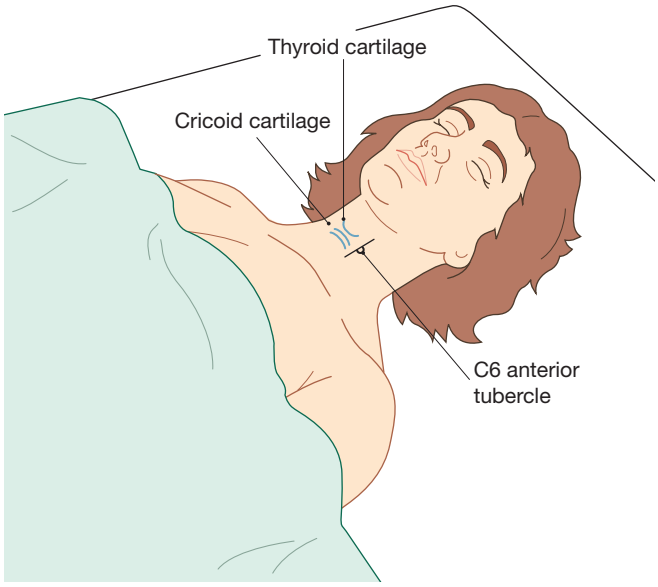

Fig. 5.2.4

With the right hand

- The patient is requested to exhale deeply before needle insertion to minimize the risk of pneumothorax.
- The needle is inserted between the fore- and middle fingers of the other hand, directly perpendicular to the floor.
- When contact with the transverse process of C7 is reached, the injecting hand is steadied and the needle is withdrawn 2 mm.
- The hub of the needle is held in place with the other hand.
- After negative aspiration, lidocaine (lignocaine) 1%, 0.5 ml, is injected.
- The patient is questioned about sensation, and any change in level of consciousness is noted.
- If negative, the same procedure is repeated and 0.5 ml boluses are given until 5–8 ml is injected.
- The needle is withdrawn and the patient is immediately put in the sitting position.

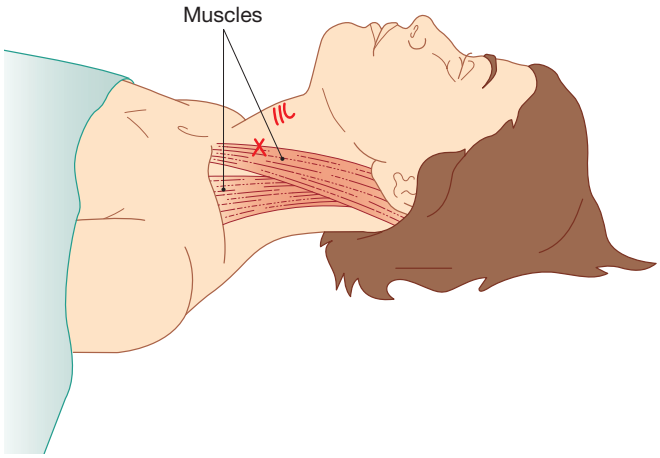

Fig. 5.2.5

Fig. 5.2.6

- Monitors should be left attached and i.v. access left in situ for at least 30 minutes. Blockade of the recurrent laryngeal nerve is less likely with the C7 approach but it is wise to advise the patient not to eat or drink for 2 hours.
- Note: if aspiration of blood occurs during the block, the needle is removed and cleared, keeping the left fore- and middle fingers in place. It is then reinserted and the block is continued as above.
- If there is pain on injection and/or paresthesia, it is likely that the brachial plexus may have been contacted, the needle is withdrawn and the landmarks are rechecked.
- Ultrasound may aid placement of the needle.

Confirmation of a successful block

- Temperature increase >1° on the side of block. The temperature should begin to rise in the finger of the blocked side within 3 minutes of injection.
- Ptosis of eyelid.
- Miosis of pupil.
- Unilateral blockage of nose on side of block.
- Unilateral flushing of conjunctiva of eye on side of block.
- Relief of sympathetically maintained pain.

Tips

- Some workers advocate targeting the **ventrolateral aspect of the C7 vertebral body instead of its transverse process.** The needle is directed 15–20° medially. With the aid of fluoroscopy, ultrasound or CT, the vertebral body is contacted just medial to the insertion of the longus colli muscle. The needle is then withdrawn 2 mm, stabilized, and 1 ml of non-ionic contrast medium is injected.
- The external jugular vein usually crosses the SCM muscle at the level of C6.
- Skin infiltration prior to block should be avoided if possible, as this will make landmarks more difficult to locate.
- An extension set may be inserted between the needle and syringe for better stability of needle, but an assistant is then required to continue the procedure as described above.
- Lidocaine (lignocaine) 1% 10 ml may be given if a previous block failed to relieve sympathetically maintained pain in the presence of correctly placed solution. This may improve tracking of the solution caudally to produce more effective blockade of the stellate ganglion.

Potential problems

- Intra-arterial injection or intrathecal injection may result in immediate convulsion and/or loss of consciousness with possible CVS collapse.
- Hematoma may occur (avoid performing block on patients who have coagulopathy).
- Pneumothorax may occur (more likely with C7 approach).
- Recurrent laryngeal nerve block may occur and it is prudent to advise the patient about possible hoarseness post blockade. Bilateral stellate ganglion blockade should be avoided for the same reason.
- Phrenic nerve block may occur and it is prudent to caution the patient about possible shortness of breath following blockade of the stellate ganglion.

5.3 LUMBAR SYMPATHETIC BLOCK

Anatomy

The lumbar sympathetic chain is located in the prevertebral fascia, which lies on the anterolateral aspects of the vertebral bodies. The psoas muscle separates the lumbar sympathetic chain from the lumbar somatic nerves. A single injection of local anesthetic at the level of L2 will usually provide a complete block of postganglionic sympathetic efferents to the lower extremity because the lowest preganglionic sympathetic outflow to the chain is at the level of L2.

Equipment

- 2 ml, 5 ml, and 10 ml syringes
- 30 G needle
- Two 15 cm 22 G needles
- Extension set (optional)
- ECG, BP, SpO_2 monitors
- Skin temperature monitor (two probes)
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy

Drugs

- Lidocaine (lignocaine) 1%, 5 ml for skin infiltration
- Lidocaine (lignocaine) 1%, 15–20 ml (or its equivalent) for block
- Phenol 6%
- Radio-opaque contrast medium
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 5.3.1*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Each temperature probe is attached to the plantar aspect of the big toe (*Fig. 5.3.2*).
- Resuscitation equipment and drugs are checked and made ready for use.
- The thoracolumbar midline and an area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- The twelfth rib is identified and a line is drawn along its inferior border (*Fig. 5.3.3 a,b*).
- The iliac crests are palpated and the intercrestal line is identified (this corresponds with the inferior aspect of the spinous process of L4 or may lie in the L4–5 interspace).
- The spinous processes are counted until L2 is identified and confirmed with fluoroscopy.
- **The insertion point lies 8 cm lateral to the L2 spinous process and is also marked** (*Fig. 5.3.4*).
- A skin wheal is raised at one of the marked sites and the area is infiltrated with lidocaine (lignocaine) 1%.
- At a 30° angle to the frontal plane, a 22 G 15 cm needle is advanced slightly cephalad towards the lower

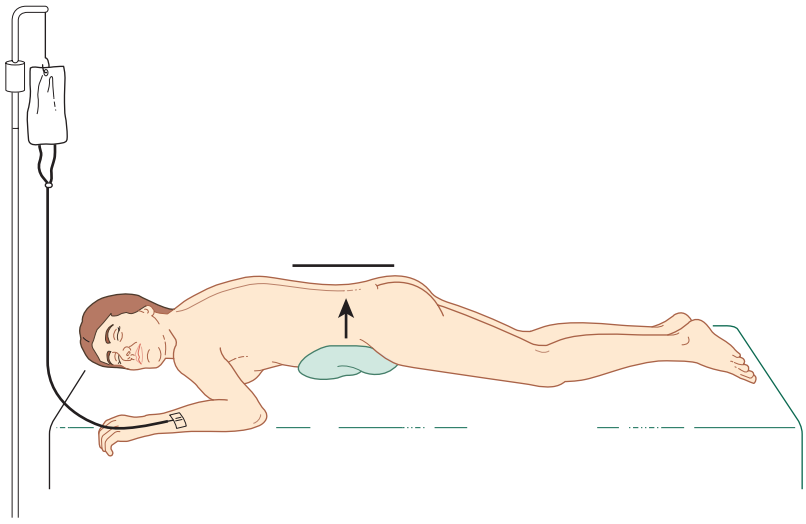

Fig. 5.3.1

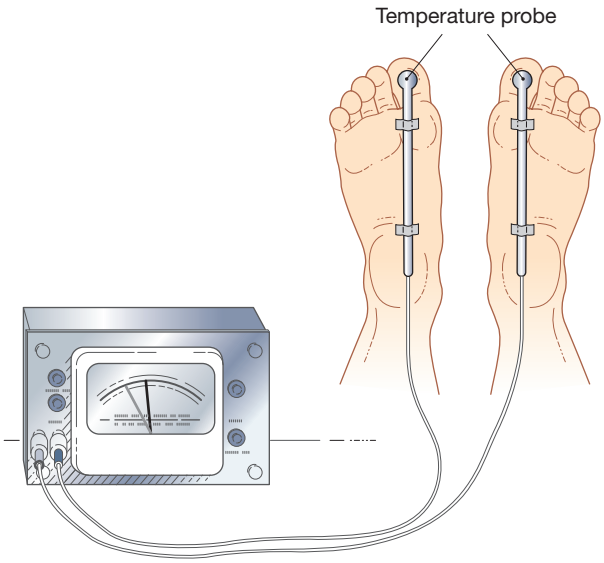

Fig. 5.3.2

portion of the L2 vertebral body (*Fig. 5.3.5*), until its vertebral body is contacted at a depth of about 7–9 cm and confirmed with fluoroscopy. If the needle contacts bone at a more superficial level, it is probable that it has come into contact with the transverse process and it will need to be repositioned.

- The needle depth is noted.
- The needle is then withdrawn to the subcutaneous tissue and, with the aid of fluoroscopy, it is re-advanced, this time at an angle 45° to the frontal plane, until the previous depth (as noted) is reached. It should slip past the vertebral body at a depth about 1–2 cm deeper than the first depth mark (*Fig. 5.3.6*).
- After negative aspiration, the fluoroscopic image is observed as a small amount of non-ionic radio-contrast medium is injected. The correct placement of the needle is indicated by the presence of a layer of contrast medium in a thin line along the anterior border of the vertebral column (*Figs 5.3.7–5.3.9*).
- After further negative aspirations, 5 ml of lidocaine (lignocaine) 1% is injected. The patient is questioned about pain relief and observations of skin temperature are made. There should be little resistance to injection, similar to resistance felt when injecting through an epidural needle. If resistance is encountered, or if the injection is painful, the needle should be repositioned. A unilateral rise in skin temperature indicates a successful block.
- After 10 minutes the patient is questioned about pain relief and any symptoms of somatic nerve blockade. Sensory and motor functioning of the lower extremities is checked. The procedure should be abandoned if there is evidence of somatic blockade.

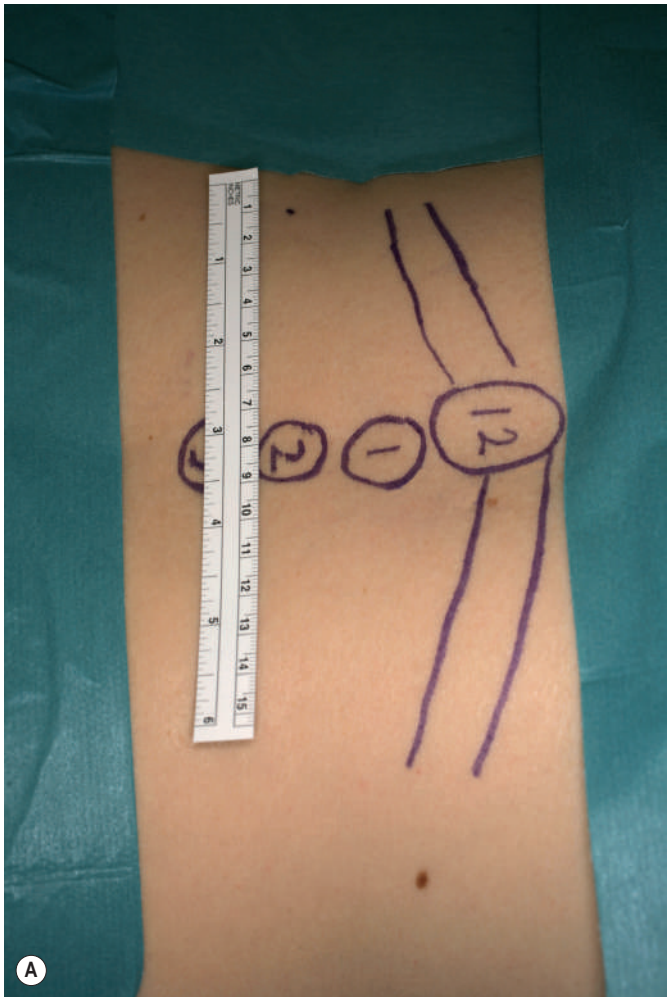

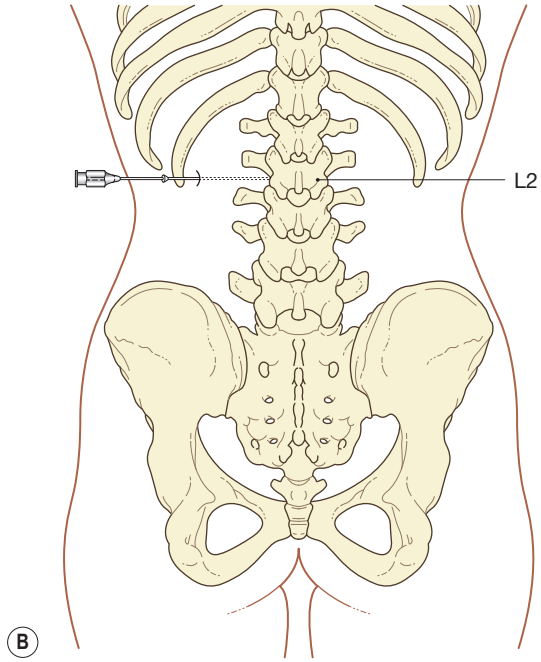

Fig. 5.3.3

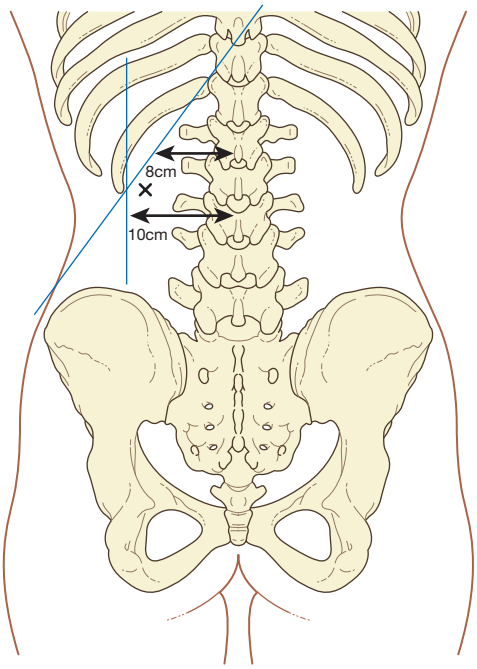

Fig. 5.3.4

Fig. 5.3.5

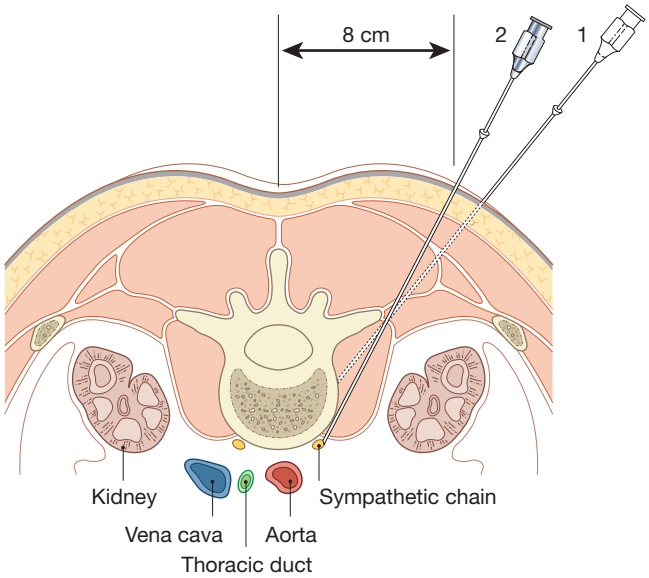

Fig. 5.3.6

- Neurolysis may be achieved by leaving the needle in place after block has been confirmed and injecting 5 ml phenol 6%. To avoid leaving alcohol in the needle tract, the needle is then cleared with air or local anesthetic 1 ml, and removed. However, because of the very high incidence of genitofemoral neuralgia that can occur post-neurolytic lumbar sympathetic block, the benefit versus risk should be considered carefully.
- Monitors should be left attached and i.v. access left in situ for at least 30 minutes.

Confirmation of a successful block

- Increase in skin temperature on the plantar surface of the foot to about 35 °C; temperature should begin to rise in the foot on the side of the block within 3 minutes of injection of local anesthetic.
- Relief of sympathetically maintained pain in the lower limb.

Tips

- If fluoroscopy is not available ultrasound may aid placement of the needle. A line 10 cm from the midline is drawn parallel to the midline; the lowest rib is identified and a line is drawn along its inferior border.
- The point of intersection of these lines should be lateral to the L2 vertebral body.
- Consideration should be given to performance of the block under CT control if the block is unsuccessful.
- Repeated blocks may bring about gradual improvement in sympathetically maintained pain.
- Immediate physiotherapy after blockade may improve the outcome.

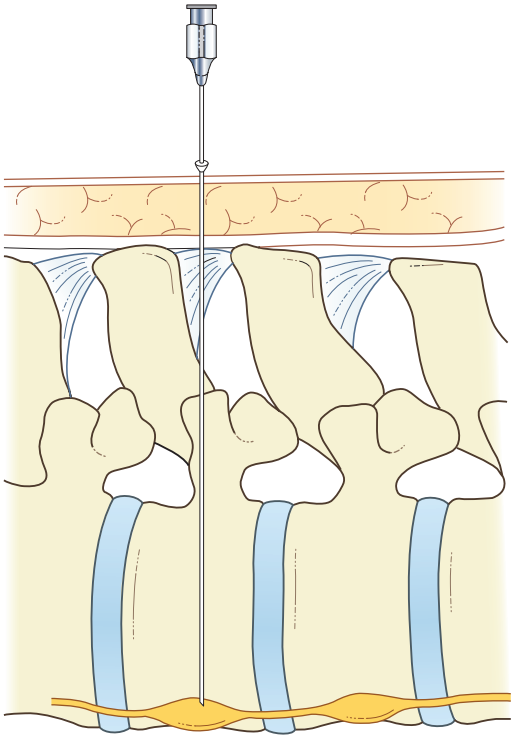

Fig. 5.3.7

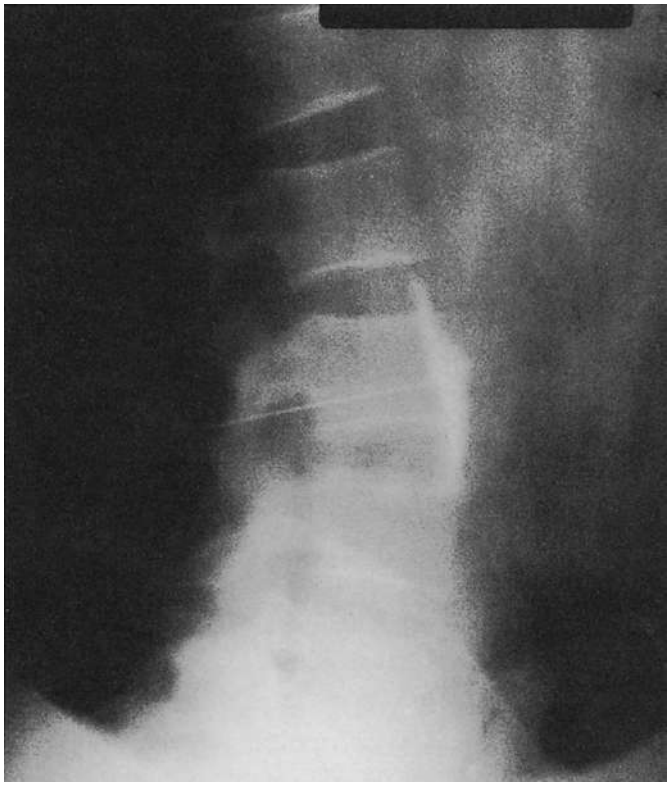

Fig. 5.3.9

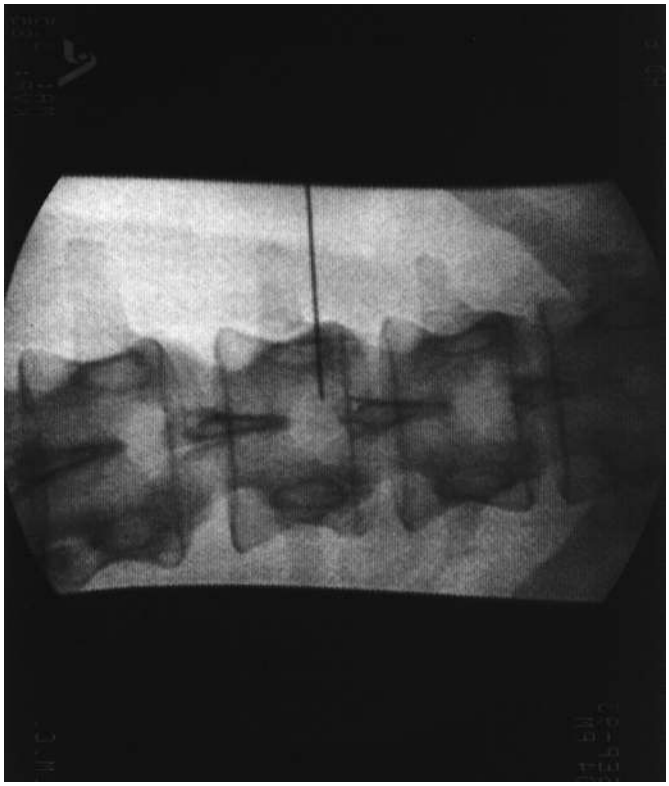

Fig. 5.3.8

Potential problems

- If the needle tip is placed too superficially, the tip may come to lie in the intervertebral foramen and injection may result in a subarachnoid block, an epidural block, or a somatic nerve block. Confirmation of needle position using lateral fluoroscopy is therefore recommended.
- Genitofemoral neuralgia may occur in 5–10% of patients post-neurolytic block causing pain in the groin.
- Perforation of the aorta or the inferior vena cava is possible and retroperitoneal hematoma may occur. Consequently the block should be avoided in patients with coagulopathy.
- Intravascular injection may occur.
- Perforation of the kidney or ureter is usually of no clinical significance unless neurolytic agents are used.
- Perforation of the intervertebral disc may occur. This also is usually of no clinical significance but may produce a septic discitis if bacterial contamination occurs.
- Postural hypotension, secondary to sympathetic blockade, may occur.
- Injection of neurolytic solution into the psoas muscle may cause rhabdomyolysis.
- Patients in the prone position should be monitored carefully when intravenous sedation is administered.

5.4 CELIAC PLEXUS BLOCK—RETROCRURAL APPROACH

Anatomy

The celiac plexus is flat and lies against the crus of the diaphragm, surrounding the root of the celiac and mesenteric arteries and anterior to three vertebral bodies centered at L1. Posteriorly on the left side is the aorta, and on the right is the inferior vena cava. The kidneys lie lateral and the pancreas anterior to the celiac plexus (Fig. 5.4.1 a,b).

The celiac plexus is made up of pre- and postganglionic sympathetic and parasympathetic nerve fibers. Postganglionic sympathetic fibers are supplied from the paired celiac ganglia. Preganglionic sympathetic efferents from the thoracic sympathetic chain are supplied via the greater and lesser splanchnic nerves. The intra-abdominal viscera are supplied by postganglionic sympathetic fibers that have synapsed in the celiac ganglia (Fig. 5.4.2). The vagus nerve also supplies parasympathetic nerve fibers. Via the celiac plexus dorsal root, ganglion cells innervate the whole of the abdominal viscera, including the liver, spleen, kidneys, suprarenal glands, and intestines, with the exception of the pelvic organs, the rectum, and the left half of the colon.

Pain originating from the viscera is often vague and poorly localized as a result of convergence of neurons in the dorsal horn and crossing over the midline of some of the visceral afferents.

There are two main approaches to celiac plexus blockade. One approach places the two needles posterior to the crura of the diaphragm, the retrocrural approach. The retrocrural approach to the celiac plexus also targets the splanchnic nerves to produce a splanchnic nerve block if required. The other approach places a needle anterior to the crus of the diaphragm on the right, the anterocrural approach (Fig. 5.4.3), as discussed in Section 5.5.

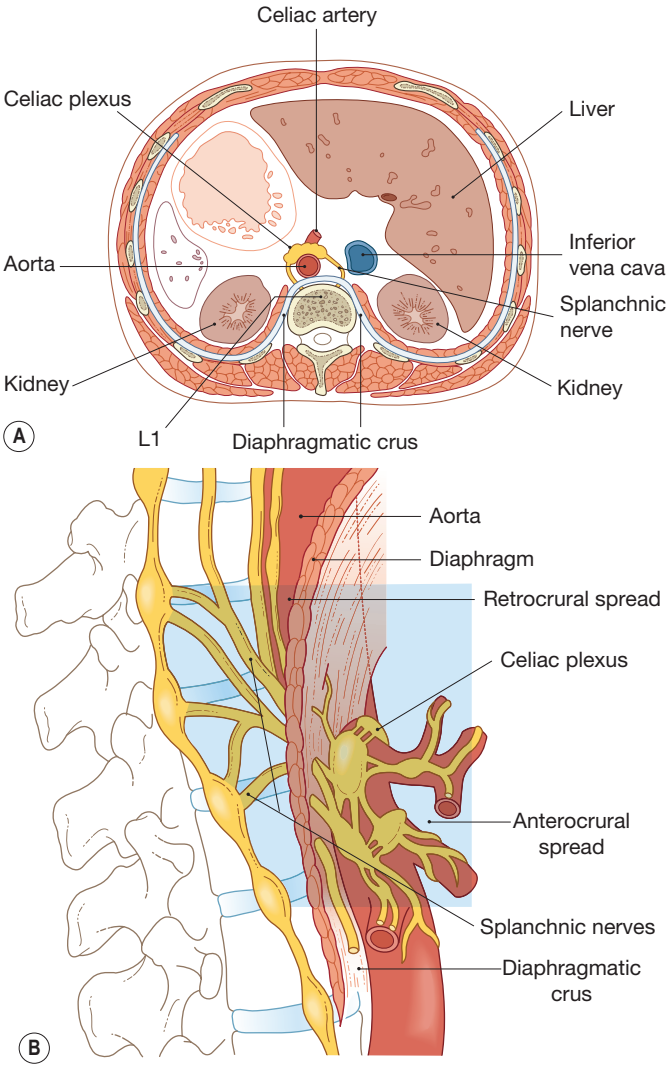

Fig. 5.4.1

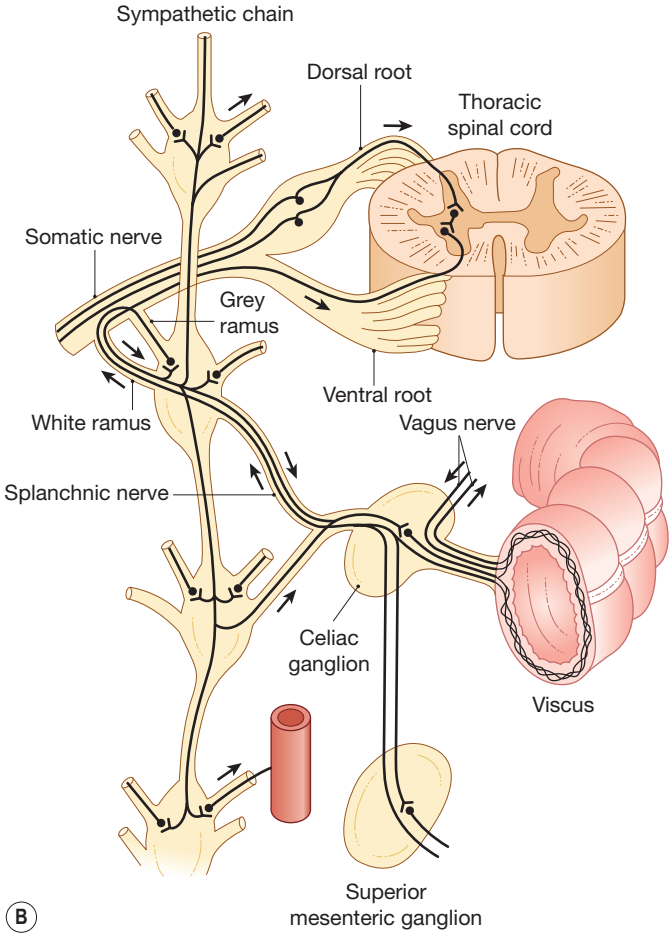

Fig. 5.4.2

Equipment

- 2 ml, 5 ml, and 10 ml syringes
- 30 G needle
- Two 15 cm 22 G needles
- Extension set (optional)
- ECG, BP, SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy

Drugs

- Mild sedative
- Lidocaine (lignocaine) 1%, 5 ml for skin infiltration
- Lidocaine (lignocaine) 1%, 15–20 ml (or its equivalent) for block
- 6% aqueous phenol or 50–75% alcohol (ethanol). Our suggestion: mix 2 parts absolute alcohol with one part 1% lidocaine. This will help reduce the incidence and severity of pain following injection. In addition, precede all alcohol injections with 3–4 ml 1% lidocaine
- Non-ionic radio-opaque contrast medium
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 5.4.4*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- Mild sedation may be induced.

RETROCRURAL APPROACH

- A 15 cm 22 G spinal needle is selected. A slight curve at the needle tip, away from the bevel direction, may be created, which allows the needle to be redirected during placement.
- An AP view of the upper lumbar/low thoracic spine is obtained and the C-arm is adjusted to superimpose the T12–L1 endplates.
- **A skin wheal is raised at the lower border of the twelfth rib on the right just above the level of the L1 transverse process** (*Fig. 5.4.5 a*). The needle is inserted at this site and advanced at an angle 30° from perpendicular inward until the L1 body is contacted just below the upper endplate (*Fig. 5.4.5*). The curve of the needle is turned laterally and the needle is advanced along the upper portion of the body. Once the needle has slipped a few millimeters past the lateral aspect of the L1 body, a lateral view is obtained. The curve of the needle is directed inward toward the body and advanced until the tip lies at the anterior border of the body, near the upper endplate, in a direct lateral view.
- Using "live" fluoroscopy, 1 ml non-ionic contrast medium is injected. The dye should remain against the anterior aspect of the bodies in the lateral view (*Fig. 5.4.6 a*).
- If dye is seen spreading dorsally toward the neural foramina (*see Fig. 5.4.6 b*), the needle should be withdrawn and repositioned at a higher level.
- An AP view is obtained, which should demonstrate dye spread against the lateral aspect of the bodies (*Fig. 5.4.6 c*).
- Spread more laterally indicates injection within the psoas muscle, in which case the needle should be repositioned more medially and anteriorly. 3 ml 1% lidocaine is then injected. The dye shadow will

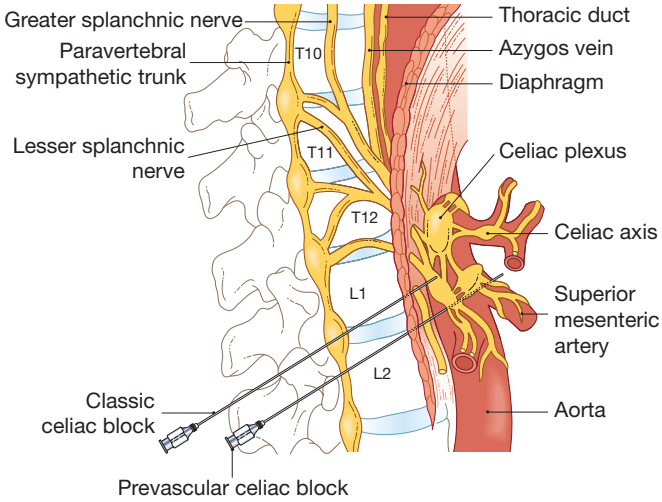

Fig. 5.4.3

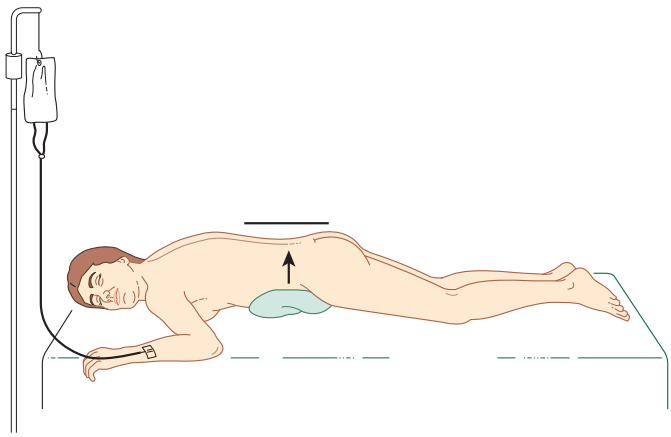

Fig. 5.4.4

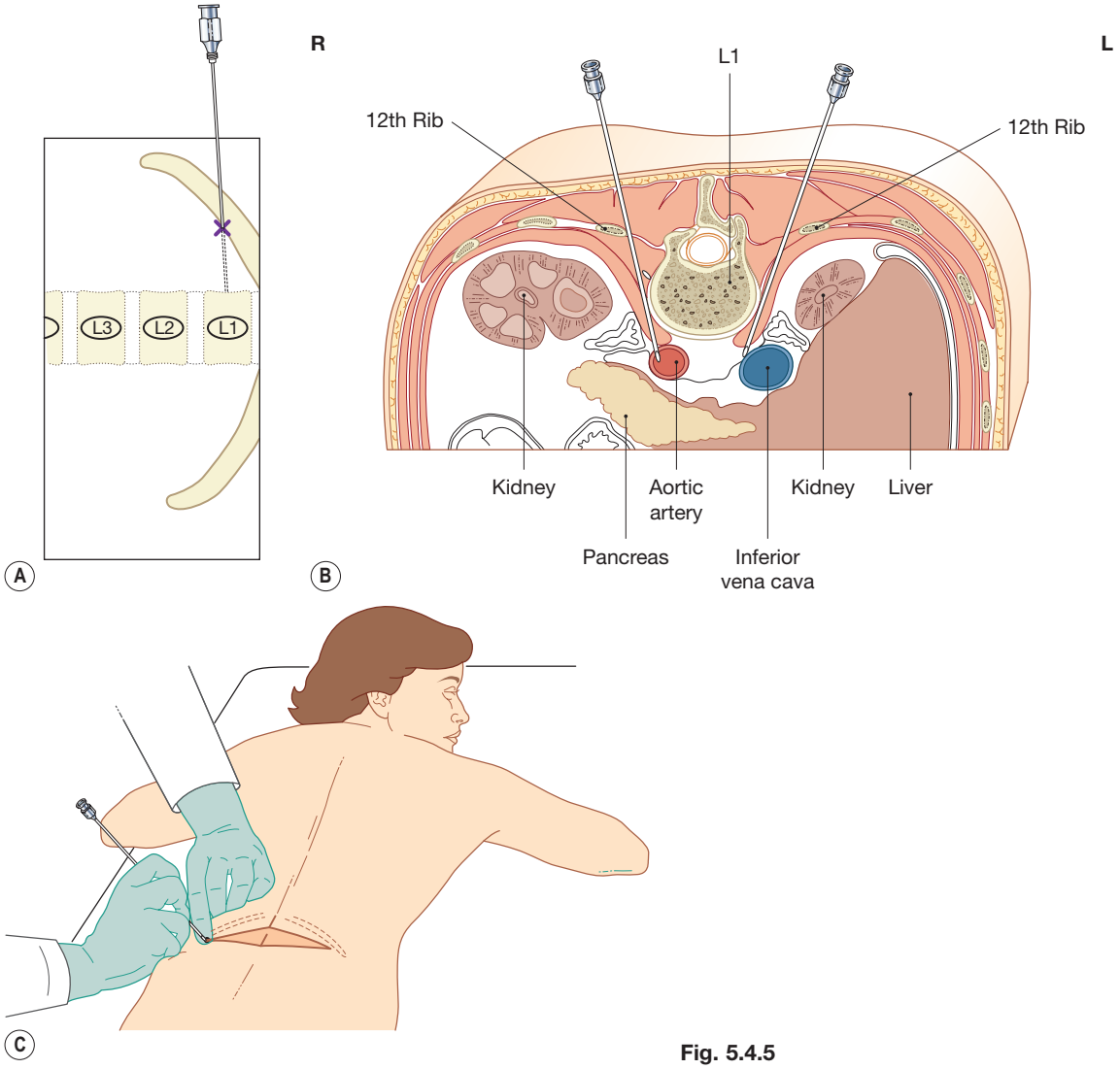

Fig. 5.4.5

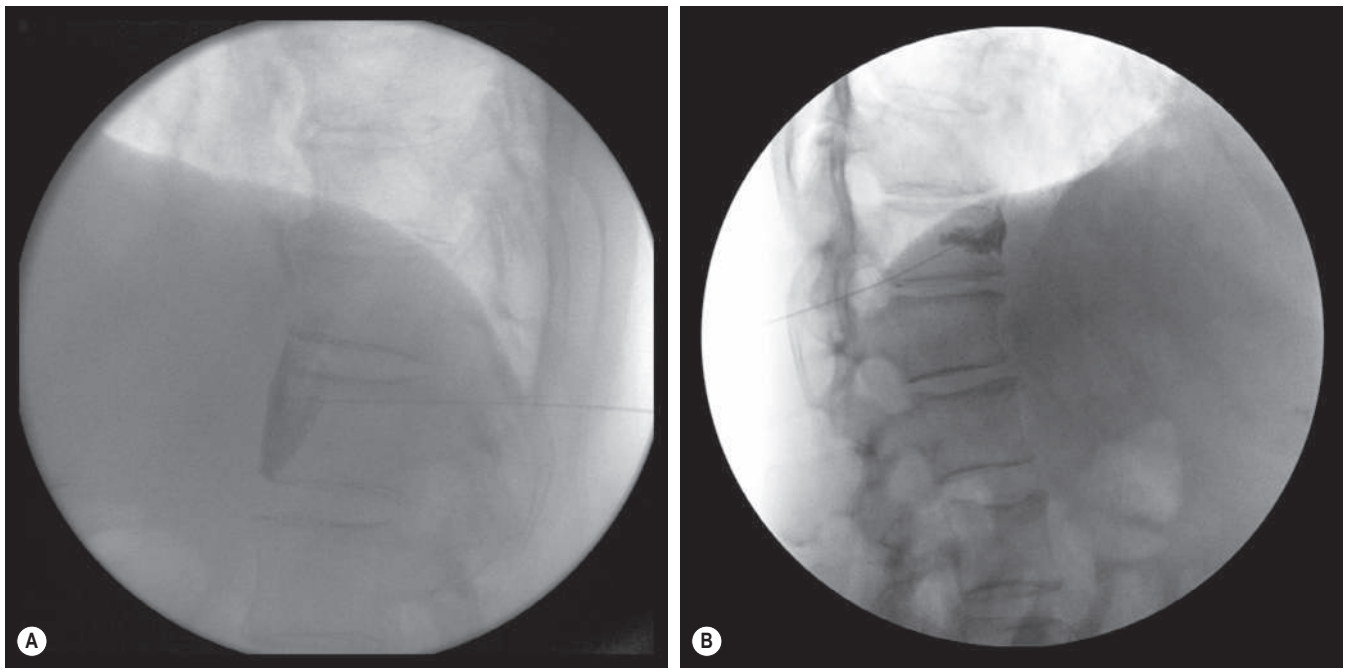

Fig. 5.4.6

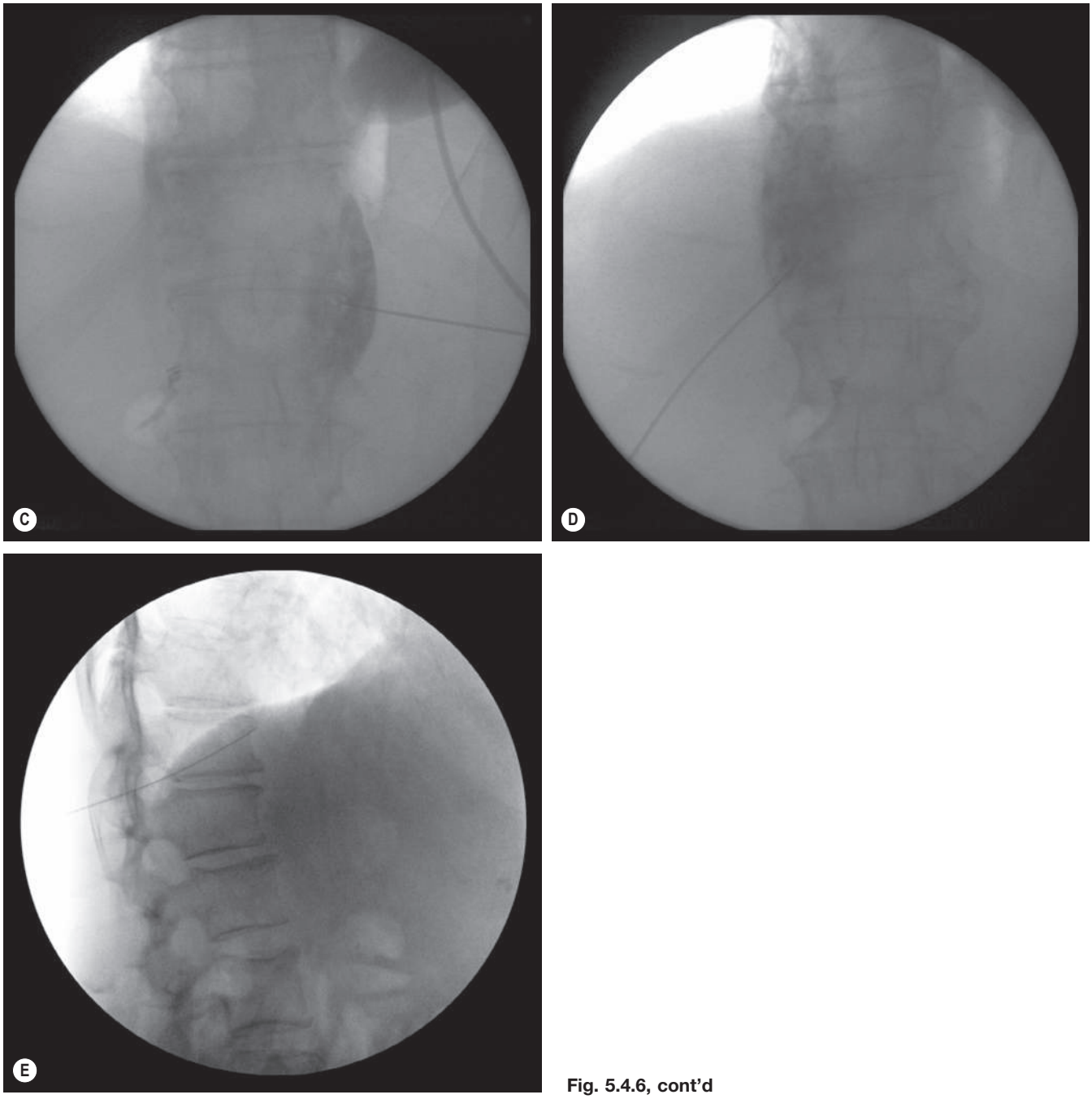

Fig. 5.4.6, cont'd

be seen to expand superiorly, spreading to the thoracic levels to contact the splanchnic nerves (*see Fig. 5.4.6 d*).

- After confirming negative aspiration for blood, 15–20 ml alcohol or phenol is injected. The needle is cleared with 1 ml lidocaine prior to removal.
- Alternatively, the needle can be advanced more cephalad to a position at the anterior border of T12 preferably near either the lower or upper endplate (*Fig. 5.4.6 e*).
- Injection near the mid-point of the body is more likely to result in dorsal spread of the drug toward the neural foramen. More cephalad placement is a bit more difficult technically, but places the needle closer to the splanchnic nerves.
- The procedure is repeated in an identical manner on the left side.

Confirmation of a successful block

- Relief of upper abdominal pain.

Tips

- After injection of non-ionic radio-contrast medium, a blush will indicate injection into muscle. If visible contrast medium disappears immediately it is likely that intravascular injection has occurred.
- Consideration should be given to performance of the block under CT control if the block is unsuccessful.
- Placement of the needle anterior to the diaphragmatic crus can also be achieved via insertion through the abdominal wall.

Potential problems

- The position of each needle tip should always be confirmed with fluoroscopy before injection of neurolytic agent as it may lie in the peritoneal cavity, within a viscus or intravascularly. If a needle tip is placed too superficially, the tip may come to lie in the intervertebral foramen and injection may result in an epidural block or a somatic nerve block. Injection of neurolytic solution into the psoas muscle may cause rhabdomyolysis.
- Perforation of the aorta or the inferior vena cava is possible and consequently the block should be avoided in patients with coagulopathy. Dissection of the aorta may occur as a result of direct damage during the block. Retroperitoneal hematoma may occur and for this reason also the block should be avoided in patients with coagulopathy.
- Orthostatic hypotension may occur as a result of sympathetic blockade for up to 3 days after a neurolytic block. Diarrhea may occur also and hydration of the patient should be monitored.
- Pneumothorax may occur.
- Transient motor paralysis and paraplegia may occur after the block, probably as a result of spasm of the segmental arteries.
- Perforation of the intervertebral disc may occur, but this also is usually of no clinical significance.
- Perforation of the kidney or ureter is usually of no clinical significance unless neurolytic agent is injected.
- The thoracic duct may be damaged (possibly causing chylothorax, or lymphedema).
- Abdominal and chest discomfort may be experienced for 30 minutes after injection of alcohol.
- There may be a detectable odor from the breath after alcohol injection.
- Patients in the prone position should be monitored carefully when intravenous sedation is administered.

5.5 CELIAC PLEXUS BLOCK—ANTEROCRURAL APPROACH

Anatomy

The anterocrural approach places a needle anterior to each crus of the diaphragm. The needles are inserted more medially and directed at a larger angle towards the midline until they come to lie in the retroperitoneal compartment between the aorta, and the inferior vena cava dorsally, and the pancreas ventrally (**Fig. 5.5.1**). Fluoroscopic imaging is necessary for accurate placement of the anterocrural needles using this approach.

Equipment

- 2 ml, 5 ml, and 10 ml syringes
- 30 G needle
- 15 cm 22 G needle (penetration of the diaphragmatic crus is easier with a large gauge needle)
- Extension set (optional)
- ECG, BP, SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy

Drugs

- Mild sedative
- Lidocaine (lignocaine) 1%, 5 ml for skin infiltration
- Lidocaine (lignocaine) 1%, 15–20 ml (or its equivalent) for block
- 6% aqueous phenol or 50–75% alcohol (ethanol). Our suggestion: mix 2 parts absolute alcohol with one part 1% lidocaine. This will help reduce the incidence and severity of pain following injection. In addition, precede all alcohol injections with 3–4 ml 1% lidocaine
- Non-ionic radio-opaque contrast medium
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 5.5.2*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- Mild sedation may be induced.

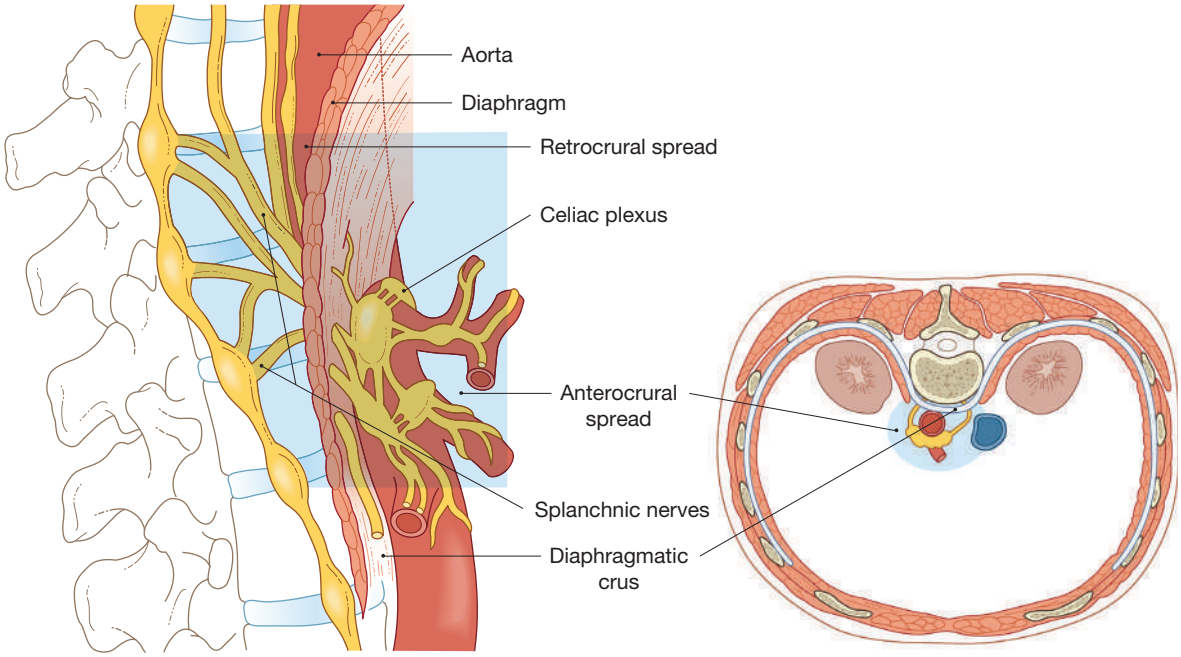

Fig. 5.5.1

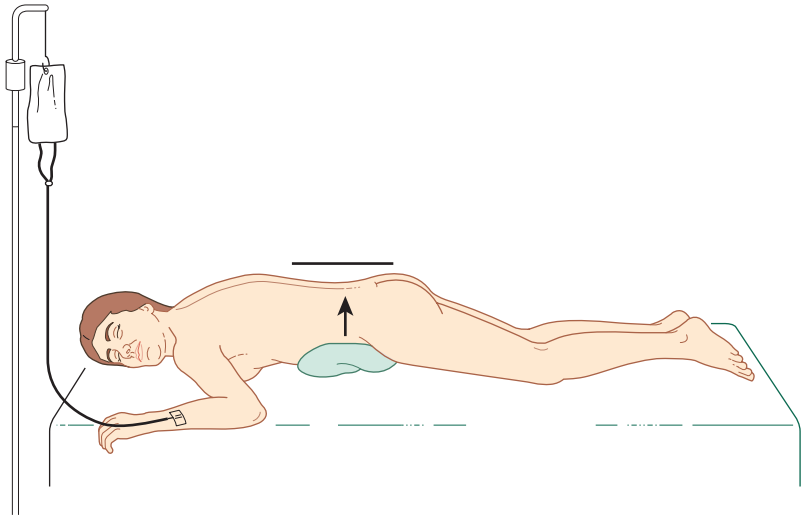

Fig. 5.5.2

- The thoracolumbar midline and area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- The twelfth rib and L1 are identified and confirmed with fluoroscopy.

ANTEROCRURAL APPROACH

Right side

- A 15 cm 22 G spinal needle is selected. A slight curve at the needle tip, away from the bevel direction, may be created which allows the needle to be redirected during placement.
- An AP view of the upper lumbar/low thoracic spine is obtained and the C-arm is adjusted to superimpose the T12–L1 endplates.
- **A skin wheal is raised at the lower border of the twelfth rib on the right just above the level of the L1 transverse process. The needle is inserted at this site and advanced at an angle 30° from perpendicular inward until the L1 body is contacted just below the upper endplate.**
- The curve of the needle is turned laterally and the needle is advanced along the upper portion of the body.
- Once the needle has slipped a few millimeters past the lateral aspect of the L1 body, a lateral view is obtained.
- The needle is advanced until the tip is 1.5–2 cm anterior to the anterior border of the L1 body. The needle is aspirated and if negative, 1 ml non-ionic contrast is injected. Dye spread should be in an amorphous pattern (*Fig. 5.5.3*).
- If aspiration is negative, 3 ml 1% lidocaine (lignocaine) is injected. If no nerve block is noted after 10 minutes, this is followed by 15–20 ml alcohol. If phenol is used, the lidocaine is not needed.
- 1 ml 1% lidocaine is injected before removing the needle to clear it.

Left side

- **The same procedure is repeated on the left.**
- The needle is positioned 1.5–2 cm anterior to the anterior border of the L1 body. It is then usually within the aorta, and aspiration is positive for arterial blood.
- The needle is advanced forward until aspiration is negative for blood (*Fig. 5.5.3 b*).
- 1 ml contrast is injected. The pattern is generally amorphous anteriorly, but a straight border of dye along the anterior surface of the aorta may be seen (*Fig. 5.4.4*).
- Aspiration is repeated and, if negative, 3 ml 1% lidocaine (lignocaine) is injected. If no nerve block is noted after 10 minutes, this is followed by 15–20 ml alcohol. If phenol is used, the lidocaine is not needed.
- 1 ml 1% lidocaine is injected before removing the needle to clear it.
- Monitors should be left attached and i.v. access left in situ for at least 30 minutes.

Confirmation of a successful block

- Relief of upper abdominal pain.

Tips

- After injection of non-ionic radio-contrast medium, a blush will indicate injection into muscle. If visible contrast medium disappears immediately it is likely that intravascular injection has occurred.
- Consideration should be given to performance of the block under CT control if the block is unsuccessful.

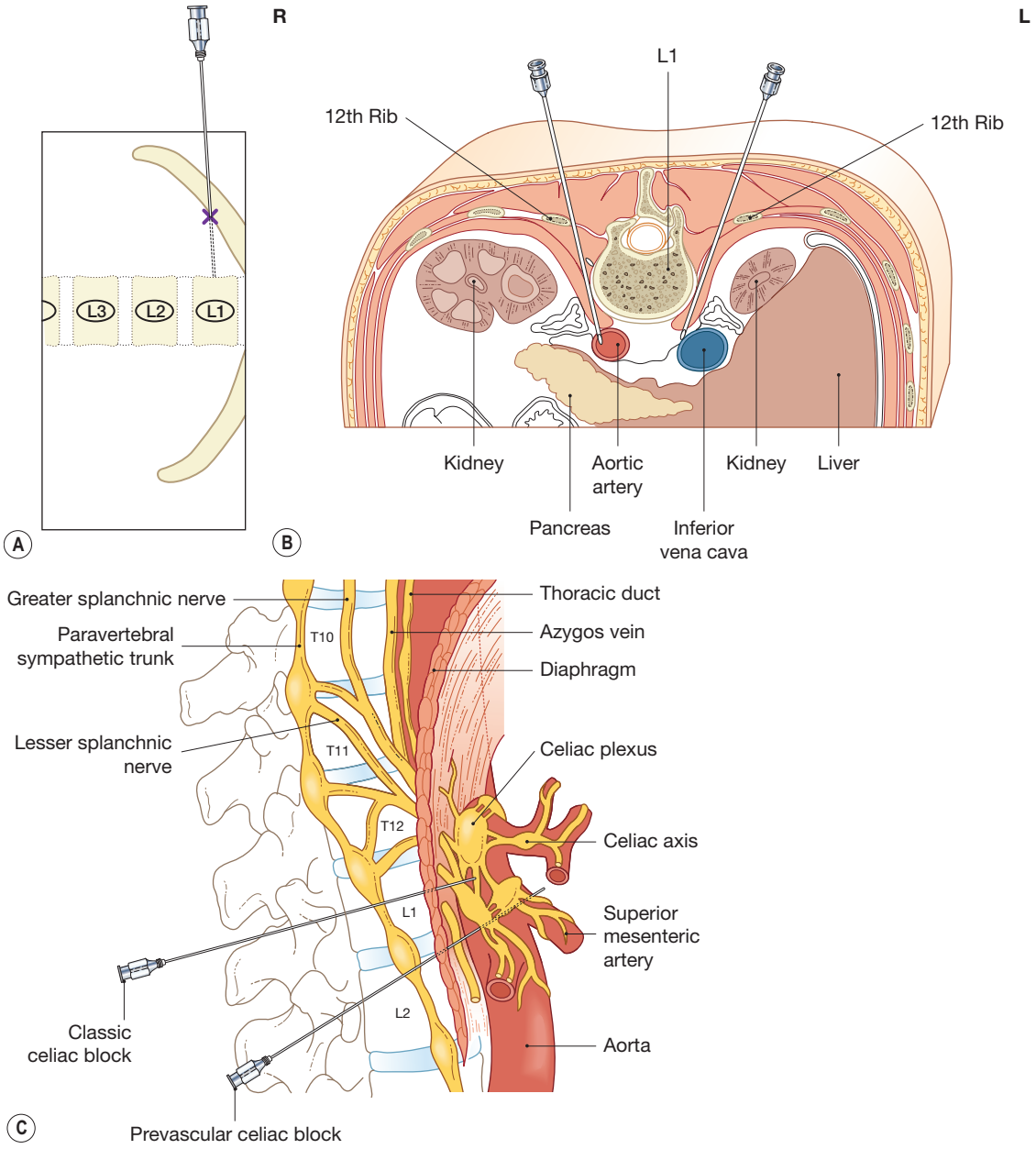

Fig. 5.5.3

Potential problems

- The position of each needle tip should always be confirmed with fluoroscopy prior to injection of neurolytic agent as it may lie in the peritoneal cavity, within a viscus or intravascularly. If a needle tip is placed too superficially, the tip may come to lie in the intervertebral foramen and injection may result in epidural block or a somatic nerve block. Injection of neurolytic solution into the psoas muscle may cause rhabdomyolysis.
- Perforation of the aorta or the inferior vena cava is possible and consequently the block should be avoided in patients with coagulopathy. Dissection of the aorta may occur as a result of direct damage during the block. Retroperitoneal hematoma may occur and for this reason also the block should be avoided in patients with coagulopathy.
- Orthostatic hypotension may occur as a result of sympathetic blockade for up to three days after a neurolytic block. Diarrhea may occur also and hydration of the patient should be monitored.

- Pneumothorax may occur.
- Transient motor paralysis and paraplegia may occur after the block, probably as a result of spasm of segmental arteries.
- Perforation of the intervertebral disc may occur, but this also is usually of no clinical significance.
- Perforation of the kidney or ureter is usually of no clinical significance unless neurolytic agents are injected.
- The thoracic duct may be damaged (possibly causing chylothorax, or lymphedema).
- Abdominal and chest discomfort may be experienced for 30 minutes after injection of alcohol.
- There may be a detectable odor from the breath after alcohol injection.
- Patients in the prone position should be monitored carefully when intravenous sedation is administered.

5.6 HYPOGASTRIC PLEXUS BLOCK

Anatomy

The superior hypogastric plexus is formed from pelvic sympathetic fibers of the aortic plexus and L2 and L3 splanchnic nerves. These afferent and efferent fibers innervate the pelvic viscera, including the uterus, bladder, vagina, and prostate. The plexus is located between the upper third of the first sacral vertebral body and the lower third of the fifth lumbar vertebral body, at the sacral promontory, in the retroperitoneal space (**Fig. 5.6.1 a,b**). Parasympathetic nerve fibers from S2–S4 pass through the inferior hypogastric plexus.

Equipment

- 2 ml, 5 ml, and 10 ml syringes
- 30 G needle
- Two 15 cm 22 G needles
- Extension set (optional)
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)
- Fluoroscopy

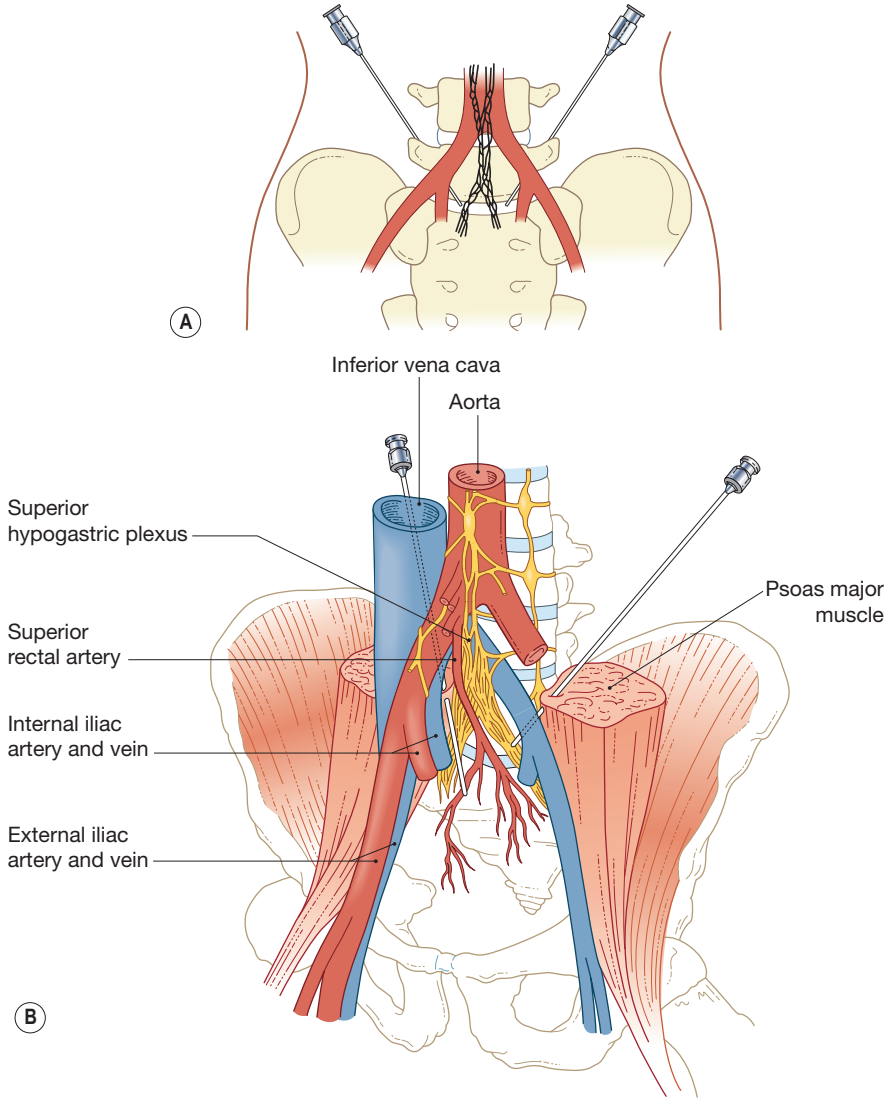

Fig. 5.6.1

Drugs

- Lidocaine (lignocaine) 1%, 5 ml for skin infiltration
- Lidocaine (lignocaine) 1%, 15–20 ml (or its equivalent) for block
- Phenol 6%
- Non-ionic radio-opaque contrast medium
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 5.6.2*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The lumbosacral midline and area 10 cm × 5 cm laterally is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- The iliac crests are palpated and the intercrestal line is identified (this corresponds with the inferior aspect of the spinous process of L4 or may lie in the L4–5 interspace).
- The spinous processes are counted until the L5–S1 interspace is identified and confirmed with fluoroscopy.
- **The insertion points lie 2 cm lateral and 2 cm cephalad to the space between the L5 transverse process and the sacrum** (*Fig. 5.6.3 a,b*).
- A skin wheal is raised at one of the marked sites and the area is infiltrated with lidocaine (lignocaine) 1%.
- At a 30° angle to the frontal plane, a 22 G 15 cm needle is advanced, aimed slightly caudad towards the L5–S1 interspace (*Fig. 5.6.3 c*), until the lower part of the L5 vertebral body is contacted at a depth of about 7–9 cm and confirmed with fluoroscopy. If the needle contacts bone at a more superficial level, it is probable that it has come into contact with the L5 transverse process or the sacrum and needs to be repositioned.
- The needle depth is noted.
- The needle is then withdrawn to the subcutaneous tissue and, with the aid of fluoroscopy, it is re-advanced, this time at an angle 45° to the frontal plane (or with slight concavity of the needle) until the previous depth (as noted) is reached. It should slip past the vertebral body at a depth about 1–2 cm deeper than the first depth mark, to lie just anterior to the upper portion of the sacrum (*Fig. 5.6.4*).
- After negative aspiration, the fluoroscopic image is observed as a small amount of non-ionic radio-contrast medium is injected (*Fig. 5.6.5*). The correct placement of the needle is indicated by the presence of a collection of contrast medium just anterior to the upper portion of the sacrum or the L5–S1 interspace (*Fig. 5.6.6*). The contrast medium usually spreads in all directions, not usually along the sacrum.
- The procedure is repeated on the other side in a mirrored fashion (*Figs 5.6.7, 5.6.8*).
- After further negative aspirations, 5 ml of lidocaine (lignocaine) 1% is injected bilaterally and the patient is questioned about pain relief. There should be little resistance to injection, similar to that felt when injecting through an epidural needle. If resistance is encountered, or if the injection is painful, the needle should be repositioned.
- After 10 minutes the patient is questioned about pain relief and any symptoms of somatic nerve blockade. Sensory and motor functioning of the lower extremities is checked. The procedure should be abandoned if there is evidence of somatic blockade.
- After confirmation of pain relief and lack of somatic block, 6 ml phenol 6% is injected through each needle using glass syringes. The needles are then cleared with air or local anesthetic 1 ml, and removed.
- Monitors should be left attached and i.v. access left in situ for at least 30 minutes.
- Ultrasound may help placement of the needle (*Figs 5.6.9–5.6.10*).

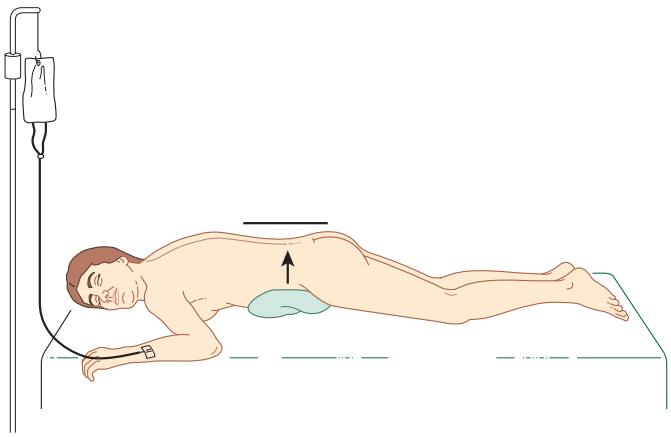

Fig. 5.6.2

Confirmation of a successful block

- Relief of lower abdominal pain.

Tips

- After injection of non-ionic radio-contrast medium, a blush will indicate injection into muscle. If this disappears immediately it is likely that intravascular injection has occurred.

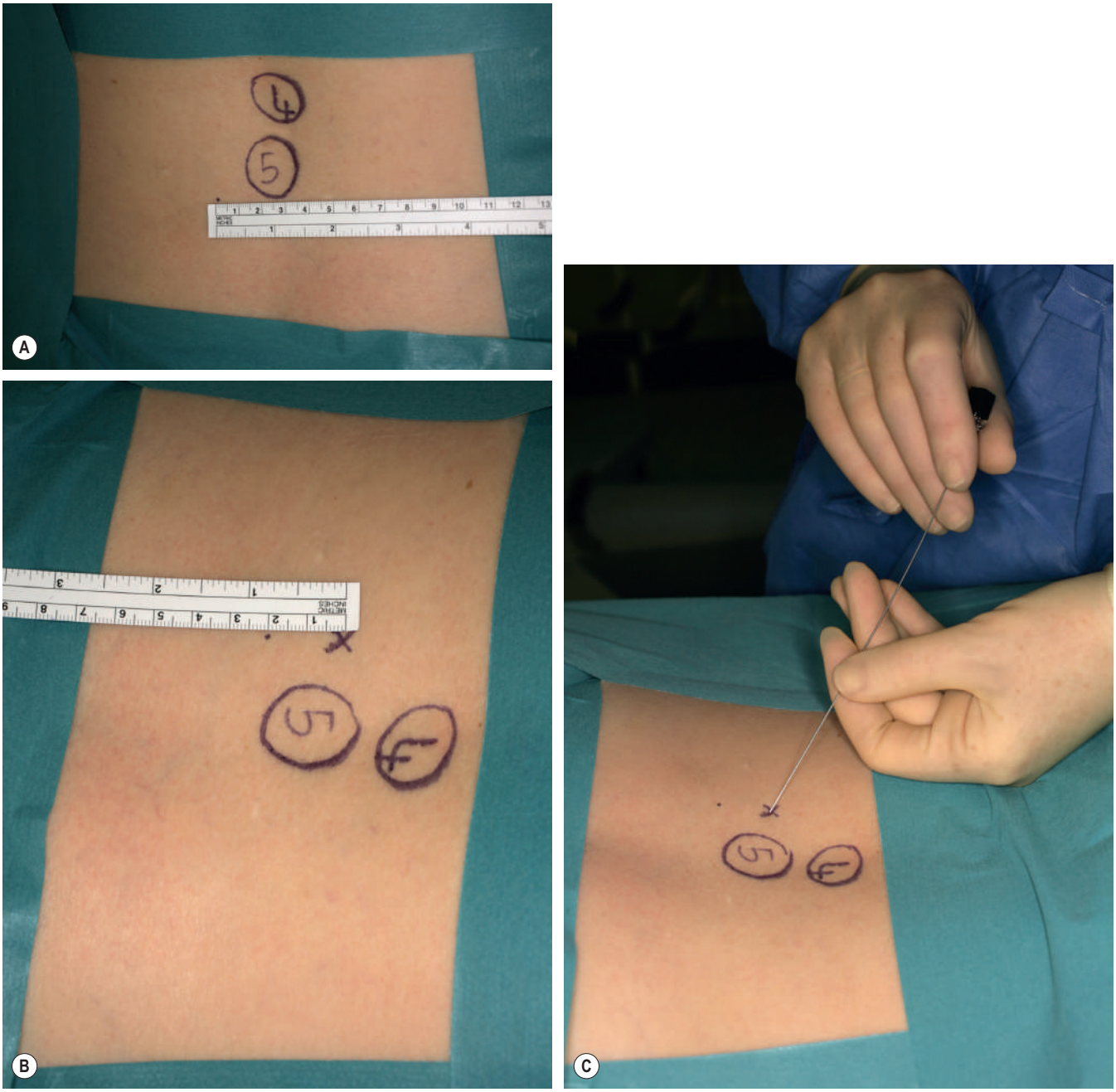

Fig. 5.6.3

- Consideration should be given to performance of the block under CT control if the block is unsuccessful (*Fig. 5.6.10*).

Potential problems

- The position of each needle tip should always be confirmed with fluoroscopy prior to injection of neurolytic agents as it may lie in the peritoneal cavity, within a viscus or intravascularly. If a needle tip is placed too superficially, the tip may come to lie in the intervertebral foramen and injection may result in an epidural block or a somatic nerve block. Injection of neurolytic solution into the psoas muscle may cause rhabdomyolysis.
- Perforation of the aorta or the inferior vena cava is possible and consequently the block should be avoided in patients with coagulopathy. Dissection of the aorta may occur as a result of direct damage during the block. Retroperitoneal hematoma may occur and for this reason also the block should be avoided in patients with coagulopathy.

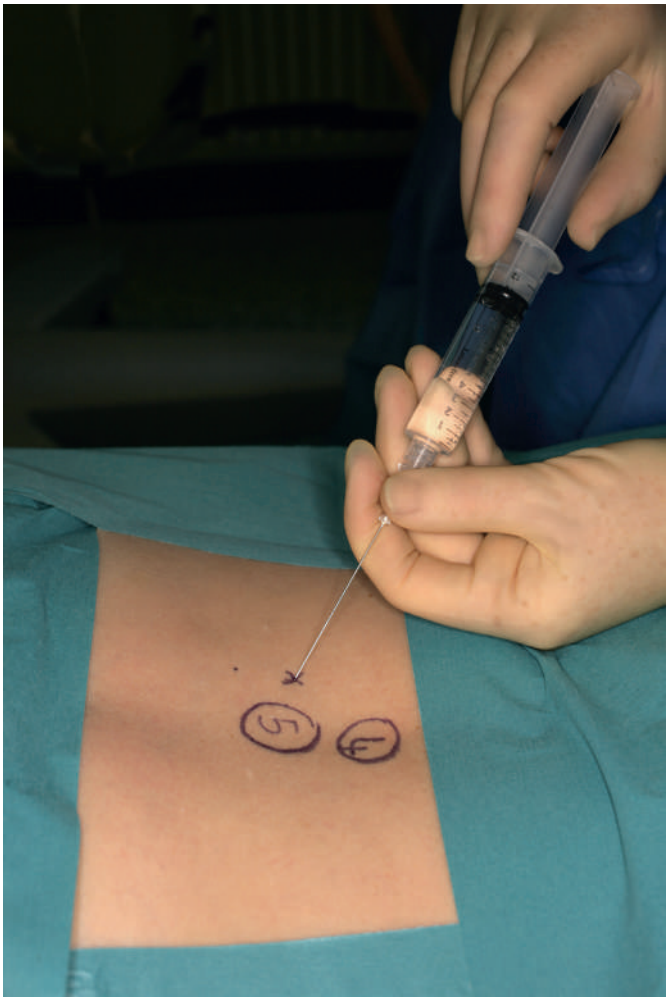

Fig. 5.6.4

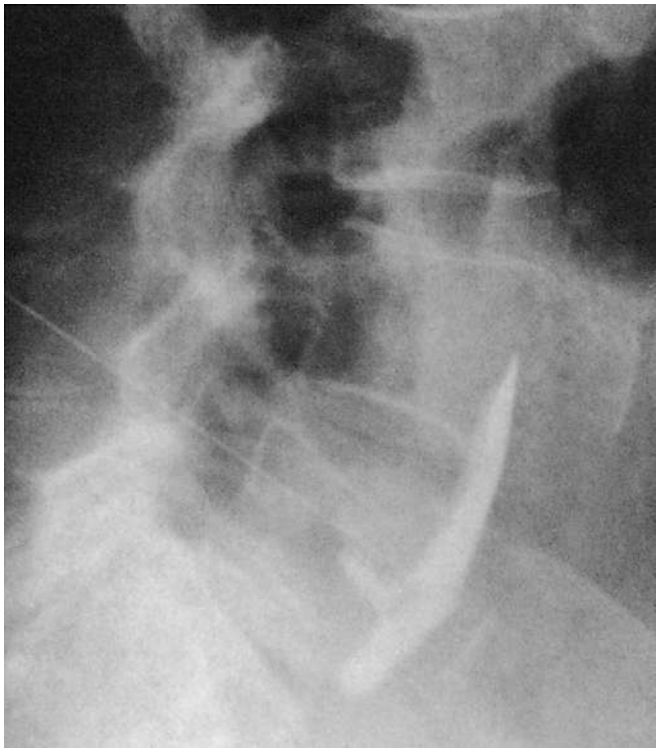

Fig. 5.6.6

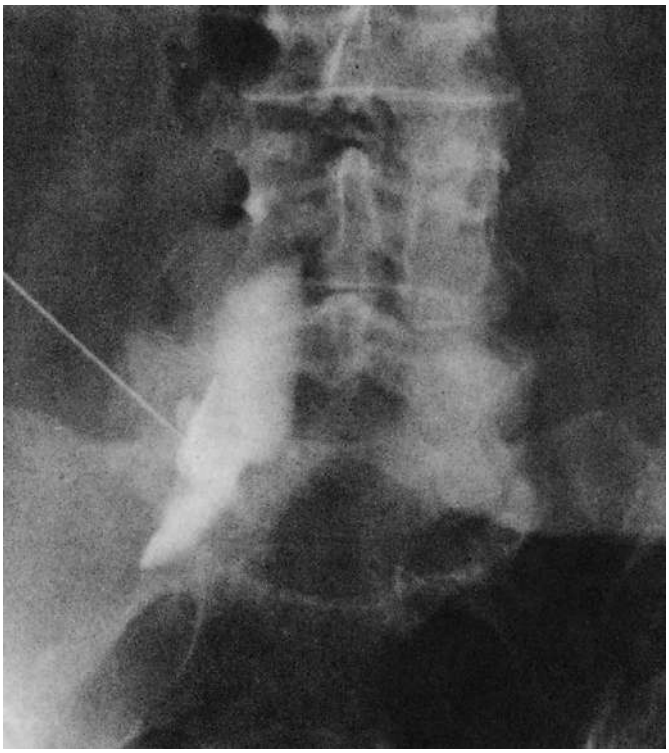

Fig. 5.6.5

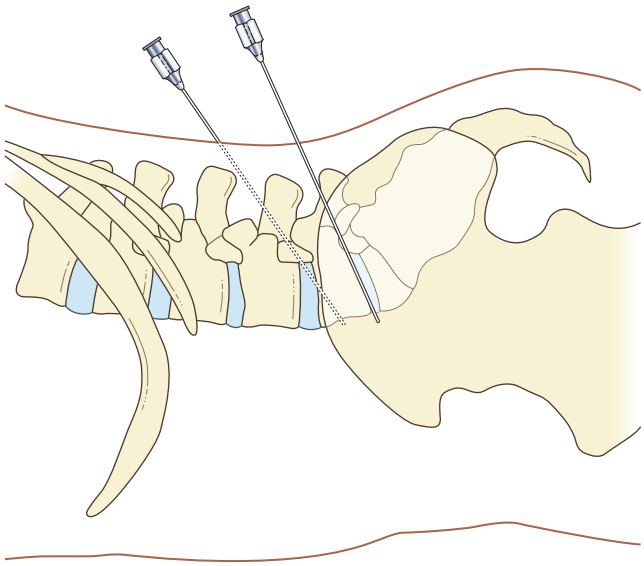

Fig. 5.6.7

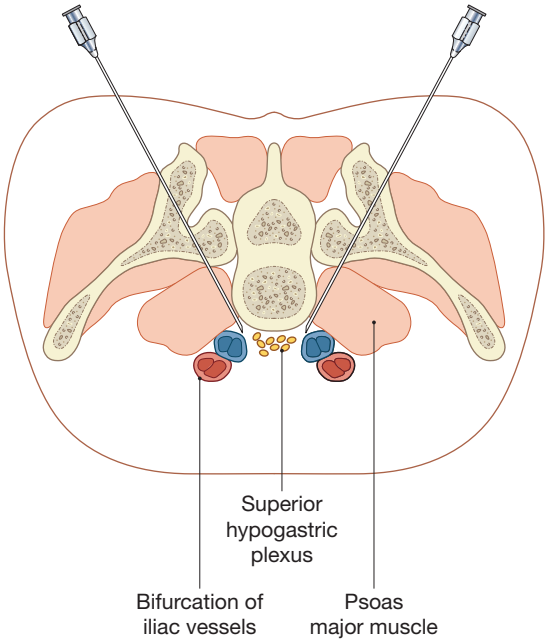

Fig. 5.6.8

- Orthostatic hypotension may occur as a result of sympathetic blockade for up to three days after a neurolytic block. Diarrhea may occur also and hydration of the patient should be monitored.
- Transient motor paralysis and paraplegia may occur after the block, probably as a result of spasm of segmental arteries.
- Perforation of the intervertebral disc may occur, but this is usually of no clinical significance.

Fig. 5.6.9

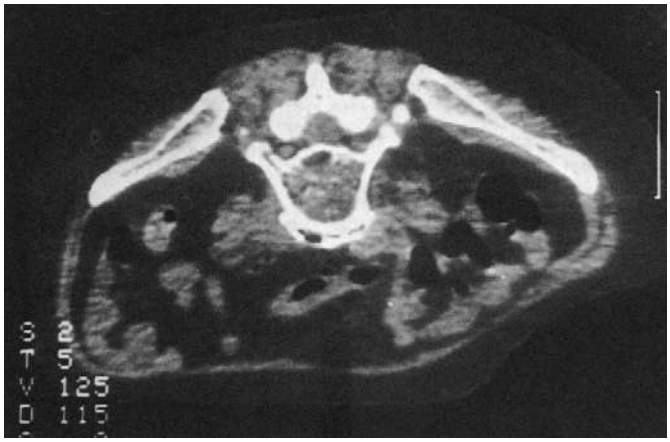

Fig. 5.6.10

5.7 GANGLION IMPAR BLOCK

Anatomy

The ganglion impar is a retroperitoneal sympathetic ganglion located at the level of the sacrococcygeal junction (Fig. 5.7.1). Above the level of this ganglion the sympathetic chains are paired. Sympathetic afferents from the perineum, distal rectum and anus, distal urethra, vulva and the distal third of the vagina converge in the ganglion impar.

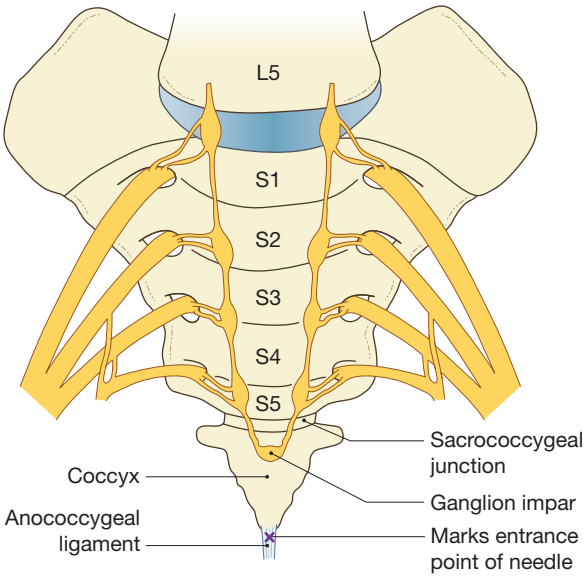

Fig. 5.7.1

Equipment

- 2ml, 5 ml, and 10 ml syringes
- 30 G needle
- 22 G spinal needle
- Extension set (optional)
- ECG, BP, and SpO_2 monitors
- Resuscitation equipment (*see Appendix 3*)

Drugs

- Lidocaine (lignocaine) 1%, 5 ml for skin infiltration
- Lidocaine (lignocaine) 1%, 15–20 ml (or its equivalent) for block
- Phenol 6%
- Non-ionic radio-opaque contrast medium
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Prone.
- Pillow under anterior superior iliac spine to flatten the normal lumbar lordosis (*Fig. 5.7.2*).

Needle puncture and technique

- Intravenous access is inserted.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The midline along the intergluteal groove and an area 10 cm × 5 cm laterally is cleaned with antiseptic

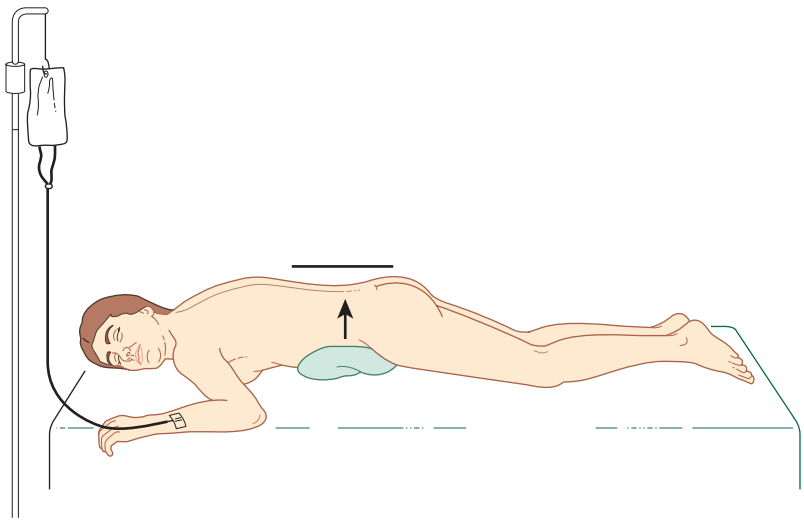

Fig. 5.7.2

solution and a fenestrated drape is placed over the sterile area.

- A skin wheal is raised at the superior aspect of the intergluteal groove, just above the anus, over the anococcygeal ligament (*Fig. 5.7.3*).
- The stylet from the 22 G spinal needle is removed, and the needle is bent with the fingers to form a 30° angle, approximately 2 cm from the hub.
- The needle is inserted through the skin wheal, with the concave curvature facing posteriorly.
- With the aid of fluoroscopy, the needle is advanced deep into the coccyx, closely approximating its anterior surface, until the tip reaches the level of the sacrococcygeal junction (*Fig. 5.7.4*).
- After negative aspiration, the fluoroscopic image is observed as a small amount of non-ionic radio-contrast medium is injected. The correct placement of the needle is indicated by the presence a small round blob of contrast medium at the anterior border of the vertebral column (*Fig. 5.7.5 a,b*).
- Lidocaine (lignocaine) 1% 5 ml is injected for ganglion blockade.

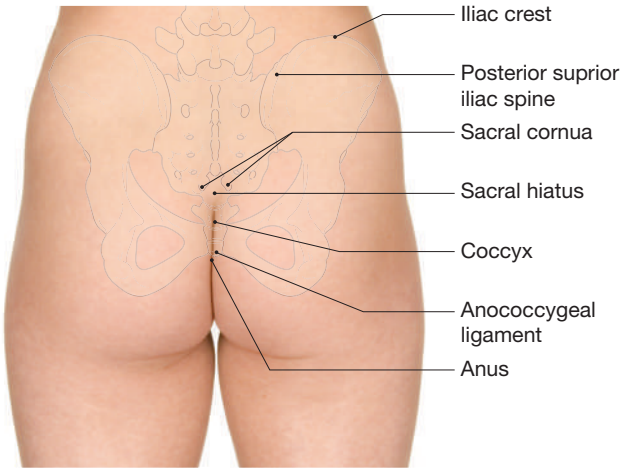

Fig. 5.7.3

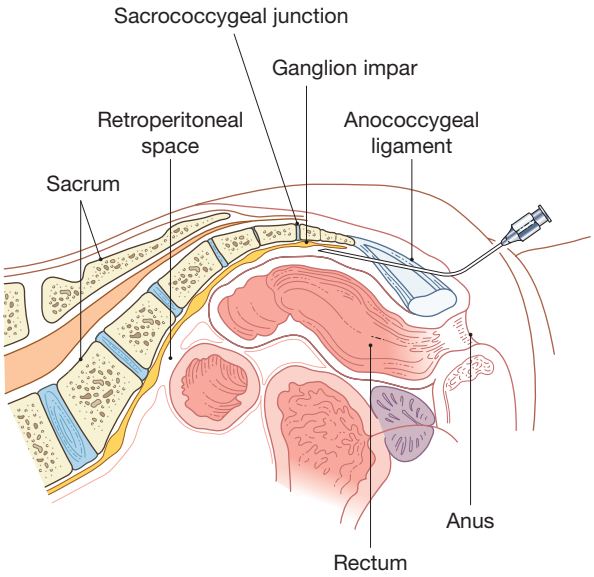

Fig. 5.7.4

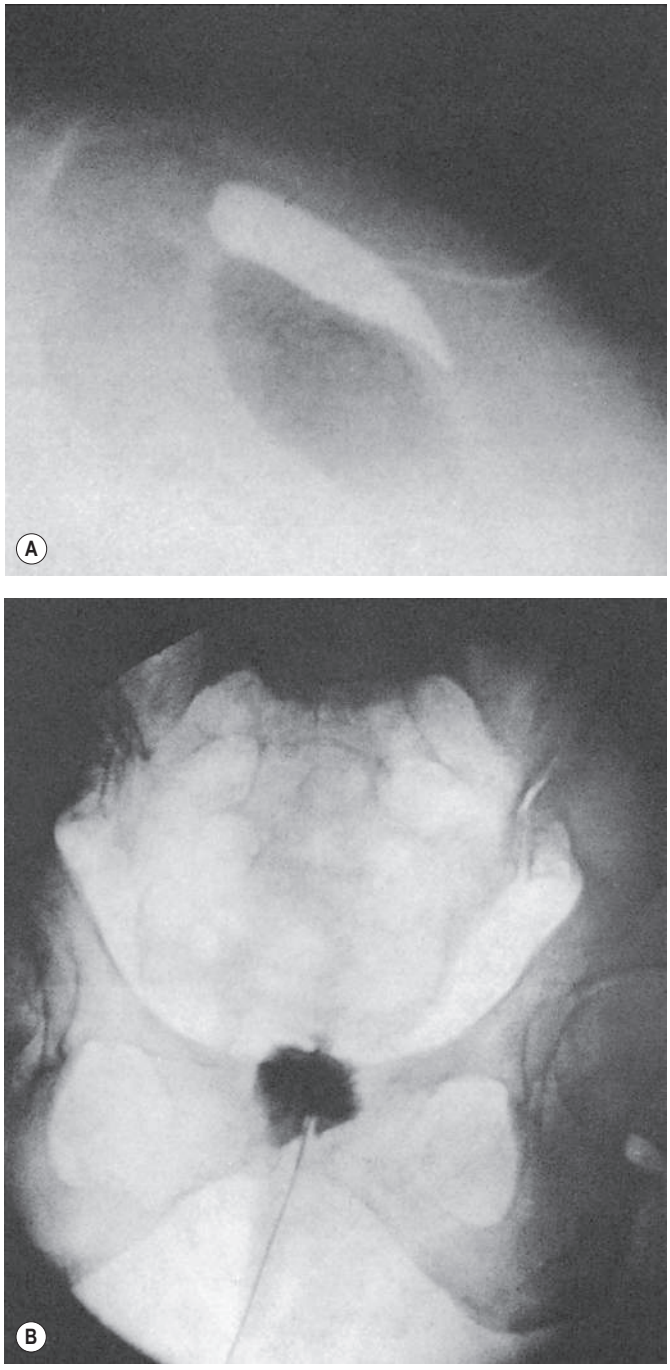

Fig. 5.7.5

- After 10 minutes the patient is questioned about pain relief and any somatic blockade. Sensory and motor functioning of the lower extremities is checked. The procedure should be abandoned if there is evidence of somatic blockade.
- After confirmation of pain relief and lack of somatic block, 5 ml of phenol 6% is injected using a glass syringe. To avoid leaving alcohol in the needle tract the needle is then cleared with air or local anesthetic 1 ml, and removed.
- Monitors should be left attached and i.v. access left in situ for at least 30 minutes.

Confirmation of a successful block

- Relief of perineal pain.

Tips

- To aid access to the anococcygeal ligament an assistant may be asked to retract the skin of the buttock; after penetration of the skin, this is no longer required.
- Exaggerated anterior curvature of the sacrococcygeal vertebral column may inhibit access and it may be necessary to bend the needle to a more acute angle.

Potential problems

- The position of each needle tip should always be confirmed with fluoroscopy prior to injection of neurolytic agent as it may lie in the peritoneal cavity, within a viscus or intravascularly. Caudal epidural placement of the needle is possible, therefore it is essential that spread of contrast material is observed to be restricted to the retroperitoneum, and that a test dose produces no somatic nerve blockade. Perforation of the rectum or periosteal injection is also possible.
- Local tumor invasion may inhibit spread of solution.
- Retroperitoneal hematoma may occur and the block should be avoided in patients with coagulopathy. Diarrhea may occur also and hydration of the patient should be monitored.
- There may be a detectable odor from the breath after alcohol injection.
- Patients in the prone position should be monitored carefully when intravenous sedation is administered.

5.8 INTRAVENOUS REGIONAL SYMPATHETIC BLOCK—UPPER LIMB

Anatomy

Peripheral sympathetic blockade is achieved by limiting the effect of the sympatholytic agent to the tissues of the affected limb using a tourniquet. Intravenous injection of an agent that releases endogenous norepinephrine (noradrenaline) from sympathetic nerve endings causes depletion of this neurotransmitter, and thereby chemical sympathetic blockade.

Equipment

- 20 ml syringe
- Two i.v. cannulae
- Pneumatic tourniquet
- ECG, BP, and SpO_2 monitors
- Skin temperature monitor
- Resuscitation equipment (*see Appendix 3*)

Drugs

- Lidocaine (lignocaine) 0.5% *without epinephrine/adrenaline*, or its equivalent
- Bretylium 1.5 mg/kg (or its equivalent, e.g. guanethedine 0.25 mg/kg)
- Saline (NaCl) 20 ml
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Supine.

Technique

- Intravenous access is inserted in the contralateral limb.
- Peripheral i.v. access is inserted in the limb to be blocked (*Fig. 5.8.1 a*).
- Monitors are attached.

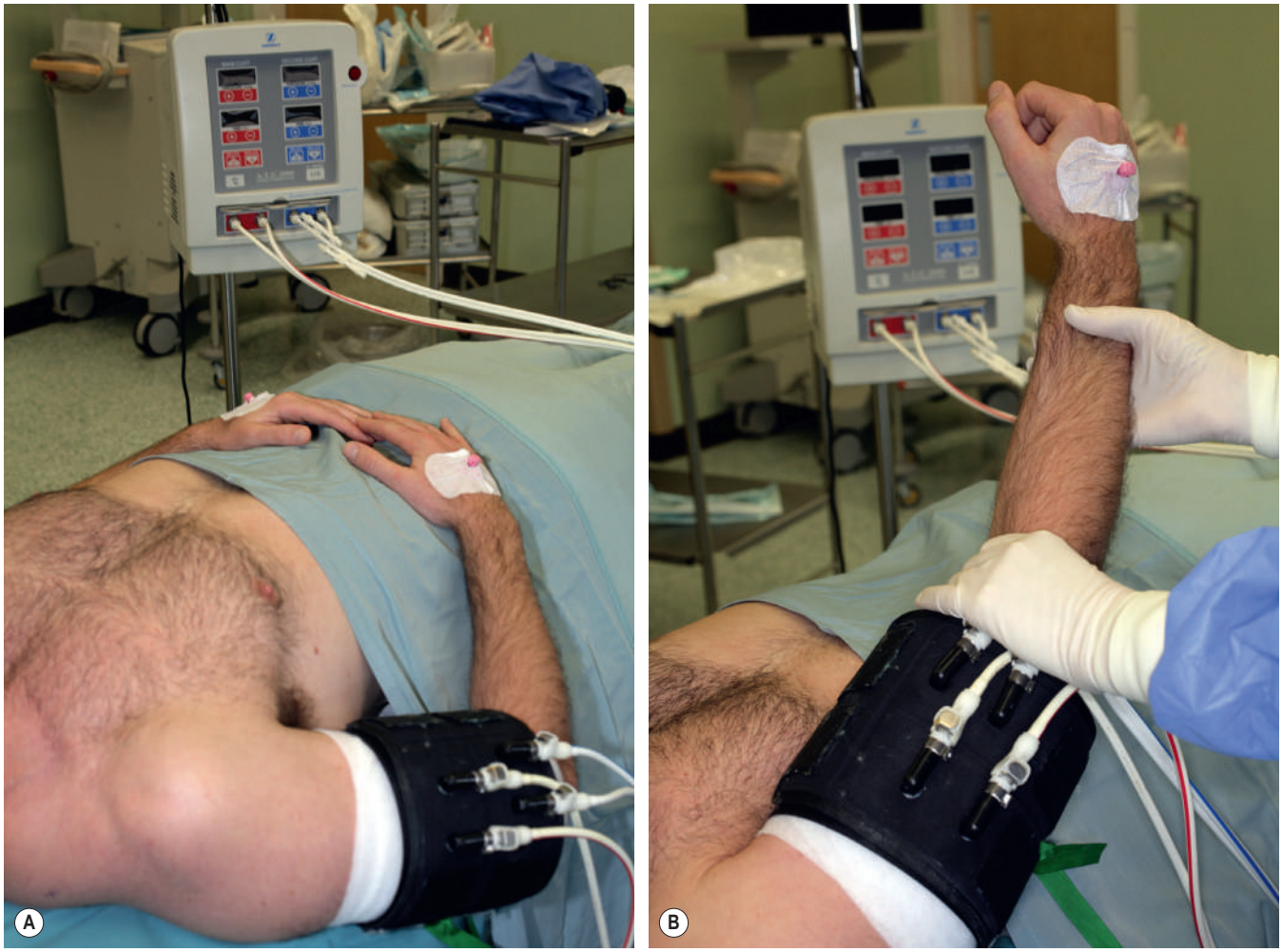

Fig. 5.8.1

Fig. 5.8.2

- Resuscitation equipment and drugs are checked and made ready for use.
- The limb is raised above the level of the heart for 2 minutes (*Fig. 5.8.1 b*).
- With the limb raised, it is exsanguinated by applying a tight wrap, e.g. Esmarch bandage.
- A thin layer of padding is applied, e.g. Velband, under the tourniquet site.
- The tourniquet is applied and the cuff is inflated to a pressure 100 mmHg higher than the systolic blood pressure (*Fig. 5.8.2*).
- The limb is then lowered. A mixture of lidocaine (lignocaine) 0.5% 15 ml (*without epinephrine/adrenaline*), bretylium 1.5 mg/kg (or guanethidine 0.25 mg/kg), and NaCl to make a total volume of 40 ml (a final lidocaine (lignocaine) solution of 0.25%), is injected through the i.v. cannula in the affected limb.
- The tourniquet is allowed to remain inflated for at least 30 minutes.

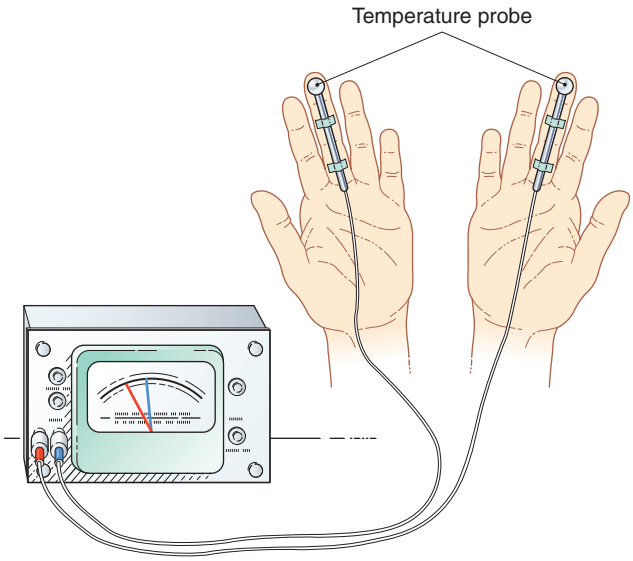

Fig. 5.8.3

- It is then deflated in one step, but left in place. Re-inflation may be required if there is a precipitous change in blood pressure.
- Monitors should be left attached and i.v. access left in situ for at least 30 minutes.

Confirmation of a successful block

- Relief of sympathetically maintained pain.
- Measurements of skin temperature of the affected limb before and after the block should demonstrate temperature increase. However, the sympatholytic effect of the drug may not be immediate (*Fig. 5.8.3*).

Tips

- If i.v. access to the affected limb is difficult due to vasoconstriction, a smear of glycerol trinitrate cream on the dorsum of the hand will usually aid i.v. insertion.
- A single or double cuff may be employed for this block but a double tourniquet may make the block more comfortable. The proximal cuff is inflated first. A few minutes after injection the distal cuff is inflated and **when inflation is complete the proximal cuff is released**.
- Retrograde cannulation, i.e. towards the periphery (*Figs 5.8.4, 5.8.5*) rather than proximally (*Figs 5.8.6, 5.8.7*), may help direct the spread of bretylium to the periphery.
- Active or passive movements of the limb may hasten the distribution of bretylium to the periphery.
- If the tourniquet inflation is painful, inhalation of nitrous oxide–oxygen mixture may improve comfort.
- Repeated blocks may bring about gradual improvement in sympathetically maintained pain.

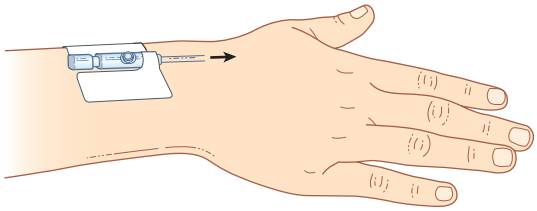

Fig. 5.8.4

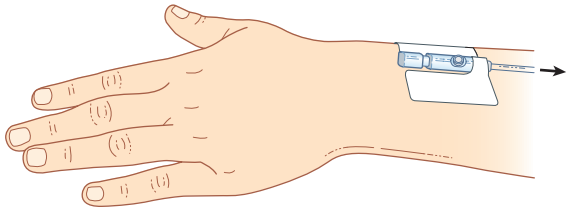

Fig. 5.8.6

Fig. 5.8.5

Fig. 5.8.7

- Immediate physiotherapy after block may improve outcome.

Potential problems

- Accidental deflation of the tourniquet early in the procedure may cause a precipitous rise in blood pressure due to the general release of endogenous norepinephrine/noradrenaline when unfixed bretylium enters the circulation. Systemic toxicity of lidocaine (lignocaine) may also occur, possibly causing seizures. Blood pressure may decrease after deflation of the cuff later in the procedure.
- The tourniquet inflation may be painful.
- A sensation of burning may occur after injection due to release of endogenous norepinephrine.
- Neuropraxia may occur (rarely) with a very tight tourniquet.
- Avoid in sickle cell anemia.

5.9 INTRAVENOUS REGIONAL SYMPATHETIC BLOCK—LOWER LIMB

Anatomy

As in the case of the upper limb, peripheral sympathetic blockade is achieved by limiting the effect of the sympatholytic agent to the tissues of the affected limb using a tourniquet. Intravenous injection of an agent releases endogenous norepinephrine (noradrenaline) from sympathetic nerve endings, which causes depletion of this neurotransmitter, and may bring about a chemical sympathetic block.

Equipment

- 50 ml syringe
- Two i.v. cannulae
- Pneumatic tourniquet
- ECG, BP, and SpO_2 monitors
- Skin temperature monitor
- Resuscitation equipment (*see Appendix 3*)

Drugs

- Lidocaine (lignocaine) 0.5% *without epinephrine (adrenaline)*, or its equivalent
- Bretylium 1.5 mg/kg (or its equivalent, e.g. guanethedine 0.5 mg/kg)
- Saline (NaCl) 30 ml
- Resuscitation drugs (*see Appendix 3*)

Position of patient

- Supine.

Technique

- Intravenous access is inserted in the contralateral limb.
- Peripheral i.v. access is inserted in the limb to be blocked.
- Monitors are attached.
- Resuscitation equipment and drugs are checked and made ready for use.
- The limb is raised above the level of the heart for 2 minutes (*Fig. 5.9.1*).
- With the limb raised, it is exsanguinated by applying a tight wrap (*Fig. 5.9.2*).
- A thin layer of padding is applied, e.g. Velband, under the tourniquet site.
- The tourniquet is applied and inflated to a pressure 100 mmHg higher than the systolic blood pressure. A second tourniquet may be applied to the calf of patients with no known predispositions to deep venous thrombosis. Inflation of this second cuff may aid limitation of spread of sympatholytic agent to the periphery (*Fig. 5.9.3*).
- The limb is then lowered and a mixture of lidocaine (lignocaine) 0.5% 25 ml (*without epinephrine*), bretylium 1.5 mg/kg (or its equivalent) and NaCl to make a total volume of 40 ml (a final lidocaine/lignocaine solution of 0.25%) is injected through the i.v. cannula in the affected limb.
- The tourniquet is allowed to remain inflated for at least 30 minutes.
- It is then deflated in one step, but left in place. Re-inflation may be required if there is a precipitous change in blood pressure.

Fig. 5.9.1

Fig. 5.9.2

- Monitors should be left attached and i.v. access left in situ for at least 30 minutes.

Confirmation of a successful block

- Relief of sympathetically maintained pain.
- Measurements of skin temperature of the affected limb before and after the block should demonstrate temperature increase (*Fig. 5.9.4*).

Tips

- If i.v. access to the affected limb is difficult due to vasoconstriction, a smear of glycerol trinitrate cream on the dorsum of the foot will usually aid i.v. insertion.
- A single or double cuff may be employed for this block but a double tourniquet may make the block more comfortable. The proximal cuff is inflated first. A few minutes after injection the distal cuff is inflated and when inflation is complete the proximal cuff is released.
- Retrograde cannulation, i.e. towards the periphery (*Fig. 5.9.5*) rather than proximally (*see Fig. 5.9.3*), may help direct the spread of bretylium to the periphery.
- Active or passive movements of the limb may hasten the distribution of bretylium to the periphery.
- If the tourniquet inflation is painful, inhalation of nitrous oxide–oxygen mixture may improve comfort.
- Repeated blocks may bring about gradual improvement in sympathetically maintained pain.
- Immediate physiotherapy after block may improve outcome.

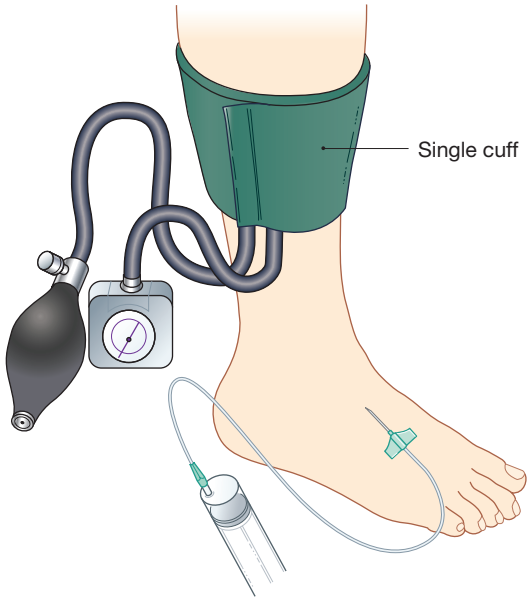

Fig. 5.9.3

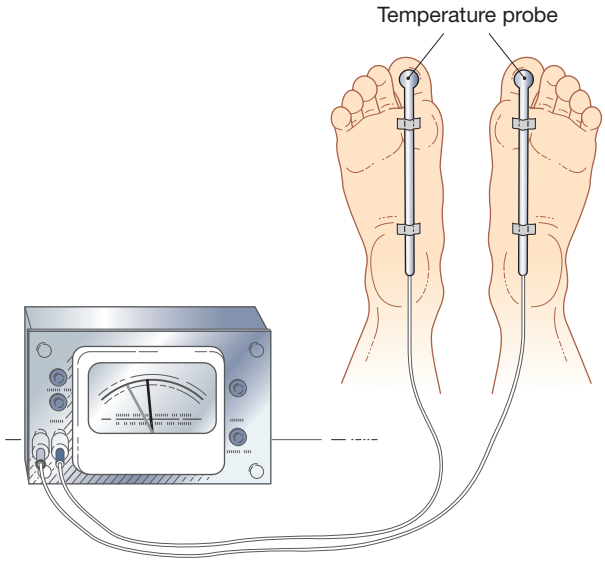

Fig. 5.9.4

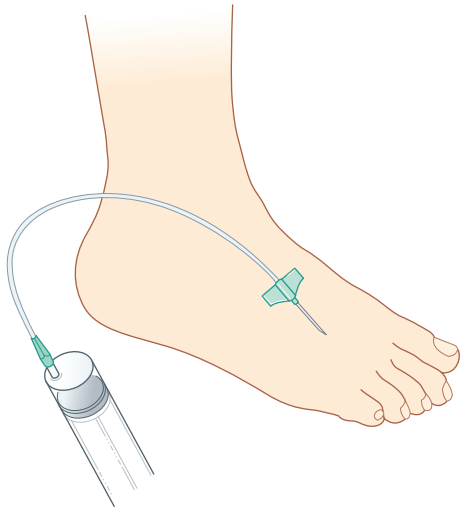

Fig. 5.9.5

Potential problems

- Accidental deflation of the tourniquet early in the procedure may cause a precipitous rise in blood pressure due to the general release of endogenous norepinephrine or epinephrine when unfixed bretylium enters the circulation. Systemic toxicity of lidocaine (lignocaine) may also occur in high doses possibly causing seizures. Blood pressure may decrease after deflation of the cuff later in the procedure.
- The tourniquet inflation may be painful.
- A sensation of burning may occur after injection due to release of endogenous norepinephrine.
- Neuropraxia may occur (rarely) with a very tight tourniquet.
- Avoid in sickle cell anemia.

MUSCLE INJECTIONS 6

Myofascial pain occurs commonly in the muscles of the upper and lower back. It is characterized by pain associated with movement of the affected muscles that develop areas of extreme tenderness, termed trigger points. Palpation of these points is usually perceived as a tight band or firm nodule in the muscle and reproduces pain that may be referred some distance from the site of palpation. Involuntary muscular contraction can occur on palpation, and snapping palpation can result in a local twitch response. Electromyography (EMG) is not reliable in diagnosing myofascial pain syndrome and it is worth remembering that this syndrome may occur in association with underlying painful disorders of the spine.

Injection of trigger points with local anesthetic, especially if repeated several times and combined with stretching exercises, may have a beneficial therapeutic effect on the pain of myofascial pain syndrome. Pain reproduction during injection, followed by relief of pain after injection, that lasts at least as long as the expected local anesthetic effect, indicates that these painful points contribute to myofascial pain syndrome.

Fibromyalgia is a pain syndrome characterized by widespread, diffuse and usually symmetrical tender areas of muscles. Bony structures, such as costochondral junctions and lateral epicondyles, produce local pain, but not referred, on palpation of tender points. Injection of these tender areas typically does not improve the pain of fibromyalgia.

Usually a dilute solution of local anesthetic suffices for beneficial effect. Bupivacaine produces more muscle degeneration than any other local anesthetic when injected into a muscle, and consequently it is usually avoided, lidocaine (lignocaine) being the usual local anesthetic of choice.

The optimum number of trigger-point injections required to produce pain relief is variable. The injection sites may themselves be painful after the local anesthetic wears off. This may exacerbate muscle spasm if too many trigger-point injections are performed. Consideration should be given to the severity of the muscle spasm, the number of trigger points, and to the sensitivity of the patient to pain when deciding on the number of injections.

6.1 TRIGGER-POINT INJECTIONS—NECK AND THORAX

Anatomy

The muscles most often involved in myofascial pain syndrome of the neck include the trapezius, rhomboid minor and major, latissimus dorsi, levator scapulae and splenius capitis (**Fig. 6.1.1**; see also Fig. 2.1.2).

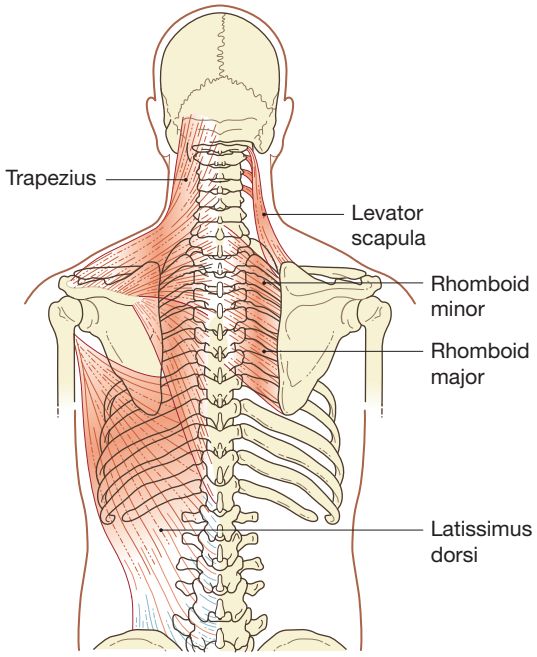

Fig. 6.1.1

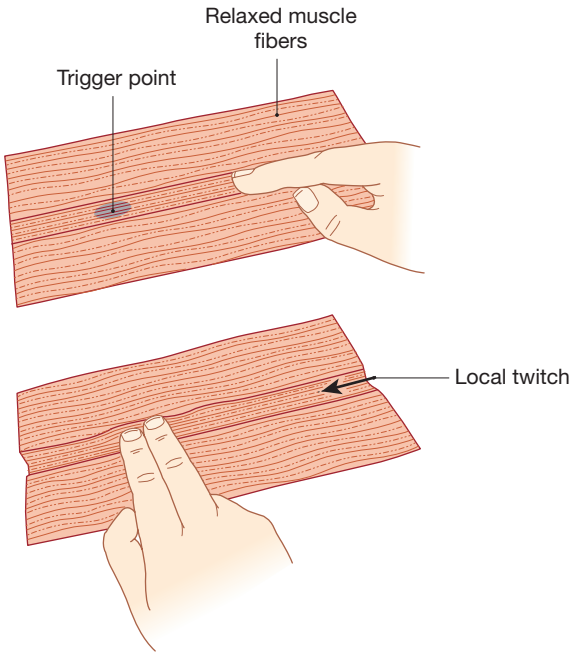

Fig. 6.1.2

Equipment

- 10 ml syringe
- 25 G needle

Drugs

- Lidocaine (lignocaine) 1% 10 ml

Position of patient

- Prone.
- Pillow under chest to allow the neck to flex.
- The sitting position is also used, but vasovagal response may follow trigger-point injections especially in young adults, and it is probably more prudent to use the prone position.

Needle puncture and technique

- The neck, shoulders, and upper posterior thorax are cleaned with antiseptic solution.
- **Trigger points in the muscles are palpated** (**Fig. 6.1.2**) **and marked** (*Fig. 6.1.3*).
- A 25 G needle with syringe attached is inserted into a trigger point (*Fig. 6.1.4*).
- After negative aspiration, 2–3 ml of lidocaine (lignocaine) 1% is injected into the trigger point while moving the needle back and forth through the muscle.
- After injection, the next trigger point is injected in the same manner (*Fig. 6.1.5 a,b*).

Confirmation of a successful injection

- Pain reproduction when the needle enters the muscle confirms correct placement.

Fig. 6.1.3

Fig. 6.1.4

Tips

- For best results, injection is carried out in a fan-like manner by repeatedly withdrawing the needle slightly and redirecting it.
- Stretching of the involved muscles by physiotherapy within the duration of the local anesthesia improves results.
- Some workers advocate massage of the area immediately after injection.

Potential problems

- Pain on injection.
- Vasovagal response (especially in young adults in the sitting position).
- Pneumothorax (especially in thin patients).

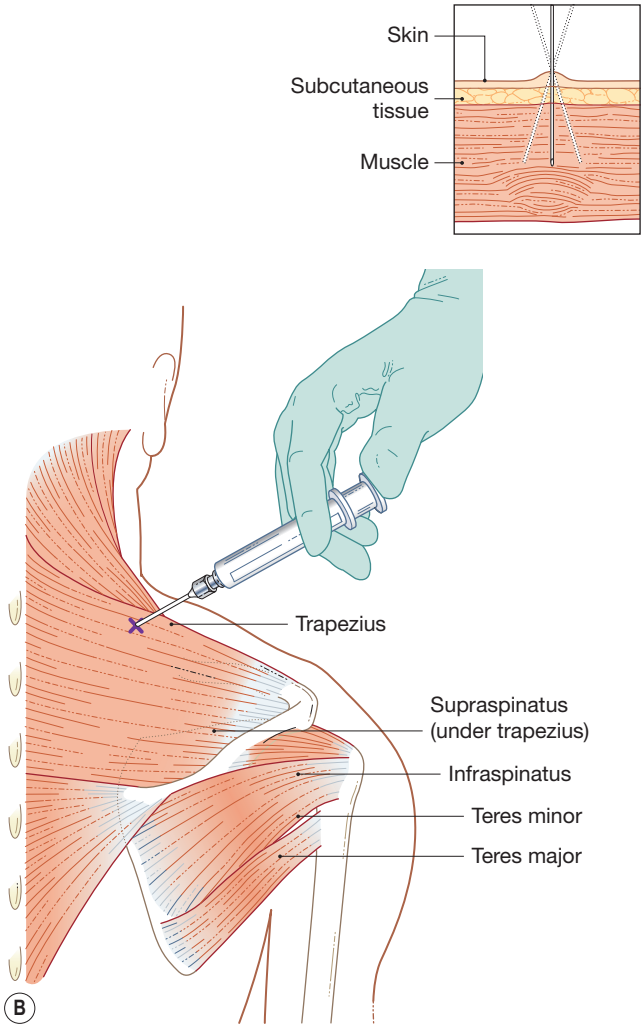

Fig. 6.1.5

6.2 TRIGGER-POINT INJECTIONS—BACK

Anatomy

The muscles most often involved in myofascial pain syndrome of the back include the erector spinae (the longissimus, iliocostalis, and spinalis columns) and the deep transversospinal (semispinalis, multifidus, and rotatores) muscles (**Fig. 6.2.1 a,b**). In the buttocks, spasm of the gluteus medius muscle may also cause significant pain.

Equipment

- 10 ml syringe
- 25 G needle

Drugs

- Lidocaine (lignocaine) 1% 10 ml

Position of patient

- Prone.
- Pillow under abdomen to straighten the normal lumbar lordosis (*Fig. 6.2.2 a*).
- The sitting position is also used, but vasovagal response may follow trigger-point injections, and it is probably more prudent to use the prone position.
- Alternatively, the semiprone position will also allow access to affected muscles (*Fig. 6.2.2 b*).

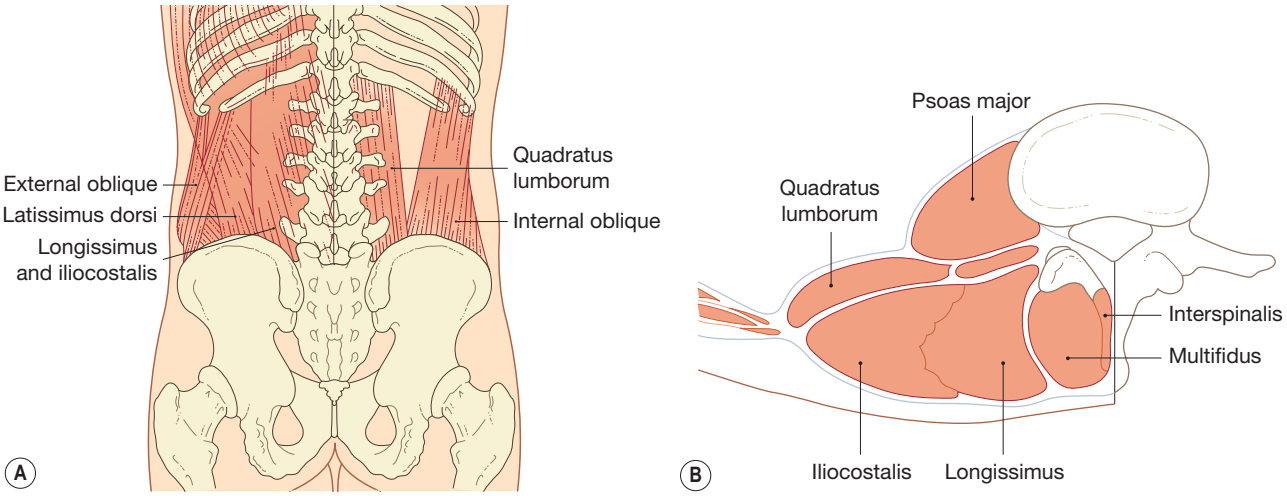

Fig. 6.2.1

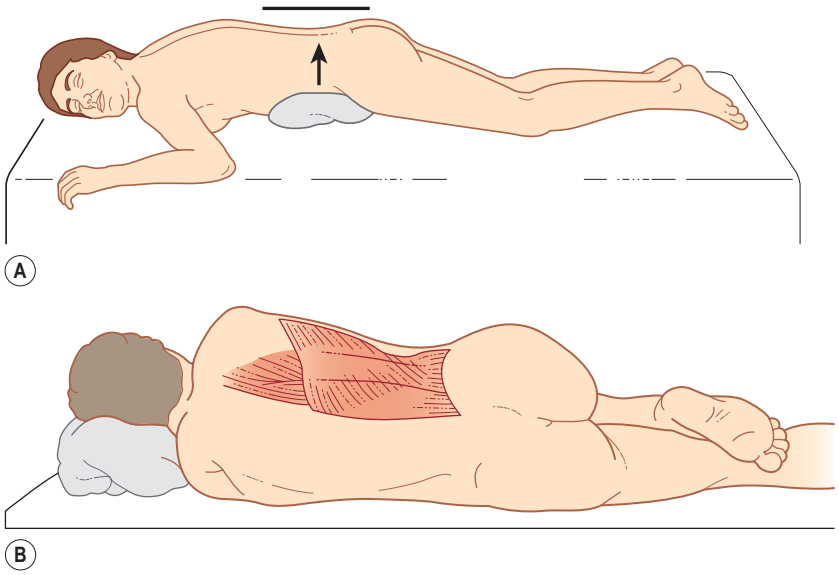

Fig. 6.2.2

Needle puncture and technique

- The midline and the surrounding area are cleaned with antiseptic solution.
- **Trigger points in the muscles are identified by palpation and marked** (*Fig. 6.2.3*).
- A 25 G needle with syringe attached is inserted into a trigger point (*Fig. 6.2.4*).
- After negative aspiration, 2–3 ml of lidocaine (lignocaine) 1% is injected into the trigger point while moving the needle back and forth through the muscle (*Fig. 6.2.5*).

Fig. 6.2.3

Fig. 6.2.4

- After injection, the next trigger point is injected in the same manner.

Confimation of a successful injection

- Pain reproduction when the needle enters the muscle confirms correct placement.

Tips

- For best results, injection is carried out in a fan-like manner by repeatedly withdrawing the needle slightly and redirecting it.
- Stretching of the involved muscles by physiotherapy within the duration of the local anesthetic improves results.
- Some workers advocate massage of the area immediately after injection.

Potential problems

- Pain on injection.
- Vasovagal response (especially young adults in the sitting position).
- Pneumothorax (especially in thin patients) is also a possibility when injecting the upper back.

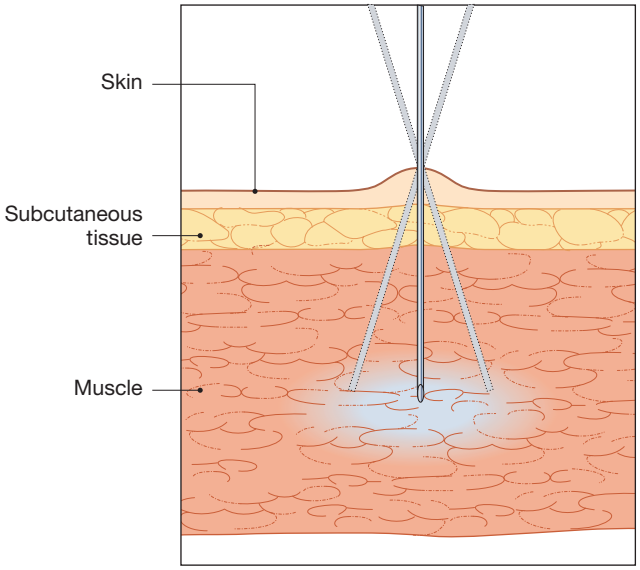

Fig. 6.2.5

6.3 GLUTEUS MEDIUS INJECTION

Anatomy

When the buttock muscles are relaxed the quadratus femoris, gemelli and gluteus medius muscles can be palpated. Spasm of the gluteus medius muscle (**Fig. 6.3.1**) may be the source of buttock pain and may respond to trigger-point injection. Unlike pirformis muscle spasm, this does not produce symptoms of sciatic nerve irritation but causes localized pain, often referred to the posterior thigh and calf.

Equipment

- 10 ml syringe
- 22 G needle

Drugs

- Lidocaine (lignocaine) 1% 10 ml

Position of patient

- Prone.
- Pillow under abdomen to flatten the normal lumbar lordosis (*Fig. 6.3.2*).

Needle puncture and technique

- The surface of the buttock and hip is cleaned with antiseptic solution.
- The posterior superior iliac spine is palpated and marked.
- The greater trochanter is palpated.
- **The insertion point of the needle lies approximately 2 cm medial and superior to the greater trochanter** (*Figs 6.3.3, 6.3.4*).
- A 22 G needle is introduced in a direction vertical to the skin and advanced until it is felt to be gripped by the tense muscle (*Fig. 6.3.5*).

Fig. 6.3.1

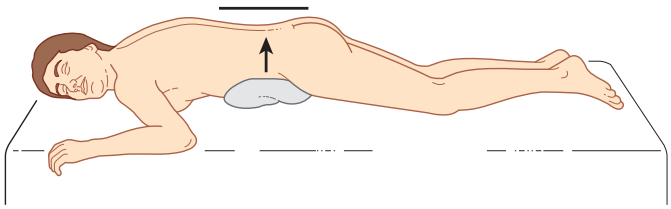

Fig. 6.3.2

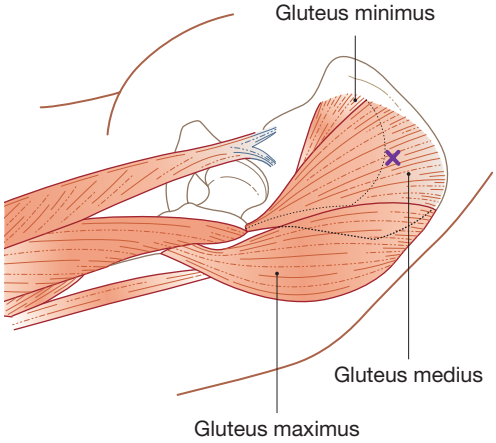

Fig. 6.3.3

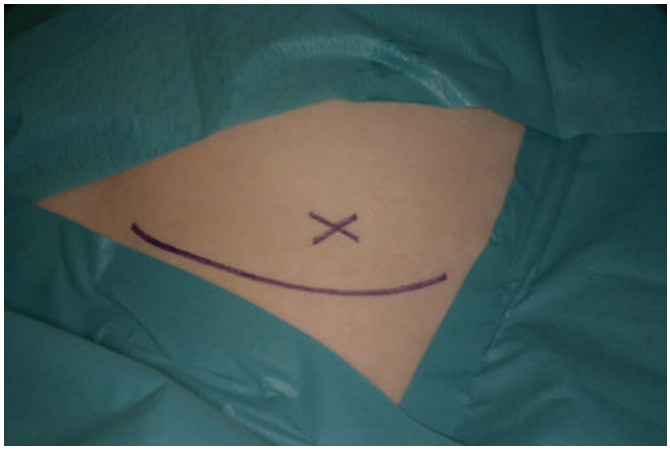

Fig. 6.3.4

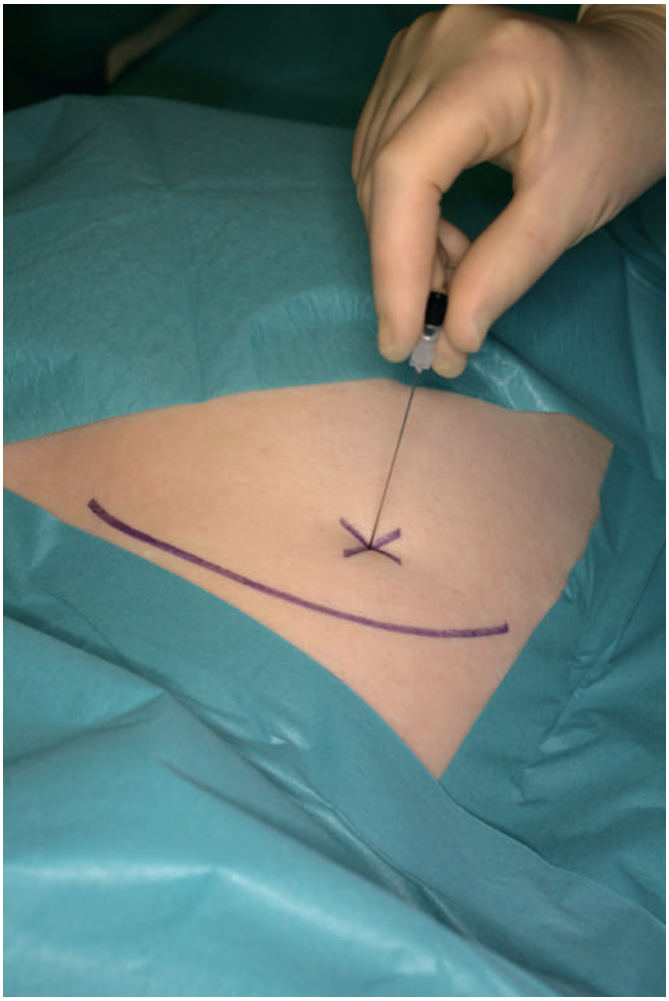

Fig. 6.3.5

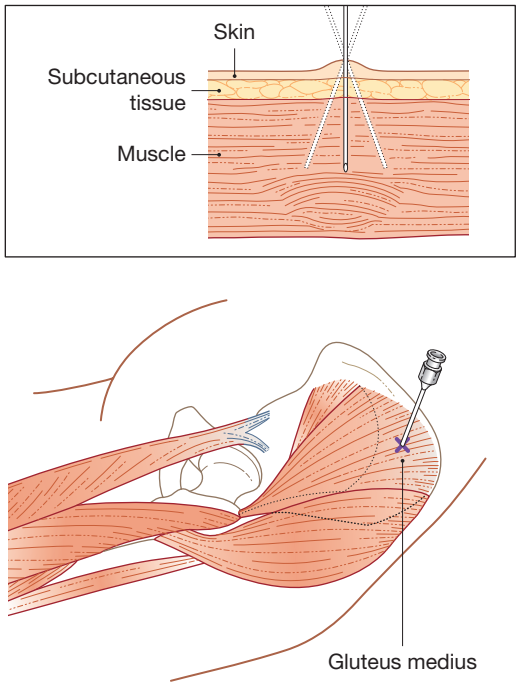

Fig. 6.3.6

- After negative aspiration, lidocaine (lignocaine) 1% 3 ml is injected in the substance of the muscle while moving the needle back and forth in the muscle (*Fig. 6.3.6*).
- The procedure is repeated if other trigger points are present in the muscle.

Confirmation of a successful injection

- Pain reproduction when the needle enters the muscle confirms correct placement of the needle.
- Relief of pain on abduction of the hip.

Tips

- For best results injection is carried out in a fan-like manner by repeatedly withdrawing the needle slightly and redirecting it (*see inset in Fig. 6.3.6*).

Potential problems

- Sciatic nerve block: although this is unusual because the injection site is not very close to the sciatic notch, it is prudent to warn the patient of the possibility.
- Infection or abscess formation.

6.4 PIRIFORMIS INJECTION

Anatomy

The piriformis muscle inserts into the pelvic surface of the sacrum from the second to the fourth segments, lateral to the anterior sacral foramina, and passes out of the pelvis through the greater sciatic foramen to insert into the superior aspect of the greater trochanter (**Fig. 6.4.1**). It overlies the sciatic nerve in the greater sciatic foramen. Contraction contributes to abduction of the lower limb. Spasm of the muscle in myofascial pain syndrome often causes pain referred to the posterior thigh and calf.

Equipment

- 2 ml syringe and two 5 ml syringes
- 25 G needle
- 22 G spinal needle, end-opening

Drugs

- Lidocaine (lignocaine) 1% 10 ml (or its equivalent)

Position of patient

- Prone.
- Pillow under abdomen to flatten the normal lumbar lordosis (*Fig. 6.4.2*).

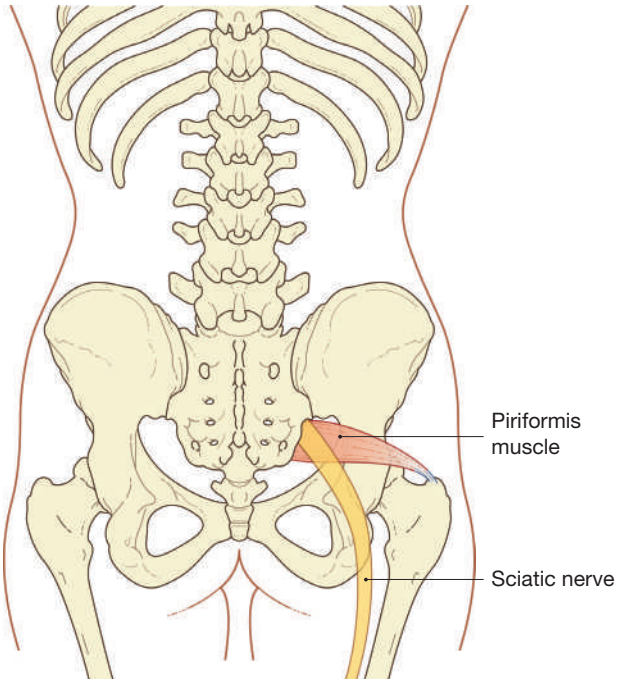

Fig. 6.4.1

Needle puncture and technique

- The surface of the buttock and hip is cleaned with antiseptic solution and a fenestrated drape is placed over the sterile area.
- The posterior superior iliac spine is palpated and marked.
- The greater trochanter is palpated and marked.
- **The insertion points of the needle lie at the points one-third and two-thirds along, and 1–3 cm below the line connecting these two marks** (*Figs 6.4.3 a,b,c*).
- The first insertion point, the medial one, is infiltrated with lidocaine (lignocaine) 1% 2 ml.
- A 22 G spinal needle is introduced in a direction vertical to the skin and advanced until it is felt to be gripped by the tense piriformis muscle, or until bone is contacted (*Figs 6.4.4, 6.4.5*).
- The end-point is a fascial click at a depth of about 4–5 cm, depending on the thickness of adipose tissue.
- The patient is questioned about the presence of pain, paresthesia, and changes in sensation in the distribution of the sciatic nerve, while the needle is being advanced.
- If these symptoms arise, the needle may be in contact with the sciatic nerve and should be repositioned. It is also possible that spasm of the muscle on needle insertion may produce these symptoms and often reproduction of pain occurs with entry into the muscle.
- After negative aspiration, lidocaine (lignocaine) 1% 5 ml is injected.
- Ultrasound may aid placement of the needle (*Fig. 6.4.6 a,b*)

Confirmation of a successful injection

- Relief of pain on abduction of the lower limb against pressure on the lateral knee in the sitting position.

Tips

- If injection is not successful in relieving the pain, it may be repeated at the lateral insertion point. This lies at a

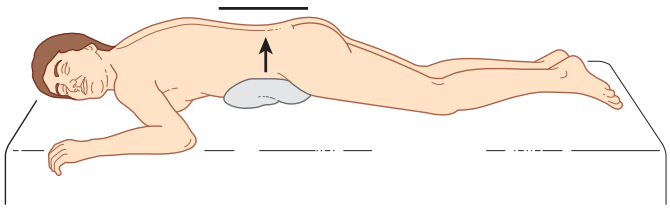

Fig. 6.4.2

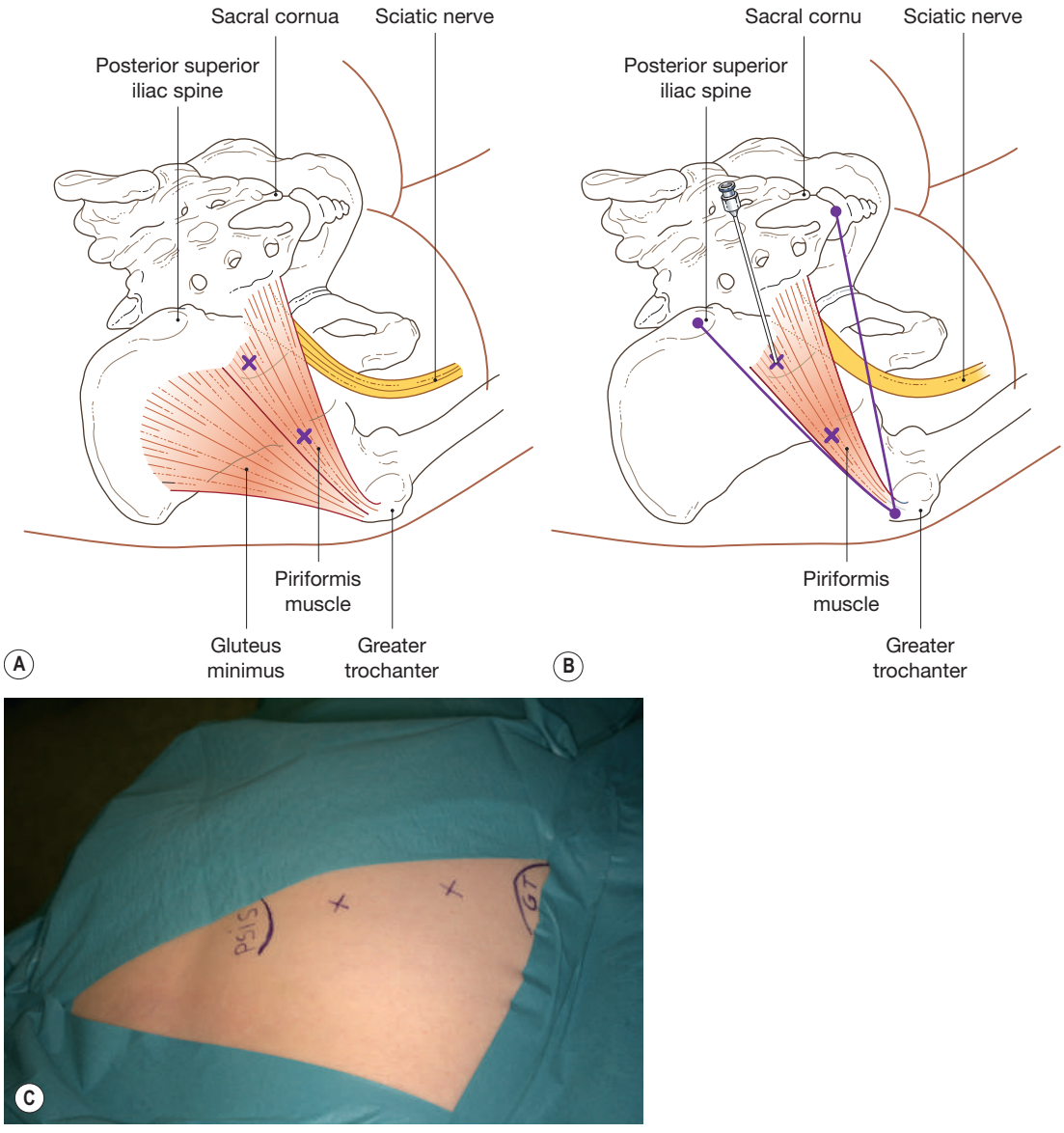
Sacral cornua
Sciatic nerve
Posterior superior iliac spine
Piriformis muscle
Gluteus minimus
Greater trochanter
A
Sacral cornu
Sciatic nerve
Posterior superior iliac spine
Piriformis muscle
Greater trochanter
B
C

Fig. 6.4.3

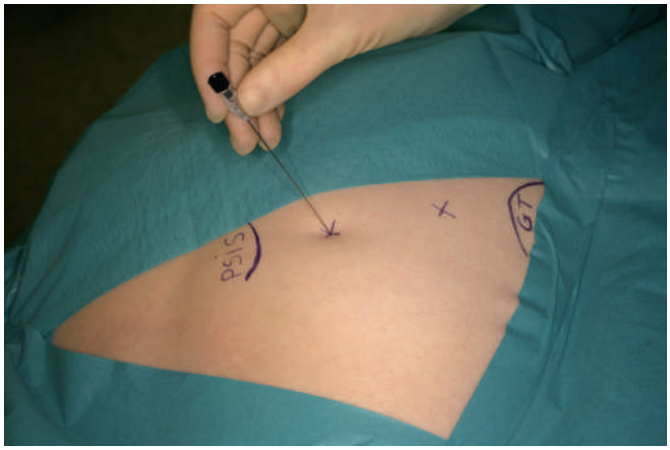

Fig. 6.4.4

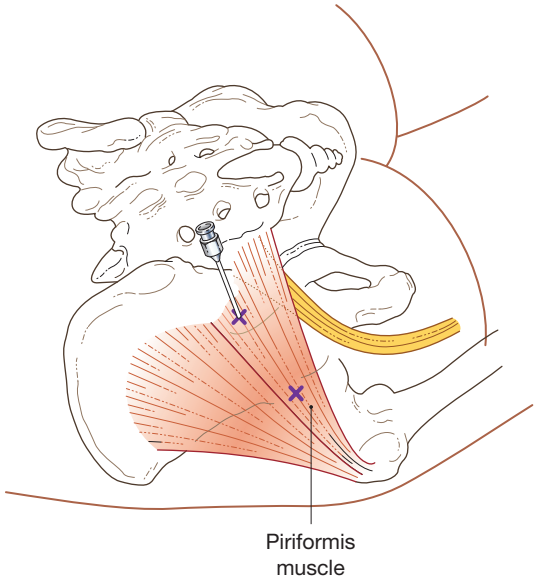

Fig. 6.4.5

point two-thirds along and 1–3 cm below the line joining the posterior superior iliac spine and the greater trochanter.

Potential problems

- Sciatic nerve block.
- Infection or abscess may occur (rarely).

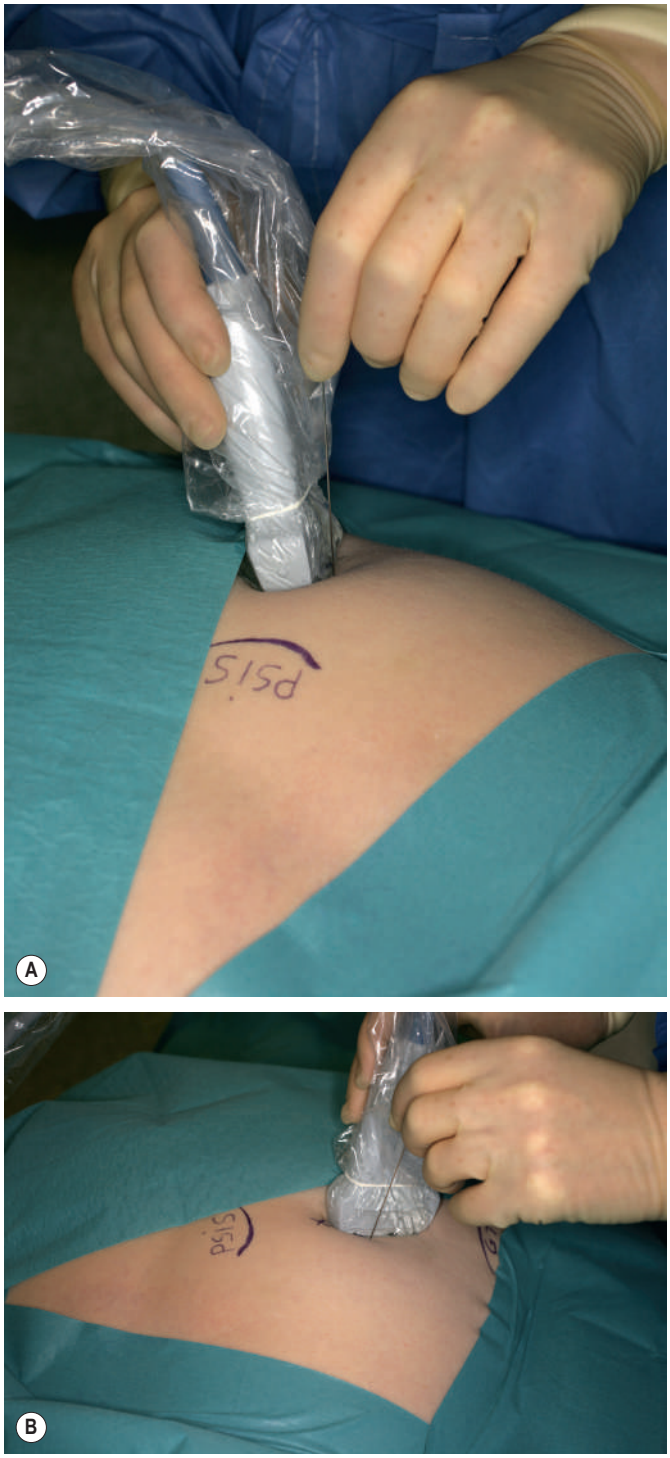

Fig. 6.4.6

TRANSCUTANEOUS ELECTRICAL NERVE STIMULATION (TENS)

7

Transcutaneous electrical nerve stimulation is thought to modify pain appreciation by stimulation of large fibers thereby blocking (or "closing the gate" to) smaller C-fibers carrying nociceptive impulses. There is also evidence that high-frequency stimulation of the skin increases latency and decreases maximum firing rates in small afferent fibers. This can produce conduction blockade in C-fibers as the current is increased, probably via potassium efflux from the axon. It is thought that a combination of these actions is responsible for the analgesia derived from the use of TENS. This is probably not related to opiate-mediated mechanisms when conventional parameters are used.

Not all pain responds to TENS. If the usual parameters do not produce pain relief, low frequency, high intensity stimulation may be tried. This means that the current amplitude is increased to a level that produces mild discomfort and muscle stimulation. Analgesia from this type of stimulation may be due to opiate-mediated mechanisms. Burst stimulation means short bursts of high frequency stimulation delivered at 1–2 Hz and may also relieve pain that is not responsive to conventional TENS.

A TENS trial may be carried out prior to giving the unit to the patient to use at home. This allows the patient to become familiar with the use of TENS, and to ensure that the pain is not aggravated by its use. A minimum of one hour is recommended as the trial period. This will indicate whether the patient is likely to respond to TENS. However, failure to respond within this time period does not necessarily mean that there will be no response if used for longer periods, or with different settings. It is important to allow the patient to use the TENS at home for a period of at least 14 days.

The TENS stimulator is a battery-operated pulse generator which has several controls. These include an on/off switch plus amplitude control, frequency control, mode selector, and width control. In addition, multichannel units have amplitude controls for each channel. The pulse generator connects to leads that then connect to electrodes, which are applied to the skin. Electrodes are applied in pairs, and are positioned so that they lie along the direction of the nerves in the area being treated, e.g. longitudinally in the limbs, but dermatomally in the trunk.

Control settings (*Figs 7.1.1, 7.1.2*)

CONTINUOUS STIMULATION

- Amplitude set to zero.
- Pulse width set to midrange.
- Switch to continuous mode.
- Increase pulse amplitude level gradually to the maximum level for comfort (sensation should be strong but not painful).
- Adjust pulse frequency to maximum level for comfort (amplitude may be reduced as pulse width is increased).
- Adjust pulse width to maximum level for comfort.
- Maintain for 45–60 minutes.

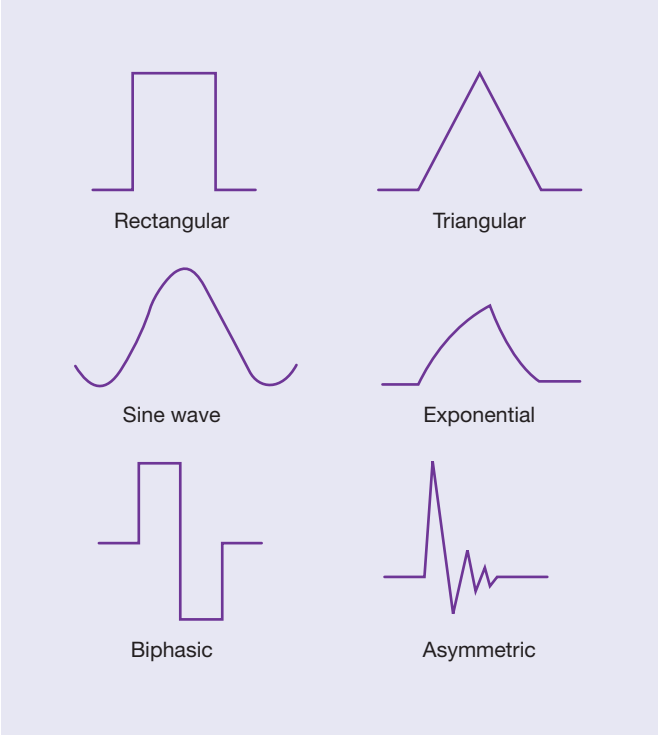

Fig. 7.1.1

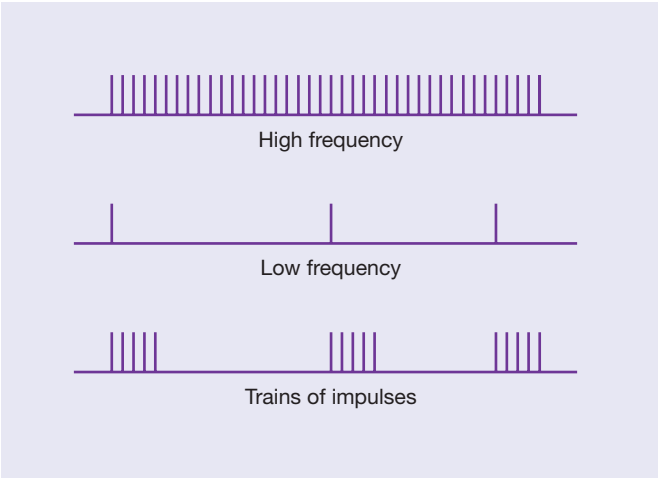

Fig. 7.1.2

Modulated settings

BURST STIMULATION

- All controls set to zero.
- Switch to pulsed mode.
- Frequency set to 1–2 Hz.
- Increase amplitude and pulse width as with continuous mode described above.
- Maintain for 45–60 minutes.

LOW-FREQUENCY HIGH-INTENSITY STIMULATION

- All controls set to zero.
- Increase amplitude to level where the muscle underlying the electrodes twitches visibly but not painfully.
- Increase frequency to 2–4 Hz.
- Maintain for short period (5–15 minutes).

All the types of stimulation should be tried for each pain, and the effects on the pain should be compared. The optimum parameters must be found by trial and error. The patient is usually advised to begin by using TENS for at least one hour three times a day. Once the effect of TENS on the pain is known it is recommended that stimulation should be discontinued after 30 minutes if the patient experiences one or more hours of analgesia from a single application. If pain relief is achieved only during stimulation, the unit can be kept on constantly. However, electrode sites should be changed every 24 hours. Occasionally, skin rash under the electrodes may occur and this problem may be minimized by frequent rotation of the electrode sites and with topical steroids. However, very few side effects are associated with the use of TENS. Electrical skin burns may occur if TENS is applied to skin with poor innervation and it is necessary to ensure that there is normal sensation prior to applying the electrodes. Allergic reaction to the electrodes or the adhesive tape has also been described, but is not common.

Use of TENS is contraindicated on areas over the anterior neck (stimulation of carotid sinus, larynx), over the pregnant uterus or in the presence of a cardiac pacemaker.

Note: Description of the insertion technique of a spinal cord stimulator or peripheral nerve stimulator is outside the scope of this text as the techniques are specific to the different types of stimulator.

APPENDIX 1

SUGGESTED CORTICOSTEROIDS

Drug name	Duration of action	Equivalent dosage (mg)	Anti-inflammatory potency (relative)	Mineralocorticoid potency (relative)
Triamcinolone[a]	12–36 h	4	5	0
Methylprednisolone	12–36 h	4	5	0.5
Dexamethasone	48 h	0.75	25	0
Hydrocortisone	12 h	20	1	2

[a]*Triamcinolone diacetate recommended for central neuroaxial injections.*

Corticosteroid injection side effects

LOCAL SIDE EFFECTS

- Atrophy of subcutaneous tissue
- Rupture of injected tendon
- Depigmentation of skin
- Infection

SYSTEMIC SIDE EFFECTS (HIGHER INCIDENCE WITH LARGER DOSES)

- Skin flushing
- Irregularity of the menstrual cycle
- Impaired glucose tolerance
- Osteoporosis
- Muscle wasting and myopathy
- Arthropathy
- Suppression of adrenal function
- Psychologic upset

APPENDIX 2

SUGGESTED NEUROLYTIC AGENTS

- Aqueous phenol 6%.
- Alcohol 100% may be diluted to 50% (pain on injection may be experienced and it is recommended that the nerve is blocked with local anesthetic prior to injection).

APPENDIX 3

RECOMMENDED RESUSCITATION DRUGS AND EQUIPMENT

SUGGESTED RESUSCITATION DRUGS

Drug	Suggested dosage (70 kg adult)	Indication
Atropine	0.2–0.4 mg i.v. increments	Bradycardia from vagal dominance
Ephedrine	5–10 mg i.v. increments	Hypotension from sympathetic block
Lidocaine (lignocaine)	50–100 mg i.v. bolus	Ventricular arrhythmias
Midazolam	1–3 mg i.v. increments	Local anesthetic; seizure activity
Diazepam	2.5–5 mg i.v. increments	Local anesthetic; seizure activity
Thiopental (thiopentone)	50–100 mg i.v. increments	Local anesthetic; seizure activity
Succinylcholine	50–100 mg i.v. bolus	Muscle relaxation; airway control

It is also recommended that the full range of drugs required for advanced cardiac life support (ACLS), including pre-filled syringes, be available in the operating room.

Suggested resuscitation equipment

- Oxygen source
- Bag and masks (full range)
- Breathing system for positive pressure ventilation
- Oro- and nasopharangeal airways (full range)
- Laryngoscopes and blades (full range)
- Endotracheal tube stylets and forceps, e.g. Magill's forceps

APPENDIX 4

DERMATOMES

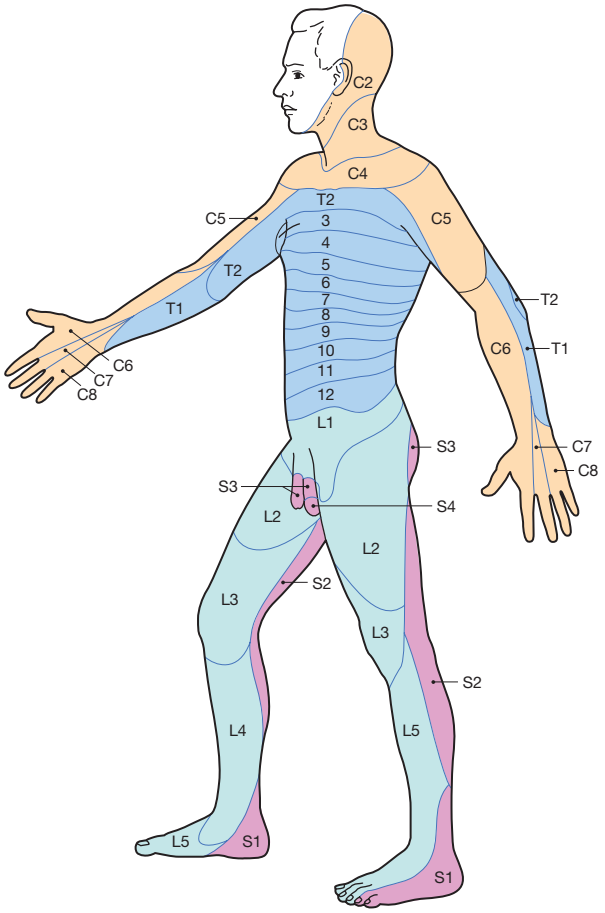

Fig. A.4.1

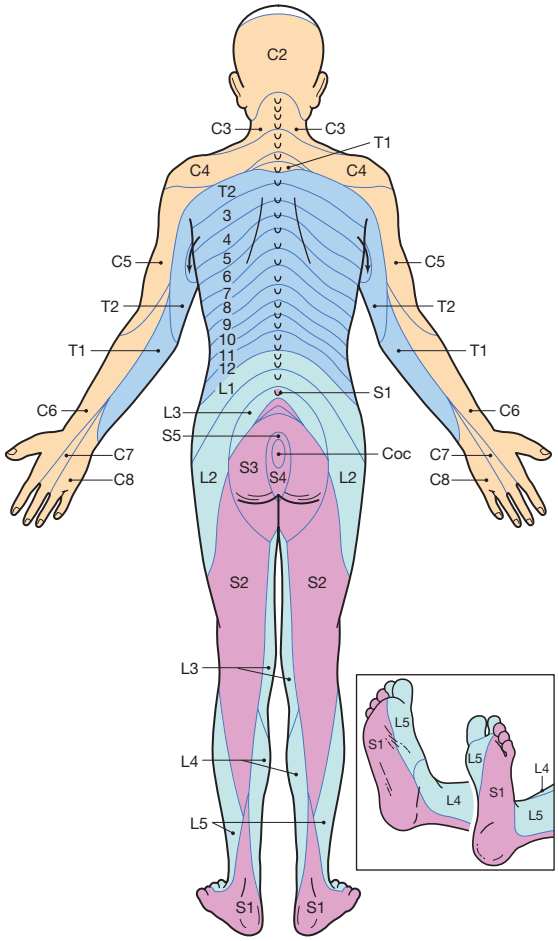

Fig. A.4.2

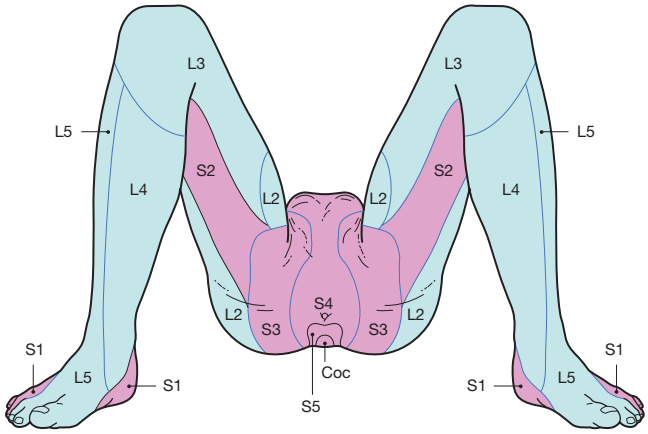

Fig. A.4.3

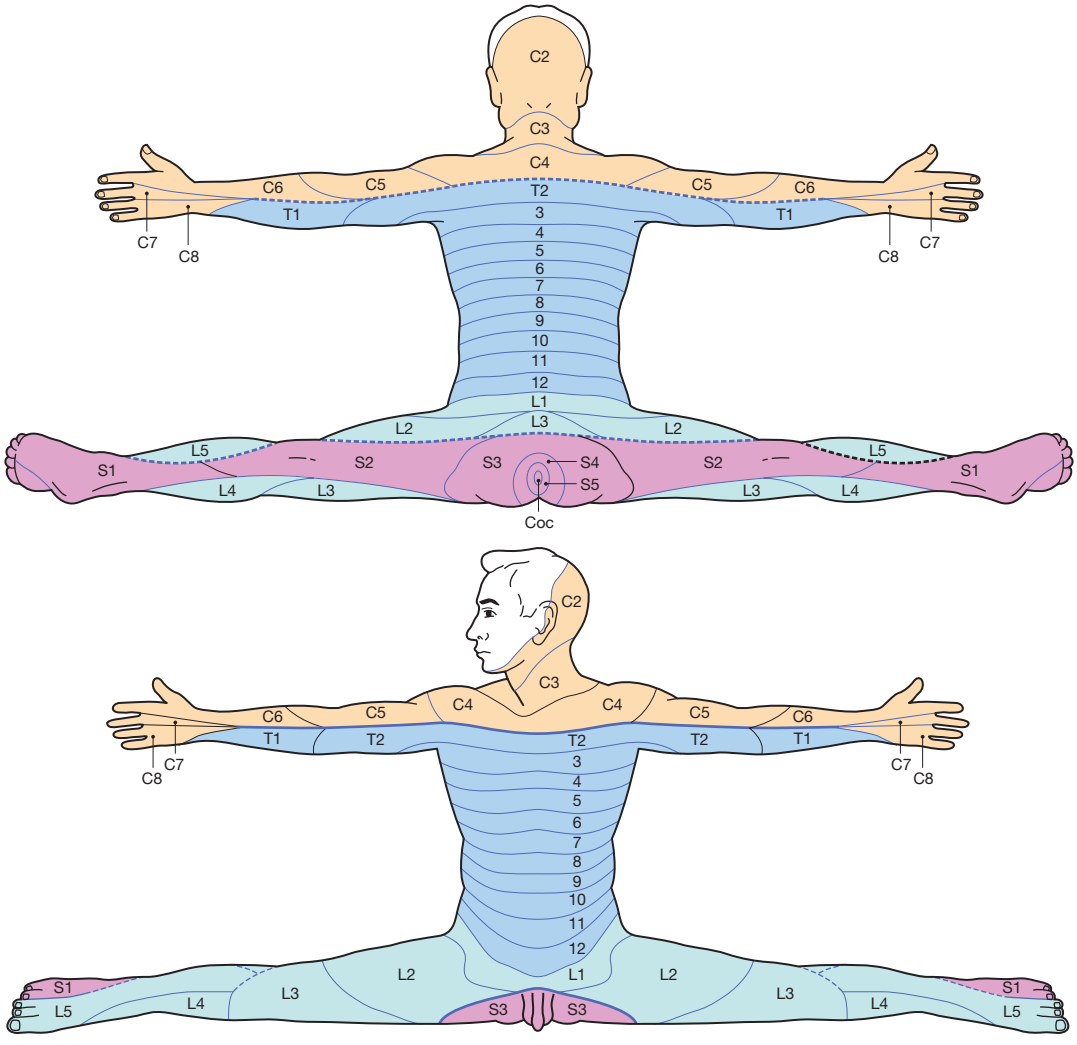

Fig. A.4.4

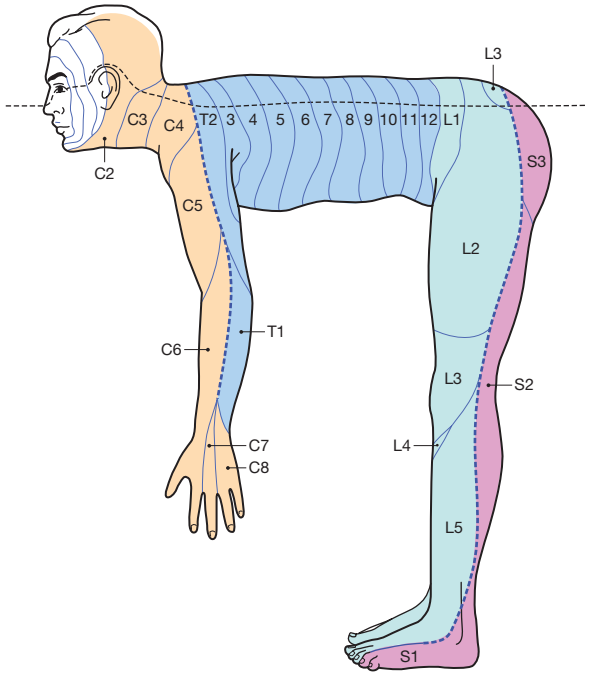

Fig. A.4.5

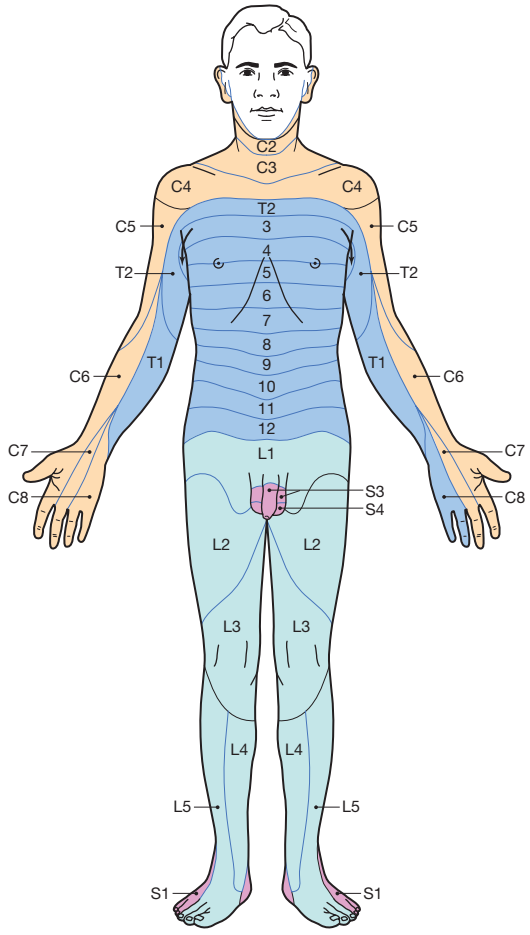

Fig. A.4.6

APPENDIX 5

SPINAL CORD SEGMENTAL MYOTOMES

Each muscle in the body is supplied by a particular level or segment of the spinal cord and by its corresponding spinal nerve.

- C5 also supplies the shoulder muscles and the muscle that we use to bend our elbow.
- C6 is for bending the wrist back.
- C7 is for straightening the elbow.
- C8 bends the fingers.
- T1 spreads the fingers.
- T1–T12 supplies the chest wall and abdominal muscles.
- L2 bends the hip.
- L3 straightens the knee.
- L4 pulls the foot up.
- L5 wiggles the toes.
- S1 pulls the foot down.
- S3, S4 and S5 supply the bladder, bowel and sex organs, and the anal and other pelvic muscles.

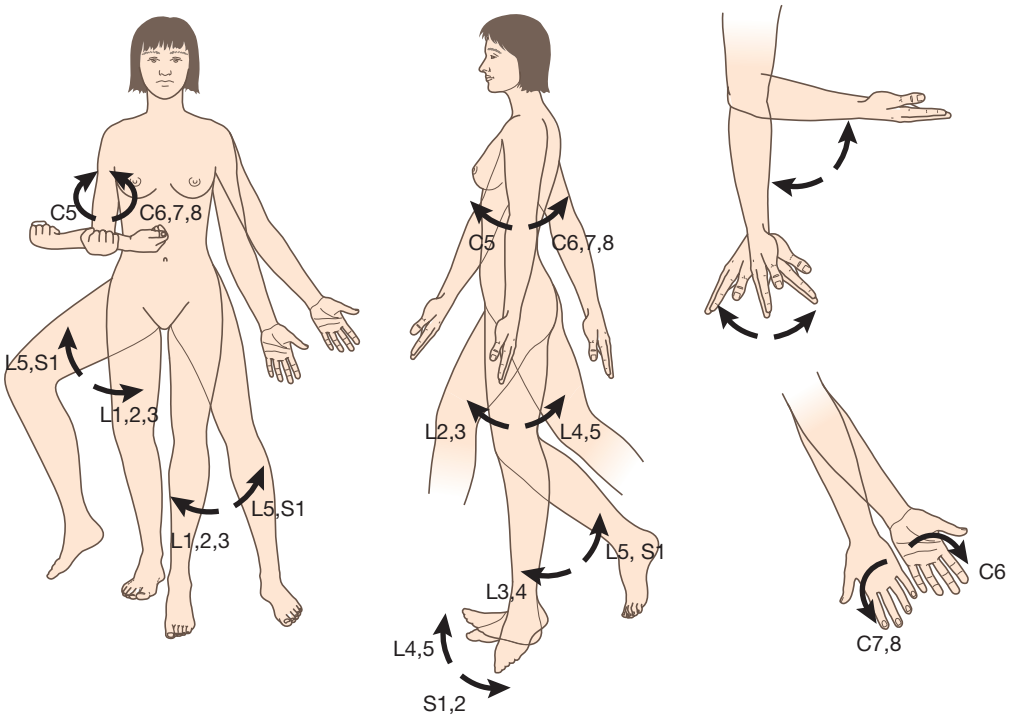

Fig. A.5.1

APPENDIX 6

LUMBO-SACRAL SPINE ANATOMY

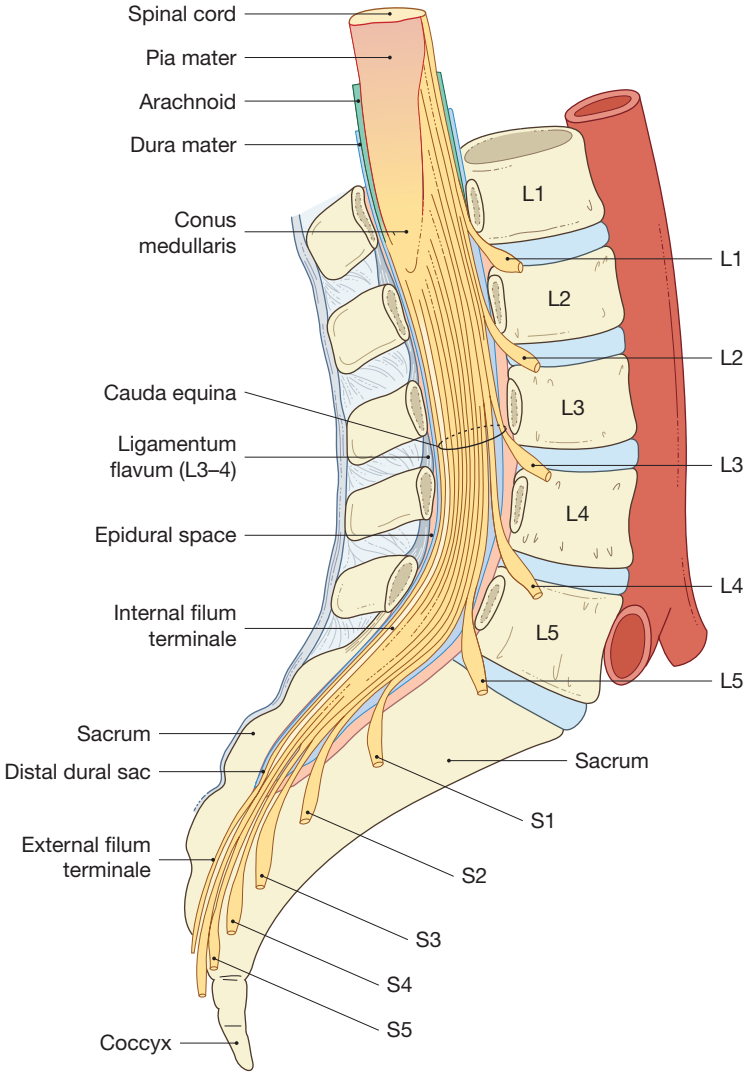

Fig. A.6.1

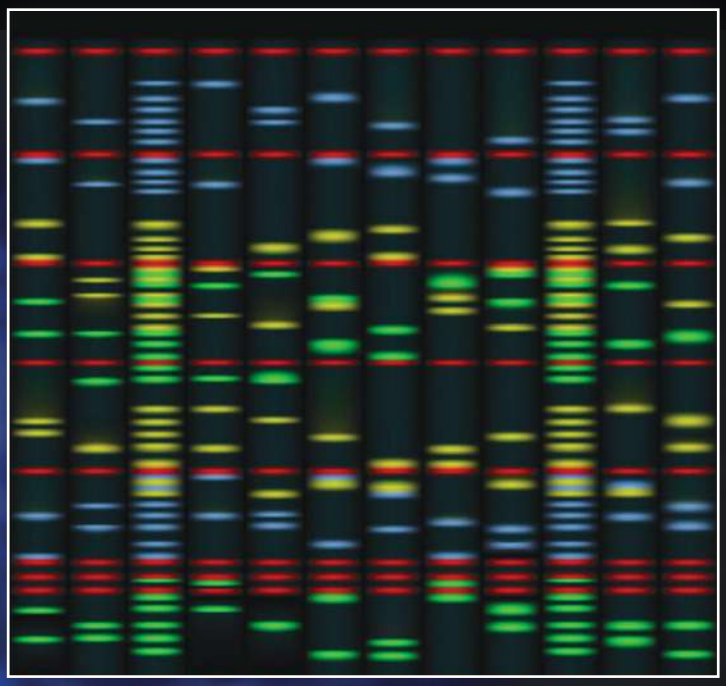

RICHARD LI

FORENSIC BIOLOGY

SECOND EDITION

CRC Press
Taylor & Francis Group

Contents

Preface xv
Acknowledgments xvii
Author xix
Introduction xxi

SECTION I Biological Evidence

Chapter 1 Crime Scene Investigation of Biological Evidence 3

1.1 Protection of Crime Scene 3
1.2 Recognition of Biological Evidence 3
1.3 Searches 7
1.4 Documentation 16
1.5 Chain of Custody 17
1.6 Collection of Biological Evidence 18
1.7 Marking Evidence 20
1.8 Packaging and Transportation 20
1.9 Final Survey and the Release of the Crime Scene 25
1.10 Crime Scene Reconstruction 29
Bibliography 30

Chapter 2 Crime Scene Bloodstain Pattern Analysis 35

2.1 Basic Biological Properties of Human Blood 35
2.2 Formation of Bloodstains 35
2.3 Chemical Enhancement and Documentation of Bloodstain Evidence 37
2.4 Analyzing Spatter Stains 39
2.4.1 Velocity of Blood Droplets 39
2.4.2 Determining the Directionality of the Stains 39
2.4.3 Determining Angles of Impact 40
2.4.4 Determining Area of Origin 41
2.5 Types of Bloodstain Patterns 44
2.5.1 Passive Bloodstains 44
2.5.2 Transfer Bloodstains 46
2.5.3 Projected Bloodstains 46
Bibliography 51

Chapter 3 Forensic Biology: A Subdiscipline of Forensic Science....53

3.1 Common Disciplines of Forensic Laboratory Services....53
3.2 Laboratory Analysis of Biological Evidence....53
3.2.1 Identification of Biological Evidence....53
3.2.2 Comparison of Individual Characteristics of Biological Evidence....54
3.2.3 Reporting Results and Expert Testimony....54
3.3 Forensic Science Services Related to Forensic Biology....57
3.3.1 Forensic Pathology....58
3.3.2 Forensic Anthropology....58
3.3.3 Forensic Entomology....60
3.3.4 Forensic Odontology....64
3.4 Brief History of the Development of Forensic Biology....65
3.4.1 Antigen Polymorphism....65
3.4.2 Protein Polymorphism....66
3.4.3 DNA Polymorphism....66
3.4.3.1 Genes and Related Sequences....66
3.4.3.2 Intergenic Noncoding Sequences....66
3.4.3.3 Human DNA Polymorphic Markers....68
3.4.3.4 Forensic DNA Polymorphism Profiling....69
Bibliography....70

Chapter 4 Sources of Biological Evidence....77

4.1 Bodily Fluids....77
4.1.1 Extracellular Nucleic Acids....77
4.2 Cells....81
4.2.1 Cell Surface Markers....81
4.2.2 Nucleated Cells....81
4.2.3 Mitochondria and Other Organelles....83
4.2.4 Cytosol....84
4.2.4.1 Messenger RNAs....84
4.2.4.2 MicroRNAs....85
4.3 Tissues....87
4.3.1 Skin....87
4.3.1.1 Biology of Skin....87
4.3.1.2 Skin as Source of DNA Evidence....87
4.3.2 Hair....88
4.3.2.1 Biology of Hair....89
4.3.2.2 Hair as Source of DNA Evidence....91
4.3.3 Bone....92
4.3.3.1 Biology of Bone....92
4.3.3.2 Bone as Source of DNA Evidence....94
4.3.4 Teeth....98
4.3.4.1 Biology of Teeth....98
4.3.4.2 Teeth as Source of DNA Evidence....100
Bibliography....102

SECTION II Basic Techniques in Forensic Biology

Chapter 5 Nucleic Acid Extraction 111

5.1 Basic Principles of DNA Extraction 111
5.1.1 Cell and Tissue Disruption 111
5.1.2 Lysis of Cellular and Organelle Membranes 114
5.1.3 Removal of Proteins and Cytoplasmic Constituents 114
5.1.4 Storage of DNA Solutions 115
5.1.5 Contamination 115
5.2 Methods of DNA Extraction 115
5.2.1 Extraction with Phenol–Chloroform 115
5.2.1.1 Cell Lysis and Protein Digestion 115
5.2.1.2 Extraction with Organic Solvents 115
5.2.1.3 Concentrating DNA 115
5.2.2 Extraction by Boiling Lysis and Chelation 117
5.2.2.1 Washing 117
5.2.2.2 Boiling 117
5.2.2.3 Centrifugation 117
5.2.3 Silica-Based Extraction 118
5.2.3.1 Cell Lysis and Protein Digestion 118
5.2.3.2 DNA Adsorption onto Silica 118
5.2.3.3 Washing 119
5.2.3.4 Elution of DNA 119
5.2.4 Differential Extraction 121
5.3 Essential Features of RNA 122
5.4 Methods of RNA Extraction 124
5.4.1 RNA–DNA Coextraction 124
5.4.2 miRNA Extraction 124
Bibliography 127

Chapter 6 DNA Quantitation 133

6.1 Slot Blot Assay 133
6.2 Fluorescent Intercalating Dye Assay 134
6.3 Quantitative PCR Assay 137
6.3.1 Real-Time Quantitative PCR 137
6.3.1.1 TaqMan Method 138
Bibliography 140

Chapter 7 Amplification by Polymerase Chain Reaction 143

7.1 Denaturation and Renaturation of DNA 143
7.2 Basic Principles of Polymerase Chain Reaction 143
7.3 Essential PCR Components 145
7.3.1 Thermostable DNA Polymerases 145
7.3.2 PCR Primers 146
7.3.3 Other Components 147

7.4 Cycle Parameters....147
7.5 Factors Affecting PCR....149
7.5.1 Template Quality....149
7.5.2 Inhibitors....149
7.5.3 Contamination....149
7.6 Reverse Transcriptase PCR for RNA-Based Assays....150
7.6.1 Reverse Transcription....150
7.6.2 Oligodeoxynucleotide Priming....151
7.6.3 Reverse Transcriptase PCR....151
Bibliography....153

Chapter 8 DNA Electrophoresis....159

8.1 Basic Principles....159
8.2 Supporting Matrices....159
8.2.1 Agarose....160
8.2.2 Polyacrylamide....161
8.2.2.1 Polymerization Reaction....161
8.2.2.2 Cross-Linking Reaction....161
8.2.2.3 Denaturing Polyacrylamide Electrophoresis....163
8.3 Apparatus and Forensic Applications....163
8.3.1 Slab Gel Electrophoresis....163
8.3.1.1 Agarose Gel Electrophoresis....163
8.3.1.2 Polyacrylamide Gel Electrophoresis....164
8.3.2 Capillary Electrophoresis....164
8.3.3 Microfluidic Devices....165
8.3.3.1 Modular Microfluidic Devices....167
8.3.3.2 Integrated Microfluidic Devices....168
8.4 Estimation of DNA Size....168
8.4.1 Relative Mobility....168
8.4.2 Local Southern Method....170
Bibliography....171

Chapter 9 Detection Methods....175

9.1 Direct Detection of DNA in Gels....175
9.1.1 Fluorescent Intercalating Dye Staining....175
9.1.2 Silver Staining....175
9.2 Detection of DNA Probes in Hybridization-Based Assays....177
9.2.1 Radioisotope Labeled Probes....177
9.2.2 Enzyme-Conjugated Probe with Chemiluminescence Reporting System....180
9.2.3 Biotinylation of DNA with Colorimetric Reporting Systems....180
9.2.3.1 Biotin....180
9.2.3.2 Enzyme-Conjugated Avidin....180
9.2.3.3 Reporter Enzyme Assay....181
9.3 Detection Methods for PCR-Based Assays....182
9.3.1 Fluorescence Labeling....182

9.3.1.1 Fluorescent Dyes 182
9.3.1.2 Labeling Methods....182
9.3.1.3 Fluorophore Detection 185
Bibliography 186

Chapter 10 Serology Concepts 189

10.1 Serological Reagents 189
10.1.1 Immunogens and Antigens 189
10.1.2 Antibodies....189
10.1.2.1 Polyclonal Antibodies....191
10.1.2.2 Monoclonal Antibodies....191
10.1.2.3 Antiglobulins 192
10.2 Strength of Antigen–Antibody Binding 193
10.3 Antigen–Antibody Binding Reactions....193
10.3.1 Primary Reactions 194
10.3.2 Secondary Reactions 194
10.3.2.1 Precipitation....195
10.3.2.2 Agglutination 197
Bibliography 198

Chapter 11 Serology Techniques: Past, Current, and Future.... 199

11.1 Introduction to Forensic Serology.... 199
11.1.1 The Scope of Forensic Serology.... 199
11.1.2 Class Characteristics and Individual Characteristics of Biological Evidence 199
11.1.3 Presumptive and Confirmatory Assays.... 200
11.1.4 Primary and Secondary Binding Assays.... 200
11.2 Primary Binding Assays 201
11.2.1 Enzyme-Linked Immunosorbent Assay (ELISA) 201
11.2.2 Immunochromatographic Assays 202
11.3 Secondary Binding Assays.... 204
11.3.1 Precipitation-Based Assays.... 204
11.3.1.1 Immunodiffusion 204
11.3.1.2 Immunoelectrophoretic Methods.... 206
11.3.2 Agglutination-Based Assays.... 208
11.4 DNA Methylation Assays for Bodily Fluid Identification....210
11.5 Forensic Applications of RNA-Based Assays and RNA Profiling.... 212
11.5.1 Messenger RNA-Based Assays....212
11.5.2 MicroRNA-Based Assays....214
11.6 Proteomic Approaches Using Mass Spectrometry for Bodily Fluid Identification 215
11.6.1 Mass Spectrometric Instrumentation for Protein Analysis 215
11.6.2 Analysis Strategies for Protein Identification 216
11.7 Microbial DNA Analysis for Bodily Fluid Identification....217
11.8 Nondestructive Assays for the Identification of Bodily Fluids.... 220
Bibliography 221

SECTION III Identification of Biological Evidence

Chapter 12 Identification of Blood 231

12.1 Biological Properties 231
12.1.1 Red Blood Cells 231
12.1.2 White Blood Cells 231
12.1.3 Platelets 231
12.2 Presumptive Assays for Identification 232
12.2.1 Mechanisms of Presumptive Assays 232
12.2.1.1 Oxidation–Reduction Reactions 234
12.2.2 Colorimetric Assays 234
12.2.2.1 Phenolphthalin Assay 234
12.2.2.2 Leucomalachite Green (LMG) Assay 234
12.2.2.3 Benzidine and Derivatives 234
12.2.3 Chemiluminescence and Fluorescence Assays 236
12.2.3.1 Luminol (3-Aminophthalhydrazide) 237
12.2.3.2 Fluorescin 238
12.2.4 Factors Affecting Presumptive Assay Results 238
12.2.4.1 Oxidants 238
12.2.4.2 Plant Peroxidases 239
12.2.4.3 Reductants 239
12.3 Confirmatory Assays for Identification 239
12.3.1 Microcrystal Assays 239
12.3.1.1 Hemochromagen Crystal Assay 239
12.3.1.2 Hematin Crystal Assay 240
12.3.2 Other Assays 240
Bibliography 241

Chapter 13 Species Identification 245

13.1 General Considerations 245
13.1.1 Types of Antibodies 245
13.1.2 Titration of Antibodies 246
13.1.3 Antibody Specificity 246
13.1.4 Optimal Conditions for Antigen–Antibody Binding 247
13.2 Assays 247
13.2.1 Immunochromatographic Assays 247
13.2.1.1 Identification of Human Hemoglobin Protein 247
13.2.1.2 Identification of Human Glycophorin A Protein 248
13.2.2 Double Immunodiffusion Assays 251
13.2.2.1 Ring Assay 251
13.2.2.2 Ouchterlony Assay 251
13.2.3 Crossed-Over Electrophoresis 251
Bibliography 255

Chapter 14 Identification of Semen 257

14.1 Biological Characteristics 257
14.1.1 Spermatozoa 257
14.1.2 Acid Phosphatase 259
14.1.3 Prostate-Specific Antigen 259
14.1.4 Seminal Vesicle–Specific Antigen 259
14.2 Analytical Techniques for Identifying Semen 260
14.2.1 Presumptive Assays 260
14.2.1.1 Lighting Techniques for Visual Examination of Semen Stains 260
14.2.1.2 Acid Phosphatase Techniques 261
14.2.2 Confirmatory Assays 264
14.2.2.1 Microscopic Examination of Spermatozoa 264
14.2.2.2 Identification of Prostate-Specific Antigen 265
14.2.2.3 Identification of Seminal Vesicle–Specific Antigen 269
14.2.2.4 RNA-Based Assays 270
Bibliography 271

Chapter 15 Identification of Saliva 277

15.1 Biological Characteristics of Saliva 277
15.1.1 Amylases 277
15.2 Analytical Techniques for Identification of Saliva 279
15.2.1 Presumptive Assays 279
15.2.1.1 Visual Examination 279
15.2.1.2 Determination of Amylase Activity 279
15.2.2 Confirmatory Assays 281
15.2.2.1 Identification of Human Salivary α-Amylase 281
15.2.2.2 RNA-Based Assays 285
Bibliography 286

Chapter 16 Identification of Vaginal Secretions and Menstrual Blood 289

16.1 Identification of Vaginal Stratified Squamous Epithelial Cells 289
16.1.1 Lugol's Iodine Staining and Periodic Acid–Schiff Method 290
16.1.2 Dane's Staining Method 293
16.2 Identification of Vaginal Acid Phosphatase 293
16.3 Identification of Vaginal Bacteria 294
16.4 Outlook for Confirmatory Assays of Vaginal Secretions 295
16.5 Menstruation 295
16.5.1 Uterine Cycle 296
16.5.2 Uterine Endometrial Hemostasis 298
16.6 D-dimer Assay 299
16.7 Lactate Dehydrogenase Assay 301
16.8 RNA-Based Assays 302
Bibliography 303

Chapter 17 Identification of Urine, Sweat, Fecal Matter, and Vomitus 307

17.1 Identification of Urine 307
17.1.1 Urine Formation 307
17.1.2 Presumptive Assays 308
17.1.2.1 The Identification of Urea 308
17.1.2.2 Identification of Creatinine 314
17.1.3 Confirmative Assays 315
17.1.3.1 Identification of Tamm–Horsfall Protein 315
17.1.3.2 Identification of 17-Ketosteroids 315
17.2 Identification of Sweat 316
17.2.1 Biology of Perspiration 316
17.2.2 Sweat Identification Assays 317
17.3 Identification of Fecal Matter 318
17.3.1 Fecal Formation 319
17.3.2 Fecal Matter Identification Assays 319
17.3.2.1 Macroscopic and Microscopic Examination 320
17.3.2.2 Urobilinoids Tests 322
17.3.2.3 Fecal Bacterial Identification 324
17.4 Identification of Vomitus 324
17.4.1 Biology of Gastric Fluid 324
17.4.2 Vomitus Identification Assays 325
Bibliography 326

SECTION IV Individualization of Biological Evidence

Chapter 18 Blood Group Typing and Protein Profiling 331

18.1 Blood Group Typing 331
18.1.1 Blood Groups 331
18.1.2 ABO Blood Group System 331
18.1.2.1 Biosynthesis of Antigens 333
18.1.2.2 Molecular Basis of the ABO System 333
18.1.2.3 Secretors 335
18.1.2.4 Inheritance of A and B Antigens 336
18.1.3 Forensic Applications of Blood Group Typing 336
18.1.4 Blood Group Typing Techniques 336
18.1.4.1 Lattes Crust Assay 336
18.1.4.2 Absorption–Elution Assay 338
18.2 Forensic Protein Profiling 338
18.2.1 Methods 339
18.2.1.1 Matrices Supporting Protein Electrophoresis 340
18.2.1.2 Separation by Molecular Weight 340
18.2.1.3 Separation by Isoelectric Point 341
18.2.2 Erythrocyte Protein Polymorphisms 342
18.2.2.1 Erythrocyte Isoenzymes 342
18.2.2.2 Hemoglobin 342

18.2.3 Serum Protein Polymorphisms 344
Bibliography 345

Chapter 19 Variable Number Tandem Repeat Profiling 353

19.1 Restriction Fragment Length Polymorphism 353
19.1.1 Restriction Endonuclease Digestion 354
19.1.2 Southern Transfer 356
19.1.3 Hybridization with Probes 357
19.1.3.1 Multilocus Probe Technique 357
19.1.3.2 Single-Locus Probe Technique 358
19.1.4 Detection 359
19.1.5 Factors Affecting RFLP Results 360
19.1.5.1 DNA Degradation 360
19.1.5.2 Restriction Digestion–Related Artifacts 361
19.1.5.3 Electrophoresis and Blotting Artifacts 362
19.2 Amplified Fragment Length Polymorphism 363
Bibliography 364

Chapter 20 Autosomal Short Tandem Repeat Profiling 369

20.1 Characteristics of STR Loci 369
20.1.1 Core Repeat and Flanking Regions 369
20.1.2 Repeat Unit Length 370
20.1.3 Repeat Unit Sequences 370
20.2 STR Loci Commonly Used for Forensic DNA Profiling 371
20.3 Forensic STR Analysis 375
20.3.1 Determining the Genotypes of STR Fragments 376
20.3.2 Interpretation of STR Profiling Results 376
20.3.2.1 Inclusion (Match) 376
20.3.2.2 Exclusion 376
20.3.2.3 Inconclusive Result 376
20.4 Factors Affecting Genotyping Results 376
20.4.1 Mutations 377
20.4.1.1 Mutations at STR Core Repeat Regions 377
20.4.1.2 Chromosomal and Gene Duplications 380
20.4.1.3 Point Mutations 381
20.4.2 Amplification Artifacts 382
20.4.2.1 Stuttering 382
20.4.2.2 Nontemplate Adenylation 383
20.4.2.3 Heterozygote Imbalance 383
20.4.2.4 Allelic Dropout 383
20.4.3 Electrophoretic Artifacts 384
20.4.3.1 Pull-Up Peaks 384
20.4.3.2 Spikes 384
20.5 Genotyping of Challenging Forensic Samples 384
20.5.1 Degraded DNA 384

20.5.2 Low Copy Number DNA Testing 386
20.5.3 Mixtures 386
Bibliography 389

Chapter 21 Sex Chromosome Haplotyping and Gender Identification 407

21.1 Y Chromosome Haplotyping 407
21.1.1 Human Y Chromosome Genome 407
21.1.1.1 Pseudoautosomal Region 407
21.1.1.2 Male-Specific Y Region 407
21.1.1.3 Polymorphic Sequences 408
21.1.2 Y-STR 409
21.1.2.1 Core Y-STR Loci 410
21.1.2.2 Multilocal Y-STR Loci 411
21.1.2.3 Rapidly Mutating Y-STR 411
21.2 X Chromosome Haplotyping 412
21.3 Sex Typing for Gender Identification 416
21.3.1 Amelogenin Locus 416
21.3.2 Other Loci 418
Bibliography 419

Chapter 22 Single Nucleotide Polymorphism Profiling 437

22.1 Basic Characteristics of SNPs 437
22.2 Forensic Applications of SNP Profiling 438
22.2.1 *HLA-DQA1* Locus 438
22.2.1.1 DQα AmpliType and Polymarker Assays 438
22.2.1.2 Allele-Specific Oligonucleotide Hybridization 440
22.2.2 Current and Potential Applications of SNP Analysis 443
22.2.2.1 Application of SNP Analysis for Forensic Identification 443
22.2.2.2 Potential Applications of SNPs for Phenotyping 444
22.3 SNP Techniques 444
22.3.1 Next-Generation Sequencing Technologies 444
22.3.2 DNA Samples, Sequencing Library, and Template Preparation 446
22.3.3 NGS Chemistry 447
22.3.4 NGS Coverage 447
Bibliography 451

Chapter 23 Mitochondrial DNA Profiling 461

23.1 Human Mitochondrial Genome 461
23.1.1 Genetic Contents of Mitochondrial Organelle Genomes 461
23.1.2 Maternal Inheritance of mtDNA 463
23.2 mtDNA Polymorphic Regions 463
23.2.1 Hypervariable Regions 463
23.2.2 Heteroplasmy 463

23.3 Forensic mtDNA Testing 465
23.3.1 General Considerations 465
23.3.2 mtDNA Screen Assay 466
23.3.3 mtDNA Sequencing 467
23.3.3.1 PCR Amplification 467
23.3.3.2 DNA Sequencing Reactions 468
23.3.3.3 Electrophoresis, Sequence Analysis, and Mitotype Designations 471
23.3.4 Interpretation of mtDNA Profiling Results 472
23.3.4.1 Exclusion 474
23.3.4.2 Cannot Exclude 474
23.3.4.3 Inconclusive Result 474
Bibliography 474

SECTION V Forensic Issues

Chapter 24 Forensic DNA Databases: Tools for Crime Investigations 485

24.1 Brief History of Forensic DNA Databases 485
24.2 Infrastructure of CODIS 485
24.3 Indexes of CODIS 487
24.4 Database Entries 489
24.5 Database Expansion 490
24.6 DNA Profiles 491
24.7 Routine Database Searches for Forensic Investigations 492
24.7.1 Case-to-Offender Searches 492
24.7.2 Case-to-Case Searches 493
24.7.3 Search Stringency and Partial Matches 493
24.8 Familial Searches 495
24.8.1 Legal and Ethical Issues of Familial Search 495
24.8.2 Familial Search Strategies 496
24.8.2.1 Identity-by-State and Kinship Index Method 496
24.8.2.2 Focusing on Rare Alleles 496
24.8.2.3 Excluding Candidates through Y-STR Screening 498
Bibliography 498

Chapter 25 Evaluation of the Strength of Forensic DNA Profiling Results 503

25.1 A Review of Basic Principles of Genetics 503
25.1.1 Mendelian Genetics 503
25.1.2 Population Genetics 504
25.1.2.1 Allele Frequency 504
25.1.2.2 Genotype Frequency 504
25.1.2.3 Heterozygosity 504
25.1.2.4 Hardy–Weinberg Principle 505
25.1.2.5 Testing for HW Proportions of Population Databases 505
25.1.2.6 Probability of Match 511

25.2 Statistical Analysis of DNA Profiling Results 512
25.2.1 Genotypes 513
25.2.1.1 Profile Probability 513
25.2.1.2 Likelihood Ratio 516
25.2.2 Haplotypes 516
25.2.2.1 Mitotypes Observed in Database 518
25.2.2.2 Mitotype Not Observed in Database 518
Bibliography 518

Chapter 26 Quality Assurance and Quality Control 523

26.1 US Quality Standards 523
26.2 International Quality Standards 524
26.3 Laboratory Accreditation 525
26.4 Laboratory Validation 525
26.5 Proficiency Testing 526
26.6 Certification 526
26.7 Forensic DNA Analyst Qualifications 527
26.8 Code of Ethics of Forensic Scientists 528
Bibliography 529

Index 533

1

Crime Scene Investigation of Biological Evidence

A forensic investigation involving biological evidence usually begins at the crime scene. The crime scene investigation process includes maintaining scene security, preparing documentation, and collecting and preserving physical evidence. A crime scene investigation requires teamwork and effort. Each team member should be assigned specific tasks (Figure 1.1).

1.1 Protection of Crime Scene

A crime scene investigation begins with the initial response to a scene (Figure 1.2). Securing and protecting the scene are important steps in a crime scene investigation (Figures 1.3 through 1.6), and this task is usually carried out by the first responding officer arriving on the scene. The entry of authorized personnel admitted to the scene should be documented using a log sheet (Figure 1.7). Suspects, witnesses, and living victims should be evacuated from the scene. If a victim is wounded, medical attention should be sought.

Appropriate supplies and devices should be used to prevent the contamination of evidence by investigators. Protective wear and devices including a face mask or shield, safety eyeglasses, a disposable coverall bodysuit, gloves, shoe covers, and a hairnet should be used (Figures 1.8 and 1.9). Exposure to bodily fluids may occur during a crime scene investigation. An investigator can be exposed to bodily fluids through the mucous membranes, skin exposure, and needlestick injuries (especially when investigating a clandestine drug laboratory scene). Therefore, biosafety procedures must be followed for the protection of personnel from infectious blood-borne pathogens such as the human immunodeficiency virus (HIV), hepatitis B virus (HBV), and hepatitis C virus (HCV); infectious aerosol tuberculosis pathogens; and other biohazardous materials.

1.2 Recognition of Biological Evidence

A preliminary survey should be carried out to evaluate potential evidence. In particular, the recognition of evidence plays a critical role in solving or prosecuting crimes. The priority of the potential evidence at crime scenes should be assessed based on each item's relevance to the solution of the case. Higher priority should be assigned to evidence with probative value to the case. For example, the evidence related to a *corpus delicti* is considered to be of the highest priority. *Corpus delicti* is a Latin term meaning "body of crime." In Western law, it primarily refers to the principle that in order for an individual to be convicted, it is necessary to prove the occurrence

Figure 1.1 A crime scene investigation team. (© Richard C. Li.)

Figure 1.2 A crime scene unit vehicle that is used to respond to crime scenes. This type of vehicle is usually outfitted with devices and supplies that investigators need when processing a crime scene, as well as evidence packaging materials, fingerprint collection kits, and DNA collection kits. Additionally, it can be equipped with a workstation for computer access, a refrigerator for storing chemicals, and a compact fuming hood for processing latent fingerprints, as well as equipment cabinets and drawers. (© Richard C. Li.)

Figure 1.3 Crime scene barrier tape is used to ensure that only investigators are admitted to the scene. (Courtesy of H. Brewster.)

Figure 1.4 A police officer guards the crime scene. (© Richard C. Li.)

Figure 1.5 Crime scene privacy screen and tent. The screen (top) and the tent (bottom) are useful devices for shielding the evidence or body from viewing by unauthorized personnel. (© Richard C. Li.)

Figure 1.6 Barricades are set up to keep crowds at a distance from the scene. (© Richard C. Li.)

of the crime. In a forensic investigation, it also refers to the physical evidence proving that a crime was committed. For example, when an individual is missing, a missing persons investigation is usually initiated. If *corpus delicti*, such as a dead body or a victim's blood at a crime scene (Figure 1.10), is discovered during the investigation, a homicide case can be established and a suspect can be charged with homicide. Higher priority should also be attached to evidence that can establish connections such as *victim-to-perpetrator linkage*. For example, items found in a perpetrator's possession may be linked to a victim. This also applies to transfer evidence based on the principles of transfer theory, also known as the *Locard exchange principle*, which theorizes that the cross-transfer of evidence occurs when a perpetrator has any physical contact with an object or another person (Figures 1.11 through 1.14). Thus, trace evidence, such as hairs and fibers, may be transferred from a perpetrator to a victim or vice versa. This explains why it is important to ensure that perpetrators and their belongings are thoroughly searched for trace evidence. Likewise, victims and their belongings should be examined for the same reason.

Victim-to-scene and perpetrator-to-scene linkages can also be established. Blood belonging to a perpetrator or a victim found at a crime scene can establish such a linkage (Figure 1.15). Additionally, reciprocal transfers of trace evidence from crime scenes can be used to link a suspect or a victim to a crime scene. A perpetrator may present a unique *modus operandi* (MO). *Modus operandi*, a Latin term commonly used in criminal investigations, refers to a particular pattern of characteristics and the manner in which a crime is committed. For example, Richard Cottingham, a serial killer known as "the torso killer," dismembered his victims and took their limbs and heads with him but left their torsos at the scene. He then set the rooms on fire before

Crime Scene Processing

1: Date _____/_____/_____ 2: Arrival Time: _____ HRS

3: Assigned Criminalist_____ 4: Assisting Criminalist(s)_____

5: Supervisor_____ Present ___Yes ___No

6: Case Number ___-_____

7: Individual/Agency Requesting Investigation _____

8: Location Address_____

a. Type of Scene ___Indoor ___Outdoor
b. Scene Secured ___Yes ___No
c. Weather Conditions_____

9: Other Personnel	Shield #	Agency	Arrival	Departure
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS
			HRS	HRS

10: Departure Time_____ HRS

Figure 1.7 An example of a log sheet for documenting authorized personnel at a crime scene. (© Richard C. Li.)

fleeing the scenes. Evidence that provides information on the MO is also vital to an investigation. A distinct MO can establish a *case-to-case linkage* for serial offender cases.

1.3 Searches

Some investigations require a search for specific items of evidence such as biological stains, human remains, and all relevant evidence. A search usually has a specific purpose. Thus, the use of search patterns can be helpful, especially in cases involving large outdoor crime scenes.

Figure 1.8 Personal protection wear and devices that are used at crime scenes. (© Richard C. Li.)

Figure 1.9 Disposable glove (left) and glove with extended cuff (right). (© Richard C. Li.)

Search patterns may include a grid, line, or zone (Figures 1.16 and 1.17). The method that is ultimately used depends on the type and size of the scene (Figures 1.18 through 1.20). Additionally, the points of entry and exit and the paths followed by a perpetrator should also be searched.

Searching for biological stains usually utilizes devices such as an alternate light source (ALS); see Figures 1.21 and 1.22. An ALS either produces a single specific wavelength of light or a desired wavelength by using specific filters. Biological materials such as blood, semen, and saliva emit fluorescent light under an ALS, which can facilitate the locating of biological materials. Additionally,

Figure 1.10 Photographic documentation of bloodstains on clothing. (© Richard C. Li.)

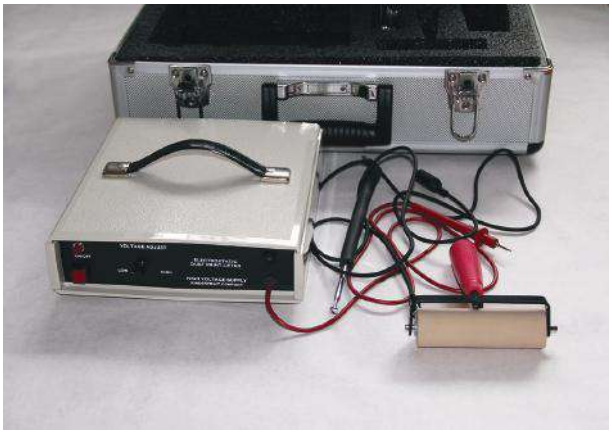

Figure 1.11 An electrostatic dust print lifting device can be utilized for processing impression evidence such as footprints and tire tracks. (© Richard C. Li.)

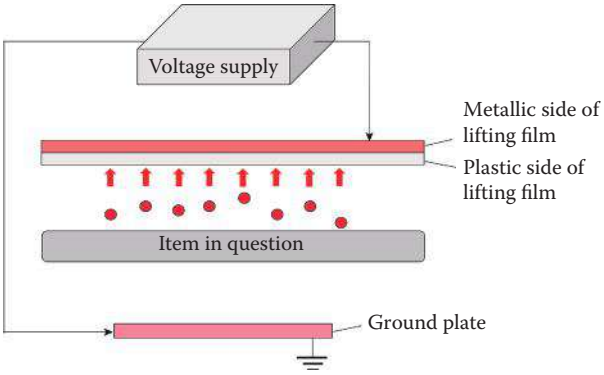

Figure 1.12 Basic components of an electrostatic dust print lifting device. The lifting film is placed on top of the item in question with the plastic side against the surface and the metallic side facing up. A ground plate is placed directly on the ground. The lifting film and ground plate are connected to the voltage supply apparatus. Once the charging voltage is turned on, the static charge transfers the dust particles from the surface to the plastic side of the lifting film. (© Richard C. Li.)

Figure 1.13 A high-intensity light-emitting diode (LED) device for locating evidence at a crime scene is particularly effective in highlighting trace evidence such as hairs, fibers, and shoe prints. (© Richard C. Li.)

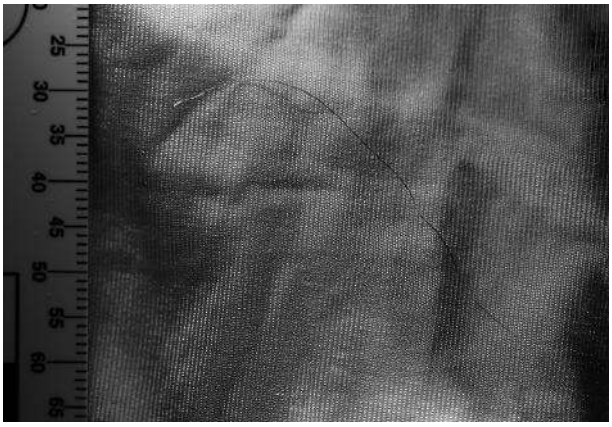

Figure 1.14 A hair found on a victim's clothing can be transferred evidence from a suspect. (© Richard C. Li.)

Figure 1.15 Finding a victim's blood on a suspect's clothing can establish a link between them. (© Richard C. Li.)

Figure 1.16 Grid search pattern for an outdoor scene. The investigators and anthropologists present are searching for human bone evidence within the grid. (Courtesy of H. Brewster.)

Figure 1.17 Line search pattern for an outdoor scene. (Courtesy of H. Brewster.)

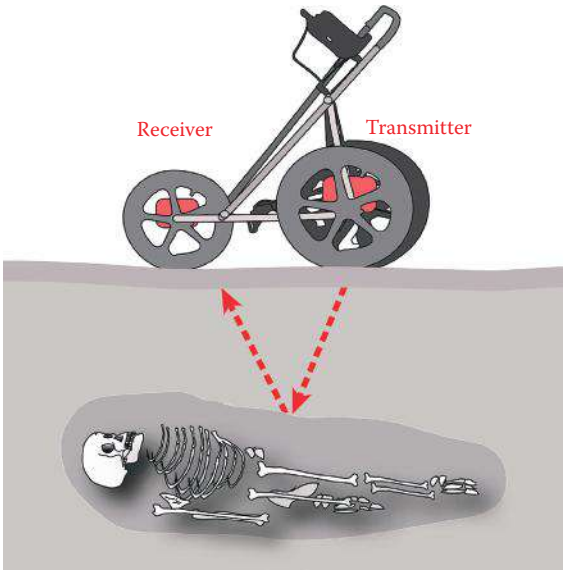

Figure 1.18 Using ground-penetrating radar (GPR) to locate clandestine graves of homicide victims. GPR uses electromagnetic waves emitted from a transmitter, which are detected by a receiver to locate clandestine burials and buried objects such as weapons embedded in soils. Images of the potential evidence are typically obtained by moving the antenna of the GPR device over the surface of the ground. (© Richard C. Li.)

Figure 1.19 Using cadaver-sniffing dogs to alert investigators to the presence of buried bodies. The odor produced by the decomposition of the human body may be sensed by cadaver-sniffing dogs. Odor-absorbing pads absorb and retain the scent of decomposed remains and can be placed at the scene for several days. Upon sniffing the pads, a cadaver dog may indicate the presence of buried human remains or may indicate that human remains were once buried in that location. However, it is not clear that this technique is reliable. (© Richard C. Li.)

Figure 1.20 Tracking dog for searching suspects. In a situation of processing a recent crime scene when the suspect may be in close vicinity to the scene, a tracking dog can be used. A tracking dog, such as a bloodhound, can potentially follow the scent from items left at the scene to locate a suspect nearby. (© Richard C. Li.)

Figure 1.21 Compact alternate light source devices, which are intended specifically for use at crime scenes, reduce search time and improve the recovery of evidence such as biological evidence and chemically enhanced latent fingerprints. (© Richard C. Li.)

Figure 1.22 LED light sources provide illumination for locating evidence such as bodily fluid stains, hair, or fibers. (© Richard C. Li.)

Figure 1.23 An example of a compact Rapid DNA device for processing DNA evidence in the field. (© Richard C. Li.)

Figure 1.24 An example of sketch documentation. (© Richard C. Li.)

field tests and enhancement reagents can be used to facilitate crime scene searching (Chapter 12). These reagents can detect and identify biological evidence. The tests are very simple, rapid, and sensitive, and thus can be used at crime scenes. For example, phenolphthalin and leucomalachite green tests can be used for detecting blood evidence. Sometimes, minute amounts of blood may be

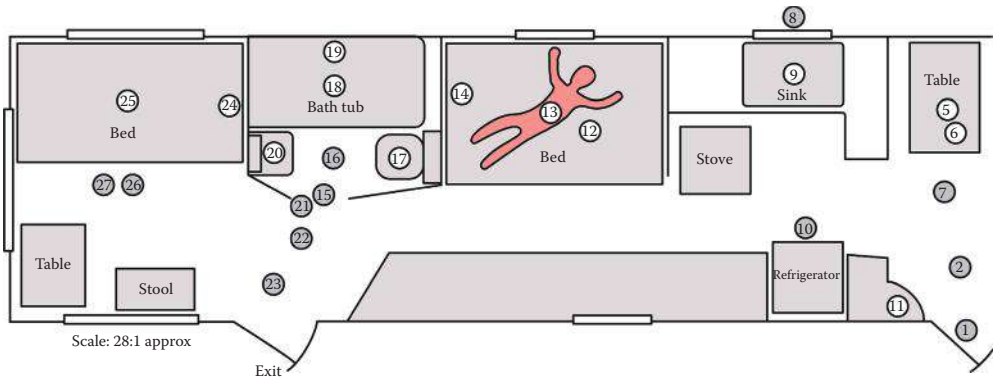

Figure 1.25 Sketches for documenting the sites of fingerprints. (© Richard C. Li.)

present at the scene as a result of attempts to clean up blood prior to the investigation. These stains may not be visible with the naked eye. Enhancement reagents such as luminol and fluorescein, which emit chemiluminant and fluorescent light upon reacting with certain biological materials, respectively, can be used. Additionally, the enhancement reagents can detect faint blood-containing pattern evidence such as faint bloody fingerprints, footprints, and other pattern evidence of physical contact such as drag marks in blood. However, precaution should be taken since these reagents are not usually very specific to blood. Certain substances such as bleach, various metals, and plants may also lead to chemical reactions with the field tests and the enhancement reagents. In these cases, the evidence collected is further tested with laboratory examination and analysis.

Recently, portable and field-deployable instruments have been developed that are capable of processing buccal swabs and potentially other evidence to produce a DNA profile on-site (Figure 1.23). It is a fully automated process, using the Rapid DNA technology (Chapter 8), that

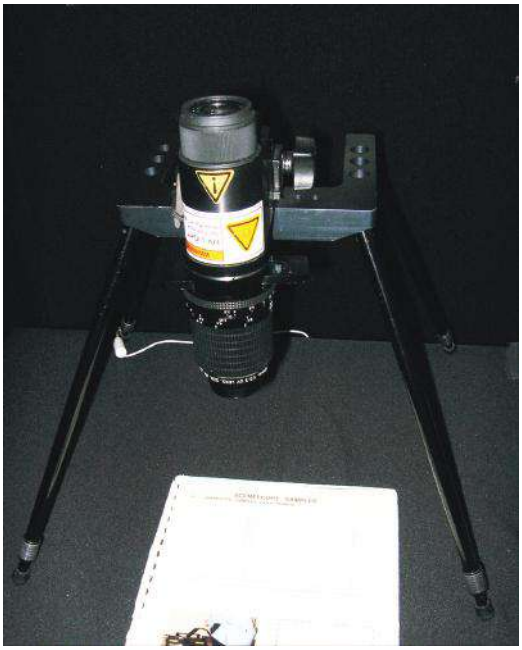

Figure 1.26 A reflected ultraviolet imaging system (RUVIS) imager for documenting close-up views of evidence such as latent fingerprints. (© Richard C. Li.)

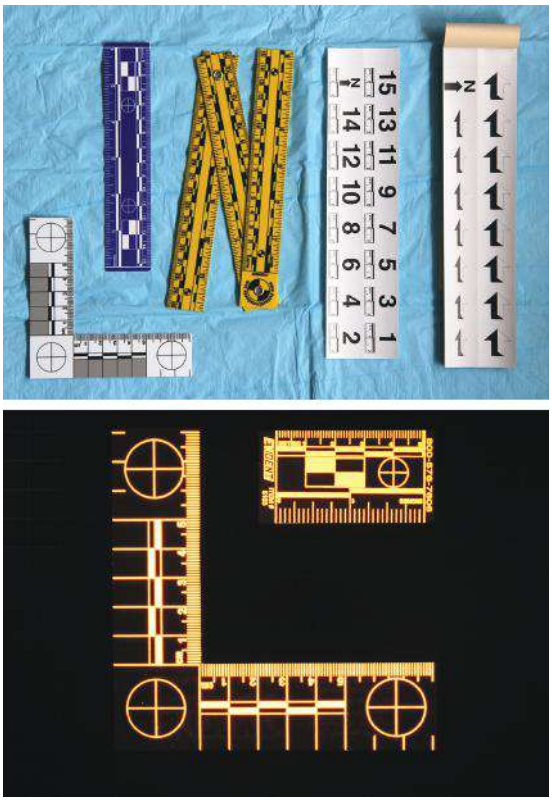

Figure 1.27 Scales for photographic documentation. Regular scales (top) can be used with visible light sources. Fluorescence scales (bottom) can be used for the documentation of fluorescing evidence after certain treatment, such as chemical enhancement, using alternate light sources. (© Richard C. Li.)

can be completed within 2 h by a trained crime scene investigator or police officer. These instruments may provide a new tool for expediting the identification of suspects and developing investigative leads at the scene. Additionally, this technology can enable law enforcement agents to rapidly determine whether the crimes were isolated incidents or part of serial crimes committed by the same offender, such as in serial burglary and arson cases. It can also be used in the identification of human remains in mass disasters.

1.4 Documentation

The conditions at a crime scene, including both the individual items of evidence and the overall scene, must be documented to provide vital information for investigators and for the courts. The most common documentation methods are drawing sketches and taking photographs and videographs. The sketch is to reflect the positions and the spatial relationships of items and persons with measurements using a scale. An investigator usually prepares a rough sketch first and later converts the rough sketch into a finished sketch (Figures 1.24 and 1.25). If bloodstains are present at the scene, the location of bloodstain patterns should be emphasized. Prior to handling and moving evidence, photographs should be taken with different views: an overall view of the entire scene, a medium-range view showing the positions and the relationships of items, and a close-up

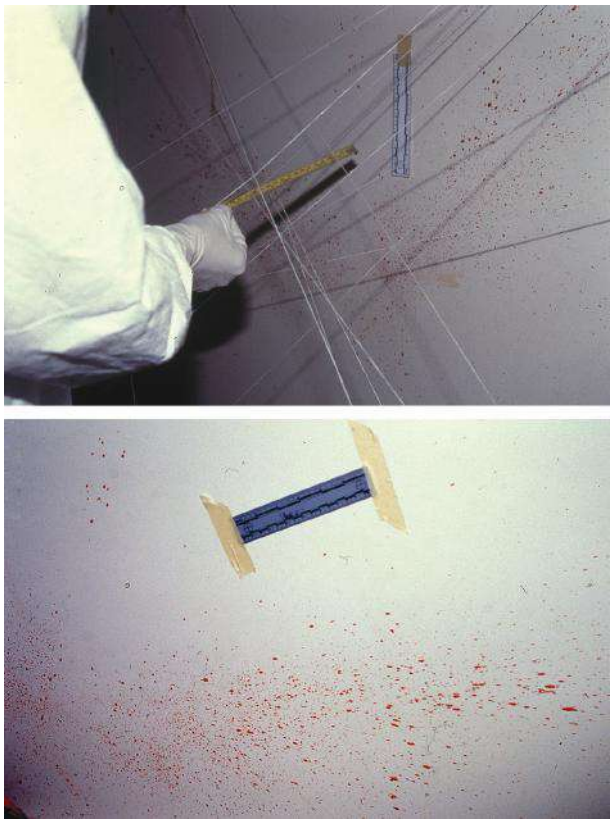

Figure 1.28 Photographic documentation of bloodstain patterns. (© Richard C. Li.)

view showing details of the evidence (Figure 1.26). Photographs should also include a measuring device such as a scale (Figure 1.27) to accurately depict the sizes of items such as bloodstains or bite marks. This can be achieved simply by placing a ruler adjacent to the evidence when it is photographed (see Figure 1.28). A photograph log sheet can be used to record the chronological order of crime scene photographs and to note filming conditions and any additional relevant information (Figure 1.29). Similar documentation should be prepared for videographs when appropriate. Additionally, written or audio-recorded notes can be used. Notes should include complete and accurate information of a crime scene investigation, such as the case identifier number, the identities of the investigators, and a description of the scene or items (e.g., location, size, and shape). Additionally, any disturbance of evidence occurring during crime scene processing should be noted.

1.5 Chain of Custody

Custody information should be recorded at each event when evidence is handled or transferred by authorized personnel. Usually, a custody form listing a specific evidence item is used to document the chain (Figure 1.30). Each individual who acquires custody of the evidence must sign a chain of custody document. An incomplete chain of custody may lead to an inference of possible tampering or contamination of evidence. As a result, the evidence may not be admissible in court.

Photography Log

Date: ____/____/____ Case Number: ____-_____ Criminalist _____
Roll # _____ Film Type & ASA: _____

EXP	Lens	f-stop	Category	Description
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				

Figure 1.29 Crime scene photography log. (© Richard C. Li.)

1.6 Collection of Biological Evidence

After the crime scene documentation is completed, the collection of evidence can be initiated. Small or portable items, such as bloodstained knives, can be collected and submitted to a crime laboratory (Figure 1.31). Large or unmovable items of evidence (Figure 1.32) can be collected and submitted in sections, such as a section of wall where bloodstains are located. Table 1.1 and Figures 1.33 through 1.38 summarize and illustrate representative collection techniques. Specific care is required for the collection of biological evidence in the following situations:

- *Bloodstain pattern evidence*: It is especially important to thoroughly document the bloodstain pattern evidence at a crime scene prior to collection. Bloodstain patterns can be especially useful in crime scene reconstruction.

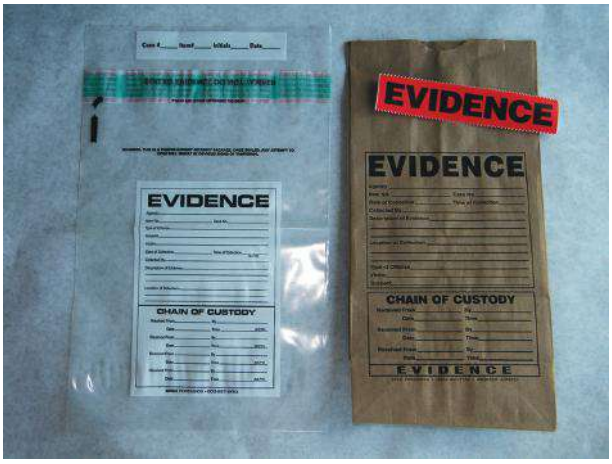

Figure 1.30 Labels with the chain of custody that are used for marking the evidence contained in the packaging. (© Richard C. Li.)

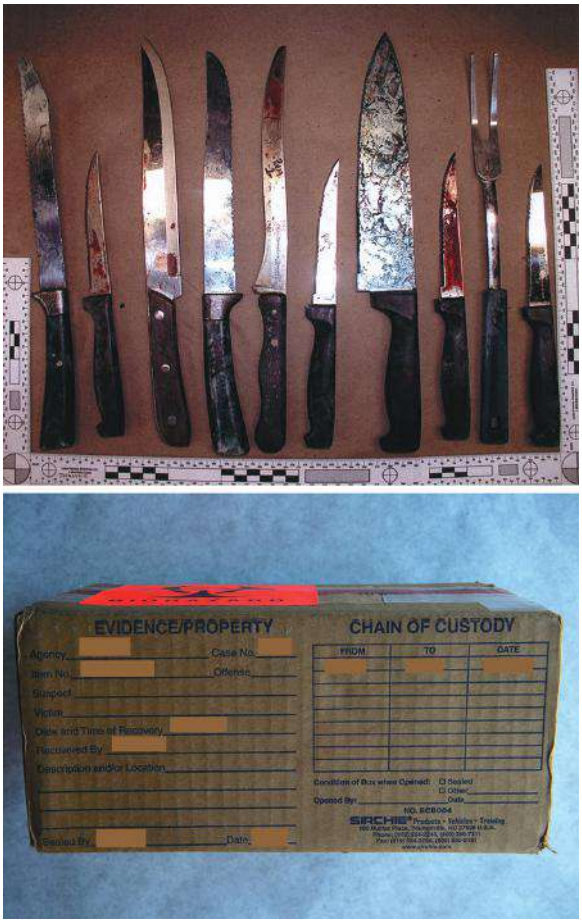

Figure 1.31 Handling sharp objects. Bloodstained knives collected and submitted to laboratories (top) and a box for packing sharp objects (bottom). (© Richard C. Li.)

Figure 1.32 A section of bloodstained carpet is collected. (© Richard C. Li.)

- *Multiple analysis of evidence*: If multiple analyses are needed for a single item of evidence, nondestructive analyses should be carried out first. For example, a bloody fingerprint should be collected for ridge detail analysis prior to collecting blood for DNA analysis.
- *Trace evidence*: Trace evidence such as hairs and fibers can be present in bloodstained evidence and should be identified and properly collected.
- *Control samples*: Control (known or blank) samples should be collected from a control area (e.g., unstained area near a collected stain).
- *Size of stain*: Polymerase chain reaction (PCR)-based forensic DNA techniques are highly sensitive and allow for the successful analysis of very small bloodstains. All bloodstains, even if they are barely visible, should be collected at a crime scene.
- *Wet evidence*: Wet evidence should be air-dried (without heat) prior to packaging to prevent the degradation of proteins and nucleic acids, which are used for forensic serological and DNA analysis.

1.7 Marking Evidence

The marking of evidence is necessary for identification purposes so that it can be quickly recognized even years later (Figure 1.39). An investigator's initials, the item number, and the case number are usually included in marking. Information can be marked on a tag, a label attached to the item, or directly on garment evidence. The marking of evidence should not be proximal to bullet holes or biological stains to prevent the mark from interfering with analyses.

1.8 Packaging and Transportation

Packaging is intended to protect and preserve evidence. All evidence should be secured and protected from possible contamination. Fragile items should be protected to prevent any damage during transportation. Exposure to heat and humidity should be avoided to protect biological evidence from degradation during transport. Various packaging methods are available

Table 1.1 Methods for Collecting Biological Evidence

Type of Evidence	Condition	Method of Collection	Procedure
Blood	Dry	Swab	Best on nonabsorbent surfaces; lightly moisten sterile swab with distilled or sterile water, rub over stain while rotating; allow to air-dry; a combination of first a moistened swabbing followed by second a dry swabbing (both swabs submitted) is recommended
		Cutting	Cut stain from item
		Scraping	Scrape bloodstain into a clean piece of paper using a clean blade; wrap sample using druggist's fold
		Lifting	Works for nonabsorbent surfaces; use fingerprint lifting tape that does not interfere with DNA testing to lift stain; lifted stain should be covered with a piece of lifter's cover
		Collect entire item	Collect if item contains bloodstain pattern; difficult to swab; requires multiple exams
	Wet	Swab	Absorb blood sample onto sterile cotton swabs; stain should be concentrated on tip and allowed to air-dry
		FTA paper	Use a sterile disposable pipet to collect liquid blood; spot on FTA paper; allow to air-dry
	Reference liquid blood sample	Venous blood collection	Collect blood in a purple-topped vacutainer tube containing ethylenediaminetetraacetic acid (EDTA) anticoagulant; refrigerate but never freeze
Semen	Dry or wet	See "Blood" above	See "Blood" above
	Condom	Collect entire item	Secure condom with a tie and place in a container in a refrigerator; submit to laboratory as soon as possible
	Various conditions	Victim sexual assault kit	Standardized kit to collect biological evidence from the body of a victim includes swabs, microscope slides, and envelopes
	Various conditions	Suspect standard kit	Standardized kit to collect biological evidence from the body of a suspect includes swabs, microscope slides, and envelopes

(*continued*)

Table 1.1 (Continued) Methods for Collecting Biological Evidence

Type of Evidence	Condition	Method of Collection	Procedure
Victim vaginal fluid	Dry or wet	See "Blood" above	Often collected from a suspect's pubic area or fingers
Saliva	Dry or wet	See "Blood" above	Often collected from a bite mark or a victim's pubic area
	Reference saliva (buccal) samples	Swab	Swab the inside of the cheek using two swabs, rotating them during collection; allow swabs to air-dry
		Filter paper	Place donor saliva sample on marked area of filter paper; allow to air-dry
Hairs	Head and pubic hairs	Lifting	Refer to dry bloodstain lifting method
		Transfer	Use forceps to transfer hair onto a piece of paper that can be folded
		Vacuum	Vacuum can be used if necessary; generally not recommended
Fingernails and scrapings	Various conditions	Clipping	Use a clean clipper to clip nails onto clean paper; wrap samples using druggist's fold
		Scraping	Scrape undersides of nails onto clean paper; wrap samples using druggist's fold
Bones	Various conditions	Freeze in container	Collect if blood is not available; collect bone with marrow if available; rib bone and vertebrae are preferred; place specimen in a container and freeze if wet
Teeth	Various conditions	Container	Collect teeth with dental pulp if possible into a container
Tissues and organs	Wet	Freeze in a container	Collect if blood is not available; place specimen in a container and freeze
DNA database samples	Various conditions	Various	Follow jurisdiction protocol for collecting samples from arrested individuals or convicted offenders or both

Source: Fisher, B., *Techniques of Crime Scene Investigation*, 7th edn., 2004, CRC Press, Boca Raton; Lee, H.C. et al., *Henry Lee's Crime Scene Handbook*, 2001, Academic Press, San Diego; National Institute of Justice. Using DNA to solve cold cases, Special Report, 2002, US Department of Justice, Office of Justice Programs, Washington, DC.

Note: Samples requiring refrigeration or freezing are noted; other samples may be stored at room temperature. Dry evidence should be packed in a porous material such as paper (envelope, bag, box) as described in the text.

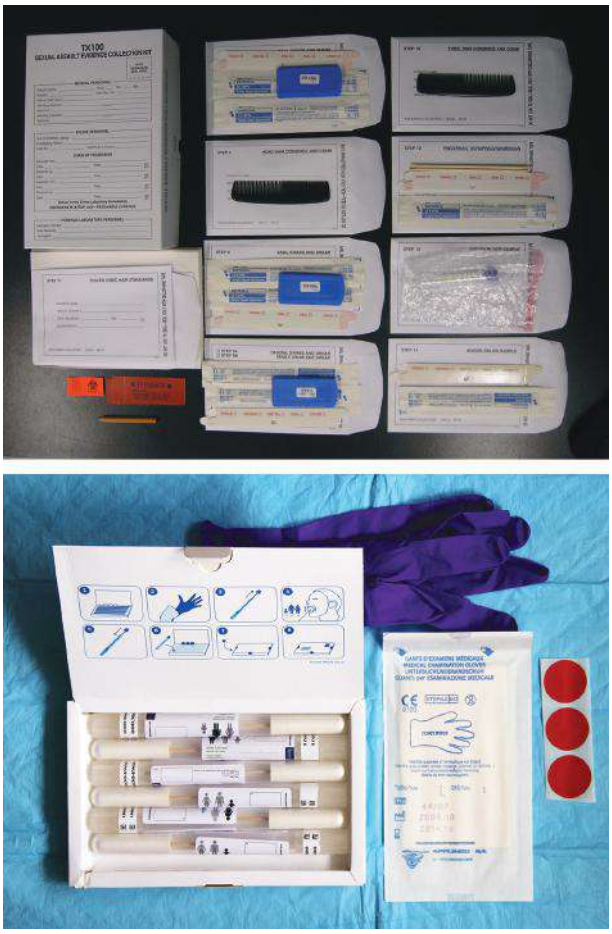

Figure 1.33 Evidence collection kits. Sexual assault evidence collection kit (top). Paternity evidence collection kit (bottom). (© Richard C. Li.)

Figure 1.34 Various types of swabs that are used for collecting biological evidence. (© Richard C. Li.)

Figure 1.35 Fingernail swabbing for recovering evidence. Fingernail swabs are often collected from individuals who are involved in a struggle in violent crimes and in digital penetration in sexual assault cases. Fingernail swabs are sampled separately from both hands. (© Richard C. Li.)

depending on the type of evidence handled (Figures 1.40 through 1.42). The following are general considerations related to the packaging of evidence:

- *Evidence from different sources*: To prevent the transfer of evidence from different sources, items of evidence should not be grouped in a single package. However, evidence may be packed in a single container if the items were found together.
- *Folding of evidence*: Folding of clothing, especially items with wet bloodstains, can transfer evidence from one part of a garment to another. If a large, dry garment must be folded, a piece of clean paper should be placed between different parts of the garment to avoid direct contact between the different parts of the garment, thereby preventing the transfer of evidence.
- *Packing materials*: Envelopes, bags, and boxes that are made of porous materials such as paper are appropriate for packaging dry biological evidence. Dry, bloodstained evidence should not be sealed in plastic bags or containers that trap moisture.
- *Liquid evidence*: Tubes containing liquid such as blood should not be frozen because the volume of a liquid expands in freezing temperatures and this expansion may lead to cracking. Tubes should be placed in plastic bags to prevent leaks in case of

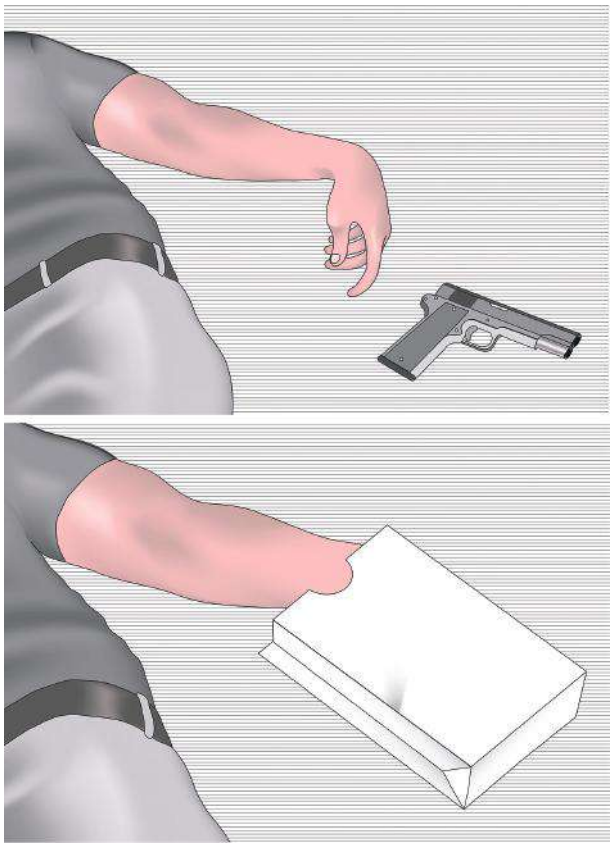

Figure 1.36 Hand bags for protecting the hands of a decedent in an alleged suicide. Gunshot residue can be found on hands after firing a weapon (top). The bagging of the hands, using paper bags and wide rubber bands, prevents the loss of gunshot residue during the transportation of the body (bottom). (© Richard C. Li.)

accidental breakage. Liquid evidence should be transported and submitted to a laboratory as soon as possible after the collection of evidence.

- *Trace evidence*: All such evidence should be wrapped in paper with a druggist's fold (Figure 1.43). The wrapped trace evidence can be packed in an envelope.

Packaged evidence should be properly labeled with a description of the evidence and sealed with evidence tape. It is important for the person packaging the evidence to initial and date across the seal to show authenticity (Figure 1.44). A seal should not be cut when a sealed evidence bag is opened. Instead, an opening can be created by cutting at an area distal from the existing seal. After analysis is complete, the evidence packaging should be resealed. Table 1.1 summarizes additional steps for packaging evidence.

1.9 Final Survey and the Release of the Crime Scene

During a final survey, a discussion with all personnel in the crime scene investigation team should be carried out to thoroughly review all aspects of the search. It is important to ensure

Figure 1.37 Blood cards are typically used for collecting blood evidence from a known source such as a suspect or a victim (top). A manual hole punch can be used to create a blood card punch for DNA extraction. Blood samples are air-dried on blood cards for storage (bottom). (© Richard C. Li.)

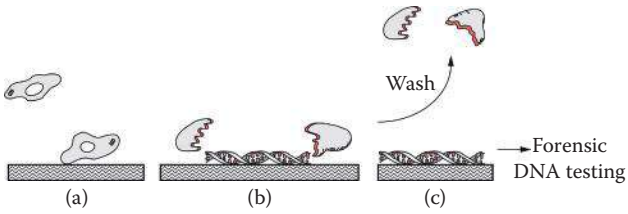

Figure 1.38 Application of Flinders Technology Associates (FTA) filter paper for the collection of biological evidence. (a) Biological fluid with cells is applied to FTA paper. (b) Cells are lysed and DNA is immobilized on FTA paper. (c) Cellular materials are washed away and DNA remaining on the FTA paper can be used for forensic testing. (© Richard C. Li.)

Figure 1.39 Photographic documentation of a knife. Note the evidence tag. (© Richard C. Li.)

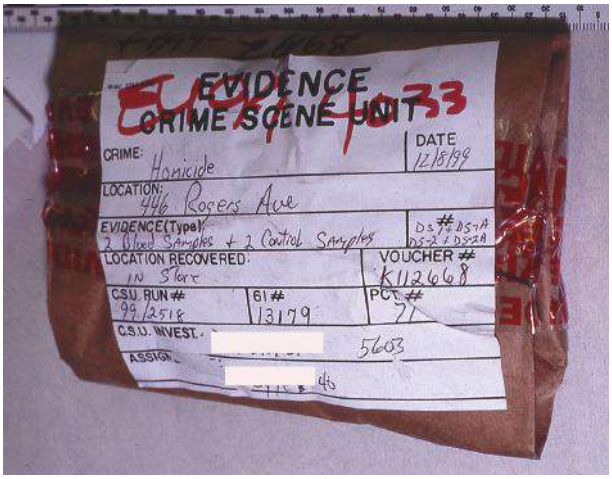

Figure 1.40 Evidence containing dried bodily fluid stains is packed in a paper bag. (© Richard C. Li.)

Figure 1.41 Alleged diluted blood collected from a pipe (left) and placed in plastic containers (right). (© Richard C. Li.)

Figure 1.42 Evidence pouch. The front of the pouch is transparent for viewing the content. The back of the pouch is made of breathable materials allowing wet evidence, such as swabs, to dry inside the pouch. (© Richard C. Li.)

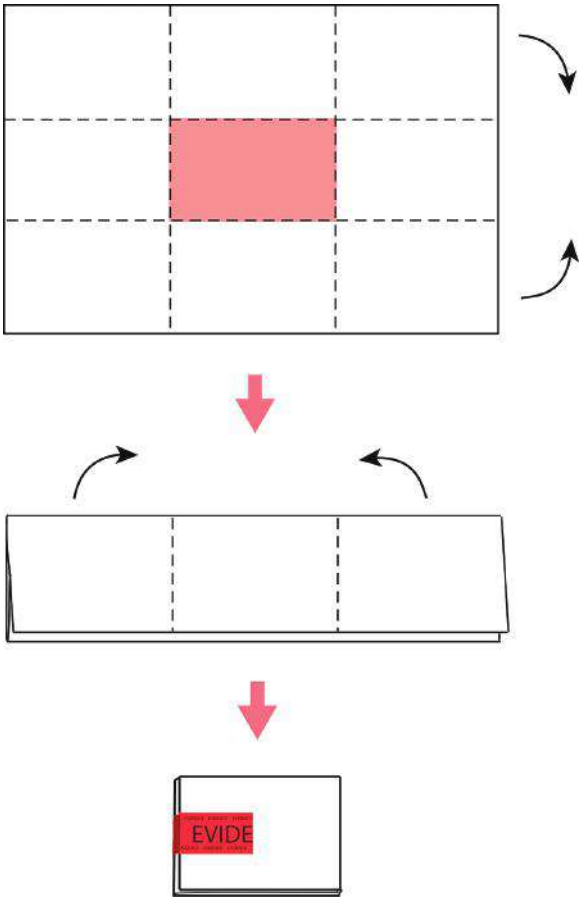

Figure 1.43 An example of a druggist's fold. Trace evidence should be deposited in the center (colored area) of the paper. (© Richard C. Li.)

Figure 1.44 Proper marking of sealed evidence. Note that the evidence packaging was cut a second time and resealed at a different location than that of the preexisting seal. (© Richard C. Li.)

that the scene has been searched correctly and completely, and that no area has been missed or overlooked. All documentation including the chain of custody document must be complete and all evidence should be collected, packed, documented, and marked. Photographs of the final condition of the scene should be taken. Once the final survey is completed, the crime scene can be released. Reentry into the crime scene may require a search warrant after the scene is released. Crime scene release documentation usually includes the time and date of release, to whom it is released, and by whom it is released.

1.10 Crime Scene Reconstruction

Crime scene reconstruction is the scientific process of determining the sequence of events and actions that occurred prior to, during, and after a crime. Reconstruction is carried out based on the information from the crime scene observations and the laboratory examination of physical evidence. The overall scientific process in reconstruction usually involves several steps. The process usually begins with the formulation of questions related to the problems that need to be solved. The questions can refer to the explanation of the specifics of the crime, for instance, "Where was the shooter's position when the shooting occurred?," "Where was the victim's position when shot?," and "What is the muzzle-to-target distance during the shooting?" In order to conduct a thorough crime scene reconstruction, all useful information is collected for review, such as photographs, videotapes, notes, sketches, autopsy reports, and analysis reports of the physical evidence. A hypothesis is then constructed based on the information obtained, which may explain the events and actions involved in a crime. The next step is making predictions that determine the logical consequences of the hypothesis. One or more predictions are selected for testing. The hypothesis is tested by conducting reconstruction experiments. One example of a reconstruction test is bloodstain pattern reconstruction in violent crimes (Chapter 2). Other examples of reconstruction may include trajectory and shooting, glass fracture, and accident reconstruction. The final step is to analyze experimental data and draw a conclusion. The experimental data are analyzed to see if the hypothesis is true or false. Additionally, the interpretation of physical evidence analysis, witness and confession statements, and investigative information should also be considered.

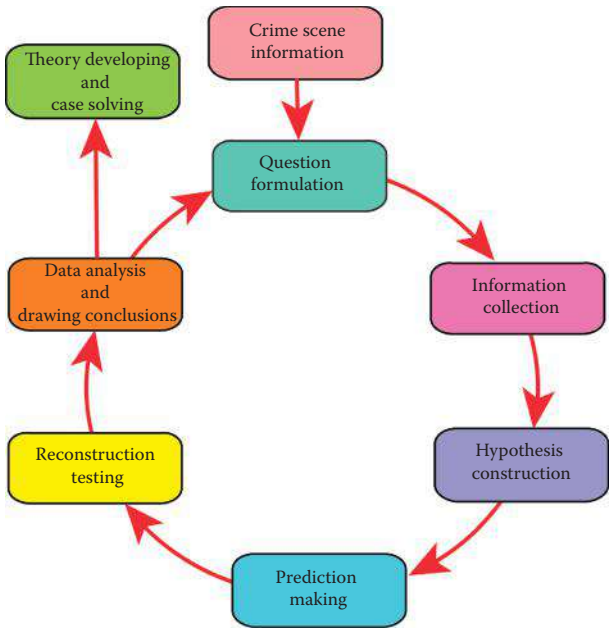

Figure 1.45 The scientific process of crime scene reconstruction. (© Richard C. Li.)

If the results of the experiment are consistent with the hypothesis, a theory can be developed that is intended to provide valuable information to the investigation and future prosecution of a case. Sometimes, forensic scientists may find that the hypothesis is inconsistent with the test results. In that case, an alternative hypothesis needs to be constructed to initiate another reconstruction process. Figure 1.45 illustrates the scientific process of crime scene reconstruction.

Bibliography

Ackermann, K., K.N. Ballantyne, and M. Kayser, Estimating trace deposition time with circadian biomarkers: A prospective and versatile tool for crime scene reconstruction. *Int J Legal Med*, 2010, **124**(5): 387–395.

Adderley, R. and J.W. Bond, The effects of deprivation on the time spent examining crime scenes and the recovery of DNA and fingerprints. *J Forensic Sci*, 2008, **53**(1): 178–182.

Barash, M., A. Reshef, and P. Brauner, The use of adhesive tape for recovery of DNA from crime scene items. *J Forensic Sci*, 2010, **55**(4): 1058–1064.

Benschop, C.C., et al., Post-coital vaginal sampling with nylon flocked swabs improves DNA typing. *Forensic Sci Int Genet*, 2010, **4**(2): 115–121.

Bond, J.W. and C. Hammond, The value of DNA material recovered from crime scenes. *J Forensic Sci*, 2008, **53**(4): 797–801.

Brownlow, R.J., K.E. Dagnall, and C.E. Ames, A comparison of DNA collection and retrieval from two swab types (cotton and nylon flocked swab) when processed using three QIAGEN extraction methods. *J Forensic Sci*, 2012, **57**(3): 713–717.

Buck, U., et al., Accident or homicide—Virtual crime scene reconstruction using 3D methods. *Forensic Sci Int*, 2013, **225**(1–3): 75–84.

Byard, R.W., et al., "Murder-suicide" or "murder-accident"? Difficulties with the analysis of cases. *J Forensic Sci*, 2010, **55**(5): 1375–1377.

Castello, A., F. Frances, and F. Verdu, Solving underwater crimes: Development of latent prints made on submerged objects. *Sci Justice*, 2013, **53**(3): 328–331.

2

Crime Scene Bloodstain Pattern Analysis

Bloodstain pattern analysis is the application of scientific knowledge to the examination and the interpretation of the morphology, the sequence, and the distribution of bloodstains associated with a crime. These analyses may determine the sequence of events; the approximate blood source locations; the positioning of the victim; and the position, the intensity, and the number of impacts applied to the blood source. They also can assist in the determination of the manner of death and can distinguish between accidents, homicides, and suicides. Bloodstain pattern analysis provides critical information for crime scene reconstructions in violent crime investigations (Chapter 1).

2.1 Basic Biological Properties of Human Blood

Blood is a bodily fluid circulating within the body. An average adult has a blood volume of approximately 8% of his or her body weight. Blood consists of a cellular portion as well as a liquid portion known as plasma (Chapter 12). The cellular portion consists of blood cells and platelets. The plasma is mostly composed of water and other substances such as proteins, inorganic salts, and other organic substances. The mass density of blood is only slightly greater than that of water. Blood can form clots (or *thrombi*) that are the result of blood coagulation (Chapter 16). Coagulation begins after an injury occurs, stopping blood loss from a damaged vessel. The normal coagulation time for 1 mL of venous blood in a glass tube is 5–15 min. The coagulation time can be affected by many factors such as blood volume and mechanical disturbance.

2.2 Formation of Bloodstains

The formation of a blood droplet is a complex event that is influenced by viscosity, surface tension, cohesion force, and gravity. Blood is viscous, and blood viscosity is a measure of the blood's resistance to flow. The viscosity of blood is approximately five times greater than that of water. During the formation of a drop of blood, blood leaks out from a blood source. The surface tension of the blood causes it to hang from the opening of a blood source and to form a pendant drop of blood (Figure 2.1). The molecules of a blood drop are held together by the cohesion force to maintain the shape of a blood drop (Figure 2.2). Surface tension causes liquids to minimize their surface. As a result, the formed blood drop is spherical. As the volume of the drop gradually increases and exceeds a certain size, it detaches itself and falls. The falling drop is also held

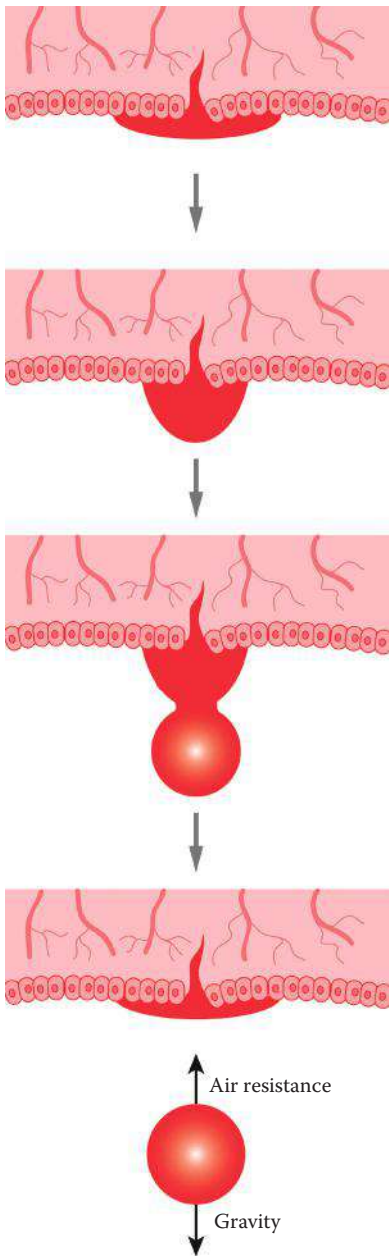

Figure 2.1 Forming a blood drop from a blood source. The blood that leaks out of the blood source forms a pendant drop of blood. As the volume of the pendant drop increases, the drop stretches in a downward direction. Eventually, the drop detaches and falls. The falling drop is largely influenced by the force of gravity and air resistance. (© Richard C. Li.)

together by surface tension. A falling blood drop is influenced by the downward force of gravity acting on the drop and the air resistance that acts in the opposite direction as the drop is in motion (Figure 2.1).

When a bloodstain lands on a surface, the shape and the size of the bloodstain is affected by the texture of the target surface. Bloodstains that land on porous or rough surfaces usually

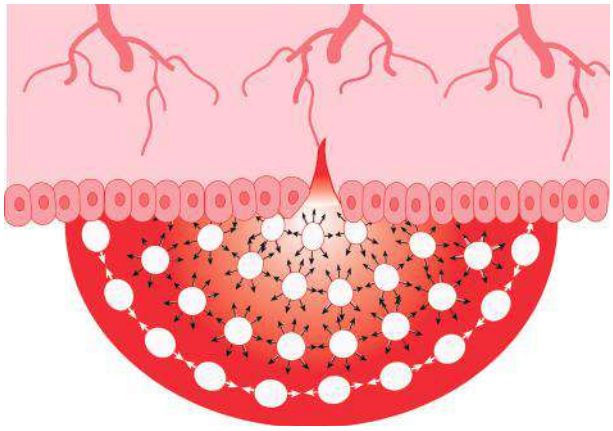

Figure 2.2 The particles of blood are attracted to each other by cohesive forces that are responsible for surface tension. As a result, a formed blood drop is spherical in order to minimize its surface area. Black arrow, cohesive force; white arrow, surface tension. (© Richard C. Li.)

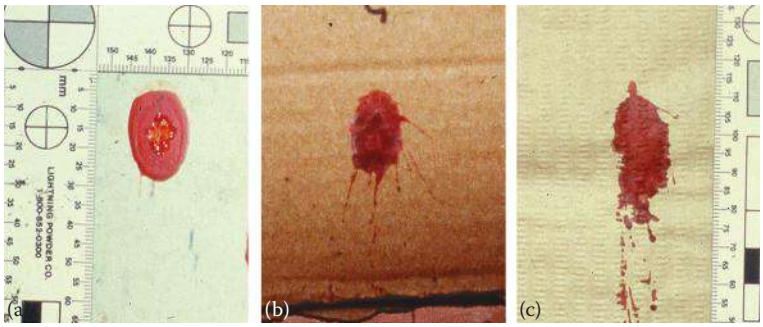

Figure 2.3 The respective morphologies of falling blood drops that land on surfaces with different textures at a 30° angle. (a) Tile, (b) cardboard, and (c) paper towel. (© Richard C. Li.)

have more distortion around the edges of the stains than those that land on smooth surfaces. A comparison of blood dropped onto different textures of target surfaces is shown in Figure 2.3.

2.3 Chemical Enhancement and Documentation of Bloodstain Evidence

Many chemical reagents react with blood to exhibit a color, a chemiluminescent light, or a fluorescent light (Chapter 12). These tests are extremely sensitive and thus are used as chemical enhancement reagents for detecting bloodstains. For bloodstain pattern analysis, the enhancement reagent is primarily used for detecting latent bloodstains such as diluted bloodstains that are visible on enhancement. A commonly used chemical reagent is luminol, which can be used for locating bloodstains at the scene. Other reagents such as phenolphthalein, leucomalachite green, and tetramethylbenzidine are not often used as enhancement reagents but rather as presumptive tests for blood. The positive reactions of all these reagents indicate the presence of blood.

Documenting bloodstain patterns at the scene is a major task of the investigation. Documenting bloodstain evidence can be done using a combination of photography, note-taking, and sketching. The general principle of crime scene documentation is described in Chapter 1. In bloodstain pattern analysis, special attention must obviously be given to bloodstains. The photographic documentation of bloodstains may be performed by multiple means, including film and digital

photography, as well as videotaping. Photographs should be taken with an overall view followed by a medium-range and a close-up view of the bloodstain patterns. A scale of measurement must be included in the photograph, which is critical for bloodstain analysis. To avoid any distortion, the photographs should be taken with the camera lens parallel to the target surface where the bloodstains are located. An overall photograph provides an overall view of the scene including the bloodstain evidence (Figure 2.4a). A midrange photograph provides more details of the bloodstain pattern compared with that of the overall photograph (Figure 2.4b). Single bloodstains should be visible in midrange images. A close-up photograph, usually taken with a macro lens, provides a detailed image of single bloodstains, which is useful for spatter pattern analysis (Figure 2.4c).

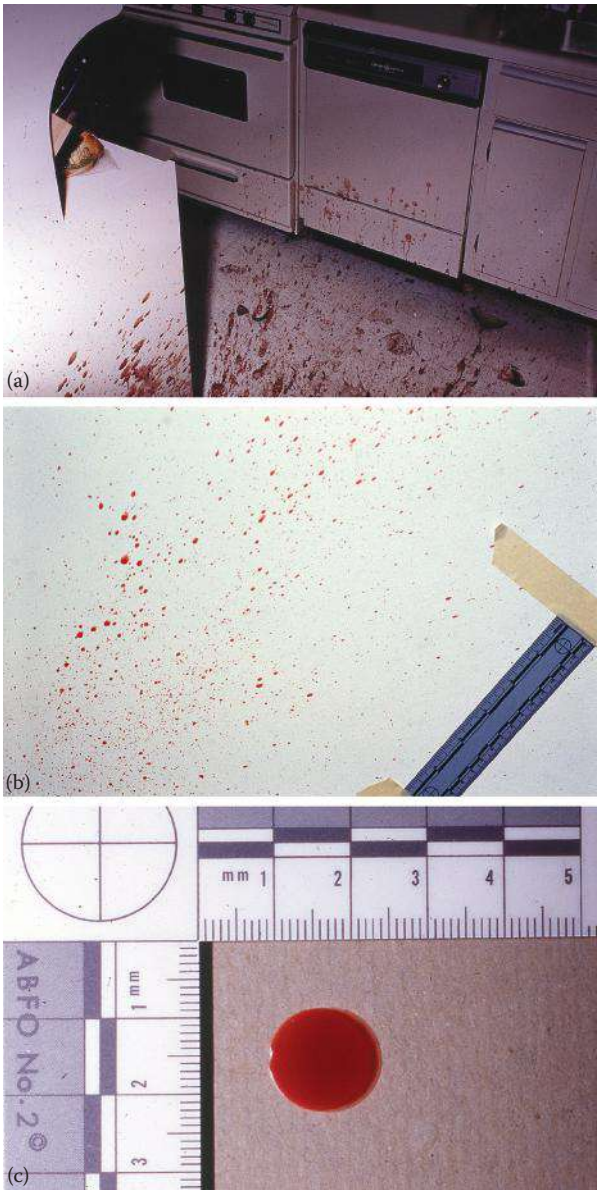

Figure 2.4 Crime scene photographic documentation. (a) Overall, (b) midrange, and (c) close-up photographs. (© Richard C. Li.)

2.4 Analyzing Spatter Stains

A *spatter stain*, based on the recommended terminology of the Scientific Working Group on Bloodstain Pattern Analysis (SWGSTAIN), is "a bloodstain resulting from a blood drop dispersed through the air due to an external force applied to a source of liquid blood." The patterns of spatter stains, including the shape and the size of the stains, are affected by the direction and the angle of impact (discussed in detail in Section 2.5.3) of the spatter stains that are projected. This information can be obtained from an analysis of the patterns of the spatter stains. Thus, it is possible to determine the area of origin (discussed in detail in Section 2.4.4) where an external force was directly applied to the blood source.

2.4.1 Velocity of Blood Droplets

The sizes of bloodstains are affected by the external force that is directly applied on a blood source. Increasing the energy of the external force will reduce the surface tension, thus decreasing the size of the droplets. Since these travelling blood droplets are driven by the energy derived from the external force, the higher the energy, the higher the velocities of the droplets. Bloodstains can be divided into three categories based on different travelling speeds. Low-velocity impact spatter is formed when a blood droplet is travelling at <1.5 m/s. The resulting stains are usually >4 mm in diameter (Figure 2.5a). As the travelling speed of blood droplets increases, the size of the spatter stain decreases. Medium-velocity impact spatter is formed when a blood source is subjected to a force associated with beatings or stabbings. The resulting stains range from 1 to 4 mm in diameter (Figure 2.5b). High-velocity impact spatter is formed when a blood source is subjected to a force associated with shooting using firearms. The resulting stains are usually <1 mm in diameter.

2.4.2 Determining the Directionality of the Stains

In this analysis, the effects of the directionality of the spatter stains projected are examined. SWGSTAIN defines the directionality to be "the characteristic of a bloodstain that indicates the direction blood was moving at the time of deposition." This analysis is applicable when the blood source is projected onto a surface at an angle of between 0° and 90°. Under this condition, the resulting spatter stain is an elongated ellipse (Figure 2.6), which is known as the *parent stain*. Additionally, *satellite stains* in the vicinity of the parent stain can be observed. As defined by SWGSTAIN, a satellite stain is "a smaller bloodstain that originated during the formation of the parent stain as a result of blood impacting a surface." More importantly, a spine is observed, which is the pointed edge away from the parent stain. When such a pattern is observed, the pointed end of the spine always points toward the direction of travel of the bloodstains.

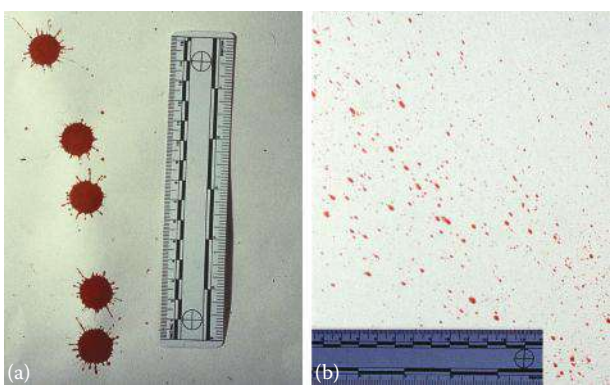

Figure 2.5 Bloodstains can be categorized based on their travelling velocities. (a) An example of a low-velocity impact spatter stain and (b) an example of a medium-velocity impact spatter stain. (© Richard C. Li.)

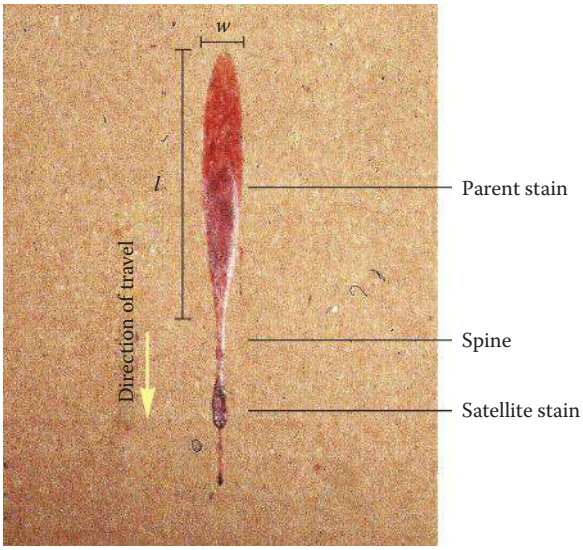

Figure 2.6 The morphology and directionality of a blood spatter stain. The arrow indicates the direction of travel. (© Richard C. Li.)

2.4.3 Determining Angles of Impact

SWGSTAIN defines the *angles of impact* to be "the acute angle (alpha), relative to the plane of a target at which a blood drop strikes the target." The shapes of the spatter stains are affected by the angle of impact. When a blood drop lands on a surface at a perpendicular angle (90°), a circular parent stain is formed (Figure 2.7), where the length and the width of the stain are equal. When a blood drop is projected onto a surface at an angle of between 0° and 90°, the stain is elongated. As the impact angle decreases, the shape of the spatter stain is more elongated (Figure 2.7), in which the length of the stain is greater than the width. It is observed that the ratio of the width and the length of the parent stain is proportional to the sine of the impact angle, which is summarized in the following trigonometric equation:

$$\sin \alpha = \frac{w}{l}$$

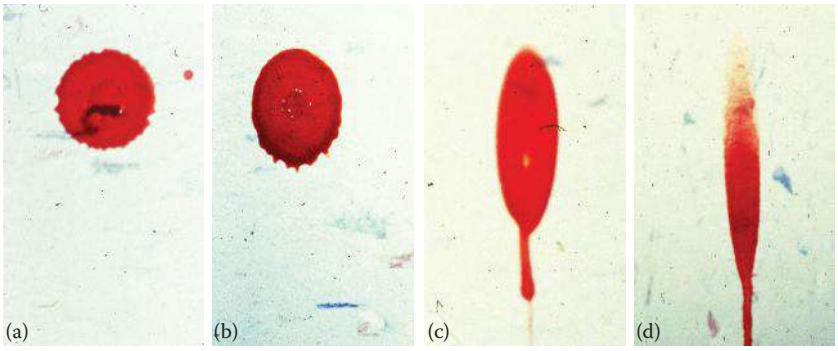

Figure 2.7 The effects of the impact angle on the shapes of blood spatter stains. Spatter stains are projected onto the surface of a ceramic tile at: (a) 90°, (b) 50°, (c) 20°, and (d) 10°. (© Richard C. Li.)

In this equation, α is the angle of impact, *l* is the length of the parent stain (major axis), and *w* is the width of the parent stain (minor axis). Thus, the angle of impact can be determined based on the relationship between the length and the width of the stain (Figure 2.8). Obviously, the measurement of the stain's axes is critical to the accuracy of the calculation of the angle of impact. To produce accurate and reproducible measurements, bloodstain pattern analysis software can be used, which superimposes a scaled close-up image of an individual bloodstain and calculates the angle of impact.

2.4.4 Determining Area of Origin

SWGSTAIN defines the *area of origin* to be "the three-dimensional location from which spatter originated." Using simple trigonometry, the area of origin can be determined based on the measurements from multiple elongated spatter stains (Figure 2.9). This can be accomplished by using the string method or the tangent method.

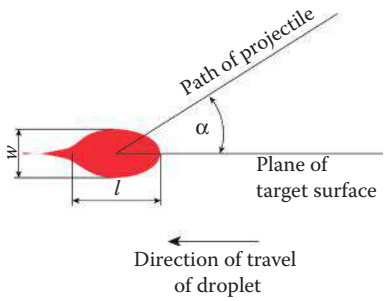

Figure 2.8 Impact angle. The angle between the path of a projectile and the plane of the target surface is shown. α, the impact angle; *l*, the length of the parent stain; and *w*, the width of the parent stain. (© Richard C. Li.)

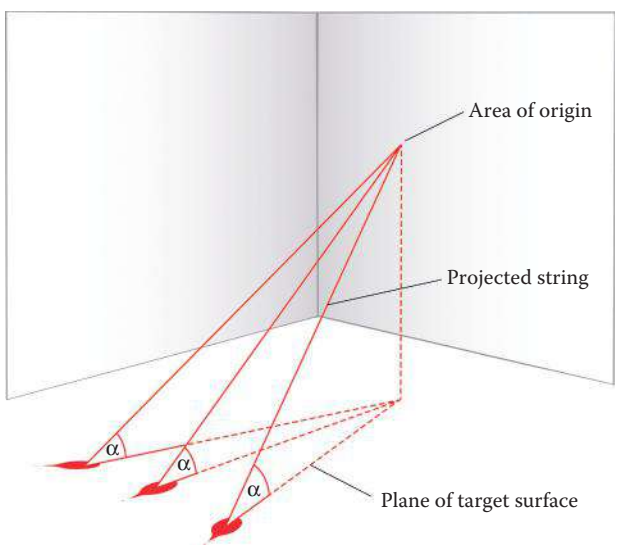

Figure 2.9 Area of origin. The area of origin is determined using the string method. Only three representative bloodstains are shown. α, impact angle. (© Richard C. Li.)

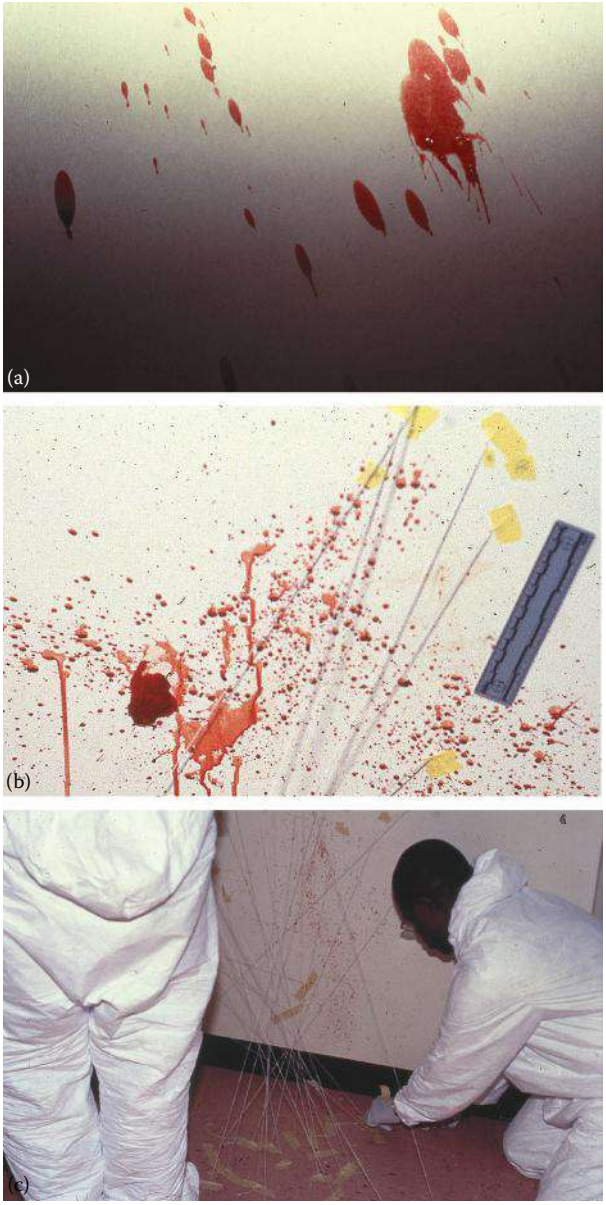

Figure 2.10 Determining the area of origin using the string method. (a) Selecting elongated spatter stains, (b) connecting strings, and (c) setting the path of the strings. (© Richard C. Li.)

In the string method, multiple (approximately two dozen) well-formed, elongated spatter stains are selected for analysis (Figure 2.10a). For each stain, the angle of impact is calculated. A piece of string is then connected between the stain and a surface with one end of the string precisely attached to the spatter stain (Figure 2.10b). The path of the string, indicating the trajectory of the stain, is set using a protractor based on the calculated angle of impact (Figures 2.10c and 2.11). This process is repeated until all the stains that have been selected are processed. For

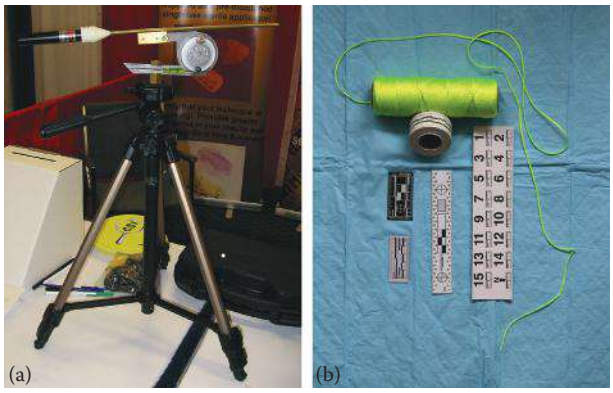

Figure 2.11 Tools for finding the area of origin. (a) Laser trajectory pointer and (b) string and scales. (© Richard C. Li.)

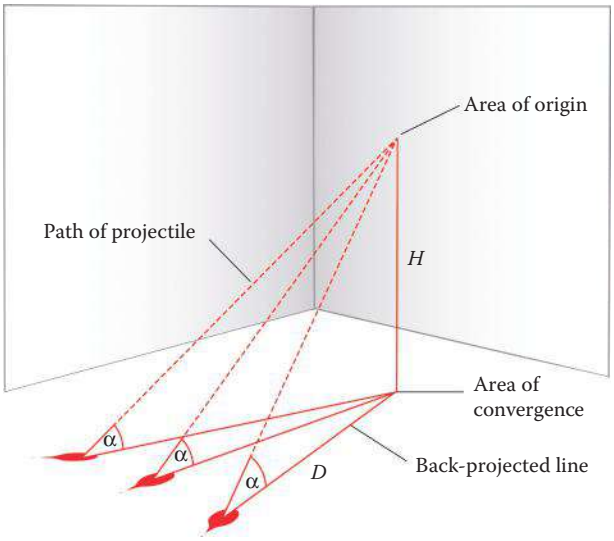

Figure 2.12 Determining the area of origin using the tangent method. Only three representative bloodstains are shown. α, the angle of impact; *H*, the height of the area of origin; and *D*, the distance from the spatter stain to the area of convergence. (© Richard C. Li.)

a spatter pattern generated from a single impact event, the strings converge. The area where the strings meet is the area of origin.

In the tangent method, the directionality of a single stain is determined first. A line is then back projected through the major axis of the bloodstain. For a single impact event, approximately two dozen stains are processed to determine the *area of convergence*. Based on SWGSTAIN's definition, the area of convergence is "the area containing the intersections generated by lines drawn through the long axes of individual stains that indicates in two dimensions the location of the blood source." Next, the angle of impact of each stain is calculated. The distance from the bloodstain to the area of convergence is measured (Figure 2.12). The height of the area of origin is calculated using the tangent function as shown:

$$H = D \cdot \tan \alpha$$

In the equation, α is the angle of impact, H is the height of the area of origin, and D is the distance from the spatter stain to the area of convergence.

2.5 Types of Bloodstain Patterns

Bloodstain patterns can be classified into three basic categories: passive, transfer, and projected bloodstains.

2.5.1 Passive Bloodstains

A *passive bloodstain* is formed due to bleeding from wounds, and the blood is deposited on a surface by the influence of the force of gravity alone. For example, a *drip stain* is formed when a falling drop of blood from an exposed wound or a blood-bearing object lands on a surface. If a blood source is moving, a *drip trail* is formed. A *drip pattern*, which is distinct from a drip stain, is formed when a liquid drips into another liquid, where one or both of the liquids are blood (Figure 2.13). As a result, secondary spatter stains are generated. As the dropping distance of the blood increases, the number of secondary spatter stains usually increases, and the size of these stains decreases. An approximate estimation of the dropping distance is possible. A *splash pattern* is formed when a volume of blood spills onto a surface (Figure 2.14). Splash patterns usually

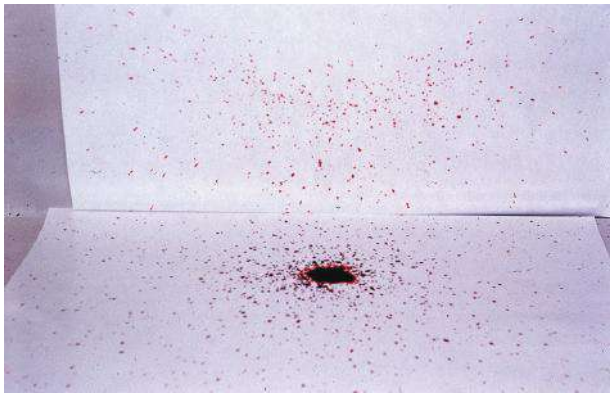

Figure 2.13 A drip pattern. The secondary spatter stains are shown. (© Richard C. Li.)

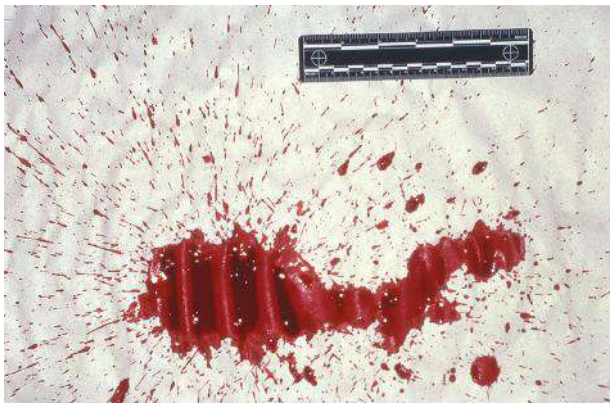

Figure 2.14 A splash pattern. Peripheral, elongated bloodstains are shown. (© Richard C. Li.)

have large stains surrounded by numerous, peripheral, elongated bloodstains. A *flow pattern* is caused by the movement of a large volume of blood on a surface either due to gravity or to the movement of the target such as a victim or postmortem disturbance. A *pool* is a bloodstain resulting from the accumulation of liquid blood on a surface (Figure 2.15). Sometimes, air bubbles in the blood may cause a *bubble ring* pattern (Figure 2.15). If blood is coagulated, gelatinous *blood clots* can be observed. Additionally, a *serum stain*, which consists of the liquid portion of the blood after a clot is formed, may also be present.

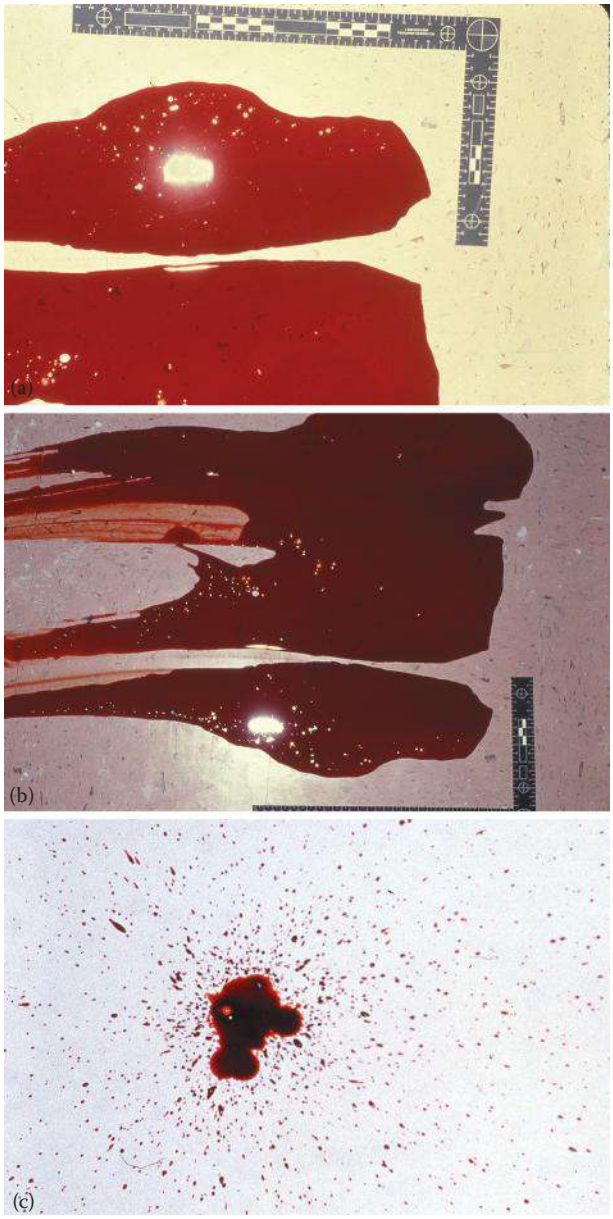

Figure 2.15 Pool and bubble ring patterns. (a) A pool pattern, (b) a disturbed pool pattern, and (c) a splash pattern with a bubble ring. Bubble rings are also present in (a) and (b). (© Richard C. Li.)

2.5.2 Transfer Bloodstains

A *transfer bloodstain*, based on SWGSTAIN, is "a bloodstain resulting from contact between a blood-bearing surface and another surface." For example, a *swipe pattern* is "a bloodstain pattern resulting from the transfer of blood from a blood-bearing surface onto another surface, with characteristics that indicate relative motion between the two surfaces." For example, bloody impressions can provide information about the shape, the size, and the pattern of the objects such as finger ridges, hands, and shoe soles. Examples of hand and shoe swipe patterns are shown in Figures 2.16 and 2.17. A *wipe pattern* is "an altered bloodstain pattern resulting from an object moving through a preexisting wet bloodstain." Examples of wipe patterns are shown in Figures 2.18 and 2.19. A *perimeter* stain, a type of wipe pattern, is a bloodstain that is disturbed before it is dried but it maintains the peripheral characteristics of the original stain (Figure 2.20). Perimeter stain patterns can be useful for the estimation of sequential events of acts. The pattern can also be used to estimate a time frame between the time of bleeding and the subsequent act. However, the drying time of a blood drop varies based on the surrounding conditions. Therefore, it is necessary to carry out a crime scene reconstruction under similar conditions to those of the scene to make such estimations.

2.5.3 Projected Bloodstains

A *projected bloodstain* is formed when a volume of blood is deposited on a surface under a pressure or a force that is greater than the force of gravity. For example, an *impact pattern* is

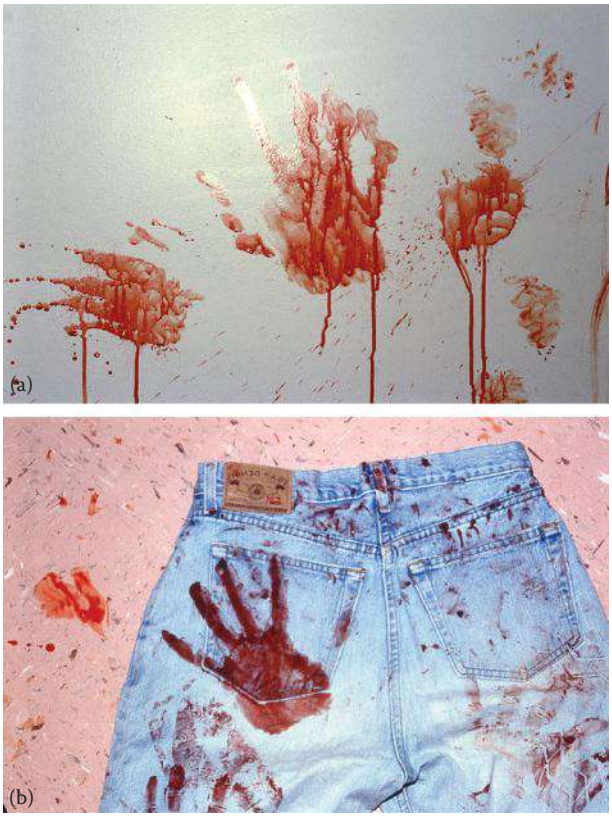

Figure 2.16 Bloody impressions. Bloody handprints are present on (a) a wall and (b) fabric. (© Richard C. Li.)

Figure 2.17 Bloody impressions. Bloody shoe prints on (a) paper and (b) fabric; and (c) bloody footprints on tile. (© Richard C. Li.)

formed when an object strikes liquid blood (Figure 2.21). A *cast-off pattern* is formed when blood drops are released from a moving blood-bearing object (Figure 2.22). Some spatter patterns are often associated with a wound penetrated by a projectile (Figure 2.23). A *forward spatter* is formed when blood drops travel from an exit wound in the same direction as a projectile, while a *back spatter* is formed when blood drops travel from an entry wound in the opposite direction of a projectile. Sometimes, internal bleeding caused by an injury may block the airway. An *expiration pattern* is formed when blood is forced by airflow through the

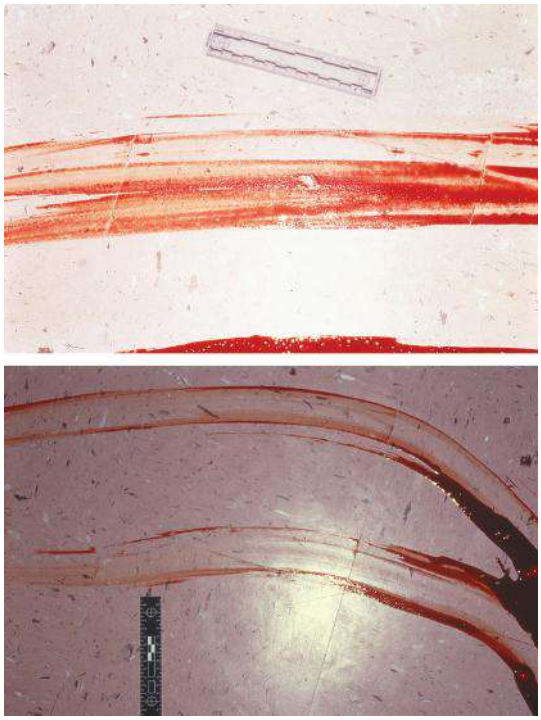

Figure 2.18 Wipe patterns. (© Richard C. Li.)

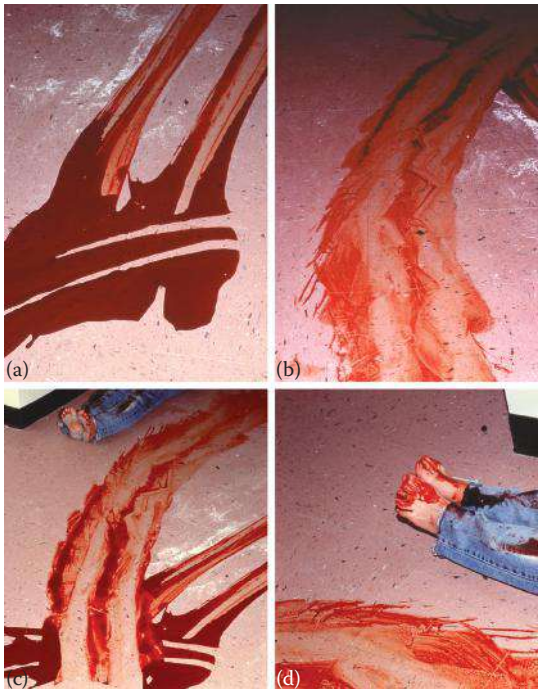

Figure 2.19 A wipe pattern caused by dragging a body through a pool of blood. (a) A pool of blood. Sections of the wipe pattern caused by dragging are shown in (b), (c), and (d). (© Richard C. Li.)

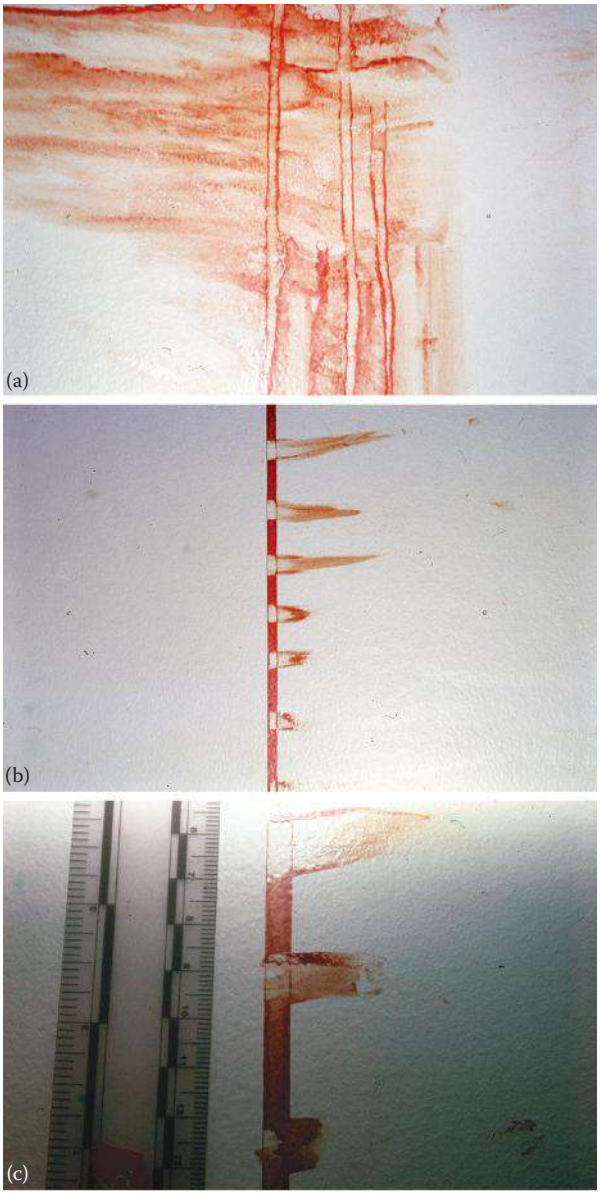

Figure 2.20 Perimeter stains. (a) Peripheral characteristics of the original stains are shown. Perimeter stains were created at different periods of time after the original stain was formed: (b) midrange view and (c) close-up view. (© Richard C. Li.)

trachea and out of the nose or mouth (Figure 2.24). An *arterial spurt pattern* is associated with wounds damaging arterial blood vessels where bloodstains are driven by arterial pressure. Although the shape of arterial patterns varies, these patterns usually have a series of large spurts with fluctuations corresponding to the systolic and the diastolic blood pressures. At a crime scene, if the projectile of bloodstains is blocked by an object, a *void pattern* is formed, which exhibits an area where there is an absence of blood surrounded by continuously distributed bloodstains.

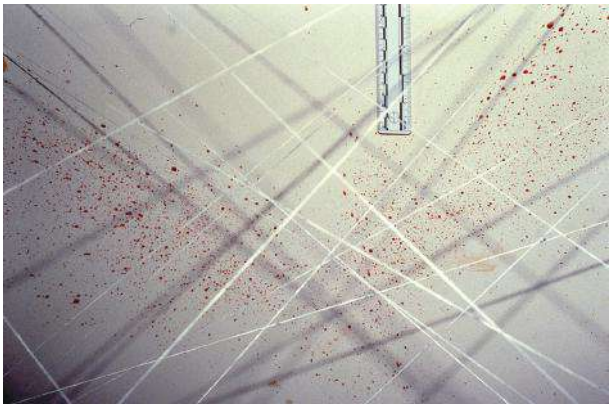

Figure 2.21 An impact bloodstain pattern as a result of using blunt force. (© Richard C. Li.)

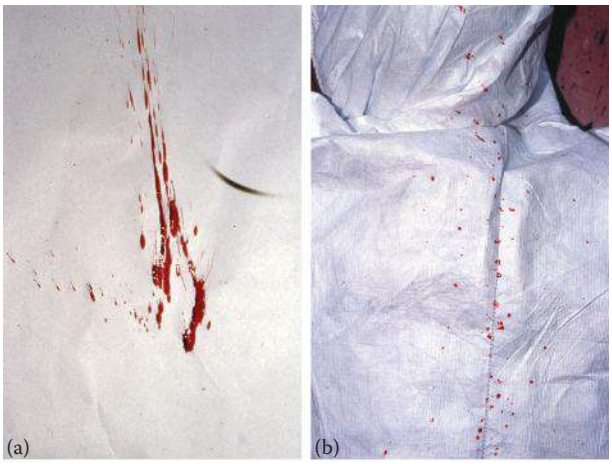

Figure 2.22 Cast-off patterns. Spatter stains are projected onto (a) a covered wall and (b) a lab coat. (© Richard C. Li.)

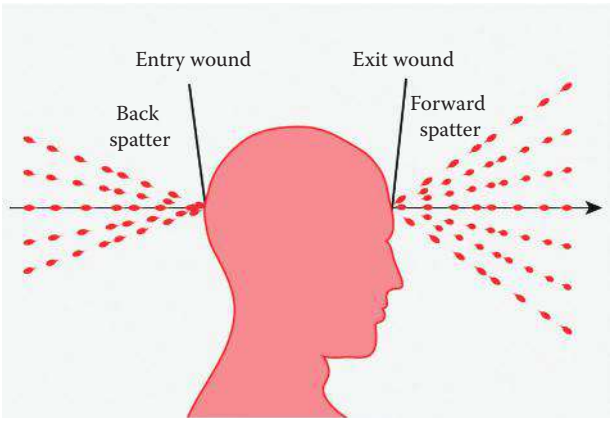

Figure 2.23 Forward and back spatter patterns. Arrow, the direction of a projectile. (© Richard C. Li.)

Figure 2.24 An expiration pattern. (© Richard C. Li.)

Bibliography

Adam, C.D., Fundamental studies of bloodstain formation and characteristics. *Forensic Sci Int*, 2012, **219**(1–3): 76–87.

Adelson, L., The coroner of Elsinore. *New Engl J Med*, 1960, **262**(5): 229–234.

Attinger, D., et al., Fluid dynamics topics in bloodstain pattern analysis: Comparative review and research opportunities. *Forensic Sci Int*, 2013, **231**(1–3): 375–396.

Behrooz, N., L. Hulse-Smith, and S. Chandra, An evaluation of the underlying mechanisms of bloodstain pattern analysis error. *J Forensic Sci*, 2011, **56**(5): 1136–1142.

Benecke, M. and L. Barksdale, Distinction of bloodstain patterns from fly artifacts. *Forensic Sci Int*, 2003, **137**(2–3): 152–159.

Buck, U., et al., 3D bloodstain pattern analysis: Ballistic reconstruction of the trajectories of blood drops and determination of the centres of origin of the bloodstains. *Forensic Sci Int*, 2011, **206**(1–3): 22–28.

Byard, R.W., et al., Blood stain pattern interpretation in cases of fatal haemorrhage from ruptured varicose veins. *J Forensic Leg Med*, 2007, **14**(3): 155–158.

Camana, F., Determining the area of convergence in bloodstain pattern analysis: A probabilistic approach. *Forensic Sci Int*, 2013, **231**(1–3): 131–136.

Connolly, C., M. Illes, and J. Fraser, Affect of impact angle variations on area of origin determination in bloodstain pattern analysis. *Forensic Sci Int*, 2012, **223**(1–3): 233–240.

Davidson, P.L., et al., Physical components of soft-tissue ballistic wounding and their involvement in the generation of blood backspatter. *J Forensic Sci*, 2012, **57**(5): 1339–1342.

de Bruin, K.G., R.D. Stoel, and J.C. Limborgh, Improving the point of origin determination in bloodstain pattern analysis. *J Forensic Sci*, 2011, **56**(6): 1476–1482.

de Castro, T., et al., Interpreting the formation of bloodstains on selected apparel fabrics. *Int J Legal Med*, 2013, **127**(1): 251–258.

Denison, D., et al., Forensic implications of respiratory derived blood spatter distributions. *Forensic Sci Int*, 2011, **204**(1–3): 144–155.

Donaldson, A.E., et al., Using oral microbial DNA analysis to identify expirated bloodspatter. *Int J Legal Med*, 2010, **124**(6): 569–576.

Donaldson, A.E., et al., Characterising the dynamics of expirated bloodstain pattern formation using high-speed digital video imaging. *Int J Legal Med*, 2011, **125**(6): 757–762.

Emes, A., Expirated blood—A review. *J Can Soc Forensic Sci*, 2001, **34**(4): 197–203.

Finnis, J., J. Lewis, and A. Davidson, Comparison of methods for visualizing blood on dark surfaces. *Sci Justice*, 2013, **53**(2): 178–186.

Fujikawa, A., et al., Changes in the morphology and presumptive chemistry of impact and pooled bloodstain patterns by *Lucilia sericata* (Meigen) (Diptera: Calliphoridae). *J Forensic Sci*, 2011, **56**(5): 1315–1318.

Illes, M. and M. Boue, Robust estimation for area of origin in bloodstain pattern analysis via directional analysis. *Forensic Sci Int*, 2013, **226**(1–3): 223–229.

Karger, B., et al., Bloodstain pattern analysis—Casework experience. *Forensic Sci Int*, 2008, **181**(1–3): 15–20.

Kettner, M., F. Ramsthaler, and A. Schnabel, "Bubbles"—A spot diagnosis. *J Forensic Sci*, 2010, **55**(3): 842–844.

Kunz, S.N., H. Brandtner, and H. Meyer, Unusual blood spatter patterns on the firearm and hand: A backspatter analysis to reconstruct the position and orientation of a firearm. *Forensic Sci Int*, 2013, **228**(1–3): e54–e57.

Larkin, B.A. and C.E. Banks, Preliminary study on the effect of heated surfaces upon bloodstain pattern analysis. *J Forensic Sci*, 2013, **58**(5): 1289–1296.

Larkin, B.A.J., et al., Crime scene investigation III: Exploring the effects of drugs of abuse and neurotransmitters on bloodstain pattern analysis. *Anal Methods*, 2012, **4**(3): 721–729.

Lee, W.C., et al., Statistical evaluation of alternative light sources for bloodstain photography. *J Forensic Sci*, 2013, **58**(3): 658–663.

Liesegang, J., Bloodstain pattern analysis—Blood source location. *J Can Soc Forensic Sci*, 2004, **37**(4): 215–222.

Maclean, B., K. Powley, and D. Dahlstrom, A case study illustrating another logical explanation for high velocity impact spatter. *J Can Soc Forensic Sci*, 2001, **34**(4): 191–195.

Makovicky, P., et al., The use of trigonometry in bloodstain analysis. *Soud Lek*, 2013, **58**(2): 20–25.

Maloney, A., et al., One-sided impact spatter and area-of-origin calculations. *J Forensic Ident*, 2011, **61**(2): 123–135.

Nikolic, S. and V. Zivkovic, Bloodstain pattern in the form of gushing in a case of fatal exsanguination due to ruptured varicose vein. *Med Sci Law*, 2011, **51**(1): 61–62.

Peschel, O., et al., Blood stain pattern analysis. *Forensic Sci Med Pathol*, 2011, **7**(3): 257–270.

Pizzola, P.A., et al., Commentary on "3D bloodstain pattern analysis: Ballistic reconstruction of the trajectories of blood drops and determination of the centres of origin of the bloodstains" by Buck et al. [Forensic Sci. Int. 206 (2011) 22–28]. *Forensic Sci Int*, 2012, **220**(1–3): e39–e40; author reply e41.

Power, D.A., et al., PCR-based detection of salivary bacteria as a marker of expirated blood. *Sci Justice*, 2010, **50**(2): 59–63.

Ramsthaler, F., et al., Drying properties of bloodstains on common indoor surfaces. *Int J Legal Med*, 2012, **126**(5): 739–746.

Randall, B., Blood and tissue spatter associated with chainsaw dismemberment. *J Forensic Sci*, 2009, **54**(6): 1310–1314.

Raymond, M.A., E.R. Smith, and J. Liesegang, Oscillating blood droplets—Implications for crime scene reconstruction. *Sci Justice*, 1996, **36**(3): 161–171.

Raymond, M.A., E.R. Smith, and J. Liesegang, The physical properties of blood—Forensic considerations. *Sci Justice*, 1996, **36**(3): 153–160.

Sauvageau, A., et al., Bloodstain pattern analysis in a case of fatal varicose vein rupture. *Am J Forensic Med Pathol*, 2007, **28**(1): 35–37.

Stephens, B.G. and T.B. Allen, Back spatter of blood from gunshot wounds. Observations and experimental simulation. *J Forensic Sci*, 1983, **28**(2): 437–439.

Striman, B., et al., Alteration of expirated bloodstain patterns by *Calliphora vicina* and *Lucilia sericata* (Diptera: Calliphoridae) through ingestion and deposition of artifacts. *J Forensic Sci*, 2011, **56**(Suppl 1): S123–S127.

Taylor, M.C., et al., The effect of firearm muzzle gases on the backspatter of blood. *Int J Legal Med*, 2011, **125**(5): 617–628.

Thanakiatkrai, P., A. Yaodam, and T. Kitpipit, Age estimation of bloodstains using smartphones and digital image analysis. *Forensic Sci Int*, 2013, **233**(1–3): 288–297.

Trombka, J.I., et al., Crime scene investigations using portable, non-destructive space exploration technology. *Forensic Sci Int*, 2002, **129**(1): 1–9.

White, B., Bloodstain patterns on fabrics: The effect of drop volume, dropping height and impact angle. *J Can Soc Forensic Sci*, 1986, **19**(1): 3–36.

Yen, K., et al., Blood-spatter patterns: Hands hold clues for the forensic reconstruction of the sequence of events. *Am J Forensic Med Pathol*, 2003, **24**(2): 132–140.

Yi, R. and Y. Wang, The experimental observation of secondary spatter bloodstain in morphology. *Chin J Forensic Med*, 2011, **26**(1): 26–29.

3

Forensic Biology

A Subdiscipline of Forensic Science

Forensic laboratories provide scientific analysis, evidence evaluation, and consultations to various criminal justice agencies for the investigation of criminal cases. Additionally, forensic laboratories provide expert testimony related to the resolution of criminal cases to the courts.

3.1 Common Disciplines of Forensic Laboratory Services

Many of the disciplines of the forensic laboratory services are commonly practiced in various municipal, county, state, and federal forensic laboratories in the United States. Forensic biology is a subdiscipline of forensic science. A full range of forensic laboratory services, known as "full service," usually includes: crime scene investigation (Figure 3.1); latent print examination (Figure 3.2); forensic biology (Figure 3.3); controlled substance analysis (Figure 3.4); postmortem toxicology (Figure 3.5); questioned document examination (Figure 3.6); firearm, toolmark, and other impression evidence examination (Figures 3.7 through 3.10); explosive and fire debris examination (Figure 3.11); and transfer (trace) evidence examination (Figures 3.12 and 3.13). Table 3.1 describes the services that are normally provided by a forensic laboratory with their respective analyses.

3.2 Laboratory Analysis of Biological Evidence

Forensic biology uses scientifically accepted protocols to analyze biological evidence. Laboratory analysis (Figures 3.14 and 3.15) utilizes scientific techniques for the examination of evidence, the reconstruction of a crime scene, the identification of biological fluids, and the comparison of individual characteristics of biological evidence.

3.2.1 Identification of Biological Evidence

The identification of biological evidence is the first step that is performed before further analyses are carried out. This includes the identification of biological fluids such as blood, saliva, and semen; this process is discussed in more detail in subsequent chapters (Chapters 12 through 14). The *identification* is based on a comparison of *class characteristics*—a set of characteristics that allows a sample to be placed in a category with similar materials. By comparing the

Figure 3.1 Crime scene investigation. Recovery of fired casings at the scene can aid in determining the position of the shooter. (© Richard C. Li.)

class characteristics of a sample with known standards of its class, biological samples can be identified.

3.2.2 Comparison of Individual Characteristics of Biological Evidence

Individual characteristics refer to the unique characteristics of both the evidence and a reference sample such as fingerprints, which share a common origin to a high degree of certainty. An example of biological evidence possessing individual characteristics is DNA polymorphisms. In the case of biological evidence, current forensic DNA profiling can compare individual characteristics of DNA evidence with a known reference sample. It is possible to determine that a biological stain originated from a particular individual, which is useful for human identification. The examination of individual characteristics of evidence can also exclude the possibility of a common origin. The specific methods utilized for the individualization of evidence are also discussed in subsequent chapters (Chapters 19 through 23).

3.2.3 Reporting Results and Expert Testimony

After the analysis of evidence is completed, a report is prepared based on the results of the analysis, which may include sections discussing the specific evidence analyzed, the method of analysis used, the results obtained, and the conclusions drawn. In the case of DNA evidence, the strength of the conclusion is usually evaluated via statistical computations (Chapter 25). A forensic scientist often serves as an expert witness whose testimony

Figure 3.2 Developing latent fingerprints by dusting methods. Using magnetic fingerprint powder (top) that is held by a magnetic applicator (middle). Fingerprints (bottom) dusted by magnetic fingerprint powder, tape lifted and preserved on a fingerprint card. (© Richard C. Li.)

provides professional opinions about the evidence analyzed. Based on the federal rules of evidence, an expert witness is qualified based on his or her knowledge, skill, experience, training, or education, and may give an opinion to the court that is relevant to the analyses conducted. However, forensic scientists must also communicate their findings to attorneys, judges, and members of a jury. This requires the translation of technical information into layman's terms.

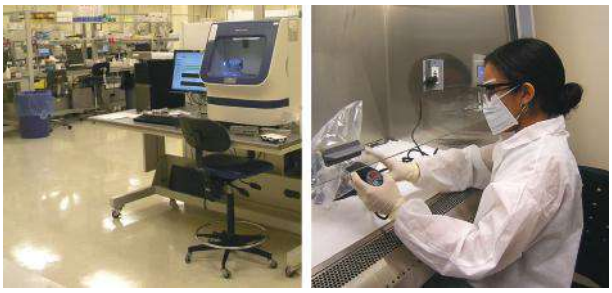

Figure 3.3 A section of a forensic biology laboratory showing an automated electrophoresis instrument used for forensic DNA profiling (left). Processing biological evidence in a biosafety cabinet (right). (© Richard C. Li.)

Figure 3.4 Gas chromatograph (top) and gas chromatograph-mass spectrometer (bottom) used for controlled substance analysis. (© Richard C. Li.)

Figure 3.5 Tissue samples (left) and gas chromatograph-mass spectrometer (right) for forensic postmortem toxicological analysis. (© Richard C. Li.)

Figure 3.6 A digital imaging system, using multiple illumination sources ranging from ultraviolet to infrared wavelengths, for examining altered and counterfeit documents (left). An electrostatic imaging system for detecting indented writing on questioned documents (right). This device generates an electrostatic image of indented writing, which is then visualized using charge-sensitive toners. (© Richard C. Li.)

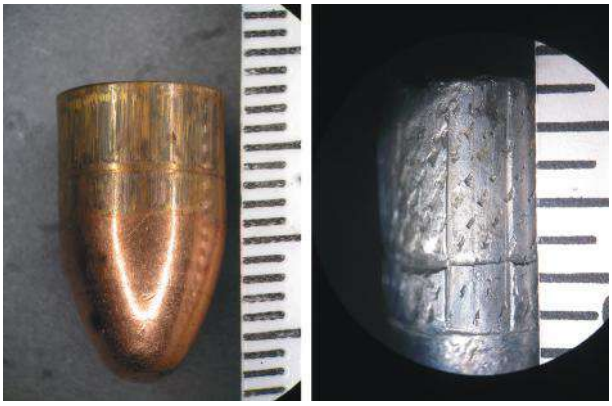

Figure 3.7 Striation marks on fired bullets can be analyzed to match a bullet to a gun. (Courtesy of P. Diaczuk.)

3.3 Forensic Science Services Related to Forensic Biology

A number of specialized forensic science services beyond those provided by forensic laboratories are routinely available to law enforcement agencies. For example, forensic services related to biological evidence and those involving more specialized analysis are available. These services are important aids to a criminal investigation and require the expertise of individuals who have highly specialized skills.

Figure 3.8 Fired hollowpoint bullets (.45 caliber ACP Winchester). Striation marks are visible on the side view of a bullet (right). (© Richard C. Li.)

Figure 3.9 A comparison microscope is used for the simultaneous comparison of two items of firearm evidence side by side. (© Richard C. Li.)

3.3.1 Forensic Pathology

When a death is deemed suspicious or unexplained, medical examiners frequently perform autopsies to determine the exact cause (Figures 3.16 and 3.17). The manner of death is classified into one of five categories based on the circumstances: natural, homicide, suicide, accident, or undetermined. Additionally, a medical examiner participating in a criminal investigation is often responsible for estimating the time of death.

3.3.2 Forensic Anthropology

Forensic anthropology is the identification and the examination of human skeletal remains (Figures 3.18 and 3.19). Skeletal remains can reveal a number of individual characteristics that can be useful in attempting to identify an individual. An examination of bones may reveal an individual's origin, sex, approximate age, race, and the presence of a skeletal injury. A forensic anthropologist may also assist in creating facial reconstructions to aid in the identification of skeletal remains or may be called on to help collect and organize bone fragments in the course of identifying victims of mass disasters such as plane crashes as well as victims in mass graves discovered after wars or genocides.

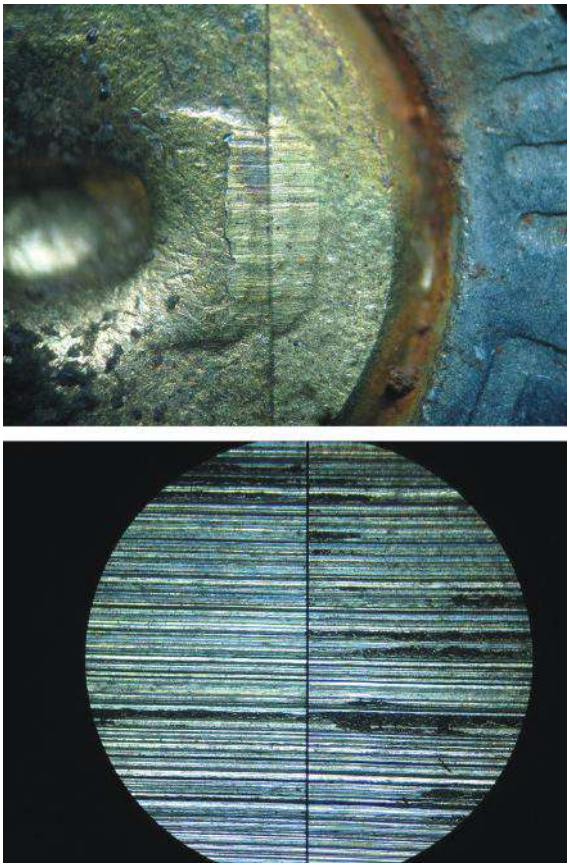

Figure 3.10 Comparing the striations between evidence and reference samples using a comparison microscope: casings (top) and bullets (bottom). (Courtesy of P. Diaczuk.)

Figure 3.11 Scanning electron microscope used in the analysis of gunshot residue and explosives. (© Richard C. Li.)

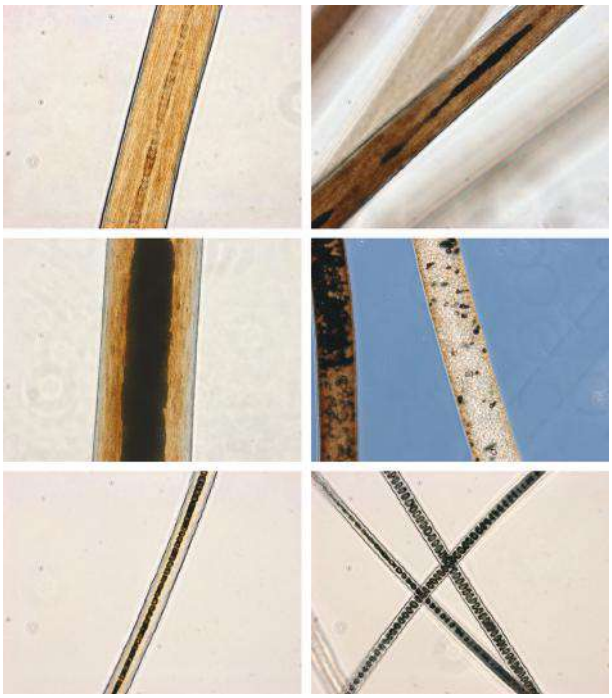

Figure 3.12 Trace evidence such as hairs can be transferred during the acts of a violent crime. The analysis and comparisons of these types of trace evidence can potentially establish a link between a suspect or a victim and a crime scene. Hairs from human (top), horse (middle left), deer (middle right), dog (bottom left), and cat (bottom right). (© Richard C. Li.)

Figure 3.13 Trace evidence: fibers. Cotton (left), nylon (middle), and polyester (right). (© Richard C. Li.)

3.3.3 Forensic Entomology

The study of insects in relation to a criminal investigation is known as forensic entomology. This forensic discipline is valuable for estimating the time of death when the circumstances surrounding the crime are otherwise unknown. The stages of development of certain insect species present in or on a body can be identified and allow a forensic entomologist to approximate how long the body was left exposed (Figure 3.20).

Table 3.1 Common Services Provided by US Forensic Laboratories

Service	Function	Method
Crime scene investigation	Evidence recognition, documentation, collection, and preservation	Crime scene responses and related endeavors are diverse and vary with case and type of evidence
Latent print examination	Analysis of friction ridge detail in fingerprints Activities include visualization, recording, comparison, storage, and recovery of latent prints	Alternate light sources, physical (powder) and chemical enhancements Direct lifts, photography, and digital imaging Use of an Automated Fingerprint Identification System (AFIS) database
Forensic biology	Identification of biological fluids (blood, semen, and saliva) DNA profiling for individualization	Serological and biochemical methods Polymerase chain reaction (PCR)-based methods Automated electrophoresis platforms Use of Combined DNA Index System (CODIS)
Controlled substance analysis	Identification and quantification of drugs present in submitted evidence	Microscopic, chemical, chromatographic, and spectroscopic methodologies Gas chromatography–mass spectrometry or infrared spectrophotometry
Postmortem toxicology	Determination of concentrations of substances and their metabolites in biological fluids or tissues	Immunoassays and chemical methods Confirmatory techniques such as gas and liquid chromatography–mass spectrometry
Questioned document examination	Investigation of forgeries, tracings, disguised handwritings, computer manipulation of images, and recovery of altered documents Analysis of papers, inks, toners, word processors, typewriters, copiers, and printers	Macroscopic and microscopic comparisons Chromatographic and spectroscopic methods
Firearm and toolmark examination	Identification of firearms, tools, and other implements (expertise achieved predominantly through experience)	Microscopic comparisons of questioned and authenticated impressions Comparison of striae on recovered bullets Use of National Integrated Ballistics Information Network (NIBIN)
Explosive and fire debris examination	Identification, recovery, and detection of bulk explosives, residues, debris, and accelerants	Microscopic, spectroscopic, and chromatographic methods Gas chromatography–mass spectrometry may be needed to adequately characterize sample
Trace evidence examination	Analysis of transferred evidence such as hairs, fibers, soil, paints, and glass	Microscopic analysis of evidence with gas chromatograph-mass spectrometers, FTIR microscopes, scanning electron microscopes, basic and advanced microscopy, and capillary electrophoresis

Source: Adapted from US Department of Justice, Office of Justice Programs, National Institute of Justice (NIOJ), *Forensic Science: Review of Status and Needs*, 1999, US Department of Justice, Washington, DC.

Figure 3.14 Photographic documentation of a bloodstained shirt after visual examination. (© Richard C. Li.)

Figure 3.15 A multiwavelength viewing and imaging device used for the examination of various types of evidence, including bodily fluid stains. (© Richard C. Li.)

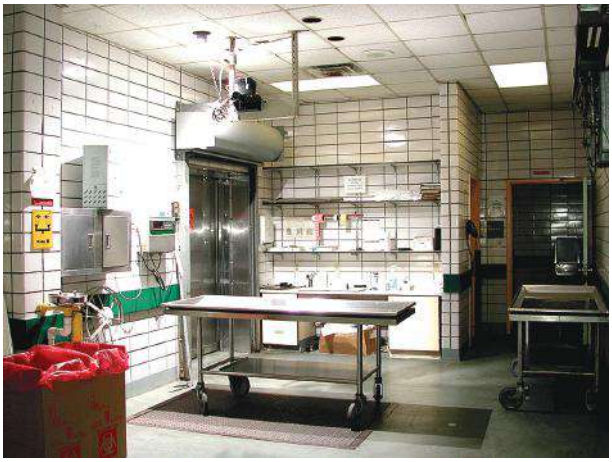

Figure 3.16 View of a forensic pathology facility. (Courtesy of G. Ledwell.)

Figure 3.17 Photographic documentation prior to autopsy (left). Preparing specimens for histological sections for forensic pathological examination (right). (Courtesy of G. Ledwell.)

Figure 3.18 Buried human skeletons recovered by forensic anthropologists. (Courtesy of H. Brewster.)

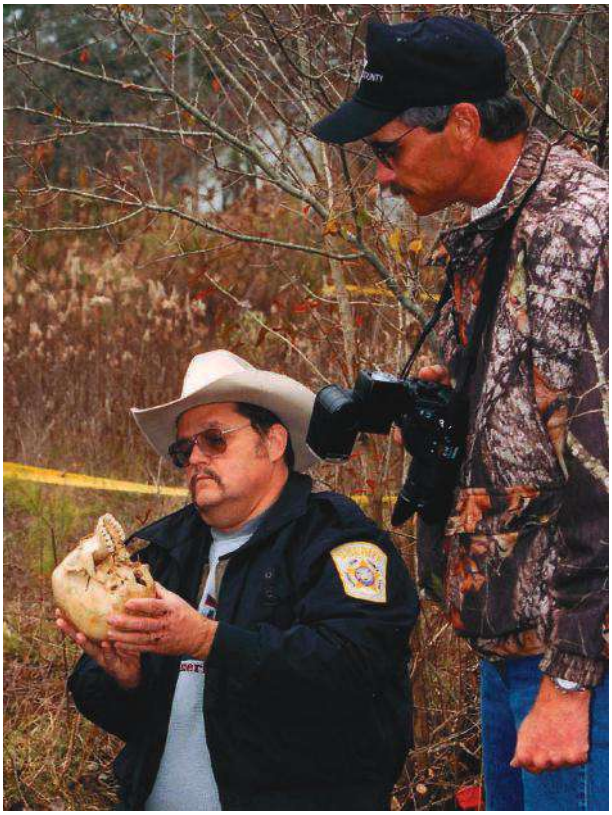

Figure 3.19 Human skull recovered at the scene. (Courtesy of H. Brewster.)

Figure 3.20 Insects found on a dead animal (left). Blow fly specimen. This insect is commonly encountered at crime scenes (right). (Courtesy of K. Wendler.)

3.3.4 Forensic Odontology

Practitioners of forensic odontology participate in the identification of victims whose bodies are left in an unrecognizable state. The characteristics of teeth, their alignment, and the overall structure of the mouth provide evidence that can identify a specific person. Dental records such as x-rays and dental casts allow a forensic odontologist to compare a set of dental remains with an alleged victim. Another application of forensic odontology in a criminal investigation is bite mark analysis (Figure 3.21). A forensic odontologist can analyze the marks left on a victim and compare them with the tooth structures of a suspect to make a comparison.

Figure 3.21 A dental cast can be utilized for identifying bite marks. (© Richard C. Li.)

3.4 Brief History of the Development of Forensic Biology

The developmental history of modern forensic biology spans three stages: (1) antigen polymorphism, (2) protein polymorphism, and (3) DNA polymorphism. Figure 3.22 illustrates this history.

3.4.1 Antigen Polymorphism

The human ABO blood groups were discovered in 1900 by Karl Landsteiner in a study of the causes of blood transfusion reactions. Landsteiner's discovery made blood transfusions feasible, and he received the Nobel Prize in 1930 when he revealed the four groups of human blood cells designated A, B, AB, and O. By the 1960s, a dozen more blood group systems had been characterized, and at least 29 systems are currently known (Chapter 18).

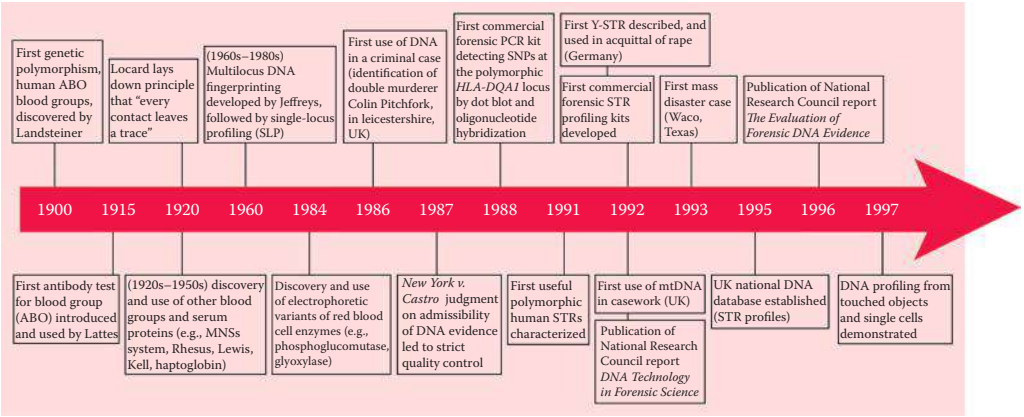

Figure 3.22 A brief history of the development of forensic biology. (From Jobling, M.A. and Gill, P., *Nat Rev Genet*, 5, 739–751, 2004. With permission.)

Subsequent studies found that the blood types in the ABO system were inherited, and the frequencies with which the four types appeared in specific human populations were found to differ. This led to the discovery of the first antigen polymorphic marker for use in human identification in forensic cases. In the past, forensic laboratories utilized blood group systems in a discipline known as forensic serology. While it is possible to exclude a suspect through the use of blood group typing, the evidence for the inclusion of a suspect is weak due to the high probability of a coincidental match between two unrelated persons.

3.4.2 Protein Polymorphism

Because of the limitations of antigen polymorphism, protein polymorphism was introduced for forensic identification (Chapter 18). Initially, a few polymorphisms in serum proteins and erythrocyte enzymes were reported. By the 1980s, however, approximately a hundred protein polymorphisms had been discovered. A few systems were commonly used in forensic laboratories, including the polymorphisms of erythrocyte enzymes, serum proteins, and hemoglobin. Blood groups and protein polymorphism analysis were combined in forensic investigations to lower the probability of a match between two unrelated individuals. However, more powerful methods were still sought.

3.4.3 DNA Polymorphism

The human genome contains all the necessary biological information for cellular and organ structure and function. It consists of the *nuclear genome* and the *mitochondrial genome* (Chapter 23 discusses the mitochondrial genome). The human nuclear genome, a set of 23 chromosomes, contains approximately 3 billion base pairs (bp). The Human Genome Project was initiated in 1990 to sequence the entire human nuclear genome. In 2003, 99% of the human genome, including the most important parts of the genome, was sequenced. Further analyses on the human genome sequences continue. The genome contains genes and intergenic noncoding sequences.

3.4.3.1 Genes and Related Sequences

Approximately 20,000–25,000 genes have been identified in the human genome, which encode the information for the synthesis of proteins. The functions of nearly half of these genes have been identified. Most encode the proteins that are responsible for the maintenance of the genome, the functioning of the cells, the immune response, and the structural proteins of cells.

Most human genes are discontinuous. The coding regions of genes are called *exons* and are separated by *introns*. During gene expression, the precursor messenger RNA transcript (pre-mRNA), consisting of both the exons and introns, is produced. The mRNA is a template for protein synthesis in which the sequence is based on a complementary strand of DNA. Through the process of splicing, the introns are removed and the exons are joined, producing the spliced mRNA form, which can be used for protein synthesis via the translation process. Other gene-related sequences include those responsible for gene transcription such as *promoter sequences*; those responsible for gene regulation such as *cis-regulatory sequences* (or enhancers); and *untranslated sequences*, which are transcribed but do not encode proteins. Figure 3.23 depicts the features of a representative human gene.

3.4.3.2 Intergenic Noncoding Sequences

More than 90% of the human genome sequence consists of intergenic noncoding sequences located between genes. The functions of these sequences are yet to be discovered. The intergenic noncoding sequences contain large quantities of various types of *repetitive DNA*, which falls into two categories: tandem repeats and genome-wide or interspersed repeats.

Tandem repeats are repeat units placed next to each other in an array. One type is called *satellite DNA* because of the observation of satellite bands containing DNA with tandem repeats

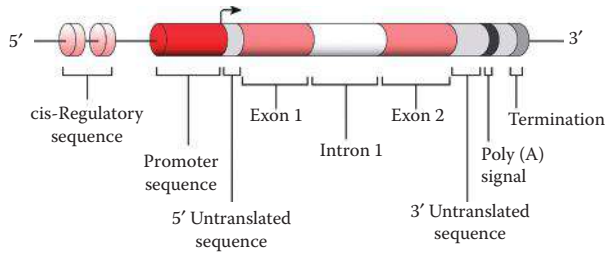

Figure 3.23 Gene structure. Transcription, which can be regulated by the cis-regulatory sequence, is initiated at the transcription start site (arrow) near the promoter. The exons, noncoding introns, and the untranslated sequences are also shown. (© Richard C. Li.)

during density gradient centrifugation. Satellite DNA can be found at centromeres and telomeres consisting of regions composed of long stretches of tandem repeats. Minisatellites and microsatellites are two other types of shorter tandem repeats. *Minisatellites*, also known as *variable number tandem repeats* (VNTRs), form arrays of tandem repeats with a *repeat unit length* from several to hundreds of base pairs. In a *microsatellite*, also known as a *short tandem repeat* (STR) or a *simple sequence repeat* (SSR), the repeat unit length can be 2–6 bp long.

Mobile elements (*interspersed repeats*) are randomly located throughout the human genome (Figure 3.24). Two human types have been characterized: *DNA transposons* and *retrotransposons*. The mobile elements change their locations, a process called *transposition*, by which these sequences are inserted into a new site in the genome. The transposition of DNA transposons is through a "cut-and-paste" mechanism. During transposition, DNA transposons are excised from one site and inserted at a new site in the genome. In contrast, retrotransposons duplicate themselves during transposition and propagate throughout the genome, which is a copy-and-paste mechanism: a copy of the original retrotransposons is generated at the new site and the original copy is retained. Additionally, the transposition of retrotransposons requires an RNA intermediate, a process called *retrotransposition*. Retrotransposons have two subtypes: *long terminal repeat* (LTR) and *non-LTR retroposons*. The non-LTR retroposons can be further divided into two subtypes: *long interspersed elements* (LINEs) and *short interspersed elements* (SINEs). Alu elements are the most abundant type of human SINE. There are more than one million copies of Alu elements in the human genome (Figure 3.25). Some members of the Alu elements are

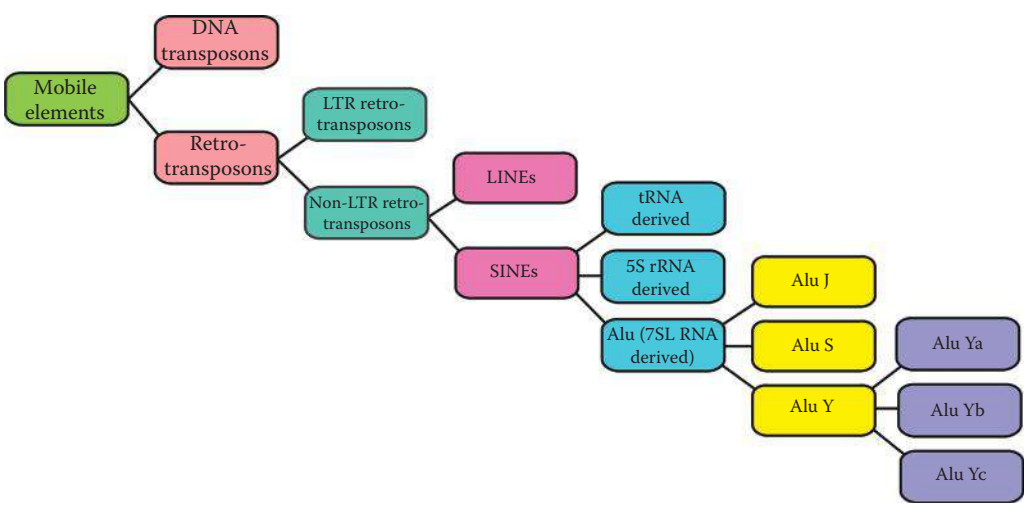

Figure 3.24 Mobile elements classification. (© Richard C. Li.)

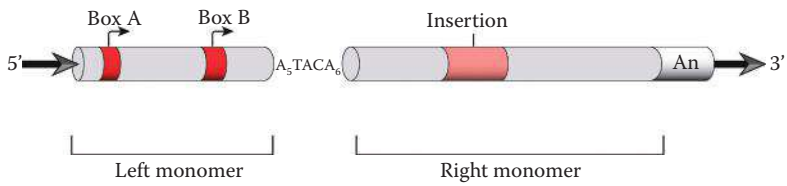

Figure 3.25 Structure of Alu elements. The name, Alu elements, was given since these elements usually contain DNA sequences cleaved by the restriction enzyme *Alu*I. The Alu element is a dimeric structure including two regions separated by an A-rich sequence. The 5′ region has boxes A and B containing RNA polymerase III promoter sequences. The 3′ region contains a 31-nucleotide insertion. There is a short polyadenylation tail located at the 3′ end of the Alu element (terminal A-stretch). Full-length Alu elements are approximately 300 bp long. (© Richard C. Li.)

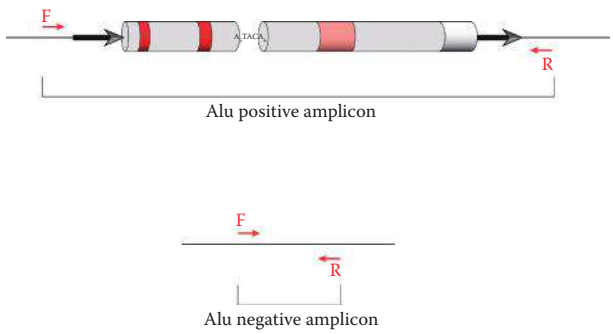

Figure 3.26 Detecting Alu element insertion polymorphism for human identification. The Alu element insertion site is amplified using the polymerase chain reaction assay. The presence and absence of the Alu element at the site can be detected based on the length of the DNA fragment analyzed. F, forward primer; R, reverse primer. (© Richard C. Li.)

polymorphic for the presence or absence of insertion that can potentially be used for forensic human identification (Figure 3.26).

3.4.3.3 Human DNA Polymorphic Markers

Most individual human genome sequences are very similar. However, variations in sequences do occur. The differences between individual genomes that occur at the DNA level are called *DNA polymorphisms*. In particular, a DNA polymorphism with alternative forms of a chromosomal locus that differ in nucleotide sequence is known as a *sequence polymorphism*. A DNA polymorphism that differs in the numbers of tandem repeat units is known as a *length polymorphism*. A DNA polymorphism can occur anywhere in a genome including genes and other chromosomal locations. Many DNA polymorphisms are useful for genetic mapping studies and hence are called *DNA markers*. DNA polymorphisms form the basis of forensic DNA profiling. The focus of this text is on the human genome, but polymorphisms also occur in the genomes of other organisms.

Most DNA polymorphisms are single nucleotide polymorphisms (SNPs) involving a single-base-pair change or a point mutation. Over one million SNPs have been identified. SNPs arise by spontaneous mutation. Most SNPs occur in noncoding regions of the genome, although some appear in coding regions as well. Other forms of DNA polymorphisms are tandem repeats such as STRs and VNTRs. Although their biological functions are unknown, STRs and VNTRs are very useful for forensic DNA analysis. Many are highly polymorphic, and the number of repeat units varies greatly among different individuals of a population. It is unlikely that two unrelated

individuals will have exactly the same combination of STR or VNTR polymorphisms if sufficient markers are examined. Thus, a resulting genetic profile can be used for human identification.

Alternative forms of DNA polymorphisms are called *alleles*. The same allele present in both homologous chromosomes is referred to as *homozygous*. Two different alleles present in homologous chromosomes are referred to as *heterozygous*. A combination of alleles at a given locus is a *genotype*. In forensic analysis, the genotype for a panel of analyzed loci is called the *DNA profile*.

3.4.3.4 Forensic DNA Polymorphism Profiling

In 1984, Sir Alec Jeffreys (Figure 3.27) developed a DNA profiling technique using a VNTR technique involving multilocus profiling and later followed by single-locus profiling (Chapter 19). This technique led to the solving of a double murder that had been committed in Leicestershire (United Kingdom) in the 1980s. The case was the first to apply DNA evidence to a criminal investigation. During the investigation, DNA profiling not only identified the true perpetrator but it also excluded an innocent suspect. This case demonstrated DNA profiling's great potential in forensic investigations.

DNA profiling offered a number of advantages compared with earlier systems. The most important is the ability of the technique to reveal far greater individual variability in DNA than can be revealed by antigen and protein polymorphic markers. The probability of two unrelated individuals having the same DNA profile is very low. The great variability of DNA polymorphisms has made it possible to offer strong support for concluding that DNA from a suspect and from a crime scene originated from the same source. This technique was subsequently implemented in forensic laboratories worldwide.

In the mid-1980s, Kary Mullis and his coworkers developed the polymerase chain reaction (PCR) technique, which amplifies a small quantity of DNA. Mullis's invention had a powerful impact on molecular biology and earned him a Nobel Prize in 1993. Since the introduction of PCR, new techniques have been developed for forensic DNA testing purposes.

Figure 3.27 Sir Alec Jeffreys. (© Richard C. Li.)

The application of PCR-based assays makes forensic DNA analysis possible when only minute quantities of DNA can be recovered from a crime scene, for example, from hairs and cigarette butts. These assays have greatly increased the sensitivity of forensic DNA testing. The first forensic application of a PCR-based assay utilizing SNPs at the *HLA-DQA1* locus (formerly called DQα) was developed in 1986 (Chapter 22). One major disadvantage of the assay was the high probability (approximately 1 in 4000) of a match between two unrelated persons. Amplified fragment length polymorphism (AFLP) at the D1S80 locus has also been implemented in forensic laboratories. The D1S80 locus is a small-size VNTR marker that can be amplified by PCR. The *HLA-DQA1* and AFLP assays were used for some years until the introduction of STR assays.

In the late 1990s, forensic laboratories started utilizing STR loci. STRs have a number of advantages compared with VNTRs. For example, STRs can be amplified by PCR because of their smaller size, which greatly increases the sensitivity of the assay. Furthermore, STR markers are as highly variable as VNTRs. With the application of multiple STR loci, the probability of a match between two unrelated persons becomes extremely low. As a result of DNA testing, perpetrators have been identified and wrongly convicted innocent people have been exonerated (Chapter 20).

In 1995, the United Kingdom established the first national DNA database for criminal investigations. By the end of 1998, several other countries, including the United States, had created their own national DNA databases. The United States has selected 13 STR loci for the Combined DNA Index System (CODIS). These national DNA databases play important roles in solving criminal cases.

Another technique known as mitochondrial DNA (mtDNA) profiling has also been used for forensic testing. mtDNA is maternally inherited genetic material and is therefore particularly useful for human identification. Each cell contains multiple copies of mtDNA. Thus, mtDNA typing can be useful for analysis when nuclear DNA is severely degraded or present in very limited amounts, such as in cases involving decomposed human remains. Alternatively, polymorphic markers at the Y chromosome have also been utilized for forensic DNA testing. Y chromosomal markers are paternally inherited so they can be used for paternity testing. These markers are also very useful in analyzing DNA from multiple contributors in sexual assault cases.

Bibliography

Aleksander, A., Forensic expertise and judicial practice: Evidence or proof? *J Eval Clin Pract*, 2012, **18**(6): 1147–1150.

Alvarez-Cubero, M.J., et al., Genetic identification of missing persons: DNA analysis of human remains and compromised samples. *Pathobiology*, 2012, **79**(5): 228–238.

Andelinovic, S., et al., Twelve-year experience in identification of skeletal remains from mass graves. *Croat Med J*, 2005, **46**(4): 530–539.

Apostolov, A., S. Hristov, and E. Angelova, DNA identification of biological traces and interpretation in a sexual assault case. *Am J Forensic Med Pathol*, 2009, **30**(1): 57–60.

Aquila, I., et al., The role of forensic botany in crime scene investigation: Case report and review of literature. *J Forensic Sci*, 2014, **59**(3): 820–824.

Asari, M., et al., Multiplex PCR-based Alu insertion polymorphisms genotyping for identifying individuals of Japanese ethnicity. *Genomics*, 2012, **99**(4): 227–232.

Auvdel, M.J., Comparison of laser and ultraviolet techniques used in the detection of body secretions. *J Forensic Sci*, 1987, **32**(2): 326–345.

Baraybar, J.P., When DNA is not available, can we still identify people? Recommendations for best practice. *J Forensic Sci*, 2008, **53**(3): 533–540.

Bauer, C.M., et al., Molecular genetic investigations on Austria's patron saint Leopold III. *Forensic Sci Int Genet*, 2013, **7**(2): 313–315.

Bogdanowicz, W., et al., Genetic identification of putative remains of the famous astronomer Nicolaus Copernicus. *Proc Natl Acad Sci U S A*, 2009, **106**(30): 12279–12282.

4

Sources of Biological Evidence

Biological evidence analysis is one of the standard forensic examinations in the investigation of a wide variety of crimes. In particular, DNA evidence facilitates investigators' efforts to link offenders to crime scenes by matching DNA profiles. DNA evidence can also be used to eliminate suspects. The DNA evidence that is routinely encountered at crime scenes can often be categorized into several groups or types. Table 4.1 lists sources of DNA that are frequently found on personal items. Figures 4.1 and 4.2 illustrate representative types of evidence that are processed and their success rates.

4.1 Bodily Fluids

Bodily fluids and their stains are useful biological evidence for forensic, serological, and DNA analysis and may be useful in solving crimes. The most common bodily fluids in forensic analysis are blood, seminal fluid, and saliva. Blood evidence, such as peripheral blood that circulates through the heart, arteries, veins, and capillaries, is one of the most common types of biological evidence that is found at crime scenes (Chapter 12). The fluid portion of blood is called *plasma*, a subcompartment of extracellular fluid, which is the bodily fluid outside cells. Blood contains various suspended blood cells. The cellular portion of the blood consists of *erythrocytes* (also known as red blood cells), *leucocytes* (also known as white blood cells), and *platelets*. Because mature human erythrocytes and platelets do not have nuclei, they are not useful sources of nuclear DNA. For forensic DNA profiling, the nuclear DNA in blood samples (Figure 4.3) is primarily isolated from leucocytes, which are nucleated. Besides peripheral blood, menstrual blood can be analyzed to investigate the possibility of the occurrence of a sexual assault crime (Chapter 16).

Other bodily fluids are *transcellular fluids*. These fluids are considered to be a subcompartment of the extracellular fluid that is contained within epithelial-lined extracellular spaces. For example, seminal fluid (Chapter 14) and saliva (Chapter 15) stains as well as vaginal secretions are analyzed for the investigation of sexual assault crimes. Sometimes, urine stains and fecal materials (Chapter 16) are related to assault crimes as well. Sweat, which is secreted from the eccrine and apocrine sweat glands in the skin (Chapter 17), and cerumen, also known as earwax (a waxy substance secreted in the ear canal), can be potentially used for human identification. Fluids present in vomitus (Chapter 17) can be potentially important for forensic investigations of violent crimes.

4.1.1 Extracellular Nucleic Acids

Blood plasma and other various bodily fluids usually contain small amounts of nucleic acids (DNA or RNA) known as extracellular nucleic acids. The nucleic acids circulating in plasma

Table 4.1 Common Items of Evidence

Evidence	Possible DNA Location	Source of DNA
Baseball bat	Handle	Skin cells, sweat, blood, tissue
Hat, bandana, mask	Inside surfaces	Sweat, hair, skin cells, dandruff, saliva
Eyeglasses	Nose, ear piece, lens	Sweat, skin cells
Facial tissue, cotton swab	Surface	Mucus, blood, sweat, semen, earwax
Dirty laundry	Surface	Blood, sweat, semen, saliva
Toothpick	Surface	Saliva
Used cigarette	Butt (filter area)	Saliva
Used stamp, envelope seal	Moistened area	Saliva
Tape or ligature	Inside or outside surface	Skin cells, sweat, saliva
Bottle, can, glass	Mouthpiece, rim, outer surface	Saliva, sweat, skin cells
Used condom	Inside surface, outside surface	Semen, vaginal cells, rectal cells
Bed linen	Surface	Sweat, hair, semen, saliva, blood
Through-and-through bullet	Outside surface	Blood, tissue
Bite mark	Skin surface	Saliva
Fingernail, partial fingernail	Scrapings	Blood, sweat, tissue, skin cells

Source: National Institute of Justice. Using DNA to solve cold cases, Special Report, 2002, US Department of Justice, Office of Justice Programs, Washington, DC.

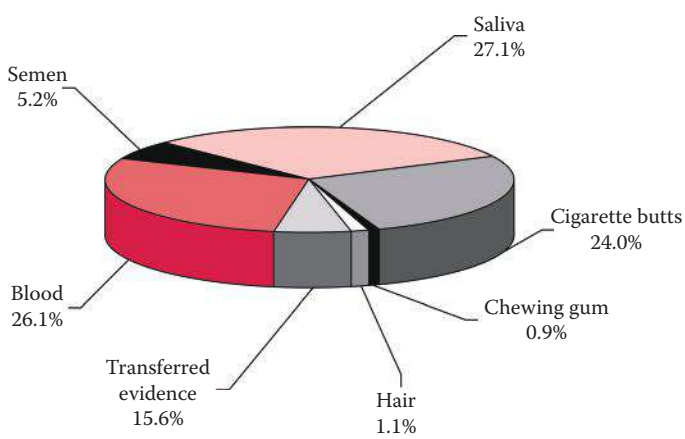

Figure 4.1 Representative types of evidence samples. Data compiled from the third quarter 2005 (July–September) results for all police forces in England and Wales. (Adapted from Bond, J.W., *J Forensic Sci*, 52, 128–136, 2007.)

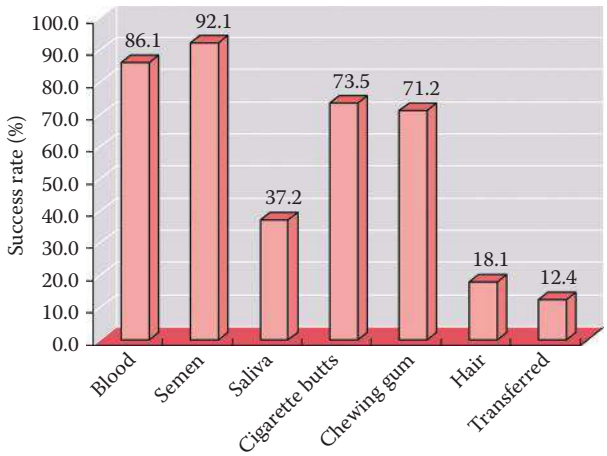

Figure 4.2 Success rate of obtaining suitable profiles from processed samples for submitting to DNA databases. Data are from the third-quarter 2005 (July–September) results for all police forces in England and Wales. (Adapted from Bond, J.W., *J Forensic Sci*, 52, 128–136, 2007.)

Figure 4.3 Biosafety cabinet for the extraction of DNA from biological evidence (left). Blood samples to be processed in a biosafety cabinet for DNA isolation (right). (© Richard C. Li.)

are referred to as *circulating* nucleic acids; the nucleic acids that are found in other body fluids, such as saliva and urine, are referred to as *cell-free* nucleic acids. Extracellular nucleic acids can remain soluble or form complexes with proteins and lipids.

The potential sources of extracellular nucleic acids are *extracellular vesicles* (EVs). EVs are endogenous vesicular structures, containing proteins and nucleic acids that are secreted by most eukaryotic cells (Figure 4.4). There are many types of EVs, including exosomes and microvesicles, which can be detected in bodily fluids. *Exosomes* are one potential source of extracellular nucleic acids. Exosomes are derived from multivesicular bodies (MVBs), which are intracellular organelles of the endocytic pathway (Figure 4.4a). MVBs fuse with the plasma membrane. As a result, the vesicles are released into the extracellular compartment as exosomes. *Microvesicles*, also called shed vesicles or ectosomes, are another possibility. Microvesicles shed from the plasma membrane and thus carry along membrane and cytosolic materials including nucleic acids (Figure 4.4b). *Apoptotic bodies* are a special type of microvesicle that is formed in apoptotic cells. During apoptosis, a process of programmed cell death, cells shrink and eventually form apoptotic bodies (Figure 4.5). Apoptotic bodies contain fragmented DNA by nucleolytic degradation that resembles similar characteristics to those observed in extracellular nucleic acids. However, messenger RNA (mRNA) within apoptotic bodies is protected from RNase degradation. These

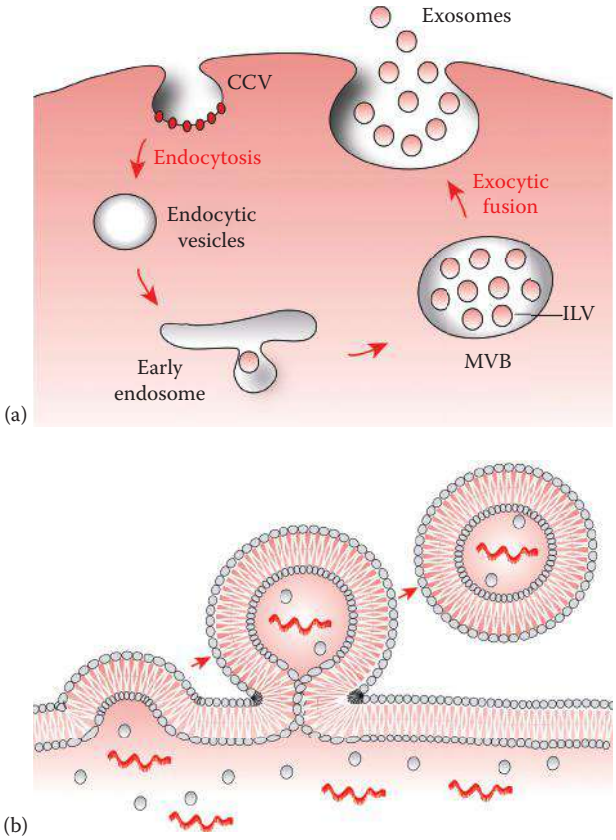

Figure 4.4 Exosomes and microvesicles released from healthy cells. (a) Exosomes are derived from the endocytic pathway. In the endocytic pathway, clathrin-coated vesicles (CCV) are formed at the plasma membrane. The endocytic vesicles are then formed through endocytosis and are transported to early endosomes. Multivesicular bodies (MVBs) containing intraluminal vesicles (ILVs) are then developed from early endosomes. Through the exocytic fusion of the MVB membrane with the plasma membrane of the cell, exosomes, containing cellular components including nucleic acids, are released to the extracellular space. (b) Microvesicles are derived from the budding of vesicles from the plasma membrane. Microvesicles are then shed from the plasma membrane carrying along cellular components including nucleic acids that are present in the cytoplasm. (© Richard C. Li.)

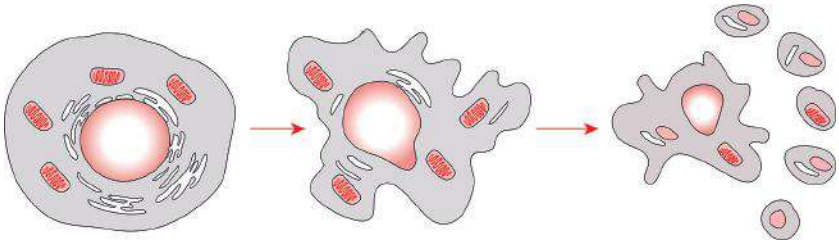

Figure 4.5 Apoptotic bodies released during apoptosis. During apoptosis, the preapoptotic cell (left) undergoes morphological changes. The early apoptotic cell (middle) shows deformation: the cell shrinks, the membrane blebs, and chromatin condense. The late apoptotic cells (right) are fragmented, releasing apoptotic bodies, which contain cytosol and organelles, and nuclear fragments. In addition, proteolytic enzymes are activated that cleave the genomic DNA into fragments. (© Richard C. Li.)

extracellular nucleic acids are potentially useful sources for the forensic identification of bodily fluids and for forensic DNA profiling.

4.2 Cells

4.2.1 Cell Surface Markers

Many of the forensic markers that are analyzed in forensic laboratories are from cells, the building blocks of the human body. All cells have membranes, also known as plasma membranes, which constitute their outer boundaries (Figure 4.6). The functions of cell membranes include exchanges with the environment, signal transduction, and structural support. The cell membrane is a phospholipid bilayer containing lipids, proteins, and carbohydrates. Membrane proteins can act as enzymes, receptors, or ion channels. Many cells also have carbohydrate-rich molecules, including proteoglycans, glycoproteins, and glycolipids, on their membrane surfaces (Figure 4.6). Many of these molecules act as cell surface antigens (Chapter 18).

4.2.2 Nucleated Cells

Probative biological evidence usually contains nucleated cells. The *nucleus* (Figure 4.7) is surrounded by a nuclear envelope and contains chromosomes and a nucleolus. A nucleolus is a dense, non-membrane-bound structure due to its high RNA content. The function of a nucleolus is to transcribe ribosomal RNA and to form ribosomes.

There are two types of cells in the body: sex cells (sperm and oocytes) and somatic cells (all other types). Spermatozoa and ova, which are formed by germ cells, are called *gametes*. In humans, each gamete is haploid, containing 22 *autosomes* (chromosomes other than sex chromosomes) plus one *sex chromosome*. In ova, the sex chromosome is always an X, while in spermatozoa it may be an X or a Y. After fertilization, the zygote, which is the fertilized egg cell formed when two gamete cells are joined, becomes diploid as a result of the fusion of the haploid spermatozoon and ovum. Most of the other cells of the body, known as *somatic cells*, are diploid. This means that they have two copies of each autosome plus two sex chromosomes, XX for females or XY for males. This results in a total of 46 chromosomes per diploid cell. The two chromosomes of a pair in a diploid cell are *homologous chromosomes*. One of the homologous chromosomes is inherited from the spermatozoon and the other from the ovum.

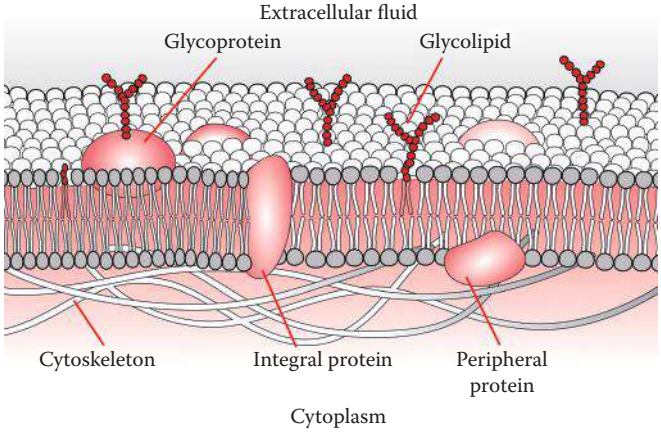

Figure 4.6 Cell membrane and carbohydrate-containing glycoproteins and glycolipids. (© Richard C. Li.)

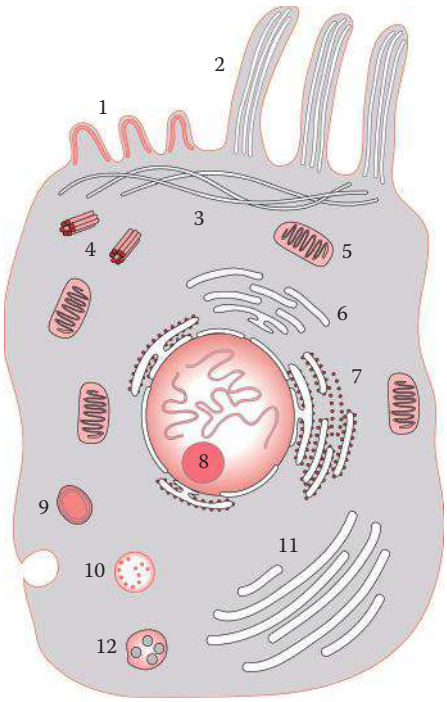

Figure 4.7 A cross-sectional view of a cell. 1, Microvilli; 2, cilia; 3, cytoskeleton; 4, centrioles; 5, mitochondrium; 6, smooth endoplasmic reticulum; 7, rough endoplasmic reticulum; 8, nucleolus; 9, peroxisome; 10, vesicle; 11, Golgi apparatus; and 12, lysosome. (© Richard C. Li.)

Although most somatic cells are diploid, exceptions exist. Some differentiated cells such as red blood cells and platelets have no nuclei and are designated *nulliploid*. A few other cells have more than two sets of chromosomes as a result of DNA replication without cell division and are referred to as *polyploid*. For example, the regenerating cells of the liver and other tissues are naturally tetraploid, while the giant megakaryocytes of the bone marrow may contain 8, 16, or even 32 copies of chromosomes.

The nuclear *chromosomes* of humans consist of complexes of DNA, histone proteins, and nonhistone chromosomal proteins. Each chromosome consists of one linear, double-stranded DNA molecule. The large amounts of DNA present in the human chromosome are compacted by their association with histones into nucleosomes and even further compacted by higher levels of folding of the nucleosomes into *chromatin* fibers. Each chromosome contains a large number of looped domains of chromatin fibers attached to a protein scaffold.

The degree of DNA packing varies throughout the cell cycle. During the metaphase of mitosis and meiosis of the cell cycle, chromatin is the most condensed. Two forms of chromatin have been defined on the basis of their chromosome-staining properties. *Euchromatin* regions are areas of chromosomes that undergo normal chromosome condensation and decondensation during the cell cycle. The intensity of staining of euchromatin is darkest in the metaphase and lightest in the synthesis (S) phase. Euchromatic regions account for most of the genome and lack repetitive DNA. Usually, the genes within the euchromatin can be expressed. *Heterochromatin* comprises the chromosomal regions that usually remain condensed throughout the cell cycle. It contains repetitive DNA and can be found at centromeres, much of the Y chromosome long arm, and the short arms of the acrocentric chromosomes (chromosomes with centromeres near one end). Genes within heterochromatic DNA are usually inactive.

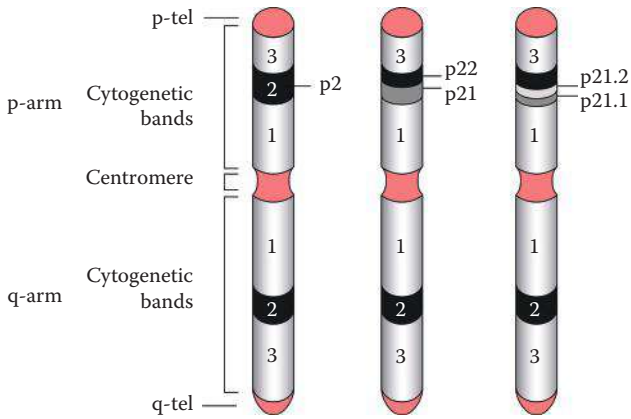

Figure 4.8 The ideogram of a human chromosome and cytogenetic banding nomenclature. The cytogenetic banding pattern is shown after chemical staining such as G-banding, which treats chromosomes with Giemsa dye. The short arm is designated p; the long arm is q. The centromere, p telomere (p-tel), and q telomere (q-tel) are also shown. G-bands at increasing resolution (from left to right) are shown. (© Richard C. Li.)

Each human chromosome has a short arm, designated p (for petit), and a long arm, designated q (for queue), separated by a centromere (Figure 4.8). *Centromeres* are the DNA sequences that are found near the points of attachment of mitotic or meiotic spindle fibers. The centromere region of each chromosome is responsible for accurately segregating the replicated chromosomes to daughter cells during cell divisions. The ends of the chromosome are called *telomeres* and they help stabilize the chromosome and play a role in the replication of DNA in the chromosome.

Chemical staining of metaphase chromosomes results in an alternating dark and light banding pattern (cytogenetic banding) that can be seen under a microscope. Each chromosome arm is divided into regions based on the cytogenic bands. This process is known as *cytogenetic mapping*. The cytogenetic bands are labeled p1, p2, p3, q1, q2, q3, and so on, counting from the centromere out toward the telomeres. At higher resolutions, subbands can be observed. For example, the cytogenetic map location of a gene termed *AMELY* (amelogenin, Y-linked; Chapter 21) is Yp11.2, which indicates its location on chromosome Y, p arm, band 11, subband 2. The visually distinct banding pattern gives each chromosome a unique appearance. Recently, the cytogenic map has been integrated with the human genome sequence to allow the determination of the positions of cytogenetic bands within the DNA sequence.

Chromosomes can be identified on the basis of the size and the positions of the centromeres and cytogenetic banding patterns. The chromosome constitution is described as a *karyotype* and can be displayed as a *karyogram*, which includes the total number of chromosomes and the sex chromosome composition (Figure 4.9). Chromosomes are numbered in order of their size, with chromosome 1 as the largest (except chromosome 21 is smaller than 22). In the cases of chromosomal abnormality, the karyotype can also reflect the type of abnormality and allow visualization of the affected chromosome bands.

4.2.3 Mitochondria and Other Organelles

The cytoplasm contains the cytosol fluid in which organelles are suspended (Figure 4.7). Multiple copies of *mitochondria* are located within the cytoplasm. Mitochondria are surrounded by phospholipid membranes that separate them from the cytosol. The mitochondria are responsible for energy production through aerobic metabolism by producing molecules containing high-energy bonds, such as adenosine triphosphate (ATP). Mitochondria have their own genome, which can be analyzed for human identification (Chapter 23).

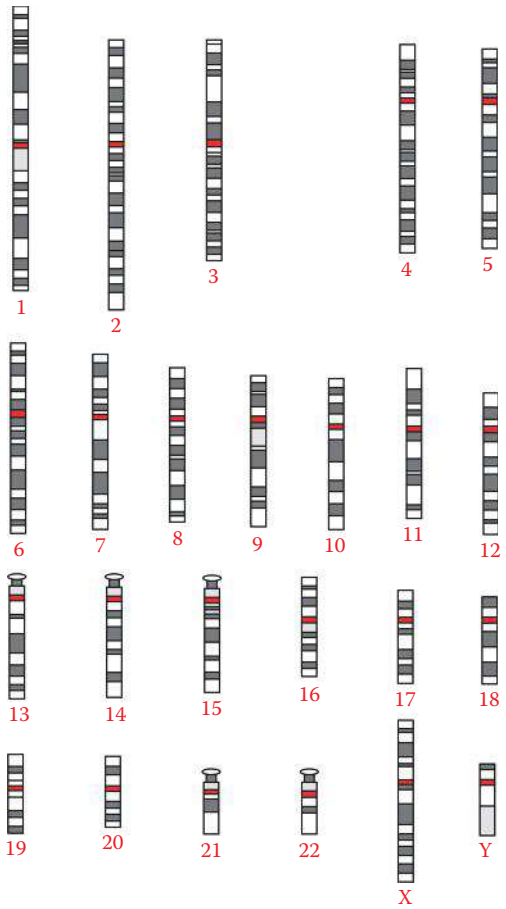

Figure 4.9 Human karyogram. The chromosomes are numbered. Cytogenic patterns show alternating dark and light bands (From www.genome.gov). Centromeres (in red) are shown.

Two types of *endoplasmic reticulum* (ER) can exist within the cytoplasm: the smooth ER (SER) is involved in lipid synthesis; the rough ER (RER) contains ribosomes on its outer surface and forms transport vesicles. The *Golgi apparatus* is responsible for the production of secretory vesicles and new membrane components, and also for the packaging of *lysosomes* (vesicles containing digestive enzymes for the degradation of injured cells). *Peroxisomes* carry enzymes that neutralize potentially harmful free radicals. Other organelles found within the cytoplasm of eukaryotic cells include the cytoskeleton, microvilli, centrioles, and cilia.

4.2.4 Cytosol

4.2.4.1 Messenger RNAs

The chromosomal DNA contains genes that encode for specific proteins. The genetic code is read as an array of triplet codes, a sequence of three bases that specifies the identity of a single amino acid. As gene expression is activated, transcription occurs in which precursor mRNA (pre-mRNA) is produced from a DNA template. After transcription, the pre-mRNA is capped, polyadenylated, and spliced to form matured mRNA. Only the matured mRNA is transported from the nucleus to the cytoplasm. Tissue-specific mRNA can be potentially used for the identification of biological evidence (Chapter 11). The proteins are synthesized in a process known as translation, in which amino acids are assembled based on the codons derived from the triplet

code of the DNA contained in the sequence of the mRNA strand. Various components including the ribosomal complex are involved in translation. The cytosol also contains many proteins that can be used for the identification of bodily fluids (Chapters 11 and 12).

4.2.4.2 MicroRNAs

MicroRNAs (miRNAs) are short RNA molecules that are 21–23 nucleotides in length. In eukaryotic organisms, miRNAs function as negative regulators of gene expression. They play roles in development and cell differentiation. Additionally, the altered expression of miRNAs can be detected in many human diseases. The biological function and the potential application of miRNAs in forensic biology are discussed in Chapter 11.

It is estimated that the human genome may encode approximately 1000 miRNAs. Based on their genomic location and structure, miRNAs can be characterized into three types: intergenic, intronic, and exonic (Figure 4.10). *Intergenic* miRNA genes are distinct transcriptional units that are found in genomic regions. In humans, most of the intergenic miRNA genes have a transcription start site and a polyadenylation site. *Intronic* miRNAs reside within the introns of protein-coding and noncoding genes. The orientation of the intronic miRNAs can be the same as that of the sense (coding) strand of a host gene or that of the antisense strand (complementary to the sense strand) of a host gene. Sense intronic miRNAs are transcribed from the same promoter as their host genes. The antisense intronic miRNAs are transcribed from their own promoters. *Exonic* miRNAs are rare in eukaryotic genomes and reside in genomic regions that overlap with an exon and an intron of a pseudogene, which is a noncoding gene or is no longer transcribed. These miRNAs are also transcribed from their host gene promoter. miRNAs can also be monocistronic or polycistronic. A *monocistronic* miRNA has a single transcriptional unit with its own promoter. In *polycistronic* miRNAs, several miRNAs reside as a cluster of transcriptional units with a shared promoter.

The biogenesis of miRNAs begins in the nucleus. miRNAs are transcribed by RNA polymerase II. Nascent transcripts, referred to as primary transcripts (pri-miRNAs), can be several hundreds to thousands of nucleotides in length. The pri-miRNAs have a hairpin secondary

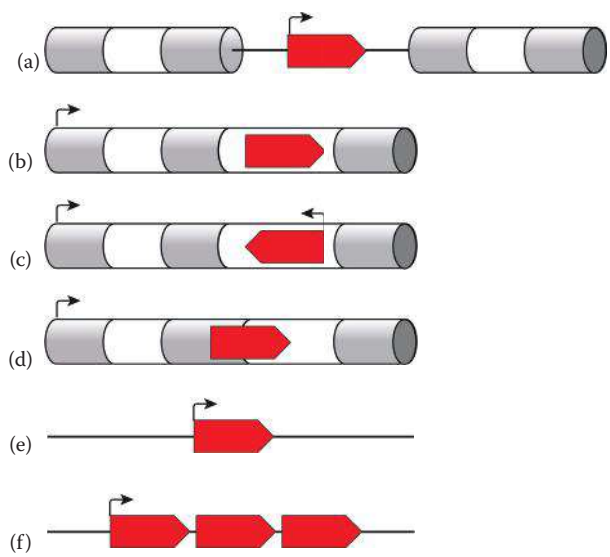

Figure 4.10 The genomic location and the structure of miRNAs. (a) An intergenic miRNA gene, (b) a sense intronic miRNA gene, (c) an antisense intronic miRNA gene, (d) an exonic miRNA gene, (e) a monocistronic miRNA gene, and (f) polycistronic miRNA genes. Exons (gray), introns (white), and miRNA genes (red) are shown. (© Richard C. Li.)

structure that is approximately 70 nucleotides long with imperfect base pairing in the stem. miRNA processing is initiated in the nucleus (Figure 4.11). The hairpin region of a pri-miRNA is then cleaved from the pri-miRNA by the nuclear RNase III endonuclease, Drosha. As a result, a pre-miRNA is formed. The pre-miRNA is transported from the nucleus to the cytoplasm by a nuclear transporter, Exportin 5. In the cytoplasm, a cytoplasmic RNAse III–like endonuclease, Dicer, cleaves the pre-miRNA to generate a double-stranded miRNA that is approximately 21–23 nucleotides in length. The mature miRNA strand is bound to the Argonaut protein to assemble the RNA-induced silencing complex (RISC). The complementary RNA strand is degraded.

A specific miRNA is designated with a prefix. The prefix for a mature miRNA is a capitalized "miR," while the prefix for a pre-miRNA is an uncapitalized "mir." The prefix "miR" is followed by a dash and a number (e.g., miR-135). Experimentally confirmed miRNAs are sequentially numbered and are deposited in the miRBase, which is a database that archives miRNA sequences and annotations.

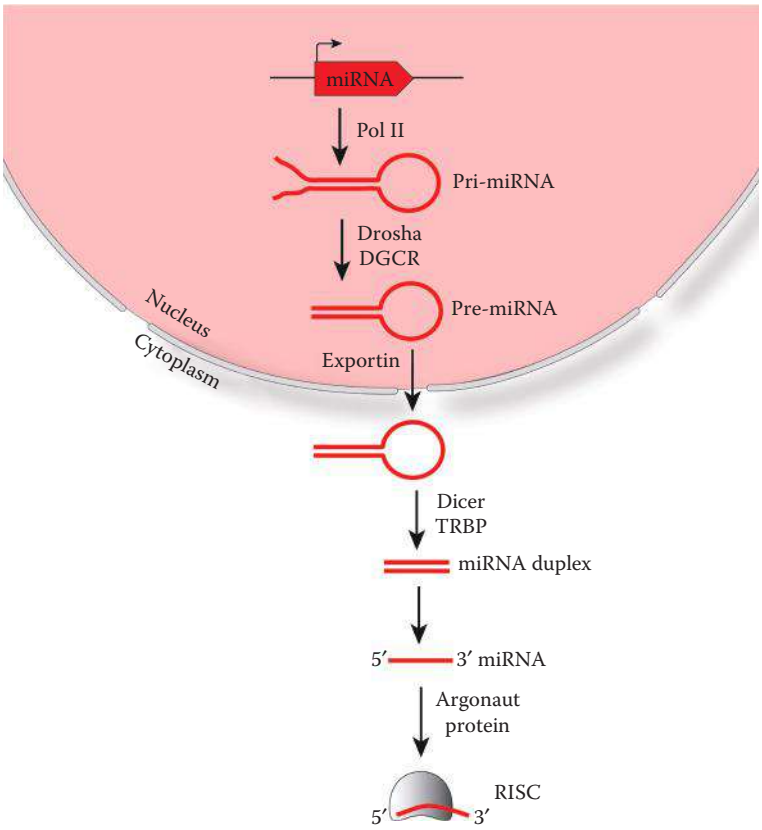

Figure 4.11 The biogenesis of miRNAs. In the nucleus, the pri-miRNA (primary transcript) is processed by Drosha (nuclear RNase for double-stranded RNA) and DGCR (nuclear double-stranded RNA-binding protein), forming the pre-miRNA. The pre-miRNA is transported from the nucleus to the cytoplasm by the exportin. In the cytoplasm, the pre-miRNA is further processed by Dicer (cytoplasmic RNase for double-stranded RNA) and TRBP (cytoplasmic double-stranded RNA-binding protein). The mature miRNA strand is bound to the Argonaut protein to assemble the RISC (RNA-induced silencing complex). (© Richard C. Li.)

4.3 Tissues

4.3.1 Skin

4.3.1.1 Biology of Skin

The biological evidence related to skin is important in forensic investigations. For example, fingerprints are ridge skin impressions that are usually collected at crime scenes, and shed skin tissue is a source of DNA for human identification. The skin covers the entire body surface. Additionally, the skin contains specialized structures that include sebaceous and sweat glands, hair follicles, and nails. The thickness of skin varies throughout the body. The skin of the dorsal area of the body is usually thicker than that of the ventral area of the body. The skin consists of different layers (Figure 4.12). The *epidermis* is the outer layer of the skin. The epidermis also contains melanocytes that produce the skin pigment melanin. The *dermis* is the middle layer of the skin. It is filled with fibrous collagen proteins secreted by fibroblasts and contains hair follicles and sweat glands. Additionally, it contains blood, lymph vessels, and nerves. The *subcutaneous layer* is the deepest layer of the skin. The subcutaneous layer consists of collagen networks and adipose tissue to prevent loss of heat.

The epidermis is a multilayered tissue that includes a number of morphologically distinct zones: the basal, the spinous, the granular, and the cornified layers (Figure 4.12). The epidermis renews continually through the proliferation and differentiation of keratinocytes. The *basal layer* contains newly formed keratinocytes that are proliferative. Epidermal differentiation begins with the migration of the keratinocytes from the basal layer toward the outer layer of skin. Once the migrating keratinocytes reach the *spinous* and *granular* layers, the keratinocytes become nonproliferating and partially differentiated. As the cells reach the *cornified* layer, these cells are filled with keratin filaments and are differentiated into corneocytes, which are dead, and terminally differentiated keratinocytes. In the course of differentiation, the cells are flattened, and all organelles including the nucleus are lost. The corneocytes are then shed from the skin surface.

4.3.1.2 Skin as Source of DNA Evidence

Evidence from skin contact, also referred to as *touched evidence*, can be collected and used for forensic DNA analysis. One example of this evidence is shed skin cells that are found on worn clothing, which is frequently encountered in crime scene investigations. For instance, a perpetrator's shed skin cells that are deposited on worn clothing are potential evidence to be collected. After collection, a DNA profile can be obtained from shed skin cells, providing the forensic evidence or "lead" that is required for the criminal investigation. Touched evidence becomes more

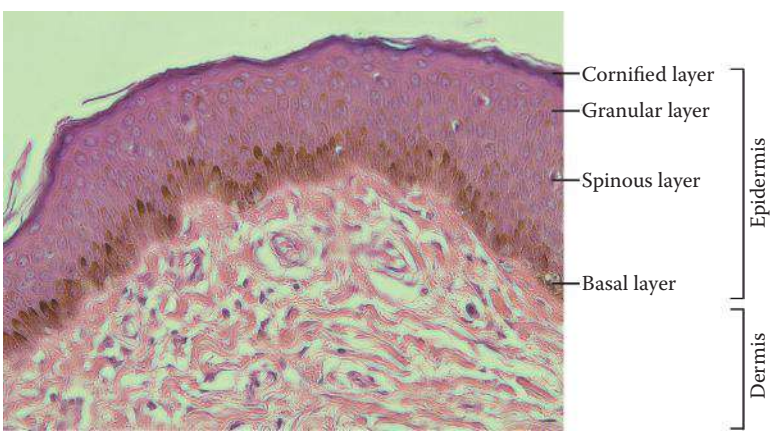

Figure 4.12 A sectional view of the skin. Multilayered epidermis tissues are shown. (© Richard C. Li.)

useful and important when no other type of evidence (e.g., a fingerprint, bloodstain, or seminal stain) can be collected at the crime scene, especially for the investigation of certain criminal cases such as cold cases and property crimes such as theft and burglary (Figures 4.13 through 4.15). Touched evidence usually contains minute quantities of nuclear DNA. DNA recovered from touched evidence is referred to as *transfer DNA*. However, the source of DNA transferred through physical contact is not well understood. Although the shed skin cells from touched evidence lack nuclei, these cells may contain DNA remnants from partial degradation during the differentiation process. These can be a possible DNA source. Additionally, the sweat glands in the skin produce sweat. When the skin makes contact with an item, a residue of sweat is left on the surface of the item. It is known that sweat contains cell-free DNA. Thus, cell-free DNA from sweat is another possible source of DNA. Furthermore, a small number of nucleated cells can be observed in touched evidence. These cells possibly originated from sweat glands and ducts.

4.3.2 Hair

Hairs, including scalp and pubic hairs, frequently constitute biological evidence that is found at crime scenes, and their identification can be of great forensic importance. Formerly, the

Figure 4.13 Evidence collection using swabbing. The evidence from skin contact for forensic DNA analysis is usually collected by swabbing. More DNA can be recovered when evidence is collected with a double swab method than with a single swab. The double swab method involves applying a moistened cotton swab followed by a second dry cotton swab onto the same target surface of evidence. The target surface is swabbed using a moistened swab first. The moisture left by the first swab is absorbed by the second dry swab. Both swabs can be pooled for DNA extraction. (© Richard C. Li.)

Figure 4.14 A worn glove as a source of DNA evidence. Shed skin cells on a worn glove can potentially generate DNA profiles. (© Richard C. Li.)

Figure 4.15 A cigarette butt as a source of DNA evidence. Shed cells left behind on a cigarette butt can be a source of DNA. A portion (~1 cm^2) of the filter paper of a smoked cigarette butt can be cut for isolating DNA. (© Richard C. Li.)

principal methods that were utilized in forensic hair analysis were limited to morphological analysis and comparisons. Since then, protein polymorphisms provide some potential for identifying individuals from single hairs. However, human hairs contain DNA and as a DNA source they may be used for forensic analysis. The development of the polymerase chain reaction (PCR) amplification technique made it possible to analyze very small quantities of DNA in hair, and the use of hair as evidence of identification has become more significant.

4.3.2.1 Biology of Hair

The human hair shaft is a keratinized cylindrical structure (Figure 4.16). The center or core of the hair is called the *medulla*, which is present in the majority of hairs. The medulla is surrounded by a *cortex*, which is the outer layer of the hair shaft. The *cuticle* consists of overlapping layers of flattened keratinized cells that protect the hair. Hairs are produced in *hair follicles* (Figures 4.17 and 4.18). Each hair follicle is located deep in the dermis (a skin layer beneath the epidermis) and opens onto the surface of the epidermis (the outer layer of the skin). The hair follicle is composed

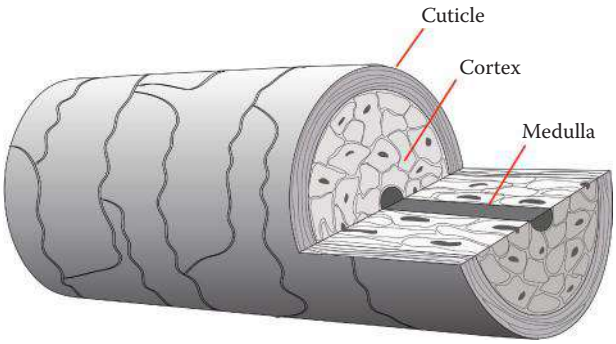

Figure 4.16 A sectional view of a hair shaft. (© Richard C. Li.)

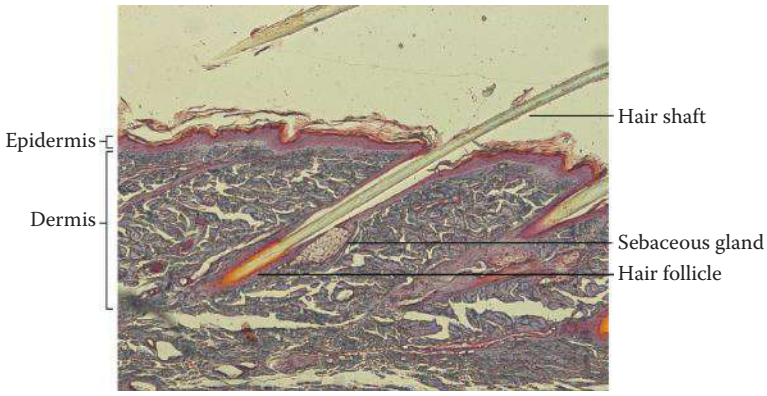

Figure 4.17 Longitudinal section view of a scalp hair follicle with accessory structures. (© Richard C. Li.)

of the *bulge* and the *bulb* regions. The bulge exports the stem cells that migrate down and give rise to bulb cells. The bulge also produces the stem cells that migrate up to form skin cells. The matrix cells, which generate the hair shaft cells, are located at the lower portion of the bulb. The *dermal papilla* is situated at the base of the bulb and contains cells that regulate hair growth, which are nourished by blood vessels and nerves. The growing hair shaft is surrounded by two concentric layers of cells, which are referred to as the inner root sheath and the outer root sheath, respectively. The entire hair follicle is surrounded by a connective tissue sheath.

Human scalp hairs grow for a few years and shed according to the hair follicle cycle (Figures 4.19 through 4.21). A human scalp hair can grow at its highest rate of approximately 1 mm per day. The growing phase of hair is called the *anagen* phase. The matrix cells undergo rapid proliferation and eventually become differentiated cells such as hair shaft cells. As a hair grows, it is pushed toward the surface of the skin and becomes longer. During the migration upward, keratinization occurs as cells are filled with fibril keratin proteins. In these keratinizing cells, nuclei are absent. However, mitochondrial remnants can be observed. By the time a hair approaches the skin surface, cell death occurs at the medulla, cortex, and cuticle. At the end of the anagen phase, the matrix cells enter the *catagen* phase and undergo cell death, thus leading to the regression of the bulb. Hair follicles then enter the *telogen* phase: the stage of rest. When another cycle begins, the follicle produces a new bulb and the telogenic hair, also known as the club hair, is pushed to the surface and shed. On average, an adult loses approximately dozens of hairs daily.

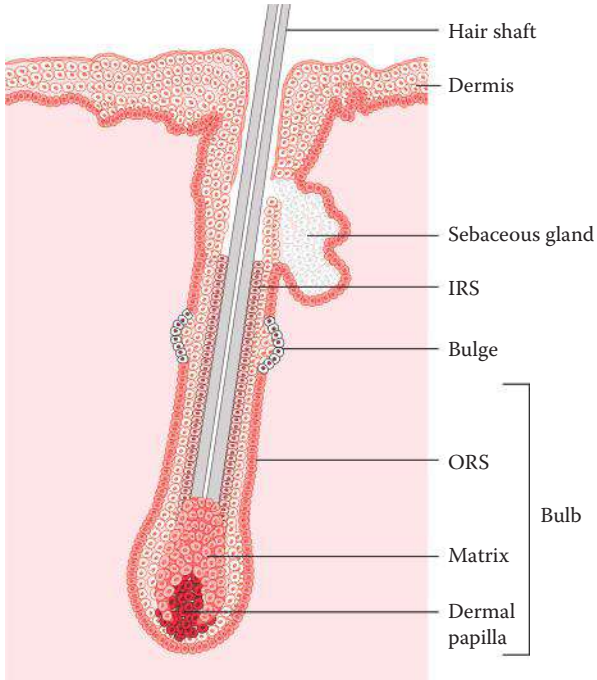

Figure 4.18 Structure of an anagen-phase hair follicle. The diagram shows the hair follicle structure including the dermal papilla, sebaceous gland, bulge, bulb, and hair shaft. Concentric layers of the outer root sheath (ORS), the inner root sheath (IRS), and the hair shaft are also shown. (© Richard C. Li)

4.3.2.2 Hair as Source of DNA Evidence

The isolation of DNA from intact hair roots is routinely used in nuclear DNA analysis. In comparison, the quantities of DNA in telogen hair roots are considerably less than the DNA that is found in the roots of anagen hairs. Nuclear DNA analysis is usually accomplished by using freshly plucked hair roots (Figures 4.22 and 4.23) because cells at the root region may contain nuclear DNA. Unfortunately, most human hairs recovered from crime scenes are shed naturally (in the telogen phase) and contain little nuclear DNA (Figure 4.24). Thus, multiple telogen hairs with roots are necessary to isolate enough nuclear DNA. However, shed hairs that are found at crime scenes may be derived from different individuals. Therefore, the ability to perform a forensic DNA analysis of a single shed hair would be highly desirable. Nuclear DNA isolation from hair shafts is still far less reliable because hair shafts contain very low amounts of nuclear DNA. In addition, variations in the amounts of DNA isolated from hair shafts are observed in a comparison between different hairs from the same head and hairs from different individuals.

A hair follicle cell contains multiple copies of mitochondria. As a result, mitochondrial DNA (mtDNA) can be successfully isolated from hair roots. Additionally, mtDNA is embedded in the keratin matrix of hair shaft cells, which protects the mtDNA molecules from degradation. Thus, mtDNA can also be isolated from hair shafts. A sequence polymorphism analysis of mtDNA from hair can be carried out. mtDNA is maternally inherited, which is useful to identify maternal relatives but cannot be used to perform paternity testing. Additionally, the mtDNA profiling results are not as discriminating as nuclear DNA profiling. Furthermore, mtDNA analysis is time-consuming. Therefore, the typing of nuclear DNA from hair would be preferable for forensic DNA analysis.

Sometimes, a mixture of more than one mtDNA sequence in the same individual is observed. This heterogeneous pool of mtDNA molecules is referred to as *heteroplasmy*. In hair, it is

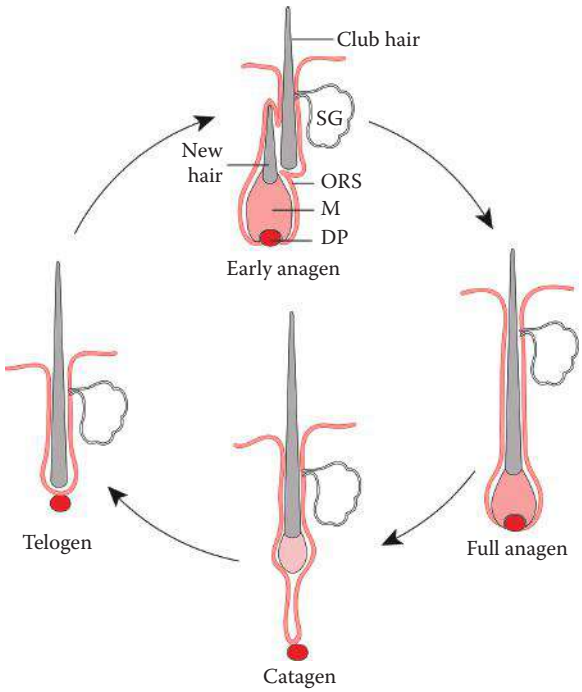

Figure 4.19 Diagrammatic representation of the scalp hair cycle. The morphological characteristics of a hair follicle are shown in three distinct phases encompassing the entire hair cycle: anagen (including early and full anagen), catagen, and telogen. DP, dermal papilla; M, matrix; ORS, outer root sheath; SG, sebaceous gland. (© Richard C. Li.)

believed that it is due to a mixture of mtDNA molecules from keratinocyte and melanocyte-derived mitochondria (Figure 4.25). Hair follicle melanocytes are formed at the beginning of each hair cycle and die at the end of the cycle. Melanocytes are located in the bulb hair follicle. Hair melanocytes play roles in hair pigmentation, which determine hair color. Melanin, which is produced by the melanocytes, is contained in an organelle called the *melanosome*. The melanosomes are transferred to neighboring keratinocytes though *dendritic processes*. In addition to melanosomes, the melanocyte mitochondria can also be transferred to keratinocytes. Thus, the keratinized cells in the hair shaft may carry more than one type of mitochondria, one from the keratinocytes and the other from the melanocytes. As a result, heteroplasmy of a mixture of different mtDNA molecules, with different DNA sequences, can occur in hairs.

4.3.3 Bone

4.3.3.1 Biology of Bone

The bodies of human remains begin to decompose shortly after death. The rate of decomposition of human remains varies greatly with environmental conditions such as climate, bacterial growth, and the presence of insects and other animal scavengers. However, soft tissues may be lost first while more stable bone tissues may remain. Identifying human skeletal remains can be applied in a variety of cases including mass fatality incidents, missing persons, fires, explosions, and violent crime cases involving skeletal remains.

An adult human skeleton consists of 206 bones (Figures 4.26 and 4.27). The shaft of a long bone, such as an arm or a leg bone, consists largely of an outer layer of *cortical* (or compact) bone, which is solid and strong. The shaft of a long bone forms a *marrow cavity*, which is filled with a specialized type of connective tissue called bone marrow. The portion at each end of a

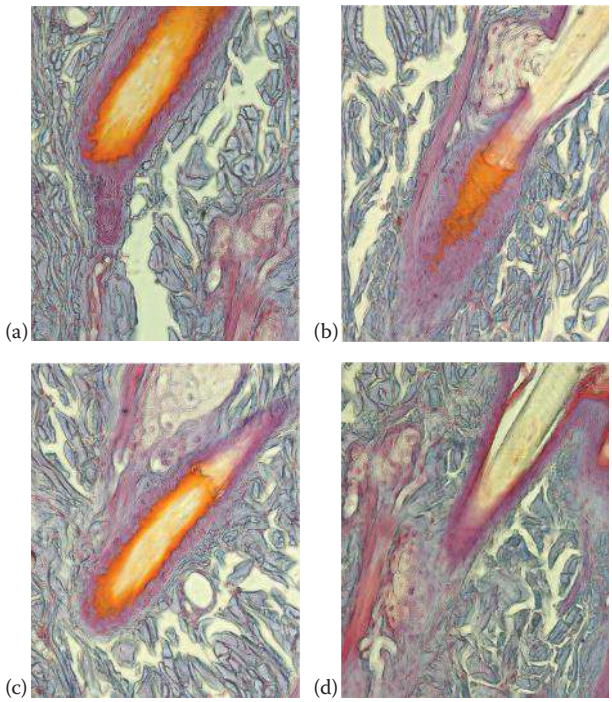

Figure 4.20 Longitudinal section view of a scalp hair follicle during the hair cycle. (a) Early anagen, (b) full anagen, (c) catagen, and (d) telogen. (© Richard C. Li.)

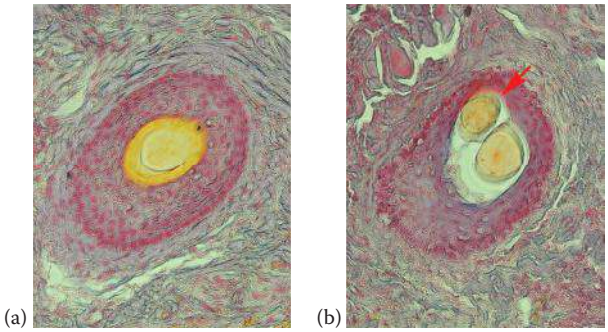

Figure 4.21 Cross-sectional view of hair follicles. (a) Early anagen and (b) full anagen. Note that the new hair (arrow) and the club hair are shown. (© Richard C. Li.)

long bone is called the *epiphysis*, which is composed largely of *cancellous* (or spongy) bone, and can bear the force of compression. A flat bone can have primarily either cortical or cancellous bone. For instance, a rib consists of primarily cancellous bone surrounded by a thin layer of cortical bone. A skull bone usually consists of largely cortical bone.

Bone, which is a connective tissue, contains a matrix and cells. The bone matrix consists of an inorganic and an organic matrix. Calcium and phosphorus are the major components of the inorganic matrix, which consists mainly of hydroxyapatite crystals, $Ca_{10}(PO_4)_6(OH)_2$. The organic matrix consists of collagens, primarily type I collagen, which are insoluble fibrous proteins. With the deposition of calcium hydroxyapatite crystals around the collagen fibrils, bone becomes a weight-bearing hard tissue.

Figure 4.22 Hair root of a pulled hair with visible soft tissue attached. (© Richard C. Li.)

Figure 4.23 Pulled dreadlocks recovered from a crime scene. (© Richard C. Li.)

Developing bones contain small numbers of *osteoprogenitor* cells. These cells can divide to produce cells that differentiate into *osteoblasts*. Osteoblast cells regulate the calcification of the bone matrix. Osteoblasts that are embedded in the bone matrix are termed *osteocytes* and are the most abundant cells in bone. Osteocytes play a role in maintaining the surrounding matrix and repairing damaged bone. Another type of cell that can be found in bone tissues is *osteoclasts*. These cells are giant cells containing 50 or more nuclei and are responsible for dissolving and recycling the bone matrix.

4.3.3.2 Bone as Source of DNA Evidence

A number of methods are used to identify human remains, for example: the identification of facial characteristics; the recognition of individualizing scars, marks, and other special body

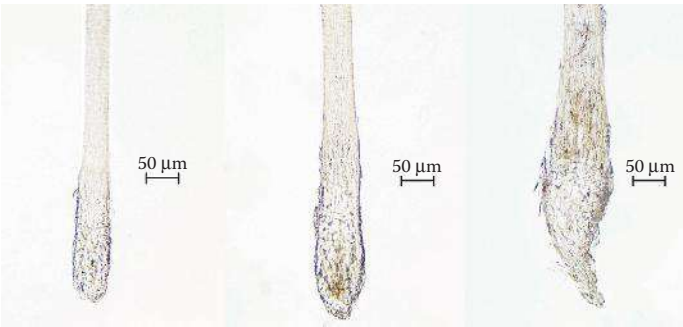

Figure 4.24 Telogen hair roots. From left to right: telogen club root with no visible soft tissue; telogen hair root with some visible soft tissue; telogen hair root with visible soft tissue. (From Bourguignon, L. et al., *Forensic Sci Int Genet*, 3, 27–31, 2008. With permission.)

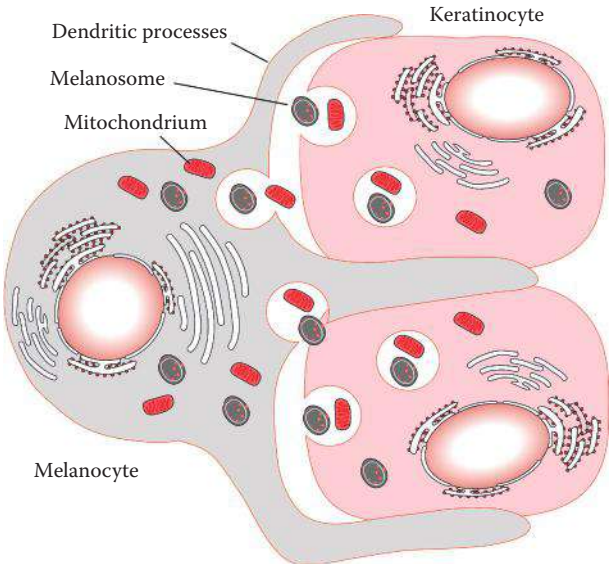

Figure 4.25 Scheme diagram of the melanosome transportation. Melanosomes are released into the extracellular space from the melanocyte dendrites through exocytosis and subsequent endocytosis by keratinocytes. During the process, mitochondria originating from the melanocytes can be potentially transported to the keratinocytes. (© Richard C. Li.)

features; the matching of dentition with premortem dental x-rays; and the comparison of fingerprints. In some situations, these methods cannot be used because of the extensive decomposition of the remains. The mass fatality terrorist attack on the World Trade Center in 2001 (Figure 4.28) serves as an example of a situation where common identification techniques may not be useful. Large quantities of compromised human skeletal fragments were recovered at the fatality site. In these cases, DNA typing is a powerful tool for identifying human remains.

Most DNA in cortical bone is located in the osteocytes. It has been estimated that there are approximately 20,000 osteocytes per cubic millimeter of calcified bone matrix. As a result, microgram quantities of DNA can potentially be obtained from a gram of bone. Thus, compact bone tissue should contain sufficient amounts of nuclear DNA for analysis. However, the skeletal fragments recovered from burial sites are often subjected to decomposition (Figures 4.29 and 4.30). During the decomposition process, both nuclear and mtDNA can be degraded.

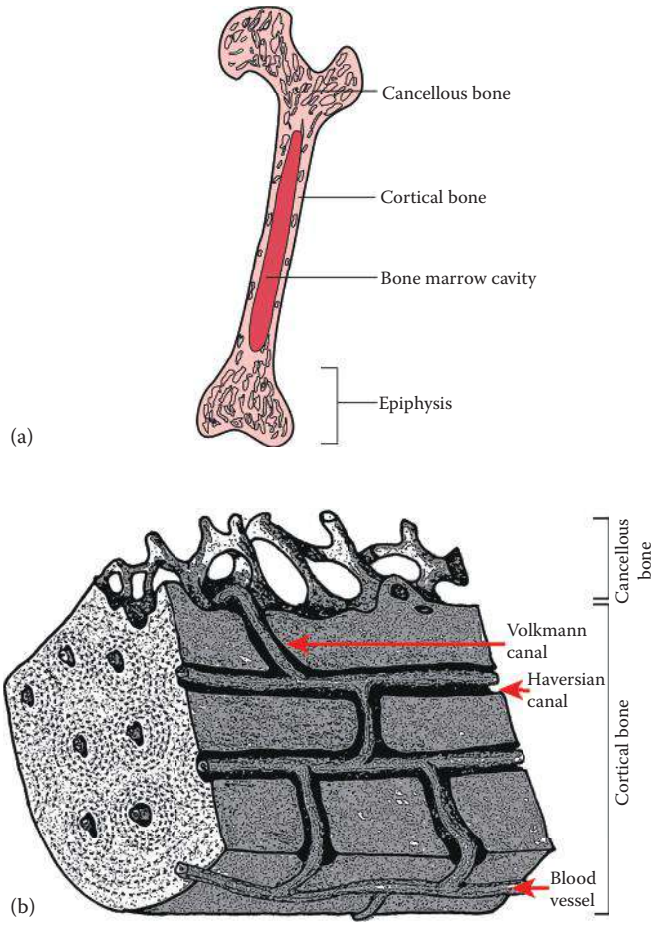

Figure 4.26 Bone structure. An adult human consists of 206 bones. A long bone, such as an arm or a leg bone, consists of an outer cylinder of cortical bone surrounding a marrow cavity. Each end of a long bone is called the epiphysis, which is composed largely of cancellous bone. Flat bones have variable structures; for example, the skull consists mainly of cortical bone whereas the spine consists mainly of cancellous bone. (a) Diagram of a long bone. (b) A bone fragment. Cortical and cancellous bones are shown. Blood vessels can be found in the Volkmann and Haversian canals. (© Richard C. Li.)

Additionally, burial conditions with high humidity and temperature promote the degradation of DNA. Thus, the identification of partial DNA profiles or the complete failure to obtain DNA profiles can occur after samples from decomposed remains are analyzed.

Processing bone samples for DNA extraction is a time-consuming task (Figure 4.31). Due to the potential for commingled remains and contaminants that interfere with forensic analysis, a bone sample initially must be cleaned prior to isolating DNA. The outer surfaces of bone fragments are usually removed by using a mechanical method such as sanding. However, to avoid cross-contamination of samples, the bone dust that is generated by sanding must be removed. Additionally, special protective equipment and safety procedures are necessary to protect analysts from exposure to blood-borne pathogens.

To obtain adequate quality and quantity of DNA from a bone sample, a high-yield DNA extraction method should be selected. The bone samples can be ground to powder to aid in DNA extraction (Chapter 5). The osteocytes containing DNA are embedded in a calcified bone matrix, which is a barrier for extracting DNA from the osteocytes. The bone matrix must be

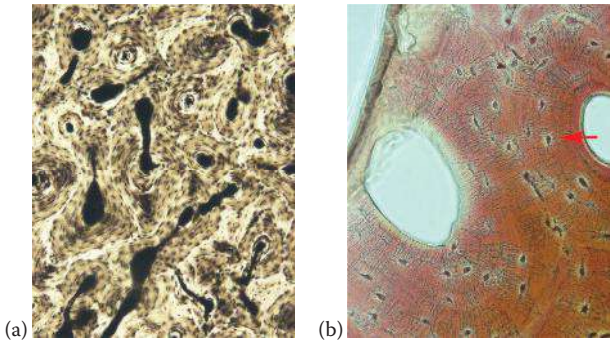

Figure 4.27 Cross-sectional view of cortical bone. (a) The functional unit of cortical bones is a cylindrical structure known as osteon. Haversian canals are shown in the center of the osteon. (b) Detailed view of an osteon. Osteocytes (arrows) are shown within osteons. (© Richard C. Li.)

Figure 4.28 Sections of the Fire Department of New York City (FDNY) Memorial Wall. Memorial to the Fallen Firefighters of 9/11, at FDNY Engine Co. 10 on Liberty Street, New York, by Rambusch Studios. (© Richard C. Li.)

Figure 4.29 Human rib bone fragments recovered from burial site. (© Richard C. Li.)

Figure 4.30 Skeletal remains, exhumed from Frombork Cathedral in Poland, are thought to be those of astronomer Nicolaus Copernicus (1473–1543). (From Bogdanowicz, W., et al., *Proc Natl Acad Sci U S A*, 106, 12279–12282, 2009. With permission.)

removed to improve the yield of DNA. A decalcification method can be utilized to dissolve calcium ions to soften the bone tissue. Additionally, the application of proteinase can be used to digest the matrix proteins, thus increasing the yield of DNA that is harvested from osteocytes.

4.3.4 Teeth

4.3.4.1 Biology of Teeth

During embryonic development, two sets of teeth begin to form. The first to appear are the *deciduous teeth* or *primary teeth*. Most children have 20 deciduous teeth, which are later replaced by 32 teeth known as the *secondary dentition* or *permanent dentition* (Figure 4.32).

Figure 4.31 Tools for cutting bone samples. Osteotomes and a mallet (top), and a rotary device (bottom). (© Richard C. Li.)

The bulk of each tooth consists of a calcified connective tissue called *dentin* (Figure 4.33). The dentin of the crown is covered by a layer of *enamel*. The rest of the dentin is referred to as radicular dentin and is covered by a layer of *cementum*, which separates the tooth from the surrounding jawbone.

Similar to bone, tooth tissue contains a matrix. The inorganic matrix of tooth contains hydroxyapatite, a calcium phosphate in a crystalline form. The organic matrices of dentin and cementum are primarily collagens. In enamel, amelogenin is the major protein of the organic matrix. Other proteins include ameloblastin and enamelin, which are also the components of the enamel organic matrix.

The interior chamber within the tooth surrounded by dentin is known as the pulp cavity. The *dental pulp*, found within the pulp cavity, is the connective tissue made up of nerve fibers, blood vessels, and various cells. The blood vessels and nerves in the pulp cavity are innervated through the root canal, a narrow tunnel located at the root of the tooth. Incisor and cuspid teeth have single roots. Bicuspids have one or two roots. Molars typically have three or more roots.

The columnar cell bodies of *odontoblasts* are located along the peripheral dental pulp. A single *odontoblast process*, arising from each cell body of the odontoblasts, projects into the dentinal tubule (Figure 4.34). Odontoblasts play important roles in the formation of dentin. Odontoblasts secrete collagens and ground substances that are the components of the

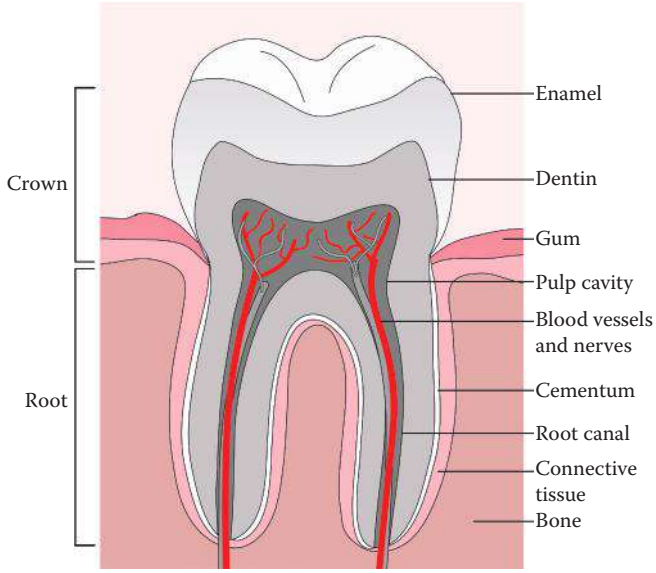

Figure 4.32 A sectional view of an adult tooth. (© Richard C. Li.)

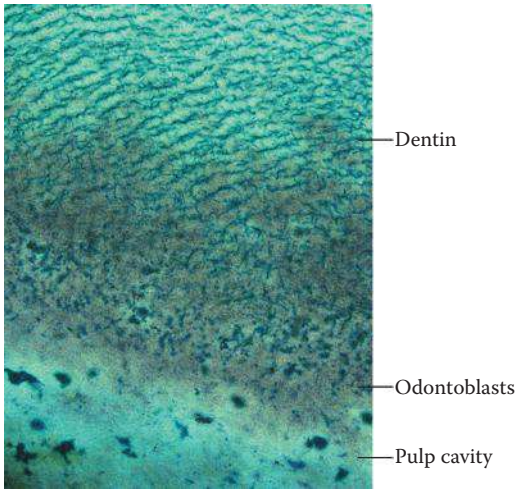

Figure 4.33 Cross-sectional view of a tooth. (© Richard C. Li.)

dentinal matrix. Additionally, odontoblasts regulate the calcification of the matrix of dentin. *Cementoblasts* are cells that play roles in forming the cementum. Cementoblasts secrete collagens and ground substances to form the extracellular matrix of the cementum. Through the process of forming cementum, cementoblasts become trapped in the extracellular matrix. The cementoblasts embedded in the cementum are referred to as *cementocytes*. *Ameloblasts* are cells that play a role in producing enamel and are subsequently lost during tooth eruption.

4.3.4.2 Teeth as Source of DNA Evidence

The characteristics of teeth, their alignment, and the overall structure of the mouth provide information for identifying a person. The use of dental records such as x-rays and dental casts can allow dental remains to be connected to a victim. Particularly in circumstances such as

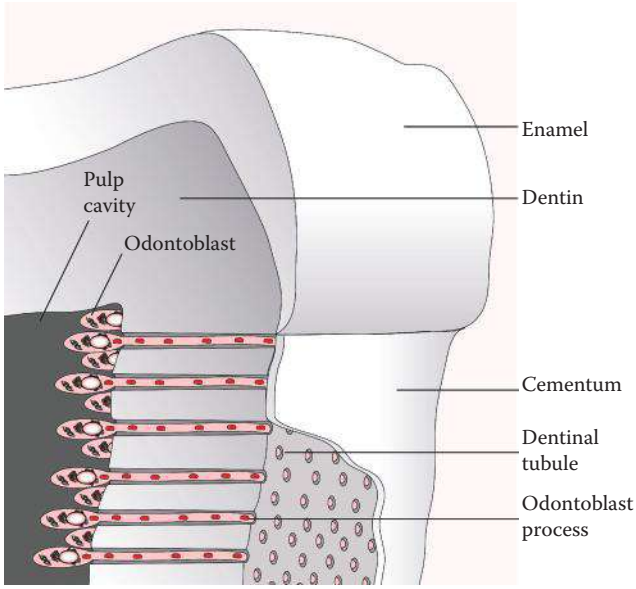

Figure 4.34 Odontoblast processes. See Section 4.3.4.1. (© Richard C. Li.)

Figure 4.35 Tools for dissecting teeth. (© Richard C. Li.) (a) A dental chisel and a mallet. (© Richard C. Li.) (b) An amalgam well. (Courtesy of Dr. Ken Hermsen.)

decomposition, an odontological comparison is possible since the dental evidence often remains intact. When no antemortem dental record is available for comparison, forensic DNA testing can be carried out for postmortem human identification. The mineralized dental structure protects DNA from degradation in cases where it may be degraded in other tissues of the body. Thus, teeth are an excellent source of DNA for forensic DNA analysis under such conditions.

Dental pulp tissue contains various cells and is a suitable source of DNA. However, when tooth evidence has been exposed to high temperature and humid environments, decomposition of the pulp tissue can occur. Thus, cementoblasts within the cementum, containing both nuclei and mitochondria, can then be utilized as a source of DNA. Additionally, odontoblast processes within dentin, containing mitochondria, can also be used. Several different methods are used to obtain dental tissues for DNA isolation (Figure 4.35). A vertical section is cut along the longitudinal axis of the tooth, which allows the dissection of the pulp, dentin, and cementum tissues

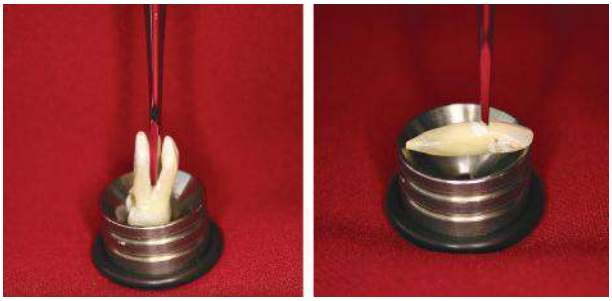

Figure 4.36 Dissecting teeth for DNA isolation. Vertical cutting and dissecting of the pulp, dentin, and cementum tissues (left) and horizontal cutting and dissecting of the root portion (right). (Courtesy of Dr. Ken Hermsen.)

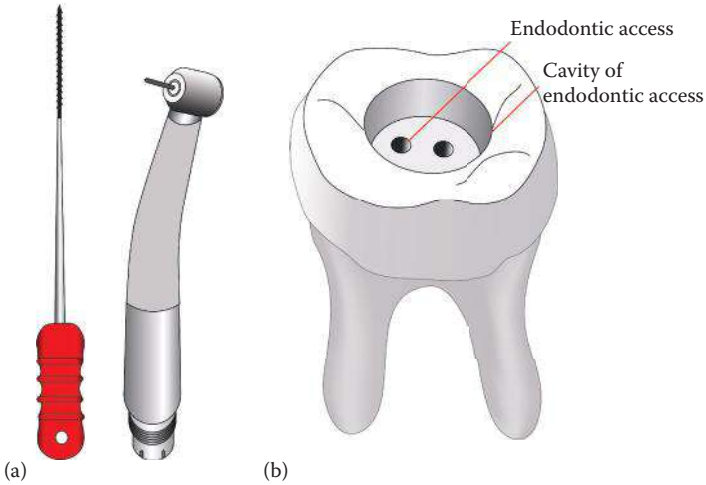

Figure 4.37 Extraction of pulp tissue using endodontic access procedures: trepanning the occlusal surface of a tooth using a dental bur mounted on a turbine (a), creating a cavity for endodontic access (b), and extracting pulp tissue using a nerve broach (a). (© Richard C. Li.)

(Figure 4.36). A horizontal section is cut through the cementum–enamel junction of the tooth (Figure 4.36). The root portion can be pulverized to a fine powder for DNA isolation. The crown can be preserved for forensic odontological comparisons if needed. Additionally, the extraction of pulp tissue can be carried out by standard endodontic access (Figure 4.37). For calcified tissues such as dentin and cementum, a decalcification step is needed to soften the dental matrix, which facilitates DNA isolation.

Bibliography

Abaz, J., et al., Comparison of the variables affecting the recovery of DNA from common drinking containers. *Forensic Sci Int*, 2002, **126**(3): 233–240.

Aditya, S., C.N. Bhattacharyya, and K. Chaudhuri, Generating STR profile from "Touch DNA". *J Forensic Leg Med*, 2011, **18**(7): 295–298.

Alakoc, Y.D. and P.S. Aka, "Orthograde entrance technique" to recover DNA from ancient teeth preserving the physical structure. *Forensic Sci Int*, 2009, **188**(1–3): 96–98.

5

Nucleic Acid Extraction

5.1 Basic Principles of DNA Extraction

Deoxyribonucleic acid (DNA) is a linear polynucleotide consisting of four types of monomeric nucleotides. Each nucleotide contains three components: a deoxyribose, a nitrogenous base, and a phosphate group (Figure 5.1). The four bases for DNA are adenine (A), cytosine (C), guanine (G), and thymine (T) (Figure 5.2). The deoxyribose is attached to the nitrogen of a base. The phosphate group is attached to the deoxyribose. In a *polynucleotide*, individual nucleotides are linked by phosphodiester bonds (Figure 5.3).

Interestingly, the first attempt to isolate DNA from humans, which involved the use of human leucocytes, was accomplished by the Swiss physician Friedrich Miescher in 1869. This study led to the discovery of the DNA molecule, which he referred to as "nuclein." Over the years, various methods for isolating DNA from blood samples were developed. However, many protocols still use the basic principles Miescher developed more than 100 years ago.

This chapter introduces basic techniques for DNA extraction used in forensic laboratories (Table 5.1). The commonly used methods isolate total cellular DNA, which is suitable for most forensic DNA analyses. However, DNA extraction procedures may vary according to the type of biological evidence, which can include cell types, substrates, the quantity of biological evidence collected, and the type of test that is being performed. Specific DNA extraction procedures for a sample of interest can be found in the literature. The extraction method of choice yields an optimal quantity, quality, and purity of DNA to satisfy forensic DNA testing needs. A sufficient quantity of DNA ensures the generation of a complete DNA profile. Poor quality of DNA, such as fragmentation due to DNA degradation, may result in a partial DNA profile or a failure to obtain a profile. Poor purity of DNA may cause interference during subsequent DNA testing. For example, DNA polymerase inhibitors interfere with DNA amplification (Chapter 7). Additional criteria for selecting proper DNA extraction methods include adaptability to automation, throughput potential, simplicity, the reduction of contamination risks, and cost-effectiveness. The most common DNA extraction protocols include basic components discussed below.

5.1.1 Cell and Tissue Disruption

In most DNA extraction protocols, enzymatic digestions, such as those with proteinase K, are used for cell and tissue disruption. The disruption process can also be carried out by boiling and by using alkali treatment and mechanical methods (Figures 5.4 and 5.5). Materials such as bone and teeth can be frozen in liquid nitrogen and then ground to a fine powder with a mortar or cryogenic grinder such as the SPEX CertiPrep® freezer mill (Figure 5.6). In such specimens, cells containing DNA

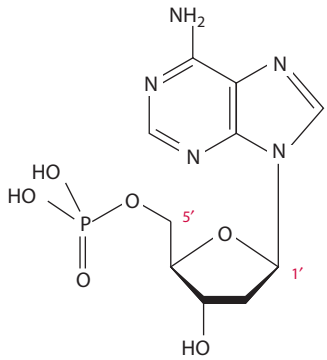

Figure 5.1 Structure of a nucleotide. Each nucleotide in a DNA polymer is made up of three components: a deoxyribose, a nitrogenous base (adenine is shown as an example), and a phosphate group. The 1′ carbon of the deoxyribose is attached to the nitrogen of a nitrogenous base. A phosphate group is attached to the 5′ carbon of the deoxyribose.

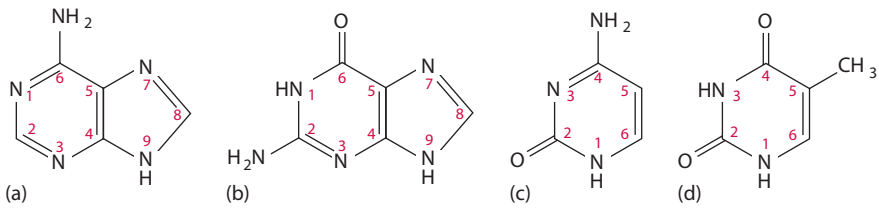

Figure 5.2 Chemical structures of nitrogenous bases: (a) adenine, (b) guanine, (c) cytosine, and (d) thymine.

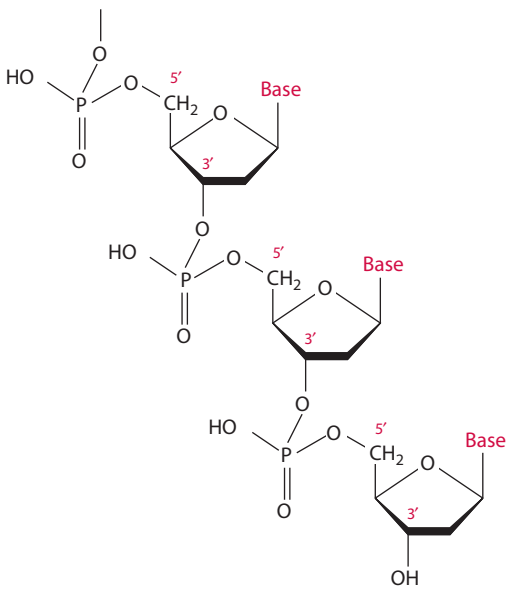

Figure 5.3 DNA polynucleotide chain. Individual nucleotides are linked by phosphodiester bonds between their 5′ and 3′ carbons.

Table 5.1 Three Basic DNA Extraction Methods

Method	Forensic Application	Purity	DNA Strand	DNA Size	Throughput	Note
Solvent-based	RFLP- and PCR-based assay	High	Double stranded	Large	Time-consuming; difficult to adapt for automation	Uses toxic solvent; multiple transfer steps between tubes
Boiling	PCR-based assay	Low	Single stranded	Small	~30 min per sample	No transfer required
Silica-based	RFLP- and PCR-based assay	High	Double stranded	Large	~1 h per sample; amenable to automation	Minimum transfer required

Figure 5.4 An automated mechanical disruption device for efficient disruption of tissues. It can be used for high-throughput applications involving sample preparation for the isolation of nucleic acids. (© Richard C. Li.)

Figure 5.5 Tissue disruption using a pressure-generating instrument also known as a barocycler. During the process, a barocycler is utilized to apply alternating cycles of high and low pressures onto specimens placed in single-use processing containers. The technique is known as pressure-cycling technology (PCT). (© Richard C. Li.)

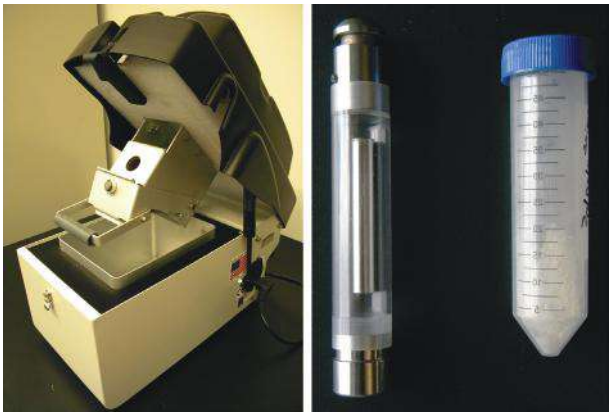

Figure 5.6 Tissue disruption using cryogenic grinding. A cryogenic mill (left), also known as freezer mill, can be used for the pulverization of hard tissue such as bone samples into fine powder prior to DNA extraction. During a grinding cycle, an impactor (right) moves back and forth, under a magnetic field, inside a grinding vial, grinding the sample to a fine powder. The grinding process is carried out at low temperatures (approximately –196°C), using liquid nitrogen to protect temperature-sensitive analytes such as DNA. (© Richard C. Li.)

are embedded in a calcified matrix that is responsible for the rigid structure of bone and teeth. In order to achieve a high yield of DNA extraction, it is necessary to remove the calcified matrix. A decalcification process removes calcium ions from the matrix, thus making the specimen suitable for DNA extraction. The most commonly used decalcifying agents for DNA extraction purposes are chelating agents such as ethylenediaminetetraacetic acid (EDTA), which sequesters the calcium ions. Decalcification is a lengthy procedure. Bone or teeth powder is usually treated with the decalcifying agent overnight or for a longer period of time, depending on the size and source of the sample.

5.1.2 Lysis of Cellular and Organelle Membranes

During or after tissue disruption, membranes—including those of cells, nuclei, and mitochondria—are lysed in order to release DNAs, including nuclear and mitochondrial DNA. The lysis can be carried out using salts and chaotropic agents (Section 5.2.3) such as guanidinium salts and detergents such as sarkosyl and sodium dodecyl sulfate (SDS). These substances can destroy membranes, denature proteins, and dissociate proteins such as histones from DNA. The lysis procedure is usually carried out in a buffer, such as Tris, in order to maintain a pH where endogenous *deoxyribonucleases* (DNases) remain inactive. DNases are a type of nuclease that catalyzes the cleavage of phosphodiester bonds of DNA. Endogenous DNases are located in cytoplasmic lysosomes and play a role in degrading the DNA of invading viruses. When cells are lysed, the DNases that are also released can degrade the extracted DNA. Chelating agents such as EDTA or Chelex® (Section 5.2.2) can therefore be used to chelate the divalent cations that are the cofactors of DNases, in order to inhibit DNase activities. Furthermore, reducing agents such as mercaptoethanol or dithiothreitol (DTT) can be used to inhibit the oxidization processes that can damage DNA.

5.1.3 Removal of Proteins and Cytoplasmic Constituents

After lysis, cytoplasmic constituents, such as proteins and liquids that interfere with DNA extraction, are removed. Proteins and lipids are usually removed by one or more rounds of extraction with organic solvents such as phenol–chloroform mixtures (Section 5.2.1). Another strategy to remove cytoplasmic constituents is to utilize the reversible binding of DNA to a solid material such as silica, which selectively binds DNA in chaotropic salt solutions. The proteins and cytoplasmic constituents can then be removed through washing steps.

5.1.4 Storage of DNA Solutions

Purified high-molecular-weight DNA is usually stored in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). EDTA is usually included in storage solutions to chelate divalent cations and thereby inhibit DNases. Such a DNA solution may be stored at 4°C or −20°C. For long-term storage, −80°C is recommended. Frequent freezing or thawing cycles should be avoided because temperature fluctuations may cause breaks of single- and double-stranded DNA. Additionally, DNA containing impurities such as extracts generated via the Chelex method are much less stable. The presence of heavy metals, such as cadmium and cobalt, can cause breakage of phosphodiester bonds in the molecules. Additionally, unsuccessful removal of exogenous or intracellular free radicals, such as a hydroxyl radical (HO·), which have unpaired valence electrons, can cause DNA damage such as strand breaks. Furthermore, contamination by DNases may lead to subsequent degradation of DNA.

5.1.5 Contamination

Contamination is usually caused by the introduction of exogenous DNA to an evidence sample. It can occur between samples, between an individual and a sample, between other organisms and a sample, or between amplified DNA and a sample. Certain procedures can be utilized to prevent the occurrence of contamination. For example, evidence and reference samples should be processed separately in different rooms to avoid sample-to-sample contamination. In situations where space is limited, the evidence sample should be processed before the reference sample. Additionally, evidence samples should be processed and extracted for isolating DNA in separate areas (or at different times) from DNA amplification areas.

Solutions and test tubes used for extraction should be DNA-free, and aerosol-resistant pipet tips should be used during the extraction process. Additionally, the levels of contamination should be monitored by using *extraction reagent blanks* (having reagents but no samples), which monitor contamination from the extraction.

5.2 Methods of DNA Extraction

5.2.1 Extraction with Phenol–Chloroform

This method is also called organic extraction. Major steps include the following:

5.2.1.1 Cell Lysis and Protein Digestion

These steps can be achieved by digestion with proteolytic enzymes such as proteinase K before extraction with organic solvents.

5.2.1.2 Extraction with Organic Solvents

The removal of proteins is carried out by extracting aqueous solutions containing DNA with a mixture of phenol:chloroform:isoamyl alcohol (25:24:1). Phenol is used to extract the proteins from the aqueous solution. Although phenol has a slightly higher density than water, it is sometimes difficult to separate it from the aqueous phase. Therefore, chloroform is utilized as it has a higher density than phenol. As a result, the phenol–chloroform mixture forms the organic phase at the bottom of the tube and is easily separated from the aqueous phase. Isoamyl alcohol is often added to the phenol–chloroform mixture to reduce foaming. During partition, DNA is solubilized in the aqueous phase, while lipids are solubilized in the organic phase. Proteins are located at the interface between the two phases (Figure 5.7).

5.2.1.3 Concentrating DNA

Two common methods for concentrating DNA are ethanol precipitation and ultrafiltration. In the first method, the DNA is precipitated from the aqueous solution with ethanol and salts. Ethanol depletes the hydration shell of DNA, thus exposing its negatively charged phosphate groups. The

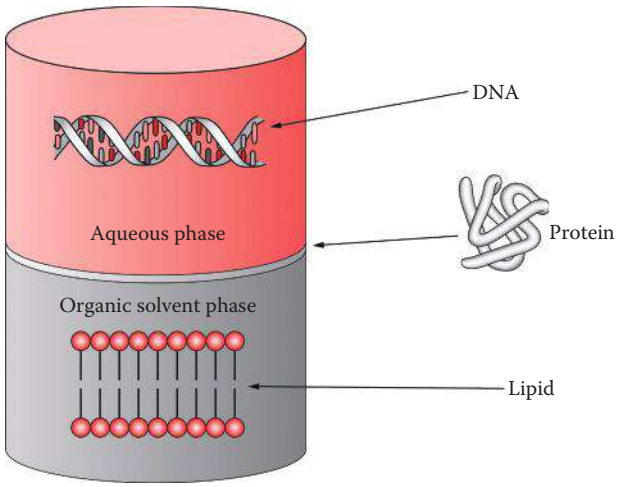

Figure 5.7 DNA extraction using organic solvent. The DNA is contained in the aqueous phase, while cellular materials such as lipids are contained in the organic-solvent phase. Proteins remain in the barrier between the two phases. (© Richard C. Li.)

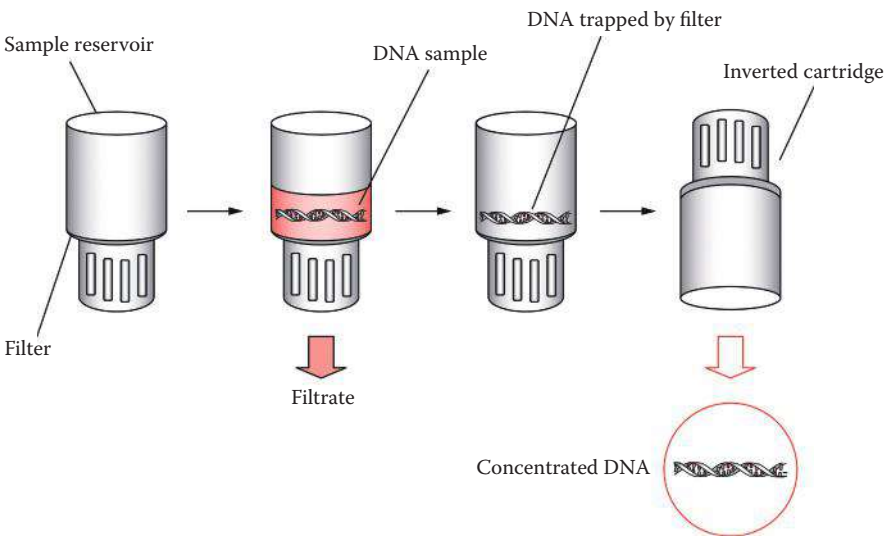

Figure 5.8 Concentrating DNA solutions using filtration devices. DNA samples are loaded into the reservoir. The liquid is filtered by centrifugation and the DNA becomes trapped in the membrane. The cartridge is then inverted to recover the trapped DNA by centrifugation. (© Richard C. Li.)

precipitation can only occur if sufficient quantities of cations are present in the solution. The most commonly used cations, such as ammonium, lithium, and sodium, neutralize the charges on the phosphate residues of DNA, forming a precipitate. Ultrafiltration is an alternative to ethanol precipitation for concentrating DNA solutions. The Microcon® and Amicon® are centrifugal ultrafiltration devices that can concentrate DNA samples (Figure 5.8). A proper Microcon® unit can be selected with a nucleotide cutoff equal to or smaller than the molecular weight of the DNA fragment of interest. Usually, the cutoff size is 100 kDa for forensic DNA samples.

Phenol–chloroform extraction yields large, double-stranded DNA and can be used for either restriction fragment length polymorphism (RFLP)-based or polymerase chain reaction

(PCR)-based analysis. However, the organic extraction method is time-consuming, involves the use of hazardous reagents, and requires the transfer of samples among tubes.

5.2.2 Extraction by Boiling Lysis and Chelation

This technique, also called the Chelex® extraction method, was introduced in the early 1990s to forensic DNA laboratories. It usually includes the following steps:

5.2.2.1 Washing

This step removes contaminants and inhibitors that may interfere with DNA amplification. For example, heme compounds found in blood samples should be removed from blood samples because it inhibits DNA amplification.

5.2.2.2 Boiling

Cells are suspended in a solution and incubated at 56°C, where DNases are not active, for 20 min. This preboiling step softens cell membranes and separates clumps of cells from each other. The cells are then lysed by heating to boiling temperature in order to break open the membranes and to release the DNA. Additionally, the lysis of cells releases all of their cellular constituents, including DNases. The DNase degradation of the extracted DNA can be blocked by applying a chelating resin (Chelex® 100) during the extraction process. Chelex® 100 is an ion-exchange resin composed of styrene divinylbenzene copolymers. The paired iminodiacetate ion groups in Chelex® 100 act as chelators by binding to divalent metal ions such as magnesium. Magnesium is a cofactor of endogenous DNases. Thus, sequestering magnesium in the solution using Chelex® 100 protects DNA from degradation by DNases (Figure 5.9).

5.2.2.3 Centrifugation

Brief centrifugation is performed to pull the Chelex® 100 resin and cellular debris to the bottom of the tube. The supernatant is used for DNA analysis. Carrying the Chelex® 100 resin over into the DNA amplification solutions should be avoided because the resin chelates magnesium, which is a necessary cofactor for DNA polymerases used for amplification.

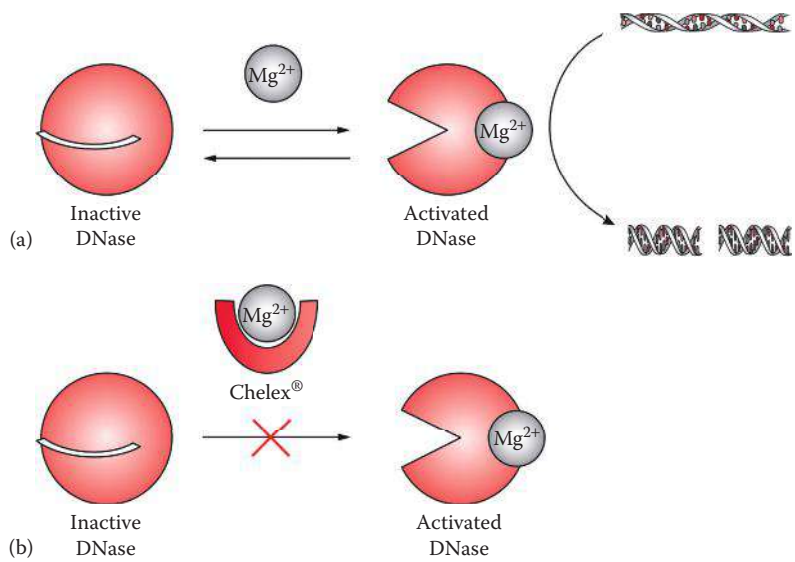

Figure 5.9 Utilizing Chelex® for preparing DNA samples. (a) Cations such as Mg^{2+} are required for the activity of endogenous DNase, which degrades DNA. (b) Chelex® prevents DNA degradation from endogenous DNase by sequestering the Mg^{2+}. (© Richard C. Li.)

This method is simple and rapid and uses only a single tube for extraction, thus reducing the risks of contamination and sample mix-ups. However, the heating step of this method disrupts and denatures proteins and also affects the chromosomal DNA. The resulting DNA extracted from the solution is fragmented single-stranded DNA. Thus, the DNA extracted is not suitable for RFLP analysis because RFLP requires double-stranded DNA samples. The DNA obtained by lysis and chelation can only be used for PCR-based DNA analysis.

5.2.3 Silica-Based Extraction

The method of adsorbing DNA molecules to solid silica surfaces is used for extraction (Figures 5.10 and 5.11). The method is based on the phenomenon that DNA is reversibly adsorbed to silica—silicon dioxide (SiO_2)—in the presence of high concentrations of chaotropic salts. Chaotropic salts can disrupt hydrogen bonding, affecting the three-dimensional structures of macromolecules. These salts are used to denature proteins. Additionally, chaotropic salts can facilitate the adsorption of DNA to silica.

In aqueous solutions, the hydration shells of nucleic acids shield the negative charges of phosphates at the phosphodiester backbone of nucleic acids. As a result, nucleic acids are usually hydrophilic in aqueous solutions. In the presence of chaotropic salts, nucleic acids become hydrophobic. The dehydrating effect caused by chaotropic salts allows the phosphate residues to become available for adsorption to the silica surface. Common chaotropic salts utilized for DNA extraction include guanidinium salts such as guanidinium thiocyanate (GuSCN) and guanidinium hydrochloride (GuHCl). GuSCN is a more potent chaotropic salt and also facilitates cell lysis and DNA adsorption. This technique usually includes the following steps:

5.2.3.1 Cell Lysis and Protein Digestion

This is carried out by proteinase K digestion. The cell membranes are broken open, and DNA is released.

5.2.3.2 DNA Adsorption onto Silica

This step utilizes silica as the stationary phase in a membrane configuration to which the DNA in the cell lysate binds. Adsorption of the DNA to the silica occurs in the presence of high concentrations of chaotropic agents (some protocols adjust pH conditions to enhance adsorption). Under these conditions, cellular materials and other contaminants that can inhibit DNA amplification reactions are not retained on the silica membrane. The adsorbed DNA is largely double stranded.

Figure 5.10 A silica-membrane spin column. (© Richard C. Li.)

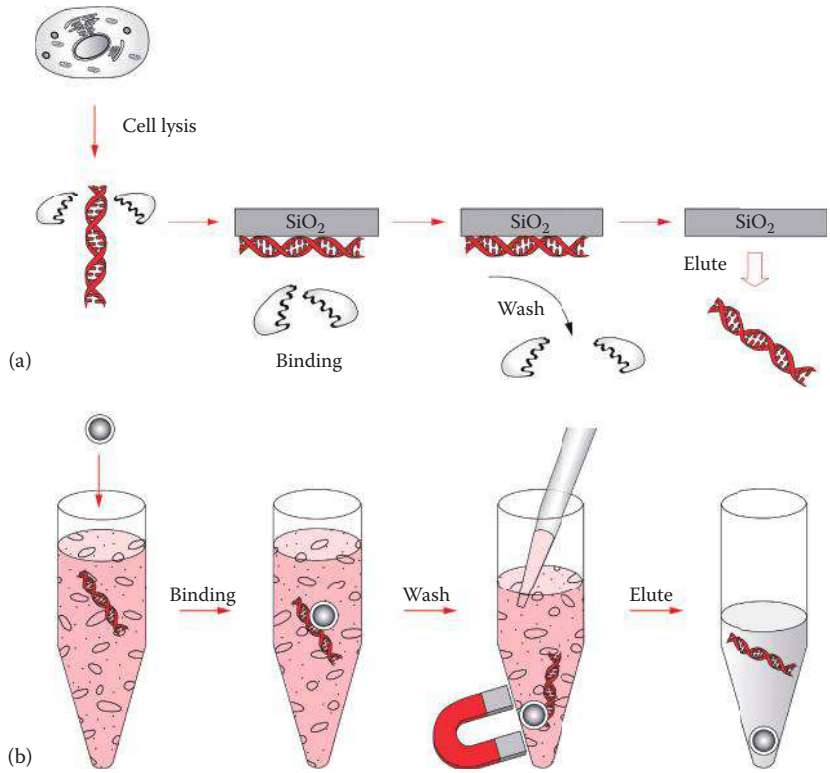

Figure 5.11 Silica-based DNA extractions. (a) Cells are lysed in the presence of proteinase. The DNA then binds to the silica matrix. A washing step removes unbound cellular materials and salts from the matrix. The purified DNA is then eluted for use in downstream applications. (b) Using silica-coated paramagnetic particles for DNA extraction. The particle is added into a lysate, and then DNA binds to the silica surface of the particle. The particle is then captured by a magnetic field. Afterwards, the supernatant is removed, and cellular materials are washed away. DNA elution follows. See Section 5.2.3. (© Richard C. Li.)

5.2.3.3 Washing

This step removes chaotropic agents and other contaminants. An ethanol-based wash solution is used. This wash solution does not remove DNA from the silica. The chaotropic agents and contaminants that are present in the solution can be removed using the ethanol-based wash solution.

5.2.3.4 Elution of DNA

The adsorbed DNA can be eluted by rehydration with aqueous low-salt solutions. The eluted DNA is double stranded and can be used for a wide variety of applications.

Silica-based extraction methods yield high-quality DNA. Silica membrane devices can also be adapted for automation; for example, this can be done by using 96-well silica membrane plates and a variety of robotic platforms. Another type of device utilizes silica-coated paramagnetic particles that adsorb DNA in the solution (Figure 5.11b). A magnet is used for particle capture instead of centrifugation or vacuum filtration. The magnetic particles can be resuspended during the wash steps, and the solution containing contaminants and cellular materials is then discarded. DNA is eluted after washing. This device can also be adapted to automated, high-throughput methods.

Over the years, high-throughput silica-based procedures have been developed to process large numbers of samples in parallel. Some of these methods are adapted for automated DNA extraction platforms (Figures 5.12 and 5.13).

Figure 5.12 Automated bench-top DNA purification systems that enable the isolation of genomic DNA from a wide variety of forensic samples. A low-throughput sample-preparation system that can process up to 12 samples (left) and a low-to-medium throughput sample-preparation system (right). (© Richard C. Li.)

Figure 5.13 Integrated platforms enabling automated DNA sample preparation and liquid handling for subsequent assay setup. These systems utilize silica-coated bead chemistries and can achieve moderate to high throughput of processing (1–96 samples) with bar coding for sample tracking. A Hamilton (top) and a Qiagen system (bottom). (© Richard C. Li.)

5.2.4 Differential Extraction

This method is very useful for the extraction of DNA from biological evidence derived from sexual assault cases, such as vaginal swabs and bodily fluid stains. These types of evidence often contain mixtures of spermatozoa from a male contributor and nonsperm cells such as epithelial cells from a female victim. Mixtures of individual DNA profiles can complicate data interpretation.

This method selectively lyses the nonsperm and spermatozoa in separate steps based on the differences in cell-membrane properties of spermatozoa and other types of cells. Thus, the DNA from spermatozoa and nonsperm cell fractions can be sequentially isolated.

First, the differential extraction procedure involves preferentially lysing the nonsperm cells with proteolytic degradation using proteinase. Sperm plasma membrane contains proteins cross-linked by disulfide bonds. The membrane exhibits a much higher mechanical stability than nonsperm cells and is thus resistant to proteolytic degradation. The nonsperm DNA is released into the supernatant and the liquid containing it (the nonsperm fraction) is extracted, yielding a fraction that predominantly contains DNA from nonsperm cells.

To lyse the sperm cells, it is necessary to cleave the disulfide bonds in addition to proteolytic digestion. The application of DTT, a reducing agent, is an approach that can be used for cleavage. In the presence of DTT and proteinase K, the sperm plasma membrane is then lysed (Figure 5.14). Subsequently, DNA from the sperm cells can be extracted (Figure 5.15).

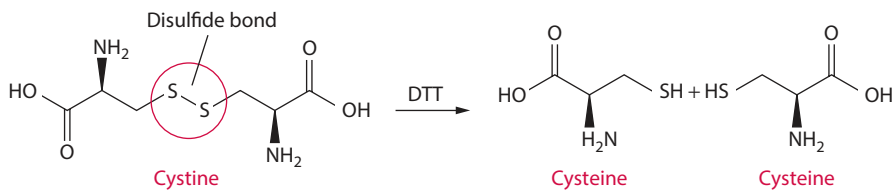

Figure 5.14 DTT reaction. The breaking of disulfide bonds in cystine residues is carried out by adding a reducing agent such as DTT.

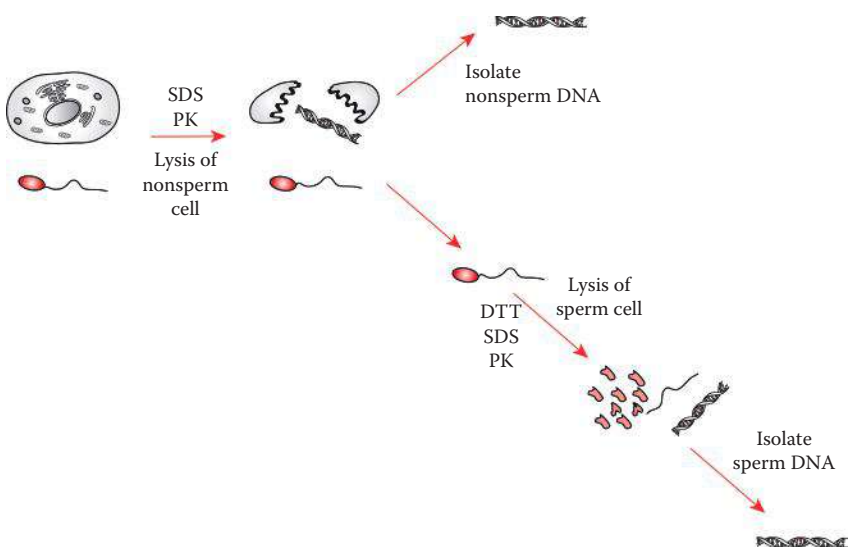

Figure 5.15 Differential extraction process used to separate sperm cells from nonsperm cells. Nonsperm cells are lysed in the presence of SDS and proteinase K (PK); sperm cells are resistant to such conditions. The nonsperm cell DNA is extracted. The sperm cells are then lysed separately in the presence of SDS and proteinase K plus DTT to extract the sperm DNA. (© Richard C. Li.)

This process isolates sperm and nonsperm cell DNA separately for obtaining DNA profiles from male and female contributors, respectively. However, the non-sperm-cell DNA and sperm-cell DNA may not be completely separated from one another; this can happen, for example, if the sperm cells have already lysed due to poor sample conditions. Some sperm DNA may be present in the non-sperm-cell fraction. Additionally, if a mixture has an abundance of non-sperm cells and fewer sperm cells, non-sperm-cell DNA may be detected in the sperm fraction. Thus, new methods that can overcome these problems are highly desired.

5.3 Essential Features of RNA

Like DNA, *ribonucleic acid* (RNA) is a linear molecule containing four types of nucleotides linked by phosphodiester bonds (Figures 5.16 and 5.17). However, certain properties of RNA differ from those of DNA. Unlike DNA containing deoxyribose (see Figure 5.18), the sugar residue

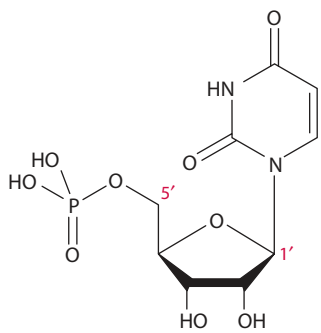

Figure 5.16 The structure of ribonucleotide. Each ribonucleotide in a RNA polymer consists of three components: a ribose, a nitrogenous base (uracil is shown as an example), and a phosphate group. The 1′ carbon of the ribose is attached to the nitrogen of the nitrogenous base. A phosphate group is attached to the 5′ carbon of the ribose.

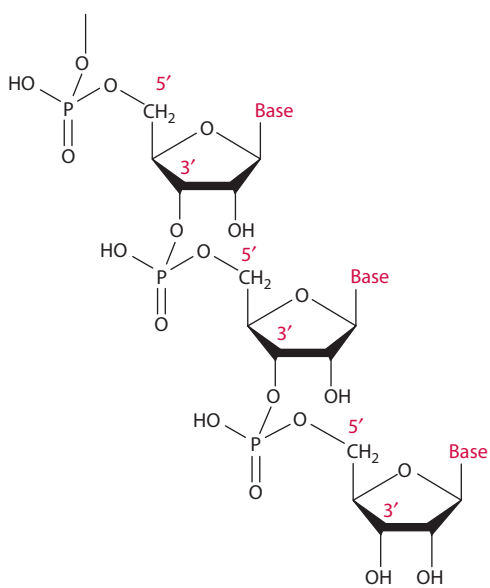

Figure 5.17 Chemical structure of RNA (polyribonucleotide) chain.

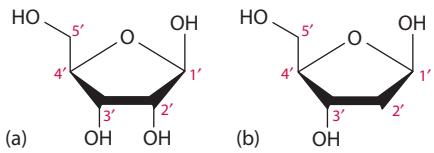

Figure 5.18 Comparison of (a) ribose and (b) deoxyribose. The hydroxyl group attached to the 2′ carbon of ribose is replaced by a hydrogen group in deoxyribose.

of RNA is a ribose, which has a hydroxyl (OH) group at the 2′ carbon position. Therefore, RNA has a lower pKa, which means that it is more acidic than DNA. RNA contains uracil (U) in the place of thymine in DNA (Figure 5.19). RNA is typically found in cells as a single-stranded molecule, while DNA is double stranded. Polyribonucleotides can form complementary helices with DNA strands by base pairing (with the exception of uracil pairs with adenine) (see Figure 5.20).

Total RNA contains all the RNA of the cells, including RNAs involved in protein synthesis and posttranscriptional modification, as well as regulatory RNAs. The RNA that carries codes from a DNA template is called messenger RNA (mRNA) and usually contains a cap and polyadenine tail at the 5′ and 3′ ends of the molecule, respectively (Figure 5.21).

The stability of RNA is an issue for mRNA-based forensic analysis. Single-stranded RNA is chemically less stable than DNA. The 2′ OH group of RNA can react with its phosphodiester backbone, potentially causing the nonenzymatic hydrolysis of an RNA molecule. Several factors, such as moisture, UV light, high temperature, and extreme pH, can facilitate the nonenzymatic

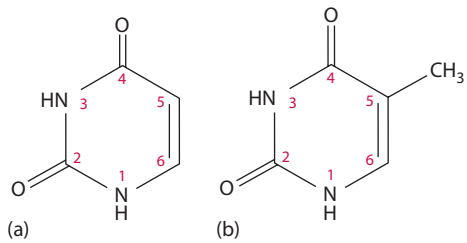

Figure 5.19 Comparison of (a) uracil and (b) thymine. Thymine has a methyl group attached at the 5-carbon position.

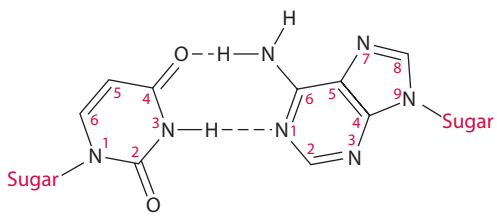

Figure 5.20 Base pairing of uracil (left) and adenine (right).

Figure 5.21 Structure of mRNA. Eukaryotic mRNA is modified at the 5′ end (cap) and 3′ end (polyadenine tail) needed for protein synthesis. Start and stop codons are required for the initiation and termination of protein synthesis. (© Richard C. Li.)

hydrolysis of RNA phosphodiesters, which leads to the degradation of RNA. Moreover, endogenous *ribonucleases* (RNases), present within cells, represent the major factor causing RNA degradation. Environmental microorganisms are also a common source of RNase contamination. RNases are very stable and usually do not require cofactors to carry out enzymatic reactions. Small amounts of RNases are sufficient to degrade RNA. RNA can be protected by adding RNase inhibitors such as diethylpyrocarbonate. Additionally, using RNase-free laboratory supplies and reagents and wearing disposable gloves while handling samples and reagents can reduce the risks of RNase contamination.

5.4 Methods of RNA Extraction

5.4.1 RNA–DNA Coextraction

For most forensic applications, total RNA is usually isolated. The procedure for isolating total RNA is simpler than that for mRNA. The total RNA isolated usually contains a sufficient amount of mRNA for subsequent reverse transcriptase PCR (RT-PCR; see Chapter 7). However, if the quantity of a target mRNA is very low, the procedures specifically for isolating mRNA should be used. There are a variety of protocols for isolating total RNA. However, RNA–DNA coextraction methods allow for the simultaneous extraction of high-quality DNA and RNA for forensic DNA analysis and bodily-fluid identification (Chapter 11), respectively, without the consumption of additional samples when extracted separately.

Biological samples are first lysed in lysis buffer. The lysis buffer usually contains chaotropic salts that can facilitate the lysis of cells and denature proteins, thus inactivating endogenous RNases to protect RNA. The lysate is then passed through a silica membrane column. A high-salt environment in the lysate allows selective binding of silica to genomic DNA over RNA. The column is washed, and purified DNA is then eluted.

The extraction of RNA is achieved by utilizing a high concentration of chaotropic salts that are already present in the lysate, along with ethanol, in order to decrease the hydrophilic property of RNA and increase its affinity for silica. After ethanol is added to the flow-through that passed the first column, the sample is then applied to a second silica column, where total RNA molecules, longer than 200 nucleotides, bind to the silica membrane. RNA including mRNA is then eluted (Figure 5.22). This method does not isolate small RNAs such as miRNA, rRNA, and tRNA, which comprise 15%–20% of total RNA.

5.4.2 miRNA Extraction

miRNAs (Chapter 4) are low-molecular-weight RNAs ranging between 15 and 30 nucleotides in length. Conventional methods routinely used for extracting total RNA do not effectively recover small RNAs; thus, they are not suitable for isolating miRNA. One approach to extracting small RNA molecules including miRNA involves two steps: organic-solvent extraction to isolate total RNA and solid-phase extraction to enrich small RNA.

To begin the process, tissues are disrupted and cells are lysed. The lysis reagents also inactivate RNases to protect RNA. The lysate is then extracted with an organic solvent, such as phenol and chloroform, with a high concentration of chaotropic salts such as GuSCN. The partitioning of DNA and RNA between the organic phase and the aqueous phase is determined by the pH during the organic extraction. At an acidic pH, DNA partitions to the organic phase. The pKa of RNA is usually lower than DNA; thus, RNA remains hydrophilic and is retained in the aqueous phase under this condition. This step removes most of the DNA, proteins, and other cellular components from the lysate into the organic phase.

The second step, silica-based extraction, further purifies and enriches small RNAs. It is achieved by utilizing the high concentration of chaotropic salt that is already present in the lysate, along with ethanol. The size-fractioning of RNAs is achieved using different ethanol

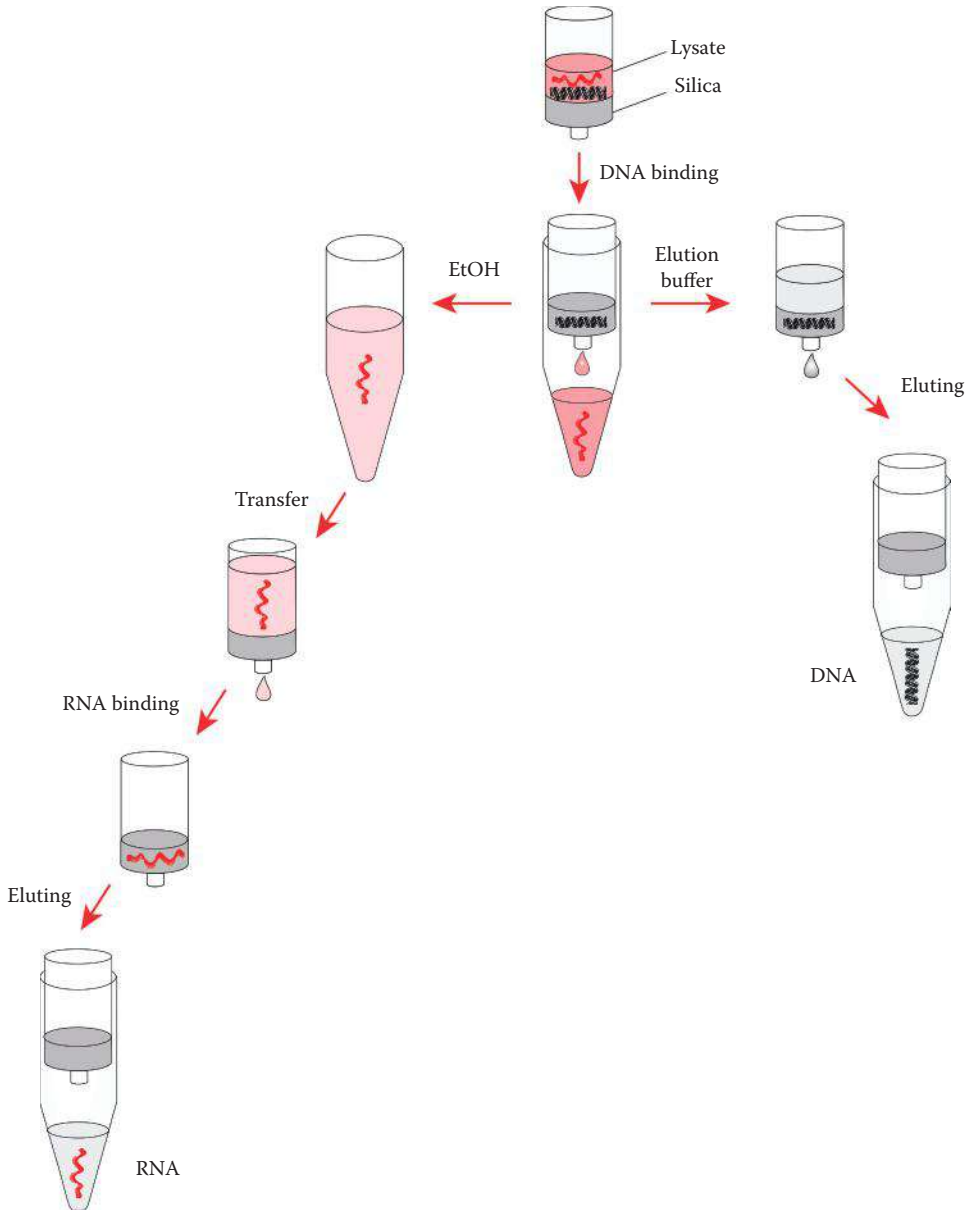

Figure 5.22 RNA and DNA coextraction using a silica-based method. See Section 5.4.1. (© Richard C. Li.)

concentrations. At a low ethanol concentration, the solvent-extracted lysate is passed through a silica membrane filter. Large RNAs are retained on the silica membrane, while small RNAs are not and are collected in the filtrate. The filtrate is passed through a second silica membrane filter. At a high ethanol concentration, the small RNAs are retained on the silica membrane. The small RNAs are then eluted in a low ionic strength solution (Figure 5.23). RNA less than 200 nucleotides, including miRNA, can be obtained.

After RNA is extracted, the small amounts of RNA can be quantified using a quantitative RT-PCR (Chapter 7) or fluorescent intercalating dye assay (Chapter 6). Purified RNA can be

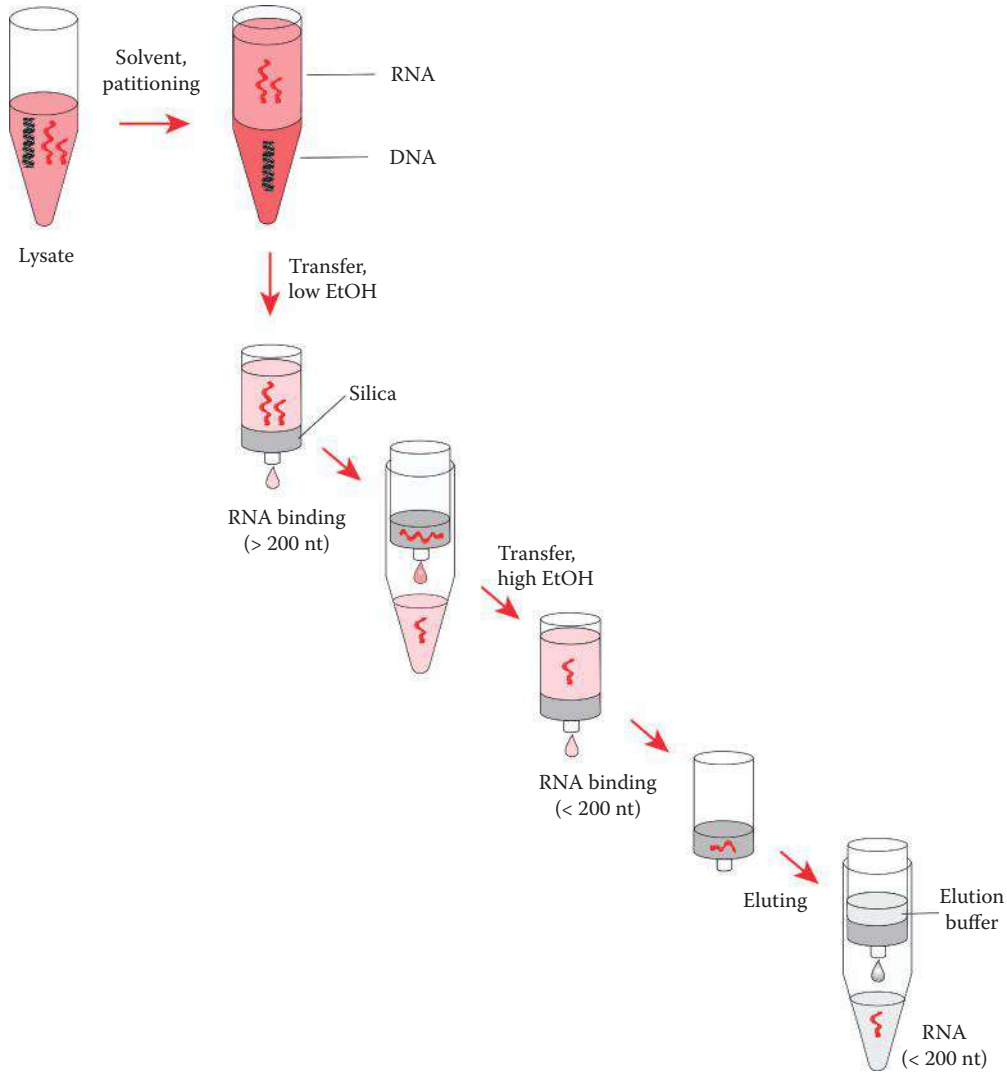

Figure 5.23 Schematic illustration of an miRNA extraction method. See Section 5.4.2. (© Richard C. Li.)

stored at −80°C in RNase-free water or Tris-EDTA buffer (pH 7). Single-use aliquots are preferred when possible to avoid multiple freeze–thaw cycles.

The integrity of extracted RNA is a great concern for RNA-based forensic analysis. RNA quality is traditionally assessed using the 28S:18S rRNA ratio. There are some conditions where this method has been shown to be inconsistent; that is, in which the 28S:18S rRNA ratio does not reflect the true level of RNA degradation. Recently, the *RNA integrity number* (RIN) has provided an effective method for determining RNA quality. RIN is a software algorithm that takes the electrophoretic RNA measurements into account in order to assign integrity values to RNA samples. RIN ranges from 1 to 10, with 10 being the most intact. Thus, the RIN method facilitates the assessment of the integrity of RNA samples. As reflected by RIN, RNA degradation has a negative influence on the reproducibility of the results of RNA-based forensic analysis.

6

DNA Quantitation

Determining the amount of DNA in a sample is essential for polymerase chain reaction (PCR)-based DNA testing (Chapter 7). Because PCR-based DNA testing is very sensitive, a narrow concentration range is required for amplification to be successful. If too much DNA template is used in PCR, the resulting artifacts may interfere with data analysis and interpretation. On the other hand, very low amounts of DNA template may result in a partial DNA profile or failure to attain a profile. Samples with poor purity may contain PCR inhibitors that lead to a failure of DNA amplification. Thus, a test that can measure the quality and quantity of the DNA template of a sample is desirable.

Additionally, forensic samples containing DNA are often mixtures that may include nonhuman DNA. For instance, microbial DNA may be present. Thus, it is necessary to use human-specific DNA quantitation methods to selectively determine the amount of human DNA present. In the United States, quality-assurance guidelines require the use of a quantitation method that estimates the amount of human nuclear DNA in crime-scene evidence samples.

In this chapter, the slot blot, intercalating dye, and quantitative PCR methods are introduced. The quantitative PCR method is the most sensitive of the three methods. It is the only method that can detect PCR inhibitors. Both the slot blot and the quantitative PCR methods can detect human and nonhuman primate DNA. Current technology cannot distinguish between human and nonhuman primate DNA.

6.1 Slot Blot Assay

Historically, the slot blot assay was used to detect human genomic DNA in a sample. Methodically, the slot blot assay works based on the following principles. Prior to the quantitation, an alkaline solution is added to the genomic DNA sample, which denatures DNA. Generating single-stranded DNA is necessary for DNA to be cross-linked onto a nitrocellulose membrane. The DNA sample is then spotted, using a slot blot device, onto a nitrocellulose membrane (Figure 6.1). The single-stranded DNA is then immobilized onto a nylon membrane. The targeted sequence is revealed by hybridization with a labeled 40-nucleotide probe complementary to a primate-specific α-satellite DNA sequence at the D17Z1 locus (Figure 6.2). In humans, the α-satellite DNA sequences are highly repetitive sequences located near the centromeres of chromosomes. These sequences are usually distinct for each chromosome.

Three detecting schemes of the slot blot assay have been developed. Initially, the D17Z1 probe was labeled with radioisotopes that could be visualized by exposing the slot blot membrane to x-ray film. The hazardous radioisotope detection method was then replaced by alkaline phosphatase-labeled and biotinylated probes. The alkaline phosphatase-labeled probe can be coupled with chemiluminescent detection (Lumi-Phos Plus kit, Lumigen, Inc.). The biotinylated probe

Figure 6.1 A slot blot device. (© Richard C. Li.)

(QuantiBlot Human DNA Quantitation Kit, Applied Biosystems) can be coupled with either colorimetric or chemiluminescent detection.

In colorimetric detection, the biotin moiety of the probe is bound to streptavidin. The streptavidin is conjugated with horseradish peroxidase, which catalyzes the oxidation reaction of tetramethylbenzidine (TMB), a substrate, forming a blue precipitate. With the chemiluminescent detection method, the horseradish peroxidase catalyzes the oxidation reaction of a substrate such as luminol, emitting photons that can be detected by exposure to x-ray film. The sensitivity of chemiluminescent detection is slightly higher than that of the colorimetric detection. The detection mechanisms will be discussed in Chapter 9.

The detected signal intensity is proportional to the concentration of the DNA sample in question. Quantitative measurements can be made by comparing an unknown sample to a set of standards with known DNA concentrations (Figure 6.3).

The assay typically quantifies DNA over the range of 150 pg–10 ng. However, the quantitation results are manually read, and conclusions are based on subjective judgments. Additionally, the D17Z1 sequences of humans and other primates share homology. The probe cannot distinguish between human and other primate genomic DNA. This is not a great concern because cases involving nonhuman primates are rare. Nevertheless, the probe does not cross-react with all other species. This assay has been replaced by recently developed quantitative PCR assays (see Section 6.3).

6.2 Fluorescent Intercalating Dye Assay

Small quantities of DNA can also be quantified by using a fluorescent intercalating dye method. Intercalating dyes, usually planar molecules, can slide themselves in between base pairs of DNA without breaking the DNA double helix. The Quant-iT™ PicoGreen® dsDNA reagent (Invitrogen) is a fluorescent intercalating dye that stains double-stranded DNA (dsDNA) for quantitation in a sample (Figure 6.4a). The detection limit of this method is approximately 250 pg. Intercalating dyes, not specific to human DNA, bind to all DNA molecules. Therefore, fluorescent intercalating dye assay can be utilized for the quantitation of known reference samples. For instance, DNA database samples from known sources can be quantified using this method. The assay has also been adapted for automation and is, thus, a high-throughput method.

DNA samples are simply added to a solution containing the fluorescent intercalating dye. The fluorescence, proportional to the quantities of DNA (Figure 6.4b), is measured using a standard spectrofluorometer with excitation and emission wavelengths of the light source. A standard

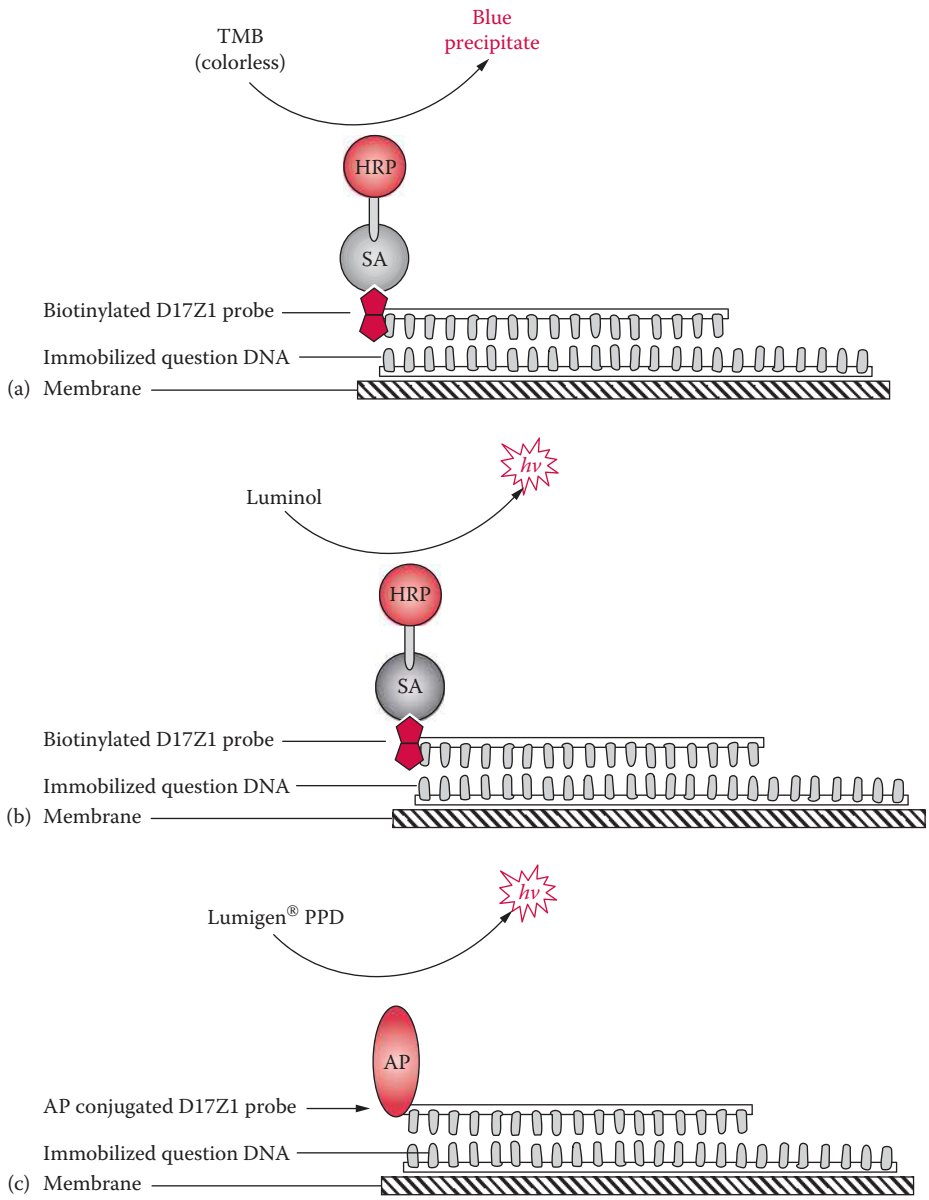

Figure 6.2 Slot blot assay. Questioned DNA is immobilized onto a solid-phase membrane then hybridized with a biotinylated D17Z1 probe. The detection of the hybridization is carried out by (a) streptavidin (SA) and horseradish peroxidase (HRP) conjugate, and a colorimetric reaction is catalyzed by HRP using tetramethylbenzidine (TMB) as a substrate; (b) SA and HRP conjugate, and a chemiluminescent reaction is catalyzed by HRP using Luminol as a substrate; (c) immobilized DNA is hybridized with an alkaline phosphatase (AP)-labeled D17Z1 probe. The detection of the hybridization is carried out by a chemiluminescent reaction catalyzed by AP using Lumigen® PPD as a substrate. (© Richard C. Li.)

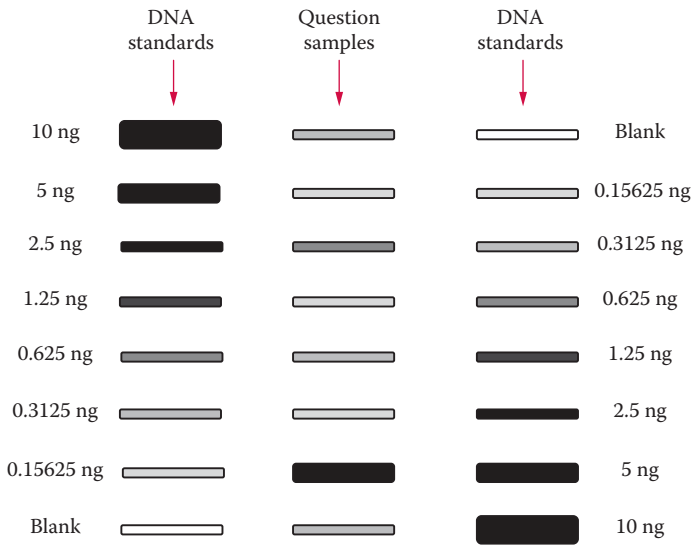

Figure 6.3 Human DNA quantitation using the slot blot assay. Standards with known amounts of human DNA are applied, and unknown samples and a set of standards are compared. The quantities in the unknown samples are estimated by visual comparison to the standards. (© Richard C. Li.)

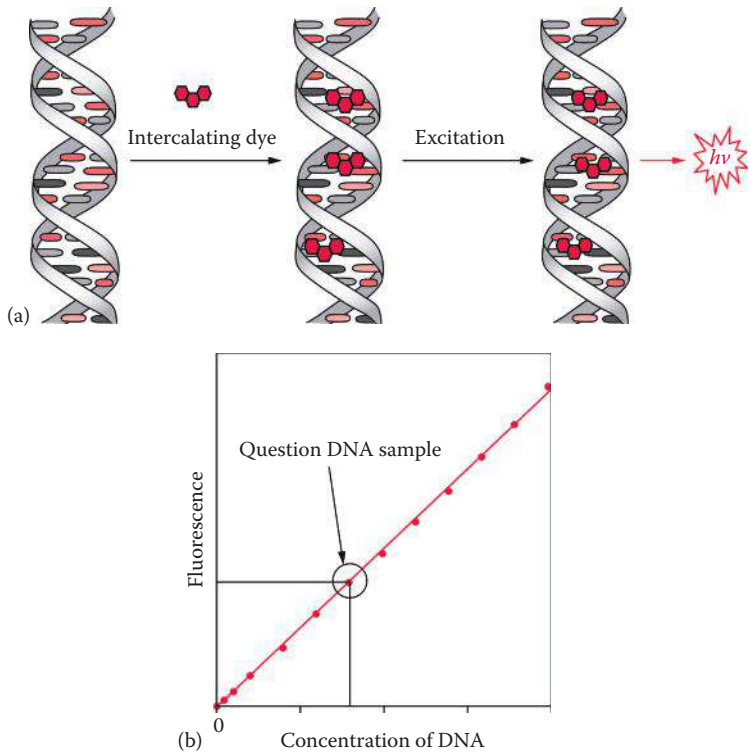

Figure 6.4 DNA quantitation using intercalating dye. (a) Fluorescent dye intercalates into DNA. The fluorescence can be measured upon applying an excitation light source. (b) A standard curve can be constructed using known amounts of DNA standards. The amount of questioned DNA can be determined by comparing the standard curve. (© Richard C. Li.)

curve is first created using samples containing known amounts of DNA. The assay is then performed for the unknown samples and the quantities of DNA in the samples are determined by comparing the results to the standard curve.

6.3 Quantitative PCR Assay

Based on the principle of PCR amplification (Chapter 7), the amount of PCR product amplified correlates with the initial concentration of DNA templates. Thus, the DNA concentration of a sample can be determined. There are two types of quantitative PCR methods. *End-point PCR* methods measure the quantity of amplified product at the end of PCR. Usually, the fluorescence is emitted by the dyes that intercalate into the double-stranded DNA. The quantity of amplified DNA is measured from the amount of fluorescence emitted from dyes such as SYBR (Figures 6.5 and 9.1b).

Real-time PCR methods can quantify the amplified DNA during the exponential phase of PCR (Chapter 7). The quantitation result is not affected to a significant extent by slight variations in PCR conditions. Thus, the precision of the quantitation of target sequences is improved with this method.

6.3.1 Real-Time Quantitative PCR

Real-time quantitative PCR (qPCR) was developed in the early 1990s, and it analyzes the amplification of a target sequence at each cycle of PCR. A fluorescent reporter is used to monitor the accumulation of amplified products during PCR. The fluorescence signals of the reporter molecule increase as amplified products accumulate with each cycle of PCR. qPCR is commonly used because of the following advantages:

- Better objectivity than the QuantiBlot method
- Increased sensitivity with a large dynamic range (30 pg–100 ng)
- More accurate measurements of small quantities of DNA in samples
- Fewer laboratory manipulations; amenable to automation
- Ability to detect PCR inhibitors (Section 6.3.1.1)

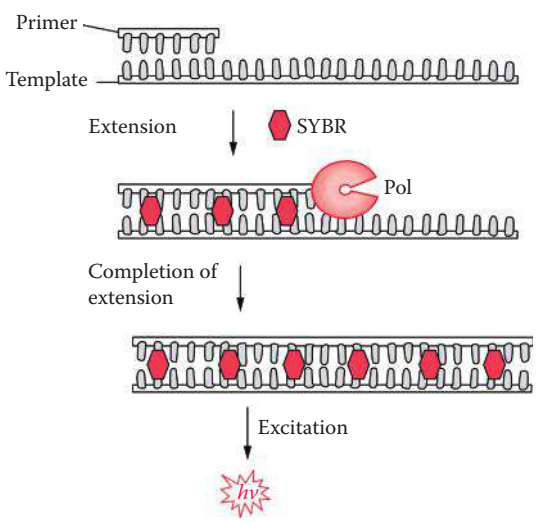

Figure 6.5 End-point PCR using SYBR Green detection. During the extension phase of PCR, in which DNA synthesis occurs, the dye binds to the double-stranded amplicons. Upon excitation, the emission intensity of the dye can be measured. Pol represents Taq polymerase. (© Richard C. Li.)

The technique is amenable to multiplexing to detect more than one type of DNA target sequence in a single reaction. Commercial qPCR kits for human DNA and Y-chromosome DNA quantitation are available. Additionally, the qPCR method for mtDNA quantitation is possible. qPCR uses commercially available fluorescence-detecting thermocyclers to amplify specific DNA sequences and measure their concentrations simultaneously. The fluorescent reporter can be a nonspecific intercalating double-stranded DNA-binding dye or a sequence-specific fluorescently labeled oligonucleotide probe (Chapter 9). The target sequences are amplified and detected by the same instrument, and the reporter fluorescence is monitored externally. Thus the reaction tubes do not need to be opened. This minimizes aerosol contamination and reduces the risk of false-positive results. A widely used qPCR probe technique is the TaqMan method (Applied Biosystems).

6.3.1.1 TaqMan Method

This method utilizes the 5′ exonuclease activity of *Taq* polymerase to cleave the probe during PCR (also known as the 5′ exonuclease assay). The probe is designed to anneal to the target sequence between the upstream and downstream primers and is added to the PCR mixture together with primers (Figure 6.6). The probe T_m (melting temperature; see Chapter 7) should be higher than the amplification primer T_m. A minor groove binder (MGB), such as dihydrocyclopyrroloindole tripeptide, is often linked at the 3′ end of the probe (Figure 6.6). A conjugated MGB binds to the minor groove of a B-form DNA helix (Figure 6.7), which is stabilized by van

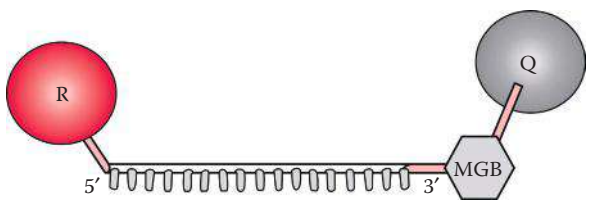

Figure 6.6 TaqMan probe. Each probe is labeled with a reporter dye (R) on the 5′ end and a fluorescence quencher (Q) on the 3′ end. MGB represents a minor groove binder. (© Richard C. Li.)

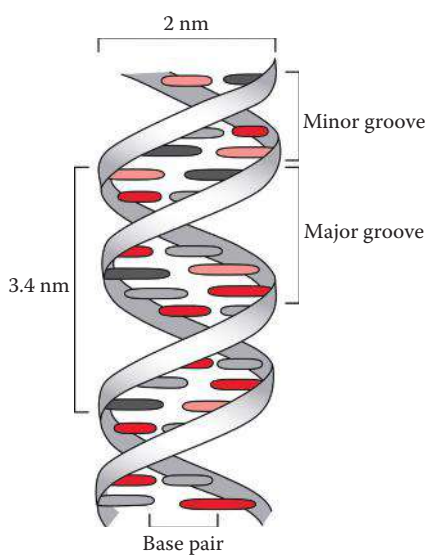

Figure 6.7 DNA double helix. The double-helical structure of B-DNA, the most common form of DNA, is shown. With a helical diameter of 2 nm, each turn of the helix takes 3.4 nm, which corresponds to 10 base pairs per turn. The major and minor grooves are shown. (© Richard C. Li.)

der Waals forces. As a result, conjugating an MGB into a probe increases the T_m values, allowing for the use of shorter probes. As a result, having the probe T_m higher than that of the primers ensures that the probe is fully hybridized during primer extension. The oligonucleotide probe is labeled with both a *reporter* fluorescent dye, usually 6-carboxyfluorescein (6-FAM) or tetrachlorofluorescein (TET) at the 5′ end; and a nonfluorescent *quencher* moiety, such as tetramethylrhodamine (TAMRA), usually at the 3′ end or any thymine position. While the probe is intact, the quencher greatly reduces the fluorescence emitted by the reporter via fluorescent resonance energy transfers (FRET). FRET is a distance-dependent interaction between two molecules in which the excitation energy is transferred from a photon donor molecule (reporter) to an acceptor molecule (quencher) without emission of a photon. During the extension phase of the PCR cycle, the 5′ exonuclease activity of Taq polymerase cleaves the reporter dye from the probe. Because the reporter dye is no longer in close proximity to the quencher, the FRET is disrupted and the probe begins to fluoresce (Figure 6.8). The intensity of fluorescence can be measured (Chapter 9) and is proportional to the amount of target DNA synthesized during the PCR.

In this assay, the rate of accumulation of amplified DNA over the entire course of a PCR is generated. The greater the initial concentration of target templates in a sample, the fewer cycles required to reach a particular quantity of amplified product. The initial concentration of target templates can be expressed using the *cycle threshold* (C_T). C_T is defined as the number of PCR cycles required for the fluorescent signal to cross a threshold of amplification where the signal exceeds background level or baseline noise. A plot of C_T against the $\log_{10}$ of the initial concentration of a set of DNA standards yields a straight line as a standard curve (Figure 6.9). The target sequences in an unknown sample can be quantified by comparing to the standard curve. Additionally, qPCR has the ability to detect PCR inhibitors (Chapter 7) that interact with DNA or with DNA polymerases. The presence of PCR inhibitors in the DNA extracts can be measured by monitoring the amplification of the internal positive control (IPC). Most human DNA quantitation kits contain a known amount of exogenous DNA as IPC that can be fortified to the sample and amplified. Monitoring the amplification of IPC enables the detection of PCR failure due to inhibition when the IPC's C_T value is higher than that of an uninhibited PCR reaction.

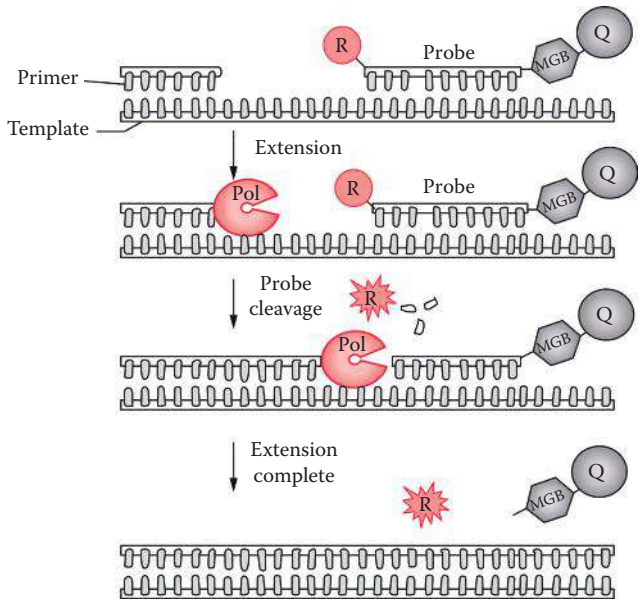

Figure 6.8 Real-time PCR with TaqMan detection. The TaqMan probe is shown. During extension, the 5′ nuclease activity of Taq polymerase (Pol) cleaves the probe. Reporter dye is released during each cycle of PCR. (© Richard C. Li.)

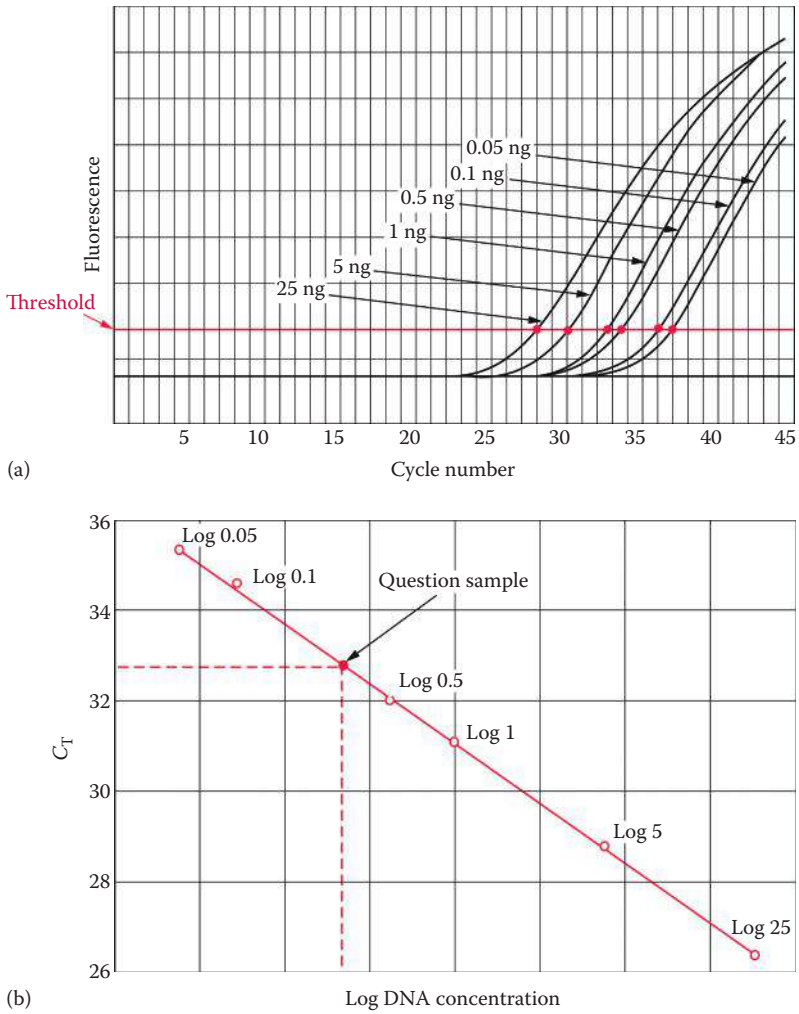

Figure 6.9 Real-time quantitative PCR. (a) Amplification curves for a dilution series of standards with known quantities of DNA. C_T is the cycle threshold at which the amplification curve crosses the threshold, as indicated by the red line. (b) A standard curve based on data obtained from the amplification curves. The quantity of DNA in a questioned sample can be determined from the standard curve. (© Richard C. Li.)

Bibliography

Ahn, S.J., J. Costa, and J.R. Emanuel, PicoGreen quantitation of DNA: Effective evaluation of samples pre- or post-PCR. *Nucleic Acids Res*, 1996, **24**(13): 2623–2625.

Alonso, A. and P. Martin, A real-time PCR protocol to determine the number of amelogenin (X-Y) gene copies from forensic DNA samples. *Methods Mol Biol*, 2005, **297**: 31–44.

Alonso, A., et al., Specific quantification of human genomes from low copy number DNA samples in forensic and ancient DNA studies. *Croat Med J*, 2003, **44**(3): 273–280.

Alonso, A., et al., Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Sci Int*, 2004, **139**(2–3): 141–149.

Andreasson, H. and M. Allen, Rapid quantification and sex determination of forensic evidence materials. *J Forensic Sci*, 2003, **48**(6): 1280–1287.

7

Amplification by Polymerase Chain Reaction

7.1 Denaturation and Renaturation of DNA

The DNA double helix is stabilized by chemical interactions. Base pairing of the two strands involves the formation of hydrogen bonds that provide weak electrostatic attractions between electronegative atoms. An adenine always pairs with a thymine (two hydrogen bonds), and a cytosine pairs with a guanine (three hydrogen bonds) (see Figure 7.1). Base stacking involves hydrophobic interactions between adjacent base pairs and provides stability to the double helix.

Double-stranded DNA is maintained by hydrogen bonding between the bases of complementary pairs. *Denaturation* occurs when the hydrogen bonds of DNA are disrupted, and the strands are separated. A *melting curve* can be obtained from measuring DNA denaturation by slowly heating a solution of DNA. As shown in Figure 7.2, an increasing temperature increases the percentage of the DNA that is denatured. The temperature at which 50% of DNA strands are denatured is defined as the *melting temperature* (T_m). The value of T_m is affected by the salt concentration of the solution, but can also be affected by nucleotide content, high pH, and length of the molecule.

Nucleotide content affects the value of T_m because GC pairs are joined by three hydrogen bonds while the AT pairs are joined by only two. Increasing the GC content of a DNA molecule increases the T_m. Excessively high pH causes the hydrogen bonds to break and the paired strands to separate. Finally, the length of the molecule also affects the T_m simply because a longer molecule of DNA requires more energy to break more bonds than a shorter molecule.

The single strands in a solution of denatured DNA can, under certain conditions, reanneal into double-stranded DNA. The process is called *renaturation* and two requirements must be met for it to occur. First, sufficient amounts of charged molecules, such as salts, must be present in the solution to neutralize the negative charges of the phosphate groups in DNA. This prevents the complementary strands from repelling each other. Additionally, the temperature must be high enough to disrupt hydrogen bonds that formed randomly between the bases of DNA strands. However, excessively elevated temperatures can disrupt the base pairs between the complementary DNA strands.

7.2 Basic Principles of Polymerase Chain Reaction

The *polymerase chain reaction* (*PCR*) allows the exponential amplification of specific sequences of DNA to yield sufficient amplified products, also known as *amplicons*, for various downstream

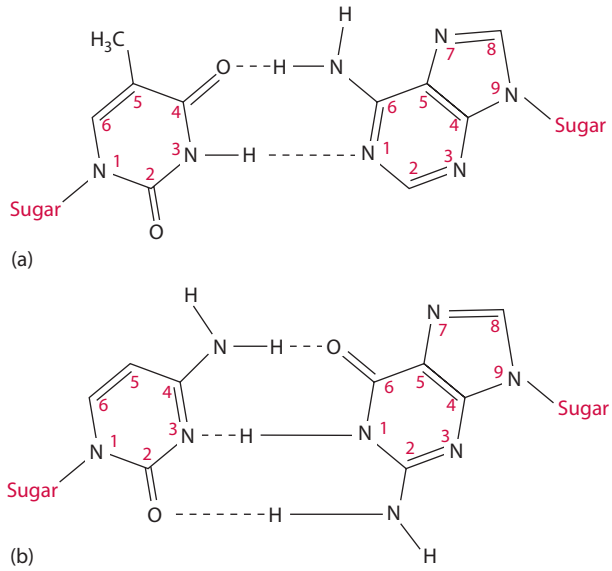

Figure 7.1 Base pairing between two DNA strands. (a) An adenine (right) pairs with a thymine (left). (b) A cytosine (left) pairs with a guanine (right).

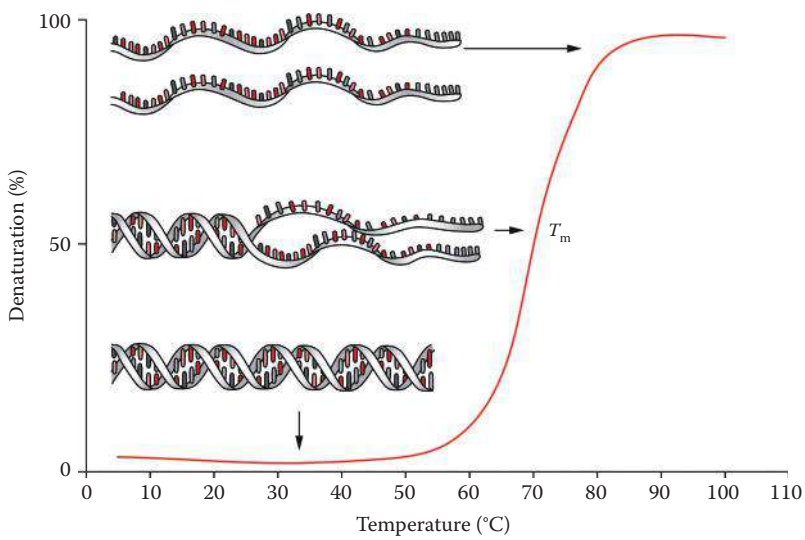

Figure 7.2 Melting curve of DNA. The degree of DNA denaturation is increased by increasing the temperature. T_m and possible shapes of a DNA molecule are shown. (© Richard C. Li.)

applications. The technique is highly sensitive and can amplify very small quantities of DNA. Therefore, it can be utilized for the analysis of samples of limited quantity. PCR-based assays are rapid and robust. Thus, PCR forms the basis of many forensic DNA assays such as DNA quantitation (Chapter 6), short tandem repeat (STR) profiling (Chapter 20), and mitochondrial DNA (mtDNA) sequencing (Chapter 23).

The concept of synthesizing DNA by a cycling process was first proposed in the early 1970s. In the mid-1980s, PCR technology was finally developed by Kary Mullis and his coworkers (Cetus Corporation) to amplify the β-globin gene for the diagnosis of sickle-cell anemia. In the late 1980s, a

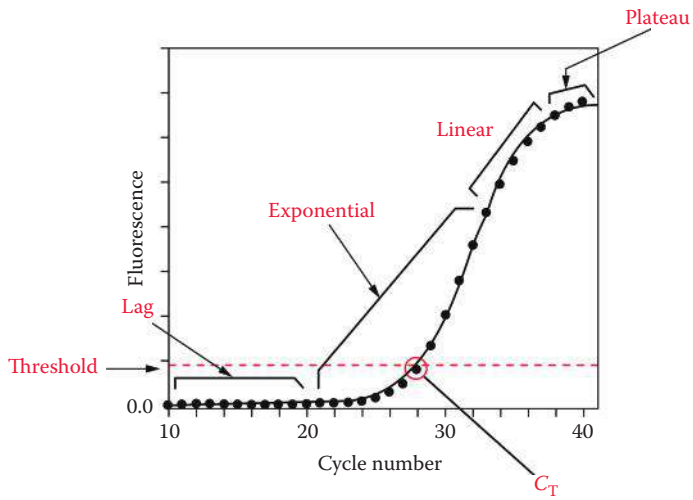

Figure 7.3 PCR amplification curve. A fluorescence signal results from the accumulation of amplicon. The location of the signal threshold is indicated as a dashed line. The C_T value for the curve is the cycle at which the amplification curve crosses the threshold. (© Richard C. Li.)

thermostable polymerase from *Thermus aquaticus* was utilized for PCR. This step greatly increased efficiency and allowed the process to be automated. The result was a powerful impact on molecular biology. In 1993, Mullis was awarded the Nobel Prize for the invention of PCR technology.

In the early 1990s, the technique for the simultaneous amplification and detection of the accumulation of amplicons at each PCR cycle was developed, and the concept of *real-time PCR* was born. This process allows the monitoring of amplicon production at each cycle of the PCR process. The fundamental processes were studied by characterizing the amplification kinetics of PCR using a graph that plotted the amount of amplicon yield at each cycle versus the cycle number (Figure 7.3).

An S-shaped amplification curve is obtained and divided into an exponential phase, a linear phase, and a plateau. During the exponential phase, the amplicon accumulates exponentially. It was revealed that the amplicon accumulation during PCR was correlated to the starting copy number of DNA template. Thus, the amount of amplicon produced during the exponential amplification phase can be used to determine the amount of starting material. This relationship can be further examined using a plot of cycle numbers versus a log scale of the serial dilution of the starting concentration of DNA template, which results in a linear relationship (see Chapter 6). It demonstrates that fewer cycles are needed if larger quantities of starting DNA template are present. The slope of this linear curve (Figure 6.9b) is known as the *amplification efficiency*. The exponential phase continues until one or more of the components (Section 7.3) in the reaction become limited. At this point, the amplification efficiency decreases, the amplicon no longer accumulates exponentially, and PCR enters the linear phase of the curve. At the plateau phase, no more amplicon is accumulated due to the exhaustion of reagents and polymerase.

7.3 Essential PCR Components

A PCR reaction requires thermostable DNA polymerases, primers, and other components, as described below (Figure 7.4).

7.3.1 Thermostable DNA Polymerases

A wide variety of DNA polymerases are available. They vary in fidelity, efficiency, and ability to synthesize longer DNA fragments. Nonetheless, *Taq* polymerase is the most commonly used

Figure 7.4 A liquid handling workstation for automated PCR assay setup. (© Richard C. Li.)

enzyme for routine PCR applications (0.5–5 units per reaction). Currently, AmpliTaq Gold™ DNA polymerase (Applied Biosystems) is the most common DNA polymerase for forensic applications.

PCR reactions are usually set up at room temperature. Nonspecific annealing between primers and template DNA can occur, resulting in the formation of nonspecific amplicons. Additionally, annealing between the primers can occur to form primer dimers. Nonspecific annealing interferes with PCR amplification by reducing the amplification efficiency of the specific sequences of interest.

Such interference can be minimized by a *hot-start PCR* approach. The AmpliTaq Gold™ DNA polymerase, a modified enzyme, remains in an inactive form until activated with a pH below 7 prior to the PCR cycling in which the inhibitory motif is inactivated. The pH of the buffer system used in the PCR reaction is temperature sensitive; increasing the temperature decreases the pH of the solution. Thus, the activation of the enzyme can actually be carried out by a heating step at 95°C prior to the start of the cycling. During the heating process, the DNA strands also denature, which can prevent the formation of nonspecific PCR products.

7.3.2 PCR Primers

PCR primers are the oligonucleotides that are complementary to the sequences that flank the target region of the template. A pair (forward and reverse) of primers (typically 0.1–1 μM) is required. Properly designed primers are critical to the success of a PCR reaction. Computer software such as Primer3 is available to assist and optimize the designing of primers.

A primer must be specific to the target sequence; otherwise, nonspecific products that might interfere with the proper interpretation of a DNA profile might be produced. The primers within a pair should have similar T_m values. The estimated T_m values of a primer pair should not differ by more than 5°C. The T_m of an oligonucleotide primer can be predicted and calculated using the following equation:

$$T_m = 81.5°\text{C} + 16.6\left(\log_{10}[K^+]\right) + 0.41\left(\%\ [\text{G}+\text{C}]\right) - \left(\frac{675}{n}\right)$$

where:

$[K^+]$ = concentration of the potassium ion
$[\text{G}+\text{C}]$ = GC content (%) of the oligonucleotide
n = number of bases in the oligonucleotide

This equation shows that the T_m can be affected by primer base composition (GC content) and primer length. The GC content of an oligonucleotide primer should be 40%–60%, and the length of an oligonucleotide primer should be 15–25 base pairs, although longer primers can be used.

A primer should not contain self-complementary sequences that may form hairpin structures interfering with the annealing of primers and the template. Additionally, the primers in a pair should not share similar sequences, to avoid the primers from annealing to each other. These annealed primers may then be amplified during PCR, creating products known as *primer dimers*, which compete with the target DNA template for PCR components.

Multiplex PCR is the simultaneous amplification of more than one region of a DNA template in a single reaction to achieve high-throughput analysis. Multiplex PCR consists of multiple primer sets in a single reaction to produce amplicons of multiple target DNA regions. The primers should be designed to yield proper sizes of amplicons to be resolvable in downstream separation and detection procedures such as electrophoresis. To prevent preferential amplification of one target sequence over another, the annealing temperatures should be similar among multiplex PCR primer pairs. Additionally, the primers should lead to similar amplification efficiencies among the loci to be tested. Forensic applications of multiplex PCR, such as autosomal and Y-chromosomal short tandem repeat analysis, are discussed in Chapters 20 and 21.

7.3.3 Other Components

Essential components include template DNA with target sequences in either linear form (nuclear genomic DNA) or circular form (mitochondrial DNA). Both single- and double-stranded DNA can be used as a template for PCR. Typically, 1–2.5 ng of template DNA is utilized for forensic applications using PCR.

Deoxynucleoside triphosphates (dNTPs) are the substrates for DNA synthesis. A PCR assay usually contains equal molar amounts (typically, 200 μM) of dATP, dCTP, dGTP, and dTTP.

Divalent cations, such as Mg^{2+}, are required for the enzymatic activity of DNA polymerases. A PCR assay usually contains 1.5–2.5 mM Mg^{2+}. Monovalent cations, such as K^+ (50 mM), are usually recommended, and a buffer is often utilized to maintain pH between 8.3 and 8.8 at room temperature.

Controls should be used to monitor the effectiveness of PCR amplification. A *positive control* shows that PCR components such as reagents and PCR cycle parameters are working properly during a PCR. A standard DNA template should be used as a positive control and amplified with the same PCR components used on the rest of the samples. The amplification negative control and extraction reagent blank are discussed in Section 7.5.4.

7.4 Cycle Parameters

PCR cycling protocols may vary according to the type of analysis. Figure 7.5 shows representative PCR cycling conditions commonly used by forensic DNA laboratories. PCR utilizes a number of cycles for the replication of a specific region of a DNA template. During each cycle, a copy of the target DNA sequence is synthesized. A PCR cycle consists of three elements: denaturation, annealing, and extension. Precise and accurate temperatures at denaturation, annealing, and extension are critical to achieve a successful amplification. At the beginning of each cycle, the two complementary DNA template strands are separated at high temperatures (94°C–95°C) in a process called *denaturation*. The temperature is then decreased to allow *annealing* between the oligonucleotide primers and the template. The temperature for annealing is usually 3°C–5°C lower than the T_m of the oligonucleotide primer.

The annealing temperature is critical. If it is too high, a very low quantity of amplicon is yielded because of the failure of annealing between the primer and the template. If the annealing temperature is too low, nonspecific amplification can occur. Next, optimal temperature for DNA polymerase is reached, thus allowing for DNA replication (*extension*). By the end of each cycle, the copy number of the amplicon is nearly doubled (Figure 7.6).

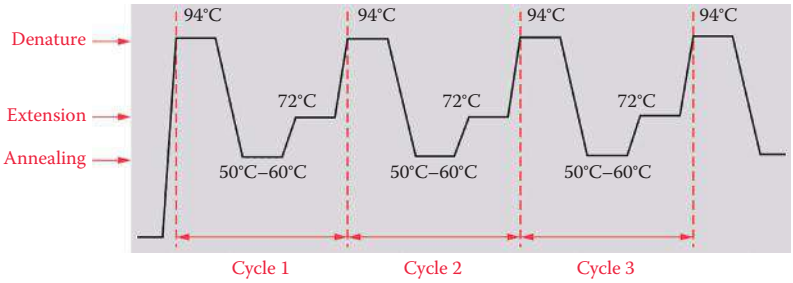

Figure 7.5 Temperature parameters during thermal cycling of the PCR process. The first three cycles are shown. (© Richard C. Li.)

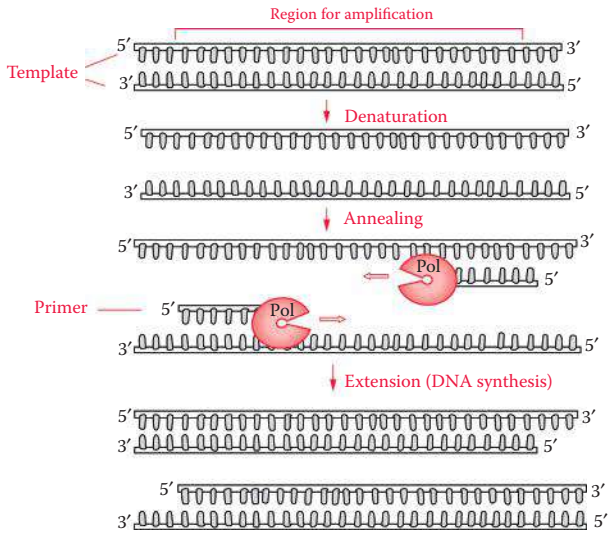

Figure 7.6 First cycle of PCR. The amplified region is determined by the positions of the primers. The direction (5′ to 3′) of DNA synthesis is indicated by arrows. Pol represents *Taq* polymerase. (© Richard C. Li.)

The number of cycles needed for PCR depends primarily on the number of copies of starting DNA template. The relationship can be expressed as the following equation:

$$N_x = N_0(1+E)^x$$

where:

X = number of PCR cycles
N_x = copy number of the amplicon after x cycles of PCR
N_0 = initial copy number of the template
E = amplification efficiency of the *Taq* polymerase

For example, in a 28-cycle PCR amplification, the DNA template can theoretically be amplified by a factor of approximately 10^8. If the cycle number is increased to 34, a factor of 10^{10} can theoretically be reached.

PCR amplification can be carried out using an instrument known as a *thermal cycler* (Figure 7.7). Various types of thermal cyclers differ in the number of samples they can process, the sizes of the sample tubes, and temperature control features.

Figure 7.7 Thermal cyclers. A thermal cycler can provide rapid temperature changes as desired to carry out PCR. A PCR thermal cycler (left) and a real-time PCR instrument (right). (© Richard C. Li.)

7.5 Factors Affecting PCR

7.5.1 Template Quality

It is important to prevent degradation of the DNA during the collection and processing of evidence. Degradation causes DNA to break into smaller fragments. If the damage occurs at a region to be amplified, the result can be failure in PCR amplification. In a degraded sample, the longer the amplicon length, the higher the risk of failure in PCR. Low copy number (LCN) of DNA template is often encountered in forensic samples. When amplifying very low levels (approximately 100 pg of DNA) of a template, the following phenomenon is often observed: one of the two alleles fails to be detected from a heterozygote and can falsely be identified as a homozygote. This phenomenon is also known as the *stochastic effect* in which the two alleles in a heterozygous individual are unequally detected at a low level of starting DNA template. Approaches such as increasing the cycle number, from 28 to 34 (Section 7.4), have been introduced to address the LCN problem.

7.5.2 Inhibitors

Inhibitors, if present, can interact with the DNA template or polymerase, causing PCR amplification failure. The presence of PCR inhibitors can be detected using an internal positive control (Section 6.3.1.1). A number of PCR inhibitors commonly encountered in evidence samples include heme molecules from blood, indigo dyes from fabrics, and melanin from hair samples. Thus, it is important to remove PCR inhibitors during DNA extraction. If PCR inhibitors are not eliminated during the extraction process, additional procedures such as the use of centrifugal filtration devices can be used. Centrifugal filtration devices can separate molecules by size. After the centrifugation step, small molecular weight inhibitors are filtered by passing through the membrane and are discarded. Alternatively, increasing the amount of DNA polymerase or adding bovine serum albumin (BSA) in the reaction can overcome the inhibition effects.

7.5.3 Contamination

PCR is a highly sensitive method; therefore, procedures that minimize the risk of contamination are necessary. To prevent contamination, pre- and post-PCR samples should be processed in separate areas or at different times. Additionally, reagents, supplies, and equipment used for pre- and post-PCR steps should be separated as well. Protective gear should include laboratory coats and disposable gloves. Facial masks and hair caps may be used if necessary. Aerosol-resistant pipet tips and DNA-free solutions and test tubes should also be used.

The levels of contamination must be monitored using controls. Extraction *reagent blanks*, which contain all extraction reagents but no sample, monitor contamination from extraction to PCR. Contamination detected in an extraction reagent blank but not in an amplification-negative control indicates that the reagents used for extraction are contaminated. Amplification-*negative*

controls, which contain all PCR reagents and no DNA template, monitor contamination of the amplification. Contamination observed in an amplification-negative control but not in an extraction reagent blank indicates that the contamination occurred during the amplification step. Contamination observed in both an amplification reagent blank and an extraction-negative control indicates that amplification reagents are contaminated. A collection of DNA profiles of each member of a laboratory should be readily available for comparisons. Sources of laboratory contamination can be identified by comparing results with an analyst's DNA profile.

7.6 Reverse Transcriptase PCR for RNA-Based Assays

The pathway for the flow of genetic information is called the *central dogma*—a term coined by Francis Crick in 1956. According to the central dogma, parental DNA serves as the template for *DNA replication*. With RNA synthesis or *transcription*, the process is carried out using the DNA as a template. Conversely, RNA chains can be used as templates for the synthesis of a DNA strand of complementary sequence, in which the end product is referred to as *complementary DNA (cDNA)*. Protein synthesis, also known as *translation*, is directed by an RNA template (Figure 7.8).

The flow of genetic information from RNA to DNA is referred to as *reverse transcription*. It was discovered independently by David Baltimore and Howard Temin in 1970, and they shared the Nobel Prize for their work. Reverse transcription is carried out by a *reverse transcriptase* that forms the basis of reverse transcriptase PCR as described below.

Reverse transcriptase PCR (RT-PCR) is highly sensitive and can be used to detect very small quantities of mRNA. It can be utilized to measure levels of gene expression even when the RNA of interest is expressed at very low levels. Detecting mRNAs of tissue-specific genes can be utilized for bodily fluid identification in forensic investigations (Chapter 11). During an RT-PCR process, a single-stranded cDNA is synthesized from a template mRNA using reverse transcription. The cDNA is then amplified by PCR for detection and analysis.

7.6.1 Reverse Transcription

The synthesis of single-stranded cDNA from an mRNA template is catalyzed by reverse transcriptase. Reverse transcriptases share many features in structure and function with DNA polymerases. The catalytic function of cDNA synthesis requires a primer that anneals to a complementary mRNA template. The primer can be either RNA or DNA; however, DNA primers are more efficient than RNA primers. During the elongation of the primer, reverse transcriptase incorporates the corresponding deoxyribonucleotide triphosphate according to the rules of base pairing with the RNA template. The RNA template is then degraded by an intrinsic RNase H activity of reverse transcriptase during the reverse transcription reaction. The retroviral RNase H is a domain of the viral reverse transcriptase enzyme. It is a nonspecific endonuclease that cleaves RNA.

Several reverse transcriptases derived from retroviruses can be used to generate cDNA from an RNA template. The most common reverse transcriptases used for cDNA synthesis are

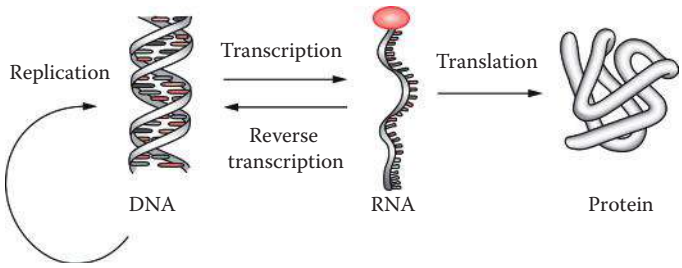

Figure 7.8 Pathway for the flow of genetic information. (© Richard C. Li.)

encoded by the *pol* gene from the avian myeloblastosis virus (AMV) and the Moloney strain of the murine leukemia virus (MMLV). These enzymes are genetically engineered reverse transcriptases but lack RNase H activity. Genetically engineered reverse transcriptases that are stable at higher temperatures (up to 60°C) are also produced. Increased thermal stability of reverse transcriptase allows the reverse transcription to be carried out at a higher than ambient temperature, which eliminates RNA secondary structures and improves the specificity of the reaction and the yields of the synthesis of full-length cDNA.

7.6.2 Oligodeoxynucleotide Priming

Oligodeoxynucleotide primers are an essential component for a reverse transcriptase reaction. Different priming strategies can be utilized to synthesize cDNA from a particular target mRNA or from all mRNA in a sample. *Gene-specific primers* are designed to hybridize to a particular mRNA sequence for the conversion of a specific gene sequence into cDNA. *Universal primers* can hybridize to any mRNA sequence in a sample to convert all mRNAs to cDNA. Two types of universal primers can be used: oligo (dT) and random hexamer primers. An *oligo (dT) primer* can hybridize to the 3′ termini poly (A) tails of eukaryotic mRNAs. Reverse transcriptases with oligo (dT) only synthesize cDNA from transcribed genes. *Random hexamer primers* are another type of universal primer. Random hexamers are nonspecific primers that can hybridize, at multiple sites, to any RNA sequence including non-mRNA templates such as ribosomal RNA.

7.6.3 Reverse Transcriptase PCR

During an RT-PCR process, a single-stranded cDNA is synthesized from a template mRNA using reverse transcription. The cDNA is then amplified by PCR with a pair of oligonucleotide primers corresponding to a specific sequence in the cDNA (Figure 7.9).

Two strategies of RT-PCR exist: a one-step and a two-step RT-PCR (Figure 7.10). One-step RT-PCR combines the reverse transcription reaction and PCR in a single tube. Only gene-specific

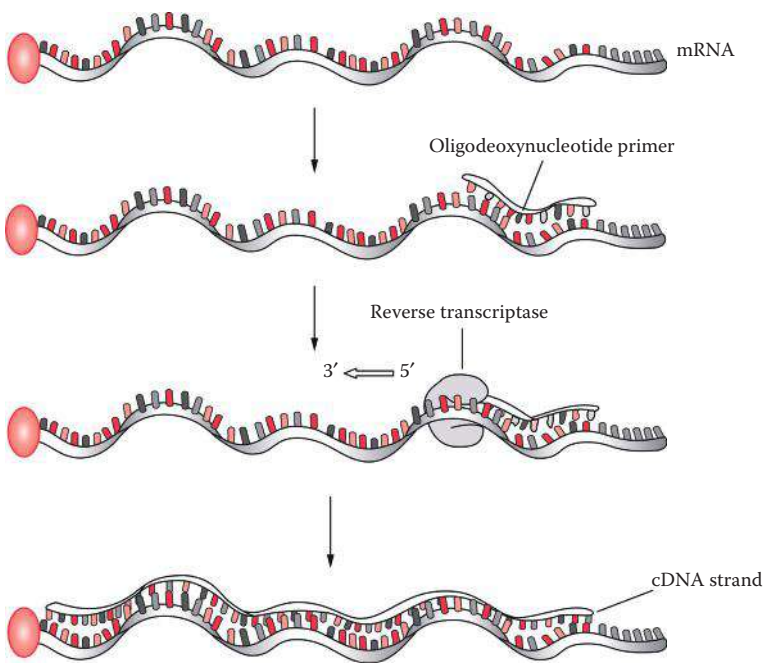

Figure 7.9 cDNA synthesis. The synthesis of a DNA strand transcribed from mRNA can be carried out using a primer and reverse transcriptase. A gene-specific primer that amplifies a specific target sequence is shown. (© Richard C. Li.)

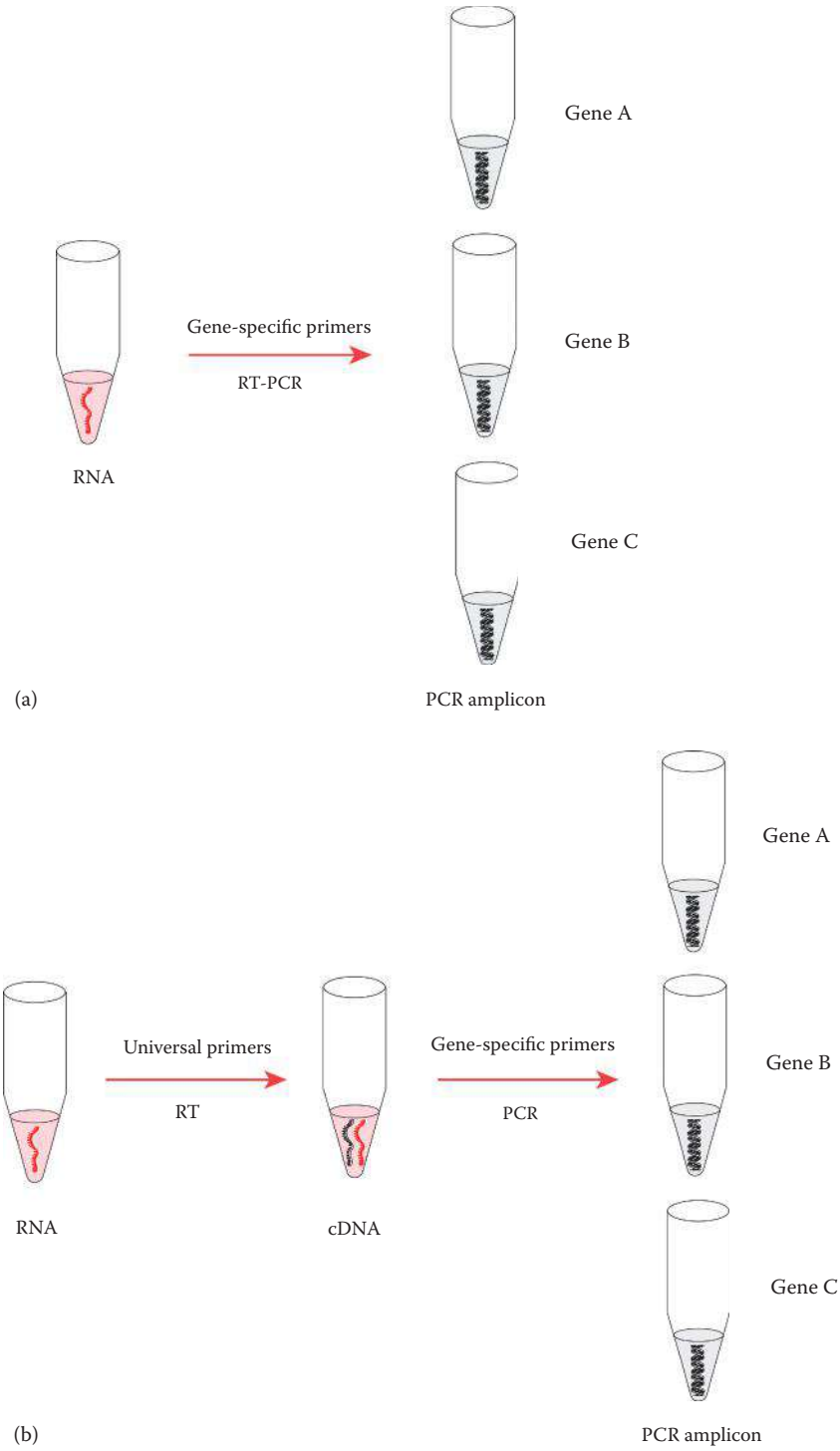

Figure 7.10 RT-PCR strategies. (a) One-step RT-PCR and (b) two-step RT-PCR. (© Richard C. Li.)

primers can be used for the reverse transcription reaction and PCR. RNAs from either total RNA or mRNA can be used. This approach simplifies the reaction setup as it is useful for processing large numbers of samples. The tubes are not opened between reverse transcription and PCR, thus minimizing the risk of pipetting errors and carryover contamination.

During a two-step PCR, the reverse transcription reaction and PCR are carried out in separate tubes. Following the reverse transcription reaction, the cDNA is transferred to a separate tube for the PCR amplification. Either oligo (dT) or random hexamer primers can be used for the reverse transcription reaction. Total RNA or polyadenylated RNA can be used for the reverse transcriptase reaction. This allows for the ability to convert all the messages in an RNA sample into cDNA. The two-step RT-PCR is useful for analyzing multiple mRNAs from a single sample.

Two types of PCR methods can be utilized for analyzing amplified products. The end-point PCR method measures the amount of amplified product synthesized during PCR at the end of the PCR amplification. The detection of the amplified product indicates the presence of the mRNA of interest. With the real-time PCR method, the amplified product is quantified during the exponential phase of PCR. The hot-start PCR strategy is typically used to increase sensitivity, specificity, and yield. Usually, the hot-start strategy is carried out at a high temperature, for example 94°C, prior to PCR cycling. Under this condition, *Taq* DNA polymerase is activated. Additionally, the RNA-cDNA hybrid is denatured, and reverse transcriptase is inactivated at 94°C. Multiplex RT-PCR assays have been developed that can detect multiple bodily fluid mRNAs from single or mixed stains (Chapter 11). The commonly used RT-PCR for forensic identification of bodily fluids is the two-step RT-PCR. The end-point PCR is usually utilized with the forward primers labeled with florescent dyes at the 5′ end. The amplified products are separated and detected in a standard capillary electrophoresis instrument.

Bibliography

Aboud, M., H.H. Oh, and B. McCord, Rapid direct PCR for forensic genotyping in under 25 min. *Electrophoresis*, 2013, **34**(11): 1539–1547.

Afonina, I.A., et al., Minor groove binder-conjugated DNA probes for quantitative DNA detection by hybridization-triggered fluorescence. *Biotechniques*, 2002, **32**(4): 940–944, 946–949.

Alaeddini, R., Forensic implications of PCR inhibition—A review. *Forensic Sci Int Genet*, 2012, **6**(3): 297–305.

Ballantyne, K.N., R.A. van Oorschot, and R.J. Mitchell, Comparison of two whole genome amplification methods for STR genotyping of LCN and degraded DNA samples. *Forensic Sci Int*, 2007, **166**(1): 35–41.

Ballantyne, K.N., R.A. van Oorschot, and R.J. Mitchell, Increased amplification success from forensic samples with locked nucleic acids. *Forensic Sci Int Genet*, 2011, **5**(4): 276–280.

Ballantyne, K.N., R.A. van Oorschot, and R.J. Mitchell, Locked nucleic acids in PCR primers increase sensitivity and performance. *Genomics*, 2008, **91**(3): 301–305.

Ballantyne, K.N., et al., Decreasing amplification bias associated with multiple displacement amplification and short tandem repeat genotyping. *Anal Biochem*, 2007, **368**(2): 222–229.

Bassam, B.J., et al., Nucleic acid sequence detection systems: Revolutionary automation for monitoring and reporting PCR products. *Aust Biotechnol*, 1996, **6**: 285–294.

Bekaert, B., et al., Automating a combined composite-consensus method to generate DNA profiles from low and high template mixture samples. *Forensic Sci Int Genet*, 2012, **6**(5): 588–593.

Birch, D.E., et al., Simplified hot start PCR. *Nature*, 1996, **318**: 2.

Bloch, W., A biochemical perspective of the polymerase chain reaction. *Biochemistry*, 1991, **30**(11): 2735–2747.

Bright, J.A., et al., The effect of cleaning agents on the ability to obtain DNA profiles using the Identifiler and PowerPlex® Y multiplex kits. *J Forensic Sci*, 2011, **56**(1): 181–185.

Brotherton, P., et al., Preferential access to genetic information from endogenous hominin ancient DNA and accurate quantitative SNP-typing via SPEX. *Nucleic Acids Res*, 2010, **38**(2): e7.

Budowle, B., A.J. Eisenberg, and A. van Daal, Validity of low copy number typing and applications to forensic science. *Croat Med J*, 2009, **50**(3): 207–217.

8

DNA Electrophoresis

It is necessary to separate various sizes of DNA fragments so that the DNA fragment in question can be identified and analyzed. This can be achieved by electrophoresis, a process in which fragments are separated based on the migration of charged macromolecules in an electric field.

8.1 Basic Principles

DNA is a negatively charged molecule in an aqueous environment, with the phosphate groups of DNA nucleotides carrying negative charges. DNA molecules migrate from the negative electrode (cathode) toward the positive electrode (anode) in an electric field during electrophoresis. The *electrical potential*, a measure of the work required to move a charged molecule in an electric field, is the force responsible for moving the charged macromolecules during electrophoresis.

The *electrophoretic mobility* of macromolecules is primarily determined by their *charge-to-mass ratios* and their shapes. However, because the phosphate group of every DNA nucleotide carries a negative charge, the charge-to-mass ratio of DNA molecules is almost the same even if the length of the DNA fragment varies. Additionally, forensic DNA testing usually analyzes linear DNA molecules such as double-stranded linear DNA for variable number tandem repeat (VNTR) analysis (Chapter 19) and single-stranded DNA for short tandem repeat (STR) analysis (Chapter 20). The shapes of linear DNA molecules are similar. Therefore, the electrophoretic separation of different-sized DNA fragments, through a series of pores of the supporting matrix in which the fragments travel, is based more on their sizes than their shapes.

It is obviously easier for smaller molecules to migrate through the pores of a supporting matrix than larger molecules; this is why smaller molecules migrate faster through the matrix. Hence, the electrophoretic mobility increases as the size of the DNA molecule decreases. Conversely, larger DNA molecules migrate much more slowly because they experience more friction and collisions as they travel through the net of pores in a matrix. Therefore, the separation of DNA molecules with different sizes can be accomplished. Figure 8.1 depicts the models for DNA electrophoresis.

8.2 Supporting Matrices

Although the actual separation occurs in an aqueous phase, most variants of electrophoresis use a physical support material also called a *matrix*. As discussed in Section 8.1, the matrix can be used as a molecular sieve for the separation of DNA molecules, and it also reduces diffusion and convection during electrophoresis. Agarose and polyacrylamide are commonly used electrophoresis matrices because of their good reproducibility, reliability, and versatility.

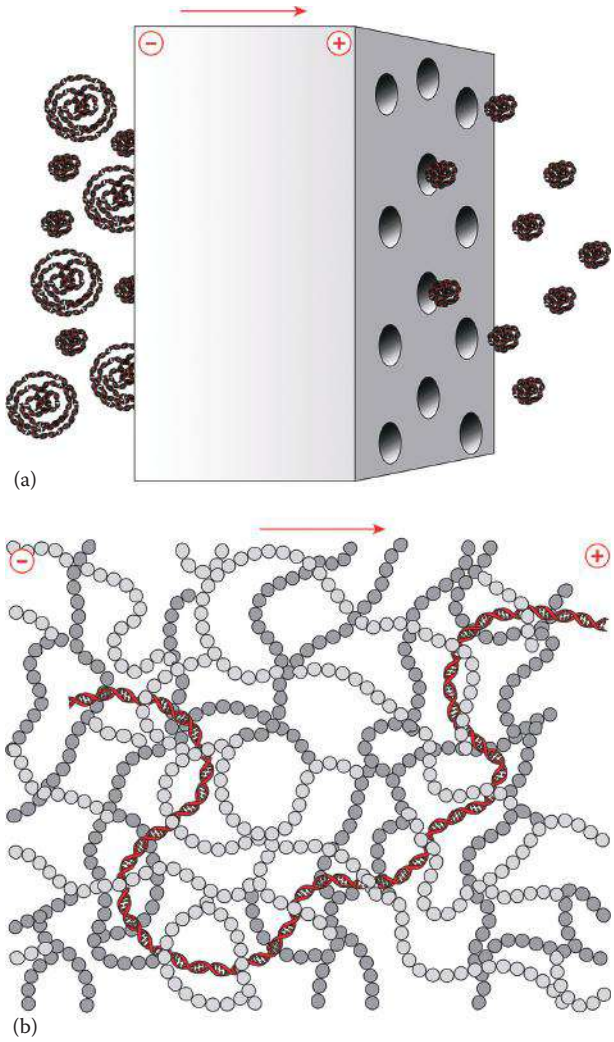

Figure 8.1 Schematic illustrations showing some of the nucleic acid separation mechanisms. (a) Ogston sieving model for the separation of DNA fragments through electrophoresis. Based on this model, the matrix used for gel electrophoresis consists of randomly distributed pores. DNA molecules are considered to behave as globular objects. During the migration, smaller DNA molecules migrate faster than their larger counterparts through the matrix. However, this model does not apply to DNA molecules with a radius much larger than the pore size of the matrix. (b) A reptation model presents an alternative mechanism of gel electrophoretic separation of DNA molecules. Based on this model, DNA molecules are flexible and can migrate through the pores of the gel matrix through a reptation process that is driven by an applied electric field. This model can also apply to capillary electrophoresis separation of DNA molecules in entangled polymer solutions. +, anode; –, cathode; arrow, direction of migration. (© Richard C. Li.)

8.2.1 Agarose

Agarose is a linear polymer composed of alternating residues of D- and L-galactose (Figure 8.2). A gelatinized agarose forms a three-dimensional sieve with pores from 50 to 200 nm in diameter (Figure 8.3). DNA fragments ranging from 50 to 20,000 base pairs in size are best resolved in agarose gels.

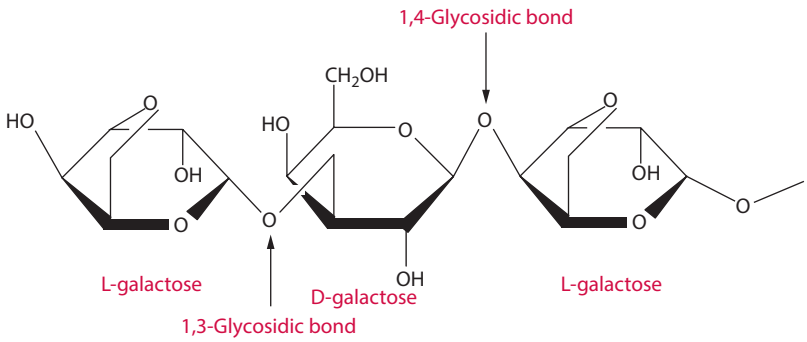

Figure 8.2 Chemical components of agarose, a linear polysaccharide composed of alternating units of D- and L-galactose. Often, the L-galactose residue has an anhydro bridge between the 3 and 6 positions and is called 3,6-anhydro-L-galactose.

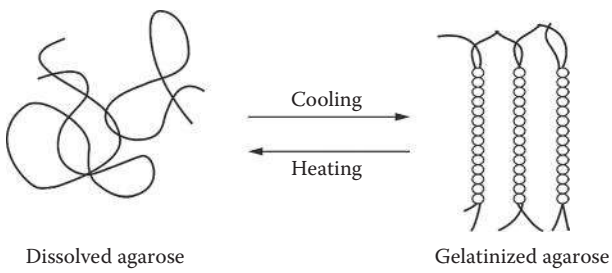

Figure 8.3 Agarose chains. Gelatinized agarose consists of a three-dimensional network in which chains of agarose form helical fibers that aggregate into supercoiled structures. (© Richard C. Li.)

The electrophoretic mobility of double-stranded DNA through agarose gel matrices is inversely proportional to the $\log_{10}$ of the size of DNA fragments ranging from 50 to 20,000 base pairs. The electrophoretic mobility of a linear DNA fragment is also inversely proportional to the concentration of agarose in the gel. A low concentration of agarose in a gel forms larger pores, allowing the separation of larger DNA fragments.

8.2.2 Polyacrylamide

A polyacrylamide gel matrix is very effective for the separation of smaller fragments of DNA (5–500 base pairs). Additionally, a single nucleotide difference in the length of a DNA fragment can be resolved with this type of gel. Polyacrylamide produces much smaller pore sizes than agarose gels and thus has a much higher resolving power than agarose gels for low-molecular-weight DNA molecules. A polyacrylamide gel matrix is formed by polymerization and cross-linking reactions.

8.2.2.1 Polymerization Reaction

Long linear chains of polyacrylamide are polymerized from acrylamide monomers (Figure 8.4). This polymerization reaction is initiated in the presence of free radicals that are generated from the reduction of ammonium persulfate (APS) by *N*,*N*,*N*′,*N*′-tetramethylethylene diamine (TEMED). See Figure 8.5.

8.2.2.2 Cross-Linking Reaction

As shown in Figure 8.5, three-dimensionally cross-linked polyacrylamide chains can be formed with the use of cross-linking agents, such as *N*,*N*′-methylenebisacrylamide (BIS; Figure 8.4).

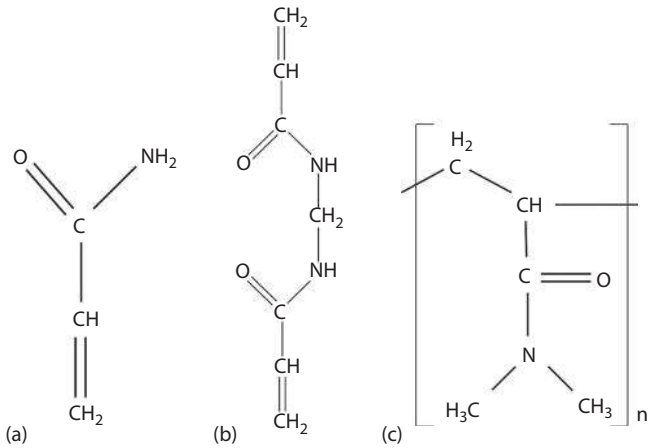

Figure 8.4 Chemical structures of (a) acrylamide, (b) *N*,*N*′-methylenebisacrylamide (BIS), and (c) polydimethylacrylamide (a linear polymer for capillary electrophoresis).

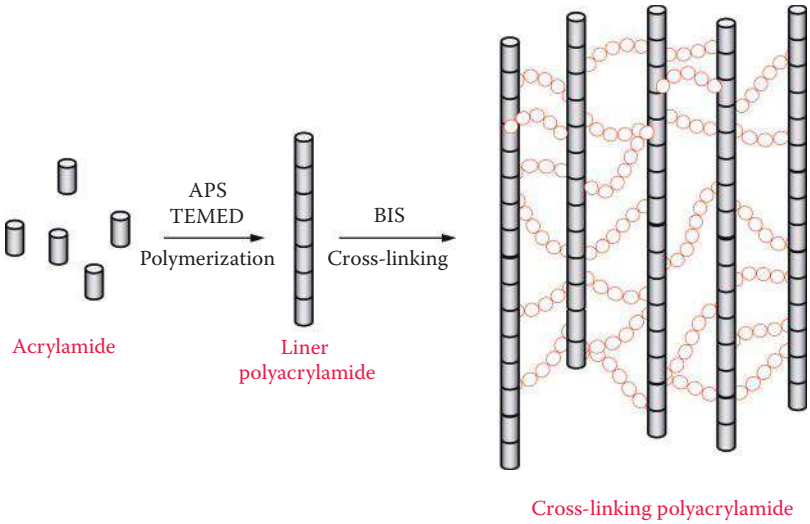

Figure 8.5 Formation of polyacrylamide gel. Polymerization is followed by cross-linking in which the pore size of the gel is determined by its degree of polymerization and cross-linking. (© Richard C. Li.)

The porosity of the resulting gel is determined by the lengths of these chains and the degree of cross-linking between them. Therefore, the sizes of the pores that are formed can be adjusted by altering the concentrations of the acrylamide monomer and the cross-linking reagent.

Polyacrylamide can also be used in a capillary electrophoresis matrix. It is very difficult to insert a cross-linked polyacrylamide matrix into a capillary. Therefore, a solution of a linear polymer (non-cross-linked) of polydimethylacrylamide (Figure 8.4) is used as a matrix for capillary electrophoresis. For instance, POP-4 (4% un-cross-linked polydimethylacrylamide polymer) is used for fragment analysis, such as forensic STR analysis (Chapter 18), in which fluorescently labeled DNA amplicons are separated using electrophoresis and are identified based on their sizes by comparing them with a size standard. POP-6 (6% un-cross-linked polydimethylacrylamide polymer) is used for sequencing, such as mitochondrial DNA (mtDNA) sequencing (Chapter 21). These un-cross-linked polymers are commercially available (Applied Biosystems).

8.2.2.3 Denaturing Polyacrylamide Electrophoresis

Polyacrylamide-based electrophoresis can be carried out with both double- and single-stranded DNA. Electrophoresis that is performed under the conditions of single-stranded DNA is called *denaturing electrophoresis*. Denatured DNA migrates through the gel as linear molecules at a rate independent of the base composition and sequence. With short single-stranded fragments, molecules differing in size by a single nucleotide can be separated. This extraordinary sensitivity to size is extremely useful for the separation of small DNA fragments that are used for DNA sequencing and STR analysis in forensic applications. The denaturing condition can be best achieved by adding chemicals such as urea or formamide to the matrix or by increasing the temperature or pH during electrophoresis.

8.3 Apparatus and Forensic Applications

Agarose electrophoresis is carried out in a slab gel. Agarose gel can be prepared in a variety of shapes and sizes and can be run in different configurations. The most common gel for forensic use is the horizontal slab. Polyacrylamide can be used either in a slab gel or in a capillary electrophoretic apparatus.

8.3.1 Slab Gel Electrophoresis

8.3.1.1 Agarose Gel Electrophoresis

One forensic application of agarose gel electrophoresis is restriction fragment length polymorphism (RFLP) analysis of VNTR loci (see Chapter 19). An agarose gel is used to separate the DNA fragments by size ranging from 500 to 20,000 base pairs of commonly used VNTR loci for forensic testing. This type of electrophoresis is done under nondenatured conditions (double-stranded DNA).

Figure 8.6 shows an example of the apparatus used for gel electrophoresis of DNA. A slab (3–4 mm) of agarose gel is prepared by allowing liquid agarose to gelatinize in a cast. The gel is submerged in a buffer tank filled with an electrophoretic buffer with proper ionic strength, which is necessary to achieve efficient electrical conductance. The gel contains small wells for loading samples. The DNA samples are mixed with a gel loading buffer prior to loading them into the gel. The gel loading buffer contains dyes that add color to the sample to facilitate the process of loading. Dyes such as bromophenol blue and xylene cyanol FF migrate toward the anode during electrophoresis. In most forensic applications, these dyes migrate faster than DNA fragments. The dyes are visible and can thus be used to track the progression of electrophoresis. The electrophoresis can be stopped as the dye front reaches the bottom of the gel (anode side).

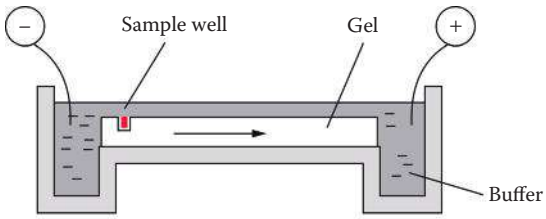

Figure 8.6 Horizontal agarose slab gel apparatus. The gel can be prepared by heating an agarose suspension to dissolve it. The agarose solution can then be poured into a gel cast and allowed to cool until gelatinized. An electrophoresis buffer is poured over the gel to submerge it. The samples, usually mixed with loading buffer containing dye to allow visualization, are loaded into the wells of the submerged gel using a pipet. The electrodes are indicated. (© Richard C. Li.)

The samples are loaded into the sample wells situated proximal to the cathode. An electric field is applied and the negatively charged DNA molecules migrate through the agarose toward the anode. Slab gel electrophoresis is capable of running multiple samples simultaneously. The sample is made visible by using a fluorescent intercalating reagent that can make DNA fluoresce under ultraviolet (UV) light (Chapter 9), either using the fluorescent intercalating reagent incorporated into the gel or staining the gel after electrophoresis. The separated DNA samples appear as bands and the sizes of the DNA fragments can be estimated by comparing them with the sizes of standards run concurrently.

8.3.1.2 Polyacrylamide Gel Electrophoresis

The forensic application of this apparatus is the separation of STR fragments and DNA sequencing reaction products of mtDNA. The sizes of DNA fragments that can be separated range from 100 to 500 base pairs—much smaller than what can be separated efficiently with agarose gels. Single-nucleotide resolution to distinguish similarly sized fragments can be achieved with this technique under denatured conditions with only single-stranded DNA. Polyacrylamide thin slab gels (0.75–1.5 mm) are usually used. Samples are loaded into wells flanked by two pieces of glass. Therefore, polyacrylamide gels are usually run in a vertical configuration (Figure 8.7). The detection of DNA bands in polyacrylamide gels will be described in Chapter 9. The sizes of DNA fragments can be calculated by including an internal size standard. As with agarose gel, polyacrylamide gel electrophoresis is capable of running multiple samples simultaneously and high throughput (a measurement of the rate that a sample is processed by a given analysis) can be achieved. However, caution should be taken to ensure that cross-contamination does not occur from the spilling of samples from adjacent wells. Additionally, polyacrylamide gels are more difficult to prepare than agarose gels.

8.3.2 Capillary Electrophoresis

Capillary electrophoresis (Figures 8.8 through 8.11) is a newer method than the slab gel method, and can be utilized to separate charged macromolecules, such as DNA, RNA, polysaccharides, and proteins.

One essential component of the capillary electrophoresis instrument is the capillary, a thin hollow tube made of fused silica (which typically has a diameter between 50 and 100 μm and a length between 10 and 50 cm). The capillary contains a translucent detection window for the instrument to detect signals from the labeled DNA fragments during electrophoresis.

Linear polydimethylacrylamide is used as the matrix. Capillary electrophoresis is conducted under denatured conditions for forensic applications such as STR analysis and mtDNA sequencing analysis. The denatured condition is achieved by including urea in the electrophoresis matrix and formamide during sample preparation. The injection of samples into the capillary is performed by an autosampler using an *electrokinetic* mechanism (an injection based on the charge of molecules). Only small quantities of sample are required for each injection, and any remaining sample can be saved in case an analysis needs to be repeated.

During electrophoresis, the capillary is connected to buffer reservoirs that are connected to electrodes. The efficient heat dissipation property of thin capillaries allows the separation to be performed at higher voltages, as the electric field (typically 200–300 V/cm) is much higher than that of a slab gel platform. Thus, the separation in capillary electrophoresis is rapid. During electrophoresis, DNA fragments migrate through the capillaries toward the anode. Similar to gel electrophoresis, DNA fragments are separated according to their sizes, with the shorter fragments moving faster than the longer fragments. During capillary electrophoresis, linear polymers in the solution can act as obstacles to the migrating DNA fragments. Thus, electrophoretic separation models (Figure 8.1) developed for gels are applicable for capillary electrophoresis. Currently, capillary electrophoresis can separate DNA fragments of up to 1000 nucleotides with single-nucleotide resolution.

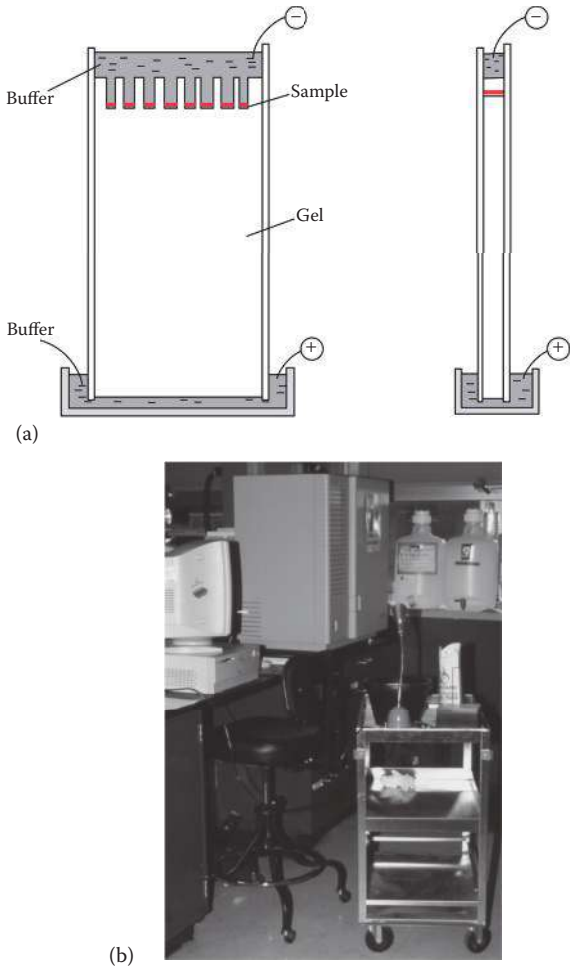

Figure 8.7 Vertical slab polyacrylamide gel. (a) Front and side view of vertical slab polyacrylamide gel apparatus. (b) Automated gel electrophoresis instrument, ABI PRISM 377® Genetic Analyzer, which was used in forensic laboratories but is now discontinued. (© Richard C. Li.)

Capillary electrophoresis instruments are equipped with detection systems utilizing laser excitation sources that excite the fluorescently labeled DNA fragments. They also include fluorescence detectors that record the signals emitted from the labeled DNA fragments (Chapter 9). In capillary electrophoresis, samples can only be analyzed sequentially so the throughput is much more limited than with slab gels. This disadvantage can be overcome by utilizing capillary array systems that can run up to 16 capillaries at one time (Figure 8.10).

8.3.3 Microfluidic Devices

Forensic DNA analysis includes the extraction of DNA from a sample, DNA quantification, polymerase chain reaction (PCR) amplification, capillary electrophoresis, data collection, and genotyping. Various microfluidic devices have been developed for forensic DNA analysis. Microfluidic devices control the movement of samples and reagents in a small, geometrically constrained environment and carry out biochemical reactions and analysis. These devices are made using microfabrication technology. This technology fabricates miniature structures historically used for integrated circuit fabrication in semiconductor manufacturing. The microfluidic

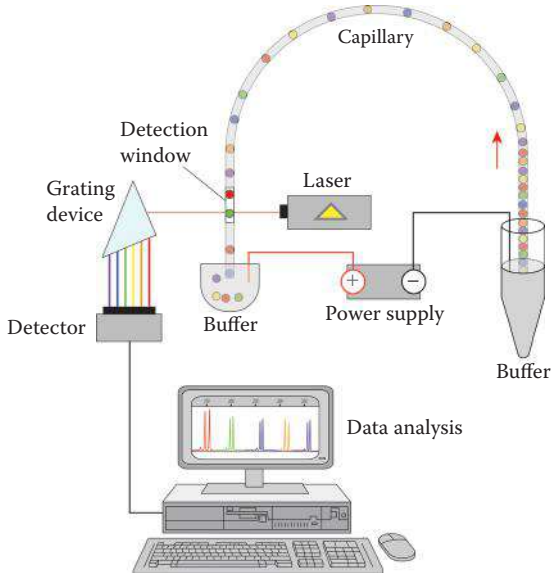

Figure 8.8 Essential components of a capillary electrophoresis system. The system includes a capillary, buffer reservoirs, two electrodes, a laser excitation source, a fluorescence detector, and an autosampler that holds the sample. Sample injection, electrophoresis, and data collection are automated and controlled by a computer. (© Richard C. Li.)

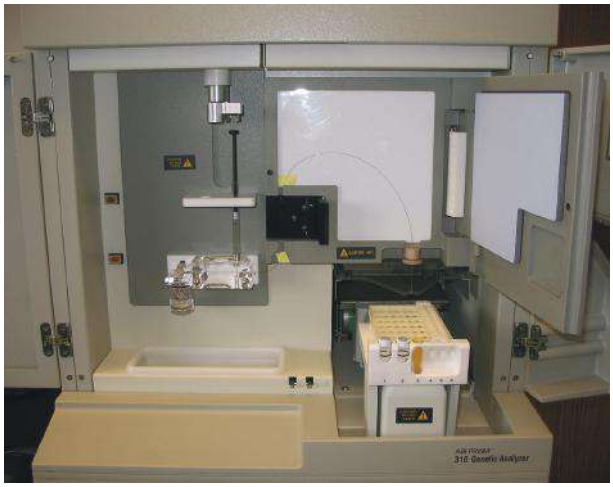

Figure 8.9 A single-capillary electrophoresis instrument, ABI PRISM 310® Genetic Analyzer. This was used in forensic laboratories but is now discontinued. (© Richard C. Li.)

devices can be classified as modular and integrated devices. Overall, the microfluidic device has many advantages compared with conventional techniques. The reaction volume required in microfluidic devices is usually in the nanoliter range, which decreases reagent and sample consumption. Additionally, due to the high surface area-to-volume ratio of the system, the efficiency of the system is greatly improved. Lastly, it can be automated, which makes the device a potential new platform for forensic DNA analysis.

Figure 8.10 Photo of a multicapillary electrophoresis instrument. ABI PRISM 3500® Genetic Analyzer. (© Richard C. Li.)

Figure 8.11 Components of the ABI PRISM 3500® Genetic Analyzer. A capillary array (left), electrophoretic buffer chambers (middle), and an autosampler (right). (© Richard C. Li.)

8.3.3.1 Modular Microfluidic Devices

The modular devices are separate devices for DNA extraction, quantitation, amplification, electrophoresis, and so on. Each processing step is carried out on different devices. The modular design is flexible, allowing laboratories to choose the best device suited to the procedure for each step. For example, an electrophoresis microfluidic device has been developed (Figure 8.12). The electrophoretic assay is carried out on a modular microfluidic chip format using the same basic principle of capillary electrophoresis. The single-use chip contains wells for samples, a sieving polymer matrix for electrophoresis, and size standards. The wells are connected through microchannels. When the interconnected wells and channels are filled, the chip becomes an integrated electrical circuit. Each well is then connected to an electrode with an independent power supply. DNA molecules, carrying negative charges, are

Figure 8.12 A modular instrument for gel electrophoresis. Left: Agilent 2100 Bioanalyzers contain 16-pin electrodes that fit into the wells of a chip. Right: The chip can accommodate sample wells, gel wells, and a well for a size standard. (© Richard C. Li.)

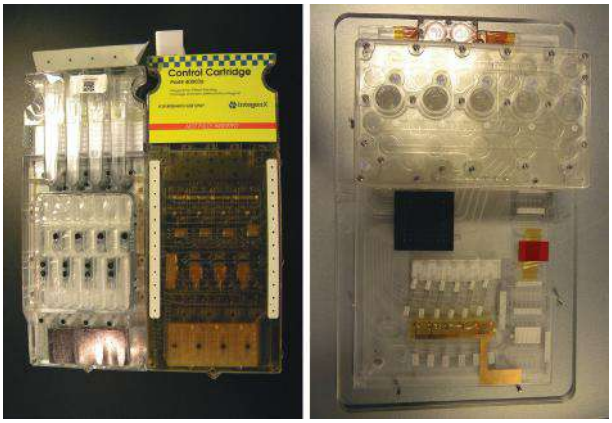

Figure 8.13 Integrated instruments for rapid DNA analysis. Cartridges developed by IntegeneX (left) and Zygem (right). (© Richard C. Li.)

electrophoretically separated based on their size. Smaller molecular weight fragments migrate faster than larger fragments. The polymer matrix contains fluorescent dye molecules that intercalate with DNA in a sample, which can be detected by laser-induced fluorescence. The duration of electrophoresis is approximately 30 min. The sensitivity is approximately 0.1 ng/μL; thus, sample consumption is small. The device provides the sizing and quantitation information of the nucleic acid fragment. In addition to DNA, the device can also be used for the analysis of RNA and protein. In the forensic laboratory, it has been used for the quantitation of amplified mtDNA product to monitor successful amplification and to normalize the quantities of DNA template for cycle sequencing.

8.3.3.2 Integrated Microfluidic Devices

It is possible to integrate several different modular devices into a single device (Figure 8.13). These devices are also known as "micro-total analysis" systems. An integrated system is a fully automated process. It also reduces the risk of contamination from laboratory sources. Full integration of all the constituent steps required for forensic STR DNA analysis has been achieved by a *rapid DNA* instrument. This instrument integrates various microfluidic devices including DNA extraction, PCR amplification, electrophoresis, detection, and genotyping steps into a single process on a cartridge. The cartridge, made from a variety of materials such as polycarbonate, incorporates pumps and valves controlling fluidic movement for the delivery of samples and prefilled reagents to the chambers for processing. DNA extraction usually utilizes silica-coated magnetic particles that can be collected by a magnet for DNA extraction. Microquantitative PCR utilizes the Taqman real-time quantitative PCR technique. During amplification, multiplex PCR is accomplished using commercial STR kits. Additionally, the cartridge also integrates a microcapillary electrophoresis chip for the separation of amplified products. The fluorescently labeled DNA fragments are detected based on laser-induced fluorescence. Data analysis and genotyping are then carried out to produce a DNA profile, which is compatible with the CODIS DNA database (Chapter 24). Rapid DNA instruments are fully automated platforms that are designed to process a single-source sample such as a buccal swab to generate a DNA profile and database search in less than 2 h. The goal is for these instruments to be deployed in field tests at the scene or at the booking stations operated by trained law enforcement agents.

8.4 Estimation of DNA Size

8.4.1 Relative Mobility

The *relative mobility* (R_f) of a DNA molecule during electrophoresis can be calculated as the distance of band migration divided by the distance of tracking dye migration. The DNA

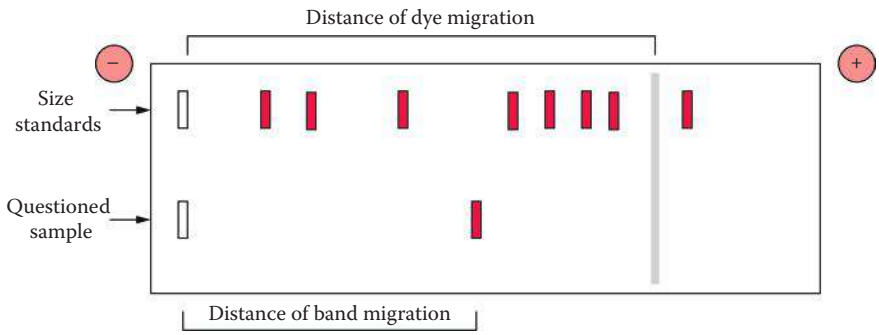

Figure 8.14 Measurement of the electromobility of DNA. Questioned DNA samples are analyzed concurrently with a set of standard DNA fragments of known size. The distances of DNA migration can thus be measured. The samples are usually mixed with a gel loading buffer prior to loading samples to the gel. The buffer contains dyes that add color to the sample to facilitate the process of loading. Dyes such as bromophenol blue and xylene cyanol FF migrate toward the positive electrode during electrophoresis. They can be used for tracking purposes as well. (© Richard C. Li.)

band migration is the distance from the sample origin to the center of the band. The tracking dye migration distance extends from the sample origin to the center of the dye band (Figure 8.14). To estimate the size of the DNA, standards containing DNA of known size and questioned samples are run on the same gel at the same time. The standards can be used to estimate the size of an unknown DNA molecule. A plot of $\log_{10}$ base pair of the standards versus R_f for a given gel can be constructed. A linear relationship, over a size range, between $\log_{10}$ base pair of the DNA molecule and the R_f can be observed. The R_f of the test sample is interpolated on the plot from which the size of an unknown DNA molecule can be determined (Figure 8.15).

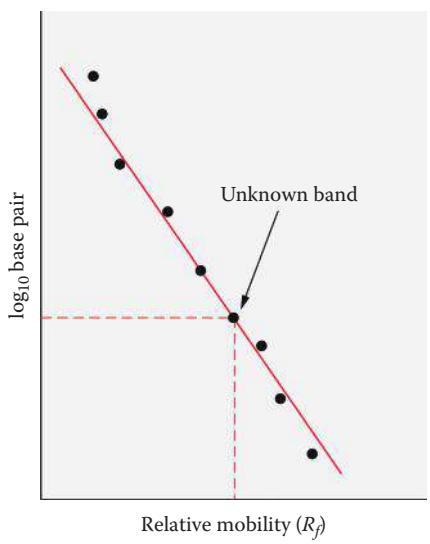

Figure 8.15 Estimation of the size of a DNA fragment. A plot of size standards versus relative migration during electrophoresis is shown. The size of the questioned DNA fragment can be estimated using this plot. (© Richard C. Li.)

8.4.2 Local Southern Method

The size of a DNA fragment is determined by an internal size standard that is a set of synthetic fragments with known molecular weight. The standard is labeled with a different colored dye so that it can be spectrally distinguished from DNA fragments of an unknown size (Figure 8.16). The sample including the internal size standard is then mixed in with DNA samples and is analyzed by electrophoresis. To determine the sizes of DNA fragments, a standard curve using the internal size standards must be established based on the reciprocal relationship between the electrophoretic mobility and the sizes of DNA fragments. However, this relationship is not exactly linear; instead, it appears to be sigmoidal (Figure 8.17). Therefore, the *Local Southern method*, described by Sir Edwin Southern, is used to generate standard curves for determining the sizes of DNA fragments. The equation is as follows:

$$L = \left[\frac{C}{M - M_0}\right] + L_0$$

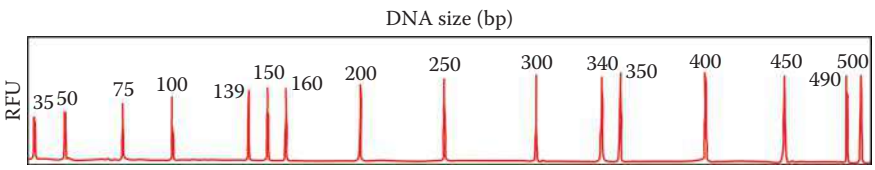

Figure 8.16 Electropherogram of a synthetic molecular weight size standard, GeneScan™ 500 size standard (Applied Biosystems). RFU, relative fluorescence units. (© Richard C. Li.)

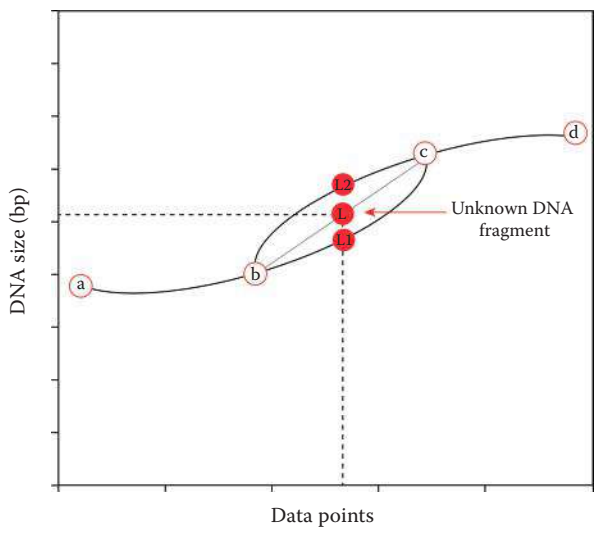

Figure 8.17 Estimation of unknown DNA fragment sizes using the Local Southern method. The migration of ssDNA fragments (20–1200 base pairs) is not linear over the entire fragment size range and is somewhat sigmoidal. To determine the size of a DNA fragment in question, two curves are generated. The first curve is generated using three standard points: two points (a and b) that are smaller and one point (c) that is larger than the questioned fragment. As a result, a fragment size (L_1) is determined. The second curve is then generated using an additional three standard points: one point (b) that is smaller and two points (c and d) that are larger than the questioned fragment. Thus, a fragment size, L_2, is determined. The two values (L_1 and L_2) are averaged to determine the size (L) of the questioned fragment. (Adapted from Southern, E.M.,*Anal Biochem*, 100, 319–323, 1979.)

In the equation, L is the size of the unknown fragment; M is the mobility of the fragment; and C, M_0, and L_0 are constants that are obtained from the fragments of known size. To determine the size of an unknown fragment, two curves are generated based on the equation. Each curve utilizes three data points. The first curve utilizes two size standards that are smaller and one size standard that is greater than that of the unknown fragment. As a result, size L_1 is obtained. The second curve utilizes a size standard that is smaller and two size standards that are greater than that of the unknown fragment. Thus, size L_2 is obtained. Finally, the average of L_1 and L_2 is defined as the size of the unknown fragment. In this method, only the size standards that are in the proximity of the unknown fragment are analyzed. However, the accuracy of the standard curve relies on the proper separation of the standards during electrophoresis.

Bibliography

Aboud, M.J., M. Gassmann, and B.R. McCord, The development of mini pentameric STR loci for rapid analysis of forensic DNA samples on a microfluidic system. *Electrophoresis*, 2010, **31**(15): 2672–2679.

Biega, L.A. and B.W. Duceman, Substitution of H_2O for formamide in the sample preparation protocol for STR analysis using the capillary electrophoresis system: The effects on precision, resolution, and capillary life. *J Forensic Sci*, 1999, **44**: 3.

Bienvenue, J.M., et al., An integrated microfluidic device for DNA purification and PCR amplification of STR fragments. *Forensic Sci Int Genet*, 2010, **4**(3): 178–186.

Buel, E., M.B. Schwartz, and M.J. LaFountain, Capillary electrophoresis STR analysis: Comparison to gel-based systems. *J Forensic Sci*, 1998, **43**(1): 164–170.

Butler, J.M., C.M. Ruitberg, and P.M. Vallone, Capillary electrophoresis as a tool for optimization of multiplex PCR reactions. *Fresenius J Anal Chem*, 2001, **369**(3–4): 200–205.

Butler, J.M., et al., Rapid analysis of the short tandem repeat HUMTH01 by capillary electrophoresis. *Biotechniques*, 1994, **17**(6): 1062–1064, 1066, 1068, passim.

Butler, J.M., et al., Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*, 2004, **25**(10–11): 1397–1412.

Chen, Y., et al., Sample stacking capillary electrophoretic microdevice for highly sensitive mini Y short tandem repeat genotyping. *Electrophoresis*, 2010, **31**(17): 2974–2980.

Choi, J.Y. and T.S. Seo, An integrated microdevice for high-performance short tandem repeat genotyping. *Biotechnol J*, 2009, **4**(11): 1530–1541.

Fernandez, C. and A. Alonso, Microchip capillary electrophoresis protocol to evaluate quality and quantity of mtDNA amplified fragments for DNA sequencing in forensic genetics. *Methods Mol Biol*, 2012, **830**: 367–379.

Frazier, R.R., et al., Validation of the applied biosystems Prism 377 automated sequencer for the forensic short tandem repeat analysis. *Electrophoresis*, 1996, **17**(10): 1550–1552.

French, D.J., et al., Interrogation of short tandem repeats using fluorescent probes and melting curve analysis: A step towards rapid DNA identity screening. *Forensic Sci Int Genet*, 2008, **2**(4): 333–339.

Gale, N., et al., Rapid typing of STRs in the human genome by HyBeacon melting. *Org Biomol Chem*, 2008, **6**(24): 4553–4559.

Gill, P., P. Koumi, and H. Allen, Sizing short tandem repeat alleles in capillary array gel electrophoresis instruments. *Electrophoresis*, 2001, **22**(13): 2670–2678.

Greenspoon, S.A., et al., A forensic laboratory tests the Berkeley microfabricated capillary array electrophoresis device. *J Forensic Sci*, 2008, **53**(4): 828–837.

Hagan, K.A., et al., A valveless microfluidic device for integrated solid phase extraction and polymerase chain reaction for short tandem repeat (STR) analysis. *Analyst*, 2011, **136**(9): 1928–1937.

Horsman, K.M., et al., Forensic DNA analysis on microfluidic devices: A review. *J Forensic Sci*, 2007, **52**(4): 784–799.

Hurth, C., et al., An automated instrument for human STR identification: Design, characterization, and experimental validation. *Electrophoresis*, 2010, **31**(21): 3510–3517.

Isenberg, A.R., et al., Analysis of two multiplexed short tandem repeat systems using capillary electrophoresis with multiwavelength fluorescence detection. *Electrophoresis*, 1998, **19**(1): 94–100.

9

Detection Methods

A variety of techniques are available for the detection of DNA fragments in forensic DNA analysis. For example, the direct detection of DNA fragments in a gel can be achieved via staining. The detection of DNA probes in a hybridization-based assay can be performed with radioisotopes, colorimetric assays, and chemiluminescence labeling. For polymerase chain reaction (PCR)-based assays, DNA primers can be labeled directly with fluorescent dyes.

9.1 Direct Detection of DNA in Gels

This section describes two simple and rapid detection methods used for detecting DNA: staining agarose gels with fluorescent intercalating dyes and staining denatured polyacrylamide gels with silver.

9.1.1 Fluorescent Intercalating Dye Staining

The location of DNA in an agarose gel can be determined directly by staining with low concentrations of fluorescent intercalating dyes (Chapter 6) such as ethidium bromide (Figures 9.1 through 9.3). These techniques allow the detection of DNA bands as small as 10 ng in agarose gels. Staining of DNA in agarose gels can be achieved by including an intercalating dye in the gel or staining the gel after electrophoresis in a dye-containing solution, followed by a washing step known as de-staining to reduce nonspecific staining. Historically, ethidium bromide-stained gels were photographed with a standard ultraviolet transilluminator at approximately 300 nm using a camera with an orange filter, whereas gels can currently be documented using digital techniques. When examining a DNA-containing gel under a UV lamp, the eyes and the skin should be protected from UV exposure. Ethidium bromide is a mutagen and a potential carcinogen. Ethidium bromide should be handled according to the Material Safety Data Sheet and safety protection wear should be used while handling the chemical. Ethidium bromide waste is usually disposed of as hazardous waste and is handled in accordance with laboratory guidelines. Alternatives to ethidium bromide are available, for example, fluorescent dyes such as SYBR® stains (Molecular Probes; Figures 9.2 and 9.3). Some of these alternatives are less mutagenic and have better performance than ethidium bromide.

9.1.2 Silver Staining

Electrophoretically separated DNA fragments can also be detected with silver nitrate staining. Silver staining of polyacrylamide gels has been used for the amplified fragment length polymorphism (AFLP; see Chapter 19) method of variable number tandem repeat (VNTR; see

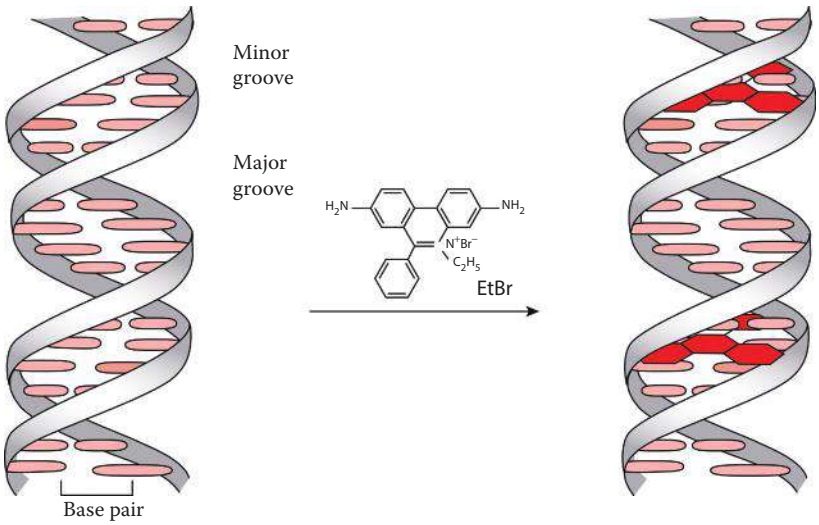

Figure 9.1 Schematic illustration showing binding of an intercalating agent, ethidium bromide (EtBr) to DNA. It intercalates into the minor groove of DNA. (© Richard C. Li.)

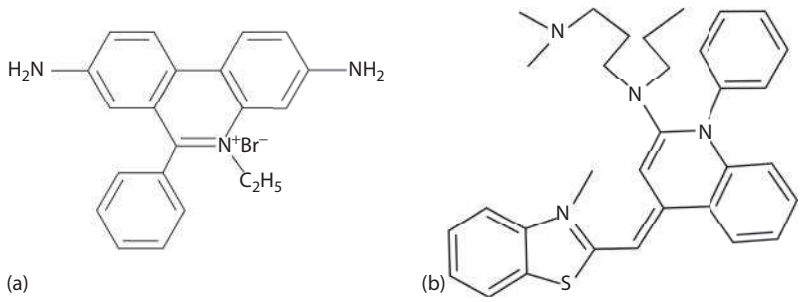

Figure 9.2 Intercalating agents used as nucleic acid stains. (a) Chemical structure of ethidium bromide. Ethidium bromide has UV absorbance. The emission maximum of the DNA–dye complex in aqueous solution is approximately 590 nm. (b) Chemical structure of SYBR Green 1. The stain is a cyanine dye that binds to DNA and is used as a nucleic acid stain. The excitation and emission maxima of the DNA–dye complex are 494 and 521 nm, respectively. The stain preferentially binds to double-stranded DNA rather than single-stranded DNA and RNA.

Chapter 19) profiling. The sensitivity of silver staining is approximately 100 times higher than that obtained with ethidium bromide, and silver staining is less hazardous than ethidium bromide detection. Also, the developing chemicals are readily available at low cost.

Silver staining involves processing a gel followed by exposure to a series of chemicals. First, the gel is submerged in a silver nitrate solution. Silver ions are positively charged, and DNA is negatively charged. Therefore, silver ions bind to the DNA and are subsequently reduced using formaldehyde to form a deposit of metallic silver on the DNA in the gel (Figure 9.4). A photograph of the gel with images of the silver-stained DNA strands is kept as a permanent record. Alternatively, the gels may be sealed and preserved for record purposes. One disadvantage of this method is that silver stains RNA and proteins along with DNA. The presence of restriction enzymes and polymerase should therefore be minimized. Additionally, bands from both complementary DNA strands may be detected in a denatured polyacrylamide gel, which leads to a two-band pattern.

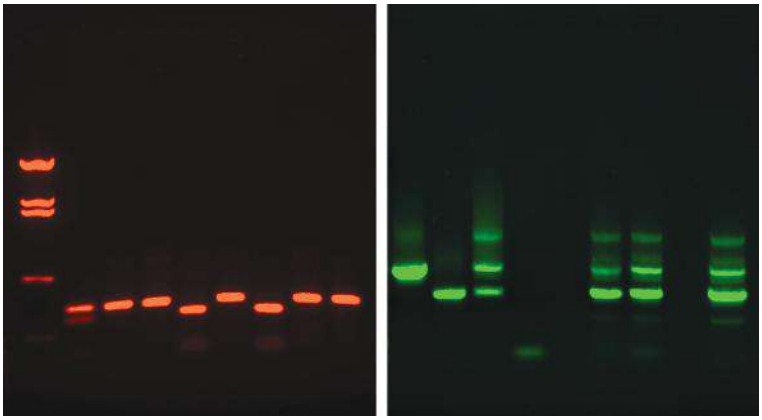

Figure 9.3 A DNA–dye complex emitting fluorescence on UV-light exposure. Ethidium bromide–containing agarose gel (left). SYBR Green 1–containing agarose gel (right). (© Richard C. Li.)

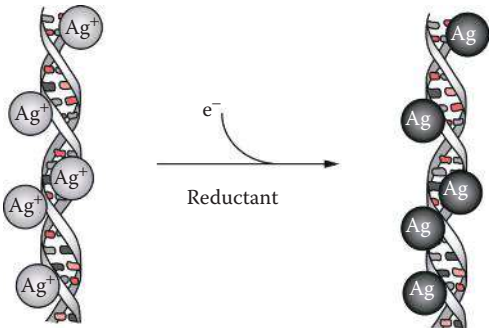

Figure 9.4 Silver staining of DNA. Silver (Ag^+) ions bind to DNA and are reduced to metallic silver (Ag) to form dark particles. (© Richard C. Li.)

9.2 Detection of DNA Probes in Hybridization-Based Assays

9.2.1 Radioisotope Labeled Probes

Radioisotope probe labeling was used for early versions of VNTR testing and DNA quantitation. Labeling can be accomplished in several ways (Figure 9.5). For example, nick translation incorporates labeled deoxyribonucleotides (dNTPs) into double-stranded DNA. DNase I is used to introduce single-strand nicks within the DNA fragment to be labeled. Next, DNA Polymerase I recognizes the nicks and replaces the preexisting nucleotides with new strands containing labeled dNTPs, resulting in the generation of ^{32}P-labeled double-stranded DNA molecules. ^{32}P, with a half-life of approximately 14 days, is the most common radioisotope used in this technique. Nick translation can utilize any dNTP labeled with ^{32}P.

Prior to hybridization, the probe is denatured into single-stranded fragments by boiling for a few minutes followed by rapid cooling on ice. After the hybridization process, these probes can be visualized by exposing the DNA-containing membrane to an x-ray film. The radioactive object is commonly placed in an x-ray cassette. The energy released from the decay of the radioisotopes is absorbed by the silver halide grains in the film emulsion and forms a latent image. A chemical development process amplifies the latent image and renders it visible on film (Figure 9.6). Because most ^{32}P emissions pass through the thin film emulsion without

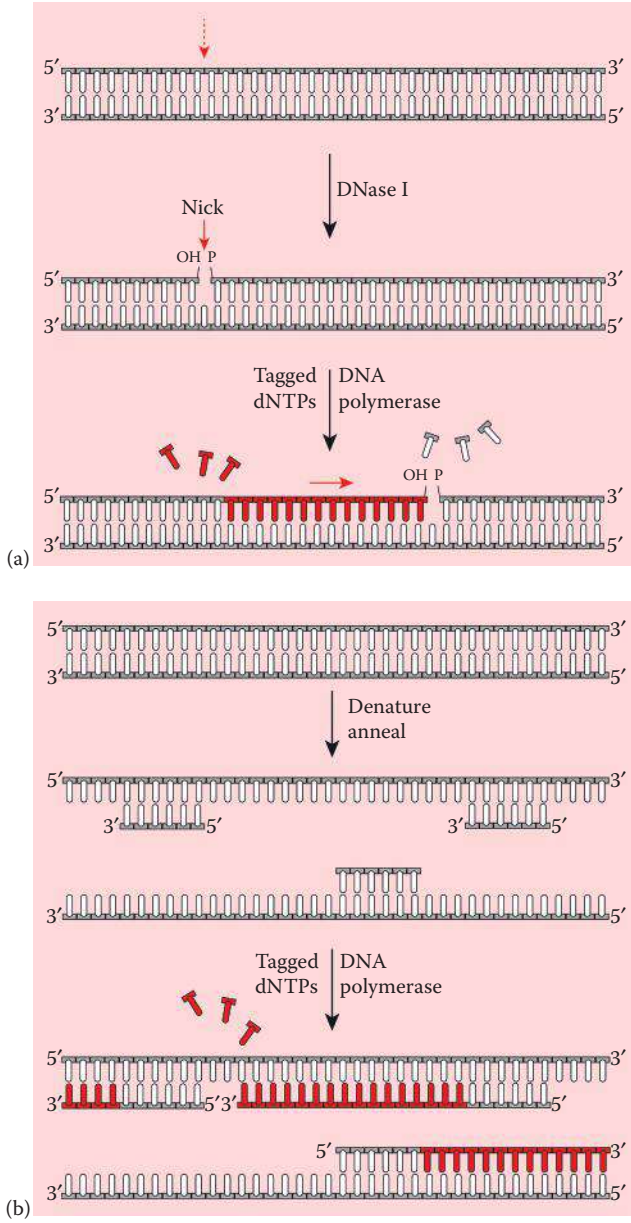

Figure 9.5 DNA probe labeling using dNTPs. (a) Nick translation method. A DNase is utilized to create nicks along a DNA strand. Tagged nucleotides (one or more than one of the dNTPs) are then incorporated into the DNA strand using a DNA polymerase. DNA synthesis extends from the 5′ to 3′ end and the original strand is degraded. (b) Random primer labeling method. A template DNA is denatured and separated into single strands. A random mixture of hexameric deoxynucleotide fragments, as primers, is then annealed to the template strand. The tagged nucleotides are then incorporated by the Klenow fragment of the DNA polymerase. (© Richard C. Li.)

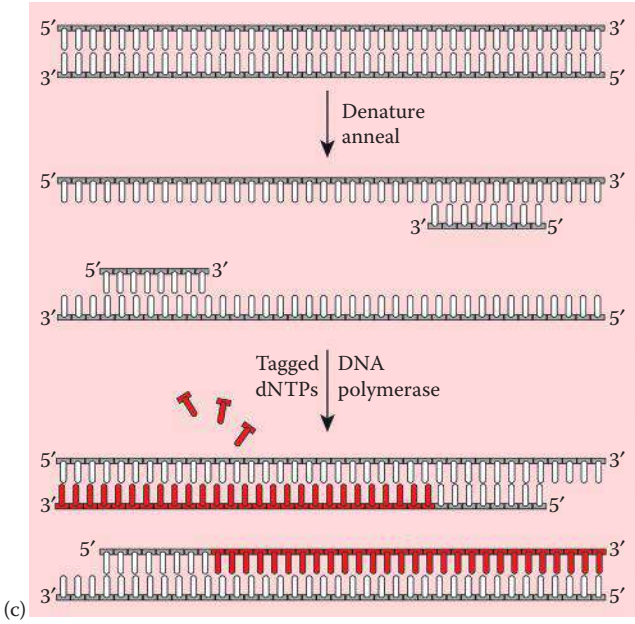

Figure 9.5 (Continued) (c) PCR method. DNA templates are denatured and then annealed to primers flanking the region of interest. During the DNA synthesis phase of a PCR cycle (see Chapter 7), tagged nucleotides are incorporated into the DNA strand using a Taq DNA polymerase, producing new double-stranded DNA. The process is repeated for many cycles. Only the first cycle is shown. (© Richard C. Li.)

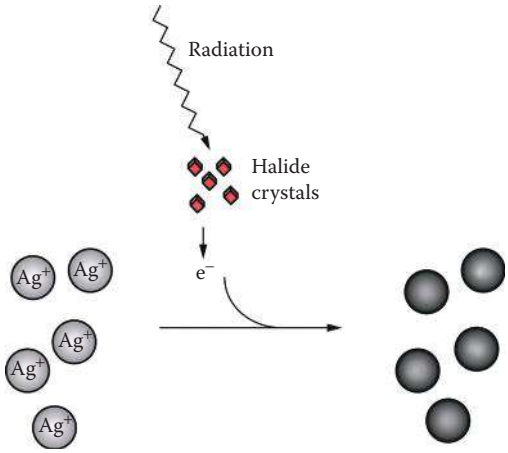

Figure 9.6 Autoradiography. Exposure to radiation causes halide crystals to release electrons, thus reducing silver (Ag^+) ions to metallic silver (Ag). (© Richard C. Li.)

contributing to the final image, the detection process may require long exposure times. Signal intensity can be enhanced, however, by using intensifying screens at low temperatures. The screens emit photons upon receiving radioactive β-particles, thus further enhancing signals. Additionally, highly sensitive x-ray film can be used. One disadvantage of using ^{32}P is that it is a safety hazard. Additionally, autoradiography is a lengthy process. Therefore, nonradioisotopic detection methods have become popular alternatives.

9.2.2 Enzyme-Conjugated Probe with Chemiluminescence Reporting System

The use of alkaline phosphatase (AP)-conjugated probes with chemiluminescent substrates comprises a highly sensitive nonradioisotopic detection system.

Alkaline phosphatase can cleave the phosphate groups from a variety of substrate molecules. Its enzymatic activity can be measured using dioxetane-based chemiluminescent substrates such as Lumigen® PPD (Figure 9.7). The Lumi-Phos Plus kit of Lumigen Inc. contains this substrate and can serve as a detection system for slot blot assays for DNA quantitation (Chapter 6) and RFLP assays for VNTR profiling (Chapter 19). AP catalyzes the cleavage of the phosphate ester of Lumigen® PPD, resulting in the release of a photon (Figure 9.8). The Lumigen® PPD substrate yields a long-lasting light emission that can be detected by exposure to x-ray film. This system provides a highly sensitive chemiluminescent detection method for AP-conjugated DNA probes in solution or on a solid matrix such as a membrane (Table 9.1).

9.2.3 Biotinylation of DNA with Colorimetric Reporting Systems

9.2.3.1 Biotin

Biotin, also known as vitamin H, is a water-soluble molecule found in egg yolk (Figure 9.9). It can be incorporated onto oligonucleotide probes without interfering with the ability of probes to hybridize because of its small size (molecular weight: 244.31 u). Signals from a biotinylated probe can be detected with an enzyme-conjugated avidin system. Two steps are required to detect biotin-labeled probes. First, an avidin conjugate consisting of a reporter enzyme is added. Then, the reporter enzyme is assayed with substrates.

9.2.3.2 Enzyme-Conjugated Avidin

Avidin is a glycoprotein found in egg white; it binds to biotin with extremely high affinity. Thus, a biotin–avidin complex is very stable. However, avidin detection has a high background due to nonspecific binding. The nonspecific binding can be reduced by replacing avidin with its

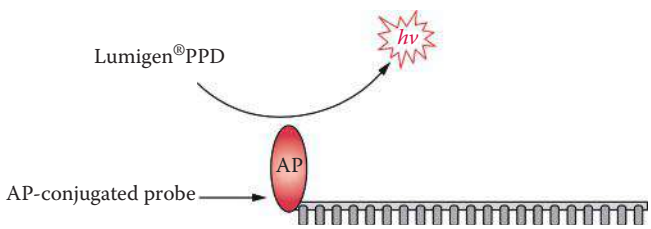

Figure 9.7 Detection system using AP-conjugated probe. Chemiluminescence is generated by using the Lumigen® PPD as an AP substrate. (© Richard C. Li.)

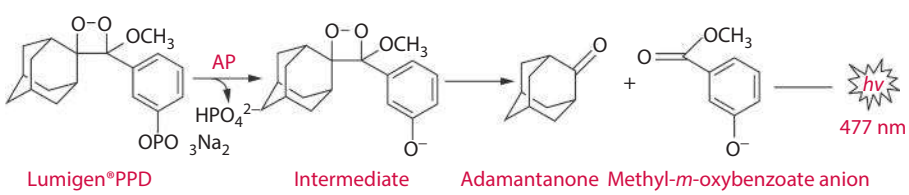

Figure 9.8 Lumi-Phos Plus contains Lumigen® PPD (4-methoxy-4-[3-phosphatephenyl]spiro[1,2-dioxetane-3,2′-adamantane], disodium salt). AP catalyzes the removal of the phosphate group of Lumigen® PPD and generates a chemiluminescent intermediate that is subsequently broken down to the excited state of methyl-*m*-oxybenzoate. The decay of the excited state end-product releases a photon.

Table 9.1 Forensic Applications of Enzyme Reporting Systems for Detecting DNA

Reporter Enzyme	Labeling	Detection Mechanism	Substrate	Forensic Application
AP	AP-conjugated probe	Chemiluminescent	Lumigen®PPD	RFLP; blot assay for DNA quantitation
Horseradish peroxidase (HRP)	Biotinylated probe, recognized by streptavidin conjugated with HRP	Colorimetric/ chemiluminescent	TMB/luminol	Blot assay for DNA quantitation
	Biotinylated primer, recognized by streptavidin conjugated with HRP	Colorimetric	TMB	Reverse blot assay for DQA1 typing and for mtDNA typing

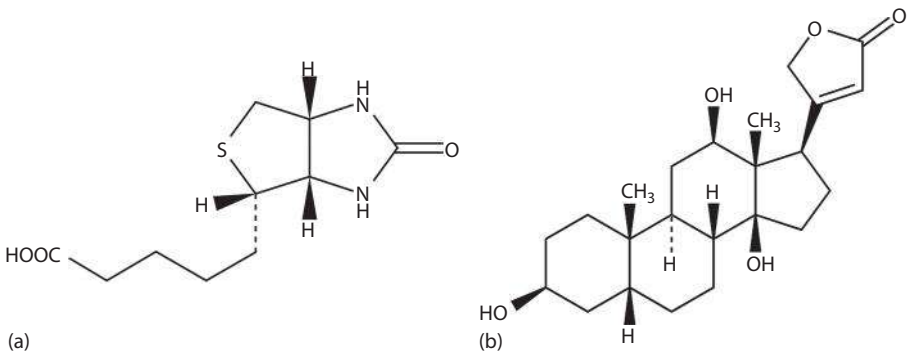

Figure 9.9 Nonradioactive tags for labeling and detecting nucleic acids. (a) Chemical structure of biotin, also known as vitamin H. Molecular formula: $C_{10}H_{16}N_2O_3S$. Molecular weight: 244.31. (b) Chemical structure of digoxigenin. Molecular formula: $C_{23}H_{34}O_5$. Molecular weight: 390.51.

streptavidin counterpart from *Streptomyces avidinii*. To detect binding, an enzyme-conjugated streptavidin such as horseradish peroxidase (HRP)-conjugated streptavidin can be used. HRP isolated from horseradish roots contains heme residues that catalyze the oxidation reactions of substrates (Chapter 12).

9.2.3.3 Reporter Enzyme Assay

HRP can be assayed with colorimetric, chemiluminescent, or fluorogenic substrates. One common colorimetric substrate for forensic DNA testing is 3,3′,5,5′-tetramethylbenzidine (TMB), which is oxidized by the peroxidase to form an insoluble precipitate of intense blue at an acidic pH (Figure 9.10). Because the colored precipitate is difficult to remove from membranes, TMB is not suitable if reprobing for RFLP analysis is required. This technique has been used for forensic DNA testing such as slot blot assays for DNA quantitation (Table 9.1). A chemiluminescent substrate such as a luminol-based reagent can also be utilized with

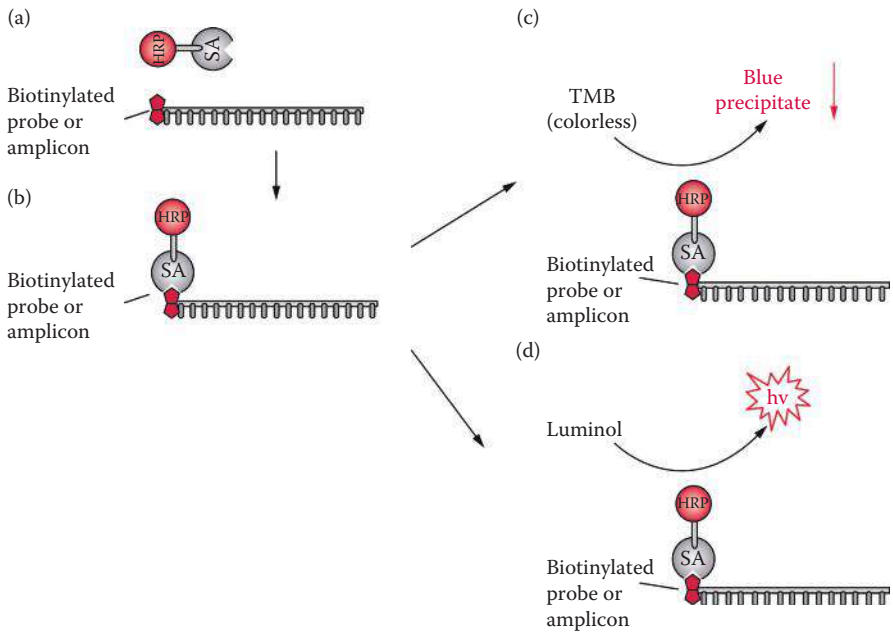

Figure 9.10 Detection system using biotinylated DNA probes with colorimetric and chemiluminescent reactions. (a) Biotinylated probe is incubated with a streptavidin (SA) and horseradish peroxidase (HRP) conjugate complex. (b) Biotin is recognized by the complex. Reporter enzyme assays can be carried out using either a colorimetric reaction with a TMB substrate (c) or a chemiluminescent reaction using luminol analogs (d). (© Richard C. Li.)

HRP. The peroxidase catalyzes the oxidation of luminol to form a chemiluminescent product (Figure 9.10).

9.3 Detection Methods for PCR-Based Assays

9.3.1 Fluorescence Labeling

9.3.1.1 Fluorescent Dyes

The advantages of fluorescence detection methods include a higher sensitivity and broader dynamic range than comparable colorimetric detection methods. Furthermore, they have the capacity for simultaneous analysis of complex samples such as multiplex PCR products with different fluorescent labels (Figure 9.11), allowing the distinction of various amplicons. Commonly used fluorescent dyes in DNA labeling emit fluorescence in the range of 400–600 nm (Figures 9.12 and 9.13).

9.3.1.2 Labeling Methods

Fluorescent dye labeling can be incorporated into a DNA fragment using a 5′-end fluorescently labeled oligonucleotide primer (Figures 9.14 and 9.15). The dye-labeled primer method is usually used for STR profiling (see Chapter 20) in which only one primer from each primer pair is labeled; therefore, only one strand can be detected. The two-band pattern observed with silver staining does not appear with this method. Additionally, dye-labeled primers allow multiplex PCR amplifications in the same tube.

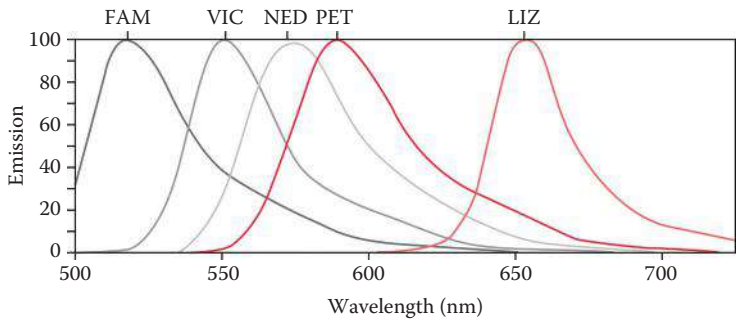

Figure 9.11 Emission spectra of common fluorescent dyes used for forensic DNA analysis. (© Richard C. Li.)

Color	Fluorescent dye	Globalfiler	AmpFlSTR Yfiler	PowerPlex fusion system	PowerPlex Y23 system
Purple dye	SID	☑			
Blue dyes	FAM	☑	☑		
	Fluorescein			☑	☑
Green dyes	JOE			☑	☑
	VIC	☑	☑		
Yellow dyes	NED	☑	☑		
	TMR (TAMRA)			☑	☑
Orange dyes	CC5			☑	☑
	LIZ	☑	☑		
Red dyes	PET		☑		
	ROX (CXR)			☑	☑
	TAZ	☑			

Figure 9.12 Examples of fluorescent dyes and their applications in STR kits. *Note*: Globalfiler and AmpFlSTR Yfiler are manufactured by Applied Biosystems. PowerPlex Fusion system and PowerPlex® Y23 System are manufactured by Promega.

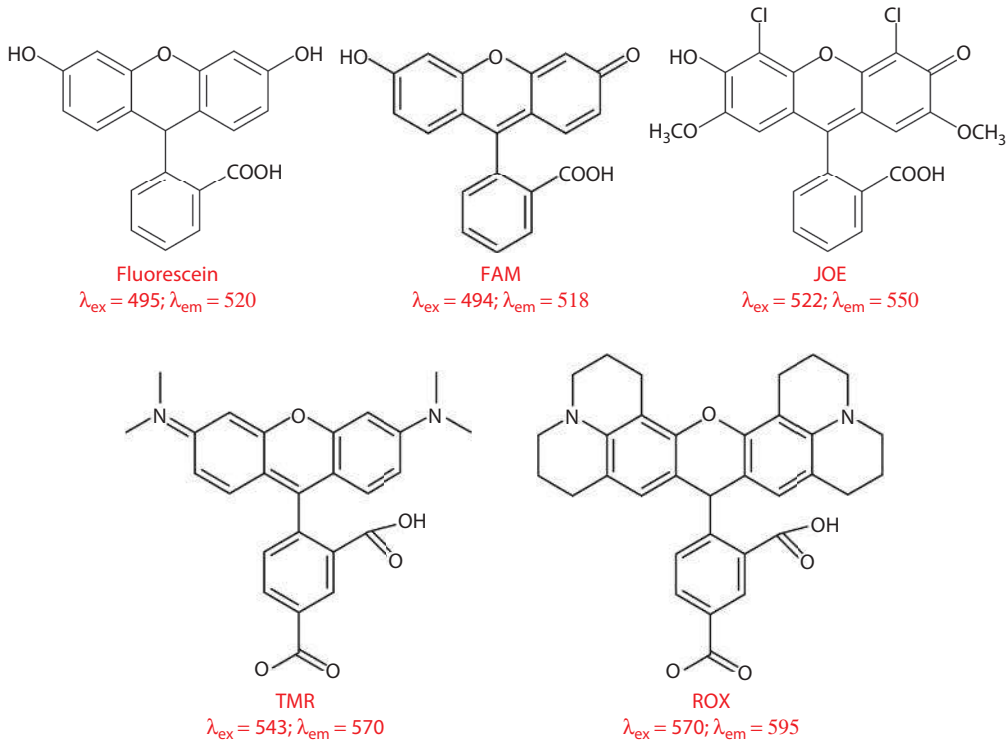

Figure 9.13 Chemical structures of representative fluorescent dyes. Fluorescein, 3′,6′-dihydroxyspiro[2-benzofuran-3,9′-xanthene]-1-one; FAM, 5(6)-carboxyfluorescein; JOE, 6-carboxy-4′,5′-dichloro-2′,7′-dimethoxy-fluorescein; TMR, 5-carboxytetramethylrhodamine; ROX, 5(6)-carboxy-X-rhodamine; λ_{ex}, peak excitation wavelength (nm); λ_{em}, peak emission wavelength (nm).

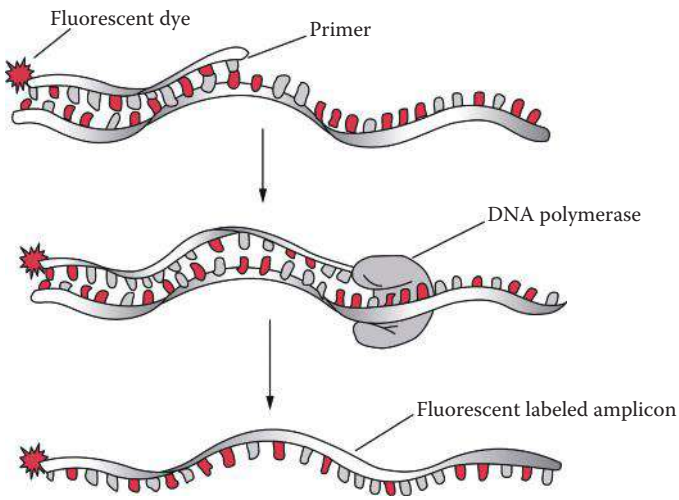

Figure 9.14 A fluorescent dye–labeled primer can be used for the amplification of DNA. The dye is conjugated at the 5′ end of the primer. The amplified product is fluorescently labeled. (© Richard C. Li.)

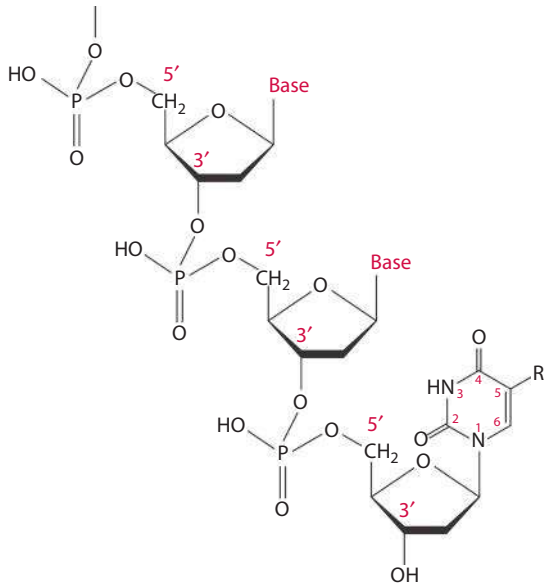

Figure 9.15 Direct labeling of nucleic acids with fluorescent dye. Labeled cytosine is shown as an example. R, fluorescent dye.

Alternatively, fluorescent dye labeling of DNA fragments can be carried out by incorporating fluorescently labeled dideoxynucleotides (ddNTPs) in the PCR product. This labeling method is usually used in DNA sequencing such as mtDNA sequence profiling (Chapter 23).

9.3.1.3 Fluorophore Detection

The *fluorophore* is a component of a florescent dye molecule that causes the molecule to be fluorescent. First, a laser strikes a fluorophore covalently linked to the end of a DNA fragment. An electron of the fluorophore is then excited, rising to an excited state from a ground state. The excited electron then descends to the ground state and releases a photon (Figure 9.16). The emitted photon has a longer wavelength (emission spectrum) than that of the excitation photon (excitation spectrum). The wavelengths of excitation and emission spectra (Figure 9.17) are largely dependent on the chemical structure of the fluorophore.

Lasers are commonly used as *excitation sources* because laser light emissions have high intensity and are monochromatic (single wavelength). The argon ion gas laser is frequently used in applications such as fluorescence-labeled STR and mtDNA sequence analysis because the excitation wavelength of commonly used fluorescent dyes matches the wavelength of the argon laser.

Optical filters are used to filter out undesired light and to allow only one particular wavelength to pass through. An essential optical filter consists of three components: an excitation filter, a dichroic beam splitter, and an emission filter. The excitation filter selectively transmits light from an excitation source. The light is then directed by the dichroic beam splitter to DNA molecules labeled with fluorescent dyes. The light emitted from a fluorophore is also transmitted by the dichroic beam splitter toward the detector. The emission filter selectively blocks undesired light, thus transmitting a specific wavelength of the emitted fluorescence. The light intensity emitted from a fluorophore is detected using a photosensitive device such as a charge-coupled device (CCD). The signal from the fluorophore is collected and converted to an electronic signal expressed in an arbitrary unit such as a *relative fluorescence unit* (RFU). Signals from multiple fluorophores in the same sample can be recorded separately using optical filters and a mathematical *matrix* (fluorophore separation algorithm). The function of a matrix is to subtract the

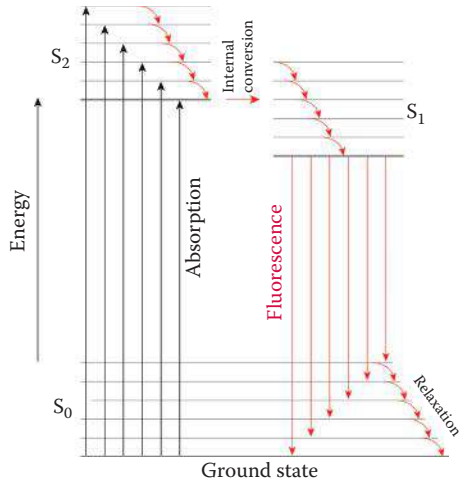

Figure 9.16 A Jablonski diagram illustrating the absorption of a single photon and subsequent emission of fluorescence. During an excitation stage, a single photon with the appropriate energy excites a molecule to the first excited energy level (S_2). In fluorescence reactions, the energy is supplied by an external light source such as a laser, while in chemiluminescence reactions, the energy is derived from a chemical reaction. The excited molecule is not stable and relaxes to its lowest vibrational level in the excited state (S_1). The molecule then returns to the ground state (S_0), emitting a photon. (© Richard C. Li.)

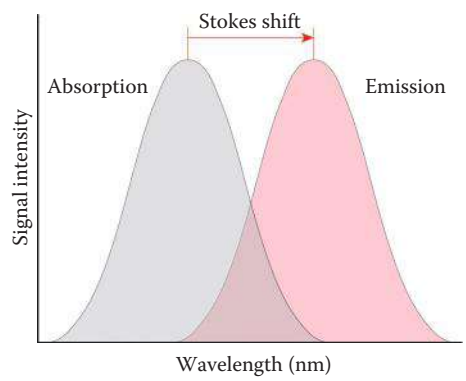

Figure 9.17 A fluorescence excitation spectrum. In fluorescence reactions, the energy of the photon emitted is lower than that of the excitation photon. As a result, the wavelength of an emission photon is longer than that of an excitation photon. The difference in wavelength is known as the Stokes shift. (© Richard C. Li.)

color overlaps of the various fluorescent dyes. The matrix can be established by the calibration of fluorescently labeled standards.

Bibliography

Bassam, B.J., G. Caetano-Anolles, and P.M. Gresshoff, Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem*, 1991, **196**(1): 80–83.

Bregu, J., et al., Analytical thresholds and sensitivity: Establishing RFU thresholds for forensic DNA analysis. *J Forensic Sci*, 2013, **58**(1): 120–129.

Bronstein, I. and P. McGrath, Chemiluminescence lights up. *Nature*, 1989, **338**(6216): 599–600.

10

Serology Concepts

10.1 Serological Reagents

10.1.1 Immunogens and Antigens

A foreign substance that is capable of eliciting antibody formation when introduced into a host is called an *immunogen*. Natural immunogens are usually macromolecules such as proteins and polysaccharides. Other molecular structures can also act as immunogens, for example, glycolipids (such as A, B, and O blood group antigens) and glycoproteins (such as Rh and Lewis antigens). However, they must be foreign to their hosts. The molecular structure of an immunogen, usually a small portion recognized by an antibody, is called an *epitope* or determinant site. An immunogen usually consists of multiple epitopes and is thus considered *multivalent* (Figure 10.1). Each epitope can elicit the production of its own corresponding antibody.

An *antigen* is a foreign substance that is capable of reacting with an antibody. All immunogens can be considered antigens, but not all antigens can elicit antibody formation. *Hapten* is one example of a substance that is antigenic but not immunogenic. Haptens are chemical compounds that are too small to elicit antibody production when they are introduced to a host animal. However, a hapten can be coupled to a *carrier*, usually a macromolecule, to produce antibodies. A hapten-conjugated carrier can become immunogenic to elicit the formation of an antibody specific to the hapten. The resulting antibody can bind to free haptens. Certain controlled substances such as cocaine and amphetamines are haptens and can be detected through corresponding antibodies for forensic toxicological analysis.

10.1.2 Antibodies

Antibodies, also known as *immunoglobulins*, are capable of binding specifically to antigens and are designated with an Ig prefix. The five major classes of immunoglobulins are designated IgG, IgA, IgM, IgD, and IgE. Additionally, IgG immunoglobulins can be further divided into four subclasses (IgG1–IgG4) and IgA immunoglobulins can be further divided into two subclasses (IgA1 and IgA2). Thus, there are a total of nine immunoglobulin isotypes in humans. *Isotypes* are the immunoglobulins that differ based on the molecular variations in the constant domains of the heavy and light chains (Figure 10.2). IgD, IgE, and IgG are usually monomers. IgM can be a membrane-bound monomer or a cross-linked pentamer (secreted form). IgA can be a monomer, dimer, or trimer. In an immune response, an initial exposure to an immunogen elicits a primary response, producing IgM immunoglobulins. Further exposure to the immunogen can elicit a secondary response, producing IgG, IgA, IgE, and IgD immunoglobulins. IgG is the most abundant immunoglobulin in serum. A majority of serology tests are based on the IgG immunoglobulins.

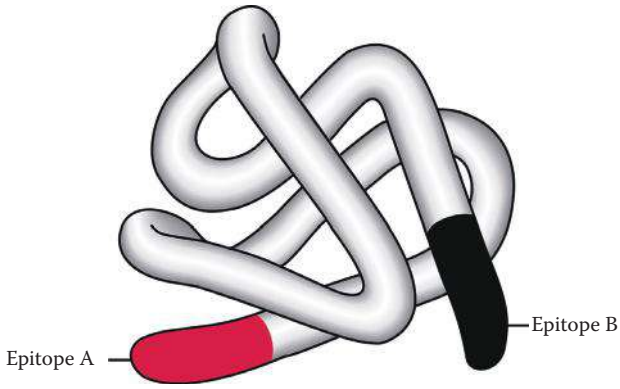

Figure 10.1 Multivalent immunogen. A protein with two different epitopes is shown. (© Richard C. Li)

Immunoglobulins have many similarities in their molecular structures. Figure 10.2 illustrates a diagram of the IgG molecule. The structures of immunoglobulins were first revealed by Gerald Edelman and Rodney Porter who shared a Nobel Prize in 1972. Immunoglobulins are composed of four polypeptide chains: two heavy (H) chains and two light (L) chains. The polypeptide chains are linked by disulfide bonds into a Y-shaped complex. The H chain can be divided into fragment antigen-binding (*Fab*) and fragment crystallizable (*Fc*) fragments. The L chain consists of a Fab fragment only. A typical antibody has two identical antigen-binding sites and is thus considered *bivalent*. The antigen-binding activity is located within the Fab fragments. In particular, the N-terminal ends of the L and H chains together form antigen-binding sites. At the amino acid sequence level, both H and L chains have *variable and constant domains* (Figure 10.2). The variable domains are located at the N-terminal ends of the immunoglobulins. Additionally, three small *hypervariable regions* are located within the variable domain of each chain.

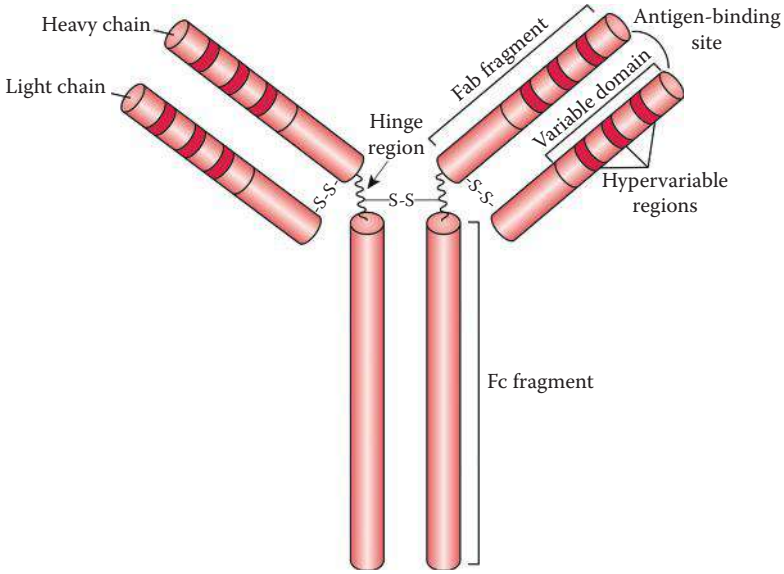

Figure 10.2 Immunoglobulin IgG structure. Immunoglobulin IgG is composed of light and heavy chains that contain variable domains. The remaining portions of chains form the constant domains. The variable domains of both light and heavy chains contain three hypervariable regions that form the antigen-binding site of the immunoglobulin. (© Richard C. Li.)

The diversity in the amino acid sequences of these hypervariable regions determines the specificity of the antigen-binding sites. The hinge regions provide flexibility to the antibody molecule and are important for the efficiency of the binding and cross-linking reactions. The basal portion of the H chain consists of *Fc* fragment.

The binding affinity and specificity of antibodies make them useful reagents for serological testing. Two types of antibodies are commonly used: polyclonal and monoclonal antibodies.

10.1.2.1 Polyclonal Antibodies

To produce an antibody, an immunogen is usually introduced into a host animal. A multivalent immunogen is capable of eliciting a mixture of antibodies with diverse specificities for the immunogen. As a result, a polyclonal antibody is produced by different B lymphocyte clones in response to the different epitopes of the immunogen. Antibodies can be circulating (in blood or other bodily fluids) or tissue bound (in cell surface antibodies). Circulating immunoglobulins are referred to as humoral antibodies.

The blood from an immune host is drawn and allowed to clot, resulting in the formation of a solid consisting largely of blood cells and a liquid portion known as *serum* containing antibodies (Figure 10.3). Such a preparation of humoral antibodies is also called a *polyclonal antiserum*. Depending on the type of animal used, the antibodies produced are classified as avian (B), rabbit (R), or horse (H) type. The characteristics of polyclonal antibodies may vary if they are produced from different individual host animals of the same species. Variations in reactions among different sources of antibodies should be monitored by quality-control procedures.

10.1.2.2 Monoclonal Antibodies

To produce a monoclonal antibody, spleen cells are harvested from a host animal, such as a mouse, inoculated with an immunogen (Figure 10.4). Next, the plasma cells of the spleen, which produce antibodies, are fused with myeloma cells. Since only a small population of cells fuse, a selection step is needed to allow only fused cells to grow (Figure 10.4). The fused cells, called *hybridoma* cells, are immortal (proliferate indefinitely) in cell cultures. Pools of hybridoma cells are diluted into single clones and are allowed to proliferate. The clones are then screened for the specific antibody of interest.

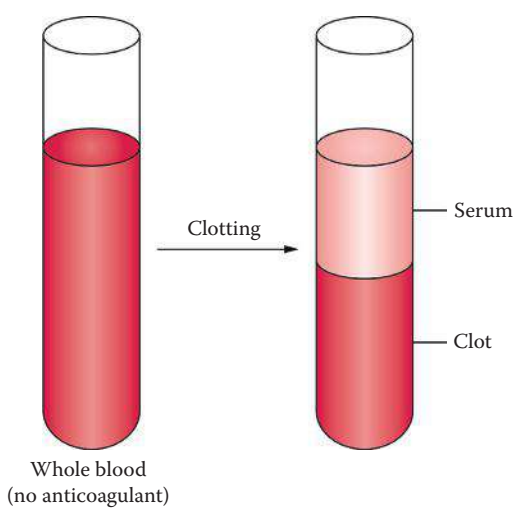

Figure 10.3 Serum component of blood. The blood of an immunized animal is collected in the absence of an anticoagulant and is allowed to clot. The resulting liquid portion of the blood is serum. (© Richard C. Li.)

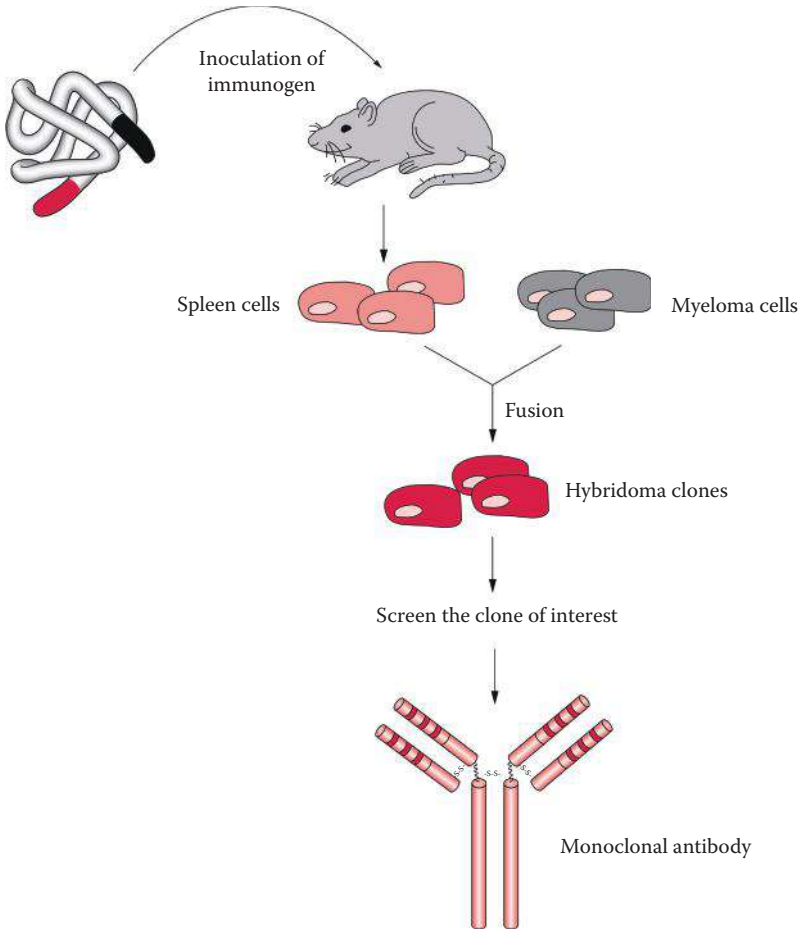

Figure 10.4 Preparation of a monoclonal antibody. A mouse is immunized with an immunogen, and its spleen cells are fused with myeloma cells to generate hybridoma cells. The clone that synthesizes and secretes the monoclonal antibody of interest can then be identified. (© Richard C. Li.)

The desired hybridoma clone can be maintained indefinitely, and it produces a monoclonal antibody that reacts with a single epitope. The hybridoma-derived monoclonal antibodies are specific and homogenous, and can be obtained in unlimited quantities. Monoclonal antibodies have been utilized in many serological assays, as discussed in Chapter 11. However, they have certain limitations in serology assays. For instance, monoclonal antibodies react with only a single epitope of a multivalent antigen and, therefore, cannot form cross-linked networks in precipitation assays (see Section 10.3.2.1).

10.1.2.3 Antiglobulins

Immunoglobulins are proteins that can also function as immunogens. If a purified foreign immunoglobulin or a fragment of a foreign immunoglobulin is introduced into a host, the antibodies produced are known as *antiglobulins*. Antiglobulins that are specific to a particular isotype can be produced in laboratories. In addition to specific antiglobulins, it is possible to produce nonspecific antiglobulins, which recognize an epitope that is common to all isotypes of an immunoglobulin class, such as the Fc portion of the heavy chain of all subclasses of human IgG. Antiglobulins are important reagents in many serological tests.

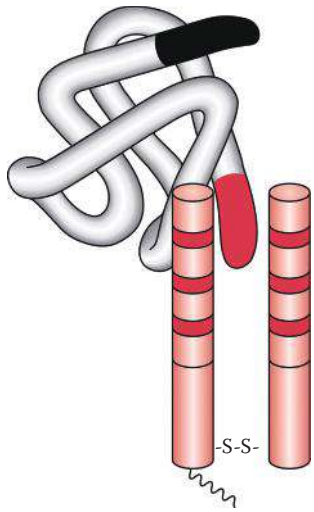

Figure 10.5 Affinity is a measure of the interaction between a single epitope on an antigen and a single binding site on an antibody. (© Richard C. Li.)

10.2 Strength of Antigen–Antibody Binding

The binding of an antigen to its specific antibody is mediated by the interaction between the epitope of an antigen and the binding site of its antibody. Noncovalent bonds can be formed during antigen–antibody binding. Various forces act cooperatively during antigen–antibody binding. These include hydrogen bonding, hydrophobic interactions that exclude water molecules from the area of contact, and van der Waals forces arising from the asymmetric distribution of the charges of electrons. The binding process occurs rapidly and the formation of the antigen–antibody complex is reversible. Such binding occurs at short distances when the antigen and antibody are in close proximity. Additionally, the strongest binding occurs only if the shape of the epitope fits the binding site of the antibody. The strength of the interaction between the antigen and the antibody depends on two characteristics, designated affinity and avidity.

Affinity is the energy of the interaction of a single epitope on an antigen and a single binding site on a corresponding antibody (Figure 10.5). The strength of the interaction depends on the specificity of the antibody for the antigen. Nevertheless, antibodies can bind with lower strength to antigens that are structurally similar to the immunogen. Such binding is known as *cross-reaction*.

Avidity is the overall strength of the binding of an antibody and an antigen (Figure 10.6). Since an antigen is usually considered multivalent and an antibody is bivalent, the avidity reflects the combined synergistic strength of the binding of all the binding sites of antigens and antibodies rather than the sum of individual affinities. It also reflects the overall stability of an antigen–antibody complex that is essential for many serological assays.

10.3 Antigen–Antibody Binding Reactions

The binding of an antigen to an antibody is an equilibrium reaction consisting of three types of reactions. The primary and secondary reactions form the basis for many forensic serological assays and will be discussed in the following subsections. The third type is called the tertiary reaction. It is used to measure *in vivo* immune responses such as inflammation and phagocytosis. Because most forensic serology tests are *in vitro* assays, the tertiary reaction is not commonly utilized in forensic serology testing and will not be discussed here.

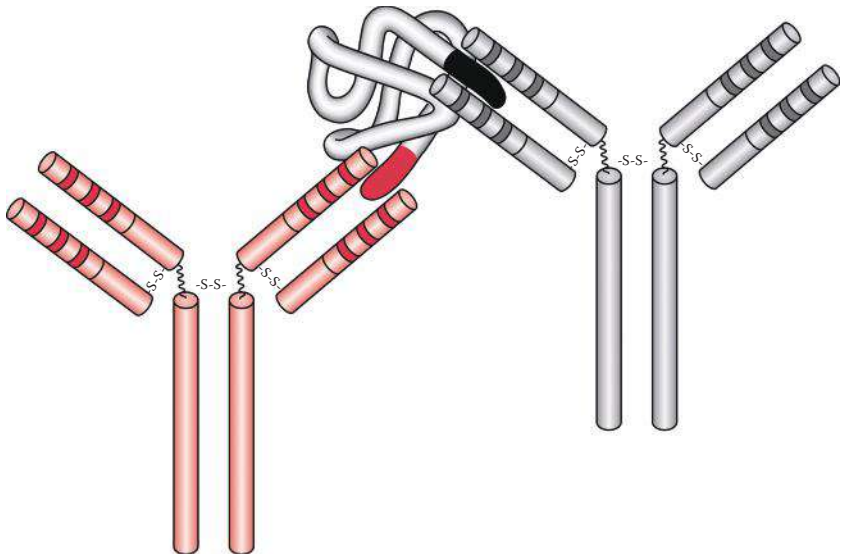

Figure 10.6 Avidity is a measure of the overall strength of the binding between antigens and antibodies. (© Richard C. Li.)

10.3.1 Primary Reactions

A primary reaction is the initial binding of a single epitope of an antigen (Ag) and a single binding site of an antibody (Ab) to form an antigen–antibody complex (Figure 10.7). This rapid and reversible binding reaction can be expressed as

$$\mathrm{Ag} + \mathrm{Ab} \rightleftarrows \mathrm{AgAb}$$

At equilibrium, the strength of the interaction can be expressed as the affinity constant (K_a) that reflects the affinity of binding, where:

$$K_a = \frac{[\mathrm{AgAb}]}{[\mathrm{Ag}][\mathrm{Ab}]}$$

The square brackets indicate the concentration of each component at equilibrium. K_a is the reciprocal of the concentration of free epitopes when half the antibody-binding sites are occupied. Thus, a higher K_a corresponds to a stronger binding interaction.

Techniques such as enzyme immunoassays, immunofluorescence assays, radioimmunoassays, and dye-labeled immunochromatography can measure the concentrations of antigen–antibody complexes formed by primary reactions (Chapter 11). These techniques are the most sensitive for detecting the presence of an antigen and an antibody in a sample. Additionally, many forensic serology assays are based on the detection of primary reactions and will be discussed in Chapters 12 through 17.

10.3.2 Secondary Reactions

The primary reaction between an antigen and an antibody is often followed by a secondary reaction. The three types of secondary reactions are precipitation, agglutination, and complement fixation. The techniques that detect secondary reactions are usually less sensitive but easier to perform than primary reaction assays. The precipitation and agglutination reactions form the basis for many serologic assays performed in forensic laboratories. These reactions will be discussed in the following subsections in detail. The third type of reaction is called *complement fixation*. If an antigen is located on a cell surface, the binding of the antigen and the antibody

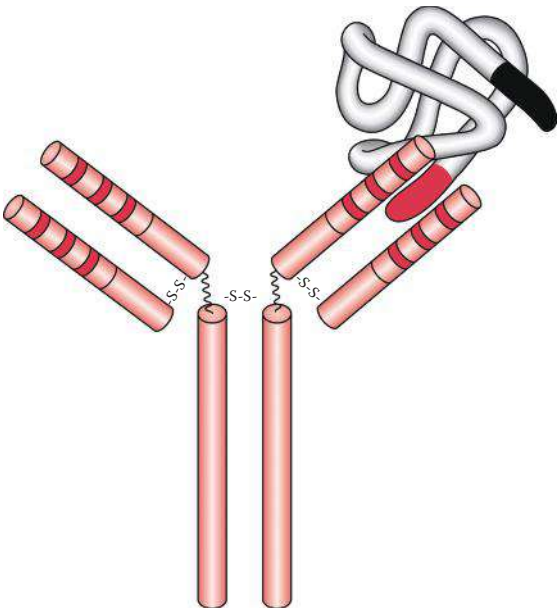

Figure 10.7 Primary reaction. Initial binding forms an antigen–antibody complex. (© Richard C. Li.)

may activate the classical complement pathway and lead to cell lysis, also known as a complement fixation reaction. The detection of this type of reaction is not commonly used in forensic serology.

10.3.2.1 Precipitation

If a soluble antigen is mixed and incubated with its antibody, the antigen–antibody complexes can form cross-linked complexes at the optimal ratio of antigen-to-antibody concentration. The cross-linked complex is insoluble and eventually forms a precipitate that settles to the bottom of a test tube. Antibodies that produce such precipitation are also called *precipitins*.

This precipitation reaction can be characterized by examining the effect of varying the relative ratio of antigen and antibody. If an increasing amount of soluble antigen is mixed with a constant amount of antibody, the amount of precipitate formed can be plotted. A precipitin curve (Figure 10.8)

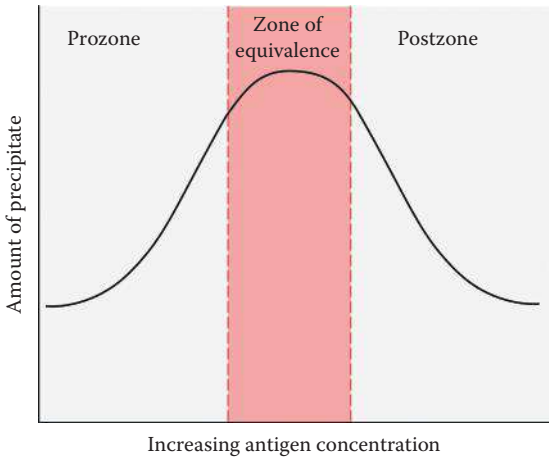

Figure 10.8 Precipitin curve. (© Richard C. Li.)

illustrates the results observed when antigens and antibodies are mixed in various concentration ratios. The curve can be divided into three zones known as the prozone, the zone of equivalence, and the postzone.

10.3.2.1.1 Prozone

At this zone, the ratio of antigen–antibody concentration is low. In other words, the antibody is in excess. Each antigen molecule is rapidly saturated with antibody, thus preventing cross-linking (Figure 10.9a). No precipitate is formed at the prozone stage.

10.3.2.1.2 Zone of Equivalence

As the concentrations of antigen increase, the amount of precipitate increases until it reaches a maximum. The amount of precipitation depends on the relative proportions of antigens and antibodies present. The maximum precipitation occurs in what is called the zone of equivalence. In the zone of equivalence, the ratio of antibody to antigen concentration is optimal and

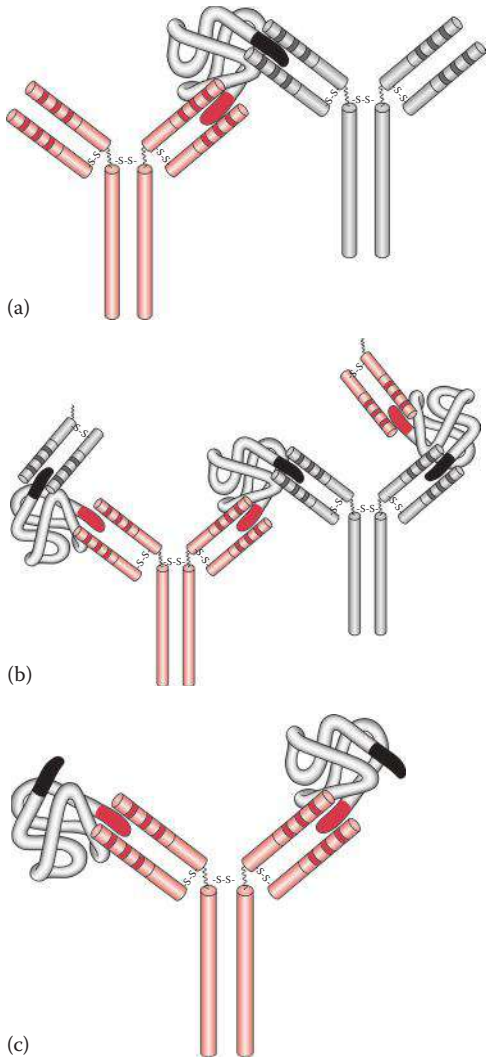

Figure 10.9 Antigen–antibody binding in (a) prozone, (b) zone of equivalence, and (c) postzone. (© Richard C. Li.)

precipitation occurs as a result of forming cross-linked networks (Figure 10.9b). Precipitation assays are usually carried out under the condition of the zone of equivalence, forming a sufficient quantity of precipitation to be detected.

10.3.2.1.3 Postzone

With the addition of more antigens, the ratio of antigen–antibody concentrations is high. In other words, the antigen is in excess. The amount of precipitate decreases and eventually diminishes. Each antibody molecule is saturated with antigen molecules (Figure 10.9c). Cross-linkage cannot form, and precipitation does not occur.

10.3.2.2 Agglutination

As discussed in the previous subsection, precipitation reactions involve soluble antigens. If the antigens are located on the surfaces of cells or carriers (carrier cells such as sheep erythrocytes,

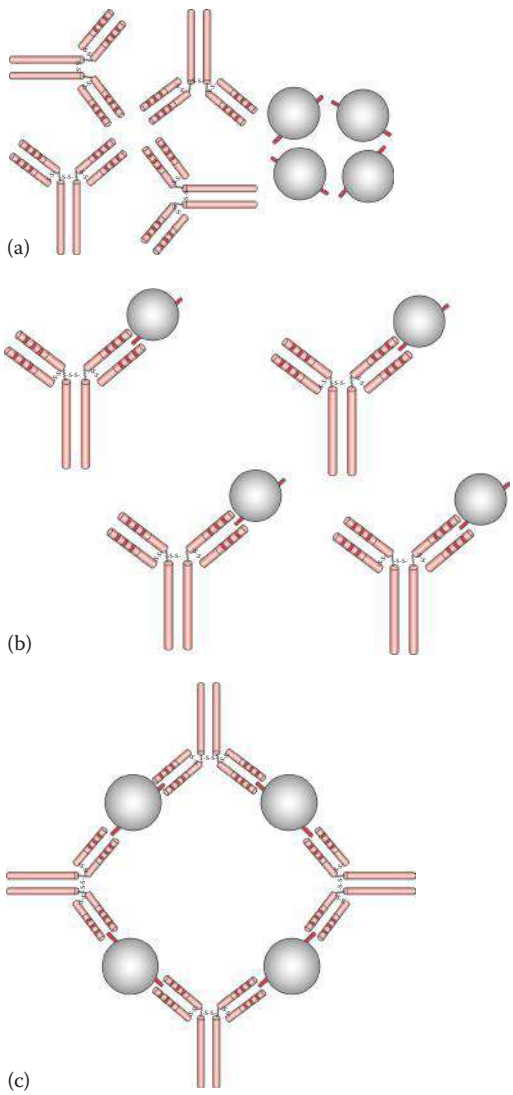

Figure 10.10 Agglutination reaction: (a) antigens are mixed with antibodies, (b) an antigen–antibody complex is formed during initial binding, and (c) the lattice is formed. (© Richard C. Li.)

bacteria, or latex particles), antibodies can bind to the surface antigens and can form cross-links among cells or carriers, causing them to aggregate. This aggregation is referred to as *agglutination*. In agglutination, a visible clumping can be observed as an indicator of the reaction of the antigen and the antibody. If the antigen is located on an erythrocyte, the agglutination reaction is designated *hemagglutination*. Agglutination is a two-step process that includes initial binding and lattice formation (Figure 10.10a).

10.3.2.2.1 Initial Binding

The first step of the reaction involves antigen–antibody binding at a single epitope on the cell surface (Figure 10.10b). This initial binding is rapid and reversible.

10.3.2.2.2 Lattice Formation

The second step involves the formation of a cross-linked network resulting in visible aggregates that constitute a lattice (Figure 10.10c). This involves an antibody binding to multiple epitopes because each antibody has two binding sites and antigens are multivalent. Lattice formation is a much slower process than the initial binding step. The cross-linking of cells requires physical contact. Additionally, an antibody must bind to epitopes on two different cells. The ability to cross-link cells depends on the nature of the antibody. Antibodies that produce such reactions are often called *agglutinins*.

Additionally, a *complete antibody* is capable of carrying out both primary and secondary interactions that result in agglutination. An antibody that can carry out initial binding but fails to form agglutination is called an *incomplete antibody*. This type of antibody is believed to have only one active antigen-binding site and is thus not capable of agglutination. It is potentially caused by the presence of steric obstruction due to the conformation of the incomplete antibody molecule, preventing the binding of antigens at the second binding site. However, other incomplete antibodies have two active sites but cannot bridge the distance between cells, thus failing to form lattices. Certain antibodies such as IgG are small and lack flexibility at the hinge region, and this may prevent agglutination. In contrast, the large IgM antibodies produce agglutination much more easily than IgG. Agglutination reactions have a wide variety of applications in the detection of antigens and antibodies. Such assays have high degrees of sensitivity and have been used for many years in forensic serology.

Bibliography

Baxter, S.J. and B. Rees, The use of anti-human haemoglobin in forensic serology. *Med Sci Law*, 1974, **14**(3): 159–162.

Dodd, B.E., Some recent advances in forensic serology. *Med Sci Law*, 1972, **12**(3): 195–199.

Fletcher, S. and M.J. Davie, Monoclonal antibody: A major new development in immunology. *J Forensic Sci Soc*, 1980, **20**(3): 163–167.

Giusti, G.V., Leone Lattes: Italy's pioneer in forensic serology. *Am J Forensic Med Pathol*, 1982, **3**(1): 79–81.

Goldsby, R., Kindt, T, and Osborne, B., *Immunology*, 6th edn. 2007, New York: W.H. Freeman and Company.

Lee, H.C., et al., Bits and pieces: Serology, criminalistics, anthropology and odontology. *J Forensic Sci Soc*, 1991, **31**(2): 293–296.

Patzelt, D., History of forensic serology and molecular genetics in the sphere of activity of the German Society for Forensic Medicine. *Forensic Sci Int*, 2004, **144**(2–3): 185–191.

Tietz, N. (ed.), *Fundamentals of Clinical Chemistry*, 1987. Philadelphia, PA: W.B. Saunders Company.

11

Serology Techniques
Past, Current, and Future

11.1 Introduction to Forensic Serology

11.1.1 The Scope of Forensic Serology

Forensic serology is the component of forensic biology that deals with the examination and identification of biological evidence. In particular, it focuses on determining the presence and identification of various bodily fluids such as blood, semen, and saliva in a questioned sample.

Bodily fluid stains are commonly associated with violent criminal cases. For instance, the identification of blood evidence (Chapter 12) is often necessary for investigating cases involving homicide, aggravated assault, sexual assault, and burglary. Proving the presence of blood has probative value or may corroborate allegations of violent acts. This evidence can also be used for investigative purposes. For example, the forensic DNA analysis of evidence such as a victim's blood on a suspect's weapon can establish a link between a victim and a suspect. The identification of semen (Chapter 14) and saliva (Chapter 15) is important for the investigation of a sexual assault case. For instance, a stain from a victim's clothing would be processed with forensic DNA analysis in a sexual assault case so that the DNA profile of the stain could be compared with that of an alleged suspect. Matching DNA profiles prove that the suspect's DNA was found on the victim's clothing, establishing a link between the suspect and the victim. Additionally, the identification of a biological stain as a semen stain through forensic serology testing proves that semen was found on the clothing taken from the victim and that the suspect is the source of the DNA from the semen stain. This evidence can then be used in court to support allegations that a sexual act occurred. Likewise, the presence of a suspect's saliva stains on a victim's genital area may corroborate an alleged oral copulation. The identification of bodily fluids other than blood, semen, and saliva will be discussed in Chapters 16 and 17.

11.1.2 Class Characteristics and Individual Characteristics of Biological Evidence

The identification of an unknown fluid sample is based on a comparison of the class characteristics (Section 3.2.1) of a sample with known standards of its class. Forensic identification typically involves bodily fluids, such as blood, semen, and saliva. If the presence of the bodily fluid is confirmed, the individual characteristics (Section 3.2.2) of the biological evidence are then determined to find out whether or not a bodily fluid sample has come from a particular individual. Today, this analysis can be achieved through forensic DNA analysis. The most commonly

utilized forensic DNA analysis is the short tandem repeat (STR) analysis used for human identification. Additionally, Y-chromosomal STR analysis is often utilized for the investigation of sexual assault crimes. Furthermore, mitochondrial DNA analysis is used for the identification of human remains. Therefore, establishing the probative value of a sample requires both the identification of its class characteristics and the individualization of its contributor. Therefore, the identification of bodily fluids cannot be replaced by forensic DNA analysis.

11.1.3 Presumptive and Confirmatory Assays

The identification of bodily fluids can be carried out using presumptive and confirmatory assays to identify the type of bodily fluid in question. The advantages of *presumptive assays* are that these assays are sensitive, rapid, and simple. A positive reaction of a presumptive assay indicates the possibility of the presence of the bodily fluid in question. However, presumptive assays are not very specific. Therefore, they should not be considered conclusive for the presence of a type of bodily fluid. In contrast, a negative assay indicates that the questioned bodily fluid is absent. Thus, presumptive assays can be used as a screening method and for narrowing down bodily fluid stains prior to other types of analyses, such as forensic DNA testing. Moreover, these assays can be used as a search method to locate bodily fluid stains at the crime scene. Additional assays, such as confirmatory assays, should be conducted afterward if necessary.

Confirmatory assays are more specific for the bodily fluid in question. These assays are utilized to identify bodily fluids with higher certainty than presumptive assays. For example, bloodstains are commonly associated with criminal investigations. A reddish-brown stain that has been identified through visual examination is usually tested by using presumptive assays. If the result of the presumptive assay on the alleged bloodstain is positive, the stain is then further analyzed by forensic DNA analysis. This approach *indicates* the presence of blood. Confirmatory assays are performed when a sample has to be *identified* as blood (Figure 11.1). Additionally, the human or animal origin of blood evidence can be determined if necessary and the techniques to do so will be discussed in Chapter 13.

11.1.4 Primary and Secondary Binding Assays

Traditionally, the detection and measurement of the antigen–antibody binding reactions serve as the bases of forensic serology. These assays fall into two categories: primary and secondary binding assays. Recall that the primary binding assays involve the initial binding between a single epitope of an antigen and a single binding site of an antibody (see Section 10.3.1). Therefore, they are very sensitive. Although the secondary binding assays are less sensitive than the primary binding assays, they are easier to perform than primary assays. The secondary assays consist of precipitation-based and agglutination-based assays. Precipitation-based assays have been used for species identification. Agglutination-based assays are more sensitive

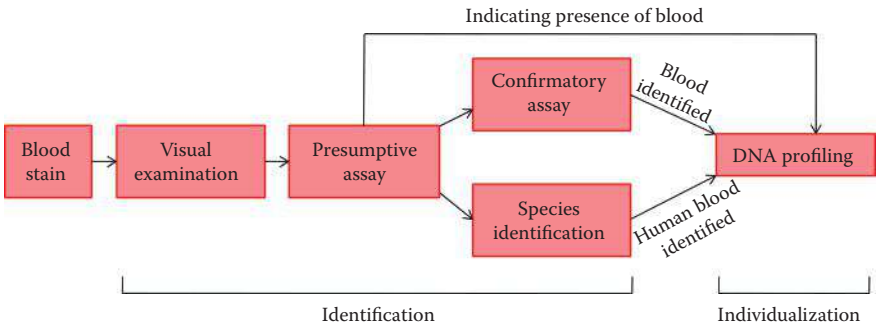

Figure 11.1 An example of work flow for processing blood samples. (© Richard C. Li.)

than precipitation-based assays. Agglutination-based assays, which detect antigens located on the surface of cells or carriers, are normally applied to blood group typing. Recently, emerging techniques such as RNA, proteomic, and DNA methylation assays, and fluorescence as well as Raman spectroscopy can potentially be used for the identification of bodily fluids.

11.2 Primary Binding Assays

11.2.1 Enzyme-Linked Immunosorbent Assay (ELISA)

An enzyme-linked immunosorbent assay (ELISA) is an immunoenzyme assay that can be used to detect and measure the antibody or antigen in question. The most common ELISA that is used in forensic serology is the antibody-sandwich ELISA (Figure 11.2). It is utilized to detect the prostate-specific antigen (PSA) to identify seminal stains and amylase for the identification of saliva (Chapters 13 and 14).

An antibody (usually monoclonal) coating is formed by nonspecific adsorption onto a solid phase such as the wells of a polystyrene plate. A sample containing the antigen to be tested is then added and binds to the solid-phase antibody. Subsequently, a second antibody is added to form an antibody–antigen–antibody sandwich complex. The second antibody binds to different epitopes of the antigen. Next, an enzyme-labeled antiglobulin (Chapter 10) is added to bind the sandwich. Subsequently, the excess enzyme-labeled antiglobulin is removed by washing and the bound antiglobulin can be detected.

A number of enzymes such as alkaline phosphatase and horseradish peroxidase have been used as reporting enzymes (Chapter 9) to label the antiglobulin for ELISA. The enzyme catalyzes the substrate and produces colorimetric or fluorometric signals. The intensity of the signals can

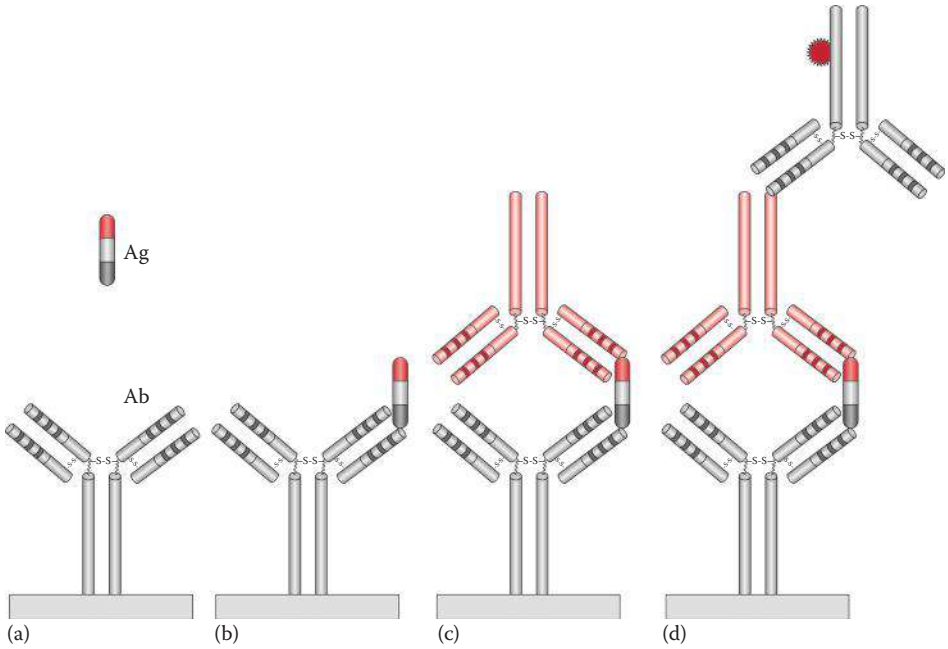

Figure 11.2 ELISA assay. (a) Sample containing Ag (antigen) is applied to a sample well where Ab (antibody) is immobilized. (b) Ag binds to an immobilized Ab to form an Ag–Ab complex. (c) A second Ab to a different epitope is added to form an Ab–Ag–Ab sandwich. (d) Labeled antiglobulin binds to the sandwich. The bound antiglobulin can be detected by various reporting schemes. (© Richard C. Li.)

be detected spectrophotometrically and is proportional to the amount of bound antigen. The amount of antigen can be quantified by comparing the standard with known concentrations. Recall that antiglobulin can recognize an epitope that is common to all isotypes within an antibody class (Section 10.1.2.3); the enzyme-labeled antiglobulin is a universal reagent regardless of the antibody used, as long as they are of the same class.

Alternatively, antibodies in a sample can also be detected and may be quantified by an ELISA system in which the antigen is bound to a solid phase instead of the antibody. After an antibody in a sample binds to the solid-phase antigen, an enzyme-labeled antiglobulin is added to bind to the bound antibody. The bound antiglobulin can be detected and measured by the addition of an enzyme substrate. The enzymatic catalytic reaction is similar to the antibody sandwich ELISA procedure previously described.

11.2.2 Immunochromatographic Assays

Figure 11.3 depicts a test using an immunochromatographic membrane device. A dye-labeled monoclonal antibody is contained in a sample well and a polyclonal antibody for the antigen (or a second monoclonal antibody to a different epitope of the antigen) is immobilized onto a test zone of a nitrocellulose membrane. An antiglobulin that recognizes the antibody is immobilized onto a control zone.

The assay is carried out by loading a sample into the sample well. The antigen in the sample binds to the dye-labeled antibody already in the sample well to form an antigen–antibody complex. The complex then diffuses across the nitrocellulose membrane until it reaches the test zone. The antibody immobilized at the test zone traps the antigen–antibody complex to form an antibody–antigen–antibody sandwich. The presence of the antigen in the sample results in a colored vertical line at the test zone. (Figure 11.4)

The immunochromatographic device also utilizes a control zone to ensure that the device works properly and that the sample has diffused completely along the test strip. Unbound

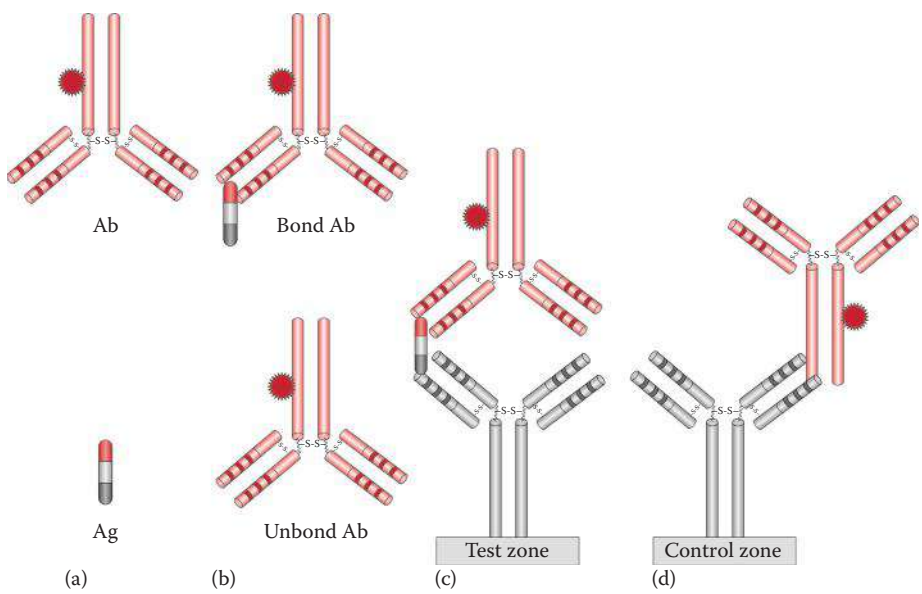

Figure 11.3 Immunochromatographic assay. (a) Sample containing Ag (antigen) is loaded in a sample well. (b) Ag binds to a labeled Ab (antibody) to form a labeled Ab–Ag complex. (c) At the test zone, the labeled Ab–Ag complex binds to an immobilized Ab to form a labeled Ab–Ag–Ab sandwich. (d) At the control zone, a labeled Ab binds to an immobilized antiglobulin and is captured at the control zone. (© Richard C. Li.)

monoclonal antibodies diffuse across the membrane until they reach the control zone where they are trapped by the immobilized antiglobulin. This antibody–antiglobulin complex at the control zone also results in a colored vertical line (Figure 11.4). The test is considered valid only if the line in the control zone is observed. The presence of an antigen results in a line at both the test zone and the control zone, while the absence of an antigen results in a line in the control zone only (Figure 11.4). This method is rapid and simple and thus can be used as a screening test in laboratories and as a field test at crime scenes to identify semen, saliva, and species (Table 11.1).

However, a false-negative result may be obtained if a sample contains a very high concentration of an antigen (Figure 11.5). Under this condition, the dye-labeled antibody will become saturated with the antigen to form antigen–antibody complexes. The unbound antigen will diffuse along with the

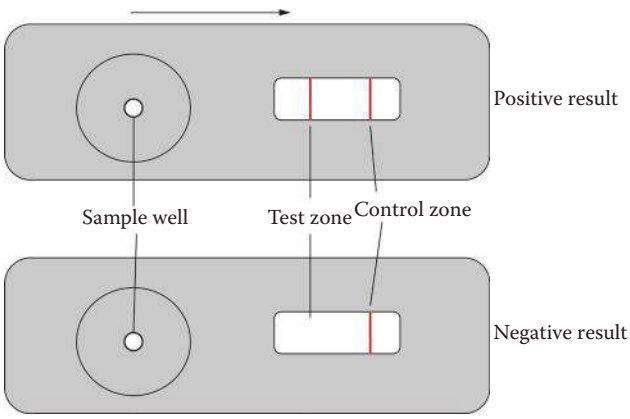

Figure 11.4 Immunochromatographic device. Positive and negative results are shown. (© Richard C. Li.)

Table 11.1 Common Immunochromatographic Assays for Forensic Applications

Assay	Antigen	Labeled Antibody	Immobilized Antibody	Forensic Application
ABAcard® HemaTrace® (Abacus Diagnostics)	Hemoglobin (Hb)	Monoclonal antihuman Hb antibody	Polyclonal antihuman Hb antibody	Blood and species identification
RSID™-Blood (Independent Forensics)	Glycophorin A (GPA)	Monoclonal antihuman GPA antibody	Monoclonal antihuman GPA antibody	Blood and species identification
RSID™-Saliva (Independent Forensics)	Human salivary α-amylase (HAS)	Monoclonal antihuman HAS antibody	Monoclonal antihuman HAS antibody[a]	Saliva identification
One-Step ABAcard PSA® (Abacus Diagnostics)	Prostate-specific antigen (PSA)	Monoclonal antihuman PSA antibody	Polyclonal antihuman PSA antibody	Semen identification
RSID™-Semen (Independent Forensics)	Semenogelin (Sg)	Monoclonal antihuman Sg antibody	Monoclonal antihuman Sg antibody[a]	Semen identification

[a] The epitope recognized by the immobilized antibody is different from that of the labeled antibody.

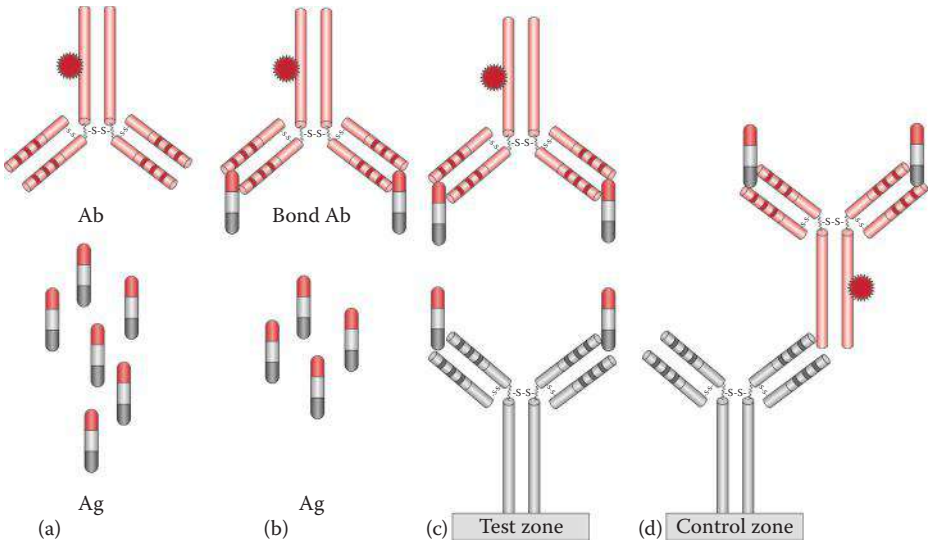

Figure 11.5 High-dose hook effect of immunochromatographic assay. (a) Ag (antigen) in sample is loaded in the sample well. (b) Ag binds to a labeled Ab (antibody) to form a labeled Ag–Ab complex. (c) At the test zone, free Ag binds to an immobilized Ab to form an unlabeled Ab–Ag complex and prevents the formation of a labeled Ab–Ag–Ab sandwich. (d) At the control zone, a labeled Ab binds to immobilized antiglobulin and is captured at the control zone. (© Richard C. Li.)

antigen–antibody complex toward the test zone. At the test zone, the unbound antigen will compete with the antigen–antibody complexes for the antibody immobilized at the test zone. Since the antigen is in excess, the unbound antigen binds to the immobilized antibody, preventing the formation of the antibody–antigen–antibody sandwich. The resulting reading appears as a negative result. This artifact is known as the *high-dose hook* effect. To prevent the high-dose hook effect, a smaller volume of sample can be applied, or the sample can be diluted to reduce the amount of antigen applied.

11.3 Secondary Binding Assays

11.3.1 Precipitation-Based Assays

These techniques are based on the precipitation reaction and are used primarily for species identification in forensic laboratories.

11.3.1.1 Immunodiffusion

Immunodiffusion is a passive method in which an antigen or an antibody or both are allowed to diffuse and therefore a gradient, from low to high concentration, is established for an antigen or an antibody or both. As a result, precipitation occurs due to the interaction between an antigen and an antibody. The assay can be carried out in a liquid or a semisolid medium such as agarose gel. The semisolid medium can stabilize the diffusion process and reduce interference such as convection, which is the movement of molecules within liquids. The two types are single immunodiffusion and double immunodiffusion.

11.3.1.1.1 Single Immunodiffusion

A concentration gradient is established for either an antigen or an antibody. One serology technique based on this principle is radial immunodiffusion—a single diffusion method in which a concentration gradient is established for an antigen. The antibody is uniformly distributed in

the gel matrix (Figure 11.6). The antigen is loaded into a sample well and allowed to diffuse from the well into the gel until a precipitation reaction occurs. The precipitate ring around the well is observed; the area within the ring of precipitate is proportional to the amount of antigen loaded in the well. Standards using known concentrations of antigen can be included in the same assay along with the samples, and a standard curve can be plotted. The amounts of antigen in samples can then be quantified by comparing the results with the standard curve.

11.3.1.1.2 Double Immunodiffusion

The second type of assay is double diffusion in which a concentration gradient is established for both an antigen and an antibody. The most common examples are the ring assay and the Ouchterlony assay. The *ring assay* can be performed in a test tube or capillary tube (Figure 11.7). An antiserum, a denser phase, is placed in a small tube. An antigen solution is carefully layered on top of an antibody solution without mixing. Both the antigen and the antibody will diffuse toward each other. In a positive reaction, a ring of precipitate can be observed at the interface of the two solutions. A negative reaction is indicated by a lack of precipitation. The assay requires positive and negative controls along with questioned samples.

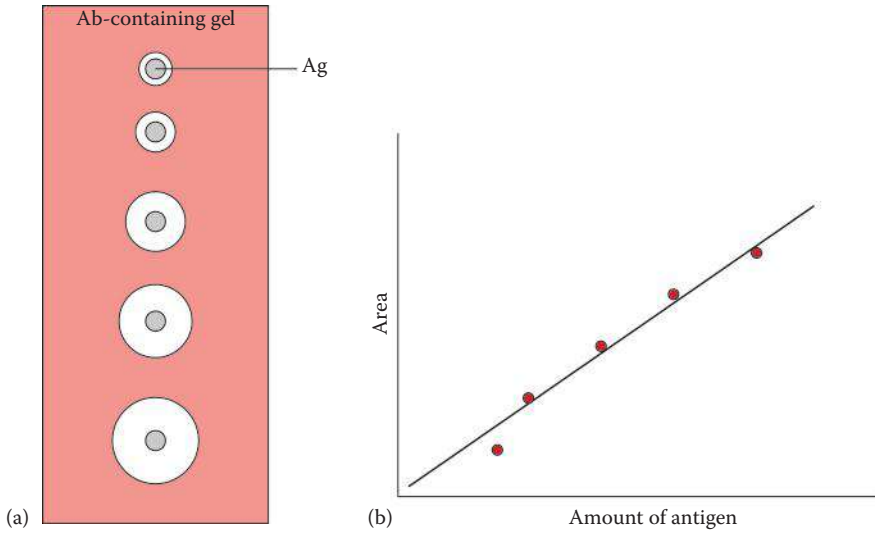

Figure 11.6 Radial immunodiffusion assay. (a) Immunodiffusion of antigens from a well into an antibody-containing gel. (b) Standard curve based on results of standards with known amounts of antigens. (© Richard C. Li.)

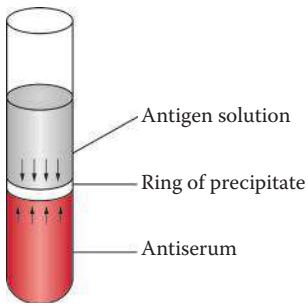

Figure 11.7 Ring assay. (© Richard C. Li.)

The *Ouchterlony assay* is named after the Swedish immunologist, Örjan Ouchterlony, who developed it. The assay can be performed in an agarose gel supported by a glass slide or polyester film (Figure 11.8). Wells are created by punching holes in the gel layer at desired locations. Often, a pattern with six wells surrounding a center well is used. The antibody is loaded in the central well while the questioned samples and the controls are loaded in the surrounding wells. The double diffusion of the antigen and the antibody from the wells is allowed to occur during incubation. If the reaction is positive, a precipitate line between wells can be observed at the end of incubation. The precipitate can be stained, enhancing visibility to aid observation. A single assay can compare more than one antigen to determine whether the antigens in question react the same way or differently with the antibody (see Chapter 7). This method is sometimes used to determine whether samples have come from the same or different origins.

11.3.1.2 Immunoelectrophoretic Methods

Diffusion techniques can be combined with electrophoresis to enhance test results. Electrophoresis separates molecules according to the differences in their electrophoretic mobility.

11.3.1.2.1 Immunoelectrophoresis

Immunoelectrophoresis (IEP) is a two-step procedure that can analyze a wide range of antigens. This technique uses electrophoresis to separate the antigen mixture prior to immunodiffusion (Figure 11.9). In the first step, the antigens in a sample are separated using agarose gel electrophoresis.

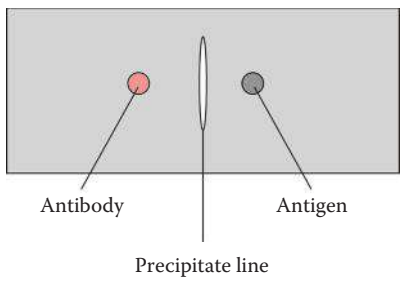

Figure 11.8 Ouchterlony assay. (© Richard C. Li.)

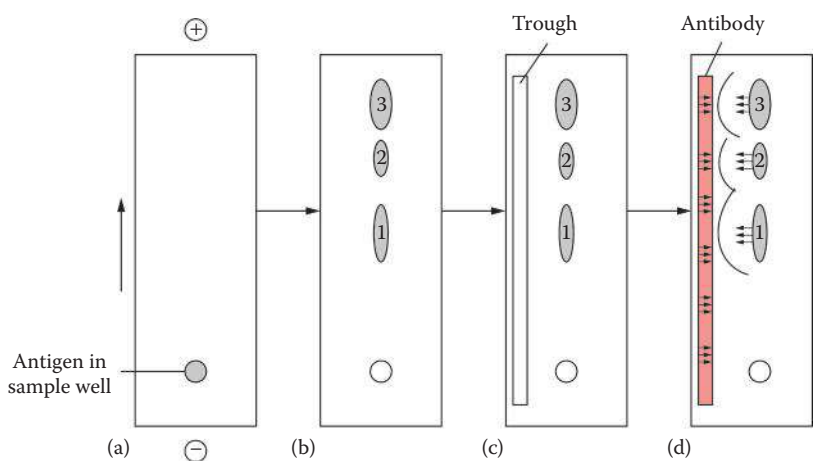

Figure 11.9 IEP assay. (a) Electrophoresis of antigens is carried out. (b) Various antigens are separated after electrophoresis. (c) Trough is cut. (d) Antibody is applied to the trough allowing diffusion to occur and forming precipitate lines. +, anode; −, cathode. (© Richard C. Li.)

Then, a trough is cut in the gel parallel to the array of antigens separated by electrophoresis. In the second step, an antibody is loaded in the trough and the gel is incubated for double diffusion. An arc-shaped precipitate line can be observed for a positive reaction. Multiple precipitate lines can occur if more than one antigen reacts with the antibody. The shapes, intensities, and locations of the precipitate lines of a known control and a questioned sample can be compared.

11.3.1.2.2 Crossed Immunoelectrophoresis

Crossed immunoelectrophoresis (CRIE), also known as two-dimensional IEP, is a modification of IEP. The first step utilizes electrophoresis to separate antigens contained in a sample (Figure 11.10). A strip of gel containing separated antigens is cut for the second round of electrophoresis. The gel including the gel strip is turned at a 90° angle and is further separated by a second-dimension electrophoresis. This drives the antigens from the gel strip into an agarose gel that contains uniformly distributed antibodies. Following the second-dimension electrophoresis, an arc-shaped precipitate line is formed. The area of the arc can be measured, and a sample can be identified by comparison with a known standard. This technique is more sensitive than IEP.

11.3.1.2.3 Rocket Immunoelectrophoresis

An antibody-containing agarose gel is used. The antigen is loaded into the well (Figure 11.11). Electrophoresis then drives the antigen from the well into the agarose gel. In a positive reaction,

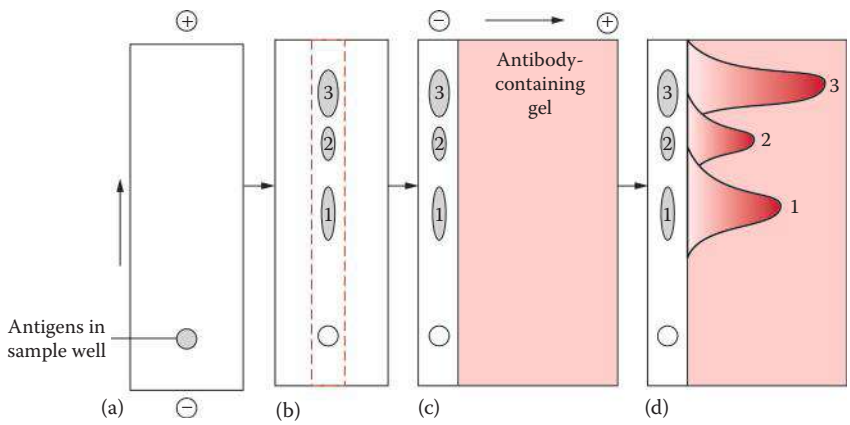

Figure 11.10 CRIE assay. (a) Electrophoresis of antigens is carried out. (b) Various antigens are separated after electrophoresis. A strip of gel containing separated antigens is excised and then used for second-dimension electrophoresis. (c) As shown, second-dimension electrophoresis is carried out to drive the antigens into an antibody-containing gel; (d) arc-shaped precipitate lines are formed as a result. (© Richard C. Li.)

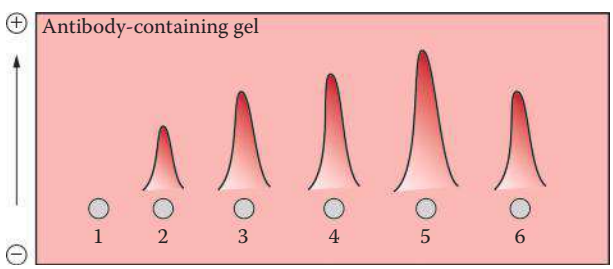

Figure 11.11 Rocket immunoelectrophoresis assay. Antigens are driven from the wells into an antibody-containing gel forming precipitate lines. (© Richard C. Li.)

a rocket-shaped precipitate line can be observed. The height of the rocket is in proportion to the amount of antigen in the sample. Quantitation can be achieved by comparing standards and the sample in the same gel.

11.3.1.2.4 Crossed-Over Immunoelectrophoresis

This technique is also known as counterimmunoelectrophoresis (CIE). Two arrays of opposing wells are created by punching holes in the agarose gel (Figure 11.12). The antibody and samples are loaded in opposing wells arranged by pairs. Electrophoresis is used to drive the antigen and the antibody toward each other. The wells containing the antibody should be proximal to the anode and the wells containing the samples should be proximal to the cathode. During gel electrophoresis, the antigen, which is usually negatively charged, migrates toward the anode. The antibody migrates in the opposite direction as a result of electroendosmosis—a phenomenon in which the movement of molecules is caused by fluid flow. A precipitate line is formed between the opposing wells if the antigen reacts with its specific antibody.

11.3.2 Agglutination-Based Assays

Agglutination reactions can be used as forensic serological assays such as for blood group typing (Chapter 18) and menstrual blood identification (Chapter 16). Agglutination assays are qualitative, indicating the absence or presence of antigens or antibodies. Semiquantitative results can be obtained by titration (diluting the antigen or antibody). Many types of agglutination reactions are available; only those used in forensic serology are discussed next.

Direct agglutination assays involve reactions in which an antibody interacts with antigens originally located on cell surfaces. In a *hemagglutination* reaction (Figure 11.13a), an antibody binds to the antigens located on erythrocytes. This method is used for the identification of blood types, for example, the testing of erythrocytes for ABO blood group typing. The assay can be carried out on a glass slide and the agglutination, as indicated by forming cell clumps, can be observed under a microscope (Chapter 18). The assay can also be carried out in a test tube. In a positive reaction, agglutinated cells are formed on the bottom of the test tube. Test tubes can also be centrifuged and then swirled to determine whether the cell clumps can be resuspended. Agglutinated cells cannot be resuspended. In a negative reaction, unagglutinated cells can be resuspended.

In *agglutination inhibition assays*, the presence of an antigen in question is indirectly detected (Figure 11.13b). If a known antigen is added to a mixture consisting of antigen-containing cells and antibodies, the added antigen will compete with the antigen located on the cell surfaces for antibody binding and inhibit the agglutination reaction. Another indirect agglutination assay is the absorption–elution assay (Chapter 18), which can be used for ABO blood group typing.

Passive agglutination assays are different from direct agglutination assays, in which the antigen is coated on the surface of carrier cells such as tannic acid–treated sheep erythrocytes. Tannic acid is considered to be a fixative. Treating with tannic acid can stabilize the carrier cells

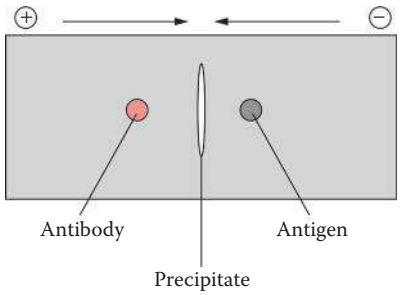

Figure 11.12 Crossed-over immunoelectrophoresis assay. (© Richard C. Li.)

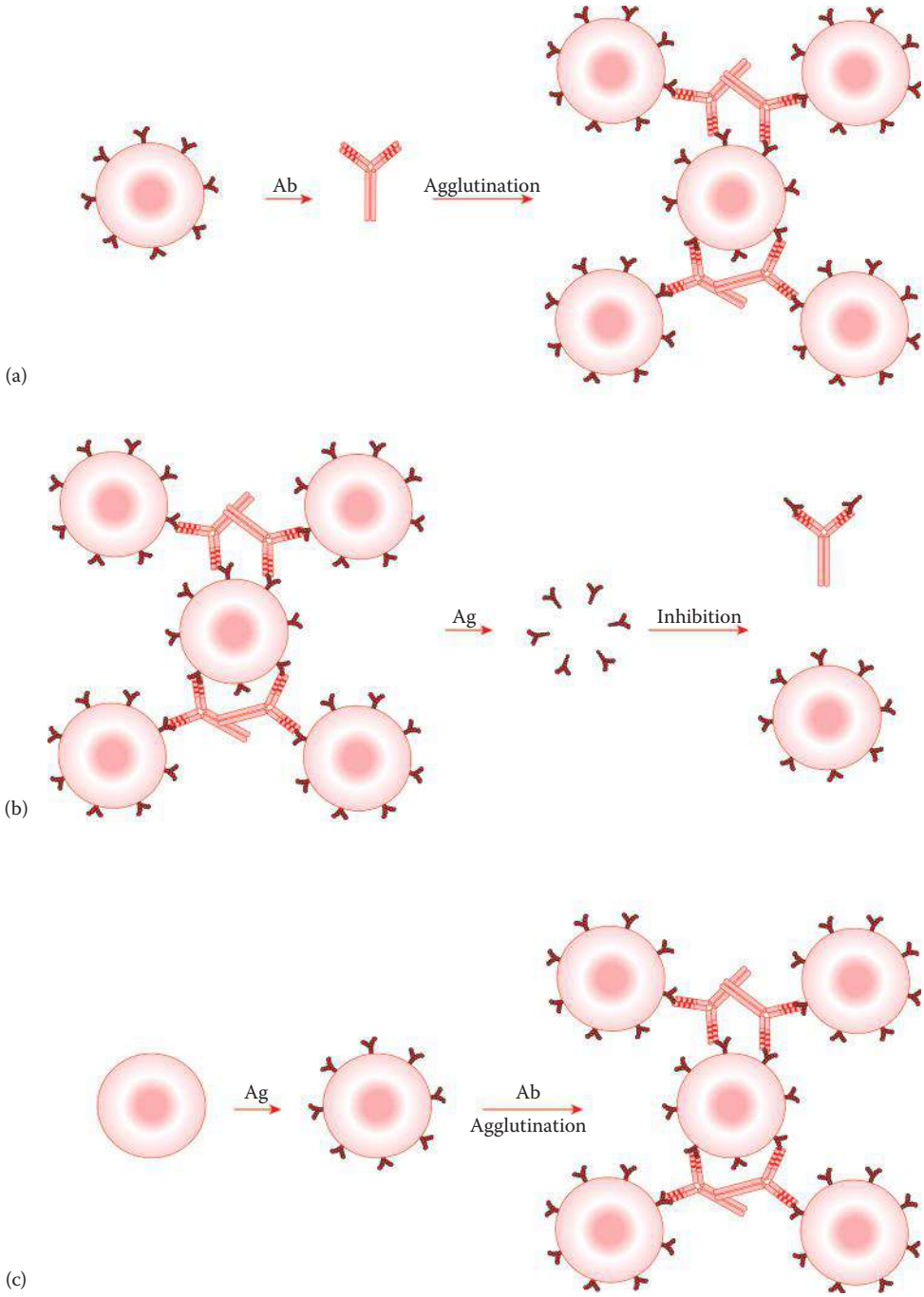

Figure 11.13 Agglutination-based assays. (a) Direct agglutination assay and (b) agglutination inhibition assay. Hemagglutination reactions are shown; antigens are located on erythrocytes. (c) Passive agglutination assay. (© Richard C. Li.)

for a subsequent coating of antigens. The carrier cells are then incubated with samples containing antigens. A coating of antigens can be formed on the surfaces of carriers. Such antigen-coated carrier cells can then be agglutinated using an antibody that is specific to the absorbed antigens (Figure 11.13c).

11.4 DNA Methylation Assays for Bodily Fluid Identification

In the eukaryotic genome, methylation occurs at the cytosine residues commonly in the CpG dinucleotide sequences of both DNA strands. The "p" in CpG refers to the phosphodiester bond between cytosine and guanine. Cytosine methylation is carried out *in vivo* by methyltransferase (DNMT). DNMT catalyzes the transfer of a methyl group (CH_3) from S-adenosylmethionine (SAM) to the C^5 position of cytosine (Figure 11.14). As a result, cytosine is converted to 5-methylcytosine. Cytosine methylation is observed throughout the eukaryotic genomes in both the coding and intergenic regions located between genes. However, cytosine methylation is rare at CpG islands, which are stretches of DNA containing a high frequency of CpG dinucleotides in the regulatory regions of genes (Figure 11.15). It is implicated that cytosine methylation alters the chromatin structure and causes chromatin condensation. As a result, gene expression is usually inhibited by cytosine methylation.

Genomic loci, known as *tissue-specific differentially methylated regions* (TDMRs), have been identified. TDMRs show differential DNA methylation patterns between different tissues. For example, semen-specific TDMRs are consistently hypomethylated in spermatozoa, which can be used to identify semen samples. Other TDMRs display varying degrees of methylation in different types of bodily fluids. This suggests that differential DNA methylation assays may be potentially useful for bodily fluid identification.

The detection of DNA methylation can be carried out by a number of methods. *Methylation-sensitive restriction enzyme digestion polymerase chain reaction* (MSRE-PCR) can be used for the

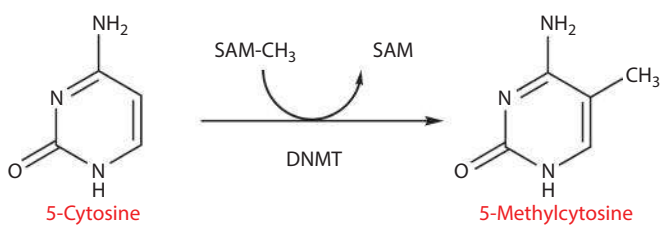

Figure 11.14 Eukaryotic cytosine methylation catalyzed by methyltransferase.

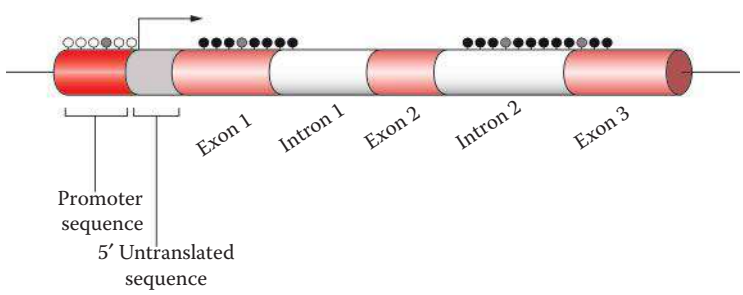

Figure 11.15 Methylation at the cytosine residues in CpG dinucleotide sequences. Note that the CpG dinucleotide sequences are usually unmethylated in the promoter regions. Black circle: methylated. Clear circle: unmethylated. Gray circle: partial methylated. (© Richard C. Li.)

rapid detection of DNA methylation. This method utilizes the *methylation-sensitive restriction enzyme* (MSRE). MSREs cleave DNA at unmethylated restriction sites, but they are not able to cleave once a cytosine residue is methylated. Thus, the methylated DNA remains intact. After PCR amplification, the methylated DNA is amplified and the amplicon is detected. In contrast, the unmethylated DNA is cleaved by the MSRE and cannot be amplified (Figure 11.16). *Bisulfite sequencing* is another widely used method. The genomic DNA is treated with sodium bisulfite, which catalyzes the hydrolytic deamination of unmethylated cytosines (Figure 11.17). As a result, unmethylated cytosines are converted to uracils, while methylated cytosines are resistant to conversion and remain unchanged. Next, uracils are replaced by thymines during PCR. After sequencing, the methylation sites can be deduced at a single-nucleotide resolution by comparing them with the reference sequence (Figure 11.18). *Methyl-DNA immunoprecipitation* (MeDIP) is a useful method for isolating methylated DNA. In this method, DNA is first sheared into fragments. The fragmented DNA is then treated with an antimethylcytosine antibody that binds to methylated cytosines. Methylated DNA fragments can then be isolated by immunoprecipitation for further analysis (Figure 11.19).

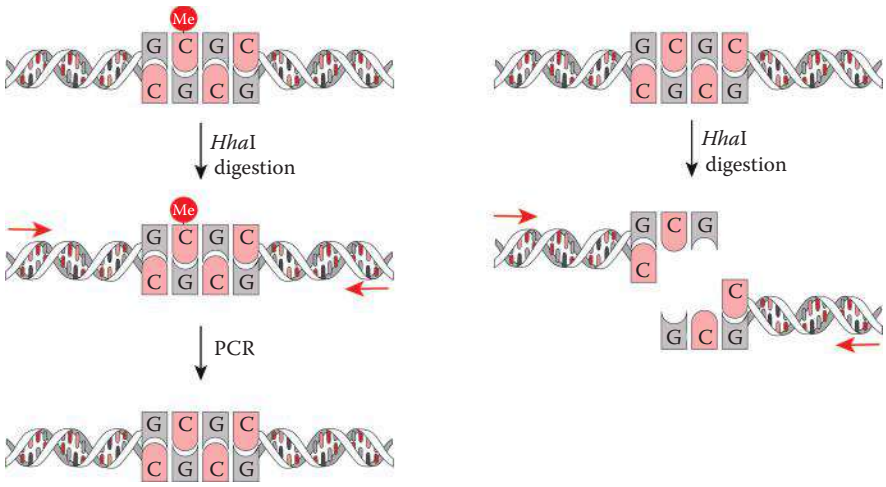

Figure 11.16 Methylation-sensitive restriction enzyme digestion PCR (MSRE-PCR) for the detection of DNA methylation. This method utilizes the methylation-sensitive restriction enzyme (MSRE) such as *Hha*I. *Hha*I cleaves DNA at the unmethylated sequence GCGC but is not able to cleave once a cytosine residue is methylated (GCmGC). The *Hha*I-treated DNA is then PCR amplified. The *Hha*I uncleaved DNA results in an amplicon, indicating the presence of a methylated cytosine in the sequence. The *Hha*I cleaved DNA cannot be amplified, indicating the absence of methylated cytosine in the sequence. Me, 5-methylcytosine site; red arrow, PCR primers. (Adapted from Frumkin, D. et al., *Forensic Sci Int Genet*, 5, 517–524, 2011.)

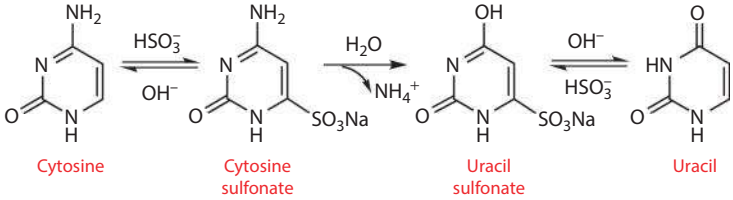

Figure 11.17 Bisulfite conversion of 5-methylcytosine.

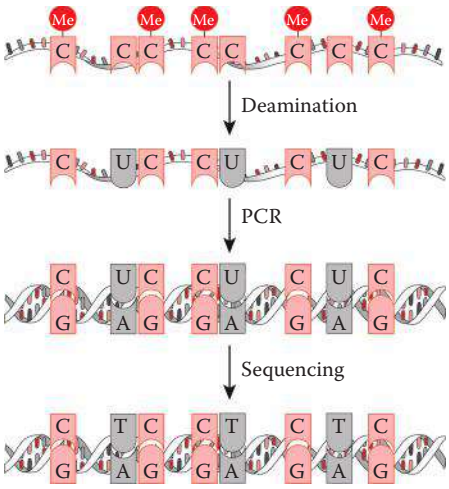

Figure 11.18 Schematic illustration of bisulfite sequencing of cytosine methylation. In a deamination reaction, unmethylated cytosines are converted to uracil. PCR and DNA sequencing are then carried out. The retention of C indicates that the site was methylated; the conversion of C to T indicates that the site was unmethylated in the DNA sample. Me, 5-methylcytosines site. (© Richard C. Li)

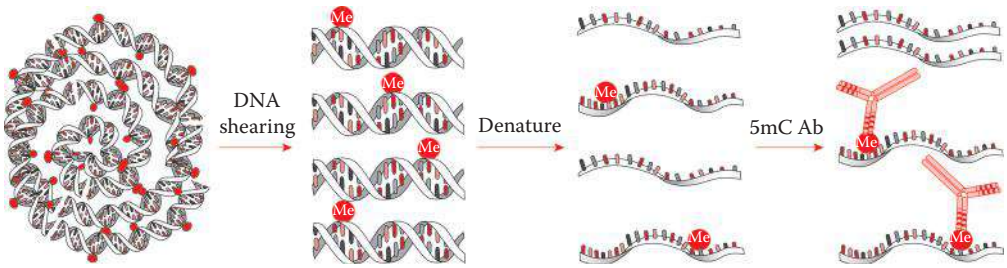

Figure 11.19 Methylated DNA immunoprecipitation (MeDIP) for the detection of methylated DNA. Genomic DNA is randomly and mechanically sheared into fragments. Fragments are then denatured to single-stranded DNA. The fragments containing the methylated region can be immunoprecipitated using an antibody that specifically recognizes 5-methylcytosine. The methylated DNA sequence in the immunoprecipitated fragments can then be determined by DNA sequencing and methylation analysis. Me, 5-methylcytosine site; 5mC Ab, 5-methylcytosine antibody. (© Richard C. Li.)

11.5 Forensic Applications of RNA-Based Assays and RNA Profiling

11.5.1 Messenger RNA-Based Assays

mRNA-based assays have been developed to identify bodily fluids for forensic investigation. In forensic pathological investigation, they can potentially be used for wound age estimation and the age of biological stains. The assays are based on the expression of certain genes in certain cell or tissue types. Candidate tissue-specific genes that may be useful for the identification of bodily fluids have been identified. Thus, the techniques that are used in the identification of bodily fluids are based on the detection of specific types of mRNA that are expressed exclusively in certain cells. The *tissue-specific genes* that are utilized for bodily fluid identification are summarized in Table 11.2. Additionally, *reference genes* that are constitutively expressed housekeeping genes are utilized as internal controls. The amount of mRNA can be assessed by normalizing the target gene to the expression level of the reference genes. Compared with conventional assays that are used for bodily fluid identification, the mRNA-based assay has higher specificity and is amenable to automation. However, one limitation is that the mRNA is unstable due to degradation by

Table 11.2 Representative Markers of mRNA-Based Assays for Bodily Fluid Identification

Bodily Fluid	Gene Symbol	Description
Blood	*ALAS2*	Aminolevulinatesynthase 2
	AMICA1	Adhesion molecule, interacts with CXADR antigen 1
	ANK1	Ankyrin1
	CD3G	CD3 gammamolecule
	CD93	CD93 molecule
	HBA1	Alpha 1 hemoglobin
	HBB	Beta hemoglobin
	PBGD	Porphobilinogen deaminase
	SPTB	Beta spectrin
Saliva	*HTN3*	Histatin 3
	MUC7	Mucin 7
	STATH	Statherin
Semen	*KLK3*	Kallikrein 3 (prostate-specific antigen)
	PRM1	Protamine 1
	PRM2	Protamine 2
	SEMG1	Semenogelin 1
	TGM4	Transglutaminase 4
Vaginal secretions	*CYP2B7P1*	Cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1
	DKK4	Dickkopf homolog 4
	FUT6	Fucosyltransferase 6
	HBD1	Beta defensin 1
	IL19	Interleukin 19
	MUC4	Mucin 4
	MYOZ1	Myozenin 1
	SFTA2	Surfactant associated 2
Menstrual blood	*MMP7*	Matrix metalloproteinase 7
	MMP11	Matrix metalloproteinase 11
Skin	*CDSN*	Corneodesmosin
	LOR	Loricrin

(*continued*)

Table 11.2 (Continued) Representative Markers of mRNA-Based Assays for Bodily Fluid Identification

Bodily Fluid	Gene Symbol	Description
Reference gene	*18S rRNA*	18S ribosomal RNA
	ACTB	Beta actin
	G6PDH	Hexose-6-phosphate dehydrogenase
	RPS15	Ribosomal protein S15

Sources: Adapted from Bauer, M. and Patzelt, D., *J Forensic Sci*, 47, 1278–1282, 2002; Hanson, E.K. and Ballantyne, J., *Sci Justice*, 53, 14–22, 2013; Lindenbergh, A., Maaskant, P., and Sijen, T., *Forensic Sci Int Genet*, 7, 159–166, 2013; Nussbaumer, C., Gharehbaghi-Schnell, E., and Korschineck, I., *Forensic Sci Int*, 157, 181–186, 2006.

endogenous ribonucleases. Additionally, bodily fluid stains collected from crime scenes are often exposed to ultraviolet (UV) light, moisture, and high temperature, which can promote mRNA degradation. Nevertheless, the successful detection of mRNA from aged samples is possible.

11.5.2 MicroRNA-Based Assays

The biological function of microRNAs (miRNAs) is to regulate gene expression. The mature miRNA strand associated with the RNA-induced silencing complex (RISC) binds to its target mRNA. In animal cells, the mature miRNA forms a base pairing with its complementary sequence of the *miRNA responsive element* (MRE) within the 3′ untranslated regions (3′UTRs) of the target mRNA (Figure 11.20). A single miRNA can bind different mRNA transcripts encoded by multiple genes. Animal miRNAs usually form a base pairing with their target mRNAs through partial complementary sequences, which lead to translation repression to inhibit protein synthesis. If the base pairing is exactly complementary to its target mRNA sequence, then the cleavage and the degradation of the target mRNA occurs. As a result, the synthesis of the protein encoded by the mRNA is reduced. Thus, miRNAs play important roles in cellular function.

miRNAs have potential applications in forensic identification. Recent studies have revealed that miRNAs can be detected from various bodily fluids such as blood, saliva, semen, vaginal secretions, and menstrual blood. The type of miRNA expressed is not exclusively tissue- or cell-type specific. However, the expression patterns of specific miRNAs are unique for various bodily fluids. Thus, a bodily fluid in question can potentially be identified using multiple differentially expressed miRNAs as markers (Table 11.3). miRNA markers for bodily fluid identification have their advantages compared with those of mRNA. Mature miRNAs are much

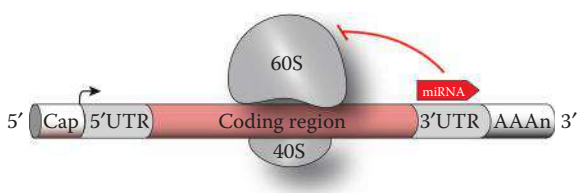

Figure 11.20 Schematic illustration of the roles of miRNA in translational inhibition. In animal cells, miRNAs bind to 3′UTR of their target mRNA with partially complementary Watson–Crick base pairing. As a result, the binding of miRNA to the target mRNA inhibits translation of the target mRNA. Eukaryotic 40S and 60S ribosomal subunits are shown. (© Richard C. Li.)

Table 11.3 Differentially Expressed miRNA Markers for Bodily Fluid Identification

Biological Fluid	miRNA Set
Blood	miR16; miR451
	miR16; miR486
	miR20a; miR106a; miR144; miR185
	miR126; miR150; miR451
Semen	miR10b; miR135b
	miR10a; miR135a; miR943; miR507; miR891a
Saliva	miR205; miR658
	miR200c; miR203; miR205
Vaginal secretions	miR124a; miR372
Menstrual blood	miR214
	miR412; miR451
Reference gene	*RNU6-2* (U6 small nuclear 2 RNA)

Sources: Adapted from Courts, C. and Madea, B., *J Forensic Sci*, 56, 1464–1470, 2011; Hanson, E.K., Lubenow, H., and Ballantyne, J., *Anal Biochem*, 387, 303–314, 2009; Wang, Z. et al., *Forensic Sci Int Genet*, 7, 116–123, 2013; Zubakov, D. et al., *Int J Legal Med*, 124, 217–226, 2010.

smaller, approximately 21–23 nucleotides in length, than mRNA. As a result, miRNAs are less susceptible to degradation and more stable than mRNAs. Additionally, the sensitivity of the miRNA-based method is much higher than that of the mRNA-based methods.

11.6 Proteomic Approaches Using Mass Spectrometry for Bodily Fluid Identification

The forensic application of proteomic approaches is to identify biomarker proteins derived from bodily fluids to determine the type of bodily fluid in question for forensic investigations. Mass spectrometry (MS) is a highly sensitive and rapid technique for protein identification from complex biological samples.

11.6.1 Mass Spectrometric Instrumentation for Protein Analysis

A mass spectrometer is an analytical technique to identify a molecule by measuring the ratio of the mass (m) to the charge (z) of a charged molecule (Figure 11.21). A typical mass spectrometer instrument analyzes ionized molecules while in its gas phase. A mass spectrometer consists of an *ion source* that converts analytes into gaseous phase ions, through a process known as *ionization*, by gaining a positive or a negative charge from a neutral species. The ions are then introduced into the *mass analyzer* of a mass spectrometer, where ions are accelerated under electric fields and separated based on their m/z ratio. The separated ions are detected by an *ion detector* that records the number of ions at each m/z value. The result of molecular ionization,

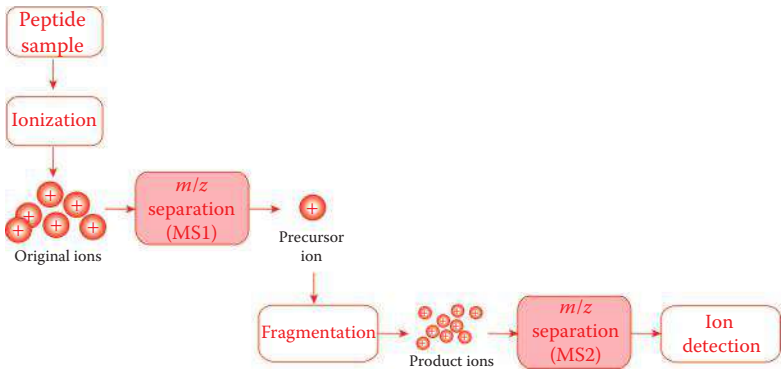

Figure 11.21 Diagram of tandem mass spectrometry. For the initial mass analysis of thermolabile analytes including peptides and proteins, soft ionization techniques such as matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) can be used to produce original ions. MALDI utilizes a laser beam to trigger ionization and facilitate the vaporization of the analyte. In the ESI method, a fine mist of droplets is formed in the presence of a high electric field, producing charged ions. In the first round of mass analysis, ions are separated based on their *m/z* ratio. A particular ion, called a precursor ion, is then selected for fragmentation, generating product ions. The fragmentation can be carried out using collision techniques such as collision-induced dissociation (CID) by multiple collisions with rare gas atoms. Additionally, a new fragmentation technique, electron-capture dissociation (ECD), can be used in which the fragmentation is induced through the capture of a thermal electron by a protonated peptide cation. The product ions are analyzed by the second round of mass analysis and detected by an ion detector. MS, mass analyzer. (© Richard C. Li.)

ion separation, and ion detection is a spectrum that can be used to determine the mass (or molecular weight) and potentially the structure of the molecule.

Furthermore, biomarker proteins can be identified by *peptide sequencing* using the tandem mass spectrometry (MS/MS) mode of operation in which a mass spectrometer uses two or more mass analyzers. After an initial mass analysis, individual peptide ions in the mass spectrum can be selected. The selected ions are known as *precursor ions*. A precursor ion is then subjected to a second round of fragmentation through collision and is broken into smaller ions known as *product ions*. The resulting product ions are further analyzed by MS. Thus, the corresponding protein can be identified through searching a protein database, based on the mass of a specific peptide obtained.

11.6.2 Analysis Strategies for Protein Identification

To date, various MS-based protein identification strategies have been developed, which can be divided into two categories: "top-down" and "bottom-up" strategies (Figure 11.22). In the *top-down* strategy, intact proteins in a complex mixture are fractionated and separated into less complex protein mixtures or single proteins. Intact proteins are then analyzed using MS. A liquid chromatography–mass spectrometry (LC–MS) system can be used for such an application. The advantage of this approach is that an entire protein is analyzed for identification purposes. However, the throughput and efficiency of this approach is still a major challenge for large-scale sample analysis in forensic applications.

The *bottom-up* strategy is commonly used for high-throughput analysis of small peptides derived from highly complex samples. Current MS usually cannot resolve the mass measurement of large proteins; rather, MS is ideal for the analysis of small peptides. One approach is the *sort-then-break* approach in which proteins in a complex mixture are first isolated and then enzymatically cleaved. The target proteins are isolated from their complex biological source using 1-D or 2-D gel electrophoresis (Chapter 18). The isolated proteins are then enzymatically

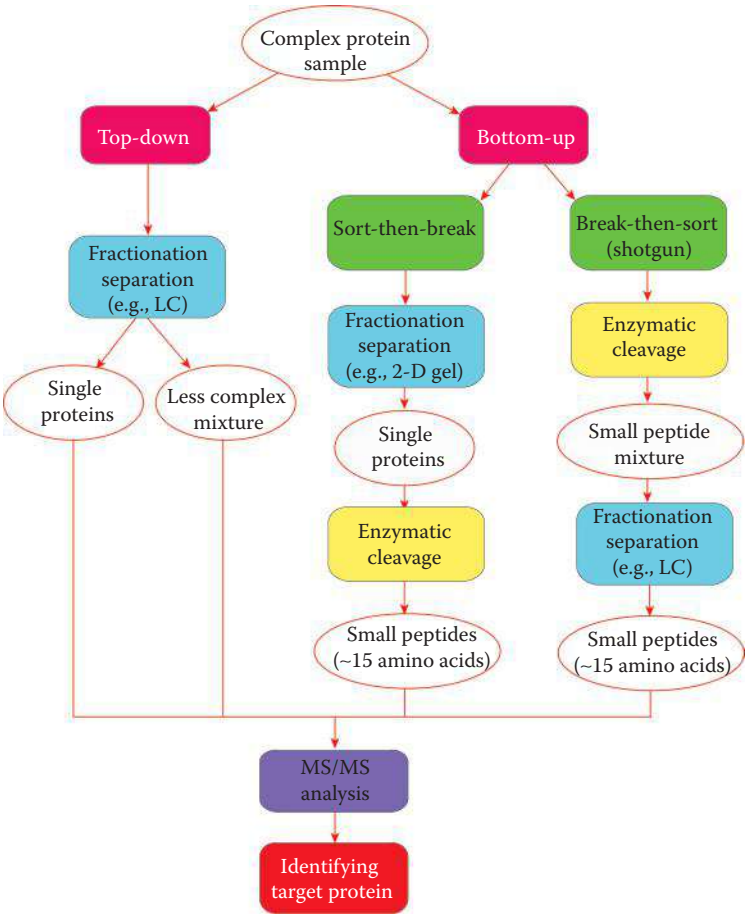

Figure 11.22 Top-down and bottom-up strategies for MS-based protein identification. LC, liquid chromatography; MS/MS, tandem mass spectrometry. (© Richard C. Li.)

cleaved into small peptides, approximately 15 amino acids in length, using proteases such as trypsin. These peptides are then analyzed by peptide sequencing using tandem MS to identify the target protein.

Break-then-sort, also known as the *shotgun* approach, is another approach of the bottom-up strategy. In this approach, proteins are cleaved first and the resulting peptide mixture is then separated by LC and analyzed by tandem MS (usually coupled to the LC). In a complex mixture, a peptide mass measurement alone is not sufficient for the identification of a target protein. Database searches are carried out utilizing software that compares the observed MS peptide spectra against all candidate spectra of the proteins that are present in a protein sequence (or translated from genomic DNA sequences) database for possible matches. Thus, a biomarker protein can be identified.

11.7 Microbial DNA Analysis for Bodily Fluid Identification

Human bodies harbor a large number of microbes. It is estimated that there are over 10 times more microbial cells than human cells in and on the human body, in the nose, mouth, throat, intestine, vagina, and skin. This microbial community is referred to as the human *microbiota*.

It is known that the microbiota has an important influence on human health. A decade ago, the concept of the *microbiome* was proposed by Nobel laureate Joshua Lederberg to identify and characterize the microbiota, their genomes, and their environmental interactions in a defined community. Traditionally, the studies of microbial genomes rely on cultivated individual species. The disadvantage of cultivation-based methods, however, is that large numbers of microbial species cannot be successfully isolated because many microbial species require specific growth conditions. Based on the advances in DNA sequencing technologies, a new approach, known as *metagenomics*, has been developed. Instead of studying the genome of an individual cultivated microbial species, metagenomics allows the study of the microbial genome of complete microbial communities harvested directly from natural environments. Metagenomics has been a powerful tool to analyze the human microbiome. Recently, human microbiota have been surveyed, revealing that microbial diversity is greatest in the teeth and the intestines, intermediate on the skin and the inside surface of the cheek, and lowest in the vagina (Table 11.4). More interestingly, each habitat is characterized by a small number of highly predominant bacterial taxa. The predominant taxa of each habitat can potentially be utilized for the forensic analysis of biological evidence. Bodily fluids identification is an important task in the forensic analysis of biological evidence. The detection of predominant taxa of bacteria can aid in the forensic identification of a particular type of bodily fluid (Table 11.4).

Table 11.4 Representative Microbiota that are Potentially Can Be Used for Forensic Identification

Body Habitats	Estimated Number of Microbial Species	Predominent Taxa	Potential Forensic Applications	DNA Marker	Further Reading
Cheek	800	*Streptococcus* spp.	Saliva identification	*gtf*	Nakanishi et al. (2011)
			Expirated blood	*gtf*	Donaldson et al. (2010)
				16S rRNA	Power et al. (2010)
Vagina	300	*Lactobacillus* spp.	Vaginal secretion identification	16S rRNA	Akutsu et al. (2012)
				16S–23S ISR	Fleming and Harbison (2010)
					Giampaoli et al. (2012)
Gastrointestinal tract	4000	*Bacteroides* spp.	Fecal matter identification	*rpoB*	Nakanishi et al. (2013)
Skin	1000	*Staphylococcus* spp.	Touched evidence identification	16S rRNA	Fierer et al. (2010)

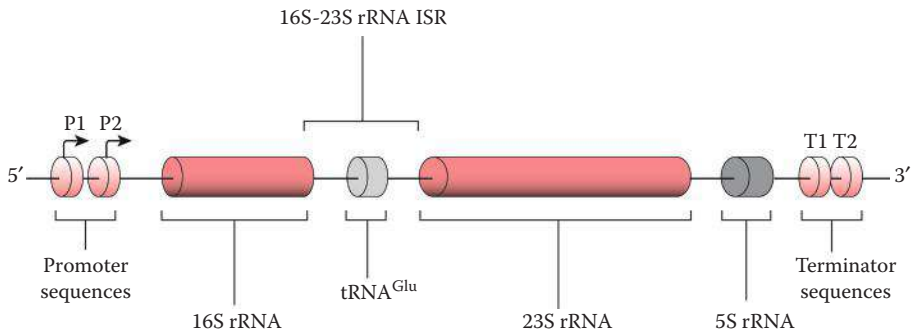

Figure 11.23 Schematic diagram of an *Escherichia coli* rRNA operon. *E. coli* has seven rRNA operons. A typical rRNA operon is shown. The rRNA operon encodes for three ribosomal RNA (rRNA) genes: the 5S and 23S genes code for the large ribosomal subunit, while the 16S gene encodes for the small ribosomal subunit. *E. coli* encodes one tRNA gene in the rRNA operon. The intergenic spacer regions (ISRs) and the promoter and terminator sequences are also shown. The rRNA operon is transcribed into a single precursor RNA transcript that is cleaved into separate rRNA and tRNA transcripts. The diagram is not to scale to the physical map of rRNA operons. (© Richard C. Li.)

The detection of bacteria can be achieved by utilizing DNA markers. The most commonly used DNA markers are the essential genes present in all bacterial species. These markers have highly variable regions that can be used to distinguish various bacterial species. The DNA sequences within these variable regions can also suggest taxonomic relationships between different taxa. Additionally, DNA markers contain conserved regions that are necessary for use as primer-binding sites for PCR-based detection.

For example, the rRNA genes (Figure 11.23), encoding for the ribosomal RNA, are conserved in all living organisms including bacteria. The bacterial rRNA operon contains three rRNA genes: the 5S and 23S rRNA genes encode the RNA components of the large subunit of the ribosome, while the 16S rRNA gene encodes the RNA component of the small subunit of the ribosome. The 16S rRNA gene is the most commonly used marker for the taxonomic identification of bacteria. The bacterial 16S rRNA genes are similar in length, approximately 1.5 kb. They contain highly conserved regions that can be used as primer-binding sites for the amplification of the adjacent DNA regions. Additionally, they contain variable regions that allow for taxonomic identification. In forensic applications, they are used as a marker for the identification of vaginal bacteria as an indicator of vaginal secretions (Chapter 16). However, the number of variable regions in the 16S rRNA genes of some bacteria species is limited. Furthermore, closely related species have high levels of sequence similarity to the 16S rRNA genes. Thus, the 16S rRNA gene marker is not adequate for distinguishing closely related species.

DNA markers of bacteria identification can also be located in noncoding regions of DNA. The *intergenic spacer region* (ISR) between the 16S rRNA and 23S rRNA genes in the rRNA operon is another commonly used marker. The 16S–23S rRNA ISR contains both length and sequence variations between species. These variations are partially caused by the number and type of tRNA genes that this region of operon contains. For example, most gram-negative bacteria contain both a copy of the $tRNA^{ala}$ and the $tRNA^{ile}$ genes, while some contain only a copy of the $tRNA^{glu}$ gene. In contrast, most gram-positive bacteria have no tRNA gene; some contain either a copy of the $tRNA^{ala}$ gene or the $tRNA^{ile}$ gene or both. Thus, heterogeneity in the length and in the sequence of the 16S–23S rRNA ISR allows bacteria to be identified at the species level.

In addition to rRNA genes, many other genes are utilized as DNA markers for bacterial species identification. For example, the *rpoB* gene encodes the β subunit of bacterial RNA polymerase and is used as a marker for the forensic identification of fecal matter (Chapter 17). The sequence variation of the *rpoB* gene allows the distinguishing of species when they are not distinguishable, using 16S rRNA gene sequences. Several oral streptococcal species that produce extracellular polysaccharides contain the *gtf* gene, which encodes glucosyltransferase. Thus, the *gtf* gene is a useful marker for oral streptococcal species identification as a potential indicator of saliva in forensic investigations. However, this marker cannot be used to identify bacteria other than *streptococcus*. Nevertheless, combining several DNA markers can be used for the identification of bacteria species.

11.8 Nondestructive Assays for the Identification of Bodily Fluids

Most of the bodily fluid identification techniques mentioned previously consume a portion of the evidence. When the amount of sample evidence is very limited, such destructive assays may consume the evidence that is also needed for subsequent forensic DNA analysis. Moreover, most of these tests are used for a single type of bodily fluid. Tests for multiple bodily fluids in question will consume additional amounts of the evidence. Therefore, nondestructive identification techniques are highly desired. Additionally, it is important to develop an assay that can be utilized to test multiple types of bodily fluids. Furthermore, a portable device that can identify bodily fluids at a crime scene is very useful. It provides investigators with test results immediately to aid the crime scene investigations. Two techniques have been developed that can potentially be used for nondestructive bodily fluid identification: fluorescence spectroscopy and Raman spectroscopy.

Fluorescence spectroscopy is one example of a nondestructive technique. Fluorescence is the emission of light by a fluorophore. A fluorophore is a moiety in a molecule that fluoresces on absorbing the energy from an excitation light source or radiation. The emitted light usually has a longer wavelength and lower energy than the absorbed radiation. Constituents present in bodily fluids, such as nucleic acids, proteins, lipids, and metabolites, can exhibit fluorescence. The unique composition of a bodily fluid emits characteristic emission spectra, thus making it identifiable. Using multiple wavelengths of an excitation light source, fluorescent emissions at a wide range of wavelengths can be detected, which allows for the identification of various bodily fluids. This technique is sensitive and rapid. It does not utilize chemical reagents and the detection does not require physical contact with a sample. After identification, the same biological sample can be used for forensic DNA analysis. Additionally, portable fluorescence instruments can potentially be used at a crime scene for bodily fluid identification. However, it is not clear if exposure to an excitation light source can damage DNA evidence.

Raman spectroscopy is another example of a nondestructive technique. Raman spectroscopy utilizes a near-infrared excitation light source and measures the scattering of laser light caused by the vibrating molecules of a sample. A typical Raman spectrum provides information about the molecular structure and the "signature" based on the properties of the constituents of a sample. Thus, it is possible to obtain unique Raman spectra for a particular type of bodily fluid. Raman spectroscopy is highly sensitive. The measurements can be carried out on very small amounts of sample, ranging from a few picoliters to femtoliters. This technique does not require any chemical reagent. Additionally, it does not consume the sample, which can be used for subsequent DNA analysis. Raman spectroscopy has been utilized for various forensic purposes including the identification of drugs, trace evidence such as fibers, and questioned document evidence such as inks and paints. It can potentially be used for the nondestructive identification of various bodily fluids. A portable Raman spectrometer will allow bodily fluids to be identified at crime scenes.

12

Identification of Blood

12.1 Biological Properties

Blood constitutes about 8% of the human body weight of a healthy individual. *Plasma* is the fluid portion of the blood. The cellular portion of the blood consists of *red blood cells*, *white blood cells*, and *platelets*, all of which are suspended in the plasma (Figures 12.1 and 12.2).

12.1.1 Red Blood Cells

These cells are also called *erythrocytes*. Their life span in humans is approximately 3–4 months. Additionally, mature human erythrocytes do not have nuclei, and therefore lack nuclear DNA. Erythrocytes consist of hemoglobin—proteins that are responsible for the transportation of oxygen. Most adult human hemoglobin consists of four polypeptide chains, two α chains and two β chains. Thus, adult hemoglobin is designated as $\alpha_2\beta_2$. Other forms of hemoglobin will be discussed in Chapter 18. Under normal physiological conditions, each hemoglobin subunit contains a heme moiety that binds to oxygen (Figure 12.3). A heme molecule consists of an organic component known as protoporphyrin IX and a *ferrous* (Fe^{2+}) iron ion (Figure 12.4). A heme molecule is also known as *ferroprotoporphyrin*. The ferrous ion of heme forms four bonds with the nitrogens of protoporphyrin IX, along with a fifth bond with a hemoglobin chain and a sixth bond with a molecule of O_2. Other chemicals such as carbon monoxide and cyanide also bind to the ferrous iron of the heme molecule and can cause chemical asphyxia. Heme groups are also present in the blood of various animals and in other proteins such as myoglobin in muscles and neuroglobin in the brain.

12.1.2 White Blood Cells

Also called leucocytes, white blood cells are subdivided into three types: granulocytes, lymphocytes, and monocytes. White blood cells are involved in defending the body against infection. They have nuclei and thus represent the main sources of nuclear DNA from the blood.

12.1.3 Platelets

These cells are also known as thrombocytes, and they play a role in blood clotting (Chapter 16). Platelets aggregate at sites of vascular and blood vessel injury. Like erythrocytes, they lack nuclei.

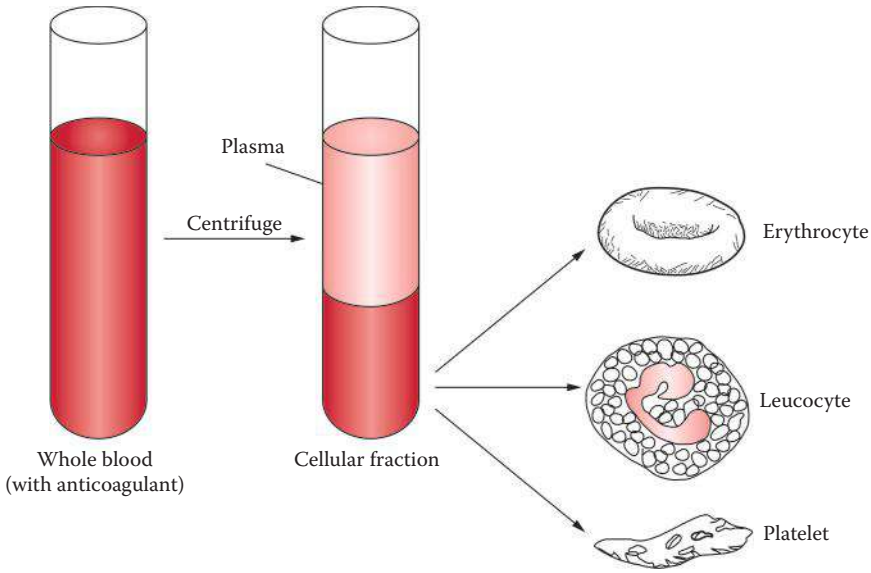

Figure 12.1 Basic composition of blood. Blood can be separated into two phases in the presence of an anticoagulant. The liquid portion called plasma accounts for approximately 55% of blood volume. The cellular elements include erythrocytes, leucocytes, and platelets. (© Richard C. Li.)

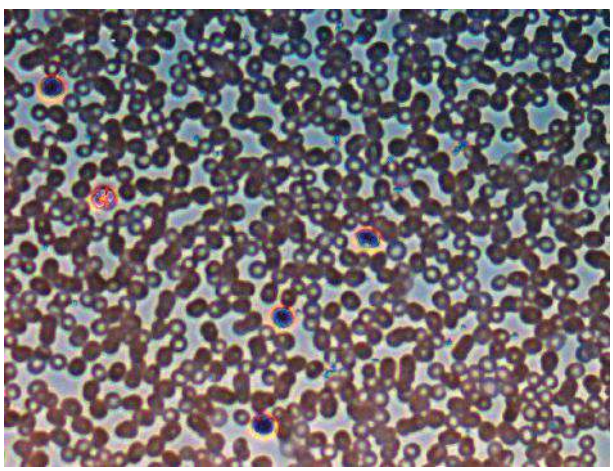

Figure 12.2 Blood cells. (© Richard C. Li.)

12.2 Presumptive Assays for Identification

12.2.1 Mechanisms of Presumptive Assays

Presumptive blood assays are designed to detect traces of blood. These assays are based on the basic principle of the oxidation–reduction reaction catalyzed by the heme moiety of the hemoglobin. As a result, colorless substrates catalyzed by heme undergo an oxidation reaction, causing either chemiluminescence, fluorescence, or a change of color. These assays are very sensitive and can detect blood in samples with 10^{-5}–10^{-6}-fold dilutions. A positive reaction indicates the possible presence of blood. Additionally, most of these assays do not interfere with forensic DNA analysis.

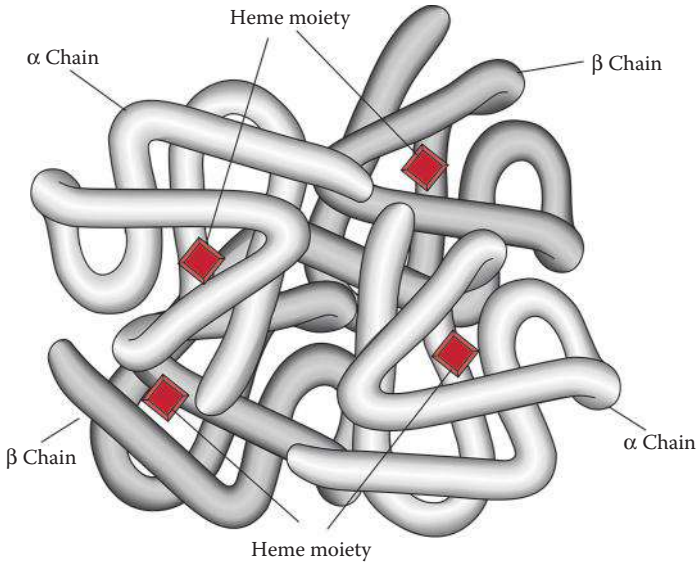

Figure 12.3 Human adult hemoglobin. Four subunits, two α and two β chains, of human adult hemoglobin are shown. Each hemoglobin subunit contains a heme moiety. (© Richard C. Li.)

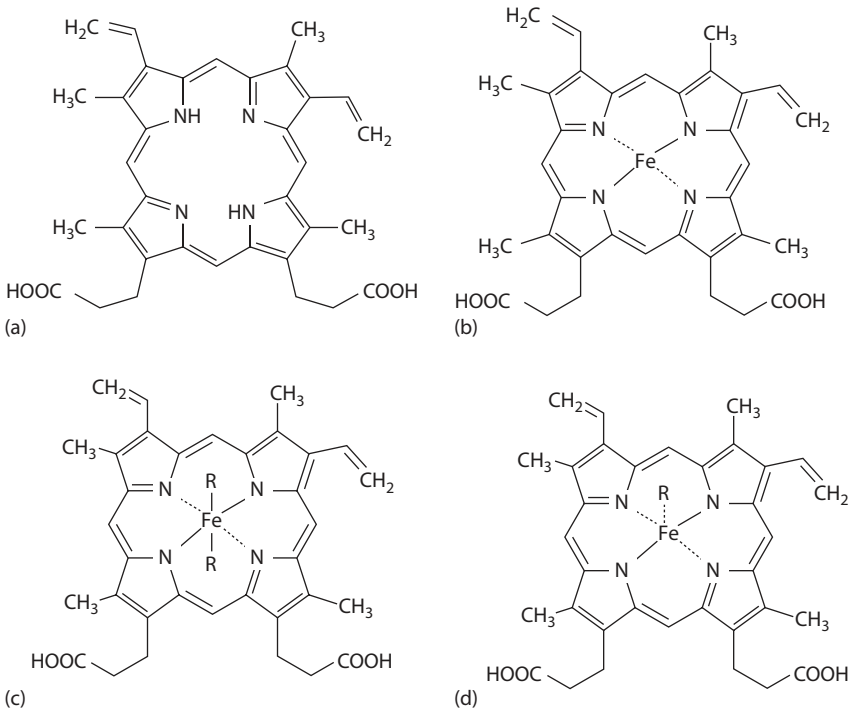

Figure 12.4 Chemical structures of heme and its precursor and derivatives. (a) Protoporphyrin IX. (b) Heme (ferroprotoporphyrin). (c) Hemochromagen, R = pyridine (pyridineferroprotoporphyrin). (d) Hematin hydroxide, R = OH (ferriprotoporphyrin hydroxide); hematin chloride, R = Cl (ferriprotoporphyrin chloride).

12.2.1.1 Oxidation–Reduction Reactions

An oxidation–reduction reaction involves a change in the oxidation state of a molecule. Specifically, the *oxidation* of a molecule means that the molecule has lost electrons, and the *reduction* of a molecule means that the molecule has gained electrons. Chemicals that can be reduced and therefore gain electrons from other molecules are called *oxidants*. In contrast, *reductants* are chemicals that can be oxidized and therefore lose electrons to other molecules. In biochemical reactions, oxidation often coincides with a loss of hydrogen. Figure 12.5 depicts an example of an oxidation–reduction reaction for blood identification. In presumptive assays, heme is utilized as the catalyst, and hydrogen peroxide is utilized as the oxidant for the reaction. In the presence of heme, a colorless substrate is oxidized, yielding a product with color, chemiluminescence, or fluorescence.

12.2.2 Colorimetric Assays

Many procedures are available for detecting heme in blood through color reactions. The most common agents are phenolphthalin, leucomalachite green, and benzidine derivatives. The color reactions produced by these assays can be observed immediately with the naked eye.

12.2.2.1 Phenolphthalin Assay

Phenolphthalein, a member of a class of indicators and dyes, is used in titrations of mineral and organic acids as well as most alkalis. The phenolphthalin assay for blood identification is also known as the Kastle–Meyer test. Kastle published a study in 1901, presenting the results of a reaction in which phenolphthalin, a colorless compound, is catalyzed by heme with hydrogen peroxide as the oxidant (Figures 12.6 and 12.7). The oxidized derivative is phenolphthalein, which appears pink under alkaline conditions.

12.2.2.2 Leucomalachite Green (LMG) Assay

Malachite green is a triphenylmethane dye. The leuco base form of malachite green is colorless and can be oxidized by the catalysis of heme to produce a green color. The reaction is carried out under acid conditions with hydrogen peroxide as the oxidant (Figures 12.8 and 12.9).

12.2.2.3 Benzidine and Derivatives

Historically, *benzidine* was used as an intermediate in dye manufacturing (Figure 12.10). Subsequently, it was used as a presumptive assay for the presence of blood after the discovery that the oxidation of benzidine can be catalyzed by heme to produce a blue to dark blue color (carried out in an acid solution). Since the blue color may eventually turn brown, the reaction must be read

$$\underset{\text{(Colorless)}}{AH_2} + H_2O_2 \xrightarrow{\text{Heme}} \underset{\text{(Color)}}{A} + 2H_2O$$

Figure 12.5 Oxidation–reduction reaction as the basis for presumptive assays for blood identification. AH_2, substrate; A, oxidized substrate.

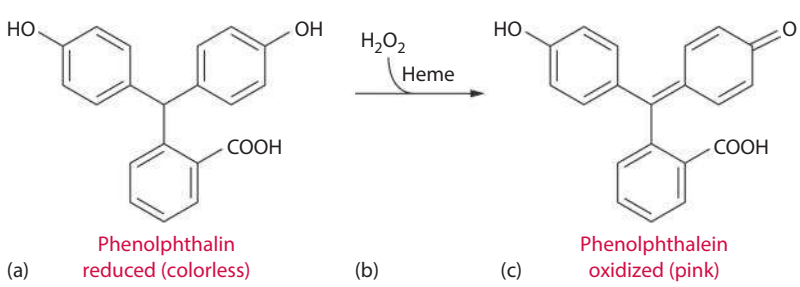

Figure 12.6 Chemical reaction of phenolphthalin assay.

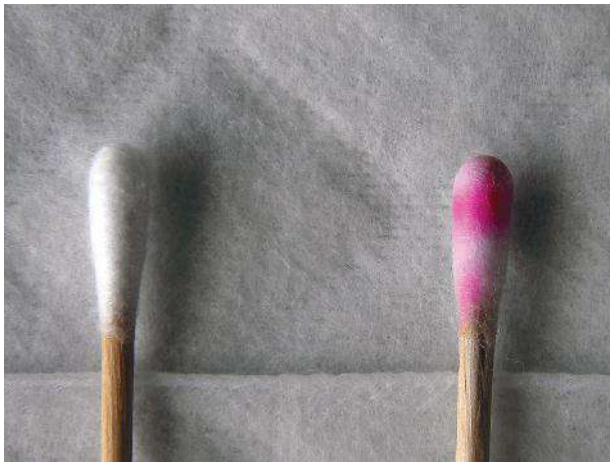

Figure 12.7 Photograph of phenolphthalin assay results. Negative (left) and positive (right) reaction. (© Richard C. Li.)

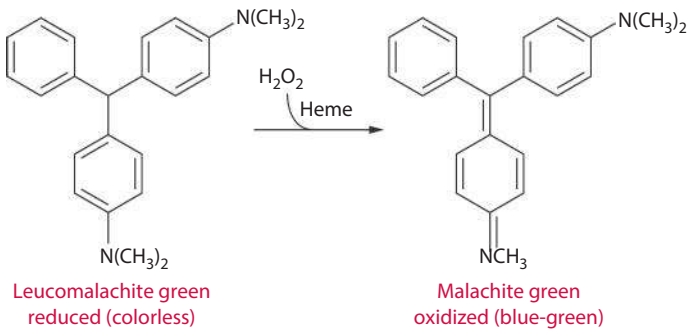

Figure 12.8 Chemical reaction of leucomalachite green assay.

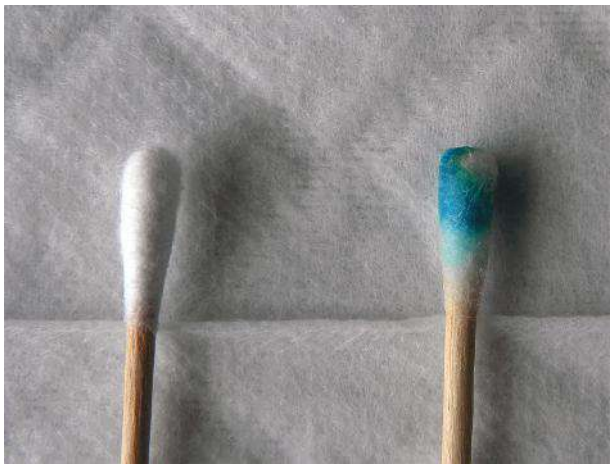

Figure 12.9 Photograph of leucomalachite green assay results. Negative (left) and positive (right) reaction. (© Richard C. Li.)

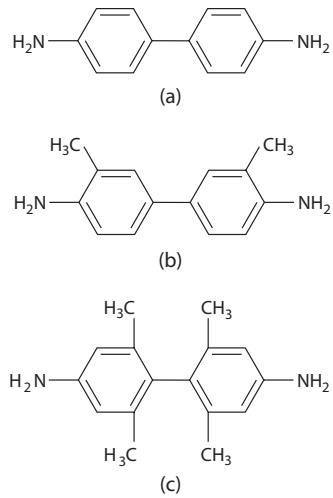

Figure 12.10 Chemical structures of benzidine and derivatives: (a) benzidine; (b) orthotolidine; (c) tetramethylbenzidine.

immediately. Benzidine was found to be a carcinogen and is therefore no longer used for forensic testing. *Orthotolidine* is a dimethyl derivative of benzidine. Its oxidation reaction can be catalyzed by heme to produce a blue color reaction under acidic conditions (Figure 12.11). Orthotolidine is also considered a potential carcinogen based on animal studies, and for this reason it has been replaced by tetramethylbenzidine. *Tetramethylbenzidine* (TMB) is a tetramethyl derivative of benzidine. The oxidation of TMB can be catalyzed by heme to produce a green to blue-green color under acidic conditions. TMB continues to be used. The Hemastix® assay kit (Miles Laboratories) is a TMB-based assay that utilizes a TMB-containing strip device. A test is carried out by applying a moistened sample to a Hemastix® strip. The appearance of a green or a blue-green color indicates the presence of blood.

12.2.3 Chemiluminescence and Fluorescence Assays

Other organic compounds whose oxidation products have chemiluminescent or fluorescent properties are utilized for testing. In the *chemiluminescence assay*, light is emitted as a product of a chemical reaction. In this category, luminol produces chemiluminescence when blood is present. In contrast, a *fluorescence assay* requires the exposure of an oxidized product, such as fluorescin, to a particular wavelength of an excitation light source. The fluorescence is then emitted at longer wavelengths than that of the excitation light source.

One advantage of chemiluminescent and fluorescent reagents is that they can be sprayed over large areas where latent bloodstains are potentially located. A positive reaction identifies blood and also reveals the patterns of bloody impressions such as footprints and fingerprints. These methods are very sensitive and can pinpoint the locations of even small traces of blood. Additionally, they are useful for detecting blood at crime scenes that have been cleaned and show no visible staining. One disadvantage of chemiluminescent and fluorescent reagents is that precautions must be taken if stains are very small or have been washed. The spraying of

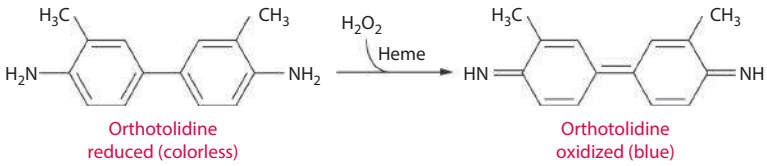

Figure 12.11 Chemical reaction of orthotolidine assay.

presumptive assay reagents may further dilute a sample and thus lead to difficulty in isolating sufficient amounts of DNA for forensic DNA analysis.

12.2.3.1 Luminol (3-Aminophthalhydrazide)

Luminol is usually utilized as a chemiluminescent reagent. The oxidation reaction of luminol catalyzed by heme produces light in the presence of an oxidant (Figures 12.12 through 12.14). The light emitted from a positive reaction can only be observed in the dark, which limits the

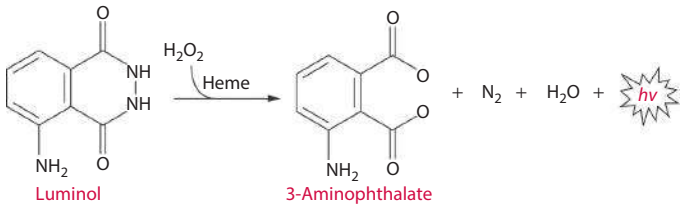

Figure 12.12 Chemical reaction of luminol assay.

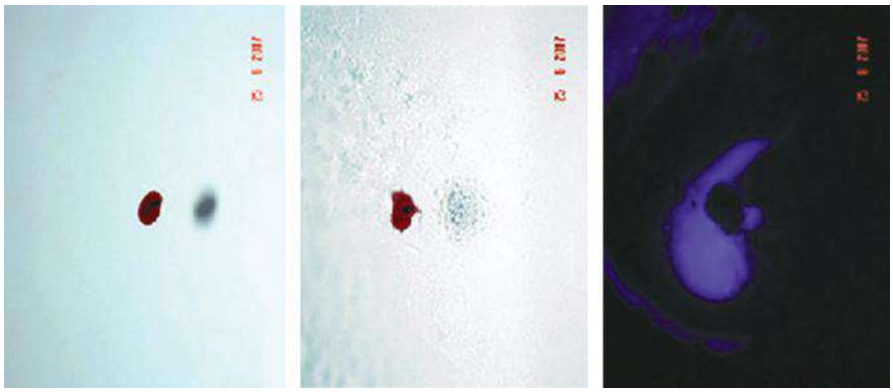

Figure 12.13 Results obtained from testing a drop of blood on a glass plate with luminol. From left to right: drop of blood; drop of blood after spraying with luminol; drop of blood after spraying with luminol and viewing in the dark. A blue chemiluminescence indicates the presence of blood. (From Bergervoet, P.W., et al., *J Hosp Infect*, 68, 329–333, 2008. With permission.)

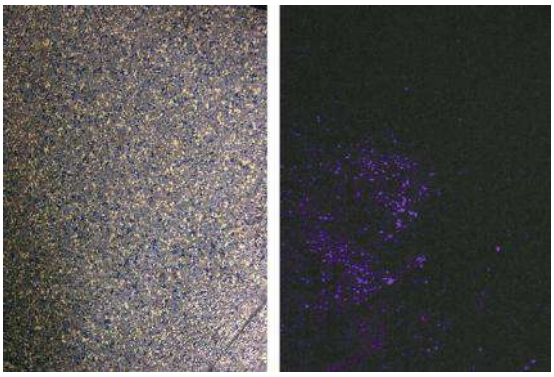

Figure 12.14 Detecting latent bloodstains on a floor with luminol. Left: with the light on; right: immediately after the light is switched off. A blue chemiluminescence indicates the presence of blood traces. (From Bergervoet, P.W., et al., *J Hosp Infect*, 68, 329–333, 2008. With permission.)

applications of luminol. The photodocumentation of a luminol-enhanced pattern should be done immediately before it fades away.

12.2.3.2 Fluorescin

Fluorescin is another reagent that is used to test for the presence of bloodstains at a crime scene (Figures 12.15 and 12.16). When oxidized and catalyzed by heme, fluorescin demonstrates fluorescent properties. Usually, fluorescin-sprayed stains are exposed to light in the range of 425–485 nm using an alternate light source device. In a positive reaction, the oxidized fluorescin emits an intense yellowish-green fluorescent light, which indicates the presence of a bloodstain. The light emitted from fluorescin-sprayed stains lasts longer than that of luminol.

12.2.4 Factors Affecting Presumptive Assay Results

The catalytic assays discussed in the previous sections are not specific to blood only, which can possibly lead to the observation of false-positive or false-negative results (Figure 12.17).

12.2.4.1 Oxidants

Chemicals that are strong oxidants may cause a false-positive reaction. Such chemicals can catalyze the oxidation reaction even in the absence of heme and result in a false-positive reaction. Certain metal salts, such as copper and nickel salts, household bleaches and cleaners that contain hypochlorite ions, and hair-coloring products that contain hydrogen peroxide, work as oxidants. To address this problem, a two-step catalytic assay should be performed. The substrate

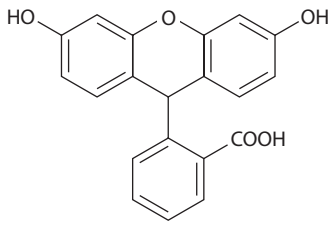

Figure 12.15 Chemical structure of fluorescein.

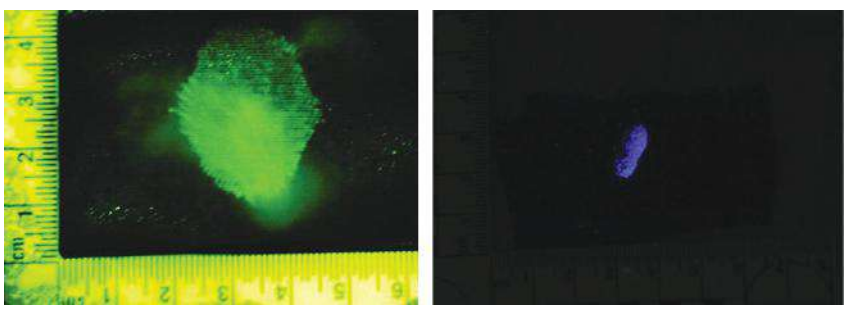

Figure 12.16 Detecting diluted bloodstains on black cotton fabric with fluorescin (left) and Bluestar (right). (From Finnis, J., Lewis, J., and Davidson, A., *Sci Justice*, 53, 178–186, 2013. With permission.)

$$\underset{\text{(colorless)}}{AH_2} + H_2O_2 \underset{\text{(3) Reductant}}{\overset{\text{(1) Oxidant; (2) Peroxidase}}{\rightleftharpoons}} \underset{\text{(color)}}{A} + 2H_2O$$

Figure 12.17 Factors affecting presumptive assay results. Strong oxidant and peroxidase may cause false-positive results; reductant may cause false-negative results.

is applied first to the sample in question. A color change occurring before the addition of hydrogen peroxide indicates a false-positive result due to a possible oxidant in the sample. If a color change is observed after the addition of hydrogen peroxide, the result is a true positive.

12.2.4.2 Plant Peroxidases

Many types of plants such as horseradish contain peroxidases. Plant peroxidases may also catalyze oxidation reactions and lead to false-positive results. However, plant peroxidases are usually heat sensitive and may be inactivated by high temperatures. Because the heme molecule is relatively stable at high temperatures, samples can be retested after heating. This will inactivate any plant peroxidases in a sample.

12.2.4.3 Reductants

Although not common, a false-negative result can occur when a strong reductant is present in a sample. Strong reductants such as certain metal ions including lithium and zinc may inhibit the oxidation reaction.

12.3 Confirmatory Assays for Identification

12.3.1 Microcrystal Assays

Microcrystal assays apply chemicals to treat bloodstains, forming crystals of heme molecules. The morphologies of the resulting crystals are distinctive for heme and can be compared with a known standard using a microscope. A positive microcrystal assay strongly indicates the presence of blood. However, confirmatory assays are usually not as sensitive as presumptive assays. Additionally, these assays cannot distinguish between human and animal blood.

12.3.1.1 Hemochromagen Crystal Assay

Hemochromagens are heme derivatives in which the ferrous iron of the heme forms two bonds with nitrogenous bases (Figure 12.4). The method for forming hemochromagen crystals was documented in 1864. Since then, various modifications have been reported. The *Takayama crystal assay*, published in 1912, has been the method preferred by many forensic laboratories. A bloodstain is treated with pyridine and glucose (a reducing sugar that is capable of reducing ferric ion) under alkaline conditions to form crystals of pyridine ferroprotoporphyrin (Figure 12.18).

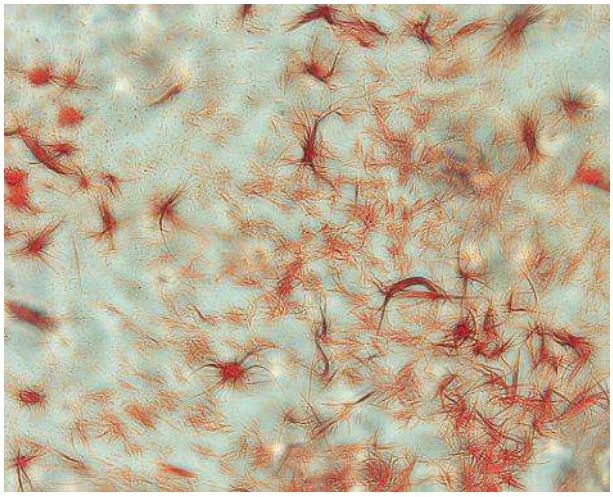

Figure 12.18 Microcrystal assays using the Takayama method. (© Richard C. Li.)

Figure 12.19 Microcrystal assays using the Teichmann method. (From James, S.H., Nordby, J.J., and Bell, S., *Forensic Science: An Introduction to Scientific and Investigative Techniques*, 4th edn., CRC Press, Boca Raton, 2014. With permission.)

12.3.1.2 Hematin Crystal Assay

This assay is also known as the *Teichmann crystal assay*. In 1853, Teichmann documented a method of forming crystals of blood specimens. When blood specimens are treated with glacial acetic acid and salts, and subsequently heated, hematin chloride (ferriprotoporphyrin chloride), a prismatic brown-colored crystal, is formed (Figure 12.19). Hematin (Figure 12.4) is a heme derivative; its iron is in the *ferric* (Fe^{+3}) state. This hematin assay has a similar sensitivity and specificity as hemochromagen assays. The hematin assay has the advantage of being more reliable than hemochromagen assays for aged blood samples.

12.3.2 Other Assays

Additional techniques may be used to confirm the presence of hemoglobin. For example, *chromatographic* and *electrophoretic methods* can identify hemoglobin by its mobility characteristics. *Spectrophotometric methods* for identifying hemoglobin are based on measurements of the

Table 12.1 Application of RT-PCR Assay for Blood Identification

Gene Symbol	Gene Product	Description	Further Reading
HBA1	Hemoglobin α1	Hemoglobin α1 chain (abundant in erythrocytes)	Waye and Chui (2001)
PBGD[a]	Porphobilinogen deaminase	Erythrocyte-specific isoenzyme of heme biosynthesis pathway	Gubin and Miller (2001)
SPTB	β-Spectrin	Subunit of major protein of erythrocyte membrane skeleton	Amin et al. (1993)

Source: Adapted from Juusola, J. and Ballantyne, J., *Forensic Sci Int*, 152, 1–12, 2005; Nussbaumer, C., Gharehbaghi-Schnell, E., and Korschineck, I., *Forensic Sci Int*, 157, 181–186, 2006.

[a] Also known as hydroxymethylbilane synthase (HMBS).

characteristic light spectra, with peak absorbance at 400–425 nm, absorbed by hemoglobin and its derivatives. Finally, *immunological methods* utilize antihuman hemoglobin antibodies. This antibody can be used to detect human hemoglobin and thus indicate the presence of human blood (see Chapter 13).

Recently, *RNA-based assays* have been developed to identify blood. These assays are based on the fact that certain genes are specifically expressed in certain cell types (Chapter 11). Thus, the techniques used in the identification of blood are based on the detection of specific types of messenger RNA (mRNA) that are expressed exclusively in erythrocytes. These assays utilize reverse transcriptase polymerase chain reaction (RT-PCR; see Chapter 7) methods to detect the gene expression levels of mRNAs for blood identification. Table 12.1 summarizes the tissue-specific genes utilized for blood identification. Compared with conventional assays that are used for blood identification, the RNA-based assays have higher specificity and are amenable to automation. However, one limitation is that RNA is unstable due to degradation by endogenous and environmentally born ribonucleases.

Bibliography

Amin, K.M., et al., The exon-intron organization of the human erythroid beta-spectrin gene. *Genomics*, 1993, **18**(1): 118–125.

Andersen, J. and S. Bramble, The effects of fingermark enhancement light sources on subsequent PCR-STR DNA analysis of fresh bloodstains. *J Forensic Sci*, 1997, **42**(2): 303–306.

Anderson, S.E., G.R. Hobbs, and C.P. Bishop, Multivariate analysis for estimating the age of a bloodstain. *J Forensic Sci*, 2011, **56**(1): 186–193.

Anderson, S., et al., A method for determining the age of a bloodstain. *Forensic Sci Int*, 2005, **148**(1): 37–45.

Andrasko, J., The estimation of age of bloodstains by HPLC analysis. *J Forensic Sci*, 1997, **42**(4): 601–607.

Anslinger, K., et al., Ninhydrin treatment as a screening method for the suitability of swabs taken from contact stains for DNA analysis. *Int J Legal Med*, 2004, **118**(2): 122–124.

Arany, S. and S. Ohtani, Age estimation of bloodstains: A preliminary report based on aspartic acid racemization rate. *Forensic Sci Int*, 2011, **212**(1–3): e36–e39.

Barnett, P.D., et al., Discussion of "Effects of presumptive test reagents on the ability to obtain restriction fragment length polymorphism (RFLP) patterns from human blood and semen stains". *J Forensic Sci*, 1992, **37**(2): 369–370.

Bauer, M., S. Polzin, and D. Patzelt, Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: A possible indicator of the age of bloodstains? *Forensic Sci Int*, 2003, **138**(1–3): 94–103.

Bergervoet, P.W., et al., Application of the forensic Luminol for blood in infection control. *J Hosp Infect*, 2008, **68**: 329–333.

Botoniic-Sehic, E., et al., Forensic application of near-infrared spectroscopy: Aging of bloodstains. *Spectroscopy* (Santa Monica), 2009, **24**(2): 42–48.

Boyd, S., et al., Highly sensitive detection of blood by surface enhanced Raman scattering. *J Forensic Sci*, 2013, **58**(3): 753–756.

Bremmer, R.H., et al., Remote spectroscopic identification of bloodstains. *J Forensic Sci*, 2011, **56**(6): 1471–1475.

Bremmer, R.H., et al., Forensic quest for age determination of bloodstains. *Forensic Sci Int*, 2012, **216**(1–3): 1–11.

Budowle, B., et al., The presumptive reagent fluorescein for detection of dilute bloodstains and subsequent STR typing of recovered DNA. *J Forensic Sci*, 2000, **45**(5): 1090–1092.

Caldwell, J.P., W. Henderson, and N.D. Kim, ABTS: A safe alternative to DAB for the enhancement of blood fingerprints. *J Forensic Sci*, 2000, **45**(4): 785–794.

Cheeseman, R., Direct sensitivity comparison to the fluorescein and luminol bloodstain enhancement techniques. *J Forensic Ident*, 1999, **49**(3): 8.

Cheeseman, R. and L.A. DiMeo, Fluorescein as a field-worthy latent bloodstain detection system. *J Forensic Ident*, 1995, **45**(6): 1.

13

Species Identification

Chapter 12 discussed the principles of the identification of blood. If a stain is identified as blood, the evidence can be tested to determine whether the blood is of human origin. If the bloodstain is nonhuman, further analysis is usually not necessary.

Before forensic DNA techniques were implemented, species identification was largely determined by serological methods. Currently, most forensic laboratories perform DNA quantitation prior to DNA profile analysis. The quantitation method specifically detects higher-primate DNA. The presence of DNA measured by the quantitation assays concurrently identifies a sample as being of human origin (since crimes involving primate blood are extremely rare). Thus, species identification is usually not performed in forensic laboratories.

Nevertheless, species identification assays can be useful for screening to exclude or eliminate nonhuman samples unrelated to an investigation. Thus, it is practical for small laboratories to eliminate unnecessary analyses due to time and budget constraints. Additionally, species identification kits such as immunochromatographic devices allow field testing by crime scene investigators.

In cases involving the killing, trading, and possession of products derived from species that are protected from illegal hunting, it may be necessary to identify the animal species prior to further analysis. Species identification using DNA analysis can be performed using commonly used loci at the mitochondrial cytochrome b gene (*Cytb*), the cytochrome c oxidase I gene (*COI*), and the D-loop region. This type of identification is usually within the scope of wildlife forensic science and is thus not discussed here.

13.1 General Considerations

Most assays for species identification are based on serological techniques, including primary and secondary binding assays. The most common primary binding assays are immunochromatographic assays. The most commonly used secondary binding assays are precipitation-based assays that rely on the binding of an antigen to an antibody, causing the formation of visible precipitation. These precipitation-based assays include ring assays, Ouchterlony assays, and crossed-over immunoelectrophoresis. These assays utilize antihuman and antianimal antibodies to identify human and animal species, respectively.

13.1.1 Types of Antibodies

An antihuman antibody that is used in the identification of human samples can be made by introducing human serum into a host animal, which then produces specific antibodies against the human serum proteins. Antibodies produced from different species of host animals may

produce variations in the characteristics of reactions. Since albumin is the most abundant protein in human serum, the antihuman antibody that is produced reacts strongly with human albumin. Albumin is a protein that plays important roles in the maintenance of the vascular circulating fluid and the transportation of various substances such as nutrients, hormones, and metabolic products. Blood is drawn from the host animal and the serum portion is collected. The collected serum is a polyclonal antihuman antiserum containing a mixture of antibodies against various human serum proteins. Likewise, an antibody against animal serum proteins can also be made to identify animal species of interest.

Other antibodies such as antihuman hemoglobin (Hb) antibodies can also be used to identify the human origin of a sample. Hb is an oxygen-transport protein that is found in erythrocytes (Chapter 12). Purified Hb can be used to generate monoclonal and polyclonal antihuman Hb antibodies. Likewise, antibodies recognizing glycophorin A (GPA), a human erythrocyte membrane antigen (see Section 13.2.1.2), can also be produced in a similar manner.

13.1.2 Titration of Antibodies

Recall that the ratio of antigen to antibody is critical for the success of a secondary reaction (Chapter 10). An extreme excess of antigen or antibody concentrations can inhibit secondary reactions. The prozone and postzone phenomena must be considered, and the concentrations of antigen and antibody must be carefully determined for forensic serology assays. For instance, in the prozone situation, a false-negative reaction may occur due to the presence of a high concentration of antibody.

Quality-control procedures can be used to estimate the amount of a specific antibody that is present, often via titration (Figure 13.1). To titrate an antiserum, a series of dilutions are made and each dilution is then tested for activity using precipitation or agglutination methods. The reciprocal of the highest dilution giving a positive reaction is known as the *titer*. This reflects the amount of antibody in the antiserum. Additionally, the polyclonal antiserum is a mixture of antibodies; thus, the reaction of the antiserum may vary from animal to animal (of the same species). Each lot or batch of antiserum must be validated by titration.

13.1.3 Antibody Specificity

In addition to the titer, the specificity of the antihuman antibody must be tested. Most antihuman antibodies usually have cross-reactivity with higher primates. This is not a great

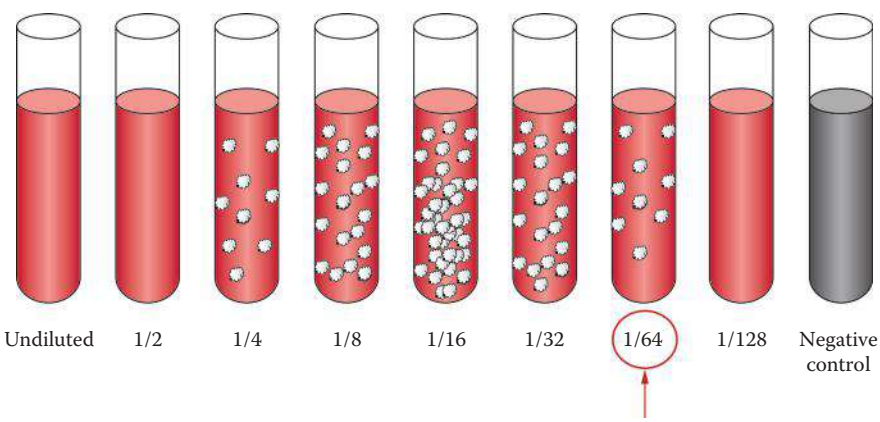

Figure 13.1 Titration of antibodies. Serum is serially diluted and a constant amount of antigen is applied to each tube. The mixture is incubated, allowing agglutination to occur. The reciprocal of the highest dilution giving a positive agglutination reaction is 64 (the titer). (© Richard C. Li.)

concern because crimes involving nonhuman primates are very rare. Nevertheless, the antihuman antibody must not cross-react with other commonly encountered animals. Antisera and positive control samples must be validated for cross-reactivity. Tissue specificity must also be validated. The antiserum against human serum is usually reactive with other human biological fluids such as semen and saliva.

13.1.4 Optimal Conditions for Antigen–Antibody Binding

A number of factors can affect antigen–antibody binding. For example, increasing ionic strength can inhibit the binding of an antigen and an antibody. Stronger inhibition is usually observed for ions with large ionic radii and small radii of hydration. It is believed that the lower degree of hydration permits interactions of ions and the antibody-binding site, leading to inhibition. A proper buffer system must be selected in serological assays to ensure reliable results. The introduction of polymers can facilitate precipitation in secondary binding reactions because the presence of a polymer in a solution decreases the solubility of proteins. Linear hydrophilic polymers with high molecular weights (e.g., polyethylene glycol) are preferred. Additional factors such as temperature and pH can also affect antigen–antibody binding.

13.2 Assays

Samples can be prepared by cutting out a portion of a stain or scraping stains from a surface. A sample is usually extracted with a small volume of saline or buffer. The extracted sample can be tested using the assays described in the following subsections. Controls should be included, for example, by using a known human serum as a positive control and an extraction blank as a negative control.

13.2.1 Immunochromatographic Assays

Immunochromatographic assays are rapid, specific, and sensitive and can be used in both laboratory and field tests for species identification. Two types of assays are discussed, including those based on the detection of human erythrocyte proteins. Chapter 11 discusses the principle of immunochromatographic assays in more detail.

13.2.1.1 Identification of Human Hemoglobin Protein

Commercially produced immunochromatographic kits such as the Hexagon OBTI (Human Gesellschaft für Biochemica und Diagnostica mbH, Wiesbaden) and the ABAcard HemaTrace® (Abacus Diagnostics, California) are available. They utilize the antibody–antigen–antibody sandwich method by using antibodies that recognize human Hb. The ABAcard HemaTrace assay utilizes a labeled monoclonal antihuman Hb antibody contained in a sample well, and a polyclonal antihuman Hb antibody immobilized at a test zone of a nitrocellulose membrane. Additionally, an antiglobulin that recognizes the antibody is immobilized onto a control zone (Figure 13.2).

A sample can be prepared by cutting a small portion (2 mm diameter) of a stain or a swab. Each sample is extracted for 5 min in 2 mL of extraction buffer. A longer extraction time may be used for older stains. The samples are loaded into the sample well, and the antigen in the sample binds to the labeled antibody in the well to form an antigen–antibody complex, which then diffuses across the nitrocellulose membrane. At the test zone, the solid-phase antihuman Hb antibody binds to the antigen–antibody complex to form a labeled antibody–antigen–antibody sandwich.

The ABAcard HemaTrace® uses a pink dye that is visualized in a positive result as a pink horizontal line at the test zone (Figure 13.3). In the control zone, unbound labeled antihuman

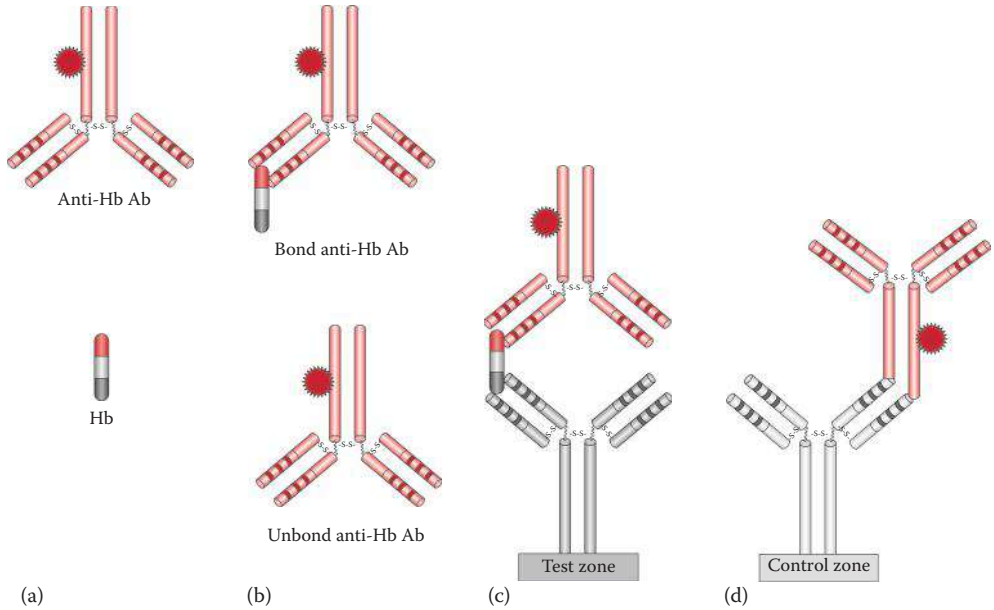

Figure 13.2 Immunochromatographic assays for the identification of Hb in human blood. (a) In a sample well, Hb in a blood sample is mixed with a labeled anti-Hb Ab. (b) The Hb binds to the labeled anti-Hb Ab to form a labeled Ab–Hb complex. The complex diffuses toward the test zone. (c) At the test zone, the labeled Ab–Hb complex binds to an immobilized anti-Hb Ab to form a labeled Ab–Hb–Ab sandwich. (d) At the control zone, the labeled anti-Hb Ab binds to an immobilized antiglobulin and is captured. Ab and Hb represent the antibody and hemoglobin, respectively. (© Richard C. Li.)

Hb antibody binds to the solid-phase antiglobulin. This antibody–antiglobulin complex at the control zone also produces a pink horizontal line. The test is considered valid only if the line in the control zone is observed. The presence of human Hb results in a pink line at both the test and control zones. The absence of human Hb results in a pink line in the control zone only. A positive result can appear in less than a minute.

Validation studies have revealed that the sensitivity of the ABAcard HemaTrace® can be as low as 0.07 μg/mL of Hb. The normal blood Hb concentration is 14–18 and 12–16 g/dL among males and females, respectively. This assay is more sensitive than the Kastle–Meyer assay (Chapter 12). Additionally, the assay is responsive to aged stains and degraded materials. Specificity studies have shown that it is specific for blood of higher primates, including humans. However, it is also responsive to seminal stains, and oral, vaginal, anal, and rectal swabs. It is believed that these biological fluids contain very low amounts of Hb, which can still be detected by highly sensitive assays. However, if the concentration of blood is too high, a false negative can result due to the high-dose hook effect described in Chapter 11.

13.2.1.2 Identification of Human Glycophorin A Protein

Commercially produced immunochromatographic kits such as RSID™-Blood (Independent Forensics, Hillside, IL) use antibodies that recognize human GPA (Figure 13.4). A labeled monoclonal antihuman GPA antibody is contained in a sample well, and a second monoclonal antihuman GPA antibody, to a different epitope of GPA, is immobilized onto a test zone of the membrane. An antiglobulin that recognizes the antibody is immobilized onto a control zone (Figure 13.5).

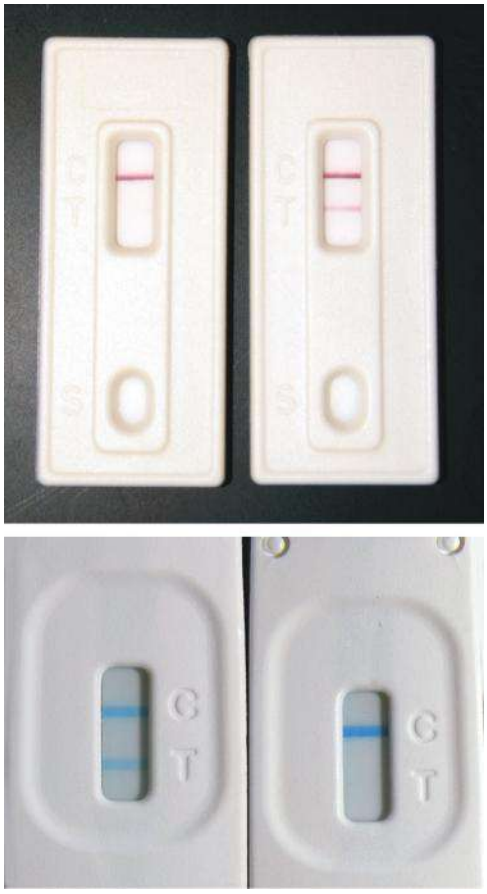

Figure 13.3 Human blood identification using immunochromatic devices. Top: In an assay using an ABAcard HemaTrace device, the negative (left) and positive (right) results are shown. The "C" band indicates that the test is valid. The "T" band indicates the presence of human blood. The sample well is labeled "S." (© Richard C. Li.) Bottom: The positive (left) and negative (right) results are shown using a Hexagon-OBTI device. The "C" band indicates that the test is valid. The "T" band indicates the presence of human blood. (From Ramsthaler, F., et al., Postmortem interval of skeletal remains through the detection of intraosseal hemin traces. A comparison of UV-fluorescence, luminol, Hexagon-OBTI(R), and Combur(R) tests. *Forensic Sci Int*, 209, 59–63, 2011. With permission.)

The sample can be collected by cutting out a small portion of a stain or a swab. The sample is then extracted overnight in an extraction buffer. The extract is removed and mixed with a running buffer. The assay is carried out by loading the extracted sample into the sample well. Again, the presence of GPA results in a pink line at both the test zone and the control zone, while the absence of GPA results in a pink line in the control zone only. The test is considered valid only if the line in the control zone is observed. A result can be read after 10 min.

Validation studies revealed that the sensitivity of the RSID kit can be as low as 100 nL of human blood. Species specificity studies showed no cross-reactivity with various animal species, including nonhuman primates. Biological fluid specificity studies revealed that the kit is not responsive to other human biological fluids such as semen, saliva, urine, milk, and amniotic and vaginal fluid. No high-dose hook effects were observed in samples containing up to 5 μL of blood.

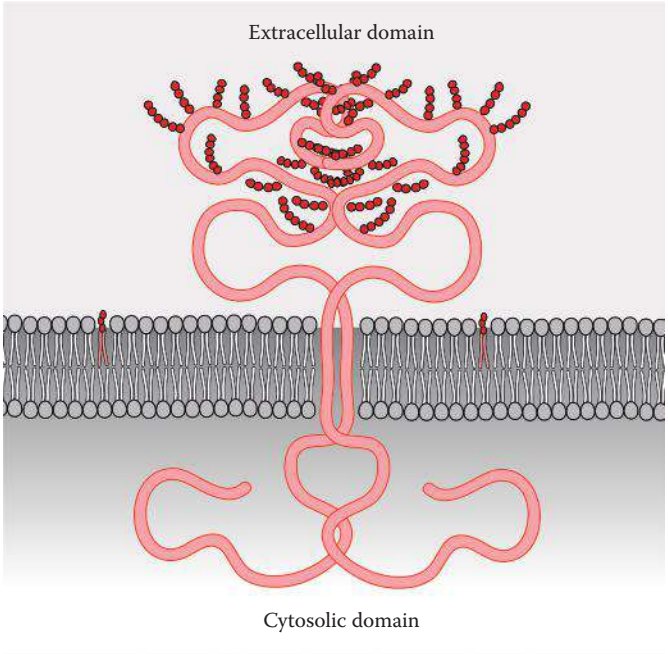

Figure 13.4 Diagram of the structure of glycophorin (GPA) protein. GPA is a transmembrane protein on the human erythrocyte membrane. A GPA dimer is shown. The extracellular domain of the GPA is glycosylated with carbohydrate side chains (red). Various GPA epitopes are antigenic determinants of several blood group systems. (© Richard C. Li.)

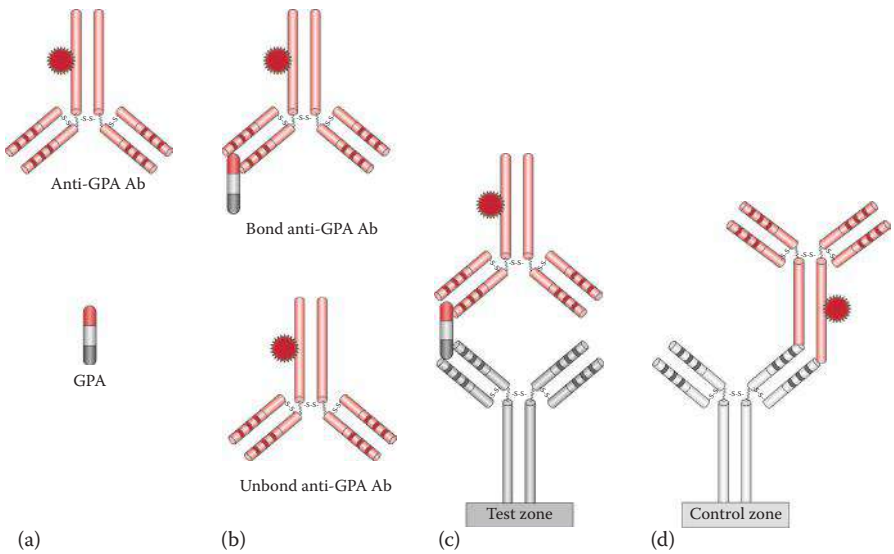

Figure 13.5 Immunochromatographic assays for the identification of GPA in human blood. (a) In a sample well, GPA in a blood sample is mixed with a labeled anti-GPA Ab. (b) The GPA binds to the labeled anti-GPA Ab to form a labeled Ab–GPA complex. (c) At the test zone, the labeled Ab–GPA complex binds to an immobilized anti-GPA Ab to form a labeled Ab–GPA–Ab sandwich. (d) At the control zone, the labeled anti-GPA Ab binds to an immobilized antiglobulin and is captured. Ab and GPA represent the antibody and GPA protein, respectively. (© Richard C. Li.)

13.2.2 Double Immunodiffusion Assays

13.2.2.1 Ring Assay

Chapter 11 discussed the basic principle of ring assay in detail. In this double immunodiffusion assay, an antihuman antibody reagent is placed at the bottom of a test tube and a bloodstain extract is placed on top of the bottom layer, as illustrated in Figure 13.6. The procedure is described in Box 13.1. In a positive reaction, a white precipitate between the two layers observed after several minutes indicates that a sample is of human origin. If the bloodstain extract is not human, no precipitation should appear.

13.2.2.2 Ouchterlony Assay

The basic principle of this double immunodiffusion assay has also been discussed in Chapter 11. The procedure of the Ouchterlony assay is described in Box 13.2. In a positive reaction, a line of precipitate will form between each antigen well and antibody well. This assay can also determine the similarity of the antigens (Figure 13.7). During the diffusion process, different antigen–antibody complexes migrate at different rates. Consequently, a separate line of precipitate will appear in the gel for each antigen–antibody complex. In an assay in which two antigens are loaded in adjacent wells and an antibody in the third well, the following results can be observed:

- If the two antigens are identical, the two lines will become fused. This phenomenon is referred to as *identity*.
- If the two antigens are totally unrelated, the lines will cross each other but not fuse; this is known as *nonidentity*.
- If the two antigens are related (share a common epitope) but are not identical, the lines will merge with spur formation. The spurs are continuations of the line formed by the antigen due to its unique epitope. This phenomenon is known as *partial identity*.

Thus, in a species test, a positive result is noted when the precipitate lines for the positive controls and the samples fuse. No spur formation should be observed.

13.2.3 Crossed-Over Electrophoresis

This method is a combination of immunodiffusion and electrophoresis (also see Chapter 11). The procedure for the assay is described in Box 13.3. With this technique, a sharp precipitate band is visualized in a positive reaction (Figure 13.8). However, false-negative results can occur due to the postzone phenomenon, in which excess antigen may inhibit precipitation. In this

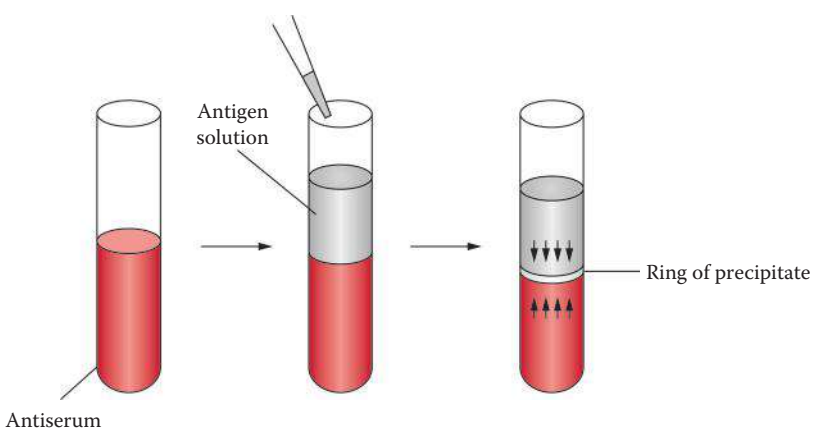

Figure 13.6 Ring assay. The antigen solution is carefully applied over the antiserum solution. Incubation may be necessary in order to form a ring of precipitate. (© Richard C. Li.)

BOX 13.1 RING ASSAY PROCEDURE

Sample preparation and extraction

- Extract a portion of a stain with saline at 4°C overnight.

Controls

- Include a positive control (known human serum sample) and a negative control (extraction blank).

Loading of antibody and samples

- Spin the antihuman antibody in a microfuge and transfer the supernatant into test tubes or capillary tubes (depending on the volume of the stain and the antiserum extracted).
- Place the sample carefully over the top of the antiserum solution, which is usually denser than the sample.

Immunodiffusion reaction

- Carry out the reaction at room temperature.
- In a positive reaction, white precipitate between the two layers can be observed after several minutes. This indicates that the sample is of human origin. No precipitate is formed if a bloodstain extract is from a nonhuman origin.

BOX 13.2 OUCHTERLONY ASSAY PROCEDURE

Sample preparation and extraction

- Cut out a small portion (approximately 5×5 mm) of a stain or a portion of a swab.
- Extract at room temperature in 100 μL of water for 30 min. The extract can be diluted if necessary. Alternatively, a very small piece of the stain or swab can be inserted directly into the well.

Controls

- Positive (known serum)
- Negative (extraction blank)
- Substrate controls (extraction of substrate from unstained area) if applicable

Agarose gel preparation

Heat a suspension of agarose (4%) until liquefied. Cool the solution in a water bath at 55°C. Pour the agarose onto a piece of glass slide and let the gel solidify to a thickness of about 2–3 mm. Alternatively, a polyester support film such as GelBond (Cambrex, New Zealand) can be used as a gel support. The agarose should be poured onto the hydrophilic side of a piece of GelBond film (6×9 cm). Punch wells consisting of a central well surrounded by four wells using a template.

Loading antibodies and samples

- Apply antihuman antibody to the central well. Apply the positive control to one of the surrounding wells.
- Apply the sample(s) in question next to a positive control.
- Apply negative and substrate controls to the remaining wells; only one negative control is needed per gel.

Immunodiffusion reaction

Incubate the plate overnight in a moisture chamber at 37°C.

Staining

Soak the gel overnight in saline solution and then soak it in deionized water for 10 min. Repeat once. Dry the gel between paper towels with a weight on top for 30 min. Dry in an oven for 30 min. Stain the gel with Coomassie blue. Stained precipitate bands appear blue.

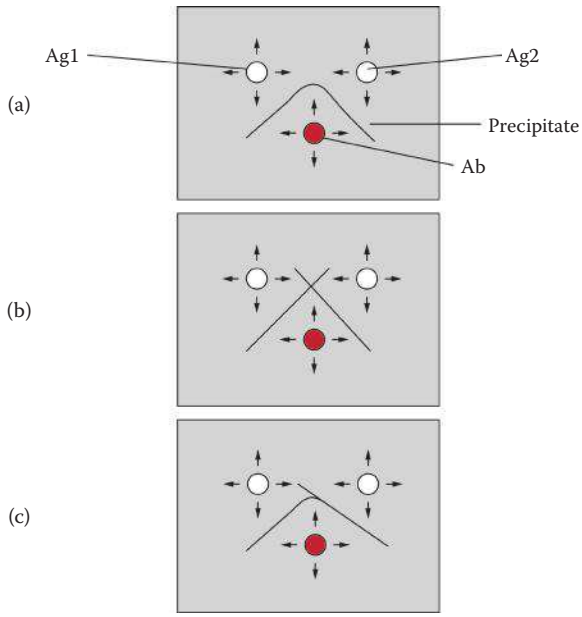

Figure 13.7 Results of Ouchterlony assay. Ab indicates an antibody. Ag1 and Ag2 are the antigen samples in question. (a) Identity. The two antigens are identical (fused line). (b) Nonidentity. The two antigens are unrelated (spur). (c) Partial identity. The two antigens are related but not identical (fused line with spur). (© Richard C. Li.)

situation, the sample can be diluted and the assay can be repeated. False-negative results can also occur due to simple mistakes made during electrophoresis:

- Electrophoresis is carried out in the opposite direction, which results in samples running off the gel.
- Electrophoresis is carried out using an incorrect buffer system, affecting antigen–antibody binding. The amount of current applied during the electrophoresis is too strong and generates heat and denatures proteins.

BOX 13.3 CROSSED-OVER ELECTROPHORESIS PROCEDURE

Sample preparation and extraction

- Cut out a small portion (~5 × 5 mm) of a stain or a portion of a swab.
- Extract at room temperature in 100 μL of water for 30 min. The extract can be diluted if necessary. Alternatively, a very small piece of the stain or swab can be inserted directly into the well.

Controls

- Positive (known serum)
- Negative (extraction blank)
- Substrate (extraction of substrate from unstained area) if applicable

Agarose gel preparation

Heat a suspension of agarose (4%) until liquefied. Cool the solution in a water bath at 55°C. Pour the agarose onto a piece of glass slide and let it solidify. Alternatively, a polyester support film such as GelBond can be used as a gel support. The agarose should be poured onto the hydrophilic side of a piece of GelBond (6 × 9 cm). Punch small wells (about 1–2 mm) in rows using a template.

Loading antibodies and samples

- Apply antihuman antibody in one row of wells.
- Apply samples in the other row of wells. Apply the positive, negative, and substrate controls.

Electrophoresis

Submerge the agarose gel in an electrophoresis tank in proper orientation. The wells containing antihuman antibody should be closest to the anode (positive electrode) and the wells containing samples should be closest to the cathode (negative electrode). During electrophoresis, the antibody in the antiserum should migrate toward the cathode while the antigen migrates toward the anode. Electrophoresis is carried out at 10 V/cm for 20 min.

Staining

Soak the gel overnight in a saline solution and then soak it in deionized water for 10 min. Repeat once. Dry the gel between paper towels with a weight on top for 30 min. Dry in an oven for 30 min. Stain the gel with Coomassie blue. Stained precipitate bands appear blue.

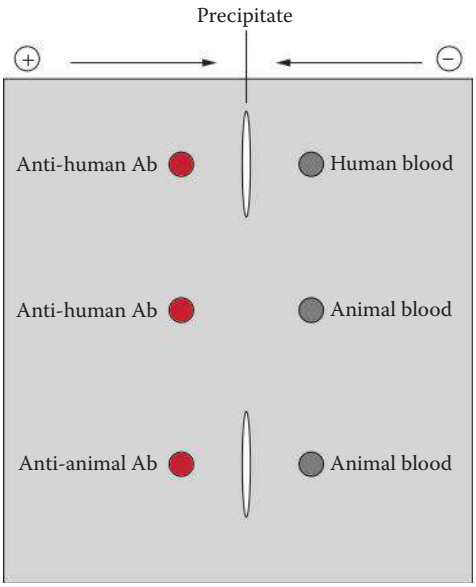

Figure 13.8 Results of crossed-over electrophoresis. A precipitate line is formed between a human blood sample and an antihuman antibody. No precipitate line is formed when the antihuman antibody is tested for an animal blood sample. A precipitate line is formed between the animal blood sample and the antianimal antibody. (© Richard C. Li.)

Bibliography

Allison, A.C. and J.A. Morton, Species specificity in the inhibition of antiglobulin sera: A technique for the identification of human and animal bloods. *J Clin Pathol*, 1953, **6**(4): 314–319.

Berger, B., et al., Validation of two canine STR multiplex-assays following the ISFG recommendations for non-human DNA analysis. *Forensic Sci Int Genet*, 2014, **8**(1): 90–100.

Bhatia, R.Y., Specificity of some plant lectins in the differentiation of animal blood. *Forensic Sci*, 1974, **4**(1): 47–52.

Coomber, N., et al., Validation of a short tandem repeat multiplex typing system for genetic individualization of domestic cat samples. *Croat Med J*, 2007, **48**(4): 547–555.

Dawnay, N., et al., Validation of the barcoding gene *COI* for use in forensic genetic species identification. *Forensic Sci Int*, 2007, **173**(1): 1–6.

DeForest, P., R. Gaensslen, and H.C. Lee, *Forensic Science: An Introduction to Criminalistics*, 1983. New York: McGraw-Hill.

Dorrill, M. and P.H. Whitehead, The species identification of very old human blood-stains. *Forensic Sci Int*, 1979, **13**(2): 111–116.

El-Sayed, Y.S., et al., Using species-specific repeat and PCR-RFLP in typing of DNA derived from blood of human and animal species. *Forensic Sci Med Pathol*, 2010, **6**(3): 158–164.

Grobbelaar, B.G., D. Skinner, and H.N. van de Gertenbach, The anti-human globulin inhibition test. In the identification of human blood stains. *J Forensic Med*, 1970, **17**(3): 103–111.

Hochmeister, M.N., et al., Validation studies of an immunochromatographic 1-step test for the forensic identification of human blood. *J Forensic Sci*, 1999, **44**(3): 597–602.

Imaizumi, K., et al., Development of species identification tests targeting the 16S ribosomal RNA coding region in mitochondrial DNA. *Int J Legal Med*, 2007, **121**(3): 184–191.

Johnston, S., J. Newman, and R. Frappier, Validation study of the Abacus Diagnostics ABAcard® HemaTrace® membrane test for the forensic identification of human blood. *Can Soc Forensic Sci*, 2003, **36**(3): 173–183.

Juusola, J. and J. Ballantyne, Messenger RNA profiling: A prototype method to supplant conventional methods for body fluid identification. *Forensic Sci Int*, 2003, **135**(2): 85–96.

Kanthaswamy, S., et al., Quantitative real-time PCR (qPCR) assay for human-dog-cat species identification and nuclear DNA quantification. *Forensic Sci Int Genet*, 2012, **6**(2): 290–295.
Karlsson, A.O. and G. Holmlund, Identification of mammal species using species-specific DNA pyrosequencing. *Forensic Sci Int*, 2007, **173**(1): 16–20.
Kun, T., et al., Developmental validation of Mini-DogFiler for degraded canine DNA. *Forensic Sci Int Genet*, 2013, **7**(1): 151–158.
Lawton, M.E. and J.G. Sutton, Species identification of deer blood by isoelectric focusing. *J Forensic Sci Soc*, 1982, **22**(4): 361–366.
Lee, H.C., Identification and grouping of bloodstains, in R. Saferstein (ed.), *Forensic Science Handbook*, pp. 267–337, 1982. Englewood Cliffs, NJ: Prentice Hall.
Lee, H.C. and P.R. De Forest, A precipitin-inhibition test on denatured bloodstains for the determination of human origin. *J Forensic Sci*, 1976, **21**(4): 804–810.
Lowenstein, J.M., et al., Identification of animal species by protein radioimmunoassay of bone fragments and bloodstained stone tools. *Forensic Sci Int*, 2006, **159**(2–3): 182–188.
Nakaki, S., et al., Study of animal species (human, dog and cat) identification using a multiplex single-base primer extension reaction in the cytochrome b gene. *Forensic Sci Int*, 2007, **173**(2–3): 97–102.
Ogden, R., Forensic science, genetics and wildlife biology: Getting the right mix for a wildlife DNA forensics lab. *Forensic Sci Med Pathol*, 2010, **6**(3): 172–179.
Parson, W., et al., Species identification by means of the cytochrome b gene. *Int J Legal Med*, 2000, **114**(1–2): 23–28.
Ramsthaler, F., et al., Postmortem interval of skeletal remains through the detection of intraosseal hemin traces. A comparison of UV-fluorescence, luminol, Hexagon-OBTI(R), and Combur(R) tests. *Forensic Sci Int*, 2011, **209**: 59–63.
Sato, I., et al., Forensic hair analysis to identify animal species on a case of pet animal abuse. *Int J Legal Med*, 2010, **124**(3): 249–256.
Shaler, R.C., Modern forensic biology, in R. Saferstein (ed.), *Forensic Science Handbook*, pp. 525–614, 2002. Upper Saddle River, NJ: Pearson Education.
Sivaram, S., et al., Differentiation between stains of human blood and blood of monkey. *Forensic Sci*, 1975, **6**(3): 145–152.
Spear, T.F. and S.A. Binkley, The HemeSelect test: A simple and sensitive forensic species test. *J Forensic Sci Soc*, 1994, **34**(1): 41–46.
Tabata, N. and M. Morita, Immunohistochemical demonstration of bleeding in decomposed bodies by using anti-glycophorin A monoclonal antibody. *Forensic Sci Int*, 1997, **87**(1): 1–8.
Tanaka, Y., et al., Enzyme-linked immunospot assay for detecting cells secreting antibodies against human blood group A epitopes. *Transplant Proc*, 2003, **35**(1): 555–556.
Tobe, S. and A. Linacre, Species identification of human and deer from mixed biological material. *Forensic Sci Int*, 2007, **169**(2–3): 278–279.
Tobe, S.S., et al., Recovery of human DNA profiles from poached deer remains part 2: Improved recovery protocol without the need for LCN analysis. *Sci Justice*, 2013, **53**(1): 23–27.
Tsutsumi, H. and Y. Katsumata, Forensic study on stains of blood and saliva in a chimpanzee bite case. *Forensic Sci Int*, 1993, **61**(2–3): 101–110.
van Asch, B. and A. Amorim, Capillary electrophoresis analysis of a 9-plex STR assay for canine genotyping. *Methods Mol Biol*, 2012, **830**: 231–240.
van Asch, B., et al., A new autosomal STR nineplex for canine identification and parentage testing. *Electrophoresis*, 2009, **30**(2): 417–423.
van Asch, B., et al., Forensic analysis of dog (*Canis lupus familiaris*) mitochondrial DNA sequences: An interlaboratory study of the GEP-ISFG working group. *Forensic Sci Int Genet*, 2009, **4**(1): 49–54.
Virkler, K. and I.K. Lednev, Blood species identification for forensic purposes using Raman spectroscopy combined with advanced statistical analysis. *Anal Chem*, 2009, **81**(18): 7773–7777.
Whitehead, P.H. and A. Brech, A micro-technique involving species identification and ABO grouping on the same fragment of blood. *J Forensic Sci Soc*, 1974, **14**(2): 109–110.
Wictum, E., et al., Developmental validation of DogFiler, a novel multiplex for canine DNA profiling in forensic casework. *Forensic Sci Int Genet*, 2013, **7**(1): 82–91.
Wilson-Wilde, L., et al., Current issues in species identification for forensic science and the validity of using the cytochrome oxidase I (COI) gene. *Forensic Sci Med Pathol*, 2010, **6**(3): 233–241.

14

Identification of Semen

14.1 Biological Characteristics

A typical ejaculation releases 2–5 mL of semen, which contains seminal fluid and sperm cells (*spermatozoa*). A normal sperm count ranges from 10^7 to 10^8 spermatozoa per milliliter of semen. The spermatozoa are formed from spermatogonia in the seminiferous tubules of the testes. This process of generating spermatozoa is referred to as spermatogenesis (Figure 14.1). The spermatozoa are then transported and stored in the tubular network of the epididymis where they undergo functional maturation (spermatogenesis and maturation take approximately 3 months). The epididymis joins the ductus deferens, which transports matured sperm from the epididymis to the ejaculatory duct. From there, spermatozoa follow the ejaculatory ducts into the prostatic urethra where they are joined with secretions from the prostate. Figure 14.2 illustrates the anatomy of the male reproductive system.

Seminal fluid is a complex mixture of glandular secretions. A typical sample of seminal fluid contains the combined secretions of several accessory glands. Seminal vesicle fluid accounts for approximately 60% of the ejaculate. Various proteins secreted by the seminal vesicles play a role in the coagulation of the ejaculate. Additionally, seminal vesicle fluid contains flavin, which causes semen to fluoresce under ultraviolet light, often utilized when searching for semen-stain evidence.

Prostatic fluid secretions account for approximately 30% of the ejaculate. The components of this fluid are complex as well. This portion of semen contains high concentrations of *acid phosphatase* (AP) and *prostate-specific antigen* (PSA). Both are useful markers for the identification of semen in forensic laboratories. The epididymis and the bulbourethral secretions each account for approximately 5% of the ejaculate.

A *vasectomy* is the surgical removal of a bilateral segment of the ductus deferens. The surgery prevents spermatozoa from reaching the distal portions of the male reproductive tract. However, a vasectomized male can still produce ejaculate that contains only seminal vesicle fluid, prostatic fluid, and bulbourethral fluid. The condition by which males have abnormally low sperm counts is known as *oligospermia*. *Azoospermia* is a condition that causes males to produce no spermatozoa. However, the secretion of seminal fluid is not affected in males who have these conditions. DNA derived from epithelial cells can be isolated from the seminal fluids of these individuals.

14.1.1 Spermatozoa

A human spermatozoon has three morphologically distinct structures: the head, the middle piece, and the tail (Figure 14.3). The head contains a nucleus with densely packed chromosomes. At the tip of the head is the *acrosomal cap*, which is a membranous compartment

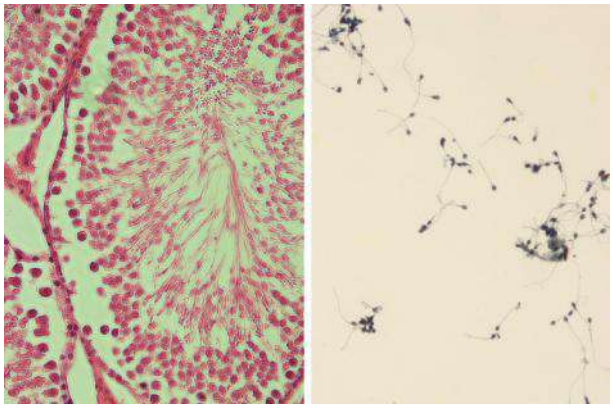

Figure 14.1 Spermatogenesis. In the seminiferous tubules (left) of the testes, spermatogonia (located at peripheral area of the seminiferous tubules) are differentiated to spermatids (located at the center of the seminiferous tubules). The spermatids are eventually differentiated to matured spermatozoa (right). (© Richard C. Li.)

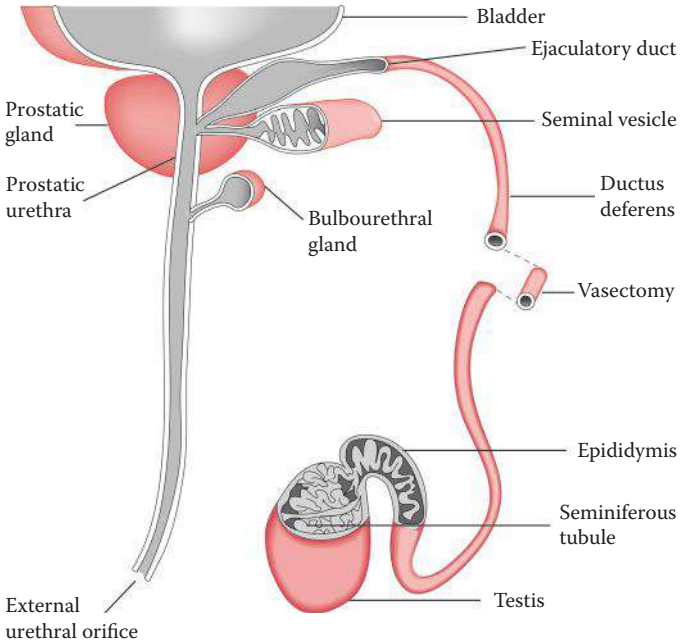

Figure 14.2 Male reproductive system and accessory glands (unilateral view). (© Richard C. Li.)

containing enzymes essential for fertilization. The head is attached to the middle piece through a short neck where the mitochondria that provide the energy for moving the tail are located. The tail or flagellum is responsible for spermatozoon motility. In contrast to other cell types, a mature spermatozoon lacks various intracellular organelles such as an endoplasmic reticulum, Golgi apparatus, lysosomes, and peroxisomes. In a normal male, at least 60% of spermatozoa have normal morphology, so morphological abnormalities can often be observed.

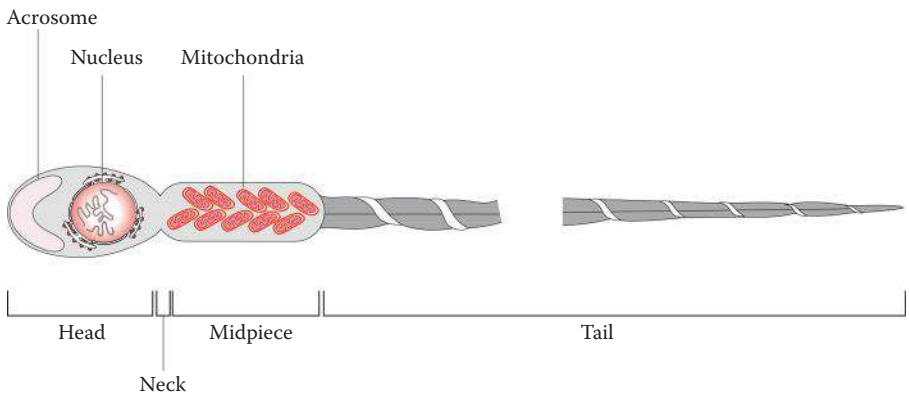

Figure 14.3 Structure of spermatozoon. (© Richard C. Li.)

14.1.2 Acid Phosphatase

Acid phosphatase (AP) consists of a group of phosphatases with optimal activity in an acidic pH environment. The greatest forensic importance of AP is that the prostate-derived AP contributes most of the AP activity present in semen. AP levels in semen are not affected by vasectomies. AP isoenzymes are also found in other tissues (Section 16.2).

The half-life of AP activity at 37°C is 6 months. However, the half-life is decreased if a sample is stored in a wet environment. AP activity can be detected from dry seminal stains stored at –20°C up to 1 year. Low levels of prostatic AP are present in the sera of healthy males. Elevated levels of prostatic AP found in serum are useful in diagnosing and monitoring prostate carcinoma. Many AP tests utilized in clinical testing may be used to identify semen for forensic applications.

14.1.3 Prostate-Specific Antigen

Prostate-specific antigen (PSA) is a major protein present in seminal fluid at concentrations of 0.5–2.0 mg/mL. PSA is produced in the prostate epithelium and secreted into the semen. PSA can also be found in the paraurethral glands, perianal glands, apocrine sweat glands, and mammary glands. Thus small quantities can be detected in urine, fecal material, sweat, and milk. PSA can also be found at much lower levels in the bloodstream. An elevated plasma PSA is present in prostate cancer patients, and it is widely used as a screening test for this disease. PSA is also elevated in cases of benign prostatic hyperplasia and prostatitis. The synthesis of PSA is stimulated by androgen, a steroid hormone.

PSA is a protein that has a molecular weight of 30 kDa and is thus also known as *P30*. It is responsible for hydrolyzing *semenogelin* (Sg), which mediates gel formation in semen (Section 14.1.4). PSA is a member of the tissue *kallikrein* (serine protease) family and is encoded by the *KLK3* locus located on chromosome 19. In addition to PSA, other tissue kallikreins encoded by *KLK2* and *KLK4* loci are expressed in the prostate. The half-life for PSA in a dried semen stain is about 3 years at room temperature. The half-life is greatly reduced when a sample is stored in wet conditions.

14.1.4 Seminal Vesicle–Specific Antigen

Human *seminal vesicle–specific antigen* (SVSA) includes two major types, semenogelin I (SgI) and semenogelin II (SgII), and constitutes the major seminal vesicle–secreted protein in semen. On ejaculation, SVSA forms a coagulum that is liquefied after a few minutes due to the degradation of SVSA by PSA. In humans, both SgI and SgII are present in a number

of tissues of the male reproductive system, including the seminal vesicles, ductus deferens, prostate, and epididymis. They are also present in several other tissues such as skeletal muscle, kidney, colon, and trachea. They have also been found in the sera of lung cancer patients.

The use of Sg as a marker for semen identification instead of PSA presents certain advantages. The concentration of Sg in seminal fluid is much higher than that of PSA, and this is beneficial for the sensitivity of detection. Sg is present in seminal fluid and absent in urine, milk, and sweat, where PSA can be found. Although Sg compounds are present in skeletal muscle, kidney, and colon, this is not a great concern because these tissue samples are not routinely collected for semen detection in sexual assault cases.

14.2 Analytical Techniques for Identifying Semen

The location of semen stains is usually carried out through visual examination. Particularly, the application of alternate light sources (ALSs) can facilitate searches for semen stains. The presumptive identification of semen is largely based on the detection of the presence of prostatic AP activity in a sample. However, most presumptive assays cannot completely distinguish prostatic AP from nonprostatic AP. Confirmatory assays for the identification of semen are available, including the microscopic examination of spermatozoa, the identification of PSA and SVSA, and the RNA-based assay.

14.2.1 Presumptive Assays

14.2.1.1 Lighting Techniques for Visual Examination of Semen Stains

Lighting techniques can be used to aid in searching for semen stains. A dried semen stain fluoresces under certain light sources such as ALSs or argon lasers. ALSs are most commonly utilized for the visual examination of semen stains (Chapter 1; Figures 14.4 and 14.5). Excitation wavelengths between 450 and 495 nm can be used, allowing for the visualization of fluorescence with orange goggles. However, this approach is not specific for semen. Other bodily fluid stains, such as saliva and urine stains, can also fluoresce with less intensity. Additionally, the intensity of the fluorescence can be affected by different colors of substrates, and the material, such as clothing, where semen stains have been deposited.

Figure 14.4 A tabletop ALS device (left) for the detection of semen stains (right). (© Richard C. Li.)

Figure 14.5 Examining a garment for semen stains. (a) A potential semen stain (labeled) is found on a garment, and (b) the stain fluoresces when irradiated with an ALS device. (© Richard C. Li.)

14.2.1.2 Acid Phosphatase Techniques

14.2.1.2.1 Colorimetric Assays

Colorimetric assays can be used for the presumptive identification of semen. The AP contained in semen can hydrolyze a variety of phosphate esters. It catalyzes the removal of the phosphate group from a substrate (Figure 14.6). Subsequently, an insoluble colored precipitate at sites of acid phosphatase activity is formed with a stabilized diazonium salt (usually in the form of zinc double salts).

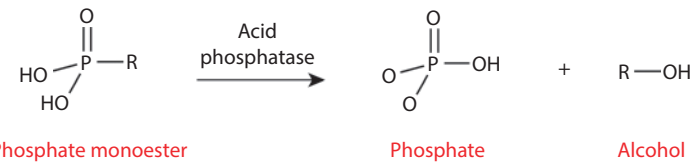

Figure 14.6 Reaction catalyzed by acid phosphatases (EC 3.1.3.2). The optimal pH of the reaction is usually under pH 7.

However, interference during a test by nonprostatic AP isoenzymes (multiple forms of AP), such as contamination by AP commonly present in vaginal secretions (Chapter 16), can create problems in specimens collected from victims. Thus, it is desirable to be able to increase the specificity of the assay for prostatic AP. One solution is the application of substrates that are hydrolyzed rapidly by the prostatic enzyme and at a slower rate by the other forms of AP isoenzymes. For example, α-naphthylphosphate and thymolphthalein monophosphate are more specific to prostatic AP than phenyl phosphate and 4-nitrophenyl phosphate (Figure 14.7). The most common method for forensic applications is the use of α-naphthyl phosphate as a substrate. In the presence of AP, α-naphthylphosphate is hydrolyzed to phosphate and α-naphthol. Subsequently, the Fast Blue B, a stabilized diazonium salt, is added to carry out an azo coupling reaction, producing a purple azo dye (Figures 14.8 through 14.10).

Prostatic AP is water soluble. Thus, a moistened cotton swab or piece of filter paper can be used to transfer a small amount of sample from a stain by briefly pressing onto the questioned stain area. The α-naphthylphosphate reagent is added to the swab or filter paper followed by the addition of Fast Blue B reagent. If a purple coloration develops within 1 min, the test is considered a positive indication for semen. Color that develops after more than 1 min may arise from the activity of nonprostatic AP.

Additionally, the prostatic enzyme is strongly inhibited by dextrorotatory tartrate ions. Thus, these inhibitors, particularly tartrate, allow a distinction to be made between prostatic AP and other AP isoenzymes. Prostate and vaginal acid phosphatase can also be distinguished by using gel electrophoresis (Chapter 16).

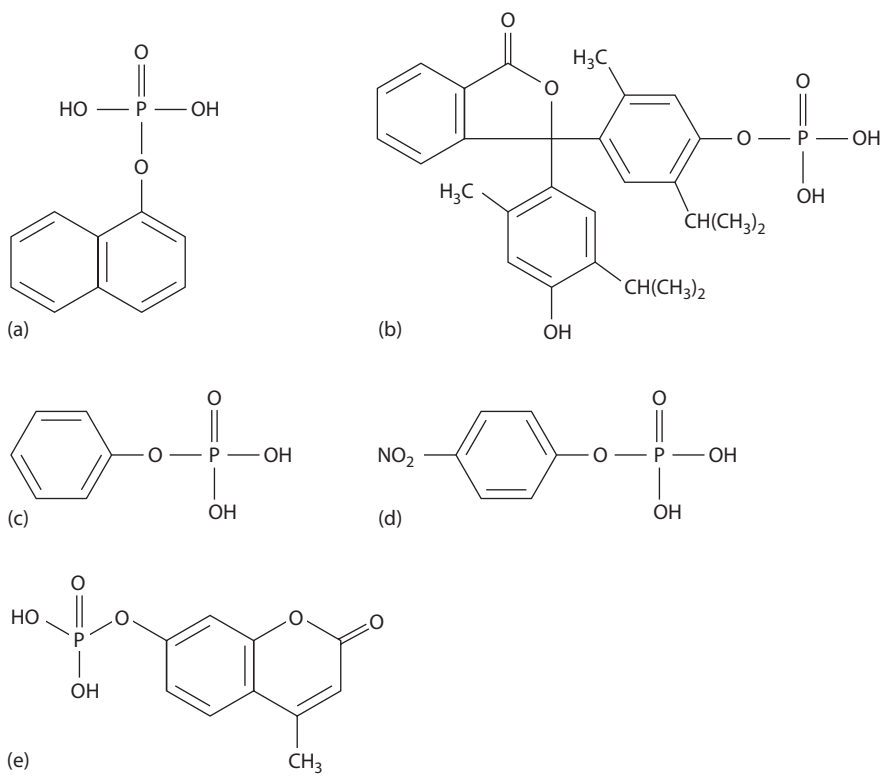

Figure 14.7 Chemical structures of acid phosphatase substrates. (a) α-Naphthyl phosphate, (b) thymolphthalein monophosphate, (c) phenyl phosphate, (d) 4-nitrophenyl phosphate, and (e) MUP.

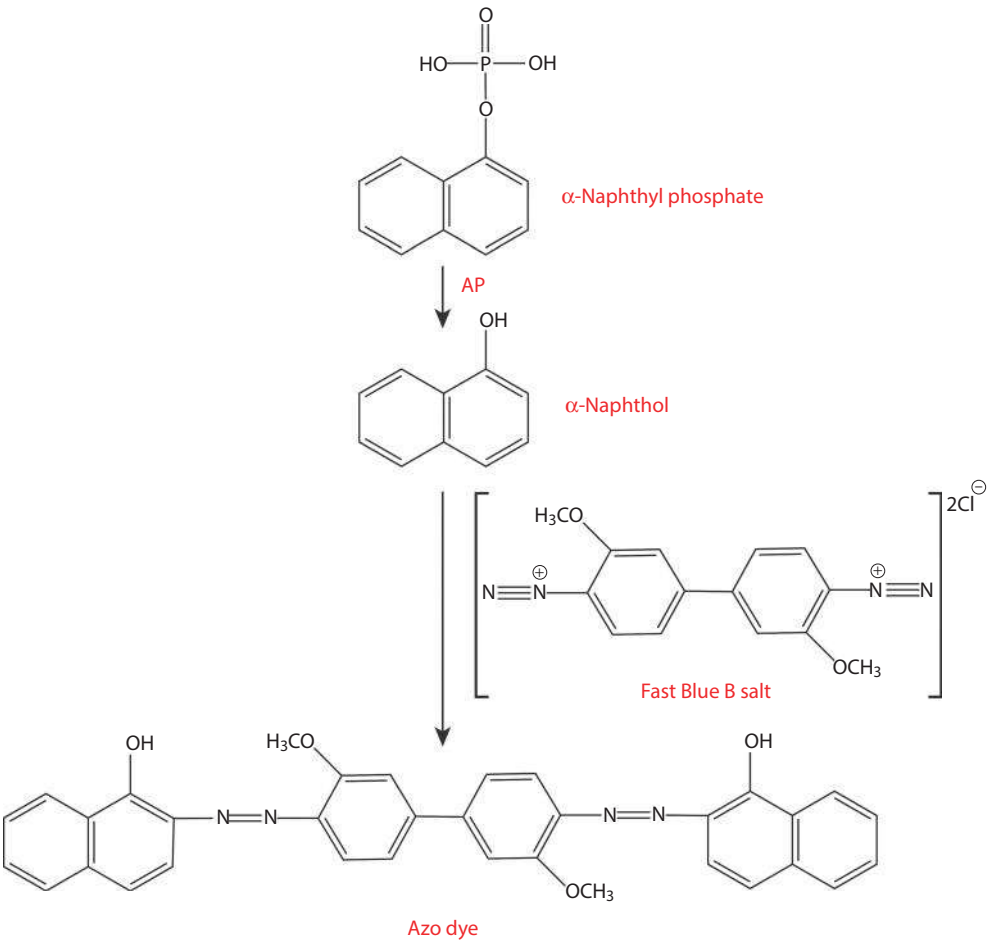

Figure 14.8 A colorimetric acid-phosphatase assay. In this assay, α-naphthylphosphate is hydrolyzed by acid phosphatase to phosphate and α-naphthol. The α-naphthol is subsequently converted into a purple azo dye with a diazonium salt such as Fast Blue B salt. AP, acid phosphatase.

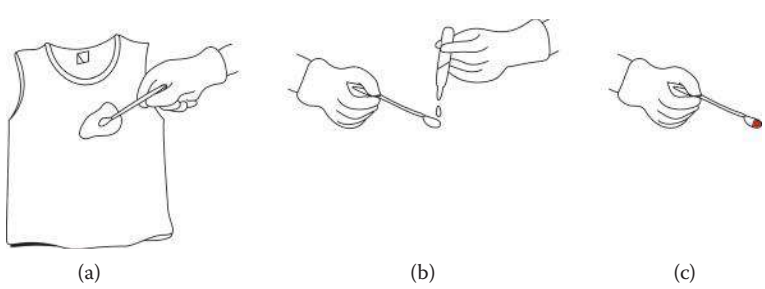

Figure 14.9 AP colorimetric assay. (a) A small amount of sample from the stain is transferred using a moistened cotton swab. (b) The substrate reagent is applied, followed by adding Fast Blue B reagent. (c) Purple coloration indicates a positive reaction. (© Richard C. Li.)

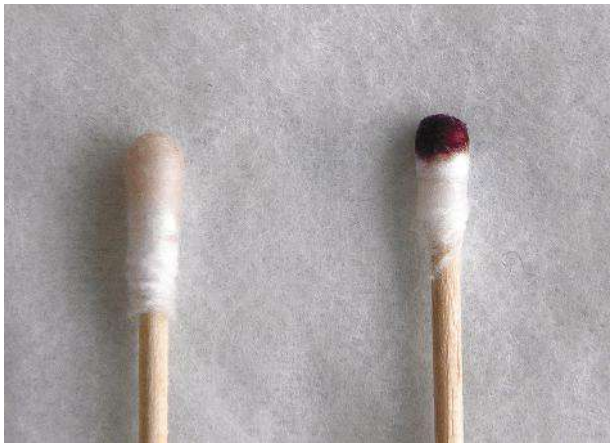

Figure 14.10 Photo of a colorimetric acid-phosphatase assay using α-naphthylphosphate as a substrate. (© Richard C. Li.)

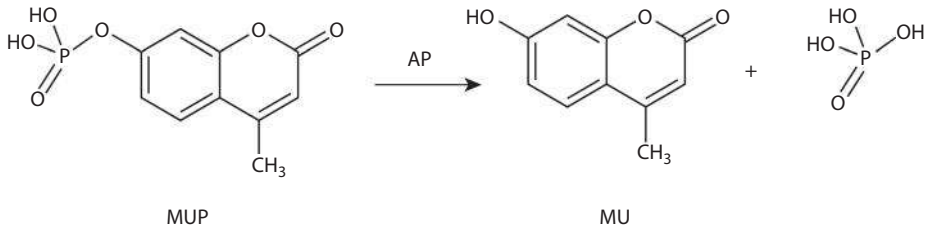

Figure 14.11 The principle of MUP assay for detecting acid phosphatase activity. In the presence of acid phosphatase, 4-methylumbelliferone phosphate (MUP) is hydrolyzed, forming phosphate and 4-methylumbelliferone (MU), which fluoresces. AP, acid phosphatase.

14.2.1.2.2 Fluorometric Assays

Fluorometric methods are more sensitive than the colorimetric detection of AP and are used for semen stain mapping. AP catalyzes the removal of the phosphate residue on a 4-methylumbelliferone phosphate (MUP) substrate (Figure 14.11), a reaction that generates fluorescence under ultraviolet light. A piece of moistened filter paper, marked for proper orientation and identification, is used for transferring the prostatic AP. The evidence to be tested, a garment for example, is covered by the filter paper. Gloved hands are used to press the filter paper onto the stained area, ensuring that the evidence is in close contact with the paper. The filter paper is lifted from the evidence and examined in a dark room using long-wave ultraviolet light to detect any background fluorescence, which is then marked on the paper. The paper can then be sprayed with MUP reagent in a fume hood. The AP reaction on the paper can be visualized immediately. Areas where semen is present can be visualized as fluorescent areas on the filter paper (Figure 14.12).

14.2.2 Confirmatory Assays

14.2.2.1 Microscopic Examination of Spermatozoa

The cells from a questioned stain on an absorbent material can be transferred to a microscope slide by extracting a small portion of a stain with water, followed by gentle vortexing. The suspension is then transferred to a slide and evaporated at room temperature or fixed with low heat. Alternatively, it can be transferred by dampening the stain with water and rubbing or rolling it onto a microscope slide.

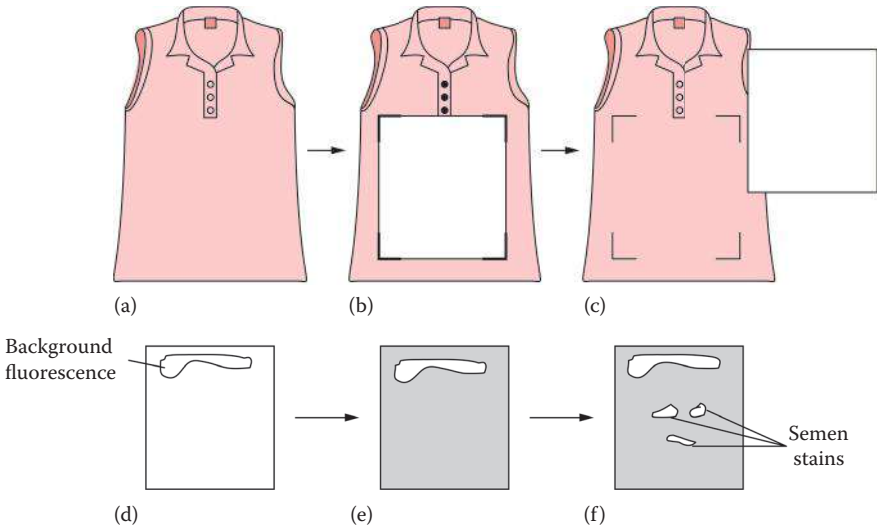

Figure 14.12 Fluorometric assay of acid phosphatase for locating semen stains. Evidence item (a) is closely covered by a piece of moistened filter paper (b) to allow transfer of a small amount of stain. The orientation of the paper is marked (c). The paper is lifted. The background fluorescence is marked under UV light (d). The filter paper is treated with MUP (e). The presence of fluorescence under ultraviolet light indicates semen stains (f). (© Richard C. Li.)

Microscopic identification of spermatozoa provides the proof of a seminal stain. Histological staining can facilitate microscopic examination. The most common staining technique is the *Christmas tree stain* (Figure 14.13). The red component known as Nuclear Fast Red (NFR) is a dye used for staining the nuclei of spermatozoa in the presence of aluminum ions. The green component, picroindigocarmine (PIC), stains the neck and tail portions of the sperm. The acrosomal cap and the nucleus stain pink-red, and the sperm tails and the midpiece stain blue-green. Epithelial cells, if present in the sample, appear blue-green and have red nuclei. Additionally, fluorescent detection utilizing SPERM HY-LITER Fluorescent Staining Kit can facilitate the identification of spermatozoa.

Laser capture microdissection (LCM) has been shown to be an effective technique for separating spermatozoa from nonsperm cells (i.e., epithelial cells from the victim) on a glass slide (Figure 14.14). This technique involves using a thin layer of a thermosensitive polymer that is placed on the surface of an LCM cap. Once spermatozoa are identified on the slide under a microscope, a polymer-containing LCM cap is placed over the spermatozoa on the slide. An infrared laser melts the polymer and causes it to adhere only to the targeted spermatozoa. The spermatozoa are then lifted off the slide. This allows spermatozoa to be separated and placed into snap-cap tubes for forensic DNA analysis.

14.2.2.2 Identification of Prostate-Specific Antigen

Over the years, a number of methods have been utilized to detect PSA: immunodiffusion, immunoelectrophoresis, enzyme-linked immunosorbent assay (ELISA), and immunochromatographic assays. ELISA and immunochromatographic assays have been found to be the most sensitive methods (Chapter 11).

14.2.2.2.1 Immunochromatographic Assays

Commercially produced immunochromatographic kits such as the PSA-check-1 (VED-LAB, Alencon), Seratec® PSA Semiquant (Seratec Diagnostica, Göttingen), and One Step ABAcard

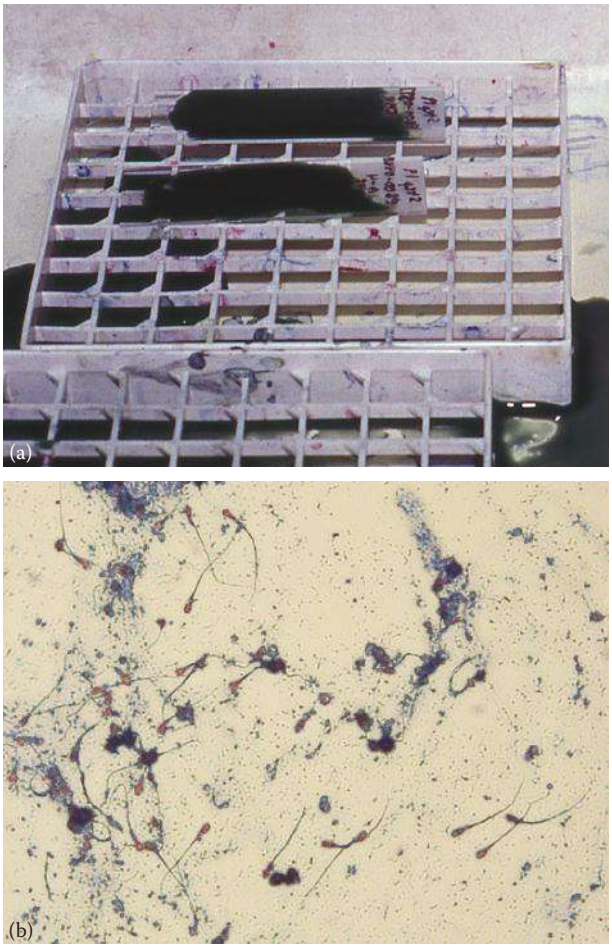

Figure 14.13 Human spermatozoa stained with Christmas tree stain. (a) Staining spermatozoa on a microscope slide and (b) stained spermatozoa. (© Richard C. Li.)

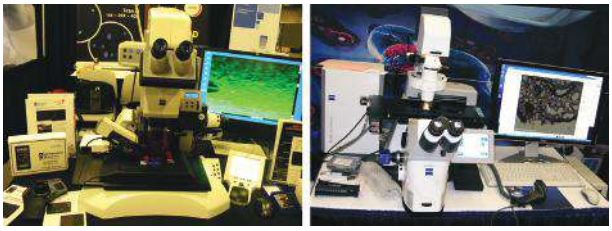

Figure 14.14 A microscopic device for fluorescent detection utilizing SPERM HY-LITER Fluorescent Staining Kit (left) and a laser-capture microdissection device for separating sperm cells from other types of cells (right). (© Richard C. Li.)

PSA® (Abacus Diagnostics, California) are available. These devices utilize antihuman PSA antibodies. In the ABAcard PSA® assay, a labeled monoclonal antihuman PSA antibody is contained in a sample well, a polyclonal antihuman PSA antibody is immobilized on a test zone of a nitrocellulose membrane, and an antiglobulin that recognizes the antibody is immobilized on a control zone (Figures 14.15 and 14.16).

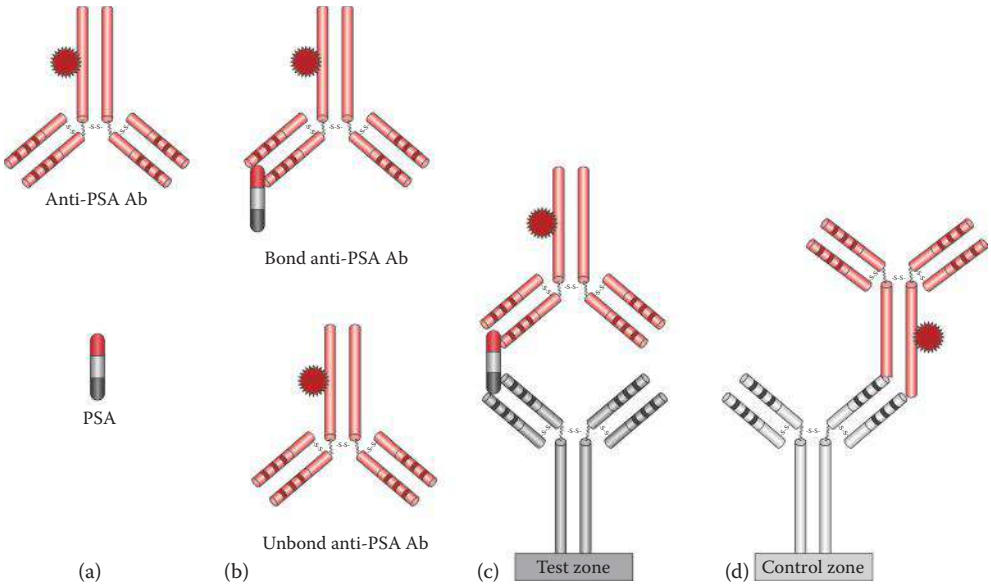

Figure 14.15 Immunochromatographic assays for identification of PSA in semen. (a) In a sample well, PSA in a semen sample is mixed with labeled anti-PSA Ab. (b) The PSA binds to the labeled anti-PSA Ab to form a labeled Ab–PSA complex. (c) At the test zone, the labeled complex binds to an immobilized anti-PSA Ab to form a labeled Ab–PSA–Ab sandwich. (d) At the control zone, the labeled anti-PSA Ab binds to an immobilized antiglobulin and is captured at the control zone. Ab and PSA represent antibody and prostate-specific antigen, respectively. (© Richard C. Li.)

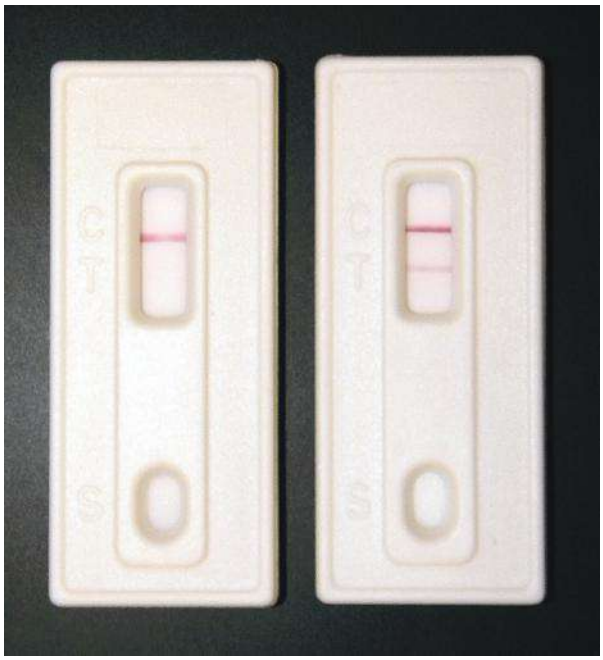

Figure 14.16 Semen identification using an immunochromatic device (ABAcard PSA). The negative (left) and positive (right) results are shown. The "C" band indicates that the test is valid. The "T" band indicates the presence of human blood. The sample well is labeled as "S". (© Richard C. Li.)

The assay is carried out by loading an extracted sample into the sample well. The antigen in the sample binds to the labeled antibody in the sample well to form an antigen–antibody complex. The complex then diffuses across the nitrocellulose membrane. At the test zone, the immobilized antihuman PSA antibody binds with the antigen–antibody complex to form an antibody–antigen–antibody sandwich. The ABAcard PSA® uses a pink dye that allows for the visualization of a positive test with a pink line at the test zone. In the control zone, unbound labeled antihuman PSA antibody binds to the immobilized antiglobulin. This antibody–antiglobulin complex at the control zone also results in a pink line. The test is considered valid only if the line in the control zone is observed.

The presence of human PSA results in a pink line at both the test and control zones. The absence of human PSA produces a pink line in the control zone only. A positive result can appear within 1 min; a negative result is read after 10 min. However, the high-dose hook effect, an artifact that may cause false-negative results (Chapter 10), occurs when high quantities of seminal fluid are tested.

14.2.2.2.2 ELISA

The ELISA method can be used to detect PSA with anti-PSA antibodies. The most common method used in forensic serology is antibody sandwich ELISA, in which an antibody–antigen–antibody sandwich complex is formed (Figure 14.17). The intensity of the signal can be detected and is proportional to the amount of bound antigen. The amount of PSA can also be quantified by comparing a standard with known concentrations. Although this method is specific and highly sensitive, it is time-consuming. Chapter 11 discusses the principle of ELISA in further detail.

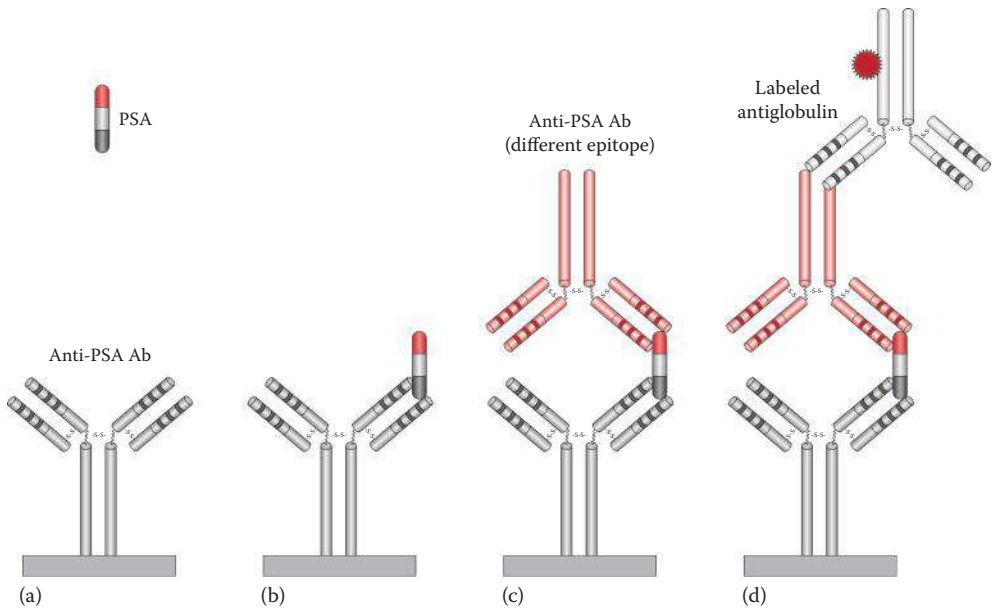

Figure 14.17 Use of ELISA for identification of PSA in semen. (a) Sample containing PSA is applied to polystyrene tubes in which anti-PSA Ab is immobilized. (b) The PSA binds to immobilized Ab to form a PSA–Ab complex. (c) A second anti-PSA Ab, specific for a different epitope of PSA, is added to form an Ab–PSA–Ab sandwich. (d) A labeled antiglobulin then binds to the Ab–PSA–Ab sandwich. The bound antiglobulin can be detected by various reporting schemes. Ab and PSA represent antibody- and prostate-specific antigen, respectively. (© Richard C. Li.)

14.2.2.3 Identification of Seminal Vesicle–Specific Antigen

14.2.2.3.1 Immunochromatographic Assays

Commercially produced immunochromatographic kits include the RSID®-Semen test (Independent Forensics, Hillside, IL) and the Nanotrap Sg. In the RSID®-Semen assay, a labeled monoclonal anti-Sg antibody is contained in a sample well, and a second monoclonal anti-Sg antibody, to a different epitope of Sg, is immobilized on the test zone of the membrane. An anti-globulin that recognizes the antibody is immobilized on a control zone (Figure 14.18).

The sample can be prepared by cutting a small portion of a stain or a swab and is extracted for 1–2 h in an extraction buffer (200–300 μL). Approximately 10% of the extract is removed and mixed with the running buffer. The assay is carried out by loading an extracted sample into the sample well. The antigen in the sample binds to the labeled anti-Sg antibody in the sample well to form a labeled antibody–antigen complex that then diffuses across the membrane. At the test zone, the solid-phase anti-Sg antibody binds with the labeled complex to form a labeled antibody–antigen–antibody sandwich. The antigen in the sample produces a pink line at the test zone. In the control zone, unbound labeled anti-Sg antibody binds to the solid-phase antiglobulin. This labeled antibody–antiglobulin complex at the control zone also results in a pink line. The presence of Sg generates a pink line at both the test and control zones. The absence of Sg results in a pink line in the control zone only. Results may be read after 10 min.

Validation studies have revealed that the sensitivity of the RSID-Semen kit for detecting seminal fluid can be as low as a 5×10^4-fold dilution. Species specificity studies have shown no cross-reactivity with various animal species including ruminants and small mammals. Bodily fluid specificity studies have also shown that the assay is not responsive to human blood, saliva,

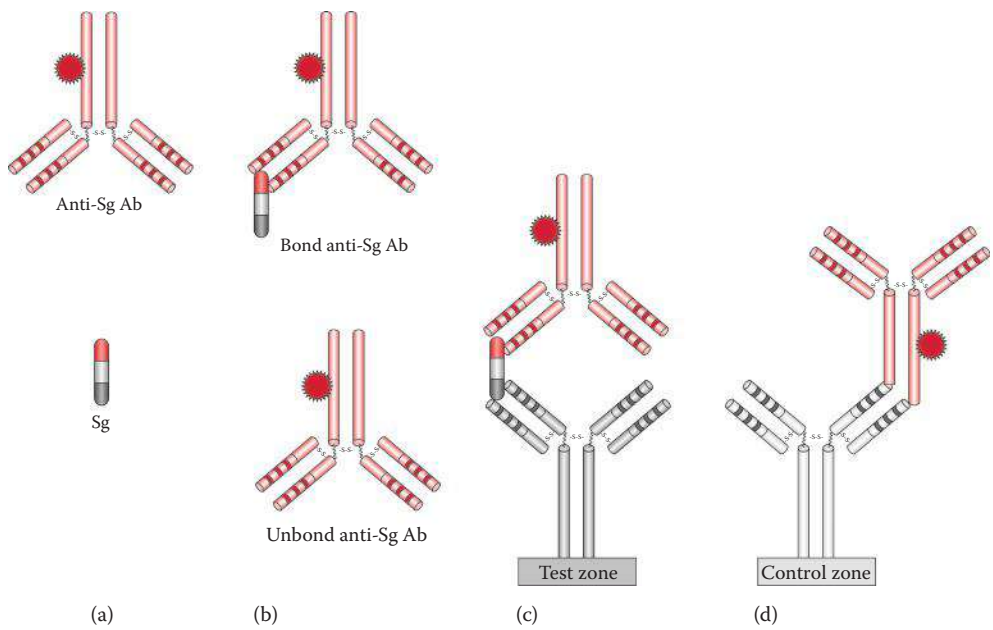

Figure 14.18 Immunochromatographic assays for identification of semenogelin (Sg) in semen. (a) In a sample well, Sg in a semen sample is mixed with labeled anti-Sg Ab. (b) Sg binds to the labeled anti-Sg Ab to form a labeled Ab–Sg complex. (c) At the test zone, the labeled complex binds to an immobilized anti-Sg Ab to form a labeled Ab–Sg–Ab sandwich. (d) At the control zone, the labeled anti-Sg Ab binds to an immobilized antiglobulin and is captured at the control zone. Ab represents antibody. (© Richard C. Li.)

urine, sweat, fecal matter, milk, or vaginal secretions. The assay results are not affected by condom lubricants or spermicides such as nonoxynol-9 and menfegol. However, the high-dose hook effect occurs when more than 3 μL of seminal fluid is tested.

14.2.2.3.2 ELISA

Identification of Sg for semen detection has also been carried out with ELISA. Anti-Sg antibodies are utilized. An antibody–antigen–antibody sandwich complex is formed (Figure 14.19). The intensity of the colorimetric or fluorometric signals can be detected spectrophotometrically and is proportional to the amount of bound antigen. The amount of Sg can be quantified by comparing a standard with known concentrations.

14.2.2.4 RNA-Based Assays

RNA-based assays (Chapter 11) have been developed to identify semen. The assays are based on the expression of certain genes in certain cell or tissue types. Thus, the techniques used in the identification of semen are based on the detection of specific types of mRNA expressed exclusively in spermatozoa and in certain cells of male accessory glands. These assays utilize reverse transcriptase polymerase chain reaction (RT-PCR; see Chapter 7) methods to detect gene expression levels of mRNAs for semen identification. Table 14.1 lists the tissue-specific genes utilized for semen identification. Compared to conventional assays used for semen identification, the RNA-based assay has higher specificity and is amenable to automation. However, one limitation is that the RNA is unstable due to degradation by endogenous ribonucleases.

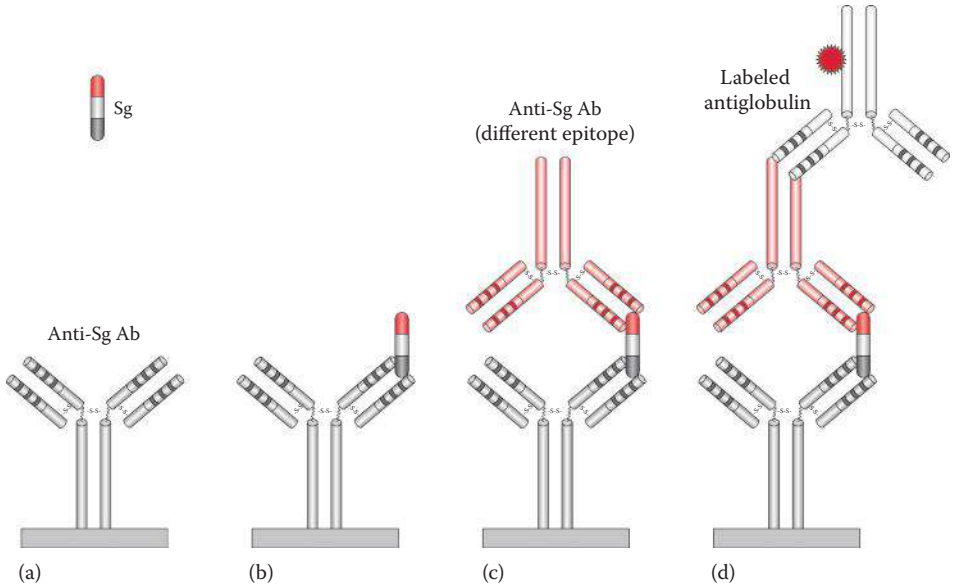

Figure 14.19 Use of ELISA for identification of semenogelin (Sg) in semen. (a) Sample containing Sg is applied to polystyrene tubes in which anti-Sg Ab is immobilized. (b) Sg binds to immobilized Ab to form a Sg–Ab complex. (c) A second anti-Sg Ab, specific for a different epitope of Sg, is added to form an Ab–Sg–Ab sandwich. (d) A labeled antiglobulin then binds to the Ab–Sg–Ab sandwich. The bound antiglobulin can be detected by various reporting schemes. Ab represents antibody. (© Richard C. Li.)

Table 14.1 Application of RT-PCR for Semen Identification

Gene Symbol	Gene Product	Description	Further Reading
KLK3	Kallikrein 3	Also called prostate-specific antigen (PSA)	Gelmini et al. (2001)
PRM1	Protamine 1	DNA-binding proteins involved in condensation of sperm chromatin	Steger et al. (2000)
PRM2	Protamine 2	DNA-binding proteins involved in condensation of sperm chromatin	Steger et al. (2000)

Source: Adapted from Juusola, J. and Ballantyne, J., *Forensic Sci Int*, 152, 1–12, 2005; Nussbaumer, C., Gharehbaghi-Schnell, E., and Korschineck, I., *Forensic Sci Int*, 157, 181–186, 2006.

Bibliography

Ablett, P.J., The identification of the precise conditions for seminal acid phosphatase (SAP) and vaginal acid phosphatase (VAP) separation by isoelectric focusing patterns. *J Forensic Sci Soc*, 1983, **23**(3): 255–256.

Allard, J.E., The collection of data from findings in cases of sexual assault and the significance of spermatozoa on vaginal, anal and oral swabs. *Sci Justice*, 1997, **37**(2): 99–108.

Allard, J.E., et al., A comparison of methods used in the UK and Ireland for the extraction and detection of semen on swabs and cloth samples. *Sci Justice*, 2007, **47**(4): 160–167.

Allen, S.M., An enzyme linked immunosorbent assay (ELISA) for detection of seminal fluid using a monoclonal antibody to prostatic acid phosphatase. *J Immunoassay*, 1995, **16**(3): 297–308.

Allery, J.P., et al., Cytological detection of spermatozoa: Comparison of three staining methods. *J Forensic Sci*, 2001, **46**(2): 349–351.

Anoruo, B., et al., Isolating cells from non-sperm cellular mixtures using the PALM microlaser micro dissection system. *Forensic Sci Int*, 2007, **173**(2–3): 93–96.

Anslinger, K., et al., Sex-specific fluorescent labelling of cells for laser microdissection and DNA profiling. *Int J Legal Med*, 2007, **121**(1): 54–56.

Astrup, B.S., et al., Detection of spermatozoa following consensual sexual intercourse. *Forensic Sci Int*, 2012, **221**(1–3): 137–141.

Auvdel, M.J., Comparison of laser and high-intensity quartz arc tubes in the detection of body secretions. *J Forensic Sci*, 1988, **33**(4): 929–945.

Baechtel, S., The identification and individualization of semen stains, in R. Saferstein (ed.), *Forensic Science Handbook*, pp. 369–374, 1988. Englewood Cliffs, NJ: Prentice Hall.

Bauer, M. and D. Patzelt, Protamine mRNA as molecular marker for spermatozoa in semen stains. *Int J Legal Med*, 2003, **117**(3): 175–179.

Baxter, S.J., Immunological identification of human semen. *Med Sci Law*, 1973, **13**(3): 155–165.

Berti, A., et al., Expression of seminal vesicle–specific antigen in serum of lung tumor patients. *J Forensic Sci*, 2005, **50**(5): 1114–1145.

Bitner, S.E., False positives observed on the Seratec(R) PSA SemiQuant cassette test with condom lubricants. *J Forensic Sci*, 2012, **57**(6): 1545–1548.

Blake, E.T. and G.F. Sensabaugh, Genetic markers in human semen: A review. *J Forensic Sci*, 1976, **21**(4): 785–796.

Chapman, R.L., N.M. Brown, and S.M. Keating, The isolation of spermatozoa from sexual assault swabs using proteinase K. *J Forensic Sci Soc*, 1989, **29**(3): 207–212.

15

Identification of Saliva

15.1 Biological Characteristics of Saliva

The human salivary glands produce 1.0–1.5 L of saliva daily. About 70% of saliva is produced from the submandibular salivary glands, 25% from the parotids, and 5% from the sublingual salivary glands (Figure 15.1). Although a continuous basal level of saliva secretion is maintained, a large amount of saliva is produced during eating. Saliva is largely water containing small quantities of electrolytes, proteins, antibodies, and enzymes. The digestion of carbohydrates in the diet begins in the oral cavity, where amylase in the saliva breaks down carbohydrates such as starch. Thus, detecting amylase indicates the presence of saliva.

15.1.1 Amylases

Amylases are enzymes that cleave polysaccharides such as starches, which are composed of D-glucose units connected by α1→4 linkages. Starches contain two types of glucose polymers: *amylose* and *amylopectin* (Figure 15.2). Amylose consists of long, linear chains of glucose residues connected by α1→4 linkages. Amylopectin is highly branched and consists of linear chains of glucose residues connected by α1→4 linkages with the branch points connected by α1→6 linkages. Both linear amylose and amylopectin can be hydrolyzed by amylase by cleaving the chains at alternate α1→4 linkages. Amylase cleaves off one maltose (two glucose units) at a time. However, the α1→6 linkages at the branch points are not cleaved by the amylase.

Two types of amylases are characterized. β-*Amylases* found in plant and bacterial sources cleave only at the terminal reducing end of a polysaccharide chain. The end of a chain with a free anomeric carbon (not involved in a glycosidic bond) is called the reducing end. Human α-*amylases* cleave at α1→4 linkages randomly along the polysaccharide chain.

Human α-amylases have two major isoenzymes (multiple forms that differ in their amino acid sequences). Human *salivary* α-*amylase* (HSA) is encoded by the *Amy1* locus, synthesized at the salivary glands and secreted into the oral cavity. Human *pancreatic* α-amylase (HPA), encoded by the *Amy2* locus, is synthesized by the pancreas and secreted into the duodenum through the pancreatic duct. The amino acid sequences of the HSA and HPA are highly homologous. Therefore, monoclonal antibodies against HSA also cross-react with HPA. However, HSA is inactivated by acids in the stomach, while most HPA is inactivated in the lower portions of the intestine, and some amylase activity remains in the feces.

Amylase activity is found in various bodily fluids including semen, tears, milk, perspiration, and vaginal secretions. Most amylase present in normal serum consists of HPA and HSA. The amylases are small molecules and can pass through the glomeruli of the kidney (Chapter 17).

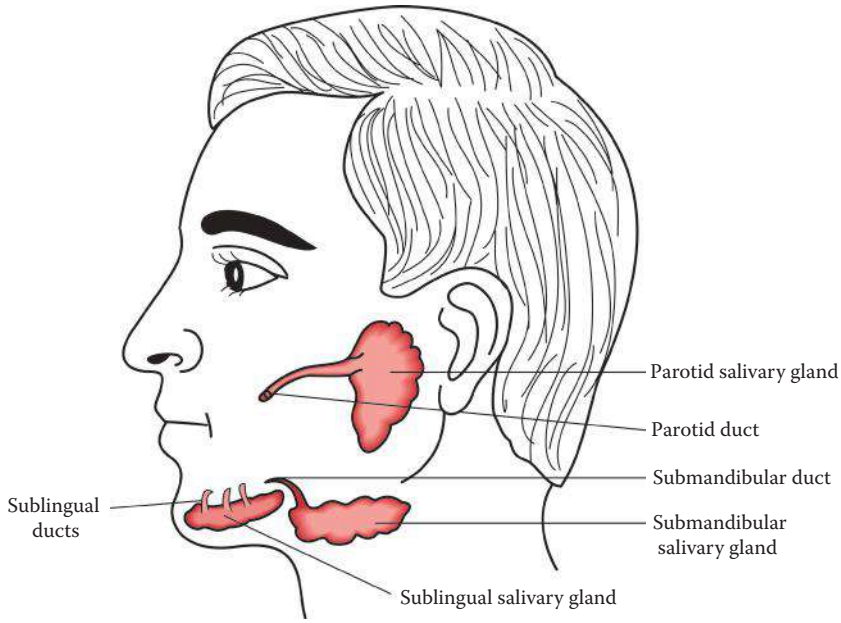

Figure 15.1 Human salivary glands. (© Richard C. Li.)

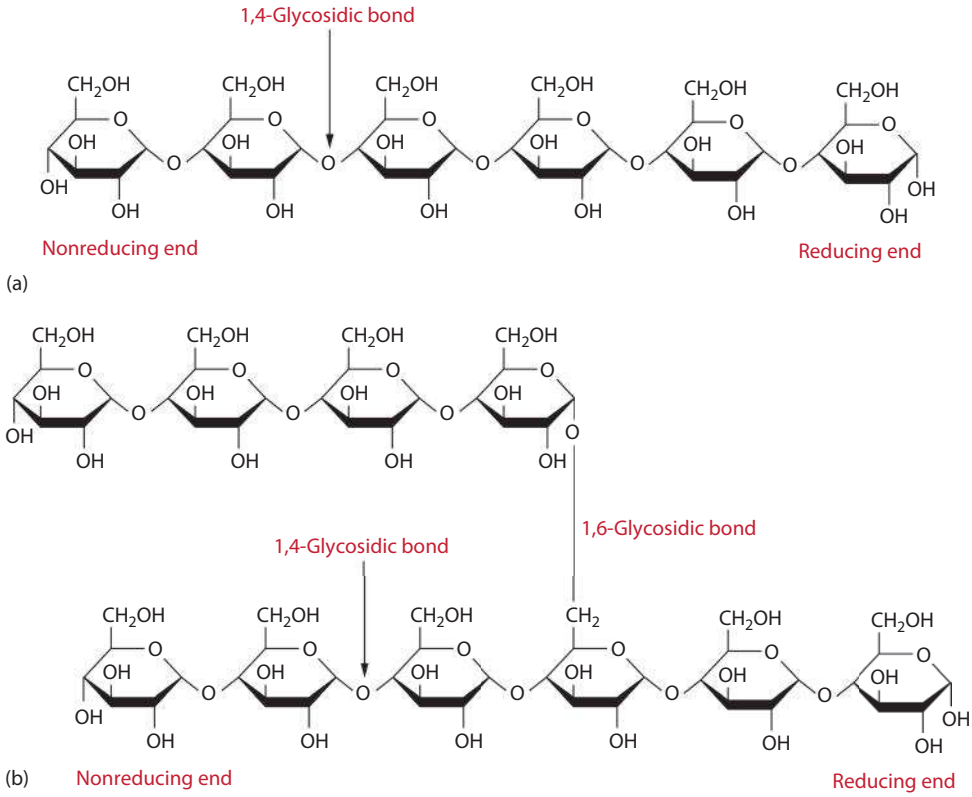

Figure 15.2 Chemical structures of polysaccharides found in starch: (a) amylose and (b) amylopectin.

Thus, the amylase present in urine is derived from plasma. Amylase can be inactivated under boiling temperatures and strong acidic and alkaline conditions. Based on various studies, its stability varies from a few weeks to several months.

15.2 Analytical Techniques for Identification of Saliva

The use of alternate light sources facilitates the search for and the visual examination of saliva stains. The identification of saliva is largely based on detecting the presence of amylase in a sample. Two types of amylase assays can be utilized. The first type measures the enzymatic activity of total amylase. This type of assay cannot distinguish HSA from other amylases including HPA and nonhuman amylases, such as those from plants, animals, and microorganisms. The second type of assay, which includes direct detection of HSA proteins and RNA-based assays, is more confirmatory than enzymatic assays.

15.2.1 Presumptive Assays

15.2.1.1 Visual Examination

The lighting techniques used to search for semen stains can be utilized in searching for saliva stains. For example, a 470 nm excitation wavelength can be used with orange goggles to allow visualization of fluorescence. However, the fluorescence of a saliva stain is usually less intense than that of a seminal stain (Figure 15.3). Microscopic examination with proper histological staining can also be performed to identify the buccal epithelial cells, indicating the presence of a saliva stain (Figure 15.4).

15.2.1.2 Determination of Amylase Activity

15.2.1.2.1 Starch–Iodine Assay

Iodine (I_2) is used to test for the presence of starch. The amylose in starch reacts strongly with iodine to form a dark blue complex, while amylopectin develops a reddish-purple color. In the presence of amylases, starch is broken down to mono- or disaccharides. Consequently, such colors do not develop when iodine is added.

Figure 15.3 Examining saliva stains using an ALS technique. (© Richard C. Li.)

Figure 15.4 Buccal epithelial cells. (© Richard C. Li.)

A common configuration of the method is the radial diffusion assay (Figure 15.5). An agar gel containing starch is prepared. A sample well is created by punching a hole in the gel and an extract of the questioned sample is placed into the well. If amylase is present in the sample, it diffuses from the sample well and hydrolyzes the starch in the gel. The gel is then stained using an iodine solution. A clear area in the gel indicates amylase activity, and the size of the clear area is proportional to the amount of amylase in the sample. A linear standard curve

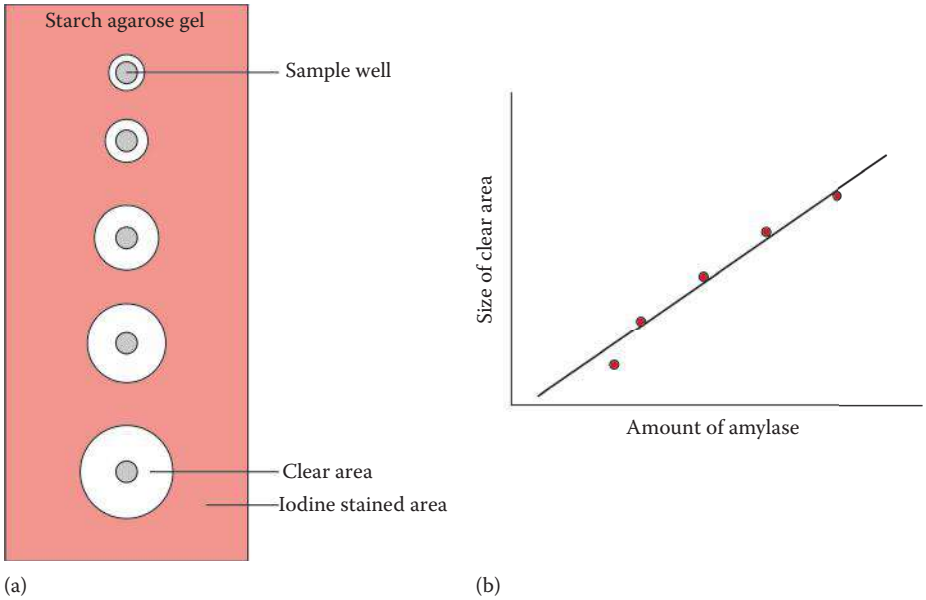

Figure 15.5 Radial diffusion assay for the identification of amylase. (a) Known amounts of amylase standards are applied to the well and allowed to diffuse. (b) The size of the clear area arising from amylase activity is plotted. The standard curve can be used to determine the amount of amylase in a questioned sample. (© Richard C. Li.)

(in log scale) can be prepared using a series of standard amylases with known concentrations. The amount of amylase can be quantified by comparing the results with the standard curve. However, this assay is not specific to HSA and can produce false-positive results. Moreover, the quantities of amylase can be used for determining the amount of sample needed for Y-STR analysis (Chapter 19), which now can be determined by using Y-chromosome-specific quantitative PCR assays (Chapter 6).

15.2.1.2.2 Colorimetric Assays

Dye-labeled amylase substrates such as dye-conjugated amylose or amylopectin are utilized. These substrates are not soluble in water. In the presence of amylase, the dye-containing moieties are cleaved and are soluble in water to produce a color. The degree of coloration, which can be measured colorimetrically by spectrophotometric methods, is proportional to the amount of amylase in the sample. Most of these assays are not HSA specific, although their specificity can be tested by using inhibitors that preferentially inhibit HSA, such as α-amylase inhibitors derived from the seeds of the wheat plant, *Triticum aestivum*. These amylase activity assays are considered presumptive, which means they are not conclusive for the presence of saliva in a sample.

While many substrates are available, Phadebas reagent (Pharmacia) is usually used in forensic laboratories. Produced in a tablet form, it is used to detect α-amylase in specimens for clinical diagnostic purposes. A small portion (approximately 3 mm^2) of a sample is cut and placed in a tube and incubated for 5 min at 37°C. One Phadebas tablet is added to each tube and mixed. Samples are then incubated for 15 min at 37°C, and the reaction is stopped at an alkaline pH by the addition of sodium hydroxide. The amylase substrate is an insoluble blue dye conjugated to starch. Amylase hydrolyzes the substrate to generate a blue color that can be measured at 620 nm using a spectrophotometer. The optical density of the supernatant is read and can be converted to amylase units by comparing to a standard curve.

The amylase assay can also be used for amylase mapping as a method of searching for possible saliva stains (Figures 15.6 and 15.7). These assays are based on the principle that amylase is water soluble and can be transferred from evidence to filter paper and then analyzed via colorimetric assay. This procedure is also referred to as a *press test*. The sensitivity of the method is similar to that of the test tube method.

The substrate can be prepared by evenly spraying the Phadebas reagent on a sheet of filter paper and allowing it to air-dry. The dried substrate-containing paper can be used immediately or stored until needed. To perform amylase mapping, a piece of paper is placed over the entire area to be tested (the item to be tested must be fairly flat to ensure good contact with the paper). The paper is dampened slightly by spraying with distilled water. An outline may be drawn on the paper to aid in locating stains. A piece of plastic wrap is placed on top to prevent the paper from drying during the assay, and a weight is applied to ensure close contact of substrate and evidence. The test is observed every minute for the first 10 min, and every 5 min thereafter up to 40 min, when a positive reaction should appear as a light blue area.

The SALIgAE® kit (Abacus Diagnostics), another commercially available colorimetric assay, has been validated for saliva identification (Figure 15.8). Its manufacturer also produces the SALIgAE spray kit, which can be used for amylase mapping.

15.2.2 Confirmatory Assays

15.2.2.1 Identification of Human Salivary α-Amylase

15.2.2.1.1 Immunochromatographic Assays

Commercially produced immunochromatographic kits include the RSID®-Saliva kit (Independent Forensics). A labeled monoclonal anti-HSA antibody is contained in a sample well. A second monoclonal anti-HSA antibody is immobilized onto a test zone of a membrane, and an antiglobulin that recognizes the antibody is immobilized onto a control zone (Figures 15.9 and 15.10).

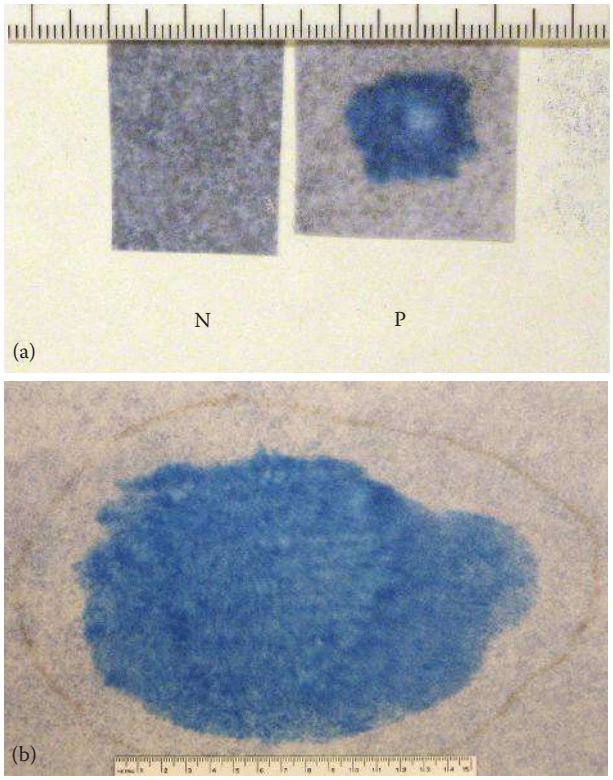

Figure 15.6 Amylase colorimetric assay using Phadebas reagent. (a) A spot test for saliva and (b) amylase mapping result showing a saliva-stained area. N, a negative result; P, a positive result. (© Richard C. Li.)

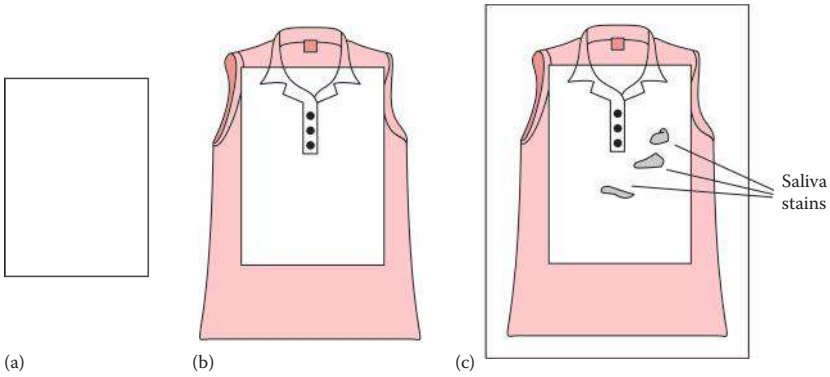

Figure 15.7 Amylase mapping for saliva stains. (a) Amylase substrate is sprayed on a sheet of filter paper. (b) Substrate-containing paper is placed over the area to be tested. The orientation of the filter paper is marked to aid in locating the stain. (c) The filter paper is dampened by spraying it with water, and plastic wrap is placed on top to prevent the paper from drying. Blue color indicates a positive reaction. (© Richard C. Li.)

Figure 15.8 Saliva identification using SALIgAE reagent. P, a positive result; N, a negative result; S, a positive control. (© Richard C. Li.)

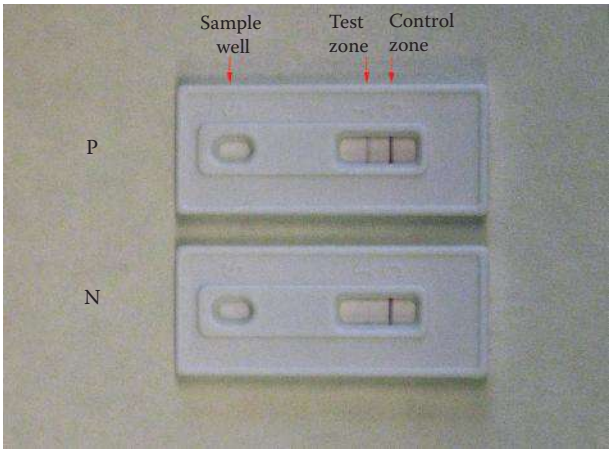

Figure 15.9 Identification of saliva using immunochromatographic assays. P, a positive result; N, a negative result. (© Richard C. Li.)

A sample can be prepared by cutting out a small portion of a stain or a swab. Each sample is extracted for 1–2 h in 200–300 μL of an extraction buffer. Approximately 10% of the extract is removed and mixed with a running buffer. The assay is carried out by loading an extracted sample into the sample well where the antigen in the sample binds to the labeled anti-HSA antibody in the well to form a labeled antibody–antigen complex. The complex then diffuses across the membrane to the test zone, where the solid-phase anti-HSA antibody binds with the labeled complex to form a labeled antibody–antigen–antibody sandwich.

The presence of antigen in the sample results in a pink line at the test zone. In the control zone, unbound labeled anti-HSA antibody binds to the solid-phase antiglobulin. The labeled antibody–antiglobulin complex at the control zone also results in a pink line. The test is

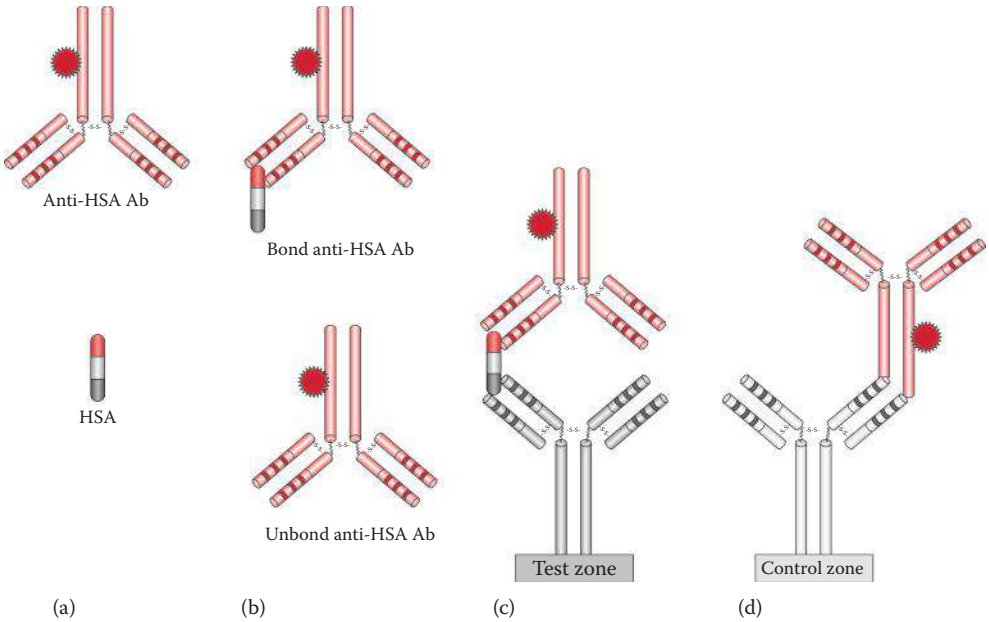

Figure 15.10 Immunochromatographic assay for the identification of HSA in saliva. (a) Sample containing amylases is loaded in a sample well. (b) Antigen binds to a labeled anti-HSA Ab to form a labeled Ab–HSA complex. (c) At the test zone, the labeled complex binds to an immobilized antihuman HSA Ab to form a labeled Ab–HSA–Ab sandwich. (d) At the control zone, the labeled anti-HSA Ab binds to an immobilized antiglobulin and is captured at the control zone. Ab, an antibody; HSA, human salivary α-amylase. (© Richard C. Li.)

considered valid only if the line in the control zone is observed. The presence of HSA results in a pink line at both the test zone and the control zone, while the absence of HSA results in a pink line in the control zone only. A result can be read after 10 min.

The sensitivity of the RSID®-Saliva kit can be as low as 1 μL of saliva. Additionally, the assay is responsive to samples extracted from saliva stains on both smooth and porous surfaces. In terms of species specificity, the kit has no cross-reactivity with various animal species, including monkeys (tamarin and callimico). As for bodily fluid specificity, it has also been shown that it is not responsive to human blood, semen, or urine. The high-dose hook effect, which creates an artifact that may cause false-negative results as described previously (Chapter 11), is not observed when up to 50 μL of saliva is tested. This method is rapid, specific, and sensitive and can be used in both laboratory and field analysis.

15.2.2.1.2 Enzyme-Linked Immunosorbent Assay (ELISA)

This method can be used to detect and to quantify a sample with the use of an anti-HSA antibody. The most common configuration in forensic serology is the antibody–antigen–antibody sandwich (Figure 15.11). ELISA utilizes reporting enzymes to produce colorimetric or fluorometric signals. The intensity of the signal can be detected spectrophotometrically and is proportional to the amount of bound antigen. The amount of HSA can be quantified by comparison with a standard of known concentration. This method is specific and highly sensitive in detecting HSA, but it is time-consuming. Chapter 11 discusses the ELISA principle in further detail.

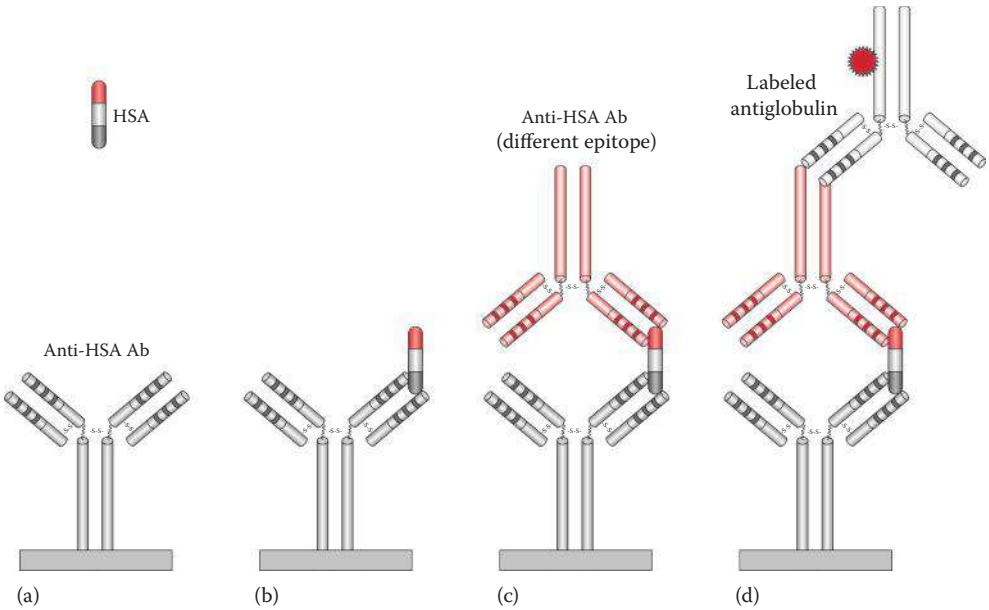

Figure 15.11 ELISA for identification of HSA in saliva. (a) Sample containing antigen is applied to polystyrene tubes where anti-HSA Ab is immobilized. (b) Antigen binds to immobilized Ab to form HSA–Ab complex. (c) Second anti-HSA Ab, specific for a different epitope of HSA, is added to form Ab–HSA–Ab sandwich. (d) Labeled antiglobulin then binds to the sandwich. The bound antiglobulin can be detected by various reporting schemes. Ab, an antibody; HSA, human salivary α-amylase. (© Richard C. Li.)

15.2.2.2 RNA-Based Assays

RNA-based assays (Chapter 11) have been developed recently for the identification of saliva. They are based on the expression of certain genes in certain cell or tissue types. Thus, the techniques used in the identification of saliva are based on the detection of specific types of mRNA expressed exclusively in certain cells in the oral cavity. These assays utilize reverse transcriptase polymerase chain reaction (RT-PCR; see Chapter 7) methods to detect gene expression levels of mRNAs for saliva identification. Table 15.1 summarizes the tissue-specific genes utilized for saliva identification. Compared to conventional assays used for saliva identification, RNA-based assays present higher specificity and are amenable to automation. However, one limitation is that the RNA is unstable because of degradation by endogenous ribonucleases.

Table 15.1 Application of RT-PCR Assay for Saliva Identification

Gene Symbol	Gene Product	Description	Further Reading
HTN3	Histatin 3	Histidine-rich protein involved in nonimmune host defense in oral cavity	Sabatini et al. (1993)
STATH	Statherin	Inhibitor of precipitation of calcium phosphate salts in oral cavity	Sabatini et al. (1990)

Source: Adapted from Juusola, J. and Ballantyne, J., *Forensic Sci Int*, 152, 1–12, 2005.

16

Identification of Vaginal Secretions and Menstrual Blood

The identification of vaginal secretions and menstrual blood is important for the investigation of sexual assault cases. Such identification can help to corroborate allegations of sexual assault. For example, in a sexual assault investigation, a stain was observed after examining the suspect's clothing. Subsequently, forensic DNA analysis revealed that the DNA of the stain originated from the victim, establishing a link between the victim and the suspect. However, the defendant may assert that the victim's DNA originated from a sweat stain as a result of casual contact and deny any criminal act. If vaginal secretions were found in the stain, the evidence has probative value to corroborate an allegation of a sexual act. Additionally, postcoital drainage stains on clothing or bedding are often recovered at crime scenes. These stains usually consist of a mixture of semen and vaginal secretions. In such cases, the presence of vaginal secretions in these stains indicates the occurrence of sexual intercourse. Sometimes, vaginal secretions can be transferred onto a perpetrator during a sexual assault. For example, the presence of vaginal secretions on a suspect's genital area can indicate the occurrence of sexual intercourse. Furthermore, the presence of vaginal secretions on an object can corroborate an allegation of vaginal rape with a foreign object.

If a sexual assault victim is in menses when an assault occurred, blood evidence, such as the victim's bloodstains located on the suspect's clothing, may be recovered at the scene. The defense may argue that the bloodstains resulted from an injury and deny that any sexual act occurred. In this case, the identification of menstrual blood would corroborate an alleged rape. Therefore, it is necessary to distinguish between peripheral and menstrual blood in investigations of sexual assault where blood evidence is found at the scene.

16.1 Identification of Vaginal Stratified Squamous Epithelial Cells

A normal human vagina is covered by the *squamous mucosa*, which is composed of stratified squamous epithelial tissue (Figure 16.1a). Lying under the squamous mucosa is the *submucosa*, which contains an abundance of connective tissue and capillaries. Below the submucosa is the *muscularis*, which is made up of smooth muscle.

The squamous mucosa consists of multiple layers of cells (Figure 16.1b). At the *basal layer* of the squamous mucosa, basal cells are anchored to the basement membrane that separates the squamous mucosa from the submucosa. The basal cells are small in size with relatively large nuclei and are highly proliferative. As the cells migrate up from the basal layer to the *parabasal layer*, the cells undergo differentiation. At the *intermediate layer*, the cells are flattened and their nuclei are compressed. As the epithelial cells reach the apical layer, the *superficial layer*, the cells are fully differentiated with small and dense nuclei.

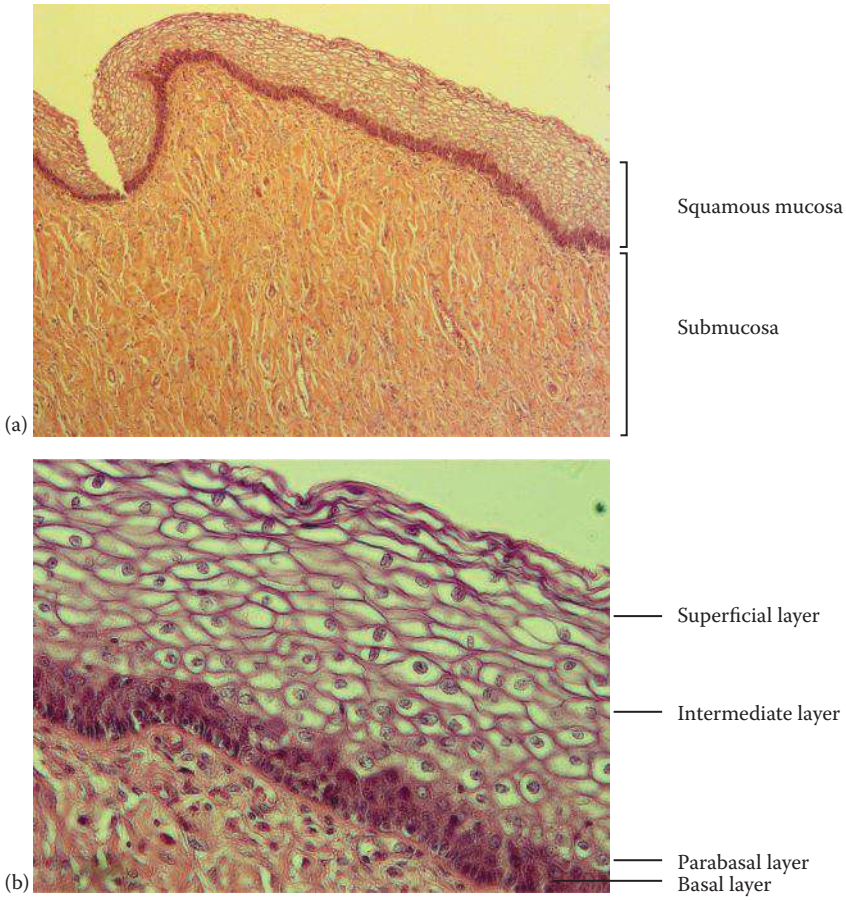

Figure 16.1 Histological structure of human vaginal tissues. (a) The squamous mucosa and the submucosa. (b) Multiple layers of cells in the squamous mucosa. (© Richard C. Li.)

Vaginal epithelial cells do not accumulate keratin, which is different from skin cells (Chapter 4). The intermediate and superficial layer cells contain abundant glycogen in their cytoplasm. The presence of glycogen in these cells is an indication of normal development and the differentiation of the vaginal epithelial cells. The apical surface of the vaginal squamous mucosa is usually covered by mucus that is secreted from glands that are located deep in the epithelium. The cells of the apical layers are eventually sloughed and are continuously replaced by the cells of deeper layers. When the lining of the vagina is swabbed during the collection of evidence, glycogenated cells with small numbers of parabasal cells are usually recovered (Figure 16.2). In addition to the vagina, glycogenated squamous epithelial cells are found in the linings of the oral cavity, pharynx, esophagus, anus, and the apex of the urethra. The differentiation of the vaginal epithelial cells requires estrogen. However, in premenarche and postmenopausal women, estrogen levels are very low; thus, the vaginal epithelial cells only differentiate to the parabasal cells. Primarily, parabasal cells are found in specimens from these individuals.

16.1.1 Lugol's Iodine Staining and Periodic Acid–Schiff Method

Lugol's iodine solution, named after the French physician Jean Lugol, is originally used as an antiseptic that is applied to skin or tissue to prevent infection. In forensic applications, it is utilized for the identification of glycogenated vaginal epithelial cells. The technique is based on the

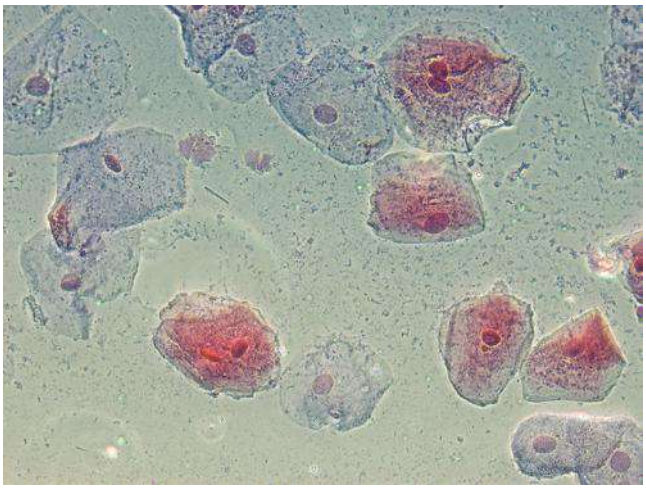

Figure 16.2 Vaginal stratified squamous epithelial cells. (© Richard C. Li.)

principle that iodine reacts with intracellular glycogen to exhibit a color. Glycogen is the principal storage form of glucose in animal and human cells and is found in the granules in the cytoplasm of the cells of many tissues. In addition to squamous epithelial cells, glycogen is also found in hepatocytes, which have the highest glycogen content, as well as muscle cells. Glycogen is a polysaccharide composed of D-glucose units. Similar to the amylopectin in plant cells (Chapter 15), glycogen is a branched polysaccharide consisting of linear and branched chains (Figure 16.3). The D-glucose residues of the linear chain are connected by α1→4 glycosidic bonds, while the branching points are connected by α1→6 glycosidic bonds. However, glycogen is more branched than

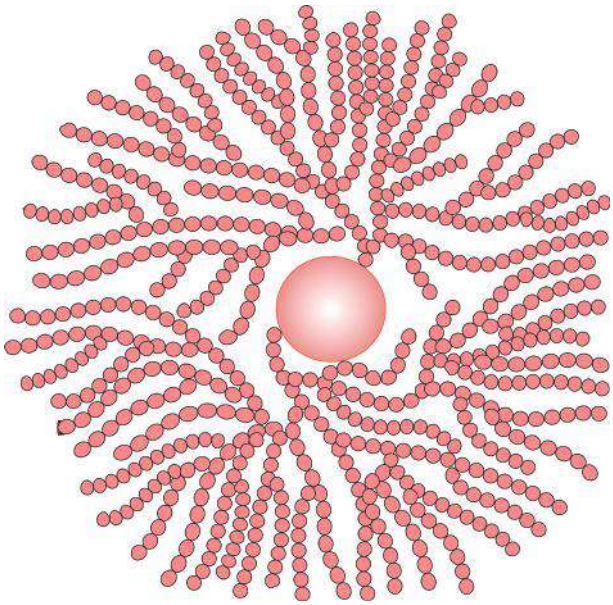

Figure 16.3 Structure of glycogen. Glycogen is a highly branched polysaccharide composed of glucose. Glycogen has a similar structure to amylopectin in starch (see Chapter 15). However, glycogen is more branched than starch. Additionally, glycogen contains a protein known as glycogenin at the center of its structure. (© Richard C. Li.)

amylopectin. Branch points occur approximately every 10 glucose residues in glycogen while they occur every 25 glucose residues in amylopectin. Lugol's stock solution is an aqueous solution of 5% iodine (I_2) and 10% potassium iodide (KI). Potassium iodide allows the iodine to be soluble in water through the formation of the triiodide ion. For staining the vaginal epithelial cells, 5% of the stock solution is usually used as the working solution. Iodine atoms fit into the helices of glycogen to form a dark brown glycogen–iodine complex (Figures 16.4 and 16.5). Vaginal epithelial cells can also be stained using the periodic acid–Schiff method. The cytoplasm of the vaginal epithelial cells is stained magenta, and the nucleus is stained purple.

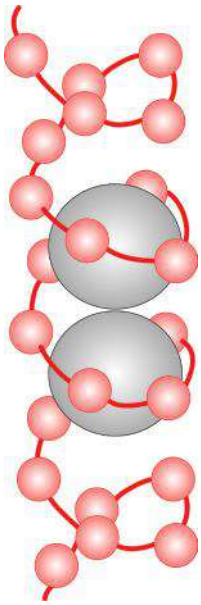

Figure 16.4 Diagram of a glycogen–iodine complex. In the complex, the glycogen chain forms a helix structure with six monosaccharide residues (red) per turn. Iodine molecules (gray) fit in the helix to form the glycogen–iodine complex, exhibiting a color. (© Richard C. Li.)

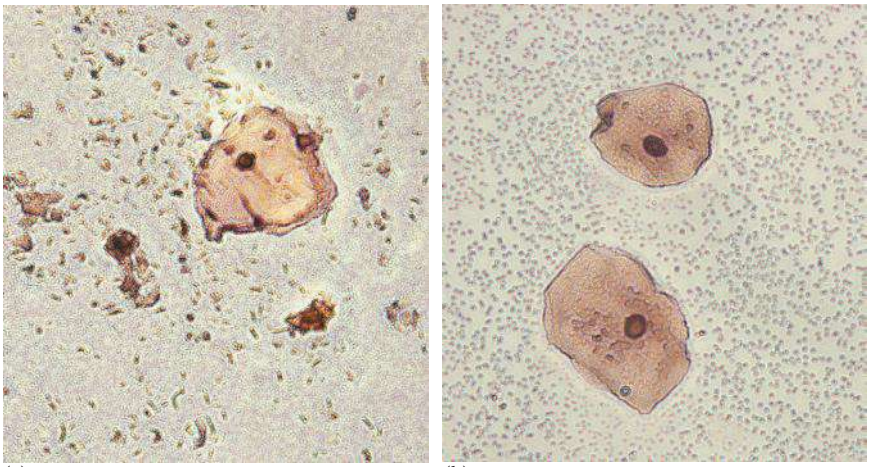

(a) (b)

Figure 16.5 Epithelial cells stained with Lugol's iodine solution. (a) Vaginal and (b) buccal epithelial cells. (© Richard C. Li.)

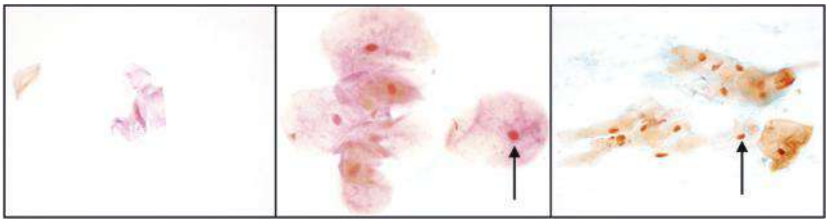

Figure 16.6 Staining of epithelial cells with Dane's method. The results obtained using methanol fixation followed by Dane's staining on skin (a), buccal (b), and vaginal epithelial (c) cells are shown. The arrows indicate red nuclei in buccal and orange nuclei in vaginal cells. (From French, C.E., et al., *Forensic Sci Int*, 178, 1–6, 2008. With permission.)

16.1.2 Dane's Staining Method

Evidence containing skin, buccal, and vaginal epithelial cells is often recovered in forensic investigations, particularly in sexual assault cases. Thus, it is necessary to distinguish between these cells. Skin, buccal, and vaginal epithelial cells belong to stratified squamous epithelium. Differentiated skin epithelial cells are keratinized and are classified as keratinizing squamous epithelial cells, while buccal and vaginal epithelial cells are nonkeratinizing squamous epithelial cells. Additionally, the skin epithelial cells lose nuclei and other cellular organelles during differentiation. In contrast, buccal and vaginal cells contain nuclei. Based on their morphology, it is possible to distinguish skin epithelial cells from buccal and vaginal epithelial cells. However, buccal and vaginal cells are morphologically indistinguishable from each other. Although Lugol's iodine solution and periodic acid–Schiff can stain vaginal epithelial cells, these stains cannot distinguish vaginal from other glycogenated epithelial cells in the oral mucosa and the urinary tract. Recently, Dane's staining method has been developed to distinguish all three types of cells. Skin cells are stained red and orange; buccal cells are stained predominantly orange-pink with red nuclei; and vaginal cells are stained bright orange with orange nuclei (Figure 16.6).

16.2 Identification of Vaginal Acid Phosphatase

Acid phosphatases are a group of enzymes that are capable of hydrolyzing a variety of small organic phosphomonoesters under acidic conditions. To date, at least five different acid phosphatase isoenzymes have been identified in human tissues: erythroid acid phosphatase (encoded by the *ACP1* gene), lysosomal acid phosphatase (encoded by the *ACP2* gene), prostate acid phosphatase (encoded by the *ACPP* gene, also known as *ACP3*), macrophage acid phosphatase (encoded by the *ACP5* gene), and testicular acid phosphatase (encoded by the *ACPT* gene). Human prostatic acid phosphatase is found in large quantities in seminal fluid and is used as a biomarker for semen identification (Chapter 14). The prostatic acid phosphatase is a homodimer containing two identical subunits with a molecular weight of 50 kDa. Small amounts of acid phosphatase can be detected in vaginal fluid, which is produced in normal cervical epithelial cells. However, the molecular characteristics of vaginal acid phosphatase are still not known. Historically, vaginal acid phosphatase has been used as a biomarker for the identification of vaginal secretions using acid phosphatase catalytic assays (Chapter 14). In sexual assault investigations, it is important to distinguish vaginal acid phosphatase from prostate acid phosphatase originating from semen exposure. These two enzymes have identical molecular weights, enzymatic specificities, and responses to the same inhibitors. Nevertheless, vaginal and prostate acid phosphatases can be distinguished using agarose electrophoresis. Based on their electrophoretic mobility, bands of vaginal and prostate acid phosphatases can be separated. The prostate acid phosphatase

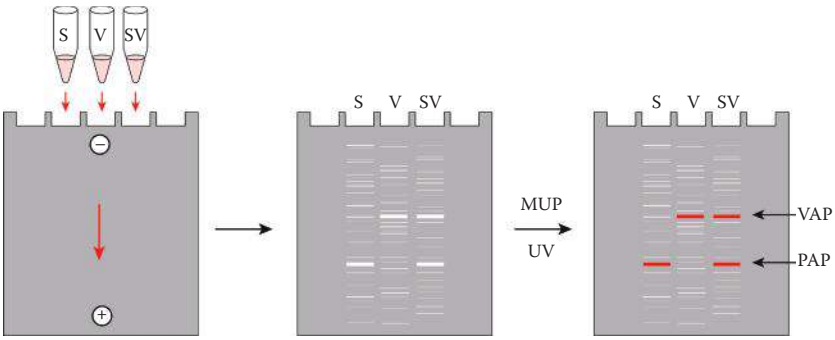

Figure 16.7 Distinguishing prostate and vaginal acid phosphatases using polyacrylamide gel electrophoresis. Bands of acid phosphatases are detected using 4-methylumbelliferyl phosphate (MUP). MUP is a substrate of acid phosphatases, resulting in a product that fluoresces in UV illumination (see Chapter 14). The prostate acid phosphatase has higher electrophoretic mobility toward an anode (a positively charged electrode) than vaginal acid phosphatase. S, semen; V, vaginal secretions; SV, mixture of semen and vaginal secretions; PAP, prostate acid phosphatase; VAP, vaginal acid phosphatase; MUP, 4-methylumbelliferyl phosphate; UV, ultraviolet illumination. (© Richard C. Li.)

has higher electrophoretic mobility toward an anode (a positively charged electrode) than the vaginal acid phosphatase (Figure 16.7). Thus, the presence of vaginal acid phosphatase can be determined.

16.3 Identification of Vaginal Bacteria

Lactobacillus can be found in the respiratory, the gastrointestinal, and the urogenital tract of healthy humans and animals. *Lactobacillus* taxa are the predominant bacteria in the vagina of women of reproductive age, and they play an important role in protecting the host against invasive pathogenic organisms. *Lactobacillus* consists of rod-shaped, nonmotile, and non-spore-forming gram-positive bacteria (Figure 16.8). Since these bacteria survive on carbohydrates, *Lactobacillus* bacteria produce lactic acid. As a result, a low pH environment is established in

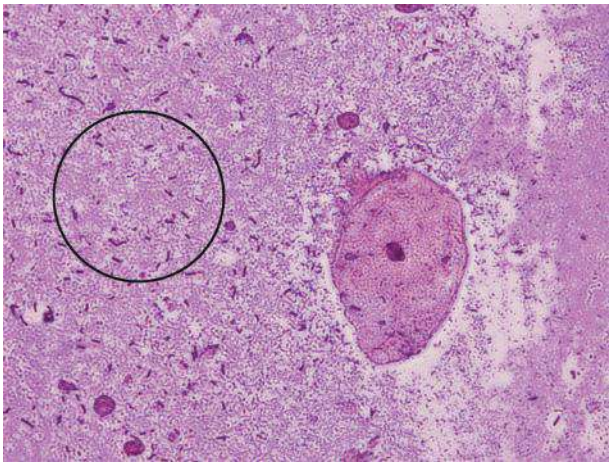

Figure 16.8 Vaginal *Lactobacillus* bacteria (circled). (© Richard C. Li.)

the lumen of the vagina, which restricts the growth of pathogenic organisms. Such vaginal lactic acid–producing bacteria were thought to be *Lactobacillus acidophilus*. In the 1980s, it was determined that *L. acidophilus* was not a single species, but actually a group of related species known as the *L. acidophilus* complex. The species of the complex can now be distinguished based on their DNA sequences. The most frequently occurring *Lactobacillus* species found in the vagina are *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii*. Among them, *L. iners* is the most common species of *Lactobacillus* in women. *Lactobacillus* can be identified based on the sequences of DNA markers such as the 16S rRNA gene and the intergenic spacer region between the 16S rRNA and 23S rRNA genes (Chapter 11). In forensic applications, *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii* can be detected in vaginal secretions. Thus, the identification of *Lactobacillus* taxa can potentially be utilized for the forensic identification of vaginal fluid. However, the presence of *Lactobacillus* taxa is not specific enough to vaginal fluid. Some studies show that *Lactobacillus* taxa are present in semen while others show that *Lactobacillus* can be found in female urine. The openings of the female urogenital system are in close proximity; thus, it is possible to have partially overlapping microbiota between the urine and vaginal secretions. Thus, the identification of multiple bacterial species, at least the four *Lactobacillus* species previously mentioned, should be carried out to distinguish these samples.

16.4 Outlook for Confirmatory Assays of Vaginal Secretions

The identification of the vaginal stratified squamous epithelial cells provides important probative evidence in forensic investigations. However, the existing methods described earlier can sometimes give false-negative or false-positive results. In some situations, these assays also cross-react with other types of bodily fluids. Thus, none of these assays is confirmatory. A useful identification method for vaginal secretions should be able to distinguish vaginal secretions from other bodily fluids and should be easy to perform. For example, nondestructive confirmatory identification methods such as fluorescence spectroscopy and Raman spectroscopy (Chapter 11) can potentially be useful for the identification of vaginal secretions. Recently, the analysis of tissue-specific gene expression has been utilized for the identification of vaginal secretions. Using the reverse transcription polymerase chain reaction (RT-PCR) technique (Chapter 7), the mRNAs of the tissue-specific genes of vaginal epithelial cells can be detected. For example, two commonly used markers for vaginal secretion identification are *MUC4* and *HBD1*. *MUC4* encodes a mucin protein that is a major component of vaginal mucus, and *HBD1*, the human β defensin 1, encodes a vaginal antimicrobial peptide. Both *MUC4* and *HBD1* are expressed in vaginal epithelial cells and are considered reliable markers of vaginal fluid. Additional mRNA and miRNA markers are described in Tables 16.1 and 16.2.

16.5 Menstruation

Menstruation is the periodic discharge of blood and the elimination of the degenerated lining of the endometrium from the uterus of nonpregnant women. From menarche to menopause, women may menstruate up to 400 times during their reproductive age. The uterus plays an important role in preparing the uterine endometrium for the possible implantation of a developing embryo. The linings of the uterus are composed of the myometrium and the endometrium. The *myometrium* consists of the muscle fibers of the uterus. The *endometrium* consists of the simple columnar epithelium and the stroma (Figure 16.9). The simple columnar epithelium is formed by single-layered elongated cells located at the apical surface of the endometrium. The stroma consists of connective tissues as well as spiral arteries. Spiral arteries are small arteries that ascend through the endometrium and form a coil-like structure, which supplies blood to the endometrium.

Table 16.1 Representative Markers of mRNA-Based Assays for Vaginal Secretions and Menstrual Blood Identification

Biological Fluid	Gene Symbol	Description
Vaginal secretions	*CYP2B7P1*	Cytochrome P450, family 2, subfamily B, polypeptide 7, pseudogene 1
	DKK4	Dickkopf homolog 4
	FUT6	Fucosyltransferase 6
	HBD1	β Defensin 1
	IL19	Interleukin 19
	MUC4	Mucin 4
	MYOZ1	Myozenin 1
	SFTA2	Surfactant associated 2
Menstrual blood	*MMP7*	Matrix metalloproteinase 7
	MMP11	Matrix metalloproteinase 11

Source: Adapted from Bauer, M. and Patzelt, D., *J Forensic Sci*, 47, 1278–1282, 2002; Hanson, E.K. and Ballantyne, J., *Sci Justice*, 53, 14–22, 2013; Juusola, J. and Ballantyne, J., *Forensic Sci Int*, 152, 1–12, 2005; Nussbaumer, C., Gharehbaghi-Schnell, E., and Korschineck, I., *Forensic Sci Int*, 157, 181–186, 2006.

Table 16.2 miRNA Markers for Vaginal Secretions and Menstrual Blood Identification

Bodily Fluid	miRNA Marker	Sequence
Vaginal secretions	miR124*a*	UAAGGCACGCGGUGAAUGCC
	miR372	AAAGUGCUGCGACAUUUGAGCGU
	miR617	AGACUUCCCAUUUGAAGGUGGC
	miR891a	UGCAACGAACCUGAGCCACUGA
Menstrual blood	miR214	UGCCUGUCUACACUUGCUGUGC
	miR412	ACUUCACCUGGUCCACUAGCCGU
	miR451	AAACCGUUACCAUUACUGAGUU

Source: Adapted from Hanson, E.K., Lubenow, H., and Ballantyne, J., *Anal Biochem*, 387, 303–314, 2009; Wang, Z., et al., *Forensic Sci Int Genet*, 7, 116–123, 2013.

16.5.1 Uterine Cycle

The endometrium can be divided into two zones: the functionalis and the basalis (Figure 16.10). The *functionalis* is the luminal part of the endometrium. It is the zone of cyclic changes in the endometrium and is shed during menstruation. The *basalis* is the basal part of the endometrium and is not shed during menstruation. This zone produces cells to regenerate the functionalis during the next cycle.

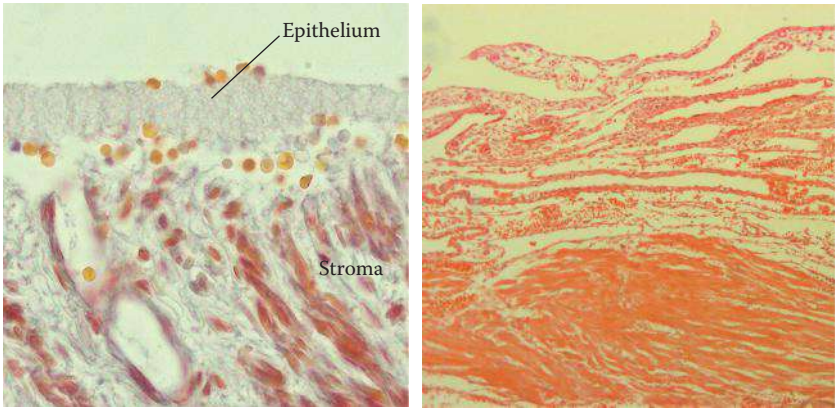

Figure 16.9 Uterus endometrium. (a) The proliferative phase and (b) the menstrual phase. (© Richard C. Li.)

During the uterine cycle, repetitive physiological changes occur in the functionalis (Figure 16.10). The cycle is divided into three phases: the menstrual, proliferative, and secretory phases. The first day of menstrual bleeding is considered the onset of the menstrual phase. During the menstrual phase, the functionalis degenerates and is sloughed off from the uterine wall and bleeding occurs, known as menses. During the proliferative phase, the functionalis begins regeneration in which the spiral arteries are proliferated. During the secretory phase, the spiral arteries are further developed and coiled. In the absence of pregnancy, a decrease in the progesterone level leads to the constriction of the spiral arteries. As a result, the functionalis becomes ischemic (insufficient blood flow), leading to hypoxia (low levels of oxygen in tissue). In addition, the activation of an enzymatic degradation process causes the destruction and

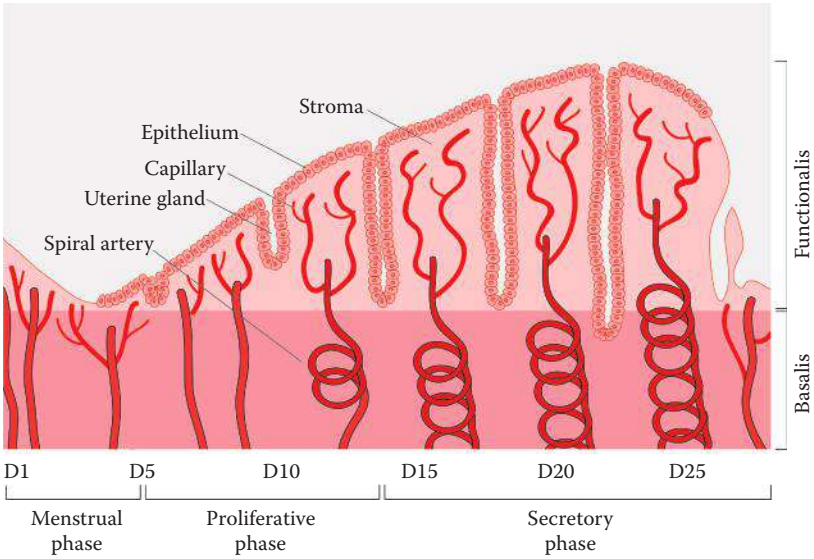

Figure 16.10 Changes in the functionalis of the uterine mucosa during a uterine cycle. D, day. (© Richard C. Li.)

shedding of the functionalis during menses. Menstrual fluid contains blood, the functionalis layer tissue, and mucus.

16.5.2 Uterine Endometrial Hemostasis

The cessation of menstrual bleeding is achieved by endometrial *hemostasis* that is initiated when injury occurs due to the shedding of the endometrium (Figure 16.11). Hemostasis begins with platelet activation and aggregation to form platelet plugs at the site of injury. Additionally, the blood coagulation cascade is activated to produce *thrombin*. Thrombin, a serine protease, converts soluble fibrinogen into fibrin. Fibrin, a protein involved in blood clotting, aggregates with the platelet plugs and leads to the cessation of bleeding by forming blood clots, known as *thrombi*. Under normal physiological conditions, uterine endometrial hemostasis is a balanced process between blood coagulation and clot dissolution to control blood loss and to prevent clot accumulation within the uterus. As a result, the balance of these two processes allows the removal of tissue fragments from the uterus cavity in order to reduce the risks of infection. Blood clots are prevented from accumulating during menstruation by forming low amounts of platelet plugs and synthesizing coagulation factor inhibitors that inhibit blood coagulation. Additionally, *fibrinolysis* is activated, during which thrombus is broken down by a protease known as *plasmin*. Plasmin cleaves fibrin, generating soluble degradation products. As a result, fibrinolysis can inhibit blood clot formation.

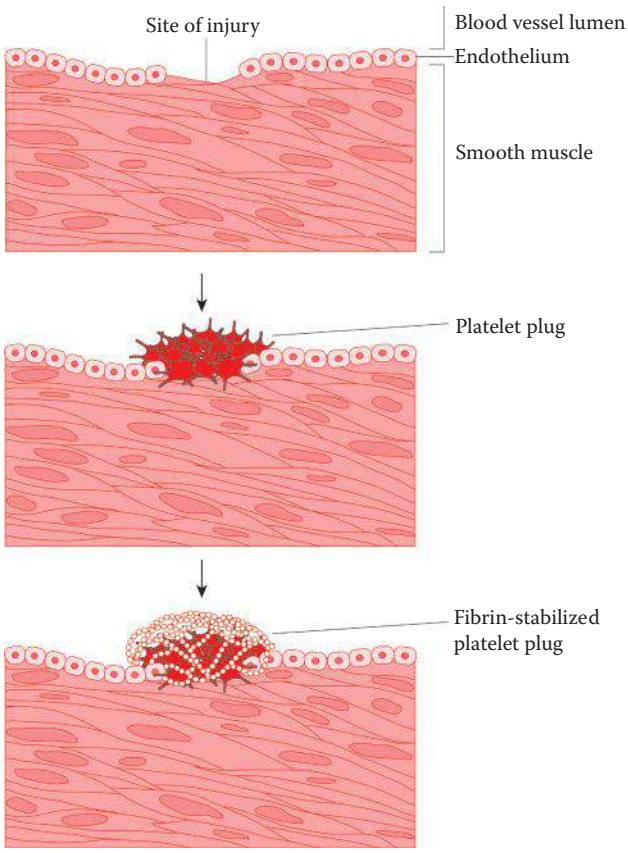

Figure 16.11 Diagram of hemostasis. Hemostasis occurs at the site of injury of a blood vessel. A platelet plug is formed as a result of the aggregation of platelets. The plug is further stabilized by the formation of a fibrin clot over the platelet plug. (© Richard C. Li.)

16.6 D-dimer Assay

During fibrinolysis, cross-linked fibrin is cleaved by plasmin, producing a degradation product known as *D-dimer* (Figure 16.12). Assays for D-dimer fragments have been utilized in the forensic identification of menstrual blood. A number of different formats of D-dimer assays can be used. In an enzyme-linked immunosorbent assay (ELISA; Chapter 11), antibodies bind to the D-dimer antigens on the solid phase. The D-dimer–antibody complex is subsequently analyzed using an antibody-based detection system. This method is highly sensitive, but is time-consuming. Latex agglutination assays are based on the interaction of antibodies and D-dimers that are located on carriers to form aggregates during the agglutination process (Chapter 11). However, the magnitude of the agglutination response is manually read and conclusions are based on subjective judgments. Immunochromatographic assays (Chapter 11) utilize monoclonal antibodies specific to D-dimers, which have been developed recently (Figures 16.13 and 16.14). This immunochromatographic assay is very specific, sensitive, and rapid, and can be completed within minutes. The immunochromatographic devices are portable, and thus can potentially be used at crime scenes. The D-dimer assays can positively identify menstrual blood samples. Although peripheral blood contains low levels of D-dimer, these assays do not show positive reactions with peripheral blood. Thus, menstrual blood can be distinguished from peripheral blood using D-dimer assays. However, postmortem blood also contains these D-dimers, which are detected by these assays. Although the detection of postmortem blood would complicate the interpretation of results, postmortem blood is not often encountered in sexual assault cases.

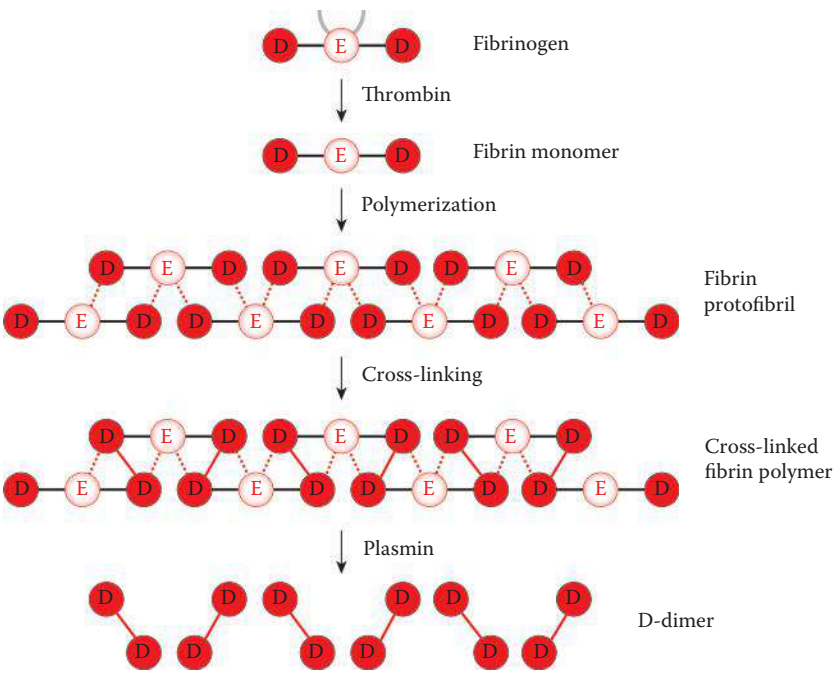

Figure 16.12 The formation and the degradation of a fibrin polymer. The activation of coagulation ultimately generates thrombin, which catalyzes the conversion of fibrinogen to fibrin by cleaving the fibrinopeptides (gray). A fibrin monomer contains an E domain and two D domains. The fibrin monomers are held together by noncovalent bonds (dotted red) between the D domains and E domain to form a fibrin polymer. Fibrin polymers are then covalently linked (solid red) to form a cross-linked fibrin polymer, which plays a role in forming clots. During fibrinolysis, plasmin cleaves the cross-linked fibrin at multiple sites giving rise to fibrin degradation products including D-dimer. (© Richard C. Li.)

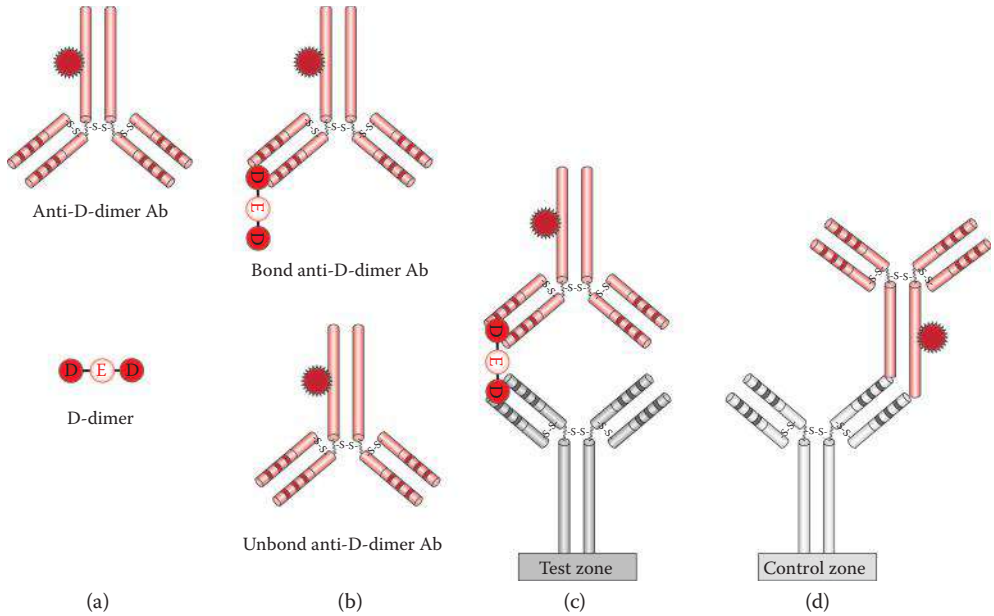

Figure 16.13 Immunochromatographic assays for the identification of D-dimer in menstrual blood. (a) In a sample well, D-dimer antigen in a menstrual blood sample is mixed with labeled anti-D-dimer Ab. (b) The D-dimer binds to the labeled anti-D-dimer Ab to form a labeled Ab–antigen complex. (c) At the test zone, the labeled complex binds to an immobilized anti-D-dimer Ab to form a labeled Ab–antigen–Ab sandwich. (d) At the control zone, the labeled anti-D-dimer Ab binds to an immobilized antiglobulin and is captured at the control zone. Ab, antibody. (© Richard C. Li.)

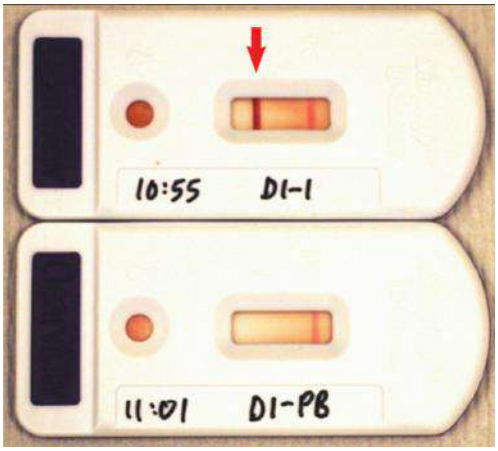

Figure 16.14 Detecting D-dimer using immunochromatographic assays. The results obtained using the Clearview Simplify D-dimer devices (Alere, Cheshire, UK) on menstrual (top) and peripheral blood (bottom) are shown. A positive result (arrow) indicates the presence of D-dimer. (From Baker, D.J., Grimes, E.A., and Hopwood, A.J., *Forensic Sci Int*, 212, 210–214, 2011. With permission.)

16.7 Lactate Dehydrogenase Assay

Lactate dehydrogenase (LDH) is an enzyme that plays an important role in glycolysis. In humans, LDH catalyzes the reversible reduction of pyruvate into lactate when the amount of oxygen is limited. LDH is a tetrameric enzyme consisting of three different types of subunits. The A subunit, also known as the M subunit, is encoded by the *LDHA* gene and is primarily expressed in skeletal muscle. The B subunit, also known as the H subunit, is encoded by the *LDHB* gene and is primarily expressed in cardiac muscle. The C subunit, encoded by the *LDHC* gene, is expressed restrictively in the testes. LDHs are found in various human tissues. Five isoenzymes can be found in blood (Figure 16.15). LDH1 consists of four identical B subunits; LDH2 consists of one A and three B subunits; LDH3 consists of two A and two B subunits; LDH4 consists of three A and one B subunits; and LDH5 consists of four identical A subunits. The five isoenzymes can be separated using electrophoresis (Figure 16.16) and detected using a colorimetric assay (Figures 16.17 and 16.18). According to their electrophoretic mobility, five bands can be identified. LDH1 has the highest electrophoretic mobility (toward an anode that is a positively charged electrode) and LDH5 has the lowest electrophoretic mobility. In peripheral blood, LDH1, 2, and 3 are the predominant forms of the isoenzymes and LDH4 and 5 are the minor forms of the isoenzymes. In contrast, LDH4 and 5 are consistently the predominant isoenzymes in menstrual blood, while the amounts of LDH1, LDH2, and LDH3 vary. Thus, menstrual blood can be distinguished from peripheral blood. LDH was historically used as a marker for the forensic identification of menstrual blood.

Figure 16.15 Isozymes of lactate dehydrogenase. Human lactate dehydrogenases are composed of four subunits. Five types of isozymes with their subunits are shown. A, LDH A subunit; B, LDH B subunit.

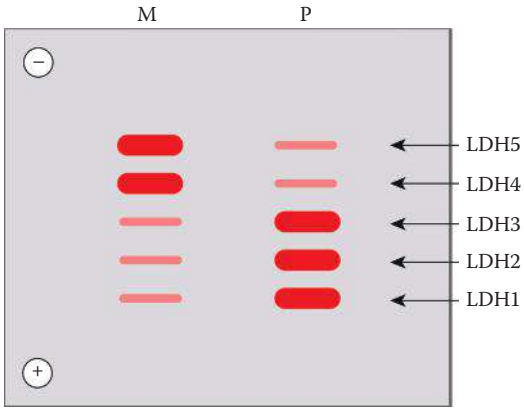

Figure 16.16 Electrophoretic separation of lactate dehydrogenase (LDH) isozymes. Extracted blood samples are loaded onto a cellulose acetate membrane. Electrophoresis is then carried out. Separated bands of LDH isozymes are detected using a colorimetric assay (see Figure 16.15). M, menstrual blood; P, peripheral blood. (© Richard C. Li.)

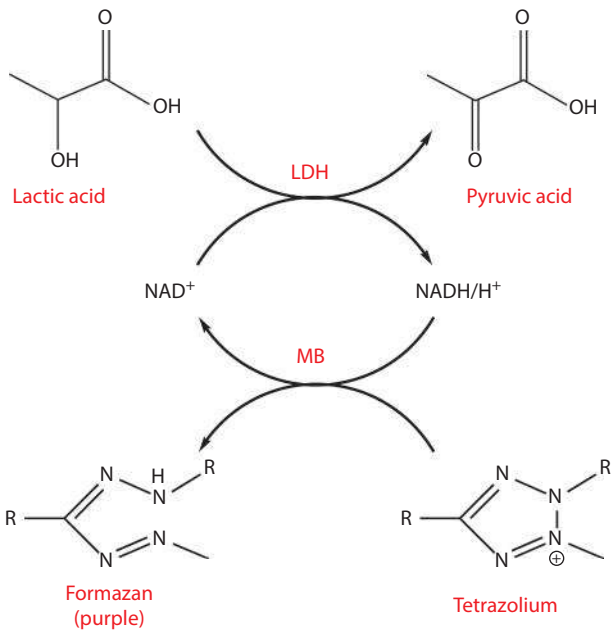

Figure 16.17 Detecting LDH activity using a colorimetric assay forming a colored dye. At the start of a cascade reaction, the LDH catalyzes the conversion of lactic acid to pyruvic acid. Nicotinamide adenine dinucleotide (NAD^+) is then reduced to $NADH/H^+$. Subsequently, Meldola's Blue (MB; 8-dimethylamino-2-benzophenoxazine), an electron carrier, transfers H/H^+ from $NADH/H^+$ to a tetrazolium salt. The pale yellow tetrazolium salt, nitroblue tetrazolium (NBT), is reduced to form a purple formazan dye.

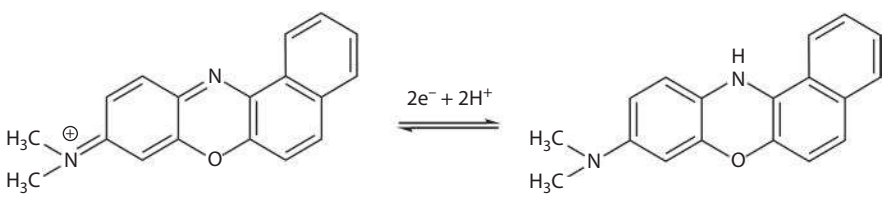

Figure 16.18 The reduction and oxidation of Meldola's Blue. Meldola's Blue is an electron carrier, which transfers hydrogen atoms (coupled to electron transfer).

16.8 RNA-Based Assays

Matrix metalloproteinase (MMP) genes are considered tissue-specific markers for human endometrium tissues. MMPs are zinc-dependent endopeptidases that degrade extracellular matrix components. Additionally, they cleave other proteins such as cytokines, chemokines, and growth factors. The *extracellular matrix* (ECM) is the extracellular space of tissue that is filled by macromolecules such as collagens, laminins, fibronectins, and proteoglycans. The ECM can be divided into two categories: the interstitial matrix and the basement membrane. The interstitial matrices are located in the intercellular spaces. The basement membranes are thin layers of macromolecule fibers that usually lie under the epithelium and the endothelium. Both the interstitial matrices and the basement membranes provide structural support to the cells. It is proposed that MMPs play an important role in the degradation of ECM, leading to the destruction of endometrium tissues during the uterine cycle. To date,

a total of 23 MMPs, divided into five subgroups, have been found in humans. The most commonly used markers for the forensic identification of menstrual blood are *MMP7* and *MMP11* (Table 16.1).

Many MMPs are expressed in human endometrium throughout the uterine cycle. The patterns of the MMP gene expressions are correlated with their functions in endometrium tissue breakdown during menstruation. *MMP7* is predominantly expressed in epithelial cells, while *MMP11* is expressed in the stromal cells of the endometrium. Both *MMP7* and *MMP11* mRNA expressions are elevated at the menstrual phase and remain at high levels during the proliferative phase. It is also known that MMPs' mRNA may be elevated in postpartum, wound healing, and metastatic cancer conditions, which may potentially lead to a false-positive identification of menstrual blood.

In menstrual blood samples, among all of the MMP genes tested, *MMP11* is the most sensitive and specific marker for distinguishing menstrual blood from peripheral blood. Using the RT-PCR technique (Chapter 7), *MMP11* mRNA can be detected in menstrual blood from the first to the eighth day of menstruation but it is absent in peripheral blood and vaginal secretions. Therefore, *MMP11* can be used as a marker for the identification of menstrual blood. Likewise, *MMP7* is also a useful marker for menstrual blood identification. However, the *MMP7* mRNA level in menstrual blood is less than that of *MMP11*. Additional markers such as miRNA markers for menstrual blood identification are included in Table 16.2.

Bibliography

Ablett, P.J., The identification of the precise conditions for seminal acid phosphatase (SAP) and vaginal acid phosphatase (VAP) separation by isoelectric focusing patterns. *J Forensic Sci Soc*, 1983, **23**(3): 255–256.

Adams, E.G. and Wraxall, B.G., Phosphatases in body fluids: The differentiation of semen and vaginal secretion. *Forensic Sci*, 1974, **3**: 57–62.

Akutsu, T., et al., Evaluation of latex agglutination tests for fibrin-fibrinogen degradation products in the forensic identification of menstrual blood. *Leg Med (Tokyo)*, 2012, **14**(1): 51–54.

Akutsu, T., et al., Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid. *Leg Med (Tokyo)*, 2012, **14**(3): 160–162.

Albrecht, K., et al., Immunohistochemical distribution of cyclic nucleotide phosphodiesterase (PDE) isoenzymes in the human vagina: A potential forensic value? *J Forensic Leg Med*, 2007, **14**(5): 270–274.

Alsawaf, K. and A.T. Tu, Isotachophoretic analysis of bloodstains: Differentiation of human, menstrual, bovine, and ovine bloods. *J Forensic Sci*, 1985, **30**(3): 922–930.

Baker, D.J., E.A. Grimes, and A.J. Hopwood, D-dimer assays for the identification of menstrual blood. *Forensic Sci Int*, 2011, **212**(1–3): 210–214.

Bauer, M. and D. Patzelt, Evaluation of mRNA markers for the identification of menstrual blood. *J Forensic Sci*, 2002, **47**(6): 1278–1282.

Bauer, M. and D. Patzelt, Identification of menstrual blood by real time RT-PCR: Technical improvements and the practical value of negative test results. *Forensic Sci Int*, 2008, **174**(1): 55–59.

Benschop, C.C., et al., Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? *Int J Legal Med*, 2012, **126**(2): 303–310.

Divall, G.B. and M. Ismail, Lactate dehydrogenase isozymes in vaginal swab extracts: A problem for the identification of menstrual blood. *Forensic Sci Int*, 1983, **21**(2): 139–147.

Ferri, G., et al., Successful identification of two years old menstrual bloodstain by using MMP-11 shorter amplicons. *J Forensic Sci*, 2004, **49**(6): 1387.

Fleming, R.I. and S. Harbison, The use of bacteria for the identification of vaginal secretions. *Forensic Sci Int Genet*, 2010, **4**(5): 311–315.

French, C.E., et al., A novel histological technique for distinguishing between epithelial cells in forensic casework. *Forensic Sci Int*, 2008, **178**(1): 1–6.

Fujita, Y., I. Tokunaga, and S.I. Kubo, Forensic identification of a vaginal fluid and saliva mixture through DNA analysis. *Acta Criminologiae et Medicinae Legalis Japonica*, 2003, **69**(2): 48–52.

17

Identification of Urine, Sweat, Fecal Matter, and Vomitus

17.1 Identification of Urine

The identification of the presence of urine is useful for the forensic investigation of an alleged sexual assault and harassment involving urination. More importantly, the identification of urine stains can aid in the investigation of homicides involving either ligature or manual strangulation. In these incidents, strangulation victims often involuntarily excrete urine prior to death. The locations of urine stains at the crime scene provide useful information to determine the site where the violence has occurred.

17.1.1 Urine Formation

In humans, the urinary excretory system eliminates soluble toxic wastes that are cellular metabolic by-products. The urinary system consists of the kidney, the ureter, the urinary bladder, and the urethra (Figure 17.1). The formation of urine takes place in the kidneys and, in particular, in the nephrons. The *nephron* is the basic structural and functional unit of the kidney (Figures 17.2 and 17.3). Each nephron is composed of a glomerulus, a Bowman's capsule, and a renal tubule. The glomerulus is formed by a network of capillaries and is surrounded by a Bowman's capsule. *Filtration* is the first step in urine formation. As blood flows through the glomeruli, much of its fluid, except cells and large molecules, is filtered through the capillaries into the Bowman's capsule. The glomerular filtrate, the preliminary form of urine, consists of water, salts (largely sodium and potassium ions), glucose, and waste products such as urea. The filtrate is then passed through the renal tubule where *reabsorption* occurs. During the reabsorption process, water, glucose, nutrients, and ions such as sodium are reabsorbed back into the blood. The last process of urine formation is *secretion*. Secretion occurs at the distal and the collecting tubules of the nephron where ions (such as hydrogen and potassium ions), ammonia, and certain metabolites are secreted from the blood into the lumen of the renal tubule to be eliminated in the urine. The urine is drained from the kidneys through the ureters and is stored in the bladder before it is finally excreted through the urethra.

Urine is an aqueous solution consisting largely of water. Urea is the most abundant waste product in urine, resulting from the elimination of ammonia that is produced from the metabolic process of amino acids. The average urea concentration in human urine is approximately 9 g/L. Other major components are creatinine, uric acid, and ions such as phosphate, sulfate,

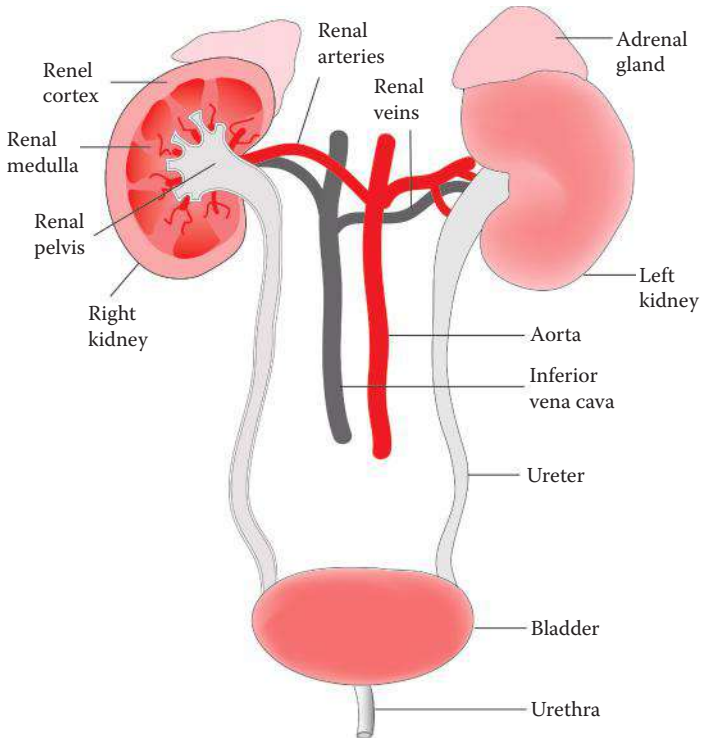

Figure 17.1 Human urinary system. A longitudinal view of the right kidney is shown. (© Richard C. Li.)

chloride, sodium, and potassium. Healthy individuals also excrete a small amount of protein. Urine has a characteristic yellow color. Subsequent to excretion from the body, urine acquires an odor that results from the ammonia that is released from the breakdown of urea by bacteria.

17.1.2 Presumptive Assays

In forensic identification, urine stains can be located by visual examination based on the characteristic yellow color of urine and the detection of the distinctive odor of urine stains. Under alternative light sources, urine stains emit a fluorescent light that facilitates the locating of urine stains in clothing and bedding. Chemical analysis can be carried out to detect the major inorganic anions in urine such as phosphate and sulfate as well as the major organic compounds in urine such as urea, creatinine, and uric acid. These assays are summarized in Table 17.1. However, these assays are not specific to urine. Other bodily fluids, such as sweat, also contain these chemical components.

17.1.2.1 The Identification of Urea

The *para-dimethylaminocinnamaldehyde* (DMAC) assay is simple and rapid and is the most commonly used presumptive assay for the forensic identification of urine stains (Figure 17.4). The DMAC assay can be performed using two different methods: the colorimetric and the fluorometric methods (Figure 17.5).

In the colorimetric method, a portion of a stain (~1 cm^2) is cut and is extracted with 1 mL of distilled water. The extraction is transferred onto a piece of filter paper and is allowed to dry. One drop of 0.1% DMAC solution is then added to the filter paper. DMAC reacts specifically with urea, if present, producing a pink-colored (or magenta-colored) product. DMAC does not react with creatinine, ammonia, or uric acid. The appearance of a pink color within 30 min after applying the DMAC reagent is considered a positive reaction. No color change within 30 min is

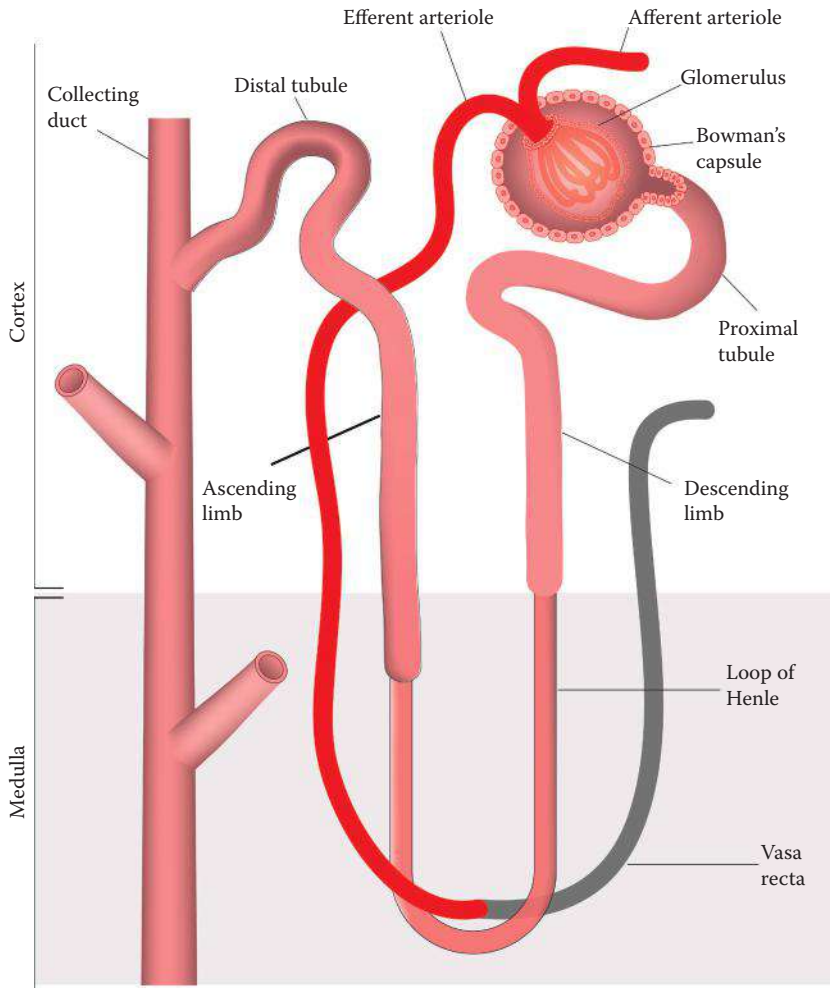

Figure 17.2 Diagram of the human nephron. (© Richard C. Li.)

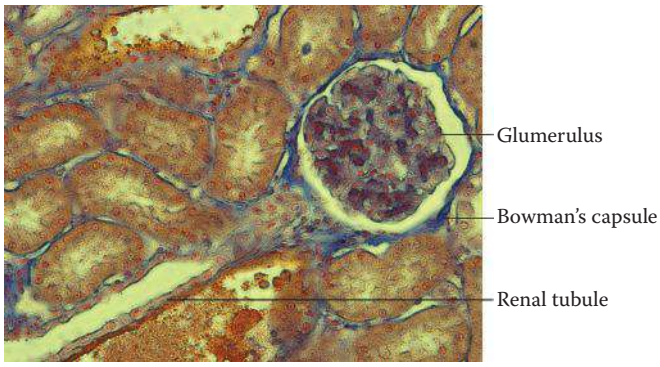

Figure 17.3 Section view of the human nephron. (© Richard C. Li.)

Table 17.1 Assays for Identification of Urine and Sweat

Biological Fluid	Assays	Identification	Assay	Method	Further Reading
Urine	Presumptive assays	Locating stains	Visual examination	Alternate light source	Wawryk and Odell (2005)
		Urea	DMAC assay	Colorimetric method	Ong et al. (2012)
				Fluorometric method	Farrugia et al. (2012)
			Microscopic crystal assay	Microscopic examination of urea crystals and urea nitrate crystals converted from urea	Nickolls (1956)
				Microscopic examination of crystals using xanthydrol	Biles and Ziobro (1998)
			Urease assay	The enzyme catalyzes the breakdown of urea and releases ammonia and carbon dioxide. The ammonia is detected using an acid-base indicator, bromthymol blue, which exhibits a blue color	Bedrosian et al. (1984)
				The ammonia is detected using Mn and Ag nitrates, which exhibit a black color	Gaensslen (1983)
			Chromatography	Gas chromatograph–mass spectrometer, thin-layer chromatography	Lum (1991)
		Creatinine	Colorimetric assay	Jaffe's test, using picric acid to convert creatinine to a red creatinine picrate compound	Greenwald (1930)
				Using potassium ferricyanide to produce a blue nitroprusside compound	O'Leary et al. (1992)
			Chromatography	Gas chromatograph–mass spectrometer	Lum (1991)

		Uric acid	Uricase assay	Measuring the disappearance of the absorption of uric acid at 293 nm after treatment with uricase that catalyzes the oxidation of uric acid to 5-hydroxyisourate	Kageyama (1971)
	Confirmatory assays	Tamm–Horsfall glycoprotein	Immunological assays	RIA, crossover electrophoresis, Ouchterlony, ELISA, and immunochromatographic methods	Tsutsumi et al. (1988)
		17-ketosteroids	Mass spectrometer	Liquid chromatograph–mass spectrometer	Nakazono et al. (2002)
Sweat	Presumptive assays	Lactate, lactic acid, urea, and single amino acids	Various	SEM coupled with EDX and Raman microspectroscopy	Sikirzhytski et al. (2011)
	Confirmatory assays	Dermcidin	Immunological assay	ELISA	Sagawa et al. (2003)
			RNA-based assay	RT-PCR	Sakurada et al. (2010)

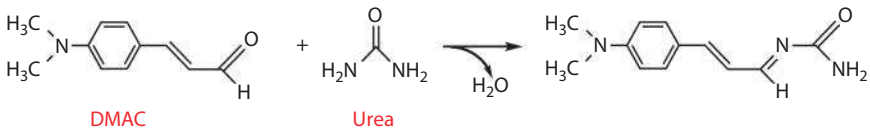

Figure 17.4 Chemical reaction of DMAC assay.

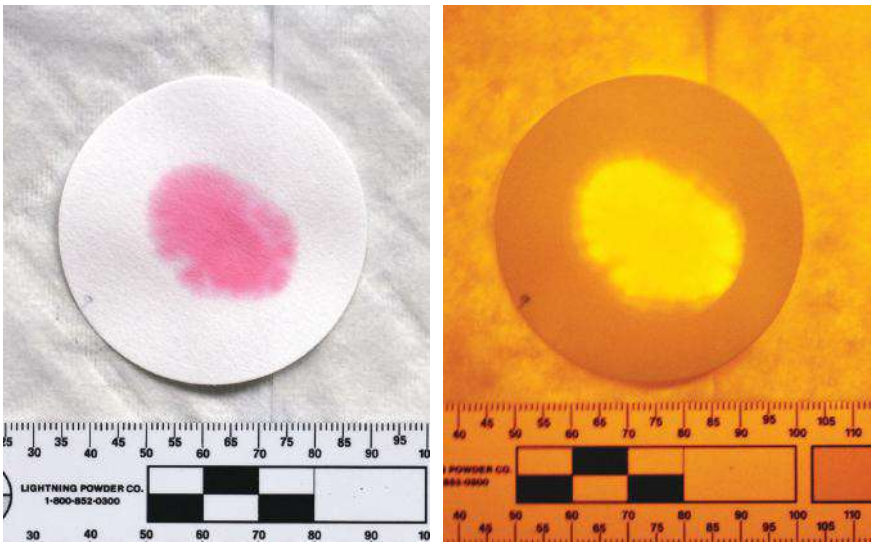

(a) (b)

Figure 17.5 Positive reaction of DMAC assay of a urine stain. (a) Colorimetric and (b) fluorometric method. (© Richard C. Li.)

considered a negative reaction. However, this method is not specific to urine, as other bodily fluids such as saliva, semen, sweat, and vaginal secretions can also give positive reactions. A diluted DMAC solution, from 0.1% to 0.05%, can maintain appropriate sensitivity to urine stains and minimize false-positive reactions caused by the detection of low levels of urea that are present in other bodily fluids.

The fluorometric method is useful for locating urine stains on large pieces of evidence such as clothing and bedding (Figure 17.6). Additionally, this method can detect patterns of urine stains (Figure 17.7), which can be useful in crime scene reconstructions. In the fluorometric method, the evidence to be examined, such as a garment, is covered by a sheet of filter paper that has been preabsorbed with the DMAC solution. The evidence and the filter are then wrapped together in a sheet of aluminum foil and are left overnight in a press, ensuring that the evidence is in close contact with the filter paper. Alternatively, the layers can be heated for 30 s using an iron. Using a light source at 473–548 nm, the DMAC-treated urine stain fluoresces. The fluoresced stain is best observed with a 549 nm filter. However, colored fabrics interfere with the assay since dyes and pigments can inhibit the fluorescence.

In addition to the DMAC test, the identification of urea that is present in urine stains can be carried out by urease assays. Ureases catalyze the breakdown of urea, thus releasing ammonia and carbon dioxide (Figure 17.8). The ammonia is detected using an acid-base indicator, bromthymol blue, which exhibits a blue color. Alternatively, the ammonia can be detected by manganese and silver nitrates, which exhibit a black color. Additional assays such as microscopic crystal and chromatographic assays are summarized in Table 17.1.

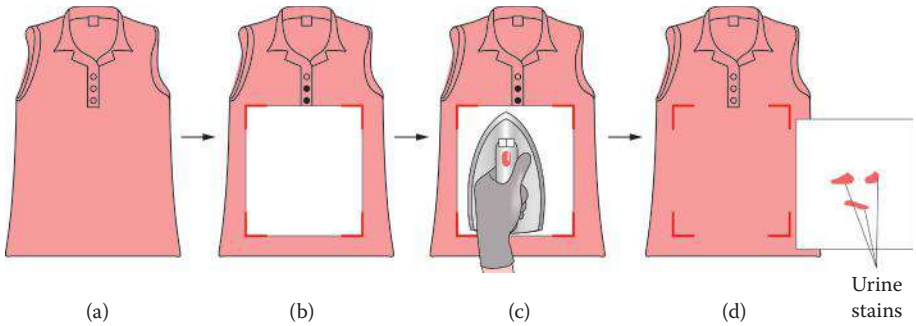

Figure 17.6 Fluorometric DMAC assay of urine stains. Evidence item (a) is closely covered by a piece of moistened filter paper containing DMAC. The orientation of the paper is marked (b). The layers are pressed to allow the transfer of a small amount of the stain onto the paper and are heated using an iron (c). The paper is then lifted. The DMAC-treated urine stain fluoresces under a light source at 473–548 nm (d). (© Richard C. Li.)

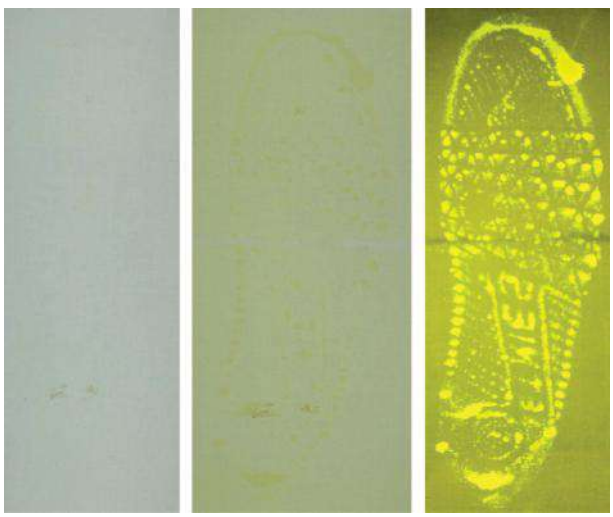

Figure 17.7 Fluorometric DMAC enhancement of a footwear impression in urine on white cotton fabric. From left to right: urine impression, DMAC-treated urine stains under white light, and DMAC-treated urine stains using an excitation light source. (From Farrugia, K.J., et al., *Forensic Sci Int*, 214, 67–81, 2012. With permission.)

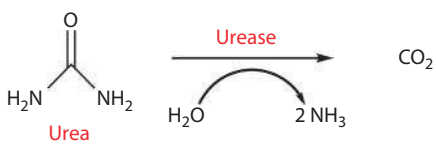

Figure 17.8 Biochemical reaction of urease.

17.1.2.2 Identification of Creatinine

Creatinine is produced during normal muscle cell metabolism (Figure 17.9). During this metabolism, phosphocreatine, an energy-storing molecule in muscle cells, breaks down to form creatine. Creatine is then metabolized to creatinine, which is released from the muscle cells into the blood. Serum creatinine is largely filtrated by the renal glomeruli. A small amount of creatinine is secreted by the renal distal tubules. The amount of creatinine excreted in urine is proportional to the muscle mass of an individual. However, creatinine is not only present in urine. It is also present in other bodily fluids such as blood and semen.

The creatinine present in urine can be detected using the Jaffe color test (Figure 17.10). In this test, picric acid is used to convert creatinine, under alkaline conditions, to form creatinine picrate, which is a bright red product. Additional tests for urine creatinine are summarized in Table 17.1. Recently, a Uritrace device (Abacus Diagnostics) has been made commercially available for the detection of creatinine.

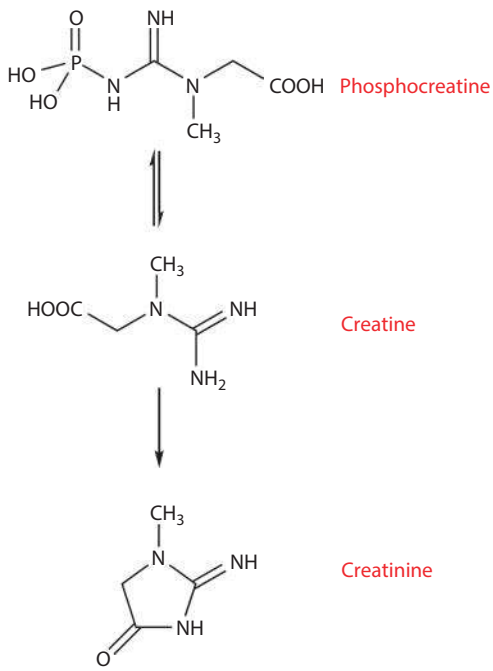

Figure 17.9 Schematic view of the formation of creatinine.

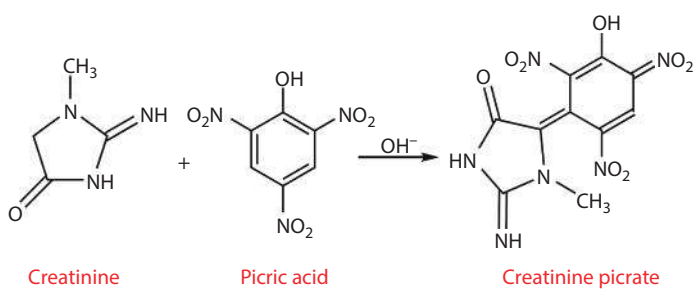

Figure 17.10 Chemical reaction of Jaffe's assay.

17.1.3 Confirmative Assays

17.1.3.1 Identification of Tamm–Horsfall Protein

Tamm–Horsfall protein (THP), also known as *uromodulin*, is the most abundant protein in urine, and accounts for 40% of the urine proteins. THP is exclusively synthesized in the epithelial cells of Henle's loop. THP is secreted from the apical plasma membrane of the epithelial cells into the lumen. Under physiological conditions, an adult excretes 20–100 mg of THP into urine daily. The biological function of THP is not fully understood. It is speculated that it prevents the body from contracting urinary tract infections and from forming renal stones.

THP is a urine-specific biomarker for forensic urine identification. It can be detected using an enzyme-linked immunosorbent assay (ELISA). More recently, it can be detected using an immunochromatographic assay, RSID-Urine, which utilizes a polyclonal rabbit antibody that is specific to THP (Figures 17.11 and 17.12). This test is rapid and simple and thus can be used as a screening test in laboratories and as a field test at crime scenes to identify urine. The detection limit of RSID-Urine for THP is 0.5 μL of urine. Although the sensitivity of RSID-Urine is lower than that of ELISA detection of THP, it is sufficient for forensic applications. These methods are specific to THP that is present only in urine while no detection is found in other bodily fluids such as plasma, saliva, semen, vaginal fluid, or sweat.

17.1.3.2 Identification of 17-Ketosteroids

In humans, urine contains derivatives of 17-ketosteroids such as androsterone, dehydroepiandrosterone (DHEA), and etiocholanolone. *Androsterone* is a steroid hormone with a weak potency of testosterone. *DHEA* is a metabolic intermediate in the biosynthesis of the gonadal steroids. DHEA also has a potential function as a steroid hormone. *Etiocholanolone* is a metabolite of testosterone. These compounds are present in urine as conjugates, in which 17-ketosteroids

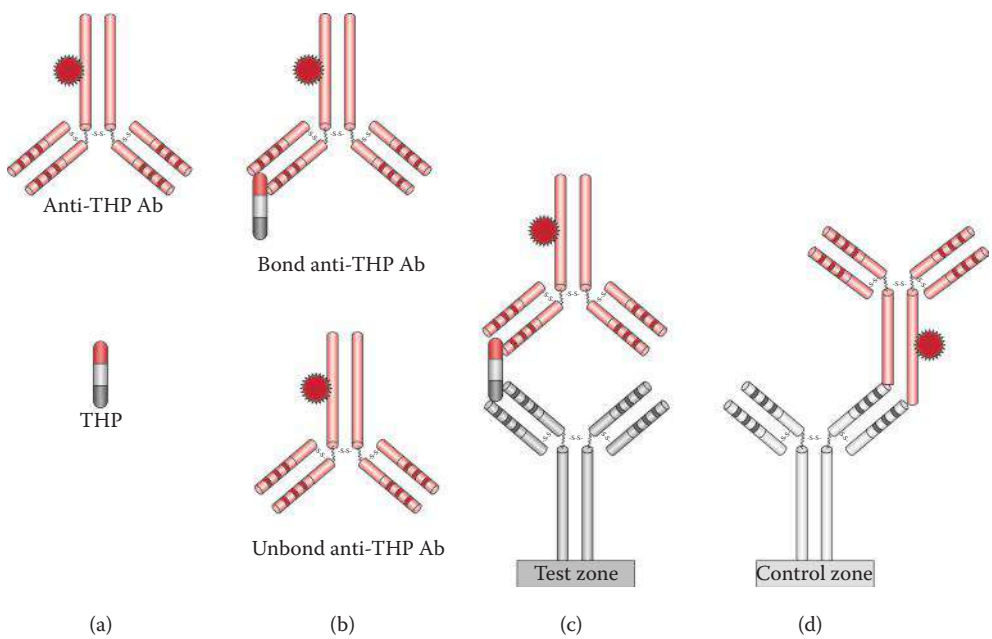

Figure 17.11 Immunochromatographic assays for the identification of THP in urine. (a) In a sample well, THP in a urine sample is mixed with labeled anti-THP Ab. (b) The THP binds to the labeled anti-THP Ab to form a labeled Ab–THP complex. (c) At the test zone, the labeled complex binds to an immobilized anti-THP Ab to form a labeled Ab–THP–Ab sandwich. (d) At the control zone, the labeled anti-THP Ab binds to an immobilized antiglobulin and is captured at the control zone. Ab, antibody; THP, Tamm–Horsfall protein antigen. (© Richard C. Li.)

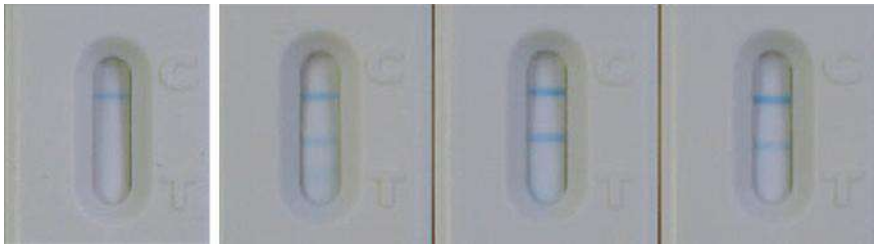

Figure 17.12 Results obtained using immunochromatographic assays (RSID™-Urine; Independent Forensics, Lombard, IL) for the identification of THP in urine. From left to right: negative control, undiluted, 1:10 diluted, and 1:100 diluted urine samples. A visible blue line at the test (T) zone and control (C) zone indicates a positive result. The blue line at the control zone only indicates a negative result. (From Akutsu, T., Watanabe, K., and Sakurada, K., *J Forensic Sci*, 57, 1570–1573, 2012. With permission.)

are modified in the liver through glucuronidation and sulfation. In glucuronidation, glucuronic acids are added to 17-ketosteroids while in sulfation, sulfates are transferred to 17-ketosteroids. As a result, these conjugates, containing charged moieties, are more water soluble than the non-conjugated 17-ketosteroids. Thus, they are excreted into urine to be eliminated from the body. The five major components of the 17-ketosteroid conjugates present in human urine are androsterone glucuronide, androsterone sulfate, DHEA sulfate, etiocholanolone glucuronide, and etiocholanolone sulfate (Figure 17.13). Thus, the analysis of 17-ketosteroid conjugates in urine stains is useful for the identification of human urine stains. These five 17-ketosteroid conjugates can be identified using liquid chromatography–mass spectrometry (LC-MS). However, some of these 17-ketosteroid conjugates are also detected in serum. Therefore, the presence of all five conjugated 17-ketosteroids in a sample is required to identify a urine stain. Additionally, the profiles of the 17-ketosteroid conjugates are human specific and are distinguishable between humans and animals.

17.2 Identification of Sweat

Sweat is the least common bodily fluid analyzed in forensic laboratories compared with others that have been mentioned in previous chapters. However, sweat identification is still useful for forensic investigations. For example, forensic DNA analysis allows the generation of DNA profiles from trace biological evidence such as fingerprints. Identifying sweat can be important for the analysis of these types of evidence.

17.2.1 Biology of Perspiration

Humans have two types of secretory sweat glands: the eccrine and the apocrine sweat glands (Figure 17.14). *Eccrine* sweat glands are distributed almost all over the body and are controlled by the sympathetic nervous system. Eccrine sweat glands play a role in regulating body temperature. When the body temperature rises, eccrine sweat glands secrete a watery sweat to the skin's surface where heat is carried away through the evaporation of the sweat to maintain normal body temperature. In humans, *apocrine* sweat glands, which are associated with hair follicles, are usually restricted to the underarm and genital areas and are controlled by emotional stress. Apocrine sweat glands are inactive until puberty. This type of sweat gland secretes an oily sweat that is odorous after being processed by skin bacteria. The majority of sweat evidence that is analyzed in forensic laboratories is sweat stains secreted from eccrine glands. Sweat contains water, minerals, lactate, and urea. Its biochemical composition varies

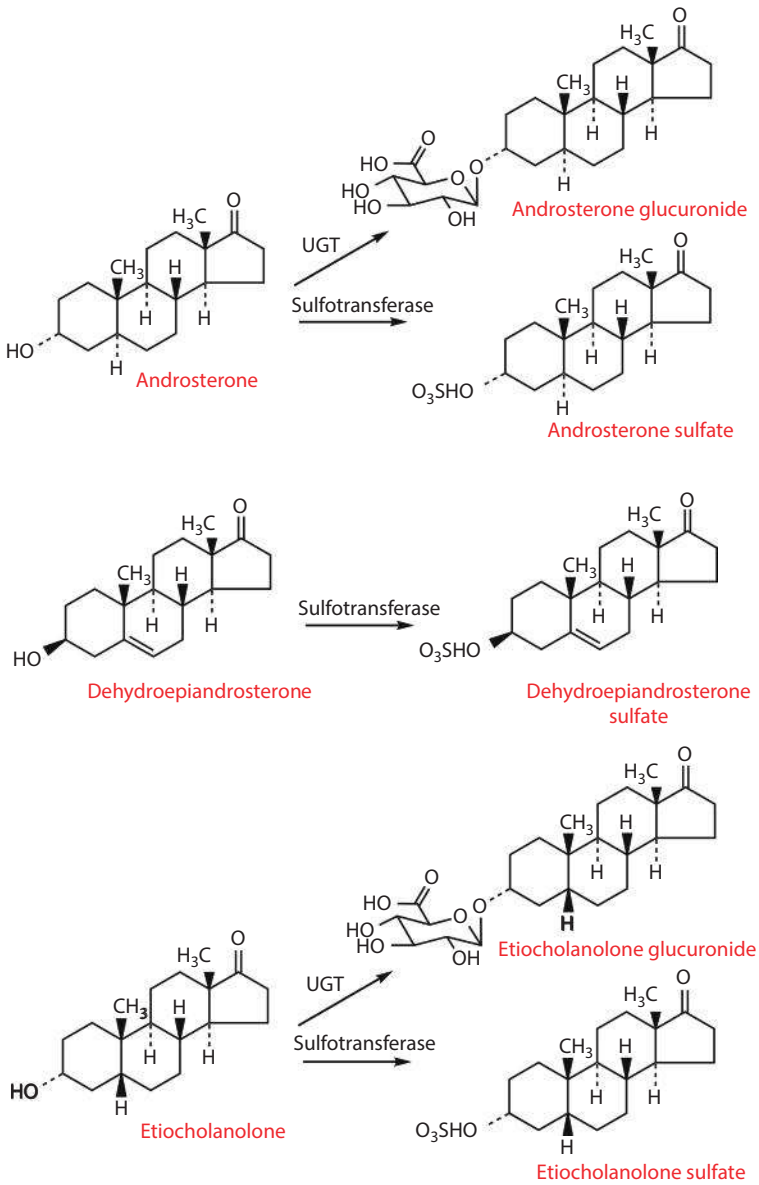

Figure 17.13 Structural formula of 17-ketosteroids and 17-ketosteroid conjugates present in human urine. UGT, UDP glucuronosyltransferase. (© Richard C. Li.)

among individuals and their physical activities. Sweat contains low levels of constituents that are also present in other bodily fluids such as urine. Thus, it has been considered a difficult bodily fluid to identify.

17.2.2 Sweat Identification Assays

Sweat evidence has been analyzed using presumptive assays (Table 17.1) such as elemental analysis using scanning electron microscopes coupled with energy dispersive x-ray spectroscopy in the detection of lactic acid. Since sweat contains many inorganic and organic compounds that are also present in other bodily fluids, these assays are not specific to sweat. Raman

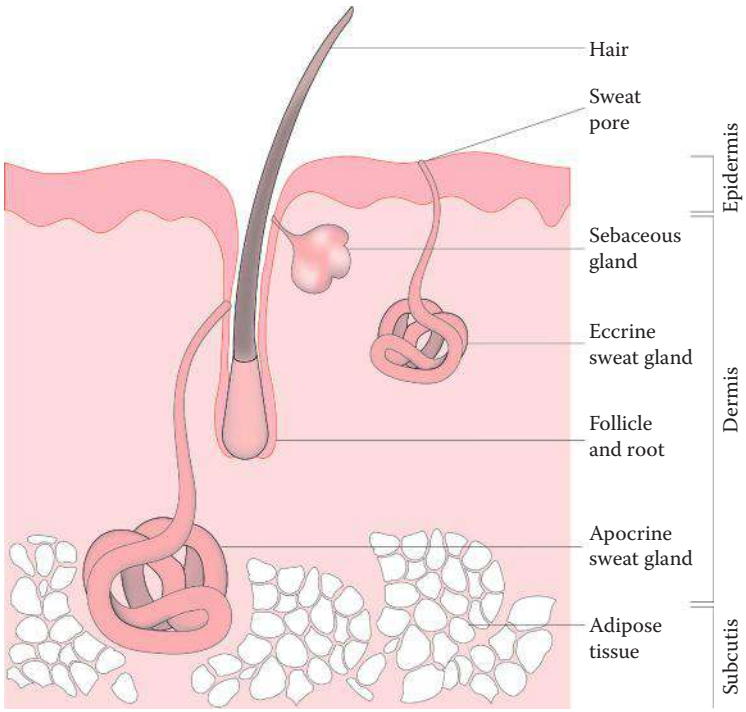

Figure 17.14 Diagram of the human eccrine and apocrine sweat glands. (© Richard C. Li.)

microspectroscopy is potentially useful for the identification of sweat for forensic purposes, which is based largely on the profiles of lactate, lactic acid, urea, and single amino acids in urine.

Recently, *dermcidin* has been identified as a potential biomarker of human sweat. Dermcidin belongs to a class of human antimicrobial peptides of the innate immune defense system and plays an important role in protecting epithelial barriers from infections. Dermcidin, specifically expressed in eccrine sweat glands (Figure 17.15), is secreted into the sweat and is transferred to the epidermal surface. In forensic applications, dermcidin can potentially be utilized as a biomarker for the confirmatory assay of sweat identification. The detection of dermcidin in sweat stains can be performed using ELISA assays utilizing antibodies specific to human dermcidin. This method is highly sensitive and is able to detect dermcidin in sweat samples that are diluted 10,000-fold. Dermcidin is encoded by the *DCD* gene. Its mRNA can be detected using reverse transcription polymerase chain reaction (RT-PCR) assays (Chapter 7) that can detect *DCD* mRNA in 10 μL of sweat sample. Dermcidin assays are also specific to sweat as dermcidin is not detected in other bodily fluids such as semen, saliva, and urine.

17.3 Identification of Fecal Matter

The examination of feces has been used in criminal investigations for over a century. Specifically in 1948, a burglary case was reported using fecal analysis to link the shoes of a suspect to a crime scene. One of the aspects of fecal analysis is to determine a common origin of the reference sample and the fecal evidence, thus potentially linking a suspect to a crime scene. Today, the individual characteristics of a fecal sample can be effectively determined using forensic DNA analysis of sloughed intestinal epithelial cells that are present in fecal matter. The identification of fecal matter is valuable in providing important information for a criminal investigation. For example, the presence of fecal matter may corroborate a sexual

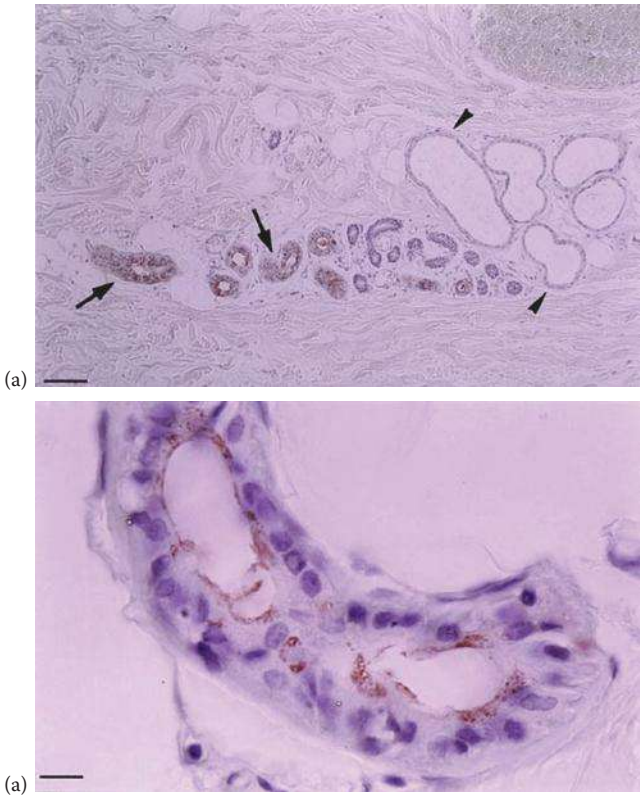

Figure 17.15 Localization of the dermcidin peptide in the sweat glands. Skin tissue sections are treated with a dermcidin antibody and stained. (a) Both the eccrine and apocrine sweat glands are shown. Only cells of the eccrine gland in the skin express dermcidin (brown stained). Arrows: eccrine sweat glands. Arrowheads: apocrine sweat glands. Scale bar: 100 μm. (b) Close-up view of the eccrine gland. The presence of the dermcidin peptide is observed (brown). Scale bar: 10 μm. (From Sagawa, K., et al., *Int J Legal Med*, 117, 90–95, 2003. With permission.)

assault involving sodomy, assault with fecal matter, vandalism, and burglary during which the perpetrator defecated at the scene.

17.3.1 Fecal Formation

Feces are a type of waste matter that is the direct result of food that has been processed by the digestive system (Figure 17.16). Human feces contain undigested foodstuffs, sloughed intestinal epithelial cells, intestinal bacteria, bile pigments, electrolytes, and water. Feces are formed in the intestines during the last phase of digestion. Feces first enter the colon in liquid form. Most of the nutrients are absorbed on the surface area of the small intestine. In the large intestine, water, sodium, and chloride are absorbed on the surface of the lumen. The remaining luminal contents are converted into feces. Food stays for approximately 2–6 h in the stomach. It takes an additional 3–5 h to travel through the small intestine and 12–24 h to travel through the large intestine.

17.3.2 Fecal Matter Identification Assays

A fecal analysis includes macroscopic and microscopic examination, chemical tests, and fecal bacterial identification.

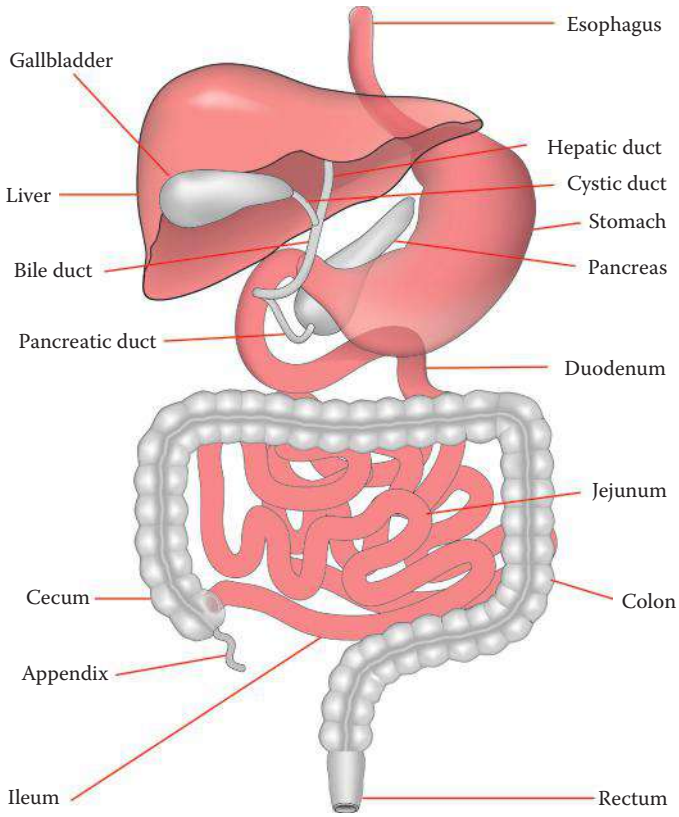

Figure 17.16 Diagram of the human digestive system. The esophagus, stomach, liver, gallbladder, pancreas, small intestine (duodenum, jejunum, and ileum), and large intestine (cecum, colon, appendix, and rectum) are shown. (© Richard C. Li.)

17.3.2.1 Macroscopic and Microscopic Examination

The color and odor of human feces are useful characteristics for fecal identification. The normal brown color of feces primarily results from the presence of urobilinoids, which are heme catabolic by-products. The characteristic odor of feces is caused by the metabolic by-products of the intestinal bacterial flora. Indole, skatole, and hydrogen sulfide are the compounds that are responsible for the odor of feces.

When fecal stains are analyzed, the microscopic examination of the feces can be performed on an aliquot of fecal suspension. The presence of characteristic undigested foodstuffs can indicate human feces. Fecal matter can be transferred from samples of clothing by scraping with a sterile stainless steel spatula. The fecal matter is then hydrated in 6% formalin solution for 24–48 h prior to microscopic examination. Undigested foodstuffs such as vegetable fragments and meat fibers are often present in human feces. Vegetable fragments (Figure 17.17) are often undigested vegetable dermal tissues that cover and protect the plant and fragments of vascular tissues that play roles in transporting water and nutrients throughout the plant. The types of vegetables can be identified by comparing the observed fragments with known plants. Meat fibers (Figure 17.18) are undigested animal skeletal muscle fibers. These fibers have characteristic striations, usually rectangular in shape, that are used for comparison with known animal tissues.

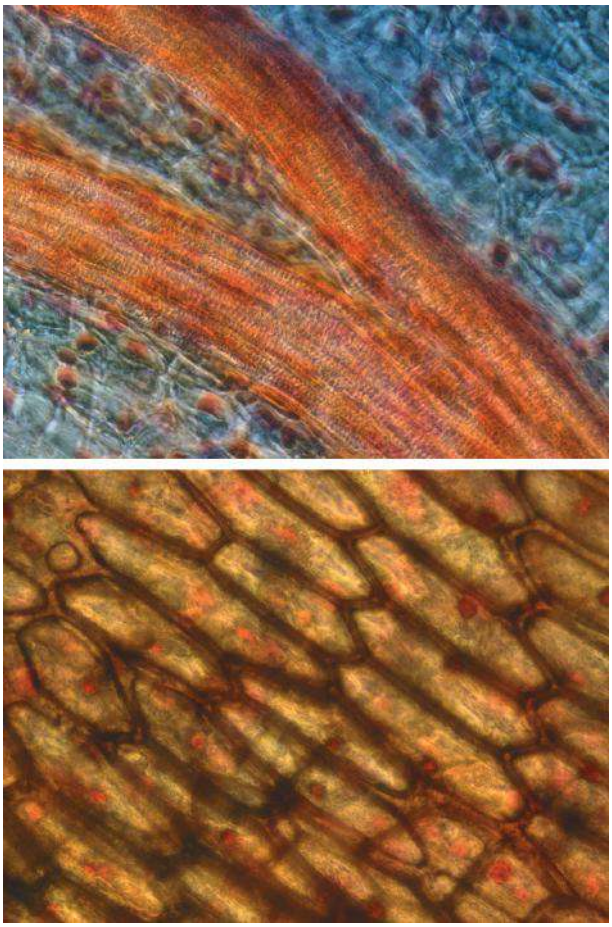

Figure 17.17 Scallion vascular tissue (top) and dermal tissue (bottom). (© Richard C. Li.)

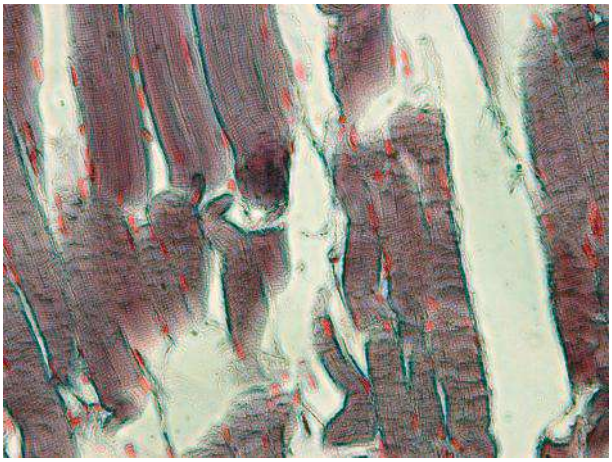

Figure 17.18 Cattle meat fibers. (© Richard C. Li.)

17.3.2.2 Urobilinoids Tests

The forensic analysis of fecal matter often involves the identification of fecal stains on swabs and clothing from which small amounts of sample may be available. Chemical tests can be useful for the analysis of fecal stains. By far the most common chemical test performed on feces is the detection of urobilinoids. *Urobilinoids*, including urobilin and stercobilin, are generated from the degradation of heme and are excreted into feces.

The average lifetime of erythrocytes is approximately 3–4 months. Erythrocytes are continuously undergoing hemolysis in which erythrocytes are naturally broken down and are usually processed in the reticuloendothelial system of the spleen. Hemoglobin is released daily during the hemolysis process and is degraded into heme, globin, and iron. Other sources of heme are derived from the degradation of erythrocyte precursors in bone marrow and other heme-containing proteins such as myoglobin and cytochromes. In the peripheral tissues, heme undergoes catabolism to form *bilirubin* (Figures 17.19 and 17.20). Bilirubin is further converted to *urobilinogen* in the intestine. A portion of urobilinogen is reduced to *stercobilinogen*. In the large intestine, the spontaneous oxidation of urobilinogen and stercobilinogen results in the formation of

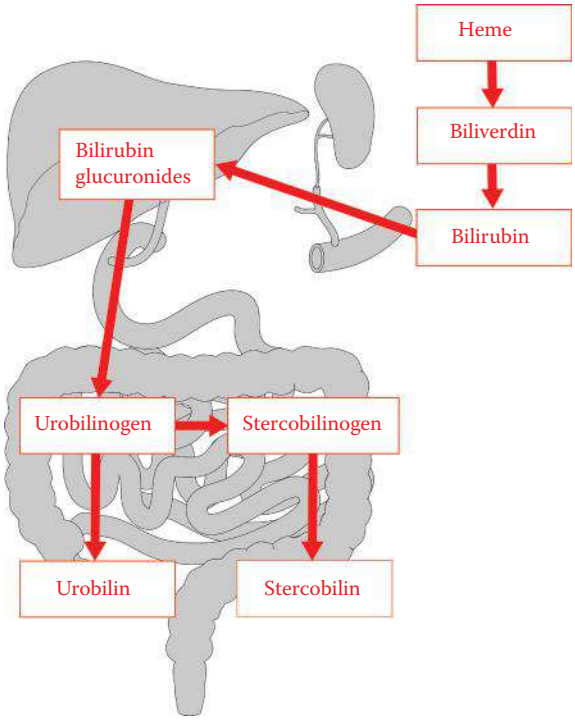

Figure 17.19 The formation of urobilinoids. Aged erythrocytes are disposed in the spleen, releasing hemoglobin that is broken down to heme. The heme is converted to biliverdin and is subsequently reduced to bilirubin. The bilirubin is then released into the bloodstream where it is bound to albumin, which cannot be filtrated at the glomeruli. The bilirubin is transported through the bloodstream to the liver where it is conjugated with glucuronic acid, forming water-soluble bilirubin monoglucuronide and diglucuronide. The bilirubin glucuronides are excreted into the bile and are subsequently excreted into the small intestine. In the intestines, the glucuronic acid of the conjugated bilirubin is removed. The unconjugated bilirubin is metabolized by intestinal bacteria, forming urobilinogen. A portion of the urobilinogen is further metabolized to stercobilinogen. The urobilinogen and the stercobilinogen are oxidized by intestinal bacteria, forming urobilin and stercobilin, respectively, which are excreted into the feces. (© Richard C. Li.)

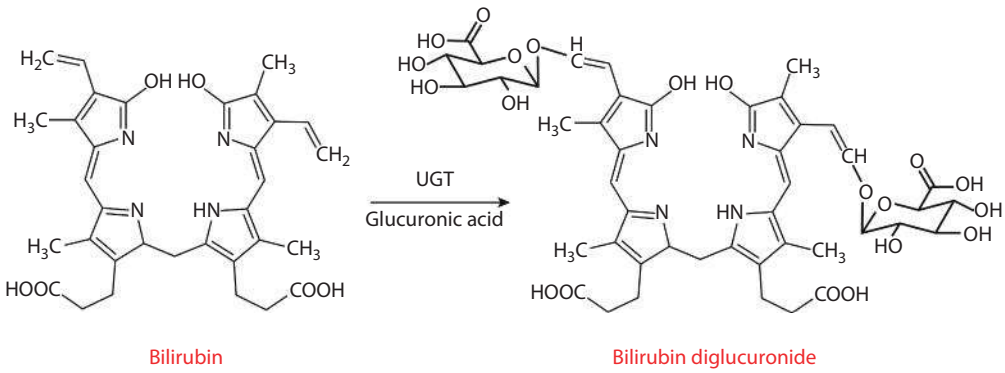

Figure 17.20 The formation of bilirubin glucuronides in the liver. UGT, UDP glucuronosyltransferases.

urobilin and stercobilin (Figure 17.21), respectively. These compounds are brown colored and are responsible for the characteristic color of feces.

The urobilinoids can be detected using the Schlesinger and Edelman tests. In the *Schlesinger* test, a sample is mixed with saturated zinc acetate in ethanol solution to form aurobilinoid–zinc chelation complex that emits a characteristic green fluorescence under ultraviolet light. The *Edelman* test is a variation of the Schlesinger test. A sample is treated with a mercuric salt solution to yield a pink-colored compound. Further treatment with a zinc salt produces fluorescence. However, less fluorescence is observed in the Edelman test than in the Schlesinger test. Inconclusive and inconsistent results are often obtained using these tests where fecal material sometimes gives no visible fluorescence. Additionally, the intensity of the fluorescence observed varies between samples. The reliability and selectivity of the tests can be increased using a spectrometric measurement of the fluorescence detection of fecal urobilinoids based on the principle of the Schlesinger test. A dry sample is treated with 1 mL of zinc acetate solution (1% zinc acetate methoxyethanol solution and 0.2% Tris). The suspension is then sonicated for 5 min, heated at 100°C for 10 min, cooled, and centrifuged. The presence of urobilinoids can be detected using excitation and emission maxima at 507 and 514 nm, respectively.

The disadvantages of the Schlesinger and the Edelman tests are their low species specificity as both tests cannot distinguish between human and other mammalian fecal materials. Moreover, under normal circumstances, up to 5% of urobilinogen is transported to the kidney and oxidized to urobilin in urine. Under some pathological conditions such as hepatic function disorders, the level of urobilin in the urine can be very high. Both the Schlesinger and Edelman tests respond to the urobilin of urine stains as well.

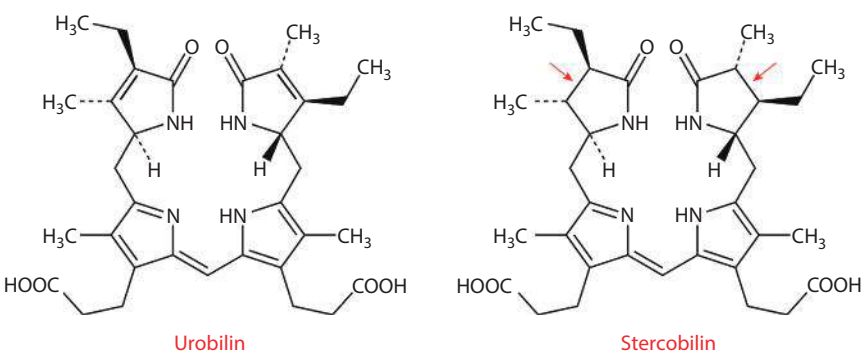

Figure 17.21 Urobilinoids: urobilin and stercobilin. Arrow, site of reduction.

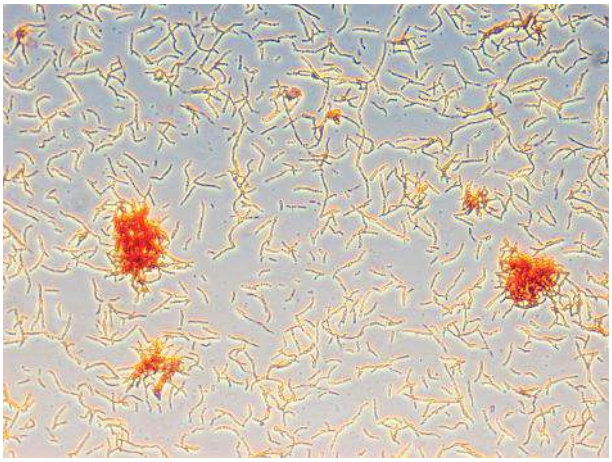

Figure 17.22 Intestinal bacteria. (© Richard C. Li.)

17.3.2.3 Fecal Bacterial Identification

The human intestinal microbiota contains more than 4000 bacterial species (Figure 17.22). Traditional methods for detecting fecal materials utilize the cultivation of fecal indicator bacteria such as *Escherichia coli* or *Enterococci* spp. However, these fecal indicator bacteria constitute only a small portion of the fecal microbiota, and thus may not be adequate for forensic identification purposes. *Bacteroides*, however, accounts for approximately 30% of fecal microbiota that are the predominant bacteria in human feces. *Bacteroides* can potentially be used for the forensic identification of feces. *Bacteroides* is a genus of rod-shaped, anaerobic gram-negative bacteria. These bacteria play a role in digesting complex carbohydrates and other substances that cannot be digested by human enzymes. The identification of *bacteroides* can be carried out by detecting specific DNA sequences of the *rpoB* gene, which encodes the β subunit of bacterial RNA polymerase (Chapter 11). The presence of the species-specific DNA sequence can be detected by RT-PCR utilizing primers that are specific to the target species but not to other species. Thus, only a targeted species can be amplified if it is present. Two fecal predominant bacteroides species, *B. uniformis* and *B. vulgatus*, can be detected in feces. *B. uniformis* is not detectable in blood, saliva, semen, urine, vaginal fluids, or on skin surfaces. Therefore, *B. uniformis* is considered as a specific indicator bacterium for forensic fecal identification. Sometimes, *B. vulgatus* can also be detected in vaginal fluid samples. Therefore, precaution should be taken in interpreting the results obtained using a *B. vulgatus* assay. Additionally, this method alone cannot discriminate between human and animal feces. Furthermore, fecal microbial populations can be affected by the host's diet. Individuals who consume saturated fats and proteins, which are abundant in Western diets, have predominantly *Bacteroides* species in their feces. However, individuals who consume a low-fat and carbohydrate-rich diet have predominantly *Prevotella* species, also a genus of gram-negative bacteria, in their feces.

17.4 Identification of Vomitus

17.4.1 Biology of Gastric Fluid

Gastric fluid can be found in stains derived from stomach wounds. Most often, gastric fluid is from vomitus found at a crime scene or as dried stains on clothing. Vomiting is the forceful expulsion of the contents of the stomach through the mouth (Figure 17.23). It is usually preceded by salivation, sweating, and the sensation of nausea. Vomiting usually begins with a deep

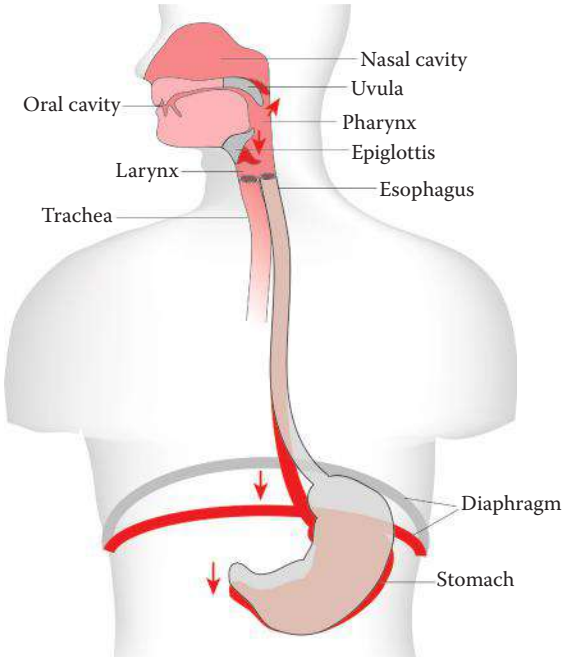

Figure 17.23 Act of vomiting. The contraction of the diaphragm and the stomach during the phase of vomiting is shown (in red). Additionally, the positions of the epiglottis and the uvula during vomiting are shown (in red). As a result, the stomach contents are expelled through the mouth. (© Richard C. Li.)

inhalation and the closure of the glottis. Subsequent contractions of the diaphragm and the abdominal muscles compress the stomach. The gastric contents are then forced upward through the relaxed sphincters and the esophagus, and are expelled through the mouth. Since the glottis is closed, vomitus usually does not enter the respiratory tract. Although the uvula is usually raised to close the nasal cavity, vomitus sometimes enters the nose. Vomiting can be caused by a wide variety of pathological conditions. Vomiting may be a specific response to acute intoxication in homicidal poisoning cases or it may be caused by trauma in a violent assault. Thus, the forensic identification of the gastric fluid can corroborate a criminal act.

The stomach stores ingested food until it can be emptied into the small intestines. When food enters the stomach, hydrochloric acid is secreted in large quantities, which facilitates the initial degradation of proteins. The stomach also secretes mucus that lubricates the gastric surface to protect the epithelium from acidic environments. Hormones such as gastrin, which are found in the gastric fluid, regulate acid secretion and gastric movement. A number of enzymes are secreted into the gastric fluid, including lipase, which plays a role in lipid hydrolysis, and gelatinase, which can hydrolyze gelatin. The stomach also secretes pepsinogens, which are enzyme precursors, into the gastric fluid. In the stomach, pepsinogens are activated by hydrochloric acid into pepsin, which is largely responsible for the digestion of proteins.

17.4.2 Vomitus Identification Assays

Vomitus is highly acidic and tends to be malodorous. The color of vomitus may be of forensic interest. Fresh blood in the vomit is usually bright red and suggests bleeding due to injuries, while dark red blood clots suggest bleeding in the stomach due to pathological conditions such as an ulcer. A microscopic examination (Section 17.3.2.1) can be performed to identify recently ingested food particles that are present in a sample. The presence of gastric fluid in vomitus can also be identified by the detection of pepsins secreted from the stomach. This identification test

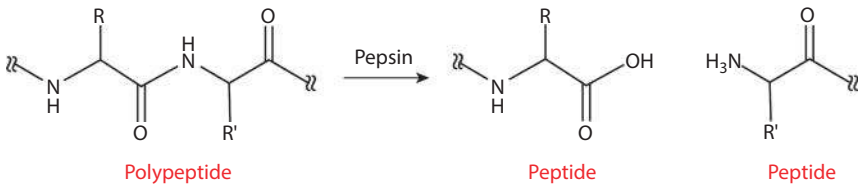

Figure 17.24 The cleavage of polypeptide by pepsin. Pepsin cleaves polypeptide at sites of aromatic amino acids. R and R′: tryptophan, phenylalanine, or tyrosine.

is based on the proteolytic activity of pepsins. Pepsins are endopeptidases that cleave primarily on peptide bonds in the middle of the protein. Aromatic amino acids such as tryptophan, phenylalanine, and tyrosine are the preferred targeted amino acids for the cleavage reaction by pepsins (Figure 17.24). In the pepsin-proteolytic assay, fibrin blue is used as a substrate for pepsin. Fibrin blue is an insoluble protein–dye complex that is colorless. In the presence of vomitus, pepsin cleaves fibrin blue and releases a chromophore that is soluble in water and exhibits a blue color. In the assay, a fibrin blue–containing agarose gel is utilized. The sample from vomitus is loaded onto the gel plate. After incubation, a blue ring around the sample can be observed as a result of the enzymatic reactivity of pepsin. The amounts of pepsin in a sample can be quantified. The results on the gel plate can be photographed and the dried gel plate can also be preserved as evidence. This method can determine the pepsin content of fresh and aged forensic samples. Bodily fluids other than vomitus do not show positive reactions with the use of this method. However, this method cannot distinguish the vomitus of humans from that of other vertebrates.

Bibliography

Akutsu, T., K. Watanabe, and K. Sakurada, Specificity, sensitivity, and operability of RSID-urine for forensic identification of urine: Comparison with ELISA for Tamm–Horsfall protein. *J Forensic Sci*, 2012, **57**(6): 1570–1573.

Akutsu, T., et al., Evaluation of Tamm–Horsfall protein and uroplakin III for forensic identification of urine. *J Forensic Sci*, 2010, **55**(3): 742–746.

Bedrosian, J.L., M.D. Stolorow, and M.A. Tahir, Development of a radial gel diffusion technique for the identification of urea in urine stains. *J Forensic Sci*, 1984, **29**(2): 601–606.

Biles, P.V. and G.C. Ziobro, Identification of the source of reagent variability in the xanthydrol/urea method. *J Assoc Official Anal Chem Int*, 1998, **81**(6): 1155–1161.

Bock, J.H., M.A. Lane, and D.O. Norris, *Identifying Plant Food Cells in Gastric Contents for Use in Forensic Investigations: A Laboratory Manual*, 1988. Washington, DC: US Department of Justice, National Institute of Justice.

Brunzel, N.A., *Fundamentals of Urine and Body Fluid Analysis*, 2nd edn., 2004. Philadelphia, PA: W.B. Saunder.

De Araújo, W.R., M.O. Salles, and T.R.L.C. Paixão, Development of an enzymeless electroanalytical method for the indirect detection of creatinine in urine samples. *Sensor Actuat B-Chem*, 2012, **173**: 847–851.

Farrugia, K.J., et al., Chemical enhancement of footwear impressions in urine on fabric. *Forensic Sci Int*, 2012, **214**(1–3): 67–81.

Farrugia, K.J., et al., A comparison of enhancement techniques for footwear impressions on dark and patterned fabrics. *J Forensic Sci*, 2013, **58**(6): 1472–1485.

Fujishiro, M., et al., Identification of human urinary stains by the uric acid/creatinine quotient and HPLC chromatogram. *J Showa Med Assoc*, 2008, **68**(3): 175–181.

Gaensslen, R.E., Identification of urine, in *Sourcebook in Forensic Serology, Immunology, and Biochemistry*, pp. 191–195, 1983. Washington, DC: US Government Printing Office.

Greenwald, I., The chemistry of Jaffe's reaction for creatinine: VI. A compound of picric acid with two molecules of creatinine. Its combinations with acid and alkali. *J Biol Chem*, 1930, **86**: 333–343.

18

Blood Group Typing and Protein Profiling

18.1 Blood Group Typing

18.1.1 Blood Groups

For the purposes of this text, blood groups are defined as antigen polymorphisms present on erythrocyte surfaces. Human erythrocyte surface membranes contain a variety of blood group antigens. Transfusion reactions occur when an incompatible type of blood is transfused into an individual, which can lead to severe symptoms or even death. Karl Landsteiner discovered the first blood group, known as the ABO system, in the early 1900s, while studying transfusion and transplantation. The discovery made blood transfusions feasible, and Landsteiner was awarded the Nobel Prize in 1930.

The International Society of Blood Transfusion currently recognizes 29 blood group systems, which include hundreds of antigen polymorphisms (Table 18.1). From the 1950s to the 1970s, the structures and biosynthesis pathways of many blood group antigens were determined. The genes for most of these blood group systems have been identified as well. The isolation of blood group genes has made it possible to understand the molecular mechanisms of the antigenic characteristics of the blood group systems.

The ABO system of antigens in human erythrocytes is the most commonly used blood group system for forensic applications. Forensic laboratories also use others, including the Rh, MNS, Kell, Duffy, and Kidd systems.

18.1.2 ABO Blood Group System

In the ABO blood group system, two types of antigens, designated A and B, give rise to four blood types:

- Type A individuals have the A antigen.
- Type B individuals have the B antigen.
- Type AB individuals have both A and B antigens.
- Type O individuals have neither A nor B antigens.

The antigens may be found in other bodily fluids as well as blood, such as amniotic fluid, saliva, and semen as well as many organs including the kidney, pancreas, liver, and lungs.

Table 18.1 Blood Group Systems

Number	Name	Symbol	Number of Antigens	Gene Names	Chromosomal Location
001	ABO	ABO	4	*ABO*	9q34.2
002	MNS	MNS	43	*GYPA*, *GYPB*, *GYPE*	4q31.21
003	P	P1	1		22q11.2-qter
004	Rh	RH	49	*RHD*, *RHCE*	1p36.11
005	Lutheran	LU	20	*LU*	19q13.32
006	Kell	KEL	25	*KEL*	7q34
007	Lewis	LE	6	*FUT3*	19p13.3
008	Duffy	FY	6	*FY*	1q23.2
009	Kidd	JK	3	*SLC41A1*	18q12.3
010	Diego	DI	21	*SLC4A1*	17q21.31
011	Yt	YT	2	*ACHE*	7q22.1
012	Xg	XG	2	*XG*, *MIC2*	Xp22.33, Yp11.3
013	Scianna	SC	5	*ERMAP*	1p34.2
014	Dombrock	DO	5	*DO*	12p12.3
015	Colton	CO	3	*AQP1*	7p14.3
016	Landsteiner-Wiener	LW	3	*ICAM4*	19p13.2
017	Chido/Rodgers	CH/RG	9	*C4A*, *C4B*	6p21.3
018	H	H	1	*FUT1*	19q13.33
019	Kx	XK	1	*XK*	Xp21.1
020	Gerbich	GE	8	*GYPC*	2q14.3
021	Cromer	CROM	12	*DAF*	1q32.2
022	Knops	KN	8	*CR1*	1q32.2
023	Indian	IN	2	*CD44*	11p13
024	Ok	OK	1	*BSG*	19p13.3
025	Raph	RAPH	1	*CD151*	11p15.5
026	John Milton Hagen	JMH	1	*SEMA7A*	15q24.1
027	I	I	1	*GCNT2*	6p24.2
028	Globoside	GLOB	1	*B3GALT3*	3q26.1
029	Gill	GIL	1	*AQP3*	9p13.3

Source: Adapted from Daniels, G.L., et al., *Vox Sang*, 87, 304–316, 2004.

18.1.2.1 Biosynthesis of Antigens

All individuals generate the *O antigen*, also known as the *H antigen*. The O antigen is synthesized by *fucosyltransferase*, a fucose transferase encoded by the *FUT* genes, which adds a fucose on the end of a glycolipid (in erythrocytes) or glycoprotein (in tissues). An additional monosaccharide (Figure 18.1) is then transferred to the O antigen by a transferase encoded by the *ABO* locus. The specificity of this enzyme determines the ABO blood type (Figure 18.2):

- The *A* allele produces the A-transferase, which transfers N-acetylgalactosamine to the O antigen and thus synthesizes the A antigen.
- The *B* allele produces the B-transferase, which transfers galactose to the O antigen and thus synthesizes the B antigen.
- The *O* allele has a mutation (small deletion), which eliminates transferase activity, and no modification of the O antigen occurs.

As a result, the A and B antigens differ in their terminal sugar molecules. Subgroups of blood types A and B have been described. The most important are the A_1 and A_2 antigens. Both A_1 and A_2 (and A_1B and A_2B) cells react with anti-A antibodies. However, A_1 cells react more strongly than A_2 cells. The apparent difference between A_1 and A_2 is that each A_1 cell contains more copies of the A antigen than A_2 cells.

18.1.2.2 Molecular Basis of the ABO System

A- and B-transferases are encoded by a single gene, *ABO*, on chromosome 9. The *ABO* gene (approximately 20 kb) is organized into seven exons. Most of its coding regions are located in exons 6 and 7 of the *ABO* locus, including the domain responsible for catalytic activity (Figure 18.3). The gene products of the *A* and *B* alleles differ by four amino acid substitutions (Table 18.2). In particular, amino acid residues at positions 266 and 268 are more important in determining the enzymatic property of a transferase.

The A^1 allele and A^2 allele differ in a single nucleotide deletion upstream from the translation stop codon. The resulting reading-frame shift in the A^2 allele abolishes the stop codon, yielding a product with an extra 21–amino acid residue at the C-terminus.

Subgroups of blood types O have also been reported. The sequence of the O^1 allele has a deletion of a single nucleotide at exon 6. This nucleotide deletion leads to a reading-frame shift generating a truncated protein, which lacks the catalytic domain. While the O^{1y} allele also has a single nucleotide deletion, it differs from O^1 by nine nucleotides within the coding sequence. O^1 and O^{1y} have identical phenotypes. There is also an O^2 allele, which is inactivated by a substitution

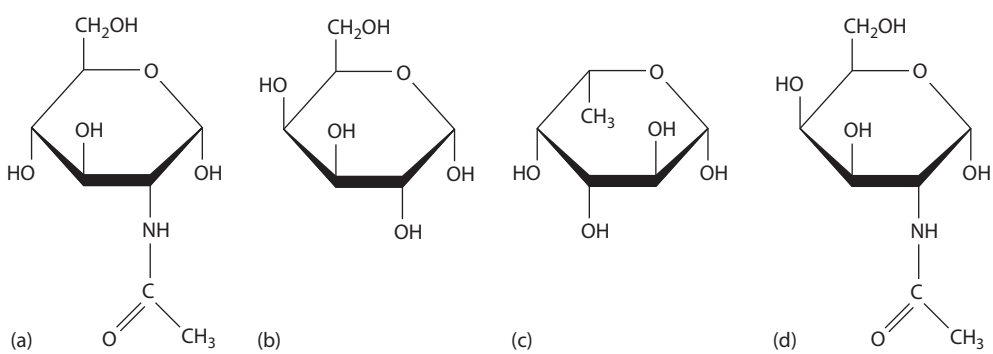

Figure 18.1 Chemical structures of (a) *N*-acetylglucosamine, (b) galactose, (c) fucose, and (d) *N*-acetylgalactosamine.

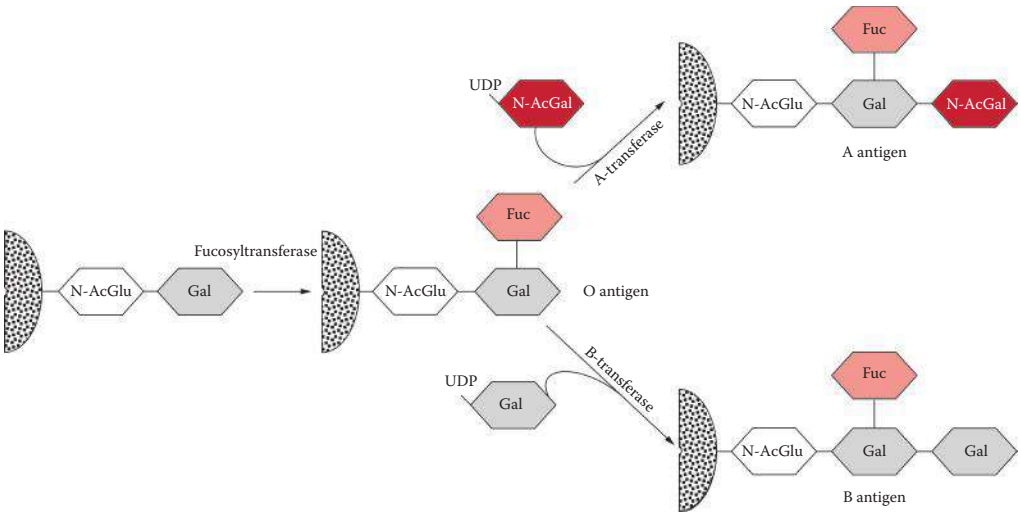

Figure 18.2 Biosynthesis of ABO antigens. O-antigen biosynthesis is catalyzed by fucosyltransferase. A-antigen biosynthesis is catalyzed by the A-transferase that transfers the N-acetylgalactosamine from the donor and uridine diphosphate (UDP)-*N*-acetylgalactosamine to the O antigen. B-antigen biosynthesis is catalyzed by the B-transferase that transfers the galactose from UDP-galactose to the O antigen. N-AcGlu, *N*-acetytglucosamine; Gal, galactose; Fuc, fucose; N-AcGal, *N*-acetylgalactosamine. (© Richard C. Li.)

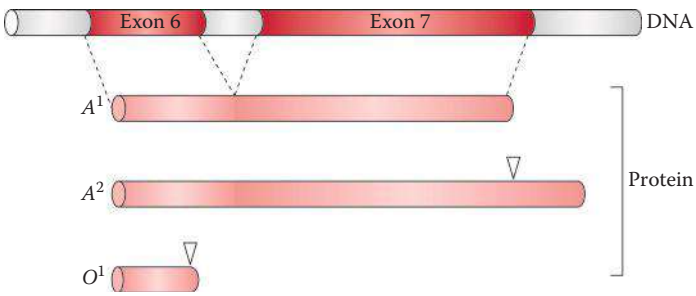

Figure 18.3 Structure of ABO gene and variants. Exons 6 and 7 are shown. The deletion mutation in A^2 and O^1 variants is indicated by an inverted triangle and leads to their A_2 and O phenotypes, respectively. (© Richard C. Li.)

Table 18.2 Amino Acid Substitutions at Four Positions in Human ABO Variants A^1, B, and O^2

ABO Variant	Amino Acid Position			
	176	235	266	268
A^1	Arginine	Glycine	Leucine	Glycine
B	Glycine	Serine	Methionine	Alanine
O^2	Glycine	Glycine	Leucine	Arginine

mutation at glycine (position 268) by arginine. Additionally, a few dozen other rare *O* alleles, which yield inactive proteins, have also been documented.

18.1.2.3 Secretors

In addition to erythrocytes, individuals whose A, B, and O antigens can be found in other types of bodily fluids are referred to as *secretors*. Eighty percent of Caucasians are secretors. As described earlier, the O antigen is the substrate for the A- and B-transferase because the A- and B-transferase can only utilize a fucosylated substrate. The O antigen is synthesized by fucosylation of the terminal galactosyl residue catalyzed by the fucosyltransferase, which is encoded by *FUT* genes.

Chromosome 19 contains two homologous genes: *FUT1* and *FUT2*. *FUT1* is expressed in tissues of mesodermal origin (embryonic tissues that serve as precursors of hemopoietic tissues, muscle, the skeleton, and internal organs) and is responsible for the synthesis of the O antigen in erythrocytes. *FUT2* is expressed in tissues of endodermal origin (embryonic tissues that are precursors of the gut and other internal organs); it is responsible for the synthesis of the O antigen in secretions.

About 20% of Caucasian individuals (called *nonsecretors*) are homozygous for a nonsense mutation in *FUT2* at amino acid position 143, resulting in a truncated protein. Bodily fluids such as the semen of type A or B nonsecretors (who carry homozygous *FUT2* mutations) contain no A or B antigens despite containing active A- or B-transferases (Figure 18.4). This can be a problem in investigating sexual assault cases when the blood type of the seminal evidence needs to be determined. However, nonsecretors have O antigens on erythrocytes synthesized by *FUT1* and thus have A or B antigens in blood. Individuals carrying very rare homozygous *FUT1* mutations produce erythrocyte O-deficient phenotypes in which the erythrocytes express no O antigens and thus express neither A nor B antigens, regardless of *ABO* genotype. Individuals

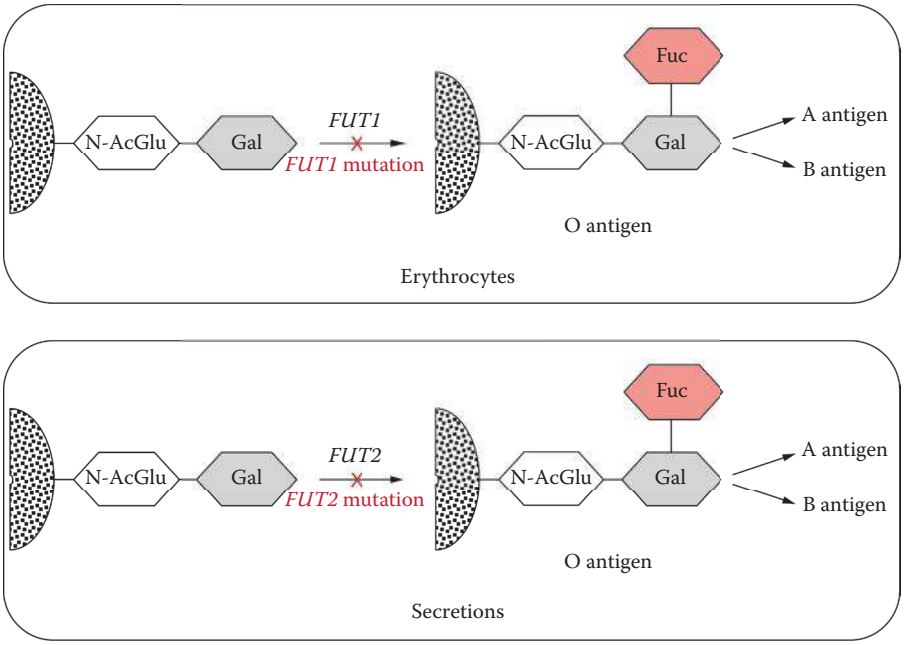

Figure 18.4 Tissue-specific O-antigen biosynthesis by *FUT1* and *FUT2* gene products. Tissue-specific O-antigen biosynthesis in erythrocytes is catalyzed by the *FUT1* gene product; in secretions, it is catalyzed by the *FUT2* gene product. The mutations abolishing the biosynthesis of O antigens are indicated. The *FUT2* mutation produces a nonsecretor phenotype. (© Richard C. Li.)

who carry both *FUT1* and *FUT2* mutations have no O antigens (nor A nor B antigens) in their erythrocytes and other bodily fluids and are known as Bombay (O_h) phenotypes.

18.1.2.4 Inheritance of A and B Antigens

A and *B* alleles are dominant. For *AO* and *BO* heterozygotes, the corresponding transferase synthesizes the A or B antigen. *A* and *B* alleles are codominant in *AB* heterozygotes because both transferase activities are expressed. The *OO* homozygote produces neither transferase activity and therefore lacks both antigens. The inheritance of *A* and *B* alleles obeys Mendelian principles (Chapter 25). For example, an individual with type B blood may have inherited a *B* allele from each parent or a *B* allele from one parent and an *O* allele from the other; thus, an individual whose phenotype is B may have the *BB* (homozygous) or *BO* (heterozygous) genotype. Conversely, if the blood types of the parents are known, the possible genotypes of their children can be determined. When both parents are type B (heterozygous), they may produce children with the genotype *BB* (B antigens from both parents), *BO* (B antigen from one parent, O from the other heterozygous parent), or *OO* (O antigens from parents who are both heterozygous). Thus, blood group typing can be used for paternity testing.

18.1.3 Forensic Applications of Blood Group Typing

The application and usefulness of blood typing in forensic identification are based on the ability to group individuals into four different types using the ABO blood system, allowing individuals to be identified. For example, if one crime scene blood sample is type B and a suspect has type A, the crime scene sample must have a different origin. However, if both the sample and the suspect are type A, the sample may have come from the same origin or from a different origin that happened to be type A.

Unfortunately, the probability that any two randomly chosen individuals have an identical blood type is very high. Approximately 42% of Caucasians have type A blood. The frequency of other blood types within the ABO system is shown in Figure 18.5. Multiple blood group systems were utilized to decrease the probability of a coincident match.

The A and B antigens are very stable and can be identified in dried blood even after many years. They can also be found in semen and other bodily fluids of secretors. Thus, in sexual assault cases, for example, the ABO type of a semen sample can be examined to identify a perpetrator.

18.1.4 Blood Group Typing Techniques

The most common assays used in forensic serology involve agglutination and include the Lattes crust and absorption–elution assays.

18.1.4.1 Lattes Crust Assay

In the early 1900s, Karl Landsteiner used his blood and blood obtained from his laboratory coworkers to test the effects of serum on erythrocytes. He discovered that naturally occurring

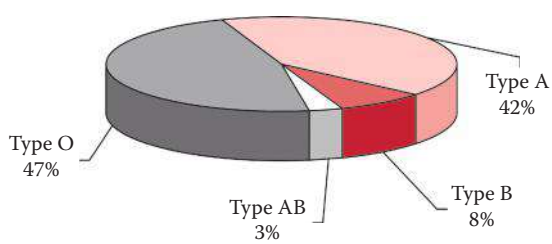

Figure 18.5 Frequency distributions of ABO types observed in American Caucasians. Different human populations may exhibit different frequencies of the four blood types. (© Richard C. Li.)

antibodies in serum caused agglutination of certain erythrocytes, and the agglutination patterns observed were designated A, B, and O. Each pattern indicated the presence or absence of a particular antigen on erythrocytes.

Shortly after birth, newborn infants develop antibodies against antigens that are not present in their own bodies. For example, type A individuals develop anti-B antibodies, type B individuals develop anti-A antibodies, type O individuals develop both types of antibodies, and type AB individuals do not develop anti-A or anti-B antibodies. When the plasma of a type A individual is mixed with type B cells, the anti-B antibodies from the type A individual cause the type B cells to agglutinate. This result forms the basis for blood group typing.

The Lattes crust assay relies on the principles of Landsteiner's experiments. It is an agglutination-based assay that utilizes the A, B, and O indicator cells to test the agglutination reaction with its corresponding naturally occurring serum antibodies in a questioned sample. The procedure for the Lattes crust assay is described in Box 18.1 and illustrated in Figure 18.6. Typical results are summarized in Table 18.3 and illustrated in Figure 18.7. Type A blood contains naturally occurring anti-B antibodies that agglutinate only with B cells. Likewise, type B blood agglutinates only with A cells, type O blood agglutinates with both A and B cells, and type AB blood does not agglutinate with any cells.

The Lattes crust assay is simple and rapid. However, one limitation is that the assay is not very sensitive and requires a large quantity of blood. Recall that successful agglutination reactions usually require intact cells. The agglutination assay of forensic samples is, therefore, difficult to carry out because blood cells lyse when they are dry. Therefore, this method is not reliable for testing old stains.

BOX 18.1 LATTES CRUST ASSAY PROCEDURE

1. Place small quantities of blood crust from a specimen on a microscopic slide and place a cover slide over the crusts. Prepare slides for A, B, and O cells separately.
2. Prepare cell suspensions with saline (0.85% NaCl in phosphate buffer, pH 7.4) for the A, B, and O cells separately.
3. Apply a few drops of the A-cell suspension and allow the cells to diffuse under the cover slip. Repeat this step for B cells and O cells.
4. Incubate the slides in a moisture chamber at room temperature for 2 h.
5. Examine results under a microscope.

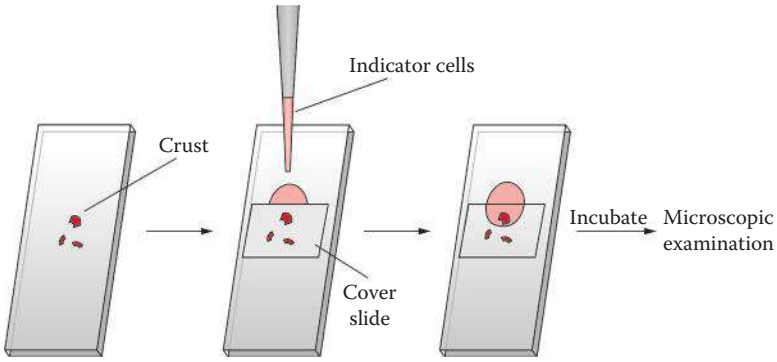

Figure 18.6 Lattes crust assay. (© Richard C. Li.)

Table 18.3 Representative Results of Lattes Crust Assay

Blood Type	Serum Antibody	Agglutination Reaction Observed
A	Anti-B	B cells
B	Anti-A	A cells
O	Anti-A, anti-B	A cells, B cells
AB	None	None

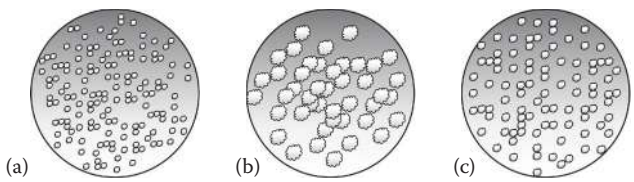

Figure 18.7 Diagram of Lattes crust assay results. (a) Indicator cells added before incubation. (b) Strong agglutination: large clumps are observed after incubation. (c) Negative agglutination: a cloudy background may be observed after incubation.

18.1.4.2 Absorption–Elution Assay

The absorption–elution assay is highly sensitive and can be used for testing dried bloodstains. This method indirectly detects the presence of antigens. The antigens are immobilized in a solid phase (Figure 18.8). At low temperatures, the antigens bind to their corresponding antibodies: anti-A antibodies, anti-B antibodies, or anti-O lectins. (The anti-O lectin is isolated from plants and reacts strongly with the O antigen present in type O blood, but has some cross-reaction with the A antigen). The excess unbound antibodies are removed by washing, and the bound antibodies are then eluted at higher temperatures (recall that antigen–antibody binding can be affected by temperature; Chapter 13). The eluted antibodies can then be identified by an agglutination assay using A, B, and O indicator cells.

Typical results of an absorption–elution assay are summarized in Table 18.4. The bloodstains containing the A antigen can bind to anti-A antibodies. The eluted anti-A antibody can form agglutination with A cells. Likewise, for type B blood, the eluted anti-B antibody can form agglutination with B cells; for type AB blood, the eluted antibodies can form agglutination with both A and B cells; and with type O blood, the eluted anti-O lectins can form agglutination with O cells.

18.2 Forensic Protein Profiling

Because of the limitations of blood group systems, inherited protein polymorphic markers have been utilized to decrease the chances of matches between two unrelated individuals. The amino acid sequences of many proteins vary in the human population. An estimated 20%–30% of the proteins in humans are polymorphic. Some of the variations in amino acid sequences affect the function of proteins, but many of them exert little or no effect on protein function. Thus, individuals can be divided into groups based on the types of protein polymorphisms. A combination of the blood group systems and protein polymorphic markers can be used for

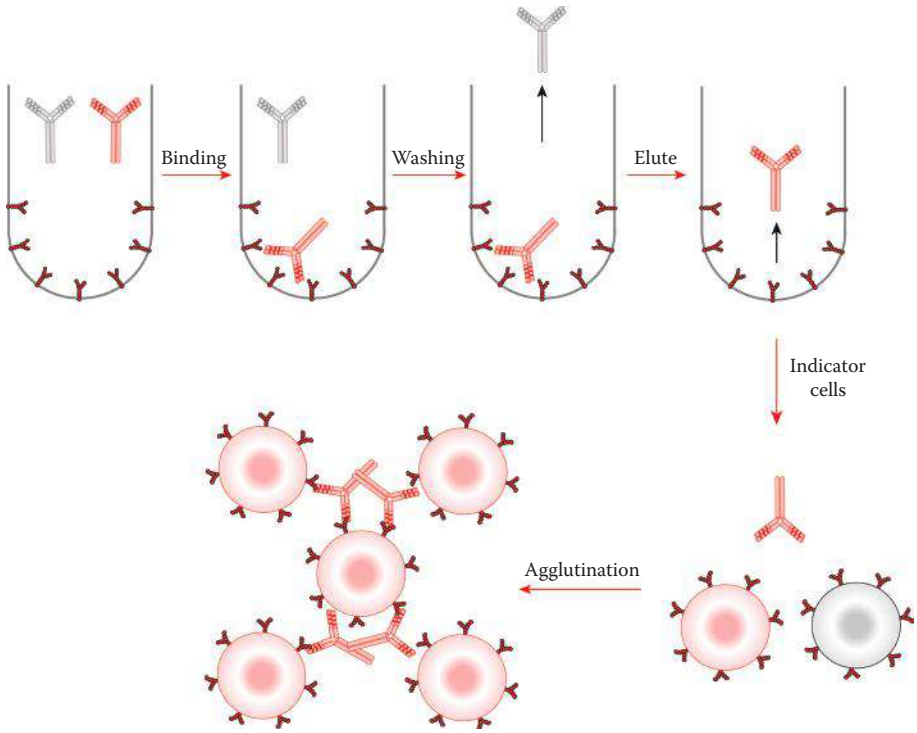

Figure 18.8 Absorption–elution assay. An antigen sample is immobilized on a solid-phase matrix. Antibodies (two different types are shown) are added. The antibody that is specific to the antigen binds. Unbound antibody is washed away. The bound antibody is then eluted. The eluted antibody is tested with indicator cells (two different types are shown). The indicator cell that is specific to the antibody eluted shows a positive agglutination reaction. (© Richard C. Li.)

Table 18.4 Representative Results of Absorption–Elution Assay

Blood Type (Stain)	Antibody Bound and Eluted	Agglutination Reaction Observed
A	Anti-A	A cells
B	Anti-B	B cells
O	Anti-O	O cells
AB	Anti-A, anti-B	A cells, B cells

criminal investigations and paternity testing. The probability that results for two unrelated persons would match is decreased to one in several hundred through use of the blood-typing and protein-profiling techniques.

18.2.1 Methods

Identification of protein polymorphisms is performed through electrophoretic separation based on the molecular weights (*M*r) and charges of the protein variants.

18.2.1.1 Matrices Supporting Protein Electrophoresis

Electrophoresis of proteins is generally carried out in a support material, also called the matrix, to separate various macromolecules. The matrix also reduces the effects of diffusion and convection on the macromolecules. Historically, protein profiling for forensic application utilizes two types of matrices: papers such as cellulose acetate; and gels composed of starch, agar, agarose, or polyacrylamide. The first polymorphic protein marker, phosphoglucomutase, was characterized by starch-gel electrophoresis. However, agarose and polyacrylamide became more commonly used in electrophoresis due to good reproducibility and reliability (Table 18.5).

18.2.1.2 Separation by Molecular Weight

An electrophoretic method is frequently utilized to resolve various proteins based on their molecular weights. Native electrophoresis, also known as nondenaturing electrophoresis, can be used to isolate proteins for studying the functions of proteins. Biological activity of the protein can be retained for further analysis. However, some proteins are not well separated in electrophoresis in their native form. Thus, it may be necessary to denature the proteins in order for them to be resolved better during separation. This process is called *denaturing protein electrophoresis*. The following additives can be used:

18.2.1.2.1 Reducing Agents

It is common to include reducing agents such as mercaptoethanol (ME), dithiothreitol (DTT), or sodium mercaptoethane sulfonate (MESNA) to denature proteins. Reducing agents cleave the disulfide bonds of proteins. As a result, protein shape becomes unfolded and linear. These agents can be used during sample preparation and can also be added to the electrophoresis buffer.

18.2.1.2.2 Detergents

Detergents disrupt noncovalent interactions within the structures of native proteins. The procedure is generally performed with sodium dodecylsulfate (SDS), a strong anion detergent that binds to most proteins in amounts proportional to the molecular weight of the protein (approximately one molecule of SDS for two amino acids). The bound SDS contributes a large net negative charge on the protein, which masks any surface charges of the native protein. As a result, the charge-to-mass ratio of the protein becomes a constant. As with reducing agents, the various native conformations of proteins change to a more uniformly linear shape when SDS is bound.

Table 18.5 Properties of Matrices Supporting Protein Electrophoresis

Supporting Matrix	Pore Size	EEO	Reproducibility	Strength	Preparation	Toxicity
Cellulose acetate	Large	High	Poor	Good	Simple	Nontoxic
Starch	Large	High	Poor	Fragile	Simple	Nontoxic
Agar	Large	High	Poor	Fragile	Simple	Nontoxic
Agarose	Large	Low	Good	Fragile	Simple	Nontoxic
Polyacrylamide	Small	Very Low	Good	High; tolerates high electronic field	Complex	Toxic

Note: Electroendosmosis (EEO) occurs when fixed charges of the supporting matrix cause liquid flow toward the electrodes. A matrix with high EEO may affect the mobilities and separation performances of proteins during electrophoresis.

Electrophoretic mobility in the presence of SDS, therefore, becomes based on *M*r rather than both *M*r and the charge. Smaller proteins move through the pores of the gel matrix more rapidly than larger proteins. As a result, the larger the size of the protein, the smaller its electrophoretic mobility.

SDS gel electrophoresis can also be used to determine the *M*r of an unidentified protein based on its electrophoretic mobility on the gel. Standard marker proteins of known molecular weight are run on the same gel and allow the estimation of the *M*r of an unknown protein. A linear plot of log *M*r values of marker proteins versus relative migration during electrophoresis allows the molecular weight of the unknown protein to be determined from the graph.

18.2.1.3 Separation by Isoelectric Point

The *isoelectric focusing (IEF)* technique can be used to separate proteins according to their *isoelectric points (pI)*. The pI is the pH value at which the net electric charge of an amino acid is zero. All proteins are composed of amino acids, and each has its own characteristic pI at which its net electric charge is zero and does not migrate in an electric field.

In IEF electrophoresis, a pH gradient is created in a gel between the electrodes, and a protein sample is placed in a well on the gel. With an applied electric field, proteins enter the gel and migrate until they reach a pH equivalent to their pI values, at which they lose mobility (Figure 18.9). IEF, based on molecular charge, is capable of producing sharper bands than denaturing protein electrophoresis and thus has a higher resolving power. The technique can detect very low quantities of proteins in samples. A pH gradient in the gel is established by utilizing materials such as carrier ampholytes or immobilines that are dispersed in the gel. Carrier ampholytes are synthetic amphoteric compounds that contain multiple weak ionizable moieties acting as either acids or bases. To establish a pH gradient, a mixture of ampholytes with slightly different pIs is directly added to an IEF gel. The pH gradient is generated by applying an electric field on the ampholyte-containing gel. Under the electric field, the negatively charged ampholytes migrate toward the anode, and the positively charged ampholytes migrate to the cathode. As a result, a gradual pH gradient is created between the anodal end of the gel (acidic) and the cathodal side of the gel (basic). Immobilines are a series of modified acrylamide monomers that can be acidic or basic. The pH gradient of an IEF gel can also be established using a gradient-forming device that changes the proportion of the immobilines added to the gel matrix mixture as it is loaded into the gel-casting apparatus.

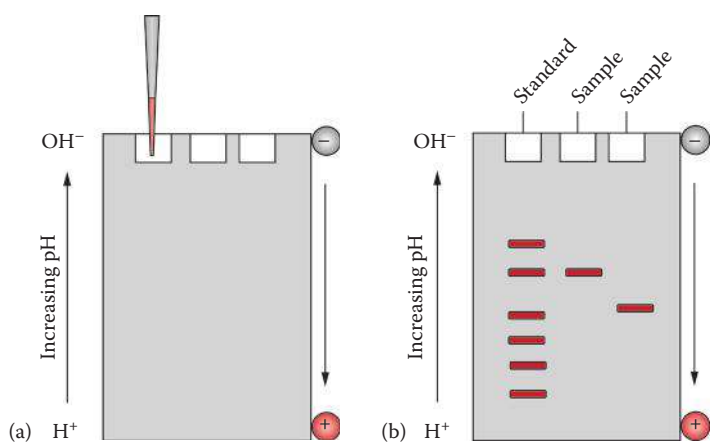

Figure 18.9 Isoelectric focusing. (a) A pH gradient is established by allowing low-molecular-weight organic acids and bases to distribute themselves in an electric field across a gel. A sample containing a protein mixture is loaded into a sample well. (b) During the electrophoresis, each protein migrates until it matches its pI. Proteins with different pIs are separated. (© Richard C. Li.)

18.2.2 Erythrocyte Protein Polymorphisms

18.2.2.1 Erythrocyte Isoenzymes

The human erythrocyte contains a number of *isoenzymes*, which are multiple forms of an enzyme that catalyze the same reaction but differ in their amino acid sequences. Individuals can be divided into groups on the basis of the different isoenzymes present in their erythrocytes. The isoenzyme type is also inherited according to Mendelian principles.

The polymorphism of erythrocyte phosphoglucomutase (PGM) was first described in the 1960s and was later successfully applied to the testing of bloodstains. PGM, an important metabolic enzyme, catalyzes the reversible conversion of glucose-1-phosphate and glucose-6-phosphate. The PGM found in erythrocytes is encoded at the *PGM1* locus at chromosome 1. The PGM encoded by *PGM1* can also be found in semen and thus can be utilized for the testing of semen samples in sexual assault cases. The protein polymorphisms of the PGM have two alleles, which result in three different phenotypes, depending on the combination of the two alleles. The success in the forensic application of PGM led to the similar use of many other erythrocyte isoenzyme polymorphisms. The most commonly used erythrocyte isoenzyme systems are listed in Table 18.6.

18.2.2.2 Hemoglobin

Recall that the use of hemoglobin (Hb) in screening and confirmatory blood tests was discussed in Chapter 12. Adult human Hb consists of two α chains and two β chains. Each polypeptide chain contains a heme group involved in oxygen binding. A very small portion of blood possesses a form of the human adult Hb consisting of two α chains and two δ chains.

More than 200 Hb variants have been identified and can be useful as markers for forensic applications. In particular, two types of human Hb variants are important in forensic testing: fetal Hb and sickle-cell Hb (Hb S). Hb S is the factor responsible for sickle-cell disease (Figure 18.10). Hb variants can be resolved using electrophoresis (Figure 18.11).

18.2.2.2.1 Fetal Hemoglobin

Humans have three forms of Hb during their development: embryonic, fetal, and adult Hb. In adults, the Hb tetramer consists of two identical α and two identical β chains. Embryonic erythrocytes contain Hb tetramers that are different from the adult form. Each embryonic Hb consists of two identical α-like chains and two identical β-like chains. The embryonic Hb is gradually replaced during pregnancy (approximately 3 months after conception) by fetal Hb, which comprises approximately 70% of the Hb in fetal blood. The fetal Hb has two identical α chains and two identical γ chains.

TABLE 18.6 Common Isoenzymes Used for Forensic Protein Profiling

Erythrocyte Isoenzyme	Protein Symbol	Number of Alleles
Phosphoglucomutase	PGM	2[a]
Erythrocyte acid phosphatase	ACP/EAP	3
Esterase D	ESD	2
Adenylate kinase	AK	2
Glyoxalase I	GLO	2
Adenosine deaminase	ADA	2

[a] Ten alleles can be observed using IEF electrophoresis.

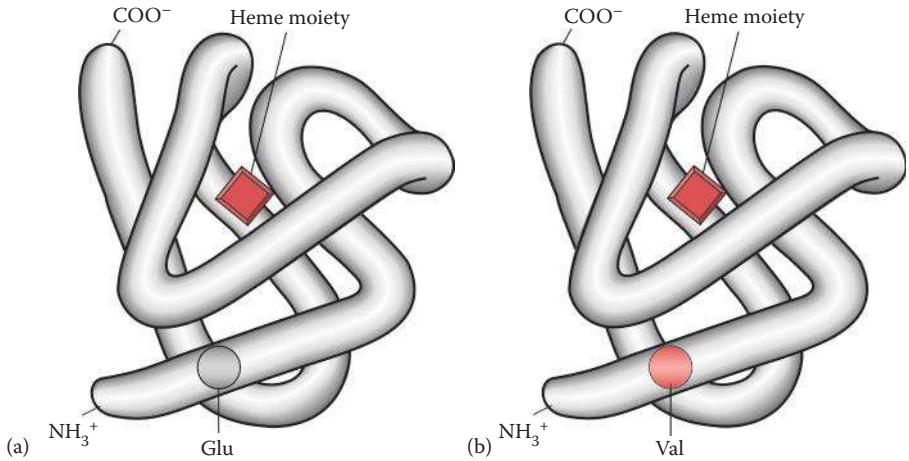

Figure 18.10 Normal and sickle-cell hemoglobin β chains. (a) Normal hemoglobin β chain contains a glutamic acid residue (Glu) at position 6 of the N-terminal of the protein. (b) At position 6 of the sickle-cell hemoglobin β chain, the glutamic acid residue is replaced by a valine (Val). (© Richard C. Li.)

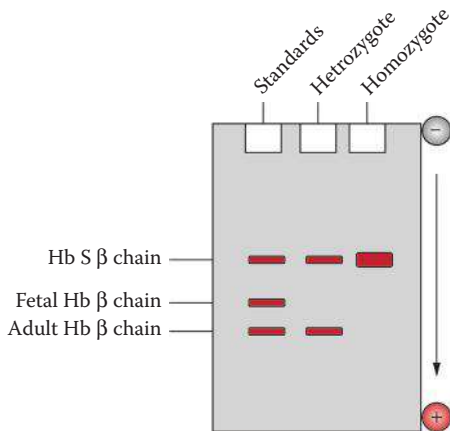

Figure 18.11 Hemoglobins resolved by isoelectric focusing electrophoresis. (© Richard C. Li.)

The embryonic and fetal Hbs have higher affinities to oxygen, required to provide the embryo and fetus with sufficient amounts of oxygen taken from maternal blood. Fetal Hb is replaced by adult Hb approximately 6 months after birth. These Hbs are encoded by their corresponding genes located at the globin gene clusters. The detection of fetal Hb in a bloodstain via electrophoresis can provide important evidence in cases of infanticide and concealed birth.

18.2.2.2.2 Hemoglobin S

Hb S polymorphism has forensic importance in identifying individuals. The Hb S polymorphism is observed in high frequencies among those of African heritage and some Hispanic populations. Such a protein polymorphic marker can provide investigational leads for the indication of the ethnic origin of a perpetrator. Hb S transports oxygen much less efficiently than

normal Hb. Individuals who are homozygous for Hb S usually die early after suffering from sickle-cell anemia and related complications. However, a heterozygous individual (an individual with a copy of the wild-type Hb allele from one parent and a copy of the Hb S allele from the other) can survive. This condition is known as the sickle-cell trait.

In the 1950s, Vernon Ingram of Cambridge University discovered the molecular mechanism of the Hb S defect. His work revealed that the Hb S bears a mutation, which changes the glutamic acid in wild type to a valine at the sixth amino acid from the N-terminal end of the β chain. This substitution of amino acids causes a major change in the structure of the β chain, which in turn results in sickle-cell anemia.

18.2.3 Serum Protein Polymorphisms

The serum portion of blood consists of a large number of proteins. The work on serum proteins for forensic purposes started in the 1950s, when variations in serum proteins were found useful for distinguishing individuals. Over the years, a number of serum proteins were characterized and applied for forensic testing. Haptoglobin (Hp) was the most widely used of the polymorphic serum proteins in forensic biology (Figure 18.12). Haptoglobin is a protein that binds and transports Hb from the bloodstream to the liver for the recycling of the iron contained in the Hb.

Immunoglobulin (Chapter 10) accounts for approximately 15% of serum protein and has been found to be highly variable. Two immunoglobulin proteins are utilized for forensic application. The γ chain protein (G_m) is the heavy chain of immunoglobulin G and the κ chain protein (K_m) is one of two types of the light chain of all immunoglobulins. Table 18.7 lists common serum group systems. All exhibit genetic variations and can be detected in bloodstains. The variants of these proteins can be determined by electrophoresis or serological methods.

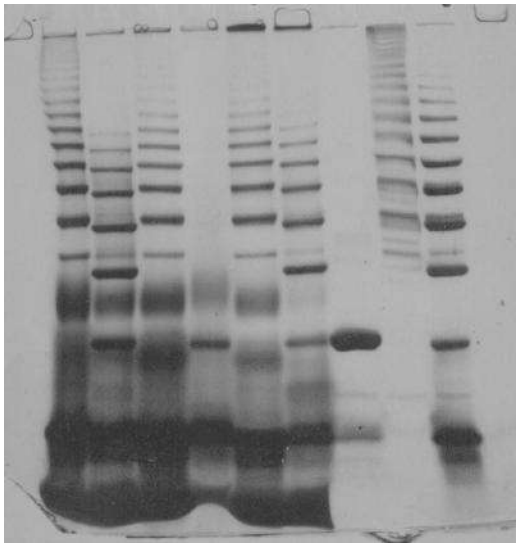

Figure 18.12 Polyacrylamide gel electrophoresis of haptoglobin proteins. From left to right: Hp2, Hp2-1, Hp2, Hp1, Hp2, Hp2-1, Hp1, Hp2, Hp2-1. An anode is at the bottom. (From James, S. and Nordby, J.J., *Forensic Science: An Introduction to Scientific and Investigative Techniques*, CRC Press, Boca Raton, FL, 2005. With permission.)

TABLE 18.7 Serum Proteins Used for Forensic Protein Profiling

Serum Protein	Protein Symbol	Gene Symbol	Chromosomal Location	Number of Amino Acids	Number of Alleles
Haptoglobin	Hp	*HP*	16q22.1	328/387	2
Group-specific component	Gc	*GC*	4q12–13	458	2
Transferrin	Tf	*TF*	3q22.1	679	3
Protease inhibitor (α1-antitrypsin)	Pi	*PI*	14q32.1	394	Many

Source: Adapted from Yuasa, I. and Umetsu, K., *Leg Med*, 7, 251–254, 2005.

Bibliography

Blood Groups

Achermann, F.J., et al., Soluble type A substance in fresh-frozen plasma as a function of ABO and secretor genotypes and Lewis phenotype. *Transfus Apher Sci*, 2005, **32**(3): 255–262.

Aki, K., A. Izumi, and E. Hosoi, The evaluation of histo-blood group ABO typing by flow cytometric and PCR-amplification of specific alleles analyses and their application in clinical laboratories. *J Med Invest*, 2012, **59**(1–2): 143–151.

Allen, F.H. Jr. and R.E. Rosenfield, Review of Rh serology. Eight new antigens in nine years. *Haematologia* (Budap), 1972, **6**(1): 113–120.

Anstee, D.J. and M.J. Tanner, Separation of ABH, I, Ss antigenic activity from the MN-active sialoglycoprotein of the human erythrocyte membrane. *Vox Sang*, 1975, **29**(5): 378–389.

Asari, M., et al., A new method for human ABO genotyping using a universal reporter primer system. *J Forensic Sci*, 2010, **55**(6): 1576–1581.

Bargagna, M. and M. Pereira, A study of absorption–elution as a method of identification of Rhesus antigens in dried bloodstains. *J Forensic Sci Soc*, 1967, **7**(3): 123–130.

Bargagna, M., M. Sabelli, and C. Giacomelli, The detection of Rh antigens (D,C,c,E,e) on bloodstains by a micro-elution technique using low ionic strength solution (LISS) and papain-treated red cells. *Forensic Sci Int*, 1982, **19**(2): 197–203.

Bassler, G., Determination of the Lewis blood group substances in stains of forensically relevant body fluids. *Forensic Sci Int*, 1986, **30**(1): 29–35.

Boorman, K.E., B.E. Dodd, and P.J. Lincoln, *Blood Group Serology*, 5th edn., vol. 4, p. 495, 1977. London: Churchill Livingstone.

Brauner, P., DNA typing and blood transfusion. *J Forensic Sci*, 1996, **41**(5): 895–897.

Brewer, C.A., P.L. Cropp, and L.E. Sharman, A low ionic strength, hemagglutinating, autoanalyzer for rhesus typing of dried bloodstains. *J Forensic Sci*, 1976, **21**(4): 811–815.

Busuttil, A., et al., Assessment of Lewis blood group antigens and secretor status in autopsy specimens. *Forensic Sci Int*, 1993, **61**(2–3): 133–140.

Camp, F.R. Jr., Forensic serology in the United States. I. Blood grouping and blood transfusion—Historical aspects. *Am J Forensic Med Pathol*, 1980, **1**(1): 47–55.

Cartron, J.P., et al., "Weak A" phenotypes. Relationship between red cell agglutinability and antigen site density. *Immunology*, 1974, **27**(4): 723–727.

Cayzer, I. and P.H. Whitehead, The use of sensitized latex in the identification of human bloodstains. *J Forensic Sci Soc*, 1973, **13**(3): 179–181.

Chatterji, P.K., A simplified mixed agglutination technique for ABO grouping of dried bloodstains using cellulose acetate sheets. *J Forensic Sci Soc*, 1978, **17**(2–3): 143–144.

Chester, M.A. and M.L. Olsson, The ABO blood group gene: A locus of considerable genetic diversity. *Transfus Med Rev*, 2001, **15**(3): 177–200.

19

Variable Number Tandem Repeat Profiling

Tandem repeats are abundant in the human genome. *Minisatellites* were first defined as a class of tandem repeats in the 1980s. Some of these repeats share a GC-rich core sequence. Subsequently, tandem repeats with higher AT contents of core sequence have also been characterized. The minisatellites are also called *variable number tandem repeats* (*VNTRs*), as shown in Figure 19.1. The repeat unit length of a VNTR can range from several to hundreds of base pairs (bp). The tandem repeat arrays can be kilobases (kb, corresponding to 10^3 bp) long, and the numbers of tandem repeat units in some VNTR loci are highly variable, leading to variable lengths of DNA fragments. A genotype is defined by a particular number of tandem repeat units at a given locus.

Table 19.1 lists the common VNTR loci used for forensic testing. To achieve high discriminating power, the VNTR loci should not be linked, which means that they should be inherited independently of each other. For example, loci located on different chromosomes or far apart on the same chromosome can be used (Section 21.2). Many VNTR loci used for forensic applications are highly polymorphic, and as many as hundreds of different genotypes per locus can be observed among the population. The discriminating power of VNTR loci used for forensic testing can be measured by *population match probability* (P_m; Chapter 25). The lower the P_m, the less likely a match will occur between two randomly chosen individuals. A P_m of up to 10^{-12} can be achieved by testing several VNTR loci.

19.1 Restriction Fragment Length Polymorphism

VNTR profiling utilizes RFLP—the first historical method used in forensic DNA testing (Figure 19.2). It utilizes *restriction endonucleases* that recognize and cleave specific sites along the DNA sequence. Cleavage of a DNA sample with a particular restriction endonuclease results in a reproducible set of restriction fragments of various lengths. Appropriate restriction endonucleases should be selected so that the genomic DNA is cleaved at sites that flank the VNTR core repeat region. The resulting fragments are then separated according to their sizes by gel electrophoresis through a standard agarose gel (Chapter 9).

The DNA is then processed using the *Southern transfer and hybridization* technique. The DNA is denatured and transferred from the gel to a supporting matrix such as a nylon or nitrocellulose membrane. The DNA immobilized on the membrane is then hybridized with a labeled *probe*. Only bands of DNA that have complementary sequences to the probe are recognized by detection systems such as autoradiography (Section 19.1.4). Using the RFLP technique, the

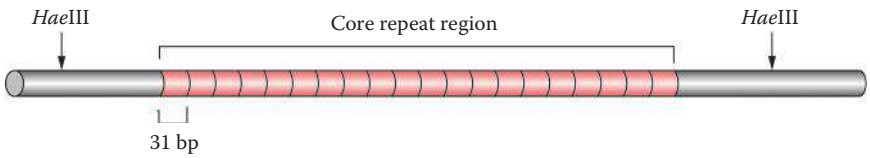

Figure 19.1 VNTR locus D2S44 (2q21.3–2q22). Each repeat unit consists of 31 bp. *Hae*III represents the *Hae*III restriction site.

Table 19.1 Common VNTR Loci

Locus	Chromosome Location	Repeat Unit Length (bp)	*Hae*III Fragment Size (kb)	Probe
D1S7	1	9	0.5–12	MS1
D2S44	2	31	0.7–8.5	yNH24
D4S139	4	31	2–12	pH30
D10S28	10	33	0.4–10	pTBQ7
D14S13	14	15	0.7–12	pCMM101
D16S85	16	17	0.2–5	3′HVR
D17S26	17	18	0.7–11	pEFD52
D17S79	17	38	0.5–3	V1

Source: Adapted from Budowle, B., et al., *DNA Typing Protocols: Molecular Biology and Forensic Analysis*, Eaton Publishing, Natick, MA, 2000; Office of Justice Programs, Future of forensic DNA testing: Predictions of the Research and Development Working Group. National Institute of Justice, US Department of Justice, 2000.

length variations among restriction sites can be detected. Most forensic applications focus on the length variations of VNTR regions located between two restriction sites.

In summary, the RFLP method includes several steps: (1) genomic DNA preparation, (2) restriction endonuclease digestion of the genomic DNA into fragments, (3) agarose gel electrophoretic separation of the DNA fragments according to size, (4) transfer of DNA fragments using Southern transfer, (5) hybridization with locus-specific probes, and (6) detection of locus-specific bands by autoradiography or chemiluminescence.

19.1.1 Restriction Endonuclease Digestion

Restriction endonucleases are enzymes that cleave the phosphodiester bond of DNA at or near specific recognition nucleotide sequences known as restriction sites. A restriction site usually is a short motif that is 4–8 bp in length. It often has a specific palindromic recognition sequence, that is, a segment of double-stranded DNA in which the nucleotide's sequence is identical with an inverted sequence in the complementary strand. Thus, double-stranded DNA is required to be cleaved by most restriction endonucleases. As a result, both sticky ends and blunt ends of restriction fragments can be generated after the cleavage (Figure 19.3).

To date, hundreds of restriction endonucleases have been described. They are traditionally classified into three types on the basis of subunit composition and enzymatic properties. Type II restriction endonucleases are most commonly used in molecular biology applications. Type II restriction endonucleases, requiring magnesium as a cofactor, usually cleave DNA at defined positions within their recognition sequences. The Enzyme Commission (EC) number of type II

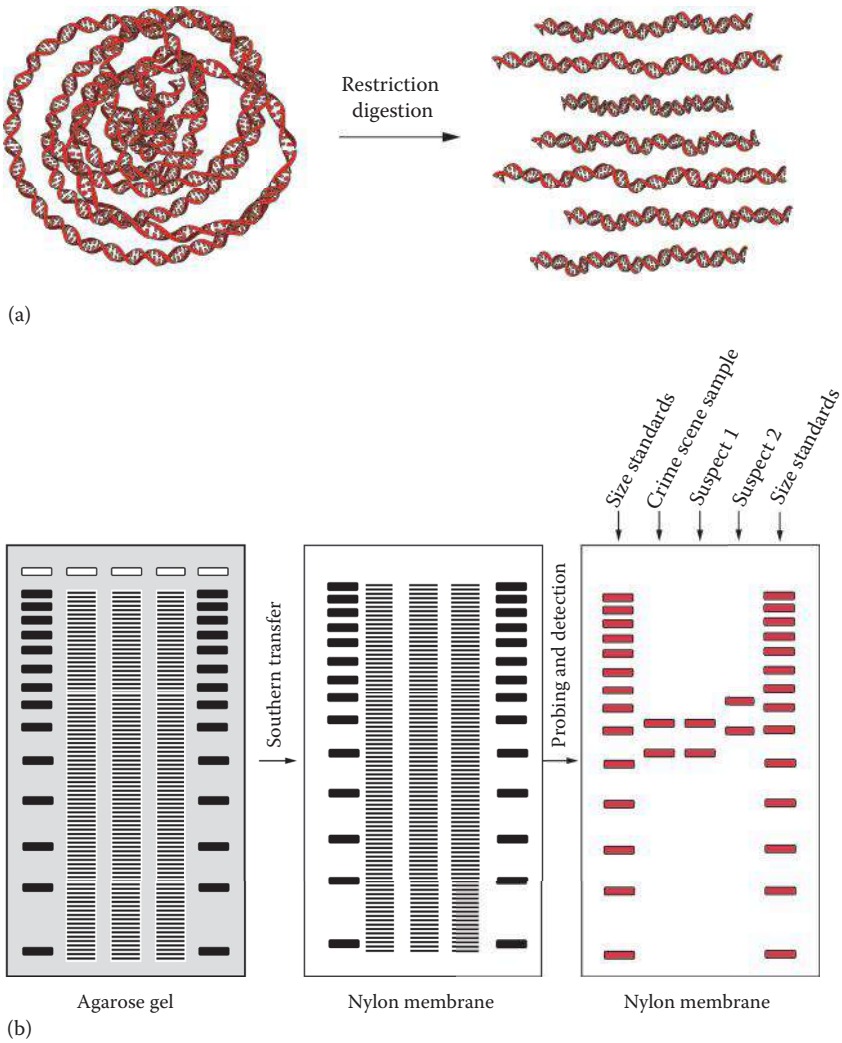

Figure 19.2 RFLP. (a) Restriction digestion generates restriction fragments with various lengths of genomic DNA. (b) Restriction fragments are separated by gel electrophoresis. DNA is transferred to a solid phase and probed. The signal is detected and the DNA fragment of interest can be observed. Band patterns of heterozygous loci of individuals are shown. (© Richard C. Li.)

restriction endonucleases is EC 3.1.21.4. The EC number is a numerical classification system of nomenclature based on the chemical reaction that is catalyzed by the enzyme. In contrast, type I (EC 3.1.21.3) and type III (EC 3.1.21.5) restriction endonucleases cleave at sites remote from their recognition site, which are not utilized for RFLP applications.

Restriction endonucleases are isolated from various bacteria. Each enzyme is named using a nomenclature system after the bacterium from which it was isolated. For example, in the restriction enzyme *Hae*III, H is from the genus name *Haemophilus*, ae is from the species name *aegypticus*, and III stands for the third endonuclease isolated from the *Haemophilus aegyptius* bacteria.

The restriction endonucleases play a role in protecting the bacteria from phage (bacterial virus) infections by using their endonucleases to destroy foreign DNA molecules. Bacterial DNA is usually methylated. In prokaryotes, the *Dam* methylase transfers a methyl group to the N6 position of the adenine in the sequence GmATC and the *Dcm* methylase transfers a methyl

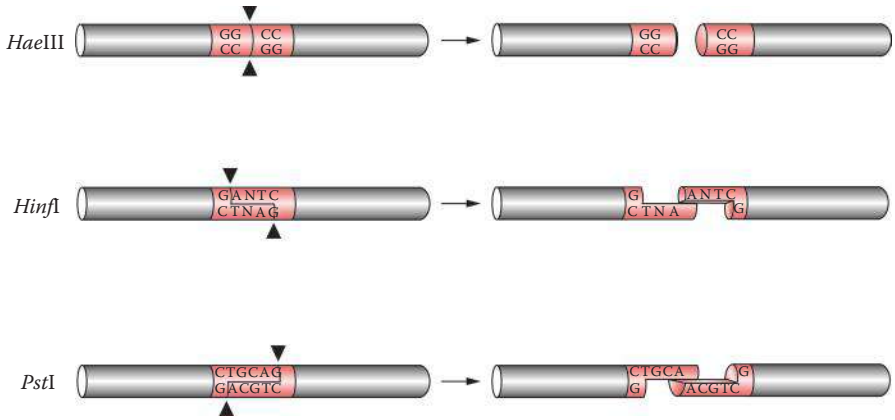

Figure 19.3 Restriction sites for *Hae*III, *Hinf*I, and *Pst*I. *Hae*III digestion produces a blunt end DNA fragment. *Hinf*I and *Pst*I digestions produce sticky ends. *N* represents any nucleotide. (© Richard C. Li.)

group to the C5 position of cytosine in the sequences CmCAGG and CmCTGG. The activities of restriction endonucleases can be influenced by DNA methylation. Many restriction endonucleases cannot cleave methylated DNA. Therefore, bacterial DNA is distinguishable from foreign DNA by the bacteria's restriction endonucleases. This phenomenon protects bacterial DNA from digestion by their own endonucleases. Note that because DNA methylation also occurs in the human genome (Chapter 11), it is important to choose the restriction endonucleases that are not affected by the methylation of human genomic DNA for RFLP analysis.

The type II restriction endonucleases were used in RFLP analysis for forensic DNA testing (Figure 19.3). In order to perform this analysis, the preferred restriction endonucleases for RFLP were those that cleave at the flanking regions of VNTR repeat units but not within the core repeat sequences of the VNTR. Most forensic laboratories used a single restriction endonuclease for a panel of VNTR loci because the DNA in evidence samples was often insufficient for performing multiple tests with different restriction endonucleases. For instance, *Hinf*I-based RFLP was commonly used in European forensic laboratories, and *Hae*III-based RFLP was used in North American and some European forensic laboratories. Other restriction endonuclease-based RFLPs such as *Pst*I were also used.

The use of these common restriction endonucleases allows the comparison of data of various laboratories. Several VNTR loci are suitable for RFLP analysis with these restriction endonucleases. For example, *Hae*III presents several advantages for forensic RFLP analysis. It recognizes a four-base sequence, 5′-GGCC-3′, and cleaves the DNA between the internal G and C residues of the recognition site (GG/CC). *Hinf*I recognizes a five-base restriction site, and *Pst*I recognizes a six-base restriction site. Hypothetically, four-base restriction sites are likely to occur more often than five- and six-base restriction sites in the human genome. Thus, *Hae*III restriction sites occur more frequently than *Hinf*I and *Pst*I sites. As a result, *Hae*III-cleaved DNA fragments are smaller than those of *Hinf*I and *Pst*I. The *Hae*III-generated VNTR allele sizes are easier to separate using conventional agarose gels, also called analytic gel electrophoresis. After electrophoresis, a smear of various sizes of DNA fragments can be observed. The analytic gel is then processed for Southern transfer. Moreover, the enzymatic activity of *Hae*III is not affected by the methylation of human genomic DNA. Its enzymatic activities also appear unaffected when a reaction proceeds under nonoptimal conditions. Additionally, low star activity is observed (Section 19.1.5.2.2).

19.1.2 Southern Transfer

Also known as *Southern blotting*, this technique was named after Sir Edwin Southern, who developed it in the United Kingdom in the mid-1970s. The method can be used to transfer

DNA from an agarose gel to a solid matrix so that it can be detected with a hybridization probe (Section 19.1.3). This method is still used today in many research laboratories. Prior to the transfer of DNA, the DNA in the gel must be denatured, under alkaline conditions such as treatment with sodium hydroxide, into single-stranded DNA. The single-stranded DNA in the gel is then transferred by capillary action to a solid matrix such as a piece of nylon membrane. The single-stranded DNA fragments transferred can be immobilized on a nylon membrane by an ultraviolet cross-linking process (Figure 19.4).

19.1.3 Hybridization with Probes

A hybridization probe of RFLP is a small segment of labeled DNA that is usually several hundred to a thousand bases in length containing the VNTR sequence. It is utilized to detect the presence of much longer target DNA sequences, in this case the VNTR sequences that are complementary to the nucleotide sequences of the probe. The probe is first denatured by heating or by exposure to alkaline conditions into single-stranded DNA. The hybridization process allows complementary pairing between the probe and the target sequence. Two types of probe techniques were developed for VNTR analysis: the multilocus probe and single-locus probe techniques.

19.1.3.1 Multilocus Probe Technique

The *multilocus probe* (*MLP*) technique can detect multiple VNTR loci simultaneously (Figure 19.5). Some VNTRs in the human genome share a short GC-rich core sequence of 10–15 bp. The MLP consists of this core sequence and hybridizes to multiple VNTRs that share these core sequences. As a result, the utilization of MLP produces a complex bar-code-like band pattern from alleles of multiple VNTR loci (Figure 19.6).

The MLP technique was pioneered by Sir Alec Jeffreys in 1984 at the University of Leicester in the United Kingdom and was called *DNA fingerprinting*. Because of its excellent discriminating power, the method was used for parentage testing in immigration disputes with great success.

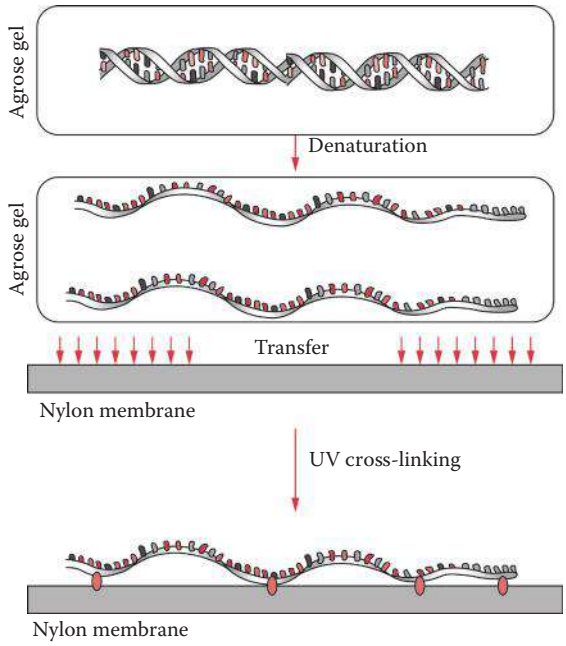

Figure 19.4 Southern blotting. DNA in agarose gel is denatured into single-stranded DNA and transferred to a solid-phase membrane where the single-stranded DNA is immobilized by ultraviolet cross-linking. (© Richard C. Li.)

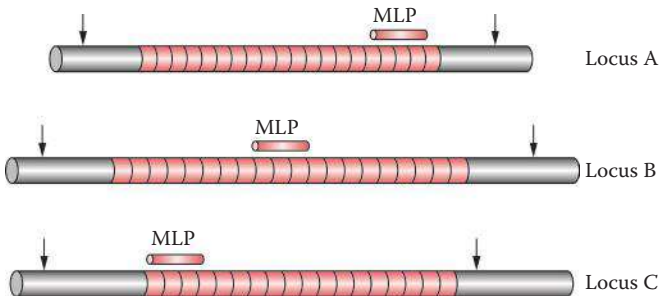

Figure 19.5 VNTR analysis using the MLP method. The technique can detect multiple VNTR loci simultaneously. Restriction sites are indicated by arrows. (© Richard C. Li.)

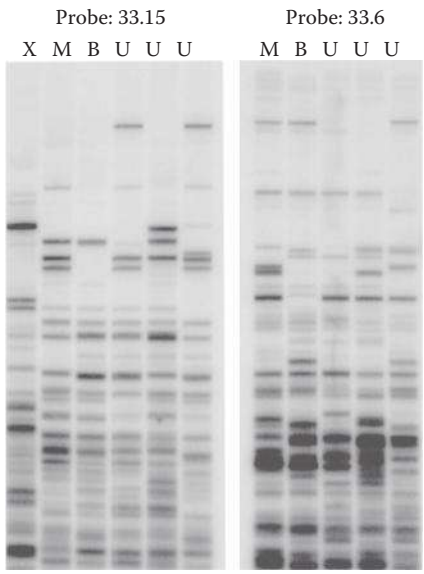

Figure 19.6 First application of DNA fingerprinting. The MLP method was used to analyze samples for an immigration case. M, mother; U, three undisputed children; B, male child in dispute; X, an unrelated individual. All bands in B can be traced back to M or U. (From Jeffreys, A.J., *Nat Med*, 11, 1035, 2005. With permission.)

However, one of the disadvantages of the MLP approach is that the interpretation of a mixed DNA sample from more than one individual is nearly impossible due to its complex DNA fingerprinting patterns. Therefore, MLP analysis was not widely utilized in forensic DNA laboratories.

19.1.3.2 Single-Locus Probe Technique

To resolve the disadvantages encountered in the MLP technique, probes that recognize the genomic DNA at the flanking regions of specific VNTR loci can be used. The probe only hybridizes to a single VNTR locus and the technique is called the *single-locus probe* (*SLP*), as depicted in Figure 19.7. SLP generates a simple pattern called a *DNA profile*, consisting of one band for a homozygote and two bands for a heterozygote per locus. In order to improve the discriminating power of the test, SLP analyses of different VNTR loci can be performed by using different probes sequentially with a single locus at one time. SLP can analyze mixed DNA samples from two or more contributors. The sizes of fragments can be estimated and converted into a numerical form suitable for databasing. Therefore, DNA profiles can be compared among different laboratories.

Figure 19.7 VNTR analysis using the SLP method. The technique can detect a single VNTR locus. Restriction sites are indicated by arrows. (© Richard C. Li.)

This technique led to the solving of a double murder case in Leicestershire in the 1980s. The case was the first to apply DNA evidence to a criminal investigation. DNA profiling identified the true perpetrator and also excluded an innocent suspect (Figure 19.8). In 1983 and 1986, two girls were raped and murdered. Crime scene evidence suggested that the two cases were committed by the same perpetrator. A young local man, Richard Buckland, was the suspect. However, the DNA evidence revealed that the semen samples from both crimes did not originate from Buckland. To solve the crimes, an investigation was carried out in which 5000 local men were asked to volunteer DNA samples for testing. Several months after the investigation, a witness tipped off police that a man named Pitchfork had paid someone for giving a blood sample as Pitchfork's. In 1987, Pitchfork was arrested. It was discovered that Pitchfork's DNA profile matched that of the crime scene evidence. He was sentenced to life imprisonment. This case demonstrated the great potential of DNA profiling in forensic investigations. Consequently, SLP became a common method in most forensic laboratories in the late 1980s–1990s.

19.1.4 Detection

To detect VNTR loci, a labeled SLP probe is hybridized to the target sequence of DNA, which has been immobilized on a solid matrix such as a piece of nylon membrane (see Southern

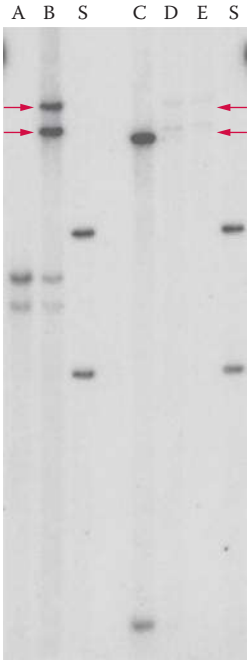

Figure 19.8 First application of DNA profiling in a criminal investigation using the SLP method. A, hair roots from the first victim; B, a mixture of semen and vaginal fluid from the first victim; C, blood from second victim; D, a vaginal swab from the second victim; E, a semen stain on clothing from the second victim; S, blood from the suspect. Alleles (arrows) are matched with the profiles of the two cases but not with the suspect profile. (From Jeffreys, A.J., *Nat Med*, 11, 1035, 2005. With permission.)

transfer, Section 19.1.2). Any unbound probes are washed away so that they do not interfere with the signal. Two types of detection systems are used for VNTR analysis. Radioisotope labeling, such as with a ^{32}P-labeled probe, can be used. The hybridized probe can be detected by exposing the membrane to a sheet of x-ray film to generate an autoradiograph. Alternatively, an enzyme-conjugated probe can also be used. Alkaline phosphatase (Section 9.2.2) is an example of an enzyme used in this type of probe. Its enzymatic activity can be detected with chemiluminescent substrates. A chemiluminescent signal can be detected by exposure to x-ray film as well.

In RFLP analysis, several loci are commonly analyzed sequentially using the same membrane. This approach has the advantage of not consuming additional DNA samples, which are often limited in forensic cases. When probing for multiple loci, multiple probes for each locus are sequentially hybridized and removed one at a time. Once the analysis of the first probe is completed, the probe is removed by a procedure called probe stripping, which is carried out under conditions such as high-temperature washing to denature the DNA strands (Chapter 17). The probe for the next locus to be analyzed is then hybridized to the same membrane, and the process is repeated for each probe to be tested.

Typically, a size standard is utilized on each gel. Band sizes can thus be estimated by comparison to these standards. However, the VNTR alleles that differ by only one or two repeat units are usually not distinguishable. For this reason, genotypes can be determined by bins but not discrete alleles. A *bin* is a range of DNA fragments that differ by only a few repeat units. A sample with a known VNTR genotype is also utilized on each gel as a positive control where historically a genomic DNA sample from cell line K562 (a human erythroleukemic cell line) was used for the positive control. DNA samples to be compared can be loaded side by side on the same gel. As a result, the patterns of VNTR fragments can be compared from sample to sample. The following possible conclusions can be made. If the VNTR fragments are at corresponding positions (profiles match), they are considered to be a match (inclusion). Chapter 25 evaluates and discusses the strengths of the results. If the profiles are different, the two DNA samples are considered to have come from different origins (exclusion).

19.1.5 Factors Affecting RFLP Results

The accuracy of VNTR profiling results can be affected by certain factors such as sample conditions, genetic mutations, and experimental artifacts appearing during the procedure. Consequently, these factors can impact data interpretation and are explained in the following sections.

19.1.5.1 DNA Degradation

RFLP analysis requires the genomic DNA to be intact. DNA degradation results in damage such as creating nicks and breaks in the strand. The more severe the degradation, the smaller the average size of the DNA fragments. When the average size of DNA fragments becomes too small, the allele may not be detected. Many VNTR tandem arrays can span several kilobases in length. In theory, large alleles are more likely to be affected by degradation than smaller alleles at a different locus.

A two-banded heterozygous profile can be observed as a one-banded homozygous profile if the larger band is not detected due to degradation. This artifact could lead to a false determination of exclusion. However, DNA degradation can be detected prior to conducting RFLP by the use of agarose gel electrophoresis, also known as a *yield gel*, used for evaluating the yield and integrity of the isolated genomic DNA. High-molecular-weight genomic DNA bands are usually observed for a typical genomic DNA sample. In contrast, a smear of low-molecular-weight DNA bands can be observed if DNA is degraded. The sizes of the DNA can be estimated by comparison to a size standard run on the same gel. Additionally, the yield of DNA can be estimated by comparing the intensity of the size standards with a known quantity of DNA.

19.1.5.2 Restriction Digestion–Related Artifacts

19.1.5.2.1 Partial Restriction Digestion

Complete restriction digestion should be achieved for RFLP analysis. If partial digestion occurs, the partially cleaved DNA strands are longer than the cleaved fragments (Figure 19.9). Thus, partial digestion results in a mixture of fragments with correct sizes and slightly larger fragments. Under these conditions, a larger uncleaved band, usually lower in intensity than the true bands, can be observed. The multibanded pattern due to the partial digestion can be observed at multiple loci analyzed in the same nylon membrane.

Detection of more than two bands in an RFLP profile may lead to a false interpretation and be incorrectly concluded to be a mixture. However, partial digestion can be detected after restriction digestion. DNA cleavage by restriction endonuclease digestion can be examined using agarose gel electrophoresis. A small portion of a sample can be analyzed. After separation using electrophoresis, a smear of various sizes of cleaved DNA fragments can usually be observed if restriction digestion is completed. Conversely, high-molecular-weight genomic DNA can still be observed from partially digested DNA samples. Additionally, comparisons can be made between the sample and uncleaved and completely cleaved standard samples of DNA. This quick assay is also called a *test gel*, used to determine if the DNA was cleaved to completion.

If partial digestion occurs, procedures such as additional purification of the DNA sample can be carried out. Additionally, optimal amounts of DNA, restriction enzymes, buffer, and proper incubation conditions should be used in achieving complete digestion.

19.1.5.2.2 Star Activity

Star activity refers to a deviation of the specificity of a cleavage site of a restriction endonuclease under certain conditions, such as a high concentration ratio of enzyme to DNA, the use of non-optimal buffers for restriction digestion, prolonged digestion time, the substitution of Mg^{2+} with other divalent cations, and the presence of organic solvents such as ethanol and concentrated glycerol. For instance, *Hae*III cleaves at the GGCC DNA sequence. When star activity occurs, the enzymatic specificity is reduced and cleaves at a sequence slightly different from GGCC. If the start site is presented at an internal location of a VNTR locus, the enzyme would cleave the GGCC sequences and additionally cleave at the internal star site (Figure 19.10). This would result in an additional band smaller than the true alleles, although the intensity of this band is usually not the same as other bands. Thus, a multiband pattern is observed. However, the star

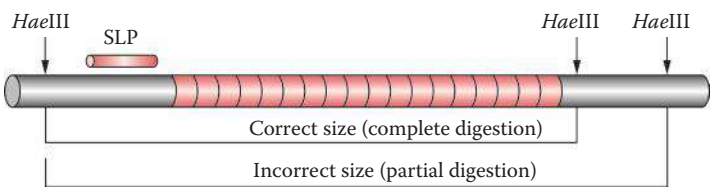

Figure 19.9 Effects of partial restriction digestion on the RFLP profile. Only the restriction fragments detectable by the probe are shown. *Hae*III restriction sites are indicated by arrows. (© Richard C. Li.)

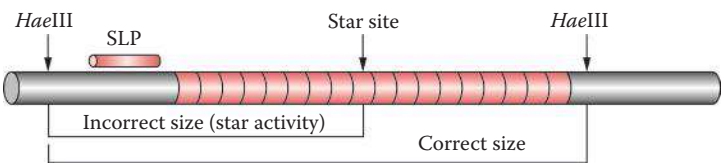

Figure 19.10 Star effects on RFLP profiles. Only the restriction fragments detectable by the probe are shown. *Hae*III restriction sites and star sites are indicated by arrows. (© Richard C. Li.)

site presented externally to a VNTR locus cannot be detected by the probe and does not affect the profiling results. Star activity can usually be avoided by carrying out restriction digestion reactions under conditions recommended by their manufacturers.

19.1.5.2.3 Point Mutations

A *point mutation* is caused by the substitution, deletion, or insertion of a single nucleotide. Point mutations at a restriction site within flanking regions may abolish the site, and the result is a band slightly larger than the true allele (Figure 19.11). The point mutation may also be present internally in a VNTR sequence. If such a point mutation creates a restriction enzyme site, the enzyme will cleave at the regular site and at the mutation site and yield two smaller bands. If the created restriction site is located internal to the probe binding region, both bands will be detected for that allele. These rare mutations obey Mendelian inheritance.

19.1.5.3 Electrophoresis and Blotting Artifacts

19.1.5.3.1 Partial Stripping

If more than one VNTR locus is analyzed sequentially using the same membrane, a probe must be removed by the stripping process before the application of the next probe. Any probe remaining on the membrane due to partial stripping may generate additional bands when the next probe is analyzed. However, bands due to partial stripping are usually faint and have the same electrophoretic mobility as the previous autoradiograph.

19.1.5.3.2 Separation Resolution Limits and Band Shifting

Agarose gel electrophoresis cannot resolve restriction fragments that differ by one or a few repeat units, especially for high-molecular-weight fragments. These bands may not be

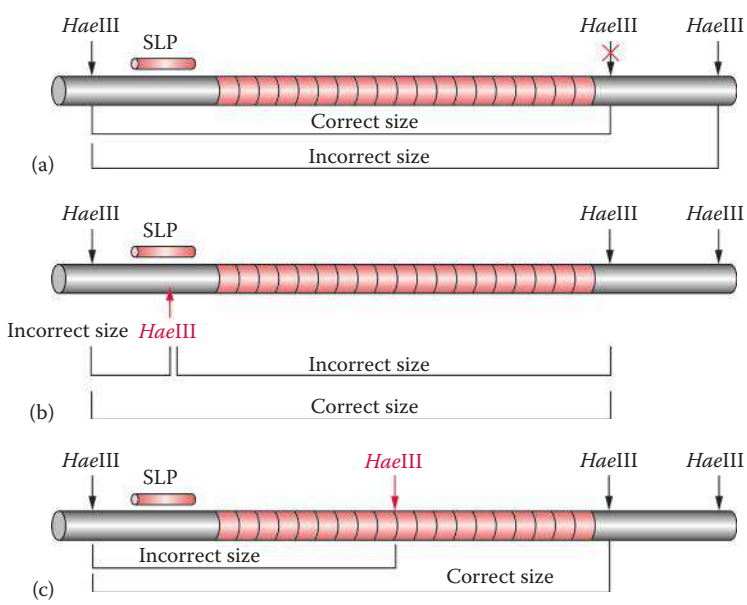

Figure 19.11 Effects of point mutations on the RFLP profile. Only the restriction fragments detectable by the probe are shown. *Hae*III restriction sites are indicated by arrows. (a) Point mutation (in red) abolishes the *Hae*III restriction site. (b) A point mutation (in red) creates an internal *Hae*III restriction site residing within the probe-binding region. (c) A point mutation (in red) creates an internal *Hae*III restriction site residing outside the probe-binding region. (© Richard C. Li.)

separated and will appear as a single band. This may lead to a false interpretation as a homozygous profile. Additionally, minor variations in the electrophoretic mobility of DNA fragments, known as band shifting, can cause two samples from the same individual to appear different.

19.1.5.3.3 Bands Running off Gel

The commonly used VNTR loci generate bands from hundreds of base pairs to 20 kb in length. The small bands have higher electrophoretic mobility and may run off the front edge of a gel during electrophoresis and fail to be detected. This phenomenon may also lead to a false interpretation as a homozygous profile. To prevent these DNA fragments from running off the gel into the buffer tank, a longer gel can be used. Alternatively, the electrophoresis can be stopped before the dye front, the furthest extent that dyes migrate, reaches the front edge of the gel.

19.2 Amplified Fragment Length Polymorphism

The RFLP analysis of VNTR profiling does not perform well for degraded or limited quantities of DNA from crime scene samples. For these reasons, an improved VNTR method was developed. Some VNTR loci have relatively short alleles (<1 kb). These loci are suitable for PCR amplification. This technique is called *amplified fragment length polymorphism* (*AFLP*). One locus, D1S80, was used by forensic DNA laboratories for AFLP analysis. Fragments in the range of 14–42 repeat units (16 bp per repeat) were amplified using the AFLP method (Figure 19.12). The amplified DNA fragments were commonly separated according to size using polyacrylamide gel electrophoresis and detected using a silver stain (Figure 19.13).

D1S80 loci are detected as discrete alleles and thus can be compared directly to an allelic ladder (a collection of common alleles used as a standard) on the same gel. This technique represented an improvement over the RFLP system. RFLP allele sizing cannot be performed with precision and the resolution limits of agarose gel electrophoresis are much lower than those of the polyacrylamide gels.

The AFLP technique requires less DNA than the RFLP method and performs better for degraded samples. The AFLP method at the D1S80 locus can be analyzed in a multiplex fashion with an amelogenin locus (Chapter 21). The amelogenin gene is used for forensic sex-typing applications. Typing the amelogenin gene enables the determination of the sex of the contributor of a biological sample.

Due to the wide variation in allele sizes at the D1S80 locus, preferential amplification may be observed. Under certain conditions, the larger alleles may not be as consistently amplified as the small alleles, which may cause lower signal intensity of the larger allele. Additionally, only one locus is analyzed in this system. Furthermore, the D1S80 locus contains two alleles that are very common in some populations. Thus, the discriminating power is reduced compared to RFLP. D1S80 was gradually replaced by multiplex STR systems in the late 1990s.

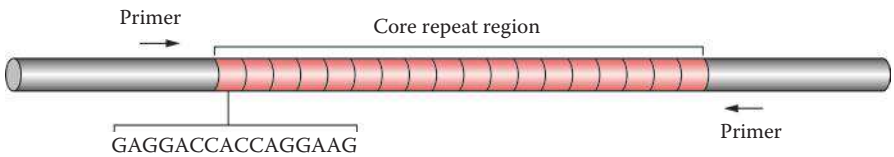

Figure 19.12 VNTR locus D1S80 (chromosome 1p). Each repeat unit is 16 bp long. PCR primers are indicated to amplify the core repeat region.

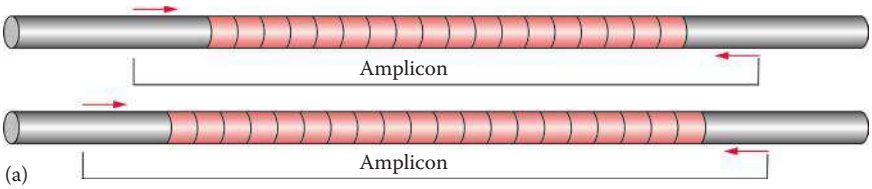

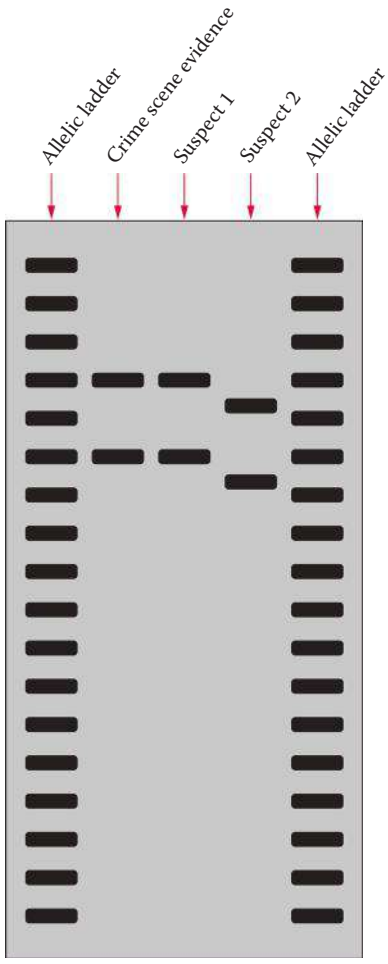

Figure 19.13 AFLP analysis. (a) Heterozygous D1S80 loci amplified by PCR. PCR primers are indicated as *arrows*. (b) Silver-stained polyacrylamide gel showing D1S80 amplicons along with allelic ladders. (© Richard C. Li.)

Bibliography

RFLP

Balazs, I., et al., Human population genetic studies of five hypervariable DNA loci. *Am J Hum Genet*, 1989, **44**(2): 182–190.

Benzinger, E.A., et al., Time course and inhibitors of *Hae*III digestion in the forensic laboratory. *Appl Theor Electrophor*, 1995, **4**(4): 179–188.

Benzinger, E.A., et al., Products of partial digestion with *Hae*III. Part 1. Characterization, casework experience and confirmation of the theory of three-, four- and five-banded RFLP pattern origins using partial digestion. *J Forensic Sci*, 1997, **42**(5): 850–863.

20

Autosomal Short Tandem Repeat Profiling

A *short tandem repeat* (STR) is a region of genomic DNA containing an array of short repeating sequences. STRs are also called *microsatellites* or *simple sequence repeats*. A STR repeat unit can be several base pairs (bp) in length. Arrays range from several to approximately a hundred *repeat units*, which are the component of repetition. The number of STR repeat units varies among individuals. The most commonly used STR loci are 100–500 bp in length, which are shorter than the smallest variable number tandem repeats (VNTRs) (approximately 1000 bp). Thus, STR loci have many advantages compared to VNTR loci:

- STR loci can be amplified by PCR.
- STR profiling can be carried out for degraded DNA samples.
- Preferential amplification is reduced at STR loci.
- The resolution of electrophoretic separation of STR fragments is superior.
- STR loci are suitable for multiplex amplification.

Additionally, STR profiling, as with VNTR profiling, is suitable for the interpretation of mixed DNA profiles from multiple individuals. Thus, STR loci are better candidates for forensic DNA testing than VNTR loci. This chapter will discuss autosomal STR profiling. Male-specific Y chromosomal STR will be discussed in Chapter 21.

20.1 Characteristics of STR Loci

More than 10^5 STRs exist in the human genome. Many STRs have been characterized and used in various types of studies such as genetic mapping and linkage analysis. Some STRs have been characterized specifically for forensic DNA profiling.

20.1.1 Core Repeat and Flanking Regions

The *core repeat region* of each STR locus contains tandemly repeated sequences. The designation of genotypes for human identification is based on the number of tandem repeat units at a STR locus, which varies among human individuals (Figure 20.1). The *flanking regions* surrounding the core repeat region are also needed for STR analysis. PCR primers complementary to these flanking regions are used, allowing the core repeat regions to be amplified.

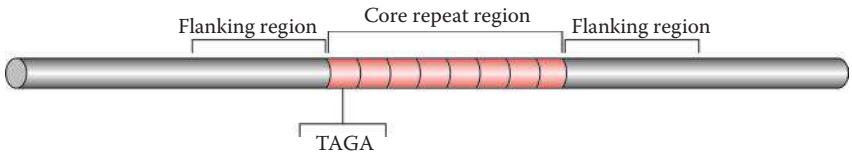

Figure 20.1 Core repeat and flanking regions of CSF1PO STR locus. It consists of eight repeating units of tetrameric nucleotides (TAGA); thus, it is designated as allele 8. (© Richard C. Li.)

20.1.2 Repeat Unit Length

Repeat unit length is the number of nucleotides in a single repeat unit. Dimeric, trimeric, tetrameric, pentameric, and hexameric repeat units appear in the human genome. For example, dimeric and trimeric repeats are very abundant, but they are not usually used for forensic applications. High frequencies of stutter peaks (Section 20.4.2.1) that interfere with genotype interpretation are observed when dimeric and trimeric repeats are amplified. On the contrary, only a few thousand pentameric repeats and a few hundred hexameric repeats exist in the human genome. The pentameric and hexameric repeats are very polymorphic. Only a few pentameric and hexameric repeats are used for forensic applications because they are less abundant in the human genome. The human genome has at least 10^4 tetrameric repeats representing approximately 9% of the total STRs. The STRs with tetrameric repeats are very polymorphic. When they are amplified by PCR, this category of STRs exhibits fewer frequencies of stutter peaks than STRs with dimeric and trimeric repeats. Therefore, the most commonly used STR loci for forensic DNA profiling are the STRs with tetrameric repeats.

20.1.3 Repeat Unit Sequences

STR loci compatible for forensic use can be divided into several classes based on their repeat unit sequences. Figure 20.2 shows representative examples of core repeat sequences. *Simple repeats* consist of tandem repeats with identical repeat unit sequences (Figure 20.2a). Allele designation is based on the number of repeat units in the core repeat region. For example, a D5S818 allele consisting of ten repeating units of the tetrameric nucleotide sequence AGAT is designated as allele 10. *Compound repeats* consist of more than one type of simple repeat (Figure 20.2b). *Complex repeats* contain several clusters of different tandem repeats with intervening sequences (Figure 20.2c).

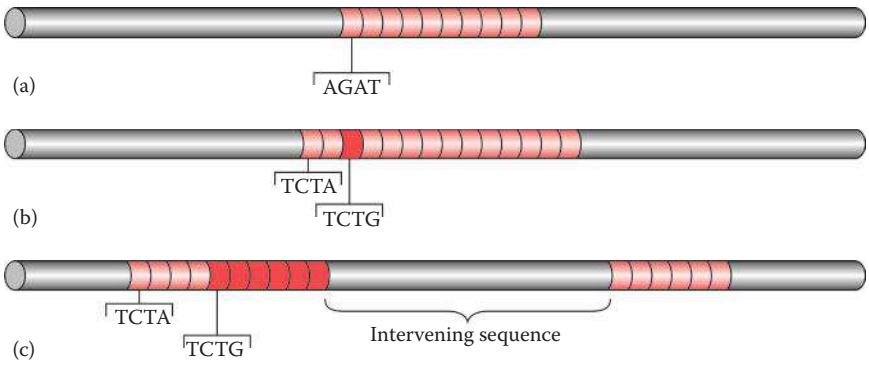

Figure 20.2 Examples of core repeat sequences. (a) A simple repeat in which D5S818 $[AGAT]_{10}$ is designated as allele 10, consisting of 10 repeating units of the tetrameric nucleotides, AGAT. (b) Compound repeats. Allele 14 of D8S1179 consists of two types of repeating units: $[TCTA]_2$, $[TCTG]_1$, and $[TCTA]_{11}$. (c) Complex repeats. Allele 24 of D21S11 contains several clusters of different tandem repeats, $[TCTA]_4$, $[TCTG]_6$, and $[TCTA]_6$, with a 43 bp intervening sequence: $[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCATA. (© Richard C. Li.)

Nonconsensus alleles with partial repeat units also appear in the population. These nonconsensus alleles, also known as *microvariants*, differ from common alleles by one or more nucleotides. They are designated by the number of consensus repeats, followed by a decimal point and the number of nucleotides of the partial repeat, for example, the TH01 allele 9.3 is 1 nucleotide shorter than allele 10.

Another type of nonconsensus allele can result from a limitation of STR analysis. These alleles have the same number of tandem repeats as common alleles but contain different sequences. These microvariants cannot be distinguished by STR profiling because their length is identical to the lengths of common alleles.

20.2 STR Loci Commonly Used for Forensic DNA Profiling

In the early 1990s, STR loci were initially utilized for genetic studies and were later applied to forensic DNA profiling. The first STR multiplex system, known as the quadruplex, was developed by Forensic Science Services in the United Kingdom. It consisted of four STR loci (F13A1, FES, TH01, and VWA) with a *population match probability* (P_m) of 10^{-4} (Figure 20.3). P_m measures the discriminating power of an STR locus used for forensic DNA analysis. The lower the P_m (i.e., the higher the discriminating power), the less likely a match will occur between two randomly chosen profiles from different individuals (Chapter 25). In 1995, the first national DNA database was established in the United Kingdom. It contained six STR loci, also known as the *second-generation multiplex* (SGM), consisting of D8S1179, D18S51, D21S11, FGA, TH01, and VWA, with a P_m value of 10^{-7}. The SGM system also included the amelogenin locus (Section 21.3.1) for determining the sexes of DNA contributors. Subsequently, four additional loci were added to SGM with a P_m as low as 10^{-13} (SGM Plus).

To allow for international data exchange, the European DNA Profiling Group (EDNAP) recommended the use of TH01 and VWA loci for all participating European laboratories in 1996. In 1998, the *European Standard Set* (ESS) of loci was established and included TH01, VWA, FGA, and D21S11 for forensic use in Europe. Thereafter, D3S1358, D8S1179, and D18S51 loci were added to the ESS. Other loci are used as well, such as SE33, which is used in Germany's database. In 1998, the US Federal Bureau of Investigation established the *Combined DNA Index System* (CODIS). It contains 13 core STR loci plus the amelogenin sex-typing locus with a P_m

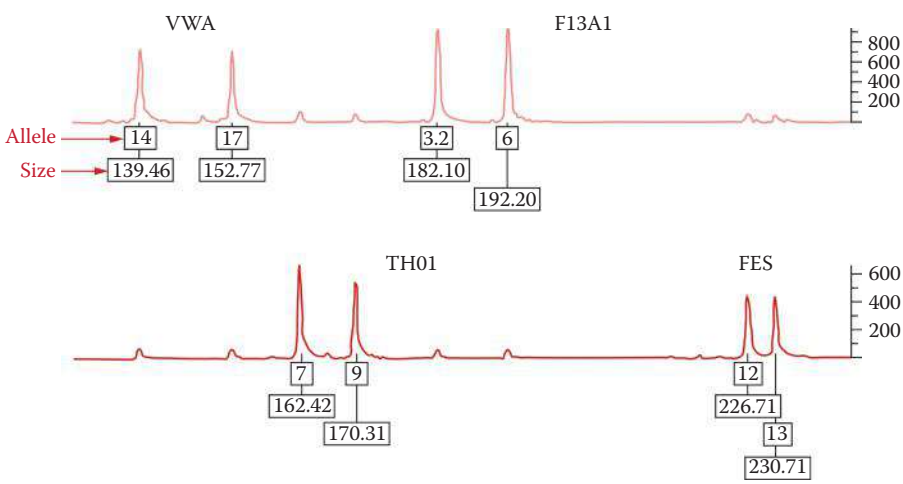

Figure 20.3 DNA profile obtained using the first STR multiplex system: the quadruplex. F13A1, FES, TH01, and VWA loci are shown. (© Richard C. Li.)

Table 20.1 Common STR Loci

Locus Symbol	Repeat Motif	Repeat Category	Cytogenetic Map Location on Chromosome	Distance from pter (Mb)	Structural Gene
CSFIPO	TAGA	Simple	5q33.1	149.4	Intron 6 of c-fms protooncogene
FGA	[CTTT] [CCTT]	Compound	4q31.3	155.5	Intron 3 of fibrinogen α chain gene
TH01	AATG	Simple	11p15.5	2.2	Intron 1 of tyrosine hydroxylase gene
TPOX	TGAA	Simple	2p25.3	1.5	Intron 10 of thyroid peroxidase gene
SE33/ ACTBP2	AAAG	Complex	6q14	89	5′ Flanking sequence of β-actin-related pseudogene 2 gene
VWA	[TCTG] [TCTA]	Compound	12p13.31	6.1	Intron 40 of von Willebrand factor gene
D1S1656	[TAGA] [TAGC]	Compound	1q42.2	230.9	Anonymous
D2S441	[TCTA] [TCAA]	Compound	2p14	68.2	Anonymous
D2S1338	[TGCC] [TTCC]	Compound	2q35	218.9	Anonymous
D3S1358	[TCTG] [TCTA]	Compound	3p21.31	45.6	Anonymous
D5S818	AGAT	Simple	5q23.2	123.1	Anonymous
D7S820	GATA	Simple	7q21.11	83.8	Anonymous
D8S1179	[TCTA] [TCTG]	Compound	8q24.13	125.9	Anonymous
D10S1248	GGAA	Simple	10q26	–[a]	Anonymous
D12S391	[AGAT] [AGAC]	Compound	12p13.2	12.5	Anonymous
D13S317	TATC	Simple	13q31.1	82.7	Anonymous
D16S539	GATA	Simple	16q24.1	86.4	Anonymous
D18S51	AGAA	Simple	18q21.33	60.9	Anonymous
D19S433	[AAGG] [TAGG]	Compound	19q12	30.4	Anonymous
D21S11	[TCTA] [TCTG]	Complex	21q21.1	20.6	Anonymous
D22S1045	ATT	Simple	22q12.3	37.5	Anonymous

Source: Ensembl Homo sapiens version 75.37 (GRCh37). Mb, megabase.
[a] Locus is not mapped to the assembly in the current Ensembl database.

of 10^{-15}. As the database grows rapidly, the chances of finding incidental matches among DNA profiles are increasing. Recently, additional loci have been added to CODIS and ESS core loci to reduce the likelihood of adventitious matches, as well as to facilitate international data sharing among law enforcement agencies and to improve the discrimination power of forensic STR analysis.

Commonly used STR loci characterized for forensic DNA profiling are summarized in Tables 20.1 and 20.2. To achieve low P_m in forensic STR profiling, desired STR loci should possess certain characteristics as described below. First, the alleles of STR loci selected should be highly variable among individuals. Second, if more than one locus is selected, the loci should not be linked to each other or inherited together (Section 21.2). The STR loci utilized are usually located at different chromosomes to ensure that they are not linked. However, loci that are far

TABLE 20.2 Core STR Loci

Locus	SGM	SGM Plus	ESS	ESS-Extended	CODIS	CODIS-Extended
Amel	☑	☑	☑	☑	☑	☑
CSFIPO					☑	☑
D1S1656				☑		☑
D2S441				☑		☑
D2S1338		☑				☑
D3S1358		☑	☑	☑	☑	☑
D5S818					☑	☑
D7S820					☑	☑
D8S1179	☑	☑	☑	☑	☑	☑
D10S1248				☑		☑
D12S391				☑		☑
D13S317					☑	☑
D16S539		☑			☑	☑
D18S51	☑	☑	☑	☑	☑	☑
D19S433		☑				☑
D21S11	☑	☑	☑	☑	☑	☑
D22S1045				☑		2
DYS391						☑
FGA	☑	☑	☑	☑	☑	☑
TH01	☑	☑	☑	☑	☑	☑
TPOX					☑	1
VWA	☑	☑	☑	☑	☑	☑
SE33						3

Source: Hares, D.R., *Forensic Sci Int Genet*, 6, e135, 2012; Gill, P., et al., *Forensic Sci Int*, 156, 242–244, 2006.

☑, minimum required loci; 1–3, recommended loci in ranked order of preference.

enough apart on the same chromosome can still be used (Figures 20.4 and 20.5), since they are not linked. Additionally, STR loci with fewer amplification artifacts such as stutter products are desired. Stutters can complicate the interpretation of profiles derived from a mixed DNA sample from more than one contributor. Moreover, STR loci with short amplicon (amplified product) lengths are preferred for multiplex STR analysis and the testing of degraded DNA samples.

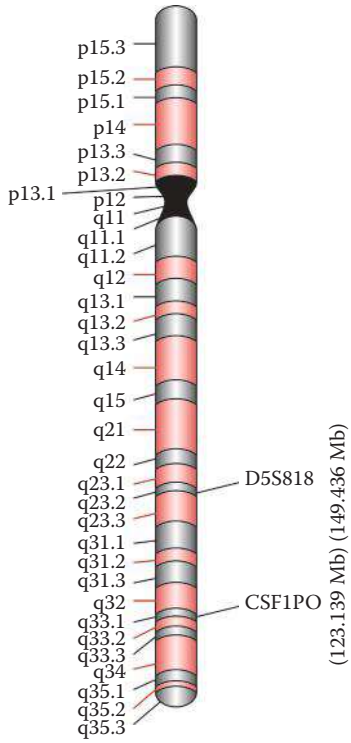

Figure 20.4 Cytogenetic map showing the locations of STR markers on chromosome 5. CSF1PO and D5S818 are separated by 26 Mb (megabases). (© Richard C. Li.)

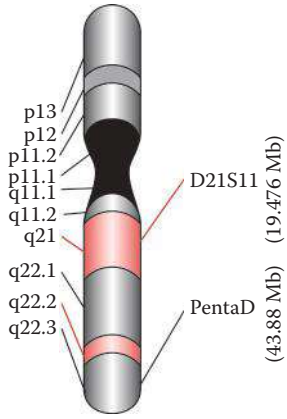

Figure 20.5 Cytogenetic map showing the locations of STR markers on chromosome 21. D21S11 and PentaD are separated by 24 Mb (megabases). (© Richard C. Li.)

20.3 Forensic STR Analysis

STR loci are amplified using fluorescent dye-labeled primers. A multiplex STR system utilizes multiple fluorescent dyes to label each amplicon. The amplicons are separated via electrophoresis. The different fluorescent dye colors are resolved by the detector, and the signals corresponding to each DNA fragment are identified using specialized computer software. The data collection process generates an *electropherogram* that shows a profile of peaks corresponding to each DNA fragment. The positions of these peaks represent the electrophoretic mobility of the DNA fragments. A small fragment, migrating faster than a large one, peaks earlier in the electropherogram than a longer fragment. The DNA fragments are sized by comparison to an internal size standard (Figure 20.6; Chapter 8). Figure 20.7 summarizes the work flow of a forensic STR analysis.

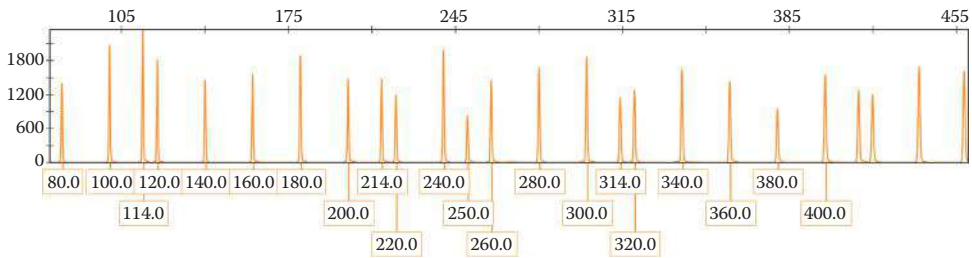

Figure 20.6 Electropherogram of GeneScan™ 500 size standard (Applied Biosystems). RFU represents relative fluorescence unit. (© Richard C. Li.)

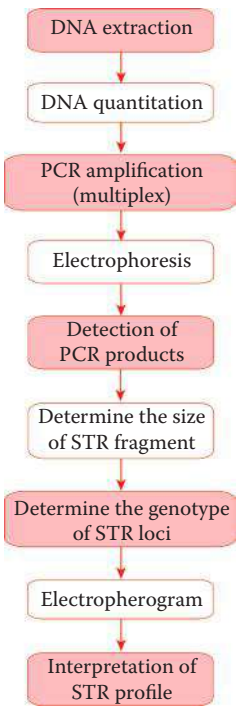

Figure 20.7 Capillary electrophoresis separation of amplified STR products. Fluorescent dye-labeled amplification products are separated and subsequently detected. Various fluorescent dye colors are resolved by the detector. The peaks corresponding to each DNA fragment are identified. (© Richard C. Li.)

The area or amplitude of the peak, expressed as relative fluorescence units (RFU) (Section 9.3.1.3), reflects the fluorescent signal intensity. The RFU value of peak height is proportional to the amount of DNA amplicons being analyzed. When the RFU value is very low, it is difficult to distinguish a signal from background noise. The manufacturer of commonly used instruments for forensic DNA analysis recommends 150 RFU as the threshold of detection. Peaks below 150 RFU should be interpreted with caution. Some forensic laboratories use lower thresholds, as low as 50 RFU, based on their own validation studies. In contrast, when the RFU value is too high, it saturates the sensitivity of an instrument as well as causing artifacts such as pull-up signals (Section 20.4.3.1). The maximum RFU allowed is usually 6000 RFU.

20.3.1 Determining the Genotypes of STR Fragments

As noted earlier, electropherograms are usually plotted as fluorescent signal intensity (in RFU) versus the sizes of the DNA fragments. The data in an electropherogram can then be converted into a genotype. The genotype for a specific STR locus is defined as the number of repeat units of the allele. STR genotype data generated from different laboratories can be compared easily and are suitable for databasing.

The genotype is determined by using an allelic ladder, which is important to achieve accurate genotype profiling. An allelic ladder is a collection of synthetic fragments corresponding to common alleles observed in the human population for a given set of STR loci (Figures 20.8 and 20.9). The ladders are compared to data obtained from an electropherogram of a questioned sample to determine the genotype. Thus, each allele in a ladder must be resolved properly in order to determine correct STR alleles for a sample. The sizes of DNA fragments of a sample are correlated to sizes of fragments for each allele in an allelic ladder in order to determine the allele designation (genotype) of a questioned sample (Figures 20.10 and 20.11). If a rare allele fails to match alleles within an allelic ladder, it is considered an *off-ladder allele*. If an off-ladder allele is present, the sample should be reanalyzed. The presence of a rare allele can be confirmed by repeating the electrophoresis process based on the characteristic electrophoretic mobility of the rare allele.

20.3.2 Interpretation of STR Profiling Results

General guidelines for the interpretation and the reporting of STR profile results were set by the Scientific Working Group on DNA Analysis Methods (SWGDAM) and the DNA Commission of the International Society of Forensic Genetics (ISFG). Typically, conclusions are categorized as inclusion, exclusion, or inconclusive result.

20.3.2.1 Inclusion (Match)

Peaks of compared STR loci, such as those between the profiles of suspect and crime scene evidence or victim and crime scene evidence, show identical genotypes. The strength of this conclusion can be evaluated via statistical analysis and is usually cited in the case report (Chapter 25).

20.3.2.2 Exclusion

The genotypes of two or more samples differ, and the profile of the sample is determined to be an exclusion, meaning that the profiles originated from different sources.

20.3.2.3 Inconclusive Result

The data do not support a conclusion of inclusion or exclusion. In other words, insufficient information is available to reach a conclusion.

20.4 Factors Affecting Genotyping Results

A number of genetics-, amplification-, and electrophoresis-related factors may affect the accuracy of genotypic profiles.

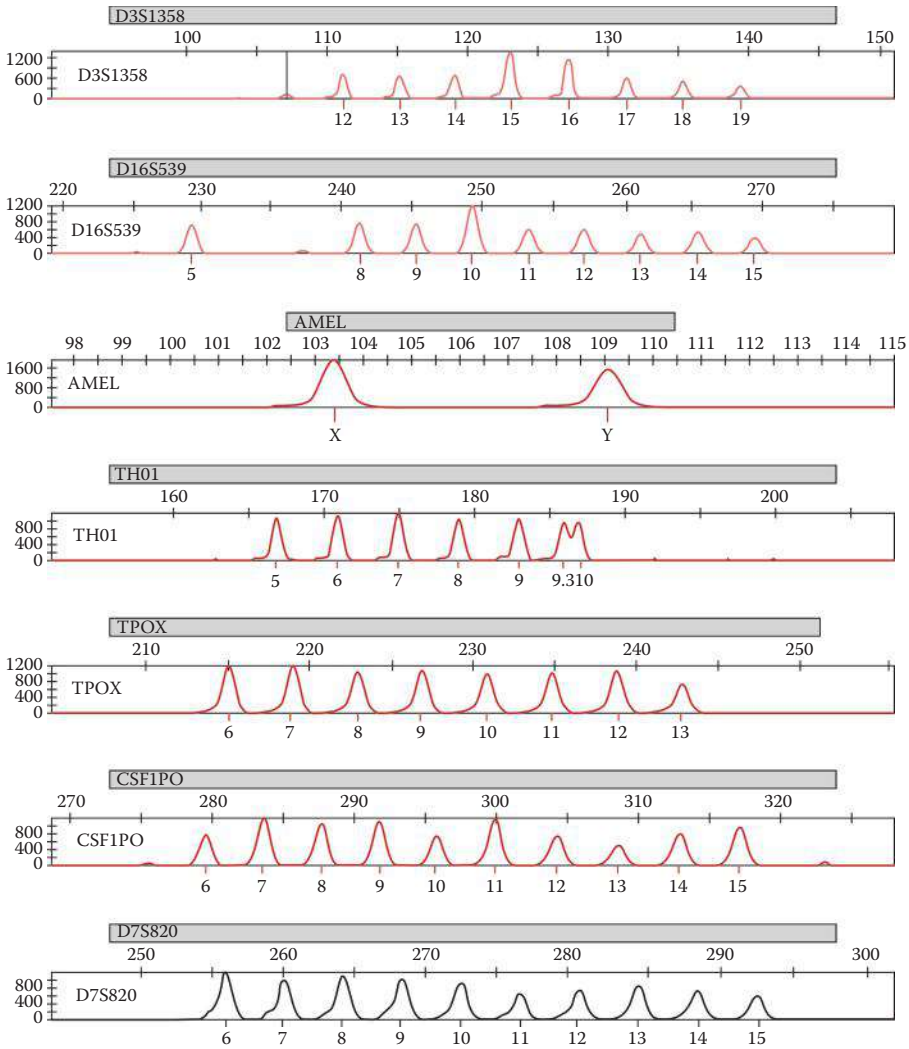

Figure 20.8 Electropherogram of the allelic ladder of the AmpFlSTR® COfiler® PCR Amplification Kit (Applied Biosystems). AMEL, CSF1PO, D3S1358, D7S820, D16S539, TH01, and TPOX loci are shown. (© Richard C. Li.)

20.4.1 Mutations

STR loci with low mutation frequencies are desired, in particular, for human identification in mass disasters and for missing person and paternity cases. However, STR mutations do occur, which can affect the profiling results.

20.4.1.1 Mutations at STR Core Repeat Regions

Mutations, usually resulting in a gain or a loss of a single repeat unit, are observed at STR loci. If a mutation occurs in the germ cells (cells that form gametes), the mutant allele will be transmitted to and be present in all cell types of the progeny. This type of inheritable mutation in germ cell lineage is called a *germ-line mutation*. The frequency of germ-line mutation can be measured by the *mutation rate*, expressed as the number of mutations per generation (germ-line transmission). The average mutation rate of commonly used STR loci is about 10^{-4} mutations per germ-line transmission. However, the mutation rate may vary among different STR loci.

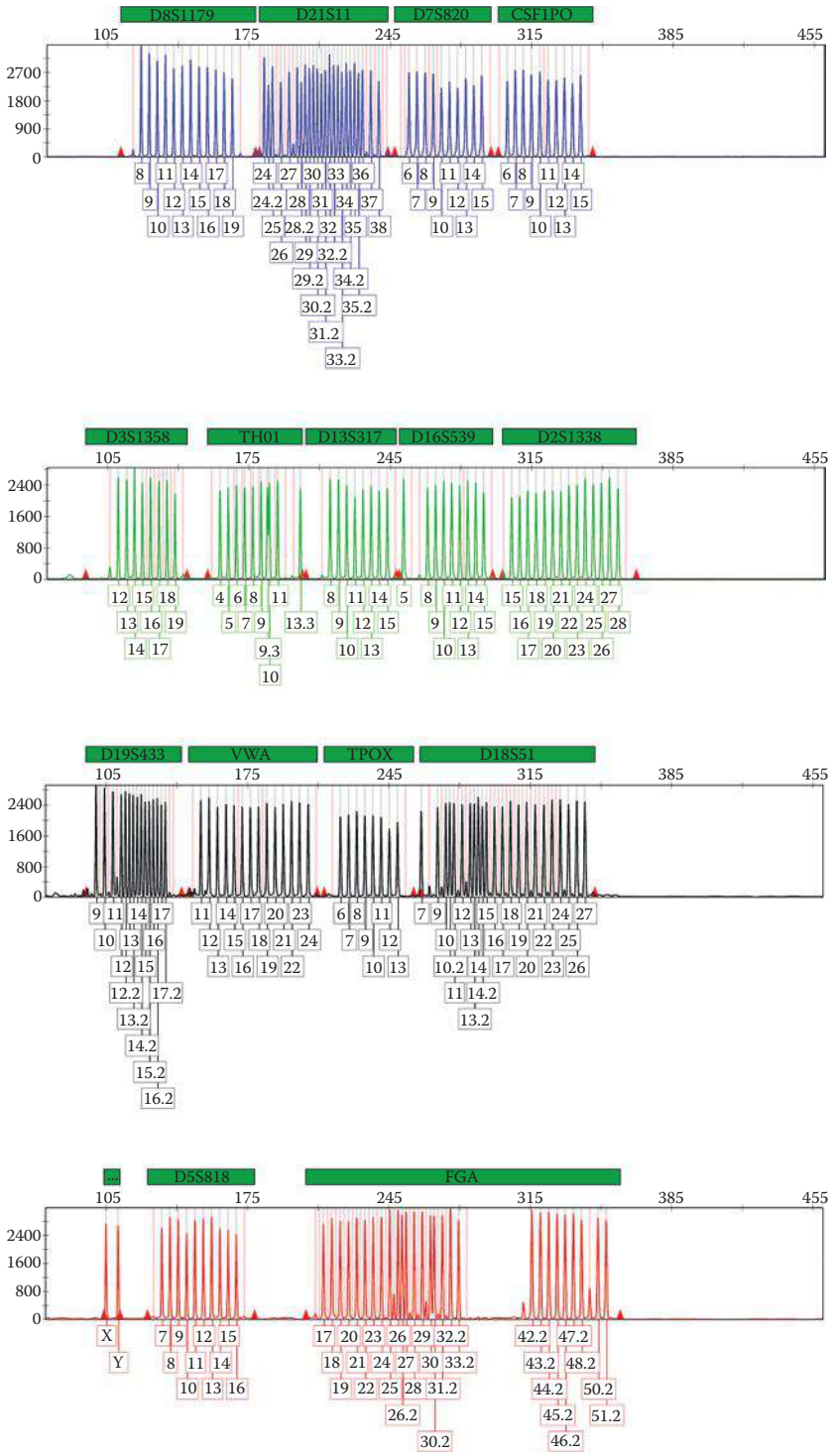

Figure 20.9 Electropherogram of the allelic ladder of the AmpFISTR® Identifiler® Plus Kit (Applied Biosystems). AMEL, CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX, and VWA loci are shown. (© Richard C. Li.)

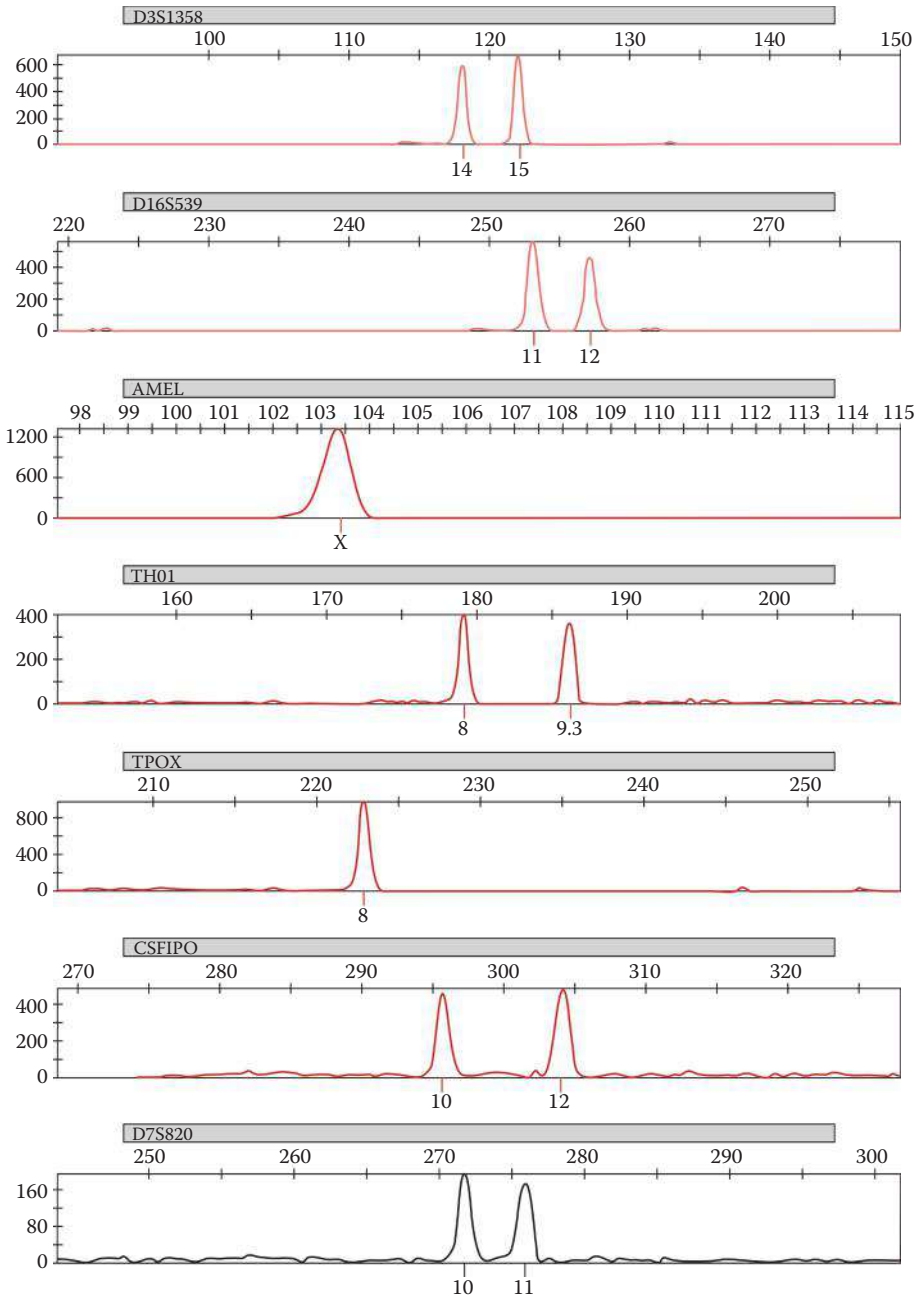

Figure 20.10 Individual DNA profile (Cofiler). The genotype of the DNA profile is shown. AMEL: X, X. CSF1PO: 10, 12. D3S1358: 14, 15. D16S539: 11, 12. D7S820:10, 11. TH01:8, 9.3. TPOX: 8, 8. (© Richard C. Li.)

In contrast, *somatic mutations* involve the mutation of only somatic cells. The germ cells are not affected, and thus a mutant allele is not transmitted to the progeny. A somatic mutation occurring at the core repeat region of an STR locus can be detected and compared to the wild-type allele. The ratio of the signal intensities of the wild-type and mutant alleles varies, depending on the number

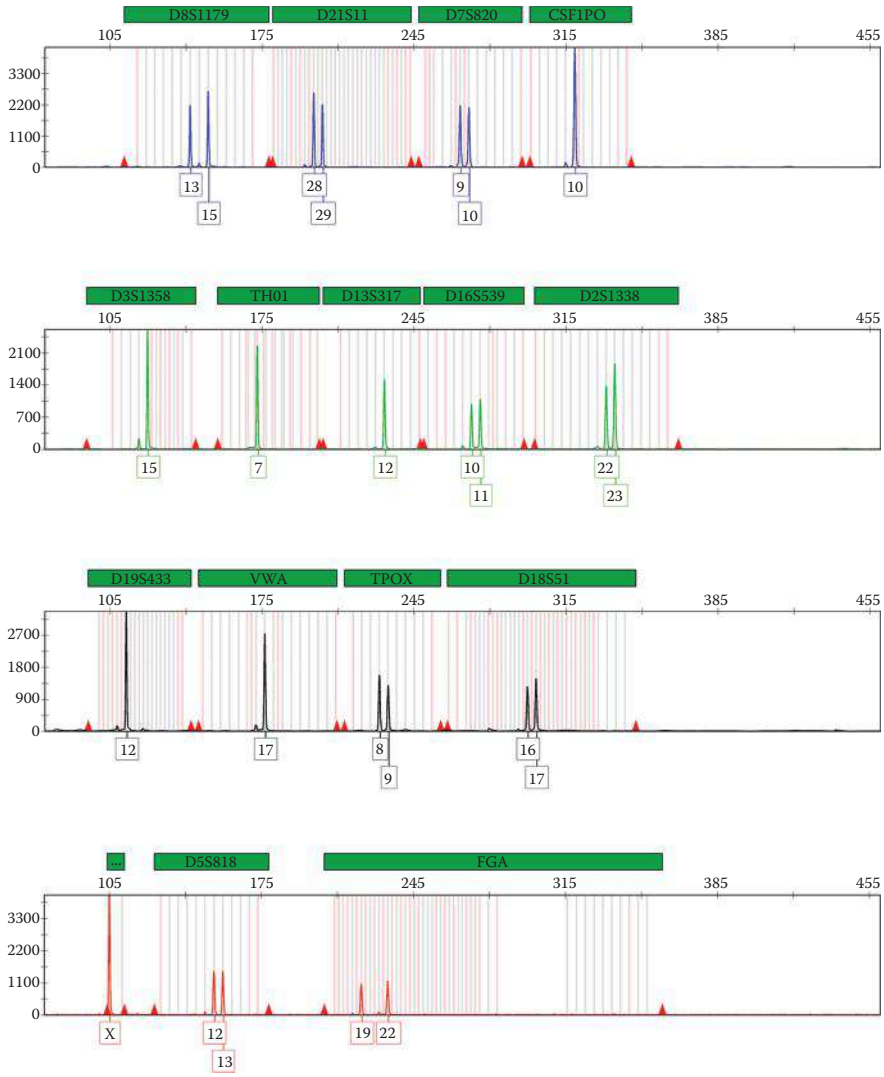

Figure 20.11 Individual DNA profile (Identifiler). The genotype of the DNA profile is shown. AMEL: X, X. CSF1PO: 10, 10. D2S1338: 22, 23. D3S1358: 15, 15. D5S818: 12, 13. D7S820: 9, 10. D8S1179: 13, 15. D13S317: 12, 12. D16S539: 10, 11. D18S51: 16, 17. D19S433: 12, 12. D21S11: 28, 29. FGA: 9, 22. TH01: 7, 7. TPOX: 8, 9. VWA: 17, 17. (© Richard C. Li.)

of mutation-carrying cells in the tissue. Somatic mutations are usually tissue specific. STR profiles from different tissues of the same individual can be compared if a somatic mutation is suspected.

20.4.1.2 Chromosomal and Gene Duplications

Duplicating one of the homologous chromosomes results in a total of three copies of a particular chromosome. This condition, called *trisomy*, is rare and often associated with genetic diseases such as Down's syndrome (chromosome 21 duplication). Duplications have also been observed in chromosomes 13, 18, and X. Other anomalies include duplication of a portion of a chromosome and a single or group of genes instead of an entire chromosome.

A duplication bearing a mutation within the STR core repeat region can affect the number of tandem repeat units. If the duplicated locus is mutated, a *triallelic* or three-peaks pattern can be

detected at a single locus, but not at loci located at other chromosomes in a multiplex STR profile (Figure 20.12). The three alleles usually have equal signal intensity (peak amplitude or peak area). Triallelic patterns at STR loci commonly used for forensic DNA analysis have been documented. Many occur at the TPOX, FGA, and CSF1PO loci. If the duplicate locus is not mutated, only two alleles will be observed in a heterozygote. However, the ratio of the peak amplitude of the alleles will be 1:2 (one copy vs. two copies including the duplicate) at that particular locus. However, STR profiles at the loci located at other chromosomes are not affected.

20.4.1.3 Point Mutations

Point mutations involve the changing of a nucleotide sequence through nucleotide substitution, insertion, or deletion. Insertion or deletion mutations affect the lengths of the core repeat regions and the amplified flanking regions of STR loci and thus affect STR profiles. Nucleotide substitution mutations (except those residing within the primer-binding regions) do not affect the length of DNA and thus do not affect STR profiles.

However, mutations occurring at the primer-binding sequences of the flanking regions of STR loci may affect genotype results. If a mutation at a primer-binding sequence prevents the primer from annealing to the template, this leads to a complete failure of the amplification of the allele. This phenomenon is known as a *null allele* or silent allele (Figure 20.13). To overcome the consequences of a null allele, an alternative primer annealing to a flanking region away from the mutated sequence can be used. Additionally, a primer with the sequence that is complementary to the known mutation can also be used. If the mutation does not completely prevent the primer from annealing but reduces the efficiency of the amplification, the resulting signal intensity of the allele is usually decreased. This problem may be solved by modifying the condition of amplification for the mutant allele.

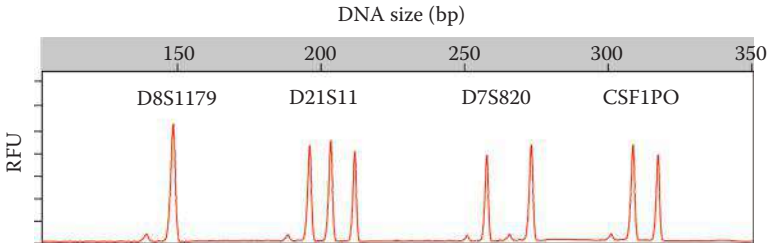

Figure 20.12 Triallele. In this example, the triallele is observed only at D21S11 and not at other STR loci. (© Richard C. Li.)

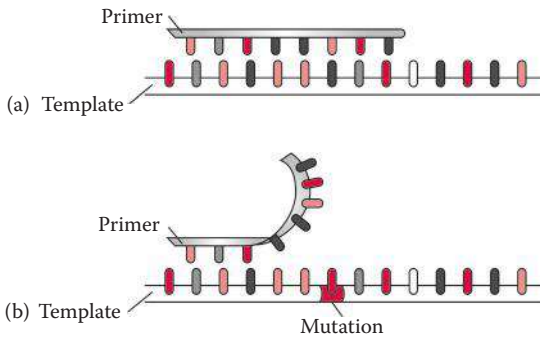

Figure 20.13 Null allele. An allele present in the sample failed to be amplified by one of the primer sets as a result of a rare mutation at the primer-binding sequence of the flanking region: (a) wild type and (b) mutation. (© Richard C. Li.)

20.4.2 Amplification Artifacts

20.4.2.1 Stuttering

A stutter is a minor allele peak, also known as a stutter peak, whose repeat units are shorter or longer than the parental allele peak (Figure 20.14). Less stuttering is observed with pentameric and hexameric repeat unit loci compared to shorter repeat unit loci. The loci that contain complex repeat sequences usually exhibit reduced stuttering. At a given STR locus, large alleles appear to yield more stutter than smaller alleles.

Commonly observed stutters are one repeat unit shorter than the parental allele. It is believed that stuttering is due to the slippage of polymerase, which may have occurred during amplification reactions (Figure 20.15). Stutters with repeat units longer than the parental allele peak can also be observed, but are very rare. The *stutter ratio* is defined as the area of the stutter peak divided by the area of the parental peak. The stutter ratio is usually less than 0.15. A ratio of greater than 0.15 should be interpreted with caution due to the potential presence of DNA from more than one contributor.

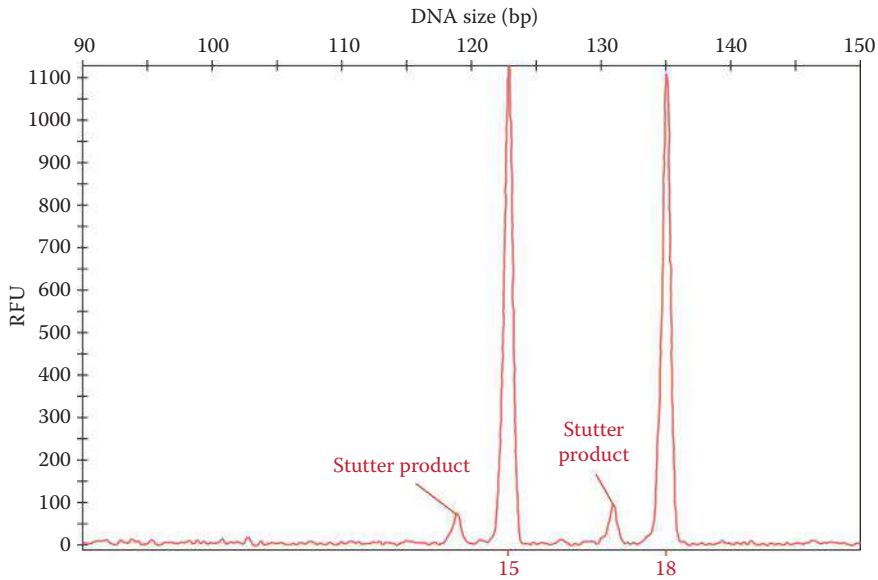

Figure 20.14 Stutter products. (© Richard C. Li.)

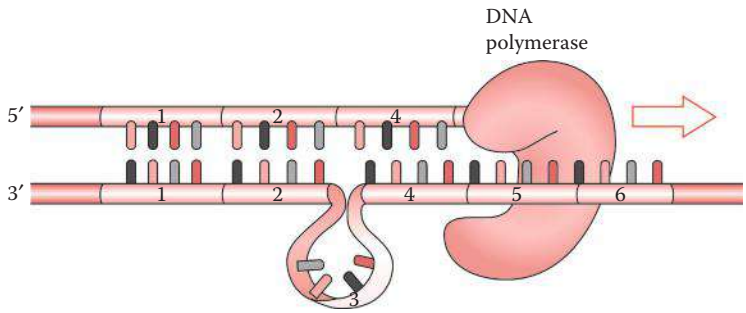

Figure 20.15 Proposed mechanism for stutter products. During the DNA synthesis step of PCR amplification, a DNA polymerase slips, and a region of the primer–template complex becomes unpaired, causing the template strand to form a loop. The consequence of this one-repeat loop is a shortened PCR product smaller than the template by a single repeat unit. (© Richard C. Li.)

20.4.2.2 Nontemplate Adenylation

During PCR amplification, DNA polymerase often adds an extra nucleotide, usually an adenosine, to the 3′-end of an amplicon. Such a phenomenon is referred to as a *nontemplate addition* resulting in an amplicon that is one base pair longer than the parental allele (designated the +A peak), as shown in Figure 20.16. As a result, an amplicon has both the –A amplicon, which corresponds to the size of the parental allele, and the +A amplicon, which represents the amplicon with the nontemplate addition. To simplify the analysis, the most commonly used multiplex STR kits utilize amplification conditions that favor the adenylation of amplicons. Thus, most amplicons in a sample contain an additional adenosine at the 3′ end (+A peak). However, partial nontemplate addition can occur when too much DNA template is utilized in PCR amplification. As a result, a mixture of –A and +A peaks is usually observed.

20.4.2.3 Heterozygote Imbalance

Heterozygote imbalance occurs when one of the alleles has greater peak area or amplitude than the other allele within the same locus in which the two alleles of a heterozygote are compared (Figure 20.17). It is believed that heterozygote imbalance may arise if the DNA sample contains unequal copies of DNA template of the two alleles for the heterozygote, or the two alleles of a heterozygote are unequally amplified, a condition known as *preferential amplification*. Preferential amplification usually refers to an event where a smaller allele is amplified more efficiently than larger ones. As a result, the presence of heterozygote imbalance interferes with the interpretation of samples with a DNA mixture derived from more than one contributor.

20.4.2.4 Allelic Dropout

Allelic dropout occurs when an allele, usually one of the heterozygote alleles, fails to be detected. To date, our understanding of what causes the dropout is very limited. The occurrence of allelic dropout can be the result of an extreme situation of preferential amplification or heterozygote imbalance. Additionally, certain mutations leading to amplification failure (Section 20.4.1.3) can cause allelic dropout.

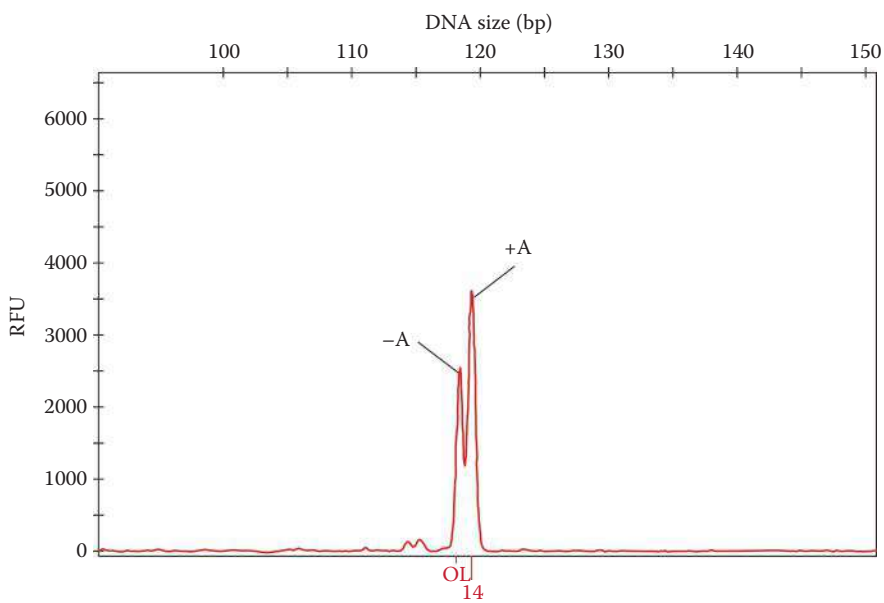

Figure 20.16 Nontemplate adenylation. OL represents the off-ladder allele. (© Richard C. Li.)

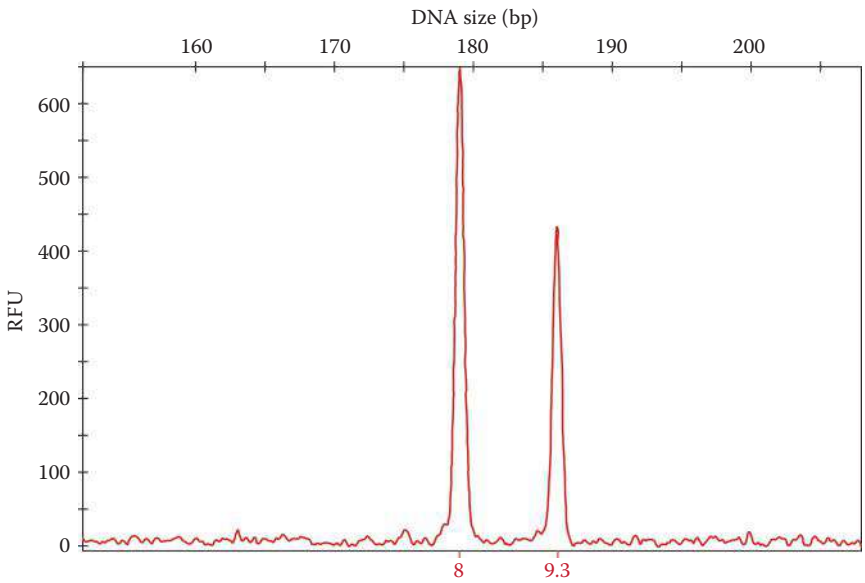

Figure 20.17 Heterozygote imbalance. The signal intensity of one allele is greater than that of the other allele within the same locus. (© Richard C. Li.)

20.4.3 Electrophoretic Artifacts

20.4.3.1 Pull-Up Peaks

A *pull-up peak* occurs when a minor peak of one color on an electropherogram is pulled up from a major allelic peak in another color (Figure 20.18) when the colors have overlapping spectra. For example, a green peak may pull up a yellow peak, or a blue peak may pull up a green peak. A pull-up peak may contribute to the inaccuracy of a profile if the position of a pull-up peak corresponds to the position of an allele. A pull-up peak often occurs when a sample is overloaded or a matrix file (a spectral calibration) is not updated. Thus, loading a proper amount of sample or installing an appropriate matrix file can prevent the occurrence of pull-up peaks.

20.4.3.2 Spikes

Spikes are sharp peaks, with similar signal intensities, that are present in all color panels of an electropherogram (Figure 20.19). Spikes are caused by air bubbles and urea crystals in the capillary of an electrophoretic platform. Voltage spikes can also contribute to spike peaks. The spikes are electrophoretic artifacts and are not reproducible. Thus, electrophoresis can be repeated to verify that the spikes occurred in a previous electrophoresis result.

20.5 Genotyping of Challenging Forensic Samples

20.5.1 Degraded DNA

Environmental exposure, such as high humidity and temperature, of biological evidence can lead to DNA degradation such as the breaking of DNA molecules into small fragments. The more severe the degradation, the more intensive the fragmentation. In forensic DNA analysis, the size range of STR amplicons is usually 100–500 bp in length. When a sample experiences some degradation, large alleles are less likely to be amplified than small alleles (Figures 20.20 and 20.21). As a result, the dropout of large alleles often occurs, leading to a partial DNA profile or even a failure in obtaining a DNA profile. To address this issue, the PCR primers can be

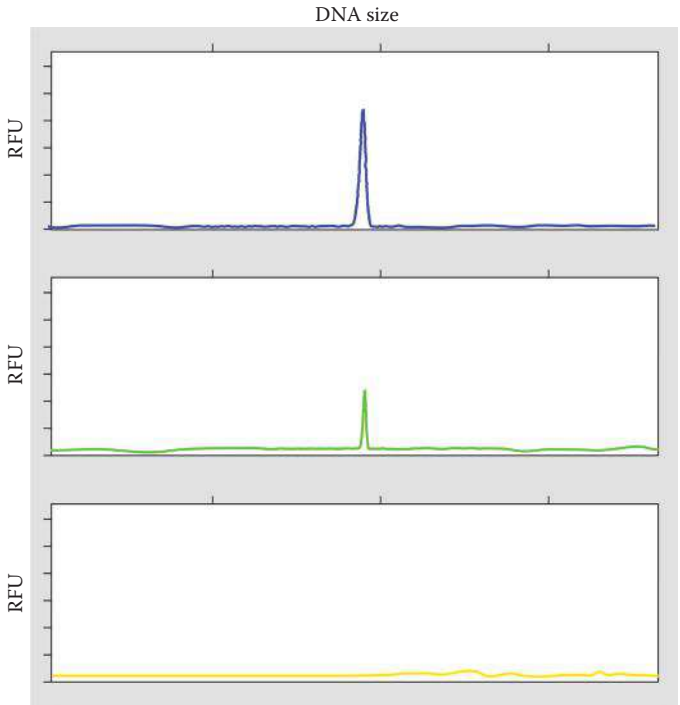

Figure 20.18 Pull-up peaks. The peaks with overlapping spectra observed in the top and middle panels are not observed in the bottom panel. (© Richard C. Li.)

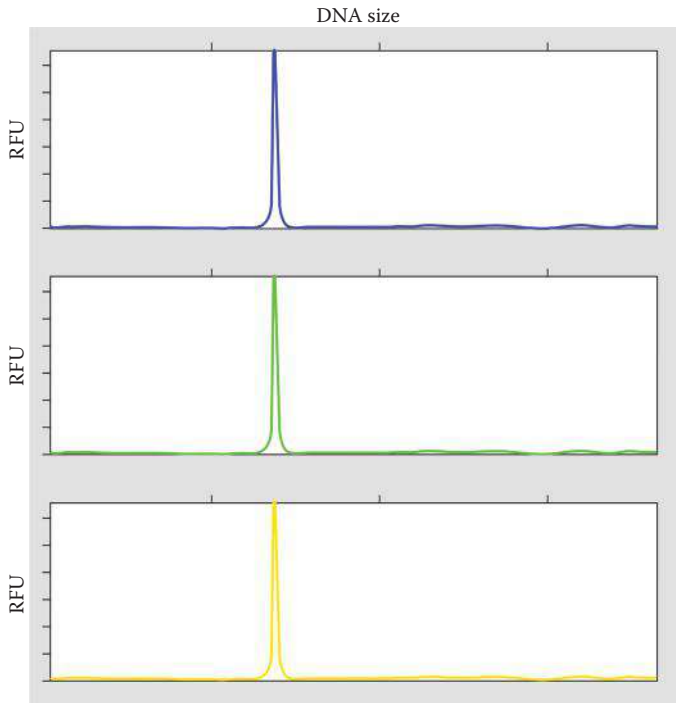

Figure 20.19 Spike peaks can be observed in various intensities. (© Richard C. Li.)

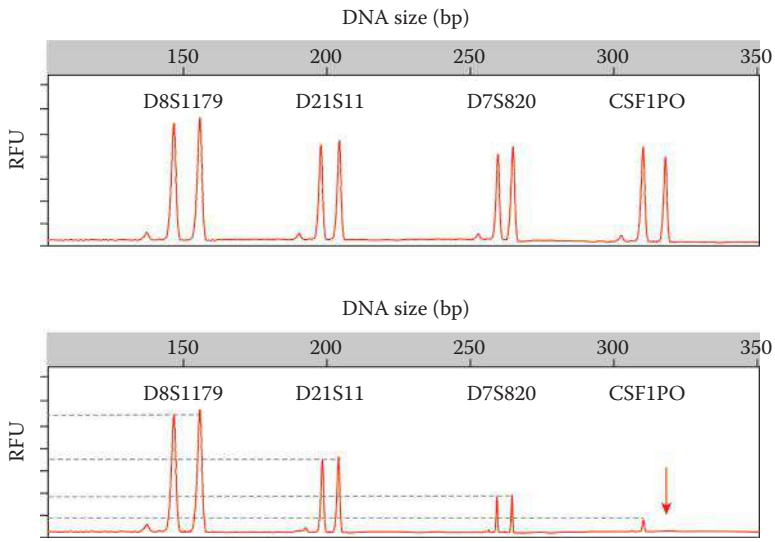

Figure 20.20 Diagram of STR profiles of samples: without DNA degradation (top) and with degradation (bottom). Arrow: allelic dropout. (© Richard C. Li.)

redesigned to anneal more proximally to the STR core repeat region than standard STR primers, yielding small amplicons also known as miniSTRs. Using the miniSTR strategy, more alleles can be detected in degraded DNA samples than using the standard STR primers.

20.5.2 Low Copy Number DNA Testing

Low copy number (LCN) DNA analysis involves the testing of very small amounts of DNA (<100 pg) in a sample. LCN DNA analysis is often needed for samples derived from evidence such as fingerprints and tools and weapons handled by perpetrators. STR analysis of extremely low levels of human DNA can be achieved by increasing the number of PCR cycles (e.g., increasing from 28 to 34 cycles) to improve the yield of amplicons, thus improving the sensitivity of the analysis.

However, this approach also increases the appearance of artifacts that can make interpretation difficult. For instance, the occurrence of allele dropout, heterozygote imbalance, and stuttering is frequently observed in LCN DNA analysis. Additionally, allele drop-in can arise from contamination. The phenomenon of allele drop-in is usually not reproducible. In an LCN DNA analysis, genotypes can be determined if identical alleles can be detected from two independent amplification reactions.

20.5.3 Mixtures

Samples of DNA from two or more contributors are commonly encountered in forensic cases such as sexual assaults in which the evidence recovered from a victim is mixed with a suspect's bodily fluids (Figure 20.22). The interpretation of DNA profiles of mixed stains is known as *mixture interpretation*, which is described below:

1. *To determine the presence of a mixture*: First, determine whether the source of the DNA in the sample came from one or more individuals by examining the number of alleles at multiple loci. The characteristics listed below usually indicate a mixture:
   a. Severe heterozygote imbalance.
   b. Stutter ratio above 0.15.
   c. Presence of three or more alleles per locus at multiple loci.

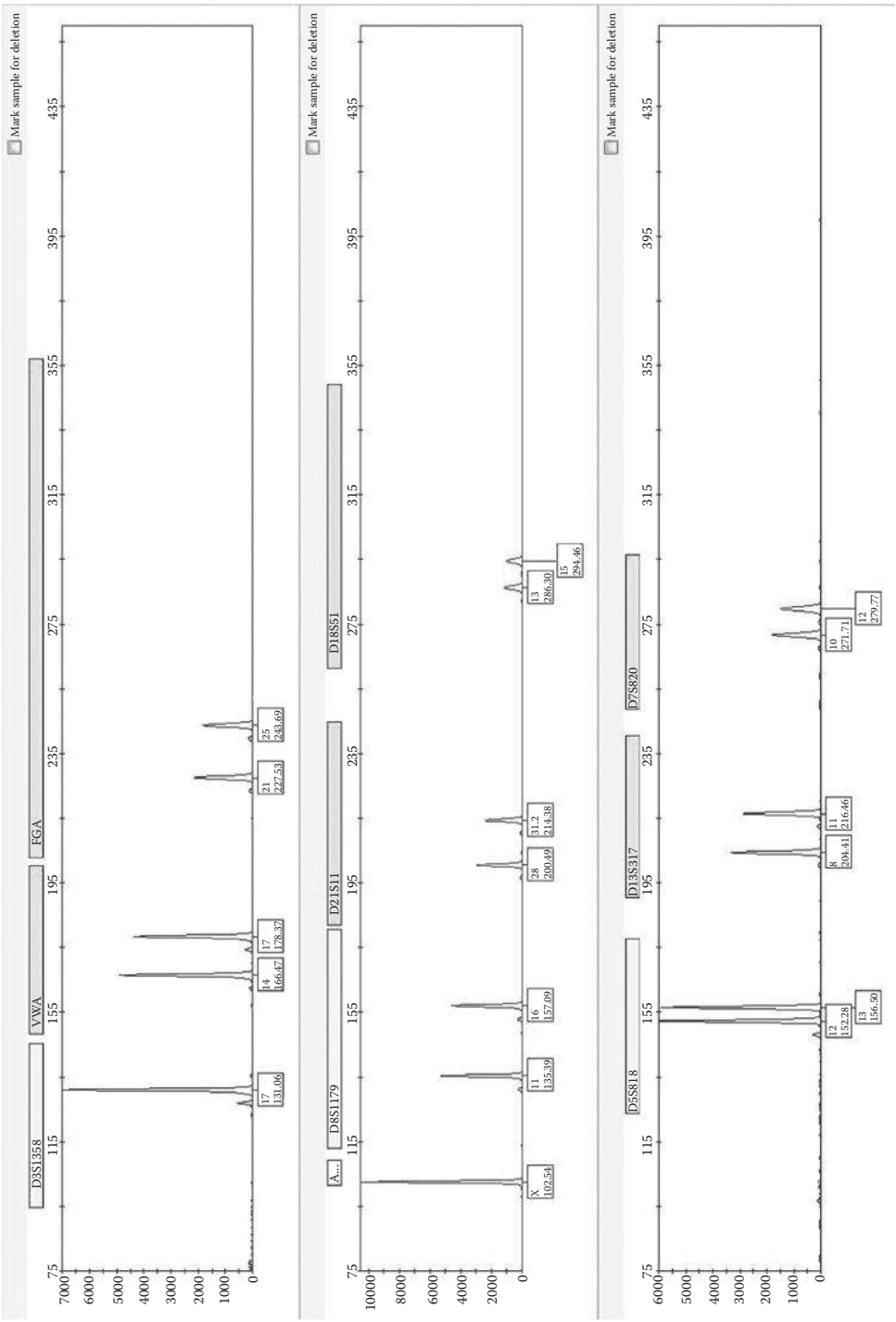

Figure 20.21 DNA profile (AmpFlSTR® Profiler Plus® PCR Amplification Kit; Applied Biosystems) of the degraded sample. (© Richard C. Li.)

Additionally, caution should be taken to distinguish the presence of a mixture and various artifacts such as stutters and nontemplate adenylation.

2. *To determine the genotypes of all alleles and to identify the number of contributors*: Note that the maximum number of alleles at any given locus is two per individual. In the case of homozygous or allele overlap, the number of alleles observed can be less than two per individual.

3. *To estimate the ratios of the contributions*: Determine the relative ratios of the contributions to the mixture made by each individual by comparing the peak areas or

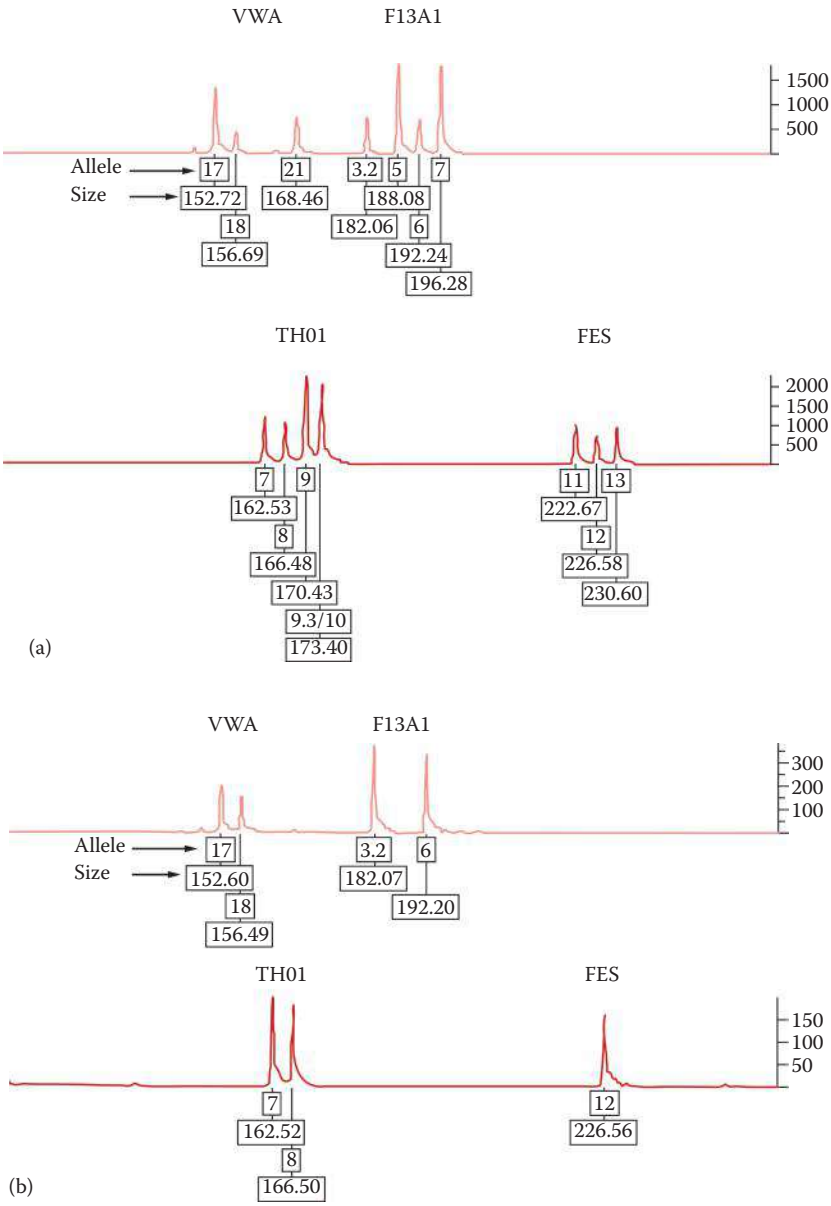

Figure 20.22 DNA profiles of mixed bodily fluids. (a) DNA profile of mixed stains from evidence, (b) DNA profile of the victim. (© Richard C. Li.)

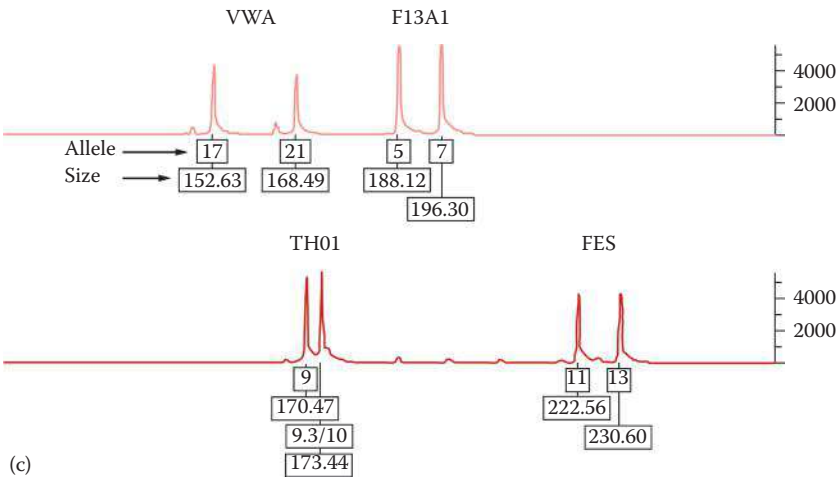

Figure 20.22 (Continued) (c) DNA profile of the suspect.

amplitudes. Amelogenin, a sex-typing marker, is useful in determining the genders of DNA contributors.

4. *To consider all possible genotype combinations*: This may be done by pair-wise comparisons to determine the allele combinations that belong to the minor contributor and those that belong to the major contributor.

5. *To compare reference samples*: The final step is to compare the genotype profiles with the genotypes of reference samples from a suspect and victim. If the DNA profile of the suspect's reference sample matches a major or minor component of the mixture, the suspect cannot be excluded as a contributor.

Detailed guidance for basic steps in mixture interpretation is provided in a number of published guidelines. For example, the Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories was recently established by the Scientific Working Group on DNA Analysis Methods (SWGDAM), which provides direction on DNA profiling and mixture interpretation. Additionally, guidance for the statistical evaluations of the mixture analysis results is also provided.

Bibliography

STR Profiling

Andersen, J., et al., Report on the third EDNAP collaborative STR exercise. *Forensic Sci Int*, 1996, **78**(2): 83–93.

Anker, R., T. Steinbrueck, and H. Donis-Keller, Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. *Hum Mol Genet*, 1992, **1**(2): 137.

Asamura, H., et al., MiniSTR multiplex systems based on non-CODIS loci for analysis of degraded DNA samples. *Forensic Sci Int*, 2007, **173**(1): 7–15.

Asamura, H., et al., MiniY-STR quadruplex systems with short amplicon lengths for analysis of degraded DNA samples. *Forensic Sci Int Genet*, 2007, **1**(1): 56–61.

Aznar, J.M., et al., I-DNASE21 system: Development and SWGDAM validation of a new STR 21-plex reaction. *Forensic Sci Int Genet*, 2014, **8**(1): 10–19.

Balding, D.J., Evaluation of mixed-source, low-template DNA profiles in forensic science. *Proc Natl Acad Sci U S A*, 2013, **110**(30): 12241–12246.

21

Sex Chromosome Haplotyping and Gender Identification

21.1 Y Chromosome Haplotyping

21.1.1 Human Y Chromosome Genome

The Y chromosome is inherited from the father and is passed on to all male offspring; this is known as *patrilineage* (Figure 21.1). Thus, the Y chromosome is unique to males. The chromosome encodes dozens of genes required for male-specific functions, including sex determination and spermatogenesis. The human Y chromosome genome contains approximately 59 million base pairs (bp) and likely contains 50–60 genes. The Y chromosome can be divided into two regions: the pseudoautosomal region and the male-specific Y region (Figure 21.2).

21.1.1.1 Pseudoautosomal Region

The *pseudoautosomal regions (PARs)* are homologous nucleotide sequences that are present on the X and Y chromosomes. There are two PARs on each X and Y chromosome: PAR1 and PAR2. In the Y chromosome, PAR1 is located on the terminal region of the short arm. The PAR1 comprises 2.6 Mb. Twenty-four genes have been identified within the PAR1. The PAR2 of the Y chromosome is located at the tip of the long arm. The PAR2 comprises 320 kb, with only four genes identified so far.

The PARs play an important role in proper segregation of the X and Y chromosomes during meiosis. During meiosis in males, the PARs allow the Y chromosome to pair with the X chromosome. Crossing over and recombination (Chapter 25) within the PARs of the X and Y chromosomes can occur. As a result, males can inherit an allele originally present on the PARs of the X chromosome, and females can inherit an allele originally present on the PARs of the Y chromosome. In particular, PAR1 plays a major role in X–Y chromosome pairing. The deletion of PAR1 results in failure of pairing and leads to male infertility. PAR2 is much shorter than PAR1, and thus shows a lower frequency of pairing than PAR1. The deletion of PAR2 exhibits less severe phenotypes than that of PAR1.

21.1.1.2 Male-Specific Y Region

The remainder of the Y chromosome is known as the *male-specific Y (MSY)* region. It was previously called the *nonrecombining Y (NRY)* region. The bulk of this region does not participate in homologous recombination. However, certain sections involve intrachromosomal gene conversion, which is the nonreciprocal transfer of genomic DNA between a pair of repeated genes on a single chromosome; in this case, the Y chromosome. About 40 megabases (Mb) within the

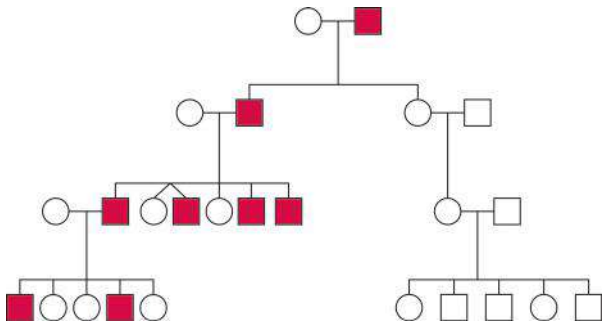

Figure 21.1 Human family pedigree showing inheritance of the Y chromosome. Females and males are denoted by circles and squares, respectively. Red symbols indicate individuals who inherited the same Y chromosome.

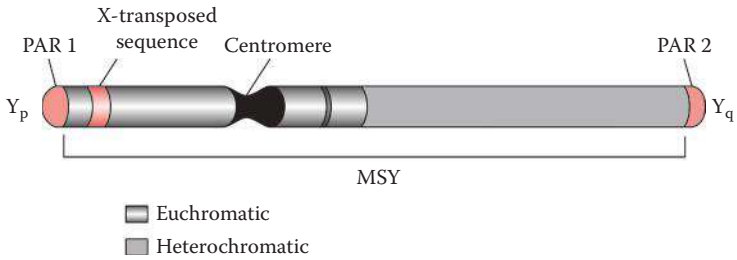

Figure 21.2 Human Y chromosome structure. PAR, the pseudoautosomal region; MSY, male-specific Y region; Y_p, the short arm of the Y chromosome; Y_q, the long arm of the Y chromosome. (© Richard C. Li.)

MSY region are heterochromatic (highly repetitive sequences) including the centromeric region and the bulk of the distal long arm. The euchromatic region, which comprises transcriptional active genes, is approximately 23 Mb. Certain sections of the euchromatic region share some homology with the X chromosome. For instance, X-transposed sequences of the Y chromosome are 99% identical to sequences within Xq21 (a band in the long arm of the X chromosome). The X-transposed sequences are the sequences, a total of approximately 3.4 Mb in length, that were transposed from the X chromosome to the Y chromosome several million years ago. Additionally, dozens of genes located in the euchromatic region share 60%–96% homology with their X chromosome counterparts. The regions sharing homology with the X chromosome should be avoided when selecting Y chromosome–specific markers for forensic DNA profiling.

21.1.1.3 Polymorphic Sequences

The Y chromosome contains an abundance of polymorphic markers. DYF155S1, also known as MSY1, is the first characterized variable number of tandem repeats (VNTR) or minisatellite at the human Y chromosome. It consists of an array of AT-rich repeats at 25 bp per unit repeat. The DYF155S1 locus is highly polymorphic. The unit repeat sequence varies through base substitution. At least five different variant types of unit repeat sequences exist. Additionally, the numbers of these units also vary, ranging from 50 to 115 repeats. Thus, the length of alleles range from approximately 1200 to 2800 bp. DYF155S1 is of considerable interest as a potential marker in forensic testing. However, the analysis method is labor intensive, making it difficult to implement in forensic casework, such as the investigation of sex crimes. Moreover, many single nucleotide polymorphisms (SNPs) and mobile elements exist at the Y chromosome. However, the discrimination power of SNPs and mobile elements is considerably less than that of Y chromosome short tandem repeats (Y-STRs). To date, Y-STRs are usually used for Y chromosome DNA testing due to the high-throughput analysis and good discrimination power (Section 21.1.2).

21.1.2 Y-STR

Y chromosome loci are very important for forensic DNA profiling, and this chapter will discuss such applications. For instance, the Y-STRs used in forensic DNA testing are male specific (for humans and certain higher primates) and are thus useful in the investigation of sexual assault cases involving male suspects. The evidence gathered in such cases usually consists of a mixture of high levels of female DNA and low levels of male DNA. The Y chromosome–specific loci can be examined without interference from large amounts of female DNA; differential extraction of sperm and nonsperm cells may not be needed. Furthermore, the Y-STR system is useful for determining the numbers of unrelated male perpetrators in sexual assault cases. The Y-STR loci used for forensic applications are located in the nonrecombining section of the Y chromosome so that patrilineage can be established. The technique can be used for paternity testing and the identification of missing persons. Finally, data interpretation can be simplified by the use of a single allele per Y-STR locus profile at most loci. Reference databases are available for estimating Y-STR haplotype frequencies among various human populations for statistical analysis of profiling results.

The major disadvantage of Y-STR loci is that their discriminating power is lowerthan that of autosomal loci. Because Y chromosome loci are linked, the product rule for statistical calculations for profile probability does not apply. Chapter 25 discusses the statistical evaluation of the strength of the matches. Also, the current Y-STR profiling cannot distinguish individuals with the same patrilineage.

More than 400 STR loci have been identified in the Y chromosome genome. The precise locations of these loci have been sequentially mapped using human genome sequencing data. Most Y-STR loci, approximately 60% of the 400 identified, are located on the long arm of the chromosome; about 22% are located on the short arm and a few are found in the centromeric region (Figure 21.3). Y-STRs in the telomeric region have yet to be identified. Only about 5% of Y-STRs are located within 5′ untranslated or intron regions of protein-coding genes. The

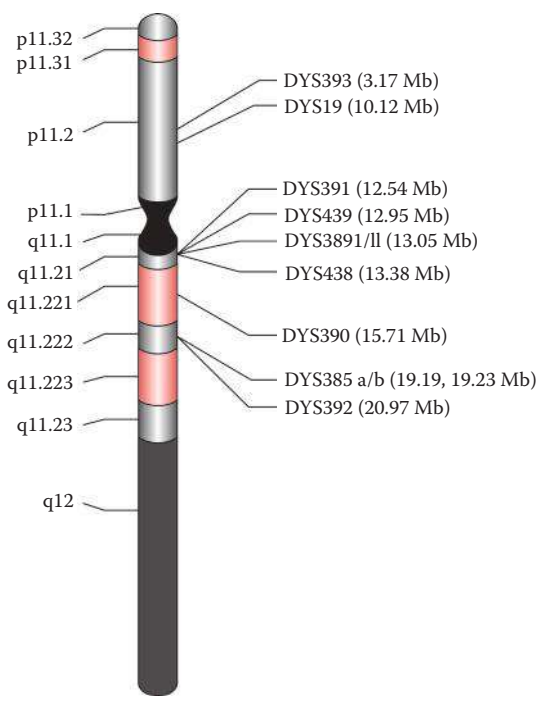

Figure 21.3 Human cytogenetic map of the Y chromosome. The Y-STRs and positions are shown (Mb = megabase). Cytogenetic patterns with alternating dark and light bands are shown. (© Richard C. Li.)

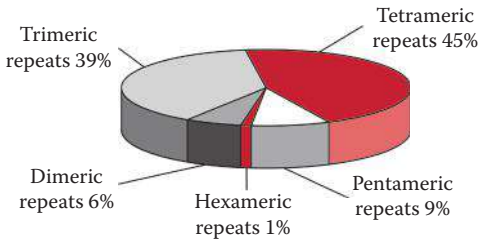

Figure 21.4 Human Y-STRs with different repeat unit length. About 400 Y-STRs have been identified and categorized according to repeat unit length. (Adapted from Hanson, E.K. and Ballantyne, J., *Leg Med (Tokyo)*, 8, 110, 2006.)

repeat unit lengths of identified Y-STRs have been analyzed. Among the 400 Y-STRs, 6% are dimeric repeats, 39% are trimeric, 45% are tetrameric, 9% are pentameric, and 1% are hexameric (Figure 21.4).

Fewer than half of the Y-STRs have been characterized. Some loci are polymorphic and are useful for forensic applications and developing new Y-STR multiplex systems. Since homologous recombination does not occur on the majority of the Y chromosome, alleles of Y-STR loci are linked (inherited together); they are referred to as *haplotypes*. As a result, the discrimination power of Y-STRs is much lower than that of autosomal STRs. The most commonly used Y-STR loci for forensic testing are described below.

21.1.2.1 Core Y-STR Loci

In 1997, the *European minimal haplotype* locus set, also known as the minimal haplotype loci, was recommended by the International Y-STR User Group for forensic applications (Table 21.1; Figure 21.5). This haplotype set includes a core set of nine Y-STR loci: DYS19, DYS385a and b, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393. In 2003, two additional loci were recommended by the Scientific Working Group for DNA Analysis Methods (SWGDAM) for forensic DNA analysis. The SWGDAM loci include the European minimal haplotype loci plus two additional loci: DYS438 and DYS439 (Table 21.1).

The application Y-STR for forensic casework has been expanded with additional Y-STR markers. The use of Y-STR loci has been facilitated by commercially available PCR amplification

Table 21.1 Common Y-STR Core Loci

Locus	EMH	US Haplotype Loci	Repeat Motif
DYS19	☑	☑	TAGA
DYS385a/b	☑	☑	GAAA
DYS389 I	☑	☑	TCTA
DYS389 II	☑	☑	[TCTG] [TCTA]
DYS390	☑	☑	[TCTG] [TCTA]
DYS391	☑	☑	TCTA
DYS392	☑	☑	TAT
DYS393	☑	☑	AGAT
DYS438		☑	TTTTC
DYS439		☑	GATA

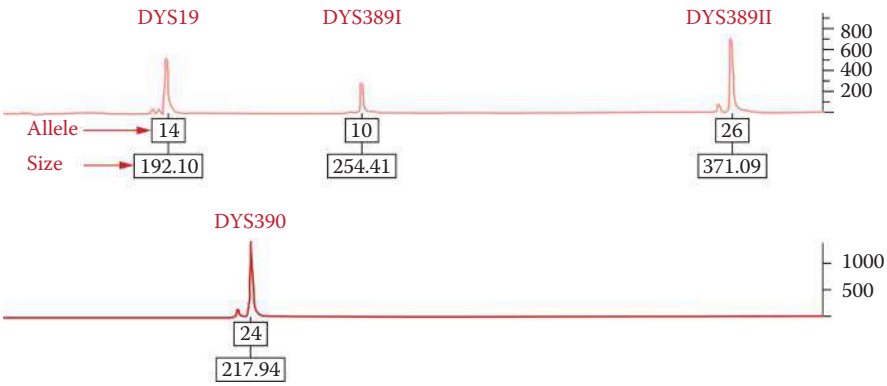

Figure 21.5 Electropherogram of Y-STR profile using an early generation of multiplex of four loci. The genotype of the DNA profile is shown (DYS19: 14. DYS389I: 10. DYS389II: 26. DYS390: 24). (© Richard C. Li.)

kits in multiplex Y-STR systems. To improve discriminating power, multiplex systems including new Y-STR loci are desired. Many new Y-STR loci are being characterized for developing new multiplex systems. Commercially available kits with more Y-STR loci are now available and have been validated for forensic use (Table 21.2; Figures 21.6 and 21.7). Some of the additional Y-STR loci are highly discriminating, allowing for further distinction between unrelated male individuals. Additionally, Y-STR loci, such as DYS570 and DYS576, with high mutation rates (see rapidly mutating Y-STR loci; Section 21.1.2.3) are included, which are useful for the discrimination of related individuals.

21.1.2.2 Multilocal Y-STR Loci

DYS385 and DYS389 are *multilocal Y-STR loci* (*MLL*). The MLL designation refers to the presence of a particular STR at more than one site on the Y chromosome due to duplication. To date, about 50 such MLL Y-STRs have been identified. Further MLL subdivisions are designated bilocal, trilocal, and so on. DYS385 and DYS389 are bilocal.

The DYS385 locus has two inverted duplicated clusters and is separated by a 4×10^4 bp interstitial region (Figure 21.8). It can be amplified by a single set of primers. One allele is observed if the duplicates are the same length. If the duplicated clusters have different lengths, they can generate two different alleles when amplified. The smaller allele is designated "a" and the larger allele is designated "b." Moreover, the DYS389 locus has two duplicated clusters with the same orientation (Figure 21.9). In a single set of PCR primers, there are two binding sites for the same forward primer at each 5′ flanking sequence of the core repeat region of DYS389. These binding sites between DYS389I and DYS389II are about 120 bp apart. Therefore, two amplicons are produced. DYS389I is designated for the smaller allele, and DYS389II is designated for the larger allele.

21.1.2.3 Rapidly Mutating Y-STR

Current Y-STR loci implemented in forensic casework have adequate resolution of males from different patrilineages. However, most Y-STR sets have limited abilities to differentiate related males who belong to the same patrilineage. Thus, current forensic Y-STR profiling is not able to exclude patrilineal relatives of the suspect. Recently, rapidly mutating Y-STR (RM Y-STR) loci have been identified and characterized. The mutation rates of RM Y-STR are above 10^{-2}, which is considerably higher than the average mutation rates of Y-STRs. The average mutation rate for the majority of the Y-STRs characterized, including the core Y-STR loci currently used in forensic casework, is approximately 10^{-4}–10^{-3} per generation. Due to the high mutation rates, RM Y-STRs can improve patrilineage resolution. For use in forensic casework, the RM Y-STR loci have potential abilities to differentiate both paternally related and unrelated males.

Table 21.2 Y-STR Loci Covered by Representative Commercial Kits

	Loci	Yfiler[a]	PowerPlexY23[b]
Core loci	DYS19	☑	☑
	DYS385a/b	☑	☑
	DYS389 I	☑	☑
	DYS389 II	☑	☑
	DYS390	☑	☑
	DYS391	☑	☑
	DYS392	☑	☑
	DYS393	☑	☑
	DYS438	☑	☑
	DYS439	☑	☑
Additional loci	DYS437	☑	☑
	DYS448	☑	☑
	DYS456	☑	☑
	DYS458	☑	☑
	DYS481		☑
	DYS533		☑
	DYS549		☑
	DYS570		☑
	DYS576		☑
	DYS635	☑	☑
	DYS643		☑
	Y-GATA-H4	☑	☑

[a] AmpFlSTR® Yfiler® PCR Amplification Kit (Applied Biosystems).
[b] PowerPlex® Y23 System (Promega).

21.2 X Chromosome Haplotyping

X-chromosomal STR (X-STR) profiling is a useful tool in kinship testing in forensic investigations. For example, males under usual circumstances have only one X chromosome; thus males are hemizygous for the X-STR loci on the X chromosome. Homologous recombination between the X and Y chromosomes is restricted to the homologous PARs (Figure 21.2). The paternal X chromosome is inherited by daughters as haplotypes. Thus, father–daughter kinship is easier to determine using X-STRs than autosomal STRs. In females, there are two copies of the X chromosomes. Homologous recombination can occur between two X chromosomes in the mother–child transmission. Although it is possible to determine mother–child kinship using X-STR analysis, the level of certainty is less than that of father–daughter kinship analysis. Nevertheless, sharing of rare X-STR haplotypes can strengthen an indication of kinship. The most useful

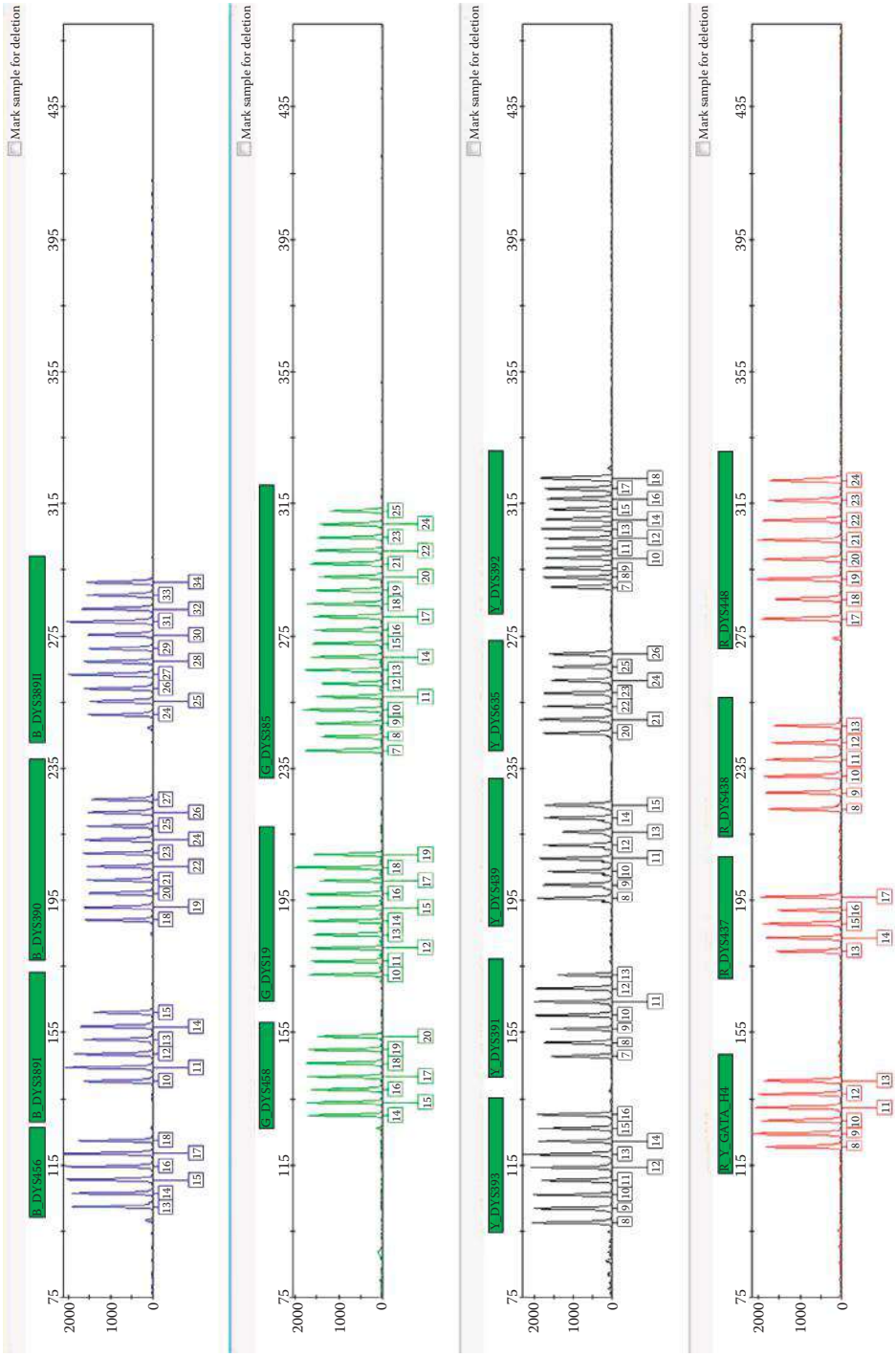

Figure 21.6 Electropherogram of the allelic ladder of AmpFlSTR® Yfiler® PCR Amplification Kit (Applied Biosystems). DYS19, DYS385 a/b, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and Y-GATA-H4 loci are shown. (© Richard C. Li.)

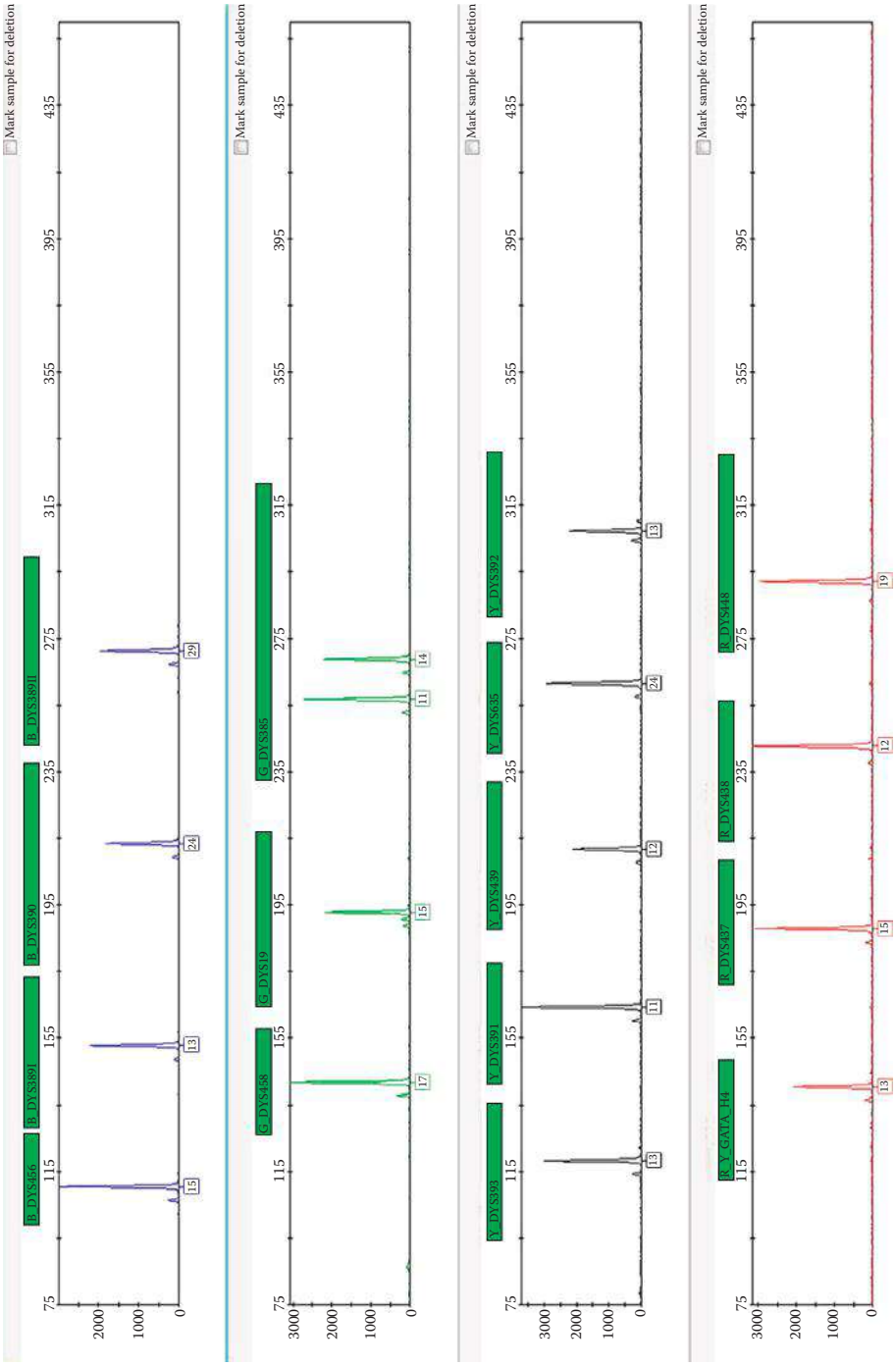

Figure 21.7 Individual Y-STR profile (AmpFISTR® Yfiler® PCR Amplification Kit). The genotype of the DNA profile is shown. DYS19: 15. DYS385 a/b: 11, 14. DYS389 I: 13. DYS389 II: 29. DYS390: 24. DYS391: 11. DYS392: 13. DYS393: 13. DYS437: 15. DYS438: 12. DYS439: 12. DYS448: 19. DYS456: 15. DYS458: 17. DYS635: 24.Y-GATA-H4: 13. (© Richard C. Li.)

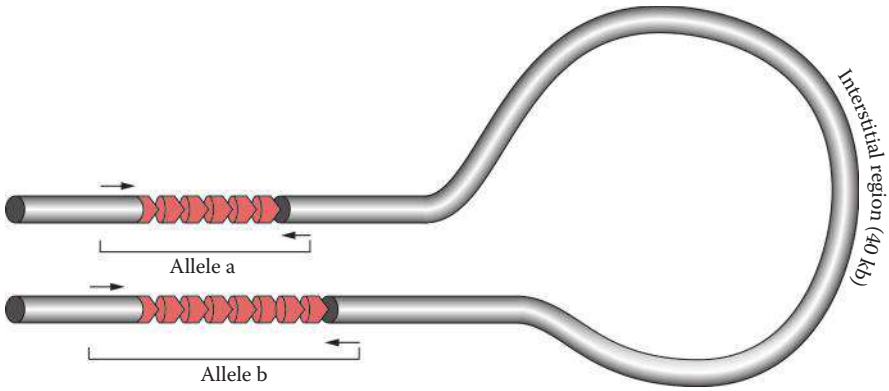

Figure 21.8 MLL Y-STR locus DYS385. At the DYS385 locus, note the two inverted duplicated regions of the Y chromosome with an interstitial region of 40 kilobases (kb). These inverted regions can be amplified with a single pair of primers (indicated by arrows) in one PCR reaction. The allele designations are described in the text. (© Richard C. Li.)

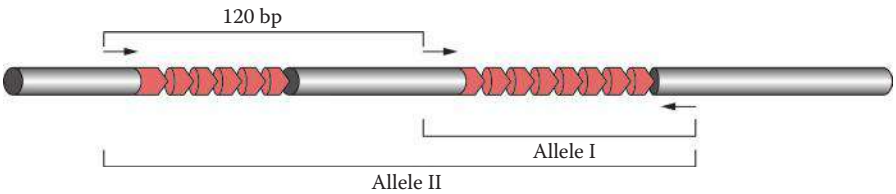

Figure 21.9 MLL Y-STR locus DYS389. At the DYS389 locus, note the two duplicated regions of the Y chromosome with the same orientation. These duplicated regions can be amplified with a single pair of primers (indicated by arrows) in one PCR reaction. The allele designations are described in the text. (© Richard C. Li.)

application of X-STR profiling, however, is in situations where autosomal STR profiling fails to determine kinship with high levels of certainty, and where Y-chromosomal or mitochondrial DNA profiling results are inconclusive.

Many X-STR loci have been identified spanning the entire human X chromosome. Increasing the number of X-STR loci to be analyzed in the inherited region would increase the degree of certainty in determining kinship. However, it is necessary to take into consideration the linkage between X-STR loci on the X chromosome. Based on Gregor Mendel's first law, the different alleles of two loci segregate independently. This is definitely true for loci localized on different chromosomes. The segregation can also be observed for loci located on the same chromosome, including the X chromosome, due to homologous recombination and meiotic crossing-over. Generally speaking, the farther apart the two loci are on the chromosome in physical distance, the more likely those loci are to segregate independently. When two loci are very close to each other, the alleles of the two loci are inherited together as a haplotype. Therefore, including additional closely linked STR loci does not necessarily improve the discrimination power in kinship analysis.

Homologous recombination events at different regions of the genome may occur with different probabilities. Thus, the tightness of linkage between two X-STR loci cannot be accurately measured based on the physical distance. In order to determine how closely two different loci are linked, the *recombination fraction* (or recombination frequency) can be used, which is the percentage of recombinants resulting from chromosomal crossover between two loci during meiosis among all the offspring. The recombination fraction reflects how often the loci are inherited

together or how often they are separated. The recombination fraction ranges from 0% to 50%. The minimum recombination fraction is 0, indicating perfect linkage (no recombinants). The maximum is 50%, meaning complete independence of the two loci (they are actually located on different chromosomes or on the same chromosome with great distance between them).

Multiplex X-STR assays have also been developed. Commonly used X-STR loci can be grouped into linkage groups. Each linkage group is presumably transmitted independently from one another. These loosely linked loci can be treated as if they were located on different chromosomes. The distribution of alleles at these X-STR loci should be in accordance with the Hardy–Weinberg principle (Chapter 25). Within each linkage group, the occurrence of recombination is negligible. Thus, loci within groups are generally considered as closely linked and are thus treated as a haplotype.

21.3 Sex Typing for Gender Identification

Sex typing of a biological sample is useful in forensic investigation. For example, the application of sex typing facilitates the identification of the victim and the offender's DNA evidence in sexual assault cases and the remains of victims in mass disasters or missing persons investigations. One commonly used sex-typing marker is the amelogenin (*AMEL*) locus. The use of multiplex PCR systems with an additional amelogenin marker, a non-STR marker, leads to potential gender determination.

21.3.1 Amelogenin Locus

This region encodes extracellular matrix proteins involved in tooth enamel formation (Table 21.3). Mutations in the *AMEL* gene can lead to an enamel defect known as amelogenesis imperfecta. Amelogenesis imperfecta is a disorder that causes abnormal formation of tooth enamel in both primary and permanent teeth. The formed tooth enamel is soft and thin; therefore it is easily damaged. The *AMEL* locus has two homologous genes: *AMELX* (Xp22.2) is located on the human X chromosome (Figure 21.10) and *AMELY* (Yp11.2) is located on the human Y chromosome (Figure 21.11). Although the genes constitute a homologous pair, they differ in size and sequence. The coding sequence of the *AMELX* gene has a size of 2872 bp and the *AMELY* gene has a size of 3272 bp in length.

The most commonly used sex-typing method at the *AMEL* locus is the detection of a 6 bp deletion at intron 3 of *AMELX* (Figures 21.12 and 21.13). This deletion is not present in *AMELY*. Sex typing can be performed using primers specifically amplifying the region of the *AMEL* locus. Primer sets were developed to amplify both alleles in a single PCR. The two most commonly

Table 21.3 Sex-Typing Makers and Homologous Genes on the Human X and Y Chromosomes

Sex-Typing Markers					
X Chromosome			Y Chromosome		
Gene Symbol	Chromosomal Location	Distance from Xpter (Mb)	Gene Symbol	Chromosomal Location	Distance from Ypter (Mb)
AMELX	Xp22.2	11.3	*AMELY*	Yp11.2	6.7
DXYS156	Xq21.31	88.9	DXYS156	Yp11.2	3.2
SOX3	Xq27.1	139.6	*SRY*	Yp11.31	2.7
STS	Xp22.31	7.1–7.3	*STSP1*	Yq11.221	17.7
TSPYL2	Xp11.2	53.1	*TSPY1*	Yp11.2	9.2–9.3

Source: Ensembl Homo sapiens version 75.37 (GRCh37). Mb, megabase.

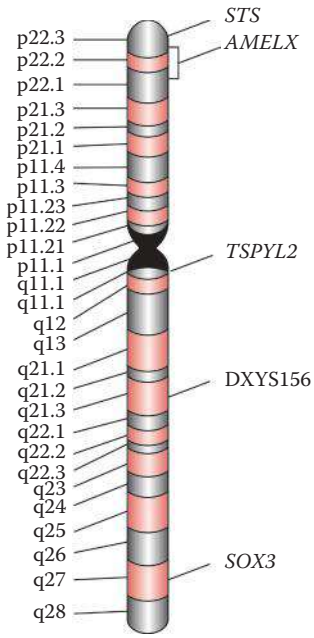

Figure 21.10 Cytogenetic map of the human X chromosome and sex-typing loci shown with physical positions (Mb = megabases). Cytogenetic patterns with alternating dark and light bands are shown. (© Richard C. Li.)

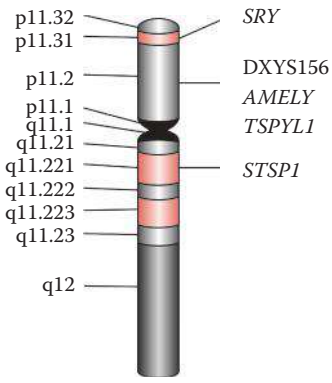

Figure 21.11 Cytogenetic map of the human Y chromosome and sex-typing loci shown with physical positions (Mb = megabases). Cytogenetic patterns with alternating dark and light bands are shown. (© Richard C. Li.)

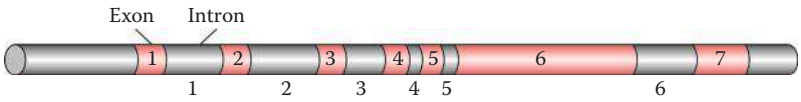

Figure 21.12 Structure of human *AMELY* gene. Exons 1 through 7 and introns 1 through 6 are numbered. (© Richard C. Li.)

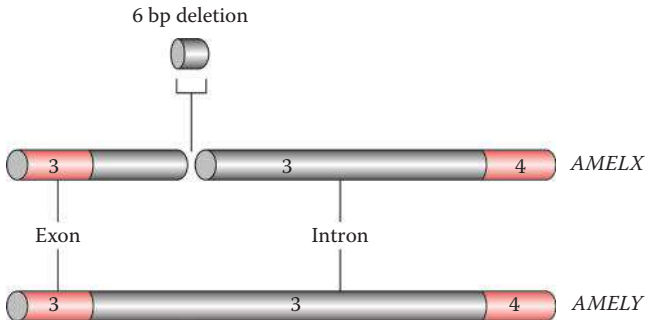

Figure 21.13 Sex typing using *AMEL* markers. A 6-bp deletion in intron 3 is present in *AMELX* but not in *AMELY* and can be resolved using electrophoresis as described in the text. (© Richard C. Li.)

used amelogenin primer sets generate amplicons of 106 and 112 bp or 212 and 218 bp for *AMELX* and *AMELY* loci, respectively. The amplicons generated from *AMELX* and *AMELY* are separated by electrophoresis. The observation of the *AMELX* fragment alone indicates a female, whereas the observation of both *AMELX* and *AMELY* fragments indicates a male (Figure 20.8). Nevertheless, DNA from primates and some rudiments can be amplified as well, but the amplicon sizes vary. The *AMEL* locus has been coamplified with other markers to provide a combined sex and identity test. Such combined tests have been used in historical D1S80 AFLP profiling (Chapter 19) and various current STR multiplex analyses.

Cases of *AMELY* null mutations have been reported where only the *AMELX* fragment is detected in *AMELY* null males. Many of them are phenotypically normal but present the *AMELX* sex types of females. Various interstitial deletions at the Y chromosome short arm and point mutations within the primer-binding sites have been identified as the possible cause of some *AMELY* null sex typing. The frequency of *AMELY* null males is rare, but is higher in Sri Lanka and India. However, the mutations may lead to incorrect identification of sex if the mutations occurred in the DNA evidence of criminal investigations such as sexual assaults or in the identification of human remains in mass disasters. In the case of a female individual, an *AMELX* mutation, which is rare, usually affects one copy of *AMELX*; however, the wild-type copy of the *AMELX* is detected. Thus, the *AMELX* mutation is less problematic in females than the *AMELY* mutation in sex typing in males.

21.3.2 Other Loci

To solve the incorrect identification problem with null *AMEL* mutations, additional candidate genetic markers are used in combination with *AMEL* analysis (Table 21.3). One useful marker is the sex-determining region Y (*SRY*) gene located on the Yp11.31 of the Y chromosome. *SRY* encodes a transcription factor that plays a role in the regulation of sex determination toward male development. The *SRY* protein contains a DNA-binding domain known as the HMG box. *SOX3*, an *SRY*-related HMG box-containing gene, has been identified at Xq27.1 of the X chromosome, which shares sequence homology with *SRY*. *SRY*-specific amplification is achieved using *SRY*-specific primers.

Another marker is the *TSPY* locus located on Yp11.2 of the Y chromosome. The *TSPY* locus encodes the testis-specific protein Y-encoded gene that is only expressed in the testis and may play a role in spermatogenesis. The *TSPY* structural gene is approximately 20 kb in length, consisting of 6 exons and 5 introns. The *TSPY* gene has multiple copies. A single Y chromosome contains several dozen copies of *TSPY* genes that are organized as a tandem array. The *TSPY* gene cluster shows copy number variation (CNV) between individuals in a population ranging from 23 to 64 copies. *DYS14* is a marker utilized for the characterization of the *TSPY* locus. The

DYS14 locus includes the partial sequence of the first exon and intron of the *TSPY* gene and shares approximately 98% sequence homology with other members of the *TSPY* gene cluster.

A *TSPY*-like (*TSPYL*) gene has been identified on the short arm of the X chromosome and is designated as *TSPYL2*. *TSPYL2* is a single-copy gene per X chromosome and is 6.3 kb in length, consisting of 7 exons and 6 introns. *TSPYL2* shares homology with *TSPY*. Additionally, several *TSPYL* genes have been identified on the autosomes. Primers are designed to amplify *TSPY* only and not amplify *TSPYL2* and other *TSPYL* gene sequences. The sensitivity of detecting *TSPY*, due to multiple copies, can be considerably higher than any single copy genetic maker such as *SRY*.

Furthermore, the DXYS156 locus, located at the pseudoautosomal region of both X and Y chromosomes, is another candidate marker used for sex typing. DXYS156 is a polymorphic pentanucleotide STR. The DXYS156 Y alleles have an additional adenine insertion in the repeat units of STR. Thus, the DXYS156 Y alleles can be distinguished from the DXYS156 X alleles because of the insertion. Additionally, the DXYS156 alleles are ethnically distributed. The DXYS156 profiling may potentially indicate the ethnic origin of an individual as an investigative lead.

Moreover, the steroid sulfatase (*STS*) gene encodes an enzyme that catalyzes the conversion of sulfated steroid precursors to biologically active steroids such as estrogens and androgens. The *STS* gene is located on the distal part of the short arm of the X chromosome within Xp22.31. The *STS* gene is 146 kb in length, consisting of 10 exons and 9 introns. The size of exons is between 120 and 160 bp long. However, the sizes of introns vary greatly, ranging from 102 bp to 35 kb. The entire coding sequence of the *STS* gene is 1542 bp in length. Deletions and point mutations of the *STS* gene have been associated with X-linked ichthyosis, a skin disease that affects males. There is an *STS* pseudogene, 100 kb in length, on the long arm of the Y chromosome known as *STSP1*. Although *STSP1* shares some sequence homologies with the *STS* gene, the pseudogene does not encode a functional gene. Primer sets are designed to amplify the target sequences within the first intron (35 kb in length) of the *STS* gene and its homologous *STSP1* pseudogene sequence. Although the homology of these target sequences of the *STS* gene and *STSP1* pseudogene is approximately 80%, amplicons with two different sizes are produced to identify *STS* and *STSP1* alleles for sex typing.

Bibliography

Y Chromosome Haplotyping

Alves, C., et al., Evaluating the informative power of Y-STRs: A comparative study using European and new African haplotype data. *Forensic Sci Int*, 2003, **134**(2–3): 126–133.

Ambers, A., et al., Autosomal and Y-STR analysis of degraded DNA from the 120-year-old skeletal remains of Ezekiel Harper. *Forensic Sci Int Genet*, 2014, **9**: 33–41.

Andersen, M.M., et al., Estimating stutter rates for Y-STR alleles. *Forensic Sci Int Genet Suppl*, 2011, **3**(1): e192–e193.

Andersen, M.M., et al., Estimating Y-STR allelic drop-out rates and adjusting for interlocus balances. *Forensic Sci Int Genet*, 2013, **7**(3): 327–336.

Asamura, H., et al., Evaluation of miniY-STR multiplex PCR systems for extended 16 Y-STR loci. *Int J Legal Med*, 2008, **122**(1): 43–49.

Ayadi, I., et al., Combining autosomal and Y-chromosomal short tandem repeat data in paternity testing with male child: Methods and application. *J Forensic Sci*, 2007, **52**(5): 1068–1072.

Ayub, Q., et al., Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res*, 2000, **28**(2): e8.

Balaresque, P., et al., Genomic complexity of the Y-STR DYS19: Inversions, deletions and founder lineages carrying duplications. *Int J Legal Med*, 2009, **123**(1): 15–23.

Ballantyne, K.N., et al., A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci Int Genet*, 2012, **6**(2): 208–218.

22

Single Nucleotide Polymorphism Profiling

22.1 Basic Characteristics of SNPs

Human genomes contain sequence polymorphisms, which are the variations in nucleotide sequences among individuals. One type of sequence polymorphism is called *single-nucleotide polymorphism* (*SNP*). An SNP constitutes a single-base-pair change originating from a spontaneous mutation that can be a base substitution, insertion, or deletion at a single site. An estimated 10 million SNPs may exist in the human genome, and among them, over a million SNPs have been identified. Thus, SNPs are the most frequent form of DNA polymorphism observed in humans. Most SNPs appear in noncoding regions and some SNPs are found in coding regions of genes (Figure 22.1).

The vast majority of SNPs are biallelic, although very rare triallelic and tetraallelic SNPs also exist. If an SNP originating from a spontaneous mutation occurs in the germ line, it can be inherited by offspring and can spread in the human population. As a result, both the wild type (the typical form of an allele occurring in nature) and a mutant allele are produced. This type of SNP is referred to as a biallelic SNP. Subsequently, a third mutation, a very rare event, may occur at the same nucleotide site. This SNP has a wild type and two different mutant alleles, which results in a triallelic SNP. It is also possible that a triallelic SNP is created by introducing two of the mutations simultaneously at a single site, which is also a very rare event.

Generally speaking, there are a number of advantages to utilizing SNP loci as markers for forensic applications. First, SNPs are abundant within the human genome; therefore, a sufficient number of SNP loci that are suitable for human identification can be selected. Second, in SNP analysis, amplified fragments (amplicons) are usually 50–100 base pairs in length and are smaller than that in short tandem repeat (STR) analysis (Chapter 20). Therefore, SNP profiling can be more useful than STR for analyzing degraded DNA samples in which genomic DNA is fragmented (Section 20.5.1). Third, SNP loci have lower mutation rates than STRs and are, therefore, useful for specialized human identification such as paternity testing. Lastly, it is possible to achieve high-throughput SNP analysis by utilizing multiplex SNP assays, which are amendable for automation.

The application of SNPs for forensic DNA analysis also has some disadvantages, in that SNP loci are not as polymorphic as STR loci. It is estimated that the analysis of 50–60 SNP loci is needed to achieve a similar level of population match probability (P_m—the lower the P_m, the less likely a match occurs between two randomly chosen individuals; see Chapter 25) using 13

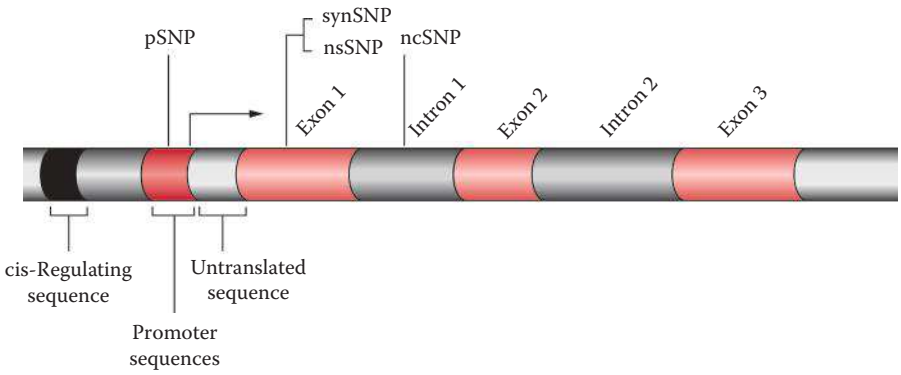

Figure 22.1 SNPs fall into several classes. Most reside in the noncoding regions of DNA and are designated as noncoding SNPs (ncSNPs). A subset of ncSNPs can also be found in introns. SNPs residing in exons are further divided into two types: The synonymous type (synSNP) is an exonic SNP that does not change the amino acid composition of the encoded polypeptide. Conversely, a nonsynonymous type (nsSNP) changes the encoded amino acid. SNPs in the promoter regions of the genome are known as promoter SNPs (pSNPs). Arrow: transcription start site. (© Richard C. Li.)

STR loci in CODIS. Moreover, most SNPs are biallelic; therefore, it is difficult to resolve mixed DNA profiles when a shared common allele exists in a mixture of more than one individual. Furthermore, most DNA databases contain STR profiles instead of SNP profiles, which means that it is not possible to carry out a database search for a matching SNP profile in cases where no suspect has yet been identified.

22.2 Forensic Applications of SNP Profiling

22.2.1 *HLA-DQA1* Locus

The first use of SNP-based profiling for forensic application involved sequence polymorphisms at the *HLA-DQA1* locus (formerly called the DQα locus). The *HLA-DQA1* gene is a member of the human leukocyte antigen (HLA) family, which contains a large number of genes involved in the immune response in humans. The *HLA-DQA1* locus is located within human HLA gene clusters on chromosome 6. The region tested for forensic use is highly polymorphic and is located at the second exon of the gene (Table 22.1).

22.2.1.1 DQα AmpliType and Polymarker Assays

The SNP profiling of the *HLA-DQA1* locus was carried out using the DQα AmpliType® kit. It was the first commercial kit, developed in the late 1980s by Cetus Corporation in Emeryville, CA. The *HLA-DQA1* panel can distinguish the following: alleles 1 (subtyped as 1.1, 1.2, or 1.3), 2, 3, and 4 (subtyped as 4.1 and 4.2/4.3, in which the 4.2 and 4.3 alleles are combined and cannot be distinguished). Therefore, a total of 28 possible genotypes from combinations of these alleles can be distinguished. Although the P_m of this SNP profiling is high (approximately 5×10^{-2}), it is useful as a preliminary test to quickly exclude suspects.

In addition to the *HLA-DQA1* locus, five additional loci—*LDLR, GYPA, HBGG,* D7S8, and *GC*—were utilized for forensic application in 1993 (Table 22.1). These loci were included in the second generation of the kit known as the AmpliType® PM PCR amplification and typing kit (also known as Polymarker), manufactured by Perkin-Elmer (Norwalk, CT). It consisted of one panel for the testing of *HLA-DQAl* and another panel for an additional five loci (Figure 22.2). Among these five additional loci, *LDLR, GYPA*, and D7S8 each have two alleles (designated

Table 22.1 Chromosomal Locations of SNP Loci Used in AmpliType® PM PCR Amplification and Typing Kit

Locus	Gene Product	Chromosome Location	Number of Alleles	Further Reading
HLA-DQA1	HLA-DQA1	6p21.3	7	Gyllensten and Erlich (1988)
LDLR	Low-density lipoprotein receptor	19p 13.1-13.3	2	Yamamoto et al. (1984)
GYPA	Glycophorin A	4q28-31	2	Siebert and Fukuda (1987)
HBGG	Hemoglobin G gammaglobin	11 p 15.5	3	Slightom et al. (1980)
D7S8	Anonymous	7q22-31.1	2	Horn et al. (1990)
GC	Group-specific component	4q11-13	3	Yang et al. (1985)

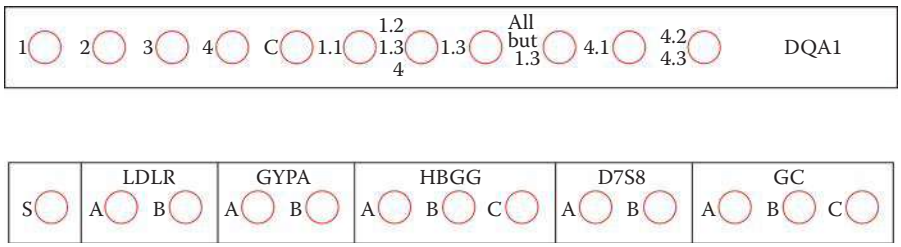

Figure 22.2 Panels of immobilized probes in the Polymarker kit. Top: *HLA-DQA1*. Bottom: additional five loci (*LDLR, GYPA, HBGG*, D7S8, and *GC*). C and S represent threshold control dots. (© Richard C. Li.)

A and B), while *HBGG* and *GC* each have three alleles (A, B, and C). As a result, the P_m of the Polymarker panels decreased to 10^{-4}.

The DNA profiles of the Polymarker system have been accepted in US courts. The Polymarker system has a number of advantages, particularly compared to variable number tandem repeat analysis (VNTR; see Chapter 19). First, the Polymarker system is a PCR-based method and is capable of analyzing a small amount of DNA sample (approximately 2 ng per analysis). Therefore, it is more sensitive than the method used in VNTR analysis. Second, the Polymarker system is an SNP assay with short amplicon sizes and, therefore, can analyze degraded DNA samples, which is not possible in VNTR analysis. Third, the Polymarker system is more rapid and less laborious than VNTR analysis. Lastly, the amplicon sizes of alleles at a given locus have identical lengths and, therefore, this assay does not exhibit preferential amplification as in amplified fragment length polymorphism (Chapter 19). However, the Polymarker system has its limitations compared to VNTR analysis. For example, compared to the VNTR analysis, the P_m of the Polymarker system is high, resulting in poor discrimination power in the comparison of two DNA profiles. Therefore, the Polymarker system is merely useful for excluding a suspect. Moreover, the SNP loci utilized in the Polymarker system are less polymorphic than VNTR loci. The limited number of alleles per SNP locus makes identifying the components of mixtures more difficult than with VNTR analysis in situations when contributors of a mixture sample share common alleles. For these reasons, the Polymarker system was replaced by STR profiling in the late 1990s.

22.2.1.2 Allele-Specific Oligonucleotide Hybridization

Both DQα AmpliType® and Polymarker kits utilize the *allele-specific oligonucleotide (ASO) hybridization* technique. This technique analyzes single-nucleotide variations, such as SNPs, at a given locus. It is based on the principle that ASO probes, usually 14–17 bases in length, hybridize to their complementary DNA sequences to distinguish known polymorphic alleles. ASO probes for multiple alleles at several loci can be arranged on the same panel to establish the presence or absence of specific alleles in PCR-amplified fragments of a DNA sample (Figure 22.3). Thus, the genotypes can be determined.

In the DQα AmpliType® and Polymarker kits, the oligonucleotides representing different alleles are immobilized to a solid matrix consisting of nylon membrane strips. Each immobilized probe at a particular site on the membrane is utilized to detect corresponding SNPs. Since the probe rather than the target DNA (as with regular blot format; see Chapter 9) is immobilized to the solid phase, the configuration used here is known as a *reverse blot* format.

The regions of the DNA in question are amplified by PCR. One of each pair of the primers is conjugated with biotin at the 5′ end (Figure 22.4). Thus, the amplicons are biotinylated for purposes of detection. This kit established multiplexing of a six-locus system, allowing simultaneous amplification of the *HLA-DQA1* locus along with *LDLR, GYPA, HBGG*, D7S8, and *GC* in

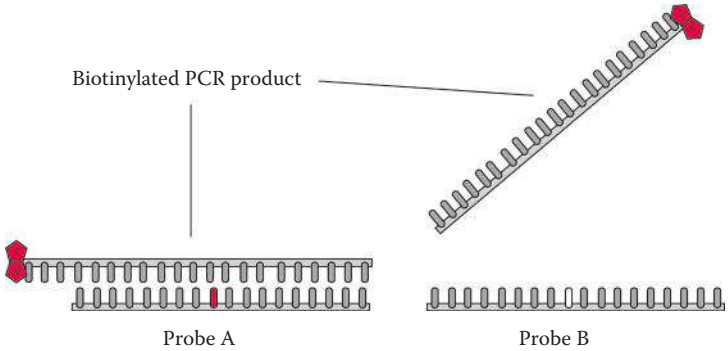

Figure 22.3 Hybridization with ASO. Two probes are incubated with the target DNA containing the SNP. Only the perfectly matched probe–target DNA can hybridize under optimal conditions.

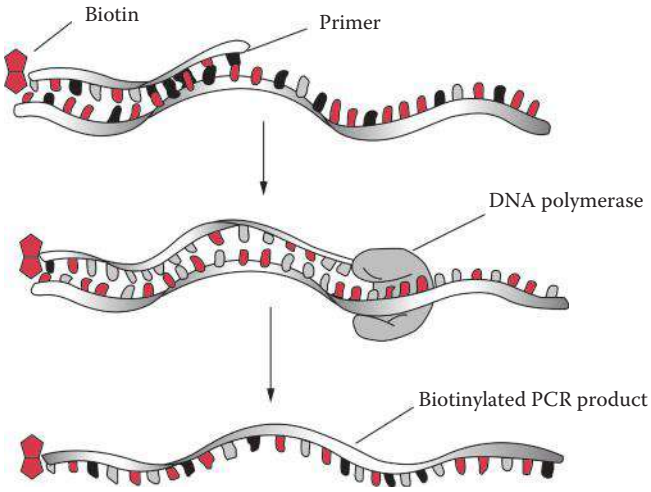

Figure 22.4 Amplification of DNA using biotinylated primers. The biotin is conjugated at the 5′ end of the primer. The amplified products are biotinylated. (© Richard C. Li.)

a single reaction for each sample. Following the denaturation of the PCR product to separate the two complementary DNA strands of the amplicons, the biotinylated strands are hybridized to the immobilized probes. Hybridization and washing conditions are established to ensure proper hybridization of the ASO probe and its target sequence. Unbound amplicons are washed away.

The presence of a PCR product bound to a specific probe can be detected by a colorimetric detection system (Figure 22.5). Since the amplicons are biotinylated, a horseradish peroxidase–conjugated streptavidin complex is utilized to bind to the biotinylated amplicons. The horseradish peroxidase then catalyzes the oxidation reaction of the colorless substrate, tetramethyl benzidine (TMB), into a blue precipitate at a designated location on the membrane strip, allowing the genotype of the sample to be determined (Figures 22.6 and 22.7).

The kit uses a threshold control (C dot in the *HLA-DQA1* panel and S dot in the panel of the other five loci) to distinguish between signal and background noise and to determine whether a

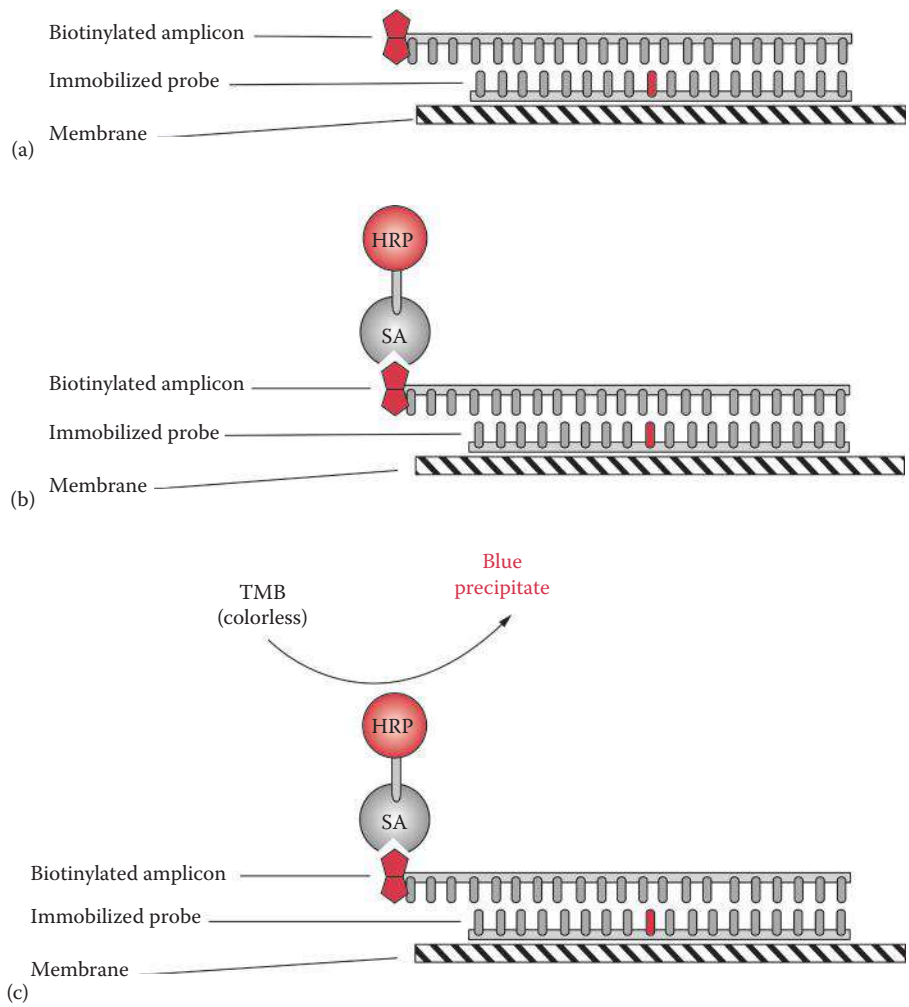

Figure 22.5 Reverse blot assay. (a) The probe is immobilized onto a solid-phase membrane and hybridized with a biotinylated PCR product having the target sequence. (b) The detection of hybridization is carried out by streptavidin (SA) and horseradish peroxidase (HRP) conjugate. (c) Colorimetric reaction is catalyzed by HRP using TMB as a substrate. (© Richard C. Li.)

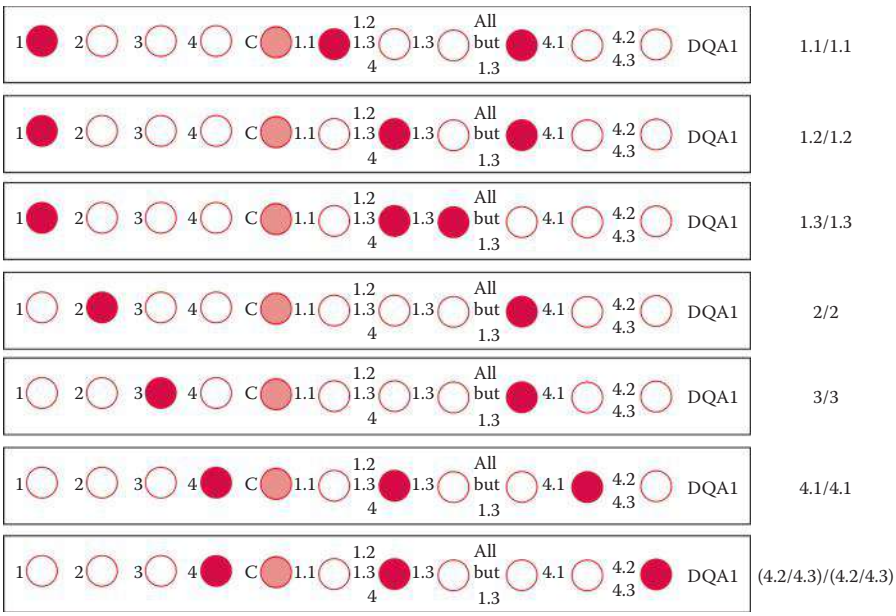

Figure 22.6 DNA typing of homozygous alleles at the *HLA-DQA1* locus using the reverse blot assay (Polymarker kit). The genotype of each sample is noted on the right. (© Richard C. Li.)

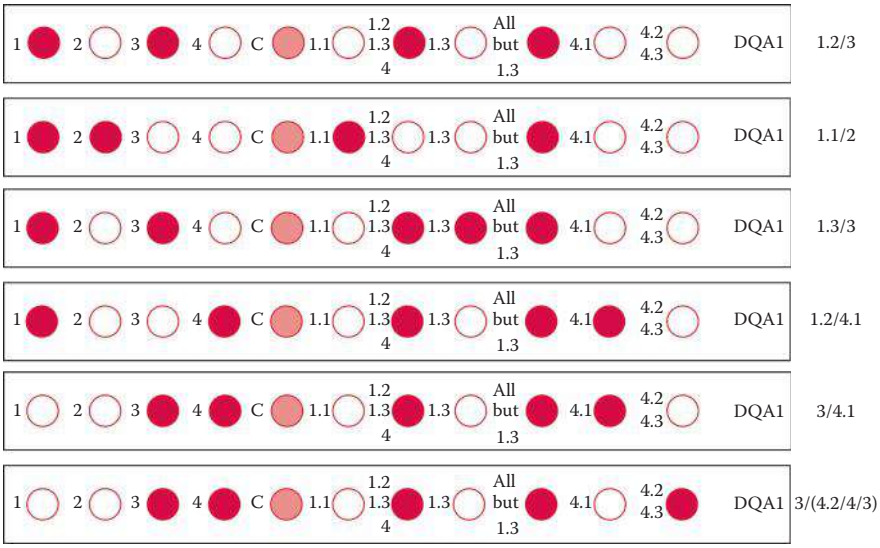

Figure 22.7 DNA typing of heterozygous alleles at the *HLA-DQA1* locus using the reverse blot assay (Polymarker kit). The representative genotypes of the samples are noted on the right. (© Richard C. Li.)

sufficient amount of DNA has been amplified to detect all the alleles present in a sample. If the signal intensities of allele dots are greater than or equal to the threshold control, the alleles are considered true. If the allele dots are less intense than the threshold control, they are considered inconclusive for the determination of full genotypes because an allele may not have been detected due to a low level of DNA template.

22.2.2 Current and Potential Applications of SNP Analysis

22.2.2.1 Application of SNP Analysis for Forensic Identification

The applications of SNP analysis for forensic human identification are summarized in Table 22.2. Autosomal SNP panels can be used for the most common types of forensic testing, including analysis of degraded DNA samples. The SNP panels for mitochondrial DNA (mtDNA) profiling are useful for identifying human remains by comparing the DNA profiles in question to those of potential relatives. This approach may serve as an alternative method to the DNA sequencing method of mtDNA analysis, which is time-consuming and laborious. SNP loci on the Y chromosome are also potentially useful markers for paternity testing because of low mutation rates. SNP loci such as ancestry informative markers (AIM) can be used to determine the ethnic origins of questioned samples to generate leads for investigations.

Table 22.2 Examples of Forensic Applications of SNP Profiling

Testing	Candidate SNP Loci	Application	Further Reading
Identity	Autosomal SNPs	Human identification	Kidd et al. (2006); Sanchez et al. (2006)
		Human identification via degraded DNA	Budowle (2004)
	mtDNA SNPs	Human identification	Grignani et al. (2006)
	Y chromosome SNPs	Paternity testing	Hammer et al. (2006)
Biogeographical origin	Ancestry informative markers (AIMs)	Ethnic group identification	Frudakis et al. (2003); Shriver and Kittles (2004)
Physical characteristics	*MC1R* (melanocortin 1 receptor gene)	Hair color identification (investigative lead)	Grimes et al. (2001)
	P (gene has role in pigmentation)	Eye color identification (investigative lead)	Rebbeck et al. (2002)
Pathology	*KCNH2* (cardiac potassium channel gene)	Determining cause of sudden death from cardiac arrhythmia long QT syndrome	Lunetta et al. (2002)
	SCN5A (encodes cardiac sodium channel gene)	Determining cause of sudden death from cardiac arrhythmia long QT syndrome	Burke et al. (2005)
Toxicology	*CYP2C19*, *CYP2D6*, *CYP3A4*, *CYP2E1* (drug metabolizing enzyme genes)	Investigation of drug overdose (including death)	Kupiec et al. (2006)

22.2.2.2 Potential Applications of SNPs for Phenotyping

One potential application of SNP analysis is in determining phenotypic information, also known as *phenotyping* (Table 22.2). The relevant SNP loci are usually *nonsynonymous SNPs* (*nsSNPs*); they reside in the exon and change the encoded amino acid, which leads to an altered phenotype. Phenotyping of a questioned sample can reveal physical characteristics of an individual, such as hair and eye color, to provide leads for investigations. A number of SNPs residing within the melanocortin 1 receptor gene (*MC1R*) are associated with red hair, fair skin, and freckles, while SNPs residing within the *P* gene, which play a role in pigmentation, are associated with eye color variations.

Phenotyping can also be employed in the area of forensic pathology. Cardiac arrhythmia long QT syndrome (LQTS) can cause sudden death. A number of LQTS-associated SNPs—for example, SNPs in *KCNH2* and *SCN5A* genes—have been shown to correlate to such deaths. Thus, these SNPs are potentially useful for investigating the causes of death. Finally, phenotyping also has applications in forensic toxicology. A number of SNPs in genes, such as *CYP2D6*, that are responsible for metabolizing drugs can serve as potential markers for postmortem investigations of drug-overdose cases.

22.3 SNP Techniques

Over the years, various techniques of SNP analysis have been developed and can be divided into several groups based on the mechanisms used: allele-specific hybridization, primer extension, oligonucleotide ligation, and invasive cleavage. In *allele-specific hybridization*, allele discrimination is based on an optimal condition allowing only the perfectly matched probe–target hybridization to form. *Primer extension* methods are based on the ability of DNA polymerase to incorporate specific deoxynucleotides (dNTPs) complementary to the sequence of the template DNA. *Allele-specific oligonucleotide ligation* is based on the condition that only the allelic probe perfectly matched to the target is ligated. In the *invasive cleavage* method, allelic discrimination is based on DNA sequence-specific cleavage by endonucleases. A number of detection methods can be utilized in SNP analysis, such as the measurements of fluorescence, luminescence, and molecular mass. Most assays are carried out in solutions or on solid matrices such as glass slides, chips, or beads. Table 22.3 summarizes the representative assays for SNP typing.

For decades, Sanger sequencing, using chain-termination chemistry, has been the standard method for DNA sequencing (Chapter 23). In recent years, *next-generation sequencing* (NGS), a rapidly developing technology, has had a profound impact on biology. Compared to Sanger sequencing, NGS is advantageous in that it can achieve substantially higher throughput, and at lower cost, than the Sanger method. For example, a human genome can now be sequenced in several days using NGS technologies. Although NGS is not yet widely utilized for forensic applications, it has great potential for forensic DNA analysis, particularly in SNP analysis.

22.3.1 Next-Generation Sequencing Technologies

NGS technologies have two categories of application: *de novo* sequencing and resequencing. In *de novo* genome sequencing, uncharacterized genomes or characterized genomes with substantial structural variations are sequenced. Sequence reads are assembled without any reference sequence. In *resequencing* applications, characterized genomes are sequenced. Sequence reads are assembled against an existing reference sequence to identify sequence polymorphisms. Thus, resequencing can potentially be used for forensic applications to detect polymorphisms associated with human identification. *Target resequencing* is a useful method of resequencing that can be utilized for forensic applications. Prior to sequencing, the genomic regions of interest from a DNA sample are selectively isolated through a method known as *enrichment*. Several target-enrichment strategies have been developed. PCR is the most widely used enrichment method.

Table 22.3 Techniques for SNP Assay

Basis of Technique	Representative Assay	Detection	Format	Notes	Further Reading
Allele-specific hybridization	Reverse Blot	Colorimetry	Membrane based	Useful for screen assay	Saiki et al. (1989)
	LightCycler® (Roche)	Fluorescence	Solution based	Applicable for automation; limited multiplexing capability	Lareu et al. (2001)
	TaqMan® (Applied Biosystems)	Fluorescence	Solution based	Applicable for automation; limited multiplexing capability	De la Vega et al. (2005)
	Molecular beacons	Fluorescence	Solution based	Applicable for automation; limited multiplexing capability	Tyagi et al. (1998)
	GeneChip® (Affymetrix)	Fluorescence	Chip based	Designed for genotyping large number of SNPs; exceeds forensic needs	Mei et al. (2000)
Primer extension	SNaPshot™ (Applied Biosystems)	Fluorescence	Solution based	Applicable for capillary electrophoresis and multiplexing	Sanchez et al. (2006)
	PinPoint (Applied Biosystems)	Mass spectrometry	Solution based	Multiplex capability lower than SNaPshot™	Haff and Smirnov (1997)
	Arrayed primer extension (APEX)	Fluorescence	Chip based	Low reproducibility	Pastinen et al. (1997)
	Pyrosequencing	Fluorescence	Solution based	May be applicable for analysis of mixed profiles; not applicable for automation; limited multiplexing capability	Ronaghi et al. (1998)
Allele-specific oligonucleotide ligation	SNPstream® UHT (Orchid-Gene Screen)	Fluorescence	Chip based	Developed for mass disaster cases	Bell et al. (2002)
	SNPlex™ (Applied Biosystems)	Fluorescence	Solution based	Applicable for multiplexing; requires higher amount of target DNA template than other PCR-based methods	De la Vega et al. (2005)
Invasive cleavage	Invader® (Third Wave Technology)	Fluorescence	Solution based	PCR not required; large amount of target DNA required	Lyamichev et al. (1999)

Source: Adapted from Sobrino, B., M. Brion, and A. Carracedo, SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci Int, 2005. 154(2-3): p. 181–94.; Budowle, B., SNP typing strategies. Forensic Sci Int, 2004. 146 Suppl: p. S139–42.

The PCR-based approach is highly effective in targeting genomic regions that are small in size. Nevertheless, a typical NGS involves several major steps, including sample, library, and template DNA preparation; template amplification; sequencing and detection; and base calling; as well as data analysis (Figure 22.8).

22.3.2 DNA Samples, Sequencing Library, and Template Preparation

For most sequencing applications, micrograms of purified DNA are needed. This requirement is still a challenge for forensic casework applications where nanograms of DNA are often obtained. Human autosomal genome sequencing usually requires converting a DNA sample into a sequencing library. Two types of sequencing libraries are usually used: mate-pair libraries and fragment libraries. A mate-pair library is often used in *de novo* sequencing applications, while the fragment libraries can be used for resequencing and forensic applications. To construct a fragment library, a DNA sample is fragmented using mechanical methods such as sonication, nebulization, or shearing; endonuclease digestion; or a transposon-based method. Subsequently, sequencing adapters containing primer-binding sites for universal PCR primers are ligated to both ends of the DNA fragments. In resequencing applications, multiplex sequencing of pooled samples is often carried out, which can improve the efficiency and reduce the costs of sequencing. For multiplex sequencing, an index tag containing a bar code can be ligated to each DNA fragment to allow it to be identified after sequencing. Indexing can minimize the risk of sample mix-ups and contaminations during the sample preparation.

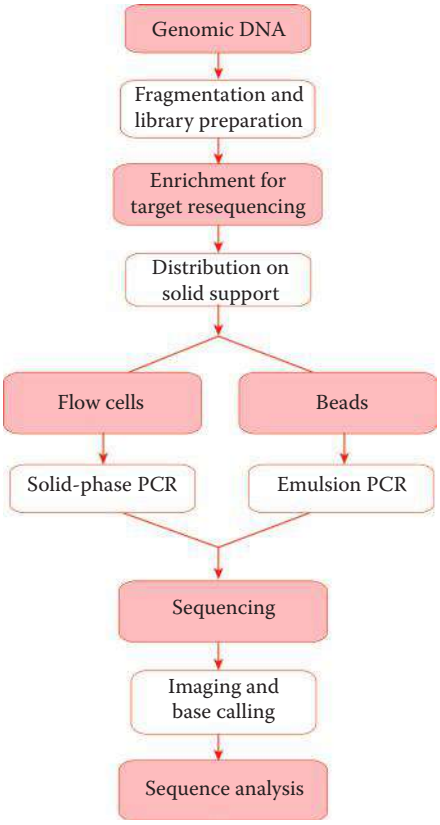

Figure 22.8 NGS work flow for potential forensic applications. (© Richard C. Li.)

Most imaging systems of sequencing platforms are not equipped to detect single-molecule fluorescence. Thus, DNA templates are amplified using either emulsion or solid-phase PCR. In the *emulsion PCR*, single-stranded DNA templates are bound to beads. DNA-bound beads are suspended in a water-in-oil emulsion, where each bead is placed in a single aqueous droplet. DNA fragments are amplified using PCR. As a result, each bead is coated with thousands of copies of the same template sequence. Subsequently, beads carrying amplified DNA are deposited on a glass slide or into individual fiber-optic wells for sequencing. In *solid-phase PCR*, both forward and reverse primers are attached to a slide (Figure 22.9). DNA templates are hybridized to immobilized PCR primers. The extended primer bends and hybridizes to a second immobilized primer, forming a bridge. During the "bridge amplification," the single-stranded template is amplified to form clusters. Solid-phase amplification can produce millions of spatially separated template clusters with free ends, allowing universal sequencing primers to hybridize for sequencing.

22.3.3 NGS Chemistry

The amplified template fragments are then sequenced. At present, there are two major categories of NGS chemistry that are routinely used: *sequencing by synthesis* and *sequencing by ligation*. The *pyrosequencing* technology is one example of NGS using the sequencing-by-synthesis chemistry. During the pyrosequencing, each nucleotide substrate is introduced one at a time. Only the correct nucleotide corresponding to the template is incorporated, and a pyrophosphate is released (Figure 22.10a). Pyrophosphate is then converted to ATP, catalyzed by sulfurylase (Figure 22.10b). ATP is then utilized by luciferase to convert luciferin to oxyluciferin, and the reaction emits light that is detected by a charge-coupled device camera (Figure 22.10c). This technology allows the generation of sequences with long read lengths. *Cyclic reversible termination* is another example of NGS using sequencing-by-synthesis chemistry. In this method, chain terminators (fluorescent-labeled dideoxynucleotides) are used to extend a primer sequence complementary to the template DNA. Each of four nucleotides is labeled with a different fluorophore. After incorporation and fluorescence detection, the terminating and fluorescent moieties are cleaved and removed (Figure 22.11). As a result, the next sequencing cycle is carried out. This process is repeated until the sequence is completed. This is the most commonly used method throughout the field of NGS.

In the sequencing-by-ligation chemistry, probes (eight nucleotides in length) are utilized for ligation reactions. Each probe contains five specific nucleotides that are complementary to the template and three nucleotides that are universal. Probes containing all possible combinations of the first five nucleotides are utilized. Each probe is fluorescently labeled according to the first two bases of the probe so that it can be identified. During the first round of sequencing reaction, only the probe that is complementary to the template sequence can hybridize. The probe is then ligated to the 3′ end of the primer. After detection, three universal nucleotides, including fluorophore, are cleaved. The ligation process is repeated several times using a new set of probes each time. The newly made complementary DNA strand is then stripped off. In the next round of sequencing, new primers that are one nucleotide shorter than the previous primer are utilized, and the ligation reaction is repeated. In this round of sequencing, different nucleotide positions are read. This process is repeated for several rounds with different primers, until all nucleotides in the template have been sequenced twice. As a result, this sequencing method has a low error rate.

22.3.4 NGS Coverage

In NGS, sequencing reads are usually not distributed evenly over the genomic regions of interest. As a result, some nucleotides will be covered by fewer or more reads than the average. Therefore, multiple reads for each nucleotide are necessary to obtain a reliable sequence. The average number of times that each nucleotide in the genomic regions of interest is sequenced is known as the

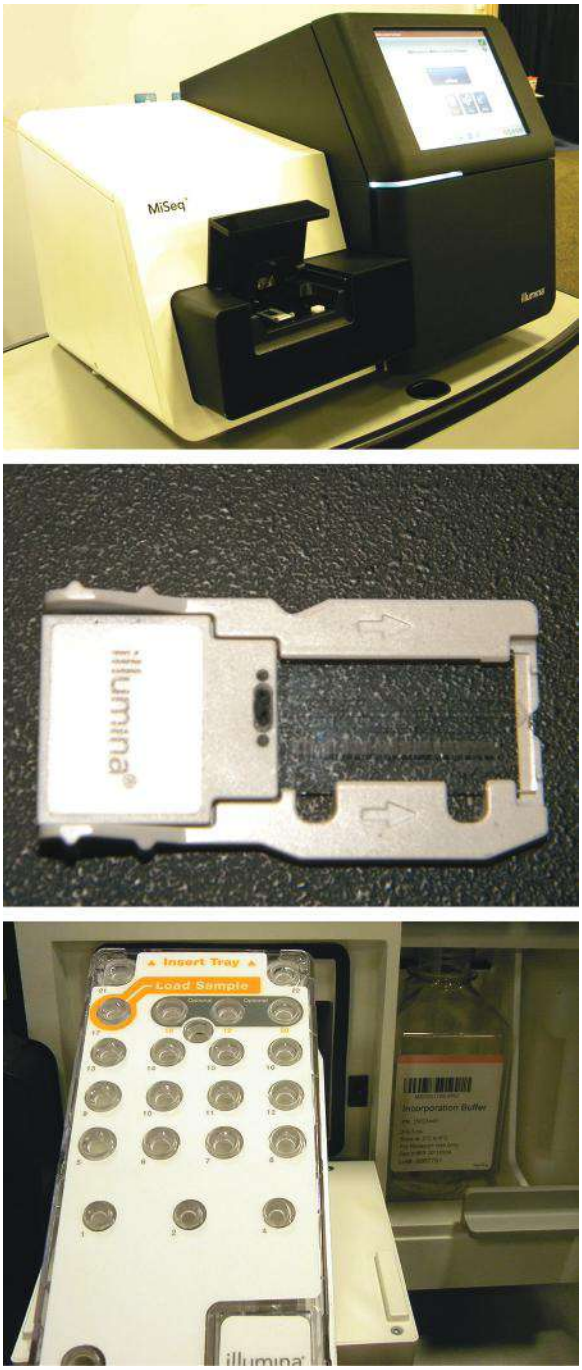

Figure 22.9 A photo of an NGS platform. A MiSeq (Illumna) desktop sequencer (top), a flow-cell device (middle) where DNA templates are distributed for PCR, and a reagent cartridge (bottom) are shown. (© Richard C. Li.)

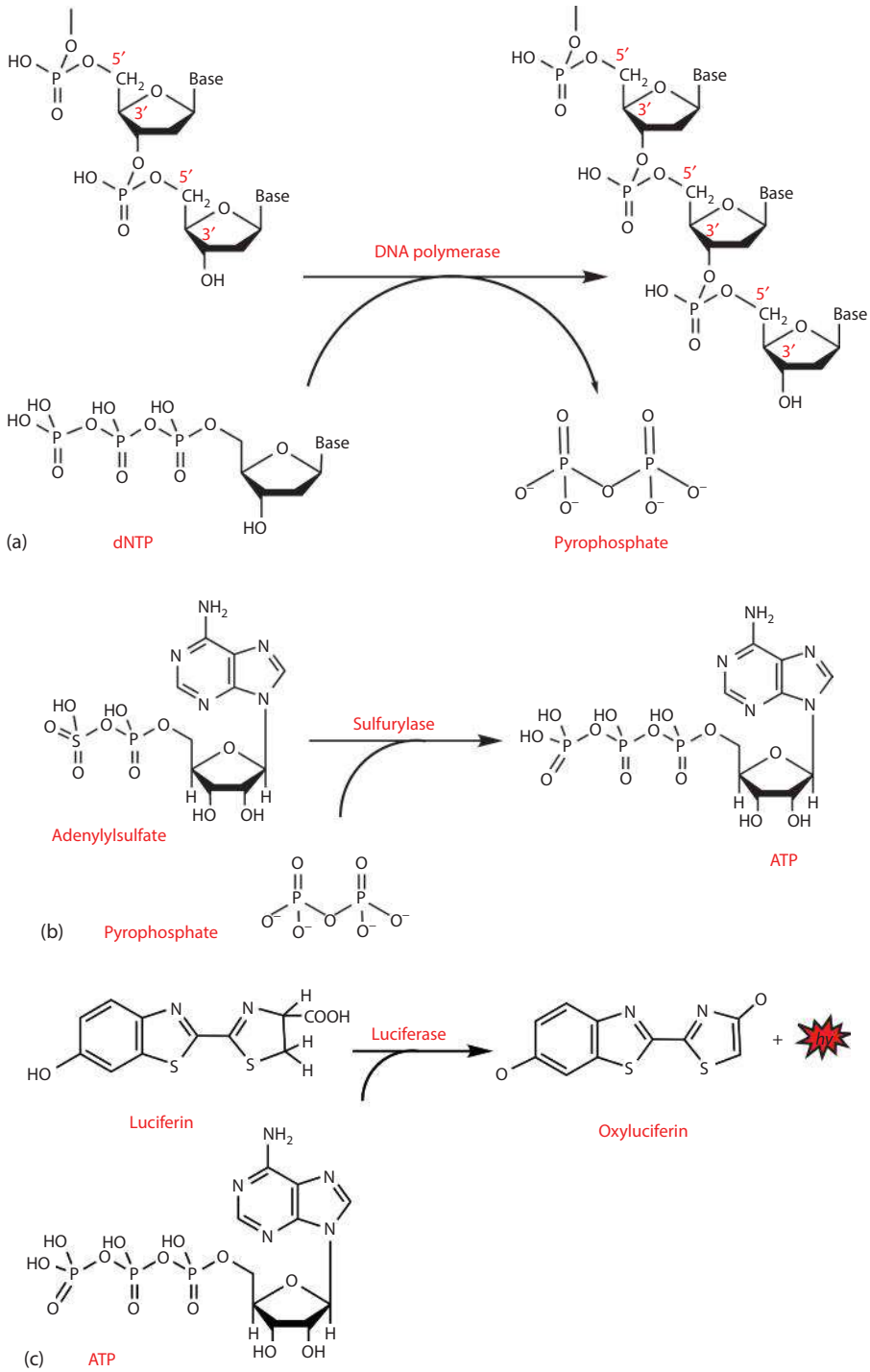

Figure 22.10 Pyrosequencing chemistry. (a) During the pyrosequencing, a nucleotide substrate is incorporated, releasing a pyrophosphate. (b) Pyrophosphate is then converted to ATP. (c) ATP is utilized by luciferase, converting luciferin to oxyluciferin, which emits light.

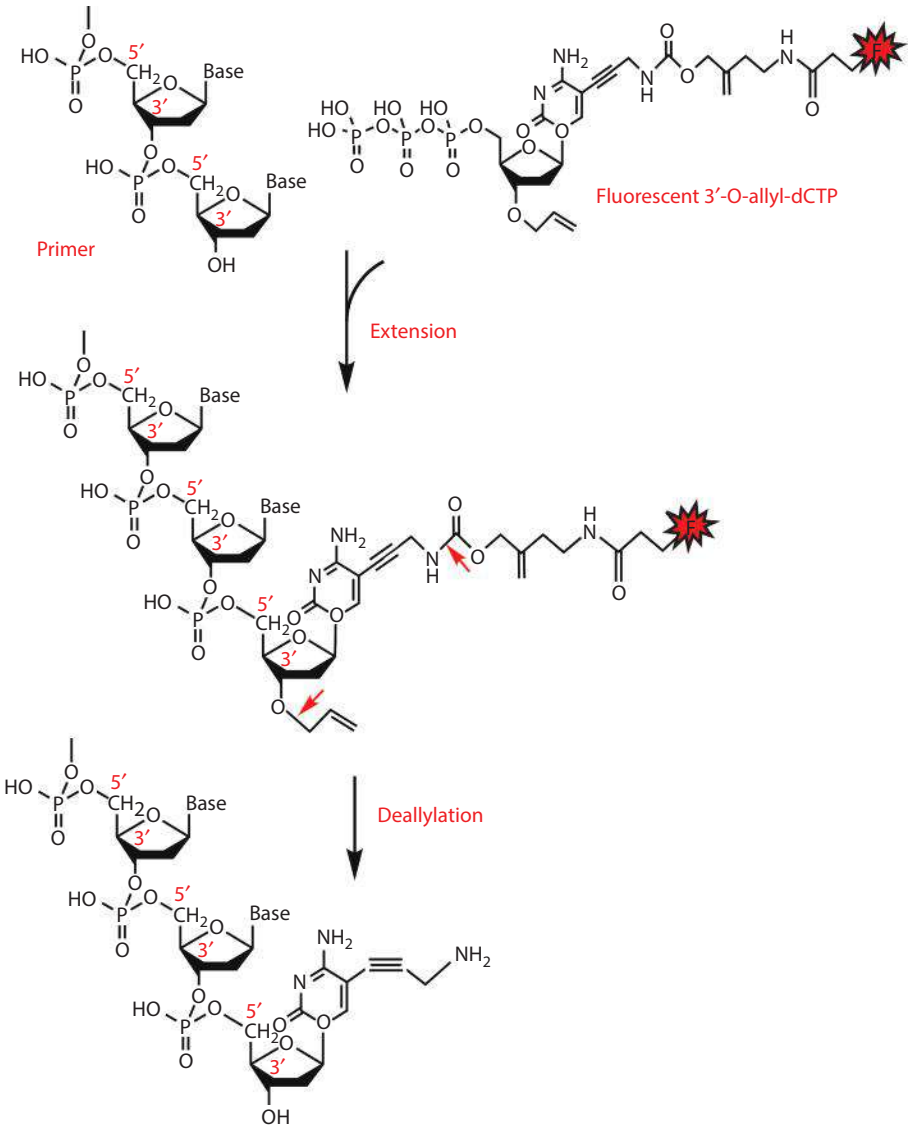

Figure 22.11 Cyclic reversible termination. A single sequencing cycle is shown. The incorporation of a C is illustrated as an example.

coverage or the *sequencing depth*. The level of coverage depends on the types of applications. For example, for human genome SNP analysis, the level of coverage is from 10 to 30. The coverage can be calculated using the following equation:

$$C = \frac{LN}{G}$$

where:

C = coverage
G = haploid genome length
L = read length
N = number of reads

Currently, NGS has not been widely used in forensic applications. One disadvantage of NGS is that the error rate of NGS is higher than that of Sanger sequencing. Although the error rate of NGS is not yet acceptable for forensic casework analysis, it can potentially be overcome by increasing the sequencing coverage.

Bibliography

Ackerman, M.J., D.J. Tester, and D.J. Driscoll, Molecular autopsy of sudden unexplained death in the young. *Am J Forensic Med Pathol*, 2001, **22**(2): 105–111.

Ali, M.E., et al., HLA-A, -B and -DRB1 allele frequencies in the Bangladeshi population. *Tissue Antigens*, 2008, **72**(2): 115–119.

Alkhayat, A., F. Alshamali, and B. Budowle, Population data on the PCR-based loci LDLR GYPA, HBGG, D7S8, Gc, HLA-DQA1, and D1S80 from Arabs from Dubai. *Forensic Sci Int*, 1996, **81**(1): 29–34.

Allah, R., L. Yang, and S.B. Li, SNPs and forensic DNA typing. *Fa Yi Xue Za Zhi*, 2007, **23**(5): 373–379.

Allwood, J.S. and S. Harbison, SNP model development for the prediction of eye colour in New Zealand. *Forensic Sci Int Genet*, 2013, **7**(4): 444–452.

Altshuler, D., et al., An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, 2000, **407**(6803): 513–516.

Amigo, J., et al., The SNPforID browser: An online tool for query and display of frequency data from the SNPforID project. *Int J Legal Med*, 2008, **122**(5): 435–440.

Arnestad, M., et al., Prevalence of long-QT syndrome gene variants in sudden infant death syndrome. *Circulation*, 2007, **115**(3): 361–367.

Babol-Pokora, K. and J. Berent, SNP-minisequencing as an excellent tool for analysing degraded DNA recovered from archival tissues. *Acta Biochim Pol*, 2008, **55**(4): 815–819.

Bandelt, H.J. and A. Salas, Current next generation sequencing technology may not meet forensic standards. *Forensic Sci Int Genet*, 2012, **6**(1): 143–145.

Barbaro, A., et al., Genetic variability of the SNPforID 52-plex identification SNP panel in Italian population samples. *Forensic Sci Int Genet*, 2012, **6**(6): e185–e186.

Bell, B., et al., Distribution of types for six PCR-based loci; LDLR, GYPA, HBGG, D7S8, GC and HLA-DQA1 in central Pyrenees and Teruel (Spain). *J Forensic Sci*, 1997, **42**(3): 510–513.

Bell, P.A., et al., SNP stream UHT: Ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Bio Techniques*, 2002, **74**(Suppl 70-2): 76–77.

Birngruber, C., F. Ramsthaler, and M.A. Verhoff, The color(s) of human hair—Forensic hair analysis with SpectraCube. *Forensic Sci Int*, 2009, **185**(1–3): e19–e23.

Blake, E., et al., Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)-DQ alpha oligonucleotide typing on biological evidence samples: Casework experience. *J Forensic Sci*, 1992, **37**(3): 700–726.

Blanco-Verea, A., et al., Forensic validation and implementation of Y-chromosome SNP multiplexes. *Forensic Sci Int Genet Suppl*, 2008, **1**(1): 181–183.

Borsting, C. and N. Morling, Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. *Forensic Sci Int Genet*, 2011, **5**(3): 236–241.

Borsting, C. and N. Morling, Reinvestigations of six unusual paternity cases by typing of autosomal single-nucleotide polymorphisms. *Transfusion*, 2012, **52**(2): 425–430.

Borsting, C., C. Tomas, and N. Morling, SNP typing of the reference materials SRM 2391b 1-10, K562, XY1, XX74, and 007 with the SNPforID multiplex. *Forensic Sci Int Genet*, 2011, **5**(3): e81–e82.

Borsting, C., C. Tomas, and N. Morling, Typing of 49 autosomal SNPs by single base extension and capillary electrophoresis for forensic genetic testing. *Methods Mol Biol*, 2012, **830**: 87–107.

Borsting, C., E. Rockenbauer, and N. Morling, Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. *Forensic Sci Int Genet*, 2009, **4**(1): 34–42.

Borsting, C., H.S. Mogensen, and N. Morling, Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. *Forensic Sci Int Genet*, 2013, **7**(3): 345–352.

Borsting, C., et al., Performance of the SNPforID 52 SNP-plex assay in paternity testing. *Forensic Sci Int Genet*, 2008, **2**(4): 292–300.

23

Mitochondrial DNA Profiling

Forensic mitochondrial DNA (mtDNA) analysis is an important tool for human identification and is especially useful for identifying victims such as missing persons and individuals in mass fatality cases. Because mtDNA is maternally inherited, the mtDNA profiles of these individuals can be compared to those of their maternal relatives, and thus these individuals can be identified. Additionally, cells contain a much higher copy number of the mtDNA genome than that of the nuclear genome. Therefore, mtDNA testing is frequently used to analyze evidence samples, such as hair shafts, that contain low amounts of nuclear DNA. Furthermore, buried bones and decomposed tissues, in which nuclear DNA may be degraded, can be tested with mtDNA analysis.

23.1 Human Mitochondrial Genome

Mitochondria are cellular organelles that serve as the energy-generating components of cells (Figure 23.1). Each cell contains hundreds of mitochondria, which have their own extrachromosomal genomes separated from the nuclear genomes. Although each human mitochondrion contains multiple copies of the mtDNA genome, the exact copy number varies for each cell. However, it is estimated that hundreds of copies of the mtDNA genome exist in most cells.

23.1.1 Genetic Contents of Mitochondrial Organelle Genomes

The human mitochondrial genome was first sequenced by Fred Sanger's laboratory at Cambridge University and was published in 1981; it is known as the *Cambridge reference sequence* (CRS). The sequence was largely derived from a placental sample from an individual of European descent and also partially from HeLa cells (a cell line derived from cervical cancer cells), as well as from bovine cells. It was later discovered, by resequencing the original mtDNA sample, that the CRS contains substitution errors at 10 nucleotide positions. The *revised Cambridge reference sequence* (rCRS) was published in 1999 and presented corrections to these substitution errors. Additionally, CRS contains a cytosine dimer at nucleotide positions 3106 and 3107, which is in fact a single cytosine nucleotide. This error is not corrected in the rCRS in order to retain the original nucleotide numbering system of CRS and thus to avoid inconsistency with the previous literature. The human mitochondrial DNA genome is circular with no beginning and end, which can make sequence comparison a potential problem. Therefore, the rCRS, with its nucleotide numbering system, is used as a reference when aligning with other mitochondrial sequences for comparison purposes.

Organelle genomes are usually much smaller than their nuclear counterparts. The human mtDNA genome consists of 16,569 base pairs (bp) containing 37 genes (Figure 23.2). Thirteen of

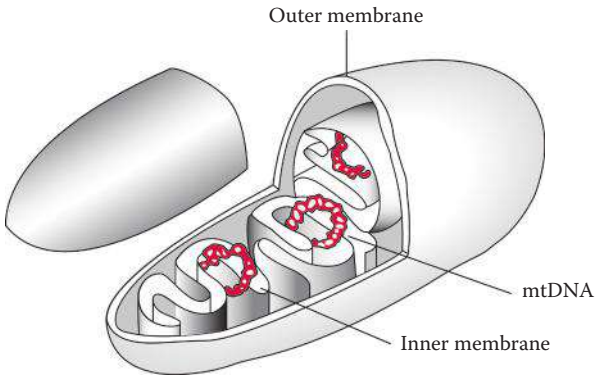

Figure 23.1 Mitochondrion. mtDNA, mitochondrial DNA. (© Richard C. Li.)

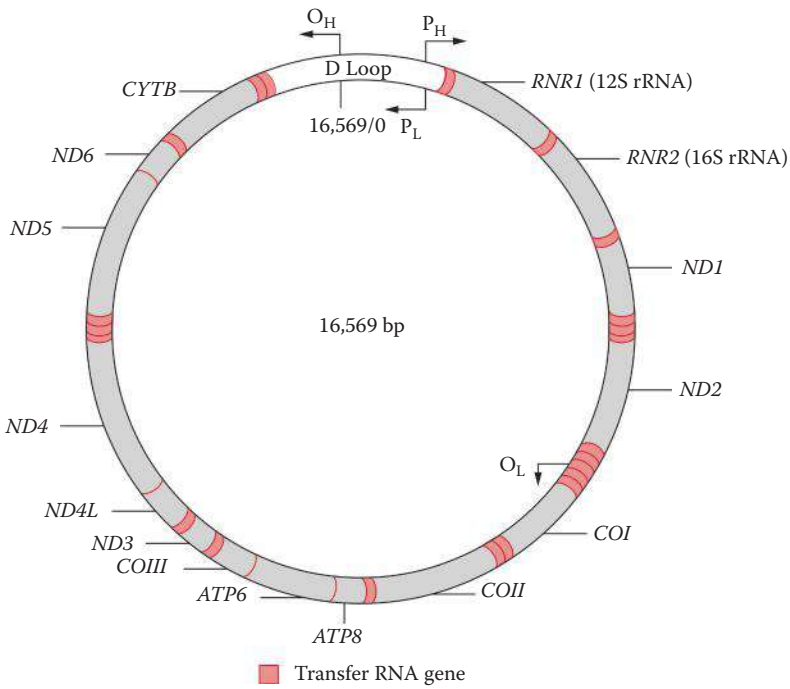

Figure 23.2 Human circular mitochondrial genome. The transcription direction for the H (heavy) and L (light) strands are indicated by arrows (P_H, P_L, respectively). The origins of replication are labeled O_H for the heavy strand and O_L for the light strand. The mitochondrial DNA genome encodes genes. ND, NADH coenzyme Q oxidoreductase complex; CO, cytochrome c oxidase complex; CYTB, cytochrome b; ATP, ATP synthase; rRNA, ribosomal RNA. Transfer RNA genes are shown as indicated. (© Richard C. Li.)

these genes code for proteins involved in the respiratory complex, a main energy-generating component in mitochondria. The other 24 specify noncoding RNA molecules required for the expression of the mitochondrial genome. The genes in the human mitochondrial genome are much more closely packed than in the nuclear genome and contain no introns. A *control region*, also known as a *displacement loop* (*D loop*), contains the origin of replication for one of the mtDNA strands but does not code for any gene products (Figure 23.2). An asymmetric distribution of

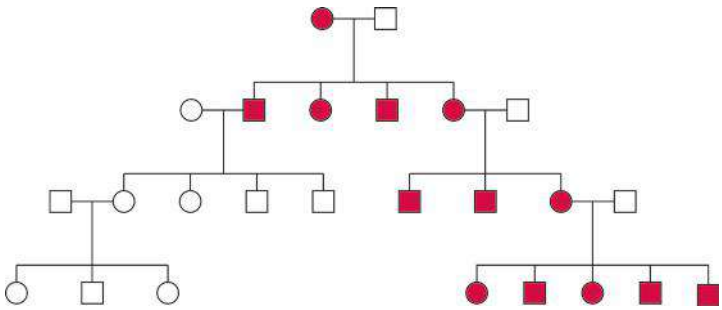

Figure 23.3 Pedigree of a human family showing inheritance of mtDNA. Females and males are denoted by circles and squares, respectively. Red symbols indicate individuals who inherited the same mtDNA.

nucleotides gives rise to light (L) and heavy (H) strands. The H strand contains a greater number of guanine nucleotides and a higher molecular weight in comparison to the L strand.

23.1.2 Maternal Inheritance of mtDNA

Maternal inheritance is typically observed for the mtDNA genome (Figure 23.3), which is inherited differently from nuclear genes. The inheritance of the mtDNA genome does not obey the rules of Mendelian inheritance and is thus called non-Mendelian inheritance.

The mitochondria of a spermatozoon are located in the midpiece (Chapter 14). At conception, only the head portion of a spermatozoon (containing a nucleus but no mitochondria) enters the egg. The fertilized egg contains the maternal mitochondria, which are transmitted to the progeny. Occasionally, the paternal mitochondria can enter the cell. However, paternal mitochondria in the spermatozoon that enter the egg are usually destroyed by the egg cell after fertilization (Figure 23.4). Therefore, the coinheritance of maternal and paternal mtDNA in a single individual is extremely rare in humans. The mtDNA sequence is identical for relatives within the same maternal lineage, a property that is useful when identifying individuals by comparing their mtDNA with that of maternal relatives.

Homologous DNA recombination (Chapter 25) has not been observed in the mtDNA genome. Thus, an mtDNA profile, also referred to as the *mitotype*, is considered a haplotype treated as a single locus. The mitochondrial genome has a higher mutation rate (up to 10 times higher) than its nuclear counterpart. The presence of mutations can be problematic in victim identification when comparing the mtDNA profiles of a victim with the relatives of the victim.

23.2 mtDNA Polymorphic Regions

23.2.1 Hypervariable Regions

The most polymorphic region of mtDNA is located within the D-loop (Figure 23.5). The three *hypervariable regions* in the D-loop are designated hypervariable region I (HV1: 16,024–16,365; 342 bp), hypervariable region II (HV2: 73–340; 268 bp), and hypervariable region III (HV3: 438–574; 137 bp). The most common polymorphic regions of the human mtDNA genome analyzed for forensic purposes are HV1 and HV2.

23.2.2 Heteroplasmy

Heteroplasmy occurs when an individual carries more than one mtDNA haplotype. Heteroplasmy may be observed with one kind of tissue and may be absent in other kinds of tissues; for example,

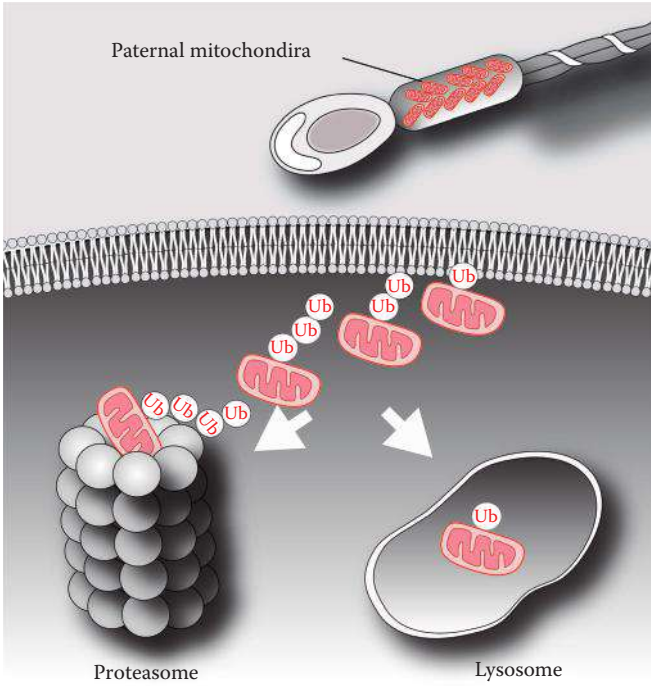

Figure 23.4 A model of uniparental mtDNA inheritance in humans. Paternal mitochondria occasionally enter the egg cell. Paternal mitochondria inside fertilized eggs are tagged by ubiquitin protein (Ub). It is proposed that the ubiquitination of sperm mitochondria leads to the degradation of paternal mitochondria in fertilized eggs. Tagged Ub can be recognized by proteasomes and lysosomes, which are cellular degradation machineries. A polyubiquitin chain with at least four ubiquitin units is needed to be recognized by the proteasomes. Ubiquitin with less than four ubiquitin units can be processed by the lysosomes. (© Richard C. Li.)

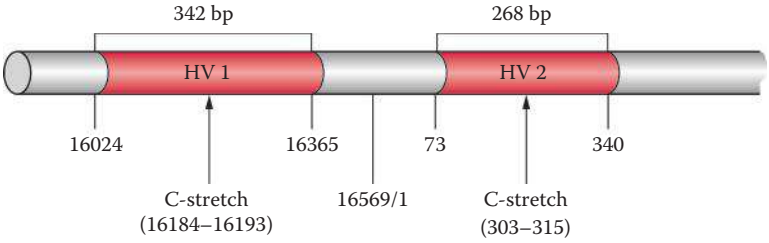

Figure 23.5 Hypervariable regions of the D-loop in mtDNA (with nucleotide positions). (© Richard C. Li.)

mtDNA heteroplasmy is commonly observed in hair samples (Chapter 4). Additionally, an individual may exhibit one mitotype in one tissue and a different mitotype in another. Thus, it is necessary to obtain and process additional samples to confirm the heteroplasmy when it is observed in a questioned sample but not in a known sample or vice versa. The two types of heteroplasmies are sequence and length heteroplasmy.

Sequence heteroplasmy is defined as the presence of two different nucleotides at a single position shown as overlapping peaks in a sequence electropherogram (Figure 23.6). Heteroplasmy usually occurs at one position, but on rare occasions it can be observed at more than one position. Hot spots for heteroplasmy have been documented at both HV1 and HV2 regions. Heteroplasmy may complicate the interpretation of mtDNA results, but its presence can also

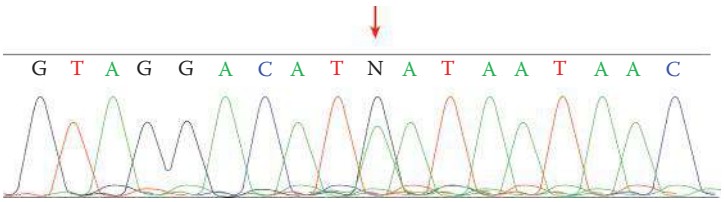

Figure 23.6 Electropherogram showing mtDNA sequence heteroplasmy at position 234R (A/G) as indicated by an arrow. N, unresolved sequence. (© Richard C. Li.)

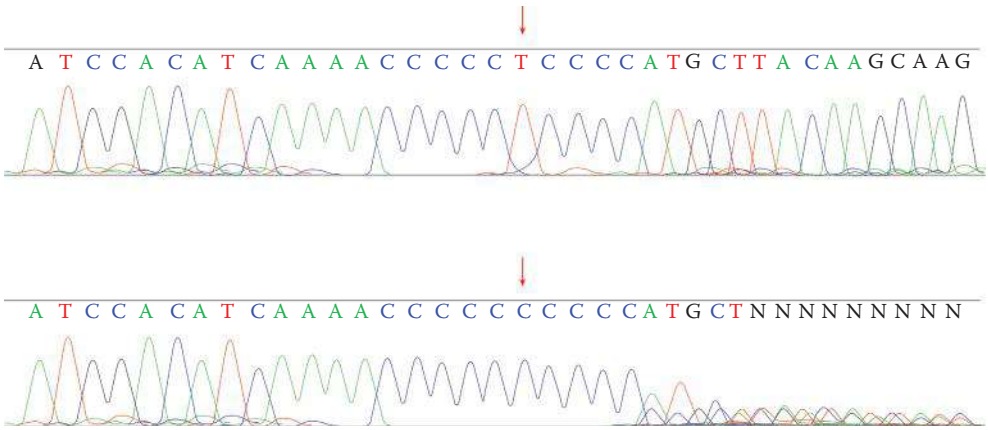

Figure 23.7 Length heteroplasmy. Electropherogram showing mtDNA length heteroplasmy at the C stretch of the HV1 region where position 16,189 is a T (top) and a C (bottom) as indicated by an arrow. N, unresolved sequence. (© Richard C. Li.)

improve the strength of a match. Both HV1 and HV2 of the human mtDNA D-loop region contain homopolymeric cytosine sequences known as *C stretches*. The HVI region contains a C stretch between nucleotide positions 16,184 and 16,193, interrupted by a thymine at position 16,189. If a base transition from T to C occurs at position 16,189 (a variant present in approximately 20% of the population), it results in an uninterrupted C stretch. A similar C stretch resides between positions 303 and 315 of the HV2 region.

Length heteroplasmies are often observed at the uninterrupted C stretches in sequencing, in which sequencing products with various lengths of polymeric cytosine residues are present. As a result, sequences downstream from the C stretch cannot be resolved (Figure 23.7). It is not clear whether the length heteroplasmy is due to replication slippage at the C stretches or results from the presence of a mixture of length variants in the cells. If length heteroplasmy occurs, alternative sequencing primers that anneal at the downstream of C stretches can be used to obtain the downstream sequences of the C stretches.

23.3 Forensic mtDNA Testing

23.3.1 General Considerations

mtDNA analysis is often used on samples derived from skeletal or decomposed remains. The surface of the sample should be cleaned to remove any adhering debris or contaminants. Bones and teeth are pulverized to facilitate extraction of the mtDNA (Chapter 5). Duplicate extractions (e.g., two sections of a single hair) are recommended if sufficient sample material is available.

mtDNA is extracted using a similar method to nuclear DNA (nuclear DNA is coextracted with mtDNA). The amount of mtDNA can therefore be estimated from the quantity of nuclear DNA obtained. Alternatively, mtDNA-specific quantization methods using real-time PCR (Chapter 6) can also be used to directly obtain measurements of mtDNA extracted.

For mtDNA sequencing, the analysis of both strands of the mtDNA in a given region must be performed to ensure accuracy. Due to the high sensitivity of mtDNA analysis, it is essential to minimize risks of contamination during the procedure. Contamination must be strictly monitored using proper controls such as extraction reagent blanks (Section 7.5.3) and amplification negative controls (samples containing all reagents except DNA template).

Finally, a positive control must also be used to monitor the success of the analysis. It should be introduced at the amplification step and remain through the sequencing process. A positive control consists of a DNA template of known sequence, such as DNA purified from the HL60 cell line.

23.3.2 mtDNA Screen Assay

One example of the assay for screening mtDNA variations is the allele-specific oligonucleotide (ASO) assay. It allows initial screening of mtDNA sequence polymorphisms and has the potential to reduce the number of samples required for mtDNA sequencing. This method is also useful for excluding or eliminating suspects from a case. However, HV1 and HV2 sequencing should be performed to obtain complete sequence information for the targeted HV regions to confirm a match.

The commercial Linear Array™ mtDNA HV1/HV2 region sequence typing kit (Roche Applied Sciences, Indianapolis) utilizes reverse ASO configuration with a panel of immobilized ASO probes that detect common polymorphic sites (Figures 23.8 and 23.9). The mtDNA is amplified at both HV1 (444 bp amplicon) and HV2 (415 bp amplicon) regions and the forward primers are biotin labeled at the 5′ ends of the oligonucleotides. Thus, the amplified PCR product (amplicon) is biotinylated. A horseradish peroxidase–conjugated streptavidin complex then binds to the biotinylated amplicon (see Chapter 9). Finally, colorimetric detection is carried out with tetramethylbenzidine (TMB) as the substrate to produce a colored precipitate at the designated location. The typing kit detects sequence variations in 19 positions within the HV1 and HV2 regions.

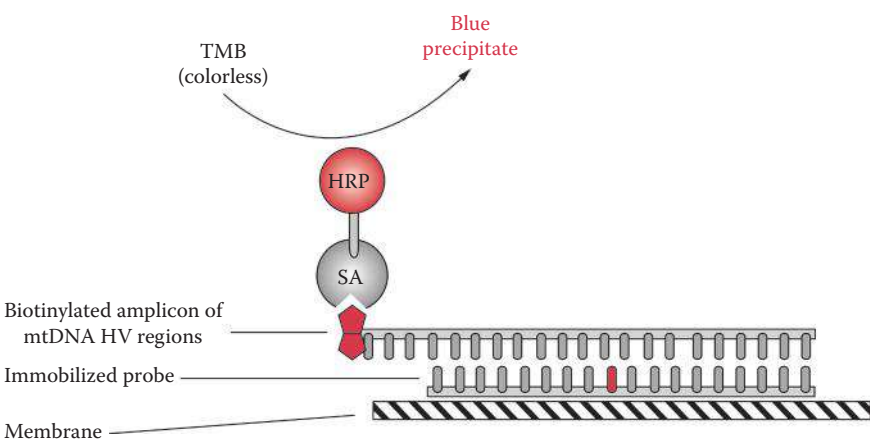

Figure 23.8 Reverse blot assay employed in mtDNA screen. A probe is immobilized onto a solid-phase membrane, and then hybridized with a biotinylated amplicon of the mtDNA HV sequences. Hybridization is detected by a streptavidin (SA) and horseradish peroxidase (HRP) conjugate. A colorimetric reaction is catalyzed by HRP using tetramethylbenzidine (TMB) as a substrate. (© Richard C. Li.)

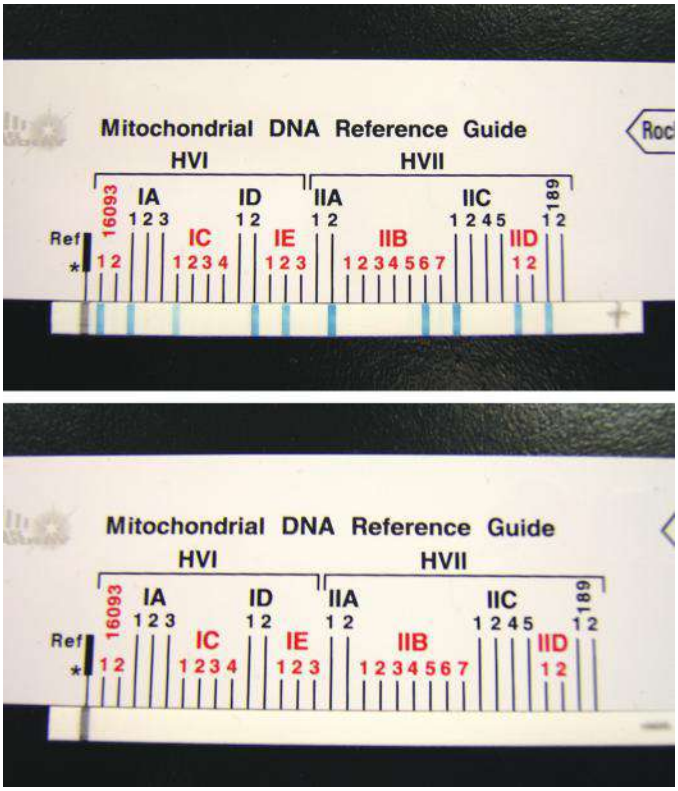

Figure 23.9 Linear array mtDNA assay results (top) and negative control (bottom). (© Richard C. Li.)

23.3.3 mtDNA Sequencing

To sequence a specific region of mtDNA, a combination of PCR amplification and DNA sequencing techniques is utilized to reduce the time and labor needed to obtain DNA sequences from genomic DNA templates. mtDNA sequencing usually consists of (1) PCR amplification, (2) DNA sequencing reactions, (3) separation using electrophoresis, and (4) data collection and sequence analysis (Figure 23.10). Chapter 25 describes the evaluation of the strength of the results via statistical analysis.

23.3.3.1 PCR Amplification

The extracted DNA samples must be amplified to yield sufficient quantities of template for sequencing reactions. PCR amplification of all or a part of the D-loop region can be carried out with various primer sets. If a sample contains intact mtDNA, the HV1 and HV2 regions can be amplified as two amplicons, each of about 350–400 bp in length. If an mtDNA sample is fragmented due to degradation, the hypervariable regions can be amplified as smaller amplicons. PCR amplification of mtDNA is usually done in 34–38 cycles. Protocols for highly degraded DNA specimens sometimes require 42 cycles. The use of higher PCR cycle numbers can improve the yield of the amplicons.

Following mtDNA amplification, a purification step is necessary to remove excess primers and deoxynucleotide triphosphates (dNTPs). This step can be performed using filtration devices such as a Microcon® to remove small molecules from the sample or using nuclease digestion with shrimp alkaline phosphatase or exonuclease I to degrade remaining primers and dNTPs. The concentration of the amplicons is important for an optimal sequencing reaction in the next phase of mtDNA sequencing. The quality and quantity of the mtDNA amplicon must be evaluated to confirm the presence or absence of amplicons and their concentrations. This can be done

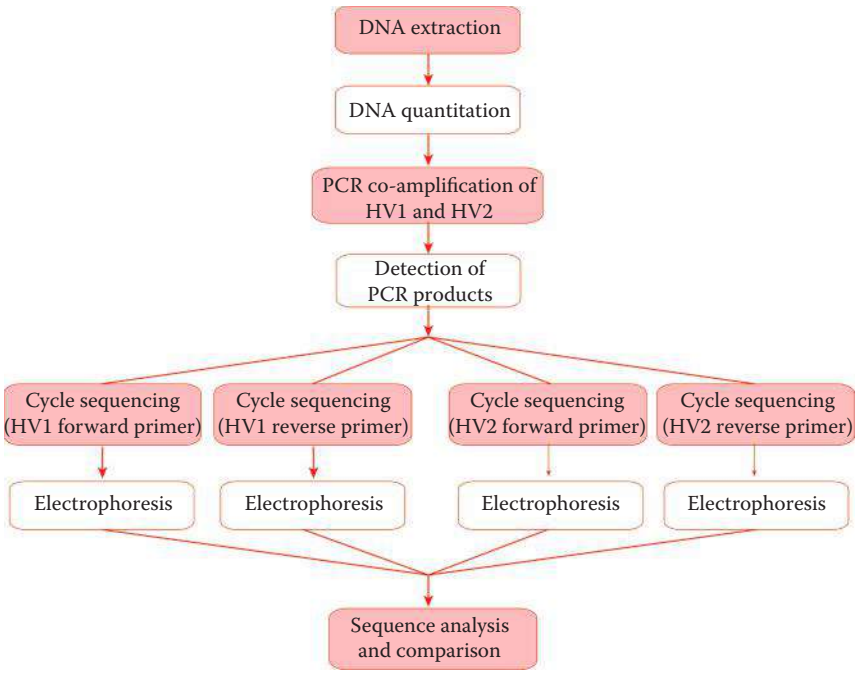

Figure 23.10 Sanger sequencing work flow for the mtDNA HV regions. (© Richard C. Li.)

using an agarose yield gel to visualize the amplicons of the sample or via capillary electrophoresis using a modular microfluidic chip device (Figure 8.12), a more informative method, for quantifying amplicons.

23.3.3.2 DNA Sequencing Reactions

The best-known DNA sequencing techniques are the *chain-termination* method and the *chemical degradation* method developed by Sanger and Gilbert (who shared the Nobel Prize for their work), respectively, in 1977. Over the years, the chain-termination method has become more common because it is suitable for automation and does not require the toxic chemicals necessary for the chemical degradation method.

23.3.3.2.1 Chain-Termination or Sanger Method

Using the chain-termination method, an oligonucleotide primer that can anneal to a single-stranded DNA template is utilized. A sequencing reaction contains DNA polymerase and the four dNTPs in order to carry out extension. The reaction also contains small quantities of dideoxynucleotide triphosphates (ddNTPs, Figure 23.11). Thus, a sequencing reaction involves a combination of extension and termination of the chain (Figure 23.12). If a ddNTP molecule is incorporated into a growing DNA chain, the absence of a 3′ OH group in the ddNTP molecule prevents the formation of a phosphodiester bond and thus disrupts the extension of the oligonucleotide chain.

The ratio of ddNTPs to dNTPs has been optimized to result in a collection of DNA fragments varying in length by one nucleotide from the primer length to the full length of the sequencing reaction product. As a result, the products of the sequencing reaction consist of a pool of various lengths of oligonucleotide chains terminated by ddNTPs. By using the four different ddNTPs, populations of DNA fragments are generated that terminate at positions occupied by every A, C, G, or T in the template strand.

The sequencing product of chain termination can be labeled with the *dye-terminator* system, in which the terminator is labeled, or with the *dye-primer* system, in which the primer is labeled.

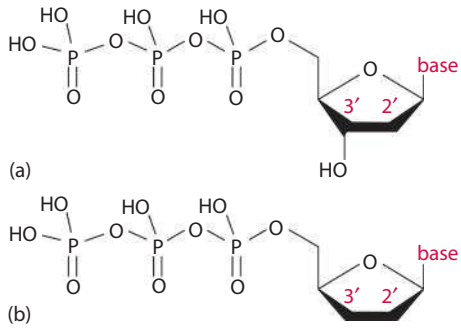

Figure 23.11 Chemical structures of dNTP and ddNTP: (a) dNTP; (b) ddNTP. Both hydroxyl groups attached to the 2′ and 3′ carbons of ribose are replaced by hydrogens.

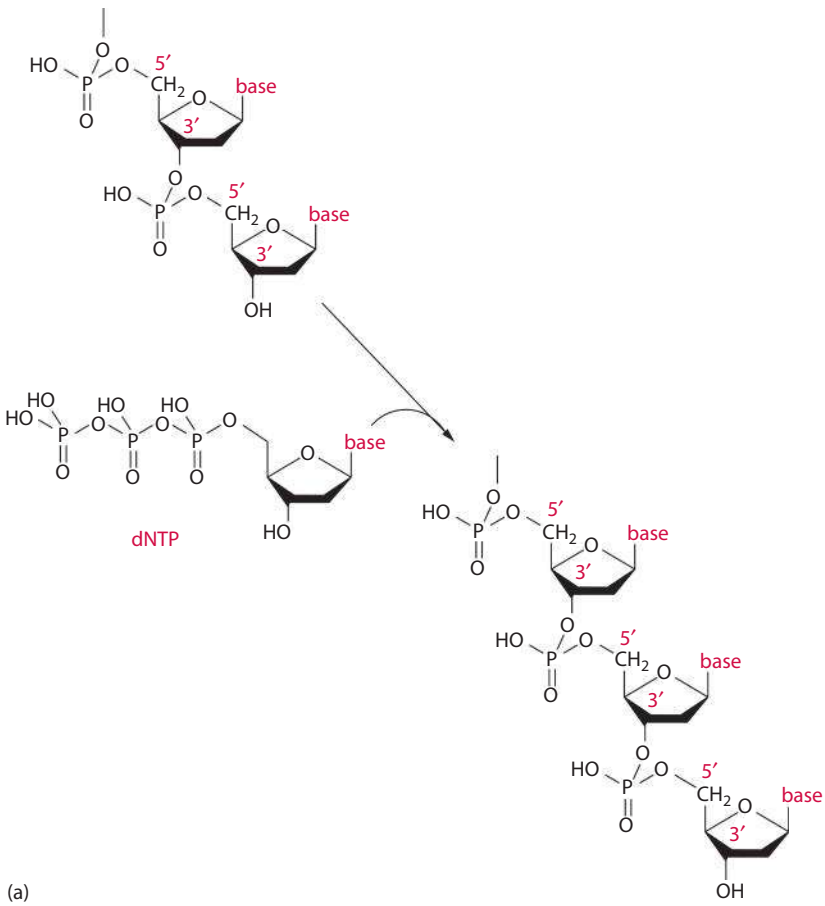

Figure 23.12 Competition between extension and termination. (a) Growing chain in extension.

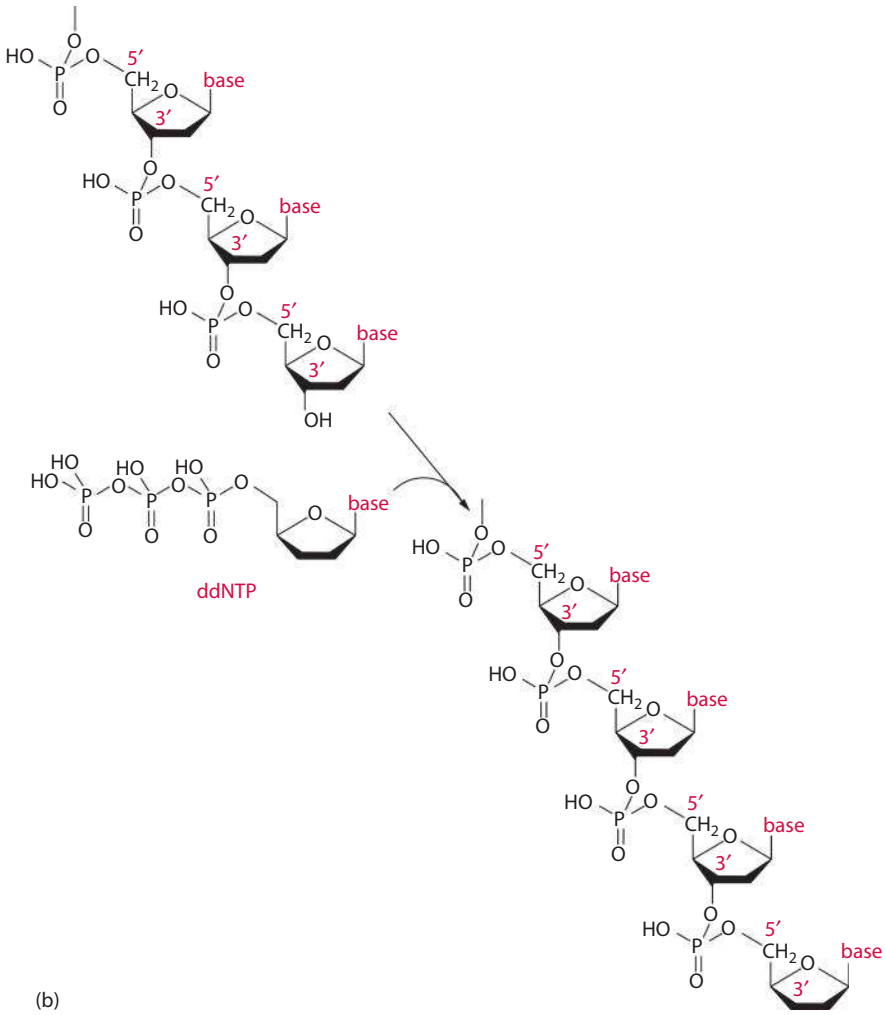

Figure 23.12 (Continued) (b) DNA strand with labeled ddNTP incorporated into a growing DNA chain that interrupts the incorporation of new nucleotides. The ddNTP usually is labeled with a fluorescent dye used for detection.

The dye-terminator system is commonly used for mtDNA sequencing in forensic laboratories. With the dye-terminator system, the ddNTPs are labeled with four different fluorescent dyes, each with a distinct spectrum. Thus, the sequencing with all four ddNTPs can be carried out in a single reaction. The labeled products of sequencing reactions are then resolved during electrophoresis and the sequencing data can be collected (Section 23.3.3.).

23.3.3.2.2 Cycle Sequencing

The chain-termination reaction is carried out using a *cycle sequencing* technique commonly used in forensic laboratories for mtDNA sequencing. Cycle sequencing, developed in the late 1980s, utilizes thermal cycling to generate a single-stranded template for chain-termination sequencing reactions. The application of thermal cycling in a sequencing reaction greatly increases the signal intensity and thus the sensitivity of the sequencing.

The sequencing reactions are carried out with multiple rounds of thermal cycling. Each cycle consists of three steps: denaturation of the double-stranded DNA template, annealing of

a sequencing primer to its target sequence, and the extension of the annealed primer by DNA polymerase. Cycle sequencing utilizes only a single primer per reaction. During the extension of cycle sequencing, the extension of the strand is terminated with the incorporation of a ddNTP (Figure 23.13). The resulting partially double-stranded hybrid, consisting of the full-length template strand and its complementary chain-terminated product, is denatured during the first step of the next cycle, thereby liberating the template strand for another round of annealing, extension, and termination.

23.3.3.3 Electrophoresis, Sequence Analysis, and Mitotype Designations

The cycle sequencing products can be separated using electrophoresis in a 4% polyacrylamide denatured gel or a POP-6 polymer (Applied Biosystems) as the matrix for capillary electrophoresis (Figure 23.14). Following data collection, sequence data analysis can be performed with the Sequencher™ software (Gene Codes Corporation, Ann Arbor). Figure 23.15 shows sequence data.

Sequencing of a region of the mtDNA genome should be performed twice. Additionally, both strands of a region of the mtDNA genome should be sequenced to reduce ambiguities in sequence determination. The sequences of evidence samples and reference samples such as that of the victim or suspect can be compared. The nomenclature used in reporting should be compatible with International Union of Pure and Applied Chemistry (IUPAC) codes.

23.3.3.3.1 Reporting Format

The rCRS is used as a reference standard to facilitate the designation of mitotypes. For reporting purposes, sequence differences relative to the rCRS are listed in data format. When a difference between an individual's sequence and that of the rCRS sequence is observed, only the position (designated by a number) and the nucleotide differing from the reference standard are recorded. In this format, nucleotides identical to the rCRS are not listed. For example, at position 228 (HV2), the rCRS has a G. If a mitotype carries an A at position 228, the individual's mtDNA sequence is described as 228A. If an unresolved sequence ambiguity is observed at a position,

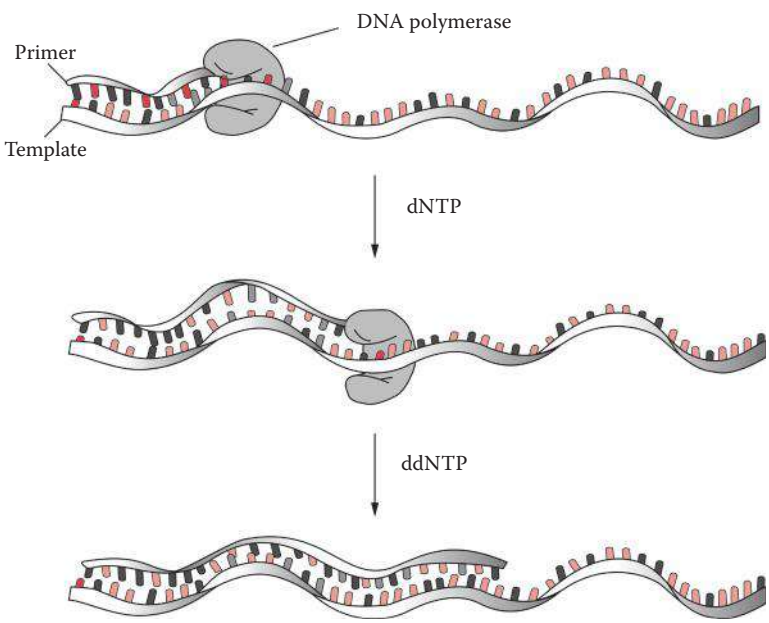

Figure 23.13 Cycle sequencing reaction. The reaction requires DNA polymerase, a template, and a primer. During DNA synthesis, the dNTP is incorporated by a new phosphodiester bond with the primer. The incorporation of ddNTP blocks further DNA synthesis of the growing chain. (© Richard C. Li.)

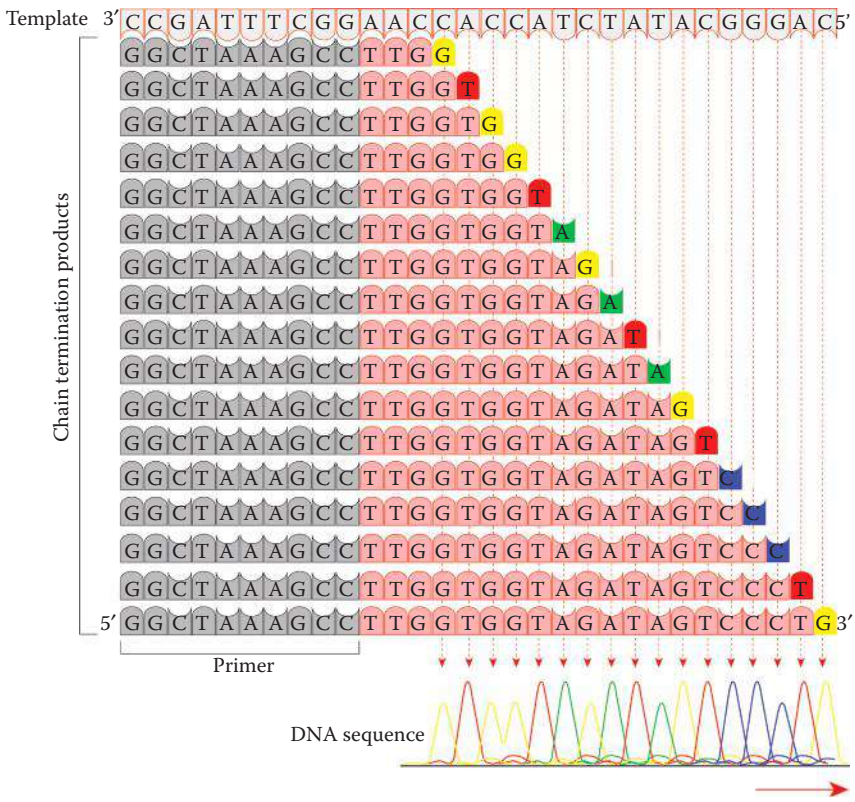

Figure 23.14 Diagram of the Sanger sequencing products. The DNA chain lengths are determined by the addition of ddNTP at different positions. The fluorescent dye-labeled cycle-sequencing products are separated using capillary electrophoresis. The fluorescent dyes are resolved by a detector, and the peaks corresponding to each DNA fragment are identified. (© Richard C. Li.)

the base number for the position is listed followed by an N. For example, 228N means an unresolved sequence ambiguity was observed at position 228.

23.3.3.3.2 Insertions

The insertion site is described by noting the position (at 5′ to the insertion) followed by a decimal point and a number. The number indicates the order of the insertions (e.g., 1 indicates the first insertion, 2 indicates the second, etc.). The base calling following the number indicates the inserted nucleotide (e.g., 524.1A, 524.2C).

23.3.3.3.3 Deletions

The deletion site designation is followed by the letter *d*. For example, a deletion at position 16,296 is recorded as 16,296d.

23.3.3.3.4 Heteroplasmic Sites

The IUPAC codes for base calling can be applied to heteroplasmic sites. For example, an A/G heteroplasmy can be designated as R, and a C/T heteroplasmy can be designated as Y.

23.3.4 Interpretation of mtDNA Profiling Results

Interpretation guidelines are used when comparing sequencing results between evidence and reference samples. General guidelines were set forth by the Scientific Working Group on DNA

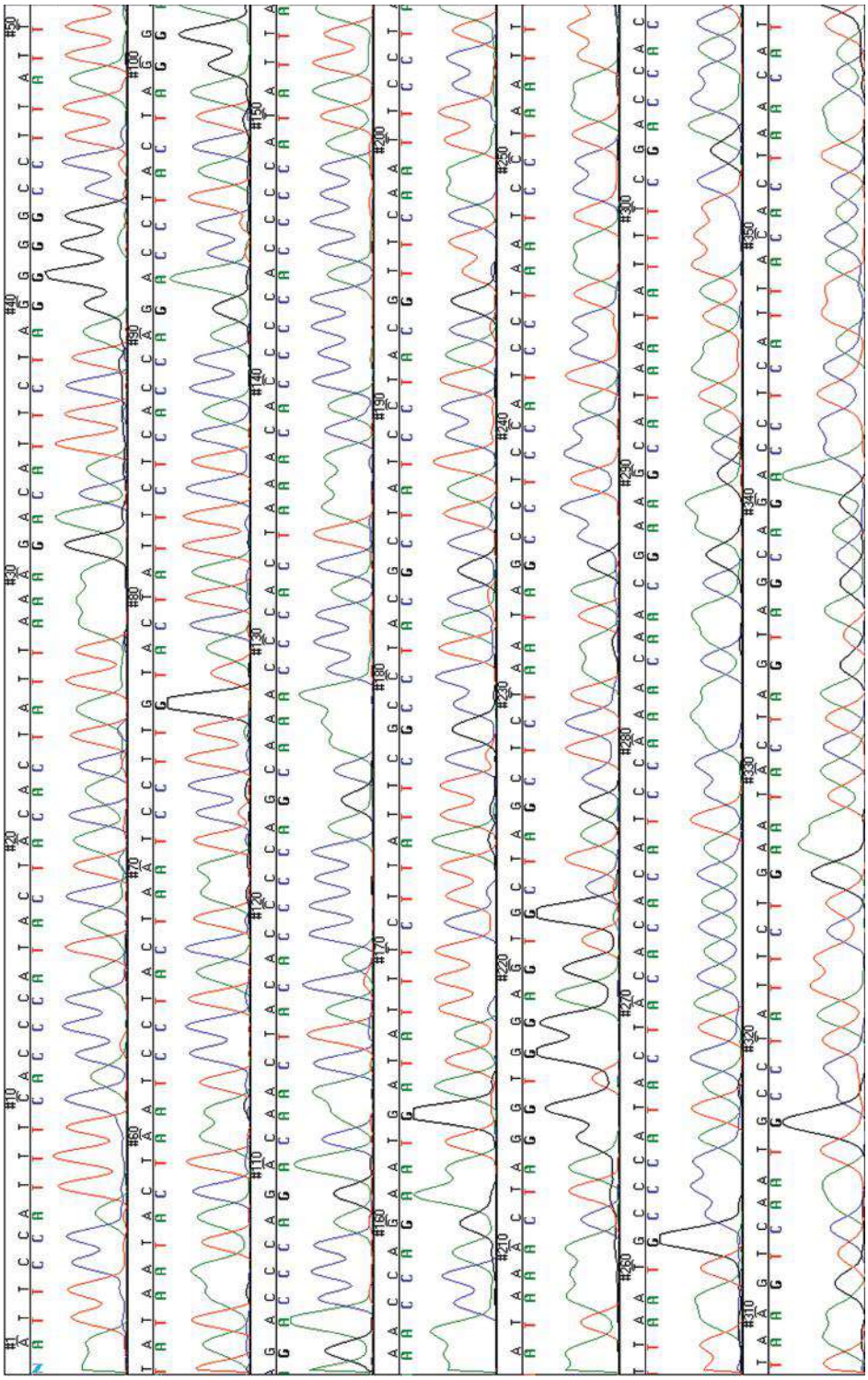

Figure 23.15 Electropherogram representing a DNA sequence presented as peaks with corresponding colors for each of the four bases. The bases associated with each peak are noted. The sequence is read from left to right (5′ to 3′). (© Richard C. Li.)

Analysis Methods (SWGDAM) and the DNA Commission of the International Society of Forensic Genetics (ISFG). In reporting mtDNA profiling results, the most common categories of conclusions are the following: cannot exclude, exclusion, and inconclusive result.

23.3.4.1 Exclusion

If the sequences of questioned and known samples are different, then the samples can be excluded as originating from the same source. It should be taken into account that higher mutation rates are found with the mtDNA genome than are found with the nuclear genome. The SWGDAM's guidelines define that the conclusion of exclusion can be made if there are two or more nucleotide differences between the questioned and known samples. Additionally, mutations seem to be more common in certain tissues. For this reason, the sources of the tissues analyzed should be taken into consideration.

23.3.4.2 Cannot Exclude

If the sequences are the same, the reference sample and evidence cannot be excluded as arising from the same source. When an mtDNA profile cannot be excluded, it is desirable to evaluate the weight of the evidence. In cases where the same heteroplasmy is observed in both questioned and known samples, its presence increases the strength of the evidence. However, if heteroplasmy is observed in a questioned sample but not in a known sample or vice versa, a maternal lineage still cannot be excluded.

23.3.4.3 Inconclusive Result

If the questioned and known samples differ by a single nucleotide, and no heteroplasmy is present, the results are considered to be inconclusive.

Bibliography

Mitochondrial DNA Profiling

Adachi, N., K. Umetsu, and H. Shojo, Forensic strategy to ensure the quality of sequencing data of mitochondrial DNA in highly degraded samples. *Leg Med (Tokyo)*, 2014, **16**(1): 52–55.

Anderson, S., et al., Sequence and organization of the human mitochondrial genome. *Nature*, 1981, **290**(5806): 457–465.

Andreasson, H., et al., Forensic mitochondrial coding region analysis for increased discrimination using pyrosequencing technology. *Forensic Sci Int Genet*, 2007, **1**(1): 35–43.

Andrews, R.M., et al., Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*, 1999, **23**(2): 147.

Anjum, G.M., et al., Pyrosequencing-based strategy for a successful SNP detection in two hypervariable regions: HV-I/HV-II of the human mitochondrial displacement loop. *Electrophoresis*, 2010, **31**(2): 309–314.

Arnestad, M., et al., A mitochondrial DNA polymorphism associated with cardiac arrhythmia investigated in sudden infant death syndrome. *Acta Paediatr*, 2007, **96**(2): 206–210.

Asari, M., et al., Differences in tissue distribution of HV2 length heteroplasmy in mitochondrial DNA between mothers and children. *Forensic Sci Int*, 2008, **175**(2–3): 155–159.

Bandelt, H.J. and W. Parson, Consistent treatment of length variants in the human mtDNA control region: A reappraisal. *Int J Legal Med*, 2008, **122**(1): 11–21.

Bandelt, H.J., M. van Oven, and A. Salas, Haplogrouping mitochondrial DNA sequences in legal medicine/forensic genetics. *Int J Legal Med*, 2012, **126**(6): 901–916.

Bendall, K.E. and B.C. Sykes, Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet*, 1995, **57**(2): 248–256.

Bendall, K.E., V.A. Macaulay, and B.C. Sykes, Variable levels of a heteroplasmic point mutation in individual hair roots. *Am J Hum Genet*, 1997, **61**(6): 1303–1308.

Bendall, K.E., et al., Heteroplasmic point mutations in the human mtDNA control region. *Am J Hum Genet*, 1996, **59**(6): 1276–1287.

24

Forensic DNA Databases

Tools for Crime Investigations

24.1 Brief History of Forensic DNA Databases

Forensic DNA databases are networks for exchanging information among law enforcement agencies to assist in solving crimes. For example, forensic DNA databases allow forensic laboratories to search DNA profiles against the databases to identify criminals. In 1995, the United Kingdom established the world's first national DNA database, NDNAD, in England and Wales. Scotland and Northern Ireland have their own databases but also submit their profiles to NDNAD. NDNAD demonstrated initial success in solving crimes. Three years later, the United States introduced its national *Combined DNA Index System* or CODIS. By the end of 1998, other countries (such as Austria, Germany, the Netherlands, New Zealand, and Slovenia) had also introduced national DNA databases. Table 24.1 describes some of these national DNA databases. This chapter will focus on CODIS.

24.2 Infrastructure of CODIS

In the United States, a pilot project of DNA databasing was initiated by the Federal Bureau of Investigation (FBI) and 14 participating state and local laboratories. Subsequently in 1994, the Congressional DNA Identification Act authorized the FBI to establish a national DNA database including "DNA identification profiles of persons convicted of crimes, and analyses of DNA samples recovered from crime scenes and from unidentified human remains." By 1997, 13 STR loci were selected and in 1998 were implemented as the core loci for the national database, known today as CODIS. All 50 states, the District of Columbia, the federal laboratories, the US Army Criminal Investigation Laboratory, and Puerto Rico contribute to CODIS.

CODIS has three hierarchical levels: the Local DNA Index System (LDIS), the State DNA Index System (SDIS), and the National DNA Index System (NDIS) (Figure 24.1). The LDIS is maintained at crime laboratories operated by police departments, sheriff's offices, and local agencies. All forensic DNA profiles originating at the local level are stored in the LDIS and are transmitted to the SDIS and NDIS. Each state maintains a SDIS, which is typically operated by a designated state laboratory. An SDIS also stores the DNA profiles generated from state laboratories. The quality assurance standards for a qualified laboratory were set up by the DNA Advisory Board in 1998. The periodic revision of the standards is now carried out by the

Table 24.1 Characteristics of National DNA Databases

Country	Year Established	Suspect Entry Criteria	Convicted Offender Entry Criteria	Removal Criteria
United Kingdom	1995	Any recordable offense that leads to imprisonment	Entered as suspect	Never removed
New Zealand	1996	No suspects entered	Relevant offense (≥7 years in prison)	Never removed unless conviction quashed
Austria	1997	Any recordable offense that leads to imprisonment	Entered as suspect	Only after acquittal
The Netherlands	1997	No suspects entered except when suspect's DNA is tested for case	Offense leading to >4 years in prison	20–30 years after conviction
Germany	1998	Offense leading to >1 year in prison	After court decision	After acquittal or 5–10 years after conviction if prognosis good
Slovenia	1998	Any recordable offense that leads to imprisonment	Entered as suspect	Depends on severity of crime
United States	1998	No suspects entered; under revision	Depends on state law	Depends on state law
Finland	1999	Offense leading to >1 year in prison	Entered as suspect	Only after acquittal
Sweden	2000	No suspects entered	Offense leading to >2 years in prison	10 years after release from prison
Switzerland	2000	Any recordable offense that leads to imprisonment	Entered as suspect	After acquittal or 5–30 years after conviction
France	2001	No suspects entered	Sexual assaults and serious crimes	40 years after conviction

Source: Adapted from Jobling, M.A. and Gill, P., *Nat Rev Genet*, 5, 739–751, 2004.

Scientific Working Group of DNA Analysis and Methods (SWGDAM). The SDIS enables local laboratories within that state to compare DNA profiles. SDIS also serves as the communication path through which the LDIS and NDIS are able to exchange messages. Communication is mediated using a secured network with encryption. Only authorized personnel, approved by the FBI, have access to CODIS network terminals and servers. NDIS is the highest level of the CODIS infrastructure, which contains DNA profiles contributed by participating local and state

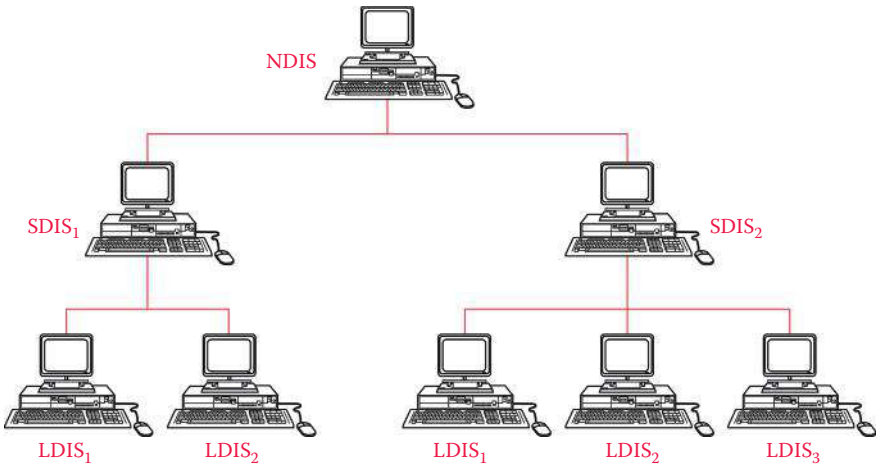

Figure 24.1 CODIS infrastructure. All DNA profiles originate at LDIS, and then enter SDIS and NDIS. SDIS allows laboratories within states to exchange DNA profiles. NDIS is the highest level of the infrastructure. It allows the participating laboratories to exchange DNA profiles on a national level.

laboratories. Additionally, NDIS stores DNA profiles generated by federal laboratories, such as those generated by the FBI and the US Army Criminal Investigation Laboratory. NDIS enables qualified state laboratories to compare DNA profiles. Searches of DNA profiles can be conducted at the national level. All DNA profiles submitted to NDIS are automatically searched weekly against the DNA profiles from other states. NDIS is administered by the FBI, which provides software, training, and support for all participating laboratories. Many law enforcement laboratories from other countries utilize the CODIS software for their own databases. However, laboratories in foreign countries do not have any access to the CODIS system. Nevertheless, a search through the CODIS database may be requested either from the FBI or International Criminal Police Organization (Interpol). Conversely, a search of an international DNA database can be arranged through Interpol.

24.3 Indexes of CODIS

The DNA profiles entered in CODIS are organized into categories known as indexes (Figure 24.2). The Convicted Offender Index contains DNA profiles of individuals convicted of crimes. The Arrestee Index contains DNA profiles of arrested individuals. It varies in SDIS databases based on each state's law permitting the collection of DNA samples from arrestees. The Forensic Index contains DNA profiles, also known as forensic profiles, derived from crime scene evidence, potentially originating from perpetrators but not including suspects.

Additionally, the FBI also established the National Missing Person DNA Database (NMPDD) Program for the identification of missing and unidentified persons at the national level. The DNA profiles entered into the NMPDD are categorized into three indexes that can be searched against each other. The Missing Person Index and the Unidentified Human Remains Index contain DNA profiles from missing persons and unidentified human remains, respectively. The Biological Relatives of the Missing Person Index contains DNA profiles voluntarily contributed from relatives. This index may also store patrilineal or matrilineal DNA profiles from the relative such as a biological father, mother, or child of the missing person to assist investigations. Additionally, a pedigree chart (a diagram showing the relationship between the missing person and relatives) can be created.

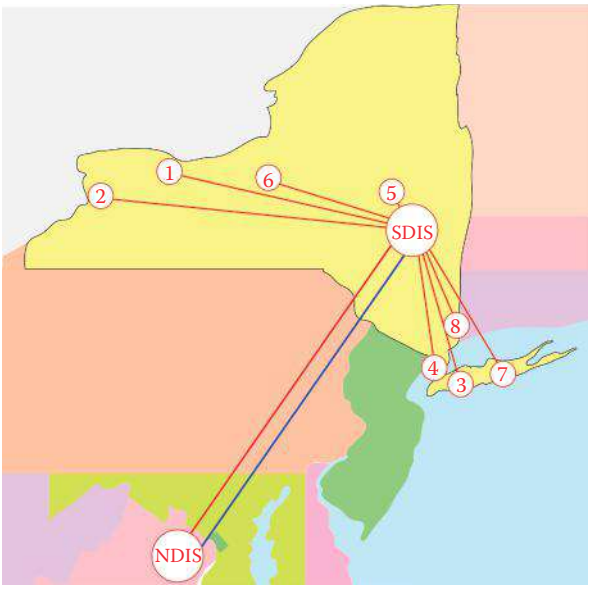

Figure 24.2 Example of an SDIS maintaining the Forensic Index (red) and Convicted Offender Index (blue). In this example, the New York State Police Forensic Investigation Center serves as the SDIS laboratory for New York State. LDIS sites: 1. Erie County Department of Central Police Services Forensic Laboratory, 2. Monroe County Public Safety Laboratory, 3. Nassau County Office of the Medical Examiner Division of Forensic Services, 4. New York City Office of the Chief Medical Examiner Department of Forensic Biology, 5. New York State Police Forensic Investigation Center, 6. Onondaga County Center for Forensic Sciences, 7. Suffolk County Crime Laboratory, and 8. Westchester County Department of Laboratories & Research Division of Forensic Sciences. (© Richard C. Li.)

In 2010, the FBI established a Rapid DNA Program Office for the purpose of developing rapid DNA technology. Rapid DNA technology is a fully automated process of performing STR analysis within 1–2 h to generate a CODIS core STR profile from a reference sample such as a buccal swab (Figure 24.3). The Rapid DNA Index System (RDIS) is the proposed index of NDIS. It shall be an integrated system capable of applying rapid DNA technology and carrying out database searches from police custody or booking units by trained police officers. The entire process, including obtaining any match from a database search, shall take less than 2 h.

The forensic DNA analysis of a reference sample, taken from an individual, takes 2–3 days if processed immediately in a forensic laboratory. The entire process includes the extraction of DNA from a sample, DNA quantification, PCR amplification, electrophoresis, and data collection and analysis. After being transported to a laboratory for processing, samples are usually stored and batched in laboratories prior to analysis. A typical turnaround time for analysis of such a sample is 1–3 months. While the sample is processed, the suspect is often released from custody. Perpetrators released on police bail may commit another crime. Thus, it is desirable to develop new technology that is capable of completing the forensic DNA analysis and database search while the suspect remains in custody. This technology would also facilitate rapid exclusion of a suspect, thus redirecting the investigation.

Rapid DNA instruments are portable, compact instruments designed to be deployed into field testing. These instruments are fully automated for processing reference samples in order to generate a DNA profile in less than 2 h. Several versions of the instruments have recently been developed by manufacturers and some of them are currently commercially available. In addition to rapid DNA instruments, rapid DNA profiling may be achieved through alternative processes. First, it can be achieved through rapid services that offer a quick turnaround time (<2 h). Second,

Figure 24.3 A rapid DNA instrument. A compact instrument that is designed to process swab samples and to produce DNA profiles within 2 h in the field. (© Richard C. Li.)

rapid DNA techniques can be applied using standard laboratory equipment but implementing specialized protocols to generate CODIS-compatible profiles in less than 2 h.

The primary goal of the rapid DNA initiative is to produce a CODIS-compatible DNA profile of a sample taken from an arrestee in police custody and to search DNA databases during the booking process in less than 2 h. This technology could be used, for example, to quickly identify other unsolved crimes or eliminate suspects, which can aid in investigative decision making. It can also potentially be useful for other investigations, such as identifying human remains in mass disasters; identifying detainees in counterterrorism applications; and, for immigration and border agents, confirming and verifying individual identifications or family relationships.

24.4 Database Entries

Currently, over 190 public law enforcement laboratories participate in CODIS across the United States. As of April 2013, CODIS contains the DNA profiles of more than 10 million convicted felons, over 1 million arrestees, and half a million crime scene samples. The proportion of the population represented in the database is approximately 3%. The NDIS is one of the largest DNA databases containing DNA profiles in terms of absolute numbers. Each CODIS entry consists of the DNA profile of the sample and the specimen identification number, as well as the information of the laboratory submitting the DNA profile and the laboratory analyst that performed the DNA analysis. The entry does not include case information or the personal information of the offenders or arrestees. Access to CODIS profiles is restricted to criminal justice agencies for law enforcement identification purposes.

The SDIS database retains samples collected from convicted offenders after DNA profiles are obtained. State policies vary in the retention of samples and DNA profiles in situations of

dismissed charges, acquittal, or no charges for arrestees. Some states, such as Virginia, destroy the DNA sample and expunge the DNA profile from the database. Other states, such as California, require a petition to destroy the sample and expunge the DNA profile from the database. There are two major reasons for sample retention. First, the sample is needed for confirmatory processes for the purpose of quality assurance and control. In confirmatory processes, the DNA analysis of the original sample is repeated and compared against the prior analysis. Second, the retention of samples allows for possible retesting if new technology becomes available. It also allows retesting of a sample for purposes of updating with expanded loci. Considerable debate has surrounded the retention of samples. It can be argued from the opposite perspective that the database could reveal private genetic information that could then be misused. The objection is based on concern for the protection of privacy rights. If a sample were made available to an unauthorized person, confidential information could be disclosed.

24.5 Database Expansion

Currently, all 50 states have authorized the collection of samples from convicted felons for DNA databasing. Over the years, the demand for the utilization of databases has increased sharply (Figure 24.4). More jurisdictions are incorporating more felonies into the lists of crimes that require DNA profiles, and some jurisdictions plan to include all felonies in such databases. Additionally, the database system is projected to include the profiles of minor criminals, because statistics show that most offenders found guilty of serious crimes were previously convicted for minor crimes. Broadening the size of the database and including samples from more types of crimes could lead to the assumption that the number of crimes solved would also increase. Although state laws vary, each state has its own statute governing the entry criteria of the database samples. More and more states are authorizing the collection of additional types of DNA samples, including individuals convicted of misdemeanor crimes and adult felony arrestees who have not yet been convicted for the offense. It is known that offenders tend to commit multiple

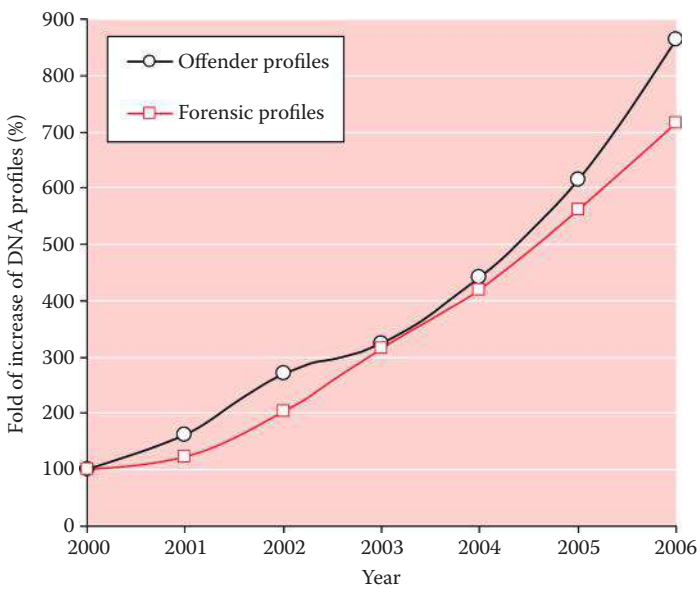

Figure 24.4 Growth of the CODIS database. In 2000, 460,365 offender profiles and 22,484 forensic profiles were entered.

crimes. Thus, including DNA profiles of arrestees in the database is beneficial in catching serial offenders and potentially preventing future crimes. Virginia is the first state to collect DNA from violent-felony arrestees, among other crimes. Over the years, more than half of states in the United States have amended their law to include all felony arrestees or a subset of arrestees for felonies that involve violence or sexual assault.

In 2009, Alonzo King was arrested in Wicomico County, Maryland, on violent assault charges. Maryland's DNA Collection Act authorizes law enforcement agencies to collect DNA samples from an individual who is arrested for violent crimes. During the booking process, under state law, King's buccal swabs were collected for forensic DNA identification. When King's DNA profile was entered into Maryland's SDIS, it matched to a crime scene DNA profile from an unsolved rape case in 2003. Based on the DNA evidence, King was indicted for first-degree rape by a grand jury. The defendant filed a motion to suppress the DNA evidence. King argued that using his DNA to investigate his connection to the 2003 rape was an unreasonable search and seizure under the Fourth Amendment, since the police had no reason to suspect his involvement in the 2003 rape. During the trial at Wicomico County Circuit Court, King's motion was denied. He was convicted of rape and sentenced to life in prison. The Court of Appeals of Maryland, the state's highest court, upheld that the DNA Collection Act was constitutional. The police had probable cause to arrest King on the assault charge, and thus collecting King's DNA sample was a reasonable search. However, the court overturned the low court and reversed King's conviction. It stated that the DNA Collection Act was inappropriately applied to King in this case. Thus, investigating King without probable cause for the unrelated rape was unconstitutional. To determine whether the Fourth Amendment permits states to collect and analyze the DNA of arrestees, a divided Supreme Court ruled 5–4 and reversed the Court of Appeals. The Supreme Court's judgment ruled that Maryland has the right to collect DNA evidence from individuals arrested for serious crimes. The ruling reflects the Supreme Court's view on the balance between the interest of criminal justice systems in solving violent crimes and an individual's interest in the Fourth Amendment, which protects them from unreasonable searches.

Some jurisdictions, such as the United Kingdom, allow DNA samples to be taken from individuals suspected of committing recordable offenses that may lead to prison sentences. In contrast, DNA profiles from suspects are not eligible for entry into CODIS. In some jurisdictions, it has been suggested that databases include many more offenders and suspects, as well as the general public. One advantage of including an entire population in a database is the ability to identify missing, kidnapped, and abducted individuals in addition to victims of major accidents and mass fatalities. Nevertheless, debates concerning the need to balance the benefits and dangers of developing a broader database will inevitably continue.

24.6 DNA Profiles

Currently, the CODIS software supports the storage and search of DNA profiles of short tandem repeat (STR), Y chromosome STR (Y-STR), and mitochondrial DNA (mtDNA). Y-STR and mtDNA profiles may only be searched within NMPDD-related indexes. The CODIS software no longer supports searches of DNA profiles generated by restriction fragment length polymorphism (RFLP) analysis. The 13 core CODIS STR loci are CSF1PO, FGA, THO1, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. DNA profiles are entered into one of the indexes, such as convicted offender, arrestee, forensic, unidentified human remains, missing person, or a relative of a missing person. There is a minimum loci requirement for each DNA profile entering CODIS: 13 core CODIS loci are required for a DNA profile entering the Convicted Offender Index and the Arrestee Index; 13 core CODIS loci and amelogenin, a sex-typing marker, are required for a DNA profile entering the Relatives of Missing Person Index; at least 10 CODIS loci are required (all 13 core loci must be tested) for the Forensic Index; and at least 8 loci and amelogenin are required (all 13 core loci must be tested)

for the Missing Person and Unidentified Human Remains Index. The DNA profiles must be generated by an accredited and audited laboratory in accordance with the FBI's Quality Assurance Standards using approved commercially available kits. As the database is growing rapidly, the chances of finding incidental matches among DNA profiles is increasing. To increase discrimination power, it has recently been proposed that CODIS core loci be expanded to include 20 or more markers (Chapter 20). Additionally, the expanded core loci share additional loci compatible with international standards for forensic DNA analysis, which facilitates the exchange of information with international law enforcement agencies.

24.7 Routine Database Searches for Forensic Investigations

The ultimate goal of the database utilization is to provide investigative leads for law enforcement in solving crimes, particularly in cases where no suspect has yet been identified. Currently, DNA profiles uploaded to NDIS are automatically searched once a week (Figure 24.5). A hit is a match made from the information provided by comparing a target DNA profile against the DNA profiles contained in the database. There are two types of CODIS hits: an *offender hit* provides the identity of a potential suspect of a crime, while a *forensic hit* reveals the linkage between two or more crime scenes. Once a hit is identified, the match is verified by the laboratories that originally processed the evidence. A verified CODIS hit can be utilized as probable cause to allow law enforcement to obtain a court order to collect a DNA sample from a suspect. Collecting DNA with a warrant ensures admissibility in court. *Investigation-aided* cases are those assisted by CODIS hits, including case-to-case matches as well as case-to-offender matches. The number of investigations aided is a useful measure of the successful application of the database (Figure 24.6). As of April 2013, over 207,800 hits had been made with CODIS assistance in more than 199,200 investigations.

24.7.1 Case-to-Offender Searches

Matches of profiles from the Forensic Index and the Offender Index reveal the identities of perpetrators of crimes. For example, in 1998, a Florida man, Leon Dundas, became a suspect in a rape case but refused to provide his DNA reference sample for testing. Dundas was killed a year later in an illegal drug deal. Thus, a postmortem sample of Dundas was obtained and his DNA profile was compared with the Forensic Index of CODIS. It was discovered that Dundas's profile

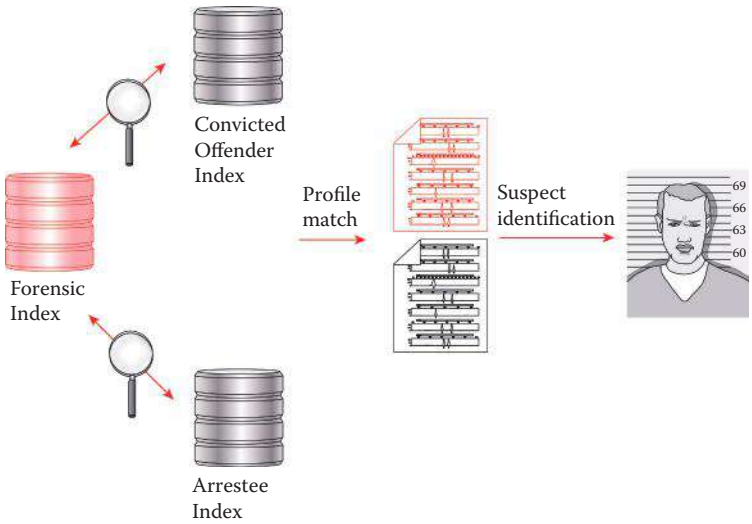

Figure 24.5 Example of a weekly routine CODIS search. (© Richard C. Li.)

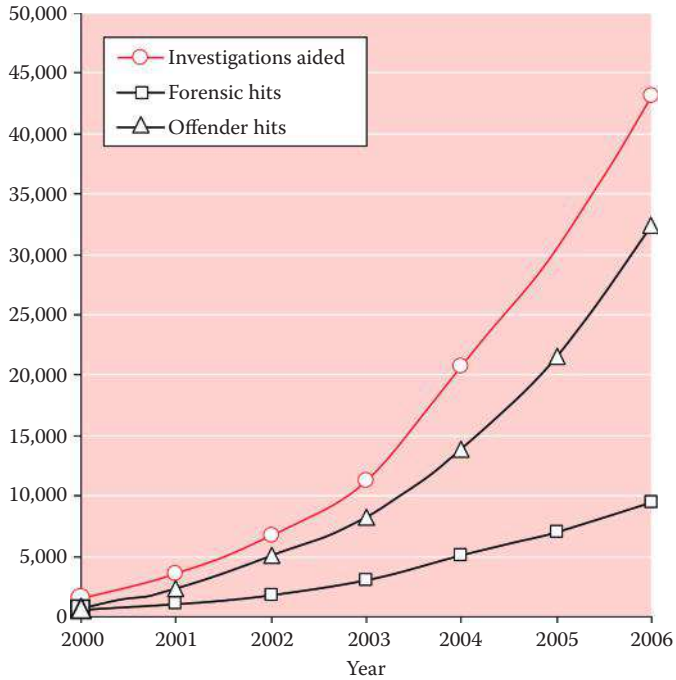

Figure 24.6 Number of hits and investigations aided by the CODIS database.

matched the DNA sample found in the rape case. Additionally, his DNA profile was linked to several other unsolved rapes in Jacksonville and Washington, DC. Such cases can now be solved by utilizing DNA databases.

24.7.2 Case-to-Case Searches

A target DNA profile from a crime scene is also searched against the profiles stored in the Forensic Index of the database. Matches of profiles among the target profile and the profile in the Forensic Index can link separate crime scenes and aid in identifying serial offenders. This helps law enforcement agencies in multiple jurisdictions to coordinate their investigations and share leads. For example, in 1996, two young girls were abducted from bus stops in St. Louis. Both girls were raped and DNA samples were collected. Both DNA profiles pointed to the same perpetrator. In 1999, the St. Louis police decided to reanalyze the samples using new STR technology through the CODIS database. The database found a match to a different rape case, to which the perpetrator, Dominic Moore, had already confessed, thus identifying him as the perpetrator of the 1996 rapes.

24.7.3 Search Stringency and Partial Matches

Database searches are carried out using the CODIS software with three stringency levels, which allow the search of complex forensic profiles against offender profiles (Figure 24.7). A high-stringency match requires an exact match in which all alleles are matched at each locus between the target and candidate profiles. A moderate-stringency match, as defined by the FBI, is a candidate match "between two single source profiles having at each locus all of the alleles of one sample represented in the other sample." In a moderate-stringency match, allelic drop-outs in a target or candidate profile are allowed, possibly resulting in a partial match at some loci. As a result, high-stringency matches are automatically included in a moderate-stringency

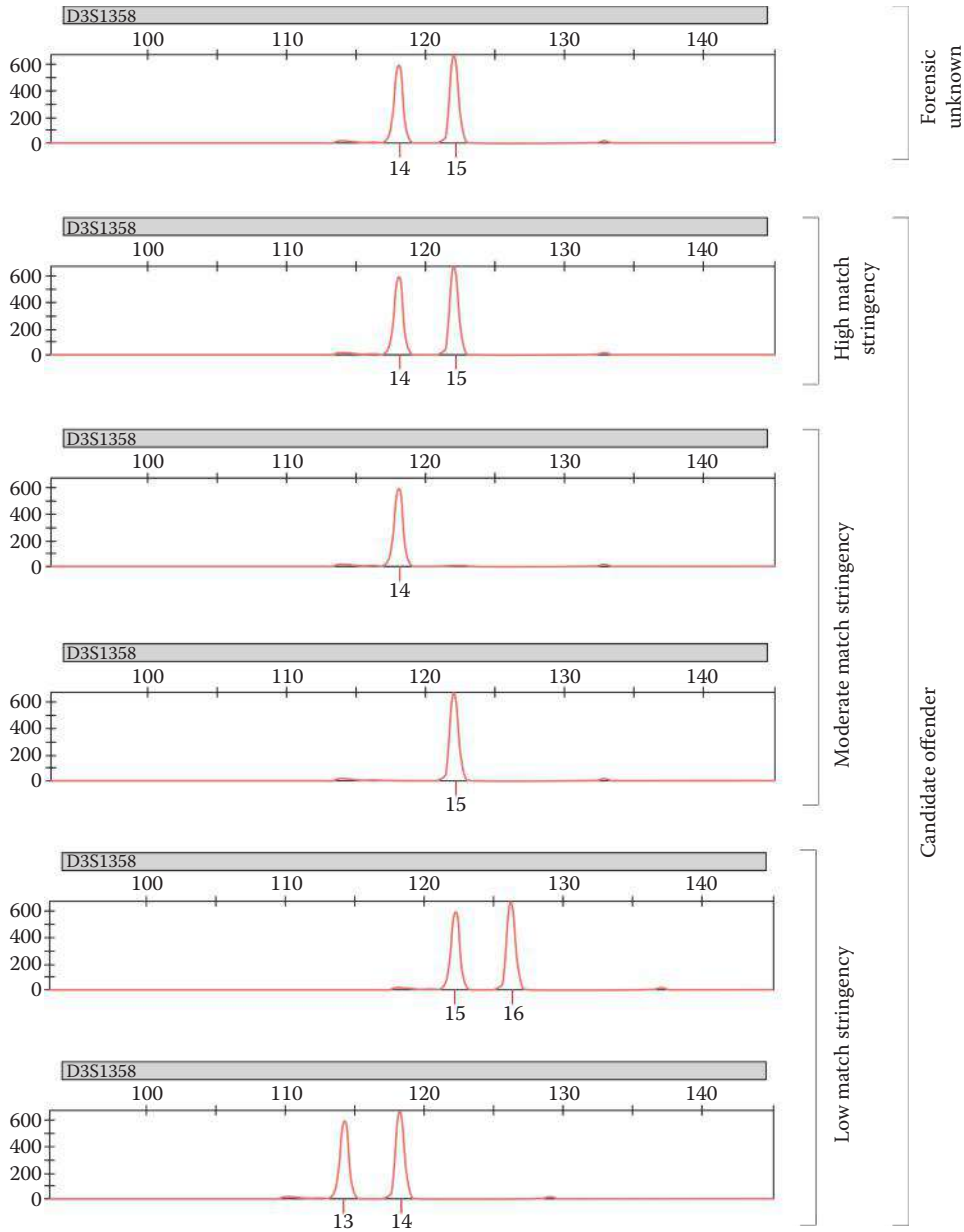

Figure 24.7 Diagram of match stringencies. A heterozygote (allele 14 and 15) at locus D3S1358 is utilized as the genotype for the forensic unknown profile. (© Richard C. Li.)

search. In a low-stringency search, both mismatches and allelic dropouts are allowed. Presently, NDIS searches are only carried out at moderate stringency. During forensic DNA analysis, DNA degradation may prevent full DNA profiles from being processed, producing only partial profiles. Additionally, mixture profiles derived from forensic samples containing DNA contributed by more than one individual may be encountered. Furthermore, due to mutations, null alleles may occur in the profiles produced with some primer sets but not other primers. Therefore, searching at moderate stringency allows the detection of matches under the situations described above.

Such moderate-stringency searches between the Forensic Index and Offender Index may occasionally generate partial match profiles. Since a partial match is not an exact match of the two profiles, further investigation is needed. For instance, additional Y-STR and mtDNA analysis is needed to eliminate unrelated individuals.

24.8 Familial Searches

Familial search, initiated in the United Kingdom, is a new method of applying databases in criminal investigations. It is known that, in the United States, nearly half of prison inmates have close relatives who have also been incarcerated. Familial search is based on the assumption that close relatives share more alleles of DNA profiles than unrelated individuals. Thus, databases may be utilized to identify perpetrators by finding a close relative, if the close relative has been convicted of a crime and is listed in the database. Familial search is an intentional search of a target crime scene profile against an offender database to obtain a list of candidate profiles that are similar to the target profile. This list may include the profile of a close relative of the perpetrator, who is the source of crime scene evidence. These matches most frequently involve siblings, parents, or children. The investigative leads produced by familial searches allow law enforcement to conduct further investigations to identify the perpetrator.

The first familial search leading to a successful prosecution was conducted in Surrey, England. In 2003, a truck driver was killed after a brick was thrown through his windshield from a bridge. The perpetrator's DNA profile was obtained from the brick. A search of the UK's national DNA database revealed no match for the perpetrator. Next, a familial search of the database was conducted. The system identified a close relative that led police to identify the perpetrator, Craig Harman, who was then convicted of manslaughter.

24.8.1 Legal and Ethical Issues of Familial Search

The use of forensic databases involves a balance of individual civil rights and the interests of the criminal justice systems. Many concerns have been raised, including the potential for these searches to violate the privacy of unrelated people whose genetic profiles happen to resemble those of individuals included in the databases. In the United States, the Fourth Amendment protects against unreasonable searches and seizures. The permissibility of familial searching under the Fourth Amendment is yet to be addressed by courts. The collection of biological material for the initial creation of a profile for law enforcement purposes is subject to the Fourth Amendment implications. In terms of familial searches, some legal experts argue that the biological materials are not collected directly from the individuals for a familial search, and that these individuals may thus be protected under the Fourth Amendment. Others argue, however, that the Fourth Amendment protects the initial creation of the profile, including the sample collection, forensic DNA analysis, and the databasing. They argue that the Fourth Amendment may not protect subsequent investigations of DNA profiles during familial searches. In addition, the CODIS database consists of a high percentage of profiles from individuals of racial minority groups, including African Americans. Familial searches can disproportionately focus on a specific racial group.

Familial searches are rare in the United States. Two jurisdictions, Maryland and the District of Columbia, have laws prohibiting the use of familial searches. Familial searches are not conducted at the NDIS. While familial searches are now being performed in several jurisdictions in the United States, policies on familial searches vary among jurisdictions. The major issues relate to criteria for privacy, information release, search approval, and the types of crimes eligible for familial searches. Currently, California, Colorado, Texas, and Virginia have state legislation permitting familial searches. Familial searches are initiated when a specific suspect is not known, and the cause for the search needs to be justified. Familial searches are usually conducted for

crimes that pose a substantial threat to public safety, typically those cases involving the most serious offenses. Additionally, familial searches are only conducted after a routine search of a DNA profile has yielded no match in the database.

In 2003, for example, a crime scene DNA sample of a closed rape and murder case, committed decades previously, was reanalyzed using forensic DNA techniques. The crime scene DNA profile was compared to the profile of a man, Darryl Hunt, who was then imprisoned for the crime. Surprisingly, the crime scene DNA profile did not match either that of Hunt or those of convicted felons in the database. However, the database search revealed a close relative of the true perpetrator and thus led law enforcement to identify Willard Brown as the perpetrator. Brown was sentenced to life imprisonment. Additionally, Hunt was exonerated after 18 years in prison.

24.8.2 Familial Search Strategies

The familial search is usually carried out using specially designed software. Although the CODIS software is not designed for familial search, it can be used for familial search through a low-stringency search, which may result in a list of candidate profiles including close relatives of individuals, such as parent–offspring or full-sibling relatives. In a large offender DNA database, similar DNA profiles from unrelated people are often observed due to shared alleles. Therefore, the candidate profile list may also include unrelated individuals whose DNA profiles are similar to the target profile. As a result, a familial search can provide a list of potential candidates consisting of hundreds of profiles, which would be too labor intensive to pursue through further investigations. Several methods can be used to determine a cutoff analytical threshold in order to limit a pool of candidate profiles and exclude unrelated individuals from familial searches (Figure 24.8).

24.8.2.1 Identity-by-State and Kinship Index Method

The *identity-by-state* (IBS) method compares the number of shared alleles and loci between a target forensic profile and the offender profiles in a database but does not take into account allele frequencies. The analytical threshold for a familial search is determined by a preset number of shared alleles or loci in order to prompt further investigation. For example, some states use 15 shared alleles as the analytical threshold to be considered as a candidate, while other states require at least one shared allele for each locus. Additionally, this method can prioritize a pool of candidates based on the highest to the lowest number of shared alleles for investigation.

The *Kinship Index* (KI) method is a likelihood ratio–based method that evaluates the familial match by comparing the probability that two DNA profiles are from related individuals to the probability that they are unrelated. The KI method analyzes the allele frequency data, including all CODIS core loci, to calculate the Combined Kinship Index. The KI may vary based on the allele frequency data across the population. Thus, the accuracy of the KI method relies on the relevance of the population data set analyzed. False inclusions or exclusions may occur if nonrelevant population data are utilized. The KI method also allows the generation of a ranked candidate list according to the probability that the individuals are related. Generally speaking, the accuracy of the KI method is higher than that of the IBS method. However, using both IBS and KI methods is better than using a single method alone for familial search.

24.8.2.2 Focusing on Rare Alleles

During a familial search, target DNA profiles may contain rare alleles with low allelic frequencies. The chance of sharing rare alleles for two close relatives is higher than for two unrelated individuals. If a target profile carries a rare allele, it can be used to narrow the pool of candidates.

For example, Jeffrey Gafoor, then 23, had been living in a suburban neighborhood in Cardiff, Wales. He had a reputation for being a loner. Besides working in his family-owned shop, he spent most of his time at home. On February 13, 1988, Gafoor entered 7 James Street in the Butetown area of Cardiff. The first-floor unit was temporarily occupied by Lynette White, a 20-year-old

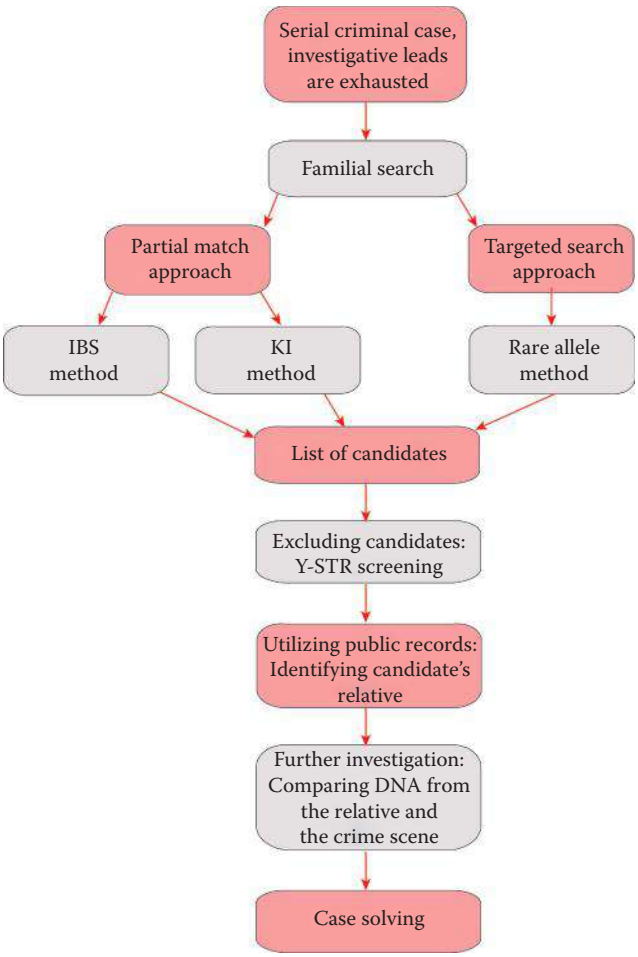

Figure 24.8 Example of a familial search workflow. (© Richard C. Li.)

woman who was working as a prostitute in the area. Gafoor went there for the purpose of receiving sexual services. He left the scene in the early hours of Valentine's Day. On Valentine's night, White's body was discovered. She suffered from more than 50 stab, cut, and slash wounds as well as defensive wounds on her hands. The body was dragged to the corner of the room adjacent to the bed, the only piece of furniture in the room.

Gafoor left blood and semen evidence at the scene. He bled during the murder and deposited bloodstains at the scene. Additionally, he left seminal stains at the scene; the semen had no measurable level of sperm and most likely came from an individual who had a medical condition known as azoospermia (Chapter 14). After the murder, Gafoor lived at the same place as he did beforehand and kept the same job for many years.

Two years later, three local men, known as the "Cardiff Three," were convicted of White's murder and sentenced to life in prison. In 1992, however, their convictions were reversed by the court of appeal. Subsequently, the crime scene evidence was tested using new techniques in forensic DNA analysis that were not available in 1988. A DNA profile was obtained from the crime scene. A familial search revealed a similar profile from a 14-year-old boy in the UK National DNA Database. In particular, there was a rare allele match between the crime scene and the candidate DNA profiles. This search led to the further investigation of Gafoor, who was the uncle of the boy. Gafoor was charged with murder.

24.8.2.3 Excluding Candidates through Y-STR Screening

The vast majority of DNA profiles in CODIS come from males. Patrilineal markers, such as Y-STR loci, are successfully utilized to screen the candidates to verify the relationship between two individuals. This analysis helps to identify first-order relatives as well as paternal half-siblings. STR loci beyond the CODIS core loci and maternal lineage makers, such as mitochondrial DNA typing systems, can be used to narrow the pool of candidates and eliminate coincidental partial matches. One or more of these methods can be incorporated to evaluate highly ranked candidates.

For example, Lonnie Franklin lived in a southern Los Angeles neighborhood. Throughout the 1980s, Franklin worked for the city as a maintenance worker and a sanitation truck operator. His neighbors recalled that he often brought prostitutes to a camper parked in the backyard of his house. He also took photos of nude women, which he kept in his garage. Nevertheless, to many of his neighbors, Franklin seemed to be a friendly person who often chatted with them.

Since 1985, the bodies of many women, most of which were prostitutes, had been dumped in an alley running along Western Avenue in a southern Los Angeles neighborhood. The victims had been shot and some had been strangled after sexual contact. DNA and ballistics analysis revealed that at least 10 of the murders had been committed by the same perpetrator. The perpetrator was known to the general public as the "Grim Sleeper" because he had taken a 14-year break during the period in which these 10 murders had taken place. The crime scene DNA profile did not match any profile in the database.

In 2008, the first familial search of the "Grim Sleeper" case was conducted by the California Department of Justice. California state law allows familial searches for high-profile cases if all other leads have been exhausted. However, all candidate DNA profiles were excluded through Y-STR screening. As a result, no candidate was identified as a possible relative of the "Grim Sleeper." In 2010, a second search targeted partial matches that shared at least 15 alleles. It produced a list of 100 candidate profiles ranked by the likelihood that they were related. Y-STR was used as a screen tool to identify possible patrilineal relatives. This time, the search generated a match to a new DNA profile of a felon, which had recently been entered into California's SDIS, in 2009. The search suggested that the candidate was a relative of the source of the DNA from the crime scenes. This search led to the identification of Franklin as a suspect in the "Grim Sleeper" murders. Police detectives conducted surveillance and collected a discarded slice of pizza and a cup used by Franklin at a local restaurant for DNA identification. The DNA profile was identical to the crime scene DNA profiles. Franklin was arrested and was charged with ten counts of murder and one of attempted murder.

Bibliography

Alonso, A., et al., Challenges of DNA profiling in mass disaster investigations. *Croat Med J*, 2005, **46**(4): 540–548.

Aoki, Y., et al., Comparison of the likelihood ratio and identity-by-state scoring methods for analyzing sib-pair test cases: A study using computer simulation. *Tohoku J Exp Med*, 2001, **194**(4): 241–250.

Asplen, C. and S.A. Lane, International perspectives on forensic DNA databases. *Forensic Sci Int*, 2004, **146**(Suppl): S119–S121.

Baker, L.E. and E.J. Baker, Reuniting families: An online database to aid in the identification of undocumented immigrant remains. *J Forensic Sci*, 2008, **53**(1): 50–53.

Balding, D.J., The DNA database search controversy. *Biometrics*, 2002, **58**(1): 241–244.

Balding, D.J., et al., Decision-making in familial database searching: KI alone or not alone? *Forensic Sci Int Genet*, 2013, **7**(1): 52–54.

Barash, M., et al., A search for obligatory paternal alleles in a DNA database to find an alleged rapist in a fatherless paternity case. *J Forensic Sci*, 2012, **57**(4): 1098–1101.

Barber, M.D. and B.H. Parkin, Sequence analysis and allelic designation of the two short tandem repeat loci D18S51 and D8S1179. *Int J Legal Med*, 1996, **109**(2): 62–65.

25

Evaluation of the Strength of Forensic DNA Profiling Results

25.1 A Review of Basic Principles of Genetics

25.1.1 Mendelian Genetics

Mendel's first law is the principle of segregation of alleles. Each pair of alleles segregates from others in the formation of gametes (mature reproductive cells such as spermatozoa and oocytes). One-half of the gamete carries one allele, and the other half carries the other allele.

Mendel's second law is the principle of independent assortment of alleles. The segregation of each pair of alleles is independent of the segregation of other pairs during the formation of gametes.

Gametes are formed during a process known as *meiosis*, in which cells with haploid chromosome numbers (23 in humans) are produced by the division of cells with diploid chromosome numbers (46 in humans). A fertilized human egg thus contains a diploid number of chromosomes. A diploid is composed of 22 pairs of autosomes and 2 sex chromosomes (XX in females and XY in males).

Based on Mendelian principles, genes on different chromosomes assort independently of one another in gamete production. Genes residing very closely together on the same chromosome are usually inherited together. Thus, they do not assort independently and are called *linked genes*. Genes distant from each other on the same chromosome are usually inherited separately. This results from an exchange of segments between a pair of *homologous* chromosomes when the chromosomes are paired during the early phases of meiotic division. These types of gene-exchange events on homologous chromosomes are collectively called *crossing over*, which results in the *recombination* of genes in a pair of chromosomes (Figure 25.1).

The Mendelian inheritances of genes can often be measured using *probabilities*. A probability is the ratio of the number of actual occurrences of an event to the number of possible occurrences. Additionally, the probability of two independent events occurring simultaneously is the product of each of their individual probabilities; this is known as the *product rule of probability*.

Mendelian principles apply to the inheritance of loci of the autosomal nuclear DNA genome commonly used for forensic testing. The Y-chromosomal genome is inherited paternally, which does not obey the rules for Mendelian principles. Mitochondria contain their own mitochondrial genomes and are inherited maternally. This maternal inheritance

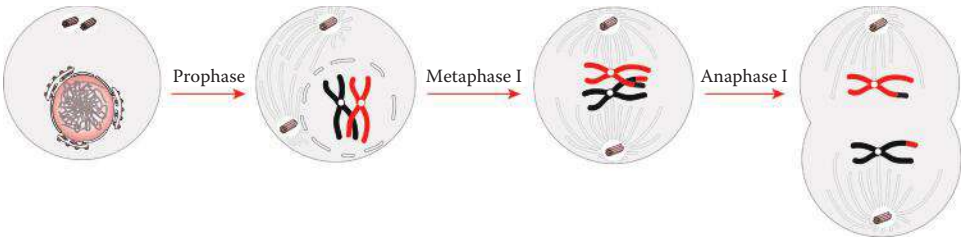

Figure 25.1 Crossing over hypothesis. Chromosomes are replicated prior to the first meiotic division. Each duplicate is called a chromatid, which forms into tetrads in the prophase. A crossover event occurs between the maternal and the paternal chromatids. Each chromatid breaks at the point of the cross and fuses with a portion of its counterpart. Chromosomes are associated with the microtubules during metaphase I and are pulled toward the spindle poles during anaphase I. The maternal and paternal homologs of one chromosome are shown. (© Richard C. Li.)

of mitochondrial genes also does not obey the rules for Mendelian principles. The inheritance of Y-chromosomal and mitochondrial genomes is referred to as non-Mendelian inheritance.

25.1.2 Population Genetics

Population genetics studies the causes of patterns of genetic variations within and among populations.

25.1.2.1 Allele Frequency

Allele frequency (p) can be calculated directly by counting the number of alleles of one type at a given locus and dividing it by the total number of alleles at that locus in a sampled population. This is called the *gene counting* method.

25.1.2.2 Genotype Frequency

Genotype frequency (P) observed at a given locus can be calculated by dividing the number of individuals with one genotype by the total number of individuals in a sampled population. Each genotype at the locus can be calculated separately. The summation of all genotype frequencies at that locus should equal 1.

25.1.2.3 Heterozygosity

Heterozygosity is the proportion of alleles, at a given locus, that are heterozygous and is calculated as

$$h = 1 - \sum_i p^2 \tag{25.1}$$

where:

h = heterozygosity
p = allelic frequency of the locus for homozygotes

The amount of heterozygosity at a locus in a sampled population is a measure of genetic variation. The higher the heterozygosity, the more variation there is at a given locus.

25.1.2.4 Hardy–Weinberg Principle

The Hardy–Weinberg (HW) principle, independently discovered by two scholars in the early 1900s, allows predictions of genotype frequencies to be made based on allelic frequencies. However, certain conditions must be met. The population must be large; mate randomly; and lack mutation, migration, and natural selection. If these conditions of the HW principle are met, the population will be in *equilibrium*, and the following results are expected:

1. The frequencies of the alleles will not change from one generation to the next.
2. Genotype frequencies can be predicted by the allelic frequencies (p^2 and q^2 for genotype frequencies of homozygotes and $2pq$ for genotype frequencies of heterozygotes). The sum of the genotype frequencies should equal 1.

$$p^2 + 2pq + q^2 = 1 \tag{25.2}$$

If the observed genotype proportions are different from those expected, one or more of the HW assumptions have been violated.

25.1.2.5 Testing for HW Proportions of Population Databases

To determine whether the genotypes of a population in question obey the HW principle, a population database can be constructed. Samples (usually 100–200 samples for STR loci) are collected and analyzed at the loci of interest. Allelic frequencies are obtained by using the gene counting method. Table 25.1 shows the allelic frequencies of CODIS loci from a population database. The *observed genotype frequencies* at a given locus, as described earlier in Section 25.1.2.2, are calculated by dividing the number of individuals with one genotype by the total number of individuals in the population sampled. The *expected genotype frequencies* are calculated using p^2, $2pq$, and q^2.

The observed and expected genotype frequencies are then compared using a chi-square test. The significance of the differences between observed and expected genotype frequencies can then be determined. The chi-square is calculated using the following formula:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{25.3}$$

where:

O_i = ith observed genotype frequency
E_i = ith expected genotype frequency
n = total number of genotypes

Chi-square (χ^2) is calculated as the sum of all genotypes of a given locus.

The chi-square value and the degrees of freedom (the number of genotypes minus the numbers of alleles) are then used to obtain a p value (not to be confused with the allelic frequency designated p) from a table of p values, and such tables can be found at the backs of most statistics textbooks. If the p value exceeds 0.05 (5% significance level), the deviation of the expected genotype frequencies from the observed genotype frequencies is not considered statistically significant. Thus, the null hypothesis that the observed genotype frequencies fit the expected genotype frequencies predicted by the HW principle is not rejected if the p value is greater than 0.05.

Table 25.1	Allelic Frequencies of 13 CODIS STR Loci		
Allele	Allelic Frequency (%)		
	African American	Caucasian	Hispanic
D3S1358	(*N*=210)	(*N*=203)	(*N*=209)
<12	0.476	0.000	0.000
12	0.238	0.000	0.000
13	1.190	0.246	0.239
14	12.143	14.039	7.895
15	29.048	24.631	42.584
15.2	0.000	0.000	0.000
16	30.714	23.153	26.555
17	20.000	21.182	12.679
18	5.476	16.256	8.373
19	0.476	0.493	1.435
>19	0.238	0.000	0.239
VWA	(*N*=180)	(*N*=196)	(*N*=203)
11	0.278	0.000	0.246
13	0.556	0.510	0.000
14	6.667	10.204	6.158
15	23.611	11.224	7.635
16	26.944	20.153	35.961
17	18.333	26.276	22.167
18	13.611	22.194	19.458
19	7.222	8.418	7.143
20	2.778	1.020	1.232
21	0.000	0.000	0.000
FGA	(*N* = 180)	(*N* = 196)	(*N* = 203)
<18	0.278	0.000	0.000
18	0.833	3.061	0.246
18.2	0.833	0.000	0.000
19	5.278	5.612	7.882
19.2	0.278	0.000	0.000
20	7.222	14.541	7.143
20.2	0.000	0.255	0.246
21	12.500	17.347	13.054

Table 25.1 (Continued) Allelic Frequencies of 13 CODIS STR Loci

Allele	Allelic Frequency (%)		
	African American	Caucasian	Hispanic
21.2	0.000	0.000	0.246
22	22.500	18.878	17.734
22.2	0.556	1.020	0.493
22.3	0.000	0.000	0.000
23	12.500	15.816	14.039
23.2	0.000	0.000	0.739
24	18.611	13.776	12.562
24.2	0.000	0.000	0.000
24.3	0.000	0.000	0.000
25	10.000	6.888	13.793
26	3.611	1.786	8.374
27	2.222	1.020	3.202
28	1.667	0.000	0.246
29	0.556	0.000	0.000
30	0.278	0.000	0.000
>30	0.278	0.000	0.000
D8S1179	(N = 180)	(N = 196)	(N = 203)
<9	0.278	1.786	0.246
9	0.556	1.020	0.246
10	2.500	10.204	9.360
11	3.611	5.867	6.158
12	10.833	14.541	12.069
13	22.222	33.929	32.512
14	33.333	20.153	24.631
15	21.389	10.969	11.576
16	4.444	1.276	2.463
17	0.833	0.255	0.739
18	0.000	0.000	0.000
D21S11	(N = 179)	(N = 196)	(N = 203)
24.2	0.279	0.510	0.246
24.3	0.000	0.000	0.000

(*continued*)

Table 25.1 (Continued) Allelic Frequencies of 13 CODIS STR Loci

Allele	Allelic Frequency (%)		
	African American	Caucasian	Hispanic
26	0.279	0.000	0.000
27	6.145	4.592	0.985
28	21.508	16.582	6.897
29	18.994	18.112	20.443
29.2	0.279	0.000	0.246
30	17.877	23.214	33.005
30.2	0.838	3.827	3.202
30.3	0.000	0.000	0.000
31	9.218	7.143	6.897
31.2	7.542	9.949	8.621
32	0.838	1.531	1.232
32.1	0.000	0.000	0.000
32.2	6.983	11.224	13.547
33	0.838	0.000	0.000
33.2	3.352	3.061	4.187
34	0.838	0.000	0.000
34.2	0.279	0.000	0.493
35	2.793	0.000	0.000
35.2	0.000	0.255	0.000
36	0.559	0.000	0.000
>36	0.559	0.000	0.000
D18S51	(N = 180)	(N = 196)	(N = 203)
<11	0.556	1.276	0.493
11	0.556	1.276	1.232
12	5.833	12.755	10.591
13	5.556	12.245	16.995
13.2	0.556	0.000	0.000
14	6.389	17.347	16.995
14.2	0.000	0.000	0.000
15	16.667	12.755	13.793
15.2	0.000	0.000	0.000
16	18.889	10.714	11.576

Table 25.1 (Continued) Allelic Frequencies of 13 CODIS STR Loci

Allele	Allelic Frequency (%)		
	African American	Caucasian	Hispanic
17	16.389	15.561	13.793
18	13.056	9.184	5.172
19	7.778	3.571	3.695
20	5.556	2.551	1.724
21	1.111	0.510	1.970
21.2	0.000	0.000	0.000
22	0.556	0.255	0.739
>22	0.556	0.000	1.232
D5S818	(N = 180)	(N = 195)	(N = 203)
7	0.278	0.000	6.158
8	5.000	0.000	0.246
9	1.389	3.077	5.419
10	6.389	4.872	6.650
11	26.111	41.026	42.118
12	35.556	35.385	29.064
13	24.444	14.615	9.606
14	0.556	0.769	0.493
15	0.000	0.256	0.246
>15	0.278	0.000	0.000
D13S317	(N = 179)	(N = 196)	(N = 203)
7	0.000	0.000	0.000
8	3.631	9.949	6.650
9	2.793	7.653	21.921
10	5.028	5.102	10.099
11	23.743	31.888	20.197
12	48.324	30.867	21.675
13	12.570	10.969	13.793
14	3.631	3.571	5.665
15	0.279	0.000	0.000
D7S820	(N = 210)	(N = 203)	(N = 209)
6	0.000	0.246	0.239

(*continued*)

Table 25.1 (Continued) Allelic Frequencies of 13 CODIS STR Loci

Allele	Allelic Frequency (%)		
	African American	Caucasian	Hispanic
7	0.714	1.724	2.153
8	17.381	16.256	9.809
9	15.714	14.778	4.785
10	32.381	29.064	30.622
10.1	0.000	0.000	0.000
11	22.381	20.197	28.947
11.3	0.000	0.000	0.000
12	9.048	14.039	19.139
13	1.905	2.956	3.828
14	0.476	0.739	0.478
CSF1PO	($N = 210$)	($N = 203$)	($N = 209$)
6	0.000	0.000	0.000
7	4.286	0.246	0.239
8	8.571	0.493	0.000
9	3.333	1.970	0.718
10	27.143	25.369	25.359
10.3	0.000	0.246	0.000
11	20.476	30.049	26.555
12	30.000	32.512	39.234
13	5.476	7.143	6.459
14	0.714	1.478	0.957
15	0.000	0.493	0.478
TPOX	($N = 209$)	($N = 203$)	($N = 209$)
6	8.612	0.000	0.478
7	2.153	0.246	0.478
8	36.842	54.433	55.502
9	18.182	12.315	3.349
10	9.330	3.695	3.349
11	22.488	25.369	27.273
12	2.392	3.941	9.330
13	0.000	0.000	0.239
TH01	($N = 210$)	($N = 203$)	($N = 209$)

Table 25.1 (Continued) Allelic Frequencies of 13 CODIS STR Loci

Allele	Allelic Frequency (%)		
	African American	Caucasian	Hispanic
5	0.000	0.000	0.239
6	10.952	22.660	23.206
7	44.048	17.241	33.732
8	18.571	12.562	8.134
8.3	0.000	0.246	0.000
9	14.524	16.502	10.287
9.3	10.476	30.542	24.163
10	1.429	0.246	0.239
D16S539	(N = 209)	(N = 202)	(N = 208)
8	3.589	1.980	1.683
9	19.856	10.396	7.933
10	11.005	6.683	17.308
11	29.426	27.228	31.490
12	18.660	33.911	28.606
13	16.507	16.337	10.337
14	0.957	3.218	2.404
15	0.000	0.248	0.240

Source: Budowle, B., et al., *J Forensic Sci*, 44, 1277–1286, 1999. With permission.

25.1.2.6 Probability of Match

The discriminating power of genetic loci used above can be measured by *population match probability* (P_m). P_m is defined as the probability of having a matching genotype between two randomly chosen individuals. The lower the P_m, the less likely a match between two randomly chosen individuals will occur. This is calculated as follows:

$$P_m = \sum_i \left(p^2\right)^2 + \sum_j \left(2pq\right)^2 \tag{25.4}$$

where:

p and q = the frequencies of two different alleles

P_m can also be used to compare the discriminating powers of different loci

Tables 25.2 through 25.4 show P_m values of loci commonly used for forensic applications, including SNP, VNTR, and STR.

Table 25.2 Heterozygosity and P_m Values of Six SNP Loci

Locus	Allele	Heterozygosity	P_m
DQA1	7	0.828	0.053
LDLR	2	0.493	0.379
GYPA	2	0.498	0.376
HBGG	3	0.508	0.360
D7S8	2	0.476	0.388
GC	3	0.592	0.235
Average		0.566	
Product			2.5×10^{-4} (1 in 4000)

Source: Office of Justice Programs, National Institute of Justice, US Department of Justice, *The Future of Forensic DNA Testing: Predictions of the Research and Development Working Group*, 2000.

Table 25.3 Heterozygosity and P_m Values of Six VNTR Loci

Locus	Bins	Heterozygosity	P_m
D1S7	28	0.945	0.0058
D2S44	26	0.926	0.0103
D4S139	19	0.899	0.0184
D10S28	24	0.943	0.0063
D14S13	30	0.899	0.0172
D17S79	19	0.799	0.0700
Average		0.902	
Product			8.26×10^{-12} (1 in 1.2×10^{11})

Source: Office of Justice Programs, National Institute of Justice, US Department of Justice, *The Future of Forensic DNA Testing: Predictions of the Research and Development Working Group,* 2000.

25.2 Statistical Analysis of DNA Profiling Results

It is desirable to evaluate the strength of DNA profiling results, particularly if two DNA profiles match. A DNA profile from crime scene evidence and a profile from a suspect may be the same for two reasons: (1) the crime scene sample may have come from the suspect or (2) the suspect happens to have the same profile as the individual who left the evidence found at the crime scene. The significance of a match between DNA profiles can be evaluated by using statistical calculations that determine the rarity of a specific DNA profile in a relevant population. The statistical evaluation of the significance can be included in a case report

Table 25.4 Heterozygosity and P_m Values for CODIS Loci

Locus	Allele	Caucasian American		African American	
		Heterozygosity	P_m	Heterozygosity	P_m
CSF1PO	11	0.734	0.112	0.781	0.081
TPOX	7	0.621	0.195	0.763	0.090
TH01	7	0.783	0.081	0.727	0.109
VWA	10	0.811	0.062	0.809	0.063
D16S539	8	0.767	0.089	0.798	0.070
D7S820	11	0.806	0.065	0.782	0.080
D13S317	8	0.771	0.085	0.688	0.136
D5S818	10	0.682	0.158	0.739	0.112
FGA	19	0.860	0.036	0.863	0.033
D3S1358	10	0.795	0.075	0.763	0.094
D8S1179	10	0.780	0.067	0.778	0.082
D18S51	15	0.876	0.028	0.873	0.029
D2S11	20	0.853	0.039	0.861	0.034
Average		0.7812		0.7866	
Product			1.738×10^{-15} (1 in 5.753×10^{14})		1.092×10^{-15} (1 in 9.161×10^{14})

Source: Office of Justice Programs, National Institute of Justice, US Department of Justice, *The Future of Forensic DNA Testing: Predictions of the Research and Development Working Group,* 2000.

(Figure 25.2). Guidelines for DNA profile interpretation such as those issued by the National Research Council, the DNA Advisory Board, and the European DNA Profiling Group can be consulted.

25.2.1 Genotypes

The approaches to performing statistical analysis are (1) calculation of profile probability and (2) use of the likelihood ratio method. Although profile probability is the most commonly used method because of its simplicity, both approaches lead to the same conclusion.

25.2.1.1 Profile Probability

Profile probability can be calculated based on the following steps:

1. The locus genotype frequency can be calculated as follows; p and q are the allelic frequencies observed in the database for a given allele:

   Locus genotype frequency for homozygotes:

$$P_i = p^2 \tag{25.5}$$

Locus genotype frequency for heterozygotes:

$$P_j = 2pq \tag{25.6}$$

2. Profile probability can then be calculated based on the product rule by multiplying all the locus genotype frequencies calculated as above.

The lower the profile probability is, the less likely that an individual chosen at random will have a coincident match with the DNA profile of the evidence sample. Table 25.5 shows calculations of profile probability from a DNA profile.

25.2.1.1.1 Structured Populations

The above calculation of profile probability is based on the assumption that a randomly selected individual is unrelated to a perpetrator. However, it is likely that the individual and the perpetrator are from the same subpopulation (groups within a population) and are thus not completely independent. Mating is more likely to occur within subpopulations than between subpopulations. As a result, the proportion of homozygotes increases and the proportion of heterozygotes decreases in a subpopulation because individuals in a subpopulation appear to be related.

DEPARTMENT OF FORENSIC BIOLOGY

LABORATORY REPORT

LAB NO: IQAS 2003, Lab # 259

SUSPECTS: Suspect 1, 2032
Suspect 2, 2033
Suspect 3, 2034

SUMMARY OF RESULTS:

Semen* was found on the crime scene sample 2031b, based on the presence of P30 antigen and sperm.

PCR DNA typing was done on crime scene sample 2031b. Results indicate the semen could have come from the suspect 1. This combination of DNA alleles would be expected to be found in approximately:

1 in greater than 1 trillion Blacks**
1 in greater than 1 trillion Caucasians
1 in greater than 1 trillion Hispanics
1 in greater than 1 trillion Asians

The DNA from crime scene sample 2031b could not have come from suspect 2 or suspect 3.

Amylase, a component of saliva, was not found on crime scene sample 2031b.

No semen was found on the control sample 2035.

Crime scene sample 2031a was not analyzed.

The DNA results in this case do not match any previous PCR (STR) DNA cases to date.

*Semen has two components: the seminal plasma (which contains the P30 antigen) and spermatozoa. Semen can be identified by detecting either P30 antigen and/or sperm.

** OCME STR database, National Research Council (1996) The Evaluation of Forensic DNA Evidence, Natl. Acad. Press, Washington DC.

Figure 25.2 An example of a laboratory report. (© Richard C. Li.)

Table 25.5 Calculation of Profile Probability of STR Profile with 13 CODIS Loci

Locus	Profile	Allelic Frequency[a]	Formula	Locus Genotype Frequency
CSF1PO	10	0.25369	$2pq$	0.165
	12	0.32512		
D3S1358	14	0.14039	$2pq$	0.0691
	15	0.24631		
D5S818	11	0.41026	p^2	0.168
	11	0.41026		
D7S820	10	0.29064	$2pq$	0.117
	11	0.20197		
D8S1179	13	0.33929	p^2	0.115
	13	0.33929		
D13S317	11	0.31888	p^2	0.102
	11	0.31888		
D16S539	11	0.27228	$2pq$	0.185
	12	0.33911		
D18S51	15	0.12755	$2pq$	0.00911
	19	0.035710		
D21S11	30	0.23214	p^2	0.0539
	30	0.23214		
FGA	23	0.15816	$2pq$	0.0436
	24	0.13776		
TH01	8	0.12562	$2pq$	0.0767
	9.3	0.30542		
TPOX	8	0.54433	p^2	0.296
	8	0.54433		
VWA	17	0.26276	$2pq$	0.117
	18	0.22194		
				Profile probability $= 2.76 \times 10^{-14}$

[a] See Table 25.1.

The effect of population structure should be considered and an appropriate correction should be made to estimate profile probabilities. The correction can be made by using the factor θ. Thus, the locus genotype frequency can be calculated as follows:

Locus genotype frequency for homozygotes:

$$P_i = \frac{[2\theta + (1-\theta)p][3\theta + (1-\theta)p]}{(1+\theta)(1+2\theta)} \tag{25.7}$$

Locus genotype frequency for heterozygotes:

$$P_j = \frac{2[\theta + (1-\theta)p][\theta + (1-\theta)q]}{(1+\theta)(1+2\theta)} \tag{25.8}$$

The θ value is 0.01 for the majority US population and 0.03 for the Native American population. Table 25.6 shows the calculation of profile probability with subpopulation correction using $\theta = 0.01$. The profile probability is approximately three times higher than the value without the correction (Table 25.5). Additional corrections can be calculated for relatives, mixed stains, or database searches using formulas provided by the National Research Council's guidelines.

25.2.1.2 Likelihood Ratio

The *likelihood ratio* (LR) method is an alternative for evaluating the strength of a match. The method allows the calculation of the probability of the DNA profile under two hypothesis:

Hypothesis 1 (H_1) —The evidence and suspect profiles originated from the same source.

Hypothesis 2 (H_2) —The evidence and suspect profiles did not originate from the same source (i.e., the suspect happens to have the same profile as that of the individual who left the evidence).

The LR is the probability of hypothesis H_1 divided by the probability of hypothesis H_2. Where Pr_1 is the probability under hypothesis H_1, and Pr_2 is the probability under hypothesis H_2, this can be expressed as:

$$\mathrm{LR} = \frac{\mathrm{Pr}_1}{\mathrm{Pr}_2} \tag{25.9}$$

The greater the numerator (Pr_1), the greater the likelihood ratio becomes. The result favors hypothesis H_1 (the evidence and suspect profile originated from the same source). Pr_1 is equal to 1 (100%) when a match occurs, and Pr_2 is equal to the profile probability. A LR of 1000 indicates that the evidence is 1000 times as probable if the evidence and suspect profiles originated from the same source.

25.2.2 Haplotypes

The term *haplotype* was first used to describe very closely linked polymorphic loci. During meiosis, alleles at neighboring loci cosegregate (both alleles segregate as a single allele) because of the close linkage of loci. The term also applies to genetic regions within which recombination events are very rare, that is, within mitochondrial and Y-chromosomal DNA. The entire mitochondrial DNA (mtDNA) sequence can be considered to be a single locus or haplotype because of the absence of recombination. Likewise, Y chromosome loci can also be considered haplotypes.

Where recombination is very rare, certain allelic combinations occur in populations much more frequently than would be expected. This phenomenon is called *linkage disequilibrium*. As a result, the HW principle cannot be applied. The two methods for evaluating the strength of

Table 25.6 Calculation of Profile Probability of STR with 13 CODIS Loci and Correction Factor ($\theta = 0.01$)

Locus	Profile	Allelic Frequency[a]	Formula	Locus Genotype Frequency
CSF1PO	10	0.25369	$\frac{2[\theta+(1-\theta)p][\theta+(1-\theta)q]}{(1+\theta)(1+2\theta)}$	0.168
	12	0.32512		
D3S1358	14	0.14039	$\frac{2[\theta+(1-\theta)p][\theta+(1-\theta)q]}{(1+\theta)(1+2\theta)}$	0.0734
	15	0.24631		
D5S818	11	0.41026	$\frac{[2\theta+(1-\theta)p][3\theta+(1-\theta)p]}{(1+\theta)(1+2\theta)}$	0.180
	11	0.41026		
D7S820	10	0.29064	$\frac{2[\theta+(1-\theta)p][\theta+(1-\theta)q]}{(1+\theta)(1+2\theta)}$	0.121
	11	0.20197		
D8S1179	13	0.33929	$\frac{[2\theta+(1-\theta)p][3\theta+(1-\theta)p]}{(1+\theta)(1+2\theta)}$	0.126
	13	0.33929		
D13S317	11	0.31888	$\frac{[2\theta+(1-\theta)p][3\theta+(1-\theta)p]}{(1+\theta)(1+2\theta)}$	0.113
	11	0.31888		
D16S539	11	0.27228	$\frac{2[\theta+(1-\theta)p][\theta+(1-\theta)q]}{(1+\theta)(1+2\theta)}$	0.187
	12	0.33911		
D18S51	15	0.12755	$\frac{2[\theta+(1-\theta)p][\theta+(1-\theta)q]}{(1+\theta)(1+2\theta)}$	0.012
	19	0.03571		
D21S11	30	0.23214	$\frac{[2\theta+(1-\theta)p][3\theta+(1-\theta)p]}{(1+\theta)(1+2\theta)}$	0.0630
	30	0.23214		
FGA	23	0.15816	$\frac{2[\theta+(1-\theta)p][\theta+(1-\theta)q]}{(1+\theta)(1+2\theta)}$	0.0473
	24	0.13776		
TH01	8	0.12562	$\frac{2[\theta+(1-\theta)p][\theta+(1-\theta)q]}{(1+\theta)(1+2\theta)}$	0.0815
	9.3	0.30542		
TPOX	8	0.54433	$\frac{[2\theta+(1-\theta)p][3\theta+(1-\theta)p]}{(1+\theta)(1+2\theta)}$	0.309
	8	0.54433		
VWA	17	0.26276	$\frac{2[\theta+(1-\theta)p][\theta+(1-\theta)q]}{(1+\theta)(1+2\theta)}$	0.120
	18	0.22194		
				Profile probability $= 7.8 \times 10^{-14}$

[a] See Table 23.1.

a match between haplotypes are (1) mitotype frequency and (2) likelihood ratios. The current most common approach for interpreting mtDNA profiles is mitotype frequency carried out with the gene counting method, that is, the calculation of the number of occurrences of a particular sequence or haplotype. The interpretation of Y chromosome profiles is similar to the interpretation of mtDNA profiles. The estimation of the frequency of a mitotype can be calculated as shown below.

25.2.2.1 Mitotypes Observed in Database

If a mitotype is observed at least once in a database, Equation 25.10 can be used. P_{mt} is the mitotype frequency, x is the number of observations of the haplotype, and n is the size of the database (the number of mitotype entries):

$$P_{mt} = \frac{x+2}{n+2} \tag{25.10}$$

Any sampling error may be addressed by a confidence interval:

$$P_{mt} \pm 1.96\sqrt{\frac{P_{mt}\left(1-P_{mt}\right)}{n}} \tag{25.11}$$

In this case, P_{mt} is the mitotype frequency, and n is the size of the database. The conservative upper bound of the frequency is usually quoted.

25.2.2.2 Mitotype Not Observed in Database

If a mitotype has not been observed in a database, Equation 25.12 can be used to calculate the mitotype frequency; α is 0.05 for a 95% confidence interval and n is the size of the database.

$$P_{mt} = 1 - \alpha^{1/n} \tag{25.12}$$

Bibliography

Aitken, C.G.G., *Statistics and the Evaluation of Evidence for Forensic Scientists*, p. 260, 1995. New York: Wiley.

Aktulga, H.M., et al., Identifying statistical dependence in genomic sequences via mutual information estimates. *EURASIP J Bioinform Syst Biol*, 2007, **2007**(1): 14741.

Allen, R.W., et al., Considerations for the interpretation of STR results in cases of questioned half-sibship. *Transfusion*, 2007, **47**(3): 515–519.

Amorim, A., A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers. *Forensic Sci Int Genet*, 2008, **2**(4): 376–378.

Andersen, M.M., P.S. Eriksen, and N. Morling, The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. *J Theor Biol*, 2013, **329**: 39–51.

Andersen, M.M., et al., Estimating trace-suspect match probabilities for singleton Y-STR haplotypes using coalescent theory. *Forensic Sci Int Genet*, 2013, **7**(2): 264–271.

Balding, D.J., The prosecutor's fallacy and DNA evidence. *Crim Law Rev*, 1994, 711–721.

Balding, D.J., When can a DNA profile be regarded as unique? *Sci Justice*, 1999, **39**(4): 257–260.

Balding, D.J. and P. Donnelly, Inferring identify from DNA profile evidence. *Proc Natl Acad Sci U S A*, 1995, **92**(25): 11741–11745.

Balding, D.J. and P. Donnelly, Evaluating DNA profile evidence when the suspect is identified through a database search. *J Forensic Sci*, 1996, **41**(4): 603–607.

Balding, D.J. and R.A. Nichols, DNA profile match probability calculation: How to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int*, 1994, **64**(2–3): 125–140.

Beecham, G.W. and B.S. Weir, Confidence interval of the likelihood ratio associated with mixed stain DNA evidence. *J Forensic Sci*, 2011, **56**(Suppl 1): S166–S171.

26

Quality Assurance and Quality Control

Quality assurance (*QA*) for forensic services requires certain processes to ensure that a service will meet laboratory requirements for the integrity of testing. A QA program must include components that address:

- Continuing education, training, and certification of personnel
- Specification and calibration of equipment and reagents
- Documentation and validation of analytic methods
- Use of appropriate standards and controls
- Sample handling procedures
- Proficiency testing
- Data interpretation and reporting
- Audits (internal and external) and laboratory accreditation
- Corrective actions to address deficiencies and assessments for laboratory competence

Over the years, many guidelines for quality assurance in forensic DNA laboratories have been established. These guidelines will be introduced in this chapter.

Quality control (*QC*) for forensic services refers to the operational procedures necessary to meet quality requirements. QC procedures may include maintenance of calibration records for equipment and instruments as well as the testing of chemical reagents and supplies used in analysis to ensure reliable results.

26.1 US Quality Standards

DNA profiling methods were first used in criminal investigations in the 1980s. By the early 1990s, emerging forensic DNA techniques had undergone detailed reviews by the *National Research Council* (*NRC*) of the National Academy of Sciences. In 1992, the first published NRC report included recommendations in the areas of (1) technical considerations, (2) statistical interpretation, (3) laboratory standards, (4) data banks and privacy, (5) legal considerations, and (6) societal and ethical issues related to forensic DNA testing. The NRC report attempted to explain the basic scientific principles of forensic DNA technology and made suggestions for applications and improvements. However, the report received negative criticism from both the forensic and the legal communities.

In 1996, a second NRC committee was formed "to update and clarify discussion of the principles of population genetics and statistics as they apply to DNA evidence." Its report stated:

> The central question that the report addresses is this: What information can a forensic scientist, population geneticist, or statistician provide to assist a judge or jury in drawing inferences from the finding of a match? To answer this question, the committee reviewed the scientific literature and the legal cases and commentaries on DNA profiling, and it investigated the criticisms that have been voiced about population data, statistics, and laboratory error. Much has been learned since the last report. The technology for DNA profiling as well as the methods for estimating frequencies and related statistics have progressed to the point where the reliability and validity of properly collected and analyzed DNA data should not be in doubt. The new recommendations presented here should pave the way to more effective use of DNA evidence. (*The Evaluation of Forensic DNA Evidence* also known as the NRC II report)

The NRC II report consisted of (1) an introduction describing the 1992 report, changes made subsequent to that report, and the validity and application of DNA typing techniques; (2) assurance of high standards of laboratory performance; (3) population genetics issues; (4) statistical issues; and (5) DNA evidence in the legal system.

In 1995, The *DNA Advisory Board* (*DAB*) was formed as a result of the DNA Identification Act (1994) passed by Congress. The DAB served from 1995 to 2000 to develop guidelines for quality assurance in forensic laboratories. During that time, the DAB provided two sets of guidelines: Quality Assurance Standards for Forensic DNA Testing Laboratories (1998) and Quality Assurance Standards for Convicted Offender DNA Databasing Laboratories (1999). These standards describe the requirements to ensure quality and integrity of the data as well as competency of laboratories.

These standards were built on the previous standards set by the *Scientific Working Group on DNA Analysis Methods* (*SWGDAM*). After the DAB's assignment ended in 2000, the SWGDAM became responsible for providing guidelines to the US forensic community. The SWGDAM was established in 1988 by the FBI Laboratory. It was initially called the Technical Working Group on DNA Analysis Methods; the name changed in 1998. The SWGDAM comprises forensic scientists from DNA laboratories in the United States and Canada. Its purpose is to facilitate forensic DNA community discussions regarding necessary laboratory methods and to share protocols for forensic DNA testing. The FBI sponsors and hosts its meetings and plays an important role in its activities.

The SWGDAM established and revised several guidelines including the Guidelines for a Quality Assurance Program for DNA Analysis, published in 1989, 1991, and 1995 (the validation section was revised in 2003); Quality Assurance Standards for Forensic DNA Testing Laboratories (2011); Quality Assurance Standards for DNA Databasing Laboratories (2011); Validation Guidelines for DNA Analysis Methods (2012); and SWGDAM Training Guidelines, published in 2001 and revised in 2013.

The SWGDAM also formed subcommittees to provide guidelines in more specific areas of forensic DNA testing, for example, SWGDAM's Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories (2010), Y-chromosome Short Tandem Repeat (Y-STR) Interpretation Guidelines (2009), and Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories (2013). The SWGDAM also organized a number of interlaboratory and validation studies for new techniques.

26.2 International Quality Standards

In the early stages of forensic DNA testing, the *International Society for Forensic Genetics* (*ISFG*) recognized the potential of DNA testing for criminal investigations and made a number of recommendations related to the forensic application of DNA polymorphisms.

The ISFG provided various recommendations for forensic DNA testing of STR, mtDNA, and Y chromosome markers for the international community. It formed a working group called the *European DNA Profiling Group* (*EDNAP*) in 1991. The EDNAP has investigated systems for DNA profiling, has organized a number of collaborative exercises for the evaluation of new methods, and has published reports of its studies.

The *European Network of Forensic Science Institutes* (*ENFSI*) formed its DNA working group about a decade ago to address issues of quality and standards for forensic DNA testing. The *Interpol European Working Party on DNA Profiling* (*IEWPDP*) also makes recommendations for applying DNA evidence to criminal investigations in Europe. Based on the EDNAP exercises and recommendations by the ENFSI and IEWPDP, the *European Standard Set* (*ESS*) for autosomal STR core loci was established. The Standardization of DNA Profiling Techniques in the European Union (STADNAP) group has been working on the selection of forensic DNA profiling systems, methods for use among European countries, and the maintenance of European population databases.

26.3 Laboratory Accreditation

Accreditation is the process used to assess the qualification of a laboratory to meet established standards. During an accreditation process, the services and performance of a laboratory are evaluated, particularly in the areas of management, operations, personnel, procedures, equipment, physical plants, security, and personnel safety procedures. The accreditation process generally involves several components such as self-evaluation, the preparation of supporting documents, on-site inspection and reports, and accreditation review reports.

Accreditation in the United States is offered by the *Laboratory Accreditation Board* of the *American Society of Crime Laboratory Directors* (*ASCLD/LAB*) for forensic laboratories performing casework. The *American Association of Blood Banks* (*AABB*) provides accreditation for laboratories performing DNA parentage testing according to the AABB's standards.

The accreditation of a forensic laboratory is granted for 5 years. To remain in compliance, a laboratory must undergo audits to evaluate its operation according to established guidelines. The FBI has published the Quality Assurance Standards Audit for Forensic DNA Testing Laboratories (2011) and the Quality Assurance Standards Audit for DNA Databasing Laboratories (2011). The areas of operating protocols, instruments and equipment, and personnel training are evaluated based on guidelines. Problems identified during an audit must be documented and actions to resolve the problems must be addressed. Annual internal audits and external audits during alternate years are required under the guidelines.

26.4 Laboratory Validation

Validation is the process of confirming that a laboratory procedure is sufficiently robust, reliable, and reproducible. A robust method maintains successful performance and can cope with errors. A reliable method produces accurate results. A reproducible method achieves the same or very similar results each time a sample is tested.

Two types of laboratory validations, developmental and internal validations, are used for modifying methods for forensic DNA analysis. An *internal validation* is required when adopting a procedure for forensic applications. Based on the SWGDAM's Validation Guidelines for DNA Analysis Methods, the internal validation is "an accumulation of test data within the laboratory to demonstrate that established methods and procedures perform as expected in the laboratory." The *developmental validation* "is the acquisition of test data and determination of conditions and limitations of a new or novel DNA methodology for use on forensic, database, known or casework reference samples." For example, the characterization of a new genetic marker requires

studies in the inheritance, the genomic location, the detection method, and the polymorphism of the genetic marker. However, developmental validation studies in the same subject but from fields other than forensic DNA analysis may be acceptable for forensic applications.

Additionally, the precision and the accuracy of the test are important criteria for developmental validation. Based on the SWGDAM, *precision* "characterizes the degree of mutual agreement among a series of individual measurements, values and/or results. Precision depends only on the distribution of random errors and does not relate to the true value or specified value." The measure of precision can be expressed as a standard deviation of test results. *Accuracy* is referred to as "the ability of a measuring instrument to give responses close to a true value," which can be assessed by a *performance check*, a quality assurance procedure monitoring the performance of instruments and equipment affecting the accuracy of forensic DNA testing.

26.5 Proficiency Testing

Proficiency testing is an important component of quality control and quality assurance. It evaluates a laboratory's performance of DNA analyses according to the laboratory's standard protocols. Proficiency testing also evaluates the quality of performance by individual analysts following laboratory protocols.

Proficiency tests of DNA analysts must be conducted every 6 months based on DNA Advisory Board Standards. The testing usually involves mock forensic case samples including questioned bodily fluid stains and reference samples. The test is assigned to an analyst for processing according to the laboratory procedures. A report must be prepared and is then reviewed. The proficiency test can be administered as either an open or a blind test. In the blind test, the analyst is not aware that he or she is being tested. Blind testing is considered a more effective means of evaluating performance.

The tests may be administered internally or by any of a number of external testing organizations. For example, Orchid Cellmark provides the *International Quality Assessment Scheme* (*IQAS*) DNA proficiency test, and the Collaborative Testing Services (CTS) Forensics Testing Program offers similar proficiency tests. These proficiency tests may include a selection of sample types (neat or mixture) for serological tests and/or DNA analysis (autosomal STR, Y-STR, and mitochondrial DNA). The College of American Pathologists (CAP) has a paternity testing proficiency program offering external proficiency tests. In Europe, the *German DNA Profiling Group* (*GEDNAP*) provides proficiency testing for participating European laboratories.

26.6 Certification

Certification is a voluntary process that recognizes the attainment of professional qualifications needed for practice in forensic services. Certification is not required, but is desired by some laboratories. In 2009, a report entitled Strengthening Forensic Science in the United States: A Path Forward was published by the Committee on Identifying the Needs of the Forensic Science Community at the National Academy of Sciences. The report recommends that certification should be mandatory for forensic science professionals, which includes written examinations, supervised practice, proficiency testing, and compliance to a code of ethics (Section 26.8). In the United States, the *American Board of Criminalistics* (*ABC*) offers three types of certification for forensic scientists. A diplomate must pass a general knowledge examination. ABC also requires a bachelor's degree in a natural science and 2 years of experience in a forensic laboratory. To obtain fellow status (higher than diplomate status), an applicant must have 2 years of experience in his or her specialty and must have met the diplomate requirements in addition to passing a written specialty examination and a proficiency test.

ABC has added a third certification that is completely separate from the other certification programs. The technical specialist certification for molecular biology was created to recognize the qualifications required for the analysis of biological materials through DNA profiling. An applicant must take a specialist examination containing questions from the general knowledge examination and a subset of questions from the forensic biology fellow examination. This certification also requires a bachelor's degree in a natural science and 2 years of experience along with successful completion of a proficiency examination within 12 months of taking the technical specialist certification examination.

26.7 Forensic DNA Analyst Qualifications

DAB's Quality Assurance Standards for Forensic DNA Testing Laboratories require that an examiner or analyst have "at a minimum a BA/BS degree or its equivalent degree in a biology, chemistry or forensic science-related area and must have successfully completed college course work (graduate or undergraduate level) covering the subject areas of biochemistry, genetics and molecular biology," as well as "course work and/or training in statistics and population genetics as it applies to forensic DNA analysis."

Additionally, SWGDAM Training Guidelines discuss the course work requirements in detail. At least nine cumulative semester hours are required for the course work covering the required subject areas. Particularly, the section in Fundamental Scientific Knowledge of the SWGDAM Training Guidelines defines required key elements to be covered in course work. The key elements also aid in evaluating the contents of the course work.

Biochemistry "refers to the nature of biologically important molecules in living systems, DNA replication and protein synthesis, and the quantitative and qualitative aspects of cellular metabolism," and may include but is not limited to:

- Structure and function of cellular components such as proteins, carbohydrates, lipids, nucleic acids, and other biomolecules
- Chemistry of enzyme-catalyzed reactions
- Metabolism
- DNA and RNA
- Protein synthesis
- Cell membrane transport
- Signal transduction

Genetics "refers to the study of inherited traits, genotype/phenotype relationships, and population/species differences in allele and genotype frequencies," and may include but is not limited to:

- Heredity
- Function of genes
- Gene expression
- Recombinant DNA
- Mitosis/meiosis

Molecular biology "covers theories, methods, and techniques used in the study and analysis of gene structure, organization, and function," and may include, but is not limited to:

- Interrelationship of DNA, RNA, and protein synthesis
- Central dogma

- Transcription, translation, replication
- Recombinant DNA techniques
- PCR
- Cloning

Course work that is similar in content and scope to those described above may be qualified as required course work. Compliance with the required course contents can be evaluated through transcripts, syllabi, and letters from the instructors.

DAB's Quality Assurance Standards for Forensic DNA Testing Laboratories also describes requirements for an analyst to meet prior to initiating independent casework analysis using DNA technology. First, an analyst must have at least 6 months of forensic DNA laboratory experience "including the successful analysis of a range of samples typically encountered in forensic casework." Second, an analyst must have "successfully completed a qualifying test before beginning independent casework responsibilities."

26.8 Code of Ethics of Forensic Scientists

Ethical codes for forensic scientists are used as guides for individuals making their decisions in distinguishing the difference between correctness and incorrectness. These codes are usually adopted by forensic organizations to regulate the profession. Failure to comply with a code of forensic practice can raise doubt in an individual's fitness for providing forensic services. Additionally, it may result in the expulsion of an individual from a forensic organization. A typical document of ethical codes for forensic scientists generally contains five sections: (1) ethics relating to the scientific method, (2) ethics relating to opinions and conclusions, (3) ethical aspects of court presentation, (4) ethics relating to the general practice of forensics, and (5) ethical responsibilities to the profession.

For the first section, the ethical application of scientific methods is discussed in detail. True scientific methods should be utilized to adequately analyse all the evidence. Such analyses should not be conducted by secret processes. Conclusions should be drawn from the analyses of evidence that appears representative, typical, or reliable.

Moreover, ethical standards relating to opinions and conclusions are very important to the forensic profession. Conclusions should always be drawn from the application of proven scientific methods. The purpose of experimental design and the interpretation of results is to reveal facts. During an analysis, experimental controls should always be utilized. If necessary, the results of analyses should be verified by repeating the analysis or using additional techniques. Explanations should be provided where inconclusive results are obtained. The opinion provided by a forensic scientist should be unbiased and should not be influenced by matters unrelated to the specific evidence in question. A forensic scientist should not choose the interpretation that is in favor of the side of his or her employer.

Pertaining to the courtroom, ethical standards relating to expert witness testimonies is a crucial component for a forensic scientist. An expert opinion is defined as a formal consideration of a subject within an individual's knowledge, skill, education, training, and experience. A forensic scientist should not extend an opinion beyond his or her competence. Appropriate terms should be used to represent the degrees of certainty of an expert opinion. A forensic scientist should not only present the evidence that supports the view of the side of his or her employer. When testifying as an expert witness, language that can be understood by lay jurors should be used to avoid misinterpretation.

Furthermore, in terms of the general practice of forensic science, a few things should be mentioned here. A forensic scientist in private practice should set a reasonable fee for the services provided, which should not be rendered on a contingency fee basis. When a different opinion

is presented by another analyst in the reexamination of evidence, both analysts should resolve their contradiction before the trial. A forensic scientist may serve, in an advisory capacity, either the prosecutor or defense in the cross-examination of another expert.

Lastly, ethical responsibilities to the profession are an important component of ethical codes. Information regarding new developments or techniques of forensic analysis should be made available to other forensic scientists. Likewise, any information regarding methods in use that may appear unreliable should be brought to the attention of others. Individual forensic scientists should not seek publicity for the association of their name with specific cases, developments, publications, or organizations solely for the purpose of gaining personal prestige.

Bibliography

Andersen, J., et al., Report on the third EDNAP collaborative STR exercise. *Forensic Sci Int*, 1996, **78**(2): 83–93.

Balazic, J. and I. Zupanic, Quality control and quality assurance in DNA laboratories: Legal, civil and ethical aspects. *Forensic Sci Int*, 1999, **103**: S1–S5.

Bar, W., et al., DNA recommendations. Further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. *Int J Legal Med*, 1997, **110**(4): 175–176.

Buckleton, J., Validation issues around DNA typing of low level DNA. *Forensic Sci Int Genet*, 2009, **3**(4): 255–260.

Budowle, B., et al., A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. *J Forensic Sci*, 2009, **54**(4): 798–809.

Carracedo, A., et al., DNA commission of the International Society for Forensic Genetics: Guidelines for mitochondrial DNA typing. *Forensic Sci Int*, 2000, **110**(2): 79–85.

Carracedo, A., et al., Results of a collaborative study of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format. *Forensic Sci Int*, 2001, **119**(1): 28–41.

Cormier, K., L. Calandro, and D.J. Reeder, Evolution of the quality assurance documents for DNA laboratories. *Forensic Mag*, 2005, 2(1): 16–19.

Council, N.R., *NRC II: The Evaluation of Forensic DNA Evidence*, 1996. Washington, D.C.: National Academy Press.

Coyle, H.M., The importance of scientific evaluation of biological evidence—Data from eight years of case review. *Sci Justice*, 2012, **52**(4): 268–270.

DNA Advisory Board, Federal Bureau of Investigation, US Department of Justice, Quality assurance standards for forensic DNA testing laboratories. *Forensic Sci Commun*, 2000, **2**(3). Retrieved from http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2000/codis2a.htm.

Duewer, D.L., K.L. Richie, and D.J. Reeder, RFLP band size standards: NIST standard reference material 2390. *J Forensic Sci*, 2000, **45**(5): 1093–1105.

Duewer, D.L., et al., NIST mixed stain studies #1 and #2: Interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. *J Forensic Sci*, 2001, **46**(5): 1199–1210.

Frank, W.E., et al., Validation of the AmpFlSTR Profiler Plus PCR amplification kit for use in forensic casework. *J Forensic Sci*, 2001, **46**(3): 642–646.

Gill, P., et al., Report of the European DNA profiling group (EDNAP)—Towards standardisation of short tandem repeat (STR) loci. *Forensic Sci Int*, 1994, **65**(1): 51–59.

Gill, P., et al., Report of the European DNA profiling group (EDNAP): An investigation of the complex STR loci D21S11 and HUMFIBRA (FGA). *Forensic Sci Int*, 1997, **86**(1–2): 25–33.

Gill, P., et al., Report of the European DNA profiling group (EDNAP)—An investigation of the hypervariable STR loci ACTBP2, APOAI1 and D11S554 and the compound loci D12S391 and D1S1656. *Forensic Sci Int*, 1998, **98**(3): 193–200.

Gill, P., et al., DNA commission of the International Society of Forensic Genetics: Recommendations on forensic analysis using Y-chromosome STRs. *Int J Legal Med*, 2001, **114**(6): 305–309.

FORENSIC SCIENCE

FORENSIC BIOLOGY

SECOND EDITION

Over the last several years, new research and developments in analysis methods and practice have led to rapid advancements in forensic biology. Identifying critical points of knowledge and new methodological approaches in the field, ***Forensic Biology, Second Edition*** focuses on forensic serology and forensic DNA analysis. It provides students and professionals with a scientific grounding in biological evidence—both the techniques utilized to identify it and the methodology to analyze it.

This second edition:

- Introduces the language of forensic biology, enabling students to become comfortable with usage and terminology
- Provides clear explanations of the principles of forensic identification and analysis of biological evidence
- Explains forensic serology and DNA techniques used in the field and the laboratory
- Discusses the benefits and limitations that apply to various forensic biology techniques
- Includes schematic illustrations to clarify concepts
- Presents three new chapters (Chapters 2, 16, and 17) created for this edition
- Adds more than two hundred new full-color figures

Covering the full scope of forensic biology, the book uses an accessible, straightforward style designed to enhance students' education and training so they are prepared, both in the laboratory and in the field.

K14098

ISBN: 978-1-4398-8970-1
90000
9 781439 889701

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

www.crcpress.com

Color Atlas of Forensic Medicine and Pathology

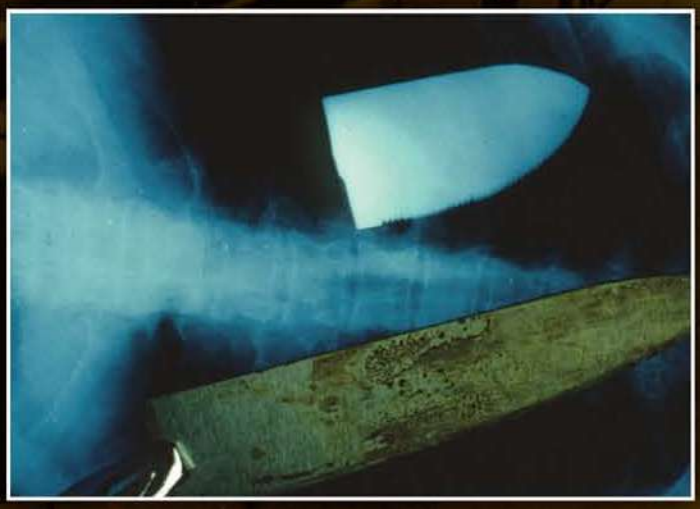

Edited by
Charles A. Catanese

CRC Press
Taylor & Francis Group

Table of Contents

Preface ix

Acknowledgments xi

Editor xiii

Contributors xv

Research Assistants and Photo Imaging Contributors xvii

1 Sudden Natural Death in a Forensic Setting 1
CHARLES A. CATANESE AND AMY V. RAPKIEWICZ

2 Therapeutic Interventions, Complications, and Accidents 57
MICHAEL J. CAPLAN AND CHARLES A. CATANESE

3 Substance Abuse and Poisoning 85
CHARLES A. CATANESE AND LAURA M. LABAY

4 Postmortem Change and Time of Death 111
CHARLES A. CATANESE, BRUCE LEVY, AND GERARD CATANESE

5 Pediatric Forensic Pathology 147
MICHAEL J. CAPLAN AND CHARLES A. CATANESE

6 Blunt-Force Injuries 189
CHARLES A. CATANESE AND JONATHAN LUCAS

7 Sharp-Force Injuries 239
CHARLES A. CATANESE

8 Gunshot Wounds 283
CHARLES A. CATANESE, THOMAS GILSON, AND THOMAS ANDREW

9 Burns 353
CHARLES A. CATANESE AND GERARD CATANESE

10 Asphyxia 373
CHARLES A. CATANESE AND BARBARA K. BOLLINGER

Index 399

Preface

While death investigation, and what we have come to understand as "forensic pathology," has been practiced in one way or another back to antiquity, official designation of forensic pathology as a subspecialty by the American Board of Pathology dates back only to 1956. The subspecialty comprises a small band of trainees in pathology and at any given time there are only 400 to 500 full-time practitioners of the specialty in the United States.

Furthermore, as a recognized subspecialty, forensic pathology is young enough that each of its full-time practitioners can trace his or her roots back to one of the six individuals who sat for that original examination back in 1956. For many of the authors of this volume, that family tree goes from Lester Adelson through Charles Hirsch and ultimately to us. We acknowledge and are proud of the fact that we stand on the shoulders of giants.

Our collective experiences as trainees and staff at the Office of the Chief Medical Examiner in New York City under the tutelage of Dr. Charles S. Hirsch has made us the forensic pathologists we are today. Our mentor's emphasis on precision and accuracy in description of findings, translation of these descriptions into language easily understood by a broad range of end users, and the public health importance of our work has left an indelible imprint that we desire to pass on to others. It is our hope that this atlas reflects these qualities.

With sincerest gratitude,

ȧ omas Andrew
Fellowship Class of 1992

Michael J. Caplan
Gerard Catanese
Bruce Levy
Fellowship Class of 1993

ȧ omas Gilson
Charles A. Catanese
Fellowship Class of 1995

Barbara K. Bollinger
Jonathan Lucas
Fellowship Class of 2000

Sudden Natural Death in a Forensic Setting

1

CHARLES CATANESE AND AMY V. RAPKIEWICZ

Introduction

This chapter offers a brief overview of some common and some not so common natural deaths that typically may occur in a medical examiner system. Also demonstrated are examples that may alter appearance of tissue such as formaldehyde fixation and variation due to different types of photographic imagery. There are also examples of normal organs in both a fresh and formaldehyde fixed state that can be used by the reader to compare with diseased organs.

Deaths under this category are often unexpected and sometimes unwitnessed. There is often a suspicion of foul play. Families may say, "But Doctor, he was in fine health. I saw him an hour ago. It cannot be natural. Somebody must have harmed him," etc. Because of the sudden, often unexpected nature of these deaths, it is best to do an autopsy to clarify exactly what happened. This decision to autopsy depends on many factors, including the decedent's age, medical history, family wishes, decedent's wishes (wills, etc.), religious beliefs, circumstance at time of death, resources of a particular system, etc. As one becomes less certain of the cause of death, the level of suspicion will increase. At some point, the decision to autopsy becomes obvious and absolutely necessary. This decision is based on experience, knowledge, and sound judgment. Not infrequently, seemingly natural deaths can have unnatural or traumatic previous circumstances; therefore, when uncertain, an autopsy is best performed. In many medical examiner systems, the majority of deaths end up being certified as natural.

Sudden death is defined in different ways. It may indicate a death that occurs within 24 hours of the onset of symptoms. It may also indicate the death occurred within 1 hour or even within seconds. There are not many diseases that can cause death within minutes of the onset of symptoms. Natural death means the manner of death is exclusively or 100% natural. If there is a 1% component of another manner of death, it is no longer natural. If there are multiple components of different manners of death commingled in a case investigation, the following rule will apply: a homicide overrides all, then an accident, then a therapeutic complication. A suicide requires the establishment of intent to do harm to oneself. For example, someone with end-stage metastatic liver cancer ingests 100 acetaminophen tablets to commit suicide. In the process of waiting to die, he decides to walk to a store. On the way, he trips in a pothole, falls in the street, strikes his head, and has an expanding subdural hemorrhage. While he is lying there waiting for EMS, a stolen car fleeing the scene of a robbery runs him over, lacerates his heart in half, and he dies within seconds. The manner of death in the case would be homicide. The death occurred as a result of being run over by a car during an illegal act. A lacerated heart is universally fatal regardless of the other violent and natural processes. The death certificate should include only the trauma from the car. If the trauma from the car was not lethal by itself, one may add "other finding" to part two of the cause of death, but the manner would remain homicide.

Heart Disease

Heart disease leading to ventricular irritability to create a lethal arrhythmia is the most significant cause of death in this category. The most common arrhythmia leading to sudden cardiac death is ventricular fibrillation. Ventricular tachyarrhythmias are most commonly seen within 12 hours of a myocardial infarction. Critical coronary atherosclerosis and hypertension are by far the leading causes of these processes. Some diseases that contribute to atherosclerosis and arteriosclerosis formation include hyperlipidemia, high blood pressure, diabetes mellitus, obesity, cigarette smoking, stress, and sedentary life style.

Having 75% or greater blockage in any of the epicardial vessels is considered critical stenosis and is consistent with being alive one second and having loss of consciousness leading to death the next. Hypertensive cardiovascular disease is usually essential in origin from an intrinsic abnormality of sodium metabolism. Other significant causes of hypertension include many types of kidney disease including adult polycystic kidney disease and renal artery stenosis. Hypertension may be sporadic and missed on routine doctor appointments. High blood pressure is also associated with small-vessel coronary artery disease, as is diabetes mellitus, which is a reasonable cause of death by itself. Once people reach a pivotal point of myocardial irritability and go

into ventricular fibrillation they usually have approximately 15 seconds of consciousness left. Prior to losing consciousness, decedents may reach up to chest or neck and mention a fluttering sensation in the chest. They may have pressure, pain, or no expectation of what is to come. Ventricular irritability associated with coronary artery ischemia is due to lack of oxygen and nutrients reaching the conducting system of the heart. If the heart is not cardioverted back to a normal rhythm within 4–6 minutes, there is usually irreversible brain damage.

Another major cause of ventricular irritability leading to fatal arrhythmia is hypertension. Concentric left ventricular hypertrophy usually defined at autopsy as having a left ventricular wall thickness greater than 1.5 cm for most average-sized adults is a known risk factor for sudden cardiac death. Left ventricular thickness is best measured approximately 2 cm below the mitral valve annulus and excludes trabeculations and papillary muscles. As the disease process causing cardiac hypertrophy advances, heart failure may ensue with chamber dilatation. Although the overall heart size is enlarged, the left ventricle wall thickness may be less than 1.4 cm. Although hypertensive disease is the major risk factor for the development of left ventricular hypertrophy, other risk factors include aortic stenosis, either congenital or acquired. The hearts of patients with hypertensive or arteriosclerotic cardiovascular disease typically show evidence of prior infarction and interstitial fibrosis. Both findings also predispose to myocardial irritability and fatal (tachy)arrythmias.

Complications other than tachyarrhythmia and pump failure of myocardial infarctions can result in sudden cardiac death; the most common include the myocardial rupture syndromes including ventricular wall and papillary wall rupture. Typically, these insults occur approximately 1 week following a myocardial infarction, the point at which there is removal of necrotic myocytes by macrophages. Hemopericardium with ensuing cardiac tamponode can occur following ventricular free wall rupture; this scenario is rapidly fatal in most cases, causing decreased venous return to the heart with jugular venous distention.

In young patients, particularly athletes, hypertrophic cardiomyopathy is not an uncommon cause of sudden death. These patients can be asymptomatic prior to the sudden event or may have past episodes of palpitations or syncope. Typically, macroscopic heart evaluation shows cardiac hypertrophy with significant asymmetry of the subaortic septal region, which poses as an outflow obstruction. Microscopic sections from this region show variable degrees of myocyte disarray, fibrosis, myocyte hypertrophy, and small-vessel disease. The disease is due to an autosomal dominant mutation in the cardiac sarcomere apparatus, most commonly the myosin heavy chain, but many mutations have been described.

Arrhythmogenic right ventricular cardiomyopathy can present with sudden unexpected death. At autopsy, the right ventricle is thinned, with microscopic evaluation showing significant transmural infiltration by fibrofatty tissue.

Myocarditis due to a variety of causes including viral, bacterial, fungal, parasitic, autoimmune, and hypersensitivity can present as sudden death. The degree of activity, myonecrosis, and the location of the inflammation (i.e., conduction system involvement) are important in determining the significance of the infiltrates. Notably, eosinophils are seen quite commonly in hypersensitivity myocarditis and can be a clue to the underlying etiology.

Dilated cardiomyopathy is common, and has many etiologies that include idiopathic arteriosclerotic disease, hypertensive cardiovascular disease, alcoholism, elevated catecholamines, myocarditis, postpartum, doxorubicin, endocrinopathies, and genetic diseases. The heart typically is enlarged with a globoid configuration. The microscopic analysis shows interstitial fibrosis.

Rare infiltrative cardiac disease such as amyloidosis, hemochromatosis, primary or metastatic tumors, and sarcoidosis can result in sudden death. Microscopic evaluation in these cases is necessary, with particular attention to nodal tissues.

S. aureus is the most common organism found in infective endocarditis (IE). *S. aureus* endocarditis is associated with the highest mortality and risk of embolism. Increasing age, periannular abscess, heart failure, and absence of surgical therapy were identified in multivariate analysis as independent poor prognostic factors for increased mortality in patients with *S. aureus* IE. Other risk factors for the development of IE include congenital or acquired anatomic valve abnormalities such as stenosis. Impaired cardiac conductivity and function with heart failure not infrequently develops in patients with multiple septic myocardial emboli and infarcts due to IE, particularly with paravalvular abscess formation. According to a recent study of a cohort of 606 cases of infective endocarditis, 99 cases have embolization, of which 32 cases involve the central nervous system (CNS) with significantly higher mortality (65%) than those without CNS emboli.

Recently, genetic abnormalities have been found to underlie many of the intrinsic abnormalities of conducting systems including Wolff–Parkinson–White syndrome (WPW) and long Q-T syndrome. Sudden death in WPW is thought to occur as a result of an induction of ventricular tachycardia via an atrioventricular reentry pathway. Long-QT syndrome can also present

with sudden death. Investigations are ongoing around the association of sudden infant death syndrome with long-QT syndrome. Recent data is suggesting that a genetic basis for the arrhythmogenic disease with the identification of the long-QT genes.

Sudden death related to cardiac valve pathology other than endocarditis is relatively uncommon, as valve replacement surgery has become a standard therapy. Patients with aortic stenosis, especially when acutely symptomatic, can experience sudden cardiac death. Most cases of aortic stenosis are caused by either rheumatic heart disease or valve calcification, which can occur on trileaflet or congentially (uni)bicuspid valves. The mechanism for death in severe aortic stenosis (valve area <1cm2) appears to be through left ventricular hypertrophy and subsequent myocardial instability. In rare instances of severe aortic valve calcification, the deposits can erode the region and involve the conduction system. Mitral valve prolapse has long been associated with sudden cardiac death. The underlying etiology is not well understood, but seems to most frequently involve a severe valve deformity with a redundant, thickened, myxomatous mitral valve and ventricular arrhythmias such as ventricular fibrillation. On histologic sectioning, the mitral valve will show deposition of acid mucopolysacchrides.

Coronary artery anomalies are not uncommon but only certain anomalies result in ischemia such as anomalous origin of a coronary artery from the opposite sinus (ACAOS), anomalous left coronary artery from the pulmonary artery (ALCAPA), ostial atresia/stenosis, and coronary artery fistulas. Left-sided ACAOS can result in acute takeoff angles with an increased risk of sudden death during or shortly after exercise. Besides the acute angle take off, there maybe ridge like defect at the coronary ostea further decreasing blood flow in times of accelerated heart rates with increased oxygen demand. Myocardial tunneling is another anomalous coronary artery distribution that maybe associated with increased arrhythmogenic potential. There is debate about the significance of this anomaly. Some still believe it may be significant when a large portion of the epicardial coronary artery dips deeply into the left ventricle wall for a considerable distance, during times of rapid muscle contraction.

Vascular Disease

Causes of sudden death associated with vascular disease include those that lead to occlusion, narrowing, or rupture of a blood vessel. Atherosclerotic aneurysms can rupture, leading to rapid loss of consciousness and death. These aneurysms can occur just about anywhere, but are by far most common in the abdominal aorta. Most abdominal aortic aneurysms occur below the renal artery. The risk of rupture increases with the size of the aneurism, smoking history, and hypertension. The annual risk of rupture over 7 cm in size is 33%. Retroperitoneal rupture is typically associated with hematoma formation, whereas rupture into the abdominal cavity can be rapidly fatal, with hemoperitoneum and shock. Patients who have a ruptured aortic aneurysm and reach the hospital have a 50% mortality rate, with the overall mortality rate greater than 85%.

Aortic dissection is characterized by an intimal tear followed by a dissection of blood within the wall of the aorta, most commonly the tunica media. Rupture of this dissecting aortic hematoma may lead to hemothoraces, hemopericardium, or fatal arrhythmia. Aortic dissection is a major cause of sudden death, mostly in patients over 50 years of age with the underlying risk factor being essential hypertension. However, pregnant women and patients with connective-tissue diseases such as Marfan's syndrome also make up a significant affected patient population. Aortic dissection can also occur following accidental or iatrogenic trauma to the aortic intima. In younger patients and those with connective tissue disease, microscopy may reveal cystic medial degeneration of the aortic media.

Most spontaneous subarachnoid hemorrhages (SAH) (90%) are caused by ruptured intracranial saccular (berry) aneurysms. SAH occurs at a peak age of 55–60 years. Rupture of an intracranial aneurysm is believed to account for 0.4 to 0.6% of all deaths. SAH is associated with a greater than 50% mortality rate. Some hospital-based studies suggest that approximately 10% of patients with aneurismal SAH die prior to reaching the hospital, 25% die within 24 hours of SAH onset, and about 45% die within 30 days. It is not unusual to perform forensic autopsies where death was almost instantaneous and outside of a hospital. The mechanism of death in such cases is cardiac arrhythmia, which is described in greater depth later. Most intracranial aneurysms (approximately 85%) are located in the anterior circulation, predominately on the circle of Willis. Risk factors for both SAH and intracranial aneurysms are similar and include hypertension, cigarette smoking, and alcohol consumption. Atherosclerosis is an independent risk factor for the development of intracranial aneurysms. The natural history of subarachnoid hemorrhage shows that rupture often occurs when they reach a size over 7 mm. Rupture of an aneurysm releases blood directly into the cerebrospinal fluid (CSF) under arterial pressure. The blood spreads quickly within the CSF, rapidly increasing intracranial pressure. A major symptom associated with SAH includes patients describing the worst headache of one's life. Increased intracranial

pressure is associated with the Cushing's triad (hypertension, bradycardia, and abnormal respiration). SAH is associated with cerebral edema and subsequent herniation. Tonsillar and central transtentorial herniation is associated with compression of cardiovascular and respiratory centers in the medulla and as such is rapidly fatal. Other less common causes of subarachnoid hemorrhages include angiomas and arteriovenous malformations. Ruptured berry aneurysms are the most common natural cause of SAH, whereas trauma is the most common overall cause. Ruptured berry aneurysms are a leading cause of sudden death in women during sexual activity, whereas for men it is heart disease.

Cerebrovascular accidents (episodes), which include ischemic or intracerebral hemorrhage, can lead to sudden death. I recommend not using the term "accident" because there is nothing accidental about this process and its use often adds confusion in forensic proceedings. The terms "stroke" or "event" as an alternative is less confusing to nonmedical personnel. Thromboembolic events can underlie ischemic cerebral events and are associated with heart disease, valvular pathology, or carotid artery disease. Hypertension is a major risk factor for intraparenchymal hemorrhage and may lead to increased intracranial pressure, herniation, and death.

The greatest percentage of thrombi resulting in pulmonary embolism is thought to originate in the deep veins of the lower extremities. Deep-venous thrombosis can also occur in the pelvis or other locations. Fragments of blood clot may break off and embolize to the pulmonary arteries. An occlusion greater than 50 to 75% of the large pulmonary vessels results in a rise of the pulmonary artery pressure greater than 40 mmHg. This rise of pulmonary arterial pressure is accompanied by an increase in right ventricular diastolic, right atrial, and systemic venous pressures, with a decrease in cardiac output resulting in sudden death. Patients who have multiple small pulmonary emboli or in situ thrombus formation over time may present with increasing shortness of breath and right-sided heart failure. Because the lungs have dual circulation, infarctions are less common unless there is significant underlying natural disease with decreased cardiac function.

Various types of vasculitis or blood vessel inflammation can cause wall thickening, thrombosis, dissection, and rupture. Mesenteric thrombosis may be associated with polyarteritis nodosum and other autoimmune conditions.

Other Causes of Sudden Death

Rare undiagnosed brain tumors may present with sudden death. Infiltration or edema formation into the key respiratory/cardiac centers of the brain with possible herniation are two mechanisms. Early or late stage malignancies may sometimes metastasize to the heart and interfere with the conducting system, causing a fatal arrhythmia. Other causes of sudden death in patients with malignancies include cardiovascular events such as acute myocardial infarction, therapeutic complications (i.e., anaphylaxis), and metabolic derangements. Rare causes of sudden death in patients with tumors or malignancies include erosion of large vessels or visci with fatal hemorrhage. A colloid cyst of the third ventricle may lead to sudden death and is usually associated with premortem postural headaches. In certain positions, the cyst will act like a ball valve and suddenly block the flow of cerebral spinal fluid, resulting in acute obstructive hydrocephalus. One may be fine standing but develop symptoms when he or she lies down. This buildup of cerebral spinal fluid pressure can cause a fatal arrhythmia. Bacterial pneumonia with the combination of hypoxia and bacterial toxins and end products can cause sudden death.

Status asthmaticus and sudden asphyxic asthma are life-threatening forms of asthma. These cases are not unusual in a forensic setting. Status asthmaticus is defined as an acute attack of respiratory failure due to airway inflammation, edema, and mucous plugging. Sudden asphyxic asthma is due to brochospasm rather than airway inflammation. Viral infections and other causes have been implicated as precipitants of these potentially fatal complications. Grossly in both cases, the lungs may appear so much hyperaerated that at times rib indentations will show. Thick mucus plugs may obstruct the upper airways. Sudden death in asthmatic patients is thought to be secondary to fatal arrhythmia, occurring as a consequence of global hypoxia and right heart failure.

There is a condition known as sudden unexpected death in epilepsy (SUDEP). The mechanism is unclear but this phenomenon occurs in up to 18% of patients with epilepsy, presumably in those with subtherapeutic levels of anticonvulsants. Autonomic dysfunction has been proposed as a mechanism. Other mechanisms for death in patients with epilepsy include accidental/traumatic incidents such as drowning and choking that occur during a seizure. Hypoxia as a result of respiratory compromise can result in ischemic cardiac events. This may be part of the final mechanism of death in epileptic patients experiencing status epilepticus. Another interesting point to remember is that there is often very rapid rigor mortis formation in deaths directly following static epilepticus due to substantial adenosine triphosphate (ATP) depletion associated with prolonged muscle contractions from prolonged convulsions. Usually there are few pathologic findings that explain the sudden death

in epileptic patients. Autopsy findings may include bite marks to the tongue with hemorrhage or a voided urinary bladder. There may be no finding at all. These are nonspecific findings and seizure activity may also occur prior to many other nonepilepsy-related deaths.

Fatal anaphylaxis can result from exposure to insect stings, foods, latex, drugs, chemicals, and exercise. This mast cell-mediated systemic reaction results in severe angioedema and bronchoconstriction of the upper respiratory tract along with hypotension resulting in respiratory and circulatory collapse. Death caused by anaphylaxis is primarily due to airway obstruction when laryngeal edema fills the rich lymphatic supply of the epiglottic folds. Increased mast cell tryptase levels in the patient's serum can be detected that peaks approximately 15 to 60 minutes after the onset of anaphylaxis and then declines with a half-life of about 2 hours.

The mortality for gastrointestinal bleeding (GI) in the case of ruptured esophageal varices most commonly encountered in patients with portal hypertension is high. Intra-aortic balloon pumps are lifesaving procedures but only if the patient presents in a timely fashion. Other causes of fatal upper gastrointestinal bleeding include stomach and duodenal ulcers; in this scenario the source is arterial as opposed to venous in esophageal varices. Fatal lower gastrointestinal bleeding can be seen in patients with angiodysplasia, diverticulitis, and carcinoma; however, this scenario is less common than upper GI bleeding.

Mostly complications of morbid obesity are thought to underlie the association with sudden death. Hypertension, left ventricular hypertrophy, and cardiomegaly are all independent risk factors for sudden death. Postural asphyxia may occur as a result of obesity. Morbid obesity is a reasonable cause of death by itself due to stress on the heart. An individual who is three times the expected body weight has roughly three times the vasculature with three times the blood volume to pump. In times of other stress, this can have devastating consequences on the heart, with death by arrhythmia.

Waterhouse–Friderichsen syndrome was first described as occurring in patients with meningococcemia and is characterized by severe bacteremia and bilateral adrenal hemorrhages. This combination results in overwhelming shock and, if untreated, sudden death can occur. Organisms other than *N. meningitis,* such as *E.Coli,* have been reported to produce this syndrome.

Multiorgan failure and death can be seen in sickle cell anemia patients with an acute crisis. Precipitants may include infection, dehydration, hypoxia, physical excretion, vaso-occlusion, or fat embolus following bone infarction. This acute hemolytic sickling crisis results in severe hypoxemia with end organ failure. Patients with sickle cell anemia have auto-infracted spleens and are much more susceptible to encapsulated organisms such as pneumococcal bacteria. Even patients with sickle cell trait may develop crisis in times of great physical exertion with dehydration, such as basic training in the army or boot camp.

Natural disease processes may weaken the body, making fatal traumatic injury more likely. Osteoporosis from aging, Cushing syndrome, steroid use, and other natural disease processes will make bones more fragile and allow fractures to occur more easily.

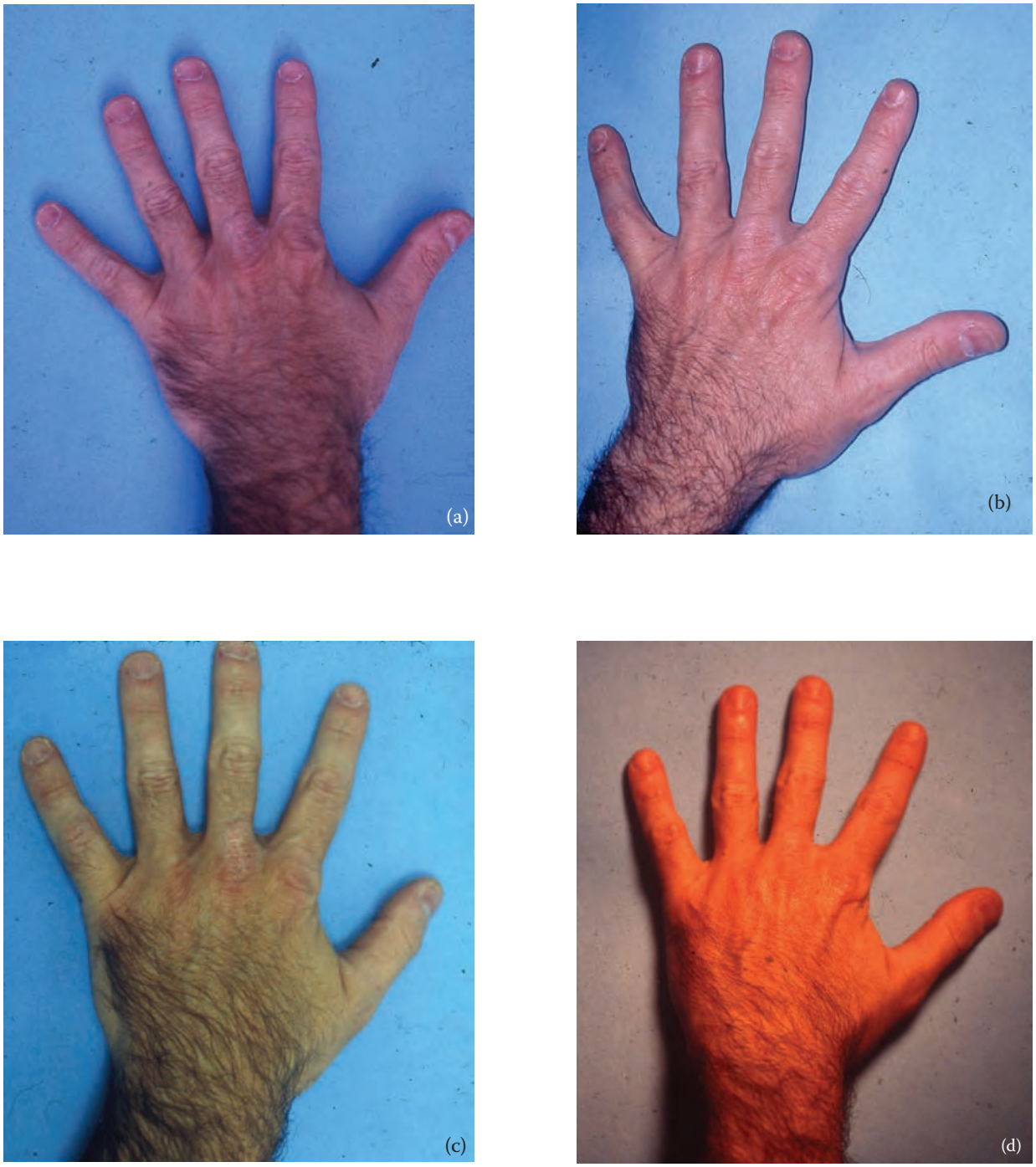

When examining photographic evidence it may be important to have knowledge of the type of camera, film, and lighting used when documenting different disease states. These four photos demonstrate different types of lighting causing variation in picture color. Figure (a) was taken in overcast sunlight, (b) with camera flash, (c) under fluorescent light, and (d) with a Tungsten filament regular light bulb. If your opinion is that a photographic image is not interpretable, it is perfectly acceptable to say that you cannot render an opinion based on this two-dimensional image.

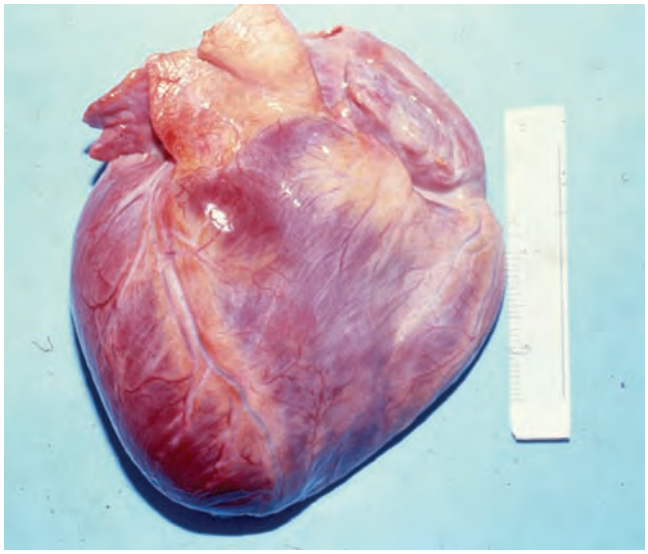

Normal fresh heart.

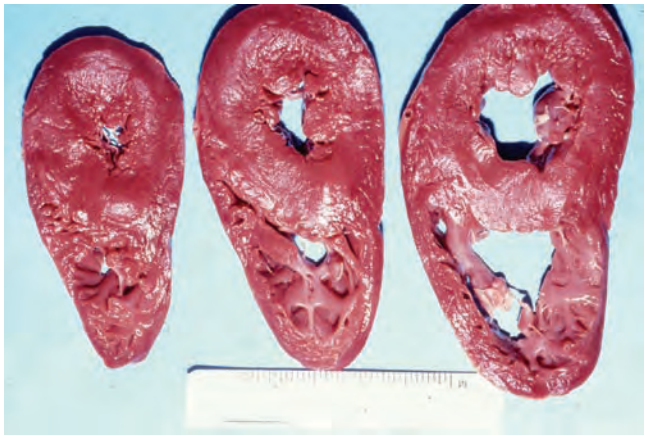

Sections of a normal fresh heart showing right and left ventricle.

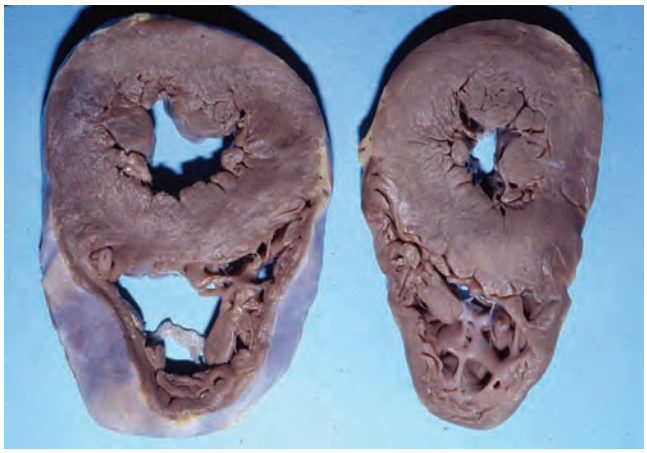

Sections of normal right and left ventricle after formaldehyde fixation.

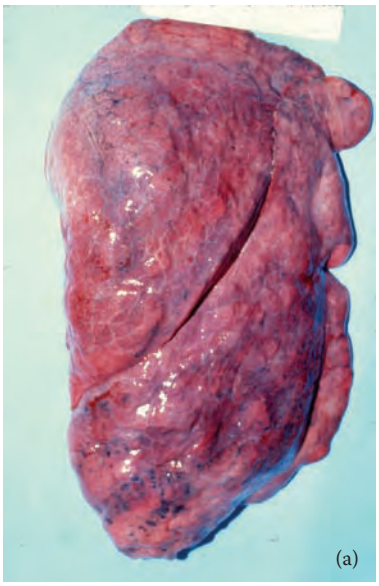

Normal fresh left lung (a) demonstrating two lobes. Right lungs (b) have three lobes.

Normal lung fixed in formaldehyde.

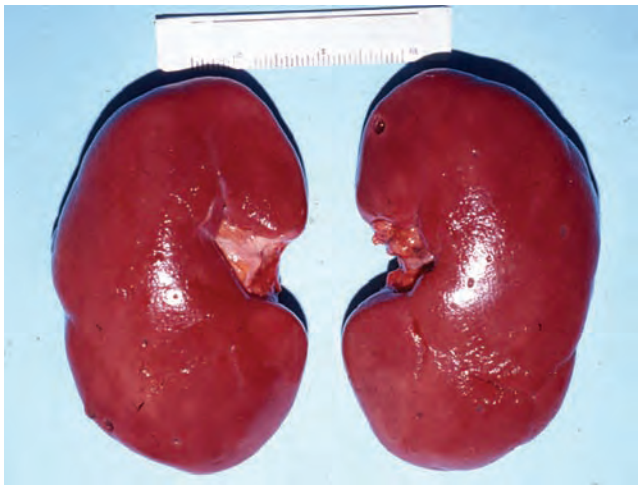

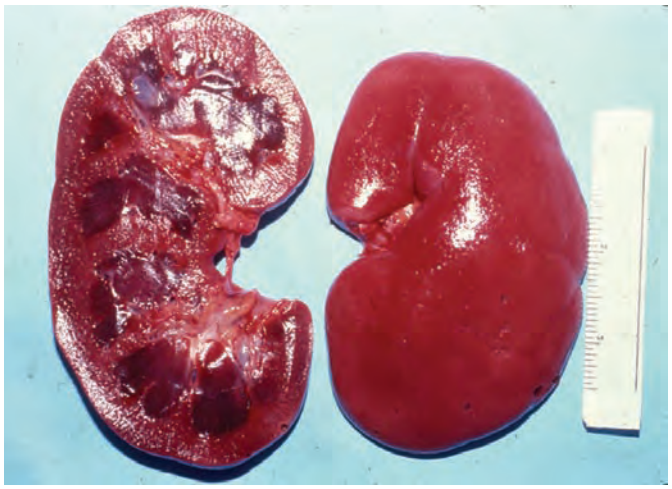

Normal fresh kidneys.

Non-fixed fresh kidneys. Note the pale discoloration resulting from fatal blood loss prior to death due to a gunshot wound.

Normal kidney fixed in formaldehyde.

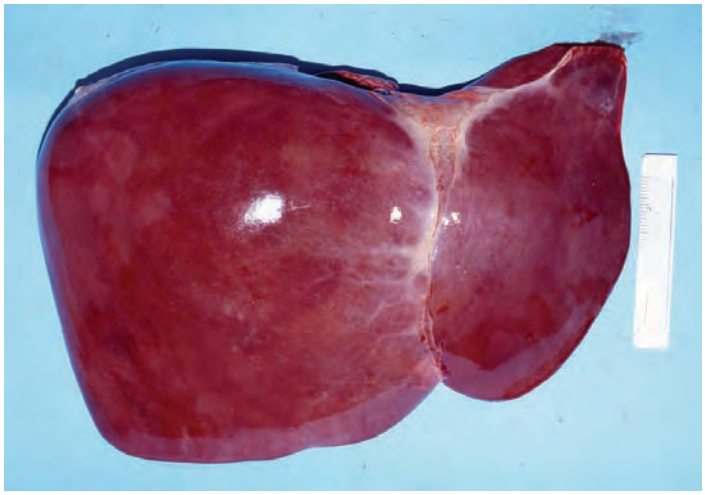

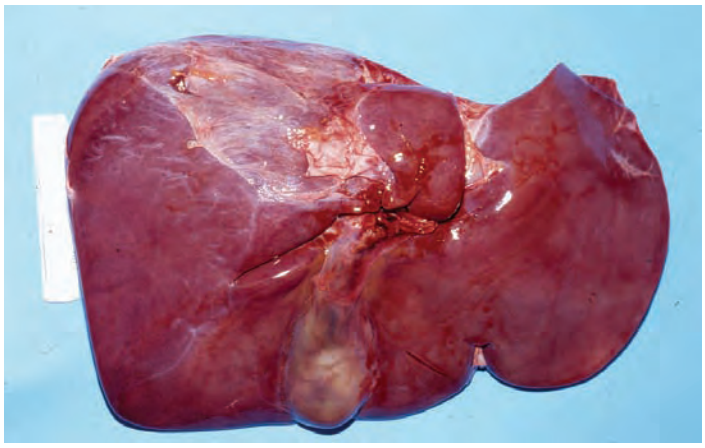

Normal fresh liver.

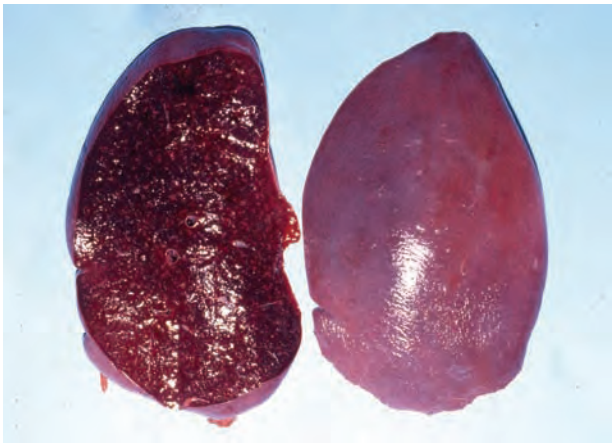

Normal fresh spleen.

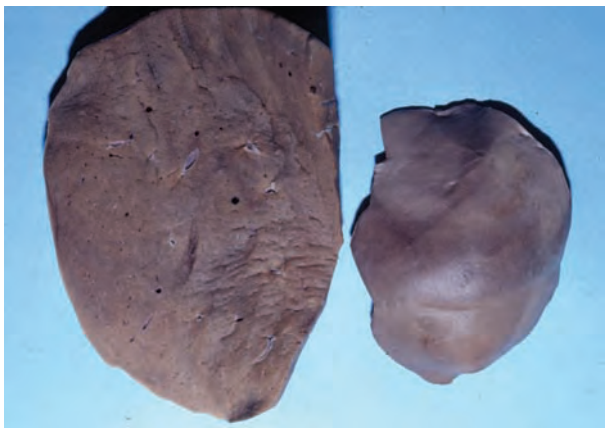

Normal liver fixed in formaldehyde.

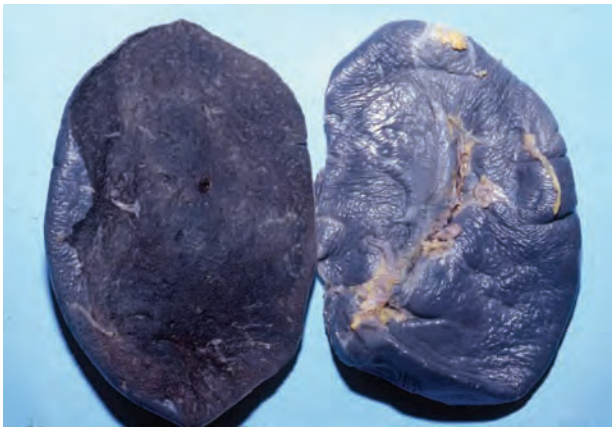

Normal spleen fixed in formaldehyde.

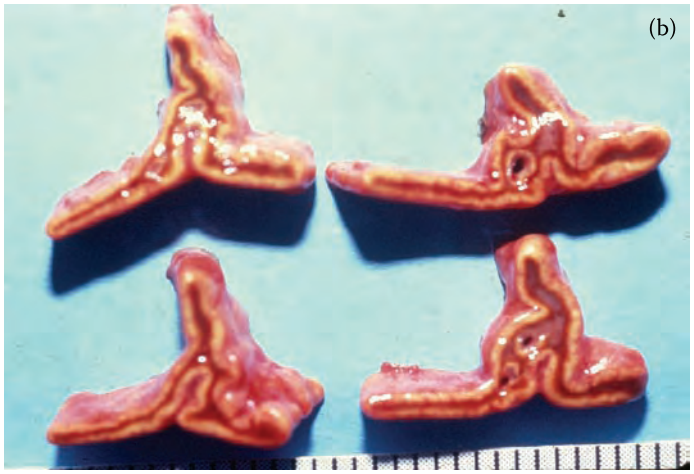

Normal fresh adrenal gland intact (a) and sectioned (b).

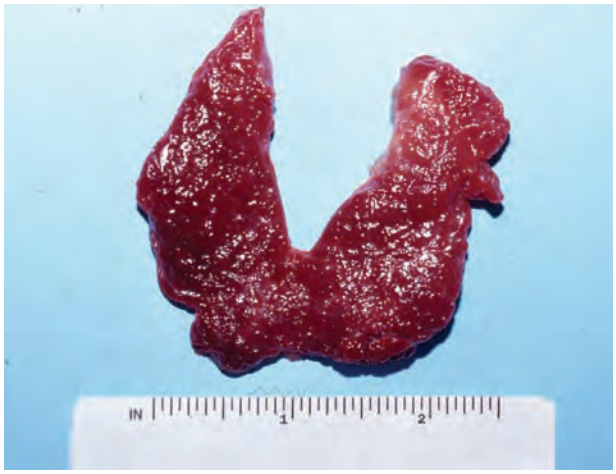

Normal fresh thyroid gland.

Normal adrenal gland section fixed in formaldehyde.

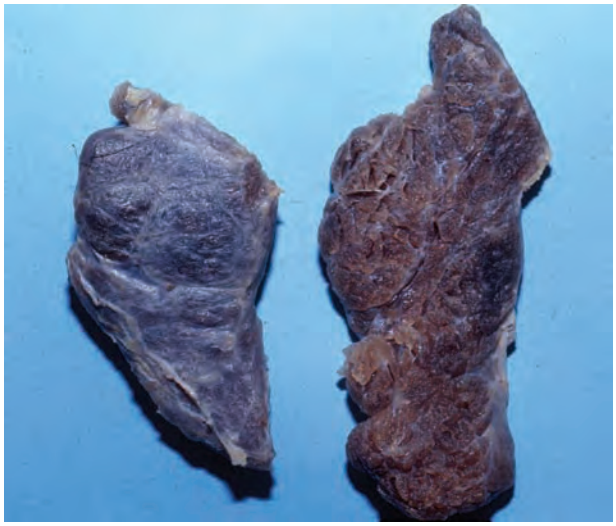

Normal thyroid fixed in formaldehyde.

Normal fresh testes.

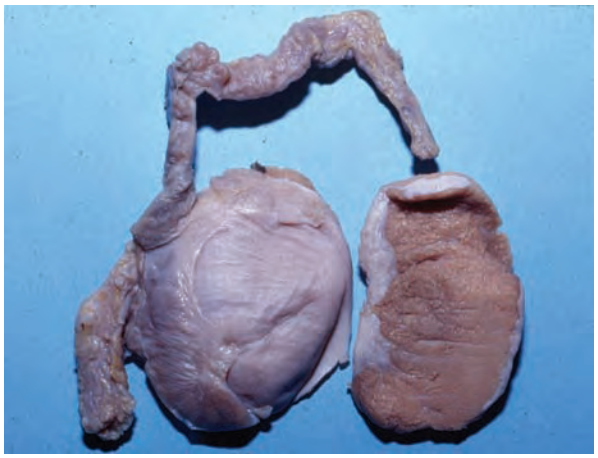

Normal testes fixed in formaldehyde.

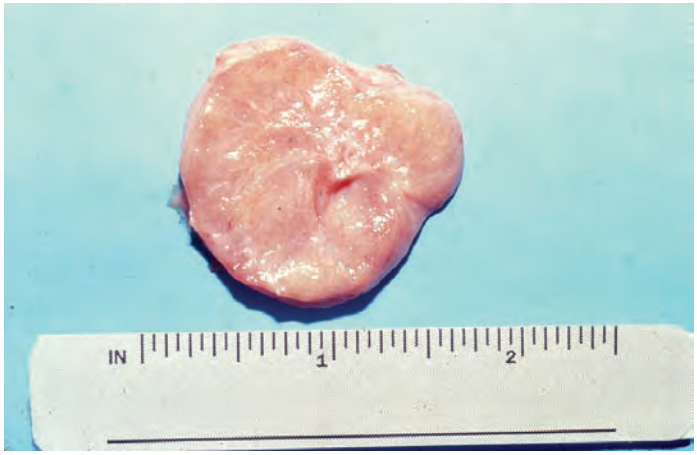

Normal fresh prostate.

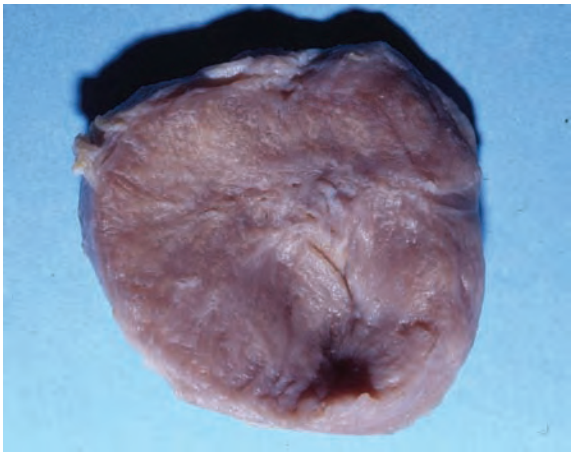

Normal prostate fixed in formaldehyde.

Normal fresh pancreas.

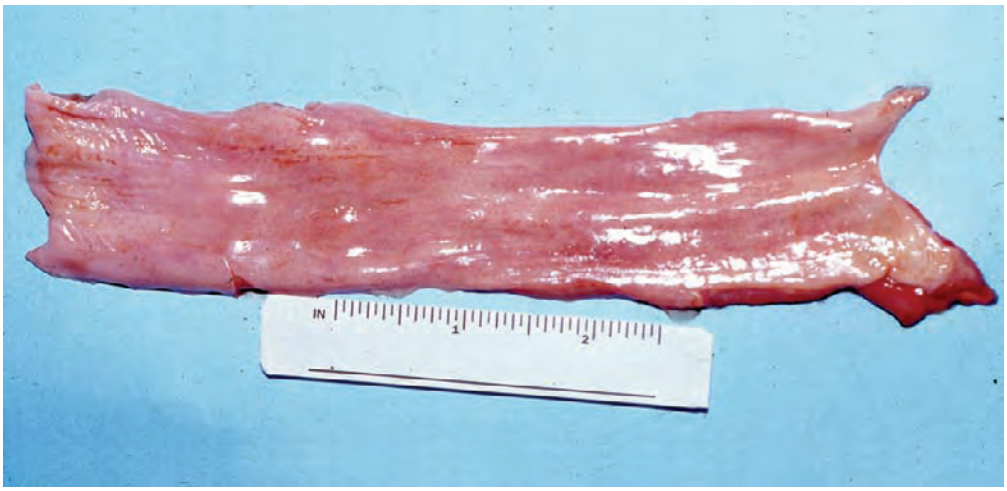

Normal fresh esophagus.

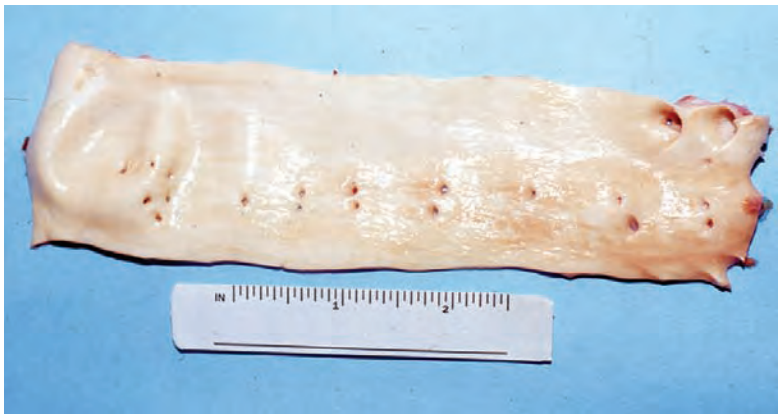

Aorta with slight atherosclerosis. Note the fatty streaks on the intimal surface.

Normal fresh bladder.

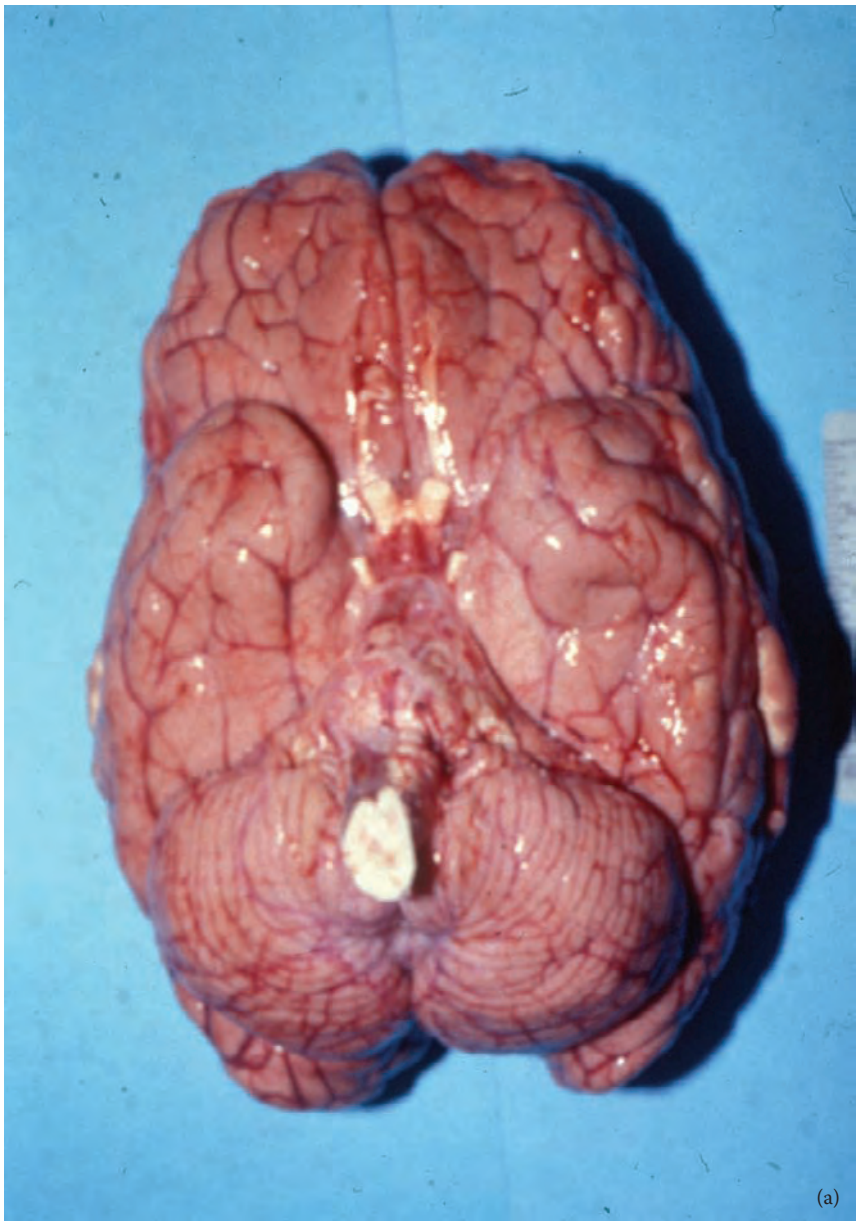

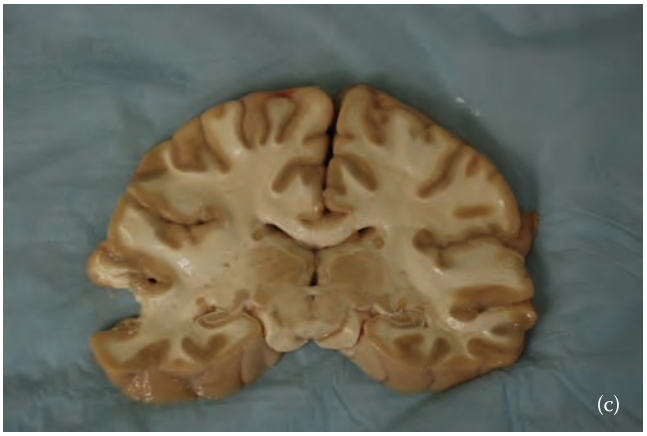

Figure (a) demonstrates congestion of fresh brain. Note the slight pink color. Figures (b) and (c) demonstrate fresh brain in an individual who exsanguinated from a ruptured aortic aneurysm. Note the pale discoloration due to blood loss.

This figure depicts a decedent with an endotracheal tube in the right side of his mouth. It also demonstrates the "purple head sign," a common finding in victims of sudden death, particularly cardiac death. The explanation for this finding is not known in entirety but is attributed to uncontrolled terminal sympathetic nervous system discharges, which open free capillary sphincters and produce a gush of capillary blood.

Morbid obesity. This is a legitimate cause of death and can stand alone on a death certificate.

Petechiae associated with heart disease and resuscitation.

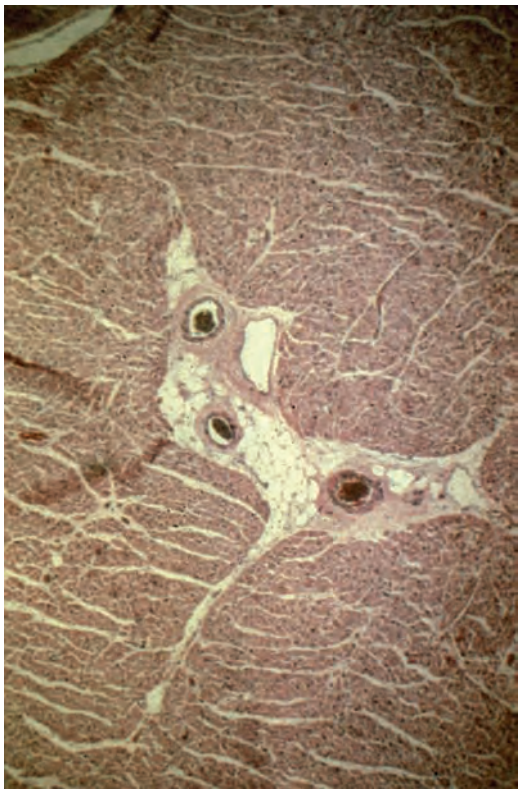

Low-power magnified view of small coronary arteries with thrombosis.

Epicardial vessel with complete occlusion by organizing thrombus. Note the adjacent epicardial hemorrhage.

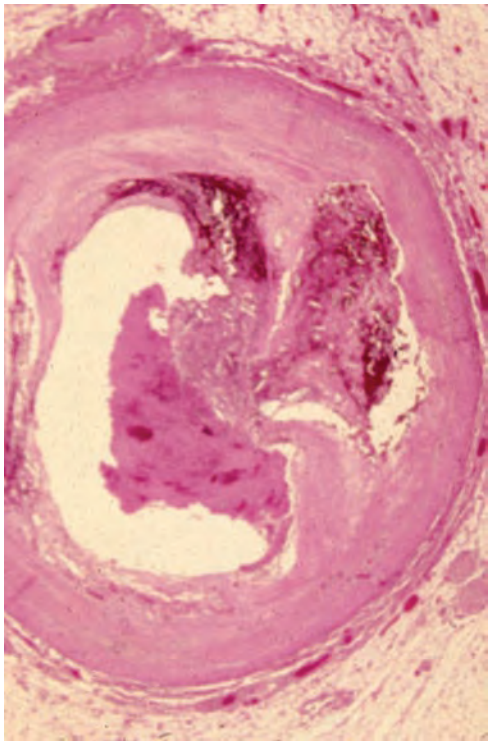

Coronary artery with a ruptured atherosclerotic plaque and thrombus formation.

Marked coronary atherosclerosis in an epicardial vessel.

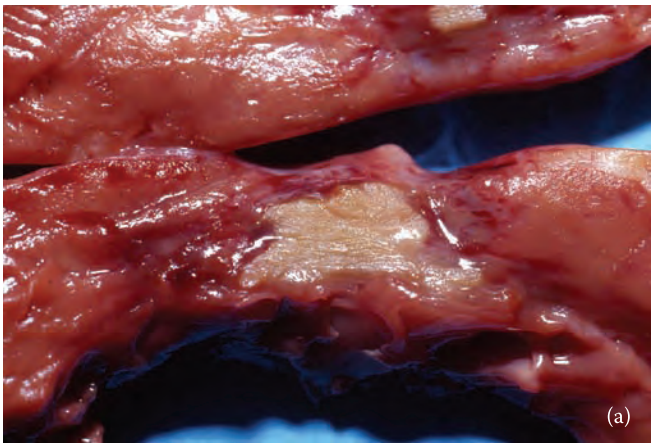

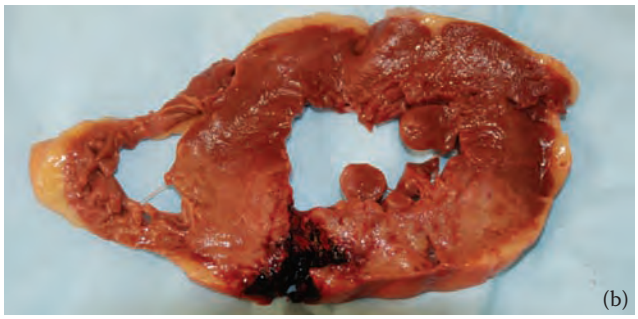

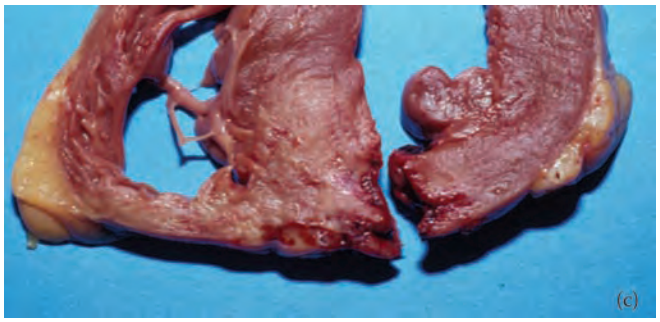

An acute myocardial infarction. Note the yellow discoloration due to necrosis. This infarction is approximately 3 to 6 days old. In Figure (b) and (c), there is a transmural acute myocardial infarction with rupture.

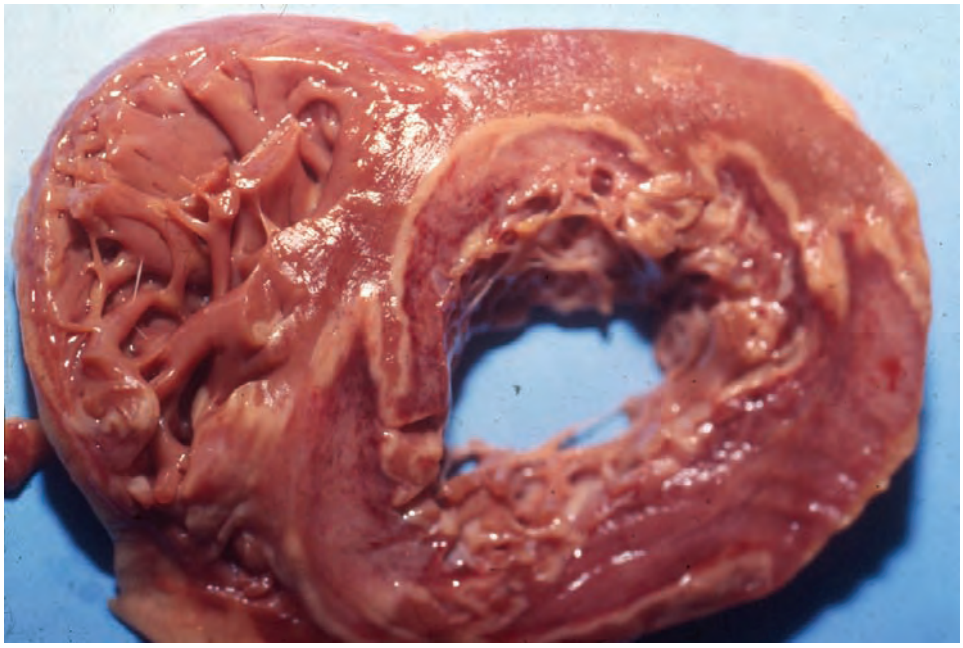

Acute myocardial infarction.

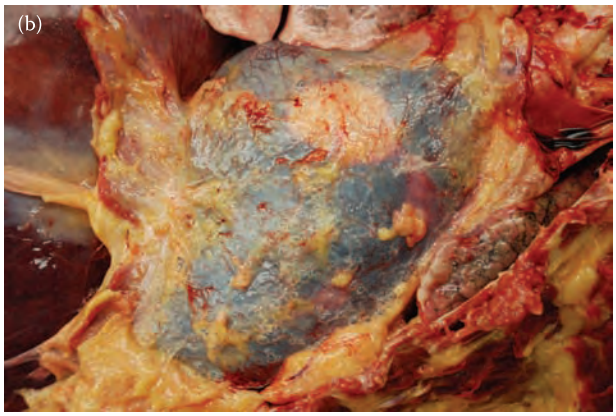

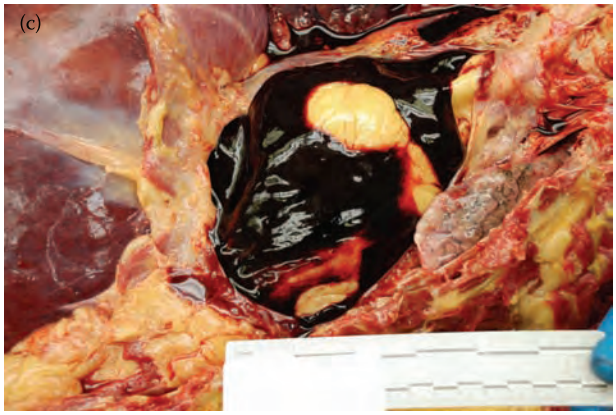

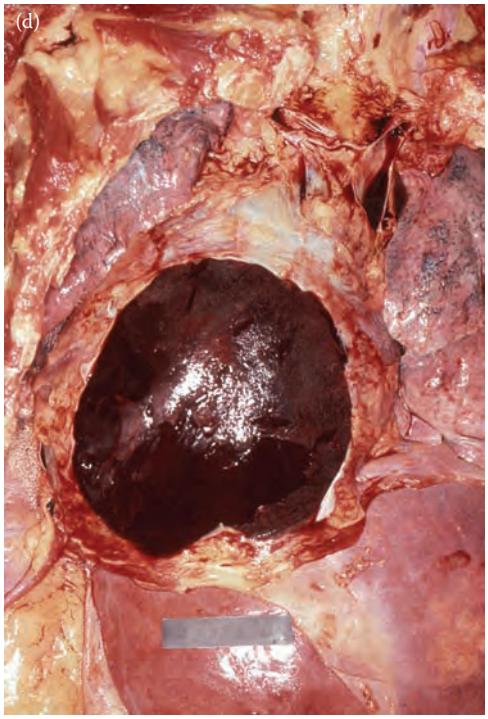

View of the thoracic cavity looking downward at the heart during autopsy (a) and (b). Note the purple discoloration of the pericardial sac due to underlying accumulation of blood. Note the two different examples with large blood clot encasing the heart after the pericardial sac was removed (c) and (d). This demonstrates a cardiac tamponade following an acute ruptured myocardial infarction.

Two examples of hearts demonstrating acute ruptured myocardial infarction. Note the adjacent hemorrhage and perforation site. This resulted in cardiac tamponade and sudden death.

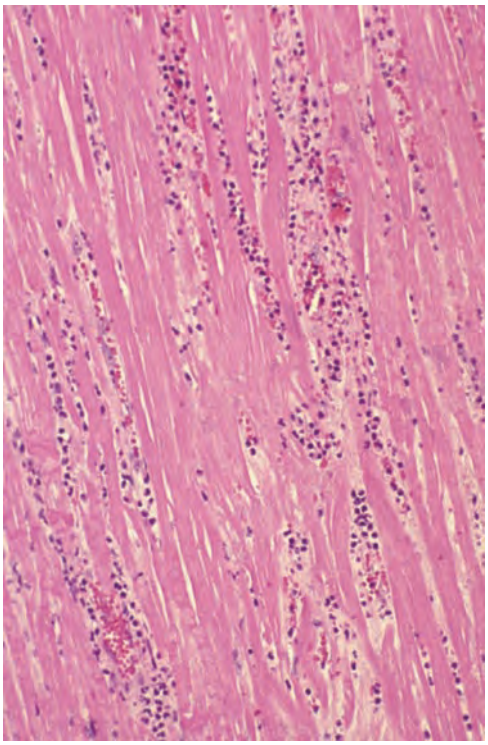

This microscopic section of heart muscle reveals an acute myocardial infarction with hemorrhage, polymorphonuclear cell infiltrates, and myocardial necrosis.

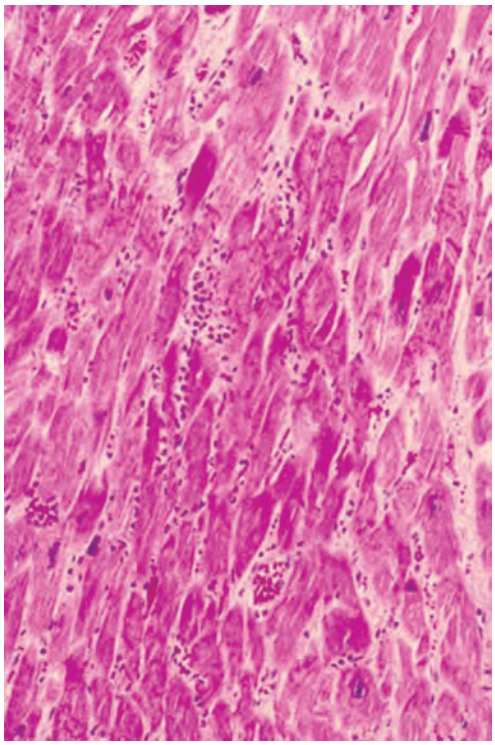

This acute myocardial infarction demonstrates coagulative necrosis with contraction bands. There are also polymorphonuclear cell infiltrates.

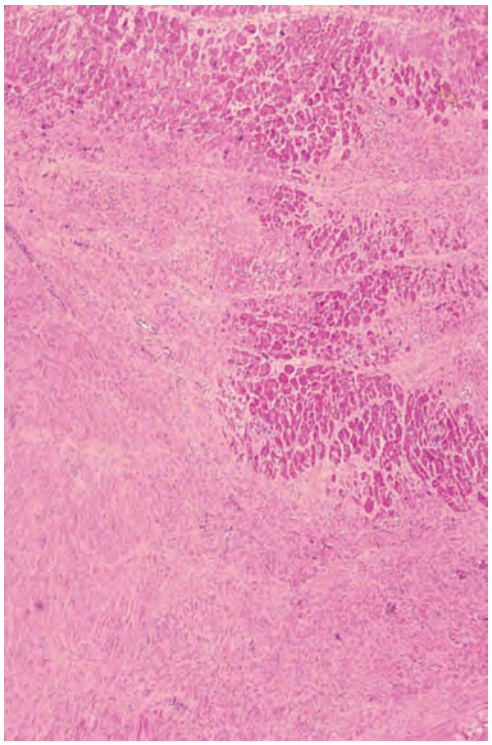

This histopathologic section of myocardium reveals an old infarction with fibrosis.

This low-power view of heart muscle reveals patchy areas of fibrosis consisting of healed hypoperfusion infarctions secondary to a remote trauma with severe shock. Also note the perivascular fibrosis.

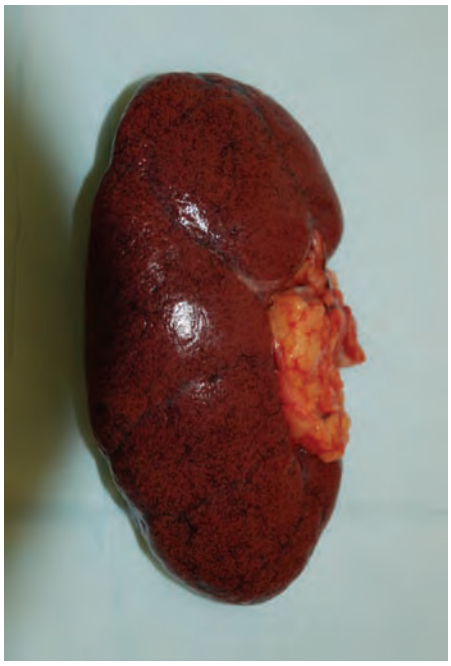

Early to moderate nephro-arteriolosclerosis.

Moderate to marked nephro-arteriolosclerosis.

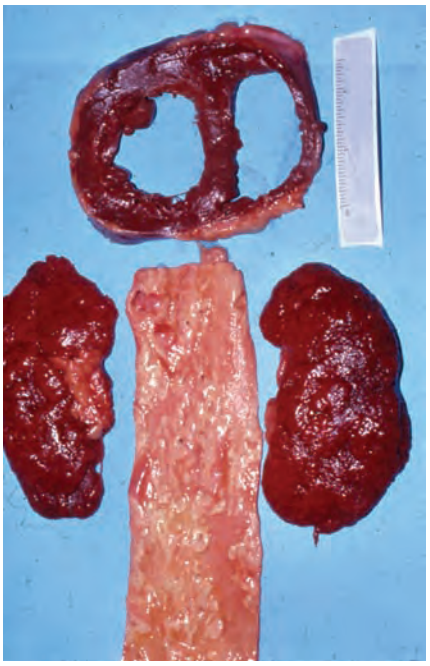

Nephro-arteriolosclerosis associated with hypertensive cardiovascular disease. Note the cardiac hypertrophy and biventricular dilatation in this failing heart. There is also moderate atherosclerosis of the aorta. Note the markedly granular subcapsular kidney surfaces and cortical scarring also associated with this process.

Arteriovenous hemodialysis grafts for treatment of chronic renal failure due to hypertensive cardiovascular disease.

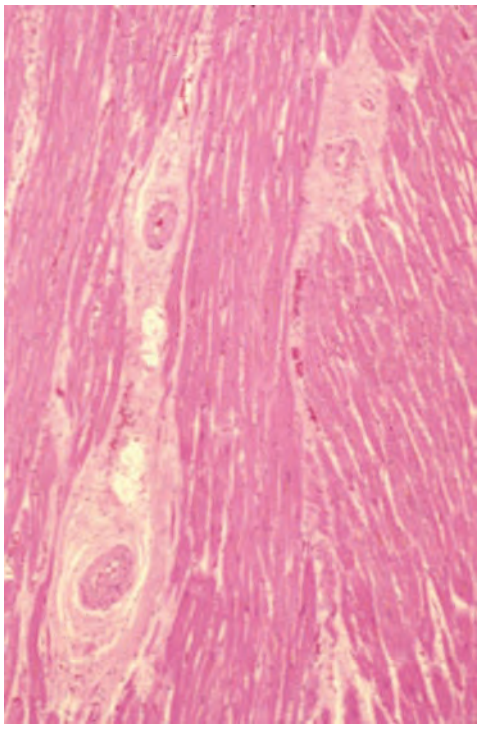

This microscopic image of myocardium reveals significant perivascular fibrosis and small vessel disease in a patient who had a long history of hypertensive cardiovascular disease and diabetes mellitus.

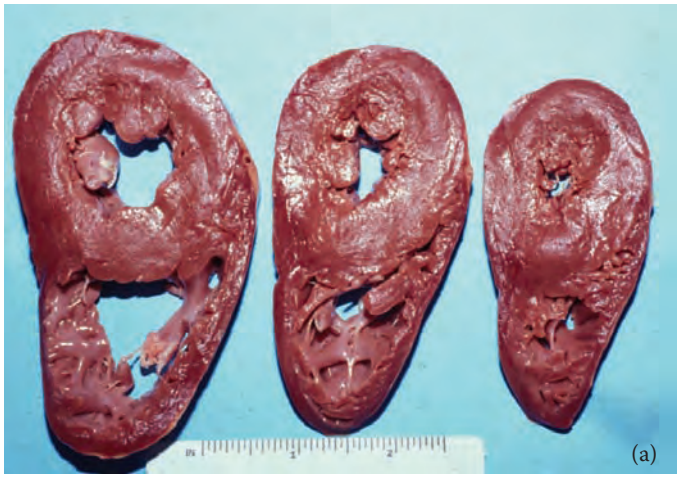

Normal cross sections of heart (a). Compare this image to the ones below. Figure (b) shows a markedly enlarged heart due to hypertensive cardiovascular disease. This heart is diffusely enlarged and shows cardiac hypertrophy. Figure (c) is a markedly hypertrophied heart with extreme concentric left ventricle hypertrophy. This individual had severe hypertension.

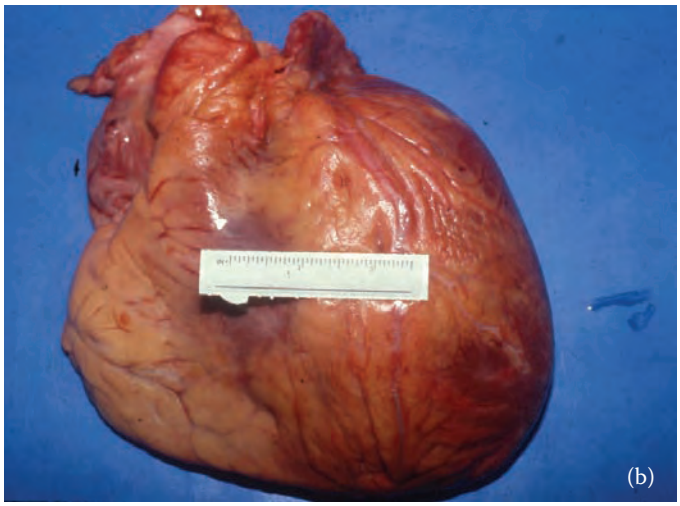

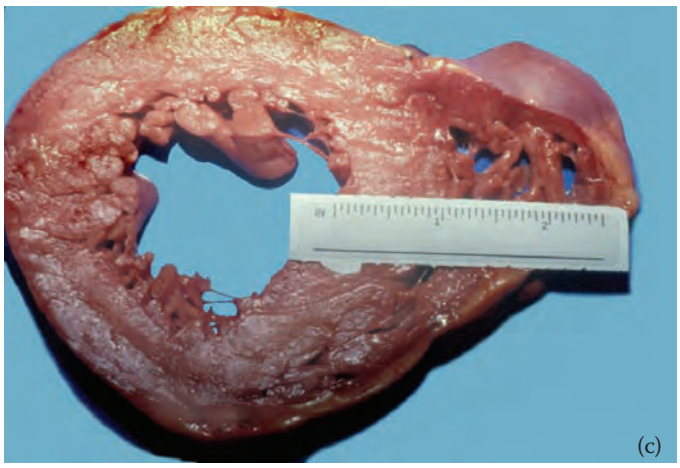

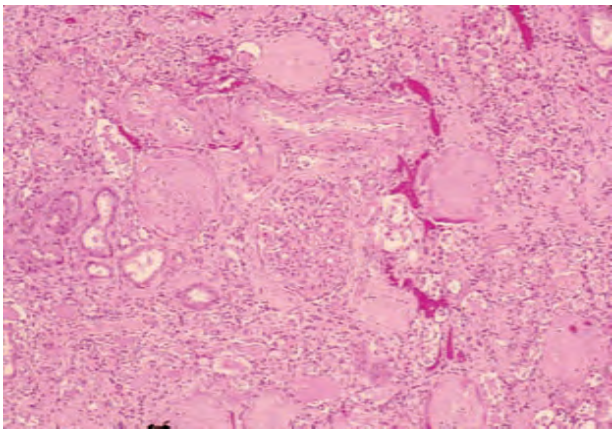

This is a microscopic section of kidney showing fibrosis, arteriolosclerosis, and nephrosclerosis due to hypertensive cardiovascular disease.

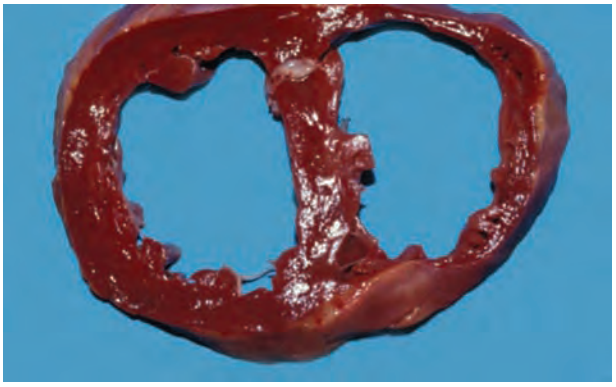

The right and left ventricles show marked dilatation. This individual died of a peripartum cardiomyopathy. A dilated cardiomyopathy or end stage hypertensive cardiovascular disease with cardiac failure will appear the same grossly.

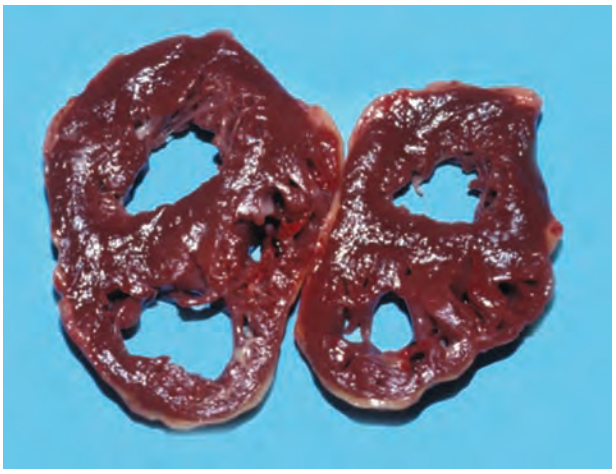

The sections of these ventricles reveal marked right ventricle hypertrophy. This individual had end-stage primary pulmonary fibrosis with cor pulmonale and cardiac failure.

The intimal lining of an aorta with marked atherosclerosis in a decedent with a longstanding history of smoking, diabetes, high cholesterol, and high blood pressure.

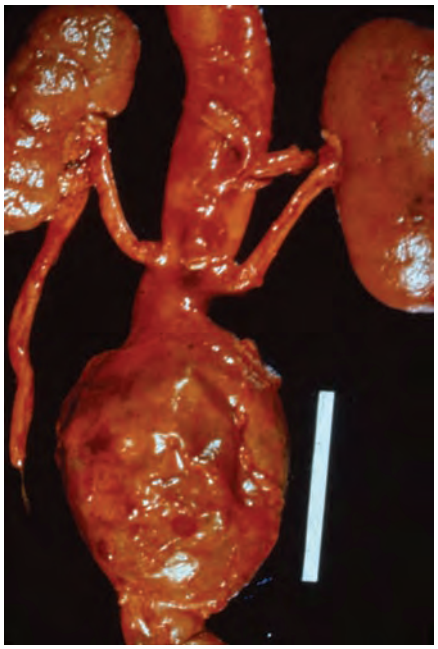

Abdominal aortic atherosclerotic aneurysm shown in its typical location inferior to the renal arteries and above the iliac bifurcation.

An abdominal atherosclerotic aneurysm with a rupture at its anterior aspect and visible thrombosis.

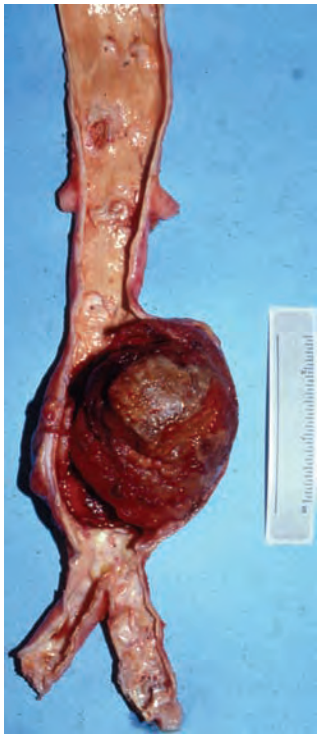

This abdominal aortic aneurysm that has been opened to remove half of the vessel wall and show the underlying intimal surface with moderate atherosclerosis except for the region of the aneurysm, which has marked atherosclerosis and a large overlying thrombus.

This abdominal aortic aneurysm has been cross sectioned to show the partial obstruction of the aneurysm by organizing thrombosis. The lumen is demonstrated by fresh red blood clot at its surface and the thrombus is demonstrated by the light grey regions adjacent to the right and left wall.

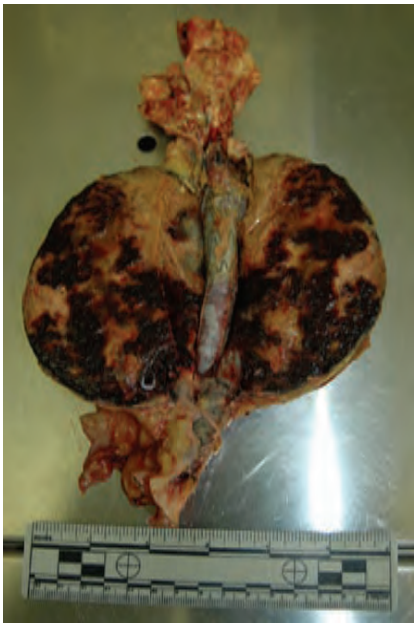

Abdominal aortic aneurysm with vascular graft repair.

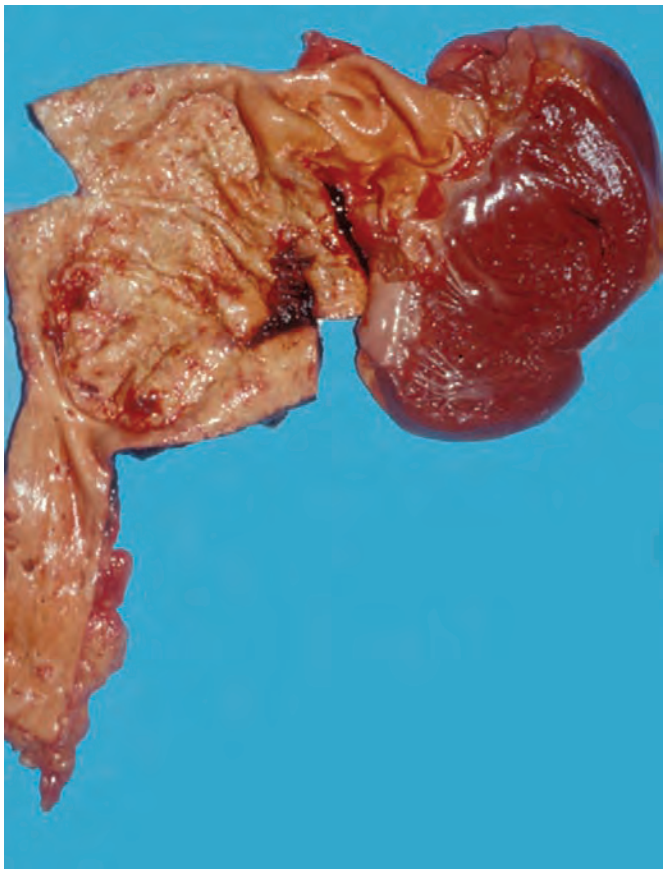

Thoracic aortic atherosclerotic aneurysm. This is not the typical location for such an aneurysm. Also note the aneurysm begins distal to the root of the aorta.

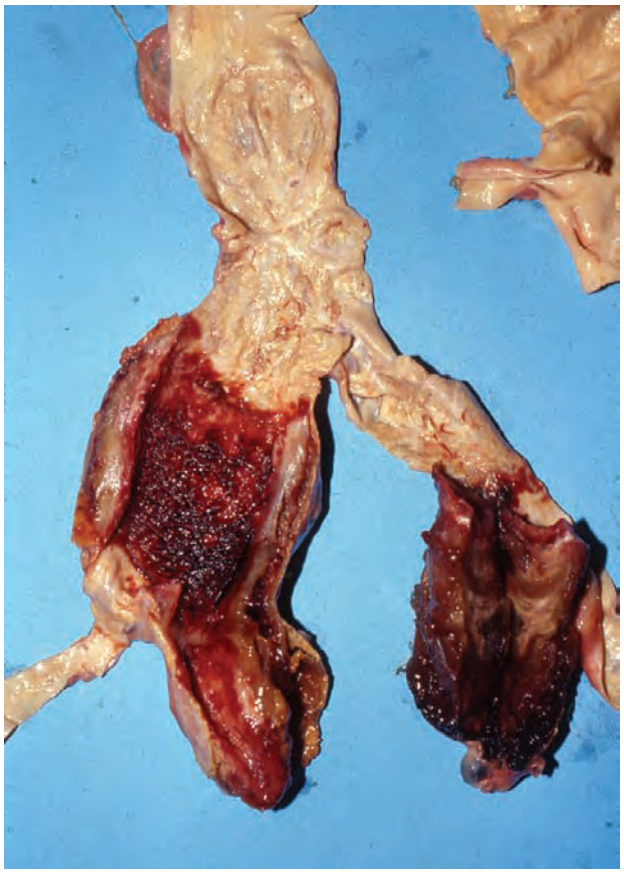

This demonstrates a decedent who had multiple atherosclerotic aneurysms including both iliac arteries.

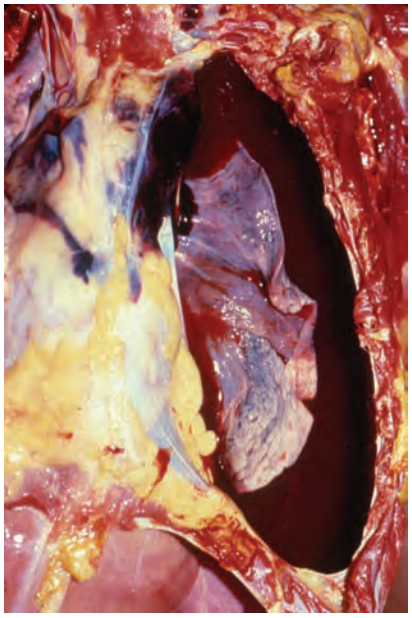

Hemothorax from a ruptured thoracic atherosclerotic aneurysm.

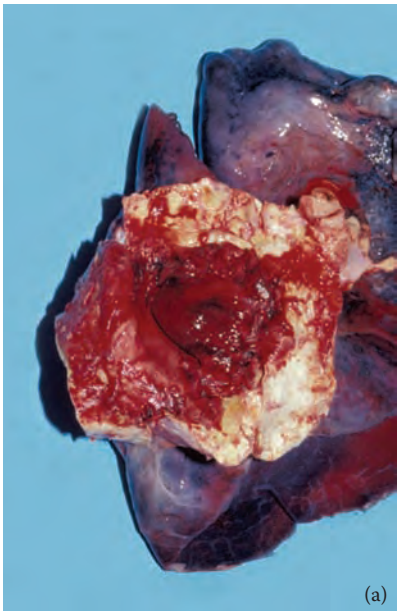

Thoracic atherosclerotic aortic aneurysm adherent to lung with rupture into the lung parenchyma causing massive hemoptysis. Note that Figure (b) demonstrates the rupture site with adherent blood clot removed.

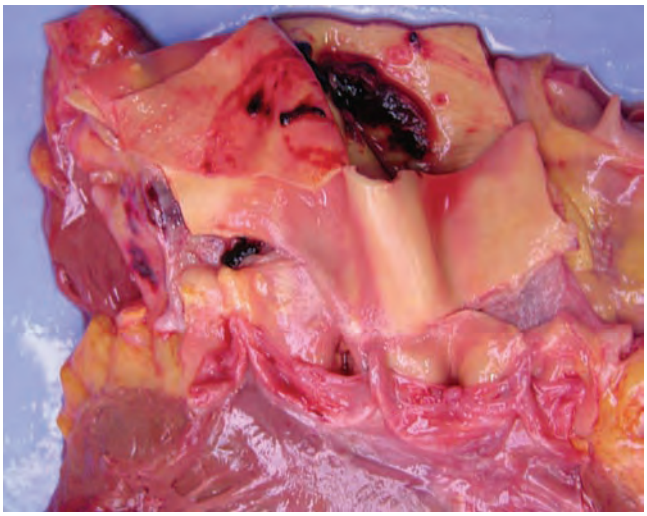

Intimal tear of the ascending aorta with dissection. Note blood tracking through the separated media.

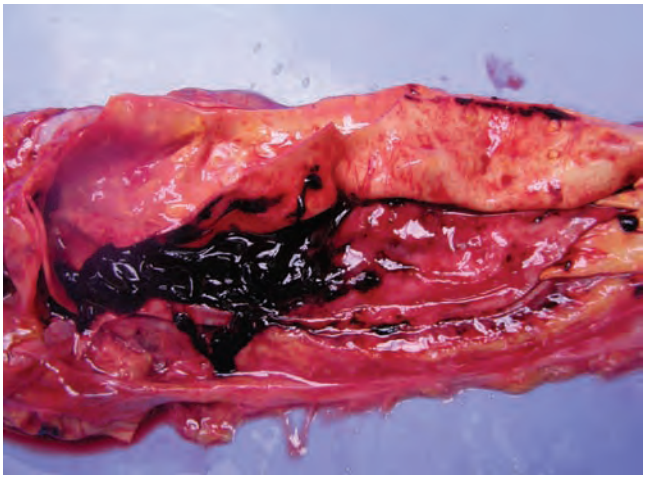

Aortic dissection with exposed separated media with blood clot in a person with Marfan's syndrome.

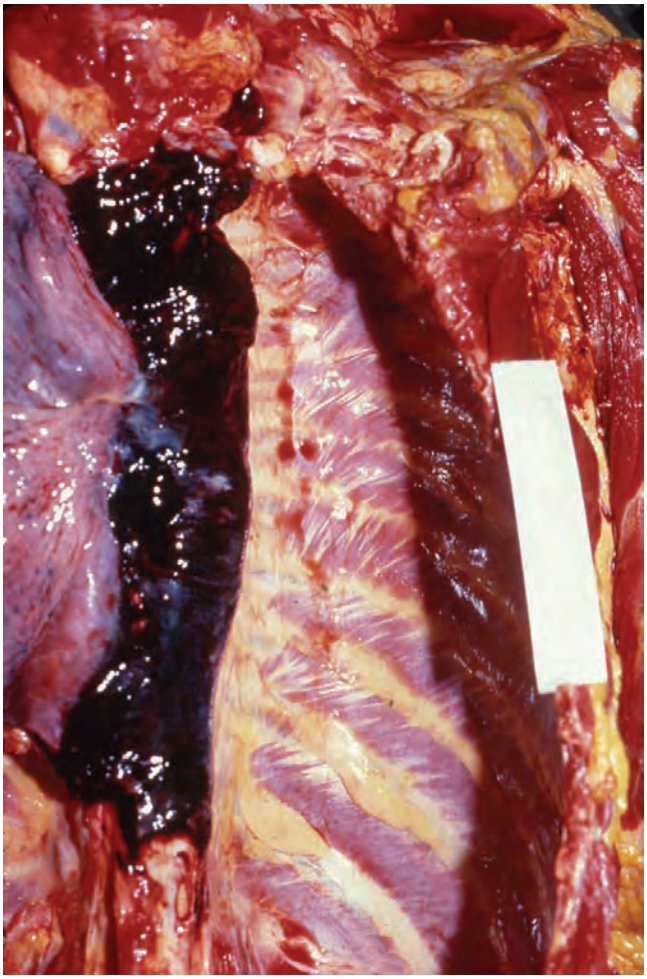

An in situ aortic dissection. Note the hemorrhage extending from the root of the aorta down the paravertebral region shown by dark red hemorrhagic discoloration. This decedent had severe hypertensive cardiovascular disease.

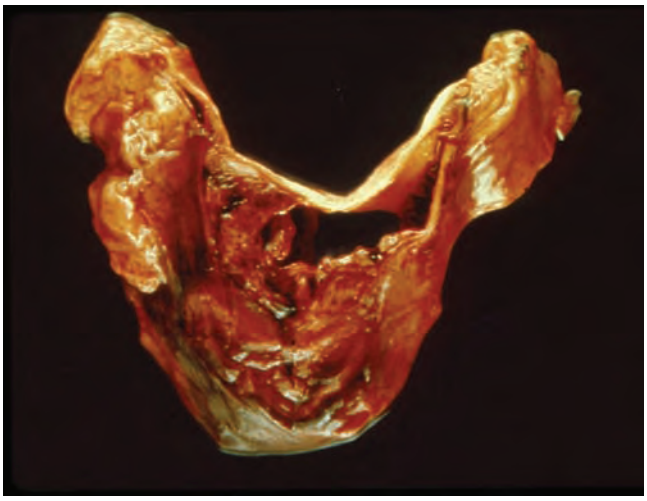

Cross section of aorta revealing a double-barrel lumen. The superior aspect of this figure shows an opened aorta with the exposed lumen. Directly inferior to this is the separated media and adventitia with a second lumen that is partially thrombosed.

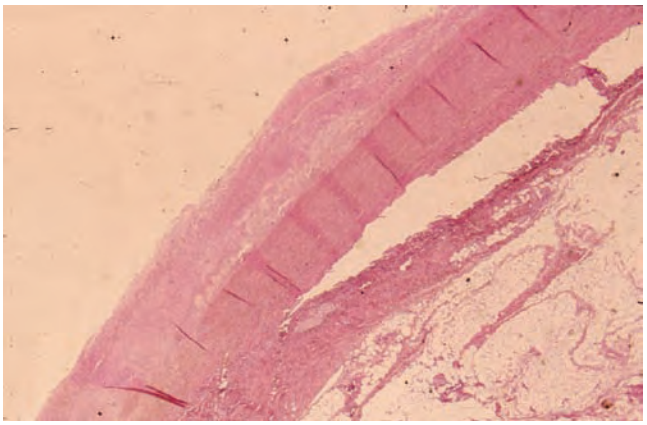

Microscopic view of an aortic dissection.

This endocardial surface shows large nonbacterial thrombotic endocarditis associated with a hypercoagulable state from metastatic adenocarcinoma. Special stains were negative for microorganisms.

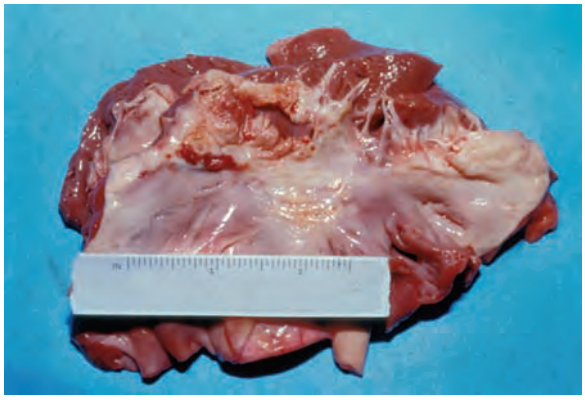

This perforated mitral valve is secondary to acute bacterial endocarditis. This individual first went to the emergency room approximately a day and a half before with the complaint of fever and chest pain. He was sent home with antibiotics and later returned with severe pulmonary edema and died shortly after.

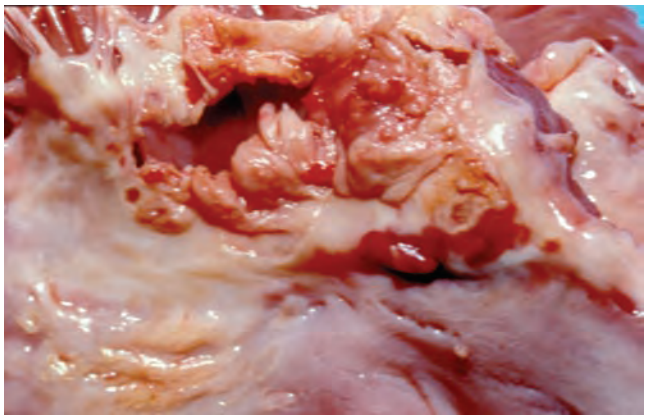

Close-up view of an acute infectious endocarditis with valve perforation. Gram stain revealed numerous gram positive organisms.

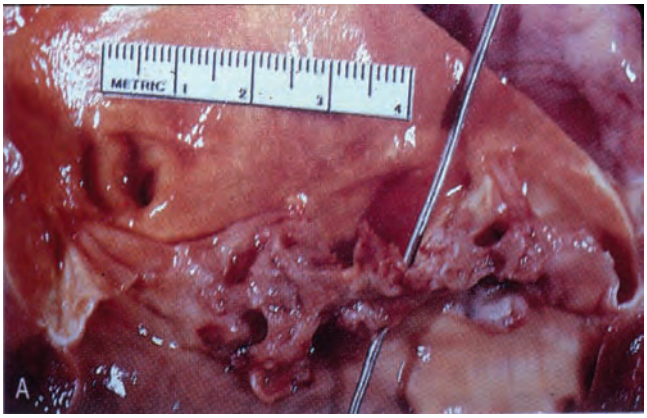

Acute infectious endocarditis with valve leaflet perforation.

Remote cardiac valve damage from rheumatic fever.

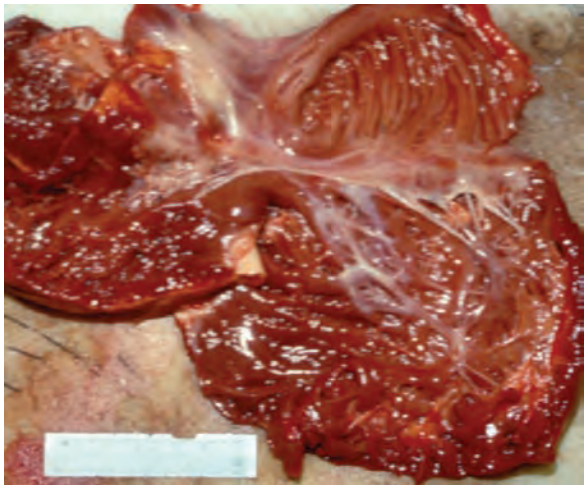

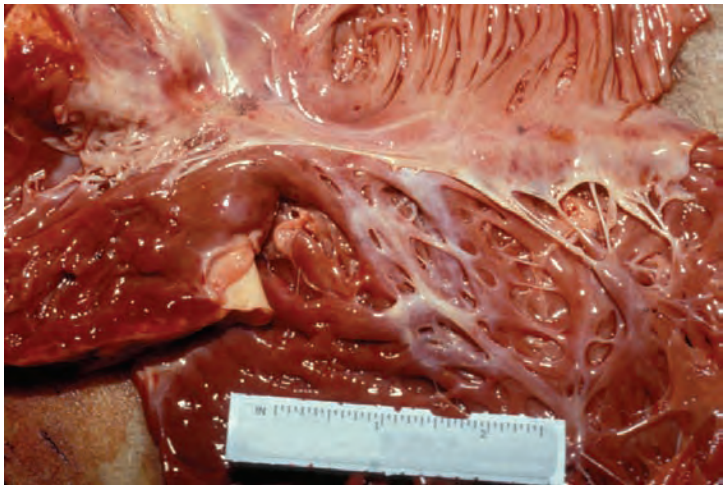

These figures are views of a remotely damaged tricuspid valve secondary to chronic intravenous drug abuse and past endocarditis. Note the fibrosis of the adjacent endocardium secondary to regurgitive turbulent blood flow. The decedent was known to have a longstanding cardiac murmur.

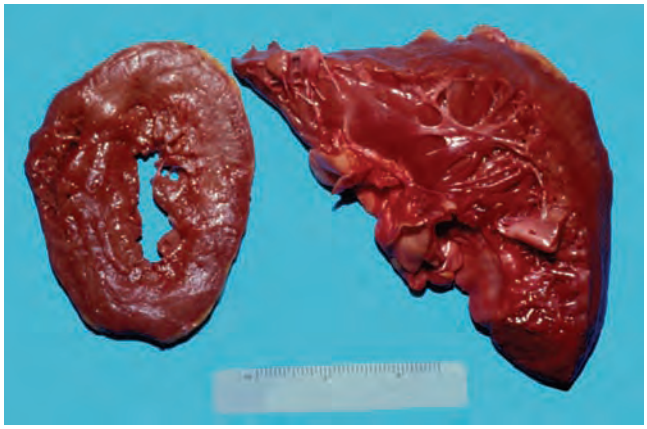

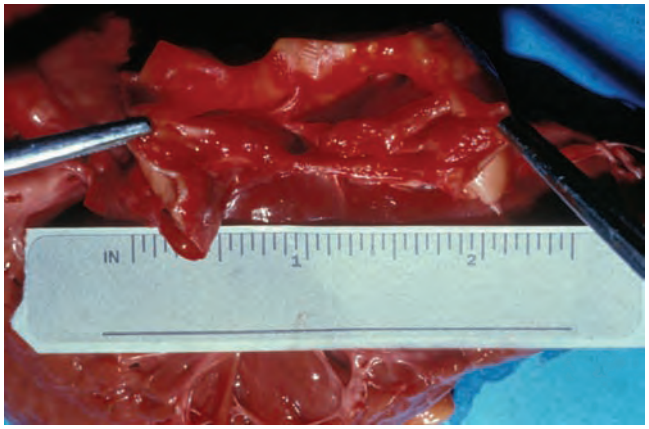

These figures show a markedly hypertrophic heart with concentric left ventricle hypertrophy and a congenital subaortic band causing marked aortic stenosis and sudden cardiac death at age 42 years. The decedent had decided years earlier not to have a valve replacement.

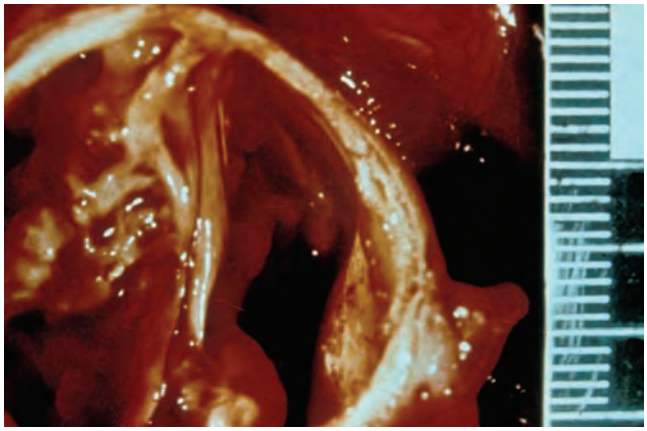

Bicuspid aortic valve. With advancing age and atherosclerosis, these valves may become markedly stenotic and increase one's risk for sudden cardiac death.

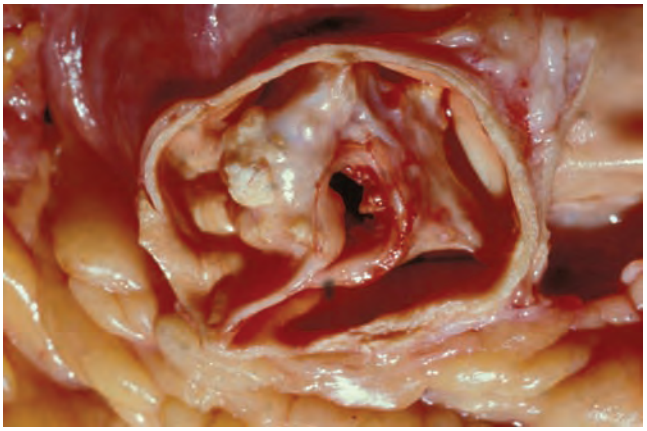

Markedly stenotic bicuspid aortic valve from an older individual with a longstanding history of atherosclerotic cardiovascular disease.

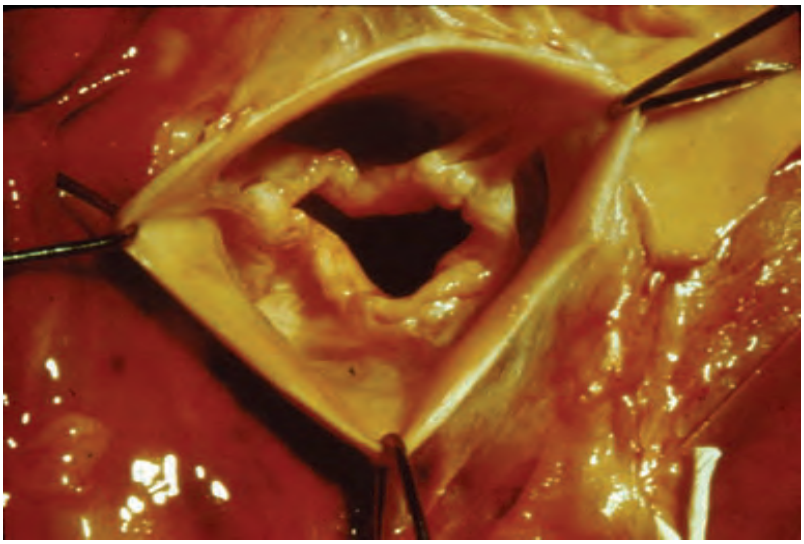

Severely stenotic and insufficient aortic valve associated with childhood rheumatic fever. Correction of aortic stenosis will decrease the risk of sudden cardiac death.

A fulminant pulmonary edema with foam extending from the mouth and nose due to congestive heart failure.

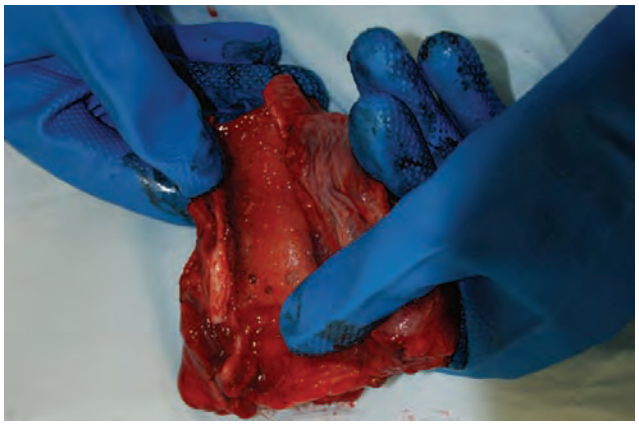

Note the frothy fluid from pulmonary edema extending into the laryngeal airway.

Pitting edema of the leg due to congestive heart failure.

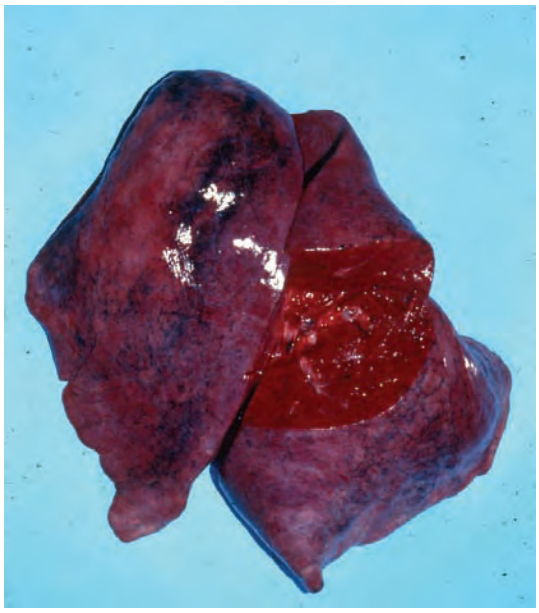

Lung with marked congestion and edema. Note the diffuse purple discoloration.

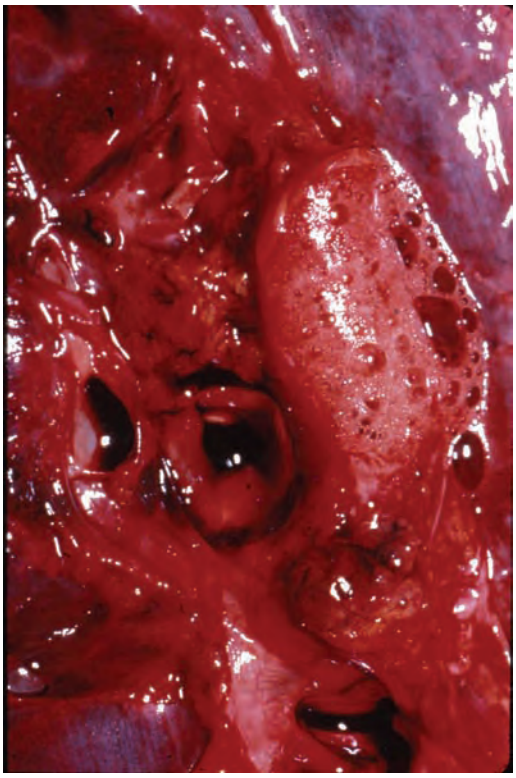

Marked pink to red frothy fluid extending from the cut parenchyma of a lung due to fulminant pulmonary edema.

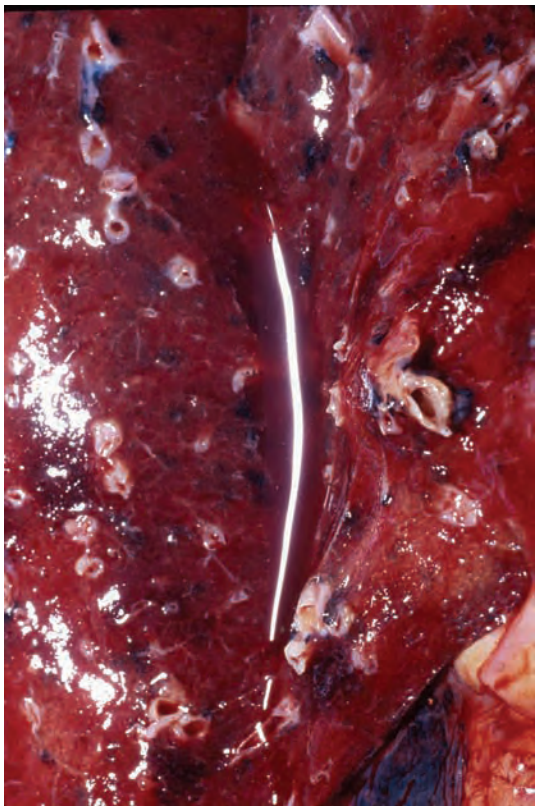

Cut section of a lung with pulmonary edema.

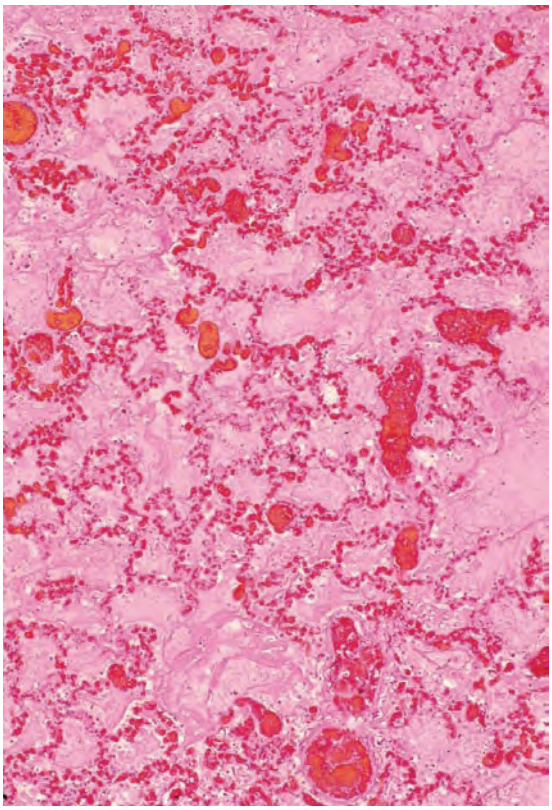

Microscopic view of lung with acute fulminant pulmonary edema.

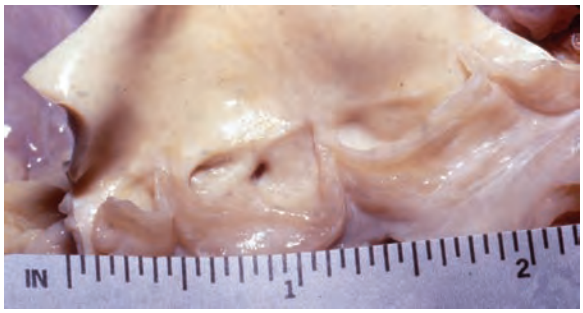

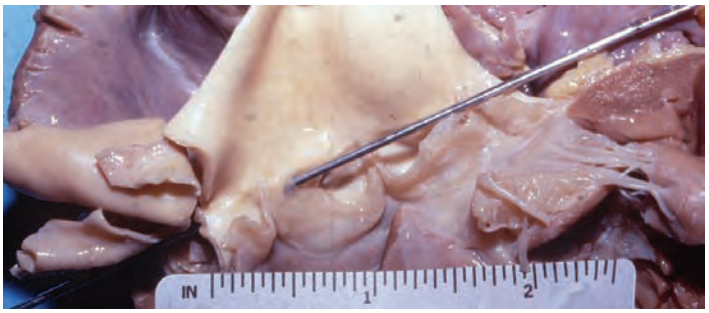

These two figures demonstrate a coronary artery anomaly with acute angle takeoff and luminal narrowing in a 15-year-old who died suddenly during a basketball game. There was no history of blunt impacts to the chest during the game. There was no past history of syncopal episodes or chest discomfort.

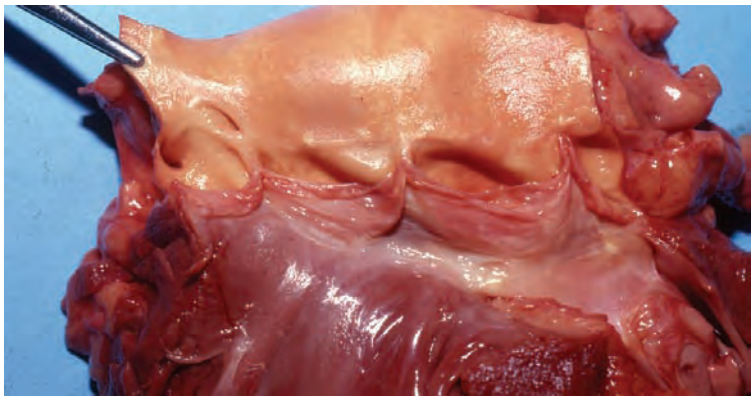

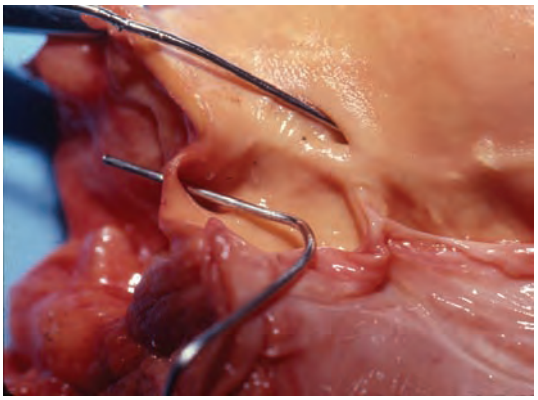

A coronary artery anomaly with acute angle takeoff and luminal narrowing. This individual died a sudden cardiac death shortly after exertion.

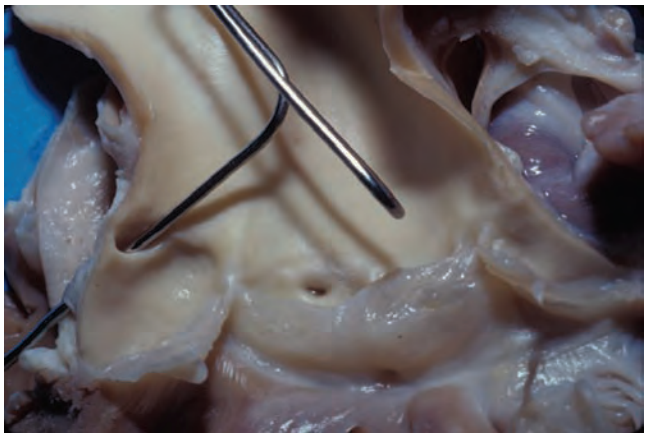

This coronary anomaly reveals a bicuspid aortic valve with superior displacement of one of the coronary ostia. This child died of trauma sustained in a motor vehicle accident and this finding was incidental.

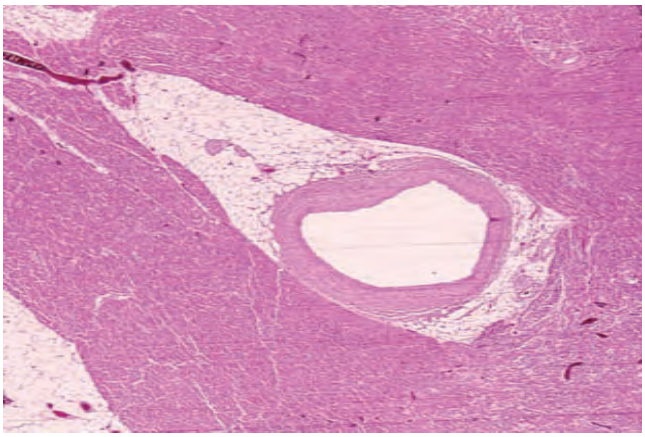

This low-power histopathologic view demonstrates "myocardial tunneling" where the left anterior descending coronary artery dipped deeply into the left ventricle wall.

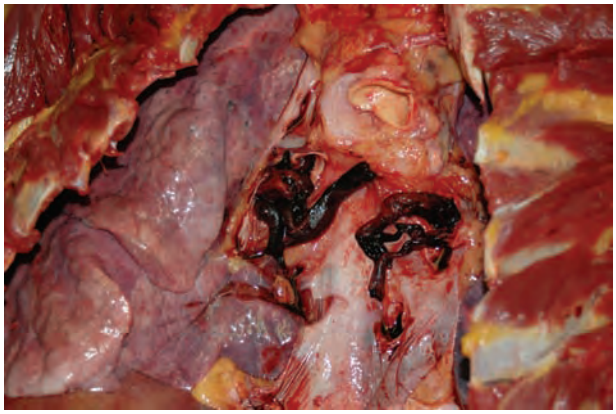

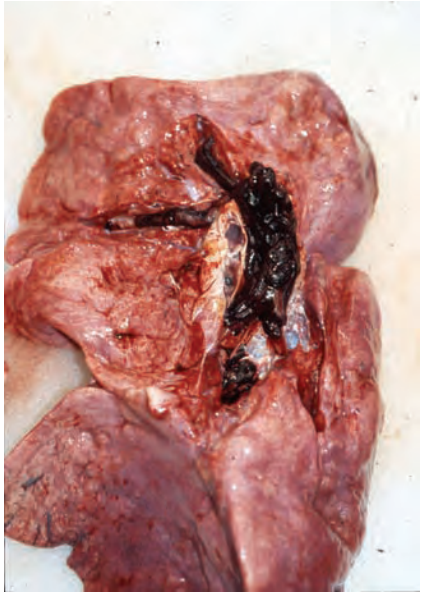

Pulmonary thromboembolus in an obese woman who was on birth control pills and smoked cigarettes.

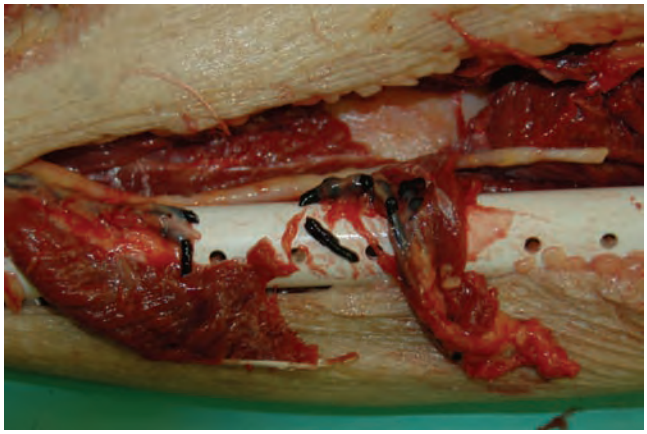

Incision of the lower leg with dissection of the gastrocnemius muscle demonstrating multiple deep venous thrombi in an individual who underwent organ donation with removal of bone and soft tissue from each leg. Note the plastic tubing left by the organ donor network.

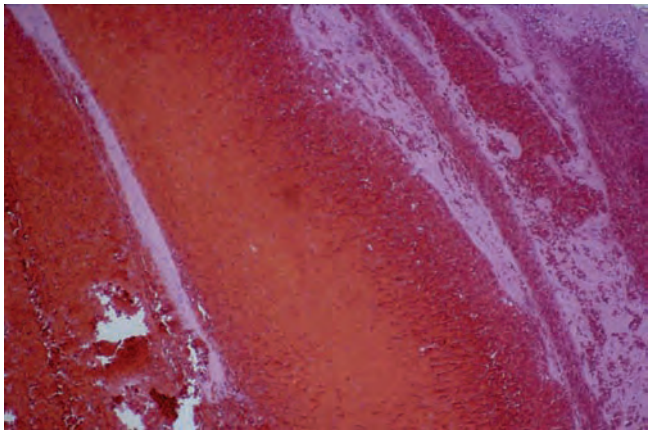

Microscopic view of a pulmonary thromboembolus. Note the laminated alternating lighter fibrin bands referred to as the lines of Zahn.

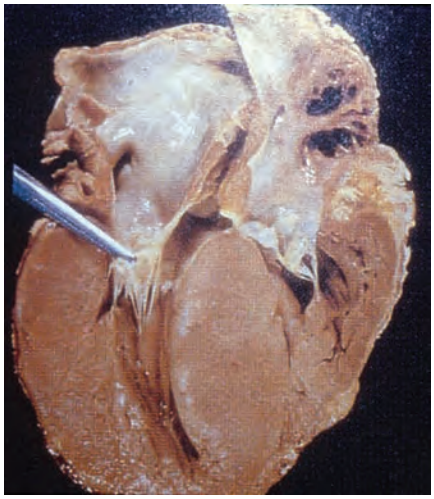

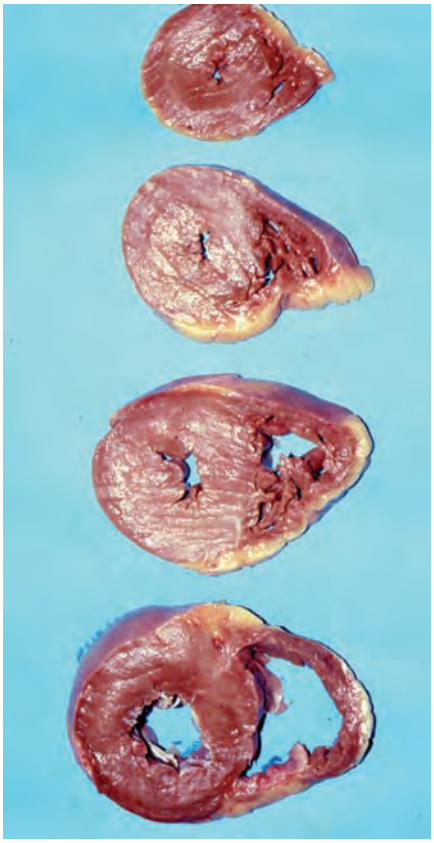

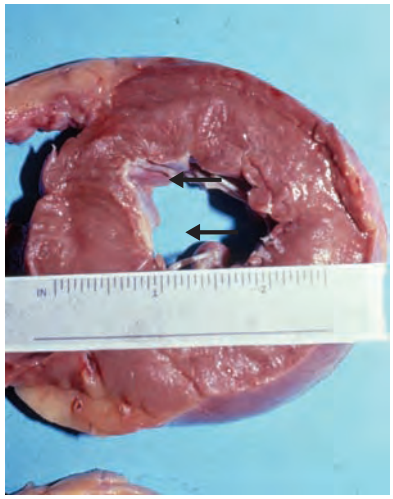

These are different views of two hearts with the classical variant of hypertrophic cardiomyopathy. Note the large degree of asymmetric left ventricle hypertrophy.

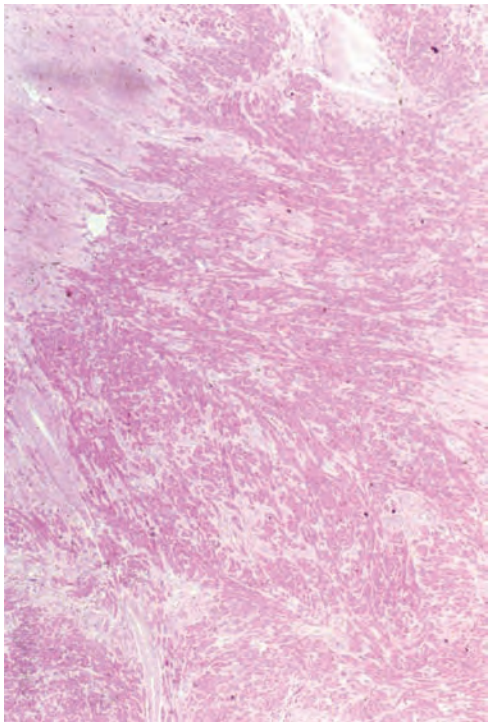

Hypertrophic cardiomyopathy with myocyte fiber disarray and fibrosis.

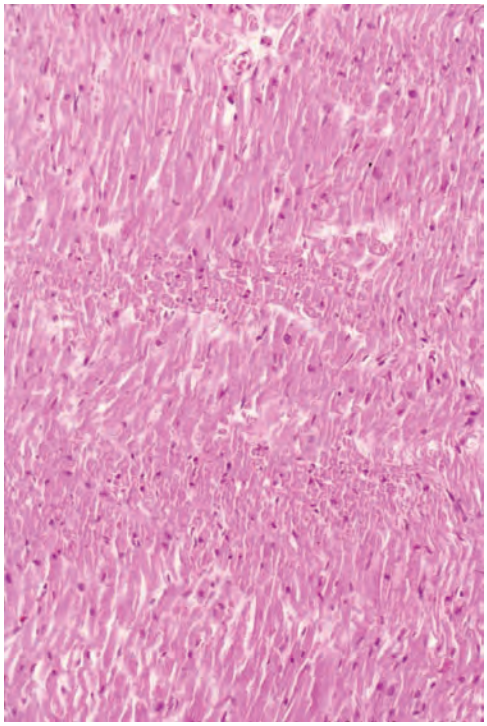

Hypertrophic cardiomyopathy with myocyte fiber bundle disorganization.

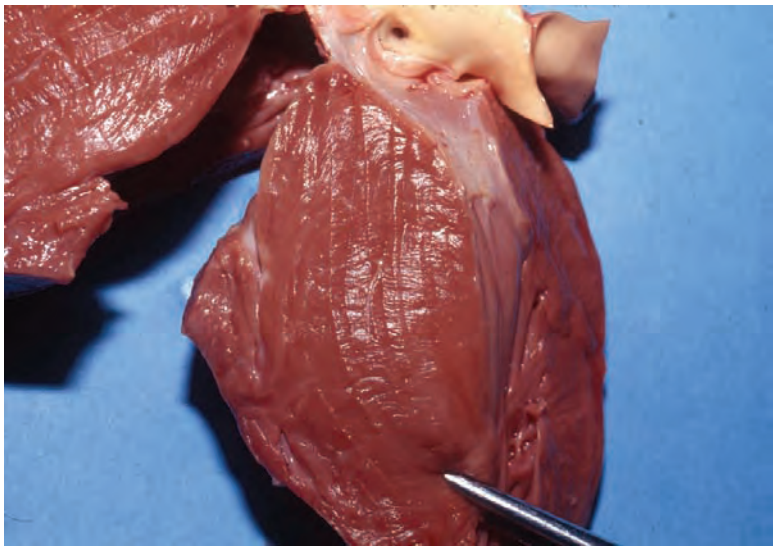

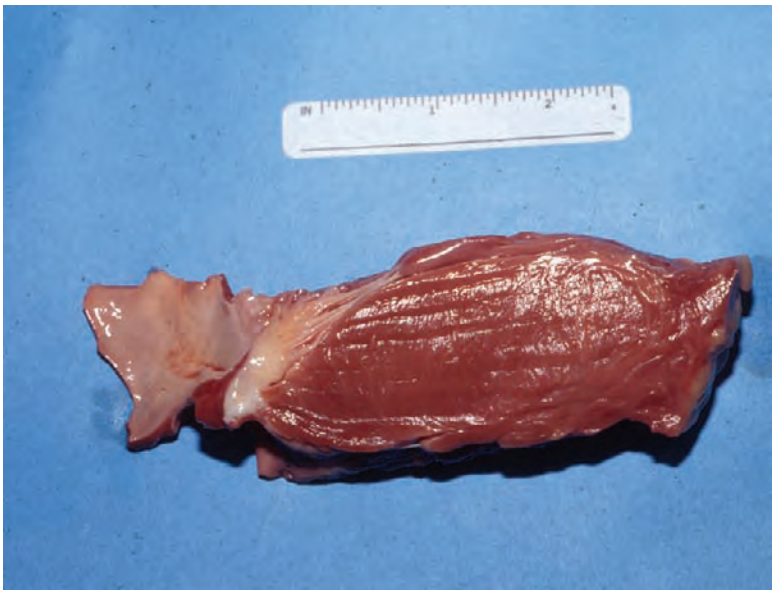

Hypertrophic cardiomyopathy. Note the marked septal hypertrophy with asymmetry.

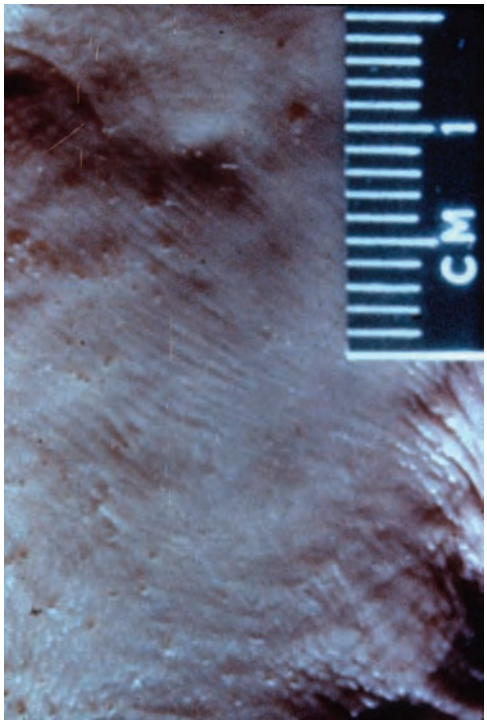

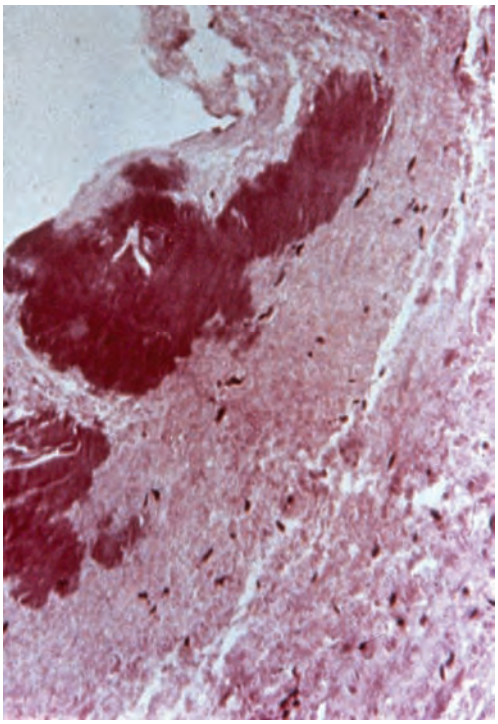

Note the waxy pink myocardial discoloration in this individual with amyloidosis.

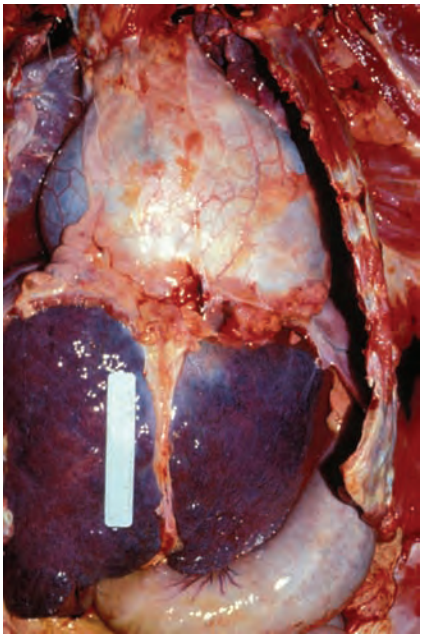

This decedent's pericardial sac was full of several hundred cc's of yellow-pink fluid. The gradual buildup of this fluid over a long period was due to advanced secondary hemochromatosis.

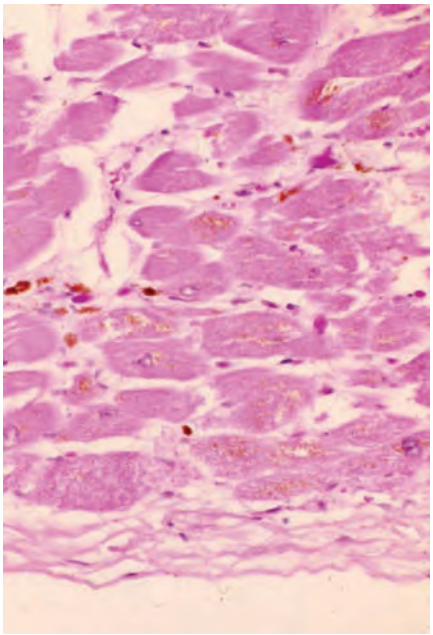

Microscopic section of heart showing iron deposition from secondary hemachromatosis associated with blood transfusions for treatment of beta-thalassemia major.

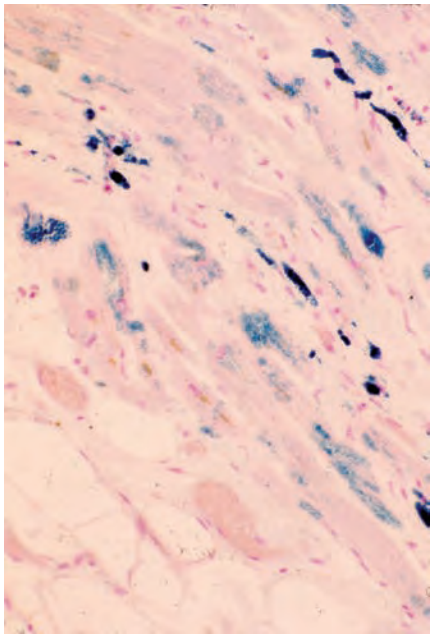

Microscopic section of heart with Prussian blue staining of iron deposits in the heart of the same individual with secondary hemochromatosis.

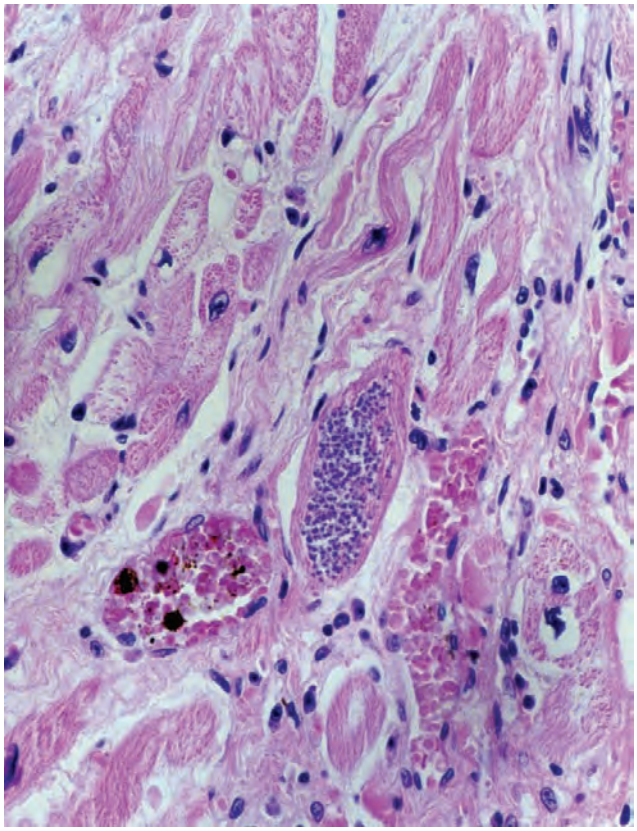

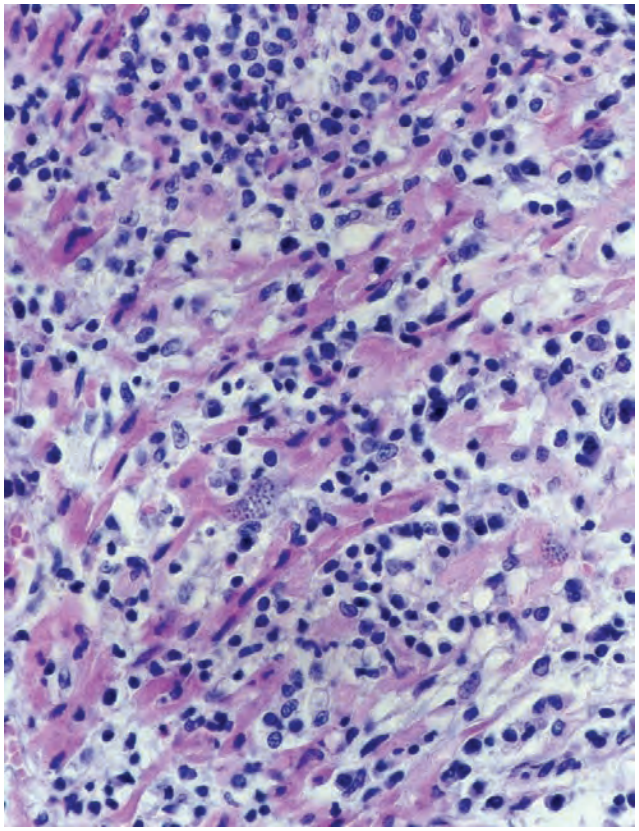

These microscopic myocardial views demonstrate Chagas' disease. Note the extensive chronic inflammation.

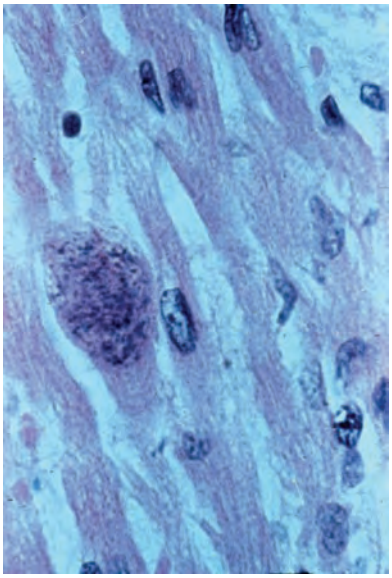

Myocarditis secondary to toxoplasmosis in an individual with acquired immunodeficiency syndrome.

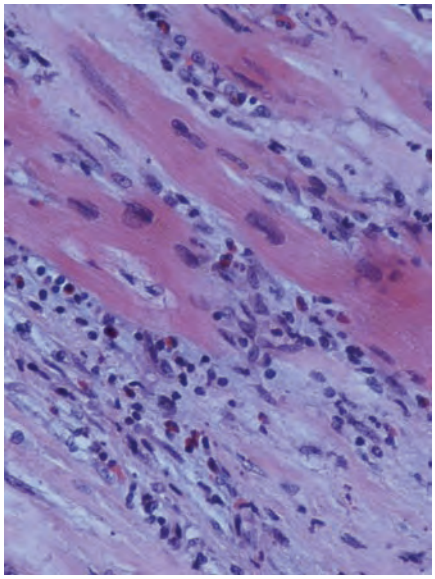

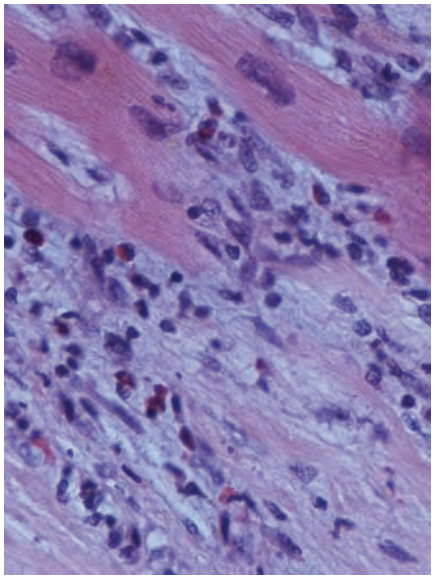

Eosinophilic myocarditis associated with a hypersensitivity drug reaction.

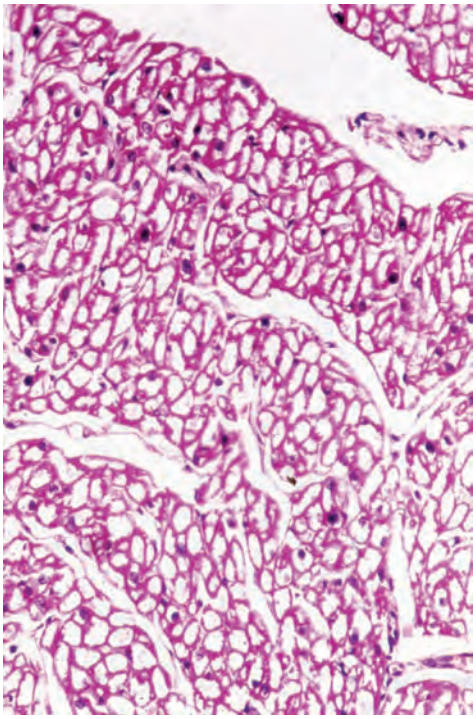

Microscopic view of myocardium in an individual who died of Pompe's disease. Note the many artifactual empty vacuoles following processing leading to glycogen loss.

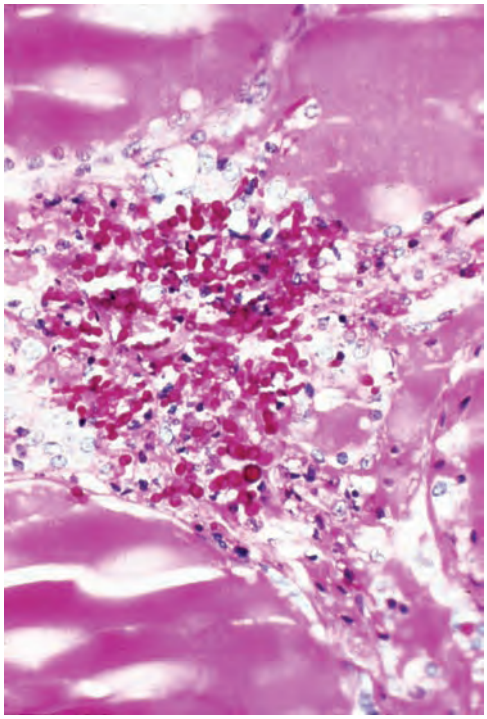

This candidal myocarditis was found in an individual with acquired immunodeficiency syndrome.

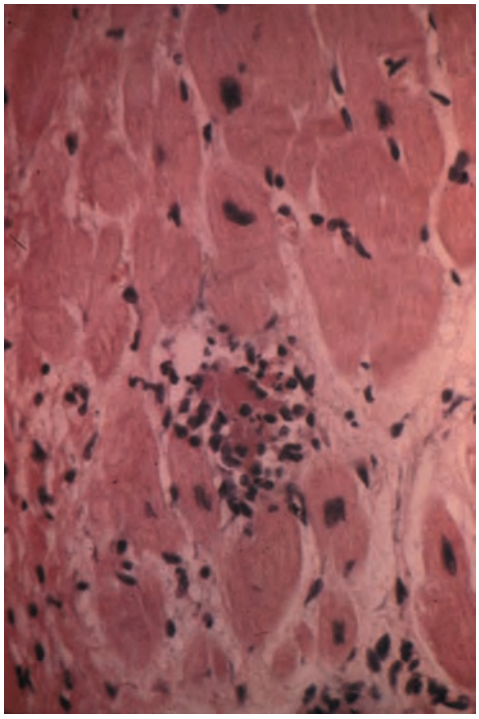

This focal viral myocarditis demonstrates myocardial necrosis with inflammation made up predominately of lymphocytes.

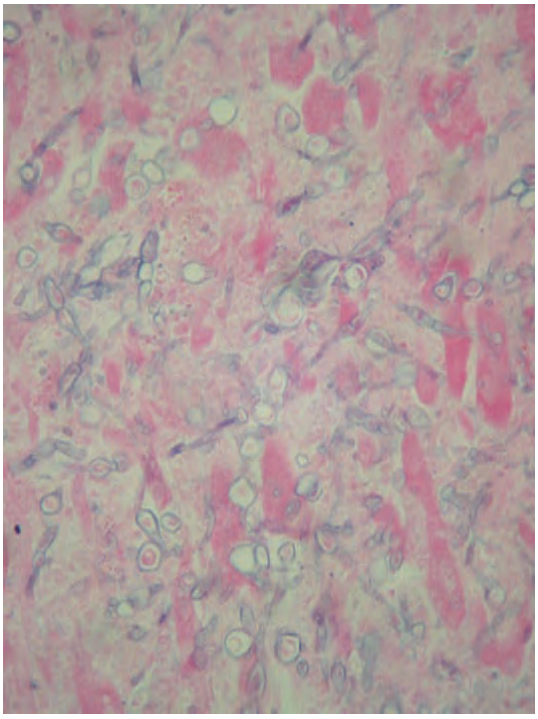

Aspergillosis myocarditis from an immuno-compromised individual who underwent a recent bone marrow transplant.

Macroscopic section of myocardium from an individual who died of cardiac sarcoidosis.

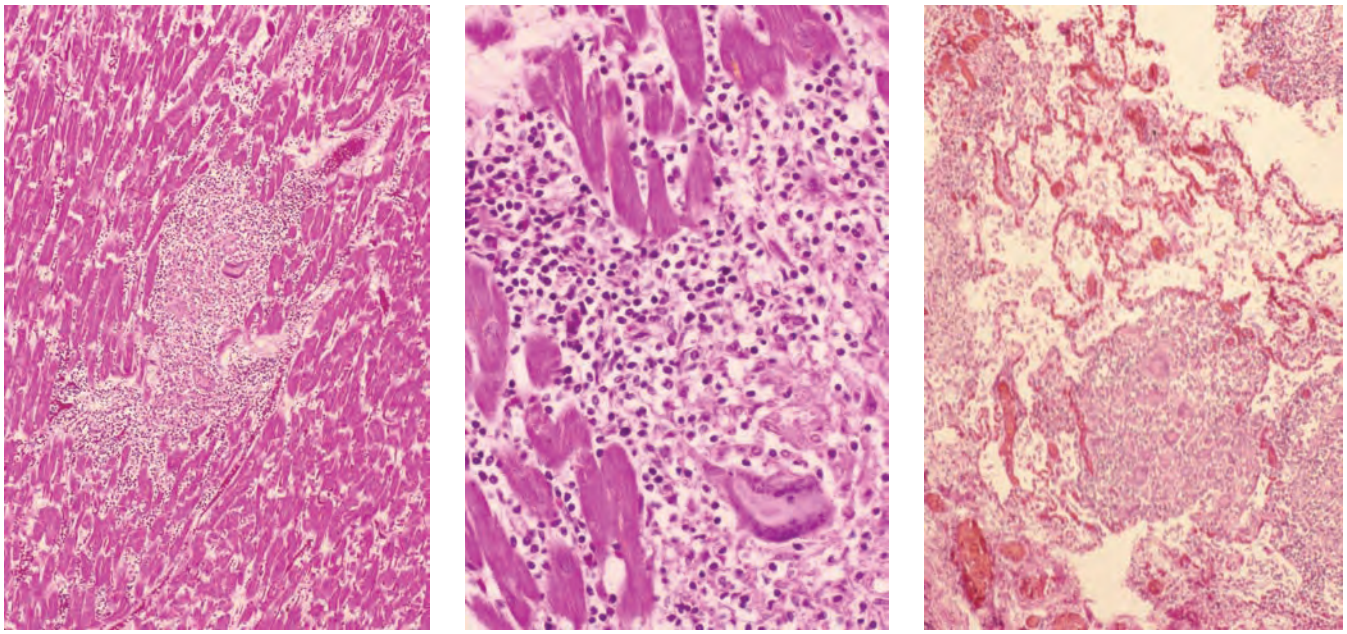

Sarcoidosis. Note the noncaseating granulomas consisting of chronic inflammation, epitheloid cells, lymphocytes, and giant cells found throughout the body. These views demonstrate examples from the heart and lungs.

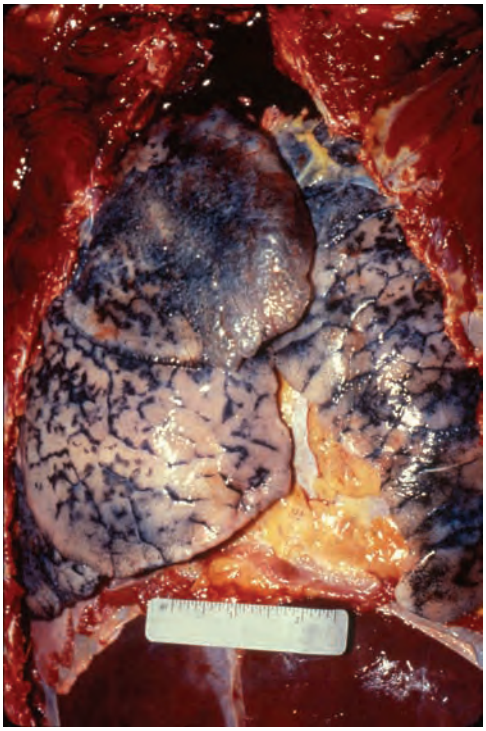

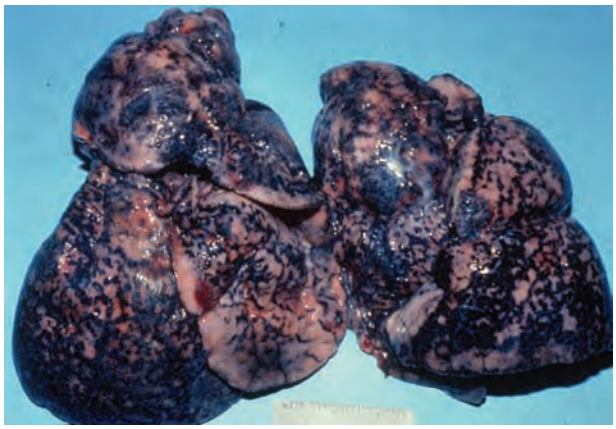

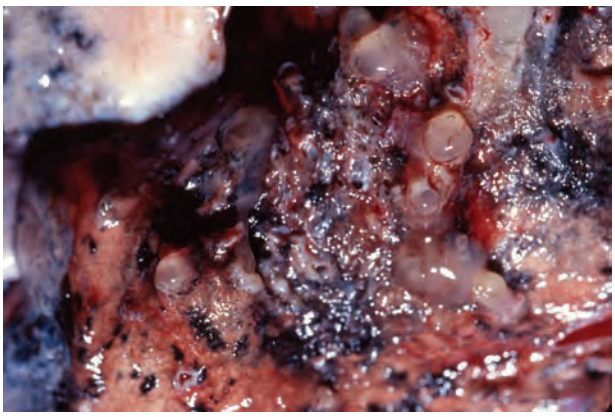

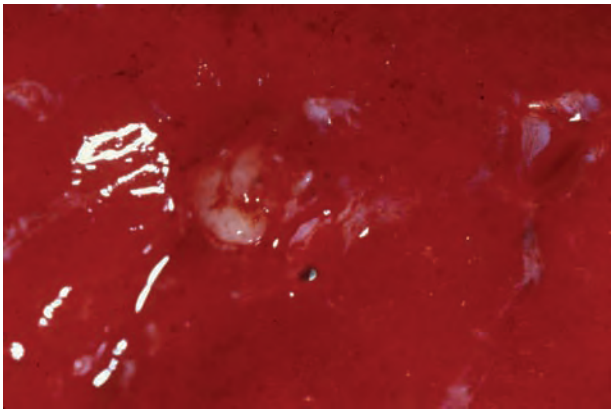

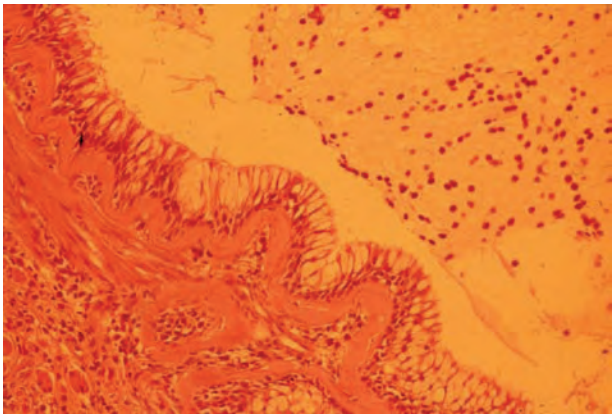

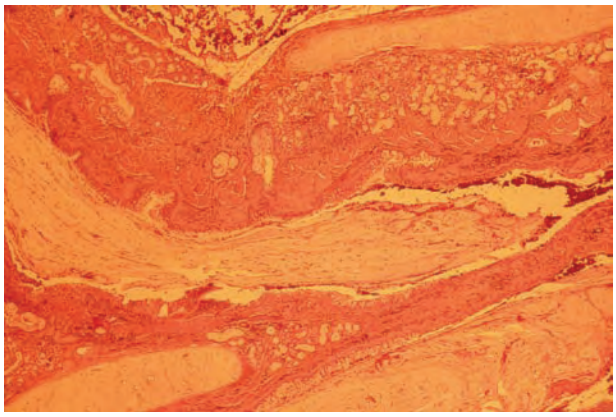

These figures demonstrate acute and chronic bronchial asthma. Note the lungs within the thoracic cavity are hyperaerated and expand to overlie the pericardial sac. Upon removal of the lung, they appear markedly hyperaerated. If these lungs were placed in a water bath they would float almost entirely on the surface. Cut section through the parenchyma reveals thick copious mucoid secretions within the bronchial distribution. Microscopic section reveals increased amounts of goblet cells and smooth muscle. There is a large mucous plug within the airway with numerous eosinophils. The parenchyma also has numerous inflammatory cells, predominantly eosinophils.

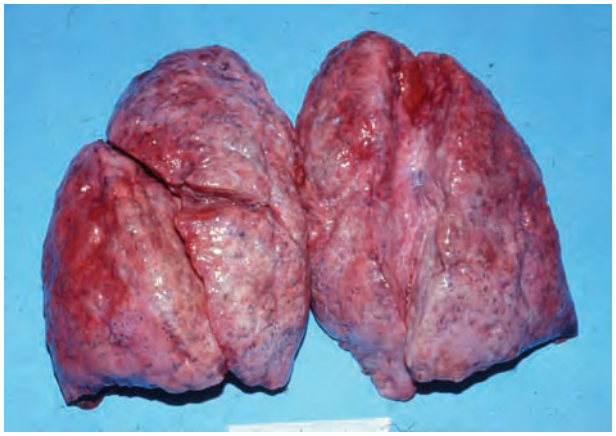

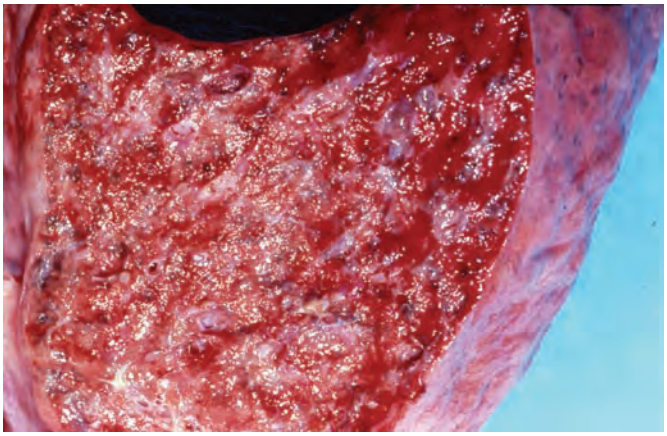

Chronic obstructive pulmonary disease (COPD) with hyperaeration and emphysematous change.

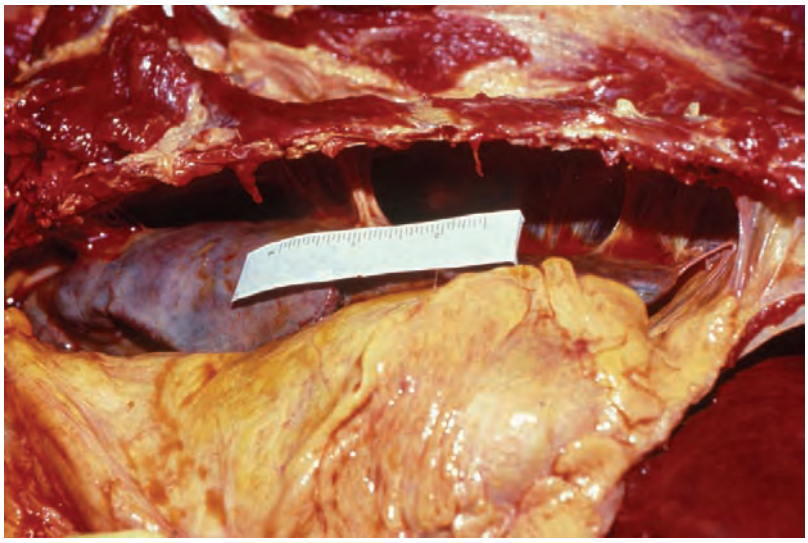

Pleural adhesions most likely due to past bouts of pneumonia.

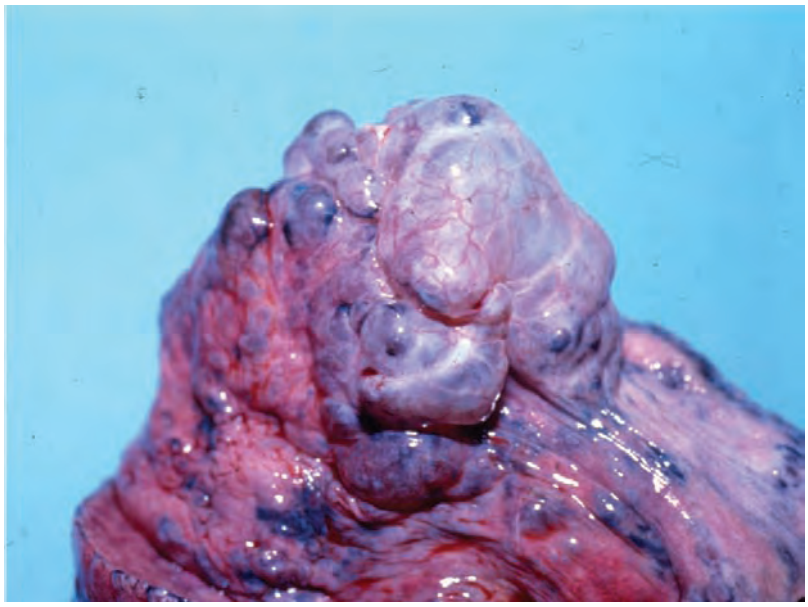

Bolus emphysema in a person with COPD. These bullae may occasionally rupture and cause a pneumothorax.

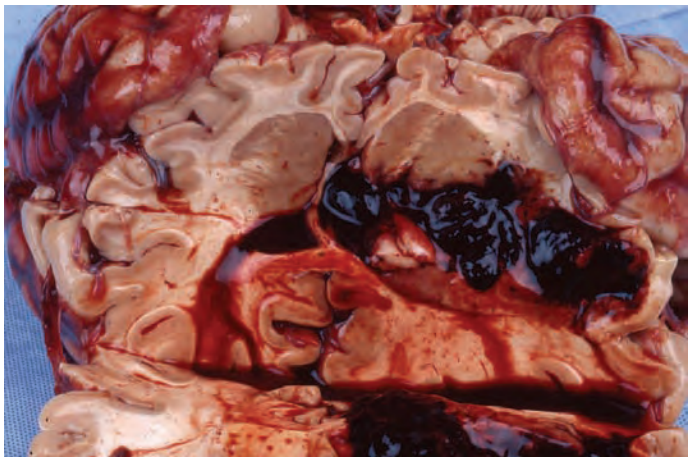

Hypertensive intracerebral hemorrhage. Mostly caused by rupture of a small intraparenchymal vessel.

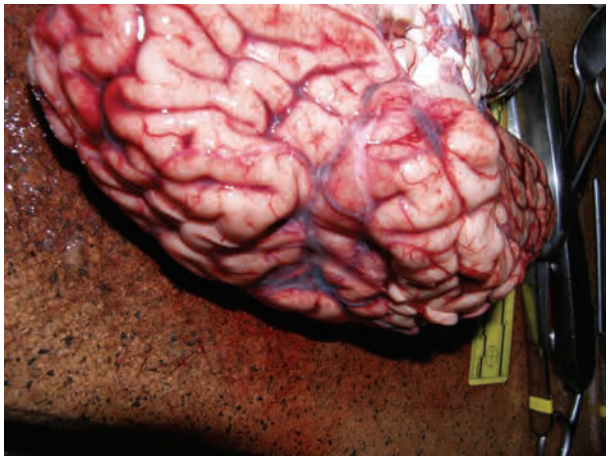

Old healed stroke. Note the indentation of the cerebral hemisphere.

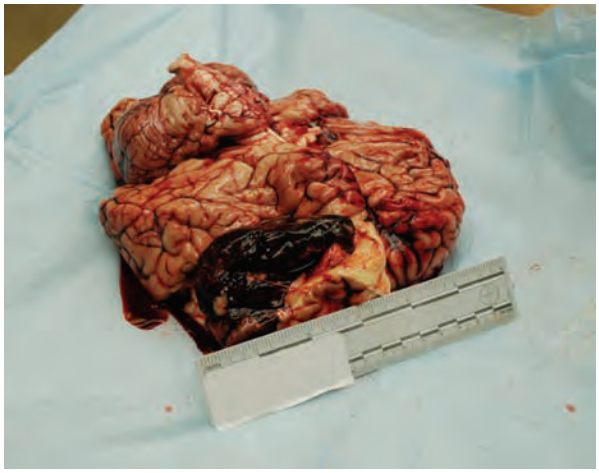

Intracerebral hemorrhage due to ruptured A-V malformation.

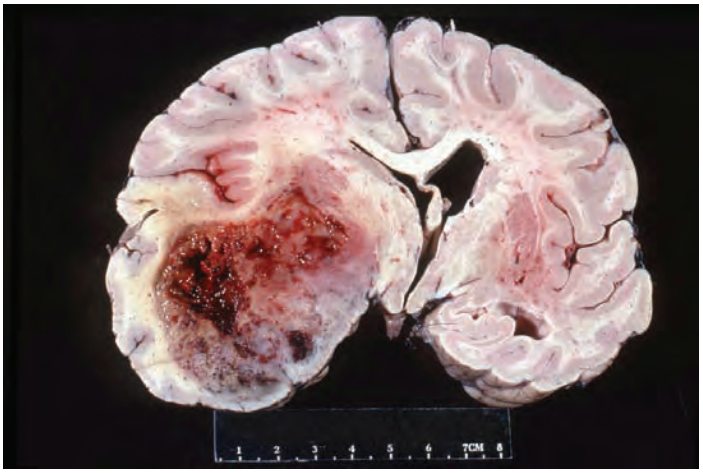

Astrocytoma causing significant compression of the surrounding structures.

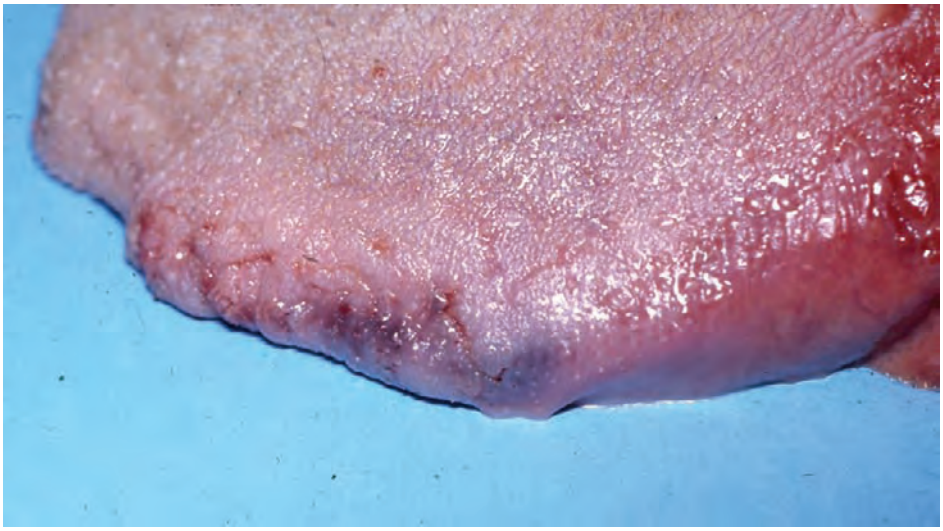

Bite marks to tongue with hemorrhage from an individual who died of epilepsy. This was the only finding at autopsy.

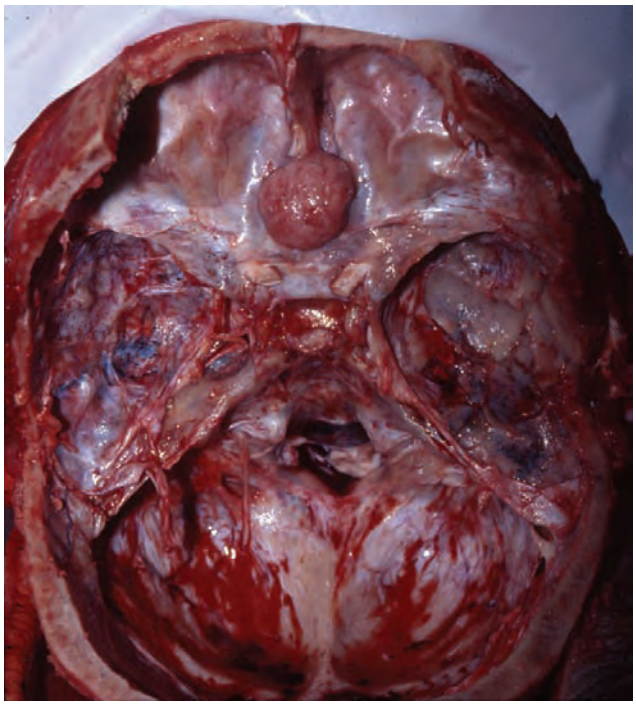

Anterior cranial fossa meningioma.

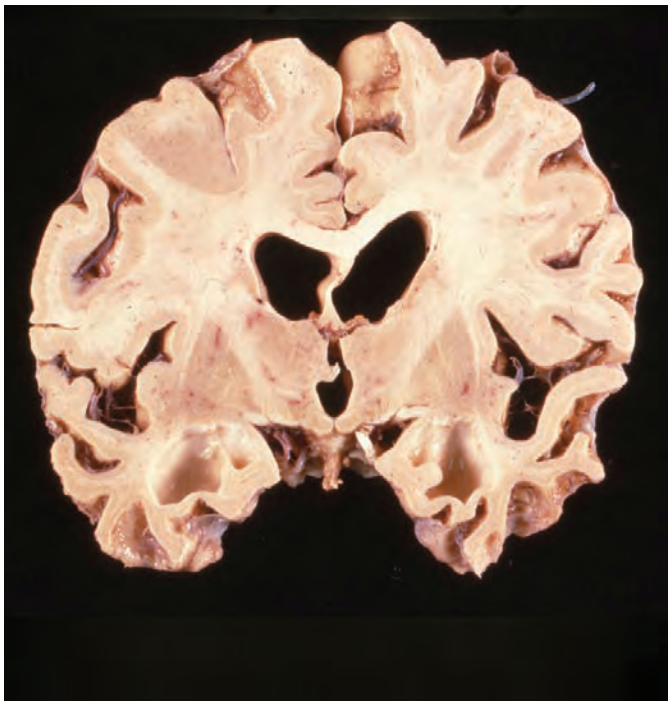

Pick's disease. Note the asymmetric atrophy of the frontal, temporal, and parietal lobes. This form of chronic dementia occurs far less frequently than Alzheimer's disease.

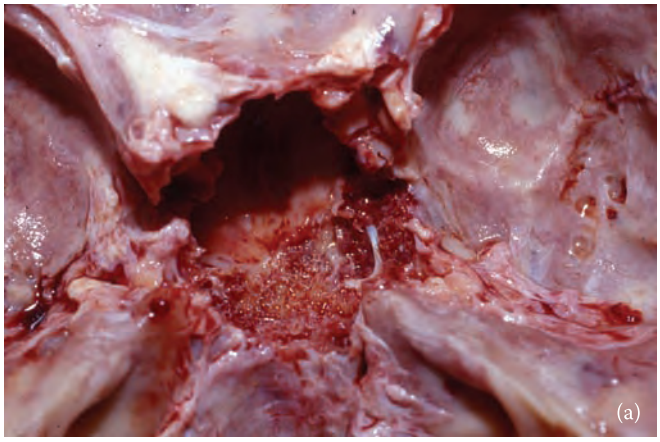

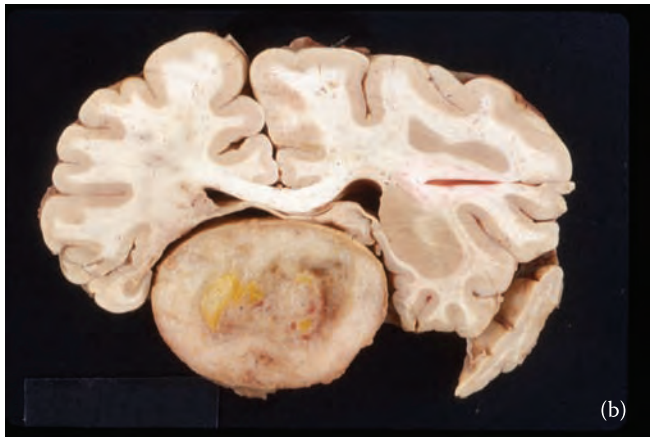

Two separate cases of pituitary adenoma. Figure (a) depicts a large erosion into the sela turcica and (b) demonstrates a large adenoma viewed on a cross section of a cut formaldehyde-fixed brain. These adenomas may vary largely in size.

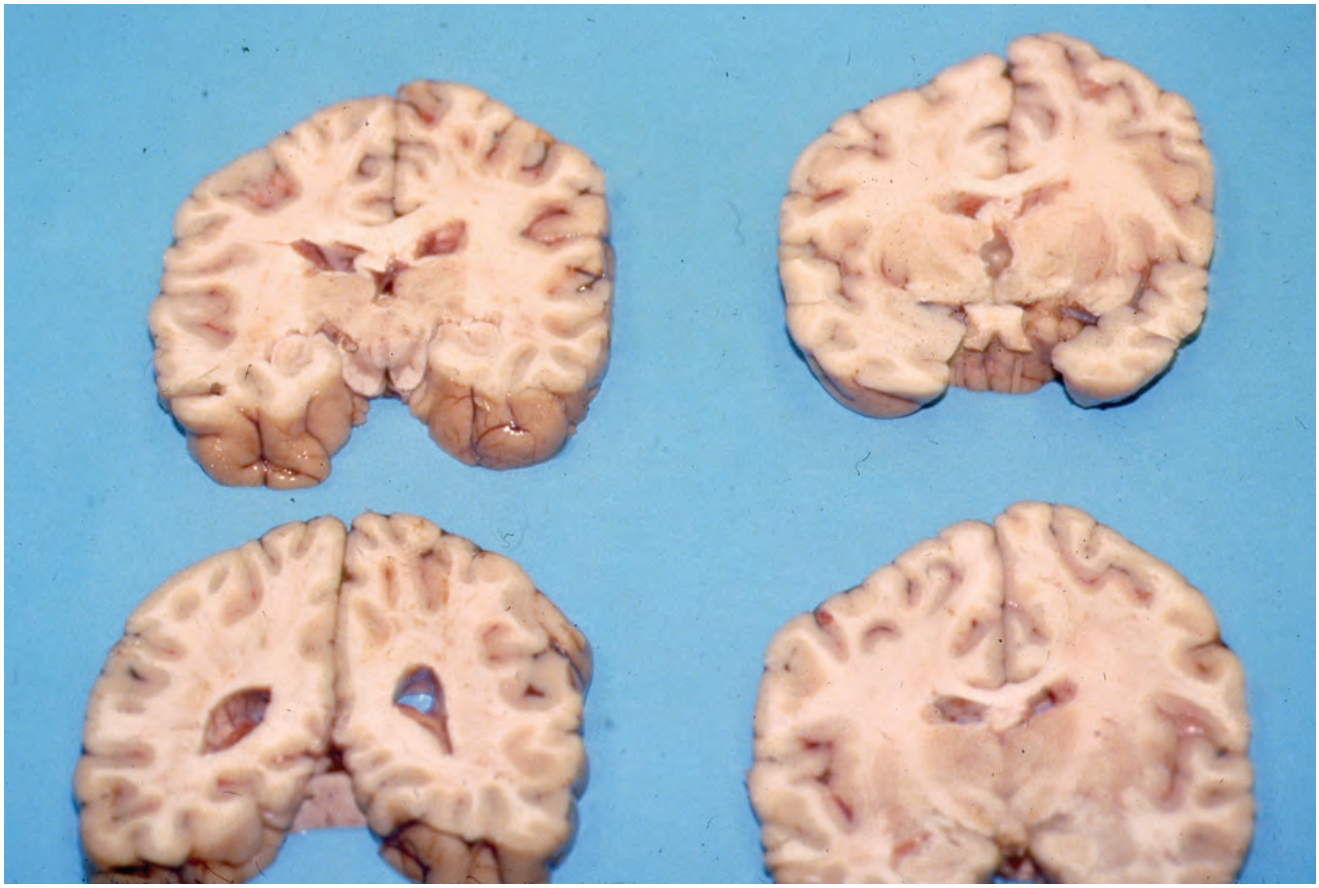

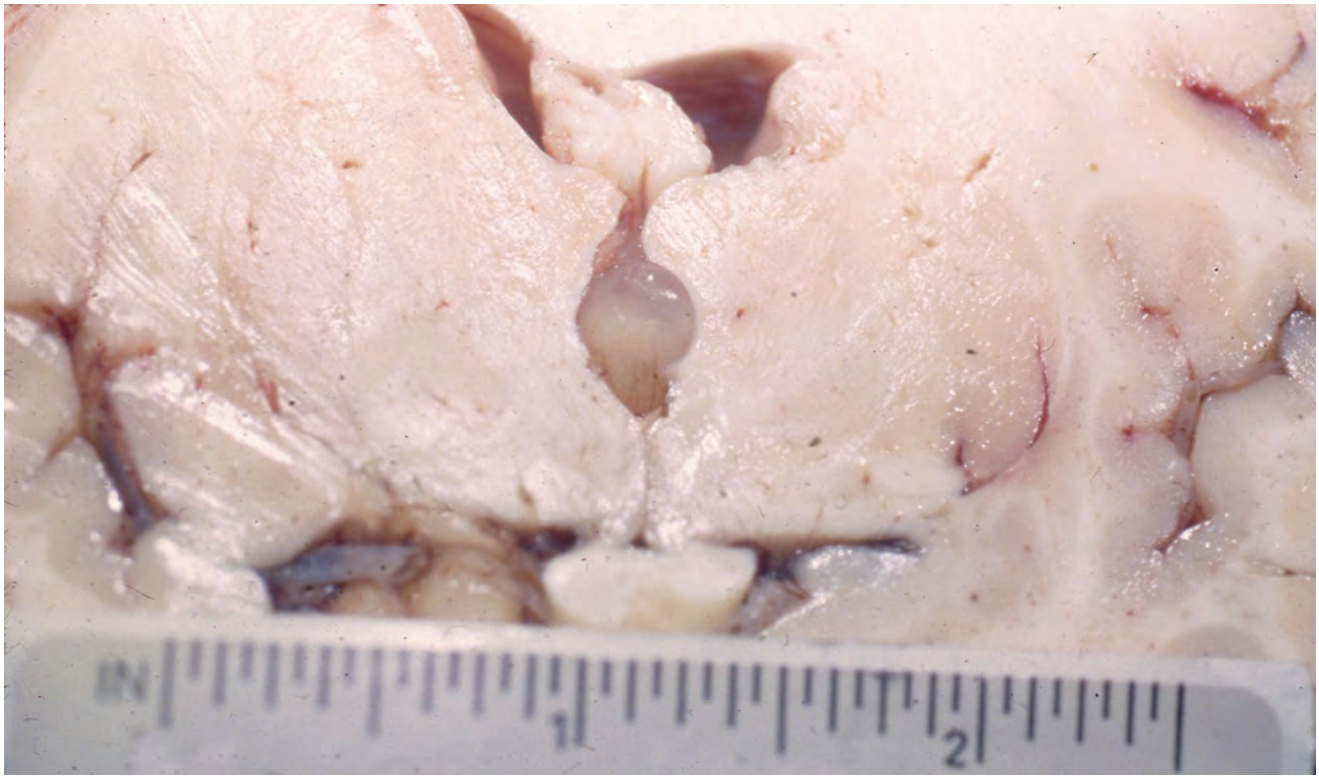

Colloid cyst of the third ventricle. This decedent had a history of severe headache with postural changes. This may be associated with sudden cardiac death following a buildup of cerebrospinal fluid pressures associated with central nervous system cardiac center disruption and fatal arrhythmia.

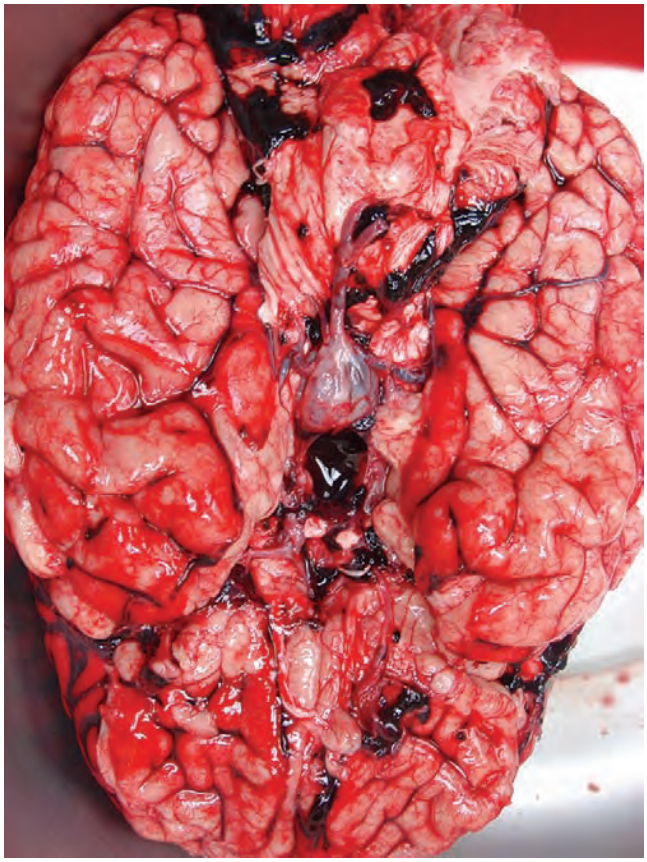

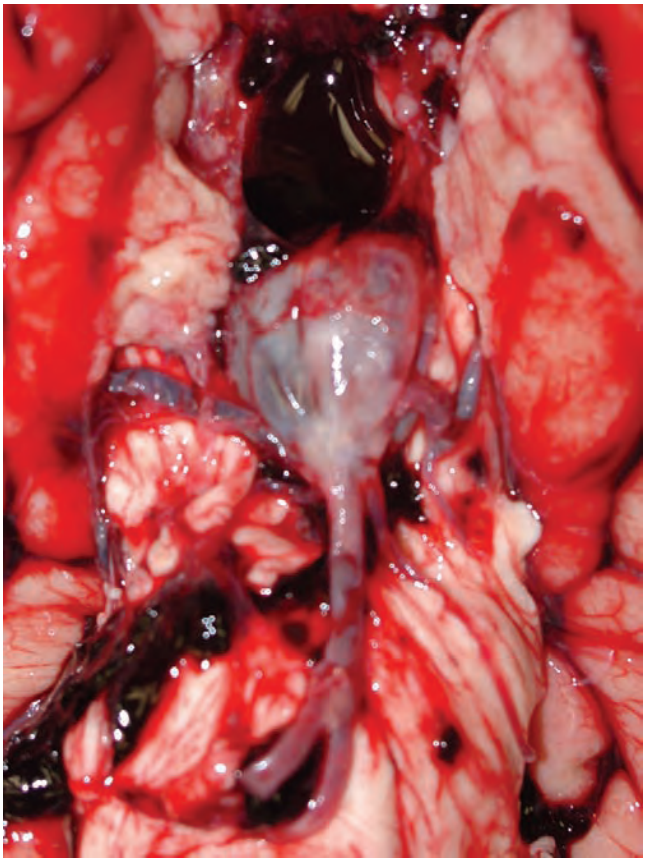

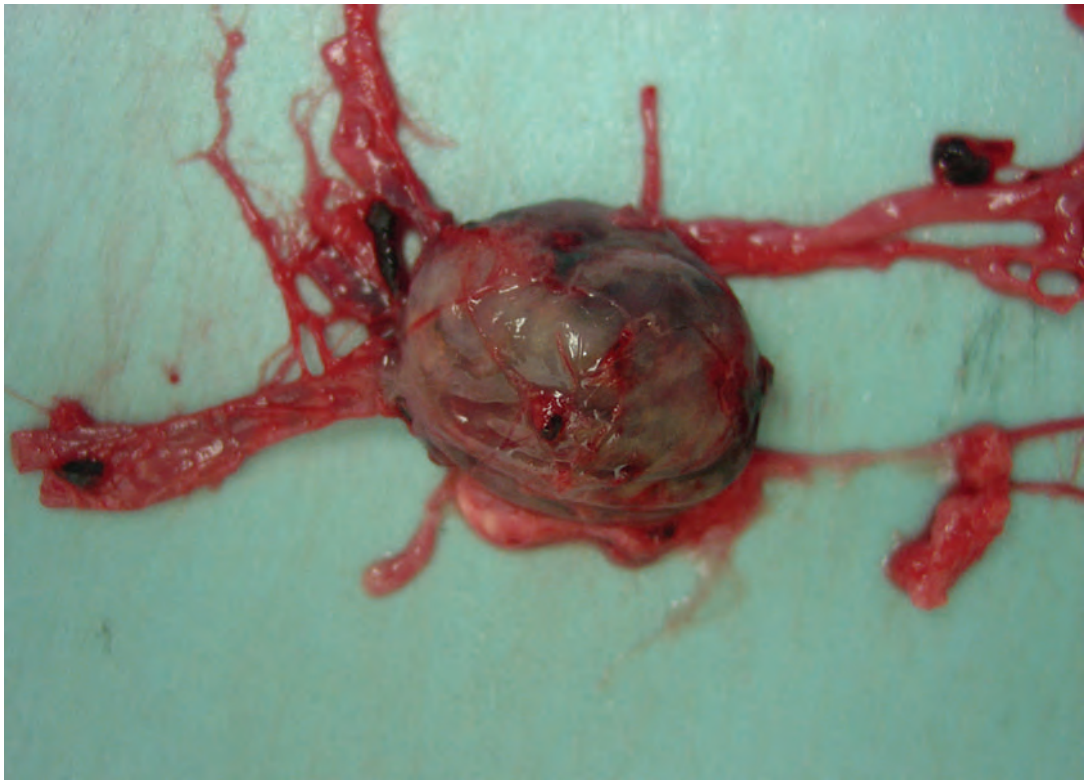

Sudden death associated with ruptured saccular cerebral artery aneurysm with subarachnoid hemorrhage. These are examples of giant berry aneurysms, which are greater than 2.5 centimeters in greatest dimension. Small saccular aneurysms may rupture the same way. The jolt of increased pressure from arterial blood through the subarachnoid space at the base of the brain may disrupt the cardiac centers, causing fatal arrthymia.

Decubital ulceration or pressure sore. This is often associated with poor patient care.

Healing ankle ulceration associated with peripheral vascular disease.

Dry gangrene associated with peripheral vascular disease due to long-term diabetes mellitus.

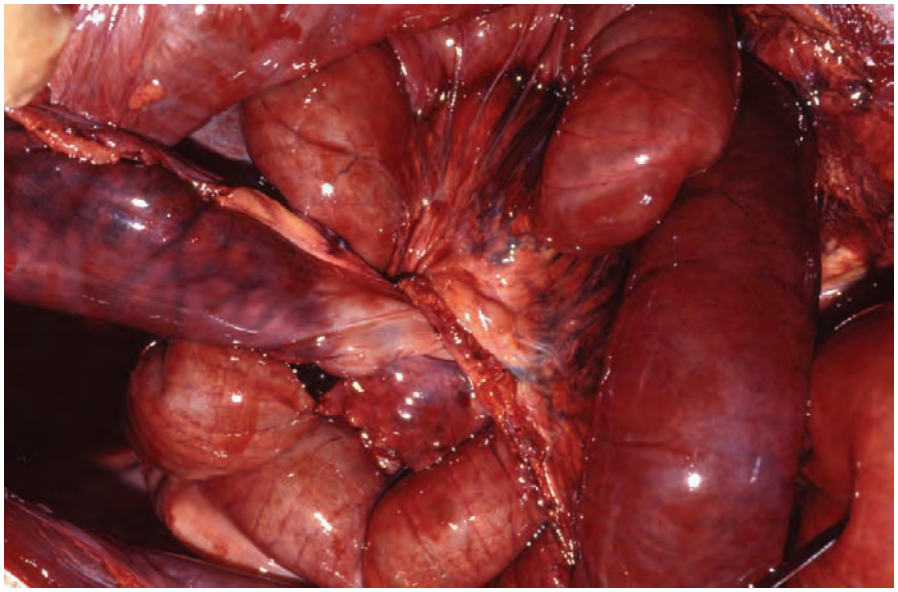

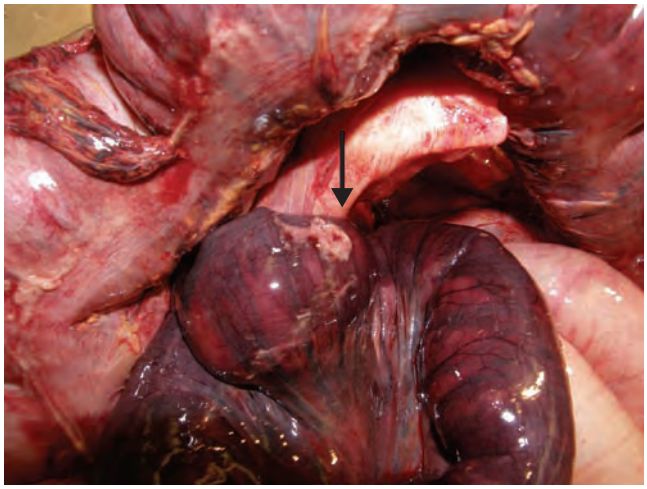

Ischemic bowel due to small intestine volvulus with vascular compromise. The arrow indicates purulent exudate at the serosal surface.

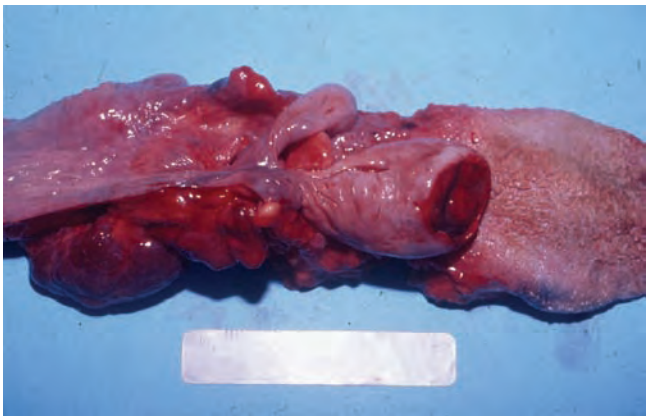

This decedent had a longstanding history of difficulty swallowing. She was seen to gesture as though she could not breathe and then collapsed. Autopsy revealed a large benign esophageal polyp that obstructed the upper airway. Note the ulceration at the tip of the polyp from constant rubbing. During this last episode she was unable to clear the obstruction.

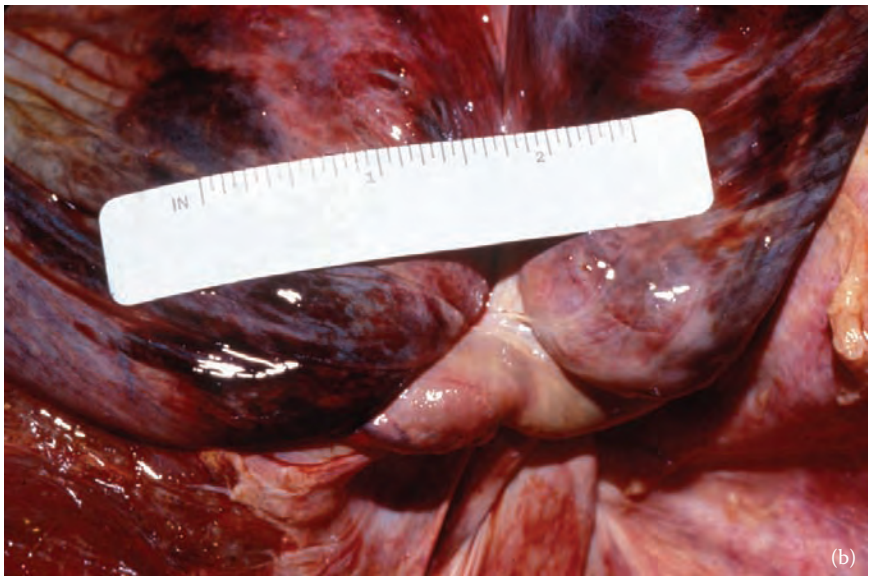

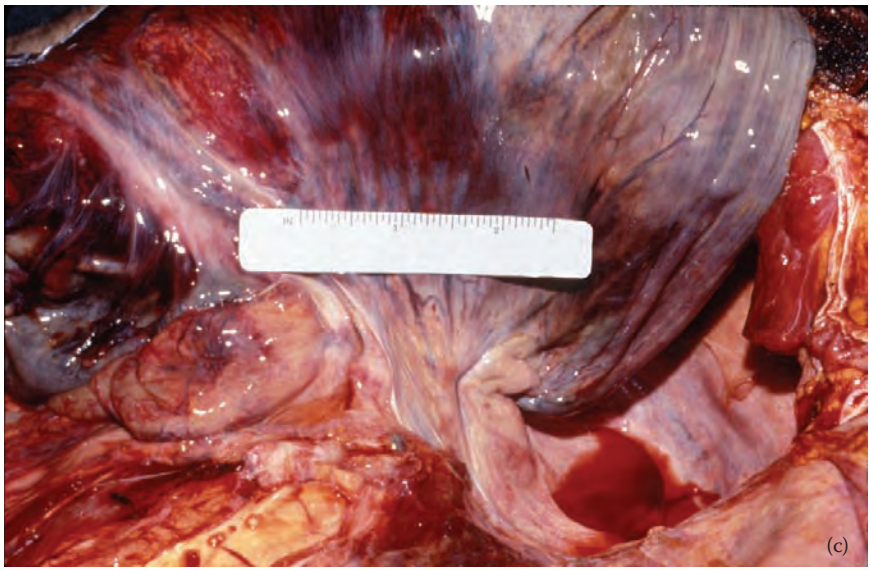

Cecal volvulus. Note the red discoloration of the serosal surface. Figure (a) demonstrates cecal enlargement due to obstruction, (b) demonstrates the cecal volvulus with rotation and obstruction, and (c) demonstrates this region untwisted.

This individual never sought medical attention for her breast carcinoma and died at home.

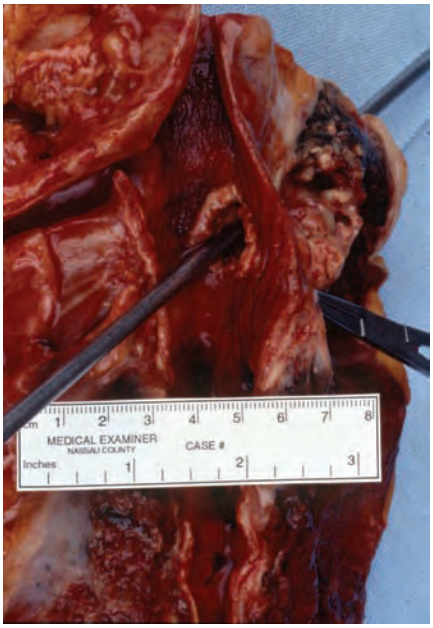

Bronchogenic carcinoma with erosion through the airway.

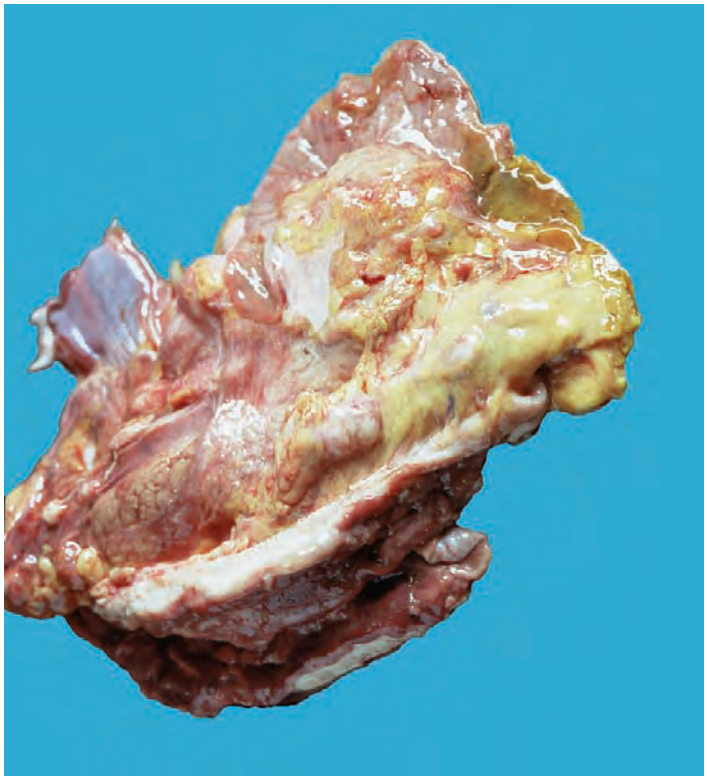

This individual never sought medical attention for his gastric cancer and died at home.

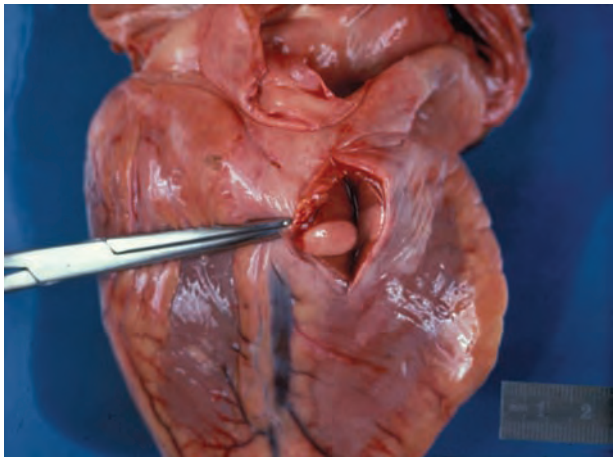

Metastatic bronchogenic carcinoma of the lung. Note the metastatic nodule in the heart muscle causing fatal arrhythmia.

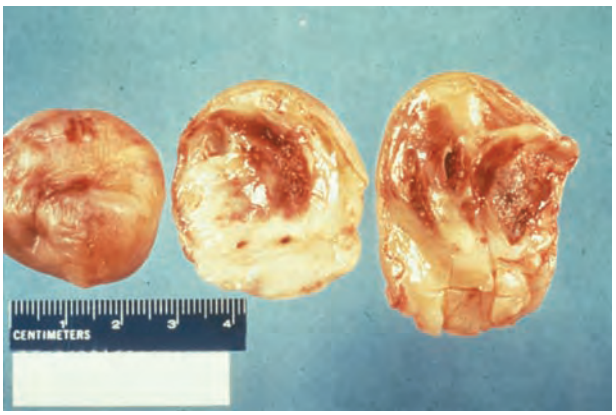

Primary rhabdomyosarcoma of the heart.

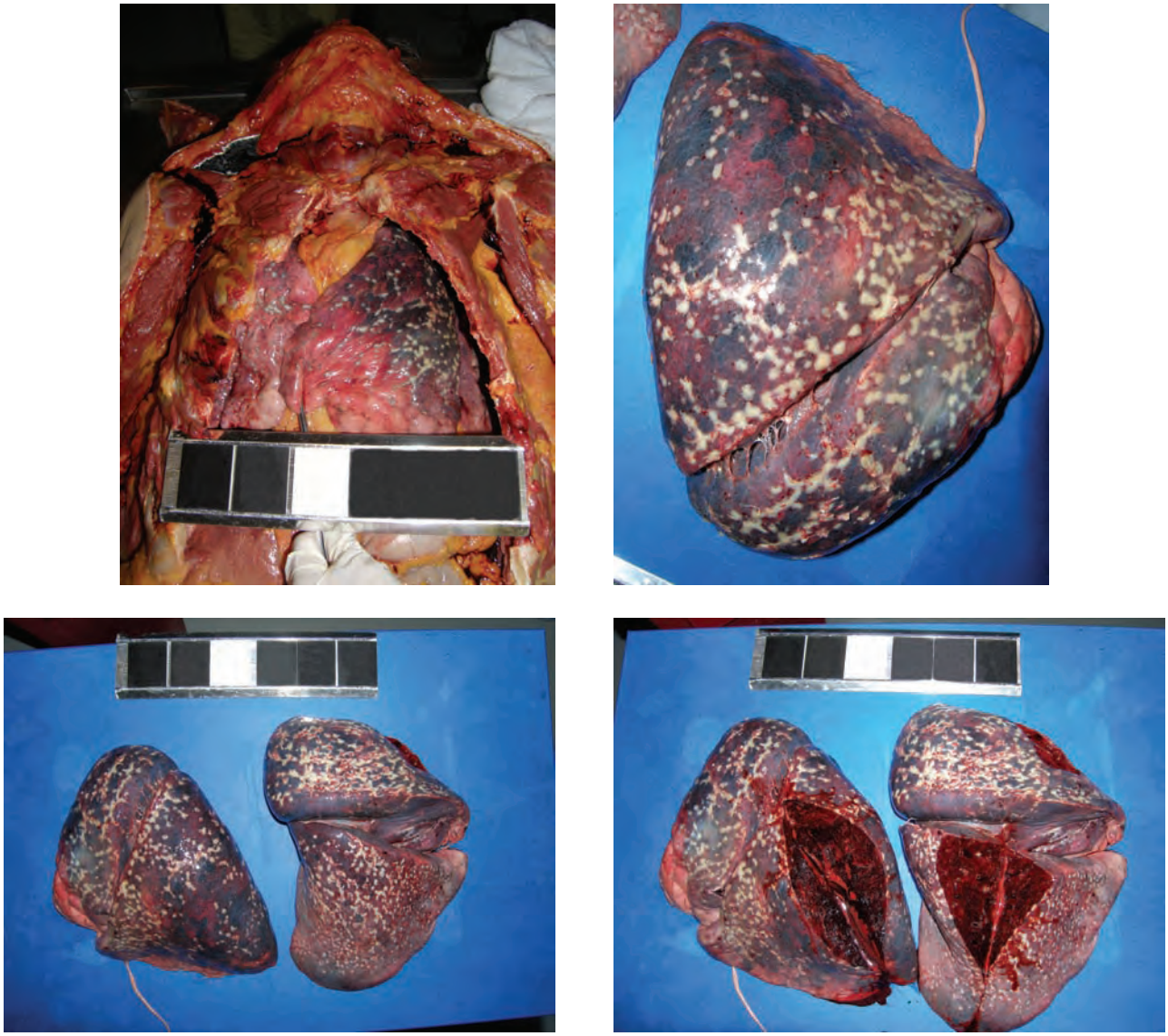

Multiple healed granulomas most likely from a past fungal infection.

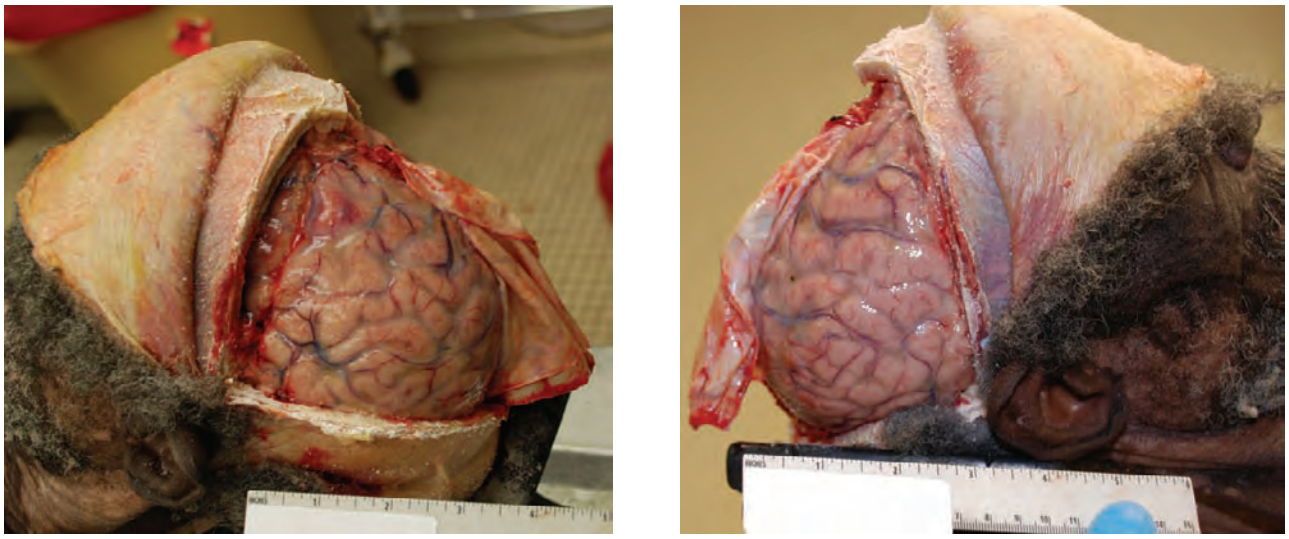

Cryptococcal meningitis in an individual with AIDS caused by intravenous drug use.

Neurofibromatosis type 1. Note the numerous neurofibromas at the surface of the body. These tumors may be found throughout the body. Death may occur by tumor growth leading to damage of adjacent structures such as the gastrointestinal track with obstruction, central nervous system with compression, renal artery with hypertension, etc.

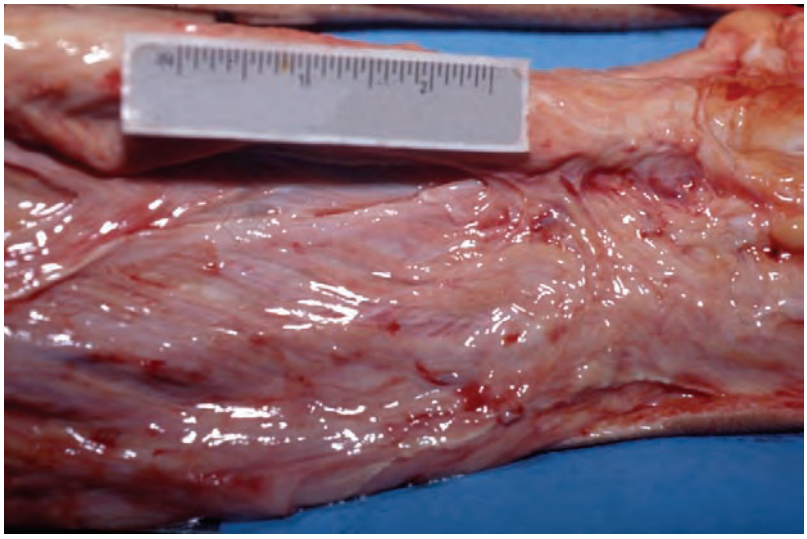

This decedent had Duchenne's muscular dystrophy and was bedridden for many years. He developed severe osteoporosis. While being removed from bed one day by a new caregiver, he sustained bilateral femur fractures and died shortly after.

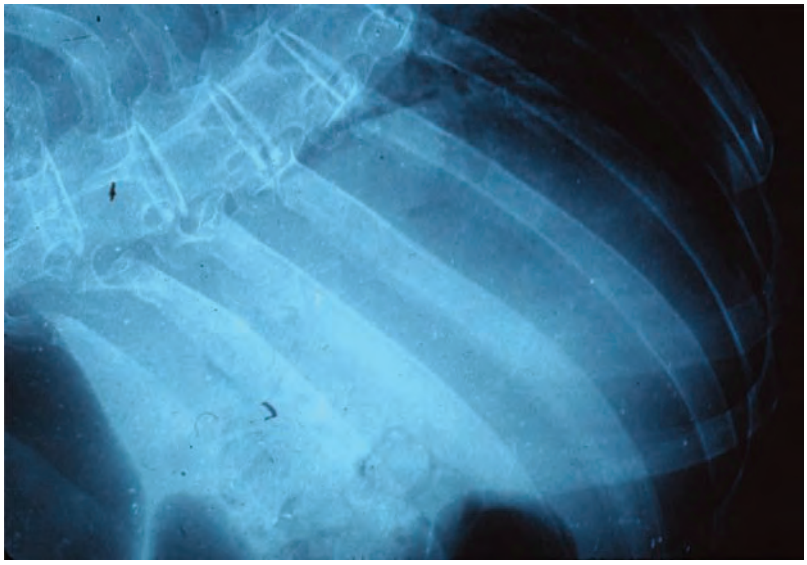

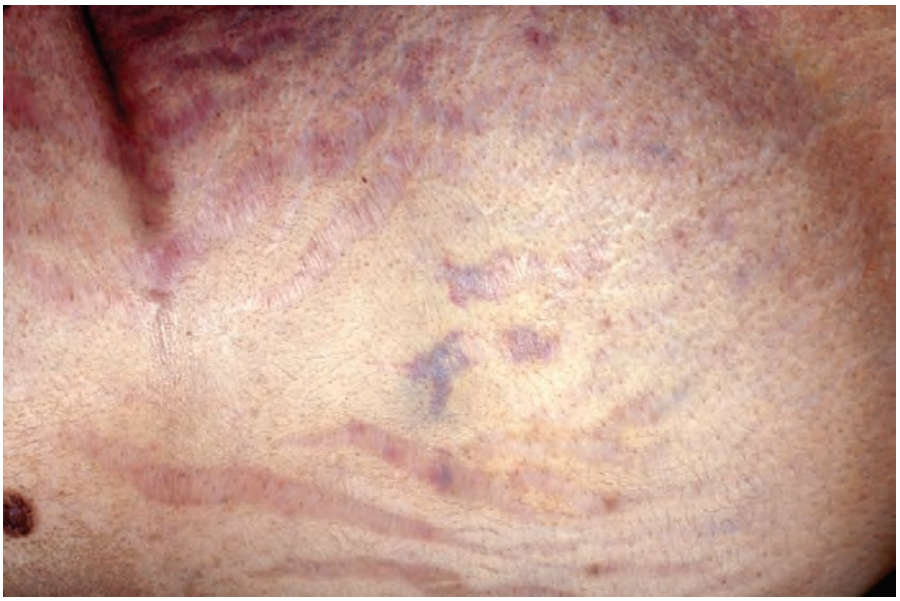

Cushing's syndrome due to hypercortisolism. Note the "moon face," "truncal obesity" and purple striates at the lower abdomen. This syndrome is also associated with glucose intolerance. From a forensic standpoint and interpreting injuries, it is important to know this is associated with osteoporosis, increased bruisability, and poor healing.

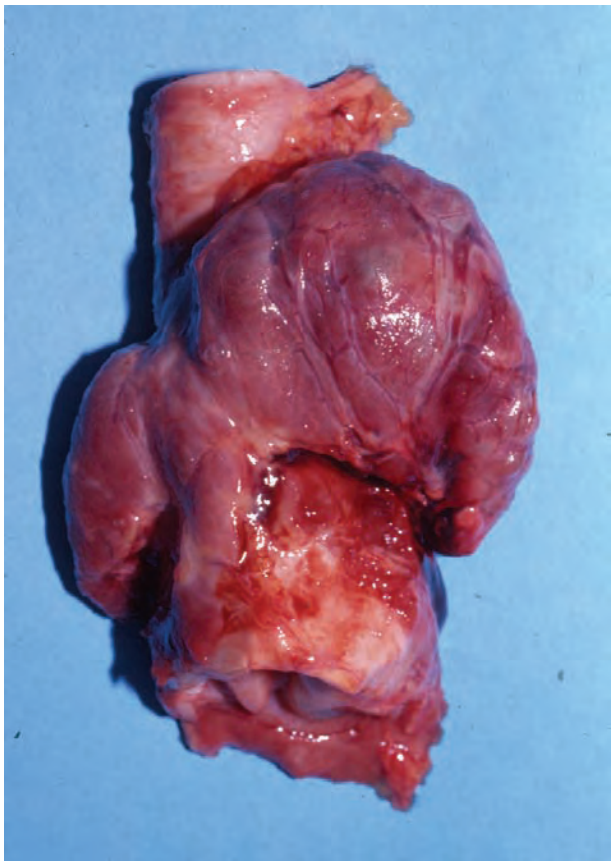

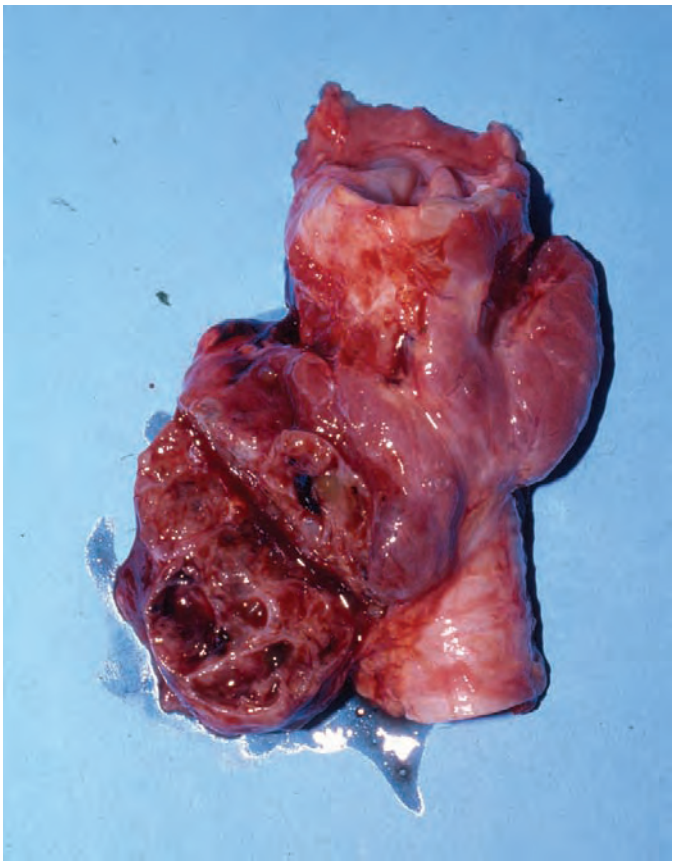

Goiter with hyperthyroidism may lead to thyrotoxicosis and sudden cardiac death due to tachyarrthymia.

This portion of gastric mucosa reveals multiple small areas of hemorrhage in an individual with hypothermia. "Leopard skinning" of the gastric mucosa and acute pancreatitis are often seen with cases of marked hypothermia followed by short-term survival.

This individual had his fingers and toes amputated years before following extensive hypothermic injury with frostbite.

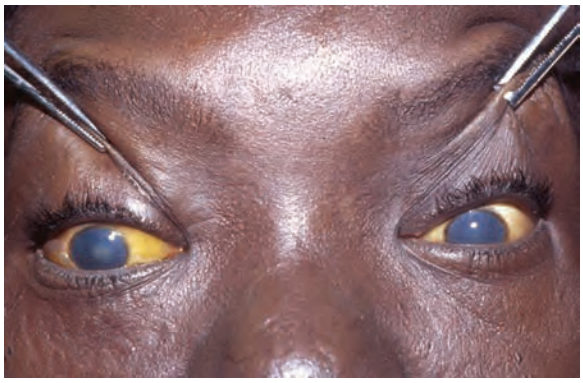

Jaundice due to hemolysis associated with sickle cell crisis.

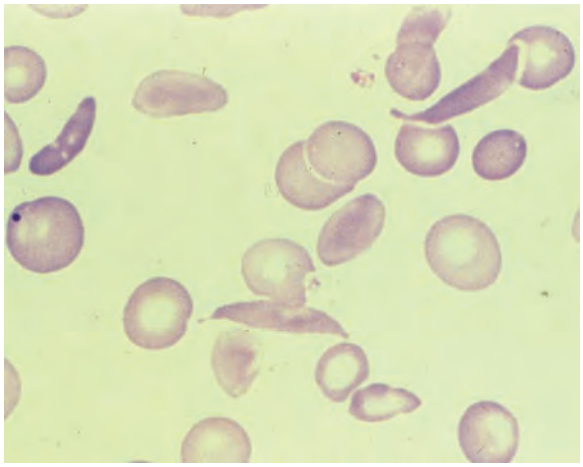

Sickle cell anemia. Peripheral blood smear showing sickled cells.

Sickle cell crisis. Note the dark red-black sludge within the gall bladder from hemolysis.

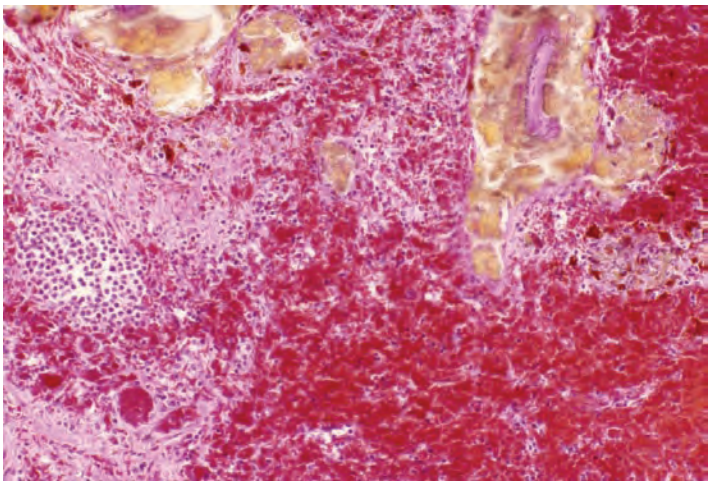

Autoinfarction of spleens from adult individuals with sickle cell anemia. These spleens are typically shrunken and firm. Note the histopathology sections with marked congestion fibrosis and Gamna-Gandy bodies from iron pigments with calcium salts.

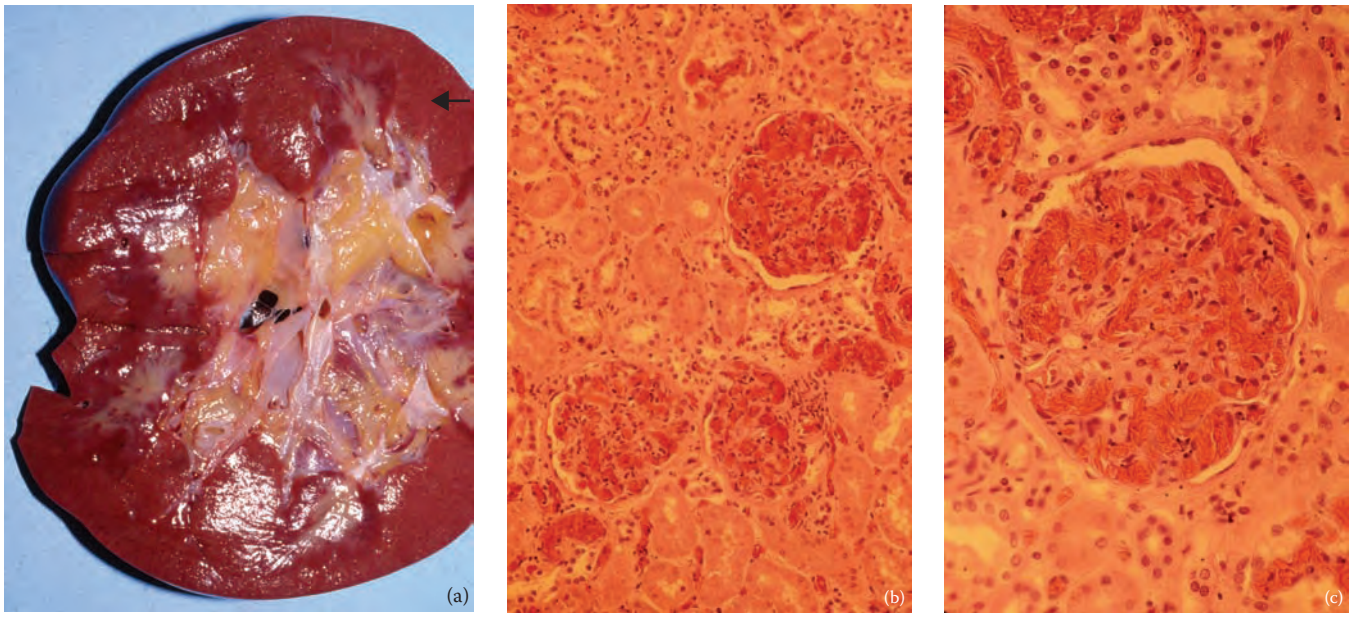

Sickle cell anemia with crisis. Kidney with papillary necrosis (a) and congestion (b) and (c).

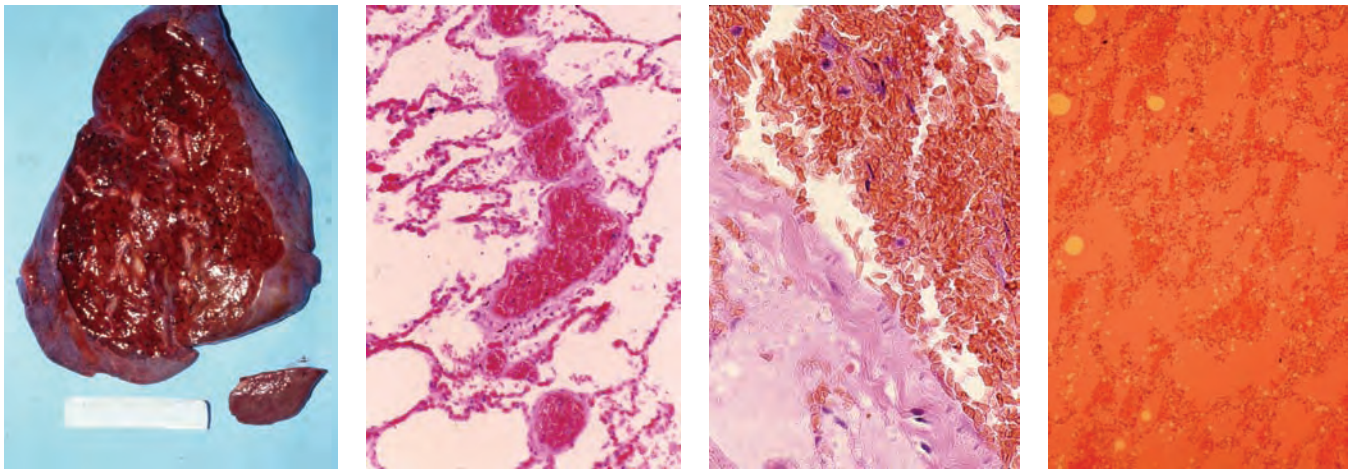

Sickle cell anemia with crisis. Lung with marked congestion and pulmonary edema.

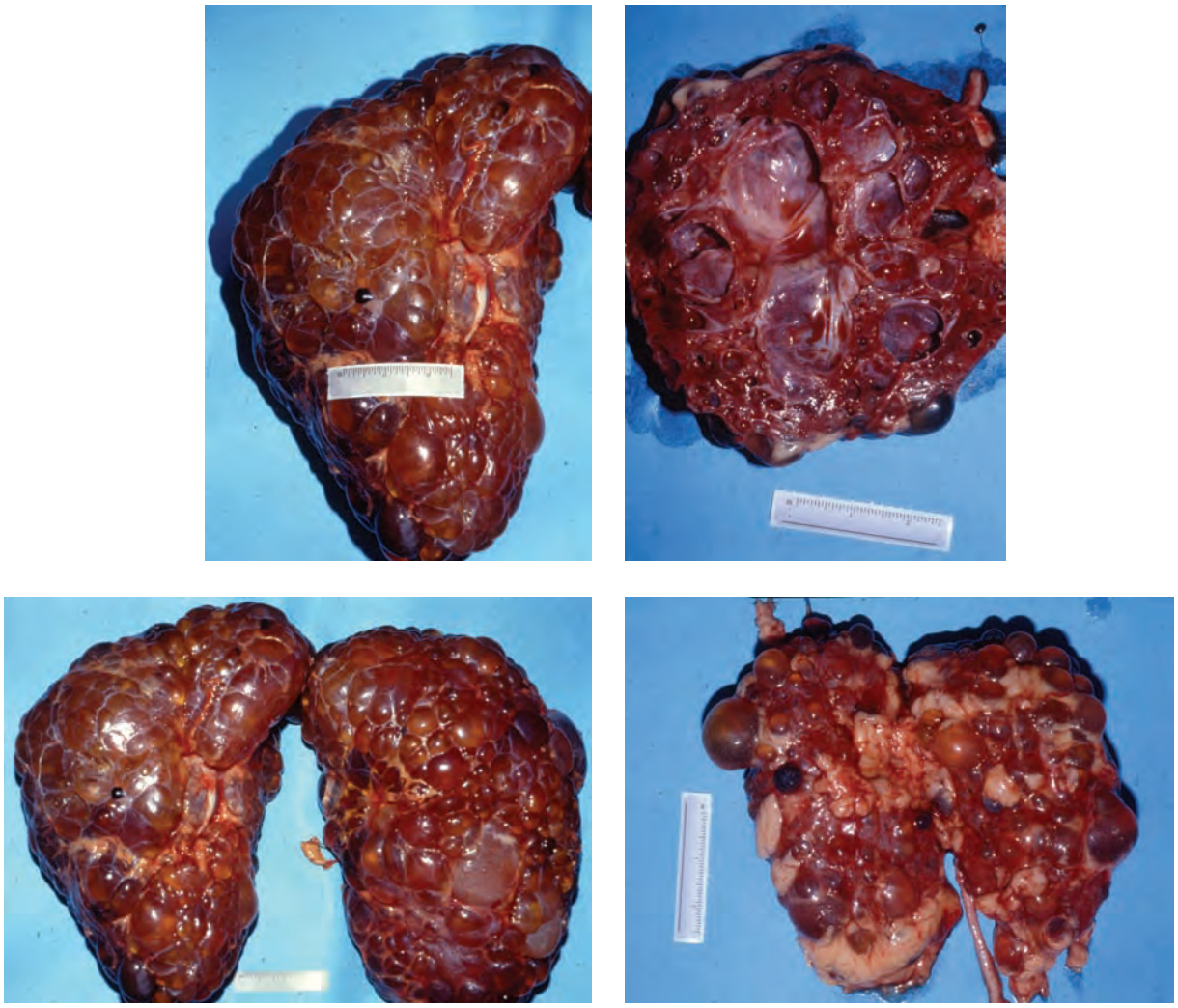

Adult polycystic kidney disease. This autosomal dominant disorder is another significant cause of hypertension. These kidneys weighed over 3000 grams each. There is also an association with cysts in other organs and intracranial berry aneurysms.

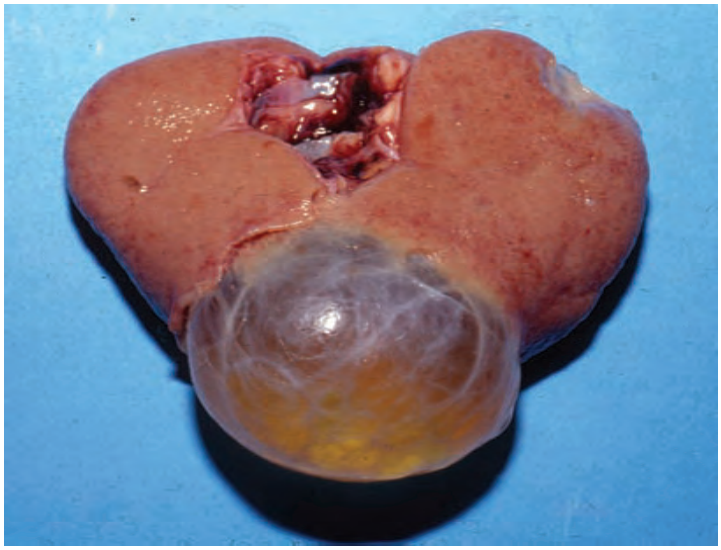

Simple benign cyst.

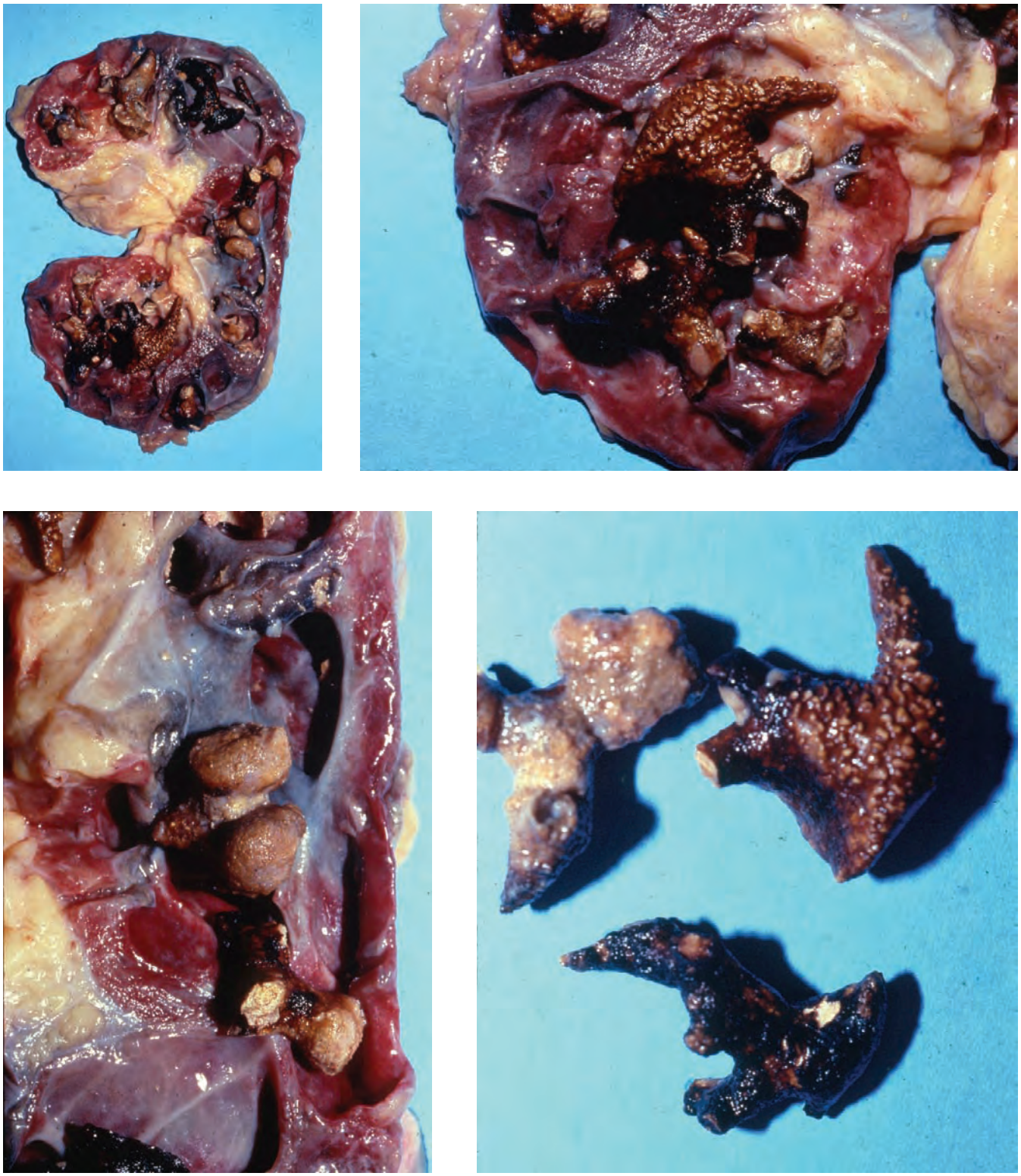

Chronic pyelonephritis with "stag horn calculi" associated with urea splitting bacterial infection such as proteus or staphylococci causing magnesium ammonium phosphate salt precipitation. It is also important to know that acute obstructive urolithiasis with urosepsis can be rapidly fatal.

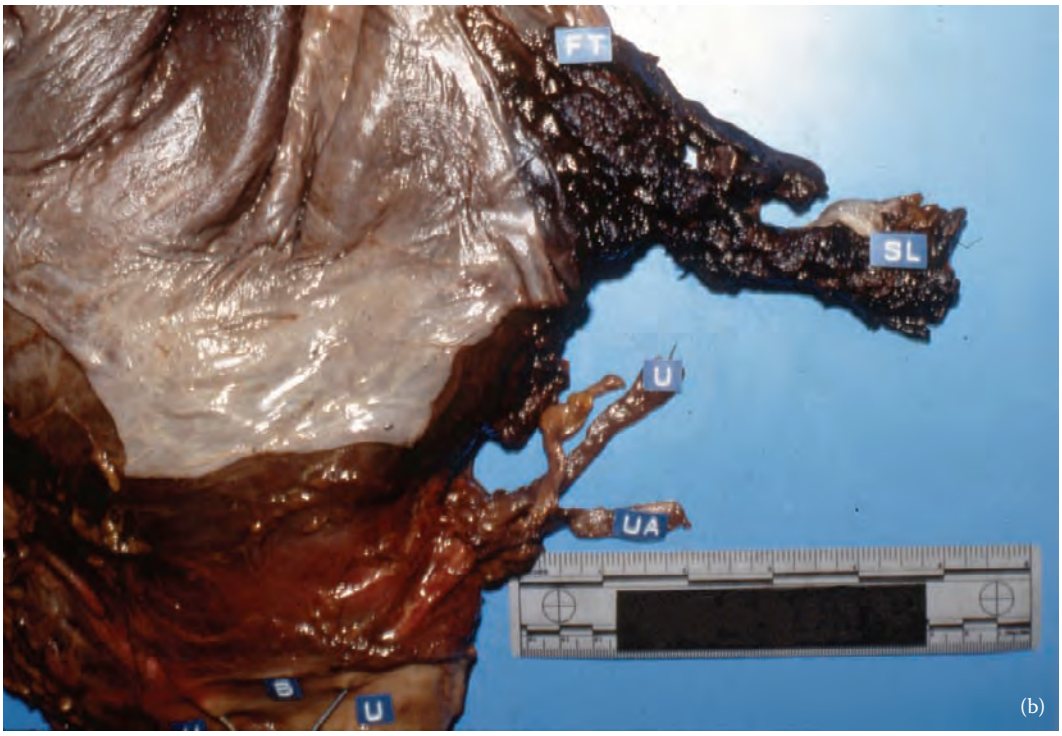

This is a postpartum death involving a 30-year-old nurse. This woman was known to be colonized by group A beta-hemolytic streptococcus (strepococcus pyogenes). Approximately 8 hours following an uneventful delivery, she began to complain of back and pelvic pain. Within 4 hours she was in full-blown shock and was refractory to resuscitative measures. Death was pronounced approximately 14 hours following her delivery. Figure (a) depicts the gravid uterus and confluent brown-black discoloration of the left adnexa, which proved microscopically to show areas of necrosis with numerous clusters of bacterial cocci and scant neutrophilic infiltrates. Postmortem cultures of multiple organs, including lungs, liver, spleen, uterus, and peritoneal fluid, all grew group A beta-hemolytic streptococcus betahemolitic streptococcus. This "toxic shock like" death due to group A streptococcus has given rise to the term "flesh-eating bacteria."

Therapeutic Interventions, Complications, and Accidents

2

MICHAEL J. CAPLAN AND CHARLES A. CATANESE

The evaluation of deaths related to complications of diagnostic and therapeutic procedures has traditionally been a problematic area for forensic pathologists and death investigators. The reasons for this are multiple: (1) these deaths often blur the boundaries dividing medicolegal and hospital death investigation; (2) many forensic pathologists do not feel comfortable approaching the often complex and highly technical issues present today in modern medicine; and (3) at least some forensic pathologists believe that these types of death belong more appropriately within the domain of the hospital autopsy. Despite this ongoing controversy, most forensic pathologists and death investigators will, at some point during their careers, be forced to handle these types of deaths.

If a patient dies either during or sometime following a diagnostic and therapeutic procedure, the first question in the diagnostic algorithm or decision tree becomes: Was the death in any way related to the procedure or was the procedure simply an artifact, with the death attributable to some underlying natural disease process or injury? For example, if a 65-year-old man who has stubbornly refused surgical intervention of a known 8-cm-diameter abdominal aortic aneurysm suddenly collapses and manages to make it to the operating room but dies before repair can be attempted, the "intraoperative death" is simply an artifact of modern medical resuscitation and transport technologies. In such a case, the medical intervention or procedure played no part in the death, and the cause of death would be worded as "ruptured abdominal aortic aneurysm due to atherosclerotic cardiovascular disease, manner: natural." On the other hand, if an otherwise healthy individual underwent an elective procedure, for example, a bunionectomy, and died on the operating table, to ascribe the death to "hallux valgus" would be as ridiculous as it was erroneous. It is for cases like this that the concept of therapeutic complication merits consideration for a place in manner-of-death certification.

Once it is determined that the death was related somehow to the procedure, the next question in the algorithm becomes: Was the death a result of a known or predictable (albeit rare or unusual) complication of a properly performed diagnostic procedure or appropriately administered therapy, or did it ensue from a procedural error? When the former scenario exists, the term *therapeutic complication* is preferred, because that is a neutral designation as opposed to the term "therapeutic misadventure," which is inflammatory and prejudicial. When the latter situation is present, the manner of death is most accurately designated *accident*. Many examples of legitimate therapeutic complications exist. In the case of a 64-year-old man who sustained a non-ST (non-Q wave) segment elevation myocardial infarct, and who was treated with the antiplatelet drug eptifibatide (Integrilin), a platelet glycoprotein IIb-IIIa antagonist, he suffered intractable pulmonary hemorrhage and ultimately succumbed to respiratory failure. Pulmonary hemorrhage has been a recently reported complication of eptifibatide therapy. No error was committed in this case; rather, the decedent experienced an unusual yet known (documented) complication of appropriate therapy for his non-ST elevation MI. Therefore, the cause of death would be worded as: "Complications of pulmonary hemorrhage following eptifibatide (Integrilin) treatment for non-ST (non-Q wave) segment elevation myocardial infarct (due to atherosclerotic heart disease)" and the manner, therapeutic complication. On the other hand, if a middle-aged woman underwent a surveillance colonoscopy and presented days later with peritoneal signs and was found to have *E. coli* bacteremia and sepsis due to peritonitis resulting from inadvertent perforation of the colon during the procedure, the death certificate would read: "*E. coli* sepsis due to peritonitis due to colonic perforation complicating surveillance colonoscopy." However, because this complication is not an accepted sequela of an appropriately performed diagnostic or therapeutic procedure, but rather, an unintentional error, the manner of death in this case is most accurately deemed accident. Admittedly, there are cases that fall within the murky zone between natural and therapeutic complication and between therapeutic complication and accident. In such cases, all that one can do is exercise one's best objective clinical judgment. It is important to recognize and remember that death certificates are not immutable documents that are "written in stone"; rather, they can be amended should more accurate information regarding the circumstances of death become available at a later time.

While the concept of therapeutic complication is an invaluable tool in the assessment of these types of deaths, the practical applicability of the term to the death certificate has enjoyed less success. Only two known jurisdictions—Cuyahoga County, Ohio, and New York City—include "therapeutic complication" as a choice in the manner of death on the death certificate. Therefore, when a death fits the criteria for therapeutic complication as the manner, but the particular jurisdiction does not include it on the death certificate, then the manner defaults to natural. One other option is to list "therapeutic complication" in parentheses after "natural," but this probably is not common practice. Thus, for the indefinite future, forensic pathologists will be resigned to using the term as a conceptual tool in the evaluation of these most challenging deaths.

These figures show one of many possible artifacts associated with organ donation that may be mistaken for injuries. Note (a) the hemorrhagic discoloration around each eye (periorbital) and (c) the plastic insert in the opened eye following tissue removal. Medical record review and clinician interview revealed that these periorbital hemorrhages were not present at the time of admission.

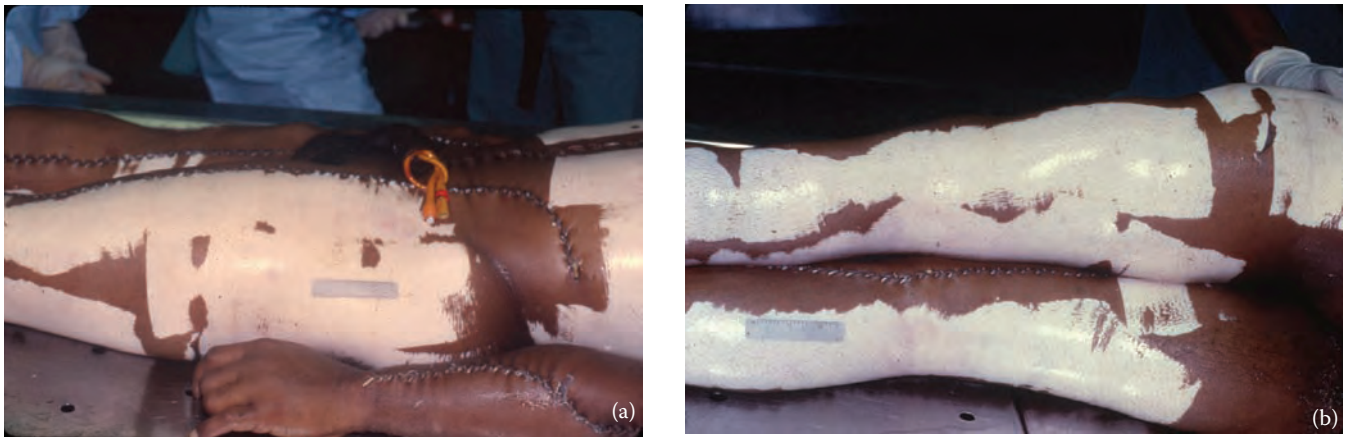

Tissue procurement of superficial skin.

Postmortem tissue procurement and full thickness skin.

Sutured incisions following long bone, soft tissue, and saphenous veins removal for postmortem tissue procurement.

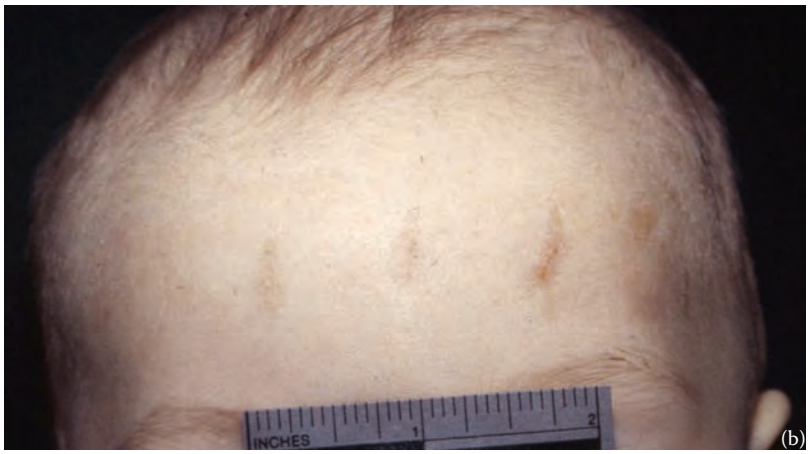

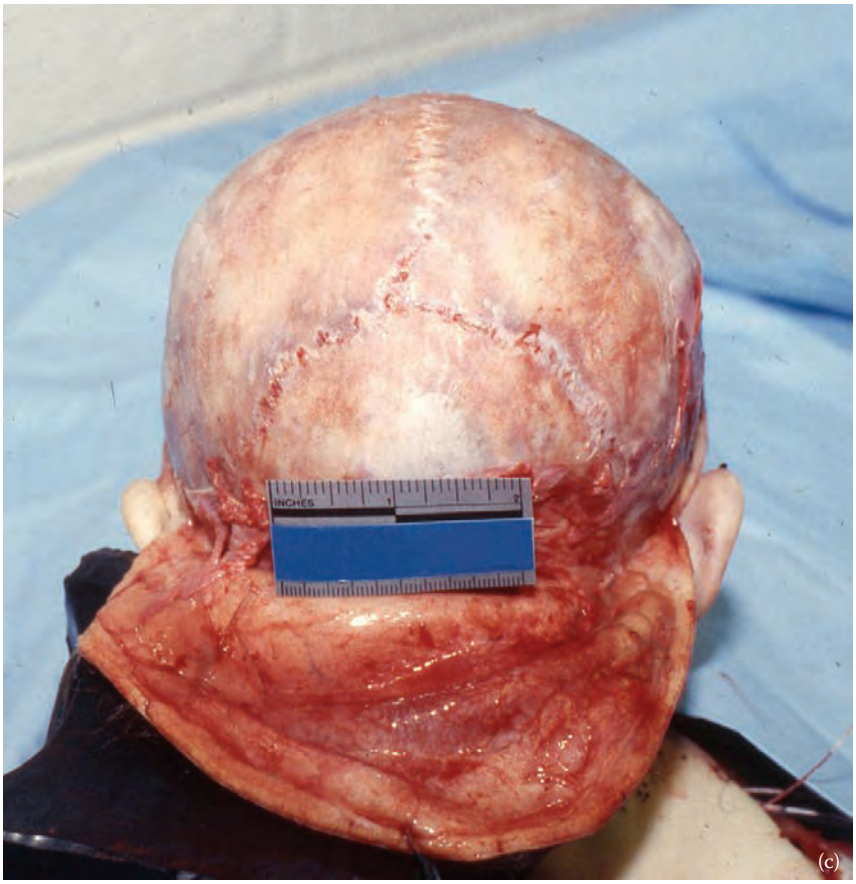

Figures (a) and (b) show ill-defined, parallel, vertically oriented gray-brown marks on the forehead. These are artifacts of transport. The child's head was stabilized with a strap while flown by helicopter from an outside hospital to a tertiary pediatric center. Figure (c) shows reflected scalp and well-defined sagittal and lamboidal sutures without features of subgaleal contusions or skull fractures.

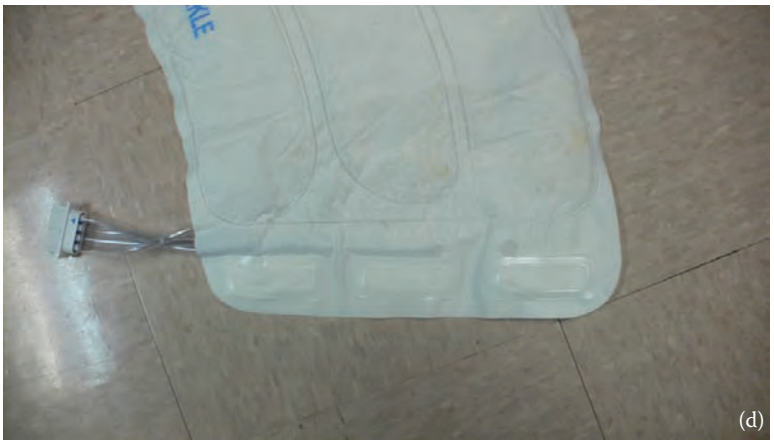

This individual was ejected from a car during a collision. She was taken to the hospital in an unconscious state and remained so until her death. There was initial speculation that she was struck by another motor vehicle due to the "tire-like" patterns on her legs (a–c). Further investigation revealed these injuries were not present at the time of arrival to the emergency department. She survived in the hospital for approximately 2 days. There was a pelvic fracture with blood seeping into the legs, extensive generalized edema, and disseminated intravascular coagulation (b). Pressure boots were placed to decrease the risk of deep venous thrombosis (d). This pattern of ecchymosis was caused by the pressure boots in conjunction with the complications of the injuries.

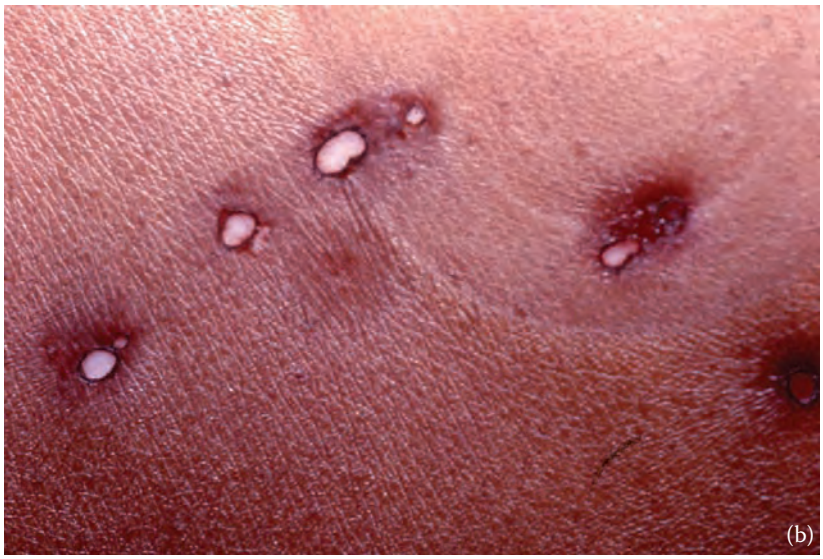

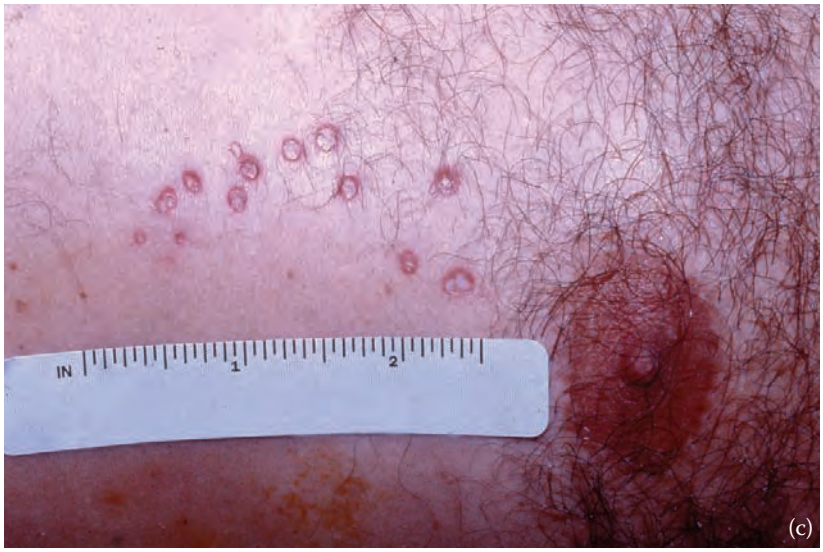

Figures (a), (b), and (c) show electrical burns caused by defibrillation with damaged cardioversion paddles during attempted resuscitation.

These figures show perimortem injuries to the anterior chest, axillary regions, and the anterior aspects of the upper arms associated with incorrect placement of a rib splitter during a thoracotomy.

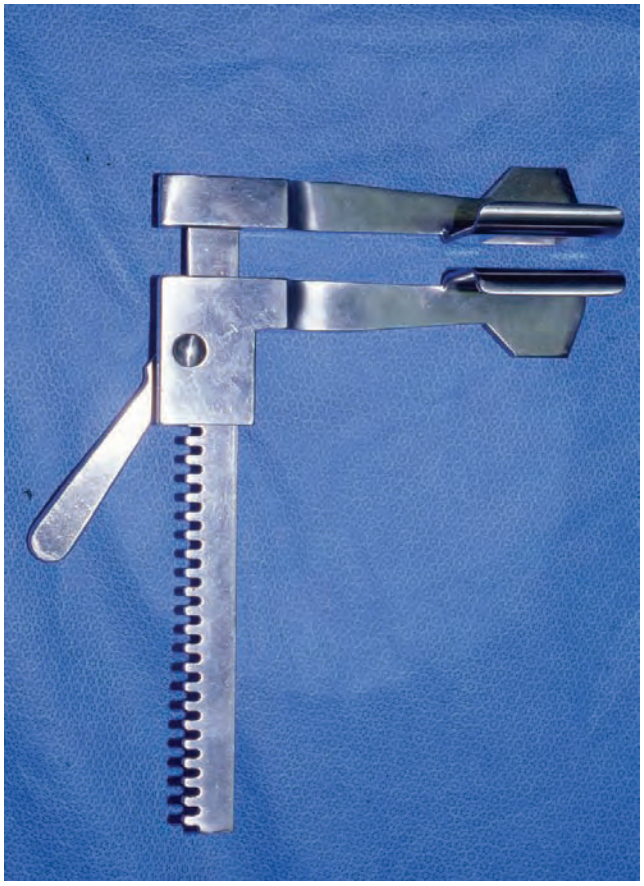

These figures demonstrate the posterior pharyngeal/esophageal cut surface with intramural hemorrhage. This is an artifact encountered frequently during resuscitative measures. The rich esophageal venous plexus tends to be a very hemorrhagic area following intubation.

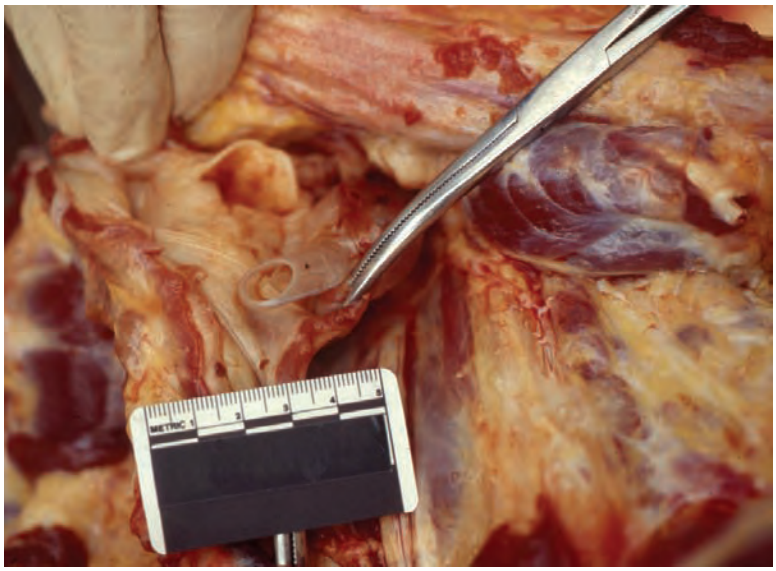

This demonstrates an esophageal intubation with the tip of the endotracheal tube protruding from the esophageal lumen. The epiglottis can be seen slightly behind the endotracheal tube, confirming an esophageal intubation.

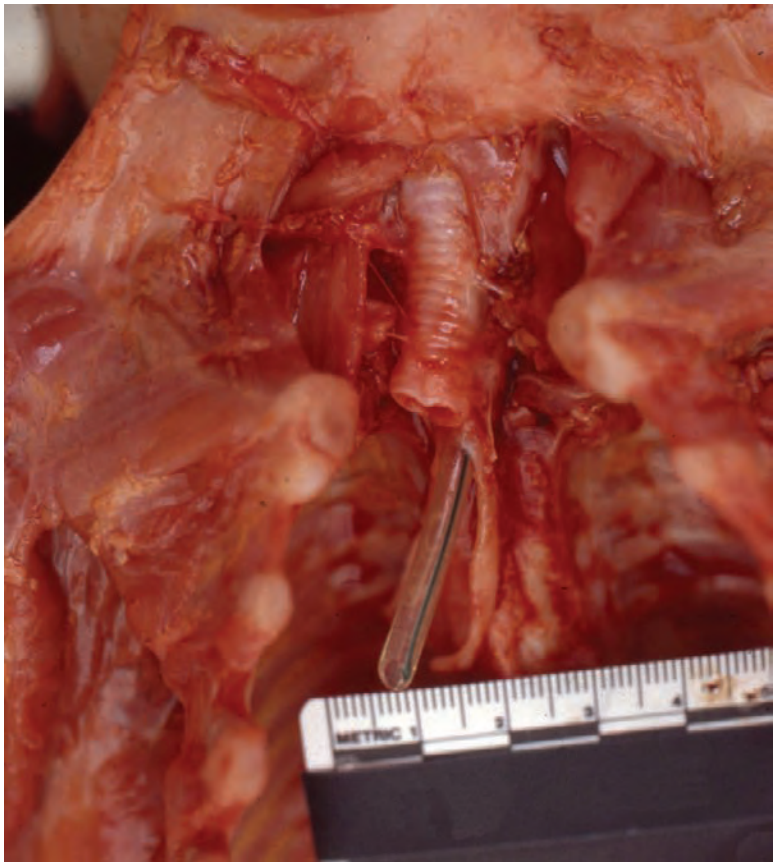

This autopsy involved an infant who succumbed to sudden infant death syndrome (SIDS). The close-up photograph depicts an esophageal intubation, with the endotracheal tube clearly within the esophageal lumen and the concentric tracheal rings visible slightly anterior to and (anatomically) to the right of the tube. Esophageal intubation is not an uncommon finding associated with resuscitation and usually plays no substantial contributory role in the death; however, it should always be documented as part of a complete autopsy and may, in some situations, be a quality control measure for Emergency Medical Services (EMS) personnel and paramedics. Esophageal intubations during an elective procedure, on the other hand, are very important and potentially causal or contributory to the death.

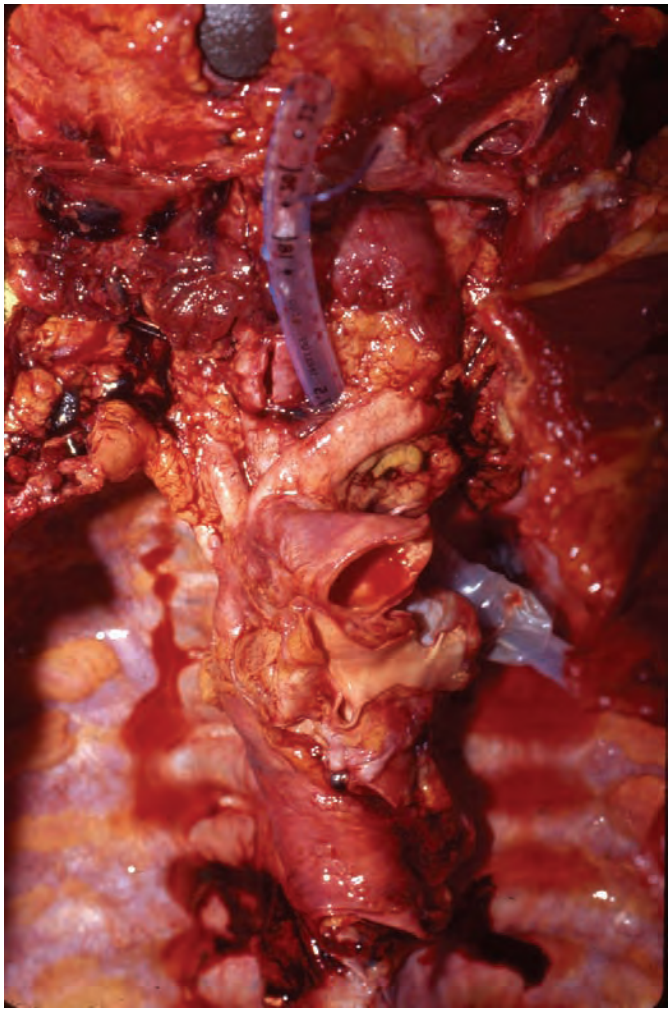

This endotracheal tube was inserted through the trachea and paratracheal soft tissues into the thoracic cavity during resuscitation.

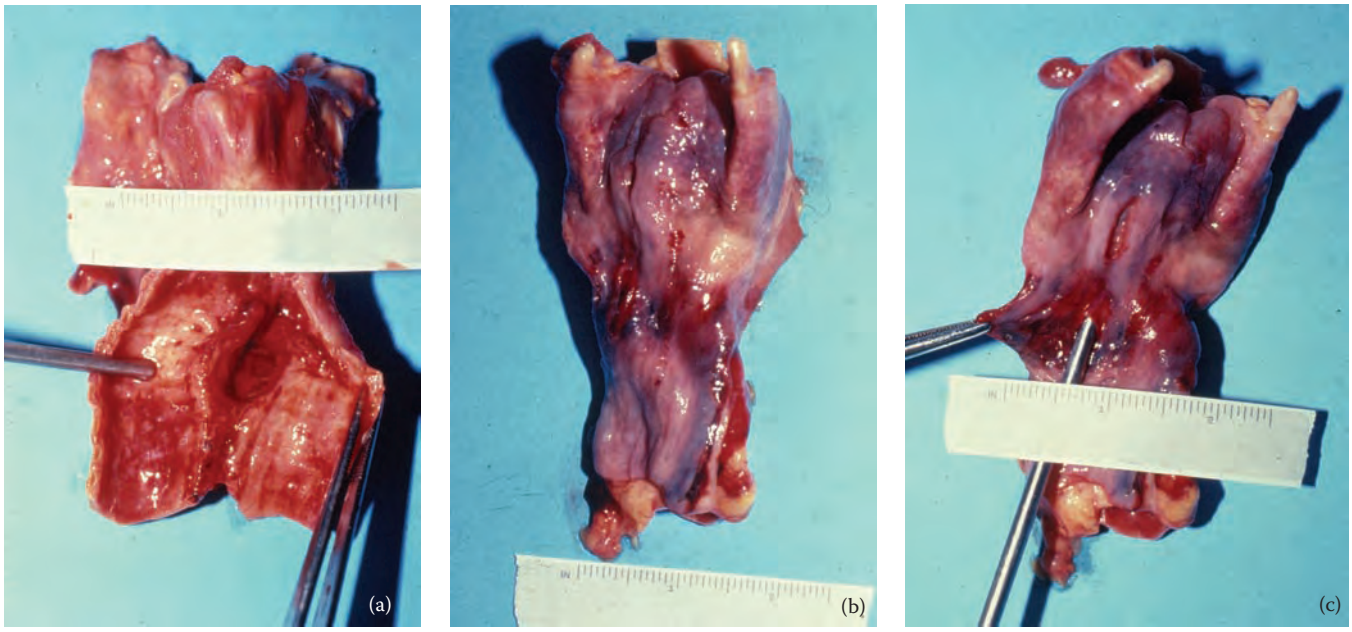

A tracheostomy tube was inserted through the posterior aspect of the trachea (a) into the esophagus (b) (traumatic tracheoesophageal fistula). Note the perforations on the posterior surface of the trachea (a) and the anterior aspect of the esophagus (b).

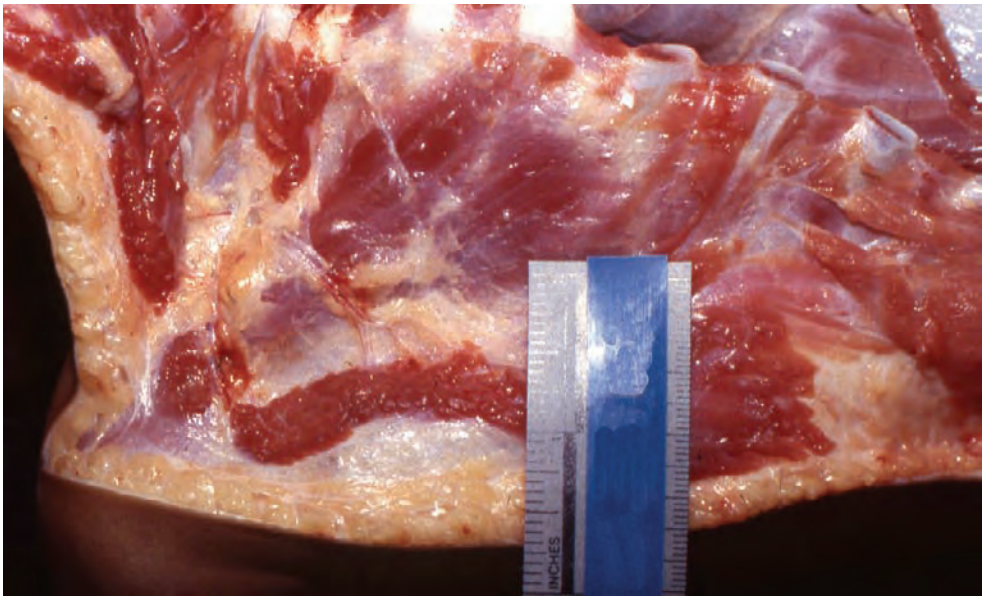

This is a view of the chest wall of an 18-month-old toddler with a history of tracheomalacia and subaortic stenosis who required a tracheostomy. During a bout of crying, the tracheostomy tube became dislodged, and after nurses attempted to reposition it, the baby rapidly developed subcutaneous emphysema, followed by bilateral tension pneumothoraces. Autopsy demonstrated marked subcutaneous emphysema, including periorbital swelling. The tube had been removed before the autopsy, precluding assessment of its placement. This particular view demonstrates air bubbles within the subcutaneous fat. It is important that all tubes remain in the body for objective postmortem (autopsy) evaluation of their placement.

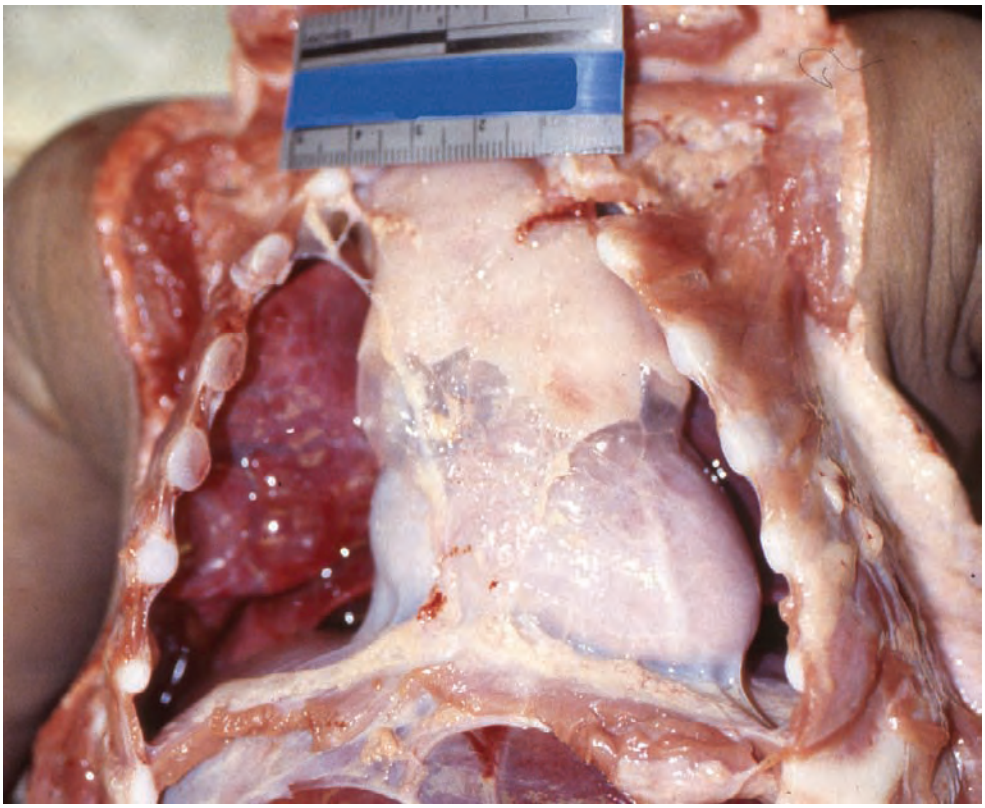

Subcutaneous emphysema. Anterior chest with air bubbles throughout the mediastinal soft tissues. This may be commonly observed with vigorous resuscitation or trauma. Palpation of these regions often reveals crepitus.

One must be careful while examining injuries altered by therapeutic intervention. This picture demonstrates a catheter placed into an injured blood vessel through a stab wound.

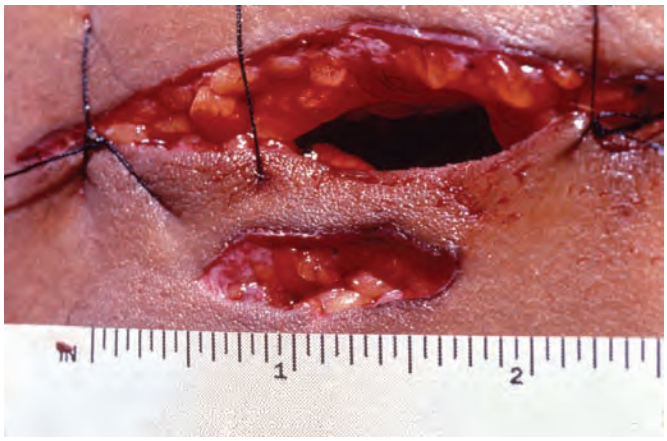

This demonstrates a stab wound adjacent to a sutured chest tube incision. This may make injury interpretation more challenging.

Healing infected incision. This individual died of septic complications of his initial trauma. Had it not been for this trauma, the infection that took his life would not have occurred. The manner of death in this case was ruled accidental.

This demonstrates a stab wound that the clinicians converted along one side (edge) into a thoracotomy incision. If at all possible, this should never be done, because it makes injury interpretation much more difficult.

This is a "cutdown site" used for vascular access by a hospital emergency department (ED) physician, created by a transverse incision within the antecubital fossa This decedent had also sustained stab and incised wounds in other parts of the body. Again, this illustrates the importance of having a complete medical record and discussions with the treating ED physicians or trauma surgeons to properly differentiate between therapeutic interventions and injuries sustained before such therapeutic interventions.

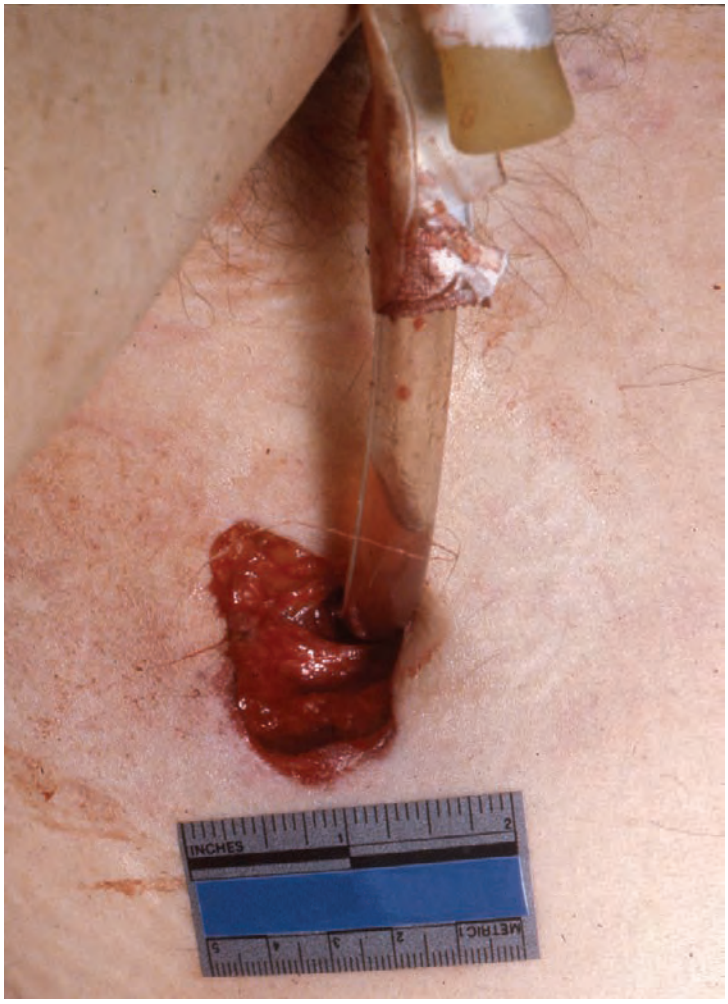

This chest tube was inserted through a previously existing perforation. Notice, however, that along the left margin of this perforation (from the 7–11 o'clock positions), there is a distinct abrasion. Further investigation established that the chest tube was inserted through a previously sustained entrance gunshot wound, thus explaining the abraded margin of this perforation.

Arteriovenous hemodialysis grafts with healing needle puncture marks.

These individuals exsanguinated in their residences following ruptures of their infected arteriovenous dialysis grafts.

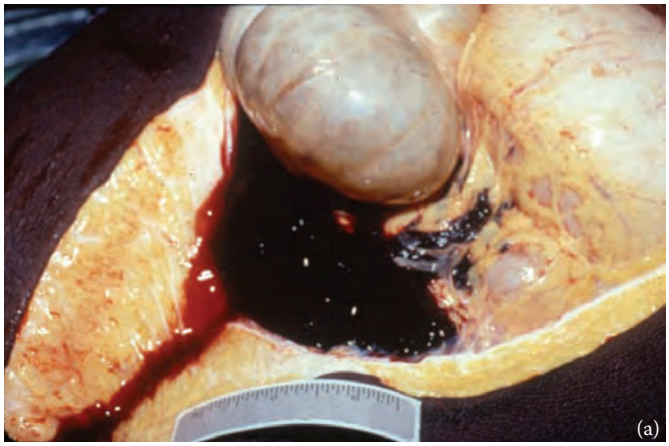

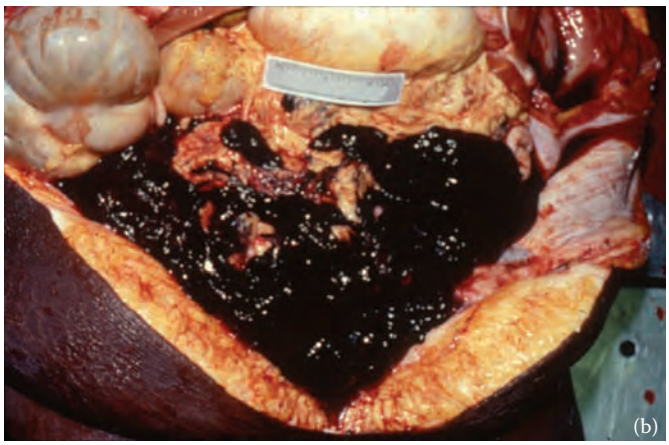

This woman underwent (a) Cesarean section (c) for an otherwise unremarkable term pregnancy hours before going into hemorrhagic shock while in the recovery room. Autopsy disclosed a large hemoperitoneum with clotted blood extending from a bleeding abdominal wall vein (a) and (b).

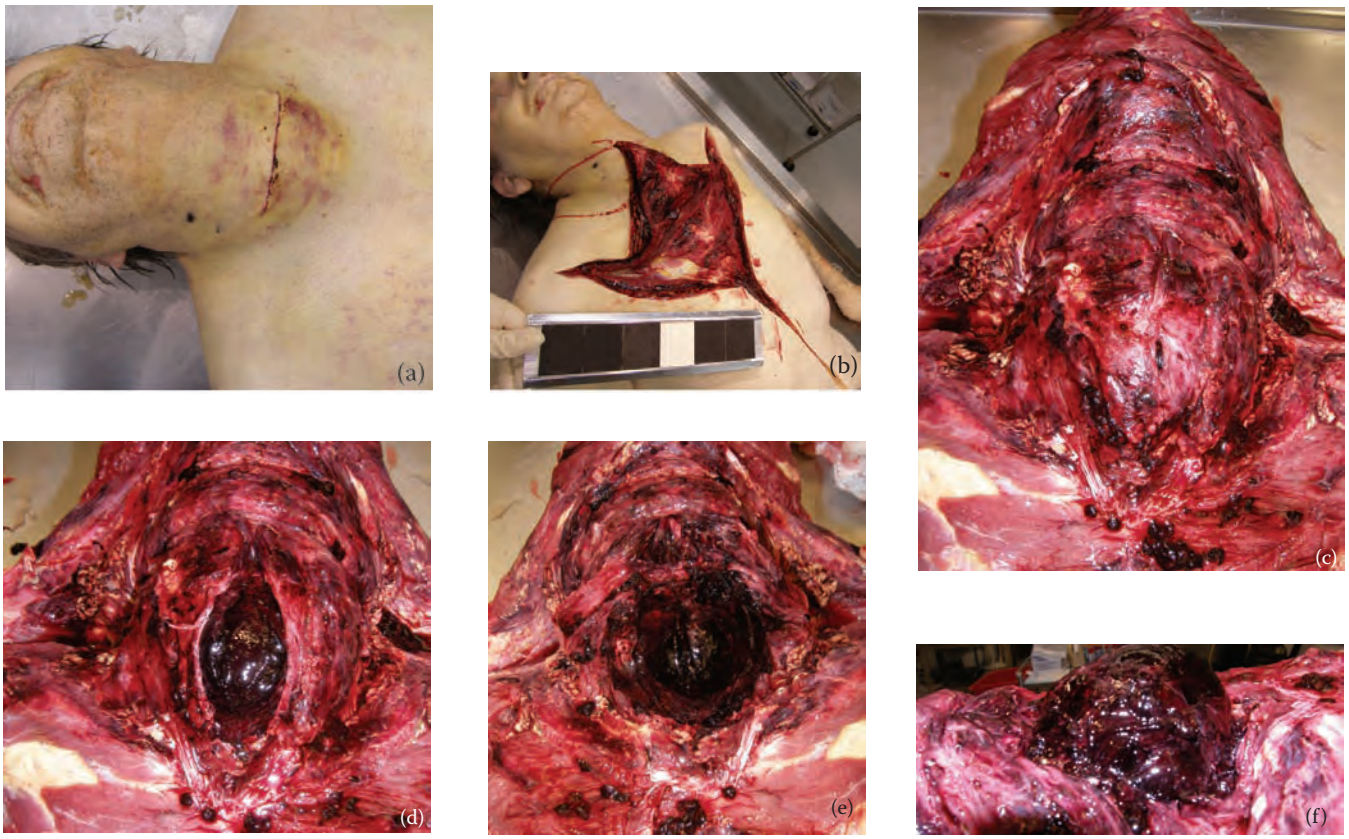

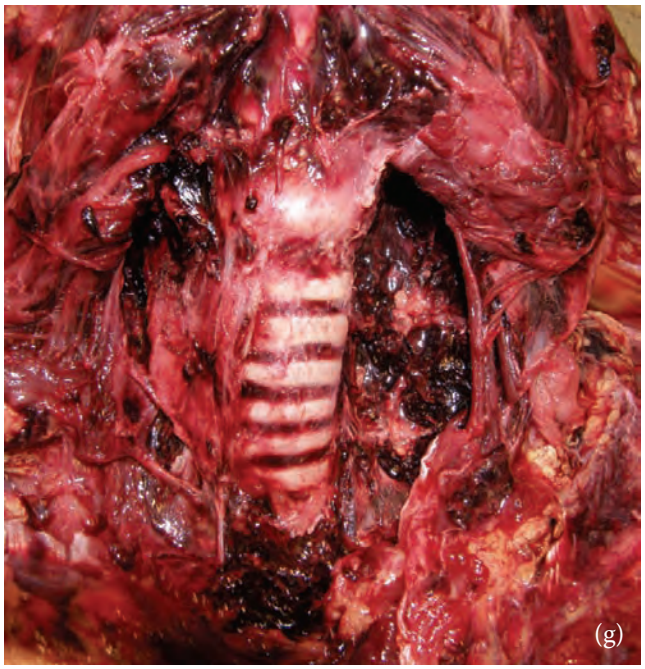

These figures show an individual who died of hemorrhagic complications following (a) a thyroidectomy. A large accumulation of clotted blood collapsed his airway while he was sedated at home. He was found unresponsive lying on a sofa. Note the hemmorhage and larger blood clot in the anterior neck and chest.

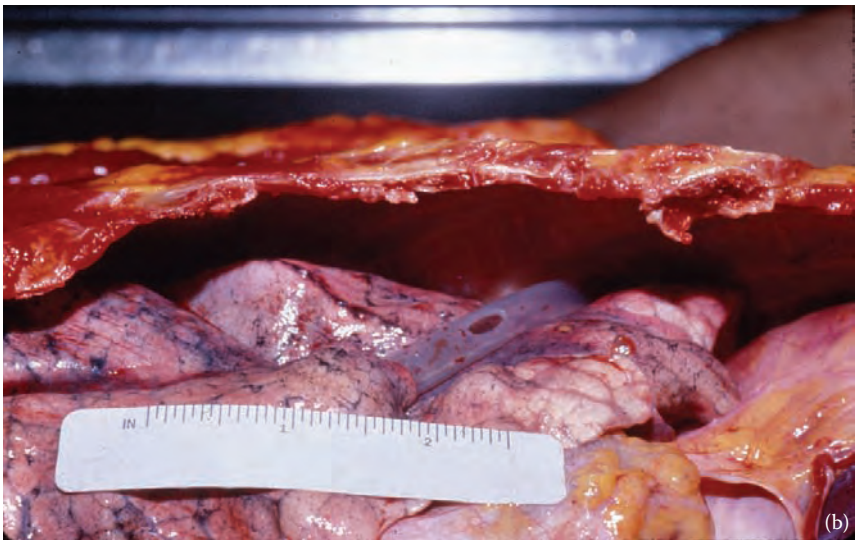

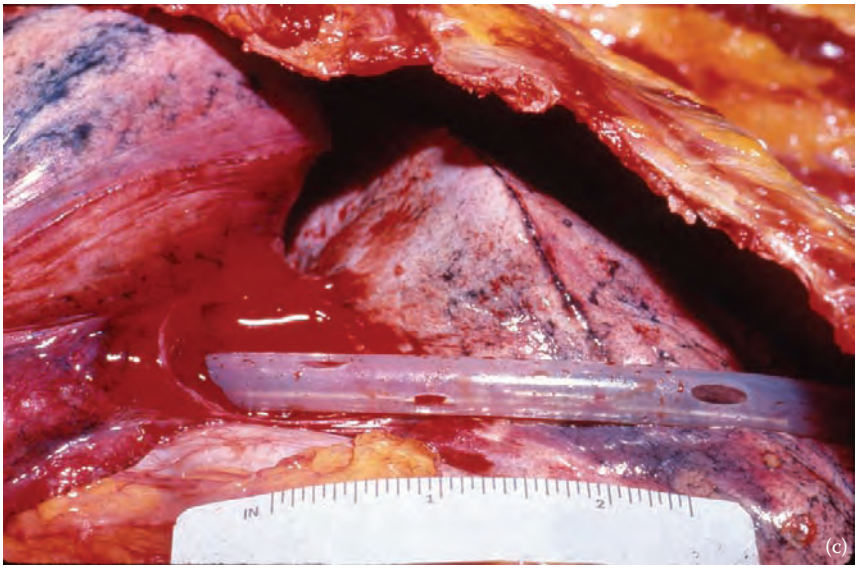

These figures show a patient who was admitted in cardiac arrest following an acute asthma attack. Chest tube placement was inserted through the lung parenchyma during resuscitation. Approximately 150 mL of liquid blood were recovered from the left hemithorax, indicating that this injury was perimortem and iatrogenic.

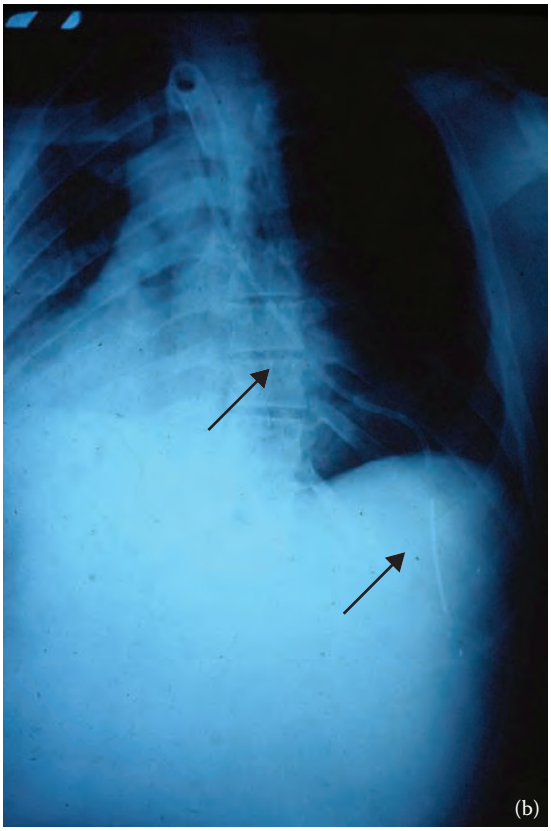

This individual underwent placement of a nasogastric tube that perforated the esophagus and entered the right thoracic cavity (hemithorax). The injury went unnoticed for many hours. The decedent was fed through the nasogastric tube, which is demonstrated by the accumulation of yellow fluid within the thoracic cavity. Autopsy revealed a fibrinous pleuritis and an early pneumonia.

Percutaneous endoscopic gastrostomy feeding tube.

Panoramic view illustrating the anterior gastric wall, with a probe through the previous PEG site, also demarcated by an arrow, and the other arrow demonstrating greater omentum with fibrinous exudate. Similarly appearing exudate can also be seen on the anterior surface of the gastric body and fundus.

This is a close-up anterior view of the stomach, demonstrating the previous PEG site through the anterior gastric wall, as well fibrinous exudate on the surface of the greater omentum, with both features demarcated by arrows.

This depicts the parietal peritoneum with a gray, shaggy, fibrinous exudate.

This case involved an elderly woman with a history of dementia who could no longer live independently and was made a ward of the state. She was hospitalized for long-term intermediate care. She had undergone percutaneous endoscopic gastrostomy (PEG) tube insertion due to insufficient oral intake (a). She underwent a tube replacement for leakage. Later that evening, she experienced acute onset of abdominal pain that was complicated by increasing abdominal distension, green discoloration of the abdominal wall site, and purulent drainage from the tube site. The remainder of her course was characterized by respiratory distress, aspiration of gastric contents, hypotension, and tachypnea, culminating in systemic inflammatory response syndrome with lactic acidosis, multisystem organ failure, and hemodynamic instability, refractory to aggressive resuscitative measures. Autopsy revealed intraperitoneal placement of the gastrostomy tube with intraperitoneal administration of enteral feedings and a collection of approximately 700 mL of turbid peritoneal fluid with extensive fibrinous adhesions. Microscopic examination confirmed an exuberant fibrinous exudate with scattered clusters of bacterial cocci and yeast, as well as birefringent material. The cause of death was certified as complications of peritonitis following intraperitoneal placement of gastrostomy tube during tube change, following percutaneous endoscopic gastrostomy insertion for insufficient oral intake related to atherosclerotic cerebral vascular disease and dementia. The manner of death was certified as accident.

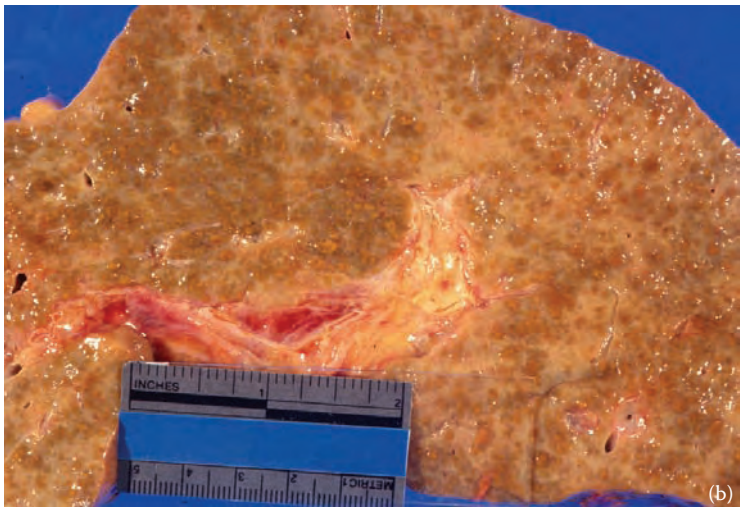

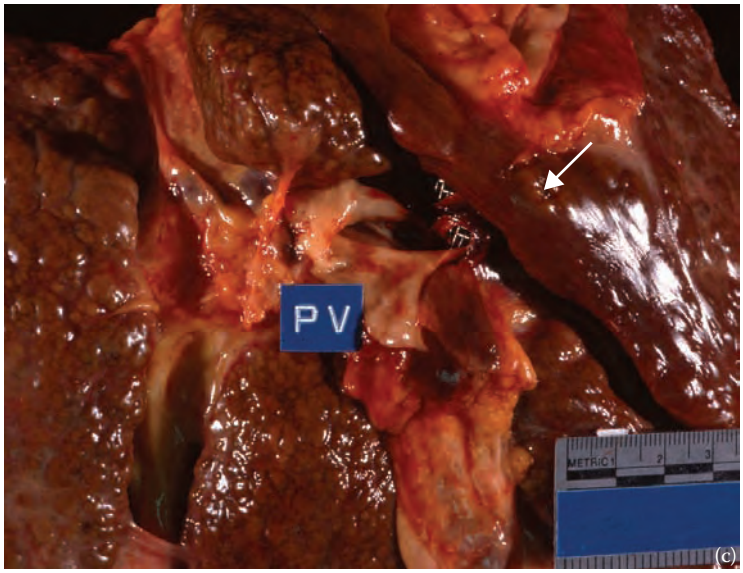

This was a middle-aged woman with primary biliary cirrhosis who underwent a transjugular intrahepatic portosystemic shunt (TIPS) for decompression of high portal venous pressure (portal hypertension). The picture (a) depicts a large hemoperitoneum, with greater omentum, stomach, and intestines floating on top of a pool of blood. There was advanced end-stage cirrhosis (b), with confluent scar enveloping and entrapping regenerative parenchymal nodules. Liver diseases such as this are associated with an increased risk of hemorrhagic complications due to coagulopathy and portal hypertension. This fatal hemorrhage resulted from laceration of a portal vein branch occurring during stent placement (c).

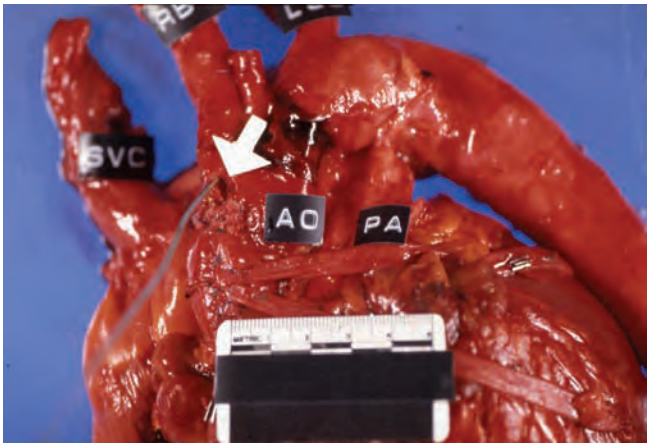

Note the probe inserted through the perforation site, which is also demonstrated with the arrow.

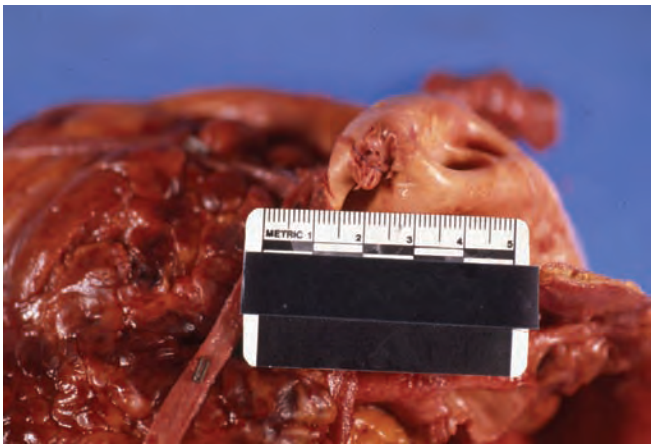

This image demonstrates the sutured cannulation site as well as one of the bypass grafts.

Close-up of perforation site demonstrated by probe and arrow.

This case involved a patient who underwent six-vessel coronary artery bypass grafting that was followed by intractable postoperative hemorrhage. The pericardial and left pleural drainage exceeded 1 liter within the first postoperative hour. The patient became hypotensive, requiring emergent re-exploration that was refractory to resuscitative efforts. Autopsy revealed an approximately 3-mm perforation of the ascending aorta, immediately above and distal to a sutured aortic cardiopulmonary bypass cannulation site, with residual bilateral hemothoraces. (left: 1,100 mL and right: 450 mL). The cause of death was certified as intractable hemorrhage with perforation at the aortic cannulation site complicating coronary artery bypass grafting for atherosclerotic coronary artery disease. Hypertensive heart disease was listed as a contributory cause (heart weight: 690 grams and body weight: 120 kgs). The manner of death was certified as therapeutic complication.

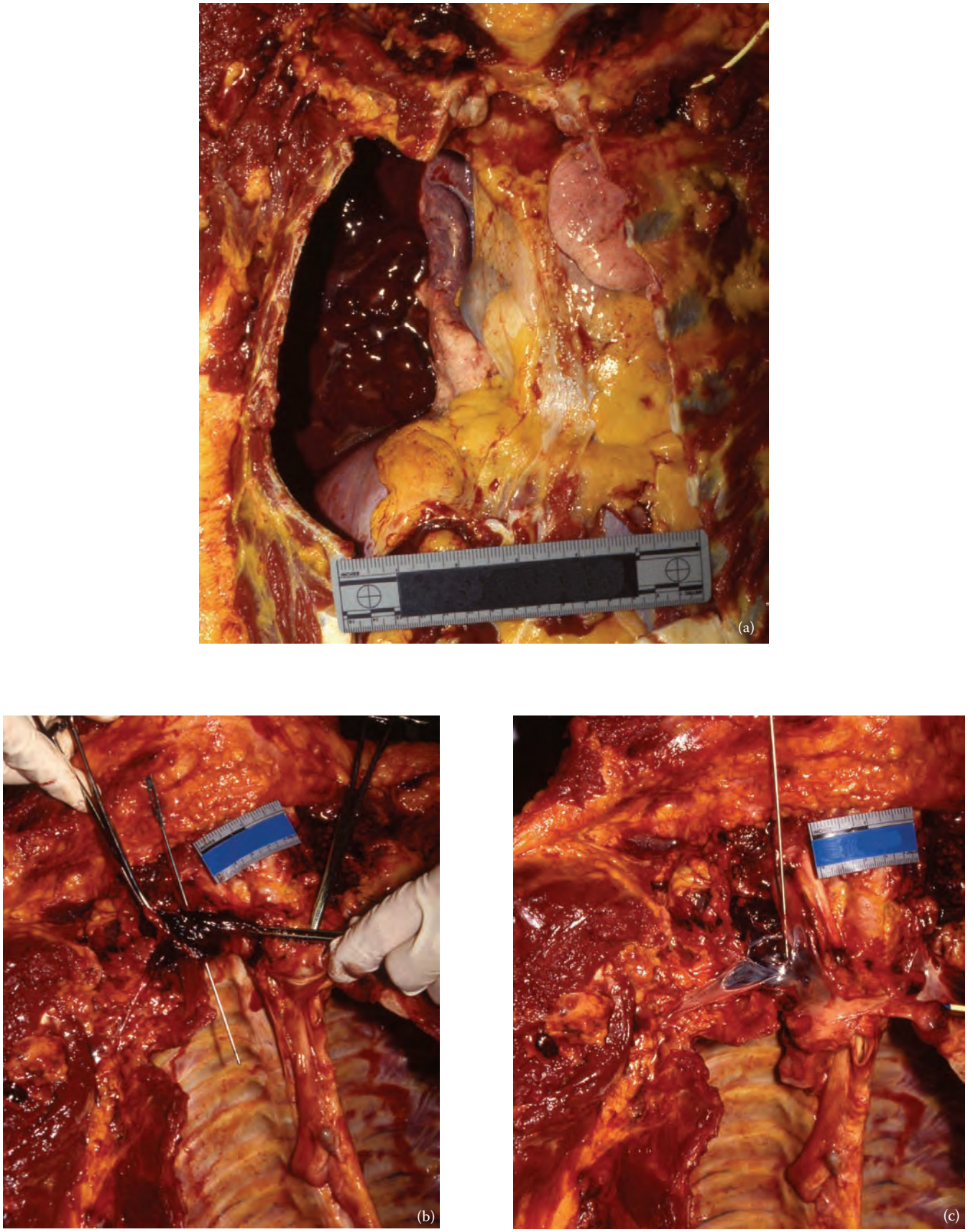

Right hemothorax (a) due to perforation of right internal jugular vein complicating catheter insertion. Note the probe (b) and (c) demonstrating the perforation through the right internal jugular vein and the hemorrhage within the anterior overlying soft tissues.

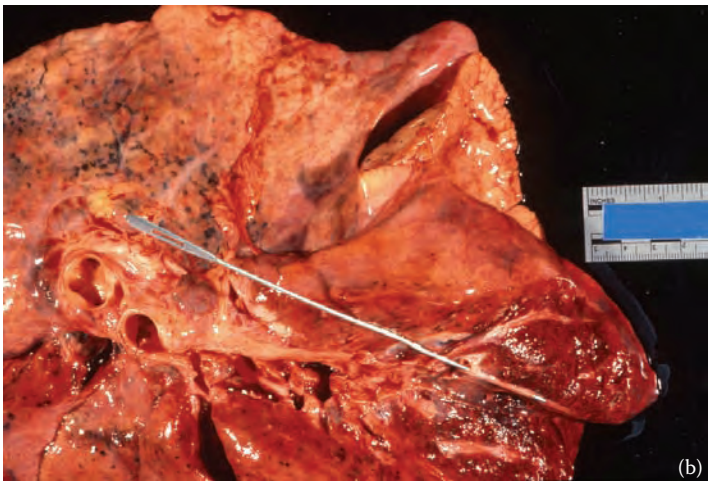

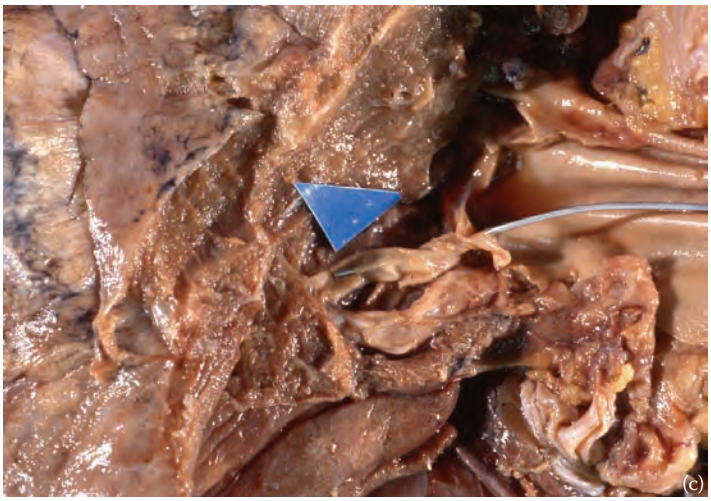

This case involved complications of a transbronchial biopsy in an individual who was HIV-positive and was slightly thrombocytopenic (platelet count: approximately 75,000/uL). The procedure was followed by extensive pulmonary hemorrhage, which culminated in respiratory compromise and death. Figure (a) demonstrates a Swan-Ganz catheter within the pulmonary arterial system. Figures (b) and (c) is a close-up view with a probe through a perforation of one of the large branches of the right main pulmonary artery with the arrow head demarcating the tip of the probe.

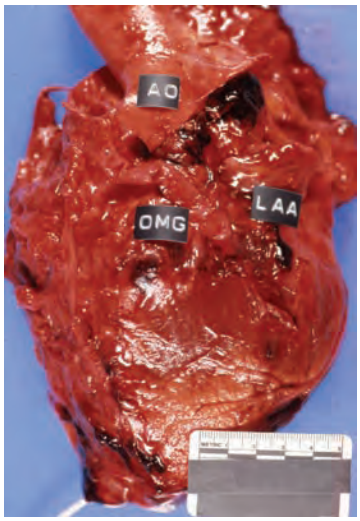

Heart with labels illustrating the aorta (AO), obtuse marginal graft (OMG), and left atrial appendage (LAA).

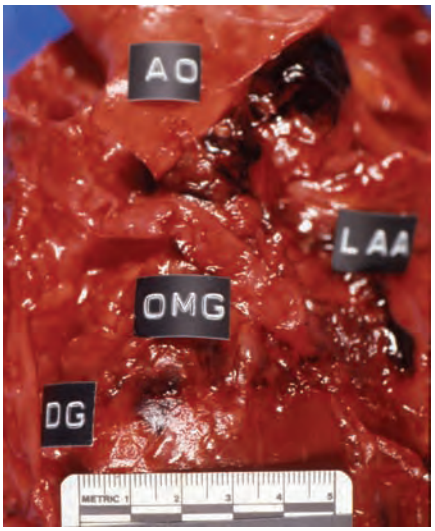

Close-up view of grafts.

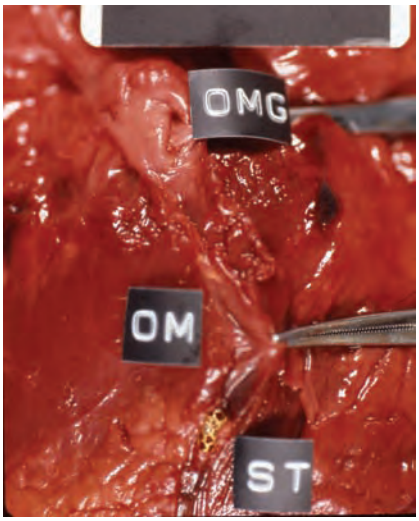

This image demonstrates the relationship of obtuse marginal graft and native obtuse marginal vessel, as well as a stent protruding through the lacerated native obtuse marginal artery.

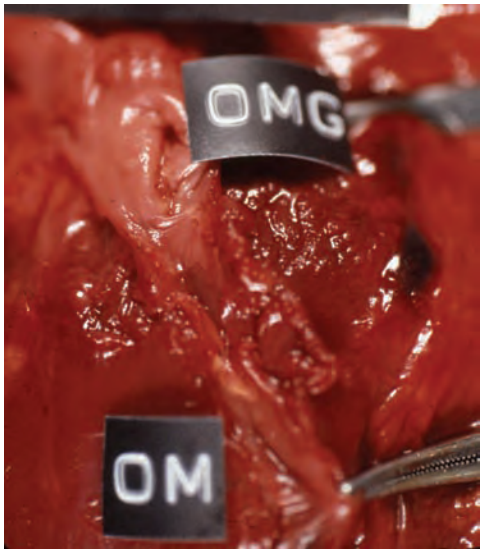

Close-up view of the junction between the obtuse marginal graft and the native obtuse marginal artery demonstrating the irregularity in the arterial wall.

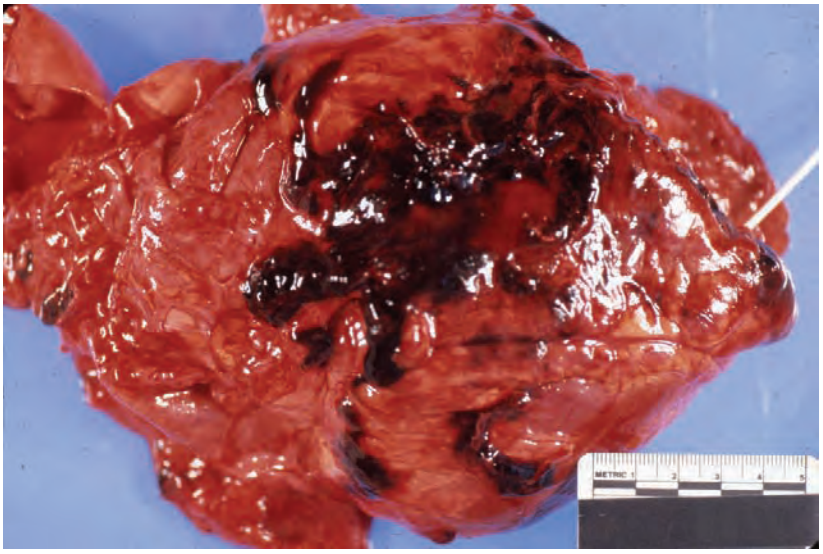

Posterior view of heart demonstrating confluent epicardial hemorrhage associated with this procedure.

These pictures demonstrate a complication of coronary artery bypass grafting. In this case, a stent perforated the native obtuse marginal branch of the left circumflex artery immediately distal to the anastomosis between the graft and native vessel.

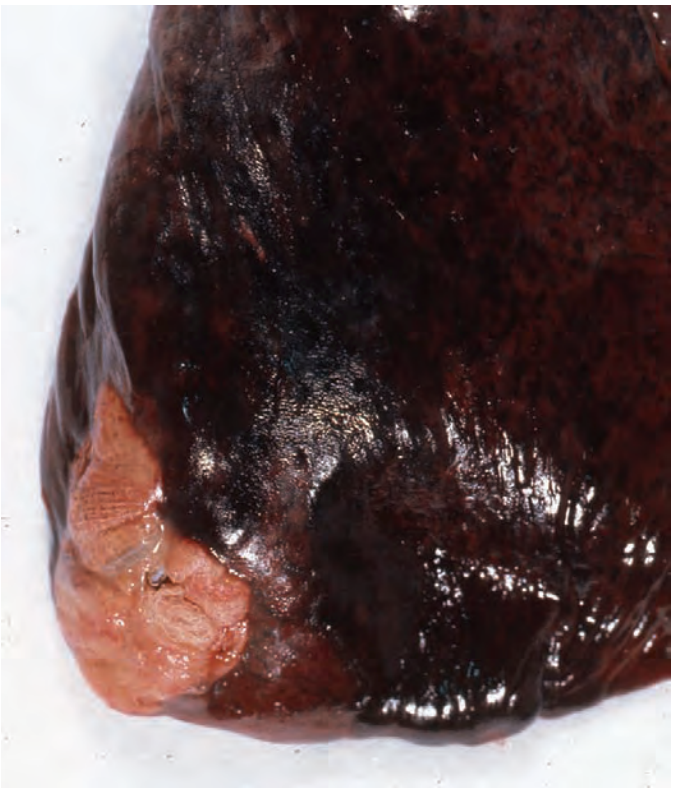

This depicts the lower lobe of the left lung with congestion and pneumonia. Note at the inferolateral aspect of the left lower lobe is a fragment of gauze that was inadvertently left behind during another operation months earlier. The gauze is adherent to the surface with overlying adhesions and adjacent purulent exudate.

This individual was in a motor vehicle accident and sustained a skull fracture involving the cribriform plate. A nasogastric tube was inserted that inadvertently penetrated the cranium. The tube was placed on intermittent suction. Aspirated brain matter is visible at the end of the tube.

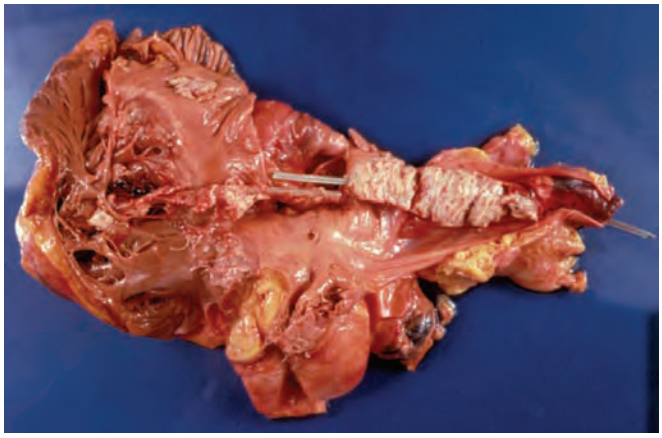

This demonstrates a large gray-tan thrombus encasing a pacemaker lead extending from the heart.

In busy trauma centers, the body is sometimes received with multiple instruments used for resuscitative efforts. This photograph was taken after the body bag was opened at the morgue. It is always important to be careful of sharps that have been inadvertently left behind.

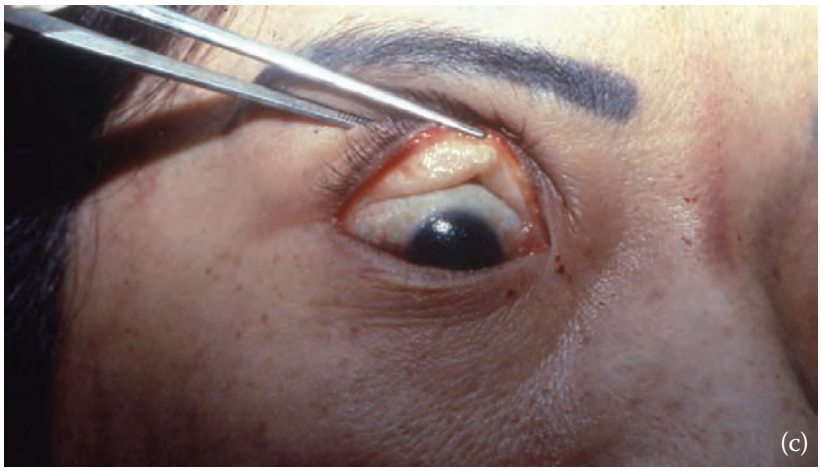

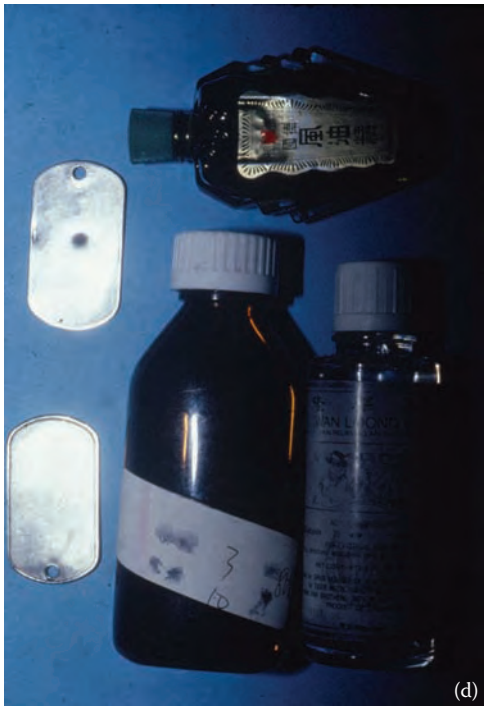

This Vietnamese woman died of Takayasu arteritis that was complicated by a ruptured aortic aneurysm. She was treated at home by her grandmother with Southeast Asian folklore remedies, including coining, which involves rubbing of hot oils and medicine onto skin (d). Note the unusual location of the coin marks at her neck (a) and (b). At first this was mistaken for manual strangulation. There were no internal neck injuries and no scleral or conjunctival hemorrhages (c).

Substance Abuse and Poisoning

3

CHARLES A. CATANESE AND LAURA M. LABAY

Introduction

The term substance abuse is most simply defined as the excessive use of one or numerous drugs. In general, most people typically have some level of access to several types of drugs, including legitimately prescribed medications, and on average, will benefit from their appropriate use. However, the term substance abuse can be applied to a variety of situations or circumstances. For example, one person may intentionally be abusing a multitude of substances including illicit drugs, prescribed medications, and over-the-counter remedies (i.e., supplements, herbals, diet aids, etc.). This is in direct contrast to another individual who may be taking only a single medication, but deliberately using it in a manner that is not consistent with normal therapeutic use. It stands to reason that when individuals begin to abuse one or more substances, and use them in an uncontrolled manner without the appropriate oversight, that they place themselves in a potentially harmful or lethal situation.

In a forensic or medicolegal setting it often becomes a matter of necessity to interpret drug findings and render an opinion regarding the toxicological, physiological, and pathological impact the drugs may have had upon an individual. To this end, it is important to evaluate and consider a host of factors. These factors are not always straightforward or quantifiable, but nevertheless include the drug's inherent physical and chemical properties, the dosage of the drug used, the frequency of the drug intake, the route of drug administration, the concentration of drug and drug metabolite found, the person's tolerance level to the drug and any medical conditions or disease states the person may have experienced.

It is important to recognize that for a particular drug to ultimately produce a toxic or lethal effect, it must be present in an individual at a sufficient concentration for a sufficient length of time. However, while references and other texts are available that help to classify drug concentrations as "therapeutic," "toxic," or "lethal," interpretation is often not so simple. For example, postmortem methadone concentrations are often challenging to interpret because the range of blood concentrations detected in people enrolled in narcotic maintenance programs may overlap the blood concentrations found in overdose or lethal situations.

Also, some drugs, depending on their physical and chemical properties and their concentrations in the body, will exert their most toxic or lethal effects in an acute manner, while others will take a longer period of time to act. In general, deaths associated with substance abuse may be related to acute or chronic complications. *Acute substance abuse* indicates that the death was related to direct toxic effects of the drug shortly after administration. In a *delayed overdose,* the drug produces damage to the body over hours or days following acute intoxication, with complications that may include coma, sepsis, brain swelling, and herniation. This process may evolve over days with the drug(s) metabolized from the body. The underlying process that set off this sequence of events is the acute intoxication.

Depending on the type of drug and its half-life there may be no or minimal amounts of drug left in the blood or other tissues at the time of death. However, the possibility exists that the drug or its metabolite(s) could still be detected and identified in the urine. In these circumstances, it is important to take into account that urine is really a pooled specimen collected in the bladder over a period of time and that a quantitative result represents only the average drug–urine concentration over the period of time that the urine was produced. Therefore, this type of specimen does not accurately reflect the blood–drug concentration at any single point in time. Rather, a positive finding of a drug or drug metabolite in urine indicates prior exposure only to that particular drug and in this regard it is relevant to consider other information (e.g., case history, medical records, anatomical findings, etc.) as well.

If a decedent suffered trauma while intoxicated and developed an *epidural* or *subdural hemorrhage*, you would expect to find the drug present in these samples even after many days. *Chronic sequela* associated with substance abuse include complications of infections such as HIV disease, hepatitis, hepatic cirrhosis, and endocarditis. Injection of ground-up oral medications may lead to pulmonary failure, with multiple foreign body granulomas and fibrosis demonstrated on histopathologic section. There are many other complications, including Wernicke's-Korsakoff's encephalopathy, alcoholic cardiomyopathy, and hardening of the arteries due to cocaine use.

Intravenous drug abuse may present with many different features. If the drugs are injected acutely in the same vein during binges, they appear as multiple fresh needle marks in a row. These are called "track marks" because they resemble a pattern of railroad ties being laid down one after another. It may be possible to detect the parent drug by excising the immediate area surrounding the injection site and submitting this section for toxicological analysis. If several injection sites are observed and a decision is made to submit more than one for testing, make certain to package the specimens in separate containers, identify the sites from which each was taken, and note their age and appearance. Alternatively, the syringe or other drug paraphernalia such as vials, pipes, or spoons, if available, can also be submitted for analysis. This is a good option if, for whatever reason, limited biological specimens exist. Once the nonbiological contents of the items are identified, directed toxicological analysis on the biological specimens can then proceed. If this is done, it is important that the items be packaged in individual containers, away from biological specimens, so that contamination does not occur.

Injecting drugs into blood vessels can lead to the introduction of many types of infections. Sharing of needles may cause transmission of the hepatitis virus or the human immunodeficiency virus. Various types of bacteria can also be introduced during injection, leading to vasculitis, cellulitis, pneumonia, and endocarditis. Acute intravenous fatalities, such as those caused when individuals inject themselves, are usually classified as *accident*. If one can demonstrate the drug was given to the individual in order to purposefully cause his or her death, the manner would be classified as *homicide*. Also, if another individual injects the drug into the decedent's arm the manner of death should be homicide.

As the vein is chronically abused, it will scar and develop a chronic track mark. Chronic track marks appear as linear scars that traverse the path of underlying veins. Histological sections from these regions often show polarizable debris from impurities found in past-injected drugs. If a chronic intravenous drug abuser consistently rotates the injection sites from one vein to another chronic track marks will not develop.

Skin-popping refers to drugs, such as heroin, injected into the subcutaneous tissues. This usually indicates a long history of drug abuse with destruction of the previously accessible peripheral veins. This may appear as fresh needle marks, usually at the surface of the thighs, or in other places without an underlying vein. These sites often get infected and lead to cellulitis, abscess formation, and scarring.

Another route of drug administration includes *inhalation*. Inhalation (i.e., snorting) of drugs may lead to mucosal reddening from direct irritation. Crack pipes may become very hot during use and burn the drug abuser's mouth and lips. Sometimes histopathology sections of the lungs reveal numerous pigment-laden macrophages from constant inhalation of smoke and debris from inhaled burning drugs such as crack. *Chronic snorting* of drugs may lead to a perforated nasal septum.

Keep in mind, it is not only the illicit drugs such as cocaine or heroin that can be inhaled; an individual may choose to *crush* and *"snort" a tablet or pill* as well**.** In some instances, residue may be observed in the nasal or oral cavity, and these areas can be swabbed and the parent drug identified though analytical testing. Individuals may also choose to inhale or "huff" commercial products such as gases, fuels, aerosols, solvents, or propellants as a means of abuse. This may be accomplished by breathing in the fumes from a rag that has been soaked in the material or by placing a bag containing the fumes over the nose and mouth. In some cases, the inhalation may take place directly from the item (i.e., aerosol can, glue bottle), as well. Users have died from hypoxia, pneumonia, cardiac failure, and aspiration of vomit. If the latter situation is relevant to the case, it is important to inform the toxicology lab of this, as most routine toxicology screens do not include these substances in the scope of their analysis. For these cases, lung tissue or tracheal air, in addition to the routine biological samples, can be collected as well.

Transdermal drug delivery systems introduce medications into the general circulation through a slow diffusion process, and may be considered a desirable alternative to taking oral medications or using substances that require repeated injection. In addition, the transdermal systems, perhaps most importantly, help to minimize the extreme blood spikes and trough levels that may be experienced with orally administered medications. Instead, blood–drug concentrations are maintained at more consistent concentrations.

Examples of commonly encountered medications delivered through transdermal systems include nicotine, hormones (i.e., contraceptives) and the potent narcotic analgesic fentanyl. However, similar to any medication, these patch-style systems may be subject to abuse. For instance, a person may make the decision to apply a multiple number of patches to the body, or through manual or chemical means remove the drug from the drug reservoir or from the adhesive matrix. It is also not uncommon for a person to chew or even swallow them. In these cases, remnants of patches may be observed in the oral cavity or the gastric contents. Other signs of this type of abuse would be if a person presented with patches applied in a manner not consistent with normal therapeutic use. For these cases, document the number of patches found on the body, their locations, any writing

or markings on the patches and any other descriptive features. Make an attempt to ascertain if all the patches represent a single acute application or if some of the patches may be from remote use.

In these types of death situations where patch use is known or suspected (i.e., old adhesive markings are noted on the body) it is important to make certain that specimens are not collected in proximity to a vein or artery that is in the immediate vicinity of a patch or patch marking. As a specific example, following the death of an individual wearing a fentanyl patch, some drug may still be present in the depot beneath the patch. If a blood specimen is collected from an area immediately beneath or near the location of the patch, the concentration of fentanyl may not accurately reflect the circulating concentration at or around the time of death.

Similar to the patch, suppositories represent another type of drug delivery system. Suppositories contain medicated material for insertion into a bodily passage or cavity such as the rectum. Once the suppository, typically a solid substance, is inserted, it will begin to dissolve over time and subsequently will deliver the medication into the body. Substance abuse using this route of administration is relatively rare.

Substance Abuse

Acute and Chronic Alcoholism

Death occurring from acute alcoholism is usually classified as *natural*, although this may not always be the case. A young adult at a party trying to impress friends may attempt to guzzle a liter of whisky. This is more than enough alcohol at one time to kill most adults. In this case, the manner of death would be accident. Sometimes depressed people with an acute triggering event, tell their family that they are going to "drink themselves to death." After several days of drinking large quantities of alcohol, they die. If you can establish the intent to do harm, the manner of death would be *suicide*. If someone at a party drinks alcohol from a funnel and a hose, by pouring the alcohol themselves, and then dies, the manner of death would be accident. If someone holds the hose and pours the alcohol for them, the manner of death should be homicide.

The deaths of individuals with a history of chronic alcoholism, including varying degrees of liver disease, intoxications, and withdrawal, are mostly certified as natural. Chronic alcoholics who stop drinking abruptly may get the "shakes" and eventually experience seizures leading to death. Withdrawal of many other drugs without other underlying natural disease will not typically cause death.

It is very difficult to give an opinion about individuals' state of mind or behavior while they were drinking or using any drug, for that matter. One has no way to know their true thought processes unless their behavior was witnessed or somehow clearly documented. Chronic alcoholics, with a high degree of tolerance, may be able to achieve a high blood alcohol level and not show visible signs or symptoms consistent with intoxication. Some alcoholics become aggressive, loud, and agitated while others become subdued and complacent.

It is important to obtain blood samples with care in cases where there is suspected alcohol use, and trauma played a role in the person's death. If one obtains a blood sample only from the pericardial sac or chest cavity, where other visceral lacerations may exist, contamination with gastric contents or other bodily fluids is a likely possibility. In turn, the measured blood alcohol level may be markedly increased as compared with the actual circulating blood concentration at and around the time of death.

As a more specific example relating to alcohol, when death is a consequence of multiple traumatic injuries there is the possibility that significant damage to the internal organs occurred. These organs include, but are not necessarily limited to, the stomach, small intestine, and liver. If those organs, now damaged, contained unabsorbed alcohol, a blood specimen collected from the chest area may readily become contaminated. As a result, the concentration of alcohol determined may not accurately reflect the circulating concentration at and around the time of the fatal event. Instead, ideally, for postmortem alcohol analysis and to mitigate some of these issues, an alternative specimen type should be harvested. These specimen types would include a vitreous specimen, a urine specimen, or blood collected from a peripheral site such as the femoral vein. This is because, in those cases where the heart blood may have become contaminated due to events such as trauma, the alternative specimen can be used to help provide an interpretable alcohol concentration.

Alternative specimens, such as vitreous fluid, are vital to collect, especially when alcohol is suspected or involved. It is important to recognize that alcohol may form postmortem, and analytical testing cannot make the distinction between this type of alcohol and alcohol that was present prior to death. Indeed, an alcohol finding may represent a combination of these two circumstances. To prevent or minimize postmortem formation once blood is collected, it should be placed into an appropriate type of specimen collection container, one that contains preservatives that help to inhibit microbial growth. If the postmortem formation of alcohol is a concern, such as in a case where marked decomposition has

occurred, a positive alcohol finding may be confirmed in the vitreous fluid, a specimen type that is more resistant to microbial growth.

Post-Mortem Redistribution

Besides contending only with the issue of contamination as caused by trauma, drugs and their metabolites may also be subject to postmortem redistribution. Postmortem redistribution is the movement of drugs among tissues, organs, and bodily fluids after death. The rate and extent of this movement varies according to several factors, including the nature of the drug, and the time interval between death and the postmortem collection time of specimens. Within the torso, the major organs constitute potential drug pools, and the gastrointestinal tract might contain considerable quantities of unabsorbed drug. Therefore, central blood is subject to redistribution from these local organs. In general, redistribution into central vessels is greater than redistribution into peripheral vessels and this is why it is preferred that final quantitative amounts be determined from a peripheral blood source. Therefore, while central blood pools are acceptable for screening purposes, it is always better to get a peripheral blood sample for quantitative confirmation work, if possible.

Cocaine

Cocaine, a Schedule II controlled substance, is found in the leaves of a South American shrub called *Erythroxylon coca* and is one of the most potent of the naturally occurring central nervous system (CNS) stimulants. First isolated in 1855, it has been used medicinally as a local anesthetic. However, because of its high potential for abuse, the use of cocaine for clinical situations has become severely limited. When cocaine is taken for illicit reasons it is either taken as the water-soluble hydrochloride salt by nasal insufflation ("snorting"), intravenous injection, or as the free-base ("crack") by smoking. Regardless of the chemical form or route of administration, once cocaine is administered it is rapidly absorbed and distributed throughout the body. Once inside the body, the dosage form of cocaine cannot be distinguished and analytical determinations are reported as the free-base form.

Cocaine is rapidly biotransformed in the body to a few major metabolites and products including benzoylecgonine, ecgonine methyl ester, ethylecgonine, and ecgonine. These metabolites are all pharmacologically inactive. Small amounts of an active metabolite, norcocaine, may also be produced. However, this product is rarely detected in blood. Cocaethylene is a pharmacologically active substance formed in the liver when cocaine and ethanol are co-ingested. The most predominant cocaine product detected in the majority of biological specimens is benzoylecgonine. Elimination on half-lives are approximately 4.5 hours for benzoylecgonine and approximately 0.8 hours for cocaine.

Effects displayed by an individual under the influence of cocaine may include dilated pupils, increased blood pressure, increased pulse rate, and increased sense of strength or invincibility. It is believed that cocaine is toxic to the cardiovascular system causing thrombosis, myocardial infarction, tachycardia, or fibrillation that may occur in cases of acute and chronic abuse.

Heroin

Heroin is a Schedule I controlled substance and a synthetic derivative of morphine. It is made by first extracting morphine from opium and then chemically treating the morphine with acetic anhydride, sodium chloride, and hydrochloric acid.

Once heroin is taken into the body, most frequently by injection, smoking, or inhalation, it is rapidly deacetylated to 6-monoacetylmorphine (6-MAM), a product that is then hydrolyzed at somewhat of a slower rate than morphine. Unlike the heroin, which has little affinity for the opiate receptors in brain tissue, both 6-MAM and morphine are pharmacologically active. Because 6-MAM is a specific product of heroin, if it is found to be present in a biological specimen, it can be concluded that the individual either used or was exposed to heroin at some point prior to death. However, due in part to the short half-life of 6-MAM, morphine is most often the predominant species detected in biological specimens. Therefore, in cases where heroin is suspected to be the lethal agent and morphine is found in the blood but 6-MAM is not, it may be of benefit to test an alternate specimen type such as urine for the presence of this heroin specific marker as well.

The primary toxic manifestations of heroin use may last for approximately 4 to 6 hours and include the same effects most commonly associated with other opioids. Some of the more common effects include drowsiness, loss of coordination, decreased blood pressure, decreased pulse and respiration, mental clouding, sedation, and sweating. At sufficiently high levels, the user may slip into a coma and may ultimately stop breathing.

Lysergic acid diethylamide (LSD)

LSD, a Schedule I semisynthetic controlled substance, is manufactured from the main precursor chemicals

lysergic acid, lysergic acid amide and ergotamine tartrate. LSD is normally taken by placing a "dot" laced with the material on the tongue. The LSD is then dissolved by the saliva and readily absorbed through the mucous membranes. This method of ingestion allows for its effects to be rapidly felt. Other means of ingestion include mixing the LSD with liquids or adding it to sugar cubes. However, the drug cannot be taken into the body by smoking, as pyrolysis destroys the LSD.

LSD is generally classified as a hallucinogen or psychedelic drug, and may produce both auditory and visual illusions. Approximately 30 to 60 minutes after ingestion, the user will experience the initial effects which, in general, may last for about 8 to 12 hours. Physiological effects are primarily sympathomimetic and may include mydriasis, hyperthermia, seizures, panic, and paranoid reactions. Flashback reactions, a brief recurrence of the LSD experience, are not uncommon in the experienced user, and may occur for weeks, months, or years after the last usage.

Death due to the pharmacological effects of LSD is rare, with most deaths occurring as a result of LSD-induced suicide and accidental trauma.

Marijuana

Marijuana, a Schedule I controlled substance, is a complex mixture of several products obtained from various parts of the *Cannabis sativa* plant, and is the most widely used illicit substance in the United States and the rest of the world. More than 400 chemical substances are found in marijuana. Sixty of these substances are called cannabinoids and are responsible for the psychoactive properties of the plant. The most relevant cannabinoid is tetrahydrocannabinol (THC), as it is the primary psychoactive ingredient in marijuana.

One of the most notable features of this drug is its long half-life, with some metabolic components exceeding 50 hours. This is because the drug is a highly lipid soluble and may undergo significant enterohepatic recirculation. In fact, the redistribution of THC from tissue to blood has been shown to be the rate-limiting step in its metabolism. In the body, THC is metabolized to two major metabolites, 11-hydroxy-THC (11-OH-THC) and tetrahydoxycarboxylic acid (THCC). The former metabolite is pharmacologically active, while the later is devoid of any pharmacological activity.

Marijuana is most frequently smoked, although it can be ingested as well. THC rapidly leaves the blood, even during a smoking period, and falls to below detectable levels within several hours. The most common physical effects are acceleration of heart rate, a moderate increase in blood pressure, a slight decrease in body temperature, reddening of the eyes, and a dryness of the mouth. The psychological effects of marijuana use include a pleasant feeling of well-being and euphoria, distortion of time, reduced ability to concentrate and memorize, and impaired short-term memory. Individuals under the influence of marijuana may have difficulties in tracking movement and demonstrate an inability to appropriately respond to stimuli. In general, this condition may persist for hours after the feelings of intoxication have dissipated, leaving users with a false sense of security concerning their abilities to safely operate a motor vehicle or machinery.

Death strictly due to the pharmacological effects of marijuana is not well documented, with most deaths occurring as a result of accidental trauma.

Methamphetamine

Methamphetamine is a Schedule II controlled substance with a very high potential for addiction and abuse. There are two different chemical forms or isomers of methamphetamine, each producing effects that differ in scope and magnitude. The l-isomer of methamphetamine may be found in over-the-counter nasal inhalers and is used for its vasoconstrictive properties. Compared with d-methamphetamine, it is a weak central nervous system stimulant. In contrast, the d-isomer may represent the licit or illicit forms of methamphetamine. In terms of legitimate or legal use, methamphetamine may be prescribed for a limited number of medical conditions such as weight loss, narcolepsy, and attention deficit disorder. However, because other less addictive and dangerous substances that do not quickly result in patient tolerance to the drug are also available, it is not that frequently prescribed.

Methamphetamine in the body undergoes demethylation to its primary active metabolite amphetamine; in most cases both methamphetamine and amphetamine will be detected. Analytical methods that differentiate the isomers of methamphetamine (and amphetamine) exist and may be employed if warranted.

People who abuse methamphetamine experience certain sequelae of such drug use. In general, the effects of methamphetamine can be broken down into three main stages. The first stage is the "high," where blood concentrations are at their greatest and people are feeling the stimulant effects of methamphetamine. The second stage is the "tweaking" period, where blood levels are on the decline and it is this period where people crave the drug and may behave in an aggressive and violent manner. The third stage is the "crash," where people feel exhausted and drained. High doses of methamphetamine can elicit restlessness, confusion, hallucinations, circulatory collapse, and convulsions.

Methadone

Methadone is a Schedule II controlled substance that is often prescribed during the process of narcotic detoxification, narcotic maintenance, and treatment programs, and to control severe and chronic pain. As compared with morphine, it produces less sedation and euphoria, but cessation of its use may result in withdrawal symptoms—not as severe as those seen with morphine, but longer in duration. Methadone works by decreasing the withdrawal symptoms felt by the narcotic abuser and, when a person attempts to reuse, the desired effects of the illicit drug are minimized.

In the body, methadone is metabolized to EDDP (2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine) and EMDP (2-ethyl-5-methyl-3,3-diphenylpyrroline). These metabolites do not possess any pharmacological activity and do not accumulate to an appreciable extent in plasma during therapy.

Methadone overdose is characterized by stupor, lethargy, pupillary constriction, hypotension, coma, respiratory collapse, and death.

3,4-Methylenedioxymethamphetamine (MDMA)

MDMA is a synthetic sympathomimetic compound with mixed stimulant, psychotropic, and hallucinogenic activities. It was used briefly as an adjunct to psychotherapy, but because of widespread abuse it has now been reclassified as a Schedule I controlled substance. The synthesis of MDMA is both complex and time-consuming. As a direct result, the final product may contain a variety of impurities that may be toxic in their own right. Also, drug manufacturers may add an illicit substance or a stimulant in an attempt to enhance the effects of the MDMA. The drug is available in tablet and powder form and it may be injected, inhaled, or ingested. Tablets of MDMA come in a litany of colors and shapes, and may be imprinted with a variety of images including things such as peace symbols, cartoon characters, butterflies, and angels.

In the body, MDMA is metabolized to many other compounds, but the main metabolite is a demethylated product called methylenedioxyamphetamine (MDA).

The effects of MDMA are related to the dose. A lower-level dose (approximately 50 mg) may result in feelings of enhanced creativity, while mid-level doses (approximately 100 mg) may make the user feel open to improved communication and empathy. The ability of the user to undergo self-exploration and analysis is achieved with the doses typically greater than 125 mg. In general, the psychological effects are much more pronounced than the physical effects. Abusers of the drug have been reported to experience longlasting neurobehavioral disorders following cessation of its use. Symptoms of MDMA toxicity include visual hallucinations, confusion, hypotension, agitation, coma, and death.

Phencyclidine (PCP)

PCP is a Schedule II controlled substance that was developed and marketed in the 1950s for use as an intravenous anesthetic. However, it was discontinued for medicinal use in 1965 as patients frequently would become agitated, delusional, and irrational, and experience the distortion of sights and sounds. In the early 1960s, PCP gained a reputation as a drug that produced certain desired feelings such as detachment and dissociation, but also caused the user to sometimes experience unwanted reactions (i.e., bad trips).

At low doses, PCP may cause changes in body awareness and produce psychological effects of euphoria, an alteration of time and space, confusion, bizarre behavior, and panic. Physical effects produced by PCP include impaired motor skills, shallow breathing, sweating, blank staring, speech disturbance, and an inability to regulate body temperature. At high doses, PCP may result in hallucinations, seizures, coma, and death. Because PCP has sedative-like effects, interactions with other central nervous system depressants, such as alcohol and benzodiazepines, may also lead to a life-threatening situation. In general, chronic PCP users may repeatedly use the drug for days at a time and during this period go without food or sleep. There appears to be no relation between plasma levels of phencyclidine and the degree of intoxication that a person may experience.

Oxycodone

Oxycodone is a Schedule II controlled semisynthetic narcotic analgesic derived from thebaine. It is used to control pain associated with such ailments as bursitis, injuries, simple fractures, and neuralgia, and is often found in combination with other drugs such as acetaminophen and aspirin. The addiction liability of oxycodone is about the same as for morphine.

Oxymorphone is a pharmacologically active metabolite of oxycodone that may be seen in blood in very low concentrations. Of interest is that oxymorphone may be prescribed as a parent drug and has a greater analgesic potency than morphine.

In overdose, oxycodone can produce stupor, coma, muscle flaccidity, severe respiratory depression, hypotension, and cardiac arrest. However, sustained-release preparations appear to produce adverse reactions, up to and including death, at lower-level concentrations, especially in combination with other central nervous

system depressants, depending on use pattern and route of administration.

Poisons

Arsenic

Arsenic is a metalloid that is present in all parts of the environment and, for example, may be found in water, soil, and sediment. In broad terms, there are two main forms of arsenic —organic and inorganic. Organic arsenic is present in food, with crustaceans and fish being some of the richest sources. These organic forms of arsenic (arsenobentaine and arsenocholine) are considered to be relatively nontoxic and will be rapidly excreted unchanged in the urine. Inorganic arsenic occurs in two oxidation states: a trivalent form (arsenite) and a pentavalent form (arsenate), with the trivalent form being more toxic than the pentavalent form, which undergoes metabolism to monomethylarsonic acid (MMA) and dimethylarsenic acid (DMA). MMA and DMA are then excreted in the urine.

Arsenic inactivates up to 200 enzymes, most notably those involved in cellular energy pathways, and DNA replication and repair. Unbound arsenic also exerts its toxicity by generating reactive oxygen intermediates that cause lipid peroxidation and DNA damage. Inorganic arsenic binds thiol or sulfhydryl groups in tissue proteins of the liver, lungs, kidney, spleen, gastrointestinal mucosa, and keratin-rich tissues such as the skin, hair, and nails.

The lethal dose of arsenic in acute poisoning ranges from 100 mg to 300 mg. Severe acute arsenic intoxication produces several well-described symptoms. Bloody vomit and diarrhea may occur within 1 to 4 hours of ingestion. Gastrointestinal volume loss is compounded by profound capillary permeability produced by arsenic's interruption of cellular energy metabolism. Cerebral edema, microhemorrhage, encephalopathy, and seizures may also arise from loss of capillary integrity. The rate-corrected QT interval (QT_c) prolongation and tachyarrhythmias may develop.

Subacute arsenic toxicity involves predominately the neurologic and cardiovascular systems. Within days to weeks after ingestion, many untreated or undiagnosed patients describe debilitating peripheral neuropathy characterized by excruciating pain and severe motor weakness. Persistent QT_c prolongation and the accompanying risk of *torsades de pointes*, a specific type of cardiac arrhythmia, occur among patients with clinically significant body burdens of arsenic.

Chronic arsenic toxicity presents itself following months or years of exposure. Some hallmark features of this type of toxicity not described above include hyperpigmentation of the skin, hyperkeratosis of the hands and feet and the appearance of Mee's lines in the fingernails.

It is important to note that death may occur in all of the above situations (e.g., acute, subacute, and chronic) if the person is exposed to a sufficiently high dose of arsenic.

Carbon Monoxide

Carbon Monoxide (CO) is an odorless, colorless gas without taste that forms as the result of the incomplete combustion of carbon-containing material. Motor vehicles, appliances, and heaters that use carbon-based fuels are major sources of exposure. However, it is important to note that natural sources of carbon monoxide also exist. These sources include fire, gases emitted from mines, marine algae, and human metabolism. Carbon monoxide is endogenously produced when hemoglobin, the molecule responsible for oxygen transport, and other heme-containing substances are degraded or broken down. Because of this, endogenous levels of CO, analytically measured as carboxyhemoglobin, are typically less than 1%. It is important to note that CO levels within the body may vary depending on several other factors as well. For example, because cigarette smoke contains CO, a smoker may exhibit carboxyhemoglobin levels as high as 8% saturation.

A person becomes exposed to CO via inhalation with the ultimate biological saturation level dependent upon several factors including CO concentration, duration of exposure and the activity level of the individual. Carbon monoxide poisoning produces hypoxia by two main mechanisms of action. First, CO binds to hemoglobin with an affinity that is greater than 200 times that of oxygen and therefore, by occupying the oxygen binding sites of hemoglobin, CO directly decreases the oxygen-carrying capacity of blood. Second, when CO binds to hemoglobin, the hemoglobin undergoes a change in its configuration so that oxygen release from the hemoglobin is hindered. Early signs of CO poisoning include headache, nausea, and vomiting. As the CO poisoning progresses, the person may experience impaired mental function, an inability to concentrate, and personality changes. Finally, the individual may develop seizures, coma, and death. Classic pathological signs that are most often associated with CO poisoning, although rarely observed, include cherry red skin and retinal hemorrhages.

Cyanide

Cyanide is a potent, rapidly acting lethal poison, and death may occur within minutes following its ingestion.

Common sources of cyanide include industrial manufacturing byproducts, plants, fruit pits, chemicals, and combustion products of certain plastics. Because of the latter, cyanide may play a role in the hypoxic events from fires.

Cyanide exerts its effects by disrupting electron transport at the cytochrome c oxidase step and this in turn adversely impacts the production of adenosine triphosphate (ATP). This break in the oxidative phosphorylation process stops the Kreb's cycle, and ultimately causes a metabolic acidosis as both pyruvic and lactic acid begin to accumulate.

The signs and symptoms of toxicity are dependent upon several factors. These factors include the form of the cyanide (e.g., gas versus solid), the route of exposure (e.g., inhalation versus ingestion), the duration and the extent of the exposure. The minimum lethal dose in an adult has been estimated to be 100 mg for hydrocyanic acid and 200 mg for potassium cyanide.

Cyanide produces a range of symptoms including dizziness, weakness, motor impairment, and mental impairment. These symptoms may progress toward slowed respiration, lactic acidosis, seizures, coma, and death.

It is important to recognize that blood concentrations of cyanide can increase or decrease during storage depending on the length of time, the temperature, and the presence of cyanogenic bacteria.

Ethylene Glycol

Ethylene Glycol is a nonvolatile liquid that is a common ingredient of automotive products such as antifreeze, de-icers, and coolants. It can also be found in some preservatives and as a glycerin substitute.

Following ingestion, ethylene glycol is rapidly absorbed, and manifestations of toxicity may be noted within approximately 30 minutes. Ethylene glycol is metabolized in the liver to several toxic metabolites including glycoaldehyde, glycolic acid, glyoxylic acid, and oxalic acid, and it is these metabolites that then may elicit central nervous system, cardiopulmonary, and renal dysfunction as well as produce a severe metabolic acidosis. The magnitude of the toxicity is dependent upon dose and the onset and success of treatment.

The effects of an ethylene glycol exposure are typically described in three main stages. In the first stage (0.5 to 12 hours post ingestion), neurological symptoms that include signs consistent with ethanol intoxication are manifested. Coma, convulsions, and possibly death may result in this stage. The second stage (12 to 24 hours post-ingestion) is often characterized by cardiopulmonary disturbances including tachycardia, tachypnea, and hypertension. In severe ingestions, congestive heart failure, pulmonary edema, and circulatory collapse may be seen. The end stage of ethylene glycol toxicity is renal failure (24 to 72 hours post-ingestion).

Methanol

Methanol is a type of alcohol that is extremely versatile, and so can be found as a component in industrial solvents, fuels, and antifreeze preparations, and as a denaturant for ethanol. Methanol may be introduced into the body through multiple routes of administration including inhalation and ingestion. Once in the body, methanol is first metabolized to formaldehyde and then to formic acid. Interestingly, although formaldehyde is considered a toxic substance, it has on average a half-life of only several minutes and it is the formic acid that ultimately produces the hallmark features of methanol toxicity, including metabolic acidosis and the loss of visual acuity. Methanol poisoning can be treated by the administration of ethanol, because both methanol and ethanol share a common metabolic pathway. In essence, the ethanol competes for binding to the alcohol dehydrogenase enzyme and indirectly limits the formation of formic acid. Acute methanol exposure may produce severe signs and symptoms of toxicity, including nausea, abdominal pain, and lethargy. In some cases, the poisoning will progress to where the person experiences an anion gap metabolic acidosis leading to coma, seizure, and respiratory collapse. It is important to recognize that methanol may be a component found in embalming fluids and, because of this, tissues or other samples that have come into contact with this type of material may test positive for the presence of methanol.

Poison Hemlock

Poison hemlock (*Conium Maculatum*) is a biennial member of the carrot family that grows wild throughout the United States, especially along roadsides. Socrates is thought to have died from ingestion of poison hemlock. Its toxicity is from several simple piperidine alkaloids including coniine, gamma-coniceine, conhydrine, N-methylconiine, and pseudoconhydrine. Coniine and gamma-coniceine are thought to have the most significant contribution to the toxic effects, which are similar to nicotine poisoning.

Thallium

Thallium is a metal that was once used in rodenticides, insecticides, and depilatories, but was banned from residential use in the 1970s. Today, it is used in the

semiconductor industry and may be found in switches and electronic devices. Thallium is readily absorbed from the gastrointestinal tract and may be detected in most of the body's tissues and fluids, including blood, brain, liver, kidney, spleen, bone, hair, and urine. The half-life of thallium in the blood is approximately 2 to 4 days, with one of the hallmark symptoms of this type of poisoning, alopecia or loss of hair, occurring after about 1 to 3 weeks. Symptoms of acute exposure include severe gastrointestinal distress, tingling of the hands and feet, paralysis, and respiratory failure. In people being chronically exposed, additional signs and symptoms include paralysis, hepatic and renal issues, and respiratory failure.

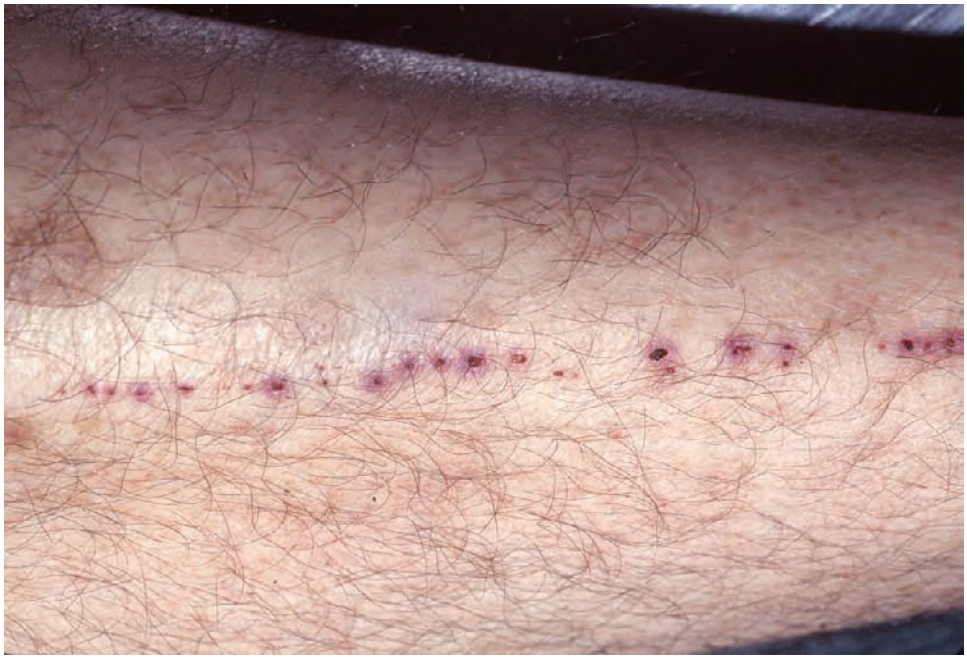

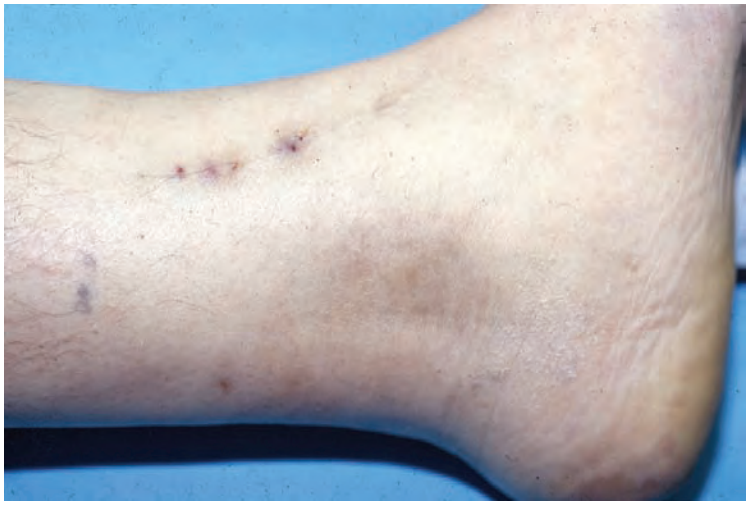

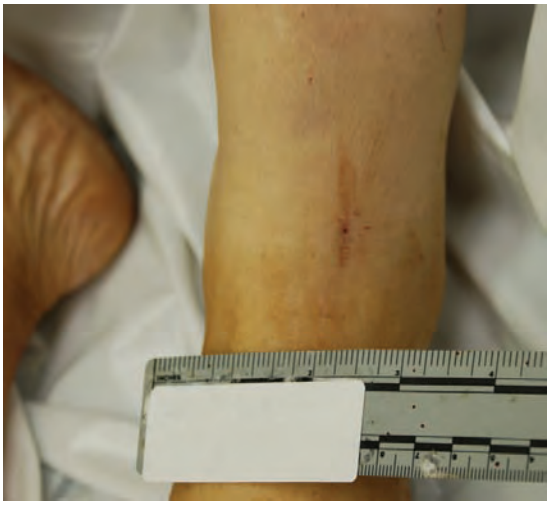

Acute track marks. Multiple fresh needle mark injection sites that traverse the path of underlying veins. Incision of these regions will often reveal underlying hemorrhage. Toxicological testing of underlying injection site tissue may reveal the parent compounds, such as heroine.

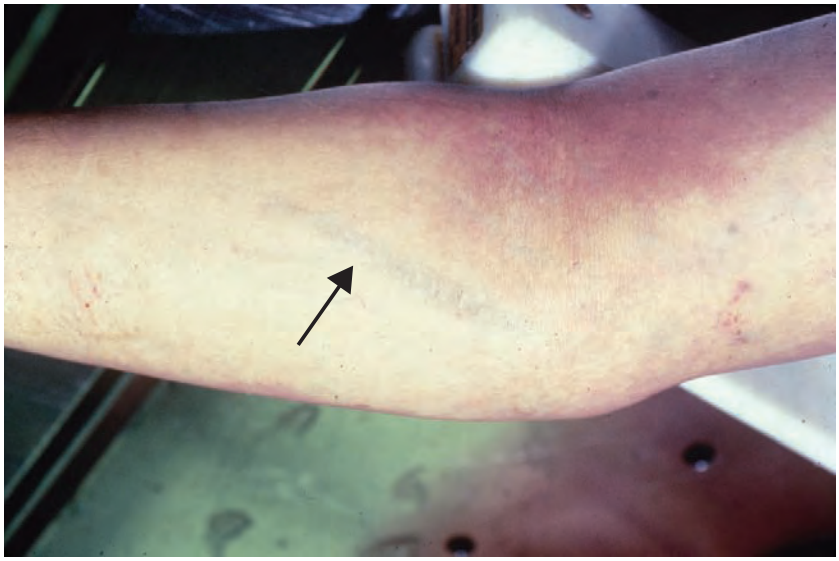

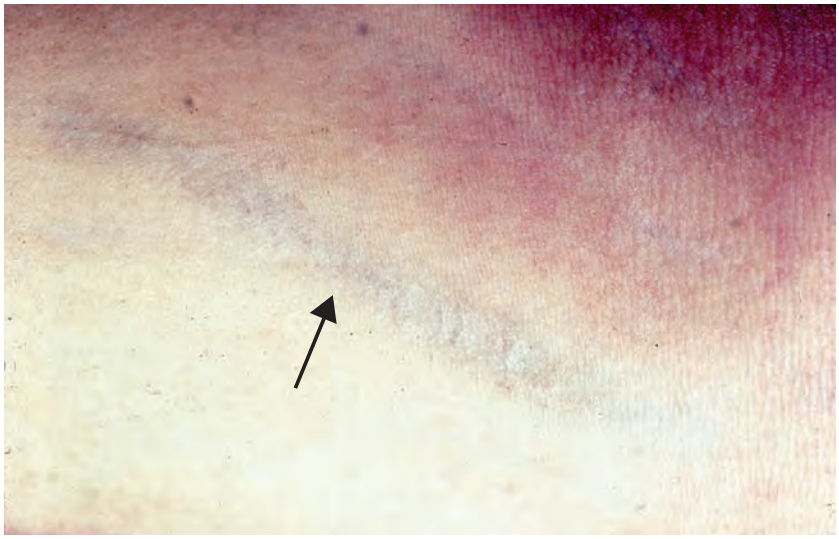

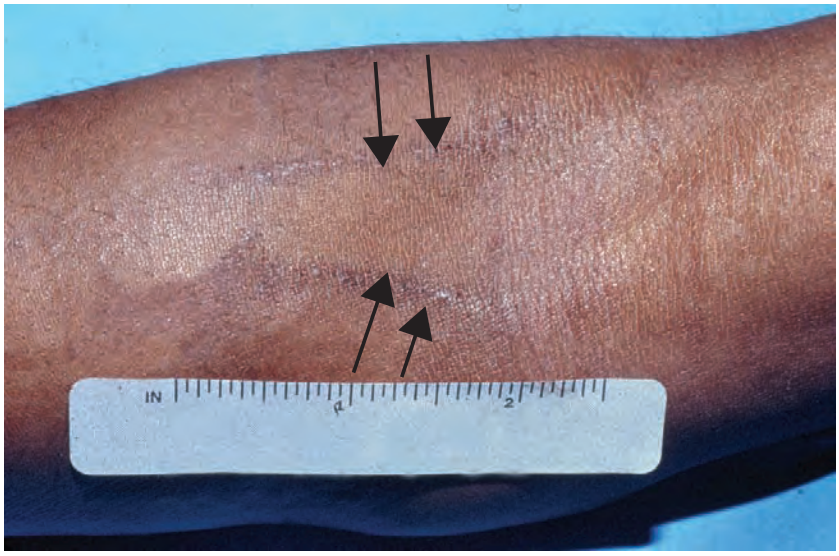

Chronic track marks. Linear scars that traverse the path of underlying veins. Microscopic sections may demonstrate inorganic debris in sub-adjacent soft tissues.

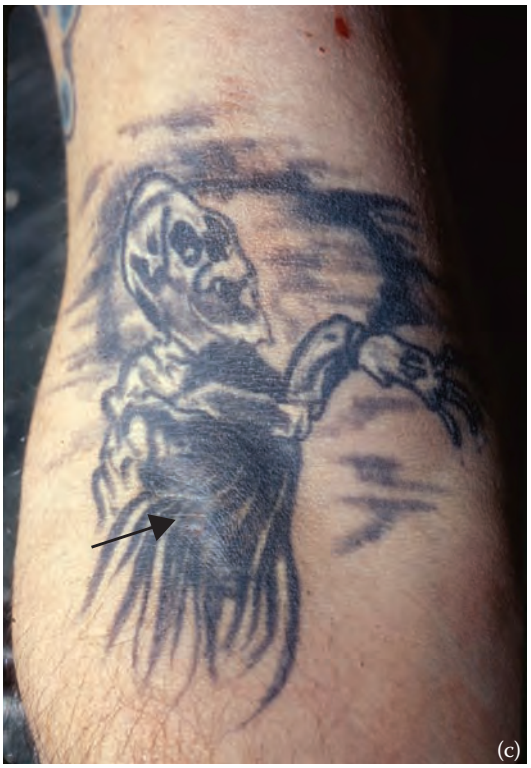

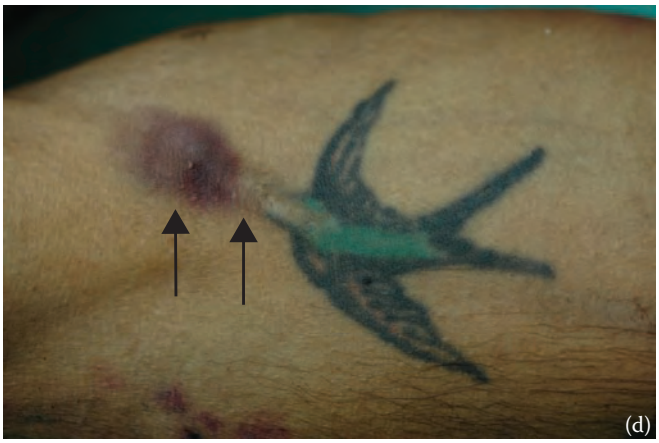

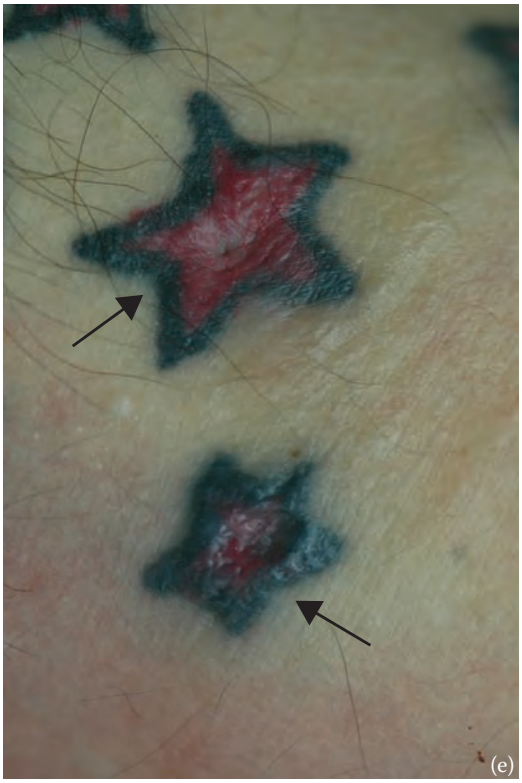

Tattoos are sometimes used to disguise intravenous drug abuse. These figures show examples of track marks in tattoos. Note the fresh injection site at the tip of the bird beak (d) and the healing injection sites at the stars (e), as well as other chronic linear track marks. Also note the fresh needle mark to the left of the tattoo (b) due to documented therapeutic intervention.

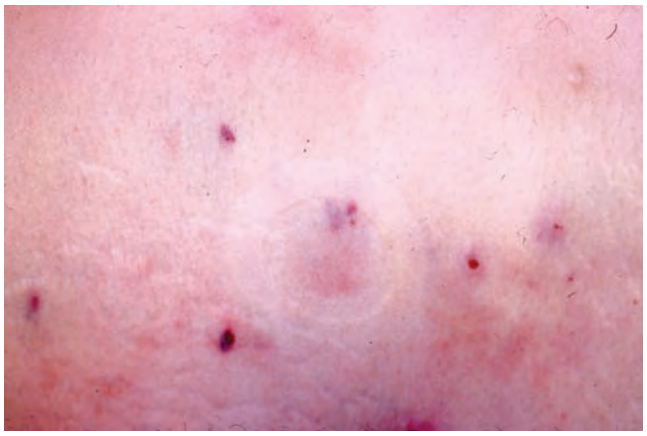

Fresh "skin popping" lesions. When drug abusers exhaust peripheral vein access they may start injecting drugs subcutaneously.

Recent healing infected skin popping lesion with adjacent cellulitis. It is common for these lesions to become infected due to the nonsterile nature of the injection equipment.

Recent healing infected skin popping lesion.

Healed and almost completely healed skin popping lesions. Note the lower wound has almost complete replacement by scar except for the central healing defect.

Remote old healed skin popping lesions.

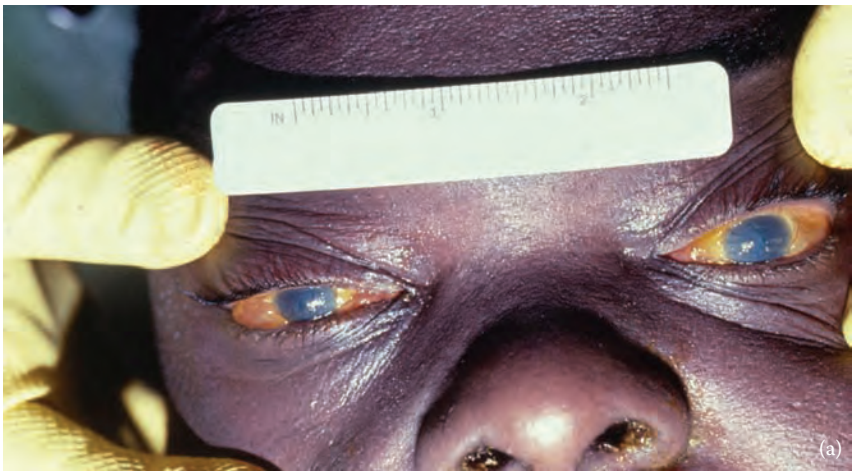

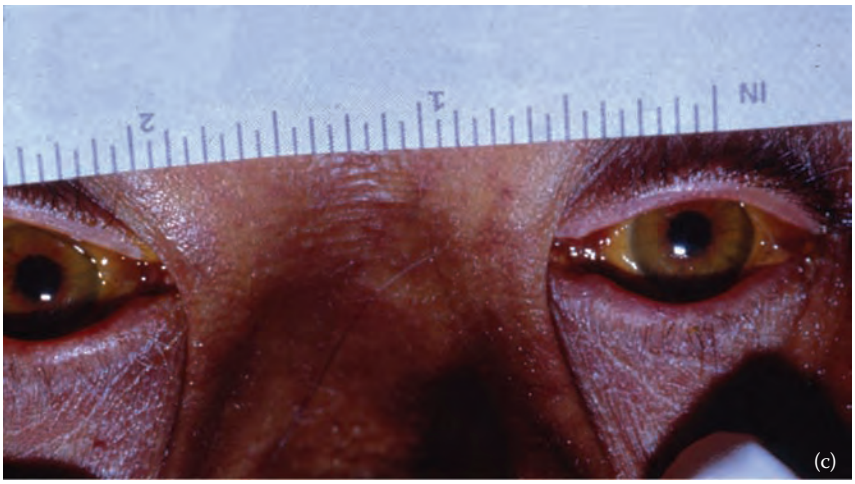

Examples of jaundice following hepatic failure due to hepatic cirrhosis as the result of chronic alcoholism; and viral hepatitis caused by chronic intravenous drug abuse. Note the yellow discoloration in the sclera and skin (a) and (c). Figure (b) demonstrates a comparison to a nonjaundiced individual. Also note the patchy areas of ecchymosis due to coagulopathy associated with the liver disease.

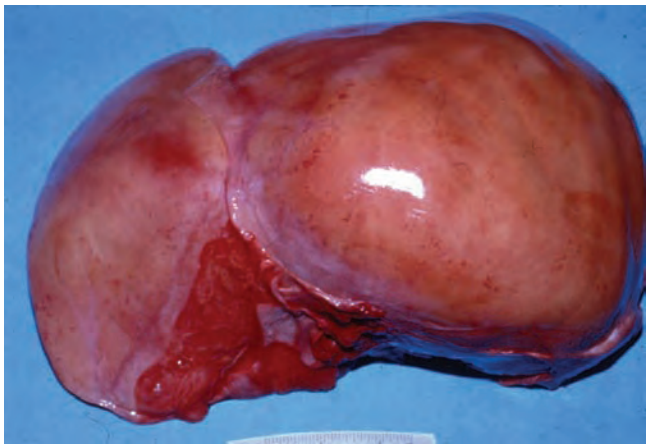

Marked hepatic steatosis due to acute alcoholism.

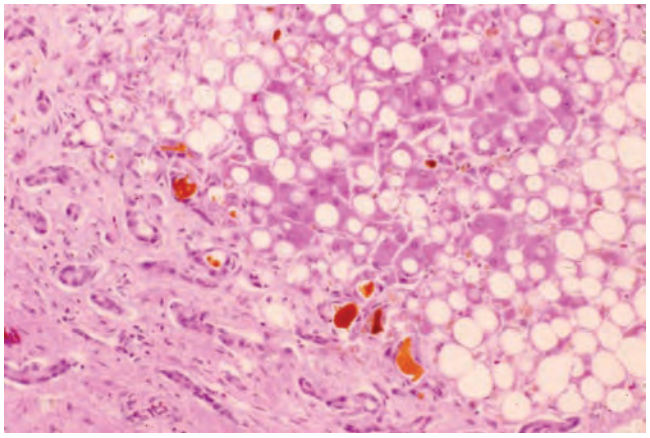

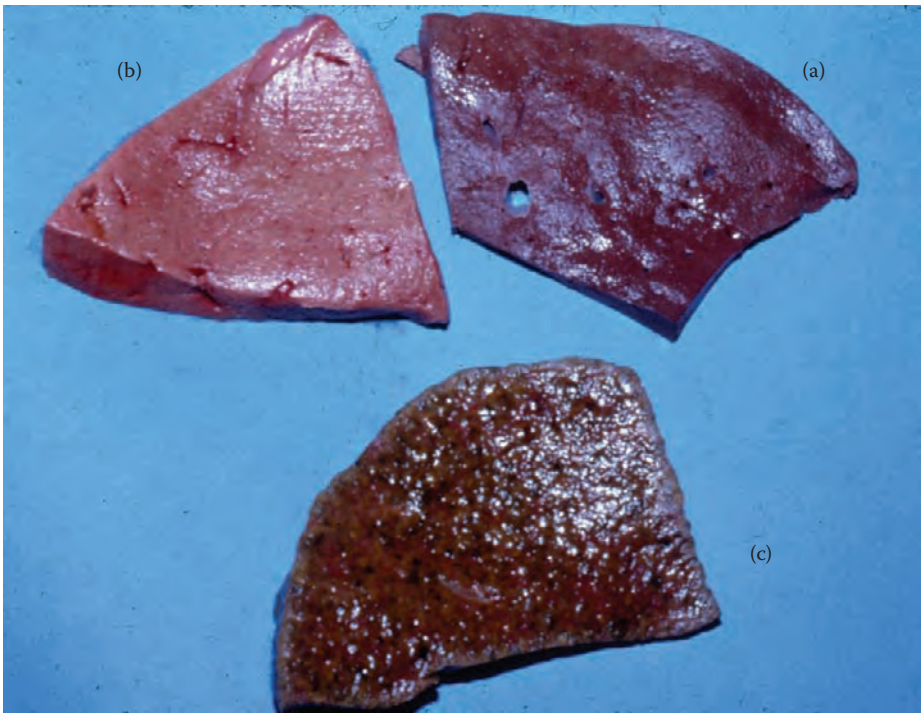

This image demonstrates cut portions of liver. Normal liver is brown (a); the yellow section demonstrates marked steatosis or fatty liver (b). The green section demonstrates micronodular cirrhosis due to chronic alcoholism with inspisated bile (c). The bile imparts this green discoloration.

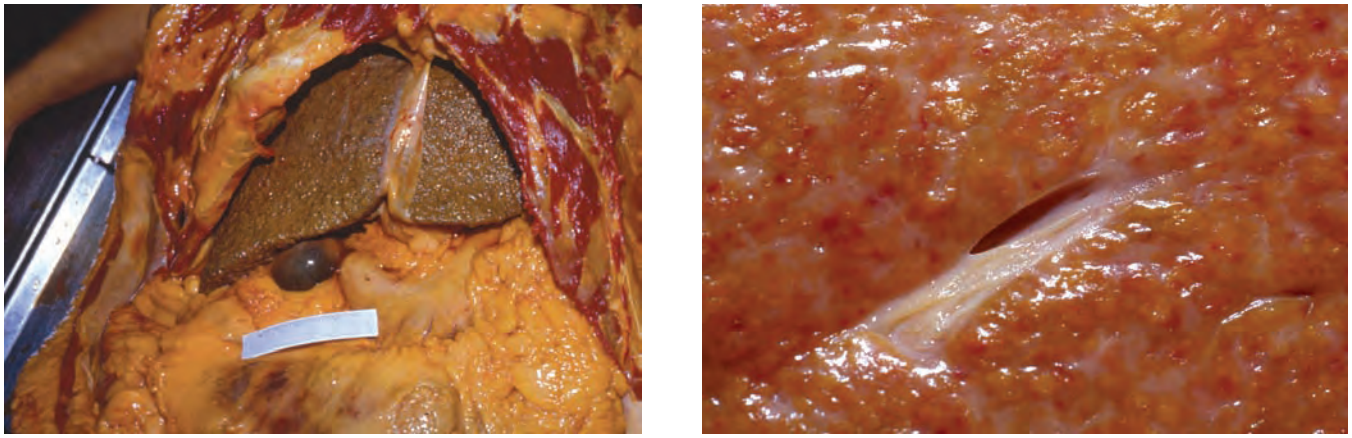

Micronodular hepatic cirrhosis with steatosis due to chronic alcoholism.

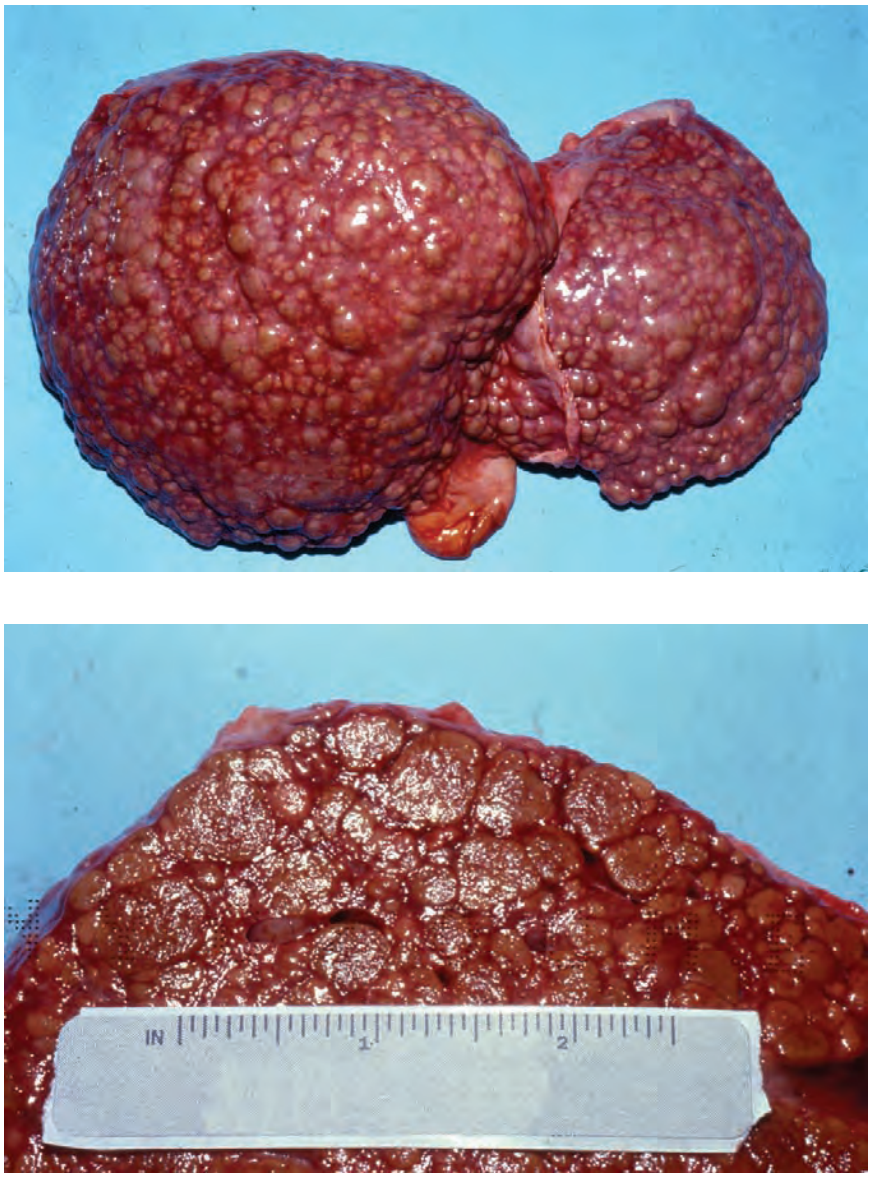

Macronodular hepatic cirrhosis due to hepatitis B infection as a result of chronic intravenous drug abuse.

Chronic alcoholic with hepatic cirrhosis and gynecomastia. This individual also had neurofibromatosis, which is demonstrated by the multiple subcutaneous nodules.

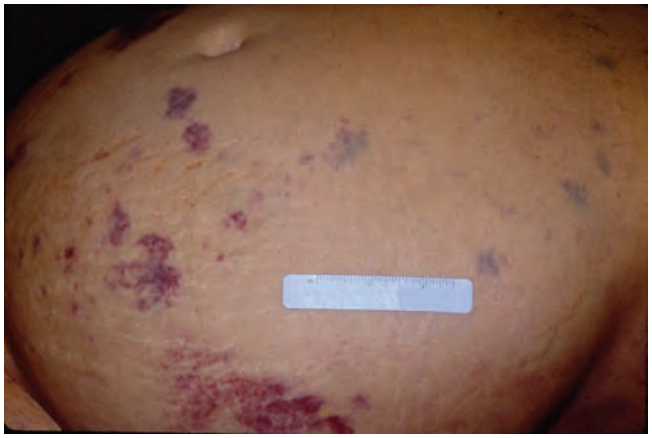

Chronic alcoholic with hepatic cirrhosis and coagulopathy. Note the multiple areas of ecchymosis and contusion caused by minimal blunt force trauma.

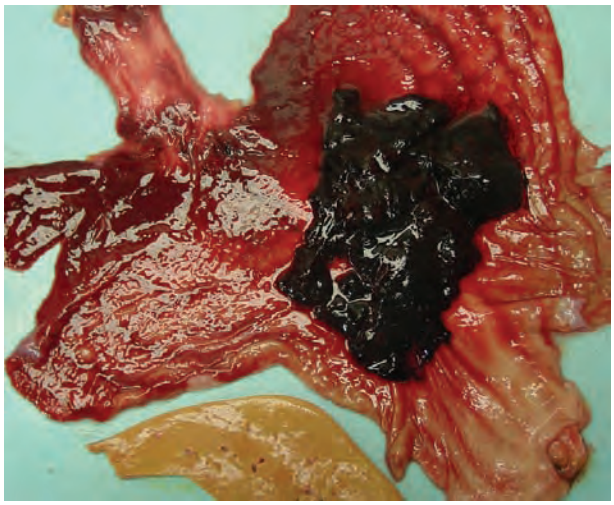

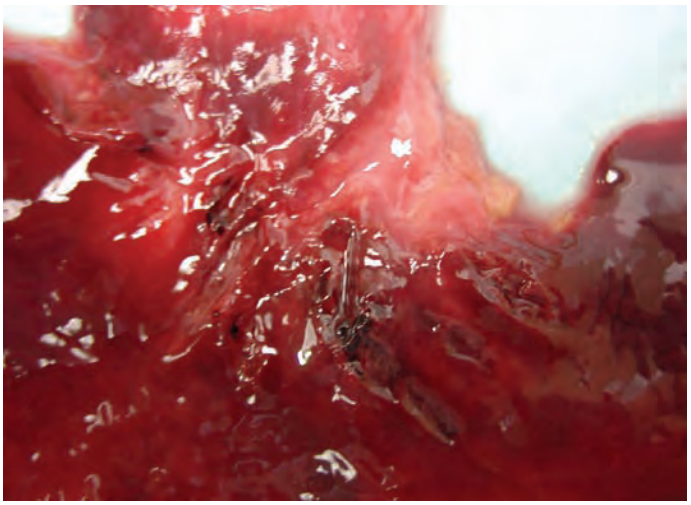

Gastrointestinal hemorrhage due to Mallory-Weiss tears in a chronic alcoholic. Note the laceration of the gastro-esophageal junction leading to death from gastrointestinal hemorrhage following multiple episodes of vomiting.

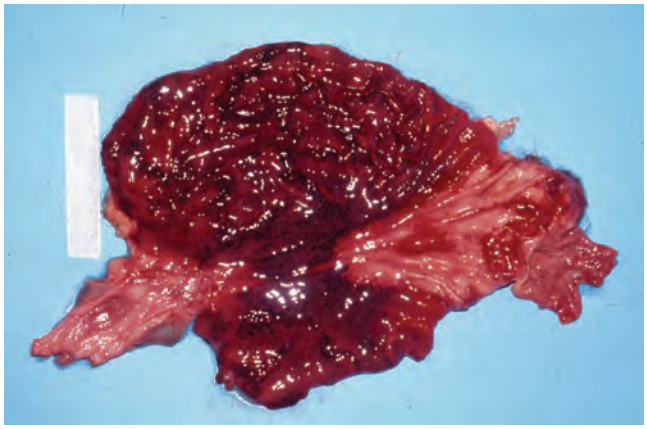

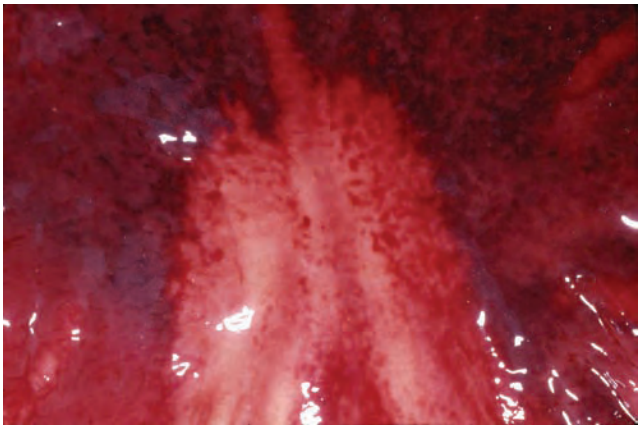

Acute gastritis associated with chronic alcoholism. Note the red discoloration of the mucosa.

Esophageal varices due to hepatic cirrhosis and increased portal hypertension. Note the dilated red to brown submucosal veins that may rupture and possibly cause exsanguination.

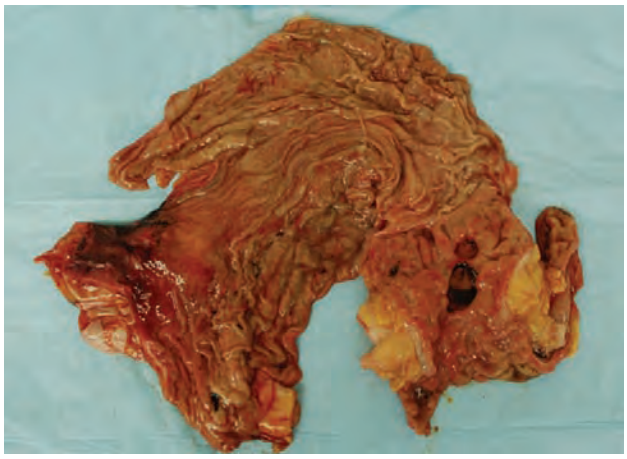

Bleeding peptic ulcers in a chronic alcoholic with gastrointestinal hemorrhage and hepatic cirrhosis.

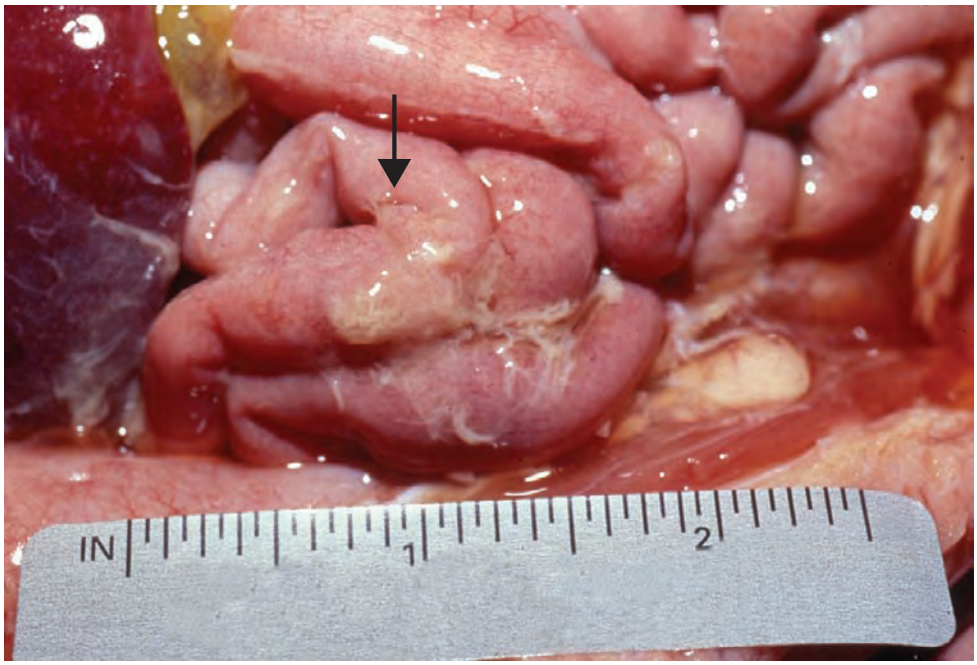

Acute peritonitis in a chronic alcoholic. Note the purulent exudate at the intestinal surface.

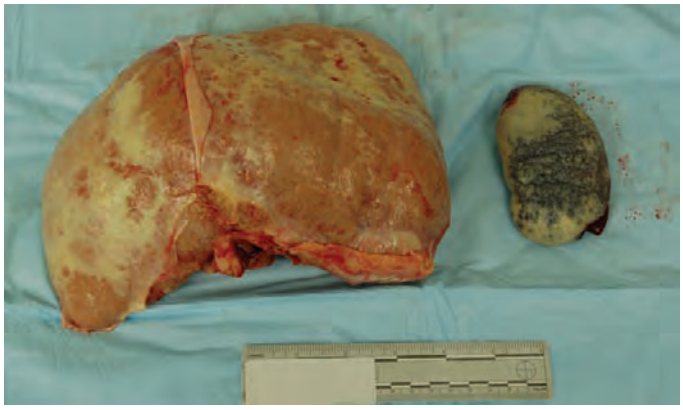

Hyaloserositis. Old healed peritonitis with fibrosis at the surface of the liver and spleen in a chronic alcoholic.

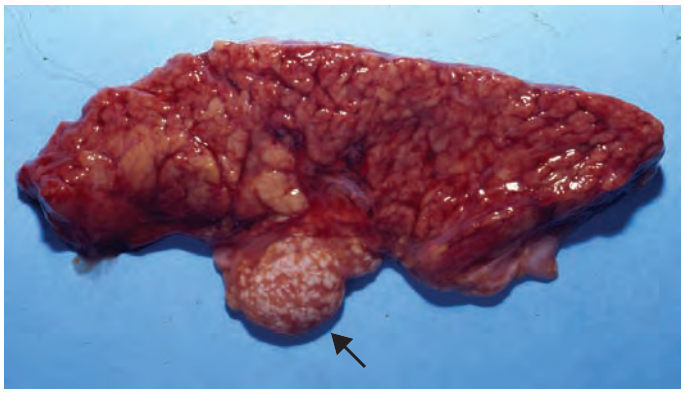

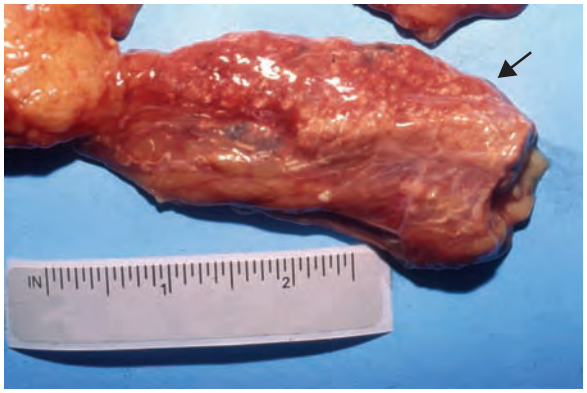

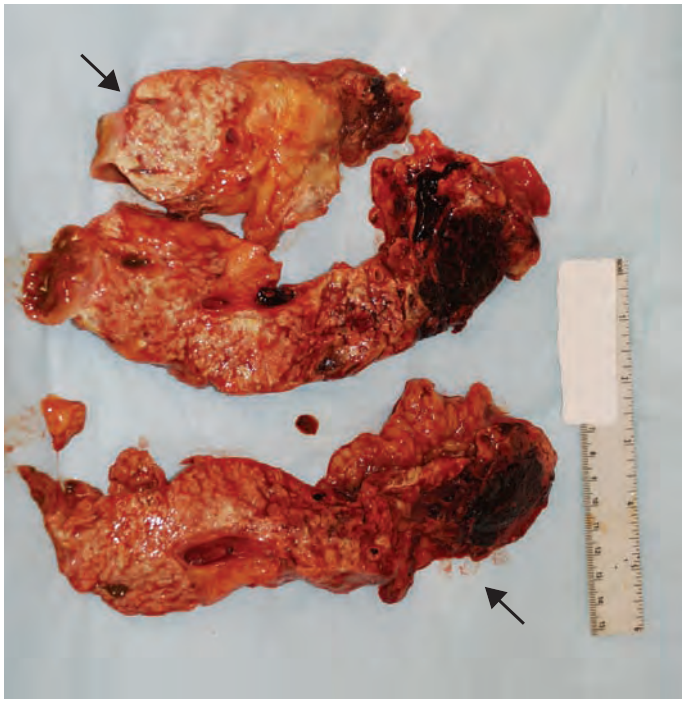

Acute pancreatitis due to chronic alcoholism. Note the white flecks due to fat necrosis and areas of hemorrhage.

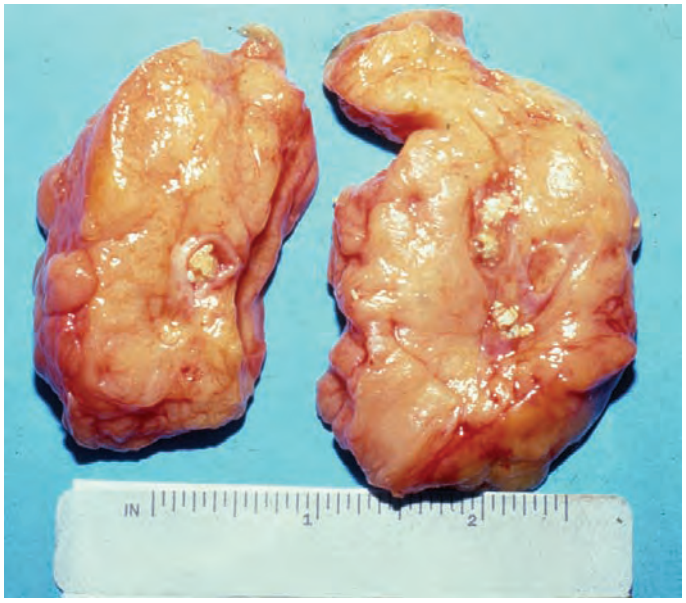

Fatty metamorphosis or atrophy of the pancreas with duct concretions in a chronic alcoholic.

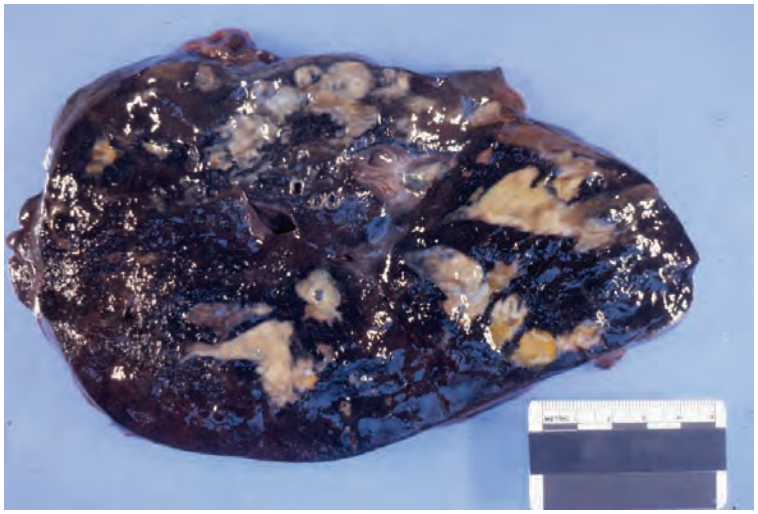

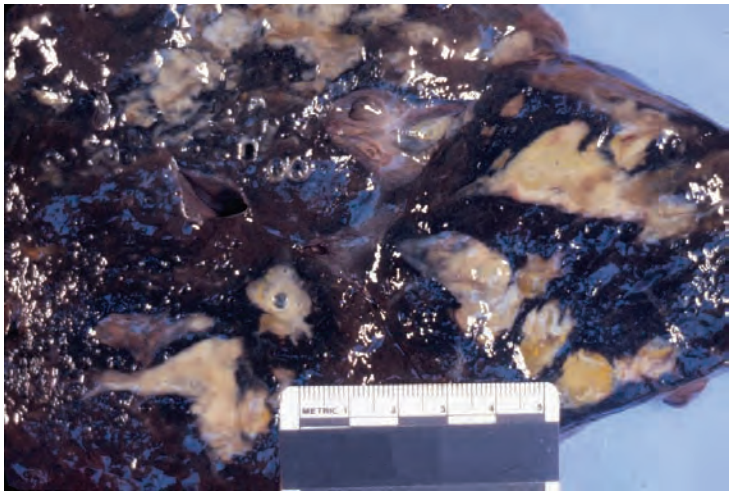

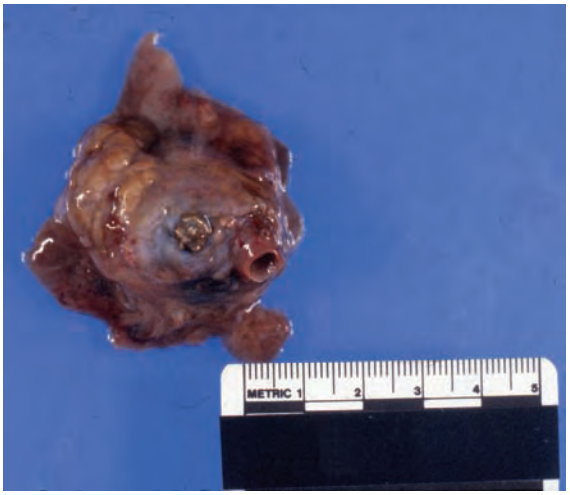

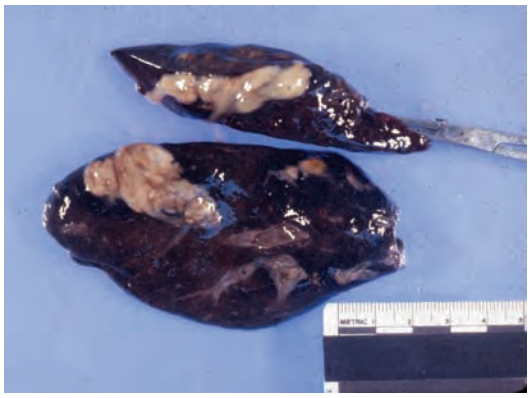

Pyogenic liver abscesses. This decedent had a history of chronic alcoholism with chronic pancreatitis and recent abdominal pain of unknown etiology. This photo depicts multiple geographic, creamy yellow-white collections of purulent material within the liver parenchyma that microscopically proved to be dense collections of neutrophils, karyorrhectic nuclear debris, and masses of faintly basophilic, fine filamentous bacteria. Samples of the abscesses were submitted to the microbiology laboratory but the best that they could do was anaerobic gram negative rods. Candidates included bacteroides, prevotella, and fusobacterium. Actinomyces would have been a consideration but they are gram positive. Pyogenic liver abscesses, while relatively rare, are known complications of chronic alcoholism and pancreatitis.

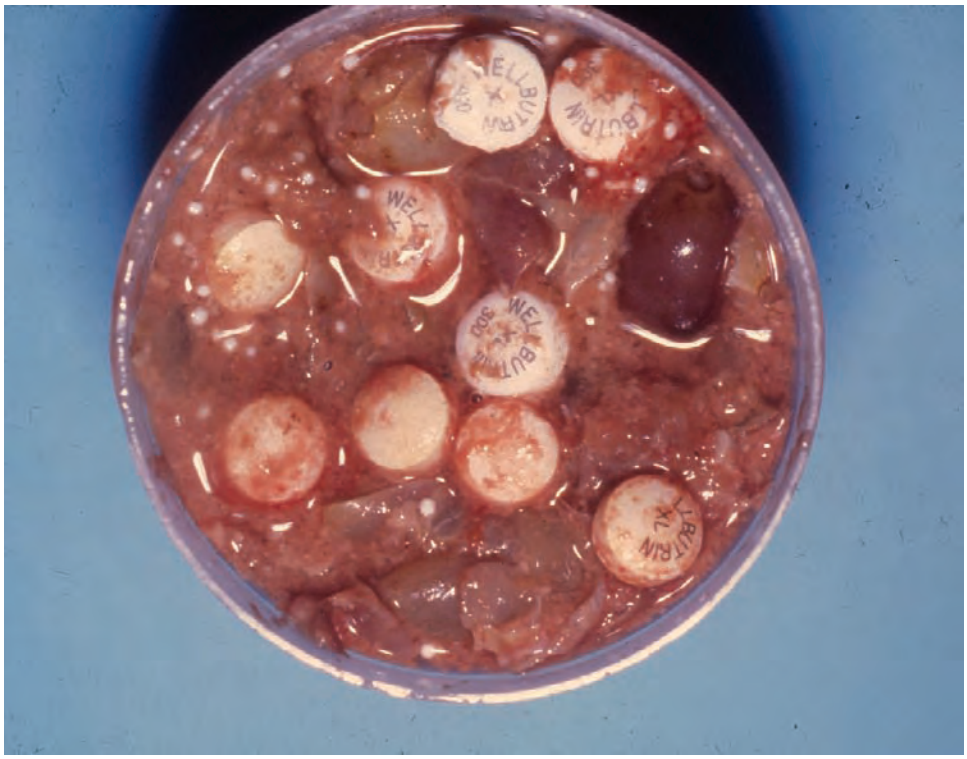

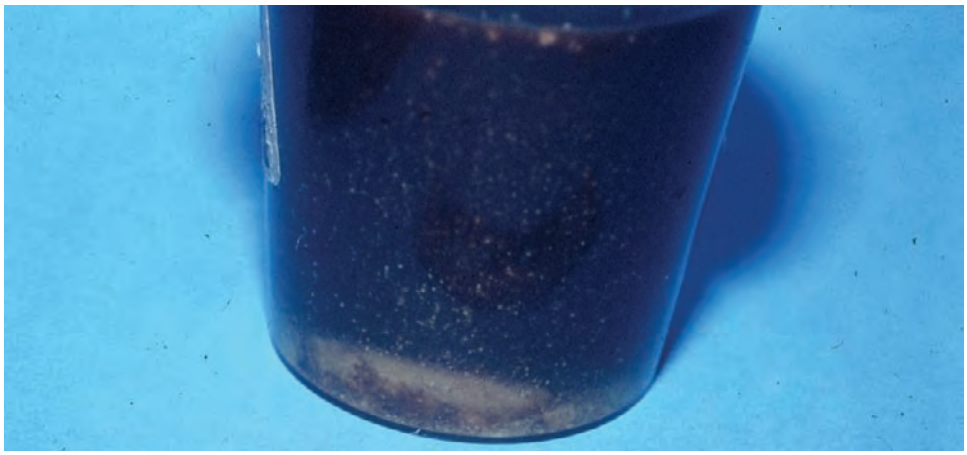

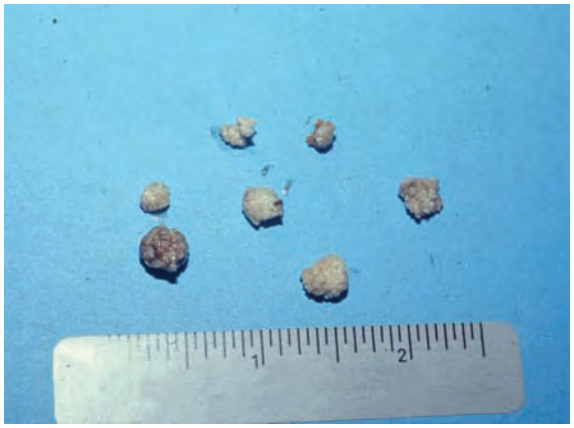

Suicidal overdoses demonstrating intact and fragmented granular pieces of partially digested pills. Note the containers with the granular white flecks at the bottom that indicate pill fragments.

Note the multiple scars demonstrating destructive behavior associated with drug abuse, mental illness, and suicide.

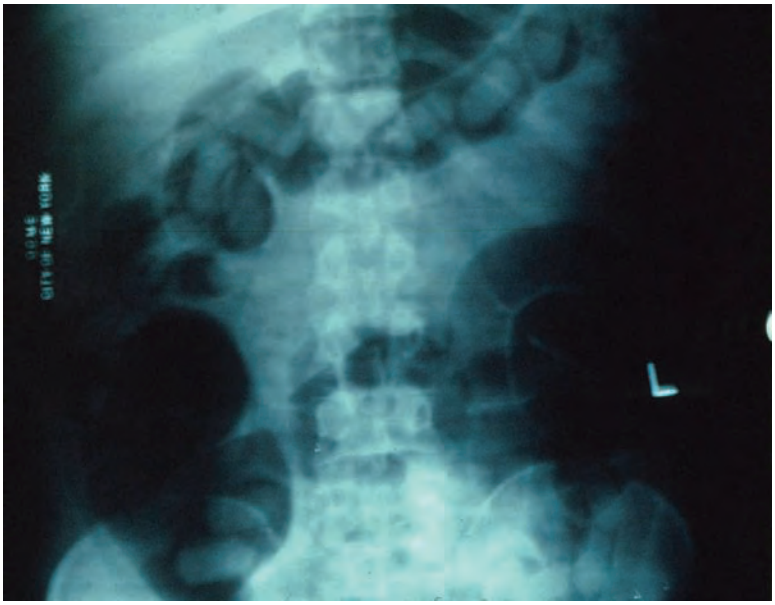

Acting as a "mule," this person ingested multiple packets of drugs to smuggle them into the country. One of the packets ruptured and the individual died of a drug overdose before the plane landed. Note the typical x-ray findings demonstrating these packets within the gastrointestinal tract.

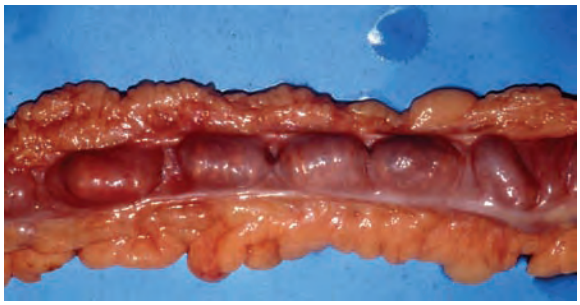

Packets of drugs discovered at autopsy in individuals who died in hotels next to an airport. Note the broken packets demonstrated at the bottom of the last picture.

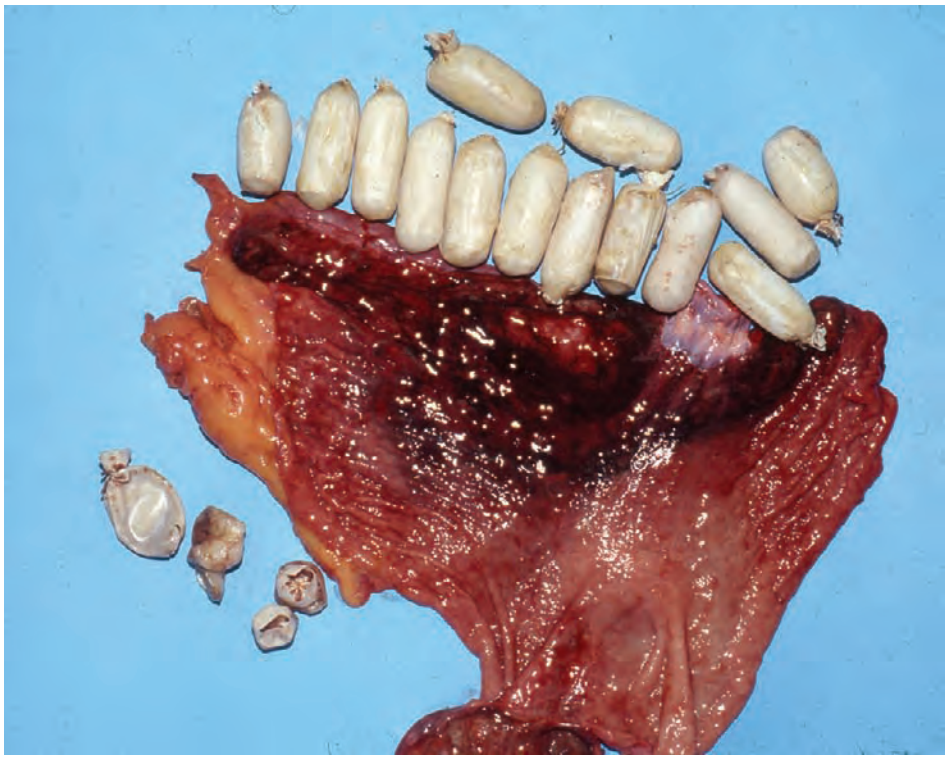

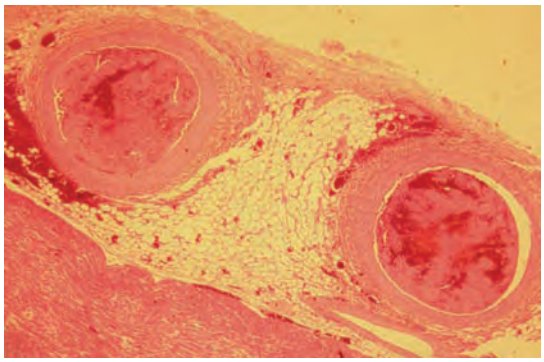

Epicardial vessels with slight atherosclerosis and occluding thrombosis associated with acute cocaine intoxication.

Carbon monoxide poisoning with cherry red lividity.

Postmortem Change and Time of Death

4

CHARLES A. CATANESE, BRUCE LEVY,
AND GERARD CATANESE

Introduction

In most jurisdictions the time of death is legally defined as the time that the person is declared or recognized to be dead. Thus, a decomposed body or skeletonized remains, clearly deceased for significant periods of time, may have an official time of death that is days, weeks, months, or even years after the actual death of the individual. However, being able to estimate the actual time of death (or the postmortem interval) can be critically important in criminal investigations, civil litigation, or settlement of the deceased's estate.

The concept of a "window of death" was developed to help answer these questions. This window can be initially bracketed by the time that the person was last known to be alive and the time he or she was declared dead. Forensic science can then apply a variety of observations and tests in an attempt to narrow that window as much as is scientifically and medically possible. Accurate interpretation of postmortem change is crucial in helping to establish the actual time of death.

Postmortem changes can be subdivided into several categories. These categories include early postmortem changes, decomposition, and skeltonization. Each has characteristics that can overlap with advancing time and are variable depending on environmental conditions and the physiologic state of the body at the time of death. The activity of insects and animals will also create artifacts that can lead to misinterpretation of the postmortem interval. Interpreting postmortem changes is not an exact science, and without a witness, one can only estimate a time period in which the death was most likely to have occurred.

Early Postmortem Changes

Generally considered to occur or evolve within the initial 24 hours after death, these consist of algor mortis, livor mortis, and rigor mortis.

Algor Mortis

The cooling of the body after death. After death the body will gradually cool until it equilibrates with the ambient temperature. Under standard climate controlled conditions, with average humidity and a room temperature of about 72° F, the body generally loses about 2° F per hour for the first 12 hours and than approximately 1° F per hour for the next 12 to 18 hours. If the ambient temperature is greater than the body temperature, algor mortis occurs in reverse and the body temperature rises until it reaches the ambient temperature.

The above calculations apply only at or near room temperature and will become variable as the ambient temperature changes. As a general rule, the rate of temperature change in a body is proportional to the difference between the temperature of the body and its environment. In other words, if the ambient temperature is below 72° F, the rate of heat loss will be increased from the numbers above. The reverse would be true for temperatures greater than 72° F. Moving air will cool faster than still air, and wet bodies will cool faster than dry bodies. Bodies found in stagnant water, flowing water, or those found buried in the ground will cool at different rates from those in the air and from each other. When the body and environment are at the same temperature, they have reached steady state and estimates of time of death based on temperature are limited to a minimum period of time.

Many other variables can influence algor mortis. In general, an elevated body temperature at the time of death will give the appearance of a shorter postmortem interval based on a measurement of the body temperature. A lower body temperature will have the reverse effect. Things that effect body temperature include infection, strokes, seizures, thyroid disease, and many other natural diseases. Whether or not the person was involved in strenuous physical activity prior to death (such as a violent struggle) will increase body temperature. Many other factors, including the age of the person, his or her overall health, and many drugs or medications can also have an effect on the calculation of postmortem interval based on this and other observations.

It is important to make certain there is consistency in how the body temperature is obtained. First, these calculations are based on a core body temperature, meaning either taking a rectal temperature or, more ideally, introducing a thermometer into the abdominal cavity beneath the liver to measure core temperature. Temperature readings on

the surface of the body, including axillary, oral, or ear temperature are unacceptable for determining a postmortem interval. The thermometer must be kept in place long enough for the body temperature to equilibrate.

Livor Mortis

The settling of blood under the effects of gravity after death. It can first be appreciated as early as 20 minutes after death in very light-skinned individuals. The color of the livor will deepen to a purple color over the first 8 to 12 hours after death. During this period the livor is typically blanching, meaning that pressure on an area of livor will cause the color to briefly leave the area, resulting in a pale mark. After about 12 hours the livor will become fixed and no longer blanch under pressure. As with algor mortis, temperature will influence the rate at which livor becomes fixed, with increased temperatures shortening the time for livor to become fixed and decreased temperatures lengthening the fixation time. In some cases, livor mortis advances to cause small visible hemorrhages in the skin known as Tardieu spots, which should be distinguished from petechiae.

The medical condition of the person and the cause of the death can also create changes in livor mortis. People who are anemic (low blood count) or sustain an injury resulting in significant loss of blood might have very faint or even absent livor mortis. It can also be difficult to appreciate livor in dark-skinned individuals.

When evaluating livor mortis, it is important to note its location, intensity, whether it is blanching or fixed, and if it is appropriate for the position of the body. If livor is inappropriate for position, you can conclude that the person was moved at some point after the onset of livor. You can also evaluate livor for its color. For example, people who die as a result of carbon monoxide or cyanide poisoning will have a bright "cherry red" color to their livor mortis. People who die in a cold environment might also have a similar red color to their livor.

Rigor Mortis

The stiffening of the body after death. It is the result of a physiochemical process within the muscles of the body that does not cause actual contraction of the muscles. Under typical conditions, rigor will first be appreciated within 30 minutes to 2 hours after death. It will progress to a maximum intensity over the first 12 hours and remain at a maximum until about 24 hours after death. It will then "pass" from the body between 24 and 36 hours after death, after which the body will remain flaccid.

As with algor mortis and livor mortis, many variables will affect the rate of development and passing of rigor mortis. In general, increased environmental or body temperatures will speed the rate of development and passing of rigor mortis. Lower environmental or body temperatures will have the reverse effect. People at the extremes of age (children and the elderly) will have different rates of appreciation of rigor mortis due to the decreased muscle mass in these age groups.

When evaluating rigor mortis, it is important to note its location and intensity and whether it is appropriate for the position of the body. If rigor is inappropriate for position you can conclude that the person was moved at some point after the onset of rigor.

Decomposition

As time advances, decomposition gradually increases. Decomposition can be subdivided into two basic categories, putrefaction and mummification. While they can be seen in isolation, careful observation will typically reveal features of both processes simultaneously.

Putrefaction

Decomposition that occurs due to the actions of bacteria. The bacteria typically break down the body from the inside out, causing many of the changes we associate with a person who has been deceased for a longer period of time. However, when there is penetrating trauma that breaks the surface of the body, bacteria from the environment can gain access and hasten putrefaction. These changes include darkening and slipping of the skin, the production of a foul-smelling gas, with bloating of the body and marbling of blood vessels. A dark colored bloody-appearing purge fluid will come from the nose and mouth and should not be confused with blood related to trauma.

This process is extremely environmentally dependent. As with the early postmortem changes, warmer temperatures accelerate this process and cooler temperatures slow it down, with many of the same factors playing a role. A body placed in a dark dumpster in the hot summer sun for half a day can reach a state of putrefaction equivalent to a body left at room temperature for several days or one in the winter cold for greater than a week. In the case of fire fatalities, charring will preserve the body and decelerate this process. A frozen body that has thawed will putrefy at a markedly accelerated rate due to spaces left by frozen ice crystals.

Mummification

Mummification occurs in a dry environment, typically such dry outside environments as a desert, or in heated indoor environments during the winter months when the relative humidity is low. Initially, there is a darkening and hardening of the skin. This will progress to produce flaking of the surface of the skin that can give extremities the appearance of a log, thus the term "tree barking." As with putrefaction, the skin will begin to split and eventually the skin and soft tissues under the skin are also lost.

Skeletonization

Both putrefaction and mummification eventually lead to skeletonization of the body. The loss of soft tissue progresses at a variable rate, depending on the environmental conditions and access of the body to a variety of animals and insects. The face and ends of the extremities are the first areas where bone is exposed, and the pelvic soft tissues are the last to be lost. While this process is usually measured in months, work performed at the Anthropologic Research Facility at the University of Tennessee in Knoxville has demonstrated that complete skeletonization of a person can occur within days under appropriate conditions.

Autolysis

Autolysis associated with decomposition occurs after cell death and is due to the actions of digestive and catalytic enzymes released from cells in the body. This term, "self-destruction," is advanced in certain organs, especially the intestinal tract and pancreas, due to their rich enzyme content. This is important to note because autolysis can mimic certain disease processes at the time of gross examination. Maceration of stillborns who were dead inside a uterus for days is a type of autolysis associated with moist sterile environments, appears as a red discoloration with skin slippage, and is not to be confused with trauma or disease.

Adipocere

A process that is rarely seen outside of exhumation cases. It is a chemical change that occurs in the fatty tissues of the body exposed to wet conditions. Adipocere takes many months to develop, and once it forms is extremely persistent, stopping the typical loss of soft tissue and skeletonization of the body.

Other Factors in Evaluating Postmortem Interval

When a deceased body has been exposed to insect or animal activity, we can make observations that might assist in the determination of the post-mortem interval. Maggots (the larval stage of blowflies) can be collected from the body and examined by an entomologist. You need to collect two sets of specimens. One set should be kept alive while the other set is killed. The entomologist will determine the exact species of fly from the matured living specimens and the likely time of death from the killed specimens. In persons who have been deceased for a prolonged period of time (months to years), an evaluation of the type and age of plants growing up through the body might also provide useful information.

Examination of the stomach and intestinal contents can also provide clues to help us in the determination of time of death. Our bodies digest food at a fairly predictable rate, depending on the quantity and types of foods consumed. Higher caloric content foods, such as fats, are digested slower than less complex foods. Larger meals remain in the stomach for a longer period of time. The rate of stomach emptying can be accelerated or slowed by natural disease, drugs, alcohol, or emotional stress. The description of the quantity, type, and condition of food in the stomach is part of the routine autopsy. This can be combined with information about that person's recent consumption of food obtained during the investigation to estimate the time from the eating of that meal until the time of death. Even without information about the last meal consumed, the stomach contents might provide clues as to when death occurred based on the types of food in the stomach.

Finally, markers found at the scene of death might provide the best information regarding the date or time of death of the individual. Observations regarding uncollected mail or newspapers, information from the telephone company about the last time the phone was used, dated sales receipts, or a noted change in habitual behavior might assist with determining the date of death. The clothing the person is wearing or the status of lighting in the home might provide clues regarding the time of the day or night that death occurred.

In conclusion, determining the postmortem interval can be difficult. Start with a broad window of death (last-known-alive to found-dead). Then use as many of the above elements as are available to attempt to narrow that window as much as possible. It is important to remember that there is variability of these observations and calculations, so keep an open mind and be willing to reevaluate your opinion based on new information as it becomes available.

Blanching lividity. Note the finger marks caused by pressing the blood away from the skin surface at the posteromedial aspect of the right thigh.

Fixed lividity at the lateral and posterior torso. Note the absence of lividity underlying the pressure region exerted by the arm against the weight of the chest. This child was discovered lying on his left side. This body was stored in a cold refrigerator for many hours prior to being autopsied.

This individual was placed in the morgue on her back the day before this examination. Note the fixed anterior lividity with partial sparing over with pressure points caused by lying on a wrinkled bedspread in her home for many hours prior to being discovered. Also note the early putrefaction consisting of slight green discoloration of the lower right abdomen.

Marked fixed lividity with dark purple spots known as "tardieu spots." Also note the vague chain pattern. It is not unusual for objects such as this to leave postmortem imprints.

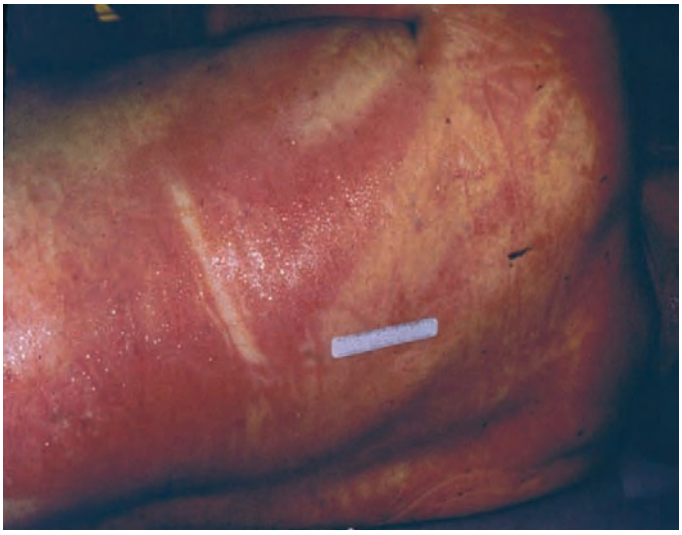

In carbon monoxide poisoning, the color of the livor mortis is more of a bright red instead of the typical red-purple color. This change can also be seen in cases of cyanide poisoning or in decedents who have died in cold environments. Note that the color differences are very subtle and it may not be possible to make a determination from visual inspection alone.

The decedent was found lying on his back with his head elevated on top of several pillows. He was in full rigor mortis at the scene and maintained this position through his arrival into the morgue.

Full rigor mortis. Note the morgue technician pulling tightly on this decedent's arm, which is stuck in a bent position.

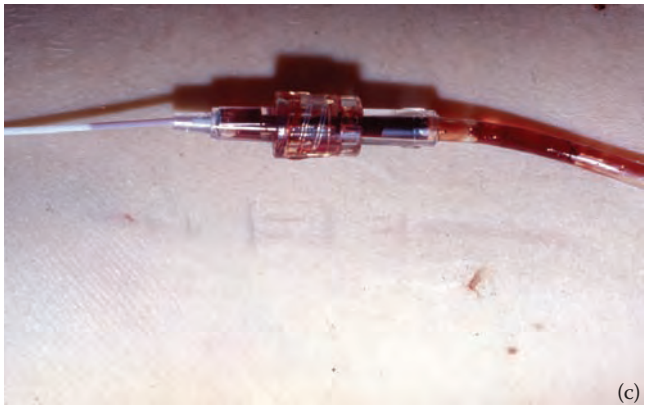

Congealing of fat. Adipose tissue at very cold temperatures will demonstrate increased viscosity. This might be mistaken for poor skin elasticity due to dehydration, which may or may not be present with congealing of fat. Note figure (b) shows indentations that have been retained following pressure exerted by fingers.

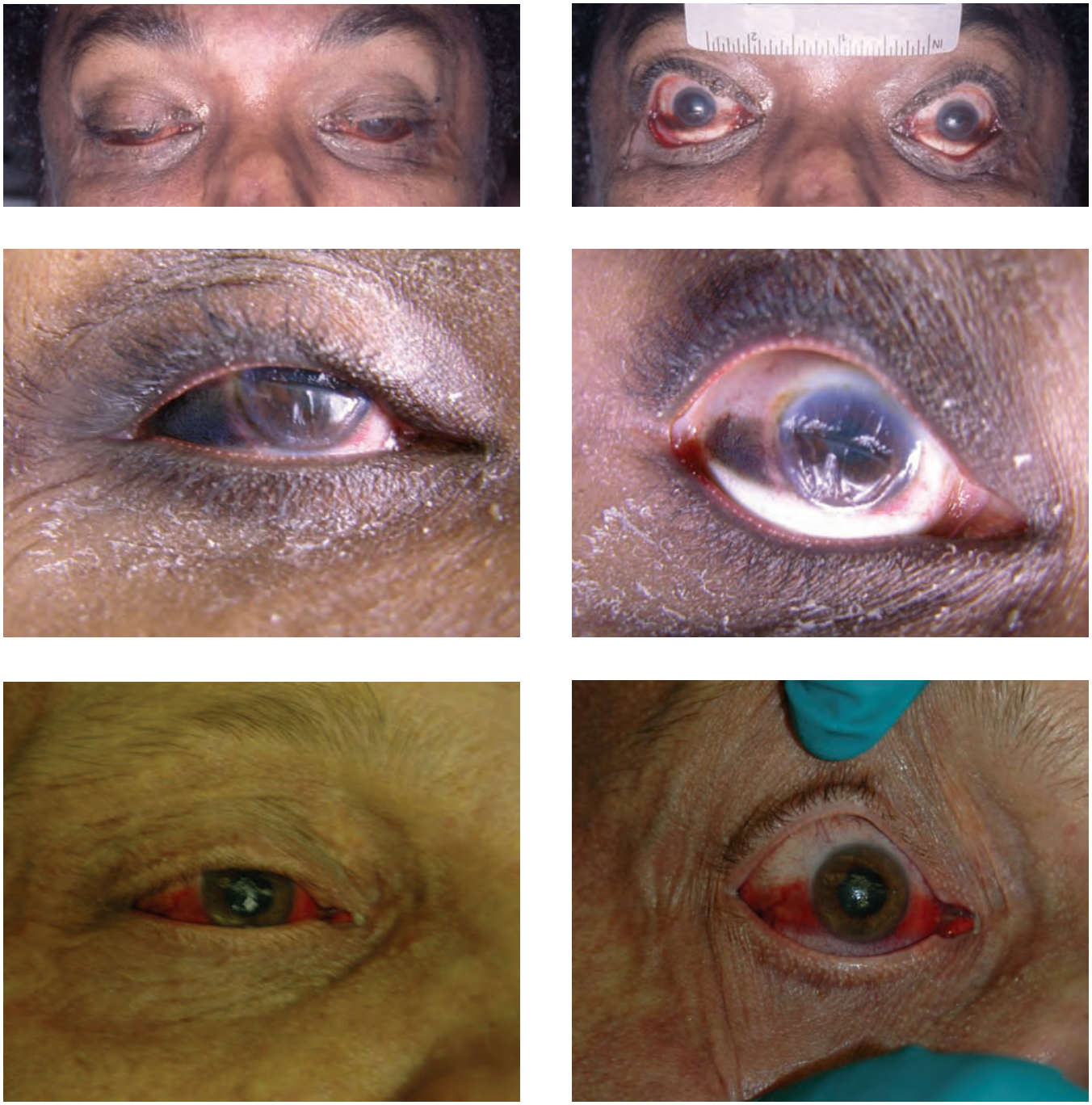

These are examples of "tache noire," which is a dark discoloration of a portion of sclera exposed to air. This is due to drying and is usually brown to red. This postmortem change may be misinterpreted as hemorrhage associated with strangulation.

This drying of the lips and tongue with dark discoloration is due to postmortem drying associated with mucosal exposure to air. This may be misinterpreted as an antemortem finding resulting from ingestion of caustic substances. The process of suicidal hanging caused the individual's tongue to stick out and become dry and dark.

This is also postmortem drying with dark discoloration of the scrotum and shaft of the penis. This may also be mistaken for an antemortem injury such as an abrasion. If there is doubt, one may make an incision to document underlying hemorrhage.

This is early mummification with drying of the hands and feet. Note the dark discoloration of the fingers with indentation from dehydration. Portions of the body with larger surface areas relative to underlying tissue mass will mummify more rapidly.

Early putrefactive change with green-brown discoloration of the lower abdomen. The first place for this to occur is typically the lower left abdomen above the cecal region.

This large hernia demonstrates early decomposition with slight green discoloration of the scrotum due to a portion of the large intestine extending into the scrotal sac.

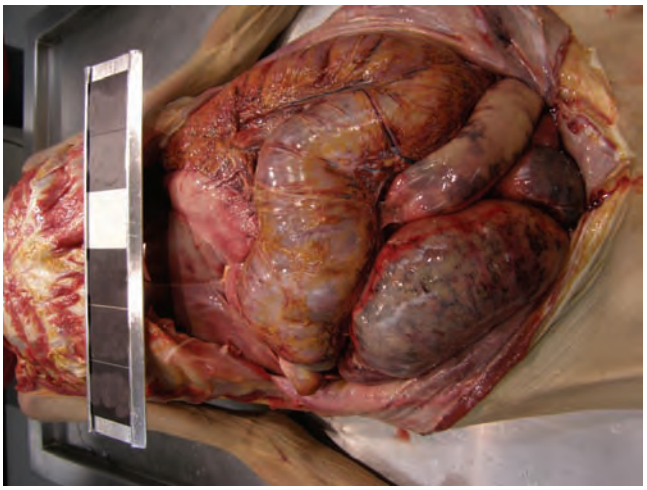

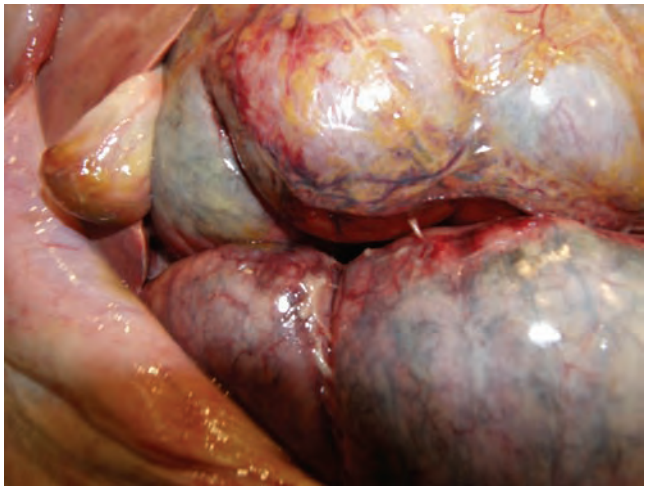

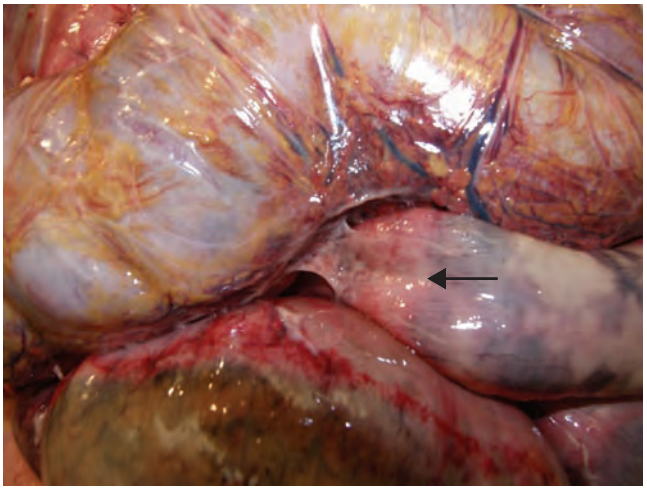

Disproportionate advancing putrefactive change of the abdomen due to acute peritonitis. Note the green discoloration of the abdomen and the absence of green discoloration elsewhere. The presence of infection with bacteria caused this regionally accelerated process. Note the parulent exudate at the surface of the intestines, indicated by the arrow.

This individual committed suicide by ingesting excess prescription medication and placing a loosely fixed plastic bag over the head. The moisture collecting at the face covered by plastic during breathing helped to create an environment where mold and mildew could form at the chin and face.

Postmortem hot water burns of the face in this individual, who was found partially submerged in the warm water of a bathtub. He suffered a cardiac event while taking a shower. It requires less heat to cause thermal damage to a dead body than to a living body. Note the sharply demarcated red region indicating his nose and mouth were below the hot water.

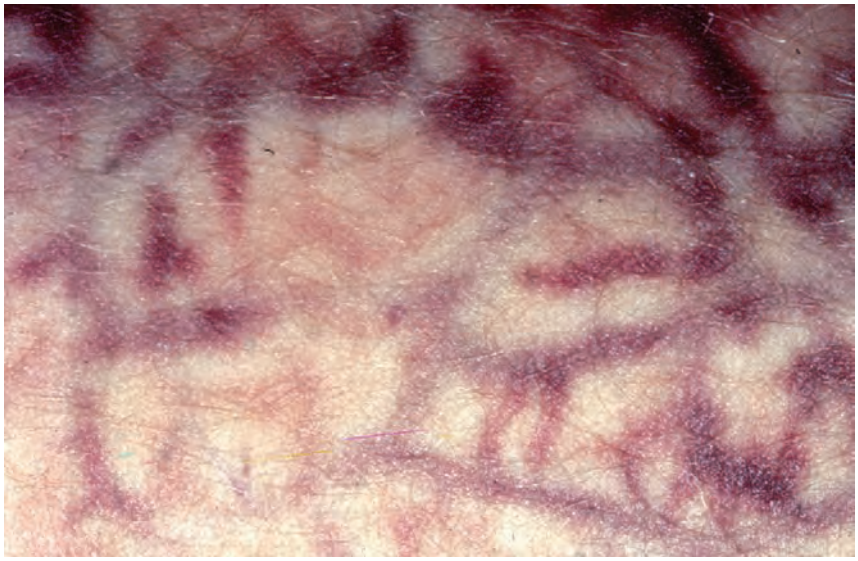

Marbling. This early presentation of putrefaction is predominantly caused by bacteria tracking through the superficial blood vessels causing pigment changes in blood and vessel walls as microorganisms digest the body.

Early to moderate putrefactive change with green-brown fluid-filled blisters. Such blisters may be mistaken for second degree thermal burns or aggressive antemortem bacterial skin infections.

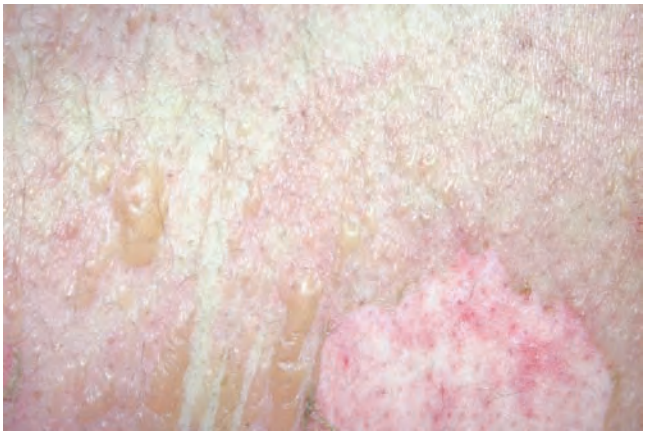

Early putrefactive change with clear yellow fluid-filled blister formation. The rupture of these blisters would appear as skin slippage.

Putrefaction with "purging." Note the red to brown fluid gurgling from the mouth and nose. This can sometimes be mistaken for an upper gastrointestinal hemorrhage.

Putrefactive change with bloating and expansion of the scrotum due to gas accumulation.

Green-brown-red discoloration with fluid-filled blister formation and skin slippage, scrotal enlargement from gas formation due to metabolism of proliferating microorganisms.

There is mummification of the fingertips with an expanded gas-filled blister and green to brown putrefied fluid in its inferior aspect.

Marbling and skin slippage; early to moderate putrefaction.

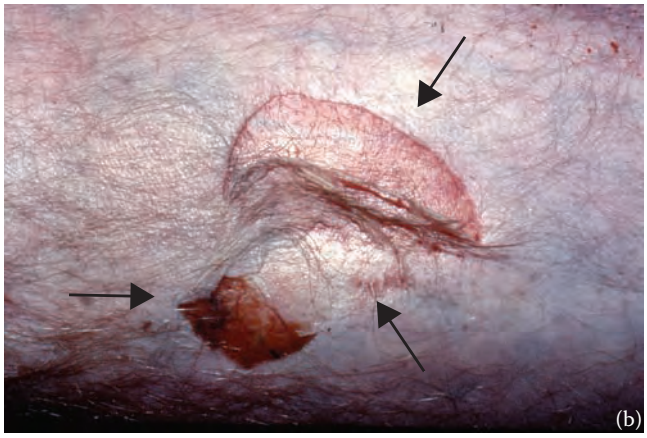

These images, (a) and (b), demonstrate skin slippage that occurred at different times. Note the regions of underlying dermis that are dry and dark, indicating older regions of skin slippage. Also note the other adjacent regions, which are moist, pale, and less dark, indicating shorter duration of underlying dermis exposure to air. This darker discoloration is due to more drying and longer exposure to the air.

Decomposing homicide victim with multiple decomposing blunt impact injuries. The combination of postmortem change with injuries may make interpretation challenging. Note the red to brown abraded contusions at the individual's face and head, with skin slippage and darker discoloration.

This individual was found lying on his left side with a mild to moderate degree of putrefaction. Note the puddle of purged fluid underneath his upper trunk and head (a). Also note the greater decomposition with green discoloration in the regions of dependent lividity where the body contacted the warm floor. Purged fluid such as this is often misinterpreted as the result of traumatic injury.

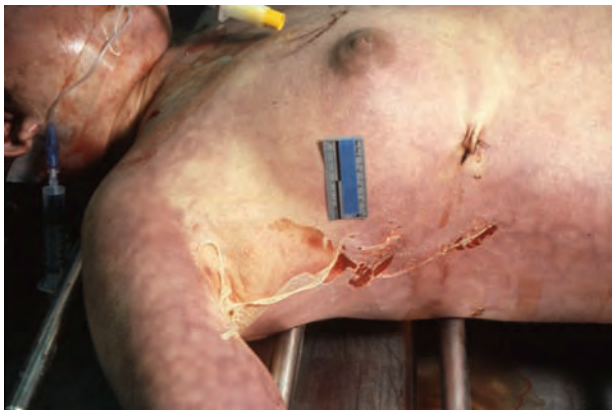

Accelerated putrefaction due to bacterial sepsis at the time of death. This picture depicts the left side of the patient's body with extensive skin slippage as well as green discoloration within the left infraclavicular area. Additionally, an endotracheal tube can be seen protruding from the mouth and taped to the side of her face. Such a depiction might lead one to conclude that overzealous paramedics had worked on a decedent who was decomposing. Marbling is also visible on the lateral and anterior aspects of the left arm. What this case really illustrates is the rapidity with which bacteria, already present within the blood stream at the time of death, can disseminate and propagate throughout the blood system, leading to accelerated postmortem putrefactive change.

Same case with front view of patient's face, also illustrating green discoloration around right infraclavicular puncture site and defibrillator marks.

This is the same case, demonstrating skin sloughage and exudation of hemolytic fluid from the external genital region.

Another case depicting a decedent with dark red-brown discoloration due to postmortem putrefactive change that was accelerated by the probable bacteremic state at the time of death.

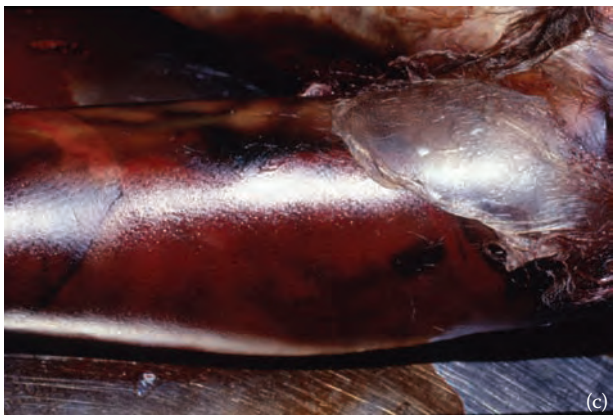

Skin slippage with confluent red to brown discoloration, moderate putrefaction. Note figure (b) shows multiple nitroglycerin patches in an individual with known significant heart disease.

Moderate to marked putrefaction with dark brown discoloration and early mummification.

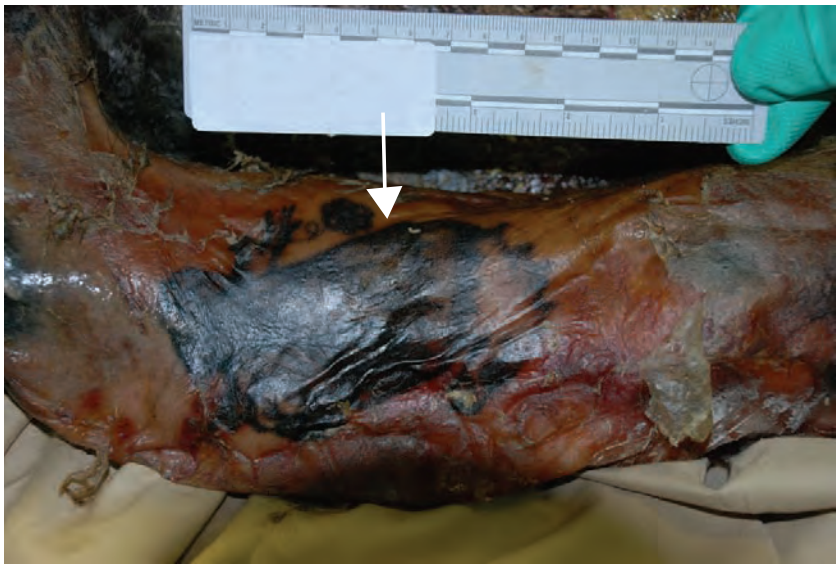

Moderate to marked putrefaction. Identification may sometimes be challenging with advancing decomposition. This individual had a tattoo on his left arm that was initially difficult to view due to putrefactive change. Cleaning of this area with removal of the superficial layers of skin made visualization and thus identification much easier.

Marked putrefactive change with dark brown discoloration and maggot feeding. Note the small circular perforations caused by maggots tunneling through skin and soft tissue.

Mummified fetuses retrieved from dried-out formaldehyde containers found within an abandoned building that once housed a doctor's office that closed more than 30 years earlier.

Advanced decomposition with mummification of the entire body. Note the dry dark leathery appearance of the skin.

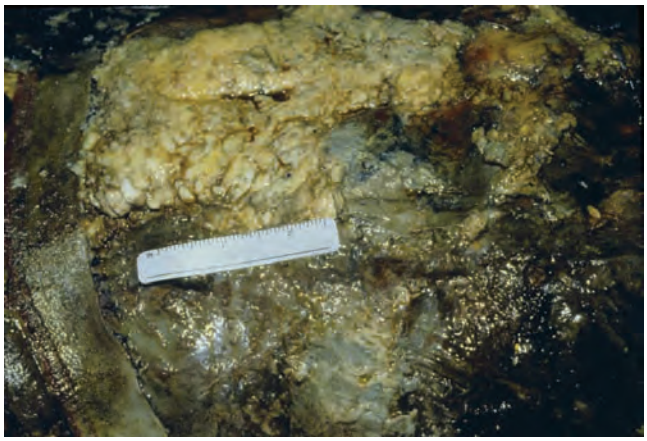

Adipocere is a decomposition process seen with immersions or damp, warm environments. The neutral fats are converted to oleic, palmitic, and stearic acids. Note the white/tan-colored adipocere, which has a waxy feel. In some areas you can see a light sheen of oil on the surface. Once formed, adipocere is resistant to further decomposition.

Skull found tucked under the bottom of a stairwell in the basement of a brownstone in Brooklyn, New York. Note the dried skin at the nose. There was resin and perforation of the cribiform plate. This was the historical remains of an Egyptian mummy skull that was most likely used in religious ceremonies. There were also commingled chicken bones, feathers, and wax.

Advanced decomposition with skeletonization from a body dumped in a wooded area approximately 1 year earlier. The last image demonstrates five toenails from inside a shoe.

This frozen individual was murdered approximately 1 month prior to being found. His body was in a slight state of putrefaction due to preservation by the freezing winter temperature. Once the body thawed out for autopsy, putrefaction advanced at a markedly accelerated rate.

Photograph taken at time of autopsy. Note the advancing putrefaction with skin slippage. Also note the scalp laceration.

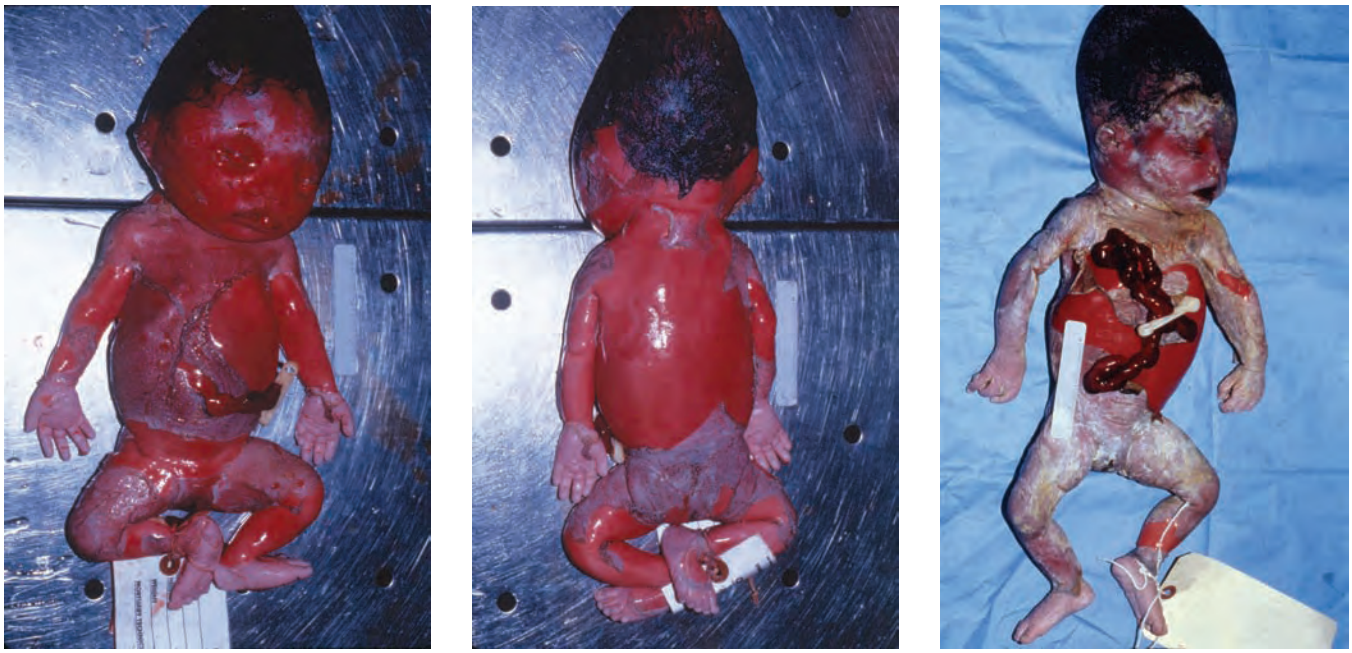

Macerated stillborns with red to brown discoloration. This type of decomposition is autolysis. The womb is normally a sterile environment and there should not be putrefaction unless there is an infection such as chorioamnionitis.

This child was discovered in the back of a garbage truck after being crushed by the trash compactor. The child was commingled with rotting food. Note the green discoloration of the skin due to putrefaction. Note the skull fracture with hemorrhage and dark discoloration due to putrefaction.

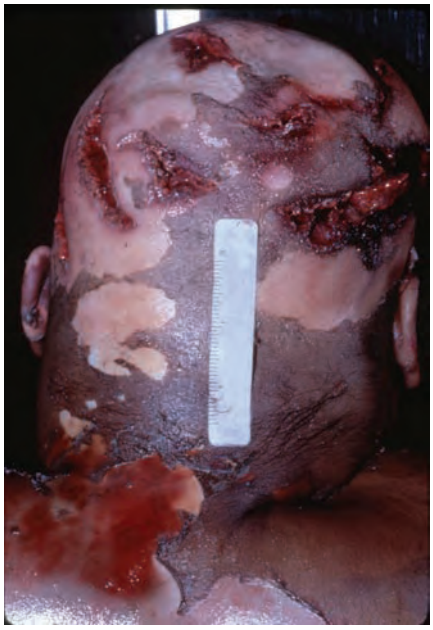

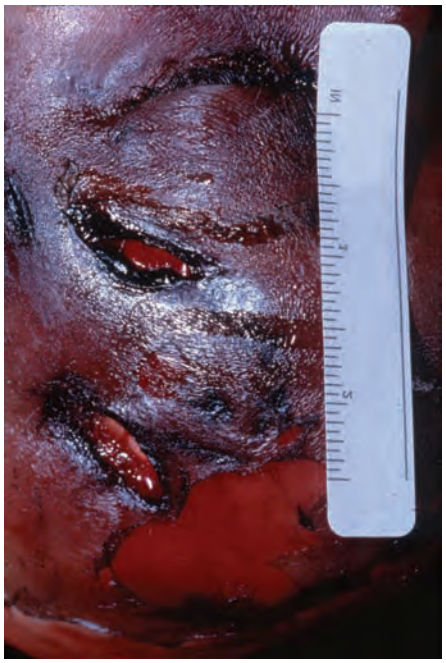

These slides show antemortem injuries that have been obscured by postmortem change. These individuals had multiple stab wounds and lacerations. Note the wound margins are dark and irregular due to drying. This may make wound interpretation challenging and at some point impossible. Also, animals and insects will often more readily feed from injured areas with exposed soft tissue and blood, further obscuring these findings.

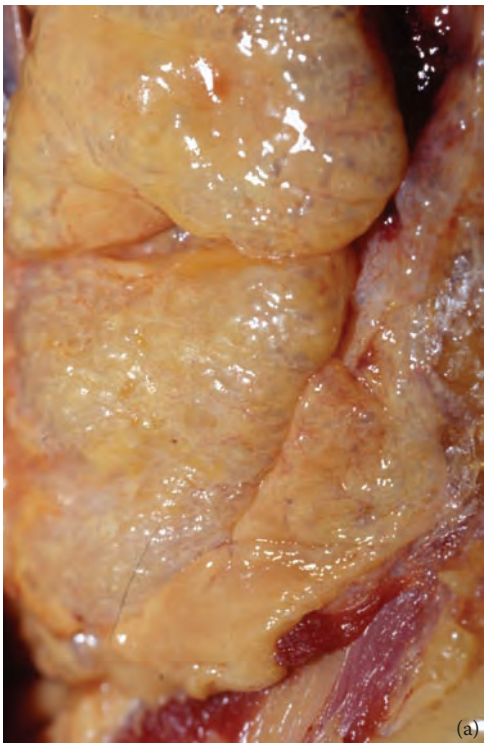
(a)

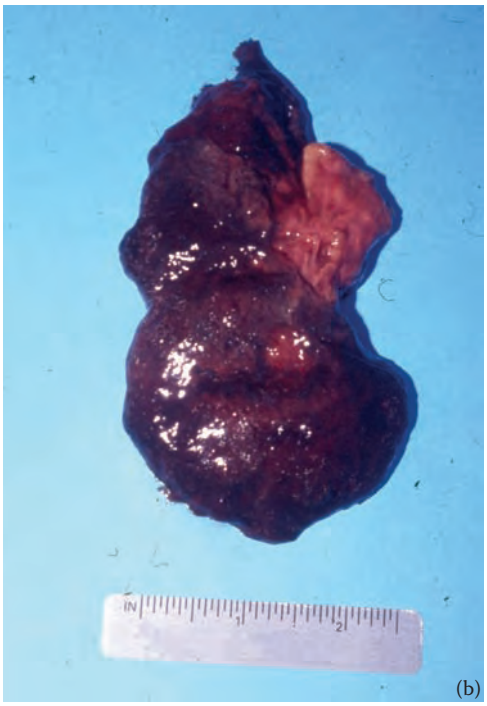
(b)

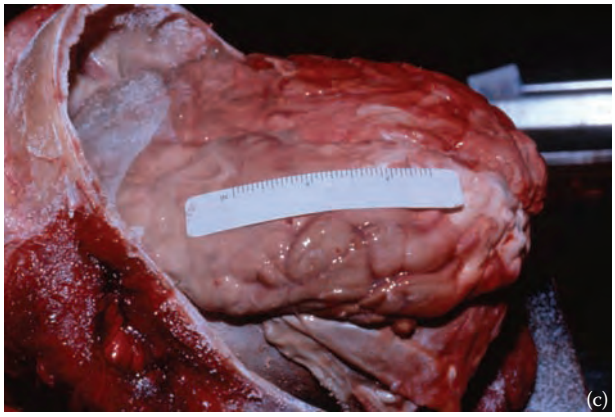
(c)

(d)

These are examples of moderate putrefactive changes of internal organs with softening, dehydration, shrinkage, and partial loss of architecture. In (a) note the crepitant adipose tissue due to expanding cavities of gas from proliferating microorganism metabolism. Also note in (c) the gray discoloration of the brain with partial loss of architecture. As central nervous system putrefaction advances, the brain will develop into a liquid, oatmeal-like consistency with few or more recognizable structures. Figure (d) is of a putrefying liver and (e) is of a putrefying heart.

(e)

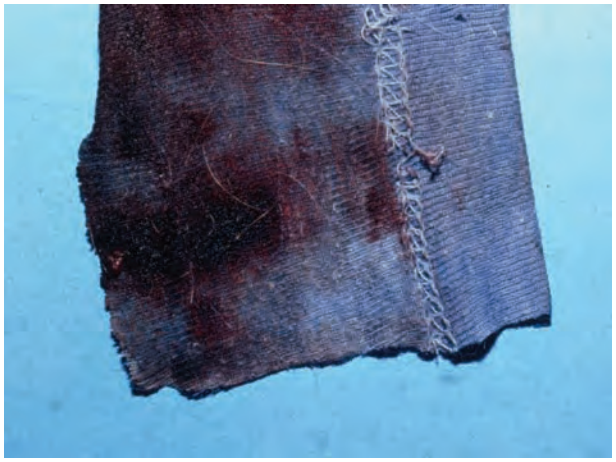

Note the fixed anterior lividity. The decedent was lying prone on a bed with a wrinkled comforter. Postmortem dog feeding occurred around the anus. Autopsy revealed the absence of large portions of intestines. Note that the torn clothing has dog hairs and blood staining. This case was first suspicious for homicidal violence. Large amounts of blood had drained from the body to the bedding, which was further spread about by the dog.

This is another example of postmortem dog feeding, which often initially involves the face, genitals, anus, and areas of wounds.

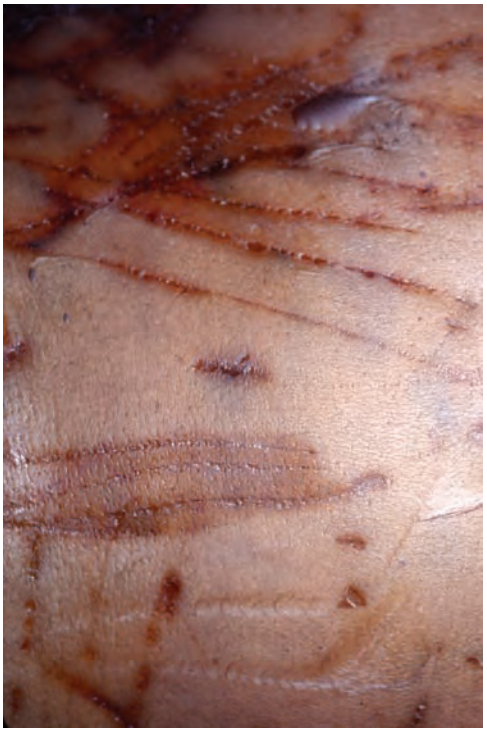

Postmortem animal feeding with dog claw abrasions.

Postmortem feeding by the decedent's dog.

Postmortem anemic superficial linear abrasions from a body retrieved from the Hudson River. This region of the body was noted to be scraping against a rocky surface just prior to retrieval from the water.

Larger carnivore activity is common in rural and suburban areas. In many cases, there may be activity from several different animals in the same environment. In this case, the damage appears to mostly be from the canine family (likely wild) with evidence of tearing of the skin and ribs. There is no obvious vital reaction, indicating the damage occurred postmortem. In these cases, one must consider whether the postmortem activity has obscured injuries sustained during life.

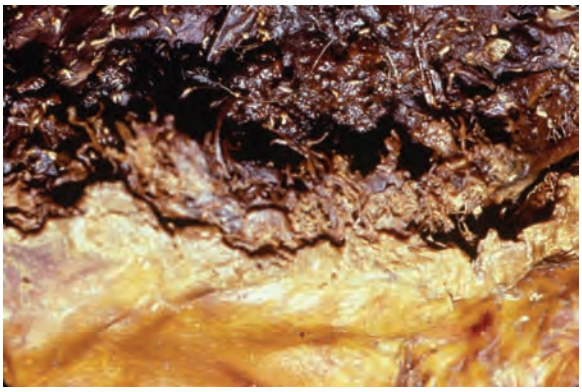

Advanced decomposition with mummification and animal feeding by the decedent's cats. Note the margin with claw marks and a scalloped border from feeding. One arm, a larger portion of the chest and most of the internal organs were absent.

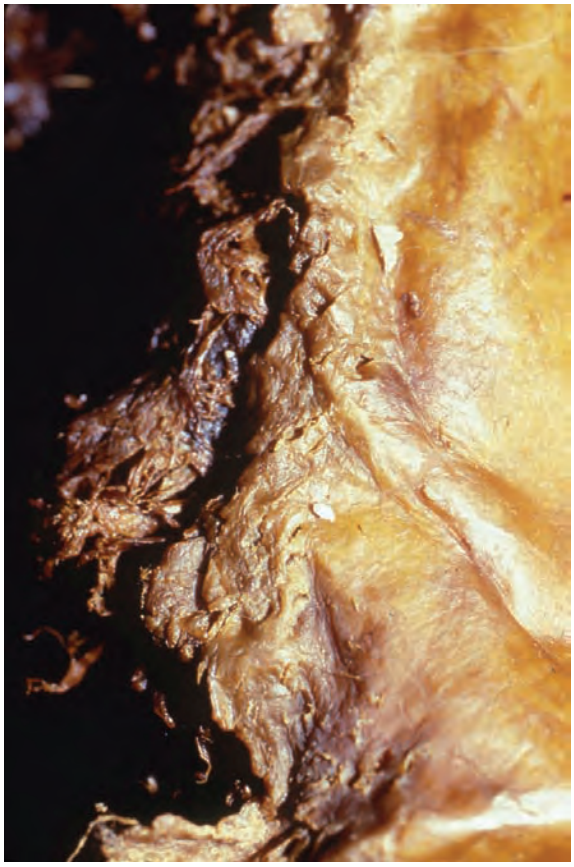

In this case the carnivore activity was caused by domestic pigs.

Postmortem feeding from fish and crustaceans found on these bodies retrieved from the ocean.

Postmortem feeding of the ear due to mice and rats.

Note the varying stages of blowfly development. Eggs have an appearance similar to grated cheese or sawdust. When the eggs hatch, maggots develop and are shown in this picture in varying sizes as small, white, and wormlike. The dark brown pupa cases from hatching flies are also apparent. This entire cycle is accelerated in hotter temperatures.

Embalmed body. Figures (a) and (b) demonstrate sutured incisions made by funeral directors to gain access for blood removal with addition of embalming fluid. Figures (c)–(f) demonstrate multiple "trocar" (a metal rod used to infuse embalming fluid) puncture marks through the body's surface and organs. These may be misinterpreted as injuries.

Pediatric Forensic Pathology

5

MICHAEL J. CAPLAN AND CHARLES A. CATANESE

Introduction

Pediatric forensic pathology is the name for the unofficial and poorly defined subspecialty of forensic pathology that focuses on the evaluation of sudden, unexpected, unexplained or traumatic deaths in children, infants, neonates, and even fetuses. By definition, therefore, this discipline requires specialized knowledge of and familiarity with the various stages in growth and development that occur during childhood, infancy, and intrauterine life, along with the differential vulnerabilities of these periods to various insults, and finally, the unique patterns of injury that characterize these different periods.

Perhaps one of the most illustrative examples of this concept is the assessment of nonaccidental or inflicted head trauma in infants and very young children. An understanding of this type of injury would not be possible without being thoroughly acquainted with the following concepts: (1) the elasticity of the infant's scalp, such that it could potentially sustain a serious blunt impact without necessarily manifesting the impact in the form of a subgaleal contusion; (2) the thin, pliable, unilaminar quality that renders the skull less susceptible to fracture but also allows it to transmit forces to the underlying brain more readily; (3) the broad, shallow skull base in infants, which facilitates rotational movement of the brain and lowers the threshold for diffuse axonal injury; (4) the incompletely myelinated infant brain, with a gray- and white-matter water content substantially higher than that of an adult, imparting a consistency of unset gelatin and making the brain more vulnerable to shearing forces; and (5) the top-heavy calvarium and the weak, underdeveloped neck muscles that fail to effectively dampen the oscillations that are initiated when rotational movement of the brain begins. It is the interface of pediatric and forensic pathology that allows the most complete and comprehensive understanding of these concepts to occur; however, very few individuals actually have such expertise in both fields. As a result, forensic pathologists may have to rely on pediatric pathologists and neuropathologists when they are evaluating difficult, complex, or problematic cases involving deaths in the pediatric population.

Of the many scenarios encountered by the pediatric forensic pathologist, none is as enigmatic as the seemingly healthy infant who is found unresponsive after being put to sleep sometime earlier by a parent, sibling, or other caretaker, either in a crib, a child or adult bed, a couch, or some other location, often in a prone (face-down) position, and with soft bedding, pillows, and clothing. Sudden infant death syndrome (SIDS), a term that was coined in the early 1960s to remove the stigma of an unforeseeable and unpreventable death of an infant from that baby's parents, has now come full circle and runs the risk of being used as a "wastebasket" term for any infant death that is not obviously explainable by a disease or injury. The time-honored external and internal autopsy findings of froth (pulmonary edema) at the nares, intrathoracic (thymic, epicardial, or visceral pleural) petechiae, moderately congested organs, and an empty urinary bladder, while characteristic, are in no way specific for or pathognomonic of SIDS, and have been observed in a wide variety of other types of deaths. Forensic pathologists and pediatricians are just beginning to gain some appreciation of the myriad of factors involved in achieving respiratory control, arousal, and autoresuscitation.

Even in this modern age of molecular medicine, SIDS remains largely an unexplored entity. Thus, the forensic pathologist probably makes the greatest contribution in this area by identifying which deaths do not lie within the spectrum of SIDS. To do this effectively demands a carefully conducted medicolegal death investigation, a meticulously performed autopsy, and the use of ancillary studies, including postmortem toxicology, microbiology, chemistry, radiography, and metabolic screening. Only after all of these studies have been completed and integrated with the autopsy, scene, and anamnestic findings can the pathologist make an opinion about the cause and manner of death.

Other situations necessitating specific pediatric forensic expertise include the assessment of sudden unexplained deaths in otherwise healthy children, either during sleep or while awake, at rest or following exertion; sudden deaths in infants or children who have undergone operations for congenital heart disease; unexpected or unexplained deaths during or following diagnostic or therapeutic procedures; complex child abuse cases involving a multiplicity of injuries separated over space and time; deaths from genetic diseases that might mimic child abuse, such as osteogenesis imperfecta; and

unexplained deaths following apparent minor trauma. In all of these scenarios, knowledge of the specific disease processes or related therapies is required in order to achieve the most complete understanding of the circumstances that culminated in the infant's or child's death. Other types of deaths affecting infants and children that are similar to those involving adults are typically fatal injuries sustained in motor vehicle crashes, either as occupants or pedestrians, although the considerable disparity in height, weight, and center of gravity often produces patterns of injury in infants and children that are distinctly different from those in adults.

Fatal child abuse presents primarily in two forms: (1) as the culmination of a repetitive pattern of inflicted injury, creating wounds of varying ages, in different parts of the body (what has been coined "the battered child syndrome"); and (2) as an impulse act by a caretaker who, following an act of urinating, defecating, or vomiting by the infant or child, lacks the self-control to prevent anger and frustration from translating into physical violence. The pediatric forensic pathologist must know what studies are necessary to perform in order to document not only what did happen, but to exclude all of the other reasonable possibilities that undoubtedly will be introduced by the defense in the adjudication of the case.

In some ways, the pediatric forensic pathologist often walks a diagnostic tightrope—deciding between SIDS and smothering, between inflicted closed-head injury and spontaneous intracranial hemorrhage, or between preeclampsia and maternal blunt trauma precipitating placental abruption. Toward that end, it is equally important to be aware of the limitations of a particular case and not offer opinions beyond what the circumstances and pathologic evidence allow. In some situations, "undetermined," while unsatisfying, is at the same time the fairest and most reasonable ruling until additional information is obtained that allows a more definitive ruling.

This child was brought to the emergency room and vigorously resuscitated. Note the marks left on the face as a result of endotracheal intubation and application of tape. It is important to document such findings and not confuse them with possible signs of child abuse, such as faint abrasions or contusions due to smothering. It is very important for clinicians to leave all tubes in place for postmortem (autopsy) confirmation of their placement.

Posterior lividity, which may be mistaken for injuries such as contusions related to child abuse.

Diaper rash. Note the red cutaneous eruptions. These indicate some degree of neglect.

Child abuse case with resuscitative efforts and perimortem sternal contusions due to chest compressions. Note other small contusions at the lateral aspect of the upper abdomen and lower chest. These were inflicted while the child was grabbed and flung about violently during the assault.

The same case showing perimortem contusions to the middle aspect of the chest due to resuscitation. Note the small amounts of subcutaneous hemorrhage overlying the sternum. Note the other segment of the "Y"-shaped incision is devoid of hemorrhage.

This case involved an infant who was found unresponsive in her crib. The case was referred to the medical examiner's office for autopsy. However, as the external examination was about to begin, this sutured midline incision was encountered. Further inquires revealed that the emergency department physician had authorized postmortem procurement of the infant's heart for heart valves. When this was discovered, the organ procurement agency was contacted, and released the heart to the medical examiner's office. The heart was found to have an anomalous origin of the left coronary artery from the right sinus of Valsalva, a rare yet documented cause of sudden death in both adults and infants. The issue of organ procurement remains a controversial and at times problematic one for the forensic pathologist. Consistently reliable methods for evaluating the coronary ostia have not been established in many tissue procurement agencies for the retrieval of hearts.

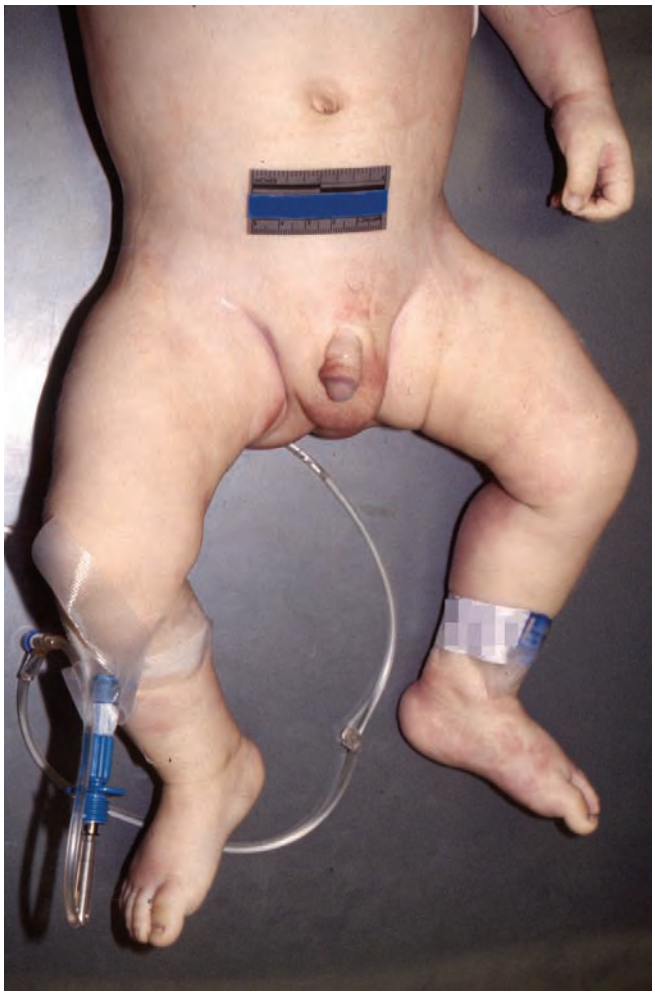

View of lower trunk, external genitalia, and lower extremities depicting a right-sided intraosseous catheter. This is a common mode of resuscitation in infants with extremely small, often collapsed blood vessels.

Hemorrhage with swelling of the thigh due to a misplaced femoral line.

Healing infected tracheotomy incision.

Scars to each wrist due to past therapy. Note image (b) has fresh needle marks as well. Therapeutic procedures should not be mistaken for injuries.

Small tear of the upper frenulum due to vigorous resuscitative efforts by an untrained person. Note the small bruise to the left side of the face, which had been caused by a fall several days earlier. The first phase of resuscitation was captured on a department store video camera when this child became lifeless following a seizure.

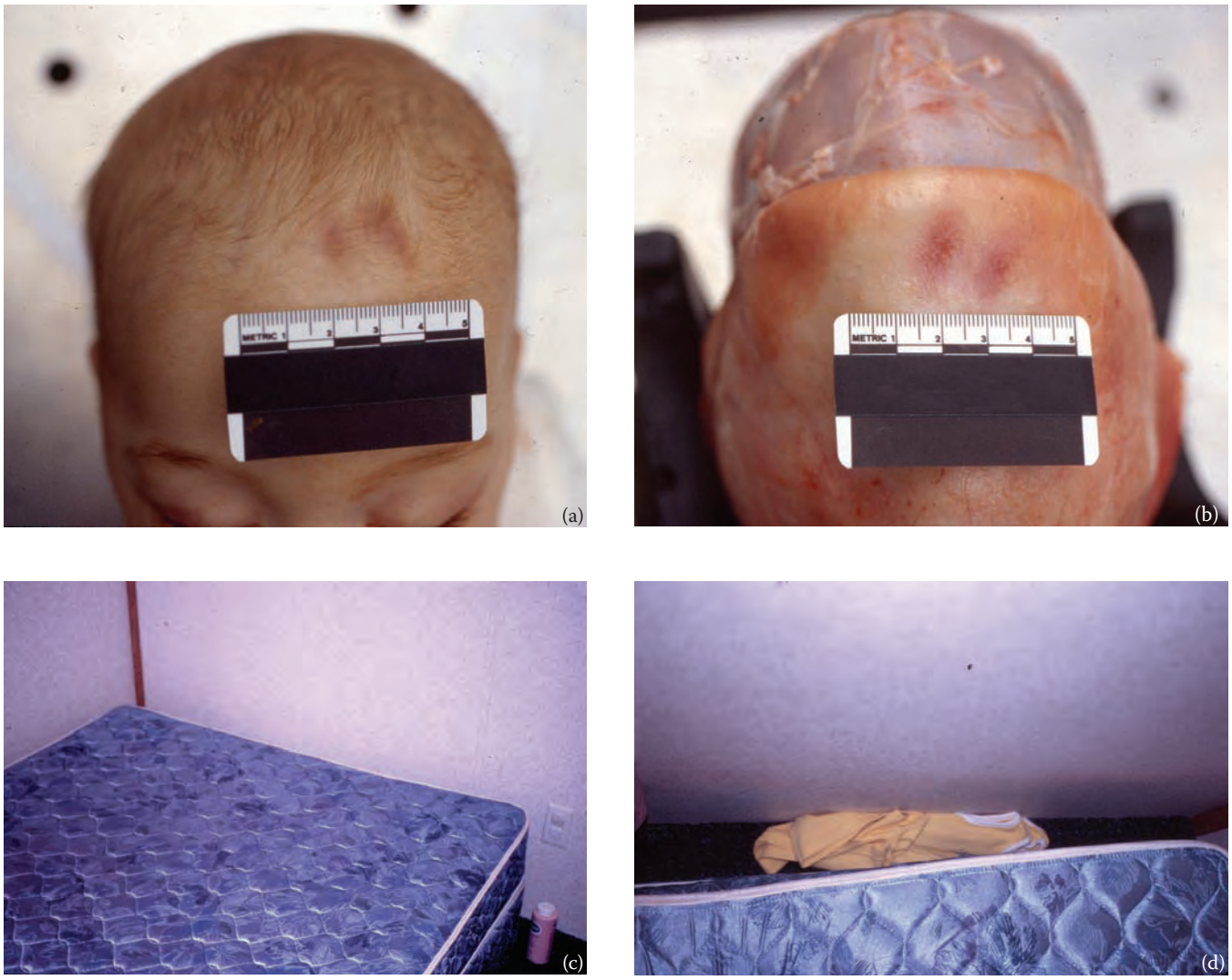

This case involved an infant who suffered an asphyxial death. He was found with his head compressed (wedged) between an adult bed and a wall. He was pulseless, apneic, and asystolic, and despite resuscitative efforts could not be revived; (b) depicts the frontal and temporal regions of the scalp with two ill-defined contusions within the frontal subgaleal region; (c) shows the relationship between the mattress and the wall. The mattress was separated from the wall by at least several inches, allowing the infant to fall into this space and become compressed; (d) is a view looking toward the floor, where the infant was found lying wedged between the wall and the mattress.

This child was brought to the Emergency Department (ED) by parents who spoke a foreign language. The child was reported to be found dead in a crib. The ED physicians believed that this death could be a result of SIDS but there was no initial scene investigation. Note the posterior patchy red lividity. There is also a faint Mongolian-type birthmark within the left gluteal region. Further scene investigation by the Medical Examiner's office and police department revealed that the child had actually slid between the crib railing and mattress, suffering an asphyxial death. There were no marks on the child's body at autopsy. The mattress was from another crib and was slightly smaller, allowing the child to slide between the railing and mattress. Note the doll demonstrating how the child was found at the scene.

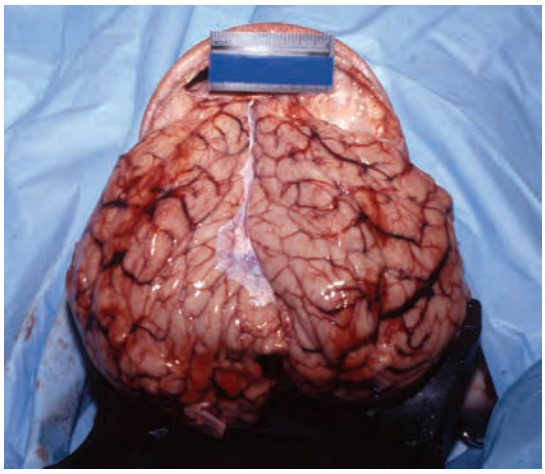

This image demonstrates brain swelling with flattening of gyri and obliteration of sulci due to hypoxic-ischemic brain injury. This 7-month-old female infant was found unresponsive in a car seat that fell off of an adult bed, with her chin tightly opposed to her chest, satisfying the criteria for positional asphyxia. She survived for several days and was ultimately declared brain dead. No other injuries were identified upon autopsy examination.

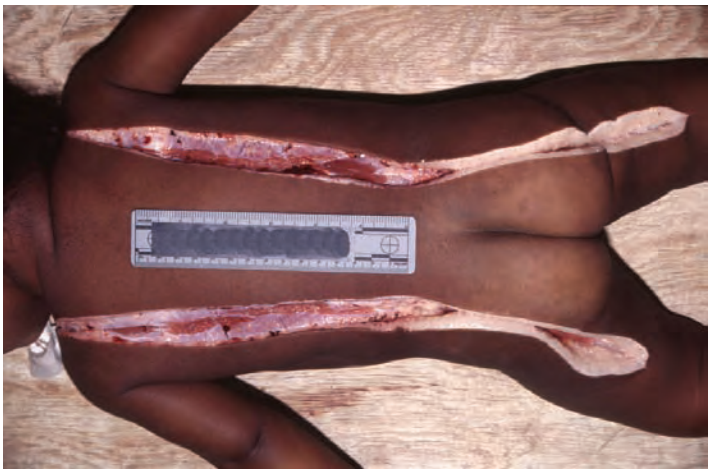

This 9-month-old child with a history of asthma was found unresponsive in her crib. Autopsy demonstrated multiple mucous plugs within the medium-sized airways, along with well-defined pathologic features of chronic asthma. No injuries were found. Death was certified as complications of bronchial asthma. Although unusual in a child this young, asthma can indeed be fatal. These bilateral vertical incisions made at the time of autopsy failed to demonstrate contusions.

This was the body of a completely healthy female infant. Her mother, who had left her unattended for an unknown period of time with her 18-month-old brother, found her submerged in a bathtub. The mother reportedly left them both in the tub without water and claimed the brother had to have turned on the faucets. When police attempted to reconstruct the scene, the older brother was unable to turn the faucets. Although the cause of death in this case was certified as drowning and the manner homicide, the mother was released on probation.

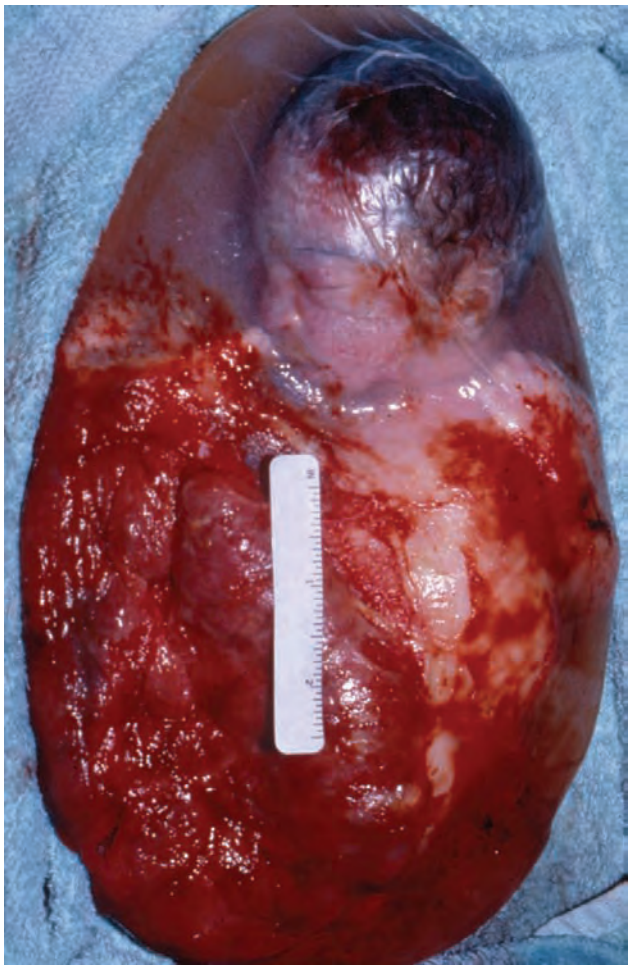

This normally developed neonate was delivered with the placenta enclosed in an intact amniotic sac. The baby died of asphyxia and the manner of death was certified as natural. The child was delivered at home and the young mother did not have medical knowledge to know that the membranes had to be immediately opened following delivery. She noted the child to be moving within this placental unit.

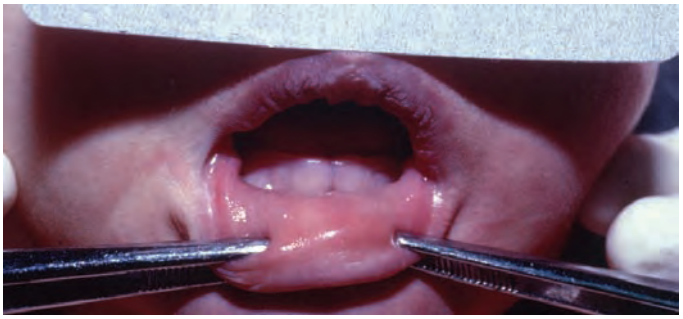

It is important to examine the frenulum and mouth region internally and externally in infants to rule out possible findings associated with smothering. Both of these cases show no injuries.

Healed frenulum laceration. This individual fell as a small child and accidentally struck her face on a kitchen cabinet.

Intact frenulum with postmortem drying of the lips. This does not indicate child abuse.

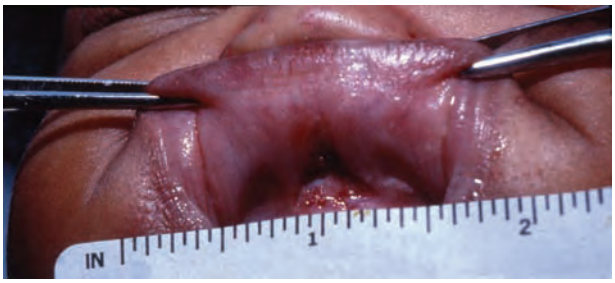

These children have varying degrees of abrasions to their face with trauma to the lips and torn frenula. All are cases of homicidal smothering. It is important to also realize that a homicidal smothering may occur without leaving marks of any kind.

These images demonstrate a case involving a several-month-old infant who was found unresponsive in her crib. She had no known previous illnesses. If possible, it is always important to photograph infants in their original clothing, which allows for more accurate reconstruction of the terminal events (i.e., stains of blood-tinged fluid.) View (a) demonstrates clear fluid emanating from the left nostril with a faintly frothy quality, probably indicative of pulmonary edema, which is a common finding in victims of SIDS, particularly when attempts are made to resuscitate. The elastic nature of infants' lungs and the vigorous resuscitation efforts collectively contribute to the formation of this edema fluid, which is often blood-tinged due to rupture of small capillaries. Oftentimes, the blood-tinged fluid is misinterpreted by police personnel as being suspicious of foul play. It is important for pathologists and emergency medical services (EMS) personnel to educate people with nonmedical backgrounds regarding the distinction between blood-tinged edema fluid and actual clotted blood.

This case involved sudden death in an otherwise healthy infant. Notice the confluent burns and abrasions on his left anterior chest wall, extending to the midline (a). These marks are indicative of defibrillation attempts from resuscitation. Additional therapeutic interventions can be seen, including an endotracheal tube, along with electrocardiogram (ECG) electro pads. Notice also the blotchy red discoloration of the forehead and face (left side slightly greater than right). While one might be tempted to conclude that this is livor mortis, it is important to realize that vigorous resuscitative efforts may also alter the patterns of vascular congestion; in the same case, (b) demonstrates an intact upper frenulum with dried blood-tinged secretion around the left side of the mouth and no trauma.

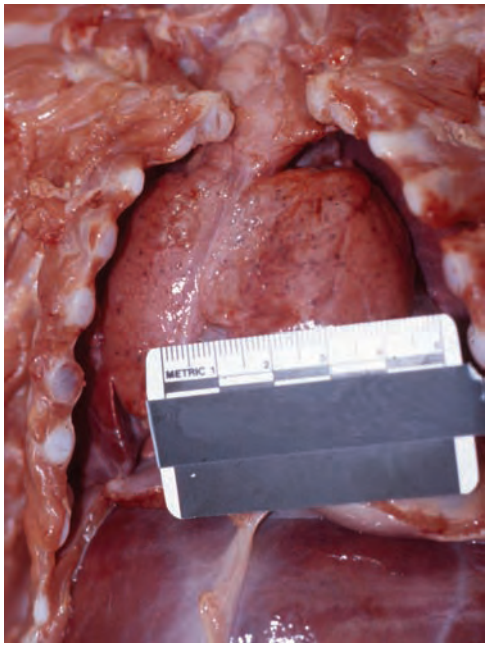

This infant was found unresponsive in his crib and could not be resuscitated. Maternal history was significant only for group B streptococcal infection. This picture depicts prominent petechiae on the anterior capsular surface of the thymus. Petechiae involving the thymic, visceral pleural, and epicardial surfaces are extremely common findings in sudden infant deaths, including deaths that have been ascribed to SIDS (up to ~85 % of SIDS fatalities). The precise mechanism underlying the formation of petechiae remains elusive, but is believed to be related to the negative pressure created by terminal gasping in infants. However, it is well known that these petechiae are not in any way specific or pathognomonic for SIDS and may be found in deaths due to many other causes, including definable natural disease conditions, accidental injuries, and inflicted injuries (homicides). In this case, the petechiae, which appear as pinpoint hemorrhages, are particularly striking.

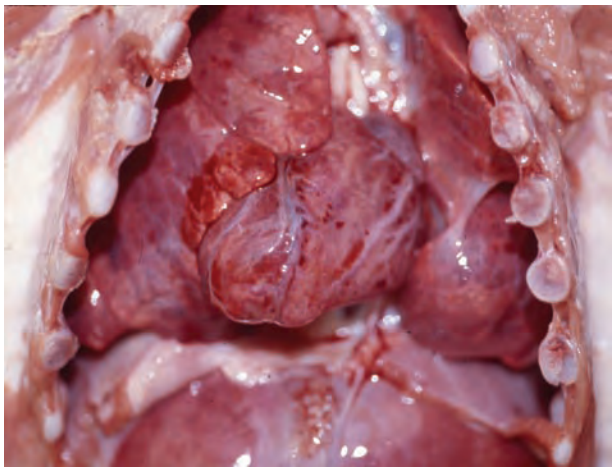

This is a view of the thoracic contents of an infant who was found unresponsive while sharing a bed with his mother while in a supine position. She called 911 and paramedics arrived to find him apneic and pulseless. He could not be resuscitated. Pertinent history included prematurity, maternal group B streptococcal infection, and a recent upper respiratory infection. The only significant findings at autopsy were a moderately cellular mononuclear leptomeningeal exudate consistent with a viral meningoencephalitis. However, there was no significant brain swelling. This fatality occurred in the early to mid-1990s, and the death was ascribed to SIDS. If a similar case were encountered today, bedsharing in an adult bed would probably be listed as a risk factor in part two of the cause of death statement and the manner of death would be undetermined. This picture depicts prominent petechiae on the anterior epicardial surface of the heart.

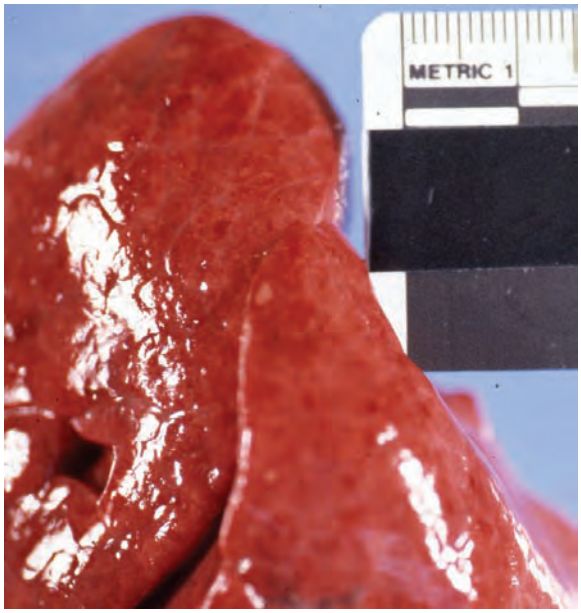

This case involved an infant who was born prematurely at 35 weeks gestation and was found prone in a bassinet with his face between a "covered" adult-sized pillow and the corner of the bassinet. The findings of this case were not sufficient to render a diagnosis of mechanical asphyxia, but asphyxia was of sufficient concern to denote the cause of death as sudden unexplained death in infancy and the manner undetermined. This picture shows sparse, inconspicuous visceral pleural and thymic petechiae, which were the only gross findings at autopsy.

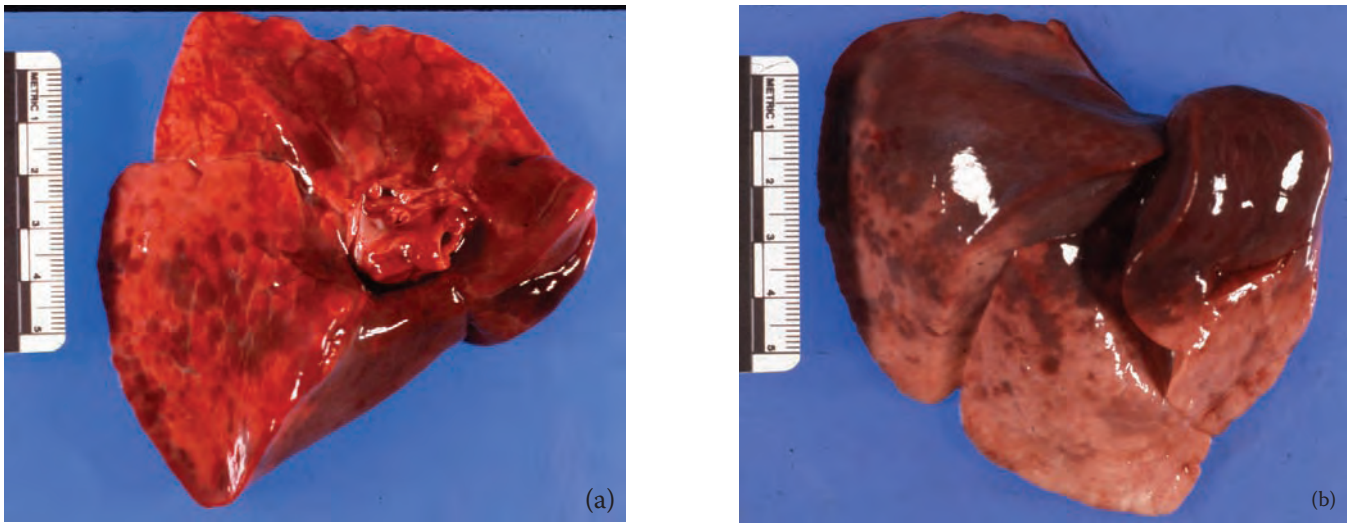

View of the mediastinal surface of the left lung (a) demonstrates confluent subpleural hemorrhages within the lower lobe, likely an artifact of vigorous cardiopulmonary resuscitation (CPR). Note the way the hemorrhages conform to anatomic boundaries, similar to what is observed in aspiration of blood; (b) demonstrates the costal surface of the lung, with blotchy subpleural hemorrhages, some slightly larger than petechiae. Note also the confluent congestion within the posterior aspects of the lung, most likely a result of postmortem hypostasis, or settling of blood due to gravity.

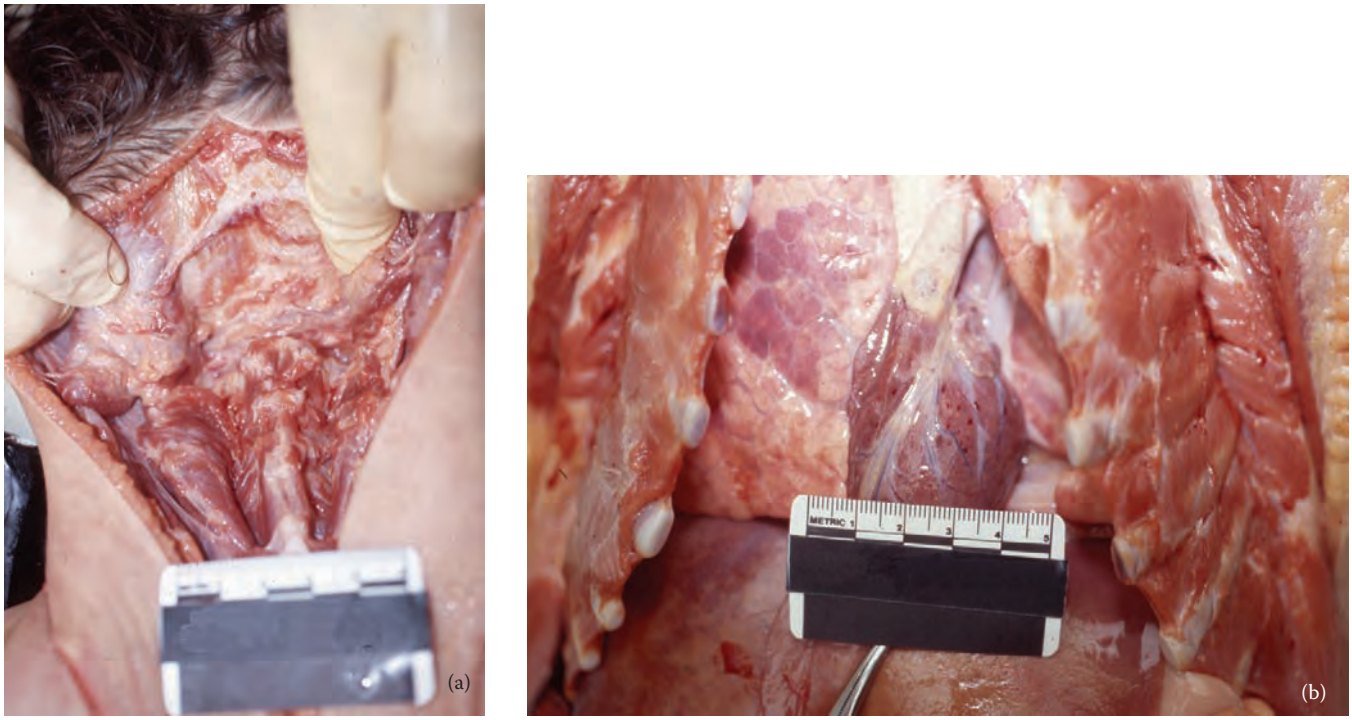

This was a case involving the enigmatic death of a 2-year-old child who was previously completely healthy. She had finished a course of amoxicillin for otitis media 3 days prior to her death. On the day before her death, she apparently fell off a "teeter-totter" and struck her head on the ground, which was a muddy, grassy surface, but she never lost consciousness and experienced no mental status changes. She could not be aroused from an afternoon nap and could not be resuscitated by pediatric life support protocol. The gross autopsy findings were unrevealing, including absence of injuries of scalp, skull, and brain. The only finding in addition to thymic, epicardial, and visceral pleural petechiae (b) was fibromuscular hyperplasia of the atrioventricular (AV) nodal artery, which was discovered by an expert cardiac pathologist who consulted on the case. This picture depicts a posterior neck dissection (a), demonstrating soft tissues and skeletal muscles that are entirely free of injury. In this case, anterior neck dissection was negative for injuries as well; (b) depicts prominent epicardial petechiae on the anterior surface of the heart. This case illustrates that the findings of intrathoracic petechiae are not confined to infants but can also be seen in young children.

This child died of complications of dehydration following a prolonged viral gastroenteritis. Note the sunken eyes.

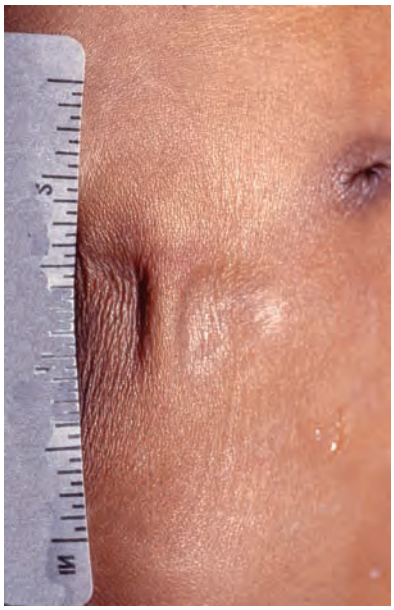

Poor skin elasticity (turgor) is demonstrated in these infants who were refrigerated in the morgue overnight. These infants were shown not to be dehydrated by vitreous analyte determination and antemortem hospital chemistry testing. The poor skin elasticity in these cases is due to the congealing of fat as a result of postmortem refrigeration.

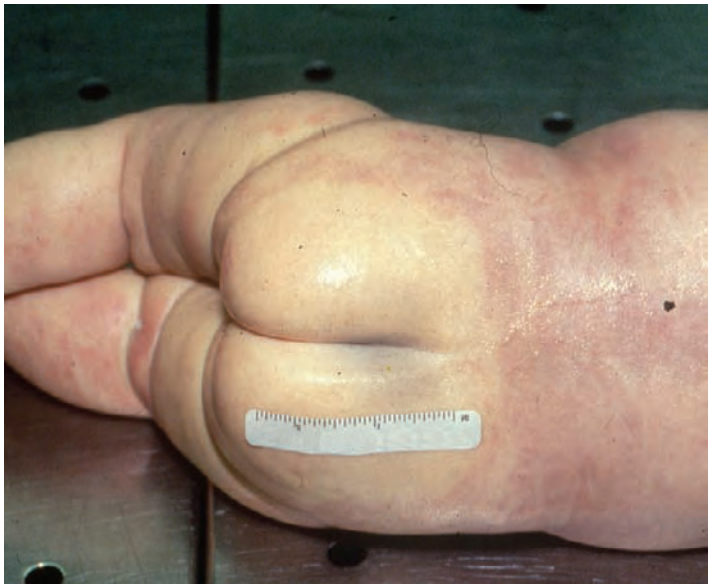

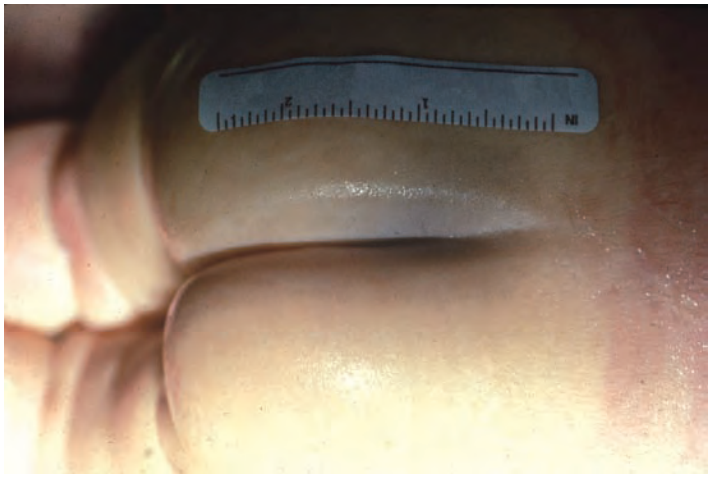

Mongolian spot. This is a birthmark and is sometimes mistaken for a contusion. During the time of autopsy it is easy to differentiate between the two by making an incision and documenting the presence or absence of underlying hemorrhage. The author has had at least one case where a contusion was demonstrated subjacent to a Mongolian spot.

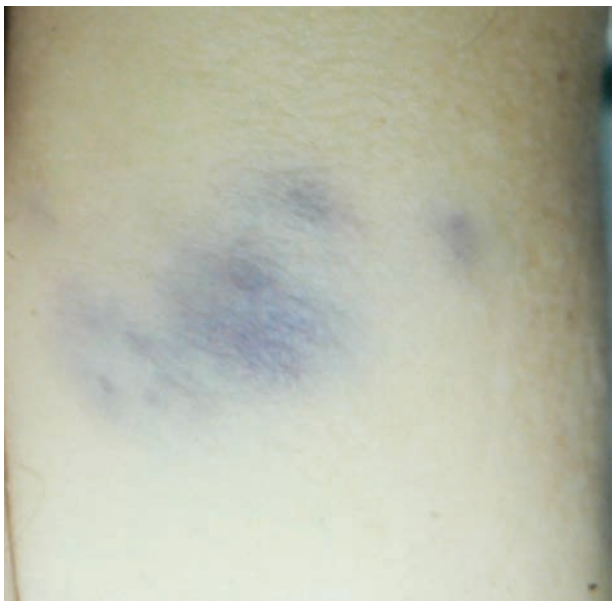

Recent contusion.

Old contusion.

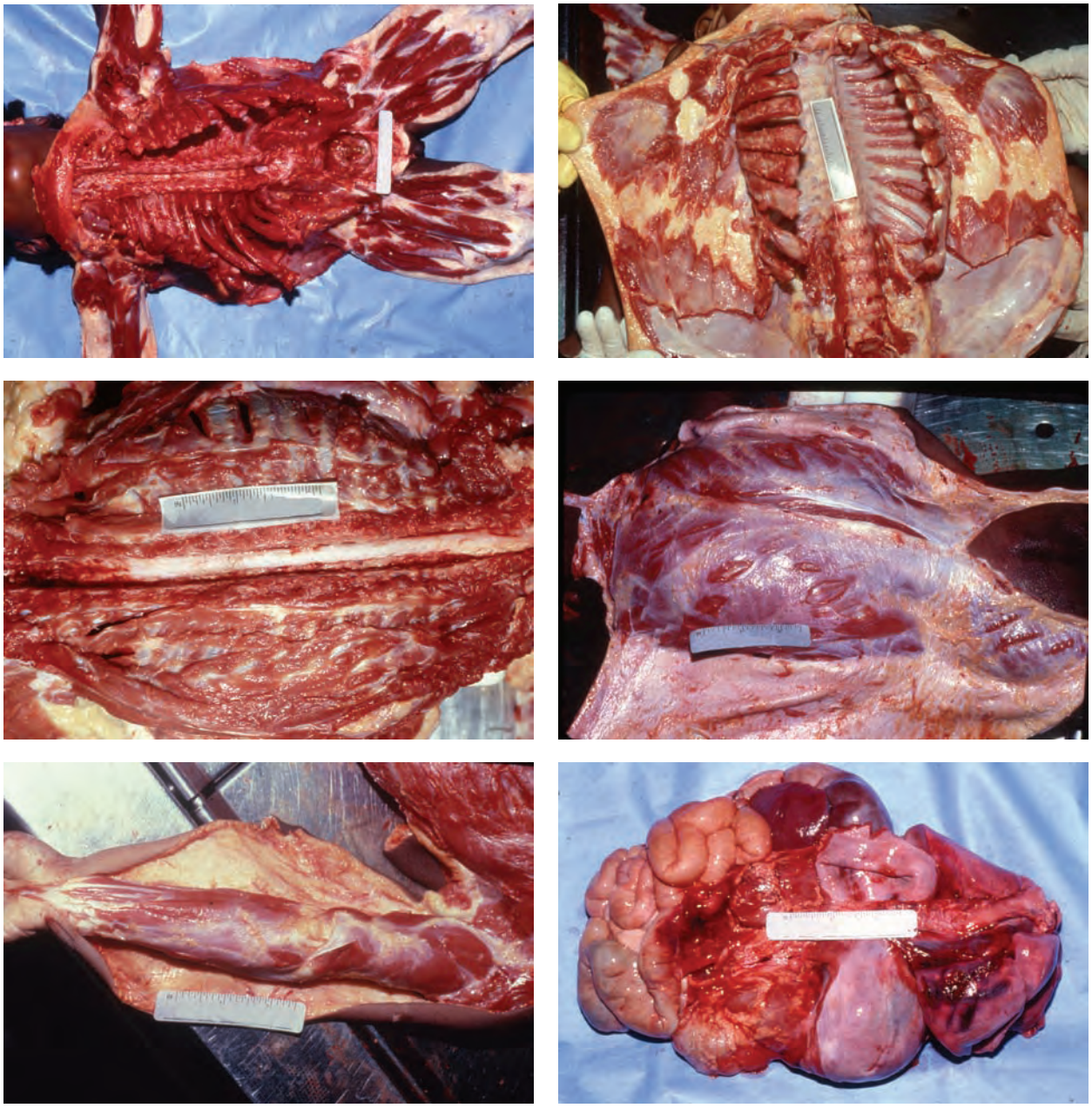

While performing autopsies on suspected child abuse cases it is important to be very thorough and it is recommended to incise the extremities and back to further document deep contusions, particularly in darker-skinned children. When examining the rib cage it is important to strip the pleural surface, separate each rib and evaluate each one separately. It is also good practice to do a posterior neck and back dissection while examining each rib as well.

This figure demonstrates how the body is sutured prior to funeral home release.

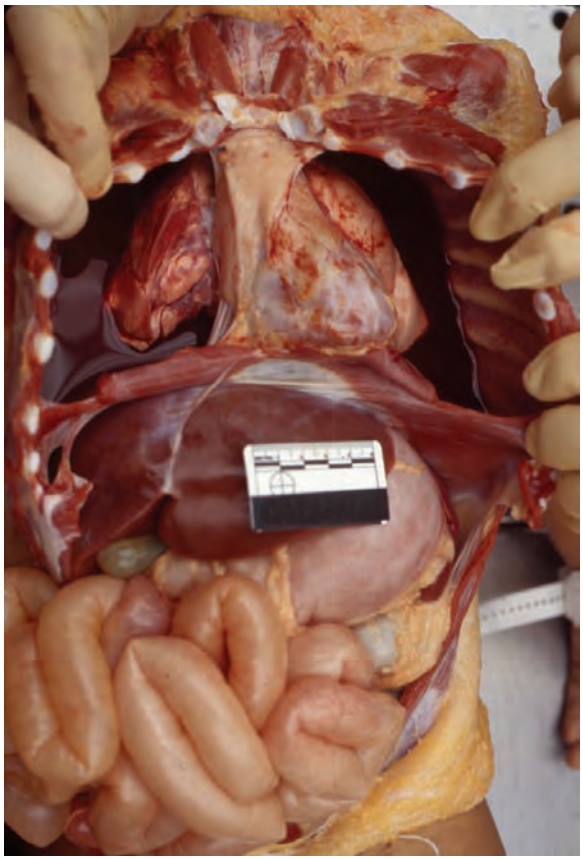

This 2-year-old child was brought to the ED by his mother's boyfriend after he was found unresponsive. The only observed external injury was a small faint contusion to the child's forehead. Autopsy revealed a fracture of the lumbar spine with a transected descending thoracic aorta and an old healed fractured clavicle. This picture demonstrates bilateral hemothoraces.

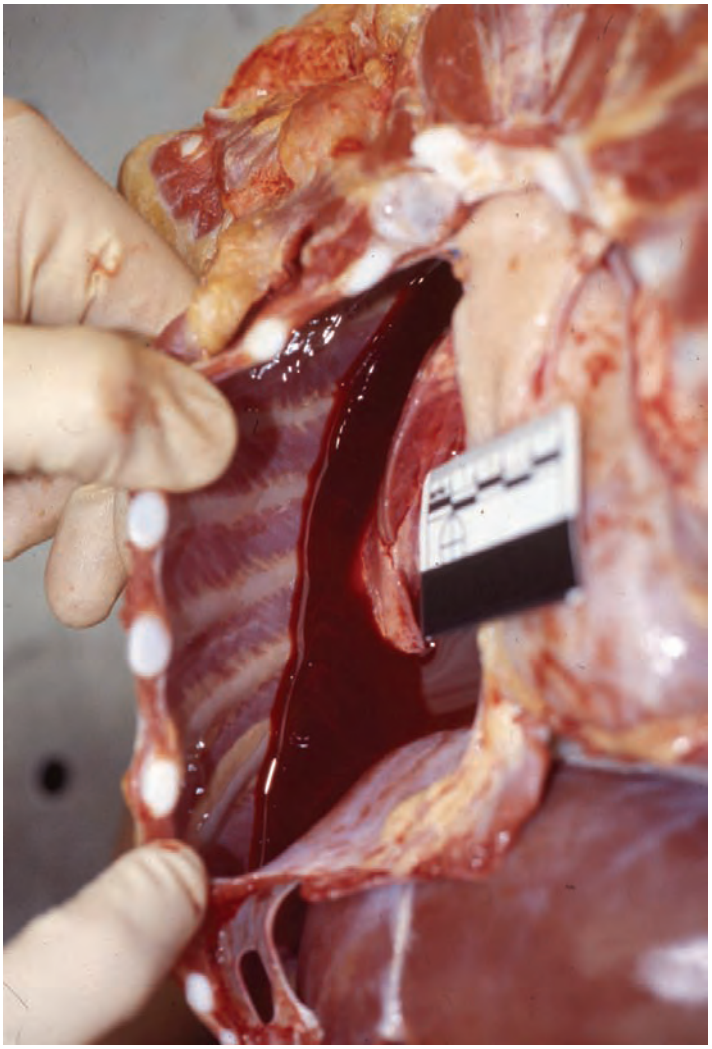

Right hemothorax due to aortic transection.

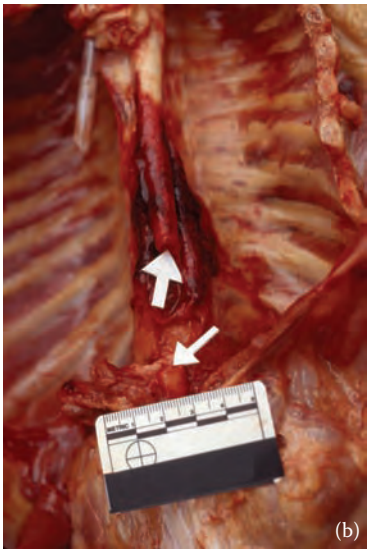

Image (a) demonstrates the volume of blood collected from the bilateral hemothoraces (approximately 165 mL in total). In a case such as this, the mechanisms of death are probably multifactorial, including not only blood loss but also a component of spinal cord or sympathetic nervous system dysfunction; (b) depicts transected free ends of lacerated descending thoracic aorta, marked with arrows; (c) is a panoramic view of lumbar fracture, which resembles a "hinge" at L1-L2.

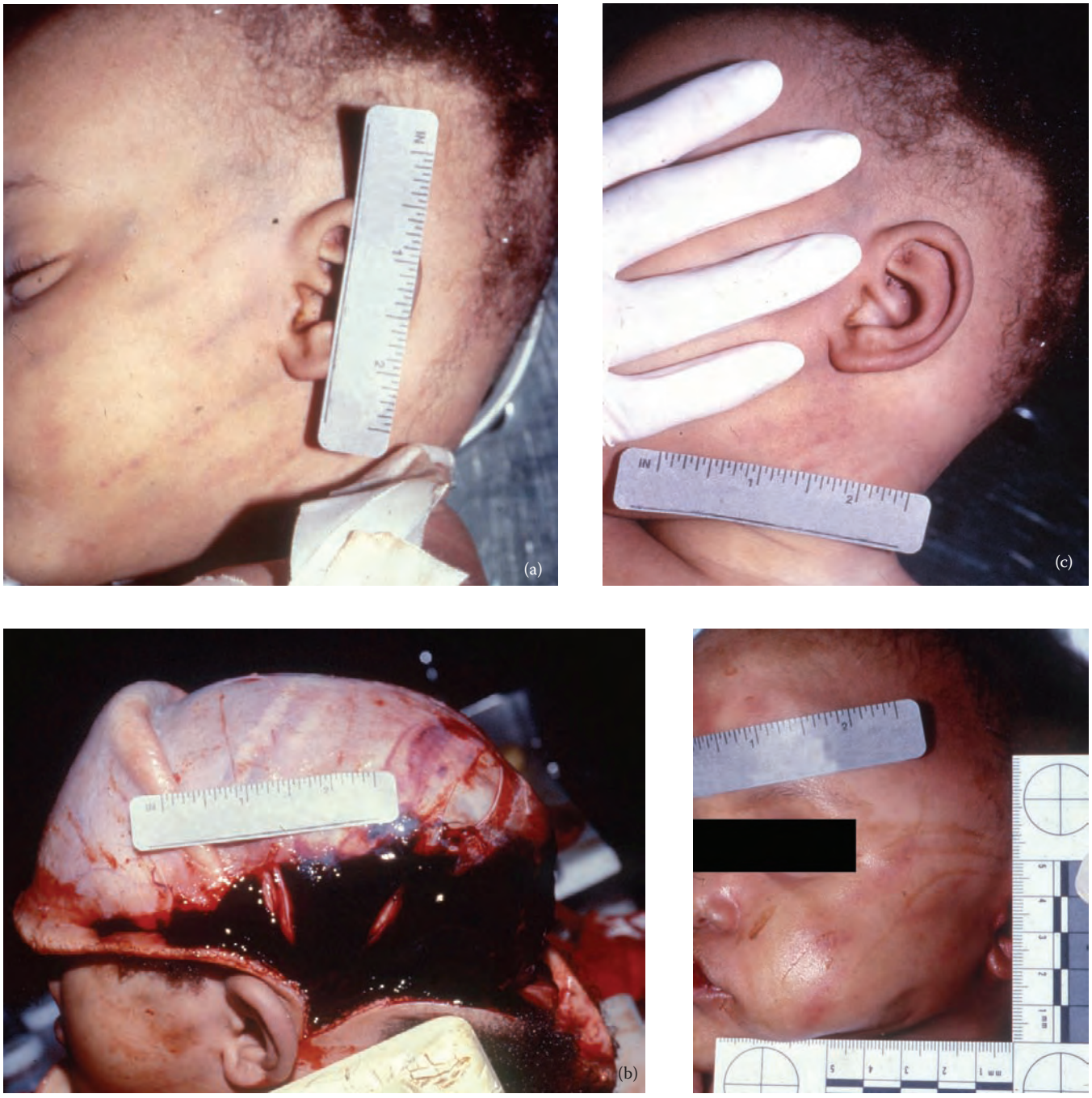

This child died from multiple blunt force injuries. (a) Note the hand pattern with parallel contusions across the child's face. The contusions are more obvious when depicted following the intermastoid incision and scalp reflection (b). Note the excessive dark red blood clot at the inferior portion of the side of the head. Also note that this is not as visible when viewed from the external skin surface. Deep contusions may be difficult to see when viewed only from the external surface of the body; (c) demonstrates the same child approximately 2 days after autopsy. Note that the characteristics of the facial bruises have changed slightly due to the extended postmortem interval.

This child sustained multiple blunt force injuries to his head, back, and extremities (a–b). The child's hair was traumatically pulled from his scalp (c). Note the multiple angulated red to brown abraded contusions on the child's back (d), consistent with a belt buckle strike. Also note, following incision of the back and extremities during autopsy (e), the large degree of red bruising within the subcutaneous soft tissues of the back. Note the pale yellow subcutaneous tissue within the posterior compartments of the lower extremities, which are atraumatic and without hemorrhage.

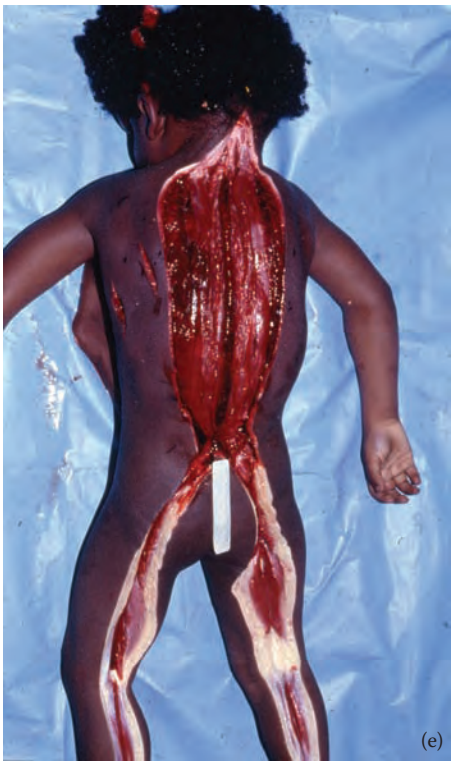

Note the multiple therapeutic interventions documented at autopsy. This documentation is important to help prevent misinterpretation of therapeutic intervention as injuries.

This case demonstrates a sutured right-sided scalp incision as part of a craniotomy for evacuation of a right-sided subdural hematoma.

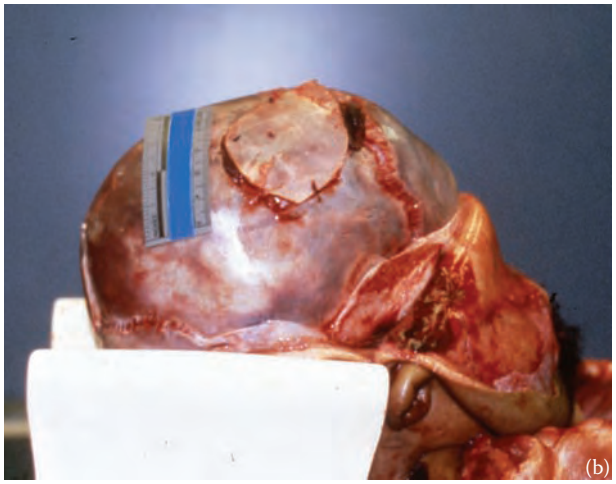

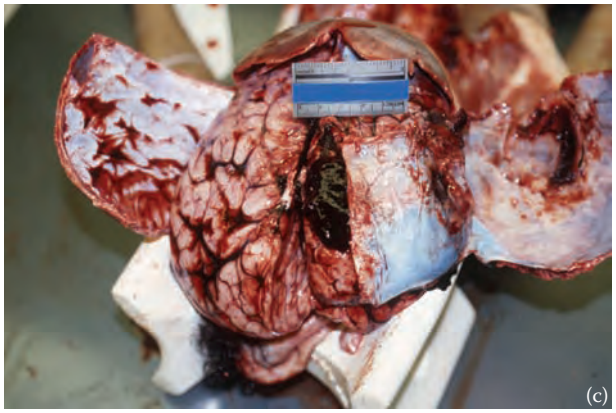

Superior view of the calvarium (a) illustrating marked widening of the coronal suture (diastasis) due to brain swelling, along with a right-sided craniotomy window; (b) right side of face and skull, including craniotomy window. Marked diastasis of the coronal suture is also seen; (c) residual right-sided subdural hemorrhage and marked brain swelling, characterized by complete flattening of gyri and effacement of sulci on the cerebral convexity of the left cerebral hemisphere. The inner surface of the left vertex dural leaf is blood-stained.

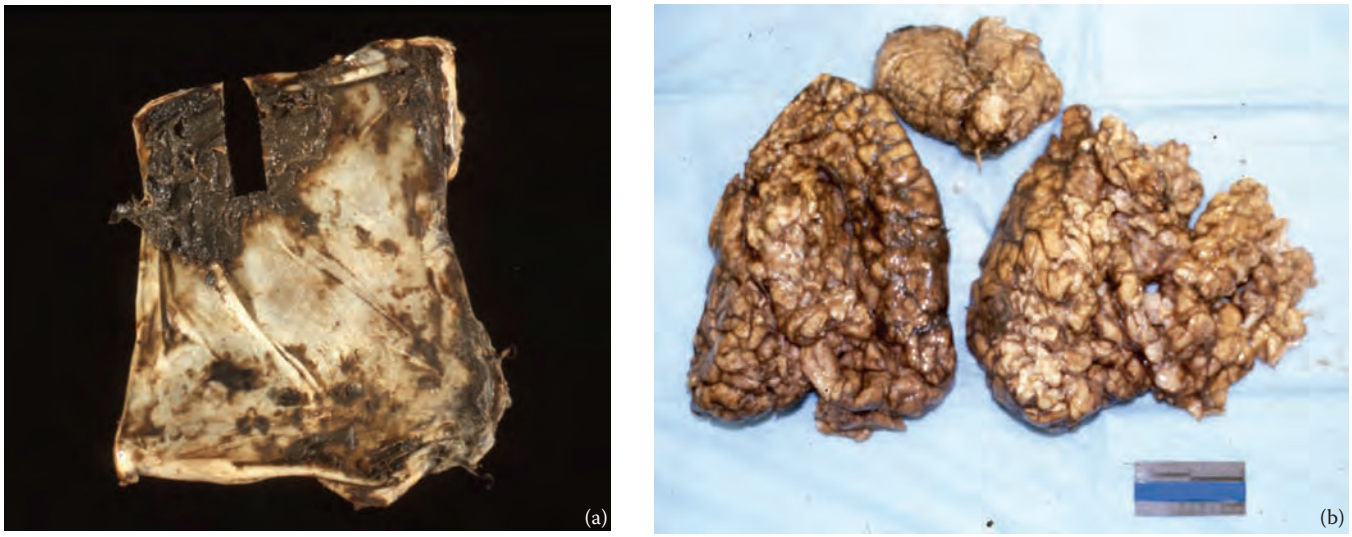

Formaldehyde-fixed portion of dura mater from an approximately 4-month-old infant with a history of tetralogy of Fallot who was found unresponsive by his mother's boyfriend. The boyfriend claimed that he had fallen approximately 3 feet from a couch onto a carpeted floor. Such explanations are often incompatible with the spectrum of injuries observed at autopsy. In this case, the adherent, thin-layered, clotted subdural blood indicates a duration of approximately 3–4 days following the injury, which correlates with the survival period of this infant after being found unresponsive. (b) Illustrates the fixed, markedly friable, fragmented brain attributable to marked hypoxic-ischemic injury, and is therefore a nonspecific finding, as it can be seen within the setting of hypoxic-ischemic brain injury of any cause.

Cross-section of the optic nerve in an infant found to have unilateral subdural hematoma and brain swelling, who underwent evacuation of the subdural hematoma. The infant never regained consciousness and was declared brain dead approximately 3 days following the surgery. Optic nerve sheath hemorrhage is correlated with retinal hemorrhages, although it is not specific for inflicted injury , and may be seen in conditions of sepsis, hemorrhagic diathesis, and other rare natural conditions.

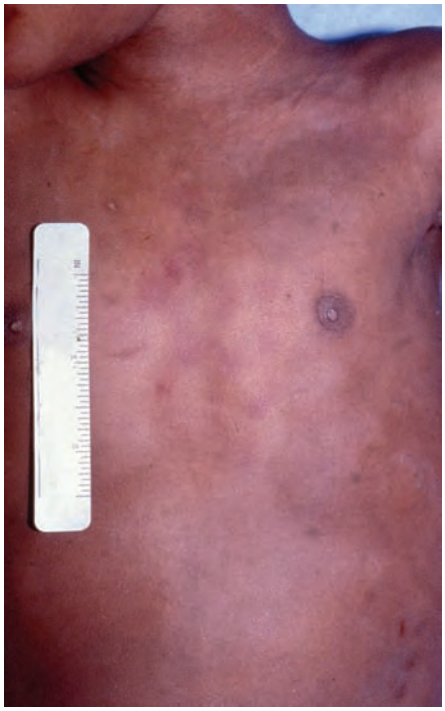

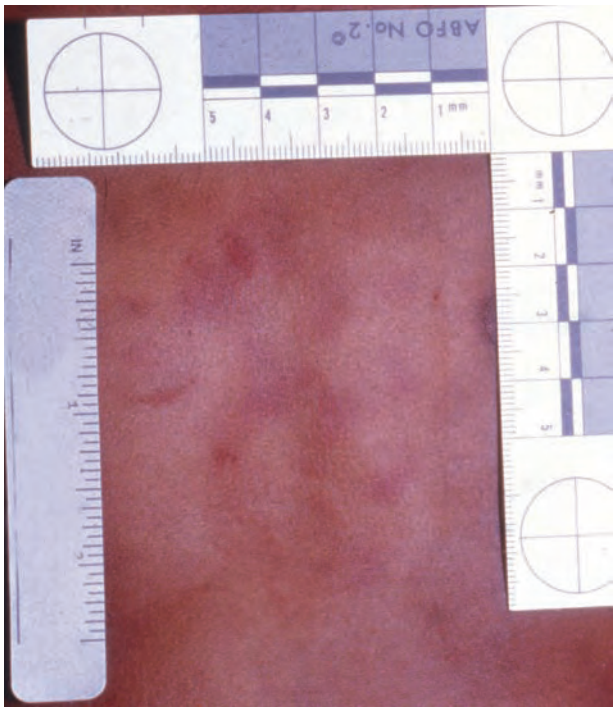

Note the vague pattern contusion of a fist on the anterior chest in this child-abuse fatality. This child died of multiple blunt impact injuries and was discovered dead at the scene in full rigor. No resuscitation was performed.

Antemortem contusions associated with child abuse combined with superimposed perimortem contusions associated with resuscitation. The resuscitative injuries overlie the sternum.

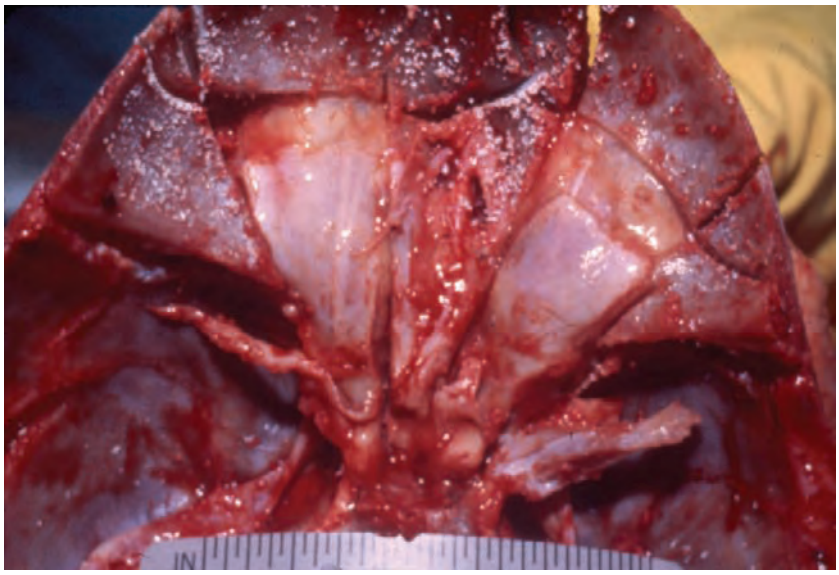

This demonstrates saw cuts through the floor of the anterior cranial fossa (orbital plate of the frontal bone) prior to removal of the eye with its attached optic nerve.

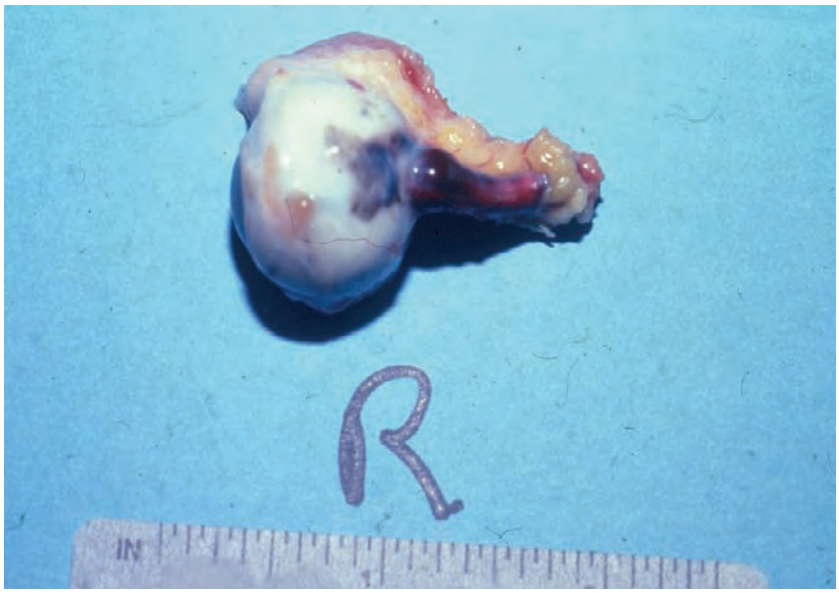

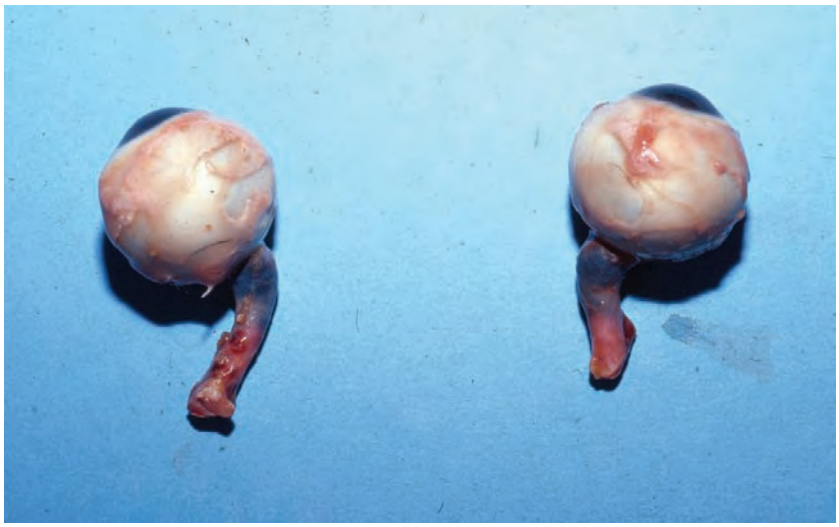

Note the optic nerve sheath hemorrhages. Both of these images demonstrate optic nerve sheath hemorrhages in small children who died from shaken-impact syndrome. Histopathology revealed retinal hemorrhages. There was no evidence of natural disease including coagulopathic syndromes or connective tissue disorders.

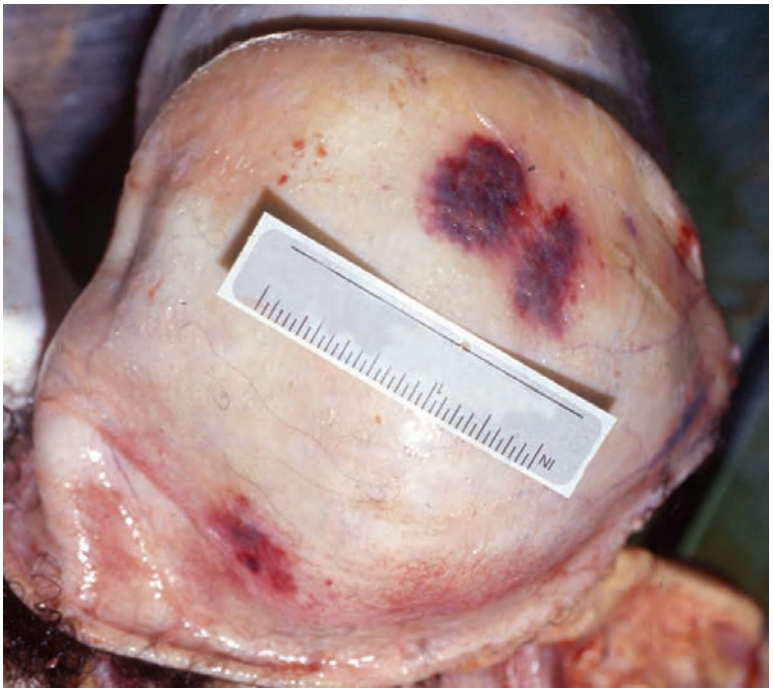

The first image of this child-abuse case demonstrates vague contusions on the child's forehead. Reflection of the scalp shows much more prominent soft tissue (subgaleal) hemorrhage.

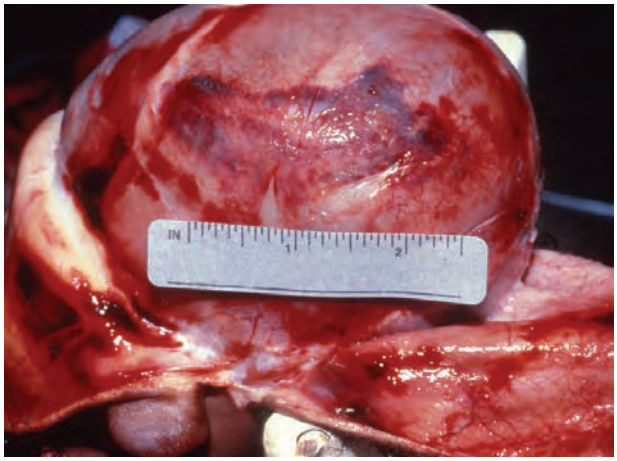

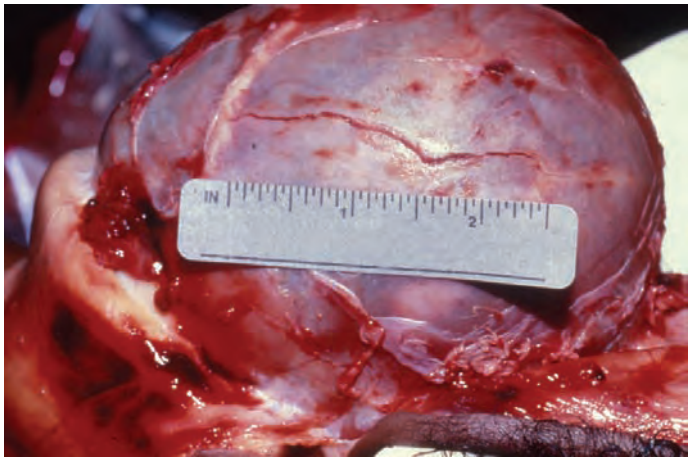

Child abuse case with reflected scalp demonstrating contusions and underlying linear skull fracture.

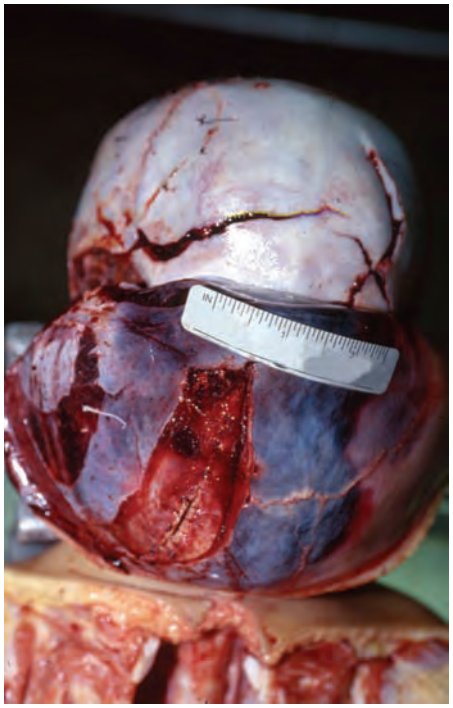

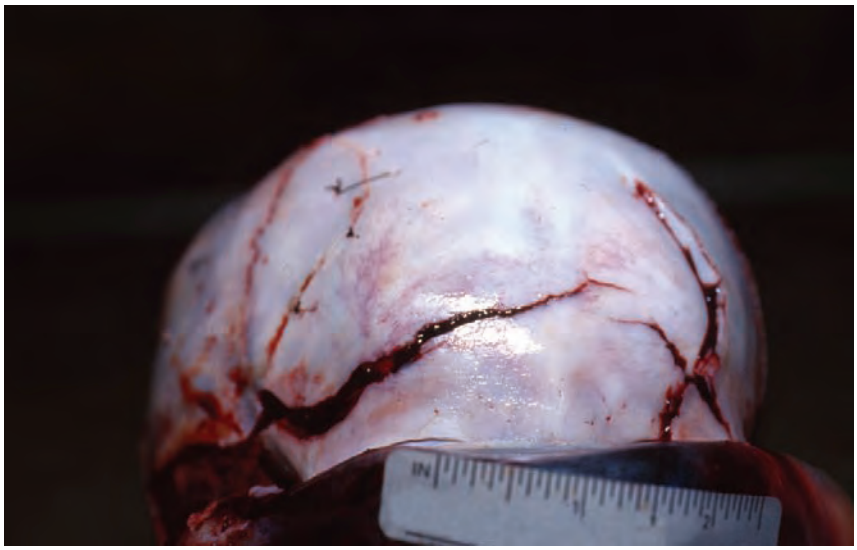

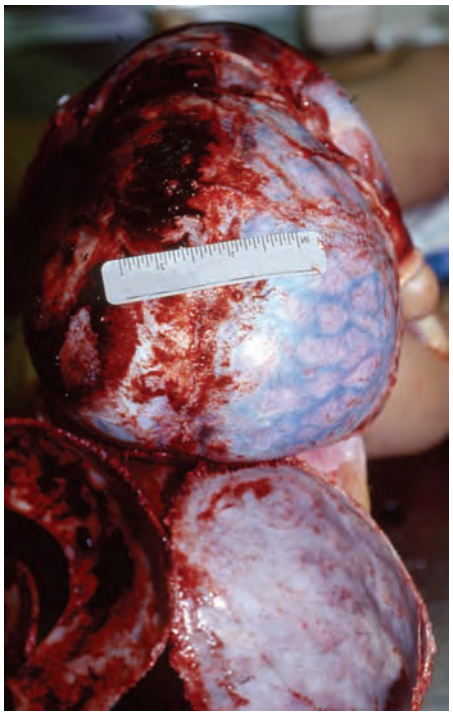

This child was visiting his grandfather with his parents from another country. The grandfather's apartment was not made "child-safe" prior to the visit. The toddler was standing and rocking a large old wooden television back and forth on a stand that was missing several screws. The television toppled over and crushed the child's head. Note the fracture pattern to the skull indicating a crush-type injury. Also note the photograph demonstrating the top of the skull reflected with underlying dura mater still overlying the brain with epidural hemorrhage. The fractures involved the temporal, frontal, and parietal bones and there was severe brain injury with brain swelling and herniation.

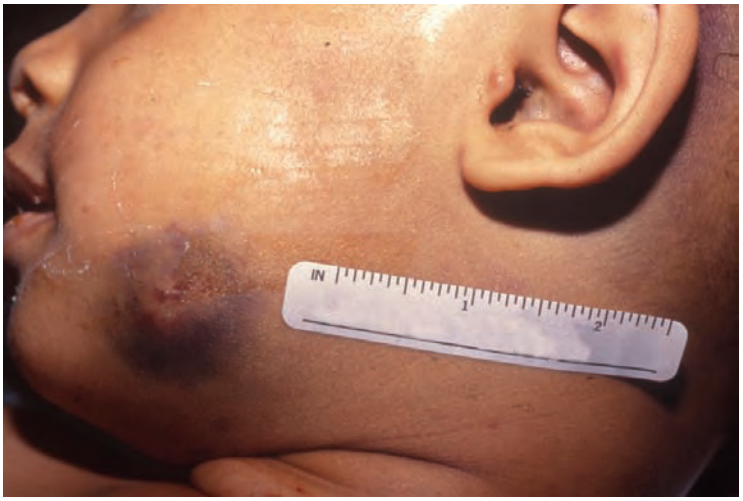

Note the large contusion to the side of the child's face from being punched.

This child was the victim of an assault and suffered a skull fracture with brain injury that caused his death. Note the area of hemorrhage to the mastoid portion of the scalp demonstrating "battle sign."

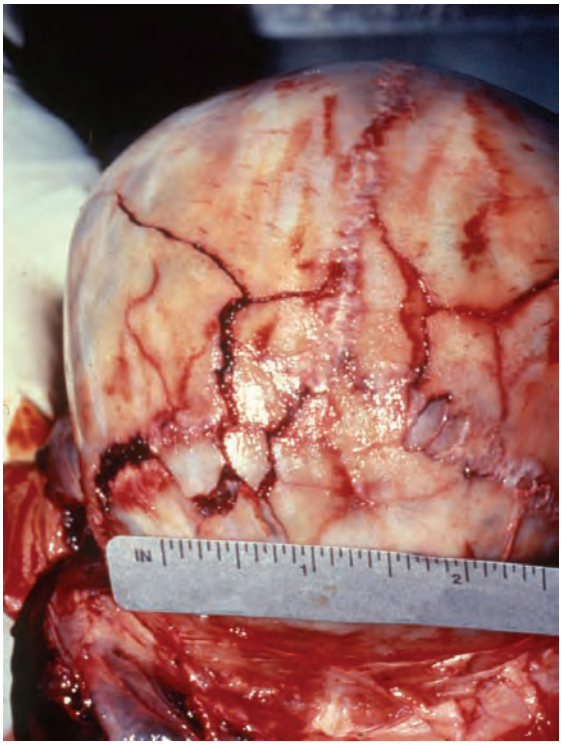

Child abuse case with comminuted skull fracture. This was most likely produced by flinging the child in a downward trajectory, and striking the child's head against a hard countertop. A fracture of this nature would take tremendous force to produce.

This demonstrates the top portion of the skull (calvarium) removed during autopsy with the dura mater still attached and clotted, partially adherent, subdural hemorrhage. Clotted adherent red-black subdural hemorrhage indicates an injury age of approximately 2–4 days.

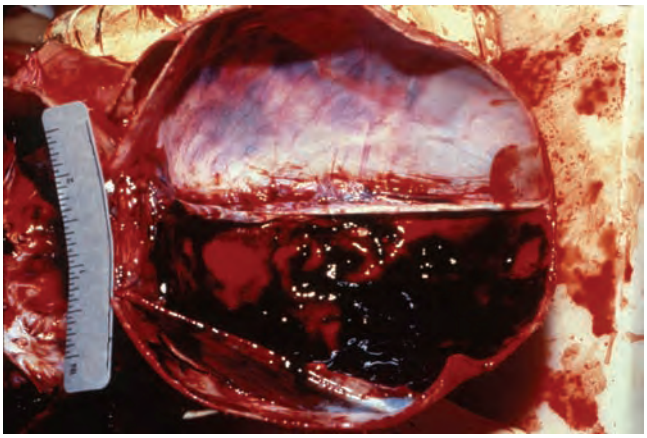

This is the top portion of the skull with adherent clotted subdural hemorrhage involving the left vertex dura (overlying the left cerebral hemisphere) with portions of dura mater separated at the posterior temporal-parietal-occipital regions, adjacent to sites of skull fracture.

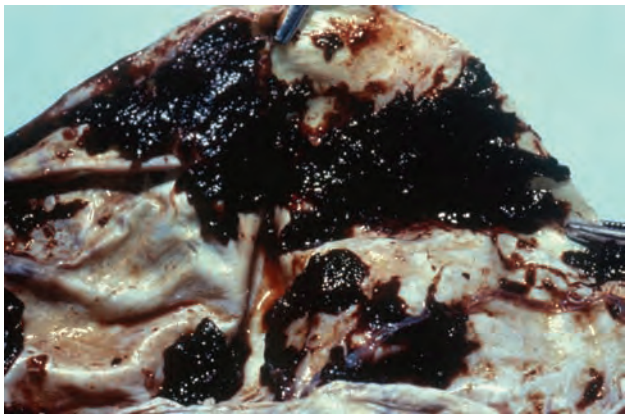

Portion of formaldehyde-fixed dura mater saved for further neuropathologic evaluation. Note the adherent clotted, black-brown subdural hemorrhage.

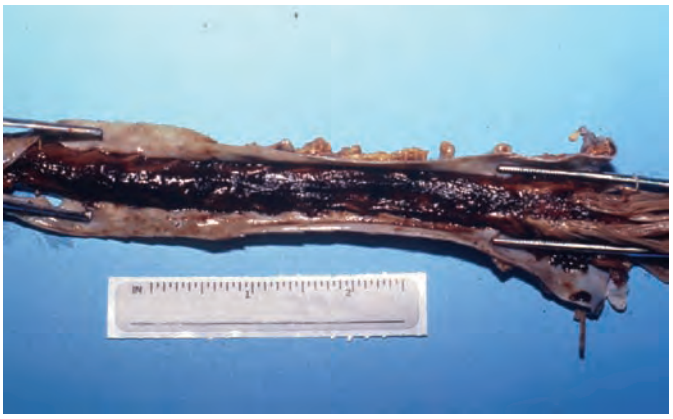

Formaldehyde-fixed portion of spinal cord removed with the dura mater demonstrating subdural hemorrhage. It is important to remove the cervical spinal cord as carefully as possible during the autopsy process in order to avoid potentially troubling postmortem artifact.

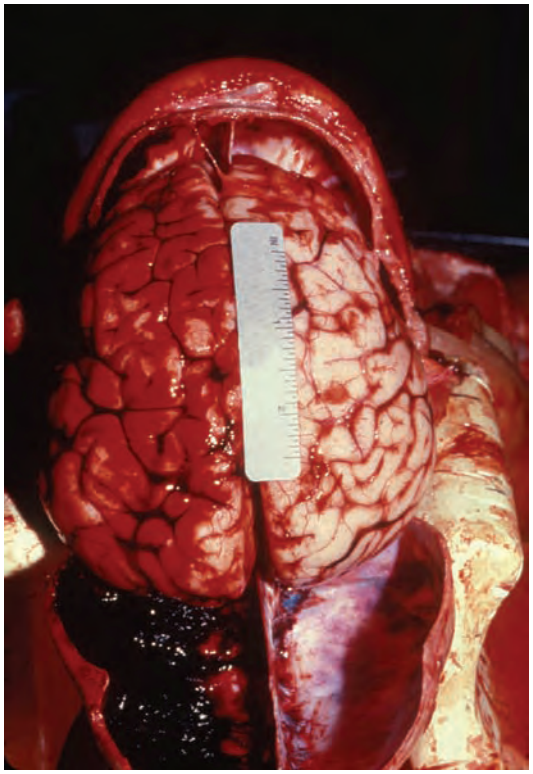

This shows the top portion of the skull (calvarium) being unroofed during the autopsy process for removal of the brain. Note that the dura mater is still adherent to the inner aspect of the skull (endocalvarial surface). There is a large left-sided adherent clotted subdural hemorrhage. There is diffuse cerebral swelling with flattening of the gyral configuration and obliteration of the sulci. There is patchy subarachnoid hemorrhage, which is greatest over the left cerebral hemisphere.

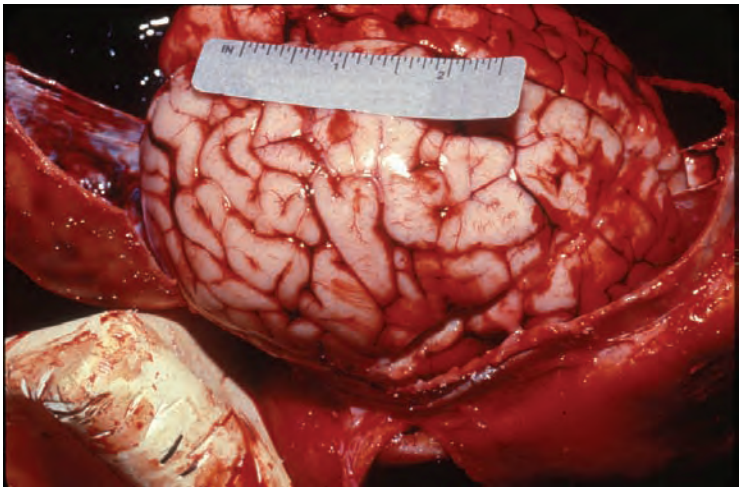

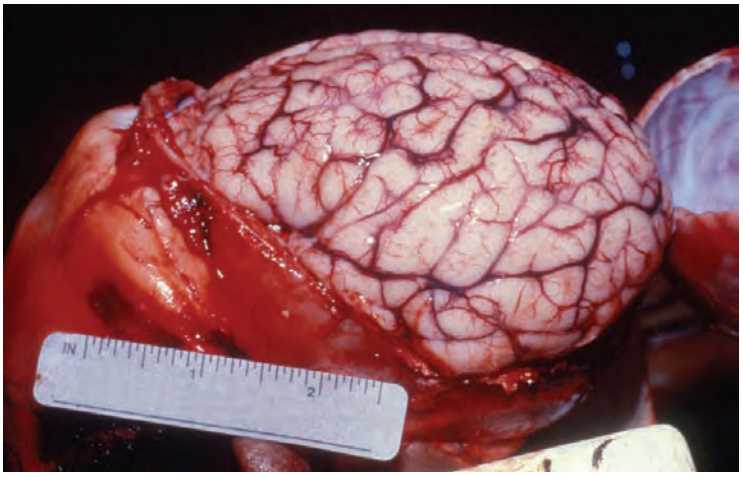

Close-up views of brain swelling with flattening of gyri and obliteration of sulci in a child who survived several days following blunt force trauma due to homicidal violence.

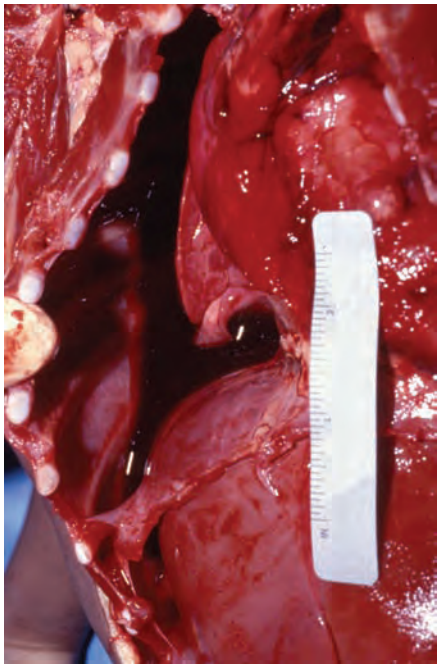

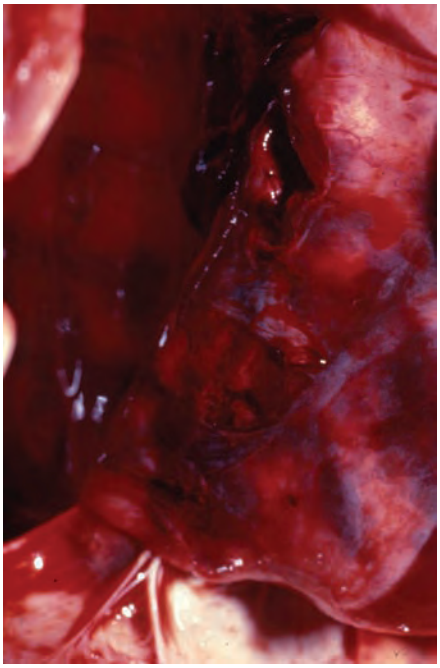

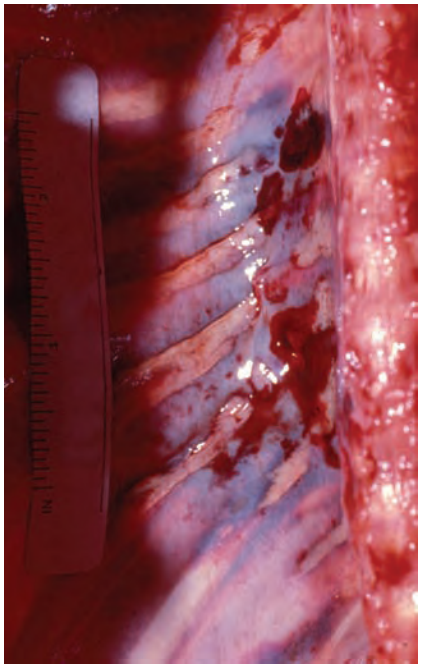

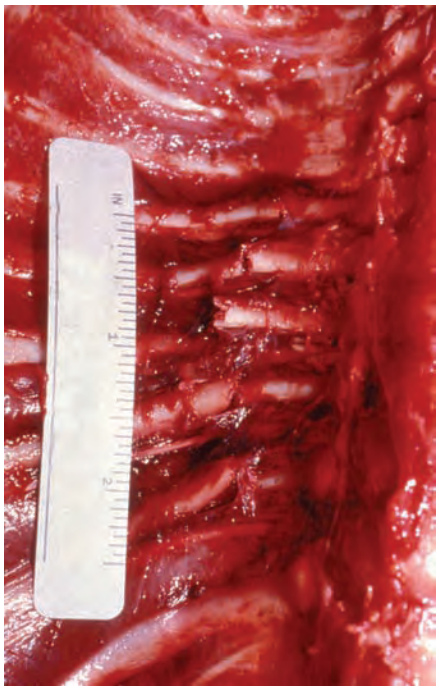

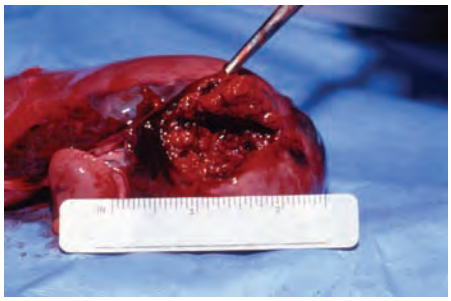

This child sustained multiple posterior rib fractures after being stomped by an adult while lying face-down on the floor. The fractured segments of the ribs penetrated the posterior aspect of the right lung, producing extensive pulmonary and intrathoracic hemorrhage. There were also large lacerations to the posterior aspect of the liver and heart.

X-ray showing multiple recent rib fractures.

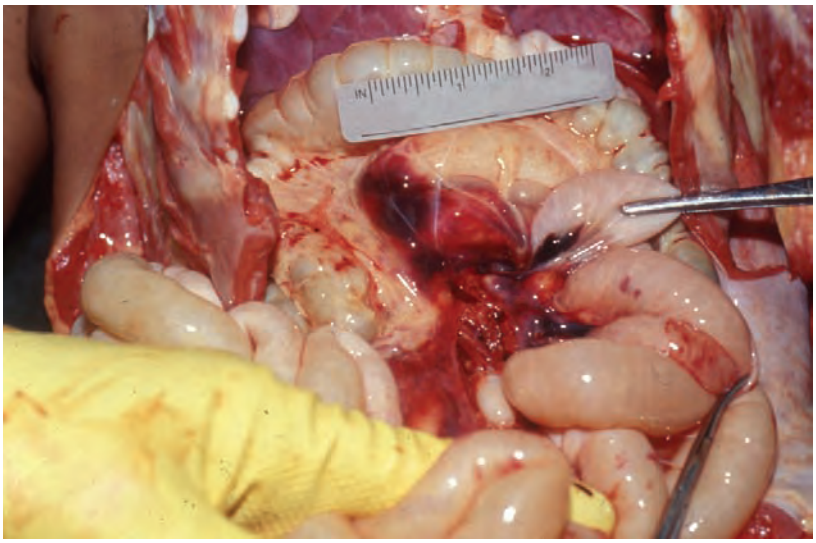

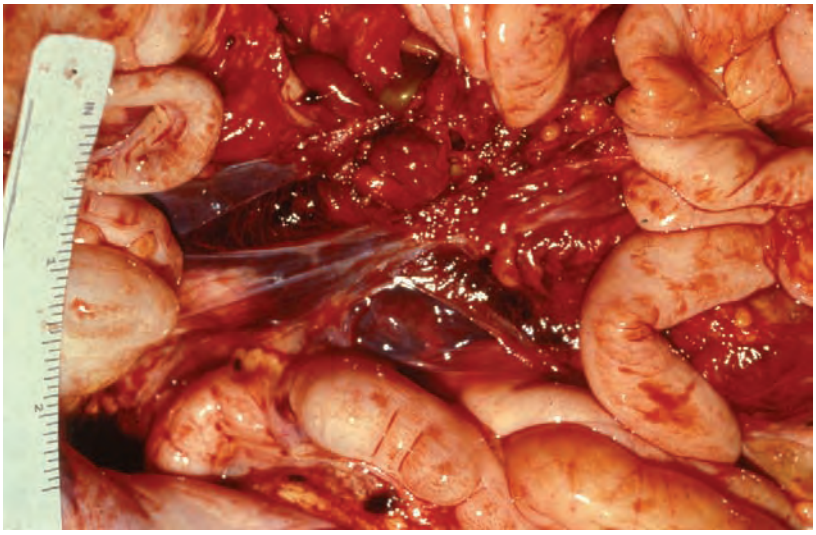

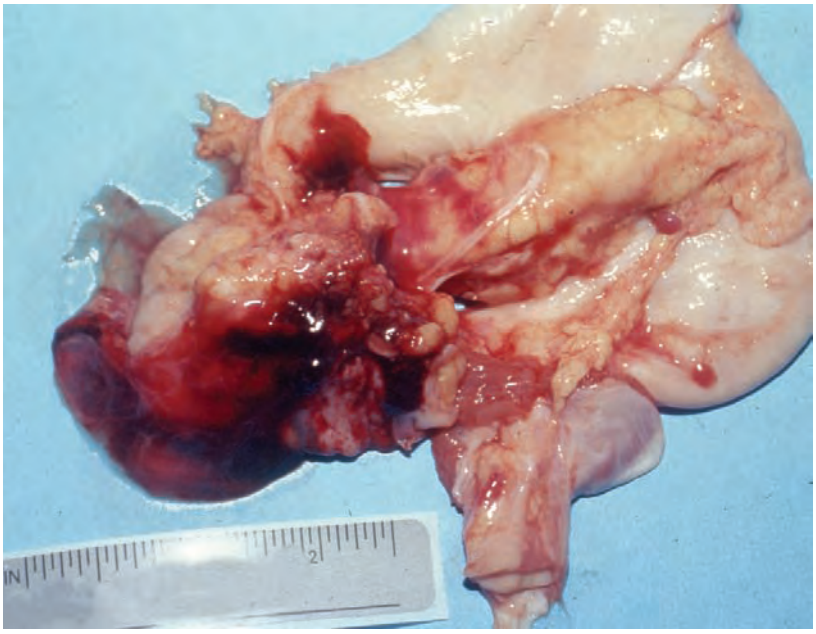

These images demonstrate multiple lacerations involving the mesentery, mesocolon, and peripancreatic soft tissues following blunt abdominal impacts associated with homicidal violence.

This 2-year-old child was found by a caretaker unresponsive in the bathtub. An autopsy revealed confluent subgaleal contusions, essentially spanning from ear to ear, along with brain swelling, although the skull was intact. Death was attributed to blunt impacts to the head with brain swelling. A component of drowning cannot be ruled out in this case. The defendant, a casual acquaintance of the mother, was convicted of second-degree murder and sentenced. Note the bilateral intraosseous catheters, used while attempting to resuscitate this child even though he was slightly over 2 years old. This image depicts the face and anterior trunk of a child with patchy contusions involving the forehead and the left paramedian aspect of the anterior chest wall. Multiple therapeutic interventions are also present. As stated previously, differentiating between the effects of therapy and antecedent injury can be problematic, and the two processes may coexist.

This depicts lower trunk and lower extremities of child, with multiple therapeutic interventions including transverse cutdown sites within both groins, bilateral intraosseous catheters, and a right groin catheter.

This image shows the perineal area of this child (a). No injuries are detected. (b) depicts the child's anus. No injuries are demonstrated. In cases such as these, it is good practice to document photographically the presence or absence of injuries, as the multiplicity of injuries may be one of the determining factors in the ultimate sentencing. Sometimes external examination may be difficult and mucosal folds may be misinterpreted as lacerations. If there is any question as to the nature of a possible lesion or injury, we recommend removing the region en bloc, opening the anorectal segment and photographing it for documentation.

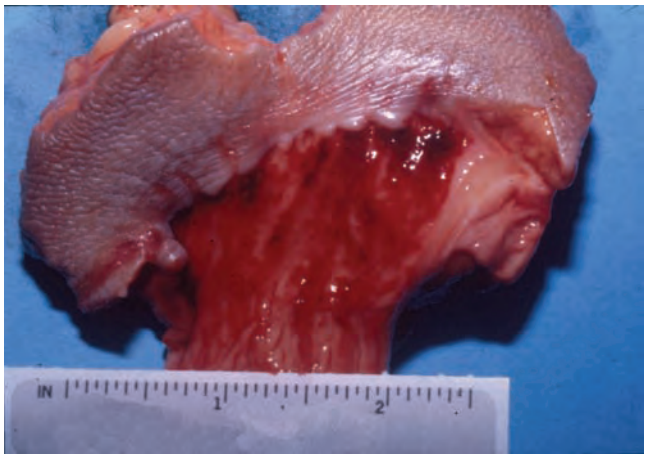

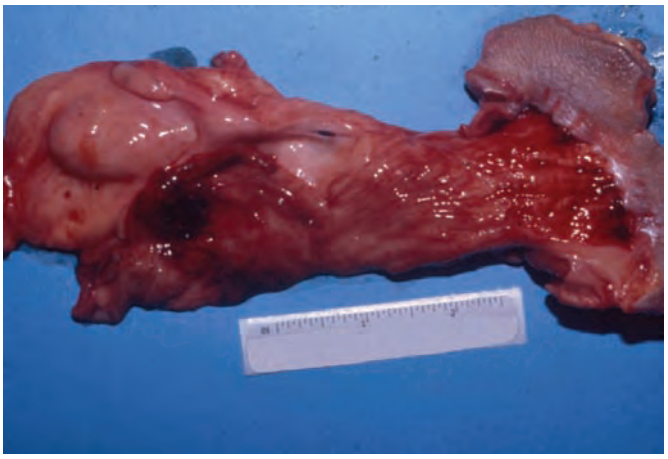

These images demonstrate sexual abuse in a young child. Note the abrasions and lacerations around the anal region. En bloc dissection of this region with the anal-rectal segment of the intestine removed. Note the internal lacerations near the anal border and the intestinal hemorrhage

Superficial, healing, criss-crossed, linear abrasions due to rubbing of Velcro jacket patches on each side of this child's neck. This indicates rough handling. This child was later killed by multiple blunt impact injuries during an assault. Note the suture and fresh needle marks at the right chest from a removed subclavian line.

(a)

(b)

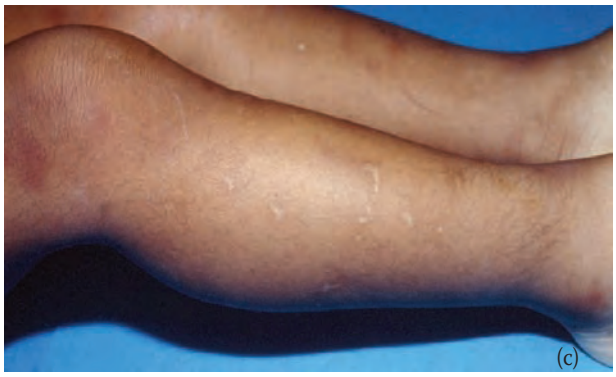

(c)

(d)

These are multiple scars, many of which were associated with chronic child abuse. When evaluating old injuries, it is important to take into context the child's age and developmental status. A child who is able to run and jump about may fall and sustain blunt-impact injuries. Sometimes it may be difficult, if not impossible, to formulate the nature of such old injuries without history or witnesses. (a) and (b) images are associated with abuse; (c) and (d) represent healed scars from accidental injuries. Note the healing injury at the nose associated with a nasogastric tube and therapy.

This demonstrates a coronal section of a humerus. Microscopic sections of long bones may be sampled in cases involving a suspicious death of an infant or young child in order to identify transmetaphyseal fractures in situations in which shearing forces have been applied (i.e., violent shaking). Unlike the case in more obvious fractures with extensive visible hemorrhage, these fractures (alternatively referred to as "bucket handle" or "corner" fractures on radiologic plain films, depending upon the view), may sometimes be more readily documented with microscopic examination.

Old healed rib fractures from a case of battered child syndrome.

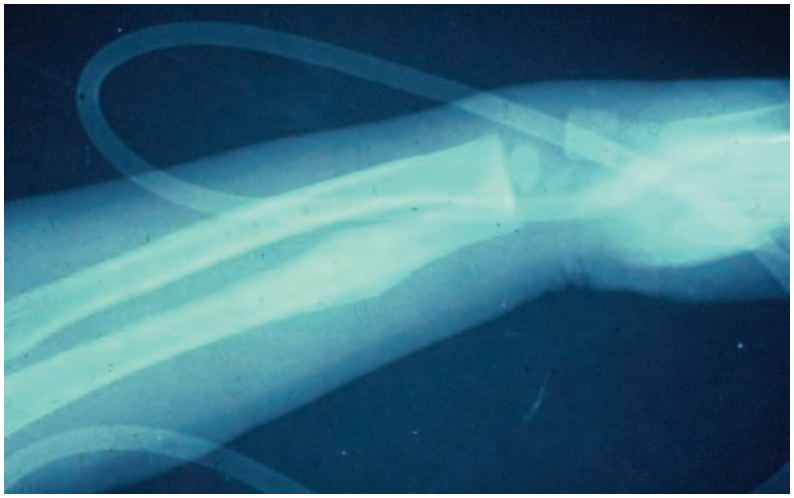

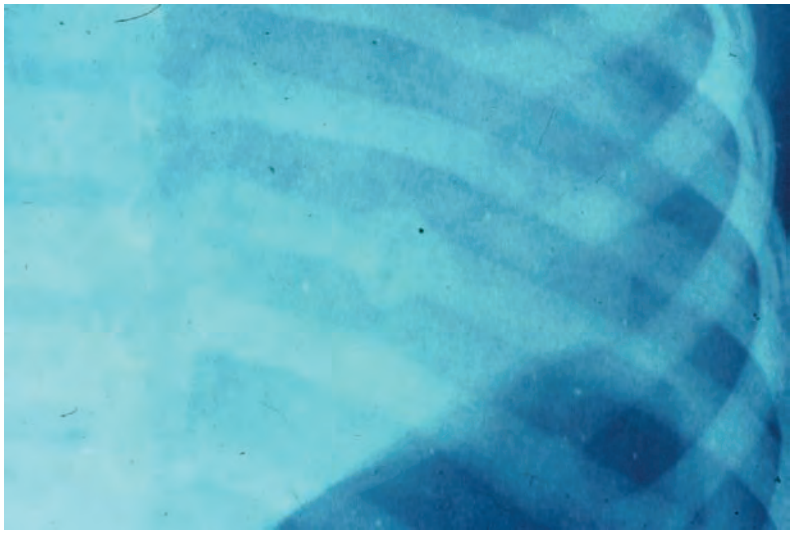

These x-rays demonstrate healing fractures.

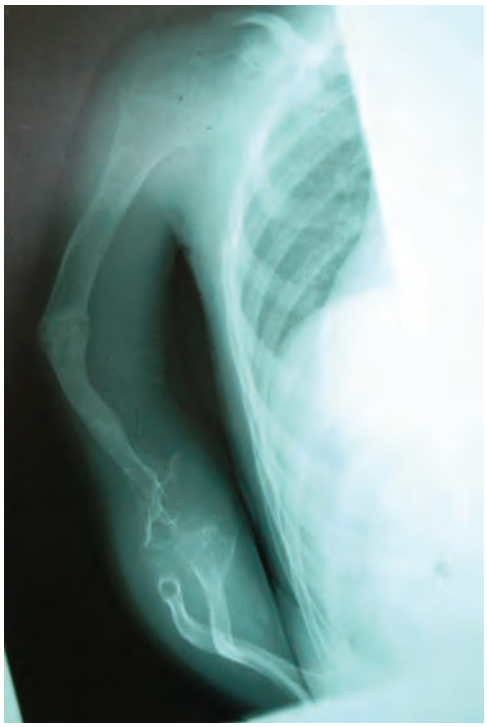

Osteogenesis imperfecta with multiple fractures. This may or may not indicate child abuse. Fractures may occur with very little force.

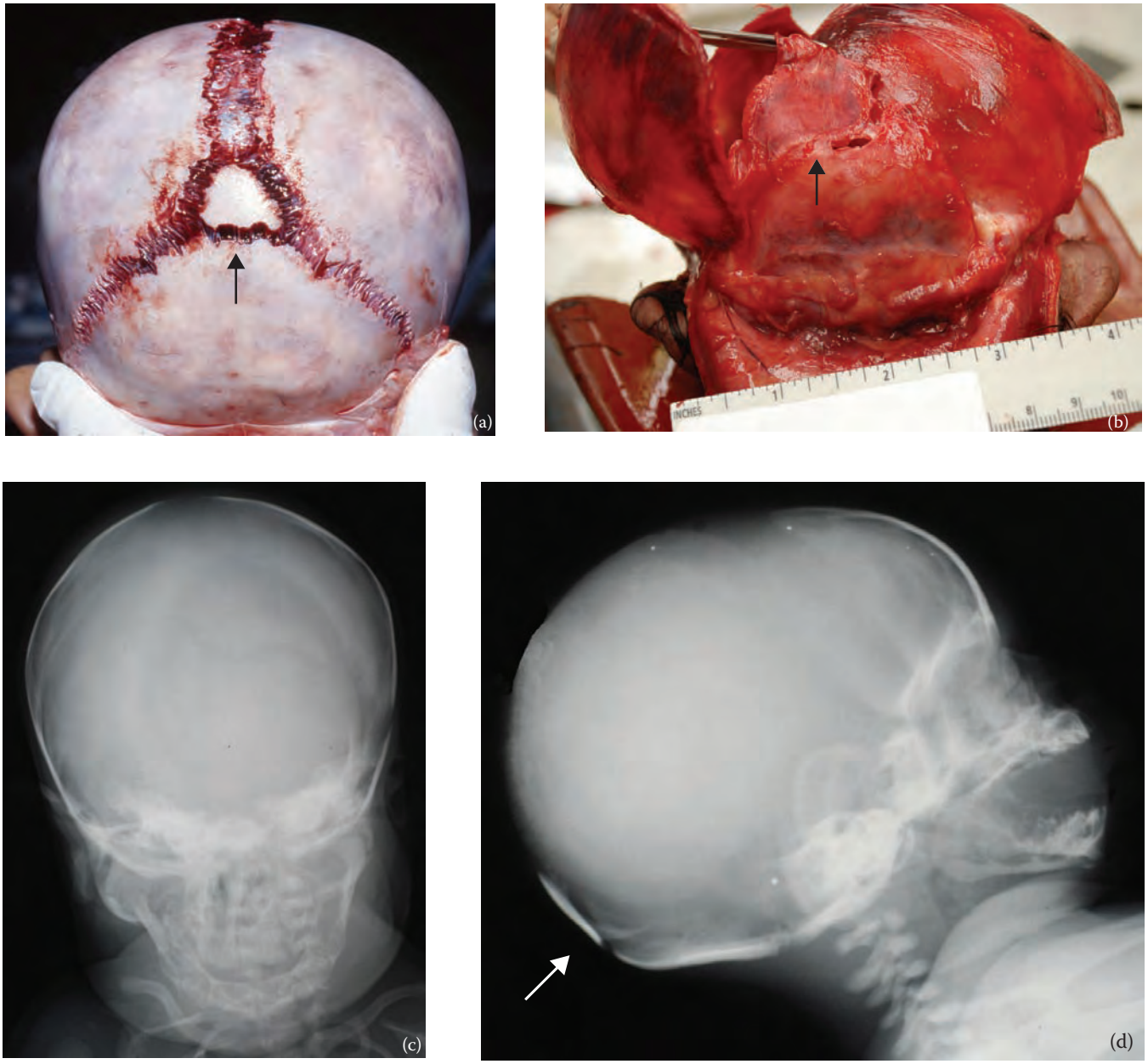

Inca bone skull anomaly. In both of these cases it was mistaken for a skull fracture. The x-rays at the bottom correspond to (b).

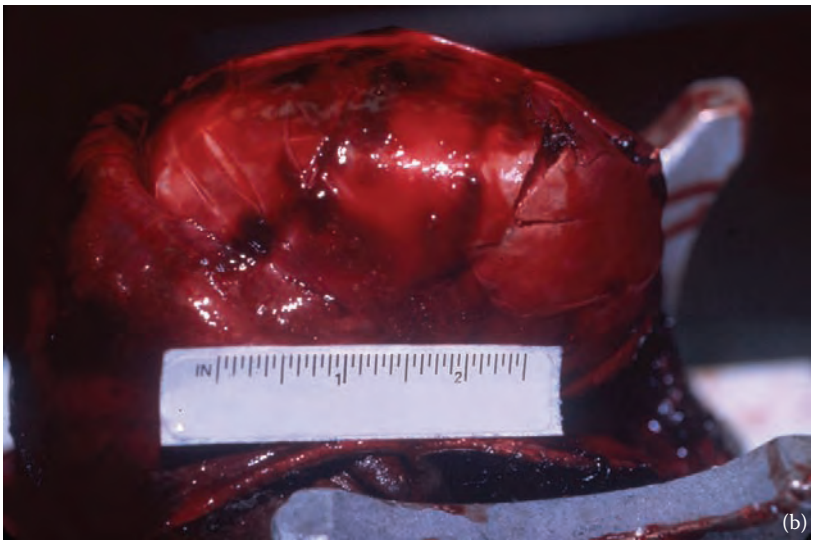

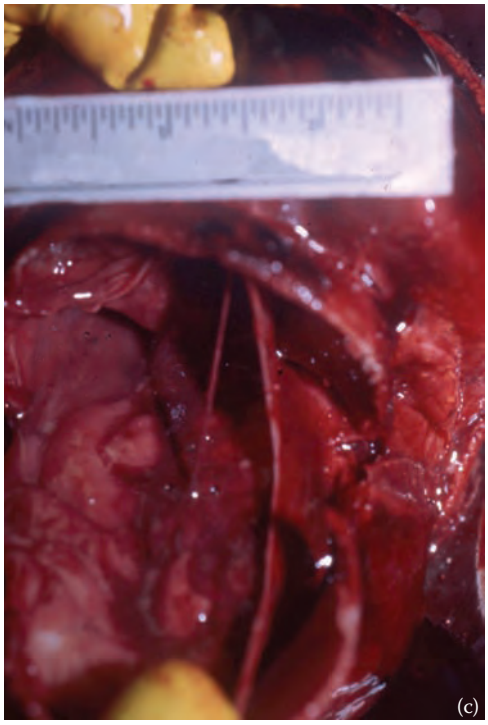

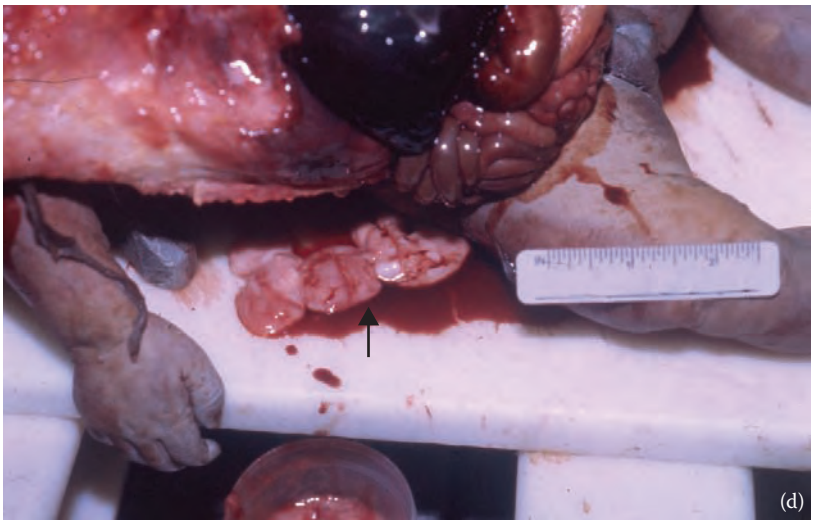

This child was retrieved from the back of a garbage truck partially tied within a torn plastic bag following compression by a trash compactor. Note the confluent hemorrhage to the scalp and the largely empty cranial vault. This child was mildly to moderately decomposed with bloating and separation of tissue planes. The brain material was squeezed from the cranium through the neck following compression of the head by the compactor. When the initial Y-shaped autopsy incision was made to the skin of the trunk, brain material leaked from the chest region (d). The child's body was found mixed with decomposing food material. Note the green discoloration from putrefaction.

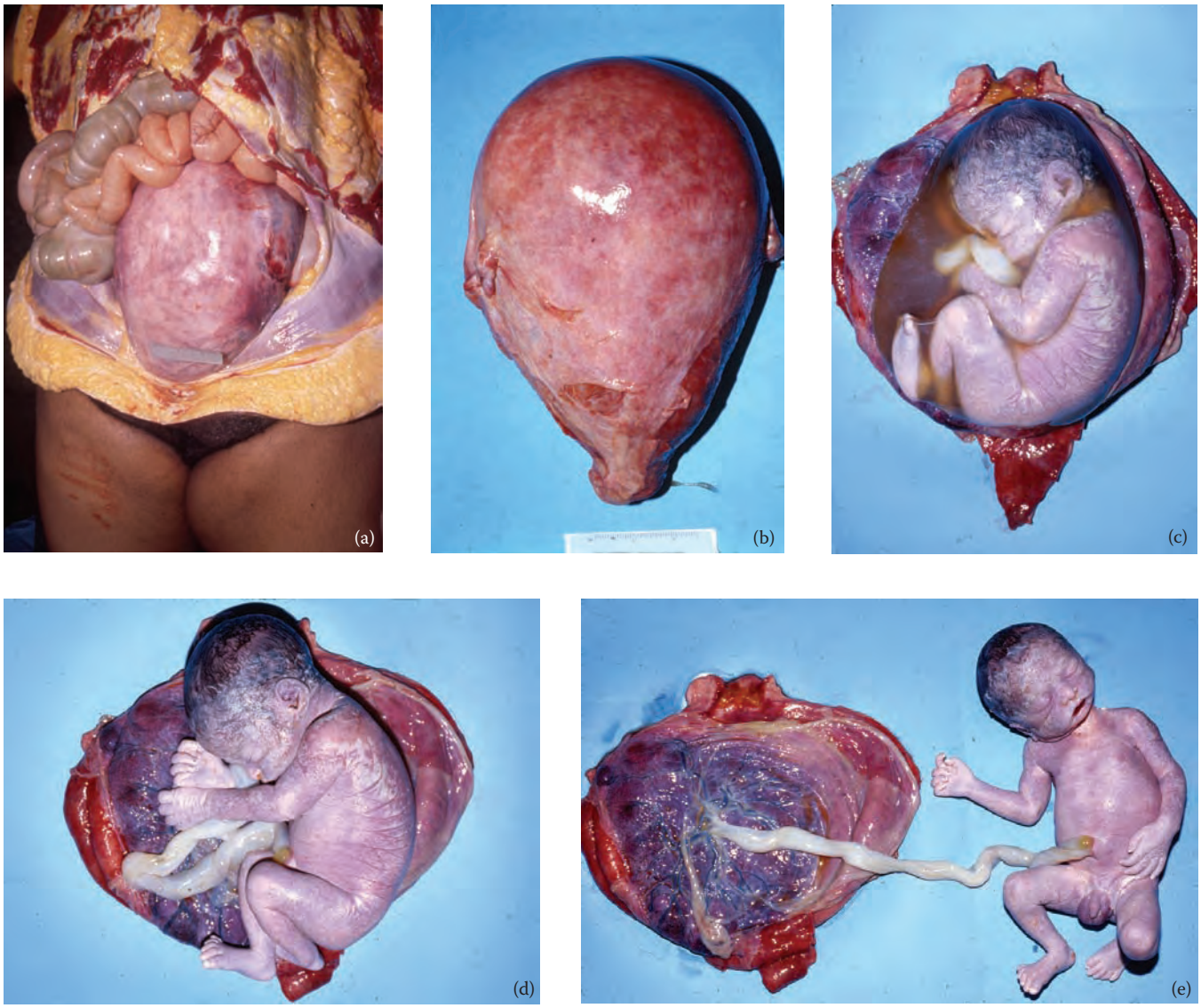

Late-second-trimester pregnancy demonstrating the uterus (a), placenta (b), and fetus (c).

Second trimester stillborn with omphalocele, illegally aborted following administration of drugs given by a nonmedical practitioner. Note the x-ray and the sunken lungs at the bottom of the water container indicating no breaths were taken.

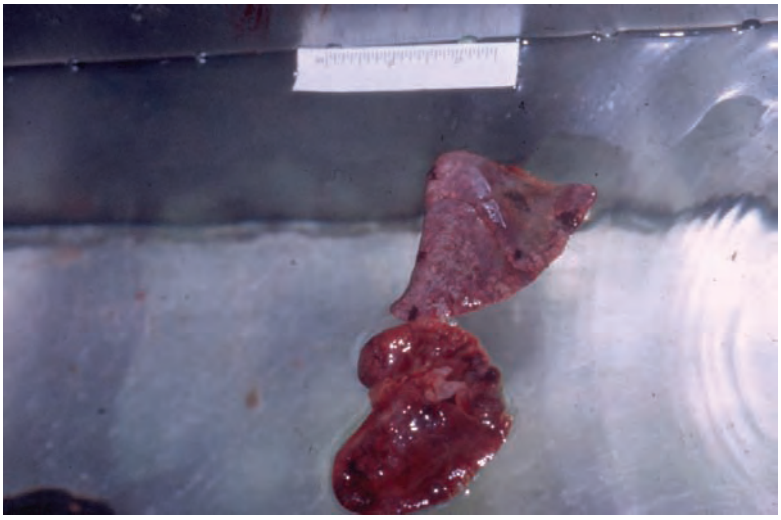

Lungs from a live birth that were expanded with air. Note how they float on top of this water bath. Admittedly, the "flotation test" to confirm independent respiration is far from foolproof and is subject to a number of artifacts and exceptions, thus limiting its utility in resolving the question of live birth versus stillbirth.

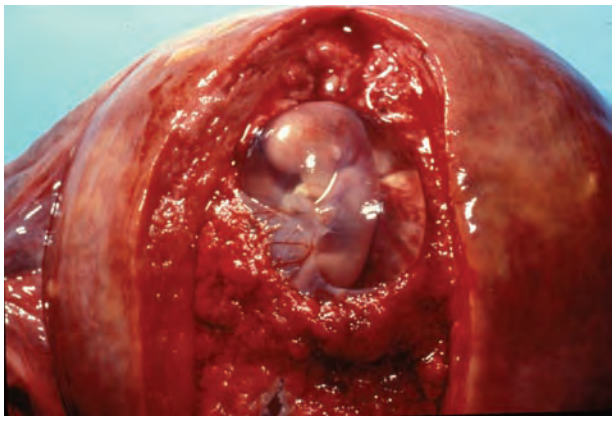

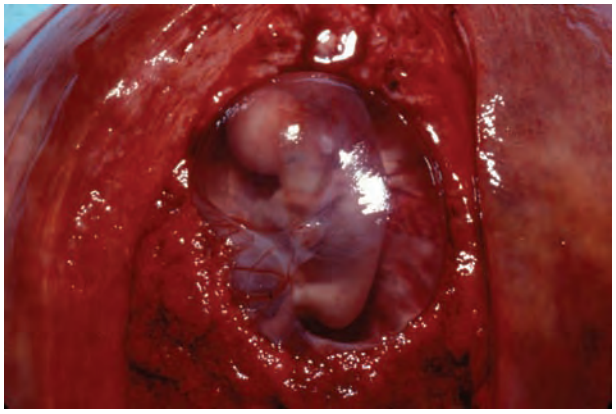

First-trimester pregnancy.

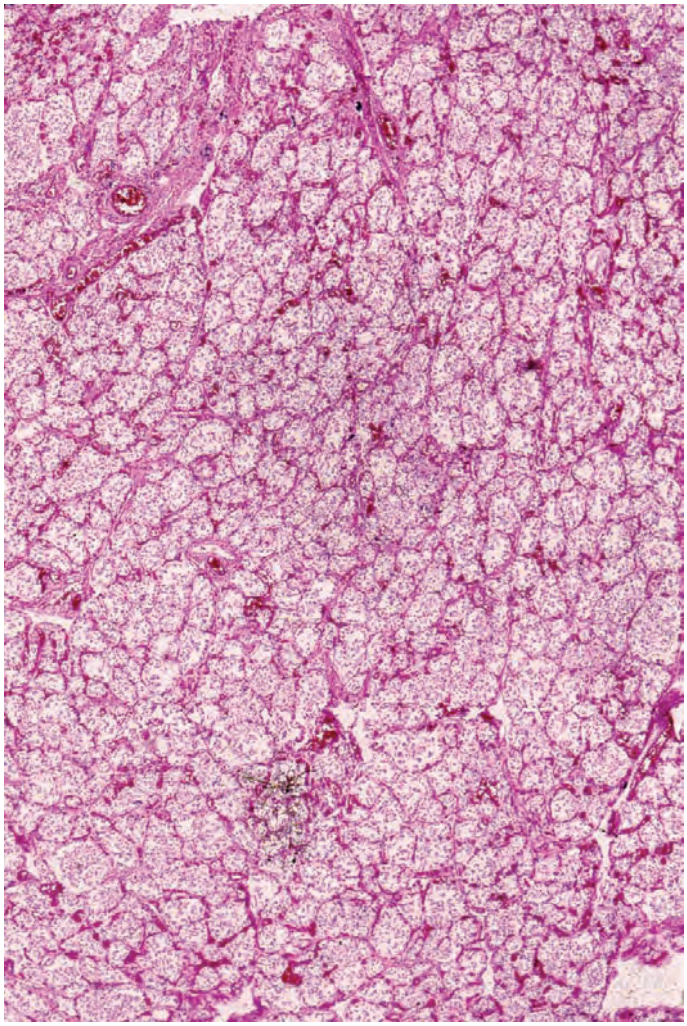

Brown fat.

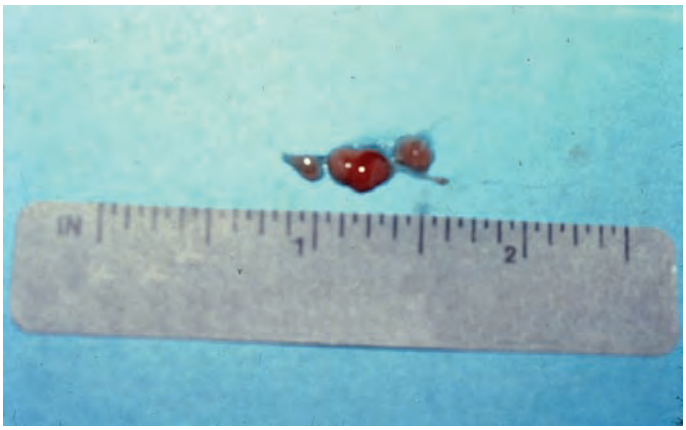

Early first-trimester pregnancy.

Blunt-Force Injuries

6

CHARLES A. CATANESE

Introduction

These are injuries produced when the body is struck with, or strikes, a blunt object. A blunt surface produces injuries by scraping, tearing, shearing, or crushing. This is in contrast to a sharp force injury, which cuts and separates the tissue as it penetrates. What injuries result from trauma are a balance of the amount of force, the area over which it is applied, and the duration of force. In general, the greater the force, smaller the area, or shorter the duration over which the force is applied, the greater the injury will be.

Blunt objects have a relatively large surface area in contrast to sharp objects, where the cutting edge has a relatively small surface area. It takes much less force to penetrate the skin with an ice pick than with the end of a baseball bat. Examples of blunt objects are fists, shoes, pipes, bricks, bats, hammers, roadways, sidewalks, cars, trains, airplanes, walls, etc. Types of blunt-force injuries include *abrasions, contusions* and *lacerations*. These may occur separately but are often present at the same time. As an example, an injury maybe described as an abraded contusion with central laceration. *Fractures* are breaks in the bone as a result of blunt force. *Avulsions* are splits in the soft tissue or soft tissue planes with or without a laceration as a result of shearing forces.

Abrasions

Abrasions are simply scrapes. They are produced as the body contacts a surface and rubs across it with sufficient force. The distance of travel can be very short (falling and scraping your knee) or very long (being dragged under a car). It is often possible to determine the direction of impact. Layers of skin are scraped away and bead up at the margin where the contact to the wound last occurred. Fragments of skin beading up at the inferior wound margin indicate a downward impact. Antemortem abrasions or those that occur during life are typically red to brown and will eventually form a scab with dried blood. Postmortem abrasions are yellow in non-lividity-dependent areas. In general, one needs a beating heart with blood pressure to produce hemorrhage and a red to brown discoloration. A postmortem scrape in a lividity-dependent region will appear red and may be difficult to differentiate from an antemortem injury.

This is important in certain cases. Lawyers will argue different ways to create reasonable doubt. For example, some may argue that an injury occurred during resuscitative efforts in the hospital, or the body was dropped from the stretcher in the morgue. Some other arguments may include: he was dead already when the second car hit him, the other perpetrator shot the body after he was dead, the child had no injuries before entering the hospital, etc.

Abrasions may also change character with increasing time. For example, initial examination of a body retrieved from water may reveal no or much less obvious injuries due to the moisture from the water at the skin's surface. As the body is stored in the morgue overnight and allowed to dry, the abrasions will darken and may become much more apparent.

Contusions

Contusions are simply bruises. These are produced following an impact where the soft tissues and blood vessels, underneath or within the skin, are torn and produce hemorrhage. Grabbing an arm tightly can produce bruising with minimal impact and greater crushing force. One needs a beating heart and blood pressure to produce a contusion. A postmortem impact may accumulate blood due to lividity in a gravity-dependent area. This type of postmortem artifact may be difficult to differentiate from an antemortem injury.

One must exercise great care when dating contusions at the time of autopsy with gross findings alone. One must account for skin color, whether the contusion is deep or superficial, the presence of hematoma, etc. Bruises go through various color changes with advancing time as the body reacts to repair the injury. This depends on the size of the injury, the physiologic state of the individual, including his or her immune and coagulation systems, the vascular efficiency adjacent to the damaged tissue, etc. Color changes range from light blue-red to dark purple then green to yellow-brown as time progresses. Inaccurate interpretations of dating contusions grossly can create significant problems in court. If one sees a variability with color ranges demonstrating

some bruises as red-purple and others as yellow-brown, it is reasonable to say that some injuries are older than others. It is much more accurate to use histopathology to date the injuries using established time ranges for advancing inflammation and healing. Random sections of injuries should be taken including the wound margins. The first several hours reveal hemorrhage with no inflammation. This is followed by iron deposition as a result of hemoglobin breakdown. Estimation of injury dating should be given in ranges.

Bruises can also change appearance as the postmortem time interval increases. A contusion will become more obvious as blood settles away from the impact site. The surrounding tissue will become paler and the contusion will be more pronounced, as in a supine body with anterior contusions in full livor. Bleeding associated with a contusion does not settle away from the impact site as lividity forms, because it is spread throughout the soft tissues and cannot drain within the vessels. Also, during vigorous emergency room resuscitation, the head and upper trunk may become congested. As lividity settles to the back of the head, injuries to the face become more prominent. Clinicians may sometimes miss these subtle contusions.

It is also important to realize that deep contusions may not be visible at external exam and may be visualized only after incisions are made. Clinicians should realize this and create a medical record that is as accurate as possible. If one documents that no bruises are present to regions of the body, yet they are found on autopsy, this may potentially weaken the case in court. The autopsy remains the most accurate way to demonstrate these findings.

Lacerations

Lacerations are tears of the skin, soft tissues, or internal organs or vessels as a result of an impact, overstretching, or crushing-type forces. These injuries are characterized by irregular margins, often with a marginal abrasion, with tissue bridging by fibrous strands and small blood vessels. If the laceration is large and gaping, tissue bridging may not be present due to the strands' being pulled apart. Skin lacerations tend to occur more often over hard surfaces, such as the scalp, knees, elbows, etc. The direction of the impact can be determined by the presence of soft tissue undermining. The underlying soft tissue adjacent to the laceration may form a pocket of separation extending in the direction of impact. A downward impact will produce undermining at the inferior aspect. These directions should be described with reference to standard anatomic planes.

Motor Vehicle Injuries

Motor vehicles include any motorized means of carrying or transporting someone. These include trucks, buses, cars, motorcycles, mopeds, snow mobiles, etc. It is always important to be as accurate as possible, including the *type of vehicle* on the death certificate for vital records. The type of vehicle is obviously important when evaluating injuries.

It is important to recognize various patterns that might help *differentiate drivers from passengers*. In high-speed collisions with unrestrained occupants, people may be ejected from the vehicle. If criminal or civil charges are filed, the living driver may indicate that the dead passenger was driving.

One should look for steering wheel impact marks to the chest, seatbelt-related abraded contusions, and pattern injuries associated with impacts to the windshield, dashboard, or vehicle roof. *Front and back windshields* are often made of laminated glass and fracture with elongated curves or splinters. *Side windows* are often made of tempered glass and fracture into small cubes. Seatbelt-patterned injuries or side impact dicing pattern injuries to the right or left side of the head will help formulate opinions about an individual's seating position. Dicing injuries to the left side of the head of an individual found next to a car with a broken left side window is evidence that he or she was the driver. One could also collect DNA samples from these regions and compare it with ejected occupants. Those not wearing seatbelts are much more likely to be ejected from the vehicle, especially in a rollover accident. Those who are ejected may impact other objects (such as a tree or pole) or may sustain crush injuries due to the vehicle's rolling over them.

Pedestrian clothing and impact sites to the body can reveal many clues with evidence about the circumstances of a collision. Questions one should ask include: Was the pedestrian standing or lying in the street? Was the individual run over or run under by the vehicle? Was more than one vehicle involved? Impact sites may reveal different front grill or tire pattern injuries. With standing pedestrian impacts, one should measure the distance between the impact site and the bottom of the foot or shoe the pedestrian was wearing at the time of the incident. It may be necessary to incise this region to visualize it better. It is good practice to include the measuring stick in the picture. This may give insight into whether the vehicle was braking before the impact occurred. As a driver jams on the brakes, the front end of a car will go downward. Fractures may occur to weight-bearing legs. Fracture patterns may be difficult to interpret when the bone is splintered and fractured into many pieces. Generally speaking, as a force passes through a bone, the fracture pattern

extends outward from the impact site similar to a skull fracture from an entrance gunshot wound producing internal beveling. With posterior standing impact, one might find stretch marks at the inguinal region opposite the impact site. This is caused by hyperextension of the hip and leg in an anterior direction.

Slow-moving vehicles tend to run over people. Fast-moving vehicles tend to run under people, meaning after being struck, the victim is tossed onto and over the hood. Flaps of skin may be torn away as a tire passes over a body. The clothing may yield significant evidence such as paint fragments. The car may have fragments of blood and hair that can be used for DNA analysis. This can be useful when there are multiple pedestrians and cars involved in collisions. Many pedestrian collisions involve children, who are sometimes impulsive and careless and may run out into oncoming traffic; the elderly, may not be quick enough to get out of the way; or individuals who have psychological histories, such as the homeless, or intoxicated individuals, who may think it's not that dangerous to cross a busy freeway.

Recent antemortem abrasions.

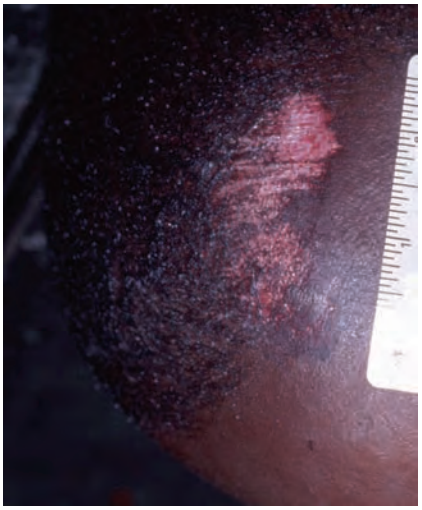

Recent deep antemortem abrasions.

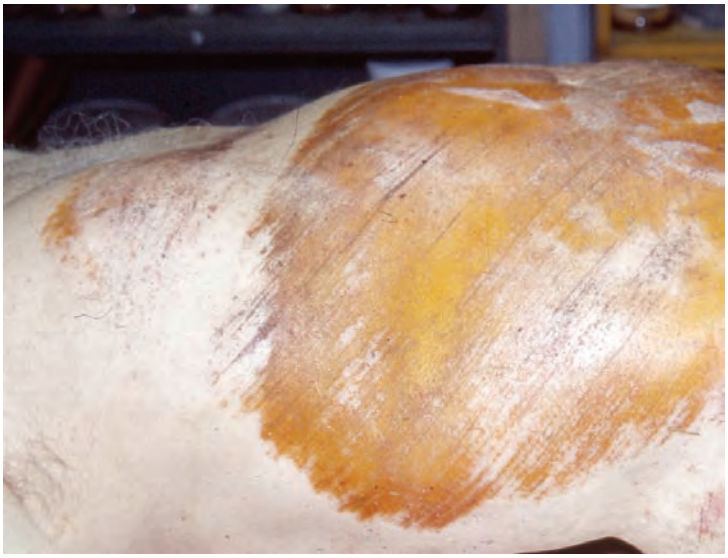

Large yellow anemic postmortem abrasion.

Recent contusion of the left hand associated with a fist fight. This individual punched the attacker before being murdered.

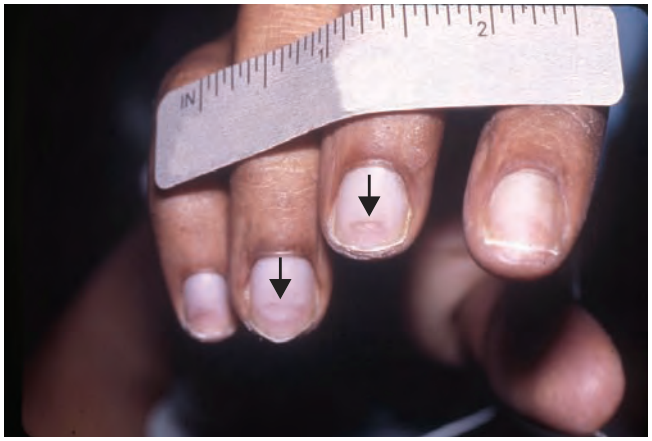

Damaged fingernails with nail bed contusions. Blunt-force injury during struggle.

Recent purple contusion of arm.

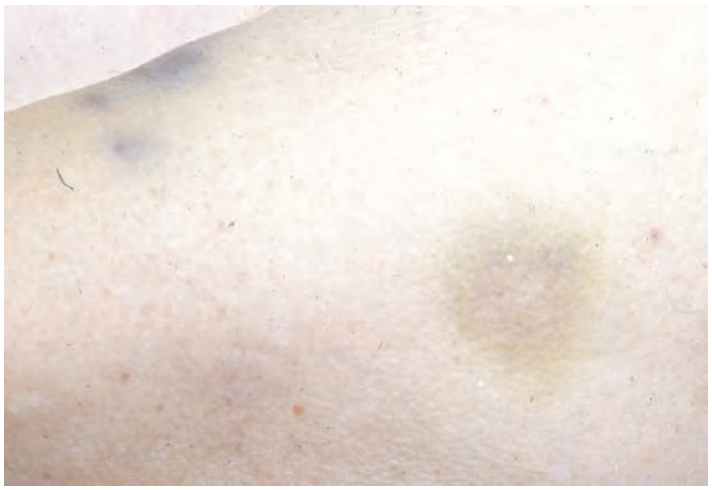

Older yellow-brown contusion.

Multiple abrasions and contusions of face. Note the bilateral periorbital ecchymosis associated with fracture of the anterior cranial fossa. This is also known as "raccoon eyes." A similar injury is Battle's sign, hemorrhage behind the ear, also a sign of basilar skull fractures. Figure (c) demonstrates a basilar skull fracture.

Recent laceration with adjacent abrasion. Note the irregular margin at the point of skin separation due to the skin's ripping or tearing apart.

Recent laceration of scrotum.

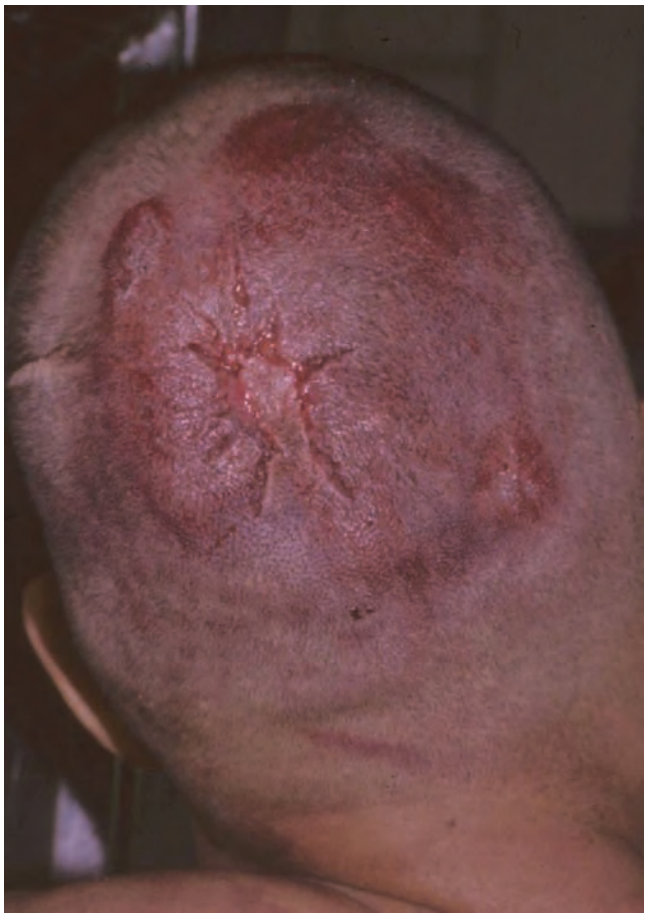

Large recent abraded contusion at the back of the head with a stellate laceration. This individual was intoxicated at a party on a rooftop. He was taking a group photograph and while backing up accidentally fell three stories to the pavement below. He had a comminuted skull fracture.

These are multiple recent lacerations and contusions at the inner aspect of the mouth following multiple blunt impacts to this individual's face with a fist.

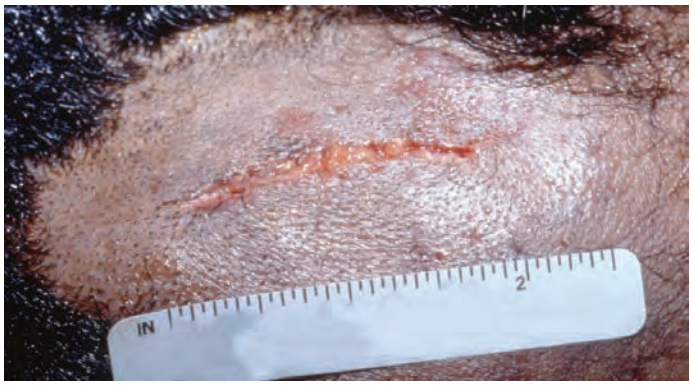

This is a postmortem anemic laceration in an individual who was stabbed to death in an apartment and later thrown out of a window.

This decomposing body has skin slippage with separation of scalp hair. Note the scalp lacerations with slight to moderate decomposition.

This individual was struck by a train. Note the soft tissue avulsion underneath the intact thick skin.

This individual died of multiple homicidal blunt- and sharp-force injuries. These are scalp lacerations in an individual with moderate decomposition. Note the extensive drying of the wound margins with clotted blood and fragments of hair. Also note the separation of the scalp from the skull with sub adjacent fractures.

Note the recent abraded contusion to the middle aspect of the chest (a). This individual had multiple layers of clothing and was reportedly stomped on by an individual with heavy boots following assaults with other weapons. The abraded contusion to the middle aspect of his chest forms a rough outline of a boot (b). The multiple layers of clothing prevented a more discernable defined boot pattern. The individual had a fractured sternum and a cardiac contusion (c).

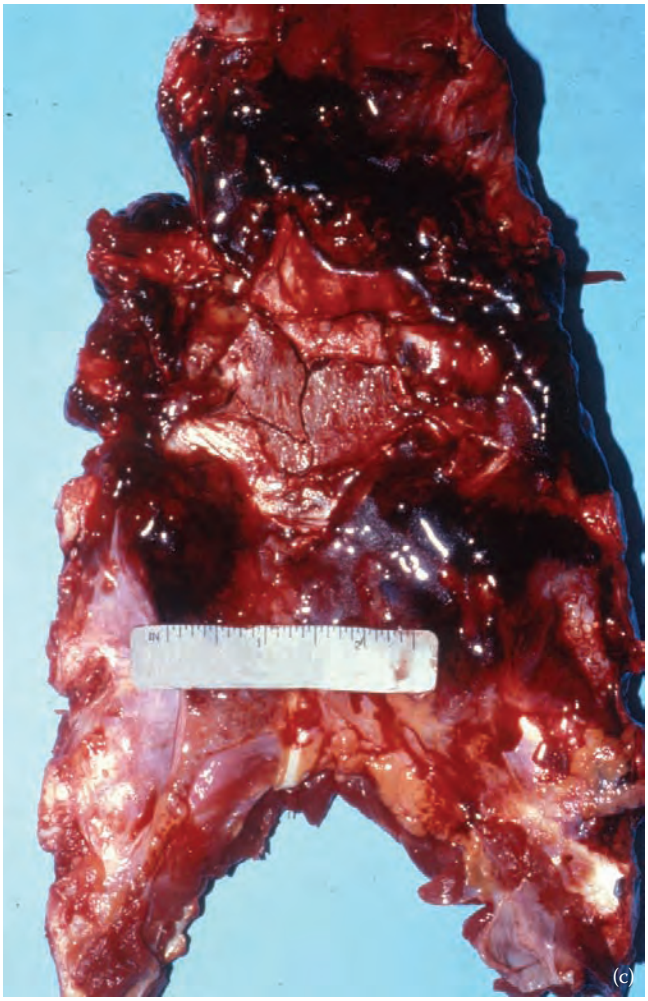

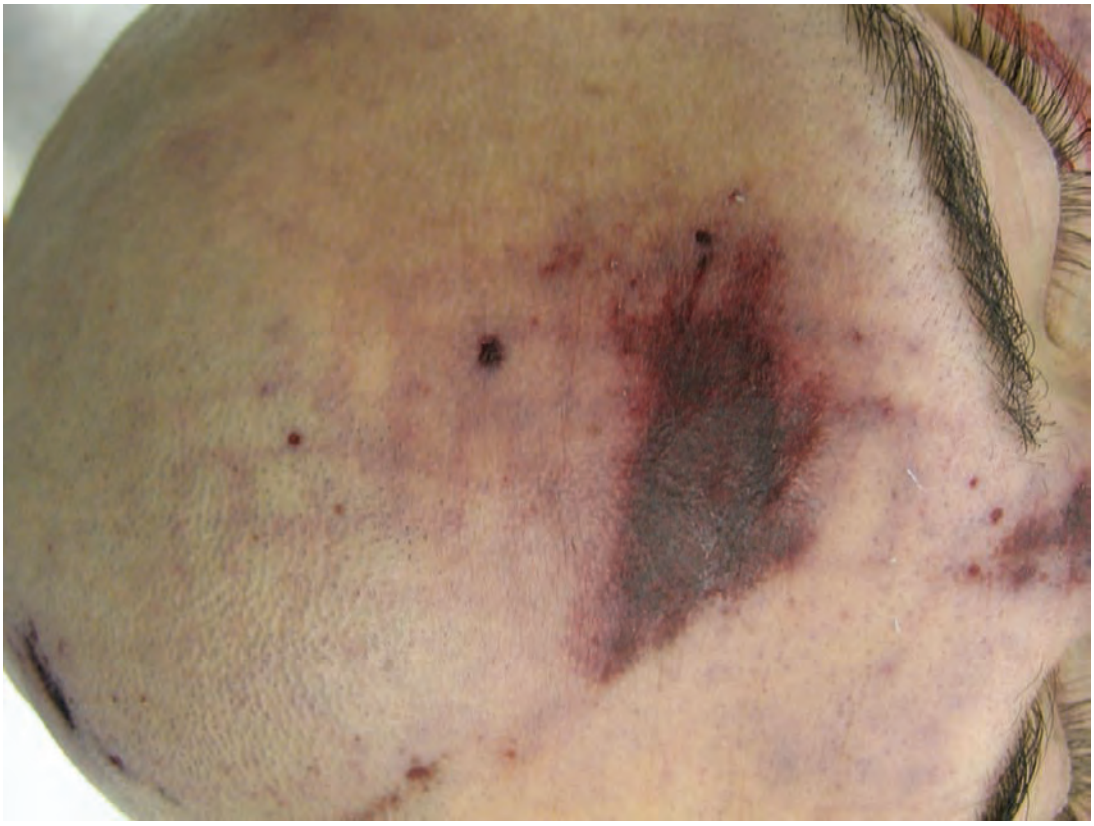

Pattern injuries associated with impacts from the bottom of a shoe while being stomped and kicked.

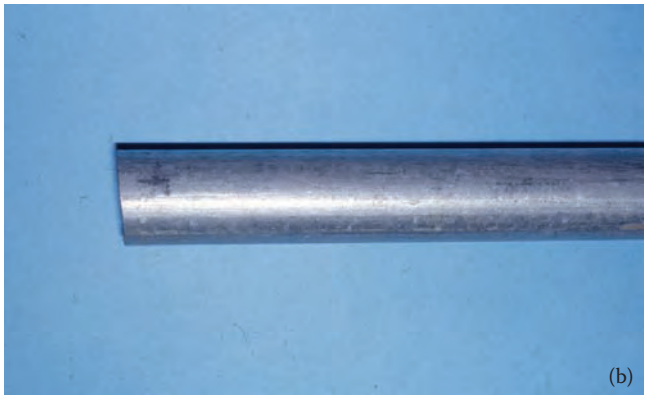

This individual was assaulted with fists and a metal pipe or rod. Note the parallel linear contusions at the posterior aspect of the left thigh (a). Depression of the soft tissues contacting the rod causes stretching at each margin with blood vessel injury and parallel linear bruises.

This individual was struck multiple times with a putter golf club. Note the pattern injury at the forehead and face, with the shaft of the club extending at the inferior aspect.

This individual was assaulted by multiple people with bottles and a baseball bat. These abraded contusions at his face and head are characteristic of a baseball bat impact. Note the oval contusion with sparing of the central aspect with overlying abrasion. Also note the abraded central region.

This elderly women had a history of an unsteady gait associated with Parkinson's disease and remote stroke. She was found in her apartment with several impact injuries to her scalp. It was initially thought by investigators that she had fallen several times and possibly suffered a heart attack. Further examination of her scalp revealed more lacerations and impacts that were initially not observed at the scene due to poor lighting and dried blood matted in her scalp hair. These scalp lacerations were produced by being struck with a hammer head.

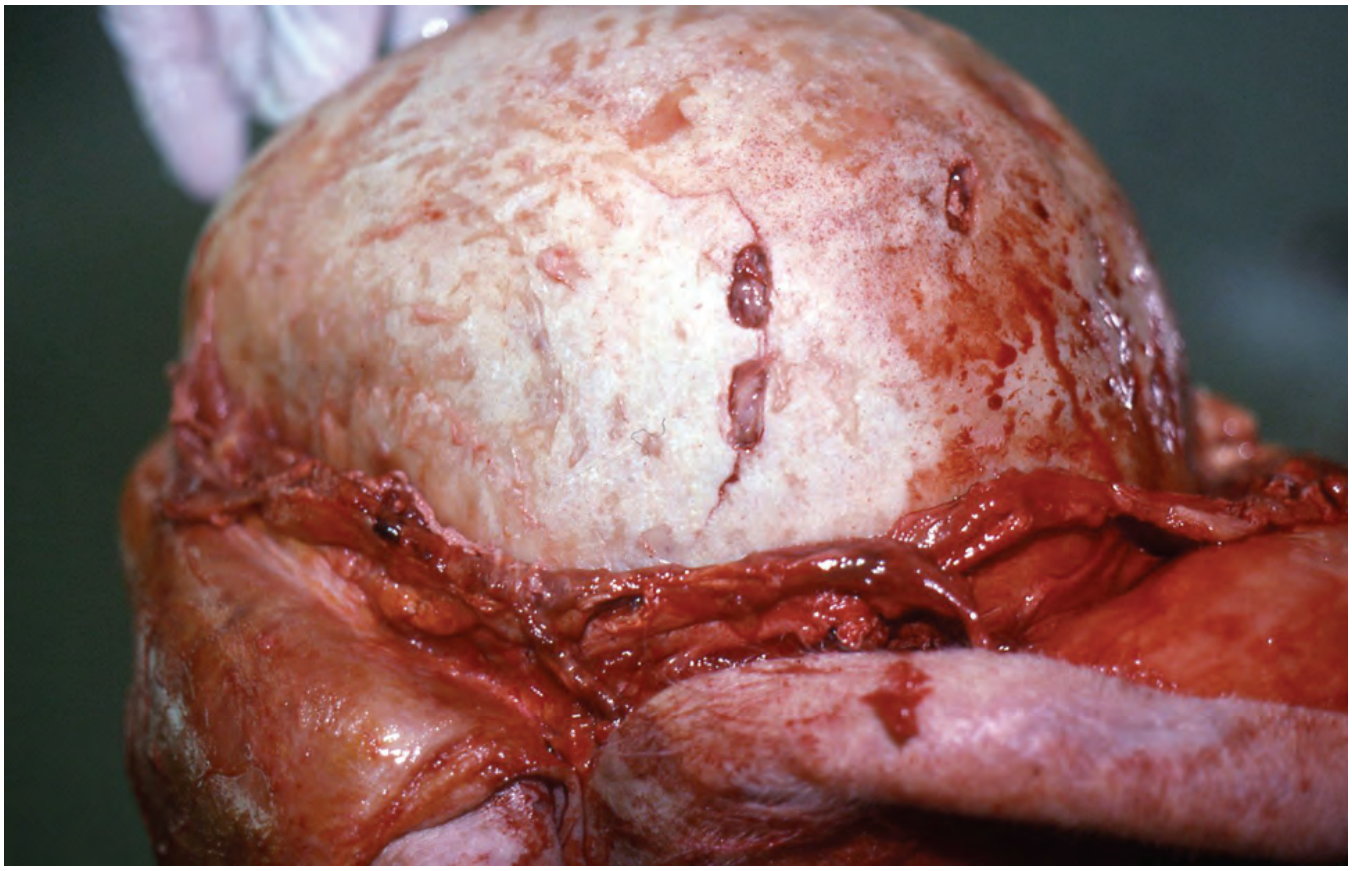

Depressed skull fractures due to multiple impacts with the head and claw of a hammer. Standard household hammer heads have a diameter of 3/4 to 1 inch, and the injuries on the skull tend to reflect this.

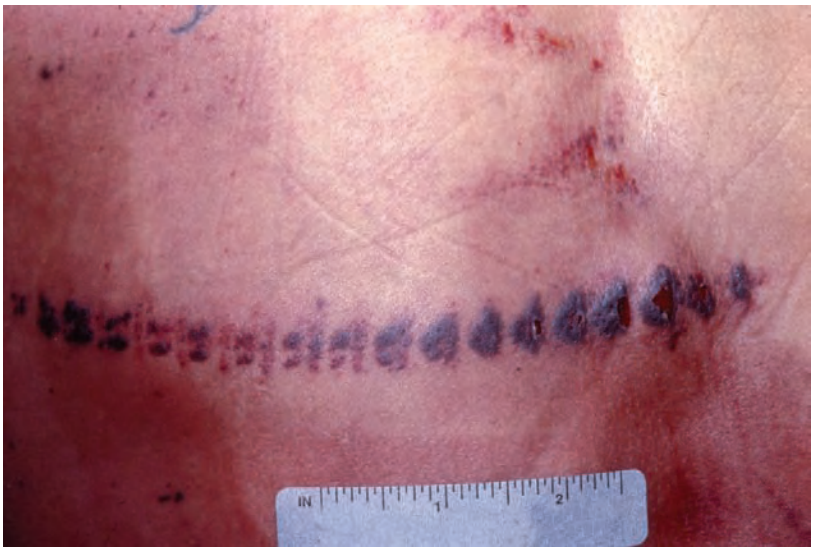

This individual was assaulted by multiple people including one with a bicycle chain. These injuries are characteristic for bicycle chain impacts.

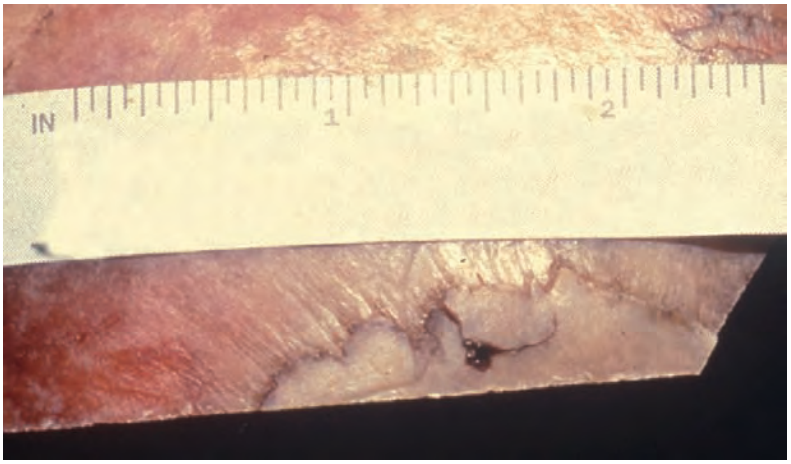

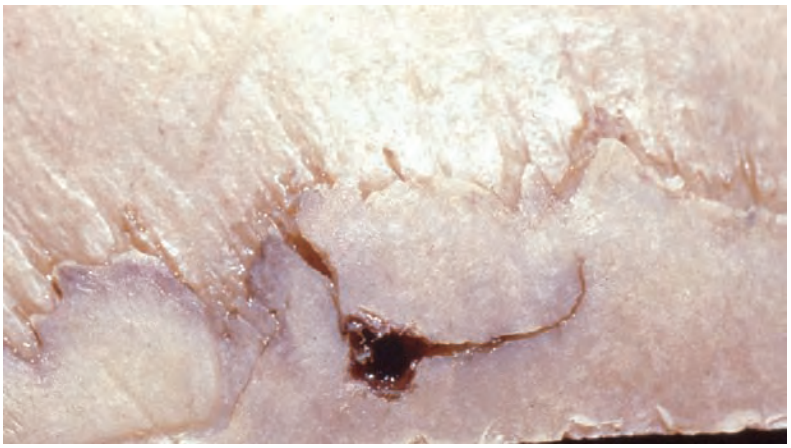

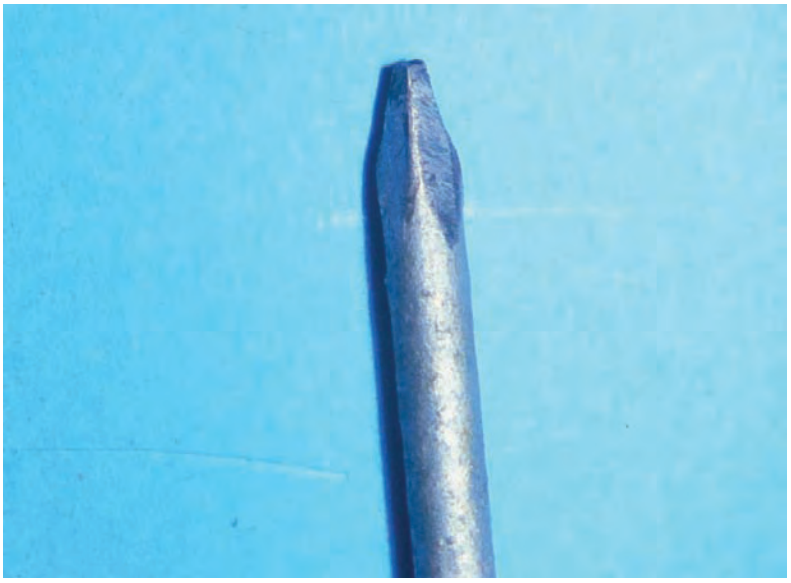

This individual was found naked from the waist down in an abandoned pipe yard. She had multiple defects to her scalp with exposed bone and extensive maggot and animal feeding. Careful inspection of the skull revealed a 1/4″ linear fracture of the superior temporal bone with a 1/16″ roughly square indentation. Further investigation of the scene revealed a board with a nail at the other side of the yard. Further testing confirmed this was the murder weapon.

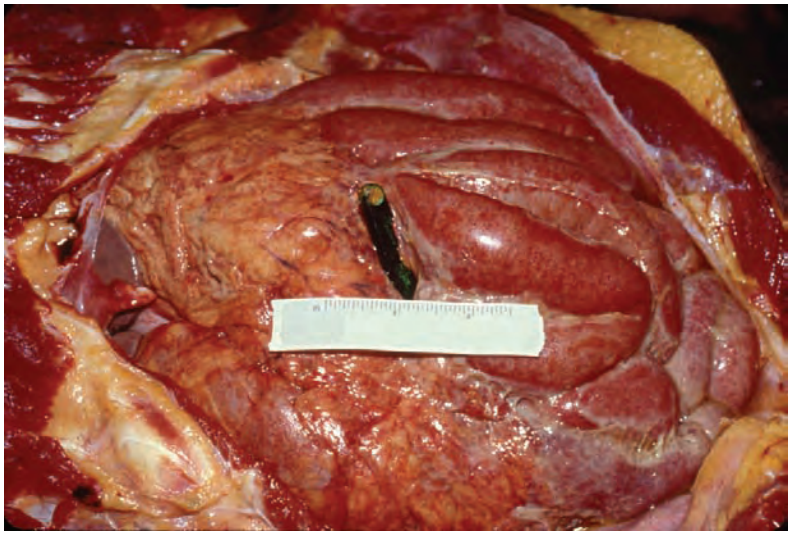

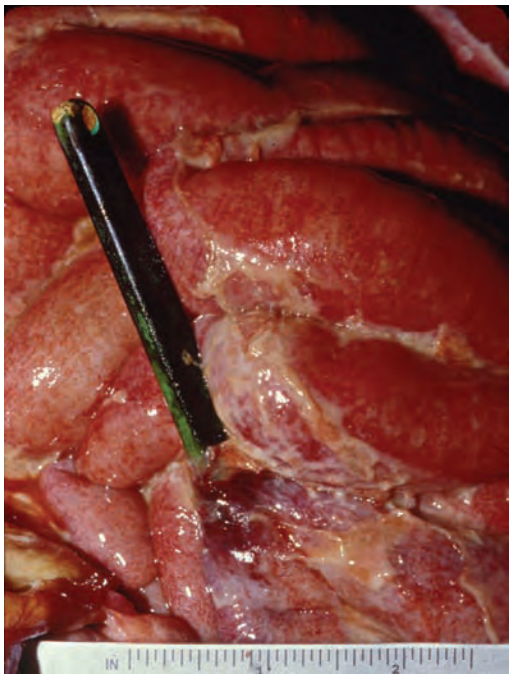

This psych patient had a history of swallowing miscellaneous objects. The blunt side of this toothbrush, in conjunction with peristalsis, eroded through the intestinal wall. This person died of peritonitis with septic complications.

Pierced ear lobe with sociopathic pirate earring.

This is a homicide case where the individual had her earring torn from her ear lobe during the assault.

Note the slit elongation of the piercing site from wearing very heavy earrings. Also note the healed linear scar to the left due to a traumatic tearing of the earring from the ear lobe with complete separation and nonplastic-surgical repair.

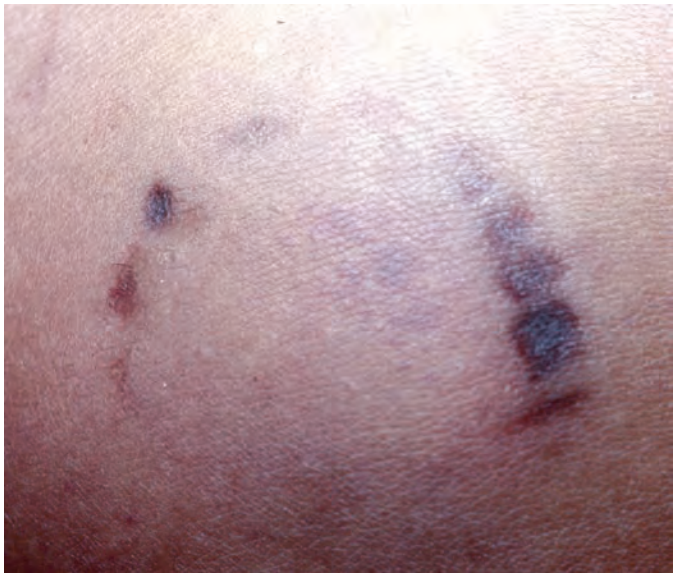

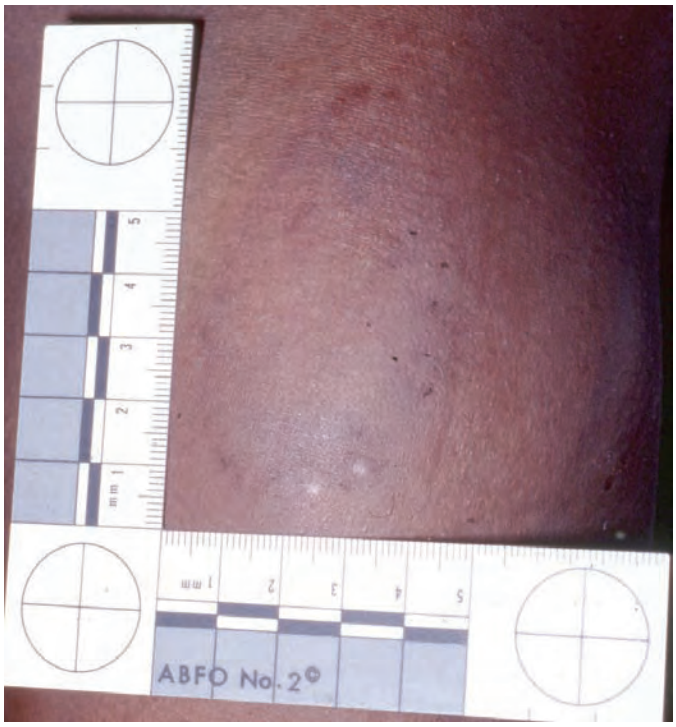

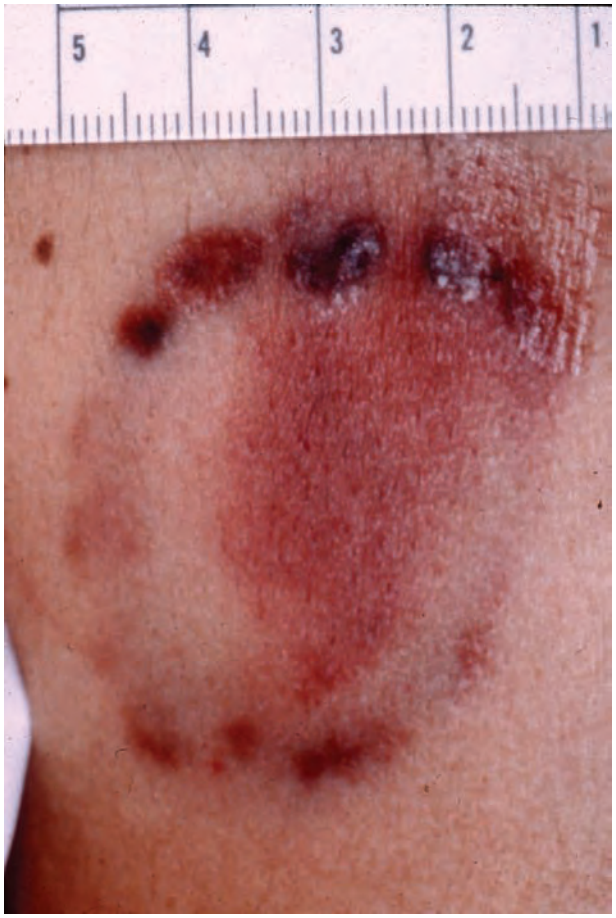

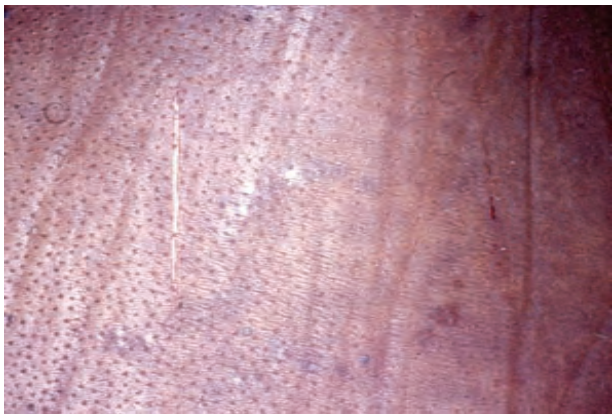

These are various examples of recent bite marks with different placement of the mouth while exerting pressure and different degrees of force applied. Note some patterns are very vague and others are extremely prominent. The typical example of a bite mark reveals a circular pattern with a central region of contusion. It is good practice to consult a forensic dentist as soon as possible whenever a bite mark is suspected. The injury should not be cleaned until swabs are taken to detect oral DNA left behind from the perpetrator. As time goes on with drying and decomposition, the injury may yield less valuable information for dental comparison and DNA analysis.

Image (a) demonstrates both a recent and a healed bite mark. The old bite mark is largely healed with hypo-pigmented deep abrasions from teeth dragging across the epidermal surface.

This individual and her roommate were found in their apartment with multiple blunt- and sharp-force injuries. Note the roughly semicircular lacerations at the superior and inferior aspect of the cheek, with the deeper lacerations of the lip revealing exposed underlying teeth. There was a large cylindrical storefront padlock within a tube sock found at the scene. There were multiple other pattern injuries to the decedent's body consistent with these roughly circular impacts.

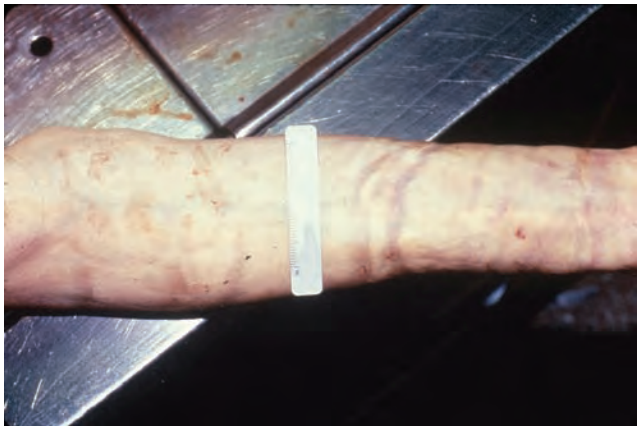

Note the furrow indentations associated with extremities bound by rope and cord prior to the victim's being stabbed to death.

This individual suffered a witnessed fatal cardiac event while standing. While falling to the ground he sustained this pattern injury by striking his head on a radiator.

This individual suffered a fatal cardiac event while riding on an escalator. He was found lying at the bottom of the escalator. Note the pattern injury to his arm from the continued rubbing of escalator treads.

This individual impacted a fence at a high rate of speed, leaving a pattern injury from the top of the fencepost's contacting his chest.

This individual was an unrestrained passenger in a motor vehicle during a head-on collision. His face went through the front windshield. He sustained multiple curvilinear lacerations and sharp-force injuries from broken glass and impact with the car roof. This type of pattern is consistent with a front windshield impact.

Note the dicing type injuries to the scalp with cubed fragments of glass embedded in the injuries. This type of injury is consistent with an impact and fracture of tempered glass, which is present in many side windows. These individuals were passengers with left-door-window head impact.

This individual was the passenger in a front seat collision. Note the abraded contusion from his seat belt.

Note the abraded contusion at the chest and arm of this individual from a steering wheel impact. He was the driver of this car in a multiple-fatality vehicular accident.

This was the driver of a car in a motor vehicle accident. Note the abraded contusion from the seat belt.

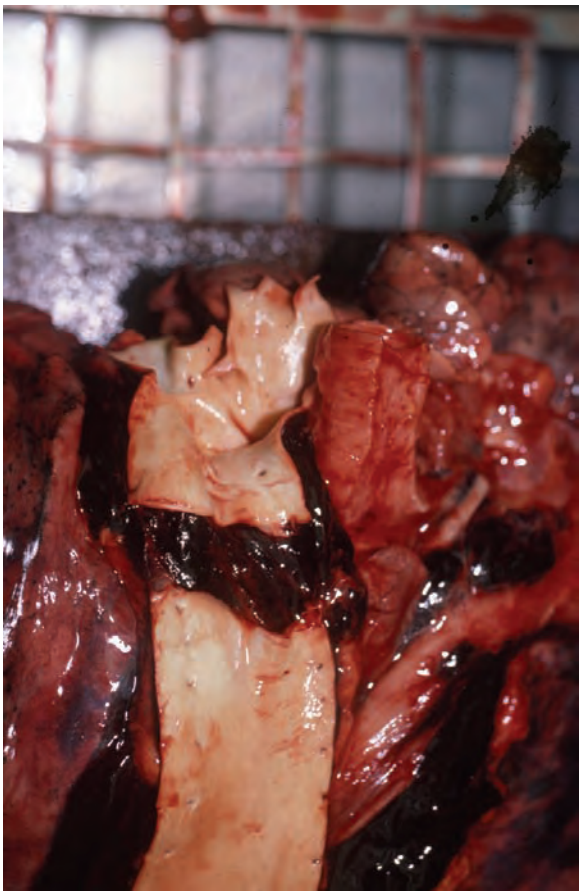

These demonstrate a transection of the aorta in the proximal descending region just distal to the ductus ligatosum. This is a common location of laceration.

This photograph demonstrates an individual who was struck at a high rate of speed and hit the pavement with great force. Note the exposed comminuted skull fracture with brain material.

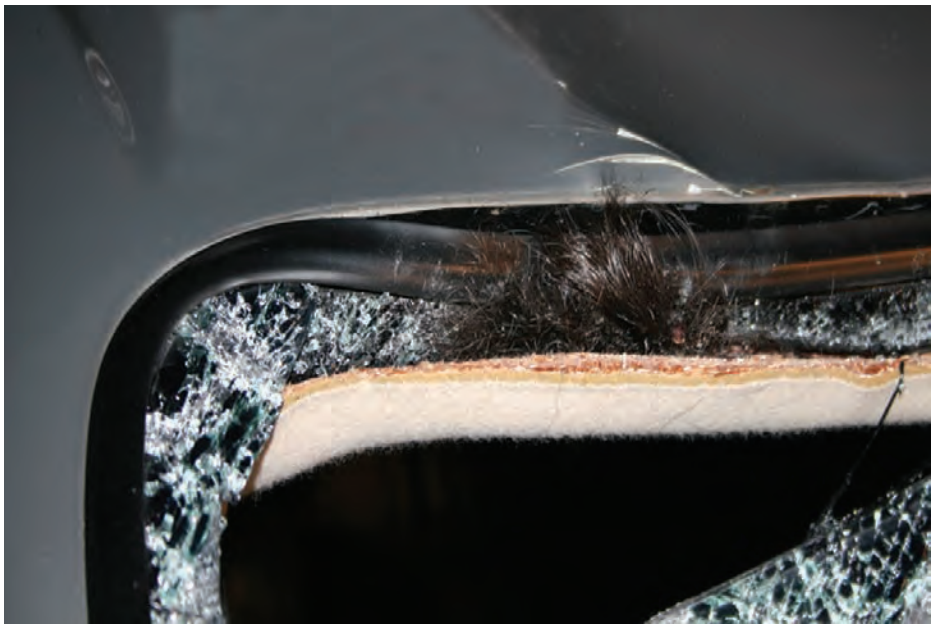

These pictures of a car show an impact from a pedestrian who was struck and run under. Note the fragment of scalp with hair embedded in the top part of the windshield and adjacent car roof. The driver of this car initially fled the scene.

Pedestrians struck by cars with grill pattern injuries.

Motor vehicle accident with decapitation.

This is another pedestrian who was struck by a car. Note the tibial fracture at the weight-bearing left leg.

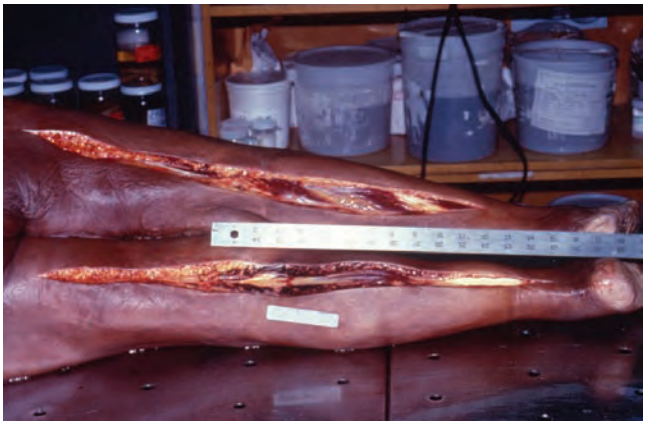

This individual was a pedestrian who was struck from behind by a car. Due to her dark skin, the contusions are not obvious from external examination alone. Incision of the posterior aspects of her leg reveal hemorrhage due to the bumper impact. It is good practice to photograph these impact sites with a ruler to demonstrate the distance from the decedent's heels. This can be matched to a particular car and to whether the driver applied brakes before striking this pedestrian. The front of the motor vehicle will go downward when the braking occurs.

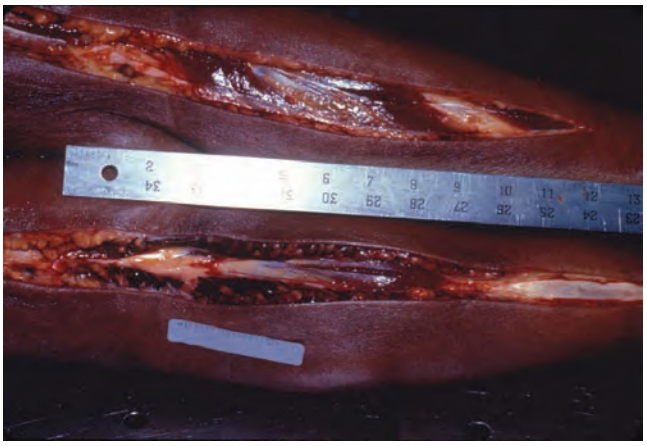

These individuals were both struck by cars and dragged across the pavement. (a) shows yellow anemic abrasions that occurred after the first impact, where the decedent sustained extensive central nervous system injury and a transected aorta. He was thrown into another lane of traffic and dragged by another car. The anemic nature of this injury and yellow discoloration suggests decreased blood perfusion. The injury in (b) shows red to brown discoloration, which is significant for vital reaction in an individual who had an intact beating heart with blood pressure.

These are multiple examples of tire imprints from people who were run over by cars. In one case, the individual was thrown into another lane of traffic after being struck. The second car denied hitting the individual, but his tire pattern was a perfect match and there was forensic evidence found on the under surface of his car.

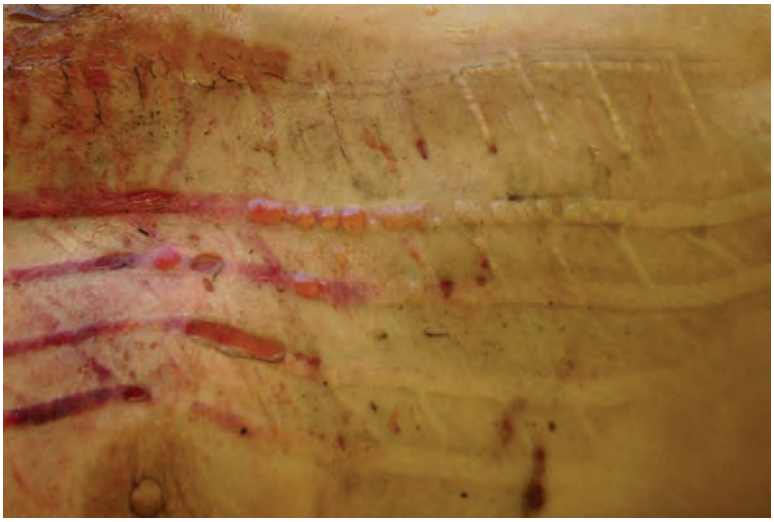

Individual trapped under a car with tire imprint to chest.

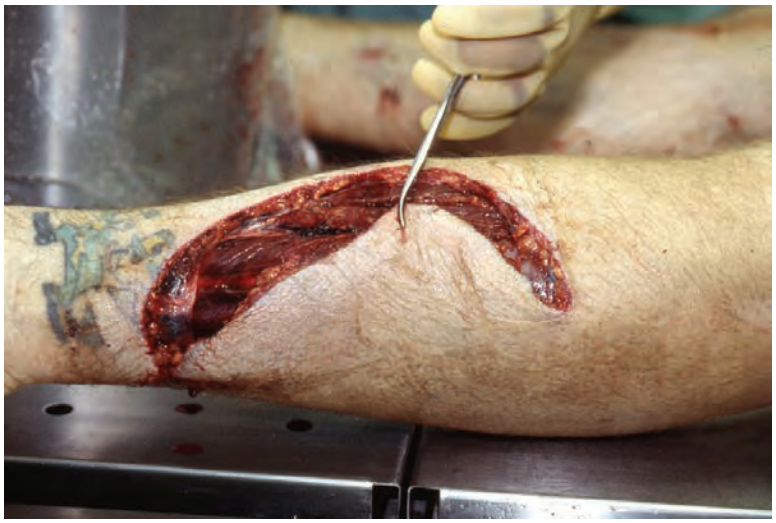

These individuals were run over by a motor vehicle while lying on the ground. Note the flap of skin being torn away from the thigh as the tire rolled over the leg.

These individuals were struck from behind by a car. Note the inguinal stretch marks caused by hyperextension of the hips and legs at the time of impact.

This fractured motorcycle helmet shows blood clot and brain material from an individual who fractured his skull after striking a pole.

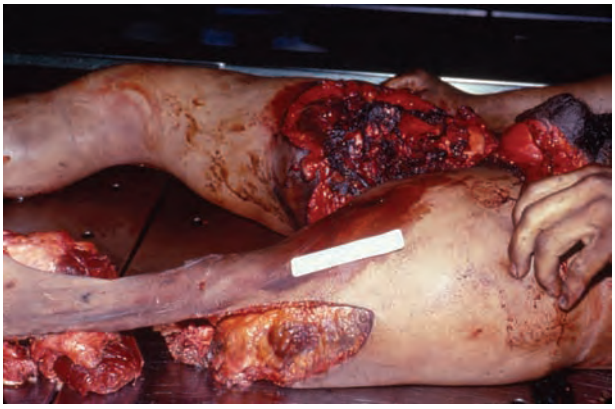

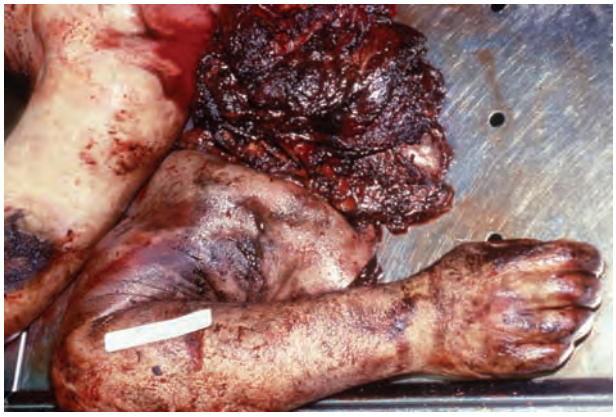

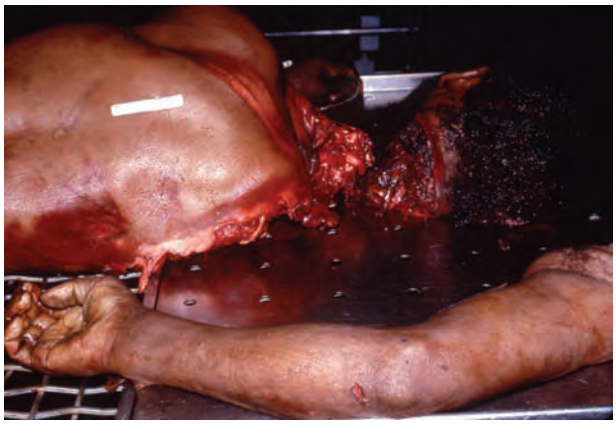

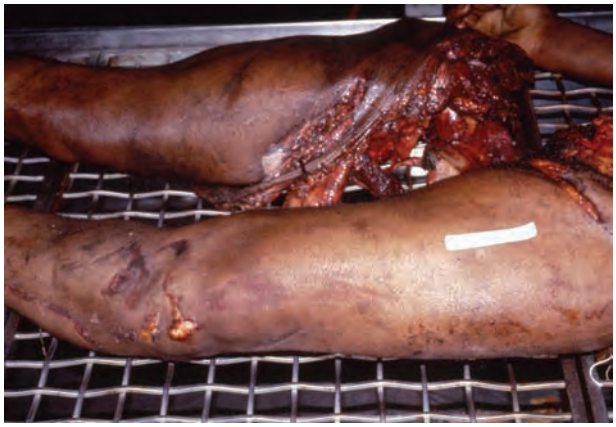

These figures demonstrate various subway fatalities where individuals were run over by trains. Note the separated body portions with extensive crush injury and axle grease from the train wheels.

This individual fell from the top of her steps and sustained multiple blunt-impact injuries including a femur fracture and a skull fracture. Note the slipper at the top steps.

This individual has a femur fracture. Note that the fractured leg is shorter and the foot is laterally rotated.

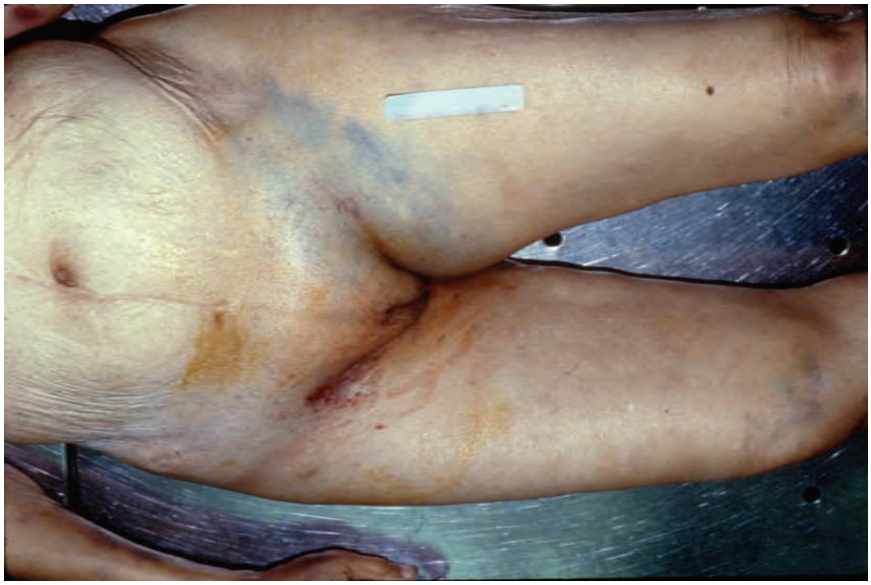

People with significant osteoporosis may fracture bones with little force. These two pictures demonstrate large areas of contusion and ecchymosis following a femur fracture secondary to a standing height fall.

Angulated irregular abrasions of back due to striking a gravel bed with great force.

This decedent jumped from a great height and fractured his right femur. Note the fractured bone protruding from the thigh. Also note the shortening of the right leg with lateral rotation.

Ring fracture. This usually occurs following a fall with initial impact to the feet or buttocks.

The decedent fell from a great height and landed on his feet. The fractured tibia was forced through the bottom of the foot upon impact.

This individual was placed in a garbage dumpster and left to decompose. The trash compactor in the truck was engaged before the body was observed. Note the garbage truck shovel laceration to the back (a). Also note the intestines forced from the anus following compactor compression (b).

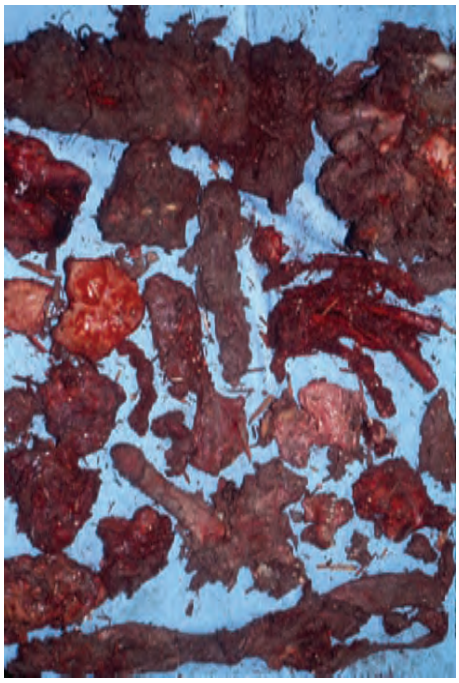

This individual was shot in the head, processed through a trash compactor, and left to decompose. Note the flattening of the body with extensive blunt-force injury and fragmentation.

This individual slipped into a terrain leveler and sustained multiple abrasions, with blunt- and sharp-force injuries including extensive fractures and internal lacerations.

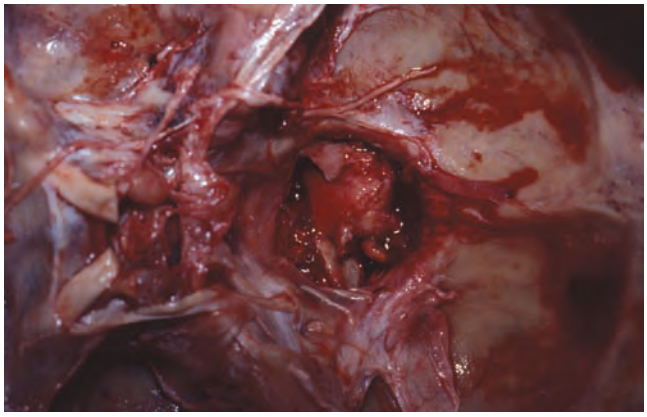

This individual had a severe blunt impact to his face that caused the atlanto-occipital ligaments to stretch and dislocate. This caused brainstem and upper cervical spinal cord injury resulting in asystole and apnea.

This picture of the base of the brain shows a medullary pontine laceration with extensive subarachnoid hemorrhage.

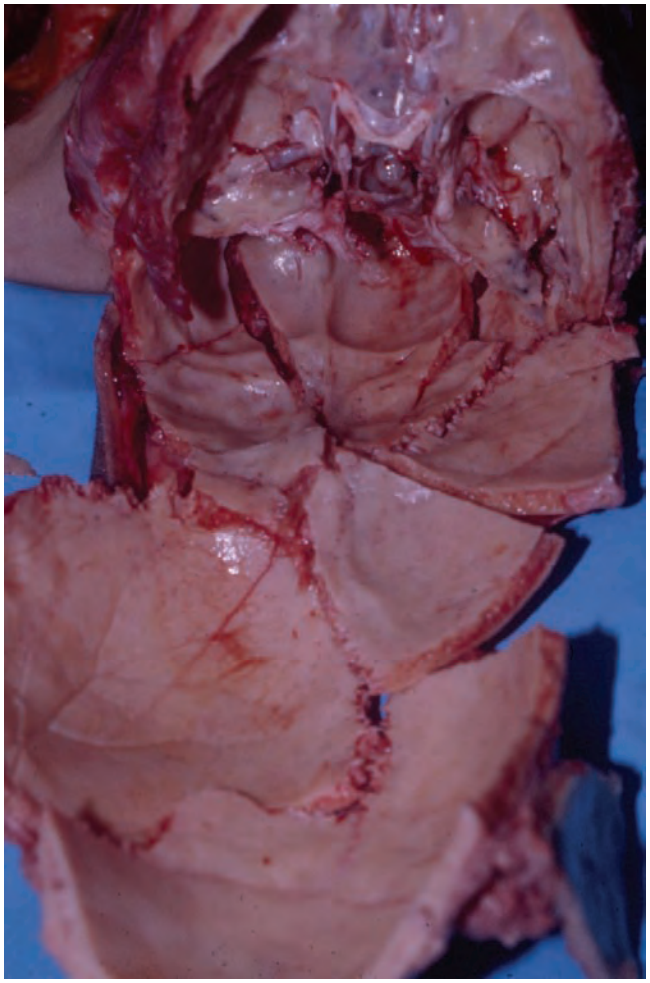

Comminuted skull fracture where the bone is broken into separate pieces.

Diastatic skull fracture showing the fracture site passes through a cranial suture.

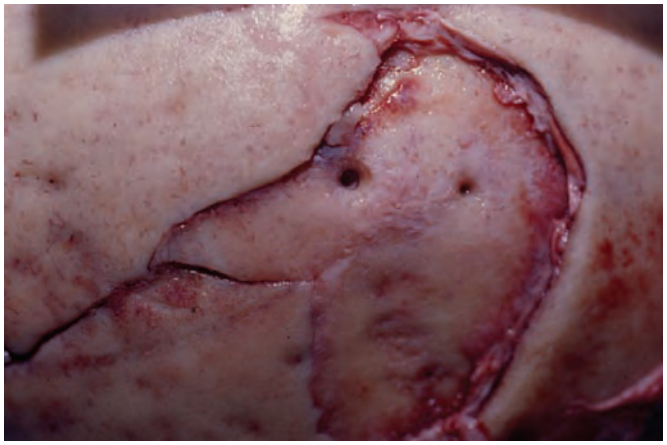

Healing skull fracture.

Frontal and temporal contusional hematomas associated with a falling impact to the back of the head in an individual with liver disease.

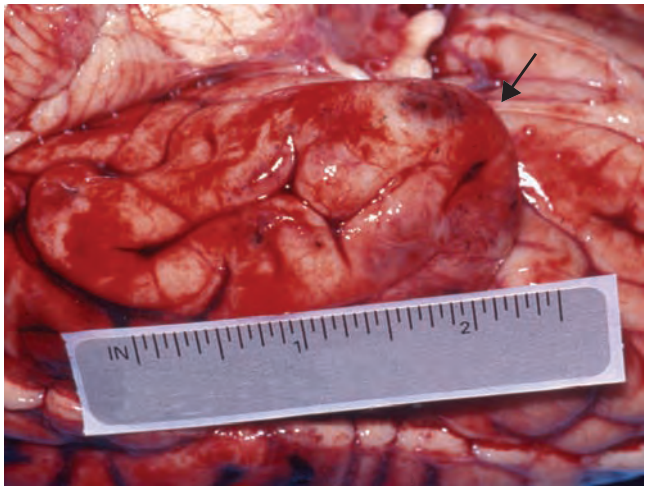

Temporal lobe with multiple splinter-type hemorrhages demonstrating typical cerebrocortical contusions.

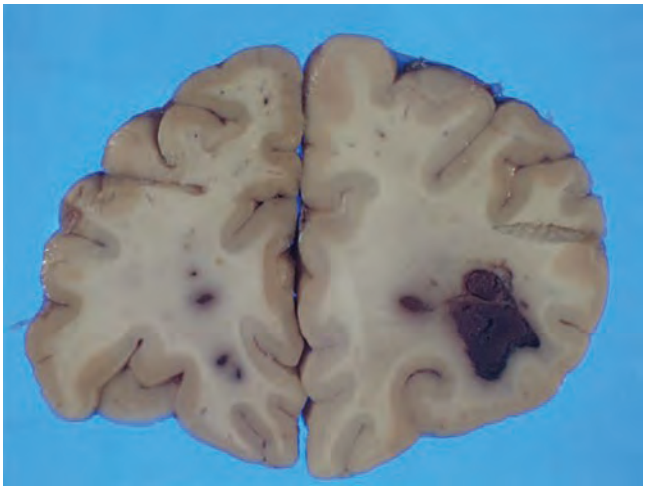

These images demonstrate gliding cortical contusions that typically occur near the white and gray matter junction following blunt-force head trauma.

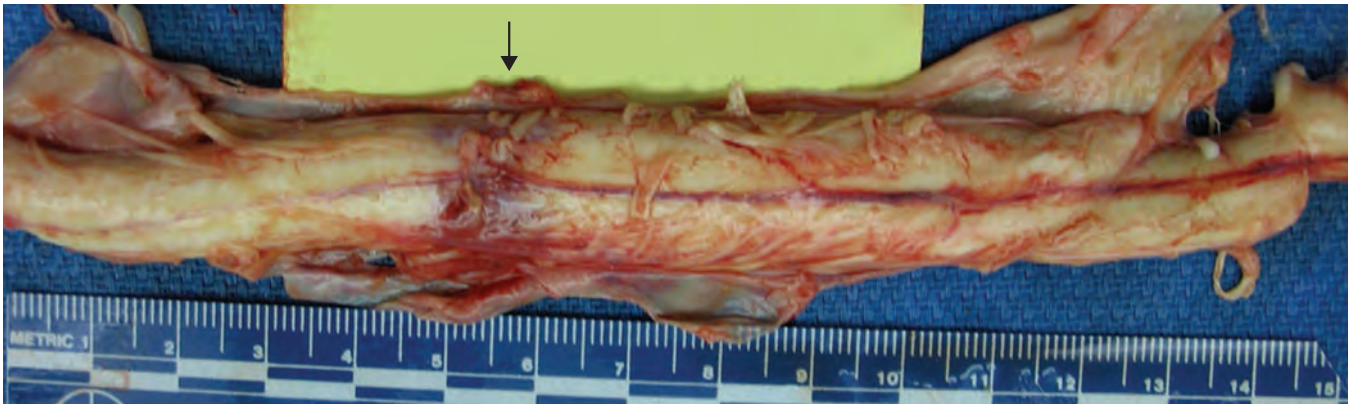

Spinal cord with a region of purple discoloration and softening demonstrating a spinal cord contusion.

Epidural hemorrhage in an individual with a temporal bone fracture and middle meningeal artery laceration. Note the clotted blood at the surface of the dura mater.

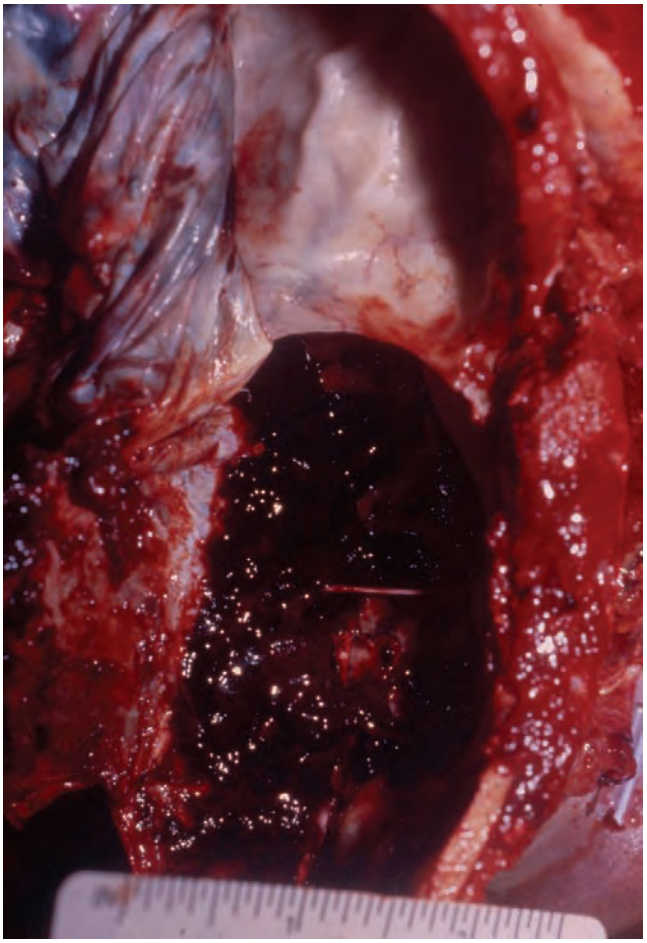

Epidural hemorrhage.

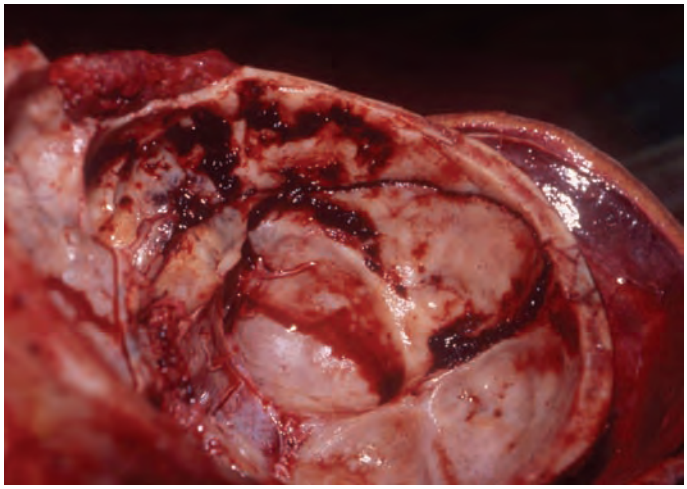

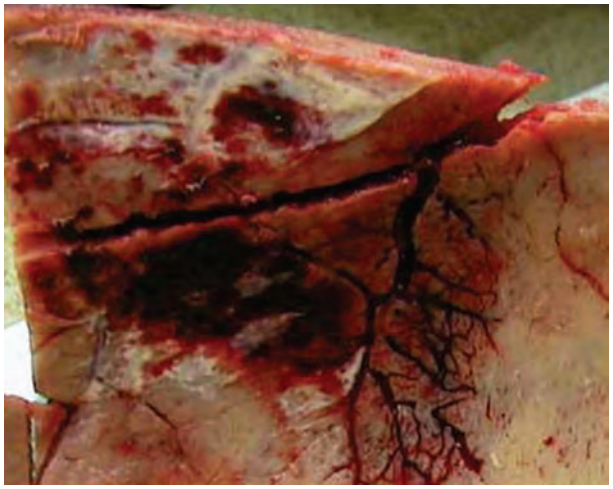

This shows the dura mater stripped from the inner aspect of the calvarium. Note the clotted adherent epidural hemorrhage within the temporal region of the skull.

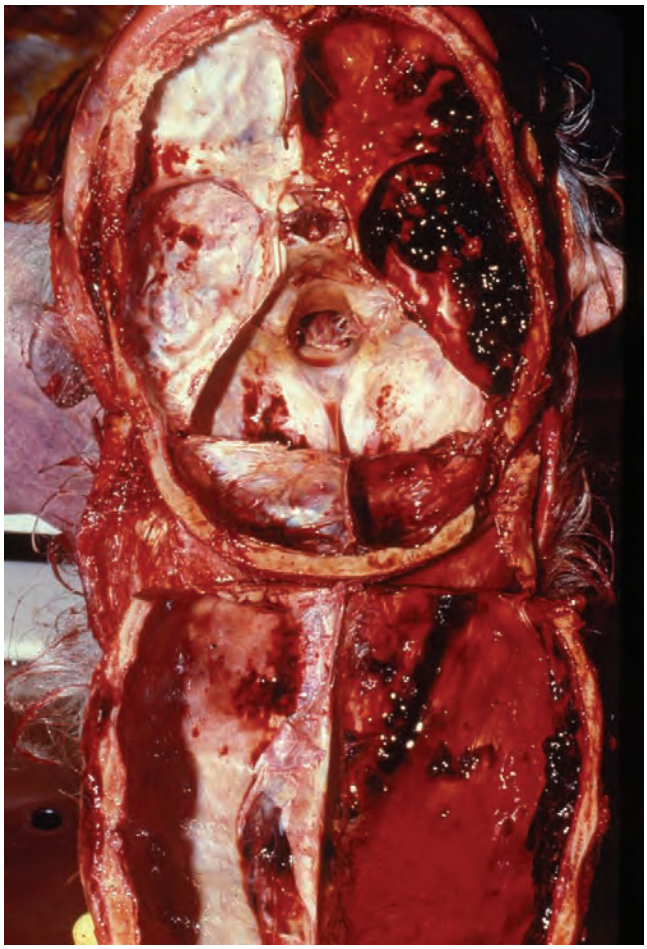

Acute right subdural hemorrhage.

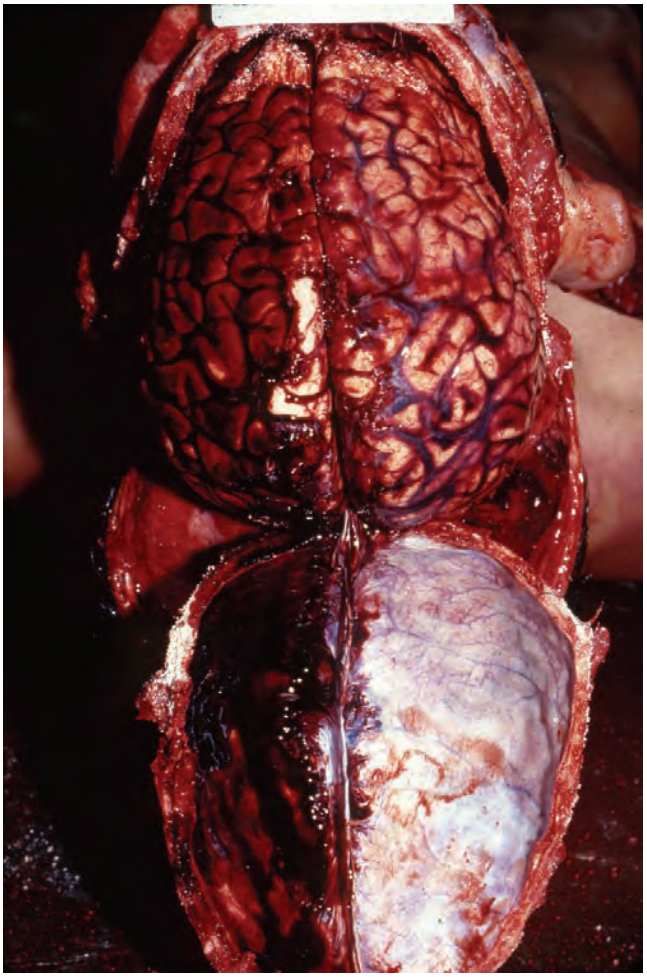

Left subdural hemorrhage that is slightly older with clotted adherent blood.

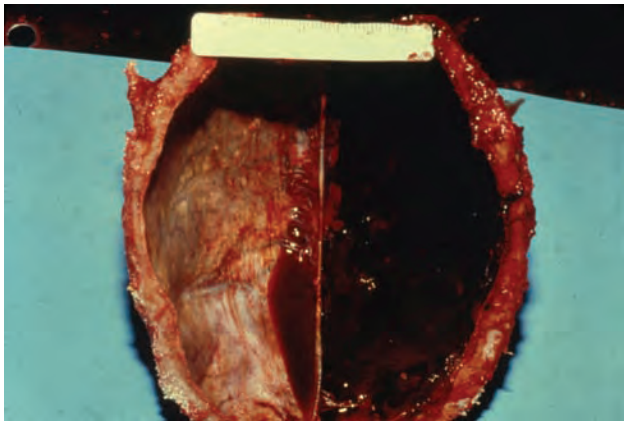

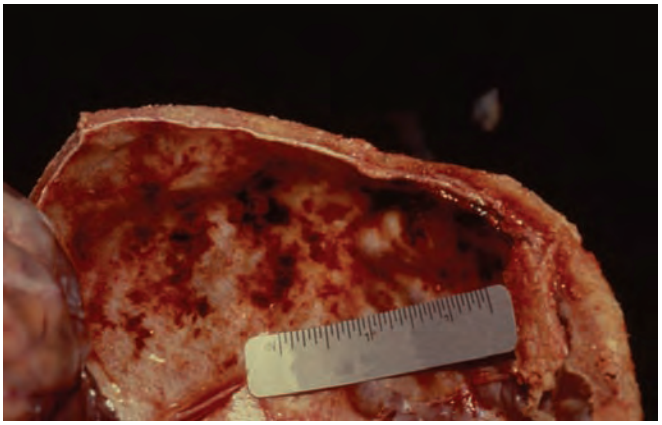

This demonstrates an older, clotted, red, adherent, subdural hemorrhage with portions of rust discoloration and membrane formation.

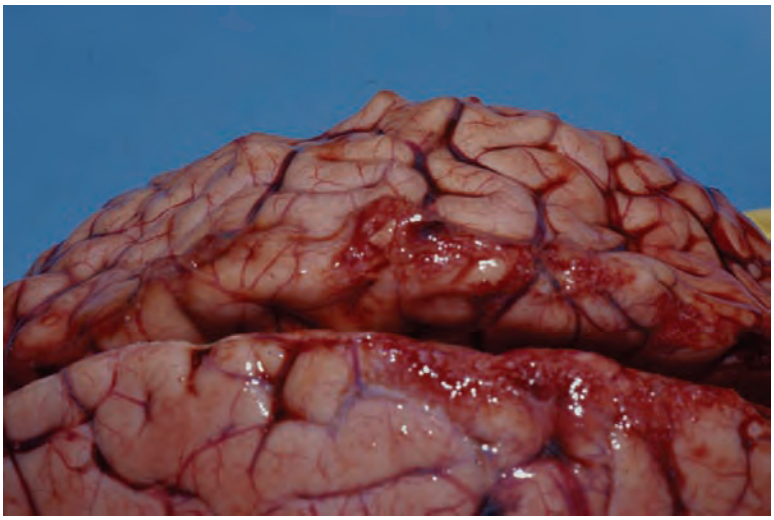

This cerebral cortex demonstrates an irregular flattened surface caused by a chronic subdural hematoma. In contrast, a chronic epidural hematoma generally leaves a flattened and less irregular cerebral cortex deformation.

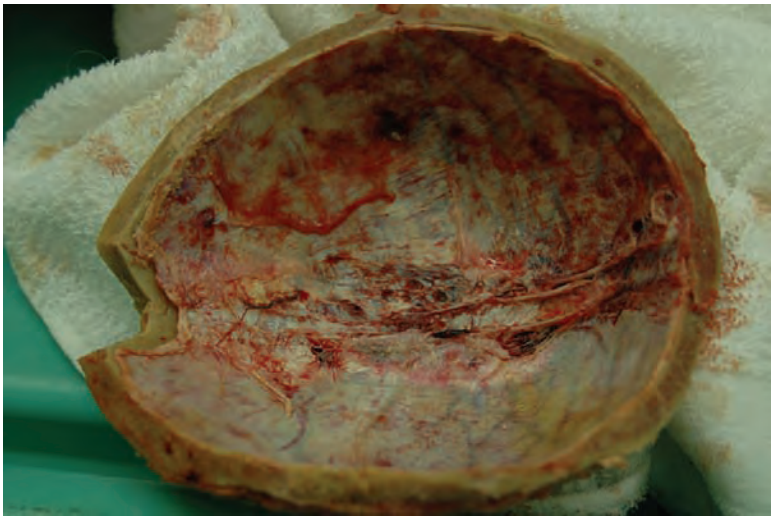

Old subdural hemorrhage with clear to rust-colored membrane formation. Note the membrane separation with beading up away from the midline caused by scraping a scalpel blade along the subdural surface.

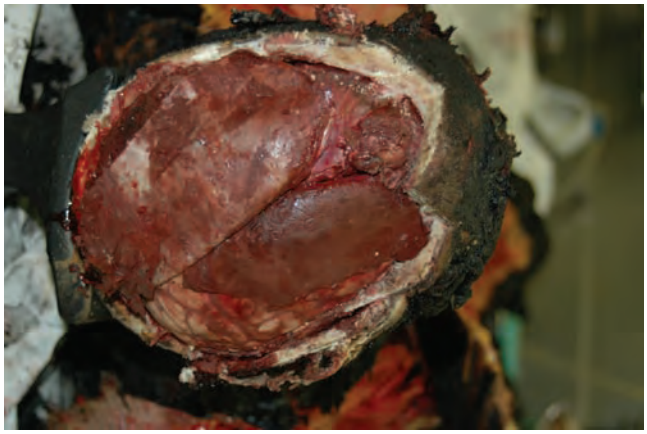

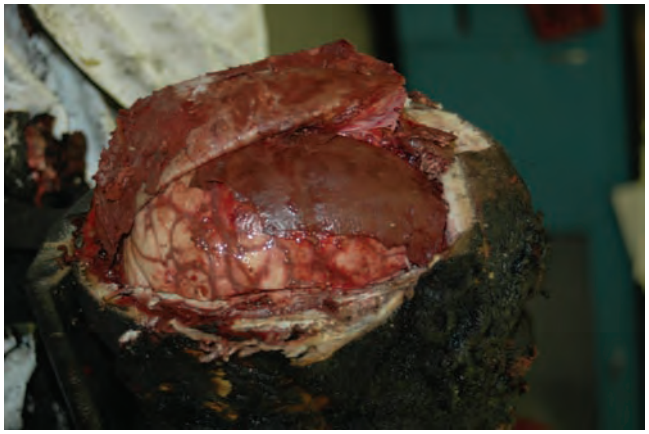

Large antemortem subdural hemorrhage altered by excessive heat due to fire.

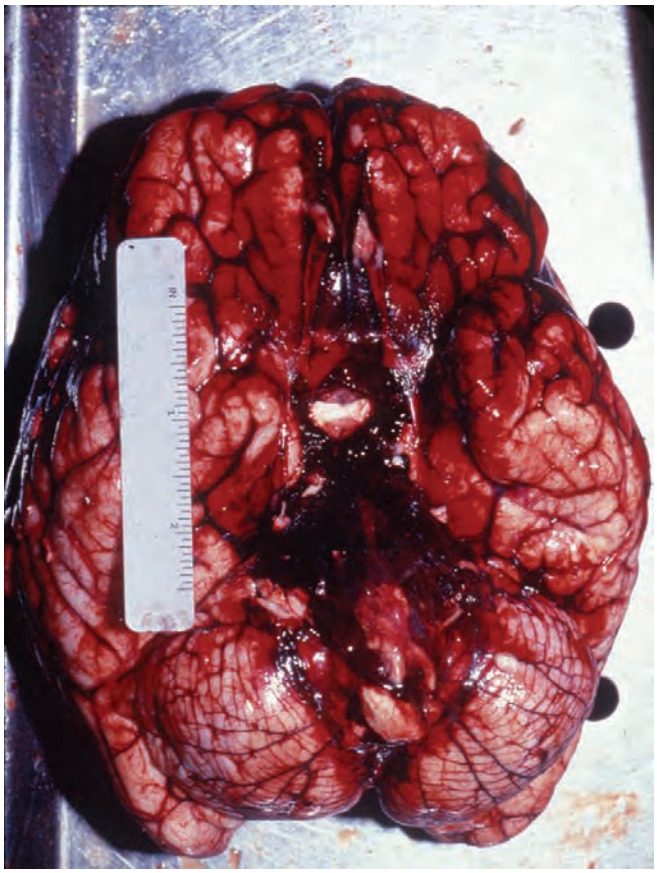

This brain demonstrates multifocal subarachnoid hemorrhage with subarachnoid hematoma at the base of the brain. This hemorrhage occurred following a blunt impact to the face causing hyperextension and rotation of the head with laceration of the right vertebral artery.
Source: Kindleburger D., Gilmore K., Catanese C., Armbrustmacher V. Subarachnoid hemorrhage due to internal carotid artery laceration following blunt impact to face with hyperextension and rotation of neck: A case report and review of the literature. *Journal of Forensic Science* 48:1366–1368, 2003.

This demonstrates a large subarachnoid hemorrhage in an individual who died several days following a traumatic head injury.

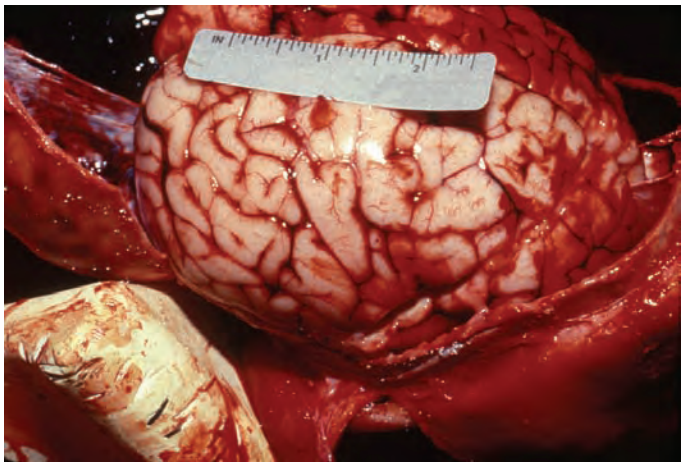

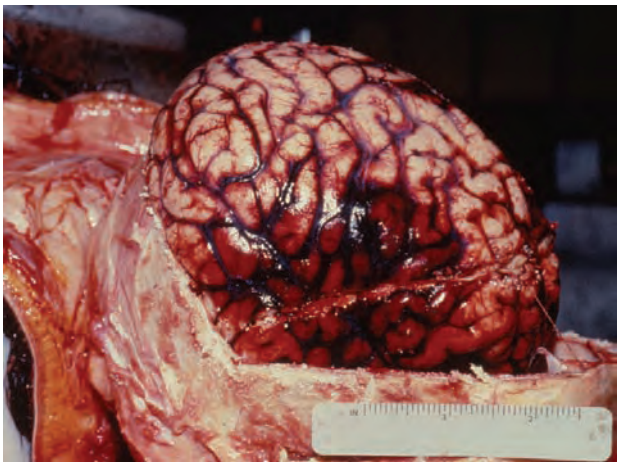

These brains are diffusely swollen with flattened gyral configuration and sulci obliteration. These individuals lived from several hours to several days after the initial insult.

Note the large areas of contusion with ecchymosis following minimal trauma in these individuals with chronic alcoholism, hepatic cirrhosis, and coagulopathy.

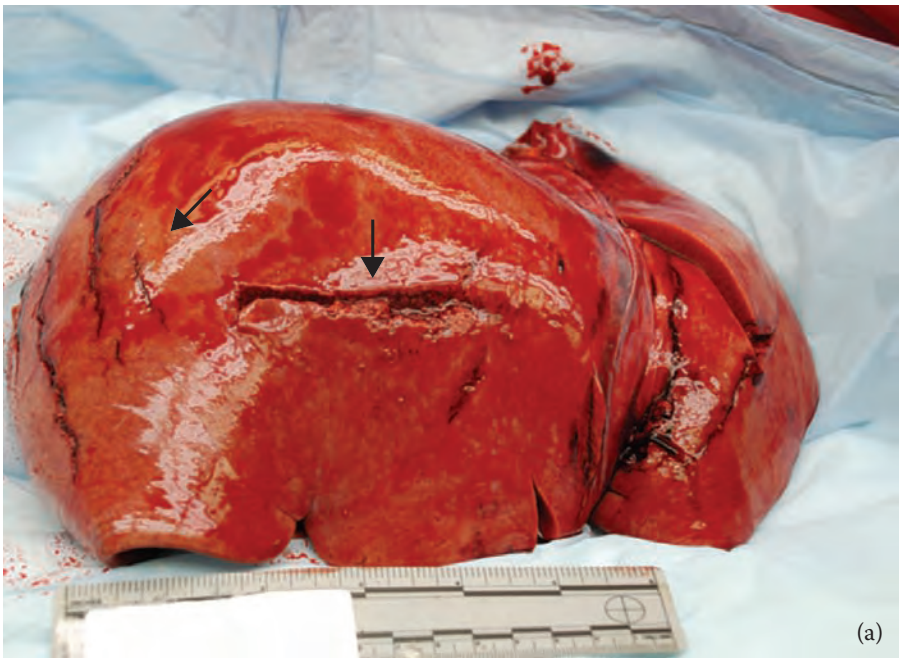

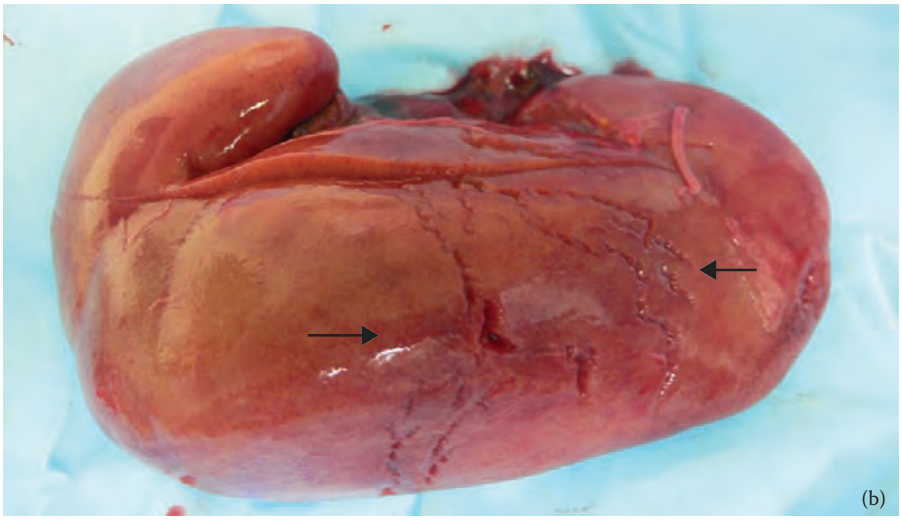

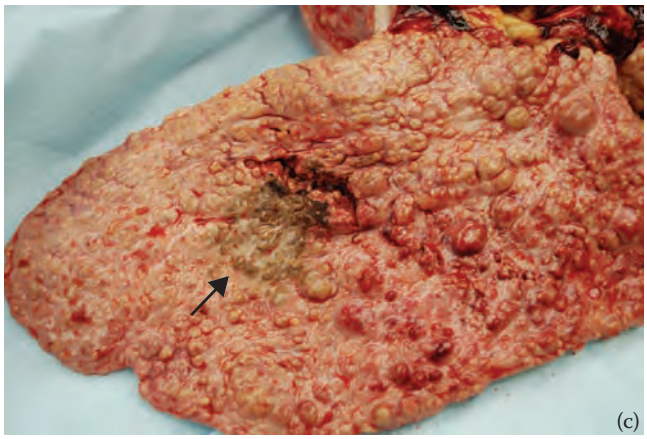

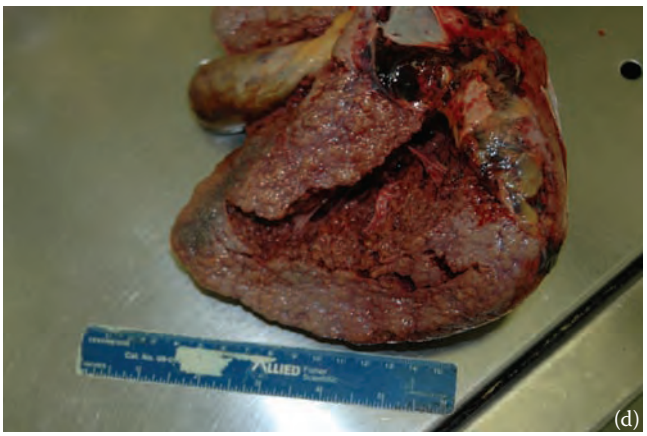

Lacerations of liver and kidney due to blunt-force trauma. Note that (c) and (d) have hepatic cirrhosis, which is less commonly associated with laceration due to fibrosis. A normal liver is the most common organ in the peritoneal cavity to lacerate (a) in association with blunt-force trauma.

This opened airway demonstrates a large piece of aspirated brain in an individual who was resuscitated following extensive blunt-force head trauma with comminuted skull fracture and communicating lacerations extending into the oral cavity. When Emergency Services arrived, the decedent was found in agonal respirations.

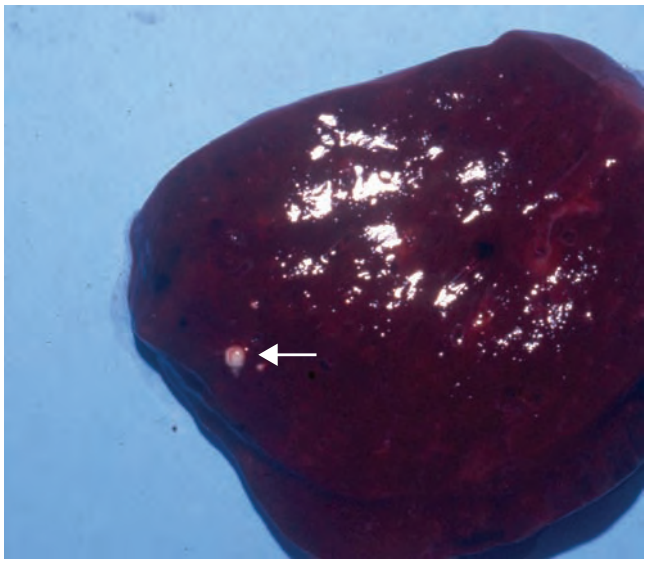

This cut section of lung demonstrates a small white to gray fragment in the parenchyma at the lower lateral aspect. Microscopic examination of this revealed intravascular embolic cerebellar tissue. This individual sustained a comminuted skull fracture with multiple central nervous system lacerations. He was vigorously resuscitated several times over the course of several hours. Brain tissue embolized through damaged blood vessels at the base of the brain.

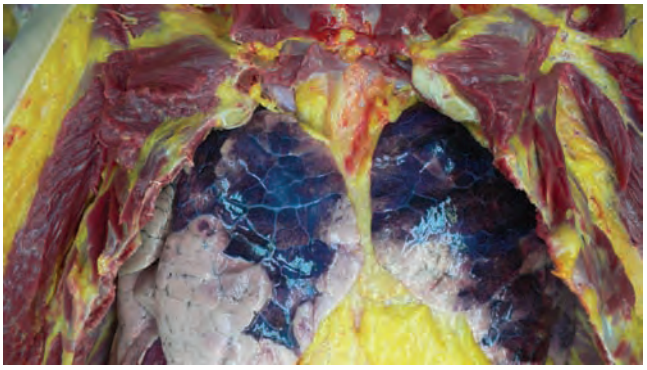

Note the patchy red to purple discoloration indicating aspiration of blood in an individual with a gunshot wound to the head, lacerations extending to the oral pharynx and residual respiratory effort before death.

This is another pedestrian who was struck by a car and survived approximately 2 days in the hospital. Note the tibial fracture with adjacent abrasions and broken skin. There is also blister formation confined to this region associated with sepsis following infection associated with this trauma.

Sharp-Force Injuries

7

CHARLES A. CATANESE

Introduction

Sharp-force injuries are defined as injuries produced by an instrument with a thin edge or point. Examples of these instruments include a knife, razor, box cutter, scalpel, sharp-edged piece of metal, broken glass bottle, broken glass window, scissor, ice pick, fork, propeller, screw driver, saw blade, axe, machete, arrow, nail, pick-axe, spiked fence post, meat cleaver, etc.

A *stab wound* is typically made by a knife blade and is defined as having a greater depth of penetration than surface dimension. An *incised wound* is a slicing-type injury where the surface dimension is greater than the depth of penetration.

Accurate, concise, and organized wound documentation is important, as with all other injuries. Each injury should have a documented location on the body, including a description of adjacent abrasions or contusions, wound dimensions, depth of penetration, and direction of penetration into the body. All injured structures should be documented, including the amount of hemorrhage both in the wound track and within body cavities. In cases where there are multiple injuries, it is acceptable to group them with ranges. It is good practice to take overall photographs of the body before and after cleaning; as well as close-up photographs of each wound.

Important aspects concerning interpretation of injury involve pattern recognition. Familiarization with this will allow opinion formulation concerning correlation of a particular instrument to a particular wound. One example involved the arrest of several suspects with different concealed weapons. The police may approach you to render an opinion about what type of weapon produced injuries so they can focus their early investigation. In one actual case, each suspect had a different instrument in his or her pocket, including a flat-edged pocket knife, a box cutter, a screw driver, and a slightly bent serrated table knife. The injuries to the decedent's body consisted of slit-like perforations with multiple adjacent parallel linear abrasions. This pattern injury is consistent with a serrated knife. Many of the images in this chapter are designed to help with pattern recognition.

Location and Direction of Injury

This should be given with reference to a particular body position, usually standard anatomic planes. Each wound should be documented by location on the body's surface, and measured from vertical and horizontal planes of reference. An example of this would be from below the top of the head or above the feet, and to the right or left of the midline. Standard anatomic planes are demonstrated with the body in an upright position with the head tilted slightly upward, the legs together, the arms at the sides, and palms facing forward. The head is superior and the feet inferior, medial is toward the midline and lateral away from the midline. The anterior or front of the body includes the face, chest, and the palms. The posterior part of the body includes the back, buttocks, etc. The direction of the wound into or through the body should be given with reference to three planes when possible, front–back, right–left, and up–down. This is important because it allows one to correlate the injuries to possible assault descriptions and help discredit or substantiate statements.

Wound Dimension

This should be documented separately for each sharp-force injury, unless there are many that can be grouped together and described in ranges. Example: There are twenty 1-inch to 2-inch, by up to 1/4-inch, stab wounds within a 5-inch × 7-inch region at the middle aspect of the right chest, which is centered 13 inches below the top of the head and 4 inches to the right of the midline.

It is good practice to document the injury as it exists on the body and then again when in a relaxed state. The important aspect is to document the wound dimension in ranges that most closely reflect the actual dimensions of the knife blade or instrument. A stab wound can be put into a relaxed state by pressing the surrounding skin toward the wound and releasing the surrounding tension, or by placing tape over the perforation site to approximate the margins. It is also acceptable to cut a square around the surrounding skin to release the tension. The skin and underlying tissues are elastic, with

different degrees of tension. This can make the surface dimension of the wound length and width slightly variable depending on location and orientation of the body. Cut marks through bone and cartilage may reflect accurate weapon dimension with tool marks and should be retained in formaldehyde.

This same concept applies for the depth of penetration, as well. Dimensions should be given in a range to account for changes in body position when examined on the autopsy table compared with the body position when the injuries were inflicted. Variables that may change this parameter while the assault is taking place include deep breaths or exhales, flexion, extension, rotation, the force used to inflict the injury, the location on the body including underlying bone or soft tissue, etc.

Adjacent Abrasions and Contusions

These may indicate body contact from the knife handle, lower part of the knife blade, or the knife hilt. This is important information in formulating an opinion as to how much force was needed to produce the injury. If the knife blade penetrates a bone and there are hilt marks adjacent to the perforation site, one can extrapolate that the knife must have been stuck into the body with great force.

Injured Organs or Structures

This information helps to allow interpretation of how functional one might be after an assault. The number and extent of internal injuries give insight into the nature of the assault and perpetrator.

For instance, "I stabbed him in the neck and he continued to chase me so I stabbed him again in the chest and abdomen." If the neck wound injured the spinal cord, you know the statement is false because the victim would no longer have voluntary movement to part of the body, as with a tendon or peripheral nerve's being cut. Also, depending on the structures damaged, the rate of blood loss may be quite variable. This would help define how fast the individual would lose blood, and what the individual might be capable of doing after the injury, and for how long. Questions like this may often come up in trial. A transected aorta would incapacitate someone more rapidly than a transected brachial (arm) artery, which would be more rapid than a transected cephalic (superficial) vein. Someone with a stab wound to the heart will often lose consciousness within minutes but still be capable of running away, defending themselves, or continuing an assault, particularly if the heart does not go into a lethal arrhythmia. It is possible for someone to get stabbed in the heart, run several blocks and shoot several people before dying.

Stab wound consisting of a flat slit-like perforation with minimal surrounding skin tension.

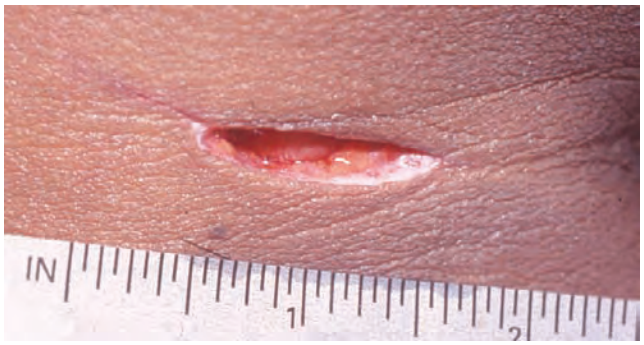

Stab wound with moderate surrounding skin tension. Note the separation of the wound margins.

Gaping stab wound consisting of an oval perforation caused by significant surrounding skin tension.

A non-serrated knife blade similar to the one used to inflict these wounds. It is important to take the surrounding skin tension into account when estimating the size and type of weapon used.

Incised wounds produced by a broken bottle.

Incised wound to the face produced by a knife.

Incised wounds produced by a box cutter.

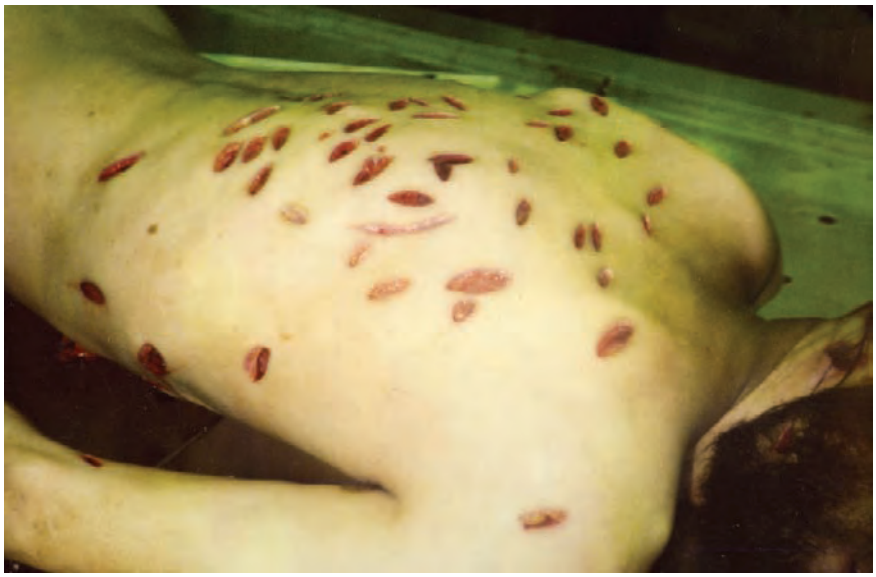

These are multiple stab wounds to the back. Note the varying dimensions and gaping nature due to the varying degrees of skin tension from underlying tissue planes in different locations.

Several of these wounds were chosen based on estimation that they were more reflective of the actual knife blade size. These wounds were then cut into a relaxed state and re-measured to help estimate the actual size of the knife blade.

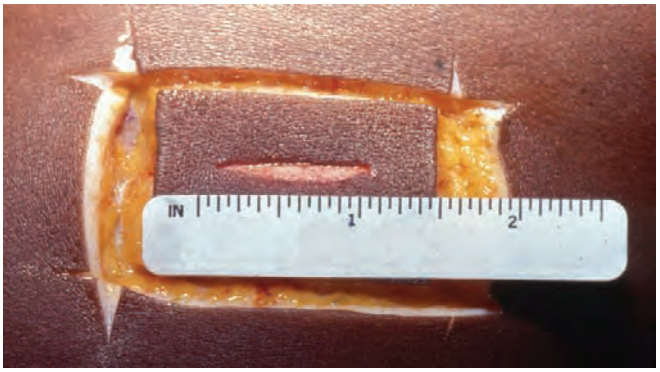

This is a close-up of one of the wounds that were cut into a relaxed state and measured. Note the sharp end at the left aspect and the blunt end at the right aspect.

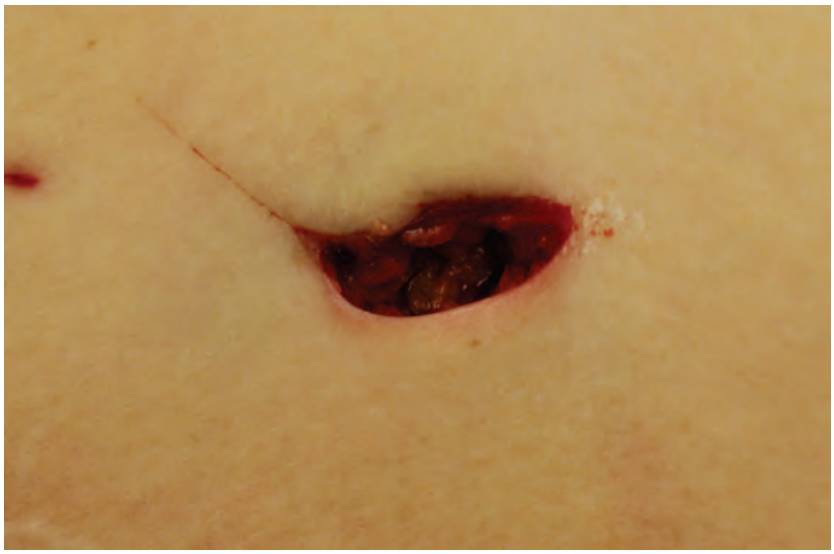

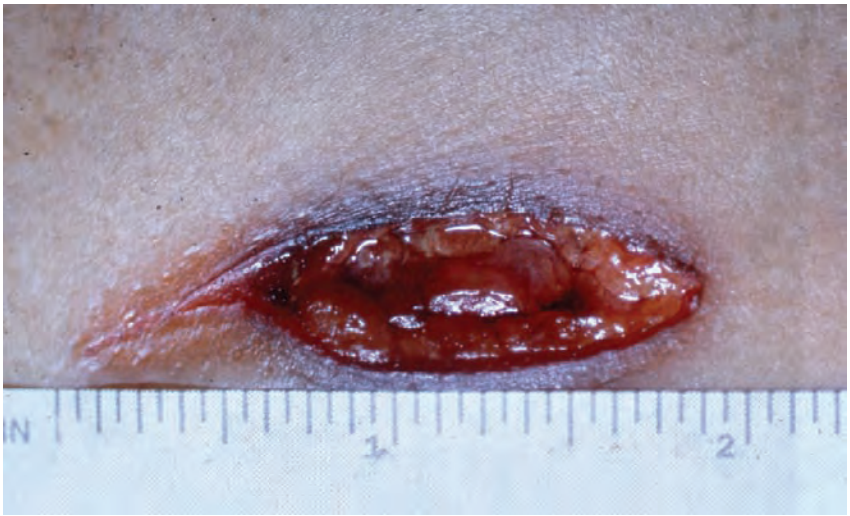

Stab wounds to body with slight drying of the margins and a "linear extension" from one end due to the sharp part of the knife blade being dragged through the wound and across the body surface, changing the dimension of the injury.

These two images demonstrate a "swallowtail"-type injury where the knife was inserted into the body in one direction, turned and then removed in another direction, producing two angulated cuts. This type of injury may be produced by the movement of the perpetrator, victim, or both.

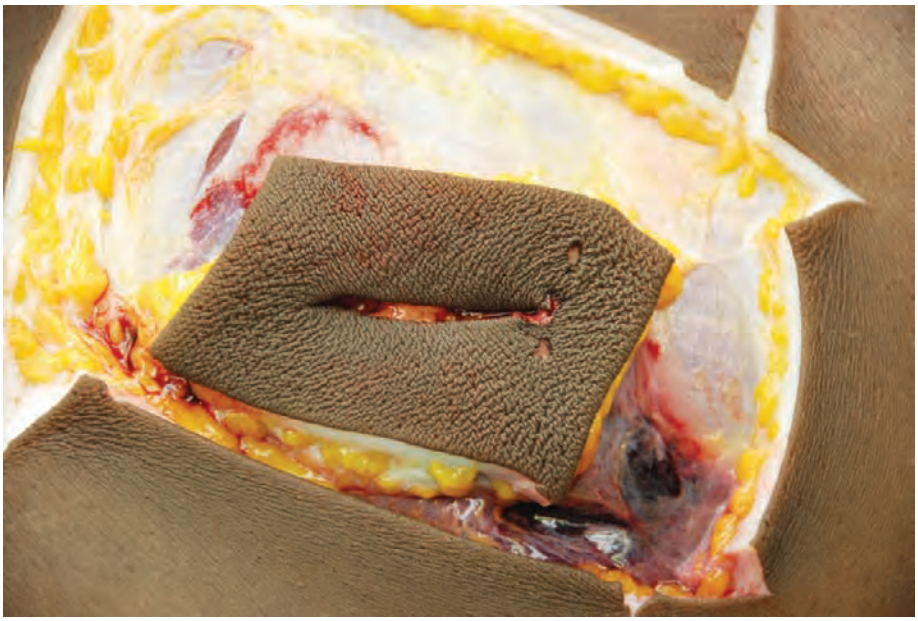

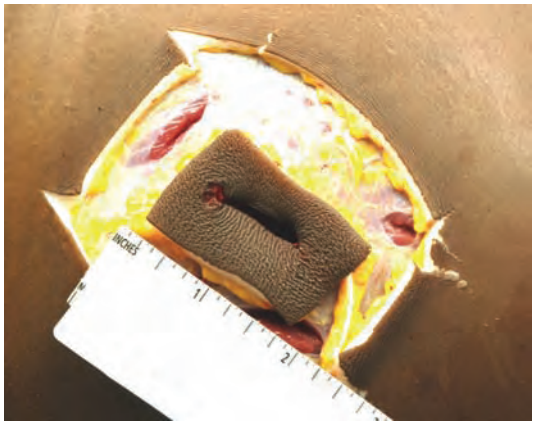

These stab wounds demonstrate adjacent abrasions from the hilt of the knife contacting the skin with great force.

Multiple self-inflected linear abrasions and incised wounds to the arms and wrists demonstrating hesitation marks in individuals who committed suicide with other lethal injuries.

Multiple linear incised wounds in an individual who lived for several hours. He died of a heart attack associated with the stress and blood loss from the self-inflicted injuries. The manner of death in this case was classified as suicide. The cause of death was listed as both multiple incised wounds and heart disease.

Multiple self-inflicted sharp-force injuries including stab and incised wounds. Note the tightly clustered orientation of these injuries, which is common in suicides.

Multiple self-inflicted stab wounds. This individual was found with her hands surrounding the knife handle in a moderate state of rigor mortis.

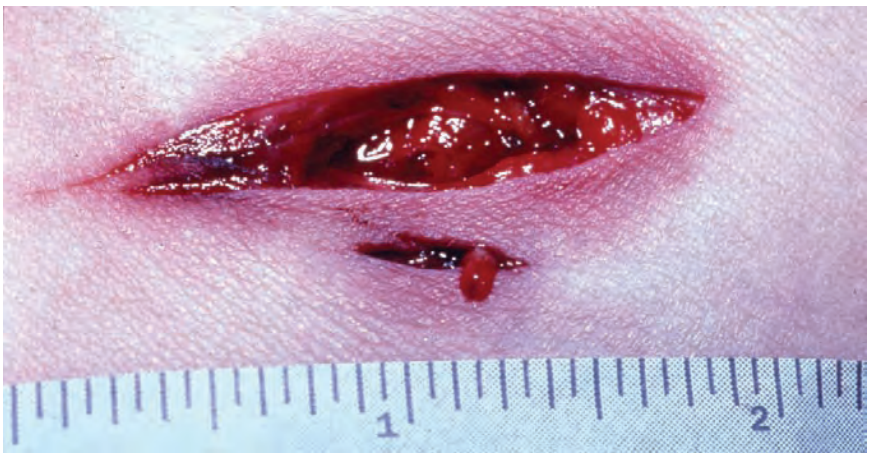

This individual was found with multiple clustered linear abrasions, incised wounds, and stab wounds to the wrists, arms, and ankles. She first attempted suicide by ingesting large quantities of acetaminophen, then inflicted the above sharp-force injuries. Note the erythema around the wounds, indicating vital reaction. There was blood noted all over the apartment. The individual later jumped out a window and died from blunt-impact injuries. The wounds shown are hesitation wounds.

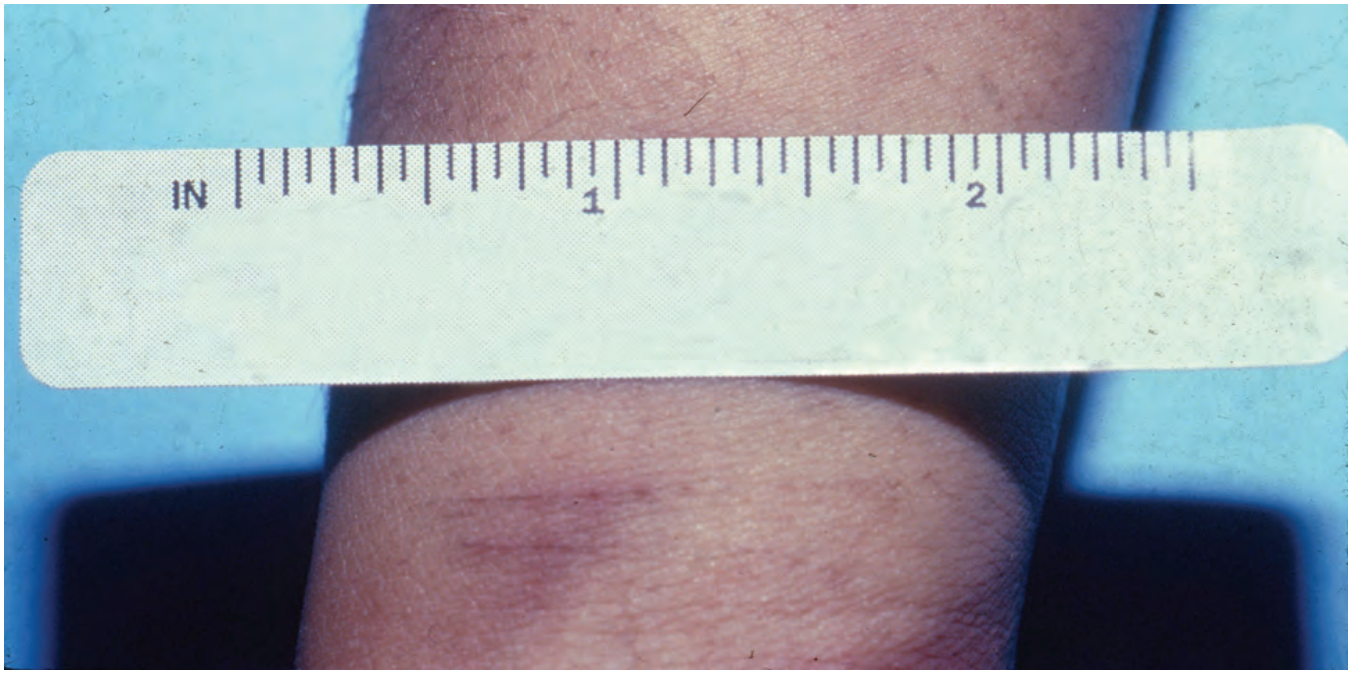

Superficial linear abrasions caused by dragging a knife blade across the wrist and hand prior to committing suicide by hanging.

Multiple clustered self-inflicted incised and stab wounds located on each side of the neck and chest in an individual who committed suicide. Most of these stab wounds were very superficial in depth of penetration and can be described as hesitation marks. Three of these stab wounds to the chest entered the thoracic cavity and penetrated the heart and lung, causing death. Note the slight degree of decomposition with wound margin drying.

Multiple clustered homicidal stab wounds in individuals who were also bound during a sexual assault. Note the location of the multiple stab wounds to the middle aspect of the individual's back, making it virtually impossible for them to have been self-inflicted.

These are multiple examples of defensive-type wounds from individuals who were attempting to ward off an assault from a knife-wielding attacker. Note the multiple linear abrasions, incised wounds, and stab wounds to the hands. Several of these injuries show cut marks between the thumb and fingers, due to the individual's attempting to grab the knife blade.

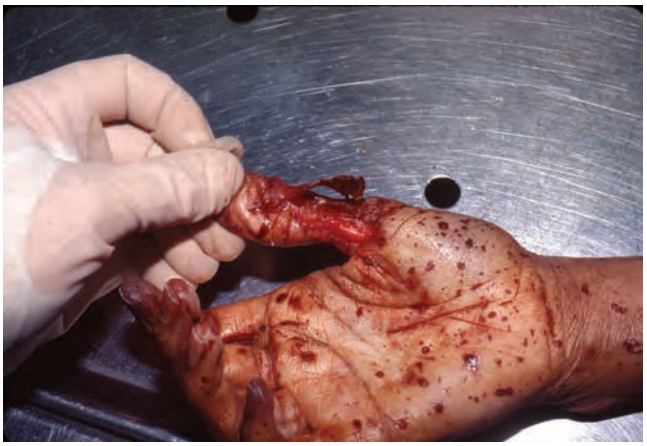

Other examples of wounds consistent with being inflicted while an individual is fending off an attack.

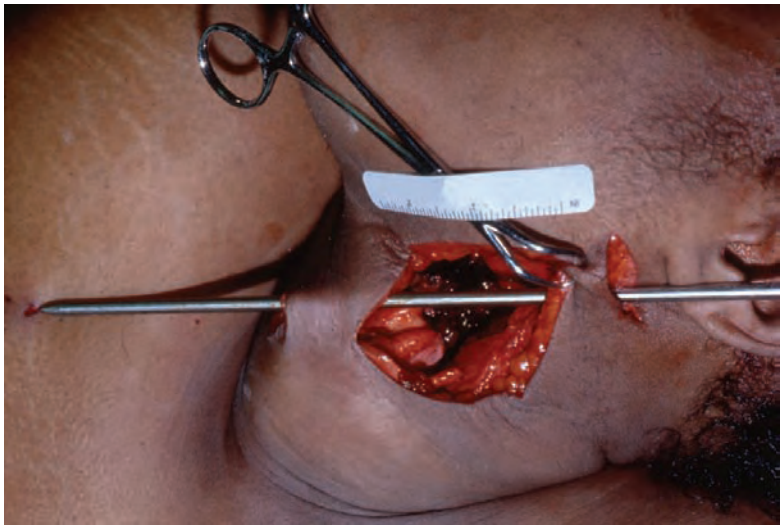

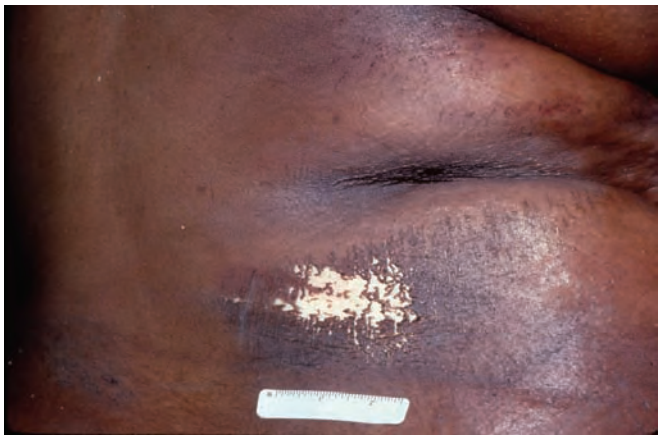

This individual was stabbed in the neck and then thrown from a window. The perpetrator described the injuries as being self-inflicted by the decedent who cut herself and then jumped out a window. The window ledge and outside wall were smeared with blood, which was easily visible from below. During autopsy, approximation of the margins revealed a stab wound just below the left ear from a knife blade that exited through her neck and stuck in her right shoulder. There was also an anemic abrasion to her back and an anemic laceration to the scalp, indicating the decedent had lost much of her blood volume and did not have a beating heart with blood pressure enough to produce hemorrhage from the injuries sustained when striking the ground.

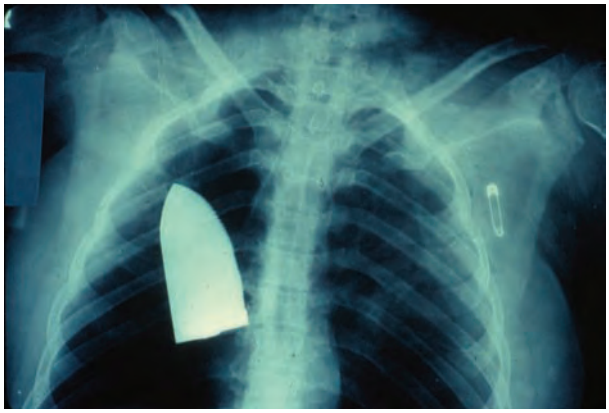

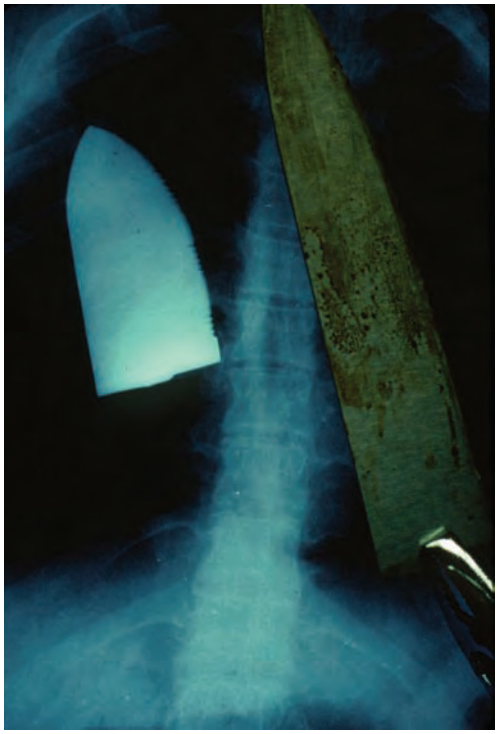

Note the two stab wounds to the left back and the broken knife blade extending from one of the wound tracks. Note the close-up view of the back, demonstrating the knife within the wound track and then with the knife removed. Note the change in the wound dimension after the knife was removed. This demonstrates the elastic nature of skin and the possible challenges of correlating a weapon to a particular wound dimension. Also note the x-ray, which demonstrates the knife blade within the body. Note the dimension of the actual knife blade compared with the dimension of the blade image on the x-ray. This is due to the angle in which the x-rays contacted the knife blade in the thoracic cavity. Weapon.

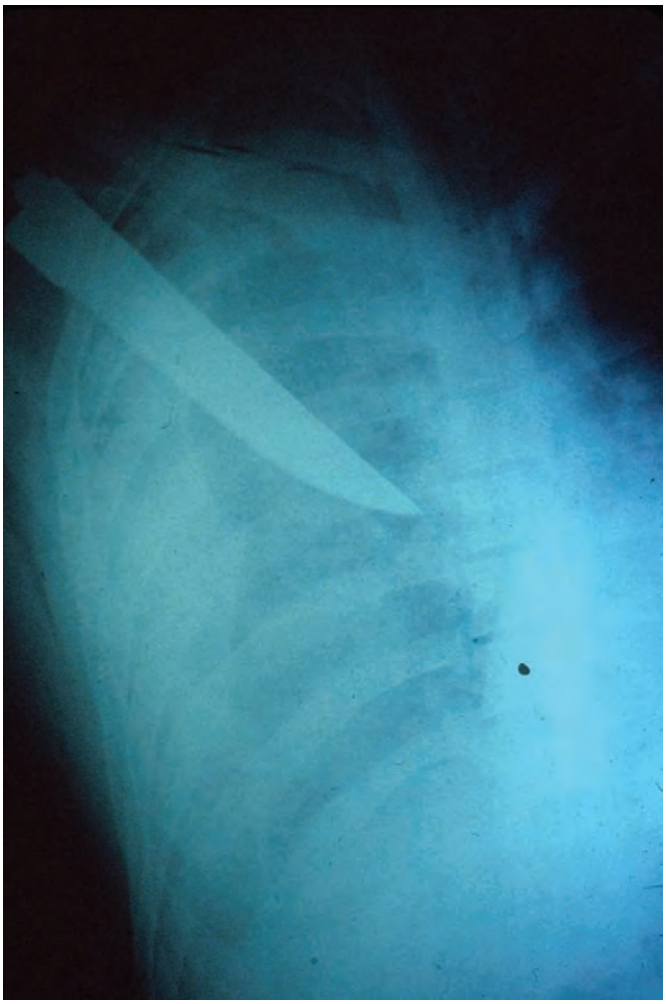

Broken knife blade left in body following a homicidal assault. This knife was stuck into this individual with great force and the blade got stuck in a bone. When the perpetrator attempted to remove it, the handle broke off and the knife blade was left behind.

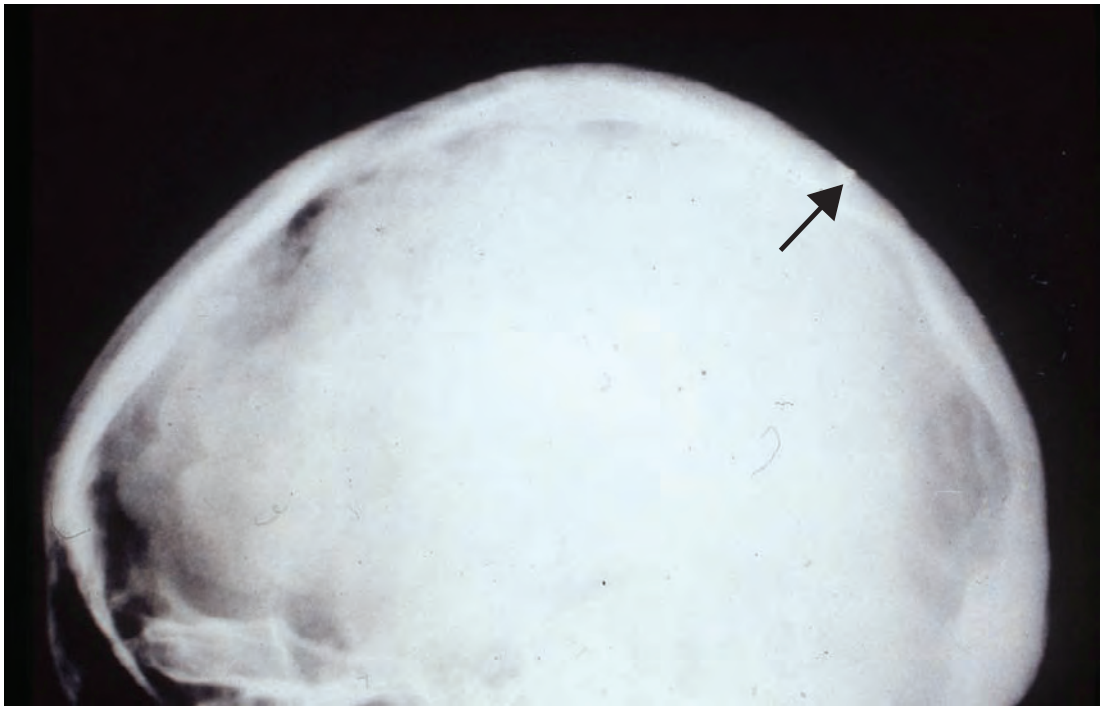

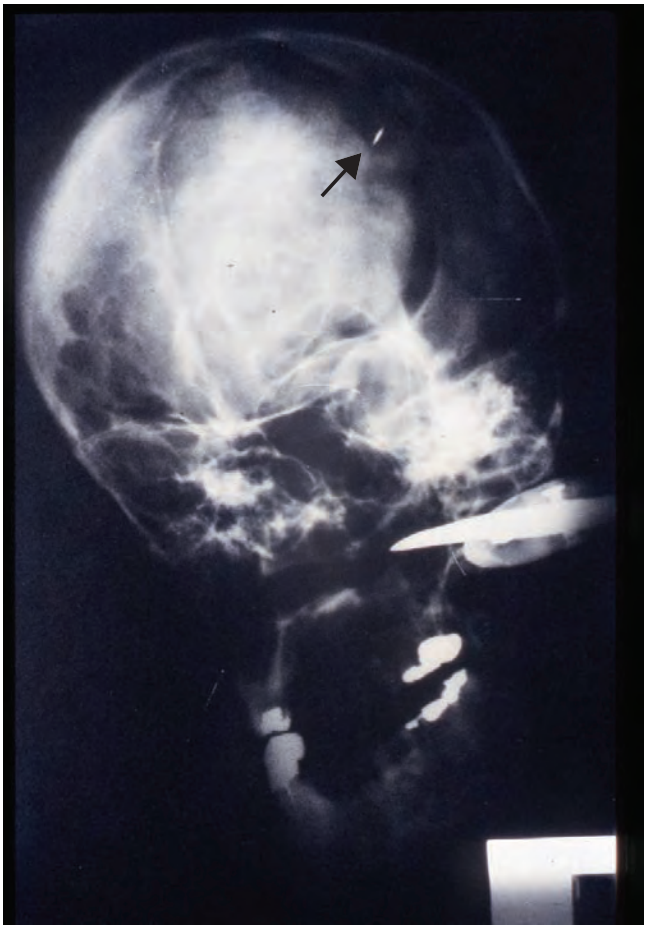

This individual was stabbed more than 75 times by a scorned lover. The tip of the knife blade broke off in the skull during this violent assault. Note the small radio-dense fragment depicted in these skull x-rays. This reinforces the notion that it is good practice to x-ray all sharp-force homicidal violence cases. A fragmented portion of knife blade can later be matched to a suspect's knife.

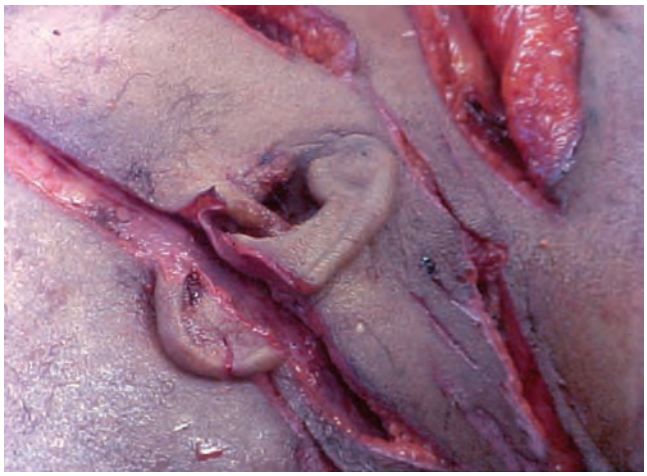

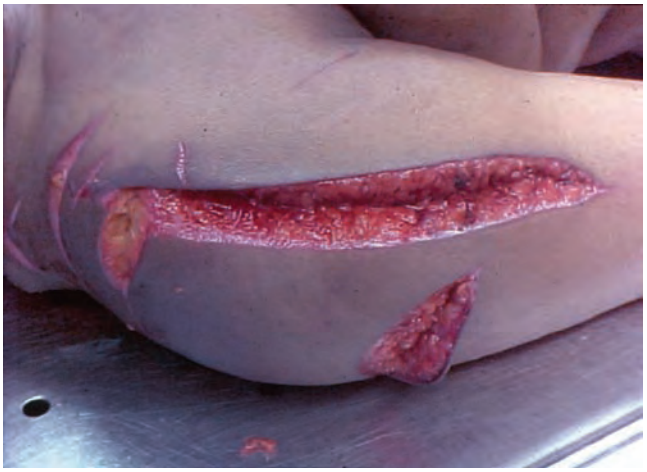

These individuals were attacked multiple times with machetes. Note the long gaping hack marks, several of which cut into and through bone. Most often, this degree of violent rage indicates the victim and the perpetrator had a relationship.

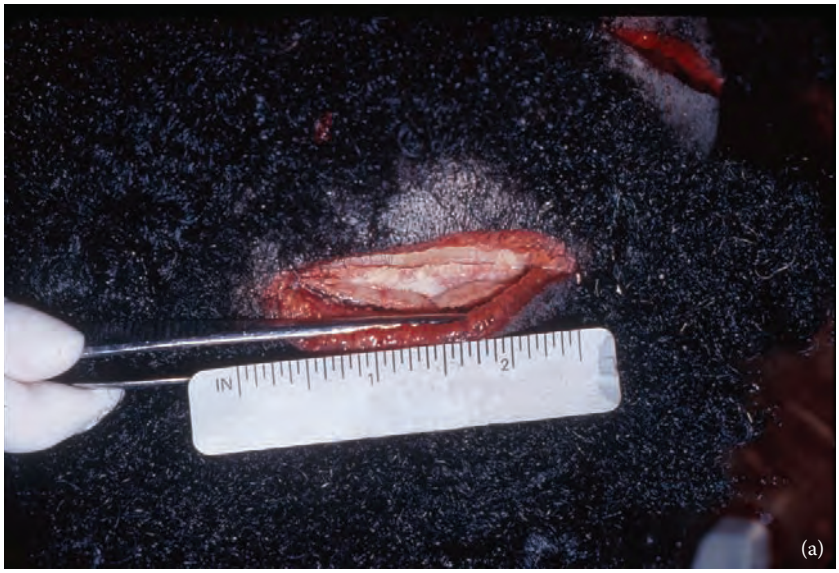

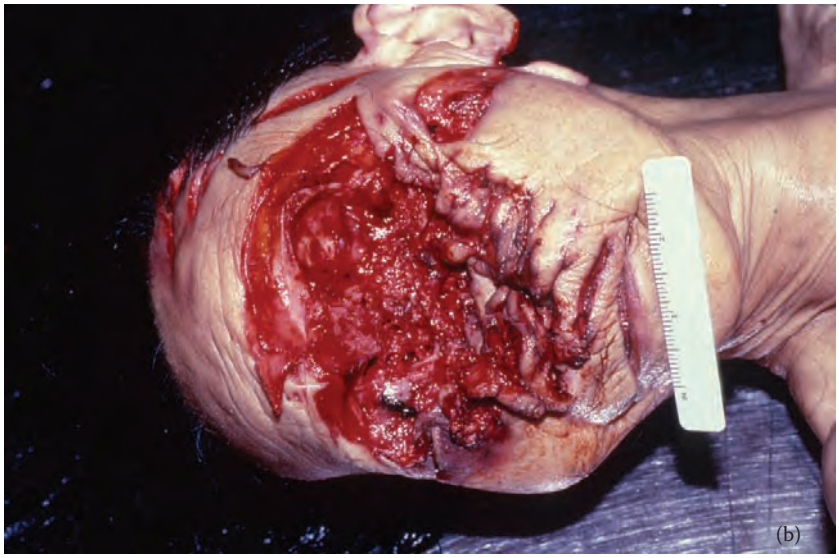

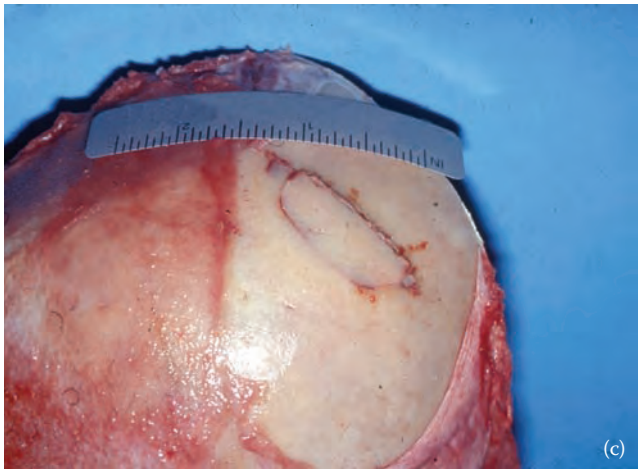

Chopping-type sharp-force injuries produced by an axe. Image (b) shows an individual whose eyes were chopped out by another person with a long history of psychosis. These types of assaults are not common. He reportedly did this because she kept staring at him.

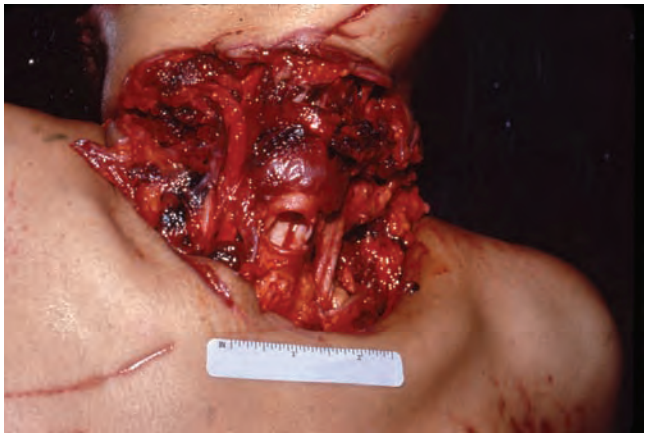

This individual was chased up several flights of stairs in an apartment complex and was cut multiple times with a box cutter similar to the one shown here. Note the blood splatter pattern on the steps. Also note that the individual was almost decapitated. This was reportedly over a dispute concerning gang-related drug dealing territory.

This individual had multiple stab wounds with adjacent abrasions consisting of multiple superficial linear parallel lines. This is a typical pattern found following a serrated knife attack. The absence of these marks does not preclude the use of a serrated knife. If a serrated knife is stuck directly into a body and not dragged across the surface, there will be no parallel linear abrasions.

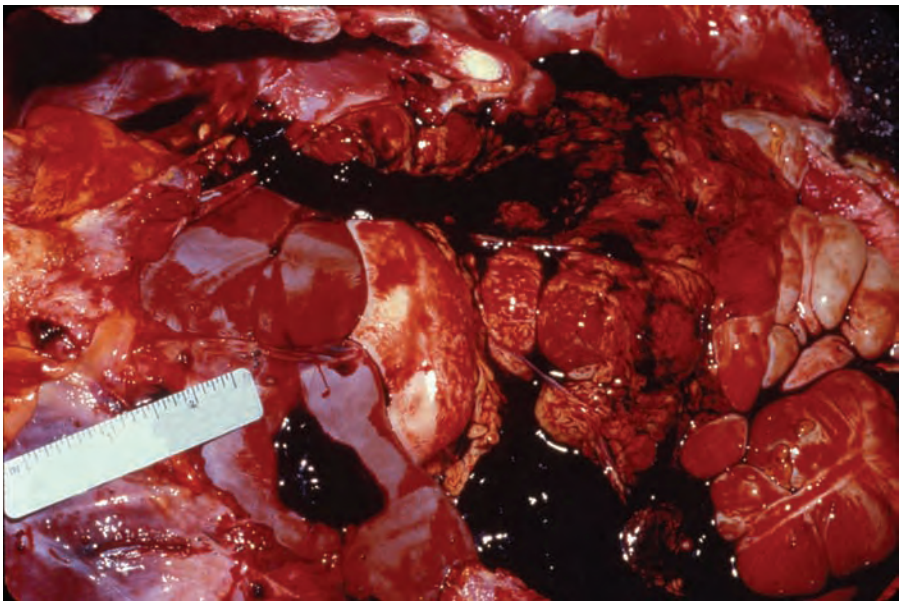

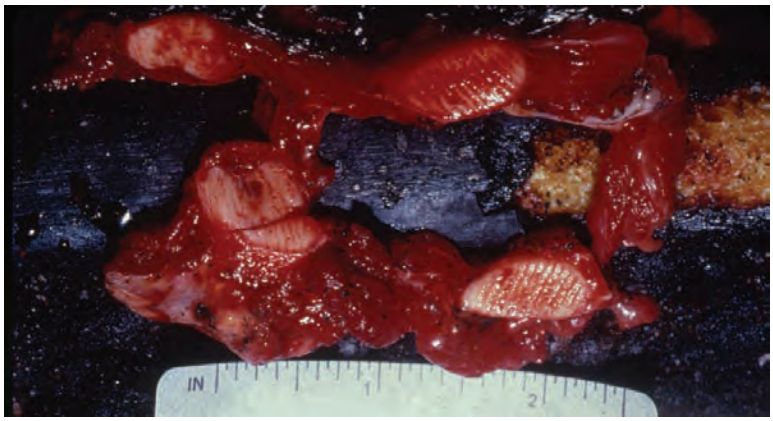

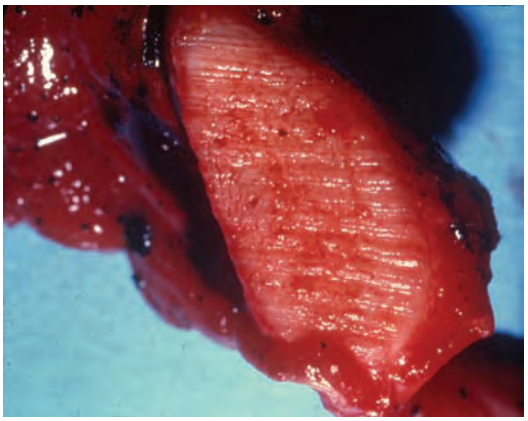

This individual was stabbed multiple times with a Ginsu-type knife. His dwelling was then set on fire and he sustained extensive thermal burns with charring of the skin. Note the costal cartilages with tool markings forming an imprint of the knife blade. Image on lower right demonstrates a knife similar to the one used in the attack.

Note this irregular incised wound with curved sharp margins produced by a knife similar to the one depicted in the previous figure.

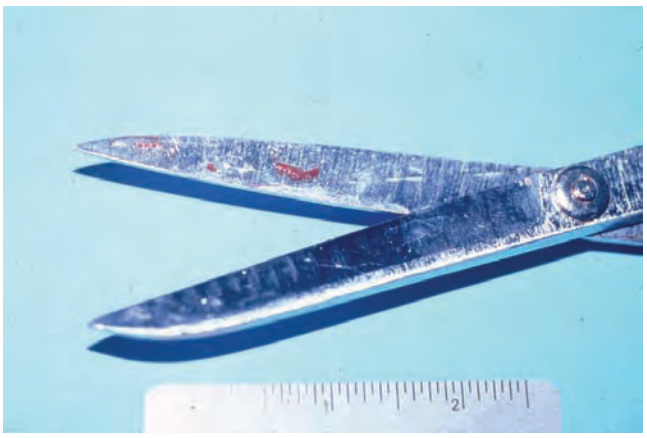

This child was stabbed multiple times with a pair of scissors. Note that the entrance defects to the anterior trunk have a slightly widened angulated character consistent with being inflicted by a pair of scissors. The posterior trunk demonstrates the pointed end of the scissor perforating the entire body.

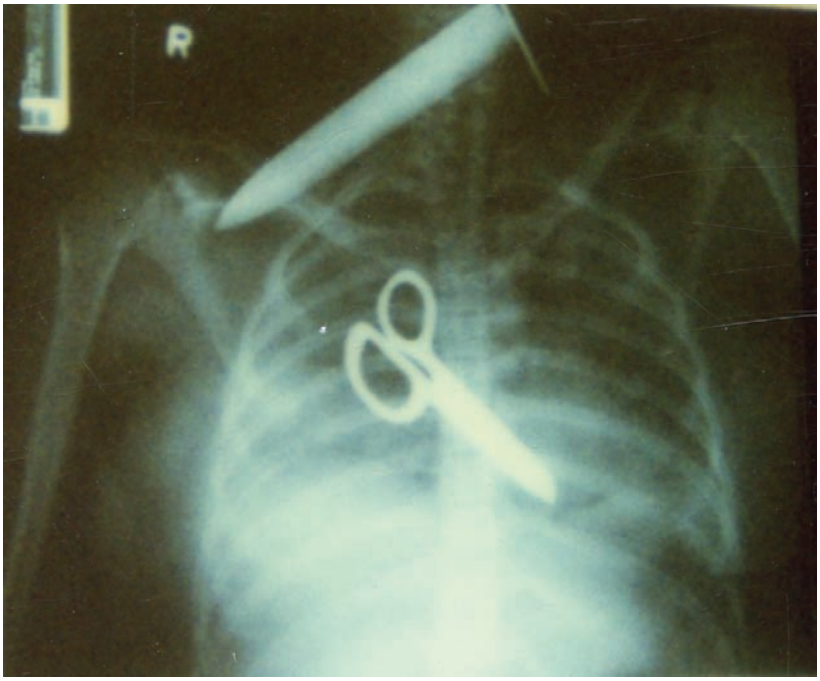

This image shows an x-ray of an adult who was stabbed with a pair of scissors that were left in the wound.

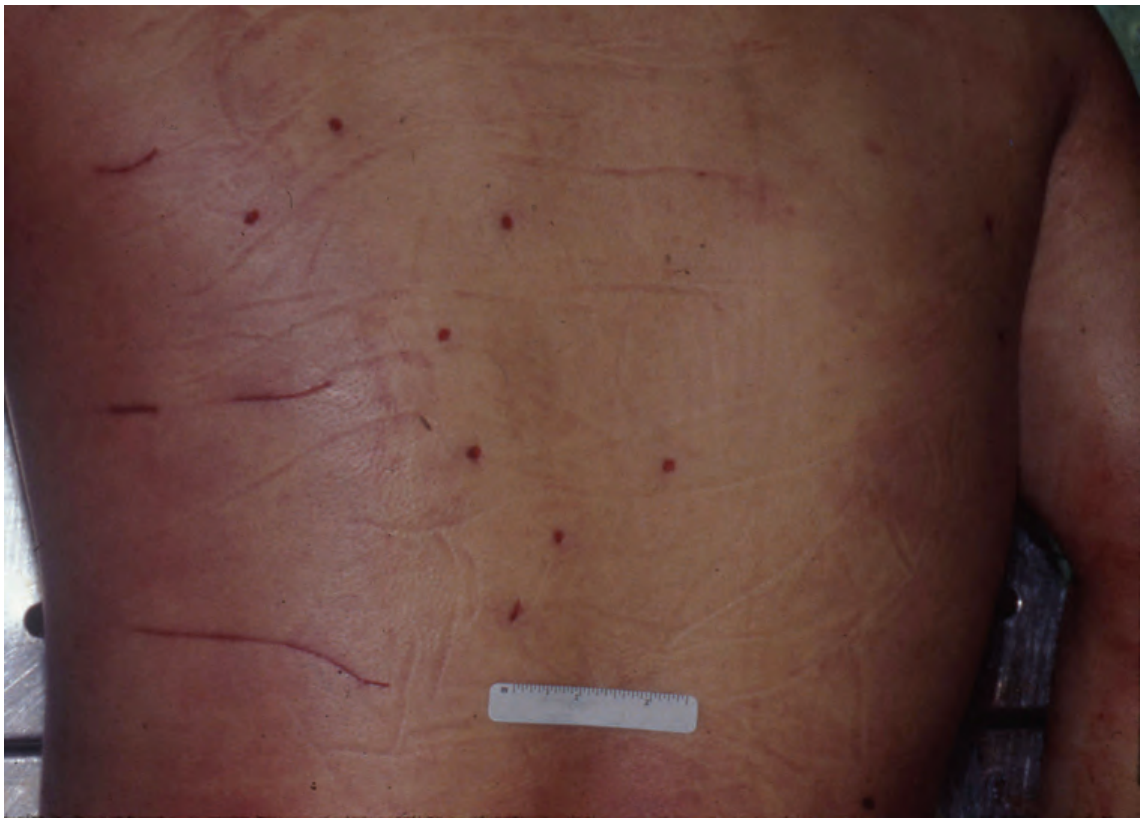

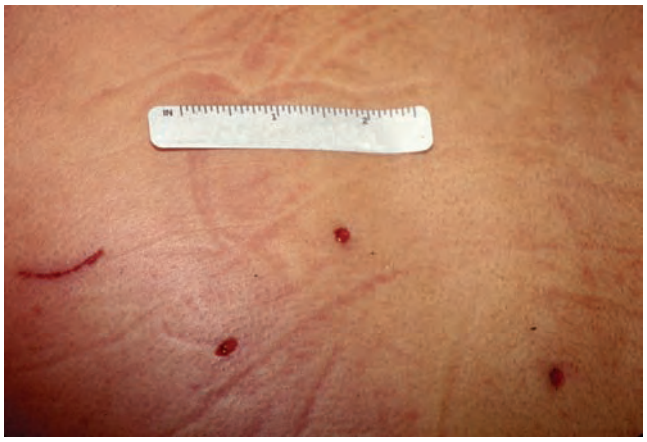

This individual was stabbed more than 75 times with an ice pick. Note the linear abrasions extending in different angles across the back as the tip of the ice pick was dragged across the skin's surface. The small dimension of these perforation wounds would take more time to lose significant blood compared with a typical knife blade. Generally speaking, the larger the wound, the faster the blood flow in a extravascular region. There was evidence of a violent struggle over a long period of time in this case.

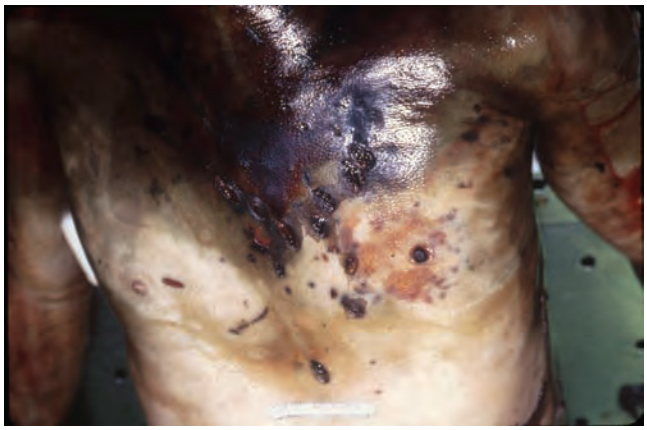

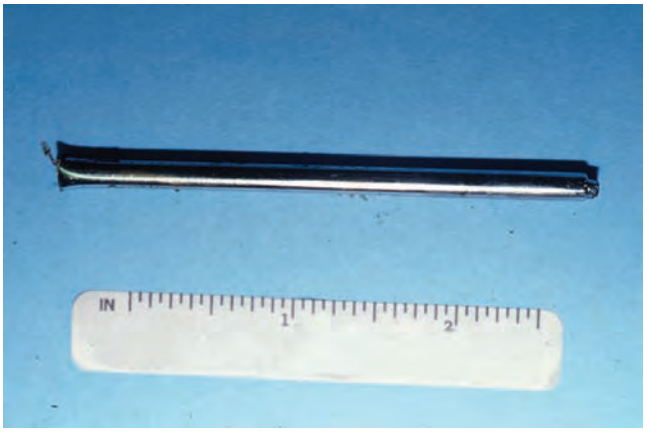

This individual was found within a crack house, in a mild to moderate state of decomposition. He sustained multiple stab wounds and puncture-type wounds to his body that were produced by broken pieces of antenna used as crack pipes. One of these antennas was jammed down his throat. He also sustained multiple blunt-impact injuries with rib fractures and other broken bones from being stomped. Note the drying and slight distortion of the wound margins due to decomposition.

This individual was shot with a hunting arrow. The entrance wound is to the medial aspect of the right chest and the exit wound was to the right back. Note the comparison of the injury to the arrow head.

This individual was struck with a bottle that produced a minimal abrasion with a greater underlying contusion and a central laceration. Note the slightly irregular margins at the laceration site due to this blunt impact. The surrounding scalp hair was shaved to demonstrate the injury in greater detail. The presence of the scalp hair at the time this injury was inflicted also served to cushion the impact site.

This is the same individual, who was then cut multiple times with the sharp broken glass end of the bottle, producing a partial transection to the carotid artery and exsanguination.

The individual had multiple curvilinear parallel abrasions and other superficial abrasions caused by a bottle top and broken glass scraping across the body surface during the assault. This initial argument began over which rival baseball team was better.

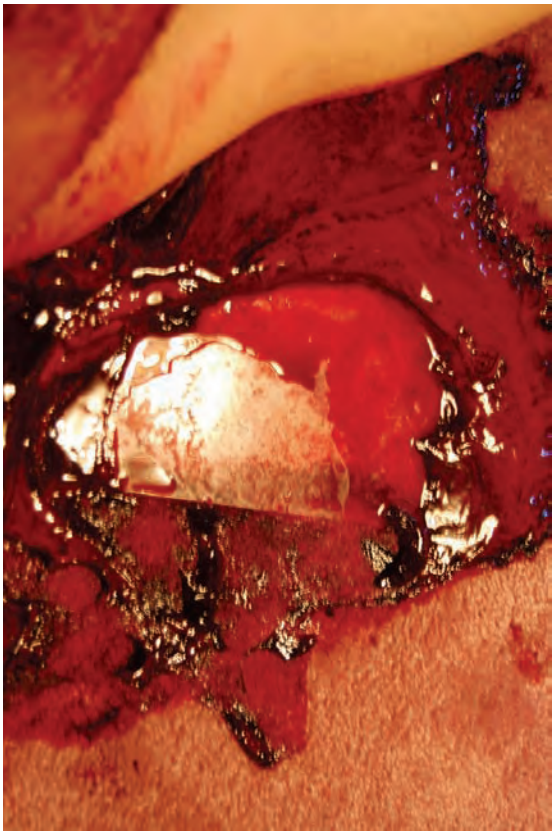

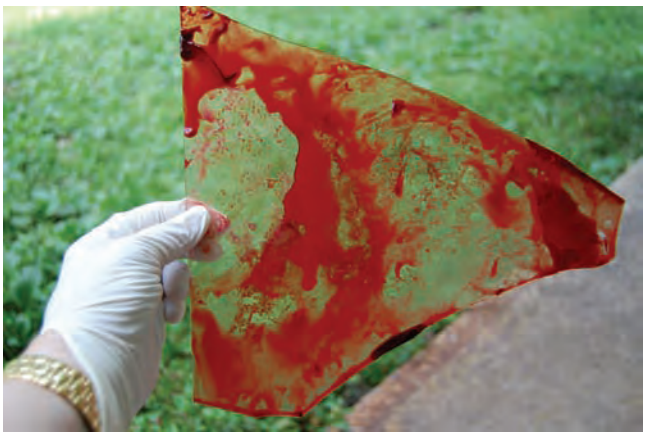

Suicide by multiple sharp-force injuries with broken glass taken from an empty picture frame.

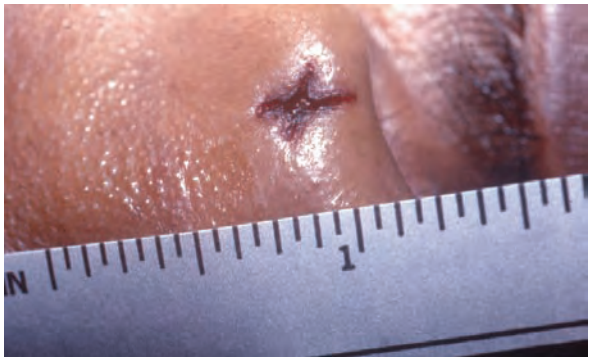

These are multiple examples of puncture wounds caused by a Phillips-head screwdriver, demonstrating the range of presentation of this type of injury.

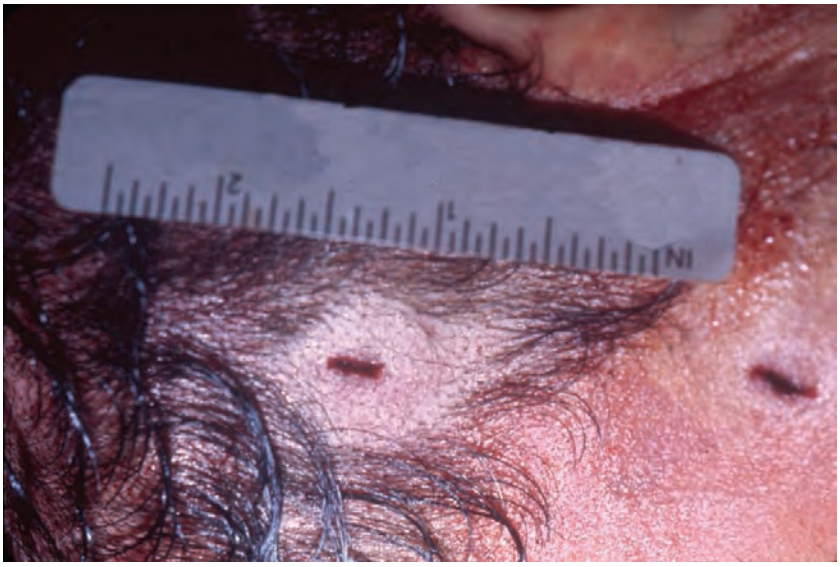

These are multiple examples of flat-head screwdriver puncture wounds.

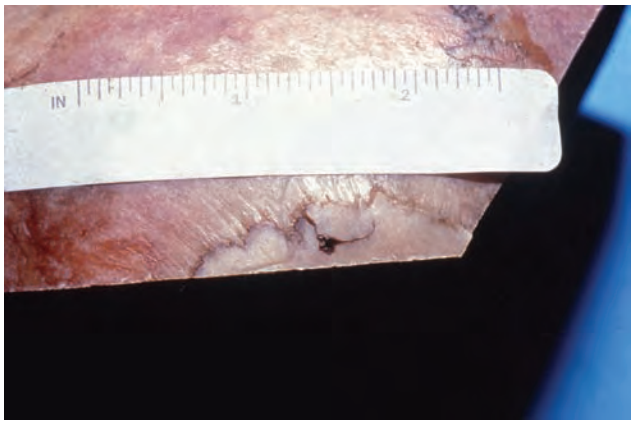

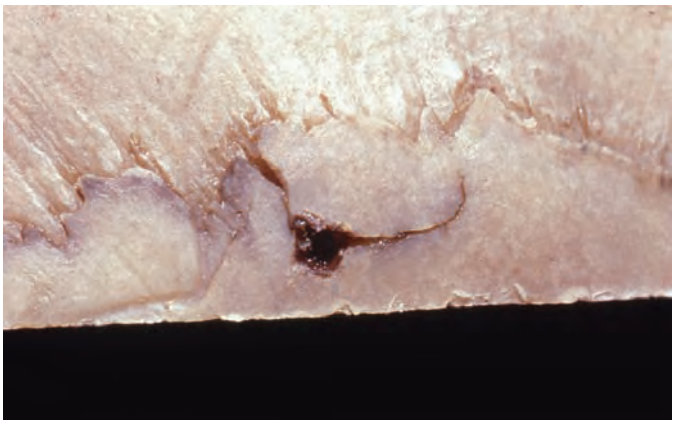

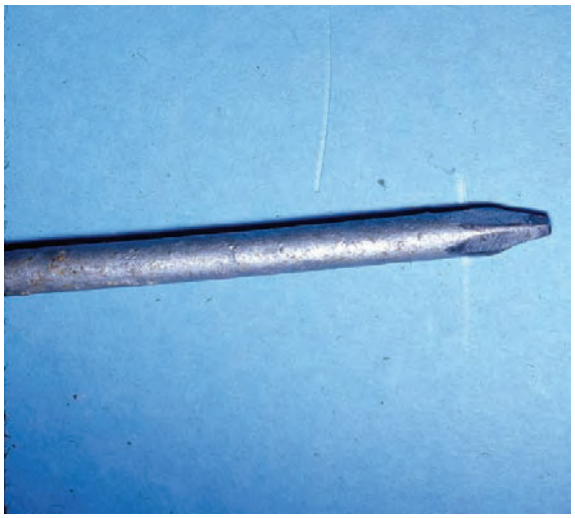

This approximately 30-year-old woman was found naked from the waist down in a construction yard. She had moderate decompositional changes with insect and animal feeding. She had multiple defects to her scalp and face with numerous maggots tracking through the underlying soft tissue. The soft tissues in this region were darkly discolored throughout. It becomes more difficult to interpret soft-tissue injuries as decomposition progresses. Careful examination of the underlying bone in such cases may often yield valuable information. Note the approximately 1/4″ linear skull fracture with a 1/16″ roughly square indentation at the superior right temporal bone. Further examination of the scene revealed a bloodstained board with a nail at the opposite side of the construction site.

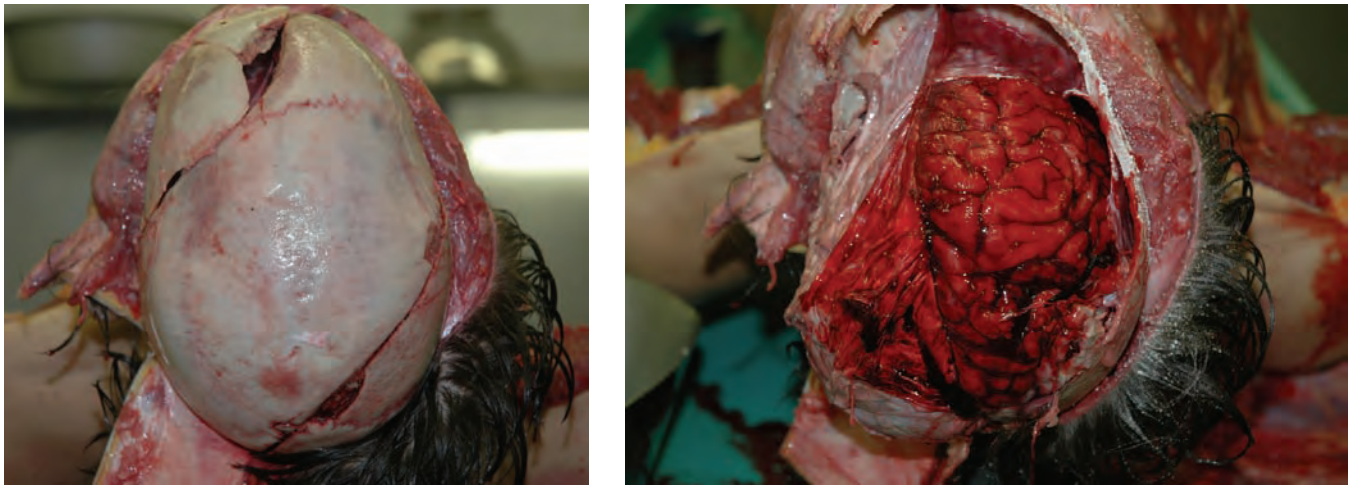

Antemortem boat propeller injuries. Note the presence of wound track hemorrhage and the parallel slicing-type injuries.

Injuries produced by a boat propeller.

Injuries produced by a boat propeller. Note the hemorrhage at the autopsy incision (lower right), demonstrating the individual was alive when these injuries occurred. Note the range of injury presentation, including superficial abrasion to deep slicing-type wounds with extensive soft-tissue damage.

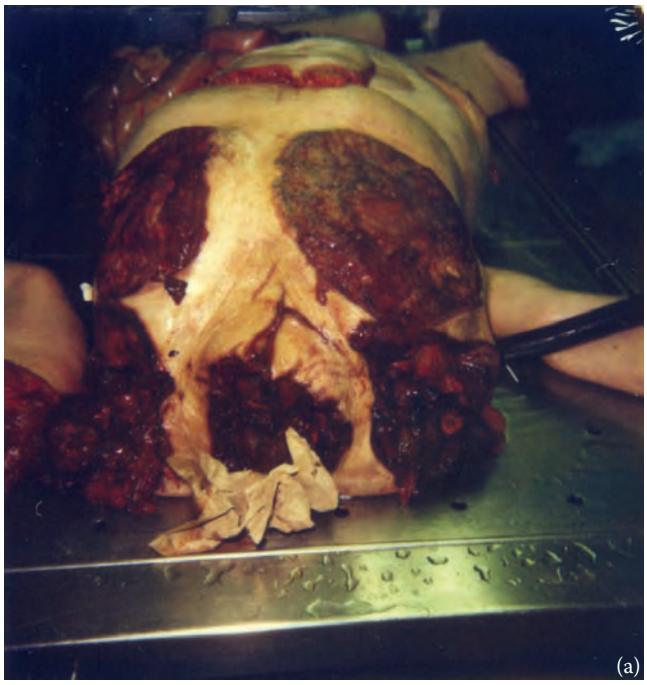
(a)

(b)

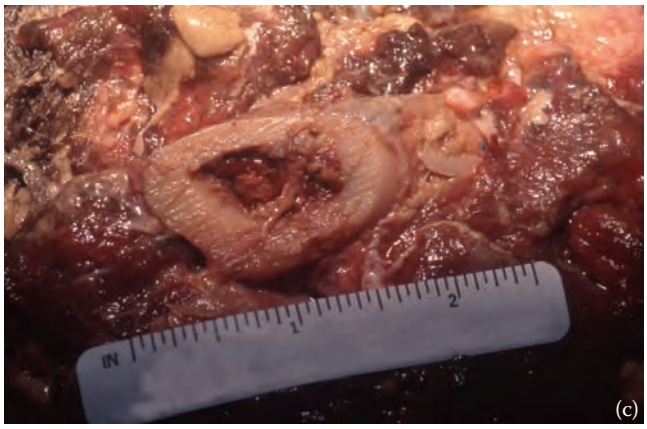
(c)

(d)

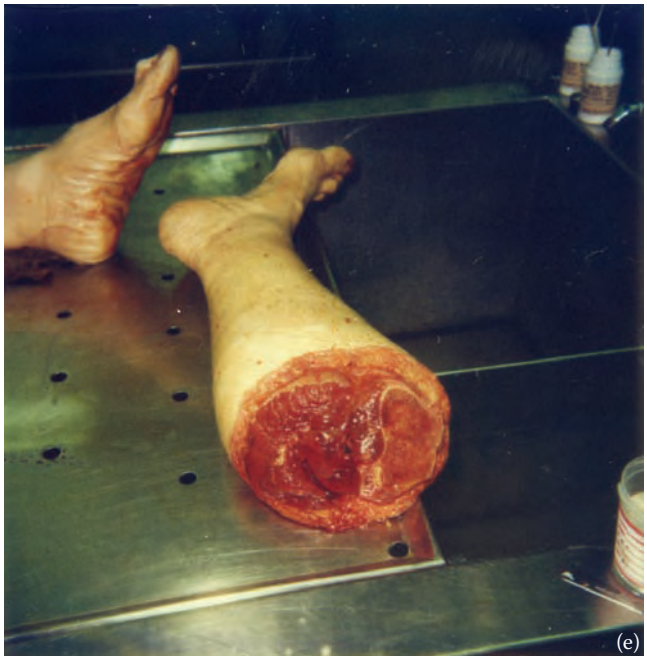
(e)

(f)

These are examples of dismembered bodies with saw cut marks through bone. It is important to save these portions of bone for possible later tool mark comparison with saw blades used during the dismemberment. Note image (a), where sexual mutilation was performed, with the breasts cut from the body. Also note image (d) of this young individual, who was cut into multiple pieces and neatly stacked on a rooftop.

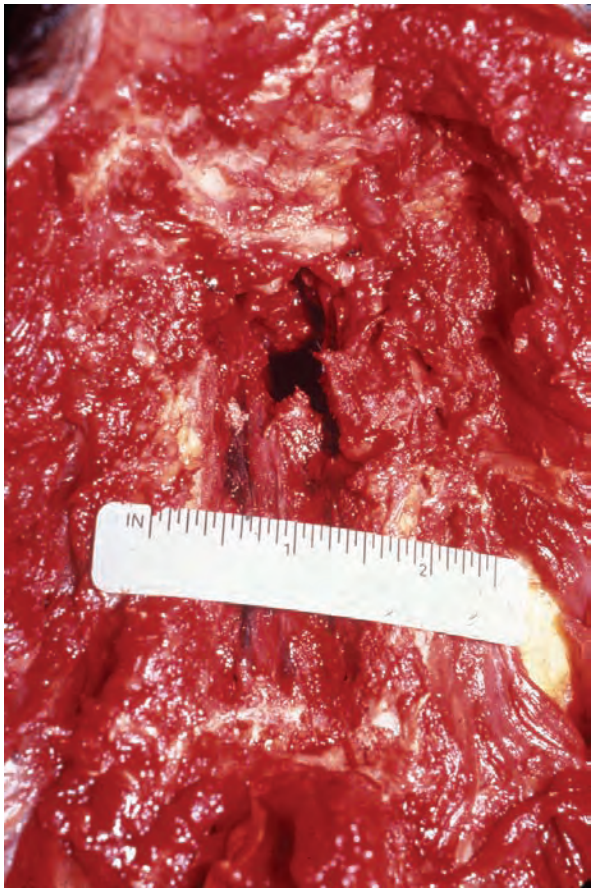

This individual was killed by being struck in the back of the neck with a pickaxe. Note the abraded margin and the perforation matching the roughly squared dimension of the pickaxe. This dispute was reportedly over not receiving back pay after complete work.

This individual committed suicide by jumping out of a window and landing on top of a spiked metal fence similar to the one shown in image (c). The hair surrounding these injuries was shaved to demonstrate the nature of the wound in greater detail. Note the pointed circular perforation site leading into a square abrasion, which was perfectly consistent with the fence spike that penetrated his chest.

This infant was found partially eaten by dogs in a building courtyard. Note the irregular nature of the wounds to the back (b–d), demonstrating a scalloped border produced by teeth and claws. Other injuries to the child's body were inconsistent with animal feeding and more likely were produced by a sharp instrument such as a knife. Note the sharp wound margins in image (e). Dogs may eat a decomposing body when they are left to starve, or they may kill living individuals, more often without eating them. It has been my experience that it is rare for a domesticated dog to eat an individual unless it is coaxed into it by exposing the hungry dog to initiate feeding activity. The dog should always be examined including the gastrointestinal contents. In one such case, neatly cut strips of soft tissue were found within the dog's stomach that were fed to the dog by the perpetrator to initiate more feeding activity.

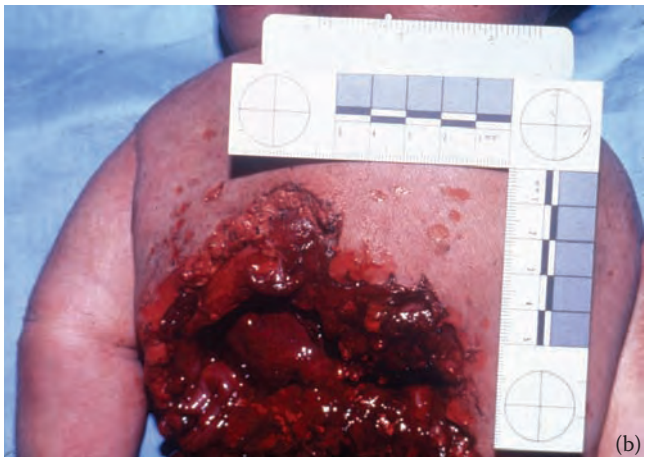

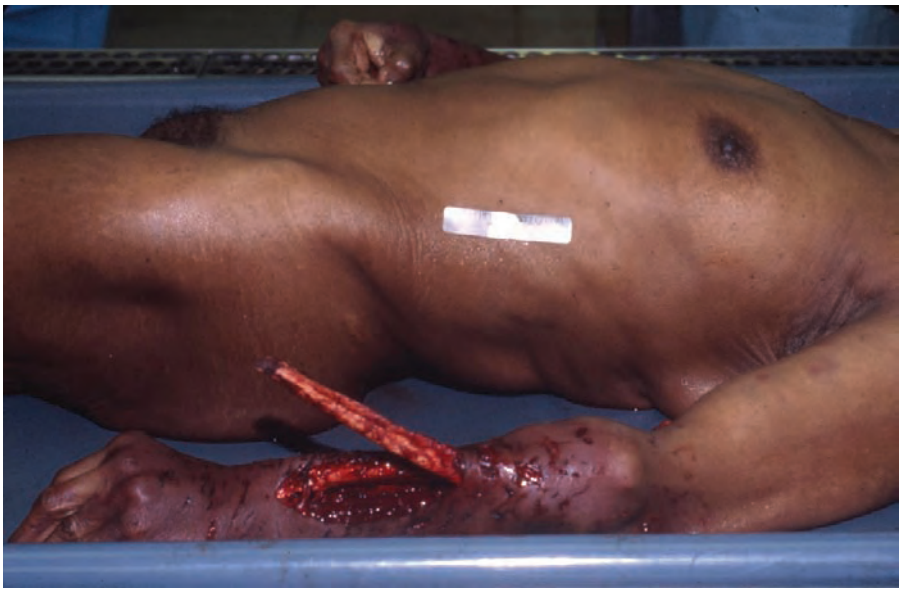

These are multiple puncture marks and lacerations with extensive adjacent hemorrhage due to an antemortem dog attack by a pit bull. This individual died as a result of blood loss.

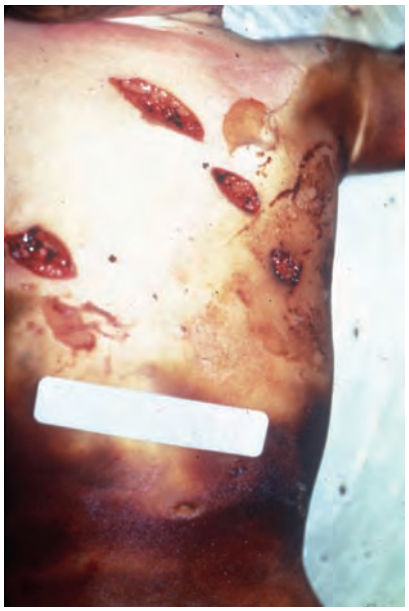

This perpetrator killed his family by inflicting multiple stab wounds and then setting their house on fire. Note the relationship between environmental factors such as heat and the change in wound characteristic. This is demonstrated to a greater degree in the following figures.

This individual was stabbed multiple times and his apartment was set on fire to conceal evidence of the crime. Note the charring of the skin with the sharply margined defect, which was from a stab wound. Internal examination revealed extensive hemorrhage throughout and surrounding the wound track. Sometimes the surface interpretation of such injuries can be challenging. Thermal damage may cause cracking of the skin, which may be misinterpreted as antemortem sharp- or blunt-force injuries.

This individual sustained multiple blunt- and sharp-force injuries due to homicidal violence. She was found many hours after the assault during a hot summer month. Note the putrefactive changes with skin slippage and green to brown discoloration. Also note the wound margins are dry and dark. Some of these wounds could obviously be classified as lacerations or stab wounds and others could not be classified due to distortion by decomposition.

Healed stab wound.

Gunshot Wounds

8

CHARLES A. CATANESE, THOMAS GILSON,
AND THOMAS ANDREW

Introduction

The evaluation of gunshot wounds is an area of importance in forensic medicine. Analysis of gunshot wounds should involve:

- Differentiation of wounds of entrance from exit
- Range of fire estimation for entrance wounds
- Determination of trajectory through the body
- Recovery of evidence (e.g. ballistics)

Whenever possible, clothing worn by a gunshot wound victim should also be examined.

Entrance and Exit Differentation

Entrance Gunshot Wounds

Most entrance wounds appear as circular perforations with a margin of abrasion, a collar of abraded skin, and the entrance defect. The margin of abrasion is produced by stretching with eventual tearing of the skin surface as the bullet enters the body. It is the most reliable feature in identifying entrance gunshot wounds. Exit gunshot wounds lack this feature, as the undersurface of the skin is stretched outward when a bullet exits the body. The margin of abrasion may provide useful preliminary information regarding trajectory of a bullet through the body. When the bullet enters the body perpendicular to the surface, the margin of abrasion is symmetric. As the bullet enters the body on an angle, the margin will be elongated on the side where the bullet first contacts the skin. The trajectory is confirmed by subsequent internal examination.

Exit Gunshot Wounds

These types of wounds typically appear as slit-like or irregular perforations without margins of abrasion. An exception to this is a shored or supported exit, which occurs when a bullet exits from the body where the skin is firmly supported. This might occur as a bullet exits when an individual is leaning against a wall, lying on the sidewalk, or even wearing tight-fitting clothing. These wounds generally have a round appearance and may have superficial abrasion around the defect. Unlike the true abrasion margin of an entrance wound, the shored exit abrasion tends to lack any significant depth and is often very irregular.

Atypical Gunshot Wounds

These are entrance wounds characterized by an irregular appearance. They may have irregular or obscure margins of abrasion. The skin adjacent to the entrance wound may show irregular abrasions or lacerations. Sometimes these perforations may appear as irregular tears. Ricochet bullets, or bullets that pass through intermediate targets, often produce atypical entrance wounds. Sometimes the jacket and the slug may separate and produce two separate irregular entrance wounds. Bullets may strike surfaces such as concrete and fragment into pieces, causing multiple irregular defects from impacts of both twisted metal and rock fragments. Fragments from upholstery, chrome, plastic, etc. may be found at the wound surface or even within the wound track. This atypical appearance may also be seen in certain areas of the body where the skin is thicker (palms and soles) or the skin contour is irregular (face). Atypical wounds may be difficult to interpret, particularly in perforating gunshot wounds where the bullet passes through soft tissue only. Fracture characteristics of bone (especially flat bones such as the skull) and lead fragmentation patterns may give great insight into establishing direction of fire.

Range of Fire Estimation

In estimating range of fire, an attempt is made to determine the distance from the end of the barrel of the gun to an entrance wound at the time the weapon is discharged. This is most reliably estimated by observing the wound and trying to re-create an observed pattern by test firing the weapon. Because the weapon is often not available, estimates based on the physical characteristics of the wound are frequently employed.

An understanding of ammunition aids in understanding range of fire estimation. The ammunition

placed into a firearm is a cartridge. A cartridge consists of a bullet, which rests atop a casing containing gunpowder. A primer located at the base of the cartridge ignites the powder. The primer is itself ignited by the mechanical action of the hammer of the firearm, when the hammer strikes the base of the cartridge. The bullet is the portion of the cartridge that exits the barrel. It is propelled by the burning gunpowder. In general, there is some residual gunpowder that does not burn. In addition to the bullet, burned and unburned gunpowder exits the barrel when a weapon is discharged. This forms the basis of range of fire and estimation.

Range of fire can be divided into three major categories. These include *close range, intermediate range,* and *distant range.* As noted, range of fire is best estimated by the use of test firings employing the known weapon with similar ammunition. Lacking these ideal conditions, estimates can be made on the basic wound characteristics as described next. Even when test firings can be performed it is worth remembering that the wounding characteristics of skin are not identical to gunpowder residue deposition on a test-firing cloth.

Close Range of Fire

This is characterized by the presence of *soot* on the adjacent skin surface or within the wound track. Soot is burned gunpowder residue with a dark powdery appearance. It is critical to remember that soot can be wiped from the body's surface, so gunshot wounds need to be examined prior to any washing, whenever possible. The presence of soot indicates the end of the barrel of the gun was held within approximately 6 to 8 inches from the body's surface. Different guns may have different ranges. The closer the end of the barrel is to the body, the denser the soot deposition. As this distance increases, the soot deposition becomes sparser. Soot may be present with stippling (see intermediate range of fire), but this is more accurately still considered close range of fire. When the end of the barrel is very close to the body, there may also be searing of the adjacent skin from burning gunpowder and flame that extends from the end of the barrel. Soot can be filtered away through intermediate targets such as clothing, car doors, walls, etc. This pattern becomes more dispersed as the distance becomes greater.

Within the close-range-of-fire category are contact gunshot wounds where the end of the barrel touches the body. These may be further subdivided into *tight contact* and *loose contact.* Some classify wounds where the barrel is very close to the body as *near contact*. The term "close contact" is redundant, potentially confusing, and best avoided.

Tight Contact Gunshot Wounds

These are produced when the barrel of the gun is held tightly against the body's surface. Most of the soot will be deposited within the wound track and may not be apparent until autopsy examination. Small amounts of soot may be present at the wound margins. Depending on the body region, type of gun, type of ammunition, and force with which the gun is pressed against the body, the wound may appear different. There are often abrasions adjacent to the gunshot wound. These are associated with the muzzle of the gun rubbing against the body's surface as the gun is discharged. These may take the form of a semicircular rim or a complete imprint pattern of the entire muzzle. When the gun is held tightly to the body, particularly over a bony surface such as the skull, there may be multiple radiating linear lacerations. These lacerations may be small or large and are more commonly associated with larger-caliber bullets. Approximation of the laceration margins reveals a central circular perforation typical of an entrance wound with abraded margins. Observation of the muscle under a tight contact wound may reveal a pink discoloration owing to the introduction of combustion products into the wound track such as nitrites and especially carbon monoxide binding with myoglobin.

Loose Contact and Near Contact

When a gun is held less tightly against the body, soot may escape more readily to the skin's surface. If the gun is discharged at an angle to the body's surface, the pattern of soot dispersion will extend over a greater surface area in the direction of fire. A partial muzzle imprint may be observed. In near contact wounds, the soot is deposited in a dense, relatively small area around the entrance and searing of the skin may be observed as noted above.

Intermediate Range of Fire

This range of fire is characterized by the presence of *stippling* defects around the entrance wound. These are small, (approximately 1/16 inch) red to brown, punctuate abrasions of the skin's surface. Stippling indicates the end of the barrel of the gun was held approximately 18 to 24 inches away from the body's surface. These defects are produced primarily by fragments of unburned gunpowder that exit the barrel with the bullet. These fragments strike the body with greater force than soot and produce permanent defects. Such defects cannot be wiped away from the body's surface. The appearance of stippling may vary depending on the type of gunpowder used in the cartridge, (e.g., ball, disk, or flake).

The term *stippling* applies to the defects in the body's surface. This pattern will become more dispersed as the distance from the weapon to the body's surface increases. Like soot, stippling can be dampened or completely filtered by clothing or other intermediate targets. Powder residue found on intermediate targets at this range of fire is characterized by small, separate, gunpowder fragments.

Soot or stippling defects may be present on an outstretched hand but not surrounding the entrance wound. For this reason, close examination of the upper extremities in gunshot wound cases may be helpful. Range of fire is estimated in the same way and the distance between the weapon and the entrance wound can be approximated by factoring in the distance from the wound to the upper-extremity findings.

Occasional "pseudostippling" may be seen around an entrance wound. This phenomenon is generally seen when a bullet passes through an intermediate target and fragments of the intermediate target strike the skin's surface (causing abrasions) around the entrance wound. As noted, such entrance defects tend to be irregular and the adjacent stippling defects exhibit wider size variation than true stippling, where the abrading gunpowder tends to produce abrasions of relatively uniform size.

Distant Range of Fire

This range of fire is characterized by the absence of gunpowder residue around an entrance gunshot wound without an intermediate target between the end of the barrel and the wound. This indicates the end of the barrel of the gun was held more than approximately 18 to 24 inches away from the body's surface. Some experts prefer to call this intermediate range, especially when the possibility of an intermediate target cannot be excluded.

Trajectory Determination

The path a bullet takes through the body may have substantial medicolegal significance. In some cases, the trajectory can be easily determined by careful dissection. In other cases, (e.g., multiple gunshot fatalities), full delineation of wound tracks may be harder to accomplish as a result of extensive internal injury. Surgical intervention may also obscure wound tracks in part. It is also worth remembering that a wound sustained in a certain body position may look slightly out of alignment with the body lying supine on the autopsy table. Superficial wounds can pose difficulties in trajectory determination as entry and exit wounds may not be easily identified. These include graze and tangential gunshot wounds.

Graze-Type Gunshot Wounds

These are produced when the bullet contacts the skin superficially at a very narrow angle, producing an elongated superficial oval abrasion. One may determine direction of fire by a semicircular margin of abrasion at the entrance side and a more irregular margin at the exit side.

Tangential Gunshot Wounds

These are produced when the bullet strikes the skin superficially at a narrow angle and creates lacerations of the underlying subcutaneous tissue and overlying skin. Direction of fire may be established by the presence of a semicircular margin of abrasion at the entrance side. Also, the laceration of the skin overlying the wound track produces tears with skin tag formation. These tags will extend outward on angles. The tips of these lacerations, at their most medial aspects (e.g. within the wound track), point toward the side from which the bullet entered the body.

Gunshot Wounds in Bone

Gunshot injuries of bone may be helpful in trajectory determination, especially in the skull and, less frequently, other flat bones. When a bullet passes through the skull, it creates a cone-shaped beveled defect. The entry point has crisp margins and the exit point is larger and represents the wider portion of the cone. Bullets entering the skull produce internal beveling and those exiting the skull produce external beveling. A tangential bullet strike of the skull may cause the bullet to fragment and result in complex beveling if a portion enters the cranial cavity and a portion shears off and remains outside the surface. In such cases, the bony defect created may exhibit internal and external beveling. Such defects are referred to as "keyhole" defects, as their appearance may resemble a keyhole.

High-Velocity Gunshot Wounds

Trajectory delineation in high-velocity gunshot wounds is similar to low-velocity injury, however, the extent of injury is far more significant and exit wounds are often dramatically large. Hunting ammunition used at high velocity will fragment more in the body than low-velocity bullets, creating a characteristic x-ray picture called a "lead snowstorm."

Evidence Recovery

Bullets retained in the body must be recovered at autopsy. Rifled firearms leave characteristic markings

on bullets that can be used for identification of a firearm by comparison with projectiles known to have been fired from that weapon. Care must be taken during dissection to avoid or minimize the creation of any additional marks on the bullet that might obscure the rifling marks. This also applies to any identifying marks inscribed place directly on the bullet after recovery. These identifying marks should be inscribed on the base of the bullet. Older bullets should be recovered at the time of autopsy. These may be enclosed in fibrous capsules at the site of lodgment and often have a dull gray appearance as a result of oxidation. Rifling marks may still be present and useful for ballistic comparison.

Shotgun Wounds

Because of differences in construction and ammunition, shotguns merit special consideration. Shotguns can be used to fire a single projectile (slug) or several pellets (shot). Shot size can vary (buckshot versus birdshot). Slugs behave essentially as single projectiles. The shotgun cartridge (or shell), contains some additional components (e.g., wadding, filler material) that might produce injuries that can be used in range of fire estimation. Soot and stippling remain useful in close- and intermediate-range determinations. The wounds produced by shotgun shells containing shot look different from conventional bullet wounds. Soot, stippling, wadding, and other components can be used for range of fire estimation, but the behavior of the pellet cluster over distance provides additional information regarding range of fire. When the shot exits the barrel, it travels initially as a tightly grouped cluster. Striking the skin, the tight cluster produces a round defect. The grouping opens up over distance with the first noticeable change in wound appearance occurring at about 3 feet. A single defect is still observed but the margins now take on a scalloped appearance. With further distance (and more pellet dispersal), wounds now consist of a central defect with scattered satellite defects surrounding them. These satellite defects increase in number with greater distance, until the central defect is lost entirely and pellets maintaining sufficient velocity strike the skin's surface individually. The dispersal pattern of pellets is not predictable once an intermediate surface intervenes. It is not possible to evaluate these wounds for range estimations and, for similar reasons; estimates of range of fire by x-ray examination are unreliable. As with other firearms, test firings are the most reliable means for range of fire approximations.

Distant entrance gunshot wound. This bullet struck the body nose end first, roughly perpendicular to the surface. Note the thin and roughly symmetric margin of abrasion surrounding the slightly oval perforation.

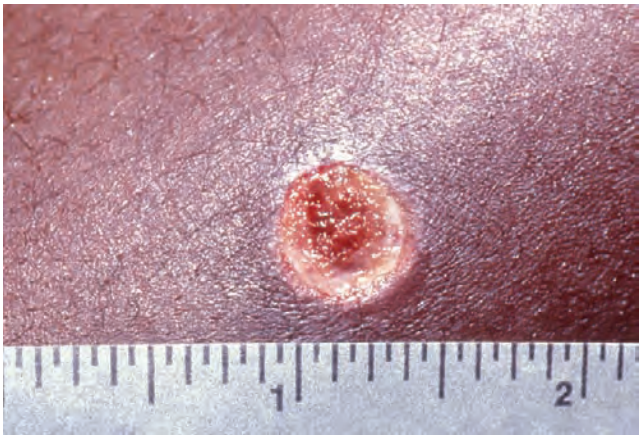

Distant entrance gunshot wound in a darker-skinned individual. Note the slightly asymmetric margin of abrasion, which is greatest at its inferior aspect. This indicates the bullet struck the body nose first, and at a slight upward trajectory. These directions are stated *relative to the body*.

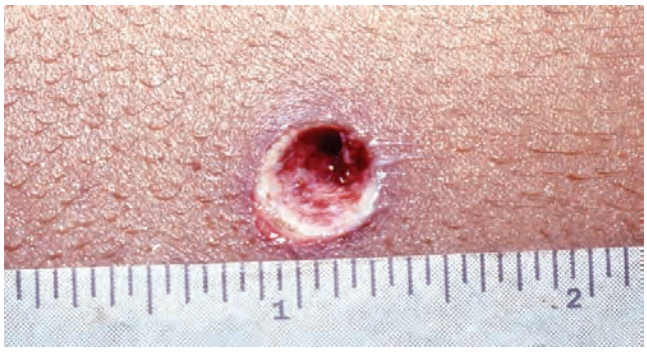

Distant entrance gunshot wound with slightly asymmetric margin of abrasion, which is greatest at the right inferior aspect. This usually indicates the bullet struck the body nose first, almost perpendicular and at a slightly upward and right-to-left trajectory. These directions are stated *relative to the body*.

Distant entrance gunshot wound with slightly asymmetric margin of abrasion greatest on its superior and left aspect. This usually indicates that the bullet struck the body nose first, almost perpendicular and at a slightly downward, left-to-right trajectory. These directions are stated *relative to the body*.

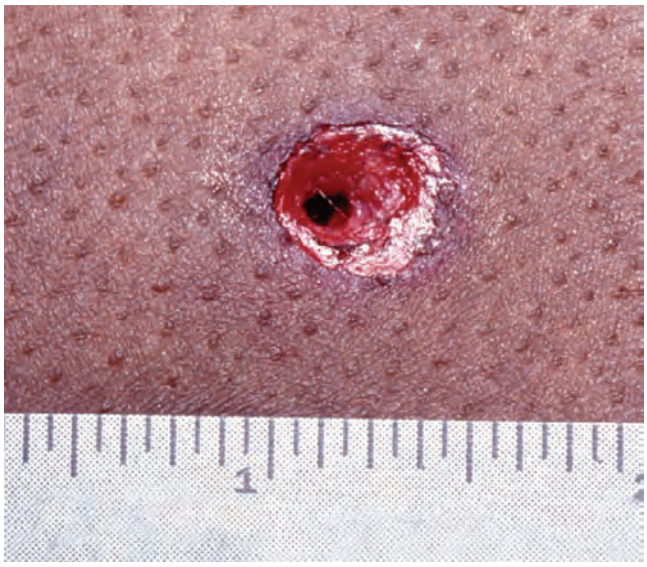

Distant entrance gunshot wound with slightly irregular, widened margin of abrasion, which is greatest at the right lateral aspects. This indicates the bullet struck the body from a slightly inferior and right-to-left trajectory. The slightly irregular nature of the margin of abrasion may sometimes be seen, as the bullet perforates clothing first.

Distant entrance gunshot wound that is slight stretched into an oval due to skin tension. Note the slightly asymmetric margin of abrasion, which is greatest at the 9 o'clock through 12 o'clock position, indicating the bullet struck the body with a slightly downward and left-to-right trajectory.

There are six separate clustered entrance gunshot wounds to this individual's head. Note the abrasion to the posterior ear helix caused by a bullet. Shaving the hair is recommended for external examination of such gunshot wounds.

Distant entrance gunshot wound to the palm of the hand through thick skin.

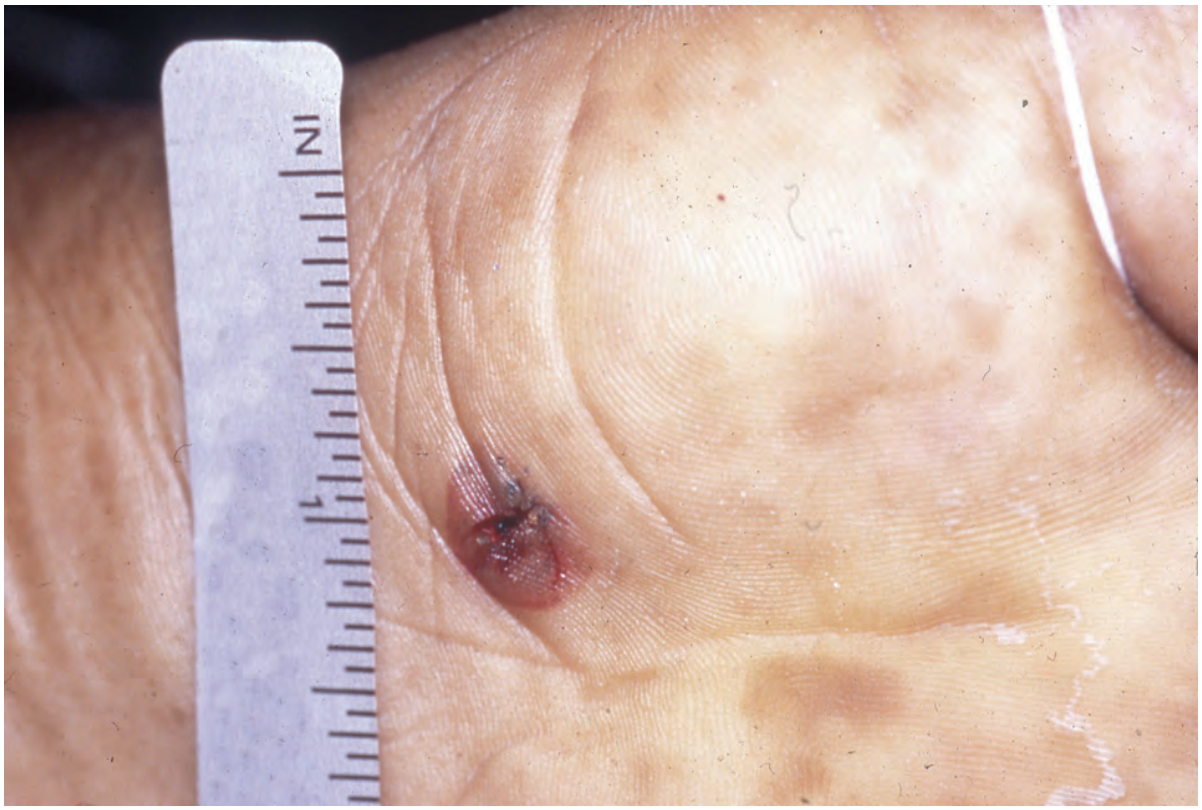

Distant entrance gunshot wound through the thick skin of the sole of the foot. This bullet first perforated the bottom of a shoe and a sock before entering the body.

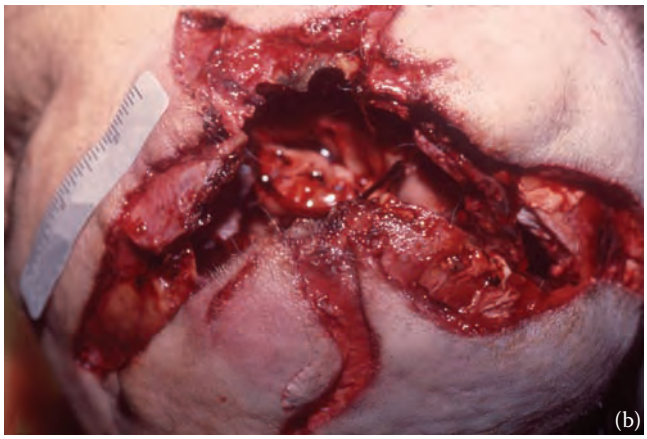

Entrance gunshot wound with surrounding stippling defects indicating intermediate range of fire. This gunshot wound (a) to the thick skin of the palm of the hand reveals typical small radiating lacerations without prominent margin of abrasion. These characteristics are typical for gunshot wounds of the palms of the hand and soles of the feet. These images (b–e) represent the same contact gunshot wound with varying degrees of tension surrounding the wound's surface. Note that with approximation of the margins (d–e) the wound forms a roughly circular perforation with margin of abrasion typical of an entrance gunshot wound. Also note the copious amounts of soot within the wound track and at the underlying bone surface. Such large radiating linear lacerations are usually associated with higher-caliber guns and tight contact of the gun muzzle on the body surface with underlying bone close to the skin. The expanding gases from burning gunpowder forced into the wound track causes such lacerations.

Contact gunshot wound to the scalp with radiating lacerations and soot deposition within the wound track and at the wound margins. These lacerations are caused by burning gunpowder, with expanding gases forced into the wound track and causing the overlying skin to lacerate.

Contact gunshot wound to the face (a). Note the stellate lacerations radiating from the perforation site with soot deposition; (b) a close-up view with a different orientation.

Contact gunshot wound to the right temporal region. Note the larger lacerations extending from the perforation site with adjacent soot deposition.

Contact gunshot wound to the eye. Note the lacerations of the eyelid and eye globe with soot deposition at the adjacent skin and within the wound track.

Contact gunshot wound to the temple. Note the small radiating lacerations from the perforation site. Note the presence of soot within the wound track and adjacent skin.

Note this tight contact gunshot wound to the right temporal region with lacerated borders and soot deposition. A .357 Magnum caused a comminuted skull fracture with head deformity and the eye protruding from the socket.

Contact gunshot wound to the scalp. (a) Note the small radiating lacerations with soot deposition. (b) Note the same wound after cleaning. The tension surrounding the wound is relieved by pressing the surrounding skin inward toward the perforation. This will approximate the wound margins and reveal a roughly circular perforation with margin of abrasion characteristic for an entrance gunshot wound. Note that even after cleaning, small amounts of soot are still present at the wound margins. Further cleaning might eliminate this gunpowder residue as well. In many cases, such as this one, it is essential to photograph the body before cleaning.

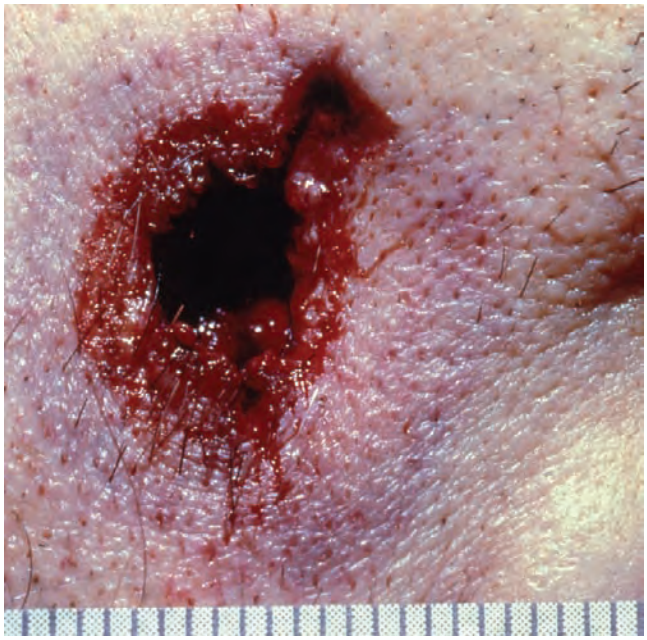

Contact entrance gunshot wound to the temple. Note the abraded imprint of the eyepiece portion of the gun at the superior aspect. The perforation site has a margin of abrasion. There is no obvious soot at the surface of the adjacent skin, but there were copious amounts of soot within the wound track. Note the pink to red discoloration surrounding the perforation due to nitrates and carbon monoxide released from burning gunpowder. These components may sometimes cause this discoloration when reacting with the underlying muscle. In this case, it is more obviously due to the decedent's light skin color.

Contact gunshot wound with muzzle imprint.

Contact gunshot wound. Note the adjacent abrasion to the perforation site due to contact with a revolver ejector rod when the gun was discharged. There are also small amounts of soot visible at the wound margins and more within the wound track.

Contact gunshot wound to the scalp with an abrasion at the superior left aspect due to the gun muzzle. There is visible soot within the wound track. Sometimes it may be difficult to visualize the presence of gunpowder residue on thick, dark, scalp hair.

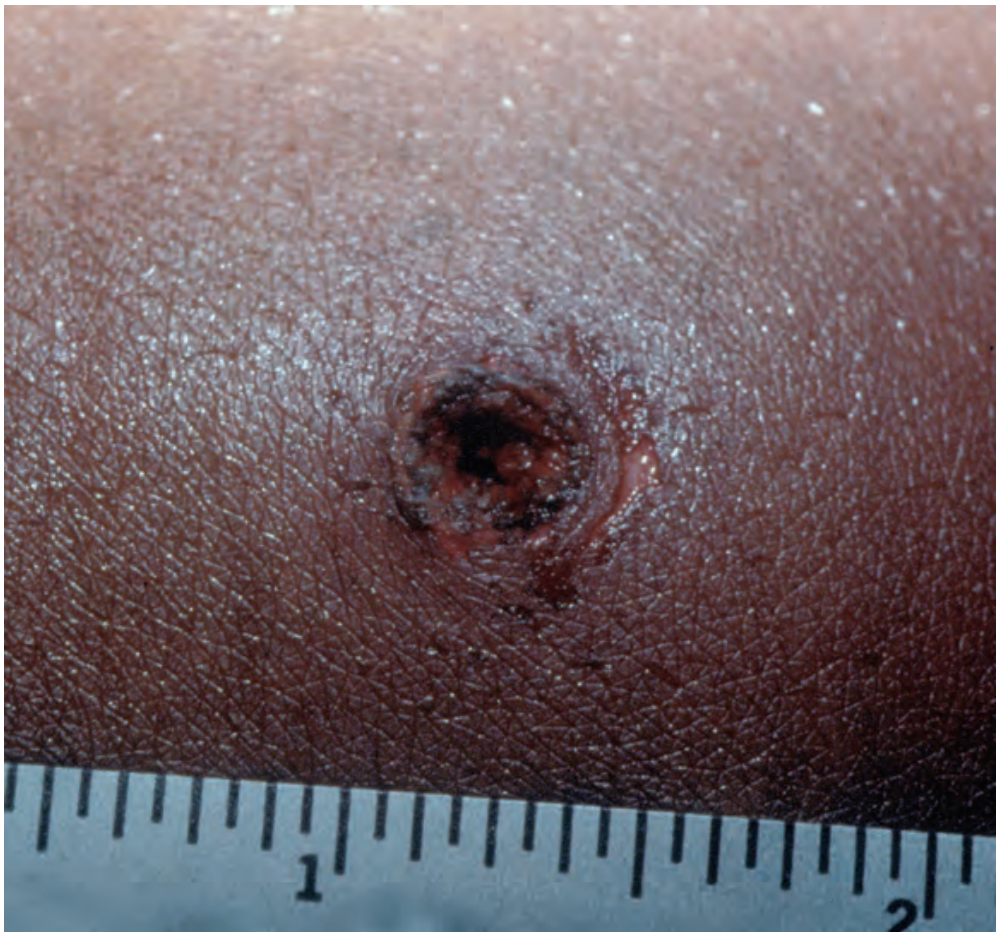

Contact gunshot wound. Note the soot at the margin of the perforation site. Note the surrounding abrasion from a partial muzzle imprint.

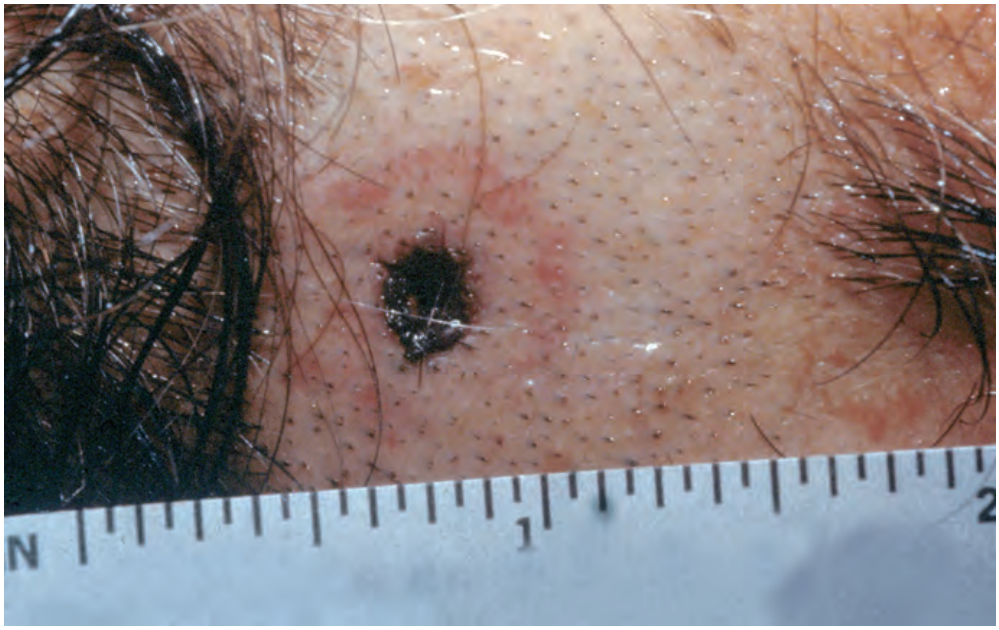

Contact gunshot wound to the right temporal scalp. Note the soot deposition surrounding the margin and the red discoloration of the adjacent skin.

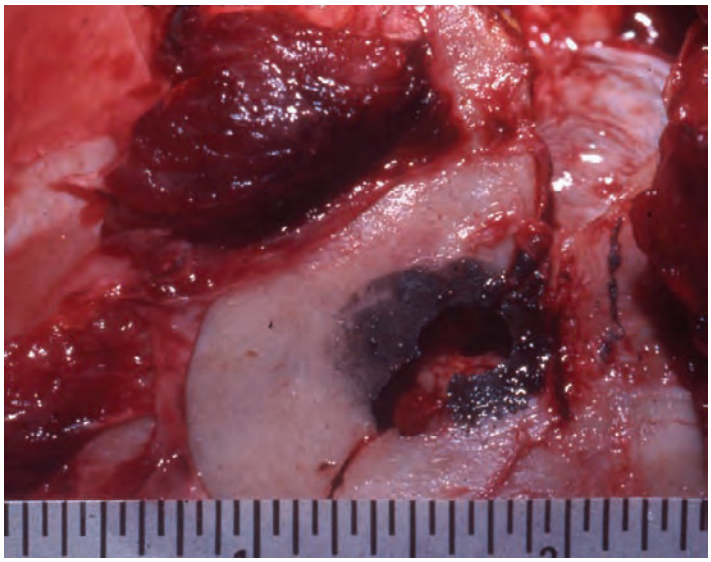

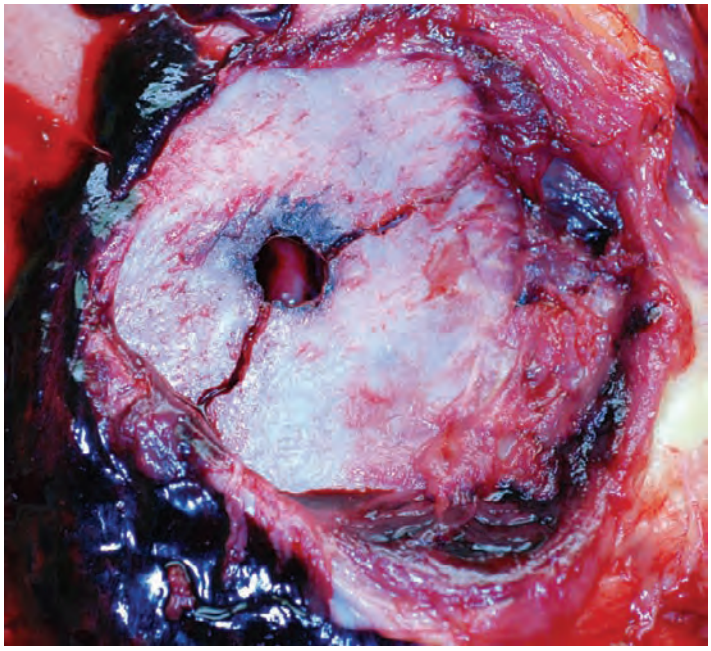

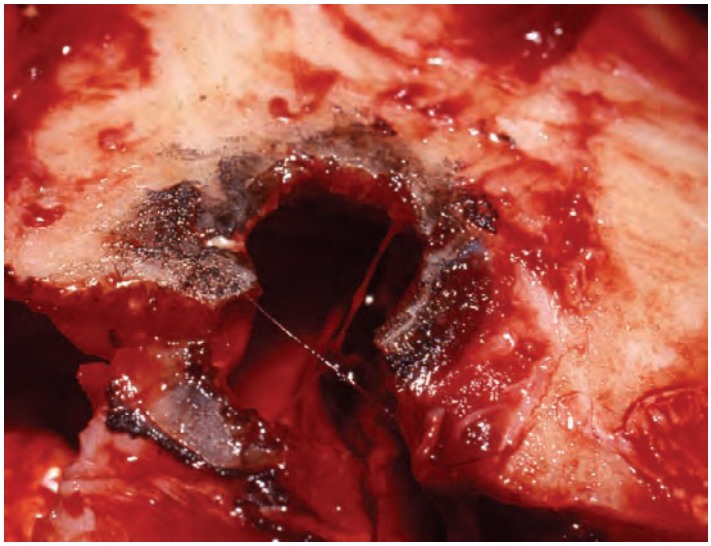

Underlying soot deposition within the wound track on the outer surface of the exposed skull. This is typically seen with a tight contact gunshot wound where soot is forced into the wound track.

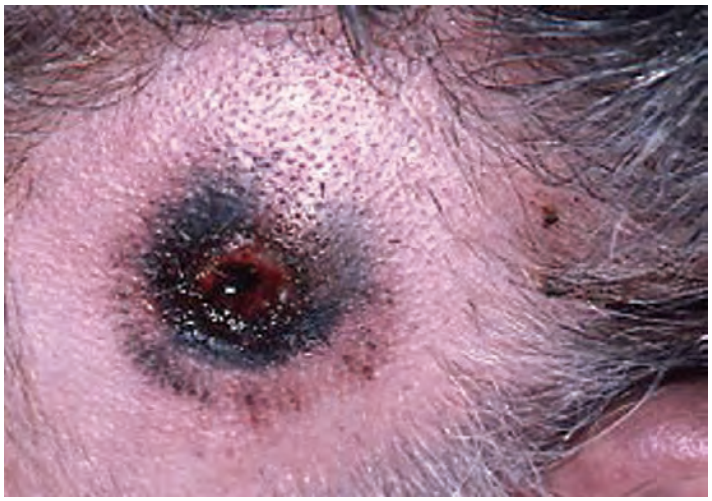

Close-range gunshot wounds with soot deposition.

Note the entrance wound is to the left where soot is visible at the adjacent skin.

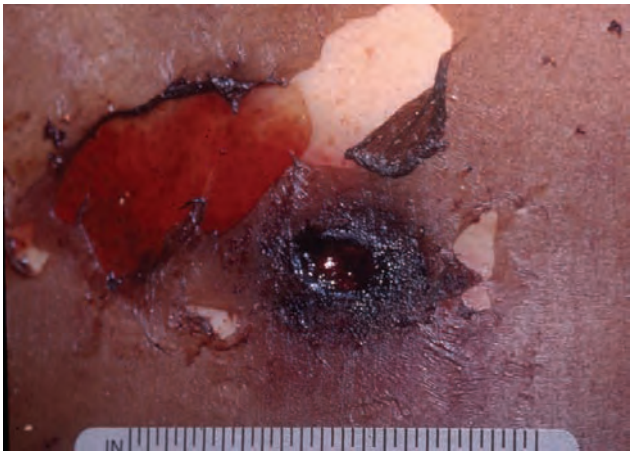

Close-range gunshot wound to a body with early decompositional changes. Note the soot deposition surrounding the perforation site. These early putrefactive changes consist of skin slippage and discoloration.

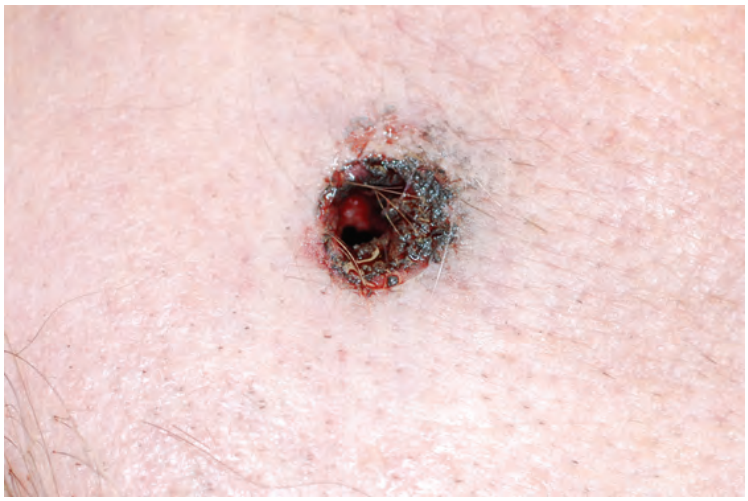

Loose contact with more soot extending to the right side of the perforation.

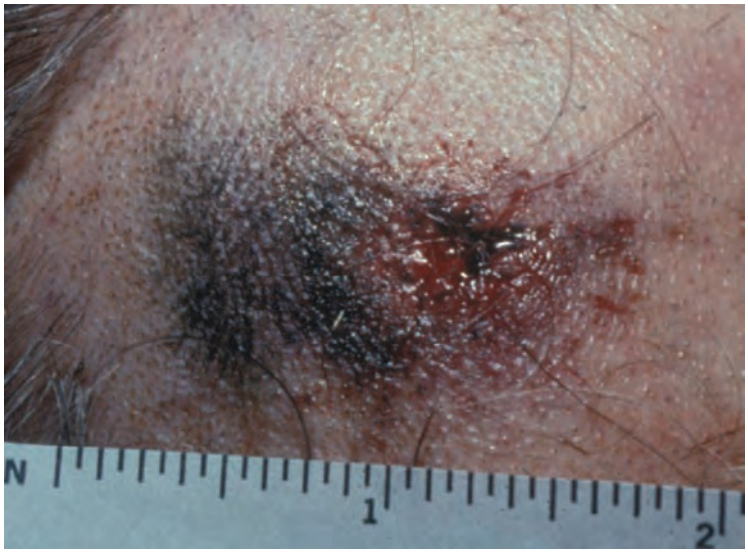

Loose contact with the muzzle more tightly applied at the right aspect. Note the greater soot deposition to the left of the perforation.

Close-range gunshot wound with soot deposition.

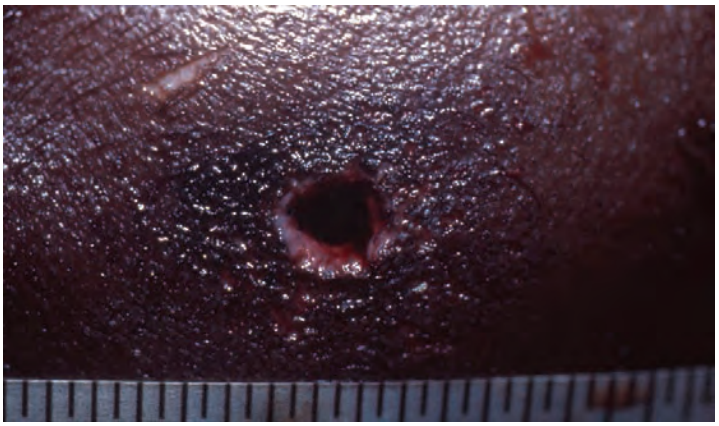

Close-range gunshot wound in a very dark-skinned individual. Note the gunpowder residue surrounding the entrance gunshot wound.

Close-range gunshot wound with a muzzle flare burn and soot encircling the perforation site.

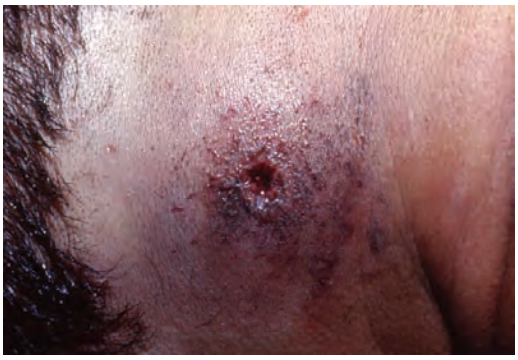

Close-range gunshot wound with sparse soot deposition surrounding the perforation site with adjacent superficial skin erosions due to muzzle flare.

Close-range gunshot wound with soot deposition, superficial skin erosions from muzzle flare with tightly compacted stippling-type defects.

Close-range gunshot wound to the scalp with sparse soot deposition, adjacent burning of the scalp due to muzzle flare, and closely packed stippling defects.

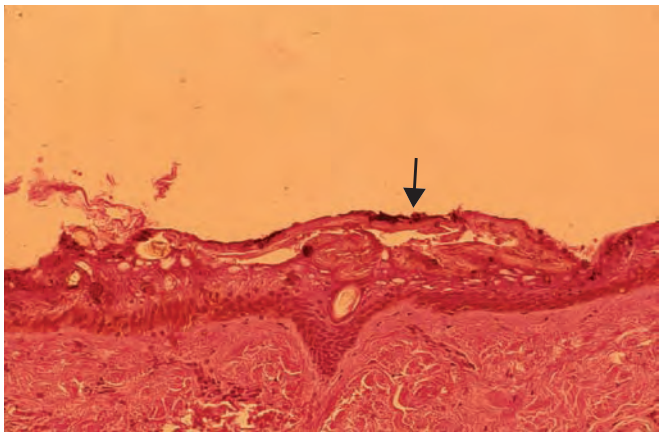

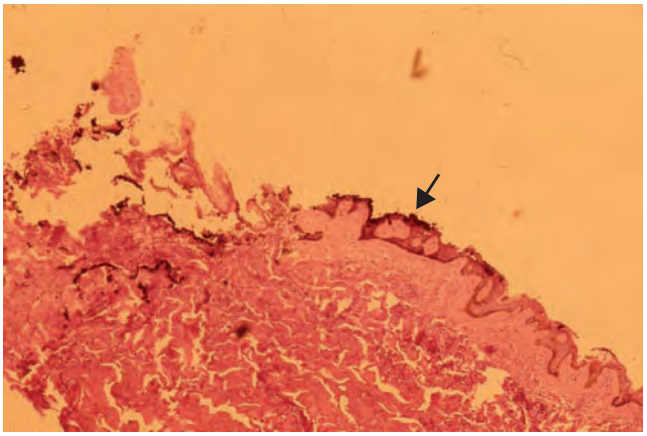

These demonstrate histopathology sections of skin with soot deposition at its surface. The soot appears as a black film on the skin's surface; unburned powder particles may also be visible.

Gunpowder residue at the fingers due to firing a handgun. This individual died from a self-inflicted gunshot wound to the head. The amount of gunpowder residue following discharge of a firearm may be quite variable and sometimes not very obvious.

Soot may also be deposited as it exits from the muzzle of the gun or from cylinder gap in revolver type handguns.

This pattern of soot was caused by holding a revolving cylinder while discharging the weapon.

Note the "blowback" blood spatter pattern on the hands of these individuals who died from self-inflicted gunshot wounds to the head. Note that in (a) the spatter pattern is partially obscured by subsequent contact deposit of blood. Also note in (d) the extent of spatter may be minimal.

All of these are articles of clothing demonstrate close-range gunshot perforations with soot deposition. It is always important to inspect the decedent's clothing. Multiple layers of heavy clothing may filter gunpowder residue from the body surface. It is also important to distinguish bullet wipe from soot deposition. Bullet wipe is a small, encircling, gray discoloration around the perforation site of the clothing due to lubricants and residue from within the barrel of the gun that adhere to the bullet surface as it passes through the barrel. Unlike soot, bullet wipe is not useful in range of fire estimations.

Clothing with soot indicating close range of fire. The individual wearing the hat had long curly black hair and there was no appreciable soot noted at the scalp.

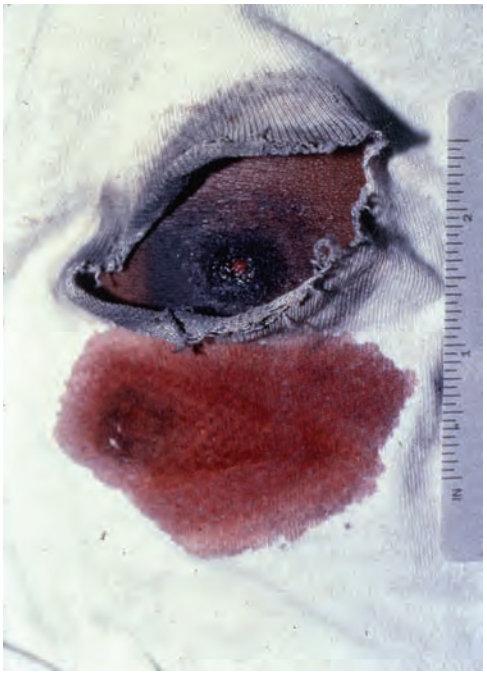

Note the large tear in the clothing from this contact wound and the minimal amount of adjacent external hemorrhage. This is an antemortem injury with more than a liter of blood observed in the thoracic cavity during autopsy.

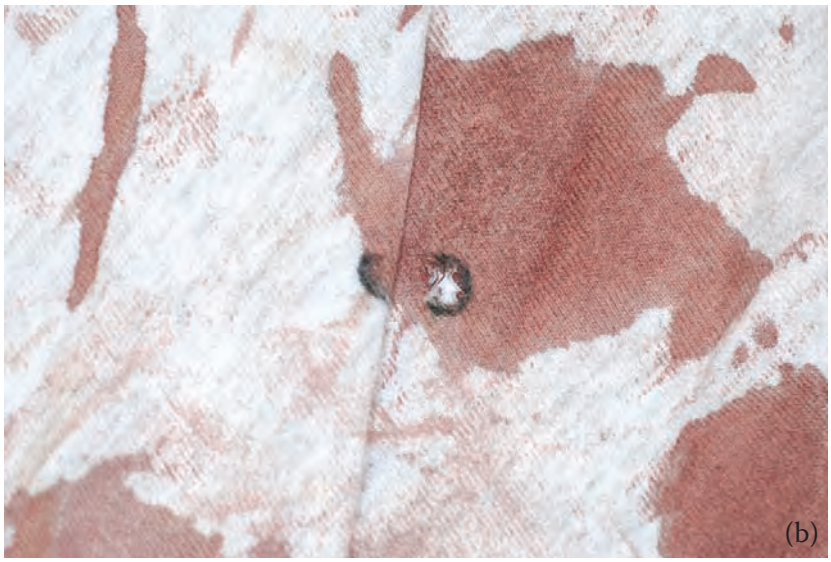

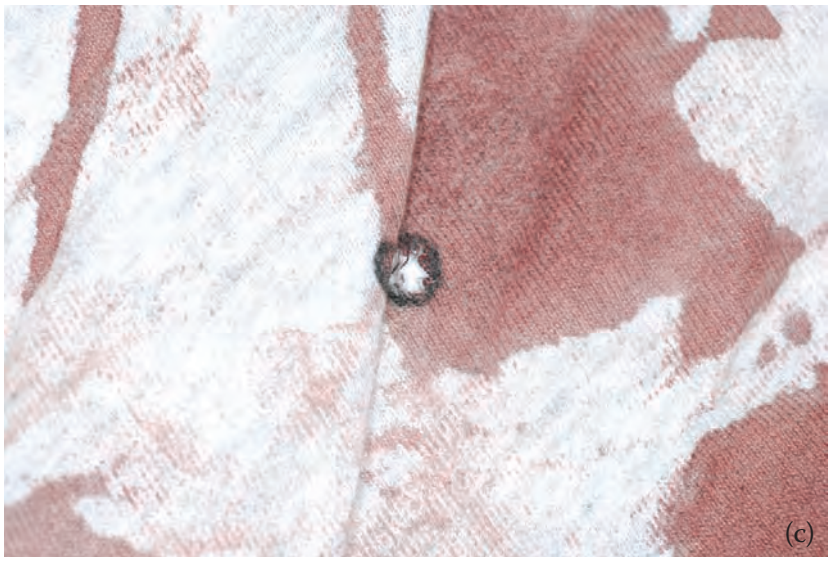

Bullet wipe. This individual had his shirt tucked in, creating folds in the cloth. Note the dark discoloration around the perforation site with the right aspect separated several inches from the site of bullet entry. By refolding the fabric and approximating the margins we can simulate how the individual actually wore his shirt (a–c).

This entrance gunshot defect is surrounded by dark clotted blood. The dark discoloration of blood, particularly on dark clothing, may make if difficult to observe soot. This is a distant gunshot wound (d).

Intermediate range of fire. Tightly packed stippling defects surrounding this entrance gunshot wound to the cheek. There is no apparent soot deposition.

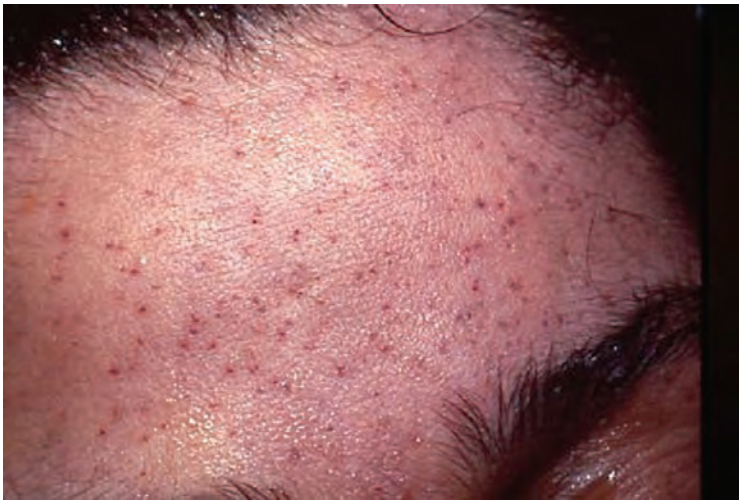

Multiple stippling defects to this individual's forehead indicating intermediate range of fire.

Intermediate range of fire with stippling defects. Stippling defects represent abrasions that, unlike soot, cannot be wiped away.

Intermediate range entrance gunshot wound to the eyebrow with stippling defects across the face. Note the irregular nature of the wound due to the location of the gunshot wound through the eyebrow ridge, with the underlying frontal bone closely subadjacent to the skin surface.

Intermediate range of fire.

Entrance gunshot wound with stippling defects indicating intermediate range of fire.

Intermediate-range entrance gunshot wound with stippling defects. The scalp is partially shaved. Note the decreased number of defects in the shaved area as a result of hair dampening the effect. The weapon was a sawed-off 30/30 hunting rifle that was reportedly discharged approximately 15 inches away from the decedent's head.

Intermediate-range gunshot wound with stippling in a dark-skinned individual.

Intermediate-range gunshot wound. Note the sparse stippling defects along the bicep and forearm with sparing of the anticubital fossa region. This indicates the decedent had his arm bent when he was shot.

This individual had stippling defects involving the posterior portion of his lateral outer ear and temporal scalp. Note the sparing of the posterior medial ear.

Stippling defects with soot deposition on the side of an outstretched hand, indicating intermediate to close range of fire.

Multiple-entrance gunshot wounds. Note the sparse stippling defects indicating intermediate range of fire at the skin anterior to the ear. The farther away a gun is discharged from the body, the more spread out the stippling defects become.

Intermediate range of fire. Note the sparse stippling defect surrounding the bullet perforation site.

A few sparse stippling defects indicating intermediate range of fire. Note the contusion surrounding the perforating gunshot wound at the arm. This dark discoloration should not be confused with gunpowder residue.

Note the multiple gunshot wounds to this individual in varying directions. She was shot multiple times while jumping in different directions in an attempt to run away. Note the contusion to her left shoulder and the stippling defects to her face. It is important to realize that the direction of fire may vary greatly.

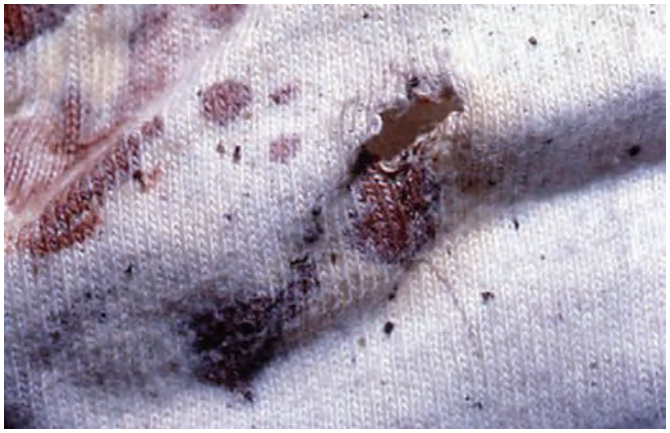

Gunshot wound perforation of clothing with multiple gunpowder residue flecks at the surrounding surface. The underlying entrance gunshot wound still revealed a sparse stippling pattern consistent with intermediate range of fire.

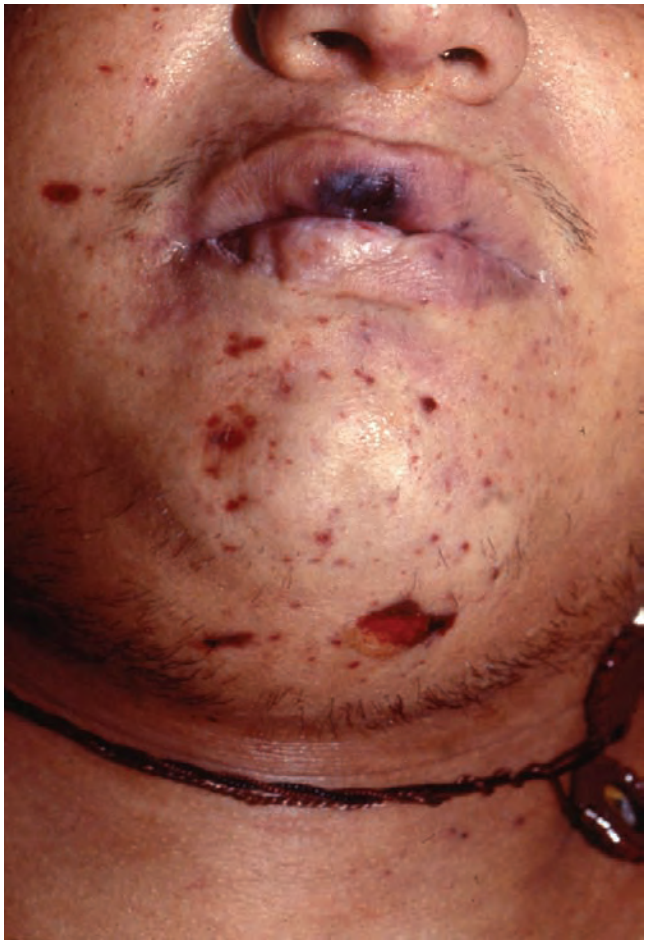

Distant gunshot wound with pseudostippling to this individual's face caused by fragmented debris as the bullet passed through a car window.

Distant gunshot wound with pseudostippling. Note the large size variation in the abrasions to the gunshot wound. True stippling defects are more uniform in size.

Atypical distant-entrance gunshot wound with surrounding contusion.

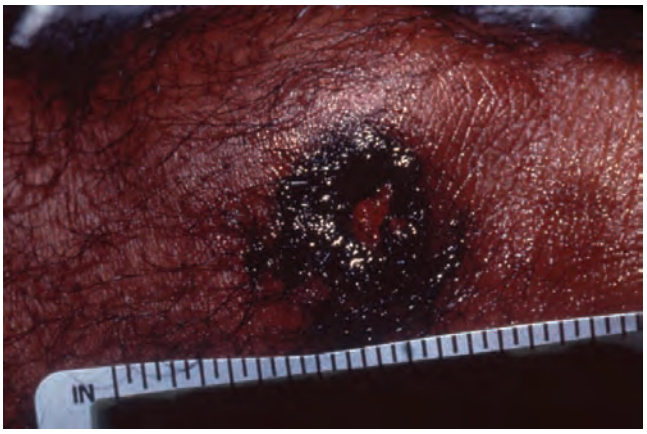

Distant-entrance gunshot wound with surrounding clotted blood. In this case, histopathology of the adjacent skin demonstrated no dark particles consistent with gunpowder residue. It may be challenging to visualize soot when it is mixed with dried blood. Dried clotted blood has a shinier character than gunpowder residue, which has a duller appearance. It may be challenging to interpret range of fire when both dried blood and soot are present simultaneously. If there is any question concerning the presence of soot, we recommend microscopic examination to help clarify this point.

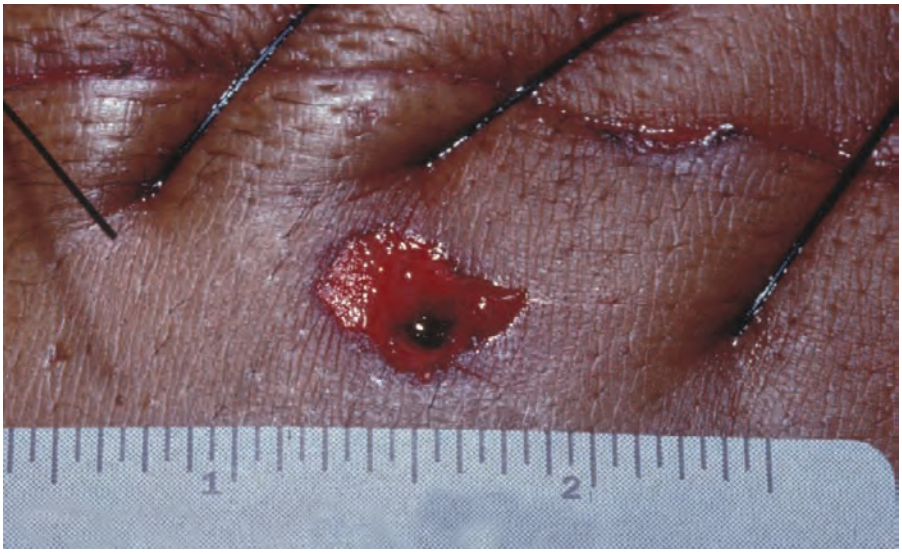

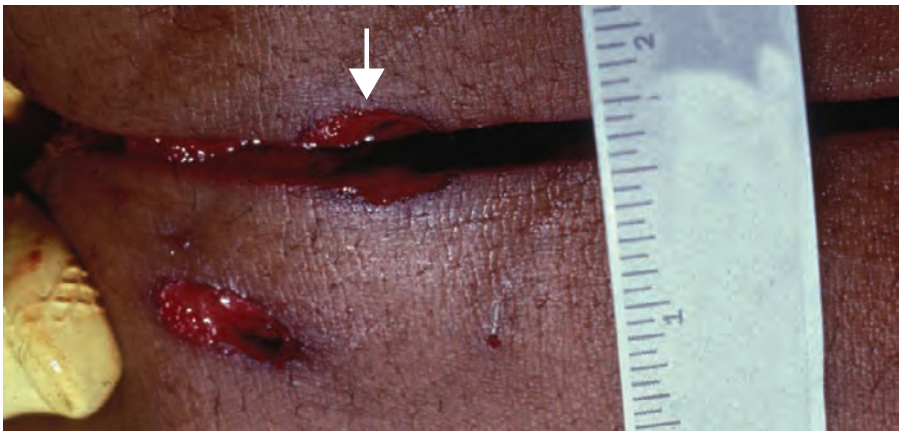

Thoracotomy incision through entrance gunshot wound. One should never cut through the wound during resuscitative efforts. No one would argue the importance of saving the individual's life. Cutting next to the wound will not affect the chances for survival. Cutting through the wound may make wound interpretation very difficult and hinder criminal proceedings. Note one gunshot wound was cut through and hidden in the suture line. The defect becomes more obvious with suture removal.

Note the large gunshot wound at the medial aspect of this thoracotomy incision.

This is a gunshot wound that is several days old. This person died of septic complications. Note the adjacent wound infection distal to the GSW site.

Exit gunshot wounds. Note the typical, often irregular, slit-like defects without margins of abrasion. The appearance is generally more irregular and less round than the entrance wound.

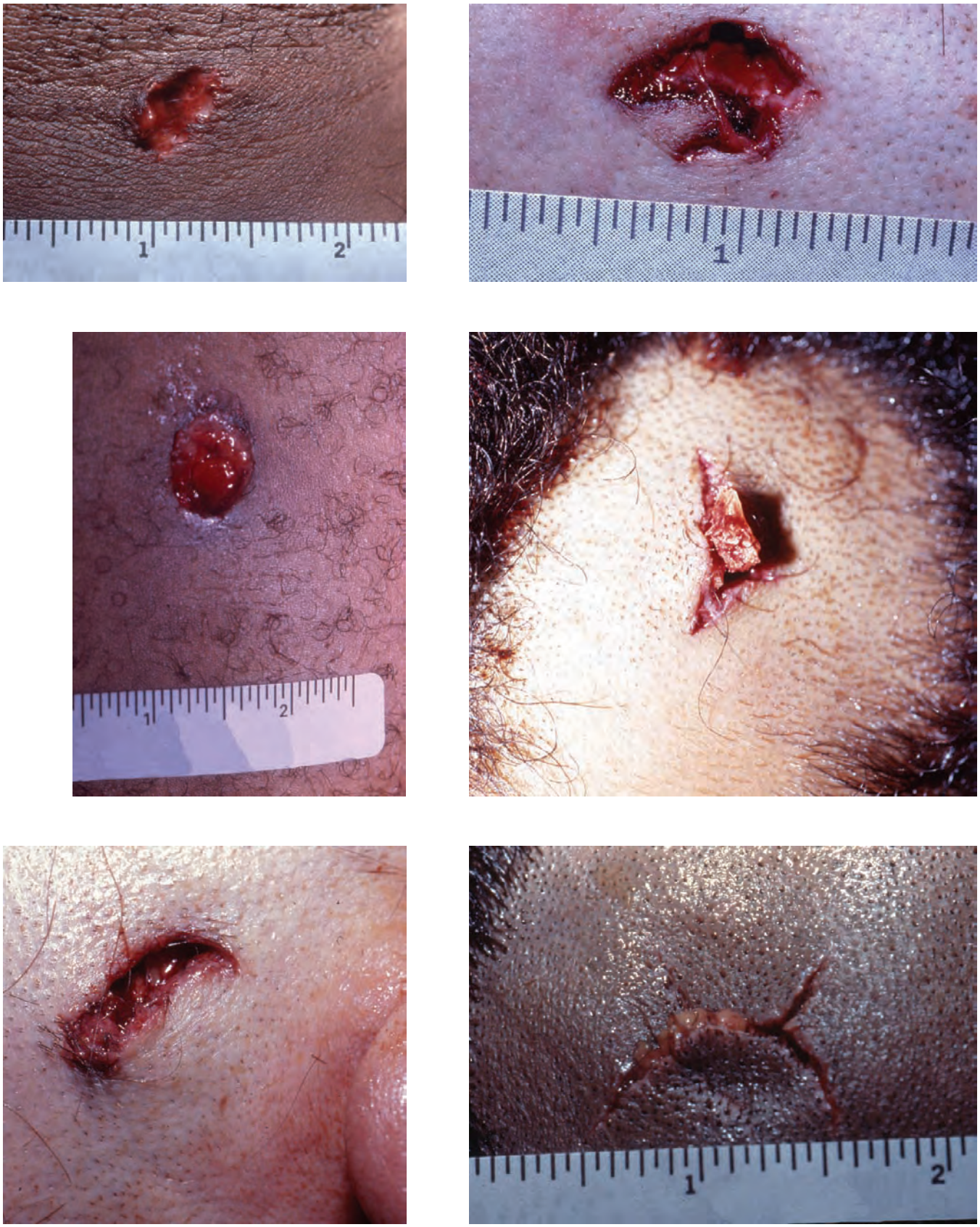

Exit gunshot wounds.

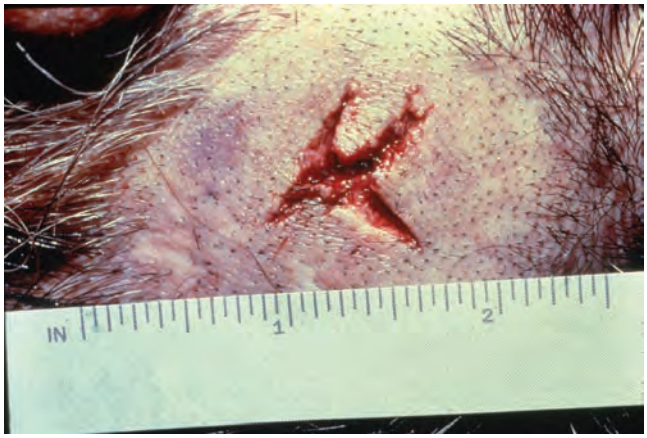

Exit wound with lacerated margins.

Exit wound from a sawed-off 30/30 hunting rifle.

Exit wound from a handgun. It is not typical to get such a large laceration from a medium-caliber bullet that already perforated the skull.

Exit gunshot wound through the right nares.

Exit gunshot wound to the lower face caused by a bullet that was markedly deformed by striking underlying bone before exiting.

Exit gunshot wound through the finger.

Exit gunshot wounds through the palm of the hand.

Exit gunshot wound through the thumbnail bed.

This individual was shot in the lower back while leaning forward. The bullet perforated the structures of the thoracic cavity and neck, producing a grazing-type gunshot wound to the tongue. The bullet then knocked out the individual's front tooth before exiting the body from the open mouth.

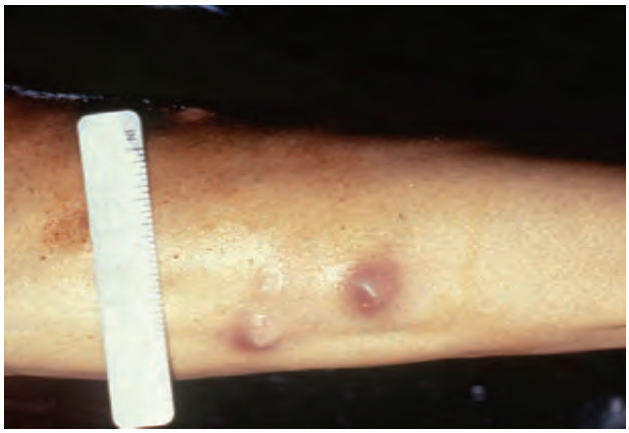

Note the lumps directly underneath the skin's surface. These are caused by bullets just beneath the surface of the skin that did not have enough kinetic energy to exit the body.

A bullet was retrieved just beneath the surface of the skin within this area of contusion.

There is a large deformed gray metal slug just underneath the surface of the skin that is partially visible underneath the broken skin.

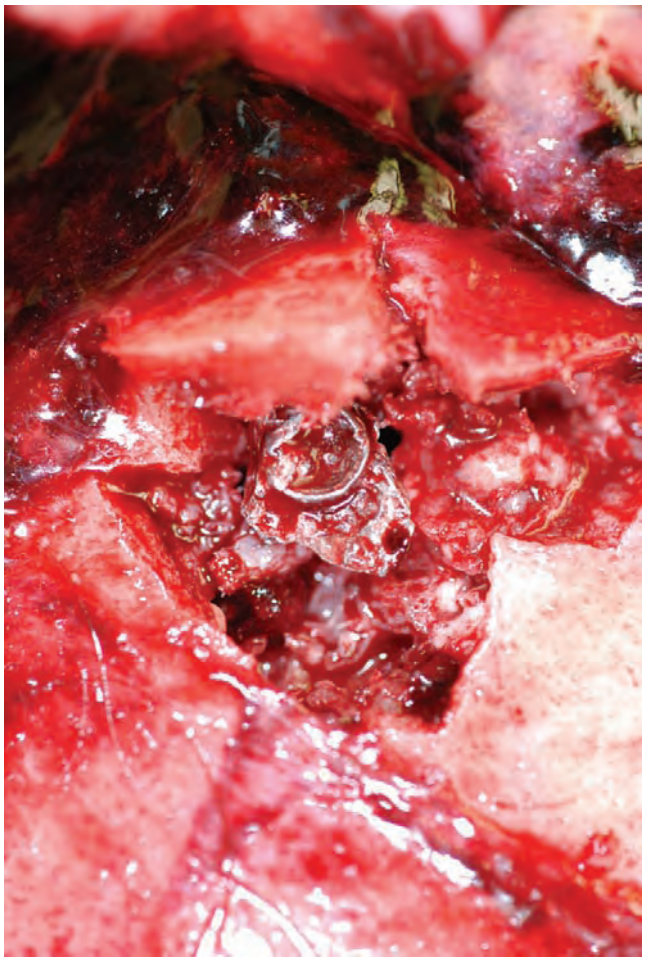

Bullet retrieved from skull fracture directly underneath the skin's surface.

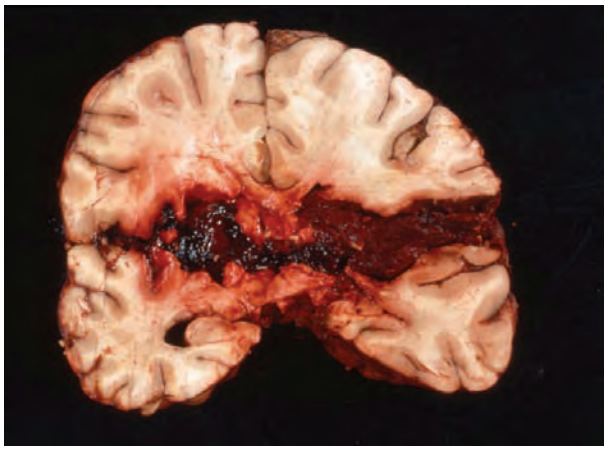

This image demonstrates a hemorrhagic pulpified wound track through the cerebral hemispheres caused by a medium-caliber handgun bullet.

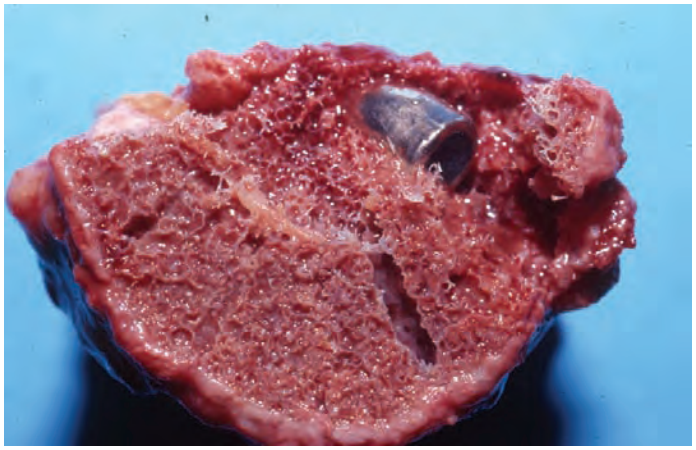

This bullet was retrieved within this portion of thoracic vertebra. It is important to carefully remove such bullets encased in bone, so as not to destroy ballistic markings on the sides of the bullet.

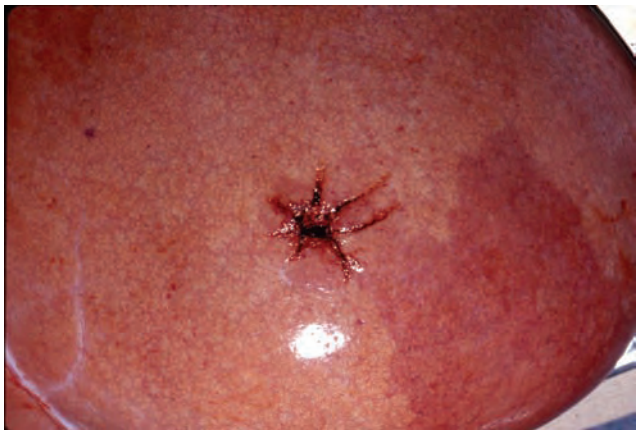

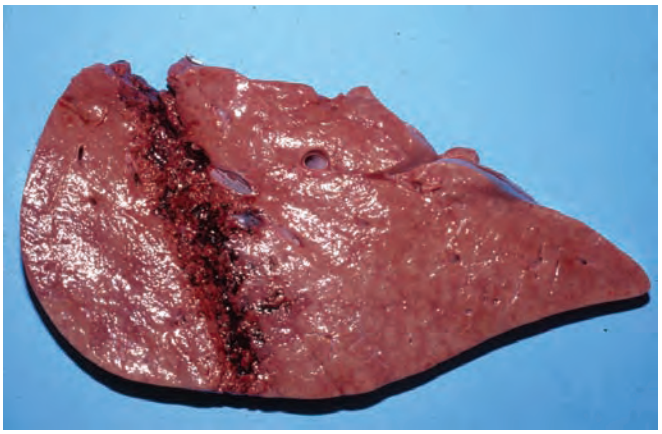

These figures demonstrate a wound track through the liver caused by a medium-caliber handgun.

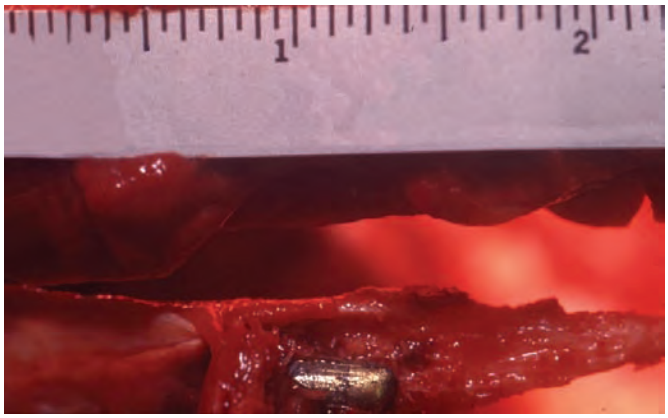

This individual had been shot several weeks earlier with a .22 caliber handgun. One of his ribs fractured while removing the breastplate during autopsy and a bullet was retrieved from within the medullary cavity.

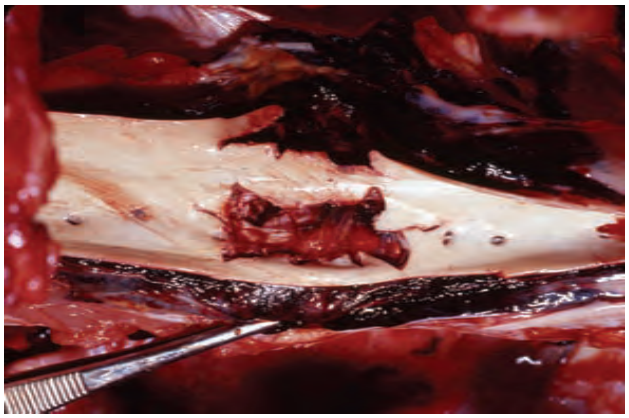

A wound track through the aorta with adjacent surrounding hemorrhage. This wound was caused by a medium-caliber weapon.

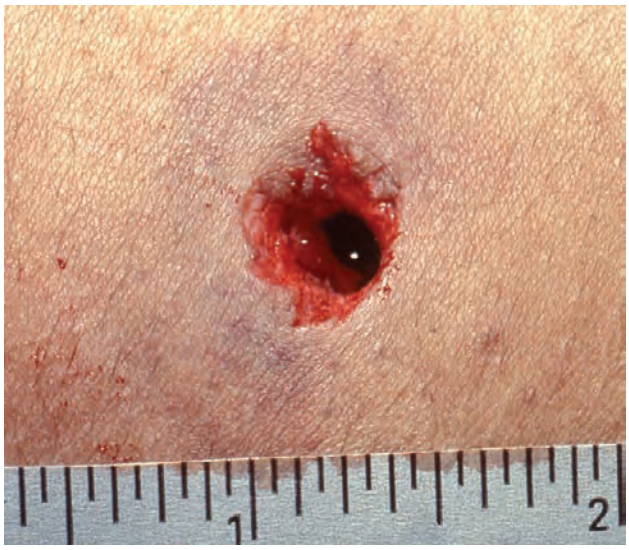

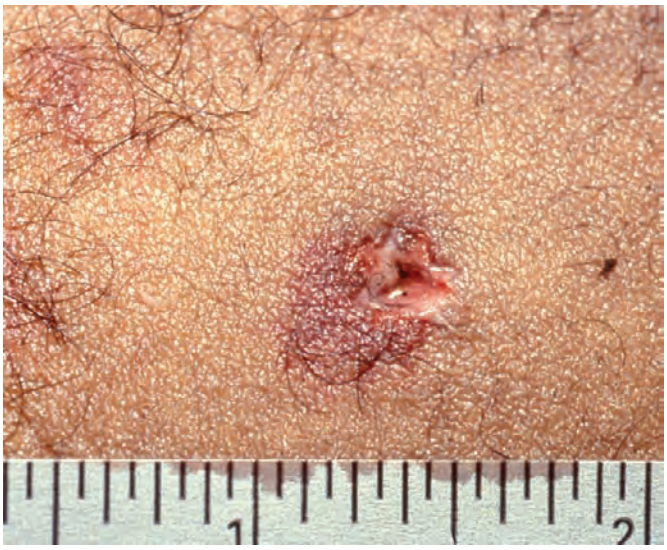

Shored exits. These wounds were produced by bullets exiting the body where the skin's surface is supported (e.g., by a firm surface or tight clothing).

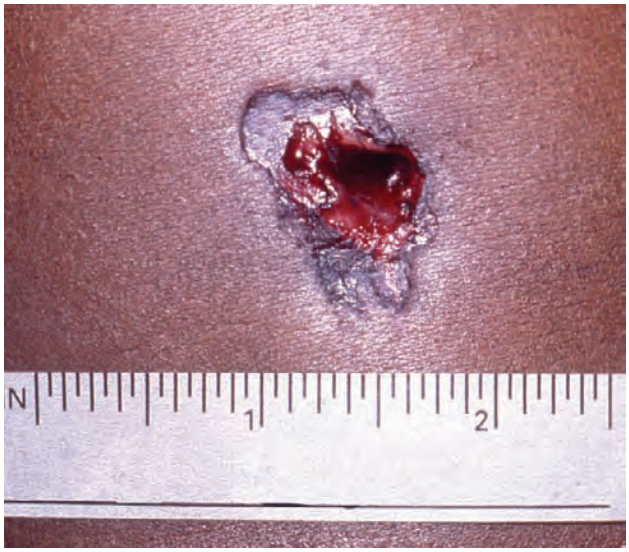

Atypical entrance gunshot wound caused by a bullet that was deformed by passing through an intermediate target. Note the large irregular abrasion caused by the deformed edges of the bullet's striking the skin surface.

Shored exit wounds.

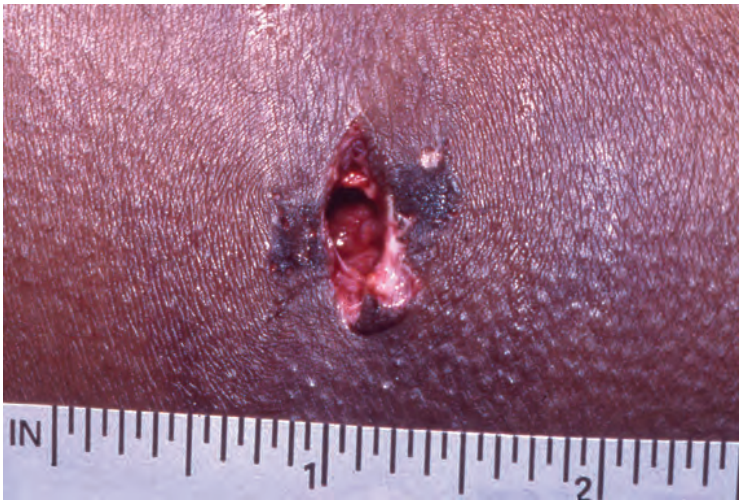

Irregular exit gunshot wound with adjacent abrasion due to its location on the body through a region of folded skin.

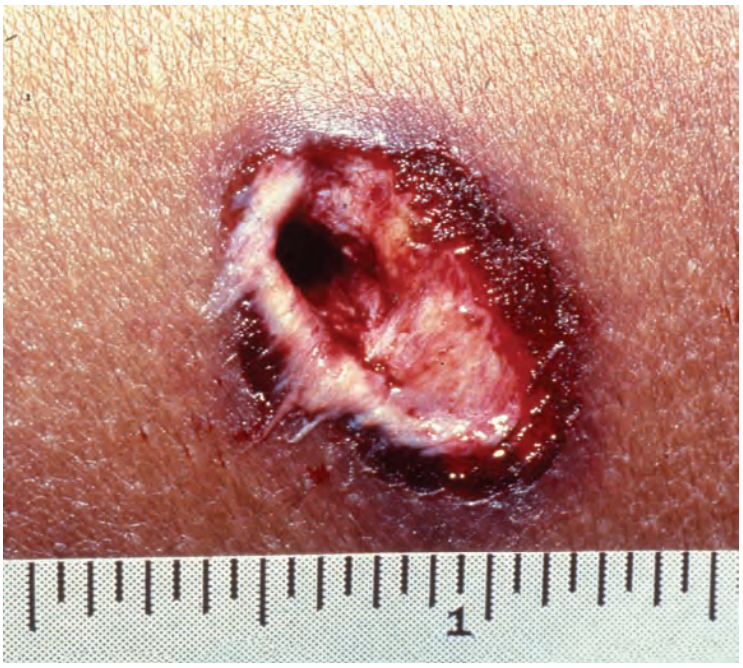

This distant-entrance gunshot wound has slight oval shape with an irregular margin of abrasion that is greatest at its inferior right aspect. This wound was produced by a ricochet bullet that first struck a brick wall. Part of the bullet had an irregular scratched surface. Also, this bullet most likely struck the body sideways.

Atypical gunshot wound. This is a re-entry gunshot wound produced from a bullet that was markedly deformed from striking a bone in another part of the body. Note the irregular nature of the perforation with the irregular adjacent abrasions.

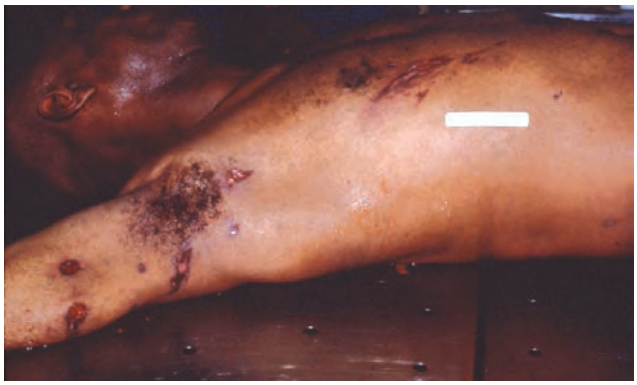

Atypical gunshot wounds from deformed and fragmented bullets that passed through an intermediate target, such as a car door and window, before entering and exiting the body. Note the markedly irregular nature of these injuries.

Multiple atypical entrance gunshot wounds produced by fragmented pieces of lead caused by .22 caliber nonjacketed bullets that struck concrete and fragmented before striking an individual who was lying on the sidewalk. Note the irregular nature of these injuries with superficial fragmented pieces of lead observed in several of the wound tracks.

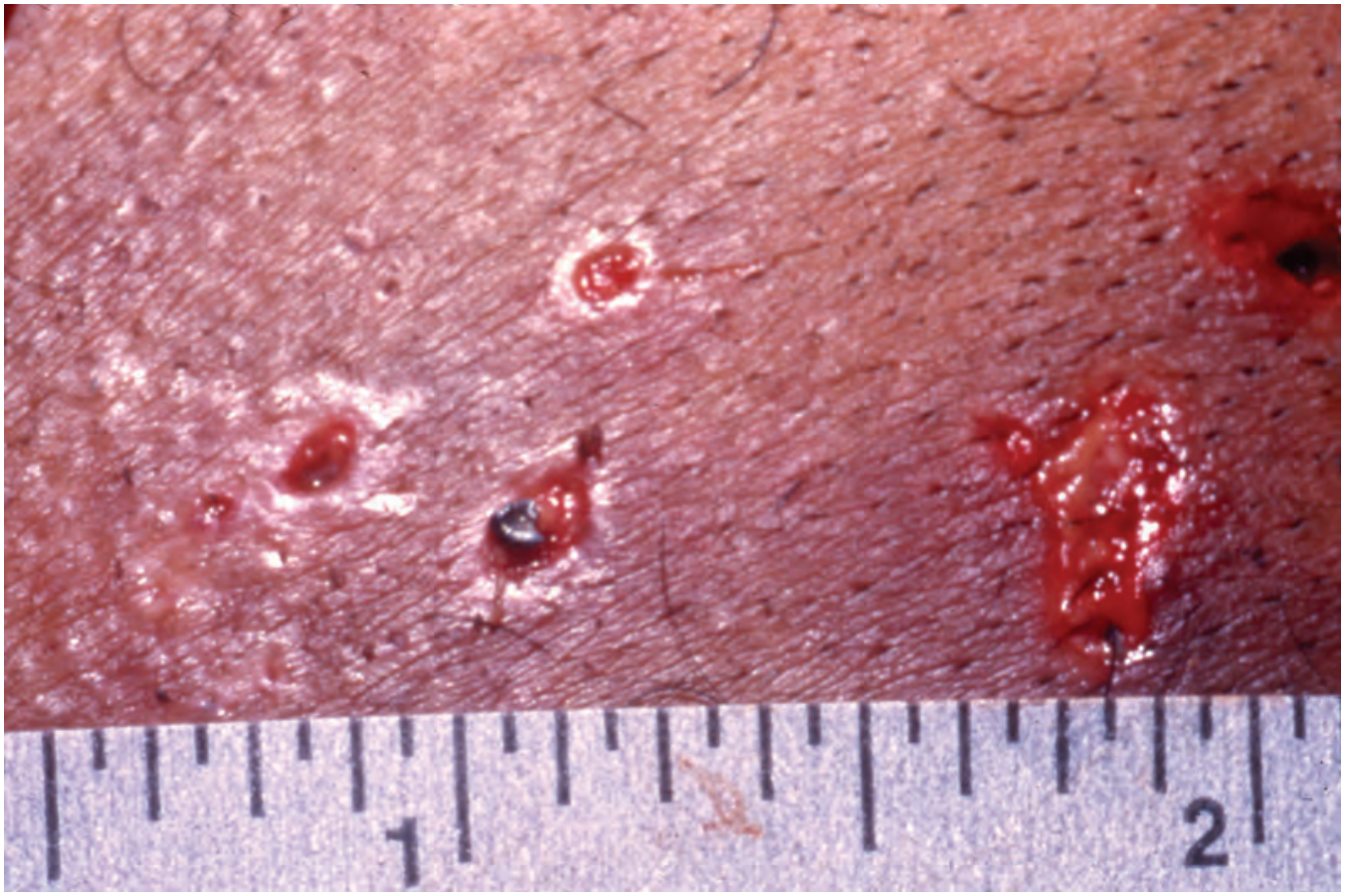

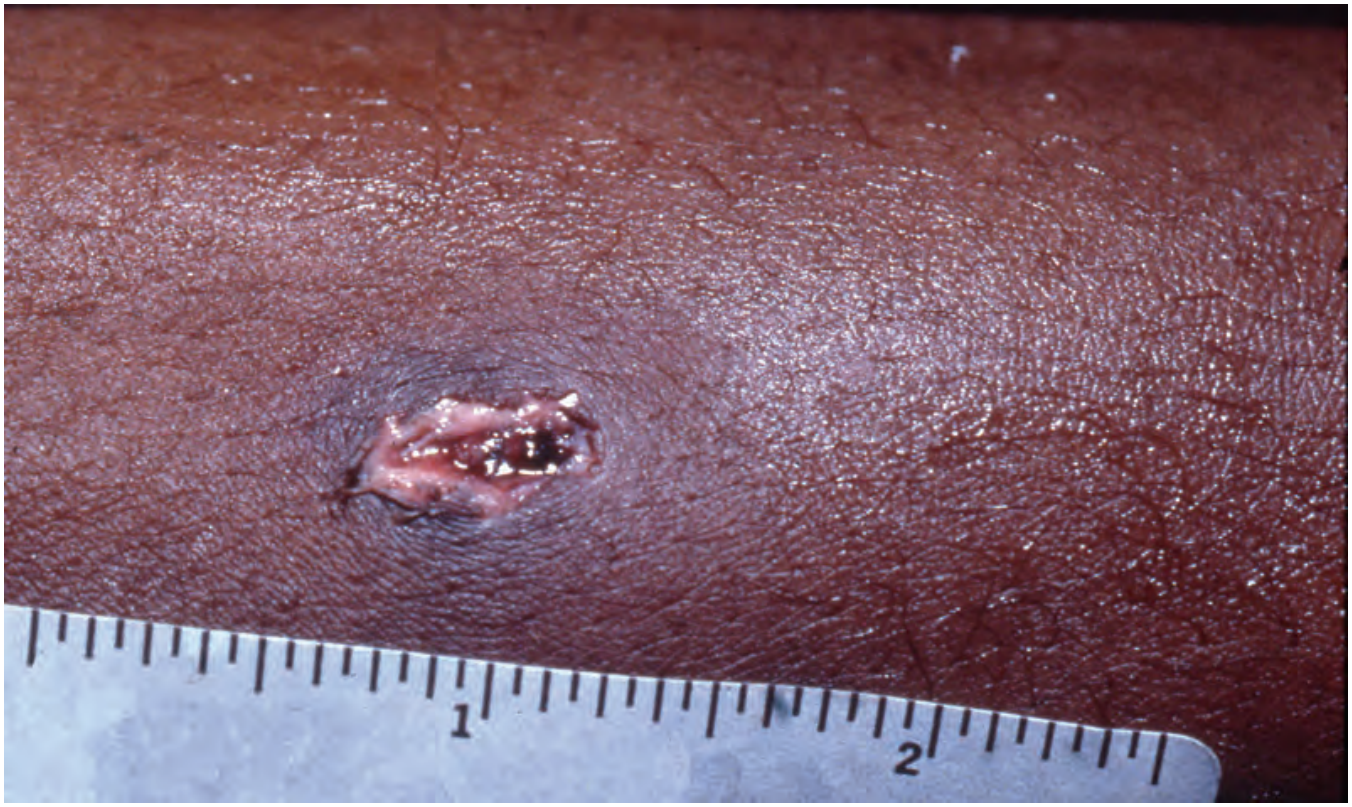

Several atypical injuries produced by fragmented nonjacketed lead bullets. The individual survived in the hospital for days. Note the healing margins and fragments of lead being pushed from the underlying soft tissue.

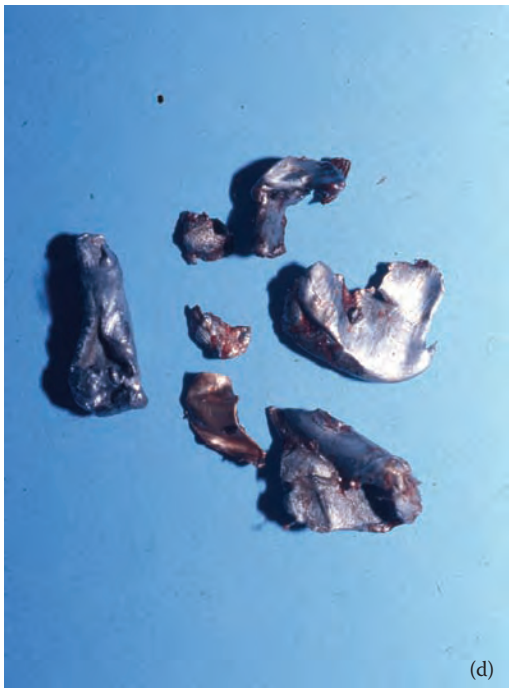

This child was shot multiple times. One of the bullets struck another object, fragmented, and produced this atypical entrance wound to his wrist (a). There were fragments of bullets retrieved from the decedent's jacket corresponding to this location (b). X-ray (c) shows multiple fragments of deformed metal retrieved just underneath the skin surface adjacent to the bones of the hand. The last image (d) shows all the bullet fragments retrieved from this individual's hand and wrist. There were no other wounds to this region of the body.

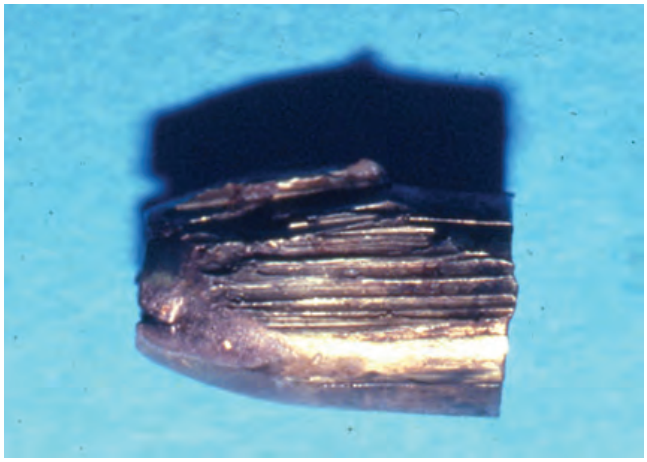

This ricochet bullet was retrieved within the abdomen of an individual with a perforated iliac artery. The bullet perforated the individual's shirt and did not strike bone while passing through the body.

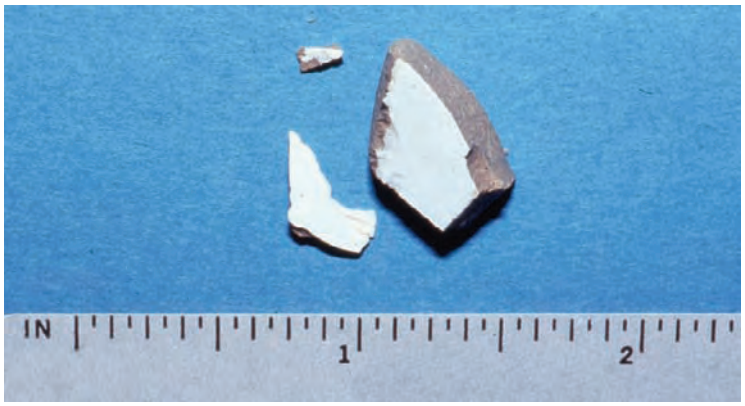

This bullet struck the ground, fragmented portions of tile, and became markedly deformed before ricocheting upward and striking the body of an individual already lying on the ground. The resulting gunshot wound was markedly atypical, producing irregular injuries to the body surface.

This deformed bullet was retrieved from an individual after it ricocheted off an intermediate target.

Intermediate-entrance gunshot wound with stippling. Note the irregular natural of the wound due to the location on the body. The bullet entered between the cartilage of the ear at a skin fold between the ear and the scalp. Note the elongated abrasion at the posterior ear due to the bullet's grazing the skin before entering the body.

Distant-entrance gunshot wound with an asymmetric margin of abrasion. This indicates that the bullet grazed the body at an upward trajectory before perforating the skin.

Graze gunshot wounds; (a) is more atypical.

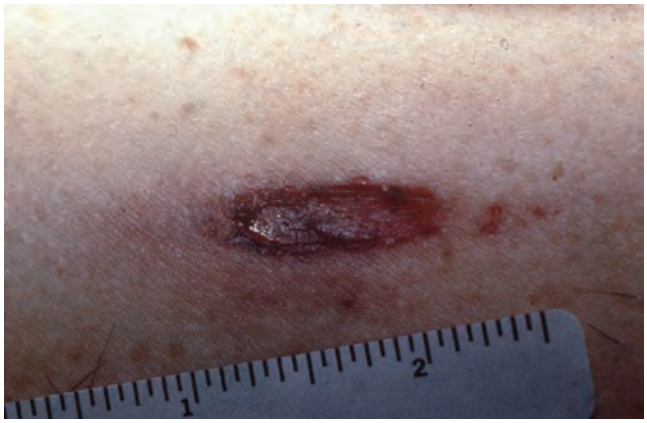

Graze gunshot wounds.

Tangential gunshot wound with a trajectory from the middle to distal aspect of the finger. Note the semicircular entrance defect at the right side of this wound overlying the proximal middle phalange.

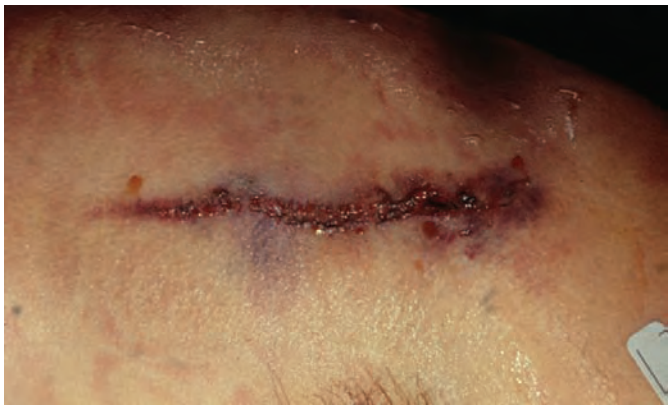

Tangential gunshot wounds to the arm and trunk. Note the skin tag formation at the wound margin produced as the bullet perforated underneath the skin's surface.

Tangential gunshot wound with skin tag formation. The direction of fire is from left to right.

Distant tangential gunshot wound. Tangential gunshot wounds are produced when the bullet strikes the body at a narrow angle, producing skin tag formation. Usually the bullet exits the body leaving an open wound through the skin's surface, connecting the path of entrance and exit perforations. These wounds are deeper than graze gunshot wounds. This wound is associated with bullet fragmentation and partial exit, which may occur when the bullet strikes the body at a surface directly adjacent to underlying bone. This is an example with bullet fragmentation, partial exit, and underlying keyhole deformity of the skull. Note the entrance side of this wound is at the anterior aspect. There is a semicircular margin of abrasion leading into this laceration. The direction of fire is from left to right.

Tangential gunshot wound of the scalp. The direction of fire is from right to left. Note the skin tag formation pointing away from the semicircular entrance site at the right side of the wound.

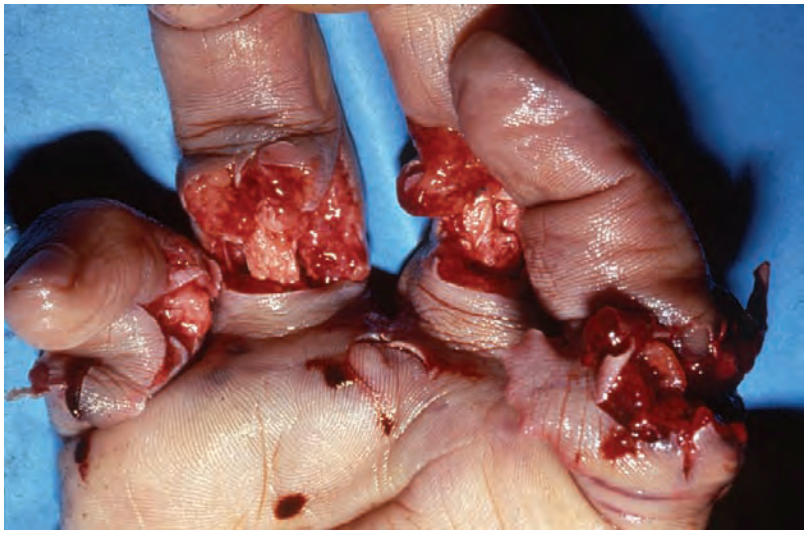

Note the tangential gunshot wound through the palmar surfaces of the fingers.

Tangential gunshot wound to the forehead. Part of this bullet entered the cranium and a portion of the bullet exited the body. There was an underlying keyhole deformity to the skull.

After this bullet perforated the fingers it produced a tangential gunshot wound to the nose before entering the body at the face.

This demonstrates an entrance and exit gunshot wound that passed superficially beneath the skin surface. The direction of fire is right to left. Note that the perforation at the right side has a more uniform oval shape with a more symmetric margin of abrasion. The exit component to this wound is more irregular.

This gunshot wound produced multiple tears and perforations to the surface of the skin as the bullet passed close to the underlying surface. Such injuries often occur in regions of the body where skin folds on itself, such as the inguinal, gluteal, and axillary regions.

This single gunshot wound passed through the superficial soft tissues of the chest, producing multiple perforations and tears to the surface. Note the irregular nature of the torn skin and irregular abrasions at the exit reentry site.

This is another example of an entry–exit–reentry wound where the bullet passed close to the underlying skin surface. Note the larger irregular abrasions connecting the exit and re-entry sites. Also note the dark discoloration due to drying.

This is an entrance and exit gunshot wound where the bullet passed very close to the under surface of the skin.

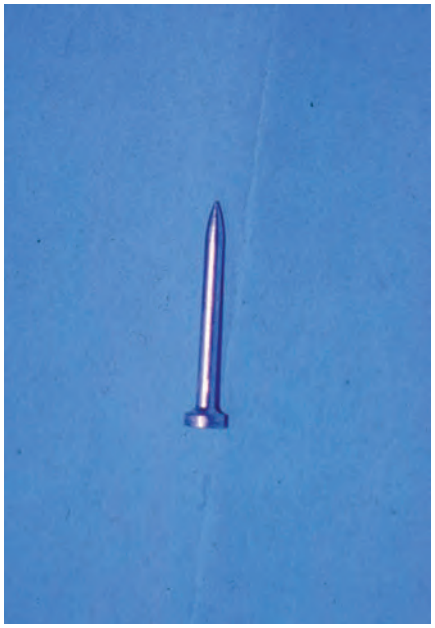

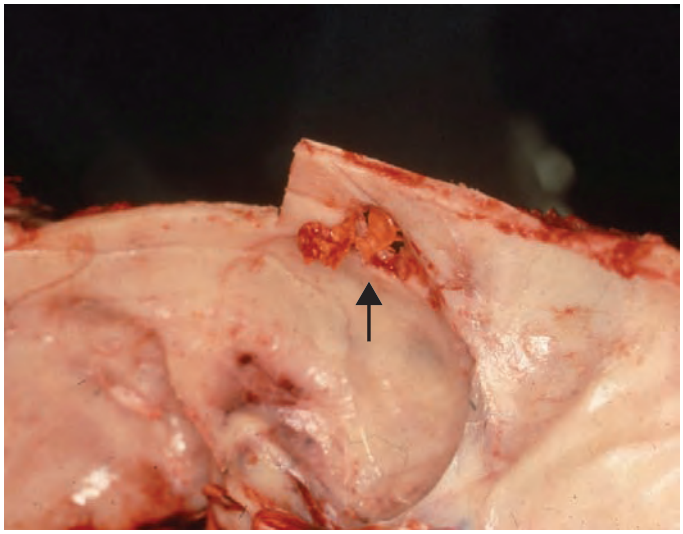
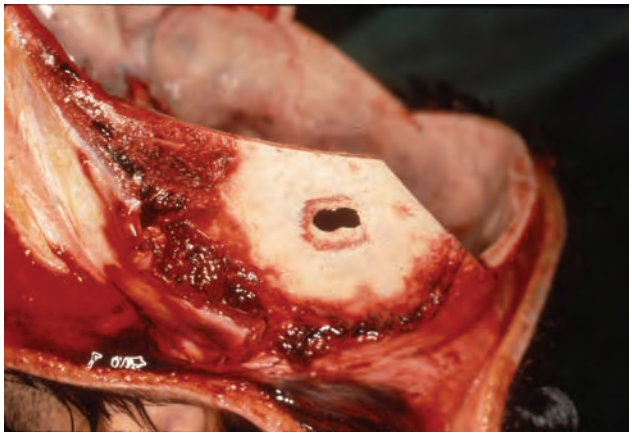

These images show different examples of nailgun wounds. Both of these cases were suicides. Both nailguns were similar to the one shown in the last image, which used gunpowder-loaded cartridges. Both entrance wounds consisted of circular perforations with symmetric margins of abrasion indistinguishable from typical entrance gunshot wounds. Note the orange plastic ejected into the entrance perforation of the temporal skull that is used in some nail guns to hold and steady the nail prior to discharge.

This individual was shot multiple times. Note the exit gunshot wound to the superior aspect of his middle chest (a). He was wearing a medallion on a string that was struck by the bullet as it exited the body, producing an imprint on the skin surface. The chest surface was most likely pressing against another object when the bullet exited. Figure (b) illustrating the medallion actually demonstrates it facing the wrong way. The medallion should be oriented with the inward curvature facing the chest. This was at first erroneously thought to be an entrance gunshot wound. Among other points arguing this to be an exit wound was the sternum fracture with bone splinters pointing in an outward anterior direction(d).

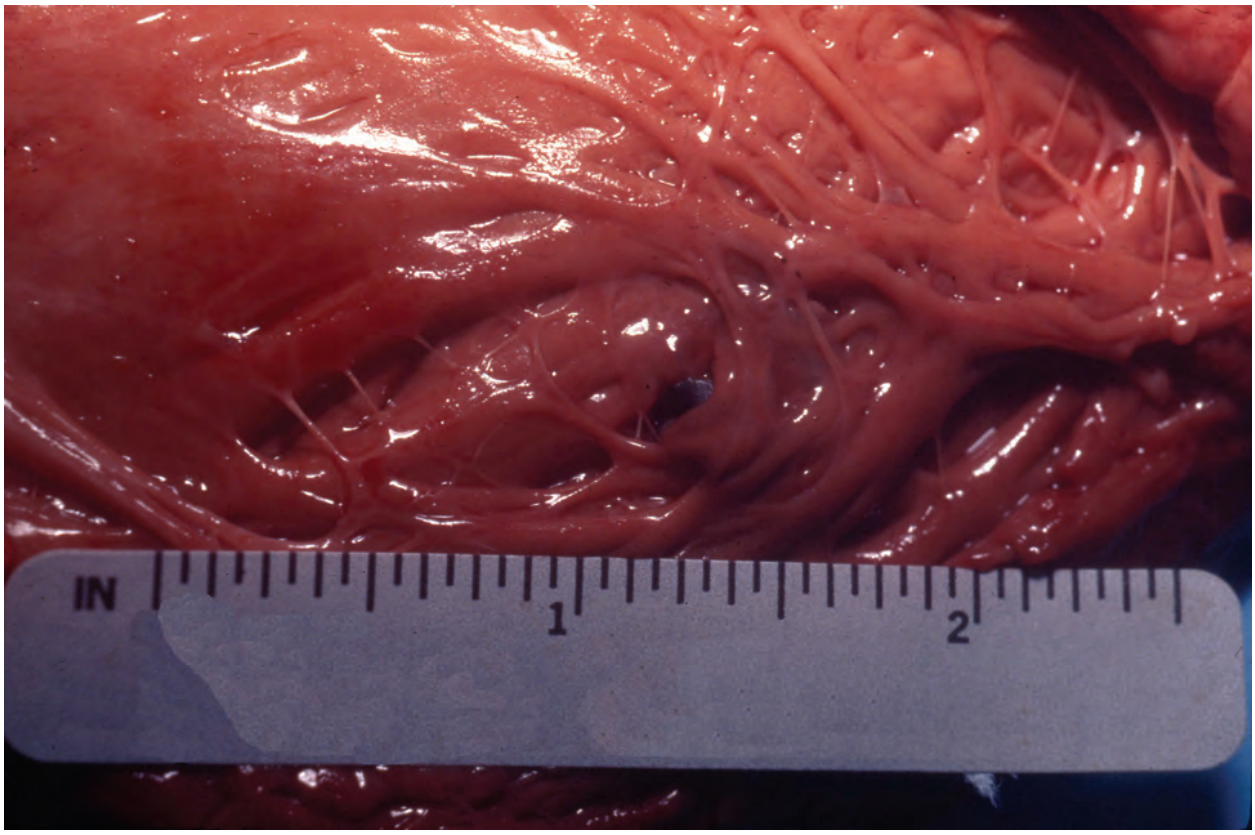

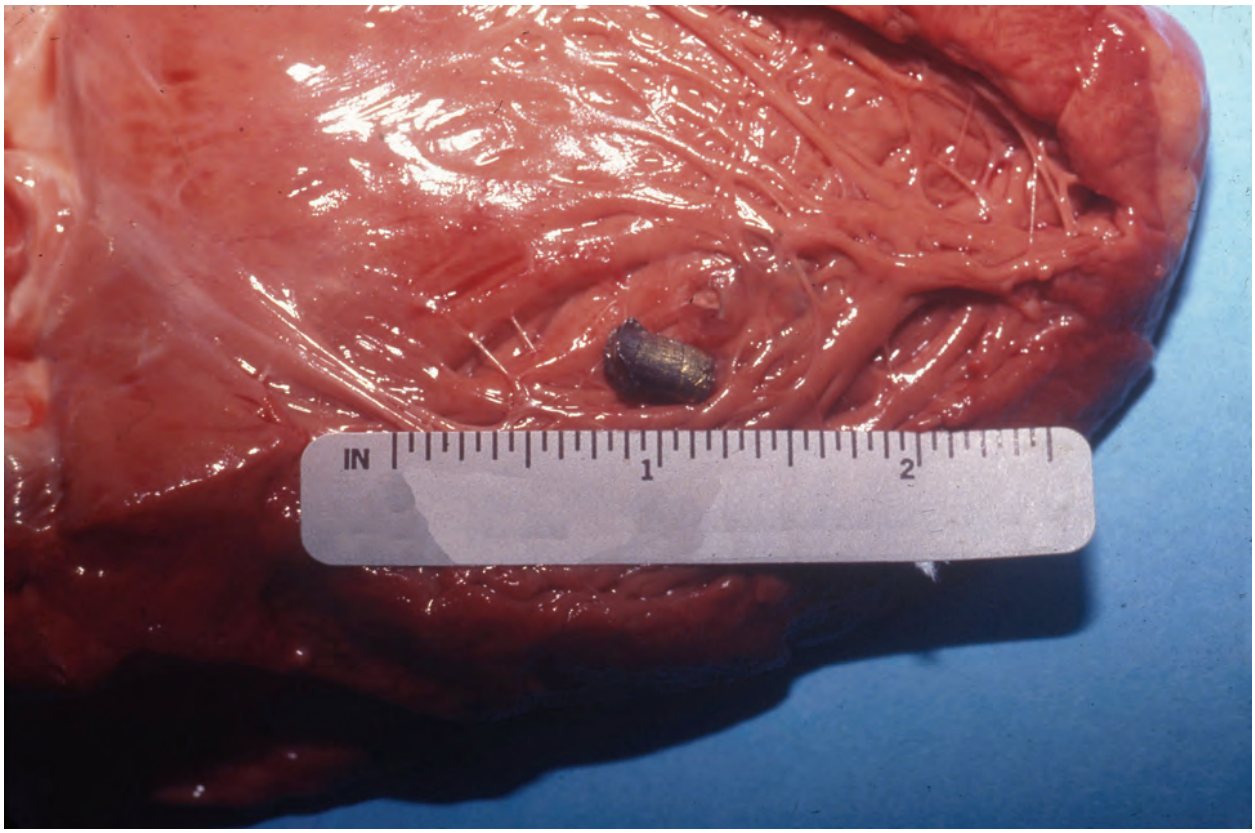

This individual was shot multiple times and survived in the hospital for approximately 2 weeks. One of the bullets entered the lung parenchyma and embolized to the heart, where it got wedged in papillary muscles of the left ventricle. He developed bronchopneumonia and his lung wound reopened, causing hemorrhage and death.

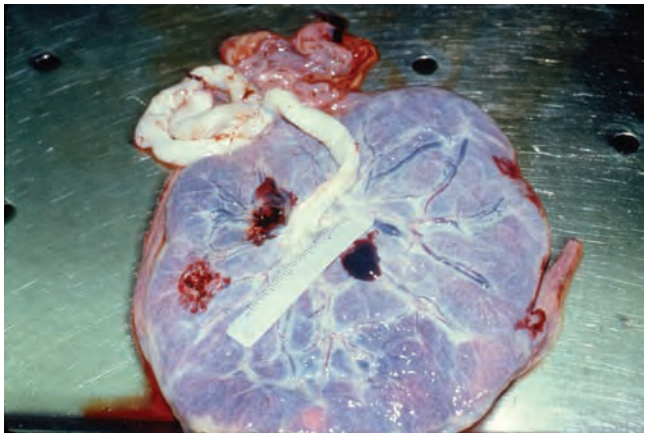

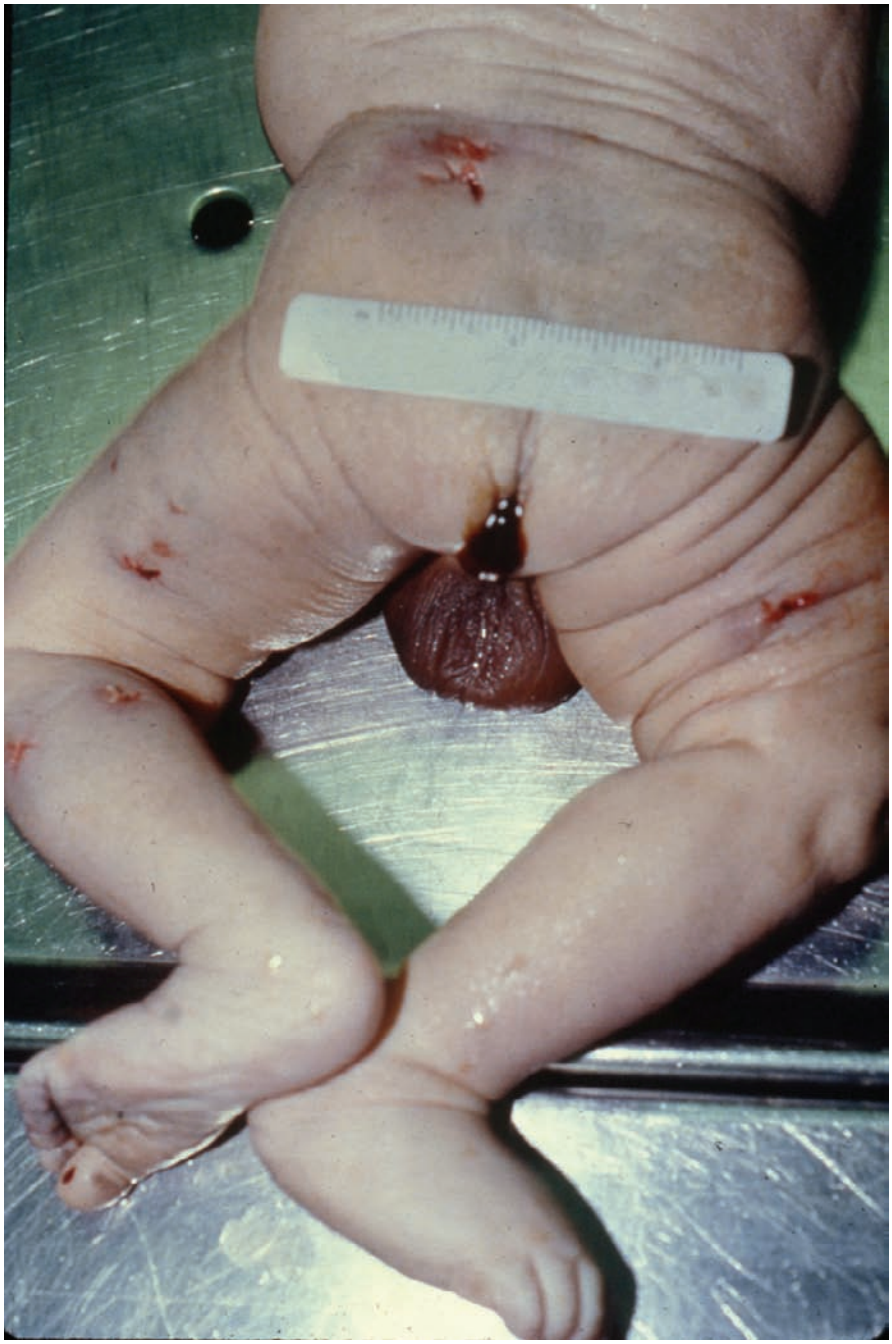

An otherwise healthy, full-term pregnant woman was shot multiple times in the abdomen. There were multiple perforating gunshot wounds to her fetus that were atypical due to the intermediate targets, including the mother, uteroplacental unit, and amniotic fluid. All the fetal wounds appeared as irregular lacerations with irregular abrasions. Direction of fire could not be determined.
Source for upper right and lower figures: Catanese C. and Gilmore K. A case report and brief analysis of fetal gunshot wound characteristics. *Journal of Forensic Science* 47:1067–069; 2002. Reprinted with permission from ASTM International.

This individual was shot in the face while wearing eyeglasses. Note the irregular nature of the entrance gunshot wound just beneath the eye (a) and the elongated abrasion extending across the cheek to the decedent's ear (b), corresponding to the eyeglass frame.

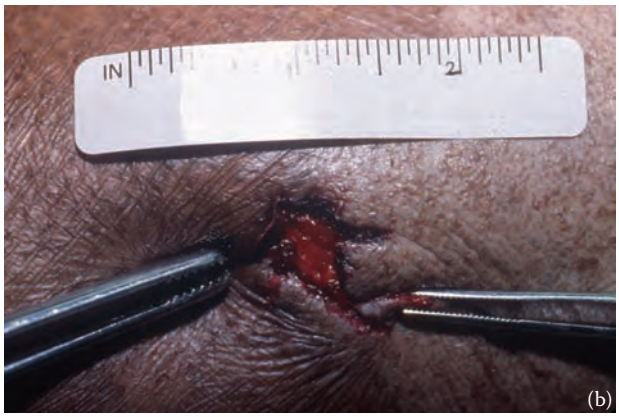

This individual was shot multiple times and had psoriatic skin disease. Note the irregular nature of this entrance gunshot wound through one of the plaques (a). The skin had small radiating lacerations and decreased abrasion similar to what is sometimes seen with gunshot wounds through thick skin found on the palms of the hands or soles of the feet (b).

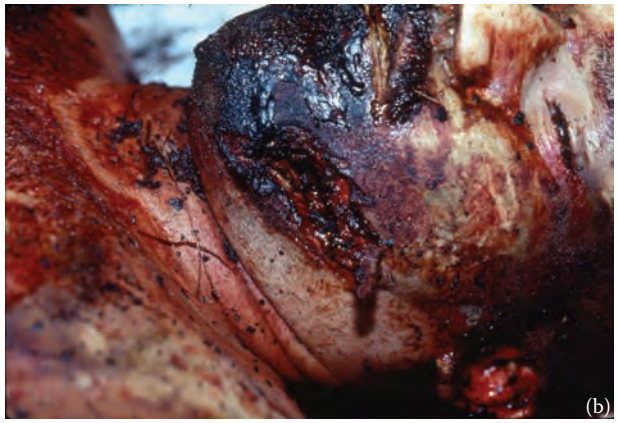

These are gunshot wounds to an individual who was set on fire to destroy evidence after being shot in his residence. His carbon monoxide blood level was negligible. Note the change in character of these injuries due to thermal effect. These thermal burns are postmortem.

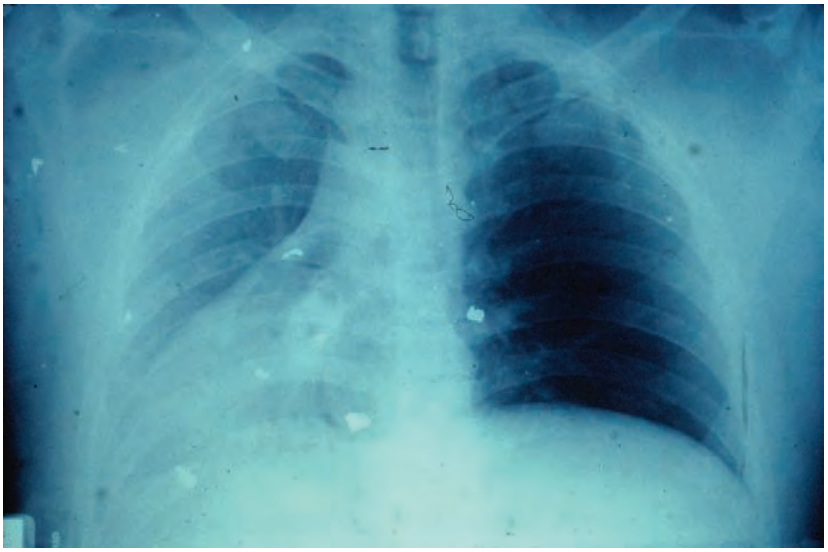

This individual was shot multiple times where the bullet struck bones and fragmented into multiple pieces. It is important to keep track of which bullets are associated with which gunshot wounds. There may be three different shooters and only one of the bullet wounds lethal. Linking the lethal bullet to a particular shooter may have different legal implications.

This individual was shot in the abdomen with a hunting rifle. Note the snowstorm effect of bullet lead fragmentation depicted on this x-ray.

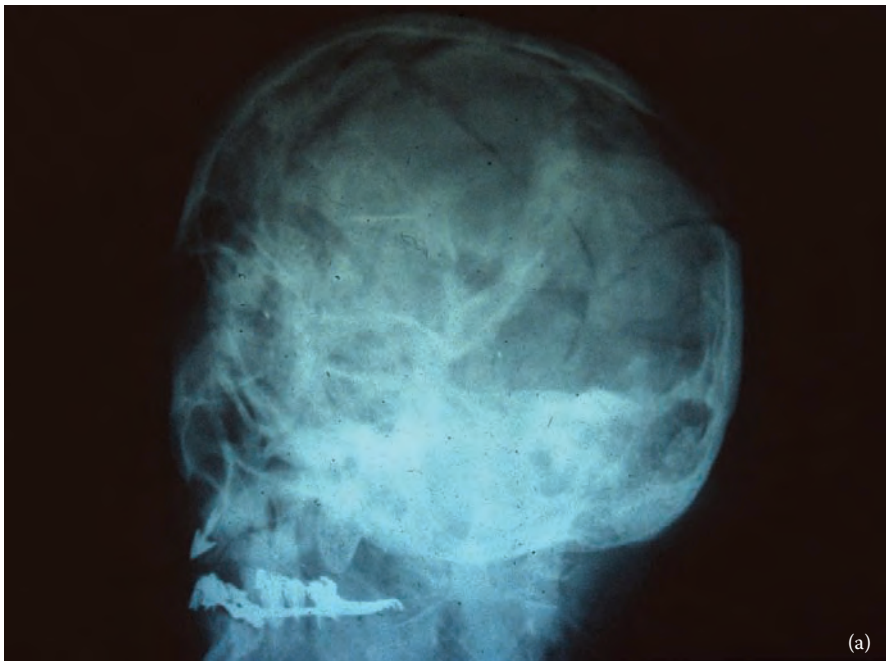

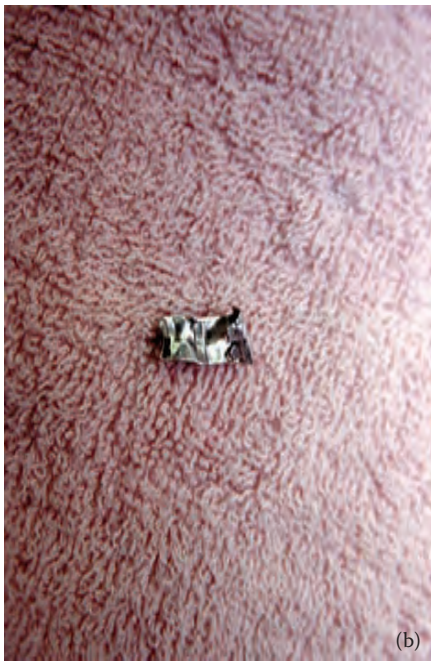

This individual (a) was shot with an aluminum-jacketed bullet. This fragment of aluminum (b) was retrieved within the wound track. Aluminum, being a relatively less dense metal, may not be apparent on x-ray, particularly when it is lying over a dense thick bone.

Note from this x-ray that one can tell the direction of travel. The bullet struck the humerus, fragmented into pieces, and then left a trail of metal fragments as it passed through the soft tissues before coming to rest.

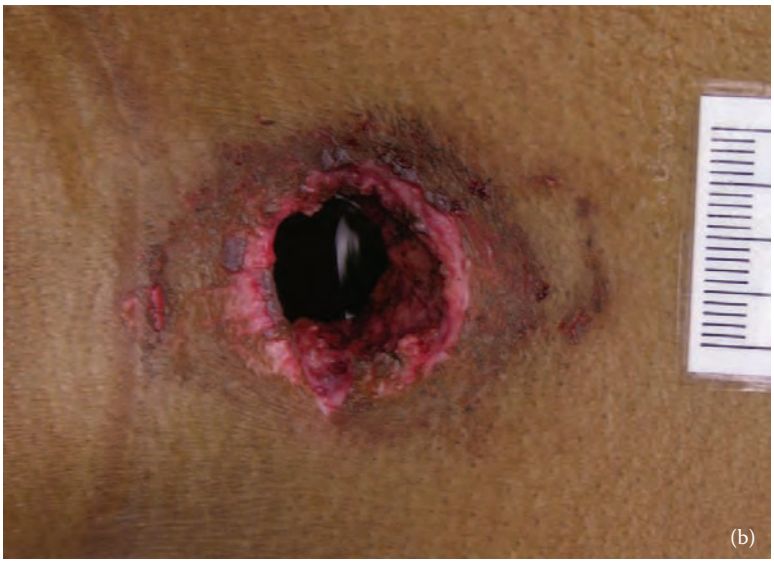

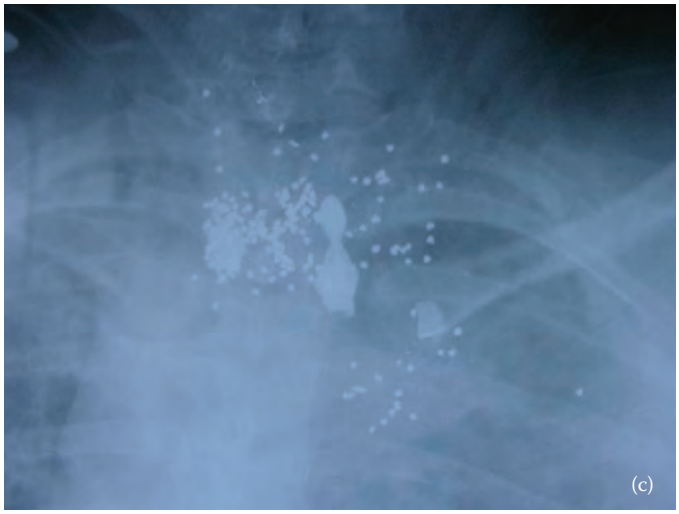

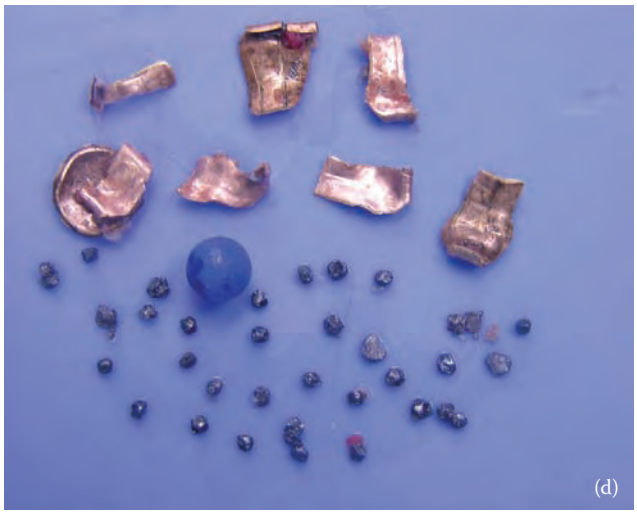

This is a contact gunshot wound to the chest with visible muzzle imprint abrasion (b). Note the blue-tipped Teflon plug Glazer safety ammunition visible within the revolver chamber (a). Note the yellow metal jacket, Teflon plug, and multiple lead pellets seen on x-ray (c), and demonstrated after removal from the body (d).

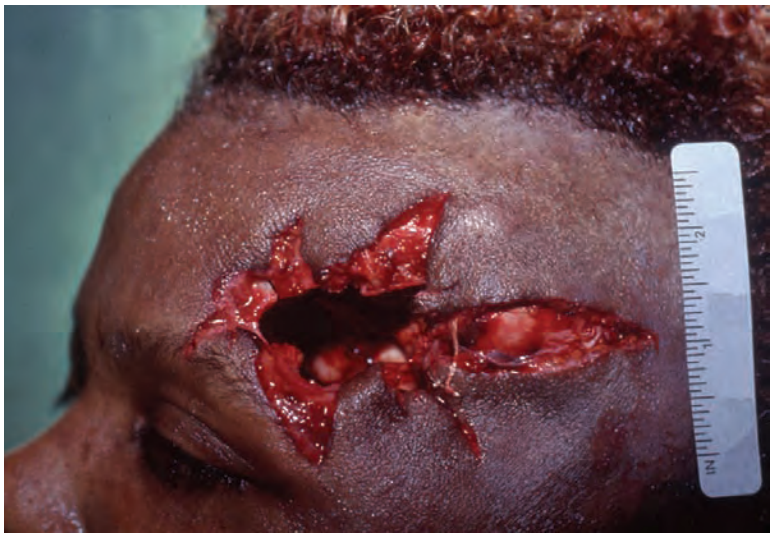

Tangential distant gunshot wound produced by Mag-safe ammunition (prefragmented).

This diagram shows a comparison of the components of Mag-safe ammunition and Glaser safety slug.

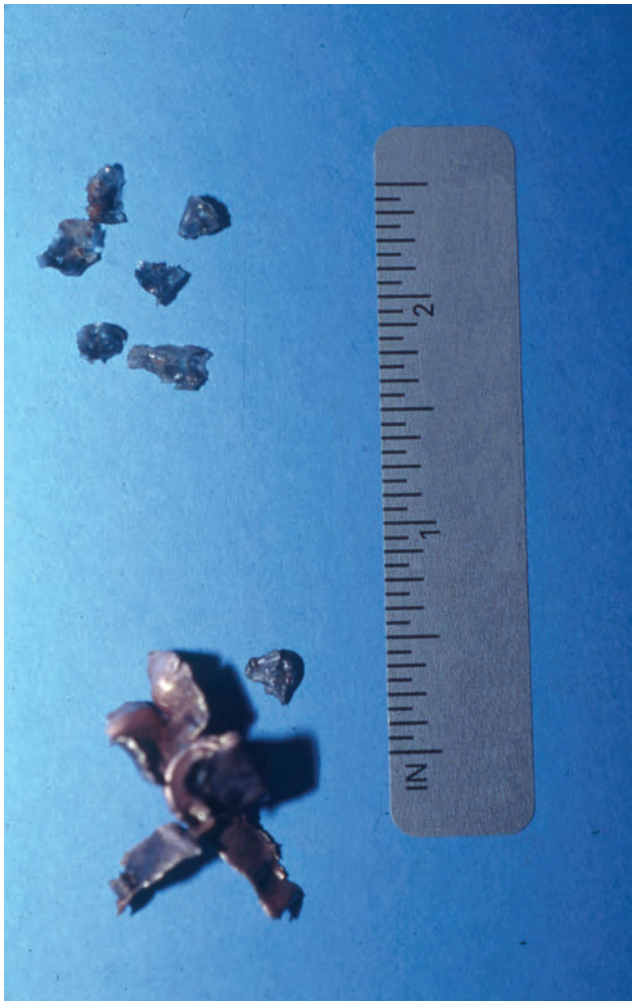

This x-ray is taken from an individual shot with Mag-safe ammunition. Note the metal jacket, gray to blue epoxy fragments, and gray metal pellets.

This demonstrates a bullet retrieved from an individual who was shot approximately 1 day prior to autopsy. Note the shiny surface of the bullet, signifying no significant oxidation.

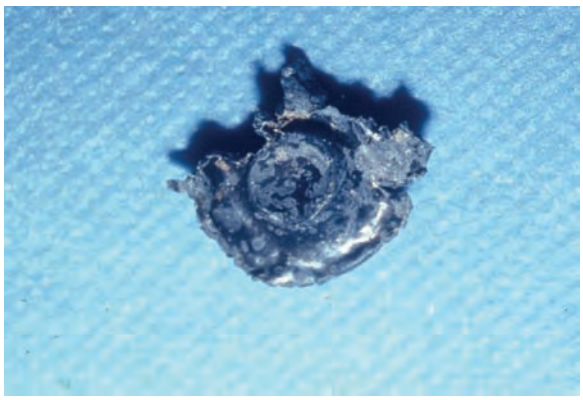

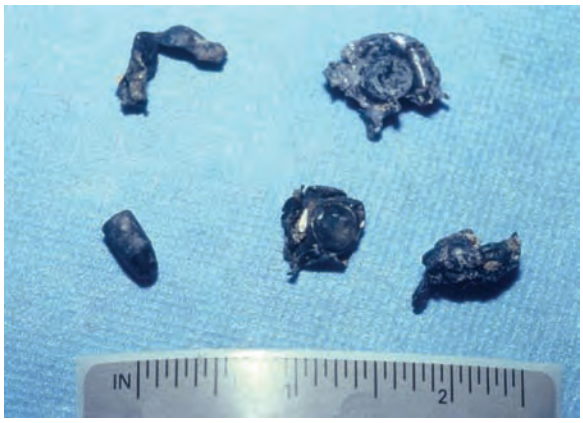

These are multiple weathered fragments of bullet retrieved from underneath and within the body of a partially skeletonized individual who was shot and thrown down a well shaft several years earlier. Note the irregular weathering marks at the surface of the bullets due to erosion.

This demonstrates a bullet retrieved from an individual who was shot over a year before. This bullet remained in his body until it was retrieved during autopsy after somebody else shot and killed him. Note the dull oxidized surface on this old bullet.

The homicide victim who owned this dog died of multiple perforating gunshot wounds. The dog sustained penetrating gunshot wounds during the attack and was autopsied to retrieve ballistic evidence.

This is a recent gunshot wound to an individual who was shot through the subcutaneous and fatty tissues of the gluteal region and thigh. Note the curvature of the wound track. This demonstrates how wound tracks may change when the body is laid on a flat autopsy table. One must keep this in mind when formulating bullet trajectories with reference to standard anatomic planes. This information may later be used as a reference to help explain possible body positions during the actual shooting. Interpreting wound track directions can become complicated when there are multiple gunshot wounds in proximity, particularly when the individual was shot while curled up in a fetal position and then later examined while spread out on an autopsy table. Directions should be stated with reference to standard anatomic planes.

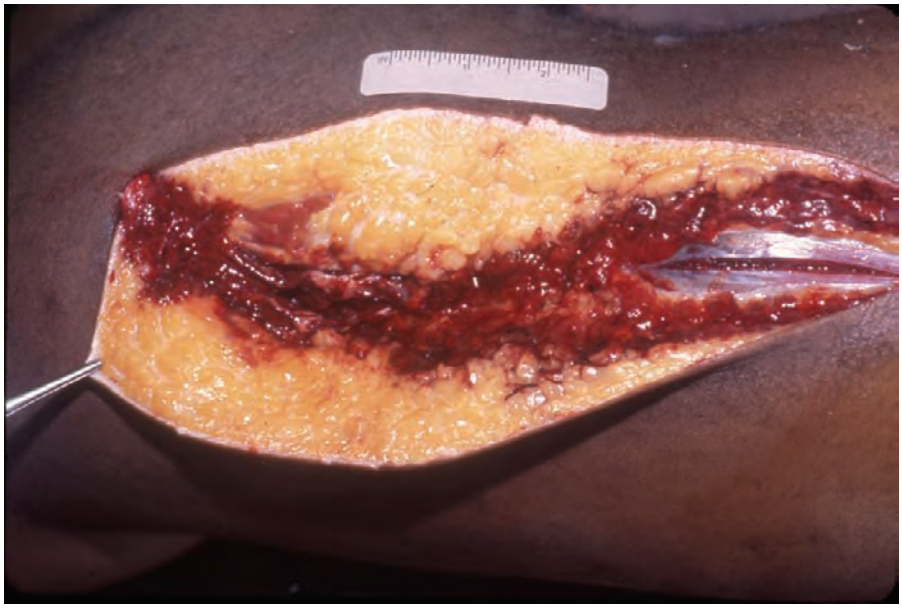

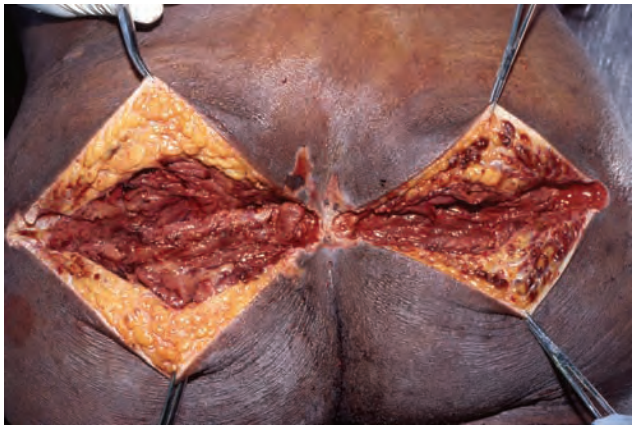

This is a healing wound track through the subcutaneous and fatty tissues of the gluteal region. Note the dull granular nature due to healing and the presence of granulation tissue.

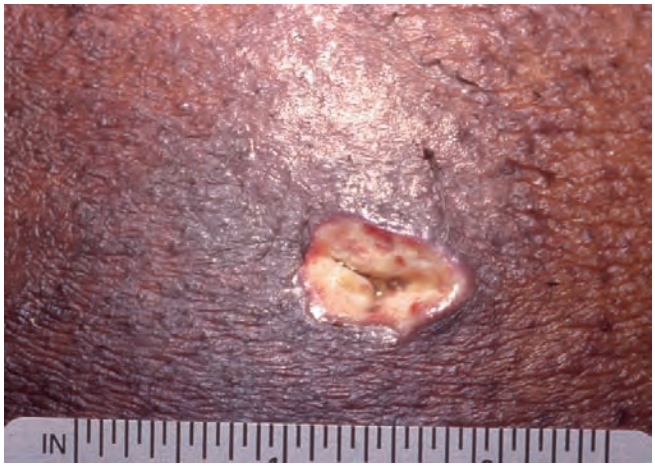

This is a healing gunshot wound that is approximately a week old.

Atypical healing gunshot wounds with fragmented pieces of lead.

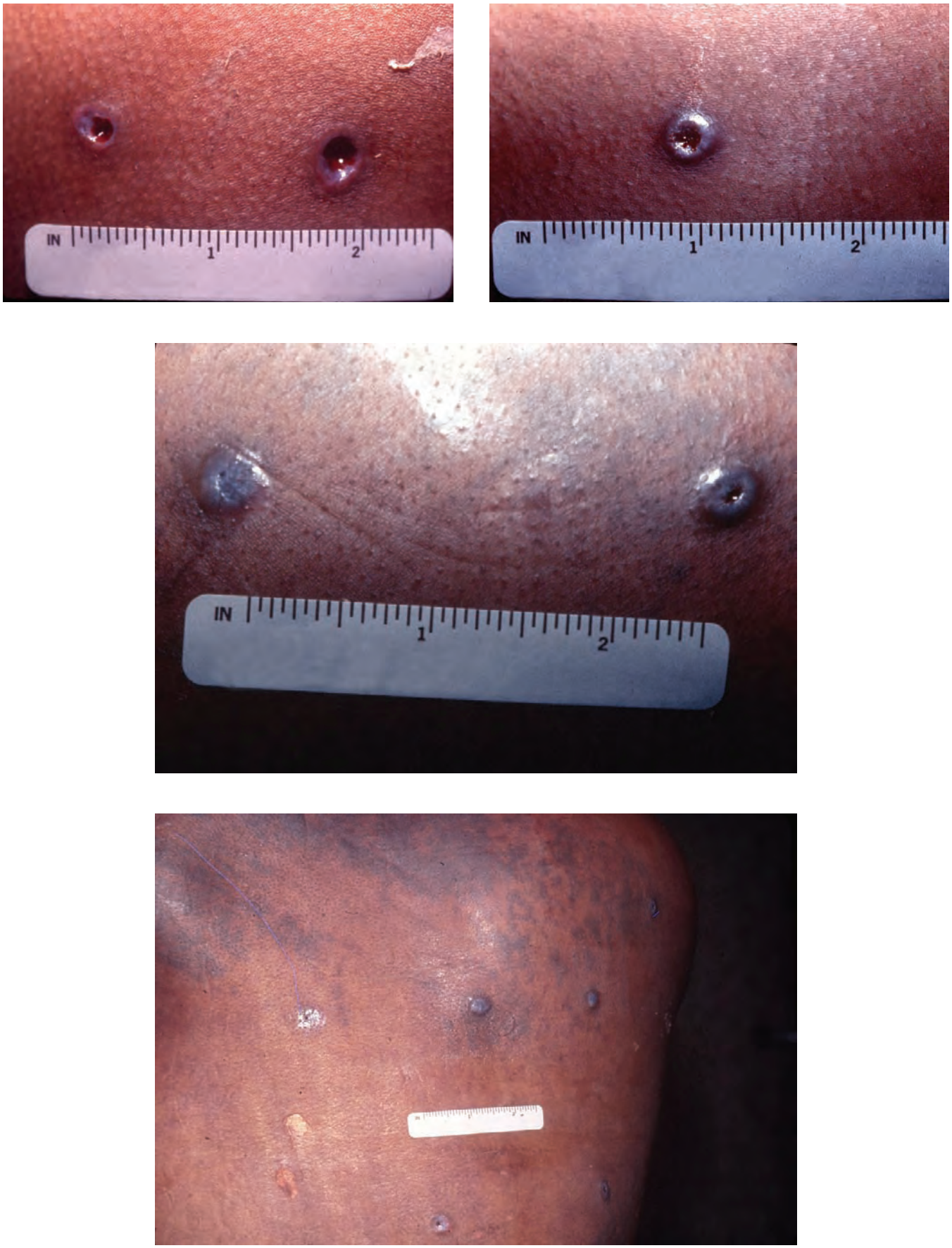

Varying stages of healing gunshot wounds.

This case involves a suicidal tight-contact intraoral .44-caliber rifle wound with extensive fractures and lacerations to the head. Note the gaping injury at the top of the head with the empty cranial vault. The brain was ejected almost entirely intact due to expanding muzzle gases.

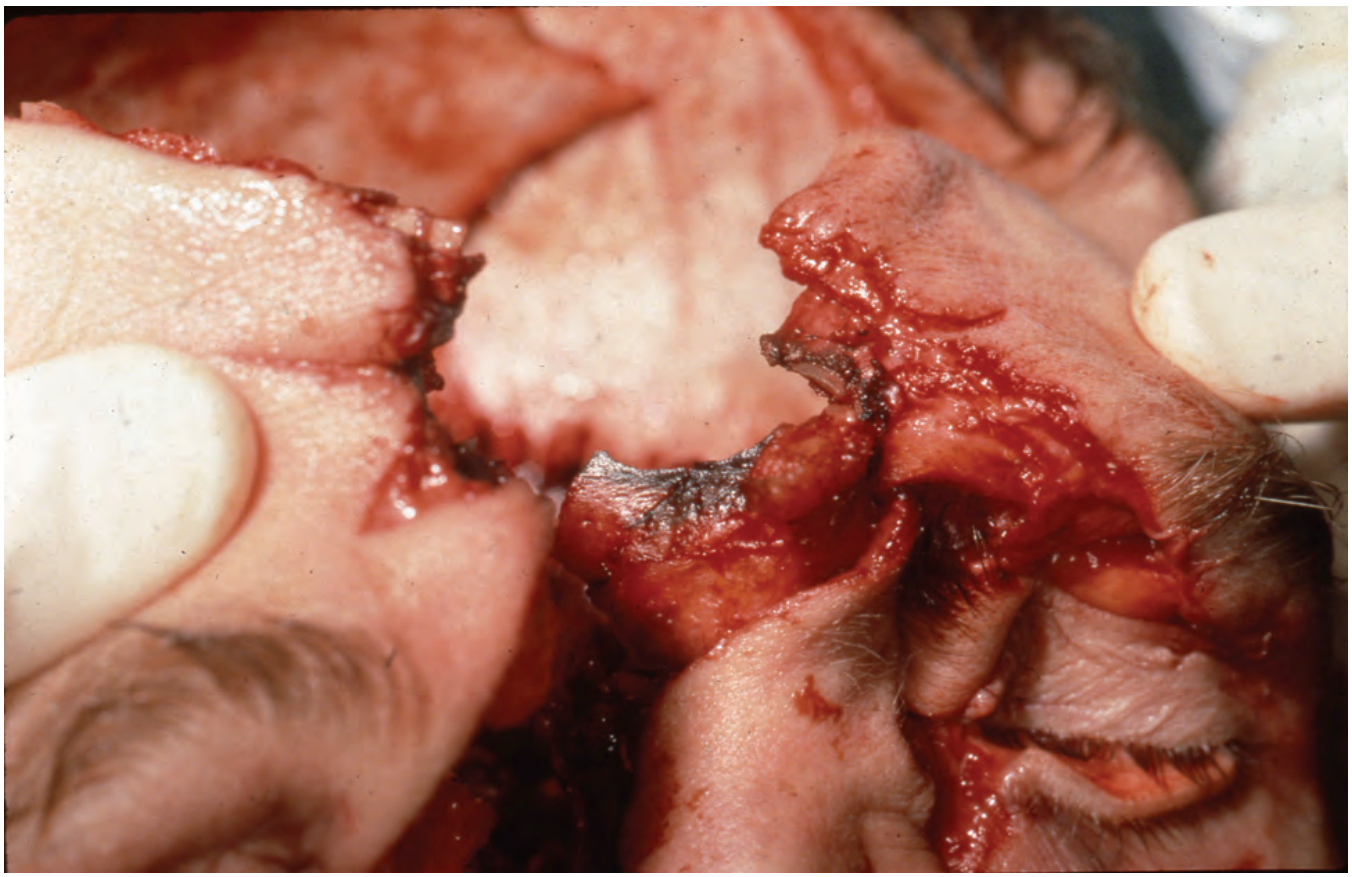

This individual died from a self-inflicted contact 16-gauge buckshot shotgun wound to his forehead. His cerebral hemispheres were ejected from the cranium almost completely intact and were discovered behind the body. Note reconstruction of this wound produces an obvious circular perforation with soot at the individual's forehead. Initially, at the scene investigation, it was thought to have been an intraoral shotgun wound, many of which are associated with stretching and lacerations of the lips and mouth region.

This individual died of a self-inflicted, contact, 12-gauge shotgun wound to the superior aspect of his neck underneath his chin. Note approximation of the wound margins produces a roughly circular perforation with soot. There were extensive lacerations to the face and head. Note the extensive stretch marks at the decedent's face. Note the blood spatter pattern and gunpowder residue on the hand.

These are contact shotgun wounds to the chest. Note the roughly symmetric margin of abrasion with small amounts of soot. More soot was observed within the wound track. Note in (a) another encircling abraded ring corresponding to where the muzzle contacted the body when the gun was discharged. Note in (b–c) there is an encircling vague pink discoloration from carbon monoxide and nitrites in the burning gunpowder that reacted with the underlying muscle and blood, producing this red discoloration.

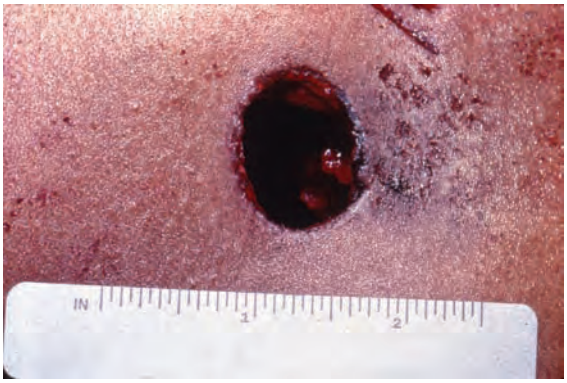

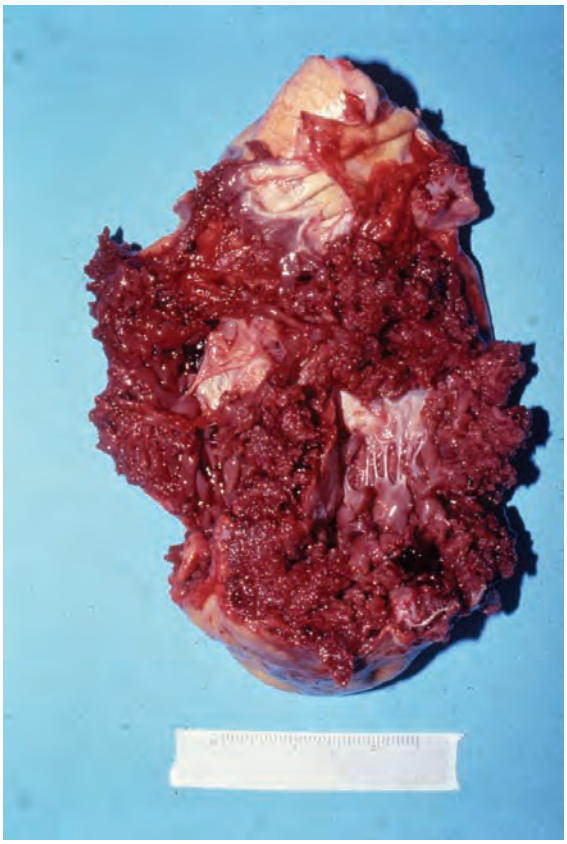

This is another example of a shotgun-slug wound. Note the extensive fragmentation and laceration to the heart. The last image shows the deformed lead shotgun slug with wadding retrieved from the wound track.

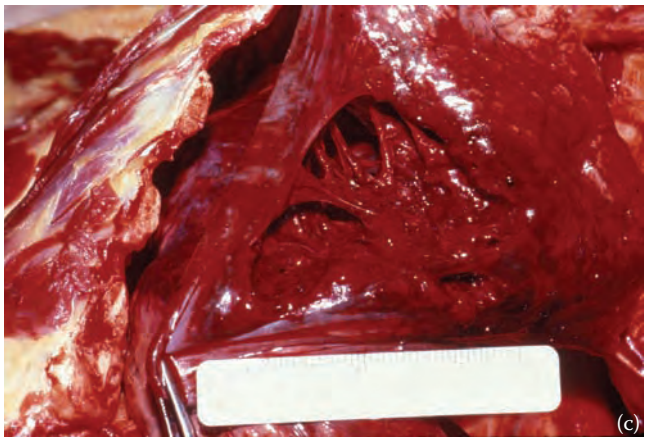

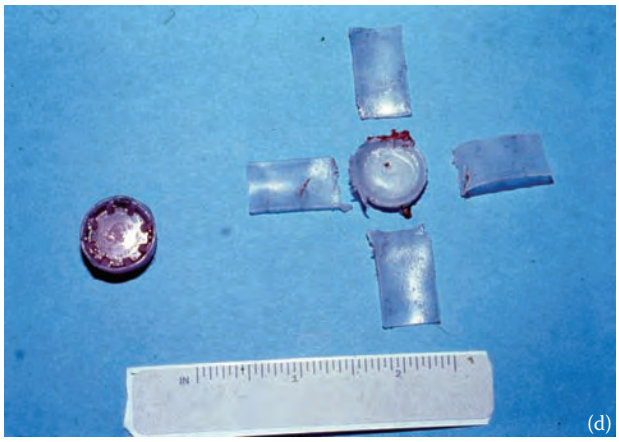

This is a .12-gauge shotgun wound through clothing showing few stippling defects and a larger abrasion adjacent to the entrance defect (a and b). The power piston was retrieved within the wound track (d). Note the abrasions adjacent to the entrance wound due to power piston impact (b). There was extensive fragmentation of the lung (c). Note the multiple exit buckshot-pellet wounds to the decedent's back (e).

Distant birdshot-shotgun wounds.

This birdshot-shotgun wound to the forearm is estimated at a range of approximately 3 to 4 feet.

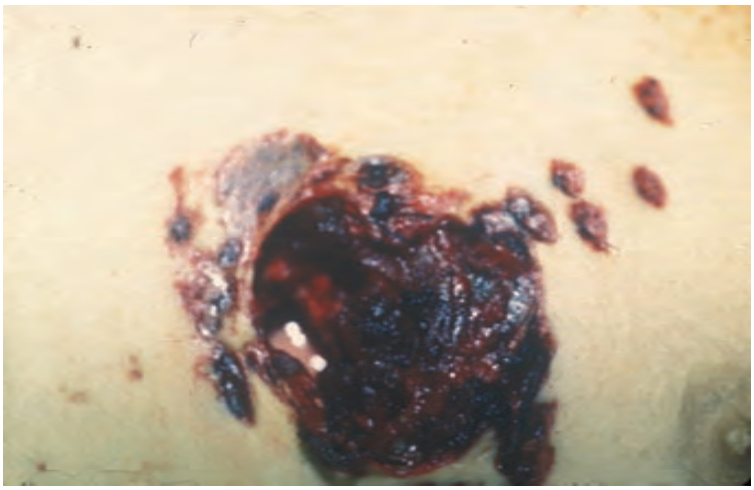

Birdshot shotgun wound with an estimated range of fire of 4 to 5 feet.

Birdshot shotgun wound with an estimated range of fire of 6 to 8 feet.

Birdshot with power piston type cup from a shotgun shell.

This individual was shot multiple times with shotgun slugs. Note the large circular perforations with margins of abrasions typical for entrance handgun wounds (a). Note the elongated margin of abrasion at the middle wound (b), indicating the slug struck the body on an angle, traveling in a left-to-right direction.

Entrance shotgun wound with an estimated range of fire of 3 to 4 feet based on the scalloped appearance of the wound margins.

Mini-14 Ruger rifle.

Flash suppressor.

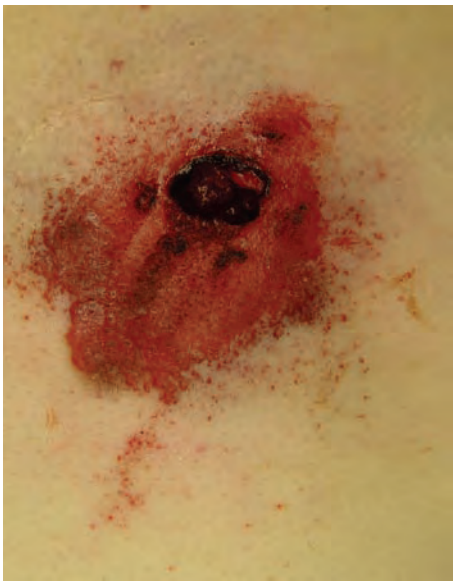

Contact Ruger mini-14 gunshot wound with a .223 bullet. Note the flash suppressor burns with soot deposition.

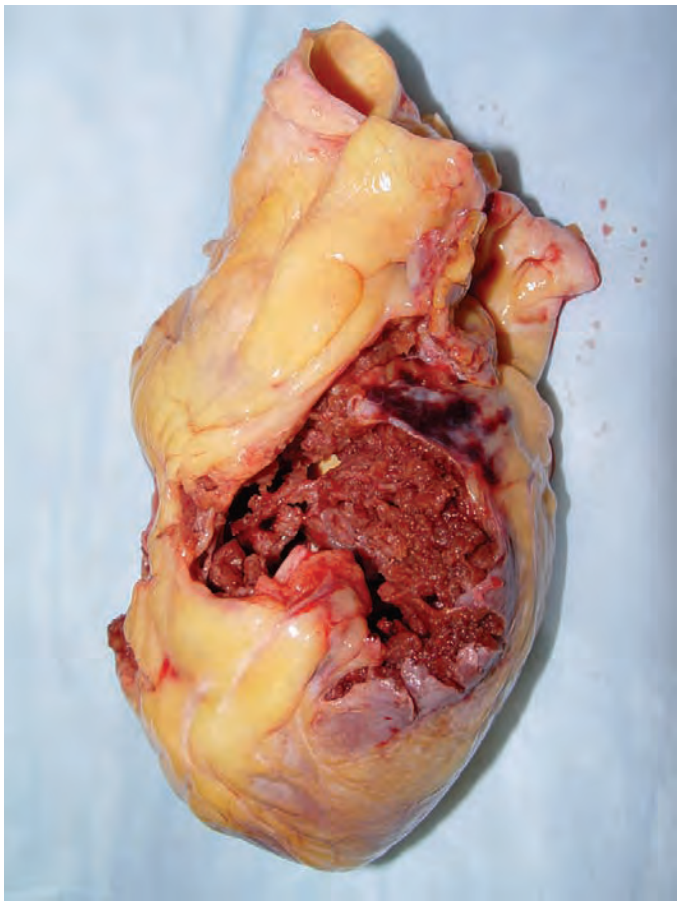

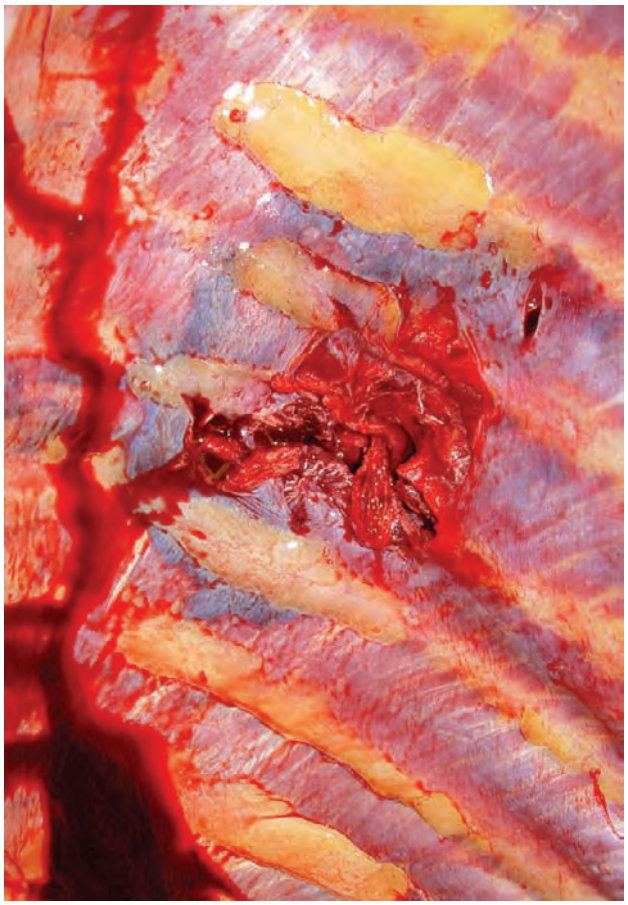

Internal injuries caused by a .223-caliber high-velocity military bullet.

Entrance gunshot wound to the roof of the mouth in the skeletonized remains of this individual with a history of depression. The remains were found with a handgun.

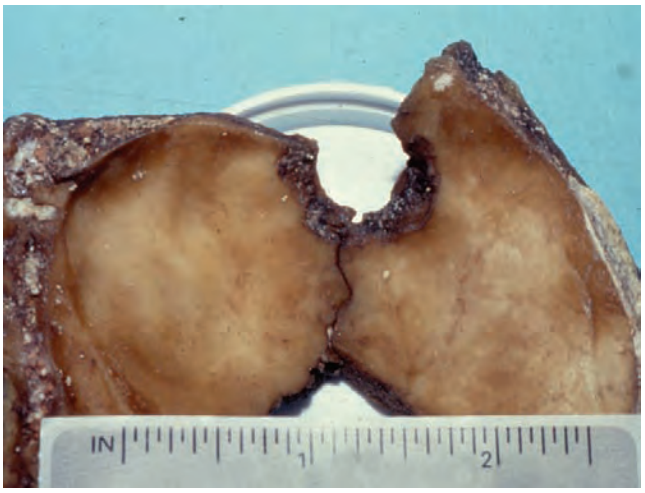

This is the inner aspect of a fractured portion of skull with a roughly circular perforation and internal beveling visible at the inner aspect of the skull indicating this to be an entrance gunshot wound.

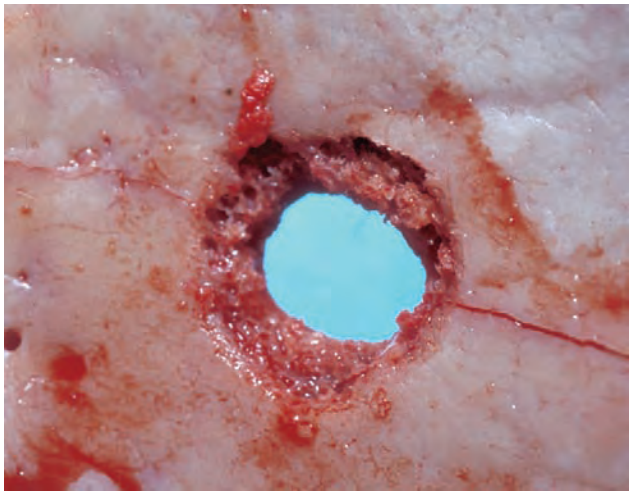

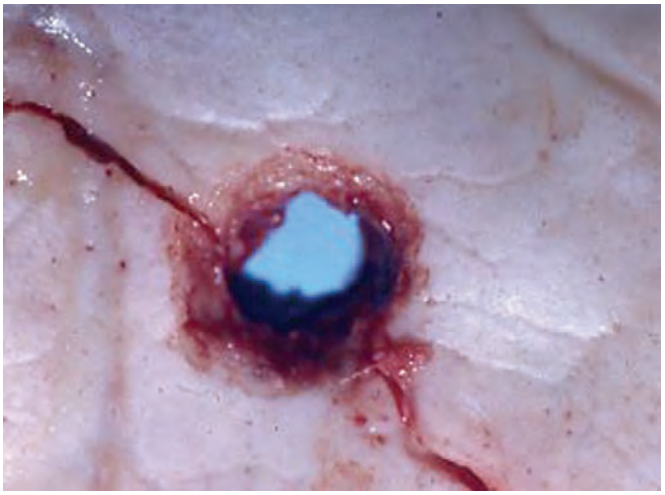

Entrance gunshot wounds to the head demonstrating internal beveling of the inner calvaria.

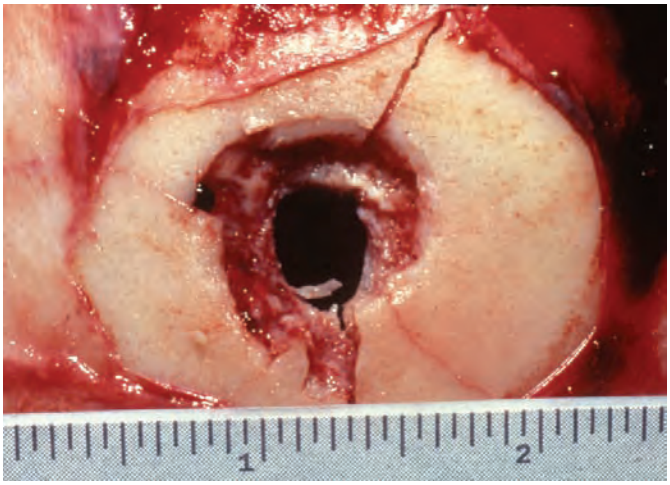

Exit gunshot wound through the skull. This image demonstrates the surface of the skull with external beveling typical for an exit gunshot wound.

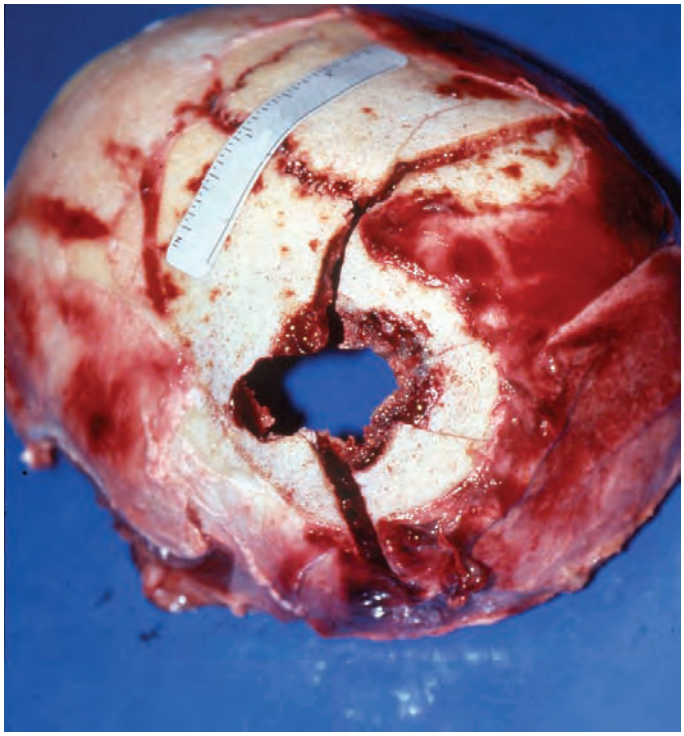

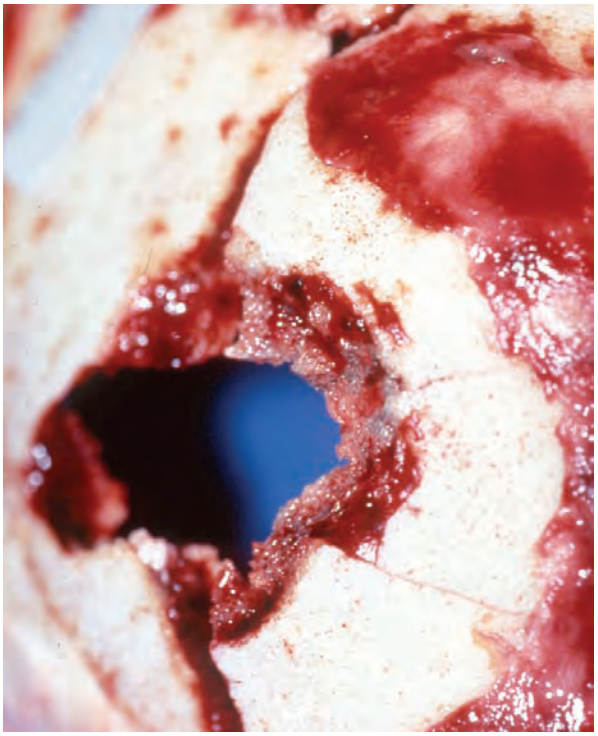

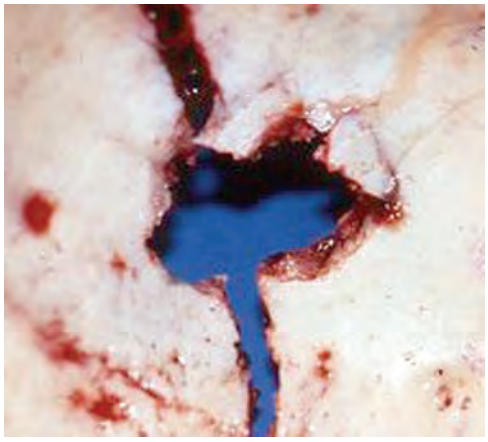

These defects at the external surface of these skulls were produced by an entrance tangential gunshot wound. This demonstrates a keyhole deformity with both internal and external beveling. The bullet typically strikes the bone tangentially, producing internal beveling at the entrance side. Then it will often fragment and partially exit the body. The exit side will have external beveling.

Burns

9

CHARLES A. CATANESE AND GERARD CATANESE

Introduction

Burns may occur following exposure to heat (thermal burns), electricity, chemicals, or radiation.

First-degree burns are the most superficial and involve the epidermis. They appear as a red discoloration of the skin. An example of this would be sunburn without blister formation. S*econd-degree burns* have deeper penetration of injury involving the epidermis and dermis causing blister formation. An example of this may occur when one touches a hot pot on a stove and pulls the hand away quickly. *Third-degree burns* are full thickness, involving the epidermis, dermis, and subcutaneous layer, and appear as collapsed blisters with skin sloughing and red-to-brown discoloration. *Fourth-degree burns* have even deeper penetration of damage, often with charring and exposed underlying tissue including bone. The bones may be fractured from intense heat and the internal organs may have a firm, discolored, and shrunken appearance.

Burns following exposure to heat are *thermal burns,* which may occur following contact with hot liquid or fire. Children may be scalded while left unattended in a bath tub. They may inadvertently hit the hot water knob to increase hot water flow. They may jump into a bath tub full of very hot water, scalding their feet. One must exercise great care in evaluating these cases as they may be a result of abuse or neglect. A child with both feet scalded with scarring around a shoe pattern is more likely the result of abuse or neglect. A child who is waiting to take a bath will not usually have shoes on and will most likely step one foot into the tub at a time. The burns will be on one foot but not usually both. Both feet scalded at once may indicate dunking into hot water as punishment.

If an individual dies and is placed in a tub full of hot water after death, the body will develop thermal injury more readily than if a living body was placed in the same water. A living body can counteract heat injury by vasodilation and circulating the heat to the body's core away from the surface, thus providing some protection to the skin. A dead body left in warm water will quickly develop postmortem thermal burns that appear as skin slippage. Decomposition may also present with skin slippage and blister formation. An individual may sustain full thickness burns to greater than half of his or her body and still be conscious without immediate death. Death often occurs later due to electrolyte imbalances or infection.

Also, mortality increases with age. Second- to third-degree thermal burns to half the body's surface would much more likely kill a senior citizen than a child. Second-degree burns are more painful than third- or fourth-degree burns due to less damage to nerve endings.

Antemortem burns may be characterized by fluid-filled-blister formation. To have fluid-filled blisters in a nondecomposing body in a nongravity-dependent area not adjacent to an area with more extensive burns with contracted tissue, one needs a blood pressure and a beating heart. Antemortem blisters also typically have a red base with surrounding erythema. This concept remains controversial in some jurisdictions. Also, a dead body from a house fire will decompose at a much slower rate than a dead body not exposed to smoke and intense heat Smoke and heat serve as preservatives. Low heat will accelerate the putrefication process. Radiant heat in a dry environment will quickly cause tissue to become firm and dehydrated, often with hair still present. This depends on the amount of heat and humidity and the duration of exposure. Also, intense heat may produce postmortem artifacts that may be misinterpreted as antemortem injury, such as an epidural hemorrhage. Epidural hemorrhage may be postmortem and is caused by heat-related contracture of the dura mater forcing blood from adjacent vessels into the epidural space. Subdural hemorrhages do not occur in this manner and are antemortem injuries.

Thermal burns sustained by fire may be accompanied by smoke inhalation. In general, most fatalities from house fires are caused by smoke inhalation. Decedents dying directly in house fires usually have mostly postmortem burns. Fire fatalities that occur outside in an open space often do not have associated significant smoke inhalation because the smoke rises rapidly and is not inhaled. In the case of a flash fire, inhaled super-heated gases damage the upper airways, including the laryngeal mucosa, and causing death from reflexive closure of the airway at the level of the vocal cords with asphyxia and eventual fatal arrhythmia. The effects of smoke inhalation are often reflected by the amount of carbon monoxide present in the blood. This depends on the nature of the burning material.

There are often other significant poisons associated with burning materials that can rapidly contribute to death, such as hydrogen cyanide. A fire victim does not need a lethal level of carbon monoxide to die of smoke inhalation. Carbon monoxide is produced from incomplete combustion of organic fuels. Carbon monoxide, in the absence of smoke, is a colorless, odorless gas that reversibly binds the hemoglobin molecule approximately 200 times greater than oxygen, resulting in hypoxia and possible death. Levels of carbon monoxide that exceed 50% saturation are considered life threatening, but may cause death with levels less than 26% saturation. Carbon monoxide levels of greater than 80% are possible. These levels need to be correlated with the physiologic disease state of the individual's body, including heart disease, for example. Someone with marked coronary-artery atherosclerosis would often require much less carbon monoxide exposure to produce death than a young healthy individual with slight atherosclerosis. Cigarette smokers may reach carbon monoxide levels of 10%. Carbon monoxide levels of 15–30% are associated with dizziness, nausea, and headache. Cherry-red lividity first becomes apparent at levels of 30–35%. The half-life for carboxyhemoglobin elimination in a resting adult at sea level is generally 4 to 5 hours. This may be reduced to 80 minutes following administration of pure oxygen, and may be further reduced to 24 minutes by using oxygen at 3 atmospheres of pressure. Primary elimination of unchanged carbon monoxide occurs by pulmonary excretion. Also, if there is more than one fatality without obvious cause, one should consider carbon monoxide poisoning.

Electrical burns may be due to low- or high-voltage exposure. The electrical current may be direct or alternating in nature. Alternating current is more likely to cause a fatal cardiac arrhythmia than direct current. High voltage is generally defined as greater than 1000 volts for alternating current and greater than 1500 volts for direct current. High-voltage burns are usually associated with extensive obvious injury. Low voltage is generally defined as being less than 1000 volts for alternating current and less than 1500 volts for direct current. Low-voltage burns may present with no visible marks to the body's surface at all. The degree of injury depends on many factors, including the duration of exposure and the amount of heat generated.

Electrical burns may also occur as a result of a lightning strike. Lightning bolts occur with an enormous short-term release of electricity, often producing minimal injuries. Lightning strikes may also produce a fern-like red pattern at the skin's surface. Patterns similar to this may be observed in high-voltage electrocution. These types of electrical discharges present with different autopsy findings. The mechanism of death is usually arrhythmia and more likely to occur if the current passes directly through the heart. Death may also occur due to asphyxia if there is interference with the central nervous system's respiratory centers or paralysis of the chest muscles. To complete an electrical circuit one needs an entrance and exit point for electricity to pass through the body. An otherwise healthy individual may be found lying barefoot and on a damp floor next to a power tool with a frayed electrical cord. This is why adequate scene investigation is crucial. It is also important to keep the electrical device as evidence to be tested, and to prevent any other fatalities. There may or may not be burns to the body's surface at autopsy.

Chemical burns are due to exposures to caustic substances. These burns most often involve injury to the skin or mucosa, leaving red discoloration or sloughing of the superficial layers. More extensive injuries may involve damage to the underlying tissue including bone. This depends on the strength and nature of the caustic substance, which include acids, bases, and other chemicals that can damage the body. Individuals may die acutely following chemical burns from many different mechanisms including hemorrhage, infection, dehydration, or they may die many years following such injuries. For instance, if an individual attempts to commit suicide by ingesting lye 20 years earlier and later develops esophageal cancer as a result of these burns, the manner of death would be suicide. Children may accidently drink caustic substances, leading to gastrointestinal perforation that may lead to adhesion and gastrointestinal obstruction many years later. In this case, the manner of death would be accidental. It is always very important to find out the initial event that starts the ball rolling in the sequence of events that eventually leads to an individual's demise.

Radiation is defined as energy distributed as waves or particles across the electromagnetic spectrum. This includes electric, radio, radar, microwaves, infrared, visible light (lasers), ultraviolet light, x-rays, gamma rays, and cosmic radiation. Waves are characterized as having long wavelengths and low frequencies, whereas particles have short wavelengths and high frequencies. The types of biological effects vary greatly depending on the type of radiation, duration of exposure, and intermediate barriers. Acute exposure to skin may range from erythema to overt necrosis with eventual epidermal atrophy and dermal fibrosis. Biological effects include cataracts, burns to the retina and skin, necrosis, fibrosis, and cancer. Generally speaking proliferating cells are affected more substantially with acute exposure as indicated by damage to the gastrointestinal and hematopoietic systems with increased risks of infection, nausea, vomiting, diarrhea, and hemorrhage. Damage to DNA may eventually lead to many different forms of cancer.

Soot within the nares due to smoke inhalation.

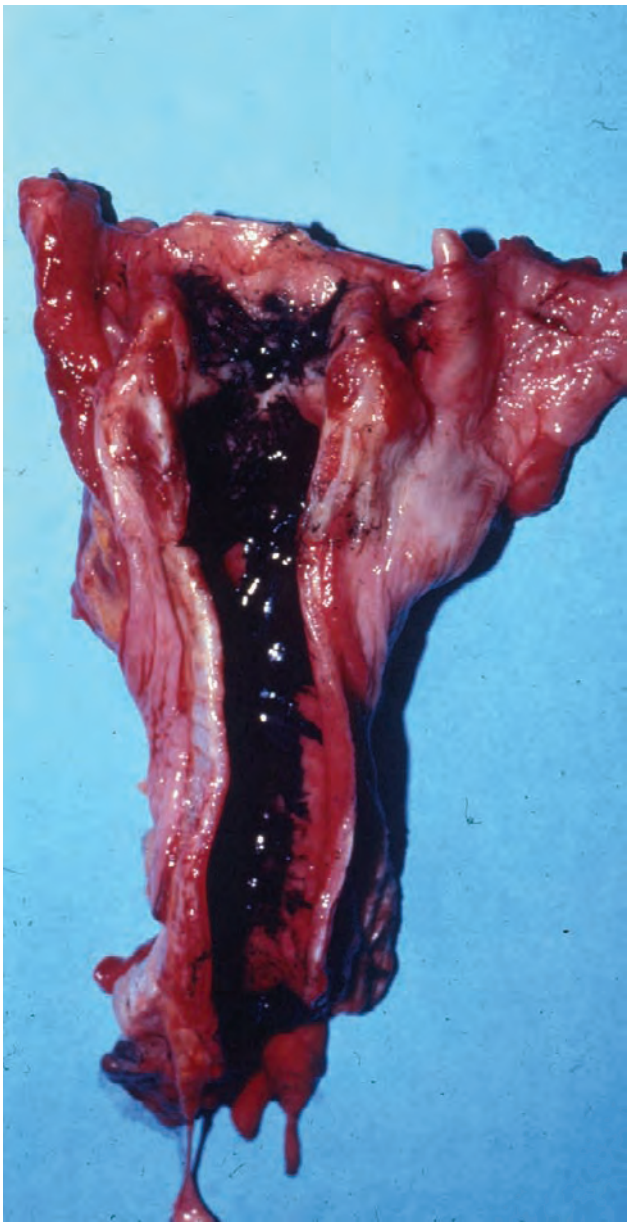

Smoke inhalation with soot deposition on the airway mucosa of the larynx and trachea.

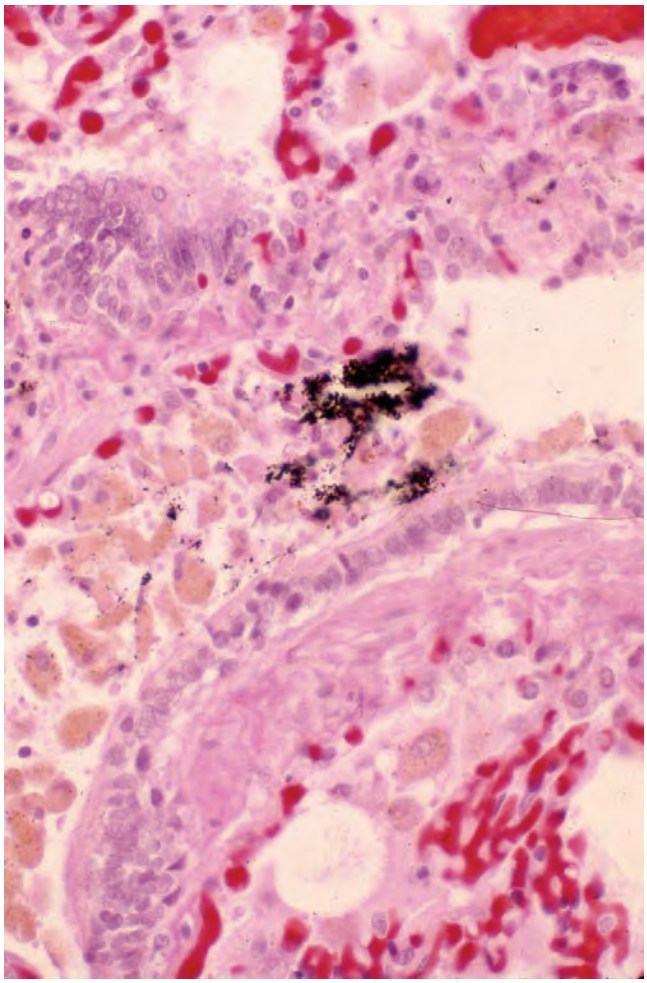

Microscopic view of lung with soot deposition in the distal bronchi.

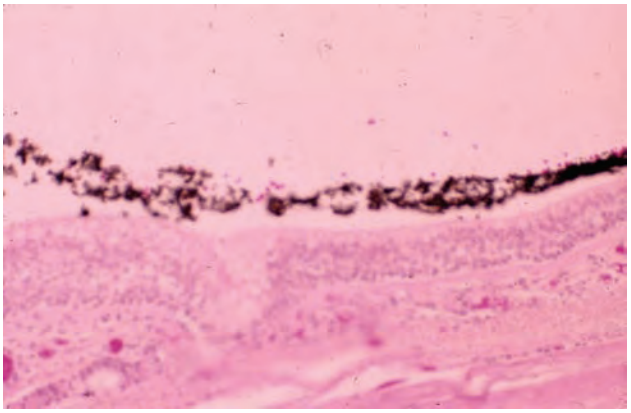

Microscopic view of airway with soot deposition at the surface of respiratory epithelium.

First-degree burn characterized by red discoloration with injury limited to the epidermis.

Second-degree thermal burns characterized by fluid-filled-blister formation. Second-degree burns include damage to both the epidermis and cermis and are often more painful than third-degree burns due to less destruction of nerve endings.

Second- to third-degree thermal burns. These areas demonstrate fluid-filled-blister formations that were interpreted as antemortem. To have fluid-filled-blister formation associated with thermal injury, one usually needs to have a blood pressure. Fluid-filled blisters generally do not occur due to putrefaction in burn victims due to the preservative effects of fire and smoke. Postmortem blister formation may occur in gravity-dependent regions without significant thermal damage or adjacent to regions with soft-tissue heat-related contractures. Also, postmortem blisters lack vital reaction.

First- to third-degree thermal burns involving the posterior aspect of the left thigh and gluteal region. The right leg is without thermal burns. First-degree thermal burns in this picture are characterized by the red discoloration without blister formation or skin slippage. Note the areas of collapsed blister formation, which are consistent with a postmortem burn. Sometimes it is difficult to interpret antemortem burns if continued heat causes fluid-filled blisters to collapse and fluid to evaporate. Most fire fatalities succumb from smoke inhalation before extensive burns occur.

Postmortem second- to third-degree thermal burn with skin slippage and collapsed blister formation.

Full-thickness or third-degree burns in an individual who lived hours after being shot in the head and having his or her residence set on fire to destroy evidence. Full thickness refers to involvement of the epidermis, dermis, and subcutaneous layers. These are often less painful than second-degree burns due to more complete damage of nerve endings in the latter.

Note the sparing of thermal injury at the bottom of the decedent's foot due to protection by a shoe.

Antemortem third-degree burns. This is self-immolation of an individual who was wearing sandals. Note the lack of thermal injury in the region protected by the shoe strap.

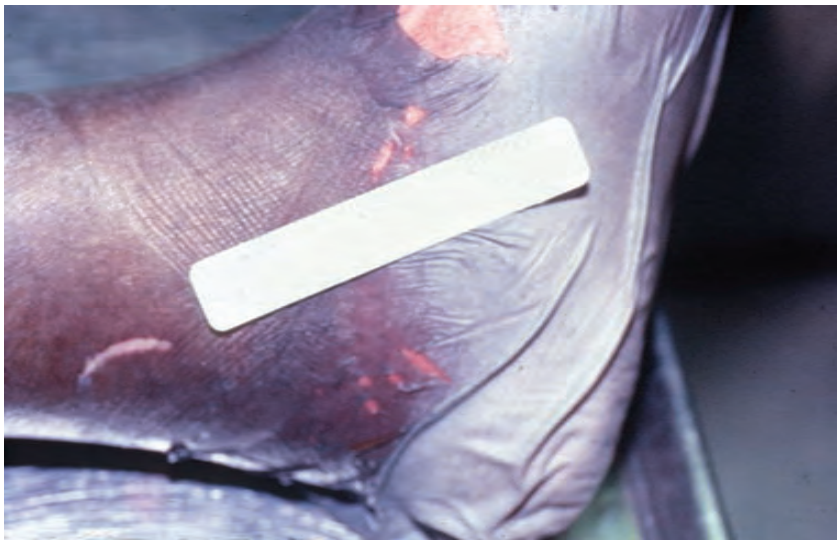

This image demonstrates postmortem second- to third-degree thermal burns to the sole of this foot. Note the wrinkled thick skin demonstrating skin slippage. There was minimal underlying fluid accumulation or blister formation.

Postmortem fourth-degree thermal burns with extensive charring and exposed muscle and bone. Note the pugilistic stance of the body with arms raised (b). Also note the postmortem skeletal fractures due to extensive heat (c).

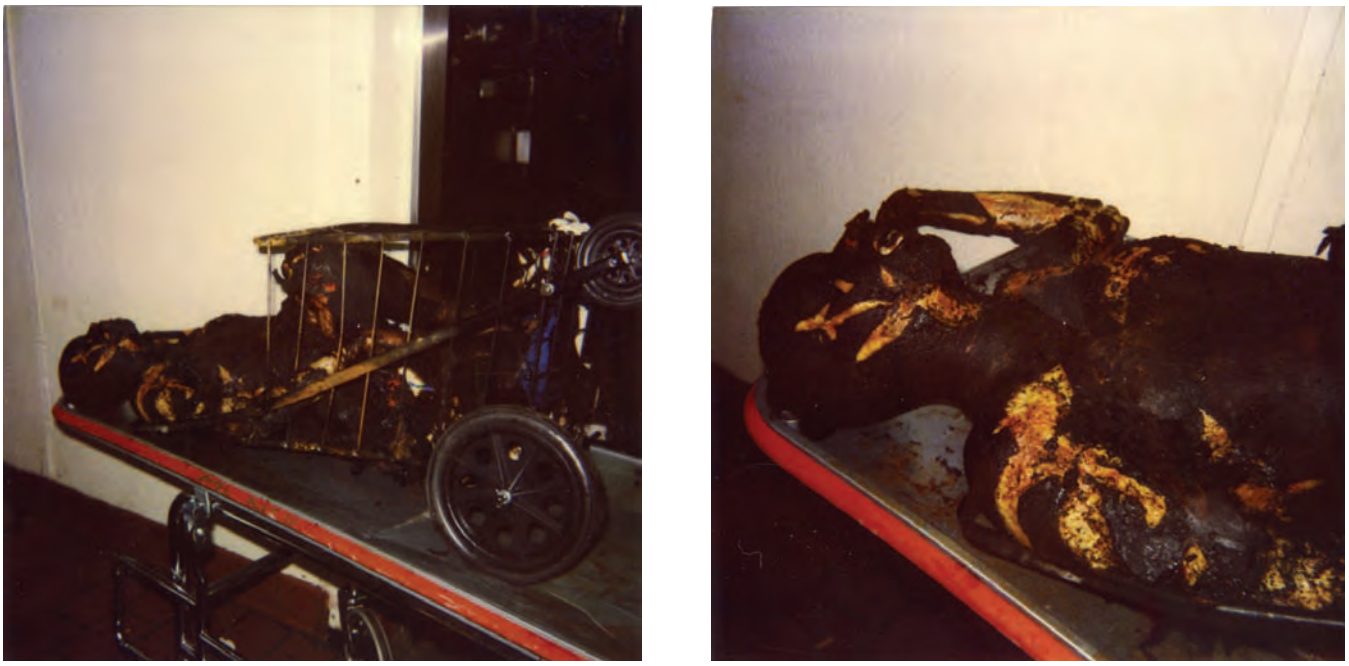

This individual was wheeled out of a building in a shopping cart and set on fire. Note the postmortem cracking of the skin due to heat exposure. This is indicated by the exposed yellow subcutaneous tissue with no hemorrhage. There were extensive fourth-degree thermal burns to the entire body.

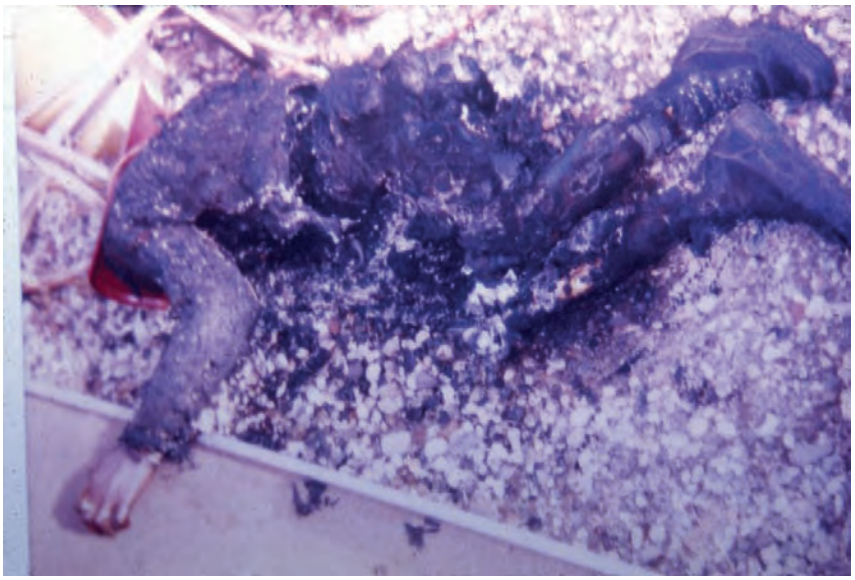

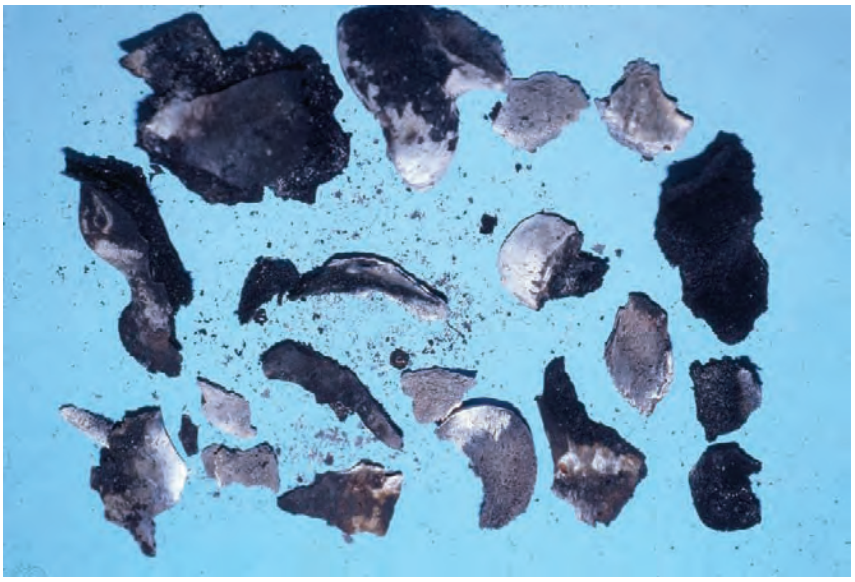

Extensive third- to fourth-degree thermal burns with partial skeletal fragmentation due to the wick effect, which refers to a self-perpetuating, low-intensity flame following ignition of certain materials contacting the body, where the skin is cracked from heat and the underlying fatty tissue is rendered into oil that is absorbed into the charred clothing, producing a wick. This low-level heat can produce extensive destruction to a body over hours.

This is a rare finding of decomposition after extensive thermal injury and exposure to smoke. The decedent was involved in a fatal fire and the body was not discovered for several days after being soaked in water following fire extinguishing. Note the microorganisms and mold growing at the body surface shown by the gray-white discoloration. Putrefactive changes were markedly inhibited due to the effects of exposure to smoke and heat.

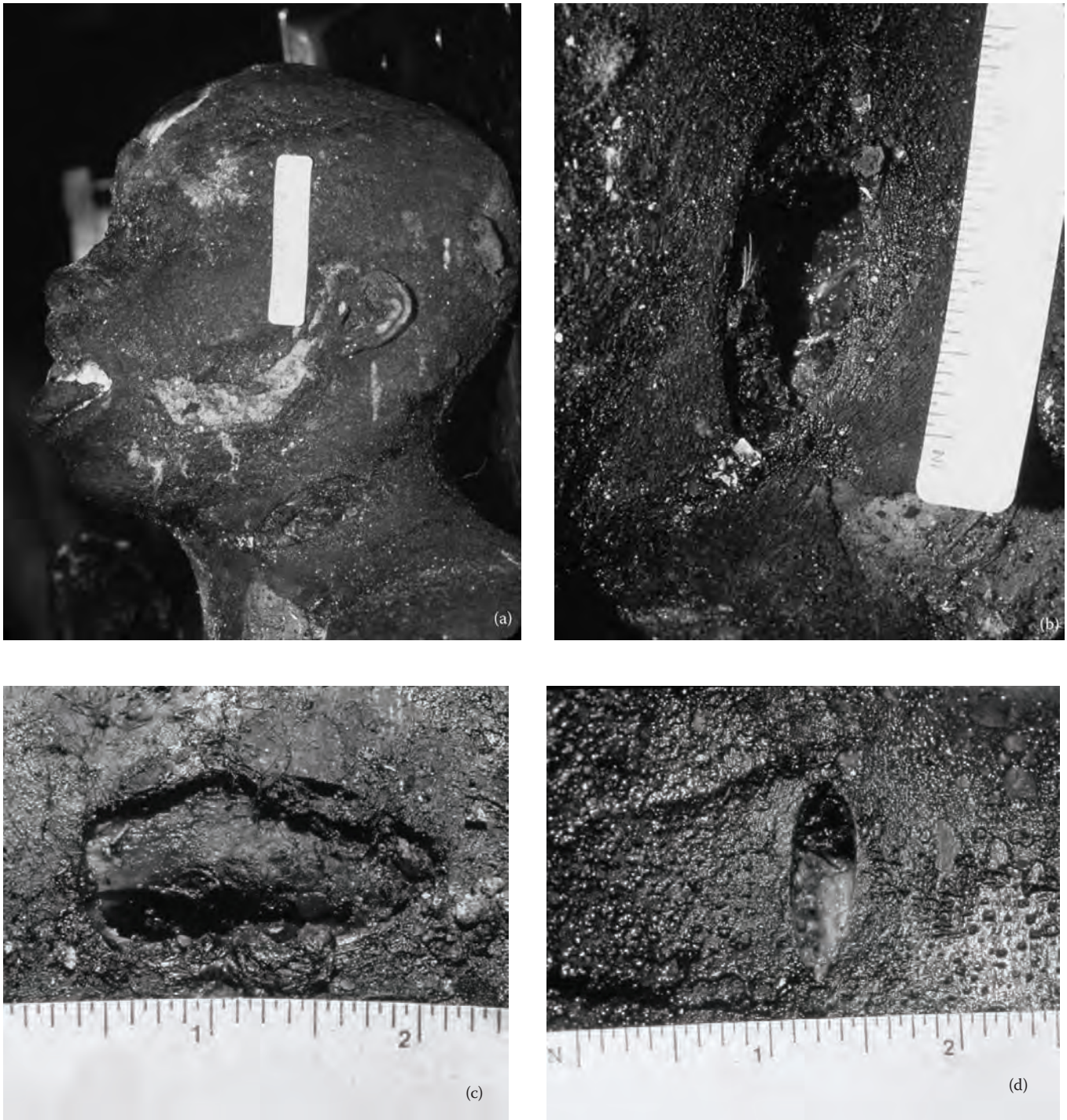

These are all examples of homicides where apartments or houses were set on fire to destroy evidence. Figures (b–d) are examples of antemortem stab wounds that have been altered by postmortem thermal burns. Note that the margins are relatively sharp and do not appear as though the skin has cracked secondary to heat exposure. Upon internal examination, these injuries become much more apparent with hemorrhage and blood accumulation. Note in (a) areas of cracked skin with exposed yellow subcutaneous tissue without hemorrhage, which is indicative of postmortem thermal injuries. Note the wound at the upper aspect of the lateral left neck with hemorrhage due to an antemortem stab wound.

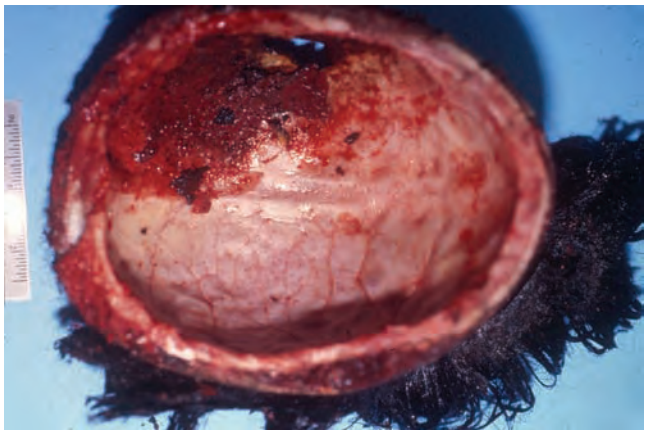

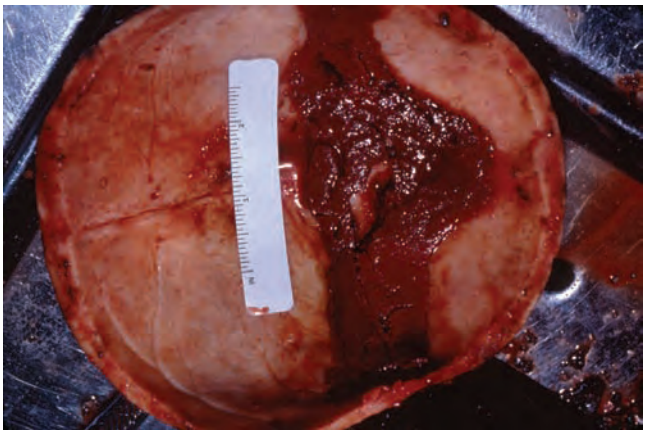

These figures demonstrate postmortem epidural hemorrhage. These can sometimes be misinterpreted as antemortem blunt force trauma. With exposure to flames, the brain and dura mater may contract, and blood may be forced from the small vessels at the inner aspect of the cranial bone and through the dural sinuses, producing epidural blood accumulation that will coagulate with heat.

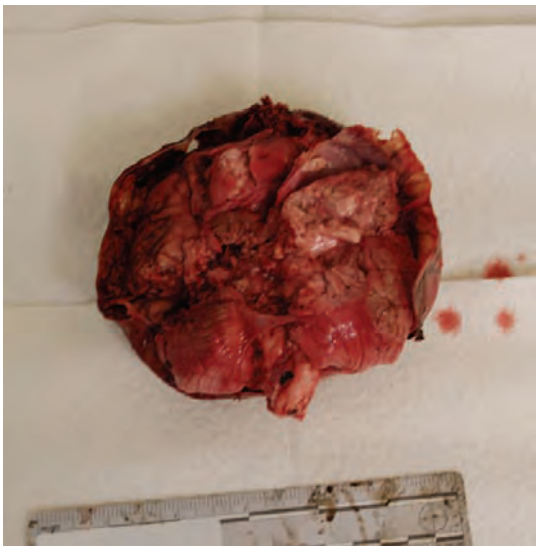

Extensive postmortem thermal injury with brain shrinkage. Organs exposed to extensive heat will decrease in size largely due to dehydration.

This individual died of an overdose and was found with dried vomitus on her face. Note the red discoloration caused by gastric acids producing burns to her face.

Chemical skin burns caused by spilling sulfuric acid. The sutured linear incisions are due to organ donation with retrieval of bone and soft tissues.

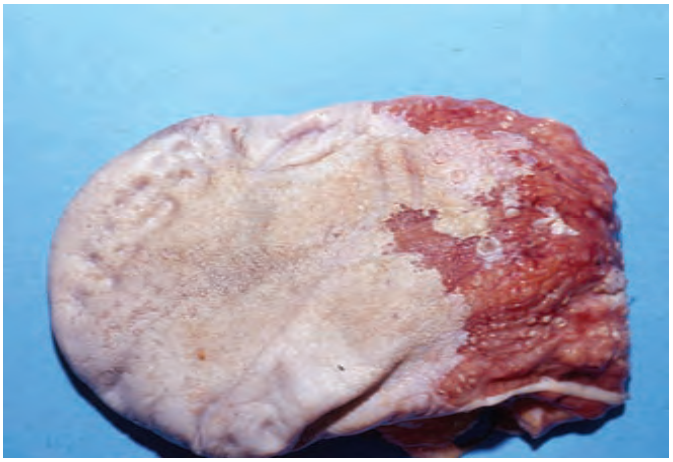

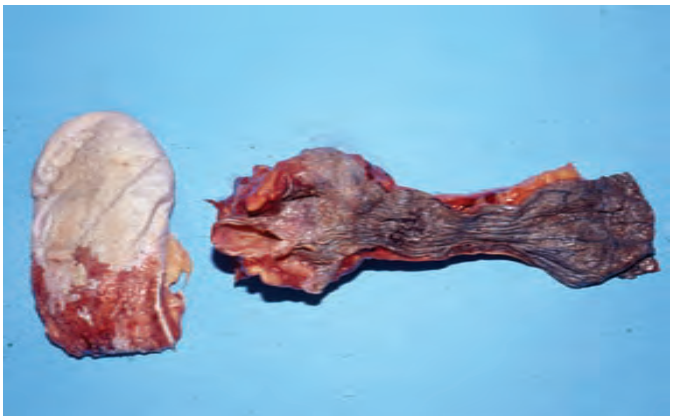

This individual drank a mixture of lye, kerosene and other caustic chemicals. Note the white discoloration from chemical burns at the lips, mouth, tongue, and esophagus.

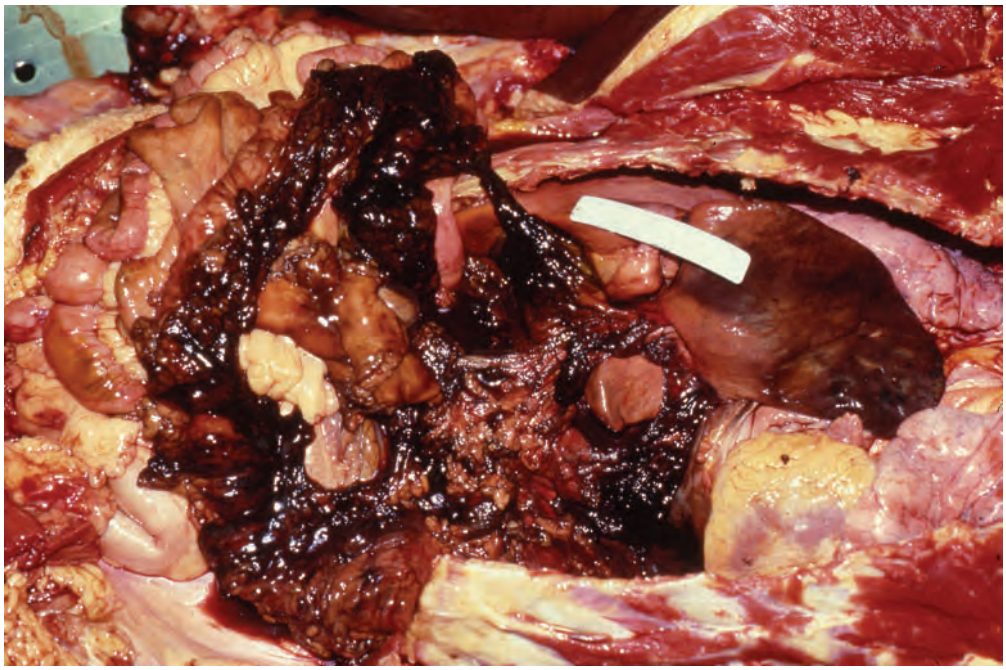

Note the red to brown discoloration in the abdominal cavity following gastric perforation and leakage of the caustic chemicals into the peritoneal cavity.

Electrical burn on the hand of this individual who reached into a ceiling and grabbed a live wire. He was standing on an aluminum ladder wearing shorts. His leg contacting the ladder completed the circuit through his heart, producing a fatal arrhythmia.

This individual fell from a subway platform onto the third rail and sustained these electrical burns.

These are examples of second- to third-degree electrical burns caused by inadvertently contacting live wires at construction sites. These individuals died as a result of a fatal cardiac arrhythmia. They fell to the ground lifeless within about 15 to 20 seconds after contact.

This individual was found lying next to an electrified subway rail with his pants down, and an electrical burn to his penis. He was lying in a puddle of urine and was markedly intoxicated at the time of his death.

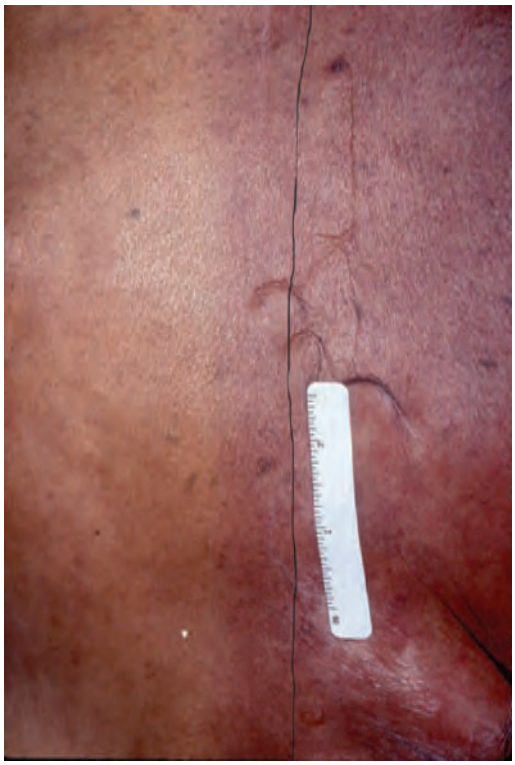

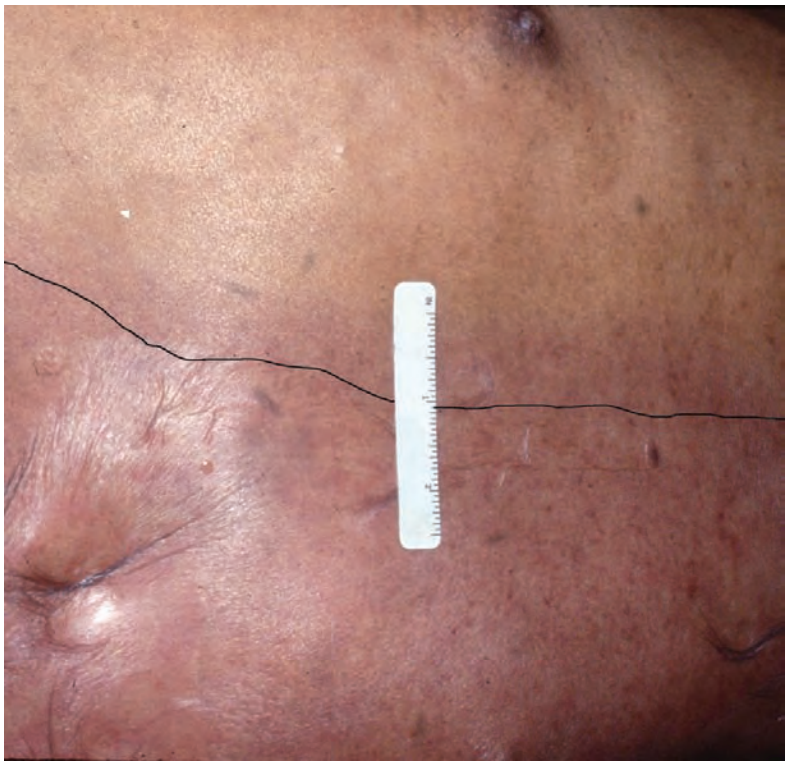

These individuals died of heart disease while taking a bath. These are examples of second- to third-degree postmortem burns due to being submerged in warm to hot water. Postmortem burns occur with less heat than antemortem burns. Note the red discoloration with skin slippage and a sharply demarcated border defining the submerged and unsubmerged areas. To help the viewer distinguish between these regions we placed a line adjacent to this demarcation. Individuals who drown in bathtubs have some contributing factor dictating why they could not keep their head above the water. They maybe neurologically compromised or intoxicated. People with seizures may accidentally drown.

Child abuse cases with homicidal scalding. Image (a) demonstrates second- to third-degree burns to both feet and ankles from repeatedly being dunked in hot water. Images (b–d) show skin grafting with therapeutic intervention in a child with fourth-degree burns who was left sitting in scalding water.

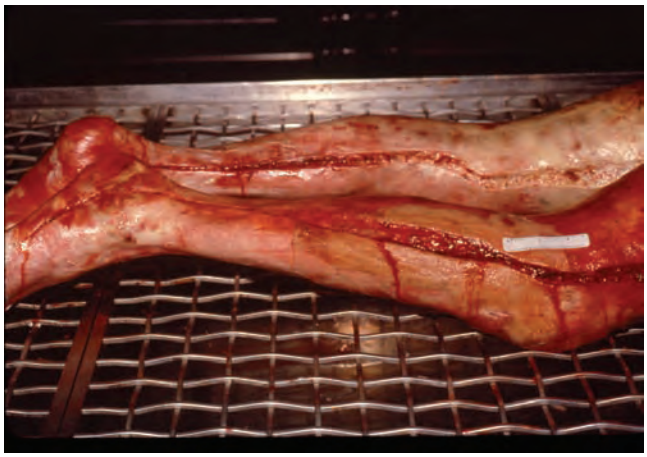

These are fasciotomy incisions made by physicians for medical therapy to relieve pressure in extremities.

Identifications may be very challenging in fire fatalities. These cases may require dental analysis for identification. Incisions may be made in the face to access the teeth. Funerals in such cases are closed casket due to the severe extensive nature of the injuries.

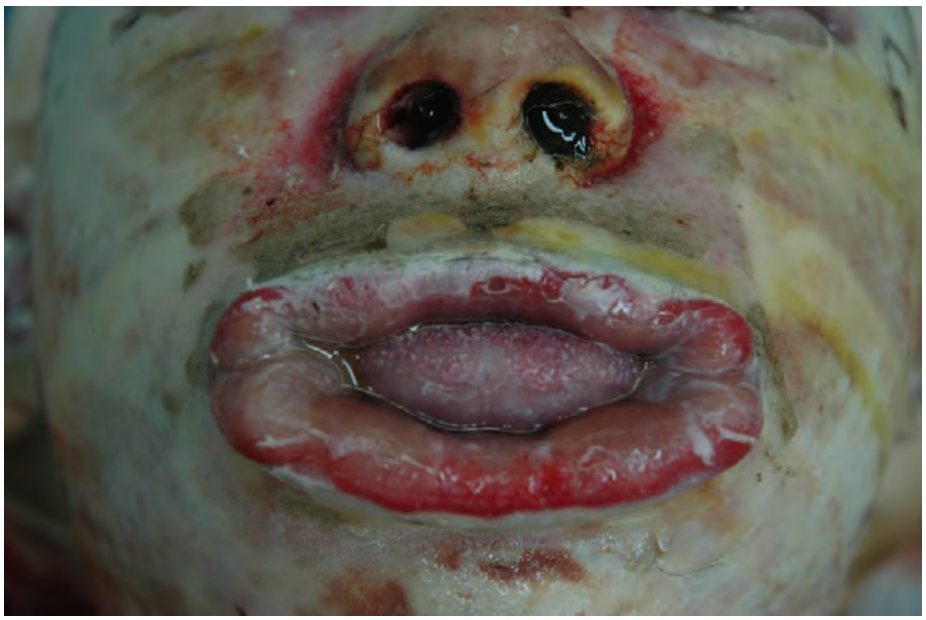

Swelling of the tongue due to inhaling super heated gases in an individual who survived for 1 day.

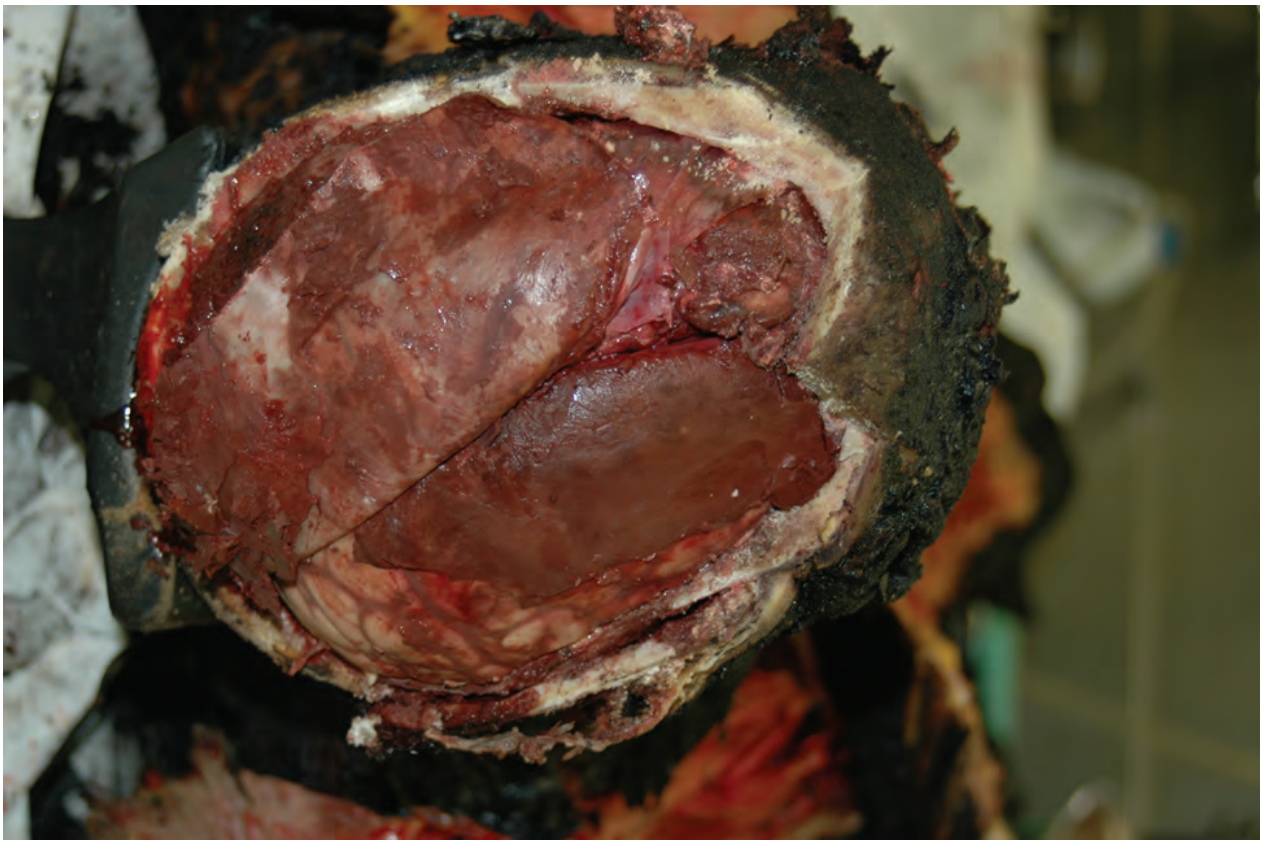

This individual crashed an airplane and sustained extensive blunt force trauma. These images demonstrate an antemortem subdural hematoma altered by extensive postmortem thermal injuries.

Asphyxia

10

CHARLES A. CATANESE AND BARBARA K. BOLLINGER

This is a general term used to describe decreased oxygen uptake or use, together with decreased carbon dioxide elimination.

Airway obstruction may occur by smothering, neck compression, foreign body aspiration, excess secretions or swelling of the airway, etc. *Smothering* is defined as external occlusion of the mouth and/or nose which prevents air exchange. Children may *aspirate foreign bodies* such as peanuts, hotdogs, popcorn, watch batteries, coins, etc. Adults who aspirate food are invariably neurologically compromised or intoxicated. Airway obstruction due to *excess mucus or swelling* as with asthma or anaphylaxis can produce asphyxia. Also various body positions may produce airway obstruction *(positional asphyxia)* as with occupants of motor vehicles trapped after collisions or intoxicated people passing out and sliding into awkward positions that prevent chest expansion and air exchange.

Autopsy findings associated with *smothering* may be very subtle or non-existent. Findings may include abrasions around the nose and/or mouth that *cannot* be explained by other means (i.e., resuscitative efforts). Great force is applied to the mouth and lips which may cause tears to the frenulum of the lip, the mucous membrane that connects the inside of the lip to the corresponding gum. Smothering may occur with the use of hands or by placing an object over a face, such as a pillow.

Depending on the degree of force applied and the structures compressed, there may or may not be petechiae present on the skin of the face, mucous membranes or eyes. Arguably, the greater the disparity in size between the perpetrator and the victim (i.e., adult and child), the less likely there will be a demonstrable injury due to the overpowering relative nature of this type of struggle. It is important that in cases of suspected smothering, experienced police interrogators and medical investigators perform the interviews. In cases where autopsy findings are very subtle, well-documented descriptions of the circumstances with specific details are extremely important.

Chemicals can produce asphyxia. *Inert gases* like methane or carbon dioxide will displace oxygen from the air and produce asphyxia by depletion or replacement of oxygen. Various *poisons* such as carbon monoxide or cyanide interfere with oxygen uptake and utilization, respectively.

Chest compression can produce asphyxia by *preventing air flow* into the lungs.

Neck compression as with *hanging and strangulation* can also produce asphyxia by obstruction of various neck structures including the airway, venous circulation and arterial circulation.

Interpretation of autopsy findings with respect to *hanging vs. strangulation* can be challenging. Each subheading below will describe the presenting classic and most common features, and then elaborate on less common features. It is important to realize that there is overlap between how the two present; depending on how the act is carried out, they may appear very similar. In establishing the manner of death, one should consider all aspects of the case including the past medical history (i.e., depression, end stage cancer, etc.), scene investigation and autopsy findings.

Hanging refers to ligature compression of the neck mitigated by the gravitational forces of the hanging head, causing partial or complete obstruction of the neck structures including blood vessels and the airway.

In a typical non-judicial suicidal hanging an individual places a ligature with a slip knot encircling the superior aspect of his or her neck. He or she secures the other end of the rope to a fixed support and allows the entire or partial body weight to pull downward, occluding the neck structures until loss of consciousness and death. In this case there should be furrow pattern that matches the overlying ligature which forms an inverted "v" mark or indentation, extending upward at the superior aspect of the neck and head. With the entire body weight pulling downward, all of the neck structures are (i.e., arterial, venous, and airway) are usually occluded at the same time and one would not expect to find petechiae in the face or eyes. There are typically no hemorrhages or fractures of the neck structures or other injuries to the body indicating a struggle. The cervical vertebrae are rarely fractured in suicidal hangings. When the body hangs for longer periods of time the furrow indentation becomes more prominent. Individuals cut down shortly after this act may have little or no furrow mark. This depends on the type of ligature used. A wide soft ligature will leave less of a mark than a narrow, more resistant ligature. If the body is left to hang for days, decomposition with stretching may eventually lead to the head being pulled away from the body.

In the case of a judicial hanging, the body is dropped from a height to produce sufficient force to fracture the upper cervical vertebrae resulting in spinal cord injury with cardiac and respiratory satiation.

An individual hanging in a sitting or lying position may have partial occlusion of the neck structures before loss of consciousness ensues. In situations where the body is adjacent to another structure, the individual may partially pull his or her body up and down, causing varying degrees of pressure–release before loss of consciousness. This will produce a similar effect to what is seen in strangulation. In these circumstances, the up and down motion of pressure release will obstruct different neck structures at different intervals. Venous circulation requires the least amount of pressure for occlusion, as compared to the arterial system and the airway. When venous circulation is obstructed without the arterial circulation, the higher pressure arterial blood beats through the capillary beds rupturing small blood vessels producing petechiae. As this process continues the hemorrhage size increases and may become confluent. These are most obvious within the sclera and conjunctivae. These movements may also produce hemorrhages or fractures to the neck structures including the airway cartilages and hyoid bone. These findings are more characteristic for strangulation, but may be seen in hangings.

Though unusual, people have been known to tie their hands behind their backs during the process of hanging themselves. The individual may have tried to complete this act several times in the past but the will to survive overpowered the will to end life. The tied hands are usually loosely tangled and not tightly tied. This will give the individual enough time to prevent himself from stopping the process. There are usually no other signs of a struggle or defensive type injuries.

With a free hanging and total body weight suspension, an otherwise healthy individual, using a slip knot, would be expected to lose consciousness within 15 to 20 seconds, and suffer irreversible brain damage within 4 to 6 minutes.

Hanging is usually suicidal, but may be accidental or homicidal.

Strangulation may be by ligature or manual. A *ligature* is something flexible that can encircle the neck, like a cord, belt or piece of clothing, etc. *Manual* strangulation refers to the use of one's hands leading to compression and blockage of the neck structures.

Manual strangulation is usually characterized by multiple irregular, angulated, abraded contusions around the neck. The marks may be curvilinear, corresponding to fingernail prints. These external marks can be somewhat variable and can range from a few to many.

Ligature strangulation is usually characterized by a horizontal furrow or mark pattern around the neck. The extent of these injuries depends on the type of ligature, how broad and soft it is, and the amount of struggle, etc.

The act of strangulation is often a very physically dominating, often non-premeditated, way of killing somebody. This act takes time and comes with the risk of injury to all those involved. The individual strangled is usually smaller and of weaker strength. Often a sexual component to the assault exists and a rape kit should be performed in all cases of suspected strangulation. Petechiae are usually present in the face and eyes. There are usually hemorrhages in the strap muscles of the neck, and there may be fractures of the laryngeal cartilages and/or hyoid bone. These fractures are more common in older victims because the cartilages are more calcified, brittle and less elastic. Older people may also have osteoporosis. Younger victims or children tend to have more flexible upper airways that often will stretch or collapse rather than fracture. Depending on how great the struggle, the amount of force used, and the type of neck compression, there may or may not be petechiae and/or hemorrhages above the ligature or region of neck compression. The presence of petechiae formation is more likely when there is a pressure–release component associated with a struggle or if less variable force is applied and if the ligature is wide with a large surface area. The presence of petechiae is less common when the force is very strong, consistent, and applied with a small surface area ligature. The latter example is more similar in nature to a hanging. Victims of strangulation often have defensive type injuries including other abrasions and contusions to their body.

It is possible to ligature strangle yourself; it is not possible to manually strangle yourself. If one is able to apply enough force to lose consciousness manually, revival occurs after the pressure is released. Continued force is necessary for death to ensue.

Someone murdered by ligature strangulation may die within a similar time frame as someone hanged; however, the time frame is usually longer. If there are multiple petechiae with hemorrhages and fractures of the neck structures, whether ligature or manual, the time frame may be much longer. This latter example usually indicates pressure–release, pressure–release over several minutes until loss of consciousness, and then continued pressure for several more minutes until death occurs. If an individual is released shortly following loss of consciousness, revival may follow.

Proper autopsy technique dictates that the *brain and visceral organs be removed prior to a layered neck*

dissection being performed. Photographs of the neck dissection in layers are recommended.

Drowning occurs when water is inhaled, filling up the alveolar spaces and preventing gas exchange. The manner of death will vary depending on how the individual came to be in the water. If the event is not witnessed the manner often remains undetermined. Water in the lungs (pulmonary edema) and the paranasal sinus is often present. The degree of pulmonary edema may vary in a fraction of the cases due to the heart beating after respirations cease. As the heart beats, before eventual asystole, some of the fluid in the lungs will be absorbed. An adult may drown in a big pool or ocean but not in a small pool or bathtub unless *neurologically compromised or intoxicated*. Homicide victims are sometimes placed in a water-filled *bathtub* to wash away evidence. The story given may be, "I found him in the tub, and he must have drowned." Common sense would dictate that an individual does not need to move very much to get his head above the water.

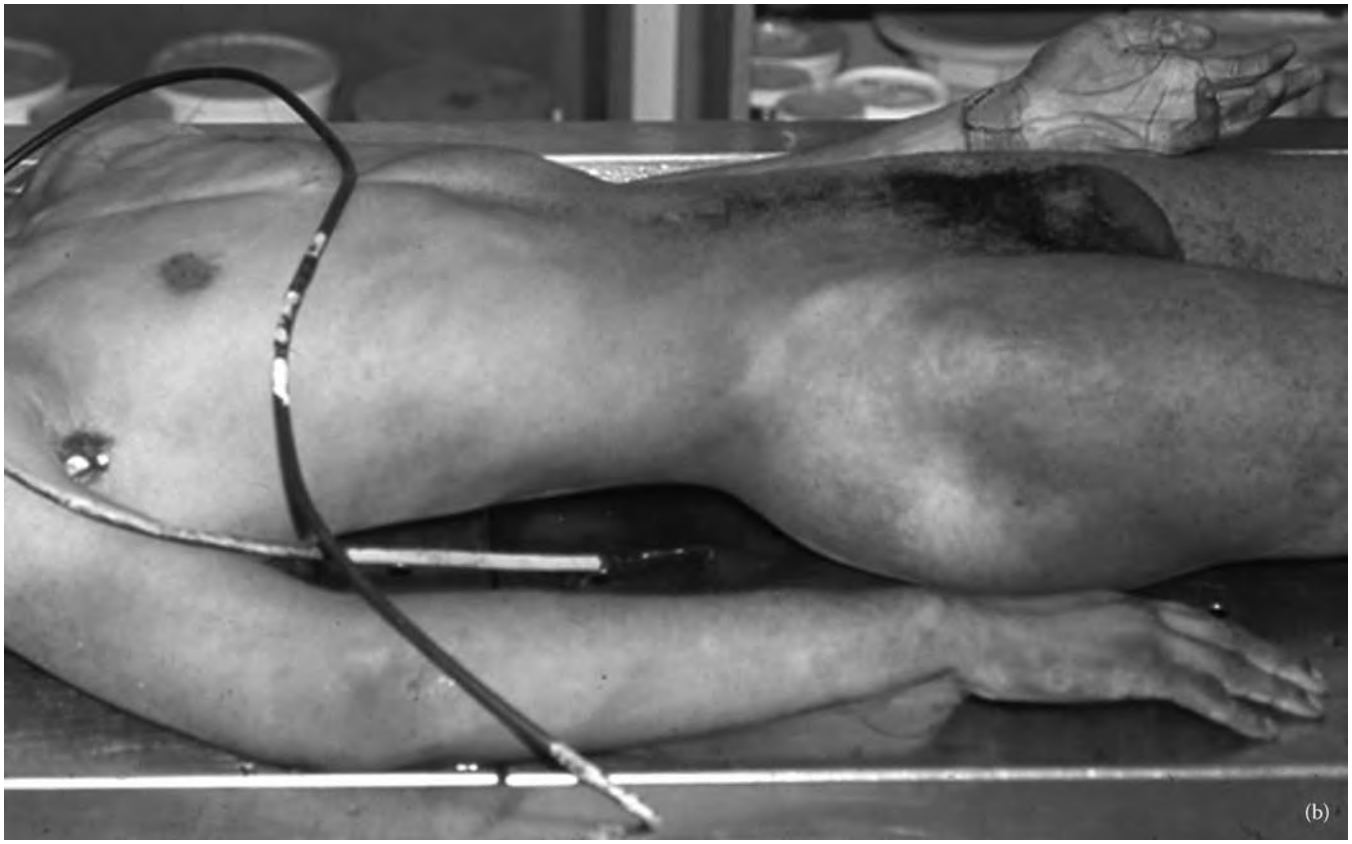

Suicidal hanging. Note the black cord (a) that matches the underlying furrow pattern at the superior aspect of the neck with upward extension at the left side of the face. Also note that the individual's lividity becomes more apparent and is fixed at the level of the upper thigh extending down to the feet (b). This fixed lividity pattern is appropriate for an individual who remained in an upright position for many hours after death. If his lividity was fixed posteriorly and not inferior; this would indicate prior scene alteration. There were hesitation marks at the wrist with blood seeping downward due to gravity (c).

Suicidal hanging with matching rope and furrow pattern circling superior aspect of the neck forming a slight inverted "V" pattern behind the right ear.

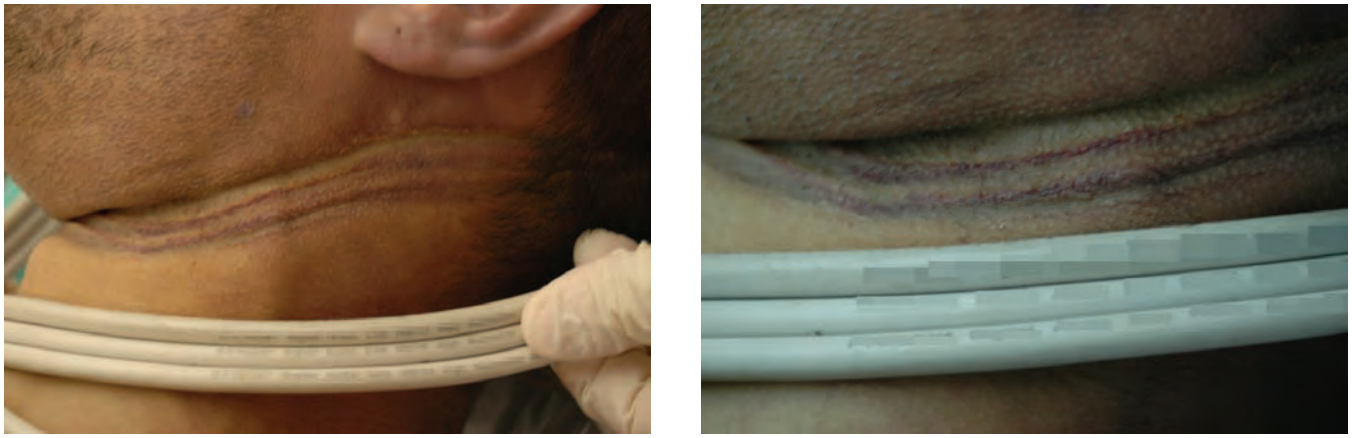

Suicidal hanging with computer cord.

Suicidal hanging with nylon bag strap.

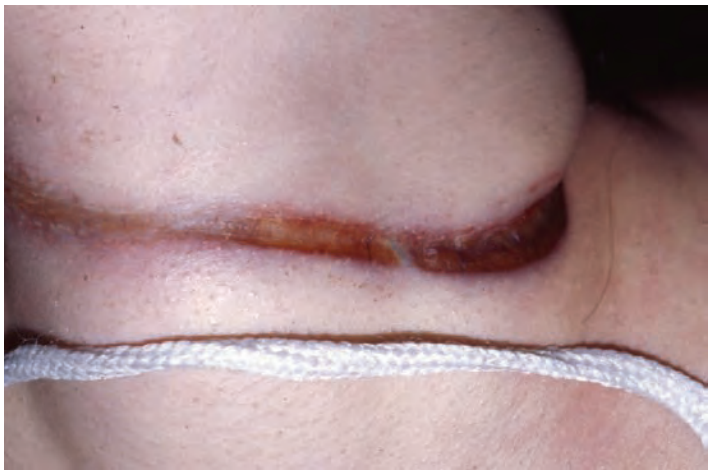

Accidental hanging in an individual with Alzheimer's dementia who enjoyed whirling around in circles while sitting on a desk chair. The person was left unattended for a short period of time by chronic care nursing staff, and got tangled in the cord from a window blind. This furrow pattern matches a blind cord as a ligature.

Suicidal hanging with chain.

Suicidal hanging by braided belt. Note the matching skin pattern.

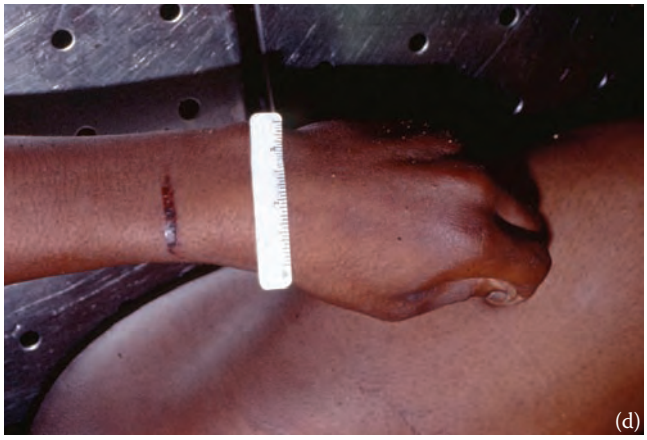

Suicidal hanging with earphone cord (a). This individual also took part of the cord and loosely wrapped his hands behind his back so he would not be able to reach up and prevent the hanging from being successful. This is not very unusual for people to hang themselves and loosely tie their hands in this fashion. Note the force of the ligature caused the tongue to protrude from his mouth (b). Also note the dark postmortem drying of the mucosa exposed to air.

Homicidal ligature strangulation in a prostitute who had large amounts of cocaine and heroin in her system. Her fingers were also noted to be crushed by a pair of pliers during the assault.

Homicidal ligature strangulations with horizontal cord marks surrounding the necks. Note the other injuries to the decedent's face in (b), including abrasions. Both of these victims were also sexually assaulted.

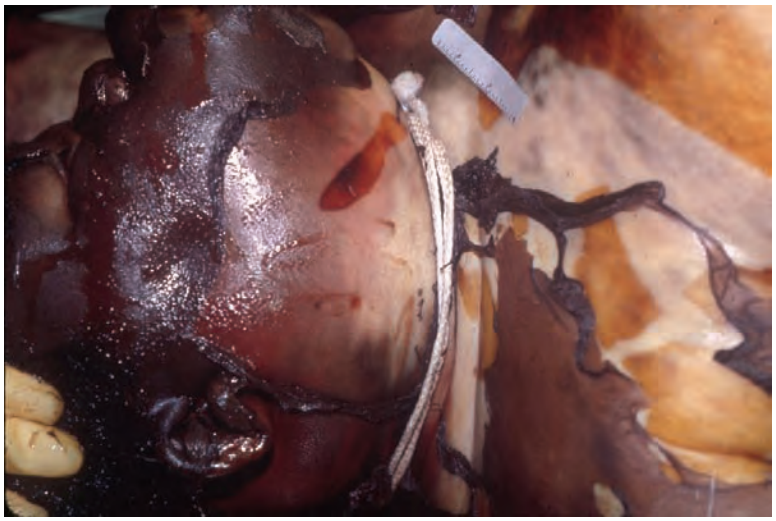

This is a suicidal ligature strangulation in an individual with a longstanding history of depression who was found in a locked secured apartment with a suicide note. Note the cord is tied tightly around his neck. There is also a moderate state of putrefactive change with skin slippage, bloating, and purging.

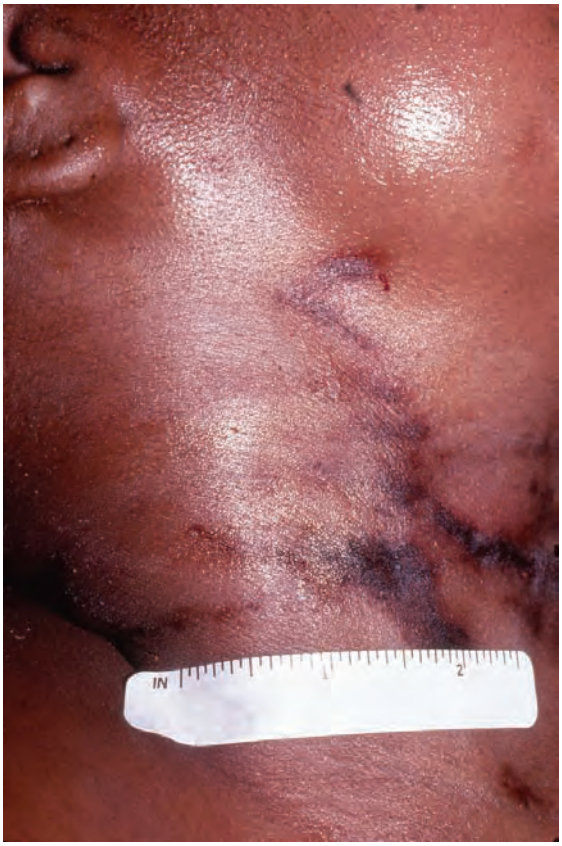

Manual strangulation. Note the multiple irregular abraded contusions surrounding the neck. There were multiple areas of strap muscle hemorrhage and a fractured hyoid bone. She was sexually assaulted and left in a stairwell. There were multiple petechiae with areas of hemorrhage in the sclerae and conjunctivae of each eye. The constellation of these findings are typical for a homicidal manual strangulation.

Manual strangulation. Note the abraded contusions to the anterior neck.

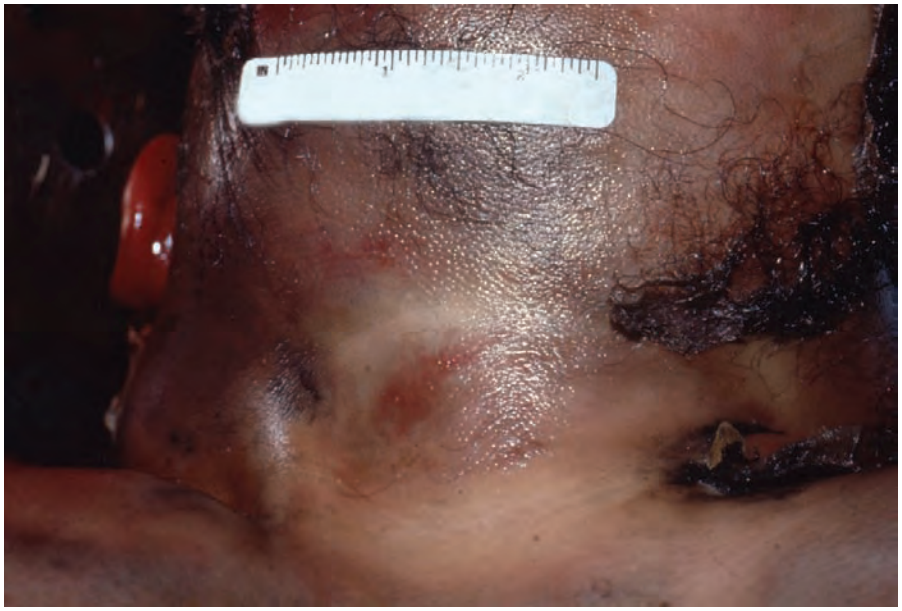

Manual strangulation. There is early decomposition with skin slippage. There are vague abrasions and superficial contusions at the neck. As decomposition progresses it may become more difficult to interpret these findings.

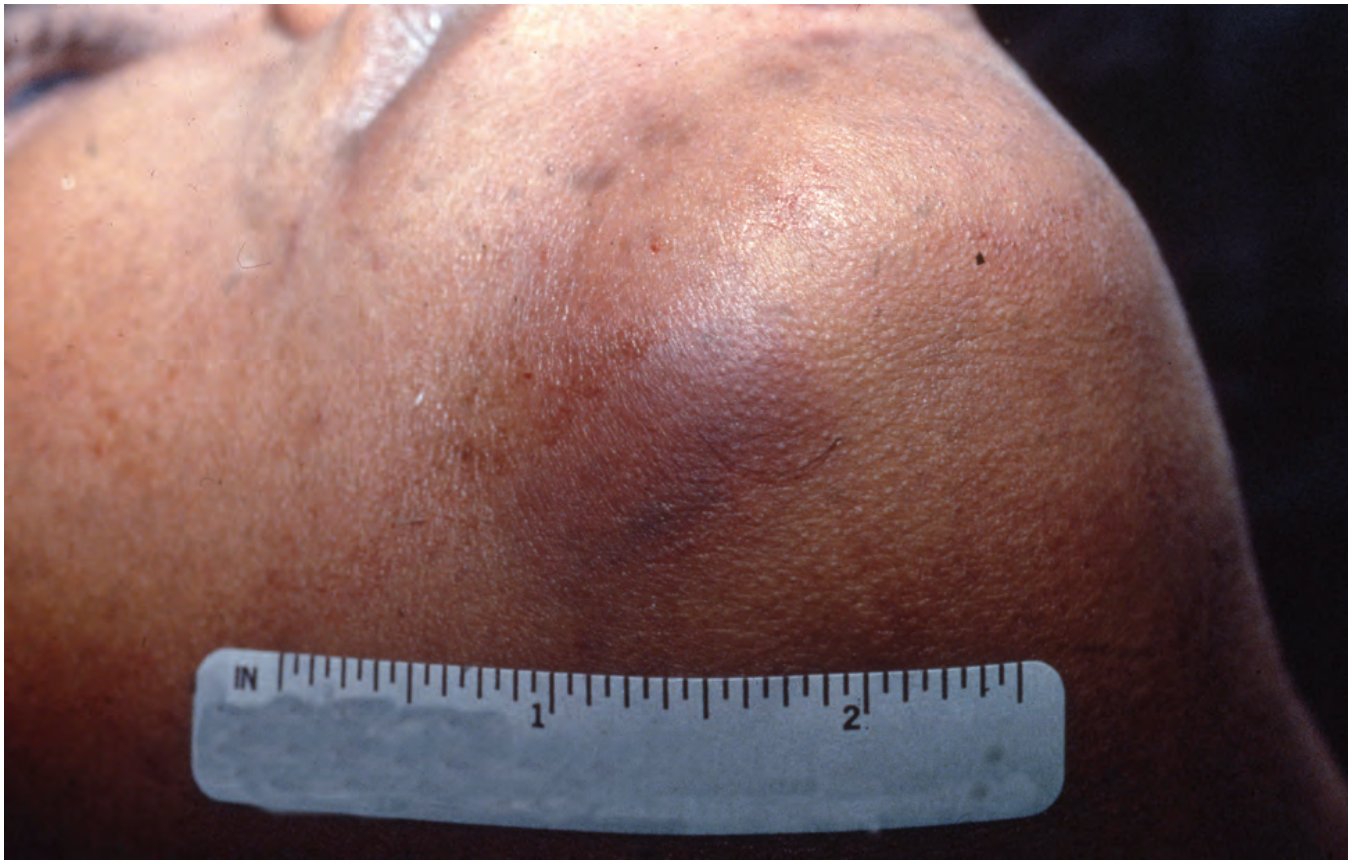

This individual was punched in the face and then yoked from behind during a sexual assault. There were no visible external injuries to the neck except this contusion at the border of the right chin. Careful internal examination after the brain and visceral organs had been removed revealed areas of hemorrhage within the anterior strap muscles and posterior paraspinal muscles. There were also petechiae with hemorrhage of the sclera and conjunctivae. This individual was found face down and the eye hemorrhages were originally thought by some to be associated with postmortem lividity. The perpetrator was caught after bragging about the assault.

This image demonstrates an eye from a person who hanged him or herself. Note there are no petechiae or scleral/conjunctival hemorrhages. The decedent's weight pulling down on the ligature produced a significant enough force to obstruct the entire blood supply to the head. Therefore, there was no pressure–release mechanism, leading to capillary rupture and hemorrhage. The absence of scleral or conjunctivae hemorrhages is more typical in hanging fatalities.

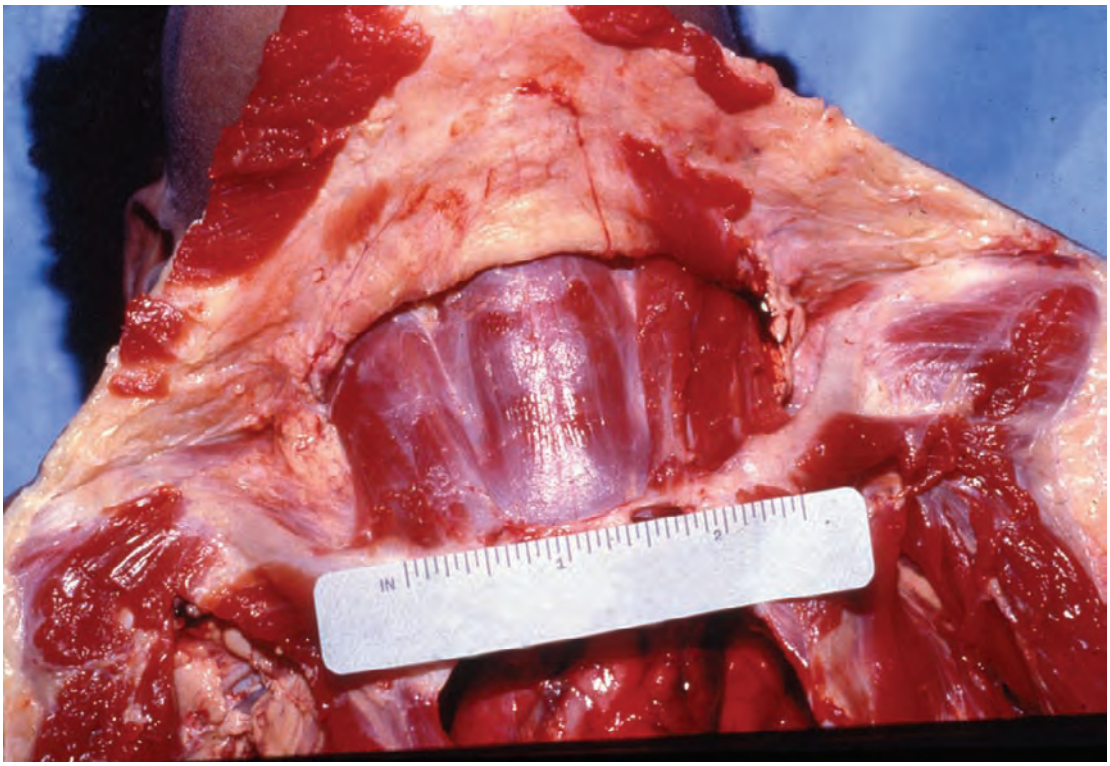

Anterior neck dissection after the visceral organs and brain have been removed. Part of the undersurface of the platysma muscle is visible at the top of this image adjacent to the yellow subcutaneous tissue of the neck. The anterior strap muscles are visible directly above the label and are free of antemortem injury. There were no hemorrhages or fractures of the remaining neck structures. The absence of these injuries is common in hanging fatalities.

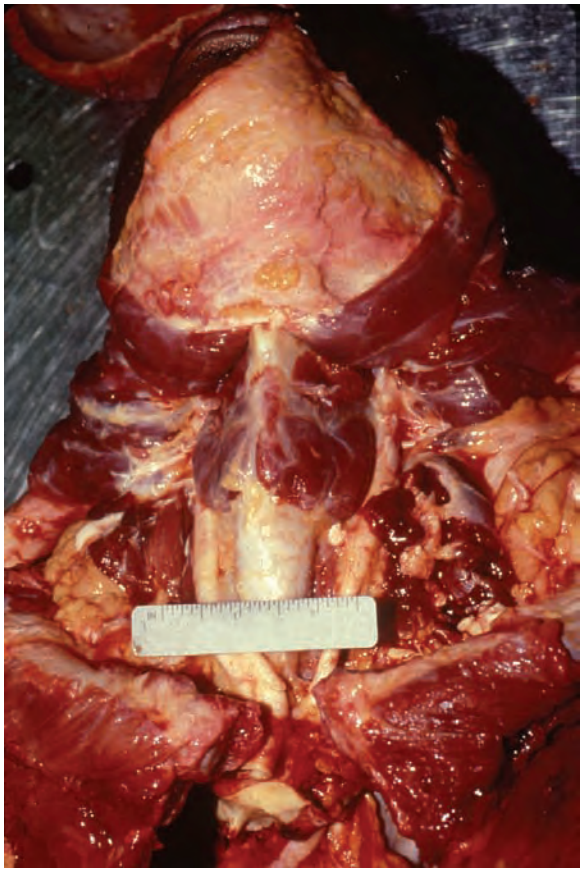

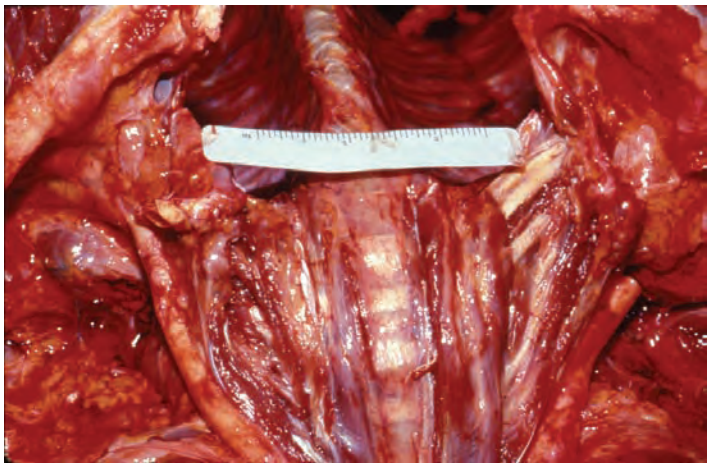

These figures demonstrate anterior neck dissections with no hemorrhages. The absence of hemorrhage is typical for most hangings.

This demonstrates a normal hyoid bone with no fractures or contusions.

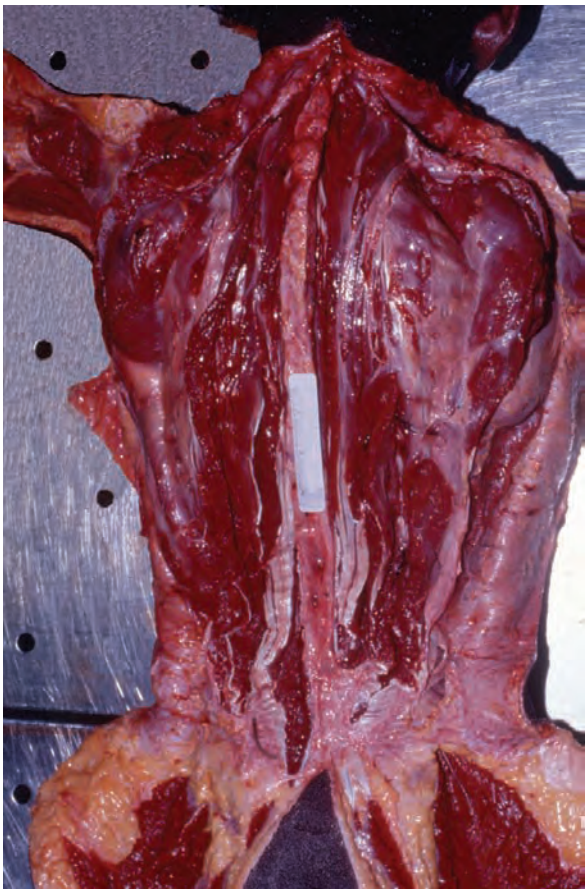

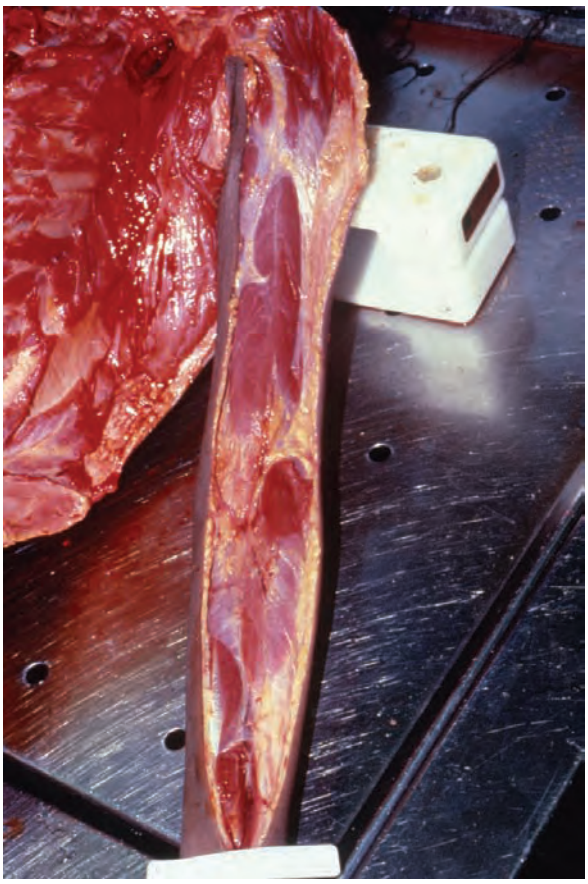

Some cases may require further evaluation including a posterior neck and back dissection. This case involved a 12-year-old African American child who was strangled. Dissection should be done in a layer-by-layer fashion until the surface of bone is exposed. It is important to dissect the arms as well, which may demonstrate contusions from being held during a struggle. There were no hemorrhages found in this case. It is often more difficult to externally visualize contusions in darker-skinned individuals.

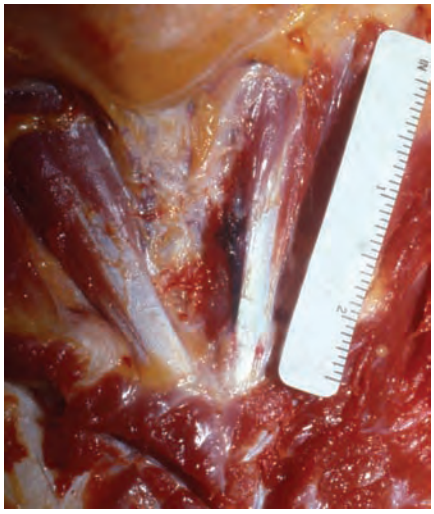

Strangulation. Hemorrhage of the medial left sternocleidomastoid muscle.

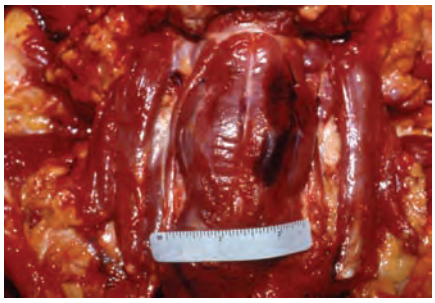

Strangulation. Note the large hemorrhage to the anterior neck structures including the left sternohyoid muscle.

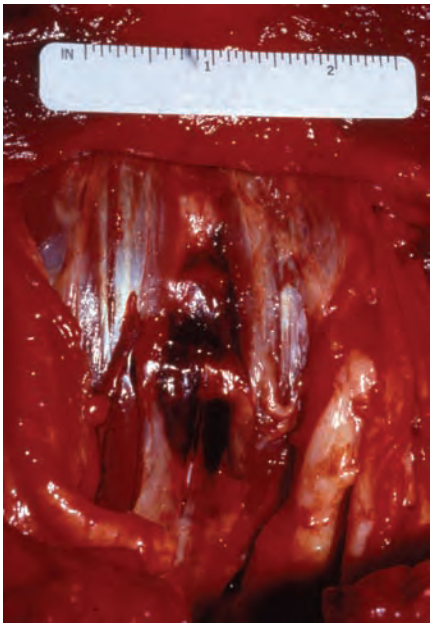

Strangulation. Note the hemorrhage in the anterior neck structures.

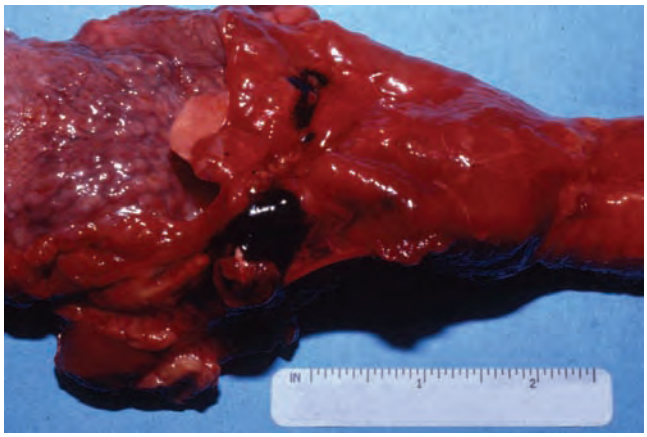

Strangulation. Note the hemorrhage overlying the left superior horn of the thyroid cartilage. The underlying cartilage was fractured.

Note the fractures of this hyoid bone with vague hemorrhages in a decedent with slight to moderate decomposition. Fractures of the hyoid bone are often found in association with homicidal strangulation. The presence of a hyoid bone fracture does not indicate that the case must be a strangulation and the absence of fractures to the hyoid bone does not indicate the decedent was not strangled. Fractures can occur as a result of a blunt impact as well.

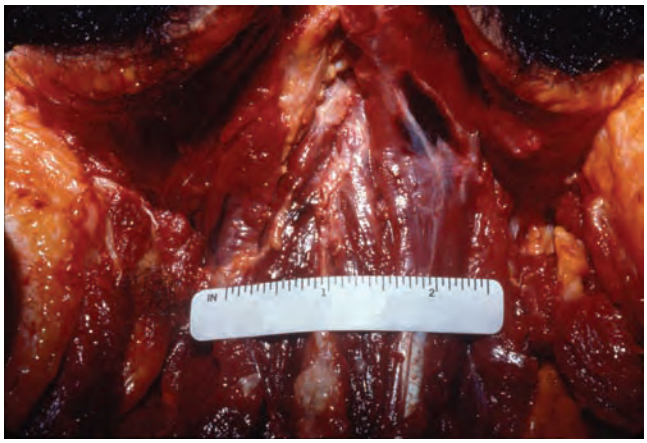

Strangulation. Posterior neck dissection with hemorrhage to the superior aspect of the semispinalis capitis muscle.

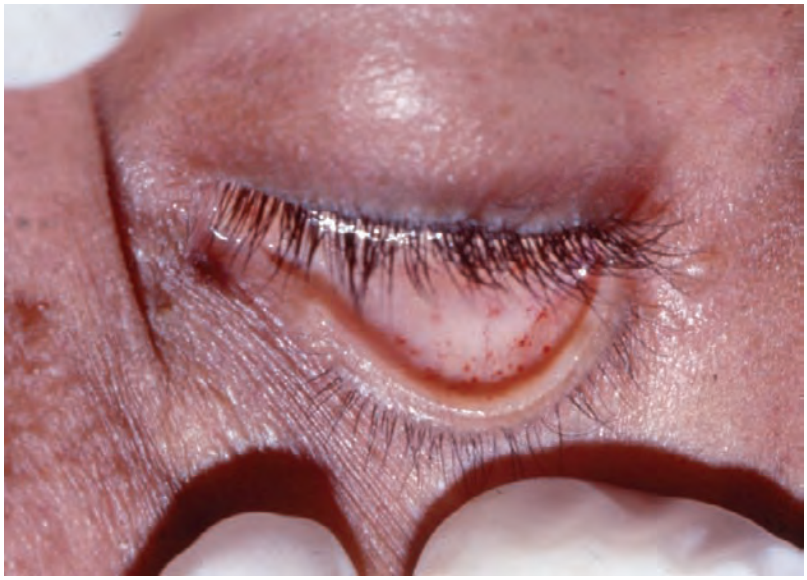

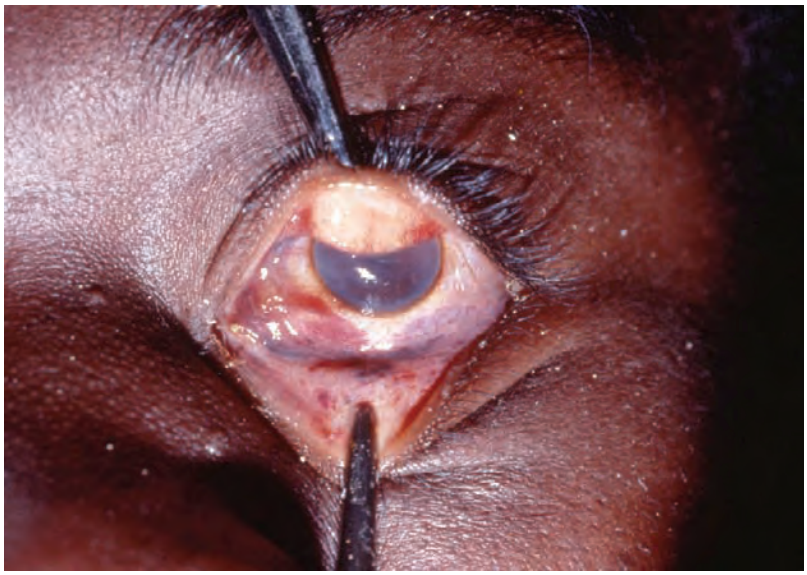

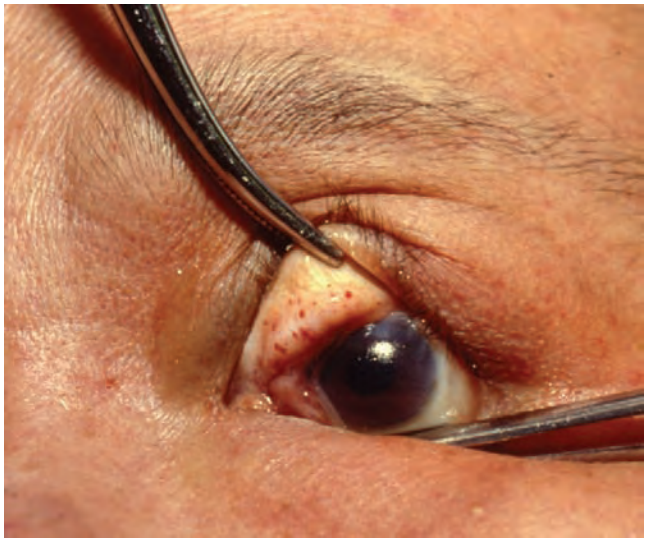

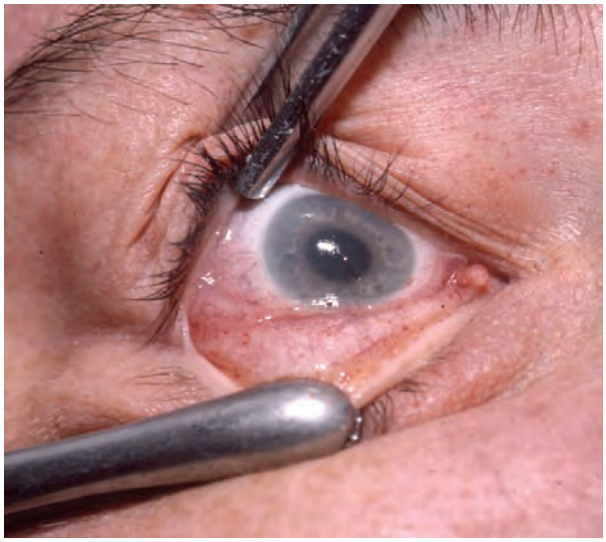

Cases of strangulation demonstrating petechiael hemorrhages of the sclera and conjunctivae.

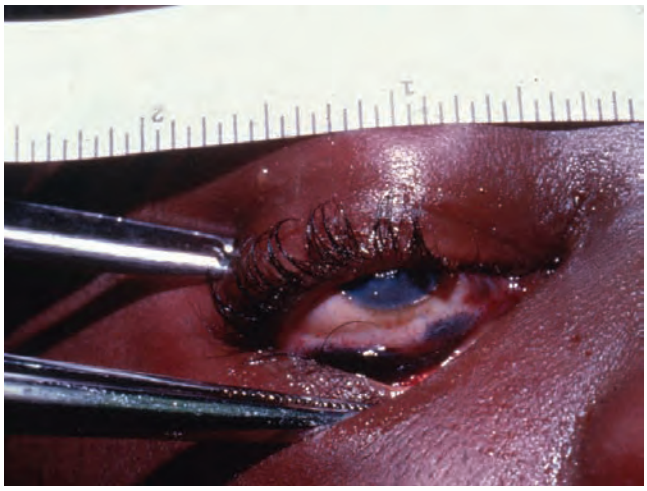

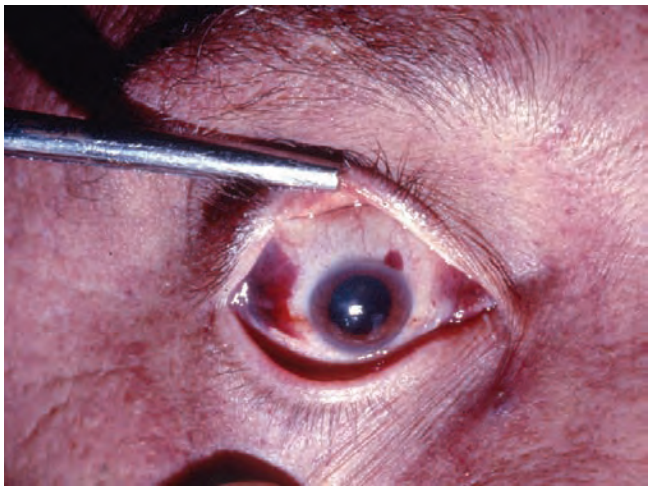

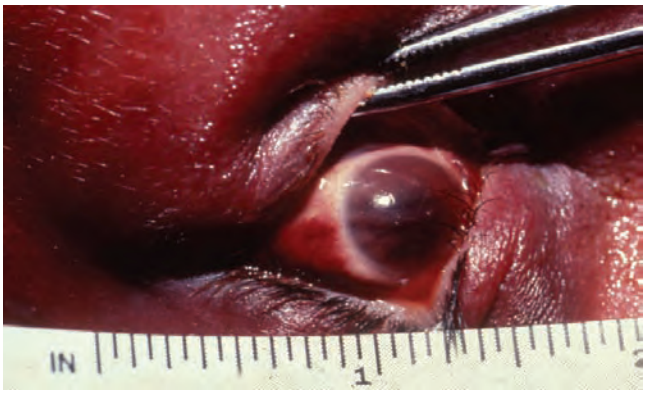

These demonstrate a range of scleral and conjunctival hemorrhages. As varied degrees of neck pressure continue after petechiae formation these hemorrhages progressively become larger and possibly confluent until death ensues.

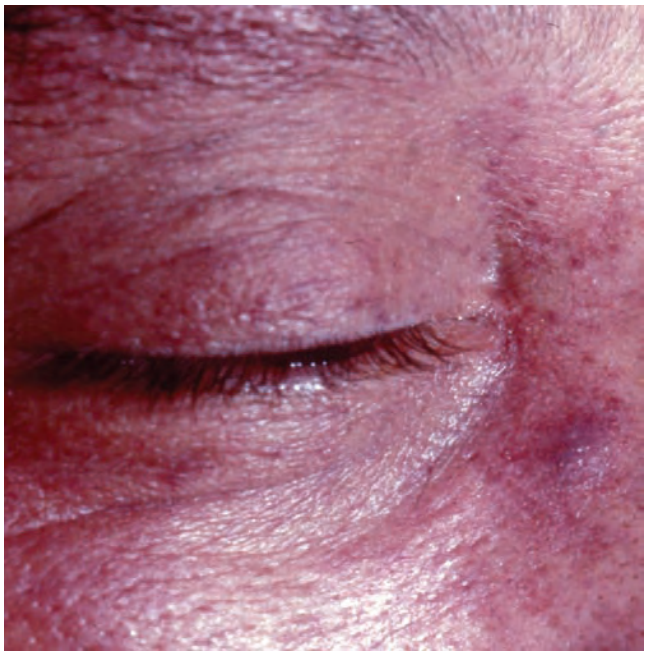

Petechiael hemorrhages of the face.

Strangulation. Petechiael hemorrhages within the mucosal surface of the mouth.

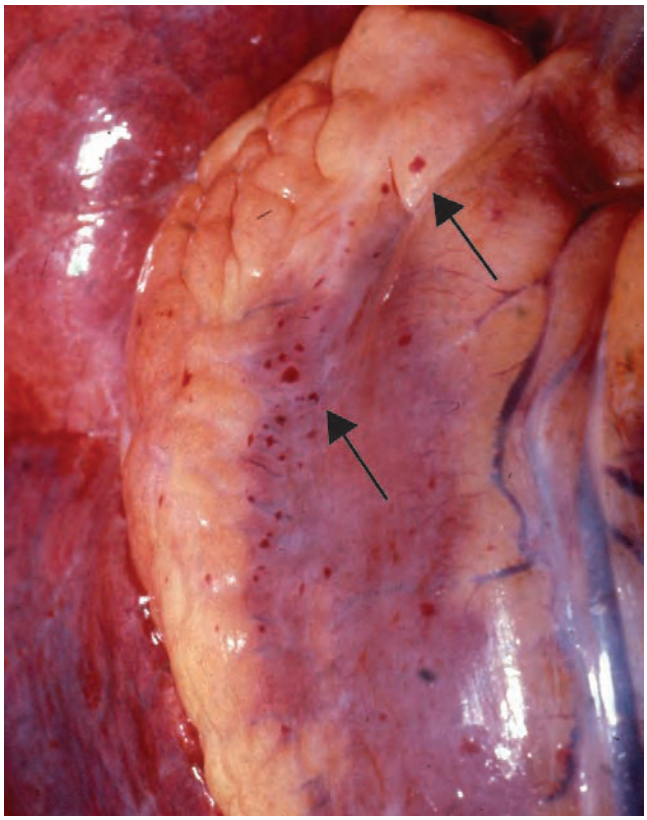

Petechiael hemorrhages at the surface of the heart associated with asphyxia due to chest compression.

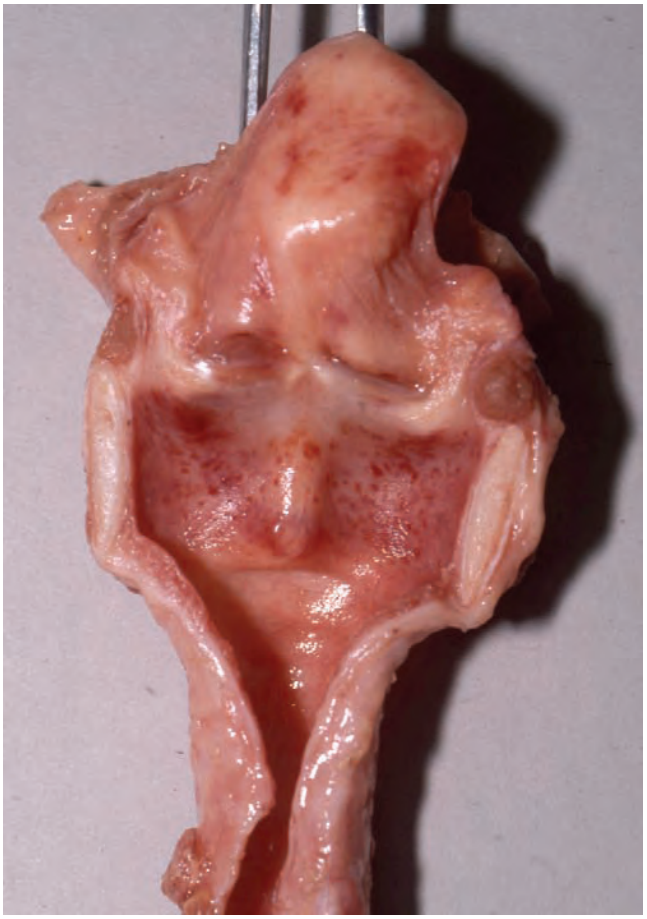

Strangulation. Petechiael hemorrhages within the mucosa of the tracheal and laryngeal cartilage.

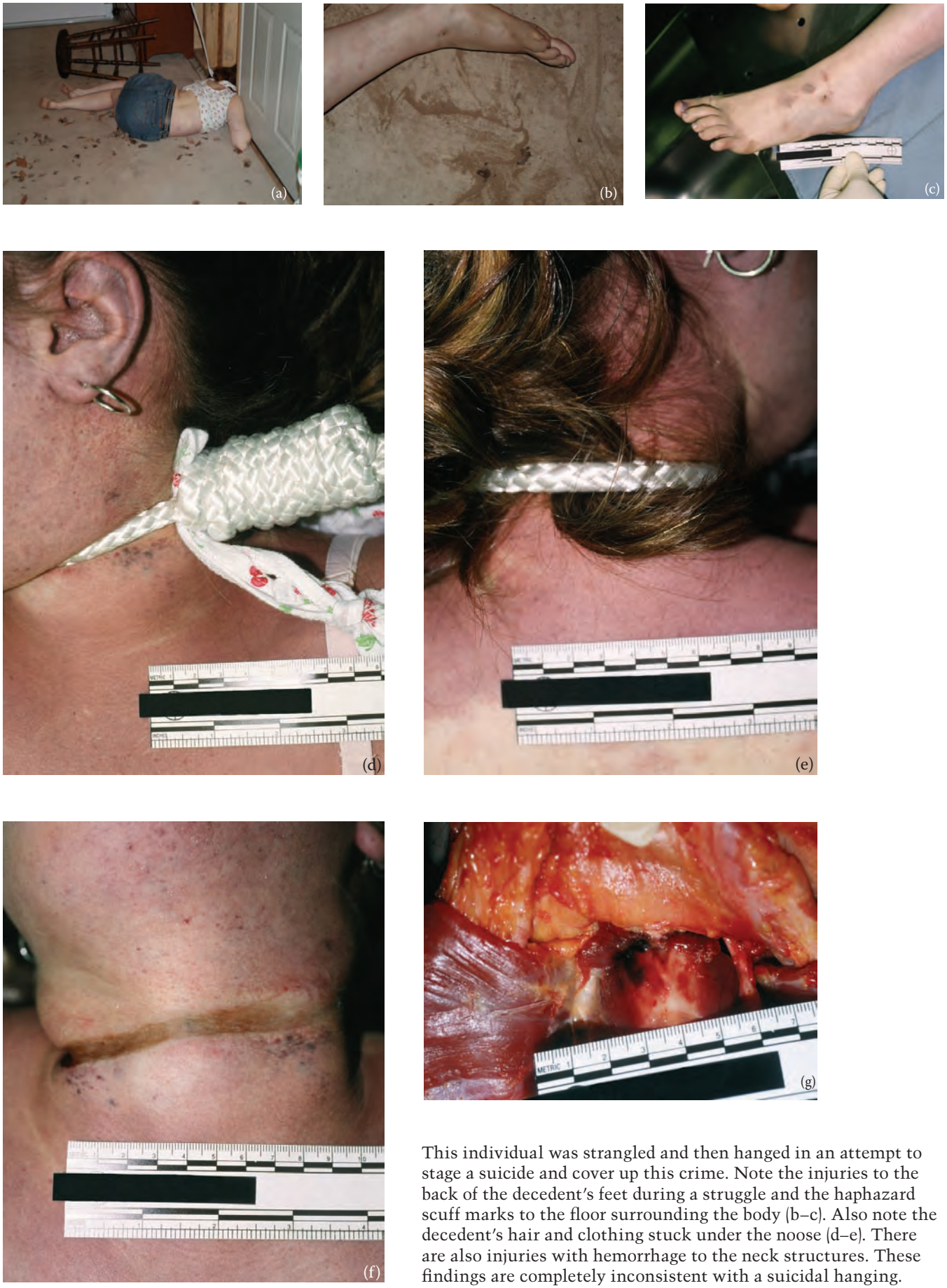

This individual was strangled and then hanged in an attempt to stage a suicide and cover up this crime. Note the injuries to the back of the decedent's feet during a struggle and the haphazard scuff marks to the floor surrounding the body (b–c). Also note the decedent's hair and clothing stuck under the noose (d–e). There are also injuries with hemorrhage to the neck structures. These findings are completely inconsistent with a suicidal hanging.

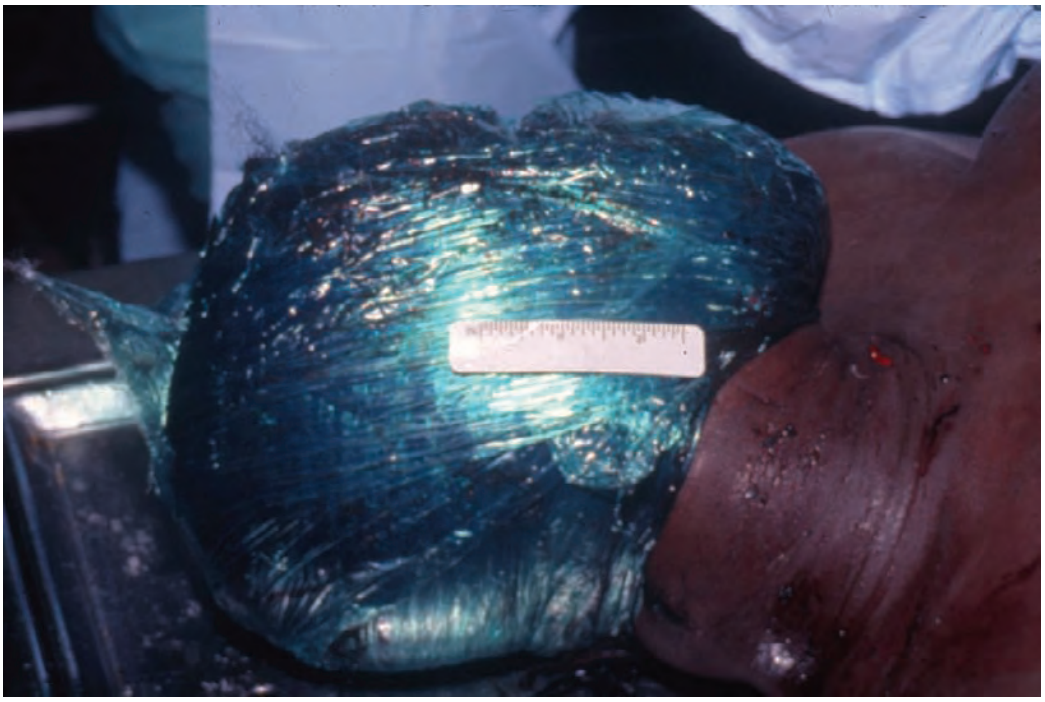

This individual was abducted, bound, and had multiple superficial stab wounds associated with a sexual assault. Her head was wrapped in multiple layers of plastic that covered her mouth and nose.

This individual committed suicide by taking multiple pills and tying a plastic bag tightly over her head. It is the author's experience that it is better to pend these cases for toxicology.

Homicidal asphyxia due to airway obstruction. This individual was found in a moderate to marked state of decomposition within her apartment. She was reportedly dealing drugs and was found tied up with a piece of cloth stuck down her mouth and throat.

Anal trauma with abrasions and lacerations due to sexual assault during strangulation. One should always assume strangulation victims have been sexually assaulted. A rape kit should always be performed in these cases.

This is an airway from a 2-1/2 year-old who was fed a hot dog by her older sibling. The large size of the hot dog piece caused it to get wedged in her throat, leading to asphyxia.

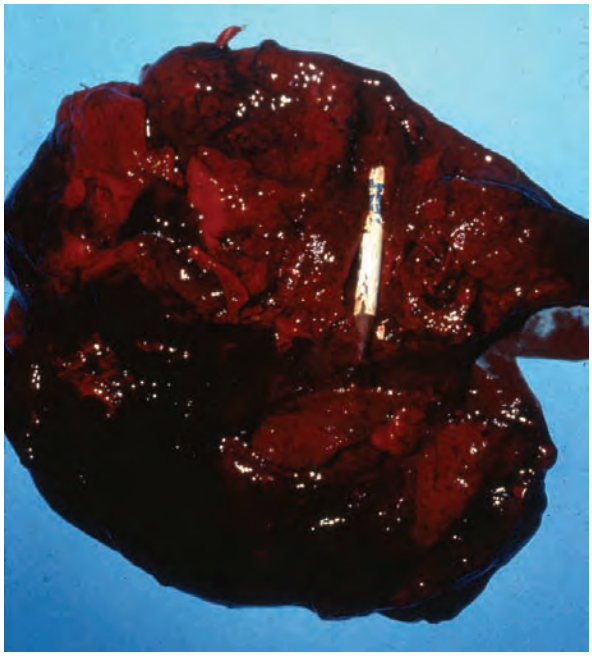

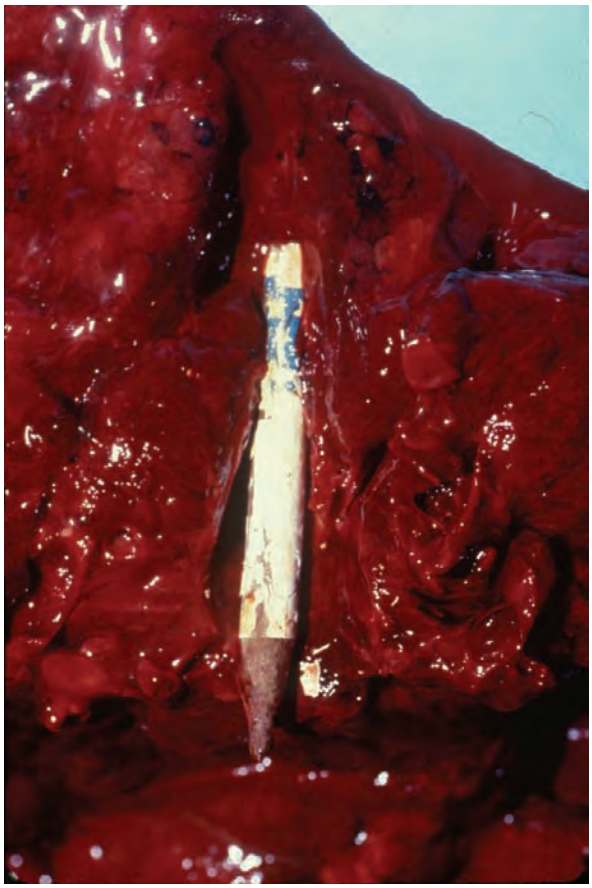

This individual had a pencil in his mouth when he collapsed at work. The pencil was inhaled and wedged in the right mainstem bronchus.

This child was found to have a pebble wedged within her right mainstem bronchus, as evident on this radiograph.

These individuals were trapped in machines that caused chest compression and inability to breathe. One was stuck in a box folding machine, another trapped under a car and another stuck under a single-person elevator lift. Note the imprint of the individual's hand (d) while he was struggling to escape, with the blister formation between these finger marks. Also note the pulmonary edema demonstrated by froth coming out of the mouth (f). This individual also had numerous petechiael hemorrhages of the upper trunk, face, eyes, and visceral organs including the heart.

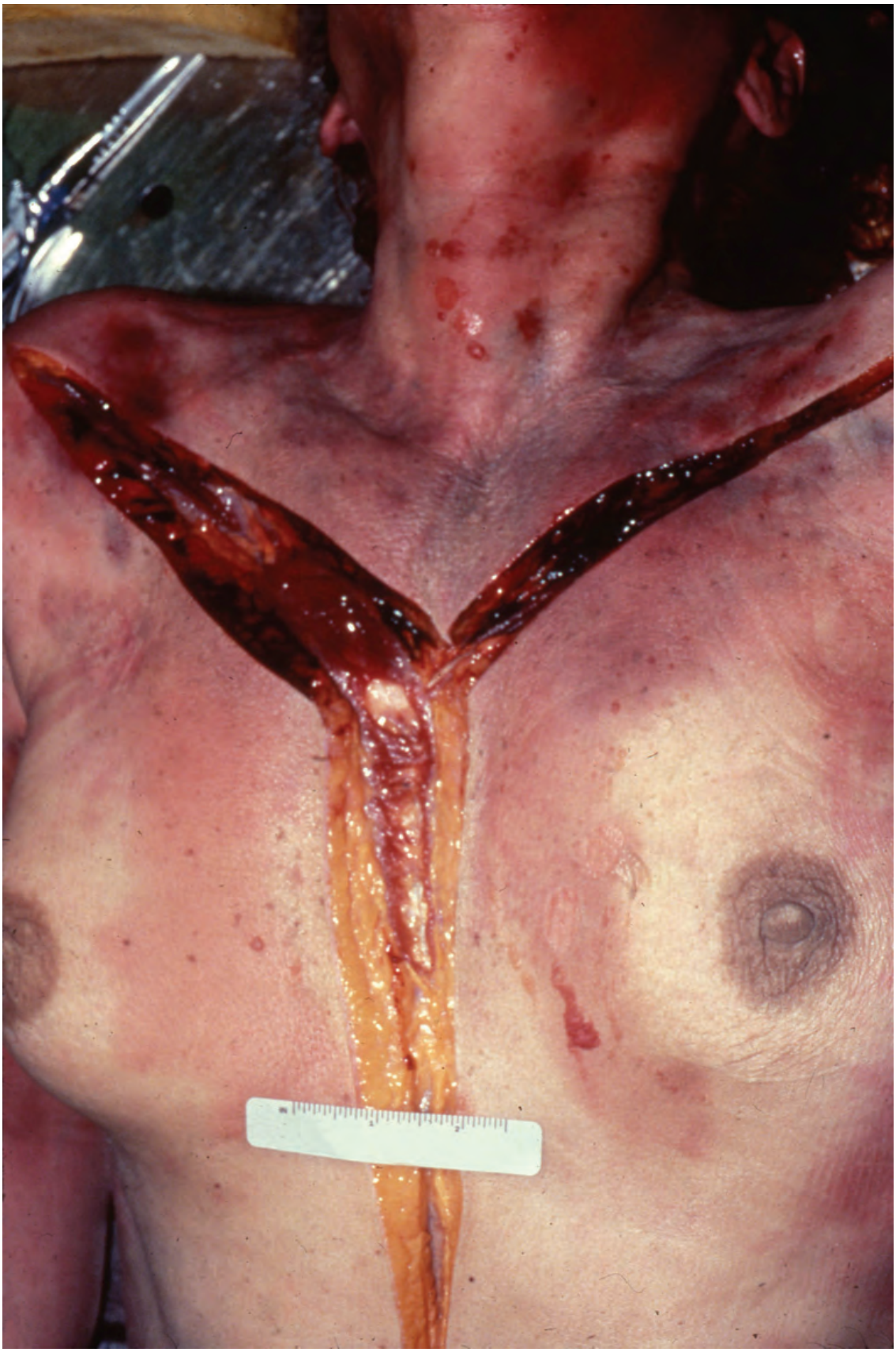

Homicidal asphyxia due to compression of chest and neck. Note the extensive hemorrhage at the superior chest visible at the superior aspect of the Y-shaped incision during autopsy. The lower aspect of this incision is yellow, anemic, and postmortem. This individual was punched, strangled and then the perpetrator sat on her chest during the assault.

Positional asphyxia. These individuals were markedly intoxicated and passed out in positions that prevented them from breathing, thus obstructing blood circulation.

FORENSICS & CRIMINAL JUSTICE

Color Atlas of Forensic Medicine and Pathology

New York City has the largest medical examiner's office in the United States, and the Brooklyn division is the busiest of the five boroughs. Charles A. Catanese received his Forensic Pathology fellowship training in New York, and then worked full time as a Medical Examiner in the Brooklyn office for more than 10 years. He has personally performed more than 4000 autopsies, including over 400 homicides. Dr. Catanese has worked through several disasters, including TWA Flight 800, AA Flight 587, and more than nine months on the World Trade Center fatalities. He is currently the Chief Medical Examiner of Orange County, New York. Drawing on his wealth of knowledge and experience in solving some of the most difficult cases a forensic examiner could encounter, he assembles hundreds of images from his own work experience to present the ***Color Atlas of Forensic Medicine and Pathology.***

Featuring twice the number of images as any other forensic pathology atlas, the book is filled with high-resolution photos that demonstrate postmortem changes of the human body and the different types of patterns produced in deaths caused by

- Natural causes
- Diagnostic or therapeutic procedures
- Substance abuse
- Poisoning
- Child abuse
- Firearms
- Blunt instruments
- Sharp instruments
- Burns
- Asphyxia

This easy-to-read atlas, created for medical and non-medical personnel, covers basic and advanced forensic concepts that relate to all manners of deaths. The carefully worded, unambiguous text describing each photo and the side-by-side comparisons of similar, yet different, pathologies make this remarkable atlas a powerful teaching tool for all those who must confront and solve the mystery of human demise. A fully searchable DVD version is also available.

4320X

ISBN: 978-1-4200-4320-4
90000
9 781420 043204

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

www.crcpress.com

Master Techniques in Orthopaedic Surgery

Soft Tissue Surgery

Steven L. Moran
William P. Cooney, III

Wolters Kluwer | Lippincott Williams & Wilkins

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 1

Why Wounds Fail to Heal

Mark H. Jensen

Steven L. Moran

Surgical wound complications can be a significant problem resulting in prolonged hospital stay, reoperations, and significant morbidity for our patients. Often the surgeon can predict preoperatively which patients will have problems healing wounds or will need additional soft tissue coverage. Surgical planning should not only include the surgical approach but also an assessment of the patient's healing risks. It is our hope that preoperative risk reduction may avoid wound healing complications.

Prevention of wound healing complications starts with an understanding of the healing process. Risk factors can be recognized and modified. Techniques and practices that are proven to prevent wound infections must be employed. This chapter will focus on understanding wound healing and identifying patients at risk for wound complications; we will also attempt to offer means to minimize risk factors for complications through surgical technique and preoperative planning.

Phases of Wound Healing

Surgically induced wounds heal in several stages. The wound passes through phases of coagulation, inflammation, matrix synthesis and deposition, angiogenesis, fibroplasia, epithelialization, contraction, and remodeling. These processes have been grouped into three main stages: inflammation, fibroplasia, and maturation. Interruption in any one of these stages can lead to wound healing complications.

The inflammatory phase of wound healing involves cellular responses to clear the wound of debris and devitalized tissue. Increased capillary permeability and leukocyte infiltration occur secondary to inflammatory mediators and vasoactive substances. Polymorphonuclear cells (PMNs) are the first cell population in the wound followed by mononuclear leukocytes which mature into wound macrophages. Inflammatory cells clean the wound of harmful bacteria and devitalized tissue. Adequate tissue oxygen tension is necessary for the release of oxygen free radicals by neutrophils. Following the initial introduction of PMNs into the wound, lymphocytes enter the wound in great number, clearing the wound of old neutrophils and secreting important cytokines and chemoattractants for fibroblasts. Fibronectin and hyaluronate deposition from fibroblasts in the first 24 to 48 hours provides scaffolding for further fibroblast migration (1,2).

The fibroblast proliferation phase starts within the initial 2 to 3 days as large populations of fibroblasts migrate to the wound. Secretion of a variety of substances necessary for wound healing and includes large quantities of glycosaminoglycans and collagen. Ground substance formed from the four main glycosaminoglycans (hyaluronic acid, chondroitin-4-sulfate, dermatan sulfate, and heparin sulfate) acts as an amorphous gel that is necessary for collagen aggregation. Collagen levels rise for approximately 3 weeks corresponding to increasing tensile strength. After 3 weeks the rate of degradation equals the rate of deposition. Angiogenesis is an important aspect of the fibroblast proliferation phase as it helps to support new cells in the healing wound.

The maturation phase starts around 3 weeks and lasts up to 2 years. It is characterized by collagen remodeling and wound strengthening. Collagen is the principal building block of connective tissue and is found in at least 13 different types. Types I to IV are the most common in the human body. Each has a distinct feature and is found in different levels in many tissues. For example, type III collagen is high in hydroxyproline and low in hydroxylysine. It is commonly found in skin, arteries, bowel wall, and healing wounds. Type I collagen is found in skin, tendon, and bone; is low in hydroxylysine content; and is the most common collagen type, accounting for more than 90% of body collagen. Early wounds are comprised of a majority of type III collagen. As the wound matures, type III collagen is replaced by type I collagen.

Collagen cross-linking improves tensile strength. There is a rapid increase in strength of the wound by 6 weeks as the wound reaches 70% of the strength of normal tissue. The wound then gradually plateaus to 80% of normal strength, but never returns to preinjury levels.

Wound re-epithelialization occurs as adjacent cells migrate through a sequence of mobilization, migration, mitosis, and cellular differentiation of epithelial cells. Wound contraction starts at about 1 week. It is facilitated by the transformation of certain fibroblasts into myofibroblasts containing Î±Ã—smooth muscle actin. These cells adhere to the wound margins as well as each other and effect contraction of the wound. These stages are imperative for proper wound healing as interruption of these processes results in chronic wound complications (1,3).

Risk Factors

The identification of patients at risk for aberrant wound healing allows the surgeon to make appropriate plans for skin closure technique, flap utilization, and postoperative wound management. This will ideally result in modification of risk factors prior to surgery. In cases of chronic diseases or nonmodifiable risk factors, patients must be informed of increased wound healing risks. The following discussion focuses on commonly encountered risk factors with recommendations to ameliorate their effects.

Diabetes

Patients with diabetes are both more likely to undergo surgery and develop perioperative complications than nondiabetic patients. This leads to longer hospital stay with higher health care costs and increased perioperative mortality. Diabetic patients have increased rates of hypertension, cardiac disease, and renal failure. These factors lead to much higher wound healing complications (3). Diabetes inhibits wound healing through many mechanisms. It is a disease affecting small vessels which are critical in supplying nutrients to the healing wound. Elevated glucose levels also affect a myriad of inflammatory systems. Neutrophil adherence, chemotaxis, phagocytosis, and intracellular bactericidal activity are all impaired. Pseuodhypoxia develops as a result of altered redox reactions and vascular permeability secondary to hyperglycemia. Furthermore, glucose is a proinflammatory mediator stimulating cytokine production and inhibiting endothelial nitric oxide levels (4). This translates

clinically into higher infectious complications. *Tight control of glucose in the perioperative period mitigates the postoperative complications seen in the diabetic patient*. This has been found in both the intensive care setting as well as routine operative cases (3,5).

Management of patients with diabetes starts in the pre-operative period. There is good evidence to suggest that a preoperative hemoglobin A1C of less than 7% drastically decreases postoperative infectious complications (4). Physicians should aggressively improve glucose management using diet, oral hypoglycemic agents, and insulin as needed. Perioperative management should include sliding scale insulin or continuous insulin infusion to maintain glucose levels below 150 mg per deciliter (3). Evidence suggests that improved outcomes are possible with tighter control of glucose levels between 80 and 110 mg per deciliter in critically ill patients (5). After dismissal from the hospital, the patient should continue their preoperative diabetic regiment.

Obesity

Obese patients have a higher rate of wound infections, dehiscence, hematomas, seromas, and pressure ulcers (6Ã—8). This is felt to be due to multiple factors including difficulty with the operation, altered immune response, increased tension with closure, increased dead space, decreased microperfusion, and decreased mobility post operatively (9,10). Patients should be encouraged to lose weight prior to elective operations. Increased levels of physical activity preoperatively translates to an improved postoperative rehabilitation process. Successful weight loss is difficult to achieve in most patients. This has prompted some surgeons to recommend gastric bypass surgery prior to certain operations such as joint replacement. This has been associated with improved outcomes in hip replacement patients with morbid obesity (11).

Smoking

Numerous studies have consistently found that smokers have significantly higher rates of wound healing complications than nonsmokers. This is related to several causes including the vasoactive effect of cigarette smoke through the sympathetic alpha receptors, increased levels of carboxyhemoglobin with a reduction in oxygen-carrying

capacity, increased platelet activation leading to microangiopathic thrombosis, increased levels of fibrinogen with decreased fibrinolytic activity, endothelial injury, and increased hemoglobin levels leading to increased blood viscosity. Regardless of the mechanism the effect is significant. Complication rates of 2.5% to 6% in nonsmokers versus 7.5% to 49% in smokers are reported in the literature.

The increased risk of wound complications in smokers is most pronounced in cases where a large area of tissue is undermined. This is most likely related to failure of the dermal and subdermal plexus supplying the resultant skin flap (Fig. 1-1). Poor outcome in smokers prompts many clinicians to postpone elective procedures until the patient has quit smoking for at least 3 to 4 weeks, particularly if the procedure involves large areas of undermining (12,13). Aggressive use of smoking cessation

programs should be implemented. Still, other surgeons feel that this would restrict too many patients from receiving necessary surgery (14,15). We prefer to delay surgery for 4 weeks prior to elective surgical procedures where soft tissue coverage may be an issue and advocate early referral to a nicotine dependence unit. If smoking cessation is not possible, the patient is informed of the likelihood of additional flap coverage for wound closure and a higher likelihood of postoperative wound complications.

FIGURE 1-1 Following total knee arthroplasty this 76-year-old diabetic smoker developed wound complications necessitating wound debridement and flap coverage.

FIGURE 1-2 Chronic steroid use leads to increased skin and vessel fragility. This skin avulsion injury occurred in this 76-year-old steroid dependent man following a fall from a chair.

Immunosuppressive Medications

Patients with intrinsic or acquired immunodeficiencies are at increased risks for wound healing complications. Patients with MHC class-II deficiency have impaired wound healing because of altered T cell immune function (16). Patient populations at risk for wound healing problems due to altered immune function include patients with hereditary, infectious, and iatrogenic immune deficiencies.

Corticosteroids inhibit wound healing by their anti-inflammatory effect. Decreased numbers of inflammatory cells are noted at the wound site; delays in collagen synthesis, fibroblast proliferation, angiogenesis, wound contracture rates, and epithelial migration are also observed (Fig. 1-2). Vitamin A has been shown to counteract the effects of steroids on wound healing in all areas except wound contraction and infection (17). The exact mechanism is not known, but may be related to the TGF-beta, IGF-I, and hydroxyproline content in the tissue (18). Factors that can help improve wound healing in immunocompromised states include prevention of malnutrition, hypoxia, endocrine disorders, anemia, and other metabolic disorders. Dead tissue, foreign bodies, tissue ischemia, and hematoma should be minimized.

Radiation and Chemotherapy

Radiation and chemotherapy are known to cause delays in wound healing. Operations in irradiated fields are particularly problematic due to dense fibrosis and decreased perfusion caused by small vessel injury (Fig. 1-3). Irradiated fields are more susceptible to infection and delayed healing. Postoperative radiation initiated after the initial 3 to 4 weeks of primary wound healing does not seem to have as marked an effect on wound healing, but can lead to contracture, wound break down, and flap necrosis (21). Neoadjuvant chemotherapy can alter wound healing, especially in cases where chemotherapy has led to neutropenia. Surgery would be ideally delayed until full recovery of platelets and leukocytes. Healing seems to proceed normally in patients who receive their chemotherapy 3 to 4 weeks after surgery, as the wound has been allowed to proceed through the first stages of healing (8,22). Certain chemotherapeutic agents, however, can have a negative effect on wound healing far greater than 4 weeks; for example, Avastin (Bevacizumab), which inhibits vascular endothelial growth factor, has an extremely long half life ranging from day 11 to 50 and has been shown to inhibit wound healing when given in the neoadjuvant setting (23). Surgeons should have a good understanding of specific complications associated with individual chemotherapeutic regiments.

Malnutrition

It has been known for many years that malnutrition has deleterious effects on wound healing. Loss of nutrients alters host immunity through decreased T-cell function, phagocytosis, complement, and antibody levels. Certain patient groups are at particular risk of malnutrition. Severe catabolic states can be induced after multi-system trauma, sepsis, and burns. Significant increases in metabolic rate

can also follow uncomplicated abdominal surgery (increase by 10%), uncomplicated injuries such as femoral fracture (20%), peritonitis (40%), and fever (10% for every 1Â°C above normothermia). This is particularly alarming considering the high portion of trauma patients who are unable to eat for extended periods leading up to and following surgery (24,25).

FIGURE 1-3 Signs of trouble in this 80-year-old gentleman who has undergone preoperative radiation with subsequent resection of a spinal tumor and coverage with a latissimus flap. Spinal tumor resection and flap harvest were preformed through separate and parallel incisions. The parallel incision in conjunction with the radiation damage has lead to ischemia within the center portion of skin bridge, which now shows signs of early ischemia with epidermolysis and ulceration.

Assessment of nutritional status becomes an important task for the surgeon. Patients at risk include those with moderate weight loss (10%Ã—20%), severe weight loss (>20%) and serum albumin <3.2 to 2.5 g/dL. Other groups include patients with trauma, burns,

gastrointestinal (GI) dysfunction, cancer, fever, and those on chemotherapy. Consideration for preoperative total parenteral nutrition (TPN) should be given to patients who are severely malnourished, however this must be weighted against the risks associated with line infection and liver failure (24). Prealbumin is a laboratory value that can be followed as a surrogate of albumin due to its shorter half life.

Postoperatively patients should be started on a diet as soon as possible. Rapid diet advancement has been shown to decrease postoperative complications and decrease hospital stay (26). In patients who are unable to restart a diet, but have a functioning GI tract, tube feeds should be administered. TPN should be started early on patients who are expected to not tolerate GI feeds. The benefit risk ratio of TPN is equal for patients with normal metabolic demands after 1 week of fasting or inability to take oral intake. Patients with hypermetabolic states can benefit from earlier initiation of TPN (25,26).

Peripheral Vascular Disease

Peripheral artery disease is present in an estimated 0.9% of patients aged 40 to 49 compared with 15% to 29% of patients over the age of 70. Risk factors for the development of PVD include diabetes mellitus, hyperlipidemia, cigarette smoking, and hypertension. Patients at risk for PVD should be screened with blood pressure measurements of the arm and ankle to determine the patient's ankle brachial index; a normal value should produce a ratio >0.9. Symptoms of PVD include claudication, rest pain, atypical leg pain, and ischemic tissue loss. Aggressive lifestyle and risk factor modification is indicated to prevent progression of disease and death from cardiovascular or cerebrovascular events. These patients should undergo cardiac risk evaluation as peri-operative myocardial infarction (MI) and stroke are very common (27).

Wounds in the setting of peripheral vascular disease heal very slowly and can be difficult to manage. Patients with nonhealing wounds in spite of maximal medical management should undergo evaluation for revascularization. Amputation may be considered in the absence of target vessels for revascularization (27). Transcutaneous oxygen measurements in peripheral occlusive disease can be

helpful to determine healing potential. $TcPO_2$ levels greater than 30 mm Hg correlate

with a high likelihood of healing whereas levels less than 15 mm Hg rarely heal (28,29).

Infection

Surgical site infections are a major impediment to wound healing and contribute to substantial morbidity and mortality. Postoperative infections increase the hospital length of stay by 7 to 10 days and increase charges by $2,000 to $5,000 (30). Rates of surgical infections vary by type of procedure, health of the patient, and skill of the surgeon. Wound classification corresponds with expected rates of infection as follows: clean (1.3%Ã—2.9%), clean-contaminated (2.4%Ã—7.7%), contaminated (6.4%Ã—15.2%), and dirty (5.1%Ã—40%) (6). Other risk factors for the development of postoperative infections include diabetes, obesity, smoking, steroids, malnutrition, colonization with *S. aureus*, preoperative hospitalization, and patient health.

The most important factors which can be used to prevent surgical site infections are timely use of preoperative antibiotics and operative technique. Other practices that are helpful include preoperative showering with antimicrobial soap, scrubbing and draping the patient with sterile drapes, hand washing, gloving, and the use of sterile gowns, masks, and hats by surgical personnel (31). These practices are aimed at reducing the amount of skin associated bacteria. Pathogens are not eliminated because approximately 20% of bacteria reside in the hair follicles and sweat glands where antiseptics do not reach. Although numerous studies have been aimed at identifying the best skin antiseptic, a systematic review failed to show superiority of one antiseptic over another (32).

Preoperative antibiotics administered within 2 hours of incision correlate strongly with the lowest rate of infections. Choices of antibiotics along with common pathogens are listed in Table 1-1. Endocarditis prophylaxis may need to be added in susceptible patients. In spite of clear benefits with the use of preoperative antibiotics, compliance is not perfect. Institutional policies should be constructed to promoting 100% compliance (33).

Hair removal is commonly performed prior to surgical procedures, however most studies have shown that this is not necessary and is associated with an increased risk of surgical site infections (34). If hair removal is preferred, clippers or depilatory creams are safer than shaving and should be used just prior to skin incision. Razors are associated with microtrauma to the skin and should not be used secondary to increased

infection rates (35).

TABLE 1-1. Antimicrobial Prophylaxis for Surgery

Nature of operation	Common pathogens	Recommended Antimicrobials	Adult dosage before surgery[1]
Cardiac	*Staphylococcus aureus,*	cefazolin or	1-2 g IV[2]
	S. epidermis	cefuroxime	1.5 g IV[2]
		OR vancomycin[3]	1 g IV
Gastrointestinal			
Esophageal,	Enteric gram-negative	*High risk[4] only:*	
gastroduodenal	bacilli, gram positive cocci	cefazolin[7]	1-2 g IV
Biliary Tract	Enteric gram-negative bacilli,	*High risk[5] only:*	
	enterococci, clostridia	cefazolin[7]	1-2 g IV

Colorectal	Enteric gram-negative bacilli,	*Oral:*neomycin	
	anaerobes, enterococci	+ erythromycin base[6] OR metronidazole[6]	
		Parenteral:	
		cefoxitin[7]	1-2 g IV
		OR cefazolin	1-2 g IV
		+ metronidazole[7]	0.5 g IV
		OR ampicillin/sulbactam	3 g IV
Appendectomy,	Enteric gram-negative bacilli,	cefoxitin[7]	1-2 g IV
non-perforated[8]	anaerobes, enterococci	OR cefazolin	1-2 g IV
		+ metrodinazole[7]	0.5 g IV
		OR ampicillin/sulbactam[7]	3 g IV

Genitourinary	Enteric gram-negative bacilli, enterococci	*High risk only*[9]*:* ciprofloxacin	500 mg PO or 400 mg IV
Gynecologic and Obstretric			
Vaginal, abdominal	Enteric gram-negative bacilli,	cefoxitin[7] or cefazolin[7]	1-2 g IV
or laparoscopic	anaerobes, Gp B strep,	OR ampicillin/sulbactam[7]	3 g IV
hysterectomy	enterococci		
Cesarean section	same as for hysterectomy	cefazolin[7]	1-2 g IV
			after cord clamping
Abortion	same as for hysterectomy	*First trimester, high risk*[10]*:*	
		aqueous penicillin G	2 mill units IV
		OR doxycycline	300 mg PO[11]

		Second Trimester:	
		cefazolin[7]	1-2 g IV
Head and Neck Surgery			
Incisions through	Anaerobes, enteric gram-	clindamycin	600-900 mg IV
oral or pharyngeal	negative bacilli, *S. aureus*	+ *gentamicin*	1.5 mg/kg IV
mucosa		OR cefazolin	1-2 g IV
Neurosurgery	*S. aureus, S. epidermis*	cefazolin	1-2 g IV
		OR vancomycin[3]	1 g IV
Ophthalmic	*S. epidermis, S. aureus,*	gentamicin, tobramycin,	multiple drops
	streptococci, enteric gram-	ciprofloxacin, gatifloxacin	topically over
	negative bacilli,	levofloxacin, moxifloxacin,	2 to 24 hours

	Pseudomonas spp.	ofloxacin or neomycin-gramicidin-polymyxin B	
		cefazolin	100 mg subcon-junctivally
Orthopedic	*S. aureus, S. epidermis*	cefazolin[12]	1-2 g IV
		or cefuroxime[12]	1.5 g IV
		OR vancomycin[3,12]	1 g IV
Thoracic (Non-Cardiac)	*S. aureus, S. epidermis,*	cefazolin or	1-2 g IV
	streptococci, enteric gram-	cefuroxime	1.5 g IV
	negative bacilli	OR vancomycin[3]	1 g IV
Vascular			
Arterial surgery	*S. aureus, S. epidermis,*	*cefazolin*	1-2 g IV
involving a	enteric gram-	OR vancomycin[3]	1 g IV

prosthesis,	negative bacilli		
the abdominal aorta,			
or a groin incision			
Lower extremity	*S. aureus, S. epidermis,*	cefazolin	1-2 g IV
amputation for ischemia	enteric gram-negative bacilli,	OR vancomycin[3]	1 g IV
	clostridia		

Reproduced from Antimicrobial prophylaxis for surgery. *Treat Guide Med Lett* 2006;4(52):83Ã—88, with permission.

1. Parenteral prophylactic antimicrobials can be given as a single IV dose begun 60 minutes or less before the operation. For prolonged operations (>4 hours), or those with major blood loss, additional intraoperative doses should be given at intervals 1Ã—2 times the half-life of the drug for the duration of the procedure in patients with normal renal function. If vancomycin or a fluoroquinolone is used, the infusion should be started 60Ã—120 minutes before the initial incision in order to minimize the possibility of an infusion reaction close to the time of induction of anesthesia and to have adequate tissue levels at the time of incision.
2. Some consultants recommend and additional dose when patients are removed from bypass during open-heart surgery.
3. Vancomycin is used in hospitals in which methicillin-resistant S. aureus and

S. epidermis are a frequent cause of postoperative wound infection, for patients previously colonized with MRSA, or for those who are allergic to penicillins or cephalosporins. Rapid IV administration may cause hypotension, which could be especially dangerous during induction of anesthesia. Even when the drug is given over 60 minutes, hypotension may occur; treatment with diphenhydramine (Benadryl, and others) and further slowing of the infusion rate may be helpful. Some experts would give 15 mg/kg of vancomycin to patients weighing more than 75 kg, up to a maximum of 1.5 g, with a slower infusion rate (90 minutes for 1.5 g). To provide coverage against gram-negative bacteria, most Medical Letter consultants would also include cefazolin or cefuroxime in the prophylaxis regimen for patients not allergic to cephalosporins; ciprofloxacin, levofloxacin, gentamicin, or aztreonam, each one in combination with vancomycin, can be used in patients who cannot tolerate a cephalosporin.

4. Morbid obesity, esophageal obstruction, decreased gastric acidity or gastrointestinal motility.

5. Age >70 years, acute cholecystisis, non-functioning gall bladder, obstructive jaundice or common duct stones.

6. After appropriate diet and catharsis, 1 g of neomycin plus 1 g of erythromycin at 1 PM, 2 PM and 11 PM or 2 g of neomycin plus 2 g of metronidazole at 7 PM and 11 PM the day before an 8 AM operation.

7. For patients allergic to penicillins and cephalosporins, clindamycin with either gentamicin, ciprofloxacin, levofloxacin or aztreonam is a reasonable alternative.

8. For a ruptured viscus, therapy is often continued for about five days. Ruptured viscus in postoperative setting (dehiscence) requires antibacterials to include coverage of nosocomial pathogens.

9. Urine culture positive or unavailable, preoperative catheter, transrectal prostatic biopsy, placement of prosthetic material.

10. Patients with previous pelvic inflammatory disease, previous gonorrhea or multiple sex partners.

11. Divided into 100 mg one hour before the abortion and 200 mg one half hour later.

12. If a tourniquet is to be used in the procedure, the entire dose of antibiotic

must be infused prior to its inflation.

Hypothermia has benefits of decreasing tissue oxygen consumption and is tissue protective in cardiac bypass and organ transplant. However, hypothermia may predispose to infection through cutaneous vasoconstriction. It has been shown that perioperative normothermia is associated with reduced rates of surgical site infections (36).

Several health care improvement initiative studies have shown that surgical site infection surveillance is effective in reducing the infection rates (37). The mechanism is unknown, but many researchers postulate that surveillance is associated with a greater awareness of sterile technique. Operative teams more rapidly recognize problems and find solutions. Surgeons should seek a program at their institution for health care improvement and operative safety. The Institute for Healthcare Improvement (IHI) is a nonprofit organization that leads health care improvement initiatives. They organized the 100,000 Lives Campaign to implement changes in United States hospitals which focused on increased compliance to four areas: appropriate use of antibiotics, removing razors from operative rooms, maintaining glucose control, and ensuring perioperative normothermia (38). Hospitals in compliance with these changes reported a 27% reduction in surgical site infections (7).

Surgical Technique

Numerous studies aimed at discovering risk factors for complications have identified surgeons as independent risk factors. This could be due to numerous factors including clinical judgment, baseline characteristics of patient populations, procedure type, and surgical technique. There is no doubt that meticulous surgical technique can improve outcomes. This is particularly important when dealing with a patient at risk for complications. Excess intraoperative blood loss has been associated with surgical site infections and poor wound healing (39). Blood loss leading to hypotension with subsequent vasoconstriction and tissue hypoxia is one explanation for increased wound healing complications. Techniques aimed at improving sterility, minimizing tissue destruction, maintaining meticulous hemostasis, improving speed and accuracy,

minimizing foreign body placement, and attention to closure are expected to reduce complications.

The choice of suture and method of suture placement can also have an impact on wound healing. There is a wide array of suture types. In general, sutures can be divided by life expectancy (permanent, absorbable, fast absorbable) and consistency (braided vs. monofilament). Permanent suture is indicated in situations where suture absorption is not desired (e.g., vasculature anastomosis). Drawbacks include development of suture abscesses and chronic infection. Stitch abscesses are less problematic with absorbable sutures, but may lose strength prior to tissue healing. Braided sutures are strong, hold well, and are easy to tie. However, they are more prone to infection than monofilament. Braided sutures may also cause tissue tearing as in the case of venous repairs. Monofilament sutures are smooth and resist infection, but are less forgiving and require more knots to resist slippage.

Placement of sutures can also contribute to wound compromise. Simple interrupted sutures and vertical matrass sutures produce less wound edge ischemia than running and horizontal matrass sutures when tissue viability is in question. Horizontal matrass sutures placed under excessive tension can further impede blood supply to wound edges resulting in infection, necrosis, or dehiscence. Care must be taken not to incorporate or impinge vasculature pedicles when placing sutures.

Conclusion

Prevention of wound complications starts with the preoperative evaluation where risk factors for wound healing are identified and modified prior to elective procedures. In the traumatic setting and in the established soft tissue deficit, patient factors such as glucose control and nutritional status can be optimized to maximize the patients healing potential. With this chapter as an introduction, the following chapters will address evaluation of traumatic wounds followed by techniques for recruiting or mobilizing surrounding tissue to close large wounds and soft tissue deficits.

References

1. Falanga V. Wound healing and its impairment in the diabetic foot. *Lancet.* 2005;66(9498):1736Ã—1743.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 2

Initial Evaluation and Management of Complex Traumatic Wounds

Patricia L. McKay

George Nanos

The principles in this chapter are forged from our experience with the treatment of high energy contaminated wounds resulting from recent overseas military conflicts. The soft tissue injuries discussed within this chapter encompass a broad spectrum of clinical presentations, both acute and chronic. These treatment protocols reflect the ever-growing body of scientific knowledge and technological advances contributing to the successful treatment of our patients. In the following illustrations, we hope to point out the common mistakes and potential pitfalls of operative and perioperative wound management and provide surgeons with a framework for successful treatment of all types of wounds.

Initial Evaluation

A thorough initial patient evaluation gives the surgeon the information by which to formulate a treatment plan with the best chance for success. A complete and accurate history must include the circumstances leading to the current wound presentation, the mechanism of injury if associated with trauma, underlying medical conditions, current occupational and socioeconomic status, and patient social habits such as smoking which may have detrimental effects on reconstructive efforts (Table 2-1).

In cases of trauma the energy level, mechanism, location, and the time course from injury to presentation are invaluable in predicting prognosis and planning treatment. Medical comorbidities such as underlying cardiopulmonary or peripheral vascular disease, endocrinopathy, neuropathic disease, immunocompromising conditions, psychiatric illness, nutritional deficits, and allergies can negatively impact the success of treatment if not identified and optimized. Medical consultation and comanagement is recommended in these cases. Tetanus status must be addressed in accordance with Centers for Disease Control guidelines.

Failure to fully recognize the scope of the wound and/or injury can have dreadful consequences to the outcome. A detailed physical examination with critical assessment of vital signs, secondary survey, and multi-system examination is required in all cases to evaluate the wound and exclude other medical conditions or associated injury that may take greater priority in treatment.

In multi-system trauma, the Advanced Trauma Life Support (ATLS) approach, beginning with airway, breathing, and circulation (ABC's) is recommended. Traumatologists or critical care specialists should be the principle coordinators of all initial medical care to ensure appropriate global management of the patient.

Critical attention should be directed to vascular status. If capillary refill and/or pulses are abnormal or absent, or if the mechanism introduces suspicion for vascular injury, Doppler testing and/or perfusion studies must be obtained. Revascularization or repair of vessels should be completed prior to, or concurrent with, surgical wound treatment. Special attention should also be paid to the neurological examination, and all deficits clearly documented. In cases of extremity trauma, a high index of suspicion for compartment syndrome is essential. Continued reevaluation is required with adjunctive use of compartment pressure monitoring if the diagnosis is in question, emergent compartment release should be performed if clinically indicated, and prophylactic release should be considered in trauma cases with revascularization procedures.

Table 2-1. Wound Treatment Pitfalls

Inadequate clinical and surgical resources for proper treatment	Failure to recognize and treat infection
Failure to recognize and optimize host factors	Wound closure with excessive tension
Failure to recognize and treat vascular compromise	Prominent bone or hardware
Inadequate debridement	

Ancillary studies aid in full comprehension of the clinical situation. Plain radiographs should be routinely obtained to evaluate for associated fracture, foreign body, exostosis, osteomyelitis, soft tissue emphysema, vascular calcifications such as those associated with Diabetes Mellitus, or other factors contributing to or resulting from the overlying wound. CT scans and MRI may provide additional valuable information. Markers of hematologic and immune status, clotting factors, electrolyte and renal function, and adequacy of resuscitation must all be accounted for based on the clinical situation.

Careful documentation of every aspect of the evaluation and treatment must be made, and generous use of medical photography is helpful not only for documentation, but also for preoperative planning purposes.

Preoperative Planning

Once the evaluation is complete, the surgeon can formulate a treatment plan and determine if surgical intervention is required. The most important question to consider is whether the surgeon and the facility possess the requisite experience, capability, equipment, and consultative and ancillary services to render optimum treatment (Fig. 2-1). The surgeon must have the technical ability and anatomic knowledge for the given location of the injury. The treating hospital and operating room must also be able to furnish the necessary instruments, implants, and wound treatment dressings.

When confronted with cases of multi-system trauma, a multidisciplinary team involving traumatologists, intensivists, medical consultants, infectious disease specialists, and wound care support personnel is generally considered to be mandatory. If unable to meet these basic requirements, the clinical situation should be carefully reviewed and the strongest consideration should be given for referral to a higher level of care.

FIGURE 2-1 Injury to multiple extremities and organ systems changes surgical options, increases the risk for systemic problems, and has greater psychological impact. Options for flaps or skin grafts are far more limited. Proper room setup and a hospital team that is prepared to manage comorbidities will facilitate delivery of appropriate care and provide the greatest likelihood of success.

Surgery

Patient positioning varies depending on the site of the wound in question. General or regional anesthesia may be employed, and consideration for peripheral nerve blocks should be considered for perioperative pain control. A sterile or non-sterile tourniquet is indispensable to ensure optimal visualization for meticulous, thorough debridement and protection of vital structures. Appropriate broad-spectrum antibiotics should be

given as scheduled throughout the surgical procedure.

Debridement

Successful surgical treatment of wounds begins with the meticulous and complete removal of foreign material, infection, and devitalized tissue to create a healthy wound bed (Fig. 2-2). In chronic wounds, we attempt to create an acute wound to promote healing. In acute injury, wounds must be extended past the zone of injury to ensure complete treatment, and failure to do so significantly limits the effectiveness of debridement. Judicious use of lavage may help remove foreign matter, but care must be taken not to extend the zone of contamination by forcing debris into the surrounding tissue. Use of a tourniquet early in the case is important to best visualize all contaminants and devitalized tissue and avoid injury to vital structures such as nerves and blood vessels. The tourniquet should be released prior to closure or dressing application to confirm removal of all devascularized tissue and excellent hemostasis.

A systematic approach to wound debridement is required and sharp debridement is the cornerstone of this surgical technique. We prefer the centripetal approach working from superficial tissues to deep, from the margins to the center of the wound. Starting at the skin edges, we meticulously work towards the deeper structures within the wound (Fig. 2-3). In general, we prefer excision of all devitalized tissue to a healthy tissue margin instead of a â€œwait and seeâ€• approach to suspect tissue as we feel this limits persistent contamination and infection. All non-viable or suspect tissue is sharply debrided from the wound until a healthy margin of viable tissue achieved. Every effort to preserve nerves and blood vessels crossing the zone of injury is made, and if they are transected these structures are carefully tagged with dyed monofilament suture and documented in the operative records so that they may be more easily visualized during later wound debridements or reconstructive efforts.

FIGURE 2-2 Palmar (A) and dorsal (B) views of an acute hand injury produced by an improvised explosive device. Jagged skin margins, necrotic tissue, foreign material, and hematoma are visible (photos courtesy of Dr. Dana Covey, CAPT, MC, USN).

FIGURE 2-3 This patient, injured by an improvised explosive device, sustained left upper (A,B) and bilateral lower extremity injuries (C,D). Excellent serial wound debridement results in smooth margins, healthy skin and muscle ready for skeletal stabilization, and delayed primary wound closure for the lower extremities.

FIGURE 2-4 Plantar wound is not ready for closure or coverage; jagged skin and eschar at the wound edge needs to be sharply debrided. Serial debridements allow the zone of injury to fully declare itself.

Identification of non-viable tissue remains a challenge, and there is no substitute for experience. In general, non-bleeding skin that appears dusky or does not blanch should be excised, creating a smooth wound margin and avoiding the creation of a ragged skin edge that is difficult for subsequent skin grafting or closure. Subcutaneous fat should be soft and yellow; hard, dusky, or gray fat should be excised (Fig. 2-4). It is also important to keep this tissue moist during deeper debridement, as it will easily desiccate. Injured blood vessels or nerves must be carefully assessed for primary or delayed repair or grafting. Smaller sensory nerve branches may not be amenable to salvage, and if so, we like to pull traction on the proximal end, cut sharply, and allow retraction into the soft tissues. If the stump cannot be retracted, we make every effort to bury it in muscle. To the greatest extent possible local soft tissues should be used to cover exposed tendons, nerves, and vessels to prevent desiccation and further

injury.

In debridement of muscle, fascia, and tendon, there are several important points to keep in mind. Muscle fascia should be stout, white, and shiny, while non-viable fascia will often appear grey or black, fragile, and stringy. When excising fascia, however, great care should be taken as neurovascular bundles may be in close proximity. Muscle should be red, shiny, of good consistency, contractile, and bleeding (Fig. 2-5). Anything to the contrary should be considered for debridement. Knowledge of anatomy and local blood supply is paramount in this endeavor as overly aggressive

debridement within muscle compartments may devascularize previously viable tissue. Tendon debridement must be carefully considered due to potential loss of function. Tendons are also easily desiccated, especially if overlying paratenon or sheath is missing.

FIGURE 2-5 A: Right hand and (B) left hip open wounds ready for further coverage. Note the healthy, smooth skin margins, clean granulating tissue bed, minimal edema, and skeletal stabilization.

Devascularized bone fragments must be removed from the wound bed, with the exception of substantial articular fragments, which should be retained in an attempt to preserve the articular surface (Fig. 2-6). Curettes, rongeurs, and burs are useful to check for punctated bleeding indicative of healthy bone that should be preserved.

Cultures of any contaminated or osteolytic bone will help guide antibiotic selection.

There are many tools available for mechanical debridement. One tool we find useful for very large wounds is the VersaJetâ„¢ Hydrosurgery System (Smith and Nephew). This device uses negative pressure from a high-speed stream of water across a small aperture at the tip to remove softer tissue and surface debris by suction. Additional lavage is not required when using this device. However, great care must be used around neurovascular structures to avoid injury.

Strict hemostasis is critical to prevent hematoma and limit further infection and morbidity due to blood loss. Suture ligatures and surgical clips should be used for larger vessels, and bovie or bipolar cautery for smaller vessels. We avoid use of braided suture when possible to avoid harboring bacteria.

Judicious use of a tourniquet is helpful to identify and control large bleeding vessels and includes release to assess hemostasis prior to closure, grafting, or dressing application. Adjunctive topical hemostatic agents are available and have been used successfully in some of our most severely war injured patients. Lavage is important for removal of foreign debris and lowering bacterial counts. Reflecting prior published research, we use plain saline for lavage without antibiotics. Based on our experience and the literature, we limit use of pulsatile lavage in the forearm and lower leg, and almost never use it in the hand or foot. Pulsatile lavage can further damage delicate tissues, exacerbating the potential for adhesions and functional loss. When pulsatile lavage is used in the hand and foot, it is usually at the half pressure setting. Usually, bulb irrigation is sufficient when combined with careful debridement.

FIGURE 2-6 During serial debridements every effort is made to preserve substantial articular cartilage fragments. This case also illustrates the value of a

team approach, after extensive surgical time to obtain the best possible ORIF of this comminuted intraarticular knee injury (A,B), the plastic surgery service performed a medial gastrocnemius pedicled rotation flap and split thickness skin grafting (C,D).

Temporary Coverage and Void Fillers

Negative pressure dressings are a great advance in the treatment of wounds not amenable to primary closure (Fig. 2-7). The wound VACÂ® (Vacuum Assisted Closure) dressing is commonly employed at our institution. It continues to debride wounds while reducing edema and local bacterial counts and it promotes growth of healthy granulation tissue. It also eliminates the need for multiple daily dressing changes, thereby reducing the patient's discomfort and nursing staff workload. We use portable suction devices supplied by the manufacturer set at the recommended 125 mm Hg intermittent suction. Wall suction can be used if a portable unit is not available, but care must be made to ensure the unit is calibrated to provide a true measurement of suction. We prefer to use less suction in the upper extremity due to the more delicate nature of the soft tissues and have found this to be effective. In the hand, use of the wound VACÂ® is limited to avoid desiccation and injury to vital nerves, blood vessels, and tendons. It can be of benefit when the underlying tissue is robust such as the thenar or hypothenar eminence, or when the more delicate tissues are absent due to the injury and debridements (Fig. 2-8). When neurovascular structures are exposed we employ standard wet to dry dressings. In general, we try to limit exposure of blood vessels, nerves, or tendons to the wound VACÂ® and try to rotate available local tissue to provide coverage prior to placement of the wound VACÂ®. Multiple wounds can be treated with the same suction tube using a foam bridging technique. However, care must be taken to cover the skin with the supplied biofilm or ioban to prevent local skin breakdown.

Adequately eliminating contamination and infection is essential to successful wound treatment. In addition to appropriate broad-spectrum antibiotic use, there are many different options available to provide local infection control that can be tailored to the clinical or surgical situation. Antibiotic bead pouches or fracture spacers have

been used effectively to provide local infection control in cases of wounds with associated high-energy fracture patterns. With comminution and bone loss soft tissue space can be maintained for future reconstruction and enhanced mechanical stability provided (Fig. 2-9). In highly resistant bacterial infection, silver impregnated films, colloidal materials, and wound VACÂ® sponges are additional options for the surgeon and have been utilized with great frequency at our institution. We presently reserve the use of silver for refractory cases due to concerns of resistance and the current high cost of this modality. For extremely large wounds with highly resistant bacterial colonization or infection that are not amenable to wound VACÂ® treatment, Sulfamylon (mafenide acetate) or Dakin's soaked wet-to-dry dressings have proven effective and resulted in successful wound closure. Infectious disease specialty assistance is recommended in such cases.

FIGURE 2-7 Use of serial debridements and negative pressure dressings prepare the wound for subsequent flap coverage.

A

B

C

D

E

F

G

H

FIGURE 2-8 This case, of a patient who sustained an isolated injury to the thumb from an AK-47 rifle round, illustrates several principles of wound management: thorough debridement preserving vital structures (A) and provisional skeletal fixation with K-wires (B); C: the use of negative pressure dressings to prepare the wound for coverage. D-H; attainment of a closed healthy soft tissue envelope prior to reconstruction, in this case a local rotation flap and full-thickness skin graft were selected.

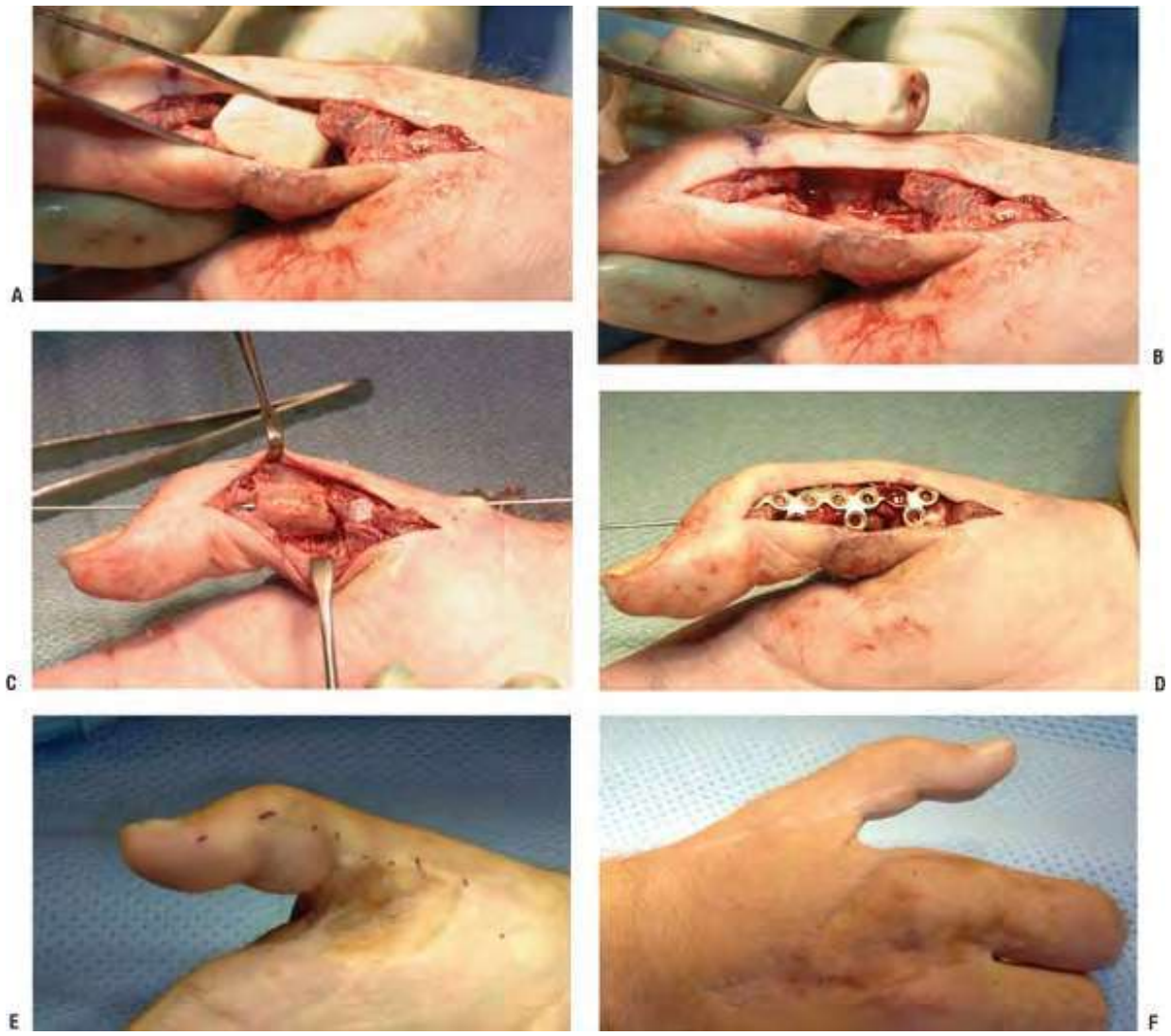

FIGURE 2-9 A,B: In cases of contaminated segmental bone loss (case from Fig. 2-8), an antibiotic impregnated cement spacer can be used to decrease infection risk and maintain space and alignment until the wound is clean enough for soft tissue reconstruction with bone grafting (C,D). E,F: A well healed soft tissue envelope, demonstrated here, is necessary prior to performing nerve and tendon grafting.

Choice of Fixation

When wounds are associated with fractures in the acute setting, provisional stabilization should be attempted to maintain soft tissue space, prevent mechanical agitation of the surrounding tissues, and optimize pain control (Fig. 2-10). In general, external fixators are preferred acutely with conversion to definitive fixation as indicated by the injury. In the setting of blast injuries, large amounts of debris are forced into the wounds with tremendous energy and the level of contamination is typically higher than that seen in most blunt open trauma. Our experience has shown the significant potential for widespread osteomyelitis when intramedullary fixation is selected for blast injured patients and we now prefer definitive treatment with an external fixation device in many cases. For the distal extremity, as well as periarticular fractures, Kirschner wires are indispensable for temporary and sometimes definitive fixation. Definitive fixation that requires significant soft tissue stripping may only compound the injury and should be entertained with extreme caution as the clinical situation dictates.

For highly contaminated wounds, or when there is concern for viability in critical areas or structures, repeat operative debridement should be planned every 24 to 36 hours until a healthy, vascularized soft tissue bed is achieved.

Timing of Final Closure

When is the wound ready to close? There are several elements that must be present to ensure success of wound closure. Certainly, all non-viable or necrotic tissue must be absent from the wound. It cannot be overemphasized that good vascular flow must be

present for healing of tissues. In addition, a tension free closure must be achieved or blood flow will be compromised at the wound margins leading to wound breakdown and dehiscence. Acute limb shortening can be considered when significant bone and soft tissue defects are present (Fig. 2-11). Soft tissue expanders can also be a valuable tool in the reconstructive phase (Fig. 2-12) One significant marker of excessive tension is blanching at the wound margins. Suture type, configuration, and binding tightness of the suture must all be considered. Mattress sutures can be tension relieving and provide nice eversion of wound edges, but one must also recognize that inappropriate orientation or overtightening of these sutures, especially in horizontal mattress fashion, can create ischemia at the wound edges. In wounds prone to tension in the closure, we typically choose nylon suture sized appropriately to the wound in vertical mattress fashion to distribute tension evenly. At times, we will also utilize â€œtrauma retentionâ€• type sutures using heavier gauge nylon to improve local tension. When excessive tension remains, it is better to leave the wound open and return to the operating room later when tension free closure can be obtained. In these cases, we prefer a â€œJacob's ladderâ€• type closure with staples and vessel loops, usually over or under a wound VACÂ® sponge or gauze dressings to create gentle traction at the wound edges and promote future closure attempts. While commercial skin traction devices are available (Fig. 2-13), we have not found them to be more effective, and they are certainly more costly.

FIGURE 2-10 A,B: Fracture stabilization is required to assist in soft tissue management and to maintain structural relationships for subsequent reconstruction. K-wires and external fixation are especially useful in traumatic, highly contaminated wounds.

A

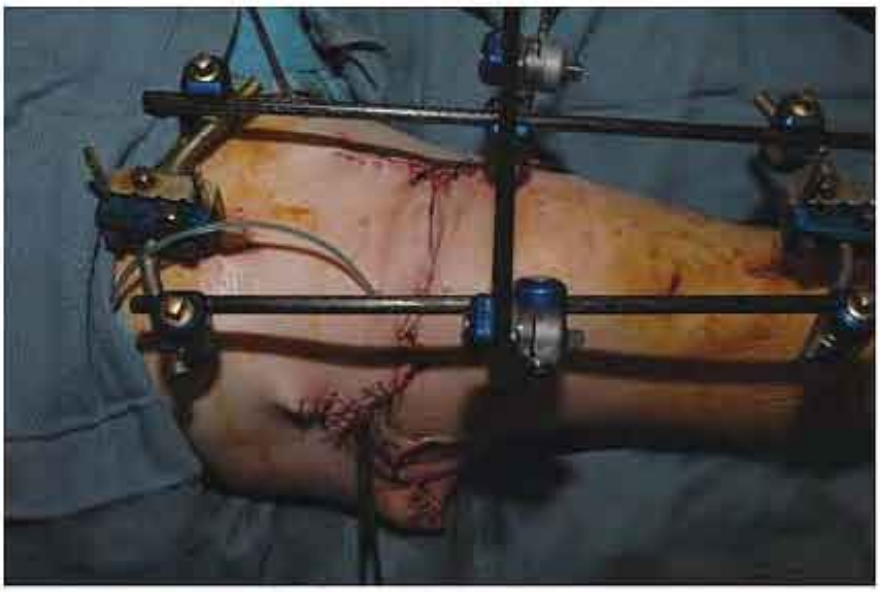

B

C

FIGURE 2-11 A-C: Large wound with a segmental femoral defect and associated loss of the anterior thigh soft tissue was managed with acute femoral shortening. Following soft tissue stabilization the femur was subsequently lengthened.

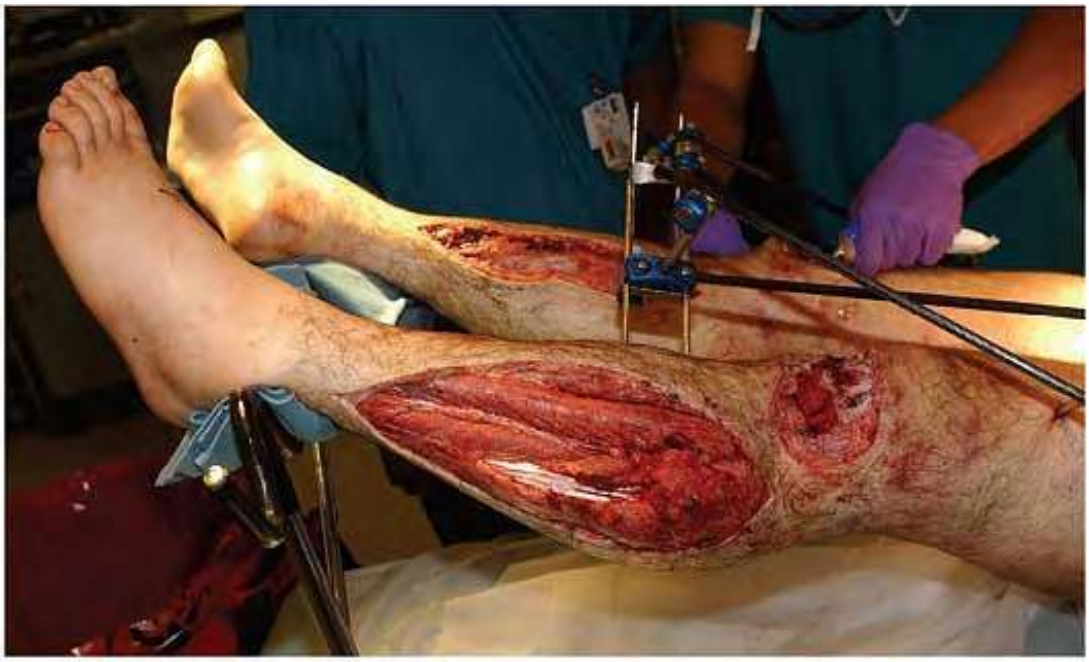

A

B

FIGURE 2-12 A,B: This patient required extensor mechanism reconstruction after an open knee injury. Soft tissue expanders were used after the initial fracture care, flap, and skin grafting, in order to create adequate skin for reconstruction and knee motion. It is important to remember that soft tissue scarring as well as

split thickness skin grafting limit skin excursion and pliability. Subsequent surgery may require release of scar or replacement of damaged skin to obtain motion in affected joints.

Absence of infection is also critical to wound closure. We routinely culture infected or suspicious wounds to help guide antibiotic therapy. Previous literature suggests a culture threshold of less than 10^5 bacteria per gram of tissue is required to allow successful wound closure. However, the utility of cultures is highly dependent on laboratory technician experience, and they have not proven to be clinically helpful at our institution. Clinical and laboratory indicators of infection such as fever, elevated white blood cell count, and elevated inflammatory markers such as C- reactive protein and erythrocyte sedimentation rate can help guide decision making. Examination of the wound may provide obvious clues like purulent material, unhealthy appearing sheen to the tissues, and foul smell. However, even in the most experienced hands, this determination can be exceedingly difficult; surgeons must rely on the clinical appearance and laboratory markers (Fig. 2-14). Current research into

other markers or mediators of inflammation and infection is ongoing, but until these efforts bear fruit, this assessment continues to be a significant challenge to the surgeon.

FIGURE 2-13 In some cases tension relieving devices can be used alone or in conjunction with negative pressure dressings to facilitate closure, as in this patient (A) with a large wound involving the anterior aspect of the lower extremity. B,C: Serial debridements and use of a Jacob's ladder with a negative pressure dressing allowed eventual primary wound closure. D: Tension relieving slits in the skin of the lower leg were also used in this case.

Postoperative Management

Postoperative wound care focuses on protecting the healing wound and optimizing medical care to ensure success. Initially, after wound closure or flap coverage, complete soft tissue rest of the involved extremity by splinting or external fixation should be considered (Figs. 2-15 and 2-16). Care should be taken when splinting to

ensure functional positions are maintained; the intrinsic plus position for the hand, full extension at the knee, and a plantigrade position for the ankle can facilitate return to function and reduce the need for secondary procedures. Provisional fracture fixation constructs must be checked to ensure integrity of the construct with modifications made as needed to control motion at the fracture site, and ongoing tissue injury.

Elevation of the extremity is critical to reduce local swelling and edema formation. When an external fixation device is present it can be tied to balanced suspension supported by an overhead trapeze. We avoid the use of slings on the lower extremity due to the potential for pressure necrosis.

P.26
P.27
P.28
P.29

If an external fixator is not present and a sling is used under the calf, it should always be tied to balanced suspension, not the overhead frame, in this manner. As the patient moves, the pressure on the calf remains constant due to the hanging weights. The skin must be assessed periodically throughout each day (Fig. 2-17).

FIGURE 2-14 Aâ€“C: In contrast to Figure 2-4, this foot wound was ready for a lateral thigh free flap as evidenced by the healthy wound bed and margins, minimal residual limb edema, and excellent flap healing.

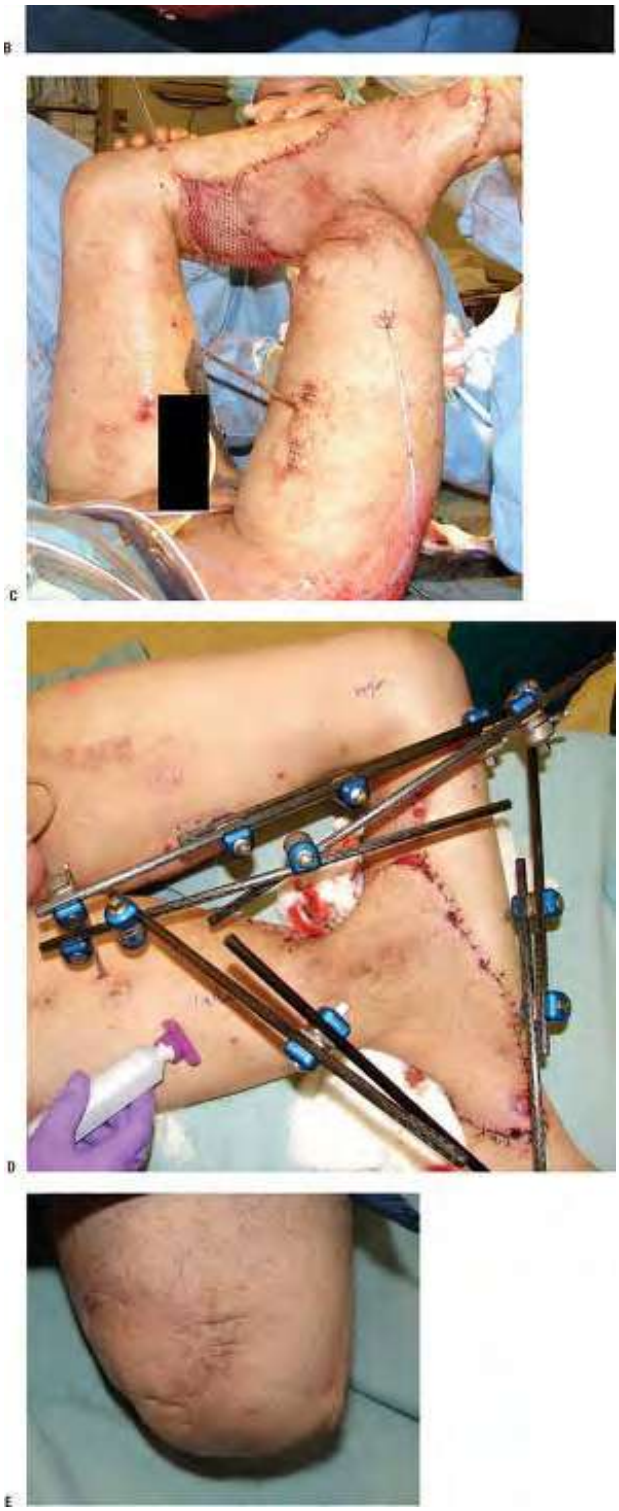

FIGURE 2-15 Patient presented with bilateral lower extremity wounds. His right below knee amputation lacked adequate soft tissue for closure and prosthetic

wear (A); his left lower extremity had a large soft tissue wound (B). The tissue that would have been discarded in a revision amputation to an above knee level was used in a â€œcross-legâ€• flap (C,D) with an excellent clinical result (E). Note the external fixator used to protect the flap after inset (D).

FIGURE 2-16 A: After continued treatment with VAC dressings and ORIF, the left upper extremity wounds from Figure 2-3A,B are ready for definitive coverage. In this case flap selection is limited by the injuries. B: Bilateral groin flaps were used to obtain coverage of both the hand and forearm wounds. To minimize tension on the pedicles, external fixators were used to stabilize the forearm to

the pelvis. C: Healed wounds with durable, healthy coverage.

FIGURE 2-17 Elevation is a valuable tool for managing edema. Care must be taken when applying elevation, particularly in the ICU and poly-trauma setting, where patients are often sedated and less mobile. This patient, with multiple bilateral lower extremity fractures eventually received a below knee amputation after development of a full thickness posterior pressure ulcer. The leg was initially elevated by a sheepskin sling tied directly to the overhead bed frame resulting in significant pressure over the posterior calf. Use of a balanced suspension device secured to an external ankle fixator, as demonstrated in the photo, may have prevented this complication.

Mobilization and edema control of the digits should begin as early as possible. Once or

twice daily motion begun within seven days of injury can significantly reduce edema and long term stiffness. Burn injuries can be particularly challenging due to pain, contracture, and poor durability of the grafted skin (Fig. 2-18). While consultation with physical and occupational therapists aids in maximizing return of function, a balance between early motion and soft tissue rest must be achieved to ensure proper wound healing. Effective communication with these vital specialists will ultimately benefit both the surgeon and patient.

Patient physiologic factors and comorbidities must continually be reevaluated and optimized in the postoperative period (Figs. 2-19 and 2-20). Markers for infection, hemodynamic status, systemic function, and nutritional status should be reassessed in accordance with the clinical picture. Wound cultures are checked and antibiotic choices are reevaluated in consultation with infectious disease specialists. When using the wound VACâ„¢ system, special attention must be made to the pressure settings and quality of suction. Nursing and ancillary staff education is invaluable. Prominent bone or hardware, especially when combined with inadequate or marginal soft tissue coverage, can lead to wound breakdown and the need for additional surgery (Figs. 2-21 and 2-22).

Pain management is not only humane, but we believe greatly contributes to the success of treatment, especially in cases requiring frequent bedside dressing changes. Reuben et al have shown a reduced incidence of CRPS with aggressive pain management. In many cases, especially trauma, our anesthesia colleagues on the pain service oversee medical management of pain, and provide regional

P.30
P.31
P.32
P.33
P.34
P.35

nerve blocks when appropriate. The mental health of the patient must also never be forgotten, especially in cases of traumatic injury, and appropriate consultation with psychiatric specialists and/or a chaplain should be considered to assist in the patient's overall well-being and ability to participate fully in their own recovery and rehabilitation.

FIGURE 2-18 This patient received severe burns to the upper extremity and chest requiring groin flap to the thumb, and multiple skin grafting procedures complicated by recurrent infections with MRSA. Upon presentation to us he had severe contracture of the long finger (A), erosion of the skin over the residual index metacarpal (B), and essentially no use of the thumb or long finger. Staged reconstruction included thumb metacarpal lengthening, completion of the index ray amputation (C), full thickness skin grafting to the long finger (D), first web-space deepening, thumb carpometacarpal joint arthroplasty, and tendon transfer

to the thumb metacarpal for abduction and opposition. The result: a supple, durable soft tissue envelope and functional hand (E,F). Attention to detail through all phases of care can reduce the need for additional procedures.

FIGURE 2-19 Underlying medical conditions where immunosuppression is present can lead to wound healing problems even after elective surgery. This rheumatoid arthritis patient underwent routine open carpal tunnel release with subsequent persistent wound drainage. Ultimately, she was diagnosed with Mycobacterium avium. The initial debridement and culture (A) failed to heal fully (B). The gelatinous appearing tenosynovium and indolent course are typical of mycobacterial infections. Successful wound healing followed appropriate antibiotic therapy and delayed primary closure.

FIGURE 2-20 A,B: Patient had compromised sensation and vascularity following multiple surgeries to the small finger for Dupuytren's disease. The patient sustained a burn to the distal phalanx on the moist heat pads used before therapy. The initial blister and mild erythema were followed by progressive distal tip necrosis; ultimately, the patient requested an amputation of the digit.

FIGURE 2-21 In this series, the patient's initial wounds were closed primarily but there was inadequate coverage of the PIPJs dorsally (A). A delayed adipofacial flap (B) and overlying full thickness skin grafts taken from his amputated lower extremity (C) were not robust enough coverage over the dorsally applied plates (D). Removal of the dorsal plates and primary closure lead to successful wound healing (E).

A

B

C

FIGURE 2-22 Patient presented with an innocuous appearing wound and persistent serous drainage 3 weeks after ORIF of an open olecranon fracture overseas (A). Careful inspection showed that the wound communicated with the prominent underlying hardware (B). Irrigation, debridement and IV antibiotics combined with removal of the prominent wire knots (C) and primary closure resulted in uneventful healing of the wound and fracture.

A

B

C

D

FIGURE 2-23 A-D: Patient suffered a blast injury resulting in an open fracture of the elbow, in addition to severe soft tissue injury and a forearm fasciotomy. He was initially managed with a spanning external fixation device and serial debridements followed by ORIF. (ORIF; pedicled Latissimus Dorsi flap and split thickness skin grafting.)

Conclusion

Treatment of wounds can be a challenging enterprise, consuming considerable time, energy, and resources. By combining the basic principles of evaluation and treatment set forth in this chapter, and with meticulous attention to detail, the surgeon will have an excellent starting point for the treatment of complicated soft tissue trauma (Fig. 2-23).

Disclaimer. The views expressed in this chapter are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government.

Recommended Reading

Anglen JO. Wound irrigation in musculoskeletal injury. *J Am Acad Orthop Surg.* 2001;9:219â€“226.

Argenta LC, Morykwas MJ. Vacuum-assisted closure: a new method for wound control treatment: clinical experience. *Ann Plast Surg.* 1997;38(6):563â€“576.

Attinger CE, Janis JE, Steinberg J, et al. Clinical approach to wounds: debridement and wound bed preparation including the use of dressings and wound-healing adjuvants. *Plast Reconstr Surg.* 2006;117(Suppl.):72â€“109S.

CDC. Preventing tetanus, diphtheria, and pertussis among adults: use of tetanus toxoid, reduced diphtheria toxoid and acellular pertussis vaccines. Recommendations of the Advisory Committee on Immunization Practices (ACIP)

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 3

Management of Simple Wounds: Local Flaps, Z-Plasty, and Skin Grafts

John B. Hijjawi

Allen T. Bishop

Skin grafting and local flap coverage have remained a common means of covering traumatic wounds. Skin grafts, local flaps, and random flaps are not as simple as direct wound closure nor are they are as â€œelegantâ€• as procedures that are found higher on the reconstructive ladder such as pedicled flaps or free tissue transfer (Fig. 3-1). However, the procedures described in this chapter remain straightforward, reliable, and time-tested and should be an essential component of any surgeon's reconstructive armamentarium.

Skin Grafts

Skin grafts are classified by source and thickness. By far the most common, durable, and successful skin grafts are autografts harvested from the patient's own skin. There are no immunologic issues, since the tissue comes from the patient's own body. Concerns over disease transmission are eliminated and expense is minimal. The only disadvantage is the creation and care of a donor site.

Skin grafts are also available as cadaveric allografts, xenografts (typically porcine skin), and most recently, cultured epithelial grafts. These materials are lifesaving

sources of temporary wound coverage in the context of massive burns. Allografts and xenografts do have the disadvantage of immunogenicity and thus impermanence since they are bound to be rejected. However, if the quality of a wound bed is questionable, preserved porcine xenografts may be an excellent option for temporary wound coverage.

Currently available cultured epithelial grafts are expensive, require time to culture, and are not as durable as autografts. Additionally, studies have shown that they are significantly more susceptible to infection than standard autografts.

Skin grafts are also classified based on thickness. The skin is composed of an outer epidermis and a deeper dermis, which is further subdivided into the reticular and papillary dermis (Fig. 3-2). All skin grafts consist of the entire epidermis and a variable amount of dermis. Split-thickness skin grafts contain only a portion of the dermis, whereas full-thickness skin grafts contain the entire dermis and dermal appendages. As a result, full-thickness skin grafts continue to support hair growth following transfer. This needs to be carefully considered when selecting donor sites in situations when a full-thickness skin graft is to be transferred to a conspicuous, previously hairless area.

Since full-thickness skin grafts include all dermal appendages, skin at the donor site will not spontaneously regenerate following harvest; therefore, these donor sites need to be closed primarily or with a split-thickness skin graft. Split-thickness donor sites retain the ability to generate epithelium and will be largely healed within several weeks if cared for properly.

Split-thickness and full-thickness skin grafts have quite different contractile characteristics on harvesting (primary contraction) and after they heal (secondary contraction). Primary contraction is

the initial contraction of a graft when it is harvested. Due to the greater proportion of elastic fibers in a full-thickness skin graft (the dermis is the location of all elastic fibers in the skin), it will undergo more primary contraction than a split-thickness skin graft when harvested. Similarly, a thicker split-thickness skin graft (e.g., 0.018 in) will undergo more primary contraction than a thin split-thickness skin graft. Most skin grafts can be stretched under minimal tension at inset to overcome this contraction, restoring their original size.

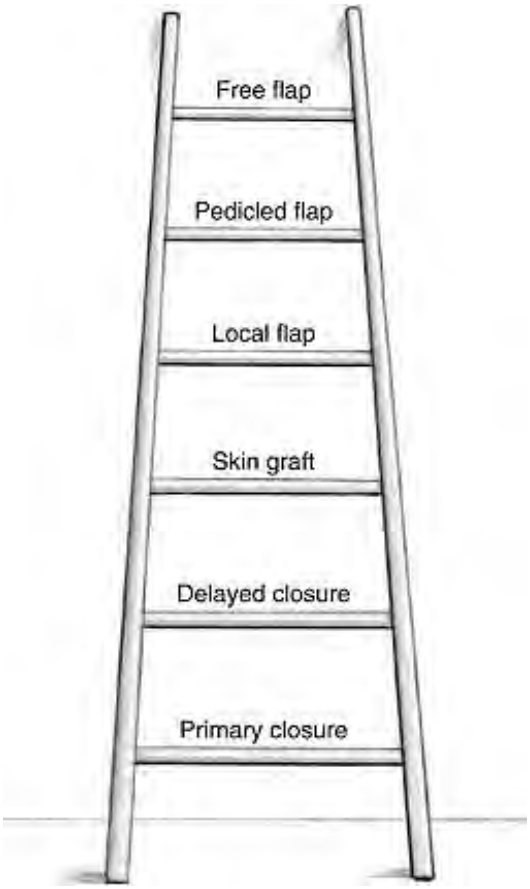

FIGURE 3-1 Reconstructive ladder. Historically, surgeons have closed wounds with the simple procedure first, moving up the rungs of the reconstructive ladder as wounds become larger and more complex.

Conversely, split-thickness skin grafts undergo more secondary contraction than full-thickness grafts. This can be exploited to provide gradual contraction of a wound over the course of several months. An example is a fasciotomy wound that is under too much tension to close primarily within the days following compartment release. A very thin split-thickness skin graft applied to the wound will undergo significantly more secondary contraction than would a full-thickness skin graft, resulting in contraction of the wound itself. After several months, this may result in a wound that is small enough to allow serial excision of the skin graft and primary closure under minimal tension. In comparison,

full-thickness skin grafts can be relied on to undergo virtually no secondary contraction in situations where this is not desirable, such as across a joint surface or in a web space.

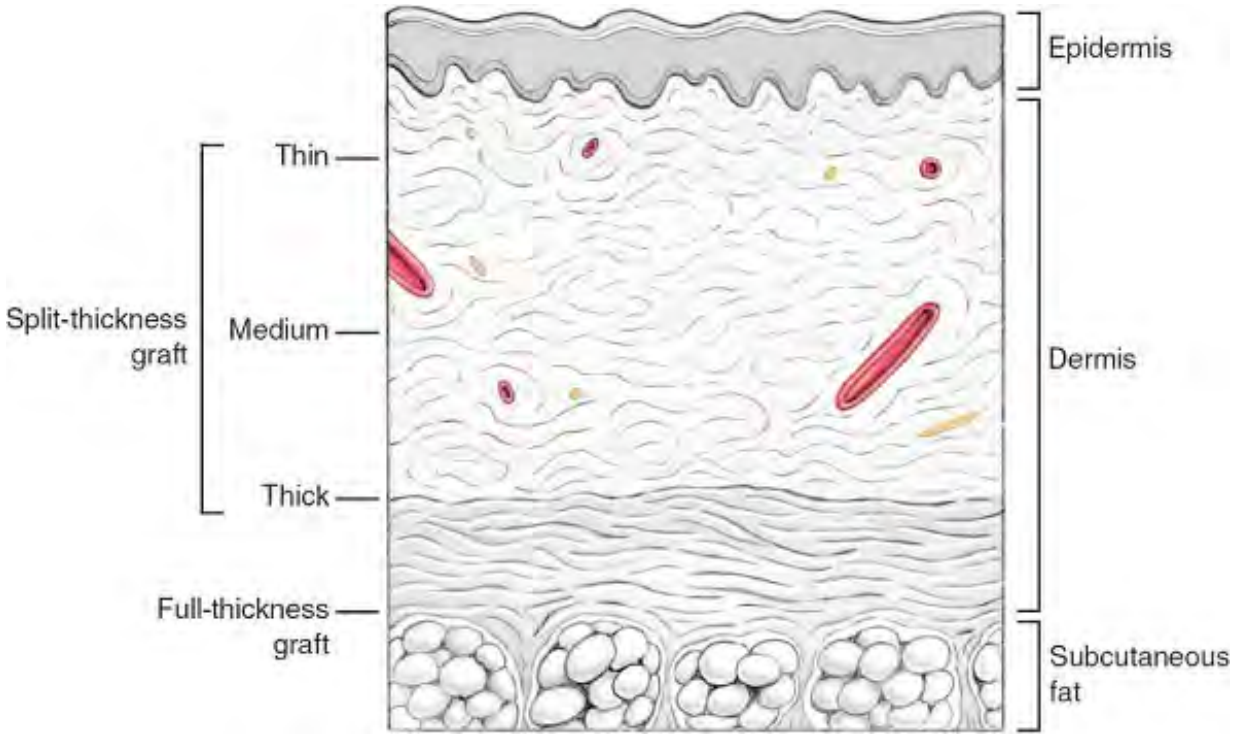

FIGURE 3-2 Skin anatomy.

Initial adhesion of a skin graft is the result of fibrin present between the graft and the recipient bed. Initial survival of a skin graft relies on plasmatic imbibition, which is the process of nutrient diffusion from the recipient site into the skin graft. Later, vascular channels within the graft line up with vascular channels in the recipient bed through the process of inosculation. Finally, long-term graft survival relies on neovascularization, or the process of new blood vessel growth into the skin graft from the recipient bed. Because skin grafts, both split thickness and full thickness, are completely reliant on the recipient bed for survival, the wound bed must be well-vascularized and free of infection to support skin graft take. Split-thickness skin grafts

have lower metabolic demands than do full-thickness grafts, and so they donâ€™t require recipient beds with as rich a blood supply. Along the same lines, full-thickness grafts take longer, from 7 to 10 days, to heal. Split-thickness grafts are generally considered healed by 5 days and should be left immobilized and dressed at least that long postoperatively. The time of healing in specific situations depends most on the quality of the recipient wound's blood supply.

Indications/Contraindications

Indications

Skin grafting may be indicated for any defect that cannot be closed primarily and that has a wound bed that can support skin graft take (Table 3-1). Skin grafts survive for the first several days through a process called imbibition. During this stage of skin graft healing, the graft obtains nutrients from the underlying wound bed through a process of diffusion. Wound beds devoid of blood flow, or with little vascularized tissue, will make for poor recipient sites. Exposed structures that will accept a graft include subcutaneous tissue, paratenon, and muscle. Other tissues, such as exposed bone, joint, tendon, and nerve, may be covered temporarily by graft used as a biologic dressing but will not support a graft for permanent coverage. Beds containing tissues of questionable viability, chronic granulation tissue, or frank infection can accept a skin graft but require thorough debridement prior to graft placement. Wounds that contain fewer than 10^5 bacteria per gram of tissue or that allow xenograft adherence within 24 hours allow successful skin grafting.

Contraindications

Skin grafts are contraindicated in areas that are exposed to repetitive trauma or that lie over osseous prominences. Bone, cartilage, and tendons denuded of periosteum, perichondrium, or paratenon cannot be covered with skin grafts as there is inadequate vascular supply to support healing of the skin graft. Controversy exists over whether skin grafts should be placed over bone, cartilage, or tendons with healthy periosteum, perichondrium, or paratenon. It is certainly possible to get skin-graft healing over such structures. However, skin grafts in these situations are rarely optimal for long-term durable coverage. In addition, skin grafting should be avoided in areas that may require secondary surgery for bone or nerve grafting, as adherence to underlying

muscle, nerve, and tendon may complicate secondary surgery. All split thickness grafts will undergo some component of contracture over time; thus, if these grafts are placed over large areas of the antecubital fossa, popliteal fossa, or olecranon, there is a risk of limitation in joint motion.

Preoperative Planning

The wound must be debrided and clean prior to attempts at skin grafting. Infection is one of the leading causes of skin graft failure. Since skin grafts are completely dependent on the wound bed they are transplanted to for nutrition, they possess no intrinsic ability to resolve infection. Quantitative wound cultures have been used for many years in some centers to determine the adequacy of a wound's microenvironment for closure. A quantitative culture revealing less than 10^5bacteria per gram of tissue has been traditionally regarded as an acceptable level of colonization below which a

wound can be closed by skin graft or local flap. However, such cultures are highly dependent on the experience of the technician performing them. A careful clinical evaluation and serial sharp debridement of clinically infected or contaminated wounds are advised.

Table 3-1. Wound Analysis

Is there too much tension across the wound for primary closure?
Is there adequate perfusion of the wound bed?
Are vital structures exposed at the base of the wound?
Is the wound infected or contaminated?
Is the wound geometry favorable for closure?

Donor site selection must also be decided before surgery. Considerations when

choosing a donor site for split thickness skin grafts include cosmesis, the thickness of skin in the donor site region, and ease of care of the donor site after graft harvest. The most common split thickness skin donor sites include the buttocks, lateral and anterior thighs, and lower abdomen. These sites are relatively easy to conceal, have thick skin resulting in less pigmentation once healed, and are readily accessible in even a bed-bound patient, easing postoperative care.

Full thickness skin grafts are most commonly harvested from the groin, antecubital fossa, volar wrist crease, medial arm, postauricular sulcus, or lower abdomen. These donor sites all exist in areas where closure can be performed within pre-existing skin creases, thus resulting in relatively inconspicuous donor sites. The hypothenar skin or plantar instep offers the unique quality of glabrous skin if needed for graft material. Always keep in mind that any amputated â€œspare partsâ€• can provide a good source of viable skin graft with no added morbidity to the patient.

Surgery

Patient Positioning

The patient's position will depend on the location of the graft to be harvested. Most frequently we harvest split thickness grafts from the upper thigh area, where they may easily be concealed under clothing A supine position with a roll underneath one hip is ideal. The majority of full-thickness grafts are harvested from the hairless skin of the groin crease, though the inner upper arm may be used as well. The patient is positioned supine for such harvests. For glaborous skin grafts, the instep of the foot may be positioned so as to allow ease of harvest. For wounds on the posterior aspect of the lower extremity or trunk, a lateral decubitus position readily exposes both the wound and a lateral thigh donor site. With careful planning it is almost never necessary to reposition a patient after harvesting the skin graft.

Split Thickness Grafts

Power dermatomes are the most common method of harvesting split thickness skin grafts, although for very small split thickness grafts hand-driven Weck blades may be more convenient.

Measure the recipient wound and choose a dermatome guard based on that

measurement (Fig. 3-3).

- Set and check the dermatome thickness (usually 0.010 to 0.015 in) with a No. 15 scalpel blade. The thin, beveled edge of the knife is about 0.010 in, whereas the thickest portion of the blade is 0.015 in thick (Fig. 3-4).
- Mark the donor site with a ruler so that you will know where the dermatome needs to â€œtouch downâ€• and â€œlift off.â€• Relying on your memory or estimating how far you will have to drive the dermatome can lead to harvesting too little graft and having to reharvest, creating an unnecessary seam in the graft. In cases of overharvesting, the extra graft can be replaced onto the donor site, but this complicates the donor site dressing.

FIGURE 3-3 A dermatome guard of appropriate width is chosen.

FIGURE 3-4 The thickness setting is double checked.

FIGURE 3-5 Countertraction is applied to avoid â€œskippingâ€• along the skin.

- Clean the donor site to remove any sticky material that will cause the dermatome to stick and apply copious amounts of mineral oil to the donor skin

and the dermatome.

- Apply countertraction to the skin in front of and behind the dermatome blade. Dermatomes, particularly when fitted with larger guards, function much more effectively on flat surfaces (Fig. 3-5).
- Activate the dermatome before â€œtouching downâ€• on the skin and plan to keep it activated until after â€œlifting offâ€• of the skin. Touch down at 45 degrees to the skin, then slightly flatten the angle between the dermatome and skin, maintaining constant firm pressure on the head of the dermatome.
- Realize that if you need to reset a hand for countertraction, you can stop the blade without lifting off of the skin and reset countertraction. The harvest can then be continued without interrupting the continuous sheet of skin graft.
- Transfer the harvested graft onto a dermal carrier with the dermis side up. This side is shinier, and has less friction when rubbed than the epidermis side. If you become confused as to which side is the dermis side, realize that the skin graft edges will always roll toward the dermis side.
- Either make several small slits in the skin graft with a scalpel (â€œpie crustingâ€•) to allow for drainage of accumulated fluid from the wound bed, or run the graft and dermal carrier through a skin graft mesher (Fig. 3-6). Most typically, grafts are meshed at a 1:1.5 ratio. This is done to allow drainage through the graft, to make grafts more conformable to the underlying wound bed, and to increase the area a graft can cover. It is not necessary, however, and many surgeons avoid it since the meshed appearance will be permanently obvious and will significantly compromise the final cosmesis.
- Fix the graft dermis side down to the wound bed with either staples or absorbable sutures such as 5-0 chromic. Traditional bolster or â€œtie-overâ€• dressings employ silk suture placed circumferentially around the skin graft and left long. They are then tied over mineral oil soaked cotton wrapped in a non-adherent dressing and placed firmly onto the skin graft (Figs. 3-7,3-8,3-9). A very convenient bolster dressing can be fashioned by placing a non-adherent Nterface dressing (Delasco, Council Bluffs, IA) over the skin graft, followed by mineral oil soaked cotton pushed firmly into the wound bed to compress the skin graft. Finally, a Reston (3M, St. Paul, MN) sponge can be cut to conform to

the wound, placed adhesive side up, and stapled to the skin surrounding the wound. This is a very stable bolster construct that resists shearing forces and provides firm compression. For skin grafts placed on an extremity, an Ace wrap can be used to protect this entire dressing.

- Appropriate splints should be applied to immobilize the recipient site until the skin graft is totally healed.

Full Thickness Grafts

Full thickness grafts are harvested after drawing an ellipse, which includes the necessary amount of skin based on the wound measurements. Closure without dogears is best achieved when the length of the ellipse is about four times the width of the ellipse. A scalpel is

used to incise the skin just through the dermis. One end of the ellipse is then lifted firmly with a single skin hook or toothed forceps and a fresh scalpel is used to elevate the skin graft just deep to the dermis, taking as little fat as possible with the skin graft. Once harvested, the full-thickness graft can be rolled over a finger or shot glass with the dermis side up, and a curved iris scissors used to remove every bit of residual fat from the dermis.

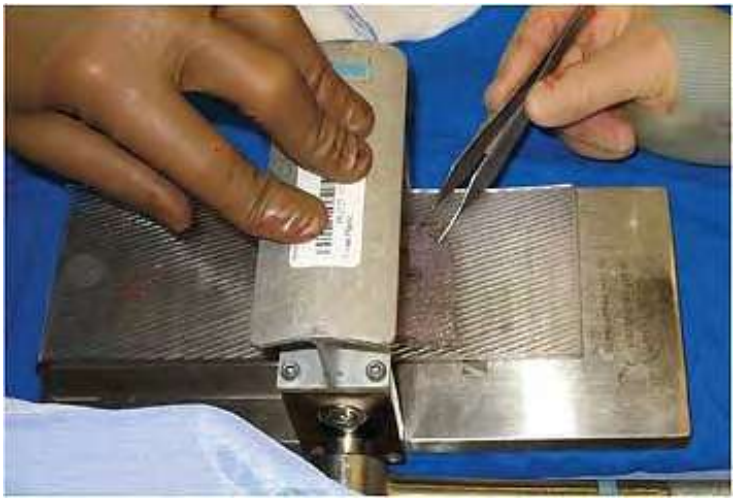

FIGURE 3-6 Meshing the skin graft dermis side up facilitates placement onto the wound.

FIGURE 3-7 The bolster or â€œtie-overâ€• dressing begins with placement of long sutures along the edge of the skin graft.

FIGURE 3-8 Cotton balls are wrapped in a non-adherent dressing and compressed onto the skin graft.

FIGURE 3-9 The completed tie-over.

Full thickness grafts are not typically meshed. Otherwise, they are fixed into position as described for split thickness grafts. They should be left undisturbed for 7 to 10 days. Donor sites are best treated with primary closure after moderate undermining of the wound edge.

Postoperative Management

Split thickness skin grafts should not be disturbed for a minimum of 5 days. The dressing can then be carefully removed to not disturb the healing skin graft, and once daily dressings are begun with an antibacterial ointment and a non-adherent dressing such as Adaptic (Johnson and Johnson, New Brunswick, NJ). Full thickness grafts should remain covered for 7 to 10 days before removing the split or bolster.

FIGURE 3-10 Donor site dressing.

Donor sites up to 100 cm^2 can be covered with Tegaderm (3M, St. Paul, MN) as long as care is taken to carefully dry the skin surrounding the donor site before applying the dressing. Serum will collect under the Tegaderm after several days, and can be left alone. If the Tegaderm leaks, a patch can be placed over the hole. The benefit of Tegaderm is a completely isolated, moist environment that is virtually painless for the patient (Fig. 3-10). The donor site is typically re-epithelialized by 2 weeks at which point the Tegaderm can be removed and replaced with once daily application of a moisturizing cream and non-adherent dressing. Alternatively, the donor site can be covered with a Xeroform (Sherwood Medical Industries Ltd., Markham, Ontario, Canada), which will dry into an eschar when exposed to air. Once dry this will be painless but is extremely sensitive until the eschar develops.

Rehabilitation

Rehabilitation of the affected extremity can begin once the surgeon is assured of stable graft take. Grafts placed over joints may benefit from 7 to 10 days of immobilization before initiating motion across the joint. For grafts placed away from a joint, normal motion can begin almost immediately if the graft has been securely bolstered to its wound bed.

Complications

Skin graft necrosis or failure is the most common complication following split thickness skin grafting. The formation of fluid under a skin graft, whether hematoma or seroma, is the most common cause of skin graft loss. As noted, the graft is completely reliant on the recipient bed for nutrition and so needs to be in complete contact with the recipient bed to survive. Therefore, precise hemostasis of the wound bed is critical before skin graft application as is firm compression of the graft onto the recipient bed through the use of bolster dressings.

Infection is the next most common cause of skin graft loss. This is best avoided by careful debridement before placement of the graft. Shearing forces can interrupt the formation of vascular connections between the graft and the recipient bed. This will ultimately lead to loss of a graft since healing relies on the formation of genuine vascular connections and is not possible through plasmatic imbibition alone. Therefore, firm immobilization of all skin grafts is critical to their healing.

Finally, an inadequately perfused recipient bed will certainly lead to skin graft loss. Common situations include patients with peripheral vascular disease, previously radiated tissues, and tendon or bone denuded of paratenon or periosteum.

Random Flaps

Random flaps, by definition, have no named or defined blood supply. They are raised in a subdermal plane and so rely on the subdermal vascular plexus of skin for circulation. To ensure adequate circulation, random flaps should be limited to a length no greater than 2.5 times the width of their base, which is the uncut border of the flap. This ratio may be even more limited in poorly perfused extremities.

Indications/Contraindications

Indications

The elasticity and slight redundancy of local tissue are critical to the success of most random flaps. They are based on local tissue so are typically contraindicated for wounds with significant surrounding soft tissue damage such as radiation wounds. The advantage of random flaps in comparison to skin grafts is that they can provide well-vascularized, full-thickness tissue for wound coverage. Far more durable to friction and repeated stress than skin grafts, local flaps are preferred for closure over vital structures such as tendons, nerves, and blood vessels. An added benefit is that they can be used to close wounds over vital structures denuded of paratenon, periosteum, or perineurium, since a well-vascularized wound bed is not required as it is for skin graft survival. Finally, despite the new scars made to raise random flaps, they are ideal in terms of color match, since they come from local skin.

Contraindications

Gross contamination or frank infection is an absolute contraindication to wound closure with local flaps. In fact, while free flaps and pedicled flaps have been shown to introduce enough independent new blood supply to overcome infectious processes like osteomyelitis, local flaps need to be placed over a clean, uncontaminated wound bed. Finally, local flaps are limited to small wounds usually less than 15 to 20 cm^2; wounds exceeding these dimensions should be covered with another method of soft

tissue coverage.

Preoperative Planning

As mentioned previously, adequate surgical debridement is a must before any attempt at closure. Tissue culture is a useful adjunct, particularly in cases of significant blunt trauma with devitalized tissue or cases of gross contamination. Often, serial wound debridement every other day for several days is necessary to obtain negative cultures.

It may not be possible to completely resolve a patient's systemic medical issues, but they should be as optimized as much as possible. Tight control of blood glucose levels in patients with diabetes, management of extremity edema, and nutritional status can all be significantly improved in many patients with several days of focused inpatient care.

Surgery

Z-Plasty

Probably the most familiar random flap to any surgeon, the Z-plasty is not actually indicated for the treatment of open wounds. Rather, the main indications for a Z-plasty include lengthening scars, interrupting linear scars with the transposition of unscarred tissue, and disrupting circumferential or constricting scars (Table 3-2). A common indication in hand surgery is to employ a Z-plasty in contracted web spaces in an effort to introduce healthy adjacent tissue relieving web contractures.

Z-plasties rely on limbs of equal length to facilitate closure. The most common Z-plasty design employs 60-degree angles, which theoretically results in a 75% increase in the length of the central limb of the Z-plasty (Fig. 3-11). Although clinically not feasible, a Z-plasty with angles of 90 degrees would result in the greatest theoretical gain of central limb length, approximately 120%.

Executing an effective Z-plasty is not simply a matter of elevating the triangular flaps and transposing them. Mobility and a tension-free closure are greatly facilitated when the tissue at the bases of the triangular flaps is also elevated.

Four Flap Z-Plasty

Most commonly used for first web space contractures, the four flap Z-plasty results in

a 150% gain in length of the original central limb, or scar contracture. Essentially, a 120-degree standard Z-plasty is drawn. Each triangular flap is then bisected, resulting in four equivalent 60-degree triangles that are raised and interdigitated as shown (Fig. 3-12).

Table 3-2. Z-Plasty Indications

Lengthening scars
Interrupting linear scars
Disrupting circumferential or constricting scars

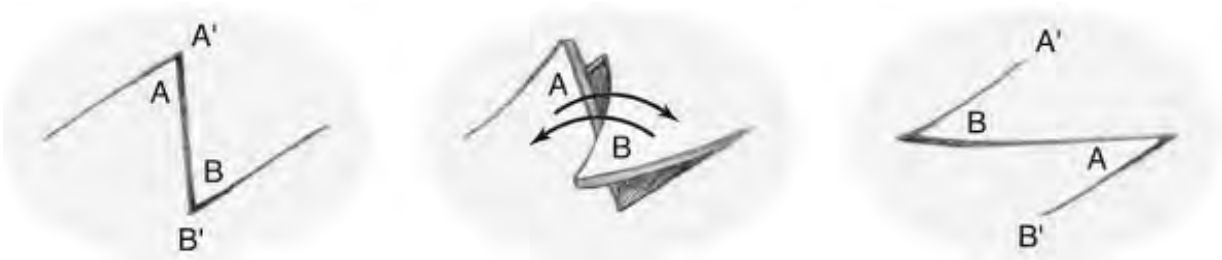

FIGURE 3-11 A basic Z-plasty.

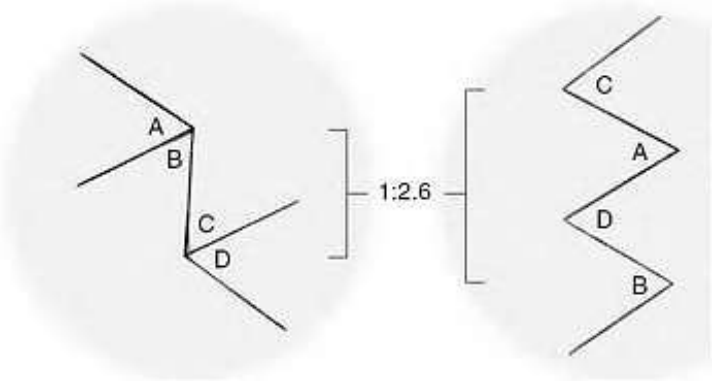

FIGURE 3-12 Four-flap Z-plasty.

Rhomboid Flap

Executing a rhomboid flap begins with converting the defect (even if it is circular) into a rhomboid to better visualize the flap design. A line equivalent in length to the limbs of the rhomboid is then drawn perpendicular to the short axis of the rhomboid (Fig. 3-13). Next, a line (B-C) is drawn at 60 degrees to the A-B line. It will be parallel to one of the limbs of the original rhomboid.

The flap is then elevated. It is also critical to elevate the skin around the base of the flap very liberally to facilitate transposition of the flap and closure of the wound. The flap should not be closed under tension; rather, further undermining of the base of the flap should be executed until the flap closes without tension.

Banner Flap

The banner flap is a type of transposition flap. A pendant or banner of skin is designed with one edge of the banner (near the base of the flap) running tangentially to the wound edge. The flap is elevated and transposed after which any redundant flap can be trimmed. It is important that the original design of the flap places the banner in an area of redundant skin. Since the banner can be designed along any border of the defect, it is helpful for eventual cosmesis if the original banner is designed within a relaxed skin tension line.

Rotational Flap

Rotational flaps are frequently employed on the dorsum of the hand, fingers, and in the scalp to close triangular defects. They are deceptively simple, and poor planning can lead to a large incision with an inadequate flap.

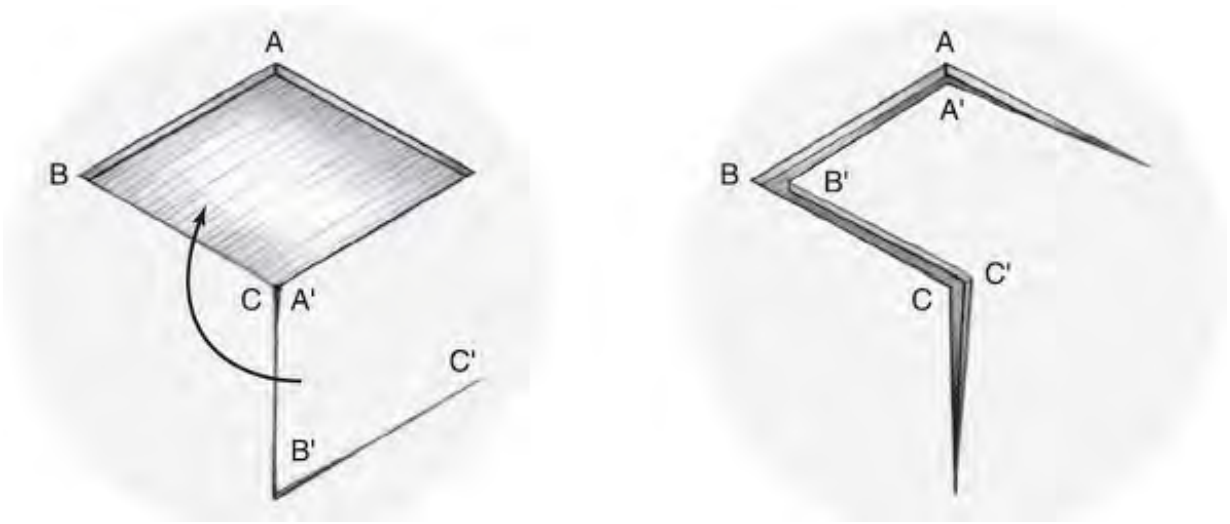

FIGURE 3-13 Rhomboid flap.

Unfortunately, relatively little advancement of the flap edge is possible without making a flap that is approximately four times greater in length than the defect length. Even with an adequately designed flap, tension can develop at a point opposite the pivot point in the base of the flap. This can be overcome by backcutting the base of the flap or excising a Burow's triangle. The pitfall here is the risk of cutting into the base of the flap, which can reduce the circulation of the flap. This added tension in the flap can lead to tip necrosis.

One simple solution for a rotation flap that seems to be under excessive tension is to simply advance the leading edge of the flap into the defect, which results in a new defect at the opposite end of the flap. Since this donor defect should be well vascularized (the flap should be taken from an area of healthy tissue), the donor site

can be skin-grafted rather than closed primarily. This strategy, while not elegant, is mush less likely to result in excessive tension on the closure.

V-Y Advancement Flap

V-Y advancement flaps are extremely useful, but their execution differs greatly from transposition flaps. V-Y advancement flaps are not elevated completely, but rather must remain connected to the subcutaneous tissue underlying the flap to maintain their viability. This is because they are incised along all skin borders and so have no connected â€œbaseâ€• through which a dermal or subdermal blood supply can provide circulation.

A V-shaped flap is designed immediately adjacent to the defect with the widest portion of the V equivalent in width to the width of the defect (Fig. 3-14). The two limbs of the V are then gradually tapered so that their length is at least 1.5 times the length of the desired advancement. The distance the edge of the flap can be advanced is somewhat determined by the laxity of the local skin.

Following advancement of the flap edge into the defect, the base of the V is closed in a linear fashion, resulting in a â€œY.â€•

Postoperative Management

On completion of the operation, a bulky non-compressive dressing is applied to the wound, which may be supported with a plaster splint if immobilization is required. A small window can be left to monitor the flap for signs of ischemia or congestion. Sutures remain for an average of 10 to 14 days, depending on the status and tension of the surrounding tissue. Ideally, flaps are inset under minimal tension, allowing the patient to begin gentle range of motion exercises immediately following surgery. Once sutures are removed, the wounds may be kept moist with a petroleum-based product, which will prevent itching during the early postoperative period.

Complications

The most feared postoperative complication following local flap reconstruction is necrosis or partial necrosis of the transferred tissue. This complication occurs because of inadequate blood supply to the flap. Flap ischemia may result from flap closure under excessive tension, hematoma, or infection.

If total flap loss develops, the wound will require closure by another means, either pedicled flap or skin graft. If partial necrosis has occurred, one can debride the necrotic portion of the flap and begin dressing changes to the underlying wound bed. Once the underlying tissues are clean of any infection and necrotic debris, the surgeon can consider allowing the wound to heal through secondary intention or attempt wound closure with an alternative reconstructive method.

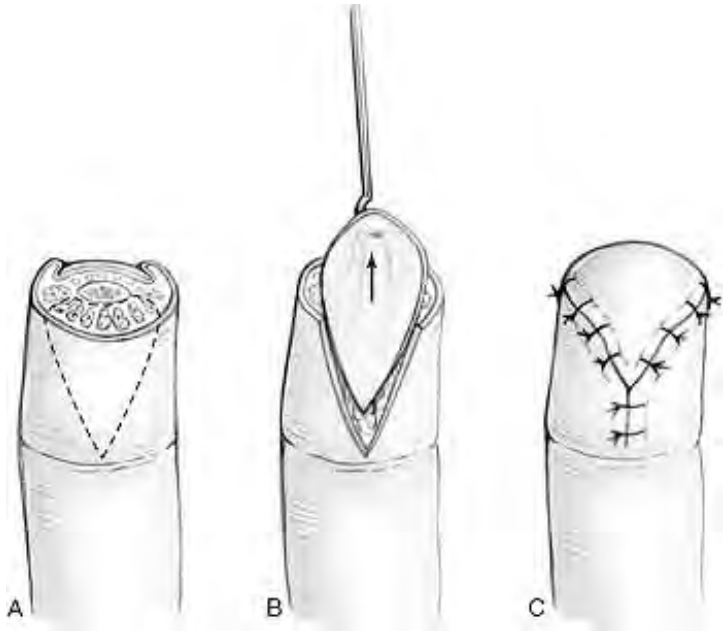

FIGURE 3-14 The Atasoy-Kleinert V-Y advancement flap. Note that the subcutaneous tissue has remained undisturbed to maintain blood supply. (Redrawn after Louis DS, Jebson PLJ, Graham TJ. Amputations. In: Green DP, Hotchkiss RN, Pederson WC, eds. *Operative hand surgery.* New York: Churchill Livingstone; 1999:48-94.)

Pearls and Pitfalls

Pearls

- Use a VAC or Reston bolster to immobilize and compress skin grafts quickly.

- Do not mesh pie-crust skin grafts on hands.
- Exploit secondary contraction of thin split-thickness grafts (0.010â€“0.012) to induce contraction of fasciotomy wounds.
- Employ Tegaderm for small donor sites (100 cm^2) to decrease pain and provide a moist wound-healing environment.
- Splint extremities to immobilize skin grafts.
- Salvage improperly planned local flaps by skin grafting the local flap donor site.

Pitfalls

- Hematoma
- Seroma
- Inadequate immobilization of grafts or flaps while healing
- Meshing grafts on the hand
- Inadequate initial wound debridement

Recommended Reading

Ablove RH, Howell RM. The physiology and technique of skin grafting. *Hand Clin.* 1997;13:163â€“173.

Chao JD, Huang JM, Weidrich TA. Local hand flaps. *J Amer Soc Surg Hand*. 2001;1:25â€“44.

Lin SJ, Hijjawi JB. Skin grafting. In: Lin SJ, Hijjawi JB, eds. *Plastic and reconstructive surgery pearls of wisdom.* New York: McGraw Hill; 2006:435â€“436.

Louis DS, Jebson PLJ, Graham TJ. Amputations. In: Green DP, Hotchkiss RN, Pederson WC, eds. *Operative hand surgery*. New York: Churchill Livingstone;

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 4

Vacuum-Assisted Closure in Extremity Trauma

Anthony J. DeFranzo

Louis C. Argenta

Indications/Contraindications

Wounds of the upper and lower extremities provide a constant challenge for orthopaedic and plastic surgeons. Complex severe injuries involving bone and soft tissue demand major reconstructive procedures. Less severe injuries exposing tendon, bone, and joints also demand innovative surgical approaches. Many patients with injured extremities have been the victim of multiple traumas and may have major associated injuries involving head, chest, and abdomen. Major blood loss, disseminated intravascular coagulopathy (DIC), high intracranial pressure, acute respiratory distress syndrome (ARDS), septicemia, and renal failure frequently complicate the treatment of these patients. Vacuum-assisted closure (VAC) has provided a way to effectively manage these wounds until definitive reconstruction can be performed. Vacuum-assisted closure has also greatly simplified reconstruction in many of these patients. Skin grafts are performed in many cases that would have otherwise required major rotational flaps or free flaps prior to the advent of VAC therapy.

The mechanics of the VAC system are straightforward. The system consists of an open cell polyurethane ether foam sponge sealed by an adhesive drape. All pores in the sponge communicate so that negative pressure applied to the sponge is applied equally

and completely to the entire wound surface. The effects of the VAC on the wound are multiple. The application of negative pressure causes the sponge to collapse toward its center. Traction forces are thus applied to the wound perimeter pulling the wound edges together progressively making the wound smaller. The VAC sponge should be cut to fit inside the wound to maximize traction forces on the wound edges. The sponge should not overlap intact skin, as skin maceration may occur. In addition, the VAC removes wound edema, appears to increase circulation and decrease bacterial counts, and significantly increases the rate of granulation tissue formation (1,6).

All wounds are thoroughly debrided prior to VAC placement. No clinical infection, purulence, or suspected osteomyelitis should be present prior to VAC placement. Grossly contaminated wounds are a contraindication for VAC placement. The VAC sponge should not be applied directly over major exposed blood vessels in the extremity status posttrauma, especially if the vessel wall has been damaged or a vessel repair has been performed. A protective interface such as Adapticw may be used. Exposed, repaired, or damaged major vessels should be covered with flaps such as muscle flaps or fasciocutaneous flaps. The VAC may be applied over intact nerves if muscle flap or adequate soft tissue coverage is not possible. A layer of Adaptic may help prevent pain caused by VAC traction especially when VAC sponges are changed. Nerve repairs should also be protected by an interface layer such as Adaptic placed under the VAC sponge. Adequate soft tissue or flap coverage of major peripheral nerves is preferred for final long-term management. The VAC may frequently allow secondary closure of wounds over major peripheral vessels or nerves by pulling together adequate soft tissue as edema is removed. Split-thickness skin grafts placed directly over major exposed vessels or nerves is possible with VAC therapy after granulation tissue occurs, but not preferred.

The VAC sponge may be placed directly on bone if the bone surface bleeds following sharp debridement by an osteotome or low-speed power burr. Occasionally, desiccation of bone may occur

with the VAC. If desiccation occurs, further debridement to punctuate bone surface bleeding with the immediate placement of Integraw Dermal Regeneration Template and then VAC placement directly over Integra may be tried. This technique has provided Integra take over bone allowing subsequent skin graft coverage. The Integra must be meshed or holes cut in the outer silicone layer before VAC sponge placement.

Relatively small areas of bone have been successfully treated in this manner. Large areas of exposed bone are better treated with pedicle muscle flaps, pedicle fasciocutaneous flaps, or free flaps. The VAC should not be placed over desiccated bone with questionable viability without punctuate bone surface bleeding.

Preoperative Planning

Wounds should be thoroughly debrided before VAC application. Once the wounds are clean, VAC therapy may be initiated. VAC sponges can initially be applied in the operating room following definitive debridement.

Surgery

To illustrate the application of the VAC device to wounds of the extremities, case studies will be shown to illustrate degloving injuries, gunshot wounds, and a variety of avulsion injuries with exposed tendons, bones, and joints. VAC treatment of traumatic wound complications such as infection and hematoma is also illustrated.

Degloving Injuries

Several degloving injuries to the hand have been treated by the VAC with varying results from 0% to 95% take of replaced degloved skin. The degree of take depends on the condition of the degloved skin and the viability of underlying tissue. The degloved skin is frequently crushed and lacerated. It may be difficult to clinically assess the viability of the degloved avascular skin. Also, roller injuries which frequently cause degloving injuries of the hand may crush as well as deglove the hand. Contusion to the hand intrinsic muscle and crush injury of blood vessels in the hand may also be difficult to evaluate. Progressive loss of viability may occur.

In all hand degloving injuries (N = 6) treated to date at Wake Forest University School of Medicine, fingers were not viable significantly beyond the proximal interphalangeal (PIP) joints and amputations were made at or just distal to that level (4). The thumb has been salvaged at full length in one case. In degloving injuries, most extensor and flexor tendons remain on the hand. If the distal phalanx is avulsed with the skin envelope the flexor digitorum profundus tendon(s) may also be avulsed from the hand and even from the forearm. Extensor peritenon and flexor tendon sheaths are frequently intact, providing vascularized coverage to the tendon.

The clinical approach has been to assess viability of the degloved skin and hand after meticulous debridement and completing the amputations at the appropriate level. The skin is thoroughly defatted, pie-crusted, and placed back on the hand with a few staples and/or sutures. The VAC is then immediately applied. The sponge is cut to fit over the grafted, degloved skin, using one or multiple contiguous pieces of sponge. A â€œhand VACâ€• kit is also commercially available. Controlled suction of 125 mm Hg pressure is applied (Fig. 4-1). This procedure takes very little operative time. If successful, the patient is spared a great deal of overall treatment time and morbidity compared to more complicated methods of reconstruction. Therefore, with a degloving injury, our treatment algorithm is as follows: (a) skin defatting and debridement, (b) wound debrideded, (c) skin applied to wound, (d) VAC applied to wound. The VAC is left in place for 4 to 6 days. If there are areas which have not revascularized, the VAC may be replaced for 2 more days for a total of 6 days. After day 6, further take with VAC therapy does not seem to occur and we have not used the VAC longer than 6 days.

If VAC therapy is totally unsuccessful, no more than 6 days are lost with the attempt at degloved skin replacement before other methods of reconstruction are initiated. The best case to date of VAC replacement of degloved skin achieved a 95% take of avulsed skin with preservation of fingers at or just distal to the PIP joints and preservation of the thumb at full length. Excellent range of motion of the metacarpal phalangeal joints was achieved. Further surgery involved outpatient revision of the amputations of the four amputated fingertips and deepening of the first web space (Fig. 4-2).

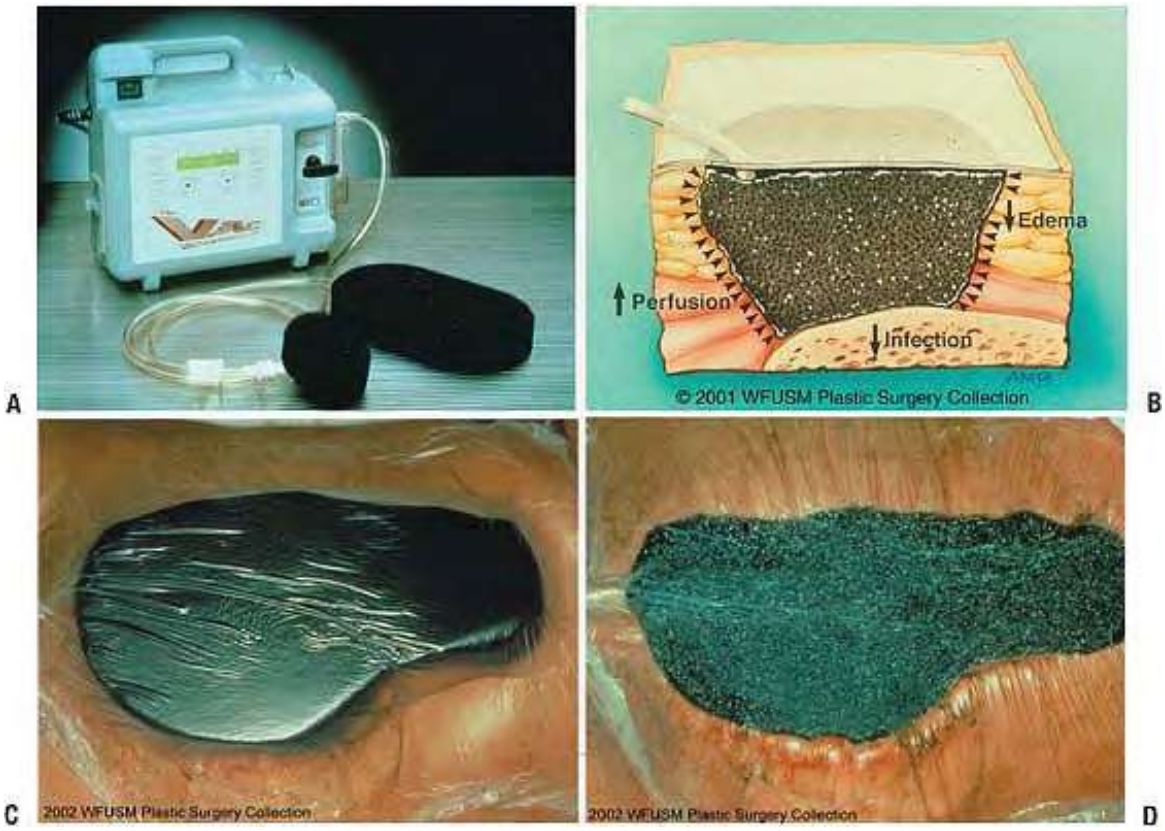

FIGURE 4-1 Mechanics of the VAC system. A: Open cell polyurethane ether foam dressing. B: A 400â€“600 micron pore diameter. C: Foam dressing cut to fit the wound. D: Wound sealed with adhesive drape. There is 50â€“125 mm Hg controlled suction applied continuously.

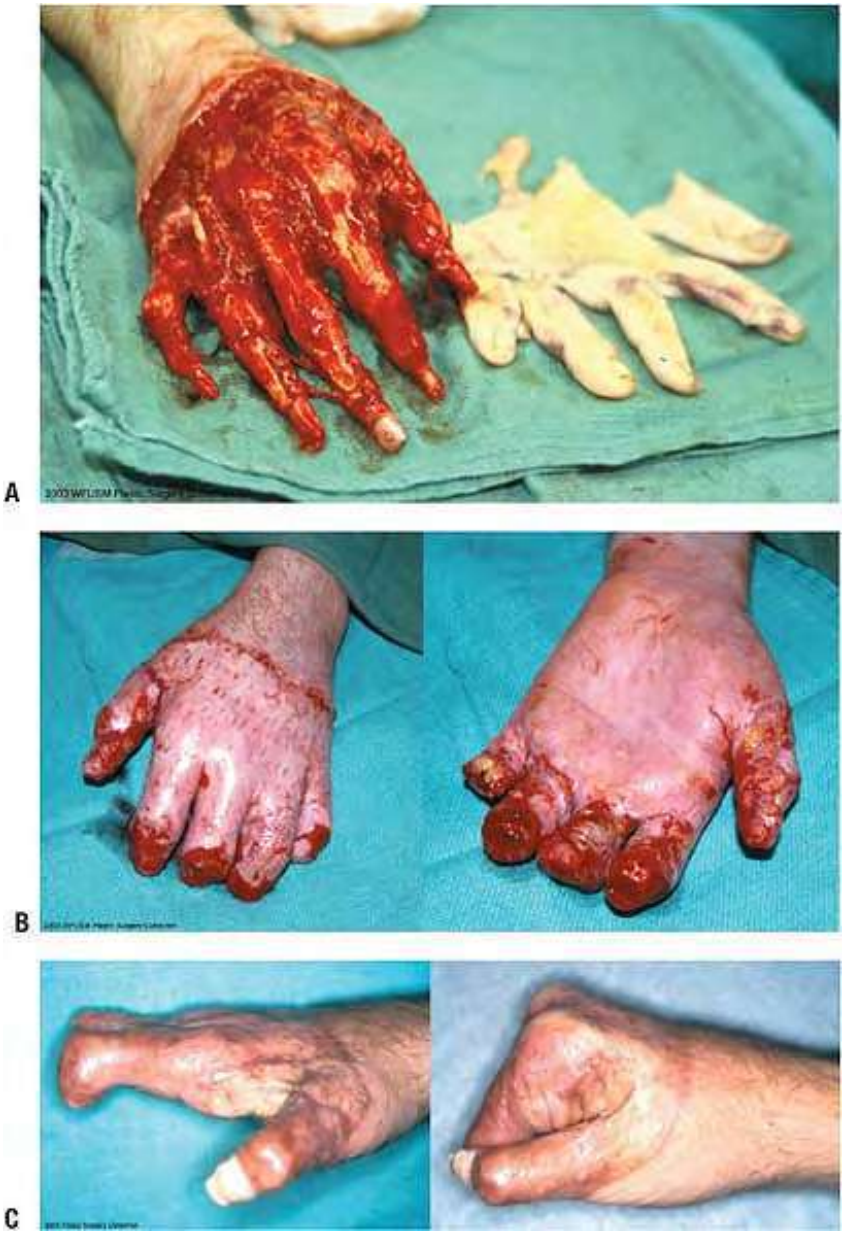

FIGURE 4-2 Degloving injury to the hand. A: Entire dorsum and all fingers avulsed to distal palm. Neurovascular bundles present to or just beyond proximal interphalangeal joints. Extensor paratenon intact; flexor tendon sheaths intact. B: Fingers surgically amputated at or just distal to PIP joints. Skin defatted, pie-crusted, and reapplied. VAC change day 4 with 95% take of reapplied degloved full-thickness skin. C: Final result with good range of motion at metacarpal phalangeal joints.

One degloving injury of the foot has been treated with the VAC. The toes were amputated with the degloved skin at the time of the injury. Metatarsals were shortened surgically so that they were covered by viable soft tissue. The skin was thoroughly defatted, pie-crusted, and returned to the foot. A few staples were used to secure the skin, and the VAC was applied for 4 days. Skin viability was 95%. No revisional surgery has been required. A small area of breakdown on the plantar surface of the heel occurred with ambulation, but healed with simple dressing changes and has not been a recurring problem with proper footwear (Fig. 4-3).

Crush Avulsion Injuries with Exposed Bone, Joints, and/or Tendons

Crush/avulsion injuries are commonly seen after industrial accidents or accidents at home with lawnmowers and farm machinery. Crush/avulsion injuries are also frequently seen with motor vehicle accidents and may be associated with other major critical injuries. Expedient closure of wounds without major reconstructive procedures with long anesthesia times may be beneficial to overall patient

care in the critically injured patient. Injuries to the upper extremity and lower extremity from the knee to the ankle result in exposure of tendons, bones, and joints. The VAC has been used on such patients only after one or more thorough operative debridements of all nonviable tissue. In some of these cases, coverage of tendon, bones, and joints has been facilitated by the use of Integraw.

FIGURE 4-3 Degloving injury to the foot. A: Entire dorsum, plantar surface, and skin of posterior heel avulsed. All toes traumatically amputated at metatarsal phalangeal joints. B: Amputations revised at metatarsal phalangeal joints. Avulsed skin defatted, pie-crusted, and replaced. The VAC is removed on day 4 with 95% of replaced skin viable. C: Final result with durable cover with proper footwear at 2 years.

Integra Dermal Regeneration Template is a clear bilayer membrane designed originally for skin replacement in burn patients. It consists of an epidermal layer comprised of a clear thin sheet of silicone (polysiloxane) and a dermal layer comprised of a porous matrix of bovine collagen fibers cross linked with chondroctin-6-sulfate from shark cartilage. The silicone sheet is removed after take of the Integra, and a split-thickness skin graft typically 8 to 12 thousandth of an inch is applied. With VAC therapy over pie-crusted or meshed Integra (noncrushing mesher), take occurs in as little as 6 days. Our practice is to maintain VAC therapy for 6 days undisturbed over Integra. Clinically, the Integra assumes a salmon-pink color when take occurs; but with 6 days of VAC therapy, a more robust red color may be achieved. We now routinely plan to perform split-thickness skin grafts 6 days after VAC placement over Integra.

Integra and split-thickness skin graft provide a pliable durable bilaminar skin reconstruction ideal to cover joint surfaces (3). With VAC therapy, Integra has taken well over a vascularized wound bed, at 6 days in most patients. In addition to simple Integra take, however, viable tendons, bones, and open joints without cover have been bridged for a distance of 1 to 2 cm. At day 6, if structures have not been â€œbridgedâ€• sufficiently, VAC therapy may be continued for a total of 10 to 12 days. Red vascular tissue can be seen spreading transversely through the Integra that is bridging a bone, joint, or tendon.

Tendons and bones can granulate in the abscence of Integra to allow split-thickness skin graft take (2). However, the use of Integra appears to provide a more durable skin cover over bone and over tendon allows tendon glide with appropriate early occupational therapy. Split-thickness skin graft applied over Integra is followed by VAC therapy for another 6 days to achieve take of the split-thickness skin graft to complete the reconstruction. The VAC has been useful in maximizing the take of split-thickness skin grafts over irregular surfaces in many of our cases involving the extremities (7). Whenever

P.54
P.55
P.56

possible, range of motion of joints and tendons is begun early at 1 to 2 weeks after successful skin graft take (Figs. 4-4,4-5,4-6,4-7).

FIGURE 4-4 A 68-year-old status post lawnmower avulsion injury. Smoker with atherosclerosis. A: Exposed medial malleolus, ankle joint, and tendon. B: VAC therapy for 2 weeks with granulation tissue covering exposed tendons and joint. No use of Integra[W]. C: Split-thickness skin graft stable for 4 years with normal footwear.

FIGURE 4-5 A 67-year-old status post motor vehicle accident with crush of lower extremity and fracture of underlying tibia treated with an intramedullary rod. A: Exposure of rod post status debridement. B: VAC therapy allowed soft tissue to cover bone by both a decrease in wound size and the formation of granulation tissue. Final skin closure was obtained with a split thickness skin graft. Wound stable at 1 year.

FIGURE 4-6 A 17-year-old with skin avulsion of the left arm status post motor vehicle accident. A: Exposed ulna is debrided to bleeding bone. B: VAC applied for 6 days with formation of granulation tissue. Integraw was then applied and VAC continued for 7 days. Finally split-thickness skin graft was applied. C: Stable cover after 3 years for definitive wound closure.

A

B

C

D

FIGURE 4-7 A 24-year-old male status post motorcycle accident with right brachial plexus injury, right clavicle and right scapula fracture, right occluded axillary artery, right pneumothorax, right open comminuted patella fracture with avulsion of patellar bone and soft tissue with an anterior open knee joint. A: Open knee joint post status debridement of nonviable patella and soft tissue. B: Soft tissue deficit following ligament reconstruction and VAC placement.C: Integraw was placed over the open knee joint. D: Granulation bridging knee joint through Integra. E: A split-thickness skin graft was applied over the Integraw after the Integra showed signs of levascularization. F: Excellent range of motion achieved at 3 months with stable cover at 2.5 years.

Gunshot Wounds

Gunshot wounds are associated with blast effect which causes major damage to bone and surrounding soft tissue. Massive wound edema develops. Reconstruction can become exceedingly difficult. When applied immediately after proper debridement, VAC therapy can remove a large amount of edemic fluid from the wound. Multiple operative debridements may be required to ensure that all nonviable bone and muscle have been removed. As much as 2 L of edema fluid have been removed from wounds of the lower extremity over a 24 hour period (5). Application of VAC can also pull retracted wound edges together and promoted the remaining wound to fill with granulation tissue covering exposed bone, tendon, and hardware (Figs. 4-8 and 4-9).

Vacuum-Assisted Closure Therapy for Posttraumatic

Wound Complications

Posttraumatic wounds are subject to higher hematoma and infection rates due to tissue devascularization and tissue contamination. If a wound dehisces, rapid edema may make wound reclosure impossible. Bone, tendon, and/or plates become exposed and their coverage is essential to a successful outcome. Vacuum-assisted closure therapy has been useful in regaining closure of posttraumatic open wound complications (Fig. 4-10).

FIGURE 4-8 A 25-year-old with a self inflected gunshot wound to the left arm. A: Debridement and plate fixation of the humerus. B: Exposed plate status post debridement with decreased wound size and granulation tissue covering plate

A

B

C

D

FIGURE 4-9 A: A 10-year-old boy with a gunshot wound to the right elbow, forearm, and hand extremity well vascularized. VAC placed on day 4 status post debridement of all nonviable tissue. Three operative debridements were required prior to VAC placement. B: Early post-operative course showing a small amount of exposed bone. The soft tissue defect was advanced to closure by VAC therapy. C: Split-thickness skin graft placed 10 days status post injury after granulation tissue covers all exposed bone. D: Wound stable 2 years. Excellent bone healing and range of motion.

FIGURE 4-10 A 32-year-old status post fractured calcaneus treated by open reduction and plate fixation. A: A post-operative hematoma resulted in wound dehiscence and exposed the plate. B: Four weeks of VAC therapy were used to obtain wound closure (3.5 weeks as an outpatient). No surgical procedure to achieve skin closure was required. Wounds stable for 3 years with normal foot wear.

Postoperative Management

We believe that the VAC sponge ideally should not be left longer than 2 days between changes to maintain proper seal and to avoid infection. Small, clean outpatient

wounds may be changed three times a week (Monday, Wednesday, Friday schedule), by visiting home nurses with a 2â€“3 day change interval.

Oral or intravenous narcotics are usually sufficient for VAC changes in adults. Infusion of Xylocaine into the sponge discontinued from suction 15 minutes before VAC change has been beneficial for local analgesia prior to VAC change. Care must be taken to administer doses of Xylocaine usually well below safe standard dose guidelines. A dose of 10 to 20 cc of 1% Xylocaine with epinephrine for a small to medium size sponge would be sufficient. Much larger wounds may not benefit from Xylocaine infusion. Children granulate more quickly than adults and may require a smaller pore size sponge (white sponge) than the standard black sponge. Granulation tissue may grow into the black sponge and cause increased pain or bleeding with the VAC sponge change in children. Every-24-hour VAC changes may occasionally be required with a standard black sponge in children. However, with granulation tissue development, split-thickness skin grafting is possible even sooner in children, and the uncomfortable course of VAC sponge changes is soon over. Children may require a brief general anesthetic or Ketamine for each VAC change in a dressing room setting staffed by an anesthesiologist.

Complications

Complications are experienced occasionally due to mechanical failure of the machine with loss of suction and unfortunate wound deterioration. The wound may become a sealed abscess cavity if suction is not properly maintained. All new machines have alarms that should signal loss of suction. Every-2-day VAC sponge changes ensures that proper machine function and seal are provided. Technical errors such as an improper seal have occurred with inexperienced personnel. In-service training is mandatory for all nursing and physician staff. Sponge overlap of the skin will cause skin maceration. Bleeding may occur with VAC change, especially in children. Bleeding can usually be controlled with pressure. Generalized bleeding from the wound usually means that sufficient granulation has occurred to allow skin grafting.

Osteomyelitis has occurred infrequently if insufficient bone debridement was performed prior to VAC therapy. Granulation tissue unfortunately may quickly cover a bony sequestrum with VAC therapy. Significant drainage from such wounds soon occurs and points to the need for further debridement of the nonviable bony sequestrum. All

questionable bone should be aggressively debrided prior to VAC placement so that all remaining bone bleeds well.

A small number of plates have become infected after VAC coverage with granulation tissue. Plate infection jeopardizes bony union and delays skin closure or skin grafting. Loose plates and screws or plates that do not conform tightly to underlying bone are likely to become infected. Relatively small plate exposures have been successfully treated with the VAC. Major plate exposure should not be treated with the VAC. Clinical judgment should dictate that a major plate exposure, especially with poor-quality surrounding tissue, must be treated with some form of flap.

Results

Vacuum-assisted closure has decreased the requirements for complex wound closure in the extremities. Fewer major pedicle flaps and fewer free tissue transfers are required. With vacuum-assisted closure of traumatic extremity wounds, the number of free tissue transfers now required at our institution is approximately 33% when compared to the number prior to VAC therapy. Long operative and anesthesia times are avoided, which is especially beneficial for critically severely injured patients.

References

1. Argenta LC, Morykwas MJ. Vacuum-assisted closure: A new method for wound control and treatment: Clinical experience. *Ann Plast Surg.* 1997;38(6):563â€“577.

2. David L, DeFranzo A, Argenta L, et al. The use of vacuum-assisted closure therapy for the treatment of lower extremity wounds with exposed bone. *Plast Reconstr Surg.* 2001;108(5):1184â€“1191.

3. DeFranzo AJ, Argenta LC, David LR, et al. Treatment of traumatic hand/upper extremity wounds with the VAC wound healing device. *Plast Surg Forum.* 2002;264â€“265.

4. DeFranzo AJ, Marks MW, Argenta LC, et al. Vacuum-assisted closure of

Originally, peripheral nerve injuries were described by Sir Herbert Seddon in 1943 as neurapraxia, axonotmesis, and neurotmesis (1). The classification was later expanded by Sunderland and further defined by Mackinnon to include six degrees of injuries (Table 5-2) (2). First degree (neurapraxia) and second degree (axonotmesis) injuries recover spontaneously, the latter at the classic rate of 1 in/month or 1 to 1.5 mm/day (3). Third degree injuries must regenerate through some amount of scar tissue, thus recovery is variable and less than normal depending on the amount of scar tissue around the nerve. In fourth degree injuries, also known as a neuroma-in-continuity, regeneration is blocked by scar tissue within the fascicles and recovery is usually very poor or does

not occur. A fifth degree nerve injury (neurotmesis) is a transection of the nerve, which always will require surgical repair, while the sixth degree nerve injury encompasses a variety of nerve injuries within a single nerve. The difficulty with surgical correction of a sixth degree injury is limiting the repair to the fascicles affected by fourth and fifth degree damage and not damaging the fascicles with the potential for spontaneous recovery (Fig. 5-1).

neurapraxia
Degree I

conduction block resolves
fast/excellent

spontaneously

axonotmesis
Degree II

axonal rupture without interruption of the basal lamina tubes
slow/excellent

Degree III

rupture of both axons and basal
slow/incomplete

lamina tubes, some scar

Degree IV
(neuroma-in-continuity)
complete scar block
none
neurotmesis
Degree V

complete transection
none

Degree VI

combination of I-V +/- normal

fascicles
mixed

Seddon	Sunderland	Mackinnon	Injury	Recovery

Table 5-2. Classification of nerve injury

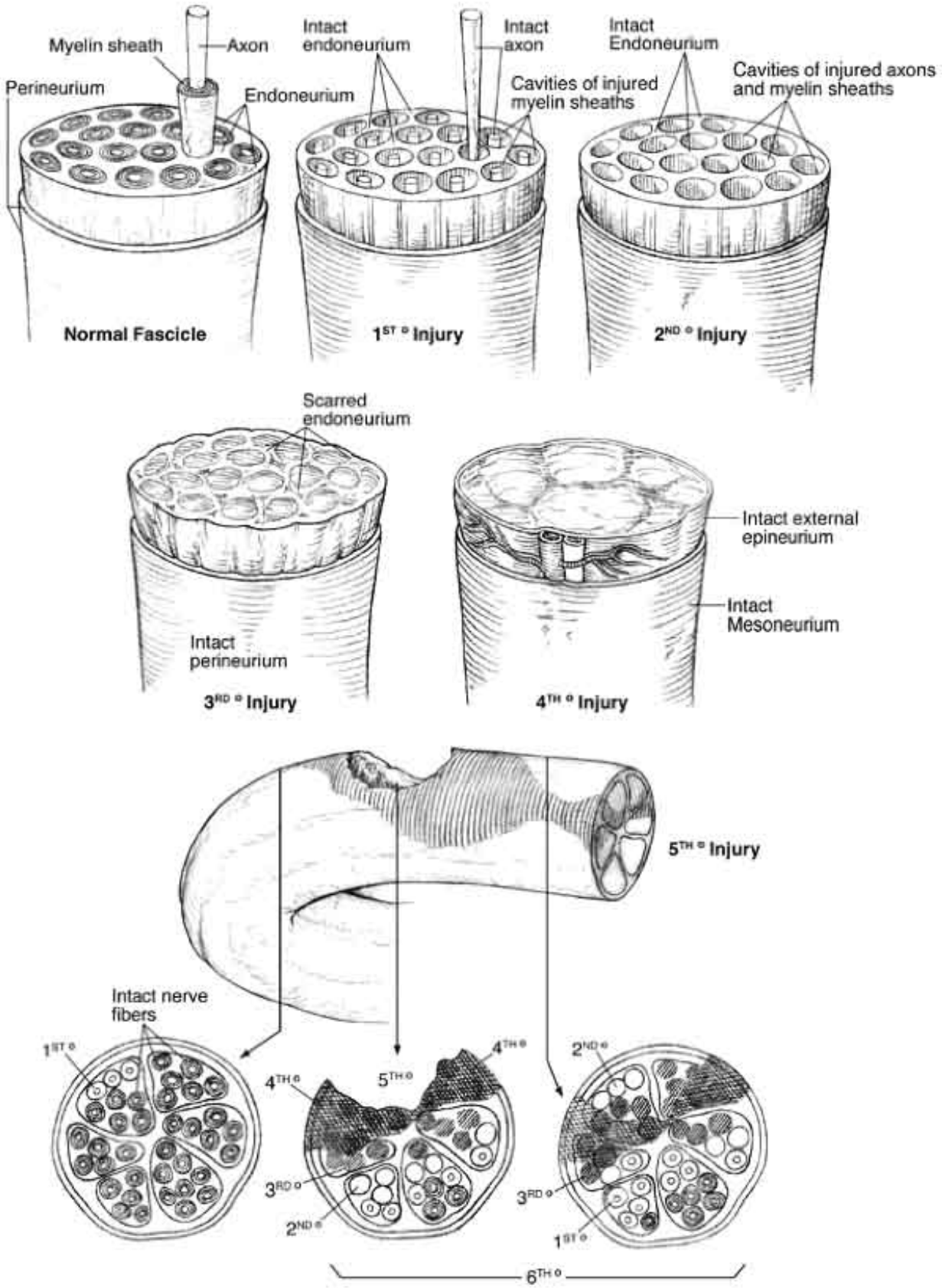
Myelin sheath
Axon
Perineurium
Endoneurium
Normal Fascicle
Intact endoneurium
Intact axon
Cavities of injured myelin sheaths
1ST ° Injury
Intact Endoneurium
Cavities of injured axons and myelin sheaths
2ND ° Injury
Scarred endoneurium
Intact perineurium
3RD ° Injury
Intact external epineurium
Intact Mesoneurium
4TH ° Injury
5TH ° Injury
Intact nerve fibers
1ST °
4TH °
5TH °
4TH °
3RD °
2ND °
2ND °
3RD °
1ST °
6TH °

FIGURE 5-1 Schematic depiction of nerve injuries from first through sixth degree. (By permission of Mayo Foundation for Medical Education and Research. All rights reserved.)

Stretch and Nerve Avulsion Injuries

Traction on a nerve can result in a mild stretch injury to the axon with preservation of the basal lamina. The Schwann cells around both myelinated and unmyelinated fibers may be locally injured and, until the cells are replaced from surrounding Schwann cells, will cause a conduction block. Such injuries are classified as first degree injuries and recover completely. As the pulling forces increase across the nerve, the basal lamina, the endoneurium, and perineurium become injured. As the nerve heals, internal scaring of the nerve may occur in some of the fascicles, leading to incomplete recovery. Grossly, the nerve will appear intact as the epineurium is usually not violated. Occasionally, neurolysis or even an internal neurolysis of the injured nerve may improve recovery as long as there is evidence that the nerve recovery is being hindered by scar surrounding the nerve, rather than internal scar to the fascicles.

Nerves that are stretched beyond the breaking point will avulse, resulting in gross disruption of the epineurium, perineurium, and endoneurium. A neuroma will typically form at the proximal end, and these injuries are treated with excision of the neuroma and primary repair when possible. Often, if the time from injury is greater than 2 weeks, even excessive mobilization of the nerve ends will not be enough to overcome the resulting gap that occurs when the unhealthy neuroma tissue at both ends is resected. A nerve graft is then needed to repair the gap.

In extreme cases, nerves can be avulsed from their insertion into the spinal cord. In the past, these injuries were treated conservatively. Once no further nerve recovery was noted, tendon transfers were used to restore the residual functional deficits (4). More recently, the introduction of nerve transfers allows for rewiring of nonfunctioning nerves by using local uninjured nerves to restore electrical continuity to the deinnervated muscles (5). Tendon transfers may be used to augment the nerve transfers once maximum recovery is achieved (6).

Nerves avulsed at the neuromuscular junction present a different problem. Nerves that are injured just prior to entering the muscle, or shortly thereafter, may still be repaired or grafted in most cases as long as a large enough nerve stump is found. Motor nerves that are avulsed from the muscle bellies are treated by implanting proximal nerve, when available, directly into the muscle with the hope that some of the fibers will find a neuromuscular junction and reinnervate at least part of the muscle. Some studies (7) show as good as M4 motor recovery 1 to 2 years after direct nerve to muscle neurotization; however, experimental studies do not support these findings. Rather, recovery is much less than a nerve coaptation would produce (8).

Crush Injuries

Crush injuries comprise the most common peripheral nerve injury to the extremity. External compression may be complicated by increased internal pressure from hematomas, fractures, and local tissue edema. When minor, this may cause a temporary neurapraxia, but with greater compression the likelihood of permanent injury increases. The most severe consequence of a crush injury is the progression to compartment syndrome. Often an early sign of impending compartment syndrome is a decrease in vibration sensibility (9). Compartment syndrome of the upper extremity and lower extremity are surgical emergencies and are reviewed separately within this text.

Nerve compression injuries may also develop distal to the actual soft tissue trauma. Local edema and inflammation after injury can exacerbate a preexisting condition, such as a mild carpal tunnel turning into an acute event after a distal radius fracture. Occasionally, on a case by case basis, surgical decompression is necessary, even though the majority will resolve spontaneously. Likewise, an anterior cruciate ligament tear of the knee requiring reconstruction or repair may precipitate an acute foot drop postoperatively despite the deep peroneal nerve being usually uninjured. Some surgeons postulated that intraoperative positioning of the limb may be a factor in the late development of palsy (10); however, we believe that local tissue inflammation probably exacerbates a preexisting condition that manifests itself as a postoperative foot drop. When conservative nonoperative measures do not lead to sufficient improvement in nerve function after 2 to 3 months, decompression of the peroneal nerve should be considered (11). Figure 5-

2 shows an algorithm for treatment of closed nerve injuries.

Penetrating Injuries

Blunt penetrating trauma is usually more locally destructive than a sharp injury. Often nearby structures such as blood vessels and tendons are injured in addition to the nerve. However, the size of a sharp laceration, such as with glass or a knife, can mislead the surgeon into underestimating the extent of injury. A seemingly small skin laceration may in fact extend under the surface and result in

a bigger injury than would be expected from the external size of the wound. Exploration is imperative if a nerve palsy is present, as the likelihood that the nerve is partially or completely transected is high. It is recommended to explore these injuries semi-electively within the first 2 weeks. The further from the time of injury, the more likely a nerve graft will be needed to overcome the resulting nerve gap. In the event of a penetrating trauma with an associated vascular injury, immediate exploration is warranted. Often the nerve injury is overlooked and not identified in the face of a more urgent vascular injury. In such cases the functional deficit may be first noticed postoperatively, when it is unclear if the nerve injury is from the inciting event, iatrogenic during the repair of the vascular injury, or secondary to edema or hematoma. While a CT scan or MRI may be helpful to evaluate for the latter, internal scarring of the nerve may not always be seen.

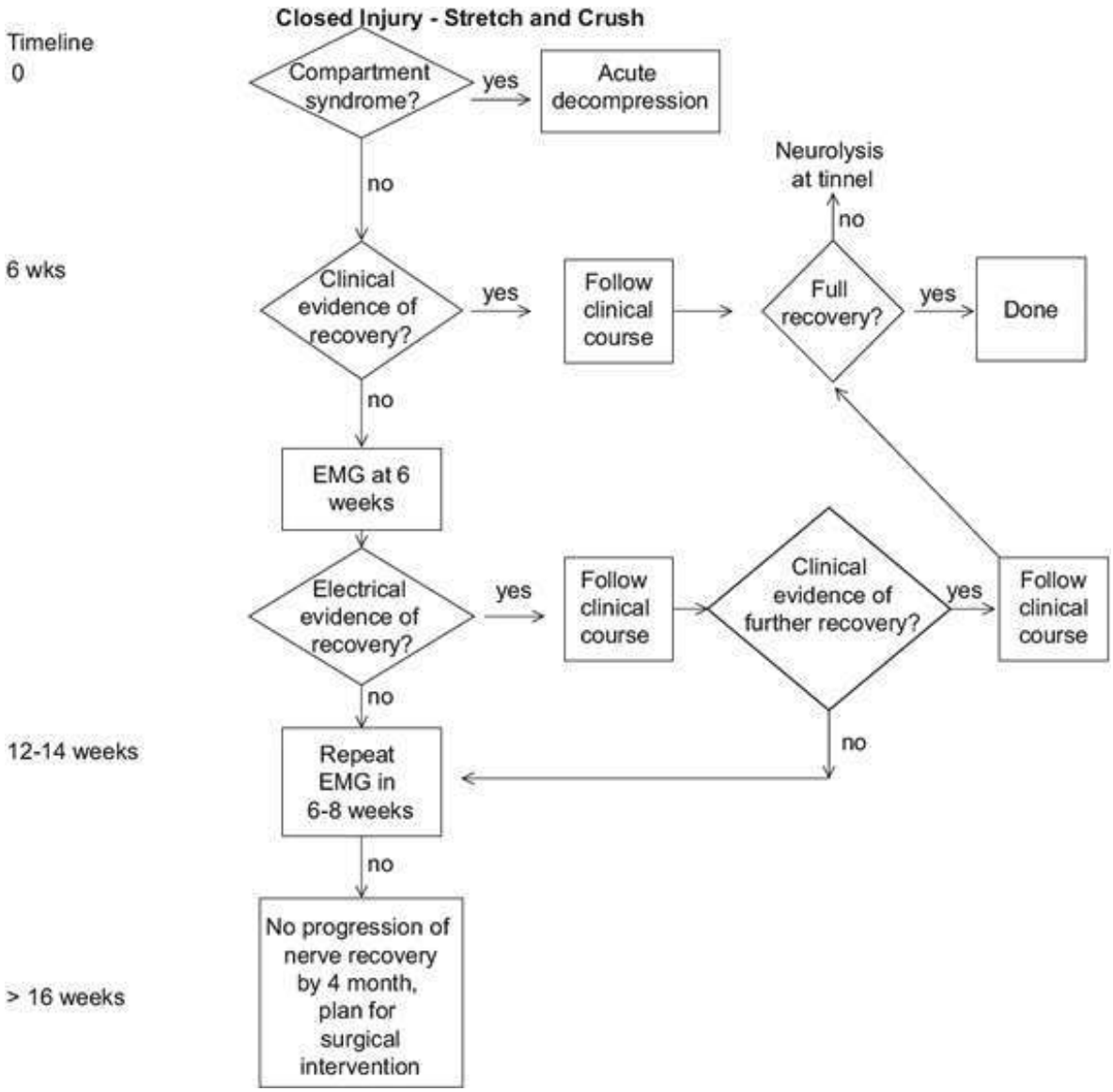

FIGURE 5-2 Algorithm for closed peripheral nerve injuries such as stretch and crush injuries.

Blunt penetrating injuries are initially treated conservatively, similar to closed crush and stretch injuries, because they are may recover spontaneously. The local tissue edema often causes a neurapraxia that resolves; however, those that do not recover after 3 months should be evaluated by electrodiagnostic studies and treated as a traction injury. Figure 5-3 shows an algorithm for managing nerve injuries and the timing of additional studies.

Two specific blunt penetrating injuries deserve special mention: gun shot wounds

and electrical injuries. Gun shot injuries present a unique problem since the trajectory of the bullet is unpredictable. The type, caliber, and velocity of the bullet each play a role in tissue destruction (12). The belief that the higher velocity bullet causes more tissue destruction can lead to early massive debridement; however, current recommendations are for judicious debridement and staged exploration,

as in crush injuries (13). In fact, a low velocity bullet like a slug that fragments and stays in the soft tissue will more likely lead to infection and local tissue destruction than an Army issue missile that passes through the tissue with relative ease (14).

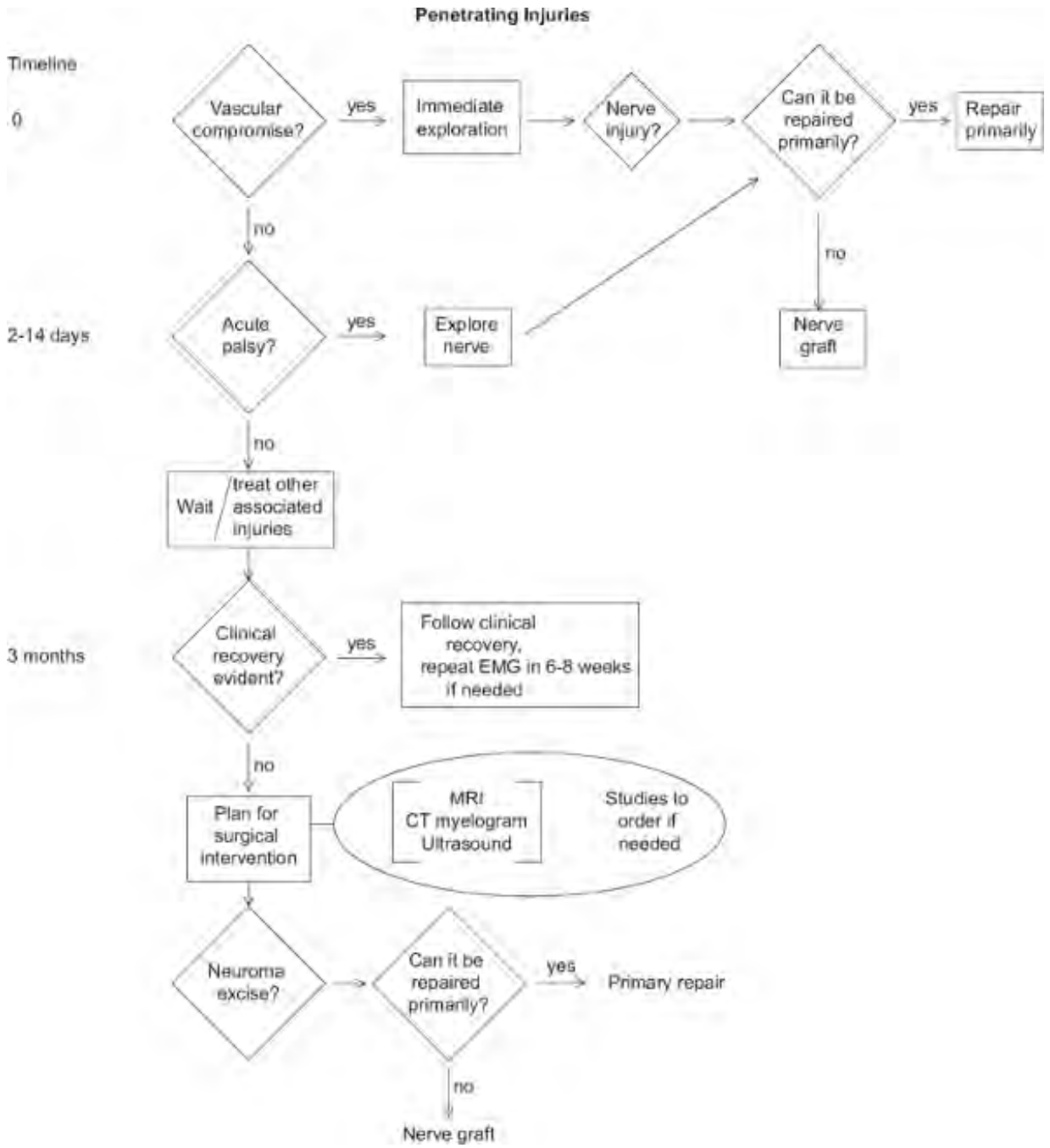

FIGURE 5-3 Algorithm for penetrating injuries.

Electrical injuries that are not severe enough to cause death from heart arrhythmia and respiratory paralysis cause vigorous nerve stimulation which leads to paralysis and vasospasm. Massive muscle contractions due to nerve stimulation or the direct triggering of striated muscles can cause muscle rupture, ligamentous tears,

fractures, and joint dislocations (15). In addition, electric current damages tissue from direct thermal heating; the coagulation necrosis is similar to a burn and is managed in the same manner (16).

One long term sequelae of gun shot wounds to the peripheral nervous system known as causalgia (complex regional pain syndrome I) was first described during the Civil War by Silas Weir Mitchell. The patients present with burning pain, paresthesias, skin atrophy, and temperature changes typical of the syndrome. This condition is also seen in electrical injury patients who survive the electrical contact. Axonal injury is usually from the direct thermal injury. Myelin injury is from direct thermal contact as well as from the electrical destruction of the myofibrils in the muscle. Post-neurological symptoms to the peripheral nervous system can vary from neuropathy to reflex sympathetic dystrophy (complex regional pain syndrome II) (17).

Paresthesias are thought to be due from perineurial fibrosis resulting in symptoms of a compressive peripheral neuropathy (18); surgical decompression can sometimes ameliorate the symptoms. Sympathectomy has been used to provide relief in those patients that respond to sympathetic blocks (19). In general nerve injuries associated with gun shot and electrical injuries are treated as closed nerve injuries.

Nerve Injury Associated with Soft Tissue Avulsion Injury

Not to be confused with an avulsion of the nerve itself, a soft tissue avulsion or degloving injury presents a unique problem of coverage. Nerves that are exposed will need coverage to prevent desiccation. The type of soft tissue coverage will depend on the location of the injury and is addressed elsewhere in this book. The coverage can be as simple as replacing missing skin with a skin graft to something more extensive that requires fascia or muscle with skin graft, and ultimately a musculocutaneous or fasciocutaneous flap. If the soft tissue is missing directly over the nerve repair, a flap is needed to cover the nerve repair; this can be either local or a free tissue transfer. The use of vacuum assisted closure dressings has changed our management of complex open fractures and may be used to cover a wound with

exposed nerves as well. If the nerves are exposed but uninjured, a protective dressing that keeps the nerves from desiccation may be used, such as a layer of AlloDermw with or without a wound vacuum assisted closure dressing, or one of the various hydrogels on the market that provides enzymatic debridement at the same time as it absorbs excess fluid from the wound. If the nerves are injured and need to be repaired, primarily or with a nerve graft, the definitive repair should be staged so that it is done at the time of the soft tissue.

Nerve Injury Associated with Bony Injury

Nerve injuries resulting from isolated closed fractures are most often due to compression from surrounding tissue edema or are due to, though less frequently, nerve laceration from the fracture ends. Iatrogenic nerve injury may also occur in the course of fracture fixation (20 ,21).

The anatomic positions of the radial, median, and ulnar nerves and their major branches make them vulnerable at several sites as they course the upper extremity, which explains the typical relationship seen between particular nerve injuries and fracture patterns (22). The more common nerve injuries with associated fracture patterns are listed in Table 5-3 and include the distal humeral shaft fractures and radial nerve palsy (23), posterior interosseous nerve injury with Monteggia fracture-dislocations (24), and median nerve and/or radial nerve injuries with supracondylar and medial epicondyle fractures in children (25). The time frame to intervention remains controversial; however, dysfunction lasting more than 3 to 4 months necessitates investigation and possible surgical exploration with neurolysis, transposition, repair, and/or reconstruction with nerve grafts or nerve transfers.

Fracture
- Humerus (mid/distal)

Radial n.
Non-operative if suspect compression

Exploration if suspect laceration
- Supracondylar

PIN

Neurolysis, possible grafting if no recovery after 3 months

Med epicondylar

Radius (prox)

Median n

Neurolysis, possible grafting if no

Radial n.

recovery after 3 months

Radius (mid/distal)

Median n.

Carpal tunnel release for neuropathy

Tibia (prox)

Peroneal n.

Peroneal n. release at fibular head for foot drop

Tibia (distal)

Tibial n.

Tarsal tunnel release for neuropathy

Injury Pattern	Nerve(s) Injured	Treatment

Table 5-3. Typical nerve injury associated with fractures

Shoulder

Axillary

Nerve reconstruction with grafts or

Musculocutaneous

nerve transfer of Ax or MC

Elbow

Ulnar n.

Ulnar nerve transposition, neurolysis of AIN,

AIN
reconstruction with nerve grafts or transfers
Hip
Sciatic n.
Neurolysis, possible graft repair, lower

Tibial n.
extremity nerve transfers

Peroneal n.

Knee
Peroneal n.
Peroneal n release at fibular head, tarsal tunnel

Tibial n.
release, possible graft repair or nerve transfer

Joint affected	Nerve(s) Injured	Treatment if conservative treatment fails

Table 5-4. Typical nerve injury associated with dislocations

For patients with chronic nerve injuries extending beyond 2 years, options for primary repair can be limited. Muscle fibrosis and motor end plate degeneration make attempts at primary muscle reinnervation unsuccessful. Instead, in these situations, tendon transfers can provide improvement in function. In comparison, nerve grafting for sensory nerve recovery is not time dependant and can be performed at any time following trauma.

Dislocations and their associated nerve injuries are listed in Table 5-4 . Ulnar nerve injury at the cubital tunnel and/or anterior interosseous nerve injuries are common in elbow dislocations (26 ,27). Shoulder dislocations have been implicated in upper

plexus nerve injuries with the axillary nerve being the most vulnerable to injury (28). In our experience with brachial plexus injuries, a second separate injury at the level of the quadrangular space is also often noted; this â€œdouble crush phenomenonâ€• (two separate nerve injuries occurring along the length of a major peripheral nerve) may account for the poor recovery that is cited in the literature with regard to axillary nerve recovery in these injury patterns (29). The double crush syndrome was originally described in the upper extremity to explain nerve problems resulting from a combination of distal nerve compression at the wrist or elbow in conjunction with proximal cervical root or thoracic outlet pathology (30). Nerves that are traumatized proximally may also become more susceptible to compression at distal sites (31).

In the lower extremity, between 10% and 25% of cases of acetabular fracture and traumatic posterior hip dislocation are associated with sciatic nerve injuries (32 ,33). Tethering of the sciatic nerve at the sciatic notch may exaggerate the traction effect on the nerve in the buttock, thus causing a stretch injury proximally and then a distal compression at the fibular head where the nerve is again tethered. The common peroneal portion of the sciatic nerve is more vulnerable because of its anatomical position and internal architecture (34).

Nerves that are not tethered, but which reside in tight fibro-osseous tunnels (such as the ulnar nerve in the cubital tunnel, the median nerve in the carpal tunnel, and the posterior tibial nerve in the tarsal tunnel), may be affected by a more proximal injury due to tissue edema. For example, the incidence of acute transient median nerve compression syndrome after distal radius fracture is estimated between 12% and 17%, and it occurs regardless of fracture type, the amount of initial displacement, the adequacy of reduction, or the method of operative treatment (35 ,36). In the majority of cases, the neuropathy resolve spontaneously; however, it certain cases, the symptoms persist after the fracture has healed. This leads to the recommendation of performing prophylactic carpal tunnel decompression whenever there is a distal radius fracture. More recently, the consensus has reversed as there is no advantage to the prophylaxis; in fact, evidence suggests there is increased morbidity to acutely decompress the carpal tunnel in the face of mild median nerve compression (37 ,38). In contrast, acute severe carpal tunnel with progressive symptoms despite preliminary fracture reduction *is* an indication for surgical release at time of distal radius fixation. Similar conditions are seen in

the lower extremity, such as acute foot drop after knee reconstruction for ligament injury or tarsal tunnel with distal tibia or maleolar fractures (39). In general, if a patient has loss of median nerve sensation following forearm fracture or injury, a carpal tunnel release is universally recommended. By contrast, a patient with foot drop following a knee injury is treated expectantly often for several months. We believe there should be *no* distinction between the upper and lower extremities and recommend peroneal nerve release at the fibular head in the case of acute foot drop, just as we would perform a carpal tunnel release for acute compression of the median nerve at the wrist.

Preoperative Assessment

Emergency evaluation of a nerve injury should include a thorough motor and sensory exam. Grossly, the muscle function of the nerve in question needs to be evaluated. For the hand, two point static and moving discrimination can be performed to determine the decrease of sensation. Alternatively, the quick and easy â€œten testâ€• sensory exam uses the patient's own subjective perception to moving light touch in order to elicit differences in sensation (40). For example, to test for a median nerve injury, both the injured and uninjured index fingers of the patient are touched at the same time over corresponding areas of each finger and the patient is asked if the subjective sensation is the same or different. This technique is particularly useful in children, is extremely sensitive, and requires no instrumentation.

In patients who present with acute motor deficits or palsy, determining if the nerve will recover spontaneously can often be confusing. The mechanism of the injury can often assist in the initial evaluation. Any sharp penetrating injury with no clinical evidence of recovery should be explored. Optimal timing is between 2 to 14 days, as long as the patient is surgically stable. Occasionally, an MRI or CT scan may be useful if there is evidence of a neuroma or neural disruption; however, any injury where there is a high index of suspicion of nerve transection should be explored and repaired. The advantage of repair within the first 2 weeks of injury is that the nerve ends have not retracted and primary repair is often possible.

Closed traction injuries and closed fractures with palsy are the most difficult to

assess. Waiting 3 to 4 months for evidence of spontaneous recovery is the gold standard. EMGs at 4 months are advised when no clinical recovery is evident. If there is no evidence of reinnervation occurring (MUPs), surgical exploration and reconstruction will be necessary. Figure 5-2 shows the algorithm for sequence of evaluation and repair. If an initial EMG shows some recovery, a follow up EMG is done 4 to 6 weeks later and correlated with clinical evidence of recovery. Once nerve recovery stops progressing as expected, surgical intervention can be planned based on the EMG results and the clinical exam without further delay.

We always prefer to be involved with the care of the patient early and follow the clinical exam and any additional testing, such as electrodiagnostic testing, ultrasound, MRI, and CT scan in order to minimize delay in recognizing a nerve deficit that will not resolve spontaneously. It is important to be aware that the injury may not be limited to only the site of the trauma, but may affect distal functioning muscle groups and may be impinged at distal sites or proximal sites. The more complex injuries will usually require a team approach, both for the operative reconstruction and for the postoperative management. A pain specialist is paramount to manage the associated chronic pain these patients often have for the more complex injuries. Physical therapy and occupational therapy during the rehabilitative period is essential to prevent joint contracture, to fabricate and adjust protective splints, and to assist in motor and sensory re-education as the recovery process is underway.

Operative Management

For acutely transected nerves primary repair remains the gold standard. Adequate resection of the nerve ends beyond the zone of nerve injury is essential to ensure healing with or without limited intraneural scaring. Large associated soft tissue wounds with varying levels of tissue destruction, like degloving injuries and crush injuries, with comminuted bony fractures will need serial debridement prior to ultimate reconstruction. In such cases primary nerve repair should be delayed until one can adequately assess the complete nerve defect. If the nerve cannot be approximated after debridement, a graft will be necessary. Nerve grafts may be used as long as the wound is clean. If the wound is significantly contaminated it is preferable to tag the ends of the nerves and return to graft when all devitalized tissue has been debrided. We do not recommend acutely repairing or grafting nerve

injuries in contaminated wound beds or in crush injuries, where multiple debridements will be required, as postoperative infection, ongoing tissue necrosis, and progressive soft tissue ischemia will all contribute to poor return of function following nerve repair.

Formal intraoperative nerve stimulation is helpful in complex closed traction injuries, especially of the brachial plexus. Proximal injuries of mixed nerves result in complete or partial deficits to several nerves, as opposed to the discrete single nerve injury pattern of distal peripheral nerve injuries. In most cases, an injury distal to the shoulder level will result in one or two discrete nerve injuries and intraoperative electrical stimulation is not necessary. In the case where several segments along the same nerve are injured, or the brachial plexus is injured, electrical stimulation is useful in guiding

intra-operative decision making. Electrical stimulation allows the surgeon to determine which fascicles within the nerve are functional and which ones are not. We limit use of formal intraoperative electrical stimulation to three major situations: (a) in cases where the preoperative EMG results were equivocal; (b) in cases where additional recovery may have occurred between the time of the previous study and the operation; and (c) in situations where the intraoperative mapping of the nerve function can affect the surgical plan. An example of a situation where the intraoperative plan may change is in the case of an incomplete neuroma of the median nerve at the wrist. If the median motor fascicle is intact, the surgeon may try to preserve those fibers and graft the sensory fibers that are injured. Alternatively, if the neuroma is nearly circumferential and the motor fibers poorly conduct to the thenar muscles, the surgeon may opt to completely resect the neuroma and primarily repair or graft the nerve defect.

A handheld 2 mA nerve stimulator will not provide as much information about nerve function as a formal intraoperative electrical stimulation, but is useful in many situations. In the first 72 hours after an acute transaction, the distal half of a motor nerve will contract when stimulated, facilitating the matching of nerves when several are cut, or when the nerve is cut close to the muscle and has begun to branch. The small motor fibers would be otherwise difficult to identify because of their small diameter, which can sometimes be as small as 0.5 to 2 mm in diameter. We find the handheld device a useful aid when performing nerve transfers; it allows

us to pick out grossly motor nerve fibers that were uninjured, which will be spliced into nonfunctioning motor nerves to restore animation. For a nerve injury that develops a partial neuroma (as in a 6th degree nerve injury), a nerve stimulator can help determine which nerve fascicles to preserve while resecting the damaged tissue, as long as the fascicles in question are mixed nerves. For injuries to purely or mostly sensory fibers, a handheld nerve stimulator is not useful and formal intraoperative nerve stimulation is the only means to evaluate sensory conduction.

Common Clinical Examples

Primary Repair for Nerve Transection

A common soft tissue injury that orthopaedic surgeons may encounter and easily repair is the laceration of a major and minor peripheral nerve as a result of a sharp penetrating injury. In the mid to distal forearm and wrist, the median nerve is fairly superficial, and thus easily injured from a deep laceration with glass or metal. The median nerve can be found deep to the palmaris tendon in patients that have a palmaris longus tendon, and just ulnar to the flexor carpi radialis tendon. Knowledge of the anatomy as well as the topography of the nerve is important in order to align the motor branches when repairing the nerve. Primary repair is the procedure of choice whenever possible and usually can be performed within the first 2 weeks after injury. Repairs performed later than this often need a nerve graft as the nerve retracts with time. The repair should be performed with as little tension as possible at the repair site in order to minimize scarring that occurs within the nerve. A small amount of tension at the repair site is acceptable and has been shown in some animal studies to stimulate neurotropic growth factors and improve healing (41). A useful tool in determining excessive tension is to bring in proximity the two ends of a nerve with a single 8-0 nylon suture. If the freshened nerve ends reach without the suture pulling through the epineurium, the tension is not excessive.

Mobilizing the nerve is crucial. Often sufficient mobilization of the nerve will eliminate the need for a graft. Because of the inherent springiness of nerve tissue, when a nerve is transected it will recoil, leaving a gap. A nerve can stretch 10% to 15% without compromising its inherent blood supply (42). If, when bringing the nerve ends together, there is still a 5 mm gap between the nerve ends, mobilizing 5

cm of nerve (usually 2.5 cm on either side, but not always) will allow the ends to come together. If a 1 cm nerve gap is present when the two freshened ends of nerve are laid in proximity, 10 cm of total nerve needs to be mobilized in order to bring the two halves together with â€œminimalâ€• tension. This does not mean that any gap should be closed primarily at all costs. If after reasonable mobilization there is a persistent gap, a nerve graft is preferable to forcing a primary repair. It is important to note that if a nerve graft is necessary, the nerve ends should be trimmed generously so that the repair will be outside the zone of injury. When a nerve graft is needed, there is no reason to limit the debridement of the remaining nerve.

External markers such as the native vessel of the nerve on the surface is used along with the different fascicular size and grouping to align the nerve (Fig. 5-4). In the past, it was believed that a

perineural repair of grouped fascicles was better than an epineural repair; however, recent studies show that either technique is equally effective (43). We prefer to use a circumferential epineural repair with 9-0 nylon interrupted sutures to minimize the fibrosis seen with perineural repairs (Fig. 5-4B). Care is taken to prevent overlap (Fig. 5-5); we prefer to have the fascicle slightly retracted with respect to the epineurium, rather than bulging out the side or overlapping internally. At the conclusion of the repair, the area under question is put through the full range of movement to verify that the repair will hold up under gentile protected motion postoperatively. Any repair that withstands gentle range of motion is considered a â€œtension-freeâ€• repair, although in fact it may be still under some mild yet acceptable tension. In larger nerves, such as the sciatic and femoral nerves, or in nerves with significant inflammation in the epineurium, we use 8-0 nylon sutures. In contrast, for cutaneous nerves and digital nerves, 10-0 nylon works best. The least number of epineural sutures to approximate and keep the repair intact through a full range of motion is used. After neurorrhaphy, the area is immobilized for 1 to 2 weeks; however, protected active range of motion is allowed and should be initiated immediately. One exception is in flexor tendon lacerations with associated digital nerve injuries. Both are repaired and the patient begins early protected range of motion under the direction of an occupational therapist. While sensibility is somewhat decreased when compared to patients who

are immobilized for 3 to 4 weeks, there is no statistical difference of sensibility at 1 year, and the advantages of early range of motion for the rehabilitation of the tendon injury outweighs the negligible delay in sensory return (44).

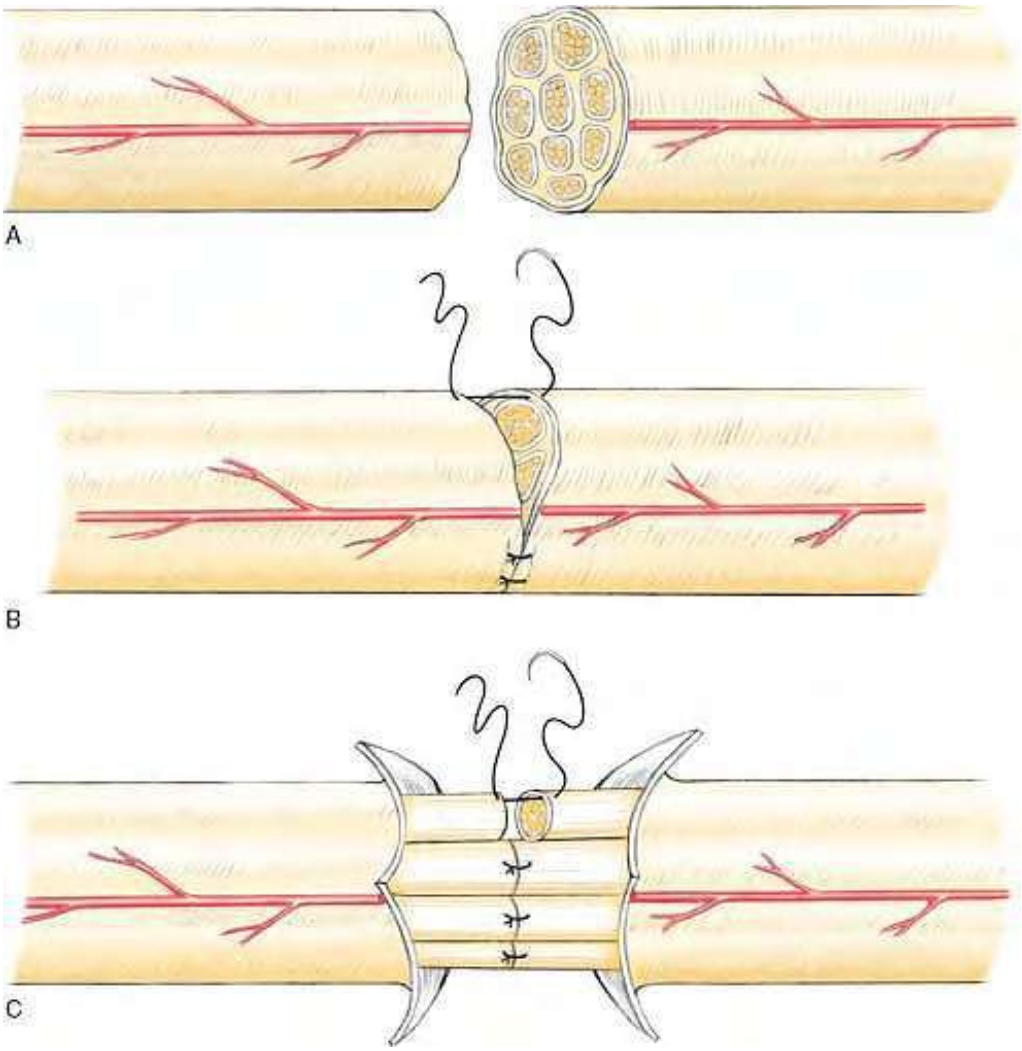

FIGURE 5-4 The two nerve ends are aligned using external markers such as the artery and the grouped fascicle pattern (A) , an epineural (B) or perineural (C) repair may be used, although we prefer the former.

Repair of Nerve Gap with Nerve Graft

While a primary repair is preferred, two neurorrhaphy sites under favorable conditions are better than a single neurorrhaphy under unfavorable conditions (45). In contaminated wounds, such as crush injuries from motor vehicle accidents, the

wound bed needs to be cleaned before any reconstruction of the resulting deformities can be performed. If a large defect to both the nerve and the soft tissue is present, nerve repair with grafts in addition to soft tissue coverage will be needed. In

areas where local muscle is unavailable, a free tissue transfer may be necessary to cover both bone and nerve repair.

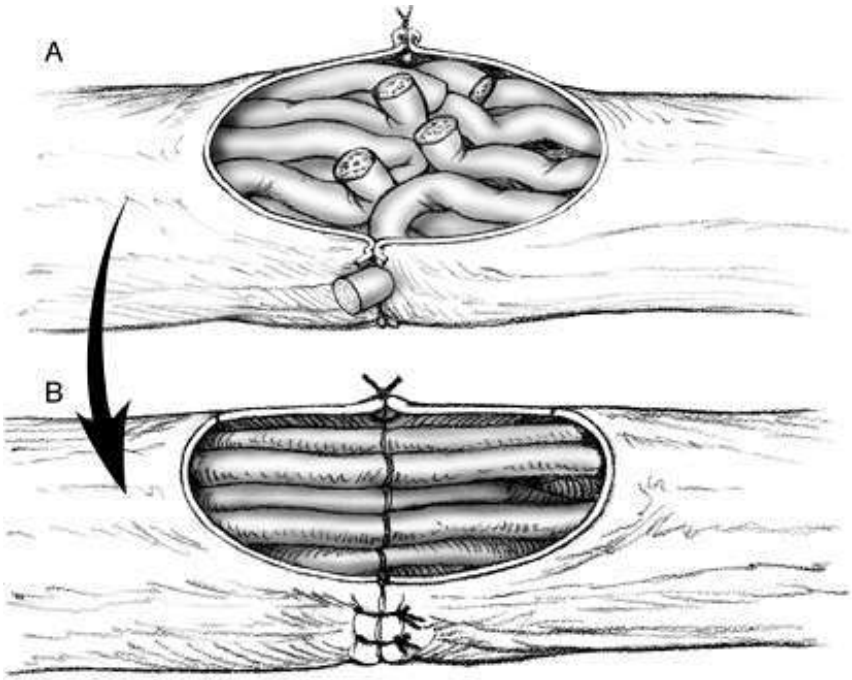

FIGURE 5-5 When aligning the nerve, make sure to trim back the fascicles so that there is no overlap when the epineurium is sutured closed (B) . It is preferable to have a small gap internally than to have fibers overlap and potentially â€œescapeâ€• from the edge or â€œfail to findâ€• the distal half (A) . (By permission of Mayo Foundation for Medical Education and Research. All rights reserved.)

A common situation which may require a nerve graft can be seen in cases of severe knee dislocation with damage to the tibial-peroneal nerve trunk. The stretch injury, if severe enough, will cause internal damage of the nerve. While EMG studies may confirm the suspicion of a third or fourth degree injury, the history and clinical findings are usually enough to warrant surgical exploration with the plan to

release the nerve at the very least from the surrounding scar tissue and compression points, and most likely excise the neuroma and graft the resulting nerve defect. If the neuroma is small, a primary repair may be possible; however, the resultant gap following neuroma excision is often not amenable to primary repair as mobilization of the sciatic nerve proximally and the tibial and peroneal nerves distally are limited by nerve branches to the underlying muscles. The resultant 2 to 3 cm gap will need to be repaired with a nerve graft. A sural nerve is the most often used donor nerve. In the case of a tibial-peroneal nerve injury described above, the nerve would be cut in half or thirds and cabled. Once again, 9-0 nylon interrupted sutures are used to repair both sites. The native artery location on the external portion of the nerve as well as the relative size of the fascicles is used to help line up the fascicles. We still prefer to do an epineural repair of the nerve to nerve graft, rather than a perineural repair of the actual fascicles.

Nerve grafts less than 10 cm long work well in vascularized wound beds. The longer the graft, the less optimal the recovery. Some people have suggested using vascularized nerve grafts when the distance is greater than 20 cm (46). However, in general, small caliber nerve grafts, such as sural and medial antibrachial cutaneous nerves, do not need to be vascularized. Large caliber nerves, such as ulnar, do need to be vascularized or the central portion will necrose and then scar, limiting its effectiveness as a graft.

Obtaining a Nerve Graft

The nerve graft serves as a guide for the proximal axon as it regrows toward the distal stump. The sural nerve is by far the most commonly used donor nerve, although other suitable donor nerves include the lateral and medial antibrachial cutaneous nerve (47 ,48) and the terminal portion of the anterior interosseous nerve (AIN) that innervates the pronator quadratus (49). The sural nerve is often chosen as a nerve donor because of its size, length, and relatively minimal donor site and minimal morbidity. Typically, harvest of the sural nerve will result in numbness over the posterior aspect of the leg and the lateral aspect of the foot and malleolus.

With the patient either prone or supine with the leg frog-legged, the sural nerve is

identified lateral to the Achilles tendon. The easiest method for harvest of the sural nerve is to use a posterior midline incision. The closure is slightly lengthier and the resulting scar may be cosmetically undesirable. The alternative is to use a stepwise technique, which requires between 4 and 6 separate horizontal

incisions along the back of the calf, each approximately 1 cm in length. The resulting scar is cosmetically more favorable and the technique adds little operative time. An endoscopic or a nerve harvester may be used to limit scars. The trick with using the nerve harvester is that contributions from the lateral and anterior sural nerves, which are cutaneous branches from the peroneal and tibial nerves, may be difficult to incorporate into the sural nerve proper and shorten the overall length that is harvested. Approximately 25 to 30 cm of graft may be harvested from a typical adult patient. The next most often used graft is the medial anti-brachial nerve. It is easily found coursing along side the bacillic vein in the upper arm. Up to 25 cm of nerve graft may be obtained.

While nerve grafts are the gold standard for motor nerve gap repairs, the disadvantage is the limited number of donor nerves available. This has led to the development of new techniques for bridging nerve gaps. For nerve gaps less than or equal to 3 cm on small caliber sensory nerves (50), the nerve may be repaired with a conduit, which may be autologous, such as a vein, or synthetic, such as one of the commercially available nerve tubes (Fig. 5-6A). The end of each nerve is trimmed until healthy fascicles are seen. The end of the nerve is inserted into the tube for approximately 2 to 4 mm and the epineurium is sutured to the end of the conduit with 9-0 nylon using a simple suture technique, a horizontal matrass suture technique, or a half buried matrass suture technique. Clinically, for nerve defects longer than 3 cm, in mixed nerves, and in pure motor nerves the recommendation is still a nerve graft. Alternatively, we have used nerve conduits as a covering over nerve repairs that required several cabled grafts into one larger nerve to serve as a temporary protective covering instead of wrapping a spatulated strip of vein (Fig. 5-6Bâ€“D).

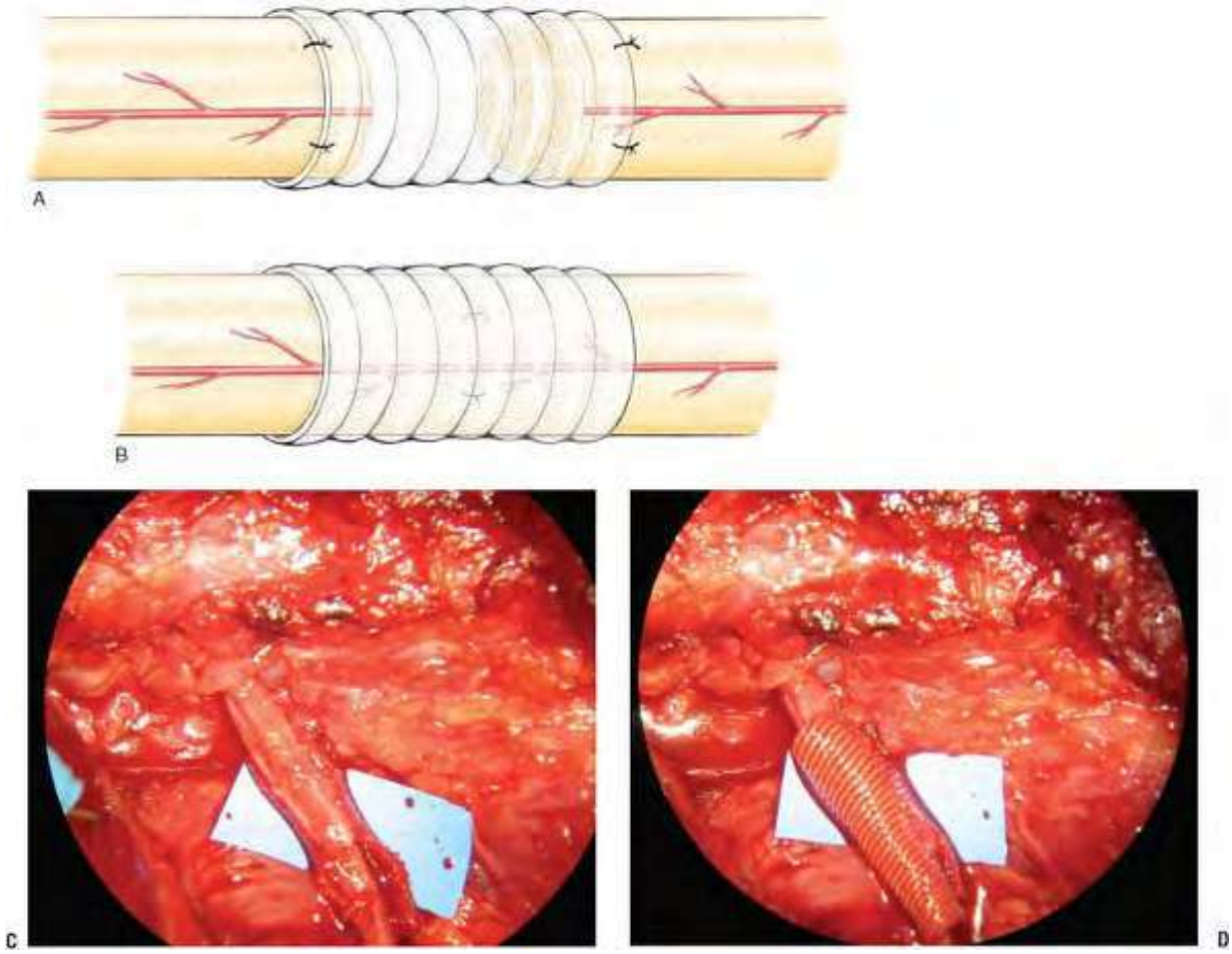

FIGURE 5-6 Neural tube may be used to bridge a gap or as a protective covering after nerve repair. When used to bridge a nerve gap (A) , the epineurium is sutured to the edge of the neural tube. Bâ€“D: When used to cover a neurorrhaphy site, the tube may be cut longitudinally and wrapped around the neurorrhaphy site. A suture is placed around the outside of the conduit to help keep it in place.

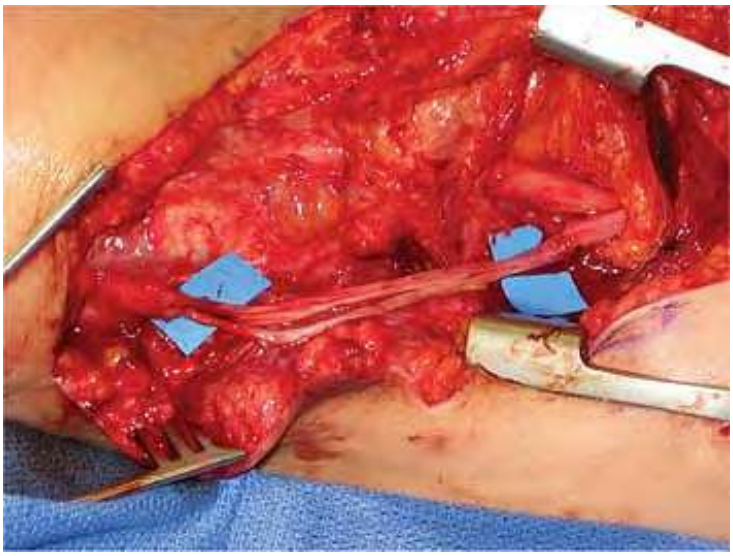

FIGURE 5-7 Three cabled sural nerves, each 2 mm in diameter, were used to bridge the defect in the radial nerve. The diameter of the radial nerve is approximately 6 mm in this location.

Neuroma of Radial Nerve with Humeral Fracture

Considerable controversy remains as to the need for early exploration in closed fracture dislocation injuries with acute nerve palsy. In a recent meta-analysis review of the literature over the last 40 years regarding radial nerve palsy after fracture of the shaft of the humerus, the prevalence of fracture associated nerve palsy was about 12% (51). Over 70% of these compressions resolve spontaneously, therefore the current recommendation is for conservative therapy with closed fractures. If after 3 months there is no clinical recovery, using high-resolution ultrasound to evaluate the injured nerve is less invasive than electrodiagnostic testing and has been used in this specific situation for effectively detecting a neuroma in the presence of a fracture (26 ,52). However, at 3 to 4 months, if spontaneous recovery is to be anticipated, there should be clinical evidence of reinnervation or electrical evidence of regeneration (motor unit potentials). If radial nerve regeneration stops at the level of the Arcade of Frohse, then decompression of the posterior interosseous nerve is indicated. In addition, open fractures, oblique fractures, and fractures that can compress or transect the radial nerve between the fracture fragments should be explored early (53).

If instead of recovery, all preoperative studies indicate a neuroma of the radial nerve in the distal arm, the nerve should be explored with the plan to resect and graft the defect. Rarely can the nerve be mobilized enough to be repaired primarily. Because the nerve and scar surrounding it are usually densely adherent and are similar in color, entering the scarred area directly is not advised. The uninjured proximal and distal ends of the nerve are exposed first. In the case of a partial palsy or radial nerve weakness, a sixth degree nerve injury is most likely. A handheld nerve stimulator can be used to stimulate fibers proximal to the injury. Those fascicles that result in movement of muscles distal to the injury are intact and these intact fibers are marked and protected from further dissection. The injured fibers are resected until normal fascicles are noted from each end. A nerve graft is used similar to before. If more than one third of the radial nerve has been resected, two or more cabled sural nerve grafts may be needed to bridge the gap (Fig. 5-7). The larger sized fascicles should be lined up and grafted directly whenever possible. 9-0 or 8-0 nylon interrupted sutures are used to repair both sites with an epineural or perineurial repair.

Postoperative Management

After any typical nerve repair, whether it is done primarily, with a nerve graft or a synthetic tube, or a primary neurorrhaphy using a nerve transfer, the affected area is splinted or placed in a sling for 2 weeks to minimize movement. The only exception is in digital nerve repairs in conjunction with a tendon repair as noted before. We begin early passive range of motion to prevent tendon adhesion and leave the patient in a dorsal blocking splint to prevent hyperextension and additional strain to the repair site.

In both motor and sensory nerve repairs, an advancing Tinel sign or â€œtinglingâ€• sign seen during the regeneration phase is followed until the patient has restoration of function or sensation. Even before there is clinical evidence of recovery, we begin desensitization of the affected area to assist in

sensory reeducation and to prevent hypersensitivity. Formal objective testing, such as Simien-Weinstein testing, and sensory grading is useful for objective data collection, but is not necessarily useful in clinical practice. Motor reeducation begins as soon as the patient has some clinical evidence of muscle contraction.

Occasionally, EMGs are used postoperatively to follow functional recovery. This is useful in cases where a primary repair may not show clinical evidence of recovery and a second stage reconstruction may be planned. In those cases, EMGs are invaluable in answering the question of whether to wait longer in the case of recovery that may not manifest in actual clinical movement yet, versus proceeding with the next phase of reconstruction, if no evidence of recovery is present.

Conclusion

Nerve injuries alone are fairly complex and when the soft tissue and bony structures in the vicinity are injured as well, the complexity increases. Bony trauma with or without soft tissue injury indicates a greater force was expended to cause the injury, thus the associated nerve injuries can be severe and at multiple levels. Avulsion and stretch injuries, crush injuries, and penetrating injuries can occur in combination or alone. The mechanism of injury varies greatly, thus no single solution to all nerve injuries exists. The deficits are often not readily apparent because of concomitant injuries or because the patient can not reliably participate in the examination. The reconstruction of both the nerves and the soft tissue will ultimately be tailored to the specific situation. Any study that can aid in the decision process, such as radiology exams and electrical diagnostic testing, should be ordered in a timely fashion. Additional assistance from the following may be needed: (a) a plastic or orthopaedic surgeon knowledgeable about peripheral nerve surgery and soft tissue coverage if the defect is more complex than initially anticipated or if extensive wound coverage issues are beyond the scope of the treating physician; (b) a pain specialist for management of chronic narcotic and nonnarcotic pain medication and nonoperative therapies such as nerve blocks and nerve stimulators for chronic pain; (c) a psychiatrist or psychologist for depression if warranted in the complex peripheral nerve injuries; and (d) occupational and physical therapists for motor and sensory reeducation, splinting, and joint mobility maintenance.

References

1. Seddon HJ. Three types of nerve injury. *Brain.* 1943;66:237.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 6

Management of Vascular Injuries Following Soft Tissue and Bony Trauma

Gustavo S. Oderich

Timothy M. Sullivan

Accidental injuries affect 2.6 million people in the United States each year. Mechanisms of injury include penetrating, blunt, and iatrogenic trauma. It is estimated that vascular trauma affects 0.2% to 4% of all injured patients. Therefore, approximately 20 to 100,000 patients will sustain a vascular injury. Over 80% of these vascular injuries will be located in extremities. Ninety percent of all vascular injuries are associated with penetrating trauma, of which 70% are due to gunshot injuries, 20% are due to stab wounds, and 10% are due to blunt trauma. Combined vascular and orthopaedic trauma is relatively uncommon, accounting for less than 1% of all cases of traumatic injuries. Vascular injuries can occur because of the superficial location of vessels, their proximity to bones, and their relatively fixed position across joints. Patients sustaining combined injuries are exposed to substantially increased risk of amputation and limb dysfunction. This chapter focuses on the indications, contraindications, preoperative planning, operative approach, and results of repair of the most common combined vascular and orthopaedic injuries.

Indications/Contraindications

Primary repair of arterial and venous injuries is indicated to control bleeding and/or to relieve limb or organ ischemia. Life-threatening injuries should be recognized early during the resuscitation phase and prioritized before proceeding with vascular and orthopaedic repair. Arterial injuries affecting only one of the tibioperoneal or forearm arteries may be treated conservatively provided that there is no evidence of distal ischemia. Nonocclusive arterial injuries (e.g., small intimal flaps or dissections) incidentally found on imaging studies can be safely observed. The benign natural history of these injuries is well documented in several large series; operative repair is required in less than 1% of patients.

Extremity trauma with complex soft tissue, vascular, and skeletal injuries poses one of the most difficult management problems. Patients should be evaluated by a multidisciplinary team. Coordinated interaction of various specialists including a vascular, orthopaedic, and plastic surgeon is of paramount importance to optimize outcome. Injuries of the head, chest, and abdomen may require additional neurosurgical and general trauma consultation. Every effort is made to balance the potential success of an arterial reconstruction and correction of the orthopaedic injury with the overall clinical status of the patient and the potential for complete functional recovery. In general, once life-threatening injuries are stabilized, treatment priorities are as follows:

- Control of bleeding and restoration of arterial inflow.
- Fracture reduction and stabilization.
- Soft tissue coverage.

Vascular repair should be prioritized over definitive orthopaedic repair. Any delay in vascular reconstruction is a gamble and may risk the only opportunity for limb salvage. Patients with stable fractures or dislocations in which minimal manipulation and length discrepancy is anticipated should be treated with immediate definitive arterial revascularization. However, patients with severely comminuted fractures and dislocations, segmental bone loss causing limb discrepancy, or severe soft tissue disruption and contamination should be treated initially with temporary intra-arterial shunts. In these cases, it is wise to delay the

definitive vascular reconstruction until wide debridement and initial skeletal repair are accomplished.

Contra-indications for vascular repair include presence of other life-threatening injuries requiring immediate attention or causing hemodynamic instability. Several predictive factors should be taken into consideration when deciding to perform revascularization versus primary amputation (Table 6-1). Overall, approximately 10% to 20% of patients with complex extremity injuries have nonsalvageable limbs and require primary amputation. Assessment of these patients should be individualized. Several factors should be taken into consideration, including the overall clinical status, severity of the arterial, neurologic, and orthopaedic trauma, and expected functional recovery. A primary amputation is considered in cases of dysvascular extremity with complex fractures and extensive soft tissue and nerve damage. Patients with major nerve transections (e.g., tibial nerve transection) and open comminuted tibiofibular fractures with arterial injuries (Gustilo III-C) have very poor functional outcome and high amputation rates; these patients are generally treated with primary amputation. Major nerve transections should be confirmed by direct visualization. Other indication for amputation is prolonged ischemia time (>12 hours) with evidence of a cadaveric extremity (e.g., mottled with absence of motor function, arterial, or venous Doppler signals).

Preoperative Planning

Prolonged ischemia time is the most important factor associated with limb dysfunction or amputation. A high index of suspicion coupled with accurate neurovascular examination is necessary for prompt diagnosis of a vascular injury. Accurate diagnosis of vascular injuries is an essential aspect in the preoperative evaluation of patients with complex extremity trauma. Prompt restoration of arterial blood flow within 6 hours from the time of initial extremity injury is the most critical factor that determines limb salvage and function.

Complete history including the mechanism of trauma, associated injuries, medical history, medications, and allergies should be recorded. Extremity vascular trauma is immediately apparent because of external bleeding, hematoma, or obvious limb ischemia. Physical examination includes inspection of the injured limb for open wounds, obvious deformities, and signs of ischemia. Distal ischemia is manifested

by the five Ps: pallor, paresthesia, paralysis, pain, pulselessness, and poikilothermia. A thorough sensory and motor examination and pulse examination should be noted. The presence of hematoma, pulsatile or not, bruit, and thrill must be documented. If distal pulses are diminished or absent, the ankle-brachial index should be determined using a hand-held Doppler device.

Signs of arterial injury have been traditionally classified into â€œhardâ€• and â€œsoftâ€• signs (Table 6-2). These correlate with the presence of a hemodynamically significant arterial lesion. â€œHardâ€• signs include

absence of distal pulses, pulsatile bleeding, expanding hematoma, palpable thrill, and audible bruit. â€œSoftâ€• signs include proximity to the vessel, peripheral nerve deficit, history of moderate hemorrhage, and *diminished* distal pulses.

Table 6-1. Factors Associated with Poor Functional Recovery or Need for Primary Amputation After Combined Vascular and Orthopaedic Extremity Trauma

Transected sciatic or tibial nerve	Multiple comminuted fractures
Transection of two of the three upper extremity nerves	Extensive soft tissue loss
Gustillo III-C orthopaedic injury	Crush injury
Below-knee arterial injury with two of the three arteries injured	Severe contamination
Prolonged limb ischemia greater than 12 hours	Elderly patients or multiple medical comorbidities

Cadaveric limb	Shock or other life-threatening injuries

Table 6-2. Clinical Signs of Vascular Injury

Hard Signs	Soft Signs
Absent distal pulses	Diminished distal pulses or ABI <0.90
Active pulsatile bleeding	Unexplained hypotension or large blood loss at the scene
Expanding hematoma	Small or moderate nonexpanding hematoma
Bruit	Injury in proximity to a major vessel
Thrill	Neurologic injury in proximity to vessel

ABI, ankle-brachial index.

The general recommendation is that patients with hard signs associated with uncomplicated penetrating trauma should undergo immediate operative exploration, without need for arteriography or duplex ultrasound. The indications for arteriography are summarized in Table 6-3. Patients with multiple penetrating injuries should be evaluated with arteriography to determine the exact location and extent of arterial lesions. Arteriography is also advised in patients with combined arterial and orthopaedic trauma, even in the presence of hard signs. While the presence of hard signs predicts major vascular injury in nearly 100% of

patients with uncomplicated penetrating trauma, less than 15% of patients with complex blunt injuries will require vascular repair. The inaccuracy of hard signs in this subgroup is explained by a combination of multiple other factors that ultimately lead to diminished pulses, including fractures, soft tissue disruption, compartment syndrome, and extrinsic arterial compression. Therefore, we generally recommend arteriography in all patients with complex skeletal trauma and hard signs of vascular injury.

We generally prefer a one-shot intraoperative arteriography technique. This avoids the 1 to 3 hour time delay required to obtain a formal arteriography in the angiography suite. An antegrade or retrograde approach in the affected limb is used for most cases. The availability of a portable C-arm and fluoroscopy permits localization of bony landmarks, selective catheterization of arterial branches, and endovascular treatment of arterial lesions. However, because fluoroscopy is usually not readily available in the emergency setting, we usually use ultrasound guidance and a micro-puncture set for arterial access. The target artery (e.g., common femoral artery or brachial artery) is accessed with a micropuncture needle (18 gauge) in an antegrade or retrograde fashion. A 0.018 inch guidewire is advanced to allow placement of a sheath. The sheath can be used for contrast injection. Alternatively, the 0.018 inch wire is exchanged into a 0.035 inch wire and a 4 French sheath is advanced. The arterial inflow proximal to the access site should be manually compressed during contrast injection using digital pressure or a tourniquet. A single hand injection of 30 mL of diluted (50:50) iso-osmolar contrast allows adequate visualization of the area of concern and distal runoff vessels.

Other noninvasive imaging modalities are duplex ultrasound, computed tomography angiography (CTA), and magnetic resonance angiography (MRA). Computed tomography angiography is now available in most centers, is quite expeditious, and permits excellent image of the arterial circulation, soft tissue, and bone. This is particularly useful for planning operative approach in patients with centrally located lesions (e.g, aorta, subclavian, or iliac arteries). However, limitations are the relatively large contrast load (150 mL) in a patient who may potentially require additional angiography.

We recommend duplex arterial ultrasound in patients with â€œsoftâ€• signs of vascular injury, including those with an ankle-brachial index of less than 0.90.

with normal pulses and no other sign of arterial injury. Physical examination excludes significant arterial injuries as reliably as arteriography or surgical exploration. This is also true for patients with posterior knee dislocations, in whom arteriography used to be obtained routinely in the 1980s. Results of contemporary series show that in the absence of hard signs, less than 5% of patients had abnormalities on the arteriography but none required operative treatment or had amputation. On the other hand, the incidence of significant vascular injuries is 70% in the presence of hard signs. In patients with normal pulse examination, nonocclusive lesions include small intimal flaps, dissections, contusions, or small pseudoaneurysms; these lesions should be treated with antiplatelet therapy only.

Table 6-3. Indications for Preoperative or Intraoperative Diagnostic Arteriography in Patients with Combined Extremity Trauma

Multiple penetrating injuries	Trajectory parallel to artery
Unclear location or extent of arterial injuries	Underlying peripheral arterial disease
Extensive soft tissue injury	
Fracture or dislocations with hard signs of vascular injury	

Surgery

Patient Positioning and General Approach

All vascular injuries can be accessed using the supine anatomic position (Fig. 6-1). Although prone position and a posterior approach may be used for isolated popliteal

artery injuries which do not extend into the superficial femoral or tibioperoneal arteries, this approach limits access to the great saphenous vein if harvesting is necessary and may require excessive manipulation of the fractured limb. A generous sterile field should be prepared and draped to allow adequate exposure, proximal and distal control, and options of extra-anatomic reconstruction if indicated. The wise surgeon should always consider the â€œworst case scenarioâ€• (e.g., axillofemoral graft). A noninjured lower extremity should be prepped circumferentially for possible vein harvesting. The ideal conduit is an autologous great or small saphenous vein. Preoperative intravenous antibiotics (e.g., a first-generation cephalosporin) should be administered.

Technique

Repair of vascular injuries can be one of the most challenging aspects of trauma management. Some basic principles of vascular repair are applicable to all vascular injuries (Table 6-4). The operative sequence consists of access, exposure, control, and repair. Initial control of external hemorrhage is usually obtained with simple digital or manual pressure. Use of surgical instruments such as hemostats is not only ineffective but risks iatrogenic injury of adjacent nerves or veins. Manual compression is maintained by a member of the surgical team until definitive proximal and distal control of the injured vessel is obtained. The hematoma should not be entered without first obtaining proximal

and distal control away from the site of injury. Access is gained using the standard exposure techniques described in the following discussion. Vascular clamps should not be applied forcefully and blindly; instead, the artery should be completely dissected, looped with a silastic vessel loop, and clamped under direct vision. One important adjunct is the use of balloon occlusion catheters. This facilitates control in areas of difficult access and avoids excessive dissection. In addition, pressure cuff tourniquets may be used in the extremities to achieve prompt vascular control. Once control is gained, the hematoma is explored with careful attention to avoid injury to adjacent structures. The extent of vascular injury is assessed and clamps are moved closer to the vascular wound. The injured vessel should be debrided and cleaned using tenotomy scissors. The extent of debridement should take into consideration the mechanism of injury, with more extensive debridements for high-

velocity gunshot wounds, and blast or crush injuries. One should avoid excising the adventitia of the vessel while dissecting or debriding the artery. Distal and proximal thrombectomy using a Fogarty balloon thromboembolectomy catheter should be performed even in the presence of relatively good prograde or retrograde bleeding. The size of the catheter is generally a 4 or 5 for the iliacs, 4 for the superficial femoral, 3 or 4 for the deep femoral, 2 or 3 for the tibials, and 2 for the pedal and smaller arteries of the forearm. Passage of thromboembolectomy catheters into the deep femoral, tibials, and smaller arteries should be done gently because of risk of arterial rupture, dissection, or intimal injury due to excessive balloon inflation. At least two â€œcleanâ€• passes should be made before considering the artery free of thrombus. Preoperative and intraoperative systemic heparinization (80 mg/kg bolus) should be used unless there is a contraindication (e.g., head injury). An activated clotting time (ACT) above 250 seconds is considered optimal. Liberal regional heparinization should be used with flushes of heparinized saline proximally and distally to prevent propagation of any thrombus.

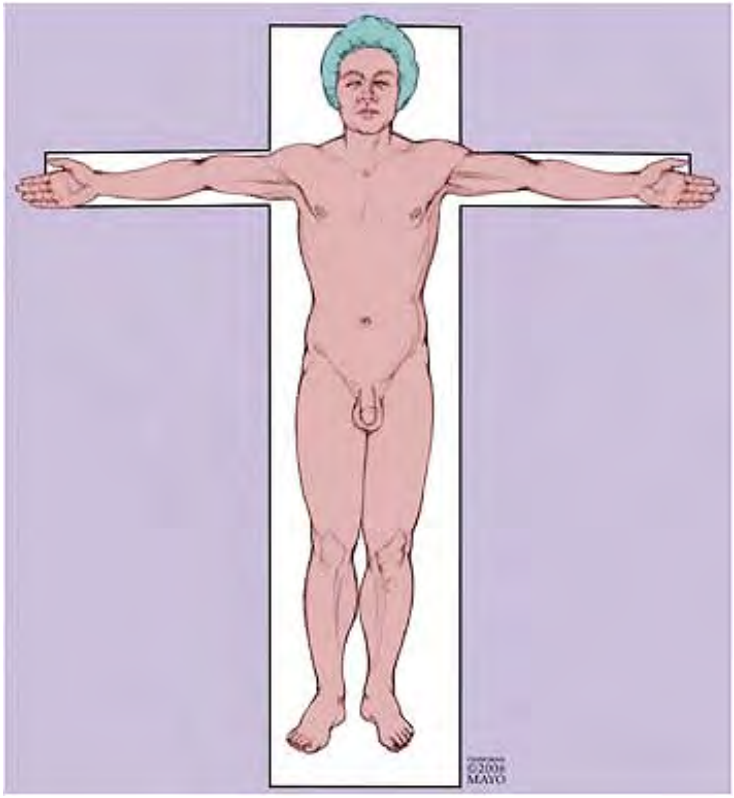

FIGURE 6-1 Patients should be positioned supine with one or both upper extremities abducted. At least one unaffected lower extremity should be circumferentially prepped for vein harvesting if indicated.

Table 6-4. Basic Principles of Repair of Vascular Injuries

Manual pressure for vascular control	Instill regional heparin after two â€œcleanâ€• passes
Access using standard vascular exposure	Perform vascular anastomosis in a tensionless fashion
Ensure proximal and distal control prior to entering a hematoma	Allow prograde and retrograde bleeding prior to completing the anastomosis
Carefully enter the hematoma, avoiding injury to adjacent structures	Complete suture line, remove distal and proximal clamp sequentially
Assess the extent of injury and presence of concomitant venous injury	Assess distal circulation with pulse examination and hand-held Doppler
Use systemic heparinization whenever possible	Obtain completion arteriography
Determine type of vascular repair	Consider need for fasciotomy
Debridement of vessel edges	Ensure adequate soft tissue coverage of vascular reconstruction
Proximal and distal Fogarty catheter thromboembolectomy	

Restoration of extremity perfusion does not always require definitive arterial reconstruction. Another important adjunct is a temporary arterial shunt placed in the proximal and distal arterial ends. This promptly controls hemorrhage and re-establishes inflow. Our preference is to use either a short or long Sundt

intraluminal arterial shunt (Fig. 6-2). The shunt is secured to the artery using silk suture or a Rummel tourniquet. Although systemic heparinization is preferred, this is not an absolute requirement for shunt placement. The main advantage of temporary shunts is the avoidance of damage to a fresh arterial repair during subsequent orthopaedic manipulations in cases of comminuted, unstable skeletal injuries.

The injured vessels are â€œset upâ€• for the reconstruction using stay sutures and clean white towels. The anastomosis should be performed without any tension using optic magnification and small monofilament suture (e.g., Prolene). In general, a 4 or 5-0 monofilament suture is used for the femoral arteries, and 6 or 7-0 for the popliteal, tibials, and smaller arteries. The type of reconstruction varies and should be tailored to both the patient condition and extent of injury (Fig. 6-3). Simple repair entails a lateral arteriorrhaphy or venorrhaphy. One should avoid excessive narrowing of the lumen and use patch angioplasty if there is any question. A segment of saphenous vein or bovine pericardium can be used for vascular patches. Vessels with injuries involving less than 1.5 cm in length can usually be re-approximated using end-to-end anastomosis. Mobilization of the artery or vein may require ligation of multiple side branches to gain enough length to allow a tensionless anastomosis. The vessel ends should be spatulated to prevent anastomotic narrowing. A running anastomosis

is used most cases, but interrupted sutures may be required in smaller vessels or pediatric cases. Before completion of the anastomosis, the proximal and distal clamps are temporarily removed to ensure adequate prograde and retrograde bleeding. Absence of back bleeding indicates thrombus formation or a technical defect and warrants further catheter embolectomy or revision. Approximately 80% of all vascular extremity injuries that occur in association with orthopaedic trauma

can be repaired using one of these three techniques: simple closure, end-to-end anastomosis, or patch angioplasty.

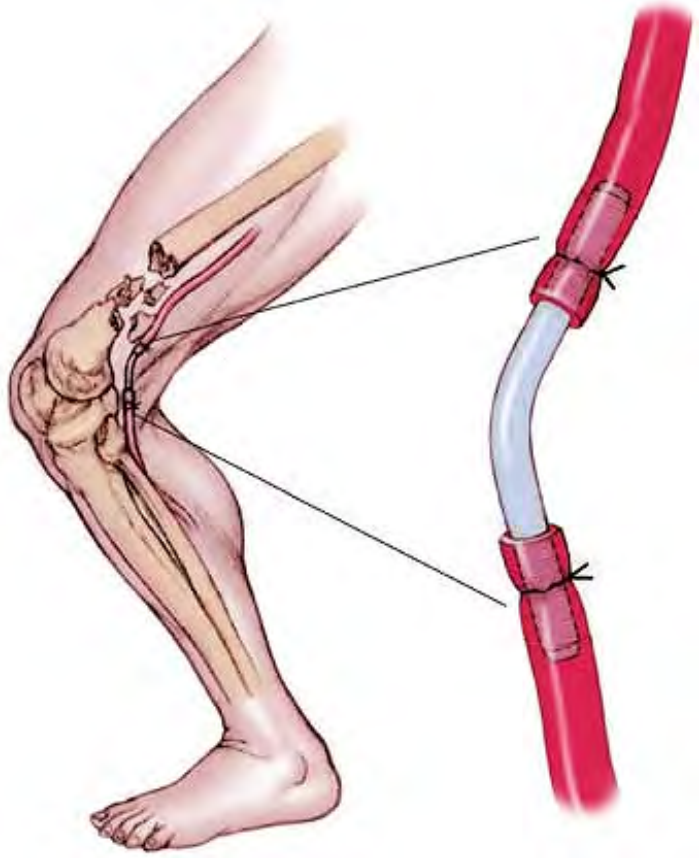

FIGURE 6-2 Intra-arterial Sundt shunt for temporary control and restoration of blood flow.

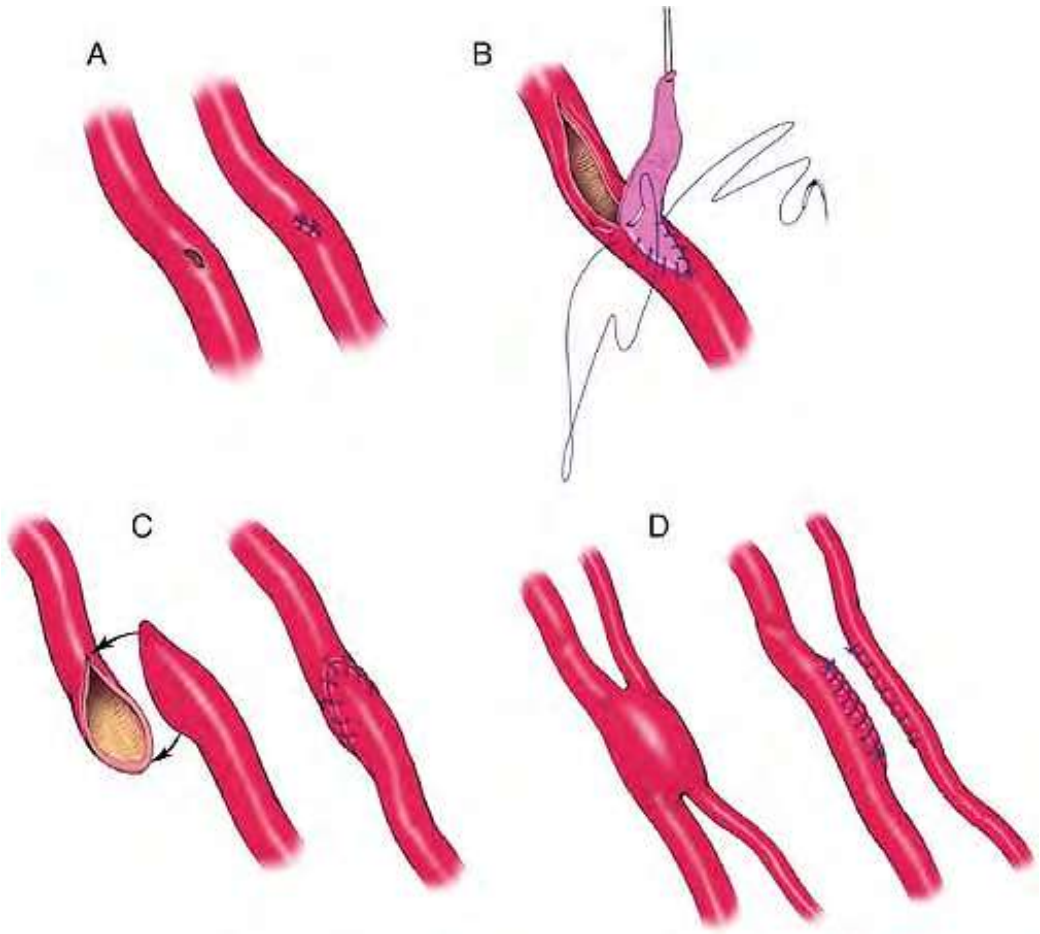

FIGURE 6-3 Types of vascular injuries and repair: simple laceration treated with arteriorrhaphy (A); use of patch angioplasty for longer lesions (B); end-to-end anastomosis (C) and lateral arteriorrhaphy (D).

Insertion of an interposition or bypass graft is required in approximately 20% of cases. Vessels with more extensive injuries of greater than 1.5 cm in length cannot be re-approximated without tension. The preferred conduit in these cases is the great saphenous vein harvested from the contralateral noninjured extremity. The ipsilateral saphenous vein should be avoided due to a high incidence of concomitant deep venous injuries (50%) and postoperative deep venous thrombosis. The great saphenous vein is ideal for vessels 6 mm or smaller. Some size mismatch is acceptable. However, for injuries of larger vessels, a larger conduit should be

selected. Spiral or panel vein grafts, or femoral vein grafts, are acceptable options. For arterial injuries of the aorta and major branches, polytetrafluoroethylene (PTFE) or dacron grafts have been used extensively with excellent results. Obviously, contamination of the surgical field represents a limiting factor in these cases. Every attempt should be made to repair both the arterial and venous injury. Although venous repair has limited patency, improved outflow decreases edema and may affect patency of the arterial repair. The traditional recommendation is that all venous injuries involving the popliteal, common femoral, axillary, and portal veins should be repaired because of poor collateral flow. This recommendation has now been extended to all major named veins.

After the vascular reconstruction is complete, the surgeon should assess the adequacy of the repair for any residual stenosis or kinks. The limb should be examined and distal perfusion documented with pulse and hand-held Doppler examination. The goal in a patient without pre-existing vascular disease is to obtain normal distal pulses at the end of the operation. Completion arteriography is important to detect any technical abnormalities that may cause early thrombosis of the repair, as well as to assess the patency of the distal run-off vessels. Finally, the vascular reconstruction should be covered with viable soft tissue. Occasionally, flaps are required to bring vascularized tissue over the reconstruction.

Specific Injuries: Anatomy and Surgical Exposure

Penetrating injuries can affect any arterial segment. Specific patterns of blunt orthopaedic and arterial trauma are well recognized (Table 6-5). The surgical anatomy and exposure of the most common arterial injuries are outlined in the following paragraphs.

Upper Extremity Injuries

Subclavian Artery

The right subclavian artery originates from the innominate artery, and the left subclavian from the aorta (Fig. 6-4). The subclavian artery is divided into three parts in relation to the anterior scalene muscle: proximal, middle, and distal. The proximal subclavian artery gives off the vertebral, internal thoracic, and thyrocervical trunk. The middle subclavian artery contains the costocervical trunk

and the dorsal scapular artery. The distal subclavian artery has no branches. A supraclavicular incision one fingerbreadth above the clavicle allows excellent exposure of the middle and distal subclavian artery (Fig. 6-5). For exposure of the proximal subclavian artery, median sternotomy is required for right-sided injuries and left anterolateral thoracotomy for left-sided injuries.

Subclavian artery injuries are uncommon and represent less than 5% of all vascular injuries. Blunt injuries may occur in association with clavicular or first rib fracture. Over 50% of patients present with massive bleeding. Subclavian artery occlusion is well tolerated in 85% of patients because of a rich collateral system. Amputation is rarely required. Arterial repair usually requires end-to-end

P.84

P.85

anastomosis or a small interposition graft using autologous vein. Overall, the mortality rate for subclavian artery injuries is 16%.

Table 6-5. Patterns of Combined Orthopaedic and Vascular Trauma

Orthopaedic Injury	Arterial Injury Location
Supracondylar humeral fracture	Brachial artery
Clavicular fracture	Subclavian and axillary artery
Shoulder dislocation	Axillary artery
First rib fracture	Subclavian artery and aorta

Femoral shaft fracture	Superficial femoral artery and above-knee popliteal artery
Posterior knee dislocation	Popliteal artery
Proximal tibiofibular fracture	Popliteal and tibioperoneal arteries

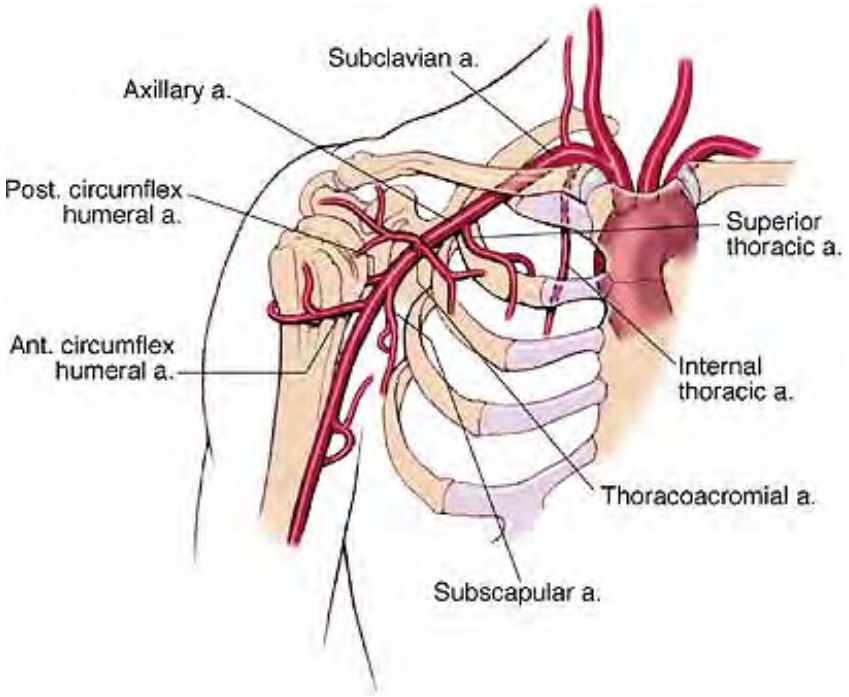

FIGURE 6-4 Surgical anatomy of the subclavian and axillary arteries.

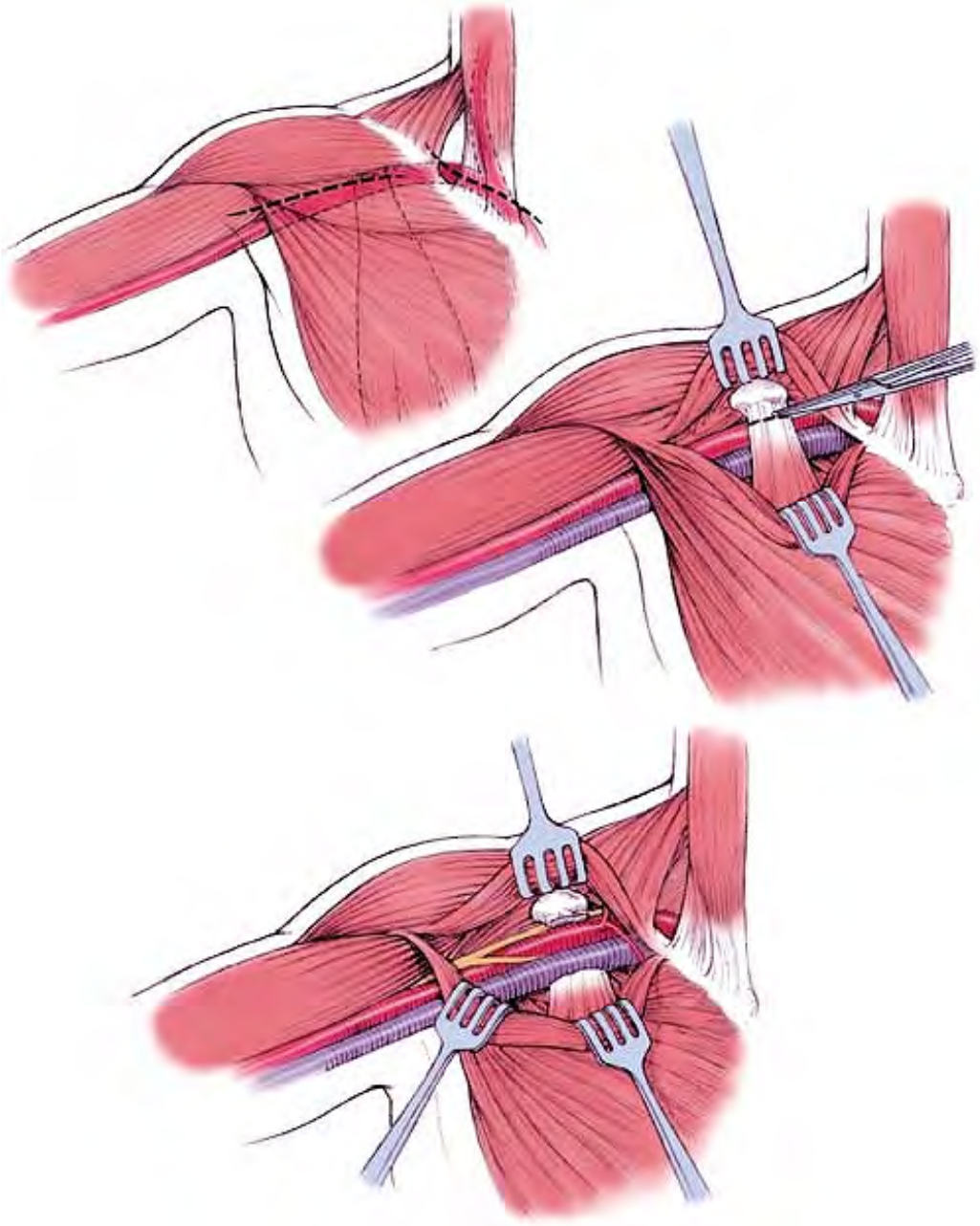

FIGURE 6-5 Preferred surgical approaches for exposure of the subclavian (supracalvicular) and axillary (infraclavicular) arteries.

Axillary Artery

The axillary artery extends from the lateral border of the first rib to the lateral border of the teres major muscle (see Fig. 6-4). The artery is divided into three parts in relation to the pectoralis minor muscle: proximal, middle (beneath), and distal. The first part gives of one branch (supreme thoracic), the second two branches (thoracoacromial and lateral thoracic), and the third three branches (subscapular, anterior and posterior humeral circumflex). The artery lies in close proximity to the axillary vein and to the brachial plexus.

Axillary artery injuries represent 5% to 10% of all arterial injuries. The vast majority (95%) are due to penetrating trauma. Although blunt injuries are rare, these occur in 1% of the patients presenting with either a fracture of the proximal humerus or anterior dislocation of the shoulder. In addition, patients who chronically use crutches may develop stenosis or occlusion from repetitive trauma. Patients with axillary artery injuries often sustain associated nerve trauma.

Exposure of the axillary artery is best achieved using an infraclavicular incision one fingerbreadth bellow the clavicle (see Fig. 6-5). Most injuries can be repaired with end-to-end anastomosis or a small interposition graft. The mortality and amputation rates are exceedingly low for axillary injuries. However, two-thirds of the patients have significant neurologic dysfunction or persistent neuralgia from associated trauma to the brachial plexus.

Brachial Artery

The brachial artery is a continuation of the axillary artery (Fig. 6-6). It starts at the lower edge of the teres major muscle and terminates approximately 2 cm below the antecubital crease, where it bifurcates into the radial and ulnar arteries. A high bifurcation of the brachial artery above the antecubital crease is found in 20% of the population. In the upper and mid-arm, the brachial artery is accompanied by the median nerve laterally and by the ulnar and radial nerves medially. The median nerve crosses anterior to the artery and is located medially at the level of the elbow joint. The brachial artery has three important branches: deep brachial artery, and superior and inferior ulnar collateral arteries.

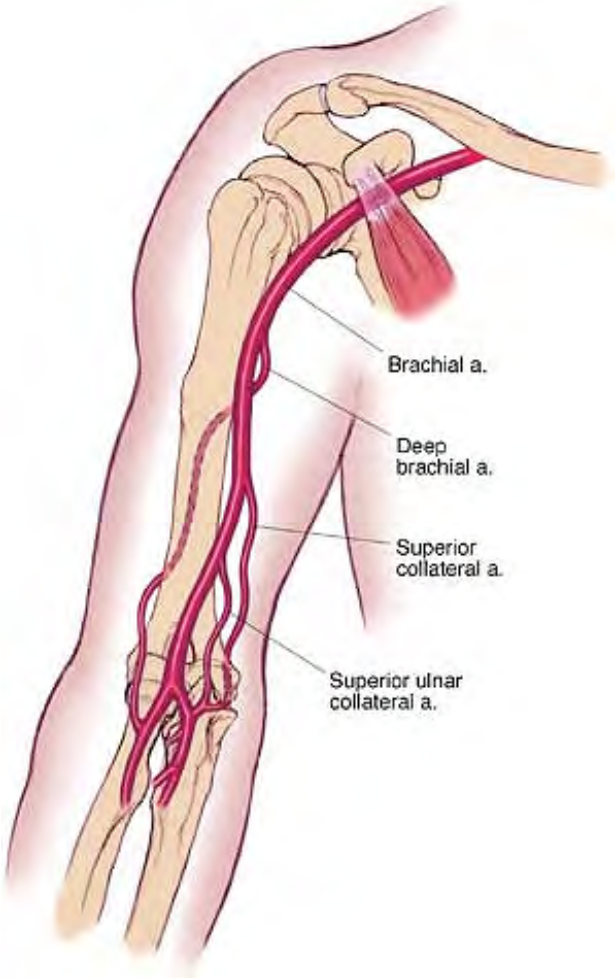

FIGURE 6-6 Surgical anatomy of the brachial artery.

The brachial artery is the most commonly injured artery of the upper extremity, accounting for 15% to 30% of all peripheral vascular injuries. Patients typically present with hand ischemia. Patients with supracondylar humeral fractures or elbow dislocations and hard signs of vascular injury should be further evaluated with arteriography to rule out brachial artery injury. The brachial artery is best exposed using a longitudinal incision along the course of the artery medial to the biceps muscle or an S-shaped incision across the antecubital fossa (Fig. 6-7). Most injuries are repaired with either end-to-end anastomosis or interposition vein graft.

Clinical outcome after repair is better than for injuries of the axillary or subclavian artery because the incidence of nerve injury is significantly less. Amputation or death is rare.

Radial and Ulnar Arteries

The ulnar artery is the larger branch of the brachial artery and gives off the ulnar recurrent arteries and the common interosseus artery (Fig. 6-8). The ulnar artery terminates in the superficial palmar arch. The radial artery gives off the radial recurrent artery, which anastomoses to the deep brachial artery. Distally, the brachial artery gives off a small branch to the superficial palmar arch and terminates in the deep palmar arch.

Injuries to the radial and ulnar artery are also common, but most often result from penetrating trauma. Complete occlusion or transection of either the radial or ulnar artery will often have no adverse effect on the circulation of the hand because of rich collateral circulation. However, patients with marked hematoma in the forearm may develop compartment syndrome requiring fasciotomy. Distal thrombectomy catheters should be handled gently because these arteries are prone to rupture or intimal injury. Repair often requires simple closure, end-to-end anastomosis, or small interposition grafts. Amputation is rare.

Lower Extremity Injuries

Common, Profunda, and Superficial Femoral Arteries

The common femoral artery originates at the level of the inguinal ligament as a continuation of the external iliac artery (Fig. 6-9). The common femoral artery is located adjacent to the femoral nerve (laterally) and common femoral vein (medially). The first branch of the common femoral artery is the superficial circumflex iliac artery, which marks the transition from external iliac to common femoral artery just below the inguinal ligament. The second branch is the superficial epigastric artery. Approximately 5 cm distal to the inguinal ligament the common femoral artery bifurcates into the superficial and deep femoral arteries. The deep (profunda) femoral artery is located in a posterolateral position. This artery gives off the medial and lateral femoral circumflex arteries and four to five perforator branches in the thigh. The superficial femoral artery follows its course

underneath the sartorius muscle (Hunter

canal) and terminates in the popliteal artery at the level of the adductor hiatus. The artery is adjacent to the saphenous nerve and femoral vein.

FIGURE 6-7 Preferred surgical approaches for exposure of the distal brachial artery and forearm vessels.

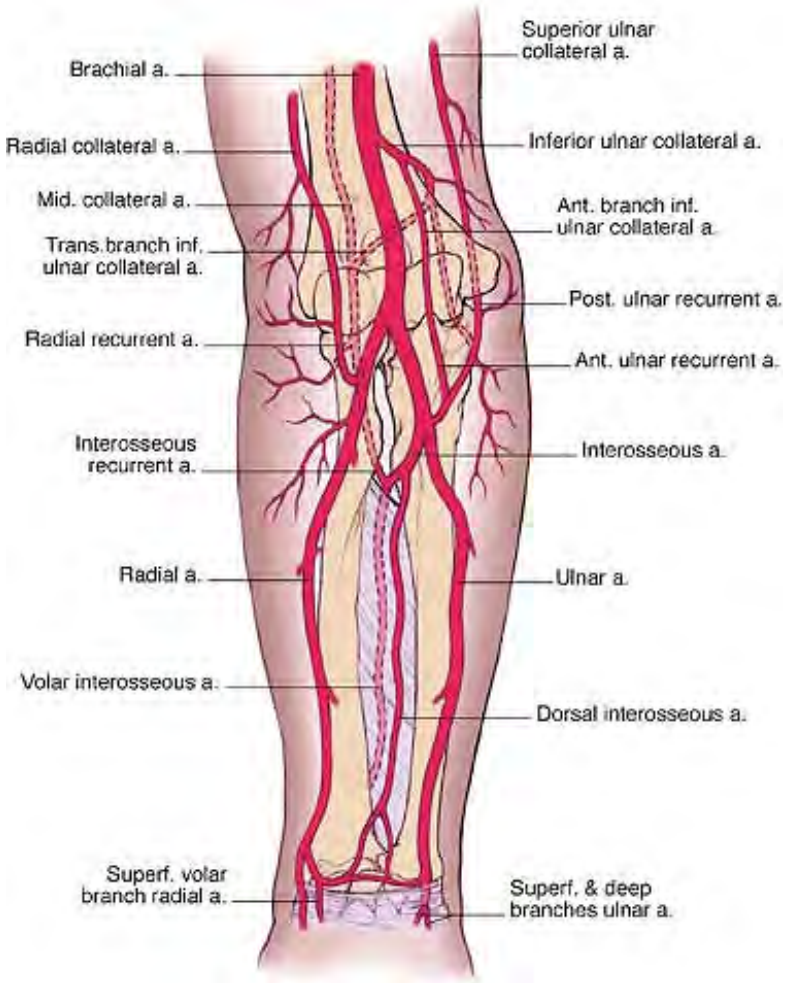

FIGURE 6-8 Surgical anatomy of the forearm arteries.

Injury to the femoral vessels accounts for one-third of all peripheral vascular injuries in civilian and military series. The most common mechanism of injury is penetrating trauma. Anterior dislocation of the femoral head is a rare cause of blunt injury. The superficial femoral artery is injured in 5% of the patients presenting with femoral shaft fracture. Bleeding is the most common presentation in cases of penetrating trauma, whereas distal limb ischemia predominates in patients with blunt injury.

Exposure of the common femoral artery and femoral bifurcation is best obtained

using a longitudinal incision two fingerbreadths lateral to the pubic tubercle (see Fig. 6-9). The incision extends from the level of the inguinal as far distally as necessary, depending on how extensive the injury is to the proximal deep and superficial femoral arteries. Rarely, a suprainguinal curvilinear incision is required for retroperitoneal exposure of the external iliac artery to achieve proximal control. Exposure of the superficial femoral artery can be achieved through a longitudinal incision along the course of the sartorius muscle. Patients with small, clean injuries of the femoral arteries may be treated with primary closure, end-to-end anastomosis, or patch angioplasty. Longer lesions require an interposition graft using autologous saphenous vein from the contralateral thigh. Long-term patency rates approach 100% for interposition grafts. Lower extremity function is predominantly determined by the extent of skeletal and neurologic injury.

Popliteal and Tibioperoneal Arteries

The popliteal artery is the continuation of the superficial femoral artery and originates at the adductor magnus hiatus (Fig. 6-10). The popliteal artery gives off multiple genicular collateral branches at the above- and below-knee level. The artery bifurcates in 90% of individuals and gives off the anterior tibial artery and the tibioperoneal trunk. The tibioperoneal trunk gives off the posterior tibial artery and the peroneal artery. The anterior tibial artery follows an anterior and lateral course, perforates the interosseous membrane, and is located in the anterior compartment of the leg along with the deep peroneal nerve. The peroneal artery is the middle branch of the three, follows its course in the deep posterior compartment, and bifurcates into the perforating branch to the anterior artery and communicating artery to the posterior tibial artery.

The posterior tibial artery is the most medial branch of the three tibioperoneal arteries and terminates as the common plantar artery below the ankle joint.

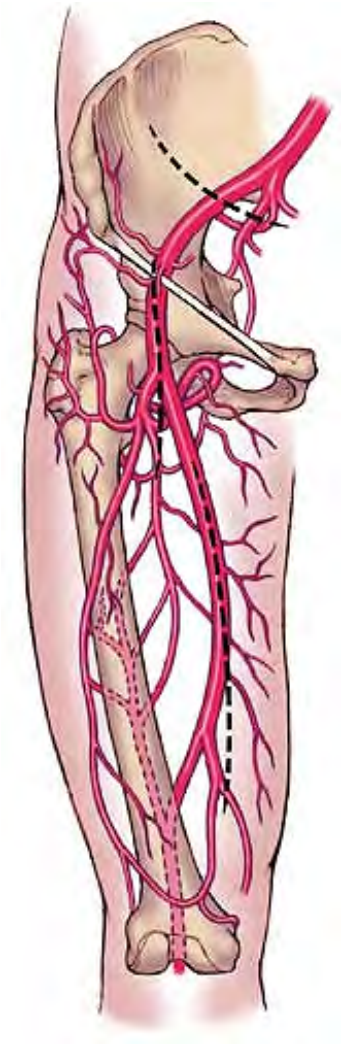

FIGURE 6-9 Surgical anatomy and preferred surgical approaches for exposure of the common, superficial, and deep femoral arteries.

Popliteal artery injuries often result from blunt trauma. Fracture or posterior dislocation of the knee is a known mechanism of injury. Most patients with popliteal artery injury present with distal ischemia because the genicular collaterals are not effective in maintaining adequate distal perfusion. Occlusion of a single tibioperoneal artery is well tolerated and does not require repair.

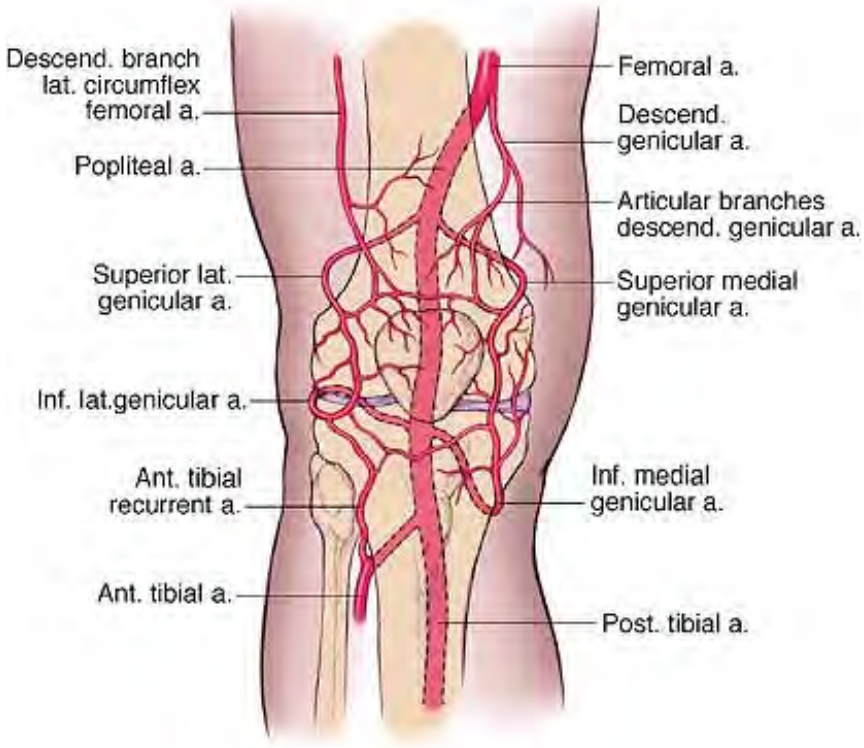

FIGURE 6-10 Surgical anatomy of the popliteal artery.

Most popliteal and tibioperoneal artery injuries can be exposed using a generous medial approach above and/or below the knee level depending on the extent of injury (Fig. 6-11). The great saphenous vein should be identified and protected. Division of the semimembranous and semitendinous muscles is often required for adequate exposure. The posterior tibial and peroneal arteries are also exposed using a medial incision along the posterior margin of the tibia. The soleus muscle is incised longitudinally allowing exposure of the mid- and distal portions of the posterior tibial and peroneal arteries. Although the origin of the anterior tibial artery can be exposed through a medial incision, an incision in the anterior compartment two fingerbreadths lateral to the tibia is required for exposure of the anterior tibial artery.

Injury to the popliteal and tibioperoneal arteries is associated with significant

morbidity. Clinical outcome is affected by the extent of skeletal and neurologic trauma, time of ischemia, and adequacy of runoff through the tibial and peroneal arteries.

Venous Injuries

The management of venous injuries is essentially identical to that for any arterial injury. Although venous repair does not yield the same satisfactory results of arterial repair and there is a higher incidence of early thrombosis, current recommendations are to attempt venous repair of any named major vein whenever possible. Large veins such as the inferior vena cava and iliac veins pose a problem in terms of conduit size mismatch. Options in these cases are use of spiral or paneled vein grafts or externally supported PTFE grafts. Injuries affecting minor veins or patients with other life-threatening injuries should be managed with primary ligation. Major veins can be exposed using the same incisions as for their arterial counterpart. Direct pressure using sponge sticks or pressure cuff tourniquets are excellent means of obtaining control in cases of bleeding. Techniques for repair should be the same as for arterial injuries with the caveat that catheter embolectomy often cannot be used because of competent valves.

Mangled Extremity

One should use clinical judgment in cases of severe trauma with complex soft tissue, skeletal, and vascular injuries. Although there are scoring systems to quantify the degree of injury, cases should be individualized. The mangled extremity severity score (MESS) grades skeletal/soft tissue damage, limb ischemia, shock, and age. However, this scoring system is inaccurate in identifying the irretrievable limb requiring primary amputation. Factors associated with need for amputation after below-knee fractures are injury of greater than three muscle compartments, occlusion of greater than two tibial arteries, and presence of cadaveric changes at initial presentation. When more than two of these predictive factors are present, none of the extremities were salvageable. Primary amputation should be considered early in patients with a dysvascular extremity and extensive soft tissue loss, as the chance of functional recovery is minimal.

FIGURE 6-11 Preferred medial knee approach for exposure of the below-knee popliteal artery and proximal tibioperoneal vessels.

Postoperative Management

Patients should be continuously evaluated for bleeding, distal limb perfusion, and development of compartment syndrome. Following massive bleeding and coagulopathy, one should aggressively resuscitate the patient and monitor laboratory results every 4 hours (complete blood count, platelets, prothrombin and partial thromboplastin times, and fibrinogen levels). We recommend that laboratory levels are kept within normal limits during the first 48 hours. Parameters include hemoglobin above 10 g/dL, platelet count above 100,000, international normalized ratio of less than 1.2, and partial thromboplastin time of less than 40 seconds. Mild abnormalities contribute to postoperative â€œoozing,â€• large hematoma, and initiate a cascade of coagulopathy that ultimately may lead to

bleeding requiring re-exploration. Assessment of vascular pulses should be done routinely with serial examination and hand-held Doppler interrogation. This is repeated every 2 hours for the first 24 hours and every 4 hours thereafter.

The injured extremity should be checked for any signs of infection. Early wound infection, particularly if associated with soft tissue necrosis, is one of the most important determinants of delayed amputation. Aggressive perioperative antibiotic therapy is recommended.

Compartment syndrome is a common occurrence in patients with combined vascular and orthopaedic trauma. Patients with ischemic limbs for more than 4 hours can be expected to have some degree of compartment syndrome. Presentation can be delayed up to 12 to 24 hours after reperfusion of the ischemic limb. Early or prophylactic fasciotomy in this setting may be associated with improved outcome. If fasciotomy is deferred at the time of arterial reconstruction, physical examination and compartment pressure measurements should be repeated frequently. Compartment pressures of higher than 30 cm H^2O strongly indicate fasciotomy.

Complications

The most common complications after vascular reconstructions are bleeding, infection, early thrombosis and development of compartment syndrome. Medical complications involving cardiac, pulmonary, renal, and neurologic systems are also common. Amputation can result because of early technical failure, venous outflow problems, infection, and fracture instability. Deep venous thrombosis is a common occurrence after combined vascular and orthopaedic trauma. Aggressive prophylaxis should be instituted in the early postoperative period. Postoperative duplex ultrasound is indicated for any sign of worsening limb edema or pain. Patients need to be followed regularly because of the potential for late complications. Arterial stenosis or thrombosis, aneurysmal degeneration of vein grafts, and chronic venous insufficiency after primary venous ligation or thrombosis of venous repair can be identified using duplex ultrasound surveillance. We generally re-evaluate patients with combined injuries 6 months after dismissal or earlier if the patient develops symptoms of chronic limb ischemia (e.g., claudication).

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 7

Management of the Soft Tissue with Shoulder Trauma

Scott F. M. Duncan

John W. Sperling

Indications/Contraindications

Shoulder trauma can result in significant soft tissue loss over some of the bony prominences of the shoulder such as the acromion and clavicle. Areas of the scapula can be exposed as well from trauma about the shoulder. Fortunately, the glenohumeral joint itself is covered with a multilayer muscle configuration. If there are concomitant fractures, nerve injuries, or vessel injuries, one must keep in mind not only the pre-existing soft tissue trauma and how to protect the healing of that, but also how to obtain the best exposure to facilitate fixing the concomitant injuries. Open fractures and neurovascular injuries require emergent intervention, or otherwise the extremity may be compromised. In such circumstances, the rescue of the threatened limb takes priority. However, again, one must keep in mind the potential need for soft tissue coverage in these patients. Fortunately, numerous flap options exist about the shoulder for soft tissue coverage (see Chapter 8).

The addition of locking plates, which are fixed angle devices for fixing proximal humerus fractures, has greatly augmented our armamentarium for treating these fractures. In significant injuries, it is not unreasonable to perform a delayed

intervention once the soft tissues have been covered or stabilized. Concomitant injuries such as rotator cuff tears can be addressed months down the line, if needed, once the other injuries and soft tissue envelope have finished healing. When the wound bed is grossly purulent, coverage and otherwise operative repair should be delayed until a clean wound can be established. This being said, it is obviously preferable to have all structures fixed prior to any type of soft tissue coverage to minimize injury to the soft tissue flap upon re-operation or re-exploration of the wound.

Preoperative Planning

A thorough examination of the shoulder and the involved extremity as well as a radiographic study of the shoulder are mandatory to recognize soft tissue and bony injuries. Neurovascular injuries may be easily missed if the more distal aspects of the extremity are not examined. Furthermore, radiographic examination must include axillary lateral views to rule out a glenohumeral dislocation. Active flexion, abduction, internal and external rotation should be tested about the shoulder when possible. Active flexion and extension about the elbow, wrist, and fingers will also help delineate more proximal neurovascular compromise. Sensation and motor function of the extremity should be documented. Radial pulses as well as ulnar pulses need to be documented as well as the time for capillary refill.

From a radiographic standpoint in the emergent setting, a computed tomography scan sometimes is required to further delineate the characteristics of bony trauma. Occasionally, an arteriogram may be needed to investigate vascular injuries. The most commonly injured nerve about the shoulder is the axillary nerve, and an examination of deltoid function and shoulder sensation should be performed. When planning the surgical exposure or exposures, one must appreciate what other injuries need to be addressed, if any, and how the incision and exposure will jeopardize any pre-existing skin or muscle trauma. Incisions should be selected that will not compromise the viability of the soft tissue

envelope and that once they have healed will not create a tethering-type scar, resulting in loss of shoulder range of motion.

Surgery

Patient Positioning

We prefer to perform surgery about the shoulder with the patient in a beach-chair position. Depending on the injuries to be addressed, supine and lateral decubitus may be used as well. However, in our experience, the beach-chair position has the advantages of allowing manipulation of the arm as needed to facilitate reduction or retraction as well as providing access to both the anterior and posterior aspects of the shoulder. These surgeries are usually done under general anesthesia, but postoperative pain management can be enhanced by the administration of interscalene blocks by the anesthesiologist. The patient is brought into the regular operating room where preparation and draping are carried out in the usual sterile fashion and manner. As with any surgery, atraumatic technique should be adhered to in an effort to lessen skin, muscle, and other soft tissue necrosis. Poorly executed shoulder surgery can result in greater functional loss than would have otherwise occurred had no surgical intervention been attempted. However, as with any surgery, risks are involved and nerve and vessel injury can occur. Excessive scarring and postoperative joint contractures may result despite the best surgical techniques.

The beach-chair position facilitates the use of fluoroscopy during the course of the procedure (Fig. 7-1). This flexibility is sometimes needed when attempting to image complex injuries. Sometimes it is necessary to extend the wound of the injury both proximally and distally to provide visibility to the area of trauma. However, most important is that an adequate view of the surgical field be created to avoid the need to perform surgical repairs in a challenged manner through a small wound. In general, we try to avoid â€œTâ€• extensions of transverse lacerations, but occasionally this may need to be done to gain access to neurovascular structures. Again, the type of shoulder injury will necessitate the surgical approach.

Technique

For proximal humerus fractures, there are two basic approaches. The first is a superior deltoid approach in which a skin incision is made in Langer's lines just lateral to the anterior lateral aspect of the acromion. This approach allows the deltoid to be split from the edge of the acromion distally for approximately 4 to 5 cm. Again, care must be taken to protect the axillary nerve. Of note is that the deltoid origin is not removed, but there is still exposure of the superior aspect of the proximal humerus.

This type of exposure is quite useful for a fixation of greater tuberosity fractures as well as the insertion of intramedullary nailing. In the beach-chair position rotation, flexion, and extension

of the humerus can enhance the exposure of the underlying structures as well as reduction of the fragments.

FIGURE 7-1 A,B: The beach-chair position helps facilitate use of intraoperative fluoroscopy.

The second approach is a long deltopectoral approach in which both the deltoid origin and insertion are preserved (Fig. 7-2). The skin incision begins just inferior to the clavicle and extends across the coracoid process and down into the area of insertion of the deltoid on the humerus. The cephalic vein should be preserved and retracted either laterally or medially depending on which is easiest for the surgeon. The cephalic vein is less likely to be injured when taken medially, given that over-vigorous retraction on the cephalic vein and deltoids can result in vein injury. If the vein is accidentally transected or significantly injured, either from the trauma or the surgical approach, ligation can be considered. In the deltopectoral approach, if more exposure is needed to gain access to the proximal aspect of the humerus, the superior part of the pectoralis major tendon insertion can be divided. This may need to be done for multipart fractures. In significant trauma injuries, the pectoralis major tendon may be

found to be disrupted and should be repaired when possible using nonabsorbable suture and/or suture anchors.

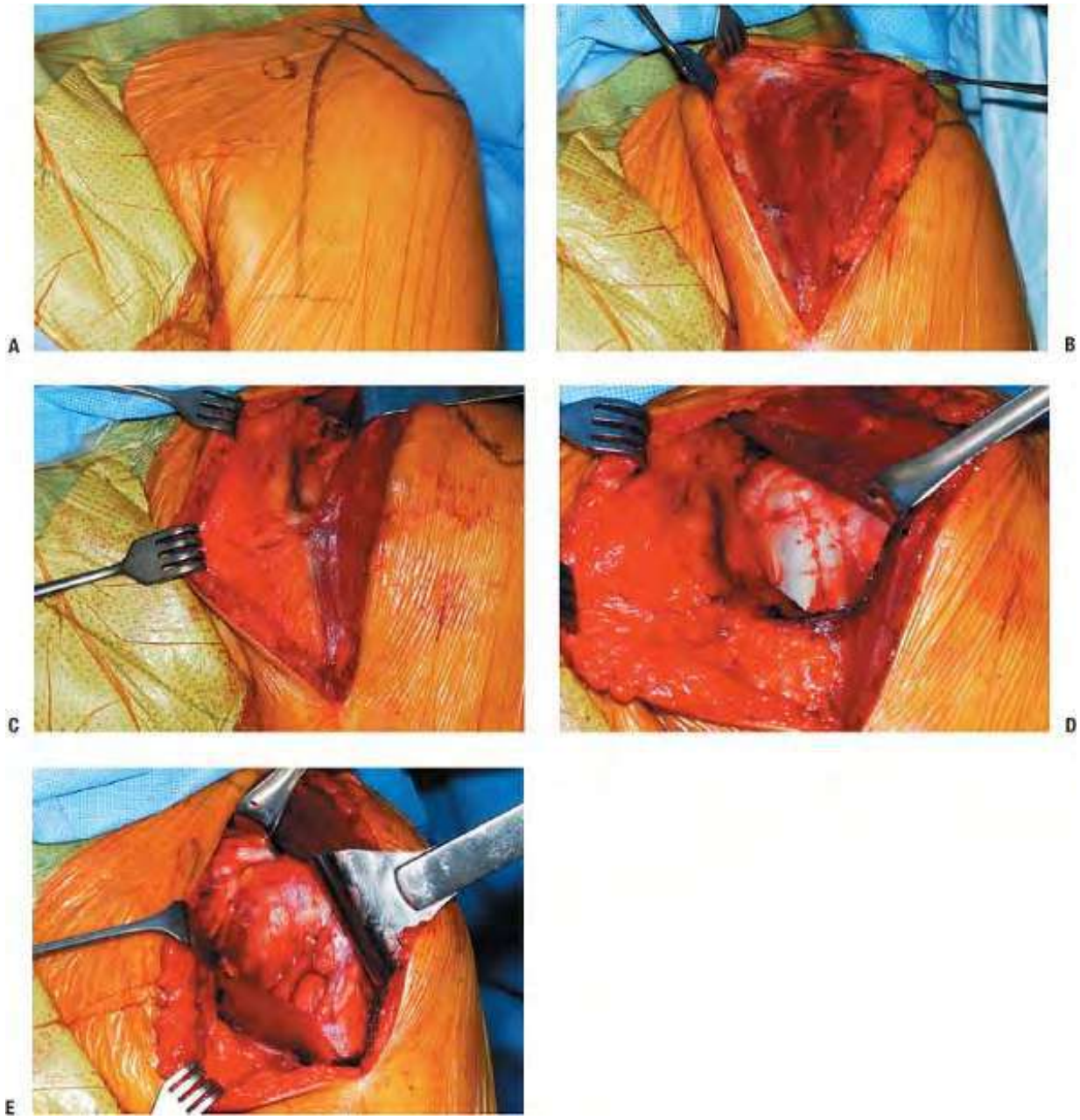

FIGURE 7-2 A: The landmarks of the shoulder are outlined and the incision is marked. B: A triangle of fat is usually present at the proximal interval between the pectoralis and deltoid. C: Typically, the cephalic vein is left within its bed

medially. D: One continues the deltopectoral interval distally. E: One places a retractor medially beneath the conjoint group and one laterally to retract the deltoid.

We do not advocate procedures that split or remove part of the acromion given the significant risk of complications. Methods of internal fixation of proximal humerus fractures go beyond the scope of what this chapter is intended to provide. However, a brief mention of various options will be discussed. As mentioned earlier in the chapter, there is an improved role for open reduction internal fixation of these fractures given the new fixed angled locking plates that have been developed. However, there still remains a role for intramedullary nailing, tension band techniques, and percutaneous pinning of proximal humerus fractures. Most important, though, is that a majority of proximal humerus fractures in our practices are still treated nonoperatively. Shoulder arthroplasty is still used in those cases where there is significant comminution and adequate fixation cannot be achieved by other means. Successful treatment of shoulder injuries requires understanding of their specific biologic and mechanical challenges. This should also encompass knowledge about the natural history and potential complications from the various treatment options.

With the beach-chair position, we take care to secure the head with use of towels to the head supporting apparatus. This can be done with prefabricated devices such as the Skytron attachment. Care needs to be taken to make sure the neck angle is appropriate. Also, when securing the head, care needs to be taken that there are no pressure points on the eyes. With regard to the anterior approach, the exposure of the shoulder joint is useful for drainage of sepsis, open reduction internal fixation of fractures, hemiarthroplasty, and reconstruction of dislocation trauma. Landmarks for this approach are the coracoid process and the deltoid insertion. The incision starts at about the level of the clavicle and goes over the coracoid process through the area where the deltopectoral groove can be palpated down to the anterior aspect of the shoulder. An alternative to this incision is an axillary incision. The shoulder must be abducted and externally rotated to make this type of approach. The downside of this in the trauma setting is that the skin flaps must be undermined extensively to visualize the area of the deltopectoral groove. However, this approach does have a cosmetic

advantage. This approach is also difficult in extremely muscular patients.

The deep surgical dissection involves identifying the short head of the biceps and coracobrachialis; these are retracted medially. The overlying fascia must be carefully incised. The long head of the biceps can be used to identify the division between the lesser and greater tuberosities. This can also be a good marker for where to begin subscapularis dissection. The arm should be kept adducted at most times, because bringing the arm into abduction brings the neurovascular structures closer to the operative field. This is especially true if any work is being done around the coracoid. Care must be taken when retracting the coracoid and coracobrachialis, as over-vigorous retraction can cause musculocutaneous paralysis. Care should be taken with the axillary nerve as well. External rotation of the shoulder increases the distance between the subscapularis and the axillary nerve. A leash of vessels on the inferior border of the subscapularis frequently requires cauterization to maintain good visualization in the surgical field.

The anterolateral approach in the trauma setting is most useful for repair stabilization of the long head of the biceps tendon as well as fixation of greater tuberosity fractures. A transverse or longitudinal incision can be used beginning at the anterolateral corner of the acromion and extending lateral to the coracoid process or going over the anterolateral aspect of the shoulder. The superficial dissection involves subcutaneous fat and deltoid fascia. This is incised in line with the skin incision. The split should not extend more than 5 cm down from the acromion as axillary nerve injury is possible. In the case of rotator cuff tears, one may consider taking the deltoid off the anterior aspect of the acromion, which also improves visualization (Fig. 7-3). Avoid inadvertent inferior retraction as well, which can also injure the axillary nerve. Bleeding will obstruct the view and is frequently encountered during this dissection. Try to obtain good hemostasis and cautery of the acromial branch of the coracoacromial artery; it usually must be electrocauterized. Also avoid excessive stripping and detachment of the deltoid. Deep dissection will reveal the rotator cuff under surface of the acromion and greater tuberosity. A stay suture in the fascia of the deltoid can be useful to prevent excessive splitting and injury to the axillary nerve.

The lateral approach provides limited access in a similar fashion to the anterolateral approach. Again, it can be useful for open reduction internal fixation of greater

tuberosity fractures, as well as for open reduction internal fixation of neck fractures using a tension band technique. The rotator cuff is easily accessible through this approach. This is a common approach for insertion of intramedullary rods into the humerus. An incision approximately 5 cm long is made from just above the acromion down the lateral aspect of the arm. Deltoid muscle is split in line with its fibers down to a level no more than 5 cm. Most surgeons insert a suture at the inferior aspect of the split to help prevent it from extending any further and potentially injuring the axillary nerve. Deep surgical dissection is continued

P.97

P.98

through the deltoid. The rotator cuff and its bursa in the subacromial space are immediately encountered. Again, the humeral neck, greater tuberosity, rotator cuff, and even the long head of the biceps, to some degree, can be accessed from this approach. Care needs to be taken to avoid injuring the axillary nerve or detaching an excessive amount of deltoid.

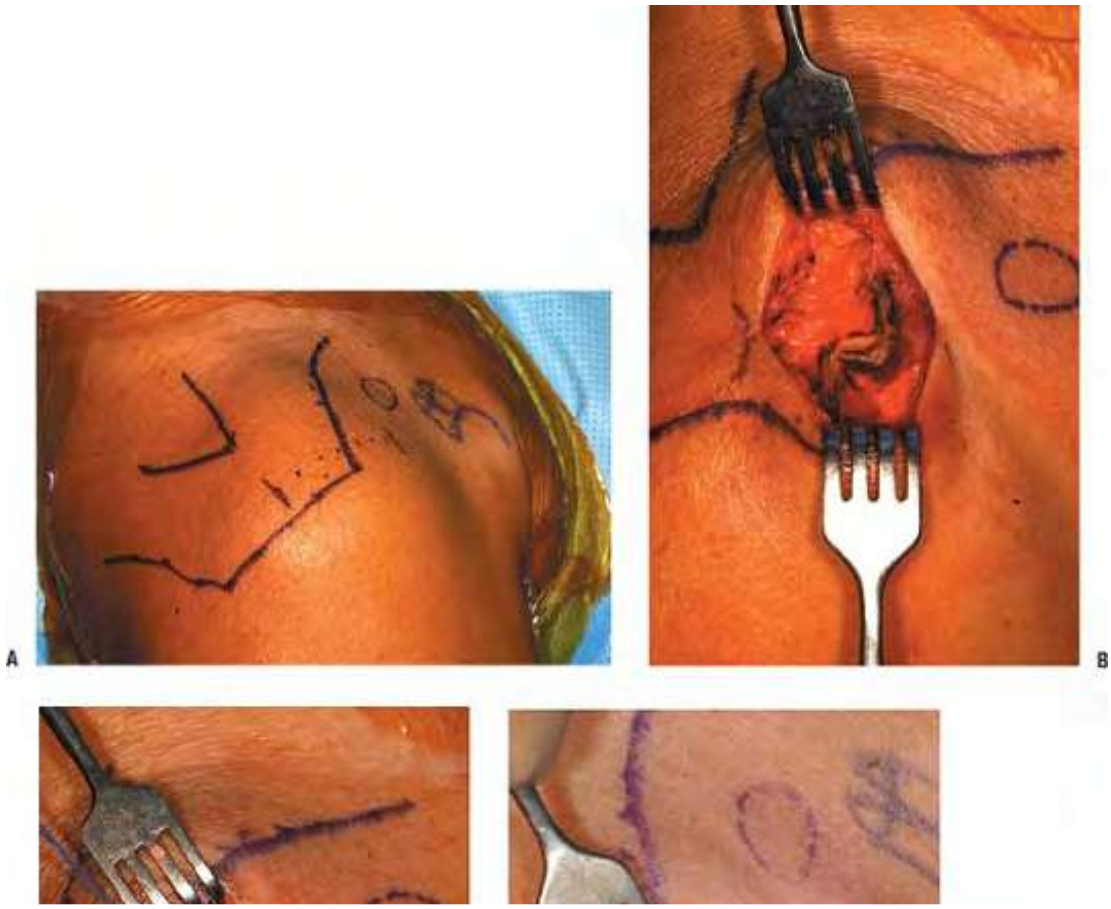

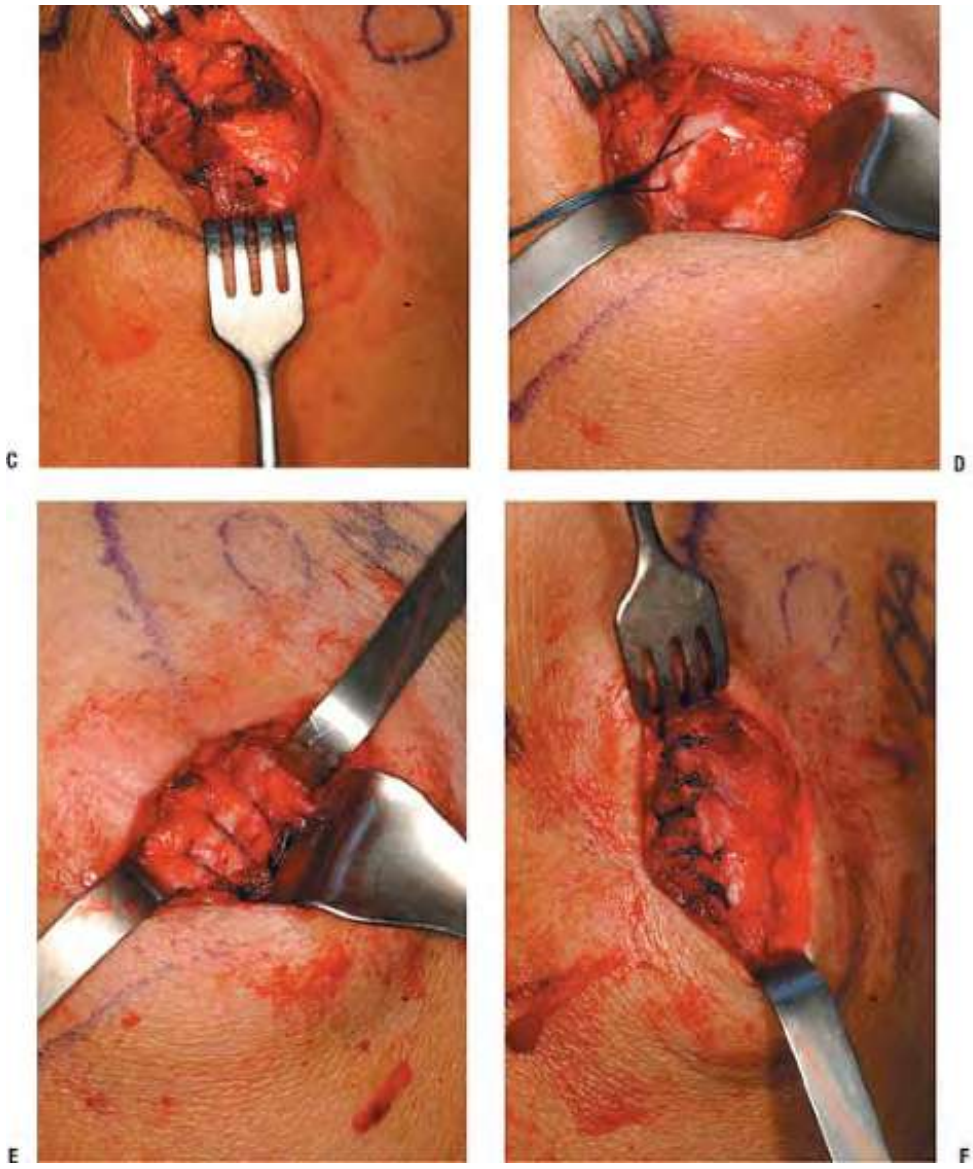

FIGURE 7-3 A: An incision is marked out parallel to the lateral border of the acromion. B: The deltoid is taken off of the acromion as outlined. C: A stitch is placed in the corner of the deltoid to assist in later repair and ensure proper alignment deltoid repair. D: The rotator cuff tear is identified and retention stitches are placed. E,F: A rotator cuff repair is performed.

The posterior approach offers access to the posterior and inferior aspects of the glenohumeral joint. In the trauma scenario, open reduction and internal fixation of

the glenoid, scapular body, and sepsis drainage can be provided. Depending on the amount of exposure required, the deltoid can be split rather than detached. The interval between the deltoid muscle and the inner lying infraspinatus muscle is identified next. The internervous interval between the infraspinatus and the teres minor is next identified and is usually developed with blunt finger dissection. Some surgeons prefer to use the interval between the two heads of the infraspinatus. By retracting each of these muscles, the inferior and posterior aspects of the glenoid can be reached.

If the joint needs to be explored, incise it longitudinally close to the scapular edge. Care needs to be taken to protect the axillary nerve as it runs in the quadrangular space beneath the teres minor. The other nerve that can potentially be injured is the suprascapular nerve as it passes around the base of the spine of the scapula and runs from the supraspinatus fossa to the infraspinatus fossa. Damage to this can cause loss of innervation to the supraspinatus and infraspinatus muscles. Excessive retraction of the infraspinatus medially can result in neurapraxia. Finally, the posterior circumflex humeral artery runs with the axillary nerve in the quadrangular space beneath the inferior border of the teres minor; and again, if the dissection is performed too low, the structure can be injured.

Postoperative Management

The shoulder is a complex joint and will rapidly become stiff. Ideally, adequate motion is needed for optimum function of the extremity for the patient. Rehabilitation depends on type of injury, stability of repairs, and amount of soft tissue injury. In most cases, the three-phase system developed by Hughes and Neer will suffice. The first phase incorporates passive-assistive exercises. Active and early resistive exercises are started in the second phase. Finally, the third phase is a program dedicated to stretching and strengthening exercises. In general, these exercises are performed three to

four times per day for 20 to 30 minutes. In the first phase, it is not uncommon for pain medication to be used. Working with a skilled physical therapist can improve patient compliance and outcome.

Phase one is started in the early posttraumatic period. If there is secure fixation of

the underlying injuries, passive exercises may be begun 24 to 72 hours after the index procedure. The surgeon must be careful not to forget about the elbow, wrist, and fingers. If these are neglected, permanent stiffness can result in these joints, and the extremity is at risk for complex regional pain syndrome (hand-shoulder syndrome). The first exercise usually consists of the pendulum or Codman's type in which the arm is rotated both outwardly and inwardly in small circles. Next, the supine external rotation with a stick exercise is used. The elbow and distal humerus must be supported with either a towel or sheets. A 20-degree shoulder abduction can facilitate performing this exercise. Gentle pulley exercises can be added to the regimen as well as extension. In general, isometrics are not started until there is evidence of some fracture healing.

Phase two exercises involve active, gentle resistive, and stretching programs. Supine active forward elevation is the first one because the effects of gravity are reduced by making the elevation easier. The patient then can be progressed to doing this exercise in a standing or sitting position. A stick in the unaffected arm can assist the involved arm with this maneuver. TheraBands can be used to strengthen the internal and external rotators as well as the anterior, middle, and posterior aspects of the deltoid. These are usually done in repetitions of 10 to 15 per session. Wall climbing exercises with stretching are also performed. Internal rotation can be one of the most difficult aspects of motion to regain in the shoulder, and the patient can work on assisting the affected extremity with the unaffected one by trying to have the hand crawl up the back of the patient.

Phase three exercises are started after solid evidence of fracture healing. The rubber strips may be replaced by rubber tubing to increase resistance. Continued aggressive stretching is performed. Light weights are usually tolerated at this time. Weights start at 1 lb and are increased in 1-lb increments, usually stopping at 5 lb. If there is significant pain with the weight, then these should be discontinued as strength can generally recover with functional activity.

Complications

Postoperative Complications

Postoperative complications, as previously mentioned, include neurovascular injury

and soft tissue loss. In the traumatized shoulder, the usual anatomic landmarks may be distorted or absent. This makes it imperative that surgeons take their time and appropriately identify neurovascular structures and internervous planes when possible. Other factors such as hardware loosening and infection can occur. In the event of excessive soft tissue loss, flap coverage may be required (see Chapter 8). Free flaps such as a latissimus or anterolateral thigh flap can be considered depending on the need for muscle versus subcutaneous and cutaneous layers. With pinning techniques, pin migration can be a problem and has been reported in the literature.

In general, complications can be broken down into vascular injury, brachial plexus injury, chest injury, myositis ossificans, frozen shoulder, avascular necrosis, nonunion and malunion, and complex regional pain syndrome. Fortunately, vascular injuries are infrequent, but they can occur and can result in loss of the extremity. Injury of the axillary artery represents about 6% of all arterial traumas and usually occurs secondary to fractures of the proximal humerus. It can be seen in older individuals with significant atherosclerosis as the vessel may tear secondary to dislocation or blunt trauma. The most common site of injury to the axillary artery is proximal to the take-off of the anterior circumflex artery. It is important to check the radial pulse, but this is no guarantee that arterial injury has not occurred. Other signs to look for include an expanding hematoma, pallor, and paresthesias. Missed arterial injuries can result in gas gangrene, amputation, and compressive neuropathies at the brachial plexus leading to permanent nerve palsies. If three is any suspicion, angiography should be performed immediately.

Brachial Plexus Injuries

Brachial plexus injuries can occur after proximal humerus fractures. This has been reported to have an incidence of 6.1%. Isolated injury to the axillary nerve is not uncommon and has been reported, but any or all components of the brachial plexus can be involved. It is, thus, extremely important to evaluate the patient initially and document any compromise in skin sensation and motor function. If a nerve injury is clinically suspected, it should be carefully followed, and electromyography and

nerve conduction velocity studies should be used to monitor for progression or failure of healing. Nerve injuries that fail to show improvement clinically and electrically may need to be explored.

Chest Injury to the Thoracic Cavity

Chest injury to the thoracic cavity can occur after fractures and trauma to the shoulder. There have been reports of intrathoracic dislocation of the head with surgical neck fractures of the humerus, as well as pneumothorax and hemathorax associated with the trauma that resulted in the proximal humerus fracture.

Frozen Shoulder

Frozen shoulder is extremely common after any type of injury or surgery to the shoulder. Even with implementation of well-organized and carefully monitored physical therapy, adhesive capsulitis can develop. Fortunately, most cases of adhesive capsulitis respond to programs of exercise and stretching. In those that fail to improve after 4 to 6 months, manipulation and lysis of adhesions can be considered. There is a risk of refracture, however. Painful and impinging hardware can also result in shoulder motion limitations; if the patient desires, once fracture union is obtained, the hardware may be removed.

Myositis Ossificans

Myositis ossificans is known to occur after fracture dislocations. It can also occur in head-injured patients. It is more commonly seen when there is a chronic unreduced fracture dislocation. Avascular necrosis of the humeral head can occur with shoulder trauma. This is more common in three- and four-part fractures, but can occur with any type of injury about the shoulder, including soft tissue injuries. Excessive periosteal stripping and excessive soft tissue dissection may also result in this complication. Nonunions and malunions can occur with any type of fracture and these may result from fractures about the shoulder. Their treatment goes beyond the purpose of this chapter.

Complex Regional Pain Syndrome

This entity, formerly known as reflex sympathetic dystrophy and before that as hand-shoulder syndrome, is cited to affect from 10% to 30% of patients who have shoulder trauma that required surgery. Soft tissue and bony injury can result in this pathologic process marked chronic pain. The patients do not necessarily have to have any type of penetrating or open trauma. The treatment of complex regional pain syndrome is early

recognition of the problem, sympathetic blocks, Neurontin, and aggressive therapy. Complex regional pain syndrome can occur in any setting with light or aggressive therapy and minimal or extensive trauma. Once the condition is recognized, treatment should be initiated. If there is any question, quantitative sudomotor axon reflex testing may be helpful.

Recommended Reading

Blom S, Dahlback LO. Nerve injuries in dislocations of the shoulder joint and fractures of the neck of the humerus: a clinical and electromyographic study. *Acta Chir Scand*. 1970;136:461â€“466.

Hughes M, Neer CS II. Glenohumeral joint replacement and postoperative rehabilitation. *Phys Ther.* 1975;55:850â€“858.

Linson MA. Axillary artery thrombosis after fracture of the humerus. *J Bone Joint Surg*. 1980;62A:1214â€“1215.

Neer CS II. Displaced proximal humeral fractures. Part I. Classification and evaluation. *J Bone Joint Surg.* 1970;52A:1077â€“1089.

Neer CS II. Displaced proximal humeral fractures: Part 2. Treatment of three-part and four-part displacement. *J Bone Joint Surg*. 1970;52A:1090â€“1103.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 8

Rotational Flaps for Rotator Cuff Repair

Emilie V. Cheung

Scott P. Steinmann

Indications/ Contraindications

Massive soft tissue tears of the rotator cuff pose significant challenges to the shoulder surgeon. Although most tears can be successfully repaired in a primary fashion, some massive tears of the rotator cuff are termed irreparable, either due to their severely contracted state or simply because of their size. Cofield has defined the term massive rotator cuff tear as greater than 5 cm, whereas others have defined it as involvement of at least two tendons.

Chronic tears, commonly seen in older patients, have evidence of attritional changes in both the tendon substance and the muscle fibers. This has been termed fatty degeneration by Gerber and can be best characterized and graded based on its appearance on T1 weighted magnetic resonance imaging (MRI) of the shoulder in the oblique sagittal plane. The presence of fatty degeneration of the musculotendinous unit may compromise the results of even a seemingly successful repair due to poor tissue quality and limited regenerative capacity. The limitations of successful tendon repair are thus a function of the length of time since injury, the degree of tissue contraction, and extent of fatty degeneration.

Management of such tears often begins with conservative management. If this fails,

surgical management may include debridement of the cuff tear, which may result in satisfactory pain relief, but will not restore strength or function. Partial repair may be possible in some instances, but this does not restore normal function. Other options include allograft or synthetic tendon augmentation, but these have met with varying success in the literature.

Long-standing rotator cuff tears result in altered kinematics of the glenohumeral joint, and rotator cuff tear arthropathy may eventually develop as a painful end-stage condition. Hemiarthroplasty can be a reasonable option in patients who have an intact coracoacromial arch limiting anterior superior escape of the humeral head. The reverse shoulder prosthesis has recently been shown to be an effective treatment option for patients with irreparable rotator cuff tears and anterior superior instability of the humeral head. At mid-term follow-up, improvements in range of motion and pain relief have been reported in the majority of cases. However, this is generally reserved for select indications in an elderly patient. The long-term results of this prosthesis have yet to be determined.

A patient who has had a previously failed massive rotator cuff repair may be challenging to treat. Tendon transfers are another option for salvaging massive rotator cuff defects. They have the distinct advantage in their ability to restore strength, in addition to providing pain relief. Local tendon transpositions historically have included superior subscapularis, teres minor, and teres major transfers. Deltoid flap reconstruction has also been reported. Our preferred method of tendon transfer for posterior superior rotator cuff tears is the latissimus dorsi transfer.

Gerber et al described the procedure of latissimus dorsi transfer as an option for the treatment of massive posterior superior rotator cuff defects in 1988. This procedure has been shown to restore an active external rotation and forward flexion moment at the shoulder, which are the primary functional deficits for this type of massive rotator cuff tear.

Latissimus transfer in cases of a failed previous rotator cuff tear results in overall improvement in function and pain relief, but is generally inferior to that observed after a primary transfer. A comparative analysis on the results of latissimus dorsi transfer for primary irreparable rotator cuff tears, versus for salvage reconstruction for a failed previous rotator cuff repair was performed by Warner and Parsons.

Patientsâ€™ satisfaction and function were superior in primary latissimus dorsi transfer when compared with patients undergoing salvage reconstruction of failed previous rotator cuff repairs.

Subscapularis tendon ruptures occur less frequently than supraspinatus and infraspinatus ruptures. However, they can occur in isolation, during anterior shoulder dislocation, or an as extension of a massive rotator cuff tear. Our preferred method of tendon transfer for irreparable subscapularis tendon tears is the pectoralis major transfer. Combined latissimus dorsi transfer with pectoralis major transfer may be indicated in massive rotator cuff deficiency involving both the posterior superior cuff and the subscapularis.

The primary indication for latissimus dorsi transfer is the loss of active external rotation due to an irreparable tear of the posterior superior rotator cuff. The patient is dissatisfied with increased functional demands, which are beyond the limitations of conservative treatment.

The primary indication for pectoralis major tendon transfer is irreparable rupture of the subscapularis, characterized by retraction of the musculotendinous unit to the glenoid, and fatty infiltration of the muscle. In a patient with posterosuperior rotator cuff tear with extension of the defect into the subscapularis, isolated latissimus transfer is contraindicated unless there is either concomitant repair of the subscapularis, or transfer of the pectoralis major, to compensate for subscapularis dysfunction.

Preoperative Planning

We begin with the history and physical examination, which are essential in diagnosis and assessment of the nature of the rotator cuff tear. Often, the patient complains of pain or dysfunction with an insidious onset. Other times, the patient complains of fatigue during forward elevation or external rotation of the shoulder. No history of trauma can be recalled and the symptoms have been chronic over months or years, which leads us to believe that the etiology is attritional. Thus, we suspect poor tendon quality and the possibility of an irreparable tear.

The physical examination of the shoulder begins by inspection of both shoulders. Deformity associated with anterior superior escape of the humeral head indicates

anterosuperior rotator cuff tear, which is a contraindication for latissimus transfer. Spinati muscle atrophy is noted, and often indicates an irreparable rotator cuff tear is present. A defect of the deltoid is a relative contraindication for this procedure. Manual motor testing of shoulder forward elevation, abduction, external rotation, and internal rotation are recorded.

The *hornblower's sign*, or external rotation lag sign (discrepancy between active and passive external rotation of greater than 30 degrees), is important to recognize. A positive sign is highly suggestive of a massive rotator cuff tear, amenable to latissimus dorsi transfer.

Another physical examination finding that is important to recognize is the *lift-off test* (the inability to actively lift the dorsum of the hand off from a resting position on the lower back), or *belly press test* (the inability to actively maintain the elbow anterior to the midline of the trunk as viewed from the side). An inability to perform these maneuvers is highly suggestive of involvement of the subscapularis in the rotator cuff tear. If a subscapularis tear is present and determined irreparable, then a pectoralis major transfer should be performed in conjunction with the latissimus transfer.

Active and passive range of motion in forward elevation, external rotation, and internal rotation are documented. Loss of passive external rotation and forward elevation is a contraindication to this procedure. Active range of motion is also measured. Patients need to be able to achieve active forward elevation to 90 degrees to be candidates for the procedure. If the patient has pseudoparalysis and minimal forward elevation, the procedure will fail. The main goal of the procedure is to promote active external rotation to eliminate the hornblower's sign and enable the patient to reach the top of the head.

It should be noted that some patients with massive cuff tears are amenable to conservative treatment, and eventually may experience minimal pain and satisfactory function, especially in the elderly

population. Latissimus dorsi transfer is more commonly performed in those individuals with higher functional demands.

Imaging studies are useful for assessing the patient who is suspected of having a painful, torn rotator cuff, unresponsive to conservative treatment. Our standard

radiographic views include a true anteroposterior (AP) view of the shoulder and an axillary view. The true AP view best demonstrates superior migration of the humeral head in the context of massive rotator cuff tears. The acromiohumeral distance (ACHD) is measured on this view. A normal ACHD value is 10.5 mm. An ACHD value of 7 mm or less suggests an irreparable tear of the infraspinatus and is a strong indicator to consider a latissimus dorsi transfer, in the context of the previously mentioned physical examination findings. The true AP and axillary views demonstrate degenerative changes of the glenohumeral joint, which would be a contraindication to tendon transfer.

Magnetic resonance imaging is commonly performed at our institution in patients with suspected large rotator cuff tears in whom we are considering surgical treatment. It is useful for assessing not only the size of the tear, but the degree of muscle atrophy and the degree of fatty degeneration. Advanced fatty degeneration and a massive tear indicate a low likelihood of successful repair, but may be considered for latissimus dorsi transfer with or without pectoralis major transfer. In many patients with massive rotator cuff tears, we are able to repair the rotator cuff tear arthroscopically. Therefore, in almost all patients who are candidates for a latissimus dorsi transfer, we recommend an attempted arthroscopic repair as the initial surgical procedure.

Latissimus Dorsi Transfer

Surgery

The patient is placed in the lateral decubitus positioning with all bony prominences well-padded. The superior approach for rotator cuff repair in Langer's lines is used. Skin flaps are raised. The deltoid is split in line with its fibers or as an alternative if greater exposure is desired; the deltoid can be reflected from the acromion with a bone chip using an osteotome (Fig. 8-1). A subacromial retractor is placed to visualize the rotator cuff defect. Excise the subacromial bursa for optimal visualization. If the tear is determined to be irreparable, proceed with latissimus dorsi tendon transfer.

A second skin incision is placed along the posterior aspect of axilla, along the lateral border of the latissimus (Fig. 8-2). Skin flaps are created. The latissimus and teres major muscles are visualized (Fig. 8-3). Dissecting distally, identify the latissimus and teres major tendons, which are in intimate contact (Fig. 8-4). The latissimus tendon is

very thin as you follow its insertion into the axilla (Fig. 8-5). The teres major tendon can be transferred with the latissimus tendon if the subscapularis is intact and if the patient has extreme weakness. However, the teres major is a bulky muscle with poor excursion. It is difficult to pass both muscles under the acromion, and often the teres major will not reach the superior aspect of the humeral head.

The latissimus tendon should be released directly off of the bone to achieve greatest length. This is best done with electrocautery (Fig. 8-6). Palpation of the radial nerve is helpful at this stage to note its location and lessen the chance of nerve injury. Note that the latissimus tendon is tendinous, without muscular fibers at its humeral attachment site. In contrast, the teres major tendon is very short, and the musculotendinous unit is very muscular at its attachment on the humerus. Internal rotation of the arm helps to bring the insertion site into view and allows for easier dissection. However, when the arm is internally rotated, the axillary and radial nerves are brought closer to the latissimus dorsi insertion, and care must be taken with this exposure and arm positioning (Fig. 8-7).

Locking nonabsorbable sutures are placed into the distal end of the latissimus tendon, with four tails left distally for tying at the end. This step may be repeated for the distal end of the teres major tendon. Ensure the muscles are gliding freely with adequate excursion. The neurovascular pedicle of the latissimus enters anteriorly, about 10 to 15 cm from its musculotendinous junction.

Identify the teres minor, triceps, and deltoid. Pass a Kelly clamp from the superior incision, over teres minor, and underneath deltoid, to exit through the inferior incision, by blunt dissection. The sutures are placed into the clamp, and the latissimus tendon is pulled up into the superior wound, over the superior posterior aspect of the humeral head (Fig. 8-8). The teres major can be passed similarly. A rongeur is used to remove residual soft tissue from the posterior superior aspect of the greater tuberosity, which will receive the transferred tendon ends. Pass the sutures through bone with a free curved needle, and tie over a bony bridge or use suture anchors (Fig. 8-9). A strong repair is required since there will be a lot of tension on the repair. It is often helpful to tie the sutures over a small metal

or absorbable washer. Suture the border of subscapularis and rotator interval to the transferred tendon. The goal is to have a strong mechanical repair, and not necessarily a watertight repair.

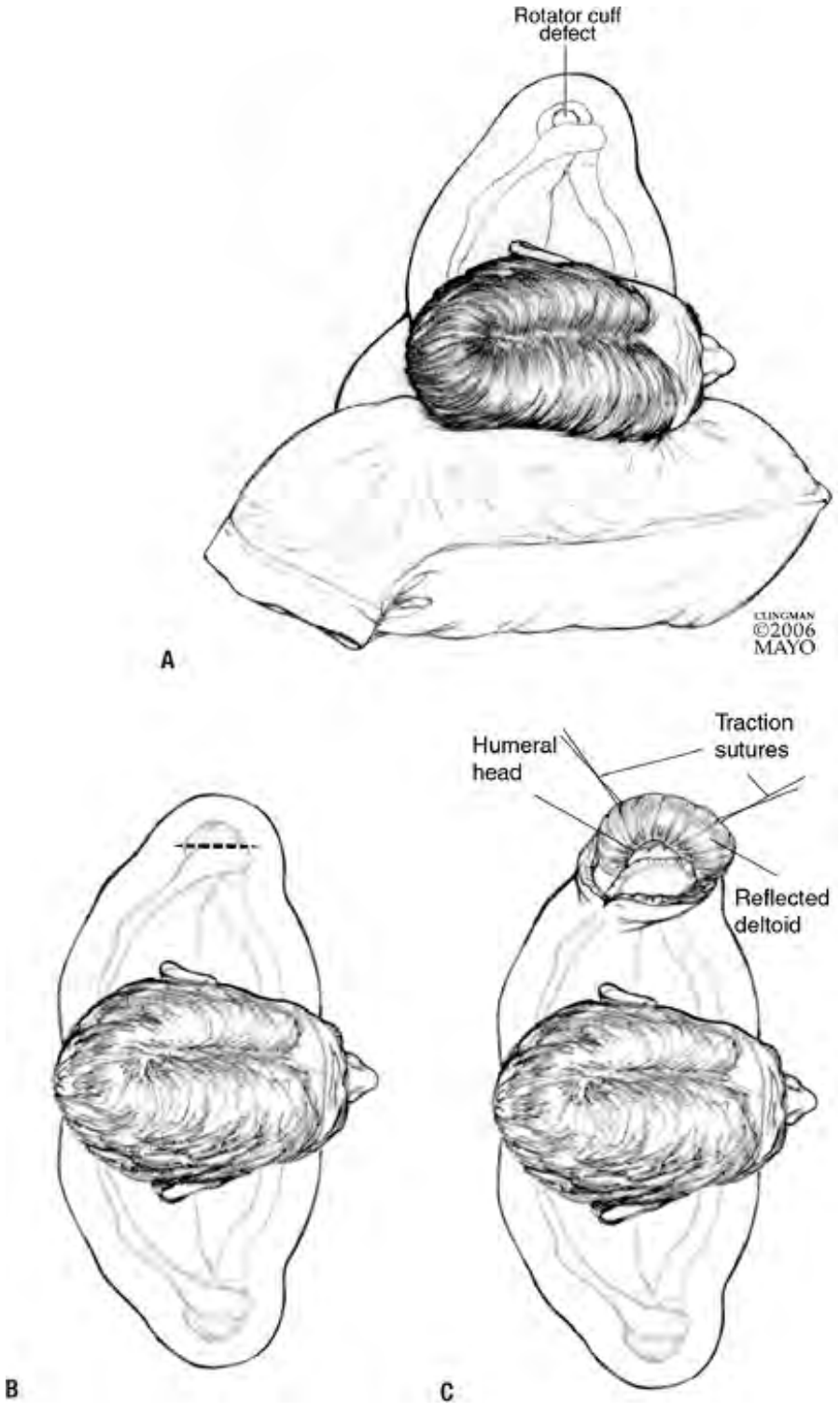

FIGURE 8-1 A: The patient is placed in the lateral decubitus position with all bony prominences well padded. B: The superior approach for rotator cuff repair

in Langer's lines is used. C: The deltoid is split in line with its fibers, or as an alternative if greater exposure is desired, the deltoid can be reflected from the acromion with a bone chip using an osteotome.

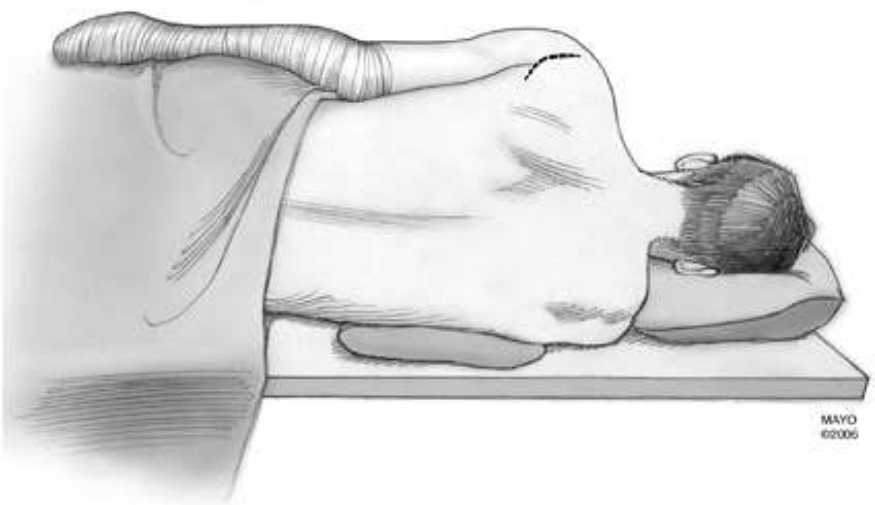

FIGURE 8-2 The second skin incision is placed along the posterior aspect of axilla, along the lateral border of the latissimus.

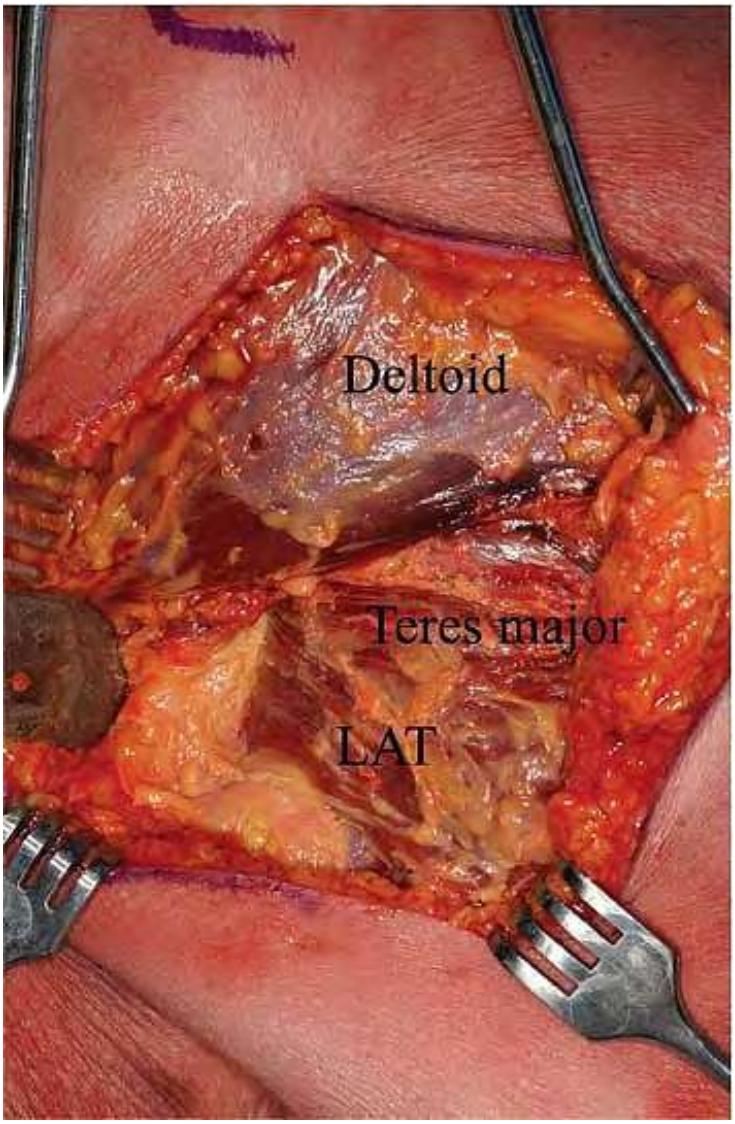

FIGURE 8-3 The latissimus and teres major muscles are visualized inferior to the deltoid muscle.

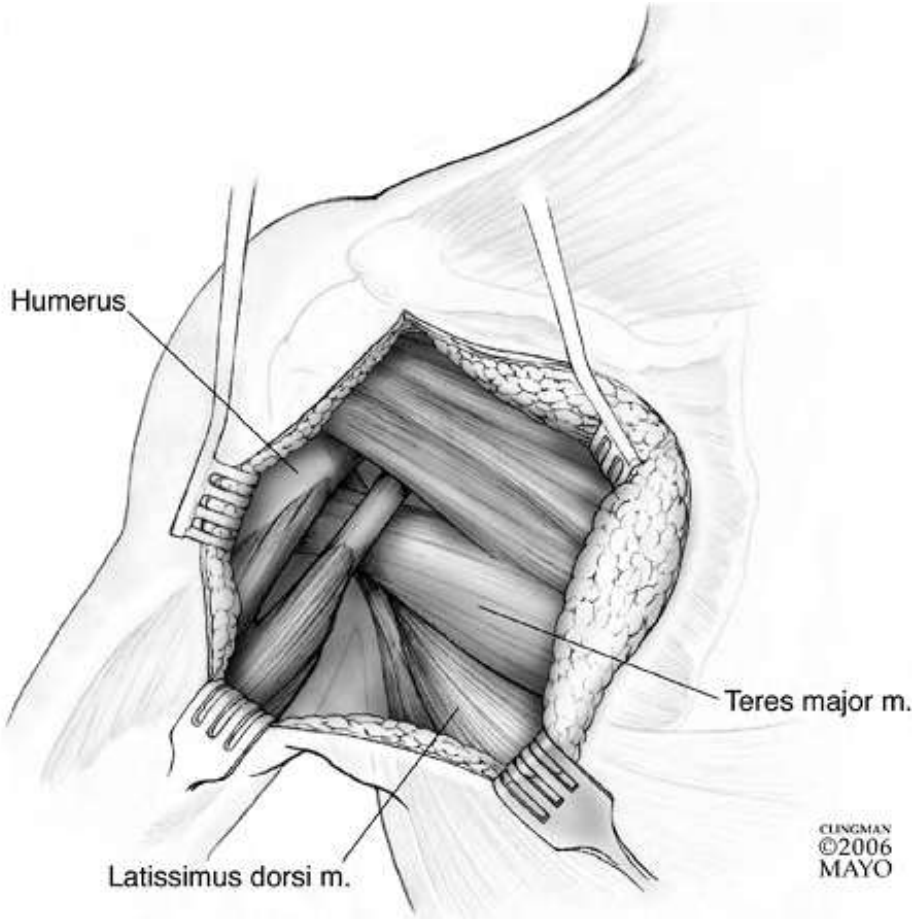

FIGURE 8-4 Dissecting distally, identify the latissimus and teres major tendons, which are in intimate contact. Note that the deltoid muscle has been removed for illustrative purposes.

FIGURE 8-5 The latissimus tendon is very thin, as you follow its insertion site into the axilla.

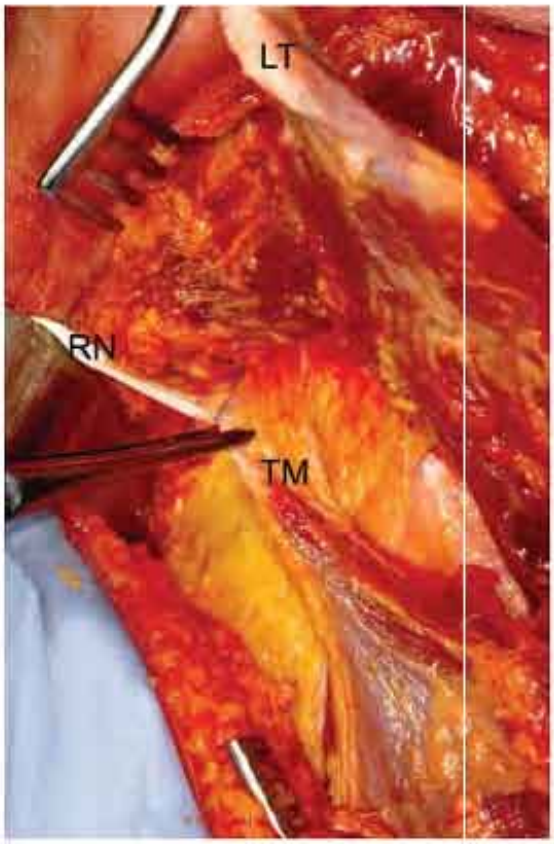
LT
RN
TM

A

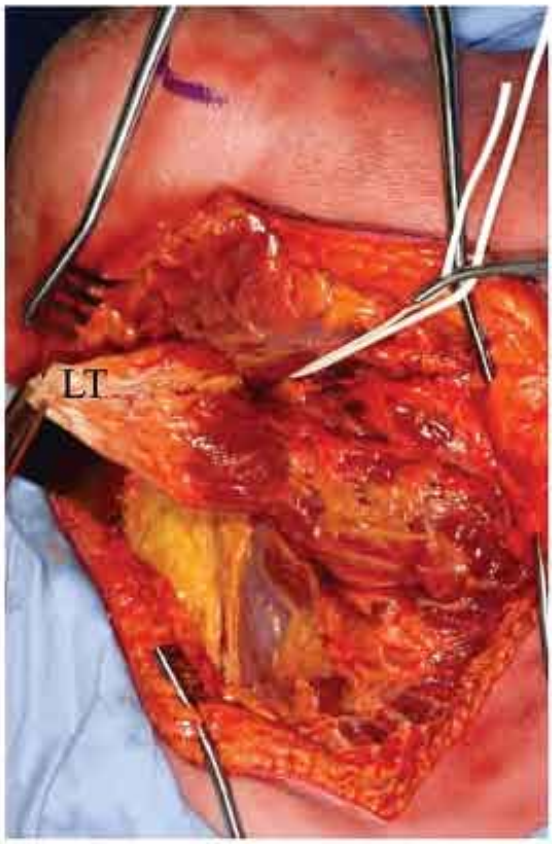
LT

B

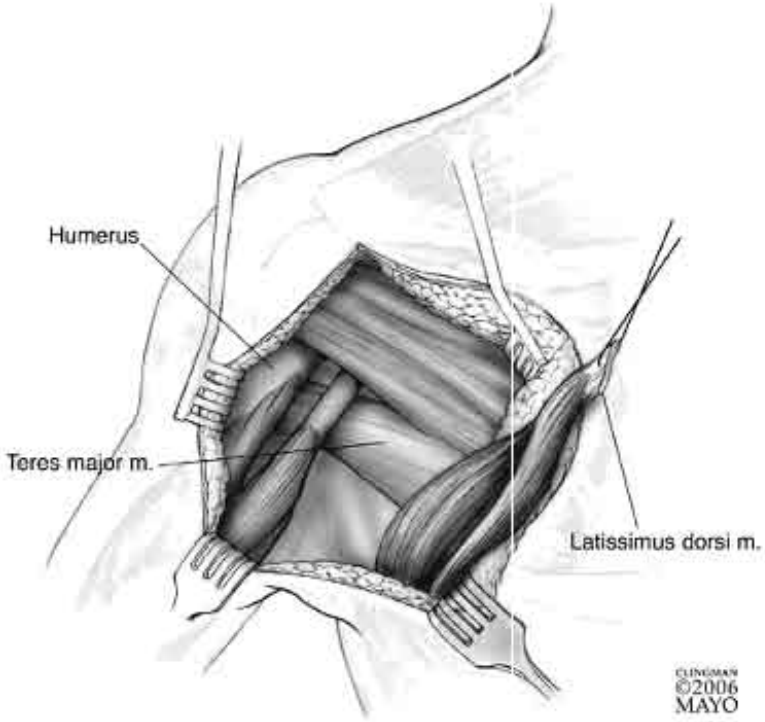
Humerus
Teres major m.
Latissimus dorsi m.
©2006
MAYO

C

FIGURE 8-6 A: The teres major tendon (TM), held with the forceps, can be transferred with the latissimus tendon if the subscapularis is intact and if the patient has extreme weakness. Note the radial nerve is identified in this photograph with the white vessel loop. B: The latissimus tendon should be released directly off of the bone to achieve greatest length. C: The latissimus tendon has been released. Note that the deltoid muscle has been removed for illustrative purposes.

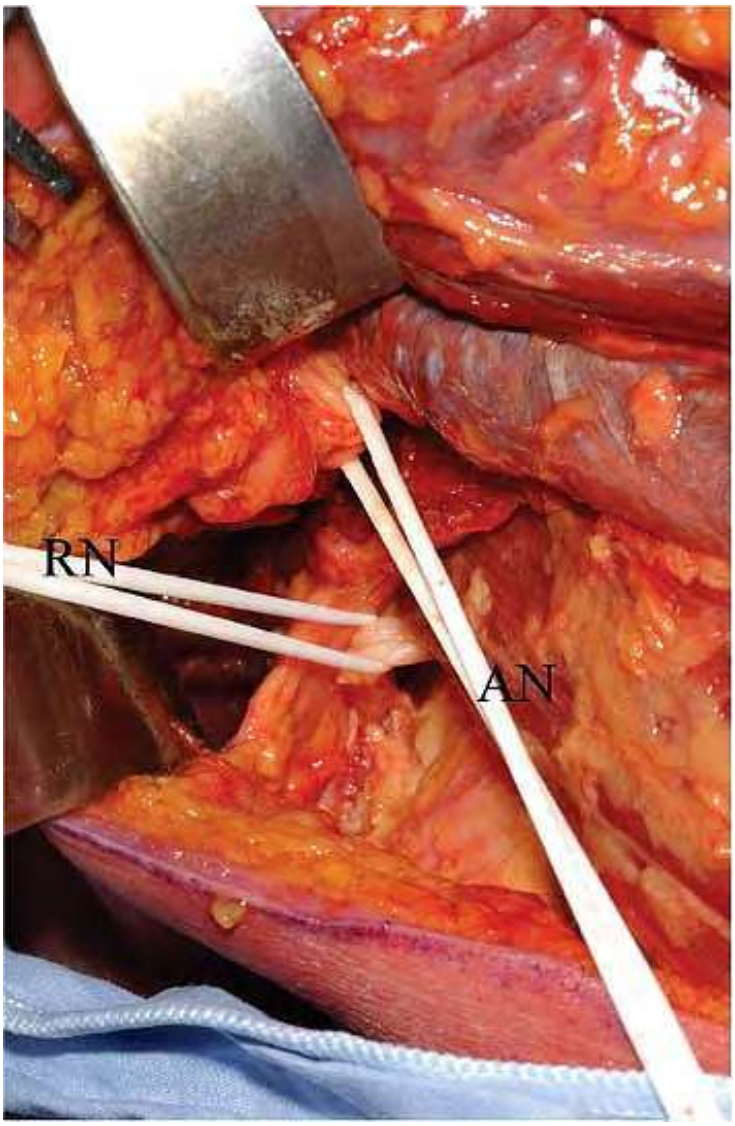

FIGURE 8-7 The axillary and radial nerves have been identified for illustrative purposes.

The deltoid with its bony attachment is repaired back to the acromion with nonabsorbable sutures. The wound is closed in layers over a drain.

Postoperative Management

The arm is placed into a pillow sling for 6 weeks to remove tension from the repair. During this period, passive flexion and external rotation are performed with the arm in the brace. Internal rotation is prohibited to prevent stress on the tendon repair. Specific training with a physical therapist is helpful to retrain the transferred musculotendinous unit to contract during active external rotation and forward flexion of the shoulder. The ability to gain voluntary control of the transferred latissimus is variable, and in many instances probably functions as an interposition. Many patients will be able to actively recruit the muscle during external rotation and forward flexion after 6 months.

Complications

Complications after latissimus dorsi transfer have been rare. Rupture of the transferred tendon has been reported. Ensuring that the tension of the repair is not due to under-excessive tension, protection of the repair with the splint, and avoiding active internal rotation for 6 weeks should prevent this complication. Refixation of the transferred tendon may be attempted.

Temporary axillary and radial nerve palsy has also been reported and perhaps avoided by careful surgical dissection and avoiding exuberant retraction of the deltoid fibers during exposure.

CLINGMAN
©2006
MAYO

A

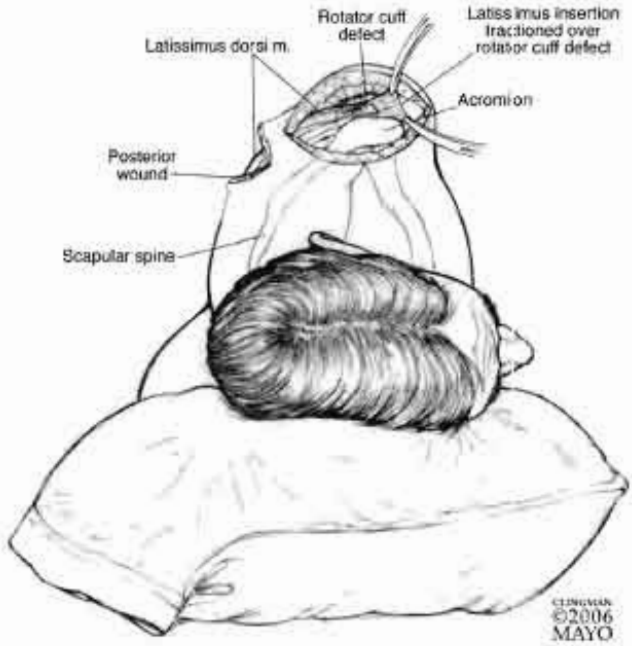
Rotator cuff defect
Latissimus insertion fractioned over rotator cuff defect
Latissimus dorsi m.
Acromion
Posterior wound
Scapular spine
CLINGMAN
©2006
MAYO

B

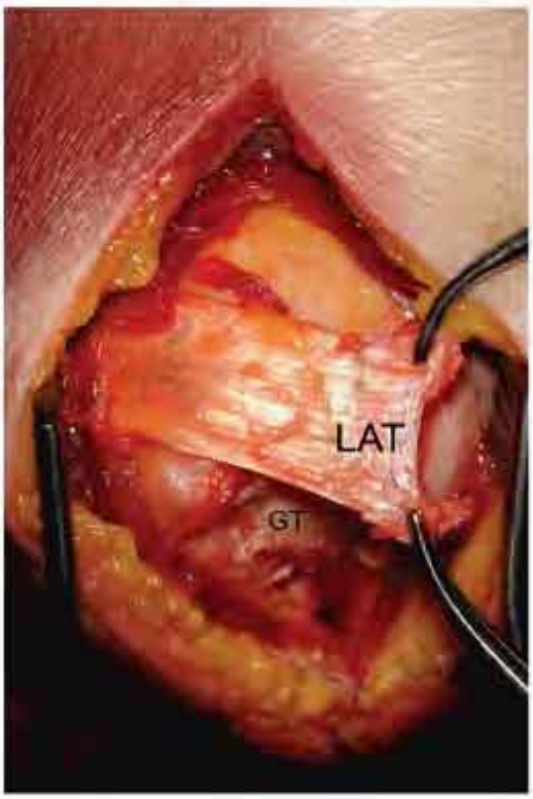
LAT
GT

C

FIGURE 8-8 Aâ€“C: The latissimus tendon has been passed under the posterior deltoid, up into the superior wound, and over the superior posterior aspect of the humeral head (HH). The teres major can be passed similarly. Note that the deltoid muscle has been removed for illustrative purposes.

FIGURE 8-9 The tendon is tied over a bony bridge at the greater tuberosity.

Pectoralis Major Tendon Transfer

Surgery

Standard beach-chair positioning is used for this surgery. A deltopectoral approach is made, and skin flaps are created (Fig. 8-10). The torn subscapularis musculotendinous unit is mobilized by releasing adhesions from the base of the coracoid, the brachial plexus, and the subscapularis fossa. A repair of the subscapularis is attempted by placement of modified Mason-Allen sutures in the remaining tendon and fascial tissue using nonabsorbable suture. The repair to the lesser tuberosity should be made with the arm in neutral rotation, if possible.

If repair is not possible, a pectoralis transfer is performed. Through the deltopectoral approach, the conjoint tendon, the tendon of the pectoralis major, and the anterior surface of the humeral head are exposed (Fig. 8-11). The tendon of the pectoralis major is exposed over its full length at the humerus, and the superior one-half to two-thirds of the tendon (depending on the size of the defect) is detached from the humerus (Fig. 8-12). The clavicular portion of the pectoralis major is taken for the transfer, and is bluntly dissected over a length of about 10 cm from the sternal portion of the pectoralis major, which remains intact. Modified Mason-Allen sutures are placed in the tendon using nonabsorbable suture (see Fig. 8-12).

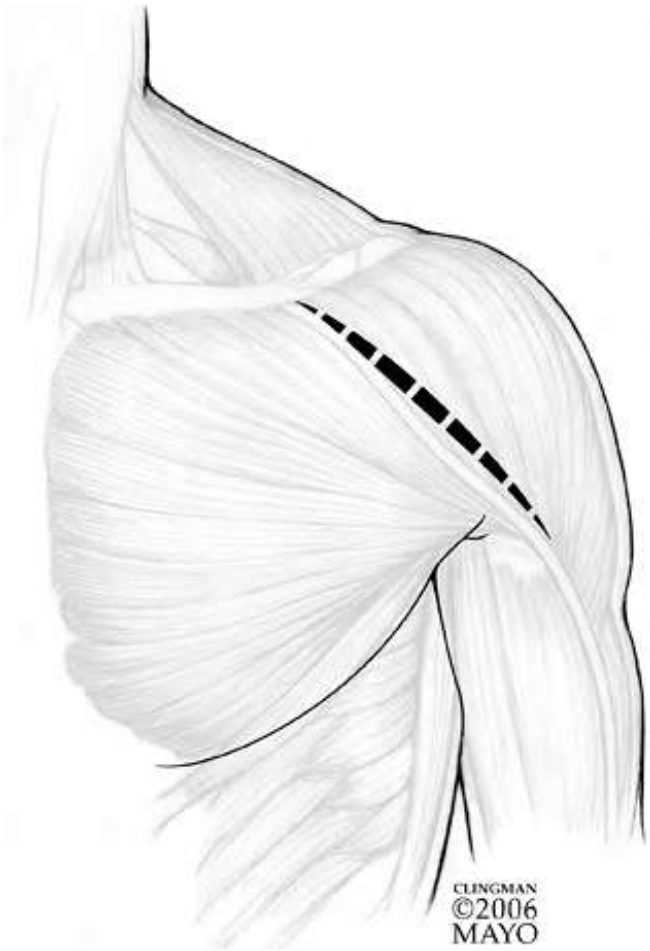

FIGURE 8-10 A deltopectoral incision is made.

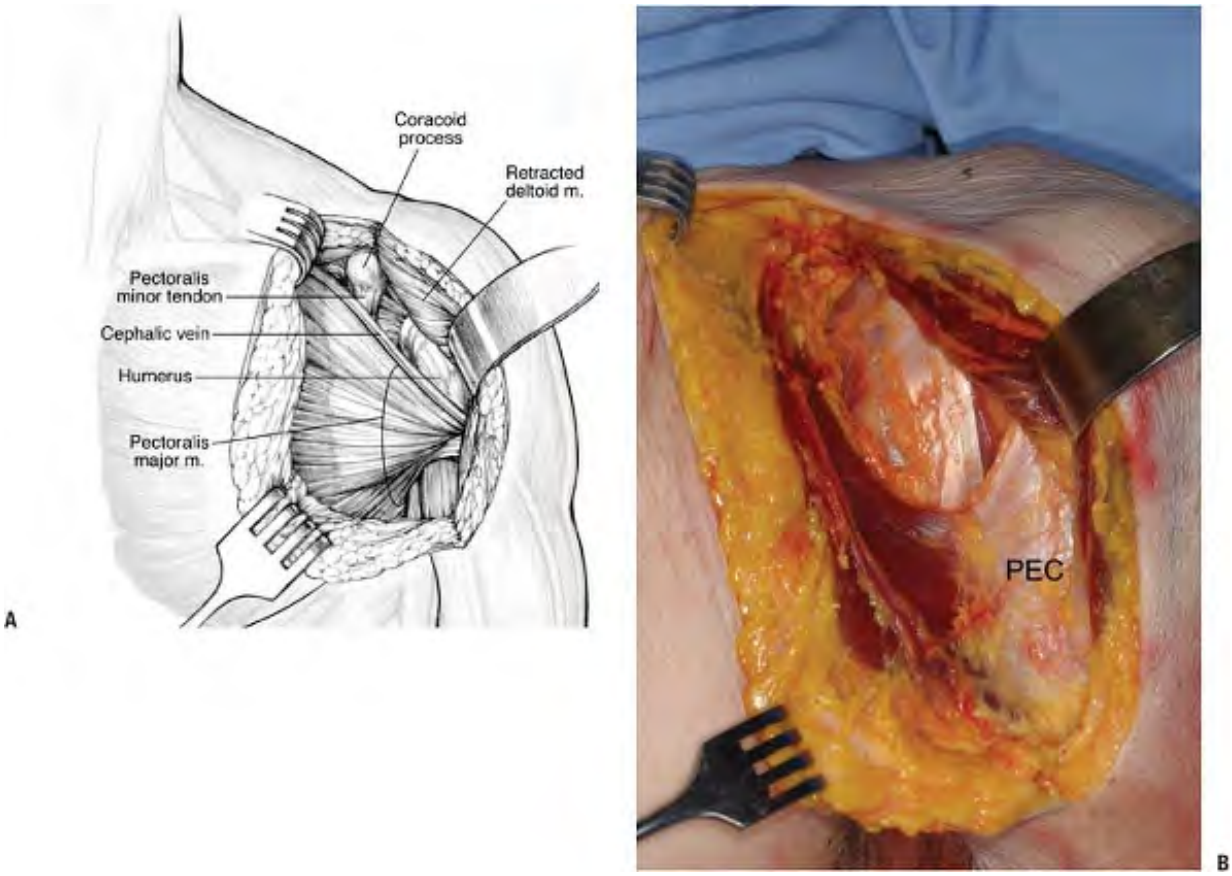

FIGURE 8-11 A,B: Through the deltopectoral approach, the conjoint tendon, the tendon of the pectoralis major, and the anterior surface of the humeral head are exposed.

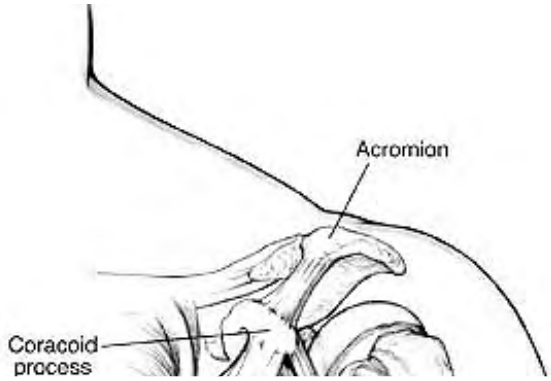

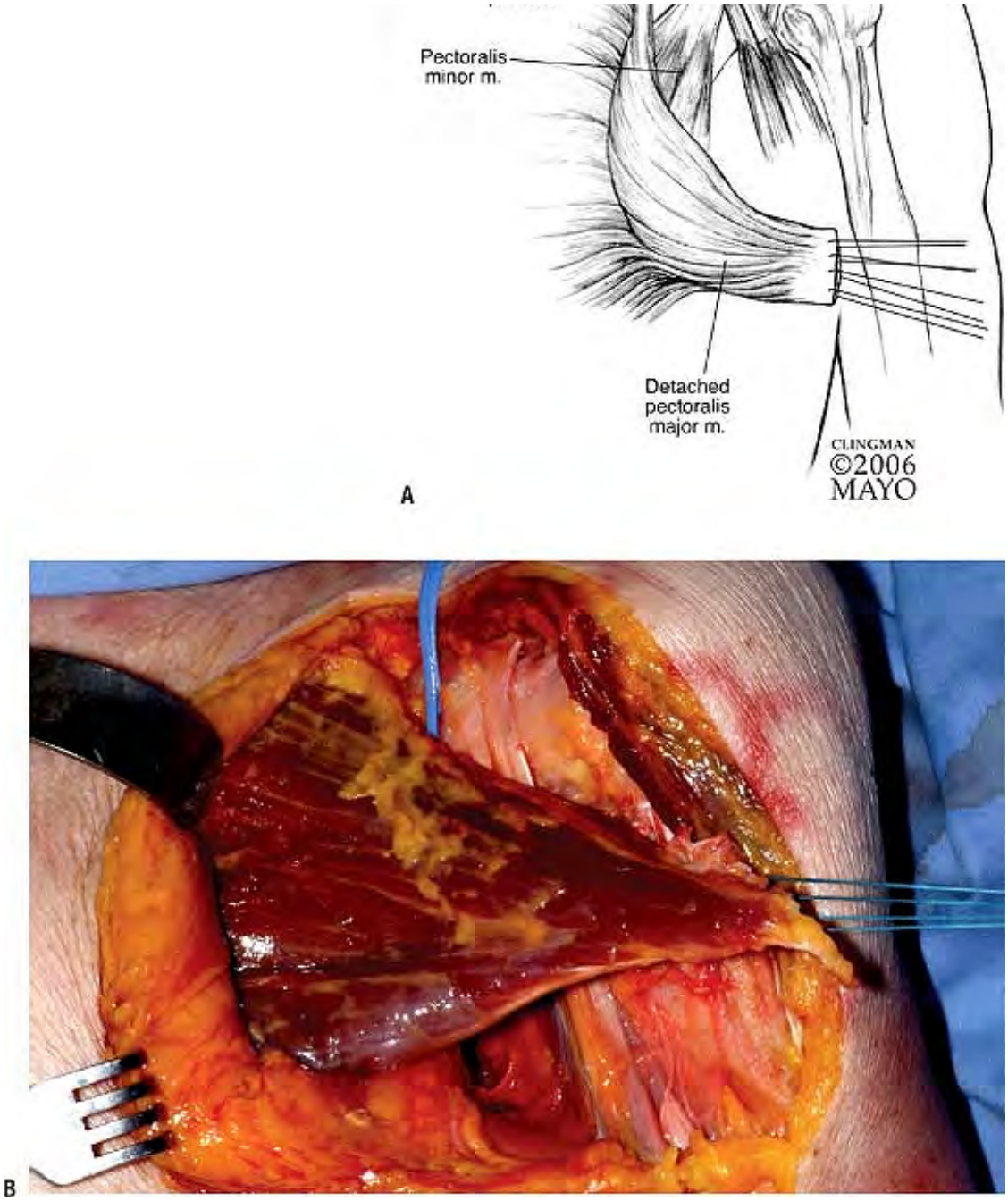

FIGURE 8-12 A: The tendon of the pectoralis major is exposed over its full length at the humerus, and the superior one-half to two-thirds of the tendon (depending on the size of the defect) is detached from the humerus. B: Locking sutures have been placed into the pectoralis major tendon.

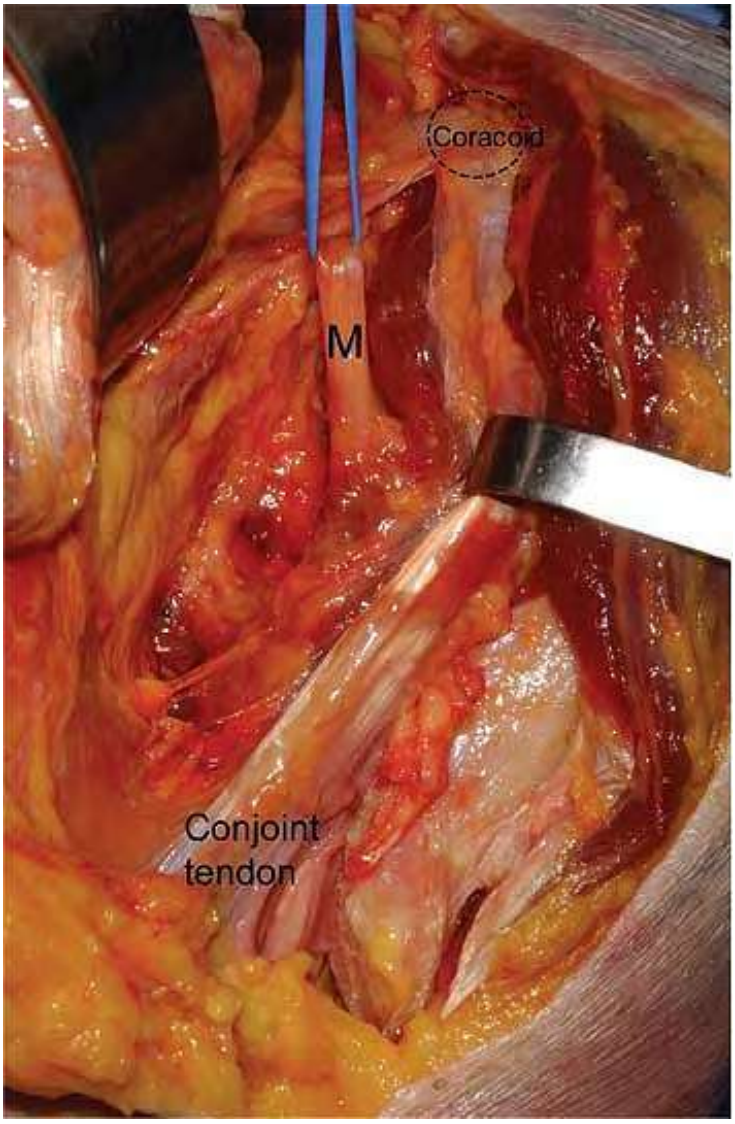

FIGURE 8-13 The musculocutaneous nerve has been identified with the blue vessel loop, between the conjoined tendon and the pectoralis major, about 5 cm distal to the coracoid.

The musculocutaneous nerve is identified by palpation, in the interval between the conjoined tendon and the pectoralis minor (Fig. 8-13). A reinsertion site for the pectoralis major tendon at the medial aspect of the lesser tuberosity is prepared with

a rongeur. Using curved forceps, the pectoralis major muscle is advanced behind the conjoined tendon in front of the musculocutaneous nerve, and the tendon is attached to the lesser tuberosity using transosseous repair, with the arm in neutral rotation (Fig. 8-14). The musculocutaneous nerve is palpated within the space between muscles to confirm that there is no tension on the nerve. If tension is found, the size of the muscle belly should be reduced. The supraspinatus is sutured to the proximal border of the pectoralis major tendon to close the rotator interval, if possible. In most patients, the tendon should be long enough to permit at least 30 degrees of external rotation of the arm after the tendon has been sutured to the lesser tuberosity. The conjoined tendon has been noted to arch forward slightly.

A proximal biceps tenodesis or release may be performed, depending on the presence of tendinopathy. Acromioplasty is not performed if the patient has an isolated subscapularis tear, or if the patient has a high likelihood of postoperative anterosuperior subluxation.

Postoperative Management

A sling is worn for 6 weeks postoperatively, and passive range-of-motion exercises are performed within the safe range of motion found intraoperatively starting on postoperative day 1. External rotation to neutral is allowed. Active range-of-motion exercises in all directions, including external rotation, are begun at 6 weeks postoperatively. Full loading is allowed at 12 weeks postoperatively.

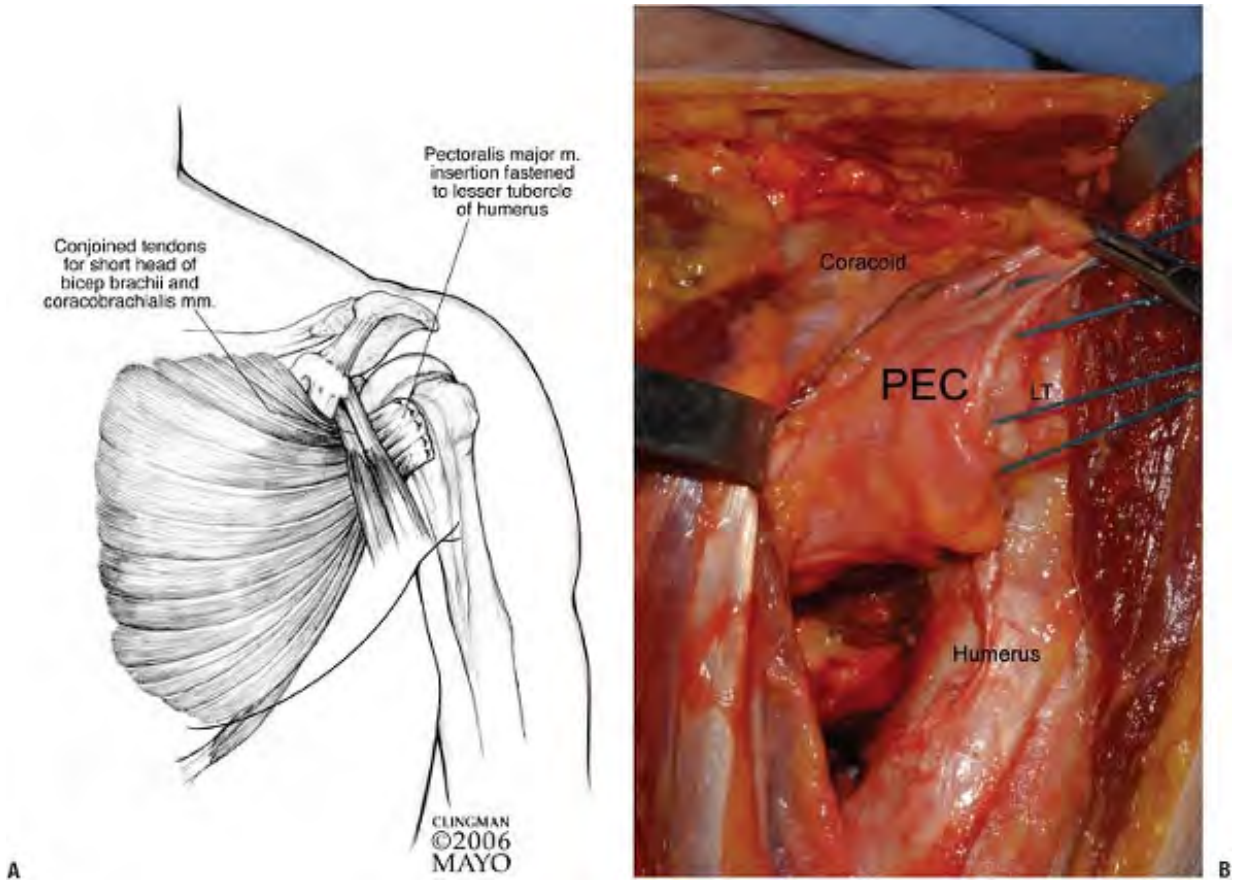

FIGURE 8-14 A: The pectoralis major muscle is advanced behind the conjoined tendon in front of the musculocutaneous nerve. For illustrative purposes, the inferior half of the pectoralis tendon is not shown. B: The tendon is attached to the lesser tuberosity using transosseous repair, with the arm in neutral rotation. For illustrative purposes, the inferior half of the pectoralis tendon is not shown.

Complications

Avulsion of the transferred pectoralis major tendon has been reported as a potential complication by Jost et al. This can be diagnosed clinically and by MRI, and has been treated with refixation of the transfer. Postoperative infection is another rare complication that should be treated with open irrigation, debridement, and a 6-week course of antibiotics. In very thin patients, a slightly more prominent anterior bulging

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 9

Surgical Exposure of the Elbow Following Bony and Soft Tissue Trauma

Emilie V. Cheung

Nathan A. Hoekzema

Scott P. Steinmann

Indications/Contraindications

The soft tissue surrounding the elbow is thin and pliable, yet durable enough to withstand constant flexion and extension. Optimal functional recovery in traumatic elbow injuries requires early and, in many cases, immediate motion. Skin slough, delayed healing, and dehiscence of skin flaps following treatment of elbow fractures can adversely affect elbow rehabilitation, delaying motion and increasing the risks of infection and nonunion. Properly planned incisions can allow for excellent exposure of the elbow while maximizing skin flap vascularity, minimizing the risks of postoperative wound complications.

Surgical exposures of the elbow typically involve a â€˜universalâ€™ posterior incision. This allows circumferential access by the creation of full-thickness skin flaps medially and laterally as needed. This allows for all intraoperative possibilities and avoids the creation of skin bridges if further surgery is needed in the future. Alternatively, separate medial and lateral skin incisions may be performed for isolated medial and lateral exposure. Prior surgical procedures should be carefully noted preoperatively.

Previous incisions may need to be incorporated in order to avoid creating narrow skin bridges which are susceptible to skin necrosis and potential wound complications.

Surgery

Patient Positioning

In our practice, the patient is routinely placed in the supine position with a small stack of towels placed underneath the ipsilateral scapula, and the arm draped across the chest after sterile preparation. The operating table should be slightly tilted away from the surgeon to help with visualization and exposure. This positioning is optimal for fixation of distal humeral fractures, radial head fractures, and medial or lateral ligament reconstruction. For medial ligament or coronoid fracture fixation, the supine position is used, but a sterile Mayo stand is utilized to position the arm more accurately for medial elbow exposure. Intraoperative fluoroscopy, if indicated, is placed on the ipsilateral side, and the elbow may be brought out laterally from the chest for the fluoroscope intermittently during the procedure. A sterile tourniquet is usually used in all elbow approaches. This allows for ease of removal if more proximal exposure of the humerus is needed. Hip rests or a bean bag may provide additional stability.

Posterior Approaches

Indications

- Triceps repair
- Combined medial and lateral approach
- Distal humerus ORIF
- Total Elbow Arthroplasty

Contraindications

Prior medial or lateral incisions.

Incision

The bony landmarks, including the olecranon process and the subcutaneous border of the proximal ulna, are marked (Figs. 9-1,9-2,9-3,9-4,9-5,9-6). The incision starts about 5 centimeters proximal to the olecranon process centered on the triceps tendon. It is then taken distally to either the lateral or medial side of the olecrenon, according to surgeon preference, and finishes distally following the subcutaneous border of the ulna. Full-thickness skin flaps are then developed, kept as thick as possible with the deep plane being the triceps fascia and epitenon proximally, and the forearm fascia and ulnar periosteum distally. If most of the surgical procedure will be on the medial side of the elbow, then the posterior incision should be placed to that side. Likewise, if a radial head fracture is being exposed, then the posterior incision should be made on the lateral side of the olecranon. If one chooses to err on the medial side of the olecranon, the incision should not be placed directly over the cubital tunnel, to avoid injuring the ulnar nerve.

Deep Dissection

The ulnar nerve is most easily found proximally between the medial intermuscular septum and the medial head of the triceps muscle. If the nerve is going to be transposed, it should be freed from proximal to distal in order to minimize damage to the nerve branches. Articular branches usually need to be sacrificed, but the first motor branch should be identified, mobilized,

and preserved. The leading edge, about 1 cm, of the distal medial intermuscular septum is removed to prevent tethering of the nerve after being anteriorly transposed. The nerve is then placed into a subcutaneous pocket anterior to the medial epicondyle. A suture is placed in the subcutaneous tissue and secured to the fascia to create a sling and prevent posterior subluxation of the nerve. If the nerve is not going to be transposed, care should be taken to avoid destabilizing the nerve in the cubital tunnel by leaving the soft tissue constraints of Osborne's fascia intact.

FIGURE 9-1 The skin is incised posteriorly with full thickness flaps raised as needed.

FIGURE 9-2 The ulnar nerve is carefully released from the cubital tunnel.

FIGURE 9-3 The medial and lateral sides of the triceps muscle are released in order to fully visualize the posterior fossa and the ulnohumeral articulation.

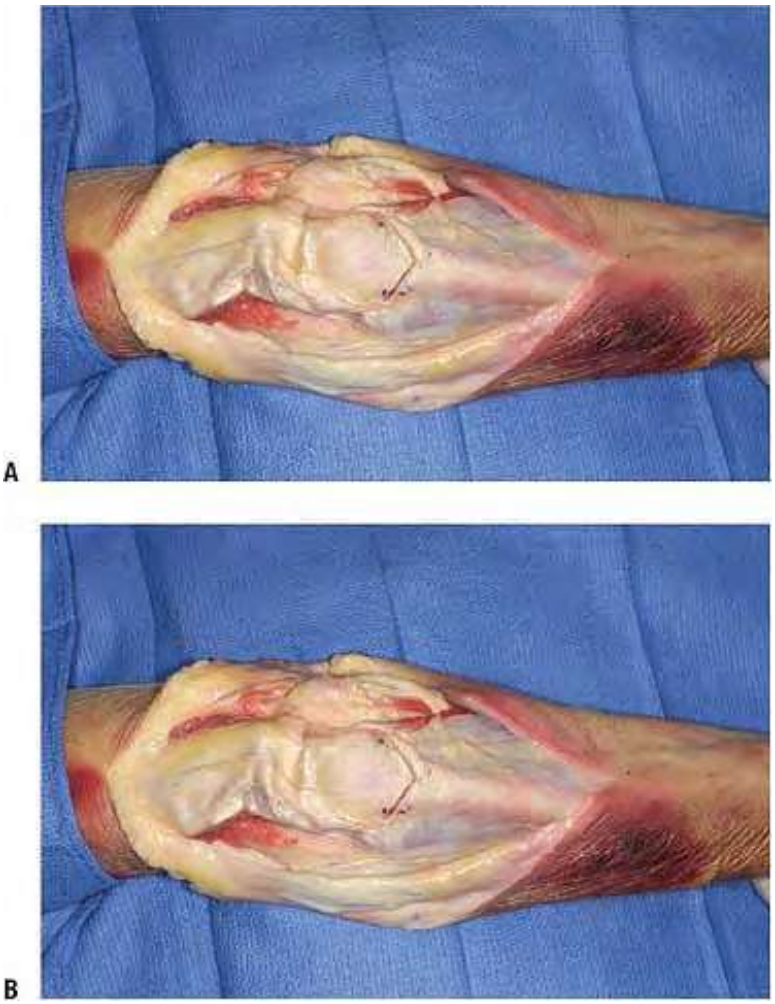

FIGURE 9-4 A,B: The olecranon osteotomy is best located at the small area of olecranon that is devoid of cartilage. A lap sponge may be threaded between the humerus and ulna to avoid inadvertent cartilage damage. The osteotomy is started with a reciprocating saw and then finished with an osteotome. Hardware may be predrilled to facilitate final osteotomy repair.

FIGURE 9-5 The osteotomied olecranon is retracted with the triceps tendon demonstrating near full visualization of the articular spool.

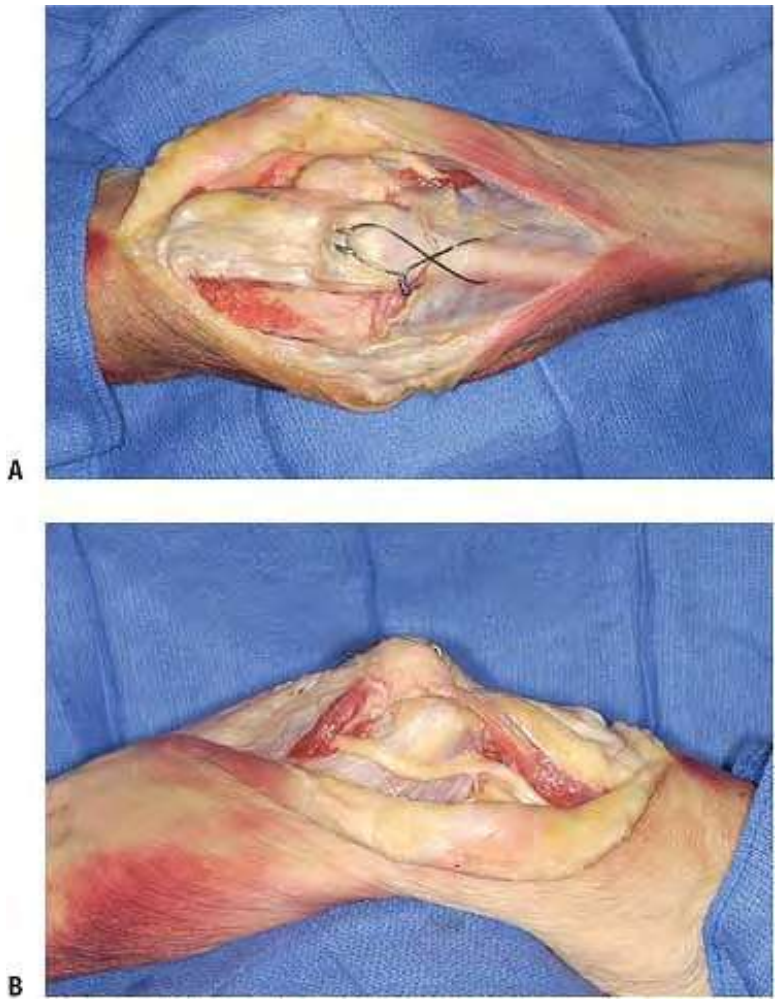

FIGURE 9-6 A,B: The completed tension band repair. The bent K-wires will then be sunk beneath the triceps tendon and the tendon repaired over them to minimize pin backout.

Olecranon Osteotomy

An olecranon osteotomy may be performed to gain full visualization of the articular surface of the distal humerus.

After identification and protection of the ulnar nerve, a capsulotomy is made at both the medial and lateral sides of the olecranon at the apex of the greater sigmoid notch to locate the olecranon â€œbare area.â€• This is the area of the olecranon where the

articular cartilage narrows and is the optimal location for the osteotomy. A lap sponge may be threaded under the olecrenon to protect the distal humeral articular cartilage from inadvertent damage. The osteotomy is created in a Chevron configuration, with the apex usually pointed distally using an oscillating saw. The saw is used for one-third of the bone cut, the second third is made with an osteotome, and the last third is cracked or fractured open with an osteotome to create the completed osteotomy. Once the osteotomy has been made, the anconeus needs to be taken off the lateral side of the olecranon fragment. Transection of the anconeus will denervate it. To avoid this problem, one technique is an anconeus flap transolecranon approach (AFT) which involves detaching the anconeus as a flap distally to proximally until the osteotomy site is reached. The osteotomy is then performed and the olecranon fragment and the anconeus are retracted proximally together to expose the distal humerus. Alternatively, the anconeus may be elevated from the proximal fragment of the olecranon osteotomy and elbow region with its distal attachment on the ulnar shaft left intact. The anconeus is then retracted laterally while the proximal fragment of the olecranon osteotomy is retracted proximally, which allows ample visualization of the distal humerus. The osteotomy site has been reported to heal reliably with few complications.

It is advantageous to place the hardware on the proposed olecranon fragment (if using a cannulated screw or Kirschner wires) prior to making the osteotomy so that the tract of the hardware is created anatomically, which will facilitate optimal alignment of the osteotomy site at the end of the procedure. We use either two Kirschner wires or a 6.5 or 7.6 mm cannulated screw and washer, with an 18 gauge tension band wire. The drill hole for placement of the wire is placed at an equal distance from the olecranon osteotomy as the distance of the osteotomy site to the tip of the olecranon. The osteotomy should be repaired by advancing the K-wires into the anterior ulnar cortex distal to the coronoid. After the K-wires have reached the anterior cortex, they are backed out about 5 mm and bent 180 degrees and tapped back until buried under the triceps tendon. The triceps may then be sutured over the wires to discourage backing out. Alternatively, a precontoured olecranon plate may be utilized for fixation.

Bryan-Morrey triceps reflecting approach

After identification and protection of the ulnar nerve, a periosteal elevator is used to

dissect the triceps muscle, including the medial and lateral margins of the triceps muscle, from the posterior humeral cortex. The triceps tendon is sharply dissected through Sharpey's fibers directly off of the olecranon starting medially and extending laterally. It is helpful to use a Beaver blade scalpel to sharply dissect the triceps tendon off of bone. The fascial overlying Kocher's interval is identified and longitudinally split. The triceps tendon and anconeus should be reflected medially to laterally, ending at Kocher's interval. The triceps may be removed with a thin wafer of bone as well, but this is not our standard practice. At the end of the procedure, the triceps tendon is repaired back to the olecranon using two transosseous drill holes placed in a cruciate configuration and one additional drill hole placed in a transverse orientation with nonabsorbable suture. The repair should be protected postoperatively by avoiding active elbow extension against resistance for a minimum of 6 weeks.

Van Goerder approach

After identification and protection of the ulnar nerve, the triceps tendon is identified several centimeters proximal to its insertion on the olecranon and a Chevron-shaped transection of the tendon is performed with the apex proximal. This is repaired at the end of the procedure with nonabsorbable suture. The repair should be protected postoperatively by avoiding active elbow extension against resistance, and avoiding passive stretching in positions of terminal elbow flexion for a minimum of 6 weeks.

Trap approach

The trans-anconeus pedicle flap (TRAP) approach requires a longer skin incision distally along the subcutaneous border of the ulna. The anconeus muscle is identified along the lateral aspect of the subcutaneous border of the ulna, and released by sub-periosteal dissection from its insertion on the ulna. The muscle is released in its entirety distally to proximally. It is reflected proximally to afford visualization of the distal humerus.

Medial Approach

Indications

- Ulnar nerve pathology

- Capsular release for stiffness
- Need to preserve the lateral ulno-humeral ligamentous complex

Contraindications

- Need for access to radial head or lateral ligaments

Incision

The medial exposure can use either the posterior or medial-posterior skin incision (Figs. 9-7 and 9-8). Once again, the skin incision can be tailored to work with need for exposure of other structures or existent scars from prior exposures.

Identification of the ulnar nerve is necessary for an adequate and safe medial exposure. The nerve should be identified proximally and subsequent dissection carried distally. In revision surgery and cases with a prior transposition, it is especially helpful to identify the nerve in a normal area prior to the dissection from scarred tissue. Once the nerve is safely identified, the medial supracondylar ridge is palpated along with the overlaying medial intermuscular septum. The medial antebrachial cutaneous nerve will be found on the fascia anterior to the septum. This should be protected to avoid postoperative neuroma formation.

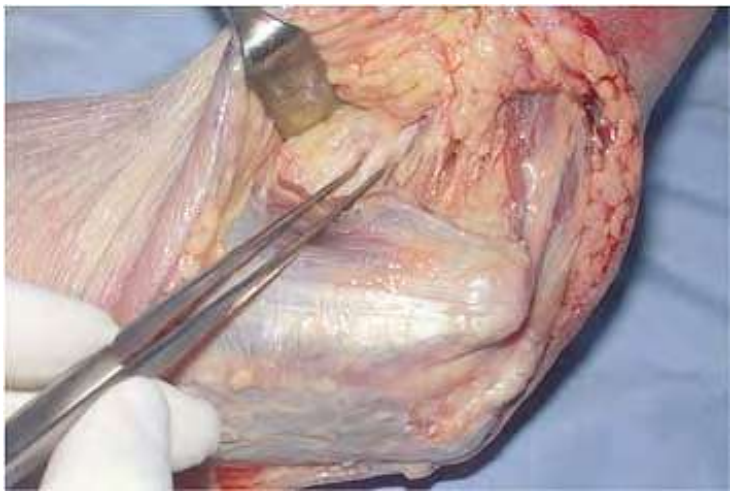

FIGURE 9-7 A skin incision is made and the subcutaneous fat is elevated from the forearm fascia. Care is taken to protect the medial antebrachial cutaneous nerve that lies in the subcutaneous adipose tissue.

The medial intermuscular septum is identified along with the medial supracondylar ridge. The brachial fascia is incised along the anterior aspect of the septum and the flexor-pronator group is released from the supracondylar ridge. Distally the flexor group is split leaving the posterior aspect of the flexor carpi ulnaris in place. The muscle group can be elevated off the anterior capsule and extend all the way across to the lateral aspect of the joint. A cuff of tissue may be left on the ridge, so that the muscle group can be repaired at the end of the procedure. The dissection is subperiosteal, deep to the brachialis so that the brachial artery and median nerve are protected.

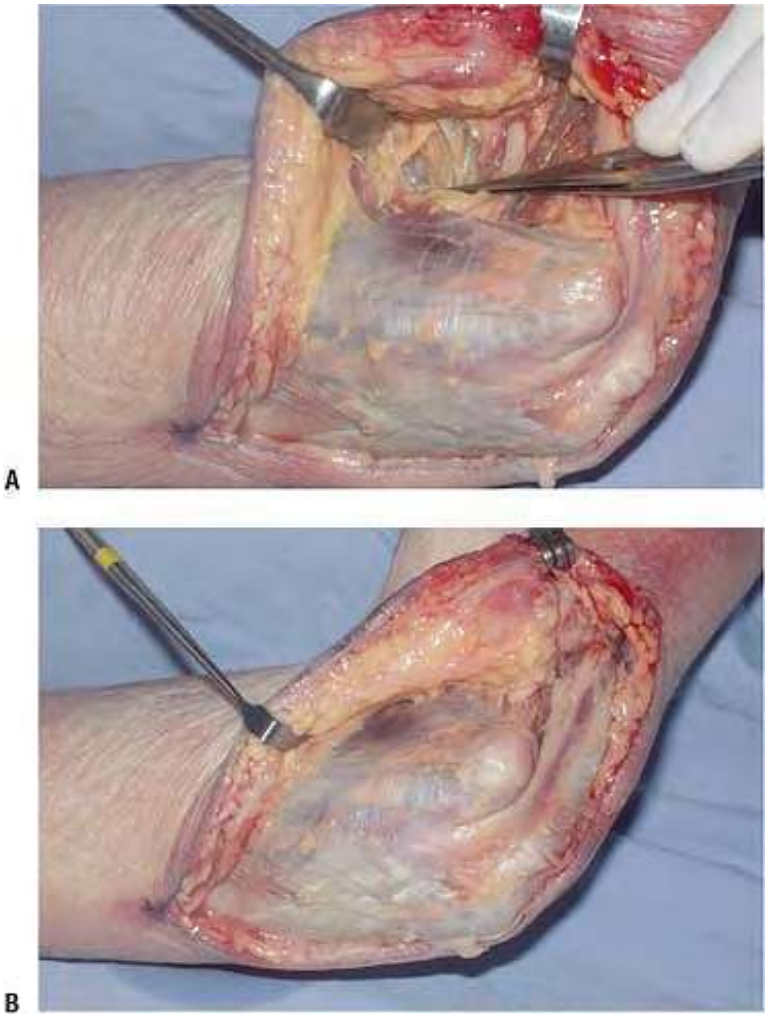

FIGURE 9-8 A,B: The median nerve is found lateral and deep to the flexor pronator group with the lateral antebrachial cutaneous nerve even more lateral.

Medial Coronoid Approach

Indications

- Fixation of coronoid fractures

Incision

The medial coronoid can be easily exposed through the floor of the cubital tunnel (Figs. 9-9,9-10,9-11,9-12,9-13). The posterior or medial approach is used to expose the ulnar nerve. The nerve is gently dissected and may be anteriorly transposed as previously described. It may be secured to a fasciocutaneous sling during fracture fixation to avoid excessive manipulation and inadvertent traction during the procedure. The nerve may be left in place if there are no concerns for neuropathy and as long as it is not destabilized. The two heads of the flexor carpi ulnaris (FCU) are split. The anterior half is retracted anteriorly, and the posterior half is retracted posteriorly in order to expose the coronoid.

P.122

P.123

The anterior band of the medial collateral ligament (MCL) is usually attached to a large anteromedial coronoid fragment. Care should be taken not to detatch the ligament from this fragment. If the capsular attachments are intact, the surgeon can judge fracture reduction based on realignment of the metaphyseal fracture fragments. The dissection of the flexor carpi ulnaris muscle fibers from the medial collateral ligament should start distally and the muscle should be brought proximally to avoid damaging the ligament and potentially destabilizing the elbow. The insertion of the MCL on the sublime tubercle should be evident. The coronoid will be in the deep portion of the wound, anterior to the ligament.

FIGURE 9-9 The ulnar nerve is released from the cubital tunnel in the standard fashion, taking care to split the FCU between the two heads. In this photograph, Osborne's ligament overlying the cubital tunnel has been released.

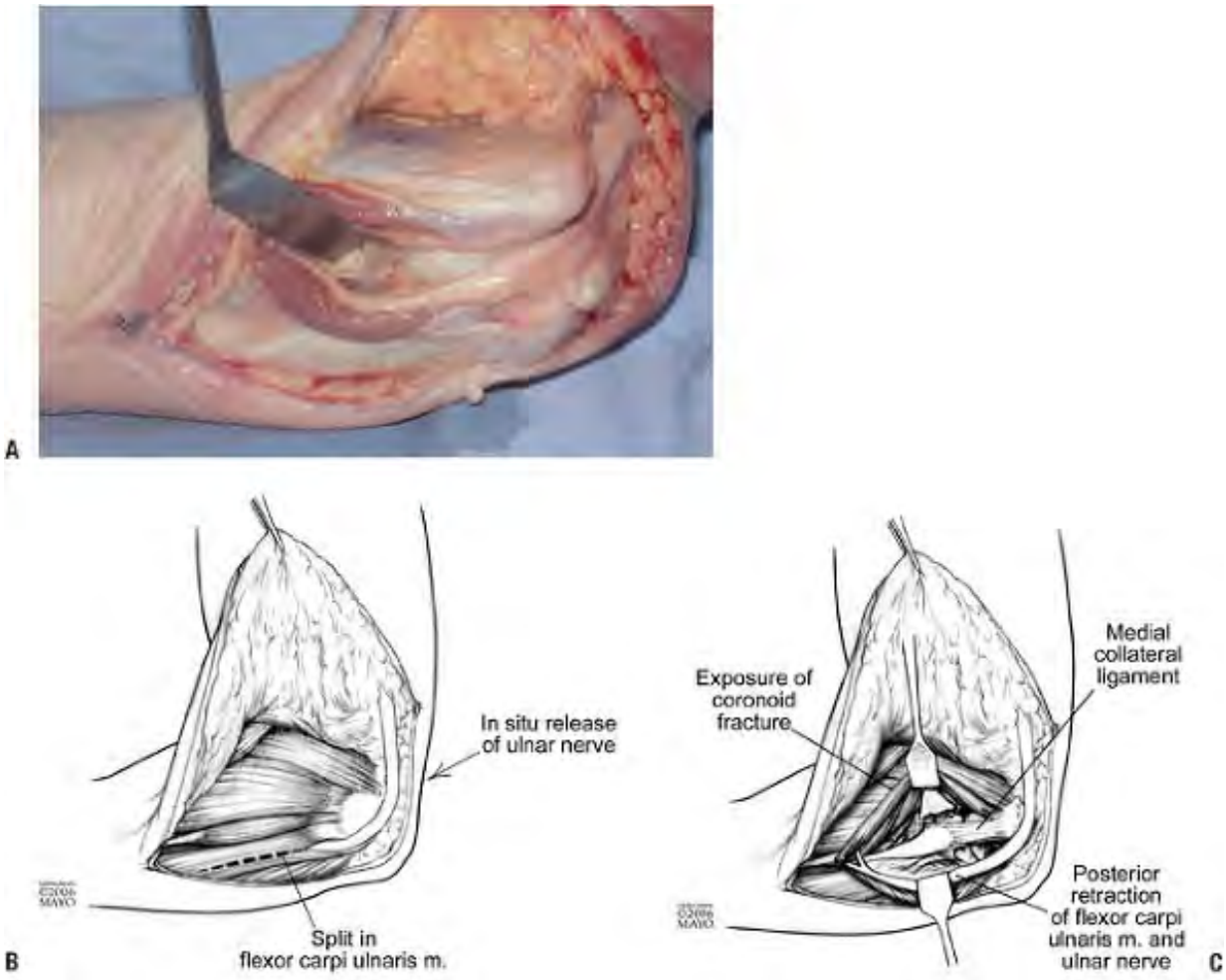

FIGURE 9-10 Aâ€“C: The nerve is gently retracted posteriorly and the ulnar head of the FCU is carefully dissected off of the medial collateral ligament (MCL), exposing the coronoid.

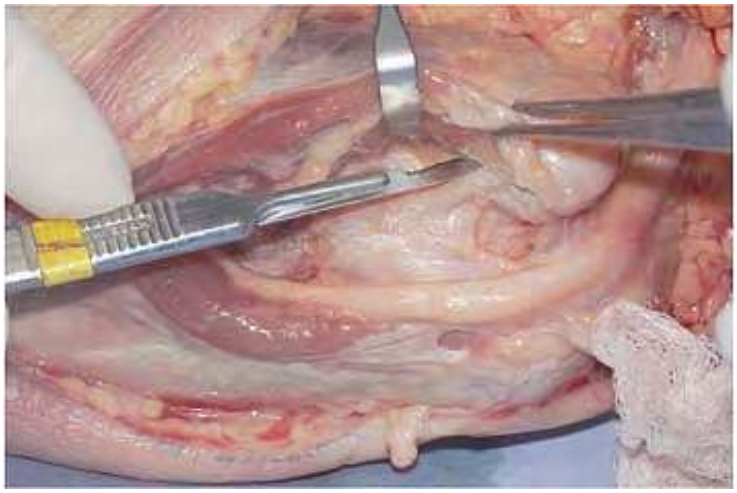

FIGURE 9-11 The medial collateral ligament is nearly fully visualized by peeling the humeral head of the FCU laterally and superiorly.

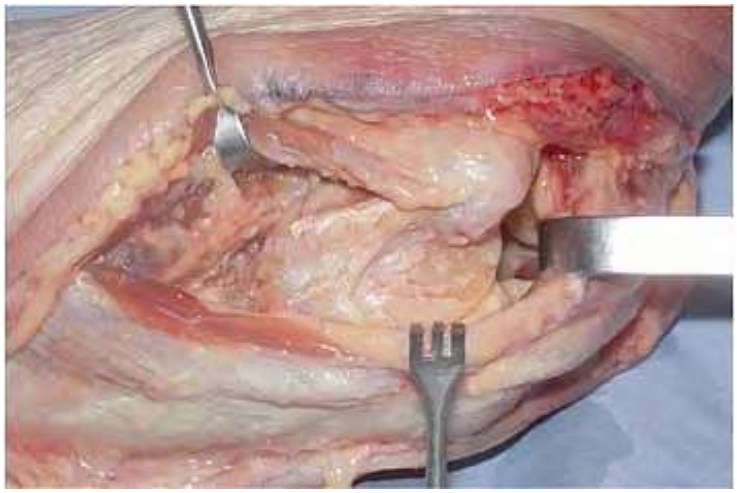

FIGURE 9-12 Utilizing the full extent of this exposure affords access to the coronoid, the medial collateral ligament, and even limited access to the posterior fossa.

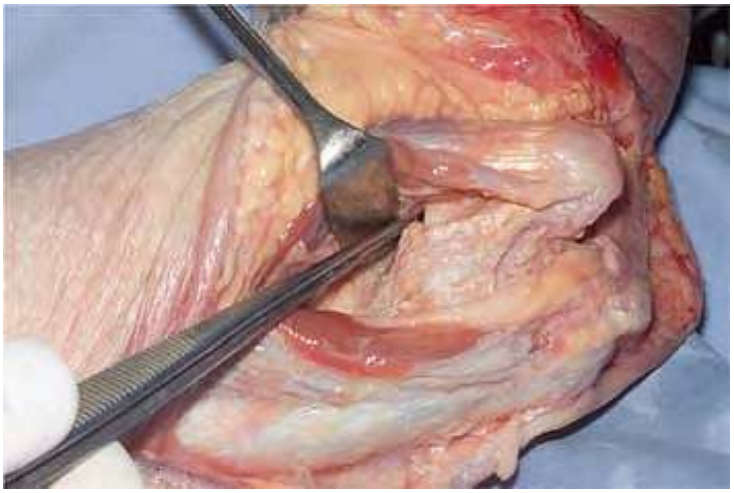

FIGURE 9-13 Entire coronoid may be seen. A full view of the MCL can be achieved including the medial epicondyle as well as the sublime tubercle.

Alternatively, if there is a large coronoid fragment, the FCU may be reflected anteriorly using subperiosteal dissection from the proximal ulna including the flexor-pronator mass proximally, as described by Taylor and Scham. Care must be taken to protect the ulnar nerve, which may need to be transposed when this approach is used.

Lateral Approaches

Indications

- Radial head ORIF, resection, or replacement
- Lateral epicondylitis debridement
- Repair of lateral ulnocollateral ligament
- Release of contractures

Contraindications

- The need to approach medial pathology
- Ulnar nerve involvement
- Medial humeral condyle fractures

Incision

The lateral approach to the elbow has become a standard means to gain access to the elbow joint (Figs. 9-14,9-15,9-16,9-17,9-18,9-19,9-20,9-21,9-22). The deep lateral approach was initially described by Kocher in 1911 and has been subsequently modified by Cohen, Hastings, and Morrey in its extensile exposure to preserve the lateral collateral annular ligament complex.

There are several different variations of the lateral exposure. The direct lateral, the LCL preserving, Kocher's approach, and the Mayo Modification have all been described. The general principle and anatomic intervals, however, are the same and are described here.

This approach to the radial head splits the lateral annular ligament complex and stays anterior to the lateral ulnar collateral ligament. The skin incision is placed either as a posterior incision with a large full-thickness lateral flap, or the incision is placed along the lateral epicondyle over the radial head to the lateral aspect of the ulna. Skin flaps are created over the antebrachial fascia.

Kocher's interval

The interval between the extensor carpi ulnaris (ECU) and the anconeus is identified. The fascia is incised from the lateral epicondyle distally following the junction of the ECU and anconeus. Care is taken to elevate the ECU anteriorly and the anconeus posteriorly. The capsule is incised along the anterior border of the lateral ulnocollateral ligament, about 1 cm above

the crista supinatorius. Care should be taken during this dissection in order to preserve the lateral collateral ligamentous complex, and avoid destabilizing the elbow.

FIGURE 9-14 A lateral or posterior skin incision may be used. The skin and subcutaneous tissues are carefully removed from the investing fascia. Special attention should be taken to avoid damage to the lateral antebrachial cutaneous nerve, which will travel within the fat at the distal aspect of this incision.

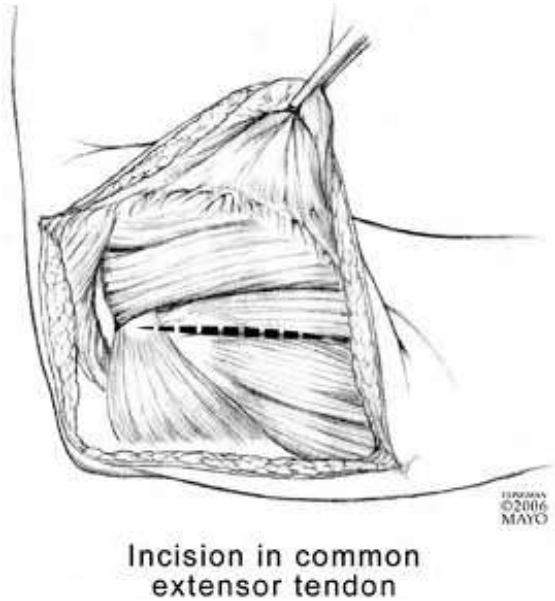

FIGURE 9-15 An alternative to the Kocher's approach is to split the common extensor group at the equator of the radiocapitellar joint. This decreases the likelihood of disrupting the lateral collateral complex.

FIGURE 9-16 A traditional Kocher incision is made between the anconeus and the extensor carpi ulnaris (ECU). Care is taken to develop this interval between the lateral collateral ligament complex (LCL) and the ECU.

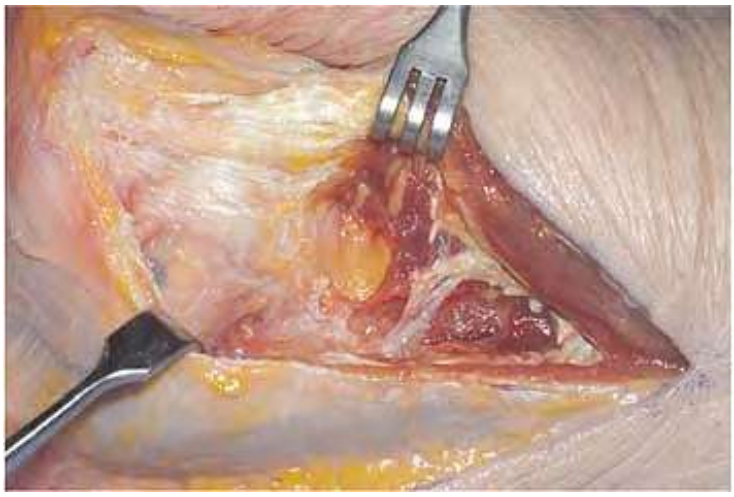

FIGURE 9-17 The extensor origin is dissected from the LCL.

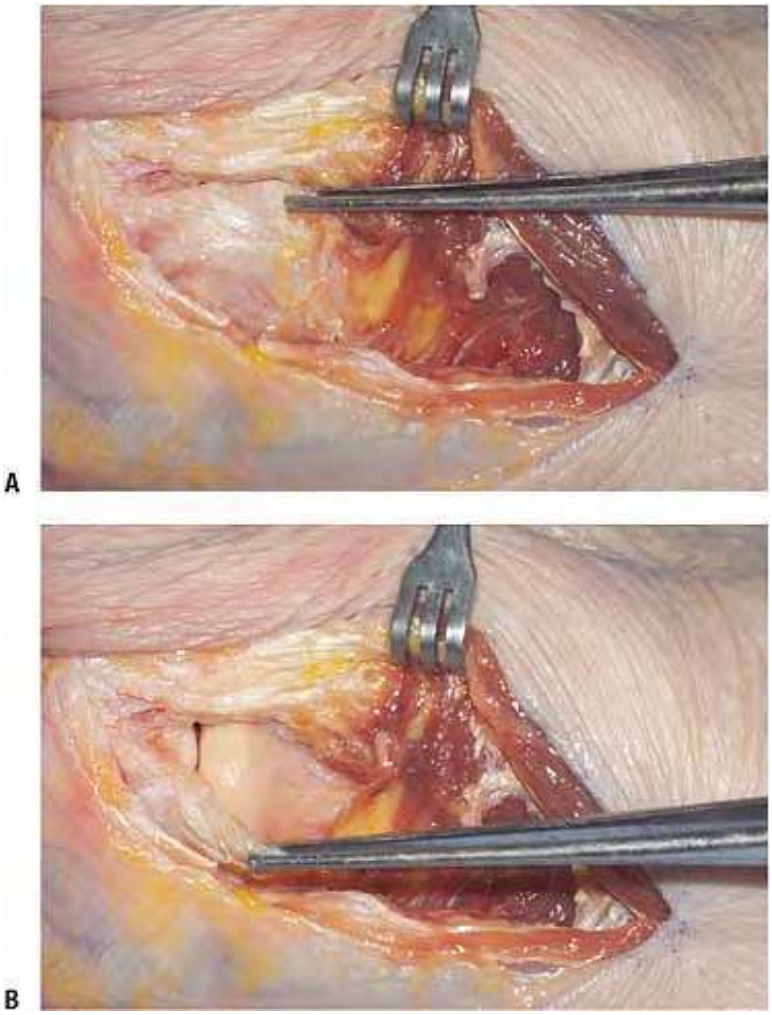

FIGURE 9-18 A,B: The capsule is incised anterior to the equator of the radial head in order to preserve the LCL complex.

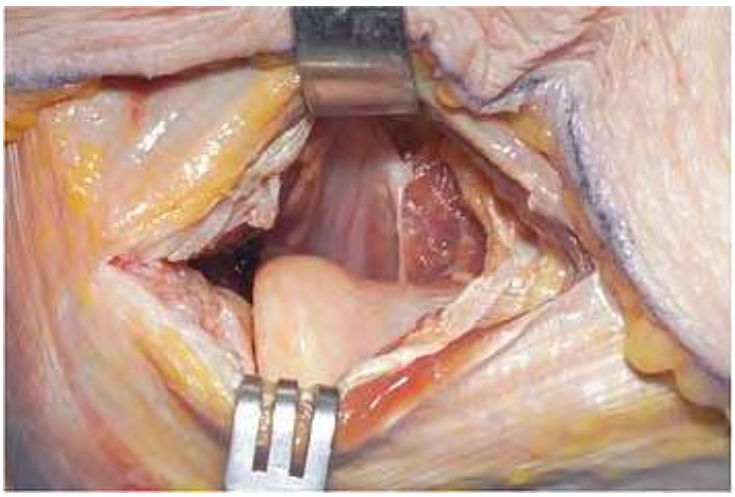

FIGURE 9-19 Excellent exposure of the radial head and neck is achieved. When retracting the anterior structures, pronating the forearm will reduce tension on the posterior interosseous nerve.

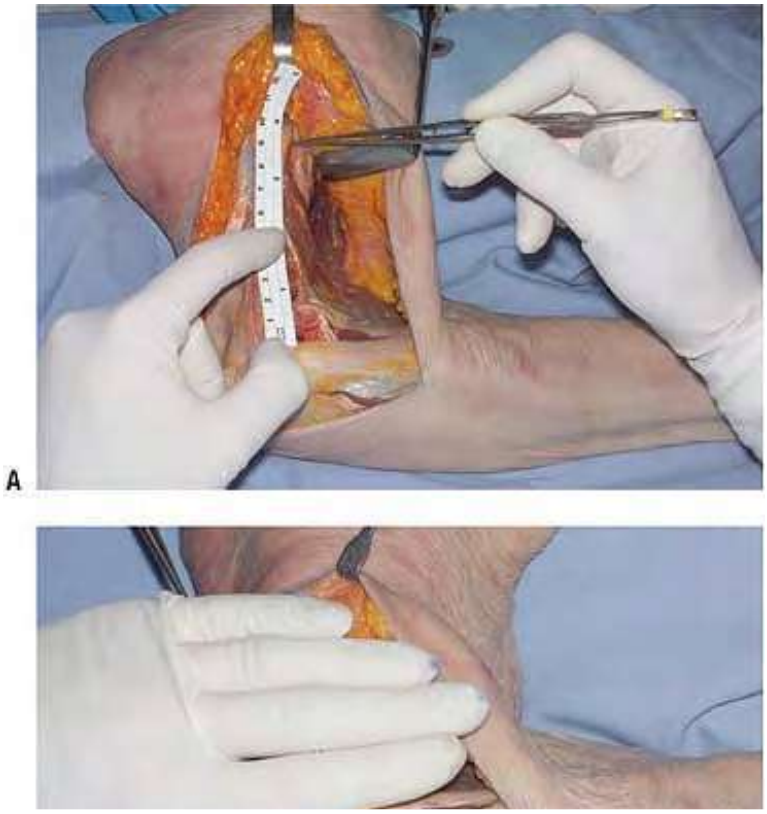

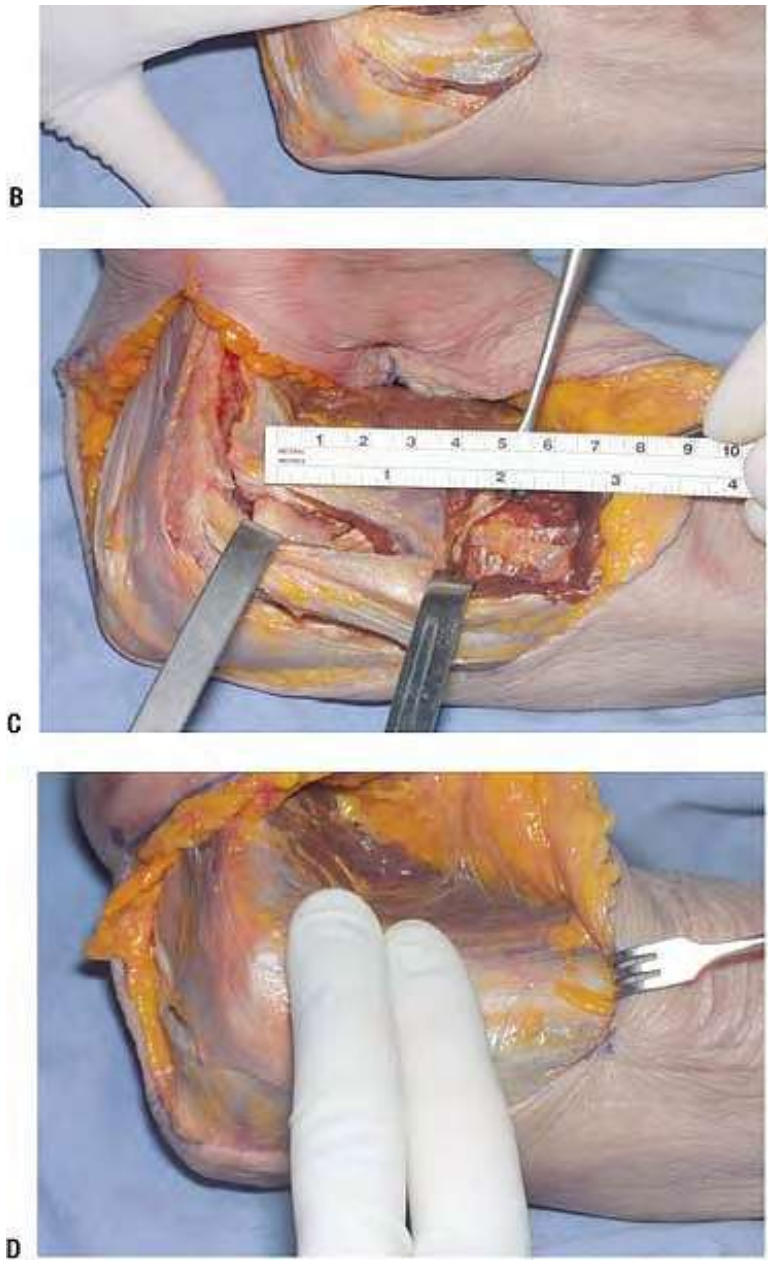

FIGURE 9-20 Aâ€“D: The radial nerve can be found proximally as close as 4 fingerbreadths (9 cm) proximally from the lateral epicondyle. The safe zone for the PIN is within 2 fingerbreadths (4 cm) distal of the radial head.

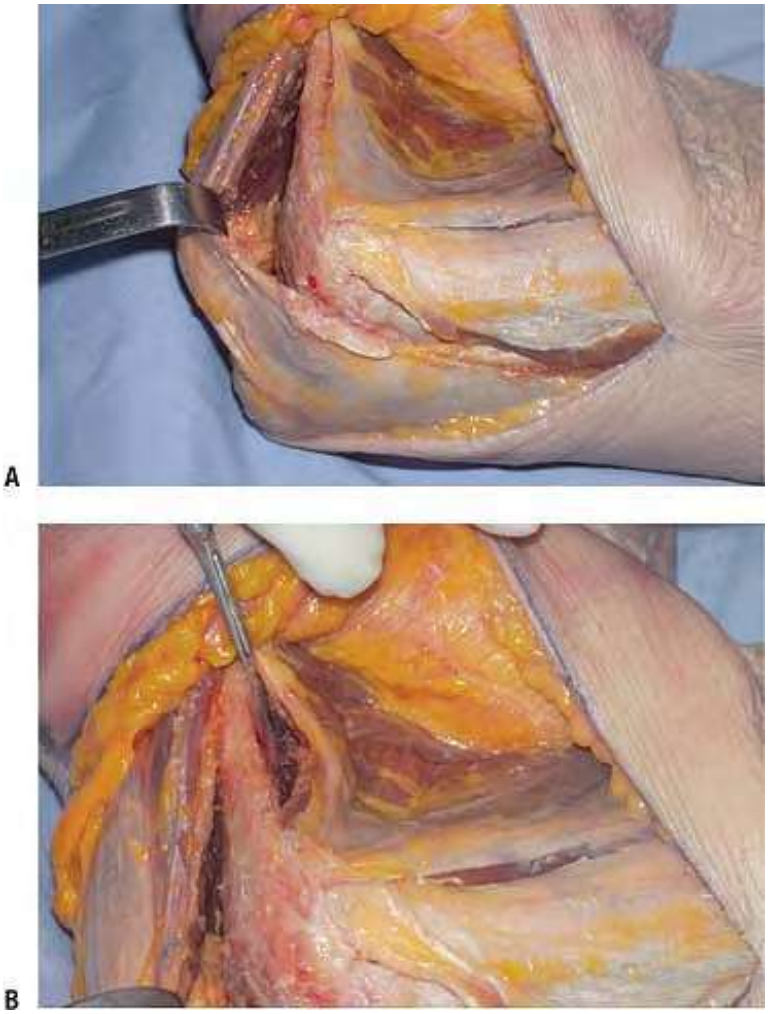

FIGURE 9-21 A,B: The lateral approach can be extended by continuing the dissection proximally on both the anterior and posterior side of the lateral humerus. Once again, one must be mindful of the LCL origin at all times.

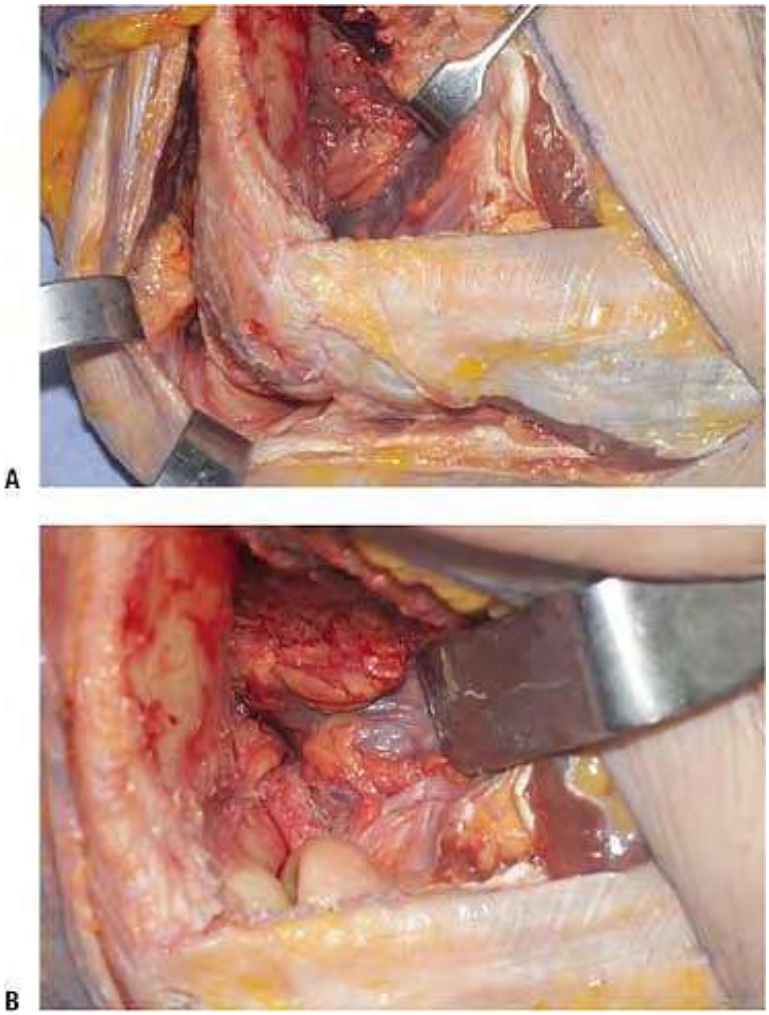

FIGURE 9-22 A,B: The brachioradialis is elevated anteriorly allowing for full visualization, even of the coronoid.

This exposure can then be extended both proximally and distally as in the case of a lateral ulnar collateral ligament reconstruction. The proximal dissection is achieved by elevating the common extensor tendon off the supracondylar ridge. The radial nerve is usually found proximally as close as 4 finger breadths (8 cm) above the lateral epicondyle, or on average (10 cm) in men from the articular surface. Distally, the posterior interosseus nerve will be found in the area of the radial neck about 2 finger breadths (4 cm) distal to the radiocapitellar joint.

Postoperative Management

Elbow rehabilitation is determined by fracture stability and concomitant ligamentous injury.

Complications

Flap necrosis, dehiscence, and wound infection are complications of any surgical incision surrounding the elbow. If the skin has been significantly traumatized prior to elbow exposure, skin slough and partial flap loss should be suspected and early wound examination should be performed to prevent wound infection. Many times partial skin loss may be treated with local wound care and dressing changes; however, if underlying vital structures are exposed, alternative means of wound closure must be considered. For exposed hardware, prosthesis, or nerve we favor aggressive wound debridement and early coverage with either the anconeus, a pedicled radial forearm fasciocutaneous flap, or latissimus flap.

Recommended Reading

Cohen MS, Hastings H. Posttraumatic contracture of the elbow. Operative release using a lateral collateral ligament sparing approach. *J Bone Joint Surg Br.* 1998;80(5):805â€“812.

Hotchkiss RN: Elbow contracture. In: Green DP, Hotchkiss RN, Pederson WC, eds. *Green's operative hand surgery*. Philadelphia: Churchill-Livingstone, 1999:667â€“682.

Oâ€™Driscoll SW. The triceps-reflecting anconeus pedicle (TRAP) approach for distal humeral fractures and nonunions. *Orthop Clin North Am*. 2000;31(1):91â€“101.

Ring D, Gulotta L, Chin K, et al. Olecranon osteotomy for exposure of fractures and nonunions of the distal humerus. *J Orthop Trauma*. 2004;18(7):446â€“449.

Ring D. Fractures of the coronoid process of the ulna. *J Hand Surg Am*.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 10

Radial Forearm Flap for Elbow Coverage

Kodi K. Azari

W. P. Andrew Lee

Indications/Contraindications

Soft tissue defects involving the posterior aspect of the elbow are not uncommon and can be challenging to manage. Tissue defects can be from trauma, burns, post-oncologic resection, pressure ulcers, extravasation injury, chronic bursitis, or chronic infection (4,7,15). In addition, elbow prosthetic devices can be exposed with devastating consequences. The goals of treating posterior elbow soft tissue defects is to provide wound closure, decrease the risk of infection, decrease edema, and allow the initiation of early rehabilitation (2,12). The soft tissue reconstruction must be aesthetically acceptable, durable, and elastic enough to allow for the constant unhindered movement of skin over the olecranon with elbow flexion and extension (7).

Many soft tissue reconstructive options exist and must be tailored to the needs of the patient's wound characteristics. Superficial wounds can be addressed by primary wound closure. Wounds with exposed â€œwhite structuresâ€• such as tendons, neurovascular structures, bone, and joint will necessitate flap coverage (9). Available flap options include regional muscle and musculocutaneous flaps, distant staged pedicle flaps, and microvascular free tissue transfer (8). Although these flaps are useful for elbow soft tissue coverage, they can carry significant morbidity. For example, regional muscle and musculocutaneous flaps necessitate the harvesting of a

functional muscle; distant pedicle flaps (such as the groin or thoracoepigastric flap) require the binding of the extremity to the flank for several weeks with significant discomfort and ensuing stiffness; and free flaps introduce the added complexity of microsurgery and prolonged surgical time (4,8). Because of the liabilities of the previously mentioned flaps, local fasciocutaneous flaps have gained popularity (10), of which the proximally based radial forearm flap is the recognized workhorse for elbow soft tissue coverage (8).

The radial forearm flap is composed of the skin, subcutaneous fat, and fascia that, if needed, can be elevated to include the entire volar surface of the forearm (13) (see Fig. 10-2). This is a reliable flap with a rich arterial supply from the septocutaneous perforator branches of the radial artery (see Fig. 10-3). The deep venous drainage for the flap is provided by the paired venae comitantes that run parallel with the radial artery and the superficial venous drainage is by branches of the cephalic vein. The radial forearm flap is versatile because it may be transposed either in the radial or ulnar direction and can be made sensate by encompassing the cutaneous nerves of the forearm. Furthermore, this flap can provide stable soft tissue coverage while still preserving elbow range of motion.

Contraindications to the use of the radial forearm flap include any injury to the radial artery, severe forearm soft tissue injuries, and inadequate collateral flow to the hand and thumb. A possible relative contraindication is recent cannulation of the superficial venous system of the upper extremity, as this may result in venous thrombus formation in the flap and subsequent venous congestion (5).

Preoperative Planning

Preparation of the Wound Bed

In cases of complex elbow wounds, it is extremely important to gain control over the wound before definitive soft tissue reconstruction. Osseous injuries and dislocations need to be stabilized with appropriate internal or external fixation and ligament reconstruction. When soft tissue injuries are present, it is mandatory to perform serial debridement of devitalized tissues until only viable tissues remains. Infections require appropriate debridement and culture-specific antibiotic coverage. It should be noted

that meeting the above requisites as quickly as possible will allow for earlier definitive vascularized wound coverage with radial forearm flap and thus afford a more successful functional outcome.

Assessment of the Arterial Vascularity of the Hand

It is imperative to assess the arterial vascularity of the hand by performing a modified Allen's test before performing the radial forearm flap. A modified Allen's test is performed by occluding the radial artery at the wrist crease and evaluating the perfusion of the digits using a hand-held pencil Doppler. In particular, one must ensure that there are Doppler signals to the thumb digital arteries once the radial artery is occluded. If the modified Allen's test shows evidence of vascular insufficiency, then one must be prepared to reconstruct the arterial tree using a saphenous vein graft or choose a soft tissue reconstruction technique other than the proximally based radial forearm flap.

Surgery

Patient Positioning

For optimal position, patients are placed supine on the operating room table. The entire upper extremity from axilla to hand is prepped and draped in standard surgical fashion, and a sterile proximal arm tourniquet applied.

Technique

The course of the radial artery is marked by drawing a line from the center of the anticubital fossa to the radial border of the proximal wrist crease (where the radial artery pulse is palpable) (Fig. 10-1). A skin island that is the shape and slightly larger than the periolecranon defect is designed and centered along the central axis of the radial artery (Fig. 10-2). The limit of the flap width is the radial and ulnar borders of the volar forearm and can extend from the antibrachial fossa to a few centimeters proximal to the distal wrist crease. The exact position of the flap is determined by a using a nonstretchable template with the antibrachial fossa as the pivot point of the vascular leash (Fig. 10-3).

Usually, the distal incision is made first and the radial artery and accompanying venae

comitantes that reside between the flexor carpii radialis (FCR) and brachioradialis (BR) tendons are isolated and divided. The cephalic vein, which lies radial to the brachioradialis tendon, is also isolated and divided.

FIGURE 10-1 The course of the radial artery and venae comitantes is marked by drawing a line from the center of the anticubital fossa to the radial border of the proximal wrist crease.

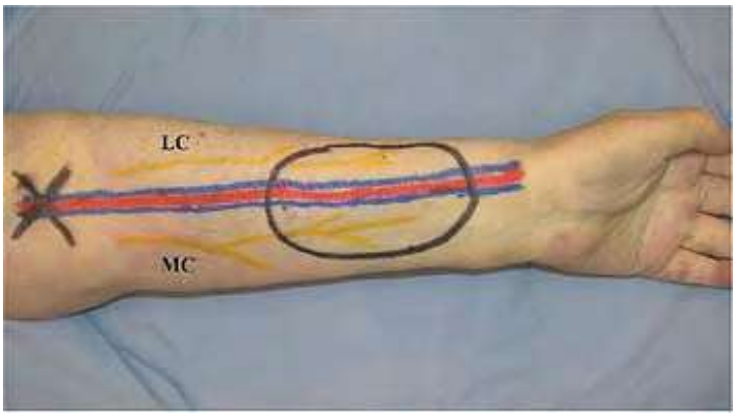

FIGURE 10-2 A skin island that is the shape and slightly larger than the periolecranon defect is designed and centered along the central axis of the radial

artery. The medial cutaneous (MC) or lateral cutaneous (LC) nerves of the forearm can be included in the skin paddle to form an innervated neurosensory flap.

The proximal incision of the flap is made and the radial artery is identified between the FCR and BR. Care should be taken not to injure the cephalic vein (or the medial or lateral cutaneous nerves if the radial forearm flap is designed as an innervated neurosensory flap) (see Fig. 10-2). Next, the ulnar incision is made and carried to the tissue plane deep to the deep forearm fascia. This subfascial plane is developed and the flap is raised from ulnar to radial off the underlying flexor carpii ulnaris, flexor digitorum sublimus, palmaris longus, and flexor carpii radialis muscles. It is imperative to preserve the paratenon of the tendons to allow skin graft coverage of the flap donor site. The dissection is carried ulnarly until the radial border of the FCR. At this important juncture the intermuscular septum carrying the septocutaneous branches of the radial artery is identified and carefully preserved (Fig. 10-4).

The radial border incision is made and the subfascial dissection is carried ulnarly. The brachioradialis muscle is retracted laterally allowing for the identification of the radial artery, intermuscular septum, and superficial radial sensory nerve. In the upper border of the flap, the cephalic vein is located and integrated in the flap for added venous drainage. The flap is next elevated and the radial artery, venae comitantes, cephalic vein, and medial or lateral cutaneous nerves (for an innervated sensory flap) are dissected proximally to an appropriate pivot point that allows for a tension free flap transfer (Fig. 10-5).

The tourniquet is released and the vascularity of flap is ascertained. In cases where large flaps are required, the perfusion of the peripheral flap skin can be evaluated by the administration of intravenous fluorescein. Initially a small test dose of 100 mg is given, and the patient's vital signs are closely monitored for 20 minutes. If there is no evidence of hemodynamic instability, then the full dose of 10 mg/kg is administered and the skin paddle evaluated 10 minutes later under ultraviolet light (Wood's lamp). The extreme areas of the flap that appear dark and do not fluoresce are sharply trimmed.

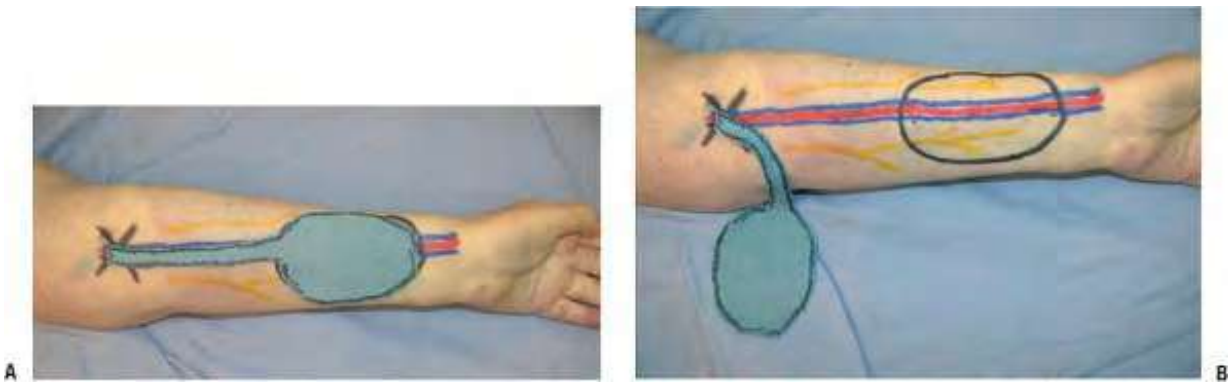

FIGURE 10-3 A,B: The exact position of the skin paddle is established by a using a nonstretchable template with the antibrachial fossa as the pivot point of the vascular leash.

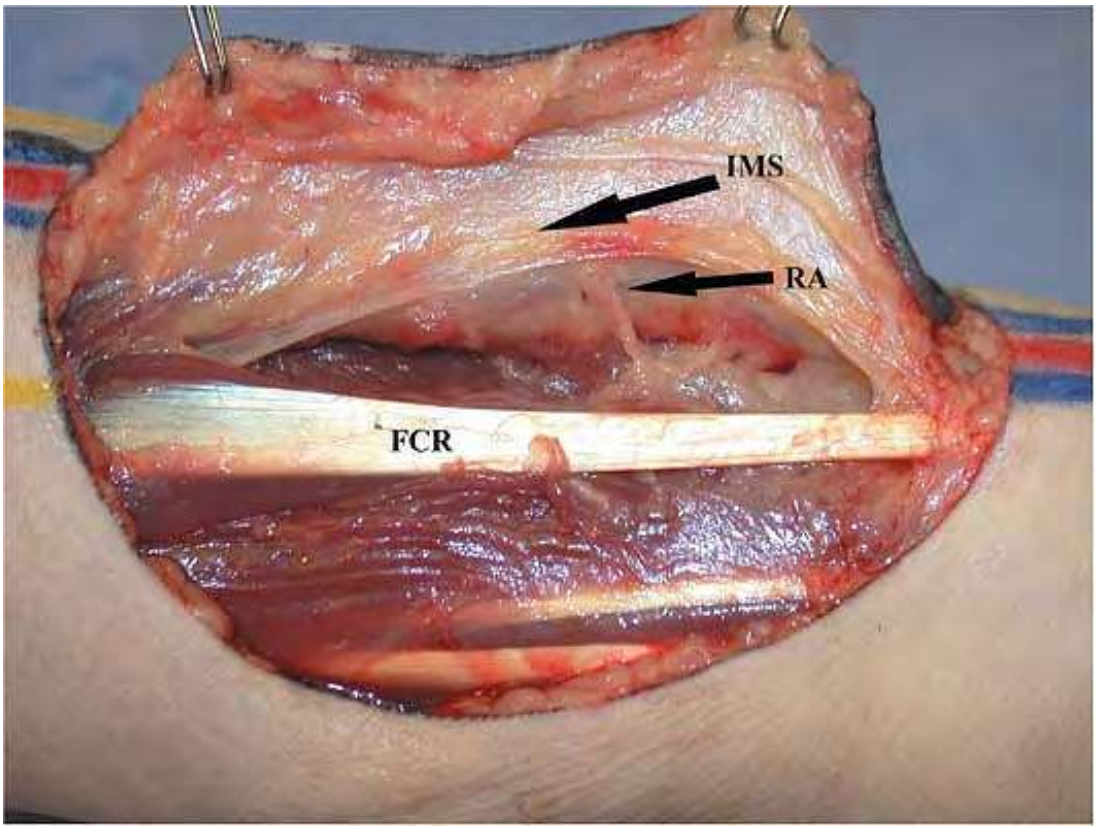

FIGURE 10-4 View from an ulnar to radial subfascual dissection demonstrating the radial artery (*RA*), intermuscular septum (*IMS*) through which septocutaneous perforator branches of the radial artery traverse, and flexor carpii radialis tendon (*FCR*).

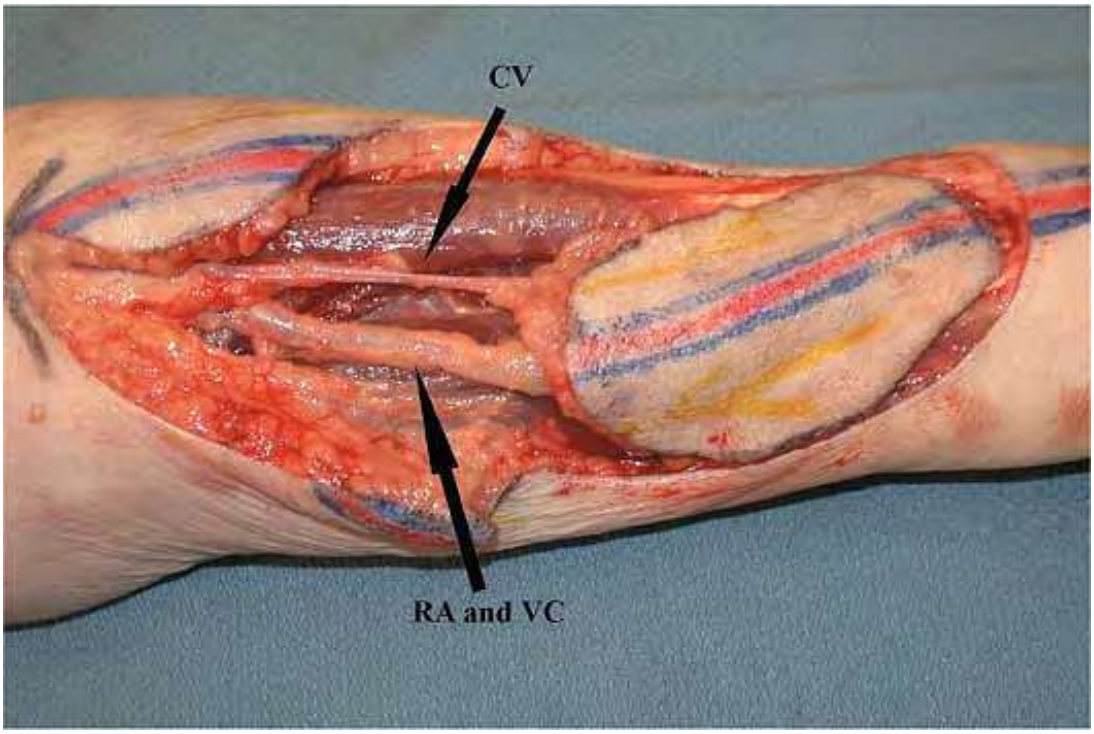

FIGURE 10-5 This figure demonstrates a flap that has been dissected proximally to an appropriate pivot point that will allow a tension free flap transfer. Note the vascular pedicle containing the radial artery (*RA*) and venae comitantes (*VC*). The cephalic vein (*CV*) was included in this flap for venous drainage augmentation.

FIGURE 10-6 Once the vascular pedicle is dissected to an appropriate pivot point, the flap can be transposed to the posterior elbow region from either an ulnar (A) or radial (B) direction.

When certain that the flap is well perfused, the flap is transferred from either an ulnar or radial direction through a subcutaneous tunnel to the elbow and inset with interrupted suture (Fig. 10-6). After obtaining meticulous hemostasis, we routinely attempt to cover the exposed tendons with local flexor tendon muscle bellies and soft tissues. This is followed by unmeshed split-thickness skin graft coverage of the entire donor site (Fig. 10-7).

A case of a proximally based neurosensory radial forearm flap for soft tissue reconstruction of the periolecranon region is illustrated in Figure 10-8.

Postoperative Management

Immobilization and elevation are absolutely critical in the postoperative management of elbow reconstruction with the radial forearm flap. A nonconstricting and bulky above-elbow splint (including fingers) is applied with a window cut out for accessibility and inspection of the flap. The elbow is held at 90 degrees, and the digits are placed in the position of safety. The digits are immobilized to prevent shear forces on the forearm donor site skin graft by tendon excursion. To help decrease postoperative edema and the possibility of the patient exerting pressure on the flap while asleep, the arm can be hung from an intravenous fluid pole. For the first 3 postoperative days,

the flap is evaluated on a frequent (every 4 to 8 hr) basis for capillary refill and the absence of venous congestion. The splint and the donor site bolster are removed on postoperative day 5 when capillary inosculation of the skin graft has occurred. Thereafter, a soft dressing is applied and the skin graft and flap

suture lines are kept moist with bacitracin ointment. Digital and elbow range of motion are gently begun and progressively increased.

FIGURE 10-7 The forearm donor site is covered with an unmeshed split-thickness skin graft.

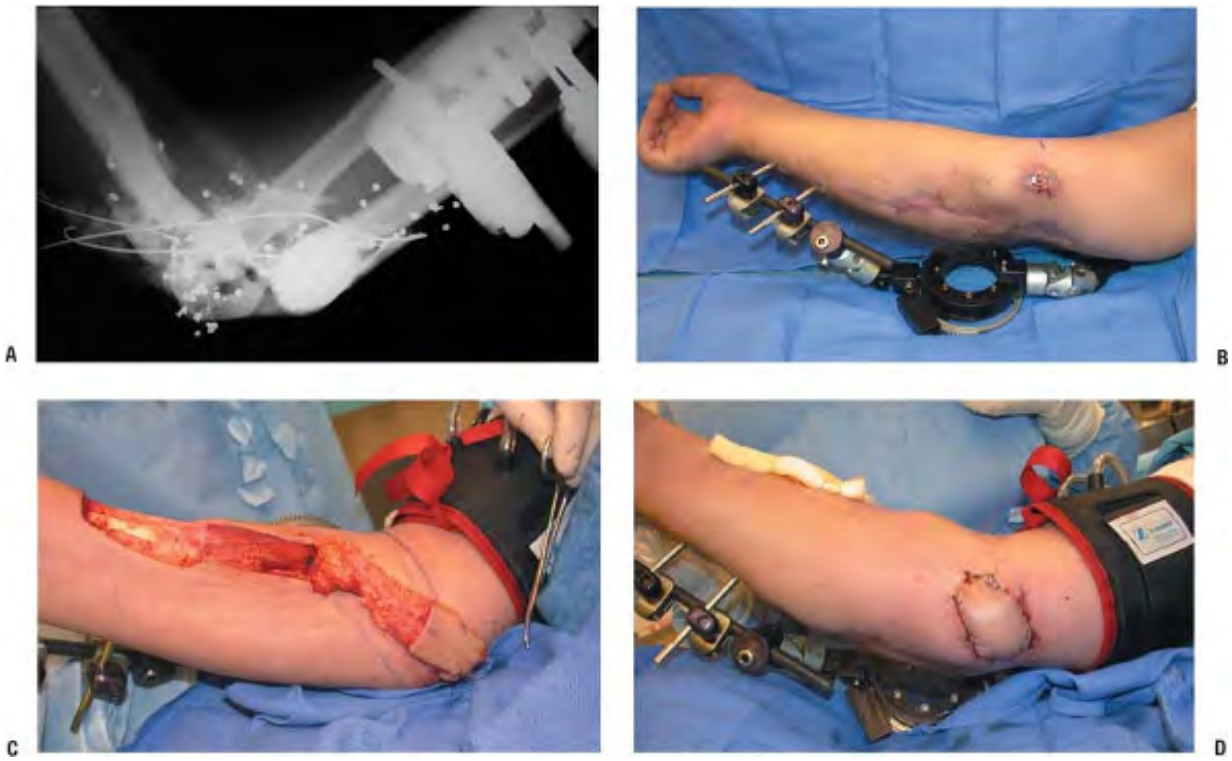

FIGURE 10-8 A 32 year old male with significant injury to the elbow from a shotgun assault. Following appropriate bone stabilization with external and internal fixation (A) he developed wound dehiscence and hardware exposure (B). An innervated neurosensory radial forearm flap was designed incorporating the medial cutaneous nerve of the forearm (C) and inset through a subcutaneous tunnel (D).

Results

Case reports of soft tissue reconstructions about the elbow region can be found in the medical literature. Hallock (6) described two cases of chemotherapy extravasation and exposure of hardware with successful healing following radial forearm flap. Thornton (13) reported a case of olecranon osteoradionecrosis treated with the radial forearm flap resulting in early elbow range of motion and complete healing. Small and Miller (11) present three cases in elderly patients to illustrate the utility of the radial forearm flap. One patient had an established infection with radial head and lateral

epicondyle ischaemia. Despite aggressive treatment this patient died 32 days postoperatively from sepsis. Another patient with a crush injury developed an ankylosed elbow that was free of pain and infection. Meland (7) reported successful pedicled radial forearm flap coverage for three patients with recalcitrant about the elbow defects in whom one or more previous flaps had failed. Tizian (15) reported the use of proximally based radial forearm flap in 14 patients. All patients showed complete healing of the periolecranon region and forearm donor site with unrestricted elbow range of motion. In a recent comparative review by Chondry and colleagues, the radical forearm flap provided excellent coverage of moderate size defects of the elbow. The flap was associated with fewer complications when compared to the pedicled latissimus dorsi flap (1).

Complications

A major, yet rarely encountered complication of the radial forearm flap is ischemia of the hand. A meticulous preoperative Doppler examination of the hand using the modified Allen's test should identify patients who are at risk. However, if arterial insufficiency of the hand is encountered once the flap is elevated, the radial artery needs to be reconstructed using an interposed vein graft. A great saphenous vein graft anastomosed end-to-side proximally and end-to-end distally is the authors preferred method.

Other major complications include flap ischemia and congestion. In both of these instances careful evaluation of the flap is mandatory to ensure that the subcutaneous tunnel is adequate, the skin paddle is not under excessive tension or external pressure, and there is no undue stretching or twisting of the vascular pedicle.

Donor site problems are one of the most common complications of the radial forearm flap. These complications can include hypertrophic scarring, poor skin graft take, and tendon loss (3,14). As mentioned previously, it is imperative to preserve the flexor tendon paratenon and attempt to cover exposed tendons such as the FCR with local muscle bellies and soft tissue advancement. If there is inadequate skin graft take and tendon exposure, local dressing changes with saline-dampened gauze are initiated to prevent tendon desiccation until the wound healing by secondary intention is complete. In rare cases, secondary flaps may be required to cover exposed tendons in

the forearm. Other reported donor site complications can include injury to the radial sensory nerve or hand swelling and stiffness (14).

Conclusion

Soft tissue defects of the posterior elbow are not uncommon and can often be difficult to reconstruct. Although a multitude of reconstruction options are available, each patient needs to be individualized. The advantages of the pedicled radial forearm fasciocutaneous flap for the management of elbow defects include ipsilateral donor site and scar, relatively simple dissection, long vascular pedicle that allows a wide arc of rotation, ability to make into a neurosensory flap by including forearm cutaneous nerves, no requirement for microsurgical expertise, ease of postoperative limb positioning, and highly reliable and vascularized tissue that is an excellent option even in the setting of osteomyelitis (7).

Recommended Reading

Mathes S, Nahai F. *Reconstructive Surgery: Principles, Anatomy, and Technique.* New York: Churchill Livingstone, 1997.

References

1. Chondry AH, Moran SL, Li S, et al. Soft tissue coverage of the elbow: an outcome analysis and reconstructive algorithm. *Plast Reconstr Surg*. 2007;119:1852.

2. Davalbhakta AV, Niranjan NS. Fasciocutaneous flaps based on fascial feeding vessels for defects in the periolecranon area. *Br J Plast Surg.* 1999;52:60.

3. Fenton OM, Roberts JO. Improving the donor site of the radial forearm flap. *Br J Plast Surg.* 1985;38:504.

4. Frost-Arner L, Bjorgell O. Local perforator flap for reconstruction of deep tissue defects in the elbow area. *Ann Plast Surg.* 2003;50:491.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 11

Pedicled and Free Latissimus Flap for Elbow and Forearm Coverage

Kenji Kawamura

Kevin C. Chung

Indications/Contraindications

The latissimus dorsi (LD) has proven to be a reliable muscle in the coverage of soft tissue defects about the shoulder and elbow. Familiarity with this muscle and its application in cases of trauma is essential for all upper extremity surgeons. The major arterial inflow and venous outflow for the LD flap is based on the thoracodorsal artery and venae comitantes. The average diameter of the thoracodorsal artery is 2.5 mm, and that of venae comitantes is 3.0 mm. The average length of maximum vascular pedicle is 12 cm, which can be obtained by dissecting the thoracodorsal vessels toward the proximal axillary artery and vein.

The indication for the use of the LD flap to the elbow and forearm is to cover a large skin and soft tissue defect that cannot be managed with local flaps. In cases in which vital muscle structures were damaged, the LD flap can provide coverage of soft tissue defects, as well as functional muscle transfer to restore of elbow, wrist, and finger motion. Unstable soft tissue coverage over elbow fracture or recalcitrant infection around the elbow may also be amenable to the LD flap because this flap provides well-vascularized tissue that can seal dead spaces and increase blood flow

to the local environment.

The pedicled LD flap can cover the forearm up to approximately 8 cm distal to the olecranon when transposed anteriorly, and about 6 cm distal to the olecranon when transposed posteriorly. Soft tissue defects over 8 cm distal from the olecranon are not suitable for the pedicled LD flap transfer, and if this muscle is to be used, the free LD flap transfer should be considered.

Contraindications for flap use include previous injury to the muscle or pedicle, such as in cases of previous thoracotomy, axillary arterial injury, or in some cases, axillary lymphadenectomy. Breast cancer surgery may injure the nerve or arterial supply to the LD muscle, rendering it fibrotic and inadequate for transfer. In such patients, palpation of a contracting LD muscle usually verifies an uninjured nerve and vascular pedicle.

The LD muscle functions as an expendable adductor, extender, and internal rotator of the arm. These functions are essential for activities of daily living for patients with contralateral shoulder girdle paralysis or extremity paralysis from spinal cord injury. One must be careful in evaluating the impact of removing the LD muscle in these situations. In general, one should not use the LD muscle for soft tissue coverage in these patients because of functional requirements.

Preoperative Planning

The LD flap can be harvested as not only a muscle flap but also a musculocutaneous flap. Muscle flap covered with skin graft is less bulky and can seal deep defects, however, musculocutaneous flap gives better aesthetic reconstruction because the skin paddle can conform to the skin texture of the upper arm. When a LD musculocutaneous flap is applied, it is recommended that the skin paddle is designed several centimeters superior to the muscle origin at the iliac crest to avoid the risk of skin paddle necrosis because the blood supply of the subcutaneous tissue near the origin is inconsistent. When greater flap reach is required, a portion of the skin paddle may be designed over the inferior margin of the muscle. This technique can facilitate a greater distal coverage of the forearm when the pedicled LD musculocutaneous flap is rotated. The entire skin paddle can survive if enough musculocutaneous perforators are included in the proximal portion of the skin

paddle design.

The arc of transposition of the LD flap depends on the location of the wound. When defects are present in the posterior aspect of the elbow and forearm, posterior arc transposition of the pedicled LD flap is recommended. Similarly, when defects are present in the anterior aspect, anterior arc transposition is recommended.

It is important to assess the strength of the LD muscle preoperatively when the LD flap is used for simultaneous soft tissue coverage and functional restoration. When a soft tissue defect is caused by a trauma, particularly associated with a motorcycle accident, the LD muscle may be paralyzed by brachial plexus injury. If the paralyzed LD muscle is planned to be used as a functional transfer, reinnervation of the LD muscle by another motor nerve, for example, intercostal nerve or accessory nerve, can be performed during operative procedures. LD muscle previously paralyzed by neurologic disorders can only be used for soft tissue coverage and should not be used for functional muscle transfer.

Preoperative angiography of the injured upper extremity will help plan for free LD flap transfer. The free flap transfer may require a more proximal vessel dissection, vein grafting, and end-to-side anastomoses in cases with a potential disruption of the recipient vessels.

For aesthetic consideration at the donor site, it is important to ask for the patient's clothing preference preoperatively for planning the skin incision. For example, a transverse incision along the bra line is recommended to conceal the donor site scar for a woman, but an oblique incision may be suitable for a patient who likes to wear backless dresses. Furthermore, the endoscopic harvesting technique may be preferred for children, those who are prone to hypertrophic scars, and for other aesthetic considerations.

Surgery

The patient, under general anesthesia, is placed in the lateral decubitus position. The injured upper limb and ipsilateral hemithorax caudal to the iliac crest are prepped and draped. The injured upper limb is supported on a sterile Mayo stand with a 90-degree abduction position of the shoulder (Fig. 11-1A). The lateral buttock and thigh are also prepared for skin grafting if necessary. A beanbag or

axillary roll with padding is placed on the opposite axilla to protect the dependent shoulder in the decubitus position, and all other pressure points are also carefully padded.

In cases of traumatic defects, the wounds are again thoroughly debrided before the flap procedure is performed. When a LD musculocutaneous flap is applied, the skin paddle that includes the perforators arising from the branch of the thoracodorsal artery is designed on the LD muscle. There is a longitudinal row of perforators 2 to 5 centimeters from the lateral border of the LD muscle, which can be easily identified with a Doppler blood flowmeter. The largest skin paddle that can be moved safely is 20 Ã— 15 cm, however, primary closure is only possible with skin paddles 8 cm in width.

The procedure is performed with Ã—2.5 loupe magnification. The incision usually begins at the posterior aspect of the axilla angled along the anterolateral border of the latissimus dorsi muscle and extends inferiorly to the iliac crest. An aesthetic incision in a woman is designed in a transverse incision along the bra line. The LD muscle is dissected medially to the paraspinous muscle and caudally to the lumbosacral fascia. The origin of the muscle is detached from the thoracic and lumbar vertebra, posterior ribs, and lumbosacral fascia. The flap is elevated superiorly toward the axilla to separate the undersurface of the muscle from the chest wall. Perforators from the intercostal and lumbar arteries are clipped. The neurovascular pedicle, which includes the thoracodorsal artery, venae comitantes, and nerve, is identified more proximally on the undersurface of the muscle. The thoracodorsal

vessels are carefully dissected toward the proximal axillary artery. The circumflex scapular vessels and the serratus anterior and teres major branches are ligated to allow maximal pedicle length. Leaving the thoracodorsal nerve intact may be considered when the flap is transferred on the pedicle, even though the flap is not used as a functional transfer. Retaining innervation minimizes postoperative muscle atrophy and will maintain muscle thickness and durability. This consideration is particularly important for distal elbow coverage when the thin muscle near the origin of the LD muscle at the iliac crest maintains its bulk.

A

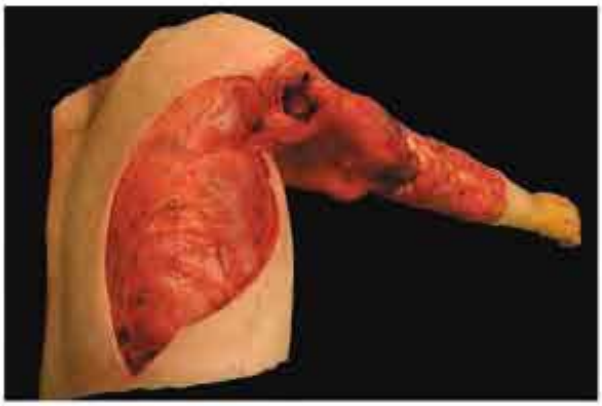

B

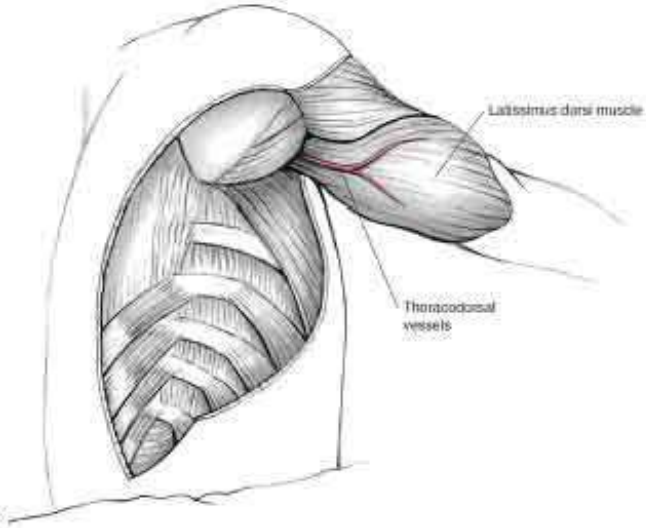

C

FIGURE 11-1 A: The patient was placed in the lateral decubitus position with the injured upper limb on a sterile Mayo stand with a 90-degree abduction position of the shoulder. B: A pedicled latissimus dorsi muscle flap was harvested and rotated posteriorly to the upper arm. A skin paddle was excised for contouring the flap on the arm. It is difficult to design a skin paddle to fit the wound geometry and it is simpler to just use the muscle only and cover with skin graft. C: Schematic drawing of a pedicled latissimus dorsi muscle transfer based on Figure 11-1B. D: Complete wound healing at 3 months postoperatively.

After the dissection of the neurovascular pedicle has been completed, the insertion of the LD muscle at the humerus is divided. Additional distal coverage with the pedicled LD flap can be obtained

by releasing its insertion from the intertubercular grooves of the humerus. When the LD flap is transferred as a free flap, vascular pedicle is divided at the juncture with the axillary artery and vein to obtain the maximum pedicle length. The thoracodorsal nerve is also divided at its proximal point when the flap is used as a functional transfer. When the LD flap is transposed to the elbow and forearm defects, the muscle can be positioned along its path by incising intervening skin or can pass under a subcutaneous tunnel (see Fig. 11-1B,C). The subcutaneous tunnel may be useful in situations when the defect size is relatively small and the transposed small muscle can be tailored to pass easily through the tunnel. But for large muscle, the subcutaneous tunnel may compress the vascular pedicle and the intervening skin should be incised to allow a tension-free passage.

The muscle is sutured to the subcutaneous tissue in the proximal portion of the wound to prevent tension on the pedicle. For functional muscle transfer, muscle tension is critical. For elbow flexor

reconstruction, the length of the transferred muscle is adjusted so that the elbow remains at 100 degrees flexion and the forearm is held in complete supination after both ends are sutured. By extending the muscle with strips of fascia lata, finger flexion or extension can also be restored.

Split-thickness skin graft is applied as needed for the transposed muscle or over the donor site (see Fig. 11-1D). Two silastic drains are placed at the donor site, and one is placed below the transposed muscle. Before closure of the donor site, fibrin sealant may be used to reduce seroma formation over the back.

When a free LD flap transfer is used for coverage of a forearm defect (Fig. 11-2A), dissection of recipient vessels should be performed before harvesting the flap to confirm the necessary length of the pedicle. A large LD muscle flap is suitable to cover a forearm defect because the flat configuration of the LD muscle allows it to be wrapped around the forearm. The LD muscle is a versatile muscle for coverage of a complex wound whereby the geometry of the wound requires two separate flaps. Splitting of the LD muscle is helpful in this situation. The thoracodorsal vessels usually bifurcate into medial and lateral branches just after entering the LD muscle. The LD muscle can be split longitudinally into halves based on these two branches. Splitting the LD muscle provides two flaps to coverage different wounds on the forearm (see Figs. 11-2B,C).

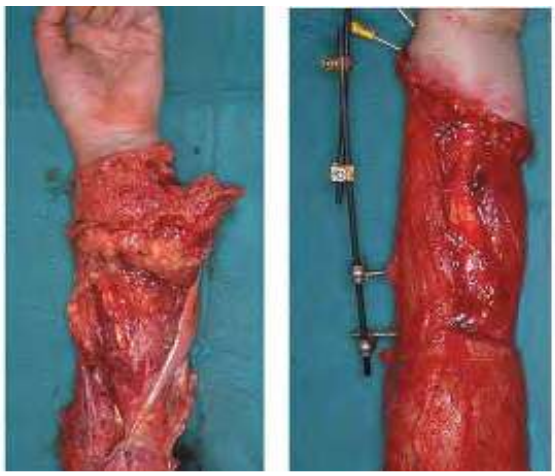

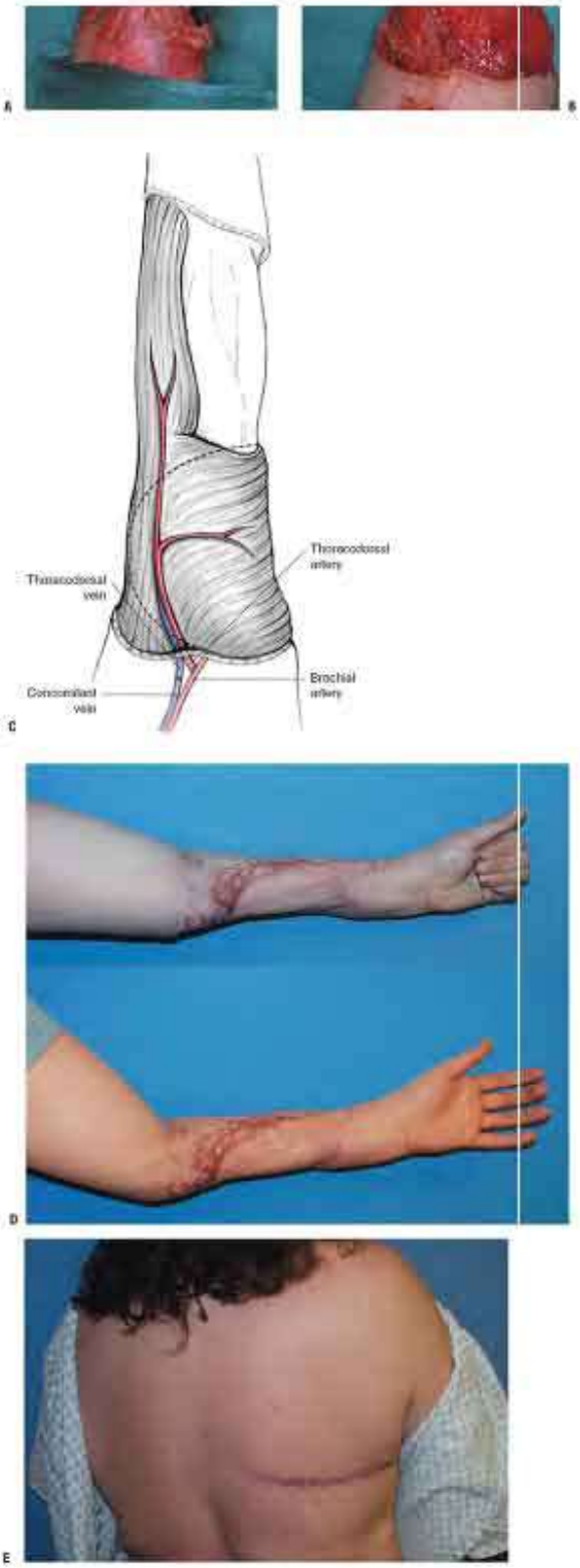
A
B
Thoracodorsal artery
Thoracodorsal vein
Brachial artery
Concomitant vein
C
D
E

FIGURE 11-2 A: Crush injury of the forearm. B: The latissimus dorsi muscle was split longitudinally into halves by identifying the arterial pedicles. One half was used to cover the soft tissue defect over the posterior elbow, and the other half was used to cover the exposed bone at the wrist and exposed tendons. The thoracodorsal artery was anastomosed to the brachial artery using an end-to-side technique, and the thoracodorsal vein was anastomosed to the brachial concomitant vein with an end-to-end technique. C: Schematic drawing of a free latissimus dorsi muscle transfer based on Figure 11-2B. D: Complete wound healing at 4 months postoperatively. There are no limitations of finger motions. E: Donor site scar resulting from an aesthetic transverse incision along the bra line.

Postoperative Management

During the first 48 hours after the operation, flap circulation is closely monitored by direct observation of the flap, which includes color, temperature, capillary refilling time, and bleeding of the flap. A Doppler blood flowmeter is useful to examine the arterial flow into the flap. If there is concern about the arterial inflow or venous outflow, the pedicle should be reexamined and the wound reexplored. Kinking and tension on the pedicle must be relieved to prevent total necrosis of the muscle. Postoperative hematoma can also compromise venous outflow.

The drain at the donor site must remain in place for 2 to 3 weeks to prevent seroma formation. Application of fibrin sealant in the donor site may reduce seroma formation, but the drains should remain in place until drainage is minimal, usually less than 25 mL/day. Compression of the donor site by thoracic elastic bandages may be effective to enhance sealing of the wound. When a seroma is diagnosed, needle aspiration is performed. Small seromas (<100 mL) may resorb spontaneously, but a large seroma needs serial aspirations.

The elbow is immobilized at 90 degrees of flexion, and the shoulder is adducted for 3 weeks after the operation to avoid wound dehiscence and tension on the pedicle. During this period, finger and wrist exercises are actively practiced (see Fig. 11-

2D). After splint removal, elbow and shoulder range of motion exercises are started. The goal of physiotherapy is to recover the full range of motion of the elbow and shoulder.

Complications

Donor site seroma or hematoma is the most commonly encountered complication at the donor site following harvest of the LD flap. Seromas can be relieved with frequent needle aspirations, but hematomas must be drained immediately to avoid secondary donor site complications relating to wound dehiscence or skin necrosis.

Scarring over the donor site is inevitable, but can be easily covered by clothing. Planning the skin incision and endoscopic harvesting technique can reduce back scar problems (see Fig. 11-2E).

Total flap necrosis is a rare complication with the LD flap method. However, partial flap necrosis at the distal portion may occur because of inconsistent blood supply to the lower third of the LD muscle. Bleeding at the distal end of the LD flap should be checked when the flap is elevated. Kinking and tension on the pedicle can cause disturbance of flap circulation and must be recognized immediately. It is also important to note potential ischemia of the transferred muscle caused by subcutaneous tunnel compression.

Power and endurance of shoulder extension and adduction may be weakened in the absence of the LD muscle. The weakness is not noticeable in most patients and will recover within a few months after removal of the LD muscle because the synergistic shoulder muscles such as teres major, pectoralis major, and subscapularis substitute the function of the LD muscle.

Hyperabduction of the arm during the operation may cause iatrogenic brachial plexus injury, which is reversible within several weeks. Care must be taken to avoid prolonged shoulder abduction during the operative procedure.

Lumbar hernia after the latissimus dorsi flap is an extremely uncommon complication and it may be misdiagnosed as lumbar seromas. Lumbar hernia can be avoided by preserving of the fascia underlying the distal latissimus dorsi aponeurosis.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 12

Fasciotomies for Forearm and Hand Compartment Syndrome

Jeffrey B. Friedrich

Alexander Y. Shin

Indications/Contraindications

Indications

The chief indication for forearm and/or hand fasciotomies is essentially any suspicion that the patient has a forearm or hand compartment syndrome. The diagnosis of upper extremity compartment syndrome is made with a combination of clinical examination and objective diagnostic measurements. The clinical examination findings for upper extremity compartment syndrome traditionally have been taught as the five â€œP'sâ€• which include pain in the forearm and hand, increased pain with passive extensive extension of the fingers, pallor, pulselessness, and paresthesias. However, one must note that not all of these criteria need be present to determine that a patient has compartment syndrome. In fact, pulselessness can be a very late or nonexistent finding in the patient who has a compartment syndrome. Therefore, reliance on these five P's should also be balanced with other clinical examination findings as well as objective diagnostic findings. In addition to the criteria mentioned, the forearm and hand suffering from compartment syndrome will typically be very tense and swollen and can often have

signs of trauma from the inciting injury, if that is the cause of the compartment syndrome. Skin changes such as blistering and a â€œshineâ€• due to significant swelling are common findings as well (Fig. 12-1).

Compartment syndrome can be caused by extrinsic pressure such as casts or splints, and this possibility should be factored into the diagnostic algorithm. The first step in evaluation with a patient who is in a cast or splint is removal or loosening of the immobilization device. Other scenarios in which the physician should suspect the development of compartment syndrome include crush injuries, supracondylar humerus fractures, two-bone forearm fractures, and any situation where there has been a period of ischemia followed by reperfusion, such as brachial arterial injury, upper arm replants, or a comatose patient who has been found with the arm in compression.

Objective data are obtained to help corroborate the diagnosis of compartment syndrome, usually in the form of intracompartmental pressure measurements. These can be taken by a variety of techniques including saline infusion pressure transducers, slit or wick catheters, or even arterial pressure monitors. There are newer diagnostic tools to diagnose compartment syndrome including near-infrared spectroscopy, electromyography, myotonometry, and laser Doppler; however, these modalities are, at this point, experimental and can only be used as correlative measures at best.

If compartment syndrome is suspected, it is essential that intervention be implemented quickly to lessen injury to the underlying muscles and nerves. Increased pressure on muscle can lead to irreversible ischemia and eventual necrosis. Late sequeli of untreated compartment syndrome of the forearm produces contractures of the volar forearm commonly known as Volkmann's ischemic contracture.

In general, prompt administration of surgical decompression of the forearm compartment will adequately treat the forearm compartment syndrome and leave a patient with minimal residual dysfunction.

FIGURE 12-1 Patient with forearm compartment syndrome demonstrating typical cutaneous findings including tense swelling and blistering.

Contraindications

While one can argue that hemodynamic instability would be a relative contraindication to upper extremity fasciotomies, it is recommended that this not be viewed as an absolute contraindication, and that once the patient is stabilized, the surgical team proceeds with forearm and/or hand fasciotomies at the earliest available time. In the worst-case scenario, fasciotomies can be performed at the bedside in the intensive care unit. In general, upper extremity compartment syndrome is an emergency and there should be no absolute contraindications to this procedure.

Preoperative Planning

When planning for surgical intervention for upper extremity compartment syndrome, the previously mentioned diagnostic modalities should be used to document increased forearm or hand pressures. In addition, if it is thought that the

patient has developed a compartment syndrome in the setting of external compression including splints or casts, these devices should be loosened or removed immediately to help diagnose the cause of the patient's symptoms as well as provide some measure of relief from the compressive insult. Often, the removal or loosening of a cast can provide adequate treatment; but even if the patient does show resolution of symptoms, practitioners are encouraged to remain vigilant and continue to closely monitor the patient's extremity until there has been a total resolution in symptoms. Other planning should include early notification of anesthesia and operating room staff and the availability of blood for transfusion if it is indicated during the surgical procedure.

In some cases coagulopathy can contribute to or be the sole cause of compartment syndrome. Excessive bleeding following fracture reduction or continued bleeding following surgery can lead to increased compartmental pressures. If a bleeding dyathesis is suspected, appropriate steps should be taken to correct the deficiency before operative intervention.

Surgery

Patient Positioning

Once anesthetized, the patient should be positioned in a way that affords adequate access to the entire affected upper extremity for both surgeon and assistant. This includes turning the operating table at a right angle to the anesthesia provider and placing the affected upper extremity on an arm board that is attached to the operating room table. Good lighting is essential, especially because the viability

of the muscles in the forearm and hand needs to be assessed at the time of fasciotomy. While pneumatic tourniquets are frequently used for upper extremity surgery, they should not be used for a forearm fasciotomy because, as stated previously, the perfusion of the affected forearm musculature must be assessed at the time of fasciotomy.

Technique for Fasciotomy of Forearm

The classic teaching of forearm fasciotomies includes two incisions, one on the volar aspect and one on the dorsal aspect of the forearm. In general, both volar

forearm compartments (superficial and deep) are decompressed via the volar incision while the dorsal compartment and the mobile wad can be adequately decompressed via the dorsal incision. A variety of incisions on the volar forearm has been proposed and has even been studied in cadavers. Ulnar sided incisions have been shown to be the safest approach to both superficial and deep volar forearm compartments, providing adequate visualization while avoiding injury to the radial artery and median nerve. Typically, the volar forearm incision extends from the antecubital fossa down to the wrist. The most common design is that of a sigmoid incision. The incision is marked beginning on the medial aspect of the elbow just anterior to the medial epicondyle, curving radially along the mid-forearm, curving back to the distal ulnar side of the forearm, and finally coming along the distal wrist crease to approximately the level of the carpal tunnel (Fig. 12-2). The purpose of this design is to, in theory, offer protection for the median nerve once decompression has been performed. The radially-based flap in the distal forearm will provide this median nerve protection even with marked edema and gaping of the wound. Once the skin incision is marked, it is incised with a scalpel. The dissection is further carried down through the subcutaneous tissues with either electrocautery or scissor dissection. Once the antebrachial fascia of the volar forearm is visualized, the fascia is incised with curved tenotomy-type scissors. The fascial division should be extended proximally and distally until the entire superficial volar forearm compartment is decompressed. Once the fascia is open, the viability of the musculature must to be assessed. If compartment syndrome is detected early and fasciotomy is employed in a rapid manner, there will typically not be any visible muscle necrosis. If, after observing the musculature for several minutes there is no reperfusion, or there is distinct necrosis of any of the forearm muscles, consideration should be given to muscle debridement at that time.

At this point only the superficial volar forearm compartment has been decompressed and the operating surgeon must then decompress and observe the musculature of the deep volar compartment as this is the area most frequently affected during a forearm compartment syndrome (Fig. 12-3). The deep volar compartment is approached through the interval between the flexor digitorum superficialis (FDS) and the flexor carpi ulnaris (FCU). The FDS muscle belly is retracted radially and the FCU muscle belly is retracted ulnarly; this allows visualization of the deep compartment and protects the median nerve and ulnar

neurovascular bundle. The deep volar fascia is opened along the entire length of the flexor digitorum profundus (FDP) and flexor pollicis longus (FPL) muscle bellies. Once the fascia is incised, the muscles will quite often appear edematous and will herniate from their respective compartments. The viability of these muscles must also be assessed at the time of the procedure. Consideration should be given to excision of any distinctly necrotic muscle at the time of fasciotomy, as this can become a nidus for infection if left in situ.

Some authors believe that decompression of the volar forearm compartments will secondarily lead to adequate decompression of the dorsal forearm compartment as well as the mobile wad. However, if one is planning to take this approach it is strongly advised that compartmental pressure readings of the dorsal forearm compartment as well as the mobile wad be taken intraoperatively to confirm

that decompression of the volar side of the forearm has indeed led to adequate decompression of the dorsal forearm.

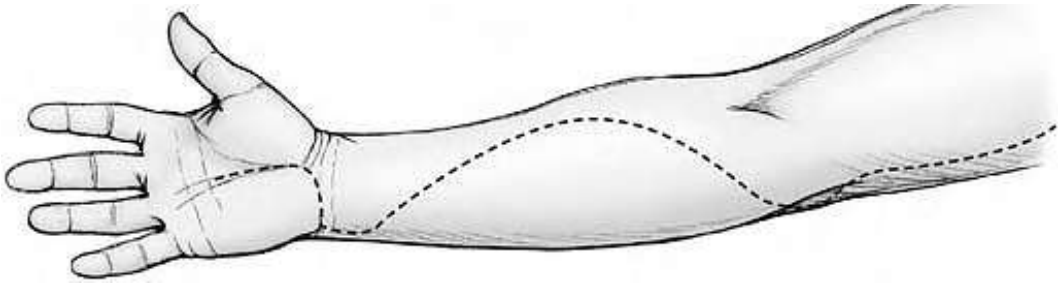

FIGURE 12-2 Incision pattern for fasciotomies of the forearm and carpal tunnel. (Adapted from Gulgonen A. Compartment syndrome. In: Green, et al, eds. *Green's operative hand surgery.* 5th ed. Philadelphia: Elsevier; 2005. With permission.)

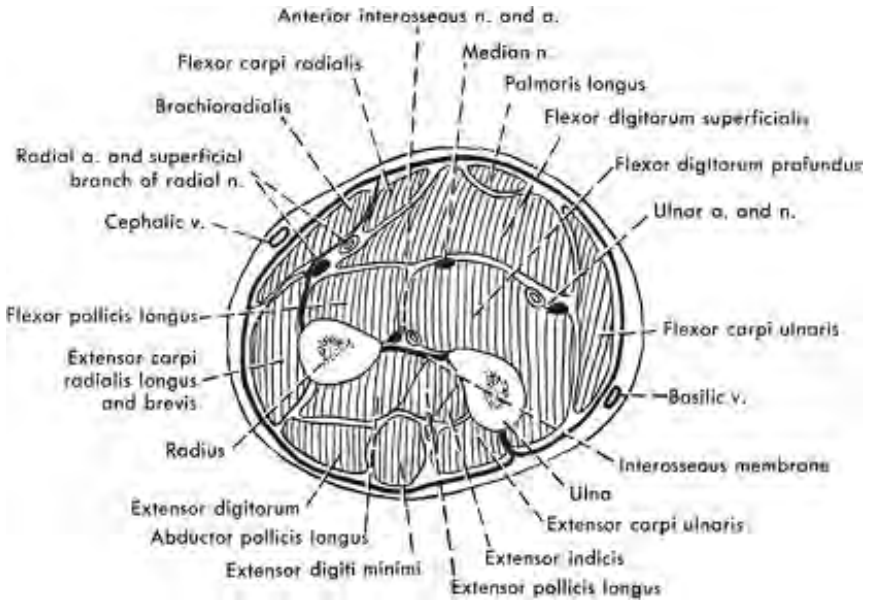

FIGURE 12-3 Cross-sectional anatomy of the mid-forearm. Note the deep position of the flexor pollicis longus and flexor digitorum profundus adjacent to the radius, ulna, and interosseus membrane. This deep position against the rigid skeleton makes these muscles particularly vulnerable to damage due to compartment syndrome (used by permission of Mayo Foundation).

After completing the volar forearm decompression, the dorsal side is now approached. The incision of the dorsal forearm is in linear fashion over the midline of the forearm dorsum. Again, this is incised with a scalpel and carried down through the subcutaneous tissue with either electrocautery or scissor dissection. The dorsal forearm fascia is incised typically with tenotomy scissors along the length of the extensor muscle bellies both proximally and distally. Once the musculature of the dorsal forearm compartment has been assessed, one must then dissect in the subcutaneous plane in the radial direction to access the mobile wad. It is here that the fascia over the brachioradialis and the radial wrist extensors (extensor carpi radialis brevis and longus) is incised with scissors and, again, the musculature is assessed.

Technique for Fasciotomy of the Hand

The hand has seven compartments: the thenar, hypothenar, carpal tunnel, and four interosseus compartments. Fortunately, all seven compartments can be decompressed with only four incisions. The carpal tunnel is opened in continuity with the forearm fasciotomy. The forearm incision is carried transversely within the wrist crease to the division of the thenar and hypothenar mounds. Here the incision is carried distally to the distal point of the transverse carpal ligament (see Fig. 12-2). After dividing the skin and subcutaneous tissues, the transverse carpal ligament is divided with scissors or a scalpel, all while continually protecting the median nerve.

The thenar and hypothenar compartments are decompressed next. These incisions are made at the border of the glabrous and hair-bearing skin of the thenar and hypothenar eminences, respectively, and are approximately 3 to 4 cm long. Once the skin and fat is opened, the fascias of these compartments are also opened with spreading scissor dissection.

Finally, the four interosseus compartments can be opened with two incisions on the dorsum of the hand. These incisions are centered over the index and ring finger metacarpal bones, and are 3 to 4 cm long. After making the index metacarpal incision, scissors are used to spread radially to open the first interosseus compartment and ulnarly to open the second interosseus space. Similarly, one enters the ring metacarpal incision to spread radially into the third interosseus compartment and ulnarly into the fourth interosseus compartment (Fig. 12-4).

At this point all of the compartments of the forearm and hand have been adequately decompressed. Primary closure is usually not possible, and not advised. One should consider delayed primary closure of the forearm and hand wounds following resolution of the patient's compartment syndrome, stabilization of the patient, and assurance that there is no underlying muscle necrosis or signs of infection. Delayed primary closure can be undertaken from 1 to 7 days following a fasciotomy, depending on the health of the patient and status of the underlying tissue. Temporary retention sutures or running elastic sutures (the â€œJacob's ladderâ€• or â€œRoman sandalâ€•) can be placed in the wound to provide some skin approximation (Fig. 12-5), but it is not advised that full closure be

attempted at this time as this can essentially recreate a compartment syndrome

due to a tight skin envelope. Another acceptable method of wound temporization is application of a negative pressure dressing, which has the dual benefit of wound fluid evacuation and prevention of further skin edge retraction.

FIGURE 12-4 Two dorsal hand incisions placed over the index and ring metacarpals for decompression of the interosseus musculature of the hand.

FIGURE 12-5 Partial coaptation of forearm fasciotomy wound edges using the â€œRoman sandalâ€• method. Vessel loops or rubber bands are weaved across the wound to bring the skin edges closer while allowing for some expansion due to postoperative edema.

Following initial fascial release, repeat examination under anesthesia is recommended at 24 to 48 hours; this allows the surgeon to verify that all necrotic muscle has been removed. If extremity edema has subsided, the skin may be closed with delayed primary closure. If there is any concern that closure is â€œtightâ€• or that closure could recreate a compartment syndrome, a â€œJacob's ladderâ€• or negative pressure dressing is reapplied to the wound and the patient is scheduled for surgical re-evaluation in 24 to 48 hours. Delayed primary closure of both palmar and dorsal wounds may be impractical in cases where the antecedent trauma has been significant. In such cases, skin grafting may be carried out over exposed muscle bellies to expedite wound coverage and patient rehabilitation (Fig. 12-6).

Postoperative Management

Postoperatively, the patient's forearm wounds should be dressed with nonadherent gauze. We most commonly use gauze impregnated with either petroleum or antibiotics to prevent wound desiccation. The patient should also be placed in a plaster or thermoplastic splint that covers the entire length of the forearm and places the hand in an intrinsic-plus position. One must take great care with the placement of this splint or cast as it can itself cause a compartment syndrome and lead to further muscle

damage despite adequate decompression of the forearm and hand compartments. In the several days between fasciotomy and wound closure, adequate wound care should be ensured. This is performed with frequent dressing changes and either whirlpool or pulse lavage type irrigation. At this time it is also wise to instruct the patient in hand motion exercises. The chief reason for these exercises is to ensure that the patient's range of motion remains relatively stable and that tendon gliding continues, especially with the flexor tendons at the wrist level.

FIGURE 12-6 Forearm fasciotomy wound immediately following skin grafting.

Complications

As with any surgery, bleeding and hematoma are possible, especially in a forearm fasciotomy in which no tourniquet is used intraoperatively. Any open wound is prone to infection, and good wound care should be instituted to prevent this problem. Finally, the most dreaded complication of forearm fasciotomy is continued muscle necrosis and eventual fibrosis despite forearm fasciotomies. In general, this will likely not occur with adequate decompression. As stated previously, the most likely reason for continued muscle necrosis following this operation is inadequate decompression of the forearm and hand compartments.

Results

In general, adequate decompression of the compartments of the forearm and hand will allow resolution of the compartment syndrome and will lead to an outcome in which the patient has no or minimal residual dysfunction, although this general scenario is not without exceptions. Most frequently, the cause of muscle necrosis and subsequent fibrosis following forearm fasciotomy is due to incomplete fasciotomy. Therefore, care must be taken intraoperatively to fully extend the fasciotomy proximally and distally, thereby decompressing the full length of all the

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 13

Soft Tissue Interposition Flaps in the Management of Heterotopic Ossification and Proximal Radioulnar Synostosis

Douglas P. Hanel

Seth D. Dodds

Indications/Contraindications

Heterotopic ossification of the elbow with loss of motion can be severely debilitating. The functional arc of elbow motion spans from 30 to 130 degrees of flexion and from 50 degrees of pronation to 50 degrees of supination. Limitations to this functional arc significantly impair the ability to perform activities of daily living. Cases of elbow ankylosis or radioulnar synostosis further minimize use of the affected extremity. Once heterotopic ossification has developed and constricts motion, it is nearly impossible to regain the lost motion with conservative measures, such as physical therapy, dynamic splinting, radiation therapy, or medication.

Surgical resection of heterotopic bone about the elbow should be considered in patients who present with an unacceptable loss of flexion/extension or pronation/supination. Excision is also warranted in cases of neurovascular impingement caused by ectopic bone. In cases of proximal radioulnar synostosis or even radial head excision, interposition materials can be used to cover exposed bone

surfaces.

While options for soft tissue interposition include silicone sheeting, fat graft, and free adipofascial flaps, pedicled myofascial flaps and allograft fascia lata have become increasingly popular. The pedicled anconeus myofascial flap is an ideal choice in those cases approached posterolaterally, whereas the pedicled brachioradialis muscle flap passed through the interosseous membrane is best suited for those cases approached from an anterior exposure. The soft tissue interposition acts as a barrier to the formation of recurrent heterotopic calcification and allegedly decreases pain with pronation and supination. It is suggested, though unproven, that pedicled graft tissue has greater potential

for sustained viability when judged against nonvascularized tissue transfers, such as subcutaneous fat or an adipose fascial graft. When properly fixed to the underlying bone, these pedicled flaps can sustain aggressive postoperative range-of-motion exercises that might dislodge less robust tissue such as fat alone or adipose-fascial grafts. We prefer the anconeus muscle pedicle flap when available and tensor fascia lata allograft when the anconeus is not available or not large enough.

Historically, contraindications to excision of heterotopic bone with or without soft tissue interposition flaps included immature ossification and an unreliable soft tissue envelope. In the past decade there have been a number of reports documenting the efficacy of early excision of heterotopic bone. It is our experience that patients suffer less soft tissue contracture and have superior function with early release of a stiff elbow. The recurrence of heterotopic bone in the posttraumatic setting has not been shown to be predicated on the timing of the excision. Outcomes of surgical excision, however, will be threatened by a poor soft tissue envelope. Once the posttraumatic or post-burn wounds have healed, soft tissue swelling has abated, and nerve recovery has plateaued, patients may safely undergo elbow contracture release.

Contraindications to pedicled soft tissue interposition flaps depend on the specific muscle selected. For an anconeus muscle interposition, previous traumatic or surgical disruption of its primary vascular supply (the medial collateral artery from the profunda brachii) jeopardizes the viability of the raised muscle flap. The brachioradialis â€œwrap aroundâ€• flap should not be raised if there is a nonfunctioning biceps brachii or brachialis, as the brachioradialis provides assistance

with elbow flexion as well as supination when the forearm is fully pronated. Assuming that the arm and forearm musculature is intact, using the anconeus or brachioradialis as a pedicled interposition flap causes little functional loss.

Preoperative Planning

Before operative release and soft tissue interposition, patients must be carefully evaluated. The history should focus on the primary complaint. Patients must verbalize appropriate frustration with their disability from elbow stiffness to warrant release. It is also imperative that patients demonstrate the willingness and capacity for intensive rehabilitation. If the presenting complaint is predominantly pain, then contracture release will be futile. A focused surgical history needs to be elicited. Previous injuries and surgeries of the involved extremity offer critical information about the status of osseous and cartilaginous structures as well as the elbow's soft tissue envelope. Operative reports from previous surgeries help understand the integrity and location of possibly transposed neurovascular structures.

Elbow range of motion, stability, pain, and functional ability are assessed. The Mayo Elbow Performance Score serves as a summary of these findings. In addition, the examination includes assessment of forearm rotation and wrist and hand function. If there is physical evidence of nerve dysfunction, electrodiagnostic studies should be obtained to confirm the location of compression and to act as a baseline of nerve function. All of our patients have had previous surgeries, and as such, a careful assessment of incisions about the elbow with regard to the palpable and radiographic location of heterotopic ossification is essential. The integrity of the skin and subcutaneous tissues should be evaluated. Lingering soft tissue swelling, edema, or erythema may all point to additional diagnoses to be contended with before deciding on heterotopic bone excision.

Standard radiographs of the elbow are obtained with oblique views to improve visualization of ectopic ossification. We routinely perform computed tomography (CT) on cases of elbow heterotopic ossification where the congruity of the articular surface is in question and in all cases with proximal radioulnar synostosis. An axial CT scan with coronal and sagittal reformatted images is currently the most helpful method to visualize the location and extent of a bony bridge between the radius and ulna (Figs. 13-1 and 13-2). Vascular study of the elbow and proximal forearm should be

considered if there are concerns about the integrity of the regional blood supply, especially in cases of pedicled soft tissue interposition. Unless there are specific historical or physical findings suggesting potential or indolent infection, we limit blood studies to those required for a prolonged general anesthetic.

The timing for intervention is somewhat nebulous and is certainly directed by physician bias. We do not believe there is sufficient scientific evidence to suggest that nuclear medicine scans or blood alkaline phosphatase levels are beneficial in the assessment or timing of surgical intervention. In reviewing the literature and comparing our experience, it would appear that waiting for fracture healing

and ectopic bone maturation, defined as well-delineated borders, is the most commonly used parameters. This would suggest that operative intervention be carried out 4 to 12 months after injury.

FIGURE 13-1 The anteroposterior and lateral elbow radiographs of a 42-year-old power lifter 8 weeks after a single incision repair of distal biceps tendon rupture. The forearm is ankylosed in mid pronation-supination.

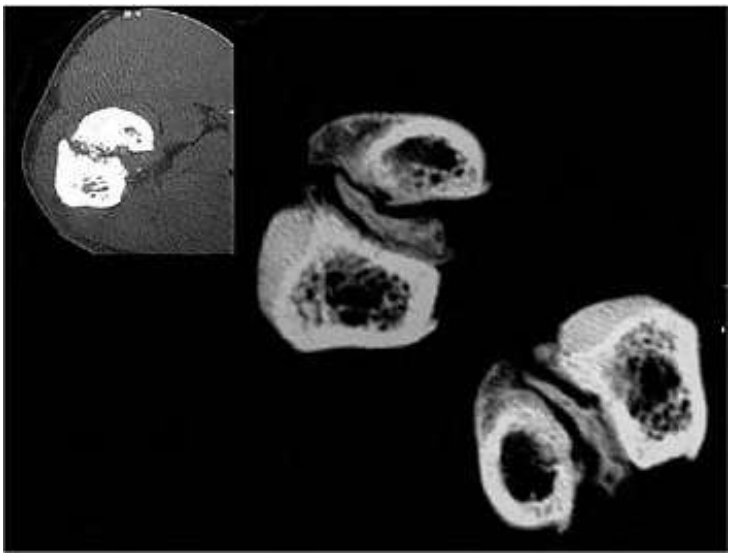

FIGURE 13-2 Coronal CT image reveals the extent of heterotopic bone involvement in the region of the bicipital tuberosity.

Excision of heterotopic ossification and soft tissue flap interposition does not require a great deal of special equipment. We have listed a few items that facilitate operative intervention and postoperative rehabilitation in Table 13-1. In addition to the tools listed in the table, we have the following items readily available: a hinged fixator, radial head prosthetics, allograft tendon for ligament reconstruction, allograft fascia lata if local interposition material is not available, and a total elbow arthroplasty set when all else fails and the patient is the appropriate age.

Surgery

Patient Positioning

Operative approach and patient positioning are determined by previous incisions, location of heterotopic bone or synostosis, and shoulder mobility. While many surgeons prefer to operate on the elbow with the affected extremity positioned across the patient's supine chest, we prefer to position

the patient with the affected extremity abducted onto a radiolucent table. The

extremity is draped free up to the clavicle. If the shoulder is mobile, external rotation will present the medial and anterior aspects of the elbow while internal rotation presents the lateral and posterior elbow. The flexibility to effortlessly alternate between lateral and medial approaches facilitates complete excision of heterotopic bone about the elbow.

Table 13-1. Soft Tissue Interposition Flaps for Elbow Heterotopic Ossification: Operative Equipment

Standard Equipment	Optional Equipment
Sterile tourniquet	Fluoroscopy
Retractors, thyroid type (long and narrow)	Hardware removal instruments
Lamina spreader	Total elbow arthroplasty
Vessel loops	Lidocaine infusion pump
Rongeurs (including Kerrison and pituitary)	Continuous passive motion machine
Osteotomes and curettes	Postoperative radiation therapy
Suture anchors	Static progressive vs. dynamic splinting
Hinged elbow distractor	
Tensor fascia lata allograft	
Closed suction drain	

Technique

Exposure

We do not use tourniquets for these cases. They are a detriment to tissue mobilization, mask small arterial bleeding, and lead to venous engorgement during prolonged cases. The initial increase in bleeding at the time of incision is mitigated by injecting the proposed incision with 0.25% bupivacaine and 1/200,000 dilute

epinephrine solution. The incision is delayed 7 to 10 minutes to allow the epinephrine to affect local capillaries.

It is our preference to use a posterior skin incision for elbow release surgery. This window allows a â€œglobalâ€• approach to the elbow and can be used to access the medial, lateral, and anterior sides of the joint. The skin incision is typically straight, passing 2 cm medial or lateral to the tip of the olecranon, but may be curvilinear to incorporate previously placed posteromedial or posterolateral surgical scars. Depending on the required heterotopic bone excision or hardware removal, the incision may extend from the proximal arm to the distal forearm. This dissection avoids injury to both the medial and the lateral brachial cutaneous nerves, preserving sensibility to the proximal forearm.

When employing this approach, it is critical to create thick soft tissue flaps. The skin incision is carried down to and includes the triceps fascia proximally and the extensor fascia of the forearm distally. This effectively creates robust fasciocutaneous flaps that can be elevated circumferentially about the elbow. The extent of elevation is dictated by the location of the joint involvement. If there is medial joint involvement, as determined by CT scan, or ulnar nerve symptoms, we address these first. If there is no involvement we go directly to the lateral elbow and proximal forearm. The dissection is carried to the medial intermuscular septum, and the ulnar nerve is identified as it passes from the anterior to the posterior compartment approximately 8 to 10 cm proximal to the medial epicondyle. Even if the ulnar nerve has been â€œtransposed anteriorlyâ€• in previous procedures, we believe that it is critical to identify the entire medial intermuscular septum, and follow it to the humerus (Fig. 13-3). In cases in which the medial intermuscular septum has been excised in part or in whole, the dissection follows the medial border of the triceps until the humerus is encountered. When the ulnar nerve is found, it should be mobilized from the cubital tunnel, preferably with a small cuff of medial triceps to protect the nerve's vascular supply. The nerve is followed into the forearm until disappeared deeply between the heads of the flexor carpi ulnaris. Any tight scar or fascial bands crossing the nerve in this dissection are divided. The medial intermuscular septum is removed in its entirety. By following the ulnar nerve proximally in the arm, the median nerve and its accompanying brachial artery can be located along the anterior margin of the intermuscular septum. Alternatively, in arms that are not densely scarred the median nerve and brachial artery can be identified in the distal arm superior to the leading

edge of the pronator teres origin as it runs medial to the substance of the biceps and brachialis muscles and just beneath the fibers of the bicipital aponeurosis. Excision of heterotopic bone involving the medial posterior and anterior elbow joint is conducted and described in greater detail in the next section.

Next, the lateral side of the elbow and involvement of the proximal forearm are addressed. In cases of anterolateral heterotopic ossification, the radial nerve is identified at mid-arm and followed distally. This is done by elevating the lateral flap in the same tissue plane as the medial dissection, between the muscular investing fascia and the muscle belly of the triceps. When the lateral intermuscular septum is encountered, the radial nerve should be identified and protected. Identifying the radial nerve in a scarred bed can be daunting. We therefore use the following strategies. First, inspect the undersurface of the flap; frequently the posterior antebrachial cutaneous branches of the radial nerve to arm and forearm are visible. These branches can be followed proximally into the lateral intermuscular septum where they are found to take off from the radial nerve proper (Fig. 13-4). If this landmark is not readily available, our second approach is to dissect the distal portion of the flap toward the lateral epicondyle. On reaching the lateral intermuscular septum, the dissection is directed cephalad. The triceps muscle belly is freed from the posterior aspect of the intermuscular septum. Small vessels and nerves seen entering the triceps muscle should be followed proximally; they will lead to the radial nerve proper. If these markers fail to lead to the nerve, cautiously proceed along the posterior intermuscular septum. Somewhere between 6 and 10 cm proximal to the tip of

the lateral epicondyle, the nerve will be encountered passing along the spiral groove to pass through the lateral intermuscular septum to enter the anterior aspect of the arm. Once the radial nerve is identified, the lateral intermuscular septum is removed and the nerve followed distally. At the level of the elbow, the nerve is easily followed into the internervous plane between the proximal aspect of the brachioradialis and the distal aspect of the brachialis. Simple blunt dissection between these two muscles just superior to the joint line will reveal the radial nerve before it dives under the supinator more distally.

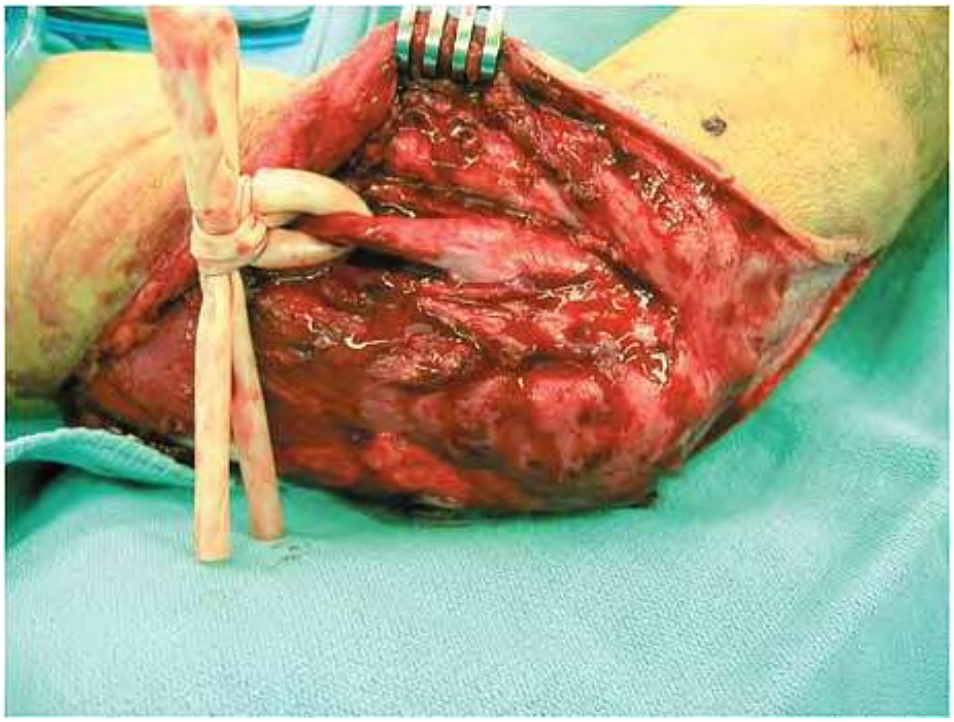

FIGURE 13-3 This previously â€œtransposed ulnar nerveâ€• sits directly on the medial epicondyle. It is essential to identify and protect the ulnar nerve before proceeding with capsular and heterotopic bone resection.

Excision of Heterotopic Bone

The excision of heterotopic bone and the scarred joint capsule follows. Anterior heterotopic bone frequently resides in the distal aspect of the brachialis muscle. It can extend medially and laterally encasing the collateral ligaments or even neighboring neurovascular structures. If a medial approach is chosen, entrance to the joint capsule and distal aspect of the brachialis can be achieved by exploiting the internervous plane between the ulnar innervated

flexor carpi ulnaris and the median innervated palmaris longus, flexor carpi radialis, and pronator teres. Alternatively, the flexor-pronator muscle mass can be sharply elevated off the medial epicondyle, carefully preserving the underlying anterior band of the medial collateral ligament. In severely contracted elbows, this exposure affords

uncompromised visualization of the anterior capsule and crossing neurovascular bundle. This elevated myofascial flap can be reattached to the medial epicondyle with multiple suture anchors or sutured down to the epicondyle through bone tunnels. If the flexor-pronator mass is released, consideration can be given to submuscular transposition of the ulnar nerve.

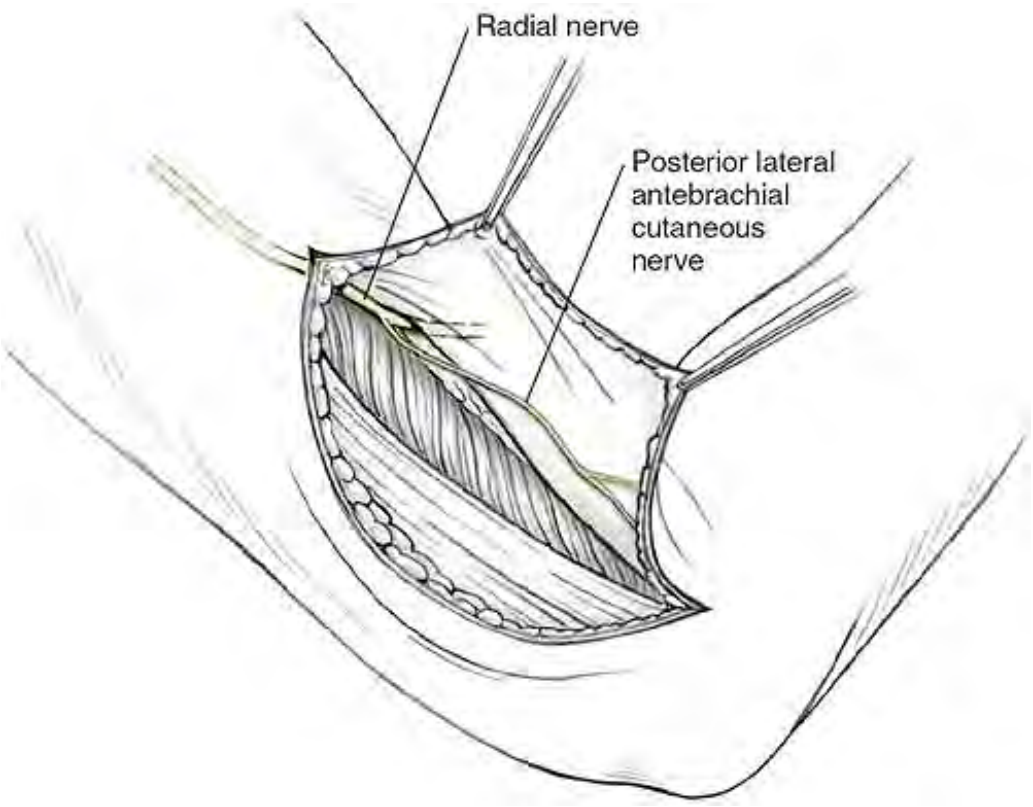

FIGURE 13-4 The posterior antebrachial cutaneous nerve is identified distally and followed proximally into the lateral intermuscular septum. This cutaneous nerve leads to the radial nerve proper. The radial nerve proper can then be followed from into the anterior compartment of the arm.

The anterolateral aspect of the elbow is approached next. If there is complete elbow flexion and extension, the dissection is directed toward the posterolateral forearm. In cases in which the anterior joint needs to be exposed, one of three intervals is used (Fig. 13-5). The first approach, and in our experience the most frequently used,

elevates the proximal most portion of the muscle taking origin from the medial epicondyle, usually the brachioradialis and a portion of the extensor carpi radialis longus. The dissection is carried medially, sweeping the brachialis from the front of the humerus. The interval between this muscle and the anterior joint capsule is developed. Heterotopic bone, if encountered, is dissected with blunt-tipped elevators and left attached to the anterior joint structures. If the relationship between these structures and the radial nerve is doubtful, the nerve is again identified in the distal arm and followed into the area of joint dissection. Under direct visualization, a blunt right-angled retractor is placed in the interval between the anterior joint dissection and the more superficially located muscle and nerve. The dissection stops distally when the coronoid process is encountered or the heterotopic bone becomes confluent the forearm bones. The entire anterior joint capsule and the heterotopic bone are removed. The coronoid fossa is cleared of soft tissues. Two other approaches to the anterolateral elbow consist of developing the interval between the anconeus and extensor carpi ulnaris, or the interval between the extensor carpi radialis brevis origin and the extensor digitorum communis. Both dissections allow easy access to the joint capsule overlying the radial head and neck. The capsular, radial, and annular portion of the lateral collateral ligament complex can be reflected or excised to expose the joint. The ulnar portion of the lateral collateral ligament, if not encased in ectopic bone, should be preserved (Fig. 13-6). More often than not, elevation of the lateral elbow complex from the ulna it is necessary. The ligamentous attachments will be reconstructed after the ectopic bone has been excised. If visualization of the anterior joint capsule is insufficient with these intermuscular approaches, the origin of the brachioradialis and extensor carpi radialis longus is elevated as described previously.

The posterior elbow joint is approached from the lateral side by elevating the triceps muscle from the lateral column of the distal humerus and developing the interval between triceps muscle and the posterior joint capsule. The triceps muscle insertion onto the olecranon is preserved. The posterior capsule, the contents of the olecranon fossa, and any bony impediments to elbow motion arising from theses posterior structures are debrided.

Proximal Radioulnar Synostosis Takedown

In cases in which pathology is limited to the proximal radioulnar joint in the vicinity of the bicipital tuberosity, the synostosis may be approached by sweeping the entire

anconeus and ulnar origin of the extensor carpi ulnaris complex off of the

posterolateral edge of the proximal ulna (Fig. 13-7Aâ€“C). The dissection along the proximal ulna is carried along the radial edge to the base of the synostosis and followed along the interosseous space until the radius is encountered. If the forearm can be pronated, there should be no danger to the posterior interosseous nerve. A pronated forearm allows the posterior interosseous nerve to lay anteriorly and medially protected in the substance of the supinator muscle. However, a supinated forearm does just the opposite. In cases where the forearm is ankylosed in supination, consideration should be given to locating the posterior interosseous nerve before the posterior dissection. This usually requires an anterior incision immediately adjacent to the ulnar border of the brachioradialis. The radial nerve is found in the interval between the brachioradialis and brachialis muscle and followed distally until it is seen to divide into the posterior interosseous nerve (PIN) and the superficial branch of the radial nerve. The PIN is followed distally into the supinator muscle. The arcade of Frohse is divided along with fibrosed muscle and scarred tissue that could lead to potential compression injury during the dissection. The posterior dissection is then resumed.

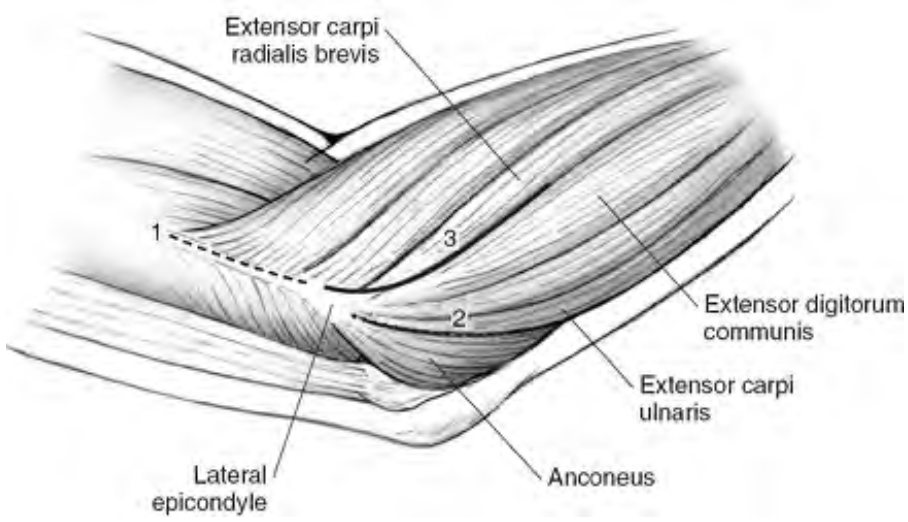

FIGURE 13-5 Three intervals used to approach the anterior lateral elbow. (1) Partial elevation of the muscles taking origin from the lateral epicondyle, the

brachioradialis, and cephalad portion of the extensor carpi ulnaris. (2) Anconeus and extensor carpi ulnaris. (3) Extensor carpi radialis brevis and extensor digitorum communis.

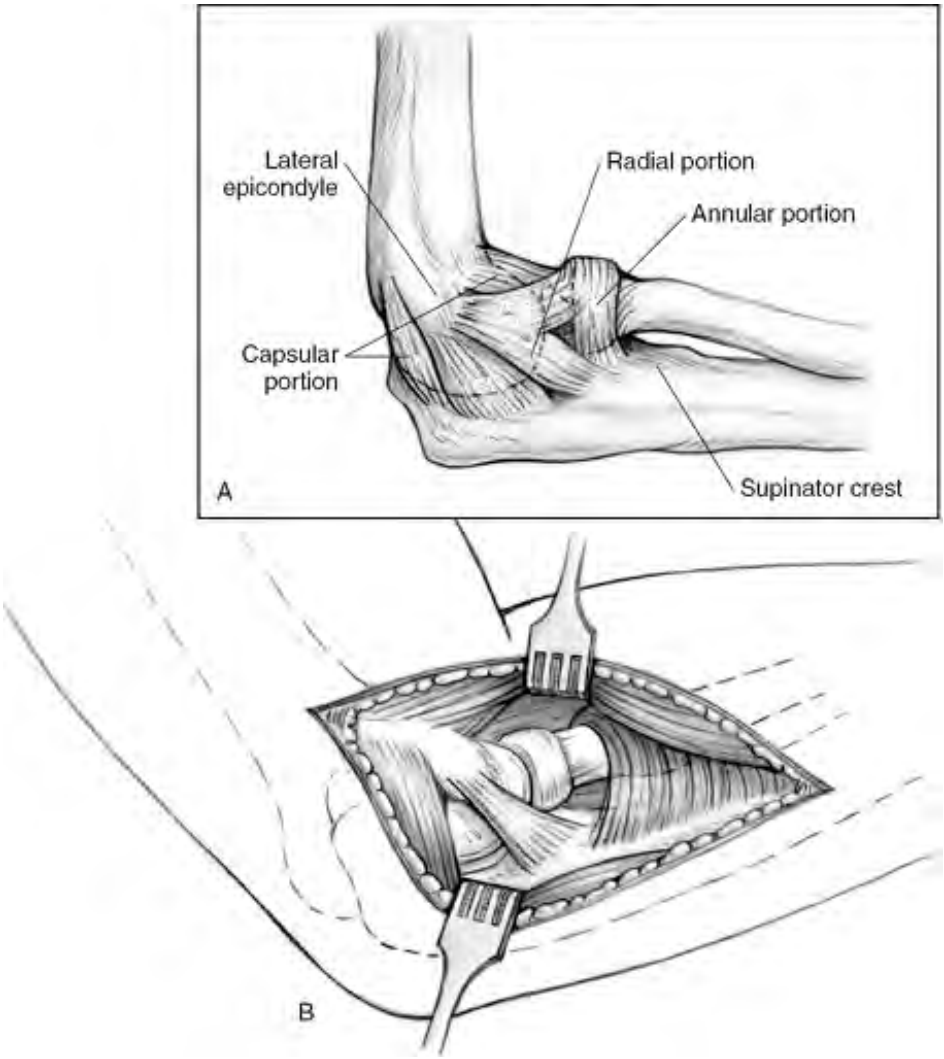

FIGURE 13-6 A: Lateral collateral ligament complex of the elbow consisting of the capsular, radial, and annular portions (*stippled*) removed to expose the lateral elbow articulations. B: The stippled portion of Figure 13-6A has been removed, exposing the radial head and neck, and the distal humerus. Preserving the ulnar portion of lateral collateral ligament (*shaded area*), when possible,

preserves elbow stability.

Once the anconeus is elevated, tissue retraction is maintained by broad right-angled retractors or broad Homan retractors. Narrow lever arm type retractors are avoided. Although anecdotal, it is thought that narrow retractors place excessive pressure on adjacent neurovascular structures, especially the PIN.

Synostosis resection proceeds using a combination of small rongeurs, osteotomes, and Kerrison rongeurs. Small lamina spreaders help retract the proximal radius and proximal ulna, opening up the space between. Once the synostosis has been released and there is freedom between the two bones, the lamina spreader offers greater direct visualization of remaining ectopic bone. The synostosis can now be excised down to the native cortices of the proximal radius and proximal ulna. It is imperative that the exposed bony surfaces be debrided to a smooth surface. Sharp edges combined with forearm rotation will penetrate and disrupt interposition flaps.

Intraoperative radiography is used to demonstrate the thoroughness of the ectopic bone excision. On completion of the resection, the surgical field is irrigated with 3 L of normal saline. Bleeding

P.158
P.159
P.160

points in soft tissues are coagulated, and bleeding points in exposed bone are sealed with bone wax. If the medullary canal of either the radius or ulna is breached, the hole is packed with allograft bone graft soaked in thrombin. This stops the marrow content from flowing into the operative field (see Fig. 13-7D,E).

A

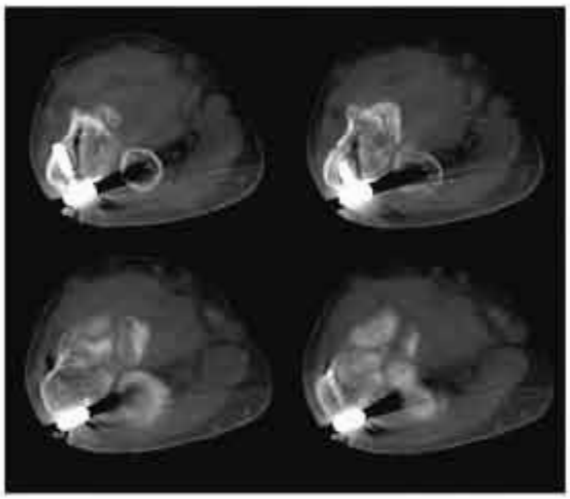

B

C

E

D

FIGURE 13-7 A: This patient developed a radioulnar synostosis after treatment of a Monteggia fracture. B: The CT scans demonstrate the location of the synostosis adjacent to the bicipital tuberosity. C: The entire anconeus and extensor carpi ulnaris complex have been elevated from the ulna (*dark arrow*), exposing the bridge of heterotopic bone between the radius and ulna (*light arrow*). D: The synostosis encased the bicipital tuberosity and tendon. The ectopic bone was removed, and the tendon left attached to a block of heterotopic bone (*dark arrow*). The resection exposed the medullary canal of the radius. The fixation plate was removed. E: The medullary canal is packed with allograft bone chips to minimize extrusion of marrow content. The biceps tendon and attached bone block are pushed into the medullary canal and secured with a stout nonabsorbable suture (not shown).

Interposition Material

Even with meticulous technique, the proximity of the proximal radius and ulna to each other makes these structures prone to recurrent synostosis. Some type of interposition material is a useful preventative adjunct. Three techniques are discussed in the following section. Each of these procedures is preceded with copious irrigation of the synostosis site and diligent hemostasis as described previously.

Pedicled Anconeus Myofascial Flap Interposition

The anconeus myofascial flap receives its blood supply from the collateral circulation about the elbow, predominantly the medial collateral artery (MCA) branch of the profunda brachii artery and venae commitantes (Fig. 13-8). The muscle is elevated from the posterior lateral border of the ulna, usually in a distal to proximal direction. The distal edge starts roughly at the junction of the proximal and mid one-third of the forearm.

A stout septum separates this muscle from the flexor carpi ulnaris. Identification of this septum and the anconeus itself may be difficult if previous surgery or initial trauma has injured the muscle. In these cases, an alternative material may be

required.

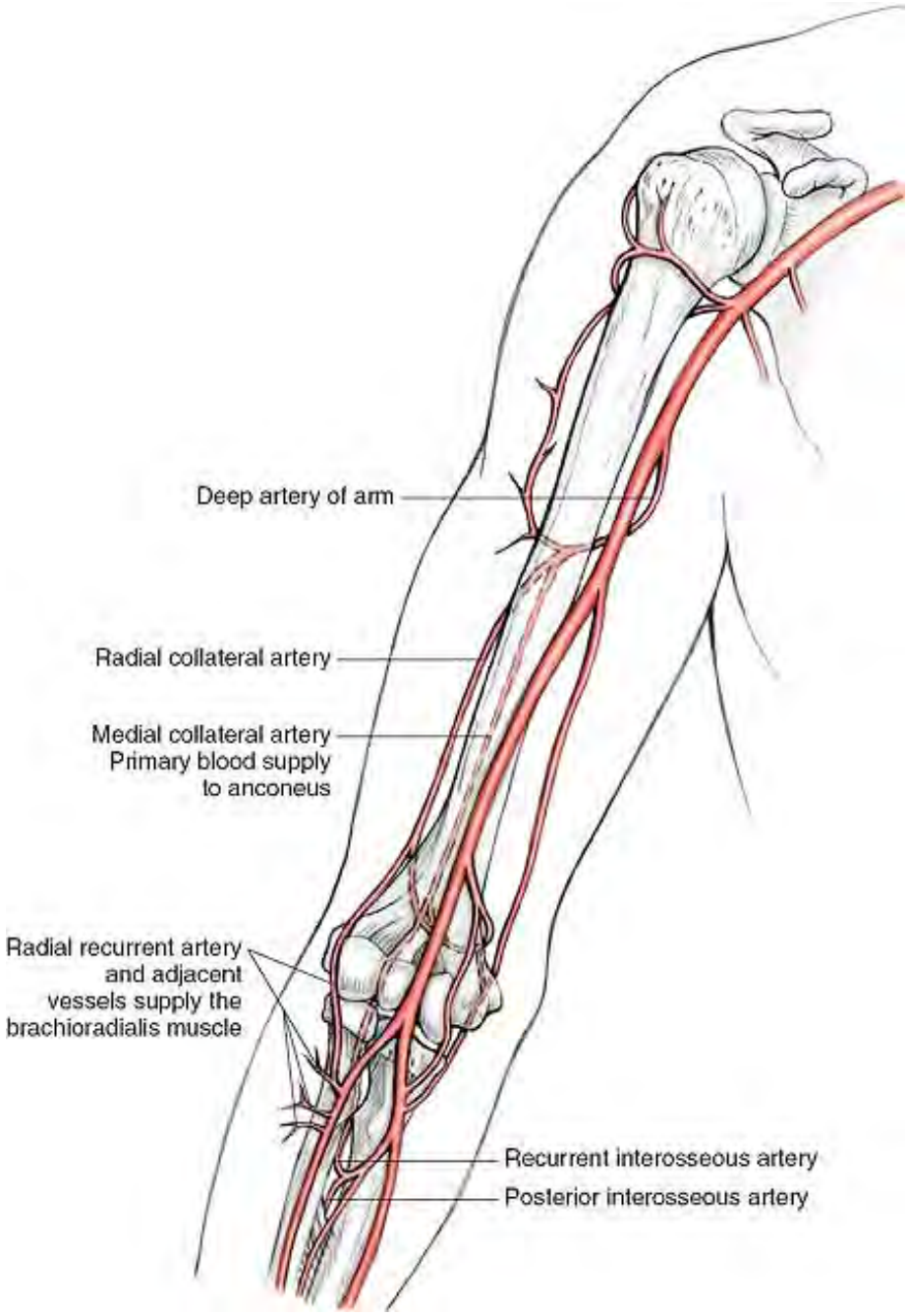

FIGURE 13-8 The arterial anatomy of the distal arm. The medial collateral artery (MCA), a branch of the profunda brachii, is the primary blood supply to the anconeus. The radial recurrent artery (RRA), a branch of the radial artery, and

small arterial branches within 3 cm of the RRA, provide the primary blood supply to the brachioradialis muscle.

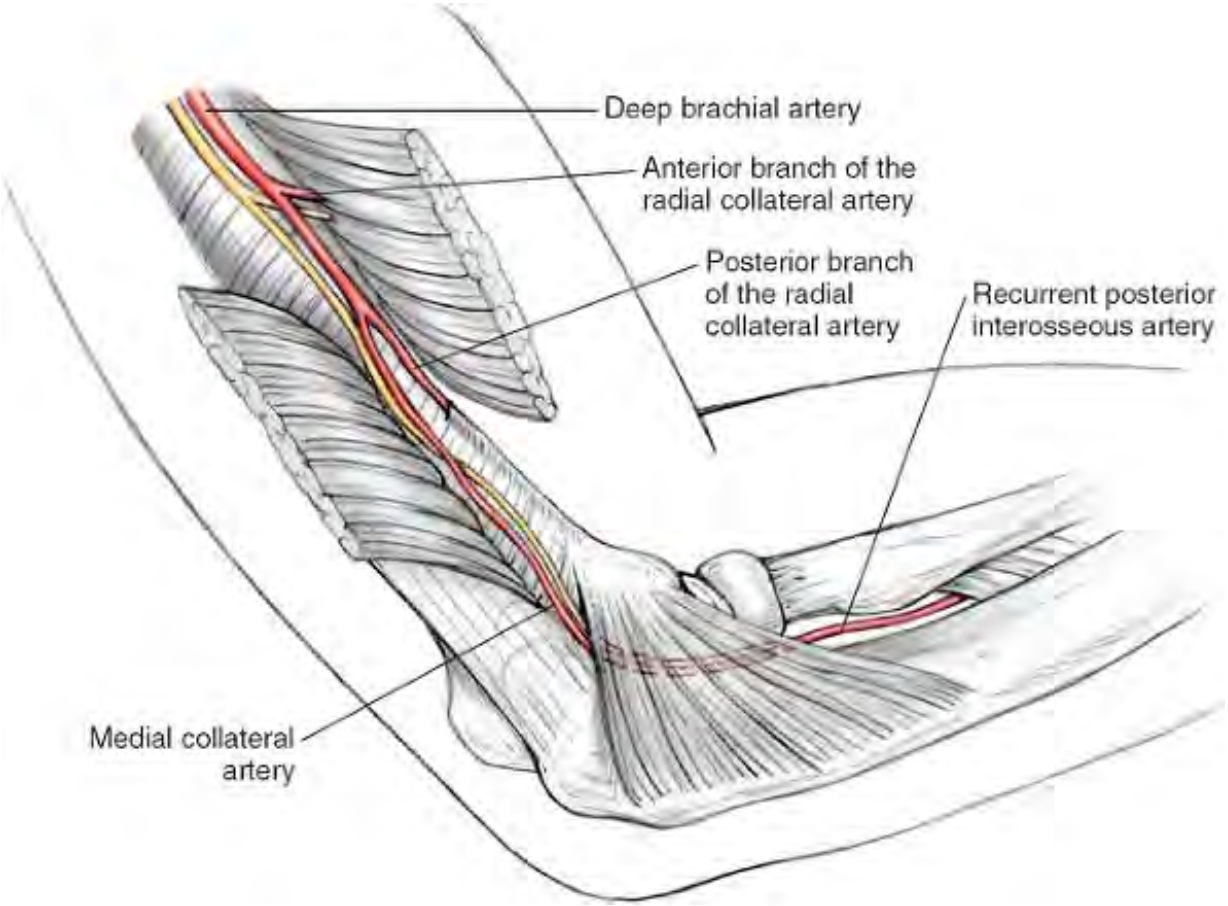

FIGURE 13-9 In addition to the medial collateral artery (MCA), the anconeus receives blood supply distally from the recurrent posterior interosseous artery (RPIA). These vessels (the MCA and RPIA) frequently form an anastomosis on the deep surface of the anconeus. The RPIA is frequently sacrificed in the process of elevating this flap. (From Schmidt CC, Kohut GN, Greenberg JA, Kann SE, Idler RS, Kiefhaber TR. The anconeus muscle flap: its anatomy and clinical application. *J Hand Surg Am.* 1999;24(2):359-369. With permission.)

Elevation of the flap is carried up to the lateral epicondyle until sufficient freedom exists to permit mobilization of the tissue's leading edge to the distal extent of the synostosis. By protecting the proximal origin and the lateral fascial attachments of the triceps, the medial collateral artery of the elbow can be preserved. The anconeus also receives blood supply distally from the recurrent posterior interosseous artery (RPIA). These vessels (the MCA and RPIA) frequently form an anatomosis on the deep surface of the anconeus (Fig. 13-9). The RPIA is usually sacrificed in the process of removing the synostosis or in elevating this flap.

The distal end of the anconeus flap is attached to the ulnar edge of the biceps tuberosity while the forearm is placed in full pronation. Sutures alone or suture anchors placed into the proximal radius or biceps tendon secure the anconeus flap with mattress or Bunnel locking stitches. Now, as the forearm is supinated, the anconeus will be drawn into the proximal radioulnar space where the synostosis previously layâ€”creating a thin, pedicled interposition flap of viable muscle and fascia (Fig. 13-10).

The utility of the anconeus myofascial interposition flap has recently been expanded by Morrey and Schneeberger to include soft tissue interposition between the capitellum and the proximal radius in cases of radial head excision or failed radial head arthroplasty (Fig. 13-11). Three interposition options are described: type I, interposition at the radiocapitellar joint; type II, the muscle is brought

posterior to the intact fibers of the ulnar portion of the lateral ulnar collateral ligament and interposed between radiocapitellar and proximal radioulnar joint; and type III, proximal radioulnar wrap. In each type of interposition, the muscle is secured with stout suture placed through drill holes.

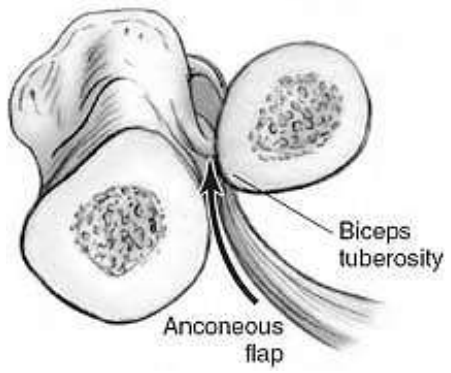

FIGURE 13-10 With the forearm placed in full pronation, the distal end of the anconeus flap is attached to the ulnar edge of the biceps tuberosity (BT) with a suture anchor or direct stitch into the biceps tendon. As the forearm is supinated (*arrow*), the anconeus will be drawn into the proximal radioulnar space.

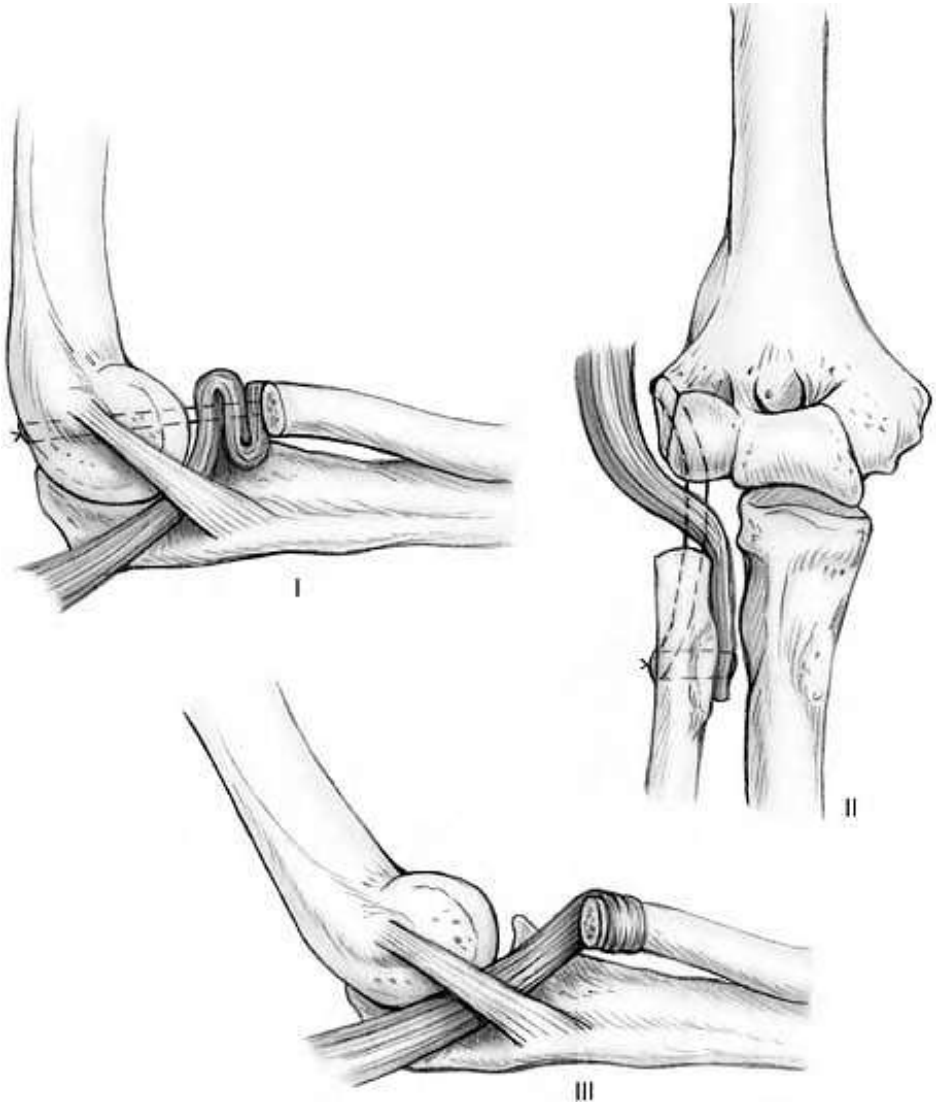

FIGURE 13-11 Three interposition options for the anconeus flap after radial head resection: type I, radiocapitellar joint: type II, radiocapitellar and proximal radioulnar joint; and type III, proximal radioulnar wrap. (From Morrey BF, Schneeberger AG. Anconeus arthroplasty: a new technique for reconstruction of the radiocapitellar and/or proximal radioulnar joint. *J Bone Joint Surg Am.* 2002;84-A(11):1960â€“1969. With permission.)

Distally based anconeus myofascial flaps composed of anconeus and some neighboring extensor carpi ulnaris have been described. This flap depends on an intact RPIA for its

blood supply, a small vessel that is easily injured during the initial fracture fixation or during resection of a synostosis. Distally based anconeus flaps may have application in the management of soft tissue defects not associated with elbow fractures such as burns or skin avulsion but should be used with caution in the setting of synostosis resection.

The proximally based anconeus pedicle flap can also be used to provide soft tissue coverage of small defects about the traumatized elbow. Three areas where this muscle will reach is the lateral epicondyle, the posterior surface of the olecranon, and the distal radial aspect of the triceps insertion into the olecranon. The muscle is elevated, rotated over the defect, and covered with split thickness skin grafts.

Pedicled Brachioradialis Myofascial Flap Interposition

The brachioradialis interposition flap is useful when resection of a radioulnar synostosis is performed through a combined anterior and posterior approach. The posterior incision and approach is the same as described previously: the anconeus-extensor carpi ulnaris interval is developed, the supinator sharply elevated from the ulna, and the synostosis excised. The anterior approach is conducted through a longitudinal incision made along the ulnar border of the brachioradialis, from the elbow flexion crease to the distal forearm. The superficial branch of the radial nerve is located in the distal forearm and carefully dissected from the undersurface of the brachioradialis. This nerve is followed proximally into the cephalad region of antecubital fossa where it is found to branch from the radial nerve. At this level the radial nerve branches to the brachioradialis, and extensor carpi radialis muscle may be encountered

and is protected. Most important, the posterior interosseous nerve branch is identified and followed into the substance of the supinator muscle. The stout fibers on the leading edge of the supinator, the arcade of Frohse, are a potential source for compression during the remainder of the procedure and are divided. Simultaneously with the dissection of the radial nerve, the radial artery and its branches to the lateral elbow are encountered and protected. The largest of these branches, the radial recurrent artery, is found crossing from ulnar to radial, just distal to the biceps tendon (see Fig. 13-8). Care is taken to preserve not only the radial recurrent artery, but also the branches of the radial artery just distal to the radial recurrent artery. While the

proximal aspect of the brachioradialis muscle belly is vascularized by the radial recurrent artery, the distal aspect of the muscle receives its blood supply from the arborization of the radial artery approximately 3 cm distal to the radial recurrent artery. In this dissection the median nerve and the ulnar artery are shielded by the medially retracted flexor-pronator muscle mass on the ulnar aspect of the proximal forearm.

Once these pertinent neurovascular structures are dissected free and protected, attention can be turned to the synostosis excision. Typically, the heterotrophic bone is found within or deep to the supinator muscle as it extends between the proximal radius and ulna. This anterior approach is helpful to visualize and debride the full extent of bridging bone in complex or revision cases of synostosis resection. After the synostosis has been taken down and the exposed bone edges covered with bone wax, the pedicled brachioradialis myofascial interposition flap is raised (see Fig. 13-12).

The tendon of the brachioradialis is divided and the muscle raised from distal to proximal. The previously identified neurovascular structures are left intact. Although the muscle can be released proximally from its origin on the lateral epicondyle and lateral intermuscular septum, this is usually not necessary; the muscle typically provides enough length for interposition between the proximal radius and ulna.

The muscle is interposed between the radius and ulna from anterior to posterior and secured in one of two ways. The first is to secure muscle to the posterior surface of the fully supinated radius using suture anchors or drill holes. When the forearm is pronated, the muscle is advanced into the interosseous space. The second method, introduced by Diego Fernandez, wraps the muscle around the proximal radius, deep to the superficial branch of the radial nerve, and is sutured to itself with locking sutures. Passing the stitches through the muscle fascia as well as through some of the tendonous fibers optimizes the fixation of this interposition flap to the bone.

In addition to being used for an interposition material, the brachioradialis can be used as a myofasciocutaneous flap to provide coverage for medium-sized elbow defects. The skin over the brachioradialis, centered on the radial recurrent artery, may be harvested up to 2 to 3 cm in width and 6 to 10 cm in length from proximal to distal. A narrow skin paddle permits primary closure of the donor site. Alternatively, the muscle alone can be elevated, inset, and split thickness skin graft is applied over it. The brachioradialis flap can be safely rotated to reach soft tissue defects involving the

antecubital fossa, the volar one-third of the forearm, and the posterolateral aspect of the elbow between the lateral epicondyle and the olecranon.

Tensor Fascia Lata Interposition

There are certainly cases of heterotopic ossification and radioulnar synostosis in which transposing a local muscle flap may be difficult or impossible. Most frequently these are cases in which the anconeus was injured during the initial traumatic event and where the extensive dissection of the brachioradialis muscle and additional scars are avoidable. In such circumstances, an alternative approach is to use tensor fascia lata (TFL) for interposition. Autograft TFL has the advantage of biocompatibility but the marked disadvantage of donor site morbidity including superficial nerve injury, additional site of scarring, and asymmetry of the thigh secondary to muscle herniation. Given the potential complications and the reality that the interposition graft (whether autograft or allograft) is not living tissue after harvest, we have abandoned the use of autograft TFL in favor of allograft.

The technique is straightforward. Following synostosis exposure and resection, the tensor fascia lata allograft is shaped to cover the exposed bone. A graft 10 to 12 cm long and 4 to 5 cm wide is required in most cases. The tensor fascia lata graft is wrapped around the most accessible bone, usually the ulna, and sutured into place using absorbable sutures (see Fig. 13-13). With a complete wrap, the graft can simply be sutured to itself or neighboring soft tissues to maintain its position. Another technique is to suture the graft to the exposed surface of the ulna, interpose it between the radius and ulna, securing to the exposed surface of the fully pronated radius with suture anchors. As the forearm supinates, the graft follows the radius and provides an effective barrier to reformation of the synostosis. The TFL, once secured to the underlying bone, creates an intact, but not water tight, sleeve

P.164
P.165
P.166

around the proximal aspect of one of the forearm bones. The goals of this type of fascial interposition are twofold: to provide painless forearm rotation and to act as a barrier to future bridging heterotopic bone.

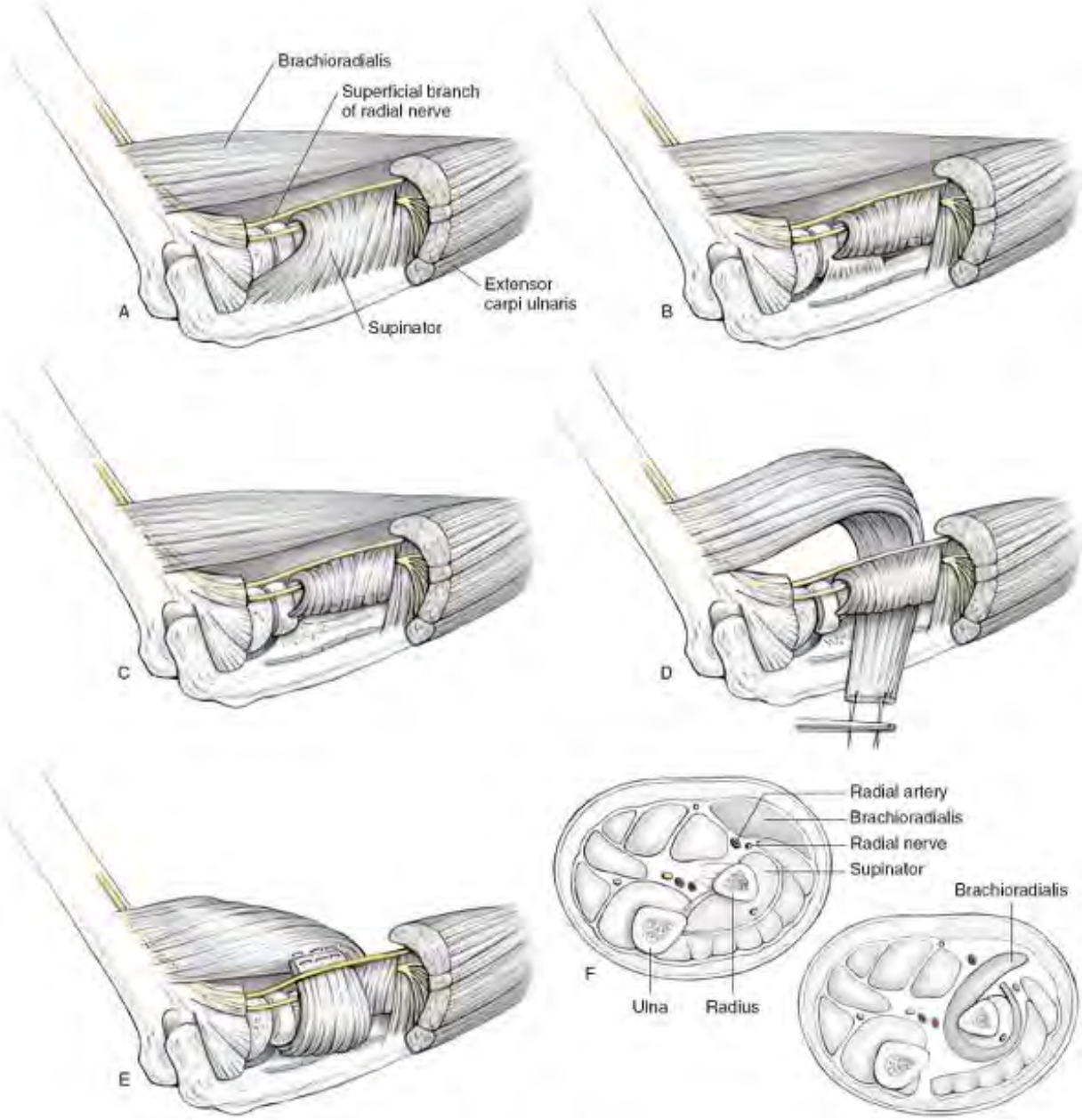

FIGURE 13-12 A: Posterior lateral approach (Kocher). Through a separate incision, the interval between the anconeus and extensor carpi ulnaris is developed. B: The supinator is sharply elevated off of the ulna exposing the radioulnar synostosis. C: The synostosis is removed and if necessary the interosseous ligament-membrane is excised. D: A separate anterior incision runs just ulnar to the brachioradialis muscle. The tendon is divided distally, the superficial branch of the radial nerve is separated from the under belly of the

muscle, and the vascular bundles entering the muscle in the proximal forearm are preserved. The muscle is passed through the interosseous space created by the resection. E: The tail of the muscle is brought deep to the radial nerve and sutured to the proximal muscle belly. F: Cross-section of the forearm. Note the muscle passes radial to the radial artery in the mid forearm, distal to the biceps tendon and deep to the superficial branch of the radial nerve. (From Fernandez DL, Joneschild E. â€œWrap aroundâ€• pedicled muscle flaps for the treatment of recurrent forearm synostosis. *Tech Hand Up Extrem Surg.* 2004;8(2):102â€“109. With permission.)

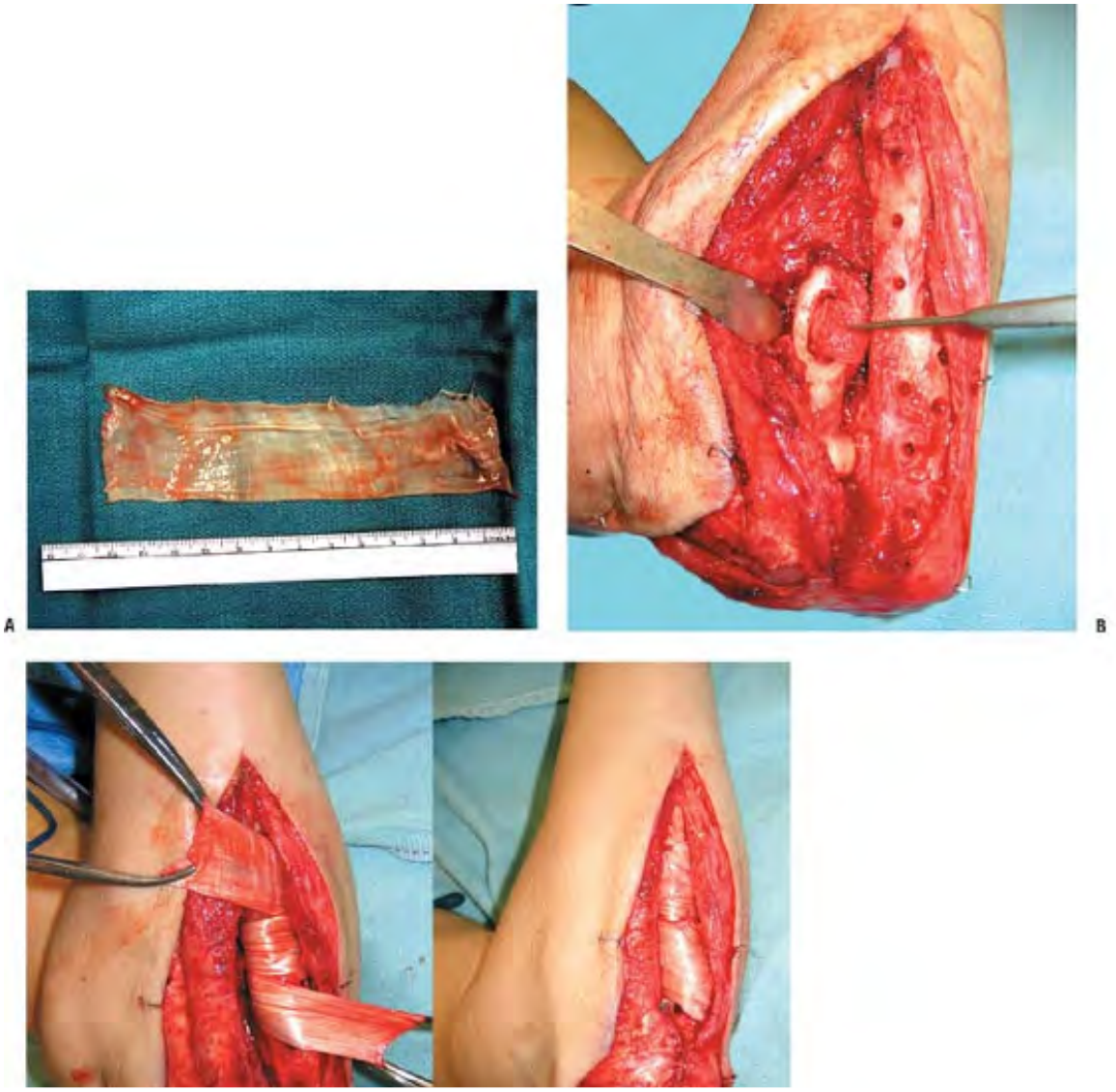

FIGURE 13-13 A: This 4 Ã— 14 cm tensor fascia lata allograft will be used as an interposition material. B: This is the forearm of the patient shown in Figure 13-7. After synostosis resection, the rough surfaces of the radius and the ulna are prone to reformation of the synostosis. C: The tensor fascia lata allograft is wrapped around the ulna and secured with stout absorbable sutures. D: Radiographs of the elbow 6 months postoperatively. E: Range of motion 6 months postoperatively.

Closure

After heterotopic bone excision from an elbow, medial or lateral instability frequently arises. If the collateral ligaments have been excised, they should be reconstructed. Allograft tendon provides an alternative to harvesting autogenous palmaris longus or plantaris for these patients. Using either form of collagen, a tendon reconstruction can

be performed with a combination of suture anchors and bone tunnels. Care should be given to reconstructing the anterior band of the medial collateral ligament and the ulnar band of the lateral collateral ligament, as these ligaments have been shown to be instrumental in the maintenance of elbow stability. If stability cannot be achieved with ligamentous reconstruction, a dynamic external fixator or dynamic traction fixator should be applied.

Closed suction drains anterior and posterior to the elbow joint are an integral part of minimizing postoperative swelling and hematoma formation. It is not necessary to place a drain in the site of a synostosis takedown, but it should be placed deep to the fascial layer. Before closing the deep soft tissues, a lidocaine or bupivacaine continuous infusion pump may be placed to help maximize postoperative pain control (in cases in which axillary catheters are not used). The catheter for a bupivacaine infusion pump should be placed in a soft tissue layer or location separate from that of the closed suction drains. Divisions in the intermuscular septi do not need to be sutured, but muscle attachments to the distal humerus or proximal ulna should be repaired. For origins and insertions of

muscle about the medial and lateral elbow, suture anchors provide sufficient stability for the bony reattachment. If the medial, flexor-pronator muscle mass has been elevated, consideration can be made to submuscular transposition of the ulnar nerve. Otherwise, a simple subcutaneous transposition maintained with a small strip of flexor-pronator fascia frequently will suffice. The fascial layer may be repaired with absorbable suture in a running or interrupted fashion. Subcutaneous and skin closure follows. The skin closure should not be under tension, nor should it be so tight as to prevent egress of drainage when early motion is initiated.

Postoperative Management

Surgical outcomes of heterotopic bone excision and interposition soft tissue flaps weigh heavily on the patient's postoperative rehabilitation. The affected upper extremity is immobilized in a soft dressing to allow immediate motion. An axillary block catheter placed by the anesthesia team can provide substantial pain relief in the immediate postoperative setting. When such a catheter cannot be placed, we have used continuous bupivacaine infusion pumps to supplement intravenous narcotic pain control.

With adequate analgesia, continuous passive motion of the elbow can be performed in flexion and extension. It is our preference to use a bedside continuous passive motion (CPM) machine for the elbow rather than the more mobile units that provide motion during ambulation. The bedside units tend to maintain a constant position, capitalizing on leverage to impart motion at the patient's elbow. We have not identified a reliable method for continuous passive forearm motion and instead rely on intermittent sessions wherein the extremity is removed from the CPM machine, the elbow flexed 90 degrees, and forearm motion addressed. Resting splints are worn while the patient sleeps at night. The position of the splint is dictated by the particular problems addressed with surgery. For patients with elbow flexion and extension problems, the limb is immobilized in maximum extension and supination. For those patients whose primary problem is forearm motion, the limb is immobilized with elbow flexed 90 degrees and the forearm in maximum supination. Patients are usually discharged on the fourth day postoperatively, and an aggressive home and outpatient therapy program is instituted.

Elbow and forearm specific exercises include active, active-assisted range of motion as well as gentle passive assisted range of motion. To minimize the â€œtrick maneuverâ€• of radiocarpal and intercarpal rotation to augment apparent pronation and supination, the wrist is secured in an immobilizer and a solid object such as a common household tool, a hammer, is grasped tightly in the hand. Doing this directs all pronation and supination efforts to the forearm. When not exercising, the patient should use a static progressive splint, especially at night. The splint is applied an hour before sleep, and the pain associated with initial application should have dissipated enough that sleep will not be interrupted. Dynamic rubber band or spring driven splints are never used. Splinting and exercises are rarely carried out past 6 months postoperatively, as there is usually no change in the range of motion past this date. Anecdotally, however, patients frequently report that even though their range of motion does not change after 6 months, their strength and ease of motion improves up to 2 years postoperatively.

Radiation treatment of the resected bed of heterotopic ossification or synostosis has been shown to be successful in preventing recurrence of ectopic bone. However, no prospective randomized trial has shown this modality to be better than resection alone. As this procedure comes with some inherent risk of wound breakdown, neuritis,

and lymphedema, and a remote risk of sarcoma, it should be discussed with the patient preoperatively. We limit radiation therapy to a single dose of 700 cGy within 36 hours of surgery and only in those patients in whom there was ankylosis of the proximal radioulnar joint. All patients are given ketorolac 30 mg IV daily for the 4 days; beyond that, non-steroidal anti-inflammatory medicines are not used.

Complications

While the release of a stiff or ankylosed elbow can provide dramatic improvement in a patient's upper extremity function, the procedure is not without its complications. It is a high-risk operation which requires strong appreciation for the anatomy of the elbow and traversing neurovascular structures. Nerve injury from retraction or surgical dissection is possible and can be minimized with frequent assessment of nerve location as well as limiting the time and rigorousness of retraction on a

nerve. Loss of fixation of the interposition material can be avoided by careful suture fixation into muscle as well as fascia.

Certainly, the most frequent complication of this procedure is recurrence of the heterotopic bone. In many cases new calcifications arise on postoperative radiographs well into the rehabilitation period. If the physical examination demonstrates no signs of motion loss, these ectopic calcifications are clinically insignificant and donâ€™t require routine followup. Some motion loss from the gains made intraoperatively can be expected. But a majority of patients who actively participate in their daily motion exercises will maintain a functional arc of motion and achieve an optimal outcome.

Recommended Reading

Alonso-Llames M. Bilaterotricipital approach to the elbow. Its application in the osteosynthesis of supracondylar fractures of the humerus in children. *Acta Orthop Scand.* 1972;43(6):479â€“490.

Bell SN, Benger D. Management of radioulnar synostosis with mobilization, anconeus interposition, and a forearm rotation assist splint. *J Shoulder Elbow Surg*. 1999;8(6):621â€“624.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 14

Principles of Hand Incisions

Marco Rizzo

William P. Cooney III

Indications/Contraindications

Because of the intricate balance of function in the hand and wrist, incisions in this area are unique. In an effort to ensure healing and maintain function, it is important to plan the incisions appropriately. The fingers have well-established approaches that allow for good visualization of deeper structures without compromising healing. Likewise, the hand and wrist are best approached while following basic principles. Poorly placed incisions result in restrictive scarring (Fig. 14-1), which may lead to diminished motion and function. In addition, the limited skin redundancy on the palmar surface of the hand may make delayed closure of edematous wounds difficult. Sound planning of incisions about the hand and wrist will help to avoid some of these difficulties. The purpose of this chapter is to review accepted incision designs for various aspects of the hand and wrist.

Preoperative Planning

Volar Approach to the Fingers

There are several commonly used approaches to the volar aspect of the fingers and thumb. The traditional Bruner incision is the approach that is most common (1). It is a

zigzag incision that angles at the flexion creases of the metacarpophalangeal (MP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints (Fig. 14-2). As the incision progresses proximally, the flexion creases of the palm are also points of direction changes. The principle is based on the fact that longitudinal incisions across these creases can generate excessive scarring and limit extension. For optimal healing of the flaps, it is best to keep the angles close to 90 degrees. Angles more acute than 60 degrees carry a higher incidence of skin necrosis.

FIGURE 14-1 Example of restrictive scar resulting from straight line incision placed over the flexor surface of the finger.

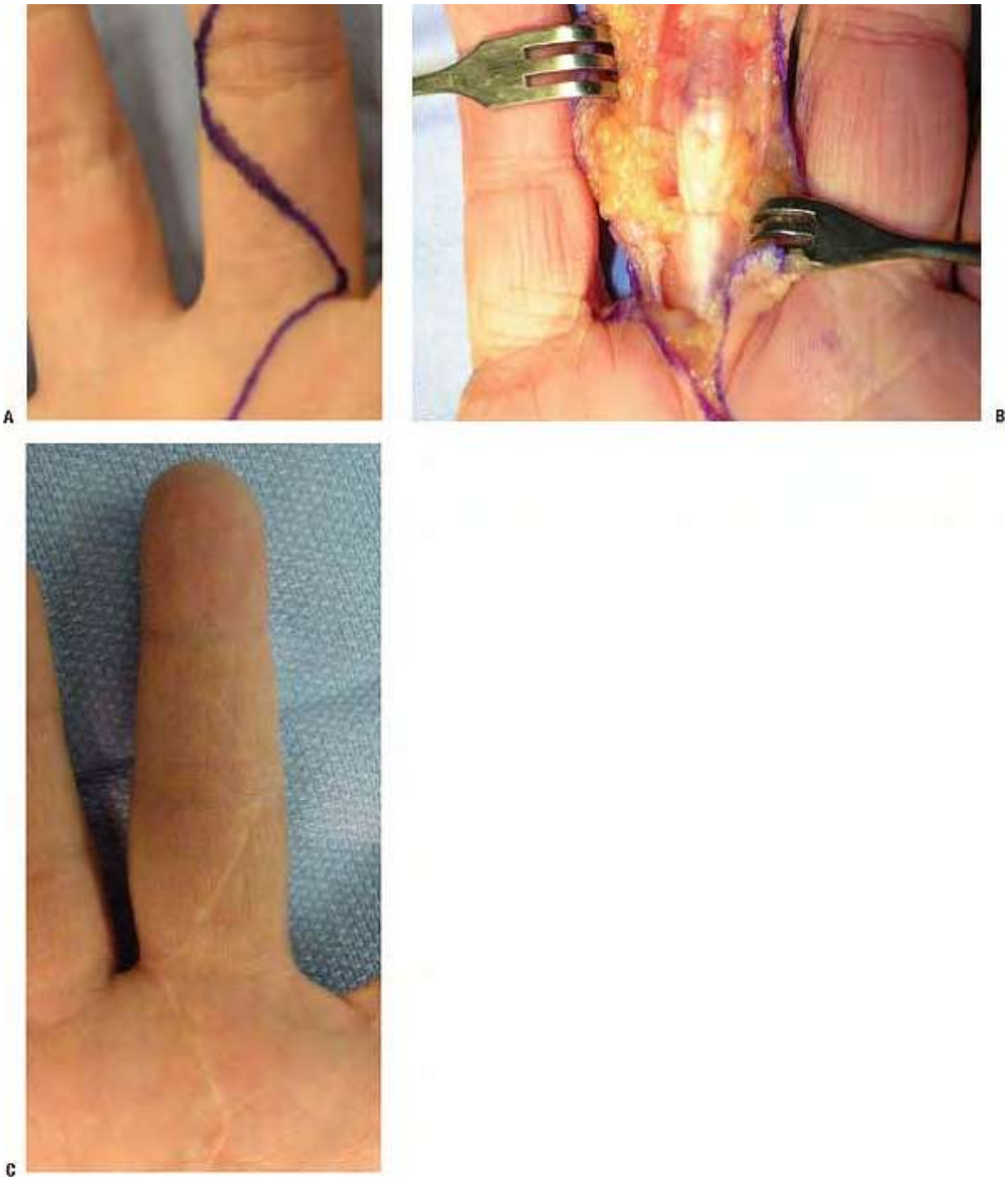

FIGURE 14-2 A: Bruner incision approach to the volar aspect of the fingers. The most commonly preferred method of crossing flexion creases is with these undulating angled incisions. B: Excellent exposure to the volar aspect of the finger can be achieved. C: These scars typically heal nicely.

An alternative to the Bruner incision in the finger is the mid-lateral approach (Figure 3A). It has the advantage of being cosmetically more appealing. The incision is based about the lateral (or medial) side of the finger and the skin is elevated to provide exposure of the underlying tissues. A good way to ensure that the surgeon is not directly over the neurovascular bundles is to connect the incisions at the apex of the flexion creases (Fig. 14-3B). This is also a preferred approach in cases of replantation surgery. Although cosmetically appealing, the challenges of this technique include elevating the skin and ensuring its viability.

In cases of flexion deformity of the fingers such as Dupuytren's contracture, a straight midline incision with z-plasties is an alternative approach to the traditional Bruner incision (Fig. 14-4). This will allow for the elongation of the incision as the flexion deformity is corrected. Angles of 60 degrees help preserve viability at the apex of the skin flap.

Incisions over the distal pulp of the finger, such as is necessary in felons, are best made directly over the area of swelling and induration. These wounds can be loosely closed or allowed to heal via secondary intention. If it appears that the incision may need to be extended proximal to the distal interphalangeal joint, a Bruner incision may be preferred for better exposure.

Dorsal Incisions About the Fingers

The dorsal skin of the hand lacks glaborous connections and has more redundancy, due to the requirements of joint flexion. Because of this redundancy, longitudinal incisions are acceptable. A longitudinal incision allows (Fig. 14-5) for protection of the venous and lymphatic drainage, while giving good visualization of the extensor mechanism (2). Lazy S or curvilinear incisions may also be designed to avoid crossing extension creases (see Fig. 14-5). Incisions over the distal interphalangeal joint can be longitudinal as well, but dissection can be limited due to the nail bed and fold distally. An alternative is the â€œTâ€• or â€œHâ€•-shaped incision that allows for visualization of the extensor mechanism and distal interphalangeal joint (Fig. 14-6).

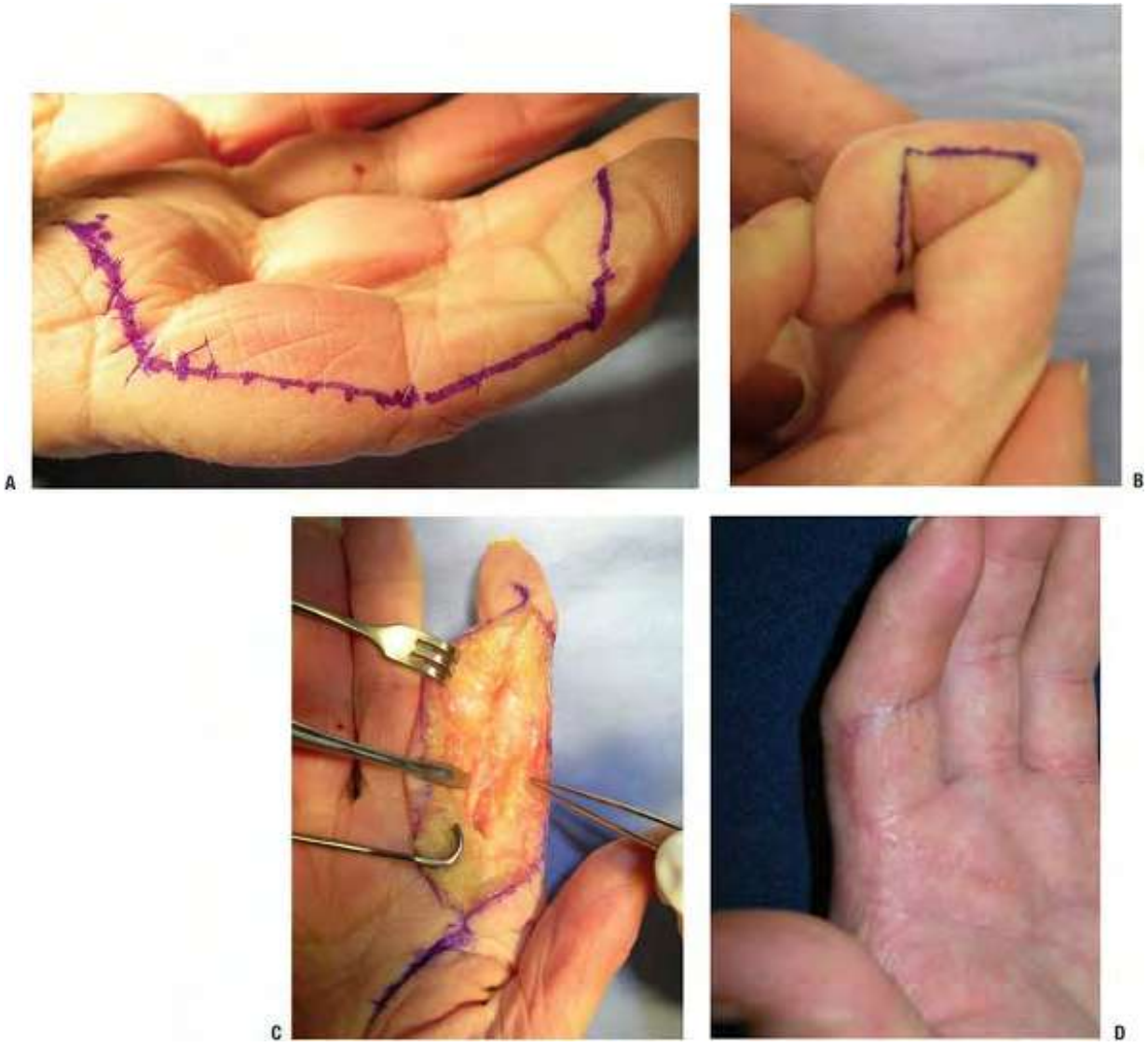

FIGURE 14-3 A: The mid-axial or lateral approach to the fingers. This technique can be a more aesthetic alternative to the Bruner incision. B: Care must be taken to connect the incision at the apex of the flexion creases in order to avoid direct dissection over the underlying neurovascular structures and prevent excessive volar scarring. C: Although it requires more dissection, excellent exposure can be achieved. D: An example of a healed mid-lateral incision.

Volar Incisions About the Hand

Transverse incisions or incisions that parallel the flexion creases offer the best cosmetic results. However, these incisions are perpendicular to the underlying vessels, nerves, and tendons. Therefore, care must be taken when performing these approaches. Incisions for trigger finger releases are a good example (Fig. 14-7). In particular, transverse incisions, in line with the metacarpophalangeal joint flexion crease, at the base of the thumb for A1 pulley release have been found to produce better aesthetic results and result in less restrictive scarring. Traditional longitudinal Bruner-type incisions are commonly used and are useful when more exposure is necessary. Again, you want to follow the basic premise of changing directions at the creases to minimize scarring. Wounds or scars

P.173

P.174

in the web spaces are at risk of developing functionally limiting scars. Depending on the patient, these can sometimes necessitate revision z-plasties and web-space deepening.

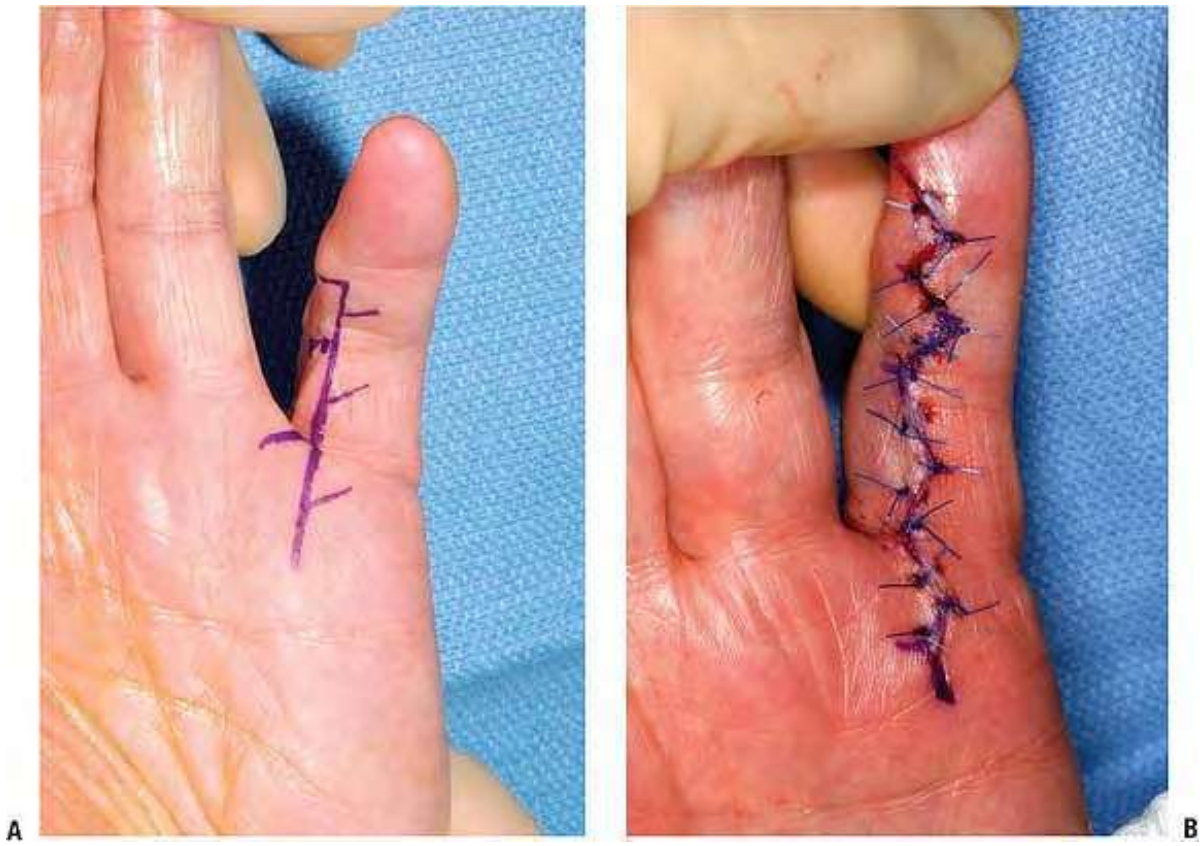

FIGURE 14-4 A: A longitudinal z-plasty approach to the volar aspect of the finger can be useful in cases of pre-existing flexion contracture, such as in Dupuytren's contracture. B: Wound closure is easily achieved with little or no tension with the finger in full extension.

FIGURE 14-5 Two common approaches to the dorsum of the fingers: a straight longitudinal and a more curved longitudinal, which avoids the extension creases.

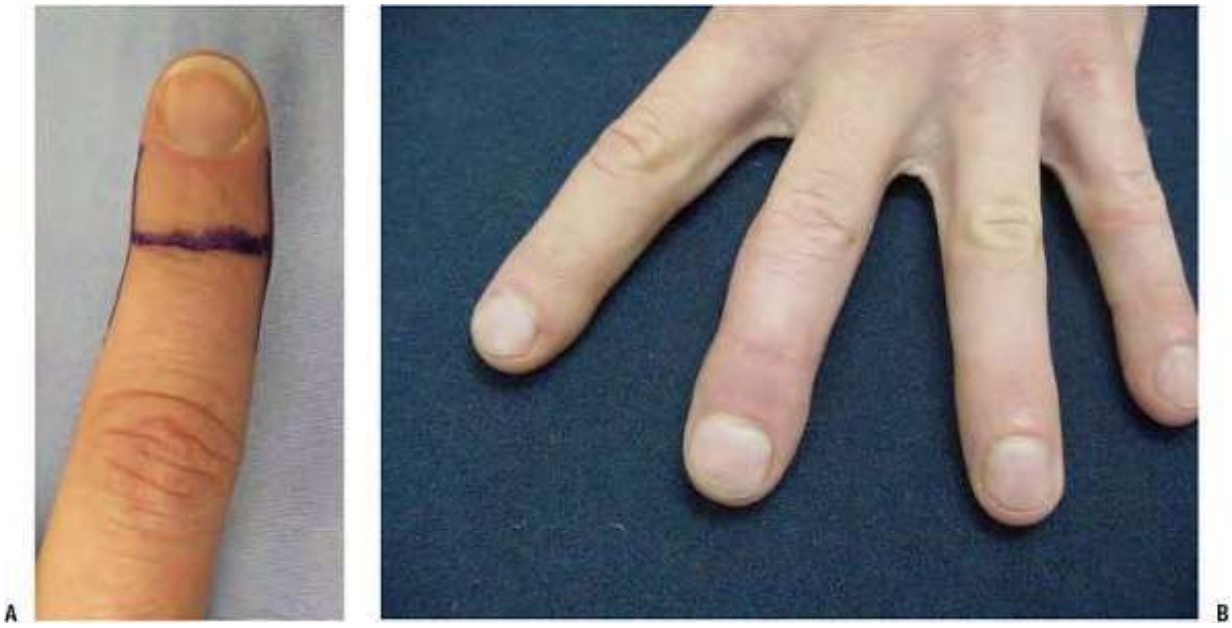

FIGURE 14-6 A: An â€œHâ€•-shaped incision over the dorsal aspect of the distal interphalangeal (DIP) joint allows for excellent exposure of the dorsal distal finger. B: The wounds typically heal very nicely.

In exposing the carpal tunnel, a longitudinal incision, preferably in or in line with the palmar crease, is commonly used. If the incision needs to be extended proximally, Bruner-type zigzag incisions can be made between the distal and proximal wrist flexion creases (Fig. 14-8). These angled incisions are the most common and useful method of crossing flexion creases in the hand and wrist. Care should be taken to stay ulnar to the palmaris longus tendon to avoid injury to the palmar sensory branch of the median nerve.

Dorsal Incisions About the Hand

Longitudinal incision can be successfully used over the dorsum of the hand without the risk of restrictive scarring. Transverse and curvilinear incisions can also be used over the dorsum of the hand if desired. The skin over the dorsum of the hand is mobile and scarring is generally less of a functional problem as the flexor tendons are the dominant force and will minimize the cosmetic and/or functional deficit. Approaches

to the carpometacarpal joint can be made longitudinally between the abductor pollicis longus and extensor pollicis brevis tendons, or they can be more radial (along the

plane of the intersection of the dorsal and volar skin) and extended proximally transversely along the wrist flexion crease, forming a â€œhockey stickâ€• appearance.

FIGURE 14-7 An incision for trigger finger approach using the palm flexion crease. These can result in cosmetically appealing scars with little or no residual deficit. In cases where greater exposure is necessary, a longitudinal Bruner-type incision is preferred.

FIGURE 14-8 An extensile carpal tunnel incision. Angulation at the wrist flexion crease will minimize scarring of the wound. By keeping the incision ulnar to the palmaris longus tendon (*dotted line*), injury to the palmar sensory branch of the median nerve can be avoided.

Incisions About the Wrist

Dorsal approaches to the wrist can be longitudinal, transverse, or curvilinear. The tissues on the dorsal aspect of the wrist are redundant and dorsal scarring is not typically a problem. These can include incisions to expose the base of the thumb. Volar incisions to the wrist can be along flexion creases or longitudinal provided the incision changes direction at the flexion creases in a modified Bruner fashion.

Incisions About the Forearm

As we extend proximally, more traditional incisions can be used. If it is a fairly short incision, such as in open treatment of distal radius fractures, a longitudinal incision will suffice (Fig. 14-9). If the incision is extended distally, a Bruner angulation is made

at the wrist flexion crease. In cases where extensive exposure is necessary, such as is required in compartment syndrome release, an undulating longitudinal incision works well. If the elbow must be crossed, this is usually done in a lazy S fashion passing to the medial aspect of the anticubital fossa.

FIGURE 14-9 A longitudinal approach to the distal radius in a case of open reduction and internal fixation of a distal radius fracture. Angulation of the incision can be performed if the incision needs to be extended distally.

Healing by Secondary Intention

In cases such as infection and Dupuytren's contracture, it may be advantageous to leave the wounds at least partially open. These wounds can heal nicely by secondary intention. Modalities such as whirlpool and wound vacuum assisted closure can be helpful in accelerating wound healing as well as minimizing risk of infection. Disadvantages include time to healing and the fact that some patients find the healing process cosmetically unappealing. Unfortunately, functional limitations such as joint

contractures can result as rehabilitation is delayed during the wound healing process. However, with good postoperative wound care, this can be an effective way to manage some of the more difficult cases or cases where it is necessary to leave the wounds open (3).

Surgery

Patient Positioning

Typically, the patient is in a supine position and a hand table is used. A tourniquet is placed in the proximal portion of the arm. We prefer the pressure to be inflated to approximately 250 mm Hg for adults and 200 to 225 mm Hg for children (generally 70â€“100 mm Hg above the patient's systolic pressure). If the surgery is isolated to the distal aspect of the finger, a finger tourniquet can be placed. In these cases, we routinely also place the upper arm tourniquet which can be used if need be. The entire extremity is prepped. Perioperative antibiotics are administered when indicated.

Technique

Loupe magnification should be used to help avoid injury to important underlying structures. The incision is typically made with a 15-blade knife, taking care to stay perpendicular to the skin and not skiving. The knife can be used to cut the dermis and epidermis. Generally, dissection should proceed from known to unknown or normal to abnormal tissues. Electrocautery, with needle-point cautery, can be used for small veins or vessels. We prefer the use of bipolar cautery. Drains should be liberally used, especially in contaminated wounds or wounds with compromised hemostasis. Options include a small silastic, penrose, or hemovac drain. The drain can be placed without suture so that it can be easily removed postoperatively to not disturb the overlying dressing. Otherwise, drains can be sutured for added protection if the surgeon wants it kept it in for a longer period of time. It is our preference to deflate the tourniquet before wound closure to evaluate hand perfusion as well as hemostasis prior to wound closure. After ensuring adequate hemostasis and viability of the hand, the tourniquet may be reinflated during wound closure. The wounds can be closed in a single layer with a nonabsorbable interrupted 4-0 or 5-0 suture such as nylon or prolene. Either horizontal or vertical mattress sutures can be used. Vertical mattresses are better for

everting the skin edges, whereas horizontal mattress sutures allow for less overall sutures. Increased ischemia of the soft tissues can occur with horizontal mattress closure. In young children, an absorbable suture such as catgut or chromic is preferred.

Postoperative Management

A thoughtfully applied hand dressing is imperative. Depending on the injury, some period of immobilization is generally recommended. Proper immobilization will allow for the skin and soft tissues to heal with minimal stretch and tension. However, this needs to be balanced with the risk of developing joint contractures. In cases of fracture care, immobilization time is prolonged compared with that for soft tissues. The dressing should be mildly compressive and supportive (Fig. 14-10). Surgeries involving the distal aspects of the digits may be adequately dressed with a tube gauze type of dressing. More proximal reconstructive surgeries will require varying periods of immobilization of the wrist and hand in the position of function. Specifically, with the wrist in mild extension, metacarpophalangeal joints flexed and interphalangeal joints extended. Elective soft tissue procedures such as carpal tunnel releases can be immobilized with the wrist in mild extension and the fingers free to move immediately. Although this is our preference, we acknowledge these can be successfully treated with simply a soft dressing. Drains are discontinued when the wound is dry.

FIGURE 14-10 An example of a compressive dressing in a patient with rheumatoid arthritis who underwent metacarpophalangeal joint arthroplasty. A: Following application of xeroform on the incision, fluffs are placed on the volar, dorsal, radial, and ulnar aspects of the wrist and hand. Gauze is lightly placed within the

web spaces of the digits. B: Abundant cast padding is applied followed by (C) plaster. In the immediate postoperative period, we prefer a splint as it will allow for swelling and will compress as the swelling improves. D: A bias wrap completes the dressing. In this example, the MP joints need to be immobilized in extension and neutral deviation. If finger motion is not contraindicated, care should be taken not to have the dressing restrict finger motion.

Complications

The most common adverse result of poorly planned incisions is scar contracture (see Fig. 14-1), which can lead to cosmetic and/or functional deficits. More difficult are skin necrosis and subsequent wound healing problems. These can lead to problems such as infection, excessive scar formation, and diminished function. Other less common complications related to inadequate or suboptimal incisions include injury to nerves, arteries, or tendons. However, with good preoperative planning and technique, these complications are usually avoidable.

References

1. Bruner JM. The zigzag volar digital incision for flexor tendon surgery. *Plast Reconstr Surg.* 1967;40:571.

2. Graham WP. Incisions, amputations, and skin grafting in the hand. *Orthop Clin North Am.* 1970;1:227.

3. McCash CR. The open palm technique in Dupuytren's Contracture. *Br J Plast Surg.* 1964;17:271.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 15

Lateral Arm Flap for Hand and Wrist Coverage

Michael Sauerbier

Goetz A. Giessler

The first description of the lateral arm flap was by Song and coworkers. They used it primarily as a free microvascular flap for small and medium-sized defects in reconstructive procedures of the head and neck. A detailed study by Katsaros et al later popularized it as an extremely versatile, easy to harvest septocutaneous flap with relatively low donor site morbidity. The vascular anatomy is reliable, and the dissection is rather quick and can be performed under regional anesthesia for reconstruction of hand and wrist defects. To date, the flap is used for a wide spectrum of defects all over the body.

The shortcomings of the original lateral arm flap were the flap's bulkiness and a pedicle of moderate length when skin flaps were centered over the middle aspect of the lateral arm. Recent anatomic studies have led to significant modifications in flap design; the skin paddle is now centered over the lower third of the upper arm extending to the lateral epicondyle (extended lateral arm flap). This produces a thin flap with a long pedicle length. The flap may also be used in a reversed fashion (reversed lateral arm flap) or as a pedicled antegrade V-Y flap (extreme lateral arm flap). A pure fascial flap has also been described. The flap may be designed to include vascularized nerve, tendon, and bone. This versatility has made the lateral

arm flap a â€œworkhorseâ€• for elective, urgent, and emergency reconstructions of the hand and wrist.

Indications/Contraindications

Indications

The lateral arm flap has a wide range of applications for closure of small and medium-sized defects in hand and wrist surgery. The classic fasciocutaneous or fascial flap serves well for reconstruction of the dorsal aspect of the hand or the thumb index web space. Patients with a thin subcutaneous fat layer are the best candidates for the flap because it is soft and easily pliable for complex three-dimensional defects.

For cases of palmar reconstruction, we prefer to use a fascial flap, as the cutaneous lateral arm flap may be too thick. This flap is then covered with a full thickness graft and made sensate by anastomising the posterior cutaneous forearm nerve to the palmar branch of the median nerve. Defects with loss of extensor tendons or metacarpal bone can be closed using a composite lateral arm flap using a distal humeral cortical segment and/or a vascularized triceps tendon graft. For all these indications, the flap has to be used as a free microvascular flap. The pedicle is long enough (see following) for flap positioning around the hand or wrist in many cases. The main alternatives to the lateral arm flap are the (osteo-) cutaneous scapular/parascapular, radial forearm, enterolateral thigh flap, and dorsalis pedis flaps. The specific features of each are provided in Table 15-1 .

Contraindications

The free lateral arm flap represents a microsurgical procedure of considerable length. Patients with poor general health who need an elective procedure should be optimized preoperatively; otherwise,

other treatment options should be considered. Previous trauma or surgical procedures in this area, such as plate osteosyntheses of the humerus, probably have destroyed the perforators and are a contraindication. Patients with considerable body fat provide a bulky lateral arm skin flap which is unsuitable for most hand and wrist reconstruction. Using the fasciocutaneous lateral arm flap

together with a skin graft can be an option in these cases. The relatively conspicuous scar of the lateral arm flap poses a relative contraindication in female patients. Whenever the donor site cannot be closed primarily, other flap options should be thought of (see previously). Hair transfer may pose a problem in head and neck reconstructions. Likewise, male patients may prove a challenge with hair transfer in hand and wrist reconstructions.

Max. skin island size (cm)
8 Ã— 25
10 Ã— 20
7 Ã— 25
15 Ã— 28
5 Ã— 8
Max. pedicle length (cm)
11
15
15
10
8
Fascia only
Yes
Yes
â€”
â€”
â€”
Tendon component
Lateral triceps tendon
Palmaris longus, half of the radial carpal flexor
Fascia lata
â€”
Short toe extensors
Muscle component
â€”
â€”
â€”

Latissimus dorsi, anterior serratus
No
Bone component
Distal humerus
Distal radius
â€”
Lateral scapular border
2nd metatarsal
Nerve component
Posterior cutaneous forearm nerve
Lateral antebrachial cutaneous nerve
Lateral femoral cutaneous nerve
â€”
Superficial peroneal nerve
Patient positioning
Supine or lateral prone
Supine or lateral prone
Supine
Lateral prone
Supine

Lateral Arm Flap	Radial Forearm Flap	Anterolateral Thigh Flap	Scapular Parascapular Flap	Dorsalis Pedis Flap

Table 15-1. Comparison of the Lateral Arm Flap to Possible Alternatives for Simple or Composite Reconstructions at the Hand and Wrist

Anatomy

The skin on the upper lateral arm receives its perfusion by septocutaneous perforators in the lateral intermuscular septum inferior to the tip of the deltoid muscle. They originate from two main branches in the middle of the upper arm between the acromion and the lateral epicondyle originating from the profunda

brachii artery. Those main branches are called the anterior and posterior radial collateral arteries (ARCA and PRCA), with the latter having a reliable longitudinal network along the septum to distal arteries. The PRCA is the nourishing artery for the lateral arm flap as it runs close the humerus in the septum between triceps brachii and the brachial muscle. The ARCA branches off between the brachial and brachioradial muscles and is ligated during flap elevation. In rare cases (about 6%), the PRCA is found to be duplicated. On its way distally to the plexus around the elbow, the PRCA gives off three to five septocutaneous perforators serving as feeder vessels for the lateral arm fascia or fasciocutaneous flap. Those surfacing perforators are easily visible during flap dissection. After reaching the subcutaneous plexus, the branches run both anteriorly and posteriorly. In this region, the PRCA anastomoses with the interosseus recurrent artery (between the lateral epicondyle and the olecranon), the recurrent radial artery (anterior to the lateral epicondyle directly above the periosteum), and the inferior ulnar collateral artery to this plexus. These multiple arterial connections allow for distal extension of the skin paddle, which can be harvested up to 12 cm distal to the epicondyle. These anastomoses also allow for the use of the flap in a reverse fashion using one of the perforators from this periarticular plexus.

The venous drainage of the lateral arm flap is by veins of the PRCA together with a rich network of subcutaneous veins including the cephalic vein, which may be included in the anterior part of

larger flaps. The cephalic vein does not need to be included in smaller flaps necessarily, but can provide adequate additional venous drainage in larger or composite flaps.

Depending on flap positioning, the overall pedicle length can reach up to 11 cm in the authorsâ€™ experience, with the artery having a diameter of 2 to 2.5 mm at its proximal origin from the brachial artery.

Bone, tendon, and nerve may be included within the lateral arm flap and used for reconstruction of composite defects. A strip of vascularized bone may be harvested from the distal, epicondylar section of the lateral humerus. Detailed studies have shown one to four nutrient vessels from the PRCA to the bone. These vessels enter the lateral humeral aspect 2 to 7 cm proximal to the lateral epicondyle. Fifty

percent of the triceps tendon may also be harvested with the flap, resulting in a maximum tendon graft of 10 Ã— 2 cm. Gosain et al demonstrated by injection studies that the blood supply to this strip is not via the lateral intermuscular septum, but rather through the triceps muscle. Thus, a cuff of muscle must be included to have a vascularized tendon segment; alternatively, an avascular triceps tendon autograft may be dissected with the flap if one does not wish to violate the triceps muscle. Finally, the posterior cutaneous forearm nerve can be included within the flap allowing for the possibility of a sensate free flap. Harvest of this nerve will result in a strip of distal forearm numbness postoperatively.

Preoperative Planning

Doppler examination can be used for centering the skin paddle over the vascular pedicle. The flap can be harvested in regional anesthesia using a sterile tourniquet (250â€“300 mm Hg). For a transplantation to the contralateral arm, general anesthesia is adequate. Loupe magnification for flap dissection is strongly recommended (2â€“4.5 times). The standard lateral arm flap is centered over a virtual axis between the tip of the deltoid muscle and the easily palpable prominence of the lateral epicondyle of the humerus (Fig. 15-1). When marking this axis in patients who are obese or with very mobile skin, the surgeon has to support redundant tissue of the posterolateral upper arm, as gravity might pull the tissues posteriorly, endangering pedicle inclusion. When flap design is continued beyond the lateral epicondyle, the central axis of the flap is continued distally from the lateral epicondyle in the direction of the distal radioulnar joint on the wrist dorsum. Harvesting the flap from the distal humerus and dorsal forearm produces a thinner flap than flaps raised at the midhumeral level. The reversed or extreme lateral arm flaps have their major indications as a pedicled flap for defect closure around the elbow and proximal forearm.

In planning a composite lateral arm flap, a precise idea about the future positioning of the several components at the recipient site is mandatory. If an osteocutaneous flap is dissected, it should be centered at the distal upper arm, as the receivable bone graft has a maximum length of about 10 cm from the lateral epicondyle proximally. It should not be wider than 25% of the humeral circumference, which is sufficient for most indications at the hand and wrist.

The maximal cutaneous flap area supported by the PRCA on the upper arm varies in size from about 5 Ã— 10 cm to 8 Ã— 20 cm. Primary closure is ideal but this is dependent on skin laxity, patient habitus, and general tissue conditions. Preoperative expansion is described in the literature with the expander being placed subfascially and medially to the lateral brachial septum; but this is certainly reserved for rare indications. Flaps less than 6 cm in width can usually be closed primarily.

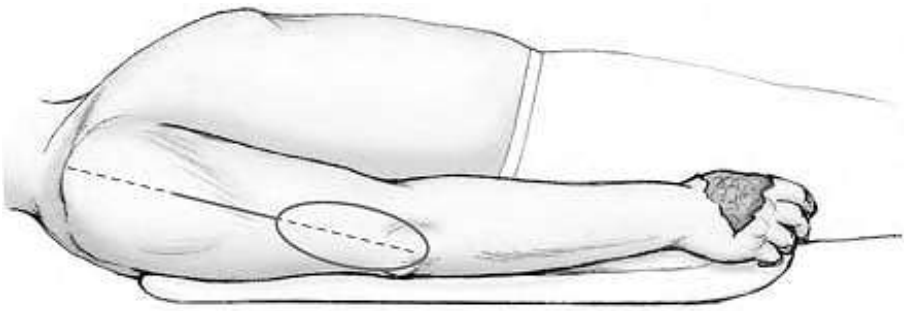

FIGURE 15-1 Planning the axis of a standard lateral arm flap.

Surgery

Patient Positioning

The patient is placed in the supine or lateral supine position with the extremity on a hand table and the arm draped fully mobile. The tip of deltoid muscle, lateral epicondyle, and dorsal aspect of the radioulnar joint are marked and connected as the flap axis (see Preoperative Planning).

Technique

Dissection is started under tourniquet control. Skin and deep fascia are dissected from the muscles together and the fascia is tacked to the skin by 4-0 resorbable sutures to prevent shear forces on the skin perforaters. In an extended or distal free lateral arm flap, it is likewise important to include the deep fascia together with a small strip of intermuscular fascia. Over the epicondylar prominence, the periosteum should be taken en bloc with the fascia to ensure including the

connections to the recurrent radial artery. Retracting the lateral head of the triceps posteriorly, the septum, its attachments to the humerus, the PRCA, and the septocutaneous perforators are visible. Once the posterior exposure is complete, the anterior incision is made and the lateral intermuscular septum is exposed from the anterior. While doing this, the surgeon often encounters the smaller, more oblique oriented septum between the brachial and the brachioradial muscle, which contains the ARCA. The identification of the PRCA perforators is done by carefully flipping the flap over along the axis to look at the lateral septum from both sides, after which all smaller branches from the ARCA are ligated. If the cephalic vein runs through the skin island, it may be dissected out more proximally to have an â€œemergency run-offâ€• and included in the flap, or left in situ. However, it is not obligatory for a successful lateral arm flap. Now the flap is very mobile and only attached to the septum (Fig. 15-2). A self-retaining retractor now can be inserted. In a fasciocutaneous or purely fascial flap, the periosteum is sharply incised bilaterally to the PRCA and sharply taken off the humerus with a periosteal elevator, as the PRCA often runs very close to it. This implies ligating smaller nourishing branches to the bone in an area of 2 to 7 cm proximal to the lateral epicondyle. The vascular pedicle and the cutaneous nerve then are developed proximally under the deltoid to gain enough pedicle length, which should match with the recipient site situation at this time. Dissecting the cutaneous nerves from the radial nerve has to be done with utmost care to prevent uncontrolled bleeding and hematoma around the nerve. Finally, the tourniquet is released, precise bipolar coagulation is performed, and the flap perfusion is checked. The pedicle is ligated only after proper dissection of the recipient vessels and shortly before microsurgical anastomosis to save ischemia time.

If an osteocutaneous composite lateral arm flap is dissected, the periosteum is not stripped off the bone, but only incised and an oscillating saw used for corticotomy (see Preoperative Planning). The release of the bone segment is carefully finished with a sharp osteotome.

The tendon harvest of a triceps tendon requires identification of the feeding branches from the PRCA into the lateral triceps head. A cuff of muscle is included with the tendon for proper perfusion (see Anatomy).

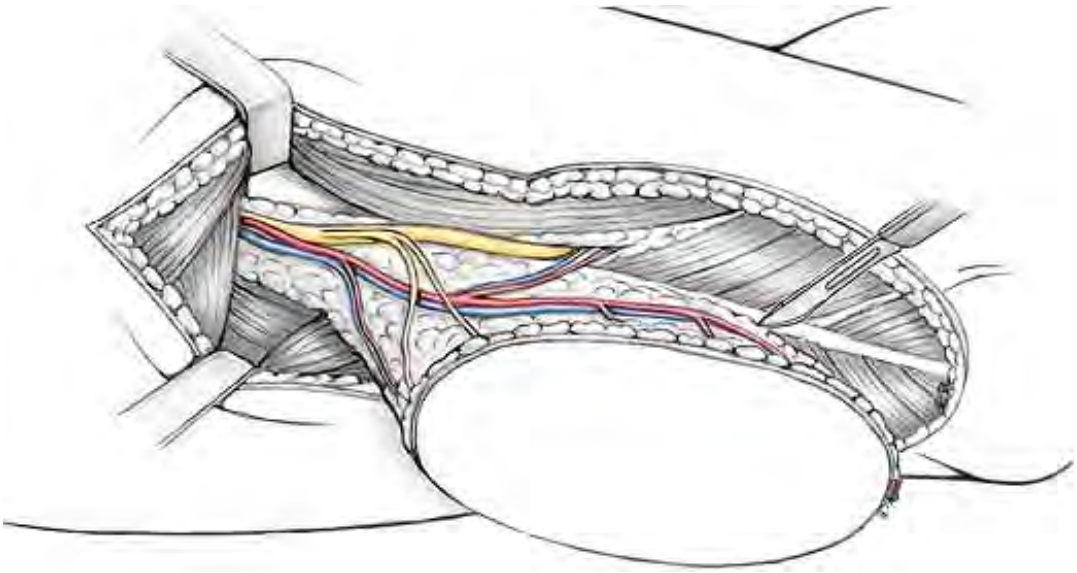

FIGURE 15-2 After the flap is isolated on the septum, the pedicle is followed proximally quite easily.

If the recipient site is prepared properly, the pedicle can be ligated with vessel clips or suture ligation. A suction drain is placed before a double-layered skin closure.

Donor Site Morbidity

Preoperatively, the upper lateral arm can be pinched to check if primary wound closure is possible after removal of the planned skin island. In flaps wider than 8 to 9 cm, primary wound closure is impossible; a skin graft must be applied, and a secondary scar correction or excision after 6 months might be necessary. In the literature, aesthetically unsatisfactory results are reported in about 27% of patients, as even the linear scar is quite visible or can show considerable hypertrophy or pigmentation changes or cause an hourglass-deformity of the upper arm. Early use of silicone sheeting, silicone gel appliance, or intralesional corticoid application should be considered. The sometimes unsightly scar is one of the reasons for restricting the use of the lateral arm flap in female patients.

Postoperative Management

At the donor site, early motion of the limb is encouraged, whereas shear forces to the scar should be avoided. Early scar therapy such as with silicone gel, sheet application, or compressive garments has proved to be successful. If an osteocutaneous flap was harvested, the patient should refrain from heavy lifting until 3 months postoperatively. Protective splints might be considered in noncompliant patients. In the case of a triceps tendon segment removal, full weight bearing (e.g., pushups) should not be allowed before 12 weeks postoperatively. If flap healing at the hand and wrist is undisturbed, physiotherapy is begun on the second postoperative day with constant monitoring of flap perfusion. If tendons were reconstructed, special rehabilitation protocols may then be initiated. Compressive gloves are useful after removal of the suture material. Flap thinning or secondary procedures should not be performed until 6 months postoperatively.

Complications

Specific complications of the lateral arm flap are rare and primarily due to poor planning. As it is harvested most often as a fasciocutaneous flap, no attempt should be made to close the fascia postoperatively, as this might lead to a muscular compartment syndrome. Even adapting the fascial edges at the proximal and distal donor site should be avoided, as this might lead to herniation of the triceps through the gap, which is less of a case if only the skin is closed in a two-layered fashion. Hematoma formation can also lead to considerable tension in that area and has to be evacuated surgically. In case of a fracture of the humerus after harvesting a bone graft, open or closed reposition and osteosynthesis should be encouraged to prevent prolonged immobilization of the elbow joint.

Complications with respect to microsurgical routines are no different than those of other free flaps. Extreme pin-cushioning and volume discrepancies due to secondary overall weight gain make flap debulking procedures necessary (e.g., aspiration lipectomy), whereas hair growth can be reduced with laser therapy.

Disturbances in forearm sensitivity (58.7%), lateral epicondylar pain (19.4%), and a hypersensitive donor site (17%) were quoted as long-term complications. To minimize this sensory forearm disturbance, the posterior cutaneous forearm nerve can be dissected carefully out of the flap. Impaired range of motion in the elbow is

reported after dissecting a composite flap with a triceps tendon graft, but is a rare complication. Primary or secondary fracture of the humerus following elevation of an osteofasciocutaneous lateral arm flap can be avoided with careful technique (oscillating saw) and removal of only 25% of the lateral cortical circumference. If a neuroma of the posterior cutaneous forearm nerve develops despite division close to the main radial nerve, it should be reexplored, resected, or buried in muscle or bone.

Illustrative Cases

Case 1

An 18-year-old man injured his left wrist with a contaminated butcher knife. Tendons and neurovascular structures were found to be intact. Treatment is demonstrated in Figure 15-3 .

FIGURE 15-3 The wound was irrigated and sutured in a small regional hospital. B: Two days later the patient developed necrotizing fasciitis and was transferred to our hospital, where the patient underwent two serial debridements with the initiation of antibiotic therapy. The defect now had a size of 12 Ã— 6 cm. The flexor tendons and median nerve were exposed. C: A distal, thin lateral arm flap was outlined from the same extremity. D: The flap was connected end-to-side to the radial artery and end-to-end to the cephalic vein and sutured into the defect. E: One year postoperative with full range of motion. F: Donor site 1 year postoperative. Note the slightly widened scar around the lateral epicondyle.

Case 2

A 70-year-old woman mutilated her left wrist in a suicide attempt and was found 3 days later in her apartment. At the time of presentation, the patient had a large soft-tissue defect in addition to desiccation of the palmar structures. Treatment is demonstrated in Figure 15-4 .

Case 3

A combine accident in a 21-year-old man resulted in a rupture of the second extensor digitorum communis, extensor indicis proprius, extensor carpi radialis longus and brevis and extensor pollicis longus tendons, three main branches of the superficial radial nerve, the radial artery and the dorsal wrist capsule, and ligaments and dorsal corticalis of the carpal bones. The whole wound was extremely contaminated with organic debris. Treatment is demonstrated in Figure 15-5 .

FIGURE 15-4 A: Intraoperative exploration showed traumatic division of the ulnar artery, median nerve, all superficial and deep flexor tendons, flexor pollicis longus, and flexor carpi radialis and ulnaris tendons. B: After a detailed debridement, all

functional structures were reconstructed except the superficial flexor tendons, which were resected. Three sural nerve grafts (each 4.5 cm) were interposed; the ulnar artery was reconstructed with a vein graft (6.5 cm). C: A fasciocutaneous, thin lateral arm flap was outlined on the ipsilateral distal upper arm. D: The harvested flap had a size of 8 Ã— 4 cm and was anastomosed end-to-end to the radial artery and the cephalic vein. E: Result 10 days after the operation. The patient started physiotherapy already on the second postoperative day.

FIGURE 15-5 A: A thorough debridement and irrigation was performed as a first stage. B: After debridement, all tendons were sutured and the wound was closed with a free fasciocutaneous ipsilateral lateral arm flap. C: The pedicle is marked with thin white arrows; the radial nerve is indicated by thick white arrows. D: It was connected end to end to the radial artery and the cephalic vein, respectively. E: Unfortunately, a persisting wound infection in the carpus made additional debridements necessary. Antibiotic beads were set into the space and an external fixator was placed. F: Finally, the distal carpal row and the bases of the second and third metacarpal bones had to be removed. G: A non-vascularized iliac crest bone graft was fixed with K-wires for a carpometacarpal arthrodesis of the second to fifth ray (white arrows). H: Postoperative radiograph. I: Due to the infection-related elongated immobilization of the hand, the extensor tendons were severely impaired in their motion, even under the flap, and had to be partially resected and tenolysed. Silicon rods were placed under the flap for a secondary extensor tendon replacement. The flap remained viable and pliable throughout all procedures.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 16

Posterior Interosseous Artery Island Flap for Dorsal Hand Coverage

Eduardo A. Zancolli

The posterior interosseous flap is an island pedicled fasciocutaneous flap from the dorsal aspect of the forearm supplied by the cutaneous branches of the posterior interosseous artery. The posterior interosseous artery is located in the septum between the extensor carpi ulnaris and the extensor indicis propius. This flap has proven to be an excellent coverage option for the dorsum of the hand and the first interdigital web. The dorsal cutaneous area of the forearm supplied by the posterior interosseous artery was initially described by Manchot in 1889 (8) and later by Salmon in 1936 (12).

The vascular anatomy of the flap was described along with 20 initial cases (17 cases to cover the first web and three cases to cover the dorsum of the hand). The dorsal cutaneous area of the forearm irrigated by the posterior interosseous artery was initially described by Manchot in 1889 (8) and later by Salmon in 1936 (12).

This island flap was presented for the first time by the author at the VIth European Hand Surgery Course in Ãœmea (Sweden) (13). In this opportunity the vascular anatomy of the flap and an experience on 20 initial cases were presented (17 cases to cover the first web and three cases to cover the dorsum of the hand). Hand outs distributed during the course were published by the author in 1993 (18). The

procedure was also presented in the same year in the XI Congress of the Argentine Society for Surgery of the Hand (awarded with the â€œJ. Goyena Prizeâ€•), in November 1985, Buenos Aires, Argentina (14). Other papers on the procedure were published by us in 1986 (15,16), 1988 (17) and 1993 (1,18). The vascular anatomy of the posterior interosseous island flap was presented by Penteado et al in 1986 (11), and other studies on the posterior interosseous flap have been published by Masquelet et al (9), Costa et al (3,4), Gilbert et al (7), Buchler et al (2), Dap et al (5), Mazzer et al (10), and Goubier et al (6). These studies have further supported the efficacy of the this flap.

Indications/Contraindications

The posterior interosseous artery island flap is capable of transporting the skin of the distal two-thirds of the forearm. The flap's pedicle length determines its coverage possibilities.

It is particularly useful for covering extensive defects on the back of the hand as far distally as the dorsal aspect of the proximal phalanges of the fingers and for obtaining a complete coverage of the first web after the release of severe adduction contractures of the thumb. Buchler et al (2) noted that the flap could reach the dorsum of the proximal interphalangeal joints. It can be indicated in different

types of pathologies such as burn sequelae, wounds on the dorsum of the hand, tumors, and congenital deformities.

Contraindications for flap use include any injury to the posterior interosseous artery or deep laceration or crush wounds to the posterior forearm in the area of the flap pedicle. The flap is not recommended in patients with diabetes, in whom small vessels may be diseased. The flap has been used successfully in smokers, but smoking certainly carries a higher risk of flap complications.

Preoperative Planning

Successful flap harvest is predicated on a thorough knowledge of the flap anatomy. Normally the anterior and posterior interosseous arteries in their course through the forearm are united through two main anastomoses, one proximal, at the level

of the distal border of the supinator muscle, and one distal at the most distal part of the interosseous space (2 cm proximally of the distal radioulnar joint) (Fig. 16-1). The island posterior interosseous island flap is supplied by the reverse arterial blood flow through the distal anastomosis once the proximal anastomosis has been ligated. Venous drainage is produced by the venae comitantes of the posterior interosseous artery.

The angiosomal territory of the posterior interosseous artery was studied in a series of 80 cadaveric forearms in 1993 (1). Ink injections performed through catheter placed in the distal part of the anterior interosseous artery stained the distal two-thirds of the posterior forearm skin through the reverse flow through the distal anastomosis. The proximal third of the forearm skin remained unstained even when larger amounts of ink were injected (see Fig. 16-5). Ink injections through the catheter placed in the proximal part of the posterior interosseous artery stained the proximal two-thirds of the posterior forearm through the direct flow through the proximal cutaneous branch (1).

Other anatomic conclusions from this study can be summarized as follows: the posterior interosseous artery usually branches from the common interosseous artery in the proximal third of the forearm in 90% of cases, it can however be a direct branch of the ulnar artery in 10% of cases (1). It pierces the interosseous membrane about 6 cm distal to the lateral epicondyle of the humerus (14,15,16,17,18) forming its origin or proximal anastomosis between the anterior and posterior interosseous arteries. It enters into the posterior compartment of the forearm below the distal edge of the supinator muscle, located at the union of the proximal with the distal two-thirds of the dorsal aspect of the forearm (60 mm distal to the lateral epicondyle). At its entrance into the posterior compartment of the forearm, the posterior interosseous artery gives off branches to the recurrent interosseous artery that anastomoses at the elbow with a descending branch from the superior profunda, with the posterior ulnar recurrent and the anastomotica magna. In its ascending course the artery runs between the lateral condyle and the olecranon (Figs. 16-1 and 16-2).

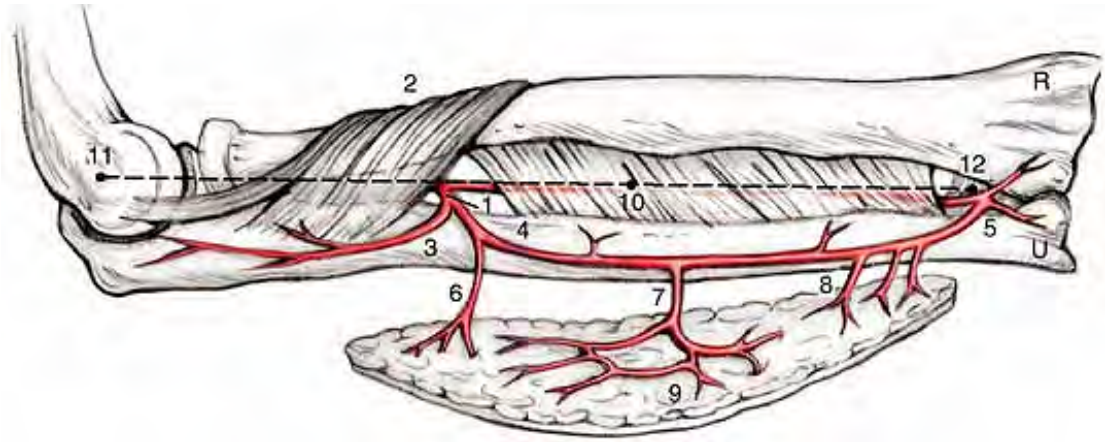

FIGURE 16-1 Anatomy of the posterior interosseous artery. (*1*) Proximal end of the posterior interosseous artery at the distal border of the supinator muscle. (*2*) The artery is emerging as a division of the common interosseous artery. (*3*) Recurrent interosseous artery. (*4*) Posterior interosseous artery following the longitudinal line X-Xâ€™, between the lateral epycondile (*11*) and the distal radio-ulnar joint (*12*). (*5*) Distal anastomosis between both interosseous arteries (2 cm proximal to the distal radio-ulnar joint). Cutaneous branches from the posterior intererosseous artery that irrigte the two distal thirds of the posterior skin of the forearm. Proximal (*9*), middle (*7*), and distal (*8*) branches. (*10*) Middle of the forearm where the middle cutaneous branch is located 1â€“2 cm distal to this point.

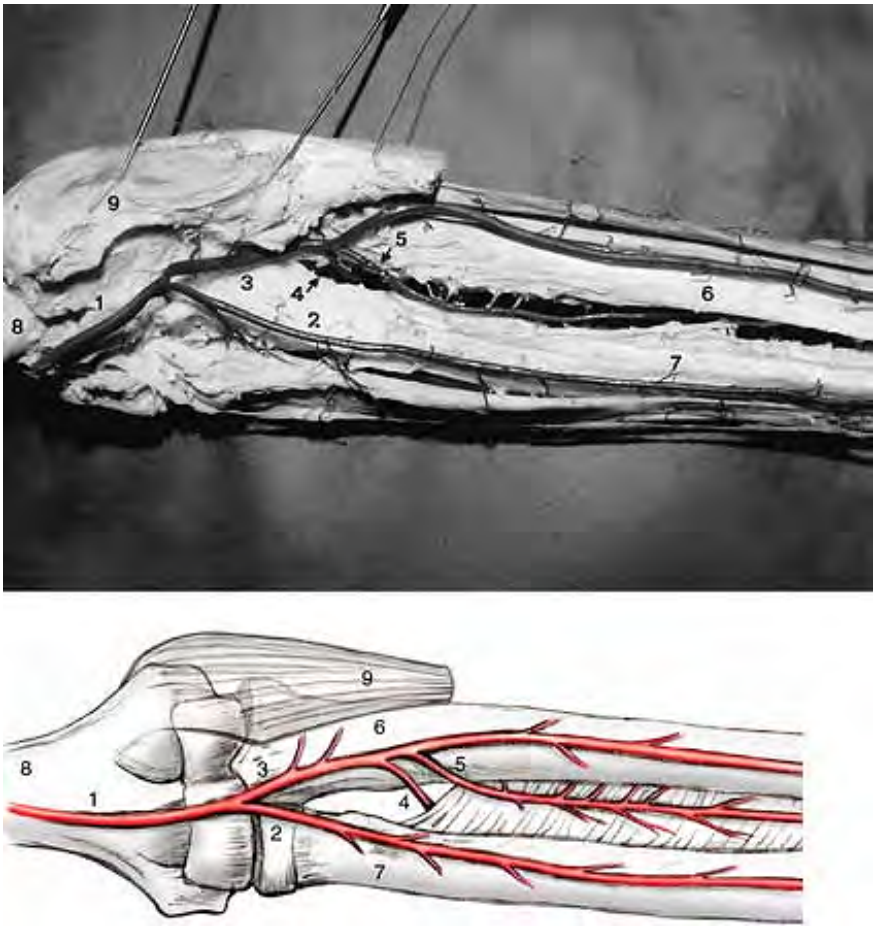

FIGURE 16-2 Cadaver specimen dissected to show the proximal origin of the posterior interosseous artery. Volar view of the proximal part of the forearm and the elbow. (*1*) Brachial artery. (*2*) Radial artery. (*3*) Ulnar artery showing the separated origin of the posterior interosseous artery (*4*) and the anterior interosseous artery (*5*). (*6*) Ulna. (*7*) Radius. (*8*) Humerus. (*9*) Medial epicondyle muscles.

From the proximal anastomosis, the posterior interosseous artery follows a line between the lateral epicondyle and the distal radio-ulnar joint. In its course it can be divided into three parts: proximal, middle, and distal (17).

In the proximal part, the artery runs deep to the abductor pollicis longus muscle and is covered by the extensor digiti minimi and the extensor carpi ulnaris muscles

and in close relation with the posterior interosseous nerve and a large venous plexus. In the middle part, it becomes superficial, just beneath the superficial antebrachial fascia in the middle third of the forearm, running between the extensor digiti minimi and the extensor carpi ulnaris muscles. In this part, the artery reduces its diameter in 90% of the cases and is usually found to be between 0.3 and 0.6 mm in diameter (1) (see Fig. 16-4).

In the distal part, the posterior interosseous artery joins with the distal end of the anterior interosseous artery, forming the distal anastomosis between the two vessels (13,15). This anastomosis is located 2 cm proximal to the distal radio-ulnar joint and very close to the periosteum of the ulnar metaphysis (Figs. 16-1 and 16-3). This anastomosis has been present in all our cadaver specimens and operative cases and represents the point of rotation of the vascular pedicle of the posterior interosseous flap. At the distal anastomosis the posterior interosseous artery enlarges between 0.9 and 1.1 mm (1).

During its course, the posterior interosseous artery gives four to six cutaneous branches which run through the septum between the extensor digiti minimi and the extensor carpi ulnaris muscles to reach the dorsal skin of the forearm, of these the principal cutaneous branches are the proximal and middle branches.

The proximal cutaneous branch emerges from the proximal part of the artery, forming a large branch that irrigates the skin over the upper third of the dorsal aspect of the forearm. It is consistently present but it has a variable origin. Of our 80 specimens, it was found as a branch of the recurrent interosseous artery in 28, a branch of the common interosseous artery in 22, and as a branch of the posterior interosseous artery in 30. This variation indicates that the proximal cutaneous branch cannot be preserved to supply the flap in all cases.

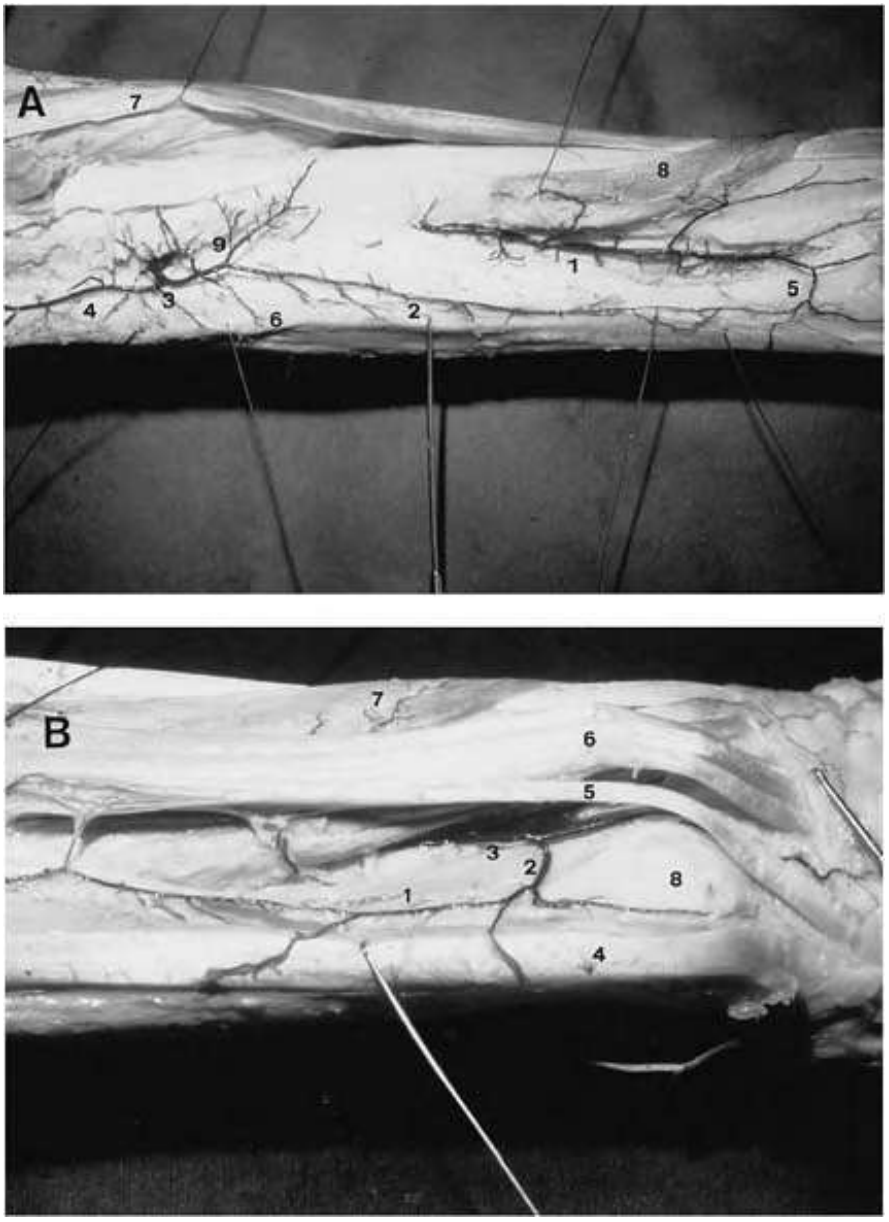

FIGURE 16-3 A: Cadaver specimen injected with latex (right forearm). Dissection of both interosseous arteries. (*1*) Anterior interosseous artery. (*2*) Posterior interosseous artery piercing the interosseous membrane (*3*) and showing the emergence of the recurrent interosseous artery (*4*). The cutaneous branches have been eliminated. (*7*) Distal anastomosis between both interosseous arteries located very close to the periostum of the distal metaphysis of the ulna. (*8*) Extensor carpi ulnaris. (*9*) Extensor carpi radialis brevis muscle. (*10*) Extensor pollicis brevis and abductor pollicis longus

muscles. (*11*) Large branch of the posterior interosseous artery to irrigate common extensor muscle. B: Posterior interosseous artery showing its proximal anastomosis (*2*) with the anterior interosseous artery (*3*). Right forearm. Several branches to the dorsal muscles of the forearm are seen. (*4*) Extensor carpi ulnaris tendon. (*5*) Extensor digiti quinti tendon. (*6*) Extensor communis digitorum tendons. (*7*) Extensor pollicis brevis and abductor pollicis longus muscles. (*8*) Distal head of the ulna.

The middle cutaneous branch originates in the middle third of the forearm where the posterior interosseous artery becomes superficial. It represents a large branch located at 1 to 2 cm distal to the middle of the forearm. It was consistently present in our cadaveric investigations and clinical cases.

No cutaneous branches were found between the described proximal and middle branches. In our cadaveric observations, one or two large interconnecting venous perforators were consistently found running together with the middle cutaneous arterial branch, uniting the superficial to the deep venous system (the venae comitants of the posterior interosseous artery) (1).

At the distal third of the posterior forearm, the posterior interosseous artery may also contribute six to eight cutaneous branches to the skin of variable diameter (1).

The posterior interosseous artery also gives several branches to the muscles of the posterior compartment of the forearm (see Fig. 16-10E) and contributes some blood supply to the periosteum of the radius and ulna.

In only two cases the continuity of the posterior interosseous artery was absent at the level of the middle of the forearm: one in an anatomic specimen and one in a clinical case. This uncommon anatomic finding is in accordance with other reports in the medical literature. Thus, Penteado et al. (11) found absence of the posterior interosseous artery in the forearm in 4 of 70 specimens. In their same series, the distal anastomosis was absent in only one case. Buchler and Frey (2) found the posterior

interosseous artery missing at the middle of the forearm in 2 of 36 cases. Despite this, angiograms and Doppler examination are not part of our standard preoperative

protocol.

Surgery

The surgical procedure is performed under general anesthesia. Pneumatic tourniquet is employed. Exsanguination is obtained by simple elevation of the upper limb for 30 seconds. Loop magnification (4Ã—3) permits a better identification of the vascular pedicles and their cutaneous branches. The patient is placed in the supine position with the forearm placed in pronation over the inferior-lateral thoracic wall with the elbow flexed in 90 degrees (Fig. 16-6). This position facilitates flap preparation and raising. The tourniquet is insufflated (250 mmHg) after the flap has been designed. The operation consists into two main steps: flap design and flap elevation.

Flap Design

Flap design (Fig. 16-7A) is strictly related to the shape and size of the recipient area. A longitudinal line (X-Xâ€™) is drawn from the lateral humeral epicondyle to the distal radioulnar joint. This line represents the course of the posterior interosseous artery in the posterior forearm. A point *A* is marked 1 cm distal to the middle of line. This point corresponds to the emergence of the middle cutaneous branch of the posterior interosseous artery and to the medial interconnecting venous perforator. The

flap is planned to be irrigated by this middle cutaneous pedicle, so the point A must be always included in the flap. A second point *B* is marked at the distal end of the line X-Xâ€™, 2 cm proximal to the distal radio-ulnar joint. This point corresponds to the distal anastomosis between both interosseous arteries and is the point of rotation of the pedicle flap.

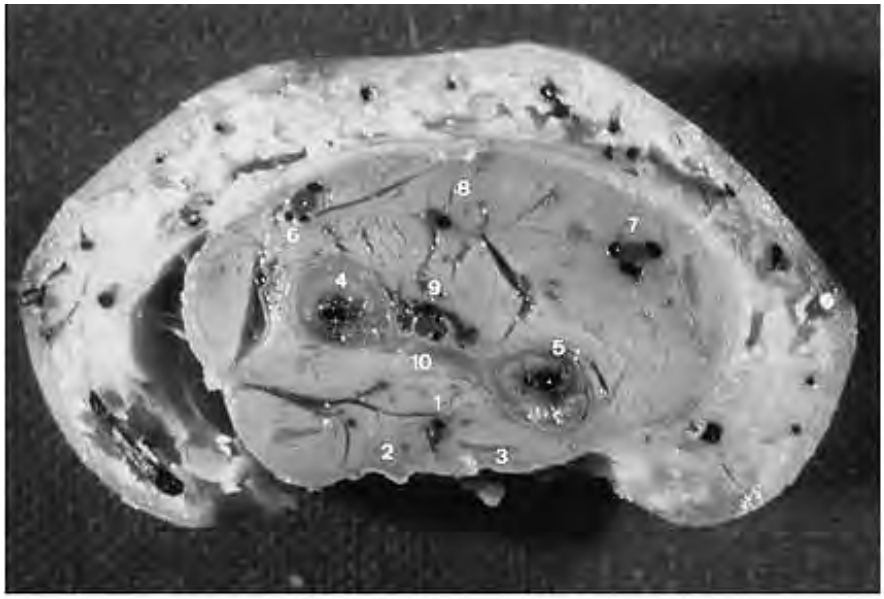

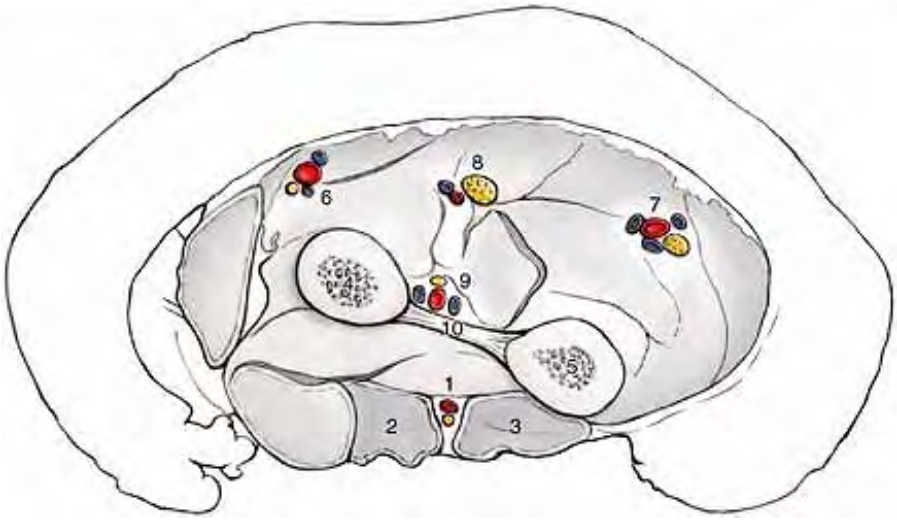

FIGURE 16-4 Transverse section of the forearm at its middle third. (*1*) Posterior interosseous arteryâ€"in its proximal courseâ€"running deep between the extensor digiti quinti (*2*) and the extensor carpi ulnaris (*3*) muscles. (*4*) Radius. (*5*) Ulna. Radial artery (*6*) and ulnar artery (*7*) both with their venae. Median nerve and median nerve artery (*8*). Anterior interosseous artery and its venae (*9*). Interosseous membrane (*10*).

FIGURE 16-5 Fresh cadaver injected with 20 mL of black ink through a catheter placed in the distal anterior interosseous artery. The distal and middle thirds of the dorsal forearm are stained.

The proximal limit of the flap can be safely placed to a point 6 cm distal to the lateral epicondyle (point *C*). The flap may be extended distally up to the wrist joint. In this case distal cutaneous branches can be included. It should be remarked that the medial vascular cutaneous pedicle is capable of irrigating a flap as large as the complete width of the posterior forearm. The center of the flap always corresponds to the line X-Xâ€™.

It was observed in our clinical cases that there is no venous insufficiency in the posterior interosseous reverse forearm flap if it is raised with the large middle venous interconnecting perforator. It ensures sufficient drainage of the subcutaneous tissue and skin into the venae comitants of the posterior interosseous artery (1).

A flap of approximately 3 to 4 cm in width will allow for direct closure of the donor area of the forearm leaving an inconspicuous scar (14,15,16) (Figs. 16-8 and 16-10). Wider flaps will need free skin graft to cover the dorsum of the forearm and may

lead to a poorer aesthetic result (Fig. 16-9).

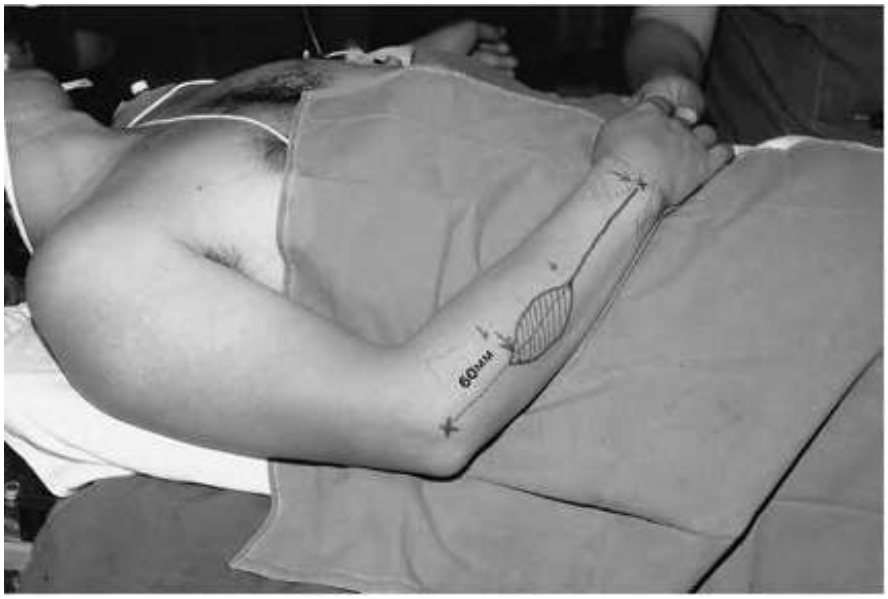

FIGURE 16-6 Position of the upper limb during surgery. In this patient a posterior interosseous artery island flap 4 cm wide has been designed to cover the first interdigital web. The flap is located at the level of the middle cutaneous branch of the posterior interosseous artery. Details on the flap design are completed in Figure 16-7.

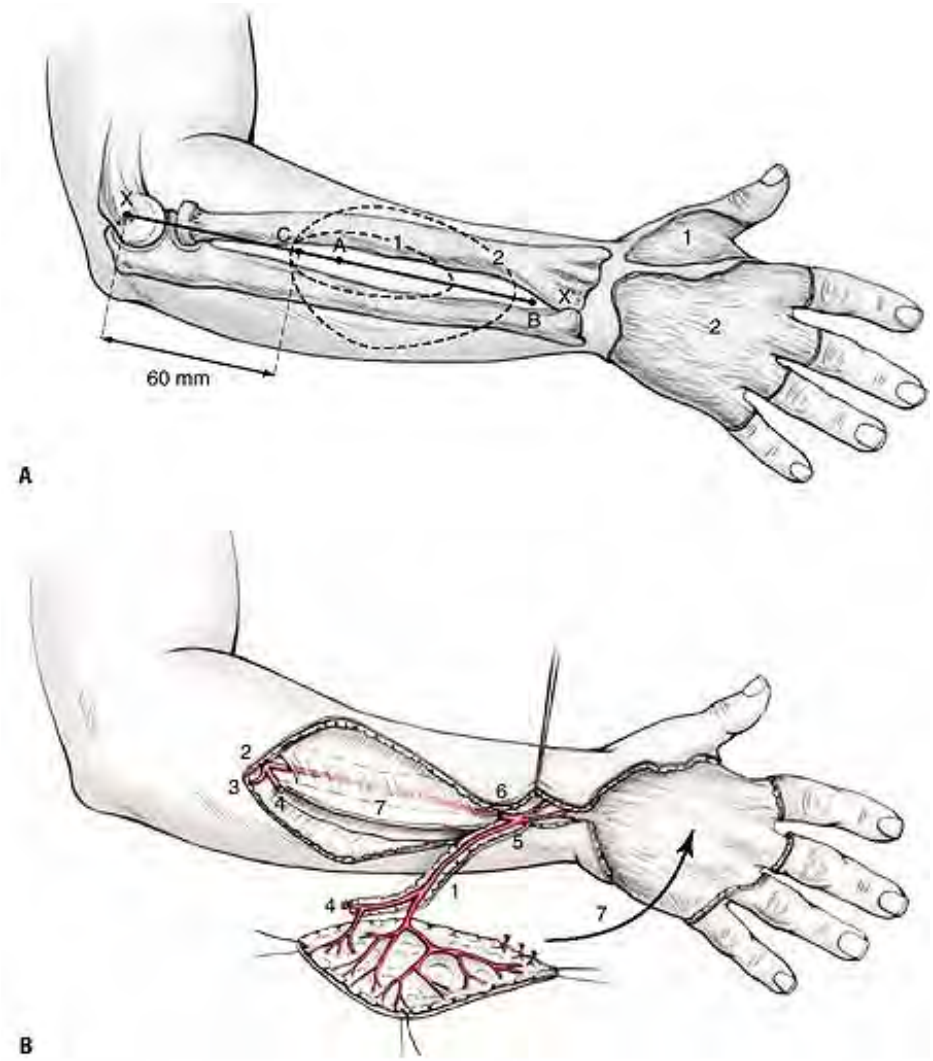

FIGURE 16-7 A: Design of the flap. The location of the island flap is in the line X-Xâ€™, between the lateral epicondyle (X) and the distal radio-ulnar joint (Xâ€™). Three points are marked in the skin. Point A corresponds to the location of the middle cutaneous branch of the posterior interosseous artery. Point B is located 2 cm proximal to the distal radio-ulnar point and corresponds with the location of the distal anastomosis; and point C is located at the distal edge of the supinator muscle where the posterior interosseous artery pierces the interosseous membrane. This point is 60 mm distal to the lateral epicondyle. The width, length, and shape of the island flap is in accordance with the defect to be covered: 1 is representing the shape, size, and location of a flap to cover the first interdigital web, and 2 represents a

flap to cover the dorsal aspect of the hand. B: Raising of the island flap. The flap has been raised with the posterior interosseous artery and its venae comitantes (*1*). In this case the middle and the proximal cutaneous branches are irrigating the flap. The intermuscular septum and a strip of fascia are raised with the posterior interosseous artery. (*2*) Common interosseous artery. (*3*) Recurrent interosseous artery. (*4*) The posterior interosseous artery has been ligated at its emergence (proximal anastomosis between both interosseous arteries). (*5*) Distal anastomosis is preserved (point of rotation of the flap). (*6*) Anterior interosseous artery. (*7*) In this case the flap is rotated to cover the dorsum of the hand. The donor area is covered with a free skin graft (see Figs. 16-8 and 16-9).

FIGURE 16-8 A: Sequela of an injury in a 16-year-old girl on the dorsum of the hand (*1*) with laceration of the long extensor tendons of the index and middle fingers and stiffness of their metacarpophalangeal joints. A posterior interosseous artery island flap was designed (*2*) with the shape of the area to be covered on the hand. Middle and proximal pedicles were preserved to irrigate the flap. Excision of the injured extensor tendons and metacarpophalangeal capsulectomies were indicated simultaneously with the island flap. The fingers were immobilized in flexion during the postoperative period. Location of the distal anastomosis (*3*). B: Finger extension obtained after a secondary surgical stage with Z plasties at the borders of the flap and tendon transfer from the flexor superficialis tendon of the ring finger, divided

into two strips, to the base of the middle and index fingers proximal phalanges. The interdigital webs were reconstructed with the flap. C: Digital flexion obtained. D: The donor area of the forearm was initially closed.

Flap Elevation

Flap elevation (see Fig. 16-7B) begins at its radial side. The dissection is carried out between the subcutaneous tissue and the superficial fascia. The incision is continued in the direction of the wrist, and the proximal and medial cutaneous branches are easily identified. The interconnecting venous perforator that accompanies the medial cutaneous branch is clearly visualized. The proximal cutaneous branch can be coagulated. Our clinical cases have shown that the flap can be safely raised with the middle cutaneous branch. The posterior interosseous artery is sectioned at its proximal origin proximal to the medial cutaneous branch. The posterior interosseous artery with its venae comitants are raised with the intermuscular septum located between the extensor carpi ulnaris and the extensor indicis propius and a very thin strip of fascia covering the muscles. Elevation is completed after sectioning the skin at the ulnar side of the flap. Finally, the flap is turned through 180 degrees to cover the recipient area: to the dorsum of the hand, or to a released first web space, or to both. Now the pneumatic tourniquet is released to observe flap circulation.

At the end of the procedure, the hand is immobilized in a neutral position of the wrist or with some dorsiflexion. In the case of coverage of the back of the hand, where the extensor tendons have been

excised and a metacarpophalangeal capsulotomy has been performed, the fingers are immobilized in flexion to prevent the recurrence of metacarpophalangeal stiffness in extension. In these cases tendon grafts can be indicated during a second surgical stage (see Figs. 16-8 and 16-9).

A

C

B

D

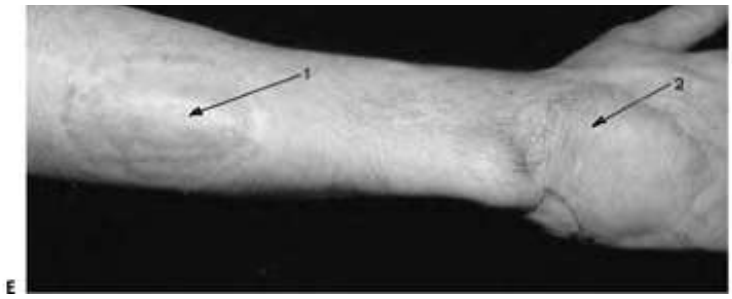

FIGURE 16-9 A: An epitheloid sarcoma with invasion of skin on the dorsal aspect of the forearm in a 20-year-old man. B: An ample excision of the affected skin, including the extensor tendons of the two last fingers, was indicated. C: Piece of the lesion. D: Immediate postoperative period after a posterior interosseous artery island flap. In a secondary surgical stage, the extensor tendons were reconstructed. E: Result after 2 years following the initial operation showing the free skin graft in the donor area (*1*) and the posterior interosseous artery island flap in the recipient area (*2*). Complete function of the fingers was obtained.

Postoperative Management

The hand and wrist are immobilized in a long plaster splint for 2 weeks following the surgery. A large window is left in the dressing to monitor the flap for signs of ischemia or venous congestion. No blood thinning agents are used during the postoperative period. Following removal of the splint, the sutures are removed and the patient may begin postoperative therapy.

Results

Our series of 80 cases (47 males and 33 females) was published in 1993 (1). In this series, 22 flaps were indicated for the coverage of skin defects in the dorsum of the hand, 37 for the reconstruction of the first interdigital web, 5 to cover the dorsum of the hand and first web space, 15 for the volar wrist, and 4 for the palm of the hand. In three cases extensor tendon grafts to extend the fingers were indicated in a second surgical stage. Seventy-six flaps were successful and four were lost. Two of these were clearly due to twisting of the pedicle. The flap provides a large

amount of skin for hand reconstruction without interfering with lymphatic drainage, venous drainage, or the integrity of the

P.202
P.203
P.204
P.205

principal vessels of the volar side of the forearm. It furnishes good-quality skin to permit secondary surgical reconstructions of tendons or the skeleton of the hand.

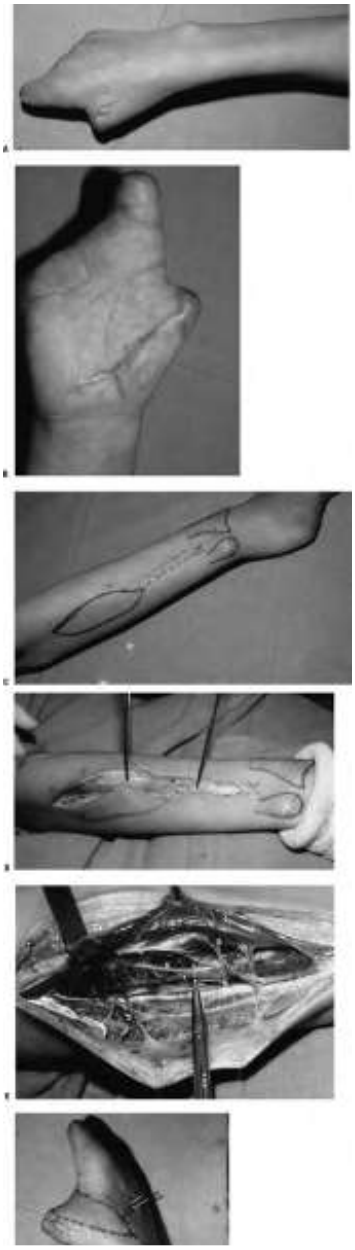

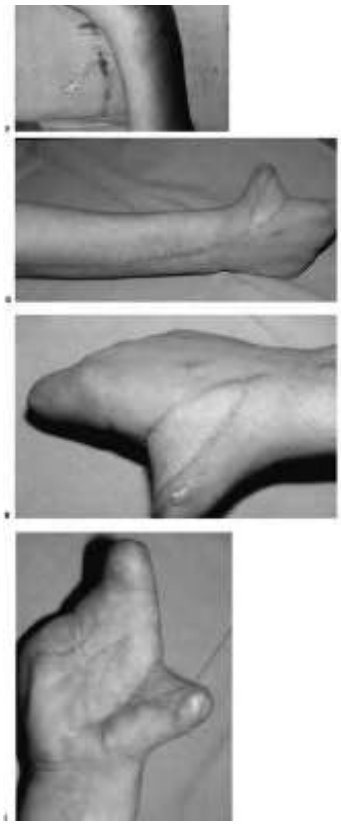

FIGURE 16-10 A,B: Sequela of a trauma with amputation of the digits and adduction contracture of the first intermetacarpal space in a 18-year-old man. C: Design of the posterior interosseous artery island flap of 3 cm width. The middle cutaneous branch was preserved to nourish the island flap. D: The initial skin incision is shown. The radial side of the flap is initially elevated from its radial side. E: Anatomy of the dissected posterior interosseous artery (*1*) with its middle (*2*) and distal (*3*) cutaneous branches to the dorsal skin of the forearm (*4*). Several arterial branches to the dorsal muscles of the forearm (*5*) and the posterior interosseous nerve (*6*) are shown. The extensor carpi ulnaris muscle (*7*) and the extensor muscles to the fingers (*8*) are separated. The hand is at the right of the picture. F: A few days after the release of the first intermetacarpal space and the transport of the island flap. Two KW are maintaining the separation of the first two metacarpals. G: Direct closure of the donor area. H,I: Result obtained.

Complications

As mentioned previously, flap loss is the most serious complication; this is most commonly due to inadvertent twisting of the pedicle or injury to the veins during dissection resulting in flap congestion. Injury to the radial nerve has also been reported with temporary paralysis to the extensor carpi ulnaris and extensor digiti minimi muscles following flap harvest. Such occurrences are rare.

References

1. Angrigiani C, Grilli D, Dominikow D, et al. Posterior interosseous reverse forearm flap: experience with 80 consecutive cases. *Plast Reconst Surg.* 1993;92(2):285â€“293.

2. Buchler U, Frey HP. Retrograde posterior interosseous flap. *J Hand Surg.* 1991;16A(2):283â€“292.

3. Costa H, Soutar DS. The distally based island posterior interosseous flap. *Br J Plast Surg.* 1988;41:221â€“227.

4. Costa H, Smith R, McGrouther DA. Thumb reconstruction by the posterior interosseous flap. *Br J Plast Surg.* 1988;41:228â€“233.

5. Dap F, Dantel G, Voche P, et al. The posterior interosseous flap in primary repair of hand injuries. *J Hand Surg.* 1993;18-B(4):437â€“445.

6. Goubier JN, RomaÃ±a C, Masquelet AC. Le Lambeaux interosseoux posterieur chez lâ€™enfant. *Chirurgie de la Main.* 2002;21:102â€“106.

7. Gilbert A, Masquelet AC, Hentz VR. Pedicle flaps of the upper limb. In: Dunitz M, ed. *Vascular anatomy: surgical technique and current indications.* 1992.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 17

Fillet Flaps in Cases of Mutilating Trauma

Bradley Medling

Michael W. Neumeister

Fillet flaps can be a very important reconstructive option in many cases of hand and lower limb trauma. Fillet flaps are, by definition, axial flaps that may provide skin, muscle, fascia, and bone. They can be harvested either as a pedicled flap or as a free flap. Careful a planning and execution of the flaps can allow for reconstruction of local defects with a flap of a similar tissue type. This technique takes advantage of spare parts that would otherwise be discarded, avoiding associated morbidity of a separate flap harvest.

Fillet flaps are most commonly employed after mutilating trauma to the hand. The concept of spare parts surgery should be broadly applied to include any tissue that would otherwise be discarded; these flaps may include bone, nerve, tendon, joints, or the whole digit. Segments of the injured part can be harvested with their blood supply intact, and thereby be used as a vascularized tissue graft or flap.

Frequently, a traumatized extremity is left without soft tissue coverage, leaving exposed vital structures such as bone, joint, tendon, or neurovascular bundles. Vascularized coverage can be salvaged from attached parts of the mangled hand that are otherwise rendered dysfunctional or from amputated parts where a flap is designed over a preserved vascular bundle.

The key component to the safe harvest of a fillet flap is an understanding of the vascular supply to tissue. Since the entire upper extremity and lower extremity can be used in a fillet fashion, the blood supply to the arm and leg must be understood. In the upper extremity, the axillary artery terminates at the lateral border of the pectoralis minor muscle becoming a brachial artery to a point just distal to the antecubital fossa. The brachial artery runs in the medial upper arm in relative proximity to the ulnar and median nerve. At the level of the elbow, the brachial artery is lateral to the median nerve. The brachial artery bifurcates into the radial and ulnar arteries, respectively, under the superficial flexor muscles of the forearm. The radial artery travels under the flexor carpi radialis muscle exiting more superficial in the forearm just lateral to the tendon of the flexor carpi radialis. The ulnar artery travels with the ulnar nerve under the muscle belly of the flexor carpi ulnaris becoming more superficial in the distal forearm. The ulnar artery travels through Guyon's canal and bifurcates into a superficial and deep branch, each of which communicates with the radial artery component of the superficial and deep palmar arches. The radial artery bifurcates at the wrist into a superficial palmar and deep dorsal branch. The superficial palmar branch courses around the base of the thenar eminence, and the larger deep branch travels under the abductor pollicis longus, extensor pollicis brevis, and the extensor pollicis longus, through the anatomical snuffbox. It then courses between the first and second metacarpals, through the adductor space to form the radial component of the superficial and deep palmar arches. The common digital vessels arise from the superficial palmar arch and travel to the level of the metacarpophalangeal joint, where these vessels now bifurcate to become the digital vessels proper (Fig. 17-1).

In the lower extremity, the superficial femoral and the profunda femoral vessels may be used as arterial inflow for fillet flaps of the thigh. The superficial femoral vessels continue distally to form the

popliteal artery, which bifurcates in the popliteal fossa to the anterior tibial and posterior tibial arteries. The posterior tibial artery subsequently bifurcates to become the posterior tibial artery and the peroneal artery. The anterior tibial artery travels in the anterior compartment above the interosseous membrane, becoming more superficial as it approaches the extensor retinaculum of the ankle. At this point, this vessel becomes the dorsalis pedis artery, which travels between the first and second metatarsal to the foot. The superficial arcuate branch provides arterial inflow to the

dorsum of the foot, while the deep branch pierces between the first and second metatarsal to form the plantar vascular arch. The common digital vessels arise from this arch, and bifurcate at the metatarsal-phalangeal joints to become the digital vessels to the toes. The peroneal artery travels beside the medial aspect of the fibula and terminates as the lateral calcaneal artery behind the lateral malleolus. The poster tibial artery travels in the poster compartment of the lower leg and exits superficial behind the medial malleolus and subsequently bifurcates into the medial and lateral plantar vessels (Fig. 17-2).

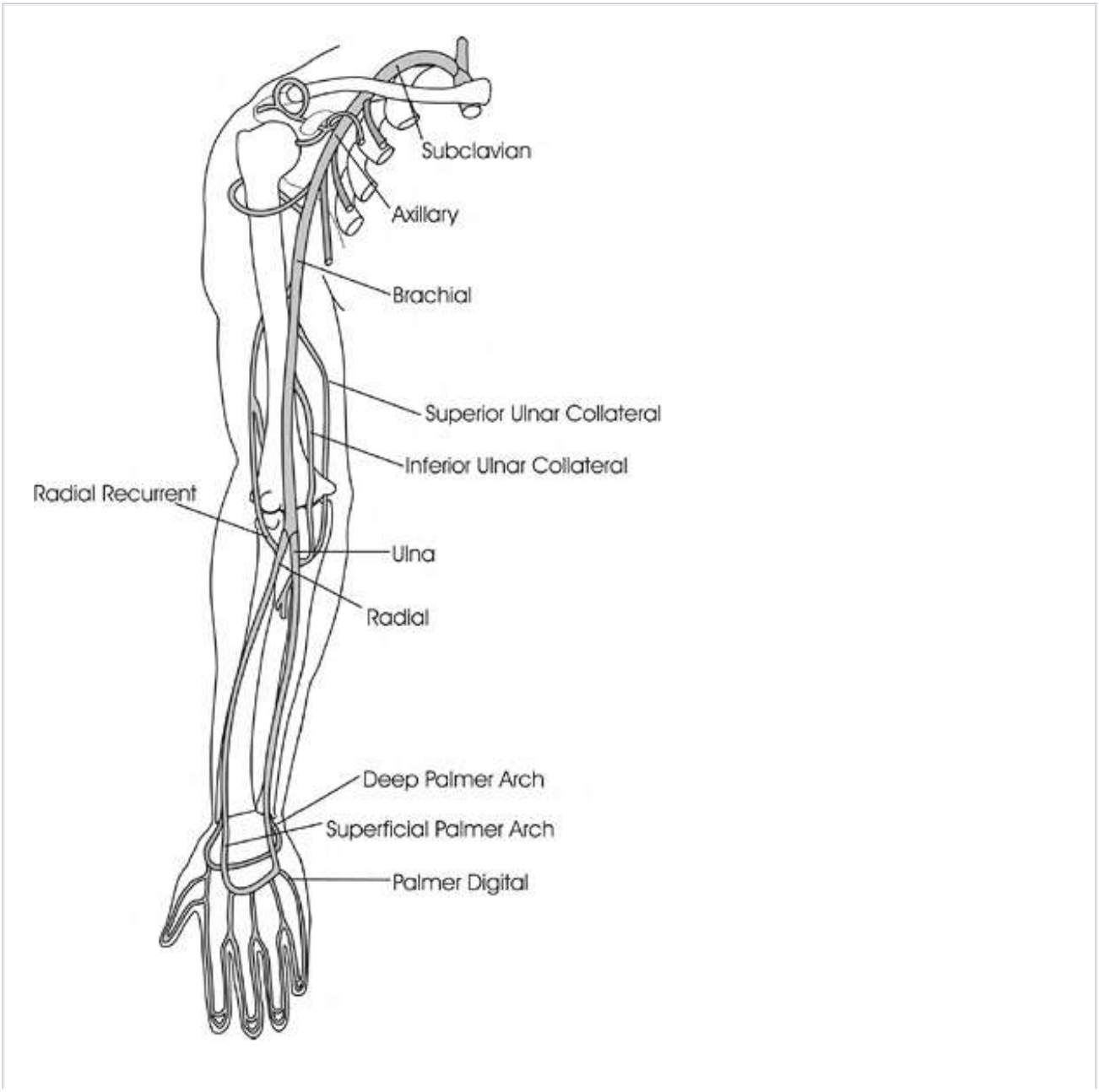

FIGURE 17-1 Upper extremity vascular anatomy. Axial pattern flaps can be designed from known vasculature and their smaller branches.

Indications/Contraindications

Before harvesting a fillet flap for soft tissue coverage, the surgeon should identify the ultimate goal of the immediate and delayed reconstructive procedures. Decisions may be less complex for some lower extremity trauma, where discarding the mutilating limb may be in the best interest of the patient from a functional point of view. Prostheses may be much more functional than attempts at salvage where stability and sensation are compromised. Conversely, mangled hands can have significant function if pinch and grasp can be restored. Reconstructive efforts may be turned toward

salvaging fingers or the thumb, rather than using their tissue as spare parts for soft tissue coverage for the remaining injured hand. The ability to restore some sensate function is a key component of the hand's role in activities of daily living. This separates the overall treatment protocols and goals of upper extremity reconstruction from that of lower extremity reconstruction (Fig. 17-3). The reconstruction process is dictated by the complexity and level of injury. Nerve regeneration, tendon and bone reconstruction, and secondary toe to hand procedures are alternatives to amputation of digits or hands and permit acceptable restoration of function. In both upper and lower extremity, however, there are opportunities that arise to use tissue that would otherwise be discarded, for definitive soft tissue closure.

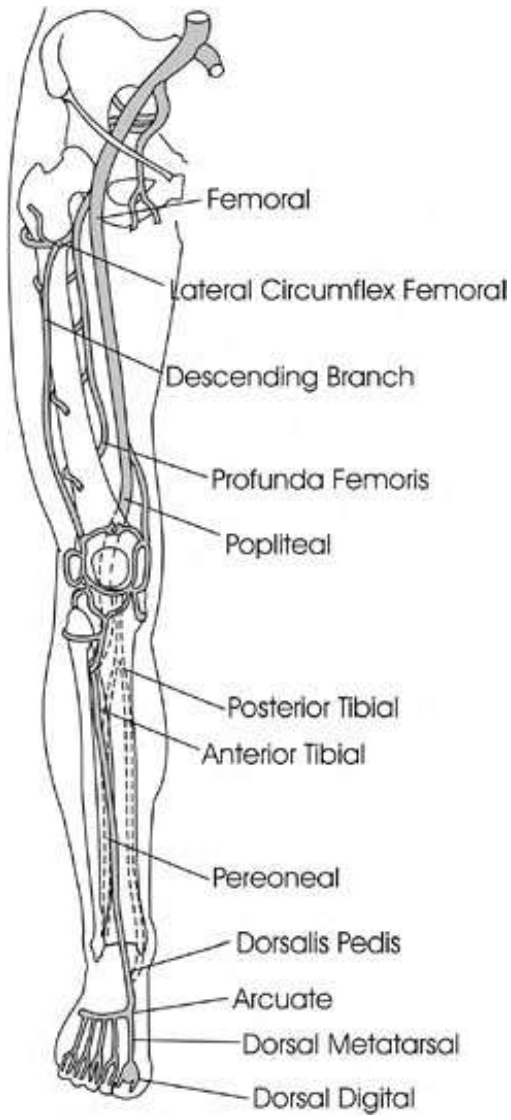

FIGURE 17-2 Lower extremity vascular anatomy.

The availability of adequate tissue often depends on the mechanism of injury to a limb. Clean guillotine lacerations or amputations, with minimal tissue loss, offer the best results at restoration of function. Fortunately, most of these injuries do not mandate that need of spare parts or fillet flaps due to the lack of soft tissue deficits. More traumatic injuries, such as crush and avulsion injuries, or electrical burns, require greater debridement and have more tissue loss. Consequently, these injuries

frequently take advantage of spare parts and fillet flap surgery for coverage.

Pedicle or free tissue fillet flaps can be used for proximal amputations that require stable coverage to better preserve function, such as elbow flexion or knee mobility. This technique is usually indicated for amputations with significant proximal tissue loss and where there is a contraindication for replantation. The remaining distal tissue is filleted, based on one neurovascular bundle, and transferred to cover the proximal exposed structures as a free tissue transfer. If there is an intact neurovascular bundle, but with segmental intervening loss of tissue in the limb, the fillet flap can be harvested as a pedicle flap, leaving the vessels and nerves intact. Indications for fillet flaps are included in Table 17-1.

FIGURE 17-3 A: Mutilated hand from corn picker injury with complete loss of all fingers. B,C: Restoration of pinch from two toe-to-hand transfers.

Table 17-1. Indications for Fillet Flap

Nonreplantable extremity with adjacent tissue loss from trauma	Availability of tissue from discarded, amputated, or nonfunctioning limb or digit
Nonsalvageable digit with proximal or adjacent digit/hand defect	Segmental loss tumor extirpation or trauma

Preoperative Planning

The preoperative planning should take into consideration patient stability, premorbid health, a history of smoking, limb function (i.e., preservation of elbow flexion), available local tissue, prostheses application, and patient occupation. Patient selection is very important when contemplating the use of fillet flaps for soft tissue coverage of limb or trunk defects. In acute, mutilating trauma, the patient must be stabilized initially as survival should be the first goal. Following the full advanced trauma life support (ATLS) workup, with primary and secondary surveys, limb salvage is planned to optimize function.

The initial surgery should entail an aggressive debridement and irrigation to make sure all tissue is viable. Many times, a second and third surgery are required to allow tissue to demarcate and to avoid infection. Before using fillet tissues, the surgeon should plan to transfer based on the blood supply to the spare tissue. The fillet flaps can either be transferred as a free flap or a pedicle flap. Digital fillet flaps are neurovascular flaps usually based on the digital arterial and nerve supply to the finger (Fig. 17-4). The surgeon will plan to use the digit that has been, in part, rendered dysfunctional as a result of the proximal trauma but where viability is still present distally through an

intact neurovascular bundle. In the elective cases, angiography may help to identify intact vascular anatomy before transfer. Angiography is rarely indicated in the acute traumatic case as the surgeon can visualize the vessels and observe the vascularity (bleeding) of the distal tissues.

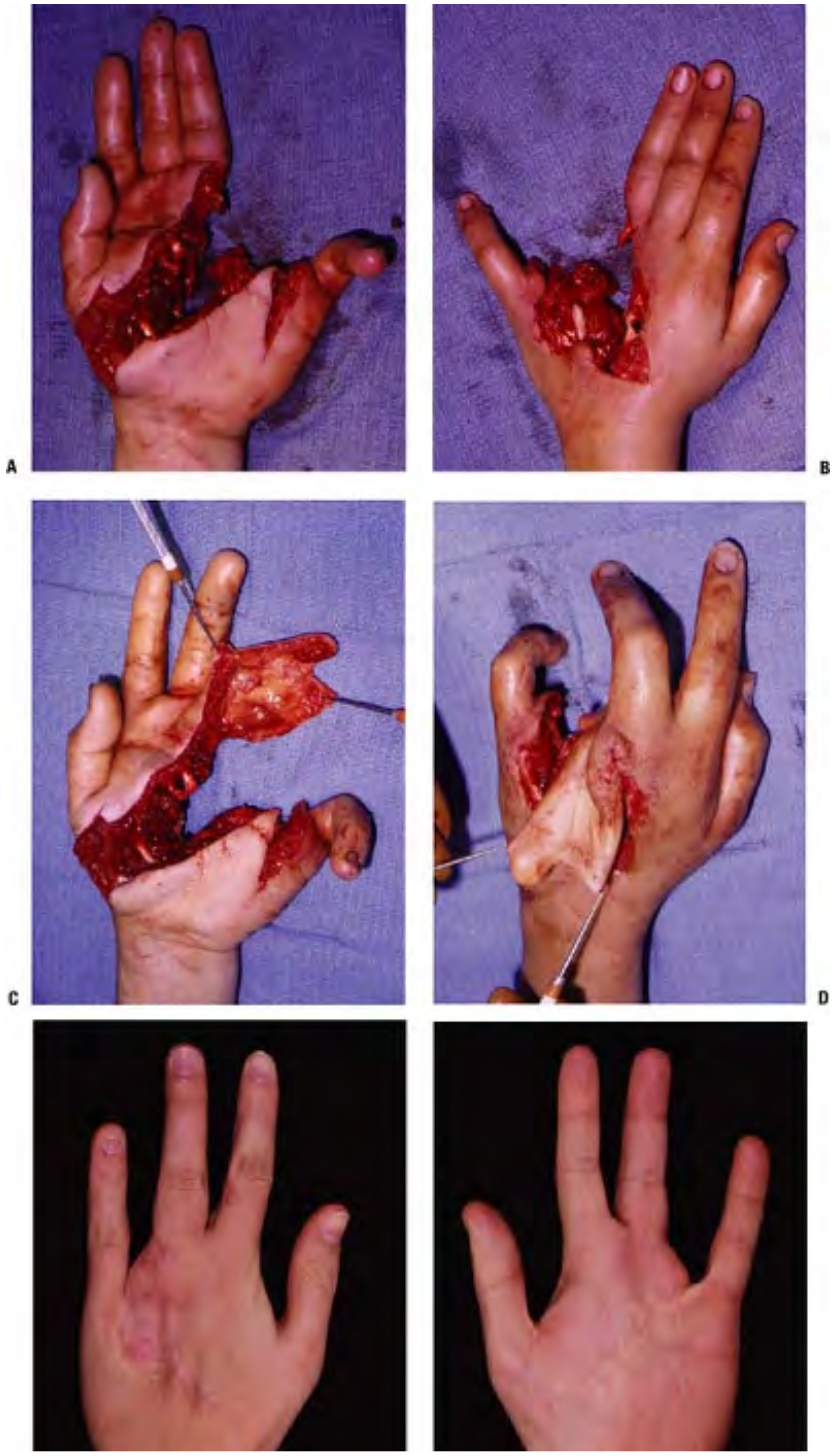

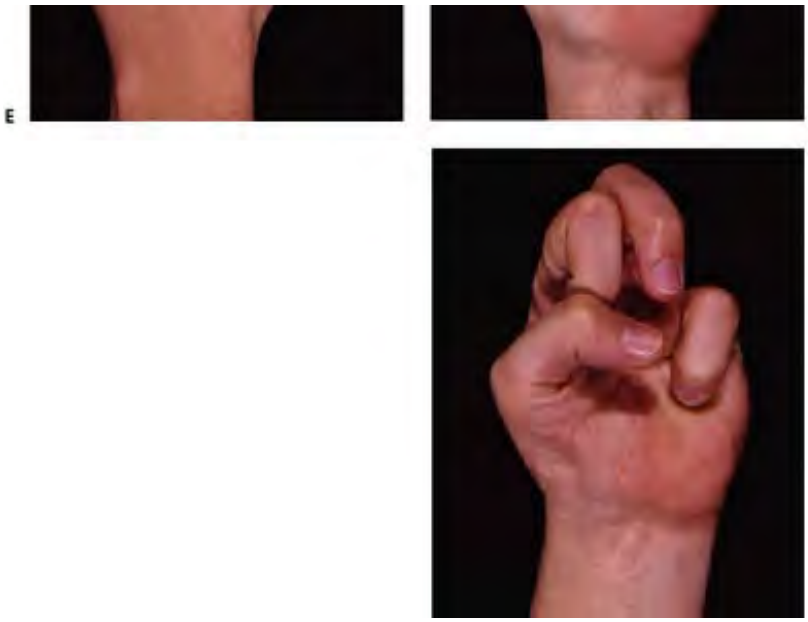

FIGURE 17-4 Aâ€“D: Tablesaw to the right hand with intrusion through the web space between the small and ring finger in an ulnar to radial direction. Complex repair of flexor/extensor tendons, neurovascular repair of the arch and common digital vessels was undertaken. The ring finger is filleted based on its radial neurovascular bundle. Resection of the head of the fourth metacarpal was performed secondary to the level of comminution. Long-term follow-up reveals good prehension. Eâ€“G: Tablesaw to the right hand with intrusion through the web space between the small and ring finger in an ulnar to radial direction. Complex repair of flexor/extensor tendons, neurovascular repair of the arch and common digital vessels was undertaken. The ring finger is filleted based on its radial neurovascular bundle. Resection of the head of the fourth metacarpal was performed secondary to the level of comminution. Long-term follow-up reveals good prehension.

Surgery

Patient Positioning

The patient should be positioned appropriately to allow dissection of any digit, arm, or leg based on the blood supply of that part. The patient is usually in the supine position

for most upper extremity and lower extremity flaps. An exception to this is the use of a fillet flap for recurrent, recalcitrant pressure sores on the ischial and para-sacral areas, for which the patient should be placed in the prone position.

Technique

The assessment of vascular integrity should be performed before inflation of the tourniquet. Although a pulse-lavage system can be employed, a simple spray bottle allows adequate force to propel smaller particulate debris from the wound (Fig. 17-5). Larger debris should be mechanically debrided with sharp dissection. The preservation of questionable tissue is acceptable if further assessment is planned in 24 to 48 hours for further debridements. Temporary coverage with dressings, xenograft, or allograft can be used in the interim to prevent desiccation in the interim (Fig. 17-6).

FIGURE 17-5 A: Pulse lavage system. B: Simple spray bottle.

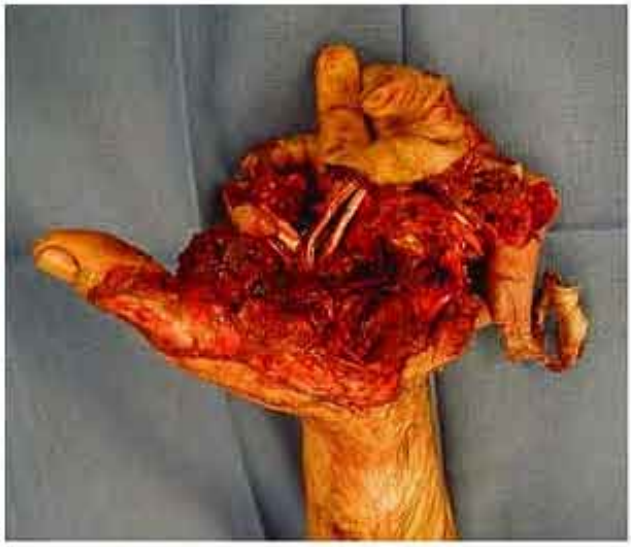

A

B

C

FIGURE 17-6 Aâ€“C: Mutilating injury necessitating temporary coverage with allograft secondary to the size of the wound, and possible need for further debridement. Dâ€“E: Mutilating injury necessitating temporary coverage with allograft secondary to the size of the wound, and possible need for further debridement.

Digital Fillet Flaps of the Hand

The hand provides an ideal setting for the use of fillet flaps for soft tissue coverage of mutilating injuries. A single finger can provide all the components required for composite reconstruction of the hand. Digital fillet flaps can be designed to cover both palmar and dorsal hand defects (Fig. 17-7). The use of adjacent digits that are rendered dysfunctional from the injury can provide coverage with glaborous skin and normal sensibility (Fig. 17-8). Similarly, microsurgical techniques can be used to move the filleted digit to an area of the hand not immediately adjacent to the flap. Idler describes the use of a filleted digit as a free flap to transfer to cover proximal hand wounds. An osteocutaneous fillet flap was described by Gainor in which the index finger with the bone intact was transferred to reconstruct the thumb. A composite, neurovascular island flap can be used in a similar fashion incorporating the distal phalanx, nail bed, and skin for sensate reconstruction of a hemi-thumb defect (Fig. 17-9).

The vascularity of the digit to be a fillet is determined by gross inspection and the use of a hand-held Doppler. The arc of rotation depends on the fillet flap's vascular pedicle. The fillet flap must reach the primary defect based on the pivot point of a pedicle. Al-Qattan described lengthening a fillet finger flap by incising the skin at the finger. He then recommends basing the flap on either the radial or ulnar vessel, which

will allow one to unfold of the flap and increase its functional pedicle length. Preoperative planning using a surgical sponge to measure the arc of rotation and the subsequent length of the flap required is a helpful tool in determining the utility of the proposed flap. Incomplete filling of the defect will result in the need for skin grafting some other form of soft tissue coverage. Split or full thickness skin grafts can be harvested from remaining spare parts. Excess tissue within the fillet flap can be used to fill in contour defects. In such cases, the distal portion of the fillet flap can be de-epithelialized and folded beneath the remaining flap to augment soft tissue bulk.

Fillet Flaps of the Legs and Arms

Fillet flaps of the legs and arms are used to cover proximal amputation stumps to prevent further shortening of the limb or to add length around the elbow or shoulder. In upper extremity trauma, these fillet flaps are based on the radial or ulnar arteries. The flap is usually a free tissue transfer because there is often an intervening defect in the vessels. The brachial artery and venae commitantes act as the recipient vessels around the elbow. In lower extremity trauma, the fillet flap can be based on either the anterior tibial or the posterior tibial vessels. Incisions in the distal amputated part are made on the opposite side of the limb from where the blood

P.216
P.217
P.218
P.219
P.220

supply is present. This keeps the blood supply to the skin, subcutaneous tissue, and muscles based axially with the medial and lateral tissue equal distances from the primary vessels. Care must be taken during the removal of the bony and muscular components of the flap not to injure the neurovascular bundles. The flap is trimmed appropriately to fit the recipient defect. The fillet flaps can be harvested as a pedicle flap or as a free tissue transfer (Fig. 17-10).

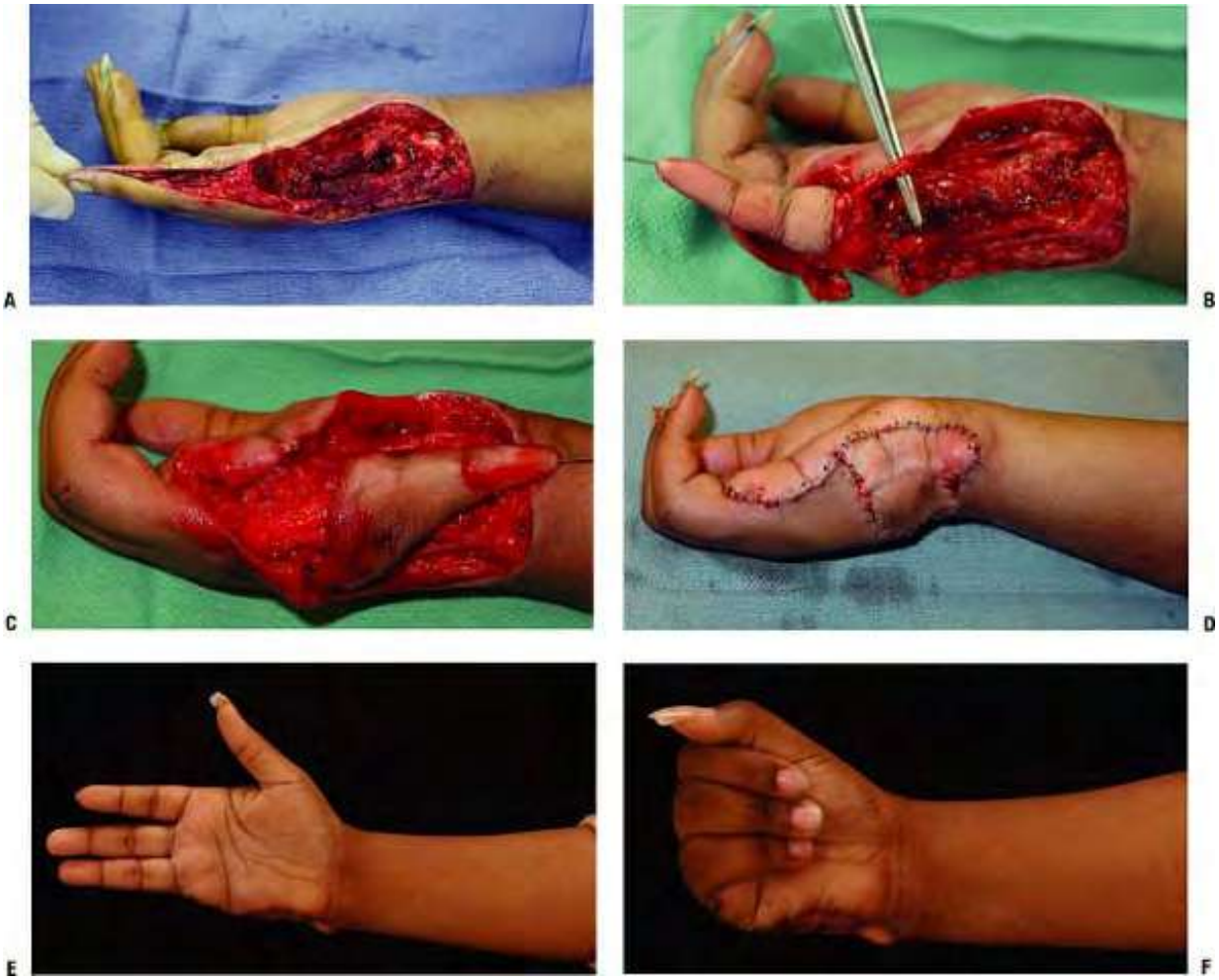

FIGURE 17-7 Fourteen-year-old patient with hypothenar defect. Aâ€“D: The small finger is filleted based on its radial neurovascular bundle for coverage. This technique provides stable, sensate closure of the defect. E,F: Three-month follow-up reveals stable coverage and good hand function.

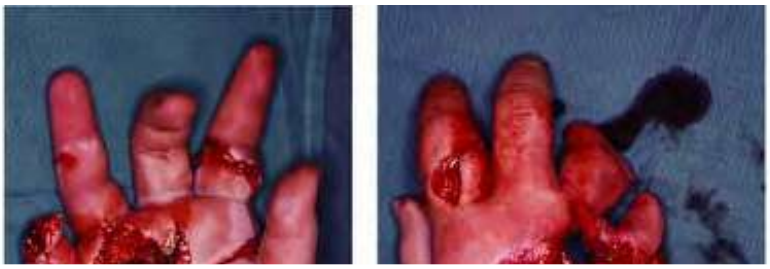

A

B

C

D

E

F

FIGURE 17-8 Aâ€“D: Punch press injury to ulnar aspect of left hand. Ring and small fingers filleted for coverage. A small skin graft is placed dorsally beyond the reach of the tip of the flap. Long-term postoperative follow-up reveals stable coverage. Eâ€“G: Punch press injury to ulnar aspect of left hand. Ring and small fingers filleted for coverage. A small skin graft is placed dorsally beyond the reach of the tip of the flap. Long-term postoperative follow-up reveals stable coverage.

FIGURE 17-9 A,B: Table saw injury to ulnar aspect of a thumb. Câ€“E: Additional injury to the middle and ring fingers. Neurovascular island flap harvested from the middle finger transferred for composite coverage of the thumb.

FIGURE 17-10 Lower extremity reconstruction after amputation. The sole of the foot is filleted and transferred proximally to cover the proximal stump after amputation.

Postoperative Care

Postoperative care depends on the area reconstructed. For upper extremity defects, bulky noncompressive dressings are used to protect lacerated or damaged digits. A window is left within the dressing to monitor the vascular status of the fillet flap. If the spare part has been transferred as a free flap, then flap monitoring is performed hourly for the first 24 to 48 hours to ensure there are no signs of arterial ischemia or venous congestion. Signs of either problem will necessitate anastomotic re-exploration to prevent flap failure.

If extensive skin grafting has been performed, motion within the digit or extremity is usually delayed for 5 to 7 days to allow for skin graft take. If a bolster dressing has been placed or if skin grafts have been secured with the use of vacuum-assisted-closure (VAC), motion may be initiated sooner.

Lower extremity wounds are also treated initially with bulky noncompressive dressings and immobilization with the application of a posterior splint. Leg elevation is essential to limit swelling and minimize the risk of venous congestion within the flap. At 1 week postoperatively, compressive wraps may be initiated to the flap and reconstructive area to help with the resolution of postoperative edema. Compressive wrapping may also aid in long-term flap contours.

Results/Complications

Flap loss is obviously the most devastating complication following fillet flap reconstruction. In such cases, alternative means of wound coverage will be necessary to cover resultant defects. Kuntscher et al. present their results from pedicle fillet flaps of fingers for dorsal, palmer, and adjacent digit defects secondary to trauma, burn wounds, Dupuytren's disease, malignant tumor resection, and diabetic gangrene. In their series of 30 patients, the fillet flap provided a stable coverage option with normal sensibility compared to the unaffected digits. The overall complication rate was 18%, with a flap loss, revision surgery, and infection rate of 7.5%.

Recommended Reading

al Qattan MM. Lengthening of the finger fillet flap to cover dorsal wrist defects. *J*

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 18

The Use of Free Flaps in Upper Extremity Reconstruction/ Anterolateral Thigh Flap

Lawrence Lin

Samir Mardini

Steven L. Moran

Indications/Contraindications

Soft tissue deficiencies within the upper extremity are common following trauma, burns, infection, and tumor extirpation. The coverage of such defects can usually be accomplished with the use of pedicled flaps or local rotational flaps. However, when defects are very large or encompass multiple structures including nerve, bone, or muscle, the use of composite free tissue transfer provides a reliable and single stage means of reconstructing complex defects.

The benefits of free tissue transfer within the upper extremity include the transfer of additional vascularized tissue to the injured area, the ability to carry vascularized nerve, bone, skin, and muscle to the injured area in one procedure, and the avoidance of any additional functional deficits to the injured limb which may be incurred with the use of a local or pedicled flap. Free flaps are not tethered at one end, as is the cases for pedicled flaps, and this allows for more freedom in flap positioning and insetting. More recent fasciocutaneous and perforator flaps also allow for primary

closure of donor sites with minimal sacrifice of donor site muscle. With current microsurgical techniques, free flap loss rates are between 1% and 4% for cases requiring elective free tissue reconstruction. Finally, the upper extremity is particularly suited for free tissue transfer as the majority of recipient blood vessels utilized for anastomosis are located close to the skin, and are of relatively large caliber.

Major indications for free tissue transfer are: (a) primary coverage of large traumatic wounds with exposed bone, joint, and tendons or hardware, (b) coverage of complex composite defects requiring bone and soft tissue replacement, (c) coverage of soft tissue deficits resulting from release of contractures or scarring from previous trauma, and (d) significant burns.

There are few absolute contraindications for free flap transfer and in many cases free tissue transfer may be the only option for upper limb salvage following significant soft tissue loss. Despite this, relative contraindications to free tissue transfer include a history of a hypercoagulable state, history of recent upper extremity DVT, and evidence of ongoing infection with the traumatic defect. Other contraindications would include inadequate recipient vessels for flap anastomosis. Disregarding

technical error, the status of the recipient vessel used for flap anastomosis may play the greatest role in flap failure; recipient vessels within the zone of injury are prone to postoperative and intraoperative thrombosis. Recipient vessels for microvascular transfer should ideally be located out of the zone of injury, radiation, or infection. In rare cases, arterial venous fistulas may be created proximally within the upper extremity or axilla using the cephalic or saphenous vein. These fistulas can be brought into the zone of injury and divided to provide adequate inflow and outflow for free tissue transfer.

Specific Indication for the Anterolateral Thigh Flap

There are many choices for free flap coverage of the upper extremity. The scapular, parascapular, latissimus dorsi myocutaneous, and lateral arm flap have long been favorites of surgeons for reconstruction of traumatic upper extremity injuries. Many of these flaps will be described in later chapters of this text. If joints are to be crossed, fasciocutaneous flaps are much preferred as muscle flaps can undergo atrophy and

restrict flexion and extension across joints or fingers (Fig. 18-1).

Classic cutaneous free flaps, such as the radial forearm flap, lateral arm flap, and scapular flap, have limitations in size, donor site morbidity, and overall thickness. Musculocutaneous flaps such as the latissimus dorsi and rectus abdominus flaps result in functional loss and donor site morbidity including, particularly in the abdomen, potential hernia formation. In addition, in coverage of joint surfaces, muscle flaps tend to undergo fibrosis and atrophy over time, which may limit muscle excursion, particularly when placed over the elbow or dorsum of the hand. Muscle is still indicated for those circumstances involving osteomyelitis or significant soft tissue contamination. More recently, chimeric flaps have been harvested to include both muscle and a large component of skin, providing the ideal coverage for many complex defects in the upper extremity.

FIGURE 18-1 A: An extensive iv infiltrate to back of hand and forearm resulted in

full thickness skin loss over the majority of the hand and forearm in this 56 year old woman. B: Soft tissue coverage was obtained with the use of a rectus abdominus free flap. C,D: Despite successful soft tissue coverage, atrophy and fibrosis within the muscle over time has lead to limitations in wrist motion resulting in decreased wrist flexion.

In recent years, the anterolateral thigh flap has become the major flap in reconstructive microsurgery, including head and neck defects and extremity wounds. It has replaced many other flaps. The skin overlying the anterior thigh region has relatively constant anatomy with the descending branch of the lateral femoral circumflex artery (LFCA), giving rise to either pure muscular (musculocutaneous) or subcutaneous (septocutaneous) perforators that supply the area. Based on their experience with over 1,500 anterolateral thigh flaps for various anatomic defects including the upper extremity, Chen et al determined that 12% were based on direct septocutaneous perforators, and 88% were based on musculocutaneous perforators. Variations in perforator anatomy can exist, which include absence of skin perforator, perforators which are too small for elevation, a perforator artery which does not run with the vein, and perforator arteries that have no accompanying vein. These anatomical variations are rare, accounting for 2% of cases, however they need to be noted by the surgeon. As proposed by Chen, an algorithm for managing anatomical variations begins with attempting to identify a more proximal perforator in the upper thigh, usually arising from the transverse branch of the LFCA, and harvesting the flap based on this perforator. Alternatively, an anteromedial thigh flap may be raised or the vastus lateralis may be taken as a musculocutaneous flap. Finally, exploration can be performed on the contralateral side as the anatomy may be different.

The anterolateral thigh flap serves as the ideal flap for upper extremity coverage due to its many considerable advantages. The flap provides a long pedicle (up to 16 cm) with suitable vessel diameters; the arterial pedicle can measure up to 2.5 mm and the two venae comitantes can measure up to 3 mm in diameter. It is also a versatile flap with the ability to incorporate different tissue components with large amounts of skin, as the flap can be harvested as a cutaneous, fasciocutaneous, or musculocutaneous flap with vastus lateralis. In addition, based on the supply of the LFCA system, a

chimeric flap incorporating the rectus femoris or tensor fascia lata can be raised to cover extensive, complex defects. The flap may be harvested as a sensate flap by including the lateral femoral cutaneous nerve or as a flow-through flap in cases of significant arterial trauma. Inclusion of thigh fascia with the flap allows its use as an interposition graft for tendon reconstruction. The thickness of the flap may be debulked primarily, optimizing the match of donor tissue for the upper extremity. Ordinary skin flaps can sometimes produce bulkiness with poor aesthetics. Thick skin paddles, such as with parascapular flaps, may interfere with motor function and flexion of the metacarpal phalangeal joints or inter-phalangeal joints. Flaps as thin as 3 mm to 5 mm have been harvested for tendon coverage. The donor site results in minimal morbidity with most sites able to be closed primarily, resulting in a linear scar and absence of any long term leg dysfunction. Lastly, its anatomic location allows for a two-team approach for flap elevation and recipient site preparation, saving considerable operative time.

Anterolateral Thigh Flap for Upper Extremity Reconstruction

Preoperative Planning

Preoperative requirements for flap consideration begin with the preparation of a clean wound bed. Radical debridement of all necrotic tissue is the most important component of a successful reconstruction. Tissue considered to be of marginal viability should be debrided early rather than performing multiple dressing changes or utilizing vacuum-assisted therapy in the hopes of rescuing traumatized tissue; such measures can lead to delayed definitive surgical reconstruction, perpetuate the inflammatory component of wound healing, perpetuate distal edema, and result in hand and limb stiffness. If the surgeon can guarantee a clean wound bed, free of any necrotic material, immediate flap coverage may be attempted in cases of acute trauma. We have found that most high energy traumatic injuries and agricultural accidents require at least one to two surgical debridements prior to definitive wound closure. Wound debridements in these cases are performed in conjunction with wound cultures for bacteria and fungal species. The ideal timing for upper limb free tissue reconstruction has been debated within the literature but should be within 72 to 96 hours of injury.

The upper extremity is evaluated for any evidence of concomitant bony or

neurovascular injury. A careful vascular evaluation is also performed and if there is any question as to the status of the inflow vessels, an angiogram or a CT angiogram may be obtained to verify inflow. Our preference is to perform the majority of arterial anastomosis in end-to-side fashion, while the veins are anastomosed in an end-to-end fashion. If consideration is being given to performing an arterial anastomosis in end-to-end fashion to either the radial or ulnar artery, the surgeon must verify a patent palmar

arch with an Allen's test or Doppler examination prior to surgery. The donor leg for the anterior lateral thigh flap should be free of concomitant soft tissue trauma. If the patient has a history of lower extremity arterial atherosclerotic disease or diabetes, examination of the lower extremity is warranted, to verify that there is an intact profundus femoral artery which gives rise to the lateral femoral circumflex system.

Patient positioning

The patient is positioned in the supine position for harvest of the anterior lateral thigh flap. The injured arm is positioned on a standard hand table. Recipient site preparation is aided with the use of an upper extremity tourniquet.

Flap harvest is usually performed under general anesthesia, though spinal/epidural block for flap elevation and axillary block of the affected extremity can be undertaken when general anesthesia is contraindicated.

Flap Elevation Surgery

General principles and steps in elevation for perforator flap include:

- Doppler mapping of the perforator,
- Design of the flap for operation,
- Identification of the perforators leading to the main pedicle,
- Intra-muscular dissection of the perforator with preservation of the nerves,
- Elevation of the flap,
- Thinning of the flap, and

- Transfer of the flap to cover the defect.

The anterolateral thigh flap is based on the lateral femoral circumflex vessel, particularly the descending branch, which courses inferiorly along the intermuscular septum giving rise to subcutaneous and/or intramuscular perforators that penetrate the fascia to supply the skin and subcutaneous tissue overlying the anterolateral thigh. A preoperative Doppler examination of the lower leg is performed to identify these perforators. A line is drawn from the anterior superior iliac spine to the lateral margin of the patella. At the midpoint of this line is a reliable perforator identifiable in almost 90% of cases (Fig. 18-2). An additional perforator can be identified in 80% of cases within the upper third of the line, again using a hand held Doppler. The skin paddle, incorporating both perforators if possible, is designed in an elliptical fashion based on the dimensions of the defect in the upper extremity.

FIGURE 18-2 Prior to flap elevation the arterial perforators to the ALT flap are identified on the leg. A line is drawn from the anterior superior iliac spine (ASIS) to the lateral margin of the patella. In 90% of cases a major perforator can be identified within 3 cm of the middle portion of this line. A second major perforator is usually identified at the upper third of this line. A: The sites for the perforators are confirmed with a handheld Doppler probe and marked with a skin scribe. B: The flap is then designed around the perforators in an elliptical fashion.

Dissection begins medially by dissecting down to the rectus femoris fascia. The fascia of the rectus femoris muscle is divided and dissection proceeds in a lateral direction until the intermuscular septum separating the rectus femoris and vastus lateralis muscles is encountered. Medial retraction of the rectus femoris muscle exposes the entire septum, allowing for visualization of the descending branch of the lateral femoral circumflex artery in addition to its branch going to the rectus femoris itself (Fig. 18-3). At this point there may be 1â€“3 major perforators exiting from the descending branch and passing through the muscle of the vastus lateralis as musculocutaneous perforators or passing directly through the septum as septocutaneous perforators to the skin. Usually, one large perforator is sufficient to supply a large skin island, and allows for easy thinning of the flap. However, more perforators may be included if the skin island is very large, and flaps as large as 15 Ã— 35 cm

have been reported. If septocutaneous perforators are present, one may simply proceed to the posterior incision, including the portion of the tensor fascia lata, and elevate the flap. If intramuscular perforators are encountered then dissection must proceed as one would for a true perforator flap, in which case the artery needs to be carefully dissected free from the surrounding muscle. A cuff of muscle may be included in the flap should the upper extremity wound require additional bulk or if the wound contains exposed bone or significant dead space (see Figure 18-3D).

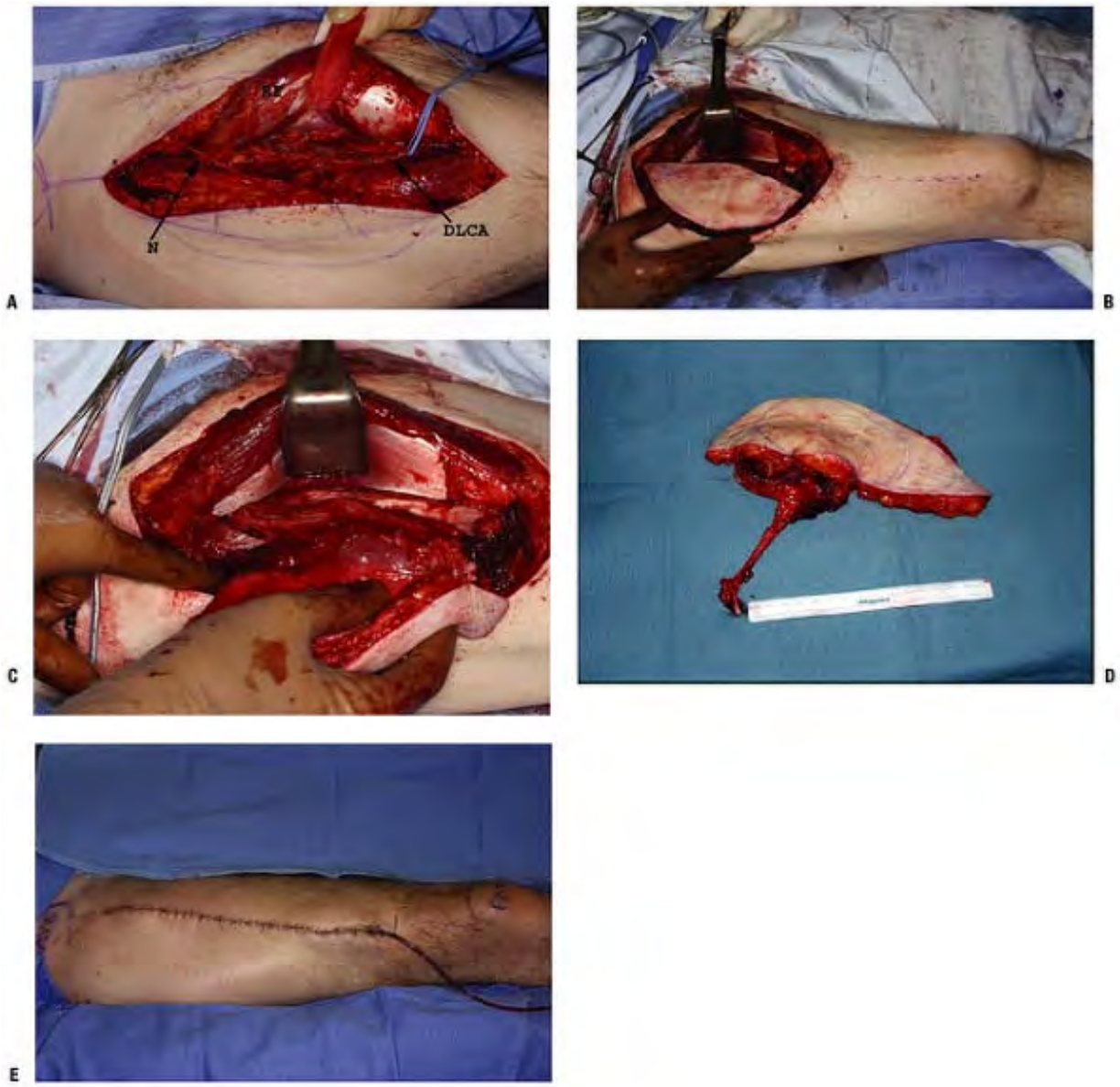

FIGURE 18-3 A: After the flap has been marked the medial incision is made first and the rectus femoris muscle is identified. The rectus femoris (RF) muscle is retracted medially to reveal the descending branch of the lateral femoral circumflex vessels (DCLA). A branch of the lateral femoral cutaneous nerve (n) can usually be identified running superior to the vastus lateralis muscle B: Once the flap perforators are identified, dissection may begin at the lateral margin of the flap. C: In this photograph 2 perforators of the lateral femoral circumflex vessels are seen passing into the vastus lateralis. These vessels may be dissected

through the muscle into the fascia of the overlying skin flap or a small cuff of muscle may be preserved around the vessels. D: The flap is now ready for transfer to the arm. The muscle cuff can be seen beneath the skin paddle. The pedicle length and large diameter allow for easy insetting and microvascular anastomosis. E: The donor site may be closed primarily over a closed suction drain.

Dissection proceeds along the descending branch of the LFCA separating venae comitantes from the artery. The pedicle can be dissected back to the main trunk of the lateral femoral circumflex artery, or divided at the descending branch. If one traces the vessels back to the origin, the motor nerve to the vastus lateralis will need to be dissected free from the arterial pedicle. This may be a tedious dissection as the nerve may pass through and around the venous pedicle. Dissection of the vessels back to the origin results in very large caliber vessels of up to 3 mm in some cases.

Preparation of the arterial recipient site is then performed. If the radial artery is going to be used, a Henry approach is usually carried out. For approach to the brachial artery a curved lazy-S incision is performed over the antecubital fossa and exposure of the brachial artery at the level of the medial forearm is performed. The flap is then transferred to the upper extremity. The tourniquet is then released in order to verify adequate arterial inflow. Once adequate arterial inflow has been verified, the anterolateral thigh flap may be divided from the lower extremity and transferred to the upper extremity. The anastomosis is then performed in end-to-side fashion using 8-0 nylon. Venous coupling may also be used to expedite the microsurgical procedure (Fig. 18-4).

The donor site is approximated primarily over closed suction drainage when flap width is less than 8 cm, or a skin graft may be used for larger width flaps. For cases involving intramuscular perforator dissection, the muscle edges are reapproximated to preserve maximal quadriceps function.

In certain cases, such as coverage of the hand, a thin flap may be required and primary defatting of the flap can be undertaken. The flap is usually thinned by excising the deep fat consisting of wide, flat fat lobules up to the junction of the superficial fat, made up of smaller, round lobules. Defatting before ligation of the pedicle allows for monitoring of flap perfusion during the thinning process. The flap

may be thinned up to 3 mm without compromise to the blood supply. A more conservative approach towards primary flap thinning should be undertaken until adequate experience is gained as flap debulking can be performed safely and easily as a secondary procedure (Fig. 18-5). Innervated flaps can be achieved by including the lateral femoral cutaneous nerve in the proximal portion of the flap. The nerve arises from the deep fascia approximately 10 cm caudad to the anterior superior iliac spine and divides into two or three branches. As previously mentioned, when added bulk is required to cover a complex wound, the anterolateral thigh flap can be harvested as a chimeric flap along with the vastus lateralis or rectus femoris muscles, the tensor fascia lata, or other skin flaps in the thigh based on the supply of lateral femoral circumflex system.

Postoperative Management

Monitoring

Flap monitoring is of paramount importance following microvascular surgery. Postoperative thrombosis of either the artery or vein can be salvaged in greater than 50% of patients if detected early. Intraoperatively, a stitch is placed over the perforator on the skin paddle to facilitate postoperative monitoring with a Doppler probe. Alternatively, an implantable Doppler probe can be placed around the vein or artery intraoperatively to allow for continuous monitoring of the anastomosis. The Doppler signal over the marked skin paddle is checked hourly with the use of a hand held Doppler while the patient remains in the Intensive Care Unit. The highest incidence of postoperative arterial thrombosis is within the first 24 hours, with the incidence of venous thrombosis occurring most frequently within the first three days. Donor site drain output is recorded and drains are kept in place until daily output is less than 30 cc. The patient is allowed to ambulate on postoperative day 2. Postoperative complications include partial to total flap loss, temporary weakness in the lower extremity which usually resolves within the first two weeks, and sensory deficit at the donor site if there has been any stretching or injury to the lateral femoral cutaneous nerve. In cases of venous congestion, exploration of the anastomosis is necessary. If prolonged venous thrombosis is experienced, the use of either TPA or leeches can sometimes be used to salvage the flap.

The use of postoperative anticoagulation is debated. Our patients are given one baby

aspirin per day, and are continued on subcutaneous heparin while they are in the hospital. Formal anticoagulation with heparin is discouraged and has been linked to an increase risk of hematoma formation and

P.229

P.230

P.231

flap loss. For those situations where arterial inflow has been difficult to establish, dextran at 25 mL/hour may be used in adults.

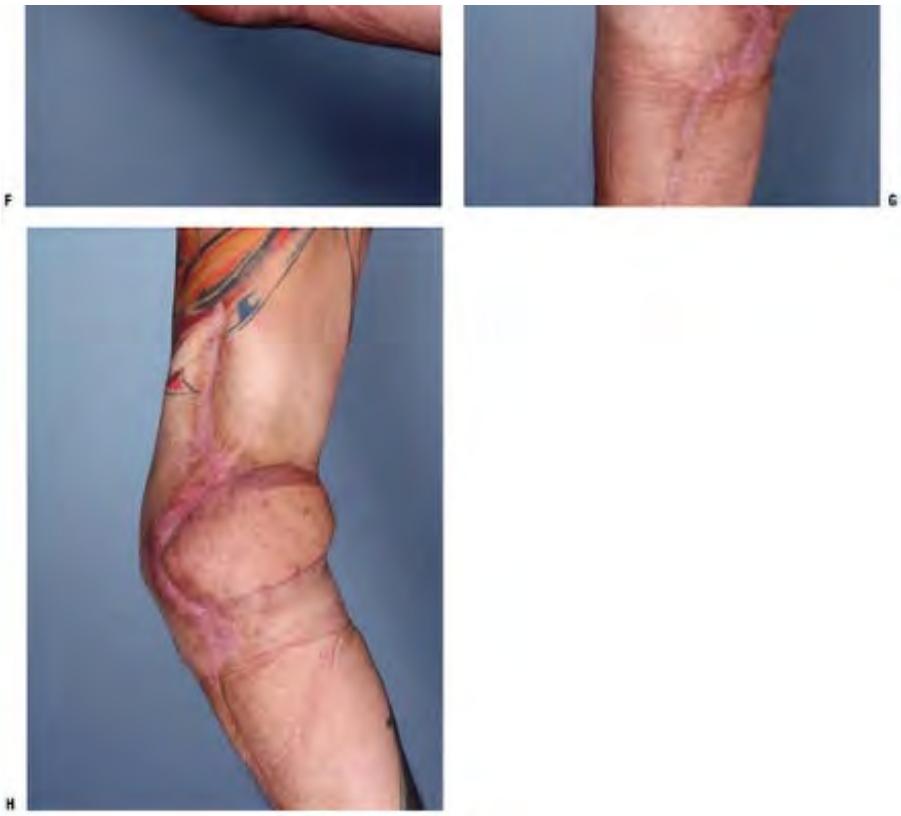

FIGURE 18-4 A: 34 year old male, status post liver transplantation, with a chronic left posterior elbow wound. The multiple posterior scares are evidence of previous attempts to close this defect. B: The skin incision for an ipsilateral ALT flap is marked on the thigh after identification of the underlying perforators. C: The flap is separated from surrounding tissue. D: The lateral femoral circumflex vessels can be seen above the surgeon's index finger entering the flap. E: In preparation for flap transfer the brachial artery is exposed at the level of the elbow. Fâ€“H: The result at 3 months postop with well healed wound, with good contour and no signs of infection.

FIGURE 18-5 Aâ€“D: This 49 year old man was involved in a farming accident which resulted in the amputation of the index thru small fingers at the level of the metacarpal bases. Soft tissue coverage was obtained with the use of a thinned ALT flap. Six months following reconstruction, the ALT flap has provided adequate contour and durability allowing patient to return to farming activities. The patient refused further surgery for finger reconstuction (C,D).

Rehabilitation

Underlying injuries usually dictate when motion may begin. We usually wait 5 days for the anastomosis to mature and for the postoperative swelling to subside. At this point if the underlying bone injury is stabilized, early mobilization is initiated. A light ace wrap is applied to the flap beginning on day five to help aid in resolution of post-operative swelling. Re-elevation of the flap for additional bone grafting or tendon grafting can be done as soon as 4 to 6 weeks. Defatting or thinning of the flap is usually delayed for 6 to 9 months after the original surgery.

Complications

The greatest complication of free flap surgery is flap loss due to arterial or venous thrombosis. The best means of preventing these problems is adequate preoperative planning. Repairs within the zone of injury, technical errors including inadequate visualization of vessels, and tension or kinking at the anastomosis site can all lead to flap failure. Ongoing infection can also result in partial flap loss. For the anterolateral thigh flap, twisting of the pedicle can occur easily due to the absence of muscle to keep its proper orientation. Marking the anterior wall of the pedicle and lifting of the flap in the air to allow the pedicle to lie in its natural orientation prior to final inset may prevent pedicle twisting. It is also important to avoid compression of the flap.

Results

Good success has been reported with the use of anterolateral thigh flaps in the upper extremity. Flap failure rates have been noted to be as low as 2%. Additional thinning procedures may be required in patients with continued bulk. Microvascular surgery and free tissue transfer within the upper extremity is facilitated by large recipient vessels, and a relatively shallow operating field.

Conclusion

With proper planning, the anterolateral thigh flap is capable of managing many complex injuries within the upper extremity. It is now often our first choice for coverage of upper extremity defects. The advantages of the ALT flap are: relatively

consistent anatomy; ease in dissection; long pedicle length; ability to achieve a thin, pliable flap; ability to provide sensate coverage through the lateral femoral cutaneous nerve; no sacrifice of major artery of the lower limb; and versatility in flap construct which allows one to harvest not only muscle, but also functional muscle such as the rectus femoris. The donor site can be closed primarily.

Recommended Reading

Breidenbach WC. Emergency free tissue transfer for reconstruction of acute upper extremity wounds. *Clin Plast Surg.* 1989;16:505â€“514.

Celik N, Wei FC, et al. Technique and strategy in anteriorlateral thigh perforator flap surgery, based on an analysis of 15 complete and partial failures in 439 cases. *Plast Reconstr Surg.* 2002;109:2211â€“2216.

Chen HC, Tang YB, et al. Reconstruction of the hand and upper limb with free flaps based on musculocutaneous perforators. *Microsurgery.* 2004.

Chen SH, Wei FC, et al. Emergency free-flap transfer for reconstruction of acute complex extremity wounds. *Plast Reconstr Surg.* 1992;89:882â€“888.

Derderian CA, Olivier WA, et al. Microvascular free-tissue transfer for traumatic defects of the upper-extremity: a 25 year experience. *J Reconstr Microsurg.* 2003;19:455â€“462.

Godina M. Early microsurgical reconstruction of complex trauma of the extremities. *Plast Reconstr Surg*. 1986;78:285â€“292.

Hamdi M, Van Landuyt K, et al. A clinical experience with perforator flaps in the coverage of extensive defects of the upper extremity. *Plast Reconstr Surg*. 2004;113: 1175â€“1183.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 19

Groin Flap Coverage of the Hand and Wrist

Jeffrey B. Friedrich

Nicholas B. Vedder

Indications/Contraindications

The pedicled groin flap is an extremely versatile and reliable flap that was initially described in 1972 by MacGregor and Jackson. It enjoys a revered place in the field of reconstructive surgery because it was one of the first axial-pattern flaps to be described and applied in humans. After several decades, many surgical techniques, including pedicled flaps, have been supplanted by new methods. However, despite the advent of microsurgery, the pedicled groin flap continues to be a venerable technique with a variety of applications.

The groin flap has been used most commonly for upper extremity reconstruction, and is indicated for soft tissue defects in which there is destruction or devitalization of soft tissue on the hand, wrist, and distal forearm. The groin flap is not designed to supplant skin-grafting on wounds that will accept a graft; rather, it is to be used when the injury or resection leaves exposed structures that will not accept a skin graft, such as tendon or bone. It has been described for soft tissue reconstruction of a number of traumatic etiologies including crush injuries, avulsions, gunshots, blast wounds, and burns. Additionally, the flap is warranted for soft tissue defects following elective resection such as contractures, hypertrophic scarring, or tumor, and has even been described as soft tissue coverage following recurrent carpal tunnel syndrome.

In the setting of trauma, the groin flap is most often used for delayed primary coverage of upper extremity wounds. Several groups have employed the flap for very early coverage (within 24â€“48 hours from injury); however, the complication rate has been found to be high. Therefore, unless a clean and healthy wound bed can be ensured initially, it is generally prudent to thoroughly debride the wound, observe it for a few days, and then apply the groin flap only after a clean, healthy wound surface is ensured.

As noted above, the groin flap will provide reliable soft tissue coverage to practically any portion of the hand as well as the distal forearm. It is particularly well-suited to the dorsal surface of the hand. The flap can certainly be used to provide coverage for the palmar surface of the hand; however, the flap can be bulky, and does not at all approximate the native, adherent, glabrous skin of the palm. This can leave the palm with a bulky flap that is prone to shearing motion when weight is borne on the palmar surface of the hand, such as seen with pushing, gripping, tool use, or twisting motions. For this reason, it is wise to consider other thinner flaps such as the reversed radial forearm flap or free fasciocutaneous flaps such as the anterolateral thigh flap for most palmar reconstructions.

A brief mention must be made of the groin flap for use as a free tissue transfer. The free groin flap was the first free flap described and can be used for many of the same indications for which other free flaps are currently used. Its reliability is somewhat limited because of the variable arterial pedicle and the short length of the pedicle. Furthermore, even in experienced hands, the free groin flap failure rate is notably higher than it is for other free flaps, thus limiting its utility.

Another chief indication for use of the groin flap is in the setting of thumb reconstruction. Often with total or near-total thumb loss, it is necessary to use free digit transfers such as the great or second toe. Unfortunately, traumatic thumb amputations can leave the patient with insufficient soft tissue of the thenar region. In these cases, the groin flap is an excellent choice to provide a soft-tissue â€œbaseâ€• on to which the digit transfer can be placed at a later date.

The contraindications for the use of the pedicled groin flap are few. In general, the flap should not be implemented in patients who are deemed to be uncooperative, or who would otherwise not be able to tolerate a 2- or 3-week interval in which their

affected hand is immobilized and attached to the groin. In the past, it was felt that elderly patients were not good candidates for this flap because of the risk of shoulder and elbow stiffness. However, a recent study has shown that elderly patients are able to tolerate the procedure with no more complications (including upper extremity stiffness) than the average young patient.

As with any flap, active tobacco use is a concern with the groin flap. Specifically, the flap can be subject to marginal necrosis in active smokers. The surgeon must balance the need for complex reconstruction with the potential morbidity that may be incurred by patients who smoke.

Previous groin surgery may or may not be a contraindication to the use of this flap. In general, inguinal hernia repair does not preclude the use of the groin flap, nor do other scars in the general vicinity such as abdominoplasty or Pfannenstiel incisions. However, a previous inguinal lymph node dissection would render this flap tenuous, or outright unsafe. The potential disruption of superficial vessels in the groin is too high to risk use of a flap based on these vessels. Additionally, patients who have had groin lymph node dissections have quite often had external beam radiation to the same region, further rendering flap dissection in this area a dangerous proposition.

Preoperative Planning

When using the groin flap in the acute posttraumatic setting, wound preparation is of paramount importance. The surgeon must perform any hand or forearm repairs that are necessary in the acute phase. This includes fracture reduction and fixation, as well as tendon, vessel, and nerve repair. Wound debridement must also be complete so as to minimize the infection risk once the site is covered with a flap. Wound debridement and cleansing may need to be done several times in the operating room prior to flap coverage, and several days of dressing changes may be necessary to ensure that the wound bed is as healthy as possible. This is especially true with crush and avulsion injuries.

In the posttraumatic setting, timing of coverage with a groin flap has become an important issue. In general, most published reports advocate use of the flap in a delayed-primary fashion, specifically between 48 hours and 7 days following injury. Others have attempted to employ the flap in a more acute setting (24â€“48 hours following injury); however, the consensus seems to be that early timing leads to an

unacceptably high rate of infection and dehiscence, and is thus not advocated. In the end, timing of flap implementation should be based on the surgeon's best clinical judgment about the mechanism of injury, and the status of the wound bed.

In the elective setting, the preoperative preparation for a groin flap is somewhat less crucial. The patient must understand that the interval between flap elevation and flap division will entail a moderate amount of disability, and will require a significant amount of diligence in regard to their self-care. Additionally, it is a good idea for the patient to make prior arrangements to have a friend or family member at home who will be able to assist them with various activities of daily living.

Surgery

The groin flap coverage surgery must be done with a general anesthetic. Additionally, it is recommended that a bladder catheter be placed as the operation can be in excess of 3 hours. If the patient is not already being administered intravenous antibiotics for prophylaxis of a traumatic wound, one dose of antibiotics should be administered 30 minutes before skin incision.

Patient Positioning

Once anesthetized, care should be taken with patient positioning. Although an ipsilateral groin flap is most commonly used, a contralateral flap is sometimes acceptable if the ipsilateral side is for some reason unavailable. Regardless of which side is chosen, a small towel roll or â€œbumpâ€• should be placed under the hip of the side that will undergo groin flap dissection to allow extension of the flap

dissection back to the flank. The upper extremity to be reconstructed should be placed on a hand table for initial wound preparation. Additionally, one should leave enough room on the side of the operating table to be able to comfortably lay the upper extremity beside the torso during groin flap inset. The other necessary step before initiating the operation is placement of a pneumatic tourniquet on the upper extremity to be reconstructed.

Technique

The operation is begun by preparation of the recipient bed. If the flap is being

employed for traumatic wound coverage, a final thorough debridement is done at this time. If it is being done for scarring or other elective soft tissue coverage, the creation of the recipient wound to be covered is done in its entirety before moving to the groin.

In preparation for flap elevation, measurements are taken to determine the required dimensions of the groin flap. This can be done in several ways. One can simply use a ruler to determine the largest dimensions of the wound, or a template can be fashioned. The paper wrapper from surgical glove packaging makes a nice template material and is readily available. This can be pressed onto the wound so that the wound fluid makes an imprint in the paper. Then the template can be made simply by cutting around the outline of the defect. Alternatively, one can use a piece of the elastic wrap used to exsanguinate the hand (Fig. 19-1). The template is then transferred to the groin. In general, the more ulnar aspects of the wound will be covered by the distal end of the flap.

When designing the flap, one must keep in mind the arterial anatomy of the flap. Specifically, the flap is supplied by the superficial circumflex iliac artery and vein (Fig. 19-2). Numerous studies have shown the artery to consistently extend to the level of the anterior superior iliac spine (ASIS) before any significant caliber diminishment. In other words, the flap is an axial pattern flap medial to the ASIS, and a random-pattern flap lateral to this point. The area lateral to the ASIS can be used as part of the flap design; however, its effectiveness is limited by its random nature. Therefore, the operator should generally design a flap that, at the portion lateral to the ASIS, has a width:length ratio between 1:1 to 1:1.5 (although the medial portion needs only be wide enough to include the axial vessels). It is possible to design a flap whose distal extent is more posterior than this limit, but a delay procedure should then be employed for the flap. A delay procedure involves incision, but not elevation

of the desired flap. The incised flap is then left in situ for a week so that the circulation at the distal end of the flap can be augmented during that time. Following that interval, the flap can then be elevated and inset in the upper extremity wound.

FIGURE 19-1 Use of a piece of the elastic exsanguination wrap to create a template of the defect to be covered with a groin flap.

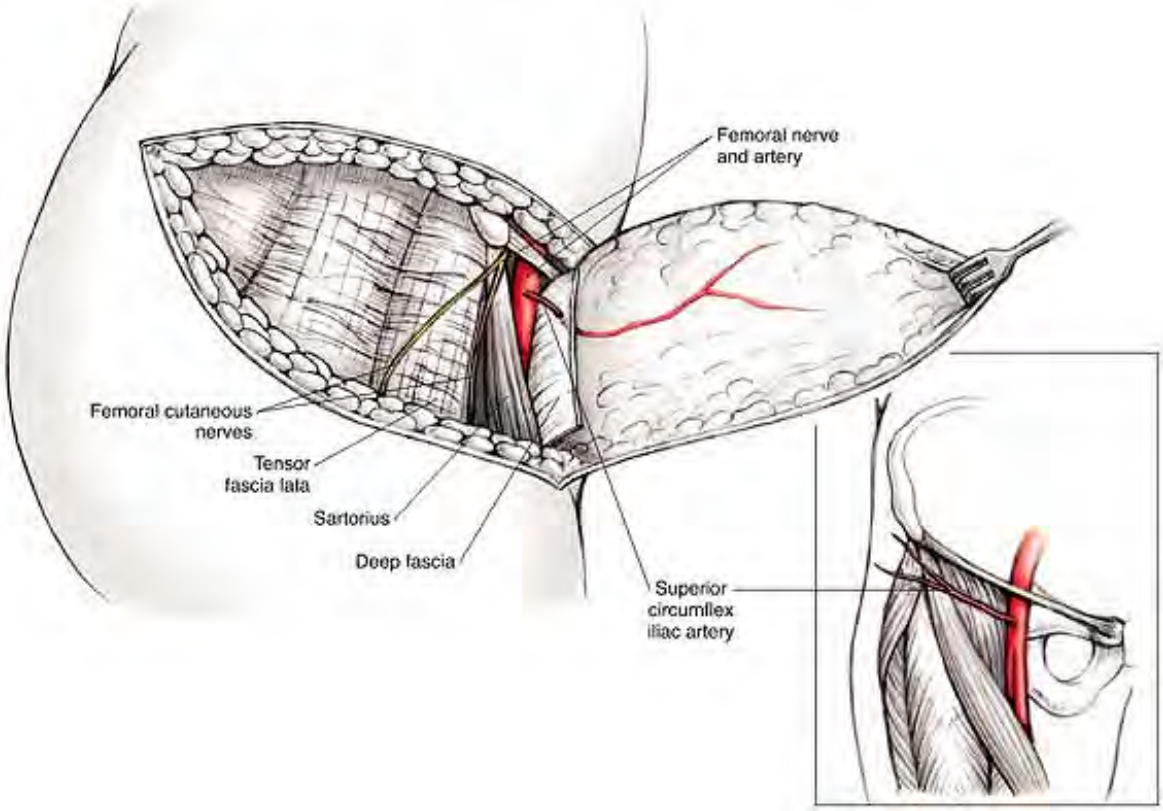

FIGURE 19-2 Detail of groin flap anatomy (right-sided), showing the pedicle (superficial circumflex iliac artery), inclusion of the superficial sartorius fascia with the flap, and the location of the lateral femoral cutaneous nerve branches as they emerge from under the inguinal ligament. (Adapted from Brown EZ Jr, Pederson WC. Skin grafts and skin flaps. In: *Green's operative hand surgery.* 5th ed. Philadelphia: Elsevier/Churchill Livingstone; 2005:1692-1702, with permission.)

When beginning the overall design of the flap, it is best to keep in mind the recommendations of Chuang, who published a set of groin flap design rules based on experience amassed via the use of over 200 groin flaps. The mainstay of these guidelines is the â€œtwo fingerbreadthsâ€• rule (the two fingerbreadths being at the level of the index and long finger distal interphalangeal joint of the operator) (Fig. 19-3). The first step is to estimate the location of the superficial circumflex iliac artery, done by marking a point two fingerbreadths below the junction between the inguinal

ligament (which runs along a line between the anterior superior iliac spine and the pubic tubercle) and the femoral artery. Once that artery location is estimated, its true path should be verified by Doppler ultrasound. The upper flap border is marked two fingerbreadths above the inguinal ligament (parallel to the artery), while the lower flap border is a line two fingerbreadths below and parallel to the artery. Also, as will be further discussed following, the superficial branch of the superficial circumflex iliac artery (SCIA) emerges from the deep fascia two fingerbreadths medial to the ASIS, then continues laterally in a subcutaneous plane.

It should be noted that if using the two fingerbreadths rule for flap design, a relatively slender flap will result medially. In reality, when the flap is centered over the SCIA, the cephalad-caudad width of the flap lateral to the ASIS can be quite substantial. However, most authors agree that flap widths up to 10 cm enable primary closure of the groin wound site, while those that are wider stand a good

chance of requiring closure with a split-thickness skin graft because of inability to achieve primary closure. It is usually possible to determine how wide a flap can safely be harvested and closed primarily by pinching the skin and subcutaneous tissue together. Flexing the operating table at the hip to relieve some of the tension can facilitate closure.

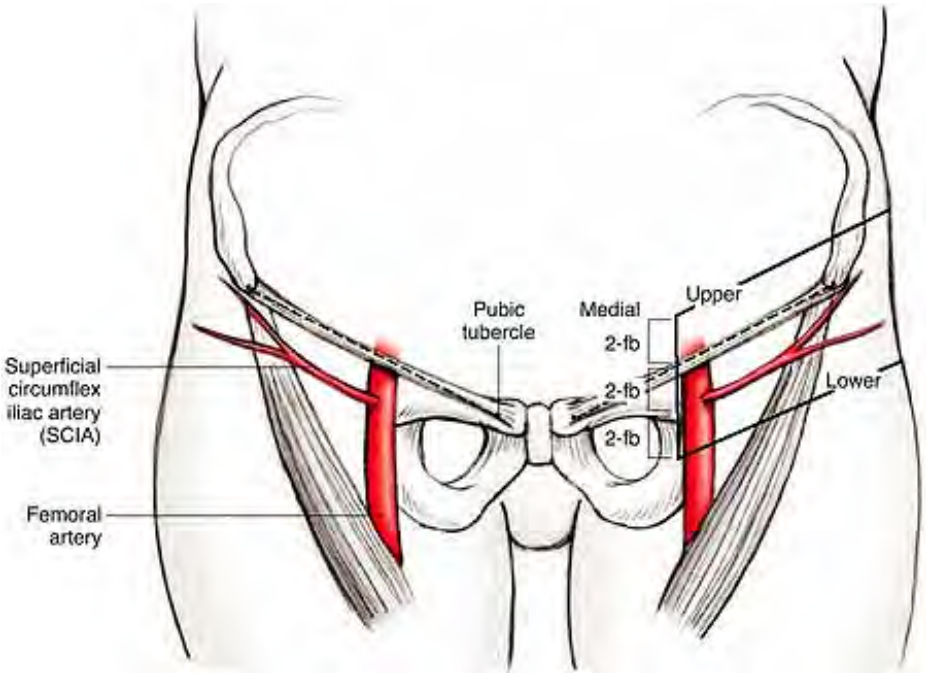

FIGURE 19-3 Drawing of the application of the two-fingerbreadth rule. The *31* on the left of the patient is the location of the inguinal ligament. The *31* immediately below that is the location of the superficial circumflex iliac artery. Note that the upper border of the flap is two fingerbreadths above the inguinal ligament, and the lower border is two fingerbreadths below the vascular pedicle (SCIA). (Adapted from Chuang DC, Colony LH, Chen HC, Wei FC. Groin flap design and versatility. *31* 1989;84(1):100â€“107, with permission.)

Once the marks are drawn, the incisions of the flap are made. The incision is carried down to the level of the external oblique fascia on the abdomen, and the muscle fascia on the leg. Then the flap dissection is begun at the lateral end in the plane between the subcutaneous fat and the muscle fascia of the thigh (below the inguinal ligament); and between the subcutaneous fat and the external oblique muscle fascia (above the inguinal ligament). The flap is most easily and safely raised with curved Metzenbaum scissors, although low-powered electrocautery can be used as well. The dissection proceeds in this fashion until reaching the lateral border of the sartorius muscle. On reaching the sartorius, the fascia over this muscle is incised at its lateral border. The sartorius fascia is then elevated from the muscle. Above the inguinal

ligament, the dissection plane remains immediately superficial to the external oblique fascia. At the medial border of the sartorius muscle, the SCIA divides into a superficial and a deep branch. By elevating the sartorius fascia, the superficial branch (which supplies the lateral extent of the groin flap) is preserved with the flap. When the medial sartorius fascia is reached, and the deep branch of the SCIA encountered, it is divided and ligated with either vascular clips or sutures. This deep branch of the SCIA is not necessary for flap survival, and it will only serve to tether the flap if more flap length is desired (Fig. 19-4).

During the flap dissection, one must remain cognizant of the lateral femoral cutaneous nerve and its course in the vicinity of the flap. The nerve emerges from under the inguinal ligament at the medial edge of the sartorius muscle (see Fig. 19-2). The nerve branches run deep to the SCIA; therefore, care should be taken to preserve the nerve branches in this area. If nerve branches are found in the sartorius muscle fascia, efforts must be made to dissect free and preserve them. Despite attempts to preserve the nerve branches, patients should be told before the procedure that nerve damage (both transient and permanent) is possible and can lead to numbness of the lateral thigh.

The endpoint of flap dissection is generally determined by the acquisition of enough flap length to be easily inset in the defect of the hand or forearm. Some references caution against proceeding more medially than the lateral border of the sartorius muscle; however, we have found that a longer flap can be raised if proper precautions are taken. That said, the absolute anatomic endpoint of flap dissection is at the lateral border of the femoral artery.

While this chapter chiefly concentrates on the use of the pedicled groin flap, a brief mention of the free groin flap will be made. The dissection of the flap proceeds in an identical manner to that described previously. However, one of the over-arching goals of free flap harvest is to obtain donor

vessels that are as long as safely possible. With the free groin flap, this involves dividing the SCIA very close to the femoral artery, and dividing the donor vein very close to the femoral vein. This must be done carefully as vascular clips or sutures placed too close to the femoral vessels can serve to narrow the lumen of the vessel. When harvesting the groin flap for free tissue transfer, one anatomic caveat must be

kept in mind: in a minority of persons (estimated at 30%), the SCIA emerges from a common vessel that bifurcates into the SCIA and the superficial inferior epigastric artery (SEIA). When this is found to be the case, the common trunk is dissected back to the femoral artery and used as the arterial donor. Both the SCIA and SEIA will reliably supply the groin skin; so when they emerge separately, the larger of the two vessels is chosen as the arterial donor. Usually, the larger vessel is the SCIA.

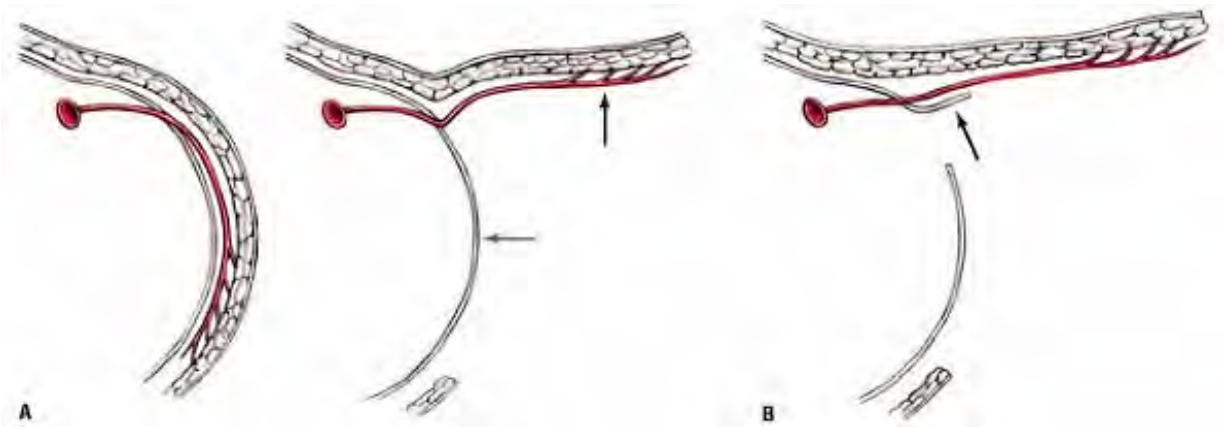

FIGURE 19-4 A: Schematic representation of vascular kinking that is possible if the sartorius fascia is not elevated with the groin flap. The illustration shows an axial view of the lower extremity. The *31* points to the superficial circumflex iliac artery (SCIA); the Î points to the fascia of the sartorius muscle. B: Schematic representation of the relatively straight course of the superficial branch of the SCIA following elevation of the sartorius fascia with the groin flap. The *31* points to the now-released sartorius fascia. (Adapted from Brown EZ Jr, Pederson WC. Skin grafts and skin flaps. In: *Green's operative hand surgery.* 5th ed. Philadelphia: Elsevier/Churchill Livingstone; 2005:1692â€“1702, with permission.)

Several steps must be taken before inset of the flap into the upper extremity. First, one must close the donor defect, because closure will be much more difficult if it is done after flap inset. This closure should optimally be a three-layered one, with 0-0 or 2-0 braided absorbable suture in the superficial fascia (Scarpa's fascia on the abdominal wall to superficial fascia on the lower extremity), 2-0 or 3-0 braided

absorbable suture in the dermis, and 3-0 or 4-0 nonabsorbable monofilament or staples in the skin. This closure is done over a suction drain, usually brought out through a separate stab incision. The closure is facilitated by flexion of the operating table at the hip, and use of towel clips or staples to temporarily approximate the skin while the sutures are being placed (Fig. 19-5). It is also important for the patient's bed to remain flexed at the hip in the immediate postoperative period to minimize tension on the suture line.

Next, the flap base should be â€œtubedâ€• before inset. Tubing of the flap diminishes the amount of exposed flap undersurface, thereby helping with postoperative wound care. The tubing is done simply by suturing the upper and lower sides of the flap together such that it forms a cylinder (Fig. 19-6). The closure of the tube can be done with a single layer of nonabsorbable monofilament sutures. One must be cautious about tubing the flap, as this can cause vascular compromise. A simple way to deal with a congested flap is to release a few sutures at the base (since this is the thickest part of the flap). In a seminal article on groin flap design, Schlenker illustrates another method to increase the circumference of the tube while simultaneously positioning the flap for inset depending on whether the hand defect is dorsal or volar. This is done by incising either the cephalad or caudad side of the flap longer than the other, then performing either a clockwise or counterclockwise closure of the donor defect. This then positions the portion of the flap that will be inset into a more optimal orientation for coverage (Fig. 19-7).

FIGURE 19-5 Flexion of the hip to facilitate groin flap donor site closure.

The other task to be done before flap inset is to thin the distal part of the flap that was originally lateral to the ASIS. This is most easily performed with curved Mayo or Metzenbaum scissors, trimming the fat down to the subdermal level. This will facilitate flap inset, and will reduce the amount of defatting that needs to be done in the secondary setting.

Once elevated, the inset of the groin flap is relatively simple. The affected hand is moved into the groin, and the flap is then inset into the recipient site, usually over a closed-suction drain. This is typically performed with a two-layered closure,

specifically 3-0 or 4-0 braided absorbable suture in the dermis, and 3-0 or 4-0 nonabsorbable monofilament suture in the skin (Fig. 19-8).

FIGURE 19-6 Groin flap following â€œtubingâ€• of the proximal portion of the flap.

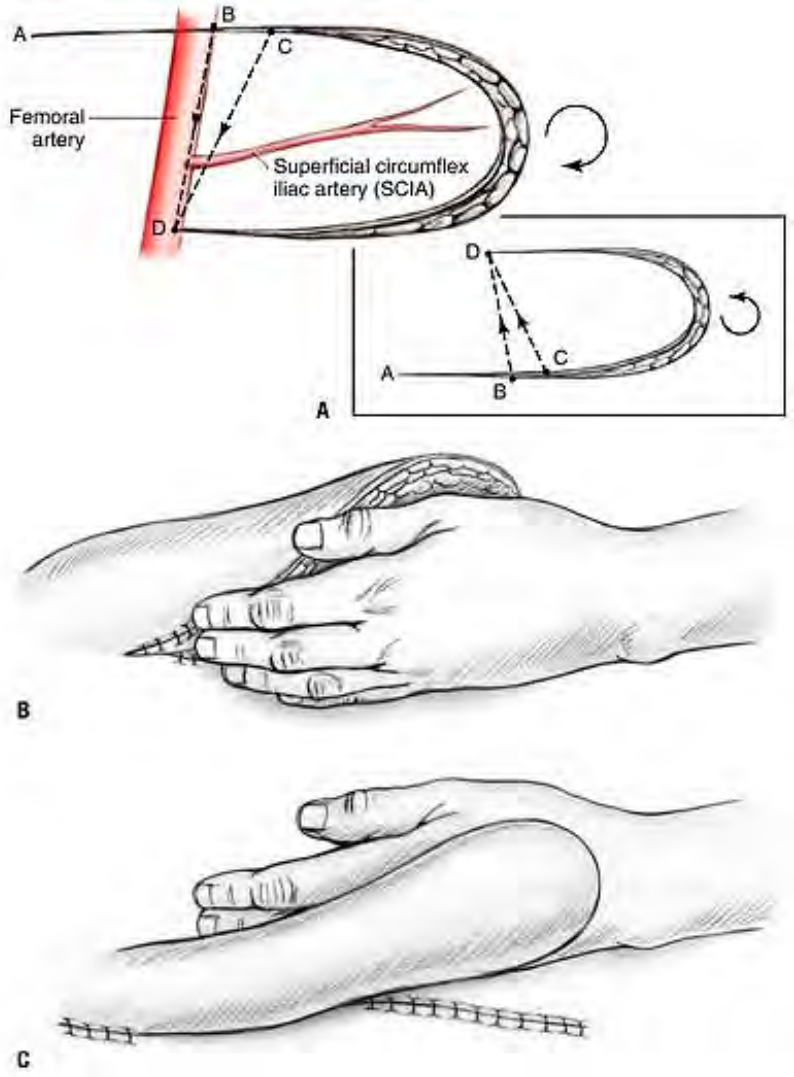

FIGURE 19-7 A: Schlenker's modification of the groin flap incisions. By making the top incision longer, and closing points B and C to point D, clockwise rotation is facilitated. Similarly, by making the lower incision longer, counterclockwise rotation is facilitated. B: By creating a flap that is rotated clockwise, and therefore facing downward, closure of a palmar defect is made easier. C: Similarly, counterclockwise rotation of the flap causes the raw surface to face upward, allowing easier closure of a dorsal hand defect. (Adapted from Schlenker JD. Important considerations in the design and construction of groin flaps. *31* 1980;5(5):353â€“357, with permission.)

FIGURE 19-8 Completion of groin flap inset on the hand.

Care must be taken with postoperative dressings and positioning. The wounds on the hand can simply be covered with antibiotic ointment. We find that it is best to cover the small exposed portions of the flap (at the base) with antibiotic (bismuth)-impregnated gauze. The closed groin wound can be dressed with a gauze dressing or covered with antibiotic ointment. Cast padding or soft towels should be placed in the affected axilla and under the hand as these areas will become moist with the arm at the side. A pillow should be placed underneath the elbow to give the arm support, and to ensure that the flap itself is just the slightest bit extended anterior and lateral to the body. It is important to avoid kinking of the flap since this will easily cause venous congestion. Finally, the donor-side leg should have a pillow placed under the knee to relieve tension on the groin flap closure. Once the patient is awake, it is important to educate them in positioning of the flap to avoid kinking.

Flap Division and Final Inset

In general, the pedicle of the groin flap is divided between 2 and 3 weeks after flap

elevation. Wei's group and others have done numerous studies using ischemic preconditioning of the flap to allow flap division at an earlier time (7â€“10 days). However, we have found that, with proper patient selection, the groin flap will be tolerated for 3 weeks just as easily as it would be tolerated for 1 week. Furthermore, we have not found any complications associated with the longer attachment times.

Flap division proceeds as follows: one should test the vascularity of the flap by firmly occluding the tubed pedicle with thumb and forefinger or by using a penrose drain clamped around the pedicle, then observing capillary refill of the flap. If there is *any* question about the perfusion of the flap, the flap division should be delayed. A hand table should be set up before the start of the case. An arm tourniquet is not necessary for this procedure, and in fact, it is discouraged because the vascularity of the flap needs to be assessed during this operation. The arm is prepared from at least the elbow down, and the entire groin area is prepared simultaneously. The flap is marked where it is thought that it will cover the remaining wound, taking care to make this marking generous. The flap is then incised with a scalpel and carried through with electrocautery. The arm is then placed on the hand table. Sometimes, there is a residual flap â€œstump,â€• which is excised to the point that the remaining groin wound can be easily closed without a large dog-ear deformity. If the stump of the SCIA is large, this should be ligated with a suture or vascular clips. The remaining groin wound is closed in a fashion similar to that of the original groin wound closure.

Finally, the remainder of the groin flap is inset to the hand. Some very judicious defatting of the open end of the flap can be undertaken at this time. This can be done with scissors, but it should not be overly aggressive. Both the donor and recipient sites should be well debrided and cleansed before closure. The closure should be loose to allow drainage, as these will be contaminated wounds. The wound is sutured with 4-0 nylon in the skin (Fig. 19-9). A soft dressing should then be applied to the

hand, usually with moist gauze in the layer next to the skin as this will encourage drainage from these contaminated wounds. We find it helpful to take advantage of the general anesthesia by manipulating the patient's elbow and shoulder before his or her awakening.

FIGURE 19-9 A: Photograph of same patient in Figure 19-6 after division of groin flap pedicle. The hand has a â€œmittenâ€• appearance after covering the finger stumps with the groin flap. B: Long-term results seen in patient from Figures 19-5 and 19-8B. The groin flap has been divided into individual flaps for each finger, and they have each been further contoured.

Postoperative Management

After the first stage of the groin flap procedure, most patients will need a significant amount of assistance from physical and/or occupational/hand therapists. The involvement of the therapist is essential to help with the transition to home life for the interval in which the hand is attached to the groin. For the purposes of practicality and cost of care, it is beneficial to send the patient home a few days after the initial stage of the procedure, rather than keeping the patient in an inpatient setting for 2 to 3 weeks. The therapist will assist the patient with normal self-care activities such as dressing (including any modifications of clothing made necessary by the attachment of the hand to the groin), hygiene, and other daily activities.

The chief purpose of the hand therapist is to ensure joint mobility to the degree that it is possible during the pedicled stage of the flap. This includes range of motion exercises with the involved shoulder and elbow, as well as the affected hand. In our experience, the groin flap is most often used to cover hand and wrist defects. Therefore, if the patientsâ€™ other hand injuries permit, we encourage them to work

on range of motion exercises with the hand, most commonly in the interphalangeal and metacarpophalangeal joints.

Postoperative wound care is relatively simple after the first stage of the surgery. The chief aim is to keep any flap surfaces that are exposed (i.e., any raw surfaces that are not covered by tubing of the flap) moist and clean. The application of antibiotic (bismuth)-impregnated gauze or antibiotic ointment serves this purpose quite well. Patients are encouraged to wash the wounds daily with soap and water and even to shower. Suture removal depends on the surgeon's preference: they can be removed in 10 to 14 days, or this can be done at the time of flap division 2 to 3 weeks after the initial procedure.

After the conclusion of the flap division procedure 2 to 3 weeks after flap elevation, physiotherapy continues to be an important part of the patientsâ€™ care. Range of motion exercises should be continued, especially in the affected elbow and shoulder. These joints can become quite stiff between the two procedures; however, this stiffness is most often transient if physical therapy is maintained. Finally, aggressive rehabilitation of the injured hand is a long-term endeavor.

Much later during the patient's care, debulking, revision, and functional improvement of the reconstruction can be taken into consideration. In many patients, the groin flap can be relatively bulky, potentially giving the reconstructed hand a â€œloaf of breadâ€• appearance. Once the flap has completely healed, this problem is relatively easily treated by partial opening of the flap followed by direct lipectomy in the subcutaneous plane, or by using suction lipectomy with a small (3 mm) suction cannula.

Complications

In the intraoperative setting, a potential complication is construction of a flap that is too wide in the cephalocaudal dimension, thereby preventing primary closure of the donor site. This is not a grave complication, but it does require closure of the donor site with a split-thickness skin graft, which obviously creates another donor site. Additionally, the appearance of the groin wound covered by a skin graft is less desirable than that of a groin wound that has been closed primarily.

The most dreaded complication of the use of the groin flap is flap loss, either partial

or total. Fortunately, the pedicled groin flap is quite hardy, making flap loss unusual if care is taken with flap design and tubing. Most series have found flap loss rates in pedicled groin flaps around 1%, and these are most often partial, rather than total flap losses. Not surprisingly, the rate of flap loss when using the groin flap as a free tissue transfer is higher than that of pedicled groin flaps. This can be attributed to the relatively increased difficulty of free groin flap dissection, and to the nature of free flaps in general. Cooper analyzed a series of 130 patients who had undergone free groin flaps for a variety of indications. The series had seven partial and nine total flap losses, and a total flap failure rate of 8.5%. This is slightly higher than most published series of free tissue transfers, where the general free flap failure rate is around 5%. Cooper's series corroborates these numbers in that they experienced an overall free tissue transfer failure rate (using flaps other than the free groin flap) of 4.2%, which is less than half of the failure rate of the free groin flap in their hands.

A complication that is of low morbidity, but can be somewhat distressing to the patient is lateral thigh numbness due to lateral femoral cutaneous nerve injury. Many of these injuries are transient, and sensation will return within several weeks. More distressing to the patient is permanent numbness of the lateral thigh. The true incidence of permanent nerve injury is difficult to gauge due to the variable amount reported in the literature (1%â€“50%). This can be prevented by careful dissection around the area of the ASIS during flap elevation, taking care to identify and protect the adjacent cutaneous nerves.

Joint stiffness of the elbow and shoulder, though usually transient, is almost a guarantee with the use of the pedicled groin flap. Joint stiffness was thought to be more pronounced and problematic in elderly (over 55 years) patients; however, Buchman published a small series of elderly patients who underwent groin flaps in which they contend that the severity of joint stiffness in this population is not any worse than that of a younger population. As mentioned in the previous section, aggressive physiotherapy is crucial to ensure that any joint stiffness is merely transient.

As with any procedure that involves a period of immobilization, deep venous thrombosis is a concern when dealing with patients who have undergone a groin flap reconstruction. Because groin flap patients are at least initially confined to bed, it is

recommended that these patients all be administered low-dose subcutaneous heparin or low-molecular-weight heparin as prophylaxis against deep venous thromboses.

Finally, there are a number of complications that are possible with just about every flap surgery, including the groin flap. Some groups have reported significant seromas in the groin donor site, but these are usually adequately treated with needle aspiration or the opening of a small part of the donor incision. Infection is certainly a possibility, especially with raw wound surfaces for a number of days. However, these are surprisingly uncommon, and are usually sufficiently treated with intravenous or oral antibiotics, along with limited wound opening.

Recommended Reading

Arner M, Moller K. Morbidity of the pedicled groin flap. A retrospective study of 44 cases. *Scand J Plast Reconstr Surg Hand Surg.* 1994;28(2):143â€“146.

Barillo DJ, Arabitg R, Cancio LC, et al. Distant pedicle flaps for soft tissue coverage of severely burned hands: an old idea revisited. *Burns.* 2001;27(6):613â€“619.

Buchman SJ, Eglseder WA Jr, Robertson BC. Pedicled groin flaps for upper-extremity reconstruction in the elderly: a report of 4 cases. *Arch Phys Med Rehabil.* 2002;83(6):850â€“854.

Cheng MH, Chen HC, Wei FC, et al. Combined ischemic preconditioning and laser Doppler measurement for early division of pedicled groin flap. *J Trauma.* 1999;47(1):89â€“95.

Cheng MH, Chen HC, Wei FC, et al. Devices for ischemic preconditioning of the pedicled groin flap. *J Trauma.* 2000;48(3):552â€“557.

Chow JA, Bilos ZJ, Hui P, et al. The groin flap in reparative surgery of the hand. *Plast Reconstr Surg.* 1986;77(3):421â€“426.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 20

Cross-Finger Flaps for Digital Soft Tissue Reconstruction

Bassem T. Elhassan

Alexander Y. Shin

Injuries of the fingertips are very common occurrences secondary to both domestic and industrial accidents and are among the most common hand injuries encountered in the emergency room and the hand practice. To function properly, the fingertip requires adequate padding with durable, sensate skin coverage. As such, any reconstructive procedure of the fingertip must include the provision of durable skin coverage, preservation of adequate sensation, maintenance of finger length, prevention of proximal joint stiffness, and restoration of cosmesis when possible.

Several techniques have been described to address tissue loss of the fingertips. These techniques include direct wound approximation, healing by secondary intention (i.e., allowing the wound to granulate), split and full thickness skin grafting, cross-finger and reverse cross (deepithelialized) finger flaps, adipofascial turnover flaps, revision amputation, volar V-Y advancement flap, Kutler's bilateral V-Y advancement, vascularized island flaps (homodigital flaps), and partial toe transfers (1,2,3,4,5,6). Each technique has its own indications and contraindications depending on the size of soft tissue loss and the concomitant injuries (exposed bone, tendon, nerve, or vessels).

Cross-finger flaps are local flaps that can reliably provide padding and durable cover for fingertip injuries and have been in use for over 50 years (7,8). The anatomy, indications and contraindications, technique, and results of treatment of the cross-finger flap and the reverse (deepithelized) cross-finger flap for coverage of fingertip injuries will be described.

Anatomy

Cross-Finger Flap

The cross-finger flap is a random local (regional) flap that is usually raised from the dorsum of the donor finger at the level of the proximal or middle phalanx. The blood supply of this flap is from the many small vessels of the subdermal and subcutaneous plexus. These plexi are fed by two major blood supplies. The first is the dorsal digital arteries that originate from the dorsal metacarpal arteries, which in turn originate from the dorsal carpal arterial arch. These branches mostly feed the plexus covering the skin and subcutaneous tissues of the dorsum of the proximal phalanx. The second blood supply is the proper digital arteries, which feed the plexus of vessels covering the skin and subcutaneous of the dorsum of the middle and distal phalanges.

When used as an innervated flap, the cross-finger flap derives its innervation from the sensory innervation of the dorsum of the hand (9,10). The superficial branch of the radial nerve and the dorsal sensory branch of the ulnar nerve innervate the dorsum of the hand up to the level of the proximal one-third of the dorsal skin of the proximal phalanx. Distal to this point, most of the innervation comes from the proper digital nerves through dorsal and distal branches (Figs. 20-1 and 20-2).

Reverse Cross-Finger Flap

The flap used for reverse cross-finger flap consists of the same tissues used for the standard cross-finger flap but without the covering skin, and thus has also been called a deepithelialized cross-finger

flap. The subdermal plexus becomes the main blood supply to this flap, fed by branches of the dorsal digital arteries and the proper digital arteries as previously described.

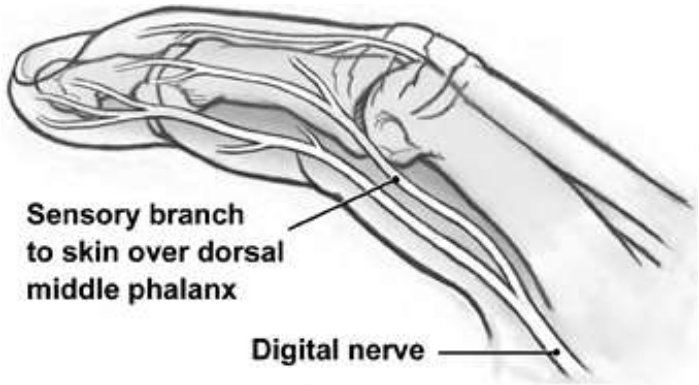

FIGURE 20-1 Innervation to the dorsum of the middle and distal phalanges from the proper digital nerves through dorsal and distal branches. (Reproduced with permission of the Mayo Foundation, 2005.)

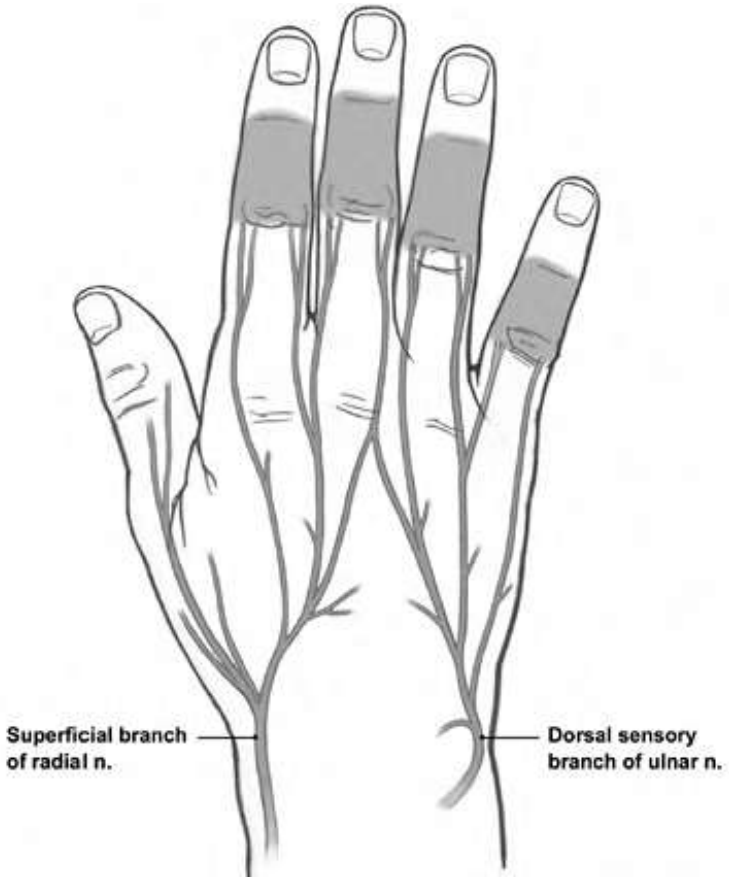

FIGURE 20-2 Innervation to the dorsum of the hand and dorsum of the proximal fingers from branches of the radial and ulnar nerves. (Reproduced with permission of the Mayo Foundation, 2005.)

Indications/Contraindications

Cross-Finger Flap

A cross-finger flap can be used any age group. In the pediatric patient, its use is reported to be successful as early as the first year of life (11). Some authors have recommended alternative tissue coverage in patients older than 50 years because of the risk of development of persistent postoperative stiffness at the proximal interphalangeal (IP) joint (12,13).

The main indication for this flap coverage is a volar fingertip wound with a major loss

of skin and subcutaneous tissue with exposed bones and/or tendons. It can also be used for volar defects of the middle phalanx.

As a random flap, the base is longer than the limbs of the flap. Thus, the entire dorsal skin of the finger can be raised and used. However, if the flap includes skin from the dorsal creases of the IP joints, the resultant dorsal scar may lead to stiffness and possible flexion contracture. It is recommended that the flap be raised between the IP creases. The flap's size therefore varies from patient to patient and depends on the size of the finger.

Contraindications include (14,15,16):

- Multiple injuries to the hand involving the potential donor finger
- Vasospastic conditions such as Raynaud disease and Berger disease
- Preexisting disabling problem such as Dupuytren contracture

Advanced age, diabetes mellitus, and rheumatoid arthritis are considered relative contraindications.

Reverse Cross-Finger Flap

A reverse cross-finger flap is indicated in adults with defects on the dorsum of the finger with major loss of skin and subcutaneous tissue and exposed bones and tendons or a nail bed defect that cannot be covered by other techniques (17).

The best donor site is the dorsum of the middle and proximal phalanges. The skin over the distal interphalangeal (DIP) and proximal interphalangeal (PIP) joints should be avoided because the subcutaneous tissue over these areas is thin and there is higher risk of developing contractures.

The size and contraindications of this flap are similar to those of the cross-finger flap.

Surgery

Cross-Finger Flap

The donor finger is adjacent to the injured finger. Typically, the donor finger is chosen

based on how the injured finger approximates to the middle phalanx of the donor finger. A majority of these flaps can be performed as outpatient surgery under regional anesthesia. Preoperative radiographs of the injured hand should be evaluated for concomitant injuries, which should be treated prior to commencement of the cross-finger flap procedure. Prophylactic antibiotic is given per the discretion of the surgeon.

Under brachial or digital tourniquet control, the injured digit is debrided and irrigated and the skin edges freshened (Fig. 20-3).

The donor and recipient fingers are placed in a position of comfort, before planning the location and the shape of the donor area. To ensure adequate soft tissue coverage, the size of the defect can be outlined on a piece of glove paper allowing 20% excess for the flap pedicle (Fig. 20-4). The outlined site is drawn on the dorsum of the adjacent donor digit. The three sides of the flap are raised as a full thickness graft on the donor finger down to but not including the paratenon (Fig. 20-5A). It is essential that the paratenon layer not be violated, as the full thickness skin graft, which is placed over the donor area, may not adhere. Meticulous hemostasis is obtained. If mobility of the raised flap is limited, or if there is kinking of the flap when the recipient site is covered, further dissection of the flap is performed by incising carefully Cleland's ligament, while protecting the neurovascular bundle that lies volar. Once elevated, the tourniquet is deflated and the vascularity of the flap and the paratenon is evaluated. The donor defect is covered with a full thickness skin graft obtained form

the anticubital fossa, upper brachium, or groin (Fig. 20-5B). The flap is sutured in place on the finger defect with nonabsorbable 5-0 suture in adults or absorbable suture in children (Figs. 20-6 through 20-9). A compressive dressing of the surgeon's choice is applied over the skin graft, followed by a bulky hand dressing with appropriate splints. The wound is kept elevated, with encouragement of movement of the uninjured digits. The dressing can be lightened at 1 week postoperatively, whereas the sutures are removed between 10 and 14 days. Between 2 and 3 weeks postoperatively, the patient is returned to the operating room, where the flap is divided, contoured, and inset (Figs. 20-10 through 20-12).

FIGURE 20-3 A-C: Front and side views of an oblique volar fingertip amputation.

Reverse Cross-Finger Flap

The reverse cross-finger flap is useful for dorsal tissue loss, unlike the standard cross-finger flap, which is indicated for volar wounds. The preparations for this flap are similar to those of the standard cross-finger flap. The differences include the hinge of the flap and the donor tissue. A thin full thickness skin flap is raised from the dorsum of the donor finger at the level of the middle phalanx, preserving the subcutaneous layer.

The hinge of the skin flap is opposite to the primary defect. This is in contradistinction to the standard cross-finger flap, where the hinge is adjacent to the site of the defect (Figs. 20-13 and 20-14). Attention should be paid to keep the dissection at the level of the dermis, below the hair follicles, and above the layer of subcutaneous veins (Fig. 20-15). The subcutaneous tissue is raised, keeping the paratenon intact, with the hinge

of the subcutaneous tissue adjacent to the defect. The subcutaneous flap is inset into the defect and the skin is placed back over the donor site (Fig. 20-16). A thin full thickness skin graft is harvested and set on the flap (Fig. 20-17). Division of the flap occurs between 2 and 3 weeks postoperatively (Fig. 20-18). A clinical example of a reverse cross-finger flap is illustrated in Figure 20-19.

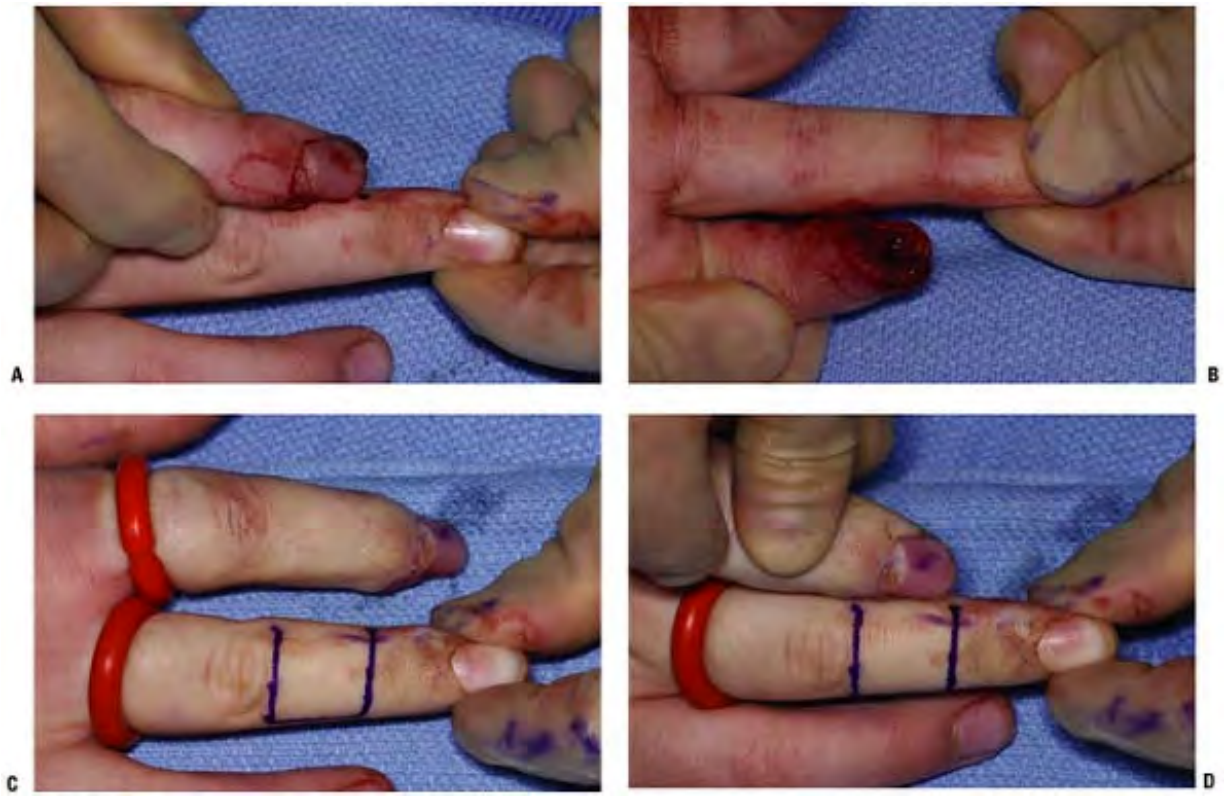

FIGURE 20-4 A-D: Intraoperative pictures showing the steps of preparation for cross-finger flap coverage of a ring fingertip defect. The injured finger is placed at the level of the planned flap from the dorsum of the middle finger at the level of the middle phalanx. The flap should be 20% larger than the actual defect to minimize tension at the time of closure.

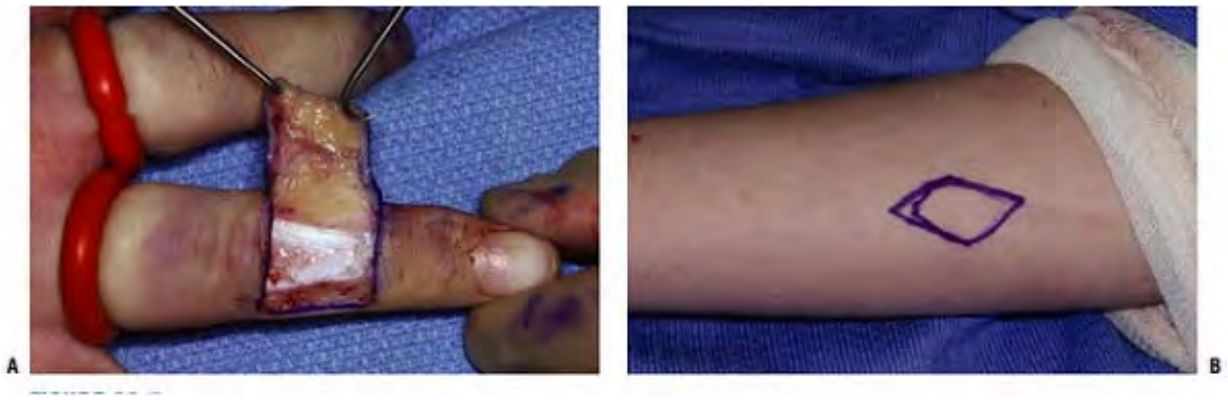

FIGURE 20-5 A: Elevation of the cross-finger flap, keeping the paratenon intact. B: Drawing the site for the full-thickness skin graft harvesting from the forearm of the involved finger. (Reproduced with permission of the Mayo Foundation, 2005.)

FIGURE 20-6 Illustration demonstrating where the cross-finger flap from the

dorsum of the middle finger is to be inset onto the amputated index tip. The full thickness graft is placed to cover the defect over the dorsum of the donor area. (Reproduced with permission of the Mayo Foundation, 2005.)

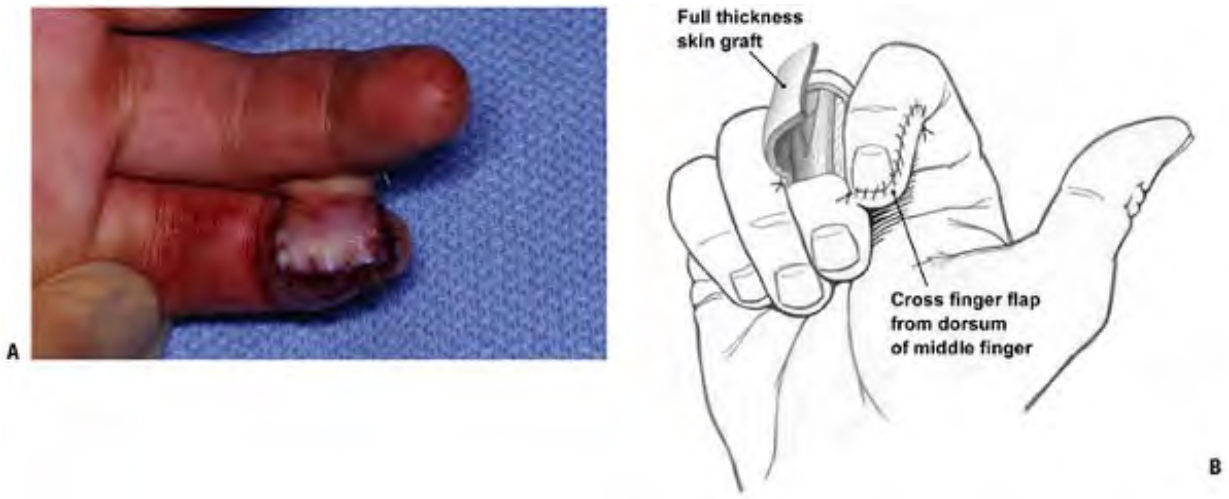

FIGURE 20-7 A: A clinical example of a cross-finger flap taken from the dorsum of the middle finger to cover the ring finger. B: Illustration demonstrating insetting of the flap. (Reproduced with permission of the Mayo Foundation, 2005.)

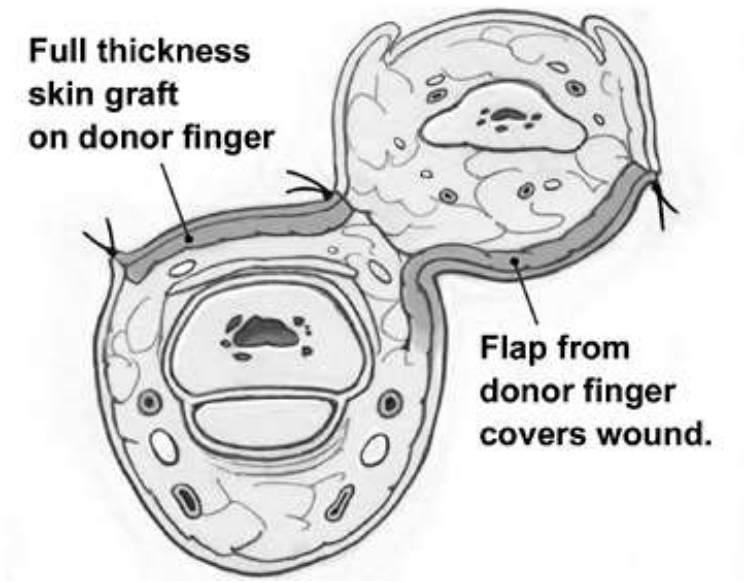

FIGURE 20-8 Illustration demonstrating cross section through donor and recipient sites. (Reproduced with permission of the Mayo Foundation, 2005.)

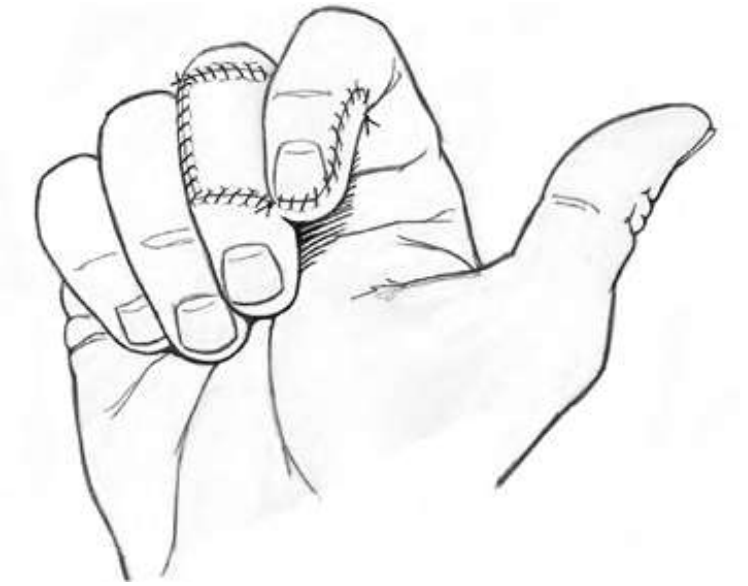

FIGURE 20-9 Illustration of the position of the finger after insetting of flap and

skin grafting of donor defect. (Reproduced with permission of the Mayo Foundation, 2005.)

FIGURE 20-10 A full thickness graft from the forearm is used to cover the defect over the donor site.

FIGURE 20-11 A,B: Division of the flap at 3 weeks postoperatively, with contouring and insetting of the flap.

FIGURE 20-12 A,B: The appearance of the finger 3 months after surgery.

FIGURE 20-13 A defect on the dorsum of middle finger at the level of the middle phalanx with exposure of the extensor tendon denuded of its paratenon is an ideal indication for a reverse cross-finger flap. (Reproduced with permission of the Mayo Foundation, 2005.)

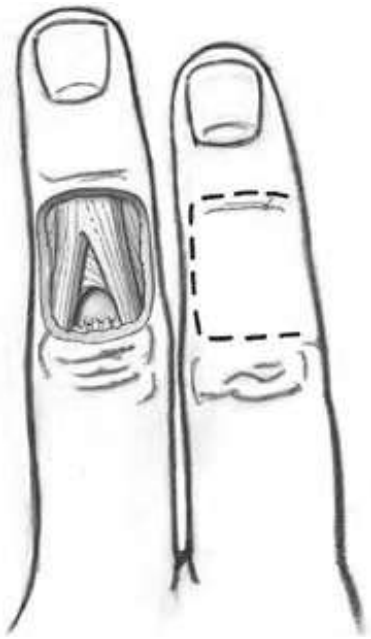

FIGURE 20-14 After the wound is debrided to fresh edges, the reverse cross-finger flap can be designed. (Reproduced with permission of the Mayo Foundation, 2005.)

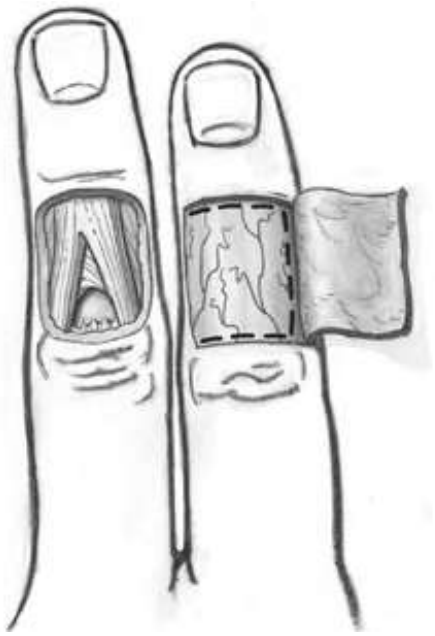

FIGURE 20-15 A thin full thickness skin flap is raised from the dorsum of the donor finger at the level of the middle phalanx, keeping the subcutaneous fat flap intact. The *31* indicate the site of the planned reverse cross-finger flap, which is divided to the level of the paratenon. (Reproduced with permission of the Mayo Foundation, 2005.)

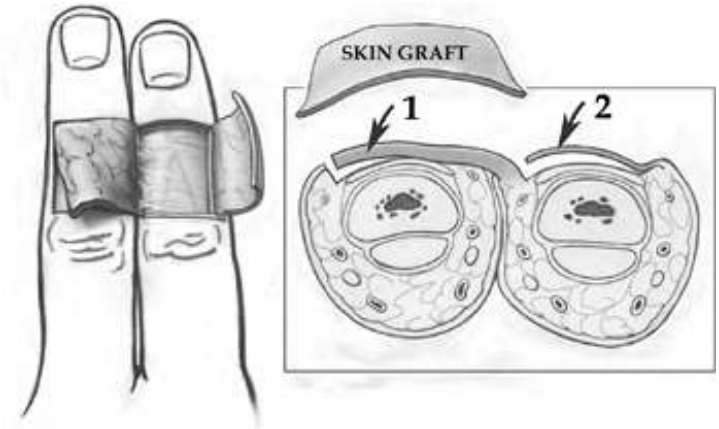

FIGURE 20-16 The hinge of the skin flap is opposite to the primary defect. The drawing shows how the reversed fat flap covers the dorsal finger skin defect (*31*). The raised thin thickness skin graft is re-placed again into its native site (*31*). Then, a thin thickness skin graft is placed over the reversed cross-finger flap. (Reproduced with permission of the Mayo Foundation, 2005.)

FIGURE 20-17 A drawing showing the appearance of the fingers after the after insetting of the reverse flap and placement of the skin graft. (Reproduced with permission of the Mayo Foundation 2005).

FIGURE 20-18 Division of the reverse cross-finger flap 2â€“3 weeks after the procedure. (Reproduced with permission of the Mayo Foundation 2005).

FIGURE 20-19 A-G: Clinical example of a reverse cross-finger flap, illustrating flap harvest and insetting. *31*

Postoperative Management and Results

After flap division, supervised finger range of motion commences, which includes combined active motion and passive stretching to eliminate stiffness and regain lost motion (see Fig. 20-9).

Postoperative infection, wound problems, and loss of flap are unusual (15).

The appearance of the skin graft and defect over the donor site are minimized or eliminated with the use of full thickness skin graft. Color mismatch between the dorsum and the pulp of the fingers is minimal, except in individuals with dark-colored skin.

The most common complications encountered with the reverse cross-finger flap include cold sensitivity and decreased sensation in the new transferred skin. In one report reviewing several follow-up studies of finger pulp loss coverage, 30% to 50% of patients developed cold sensitivity and 30% developed altered sensation regardless of the technique used (18). In another study of 54 patients followed over a 5-year period, 92% of the patients were satisfied with the result. However, 53% suffered cold sensitivity. All fingers developed protective sensation, but none had normal sensation or normal sweating (19). The altered sensory recovery varies by age. More than 90% of patients younger than 12 years had 6 mm or less two-point discrimination, compared to 40% of patients older than 40 years (20).

Atasoy (17) reported on four patients who had this reverse cross-finger flap done for nail bed defects (2 patients) with dorsal digital skin avulsions and/or extensor tendon injuries (2 patients). All patients had satisfactory results at the time of final follow-up. The patients with nail bed injury had satisfactory nail growth, and all patients had satisfactory extensor tendon function and good coverage of the defect. A potential complication of the reverse cross-finger flap is creation of an epithelial cyst. Atasoy recommended not grafting the reverse flap in case of large nail bed defect with an

intact germinal matrix, in order not to preclude the nail from growing. However, if the there is a complete avulsion of the germinal matrix, nail bed, and skin, with exposure of bone and tendon and little possibility for the nail to return, the reversed flap should be skin grafted.

In another study, Al-Qattan compared the results of deepithelialized cross-finger flap reconstruction (31 patients) versus adipofascial turnover flaps (42 patients) for reconstruction of small complex

dorsal digital defects (21). All flaps survived with no infection or hematoma. Patients who underwent reverse cross-finger flap had the following complications: flap dehiscence in one patient, stiffness of the donor finger in five patients, inclusion cyst in one patient, and significant skin graft loss in two patients.

Six patients who underwent adipofascial turnover flaps developed epidermolysis of the donor site, which resolved spontaneously. Al-Qattan recommended using the latter flap in children to avoid using of general anesthesia during flap division, and in older patients to avoid stiffness, and in multiple defects of adjacent border digits.

Conclusions

The cross-finger and reverse cross-finger flaps are reliable and dependable methods to cover volar and dorsal fingertip soft tissue defects. If carefully performed in properly selected patients, they are versatile flaps that provide a superior means of reconstruction for the injured finger with loss of significant tissue. The cross-finger flap has been used in a wide variety of patients of all ages with rewarding overall patient satisfaction.

References

1. Atasoy E, Iokimidis E, Kasdan ML, et al. Reconstruction of the amputated finger tip with a triangular volar flap. *J Bone Joint Surg.* 1970;52-A:921â€“926.

2. Fisher RH. The Kutler method of repair of finger tip amputation. *J Bone Joint Surg.* 1967;49-A: 317â€“321.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 21

Heterodigital Arterialized Flap

Shian Chao Tay

Lam Chuan Teoh

Indications/Contraindications

Hand trauma that results in exposed bone, tendon, joint, or neurovascular structures requires some type of flap coverage. Defects located around joints and web spaces may also require flap coverage to avoid contractures that could occur following split thickness grafting. The location and size of these defects can often preclude the use of small rotation or advancement flaps. Free flap coverage is an option, but the presence of surrounding infection or a wide zone of trauma may make the recipient vesselsâ€™ dissection difficult. In such situations, a heterodigital arterialized (HTA) flap may be the best option.

The HTA flap is raised from the lateral side of a donor finger together with the digital artery and a dorsal digital vein. Unlike Littler's neurovascular island flap, or Hueston's extended neurovascular island flap, the heterodigital arterialized flap's main function is to provide non-sensory reconstruction of skin defects in the hand or fingers. Thus, this flap is never harvested with the finger pulp or the digital nerve of the donor finger. This is an important feature that serves to reduce morbidity to the donor finger. The inclusion of the digital dorsal vein into the flap improves venous drainage of the flap and reduces the incidence of venous congestion, which is a well-documented complication of the classic Littler flap.

The heterodigital arterialized flap is a thin flap. In this aspect it is ideal for reconstruction in the hand and fingers as it provides near like-to-like reconstruction. The other advantages of this flap is that being regional, it obviates the need to prepare a separate surgical site and allows for almost immediate motion of the reconstructed finger. The reconstruction can be performed as a single-stage procedure (as compared to the cross-finger flap or groin flap).

In our experience, the heterodigital arterialized flap has been used to reconstruct volar, lateral, or dorsal defects in the fingers proximal to the distal interphalangeal joint, web spaces, palm, dorsum of the hand, and the thumb. The average flap dimensions are 4.1 cm (range, 1.5 to 5.5 cm) in the longitudinal axis and 2.1 cm (range, 1.0 to 3.5 cm) in the transverse axis. The HTA flap is thus a useful option to consider when faced with the problem of a relatively large defect in the hand or finger that requires non-sensory flap reconstruction. The HTA flap has been used for reconstruction of defects following infection, trauma, post-replantation of digits, electrical burns, chemical burns, and high-pressure injection injuries.

In terms of flap mobility, the reach of the heterodigital arterialized flap is limited to coverage of defects in the adjacent fingers or thumb or adjacent parts of the hand. This is principally due to the combined limitations in the reach of the dorsal vein and the digital artery of the flap. In situations where a greater reach is required, it is possible to divide the dorsal vein of the flap over the dorsum of the hand and reanastomose it in a region of healthy tissue after flap transfer. With this maneuver, the flap should have the ability to reach defects two to three fingers away, much like the traditional reach of Littler's neurovascular island flap, when pivoting at the common digital artery's offshoot at the superficial palmer arterial arch.

The HTA flap is absolutely contraindicated if there is only one functioning digital artery in the donor finger and when the vascular viability of any other adjacent fingers is threatened by flap harvesting.

It is also relatively contraindicated if the functional prognosis of the finger to be reconstructed is judged to be so poor that it might be better to amputate it. However, we recognize that social or cultural practices may still dictate reconstruction for the purpose of cosmesis or the maintenance of the â€œwholeâ€•

self. Conversely, if there is a rare situation wherein a need to perform a digital amputation exists in a hand containing a defect requiring flap reconstruction, the HTA flap can be used as a fillet flap from the amputated finger for defect reconstruction. In such a situation, the HTA flap is very useful in spare parts surgery as there is no concern for donor finger morbidity. If the defect occurs beyond the reach of the flap or in the contralateral hand, a free fillet HTA flap can be created from the finger to be amputated provided the conditions at the recipient finger allows for a free flap.

Preoperative Planning

All practical options available for reconstruction should be considered by the surgeon before deciding on the HTA flap. Risks of heterodigital donor finger morbidity should be carefully weighed against benefits of reconstruction. Often, the location, size, and nature of the defect will preclude other reconstructive options, leaving the HTA flap as the most viable option with the highest flap survival rate.

The choice of donor finger is often dictated by the location of the defect and is often an adjacent finger. The outside borders of the index and little fingers are not used as donor sites to preserve hand cosmesis and to maintain native skin for protection and sensation. This requirement is not absolute; however, we prefer the use of the middle finger, ring finger, and the ulnar border of the index finger. Theoretically, the radial border of the little finger can be used as a donor site but this is rare as the little finger is small, providing little skin, and the remaining ulnar digital artery is often absent or vestigial.

A digital Allen's test must be performed on the intended donor finger to ensure that both digital arteries are sufficiently patent. In addition, the Allen's test should also be performed on the finger that is adjacent to the donor site as this would dictate if the digital artery mobilization can proceed proximal to the point of bifurcation of the common digital artery. More often than not, a longer reach is required and ligation and division of the neighboring branch of the digital artery to the adjacent finger would need to be performed to create an arterial pedicle of sufficient length.

If the donor site is just next to the defect in the adjacent finger, also known as the

contiguous or near side of the adjacent finger (e.g., defect on ulnar side of index finger with donor site on radial side of middle finger), the pivot point can be at the bifurcation of the common digital artery, as a short pedicle would be sufficient. This pivot point can be approximately landmarked in the intermetacarpal space at the level of the distal palmer crease. The distance from this pivot point to the proximal edge of the defect can then be measured to determine the reach needed by the flap and hence the location of the proximal margin of the flap. If the flap is not coming from the contiguous side of the adjacent finger, it is likely that a more proximal pivot point would be required to ensure sufficient reach of the flap. The most proximal point would be the superficial palmer arterial arch, which can be approximated to the level of the proximal palmar crease in the relevant intermetacarpal space. In such situations, as mentioned previously, it will be necessary to ligate and divide the neighboring digital artery of the common digital artery that supplies the adjacent finger to maximize the reach of the flap.

Experience has shown that the reach of the flap increases by 10% once the digital artery is adequately mobilized. However, in the interests of safety and ensuring minimum tension in the pedicle, a 1:1 ratio should still be maintained for the length of the arterial pedicle during preoperative planning of flap surgery.

The timing for reconstruction depends on the pathology. As far as possible, early reconstruction (within 1 week) is preferred as it minimizes overall hand stiffness from prolonged disuse. For traumatic conditions, the reconstruction is ideally performed within 1 week. For post-replantation surgery with residual skin defects, the resurfacing is performed within 10 to 15 days when re-endothelialization of the vascular anastomoses is completed. Infection cases are ideally resurfaced within 4 to 7 days when the surrounding cellulitis and edema have subsided. In order to achieve this time line in infective cases, infection reversal should be achieved within one to two formal definitive excisional debridement surgeries in conjunction with appropriate intravenous antibiotics.

The surgery can be performed under general or regional anesthesia with sedation. The patient's general condition and coagulation profile should be optimized as per any reconstructive surgery. The

patient should be consented and counseled as to the surgery, the postoperative

course, and the duration of stay following surgery.

Surgery

Patient Positioning

The patient should be in the supine position, and care should be taken to protect bony prominences with suitable pressure-relieving bolsters. The upper limb with the hand requiring reconstruction should be abducted to not more than 80 degrees at the shoulder to prevent excessive stretch to the nerves in the axilla. The elbow should be flexed slightly so that the forearm can be placed transversely on a stable hand table. A pneumatic tourniquet should be applied to the most proximal part of the upper arm, and the upper limb should be cleaned and draped as far proximal until the distal edge of the tourniquet. If the surgery is expected to last more than 2 hours, urinary catheterization should be considered.

Technique

The surgery is performed under tourniquet control, which should be inflated with the upper limb elevated but not exsanguinated. This is to facilitate identification and mobilization of the digital artery and the digital dorsal vein. The defect should be thoroughly debrided and irrigated. A template of the defect is made. At this point, the digital Allen's test can be repeated to confirm the adequacy of both digital arteries in the donor finger, and also in the adjacent finger if the flap is not taken from the contiguous side of the finger adjacent to the defect.

Using the template created, the appropriate dimension of the flap is then transferred onto the donor site. The flap should be centered on the lateral or dorsolateral side of the finger, depending on the width of flap required, to ensure the inclusion of the digital artery and a dominant digital dorsal vein. The maximum width of the flap should not exceed the mid-palmer line and the mid-dorsal line (a width of approximately 3 cm). The maximum length of the flap is from the base of the finger to the distal interphalangeal joint crease (usually 4 to 5 cm depending on the size of the finger). If necessary, the flap length may be extended 0.5 cm distal to the distal interphalangeal joint crease. Next, a palmar Z incision is drawn, biased to the respective side of the finger to access the proximal part of the digital artery. This incision is carried into the palm to the level of the common digital

artery. The location of the dominant dorsal digital vein is then verified and marked to ensure that it will drain the flap. Once this is completed, the flap is ready to be harvested.

Heterodigital arterialized flap harvesting is performed with the aid of loupe magnification. The flap is first raised from its palmar margin (Fig. 21-1). If the palmar margin of the flap crosses any digital palmer flexion creases, a mini-z cut can be made at these creases, to break up the resultant scar line. Superficial palmer veins can be carefully cauterized with bipolar electrocautery and divided. As one proceeds deeper, the dissection follows the digital artery from proximal to distal starting from the bifurcation of the common digital artery. The angle of dissection toward the neurovascular bundle is an oblique plane skiving dorsal to the digital nerve, which is more central and superficial compared to the digital artery. As far as possible, the nerve should be left undisturbed in its bed with a collar of fat kept around it. Besides ensuring that the nerve is not traumatized, the collar of fat also provides additional cushioning and enhances the take of the full-thickness skin graft at the donor site. Ideally, a collar of fat or soft tissue should also be kept around the digital artery to preserve the venae comitantes. Fortunately, in the HTA flap, this is not as critical, as the dorsal vein would provide the main conduit for venous drainage.

At this point, an intraoperative occlusion test can be performed using non-traumatic microsurgical vascular clamps to determine the integrity of the contralateral digital artery in the donor finger. If mobilization of the digital artery to the level of the superficial palmer arch is necessary, a similar occlusion test should be performed on the bifurcated digital artery supplying the adjacent non-donor finger to ensure that it also has a healthy contralateral digital artery.

As mobilization of the digital artery progresses, the transverse palmer arches of the digital artery will be encountered. These are rather large and short. Sufficient room is developed and the arches can be carefully cauterized with bipolar electrocautery before division. Care should be taken to ensure that the bipolar electrocautery is applied at least 2 mm from the parent artery to prevent compromising flow in the main digital artery.

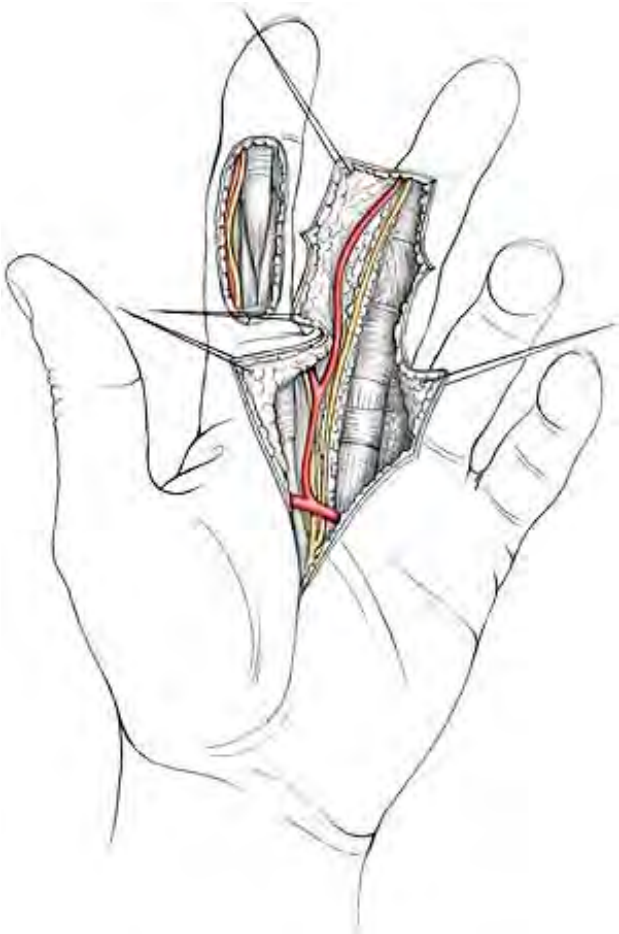

FIGURE 21-1 Completed palmar dissection of the HTA flap with the digital nerve left undisturbed in its bed with a collar of fat around it. Proximal dissection of the arterial pedicle is facilitated by a palmar Z incision into the palm.

Dorsal dissection is performed next with proximal to distal dissection and mobilization of the dominant dorsal digital vein to ensure inclusion of the vein in the flap. The flap is dissected free from the extensor tendons in the plane superficial to the paratenon. Care should be taken not to damage the dorsal skin branches of the digital artery supplying the flap. At the distal margin of the flap, the distal end of the dominant dorsal vein is ligated and divided, and the same is performed for the digital artery on the palmer side.

At this stage, the flap should be detached from distal to proximal. The adherent fibrous septa and Cleland's ligaments are divided. At the base of the digit, some dissection may be necessary to complete the division of subcutaneous tissues attached to the flap. The natatory ligament at the web space may also be divided to prevent kinking of the dorsal vein and lengthen the reach of the flap.

A generous subcutaneous tunnel is created between the donor and the recipient site. Tunneling is performed from the proximal edge of the recipient site into the palm to reach the pivot point of the flap. A hemostat forceps is used to gently enlarge the tunnel for easy flap delivery. The distal end of the flap is carefully grasped and gently maneuvered into the tunnel. A recommended method of flap delivery is to completely wrap the flap with tulle gras dressing and deliver the flap by grasping only

on the tulle gras. This serves to prevent direct grasping of the flap and also protects the flap and pedicles from being avulsed if the delivery in the tunnel is inadvertently rough or jerky. As the flap traverses the tunnel, care is taken to ensure that the digital artery and the dorsal vein do not become twisted or kinked. The flap should not be rotated more than a half turn. If there is any uncertainty, the flap should be re-delivered to ensure its correct orientation.

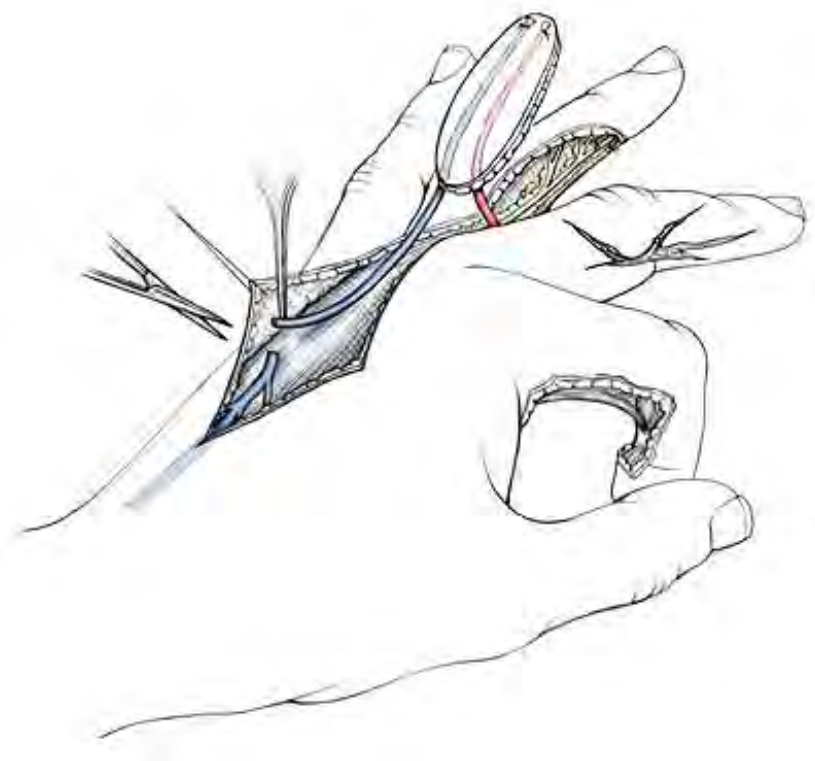

FIGURE 21-2 Completed dorsal and palmar dissection of the HTA flap with the flap completely islanded from the donor ring finger. The flap is now to be transferred. In this case, the defect requiring resurfacing is on the volar side of the index finger. The middle finger has a significant injury and is not suitable as a donor. The bi-pedicled nature of the flap will obstruct the transfer through the palmar subcutaneous tunnel. Thus, the venous pedicle of the flap has been divided over the dorsum of the hand to partially free the flap and convert it into a mono-pedicled flap.

When a longer reach of the flap is required, and the bipedicled nature of the flap is restricting the transfer, the dominant dorsal vein may be divided before the transfer (Fig. 21-2) and anastomosed to the original vein or another vein over the dorsum of the hand in an area free of adversity, after the flap has been transferred (Fig. 21-3). When this procedure is performed, it is called the â€œpartially free HTA flap.â€•

After flap transfer, the flap is provisionally held in place with a few fine sutures.

The tourniquet is then deflated to check for flap perfusion and to secure hemostasis. If flap perfusion does not return after 5 minutes or if there is flap congestion, some sutures may have to be removed. If this fails to improve flap color, kinks in the vascular pedicle may have to be located and relieved. If the vascular pedicle suffered mild inadvertent trauma during surgery, appropriate flap resuscitation measures such as warming the patient, warm saline soaks to the artery, and immersion of the arterial pedicle in 2% lidocaine or papavrine may have to be performed before adequate flap perfusion is restored.

Once adequate flap perfusion is restored, the flap is then fully inset with fine sutures, usually nonabsorbable monofilament 5-0 sutures, spaced not less than 3 to 5 mm apart. Care should be taken during the insetting to ensure that both the dorsal vein and the digital artery are not compressed or traumatized. The donor defect is resurfaced with a full-thickness skin graft. This can be taken from the medial proximal forearm, the anterior elbow, or groin if necessary. As early postoperative active rehabilitation is required, it is highly recommended to secure the skin graft with a cotton wool bolster and a meticulous tie-over dressing to prevent skin graft loss during therapy. The rest of the surgical wounds are closed with fine sutures with or without non-suction drains. A bulky, non-constrictive dressing is applied with a window left open for flap monitoring. Healthy flap perfusion should be confirmed before leaving the operating room.

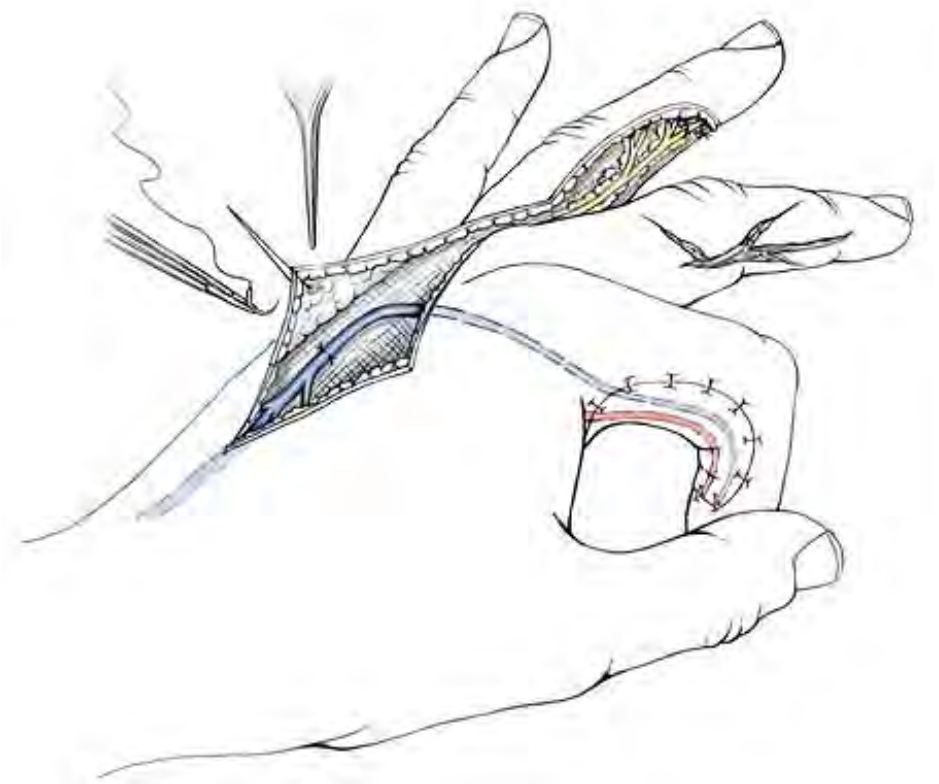

FIGURE 21-3 The partially free HTA flap after successful transfer to the defect site on the index finger. The venous pedicle of the flap has been tunnelled back for microsurgical repair to the original draining vein over the dorsum of the hand.

Results

A total of 40 HTA flaps were performed between 1991 and 2001. Thirteen of these flaps involved division and repair of the dominant dorsal digital vein to extend the reach of the flap. Total active range of motion of the donor finger was excellent in 82.5%, good in 15%, and fair in 2.5% (1 patient) according to Strickland and Glogovac's criteria for flexor tendon surgery (see Figs. 21-4, 21-5, and 21-12). The fair result was due to the development of reflex sympathetic dystrophy and associated flexion contracture. Donor finger two-point discrimination was 3 to 5 mm, except in one case (6 mm). None of the donor fingers had hypersensitivity, symptomatic neuromas, or suffered from cold intolerance.

Flap survival was 100% with no cases of flap congestion or ischaemia documented. All flaps healed primarily and provided supple coverage of the defects. Full-thickness skin graft take was similarly successful at 100% in the donor fingers. Although a contour concavity is present at the time of removal of the cotton wool bolus on the fifth postoperative day, the concavity usually fills out within

3 months after surgery. None of the cases required any further surgery to the flap such as insetting or defatting.

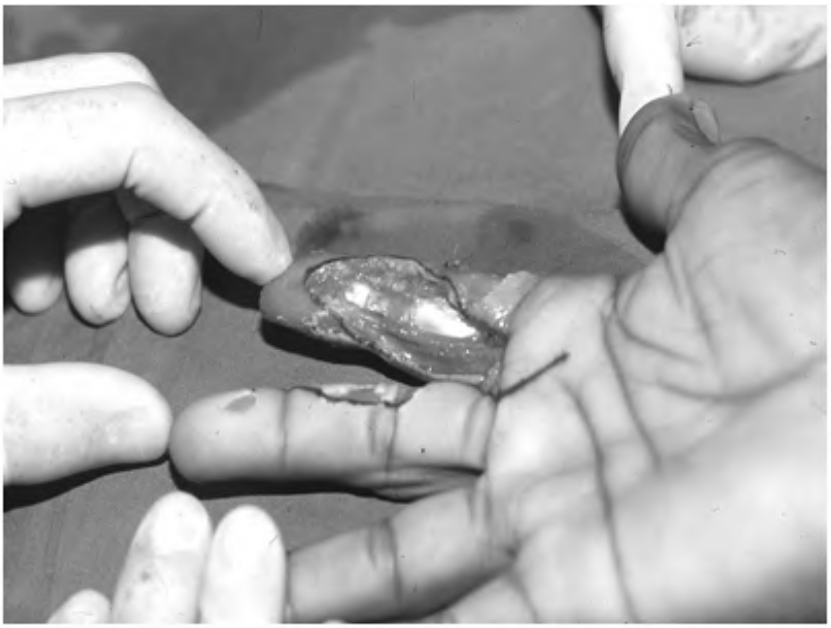

FIGURE 21-4 A sizeable defect (5 Ã— 2 cm) over the volar-ulnar aspect of the right index finger exposing bare flexor tendons and the neurovascular bundles following trauma. The HTA flap would come from the radial side of the middle finger. Thus, it would be a contiguous transfer of the HTA flap. There is no need to divide the venous pedicle to partially free the HTA flap in such a situation. (From Teoh LC, Tay SC, Yong FC, Tan SH, Khoo DBA. Heterodigital arterialized flaps for large finger wounds: results and indications. *31*2003;111(6):1905â€“1913. With permission.)

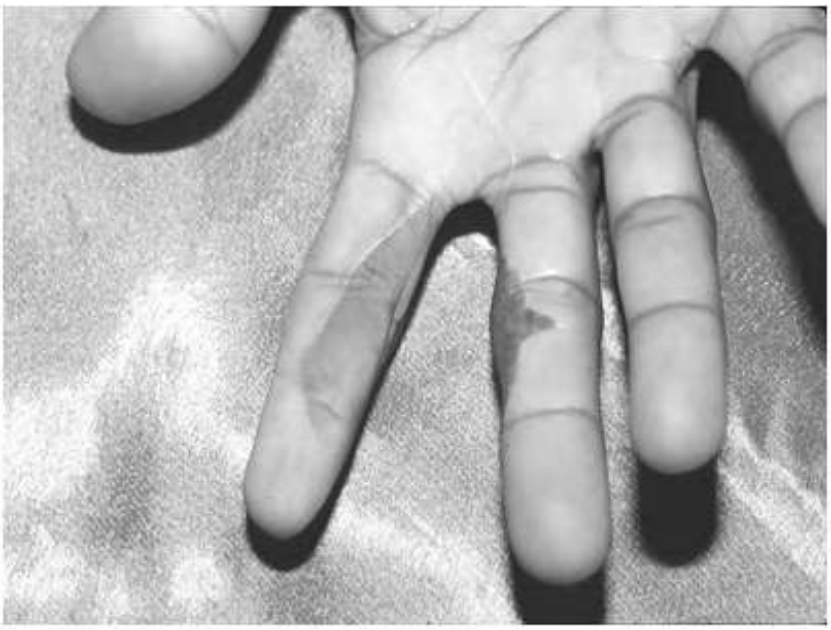

FIGURE 21-5 Cosmetic result in this dark-skinned individual at 3 months postoperatively. Patient has full extension of the donor middle finger and has excellent primary healing of the HTA flap on his index finger. (From Teoh LC, Tay SC, Yong FC, Tan SH, Khoo DBA. Heterodigital arterialized flaps for large finger wounds: results and indications. *31*2003;111(6):1905â€“1913. With permission.)

We evaluated the total active motion in 22 fingers which had undergone reconstruction with the HTA flap. According to Strickland and Glogovac's criteria, 45.4% achieved excellent and good results, with the remainder achieving fair to poor results. It should be noted that the poorer outcomes in the reconstructed fingers are due to severe insults resulting in multiple tissue involvement with concomitant damage to joints, tendons, and/or ligaments and not due to contractures of the HTA flap.

Postoperative Management

The first 24 hours are the most critical for the flap. The flap should be monitored frequently with the hand nursed in an elevated position during this time and the

patient kept nil per oral (NPO). The ambient temperature can be raised to a level appropriate for the season at least for the first night depending on the surgeon's preference. Drains, if used, should be removed within the first 24 to 48 hours. Rehabilitation should commence on the second postoperative day but can be delayed if circumstances dictate. An early and active rehabilitation program is crucial in minimizing heterodigital donor finger morbidity. Active and passive range of motion exercises are instituted for both the recipient and donor digits. Regional anesthesia in the form of infusion blocks can be maintained during the postoperative period to assist with pain control during rehabilitation. The use of a cotton bolster ensures that the full-thickness skin graft is firmly secured so as to allow progressive hand therapy within the first week. Interval splinting of the donor digit between exercises may be necessary to prevent flexion contractures.

Complications

Heterodigital donor finger morbidity is the main concern. As mentioned previously, the preservation of the digital nerve and the finger pulp on the donor finger is critical toward preserving sensory function in the donor finger. In addition, the meticulous dissection of the digital artery away from the digital nerve is crucial in minimizing inadvertent trauma to the digital nerve, which can affect sensory function in the finger pulp and produce sensitive mini-neuromas along the digital nerve. Together with careful resurfacing of the donor site using full-thickness skin grafting and institution of early active finger mobilization, stiffness of the donor and reconstructed finger can be significantly reduced.

Mild loss of total active motion can be expected in about 10% to 15% of cases. Reflex sympathetic dystrophy with flexion contracture is another serious complication that can occur. Flap failure, either partial or complete, is always a potential complication, but it has not occurred in our series. In the five cases of free HTA flap, we did not experience any vascular complications and all the free flaps survived completely. In dark-skinned races, the darker dorsal skin may be a poor cosmetic match for defects on the volar side of the hand (Fig. 21-5).

FIGURE 21-6 Excellent result and hand function with full finger flexion in both donor and reconstructed fingers. (From Teoh LC, Tay SC, Yong FC, Tan SH, Khoo DBA. Heterodigital arterialized flaps for large finger wounds: results and indications. *31*2003;111(6): 1905â€“1913.

Variants of the Heterodigital Arterialized Flap

Besides the free fillet flap, other variants of the HTA flap include the cross-finger HTA flap, which is particularly suited for coverage of dorsal defects distal to the proximal interphalangeal joint which cannot be reached by the reverse dorsal intermetacarpal flap or the Quaba flap. Another indication for the cross-finger HTA flap is in situations where proximal dissection of the digital artery in the hand is not advisable due to previous trauma or infection. The cross-finger HTA flap only requires dissection of the venous and arterial pedicle up to the level of the base of the finger and will allow transfer to any surface on the adjacent finger (Fig. 21-6). This is accessed via mid-lateral incisions raising small dorsal and volar soft tissue flaps proximal to the donor site. Care should be taken not to extend these incisions into the web space or the lateral walls of the web commissure. A similar mid-

lateral incision is performed proximal to the defect on the injured finger. The flap is directly transferred across the interdigital space and onto the defect (Fig. 21-7). This avoids the need to tunnel the pedicles and flap through the distal part of the hand and also allows the flap to reach a defect on any surface of the finger distal to the proximal interphalangeal joint. The dorsal and volar soft tissue flaps of the mid-lateral incisions are then sutured together to create a soft tissue pouch that will protect the flap pedicle and prevent separation of the digits. Division of the cross-finger

P.265

P.266

version of the HTA flap is performed at 3 weeks. In the interim, active rehabilitation of the joined fingers should be started to prevent finger stiffness.

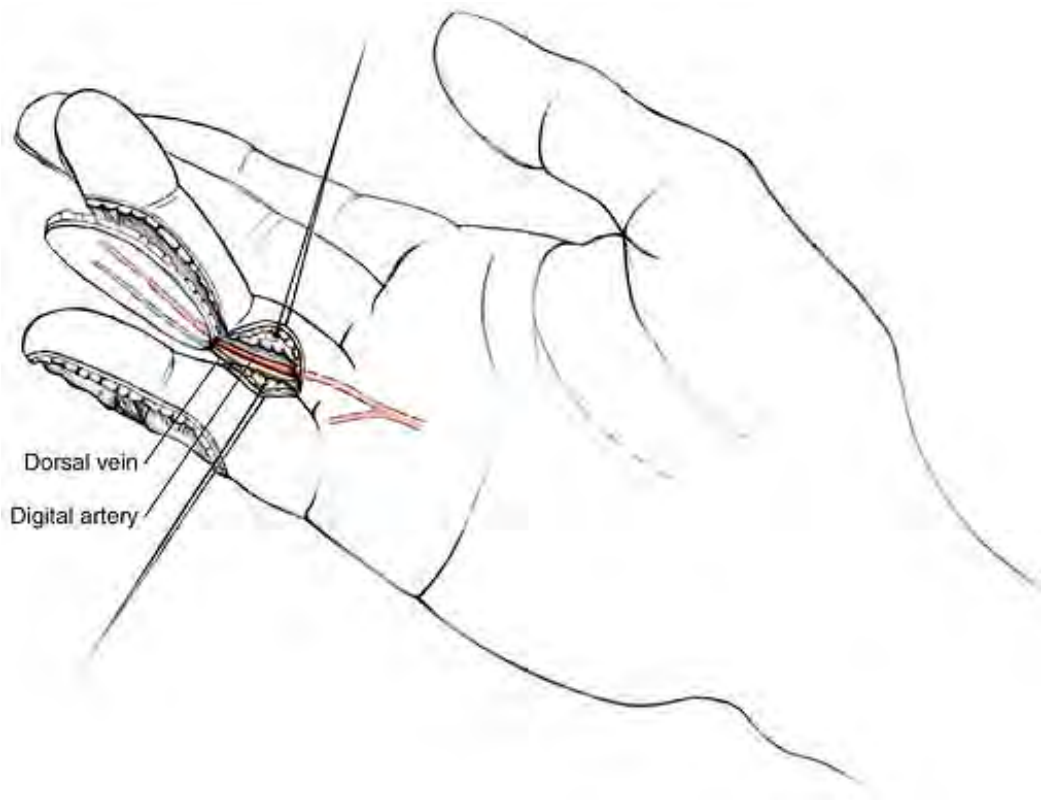

FIGURE 21-7 Cross-finger HTA flap from the ulnar side of the ring finger for transfer to a defect on the ulnar side of the distal little finger. Note that minimal pedicle length is required for this direct cross-finger transfer and

there is no need to perform any pedicle dissection in the palm.

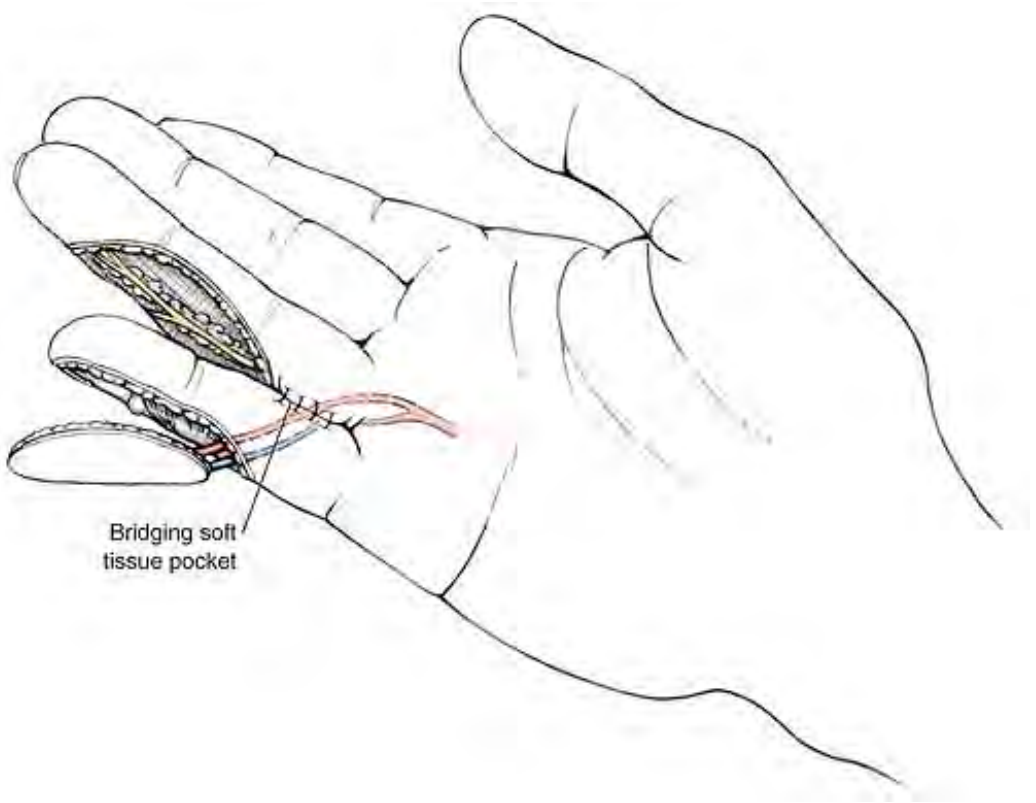

FIGURE 21-8 Cross-finger HTA flap after transfer to the defect site. The flap has been tunnelled through volar tissue of the little finger. The vascular pedicles of the flap are now protected by a soft tissue skin bridge that has been created between the base of the ring and little finger. A full-thickness skin graft with a cotton wool bolus with tie over will be applied to the donor site. The soft tissue skin bridge, together with the HTA flap vascular pedicles, will be divided at 3 weeks postoperatively when the cross-finger HTA flap has established its own vascular supply in the reconstructed little finger.

FIGURE 21-9 A patient referred from another hospital with a sizeable chronic non-healing defect on his left middle finger with concomitant injuries to all the other fingers.

FIGURE 21-10 After debridement, the true defect size is apparent (6.5 Ã— 2.2 cm) with exposed bare flexor tendons. The patient will be undergoing free HTA flap procedure with the flap being harvested from the ulnar side of the right middle finger.

FIGURE 21-11 Two weeks post free HTA flap with 100% survival of the free HTA flap with excellent primary healing in the reconstructed left middle finger. The full-thickness skin graft is also healing well on the donor site (ulnar side of the right middle finger).

The final variant is the free HTA flap. This free flap should only be used by hand microsurgeons already well experienced with the HTA flap. In occasional situations, a sizeable defect exists in a hand in which there is concomitant or prior injuries to the other fingers, rendering them unsuitable as donors of the HTA flap. In such situations, if the defect is amenable for a free flap procedure and other free flaps are not available, the HTA flap may be harvested as a completely free flap from the uninjured contralateral hand. In such situations, the free HTA flap is taken from the ulnar side of the middle finger. The indication for the free HTA flap is certainly unique and rare. First performed in 1991, we have since found cause to only have performed five such flaps. In one case, it was used to electively resurface a sizeable defect on the small finger following a successful four-finger replantation. In another case, it was used to resurface a non-healing chronic defect on the middle finger following an injury which also involved the other three fingers (Figs. 21-8 and 21-9). In two cases, there were acute concomitant injuries to the other fingers. In the last case, the free HTA flap was used as a free flow-through arterial

flap, not only to reconstruct the skin defect, but also to reconstruct a digital artery segment defect in a finger that had arterial insufficiency following an injury that also involved all the fingers. The excellent size matching of the digital arteries certainly contributed significantly to the success of the procedure. In all five cases, except for the mild cosmetic defect of the full thickness skin graft at the donor site, all the donor fingers retained full and complete range of active motion and preservation of normal pulp sensation (Figs. 21-10 and 21-11). An important reason why the free HTA flap might have such an excellent result in the donor fingers might be due to the fact that there were no other injuries in the donor hand, which greatly facilitated active rehabilitation.

FIGURE 21-12 Full flexion of the right middle donor finger demonstrated. The stiffness that is apparent in the left hand is due to prolonged immobilization and disuse as a result of the chronic non-healing wound. With successful and robust coverage of the defect in the left hand with excellent primary healing, aggressive rehabilitation of the hand can be started to alleviate the stiffness.

Recommended Reading

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 22

Homodigital Island Flap

Bradon J. Wilhelmi

Damon Cooney

Single-stage reconstruction of the finger pulp has become possible with a better understanding of hand anatomy and the advent of new surgical techniques. Single-stage reconstruction eliminates the risk of a stiff finger resulting from immobilization and the necessity of a second operation that is required with traditional regional flap reconstruction. The options for one-stage reconstruction of pulp defects include heterodigital and homodigital neurovascular island flaps. The heterodigital neurovascular island flap has the disadvantage of late or no development of sensory reorientation and paresthesia in the donor finger. Harvesting the flap from the injured finger, as is the case in a homodigital island flap, avoids unnecessary surgery to an adjacent uninjured finger as well as potentially enhancing rehabilitation by minimizing additional hand trauma. Many homodigital neurovascular flaps based on one or two volar digital pedicles have been described which can provide the fingertip with sensate coverage. The palmar advancement flap based on bilateral neurovascular pedicles has been used for reconstruction of the thumb and fingers. The interphalangeal (IP) joint is usually needed to be kept in flexion to obtain adequate advancement of the flap, which increases the complication risk for flexion contracture of the distal joint with these advancement flaps. Also, the use of the advancement flap increases the risk for dorsal skin necrosis if the dorsal branches from the proper digital artery are severed for distal advancement of the flap. The volar V-Y flap is still a suitable flap for small

fingertip defects less than 1 cm when there is sufficient skin distal to the distal interphalangeal (DIP) crease to base the flap.

Indications/Contraindications

The homodigital neurovascular flap has been described with elevation of the dorsal finger skin to repair pulp deficits which cannot be closed with the volar V-Y advancement flap. Anatomic studies have demonstrated that a dorsal arterial branch from the proper digital artery exists over the middle phalanx which runs obliquely, dorsally, and distally over the distal interphalangeal crease. The dorsal branch at the middle phalanx collateralizes with the dorsal blood supply of the finger from another dorsal branch from the proper digital artery at the level of the distal phalanx, forming a dorsal arterial arcade over the DIP joint. Also, a dorsal nerve branch can be found leaving the proper digital nerve to innervate the dorsal finger skin proximal to the nail fold. The dorsal homodigital neurovascular island flap can be based on one or both of these arteries and this dorsal nerve. When the flap is harvested on the distal-most branches, it can be called the distal dorsal homodigital artery (DDHDA) flap. This DDHDA flap can be used to close defects of 2 Ã— 2.5 cm just proximal to the tip and hyponychium (Fig. 22-1). Larger defects in this region can be reconstructed by a flap based on the dorsal branch at the middle phalanx, called the middle dorsal homodigital neurovascular island flap, or MDHDA flap. This MDHDA flap can be used to close defect 2 Ã— 4 cm in size (Fig. 22-2). Alternatively, when the loss is very distal and an antegrade homodigital neurovascular island flap will not reach the defect, a reverse homodigital island flap (RHDA flap) can be used. The RHDA flap is harvested from the lateral base of the finger and based on proximal dorsal branches of the proper digital artery (Fig. 22-3). Dorsal branches of the proper digital nerve can be used to innervate this skin paddle, coapting this nerve to the transected nerve at the tip. These homodigital flaps are contraindicated when there is significant injury involving the pedicle or the skin on which the flap is based over the dorsum of the finger. Moreover, these homodigital flaps should not be used when

a skin bridge cannot be preserved for venous drainage to the distal finger. If the harvest of the flap would result in the exposure of already injured extensor tendon, devoid of paratenon, then the flap should be avoided. Crush and electrical injuries to the dorsal finger are also relative contraindications for the homodigital flap.

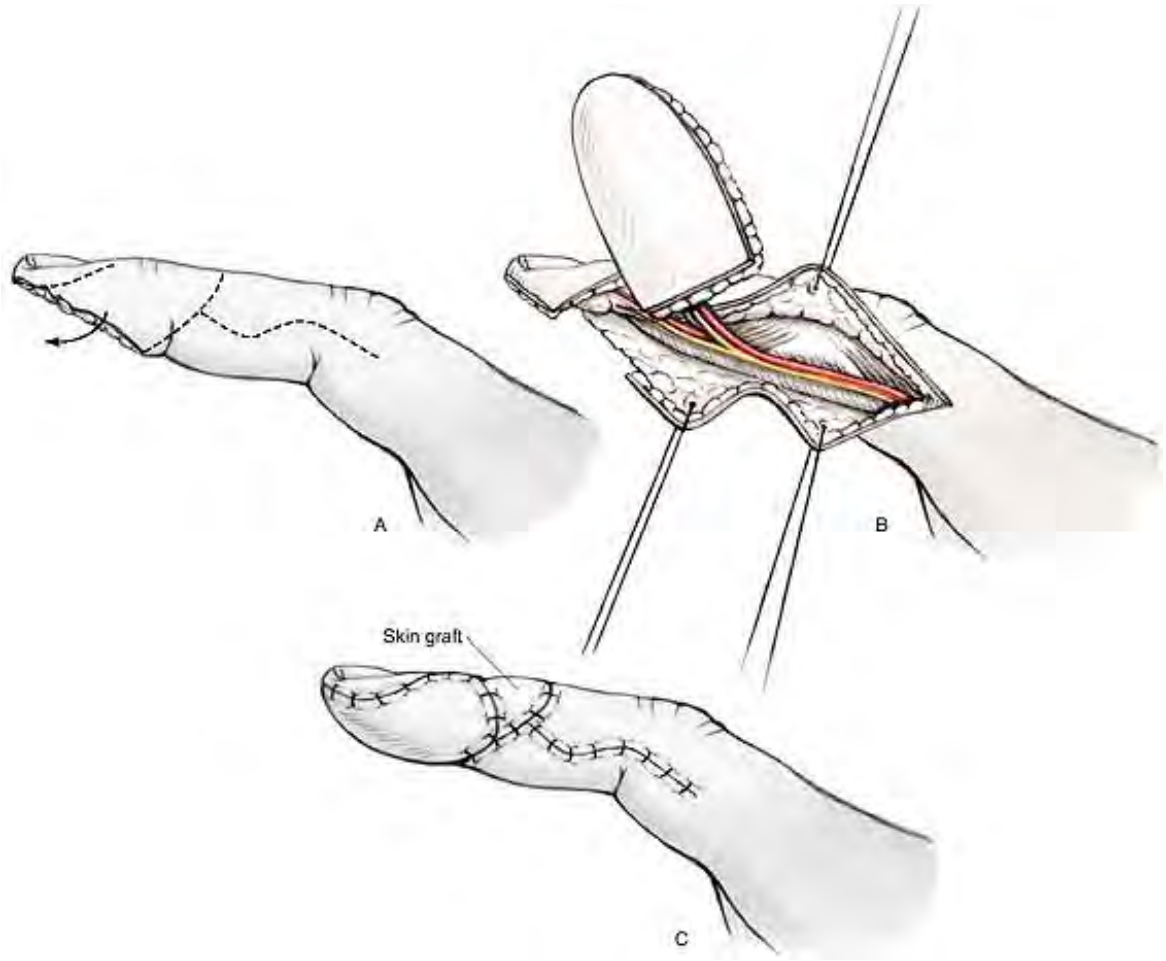

FIGURE 22-1 A: The design of the incision and distal dorsal homodigital neurovascular island (DDHD) flap. The axis of the flap curves obliquely and dorsally toward the contralateral side. B: The flap is elevated with its neurovascular pedicle and transposed obliquely into the defect and the typical wound suitable for a DDHDA flap with an intact tip and hypochium. C: The flap inset and grafted donor defect.

Preoperative Planning

The reconstruction of fingertip defects with homodigital island flaps should be performed in the operating room. The flap should be harvested with the use of a

tourniquet to facilitate identification of anatomic structures. A microscope should be available should the digital artery be injured and require repair, or if the nerve has to be coapted. Amputated remnants should be saved. These amputated parts can be used as a source for a skin graft for the tip defect if the flap cannot be used or even as the donor defect of the flap. The risks of the procedure should be explained to the patient and family. Patients need to understand that if the flap fails, they may require an amputation, as well as other risks such as loss of extension, cold intolerance, swelling, and loss of some sensation.

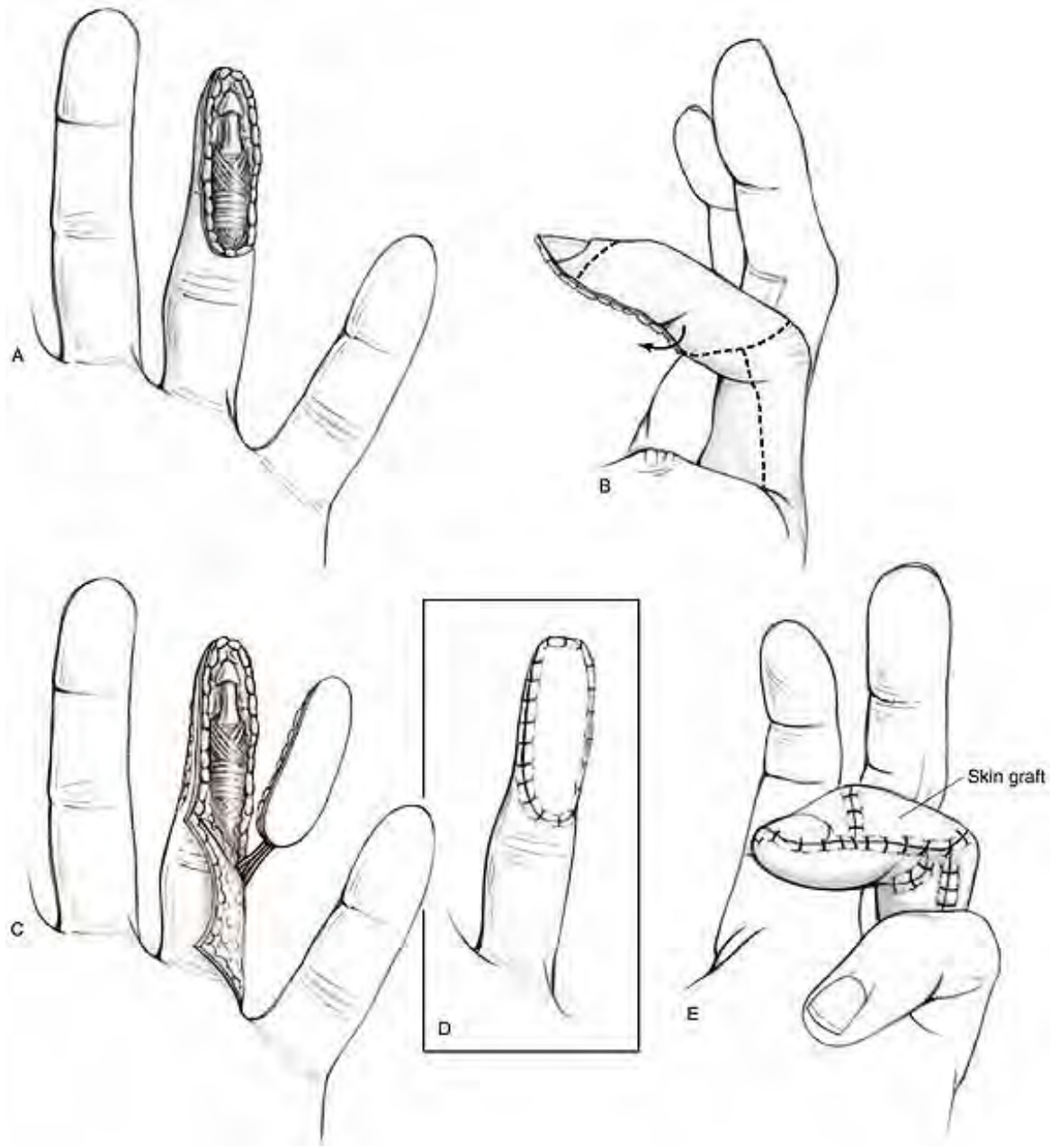

FIGURE 22-2 A,B: The rectangular flap design for the middle dorsal homodigital neurovascular island flap. C: The proximal pedicle dissection. D,E: The insetting of the flap and the grafted donor.

Surgery

Patient Positioning

The homodigital flap reconstruction of a finger defect can be performed in the operating room with the patient in the supine position with the involved hand on an arm board and an upper arm tourniquet. The arm is prepared up to the tourniquet. The procedure can be performed with a regional block, such as a bier block or axillary block, or under general anesthesia.

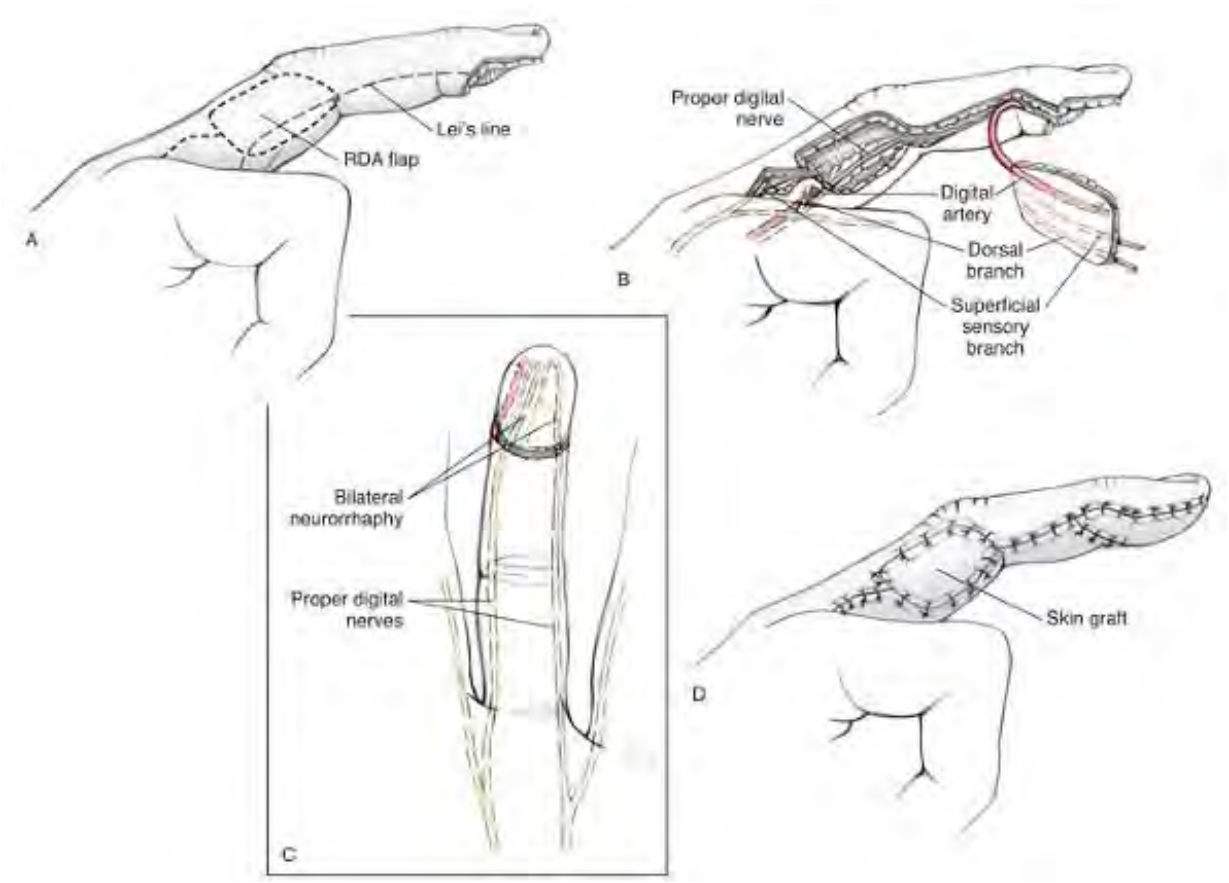

FIGURE 22-3 The required size of RDHDA flap is marked on the dorsolateral aspect of the involved proximal phalanx. A: The flap is based on the digital artery which topographically courses under the dorsoventral skin junction line. B: The vascular pedicle is ligated at its proximal end. Both the dorsal branch from the proper digital nerve and the superficial sensory branch from the corresponding radial or ulnar nerve are sectioned proximally, leaving 1 cm nerve tails attached

to the flap. C:The flap is transposed to the recipient wound. The attached sensory nerves of the flap are microanastomosed with both ulnar and radial digital nerves at the recipient wound. D: The flap donor site is resurfaced with full thickness skin graft. The pivot area of the vascular pedicle is also covered with a piece of skin to eliminate the pressure completely. (From *31*1990;46B:484. With permission.)

Technique for Distal Dorsal Homodigital Artery (DDHDA) Flap

A template of the defect is made and is used to design the skin paddle of the flap. The flap is outlined on either the ulnar-dorsal side or the radial-dorsal side of the involved digit, depending on the wound location. When either digital artery can be used, the non-opposition side of the digit is preferred. The flap design is based either over the distal phalanx and DIP joint (defect <2.5 cm) or the middle and distal phalanx, depending on the size of the defect. The distal margin of the flap should be maintained 4 mm proximal to the nail fold to avoid germinal matrix and resultant nail abnormality. Based on the width of the pulp defect, the DDHDA flap proximal incision line is marked obliquely from the volar edge of the pulp defect toward the DIP crease, reaching proximally to the dorsum of the middle phalanx of the finger. The flap may include a portion of volar tissue of the terminal phalanx if any residual pulp tissue remains intact. The extension for the distal and proximal incision lines on the dorsum of the digit is determined by the length of the designed flap. The length

of the flap should be longer than the defect to ensure adequate coverage of the flap pedicle (see Fig. 22-1). The neurovascular bundle is approached first via a zig-zag midlateral incision extending at least into the proximal interphalangeal crease. The digital pedicle is dissected proximal to distal with inclusion of surrounding fatty tissue to preserve the venae commitante with the proper digital artery to ensure venous drainage of the flap. This dissection should be performed with loupe magnification. In the proximal edge of the flap, particular attention is paid to include sufficient soft tissue between the pedicle and the flap to ensure adequate blood supply to the flap because the dorsal feeder arteries coming off the proper digital artery in the middle phalanx of the finger are usually divided during dissection. The flap is elevated in the

plane above the paratenon from the dorsum of the finger toward the pedicle, then to the volar edge (see Fig. 22-1). Once the flap elevation is complete, the flap is advanced obliquely to fill the pulp defect. The donor site on the dorsum of the finger is covered with a full thickness graft taken from the hypothenar eminence or upper forearm.

Technique Middle Homodigital Neurovascular Island (MDHDA) Flap

When a larger skin paddle (>2.5 cm and extending into the middle phalanx) is needed, the MDHDA flap is preferred. The MDHDA flap is designed as a rectangular island, with the axis of the flap parallel to the length of the volar skin defect. The rest of the procedure proceeds as described previously, except the proximal incision line is extended obliquely toward the proximal interphalangeal crease, and the dorsal digital branches at the middle phalanx are included (see Fig. 22-2).

Technique for Reverse Digital Island Flap

For very distal level injuries, when the arc of pedicle length might not reach for the DDHDA and MDHDA flaps, the reverse homodigital island flap can be used. The reverse digital island flap is designed on the dorsolateral side of the involved proximal phalanx according to the size and shape of the pulp defect (Fig. 22-3). The dissection begins proximally at the proximal digital neurovascular bundle at the base of the finger (Fig. 22-4). The dorsal nerve and artery branches from the proper digital nerve and artery are identified while preserving a thick tuft of soft tissue from the proper digital artery to the skin paddle. The skin paddle is elevated in the dorsal to volar direction. The skin over the neurovascular bundle is incised in zig-zag fashion. The proper neurovascular bundle is dissected in the proximal to distal direction, preserving the soft tissue around the proper digital artery to include the venae comitante to avoid venous insufficiency of the flap. Ideally, the proper digital nerve to the finger is preserved. Collateral flow from the contralateral digital artery to the pedicle proper digital artery should be preserved. The flap is inset and the dorsal nerve branch is coapted to

the digital nerve in the defect. Alternatively, the proper digital nerve can be included with the pedicle and repaired to the contralateral digital nerve in the defect.

However, this sacrifice of the proper digital nerve renders the finger insensate on the pedicle side, which can be avoided by using a dorsal nerve branch. The donor site at the base of the finger can usually be closed primarily, if a full thickness graft is not preferred.

FIGURE 22-4 A: Distal tip amputation of the long finger with exposure of underlying bone. In such a case, further shortening of the finger to allow for primary coverage would lead to a loss of the profundus tendon insertion and decrease finger function. B: A reverse homodigital island flap is designed on the ulnar aspect of the long finger. C: The digital artery is then dissected just distal to the proximal interphalangeal crease. D: The flap is ready for transfer. Fibrofatty tissue has been preserved around the vascular pedicle to prevent injury to the venous comitantes. The forceps hold a dorsal sensory nerve branch which has been included in the flap. E: The flap is rotated 180 degrees allowing the dorsal sensory nerve to be coapted to the injured radial digital nerve. F: The early postoperative appearance with excellent contour and well-healed donor site.

Results

Several reports with these flaps have demonstrated that the flaps have reliable arterial and venous flow. Moreover, these studies have shown good return of sensation in the 5â€“9 mm range of 2-point

discrimination. Some of the reviews have demonstrated a mild loss of extension complicating the skin grafted donor site on the dorsum of the injured fingers.

Postoperative Management

After the homodigital island artery flap, splint immobilization is unnecessary. Early mobilization can be encouraged which will lower the risk for a stiff finger postoperatively. If the RDHDA flap is used, usually the donor can be closed primarily. If a skin graft is required at the base of the finger, a splint is used postoperatively for 1 week.

Complications

Postoperative complications from the homodigital flap include loss of extension, loss of sensation, cold intolerance, and flap loss, most likely from venous compromise. The

risk of loss of extension can be minimized with early motion. Fingers that have been reconstructed with innervated homodigital flaps can achieve 5 mm 2-point discrimination, but the grafted donor site can be less sensate. There is no increased risk for cold intolerance in patients reconstructed with homodigital flaps over other options. The risk for venous compromise can be lessened by harvesting the pedicle to include soft tissue and venae comitante with the digital artery.

Recommended Reading

Atasoy E, Ioakimidis E, Kasdan ML, et al. Reconstruction of the amputated finger tip with a triangular volar flap: a new surgical procedure. *J Bone Joint Surg.* 1970;52A:921â€“926.

Flint MH, Harrison SH. A local neurovascular flap to repair loss of the digital pulp. *Br J Plast Surg.* 1965;18:156â€“163.

Foucher G, Smith D, Pempinello C, et al. Homodigital neurovascular island flaps for digital pulp loss. *J Hand Surg.* 1989;14B:204â€“208.

Joshi BB. A local dorsolateral island flap for restoration of sensation after avulsion injury of fingertip pulp. *Plast Recontr Surg.* 1974;54:175â€“182.

Lai CS, Lin SD, Chou CK, et al. Innervated reverse digital artery flap through bilateral neurorrhaphy for pulp defects. *Br J Plast Surg.* 1993;46:483â€“488.

Lucas GL. The pattern of venous drainage of the digits. *J Hand Surg.* 1984;9A:448â€“450.

Mober E. Aspects of sensation in reconstructive surgery of the upper extremity. *J Bone Joint Surg.* 1964;46A;817â€“825.

Oâ€™Brien B. Neurovascular island pedicle flaps for terminal amputations and

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 23

First Dorsal Metacarpal Artery Island Flap

GÃ¼nter Germann

Katrin Palm-BrÃ¶king

Soft tissue defects in the hand, and particularly in the thumb, frequently present difficult reconstructive problems because of the restricted availability of local tissue. Although local skin flaps have been employed and time-proven over decades, the philosophy of immediate wound closure of complex defects with exposure of tendon, bone, or joints has stimulated interest in anatomic research and the development of refined reconstructive techniques (10).

Traditional flaps have many disadvantages such as two-stage operations (i.e., cross-finger flap [17 ,24]), tedious dissections with considerable donor-site morbidity and loss of discriminative power (i.e., Littler flap [27 ,28]), or limited arc of rotation and mobility (i.e., transposition flaps [7] and the Moberg flap [1 ,8]). On the other hand, more recent microsurgical reconstructions such as the free pulp flap (7) require a significant amount of time and a familiarity with microsurgical technique. The first dorsal metacarpal artery island flap overcomes the disadvantages of traditional hand flaps and can provide a moderate-sized skin paddle on a long consistent vascular leash, allowing one-stage reconstruction for thumb and dorsal hand wounds without the need for microsurgical anastomosis.

The first dorsal metacarpal artery island flap was originally described by Hilgenfeldt in 1950 (16) and subsequently refined by Paneva-Holevic in 1968 and finally described as a

pure island flap (â€œkite flapâ€•) by Foucher and Braun in 1979 (8). Many variations of the flap have since been described and established in daily clinical use based on the consistent anatomy of the dorsal metacarpal arterial arcade (3 ,4 ,5 ,6 ,7 ,9 ,10 ,11 ,12 ,19 ,20). The majority of the flaps in recently published reports are raised from the dorsal aspect of the hand with only a few being found on the palmar surface (2 ,13 ,14).

The dorsum of the hand is supplied by the network of dorsal metacarpal arteries which are fed from the main forearm vessels: the radial, ulnar, and interosseus arteries. The first dorsal metacarpal artery (FDMA) consistently arises from the radial artery or the princeps pollicis artery and runs distally to the fascia of the interosseus muscle, frequently embedded in a fascial pocket. In most cases the first dorsal metacarpal artery divides in the middle of the second metacarpal into three terminal branches (18 ,22 ,24 ,30). The radial branch of the first dorsal metacarpal artery goes to the thumb, while the intermediate branch runs to the first web space. Distally the ulnar branch usually terminates at the level of the metacarpo-phalangeal joint and then arborizes into the dorsal skin of the index finger after giving off a perforating branch at the level of the metacarpal neck.

The exact location of the artery with respect to the first dorsal interosseous muscle may vary. Kuhlmann and de Frenne found the artery to run superficial to the muscle in 75% of specimens, while the FDMA was found within the muscle in 15% of specimens; in 10% of specimens the artery ran within and on top of the muscle during its course. Dautel and Merle identified a superficial course of the FDMA in 36% of specimens, a supra-fascial course in 23% of specimens, and a deep, intramuscular course in 56%. Despite its exact location all authors agree on the consistency of a FDMA (11 ,13 ,18 ,22 ,24 ,30) (Fig. 23-1).

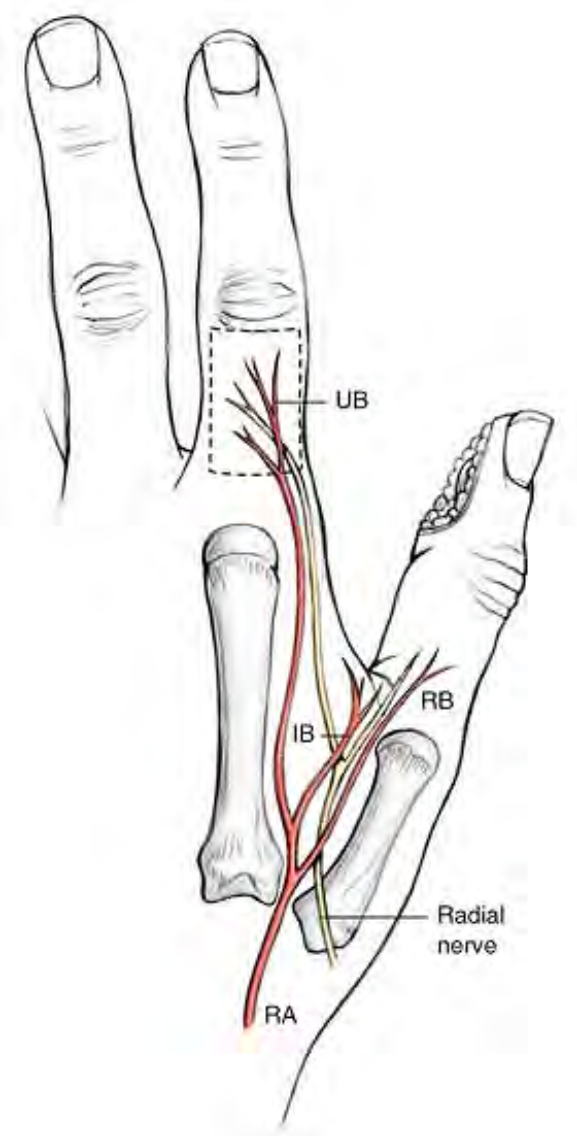

FIGURE 23-1 The three branches of the radial artery (*31*). The radial branch of the first dorsal metacarpal artery (*31*) goes to the thumb, the intermediate branch (*31*) to the first web space, and the ulnar branch (*31*) of the first dorsal metacarpal artery runs to the index finger.

Overall, the flap provides excellent sensibility, pliability, and stability for a variety of hand defects. We feel the first dorsal metacarpal island flap is one of the workhorses for reconstructive hand surgery (22 ,24 ,25 ,26 ,27).

Indications/Contraindications

Indications

Indications for the first dorsal metacarpal artery flap (â€œkite flapâ€•) are based on the following criteria:

- Etiology, location, size, and condition of the defect
- Availability of the flap (i.e., previous injuries that may compromise the vascular pedicle)
- Suitability of the flap for the defect (e.g., size, arc of rotation)
- Inclusion of other tissue components (i.e., tendon, bony segment)
- Patient's wishes (donor site morbidity, aesthetic appearance)
- Surgeon's familiarity with the technique

Based on these decision-making criteria, the flap is indicated for primary or secondary reconstructions of:

- The dorsal aspect of the thumb
- Restoration of sensibility with pulp reconstruction in the thumb
- Skin defects at the dorsal aspect of the hand within the arc of rotation of the antegrade kite flap
- Palmar defects of the index finger
- Defects of the web space

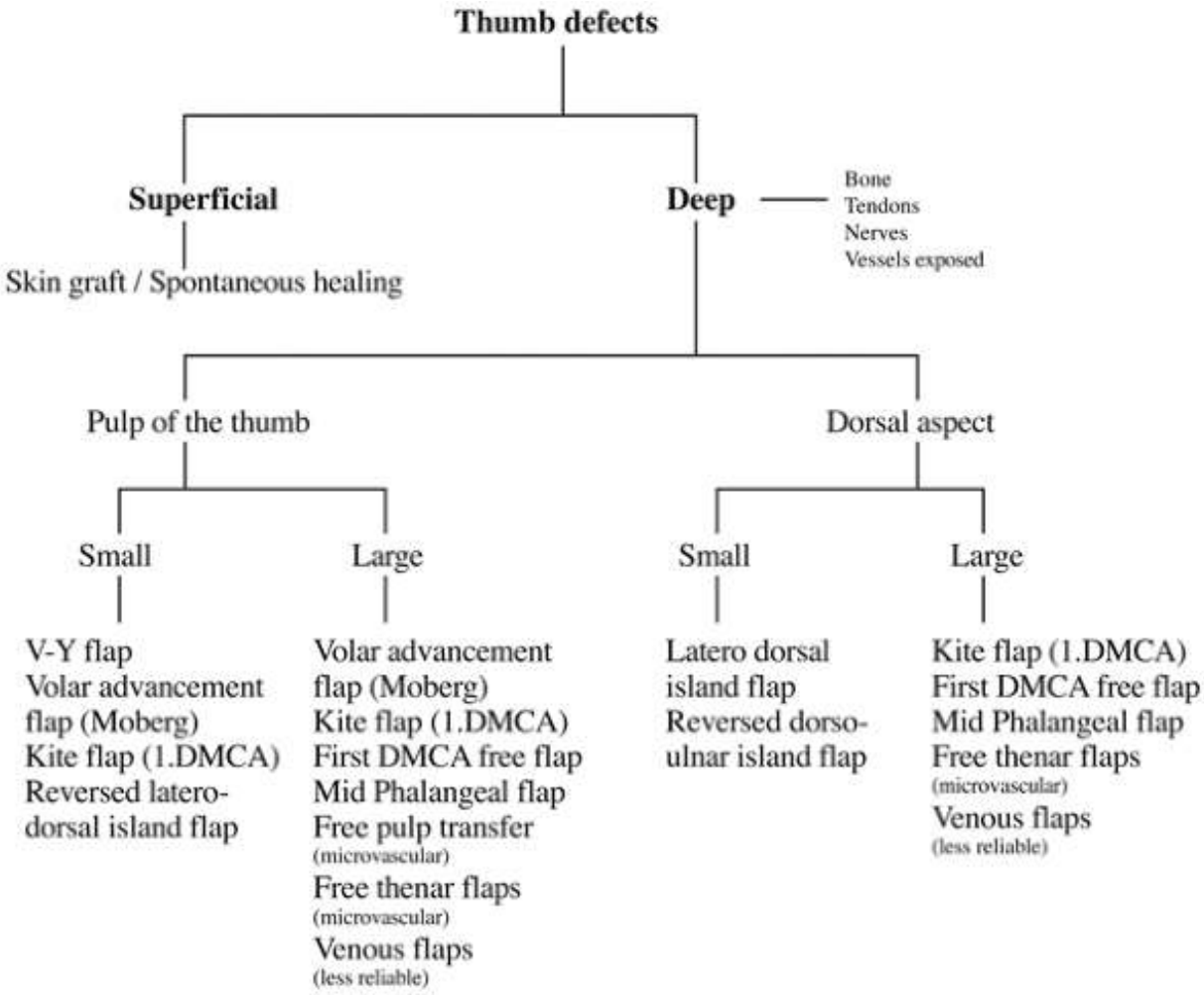

FIGURE 23-2 Decision-making algorithm for flap selection in treatment of thumb defects.

- Defects around the wrist
- Defects at the base of the thumb or the middle finger (â€œretrograde kite flapâ€•)
- Complex flaps including various tissue components
- Microvascular flap transfer based on the vascular pedicle in the snuff box

The list of indications demonstrates the versatility of the flap: however, reconstruction of the thumb takes precedence as the main indication in most centers.

Following the pioneer work of TrÃ¤nkle et al, the flap has replaced other sensate neurovascular island flaps for restoration of sensibility in the thumb (10 ,13 ,26 ,29) due to the dissection technique, the discriminative power that can be achieved, and the acceptable donor site morbidity. Based on this study, the flap is now widely used for this indication (Fig. 23-2).

Contraindications

Contraindications for flap use include acute trauma or a history of significant trauma to the donor site (15). In addition, injury mechanisms such as crush avulsions, wringer injuries, high-energy trauma (e.g., blast injuries or gun shot wounds) may limit the reliability of this flap if they have disrupted the course of the FDMA (10 ,13). If the tissue viability of the donor site cannot be estimated properly in the first hours of the trauma, wound closure should be delayed or a different flap should be used for wound coverage.

Burn injuries of the dorsum of the hand do not present a contraindication for the use of a first dorsal metacarpal artery flap in all cases. The dorsal metacarpal artery system is not damaged by deep partial-thickness burns that are excised and grafted. In 80% of patients with full-thickness burns, the dorsal metacarpal artery system is still intact. The potential for elevating a first dorsal metacarpal artery flap is therefore preserved after burn excision and grafting (11). Nevertheless, preoperative Doppler ultrasound examination is recommended before flap elevation.

Preoperative Planning

As in any surgical procedure, careful preoperative assessment and planning are key factors to success and the avoidance of complications or failure. Decision making follows an algorithmic approach in which the defect must be assessed for:

- Size
- Etiology
- Characteristics

The decision for flap selection is based on the criteria outlined previously (see indications). The patient should be thoroughly informed about the therapeutic options, and his or her expectations and wishes should also be discussed. Special attention should be paid to donor site morbidity including numbness over the radial aspect of the dorsum of the hand and the implications of a full-thickness skin graft to the donor site. The postoperative management and potential complications such as flap failure or impaired mobility of the fingers, impaired sensibility, and other possible undesirable results should be discussed with the patient. Donor-site morbidity is an important issue that should also be discussed with the patient before surgery. The functional and aesthetic morbidity of the donor site is an issue often neglected in the choice of a flap, and it must be borne in mind that the donor area of a first dorsal metacarpal artery island flap is in a very exposed position.

Clinical examination should exclude prior injuries in the area of the vascular pedicle. In most cases the artery is palpable, but Doppler ultrasound examination before flap elevation is mandatory (13) since the course of the artery is variable (see previous discussion). The course of the artery should be marked on the skin preoperatively after Doppler examination (13 ,19 ,26 ,27).

Surgery

The defect size is templated to the dorsum of the index finger. The course of the artery is marked on the dorsum of the hand with the aid of a hand-held Doppler ultrasound device. Surgery is performed using either an axillary bloc, or general or regional anesthesia. The use of a pneumatic tourniquet and loupe magnification are necessary for the dissection. The flap is harvested from the dorsum of the index finger and includes the first dorsal metacarpal artery, a branch of the superficial radial nerve, and at least one subcutaneous vein (Figure 23-3).

Dissection is performed starting with the vascular pedicle including a large subcutaneous vein which has first to be identified and mobilized. Elevation of the flap starts radially by incising the interosseus fascia at the most radial edge of the muscle. The fascia is then peeled off the muscle towards the second metacarpal. Thereby, a wide pedicle including the subcutaneous vein, a branch of the superficial radial nerve, and the vascular pedicle is created (24). A vessel loop is used to mark this pedicle. The pedicle is traced back to its origin in the snuff-box. There is no need to isolate the vascular pedicle from the

interosseus fascia, since this may injure the delicate vessels. Transillumination will verify the inclusion of the artery and the small accompanying veins in the pedicle.

Dissection now proceeds from the distal border of the flap. The flap is incised and dissected off of the paratenon of the extensor apparatus toward the proximal border of the flap. Care must be taken to preserve the paratenon in order to secure graft-take at the donor site. A critical point in the dissection is the radial aspect of the extensor hood of the metaphalangeal joint where the vessels enter the subcutaneous network of the flap. It is recommended to include a small strip of extensor hood to secure the vessel entrance. This area can be easily repaired without any functional deficit by a 4.0 polydioxanone (PDS) suture.

Once the flap is raised, the tourniquet is deflated to assess the blood supply of the flap. Occasionally, the flap may first appear pale but usually begins to turn pink within a few minutes (26 ,27). The flap may be passed through a subcutaneous tunnel into the defect (21 ,24 ,27). If the skin is too

tight, the tunnel is opened and the pedicle is skin grafted. As a sensory island flap, the first dorsal metacarpal artery island flap allows a wide arc of rotation due to its pedicle length of up to 7 cm (24).

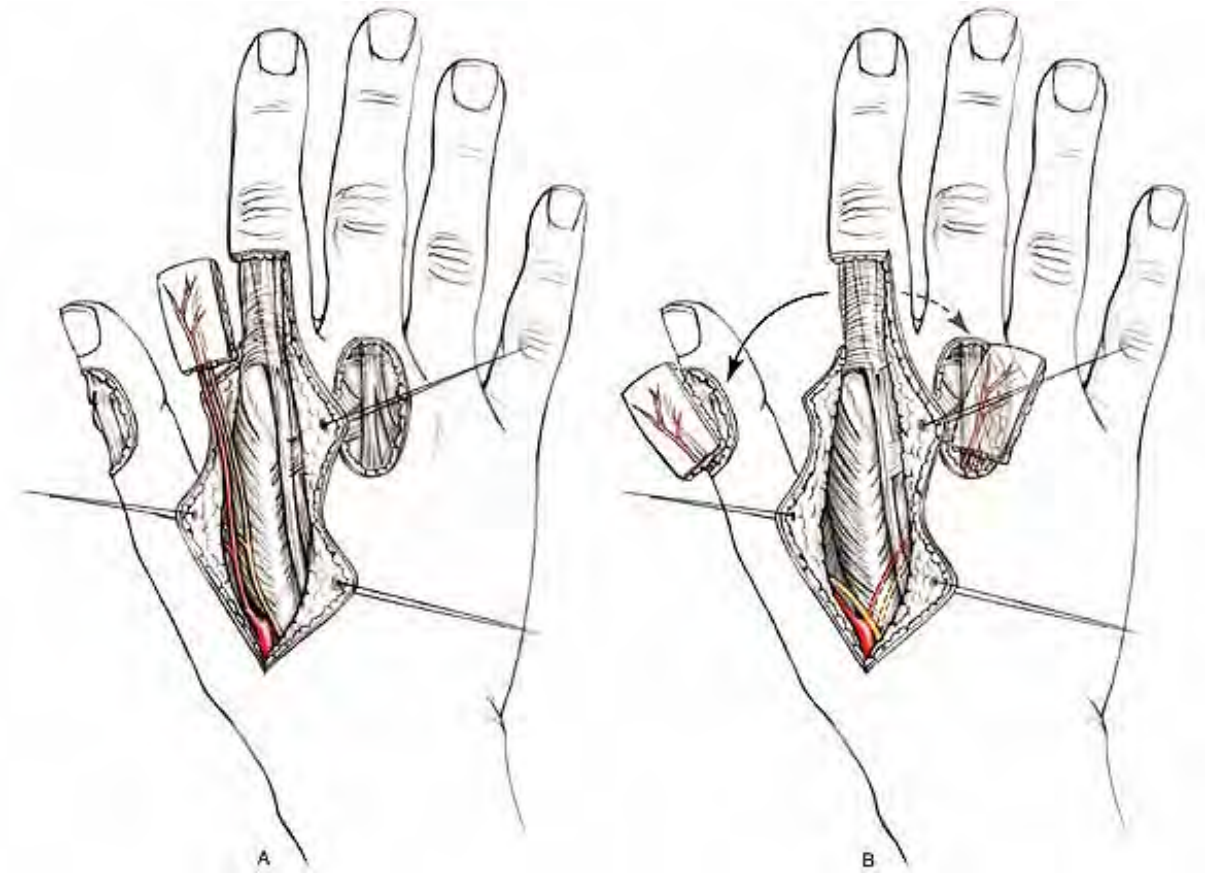

FIGURE 23-3 Schematic drawing of the dissected (A) first dorsal metacarpal artery flap and the first dorsal metacarpal artery flap with its pedicle passed distally under the skin tunnel (B) . (From Germann G, et al. *31* Berlin: Springer-Verlag; 2000. With permission.)

The donor defect of the index finger is covered with a full-thickness skin graft (19). A split-thickness skin graft may also be used; however a full-thickness skin graft will provide a better color match (23). Full-thickness grafting will provide an aesthetically acceptable appearance at the donor site within a shorter period of time, although the long-term results using a splitâ€“thickness graft are comparable (19) (Figs. 23-4 and 23-5).

Extensive experience with this flap has lead to some refinements. To avoid any tension on the pedicle, the subcutaneous tunnel has to be opened frequently. We now raise the flap constantly with a â€œcutaneous tailâ€• extending from the proximal border of the flap to the radial aspect of the second metacarpal. This extension allows opening of the tunnel and direct skin-skin closure after flap transfer. This technique does not only

provide a more stable skin closure, but also a better aesthetic appearance (see Fig. 23-4C,F). In addition, segments of the second metacarpal or the extensor indicis proprius tendon can be included in the flap for more complex reconstruction.

The first dorsal metacarpal artery flap can also be raised as a microvascular free flap (10 ,12 ,13). The flap is harvested using the same technique as the pedicled first dorsal metacarpal artery flap described previously. Dissection of the FDMA back to its origin from the radial artery provides an artery of adequate diameter for microvascular anastomosis. A long pedicle length facilitates microvascular

P.282

P.283

reconstruction outside the area of trauma; inclusion of the cutaneous nerve permits reconstruction of sensibility. We believe that dissection of this flap is easier and faster than a toe and first web-space flap (19) (Fig. 23-6).

FIGURE 23-4 A: A 63-year-old man sustained a mutilating saw injury of the hand with avulsion of the palmar aspect of the distal phalanx of the left thumb (additional injuries in the middle and small fingers). Six days after the injury: complete necrosis of the pulp of the thumb. B: Intraoperative view after radical debridement. C: Flap design: the first dorsal metacarpal artery island flap is outlined over the dorsal aspect of the proximal phalanx of the index finger. D: The first dorsal metacarpal artery island flap is elevated. Note the â€œcutaneous tailâ€• of the flap that allows primary skin-skin closure. E,F: After tunneling the flap into the defect, primary skin-skin closure is achieved. E,F: After tunneling the flap into the defect, primary skin-skin closure is achieved. G: Reconstruction of the donor site with skin graft. Hâ€“J: Aesthetic and functional outcome 2 years postoperatively. Full extension of thumb and index finger. The patient describes cold intolerance of the donor site. Two PD of the flap: 6.5 mm; 2 PD of the donor site: 6 mm.

Postoperative Management

The hand may be immobilized in a bulky dressing following surgery. A palmar plaster

splint may be used to immobilize any concomitant fractures. A window is left in the dressing for evaluation of the flap. Care is taken not to apply the dressing too tight for fear of causing arterial or venous compromise. Forty-eight to 72 hours after surgery, the flap's vascularity is ensured, and (13) physical therapy may be initiated if underlying fracture fixation is stabile. The index finger metacarpalphalangeal joint is usually immobilized for 5 to 7 days to ensure take of the skin graft over the donor site, but the proximal and distal interphalangeal joints may be mobilized as soon as possible to avoid postoperative stiffness. Occupational therapy is also initiated for sensory re-education of the flap. Initially, sensation in the flap will be perceived as coming from the dorsum of the index finger, or a dual sensation phenomenon may result. Cortical reorientation of sensation is possible with a structured sensory re-education program. The first dorsal metacarpal artery island flap allows immediate restoration of sensibility even in older patients, among whom neuronal co-adaptation of a pedicled or a free flap yields poorer results than in younger people (26 ,27).

FIGURE 23-5 A: A 44-year-old man suffered a work-related accident with a chain saw, resulting in a skin defect including a defect of the extensor pollicis longus tendon at its insertion and exposure of the interphalangeal joint. B: Design of the first dorsal metacarpal artery island flap on the dorsal of the index finger. C: The first dorsal metacarpal artery island flap prior to insertion into the defect. D: Result after insetting of the first dorsal metacarpal island flap. The flap shows perfect capillary refill. The donor site has been diminished in size by mobilization of the surrounding skin. Aesthetic

(E) and functional (F) outcome 2 years postoperatively. Excellent hand function and aesthetic appearance have been achieved. The patient returned to his previous employment without limitations.

FIGURE 23-6 A 22-year-old carpenter sustained an occupational high pressure-injection injury of the palmar aspect of the proximal interphalangeal joint of the index finger of the non-dominant right hand. A soft-tissue defect resulted, which included a large portion of the palmar aspect of the finger and exposure of both flexor tendons and neurovascular bundles. This was successfully reconstructed with a free kite flap (A) . The flap healed uneventfully with excellent aesthetic and functional result. Follow-up after 40 months (B) demonstrated a static two-point discrimination of 15 mm. Grip strength values were excellent. Time off work was 11 weeks.

Complications

Although the first dorsal metacarpal artery island flap has proven to be very reliable, various complications can occur. These complications include planning errors, complications during the operation, and postoperative difficulties. To avoid an inadequate arc of rotation or an inadequate flap size, the surgeon can add 10% to 15% more length to the pedicle, and the flap can be planned and harvested 10% to 15% larger in size (10 ,13).

Care should be taken to exert only gentle traction when passing the flap through the

subcutaneous tunnel to close the thumb pulp defect. Excessive tension can cause a partial compression of the vessels, and marginal necrosis could result. Whereas vascular insufficiency on the arterial side is frequently the result of tunnelling, a short pedicle or tight wound closure can compromise the venous outflow. In case of postoperative vascular insufficiency, sutures can be released and leech therapy may be employed.

Donor-site morbidity of dorsal metacarpal artery island flaps is low compared with axial homodigital island flaps. Neuromas are rare, and the patients report less hypersensitivity than in homodigital island flaps (13). However, one of the most important donor-site problems of the first dorsal metacarpal artery island flap is a diminished protective sensation over the transplanted skin graft, and patients may complain of cold intolerance. Hypertrophic scars present a minor donor-site problem. Functional limitation because of the donor site is negligible, and the range of motion of the donor finger reaches approximately 95% of the opposite index finger (26 ,27).

Recommended Reading

Adani R, Busa R, Castagnetti C, Bathia A, Caroli A. Homodigital neurovascular island flaps with â€œdirect flowâ€• vascularization. *Ann Plast Surg.* 1997;38(1):36â€“40.

Adani R, Squarzina PB, Castagnetti C, Lagana A, Pancaldi G, Caroli A. A comparative study of the heterodigital neurovascular island flap in thumb reconstruction, with and without nerve reconnection. *J Hand Surg.* 1994;19(5):552â€“559.

Dellon AL. Sensory recovery in replanted digits and transplanted toes. *J Reconstr Microsurg.* 1986;2:123.

Early MJ, Milner RH. Dorsal metacarpal flaps. *Br J Plast Surg.* 1987;40:333â€“341.

Gebhard B, Meissl G. An extended first dorsal metacarpal artery neurovascular island flap. *J Hand Surg.* 1995;20(4):529â€“531.

Grossman JAI, Robotti EB. The use of split-thickness hypothenar grafts for coverage

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 24

Thumb Coverage

David Elliot

The primary goals of thumb reconstruction are to restore length and sensation. A thumb should be of adequate length to allow for pinch to the index and middle fingertips and to participate in span grasp with all five digital tips. Adequate length is also essential for developing power grip. The surgeon aims to provide adequate coverage of the deeper, vital structures with soft tissue and skin of good quality. Sensibility of the tip is of particular importance, being crucial to fine pinch. A thumb â€œpostâ€• with no or little movement distal to the basal joint, and/or poor tip shape and sensation, can still work to aid in hand function but will not allow for finesse of function.

In attempting to obtain optimum reconstruction one should avoid shortening the thumb unless absolutely necessary. We encourage distal replantation or composite graft replacement, and also look to homodigital flap reconstruction whenever possible. Care should be spent in nail bed preservation and repair. Any nail greater in length than one third of its normal length should be retained, at least at primary surgery. Reconstruction of the thumb defects with the use of skin grafts from elsewhere on the body can create patches of dissimilar color, texture, and thickness and a noticeably poor cosmetic result. Finally, with respect to mobility and rehabilitation, the time-honored plastic surgical techniques of skin cross-finger, thenar, groin, and cross-arm flaps should be avoided, if possible, as they tether the injured thumb to another part

of the body, thus preventing adequate therapy, which may promote edema and stiffness.

Homodigital Flap Reconstruction

Homodigital flap reconstruction involves rearrangement of the soft tissues of the injured digit to achieve healing without seeking tissue for reconstruction from outside that digit. This concept of reconstruction has definite advantages for thumb reconstruction. In particular, it reconstructs â€œlike with likeâ€• and avoids the creation of further scarring and morbidity elsewhere on the hand or body. While homodigital reconstruction is advantageous in these respects, the availability of donor tissues within the thumb is obviously limited.

Perhaps the single most important surgical pearl to remember when using homodigital reconstruction is that the thumb and fingers have an astonishing ability to close skin defects of considerable size by a combination of wound contraction and re-epithelialization (Fig. 24-1). Thus, if a portion of the wound can heal under moist antiseptic dressings through secondary intention, the surgeon may use the homodigital flaps to cover only those portions of the wound which contain exposed vital structures or to improve soft tissue coverage over bone. The flap does not have to cover the entire defect, only the essential components of the defect.

Preoperative Planning

Preoperative planning should include AP and lateral radiographs in addition to a thorough hand exam. Concomitant arterial and nerve injuries can often be identified prior to exploration within the operating room. Many thumb injuries will occur in conjunction with other hand injuries and revascularization of injured digits takes precedence over coverage issues. Wounds should be debrided of all necrotic tissue and free of infection before embarking on soft tissue coverage.

FIGURE 24-1 A: Industrial scalpel wound of the thumb tip. B: Same digital surface after healing by re-epithelialization under moist antiseptic dressings for 3 weeks. (Reproduced from Elliot D. Specific flaps for the thumb. *Tech Hand Up Extrem Surg.* 2004;8(4):198â€“211, with permission.)

Patient Positioning

Patients are positioned supine on the operating room table. General or regional anaesthesia may be used during the surgical procedures. In some cases, thumb or wrist blocks may be used when performing smaller flaps or for debridement procedures. An upper arm tourniquet is placed prior to prepping and draping the patient. Surgery is always performed under tourniquet control to aid in visualization of vital structures.

The following sections present our preferred techniques for thumb reconstruction based on location of the original defect.

Palmar Defects

Neurovascular Tranquilli-Leali or Atasoy-Kleinert Flap

Indications/Contraindications

Historically, the original Tranquilli-Leali or Atasoy-Kleinert flap has been described to cover partial amputation defects of the distal phalanx; however, these flaps do not

work well on the thumb because of the inflexibility of the subcutaneous soft tissues. Laterally based single pedicle flaps, vascularized in the same way by the small vessels beyond the trifurcations of the digital arteries, also move poorly. The *neurovascular* Tranquilli-Leali *or* Atasoy-Kleinert flap is designed to be much larger and is islanded on both neurovascular pedicles bilaterally (Fig. 24-2). The flap is designed to extend to, or across, the IP joint crease proximally. This flap works well for minor transverse or oblique injuries of the distal phalanx with exposure of bone.

This flap can be used for stump reconstruction of any length of amputated thumb and is more useful than the original Tranquilli-Leali flap; however, it moves less freely than on a finger, partly because of the fibrous nature of the subcutaneous tissues of the thumb and partly because of what has been described as the â€œvertical dimensionâ€• of the thumb, that is its palmar-dorsal width at the tip (1). On fingers, it can be used to reconstruct defects with a palmar slope of up to 30 degrees; on the thumb it can only reconstruct defects with bone exposure which are dorsally facing, transverse, or palmar facing with less than 10 degrees of slope (Fig. 24-3).

Surgical Technique

The incisions of the V cross the interphalangeal joint crease at an angle and thus do not cause contractures (see Figs. 24-2 and 24-3). When designing the flap, one takes the V incisions out almost to the lateral nail folds distally. Having made the flap wide, the leading edge of the flap after advancement is wider than the original thumb tip. Unless the lateral corners of the flap are excised, this results in a spatulate end to the digit. Cutting off the lateral corners and allowing the resulting raw edges and tip to epithelialize not only narrows the digital tip but also rounds it to achieve a good appearance. The flap is designed as a â€œVâ€• at its proximal extremity and was conceived to close proximally as a â€œYâ€• after the flap has moved distally. Mostly, the proximal donor defect is left open to close under dressings as primary closure of the vertical limb of the â€œYâ€• tightens

the thumb too much proximally over the vascular pedicles of the flaps and compromises vascular inflow into the flap.

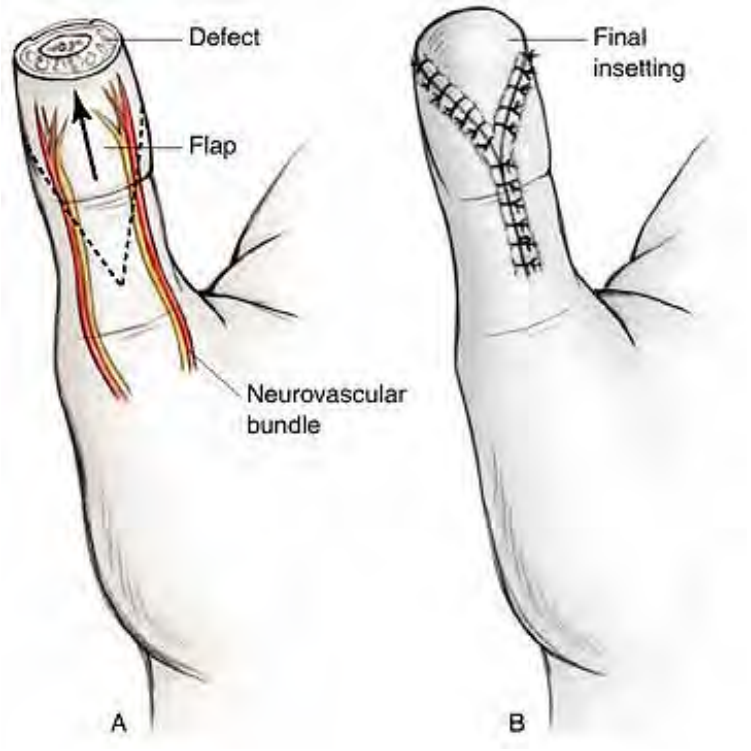

FIGURE 24-2 A,B: Illustration of the neurovascular Tranquilli-Leali flap. The flap extends proximally beyond the IP crease and includes both the neurovascular bundles.

A

B

C

D

E

F

FIGURE 24-3 A,B: Dorsal sloping crush amputation of the right thumb tip of a 70-year-old man. C:Simple folding back of the palmar tissues after excision of the nail germinal matrix to suture the skin to the proximal nail fold is not possible because of the bulk of the pulp tissue. D: A neurovascular Tranquilli-Leali flap designed to allow easier folding back of the palmar tissues and shaping of the thumb tip. The transverse skin split just beyond the â€œVâ€• was part of the injury, but only skin deep, so not affecting the vascularity of the tissues distal to it. E,F: The reconstructed shape after trimming the lateral points of the advanced flap to round the thumb tip. (Reproduced from Elliot D. Specific flaps for the thumb. *Tech Hand Up Extrem Surg.* 2004;8(4):198â€“211, with permission.)

Moberg Flaps

Indications/contraindications

With greater losses of palmar thumb tissue, more sloping palmar oblique injuries, and when the whole distal pulp of the thumb has been avulsed, the neurovascular Tranquilli-Leali flap is too small and cannot advance sufficiently. For larger thumb defects confined to the distal phalanx we favor the use of a modified Moberg flap. The Moberg flap, as described by Moberg in 1964, is an advancement flap based on both neurovascular bundles of the thumb (2). The advantages of this flap are that it returns sensation and glabrous skin to the defect site. The downside to the original flap design was the IP joint flexion contracture necessary to achieve suture of the flap distally to the nail. Our preferred modification of the Moberg flap avoids the IP joint flexion necessary with the original to advance the flap to the tip of the digit. This is, essentially, Oâ€™Brien's modification of the Moberg flap, but uses a â€œVâ€• tail proximally instead of a skin graft at the base of the thumb (Fig. 24-4) (3). This V-Y Moberg flap achieves the same excellent results as the original Moberg flap in terms of sensibility of the thumb tip, but without the restrictions of interphalangeal joint movement associated with the original flap or the need for skin graft of the Oâ€™Brien modification.

There is no danger of dorsal skin loss in raising Moberg-type flaps on the thumb, as the

dorsal skin of the thumb has a separate blood supply. Because of the safety of the V-Y Moberg flap, the speed and ease with which it can be dissected, and the excellent sensibility of the thumb tip which can be achieved with it, this technique remains our favorite thumb tip reconstruction if the defect is limited to the distal phalangeal segment of the thumb.

Technique

The flap is created laterally along the mid-lateral lines on each side of the thumb and with the tip of the proximal â€œVâ€• well back on the thenar eminence, at least as far proximally as a line drawn proximally from the ulnar border of the middle finger. This creates a large flap which incorporates the more lax subcutaneous tissues of the thenar eminence. Smaller flaps with the proximal â€œVâ€• at the base of the thumb move less easily and less far distally. The wider and longer flap avoids any need for addition of techniques to accommodate tightness at the base of the thumb. The use of this flap is considerably extended in reconstructing palmar oblique defects if the pulp distally which is denuded of skin cover remains as the leading edge of the flap, with suture of subcutaneous tissue, not skin, to the nail distally then epithelialization under dressings (Fig. 24-4).

Other Flaps

Other flaps have an infrequent use in reconstruction of the tip and palmar surface of the thumb but can be useful.

FIGURE 24-4 A,B: Preoperative views of the thumb of a 47-year-old man with a typical crush avulsion of the distal pulp. C,D: Markings of a V-Y modification of the Moberg flap shown preoperatively. E: Intra-operative view of the fully mobilized flap. F: Immediate postoperative view with the flap advanced to provide pulp cover of the distal bone. G: Final result after epithelialization of the tip under moist antiseptic dressings. (Reproduced from Elliot D, Yii NW.

Homodigital reconstruction of the digitsâ€”the perspective of one unit. *Handchir Mikrochir Plast Chir.* 2001;33(1):7â€“19, with permission.)

Lateral Pulp Flap

We devised the Lateral Pulp Flap for losses of the radial border of the tip of the index finger with exposure of bone but it is also useful on either border of the thumb tip (Fig. 24-5). This flap exploits the excess of pulp in the digital tip.

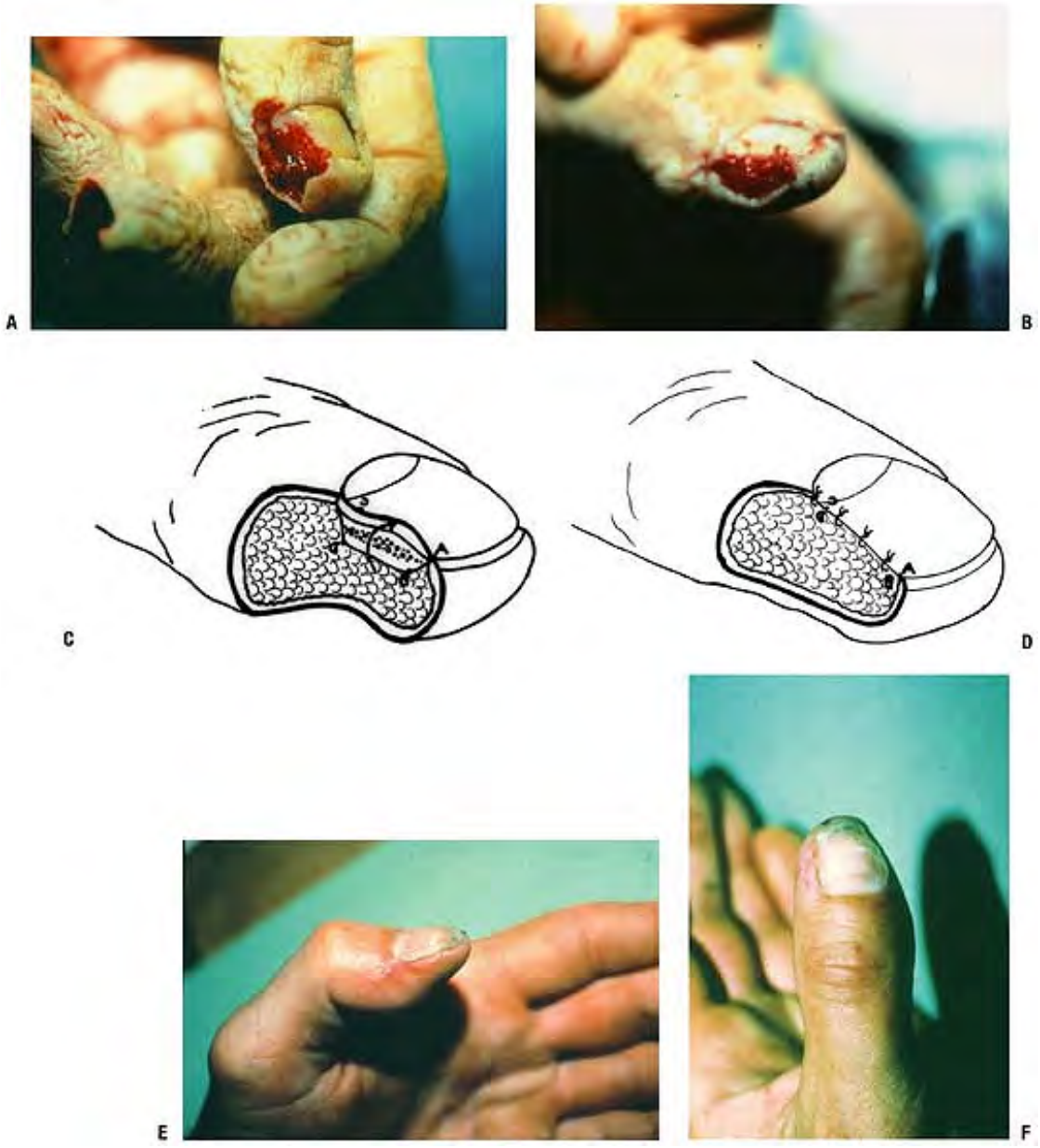

FIGURE 24-5 A: A slicing defect with loss of the radial lateral pulp, lateral nail fold, and lateral one-third of the nail of the thumb in a 50-year-old man. B: Lateral pulp transposed to cover the distal phalanx. C,D: Diagrams to show the original injury and transposition of the pulp to cover the bone of the distal phalanx. E,F: Late views of the reconstruction, 1 year after the injury. (Parts C and D reproduced from Elliot D, Jigjinni VS. The lateral pulp flap. *J Hand Surg Br.* 1993;18(4):423â€“426, with permission.)

The flap is raised by opening the tip of the digit with a fish mouth incision close to the distal nailbed and freeing the pulp attachments to the bone. The pulp then moves laterally and is lifted over the bone and sutured to the edge of the nail. The deep edge of the pulpâ€"not the superficial edgeâ€"is brought up to the nailbed to cover the bone. The pulp is then epithelialized under moist antiseptic dressings. This reconstruction creates a digital tip which is sensate but has no lateral nail fold. This seems to cause no functional problems and is less obvious than most lateral nail fold reconstructions, as these are usually too bulky.

Side-to-Side Homodigital Switch Flaps

Side-to-side homodigital switch flaps, which reconstruct or re-innervate one side of a digital tip at the expense of the other side, can be useful to resurface areas of pulp sensibility in the hand which are critical for pinch activity after localized loss of tissue or irrevocable digital nerve injury. Replacement of the radial side of the index or middle fingertips by a vascularized composite transfer from the ulnar side of the same digit is well established, both as a simple transposition and as an island flap. Use of the radial pulp of the thumb tip to reconstruct and/or reinnervate the ulnar side is logical, as the ulnar portion of the thumb tip is most involved in majority of gross and fine pinch (Fig. 24-6).

Littler Flaps

Indications/Contraindications

The classic Littler flap, as described by Littler in 1956, was designed for thumb tip and palmar thumb resurfacing. Unlike the flaps described above, the Littler flap is a heterodigital flap, which means a donor site defect will be created on the middle or ring finger. The flap utilizes tissue from either the ulnar border of the ring or middle finger; this tissue is transferred on its neurovascular bundle, thus allowing for the return of sensation to the thumb (Fig. 24-7). This differs from the heterodigital flaps described elsewhere in this text, which are designed to spare the digital nerves. This flap is easy to dissect and transfer. However, the flap can considerably downgrade the donor finger function and restoration of adequate two-point sensation after transfer

has been a problem and suffers problems of achieving â€œthumb-tipâ€• sensibility. To improve thumb- tip sensation, and avoid patient's perception of the thumb-tip as still coming from the donor finger, division of the donor finger nerves and reconnection of the nerves in the flap to those of the recipient thumb has been recommended. Nerve disconnection and reconnection extends its usefulness in younger patients but, like all neurorrhaphies, gives less satisfactory tip innervation in the middle-aged and elderly patients.

Direct contraindications for this flap include injury to the neurovascular pedicle to the donor finger. Laceration of the superficial arterial arch or common vascular bundle to the donor finger could potentially lead to compromise of the flap or donor finger following transfer. Very distal tip injuries may also be difficult to cover with this flap.

Complications

Specific complications to be noted with the use of the Littler flap include the risk of venous congestion, specifically if the nerve is dissected or divided in preparation for neurorrhaphy at the level of the thumb. In such cases leech therapy may be required for resolution of venous congestion. Poor return of thumb sensation may also occur or the inability of the patient to recorticate, in which case the sensation within the flap is always perceived by the patient as coming from the donor finger donor site. Either sensory problem may necessitate a secondary nerve graft or nerve repair, but two-point discrimination may well remain <12 mm within the flap.

Reconstruction of the Dorsal Surface of the Thumb

The Hatchet Flaps

Indications/Contraindications

For the more common, smaller defects of the dorsum of the thumb, homodigital techniques are very useful. They are generally simpler and quicker to execute and avoid injury of adjacent fingers. For these small defects the Hatchet Flap provides an easy reconstructive option. Hatchet Flaps are also very useful in replantation, as they allow one to advance the dorsal veins of the proximal part of the thumb to the replant and achieve direct vein anastomoses under good skin cover without recourse to either vein grafts or a separate skin cover procedure. These flaps can be used as far distally

as the proximal nail fold (Fig. 24-8).

Surgical Technique

The flap is designed as a random dorsal cutaneous flap (Fig. 24-8). A curvilinear incision is made over the radial (Fig. 24-9) or ulnar (Fig. 24-8) aspect of the thumb. A broad base is preserved on the opposite side of the digit. As on the palmar surface when utilizing a Moberg flap, a V-Y tail can be added to these flaps to avoid the need to skin graft the donor site proximally (Fig. 24-9)

(4). On this surface of the digits, it is necessary to close the proximal V as a Y to protect the extensor paratenon from desiccation.

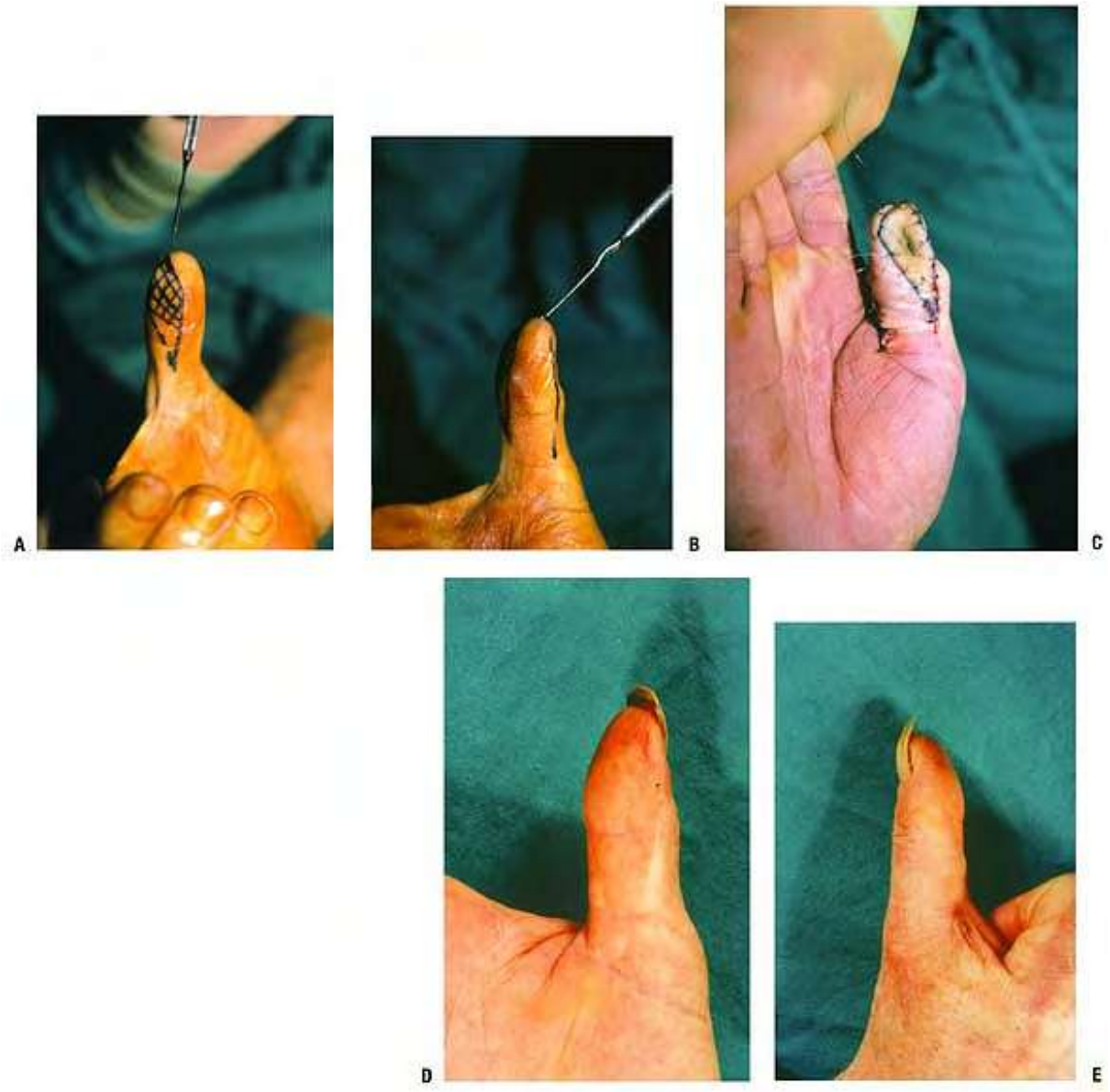

FIGURE 24-6 A: Thumb of a patient with irreparable damage to the ulnar digital nerve. A triangle of the denervated ulnar digital nerve territory is shaded, prior to excision down to bone. B: The same thumb showing the pre-operative marking of the radial switch flap. C: The thumb at surgery, after transposition of the flap and with the graft applied to the donor defect. D: Late view of the grafted donor site. E: Late view of the switch flap. (Reproduced from Elliot D, Southgate CM, Staiano JJ. A homodigital switch flap to restore sensation to the ulnar border of the thumb tip. *J Hand Surg Br.* 2003; 28(5):409â€“413, with permission)

Results

In a review of 1,077 dorsal wounds of all digits, which we treated over a 6-year period, 154 digits required flap reconstruction (5). The dorsal V-Y flap accounted for 42% of the flaps used and was the most common skin flap used on this surface of the digits, reflecting the relative incidence of defects of this size and shape.

Brunelli Flap

Surgical Technique

This homodigital flap is a useful adjunct for reconstructing smaller defects of the dorsum of the thumb. In 1993, Brunelli and his colleagues in Paris described a very easily and quickly raised flap based on the dorsoulnar artery of the thumb which is useful for reconstruction of dorsal defects of the tip of the thumb involving all or parts of the nail complex (Fig. 24-10) (6). The artery may be verified preoperatively with a handheld Doppler probe, but is very consistent running on the ulnar aspect of the extensor tendons. The flap is raised with a skin tail to reduce the risk of compression

of the pedicle just proximal and lateral to the nail fold, which is the tightest part of its course. The flap may be safely elevated to the proximal margin of the interphalangeal joint skin crease. The donor site may be closed with a skin graft or advancement flap from the dorsum of the hand (Fig. 24-10).

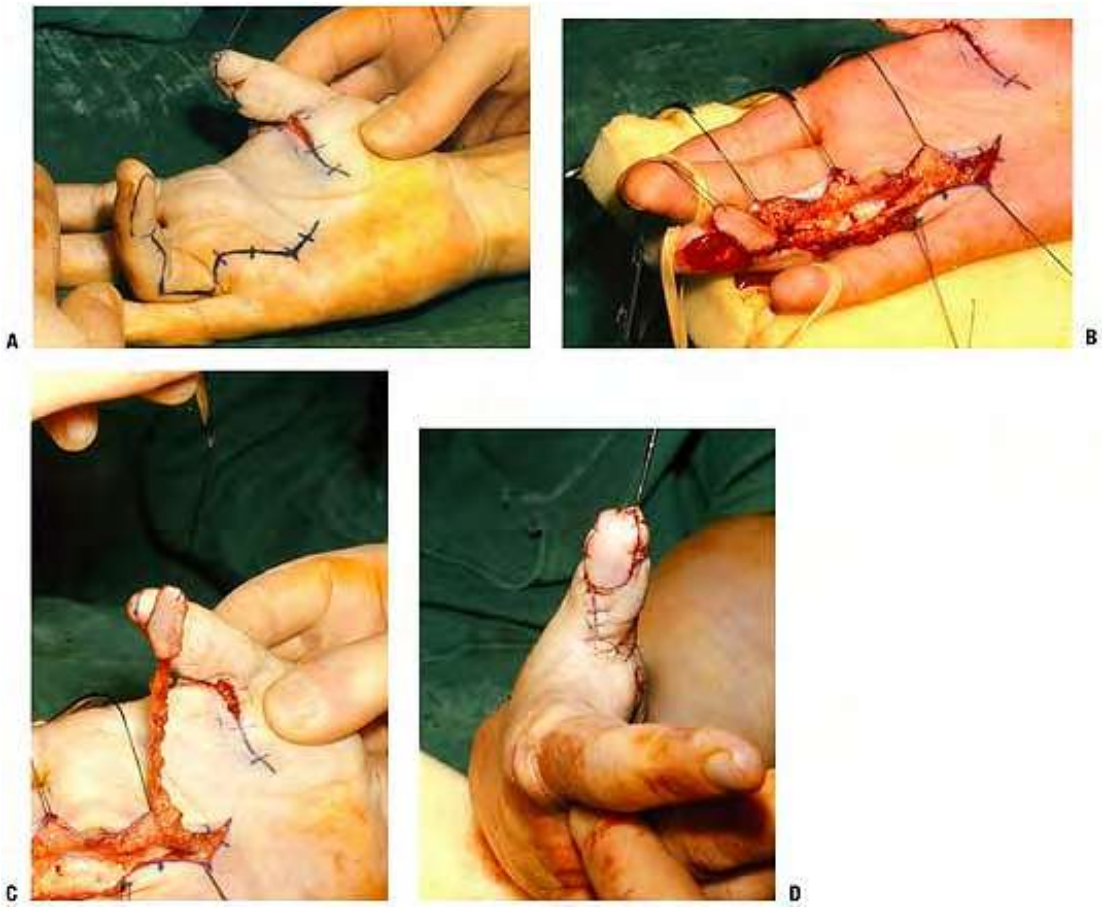

FIGURE 24-7 A: Intra-operative view showing dissection of damaged and denervated skin on the ulnar aspect of the thumb tip and the markings of a Littler flap on the ulnar side of the ring finger. B: Dissection of the pedicle of the flap. C: Demonstration of the transfer. D: Final view of the flap after tunnelling across the palm and insetting.

FIGURE 24-8 A: Preoperative view of a typical â€œsmallâ€• skin defect on the dorsum of the thumb. A simple dorsal hatchet flap is marked. B: Closure of the defect with this flap, with split skin graft reconstruction of the donor defect. (Reproduced from Elliot D. Specific flaps for the thumb. *Tech Hand Up Extrem Surg.* 2004;8(4):198â€“211, with permission.)

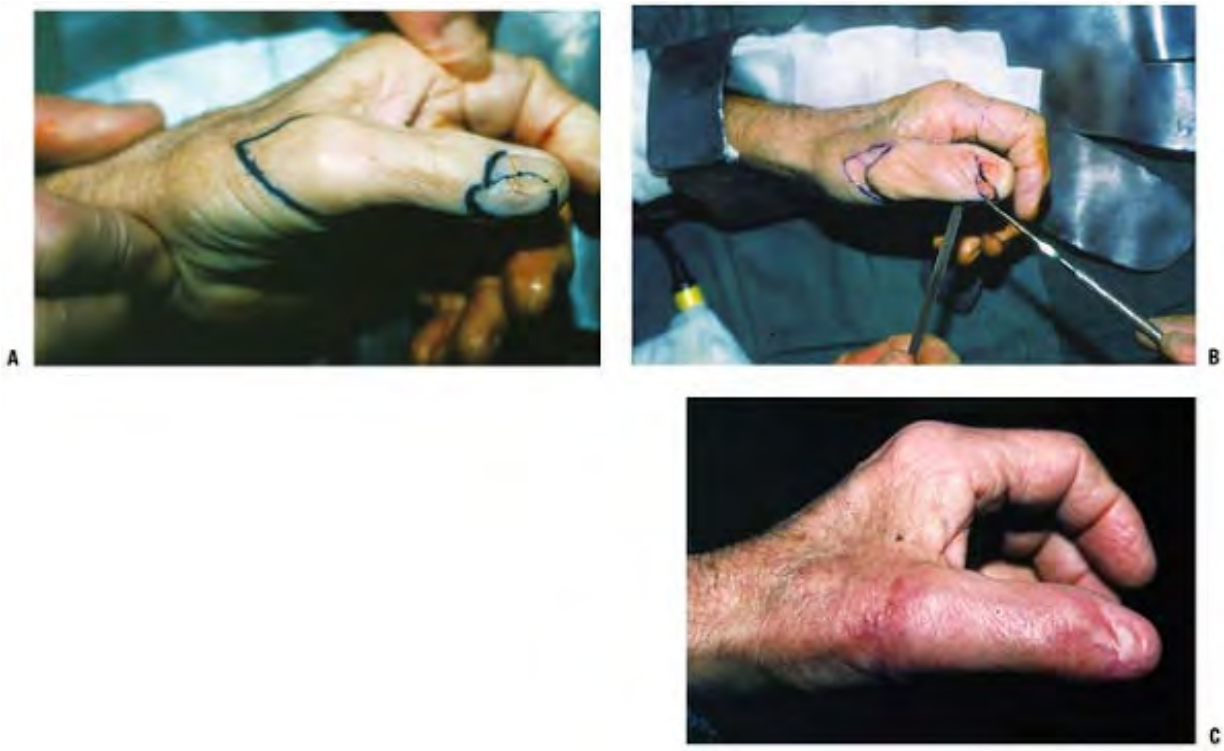

FIGURE 24-9 A: Preoperative view of the thumb of a 73-year-old man with a squamous cell carcinoma of the radial lateral fold of the nail. The thumb is marked with a dorsal V-Y hatchet flap for reconstruction. B: Intra-operative view of the flap being advanced. C: Late postoperative view. (Reproduced from Elliot D. Specific flaps for the thumb. *Tech Hand Up Extrem Surg.* 2004;8(4):198â€“211, with permission.)

This flap can be innervated for reconstruction of the thumb tip but, like the other flaps of dorsal origin, carries skin with poor innervation. It has more recently been redescribed carrying a segment of vascularized bone from the first metacarpal with it and may prove a very useful means of placing vascularized bone at the tip of the thumb without microsurgical transfer from the foot (7).

Reconstruction of Large Dorsal Defects

A very few techniques of reconstruction suffice to cover almost all full-thickness soft tissue losses on the dorsum of the thumb. Complete loss of the dorsal surface cover

can be replaced using the first dorsal metacarpal artery or with a reverse posterior interosseous artery flap; both flaps are described in detail elsewhere in the text. The latter is particularly useful if the adjacent dorsal tissues of the hand have been injured and are not available as a donor site. It can also carry vascularized bone (and tendon) to the thumb.

The posterior interosseous vessel is small and the dissection relatively difficult, requiring microsurgical instrumentation and expertise. However, this flap is mostly safe and does not involve sacrifice of a major blood vessel of the forearm. For this reason, it is preferred to the use of the radial forearm flap for larger defects of the dorsum of the thumb, although the latter is useful for complete wrap-around of the first ray. When used to reconstruct the dorsum of the thumb, the reverse posterior interosseous artery flap donor site can almost always be closed primarily and then heals with a fine, narrow scar which may be almost invisible in the hair of a male forearm.

Postoperative Management

We remove all dressings and splints from reconstructed hand injuries on the day after surgery to allow washing and dressing of the raw wounds and verification of flap viability. Although this sometimes

has to be done with care to avoid tension on flap pedicles, it allows the hand to be bathed in warm water. This both helps general cleaning and debridement of the wounds and is a comfortable environment in which to encourage very early hand mobilization. The wounds are dressed as lightly as possible while maintaining a moist antiseptic presence around any areas without skin which will heal through secondary intention. Suture lines that are not oozing are not covered. Sometimes homodigital and more distant reconstructions require that one or more joints be prevented from moving in a certain direction to take tension off the pedicle. For example, with all of the advancement flaps on the palmar surface, full extension will pull on the flap pedicles, so the thumb is splinted in slight flexion. Conversely, a slightly extended wrist position will ease tension on the pedicle of reverse posterior interosseous artery flaps. Protection of pedicles and flaps is continued during bathing and dressing changes by appropriate positioning and guarding with the other hand and between dressing changes by use of thermoplastic splints. Bathing and dressing changes are

carried out twice a day with the patients learning to do this themselves as soon as possible.

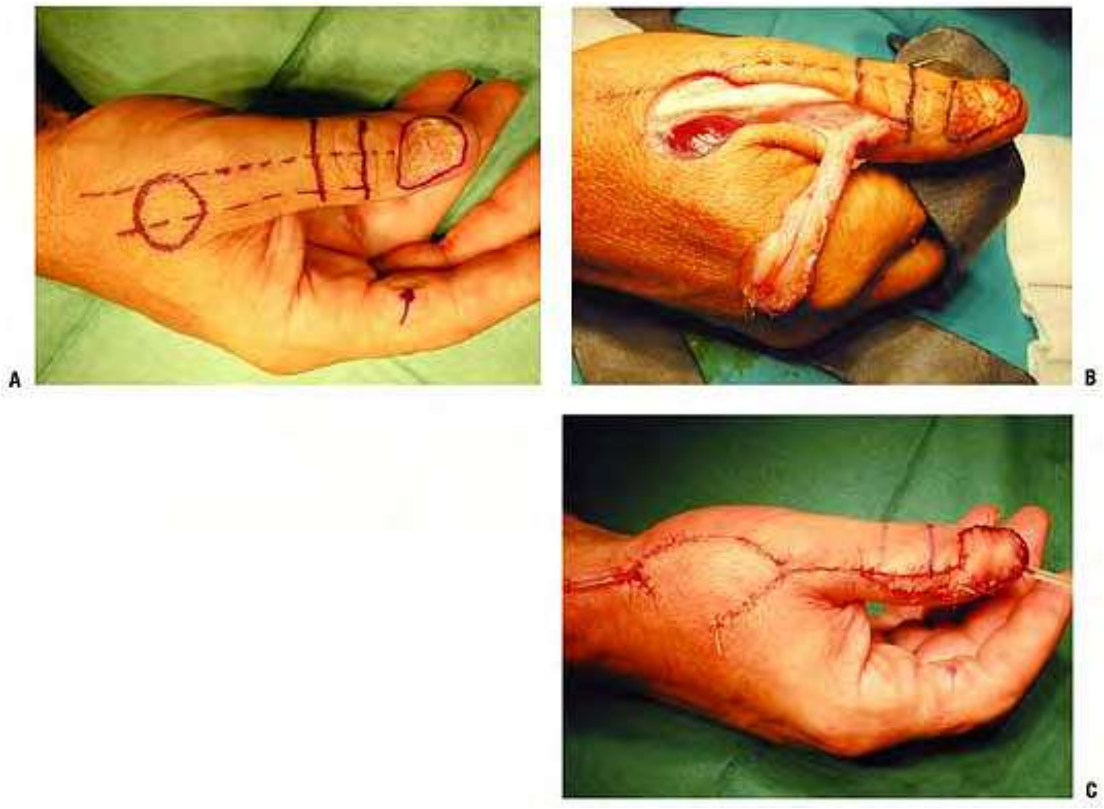

FIGURE 24-10 A: A thumb with a badly damaged nail before excision for biopsy. The Brunelli flap and its pedicle have been marked. B: The flap has been raised and shows the artery on which it is based on its undersurface. C: The flap inset after total excision of the nail complex. The flap has been taken with a â€œtailâ€• to ease pressure on the pedicle beyond its turning point. The donor site has been closed with a dorsal V-Y hatchet flap. (Reproduced with permission from Elliot D. Homodigital reconstruction of the digits: The perspective of one unit. *Indian J Plast Surg.* 2003;36:106â€“119.)

Rehabilitation

The use of homodigital flaps allows for early mobilization of the reconstructed thumb.

Motion will be restricted by concomitant tendon, nerve, and bony injuries; however with stabile skeletal fixation, most thumbs can be mobilized using limited motion and passive motion protocols under the supervision of a hand therapist within the first 72 hours following surgery. Whenever possible, we avoid techniques of bandaging, splinting, and skeletal fixation which will completely immobilize the reconstructed thumb. Over and above the early mobilization under water during dressing changes, mobilization of as much of the thumb and the rest of the hand as possible is continued from this early stage by exercises three or four times daily with the splint in place, albeit sometimes with loosening of some straps if this is possible without loss of the protection of flap pedicles. The splint provides a solid resting position for the thumb at night and gives the patient confidence that their thumb will not suffer from jarring by unexpected contact with external forces. Splints are generally used for 1 to 2 weeks. However, this period may lengthen to 4 or 5 weeks, if required to protect structures

repaired under the flap. Between activities, the hand is elevated. This regimen of elevation and enthusiastic exercising is intended to eliminate edema from the hand as quickly as possible to minimize fibrin restriction of joint and extensor tendon movements.

Complications

Major complications include flap loss, partial flap loss, infection, and iatrogenic injury to the digital arteries or nerves. Partial flap loss is often best managed by allowing the wound to heal through secondary intention. Total flap loss, while infrequent with these homodigital flaps, will necessitate the use of a heterodigital island flap such as the first dorsal metacarpal artery island flap, or a larger pedicled forearm flap, such as the posterior interosseous flap or radial forearm flap. Obviously such flaps will create further injury to the arm but are reliable means of providing soft tissue coverage. An additional downside to the use of heterodigital flaps is that they are often insensate or require a neurorrhaphy to re-establish sensation. Return of sensation in these situations may be dependent on the age of the patient. In addition, heterodigital flaps will create an injury to another digit, which may further compromise hand function.

Results

The techniques described here have proved adequate to reconstruct most of the thumb defects we have seen in our practice over the last 12 years. The principle used in the smaller cases, particularly in reconstructing the tip of the thumb, has been to capitalize on the enormous healing capacity of digital skin and assist this, when necessary, by simple local flap reconstructions. These provide the best sensation possible and the best cosmetic profile to the thumb tip. For larger defects we prefer to use homodigital techniques whenever possible, although there are upper limits to the size of defects which can be reconstructed by these. For any defect, our choice of flaps is, in part, determined by the site, size, and shape of the defect and, in part, by flap reliability and ease of use. The homodigital flaps described above have proved particularly versatile and reliable in the hands of surgeons of varying levels of experience within our unit, although requiring some familiarity with microsurgical technique and instrumentation. They are quick to perform, which can be an advantage following lengthy bony or soft tissue reconstructions. They allow early and independent mobilization of the thumb and the use of local tissue respects the cosmetic principle of reconstruction of â€œlike with like.â€• Donor site morbidity is also limited to the already injured part. When used for fingertip reconstruction, advancement homodigital flaps cannot achieve the sensibility of the original, but are durable and intrinsically more likely to restore good sensory function in the digital tips than most of the reconstructive alternatives. For larger defects, the Littler transfer, the dorsum of the index ray, and, more recently, the reverse posterior interosseous artery flap are our local work horses and are discussed in detail in subsequent chapters.

References

1. Gaul JS. A palmar-hinged flap for reconstruction of traumatic thumb defects. *J Hand Surg Am.* 1987;12:415â€“421.

2. Moberg E. Aspects of sensation in reconstructive surgery of the upper extremity. *J Bone Joint Surg Am.* 1964;46:817â€“825.

3. Oâ€™Brien B. Neurovascular island pedicle flaps for terminal amputations and

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 25

Soft Tissue Management Following Traumatic Injury to the Femur

Thomas F. Higgins

Open injuries of the femur and thigh are common. The need for major soft tissue coverage in this area in the setting of trauma is uncommon due to the generous surrounding muscular envelope. Soft tissue coverage of the femur can be a concern in certain situations such as power takeoff injuries or pedestrians struck by motor vehicles; however, the majority of these defects may be covered with skin grafts, and the need for vascularized free tissue transfer is unusual. Despite this, there are several components to soft tissue management in this area which may be used during exposure to the proximal and distal portions of the femur. Simple alterations in established techniques can lead to improved postoperative contour, improved soft tissue coverage of hardware, and decreased rates of postoperative infection and stiffness following elective and emergent surgical procedures.

This chapter covers four major areas of concern with regard to soft tissue management in the thigh:

- The approach to the proximal femoral shaft and intertochanteric zone
- The approach for the repair of supracondylar and intra-articular distal

femur fractures

- Exposure for the treatment of open femur fractures
- The management of post-traumatic arthrofibrosis of the knee with the use of a Judet quadricepsplasty

Approach to the Lateral Proximal Femur

Indications/Contraindications

Fractures of the proximal femur, whether femoral neck or intertrochanteric, constitute the majority of all femur fractures (9). The soft tissue management issues in this portion of the proximal femur are the relative dearth of coverage over the greater trochanter and the area of the femur just distal to the vastus lateralis. Frequently, plate and screw implants and laterally based orthopaedic constructs will be prominent and symptomatic, particularly in thin patients. An approach that maximizes soft tissue coverage over the lateral aspect of the proximal femur and restores the most normal anatomy will be described. Rather than a direct lateral approach cutting through the vastus lateralis, which has been frequently advocated for this operation, a simple modification provides a more anatomic dissection and a friendlier soft tissue reconstruction.

Patient Positioning

The patient is positioned as desired for ultimate reduction and fixation goals.

Surgery

A direct lateral approach begins 1 cm proximal to the vastus ridge and extends as far distally as necessary for osteosynthesis. On the coronal plane, the incision should be 1 cm posterior to the midcoronal point of the femur (Fig. 25-1). Dissection is taken down to the iliotibial band, and this is divided in line with its fibers (Figs. 25-2 and 25-3).

A gentle sweeping motion with a sponge will clear trochanteric bursa or areolar soft tissue that lies in this layer. Particularly fibrotic bursal tissue may need to be excised with scissors (Fig. 25-4). This will reveal quite clearly the proximal extent

of the vastus lateralis and its insertion at the vastus ridge and the confluence with the distal extent of the vastus medialis tendon. Rather than directly incising the vastus lateralis in the midcoronal point of the femur, a J-shaped incision is performed. Leaving several millimeters of the vastus lateralis tendon still attached to the vastus ridge, the tendon is incised from its anterior extent posteriorly until connecting with the lateral intermuscular septum (Figs. 25-5 and 25-6). The dissection is then taken through the most posterior fibers of the vastus lateralis extending from proximal to distal, just anterior to the lateral intermuscular septum. The

vastus tendon is tagged with a heavy braided nonabsorbable suture before being detached from the vastus ridge (see Fig. 25-5). As the dissection extends greater than 5 cm distal to the vastus ridge, the first perforating branch off the profunda femoris will be encountered penetrating the lateral intermuscular septum. These perforators should be ligated or cauterized. A gentle sweeping motion with an elevator from posterior to anterior along the proximal femur will allow reflection of the vastus lateralis in an anterior direction (Fig. 25-7). The periosteum and deepest soft tissues should not be stripped off the femur. For ideal visualization at this point, a Bennett retractor is placed anterior to the femur, and an assistant holds this forward (Fig. 25-8). Osteosynthesis proceeds as planned.

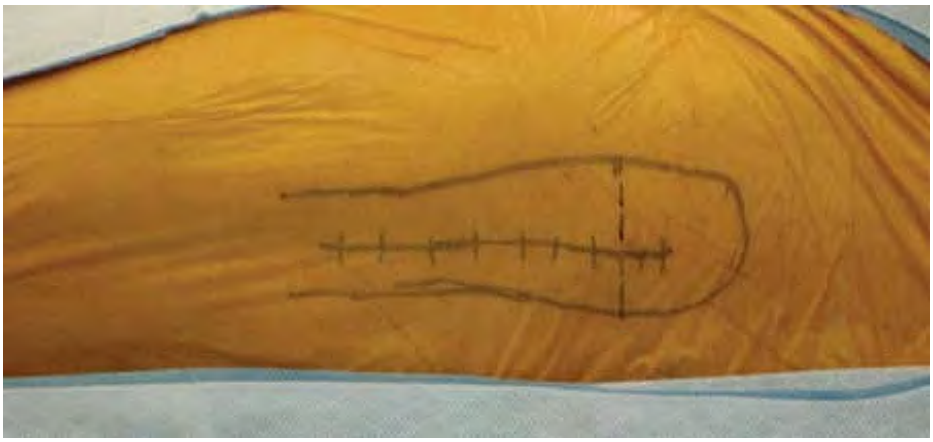

FIGURE 25-1 Incision drawn over the greater trochanter. *Dotted transverse line* represents the vastus ridge.

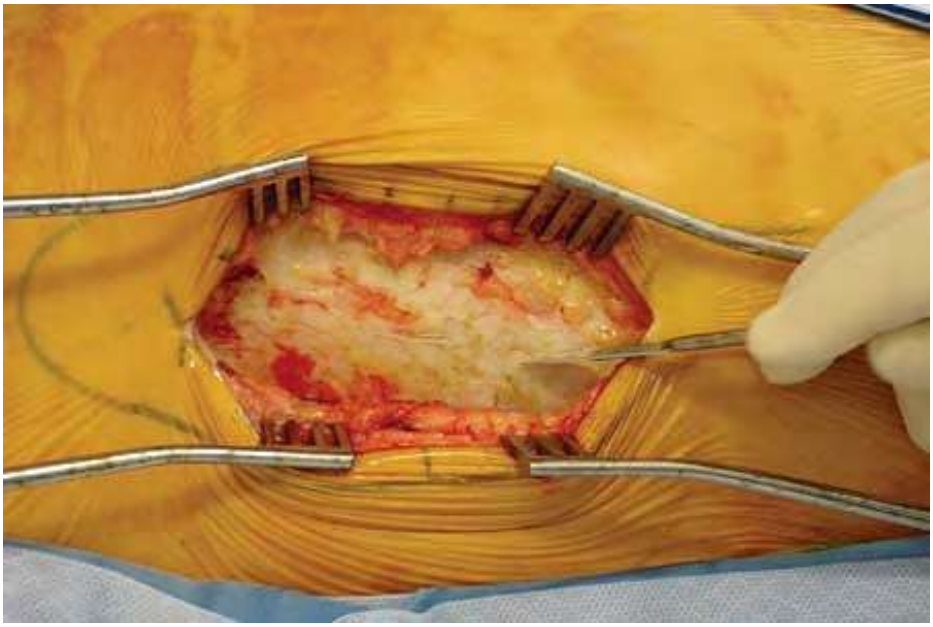

FIGURE 25-2 Iliotibial band is incised in line with its fibers.

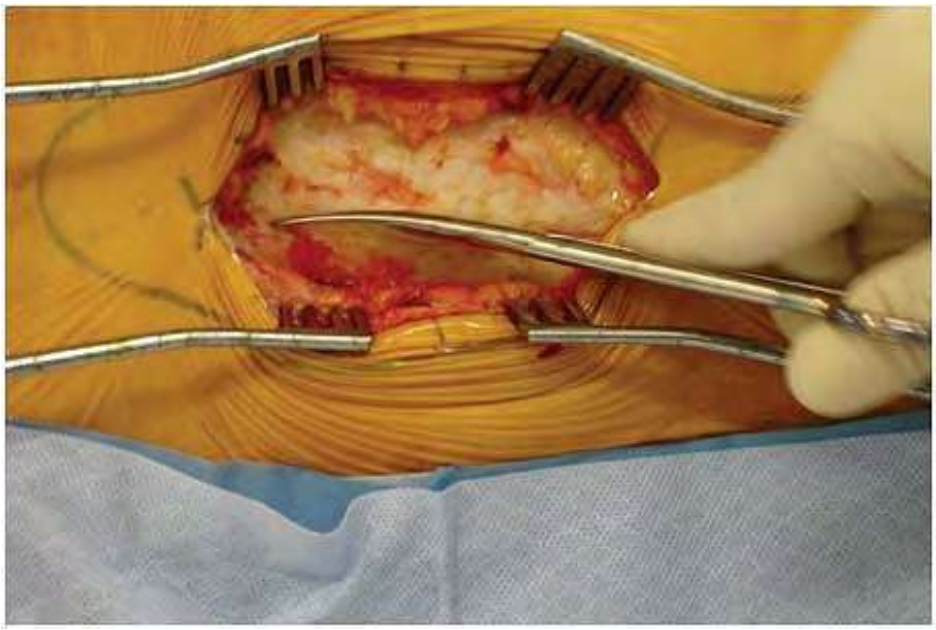

FIGURE 25-3 Iliotibial band is divided proximal to distal with curved Mayo scissors.

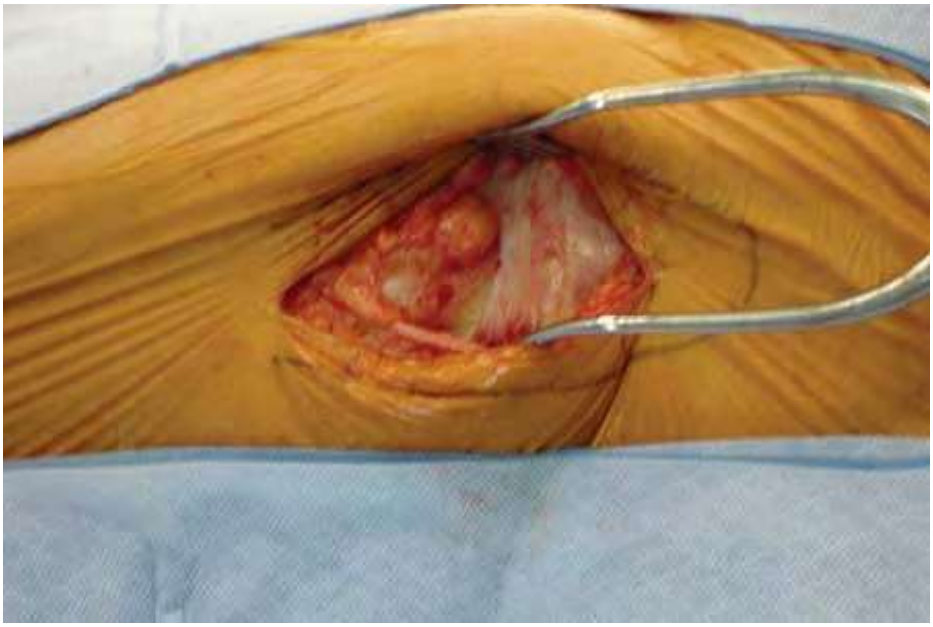

FIGURE 25-4 Fibrotic trochanteric bursa may need to be excised.

The Bennett retractors make for easy retraction and permit reduction and fixation. At the conclusion of the procedure, Bennett retractors are simply removed, and gravity allows the vastus to fall back into place, padding and directly covering the implant (Fig. 25-9). The most proximal origin of the vastus lateralis is then repaired back to its tendinous origin with a heavy braided nonabsorbable suture (Figs. 25-10 and 25-11). The posterior aspect of the vastus lateralis does not need to be repaired. The iliotibial band is repaired with interrupted nonabsorable suture, and the skin may be closed according to surgeon preference (Figs. 25-12 and 25-13).

This approach is most helpful for the placement of a dynamic hip screw, blade plate, or locking proximal femoral plate. It offers less scarring, as the muscle belly has not been interrupted. This should also be a more functional repair, as less damage presumably has been done to the vastus origin by dividing it and repairing it, rather than by directly insulting it with a midgastric approach.

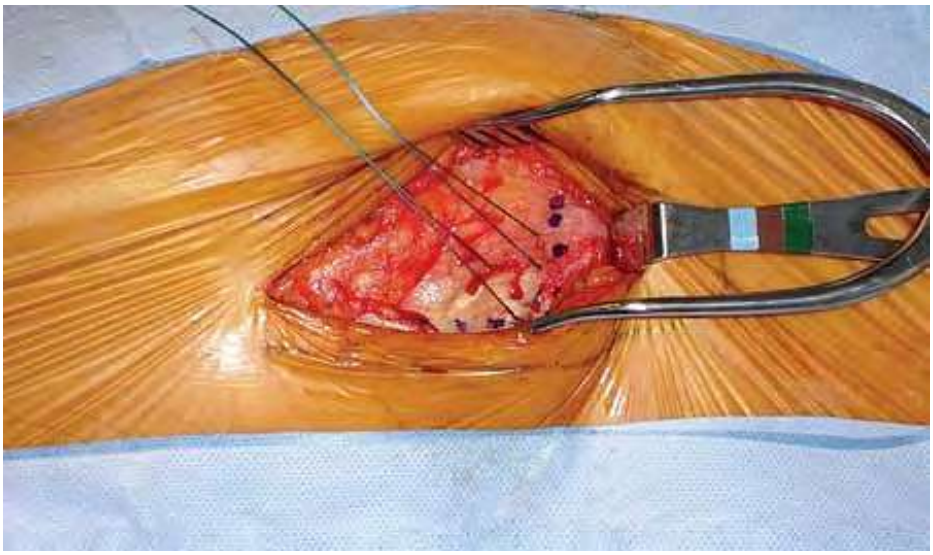

FIGURE 25-5 Proximal tendon of vastus lateralis is to be incised in a posteriorly based â€œJ.â€•

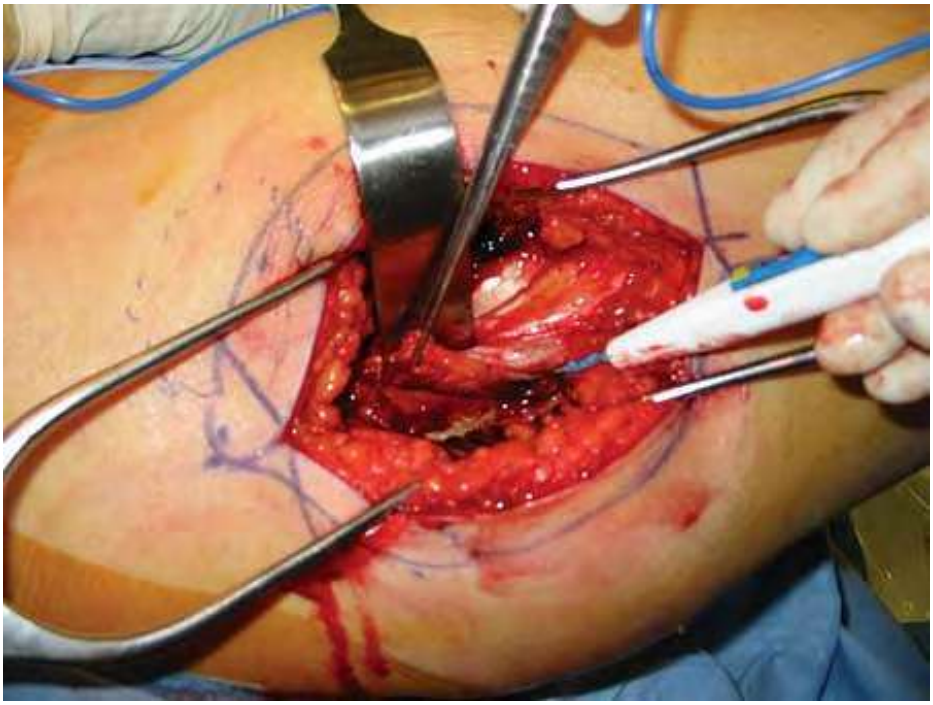

FIGURE 25-6 Vastus lateralis origin is elevated in along the â€œJ.â€•

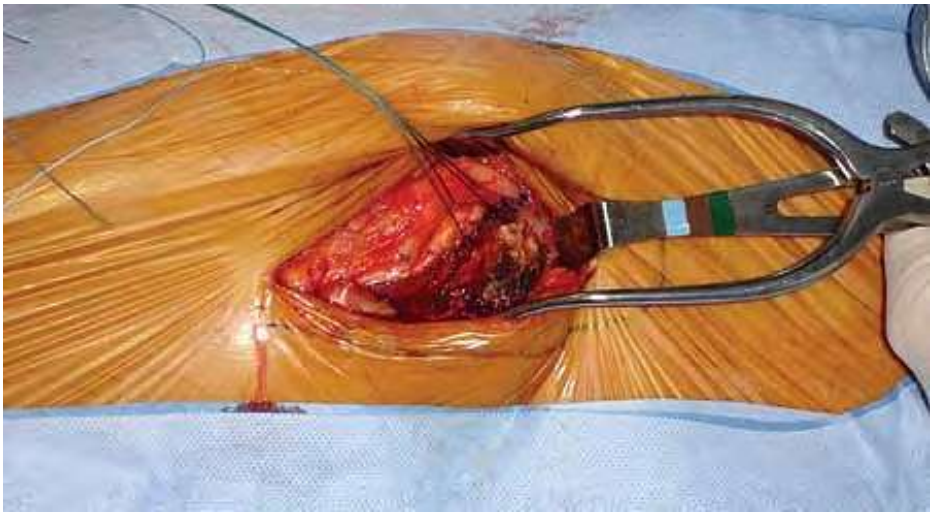

FIGURE 25-7 Lateralis is retracted anteriorly with tagging sutures.

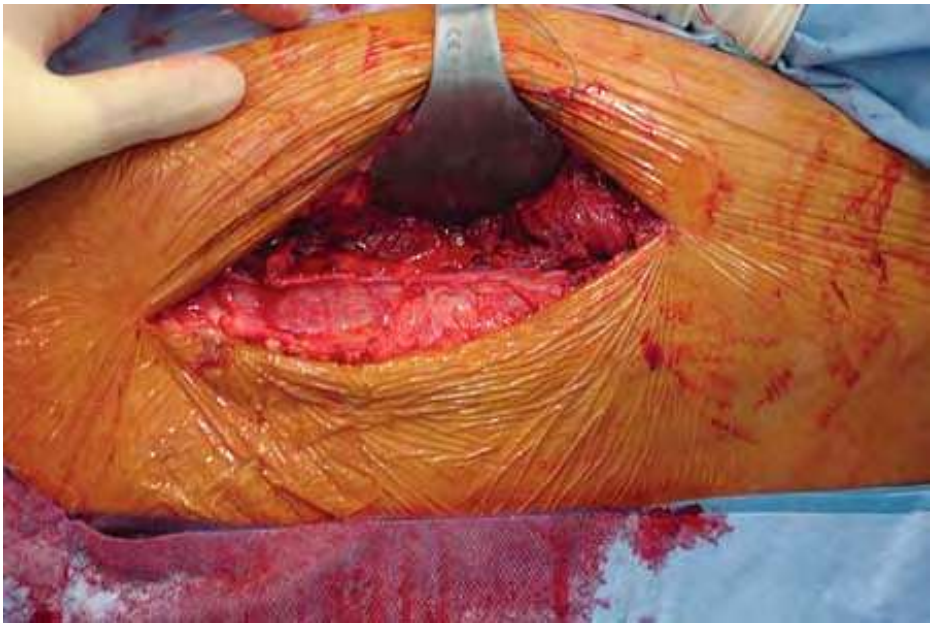

FIGURE 25-8 Bennett retractor is used to reflect the vastus anteriorly.

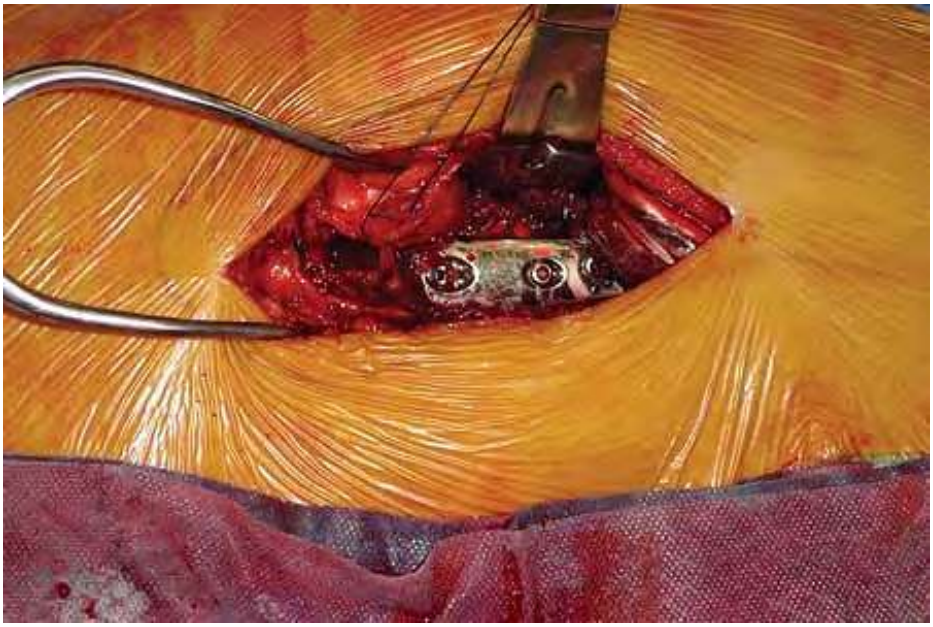

FIGURE 25-9 Osteosynthesis hardware, after placement, is covered by vastus.

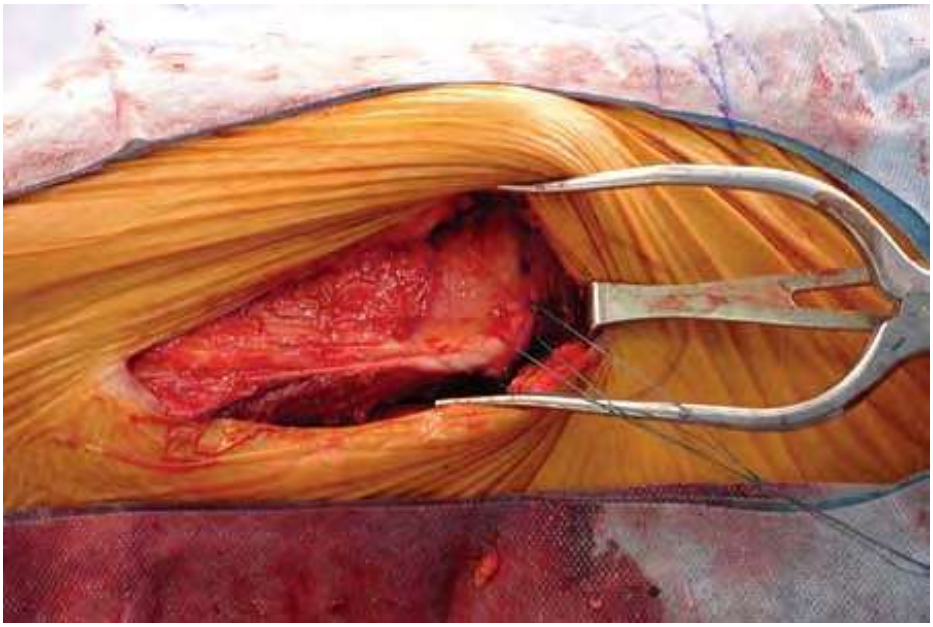

FIGURE 25-10 Vastus lateralis pulled back to its origin at the vastus ridge.

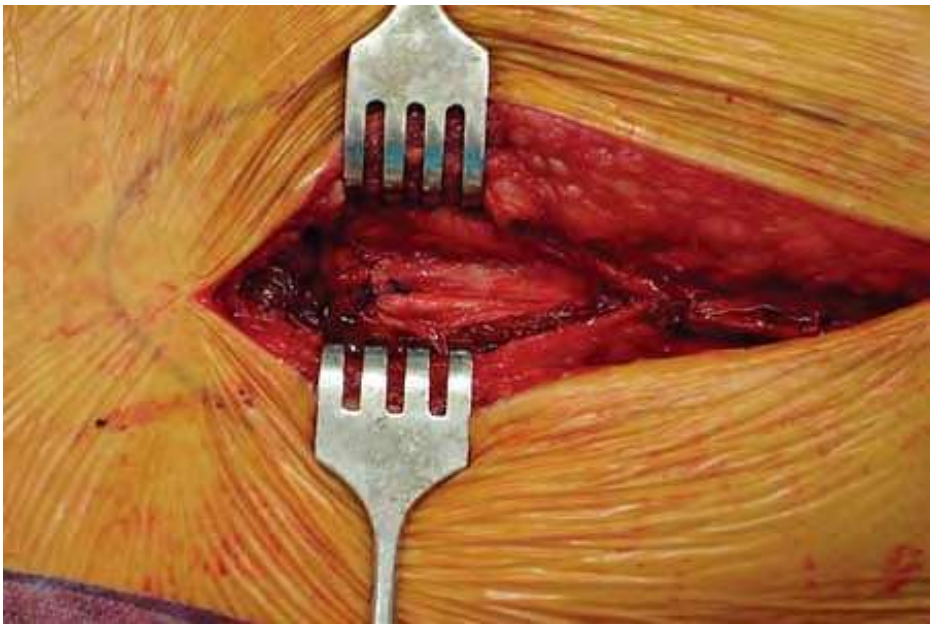

FIGURE 25-11 Through the closing iliotibial band, repair of the vastus origin may be visualized.

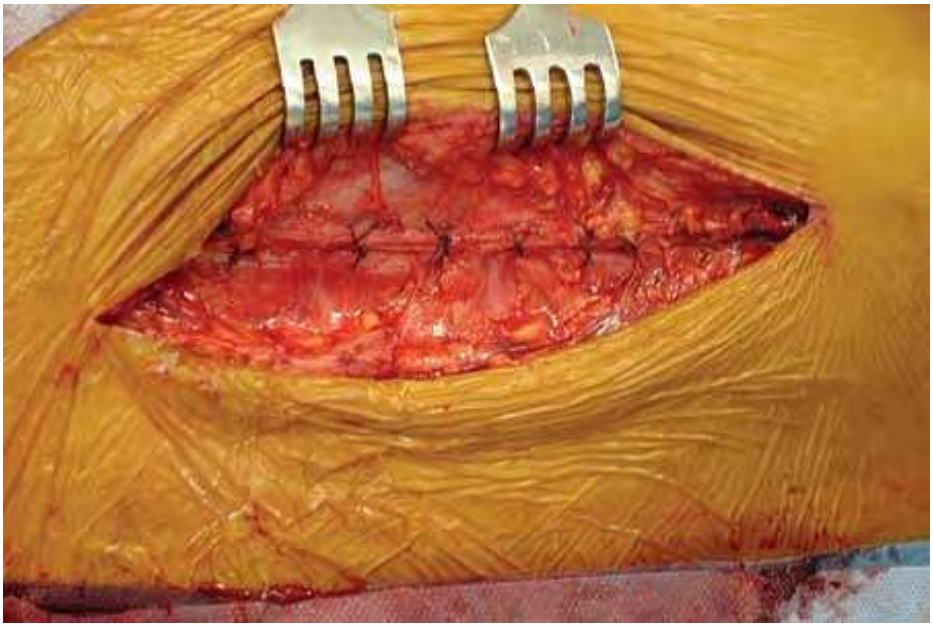

FIGURE 25-12 Iliotibial band is repaired with interrupted nonabsorbable suture.

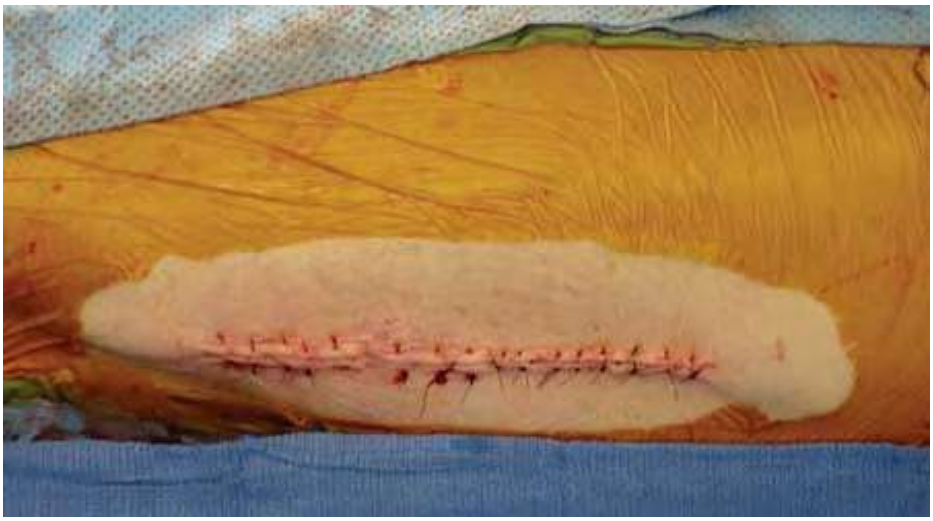

FIGURE 25-13 Skin closure at completion of intertrochanteric fracture osteosynthesis.

Postoperative Management

In the elderly population with intertrochanteric hip fractures, patients are generally allowed to bear weight as tolerated to facilitate mobilization (8). The first 2 weeks postoperatively focus on range of motion of the hip and knee, and gait training with assistive devices as necessary. At 2 weeks postoperatively, strengthening of the abductors, hip flexors, and quadriceps is added to the regimen.

In younger patients, or fractures with a subtrochanteric component, weight bearing may be protected initially, and advanced as desirable by the treating surgeon.

Complications

Complications of operations through this approach are generally related to the osteosynthesis, and these may be avoided somewhat with careful attention to achieving reduction and adherence to correct placement of implants.

Approach to the Distal Femur and Knee

Indications/Contraindications

Multiple variations on parapatellar and lateral parapatellar approaches have been described. Similar to the muscle sparing approach described for the proximal femur, the â€œswashbucklerâ€• approach described by Starr et al at Parkland Hospital (14) offers excellent visualization of the joint for anatomic fixation of interarticular fracture while preserving the vastus and quadriceps with no intramuscular dissection.

This approach offers the advantage of not interfering with or dividing the extensor mechanism while gaining adequate visualization of the joint space. With proximal and distal retraction, the intervening metaphyseal soft tissues may be largely undisturbed to enhance healing of comminuted metaphyseal bone.

Surgery

Patient Positioning

The patient is positioned supine. The lower extremity is prepared and draped free. A bump may be used under the knee to facilitate exposure and/or reduction.

Technique

The skin incision starts at the insertion of the patellar ligament and extends directly cephalad to the apex of the patella. It then extends lateral and proximal toward the lateral intermuscular septum (Fig. 25-14). A full thickness skin flap is elevated off the extensor fascia. The lateral margin of the patella and patellar ligament are identified distally. Proximally, the lateral margin of the vastus lateralis is identified.

The knee capsule is incised immediately along the lateral aspect of the patellar ligament (Fig. 25-15). Care must be taken not to cut the anterior horn of the lateral meniscus. Dissection is taken directly along the lateral margin of the patella, leaving a small cuff of tissue for eventual repair. At the superolateral shoulder of the patella, the deep dissection extends laterally toward the lateral intermuscular septum. The vastus lateralis is thus spared and is elevated anteriorly off the lateral intermuscular septum. As dissection nears the femur, perforating branches of the profunda femoris artery must be identified and ligated. These will be located every 3 to 4 cm approximately 1 cm lateral to the femoral cortex.

Right angle retractors may be placed at the level of the quadriceps tendon for medial excursion of the quadriceps and patella (Fig. 25-16). Bennett retractors may be placed over the femoral diaphysis. These

retractors in combination will allow adequate visualization of the femur for reconstruction. If possible, no retractors at all should be placed medially at the level of the metaphysis, as most techniques demand preservation of the medial soft tissues so as to not interfere with healing of metaphyseal comminution. Adequate visualization may easily be achieved with retractors placed proximal and distal to this area.

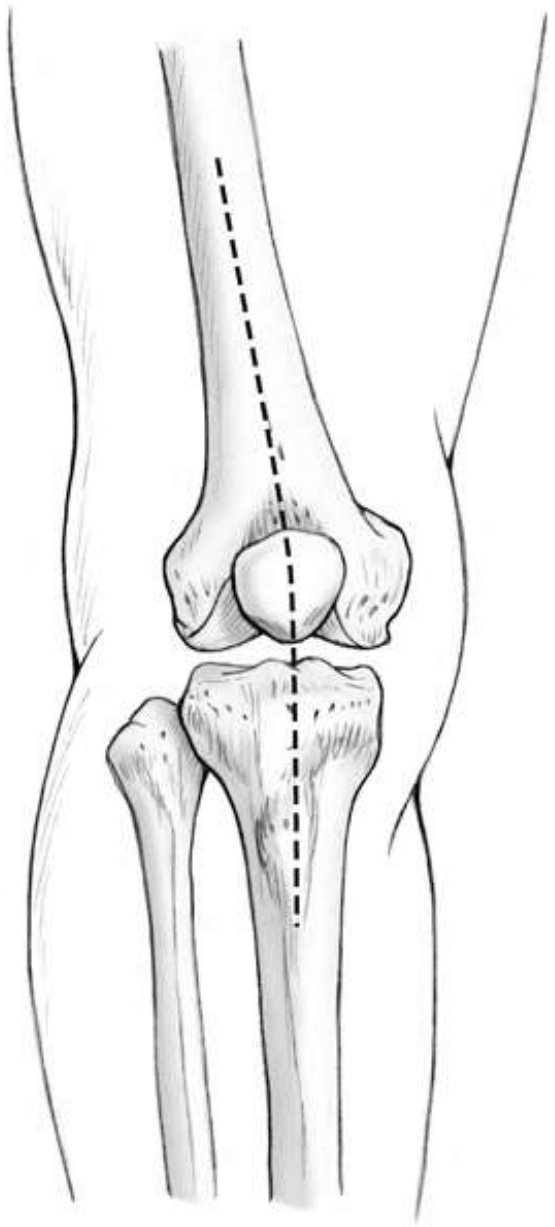

FIGURE 25-14 The skin incision for the â€œswashbucklerâ€• approach to the distal femur. (Redrawn from Starr AJ, Jones AL, Reinert CM. The â€œswashbucklerâ€•: a modified anterior approach for fractures of the distal femur. *J Orthop Trauma.* 1999;13(2):138â€“140. With permission.)

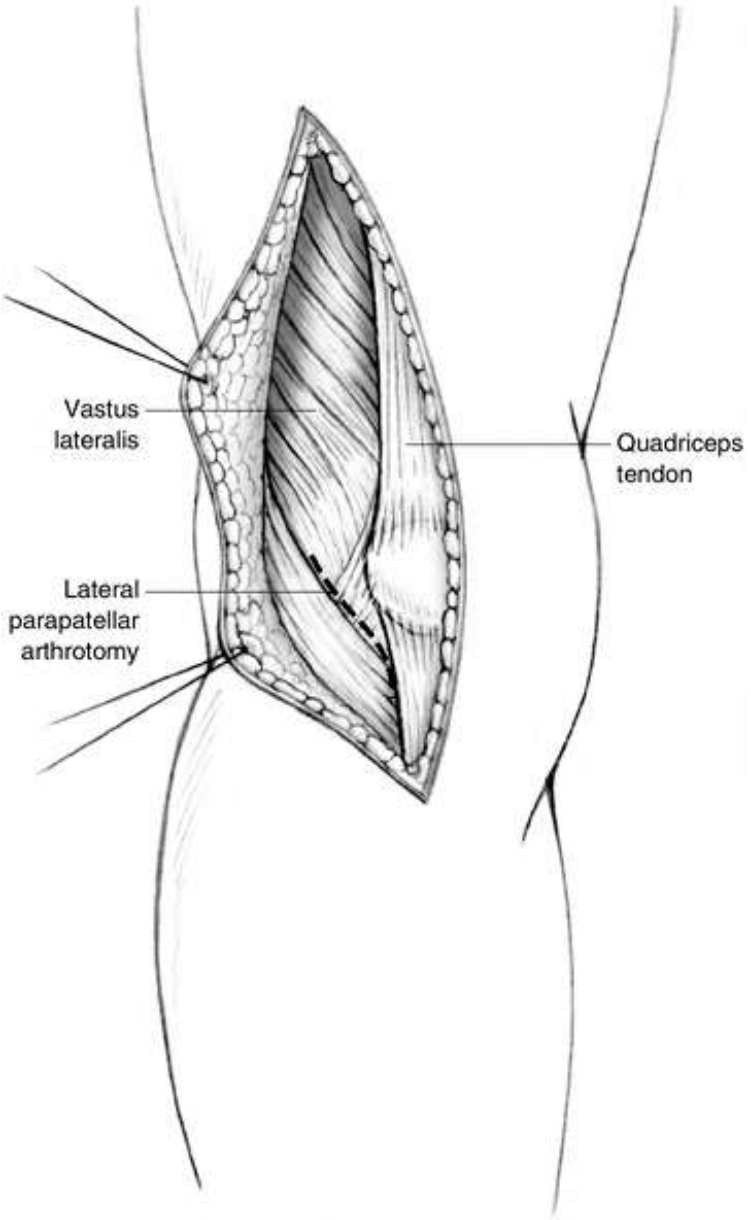

FIGURE 25-15 The lateral parapatellar arthrotomy extends to the lateral margin of the vastus lateralis. (Redrawn from Starr AJ, Jones AL, Reinert CM. The â€œswashbucklerâ€•: a modified anterior approach for fractures of the distal femur. *J Orthop Trauma.* 1999;13(2):138â€“140. With permission.)

For full visualization of the femoral condyles, the patella may be everted. This may

be facilitated by elevating a small proximal portion of the most lateral insertion of the patellar ligament off the tibia. Osteosynthesis is then performed.

Closure is achieved with repair of the lateral parapatellar retinaculum and capsule lateral to the patellar ligament. A drain may be placed in the knee and run along the lateral intermuscular septum, exiting the skin proximally. Subcutaneous tissue and skin are repaired according to surgeon preference.

Postoperative Management

Assuming stable osteosynthesis has been achieved, immediate range of motion is initiated. A physical therapist may perform passive supervised range of motion, 0 to 90 degrees, with no strengthening and no active quadriceps. Alternatively, a continuous passive motion machine may be used. Quadriceps strengthening begins at 6 weeks postoperatively, and weight bearing is advanced according to surgeon judgment.

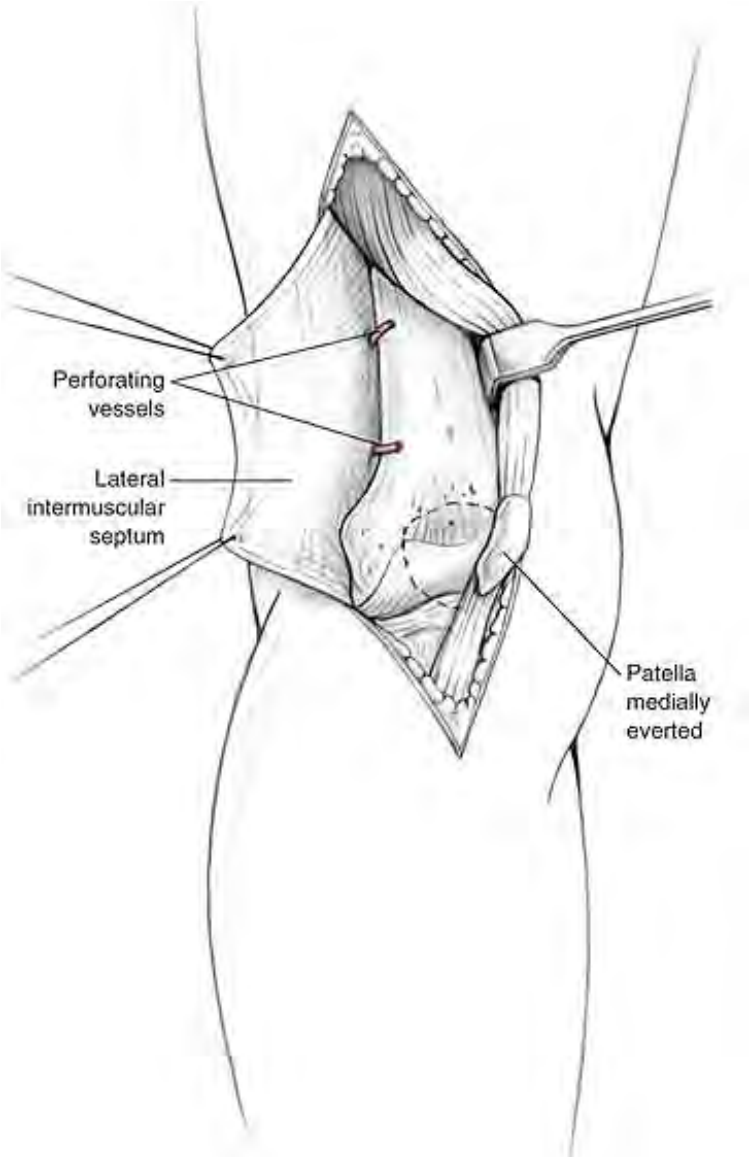

FIGURE 25-16 With quad retracted and patella everted, visualization of the distal femur is excellent. Starr AJ, Jones AL, Reinert CM. The â€œswashbucklerâ€•: a modified anterior approach for fractures of the distal femur. *J Orthop Trauma.* 1999;13(2):138â€“140. With permission.)

Complications

Potential complications of complex distal femur fractures include infection,

nonunion, and arthrofibrosis. Judet quadricepsplasty for loss of knee motion will be addressed later in this chapter.

Surgical Exposure of Open Femoral Fractures

Indications/Contraindications

Given the wide girth of soft tissues around the femur, an open fracture by definition implies a high-energy injury. There are compelling data to support early irrigation and debridement of open fractures for the prevention of infection (3,4). Small lacerations or puncture wounds are often the tell-tail signs of an open fracture and should not be overlooked; these relatively small surface injuries often disguise a significant deeper soft tissue insult (13). Open fractures of the distal femur in particular will present with a very small opening of the skin where the femoral shaft has pistoned distally to penetrate the vastus medialis obliquus or the vastus lateralis and the skin (Figs. 25-17 and 25-18).

Frequently, the treatment of open intra-articular distal femur fractures will necessitate an aggressive irrigation and debridement at the time of presentation and transarticular spanning external fixation for definitive stabilization. Even in the hemodynamically unstable trauma patient, early stabilization of the femur is mandatory. Prolonged maintenance of a patient in Hare traction is not advisable given the attendant complications of skin breakdown and sciatic nerve palsy, and so spanning external fixation is recommended over prolonged traction.

The initial management of these fractures should not include the placement of a pulsatile irrigation catheter through the small traumatic wound. This may succeed only in further embedding debris

in an open fracture site and will certainly not provide an adequate debridement of the injured soft tissues that would ultimately be responsible for the development of a deep infection. Adequate debridement of an open femoral fracture demands wound extension (5).

FIGURE 25-17 Skin staples demonstrate the length of the inadequate incision used to debride this open intracondylar distal femur fracture.

FIGURE 25-18 This inadequate index debridement leads to a long and complicated course of treatment, including a deep infection and the placement of antibiotic beads.

Preoperative Planning

Initial debridement entails some insight by the initially treating surgeon into the definitive fixation plan. Frequently, the entire approach for the eventual repair of an open femur fracture may be performed at the time of debridement, particularly if this will include plating of a distal femur fracture. Open fractures of the femoral diaphysis may still be treated quite successfully with intramedullary nailing after wound extension and debridement (13). Regardless of eventual fixation plan, proximal and distal extension of soft tissue wounds is always necessary.

Surgery

Extension of the skin laceration is performed both proximally and distally down to the investing facsia of the thigh, which is always torn due to the pre-existing trauma. Fracture exposure will often require intramuscular dissection. Specific landmarks or structures to avoid will vary depending on location of the open wounds. Extensive dissection through muscle may be required, but most of this muscle has been traumatized and may not be viable. It is exactly this tissue that must be removed for the adequate debridement of the open femoral fracture and the prevention of a potential deep infection.

If, following aggressive irrigation and debridement, it is determined that a transarticular spanning external fixation is going to be used for definitive or temporary stabilization, consideration must be given to the underlying soft tissues. In applying joint spanning external fixation, one should be careful to place the femoral pins proximal enough that the eventual internal fixation is not communicating with the pin sites if at all possible. Femoral pins may be introduced directly anterior through the quadriceps, anterolateral, or directly lateral. Tibial pins are generally introduced anteromedially. Spanning external fixation may be maintained for several weeks while waiting for physiologic or soft tissue stabilization. Nowotarski et al. demonstrated in non-spanning external fixation that patients

could safely undergo one-stage conversion to internal fixation without significant infectious problems (11). However, waiting much longer than 3 weeks may create problems in having to take down interval callus formation and makes the articular reconstruction more difficult.

Postoperative Management

In an effort to stabilize the soft tissue envelope, initiation of range of motion may be delayed after open fracture of the distal femur. If an open distal femur fracture has been spanned with an external fixator, subsequent debridements may be necessary, and the exact timing of definitive internal fixation is at the surgeon's discretion.

If a thorough debridement and immediate intramedullary nailing of an open

diaphyseal fracture has been completed, range of motion and weight bearing may be initiated right away.

Complications

The prompt and aggressive debridement of open fractures is an attempt to avoid the most obvious complication, osteomyelitis. Further complications from open fractures include knee stiffness and non-union. In cases with a segmental bone loss, or a high degree of metaphyseal comminution, patients may be informed at the time of definitive fixation that an elective bone grafting procedure will be performed 4 to 6 weeks later. The author prefers to use proximal tibial cancellous autograft, as it is biologically active, is a low morbidity harvest site when compared to iliac crest, and limits the affected area of the patient to a zone already involved in the injury. For large segmental defects, a cement spacer may be placed at the time of index fixation, with the plan to exploit the â€œbiologic membraneâ€• formed around this spacer to promote healing at the time of subsequent bone grafting (12).

Judet Quadricepsplasty

Indications/Contraindications

Frequently, patients who have suffered comminuted intra-articular injuries of the distal femur or supracondylar femur will suffer from poor range of motion secondary to post-traumatic scarring and contracture. Various methods have been proposed for the treatment of this loss of flexion at the knee. Arthroscopic lysis, closed manipulation, and quadriceps tendon lengthening have all been advocated (15). For particularly stubborn cases of post-traumatic arthrofibrosis and quadriceps contracture leading to loss of knee flexion, the author finds the Judet quadricepsplasty to be the most effective way to regain functional motion of the knee (1,2,6,7,10,16).

This approach recognizes that, in addition to the skeletal injury apparent on the radiographs, there was great soft tissue damage sustained at the time of distal femoral fracture. It is this damage to the rectus and vastus musculature that is at least in part responsible for the subsequent loss of range of motion. Arthrofibrosis and capsular contracture also contribute to post-traumatic stiffness and are

addressed by this operation.

Judet attributed the stiffness of the knee to four principle anatomic reasons. First was adhesion of the subquadricipital pouch with tethering of the patella. Second was retraction of the parapatellar fibrous elements and intra-articular adhesions. Third was adherence of the vastus intermedius muscle of the femur. Fourth was degenerative fibrosis of the quadriceps. The surgical procedure is based on a sequential release of these various parts of the contracture. Range of motion is to be assessed after each stage, and the operation proceeds until the desired results are achieved or maximum range of motion is achieved after completion of all steps.

Preoperative Planning

An epidural catheter is placed preoperatively. General and epidural anesthetic are used intraoperatively, but the regional anesthesia will help with patient compliance and tolerance with postoperative continuous passive motion.

Surgery

Range of motion is assessed on administration of the general anesthetic (Fig. 25-19). Two incisions are generally used (Fig. 25-20). The first is a medial parapatellar approach. This allows access to the medial aspect of the patellar tendon, the suprapatellar pouch medially, and the medial gutter. The medial retinaculum is released, and adhesions of the medial joint and suprapatellar pouch are excised.

FIGURE 25-19 Range of motion is assessed preoperatively.

FIGURE 25-20 Two incisions planned preoperatively for Judet quadricepsplasty.

The second incision often follows the incision made from the operative repair of the fracture (Fig. 25-21). This will usually follow the lateral aspect of the patella ligament through the lateral patellar retinaculum and proximal up along the lateral intramuscular septum to the level of the greater trochanter. In the distal segment the parapatellar and lateral retinacular tissues are released and intra-articular adhesions are excised. The patella is freed until it may be easily lifted off the femoral condyles (Fig. 25-22). Scar tissue along the medial and lateral aspects of the patellar ligament, particularly distally, would likewise need to be excised. At this point, range of motion is checked, and the quadriceps may be manipulated at each subsequent step. The vastus lateralis is elevated from the linea aspera, and perforating vessels are ligated as they are encountered. The vastus intermedius is then lifted extraperiosteally from the front and side of the femur (Fig. 25-23). Much of the intermedius will be replaced by scar and fibrotic tissue, and much of this may be excised without significant functional impact (Fig. 25-24). Frequently, one will find heterotopic bone formation within the muscle bellies, and this may be excised along with fibrotic muscle tissue in an effort to regain length. Once again, the knee is manipulated and range assessed.

Next, the vastus lateralis is released from its femoral origin superolaterally, and the rectus femoris may be released from its origin on the anterior capsule (Figs. 25-25 and 25-26). This allows distal excursion of much of the quadriceps mechanism in an effort to regain range of motion at the knee. The knee is manipulated for a final time and final range of motion is measured (Fig. 25-27). Only the skin is closed, and a drain is used (Fig. 25-28). Given the somewhat sanguinous nature of the dissection along the well-perfused margins of scar tissue, postoperative hematoma must be guarded against, and drain use is encouraged.

FIGURE 25-21 Lateral incision planned as extension (*dotted line*) off existing scar (*solid line*).

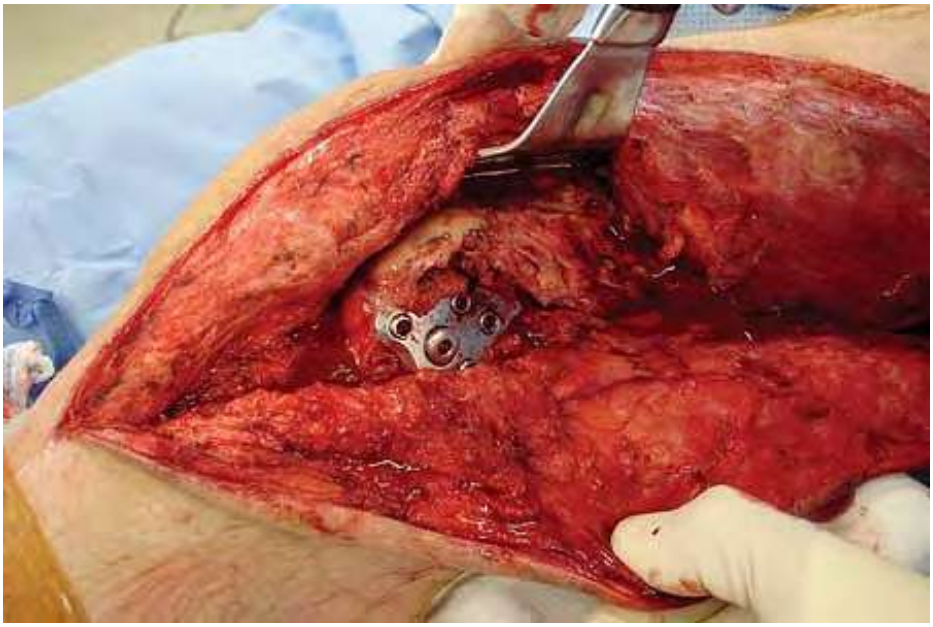

FIGURE 25-22 Supracondylar zone shows dense fibrotic scar tissue.

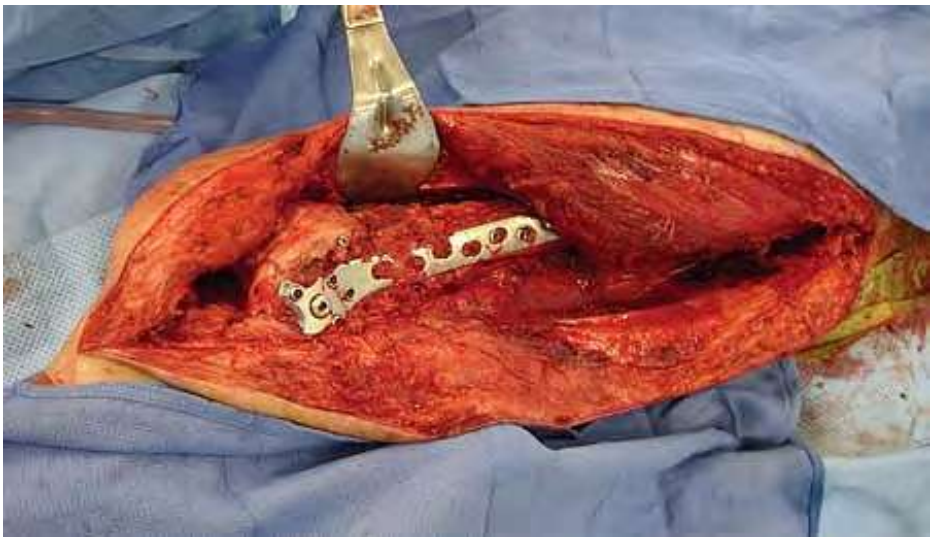

FIGURE 25-23 Scar tissue release mobilizes the extensor mechanism off the anterior femur.

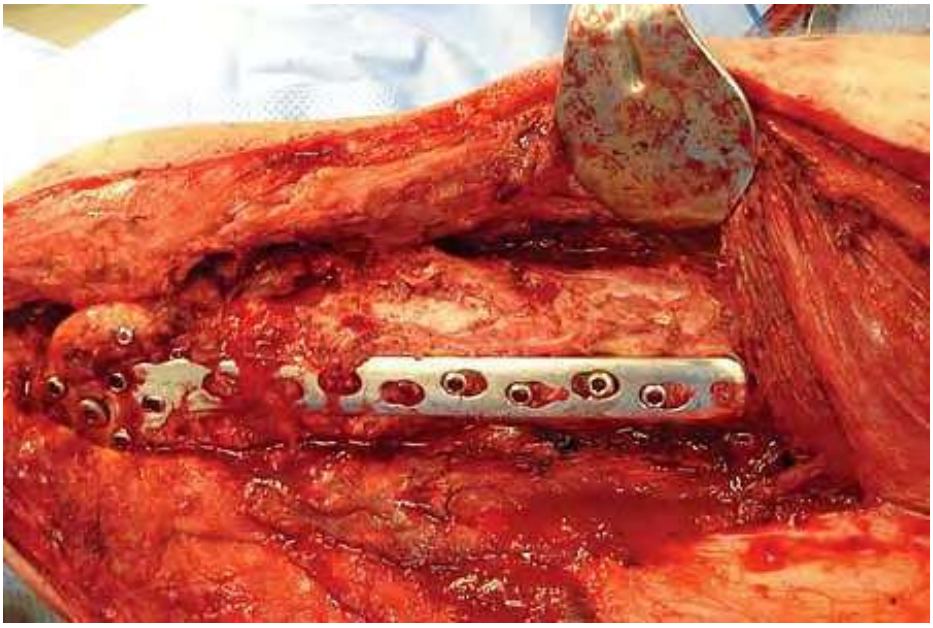

FIGURE 25-24 Heterotopic bone may scar the quadriceps to the femur and must be excised.

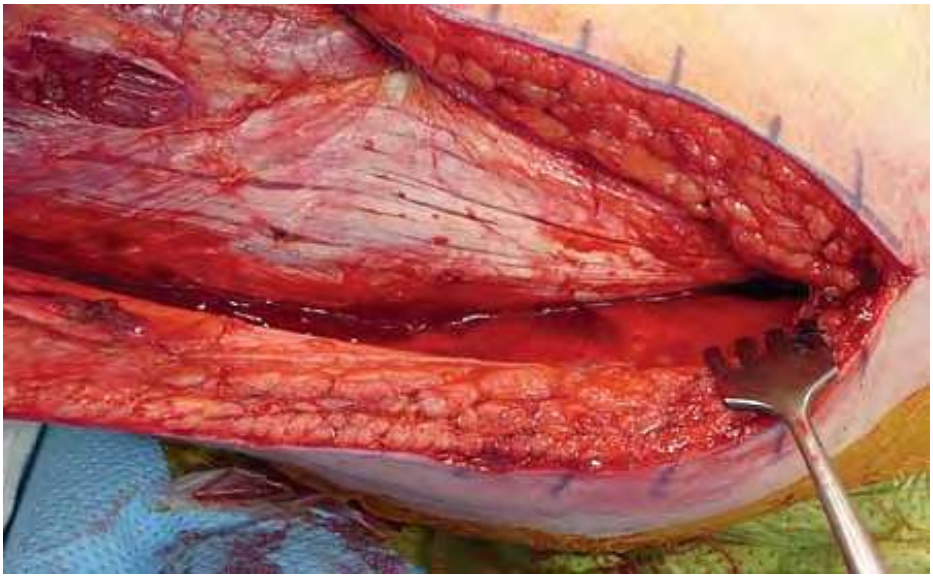

FIGURE 25-25 Disection may be extended up to the level of the vastus lateralis origin and the rectus origin anteriorly.

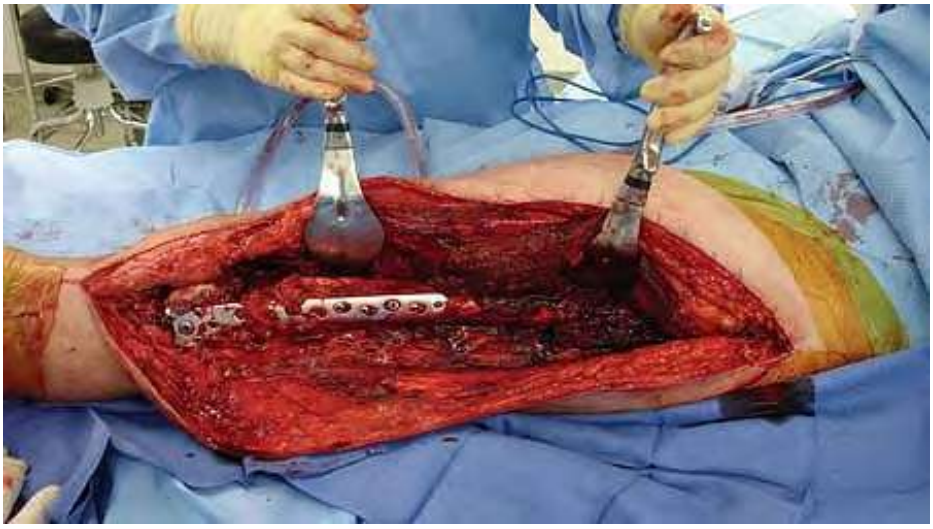

FIGURE 25-26 With release of the rectus proximally, ultimate excursion of the extensor mechanism may be achieved.

FIGURE 25-27 Ultimate range of motion should be assessed before and after closure.

FIGURE 25-28 Final skin closure of the lateral incision.

Postoperative Management

The knee may be immobilized in a position of flexion so that, on awakening from anesthetic, the patient may have some idea of what the range of motion may be achieved (Fig. 25-29). Epidural anesthetic is generally used, and continuous passive motion is initiated after the recovery room. The drain is removed in approximately 48 hours. Due to the bleeding which may be encountered in dissection of scar and muscle, we have found that blood transfusion may be necessary before discharge.

No postoperative immobilization is typically used at the time of discharge, and patients may need crutches for ambulation, as they frequently will have a functional extensor lag postoperatively. This extensor lag resolves over time.

Complications/Results

The most common immediate complication following this procedure is hematoma formation. This complication can be minimized with intraoperative hemostasis and with the routine use of closed

suction drains in all patients. Secondary complications include extensor lag, which will resolve in most cases with postoperative therapy. Recurrent stiffness and arthrofibrosis can also complicate postoperative results; however, the results of this operation in the literature have been encouraging (1,2,6,7,10,16), with total gains in range of motion between 55 to 69 degrees in the two largest series.

FIGURE 25-29 Knee is immobilized in flexion postoperatively, and continuous passive motion is initiated after the recovery room.

Summary

The soft tissue around the femur and thigh segment is adequate enough to sustain extensive trauma without requiring free tissue transfer. Some of the techniques listed in this chapter may facilitate the correct handling of the soft tissues, help to prevent infection in the setting of open fracture, or help patients to regain motion after a comminuted supracondylar injury.

References

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 26

Management of Soft Tissue Defects Surrounding the Knee and Tibia: The Gastrocnemius Muscle Flap

Steven Myerthall

David J. Jacofsky

Steven L. Moran

Defects surrounding the knee and patellar region are common and may result from trauma, tumor, or infection. One of the most important goals in the management of open wounds surrounding the knee is stable soft tissue coverage. Soft tissue coverage with muscle can seal the joint and bone from ongoing contamination and help to revascularize underlying structures, while preventing the development of late infection and nonunion. Options for soft tissue coverage are determined by the size of the wound, location, and surrounding zone of injury. The gastrocnemius muscle flap has historically been used to cover defects surrounding the knee, lower thigh, and patellar region. Its consistent vascular anatomy and superficial location have made it a workhorse for coverage of defects in this area.

The gastrocnemius is the most superficial muscle of the posterior calf. The muscle has two heads arising from the medial and lateral condyles at the femur, and the adjacent capsule of the knee; these muscles then insert into the calcaneal tendon (Fig. 26-1).

The two muscle heads, medial and lateral, unite at the level of the fibular head. Proceeding distally the muscles join with the tendon of the soleus at the mid-leg to form the Achilles tendon. The medial sural artery is the dominant vascular pedicle for the medial gastrocnemius muscle, whereas the lateral sural artery supplies the lateral head. The origin of the arteries is approximately 4 cm above the head of the fibula and both originate from the popliteal artery. Innervation to the muscles originates from the tibial nerve and lies posterior to the vascular pedicle as it enters the muscle.

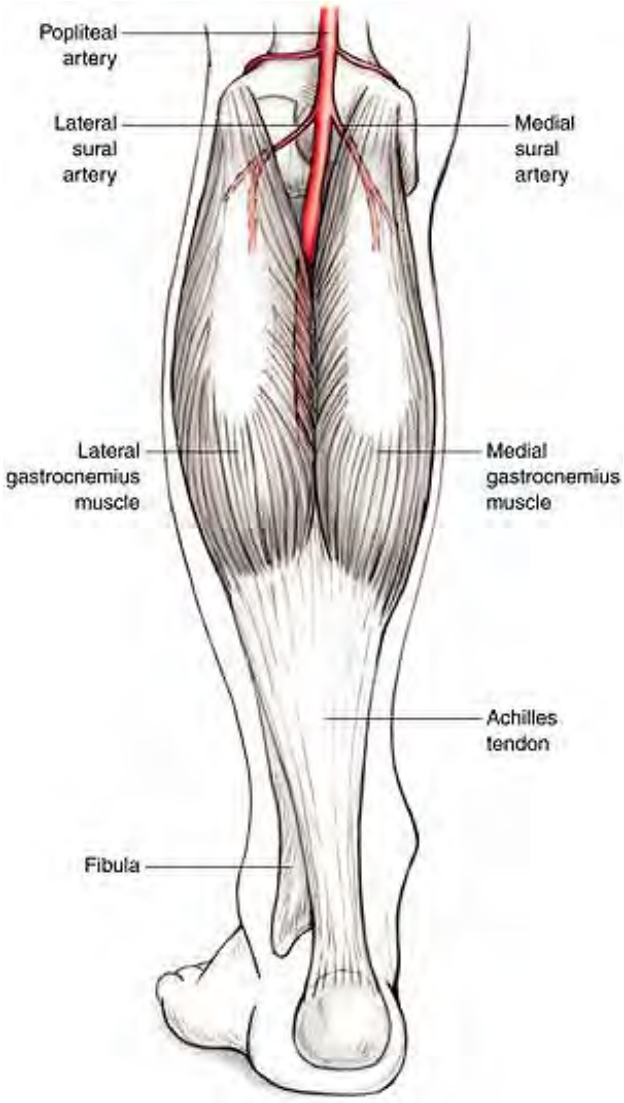

FIGURE 26-1 The medial and lateral gastrocnemius muscles as seen from the posterior approach.

Indications/Contraindications

Both the medial and lateral gastrocnemius muscles may be used independently or in conjunction with each other. The medial gastrocnemius will cover the inferior thigh, knee, and proximal tibia. The medial head of the muscle is used most frequently as a proximally based flap due to its larger size in comparison to the lateral head of the muscle. The lateral head may also be used alone or in combination with the medial head for coverage of large tibial defects or for lateral distal thigh wounds. For defects at the level of the mid-portion of the tibia, the gastrocnemius muscle may not provide adequate coverage and the soleus muscle is preferred for coverage of middle third defects. For defects involving the distal third of the tibia and ankle, free tissue transfer is usually required. The tendinous inferior margin of the gastrocnemius muscle may be used to augment the repair of an injured suprapatellar tendon.

Contraindications to the use of the gastrocnemius muscle flap include active infection and/or significant disruption of the soft tissue and/or vascular pedicle. Additional contraindications for flap use include any procedure or injury which may have traumatized or injured the sural artery, such as a previous repair of a popliteal arterial laceration or repair of popliteal aneurysm. Occasionally, severe compartment syndromes may render the muscle fibriotic and useable for transfer, but due to its superficial location and high take off of the sural artery, this is unusual. Direct trauma to the posterior calf may result in muscle destruction and should be ruled out prior to proceeding with flap elevation. Radiation to the knee following tumor extirpation may also compromise the vascular pedicle,

increasing the chance of total or partial muscle necrosis following transfer. The muscle should not be used if the soleus and contralateral gastrocnemius muscle are no longer functional, as this will create postoperative difficulty with plantar flexion. A history of a recent deep venous thrombosis within the involved extremity is a relative contraindication for the use of the flap, although we have used the flap successfully in

such situations.

Distally based gastrocnemius muscle flaps have been described for coverage of middle third and lower leg defects. Vascular supply to the muscle in these cases is supplied through crossing anastomotic arterial connections between the medial and lateral gastrocnemius muscle bellies extending across the midline raphe. Use of the flap in this manner is not common, as more reliable methods of middle third and lower leg coverage are available and include the soleus muscle flap, the sural artery flap, and free flap coverage.

Preoperative Planning

The preoperative management of severe open fractures involves stabilization of the patient, tetanus prophylaxis, and broad-spectrum antibiotics. Trauma often compromises the availability of the skin and soft tissue surrounding the knee. Nonviable tissue requires aggressive debridement. Serial debridements may be needed if tissue viability is uncertain. Once the wound bed is clean, flap coverage may proceed, ideally within 72 to 96 hours following initial injury. Preoperative physical examination should focus on determining the function of the remaining muscles of the lower leg to determine if sacrifice of the gastrocnemius could result in a significant loss of plantar flexor following muscle transfer. A vascular examination should also be performed to assess for the patency of the popliteal artery. In those patients with a history of peripheral vascular disease or a history of popliteal arterial trauma, an angiogram can confirm patency of the sural artery prior to surgery.

Surgery

Patient Positioning

Operating room setup is standard for lower extremity coverage. A leg tourniquet may facilitate dissection, especially in cases of lateral gastrocnemius harvest where the perineal nerve must be identified and protected during flap elevation.

The muscle may be harvested with the patient in prone or supine position. If the patient is prone, a stockinette or mid-posterior S-shaped incision is made and significantly facilitates exposure of the muscle's origin posteriorly over the femoral condyle. The posterior approach is made with an incision starting 5 cm above the

popliteal crease and extending down to the distal end of the muscle belly. More commonly, however, for anterior traumatic defects surrounding the patellar and tibial region, the patient is positioned in the supine position with the leg internally or externally rotated to facilitate exposure of the medial or lateral heads, respectively. Lateral decubitus positioning is also an option for lateral gastrocnemius muscle elevation.

The patient may receive either general or spinal anesthesia. In additional, depending on associated injuries, the ipsilateral or contralateral thigh should be prepared for a skin graft donor site. After sterile preparation, the entire extremity is draped and fully exposed. Appropriate preoperative antibiotics should be given 30 to 60 minutes prior to incision.

Technique

Medial Gastrocnemius Muscle Flap

A longitudinal incision is made where the separation of the bellies of the gastrocnemius and soleus muscle can be palpated. This should parallel the medial border of the tibia and can be curved proximally in a posterior direction to facilitate exposure and dissection of the proximal portion of the muscle. The incision should extend from the level of the tibial plateau to 10 cm above the medial malleolus. If the flap is to be tunneled into position on the anteromedial aspect of the leg, a skin bridge of at least 7 cm must be maintained to prevent skin necrosis. For more anterior defects, or preexisting peripatellar defects, a curved incision can be extended both proximally and distally from the defect site to a line 4 cm medial to the edge of the tibia (Fig. 26-2A,B).

During initial subcutaneous dissection, the saphenous vein is identified and preserved (Fig. 26-2C). The underlying deep investing fascia of the leg is then opened to expose the plane between the

gastrocnemius and soleus muscles (Fig. 26-2D,E). In most patients, the plantaris tendon is located in this interval and should be left intact. This plane is easily opened with blunt finger dissection. The plane between the external fascia of the muscle and

the overlying skin should also be opened using blunt dissection. Large, perforating vessel can be identified running from the muscle to the overlying skin. These vessels serve as the main blood supply to the skin paddle of a gastrocnemius fasciocutaneous flap. These vessels should be ligated in the gastrocnemius muscle flap dissection. The

sural nerve and lesser saphenous vein can be identified running between the medial and lateral heads of the gastrocnemius muscle and should be preserved.

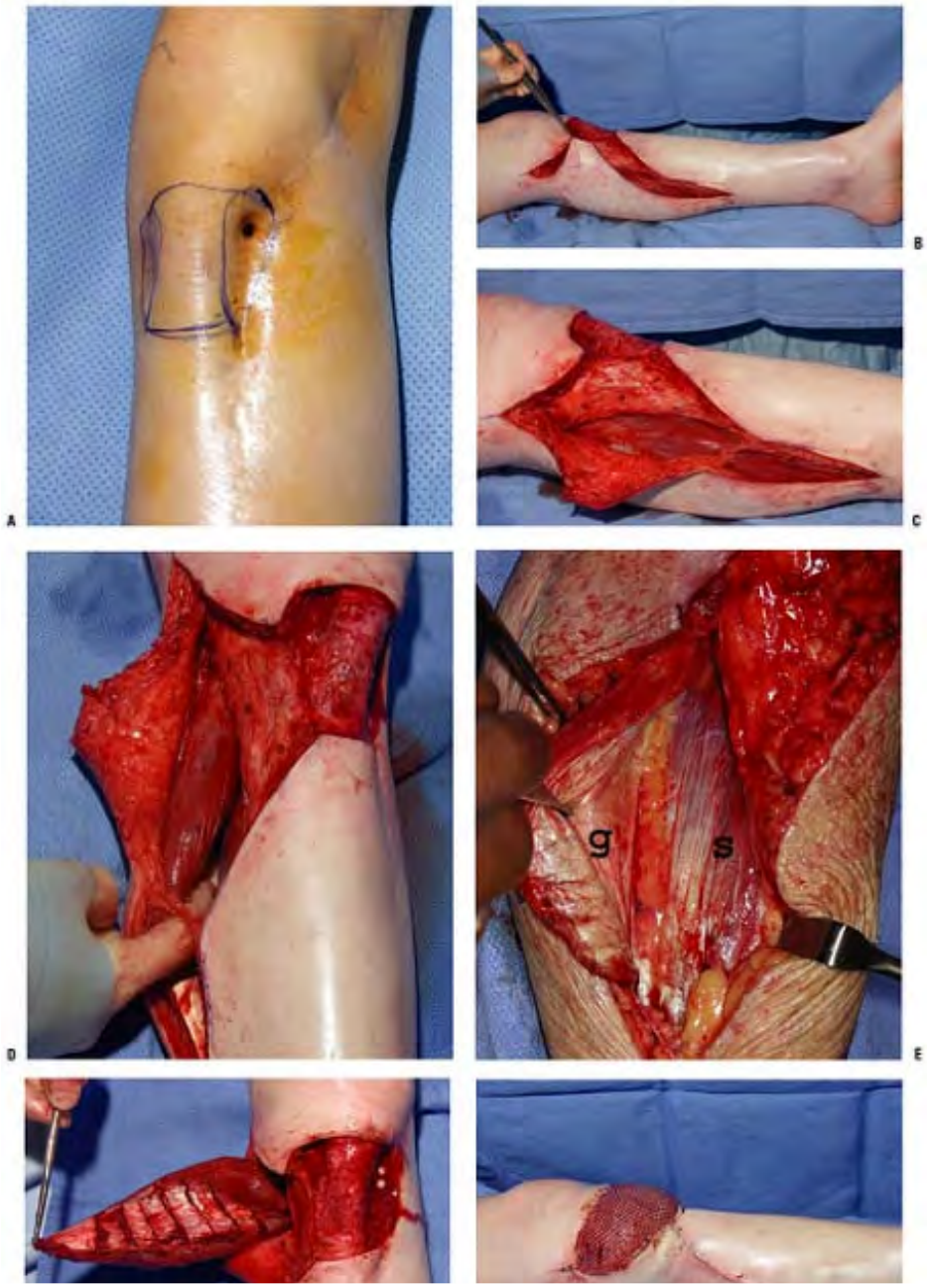

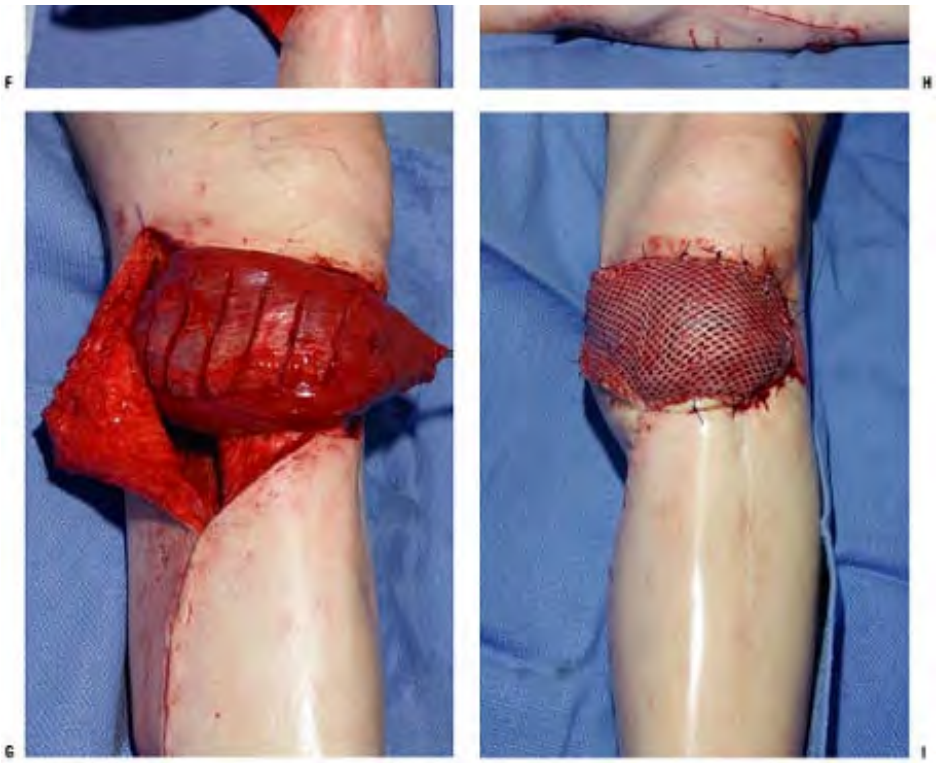

FIGURE 26-2 A: A 53-year-old patient with longstanding osteomyelitis of the tibia requiring debridement and soft tissue coverage. B: The skin surrounding the defect is excised. The bone is debrided and filled with antibiotic beads. The skin defect is incorporated into the skin incision for exposure of the gastrocnemius muscle. The skin incision is curved posterior approximately 4 cm beyond the medial border of the tibia. C: This incision gives excellent exposure of the posterior compartment. Note that the saphenous vein has been preserved at the inferior extent of the skin incision. D,E: The deep investing fascia of the leg is divided and the plane between the gastrocnemius and the soleus muscle may be identified using finger dissection. F,G: Once the muscle is mobilized, the anterior and posterior fascia may be transversely scored to allow for easier insetting. H,I: Once the muscle is inset, it is covered with a split thickness skin graft. The leg incision is closed over a suction drain. g, gastrocnemius; s, soleus.

Dissection is then continued medially between the plane of the gastrocnemius and the soleus until the midline raphe of the gastrocnemius muscle is identified. Often this

raphe will be identified by decasations of the muscle fibers distally. More proximally, the two heads may be more easily separated and identified if the raphe is not clearly visualized distally. Once the limits of the medial muscle have been clearly identified, the inferior margin of the muscle is divided from the Achilles tendon with a 2-cm cuff of tendon, which facilitates insetting of the muscle. This cuff of tendon is also very valuable for reconstruction of the terminal extensor mechanism of the quadriceps muscle.

The sural nerve will be identified within the gastrocnemius raphe and should be preserved. Using cautery, the raphe can be divided and the gastrocnemius can be separated from the Achilles tendon. At this point, the muscle may be pedicled and mobilized for coverage of the upper one-third of the lower leg. If the knee or distal thigh requires coverage, the origin of the muscle will require release. For release from the femoral condyle, posterior extension of the skin incision significantly facilitates visualization of the muscle's vascular pedicle. Skeletonization of the sural vessels from the deep surface of the muscle will also allow further mobilization of the muscle. The muscle can be transferred with either its deep or superficial surfaces exposed. Cross-hatching or removal of the fascia will allow for further mobility and advancement of the muscle without impairing blood supply (Fig. 26-2F,G).

The fascia is then inset into the defect using horizontal mattress sutures, pulling the muscle beneath the overlying skin on the far side of the defect. If there is excessive tension, sutures may be mattressed through the skin and tied over bolster dressings. Prior to inset, a 10-mm flat or round drain can be placed under the flap; a second drain should be placed at the donor site prior to closure.

A meshed skin graft should then be obtained and placed on the transposed muscle flap, followed by a bolster dressing. This dressing should include a nonadherent layer followed by a bolster with gentle compression maintained on the graft. A VAC sponge may also be applied to facilitate skin graft take to the underlying muscle (Fig. 26-2H,I). The remainder of the wound should be covered with sterile dressings and the leg placed in a knee immobilizer for support. The skin graft donor site may be dressed with a single-layer transparent dressing.

Lateral Gastrocnemius Muscle

The lateral gastrocnemius muscle will cover the inferior thigh, knee, and proximal

tibia. For harvesting of the lateral gastrocnemius the skin incision is made at the posterior midline of the lower calf extending to the lateral popliteal fossa. The proximal mark is placed 2 or 3 cm posterior to the fibula. Once again, an adequate skin bridge must be maintained between the traumatic defect and the incision to avoid skin bridge necrosis. Alternatively, the existing skin defect may be extended proximally and distally over the posterior margin of the fibula to allow for exposure of the muscle (Fig. 26-3).

Dissection begins 5 to 6 fingerbreadths below the fibular head so as not to encounter the peroneal nerve. Once the plane between the soleus and gastrocnemius is clearly identified, superficial dissection can begin between the gastrocnemius and the investing fascia of the lower leg. The common peroneal nerve must be identified at the level of the neck of the fibula and protected throughout the rest of the dissection. Since the common peroneal nerve passes between the lateral head of the gastrocnemius and the tendon of the biceps femoris, particular attention should be paid to identify the nerve during this portion of the dissection. The muscle is then separated as described for the medial gastrocnemius muscle flap, and may then be pedicled into the overlying defect site (Fig. 26-4).

The flap is inset, and the donor site closed as previously described for the medial gastrocnemius flap.

Postoperative Management

The patient should be maintained on bed rest for 5 to 7 days. The bolster dressing may be removed on postoperative day 5 to ensure adequate skin graft â€œtake.â€• Following skin graft take, Xeroform dressing changes are performed once to twice a day until the skin graft has matured, at which point the patient is encouraged to keep the skin graft lubricated with a petroleum-based lotion to prevent desiccation of the newly grafted skin. If immediate postoperative motion is essential, the origin of the gastrocnemius muscle should be divided to minimize forces across the flap in the early postoperative period. A VAC sponge may also aid in securing skin grafts if immediate motion is essential.

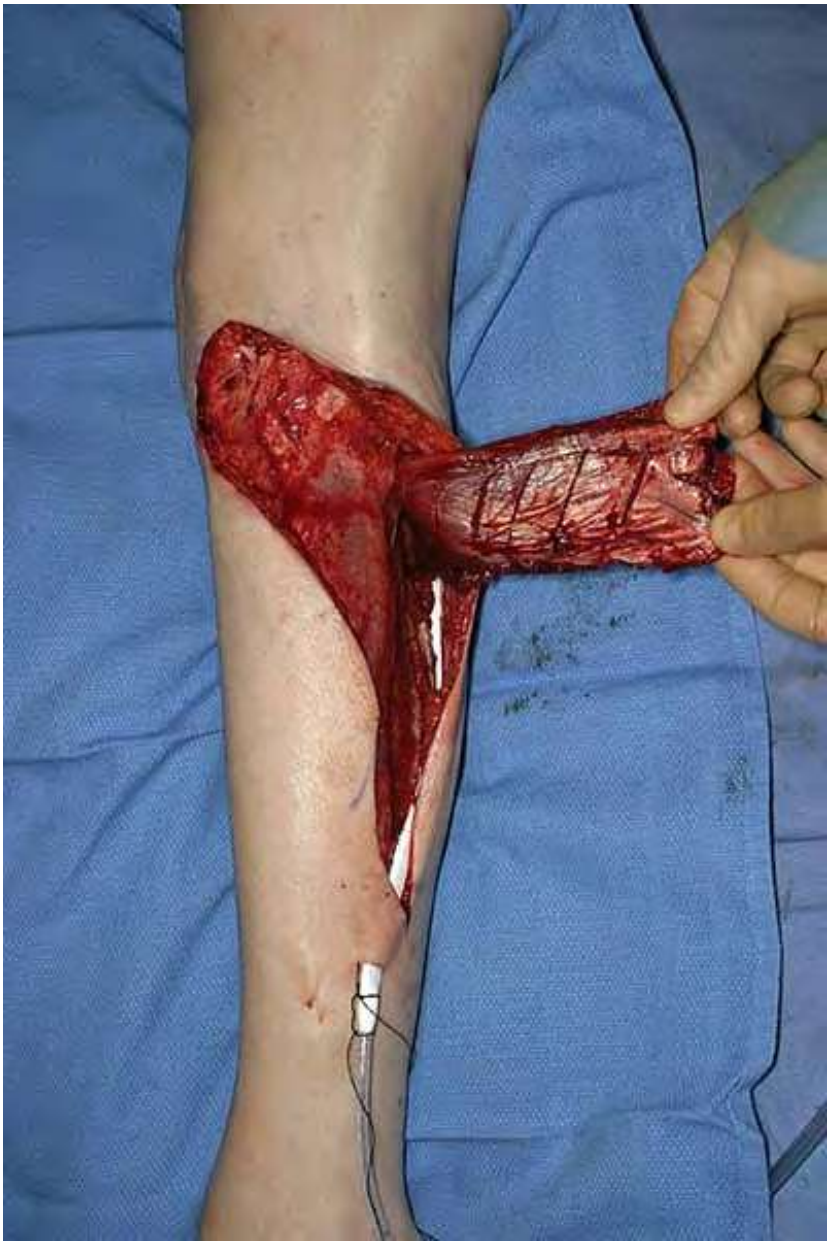

FIGURE 26-3 The incision for exposure of the lateral gastrocnemius muscle flap is made 2â€“3 cm posterior to the fibula.

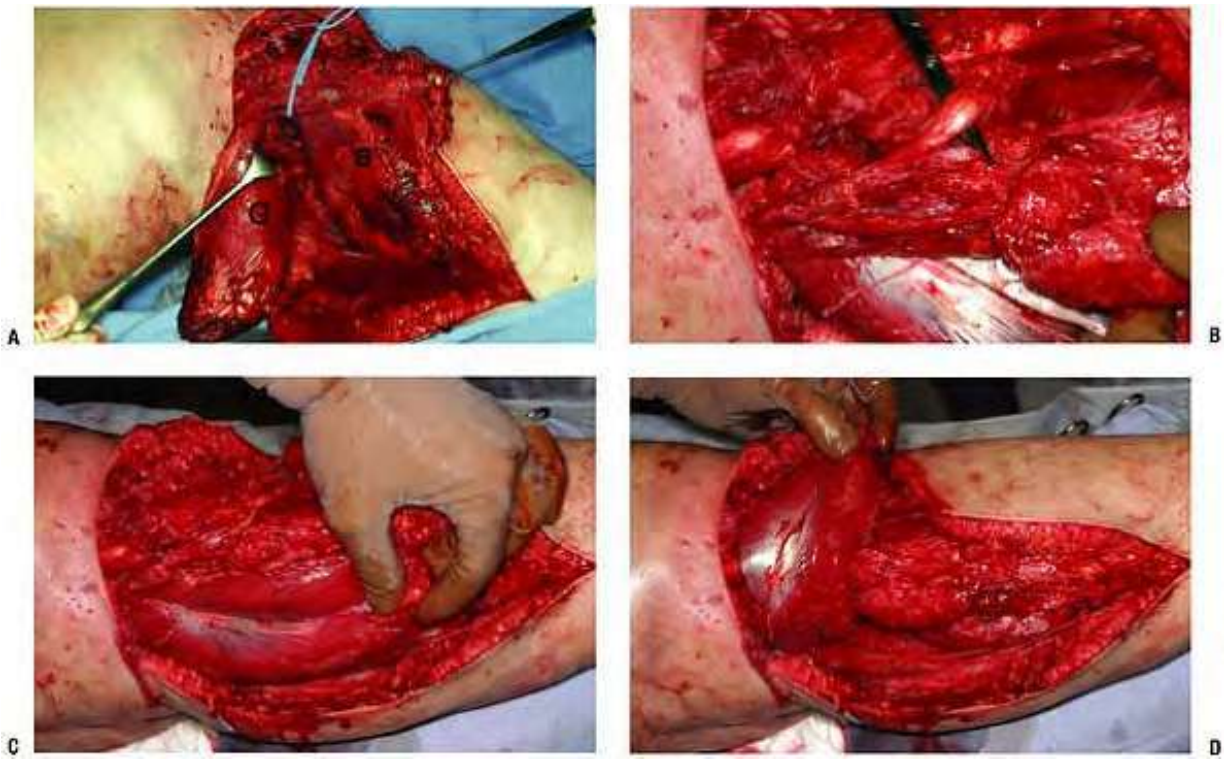

FIGURE 26-4 A: A traumatic lateral tibial and patellar defect in 28-year-old woman. The lateral gastrocnemius muscle has been harvested with a posterior and inferior incision extended from the original defect site. Note that the common perineal nerve has been identified and marked with blue vessel loop. The soleus muscle belly (*s*) can be seen below the gastrocnemius muscle (*g*). B: The common perineal nerve (seen here running above the scissors) must be identified prior to pedicling the flap into the defect site to ensure it is not inadvertently compressed or injured during flap insetting. C,D: Once the lateral gastrocnemius muscle is isolated from midline raphe, it can easily reach the lateral aspect of the peri-patellar region.

FIGURE 26-5 A 2-year follow-up of a 35-year-old man who underwent medial gastrocnemius muscle coverage for exposed tibia following a motor vehicle accident. Significant improvement in contour is obtained with long-term use of a compression wrap or compression stocking.

The drains are removed once drainage is less than 30 mL/day. Gentle range of motion should then begin at 10 to 14 days, increasing the motion by 5 or 10 degrees per day. Ambulation can begin on the postoperative day 10, with knee flexion limited with the use of a knee immobilizer. If the flap is inset under tension or if the healing environment is less than ideal, the knee immobilizer can be maintained for an additional week to prevent undo tension on the margins of the muscle flap.

Postoperative swelling is minimized with the use of elastic wraps or supportive stockings. Compression stockings worn long-term can facilitate flap contouring. Muscle

flaps undergo a process of atrophy over 6 to 12 months; patient concerns regarding scaring or flap contour should not be addressed surgically until 12 to 24 months following the initial reconstructive surgery (Fig. 26-5).

Complications

Complications following the use of the gastrocnemius muscle flap include bleeding, infection, nerve injury, and flap loss. Bleeding (hematoma) and infection are most common, though rare if meticulous hemostasis is achieved at closure and aggressive debridement of the wound is performed prior to muscle flap transfer. Suction drains placed within the donor site at the time of closure will significantly minimize the risk of postoperative hematoma formation.

Partial flap necrosis or complete flap loss is usually associated with technical error, including injury to the vascular pedicle and/or excessive tension with secondary flap ischemia. If flap ischemia is suspected intraoperatively, the vascular pedicle should be examined for signs of kinking or injury to the sural artery. Small areas of skin graft loss may be seen postoperatively, but can be conservatively managed with topical ointments and wound care without the need for further surgical intervention.

Nerve injury is uncommon with meticulous dissection and nerve preservation. Injury to the sural nerve can occur during separation of the medial and lateral heads of the gastrocnemius muscle. Injury to the common perineal nerve may occur during elevation of the lateral gastrocnemius flap. Nerve injury is best avoided by clearly visualizing the nerve prior to flap transfer. The saphenous

nerve may be ligated inadvertently with the saphenous vein during exposure of the posterior compartment; attempts should be made at nerve identification and preservation during flap dissection.

The gastrocnemius muscle provides plantar flexion of the foot and flexion of the knee joint. Removal of the muscle does not result in significant morbidity, although some loss of plantar flexion may lessen jumping and leaping ability. Either or both heads may be used with little functional deficit as long as the soleus muscle is preserved and uninjured.

Recommended Reading

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 27

Revision and Infected Total Knee Arthroplasty

Henry D. Clarke

William J. Casey III

Mark J. Spangehl

Indications/Contraindications

Soft tissue management in revision and infected total knee arthroplasty (TKA) is an extremely important factor in determining the success or failure of the procedure. Without optimal wound healing, complication rates increase and the ultimate benefit of the operation may be diminished. Despite the critical role played by the soft tissues surrounding the knee, very little scientific data exist to help guide the practicing orthopaedic surgeon in this area. Rather, most of the recommendations and guidelines are based on extrapolations from other surgical fields and anecdotal experience from experts in knee arthroplasty. In this chapter, important factors that should be considered in the preoperative, intraoperative, and postoperative settings are reviewed. In addition, surgical techniques that may minimize potential soft tissue problems and help manage complications when they do occur are also presented. The soft tissues about the knee include skin and subcutaneous coverage, as well as the extensor mechanism; therefore, in this chapter management of both of these important components is addressed.

The very circumstances that necessitate revision TKA in the settings of either infection or aseptic failure require that every patient be considered at risk for soft tissue complications. Therefore, every patient in whom revision TKA is being considered should be thoroughly evaluated preoperatively. This allows all necessary plans to be addressed in order to meet individual needs. Indeed, absolute indications and contradictions are difficult to define: rather, all factors should be considered. While any patient who undergoes revision TKA should be considered at risk for soft tissue complications, some patients will have greater risks. A healthy patient with just one prior knee surgery, with aseptic loosening, is at less risk than a patient with diabetes and peripheral vascular disease who has a chronically infected total knee replacement. However, each patient should be carefully assessed. Certainly, once the risks have been evaluated, a decision to proceed with revision TKA surgery should be based on the individual circumstances. In each case, the current symptoms, including pain and disability, presence or absence of prosthetic infection, and age and activity demands of the patient, must be evaluated in the context of the potential risks. The potential increase in symptoms, time frame for deterioration, and additional surgical problems that may be encountered by delaying surgery should also be considered. Absolute contraindications for revision TKA include irreversible medical comorbidities that raise the risk of perioperative mortality to unacceptable levels, an avascular extremity where revascularization options have been exhausted, uncontrolled sepsis, and neurologic injuries with no motor function of the extremity. Relative contraindications include failed prior soft tissue flaps, massive bone loss, recurrent prosthetic infection, extensor mechanism disruption, and unstable medical comorbidities that require optimization. When the potential risks are too great, or when other rare circumstances are encountered, such as life-threatening sepsis or an avascular limb, alternative procedures including permanent resection arthroplasty, knee arthrodesis, or amputation may be required.

Preoperative Planning

Preoperative planning is critical for minimizing soft tissue problems about the knee. Both systemic factors and the characteristics of the knee and leg should be considered before revision knee surgery.

Systemic Factors

Patient-specific medical comorbidities should be assessed and treated before surgery (Table 27-1). While optimization of these conditions appears to reduce the risk of developing wound complications after revision TKA, the increased risk of wound problems in many cases is multi-factorial and does not return to baseline levels even when medical management is optimal. For example, the increased risks of delayed wound healing and infection noted in patients with diabetes appears to be caused by both the hyperglycemia, which has been shown to inhibit collagen synthesis, disrupt fibroblast proliferation, slow capillary in-growth, and deficiencies in polymorphonuclear neutrophil cell function. Furthermore, the increased risk of soft tissue complications in patients with rheumatoid arthritis is likely due to numerous factors including the long-term sequelae of corticosteroid use, which regulates macrophage function, reduces collagen synthesis, and delays vascular in-growth, as well as direct causes of the disease itself such as skin atrophy, decreased albumin, and vasculitis.

Another important risk factor in the pre- and postoperative periods is the use of tobacco products. Nicotine and its metabolites cause vasoconstriction, which interferes with micro-circulation; however, the effects on wound healing are more profound than from vasoconstriction alone. Other constituents of tobacco smoke, such as carbon monoxide, reduce the oxygen carrying capabilities of hemoglobin, which reduces tissue oxygenation. Nicotine also appears to have a direct effect on fibroblast and immune function. The optimal time for smoking cessation appears to be 4 to 8 weeks preoperatively, but even a week of abstinence appears to reduce the risk of complications. Abstinence in the entire postoperative period is also critical. For elective procedures, all patients who smoke and have other risk factors for wound healing must stop smoking before the procedure. In some cases, professional services must be used to achieve this end. The use of nicotine delivery substitutes are controversial, as these do not eliminate the patient's use of nicotine but do eliminate multiple other toxic substances that are inhaled in tobacco smoke, such as carbon monoxide and hydrogen cyanide. Therefore, while not an optimal solution, in certain circumstances these substitutes may be preferable to reduce the overall risk.

Table 27-1. Complicating Comorbidities

Condition	Evaluation	Management
Diabetes	Blood glucose	Optimize diet and medication.
Malnutrition	Total lymphocyte count < 1,500 mm^3 Serum albumin < 3.4 g/dL	Nutritional supplements before surgery
Inflammatory arthritis	Review disease modifying medications	Hold medications, if possible, until wound healing has occurred.
Tobacco smoking	Serum kotinine level if poor abstinence/compliance	Avoid tobacco use 1 month before surgery.
Anemia	Hgb/Hct	Correct before surgery with erythropoietin injections.
Hypoxia	SaO^2	Optimize pulmonary function.
Peripheral vascular disease	Peripheral pulses, Doppler ultrasound, arteriogram	Revascularization prior to revision TKA

Obesity	BMI	Avoid significant weight loss in the immediate pre- or postoperative periods.
Cancer	History	Avoid elective surgery, if possible, while on chemotherapeutic agents.
Chronic corticosteroid use	History	> 2 years duration appears to increase risks of wound complications

BMI, body mass index; TKA, total knee arthroplasty.

The Knee

The local vascular anatomy of the knee can be unforgiving; in comparison to the hip, where wound necrosis is rarely a problem. Most of the perfusion to the skin is derived from perforator vessels that originate below the level of the deep fascia. Little subcutaneous communication occurs; rather, the perforating vessels connect at the level of the sub dermal plexus. Thus, when raising skin flaps about the knee, the dissection should be below the level of the deep fascia, rather than in the sub dermal plane to avoid damaging this superficial plexus. While the peri-patellar plexus formed by the medial and lateral, superior and inferior genicular arteries has been well described, the majority of the inflow to the skin appears to be derived from the medial side of the knee in the distribution of the saphenous and descending genicular arteries. Therefore, laterally based flaps should be avoided and any prior laterally biased vertical incision should be carefully considered when selecting the optimal incision. On the lateral side, the perforators from the

superior and inferior genicular arteries are relatively more important, and it has been reported that use of a lateral patellar release with disruption of the superior genicular artery is associated with decreased tissue oxygenation and increased lateral wound problems. Some general guidelines for selection and placement of skin incisions follow.

A single, longitudinal anterior midline incision provides the most extensile exposure and is preferred for all revision and infected total knee replacement cases. Prior incisions about the knee should be carefully evaluated. Factors that are believed to be important, although little scientific data exist to support these suppositions, include age, orientation, length, and placement in relation to other or intended incisions. A single, transverse incision can be crossed by a new perpendicular anterior incision with relatively little concern. When a single, prior longitudinal anterior incision exists, this incision should be used. If placement of this prior incision is not directly midline, the proximal and distal ends of the incision may be extended back toward the midline to reduce tension on the wound. In these cases, subcutaneous dissection should be minimized; it is critical to maintain full-thickness flaps if the subcutaneous tissues must be mobilized to allow adequate exposure of the extensor mechanism. To preserve the sub dermal plexus, these flaps should be raised at the level of the deep fascia. If the prior incision is located far from the midline, and would necessitate creation of a large laterally based subcutaneous flap, one may consider using a new incision. In these circumstances, skin bridges greater than 3 to 5 cm should be maintained between the new and old incisions. When multiple incisions, or a single anterior incision located well away from the midline, are encountered, selection of the optimal incision should be carefully considered. In some circumstances where numerous prior incisions exist, or where the skin and subcutaneous tissues have been severely damaged by prior trauma, radiation exposure, infection, or are atrophic, due to systemic conditions, preoperative soft tissue management techniques that are described in following sections, such as soft tissue expansion or soft tissue flaps, may be required. Again, little definitive data exist to guide the surgeon when these interventions are absolutely required; rather, much of this assessment is subjective.

Surgery

Soft Tissue Procedures

Sham Incision

Technique

The sham incision or delay procedure is reviewed primarily for its historical significance as an early attempt to manage patients considered at risk for wound healing problems after TKA. Approximately 10 to 14 days before the planned knee procedure, the intended skin incision is made and extended to the level of the extensor mechanism. Next, the skin is undermined on both sides of the incision to expose the extensor mechanism, as required for the upcoming knee replacement, and then the incision is closed and observed. If the wound heals without complications, the intended knee procedure is subsequently performed. If skin necrosis occurs, a soft tissue flap is performed before knee replacement. The main advantage in these cases is that the knee joint has not been violated, and there is no prosthesis at risk for infection. Furthermore, increased collateral flow caused by the incision due to the delay phenomenon was felt to reduce the risk of subsequent wound

healing problems. More contemporary soft tissue expansion techniques have, for the most part, eliminated the use of the sham incision.

Results

There is little except anecdotal evidence to support the use of this technique. Rothaus has detailed the outcomes in a small group of 12 patients, with multiple prior incisions, in whom this technique was used. In all 12 patients, the incisions healed and TKA was successfully performed without wound healing problems.

Soft Tissue Expanders

Technique

The concept of tissue expansion is not new and has been successfully used in a variety of soft tissue reconstructive procedures throughout the body. Despite the challenges posed by the unforgiving nature of the vascular supply around the knee,

the use of tissue expanders is a relatively new concept with only a limited number of large patient series.

Tissue expansion has three important benefits that aid wound healing in the overlying soft tissues. First, and probably most important, it stimulates neo-vascularization of the overlying soft tissue; this improves the capillary in-flow that directly promotes healing. Second, it physically expands the overlying skin and subcutaneous tissue allowing larger areas, due to correction of malalignment or limb lengthening, to be covered. This also reduces the tension on the wound. This expansion is not simply a stretching and thinning of the overlying soft tissue, which would leave the flap less resilient, but actually a hypertrophy of the dermis and epidermis due to fibroblast stimulation, leading to an increase in the quantity of tissue available. Third, a fibrous capsule develops around the expander, and this thick, robust, vascular tissue is very useful as a â€œpseudo-fascialâ€• layer to close over the joint and is particularly helpful in the area of the proximal medial tibia where the periosteum and soft tissue layer can be thin.

The technique for soft tissue expansion over the knee is begun approximately 6 to 8 weeks prior to the intended TKA (Fig. 27-1). Typically two to four 200 to 300 mL expanders are placed in pockets deep to the subcutaneous layer, just above the fascia (see Fig. 27-1C). These pockets are strategically placed based on the placement of prior incisions and characteristics of the overlying subcutaneous tissue and skin. The first step involves infiltrating the deep subcutaneous layer with a dilute solution composed of 1000 mL Ringer's lactate, 50 mL 1% lidocaine, and 1 mL 1:1000 epinephrine to create a pocket. Typically, injection of 250 to 300 mL of this solution is sufficient to produce the hydro-dissection that separates the skin and subcutaneous tissue from the underlying fascia to create a pocket. Also, use of the local anesthetic helps with postoperative pain relief. Next, through a short incision placed in one of the prior incisions, scissors are used to bluntly define and enlarge this plane within the deep subcutaneous tissue to create a pocket (see Fig. 27-1B). The un-inflated expander is then inserted through the same incision (see Fig. 27-1D). The access port to the expander is then tunneled through the subcutaneous tissue to an easily accessible site. The port should be accessed prior to skin closure and saline injected to ensure that the port is functional and that the expander was not damaged during the insertion. After this initial expansion, the incision is closed and a sterile dressing applied. Patients are typically placed in a

knee immobilizer for a week and allowed to weight bear as tolerated. Most patients are admitted for an overnight stay and treated with 24 hours of intravenous antibiotics. Subsequently each week, about 10% to 15% of the volume of the expander is infused with the patient as an outpatient, via the access port. If at any time the overlying skin blanches and doesnâ€™t recover after a few minutes of observation, or if the patient experiences significant pain, saline must be removed until the problem is alleviated.

At the time of the intended TKA, the expanders are extracted through the incision used for the procedure (see Fig. 27-1F). The subcutaneous tissue and skin flaps should be protected during the TKA. At the end of the procedure, a superficial drain is placed in each individual expander pocket. These drains are removed individually once output is less than 10 mL per 8 hours or 30 mL per 24 hours. This protocol appears to reduce the risk of subcutaneous hematoma formation. Early in the experience of tissue expansion about the knee when subcutaneous drains were not routinely employed, hematomas occurred more frequently and occasionally had to be drained to reduce tension on the overlying tissue. During wound closure, the expansion process may have created excessive amounts of soft tissue that need to be excised. In these cases, the edge of the flap, especially focusing on old widened scars, can be trimmed or removed. However, the tension on the wound should not be increased by resecting too much surplus tissue.

FIGURE 27-1 A: Knee with multiple prior vertical and horizontal incisions at risk for poor postoperative wound healing after TKA.B: Creation of subcutaneous pocket for soft tissue expander.C: Saline expanders demonstrating access portals for infusing saline. D: Insertion of the deflated saline soft tissue expander. E: Knee after tissue expansion process is

complete. F: Fibrous, vascular membrane that surrounds the expander cavity after removal of the expander at the time of TKA.

Results

Few large series of tissue expansion in association with TKA have been published, with most reports presenting only a handful of cases. The morbidity associated with the use of tissue expanders has been previously reported and includes infection, hematoma, expander rupture or deflation, and skin necrosis. Two significant papers have reported very favorable results. In an initial study, no significant complications occurred during the expansion process in a small group of 10 patients. In a second larger group of 29 knees, six (21%) minor wound complications occurred during expansion, including mild erythema or skin blistering. These complications were successfully managed by reducing the volume in the expander, and delaying further expansion until the problem resolved. In this same group, one major complication occurred after the insertion of the soft tissue expanders; this involved full-thickness skin necrosis in a patient with a history of radiation to the anterior knee. The knee replacement was subsequently not performed, as the patient declined to accept a prophylactic muscle flap, choosing to endure her arthritic symptoms.

Following the knee replacement, 5 (18%) of 29 patients experienced minor complications; three knees developed persistent drainage, and two knees developed subcutaneous hematomas. The persistent drainage resolved after wound compression and immobilization was initiated. The two subcutaneous hematomas required surgical evacuation, and it was these problems that prompted the use of subcutaneous drains in the expander pockets.

Soft Tissue Flap Coverage Procedures

Soft tissue flap procedures about the knee have been used both prophylactically and for salvage of wound complications about the knee. Although a large number of different techniques have been described, including simple skin grafts, random or axial pattern skin flaps, fasciocutaneous flaps, and rotational or free muscle flaps,

the most reliable and most frequently used tissues about the knee include medial gastrocnemius muscle flaps, and free latissimus dorsi or rectus abdominus flaps. In addition

to these techniques, the development of musculocutaneous perforator flaps during the past decade has provided another alternative for coverage about the knee. The principles involved in these three types of techniques are described following in ascending order of the extent of the area that can be adequately managed.

Musculocutaneous Perforator Flaps

The development of musculocutaneous perforator flaps in plastic surgery occurred during the 1990s as an extension of the generally very favorable experience with musculocutaneous flaps during the 1980s. Perforator flaps were based on the principle that neither the underlying fascial plexus, nor muscle, was required if the vessel that perforated through the muscle was carefully dissected. The major advantage of musculocutaneous perforator flaps versus musculocutaneous flaps is that the functional and cosmetic morbidity due to transfer of the underlying muscle could be avoided by transferring only the overlying skin.

Although perforator flaps can provide additional vascularized tissue in an area that may be deficient or fibrotic, their use in the setting of infection is controversial; and in these cases, traditional muscle flaps are preferred. Their use is best reserved for aseptic cases with a deficient or scarred soft tissue envelope.

Technique

The wound is first debrided to remove all necrotic tissue (Fig. 27-2A,B). Next, elevation of the perforator flap requires careful microsurgical dissection of the musculocutaneous perforator vessels (Fig. 27-2C). Due to anatomical variability of these perforating vessels, Doppler ultrasound is used to locate the perforator before elevation of the flap. Once isolated, the perforator flap can be rotated about the vascular pedicle or transferred as a free flap (Fig. 27-2D). The donor site can be managed with either primary closure or split-thickness skin grafts depending on the specific perforator flap used (Fig. 27-2E). A number of perforator flaps have proven quite reliable for

managing soft tissue problems in the lower extremity; these include flaps based on the muscle perforators of the vastus lateralis, tensor fascia lata, sartorius, gracilis, and medial gastrocnemius. Unfortunately, due to the volume of tissue required in many cases associated with failed TKA, these muscle perforator flaps often would not provide enough soft tissue to cover the defect; in these circumstances, one of the alternative techniques must be considered.

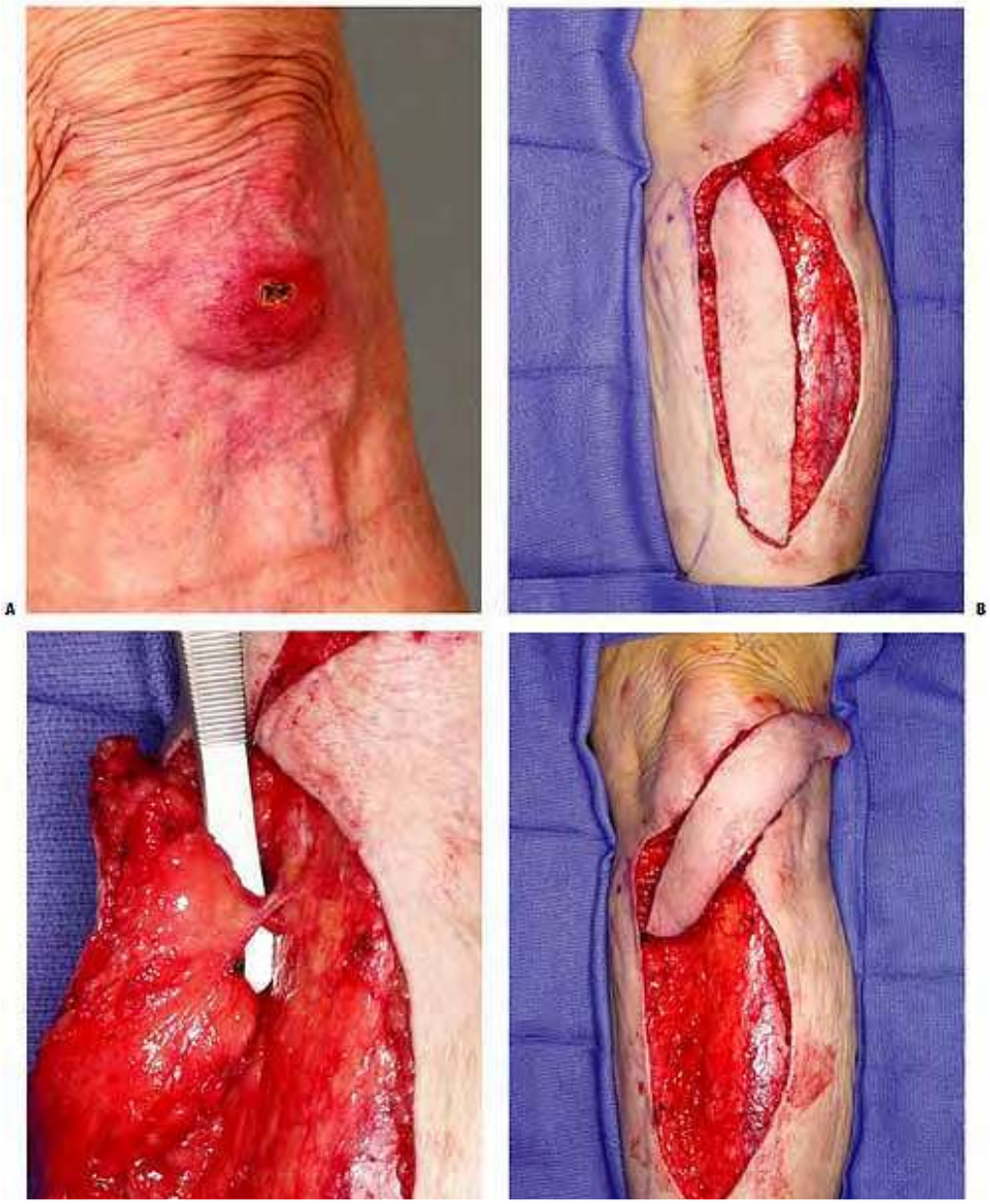

FIGURE 27-2 A: Pre-patellar bursa with central necrosis and recurrent infection that failed numerous attempts at local wound care. B: A perforator flap has been harvested after identifying the vessel, and the pre-patellar area has been debrided. C: Perforator vessel to flap. D: The flap is rotated 180 degrees on its pedicle and transferred to the area of the defect. E: Harvest site incisions have been closed primarily and the flap sutured in place.

Medial Gastrocnemius Flap

Technique

Although both the lateral and medial heads of the gastrocnemius muscle can be used to provide soft tissue coverage about the knee, both the quantity of tissue provided by the lateral gastrocnemius, and its shorter vascular pedicle, restrict its use. The lateral head is generally 3 to 4 cm shorter than the medial head, and its transfer can place the common peroneal nerve at risk for injury as the muscle is passed across the proximal fibula. Therefore, the medial gastrocnemius rotational flap has been the preferred option for treating wound problems in association with

total knee replacement (Figs. 27-3 and 27-4). First, the wound is debrided to healthy margins. Next, the medial

gastrocnemius muscle flap is developed through a posteromedial incision, separating the muscle belly away from the deep tissue superficially, and the underlying soleus muscle along its deep surface. The plantaris tendon provides an excellent landmark. The medial head is separated from the lateral head along its decussation, which does place the sural nerve at risk for injury. The vascular pedicle is based proximally on the sural artery, which is the first branch of the popliteal artery. The muscle is then divided distally at the musculotendinous junction and folded proximally. Next, the flap is tunneled under a medial skin bridge to cover the defect, where it is sutured in place (see Fig. 27-4). If added length is needed, radial scoring of the fascia along the undersurface of the muscle can be performed at 1-cm intervals to increase its arc of rotation. Once transferred, the muscle flap is covered with a split-thickness skin graft that can be performed either primarily, or at a later date; in most circumstances, the donor site can be closed primarily. Alternatively, the flap may be harvested as a musculocutaneous flap, where the entire soft tissue flap including the skin, subcutaneous tissue, fascia, and muscle is rotated about the vascular pedicle. In these cases the donor site must be skin-grafted (see Chapter 29).

FIGURE 27-3 A: Debrided wound after full-thickness necrosis over proximal tibia after TKA. B: The medial gastrocnemius is harvested through a posteromedial incision of the calf and then tunneled under a skin bridge to the debrided area. C: The muscle flap is covered with a split-thickness skin graft.

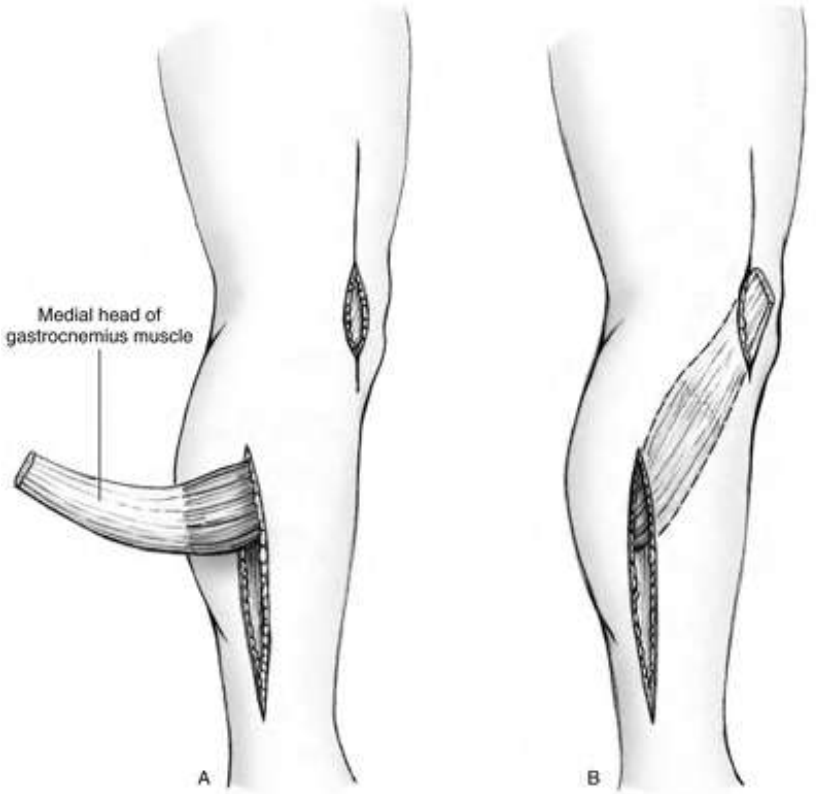

FIGURE 27-4 A: The medial head of the gastrocnemius muscle is harvested through a medial calf incision. B: The muscle is then rotated and delivered a skin bridge into the defect.

Results

In association with infected or exposed total knee prostheses where wound breakdown has occurred, use of medial gastrocnemius flaps has been quite successful, considering the complexity of the problems. In one series, Gerwin et al reported their experience from the Hospital for Special Surgery; successful salvage of an infected, or exposed total knee prosthesis with a medial gastrocnemius flap was achieved in 10 of 12 patients. Similar results have been reported by McPherson et al, with successful reimplantation in 20 of 21 patients who had chronically infected total knee replacements and compromised soft tissues about the knee.

Utilization of the gastrocnemius muscle for knee coverage results in little functional morbidity, with no significant deficits at walking speed, and only mild deficits as demand increases.

Free Latissimus Dorsi and Rectus Abdominus Muscle Flaps

In cases where the medial gastrocnemius flap has been previously used and failed, or when the extent of the soft tissue necrosis is large, alternative flaps must be considered. The free transfer of a latissimus dorsi or rectus

abdominus muscle flap has been reliable in these difficult circumstances. The muscle is harvested with its vascular pedicle and the donor site primarily closed. The recipient site is debrided to healthy wound edges and then the free muscle is sutured into the defect after the vascular anastomosis is complete. The vascular anastomosis is typically performed to the popliteal vessels.

Postoperative Management of Soft Tissue Flap Coverage Procedures

Immobilization with elevation for 7 to 10 days to reduce venous congestion is common to all flap coverage procedures. Partial weight bearing and progressive range of motion are then begun on a case by case basis. Factors to be considered include any wound drainage, marginal necrosis, venous congestion, and success of the overlying skin graft. If the flap transfer is done prior to the TKA as a prophylactic measure, then the total joint replacement may be performed after the flap has matured, which is generally a minimum of 8 to 12 weeks.

Complications of Soft Tissue Flap Coverage Procedures

While successful retention or reimplantation of the knee prosthesis has been reported, the functional results associated with these so-called successful outcomes falls far short of knee replacement in association with uncomplicated wound healing. In addition to poor functional results, other complications associated with all types of flaps can include recurrent infection or recurrent wound problems, due to marginal necrosis or complete loss of the flap, as well as skin graft problems at the donor or graft sites. Weakness caused by transfer of the donor muscle can also occur. Due to loss of the medial gastrocnemius, patients

have reduced plantar flexion strength of the involved ankle; however, given the extent of the associated knee problems, the functional limitations due to loss of the medial gastrocnemius may be relatively minor. Cosmesis can also be an issue as the flaps may be quite bulky; however, this concern must be considered relative to the presenting problem. Occasionally, debulking or thinning of a mature flap will be undertaken where atrophy of the muscle was less than anticipated.

Surgical Exposure

In addition to the specialized plastic surgery techniques that may be required in revision TKA, the orthopaedic surgeon must also be aware of the techniques that allow adequate surgical exposure to be obtained during the knee procedure in a manner that minimizes risks to the skin and extensor mechanism. If an intraoperative extensor mechanism disruption occurs during revision TKA, or if a preoperative extensor mechanism failure is known to exist, the surgeon must also be aware of the techniques used to address these problems. Simple repair, especially when associated with chronic extensor mechanism disruption, has a very poor track record, and the alternative reconstructive options are described following.

General principles to optimize wound healing and avoid complications include meticulous soft tissue handling in every case. Prolonged local tension from self-retaining retractors or from vigorous stretching due to attempts to perform mini-incision surgery should be avoided; in addition, undermining along the margins of the incision should be minimized. Intraoperatively, the importance of optimizing component positioning to avoid the need for a lateral patellar release that is associated with increased wound problems is also important. The use of a tourniquet is controversial, but evidence suggests that postoperative tissue oxygenation is reduced in patients in whom a tourniquet is used. Certainly, in patients with significant peripheral vascular disease, especially those in whom either a bypass has already been performed, or the lateral radiograph shows extensive calcification of the popliteal vessels, consideration should be made to perform the knee replacement without a tourniquet. In addition to the previously noted principles, numerous techniques exist for facilitating surgical exposure and relieving tension on the extensor mechanism to reduce the risk of iatrogenic tendon rupture or patellar fracture. These techniques, along with the results and

complications that are specific to each, are reviewed following. All of these procedures, except where noted, are performed during revision TKA with the patient positioned in the supine position. The specific techniques are described generally progressing from less extensile to more extensile.

Medial Parapatellar Arthrotomy and Proximal Tibial Peel

The medial parapatellar arthrotomy allows the optimal exposure in revision TKA and is versatile as it is compatible with each of the more extensile exposures detailed in subsequent sections. The arthrotomy is begun at the proximal end of the quadriceps tendon approximately 6 to 8 cm proximal to the superior pole of the patella (Fig. 27-5). The arthrotomy is extended distally approximately 3 to 5 mm lateral to the medial border of tendon. At the superior pole of the patella, the arthrotomy is curved around the medial

border of the patella and then continued distally along the medial edge of the patellar tendon. Next, the periosteum of the medial tibia is elevated sharply beginning at the medial border of the arthrotomy at the level of the joint. Working medially and distally, the entire medial periosteum is elevated approximately 5 to 7 cm distally. A periosteal elevator can be used once the flap of periosteum has been raised. At this stage the knee is flexed and the tibia is gradually externally rotated while the sub-periosteal elevation is continued all the way to the posteromedial corner. In many cases where this tissue plane was not violated in the primary replacement, the semi-membraneosus insertion on the posteromedial corner will be well defined, and this expansile insertion should be released in the sub-periosteal plane. As the dissection is continued to the posterolateral corner, and the tibia is externally rotated, the tibia will sublux from underneath the femur. In many cases, this will allow access to the modular tibial polyethylene. Once exposed, the tibial insert can be removed; this maneuver relaxes the flexion and extension gaps and usually improves the exposure. This exposure should be incorporated into every revision and provides the foundation for obtaining adequate visualization and access in order to safely remove the existing components and reconstruct the knee.

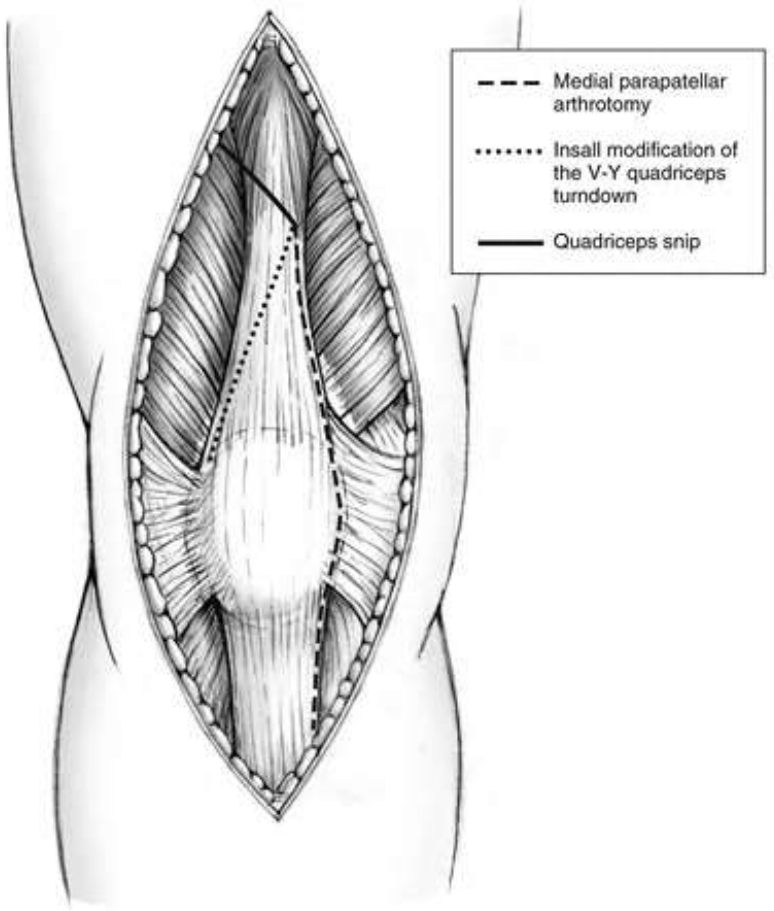

FIGURE 27-5 The medial arthrotomy (*dashed line*) is the standard exposure for revision TKA. This can be extended with a quadriceps snip (*solid line*) if the extensor mechanism is tight. The Insall modification of the V-Y quadriceps turndown (*dotted line*) allows extensile exposure in the stiff knee but is associated with postoperative extensor lag.

Quadriceps Snip

Technique

The quadriceps snip, originally promoted by Dr. John Insall, facilitates the exposure

of stiff knees in the primary and revision setting. This technique is a proximal extension of a medial parapatellar arthrotomy (see Fig. 27-5). Beginning at the proximal apex of the standard arthrotomy, the incision is extended proximally and laterally at a 45-degree angle into the fibers of the vastus lateralis. This extension not only relieves the tension on the extensor mechanism in a stiff knee that can aid in exposure of the joint, but also allows the patella to be everted more easily during patellar preparation if this is required. The arthrotomy and this proximal extension are closed in routine fashion once the arthroplasty has been performed. In distinction to many of the alternative techniques for optimizing exposure in difficult cases, there is no need to modify or restrict postoperative rehabilitation. Indeed, patients in whom the quadriceps snip has been used may participate in standard TKA

rehabilitation protocols. While this technique allows adequate exposure in most circumstances, in particularly stiff or difficult cases, additional measures may be required. In these extreme cases, a tibial tubercle osteotomy can be used even after a quadriceps snip has already been performed.

Results

Insall's clinical experience in 16 patients who had bilateral TKA with a quadriceps snip on only on one side has been reported. In these patients, no differences in quadriceps strength were observed postoperatively between the two sides. In a similar study, no difference was identified between patients who had undergone a quadriceps snip, versus those in whom a standard medial parapatellar arthrotomy had been performed.

Tibial Tubercle Osteotomy

This technique, popularized by Whiteside, facilitates exposure in even very stiff knees. In addition, the tibial tubercle osteotomy can allow access to the tibial canal that may be helpful in revision or infected TKA when well-fixed stemmed, tibial components must be removed. This technique involves first incising the medial periosteum along the medial border of the patella tendon for approximately 6 to 7 cm distal to the joint line. The osteotomy is then performed with an oscillating saw from medial to lateral, creating a wedge of bone that is

approximately 2 cm wide and 6 to 8 cm long. The osteotomized wedge should taper from a thickness of about 1 cm proximally, to the level of the anterior cortex distally. This minimizes the stress riser in the anterior tibia, especially at the distal end. In addition, at the proximal end, a step cut should be created about 1 cm distal to the joint line. This step acts as a buttress that helps to resist proximal migration of the osteotomized wedge after it has been reattached at the end of the case. Proximal migration is also resisted by the lateral periosteum and musculature that should remain attached to the osteotomized piece. During the revision knee arthroplasty, the lateral soft tissues act as a hinge that allows the osteotomized segment to be everted and displaced laterally. At the time of closure, three to four 18-gauge wires are used for fixation. The wires are passed through drill holes in the medial tibia and then brought up through the lateral aspect of the tubercle shingle. They are then brought distally and medially over the tibial crest and tightened on the medial side of the tibia. Tightening the wires pulls the osteotomy distally. The arthrotomy is then closed in a standard fashion. In most cases, we favor bypassing the tibial tubercle with a stemmed component. If a cemented stem is used, it is helpful to remember to pass the wires before insertion of cement.

Results

Advantages of the tibial tubercle osteotomy versus the V-Y quadriceps turndown include a lower incidence of extensor lag and quadriceps weakness. Whiteside reported his experience with this technique in 136 patients; only two patients experienced a residual extensor lag. Another advantage of this technique is that if rigid fixation is obtained, full weight bearing and unrestricted range of motion are allowed postoperatively. Reported complications of tibial tubercle osteotomy that are not noted with the quadriceps turndown include proximal migration of the osteotomized tubercle, patellar tendon disruption, and tibial shaft fractures. Despite these unique problems, we favor the use of tubercle osteotomy when adequate exposure cannot be obtained with a standard arthrotomy and quadriceps snip.

V-Y Quadriceps Turndown

Technique

The V-Y quadriceps turndown was originally described by Coonse and Adams and subsequently modified by Insall. In both techniques an inverted V-shaped flap incorporating the quadriceps tendon is created. Insall's modification described creating a second incision beginning at the apex of a standard medial parapatellar arthrotomy that is extended distally and laterally at a 45-degree angle along the tendinous portion of the vastus lateralis (see Fig. 27-5). In addition to allowing excellent exposure, the inverted V of tissue may be lengthened at the time of closure, forming an inverted Y shape.

Results

Although this technique allows excellent exposure and releases excessive tension on the extensor mechanism during surgery, patients may develop an extension lag postoperatively. Trousdale et al reported the Mayo Clinic experience with this technique. Patients who had undergone bilateral TKA with a V-Y quadriceps turndown in one knee, and a standard medial parapatellar arthrotomy in the contralateral knee, underwent postoperative strength testing. In these patients, the results did not show any statistical differences between the strength of the two legs. However, 5 of 14 patients with the V-Y quadriceps turndown had a persistent extensor lag. Furthermore, use of the V-Y quadriceps turndown requires alterations in the postoperative rehabilitation protocol that includes

restricted range of motion, use of a brace, and partial weight bearing for about 6 weeks. Due to the high rate of lag, and the increased recovery time associated with the use of the V-Y turndown, we favor the use of the tibial tubercle osteotomy when a quadriceps snip does not provide adequate exposure.

Femoral Peel

Technique

Windsor and Insall described the femoral peel as a useful technique for exposing an ankylosed knee. Initially, the proximal medial tibia is exposed subperiosteally, preserving the origin of the medial collateral ligament on the femur. Then, the distal femur is skeletonized by subperiosteal dissection, and the soft tissue sleeve is preserved medially and laterally.

Results

Although the soft tissue sleeve, including the medial and lateral stabilizing structures, is preserved, joint stability may be compromised by the extensive dissection; in these cases, a constrained knee prosthesis implant may be required.

Quadriceps Tendon Rupture

Technique

Fortunately, the incidence of quadriceps tendon rupture after TKA is very low with approximately 1 case per 1,000 reported, as this problem has proven to be difficult to manage. In distinction to the native knee, where primary repair of the tendon to its attachment on the superior pole of the patella with nonabsorbable sutures passed through vertical tunnels in the patella has proven quite successful, the same results have not been achieved in the setting of TKA. Therefore, the preferred technique includes a repair of the tendon with augmentation of the repair using autologous hamstring tissue or a synthetic mesh. First, the distal end of the tendon is resected back to healthy tissue. Then, two No. 5 nonabsorbable sutures are woven through the avulsed tendon in a Krackow-type technique and passed through vertical tunnels in the patella using a suture passer or straight needle. These sutures are then tied with the knee completely extended. An autologous semitendinosus or gracilis graft, or a synthetic surgical mesh, is then used to augment the repair. Postoperatively, the patient is immobilized in full extension for 6 to 8 weeks; motion is then increased by 30-degree increments at 2-week intervals, with the goal of 90 degrees by 3 months postoperatively. During this time, the leg is protected in a hinged brace. Patients with postoperative partial quadriceps tendon rupture after TKA can be adequately managed by the same nonoperative protocol as described for those patients with complete tears who have undergone surgical reconstruction.

Results

Patients with partial quadriceps tears after TKA treated nonoperatively, as described previously, have generally done well, with all seven patients in one recent large series achieving good results. In distinction, nonoperative treatment

for patients with complete quadriceps tears has been poor, with patients generally requiring drop lock knee braces to ambulate. Furthermore, patients with complete tears treated with simple repair alone also had poor results. Six of 10 patients in this group had unsatisfactory outcomes, including four reruptures. Other complications included knee recurvatum and instability, and deep infection. These generally dismal results prompted the changes in surgical techniques noted previously.

Patellar Tendon Disruption

Technique

In association with TKA, postoperative patellar tendon rupture is an uncommon but potentially catastrophic complication that presents a difficult reconstructive problem. Nonoperative management is associated with loss of extensor power in the involved extremity, usually requiring the use of a drop lock brace or walker. Surgical intervention has also been associated with poor results when simple repair has been attempted. More recently, successful salvage of a functional extremity has been made possible with allograft reconstruction using an entire extensor mechanism. The technique is performed through a midline incision when possible. A midline arthrotomy passing over the medial patella allows the native patella to be shelled out from the retinaculum. If the prior patella has fragmented and there are numerous thin fragments of bone embedded in the soft tissue sleeve, these may be left in situ to prevent causing extensive damage to the retinaculum. Distally, over the proximal tibia the native tendon remnants and retinacular tissue should be elevated sub-periosteally, both medially and laterally, for a distance of 7 to 10 cm from the joint line to expose the entire tubercle area. This exposure creates two sleeves of tissue that will be closed over the allograft at the end of the procedure. If revision TKA needs to be simultaneously performed, the procedure

is then performed at this stage. It is important to ensure that component positioning is optimal to minimize stress and shear forces on the graft.

Preparation of the host site begins with creating a trough about 6 cm long and just under 2 cm wide and deep with a burr, saw, or osteotome close to the location of the native tubercle. If possible, a bridge about 1.5 to 2 cm high should be left

between the tibial surface proximally and the proximal end of the trough, although in some revision TKA cases this may not be possible when the bone is deficient. The proximal end of the trough should have an oblique cut that creates a small overhang that will lock in the bone block of the graft and resist proximal migration due to the pull from the quadriceps.

While the host site is being prepared, a second team prepares the allograft. The fresh frozen, extensor allograft must include a tibial bone block that is at least 6 to 8cm long and 2 cm wide and deep, the entire patellar tendon, the patella, and at least 5 to 6 cm of the quadriceps tendon. Distally, a block about 6 cm long and 2 cm wide and deep is prepared. The prepared block should be slightly larger than the trough in the host tibia to facilitate a solid press fit. The proximal end of the block should also have an oblique cut that is directed proximal and posteriorly that will help lock in the block, as previously noted (Fig. 27-6). In the proximal part of the allograft, two No. 5 nonabsorbable braided sutures are stitched through the medial and lateral portions of the quadriceps tendon using a Krackow technique. The free ends of the suture should exit from the proximal end of the quadriceps tendon.

Insertion of the graft begins by weaving two additional No. 5 nonabsorbable sutures into the distal end of the native medial and lateral quadriceps tendon remnants using a Krackow technique. Next, the bone block is gently press fit and tamped into the trough. The proximal portion of the bone block should first be wedged under the oblique step in the native bone, and then the distal end gently impacted to ensure the best fit. If the bone block needs to be trimmed at this stage, either a burr or bone rongeur can be used. Once the block is tapped in place, the block is fixed with two bicortical screws that should be countersunk. Certainly, use of screws creates stress risers in the graft and theoretically increases risk of graft resorption or weakening; however, loss of fixation anecdotally appears to be reduced. Alternatively, two or three 18-gauge wires can be used, as is customary with a standard tibial tubercle osteotomy. However, in these allograft cases, the surgeon must ensure that the proximal lock between the block and trough that is produced by the oblique cuts is solid, or proximal migration may occur. Lack of the native lateral soft tissue attachments that help resist proximal migration of the block with a standard tubercle osteotomy, as well as longer time to union, are perhaps reasons why the risk of migration may be higher in these allograft cases.

After the block is secure, the sutures that were previously woven through the allograft quadriceps tendon are pulled

proximally by an assistant. While under maximum tension, these sutures are passed under, and then up through the native quadriceps tendon remnants. The native tissues are simultaneously pulled distally using the other Krackow-type sutures that were previously woven through the medial and lateral sleeves (Fig. 27-7). While the knee is held in full extension and maximal proximal and distal forces are applied, the sutures that are attached to the allograft quadriceps tendon are tied. Additional No. 5 sutures are then placed into the native and allograft tendon to stitch the graft in place. The native medial and lateral tissue sleeves are then closed over the allograft in a â€œpants over vestâ€• technique using a combination of 0 and No. 1 sutures.

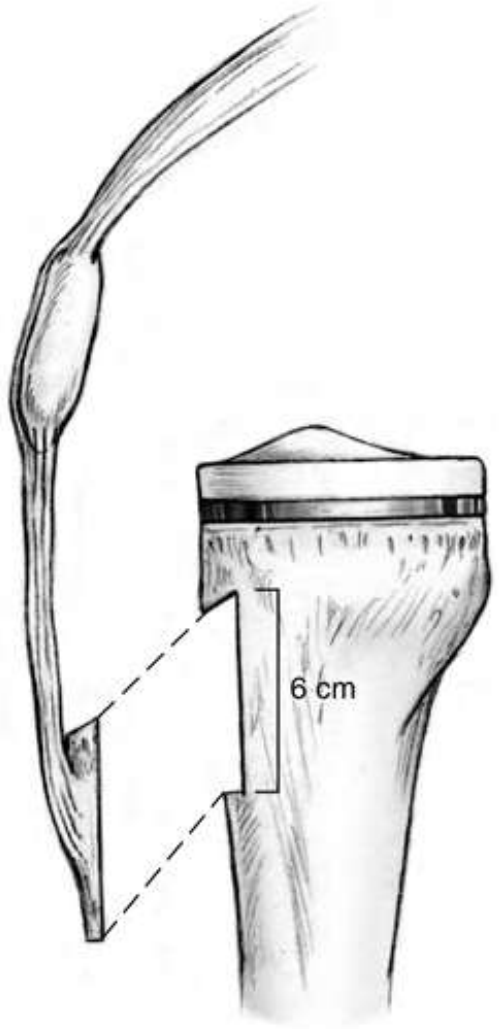

FIGURE 27-6 The trough in the host tibia should be made with a proximal overhang underneath which the oblique cut of the extensor allograft bone block can be wedged.

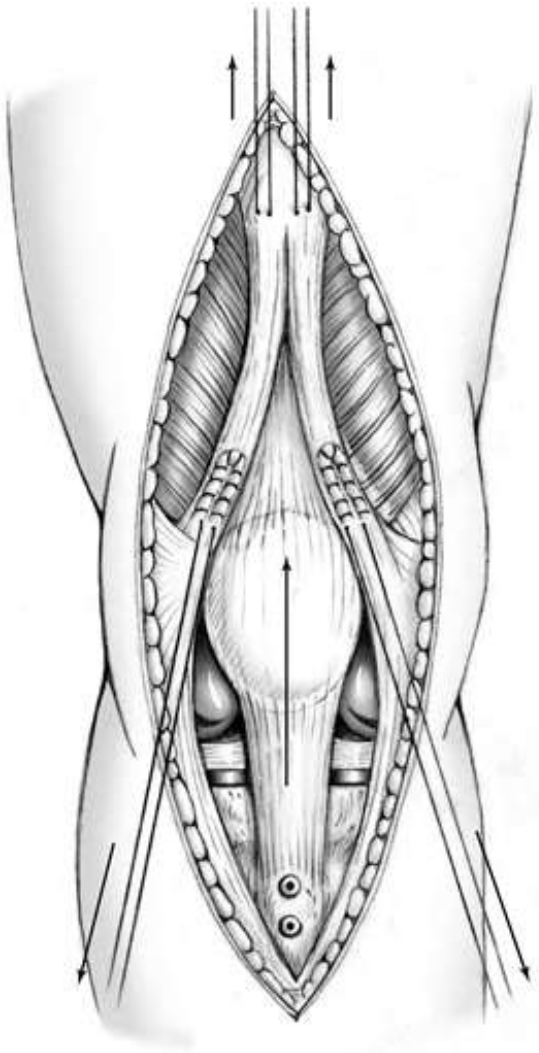

FIGURE 27-7 The extensor allograft should be tensioned and sutures tied with maximal proximal and distal directed forces applied to the allograft and native quadriceps, respectively.

Postoperatively, these patients are immobilized in full extension in a well-fitting brace or cylinder cast for 6 to 8 weeks. Factors including the patient compliance, shape of the leg, and wound care issues must be considered in selecting the optimal method of immobilization. Patients may ambulate with partial weight bearing during this period. After 6 to 8 weeks, flexion is advanced by 30 degrees

every 2 weeks, with the goal of 0 to 90 by 3 months postoperatively. Typically, a hinged knee brace with flexion stops is used to protect the patient during this period and the brace is locked in full extension during ambulation.

Results

The results of simple repair of patellar tendon ruptures after TKA have been poor with high failure rates. Furthermore, early attempts at patellar tendon reconstruction after TKA, using extensor allografts with techniques that tensioned the graft in varying degrees of flexion, have also been associated with high rates of extensor lag. Other complications of all techniques include infection and wound-healing problems. However, knee stiffness has not been a significant concern, and changes in the intraoperative technique, described previously, that emphasize tensioning the graft in

maximal flexion, in conjunction with prolonged immobilization in extension postoperatively, have been associated with better results. Burnett et al reported clinical failures, with an average lag of 59 degrees, in all seven patients in whom the graft was minimally tensioned. In the subsequent 13 patients in whom the graft was maximally tensioned, all were successes with a mean of only 4 degrees lag.

Postoperative Management

A number of factors have been reported to influence postoperative wound healing after TKA. Many of these factors include the same systemic conditions that need to be optimized preoperatively. Similarly, postoperatively, it is imperative that significant pulmonary disease and anemia, in addition to tobacco smoking, are aggressively managed to maintain optimal tissue oxygenation. Furthermore, nutritional requirements must be met, and management of diabetes should maintain blood glucose levels within tight control. When possible, immune-modifying medications should be held until primary wound healing has occurred. In addition to these factors, other factors specific to the postoperative period should be considered, especially in patients with other significant risks. Tight dressings should be avoided as these may compromise local capillary flow. Use of a continuous passive motion machine for extended periods should be avoided in any patient at risk for wound-healing problems, as flexion beyond 40 degrees is known

to reduce the tissue oxygenation along the lateral wound, and greater than 60 degrees also compromises medial tissue oxygenation. Indeed, immobilization should be instituted in patients that develop problem wounds. Development of a significant postoperative hematoma may also reduce tissue oxygenation; in cases where wound-healing problems occur in conjunction with a hematoma, surgical drainage should be carefully considered.

When wound-healing problems occur, the problem should be carefully followed and aggressively managed to avoid secondary bacterial seeding of the joint. Small areas limited to 1 to 2 mm of marginal superficial skin necrosis that involves short segments of the wound (<1 to 2 cm) may be observed and treated with local wound care in the absence of infection. In these cases, immobilization should be used and activity minimized until the margins of the wound have declared themselves. Wound breakdown or full-thickness necrosis should be debrided early, and one of the previously described soft tissue coverage procedures should be used.

Prolonged wound drainage, without wound breakdown, should also be aggressively managed. Significant serous drainage beyond 3 or 4 days should be managed with a compressive (but not tight) dressing, immobilization, and bed rest. Failure to respond within 48 hours should prompt surgical drainage if a seroma is suspected. In cases where there is significant bloody drainage, initial treatment is the same as for serous drainage. However, in this second group of patients, persistent bloody drainage after TKA is suggestive of capsular dehiscence. Failure to respond to non-surgical treatment should prompt a return to the operating room for an evacuation of the hematoma and closure of the arthrotomy. Certainly, substantial drainage beyond a week from surgery should cause concern and requires careful observation and early intervention if quick resolution doesnâ€™t occur.

Summary

In order for successful revision TKA to be accomplished, early wound healing is required. Numerous factors in the preoperative, intraoperative, and postoperative periods can influence this process. While many of these factors are beyond the control of the orthopaedic surgeon, many are not. Careful preoperative evaluation helps identify and manage potential problems to minimize the risks. Furthermore, the orthopaedic surgeon needs to be aware of the plastic and orthopaedic surgery

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 28

The Pedicled Soleus Muscle Flap for Coverage of the Middle and Distal Third of the Tibia

Salvatore C. Lettieri

Steven L. Moran

Soft tissue defects involving the middle and lower third of the leg may occur following trauma, tumor extirpation, and osteomyelitis. Anteriorly the skin and subcutaneous tissue overriding the middle third and lower third of the tibia are thin and exposed bone and tendon may result from soft tissue injury or open fractures. Historically, middle third defects have been covered with the pedicled soleus muscle flap, while attempts have been made more recently to extend to soleus flap to cover defects of the lower third of the leg.

The soleus muscle, according to the Mathes and Nahai schema, is a type II muscle, containing dominant pedicles from the popliteal, peroneal, and posterior tibial arteries and minor segmental pedicles from the posterior tibial artery. The muscle lies in the superficial posterior compartment extending the entire length of the lower leg. The soleus originates from the posterior surface of the tibia, the interosseous membrane, and the proximal third of the fibula. The muscle runs deep to the gastrocnemius muscle in the upper third of the leg (Fig. 28-1). In the middle

third of the leg, the muscle joins with the gastrocnemius muscle and is adherent to the calcaneal tendon. The soleus is a bipennate muscle with the medial and lateral muscle bellies each receiving an independent neurovascular supply; this allows the lateral and medial portions to be mobilized independently while preserving some function within the remaining soleus muscle. The medial head originates from the tibia and receives the majority of its blood supply from the posterior tibial artery. The lateral head originates from the fibula and receives the majority of its blood supply from the perineal artery, although 16% of muscles may be nourished entirely by the posterior tibial artery. The lateral and medial heads are fused proximally while a septum divides the muscle distally. This septum is an extension of the calcaneal tendon and soleus tendon. Dividing the muscle longitudinally at the level of the septum allows for the elevation of the medial and lateral hemi soleus flaps.

In the distal one third of the muscle, the soleus receives segmental arterial perforators from the posterior tibial artery (Fig. 28-2). These distal perforators may be absent in up to 26% of patients; in these cases distal perfusion to the muscle is provided by axial blood flow from more proximal perforators. The diameter and position of these distal perforators is variable but, if present and of large enough caliber, these perforators can allow for a portion of the muscle to be harvested in a reverse

fashion. The muscle may then be rotated 90 to 180 degrees, based on its distal perforators, allowing coverage of the lower third of the tibia and ankle region.

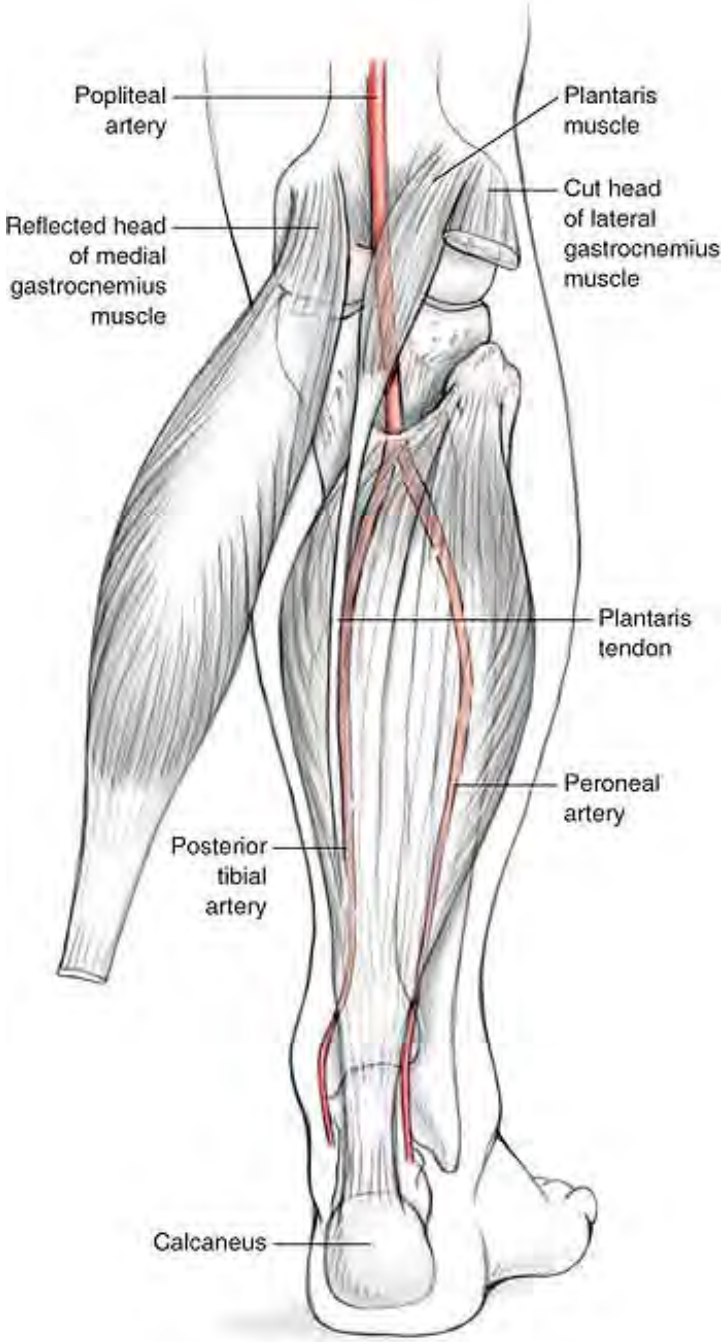

FIGURE 28-1 Schematic anatomy of the posterior leg showing the soleus located deep to the gastrocnemius muscle bellies. The plantaris tendon helps define the plane between the soleus and the gastrocnemius muscle bellies. The blood supply to the soleus is from the popliteal, posterior tibial, and perineal arteries.

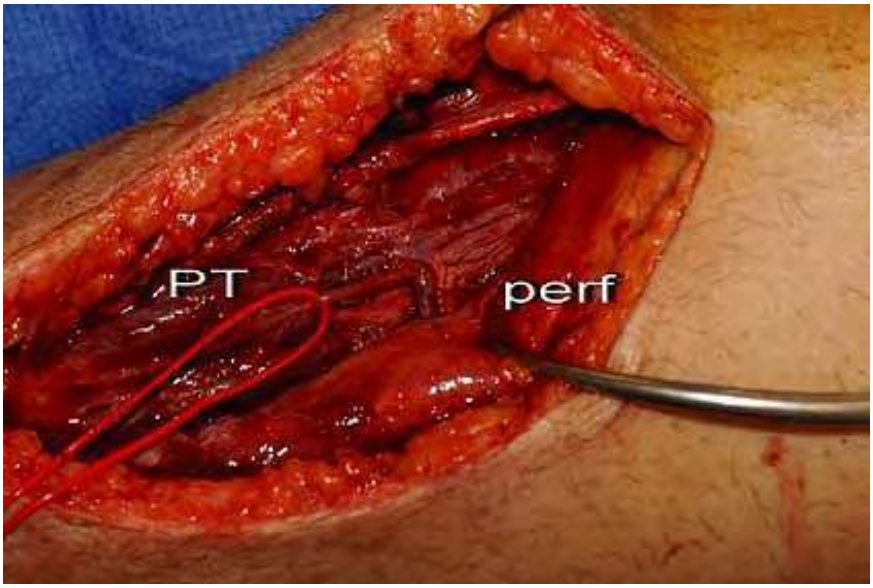

FIGURE 28-2 The posterior tibial artery (PT) as it approaches the ankle gives off several arterial perforators (perfs) to the distal soles muscle. If these perforators are of adequate caliber the muscle can be dissected free and rotated as a distally based flap.

The soleus functions to stabilize the ankle and assist in plantar flexion. The muscle works synergistically with the gastrocnemius and tibialis posterior muscle to provide plantar flexion, while the flexor hallucis longus, flexor digitorum longus, and tibialis posterior all help to provide ankle stability and resist dorsiflexion.

Indications/ Contraindications

The primary indication for the use of the soleus muscle flap is coverage of soft tissue defects in the middle third of the leg. Distal third defects have historically been covered with a free flap; however for small defects and in patients with significant comorbid disease, which would prohibit prolonged anesthesia times, a reverse soleus muscle flap may be considered as an alternative to free tissue transfer.

There are several factors which may prohibit the successful transfer of the soleus muscle and these are:

- Size of the defect
- Status of the muscle
- Status of surrounding tissue and bone
- Size and location of existing perforators

Size of the Defect

The soleus muscle has a limited surface area and a limited arc of rotation (Fig. 28-3). Large defects occupying the majority of the middle third and lower third of the leg are best covered with free tissue transfer. In addition, the distal aspect of the muscle can be unreliable if it must be stretched or inset under significant tension. In such cases, alternative methods of closure should be considered or the soleus may be used in conjunction with another flap.

The soleus can be used in conjunction with the medial or lateral gastrocnemius muscles for larger defects spanning the upper aspect of the lower leg, but this will compromise remaining plantar flexion (Fig. 28-4). When the defect is so large as to require more than just the soleus and medial gastrocnemius flaps, a free flap should strongly be considered as a means of soft tissue coverage and preservation of remaining posterior compartment function. Defects which are to be covered with a reversed soleus muscle flap should be less than 50 cm^2 while the standard soleus flap can cover most defects under 75 cm^2.

Status of the Muscle

Because the soleus muscle is closely adherent to the deep posterior surface of the interosseous membrane, tibia, and fibula, it can often be significantly traumatized following comminuted fractures of the tibia and fibula. During initial wound evaluation and debridement the muscle can often be inspected through the soft tissue defect. If the muscle is significantly lacerated by fracture fragments or contains a significant amount of intramuscular hematoma, it is most prudent to use

another flap for soft tissue coverage. In addition, any associated injury to the popliteal, peroneal, or posterior tibial arteries can adversely affect the survival of the soleus muscle.

Status of Surrounding Tissue and Bone

Preexisting damage to the surrounding skin and deeper tissue of the middle third of the leg are also a relative contraindication to soleus muscle flap use. A history of previous radiation therapy, previous surgery, or penetrating trauma to the area surrounding the middle third of the leg should alert the surgeon to potential problems with the use of this flap. The soleus muscle has been shown to provide a source of collateral arterial flow between the posterior tibial and perineal arterial systems. In patients with vascular occlusions of the perineal or posterior tibial arteries, use of the soleus muscle may further compromise limb vascularity.

Size and Location of Existing Perforators

Indications for use of the soleus for distal third defects are based on the work of Pu and include (a) defect size less than 50 cm, (b) defects located over the anterior or medial portion of the distal tibia,

P.348

P.349

P.350

(c) the presence of large perforators in distal 4 cm to 5 cm of muscle and (d) a soleus muscle which is nontraumatized on initial exploration (4,5,6). If these factors are met a reversed hemi soleus flap may be attempted for lower third defects, otherwise a free flap is chosen for distal third defects. In addition, patients who have a smoking history, peripheral vascular disease, and history of lower leg radiation therapy should be considered high risk for a reversed soleus muscle flap.

FIGURE 28-3 A-C: The soleus muscle (S) runs anterior to the gastrocnemius muscle (G) and to the area just proximal to the medial malleolus. The arc of rotation of the medially based soleus muscle is limited due to its deep origin and broad proximal attachments. This makes coverage of more laterally and inferiorly based middle third defects difficult. The solid red line shows the actual safe arc of rotation for the medially based soleus muscle flap, with the dashed line depicting the superior and inferior margins of the â€œmiddle thirdâ€• of the leg.

A

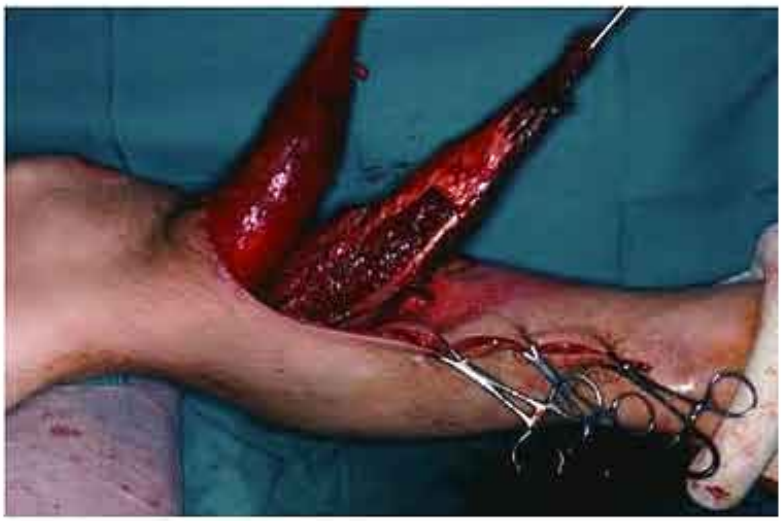

B

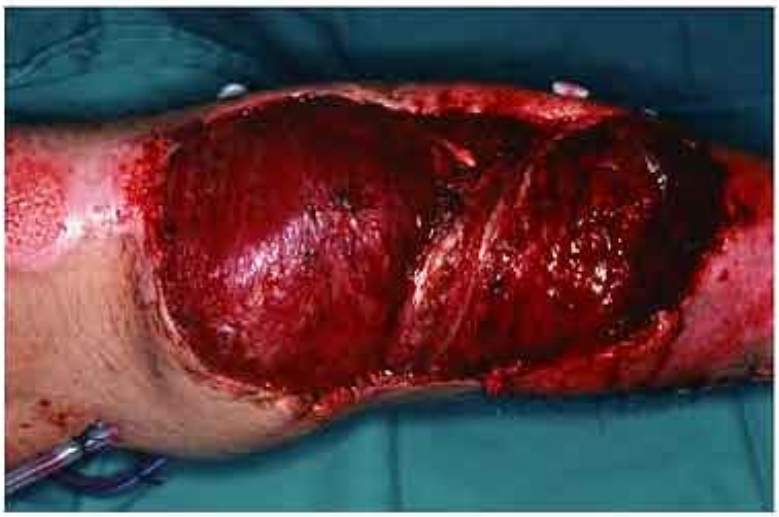

C

FIGURE 28-4 A-C: This large defect is easily covered by elevating both the medial gastrocnemius muscle and the soleus muscle.

Preoperative Planning

Close inspection of the lower leg is necessary prior to surgery. Even though the soleus occupies the same compartment as the gastrocnemius, it can be more significantly damaged in open fractures due to its adherence to the tibia and interosseous membrane. If there is significant ecchymosis or swelling of the posterior compartment, in conjunction with significant displacement of the tibia and fibula on lateral radiographs, one must assume significant damage to the soleus muscle. In this case, free tissue transfer would provide a better alternative for coverage of the middle third of the tibia. If the patient has had previous surgeries, such as fasciotomies for trauma, this may also preclude reliable rotation of the flap, since the level of the fasciotomies and also the location of the injuries may have injured the underlying flap (Fig. 28-5).

A noninvasive vascular exam should be performed on the lower leg, verifying patency of the posterior, peroneal, and anterior tibial arteries. If the patient has a history of significant peripheral vascular disease or long standing diabetes, a CT angiogram may verify patency of the posterior and perineal arteries prior to flap transfer.

Surgery

Patient Positioning and Surgical Preparation

The patient is placed in the supine position for anterior defects, the lateral decubitus position for lateral defects, and prone for posterior defect coverage. As part of the surgical preparation, the entire

lower extremity is prepped and draped in the usual fashion. For anterior midtibial defects, a sterile â€œbumpâ€• is placed beneath the distal thigh region to allow for slight external rotation of the leg and bending of the knee; this can facilitate

identification of the muscle. The muscle may be harvested with the use of spinal or general anesthetic. A tourniquet on the upper leg allows for a relatively bloodless field during muscle dissection.

FIGURE 28-5 This patient had lower leg fasciotomies performed at the time of the original tibial injury. The patient now presents with osteomyelitis and exposed hardware over the middle third of the tibia. The defect is located within the arc of coverage of the medial soleus muscle flap; however, there is significant scarring and injury to the superficial posterior compartment secondary to the fasciotomies and previous split thickness skin grafting. Because of concerns of the reliability of the soleus muscle in this situation, this patient underwent coverage of exposed hardware with a free tissue transfer.

Technique

Medial and proximally based soleus flap

The soleus is harvested through an incision which runs from the upper third of the

leg to just above the medial malleolus. The incision is made 2 cm to 3 cm posterior to the medial palpable margin of the tibia. The defect site is incorporated into this incision (Fig. 28-6). Skin bridges are avoided. Dissection is carried down to the investing fascia of the superficial posterior compartment. The saphenous vein is identified and preserved during the dissection. The deep fascia is then opened longitudinally to expose the posterior compartment.

The muscle itself is most easily identified proximally deep to the gastrocnemius muscle. The plane between the two muscles is relatively avascular with the exception of some small perforators which may be ligated. The plantaris runs in the plane between the two muscles and can be used as a landmark in cases of severe trauma where hematoma may obscure the tissue planes.

The soleus fuses with the calcaneal tendon in the middle third of the leg. Here, sharp dissection is required to separate the soleus from the common calcaneal tendon as it extends towards the heel.

P.352
P.353
P.354
P.355

This dissection is aided by placing medial traction on the edge of the Achilles tendon with the use of several clamps (Fig. 28-7). In the distal aspect, the muscle readily separates from the calcaneal tendon. Dissection from the calcaneal tendon is carried to the midline raphe for hemisoleus elevation. The midline septum does not extend the entire length of the muscle but should be used as guide for hemisoleus elevation. Once the soleus is elevated from the calcaneal tendon, the space between the gastrocnemius muscle and the soleus muscle is easily entered and there is a transverse junction point which is easily cut with scissors or a knife. This will free the entire posterior aspect of the soleus muscle.

FIGURE 28-6 A: A middle third defect in the right leg of a 66 year old diabetic woman following a Gustilo type IIIB fracture of the tibia. The fracture has been stabilized with an intramedullary rod and the wound has been covered with an antibiotic beads pouch following initial debridement at the time of fracture fixation. B: An incision is made slightly posterior to the medial tibial margin and then carried inferiorly. The open wound is incorporated into the incision to avoid creating a skin bridge. C: The soleus muscle (S) has been dissected free of the calcaneal tendon (C) and separated from the gastrocnemius muscle (G) and flexor digitorum longus muscle (FDL). D: Large perforators from the posterior tibial artery can be seen entering the muscle. E: The muscle is now divided at its distal attachment and pedicled to cover the open defect. Because the soleus is elevated off the common calcaneal tendon the superficial surface of the muscle is devoid of fascia. The muscle is â€œfanned outâ€• and inset with half buried absorbable sutures. F: The donor incision is closed up to the rotated muscle and the exposed muscle is covered with a split thickness skin graft. G,H: At 2 months the patient has a well healed wound with excellent contour.

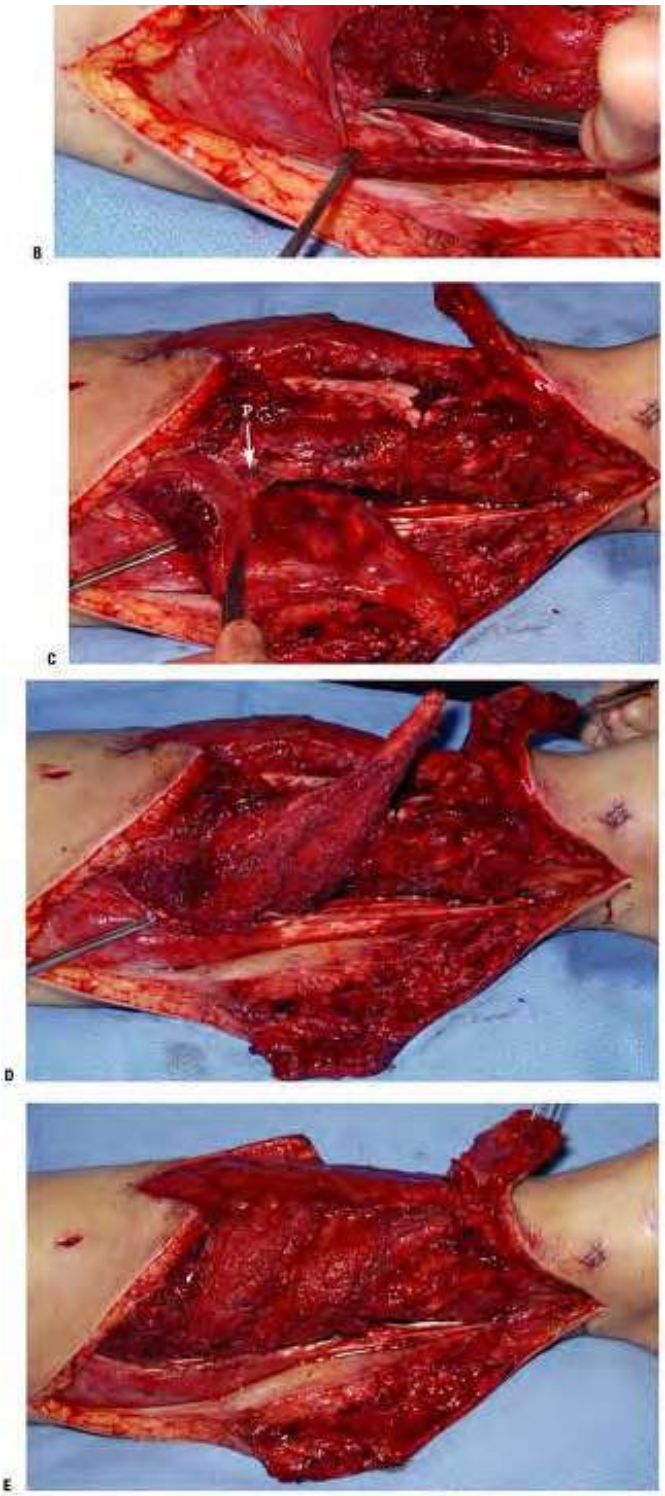

FIGURE 28-7 A: Picture of a middle to lower third Gustilo IIIB injury following a motor vehicle accident in a 58 year old man. The soleus has been separated

from the flexor digitorum longus as well as the gastrocnemius muscle. The muscle has just been divided distally and a stay suture has been placed in the distal aspect of the muscle. B: A Carroll elevator is used to facilitate dissection of the soleus muscle off the common calcaneal tendon as medial traction is applied to the edge of the tendon with Alice clamps. C: The muscle is dissected proximally until enough length is available for defect coverage. Large arterial perforators from the posterior tibial artery are preserved to perfuse the flap (p). D,E: The flap is then pedicled into position and covered with a split thickness skin graft.

The posterior tibial artery is identified in the plane between the soleus and flexor digitorum longus. Inferiorly a finger can be passed beneath the soleus muscle but superficial to the posterior tibial vessels. A cautery is then used to separate these medial attachments of the soleus to the tibia. Once the medial aspect of the muscle is mobilized, the space between the soleus and the deep posterior compartments is readily opened with blunt dissection and the lateral aspect of the muscle is identified.

Once the posterior tibial artery and nerve are identified and the superficial and deep attachments of the muscle have been mobilized, the muscle may be divided distally. Distal perforators from the posterior tibial artery and perineal vessels are then sequentially ligated to allow for mobilization of the muscle. There can be multiple small diameter perforators off the posterior tibial artery; these vessels should be clipped or ligated and not cauterized, as thermal injury may be propagated to the posterior tibial vessels.

Once the muscle has been completely elevated, it is rotated into position to cover the defect. Generally, the fascia on the deep surface is left attached to the muscle, but there will be no fascia on the superficial surface. Multiple half buried absorbable sutures can be used to â€œfan outâ€• the muscle and inset it into the defect site. The distal incision is closed over a drain up to the point of muscle rotation. The muscle should not be tunneled as this can contribute to distal venous congestion within the flap (Figs. 28-6 and 28-7).

The muscle flap is then immediately covered with a meshed skin graft taken from the upper thigh. Vaseline impregnated gauze or a Xeroform dressing are used to cover the skin graft. The leg is then placed in a large Robert-Jones type dressing with the addition of a posterior splint. A window is left in the dressing overlying the muscle so flap checks may be performed while the patient is recovering on the ward.

Lateral and proximally based soleus flap

Though the arc of rotation for this muscle is limited, the lateral approach can be advantageous for some lateral middle third defects. The incision is made just inferior to the lateral border of the fibula. The deep fascia is incised just below the fibula and the plane between the gastrocnemius muscle and soleus are created with blunt dissection. Distally the soleus is again sharply dissected from the calcaneal tendon. Deep proximal dissection involves separating the soleus from its tough attachments to the fibula. During proximal dissection one must be cognizant of the common perineal nerve running close to the fibular head. Once elevated, the muscle is inset as previously described.

Reverse soleus muscle flap

The reverse soleus muscle flap modification (Fig. 28-8) is used to cover small distal defects over the medial anterior aspect of the tibia or medial superior aspect of the medial malleolus (Fig. 28-9). The success of this operation is predicated on the presence of adequate caliber distal perforators from the posterior tibial artery. If during the surgical procedure these perforators are injured or are of insufficient quality the procedure must be abandoned and the defect should be covered with a free tissue transfer. A preoperative angiogram has been recommended by some authors to verify the position of the distal perforator prior to surgery.

Dissection is performed under tourniquet control. The same incision is made as for the proximally based soleus flap (Fig. 28-9). The incision is 2 cm medial to the medial border of the tibia. The existing wound is incorporated into the incision. The medial portion of the soleus is identified as described above and is separated from the gastrocnemius muscle, calcaneal tendon, and the flexor digitorum longus muscle. The posterior tibial artery is identified and the distal perforators are

examined. As many distal perforators should be preserved as possible but the authors have had success with the preservation of one or two perforators alone if they are of adequate caliber (vein greater than 1.5 mm and artery of 1 mm or greater). Once the perforatus are determined to be of adequate size, the soleus is then split at the level of the central raphe and divided at the junction of the proximal and middle third. The muscle is then divided longitudinally using the cautery or scissors until the inferior perforator is reached. The muscle is then rotated 90 degrees to 180 degrees. Additional arterial pedicle dissection may be required to prevent vessel kinking. The flap is then inset into the defect with half buried mattress sutures and covered with a split thickness skin graft (Fig. 28-9).

Postoperative Management

The patient should be maintained on bed rest in a posterior splint or knee immobilizer for one week. The bolster dressing may be removed on postoperative day 5 to ensure adequate skin graft â€œtake.â€• Until this time the flap ay be monitored through a window in the postoperative dressing. If the skin graft is adherent to the underlying muscle, the patient may begin to mobilize and bear weight as tolerated, barring any underlying fractures. Gentle range of motion should then begin, increasing the motion by 5 or 10 degrees per day. Ambulation can begin on the 10th post-surgical day. If the

P.357

P.358

P.359

P.360

flap is inset under tension or if the healing environment is less than ideal, the knee immobilizer can be maintained for an additional week to prevent undue tension on the margins of the muscle flap. The drains are removed once drainage is less than 30 cc a day.

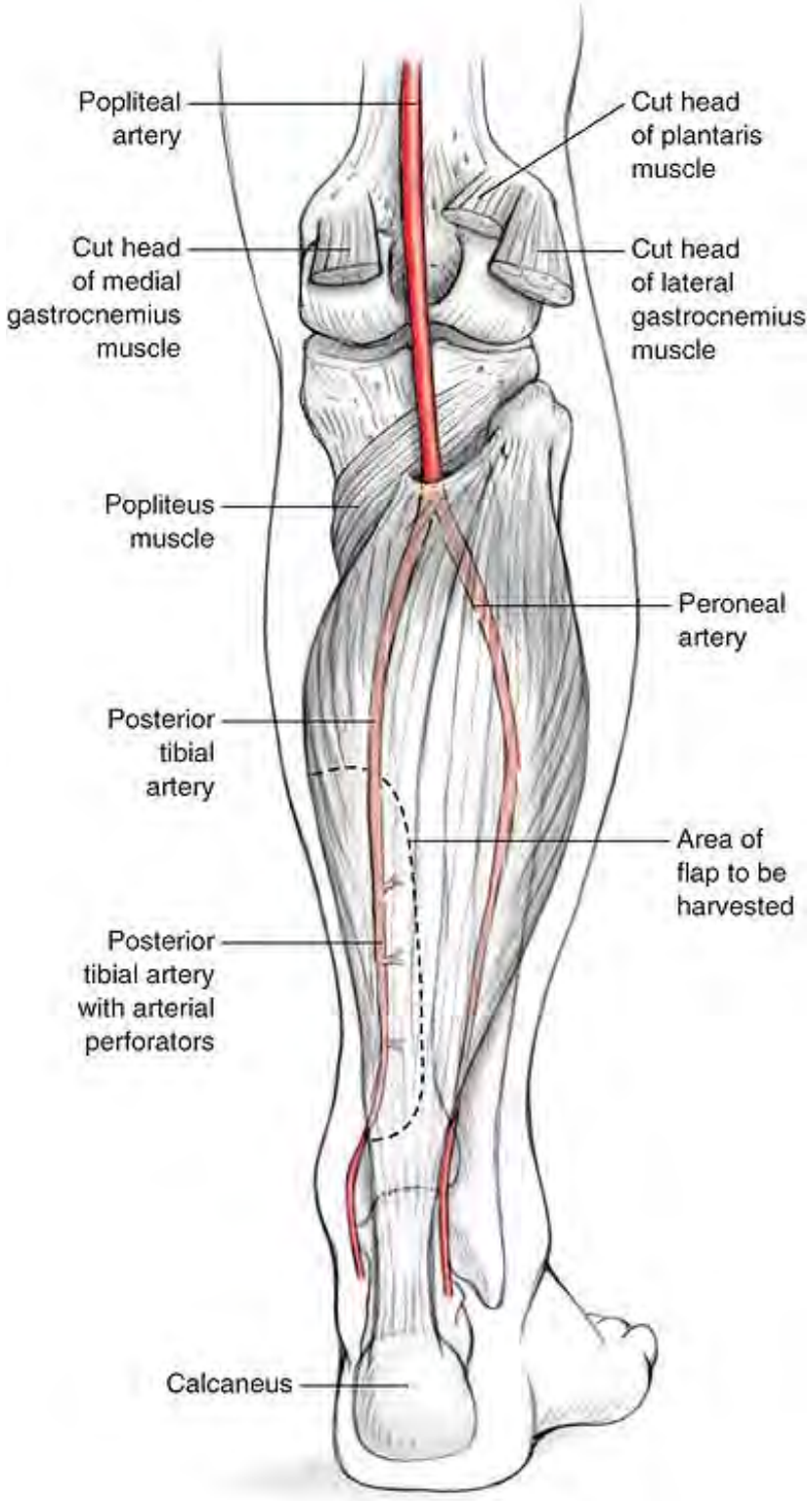

FIGURE 28-8 Schematic drawing of the distally based reverse soleus flap. The

flap's arterial supply is from perforators found at the distal portion of the posterior tibial vessels. If present these vessels may be used to supply a strip of soleus muscle which is cut from the medial margin of the soleus muscle.

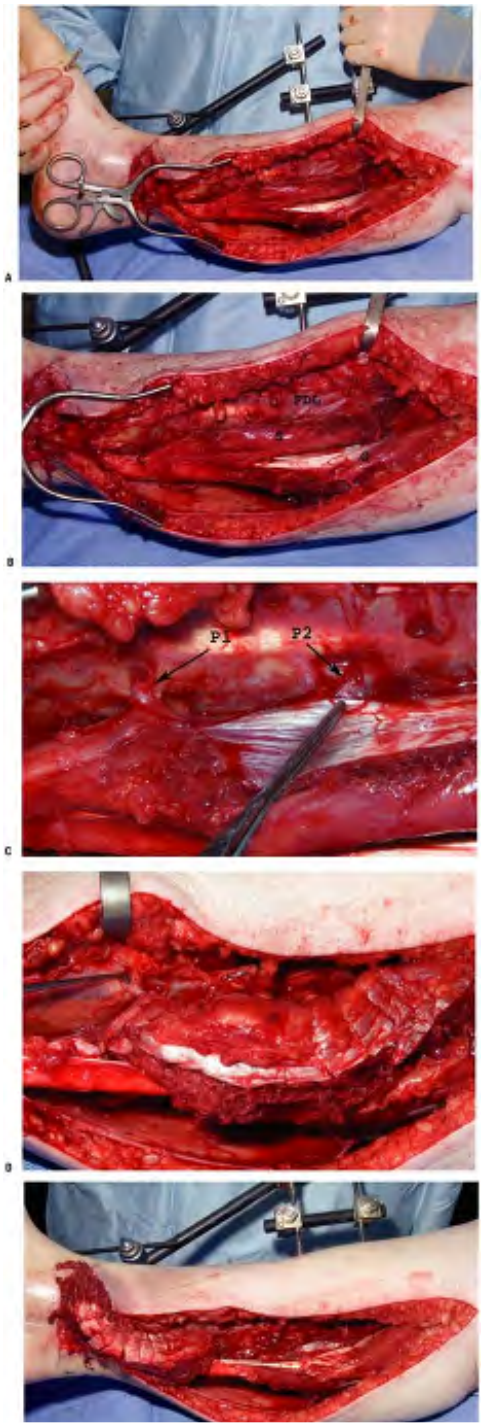

FIGURE 28-9 A: A 61 year old female with a distal defect overlying the superior medial margin of the malleolus. An incision was made at the medial posterior margin of the tibia. B: The gastrocnemius (G), soleus (S) flexor digitorum longus (FDL) and calcaneal tendon (C) are identified. C: Exploration of the posterior tibial vessels, in preparation for free tissue transfer, revealed that the patient had two (p1 and p2) large distal perforators to the soleus muscle. Because of this the patient was felt to be an excellent candidate for a distally based soleus muscle flap. D: A medial strip of soleus is then elevated from the common calcaneal tendon, gastrocnemius muscle and flexor digitorum longus muscle. The nerve hook points to the preserved distal perforator. E: The muscle is then rotated 180 degrees to cover the distal defect. Care is taken to dissect the proximal perforator back to its origin on the tibial artery to prevent kinking of the artery or vein. F,G: The muscle is covered with a split thickness skin graft and the remaining portion of the external fixator is attached to provide needed stabilization.

Once the original dressing is removed from the skin graft, dressings may be changed daily. This should include the application of topical antibiotic ointment and nonadherent gauze followed by a lightly compressive wrap to minimize edema.

Dependent leg position is limited to 30 minutes an hour for the first 2 weeks to minimize edema. On the third postoperative week the patient is fitted for compressive stockings which help the patient continue to manage lower extremity edema and help in the flap contouring.

Rehabilitation

Knee and ankle motion may begin once the skin graft is adherent to the underlying muscle bed. Weight bearing status is determined by the stability of the underlying fractures.

Results

In a study by Hallock of 29 soleus flaps, 24 of 29 flaps were used for coverage of high energy impact defects. All soleus muscle flaps in this study were based on a proximal pedicle. Complication rates were low (13.8%) and there were no cases of total flap loss (1,2). Similar results were reported by Pu, who found no cases of total flap loss when the flap was based on a proximal pedicle for coverage of middle third tibial defects (3,4,7,8).

Complications

Major complications include total and partial flap necrosis. Total flap loss can occur from injury to the vessels at the time of dissection, use of a flap which has been significantly injured with the surrounding bony trauma, or use of the flap in a situation where the posterior tibial vessels and peroneal vessels were compromised or injured. In such cases a free tissue transfer is often needed for leg salvage.

Partial necrosis most often occurs at the distal most margin of the flap. If the partial flap necrosis results in exposure of vital structures, another flap will be required for coverage; most commonly a free flap is used but local fasciocutaneous advancement flaps may be used to cover smaller defects. If the bone is completely covered and the defect is just along the periphery, the wound may be debrided and treated with dressing changes. The remaining wound can then heal through secondary intention.

Skin graft loss may occur due to infection or sheering. Such cases are treated with dressing changes until the underlying muscle bed appears capable of accepting

another skin graft. Other minor complications include hematoma in the donor bed and injury to the tibial nerve or the posterior tibial vessels during flap dissection. Both complications may be avoided with meticulous hemostasis and clear identification of the anatomical landmarks prior to muscle division. Limitations in plantar flexion are minimized if the muscle is taken in isolation.

Recommended Reading

Beck JB, Stile F, Lineaweaver W. Reconsidering the soleus muscle flap for coverage of wounds of the distal third of the leg. *Ann Plast Surg*. 2003;50;631â€“635.

Kauffman CA, Lahoda LU, Cederna PS, et al. Use of the soleus muscle flaps for coverage of distal third tibial defects. *J Reconstr Micro.* 2004;20:593â€“597.

References

1. Bos GD, Buehler MJ. Lower Extremity local flaps. *J Am Acad Orthop Surg.* 1994;2:342â€“351

2. Hallock GG. Getting the most from the soleus muscle. *Ann Plast Surg.* 1996;36;139â€“146.

3. Mathes S, Nahai F. Reconstructive Surgery: Principles, anatomy and technique, 1st ed. New York: Churchill Livingstone; 1997:1473â€“1487.

4. Pu LLQ. Soft tissue reconstruction of an open tibial wound in the distal third of the leg. *Ann Plast Surg.* 2007;58:78â€“83.

5. Pu LLQ. The reversed medial hemi soleus muscle flap and its role in reconstruction of an open tibial wound in the lower third of the leg. *Ann Plast Surg.* 2006;56:59â€“64.

6. Pu LLQ. Medial hemi soleus muscle flap: a reliable flap for soft tissue reconstruction of the middle third tibial wound. *Int. Surg.* 2006;91:194â€“200.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 29

The Sural Artery Flap

Michael Sauerbier

Thomas Kremer

Indications/Contraindications

Soft tissue defects of the lower one-third of the leg and the calcaneal region remain crucial issues. This area is easily susceptible to trauma, and defects of this area are commonly experienced in orthopaedic and trauma surgery.

This region is characterized by tightness and limited mobility of the skin and, frequently, by poor circulation. Thus, chronic and diabetic ulcers, pressure sores, unstable scars, or chronic infection following Achilles tendon rupture or trauma surgery often occur in the distal parts of the leg. Furthermore, exposure of viable structures such as neurovascular bundles, tendons, bone, or osteosynthesis material is a common clinical occurrence due to the superficial course of these structures in the lower one-third of the leg.

In these patients, vascularized soft tissue reconstruction is necessary if further soft tissue damage, with subsequent infection and a potential risk of osteitis, is to be prevented. Relatively few procedures for cutaneous coverage have demonstrated real effectiveness and an acceptable morbidity. In recent years, free microvascular tissue transplantation has become a reliable option among patients suffering from acute or chronic defects in the lower one-third of the leg. Free flaps such as the gracilis flap or

the latissimus dorsi flap, as well as fasciocutaneous flaps (e.g., anterolateral thigh flap, parascapular flap, lateral arm flap, and free groin flap) are frequently performed for limb salvage. Despite the fact that free microvascular transplantation possesses the disadvantages of extensive surgery, sophisticated equipment, general anesthesia, and high costs, this procedure often remains the only option for reconstruction of large defects in this region for limb salvage (Fig. 29-1).

For the coverage of smaller defects, alternative options such as pedicled flaps (e.g., flexor digitorum communis flap, medial plantar flap, [hemi-] soleus muscle flap, abductor hallucis flap, peroneal brevis flap, and lateral supramalleolar flap) have been developed (Table 29-1). In 1992, Masquelet et al. published their experimental work involving skin island flaps supplied by the vascular axis of sensitive superficial nerves. The distally based sural artery flap, which is perfused by reverse flow through the anastomosis between the superficial sural artery and the lowermost perforator of the peroneal artery, forms part of this group. One of the advantages of this thin fasciocutaneous flap is that it permits skin coverage with ideal contouring. Other advantages are that it can be performed quickly, the soft tissue coverage is durable, postoperative discomfort and donor site morbidity are minimal, and it is unnecessary to sacrifice major vessels. This flap has been used for the successful coverage of defects of the posterior and inferior surface of the heel, the Achilles tendon, the middle and distal one-third of the leg, and the dorsum of the foot and the lateralâ€”as well as medialâ€”malleolus (Tables 29-2 and 29-3; Figs. 29-2,29-3,29-4).

The sural artery flap is insensate and is not an option in cases in which sensitivity is an issue. Furthermore, the sacrifice of the sural nerve results in hyposensitivity of the lateral border of the foot.

Because perfusion of large sural artery flaps (> 9Ã— 12 cm) is unreliable, it is not deemed suitable for coverage of extensive defects of the lower one-third of the leg. Moreover, higher complication rates have to be anticipated in patients with comorbidities such as peripheral artery disease, diabetes mellitus, and venous insufficiency. The sural artery flap is contraindicated in patients with destruction of the vascular pedicle or the lowermost perforator of the peroneal artery (see Table 29-2).

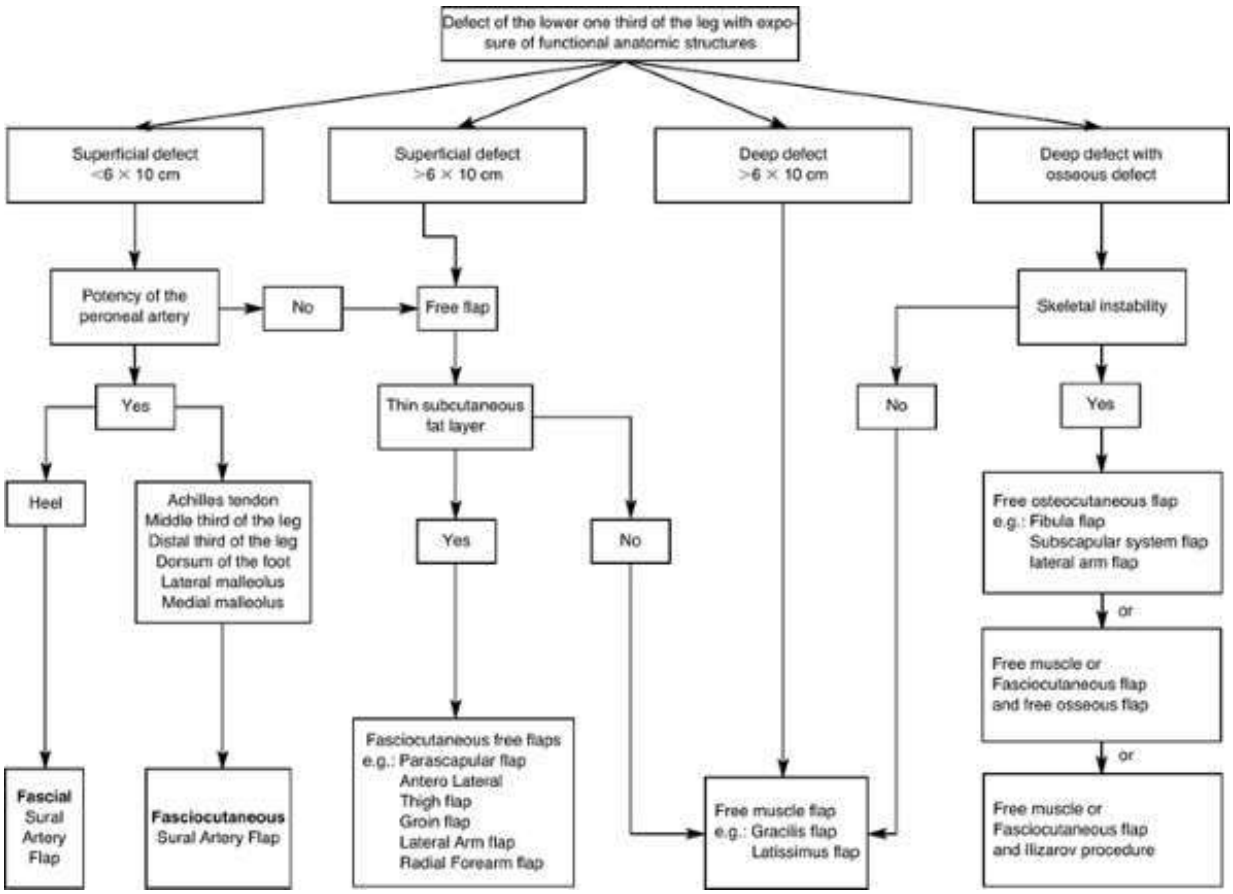

FIGURE 29-1 Treatment algorithm. Small superficial defects of the lower one-third of the leg can be treated by a sural artery flap if the defect size is smaller than 6 Ã— 10 cm. Larger superficial defects can be reconstructed by fasciocutaneous free flaps, whereas deep defects, especially in combination with osseous, defects should be reconstructed by muscle flaps. Defects in the weight bearing area of the heel should be reconstructed by fascial sural artery flaps to prevent pathologic movement of the subcutaneous layer. For alternative pedicled flaps to the lower limb, see Table 29-1.

Table 29-1. Flaps and Their Indications According to Location

Flap	Indications
Soleus flap	Middle third of the leg
Flexor digitorum communis flap	Distal third of the leg
Flexor hallucis longus flap	Distal third of the leg
Peroneus flap	Achilles tendon
Lateral supramalleolar flap	Dorsum of the foot/heel (non-weight bearing)
Dorsalis pedis flap	Foot/distal third of the leg
Extensor digitorum brevis flap	Premalleolar area/dorsum of the foot and toes
Medial plantar flap	Heel (weight bearing)

Table 29-2. Advantages and Disadvantages of the Sural Artery Flap

Advantages	Disadvantages
Constant anatomy	Increased morbidity in patients with comorbidities
Thin flap with ideal contouring	Sacrifice of the sural nerve
Quick procedure	Susceptible to venous congestion
Durable soft tissue coverage	Insensate flap
Low donor site morbidity	Split thickness skin grafts for donor site closure in large flaps (>4 cm)
Acceptable donor site scar in small flaps (<4 cm)	
No sacrifice of major vessels	

Table 29-3. Indications for Sural Artery Flap Reconstruction

Healthy patients with no comorbidities (diabetes mellitus, peripheral artery disease, venous insufficiency)

Skin and soft tissue defects up to 9 Ã— 12 cm:

- Chronic and diabetic ulcers
- Pressure sores
- Chronic infection
- Traumatic defects
- Unstable scars
- Exposure of viable structures (neurovascular bundles, tendons, bone, osteosynthesis material)

One of the following locations:

- Middle and lower third of the leg
- Posterior surface of the heel
- Inferior surface of the heel
- Achilles tendon
- Dorsum of the foot
- Lateral malleolus
- Medial malleolus

FIGURE 29-2 Male patient suffering from a combined soft tissue and osseous

defect in the heel (A). After debridement, soft tissue reconstruction was planned by a sural artery flap (B). Flap transposition was performed to the defect (C) after flap harvest (D). Split thickness skin grafting of the donor site was necessary. Postoperative course and wound healing were uneventful (E,F).

FIGURE 29-3 A 34-year-old man with a chronic ulcer of the dorsum of the foot. A sural artery flap was planned to reconstruct the defect with exposed tendons (A,B). After complete flap harvest, the skin island is transposed to the defect (C,D). The skin over the pedicle is incised and split thickness skin grafted to reduce pressure on the nutrient vessels (E,F). The functional and aesthetic results are excellent 2 weeks post surgery (G) The functional and aesthetic results are excellent 2 weeks post surgery (H,I)..

FIGURE 29-4 A 36-year-old woman with osteitis (above left). After debridement and application of an antibiotic chain (above right). Defect reconstruction was planned by a sural artery flap (upper left). After complete flap harvest, the pedicle is visible through the cural fascia, which has to be included in the flap (upper right). After transposition to the defect, wound healing was uneventful (below).

Preoperative Planning

A thorough clinical and radiologic examination of the donor and the recipient sites is mandatory, and it is first necessary to estimate the size of the recipient site defect. There are no studies regarding maximum flap dimensions with regard to safety, but small and moderate-sized defects can usually be covered satisfactorily. Defects up to 9 Ã— 12 cm have been successfully reconstructed using sural artery flaps. Radical debridement may result in an enlargement of the defect, and this should be taken into account in the preoperative planning.

An examination of the patient's peripheral function should be performed to determine the total extent of the injury. Depending on the size and location of the defect, various functional structures may be exposed or destroyed. Changes in the normal resting position of the foot and toes will help identify the loss of continuity of lower extremity tendons, and traditional functional tests will confirm the loss of active range of motion. A careful examination of sensitivity and muscle function helps to evaluate the extent of the defect in terms of destruction of sensor or motor nerve components. Additionally, peripheral perfusion should be ascertained by simple palpation of pulses or by Doppler probe. If in doubt, an angiogram should be performed. If any of these structures have been destroyed, additional reconstruction procedures must be planned. In patients suffering from chronic or infected wounds or diabetic gangrene, additional osteitis or osteomyelitis must be considered and conventional radiographs and/or magnetic resonance imaging (MRI) scans can help to prove the existence or extent of altered osseous structures.

Table 29-4. Preoperative Planning for Sural Artery Flap Reconstruction

Clinical examination
- Defect size (after debridement)
- Peripheral function (to determine destruction of functional structures [nerves, tendons, blood vessels])

MRI or conventional radiographs of the recipient site (fractures, extension of osseous or soft tissue defects, osteitis)

Examination of patency of the lesser sapheneous vein and peroneal artery
- Doppler ultrasound probe in healthy patients
- Conventional angiogram or MRI-angiogramm in comorbid patients

Detailed information of the patient
- Sacrifice of the sural nerve (hyposensitivity of the lateral border of the foot)
- Neuroma formation
- Split thickness skin grafting of the donor site

The sural artery flap pedicle should be examined preoperatively. In all patients the patency of the peroneal artery and the lesser saphenous vein is examined by palpation of pulses and Doppler probe. Additionally, the lowermost perforators between the peroneal artery and the flap pedicle can be examined 4 to 7 cm proximal to the tip of the lateral malleolus. In patients with comorbid conditions such as diabetes mellitus and peripheral vessel disease, flap perfusion may be particularly unreliable. In these patients, a conventional angiogram or an MRI-angiogram can be helpful in preoperative planning procedures.

Finally, the surgeon should provide the patient with a detailed account of the surgical procedure. Information regarding the postoperative function, the possibility of persisting complaints, and donor site morbidity should be given. In particular, the loss of sensitivity of the lateral border of the foot and the possibility of neuroma formation as well as the potential necessity of donor site closure by split thickness skin grafting

should be discussed (Table 29-4).

Surgery

Anatomy

The reverse flow sural artery flap is based on the vascular network along the sural nerve. The nutritient artery is the superficial sural artery (SSA), which arises from the popliteal artery proximal to the medial and lateral sural arteries in about 65% of cases, directly from the medial (20%) or lateral (8%) sural artery, or sometimes from the common stem of origin of the two. The artery courses posteriorly for 2 to 3 cm before joining the medial sural nerve descending between the two heads of the gastrocnemius muscle. The medial sural nerve anastomoses with the communicating branch of the lateral sural nerve to become the sural nerve. Both the medial sural nerve and the SSA pierce the crural fascia at the junction of the proximal and middle thirds of the leg to become subcutaneous. The SSA courses alongside the sural nerve to the lateral malleolus in 65% of all patients; in 35% of cases the artery fades into a vascular net at the distal one-third of the leg. Regardless of the termination, the SSA has a constant distal anastomosis from the lateral malleolar arteries, which in turn arises from the peroneal artery. The distally based sural artery flap is dependent on this anastomosis, and its perfusion is reverse flow through the SSA and its cutaneous branches at the distal two-thirds of the leg.

The venous drainage of the sural artery flap is dependent on small commitant veins and mainly on the short saphenous vein which accompanies the sural nerve at the distal two-thirds of the leg. Theoretically, a reverse flow in this vein is impossible due to of the presence of valves; however, clinical findings disprove this theory. Duplex scans of the small saphenous vein generally show a continuous or phasic reverse flow postoperatively. Several theories exist regarding these findings. Some authors stress the existence of bridges between the short saphenous vein and the commitant veins, thus bypassing the valves. Moreover, denervation of the short saphenous vein due to the surgical procedure and increased venous pressure due to the altered flow are thought to be responsible for venous dilatation, which renders the valves insufficient.

Overall, the sural artery flap exhibits a constant anatomy with reliable arterial perfusion and venous drainage.

Surgical Procedure

Surgery is performed in general anesthesia; however, in elderly patients or because of comorbidities, epidural or spinal anesthesia is also possible. Depending on the defect location, the patient is placed in prone position or in a ventral or lateral decubitus position.

The recipient site is first prepared. In patients with chronic wounds, infection, diabetic gangrene, or wound colonization, a radical debridement is mandatory. Analogous to the principles of oncologic resections, a complete resection of the entire infected or altered tissue must be performed. An insufficiently radical debridement renders the reconstructive attempts unsuccessful; therefore, compromises are unacceptable. If, after debridement, the recipient defect is too large for a sural artery flap reconstruction, a temporary wound closure (e.g., vacuum-assisted closure [VAC] therapy) should be performed and other therapeutic options such as free flaps should be planned. For this reason, it can be useful to include the possibility of free flap reconstruction in the initial preoperative patient briefing so that a one-step reconstruction can be carried out.

The authors prefer to perform flap harvest under pneumatic tourniquet (500 mm Hg) to make visualization of the anatomy easier. In addition, the surgeon should use 2.5 to 4 times magnification loupes.

As mentioned previously, the short saphenous vein and the arterial perforators of the peroneal artery on the lateral aspect of the distal leg have to be located using a Doppler ultrasound probe. Following debridement, the defect size is measured and a skin island with adequate dimensions is planned with its center along the line of the short saphenous vein. The proximal limit of the flap should not exceed a boundary of 20 cm proximal to the lateral malleolus, due to the fact that the portion of the flap that exceeds the suprafascial (subcutaneous) portion of the nerve behaves like a random extension, and thus its reliability is unsure. The skin island is planned in the middle or distal third of the leg, according to the pedicle length required (Fig. 29-5). Because of its thicker subcutaneous tissue, the skin island should be designed to be slightly larger than the recipient defect to facilitate skin-to-skin closure of the defect. The maximum size of the skin island should not exceed 9 Ã— 12 cm. The pivot point is

marked 5 cm above the tip of the lateral malleolus to include the branches that anastomose with the peroneal artery. The skin incision starts distally along the line of the pedicle to localize the vein. If the skin island is not centered over the short saphenous vein, correction of the flap position is still possible. When the pedicle has been located, the dissection proceeds in the proximal extreme of the flap; again, the short saphenous vein and the sural nerve are identified suprafascially. At this point, the short saphenous vein, the suralis nerve, and the SSA are ligated, divided, and included into the flap. The proximal stump of the nerve is coagulated and buried in the surrounding gastrocnemius musculature to avoid a troublesome neuroma over the donor area. Some authors report raising the flap and sparing the sural nerve to prevent neuroma formation and foot hyposensitivity. In our opinion this strategy is too risky for the vascular supply of the flap and is unnecessary because the morbidity of harvesting the flap with the sural nerve is minimal. In the literature, several series show no significant alteration in foot sensitivity, and in most cases, patients exhibit improved sensitivity over time.

Dissection proceeds distally at the plane below the deep fascia, so that inclusion of the pedicle in the flap is easily accomplished and the nutrient anastomosis of the flap can be preserved. Two to three constant and direct perforators from the gastrocnemius muscle to the vascular axis of the sural nerve are identified and ligated with ligature clips, because electrocoagulation of these perforators increases the risk of pedicle destruction. The fascial and subcutaneous pedicle, which includes the SSA, the sural nerve, and the short saphenous vein, should be at least 3 cm in width. The paratenon of the sural triceps muscle should be carefully preserved, especially when split thickness skin grafting of the donor site defect is necessary. The flap procedure remains simple if care is taken to avoid dissection too close to the pedicle and if dissection is only extended to the demarcated limit of 5 cm above the tip of the lateral malleolus. During elevation of the flap, near this inferior limit of dissection, efforts should be made to preserve the large perforating septocutaneous vessels.

When flap harvest is completed, transposition of the skin island to the recipient defect can be performed. Under normal conditions (e.g., a sufficient amount of elastic skin), the flap can be transposed through a subcutaneous tunnel. However, when there is any sign of tight skin or venous congestion, a decompression of the pedicle must be performed. The risk of pedicle compression increases with edema formation, surrounding scars, or skin induration and/or thick and inelastic skin. In patients with

this risk profile, two approaches should be considered. First, skin extensions (the â€œcutaneous tailâ€•) overlying the pedicle can facilitate defect closure and reduction of pressure on

the pedicle after transposition. Second, the bridge over the subcutaneous tunnel can be incised and a split thickness skin graft can be used to cover the pedicle. In the literature, the possibility of using the â€œdistally based sural artery interpolation flapâ€• is described. This involves transposition of the flap to the defect without tunneling, and the pedicle is exteriorized and skin grafted. In a second procedure, after randomization, the pedicle is removed. The authors have no experience using this approach, but there appears to be no distinct advantage of this approach over other options. Indeed, it has the disadvantage of a two-step procedure.

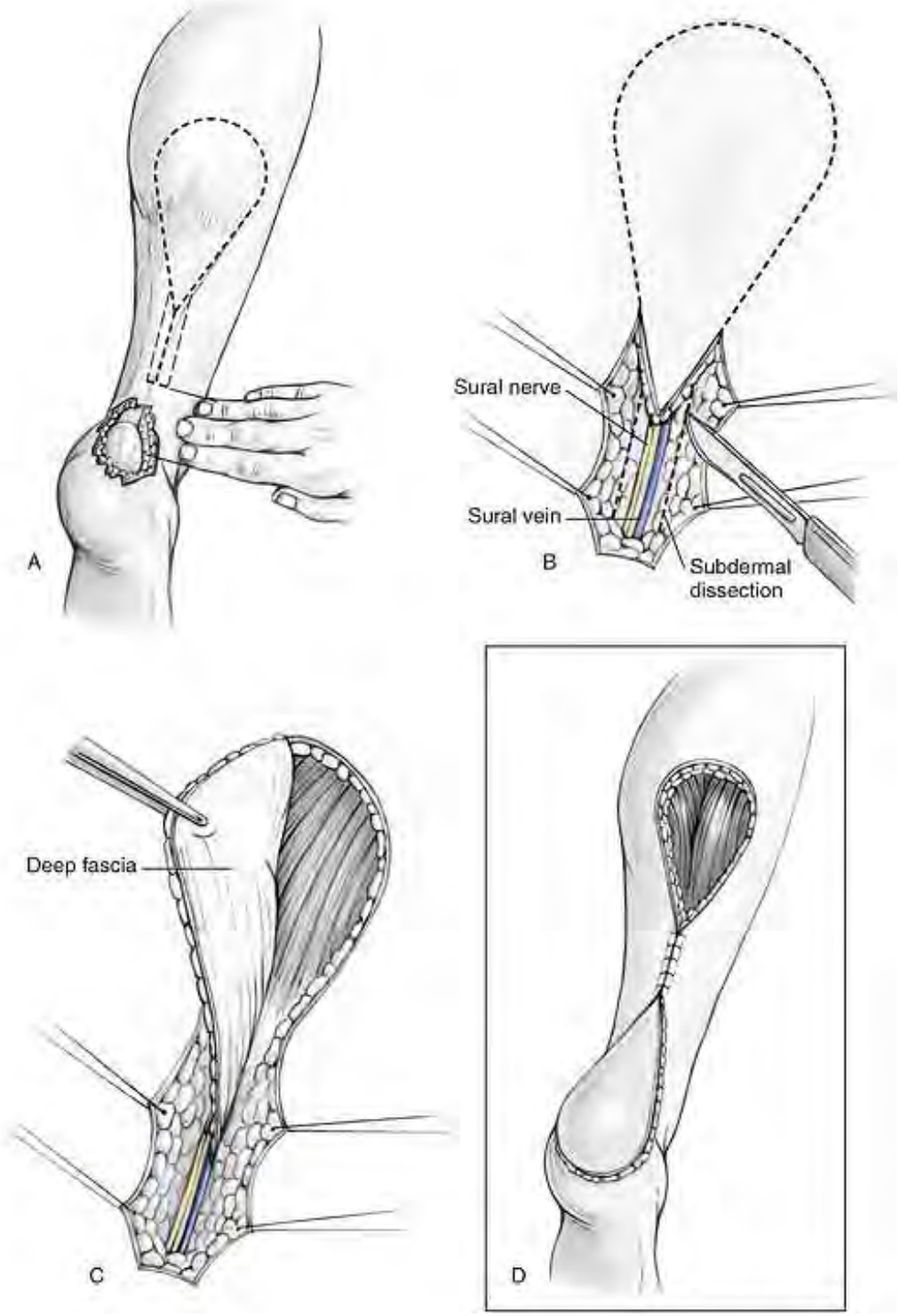

FIGURE 29-5 The skin island has to be planned according to the size of the defect with its center along the line of the short saphenous vein. The pivotal point is three fingers breadth to the tip of the lateral malleolus (A). The skin incision

starts distally along the line of the pedicle to localize the vein. The subcutaneous pedicle should be at least 3 cm wide (B; 1â€”sural vein and nerve; 2â€”subdermal dissection). Dissection proceeds along the plane below the deep fascia; thus, inclusion of the pedicle in the flap is easily possible and the nutrient anastomosis can be preserved (C: 3â€”deep fascia included). After complete flap elevation, the flap is transposed to the recipient site (D).

If signs of insufficient perfusion or venous drainage occur after flap transposition and if decompression of the pedicle is unsuccessful in resolving this problem, a suralis flap delay procedure is possible. In this procedure, the flap is replaced into its donor site bed, a powder-free glove is placed between the elevated fascia and the gastrocnemius muscle, and transposition is performed 2 to 4 days later.

In some cases, the reverse flow sural artery flap can be modified. Hence, muscle tissue can be included in the flap when the perforating vessels from the gastrocnemius muscle to the vascular axis

of the sural nerve (see previous discussion) are preserved. Flap harvest is similar to the procedure described previously. The only difference is that a fragment of the gastrocnemius muscle is taken from the lower part of the muscle without separating the fascia from the muscle fragment. Another option is to elevate the flap as a fascial flap without a skin island; this is the preferred approach of some authors in patients with heel defects. In these cases, the flap is covered by split thickness skin grafting. Furthermore, some authors describe the possibility of reconstructing defects distant from the sural artery flap donor site by performing a free microvascular sural artery flap.

Donor site closure is possible primarily when a skin island of less than 4 cm in width is harvested. In these cases, the donor site is acceptable in terms of function as well as aesthetics. Larger donor site defects have to be closed by split thickness skin grafting. In these cases, the acceptability of the donor site scar is questionable, particularly among obese patients and females.

Postoperative Management

The most important postoperative issue is prevention of compression of the vascular pedicle. This can be achieved either by an adequately elevated position of the leg and/or by the use of conventional splints with a gap over the flap; however, the authors prefer using an external fixation device. External fixation incorporates the treatment of concomitant fractures and prevention of an equinus, as well as elevated positioning of the leg.

The administration of anticoagulants in the postoperative period after pedicled flap reconstruction remains controversial. To the authorsâ€™ knowledge, there is no scientific evidence of beneficial effects of an anticoagulation therapy in terms of flap survival or thrombosis rates. Moreover, one could postulate that anticoagulants increase perioperative morbidity due to bleeding and hematoma. Nevertheless, in our center, as in many other centers, we usually use an postoperative regimen of hydroxyethyl starch (HAES 10%; 500 mL in 24 hours for 5 days) and heparin or low-molecular-weight heparin (Clexane). Moreover, all patients are treated with antibiotics (cephalosporin) perioperatively. Antibiotic treatment is prolonged if the patient shows an increased risk of postoperative wound infection or persistent infection.

Surveillance of flap perfusion in terms of arterial as well as venous flow must be performed regularly during the first postoperative days. In the authorsâ€™ department, the capillary refill is tested every hour for the first 48 hours. Thereafter, the intervals between testing are increased.

A revision procedure should be performed if any signs of poor arterial perfusion or venous congestion are identified. New positioning of the vascular pedicle or decompression of the nutrient vessels may facilitate flap salvage. Moreover, the flap may be laid back in the donor site bed if these options fail.

The patient has complete bed rest for a minimum of 5 days. After this period compression bandages are applied to the treated leg and mobilization begins. The regimen begins with hanging down the leg for 5 minutes. The tolerance of the flap to this treatment (as well as accompanying injuries) determines the next stages of mobilization. If signs of venous congestion or poor perfusion are recognized, the mobilization regimen is prolonged. In most cases, the patient is completely mobilized after 10 days. The further use of a compression garment is obligatory for a period of at least 1 year after discharge to improve contouring of the flap and avoid hypertrophic

scarring. Secondary corrections at the recipient as well as the donor site should not be performed until at least 6 months post surgery.

Complications

Partial or complete necrosis of the sural artery flap remains the most important issue concerning the success or failure of this reconstructive method. Necrosis rates between 5% and 36% have been described in the literature. A detailed analysis of the study cohorts helps to explain these different findings. When patients with comorbidities are excluded, the patient's age alone has been found to be a significant risk factor for flap necrosis. Furthermore, systemic diseases such as diabetes mellitus, peripheral arterial disease, and venous insufficiency are considered to be negative predictors for flap survival. Therefore, necrosis rates should always be interpreted in relation to the risk profile of the individual patient.

However, comorbidity with any of the above-mentioned diseases does not present an absolute contraindication to this approach, especially in view of the limited alternatives. First, it should be

kept in mind that the flap does still have a chance of success; second, the sural artery flap often represents the last chance of limb salvage in multi-morbid patients who are not suitable candidates for a free flap transfer. Some authors suggest that in cases of partial and even complete necrosis, the sural artery flap can serve as a valuable biologic dressing, so that the reconstructive surgeon sometimes experiences a well-vascularized granulating wound after debridement of the necrotic flap that allows successful split thickness skin transplantation. It seems that vascularization improves under the flap, which facilitates a skin graft on a wound bed that initially would not have been possible.

Several technical guidelines should be followed to prevent perioperative morbidity, particularly among patients with an increased risk of flap necrosis. The first of these guidelines is the use of the Doppler ultrasound in preoperative planning. Then, as described previously, the sural nerve should be included in the flap and the surgeon should pay attention to the subcutaneous layer. It can cause the flap to be bulky at the recipient site and, therefore, increases the risk of pedicle compression. The solution is to plan the flap slightly larger than the defect to facilitate skin-to-skin

closure or to perform extensive mobilization of the surrounding skin, although this in turn can cause wound healing disturbances. We recommend the use of an external fixation device to prevent pedicle compression and to facilitate postoperative care, especially in less compliant patients. Tunneling of the vascular pedicle should only be performed if elastic skin surrounds the defect; otherwise, the pedicle should be grafted or the skin paddle should be planned in a teardrop shape to facilitate pedicle coverage (see previously). Furthermore, it should be taken into consideration that the more proximal the donor site is planned, the greater the risk of necrosis.

Further complications described in the literature are not specific to sural artery flaps but relate to the underlying defect or the surgical procedure, respectively. Wound infection or persistent infection are common findings as well as hematoma, delayed healing, or persistent osteitis (which itself is usually due to incomplete debridement). As described previously, inelastic skin and induration are familiar findings especially in patients with comorbidities. Therefore, the skin surrounding the flap margins or the pedicle may require skin grafting for wound closure. Edema formation is common, especially in patients with venous insufficiency. In these cases, application of compression bandages may help improve local trophicity.

Donor site morbidity is generally low. The most common findings are neuroma of the sural nerve and scarring. In these patients, secondary corrections should be performed. Neuromas have to be resected, and the nerve stump has to be buried in the surrounding musculature.

Recommended Reading

Almeida MF, da Costa PR, Okawa RY. Reverse-flow island sural flap. *Plast Reconstr Surg.* 2002;109:583-591.

Arnez ZM, Kersnic M, Smith RW, et al. Free lateral arm osteocutaneous neurosensory flap for thumb reconstruction. *J Hand Surg [Br].* 1991;16:395â€“399.

Baumeister SP, Spierer R, Erdmann D, et al. A realistic complication analysis of 70 sural artery flaps in a multimorbid patient group. *Plast Reconstr Surg.* 2003;112:129â€“140.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 30

Fasciotomies of the Lower Extremity

Christopher J. Salgado

Amir A. Jamali

Jason Nascone

Indications/Contraindications

Compartment syndrome is the most common indication for fasciotomy. Compartment syndrome is a clinical condition with elevated tissue pressure within a closed anatomic compartment. Muscles are contained within an osseofascial compartment that has a limited capacity to expand. The causes of compartment syndrome can be divided into two major categories: decreased compartment size and increased compartment volume. Decreases in compartment size can be due to extrinsic factors such as tight dressings or casts or due to intrinsic causes such as bleeding into a compartment after injury or a postoperative coagulopathy. Increased compartment volume can occur at the macroscopic or microscopic level. Bleeding or iatrogenic infiltration of intravenous fluid into a closed compartment are both common causes of compartment syndrome. At the microscopic level, compartment volume can be increased in proportion to either increased capillary permeability and/or capillary pressure. Conditions associated with tissue damage such as burns, ischemia/reperfusion, and trauma can all lead to increased capillary permeability. Increased capillary pressure is the underlying cause of compartment syndrome due to venous obstruction or exercise.

The underlying pathologic condition leading to compartment syndrome is an elevated

tissue pressure which leads to decreased arteriolar perfusion. At this point shunting occurs, bypassing the capillary circulation which then worsens the tissue ischemia, and in turn, increases the capillary permeability and interstitial tissue pressure. This vicious cycle can quickly lead to permanent tissue damage if not treated expediently. The tissues most at risk in compartment syndrome are the nerves and muscles. If untreated, compartment syndrome can lead to Volkmann's contracture, a permanent paralysis of muscles in the compartment with scarring in a shortened position leading to the term â€œcontracture.â€• Assessment of an injured lower extremity must include a thorough evaluation of factors, which can contribute either directly or indirectly to compartment syndrome. A list of such factors is given in Table 30-1. The treatment of compartment syndrome is the correction of the underlying pathologic state and the performance of a fasciotomy. Fasciotomy is the incision of fascial compartments in order to expand the size of the compartment and to restore tissue perfusion to the contents of the compartment.

Preoperative Planning

Preoperative planning of a fasciotomy is based on an accurate and timely diagnosis of compartment syndrome. The most important data guiding this decision is the clinical examination. In the setting of an awake, unsedated patient, the diagnosis can commonly be made based on clinical grounds. The classical clinical signs of compartment syndrome are the six P's: pain, pressure, paresthesia, paralysis, pallor, and pulselessness. Note that significant muscle damage occurs prior to the onset of pallor, pulselessness, and paralysis, and that these are late findings of a missed compartment syndrome. The most sensitive clinical sign of compartment syndrome is pain. This is often described as being out of proportion to the injury. In the author's experience, the variability of pain threshold among

patients can make the assessment of â€œexpected level of painâ€• somewhat arbitrary. The presence of a greater than expected level of pain does not indicate definitive compartment syndrome and the absence of pain does not rule out the diagnosis, particularly in the setting of a possible neurologic injury. The use of sedatives and analgesics, a history of central nervous system trauma, and the possibility of peripheral nerve injury highlight the need for a high index of suspicion and the need for early intracompartmental pressure measurements. Passive stretching

of muscles within the compartment leads to elevated pressures and increased pain, another valuable tool in establishing a clinical diagnosis. Other signs of compartment syndrome include paresthesias and paralysis. Paresthesias are due to ischemic damage to the peripheral nerves running in the compartment. Decreased 2-point discrimination is the most consistent early finding. Correlation has also been reported between diminished vibration sense (256 cycles per sec) and increasing compartment pressure. On deep palpation, a firm wooden feeling is a specific sign when present. Bullae may also be observed. In later stages, the paresthesias can progress to complete anesthesia in the distribution of the peripheral nerve.

Table 30-1. Compartment Syndrome Risk Factors

History	Injury	Treatment
Crush injury	Open and closed fractures	Fluid administration
Entrapment	Arterial injury	Tourniquets
Ischemia	Venous injury	Positioning
Shock/hypotension	Gunshot wounds to extremity	MAST
Overdose/unconsciousness	Coagulopathy	Arthroscopy pumps
Tight ski boots	Shock	Jet lavage
Coumadin	Deep vein thrombosis	Revascularization

Weightlifting/overuse	Burns	Vein ligation
Knee arthroscopy	Muscle tear	Fracture tables
	Snake envenomation	Tight wound closures
	Ruptured Baker's cyst	Constrictive dressings
		Regional anesthesia

MAST, Military anti-shock trousers

Paralysis can be due to muscle ischemia, nerve ischemia, direct injury to these structures, or secondary to pain inhibition. Pulselessness is uncommon in an isolated compartment syndrome and heralds a probable vascular injury. Laboratory testing revealing a creatine kinase (CK) of 1,000 to 5,000 U/mL or higher or the presence of myoglobinuria may alert the physician to the occurrence of compartment syndrome. When the clinical picture is borderline, compartment pressure measurements must be performed as soon as possible.

A number of techniques have been employed to determine compartment pressures, including variations in size and needle design. At our institution all compartment measurements are performed with a side-port needle attached to a commercially available pressure monitor (Stryker, Kalamazoo, MI) or a standard arterial line pressure transduction line (Fig. 30-1). The Stryker pressure tonometer is widely used, and pressure measurements from the Stryker device are within 5 mm Hg of the slit catheter for 95% of all readings. Measurements with a standard 18-gauge needle are not accurate and are not recommended. Pressure measurements should be performed within all compartments and at multiple sites.

The compartment pressure data can be viewed in isolation or in relation to the patient's diastolic blood pressure. Although there is no absolute minimum compartment pressure value, most current literature indicates that the Î”P value from measured compartment pressure to diastolic blood pressure is a more valuable guide in performance of a fasciotomy. Studies suggest that the ischemic threshold of muscle is a perfusion pressure of at least 20 mmHg between the compartment pressure and the diastolic pressure. The Î”P is a direct measure of the pressure gradient between diastolic blood pressure and the tissue pressure within the compartment, indicating the presence of shunting. At our institution, a Î”P of 30 mmHg combined with increased palpable pressure is a strong indication and a Î”P of 20 mmHg an absolute indication for fasciotomy.

Therapy is begun for the treatment of compartment syndrome while preparations are made for actual surgical decompression. The affected limb(s) are placed at the level of the heart. Elevation is

contraindicated because it decreases arterial inflow and narrows the arterial-venous pressure gradient and thus worsens the ischemia. If a cast is on the affected extremity, releasing one side of the plaster cast can reduce compartment pressure by 30%; bi-valving can produce an additional 35% reduction; and cutting the cast padding may further decrease compartmental pressure by 10% to 20%. In cases of snake envenomation, administration of antivenom may reverse a developing compartment syndrome. Hypoperfusion may be corrected with crystalloid and blood products, and mannitol may reduce compartment pressures and lessen reperfusion injury.

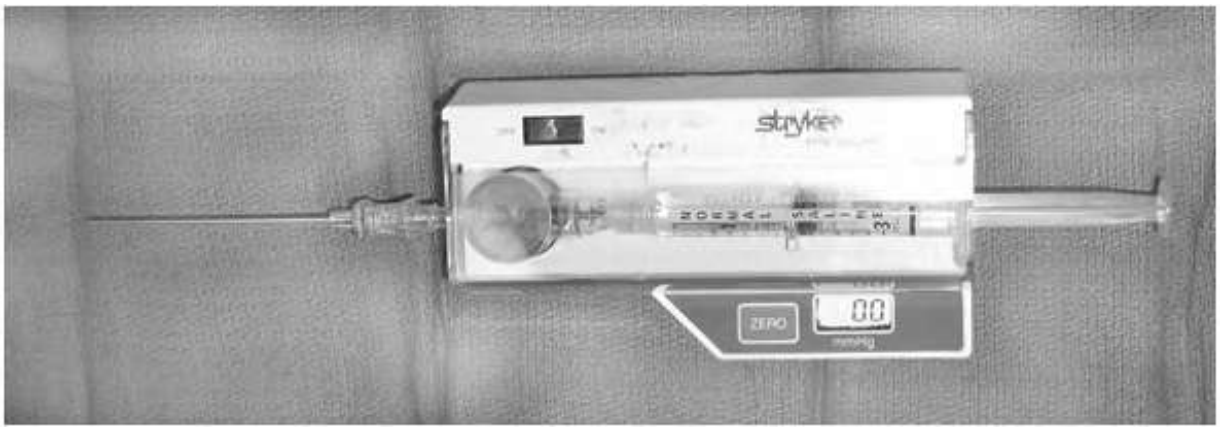

FIGURE 30-1 Stryker pressure monitor (Stryker, Kalamazoo, MI) with side-port needle.

Hyperbaric oxygen (HBO) is a valuable adjunct in the treatment of compartment syndrome. It promotes hyperoxic vasoconstriction, which reduces swelling and edema and improves local blood flow and oxygenation. It also increases tissue oxygen tensions and improves the survival of marginally viable tissue. The best results are obtained when therapy is started early after fasciotomy. Twice-daily treatments at 2.0 atmospheres absolute (ata) to 2.5 ata for 90 to 120 minutes for 5 to 7 days, with frequent examinations of the affected area, may be beneficial. This may be more practical in centers familiar with the use of hyperbaric oxygen therapy.

Vacuum-assisted closure devices (VAC) have a number of advantages in the treatment of post-fasciotomy wounds. They reduce interstitial edema and provide a one-way flow of exudate from the wound. Additionally, this therapy increases granulation tissue formation and may lead to an earlier ability to close or skin graft the wound.

Fasciotomy of the Thigh

Thigh and gluteal compartment syndromes are uncommon and may often go unrecognized. The pathophysiology and the principles of diagnosis and treatment, however, are the same as those for other compartment syndromes. Gluteal compartment syndromes are often associated with substance abuse and a prolonged period of unconsciousness or recumbency and can occur in the absence of any obvious

trauma. As a result of the large muscle mass involved, systemic manifestations of a crush syndrome are usually present. Altered mental status and metabolic abnormalities may distract from the primary problem, resulting in delayed diagnosis and treatment. The proximity of the sciatic nerve can result in compression-induced neuropathy.

Clinical Findings

Thigh compartment syndrome is rare because of the large volume required to cause a pathologic increase in the interstitial pressure. It may occur in the setting of high-energy thigh trauma such as femur fractures with an associated crush component. These patients often have pain and swelling after fixation, which may confound the diagnosis. In addition, these trauma patients are often obtunded and require substantial fluid resuscitation, increasing the risk of compartment issues. The fascial compartments in the thigh blend anatomically with muscles of the hip, potentially allowing extravasation of blood outside these compartments. Anticoagulation can be a major risk factor leading to bleeding into the thigh compartments and the development of a compartment syndrome.

Patient Positioning

The approach to thigh compartments may be medial or lateral depending on the area of injury or suspected hematoma. The thigh should be prepared from the iliac crest to the knee joint with the patient in either the lateral decubitis position or supine.

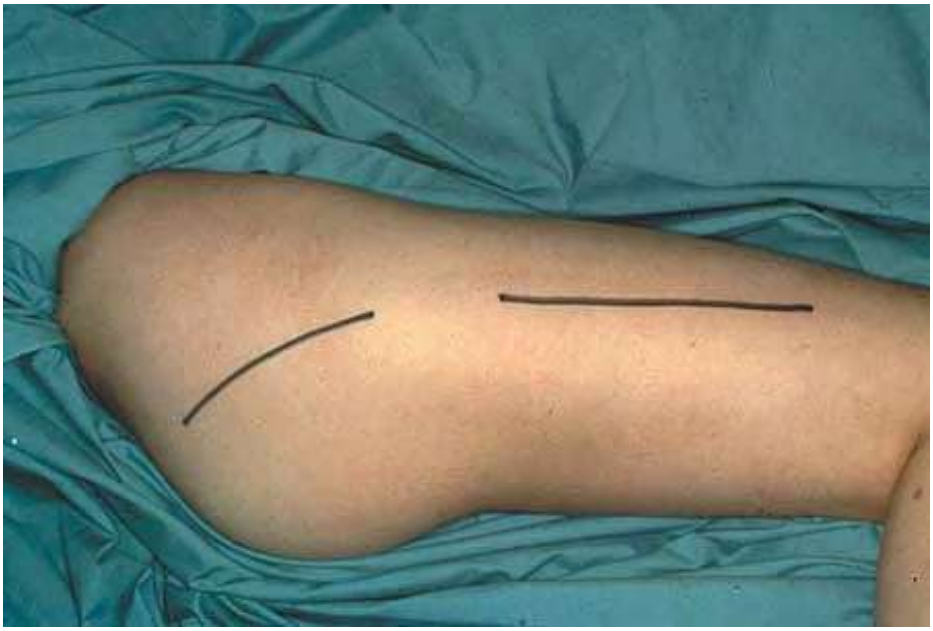

FIGURE 30-2 Skin markings for gluteal and thigh (anterior and posterior) compartment release.

Technique

For lateral and posterior compartment syndromes, the skin and subcutaneous tissues are incised beginning just distal to the intertrochanteric line and extending to the lateral epicondyle of the femur to expose the iliotibial band or fascia lata (Fig. 30-2). The iliotibial band is incised for the length of the incision. The vastus lateralis muscle is reflected superiorly and medially to expose the lateral intermuscular septum, which is incised for the length of the incision, thus freeing the posterior compartment. Caution is required to control the perforating branches of the descending branch of the lateral femoral circumflex artery traversing the lateral intermuscular septum (vessels which supply the anterolateral thigh skin), since these may retract and bleed during this portion of the decompression (Fig. 30-3). After the anterior and posterior compartments have been released, measure the pressure of the medial compartment. If elevated, then the compartment can be approached through a separate medial incision. The incision is carried along the course of the saphenous vein. Reflect the

sartorius muscle superiorly, and incise the medial intermuscular septum. Intramuscular hematomas may require release through gentle muscle splitting. The wounds are packed open and a large bulky dressing applied or, alternatively, a vacuum-assisted closure device is applied.

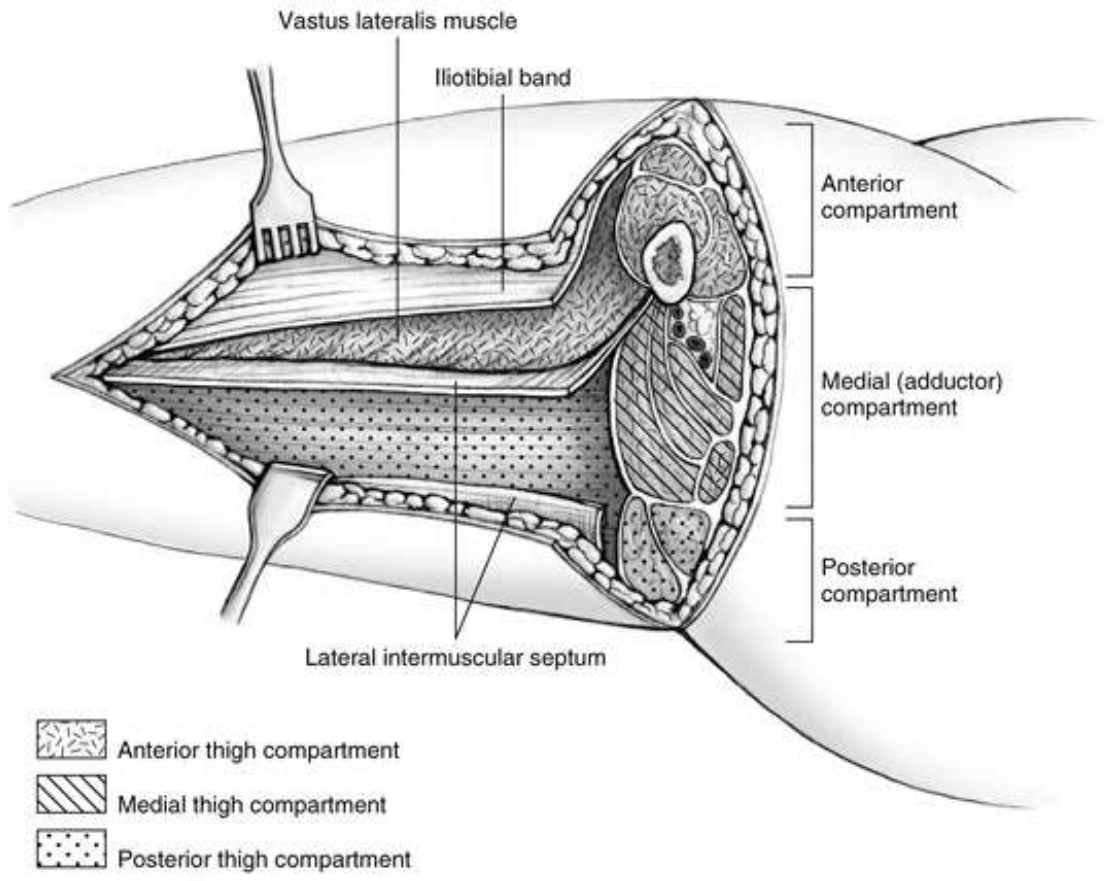

FIGURE 30-3 Schematic lateral view showing decompression of anterior compartment performed by incising the fascia latae longitudinally. The vastus lateralis is retracted medially to expose the lateral intermuscular septum, which is incised to decompress the posterior compartment.

Two to three days later, the patient is returned to the operating room for debridement of any nonviable tissue. If there is no evidence of necrotic tissue, the

skin is either loosely closed, packed once again for closure at a later date, or again covered with a vacuum-assisted closure device. Often a medial thigh fasciotomy is not needed once a lateral release is performed.

Results

Because this diagnosis is not always obvious, the surgeon must maintain a high index of suspicion. Early treatment by operative compartment release follows anatomic tracts and produces good results.

Fasciotomy of the Leg

The framework of the lower leg is composed of two long bones, the fibula and tibia, which are arranged in parallel and connected along their length by a fibrous membrane termed the interosseous membrane. These three structures together divide the leg into two anatomic sections, the anterior and posterior compartments. The anterior compartment is further divided into anterior and lateral by a thick anterior intermuscular septum. The lateral leg compartment is separated from the posterior compartment by the posterior intermuscular septum.

The anterior compartment has four muscles, the (a) extensor digitorum longus, (b) extensor hallucis longus, (c) peroneus tertius, and (d) the tibialis anterior. These muscles are supplied by the anterior tibial vessels and are innervated by the deep peroneal nerve, all traveling deep to the muscles along the interosseous membrane. The posterior compartment is divided into superficial and deep compartments by a thin fascia termed the transverse intermuscular septum. Three muscles are located in the superficial compartment: the gastrocnemius, soleus, and plantaris. The gastrocnemius and soleus join together at midcalf to form the Achilles tendon, which inserts into the calcaneal bone. The plantaris is a thin, small muscle, the tendon of which medially follows the bigger Achilles tendon to insert into the calcaneal bone. All of these muscles flex the foot in a plantar direction (the gastrocnemius also flexes the knee) with slight inversion. The muscles are vascularized by branches from the popliteal and posterior tibial artery and innervated by branches of the tibialis nerve from the popliteal fossa. The deep posterior leg compartment contains four muscles: the popliteus, flexor digitorum longus, flexor hallucis longus, and tibialis posterior. The lateral compartment contains two muscles, the peroneus longus and brevis. Their

action consists of extension and eversion of the foot. They are vascularized by vessels from the peroneal artery and innervated by the superficial peroneal nerve (Fig. 30-4).

The posterior tibial artery, after branching from the popliteal artery, descends posterior to the tibia, within the deep posterior compartment. Distally in the lower third of the leg it is more superficial, covered only by the skin and superficial fascia and parallel to the medial border of the Achilles tendon. The peroneal artery originates from the posterior tibial artery and runs laterally and downward along the posterior fibula providing four to six segmental, circular arterial branches around the fibula, nourishing the bone, periosteum, and surrounding muscles. The anterior tibial artery commences at the bifurcation of the popliteal artery passing forward between the tibialis posterior muscle and through the interosseous membrane to the deep aspect of the anterior leg compartment. After perforation of the interosseous membrane, the artery descends adjacent and along the membrane, gradually approaching the extensor retinaculum of the ankle. At the bend of the ankle joint, it becomes superficial and known as the dorsalis pedis artery.

Patient Positioning

In cases of isolated compartment syndrome of the leg, the patient is positioned in the supine position on a standard operating table. A general anesthetic is employed in most cases. A tourniquet is applied to the thigh and isolated with broad tape or a plastic isolation dressing. The leg is prepared and draped in the standard fashion. A stockinette or isolation sheet is applied to the foot to maintain the sterile field. The tourniquet is not inflated unless active arterial bleeding is encountered.

Technique

The technique for fasciotomy of the leg can be performed using either a one incision or two incision technique (Table 30-2). The two incision technique is the gold standard. Regardless of the technique,

the skin incisions must be of adequate length to decompress all affected compartments. The performance of percutaneous fasciotomies is not recommended, particularly with compartments deformed by soft tissue edema, since this technique can be fraught with complications such as iatrogenic nerve injury. In addition, the skin

itself has been shown to exert a constrictive effect on the muscle compartments in the leg.

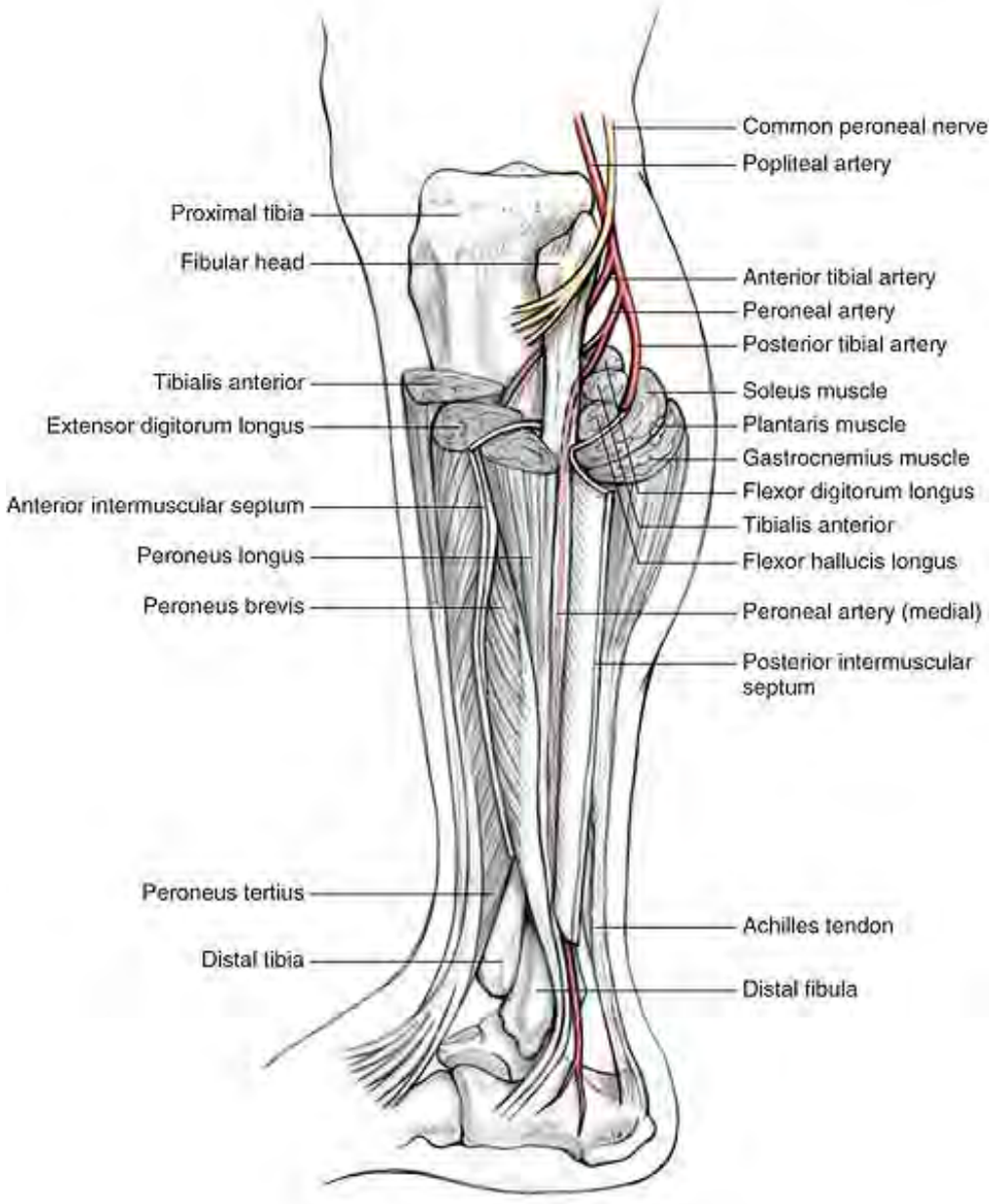

FIGURE 30-4 Lateral leg.

One Incision Fasciotomy

The planned incision is marked in line with the fibula extending to 5 cm short of either end of the fibula along the anterolateral leg. The initial step is the identification of the lateral intermuscular septum separating the lateral and anterior compartments. Make a transverse incision to expose this septum and to identify the superficial peroneal nerve just deep to the septum. Separate fasciotomies of the compartments are performed with Metzenbaum scissors. The anterior compartment is released proximally by aiming for the patella and distally by aiming for the center of the ankle in line with the tibialis anterior. Then, perform a longitudinal fasciotomy of the lateral compartment in line with the fibular shaft (Fig. 30-5). Extreme care is taken in the distal aspect of the lateral compartment at the junction of the middle and distal thirds, where the superficial peroneal nerve emerges from the lateral compartment. Direct the scissors toward the posterior lateral malleolus to stay posterior to the superficial peroneal nerve. In the case of acute compartment syndrome, we avoid subcutaneous fasciotomies. The use of the â€œsliding scissorâ€• technique decreases the ability to perform a controlled fasciotomy and should be avoided. Once the anterior and lateral compartments have been decompressed, the fibula is identified after posterior undermining of the

skin for release of the deep posterior compartments. The lateral compartment musculature is elevated off the fibula, demonstrating the posterior intermuscular septum. A longitudinal incision of this septum exposes the superficial posterior compartment. Posterior retraction of the soleus and gastrocnemius muscles then exposes the deep posterior compartment for its decompression. We have found that surgical exposure of the posterior aspect of the fibula is critical in the avoidance of bleeding from the perforating branches of the peroneal artery. Release of all four compartments through one incision, however, is not commonly performed at our institutions, and we recommend a two incision fasciotomy.

Table 30-2. Pearls for Fasciotomy of the Leg

- Check intracompartmental pressures at multiple levels within each

compartment.

- Threshold for fasciotomy is intracompartmental pressure within 20 mm Hg of the diastolic blood pressure.
- Beware of compartment syndrome secondary to intraoperative positioning of uninjured extremities.
- Pulselessness is NOT a common finding in compartment syndrome unless a concurrent vascular injury exists.
- Fasciotomy should not be performed more than 12 hours after a compartment syndrome is established.
- Compartment syndrome can occur with a late onset 2 to 4 days after the underlying event.
- Compartment pressure measurements with a standard 18-gauge needle are consistently higher than those obtained with specialized needles with a side-port needle or slit catheter.
- Compartment pressures can vary among several different points in each compartment at risk.
- In any patient at risk, baseline compartment pressures are necessary in the case of a suspicion of compartment syndrome in the future.

Two Incision Feg Fasciotomy (Gold Standard)

- Mark both incisions before making the first surgical incision.
- Beware of the superficial peroneal nerve as it emerges anteriorly within the lateral compartment at the junction of the middle and distal one-third of the fibula.
- Beware of the use of vessel loops or other elastic forms of skin tension applied at the time of fasciotomy, as these may lead to recurrence of the compartment syndrome if excessively tight.
- Close medial wound before the lateral wound to avoid need for soft tissue coverage.

One Incision Leg Fasciotomy

- Obtain complete visualization of the release of the deep posterior compartment, as extensive bleeding can occur secondary to perforating vessels on the posterior aspect of the fibula.
- Extend incision to within 5 cm of either end of the fibula on the anterolateral leg.
- Beware of the superficial peroneal nerve at the junction of the middle and distal one-third of the leg.

Two Incision fasciotomy

The two incision fasciotomy is the gold standard treatment for compartment syndrome of the leg, particularly in the hands of surgeons with limited previous experience and limited assistance in the operating room. The positioning and preparation are identical to that of the one incision fasciotomy. It is crucial to mark the medial and lateral skin incisions before making the incision to ensure an adequate skin bridge is maintained. After one single incision is made, the

skin envelope will retract in the opposite direction resulting in a narrow skin bridge anteriorly, and marking these incisions will avoid this complication. After performance of the anterior and lateral fasciotomies, a separate incision is made along the posteromedial leg, 1 to 2 cm posterior to the tibia, measuring at least 15 cm in length. The greater saphenous vein and saphenous nerve are identified and retracted anteriorly. The fasciotomy is extended as far as possible proximally and distally to the level of the medial malleolus. The soleus is then released from the posteromedial tibia with a concurrent release of the deep posterior compartment (Fig. 30-6). A common pitfall is to not adequately release the soleus muscle insertion thereby adequately releasing the deep posterior compartment. Care should be exercised when releasing the deep compartment ligating or cauterizing branches traversing the soleus muscle. The posterior tibial neurovascular bundle is just deep to this transverse intermuscular septum, which separates the deep from posterior compartments and is therefore in close proximity to the release. The skin incisions can also be slightly staggered, with the lateral more proximal (the anterior and lateral compartments are largely tendonous at their distal extent) and the medial incision made more distally. Finally, the fascia and skin are then re-evaluated for adequacy of release.

FIGURE 30-5 Intraoperative view after anterior and lateral compartment decompression. Note the intermuscular septum separating the anterior and lateral compartments.

FIGURE 30-6 Intraoperative view after release of superficial posterior compartment and before release of deep posterior compartment.

After either type of fasciotomy, compartment pressures are again checked using a sterile side-port needle attached to the non-sterile pressure monitor held by an unscrubbed assistant or, alternatively, using an arterial pressure monitor setup. The wounds are then copiously irrigated with crystalloid. Devitalized tissue if present is debrided. Vessel loops can be applied to the skin edges to prevent marked skin retraction but must be used judiciously to avoid a recurrence of the compartment syndrome. At our institution, we routinely employ a vacuum-assisted closure device over all open fasciotomy wounds to maintain a one-way flow of extravasated fluid, to encourage the formation of granulation tissue, and potentially to minimize the area needed for later skin grafting. The patient is returned to the operating room at 2 to 3 days for a repeat irrigation and debridement and partial closure. At or around postoperative day 7, the wound is again treated with irrigation and debridement. At that time, definitive closure is performed with additional skin coverage obtained with a split-thickness skin graft. We usually strive to close the medial wound first to provide bony coverage and avoid the need for soft tissue coverage, whereas the lateral wound may be easily skin grafted if appropriate. It is imperative to maintain the foot in a

neutral position to avoid equinus contracture either with external fixation or external splinting.

Results

The primary goal of fasciotomy is to prevent permanent nerve and muscle damage leading to Volkmann's contracture. The results of fasciotomy can be analyzed based on two endpoints: muscle and nerve function, and wound- or incision-related complications. A number of studies have demonstrated that the majority of patients have normal leg function if the fasciotomy is performed within 12 hours of the onset of compartment syndrome. In our experience, prompt recognition and early treatment of the compartment syndrome leads to minimal long-term complications. Despite early and aggressive fasciotomy, however, nearly 20% of patients may have persistent motor deficits at 1 year follow-up. Wound complications associated with fasciotomy include numbness and persistent ulceration at the fasciotomy site. The risk of sensory changes has been reported to be as high as 70% in some reports. The use of a VAC dressing may contribute to a lower risk of hematoma and

edema and an expeditious granulation of fasciotomy wounds that are not able to be closed. Little or no return of function can be expected when diagnosis and treatment are delayed. Tendon transfers and foot stabilization may be indicated as late treatment; but in most patients, enough scarring and contracture eventually develop in the anterior musculature to prevent foot drop. A foot drop brace (ankle foot orthosis) is indicated for the first few months until fibrosis occurs. Some patients experience persistent gastrocnemius and soleus muscle weakness thought to be due to the loss of the supporting compartment fascia.

Fasciotomy of the Foot

Anatomically, the foot consists of well-demarcated osseofascial spaces that subdivide the foot into discrete compartments. These compartments are filled with muscles, nerves, and tendons and are lined by a tight membrane (the fascia). There are four clinically relevant compartments known as the medial, central, lateral, and interosseous (Fig. 30-7). Other anatomic compartments of the foot may be identified with dyes or injection studies but are not clinically relevant. Muscles within the medial compartment are the abductor hallucis and flexor hallucis brevis, and within the

central (calcaneal) compartment lay the flexor digitorum brevis, quadratus plantae, and adductor hallucis muscle. There are four dorsal and plantar interosseous muscles between the first and fifth metatarsals, and these comprise the interosseus (intrinsic) compartment. The lateral compartment houses the abductor digiti minimi and flexor digiti minimi brevis muscle.

Compartment syndrome in the foot is commonly due to severe local trauma after fairly significant industrial, agricultural, and motor vehicle accidents in which crushing of the foot occurs. The trauma leads to bleeding from injured bones or muscles and avulsed vasculature. When the bleeding and inflammation of the muscle becomes significant, it will exceed the capacity of blood flow in and out of the small compartments.

Clinical Findings

Tense tissue bulging may be the most reliable symptom in compartment syndrome of the foot, and in the presence of this massive swelling pulses are usually not palpable. Increased pain on passive dorsiflexion of metatarsophalangeal joints is another key finding that indicates myoneural ischemia in the foot intrinsic muscles. It is imperative to maintain a high clinical suspicion based on the severity of the traumatic incident. Compartment syndromes of the foot are often associated with compartment syndromes of the deep posterior compartment.

Patient Positioning

The patient position is supine to allow easy access to the dorsum and medial aspect of the foot. A tourniquet may be used, but is not insufflated unless there is active arterial bleeding. The knee should be included in the preparation to enable better mobility of the limb during the procedure.

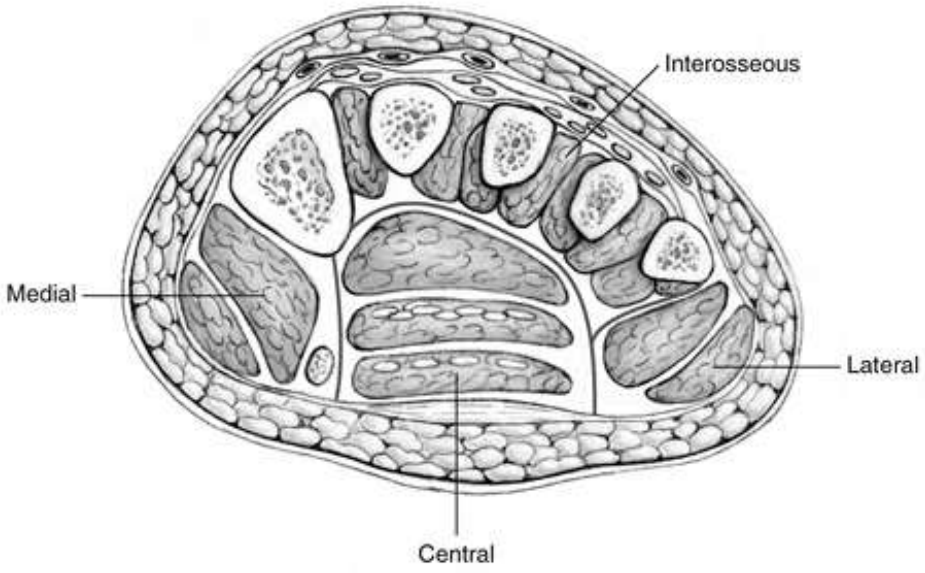

FIGURE 30-7 Schematic coronal section of right foot through base of metatarsals. Medial, central, lateral, and interosseus compartments are shown.

Technique

The same principles that apply to fasciotomy of the thigh and leg apply to the foot. Appropriate treatment for a suspected compartment of the foot is immediate fasciotomy. Debridement of marginal tissue at the time of initial fasciotomy is not advised because it is very difficult to determine muscle viability and contractility in the foot. Once the muscle is decompressed, much of it may recover after compartmental release.

Medial Approach

Effective decompression of all four compartments can be accomplished through a medial longitudinal (Henry) approach or through an additional two parallel dorsal incisions along the length of the second and fourth metatarsal bones (more common). The medial approach can be used to decompress the medial and central compartments as well as the remaining foot compartments (lateral and interosseous). The incision is

made 3 cm from the sole of the foot and extends from a point below the medial malleolus to the proximal aspect of the first metatarsal (Fig. 30-8). The posterior tibial neurovascular bundle is identified and preserved. This may be very difficult in a massively swollen foot. The fascia overlying the abductor hallucis and flexor hallucis brevis is released. Dissection is continued adjacent and deep to the first metatarsal toward the medial intermuscular septum, separating the medial compartment from the central compartment, which is opened longitudinally. The lateral plantar neurovascular bundle is found between the flexor digitorum brevis and quadratus plantae muscles and preserved. Downward retraction of the flexor digitorum brevis along with lateral dissection will allow access to the lateral intermuscular septum separating the central compartment from the lateral compartment. This septum is divided to release the lateral compartment. Blunt dissection dorsally via the central compartment will release the interosseous compartment. Scissors or sharp instruments are not used during this portion of the dissection, since it is essentially blind to avoid injury to the neurovascular bundle.

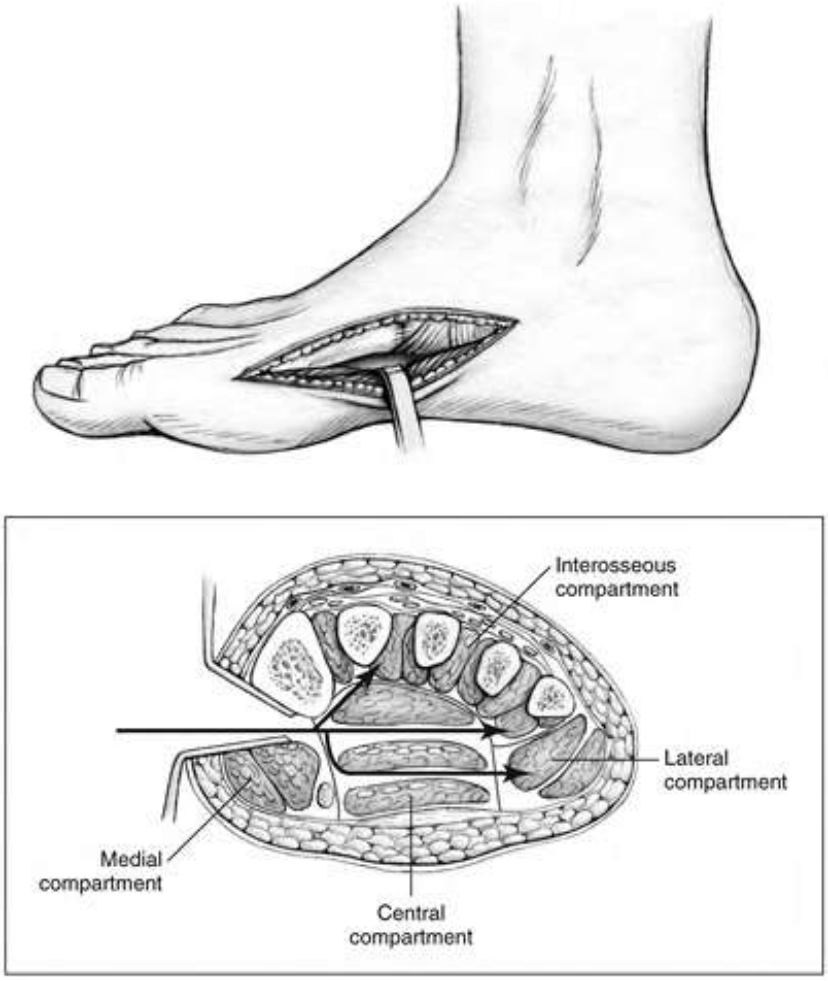

FIGURE 30-8 Medial approach for four-compartment release of foot compartment syndrome.

Dorsal Approach

This approach is more commonly used when there are concomitant metatarsal or Lisfranc fractures. Two parallel dorsal incisions are centered just medial to the second metatarsal and lateral to the fourth metatarsal shafts, maximizing the intervening skin bridge. The dorsal veins and the subcutaneous tissues are elevated laterally and medially to expose the respective interosseous musculature. Injury to the sensory nerves and extensor tendons is avoided. Caution should be used when making the

incision between first and second metatarsal to avoid iatrogenic injury to the dorsalis pedis artery. The superficial fascia is incised longitudinally, and the interosseous muscles are elevated off the metatarsals. The first dorsal and plantar interossei are stripped from the medial aspect of the second metatarsal shaft, which is then retracted medially, and the fascia of the central and medial compartment is released longitudinally deep within the inner space. The interosseous musculature is decompressed by releasing the fascia between the second and third metatarsals also through this medial incision. The lateral incision is used to decompress the interosseous muscles between the third and fourth, and fourth and fifth metatarsals, in addition to allowing access to release the central and lateral compartments (Fig. 30-9). More commonly, a separate medial incision as described previously is used to release the medial (adductor) compartment, since the medial approach only to all four compartments carries an increased risk of damage to the neurovascular bundles in the plantar aspect of the foot. These fasciotomy incisions are sometimes used for fracture fixation.

Results

Controversy exists in what results with a missed compartment syndrome of the foot. Predominately patients end up with clawing of the toes that needs to be corrected at a later date. The argument against the release is that it complicates the soft tissue envelope, limits possible incisions for reconstruction, and carries the risk of iatrogenic neurovascular injury.

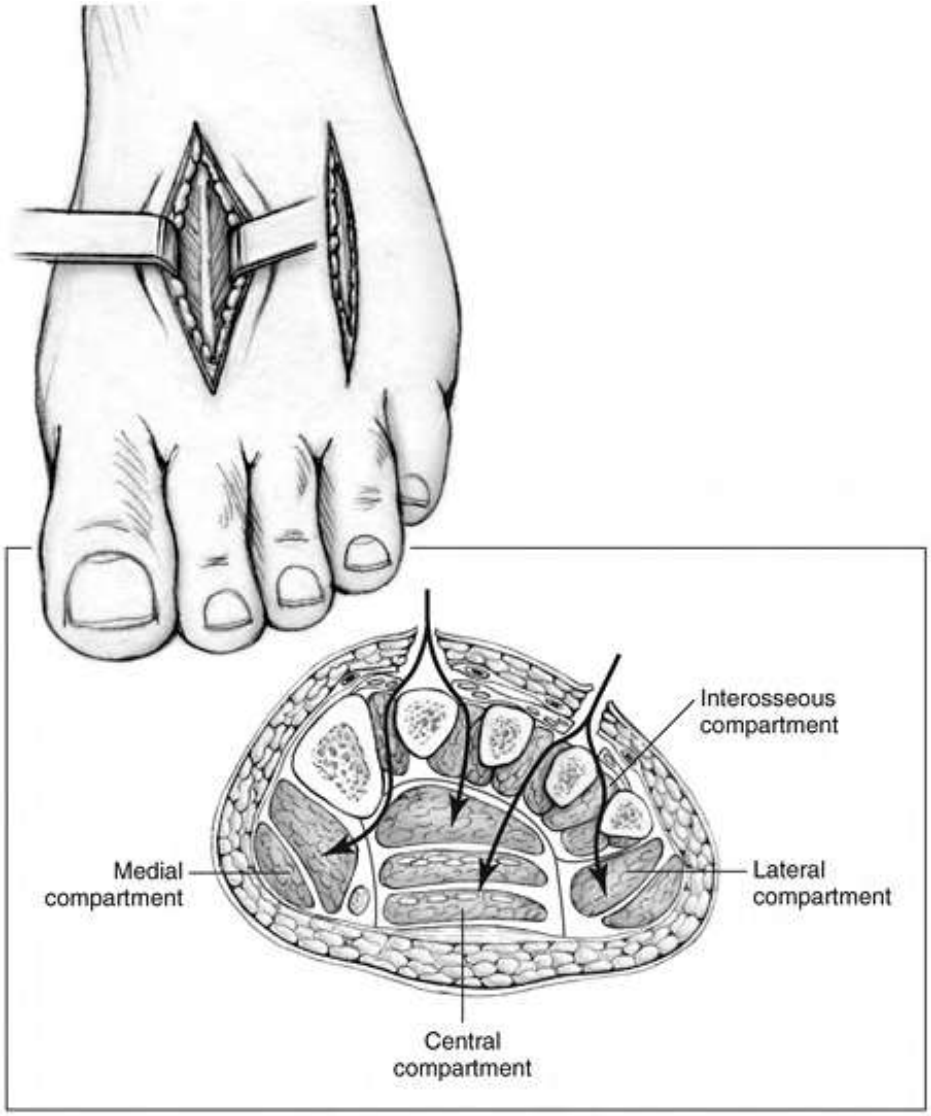

FIGURE 30-9 Dorsal approach for compartment release of foot compartment syndrome.

Treatment of a foot compartment syndrome with â€œbenign neglectâ€• is not advised, and early decompression as in the thigh and leg is recommended. A delay in diagnosis is a potentially devastating occurrence. Early reduction of dislocations and some fractures facilitates reduction in edema in the foot and decreases tissue breakdown due to pressure necrosis. Desire to obtain an early reduction must be

weighed heavily against whether the soft tissue envelope of the injured foot will tolerate additional incisions. The fasciotomy incisions may be used to facilitate reduction and fracture fixation, but caution is warranted to avoid excessive undermining of the soft tissues. Delayed fixation is feasible, but usually not before 7 to14 days post injury, due to soft tissue edema. In addition, liberal use of external fixation, particularly in the setting of a mangled foot, is advised for stabilization and edema control.

Postoperative Management

Fasciotomy sites are dressed with a wound vacuum-assisted closure sponge, and wet to dry dressing changes with normal saline solution or silver sulfadiazine depending on the degree of contamination. The patient is returned to the operating room several days later to attempt closure. When muscle necrosis ensues, the patient is brought to surgery earlier for debridement. Wound closure should not be performed until all necrotic tissue is debrided. Direct closure can be attempted when the wound approximates without tension. When the wound edges will not oppose easily, either the wound is treated conservatively with dressing changes or a vacuum-assisted closure device. The patient is brought back to the operating room at a later date for definitive closure or skin grafting. Increased intramuscular pressure can occur with closure of fasciotomy wounds secondarily and, therefore, skin grafting of the wounds is often a safer option. Hypesthesia and painful dysesthesia resulting from compartment syndrome typically resolve slowly with time. This type of neuropathic pain can by treated with diphenylhydantoin (Dilantin), gabapentin (Neurontin), or carbamazepine (Tegretol).

Complications

Although a fasciotomy incision does result in patient morbidity, the incompletely released compartment, delayed diagnosis, or unrecognized compartment syndrome has substantially higher risks. Systemic complications including acute renal failure, sepsis, and acute respiratory distress syndrome (ARDS) have been reported in some cases. Most fatalities are due to prolonged intensive care admissions with sepsis and multisystem organ failure. If fasciotomy is done within 12 hours after the onset of compartment syndrome, the prognosis is typically good. Despite early intervention, approximately 1% to 10% of all cases of compartment syndrome develop Volkmann's

contracture. Little or no return of function can be expected when the diagnosis and treatment are delayed. In no instance was benefit from fasciotomy reported after 2 or 3 days. When the procedure is done late, severe infections have been shown to develop in the necrotic tissues of many patients, frequently leading to amputations. In these situations, clinically evident by complete absence of demonstrable muscle function in any segment of the involved limb, the extremity should be splinted to maintain a functional position as muscle fibrosis and contracture develop.

Acknowledgments

The authors express their gratitude to the Department of Orthopaedics at the University of Maryland Shock Trauma and the Surgery Audio Visual Department at Cooper University Hospital (Paul Rogers) for their assistance and efforts in the preparation of this chapter.

Recommended Reading

Azar FM, Pickering RM. Traumatic disorders. In: Canale ST, ed. *Campbell's operative orthopaedics*. 10th ed. St. Louis: Mosby; 2003:1405â€“1411.

Boody AR, Wongworawat MD. Accuracy in the measurement of compartment pressures: a comparison of three commonly used devices. *J Bone Joint Surg Am.* 2005;87:2415â€“2422.

Cohen MS, Garfin SR, Hargens AR, Mubarak SJ. Acute compartment syndrome. Effect of dermotomy on fascial decompression in the leg. *J Bone Joint Surg Br.* 1991;73:287â€“290.

Finkelstein JA, Hunter GA, Hu RW. Lower limb compartment syndrome: course after delayed fasciotomy. *J Trauma.* 1996;40:342â€“344.

Fitzgerald AM, Gaston P, Wilson Y, Quaba A, McQueen MM. Long-term sequelae of fasciotomy wounds. *Br J Plast Surg.* 2000;53:690â€“693.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 31

Amputation and Stump Management

Norman S. Turner

Thomas C. Shives

Indications/Contradictions

Amputation is one of the oldest surgical procedures. Early amputations consisted of severing the extremity, and hemostasis was obtained by dipping the stump in hot oil. Techniques have dramatically improved and most of the advances have occurred during war time. Prosthetic technology now allows for amputees to run, jump, ski, swim, and be involved in competitive sports.

The indications for above and below knee amputation include life-threatening infections, malignant tumors, burns, extensive frost bite, congenital anomalies, ischemic pain, osteomyelitis, extensive trauma (including a tibial nerve ilaceration or unreconstructable vascular injury) and chronic pain. The most common indications for a below knee or above knee amputation are complications of diabetes. Functionally, the patients with below knee amputation are able to walk with prosthesis, with most patients walking within 3 months after surgery. Amputation should not be viewed as a limb salvage failure, but as a reconstructive procedure to improve function.

There are few contraindications to amputations; however, a contraindication to a below knee amputation is a non-ambulatory patient. A non-ambulating patient with a below knee amputation is at high risk for developing a flexion contracture, which can

result in increased pressure on the stump and cause ulceration. Therefore, when a patient is wheelchair bound and is not a candidate for prosthetic fitting, an above knee amputation should be considered.

Preoperative Planning

It is imperative that these patients be evaluated preoperatively to determine the vascular status of the limb. The majority of patients undergoing a below knee amputation or above the knee amputation have diabetes and have some component of peripheral vascular disease. Preoperative noninvasive vascular studies, including ankle brachial indexes, are important to determine the level of amputation. An ankle brachial index is determined by measuring the ankle systolic pressure and dividing it by the brachial systolic pressure using Doppler detection of the pulses. The severity of the arterial disease is related to decreased value of the ankle/brachial index (ABI), and a value of less than .5 is considered abnormal in people with diabetes. Also, noninvasive vascular studies using transcutaneous oxygen tension measurement (TcPo2) are useful in assisting with amputation levels. Amputations are likely to heal if the TcPo2 measurements are greater then 40 mm Hg. Patients with TcPo2 values less than 20 mm Hg are at higher risk for not healing and should be evaluated with further vascular testing and possibly an angiogram before surgical intervention. A vascular surgery consult is almost always indicated before performing an amputation. With the advances in distal bypass surgery and invasive radiologic procedures, certain patients can be successfully treated with limb salvage after vascular reconstruction.

Imaging studies are important in determining the underlying pathology. Imaging of the tibia or femur is important if there is a question regarding extension of tumor or infection into the tissues or

bone adjacent to the intended level of amputation or if there is a prosthetic device such as a total knee arthroplasty or internal fixation device in place, which may alter the surgical procedure.

If amputation is contemplated, optimizing the patient's medical condition before surgery is recommended. Literature has shown that patients with a serum albumin less than 3.5 g/dL or true lymphocyte count less than 1,500 cells per mL are at high risk for wound-healing difficulties.

Determination of the amputation level is important for both healing and function. The more distal the amputation level, the less energy required to ambulate. In elderly patients, a more proximal amputation may not allow for ambulation secondary to energy requirements. If a patient has good cognitive function, balance, and strength, then the most distal level with a realistic chance of healing should be attempted.

Surgery

In order for the below knee amputation to be performed correctly, proper attention to detail is important to improve the quality of the result. Gentle handling of the soft tissues, especially in diabetic patients, is important to minimize wound complications. The level of the amputation is determined by the extent of the infection, tumor, or the level that would provide optimal function with prosthesis. In general, the patient is positioned in the supine position for above or below knee amputations.

Full-thickness flaps should be used to minimize skin edge necrosis. Meticulous hemostasis and use of a drain is imperative to decrease the risk of a hematoma. The nerves should be divided sharply under tension to minimize the risk of a symptomatic neuroma. Also the bone ends are rasped until smooth to prevent bony prominences.

An open amputation is performed in patients with grossly contaminated wounds or in patients with extensive infection. These patients will require further surgeries to optimize the soft tissues around the stump, and then a definitive closure can be performed once this is accomplished.

Technique

Below Knee Amputation

Below knee amputation is the most commonly performed lower extremity amputation. A long posterior flap is used and brought anteriorly to cover the distal stump of the tibia, which should be 8.5 to 12.5 cm in length (Fig. 31-1). The flaps, if planned properly, will have minimal redundant skin in the corners, or â€œdog ears.â€• This will provide a good prosthetic fit.

The patient is placed supine on the operating room table. A nonsterile tourniquet is used and the leg is prepared and draped in the usual fashion. A skin marker is used to plan the flaps (see Fig. 30-1), and the flaps are drawn so that the posterior flap begins

two-thirds of the way posterior to the anterior aspect of the tibia and then extends distally and posteriorly so that the distance will be long enough to cover the tibia (Fig. 31-2). A tourniquet can be used at the discretion of the surgeon. The incision is then made through the skin and subcutaneous tissues down to the fascia. The subcutaneous nerves including the saphenous and sural nerve can be identified and divided under tension. The fascia is then incised. The anterior compartment musculature is cut with a cautery (Fig. 31-3) down to the deep peroneal nerve, which is identified and cut under tension (Fig. 31-4), and the anterior tibial artery is identified and tied with silk suture (Fig. 31-5). The superficial peroneal nerve is identified and cut under tension. The periosteum is reflected off the tibia (Fig. 31-6), and the tibia is cut 1 cm proximal from the skin incision (Fig. 31-7). A segment of the fibula is then resected 1 cm proximal to the tibial bone cut (Figs. 31-8 through 31-10). Traction is applied, and an amputation knife is used to perform the remaining portion of the amputation (Fig. 31-11). Dissection is carried deep until blood from the posterior tibial artery and vein is identified, and then the cut is beveled distally until the fascia is cut (Fig. 31-12). The anterior aspect of the tibia is beveled (Fig. 31-13). The remaining edges are rasped until smooth. The tibial nerve is identified and transected under traction (Fig. 31-14). The tibial artery and vein are identified and tied with silk sutures (Fig. 31-15). The wound is then copiously irrigated. If a tourniquet is used, it is deflated at this time. Hemostasis is obtained. The wound is then closed in layers over a drain with sutures in the fascia (Fig. 31-16). The subcutaneous layer is closed with monofilament suture, and the skin is then closed with nylon sutures in a vertical mattress fashion (Fig. 31-17). A sterile dressing is applied, and then a compressive Robert Jones dressing is applied with plaster in full extension (Fig. 31-18).

Above Knee Amputation

Above knee amputation is the second most frequently performed lower extremity amputation. Stump length is important for the lever arm control of the prosthesis.

P.391
P.392
P.393
P.394
P.395

Too long of a femoral stump, however, can lead to difficulty with fitting the prosthetic

knee joint. Therefore, the bone cut should be 10 to 12 cm proximal to distal femoral articular surface.

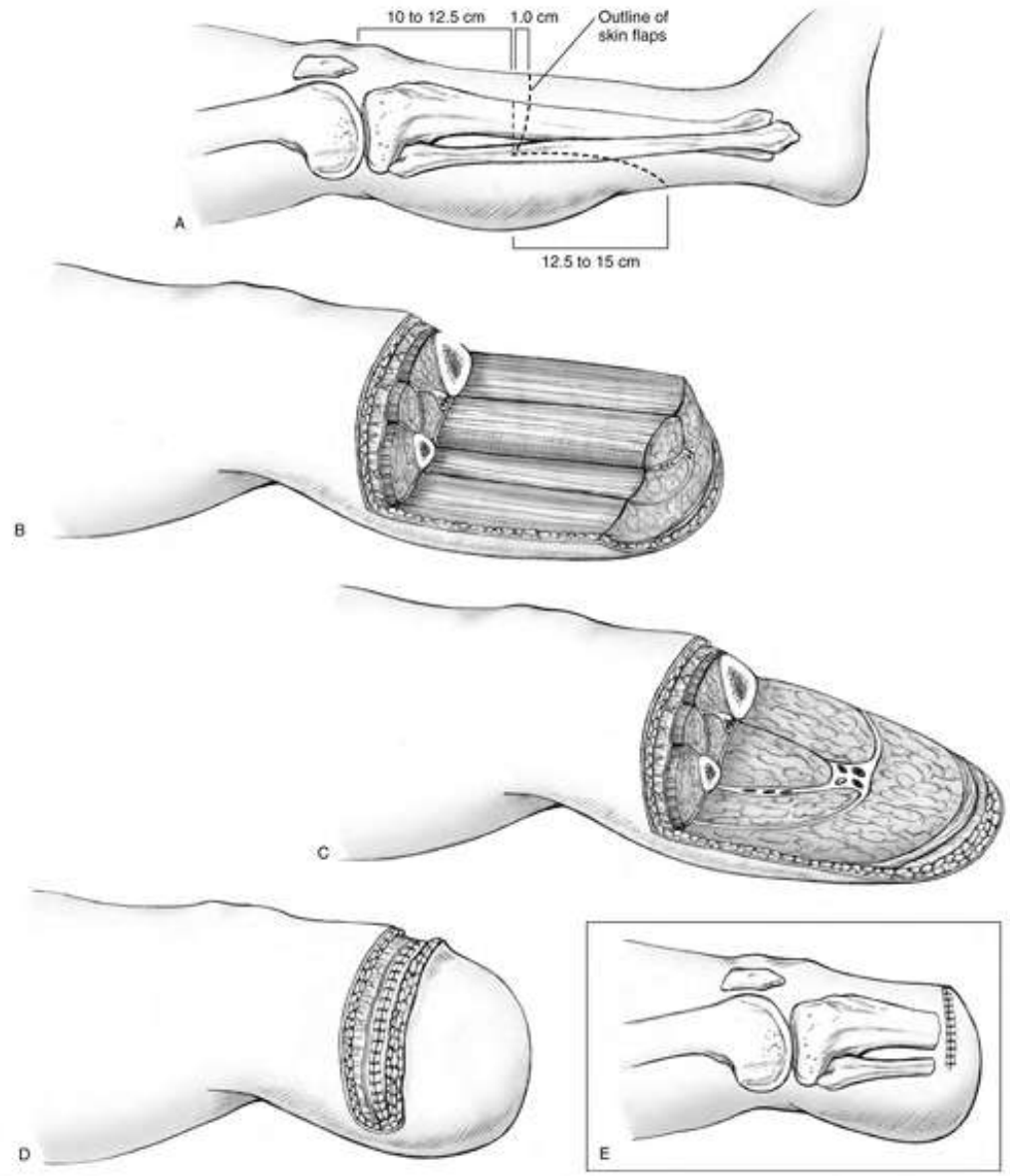

FIGURE 31-1 Below knee amputation. A: Planning of short anterior and long posterior skin flaps. B: Amputation of distal leg. C: Tailoring of posterior muscle to form flaps. D: Closure of flap to deep fascia. E: Closure of skin flaps. (Redrawn after Burgess EM, Zettl JH. Amputations below the knee. Artif Limbs. 1969;13:1. With permission.)

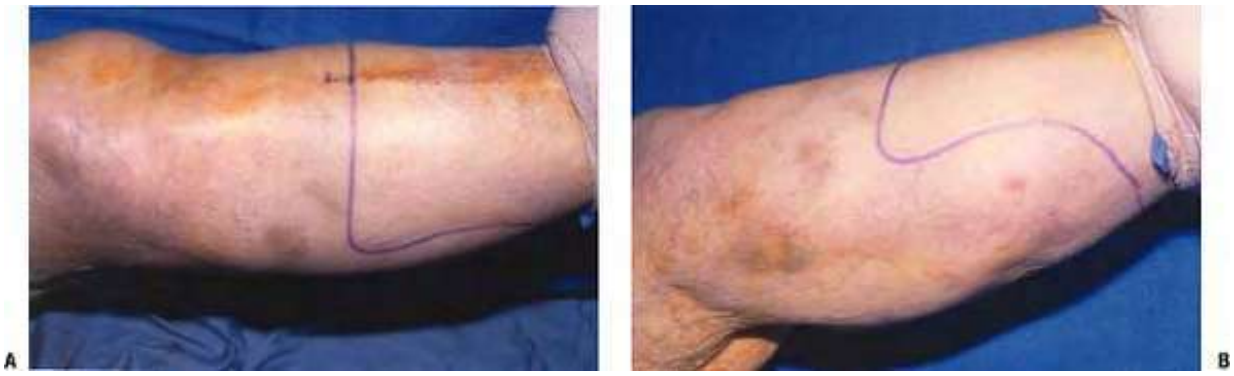

FIGURE 31-2 A,B: The long posterior flap is planned.

FIGURE 31-3 The anterior compartment muscle is dissected with cautery.

FIGURE 31-4 The deep peroneal nerve is identified and sharply transected.

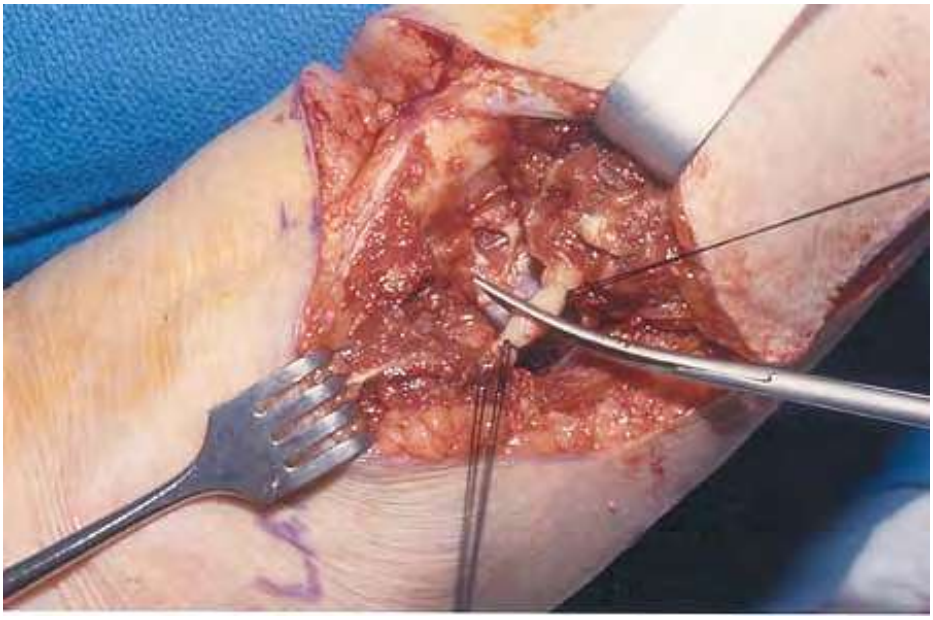

FIGURE 31-5 The anterior tibial artery and vein are ligated.

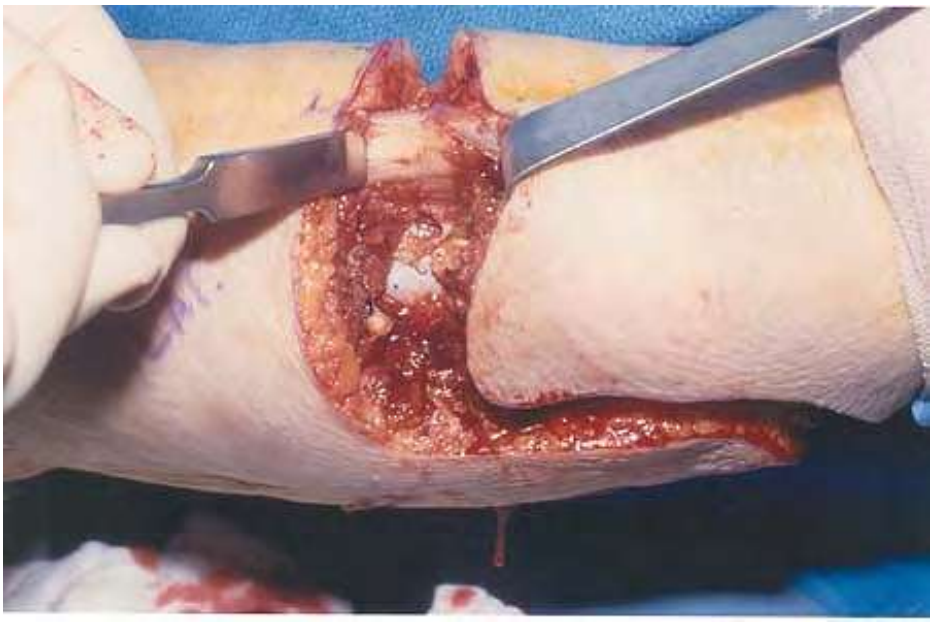

FIGURE 31-6 The periosteum is reflected off the tibia.

FIGURE 31-7 The tibia is osteotomized 1 cm proximal to skin incision.

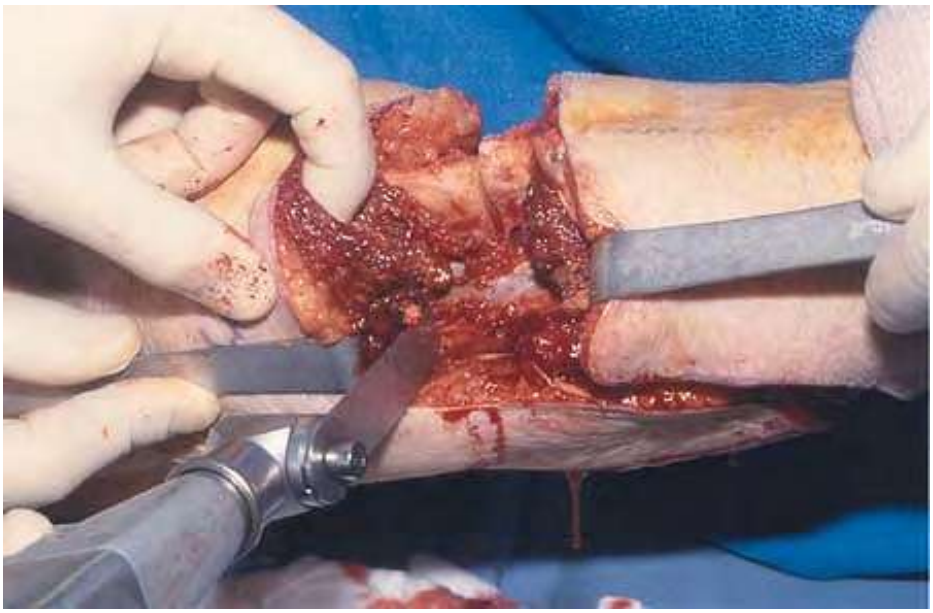

FIGURE 31-8 The fibula is identified and osteotomized 1 cm proximal to tibia osteotomy.

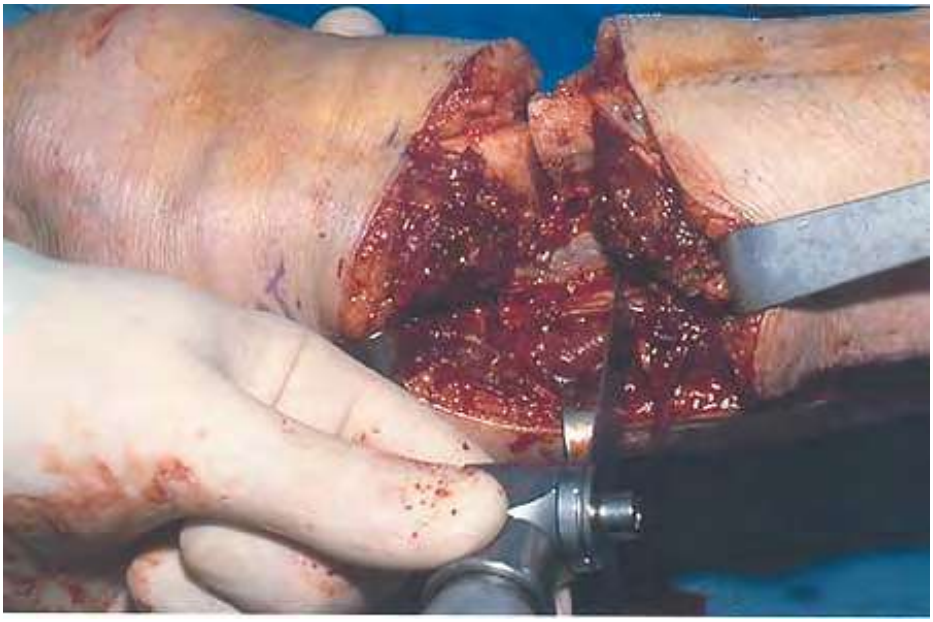

FIGURE 31-9 A second fibula osteotomy is performed 2 to 3 cm distal.

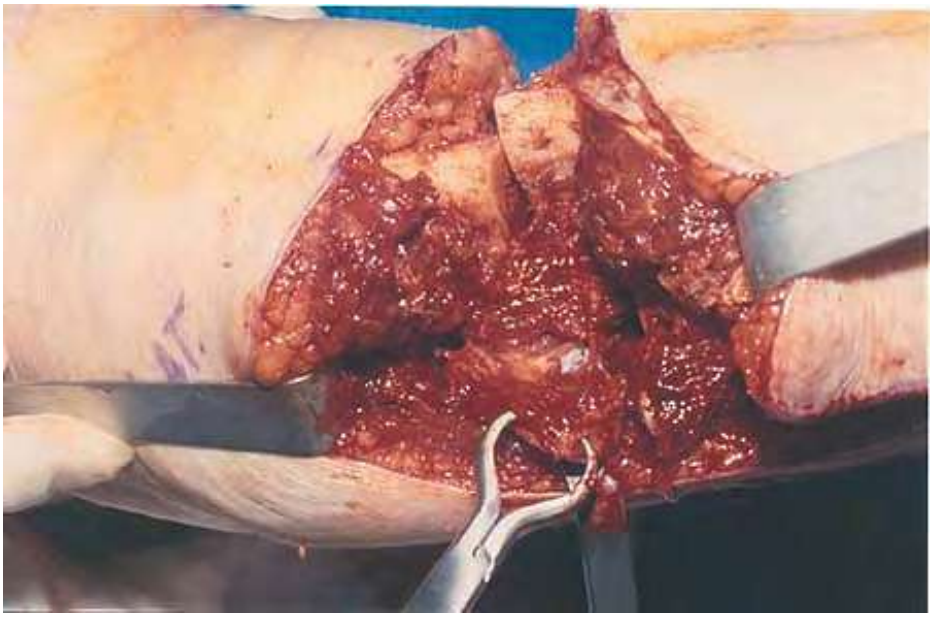

FIGURE 31-10 The segment of the fibula is removed.

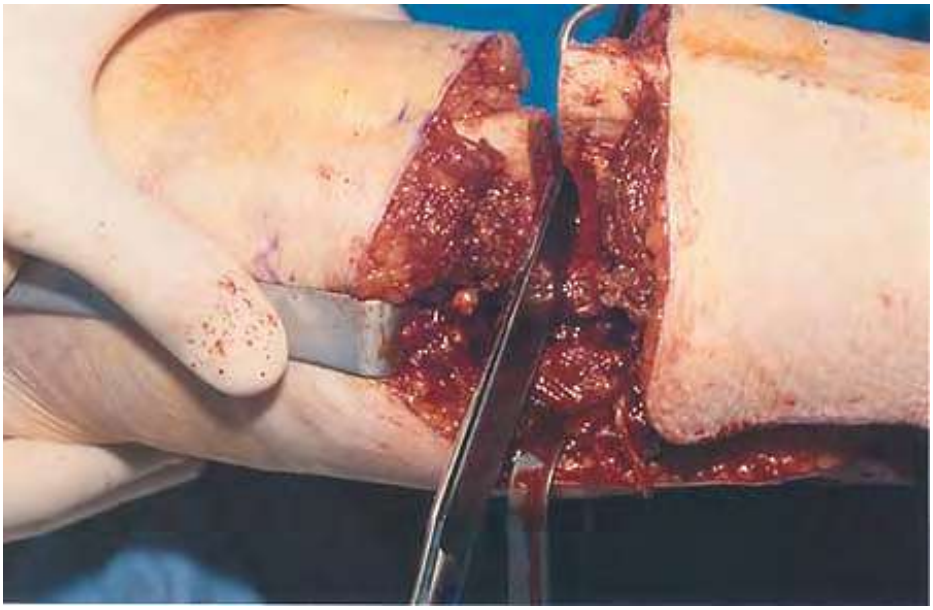

FIGURE 31-11 The amputation knife is used to sharply dissect the posterior compartment muscle.

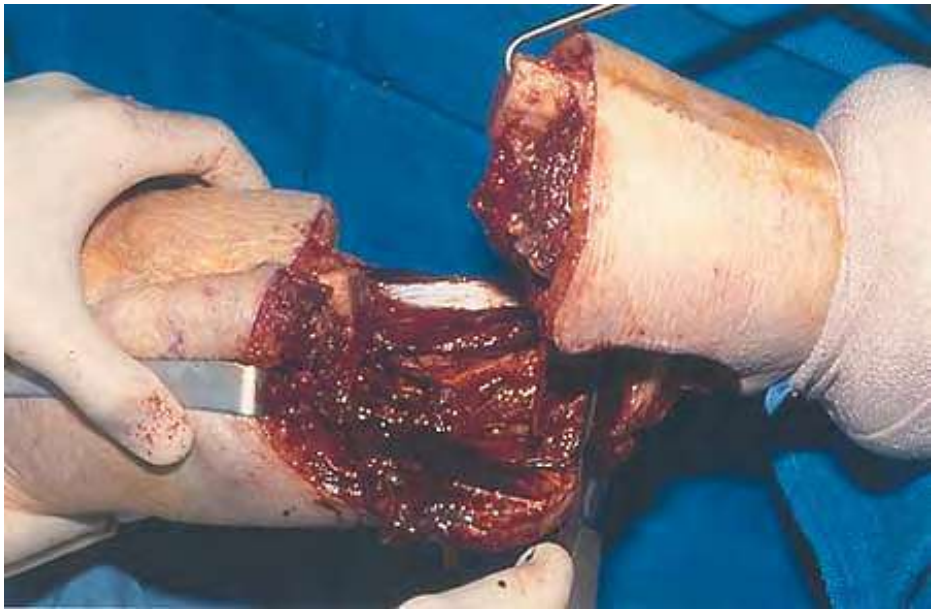

FIGURE 31-12 The posterior compartment muscle is beveled to minimize bulk.

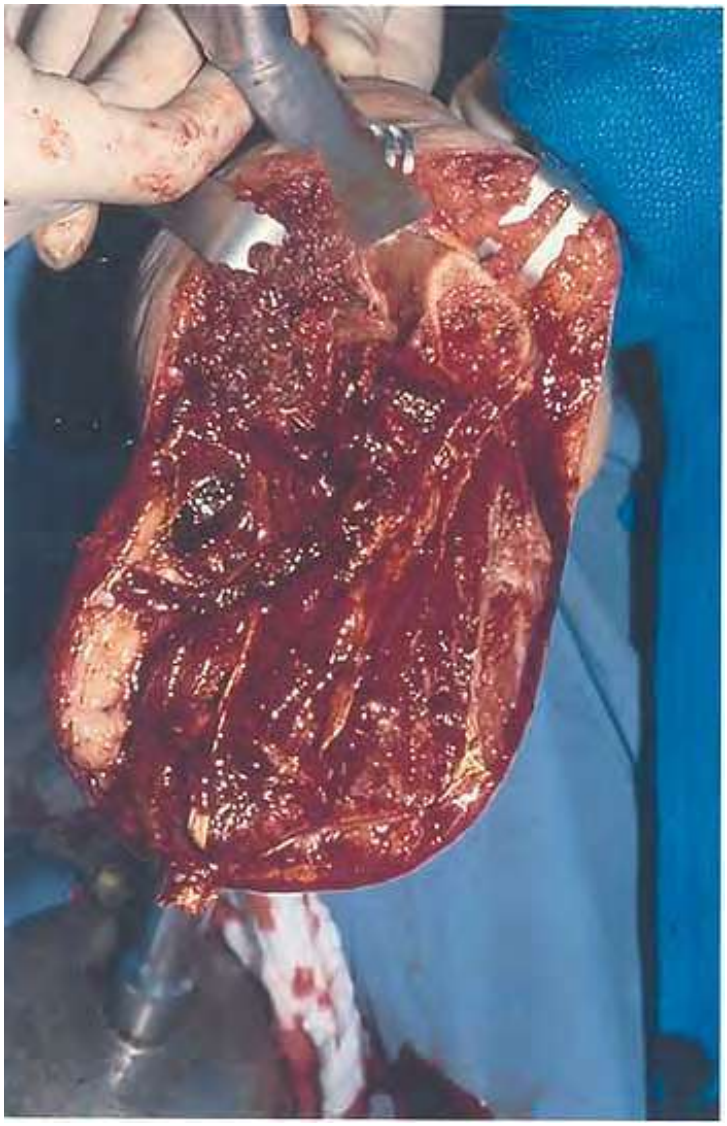

FIGURE 31-13 The saw is used to bevel the anterior tibia.

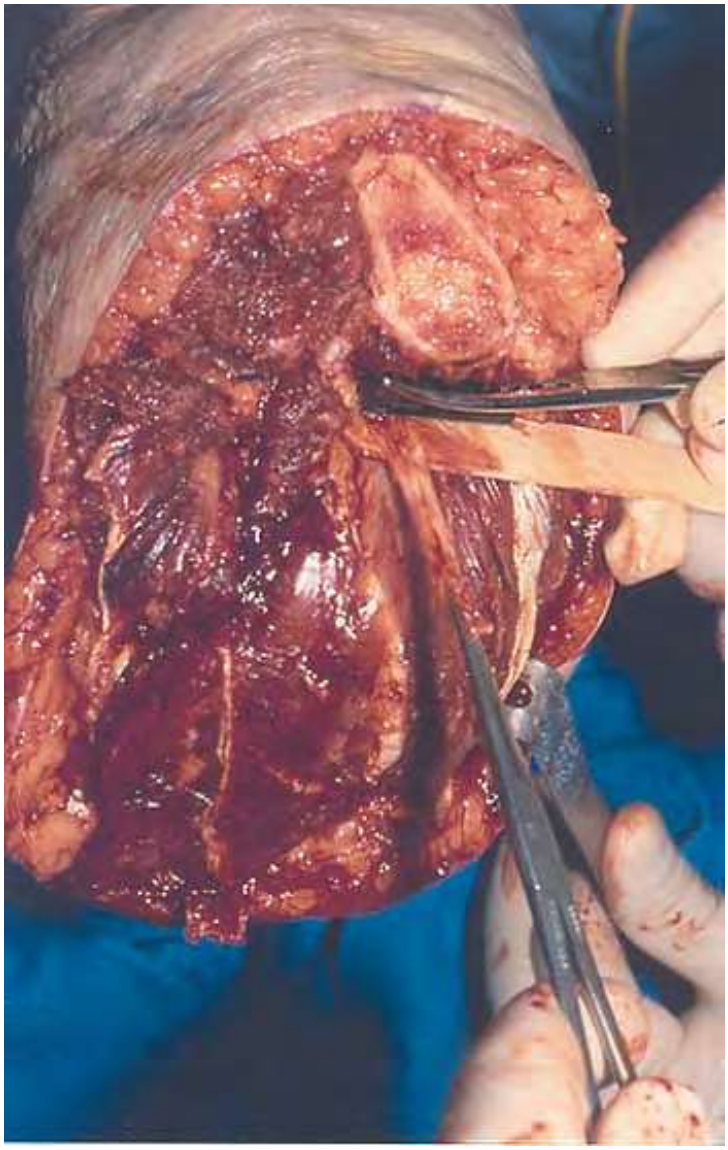

FIGURE 31-14 The tibial nerve is identified and sharply removed.

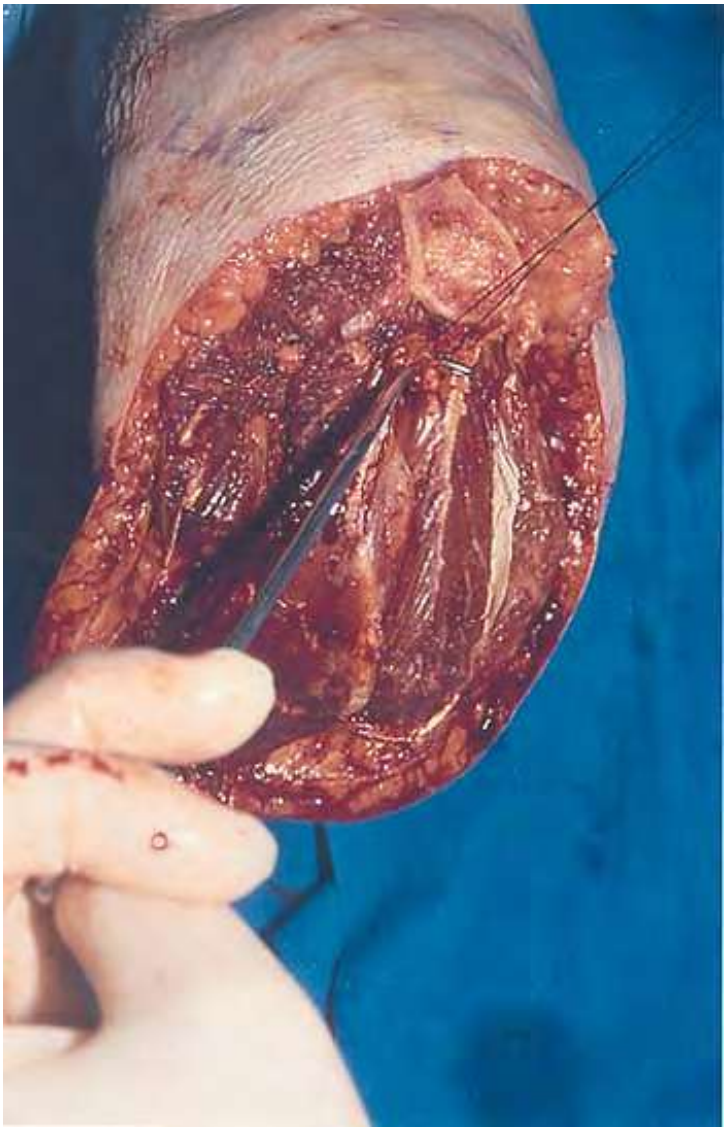

FIGURE 31-15 The tibial artery and vein are ligated.

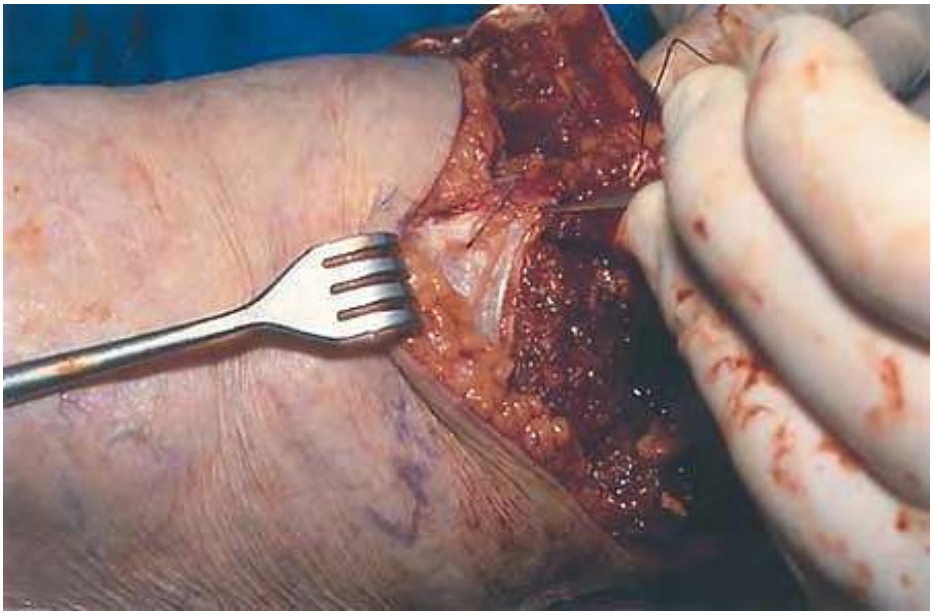

FIGURE 31-16 The fascia is closed with interrupted suture.

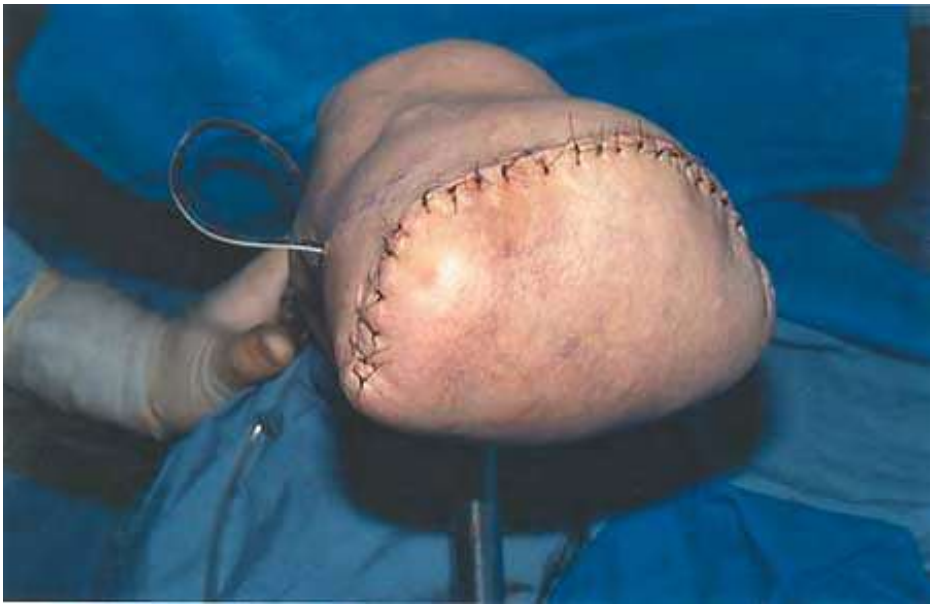

FIGURE 31-17 The skin is closed with interrupted vertical mattress suture.

FIGURE 31-18 A compressive bulky dressing with plaster is applied with the knee in full extension.

For non-ambulatory patients with ischemic disease, above knee amputations are performed with equal anterior and posterior flaps, and myodesis is not performed to prevent further vascular compromise. For patients with adequate vascular supply and the potential to ambulate, a myodesis is performed. Above knee amputation is performed with the patient in a supine position. Skin flaps are marked with a long medial flap and a shorter lateral flap (Fig. 31-19). A sterile tourniquet can be used at the discretion of the surgeon. Dissection is carried through the skin and subcutaneous tissues and down to the muscle. The muscles are then identified. The quadriceps is detached proximal to the patella, retaining some of its tendinous portion (Fig. 31-20). The vastus medialis is reflected off of the intermuscular septum, and the adductor magnus (Fig. 31-21) is detached from the adductor tubercle by sharp dissection and

reflected medially, exposing the femoral shaft. The vessels are identified (Fig. 31-22) at the level of Hunter's canal, and the artery and vein are ligated. The femur is then exposed proximally 12 to 14 cm above the condylar level and is cut with an oscillating saw approximately 10 to 12 cm above the joint line. The remaining edges are smoothed with a saw or rasp (Fig. 31-23). Small drill holes through the remaining distal femoral cortex are made with a 2.5-mm drill (Fig. 31-24). The adductor magnus tendon is then sutured with nonabsorbable suture to the lateral femur through the drill holes (Fig. 31-25). The femur is held in maximum adduction while this is being sutured to the bone as a myodesis (Fig. 31-26). The quadriceps is then brought over the bone and anchored to the posterior aspect of the femur through the drill holes (Fig. 31-27). The hip is in extension when this is done to try to minimize hip flexion. The fascia lata is then sutured to the medial fascia. Subcutaneous tissue is closed with monofilament suture, and the skin is closed with nylon suture (Fig. 31-28). Sterile dressing is applied, as well as a compressive wrap to minimize swelling.

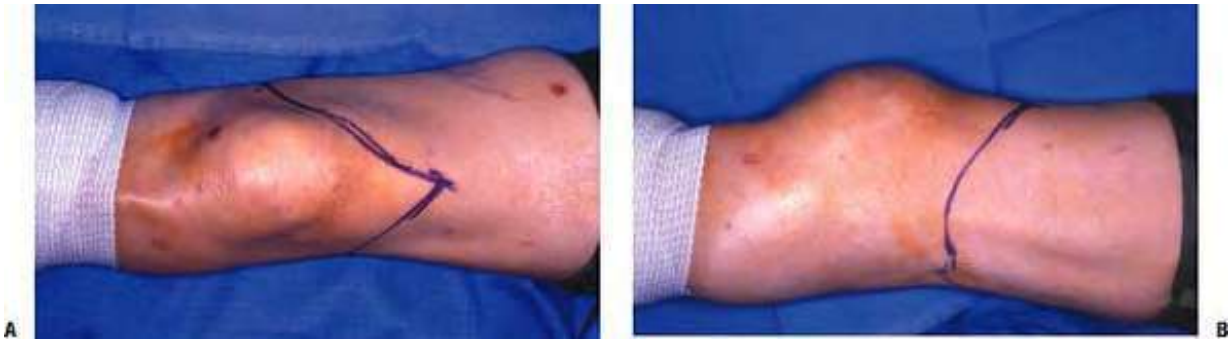

FIGURE 31-19 A,B: Skin flaps are marked with a long medial and shorter lateral flap.

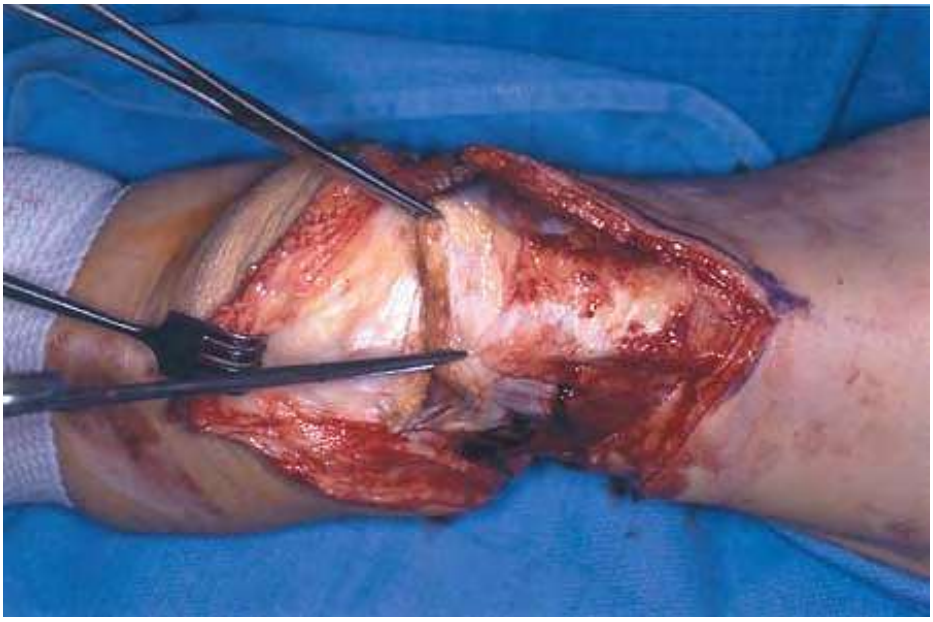

FIGURE 31-20 The quadriceps muscle is detached proximally to the patella, preserving some of the tendinous insertion.

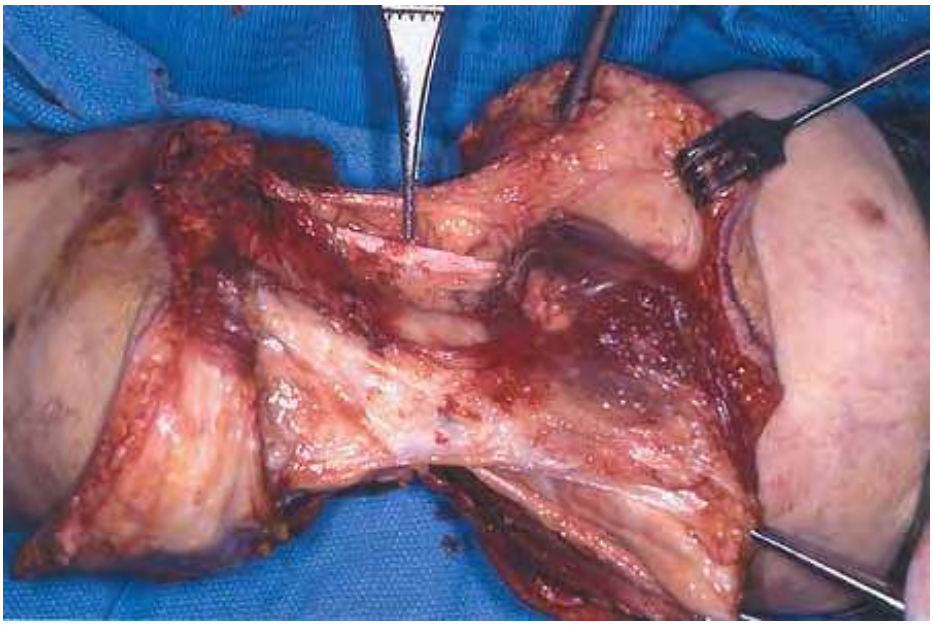

FIGURE 31-21 The adductor magnus is identified and reflected off the femur.

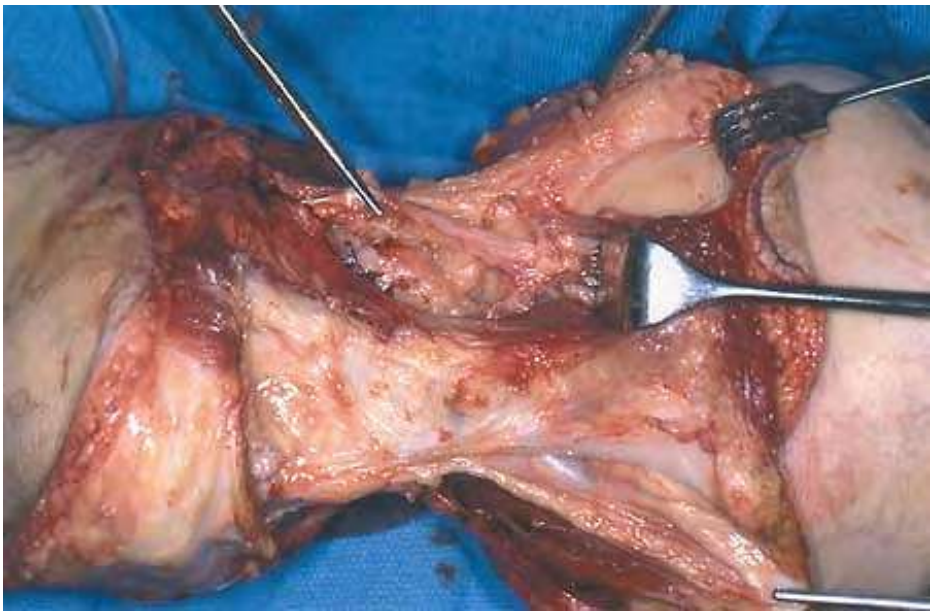

FIGURE 31-22 The femoral vessels at the level of Hunter's canal are identified and ligated.

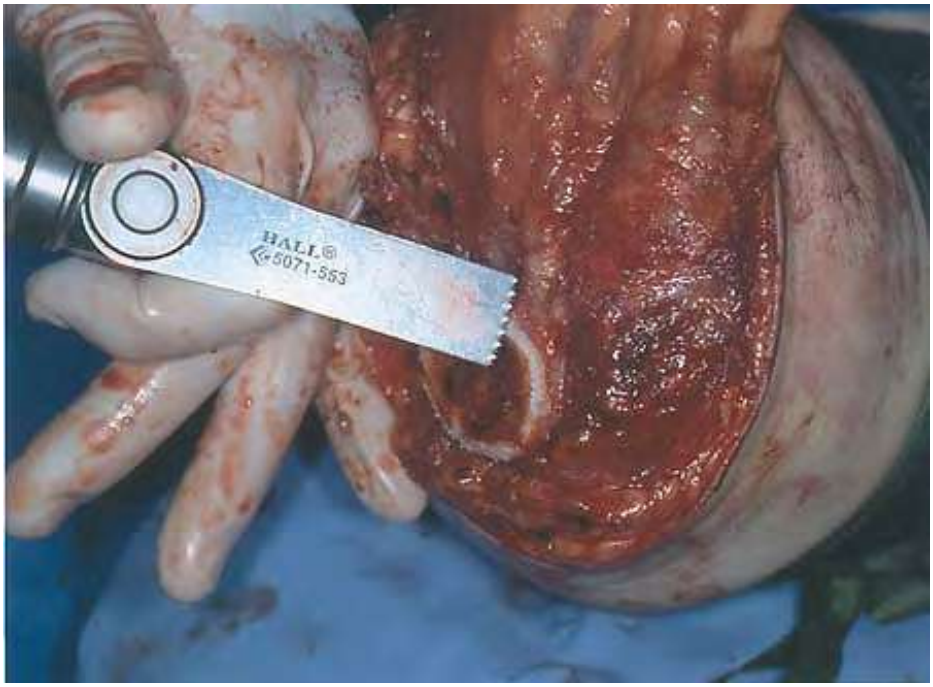

FIGURE 31-23 The femur is osteotomized and the end is rasped until smooth.

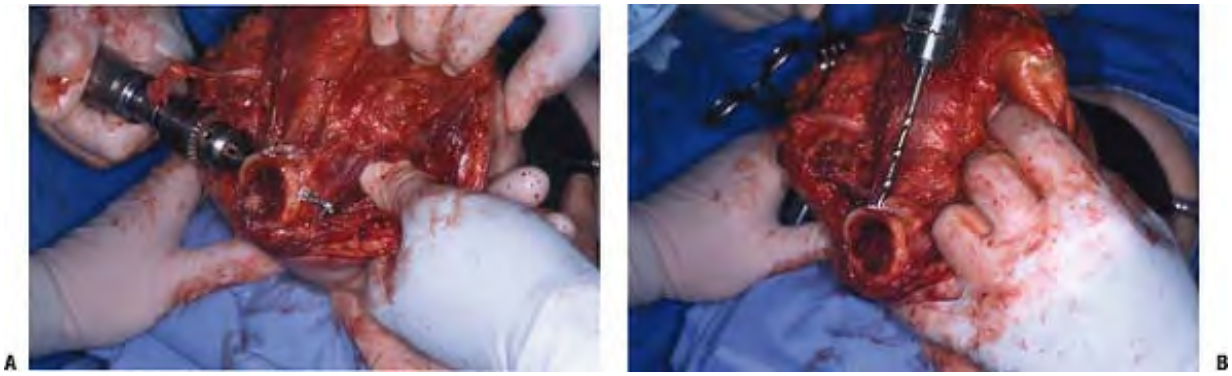

FIGURE 31-24 A,B: Four drill holes are made in the femoral stump for the myodesis.

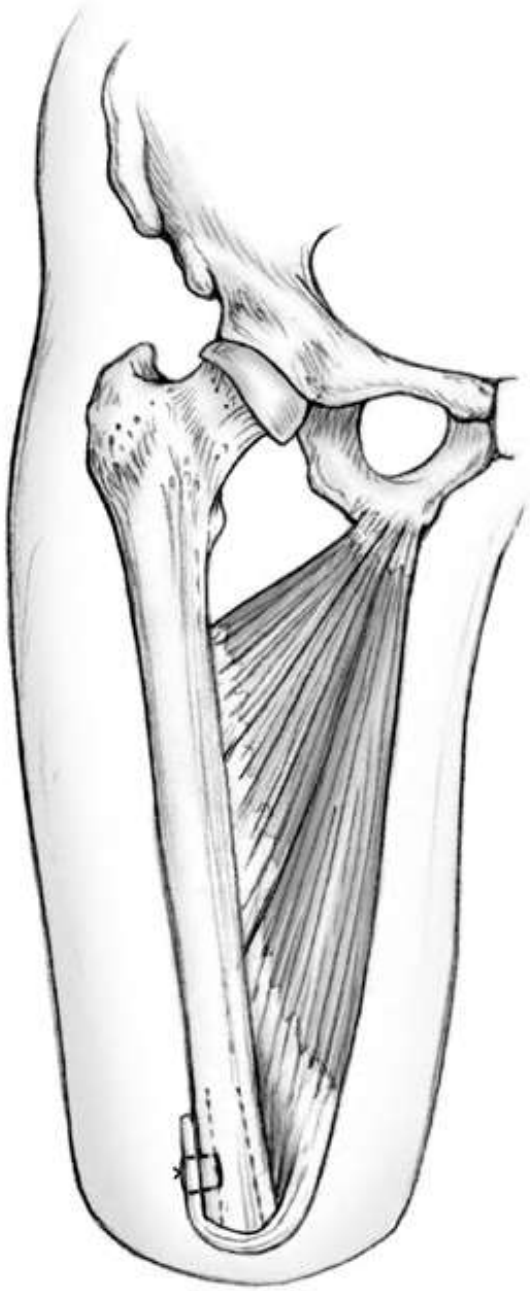

FIGURE 31-25 Attachment of adductor magnus to lateral femur. (Redrawn from Gottschalk F. Transfemoral amputations. In: Bowker JH, Michael JW, eds. Atlas of limb prosthetics: surgical, prosthetic, and rehabilitation principles. 2nd ed. St

Louis: Mosby; 1992.)

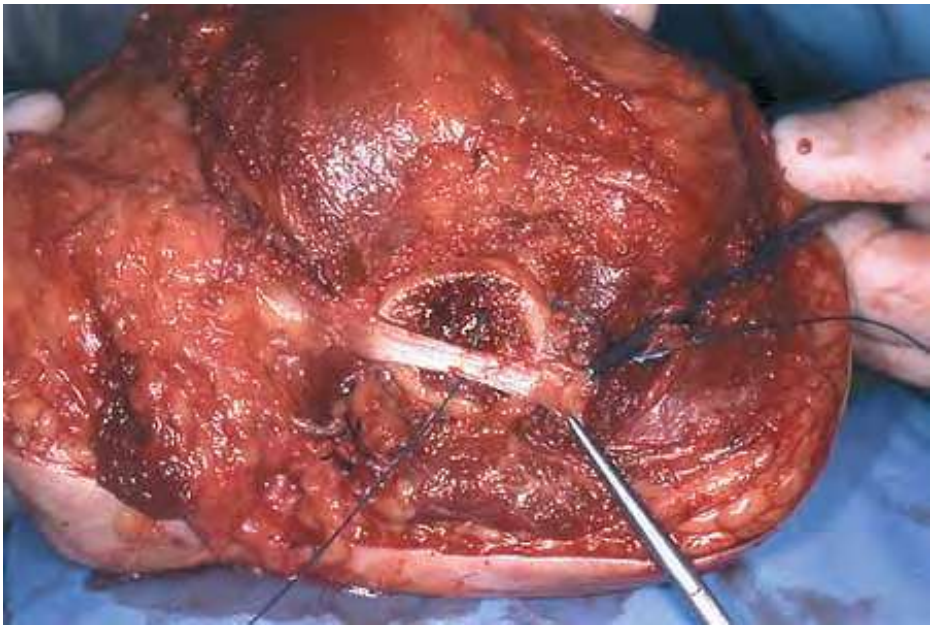

FIGURE 31-26 The adductor magnus tendon is sutured to the femur through the drill holes.

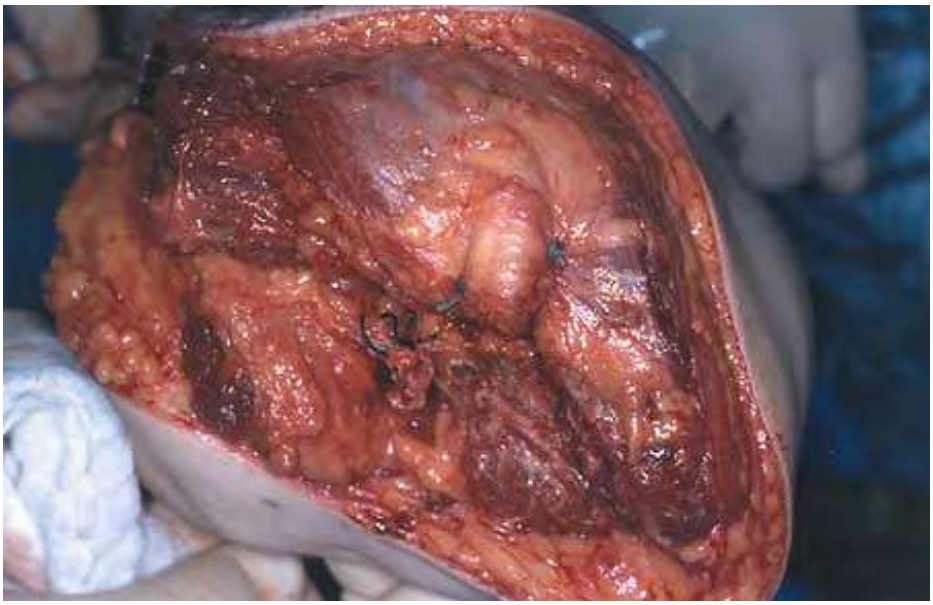

FIGURE 31-27 The quadriceps tendon is brought over the femoral stump and sutured through the drill holes.

FIGURE 31-28 The skin is closed with interrupted vertical mattress suture.

Postoperative Management

For both a below knee and an above knee amputation, a compressive dressing is used to minimize swelling. The patient is kept in bed for 24 hours. The suction drain is kept in for at least 24 hours or until the drainage has decreased to minimal output. For a below knee amputation, a compressive Robert Jones dressing is removed approximately 48 hours after the surgery. If the swelling is well controlled and the wounds are in good condition, a pilon cast can be applied with a temporary prosthetic foot. The patient is restricted to minimal weight bearing but is allowed to be up and ambulating with gait aids. The pilon cast is removed at 2 weeks after surgery and another one can be placed. At 4 weeks after surgery, the pilon cast is removed completely. Sutures are removed at 2 to 4 weeks depending on the healing of the wound. A stump protector is used between 4 to 6 weeks. Shrinker socks are used to shrink the stump, and once the stump volumes have stabilized, a permanent prosthesis can be fitted.

Above knee amputations are initially treated with a compressive dressing and a stump protector. At approximately 2 to 3 weeks after surgery, if the wounds are healed, the sutures are removed. These patients also use shrinker socks as well as a protector. Once their stump volume has stabilized, they can be fitted for their prosthesis.

Complications/Results

The most common complications of amputation surgery are soft tissue or wound healing problems. These can lead to superficial infections that can ultimately cause deep infection, and possibly progress to osteomyelitis. Such complications may result in more proximal amputations.

The treatment of wound complication initially begins with dressing changes, and frequently these patients will heal with local wound care. Occasionally, a vacuum-assisted closure device can be helpful in increasing the granulation tissue to improve the healing in a more timely fashion. Long-term complications include difficulty with

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 32

Soft Tissue Management of Ankle Fractures and Use of the Gracilis Muscle Flap

S. Andrew Sems

Matthew DeOrio

Steven L. Moran

Soft tissue management of injuries around the distal tibia and ankle region will often dictate and drive the timing and methods of definitive fixation. Injuries in this region vary from low energy ankle fractures to high energy tibial pilon fractures. The soft tissue injuries often reflect the amount of energy that was involved in creating the bony injury. Regardless of the radiographic appearance of the fractures, certain initial treatment principles hold true for both low and high energy injuries. Early fracture reduction to restore limb alignment, rotation,

and appropriate limb length combined with fracture immobilization will allow sooner resolution of soft tissue swelling and impairment that might otherwise prevent or delay early internal fixation.

Ankle Fractures

Indications/Contraindications

Operative treatment is generally recommended for all unstable ankle fractures in

which the talus is unable to be maintained in a position beneath the tibial plafond by closed reduction. Nonoperative management is appropriate for stable ankle fractures in which there is no lateral subluxation of the talus within the ankle mortise.

Comorbidities

Open reduction and internal fixation can generally be safely performed on most patients regardless of associated comorbidities. In patients with severe peripheral vascular disease, preoperative transcutaneous pressure oximetry measurements should be obtained to assess the likelihood of the patient healing the surgical wounds. TcPO2s less than 30 mm Hg have a higher risk of wound failure and fracture nonunion. Involving Vascular Medicine to assist with maximizing lower extremity perfusion is reasonable if vascular disease is a concern. The implications of insulin dependent diabetes mellitus and peripheral neuropathy should also be considered when determining operative versus nonoperative treatment but are not contraindications to surgery. Postoperative soft tissue management should be modified in patients with high risks of developing wound complications. Longer periods of cast immobilization and protected weight bearing may be indicated for patients with these comorbidities.

Nonoperative management of fractures in the nonambulatory patient is reasonable provided the fracture does not cause deformity which would result in subsequent skin breakdown. Modified techniques of internal fixation in patients who are nonambulatory may be utilized in order to maintain the anatomy around the ankle joint to prevent soft tissue compromise.

Preoperative Planning

Plain radiographs should be obtained in all patients with unstable ankle fractures prior to surgical intervention. Anteroposterior (AP), mortise, and lateral views are necessary to assess the fracture orientation. Ankle fractures tend to include the medial malleolus, lateral malleolus, and occasionally, the posterior malleolus. Recognition of the three-dimensional plane in which each fracture occurs is important so that surgical incisions are placed appropriately and the correct internal fixation is selected. While computed tomography is utilized more often for

pilon type tibial fractures, it may be used if there is not a clear understanding of the fracture pattern.

Surgical approach selection for treatment of ankle fractures depends on the location of the fractures and the quality of the surrounding tissues. Fractures of the medial malleolus are typically approached with an anteromedial incision overlying the medial malleolus. This incision can be adjusted anteriorly or posteriorly depending on the size and orientation of the medial malleolar fragment. Fractures which include both the anterior and posterior caliculus may require slight posterior adjustment of the incision while fractures that only involve the anterior caliculus may be approached through a more anteromedial incision. Fibular fractures are typically treated with a direct lateral approach, although a posterolateral approach is useful for fibula fractures which occur in the coronal plane or in cases of associated posterior malleolus fragments that may be addressed through this same incision.

The timing of surgery is dictated by the state of the soft tissue envelope. For ankle fractures, immediate fixation within the first 24 to 48 hours is feasible so long as the swelling will not compromise wound healing. Patients who present to the clinic or emergency department with ankle fractures within 48 hours from injury can often be treated with immediate open reduction and internal fixation prior to the onset of the maximal amount of soft tissue swelling. Surgery should be delayed when soft issue is so edematous that the surgeon is unable to create skin wrinkles by gently pinching the skin over both the medial and lateral malleolus. Open fractures require immediate surgical intervention for debridement of all contaminated soft tissues. Antibiotic cement bead pouches or Vacuum Assisted Closure (VAC) devices may be used until definitive fixation or final wound closure. External fixation may be necessary to restore length, alignment, and rotation of both the soft tissues and the bone prior to definitive internal fixation.

Ankle fractures should ideally be treated within 3 weeks of injury prior to early callus formation, which can create significant difficulty in fracture reduction. Patients who present in the first week following an ankle fracture and whose soft tissues are not amenable to immediate surgical fixation should be treated with a Robert Jones type compressive dressing with plaster immobilization. These patients should be encouraged to maintain strict elevation of their limb at all times. They

are then seen back in

the clinic or preoperative area approximately 7 to 10 days after their injury for soft tissue evaluation prior to surgical intervention. This dressing should be taken down and the skin should be evaluated prior to any type of anesthetic administration. In patients who are admitted to the hospital, sequential compression boots can be applied to the foot prior to the application of the Robert Jones dressing. These pneumatic boots are usually well tolerated so long as they do not extend up past the ankle.

Surgery

Medial/Lateral Approaches

Patient Positioning

The patient is positioned supine on a radiolucent table with a thigh high pneumatic tourniquet and a small bump (towel) placed under the ipsilateral hip to prevent excessive external rotation of the limb. The bump placed underneath the hip area should be adjusted so that the foot points vertically when in its resting position. A foam block or bump of towels can be utilized underneath the leg and ankle area so as to elevate the operative limb from the nonoperative limb. This facilitates obtaining lateral radiographs without manipulation of the operative leg. The intraoperative fluoroscopy unit is brought in from the contralateral side, and it is positioned perpendicular to the long axis of the patient.

Prophylactic antibiotics are utilized for internal fixation of ankle fractures. They should be administered prior to inflation of the tourniquet. Once the tourniquet is inflated, it should be maintained as long as necessary, but not for more than 2 hours. If it is impossible to complete the procedure in the allotted 2 hours, the tourniquet should be deflated and hemostasis should be obtained, and the procedure should be completed without the aid of a tourniquet. If small vessel bleeding continues to be an issue, consider moving the operative bed into a Trendelenburg position and asking the anesthesiologist to decrease the systolic blood pressure within safe limits.

Technique

Anatomic Landmarksâ€"Medial Approach

Palpate the medial malleolus to define the location of the anterior aspect of the anterior caliculus, the posterior aspect of the posterior caliculus, as well as the fracture location. The leg should be palpated to determine the longitudinal axis of the tibia.

Incision and Surgical Approach

The incision is made along the midlateral axis of the distal tibia. Beyond the tip of the medial malleolus the incision may be slightly curved anteriorly (Fig. 32-1). Be aware that excessive curvature of the incision may create difficulties in placing internal fixation in the medial malleolar fragment. Dissection should be carried through the skin and subcutaneous and hemostasis obtained as this is performed (Fig. 32-2). Care should be taken to preserve and protect the saphenous vein and nerve throughout the case. In this location, the saphenous vein is often very subcutaneous and an aggressive skin incision may lacerate the vein. With the saphenous

vein and nerve protected, the medial malleolar fragment can be retracted distally with sharp bone hooks to allow inspection of the ankle joint (Fig. 32-3). A 15 blade, used to elevate the periosteum along the fracture edges, can be utilized to ensure that the fracture is reduced in an anatomic position. Preservation of the soft tissue and minimal periosteal stripping is recommended. Following internal fixation, the subcutaneous tissue can be closed using a 2-0 absorbable suture, and the skin closed with a nonabsorbable monofilament suture (Fig. 32-4).

FIGURE 32-1 The incision is made along the mid-lateral axis of the distal tibia, curving anteriorly beyond the tip of the medial malleolus.

FIGURE 32-2 Exposure of the medial malleolar fracture line prior to elevation of the periosteum along the fracture edges.

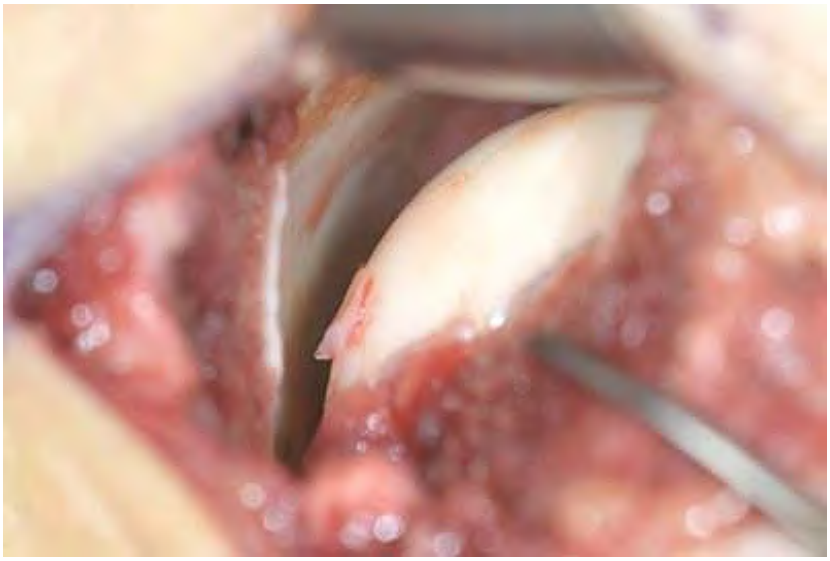

FIGURE 32-3 Retraction of the medial malleolus allows exposure of the talus and ankle joint.

Lateral Approachâ€”Lateral Malleolar Fixation

The direct lateral approach to the fibula allows placement of fixation along the lateral aspect of the fibula. This incision is positioned in line with the longitudinal axis of the fibula and may be curved anterior distally to allow access for reduction of a Chaput-Tillaux fragment (Fig. 32-5). If plate fixation along the posterior aspect of the fibula is desired, such as in a situation when an anti-glide plate is used, rather than committing excessive soft tissue stripping via a direct lateral approach, the incision can be adjusted posteriorly. The superficial peroneal nerve may cross the surgical approach in a subcutaneous location; therefore the skin incision should go no deeper than the skin (Fig. 32-6). Dissection through the subcutaneous tissues should be performed using combinations of sharp and blunt dissection with care to

identify and protect the superficial peroneal nerve if encountered. Periosteal elevation at the fracture edges with a 15 blade is recommended to ensure an anatomic reduction of the fibular fracture.

For long spiral oblique fractures in which lag screw only fixation of the fibula is to be used, this incision can be altered with an apex anterior curve along the midportion of the incision. This â€œwaveâ€• in the incision will allow anterior retraction of the soft tissues and appropriate directional placement of lag screws across the fracture site. The same concerns remain with regards to protection of the superficial peroneal nerve when this modification is performed.

Following internal fixation, closure is performed with interrupted 2-0 absorbable sutures in the subcutaneous tissue followed by nonabsorbable monofilament sutures in the skin (Fig. 32-7).

FIGURE 32-4 Skin closure is performed using a nonabsorbable monofilament suture.

FIGURE 32-5 The incision is positioned in line with the longitudinal axis of the fibula, curving anteriorly at the distal end to allow access for reduction of a Chaput-Tillaux fragment.

FIGURE 32-6 The superficial peroneal nerve may cross the incision, and care should be taken to identify and protect this structure.

FIGURE 32-7 Skin closure is performed using a nonabsorbable monofilament suture.

FIGURE 32-8 Intraoperative fluoroscopy can be utilized in either the lateral or prone position, and image quality should be confirmed prior to incision.

Posterolateral Approachâ€"Posterior Malleolus Fractures

Patient Positioning

Fractures which require fixation of the posterior malleolus can be approached through a posterolateral incision. Controversy exists regarding the size of the fragment, but internal fixation is typically recommended for fractures which include 20% to 30% of the articular surface. This approach requires the patient to be in either a lateral or prone position (Fig. 32-8). The prone position allows direct visualization and better stabilization of the limb during the approach and is therefore preferred. The patient should be placed on a well-padded radiolucent table with care to be taken to prevent hyperextension of the shoulders and neck. If the arms are placed in an abducted position, care should be taken to avoid direct compression of the ulnar nerve at the cubital tunnel. A thigh high tourniquet is utilized and can be placed prior to prone positioning. Once the patient is in the prone position, they are translated towards the foot of the bed so that the hindfoot is slightly hanging off the end of the bed to aid with reduction and fixation. The C-

arm fluoroscopy unit is placed perpendicular to the longitudinal axis of the patient and intraoperative films can be used on the patient's uninjured side to guarantee symmetry following reduction.

Technique

Palpation of the fibula and Achilles tendon is possible due to the subcutaneous location of these structures. The incision is made midway between the posterior border of the lateral malleolus and the lateral border of the Achilles tendon and will need to extend to the tip of the fibula distally and as far proximally as necessary to obtain visualization of the fibular fracture and the posterior malleolar fracture (Fig. 32-9). The short saphenous vein and sural nerve are located immediately behind the lateral malleolus and the incision is placed posterior and medial to these structures. The deep fascia of the leg is split in line with the incision and the peroneal retinaculum is incised to release the peroneus longus and brevis (Fig. 32-10). The peroneal tendons and muscles are retracted laterally and the flexor hallucis longus is elevated from its origin on the fibula and retracted medially (Fig. 32-11). This allows access to the posterior aspect of the tibia for reduction and fixation (Fig. 32-12). The peroneal tendons are retracted medial or lateral depending on the location of the fracture. For fibular fractures which require a more proximal exposure, dissection to the lateral aspect of the peroneal tendons and muscles is necessary.

Following internal fixation of the posterior malleolus fragment and/or lateral malleolar fracture through the posterolateral approach, the peroneal retinaculum is repaired utilizing a 2-0 absorbable suture (Fig. 32-13). If possible, the deep fascia of the leg is also reapproximated using the same absorbable suture. Subcutaneous tissues are closed with a 2-0 absorbable suture and the skin is closed in interrupted fashion using a 3-0 or 4-0 nylon suture (Fig. 32-14).

Postoperative Management

Following internal fixation, the patient is placed in a well-padded Robert Jones splint to provide further postoperative immobilization (Fig. 32-15). The splint is applied with the knee flexed to allow dorsiflexion

of the ankle to a neutral position and avoidance of equinus positioning. This splint is maintained for the first 10 days to allow for postoperative swelling. The patient is then converted to either a short-leg cast or a fracture boot following this initial Robert Jones dressing. The sutures are removed 2 to 3 weeks from the date of surgery, however they should remain in place if there is any concern of delayed wound healing. Weight bearing restrictions are tailored to the individual fracture pattern and in patients with excellent bony quality and excellent healing potential weight bearing may be started at 6 weeks. In patients with multiple comorbidities, poor bone quality, or concerns for delayed fracture healing, weight bearing may be limited for up to 3 months post-operatively.

FIGURE 32-9 The incision is made midway between the posterior border of the lateral malleolus and the lateral border of the Achilles tendon.

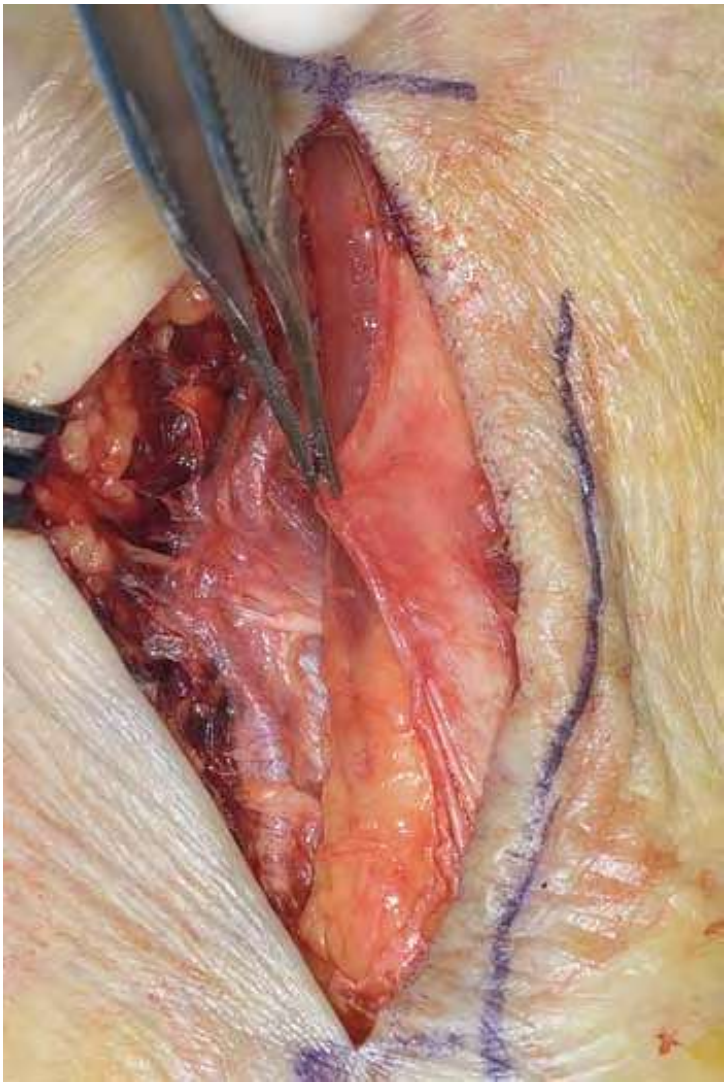

FIGURE 32-10 The deep fascia of the leg is split in line with the incision and the peroneal retinaculum is incised to release the peroneus longus and brevis.

Complications

Postoperative wound dehiscence

With appropriate soft tissue management and timing of internal fixation of ankle fractures, postoperative wound dehiscence should be a relatively infrequent

complication. Due to the relatively subcutaneous location of the internal fixation, any wound complication involving full thickness skin dehiscence or necrosis should be considered to communicate with the hardware and appropriate aggressive surgical intervention involving debridement and dressing changes should be performed.

Patients with partial skin necrosis or minor wound dehiscence that is located away from the hardware, particularly on the medial side, can be treated with local wound care including wet-to-dry dressing changes. If this is noted prior to suture removal, the sutures should be kept in place during these dressing changes, particularly near the portion of the wound adjacent to the dehiscence. Surrounding cellulitis should be treated with appropriate antibiotics.

FIGURE 32-11 The peroneal tendons and muscles are retracted laterally and the flexor hallucis longus is elevated from its origin on the fibula and retracted medially.

FIGURE 32-12 Following exposure, the reduction and fixation of the posterior malleolar fragment may be completed.

FIGURE 32-13 The peroneal retinaculum is repaired using an absorbable suture.

FIGURE 32-14 The skin is closed using a nonabsorbable monofilament suture.

FIGURE 32-15 A well-padded Robert Jones splint is placed to provide further postoperative immobilization.

Tibial Pilon Fractures

The â€œpersonalityâ€• of tibial pilon fractures is much different from that of ankle fractures. Tibial pilon fractures usually represent a much higher energy injury with more involvement of the surrounding soft tissues. While immediate open reduction and internal fixation of some ankle fractures is feasible, the same is not true for pilon fractures. The skin around the distal tibia is not very tolerant of excessive swelling with early operative fixation and therefore allowing sufficient time for solution of swelling is necessary. Initial series of tibial pilon fractures treated with immediate open reduction and internal fixation had significantly higher rates of postoperative wound complications compared with later series in which a staged protocol was utilized. Initial management of pilon fractures is with limited internal fixation of the fibula and spanning external fixation to gain appropriate limb alignment and length. This will provide appropriate soft tissue stabilization to allow for resolution of the swelling that will inevitably occur following tibial pilon

fractures.

Preoperative Planning

Initial evaluation of tibial pilon fractures consists of AP and lateral x-rays of the tibia as well as AP, lateral, and mortise radiographs of the ankle. When a pilon fracture has an associated fibula fracture, there is often shortening of the limb. After appropriate soft tissue evaluation, management consists of initial open reduction and internal fixation of the fibular fracture through a lateral or posterolateral approach in addition to application of an external fixator from the tibia to the calcaneus. This

external fixator should be positioned so that the limb length is restored and the tibiotalar joint is reduced. Distraction of the joint may be necessary to prevent â€œcartilage necrosisâ€• if the articular surface is comminuted and irregular. Preoperative planning is important because the incision for fibular fixation should be tailored to the anticipated approach for the tibial internal fixation. A posterolateral approach to the fibula may be utilized when a future anterolateral approach to the tibia is planned. When an anteromedial approach to the tibial pilon fracture is anticipated, a direct lateral approach to the fibula is appropriate. By utilizing a posterolateral approach to the fibula, an appropriate 5 or 6 mm skin bridge can be preserved between this approach and an anterolateral approach to the tibia. A direct lateral approach to the fibula should not be combined with an anterolateral approach to the tibia because the proximity of the incisions and resultant narrow skin bridge may result in skin necrosis between the incisions.

Once careful preoperative planning has determined the initial approaches for fibular stabilization, the patient should be positioned on a radiolucent table. A thigh high pneumatic tourniquet may be utilized during the initial fibular reduction and stabilization. Fixation of the fibula is generally performed prior to application of the external fixator. In the acute setting, fibular reduction is usually easily obtainable as the soft tissues have not yet contracted. In situations when difficulty is encountered gaining fibular length, the external fixator may be applied before fibular reduction. Careful construction of the external fixator will allow access to the fibula during internal fixation.

If the initial stabilization is performed within 24 hours, there is frequently minimal to moderate soft tissue swelling and fibular fixation can be performed. In the polytrauma setting, when fibular fixation may not be performed within 48 hours, careful soft tissue evaluation should be made before planning internal fixation. If the skin does not wrinkle when performing the pinch test, then fibular fixation should be delayed; however, spanning external fixation from the tibial to the calcaneus should not be delayed and this should be performed at the first possible time following injury. The external fixator improves length, angulation, and rotation of the fracture, but more importantly, it restores the appropriate soft tissue tension and can prevent further soft tissue compromise and development of fracture blisters. Tibial pin placement should be proximal enough that anticipated tibial plate fixation will not be contaminated by the pin tracts. A calcaneal transfixion pin is frequently utilized, however pins may also be placed in the talar neck or metatarsals to provide further support to the foot to prevent equinus positioning of the foot.

Following initial spanning external fixation, the patient's lower extremity should be elevated at all times to decrease soft tissue swelling. Deep venous thrombosis prophylaxis should be utilized for these now relatively immobile patients who have joint spanning external fixators in place. Repeat clinical evaluation should be performed on a weekly basis to monitor progression of a soft tissue injury. Once the swelling has subsided to the point that soft tissue wrinkles are attainable over the planned surgical incisions, internal fixation may be performed. There is frequently a 2 to 3 week interval between initial injury and resolution of soft tissue swelling. Once the soft tissues are amenable to surgical intervention, a duplex ultrasound screening examination should be performed on the injured extremity prior to removal of the external fixator and planned internal fixation. Avoidance of tourniquet should be considered in patients who have developed deep venous thrombosis in the injured extremity.

Once the soft tissue swelling has subsided, careful preoperative planning is required to determine appropriate patient positioning, operative approach, and implant choice for internal fixation. Planning consists of reviewing imaging studies, which should include a post reduction CT scan once limb length has been restored with the use of the external fixator and/or fibular fixation. Operative approaches should be selected which allow access to the fracture segment that requires anatomic

reduction. Pilon fractures with an associated medial malleolar fracture or coronal split which extends to the fibula are often best approached through a medial or anteromedial approach. Pilon fractures with significant articular comminution, particularly in the anterolateral area, can be approached through an anterolateral approach. Fractures with significant posterior comminution are best approached through a posterolateral approach as previously described for ankle fracture fixation. Two incisions are often necessary and preoperative planning is required to ensure that there is a minimum 6 cm skin bridge between the incisions to prevent skin necrosis.

Posterolateral Approach to the Distal Tibia and Fibula

The patient is positioned in a prone position on a radiolucent operative table. The patient should be translated towards the foot of the bed so that the midfoot and toes are hanging off the bed so the foot

is free. A well-padded pneumatic tourniquet may be placed around the thigh and should be placed prior to placing the patient in a prone position. The landmarks of the lateral malleolus and Achilles tendon are identified by palpation. A longitudinal skin incision is made half-way between the lateral border over the Achilles tendon and the posterior border of the lateral malleolus. The incision should extend distally until it reaches the level of the tip of the lateral malleolus. The deep fascia of the leg is incised in line with the incision, and the peroneal musculature and tendons are identified and retracted laterally. The peroneal retinaculum is incised distally to allow lateral retraction of these tendons. The flexor hallucis longus is elevated from its origin on the fibula and retracted medially to gain access to the posterior aspect of the tibia. Once reduction and fixation have been completed, the peroneal retinaculum and deep fascia of the leg are closed utilizing an absorbable suture. The subcutaneous tissue is closed with a 2-0 resorbable suture and the skin is then closed utilizing either 3-0 or 4-0 nonabsorbable monofilament sutures such as nylon.

Anterolateral Approach to the Distal Tibia

Patient Positioning

The patient is positioned supine on a radiolucent operating table and a well-padded pneumatic tourniquet is placed around the thigh. A bump is placed under the operative hip so that the foot rests vertically in the neutral position. The intraoperative C-arm is positioned on the uninjured side perpendicular to the long axis of the patient. A foam block or several folded towels can be utilized to elevate the operative limb higher than the nonoperative limb to provide a clear image with fluoroscopy. This also minimizes frequent manipulation of the leg that could potentially compromise the reduction.

Surgical Technique

For the anterolateral approach, a longitudinal skin incision is made in line with the fourth metatarsal. The incision is extended proximally as far as necessary to expose the tibia for application of a plate of the appropriate length and distally will end over the talus. The superficial peroneal nerve should be protected and retracted laterally in the wound. The subcutaneous nature of the superficial peroneal nerve requires careful skin dissection and blunt dissection once the skin has been incised. The fascia and extensor retinaculum are incised in line with the skin incisions with care taken to avoid undermining the skin flaps. The extensor digitorum longus is retracted medially and the peroneal muscles are retracted laterally. During medial dissection, avoid injury to the deep peroneal nerve and anterior tibial artery as they cross the ankle joint.

Following reduction and fixation through the anterolateral approach, the superior and inferior extensor retinaculum is repaired utilizing absorbable suture. Once the retinaculum and fascia have been reapproximated, the wound is closed with 2-0 absorbable subcutaneous sutures followed by interrupted vertical mattress nylon sutures in the skin.

Anteromedial Approach to the Distal Tibia

Preoperative Planning

The patient is positioned supine on a radiolucent operating room table and a thigh high pneumatic tourniquet is used. A foam block or set of folded towels is put

under the operative leg to elevate it higher than the nonoperative leg for lateral intraoperative fluoroscopic imaging. The foot should rest in a slightly externally rotated position; therefore; placement of bumps underneath the ipsilateral hip is unnecessary. Preoperative templating to select the appropriate implant and length prevents miscalculation and unnecessary delays in the operating room.

Technique

The skin incision parallels the tibialis anterior tendons and extends proximal enough to allow application of an appropriate length of plate. Distally, the incision curves slightly medially over the ankle joint towards the talonavicular joint. The extensor retinaculum over the tibialis anterior is incised in line with the incision, and the tibialis anterior is retracted laterally. Dissection should continue beneath the tibialis anterior on the anterior aspect of the tibia, elevating the hallucis longus and extensor digitorum tendons which are retracted laterally with the tibialis anterior. The dissection may also be carried medially towards the medial malleolus to expose fractures that extend into this region. Following open reduction and internal fixation, repair the extensor retinaculum over the tibialis anterior utilizing an absorbable size 0 suture. The subcutaneous tissue is closed using an 2-0 resorbable suture, and the skin is then closed using an interrupted vertical mattress nylon suture.

Postoperative Management

Following sterile dressing application, a bulky Robert Jones type dressing utilizing both a posterior splint, as well as a stirrup-type splint, provide added stability. The patient is hospitalized until adequate pain control has been achieved. Careful postoperative evaluation should be performed to monitor for development of possible compartment syndromes in the leg following internal fixation of pilon fractures. The patient is maintained in the initial postoperative dressing for the first 7 to 10 days, which can then be exchanged for a short leg cast for further protection of both the bony and soft tissue elements. Sutures may be removed approximately 2 to 3 weeks postoperatively once the wound has healed. Depending on the bone quality and the quality of fixation, the patient is transitioned into a fracture boot or cast when rigid fixation is a concern. Once the soft tissues have

healed, the patient should be encouraged to remove the walking boot multiple times per day to begin range of motion of the ankle. Weight bearing is not permitted until radiographic evidence of healing is seen, usually at 12 weeks from surgery. Initial weight bearing begins in a walking boot and once the patient is comfortable in the boot, they wean into a regular shoe over a several week period. After progression of weight bearing, radiographs should be obtained in 3 to 4 weeks to assure there is no displacement of the fractures.

Complications

Delayed Wound Healing/Wound Dehiscence

Soft tissue complications can be minimized when an appropriate protocol of external fixation followed by delayed internal fixation is respected. In the event that delayed wound healing with dehiscence occurs, sutures should be left in place if it is a superficial skin dehiscence. Often, these can be treated often with dressing changes and occasionally oral antibiotics if there is surrounding cellulitis. In the event that there is deep or full thickness necrosis of either the posterolateral or anterolateral approaches, careful clinical inspection should be performed to determine whether the infection communicates with the deep hardware. Due to the relatively subcutaneous nature of the anteromedial approach, full thickness break-down should be treated as a deep infection with formal irrigation, debridement, and antibiotic treatment. Full thickness dehiscence of the anterolateral or posterolateral exposures should be inspected to determine if there is communication with the deep hardware. In the event that the subcutaneous and retinacular layers appear intact, appropriate treatment methods include local debridement and dressing changes. Any infection that communicates deep with the hardware needs to be treated aggressively with formal surgical debridement and postoperative dressing changes to obtain closure. Alternatively, VAC devices may be used in open wounds to help gain granulation tissue prior to an impending soft tissue coverage operation. With any type of wound dehiscence, range of motion exercises should be halted until resolution of the wound complication. While a cast boot is convenient for inspection of the wounds, a formal cast application with window cut-outs is more effective at immobilizing the limb and allows resolution of the soft tissue complications. In the event of early deep infections, repeat serial

irrigation and debridements are necessary with intravenous antibiotic treatment to obtain suppression of the infection until union has occurred. If the infection persists following fracture union, then the hardware may be removed and further soft tissue coverage provided as necessary.

Gracilis Free Muscle Flap for Extensive Tissue Loss Following Ankle Injuries

Despite the best attempts at atraumatic tissue handling during the surgical exposure soft tissue defects surrounding the ankle may still occur. Whether due to initial soft tissue loss, or due to tissue devitalization following fixation, soft tissue deficits surrounding the ankle often require urgent coverage to avoid the complications of hardware infection, nonunion, and osteomyelitis.

For soft tissue closure of ankle defects there are many options, including the reversed soleus flap, the sural artery flap, and free latissimus dorsi muscle flap. Our preference however, in these situations, is the use of the gracilis free muscle flap. This flap provides a great deal of versatility, contours well to the ankle, and produces minimal donor site morbidity.

Indications/Contraindications

Any defect surrounding the ankle can be considered for flap coverage. The flap is limited in terms of surface area, and is not well suited for circumferential coverage of the ankle and Achilles tendon

region. Also the pedicle length can be limited if a wide zone of injury requires the use of recipient vessels higher than the midcalf.

Patient factors which may increase the risk of flap failure include atherosclerosis, smoking status, and renal failure but these are not considered absolute contraindications to surgery. Absence of adequate recipient vessels is the only absolute contraindication for flap transfer.

Preoperative Planning

Often following significant lower extremity trauma a CT angio-gram is obtained to evaluate the arterial inflow to the ankle. In severe ankle trauma the posterior or

anterior tibial arteries may be occluded or transected, limiting recipient vessel options: either vessel may be used as a recipient vessel for successful flap transfer. If an angiogram is not to be performed prior to surgery a careful assessment of the posterior and anterior tibial vessels must be performed with the use of a hand held Doppler probe.

Surgery

Patient Positioning

The patient is placed supine on the operating room table with the leg in mild abduction and the knee slightly bent. Depending on the extent of the concomitant trauma either the ipsilateral or contralateral gracilis may be used. If possible we prefer to harvest the gracilis from the same leg as the original ankle trauma.

Technique

The gracilis receives its blood supply from the medial femoral circumflex artery which originates from the profunda femoral artery. The major pedicle can be identified approximately 8 to 10 cm inferior to the pubic tubercle. The flap also has a minor arterial pedicle which enters the muscle at the level of the mid thigh. This artery originates from the superficial femoral artery. The muscle receives its nervous innervation from the anterior branch of the obturator nerve. This branch of the obturator nerve can be harvested with the muscle if there are requirements for functional muscle transfer (Fig. 32-16).

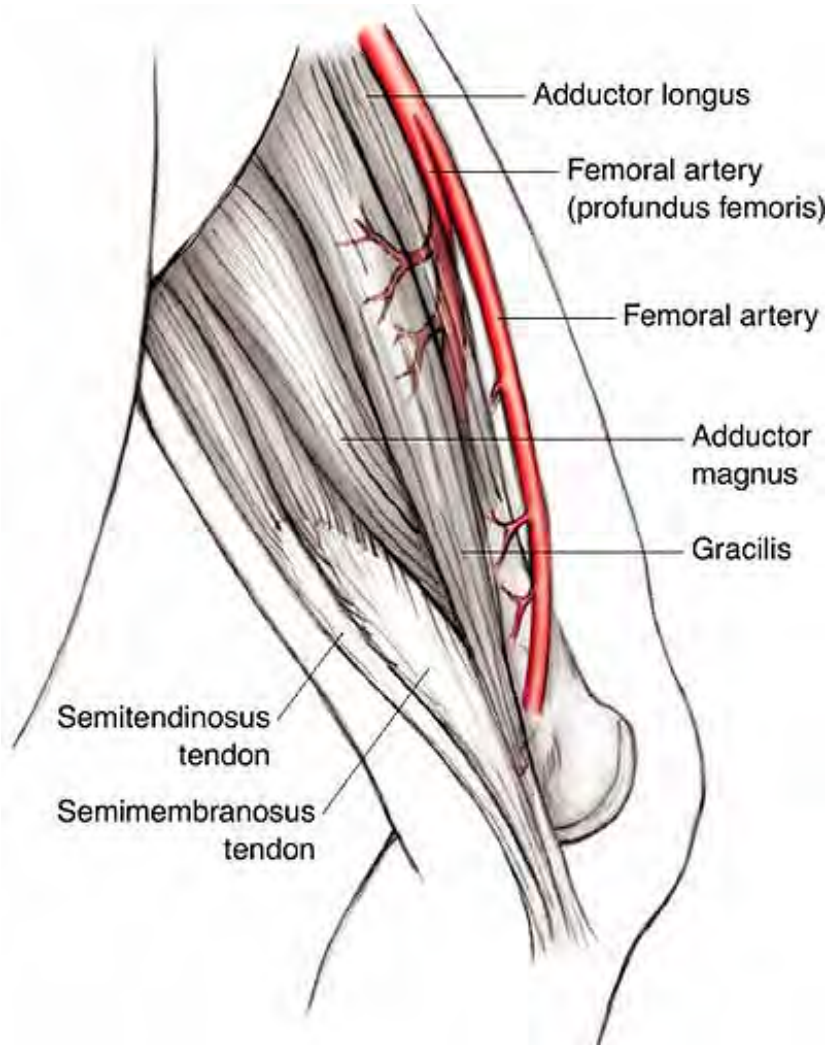

FIGURE 32-16 The gracilis muscle lies in the medial thigh. Its major pedicle is a branch of the medial femoral cutaneous artery which is a branch of the profunda femoris artery. The major pedicle can be identified entering the muscle 8 to 10 cm inferior to the pubic tubercle.

The muscle is exposed through a medial thigh incision. The incision is made 3 cm posterior to a line connecting the pubic tubercle and the medial condyle of the femur. A single long incision provides the most expedient means of identification of the gracilis muscles; however the resultant scar can be unsightly. Alternatively the

muscle may be harvested through two separate smaller incisions; a 10 to 12 cm incision may be created over the area of the vascular pedicle and a second 5 cm incision can be created at the level of the medial femoral condyle which allows for division of the muscle insertion (Fig. 32-17).

Proximally the gracilis muscle lies between the adductor longus medially and the semitendinous muscle inferiorly. It lies superficial to the adductor magnus. The pedicle lies deep to the adductor longus but runs along the superficial surface of the adductor magnus.

Once the skin incision has been made the muscle is carefully identified inferiorly at the level of the knee. Inferiorly the muscle lies posterior to the sartorius muscle and anterior to the insertion of the semimembranous and semitendinous muscles (Fig. 32-18). The gracilis can be confused with the sartorius muscle. It may be differentiated from the sartorius and semimembranous tendon by looking for the musculotendinous portion of the gracilis. At the level of the medial femoral condyle the gracilis consists of both muscle and terminal tendon, the semimembranosus is entirely composed of tendon and the sartorius is entirely muscle. Once the gracilis is identified it may be rapidly separated from the surrounding tissue with blunt dissection to the level of the minor pedicle. The minor pedicle is not divided until the major pedicle is clearly visualized (Fig. 32-19).

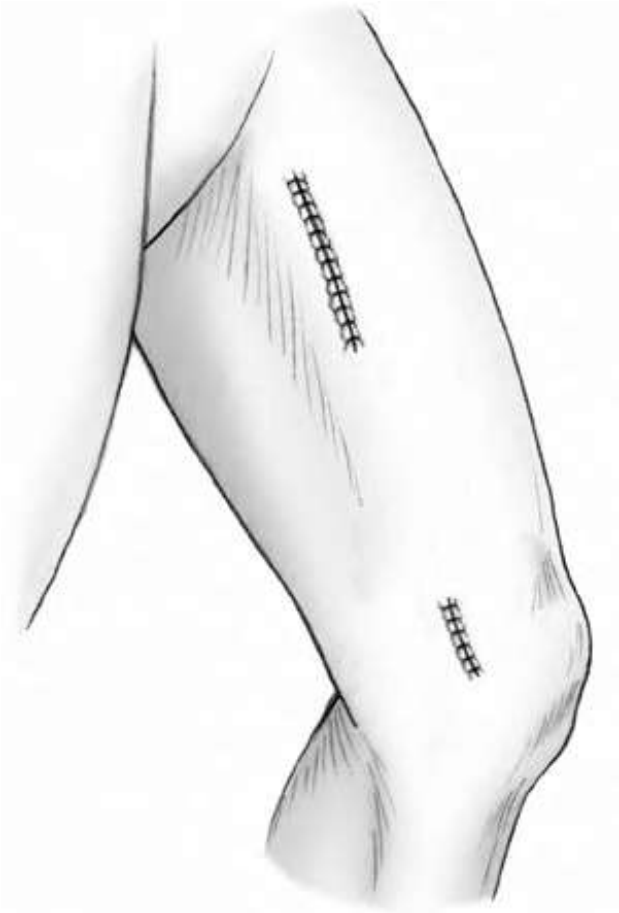

FIGURE 32-17 The muscle may be harvested through a long medial incision, or alternatively through two smaller incisions; a 10 to 12 cm incision may be created over the area of the vascular pedicle and a second 5 cm incision can be created at the level of the medial femoral condyle which allows for division of the muscle insertion.

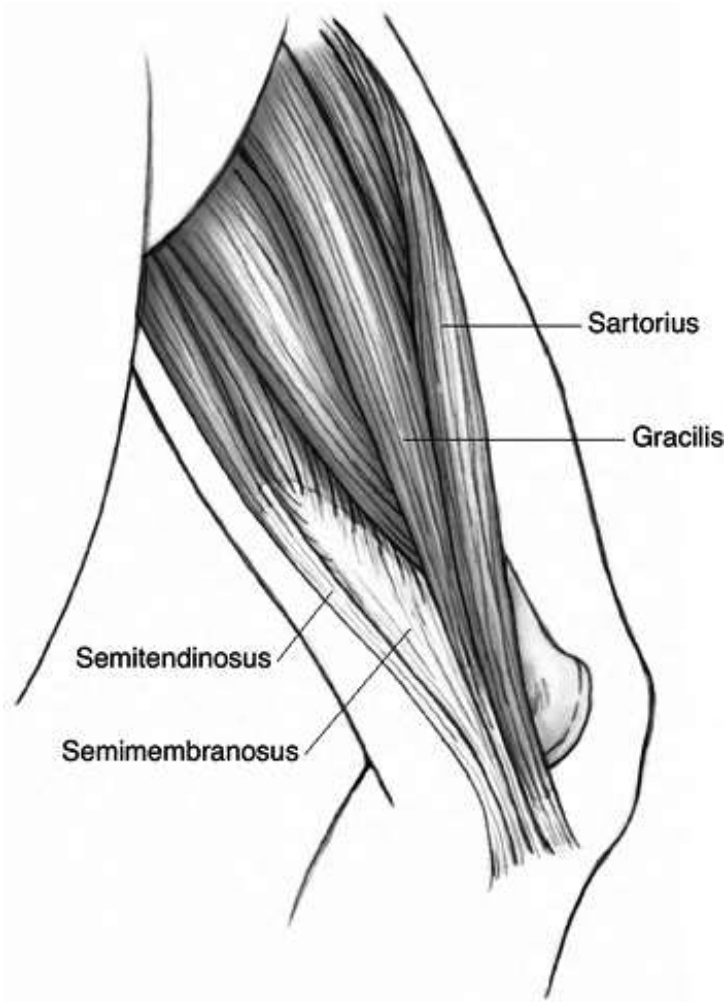

FIGURE 32-18 Inferiorly the muscle lies posterior to the sartorius muscle and anterior to the insertion of the semimembranous and semitendinous muscles. The gracilis can be confused with the sartorius muscle. It may be differentiated from the sartorius and semimembranous tendon by looking for the musculotendinous portion of the gracilis. At the level of the medial femoral condyle the gracilis consists of both muscle and terminal tendon, the semimembranosus is entirely composed of tendon, and the sartorius is entirely muscle.

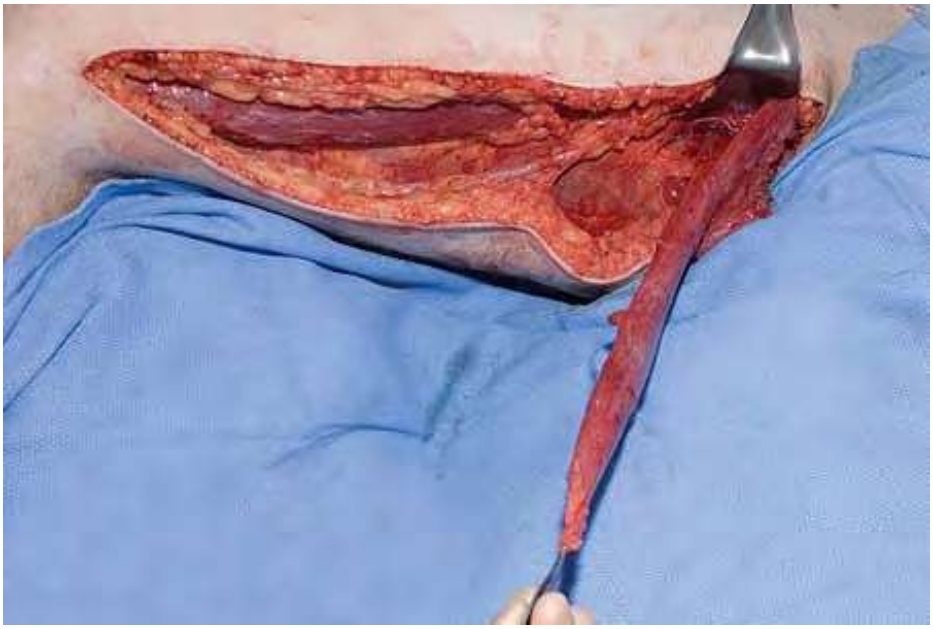

FIGURE 32-19 The muscle may be rapidly elevated back to the major pedicle. The pedicle can be seen inferior to the soft tissue retractor. The saphenous vein has been preserved as it crosses the incision midway down the thigh.

The major pedicle is identified running at the proximal lateral margin of the muscle approximately 8 to 10 cm from the pubic tubercle. The adductor longus is retracted laterally to expose the major pedicle. The fascia of the adductor magnus must be divided to allow for mobilization of the pedicle significantly. The pedicle is traced back to its origin at the profunda femoris vessels. Multiple perforating branches to the adductor muscles must be divided to gain exposure to the profundus artery, thus maximizing pedicle length of up to 6 cm. Dissection and visualization is aided with the use of lighted retractors; alternatively the adductor longus may be mobilized and retracted medially allowing visualization of the pedicle origin deep to the muscle (Fig. 32-20).

Once the major pedicle is identified and determined to be adequate for microvascular anastomosis the secondary pedicle is divided. The origin of the muscle and branch of the obturator nerve is now divided. The muscle is left to perfuse on its major pedicle until the recipient vessels have been prepared for

microvascular anastomosis.

Prior to muscle transfer a final definitive debridement of the defect site is performed. The recipient vessels are then exposed away from the original zone of injury (Fig. 32-21). Donor vessel preparation

P.414

P.415

is performed with the use of high powered operative loop magnification or the microscope to verify that the vessels are adequate for microvascular anastomosis. The anterior tibial artery is used as a recipient vessel for lateral malleolar defects while the posterior tibial artery is preferentially used for medial defects (Fig. 32-22). End to side anastomosis to the artery is preferred while an end to end anastomosis to the venous comitantes is performed in most settings. If two veins are present within the arterial pedicle both of the venous comitantes to the gracilis muscle are anastomosed to minimize the chances of postoperative venous insufficiency.

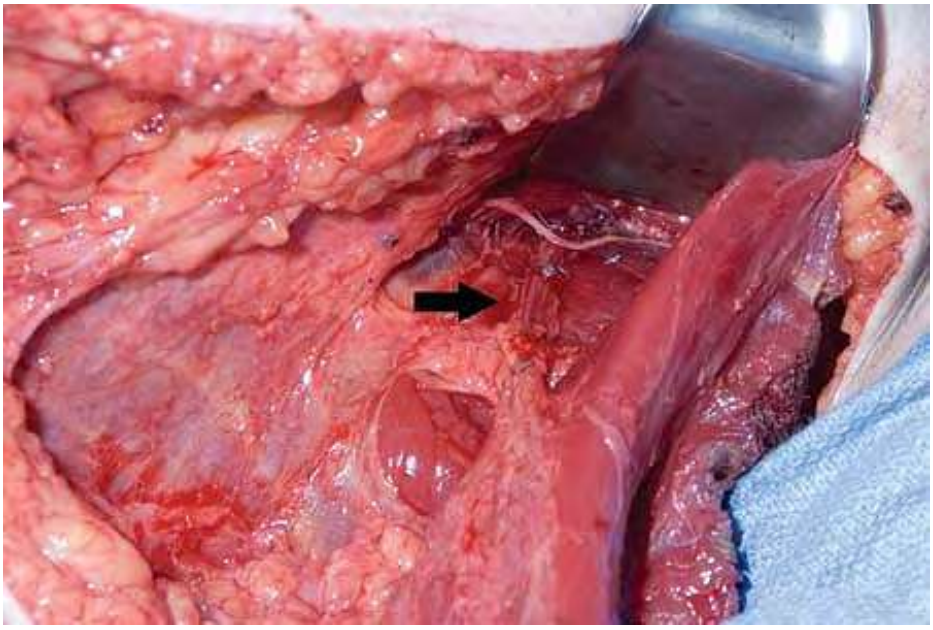

FIGURE 32-20 The pedicle (*black arrow*) is dissected back to its origin at the profunda femoris. This allows for a significant improvement in pedicle length. A branch of the obturator nerve can be seen running on top of pedicle and the

adductor magnus muscle.

FIGURE 32-21 A: A lateral malleolar defect has developed in a 48-year-old

gentleman following stabilization of a complex malleolar and calcaneal fracture. B: The anterior tibial vessels are exposed away from the zone of injury in preparation for free tissue transfer. C: The anterior vessels are identified just lateral and deep to the anterior tibialis muscle. The perineal nerve is surrounded with a vessel loop and gently freed from the underlying vessels. D: The flap at one week shows good take of the skin graft. E: Four month view shows excellent contouring of the flap with stable wound coverage allowing for normal shoe wear.

Once the anastomosis is complete the muscle is allowed to reperfuse for 20 minutes as the donor site is closed. An implantable Doppler probe is placed circumferentially around one of the two veins to allow for evaluation of anastomotic patency postoperatively. Donor site closure is performed with deep dermal sutures and an absorbable subcutaneous monofilament suture. A closed suction drain is also used to prevent postoperative seroma or hematoma formation. Flap insetting is then performed. Muscle insetting is facilitated with the use of half buried absorbable sutures placed within the muscle and pulled beneath the native skin. The epimysium of the muscle is often excised to allow for expansion of the muscle's surface area. This allows for improved contouring over the malleolus. At the

completion of insetting the muscle is covered with a meshed split thickness skin graft. The flap is covered with a Xeroform and the patient's leg is then loosely wrapped in sterile cotton and placed into a posterior splint for postoperative comfort. A large window is created in the dressing for post-operative flap monitoring. A single suture is placed superficially within the muscle over the pedicle so that anastomotic patency can be evaluated postoperatively with the use of a hand held Doppler device.

FIGURE 32-22 A: A 32-year-old gentleman presents with a defect of the lower margin of the tibia which developed following treatment of a malleolar fracture. B,C: In this case the gracilis muscle is anastomosed to the posterior tibial vessels and brought anteriorly to cover the bone and soft tissue defect. D: The muscle was harvested through two limited incisions over the medial thigh. The inferior incision is used for division of the muscle insertion while the superior incision is used for dissection of the vascular pedicle and division of the muscle origin. E,F: The wound is stable at 6 months with no further drainage or signs of ongoing infection.

Postoperative Management

Flap monitoring may be performed with either the use of an implantable Doppler probe, or with the use of a hand held Doppler probe at the bedside. Evaluation of perfusion is performed every hour for the first 24 hours by a skilled microsurgical nursing staff who are familiar knows the signs of arterial insufficiency and venous thrombosis.

On postoperative day 4 or 5 the entire dressing is removed and skin graft survival is assessed. At this point the patient may be placed into a loose Robert-Jones dressing or a removal posterior splint. The donor site drain is removed once output is less than 30 cc a day.

Complications/Results

The most devastating complication following any free tissue transfer include partial flap loss and total flap loss. If arterial insufficiency or venous insufficiency occurs postoperatively, immediate re-exploration is recommended. At the time of surgical exploration thrombosis may be identified at the anastomotic site. In such cases early revision of the anastomosis will often allow for flap salvage. If the flap is unsalvageable and flap loss occurs, an additional flap will be required. In such cases a secondary free flap, pedicled flap, or cross-leg flap can be considered. In cases of partial flap loss, occasionally the flap may be advanced distally to cover critical structures. Alternatively small areas of distal necrosis may be managed with dressing changes, skin grafting, or pedicle flap coverage.

In an examination of 50 acute traumatic and post-traumatic wounds, Redett and colleagues found the gracilis muscle effective for covering defects up to 165 cm^2. Successful free tissue transfer was performed in 93% of patients, and limb salvage was possible in 96%. Studies by Hallock have shown that the vascular pedicle is of adequate length for reconstruction of defects extending over the calcaneus and lateral malleolus (Fig. 32-21). For defects which exceed the limits of a gracilis transfer a latissimus free muscle flap can be considered. We have found the gracilis to contour nicely over time allowing for normal shoe wear following reconstruction.

Recommended Reading

Blauth M, Bastian L, Krettek C, et al. Surgical options for the treatment of

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 33

Calcaneal Fractures/Talar Neck Fractures

Michael P. Clare

Roy W. Sanders

Fractures of the calcaneus and talus are among the most complex and challenging of fractures for the orthopaedic surgeon to effectively manage, in that both are three-dimensionally unique structures with highly specialized biomechanical function and a distinctly limited surrounding soft tissue envelope. Fractures of the calcaneus and talus are generally the result of high-energy trauma, such as a motor vehicle accident or a fall from a height, and as such, the severity of fracture displacement and the extent of soft tissue disruption are directly proportional to the amount of force and energy absorbed by the limb.

Fractures of the Calcaneus

Indications/Contraindications

General Considerations

Operative treatment is generally indicated for displaced intra-articular fractures involving the posterior facet as demonstrated on computed tomography (CT) scanning. Nonoperative treatment is best reserved for non- or minimally-displaced extra-articular calcaneal fractures, and truly non-displaced intra-articular fractures as determined on CT scan.

Comorbidities

While nicotine use/dependence is not a contraindication to operative treatment, all patients who are smokers are counseled at length as to the associated risks and encouraged to discontinue tobacco use. We consider heavy smoking (â‰¥2 packs per day) as a relative contraindication to surgery. Specific contraindications for operative treatment include fractures in patients with insulin-dependent diabetes mellitus, severe peripheral vascular disease, or other major medical comorbidities precluding surgery, as well as fractures in elderly patients who are minimal (household) ambulators. Because many older patients are healthy and active well into their 70s, chronologic age itself is not necessarily a contraindication to surgical treatment.

Delayed Treatment by Necessity

Operative treatment may also be contraindicated if initial evaluation has been delayed beyond 3 or 4 weeks from the date of injury, or in certain situations in which injury severity prohibits early surgical intervention, including fractures associated with severe fracture blisters or prolonged edema; fractures with large open wounds; and fractures in patients with life-threatening injuries. Operative treatment following a prolonged delay (>4 weeks) from initial injury to definitive treatment is complicated by the fact that early consolidation of the fracture has occurred, making the fracture fragments increasingly difficult to separate to obtain an adequate reduction, and the articular cartilage may delaminate away from the underlying subchondral bone. In these instances, delayed treatment by necessity is used whereby the fracture is allowed to heal and is later managed as a calcaneal malunion following resolution of the prohibitive factors.

FIGURE 33-1 Preoperative â€œwrinkleâ€• test. Note presence of skin creases indicating that surgery may now be safely undertaken.

Preoperative Planning

In completing preparations for surgery, the surgeon should thoroughly review the plain radiographs and CT scans to gain a preliminary understanding as to the â€œpersonalityâ€• of the fracture pattern, which then allows for planning of patient position and surgical approach, as well as anticipation of specific technical steps in obtaining fracture reduction, and the necessary implants for definitive stabilization. While multiple surgical approaches have historically been described, the general consensus currently favors the extensile lateral approach for most displaced intra-articular calcaneal fractures. Percutaneous reduction techniques, such as the Essex-Lopresti maneuver, are particularly ideal for certain tongue-type fracture (Sanders type II-C) patterns.

Surgery is ideally completed within the first 3 weeks of injury before early fracture

consolidation; however, surgery should not be attempted until the associated soft tissue swelling has sufficiently dissipated as demonstrated by a positive wrinkle test. The test involves assessment of the lateral calcaneal skin with passive dorsiflexion and eversion of the injured foot. A positive test is confirmed by the presence of skin wrinkling without residual pitting edema, such that surgical intervention may be safely undertaken (Fig. 33-1). A variety of modalities may used to decrease swelling in the affected extremity. We prefer initial elevation and immobilization in a bulky Jones dressing and supportive splint, with later conversion to an elastic compression stocking and fracture boot locked in neutral flexion as the acute swelling begins to dissipate in the ensuing few days.

Extensile Lateral Approach

Surgery

Patient Positioning/General Considerations

The procedure requires use of a radiolucent table and a standard C-arm. For isolated injuries, we prefer placing the patient in the lateral decubitus position on a beanbag. The lower extremities are positioned in a scissor configuration, whereby the operative (â€œupâ€•) limb is flexed at the knee and angles toward the posterior, distal corner of the operating table, and the nonoperative (â€œdownâ€•) limb is extended at the knee and positioned away from the surgical field to facilitate intraoperative fluoroscopy. Protective padding is placed beneath the contralateral limb for protection of the peroneal nerve, and an operating â€œplatformâ€• is created with blankets and foam padding to elevate the operative limb (Fig. 33-2). The prone position may alternatively be used in the event of bilateral injuries.

The patient is given prophylactic preoperative antibiotics, and a pneumatic thigh tourniquet is used. The procedure should be completed within 120 to 130 minutes of tourniquet time, so as to minimize wound complications. If the procedure extends beyond that time, the tourniquet should be released and the remainder of the procedure performed without it. In allocating tourniquet time, the surgical approach should be completed within 20 minutes, allowing up to 60 minutes for fracture reduction, 20 minutes for implant placement, and 20 minutes for wound closure.

FIGURE 33-2 Lateral decubitus position. Note scissor-like limb configuration to facilitate intraoperative fluoroscopy.

Technique

Anatomic Landmarks

Wound complications following surgical management of calcaneal fractures remain a major source of morbidity with these injuries. The soft tissues overlying the lateral hindfoot receive blood supply from a confluence of three arterial branches: the lateral calcaneal artery, the lateral malleolar artery, and the lateral tarsal artery. Borrelli and Lashgari determined that the majority of the full-thickness flap with an extensile lateral approach is supplied by the lateral calcaneal artery, typically a branch of the peroneal artery. At the level of lateral malleolus, the artery courses parallel to the Achilles tendon and lies approximately 11 to 15 mm anterior to the terminal Achilles tendon and its insertion. Thus, strict attention to detail with respect to placement of

the incision and gentle handling of the soft tissues is of paramount importance.

Incision and Surgical Approach

The extensile lateral incision is then outlined on the skin with a marking pencil (Fig. 33-3). The incision begins approximately 2 cm proximal to the tip of the lateral malleolus, just lateral to the Achilles tendon extending toward the plantar foot. In this manner, the vertical limb of the incision will course posterior to the sural nerve and the lateral calcaneal

artery, thereby avoiding devascularization of the lateral calcaneal flap. The horizontal limb of the incision is drawn along the junction of the skin of the lateral foot and heel pad, the demarcation of which can be identified by compressing the heel. We prefer to substitute a gentle curve where these two lines intersect to form a right angle, primarily to avoid necrosis of the apical skin. The terminal portion of the horizontal limb includes a gentle curve anteriorly along the skin creases, extending over the calcaneocuboid articulation.

FIGURE 33-3 Planned extensile lateral incision.

FIGURE 33-4 Full-thickness subperiosteal flap. Note peroneal tendons contained within the flap.

With a sterile bolster placed beneath the medial ankle, the incision begins at the proximal portion of the vertical limb, becoming full-thickness at the level of the calcaneal tuberosity; dissection is specifically taken from â€œskin to boneâ€• at this level while avoiding any beveling of the skin. Scalpel pressure is again lessened beyond the apical curve, roughly at the midpoint of the horizontal limb, at which point a layered incision is again developed.

A full-thickness subperiosteal flap is then developed starting at the apex. Any use of retractors should be avoided until a sizeable flap has been raised, so as to avoid

separation of the skin from the subcutaneous tissues and periosteum. The calcaneofibular ligament and inferior peroneal retinaculum are released sharply, thus exposing the peroneal tendons. Both tendons are identified and released from the peroneal tubercle through the cartilaginous pulley, and further gently mobilized distally with a periosteal elevator, thereby exposing the anterolateral calcaneus and calcaneocuboid joint. In this manner, the peroneal tendons and sural nerve are contained entirely within the sub-periosteal flap (Fig. 33-4).

Deep dissection continues anteriorly to the sinus tarsi and anterior process, and posteriorly to the superior-most portion of the calcaneal tuberosity for â€œwindow visualizationâ€• of the posterior facet, so as to prevent rotational malalignment of the posterior facet articular surface in the sagittal plane. Three 1.6-mm Kirschner wires are then placed for retraction of the subperiosteal flap using the â€œno touchâ€• technique. In this technique, one wire is placed into the distal fibula as the peroneal tendons are slightly subluxed anteriorly; a second wire is placed in the talar neck; a third wire is placed in the cuboid as the peroneal tendons are levered away from the anterolateral surface of the calcaneus with a periosteal elevator. Thus, each Kirschner wire retracts its respective portion of the peroneal tendons and subperiosteal flap (Fig. 33-5). A small Bennett-type retractor may additionally be used at the distal margin of the sinus tarsi for further exposure of the anterolateral calcaneus.

Assessment of the Peroneal Tendons

Following fracture reduction, definitive stabilization, and final fluoroscopic images, the wound is copiously irrigated. As the previously placed Kirschner wires are manually removed, the peroneal tendons should easily reduce into the peroneal groove at the posterior border of the lateral malleolus. A Freer elevator is introduced into the tendon sheath, advanced proximally to the level of the lateral malleolus, and levered anteriorly to assess the stability of the peroneal tendon sheath and superior peroneal retinaculum. If the tendon sheath is detached from the lateral malleolus and therefore incompetent, the elevator will easily advance anterior to the lateral malleolus, indicating that a retinacular repair is required. A 3-cm incision is then made along the posterior margin of the lateral malleolus, exposing the tendon sheath and retinaculum. With the peroneal tendons held reduced in the peroneal groove, one to two suture anchors are

placed in the lateral malleolus to secure the detached tendon sheath and retinaculum. Tendon stability is then reassessed with a Freer elevator in the same manner.

FIGURE 33-5 â€œNo touchâ€• technique: K-wire placement for retraction of the full-thickness flap.

Closure Technique

A deep drain is placed exiting proximally in line with the vertical limb of the incision. Deep No. 0 absorbable sutures are then passed in interrupted, inverted fashion starting with the apex of the incision. Sutures are placed thereafter at the proximal and distal ends of the incision, and progressing toward the apex of the incision, while attempting to advance the flap toward the apex. The suture ends are temporarily clamped until all sutures have been passed (Fig. 33-6). The sutures are then hand tied in sequential fashion, starting at the ends proximally and distally, progressing toward

the apex of the incision to minimize tension at the apex.

AllgÃ¶wer-Donati Suture

The skin layer is closed with 3-0 monofilament suture using the modified AllgÃ¶er-Donati technique, again starting at the ends and working toward the apex (Fig. 33-7). The suture technique is a modified vertical mattress stitch, whereby the far end passes subcutaneous to the skin edge to minimize tension on the skin margin. Alternatively, the sutures may be passed in modified horizontal mattress fashion. In the context of an extensile lateral approach,

the knots are placed along the periphery of the incision, avoiding violation of the skin margin of the subperiosteal flap.

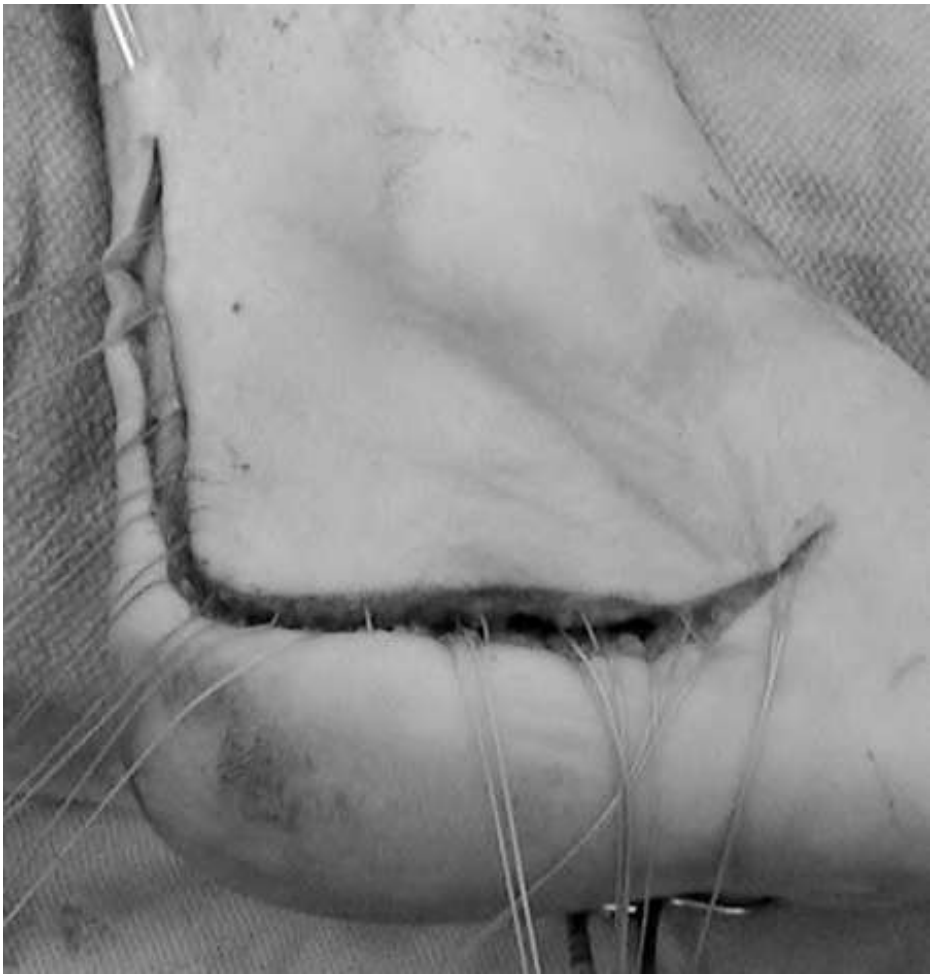

FIGURE 33-6 Deep closure. Note placement of all sutures prior to sequential tying (different patient).

FIGURE 33-7 Skin closure using modified AllgÃ¶er-Donati technique.

The suture is placed initially into the wound edge of the near end, and the needle is then reversed in the needle driver. The suture is introduced into the subcuticular edge of the far end and extended through the subcutaneous tissue of the near end, exiting approximately 1 cm from, and in line with, the initial entry point. A single reverse throw is passed to maintain the skin bridge between the entry and exit point, and the knot is slightly tensioned to approximate the skin edges. A second reverse knot is

passed to allow the knot to slide and gently evert the wound edges, followed by a forward knot to lock the stitch in place (Fig. 33-8). Following completion of wound closure, the tourniquet is deflated and sterile dressings are applied, followed by a Jones-type bulky cotton dressing and Weber splint.

Postoperative Management

The patient remains in the hospital overnight for pain control, and is converted to a short-leg non-weight bearing cast prior to discharge. The patient is converted back into an elastic compression stocking and fracture boot at 2 to 3 weeks postoperatively, and subtalar range of motion exercises are initiated. The sutures are removed once the incision is fully sealed and completely dry, typically at 3 to 4 weeks; however, they should not be removed until the wound is fully healed. We prefer that the patient sleep in the fracture boot at night until weight bearing is begun to prevent an equines contracture. Weight bearing is not permitted until 10 to 12 weeks postoperatively, at which point the fracture should be radiographically healed. The patient is then gradually transitioned to a regular shoe, and activities are advanced. In our experience, the patient should be able to return to a moderately active job at approximately 4 to 6 months postoperatively.

Complications

Delayed Wound Healing/Wound Dehiscence

Delayed wound healing or wound dehiscence is the most common complication following surgical management of a calcaneal fracture, and may occur in up to 25% of cases. Risk factors for wound complications include smoking, diabetes mellitus, open fractures, high body mass index, and a single layered wound closure. While the extensile lateral incision approximates relatively easily at the time of initial closure, wound separation may later occurâ€”typically at the apex of the incision, and even up to 4 weeks postoperatively (Fig. 33-9). The vast majority of wounds, however, will ultimately heal; deep infection and osteomyelitis develop in only 1% to 4% of closed fractures.

In the event of a wound dehiscence, all range of motion activities are discontinued to prevent further dehiscence of the wound. We prefer a fairly aggressive approach to the wound, with a management

regimen of daily whirlpool treatments, damp-to-dry dressing changes, and oral antibiotics. Other granulation-promoting wound agents may also be beneficial in this instance. Alternatively, the limb may be immobilized in a short leg cast, with a window overlying the wound for access for dressing changes. These treatment regimens will typically prove successful, so long as the wound necrosis is limited to partial thickness of the skin layer. Range of motion exercises are re-initiated once the wound seals and remains dry with the patient off antibiotics.

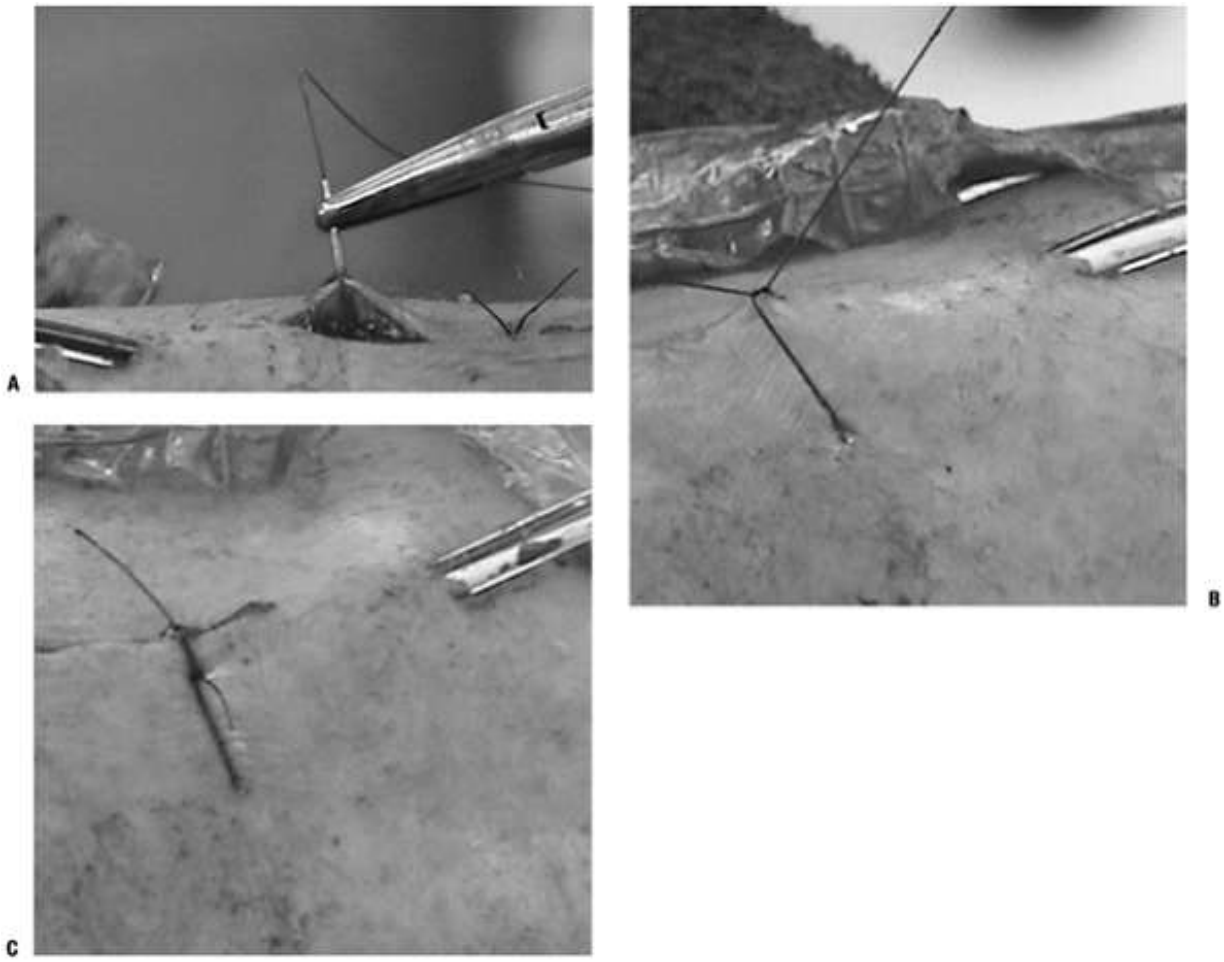

FIGURE 33-8 AllgÃ¶wer-Donati technique. Note reverse throw (B) to maintain skin bridge between suture ends.

FIGURE 33-9 Apical wound dehiscence at 4 weeks postoperatively.

We prefer use of a negative-pressure device (Vacuum-Assisted ClosureÂ®, KCI, Inc., San Antonio, TX) to promote healing in the event of a recalcitrant wound; if all other treatment methods fail, a low-profile fasciocutaneous flap such as a lateral arm flap may be required for wound coverage.

Peroneal Tendon Adhesions

Peroneal tendon adhesions and scarring may develop in up to 18% of cases, either from the extensile lateral exposure itself or from prominent screwheads adjacent to the tendons, particularly surrounding the anterior process of the calcaneus. Nonoperative treatment includes tendon massage, stretching, strengthening, and other local modalities. Peroneal tenolysis or removal of the symptomatic hardware may be required in refractory cases.

Cutaneous Nerve Injury

The most common neurologic complication with surgical treatment of calcaneal fractures is iatrogenic injury to a sensory cutaneous nerve, particularly the sural nerve. Sural nerve injury occurs in up to 15% of cases and ranges from a stretch neuropraxia, which can be transient or permanent, to complete laceration of the nerve. Clinically, the patient may experience decreased or complete loss of sensation in the lateral hindfoot, or perhaps even a painful neuroma. Initial treatment includes gabapentin or amitriptyline, shoe modifications or soft accommodative inserts, and physical therapy modalities. In the event of a painful neuroma refractory to these measures, surgical neurolysis and resection may be considered, including burial of the proximal stump into deep tissues.

Fractures of the Talar Neck

Indications/Contraindications

Operative treatment is generally indicated for all displaced fractures of the talar neck, and these fractures have traditionally been considered a surgical emergency, due to the high incidence of osteonecrosis associated with displaced fractures. More recent studies have indicated that a delay in surgical management beyond 6 to 8 hours from the time of injury does not necessarily increase the risk of osteonecrosis, such that definitive stabilization within the first 24 hours following injury is now gaining acceptance as standard of care.

Severely displaced fractures or fracture-dislocations, however, may produce sufficient tension and pressure on the surrounding skin acutely to impair local circulation and result in skin necrosis and slough, potentially leading to a catastrophic deep infection and/or osteomyelitis. Thus, a timely attempt at closed reduction of the involved fracture fragment or dislocation is of paramount importance in minimizing soft tissue

complications associated with these injuries, particularly if a delay in definitive stabilization is anticipated.

Nonoperative management is reserved for truly non-displaced fractures as confirmed on CT scan. Because of the high-energy nature of these injuries, displaced talar neck fractures occur most commonly in young adults. Thus in the majority of cases, the primary contraindication to surgery is the presence of severe life-threatening injuries where the patient is too medically unstable to tolerate surgery. In this instance, however, an attempt at closed reduction or manipulation of any dislocated fragments should be made in the emergency room to minimize the risk of skin necrosis.

Preoperative Planning

The surgeon should thoroughly review the plain radiographs and CT scans to gain an initial understanding of the fracture pattern, which then allows for appropriate preoperative planning, including: the surgical approach, anticipation of specific technical steps in obtaining fracture reduction, and the necessary implants for definitive stabilization. While we commonly use up to eight different surgical approaches in the management of talar fractures in general, we prefer dual anteromedial and anterolateral approaches for talar neck fractures.

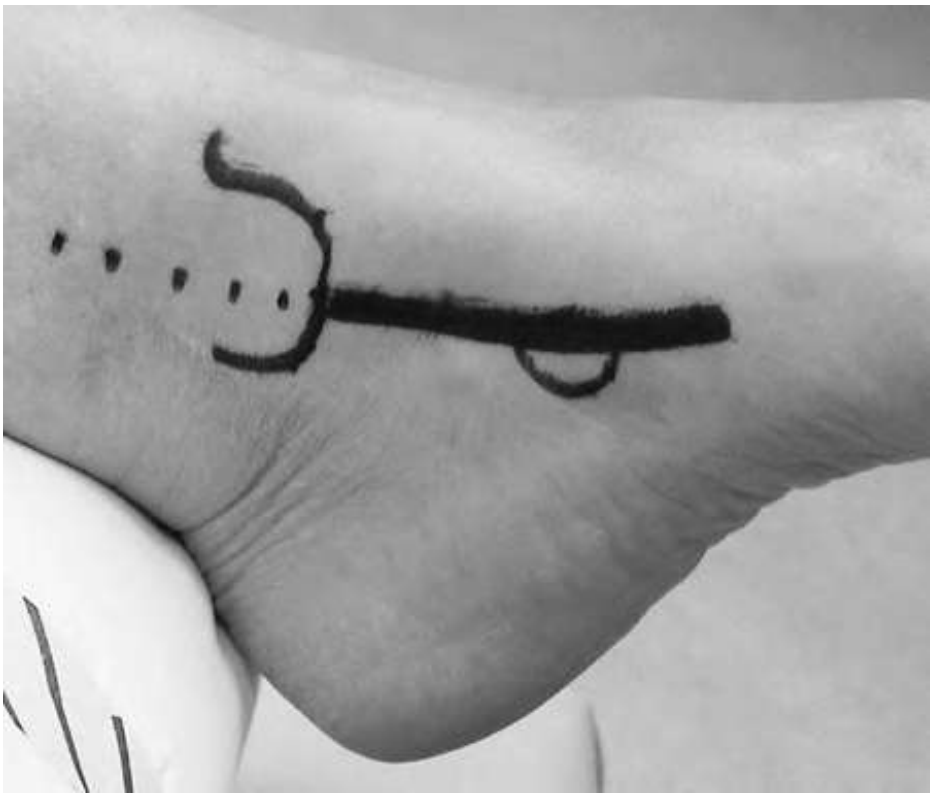

FIGURE 33-10 Planned anteromedial incision.

Surgery

Patient Positioning/General Considerations

Dual Anteromedial and Anterolateral Approaches

The procedure requires use of a radiolucent table and a standard C-arm. For isolated injuries, we prefer placing a non-sterile bolster beneath the ipsilateral hip and pelvis to allow sufficient limb exposure both medially and laterally. Protective padding is placed beneath the contralateral limb for protection of the peroneal nerve, and the contralateral limb is secured to the operating table to allow for axial plane rotation as needed for surgical exposure. The patient is given prophylactic preoperative antibiotics, and a pneumatic thigh tourniquet is employed. The procedure should be completed within 120 to 130 minutes of tourniquet time to minimize wound

complications.

Technique

Incision and Surgical Approach

The dual anteromedial and anterolateral incisions are outlined on the skin with a marking pencil. Medially, the talar neck is isolated approximately a thumb-breadth distal to the anterior tip of the medial malleolus, or midway between the medial malleolus and navicular tubercle. Thus, the anteromedial incision extends from the tip of the medial malleolus in line with the medial column of the foot to a point approximately 1 cm beyond the navicular tubercle (Fig. 33-10). In this manner, the deep dissection will course through the â€œsoft spotâ€• between the anterior and posterior tibial tendons, and posterior to the saphenous nerve and vein. The approach is potentially extensile, as it can be extended proximally to allow for a medial malleolar osteotomy for talar neck fractures extending into the talar body or posteromedial process, as well as distally for access to the entire medial column of the foot as necessary (Fig. 33-11).

FIGURE 33-11 Medial malleolar osteotomy: anteromedial approach extended proximally for exposure of talar body (different patient).

FIGURE 33-12 Planned anterolateral incision.

Laterally, the talar neck is found immediately dorsal to the sinus tarsi in line with the extensor digitorum longus and peroneus tertius tendons. The anterolateral incision actually consists of the BÃ¶hler approach, coursing from the anterolateral corner of the ankle joint in line with the extensor digitorum longus and peroneus tertius tendons toward the base of the fourth metatarsal (Fig. 33-12). This approach is also considered extensile, as it can be extended proximally to allow for exposure of the lateral talar dome, with or without a lateral malleolar osteotomy, and distally for exposure of the entire lateral column of the foot as necessary.

We prefer completing the anterolateral approach first because the majority of the comminution is typically found medially; thus, the most accurate initial indication of the extent of fracture displacement or rotational malalignment is found laterally. With sterile bolsters placed beneath the knee and ankle, the anterolateral approach is initiated, and superficial dissection continues to the extensor retinaculum and tendon sheath of the extensor digitorum longus and peroneus tertius tendons. Care is taken to avoid violation of the superficial peroneal nerve proximally, although the proximal portion of the incision rarely extends proximal enough to visualize the nerve before it begins coursing medially toward the first dorsal web space.

The tendon sheath is incised at the lateral margin of the tendons, and deep dissection is continued to the deep capsule of the ankle and subtalar joints proximally, and the extensor digitorum brevis muscle distally. The extensor brevis muscle is then traced to its origin beneath the tendons working dorsally, and subsequently reflected plantarly, thereby exposing the lateral capsule of the talonavicular joint and distal portion of the talar neck. The deep capsules of the ankle and subtalar joints are then released in line with and including the talonavicular joint capsule extending dorsally and plantarly, thereby completing a full-thickness flap (Fig. 33-13). In this manner, the foot may be adducted to expose the lateral portion of the talar head, thereby facilitating eventual placement of screws,

which may be countersunk within the talar head. We make a conscious effort to limit the extent of subperiosteal dissection into the sinus tarsi, exposing only what is necessary to obtain an anatomic reduction, to minimize further compromise of the already precarious blood supply to the talar body.

FIGURE 33-13 Anterolateral approach. Note full-thickness flap with simultaneous exposure of ankle, subtalar, and talonavicular joints.

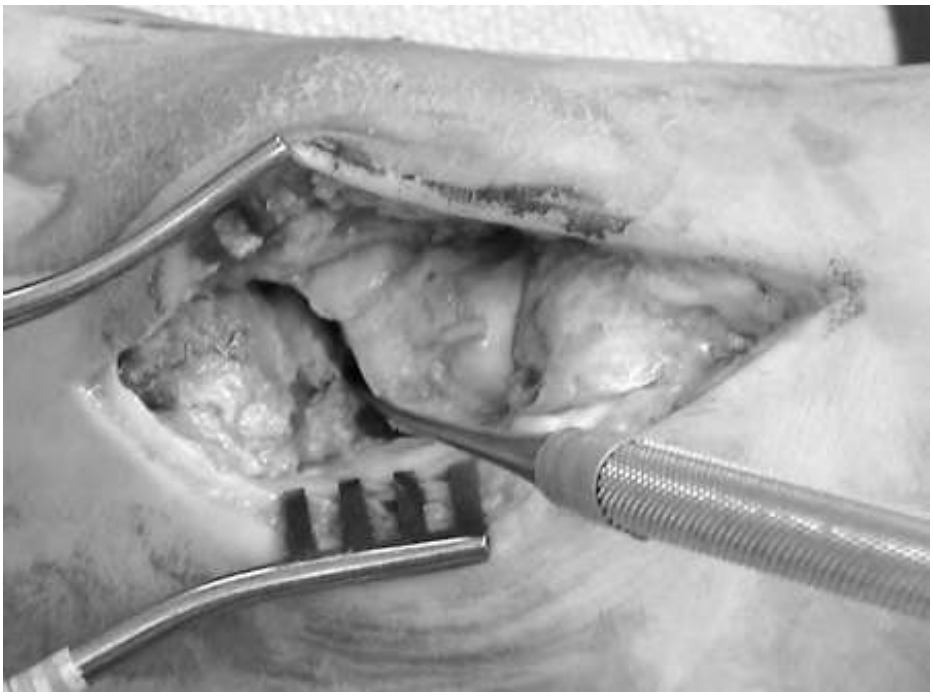

FIGURE 33-14 Anteromedial approach. Note full-thickness flap with simultaneous exposure of ankle, subtalar, and talonavicular joints.

The anteromedial approach is then initiated, and superficial dissection continues to the extensor retinaculum and deep capsules of the ankle and subtalar joints proximally, and dorsal margin of the posterior tibial tendon sheath distally. The extensor retinaculum and deep capsules are longitudinally incised, continuing along the dorsal edge of the posterior tibial tendon through the underlying talo-navicular joint capsule and spring ligament. Care is taken to avoid violation of the deltoid ligament fibers at the proximal margin of the incision.

The talonavicular joint capsule is then elevated in subperiosteal fashion off of the navicular tubercle, extending roughly to the midpoint of the navicular dorsally. The dorsal-most portion of the posterior tibial tendon insertion may also be reflected plantarly to ease soft tissue tension as necessary (Fig. 33-14). In this manner, the foot may be abducted to expose the medial portion of the talar head, again facilitating eventual screw placement. We again limit the extent of subperiosteal dissection along the undersurface of the talar neck medially in an attempt to preserve the vascular

anastamoses extending into the tarsal canal. At the completion of the anterolateral and anteromedial approaches, the fracture patterns traversing the dorsal portion of talar neck should be easily visualized.

Closure Technique

Following fracture reduction, definitive stabilization, and final fluoroscopic imaging, the wound is copiously irrigated. The deep capsular layers are closed medially and laterally with interrupted No. 0 absorbable sutures placed in figure-of-eight fashion. Laterally, the extensor digitorum brevis muscle is gently approximated distally with interrupted 2-0 absorbable sutures in similar fashion, as is the extensor retinaculum and extensor digitorum longus and peroneus tertius tendon sheath more proximally. Medially, the extensor retinaculum and posterior tibial tendon sheath are closed in identical fashion with interrupted 2-0 absorbable sutures. The subcutaneous and subcuticular layers are closed with inverted, interrupted 2-0 absorbable sutures. The tourniquet is then deflated, and the skin layers are approximated with interrupted 3-0 monofilament suture, again using the modified AllgÃ¶wer-Donati technique. Sterile dressings are applied, followed by a Jones-type bulky cotton dressing and Weber splint with the ankle in neutral dorsiflexion-plantar flexion.

Postoperative Management

The patient remains in the hospital overnight for pain control, and is converted to a short-leg non-weight bearing cast before discharge. The patient is converted back into an elastic compression stocking and fracture boot at 2 to 3 weeks postoperatively, and ankle and subtalar range of motion exercises are initiated. The sutures are removed once the incision is fully sealed and completely dry, typically at 3 to 4 weeks; however, they should not be removed until the wound is fully healed. We prefer that the patient sleep in the fracture boot at night until weight bearing is begun to prevent an equines contracture.

Weight bearing is not permitted until 10 to 12 weeks postoperatively, and the patient is gradually transitioned to a regular shoe, and activities are advanced. Postoperative radiographs are carefully

scrutinized beginning at 6 weeks following surgery for the presence of a Hawkins sign

(Fig. 33-15), suggesting revascularization of the talar body. The absence of a Hawkins sign at that time, however, does not necessarily indicate osteonecrosis, as revascularization of the talar body may not occur for up to 2 years following surgery.

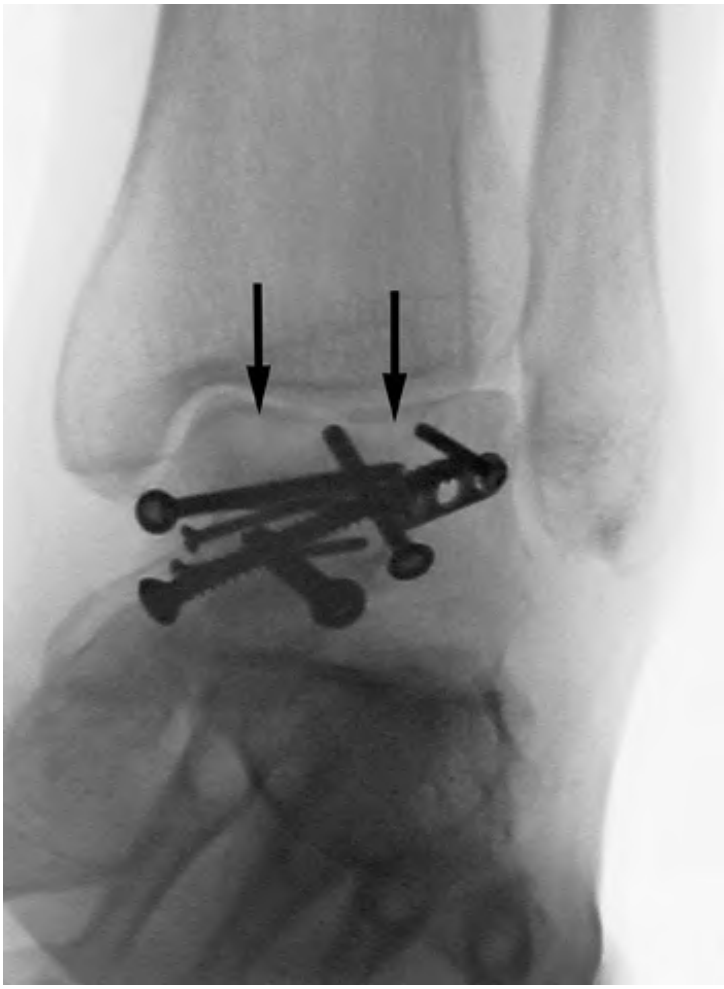

FIGURE 33-15 Hawkins sign: anteroposterior radiograph of the ankle demonstrating subchondral atrophy (arrows) at 8 weeks postoperatively following open reduction and internal fixation of a displaced talar neck fracture.

Complications

Delayed Wound Healing/Wound Dehiscence/Infection

Delayed wound healing or wound dehiscence following open reduction and internal fixation occurs in a relatively small percentage of patients, ranging from 0% to 4%. In the event of a wound dehiscence, all range of motion activities are discontinued to prevent further dehiscence of the wound. We prefer a fairly aggressive approach to the wound, with management regimen of daily whirlpool treatments, damp-to-dry dressing changes, and oral antibiotics. Other granulation-promoting wound agents, or use of a negative-pressure device may also be beneficial in this instance. Alternatively, the limb may be immobilized in a short leg cast, with a window overlying the wound for access for dressing changes. These treatment regimens typically are successful, as long as the wound necrosis is limited to partial thickness of the skin layer. Range of motion exercises are re-initiated once the wound seals and remains dry with the patient off antibiotics.

Deep infection and osteomyelitis rates are similarly low, occurring in up to 5% of closed fractures. Open injuries, however, are associated with markedly higher rates of deep infection and osteomyelitis, up to 38% in some series. Deep infection and osteomyelitis may also result from skin necrosis and slough from an unreduced closed fracture or fracture-dislocation due to excessive tension on the surrounding soft tissue envelope. Management of a deep infection includes serial surgical debridements, with local or flap coverage as needed, and long-term intravenous antibiotics. In the event of osteomyelitis, partial or complete talectomy with staged salvage arthrodesis is usually required.

Osteonecrosis

Osteonecrosis of the talar body is frequent complication following a displaced talar neck fracture, owing to the inherently tenuous blood supply. Historically, osteonecrosis rates

for displaced fractures (Hawkins type II, III, and IV fractures) treated with closed reduction and casting or pinning have ranged from 70% to 100%. Because of these factors, treatment of a displaced talar neck fracture has traditionally been considered a surgical emergency. Despite modern advances in internal fixation techniques and implants, osteonecrosis still develops in up to 30% to 50% of displaced fractures. It is well established in the literature that osteonecrosis is most related to the extent of initial fracture displacement; recent reports have suggested that the timing of

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 34

Flap Coverage for the Foot

L. Scott Levin

Alessio Baccarani

The goals of soft tissue reconstruction for the foot and ankle region are satisfactory wound coverage and restoration of function. Ancillary considerations include acceptable appearance and minimal donor site morbidity. For soft tissue coverage alone, muscle and axial fasciocutaneous flaps remain primary choices in the lower extremity. Random pattern cutaneous flaps and musculocutaneous flaps usually have more limited applications, but should be considered. Free flaps are generally the soft tissue coverage of choice for most extensive defects of the foot and of the lower third of the leg. Amputations and fillet flaps always represent a fourth possibile option when the limb cannot be preserved in its entirety. Finally, procedures such as osteotomies and/or ostectomies for the production of soft tissue â€œgainâ€• and resultant coverage of the defect are becoming more common.

Indications

Each anatomic region of the foot has certain characteristics that will influence selection of the flap to be transferred for reconstruction. The foot has special requirements for shoeing and ambulation. The reconstructive ladder for injury to the foot is based on whether there is a fracture, what part of the foot is exposed, and whether the area is weight bearing or nonâ€“weight bearing. The ankle and the

dorsum of the foot require thin, pliable soft tissue coverage for exposed tendons, bones, or joints. The plantar skin is thick and heavily keratinized, designed to resist high stress, and anchored to underlying bones and ligaments by thick fibrous connective tissue.

Topographically, the forefoot includes the dorsal areas of the metatarsals and toes. The plantar aspect includes the metatarsal heads and the instep. The hindfoot can be divided into the plantar aspect, instep, and lateral aspect of the calcaneus. The ankle can be divided into the area of the Achilles tendon and the anterior aspect of the tibiotalar joint. In the forefoot, the dorsum and the area over the toes are primarily skin and subcutaneous tissue, making the exposure of tendons and joints more probable with high-energy injury. The plantar forefoot is prone to avulsion because of the vertically oriented septa that bridge from the plantar fascia to the dermal elements of the skin. The heel pad is a very unique structure that contains cushion-like shock-absorbing chambers of fat that are not easily replaced if loss is due to such an avulsion injury.

Achilles Tendon Area

The Achilles tendon area is characterized by thin skin with little or no subcutaneous layer. Among local flaps, we consider the sural fasciocutaneous flap one of the best options for covering this area. Other flaps options include lateral supramalleolar and lateral calcaneal artery.

Ankle and Foot Dorsum

Most shallow wounds on the foot dorsum may be safely closed with a split-thickness skin graft. If necessary, exposed extensor tendons may be resected and the skin graft may be applied to the underlying periosteum, providing a simple and quick solution to the clinical problem. For management of larger wounds with bone or tendon exposure, flaps may be required. The most common local flaps used for coverage of this area are abductor hallucisâ€“abductor digiti minimi muscle flaps, extensor digitorum brevis muscle flap, lateral supramalleolar flap, and sural fasciocutaneous flap.

Plantar Forefoot

Local flaps play a major role in the management of deep wounds of the distal third of the foot. Severe injury or infection to a single toe may be best managed by toe or ray amputation, and subsequent closure by means of a plantar or of a dorsal skin flap. If additional skin is required, an adjacent toe may be filleted and trasposed for closure. Ray amputation may be necessary if the metatarsal bone is infected, injured, or devascularized. Metatarsal head ulceration is the most frequent lesion occurring in this area, especially in patients presenting with peripheral neuropathy and associated arthropathy. Many local flaps have been described to treat plantar forefoot defects. The most commonly used are neurovascular island flap (Moberg's flap), toe fillet flap, V-Y plantar flap, and suprafascial flaps medially or laterally based.

Transmetatarsal amputation provides a functional option when three or more rays have been seriously damaged, expecially in post-traumatic, ischemic, or neuropathic patients. No prosthetic or orthotic device would be necessary, and the patient may be able to wear normal shoes. Achilles tendon lenghtening should be performed in conjunction with transmetatarsal amputation to avoid equinus deformity and stump ulceration.

Plantar Midfoot

The midfoot is defined as the region between the midshaft of the metatarsals and the proximal tarsal row. It comprises the medial non-weight bearing arch as well as the more lateral weight bearing area. Small wounds in this region may be reconstructed with a variety of reconstructive options. Split-thickness skin grafts may provide adequate coverage if the transverse arch of the foot has been maintained, thus allowing the midfoot to remain a largely nonâ€“weight-bearing region. Local flap options for reconstruction of defects in this area include neurovascular island flap, V-Y advancement flap, and medially or laterally based suprafascial flaps.

Wounds larger than 4 to 6 cm generally require either free flap reconstruction or midfoot amputation to achieve a stable coverage. Clearly amputation would represent the second option if foot salvage is not indicated or possible with free tissue transfer. The two most common forms of midfoot amputation are the Lisfranc amputation and the Chopart amputation. The Lisfranc amputation is the amputation at the tarso-metatarsal joint and is associated with a high rate of equinovarus deformity. The Chopart procedure requires an intertarsal resection, just distal to the cuboid and

navicular bone. Both types of midfoot amputations will affect the patient's ability to dorsiflect and evert the residual limb because of disruption of the insertions of the peroneal and tibialis anterior tendons.

Plantar Hindfoot

Hindfoot soft tissue repair is the most challenging to the reconstructive surgeon. Reconstruction should provide durable soft tissue for safe weight-bearing, while permitting a normal ankle motion. The dualism of form and function represents a mandatory principle to be considered in the management of wounds in this area. Damage to the neurovascular and tendinous structures beneath the flexor retinaculum is an event that may impair permanently the patient's gait, mandating a below-knee amputation.

Many local flaps have been described to restore healing in the hindfoot area. The most common are intrinsic muscle flaps (abductor hallucis AH, flexor digitorum brevis FDB, abductor digiti minimi ADM); medial plantar artery flap; heel pad flaps; and sural fasciocutaneous flap. With the only exception being the sural artery fasciocutaneous flap, all the regional flaps listed require antegrade blood flow through the posterior tibial artery and its branches. Those procedures are therefore often not possible in patients affected by peripheral vascular disease.

Large hindfoot defects greater than 5 to 6 cm in patients devoid of the posterior tibial vessels should be considered for microsurgical reconstruction. Many neurosensory flaps have been described for microvascular transplantation in an effort to provide sensation to the plantar reconstruction. Lateral arm and deltoid represent two useful donor sites because of their reliable neurovascular anatomy, the ability to recover protective sensation, and because of their thickness; however, a correlation between the presence of flap sensation and the success of hindfoot reconstruction has never been established. Microvascular transplantation of muscle with skin graft coverage can provide succesful outcomes if the surgeon remembers to remove the underlying bone prominences, educates the patient on proper shoe wear, and performs frequent follow-up. We believe the use of muscle to obliterate dead space and aid in delivering antibiotics to the region is important for successful

outcomes in cases of osteomyelitis. Because of its proximity to the wound margin, the

posterior tibial artery is usually the preferred recipient vessel in hindfoot reconstruction. For hindfoot coverage, it is important to avoid scarring around the posterior tibial nerve and around the posterior tibial tendon, which may become exposed or trapped during the healing process.

Table 34-1. Surgical Reconstructive Options

Foot Region	Locoregional Flaps	Free Flaps	Amputation and Ancillary Procedures
Achilles tendon region	Sural fasciocutaneous Lateral calcaneal artery	Fasciocutaneous	Syme Below knee
Ankle and dorsum	Intrinsic muscle flaps (AH, ADM, EDB) Sural fasciocutaneous Lateral supramalleolar	Fasciocutaneous Myocutaneous	Syme Below knee
Plantar forefoot	Toe fillet flap Neurovascular island flap V-Y advancement Suprafascial flaps	Fasciocutaneous Myocutaneous Muscle	Ray Transmetatarsal
Plantar midfoot	Neurovascular island flap	Fasciocutaneous Myocutaneous	Lisfranc Chopart

	V-Y advancement Suprafascial flaps	Muscle	
Plantar hindfoot	Intrinsic muscle flaps (AH, FDB, ADM) Medial plantar artery flap Sural fasciocutaneous Suprafascial flaps (heel pad flaps)	Fasciocutaneous Myocutaneous Muscle	Syme Calcanectomy

Syme's amputation has a well-established role in the management of complex hindfoot deformities, especially in diabetic patients. The procedure involves the use of the heel pad as a soft tissue cover over the distal end of the residual tibia and fibula. Table 34-1 summarizes the main surgical reconstructive options for each region of the foot.

Preoperative Planning

Evaluation of the patient with soft tissue injury should include determination of the time of injury, mechanism, energy absorption, fracture configuration, systemic injuries, damage to the soft tissue envelope, vascularity of the extremity, sensibility, ultimate ability to salvage the foot (which is both functional and sensate), and underlying medical conditions of the patient. The principles of evaluation of orthopaedic trauma are the same for any basic medical evaluation. These principles apply whether in the outpatient clinic, emergency room, or trauma unit. An evaluation of the perfusion of the traumatized limb is of paramount importance, and if vascular (arterial) injury is suspected, a vascular surgery or microsurgery consultation should be obtained. Compartment syndrome should be considered and ruled out in any injured extremity, particularly after crush injuries. A general motor examination including the active and passive range of motion as well as a detailed sensory examination should be performed. A nerve deficit may be secondary to a spinal cord injury, nerve laceration, compartment syndrome, traction injury, or entrapment between bony fragments. The

radiologic evaluation starts with standard plain radiographic examination. Computed tomography (CT) is indicated in complex foot injuries and may give valuable information regarding soft tissue damage as well.

The wound should be inspected once, and the wound pattern and contamination noted. The next inspection of the wound should then be in the operating room under sterile conditions. Repetitive examination of open wounds in the emergency room has led to higher rates of wound infections and osteomyelitis and should be avoided. In cases of open fractures in polytrauma patients, workup of other injuries may take several hours, not to mention the need for emergent lifesaving visceral surgery that may precede definitive care for open fractures. Prophylactic antibiotics are administered and given on a regular basis until definitive wound debridement and fracture stabilization can be performed.

In summary, with any lower extremity reconstruction, three basic principles are of great importance and should therefore be carefully optimized before undertaking any reconstructive effort:

- Evaluation of underlying skeletal architecture, stabilization, and management of associated orthopaedic injuries
- Adequate wound preparation, which includes full debridement and control of any local infection prior to coverage
- Overall assessment of the patient's suitability for reconstruction and rehabilitation, including the opportunity of restoring some degree of protective sensation to the limb

Recipient Vessels

If free tissue transfer is required for wound coverage, then the last point of consideration prior to surgery should be the selection of recipient vessels for microvascular transfer as this will influence patient positioning within the operating room. A general agreement on which vessels to use has not yet been reached. Conflicting data have been reported on the survival and outcome of the transferred flaps, depending on the vessel used or the location of anastomosis proximal or distal to

the zone of injury. For example, the anterior tibial vessels may be preferred for their easy accessibility, whereas the posterior tibial vessels are strongly advocated by others due to their larger diameter.

The most important factors influencing the site of recipient vessel are the site of the injury and the vascular status of the lower extremity; it is best to ensure adequate arterial inflow and adequate venous outflow before surgery. Intraoperatively, it is imparative that the anastomosis be performed outside the zone of injury. The type of flap used, method, and type of microvascular anastomosis represent less important factors in determining the recipient vessels.

Surgery

Locoregional Flaps

As discussed in the Indications section, locoregional flaps play a very important role into the management of foot wound coverage.

Toe Fillet Flap

Anatomy

The flap is based on the medial and lateral neurovascular bundles of the toe to be amputated.

Flap Design

The toe adjacent to the wound is outlined. This island flap is better disssected with the patient in supine position under tourniquet control.

Technique

The flap is elevated beginning distally, off the distal phalanges and flexor tendons. The medial and lateral bundles are identified in the associated web spaces. A connecting incision to the wound is made. The toe is thus disarticulated at the metatarsophalangeal joint, and the dorsal skin is used for donor site closure. The flap is then rotated to the defect, ensuring a safe placement of the neurovascular structures (Fig. 34-1).

Neurovascular Island Flap

Anatomy

The flap is based on the great toe neurovascular bundle on the fibular side.

Flap Design

The flap is designed on the fibular side, centerd over the area of the neurovascular bundle. These flaps may cover wounds up to 2 to 3 cm in diameter. The use of a tourniquet facilitates a safe dissection. The donor site often requires a skin graft closure.

Technique

The flap is elevated on the lateral plantar aspect of the great toe at the level of the phalangeal periosteum. The vascular bundle can be proximally dissected in the web space to allow a longer arc of rotation. An incision is made to connect the web space to the wound, and the flap is thus transposed to the defect.

V-Y Advancement Plantar Flaps

Anatomy

These flaps are based on vertical perforating vessels throughout the plantar aspect of the foot.

Flap Design

Many V-Y advancement flaps may be designed with different orientation on the plantar foot. The use of a tourniqut facilitates a safer dissection.

FIGURE 34-1 Dorsal forefoot diabetic ulcer covered with fifth toe fillet flap shown before (A) and after (B) surgery.

Technique

The skin ajacent to the defect is incised in a V fashion (Fig. 34-2). The plantar fascia about the circumference of the flap must be incised as well. Septal attachments to the underlying metatarsal may be divided to provide further advacement. Two V-Y opposing flaps may be combined in the management of a larger wound.

Suprafascial Flaps

Anatomy

Medially or laterally based flaps of plantar skin and fat may be advanced, rotated, or transposed to cover plantar defects. Although popular in the 1970's as random flaps, they have been largely supplanted by the other techniques described. Their vascularization is based on cutaneous branches from the medial or lateral plantar arteries (Fig. 34-3).

FIGURE 34-2 Plantar forefoot pressure sore in a myelodysplastic patient treated with a V-Y advancement flap shown before (A) and after (B) surgery.

FIGURE 34-3 Plantar midfoot diabetic ulcer covered with a suprafascial rotation flap medially based shown before (A) and after (B) surgery.

Flap Design

The design of a suprafascial flap varies largely according to location of the defect. The donor site often requires a skin graft for coverage.

Technique

Medially based flaps are raised by incising laterally and elevating the subcutaneous tissues off the abductor digiti minimi muscle and plantar fascia from lateral to medial. Branches from the medial plantar nerve and artery should be preserved as they emerge from the cleft between the plantar fascia and the abductor hallucis muscle.

For the laterally based flap, the dissection is similar but requires the sacrifice of the

branches from the medial plantar artery, to allow rotation or transposition.

If the flap is designed on the heel pad to cover a small wound over the weight-bearing and posterior heel, a calcanectomy may be associated to remove bone prominences and to simplify the closure.

Intrinsic Muscle Flaps: Abductor Hallucis, Flexor Digitorum Brevis, Abductor Digiti Minimi, and Extensor Digitorum Brevis

Anatomy

The abductor hallucis muscle (AH) is vascularized proximally by branches off the medial plantar artery. The flexor hallucis brevis (FHB) is also vascularized by analogue branches. The flexor digitorum brevis (FDB) is the largest of the foot muscles, and is very useful for heel pad reconstructions. It is vascularized by branches off both the medial and lateral plantar arteries, the lateral usually being dominant.

The abductor digiti minimi (ADM) is the smallest of the muscle flaps that can be harvested. It is innervated and vascularized by the lateral plantar neurovascular bundle through branches entering its proximal portion.

The extensor digitorum brevis (EDB) is a dorsal muscle. It is vascularized by the lateral tarsal artery, which is a branch off the dorsalis pedis artery, at the level of the distal edge of the extensor retinaculum.

Flaps Design

The AH muscle with or without the medial head of the FHB and the ADM can be used to close small proximal dorsal foot wounds, but it is better qualified in the management of plantar hindfoot defects. The use of an ADM flap alone is not recommended because of its small dimensions and its short arc of rotation.

The FDB is especially qualified in heel pad reconstruction.

The EDB muscle is usually transposed proximally to cover the ankle, the dorsal foot, and the malleoli provided that the anterior tibial artery has antegrade blood flow.

Technique

When harvesting the AH, a medial foot incision is placed on the nonâ€“weight-bearing surface. The tendon is devided distally, and the muscle is separated from the medial head of the FHB, if the latter component is not required to enhance the flap dimensions. If an increase in the arc of rotation is needed, the medial plantar artery is ligated and divided distal to the branches to the abductor, and a more proximal dissection of the medial plantar artery to its origin can be accomplished.

When harvesting the FDB, a midline plantar foot incision is used to expose the muscle. The skin is elevated laterally and medially, and detached from the plantar fascia. The fascia is usually elevated with the muscle to add bulk. The four tendons are divided distally and the muscle is turned on itself, after detaching it from the quadratus plantae. The lateral plantar artery may be ligated after it passes beneath the muscle, if further mobilization is required to reach the defect. The division of the origin of the AH also increases the arc of rotation of this flap.

When harvesting the ADM, a lateral foot incision is made onto the nonâ€“weight-bearing skin. The muscle is detached from the fifth metatarsal and the tendinous insertion is divided, allowing posterior rotation of the flap. Further rotation is obtained ligating the lateral plantar artery distal to the branches to the muscle and dissecting the pedicle proximally after dividing the FDB and the AH.

When harvesting the EDB, a curvilinear incision is made on the dorsum of the foot, in continuity to the defect to be repaired. The entire dorsalis pedis pedicle is divided distally to the origin of the lateral tarsal vessels to provide the needed arc of rotation for muscle transposition. The long extensors are then dissected off the underlying short extensors muscle slips. The dissection proceeds proximally ligating the medial tarsal branches. The lateral tarsal vessels are elevated with the muscle while the origin and the tendinous extensions of the muscle are divided. The muscle may thus be rotated to the defect.

Lateral Supramalleolar Flap

Anatomy

This flap is vascularized by a perforating branch off the peroneal artery as it pierces the interosseus membrane 5 cm proximal to the tip of the lateral malleolus. Cutanous

vessels then course upward, anterior to the fibula, and anastomose with the vascular network that accompanies the superficial peroneal nerve.

Flap Design

The flap is qualified in the coverage of defects over the lateral malleolus and anterior ankle. The flap should be distally based, with the pedicle centered onto the perforator artery. Flap width includes the tissue between the fibula and tibia. The length should be 6 to 8 cm or more according to the defect. Often the flap is harvested only in its fascial component; then turned over to the defect in a book page fashion; then skin grafted. If designed in this way, the donor site can be closed directly.

Technique

The skin incision is made so that skin flaps may be elvated off the underlying fascia. The fascia is then incised anteriorly and reflected until the perforating branch is visualized. Branches of the superficial peroneal nerve are divided to allow elevation. The posterior margin is eventually incised and released from the septum between the anterior and lateral compartment of the leg.

Medial Plantar Artery Flap

Anatomy

This is a true neurosensory type A fasciocutaneous flap supplied by cutaneous fascicles from the medial and lateral plantar nerves. It is vascularized by the medial plantar artery and its vena comitans. The flap provides an invaluable amount of specialized skin that configures its chief value in heel reconstruction, because of its nonshearing, well-padded, adherent qualities.

Flap Design

Patency of anterior and posterior tibial arteries should be assessed prior to surgery. The presence of a Charcot deformity with midfoot collapse contraindicates the flap harvest. In this deformity, the instep area should remain covered by specialized plantar skin. For heel reconstruction, the flap is designed as an island centered on the medial plantar artery. It has to be outlined 2 to

3 cm proximal to the metatarsal heads, distal to the heel and medial to the lateral midsole weight bearing area. The donor site is usually closed with a split-thickness skin graft.

Technique

The flap is incised distally, exposing the medial plantar artery and nerves. After ligation of the artery, the dissection continues proximally beneath the plantar aponeurosis, including the medial plantar artery and the neurovascular bundles to the overlying fascia and skin. This requires an intraneural dissection of the midsole cutaneous branches from those fascicles supplying three and one-half digits. Often, the second common digital nerve is included with the branches to the flap because of difficulties encountered in separation. The flap is thus elevated in a distal to proximal direction in the plane between the plantar fascia and the first layer of muscles. Fascial comunications to the clefts between the underlying muscles (AH, FDB, and ADM) are cut.

Fascicles from the lateral plantar nerve may also be included in the flap following an intraneural dissection. The medial plantar artery and the fascicles from the medial plantar nerve are traced proximally to the AH muscle, which may be cut if a longer pedicle is required.

Sural Artery Fasciocutaneous Flap

Anatomy

This type A fasciocutaneous flap is innervated by the medial sural cutaneous nerve (S1-2). Its dominant vascular supply is a direct cutaneous sural artery branch that arises in the distal popliteal fossa between the two heads of the gastrocnemius muscle, and minor muscolocutaneous perforators from the gastrocnemius muscle. In most patients, the arterial supply will not be an identifiable vessel, but a â€œvascular networkâ€• that also anastomoses with the peroneal artery. The most relevant of these connections is located approximately 5 cm cephalad to the lateral malleolus. The lesser saphenous vein and its branches provide venous drainage.

Flap Design

The flap is centered between the popliteal fossa and the midposterior leg with a width up to 12 cm, but the length can be extended 20 cm to the Achilles tendon. The donor site is usually closed with a split-thickness skin graft.

Technique

The flap is raised from distal to proximal in the plane beneath the deep fascia and the above gastrocnemius muscle. The sural nerve and the lesser saphenous vein are divided distally and elevated with the flap. The pedicle is carefully dissected proximally, leaving abundant fascio-subcutaneous tissues around the neurovascular structures. It can be dissected up to 7 to 9 cm from the lateral malleolus, according to the location of the defect to be reached. After flap rotation to the defect, the skin over the pedicle is usually not sutured to avoid compression, and a skin graft is usually applied to provide coverage. Flap delay procedures and/or venous supercharging should be evaluated to avoid congestion, which commonly complicates the postoperatory course (Fig. 34-4).

Free Flaps

Latissimus Dorsi

This flap is based on the thoracodorsal artery as the major pedicle and on branches of the intercostals and lumbar arteries as secondary segmental branches. The pedicle length is 8 to 10 cm. The latissimus is innervated by the thoracodorsal nerve, which is a direct branch of the brachial plexus and enters the muscle 10 cm from the apex of the axilla. This flap's consistent anatomy and long vascular pedicle make it a common flap choice for larger defects of the foot and Achilles tendon region. For more details on flap dissection, see Chapter 11, Pedicled and Free Latissimus Flap for Elbow and Forearm Coverage. Some of the technical problems associated with this flap for lower extremity reconstruction include difficulty in positioning the patient to allow for flap elevation and simultaneous recipient site preparation. If the anterior tibial system is to be used as a recipient vessel, the ipsilateral latissimus is usually harvested; if the posterior tibial system is to be used, the contralateral latissimus should be harvested to allow for simultaneous flap elevation and recipient site exposure. In addition, in obese patients the musculocutaneous flap may be excessively thick for smooth contour over the foot (Fig. 34-5).

Rectus Abdominis

The rectus abdominis can be harvested with the patient in a supine position. This vertically oriented muscle extends between the costal margin and the pubic region and is enclosed by the anterior and posterior rectus sheats. It is a type 3 muscle (two dominant pedicles) based on the superior epigastric artery and vein and inferior epigastric artery and vein. The pedicle length is 5 to 7 cm superiorly and 8 to 10 cm inferiorly.

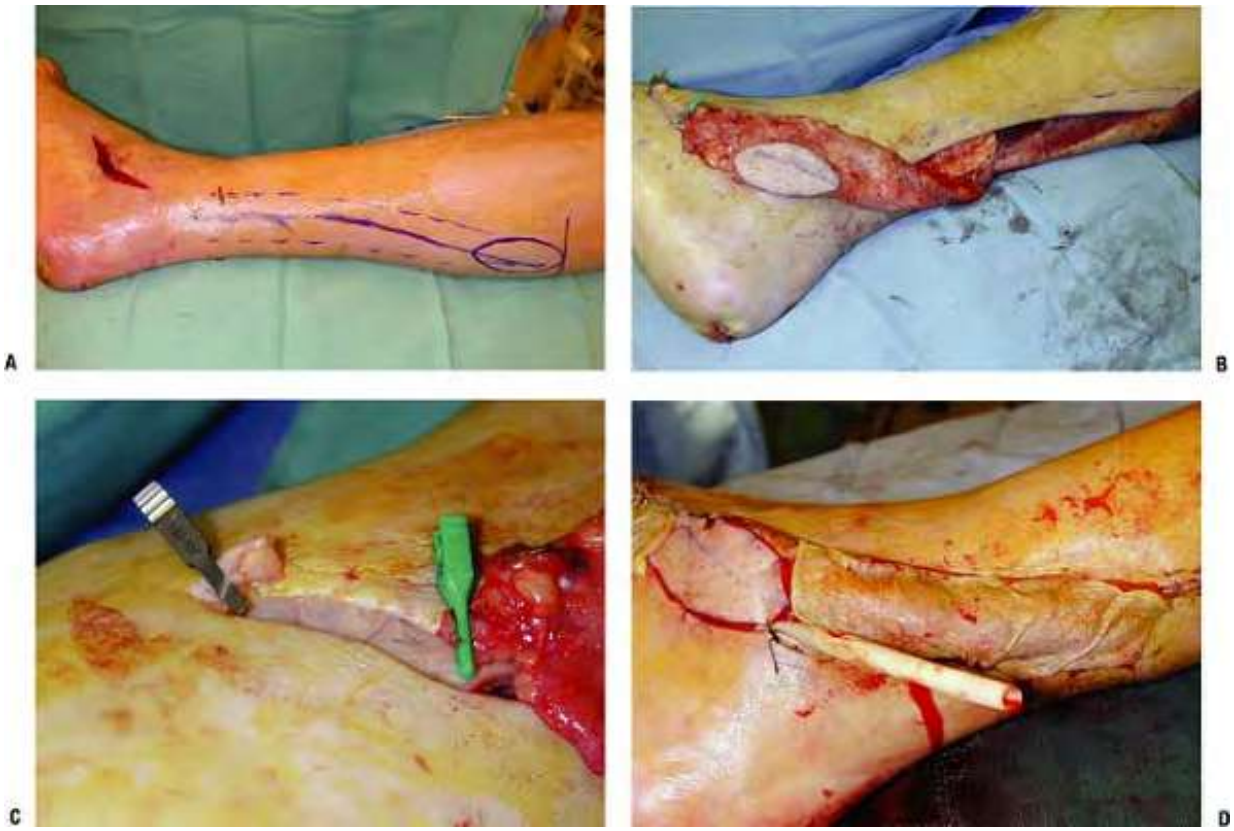

FIGURE 34-4 Wound dehiscence post triple ankle arthrodesis covered with a distally based sural flap. A: Preoperative view and flap design. B: The flap is transposed to the defect. C: The flap will be turbo-charged to improve the venous outflow (lesser saphenous vein to superficial dorsal vein). D: The pedicles are protected with split-thickness skin grafts.

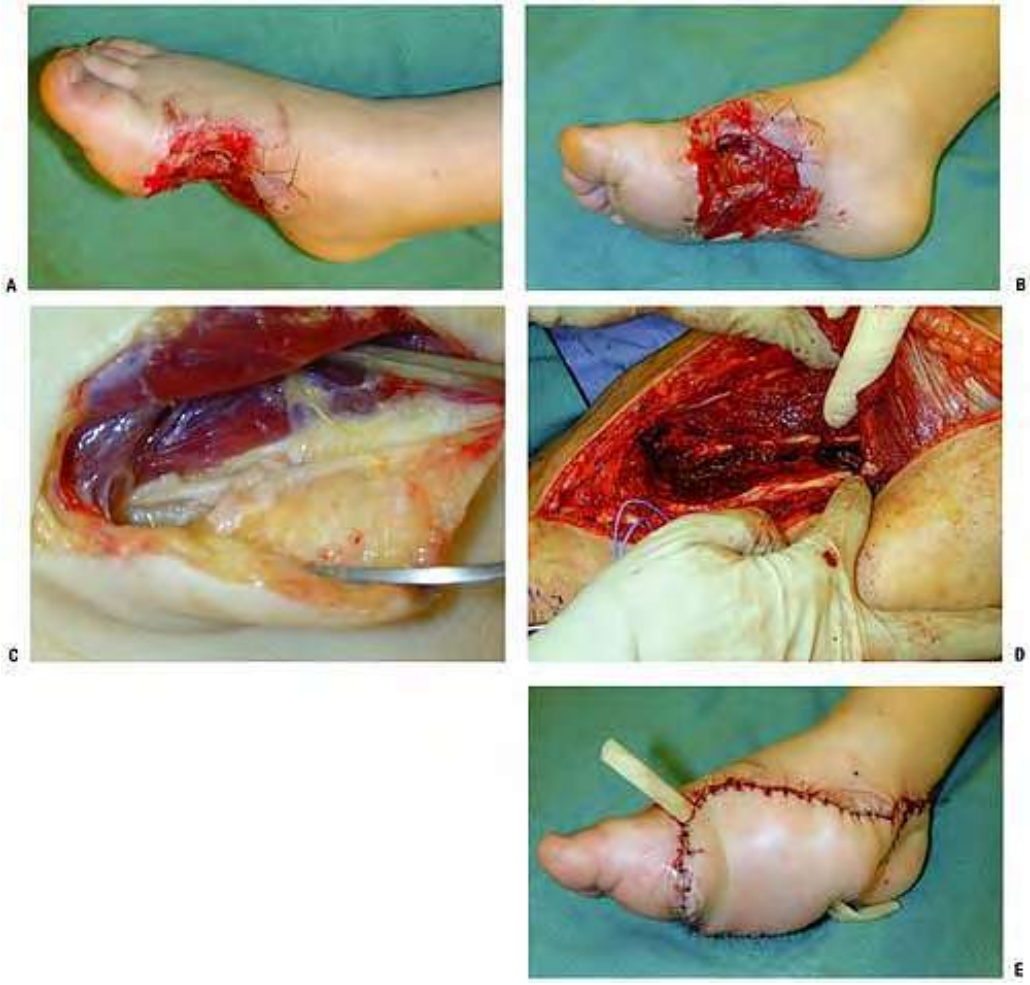

FIGURE 34-5 Medial plantar midfoot traumatic defect in a pediatric patient (lawnmower injury). The defect is reconstructed with a free latissimus dorsi flap shown before (A) and after (B) surgery. Medial plantar midfoot traumatic defect in a pediatric patient (lawnmower injury). The defect is reconstructed with a free latissimus dorsi flapshown before (A) and after (B) surgery. C,D: Latissimus dorsi flap harvesting. E: The flap is revascularized and sutured to the defect.

Each of the dominant pedicles supplies just over one-half of the muscle. There is an anastomosis between these vessels that is usually sufficient to support the

nondominant half if one of the two pedicles is ligated. Because of the larger size and easier dissection of the inferior epigastric vessel, it is usually used for free tissue transfer.

The motor innervation is supplied by segmental motor nerves from the seventh through twelfth intercostal nerves that enter the deep surface of the muscle at its middle to lateral aspects. The lateral cutaneous nerves from the seventh through twelfth intercostal nerves provide sensation to the skin territory of the rectus abdominis muscle. The size of the muscle is up to 25 Ã— 6 cm^2. The skin territory that can be harvested is 21 Ã— 14 cm^2 and is based on musculocutaneous perforatore (Fig. 34-6).

Gracilis

The gracilis is a smaller transplant and is useful for defects requiring less bulk than the latissimus or rectus. The gracilis muscle is a type 2 muscle (with a dominant pedicle and several minor pedicles). It is a thin, flat muscle that lies between the adductor longus and sartorius muscle anteriorly and the semimembranosus posteriorly. The dominant pedicle is the ascending branch of medial circumflex femoral artery and venae comitantes. The length of the pedicle is 6 cm and the diameter of the artery is 1.6 mm. The minor pedicles are one or two branches of the superficial femoral artery and venae comitantes. Their length is 2 cm and their diameter is 0.5 mm.

Motor innervation is via the anterior branch of the obturator nerve, which is located between the abductor longus and magnus muscles, and it usually enters the muscle above the level of the dominant vascular pedicle. The anterior femoral cutaneous nerve (L2â€“3) provides sensory innervation to the majority of the anterior medial thigh.

This muscle functions as a thigh adductor. The presence of the adductor longus and magnus makes it an expendable muscle.

The size of the muscle is 6 Ã— 24 cm^2. The skin territory is 16 Ã— 18 cm^2, but the skin over the distal half of the muscle is not reliable when the flap is based on its dominant vascular pedicle with division of the minor vascular pedicles. In obese patients, the musculocutaneous flap may be too bulky, necessitating use of a skin graft placed on the muscle (Fig. 34-7) (see Chapter 32).

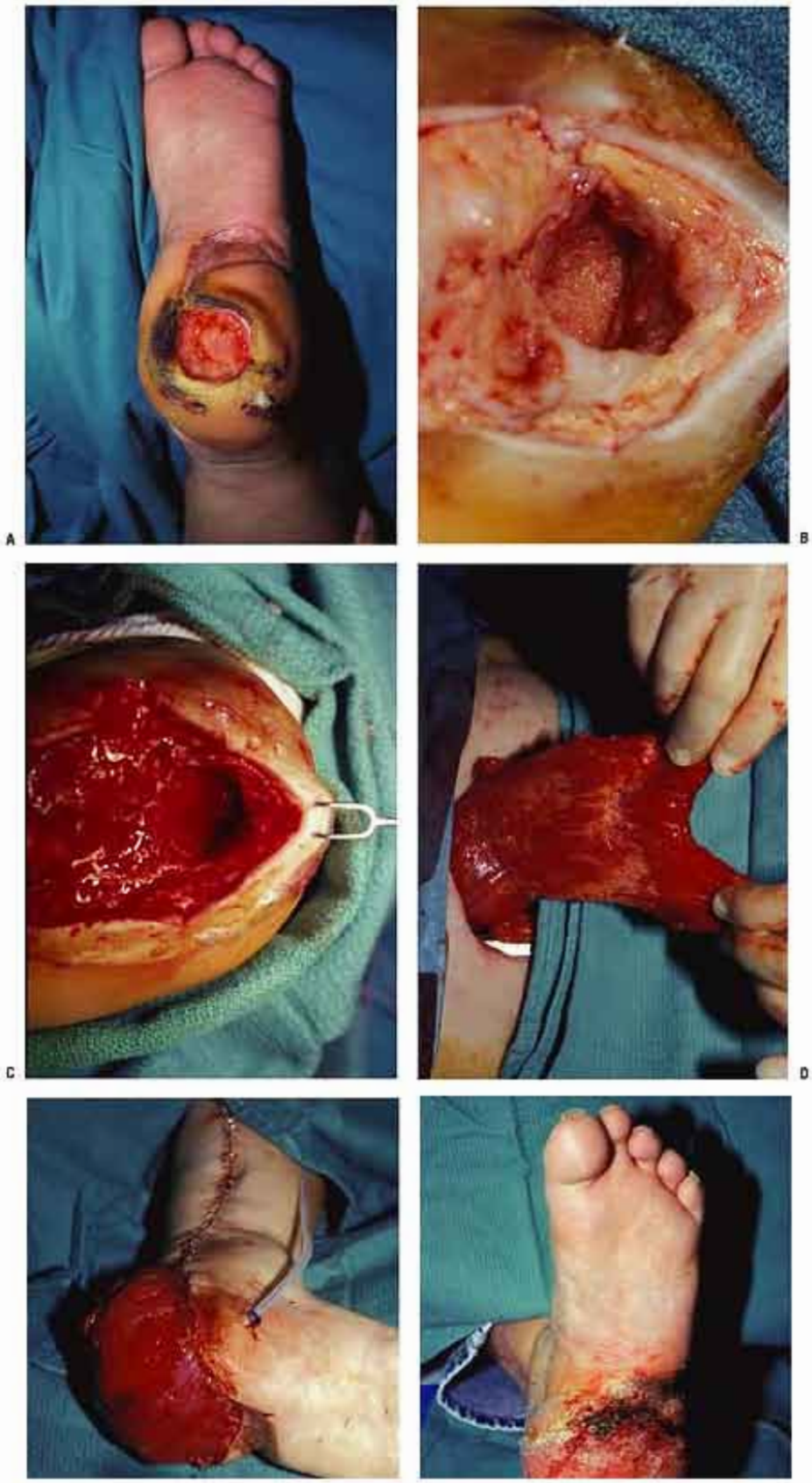
A
B
C
D

FIGURE 34-6 Calcaneal osteomyelitis on a diabetic foot. The defect was reconstructed with a free rectus abdominis flap. A,B: Preoperative view. C: After debridement. D: The rectus muscle is harvested from the abdomen. E: The flap is tailored to the defect and revascularized. F: Late postoperative view showing calcaneal salvage.

P.442
P.443

FIGURE 34-7 Gunshot wound to the distal leg-hindfoot. The defect is reconstructed with a free gracilis flap. A: Significant soft-tissue and bony defects are present. B: Debridement and external fixation. C: A myocutaneous free gracilis flap is planned. D: The free flap is harvested. E: The flap is inset to the defect and revascularized.

Radial Forearm Flap

This is a thin, well-vascularized fasciocutaneous flap on the ventral aspect of the forearm that was widely used in China before it was popularized in the Western literature. The flap is based on the radial artery, which can achieve a 20-cm pedicle and has a diameter of 2.5 mm. This length of the pedicle facilitates the microsurgical anastomosis out of the zone of injury. The venous drainage is through the venae comitantes of the radial artery, but the flap can include the cephalic vein, the basilic

vein, or both. The flap can contain the lateral antebrachial cutaneous nerve or the medial antebrachial cutaneous nerve and then serve as a neurosensory flap. The size of the flap can be 10 Ã— 40 cm^2.

A portion of the radius can be included as a vascularized bone with this flap. The advantages of this flap are a long pedicle and potential sensory innervation. The quality of the bone from the radius is mainly cortical and not of any substantial volume. Including the bone in the radial forearm flap may lead to stress fracture of the donor radius. Preliminary tissue expansion will increase the flap dimensions, and more importantly, it will allow direct closure of the donor defect (Fig. 34-8).

Scapular and Parascapular Flap

The scapular flap remains the workhorse of skin flaps. It is a thin, usually hairless, skin flap from the posterior chest and can be de-epithelialized and used as subcutaneous fascial flap, pedicled or free.

The flap is perfused by the cutaneous branches of the circumflex scapular artery (CSA) and drained by its venae comitantes. The CSA is the main branch of the subscapular artery and the main blood supply to the scapula, the muscles that attach to the scapula, and the overlying skin. The length of the pedicle is 5 cm and the diameter of the artery is 2.5 mm. The vascular pattern of this territory makes it possible to raise multiple skin flaps on a single vascular pedicle or to harvest the lateral border of the scapula as an osteocutaneous flap for a complex reconstruction.

The cutaneous territory can be 20 Ã— 7 cm^2 and can be divided in two componentsâ€”a horizontal territory (horizontal scapular flap) and a vertical territory (parascapular flap)â€”based on the branches of the CSA after the vessel courses through the triangular space. Preliminary expansion of the territory of the scapular flap will increase the flap dimensions and permit direct donor-site closure.

This flap can be combined with other flaps based on subscapular blood supply and may greatly facilitate certain complex reconstructions. These include the latissimus dorsi and serratus anterior flaps, which can supply additional skin, muscle, and bone (rib) if necessary. The primary indication for the scapular flap is a defect requiring a relatively thin, large cutaneous flap. These kinds of defects are often found in the foot. The osteoseptocutaneous free scapular flap reconstruction has been described in

the lower extremity (Fig. 34-9).

FIGURE 34-8 Post-traumatic left great toe necrosis. After debridement the soft tissue defect is restored with a free radial forearm flap. A: Preoperative view. B: After debridement. C: Radial forearm flap is tailored to cover the exposed phalanx. D,E: Four weeks postoperative showing toe salvage and flap donor site.

Postoperative Management

The success of a reconstructive foot surgery importantly relies on well-planned multidisciplinary postoperative management and rehabilitation. A nonâ€“weight-bearing regimen for at least 3 weeks is mandatory for every patient carrying plantar sutures. Elevation is required for 3 to 8 days after a local flap or a skin graft, and for 2 to 3 weeks after a free flap. Heparin regimen is advisable while the patient is in the aforementioned bed rest phase. Elastic wraps may be useful in controling edema once the patient's limb is permitted in a dependent position. Clinical findings should dictate the use of antibiotics.

The Lâ€™Enard splint (Fig. 34-10) is a useful tool to provide immobilization of the foot and ankle, and in keeping the posterior heel off the bed. It is also useful when the patient begins nonâ€“weight-bearing ambulation with crutches. Local care of all weight bearing surfaces is also recommended after suture removal, together with a frequent multidisciplinary follow-up that involves the plastic surgeon, the orthopaedic surgeon, the podiatrist and, if required, the prosthetist.

FIGURE 34-9 Dorsal foot avulsion injury in a pediatric patient (go-cart trauma). The dorsal aspect of the foot is reconstructed with a free scapular-parascapular flap. A: After debridement. B: Free scapular-parascapular flap is inset to the defect and revascularized. C: Two months postoperative view.

Postoperative care of free tissue transfer patients requires the patients to be adequately hydrated. Maintenance of proper body temperature and hematocrit is also important. Routine heparinization and anticoagulation are not used.

Flaps are usually monitored for a minimum of 5 days with a laser Doppler in addition to clinical observation. While the immediate postoperative period of 24 to 48 hours is critical, there have been occasional late failures; thus, laser Doppler monitoring should be continued for 4 or 5 days.

FIGURE 34-10 Lâ€™Enard splint.

Complications

Any flap failure requires a new detailed evaluation of the patient's local and general conditions before undertaking a new reconstruction. If a locoregional flap fails, and no other local option is available, the next step on the reconstructive ladder would be a free flap.

The success of free tissue transfer should be on the order of 95% to 99%. Acute complications usually occur in the first 48 hours and include venous thrombosis, arterial thrombosis, hematoma, hemorrhage, and excessive flap edema. Arterial insufficiency can be recognized by decreased capillary refill, pallor, reduced temperature, and the absence of bleeding after pinprick. This complication can be caused by arterial spasm, vessel plaque, torsion of the pedicle, pressure on the flap, technical error with injury to the pedicle, a flap harvested that is too large for its blood supply, or small vessel disease (due to smoking or diabetes). Management of arterial compromise requires prompt surgical intervention to restore the blood flow. Pharmacologic intervention at the time of exploration includes use of vasodilators, calcium channel blockers, and systemic anticoagulants for flap salvage presenting with

arterial insufficiency. Ultimately, if these pharmacologic agents do not relieve spasm at the level of the arterial inflow, the anastomosis should be redone to rule out intra-arterial thrombus.

Venous outflow obstruction can be suspected when the flap has a violaceous color, brisk capillary refill, normal or elevated temperature, and production of dark blood after pinprick. Venous insufficiency can occur due to torsion of the pedicle, flap edema, hematoma, or tight closure of the tissue over the pedicle. The venous outflow obstruction can result in extravasation of red blood cells, endothelial breakdown, microvascular collapse, thrombosis in the microcirculation, and finally flap death. Given the irreversible nature of the microcirculatory changes in venous congestion that occur even after short periods of time, the surgeon must recognize venous compromise as early as possible.

These complications can occur alone or in any combination. The clinical observation and the monitoring of the patient (such as with laser Doppler) should alert the surgeon who has to decide between conservative and operative intervention. Conservative treatment may include drainage of the hematoma at the bedside with release of a few sutures to decrease pressure. In cases of venous congestion, leeches may be helpful if insufficient venous outflow cannot be established despite a patent venous anastomosis. The leeches inject a salivary component (hirudin) that inhibits both platelet aggregation and the coagulation cascade. The flap is decongested initially as the leech extracts blood and is further decongested as the bite wound oozes after the leech detaches.

The donor site should be given the same attention as the recipient site during the postoperative period. Complications of the donor site include hematoma, seroma, sensory nerve dysfunction, and scar formation.

Occasionally free flaps, despite early return to the operating room for vascular compromise, do fail. Options for management include the performance of a second free tissue transfer, noting the technical or physiologic details that led to initial failure. Most of the time, free tissue transfers that fail are due to technical errors in judgment, whether they be flap harvest, compromise of the pedicle during the harvest, improper microvascular technique during anastomosis, improper insetting resulting in increased tissue tension and edema, or postoperative motion of the extremity resulting in pedicle avulsion. The next decision made by the operating

surgeon as to the management of this patient is based on several factors. If a patient required a free flap in the first place, a second free flap should be considered. If a decision is made not to redo the flap, it could be left in place using the Crane principle to see if underlying granulation will be sufficient such that skin grafting can be performed once the necrotic flap is removed.

The Crane principle can be applied to cases where a local flap or free tissue transfer that necrotizes in part or totally acts as a biologic dressing or eschar over a wound bed. If there is no infection, the eschar can be left on the wound bed to see if some healing in the form of granulation occurs underneath it. Ultimately, the eschar is removed and the granulation bed skin grafted, obviating another free tissue transfer. If wound observation shows that such a bed is not produced, then a second flap must be considered.

It is usually our preference not to follow this course, as the flap can become a source of sepsis and further compromise local tissues. Necrotic nonviable flaps should be removed and a temporary wound dressing such as an antibiotic bead pouch or wound vacuum-assisted closure (VAC) should be used. Occasionally when flaps fail in a severely compromised extremity, consideration should be given to amputation. If a second free flap is considered, errors that lead to flap compromise need to

be recognized and avoided. It may be prudent to obtain an arteriogram, evaluate the coagulation profile, and research other issues that might have led to failure.

Recommended Reading

Arnold PG, Yugueros P, Hanssen AD. Muscle flaps in osteomyelitis of the lower extremity: a 20-year account. *Plast Reconstr Surg*. 1999;104:107.

Attinger CE, Ducic I, Zelen C. The use of local muscle flaps in foot and ankle reconstruction. *Clin Podiatr Med Surg*. 2000;17(4):681.

Colen L, Uroskie T. Foot reconstruction. In: Mathes SJ, ed. *Plastic surgery.* 2nd ed. Vol.6. Philadelphia: Elsevier; 2006.

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 35

Soft Tissue Coverage of the Pelvis and Sacrum: Hemipelvectomy and Pedicled Flap Coverage

Alex Senchenkov

Robert Esther

Franklin H. Sim

Steven L. Moran

Extensive defects of the pelvis and sacrum can result from tumors, ablation, and severe trauma. Due to limited amounts and relative immobility of the pelvic soft tissue, these defects may pose a serious reconstructive challenge. Until the late 1970s, the majority of large pelvic tumors were treated with external hemipelvectomy. Advances in imaging, chemotherapy, and radiation therapy, as well as improvements in resection and reconstructive techniques, have greatly reduced the need for radical lower extremity amputations, allowing limb preservation in a majority of cases.

Historically, buttock tumors were not amenable to a classic hemipelvectomy and just a few decades ago were considered unresectable. Likewise, extensive buttock defects inflicted by trauma, infection, or end-stage pressure ulcers in paraplegics

could not be effectively reconstructed. Secondary intention healing frequently resulted in protracted hospital course, extensive scarring, contractures, and unstable soft tissue coverage. Many of these patients were bound to years of ongoing wound care and immobility.

External hemipelvectomy denotes removal of the hemipelvis with affected lower extremity by disarticulation of the pubic symphysis and the sacroiliac joint. Because external hemipelvectomy resulted in major functional impairment, limb-sparing procedures removing part or all innominate bone with preservation of the extremity have been advocated. These pelvic resections are referred to as internal hemipelvectomies.

Large, composite pelvic defects associated with internal hemipelvectomies are more challenging to reconstruct than the soft tissue defect typically created in external hemipelvectomy patients for two main reasons. First, following removal of bony hemipelvis in external hemipelvectomy, a large amount of soft tissue of the buttock or proximal thigh becomes available for reconstruction. Second, a decrease of the pelvic volume obliterates the dead space.

Sacral resections are performed as a part of extended external hemipelvectomy for musculoskeletal sarcomas and, as such, reconstructed as a part of hemipelvectomy closure. Isolated sacral defects result from composite pelvic resections for locally advanced anal and rectal malignancies or tumors intrinsic to the sacrum such as sacral chordomas and sarcomas.

Indications/Contraindications

When embarking on treatment of pelvic sarcomas, three important questions should be borne in mind.

- Is this patient operable, i.e., can the individual medically withstand a major oncologic resection?
- Is this tumor resectable, i.e., can this patient be rendered disease-free surgically?
- Can the residual defect or deformity from the proposed resection be reconstructed in a functionally satisfactorily manner with stable soft tissue

coverage?

The answers to these questions have to be determined during preoperative evaluation by the surgical oncologist, reconstructive surgeon, and anesthesiologist. Resection of the tumor with negative margins is the only reliable means of obtaining a cure in cases if tumor.

Internal hemipelvectomy is indicated in cases of localized tumor where margin negative resection of the tumor is possible with preservation of the lower extremity. If clean margins cannot be achieved, external hemipelvectomy should be performed. Main indications for external hemipelvectomy are: large tumors involving multiple compartments unresponsive to neoadjuvant therapies, contamination of compartments from pathologic fracture, or failed previous resection, a nonviable extremity. Nononcologic external hemipelvectomy may be performed in the cases of uncontrolled pelvic osteomyelitis, traumatic hemorrhage, and failed aorto-femoral revascularizations (Table 35-1).

Wound complication rates following hemipelvectomy are notoriously high and have been reported to range from 20% to 80%. Proper technical execution of the procedure and the use of well-designed skin and muscle flaps can minimize postoperative wound morbidity.

Although infrequently, pelvic and sacral resections are performed en bloc with pelvic visceral structures for locally advanced rectal and gynecologic malignancies eroding or invading the skeletal pelvis. When such pelvic resection involves removal of a part of the pelvis or sacrum, it is referred to as *composite resection*. Any type of external hemipelvectomy performed in continuity with

visceral structures is known as *compound hemipelvectomy*. Due to the aggressive nature of these tumors, the disease has to be limited to the pelvis and extensive imaging is required to select the patients that can benefit from these extensive operations.

Table 35-1. Basic Tumor Flap Principles

- Safe oncologic margins are the primary requirement
- Reconstruction does not take precedence over adequate, safe resection
- Adequate soft tissue coverage of bony reconstruction/prosthesis and neurovascular structures
- Healed surgical wound
- Durable
- Minimal donor morbidity
- Appropriate function, contour

Primary sacral tumors such as chordomas and sarcomas are relatively uncommon. The majority of these tumors are low-grade malignancies. They infrequently metastasize and therefore local control becomes important. Reconstruction of these defects with flaps facilitates optimal postoperative wound healing.

Preoperative Planning

Prior to surgical resection, patients should undergo local and systemic staging studies. Musculoskeletal malignancies have a propensity to pulmonary spread. Therefore, a chest CT is mandatory to screen for systemic disease. An MRI (and plain radiographs for primary bone tumors) is sufficient for gauging the local extent of disease and response to treatment. A CT of the pelvis is often useful to complement the MRI as this area is difficult to image. Surgical planning relies on MRI images taken before and after neoadjuvant therapies. Pretreatment MRI images may be helpful in distinguishing radiation-induced reactive changes from actual tumor tissue.

Most patients with high grade bone malignancies will undergo some form of neoadjuvant treatment, including chemotherapy and/or radiation therapy prior to tumor resection. Typically, primary bone sarcomas such as osteosarcoma are treated with several cycles of preoperative chemotherapy, surgery, and then

several additional cycles of chemotherapy. Radiation therapy also has an established role in treatment of soft tissue sarcomas.

The treatment team must choose between pre- and postoperative radiation therapy. Both approaches have advantages and shortcomings. Preoperative treatment requires a smaller area of treatment, creation of a fibrous rind around the tumor, and often causes tumor shrinkage, leading to an improved ability to obtain wide margins without sacrificing vital structures. Preoperative radiation's disadvantages include a higher rate of wound problems and less viable tumors available for pathologic examination. Postoperative radiation has the advantage of earlier surgery, viable tumors for pathologic study, and fewer wound complications. Treatment volumes however are increased and there is a delay in administering treatment to allow time for adequate healing of operative wounds. We prefer preoperative radiation for pelvic and retroperitoneal sarcomas.

Brachytherapy requires proper reconstruction planning so that flaps do not interfere with catheter placement (Fig. 35-1). Afterloading catheters should be evenly spaced and sutured in place to the tumor bed with fast absorbable sutures to prevent their displacement during postoperative therapy. Alternatively, VAC dressing can be used as a temporary coverage of brachytherapy catheters, followed by delayed primary reconstruction of the defect after completion of brachytherapy. Intraoperative radiation therapy (IORT) is another means of augmenting a preoperative radiation therapy regimen, allowing for directed treatment at close intraoperative margins.

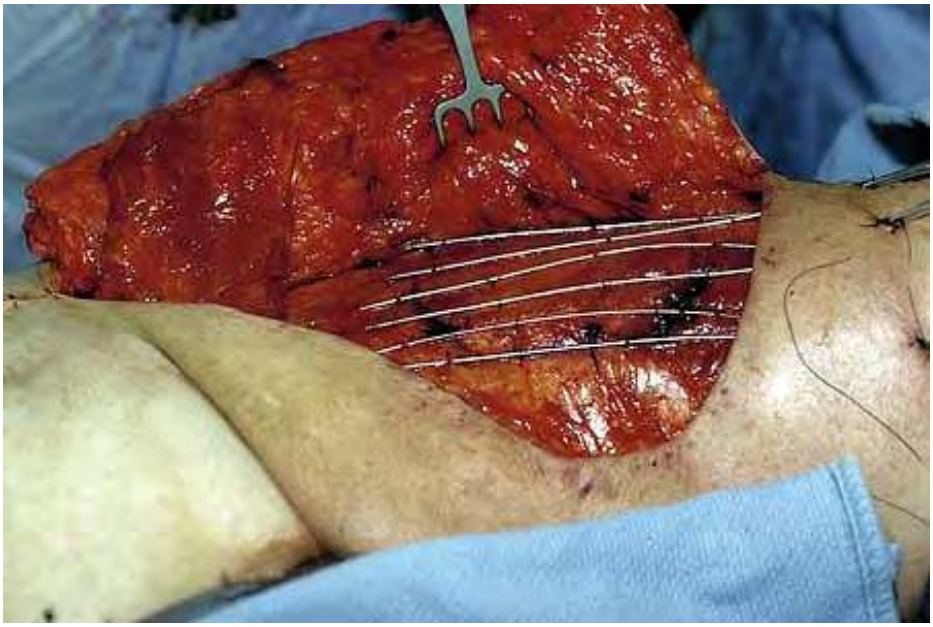

FIGURE 35-1 Placement of afterloading brachytherapy catheters under inferiorly based TRAM flap in treatment of recurrent sarcoma of the thigh.

Table 35-2. Principles of Pelvic Reconstruction

- Routine use of preoperatively placed ureteral stents aids in the identification of ureters intraoperatively.
- Patients should receive both antibiotic and mechanical bowel prep.
- Standard hemipelvectomy flaps provide adequate soft tissue coverage, and there is no difference in postoperative wound complications related to flap design.
- Abdominal and pelvic wall reconstruction is not necessary if the muscle with its investing fascia is a component of an external hemipelvectomy flap, but the abdominal wall should be

reconstructed in cases of internal hemipelvectomy.

- Omentum should be interposed between the intestine and hemipelvectomy flap whenever possible to wall off intra-abdominal contents in the event of hemipelvectomy flap necrosis.
- Precise surgical technique, thorough hemostasis throughout the entire hemipelvectomy wound, wide drainage with multiple closed suction drains, dÃ©bridement of all nonviable tissue off hemipelvectomy flap, and meticulous layered closure of the operative wound are the key to successful postoperative healing.
- Hemipelvectomy wound complications are common and related to the extensive nature of the procedure and the level of vascular ligation. Infected wounds have to be aggressively dÃ©brided until control of the wound is achieved.
- Secondary intention healing with wet-to-moist dressing changes and VACÂ® therapy is the most common approach to complicated hemipelvectomy wounds.
- Tertiary closure is reserved for the patients with healthy granulating wounds and extensive defects who are otherwise good operative candidates.
- Local tissue rearrangement by advancement of the skin flaps is the most common form of tertiary intention closure.
- Contralateral inferiorly based rectus abdominis muscle or musculocutaneous flap is the flap of choice for tertiary closure of large hemipelvectomy wounds. To preserve contralateral rectus abdominis muscle, contralateral ostomies should be avoided by careful preoperative planning and communication between different surgical specialties involved in this procedures.
- Hemipelvectomy reconstruction with a microvascular fillet flap obtained by ex-vivo anatomic exploration of an amputated extremity is an underutilized reconstructive option and should be considered in the cases of paucity or poor quality of local tissues and vascular ligation above the bifurcation of common iliac vessels that is plagued with the high rate of flap necrosis rate.

Preoperatively, the patients with large pelvic tumor undergo mechanical bowel preparation and intravenous antibiotic coverage. Ostomy sites must be preoperatively marked in accordance with anticipated flap use because inappropriate colostomy or ileostomy placement may burn an important reconstructive bridge and prevent rectus abdominis flap elevation. Involvement of several surgical services such as urological, colorectal, vascular, spine, and plastic surgery is common. The patient is positioned on a bean bag prior to induction of general anesthesia. Large bore intravenous access is established in an event of rapid blood loss. We liberally use ureteral stents that facilitate intraoperative identification of the ureters. After placement of the stents and Foley catheter, the patient is placed in the â€œsloppyâ€• lateral decubitus position and is secured with the bean bag. This position is preferred for internal or external hemipelvectomy because it permits a wide skin preparation and an easy access to the abdomen, buttock, and perineal regions. If additional procedures on spine, sacrum, or rectum need to be performed, intraoperative repositioning of the patient will be required (Table 35-2).

Low sacral resection can also be performed in the â€œsloppyâ€• lateral decubitus position (abdominolateral sacral portion) or a full lateral position with the hip and knee joints in 90-degree flexion. When combined abdominal exploration may be required to deal with the intrapelvic anterior component of the tumor, we start the abdominal portion of the operation supine and later reposition the patient for the posterior, sacral stage of the procedure. Plastic surgeon performs an initial marking and flap dissection as dictated by an anticipated defect.

Surgery

External Hemipelvectomy Reconstruction

Pelvic reconstruction following external hemipelvectomy is principally accomplished with three pedicled flap designs: posterior, long anterior, and total thigh fillet flaps. The vast majority of hemipelvectomy defects can be closed with these flaps which constitute the *first choice for hemipelvectomy flap reconstructions*. If these standard hemipelvectomy flaps are unusable due to very proximal vascular ligation, causing flap ischemia, division of the flap origin during tumor resection or previous procedure, or extensive radiation damage, then

alternative flaps must be used for coverage. These *second-line reconstructive options* include contralateral inferiorly based vertical rectus musculocutaneous (VRAM) flaps, microvascular lower extremity fillet flaps, or standard free flaps depending on the defect

configuration. Likewise, the second-line reconstructive options are useful for closure of hemipelvectomy wounds in the setting of postoperative wound complications (Table 35-3).

Table 35-3. Purpose of Soft Tissue Flaps in Tumor Surgery

- Covers exposed neurovascular structures
- Coverage for endoprostheses or allografts
- Addresses functional deficits (neurotized flaps)
- Obliterates dead space
- Enhances healing of irradiated wounds

Posterior Hemipelvectomy Flap

The classic hemipelvectomy technique relies on pelvic exploration, ligation of the common iliac vessels, division of the pelvic rim by disarticulation of the pubic symphysis and the sacroiliac joint, and creation of the posterior fasciocutaneous flap to achieve soft tissue closure. It was initially recommended that gluteal muscles be left with the specimen. This fasciocutaneous hemipelvectomy flap was based on relatively poor random blood supply due to ipsilateral ligation of the common iliac vessels and was further compromised by removal of the gluteal muscles that greatly increased wound complication rates.

Three modifications of this classic technique aimed to decrease high wound

complication rates:

- Incorporation of gluteal muscles in the hemipelvectomy flap
- Whenever oncologically appropriate, ligation at the level of external iliac vessels with reservation of the internal iliac vessels to improve the flap blood supply
- Limited resection of the bony pelvis that allows preservation of the sacral perforators

With these modifications, the posterior hemipelvectomy flap is designed as a musculocutaneous flap based on the superior and inferior gluteal vessels (Fig. 35-2). Preservation of the gluteal muscle decreases posterior flap necrosis rates and makes the construction of a long, viable posterior flap that would reach up to or above the level of umbilicus possible. Impact of the level of vascular ligation on hemipelvectomy wound outcomes has been a point of controversy. Several reports from Karakousis et al suggested that the level of vascular ligation does not affect the posterior hemipelvectomy flap viability and the rate of postoperative wound complications. These authors believed that there was an adequate blood supply of the gluteal muscle through small arterial branches along its sacral origin, which was sufficient to sustain the viability of the flap unless resection of the edge of the sacrum was oncologically necessary. In our experience, we found 2.7-fold higher rates of posterior hemipelvectomy flap necrosis in the patients that had ligation at the level of common iliac vessels. This finding was independent from sacral resection performed during extended hemipelvectomy in some of these patients (1).

Long Anterior and Total Thigh Fillet Hemipelvectomy Flaps

One of the major limitations of the posterior flap external hemipelvectomy is its inability to deal with the advanced tumors of the buttock and posterior pelvis in an oncologically sound manner. In 1953, Bowden et al described utilization of the skin of the femoral triangle based on the preserved segment of the superficial femoral artery for closure of the hemipelvectomy performed for the sarcoma of the

buttock. However, it was the critical need for soft tissue reconstruction of the advanced decubiti and the infection of bony pelvis in paraplegic patients that led to increased utilization of the soft tissue obtained from high amputations. The total thigh flap was proposed by Georgiade et al as a last-resort reconstructive option for such patients in the 1950s and subsequently gained wide-spread use. This principle was subsequently applied for coverage of the hemipelvectomy defects whereby a musculocutaneous flap of the anterior thigh compartment was elevated based on the superficial femoral artery. The technique was further refined by Sugarbaker et al, who also demonstrated that the anterior flap can be used as a sensate island flap based on the superficial femoral vessels and saphenous nerve.

Standard long anterior flap hemipelvectomy includes the bulk of quadriceps femoris muscle (Fig. 35-3). A total thigh fillet flap utilizing the majority of the thigh musculature can also be designed as a variation of the anterior hemipelvectomy flap technique (Fig. 35-4). Anterior hemipelvectomy flap is an axial pattern musculocutaneous flap based on the branches of femoral vessels, including lateral

and medial circumflex arteries which arise from branches from the profundus femoris artery. The latter perforate the adductor magnus muscle to the posterior and lateral compartments of the thigh and play a role in supporting a total thigh fillet flap.

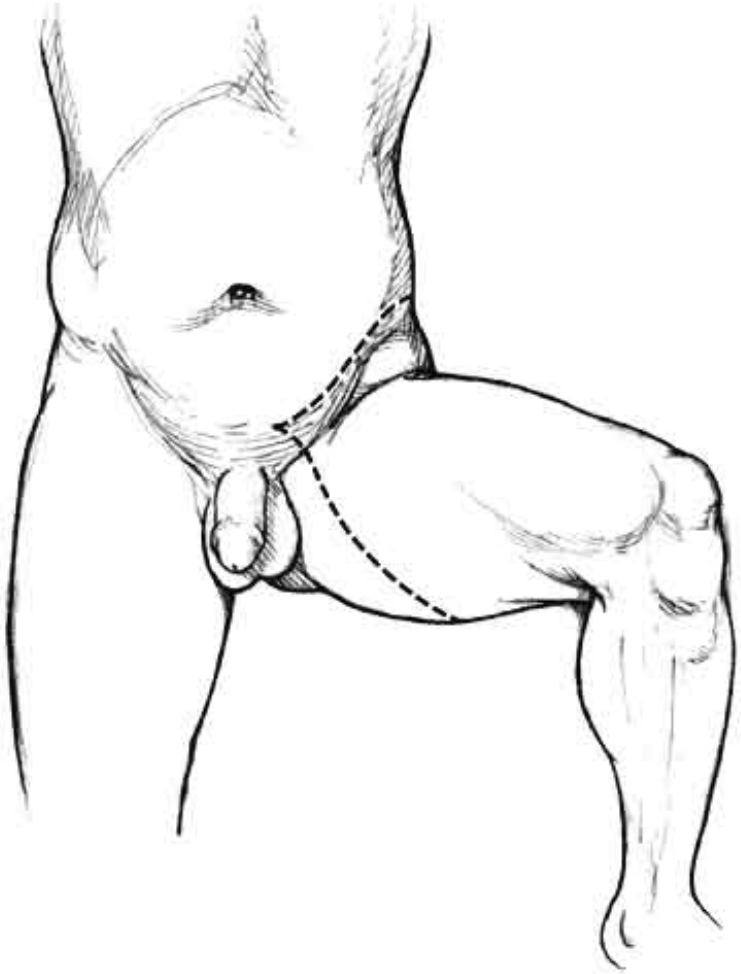

A

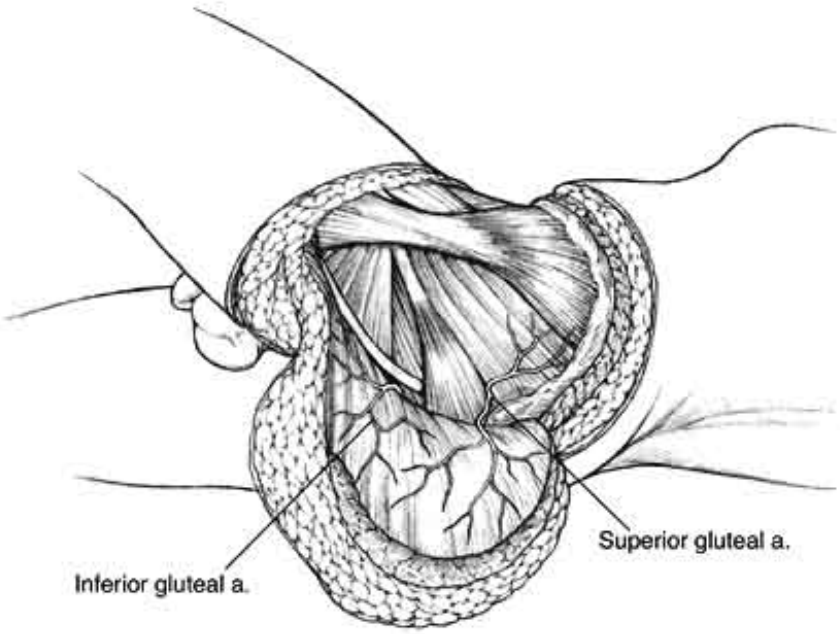

B

FIGURE 35-2 A: Skin markings of external hemipelvectomy with utilization of the posterior flap. B: Musculocutaneous design of the posterior hemipelvectomy flap based on the superior and inferior gluteal vessels.

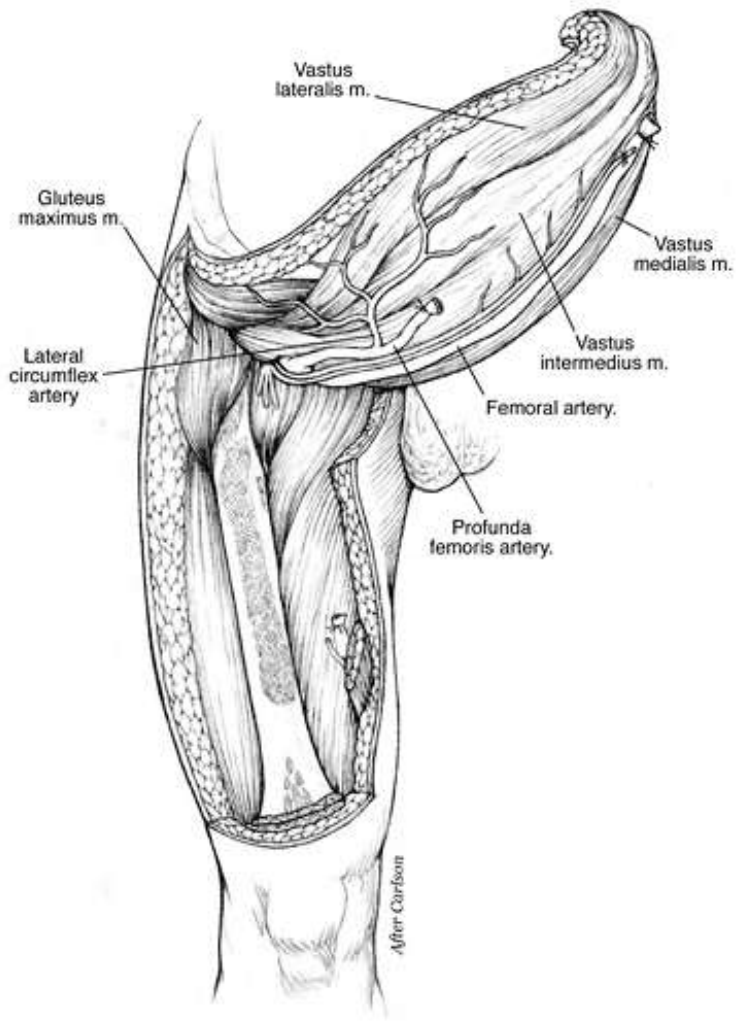

FIGURE 35-3 Long anterior hemipelvectomy flap based on the branches of profunda femoris and superficial femoris vessels.

The skin of the anterior thigh down to the knee is innervated by the lateral and

anterior femoral cutaneous nerves. These nerves can be preserved to provide sensory flap coverage of the hemipelvectomy defect. Both anterior flap and total thigh fillet flap hemipelvectomy provide well-vascularized and sensate immediate coverage of the hemipelvectomy defect. An effort should be made to preserve innervation of this flap by protecting sensory nerves during the dissection (Fig. 35-5). Both long anterior and total thigh fillet hemipelvectomy flaps are sufficient to cover even very extensive hemipelvectomy defects (as well as spinopelvic resections) with tissues that have rich axial-pattern blood supply.

Second-Line Hemipelvectomy Reconstructions

Standard hemipelvectomy flaps either posteriorly or anteriorly based provide reliable reconstruction in a vast majority of the cases because removal of the hemipelvis creates relative soft tissue excess for three reasons:

- Reduction of pelvic volume decreases requirement for the size of soft tissue envelop.
- Skeletal resection eliminates the issue of the dead space.
- Tissues or the proximal part of the amputated lower extremity can be used for hemipelvectomy defect reconstruction.

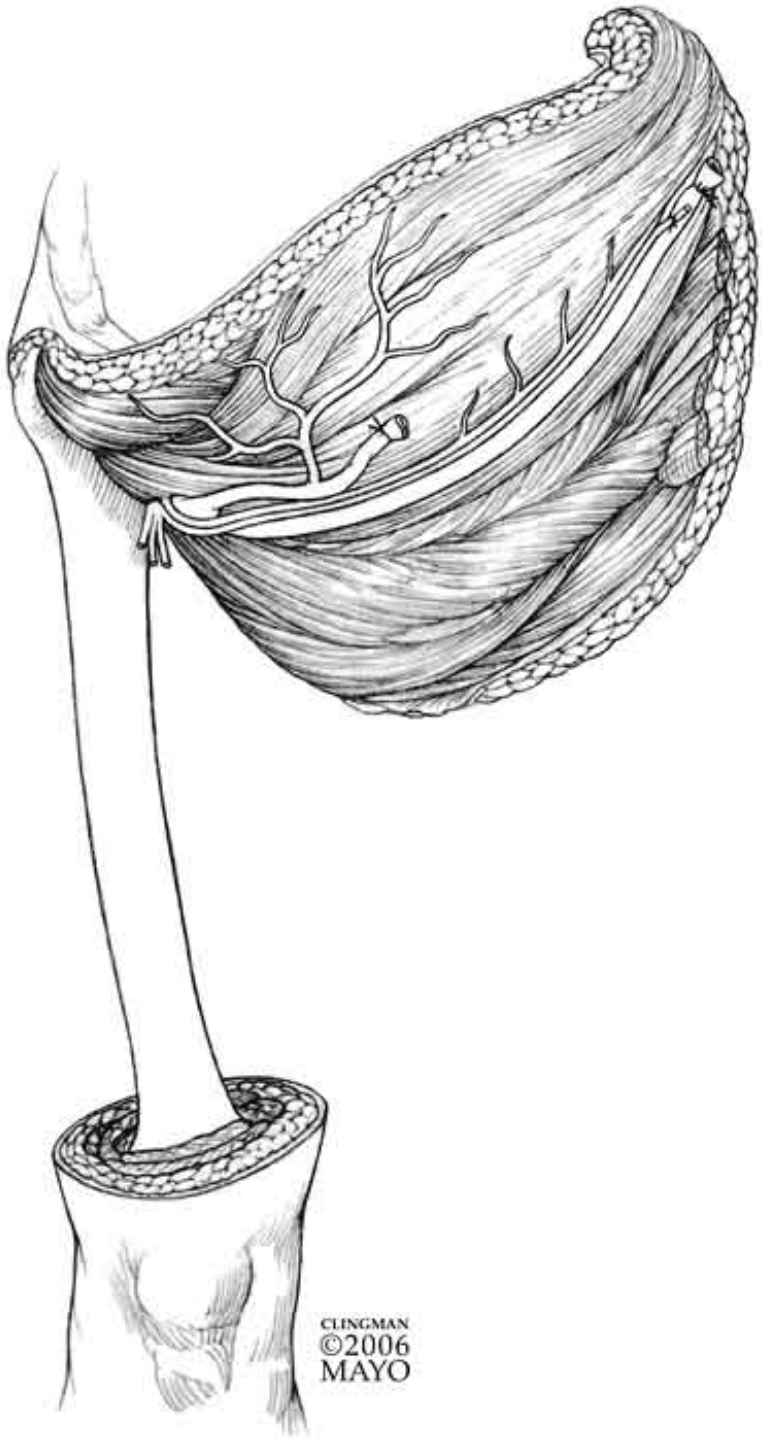
CLINGMAN
©2006
MAYO

FIGURE 35-4 Total thigh fillet flap utilizes the majority of the thigh musculature.

For these reasons, external hemipelvectomy itself is the ultimate solution to soft tissue coverage problems. The difficulty arises when common iliac vessels have to be ligated for oncologic reasons, which precludes creation of the anterior hemipelvectomy flap, and, at the same time, tissues of the buttock are not suitable for posterior flap design due to tumor involvement or sequela of previous operations or radiation therapy. In the past, such patients were considered unresectable. With the advent of microvascular tissue transfer however, a suitable and tumor-free block of tissue can be recovered by ex-vivo anatomic exploration of an amputated extremity and transferred as a free flap to achieve coverage of the hemipelvectomy defect.

Internal Hemipelvectomy Reconstruction

Internal hemipelvectomy involves total or partial removal of the innominate bone with preservation of the lower extremity. This operation provides good local tumor control and acceptable functional outcome. Several types of internal hemipelvectomy as proposed by Enneking and Dunham exist based on the part of innominate bone resected: Type 1â€”ileum; Type 2â€”periacetabular region; Type 3â€”pubic bone; and, described by some, Type 4â€”ileum and sacral ala (Fig. 35-6). This classification

P.457
P.458
P.459

provides the basis for the surgical oncologic approaches to malignant tumors of the nonvisceral pelvis.

A

B

C

D

FIGURE 35-5 A: Total thigh fillet flap provides abundance of well vascularized soft tissue. B: The flap provides excellent contour and adequate posterior reach. C,D: Postoperative result following extended hemipelvectomy for the tumor of the buttock.

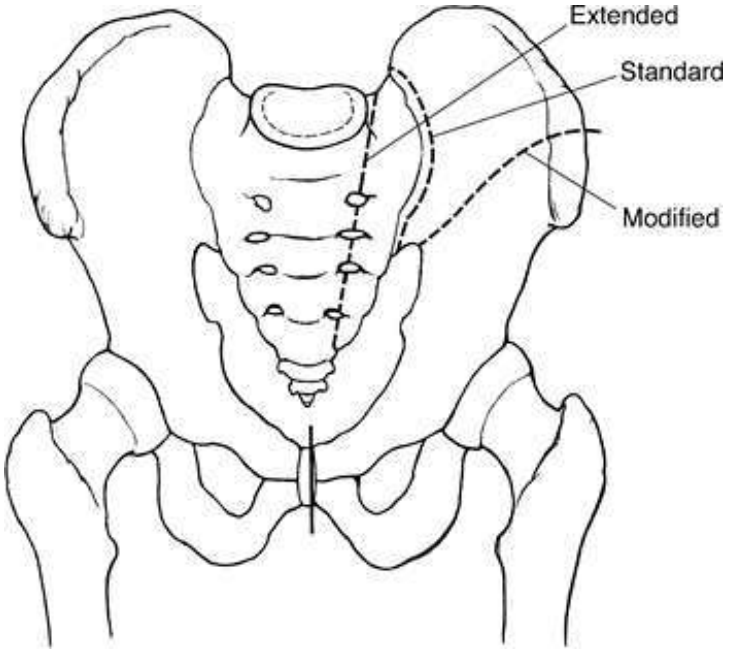

FIGURE 35-6 Schematic drawing of the pelvis showing osteotomy lines for standard modified and extended hemipelvectomies.

Introduction of internal hemipelvectomies increases requirements for soft tissue reconstruction since the pelvic volume is not reduced and lower extremity cannot be filleted. Additionally, a large amount of the dead space is created and often requires obliteration with a flap, especially in the settings of preoperative radiation. Several regional pedicled flap options for filling these defects exist (Table 35-4). Muscle flaps and de-epithelialized musculocutaneous flaps can be used effectively for dead space obliteration. De-epithelialized flaps create more bulk and are used for filling larger defects. Because fat is a part of the skin paddle,

the flap retains volume long term. In addition to dead space obliteration, these flaps bring blood supply into relatively ischemic radiated tissues of the operative site and thus enhance stable coverage and promote wound healing. Although flap demands for bony sarcomas without soft tissue extension are not high, however, in the cases of soft tissue sarcomas or soft tissue involvement from the bony sarcomas, flap reconstruction becomes critical for the coverage of exposed skeletal structures. Musculocutaneous, muscle pedicled, or free flaps can be used. When the skin is not part of such a flap, a split-thickness skin graft can be used for coverage of exposed muscle of the flap.

Sacral Reconstruction

Primary sacral tumors are often low-grade lesions such as chordoma. They do not metastasize, but have a tenacious local course. Chordomas and locally-advanced rectal cancers are the two most commonly encountered sacral tumors. Sacral surgery presents a major challenge due to regional anatomic complexity, technical difficulty in obtaining clear margins, functional impairment, and an often prolonged postoperative course due to poor wound healing. Sacrectomy is a procedure that is uncommon outside of specialized cancer centers. Although distal sacral resections are safely performed through the posterior approach, larger tumors, tumors of rectal origin, and resections proximal to S3 require an initial abdominal exploration to dissect visceral and neurovascular structures, perform formal visceral resection when required, assure hemostasis, and complete anterior sacrectomy dissection. Sacrectomy is completed through the posterior approach.

Soft tissue reconstruction becomes important in preventing postoperative wound complication following sacrectomy. Two flaps are commonly utilized: pedicled omentum and rectus abdominis muscle/musculocutaneous flaps. Omentum is preferred for smaller defects (the defects *smaller than the surgeon's fist*), and the rectus flap is used for larger ones. Inferiorly based vertical rectus abdominis musculocutaneous (VRAM) flap is the most commonly used reconstructive technique.

The flap accomplishes two main objectives: it obliterates large dead space and provides tension-free closure of the sacrectomy wound. In the cases of preoperative radiation, well-vascularized tissue of the flap enhances local

circulation and further promotes wound healing. Implementation of this practice in our institution led to a marked decrease in sacrectomy wound complication rates. Careful preoperative planning and communication between different services is critical in executing these operations. VRAM flap as a reconstructive option must always be borne in mind in the light of previous abdominal incisions as well as ostomies, drains, and feeding tube placements. The VRAM flap represents a very important surgical technique in oncologic pelvic reconstruction.

If soft tissue defect is too large for sole VRAM coverage, fasciocutaneous V-to-Y advancement flap provides additional recruitment of the local tissues. If aforementioned techniques are not sufficient to achieve an immediate coverage of sacrectomy-buttockectomy defect, microvascular tissue

transfer should be performed with a free flap of appropriate dimensions. Finally, anterior flap external hemipelvectomy is reserved as a last resort reconstructive operation.

Table 35-4. Flap Options for Internal Hemipelvectomy Defect Reconstruction

- Rectus abdominis muscle based are inferiorly based ipsilateral or contralateral:
    - Vertical rectus abdominis musculocutaneous (VRAM) flap
    - Transverse rectus abdominis musculocutaneous (TRAM) flap
    - Rectus abdominis muscle flap with or without skin graft
- Rectus femoris muscle flap
- Latissimus dorsi muscle or musculocutaneous flap
- Vastus lateralis muscle flap
- Tensor fascia lata muscle or musculocutaneous flap
- Anterolateral thigh flap

Technical Elements of Pelvic Flap Surgery

Omentum Pedicled Flap

Omentum is readily available during laparotomy although it can also be harvested laparoscopically. Its blood supply is based on the right or left gastroepiploic artery. Significant mobilization and extent of reach can be gained by basing the flap on left gastroepiploic vessels, dividing the short gastric vessels along the greater curvature of the stomach, and further dividing middle omental artery to release a long vascular pedicle (Fig. 35-7). The latter maneuver releases omentum to its fullest length, which may be important for the reconstruction of the larger pelvic defects in the event that the VRAM flap is unavailable.

Rectus Abdominus Muscle and Musculocutaneous Flaps

The rectus abdominus muscle can be elevated alone or with a skin paddle. The skin paddle may be oriented transversely or vertically.

The transverse rectus abdominis musculocutaneous (TRAM) or vertical rectus abdominis musculocutaneous (VRAM) flaps can be utilized as either pedicled or microvascular free flaps. The rectus abdominus muscle has a dual blood supply from the superior and inferior epigastric vessels. For the purposes of pelvic soft tissue reconstruction, the rectus abdominus muscle flaps are based on the inferior epigastric vessels when pedicled to cover the pelvis or sacrum.

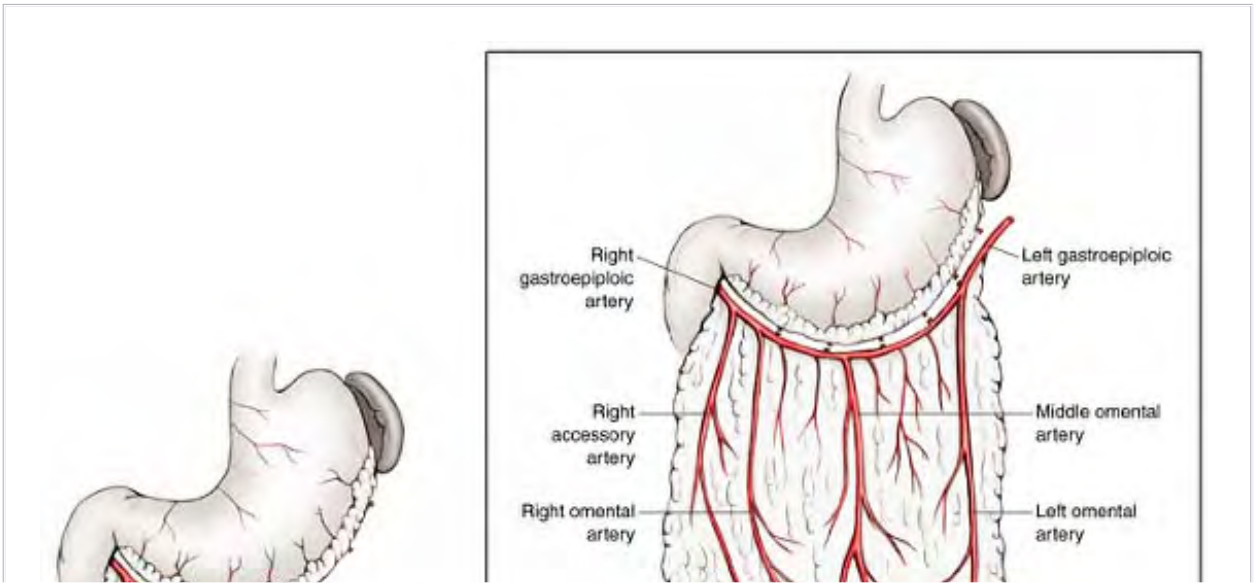

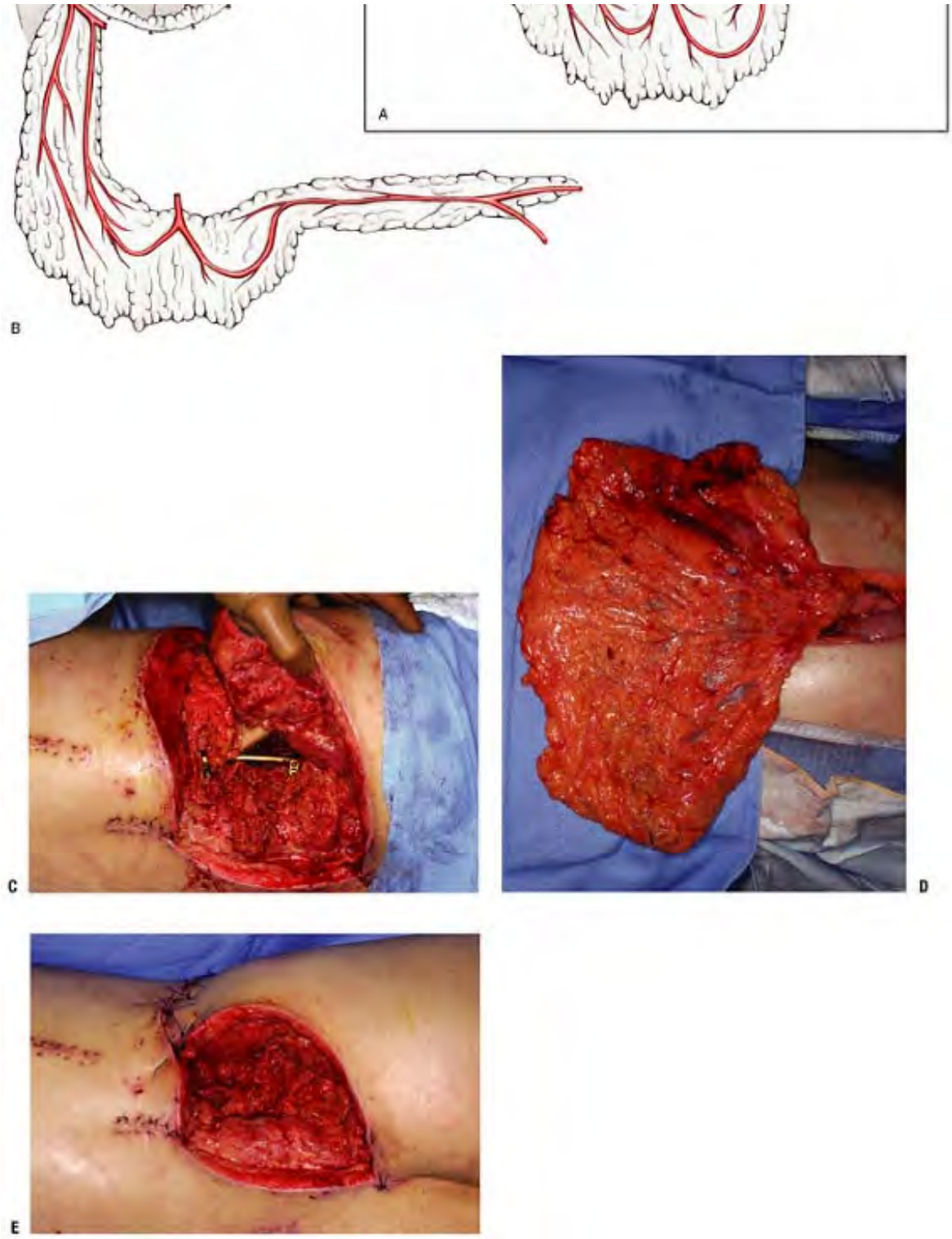

FIGURE 35-7 A: Omentum can be used for coverage in difficult situations following pelvic resection and soft tissue deficit. While the omentum lacks

significant structural strength it provides a well vascularized bed which can support skin grafting. The division of the left gastroepiploic arch allows the omentum to be mobilized off the greater curvature of the stomach. The right gastroepiploic vessels are preserved to supply the flap. B: The omentum may be lengthened by dividing the omental arcade along the lines within the illustration, creating a long vascularized pedicle. Câ€“E: Clinical example of a 16-year-old woman with exposed spinal hardware (C) following sacrectomy. The omentum was harvested through a midline anterior incision (D), lengthened as shown in (B) and then passed through the back to provide coverage for the spinal hardware.

Several key elements of preoperative planning are important in preparation for this procedure. A detailed history of previous intra-abdominal surgery needs to be obtained, because the inferior epigastric vessels may be divided during operations such as appendectomy, inguinal hernia repair, C-section, and colostomy creation (Fig. 35-8). The pedicle may also be damaged by radiation therapy. If there is any question about the integrity of the vascular pedicle, duplex evaluation should be performed.

Flap elevation is performed prior to abdominal exploration. We open the rectus sheath close to midline and dissect it off the rectus abdominis muscle (Fig. 35-8Aâ€“C). In this part of procedure, care should be taken to dissect inscriptions due to their proximity to the underlying vessels that run on the undersurface of the muscle. Once the flap has been dissected for coverage of the sacrum and posterior defects, the flap is placed in a plastic bag or wrapped with towels and placed over the anterior

P.462
P.463
P.464

portion of the lumbar spine and sacrum. The abdominal portion of the procedure is then completed and the abdomen is closed. If a large fascial defect remains following VRAM harvest, this is reconstructed with synthetic mesh or allograft fascia. Following abdominal closure, the patient is placed in the prone jackknife position for the sacral portion of the operation, and the flap is easily visualized

after removal of the specimen (Fig. 35-8Dâ€“F). Inset should ideally be performed with no tension to prevent sacral closure breakdown when the patients starts to sit.

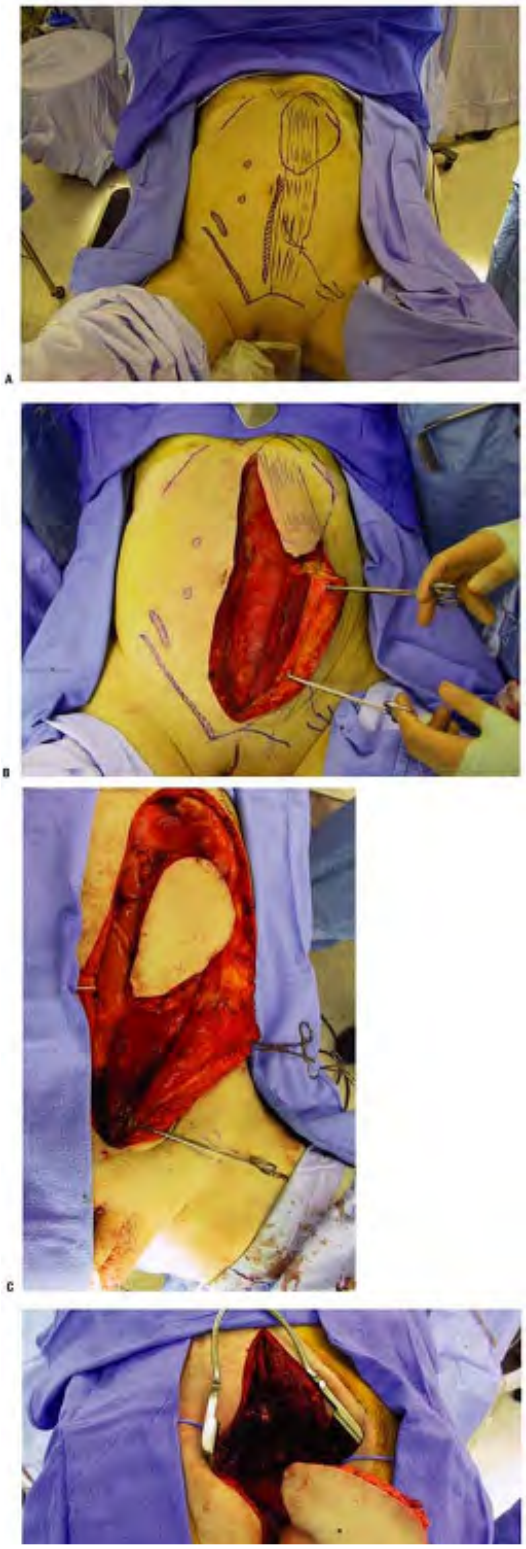

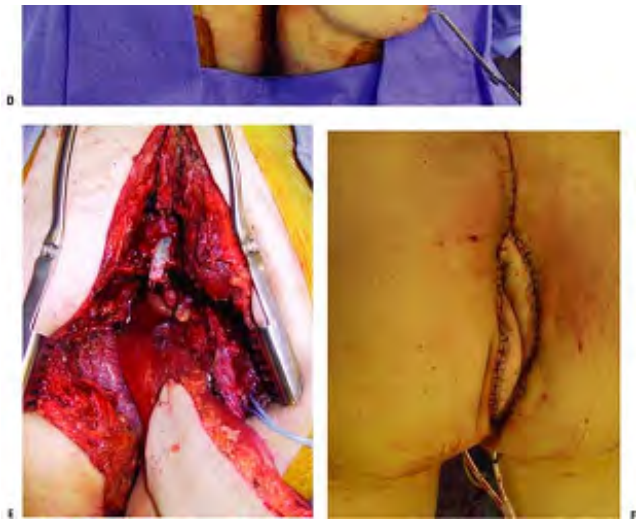

FIGURE 35-8 A: VRAM flap is planned for reconstruction of the defect from *en bloc* abdomino-perineal resection with sacrectomy. Skin paddle is designed to overlie rectus abdominus muscle. B: The rectus sheath is open and the rectus abdominis muscle is exposed. The anterior fascia is preserved with the skin paddle. C: The flap elevation is complete. The rectus abdominus muscle has been divided superiorly from its insertion into the ribs. The inferior epigastric vessels have been mobilized to prevent kinking as the flap is passed into the abdomen. D: The flap is placed into the abdomen and the anterior incision is then closed. The patient is then placed into the prone position and the sacrectomy is completed. At this point the VRAM flap is retrieved from the pelvis. E: Prior to its inset the muscle and vascular pedicle are examined to ensure there is no evidence of kinking or twisting. Note exposed dural sack. F: VRAM inset is complete with generous amount of skin remains to accommodate for sitting.

Other Flap Options

In many instances, the safest approach to patients with pelvic sarcomas is a laparotomy and, when it is performed, omentum becomes a valuable reconstructive material. Omentum, if present, should always be interposed between hemipelvectomy flap and intra-abdominal viscera. This provides an additional

protective layer in the event of hemipelvectomy wound problems. Omental flaps can be effectively used for dead space obliteration.

Although omentum and VRAM are truly the workhorses of pelvic reconstruction, several other flaps may become important as second line reconstructive options. Rectus femoris, tensor fascia lata, and anterolateral thigh flap can be used as pedicled flaps. Latissimus dorsi and anterolateral thigh flaps can be transferred as microvascular flaps depending on soft tissue requirements and availability of recipient vessels.

V-to-Y Advancement Gluteal Flaps

This versatile reconstructive technique can be performed in escalating complexity and is usually tried in the following sequence:

- Unilateral fasciocutaneous
- Bilateral fasciocutaneous
- Unilateral musculocutaneous in conjunction with contralateral fasciocutaneous
- Bilateral musculocutaneous
- VRAM and bilateral fascio- or musculocutaneous advancement flaps (Fig. 35-9)

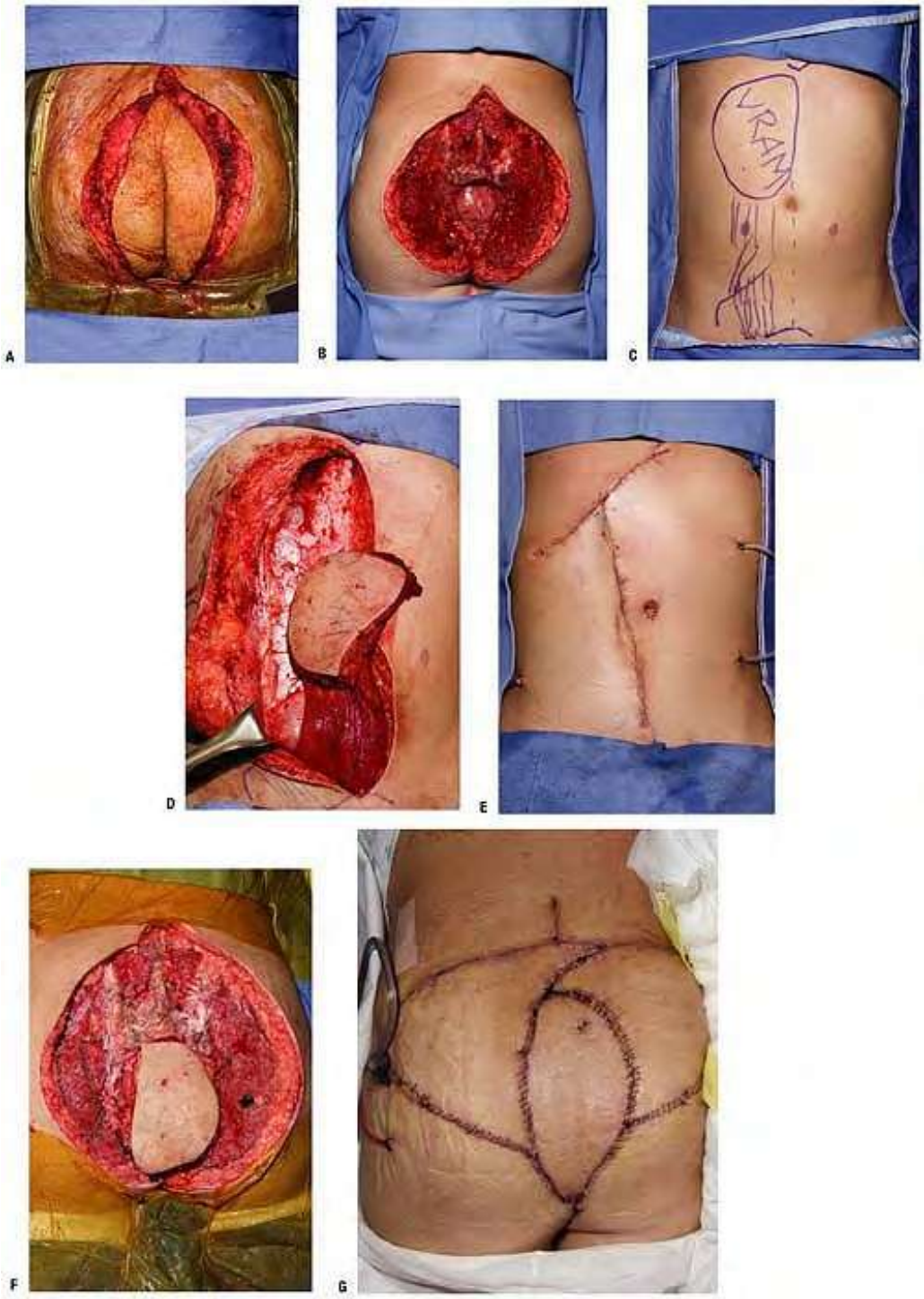

FIGURE 35-9 A,B: Combination of inferiorly based pedicled VRAM flap and bilateral fasciocutaneous V-to-Y advancement flaps for closure of a large sacral defect after wide local excision of the sacral sarcoma. Câ€“E:

Inferiorly based VRAM is elevated with large skin island based over superior portion of rectus abdominus muscle. The abdomen is closed with the aid of extensive undermining. The umbilicus is repositioned beneath the skin flaps to lie in the midline. F: VRAM flap allows for soft tissue coverage over sacrectomy defect and allows for reconstruction of pelvic floor; however large soft tissue defect remains at margins of sacrectomy incision. G: Remaining portion of sacral wound is closed with use of bilateral fasciocutaneous flaps. Flaps are elevated with fascia only to avoid injury to gluteus muscles to maximize post-operative ambulatory ability.

The surgeon must be careful in utilizing this flap in an ambulatory patient as it may affect functional performance. It is safer to reserve musculocutaneous gluteal flaps for paraplegic individuals. For large buttock defects, lateral donor sites of the V-to-Y advancement flaps sometimes may need to be temporarily covered with Vacuum-Assisted Closure (VACÂ®) dressings to be closed later when the edema subsides, skin is grafted, or allowed to heal by secondary intention.

In the face of insufficient amounts of local tissues, the donor part of the V-to-Y advancement flap can be skin grafted or covered with a wound VAC to allow healing by secondary intention.

Abdominal Wall Reconstruction in Pelvic Surgery

Integrity of the abdominal wall has to be maintained to prevent postoperative hernias. For a standard external hemipelvectomy defect without extensive resection of the abdominal wall, in the patients with musculocutaneous design of hemipelvectomy flap, no specific reconstruction of the abdominal wall is necessary other than meticulous layered closure of the hemipelvectomy wound. We did not observe postoperative hemipelvectomy hernias under these circumstances because sturdy fascia of the anterior compartment of the thigh or gluteus muscle along with muscular bulk of the flap effectively withstands pressure of the intra-abdominal viscera. Conversely, abdominal wall reconstruction becomes important for internal hemipelvectomy defects because frequently in these patients there is a very clear area of weakness in the lower abdominal wall. This situation is also aggravated by

higher functional level of the internal hemipelvectomy patients.

Reconstructive efforts are guided by the extent of the abdominal wall deficit and degree of operative contamination. Time-tested surgery principles of ventral hernia repair apply. Hernia repair should be tension free with liberal use of prosthetic mesh materials. Well-incorporated prosthetic mesh is superior to nonneurotized autologous options because it secures the dimensions of the abdominal

wall and prevents postoperative bulge and recurrent hernia formation. The downside of prosthetic mesh reconstruction is its propensity to infection in the presence of intraoperative contamination. Although prosthetic mesh can be used in clean-contaminated cases, most surgeons are reluctant to use nonabsorbable mesh in this setting and prefer either absorbable mesh such as Vicryl or a biologic substitute such as AlloDermÂ®, processed cadaveric human acellular dermis (Life Cell Corporation, Branchburg, NJ, USA) or SurgisisÂ®, a product obtained by the processing of porcine small intestine sub-mucosa (Cook Surgical, Bloomington, IN, USA). VicrylÂ® mesh is a temporary abdominal reconstruction solution that permits a conversion of the contaminated situation into a clean one at the time of the second procedure. The latter aims for definitive reconstruction, but requires utilization of costly biologic materials and postoperative formation of abdominal wall hernia or bulge continues to be a problem.

Postoperative Management

Postoperatively, patients usually require an ICU admission after major pelvic resections or hemipelvectomies. They are commonly kept on mechanical ventilation overnight until intravascular volumes are stabilized and gas exchange is adequate. To optimize postoperative recovery as well as systemic and flap perfusion the patient needs to be kept well hydrated. Adequate urine output needs to be maintained. The patient with pelvic or sacral flap is usually kept on a Clinitron bedÂ® to avoid pressure injury to the flap. Recovery of bowel function, ostomy care, DVT prophylaxis, and wound care is similar to that for a general surgery patient undergoing abdominal procedures.

Physical and occupational therapy services are involved early in the care of these patients. Even when the patients are on bed rest, range of motion exercises on

nonoperated extremities and activities of daily living should be started. When it is safe, from a surgical perspective, to resume ambulation, the patients are evaluated for short-term rehabilitation placement.

Results

Our experience shows that the outcome of external hemipelvectomy wounds was not dependent on hemipelvectomy flap selection. Reliability of its blood supply determines the success of the reconstruction. Historically, classic hemipelvectomy flaps that are based on the random circulation are exceedingly prone to necrosis. Preservation of the gluteal muscle as a part of the flap decreased, but did not solve, the problem of hemipelvectomy flap necrosis. We also observed that preservation of the branches of the internal iliac vessels providing direct blood supply to the gluteus maximus muscle further decreased hemipelvectomy flap necrosis rates. Hemipelvectomy wound outcomes therefore can be improved by increased utilization of microvascular fillet flap transfer that provides superb blood flow to the transferred tissue, not comparable with the periphery of a random hemipelvectomy following common iliac vessel ligation.

Complications

Wound complications such as surgical site infection and skin flap necrosis are the most common postoperative complications in musculoskeletal oncology. There are multiple reasons for the high rate of wound complications following major resections of musculoskeletal malignancies. Despite the fact that the vast majority of these procedures are clean cases with only occasional visceral resection that is applicable for pelvic tumors, wound complication rates are well beyond what one would expect. In modern musculoskeletal oncology practice, many sarcoma patients undergo neoadjuvant (preoperative) radiation and chemotherapy. Soft tissue complication rates approach 40% to 50 % in this group of patients. Moreover, even without neoadjuvant therapy, the duration and operative extent of the procedures significantly influence rates of postoperative wound complications. As noted earlier, wound complications that delay the administration of postoperative adjuvant therapies may have serious oncologic consequences. Rapid, uncomplicated wound healing is essential for patients who require postoperative adjuvant therapy, highlighting the need for early and close collaboration with

plastic surgeons.

A hematoma can occur after any oncologic surgery, ranging from incisional biopsy to extensive surgery. All tumor resections should utilize at least one suction drain, although larger resections may require several drains placed deep and in more superficial tissues. The pressure from the collection

of blood products in a wound can compromise soft tissue reconstructions and may become infected. Drain sites and tracks are contaminated and usually are resected at the time of definitive tumor removal. For this reason, drains should exit the skin in line with the longitudinal incision used for biopsy. It is important to remember that drain placement can be as important as where one makes an incision. Drains should come out in line and close to an incision.

Hematomas are especially problematic following incisional biopsy, potentially compromising future resection and successful local control of a tumor. A hematoma can spread neoplastic cells beyond the immediate area and even into surrounding compartments if a transverse incision is used. Rigorous attention to hemostasis is therefore essential. Compressive dressings bolstered with plaster can also help minimize hematomas in the immediate postoperative period.

Seromas can occur following tumor resections, again potentially compromising soft tissue reconstructions and wound closures. Most surgeons therefore take a conservative approach to removing drains, allowing output to decrease (to less than 30 mL/day) and remain low before removal. Drains usually remain in place for a minimum of several days, but often are left in for much longer in large or irradiated wound beds. As with other facets of patient care, good communication with plastic surgery is important regarding the timing of drain removal. It is recommended that patients going home with suction drains remain on oral antibiotics.

Oncology patients often have a compromised nutritional status, making them prone to wound infections. Moreover, these patients are subject to the immunosuppressive actions of chemotherapy agents and the local wound effects from radiation. As many oncologic reconstructions rely on allografts, endoprostheses, or combinations of the two, surgeons should approach wound infections extremely aggressively. This approach entails a low threshold for surgical

dÃ©bridement. As with all surgical procedures, orthopedic oncology patients should receive pre- and perioperative intravenous antibiotics as well as postoperative coverage when appropriate.

Sacrectomy is prone to wound complications that are related to the extent of the procedure, intraoperative contamination, use of preoperative radiation therapy, local tissue ischemia, hematoma, positional pressure, dead space, and tight closure. In the setting of preoperative radiation, wound complication rates are as high as 50%. These factors can be minimized by meticulous surgical technique, obliteration of the dead space with a flap, and postoperative use of a Clinitron bedÂ®.

Over the past 20 years, 160 external hemipelvectomies were performed in our institution. External hemipelvectomy has been associated with high morbidity but low mortality (5% to 7%). Overall, 54% of patients had at least one complication. Hemipelvectomy wound morbidity was the most common postoperative complication. Thirty-nine percent of patients experienced wound infection and 26% had hemipelvectomy flap necrosis. Wound complications were managed with serial dÃ©bridement until control of the wound was achieved; however, this may result in a sizable defect.

In a delayed reconstruction setting, the amputated extremity is no longer available for tissue procurement and the contralateral pedicled VRAM flap becomes critical in closure of such defects. This flap provides a superb reconstruction and in fact is the flap of choice for postoperative hemipelvectomy wounds. In a very rare circumstance, when VRAM is unavailable and the hemipelvectomy wound is so large it cannot be closed by local tissue rearrangement or a skin graft, a free flap such as a contralateral anterolateral thigh flap or latissimus dorsi may be required. One has to consider a paucity of recipient vessels and the need for vein grafts or saphenous arteriovenous loop that significantly increase the risk of flap failure.

Oncologic outcomes of musculoskeletal tumors are largely dependent on tumor pathology. Limb salvage is possible in the vast majority of sarcoma patients. In those patients who still require proximal amputations such as external hemipelvectomy, stable soft tissue coverage is almost uniformly achieved. Despite postoperative wound complications related to the operative extent of the resection, intraoperative contamination, and preoperative radiation therapy, the

vast majority of the patients heal the surgical wounds. Both successful wound healing and high rates of limb salvage became possible due to advances in reconstructive surgery over the past three decades. In our practice, we emphasize early involvement of a plastic and reconstructive surgeon in care of a sarcoma patient.

Pearls and Pitfalls

- Plastic surgery consultation early
- Early, durable reconstructions are important for functional recovery and also prevent delays in administration of postoperative chemotherapy and radiation therapy
- All surgeons should be present in the operating room at the start of a combined case

- Donâ€™t use an Esmarchâ€”exsanguinate, if necessary, by elevating extremity and digital arterial compression
- Use compressive Robert Jones dressings when feasible, even after soft tissue resections

References

1. Senchenkov A, Moran SL, Petty PM, et al. Predictors of complications and outcomes of external hemipelvectomy wound: account of 160 consecutive cases. *Am Surg Oncol.* 2008;15(1):355â€“363. Epub 2007, Oct 23.

Recommended Reading

Abramson DL. Single-stage, multimodality treatment of soft-tissue sarcoma of the extremity. *Ann Plast Surg.* 1997;39:454â€“460.

Aflatoon K, Manoso MW, Deune EG, et al. Brachytherapy tubes and free tissue

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 36

Management of Major Upper Limb Amputation

Tsu-Min Tsai

Huey Y. Tien

Indications/Contraindications

Replanting the upper limb above wrist level involves a large quantity of muscle and, consequently, is regarded as major replantation. Because muscle does not tolerate anoxia well, the success of arm replantation depends on the effects of ischemia, and warm ischemic time is much more detrimental than cold ischemic time. The tissue damage caused by 1 hour of warm ischemia is equal to that caused by 20 hours of cold ischemia. When warm ischemic time is more than 6 hours, the success rate is markedly decreased and the complication rate is significantly increased. When arm amputation is incomplete, there might be a venous blood regurgitation that causes capillary refill to be present. In this situation, health care personnel may be unaware that the injured limb is devascularized, and therefore fail to keep the injured limb in a cool environment. This oversight further lengthens warm ischemic time and jeopardizes the survival of the injured limb. Furthermore, one should remember that on-table time to revascularization is often warm ischemia time, as well. Daigle and Kleinert reported that the average warm/total ischemia times were 4.8/14.8 hours in failed replantation, which is significantly higher compared with the average of 1.1/7.5 hours in successful replantation.

As a rule, the more proximal the amputation, the poorer the prognosis. Although some debate its efficacy in cases of amputation proximal to the mid forearm, the upper limb should be replanted if at all possible, both for function and cosmesis. However, there are some conditions that contraindicate arm replantation, including:

- Significant associated injury
- Extensive injury to the affected limb or to the amputated part
- Severe medical illness

Significant associated injury to major organs prolongs the patient's time in surgery and, therefore, increases blood loss and the need for transfusion. Limb replantation is clearly contraindicated if it may jeopardize the life of the patient. Injury factors that require significant consideration are:

- Avulsion versus guillotine injury
- Ischemia time elapsed
- Quantity of tissue loss
- Presence of multiple levels of injury
- Degree of wound contamination

Chronic illness also may complicate, and even contraindicate, limb replantation. For example, vessel anastomosis may be more difficult among patients with diabetes, who have a higher prevalence of arteriosclerosis. Diabetes, autoimmune disorders, and prolonged steroid use increase patientsâ€™ susceptibility to infection. Serious illnesses, including organ failure, preclude replantation surgery.

In addition to severity of injury to the limb and chronic medical illness, one also needs to take patient factors into consideration. Although there is no absolute cut-off age for replantation, the prognosis for a favorable functional outcome decreases with the increase of age. Children, as a rule, tend to be excellent

candidates for major upper limb replantation. They have an advantage over adults in tissue regeneration generally, and especially in nerve regeneration. Patient factors in addition to age include the patient's occupation, hobbies, and wishes.

Preoperative Planning

Of course, always treat any life-threatening associated injury first. As soon as possible, control bleeding on the amputated stump with pressure dressings. Do not use any clamp instrument to try to stop bleeding because this increases the chance of damaging the vital structures (e.g., nerve) next to the bleeding vessels. The cut arterial end, however, can be tied to stop bleeding. Preserve the amputated part in a plastic bag, and cool it in ice slurry to just above freezing, usually 4Â°C. Then arrange rapid transportation to the replantation center.

In cases of incomplete amputation, wrap the limb snugly with gauze and elastic bandage so that bleeding stops, but not so tightly that blood flow is cut off to healthy tissue. Then, splint the injured area and place a regular ice pack on it. Special attention should be paid for any sign of hypovolemia.

While the patient is prepared for surgery or during transportation, the amputated part can be perfused with organ preservation solution, for example, Tsai's solution or UW solution, as shown in Table 36-1. Perfusion may have several benefits over simple immersion for large tissue parts, including the physical benefits of more rapidly cooling deep tissues and flushing stagnant blood. If the ischemic time is more than 6 hours, and especially if the amputation is incomplete, the amputated part also can be perfused with 1 unit of heparinized arterial blood obtained from the patient (Fig. 36-1). When the amputated part is brought to the operating room for cleaning and initial debridement, continue to keep it in a cool condition at all times.

Surgery

The patient is taken to the operating room and placed on the operating table in the supine position. Because arm replantation surgery may take as long as 16 to 18 hours, general anesthesia is practically

always required. The patient is given a dose of intravenous antibiotics 30 minutes

before surgery. The patient's injured limb is prepared and draped in standard sterile fashion. An upper-arm tourniquet is inflated to 100 mm Hg above the patient's systolic pressure to minimize bleeding. If necessary, the contralateral lower limb is also prepared and draped for possible skin, nerve, and/or vein grafting; it may even be used to provide a free flap.

Table 36-1. Solutions Used in Arm Replantation

Solution	Content	Concentration	Preparation
Tsai's solution	Injectable lactated Ringer's, heparin, preservative-free Lidocaine		Lactated Ringer's 150 mL + heparin 3,000 units + 1% Lidocaine 30 mL
Papaverine	Papaverine	1.5 mg/mL	Papaverine 1 mL (30 mg) + normal saline 19 mL
LMD/heparin drip	LMD, heparin	10 units/mL	LMD 500 mL + heparin 5,000 units

LMD, low-molecular-weight dextran.

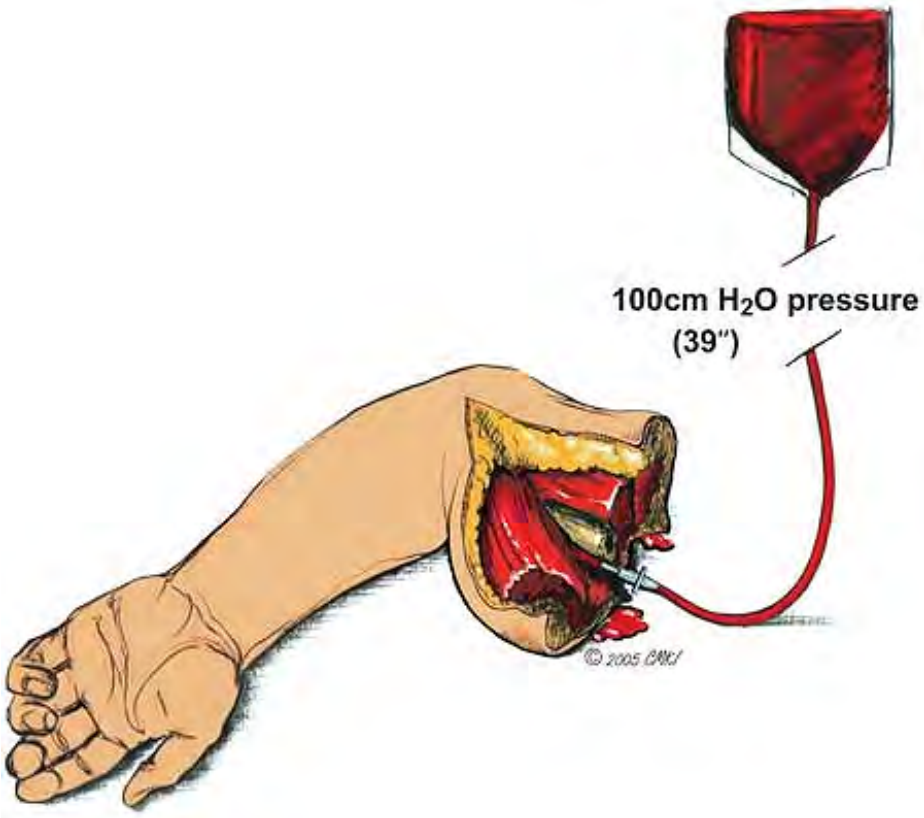

FIGURE 36-1 Perfusion of an amputated limb with heparinized blood. The blood bag is hung 100 cm above the amputated limb and the infusion flows by gravity.

The order for arm replantation is quite different from finger replantation. Before proceeding with replantation, lactated Ringer's solution or plasma is used to flush organ preservation solution from the limb. The recommended order of surgical procedures is as follows:

- Arterial shunt or perfusion to shorten anoxia time
- Thorough debridement
- Bone fixation
- Artery and vein repair, with or without vein graft

- Repair of nerves
- Repair of muscle and tendon
- Skin grafting or flap for coverage
- If indicated, fasciotomy

The artery from the amputated limb is cannulated with a standard vascular shunt to establish the connection between the proximal and distal artery and to intermittently perfuse the devascularized limb. A segment of regular intravenous infusion tube can also be used if no vascular shunt device is available. During perfusion, if the patient becomes hypotensive, temporarily clamp the perfusion. After 20 minutes, the venous shunt can be performed to minimize blood loss.

Thorough debridement is essential for a successful replantation. The injury may occur in an environment such as a farm, an automobile accident, or water sport in which the wound would be severely contaminated. Any crushed, grossly contaminated tissue must be removed. Muscles distal to the amputation level that have lost their contractility must be debrided. Effective debridement converts a dirty, crushed wound to a clean and guillotine-like wound. Residual dead tissue provides a substrate for bacterial growth after major limb replantation. The resulting gas gangrene or septicemia from postoperative infection may be life threatening. In addition, bone infection may result in non-union, delayed functional recovery, and secondary amputation.

Bone shortening is almost always the rule; usually a shortening of at least 2 to 3 cm is required. Bone shortening not only provides more healthy bony structure for an improved chance of bone

healing, but also helps to release tension on the soft tissue and eliminates the need for vein grafting, nerve grafting, or tendon grafting.

Either internal or external bone fixation can be used, depending on the surgeon's preference. Internal fixation can be chosen for more definitive treatment, especially if there is minimal concern for infection. Plates and screws provide better stability and are most commonly used. Kirschner wires can be used when the

amputation is around wrist level. Although intramedullary nailing is an option, it is rarely used. External fixation should be considered when there is a high risk of infection or when internal fixation is contraindicated. External fixation can be used for temporary fracture stabilization until soft tissue is healed. When the limb is viable, internal fixation can be implemented. Occasionally, however, external fixation can be a problem when free tissue transfer is required for soft tissue coverage. If this is the case, external fixation should be switched to internal fixation at the time of the free tissue transfer procedure or should be avoided from the beginning, if at all possible.

One should repair as many arteries and veins as possible, although there is no magic number for optimal results. In general, at least one artery and two vein anastomoses are required. The two-to-one ratio also holds when greater numbers of anastomoses are performed: Repair two veins for each arterial repair. The stump of the vessel on each end is gently debrided to remove crushed or severely contused vessel. Special attention is paid in cases of avulsion to make sure there is no hidden intima injury that might cause delayed thrombosis and, eventually, failure of replantation. The adventitia at the stump is then carefully trimmed to provide better exposure for anastomosis. A syringe with a small angiocath filled with Tsai's solution (150 mL lactated Ringer's + 3,000 units heparin + 30 mL 1% preservative-free Lidocaine) is used to flush the vessel lumen to remove any blood clot or debris (see Table 36-1). Vascular anastomosis is performed under a microscope with 8-0 or 9-0 nonabsorbable suture in interrupted fashion. For a more proximal amputation, for example, at the upper arm, anastomosis can be achieved with 6-0 or 7-0 suture. Clamps are removed after each anastomosis is completed, and the replanted part is observed for return of circulation. In the past, we would give a bolus dose of heparin (3,000 units) followed by a low-molecular-weight dextran (LMD)/heparin drip (500 mL LMD + 5,000 units heparin at 20 mL/hr) for anticoagulation. Nowadays, after the bolus of heparin, we use a low fractionated heparin (e.g., enoxaparin sodium injection or Lovenox 30 mg given subcutaneously) due to its efficacy and simplicity. If there is any sign of vessel spasm, the replanted part should be kept warm with warm normal-salineâ€“moistened gauze or abdominal pads. Occasionally, papaverine (1.5 mg/mL) can be used to reverse vasospasm (see Table 36-1).

If the gap of the arterial defect is too big for primary anastomosis without

increased tension, arterial reconstruction with a vein graft should be considered. The donor site of the graft can be a superficial vein from the contralateral arm or a saphenous vein from the lower extremity. The vein graft should be reversed routinely to prevent the valve effect. Venous anastomosis is performed in the same manner as arterial repair, but doesnâ€™t require reversal if a vein graft is used. As many veins as possible should be repaired.

Careful monitoring is necessary to avoid a systemic effect from revascularization (i.e., hyperkalemia, increased creatinine, and myoglobinuria). In complete amputation, the first 100 to 200 mL of venous blood, carrying with it toxic metabolites, is allowed to bleed out without going back to systemic circulation. For incomplete amputation, especially when part of the venous system is intact, the repaired artery is intermittently clamped to allow adjustment of systemic circulation.

All major nerves should be repaired, including the radial, median, and ulnar nerves and their branches. Epineurial repair with interrupted 8-0 or 9-0 nonabsorbable suture under a microscope is recommended. If a nerve graft is required, it can be obtained from the lateral or medial antebrachial cutaneous nerve or the sural nerve, depending on the length and size of the nerve graft desired.

Different techniques of tendon suturing can be chosen, according to the surgeon's preference. Strong tendon repair with early protective range of motion provides the best chance of functional recovery. If muscle and tendon are severely damaged, a tendon transfer is performed.

The authors recommend that fasciotomy be performed in anticipation of postischemic swelling. In the forearm, both the volar and dorsal compartments must be opened (Fig. 36-2). The incision on the volar distal forearm can be made along the ulnar side of the distal forearm to minimize tendon exposure. In the hand, all nine compartments (thenar, hypothenar, three volar, and four dorsal interosseous) are opened. This can be achieved with four incisions: one each on the side for the thenar and hypothenar eminences, and two dorsal incisions along second and fourth metacarpal bones in order to approach all interosseous compartments (Fig. 36-3).

Soft tissue coverage for exposed vital structures becomes easier after bone

shortening and, in some cases, no extra skin is required. In most cases, primary wound closure can be accomplished with skin

grafting only. Occasionally, especially in complicated amputations, additional soft tissue coverage is required. This can be achieved by forming a pedicled or free vascular flap. It can be done either in an immediate, emergent setting or in a semi-emergent setting (i.e., a few days after initial surgery).

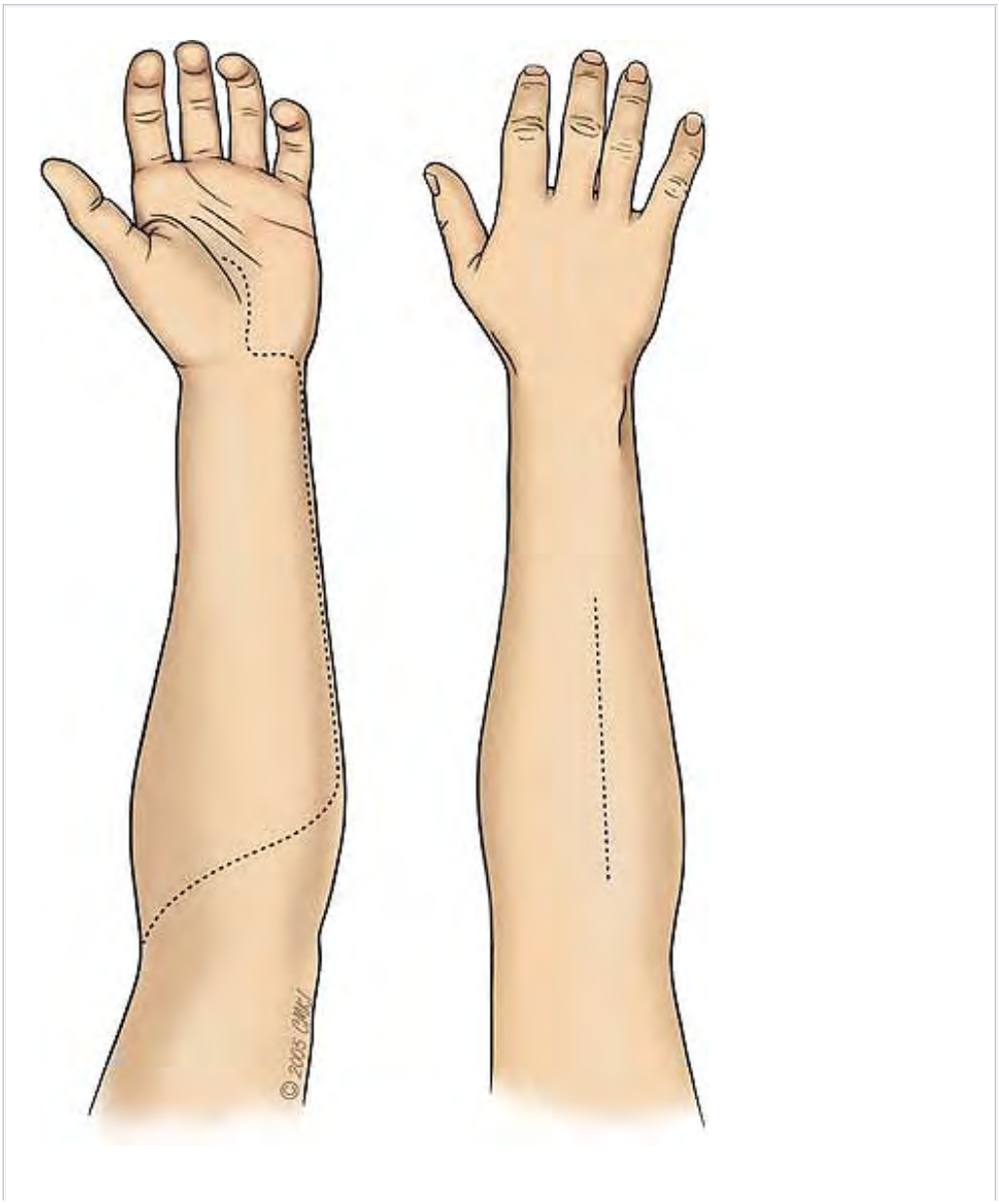

FIGURE 36-2 Volar distal forearm incision along the dorsal side.

Postoperative Management

The patient is admitted and is given nothing by mouth for at least 12 hours in case it becomes necessary to return to the operating room. The replanted limb is elevated to heart level with an intravenous pole or pillows and is monitored at least every hour for color, temperature, blanching, and capillary refill. Among them, temperature is the most sensitive parameter. A temperature drop of

more than 2Â°C represents an early sign of vascular compromise. It is also important to make sure that the capillary refill is the result of arterial flow, not venous congestion. The patient is instructed to refrain from smoking and caffeine products for 3 weeks after surgery because they may decrease circulation to revascularized tissue. Dressing change is avoided in the first 3 days to minimize the chance of vasospasm.

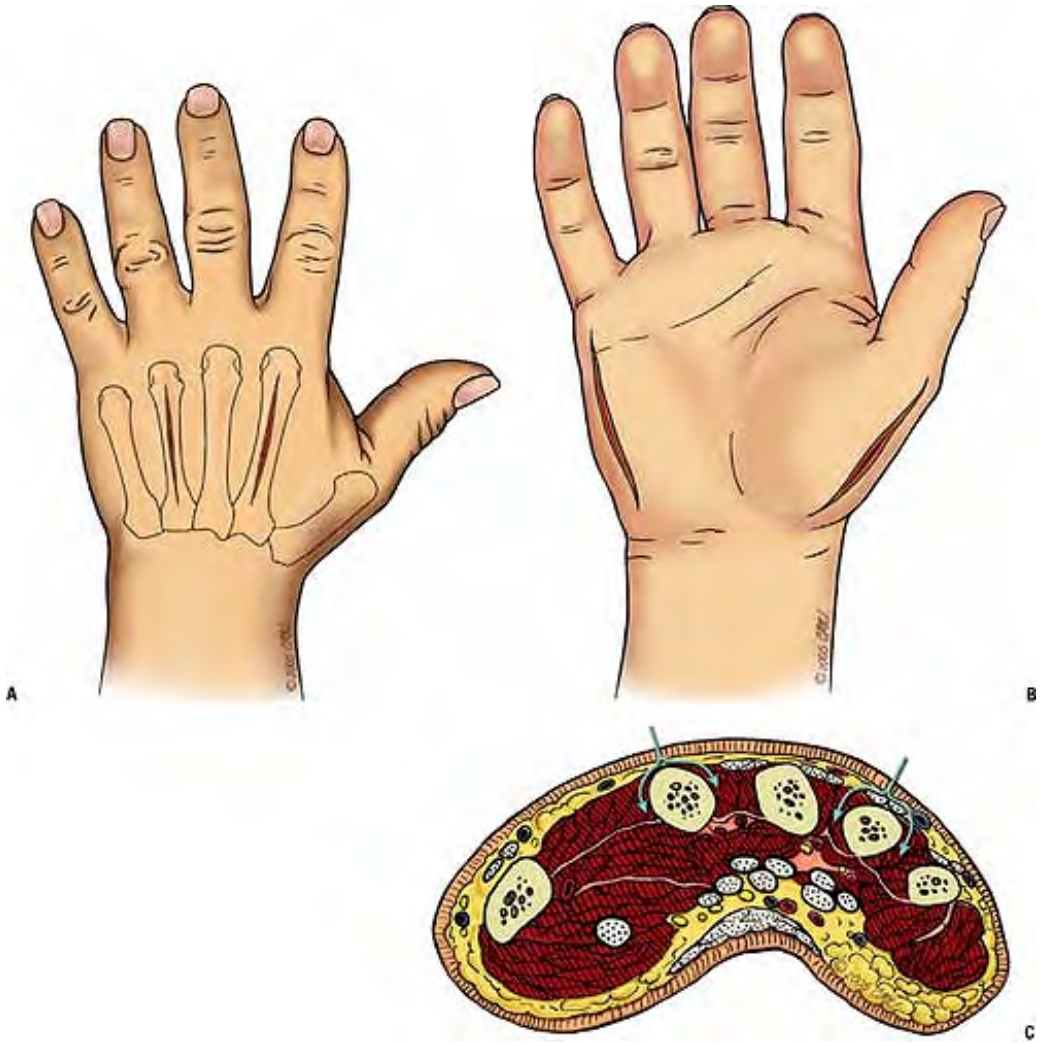

FIGURE 36-3 A: Dorsal incisions along the 2nd and 4th metacarpal bones. B: Incisions on the thenar and hypothenar eminences. C: Interosseous compartments.

By tradition, heparin and LMD are routinely given intraoperatively and postoperatively. However, a new trend is to give one dose of heparin when revascularization is completed or, sometimes, not giving any heparin at all. In these cases, patients are given only aspirin postoperatively. Heparin and LMD are given only when there is any doubt about the efficacy of circulation. Prostaglandin E_1 is also reported to provide some benefit because of its effects of general vasodilation, leukocyte adhesion inhibition, and platelet aggregation inhibition.

The dose given is 40 to 120 Âµg per day for 3 to 10 days.

Hyperbaric oxygen (HBO) treatment is very useful when the replanted limb shows marginal circulation. According to Nylander et al, HBO treatment significantly reduces phosphorylase activity, a sensitive marker for muscle damage, in the post-ischemia phase. Hyperbaric oxygen treatment is given at 2.5 atm of absolute pressure for 45 minutes, and usually three treatments are necessary.

In the weeks and months that follow replantation, additional surgeries may be required to improve the function of the replanted limb. These procedures may include debridement, tenolysis, and improvements in coverage.

Complications

Postoperative complications include persistent ischemia from the no-flow phenomenon, hyperkalemia/metabolic acidosis, myoglobinuria/renal failure, tissue necrosis, and wound infection. Any of these complications may either jeopardize the survivability of the replanted limb or endanger the life of the patient. The no-flow phenomenon is mainly due to long ischemic time and is often irreversible. Although free-radical scavengers (e.g., allopurinol) can be used to treat this condition, the result is often unsatisfactory.

During surgery, frequent test measurements should be used to monitor for the development of hyperkalemia and metabolic acidosis. High alertness and close monitoring of serum potassium during surgery, especially right after the vascular clamp is released, are key. The vessel should be intermittently clamped in the first couple of minutes of revascularization to allow time for adjustment. It's also important to maintain adequate hydration and avoid high potassium-content IV fluid. If it does develop, hyperkalemia/metabolic acidosis can be treated with alkalization and IV insulin. According to Waikakul et al, serum potassium concentration in the amputated segment was the best objective predictor of replantation success. When it is higher than 6.5 mmol/l 30 minutes after reperfusion, replantation should be avoided. A high systemic venous serum potassium concentration was also found before clinical signs of the reperfusion syndrome were seen.

Myoglobinuria and renal failure can be prevented with adequate hydration, alkalinization of urine, osmotic diuretics (mannitol), and free-radical scavengers. Frequent urine myoglobin monitoring is required until two consecutive results are within normal range. Tissue necrosis and infection are mainly due to inadequate debridement; therefore, adequate debridement is essential.

During the last 10 years, cases of myoglobinuria or renal failure are rarely seen following tissue perfusion or shunt anoxia time. Compared with minor replantation, patients with major replantation require longer hospital stays, more surgical procedures, and the functional results are not always satisfactory. The process also has a high psychiatric impact to both patient and family members.

Ultimately, the most important measure of outcome is the recovery of hand function. If there is no recovery of hand function, forearm amputation and a switch to below-elbow prosthesis can be considered. In one study from our institute, 8 of 34 patients with apparently successful major replantation underwent secondary amputation. Among them, four amputations were done because of poor function, one because of dysfunctional pain, and one because of poor cosmesis. The other two were part of planned staged procedures to preserve the length of the stump or joint. In general, a replanted limb tends to have a better functional outcome compared with a prosthesis. A replanted limb typically has better performance in gross prehensile activities and allows pronation and supination that prostheses do not offer.

In summary, arm replantation remains a major challenge to surgeons. The surgeon needs to consider the safety of the patient, to preserve limb viability, and to reconstruct the limb function. The final goal of a successful replantation is to achieve adequate function without pain. In addition, cosmesis must not be ignored, and secondary amputation sometimes is required. According to our experience, ideal conditions do not guarantee ideal results, although more distal amputation and shorter ischemic time predict a more favorable outcome.

Case Studies

A 45-year-old man sustained a blast injury resulting in total amputation at proximal upper arm level. Total ischemic time was 11 hours before revascularization was accomplished (Fig. 36-4).

A 17-year-old boy was involved in a mining accident and sustained bilateral forearm amputations (Fig. 36-5).

P.478
P.479
P.480
P.481

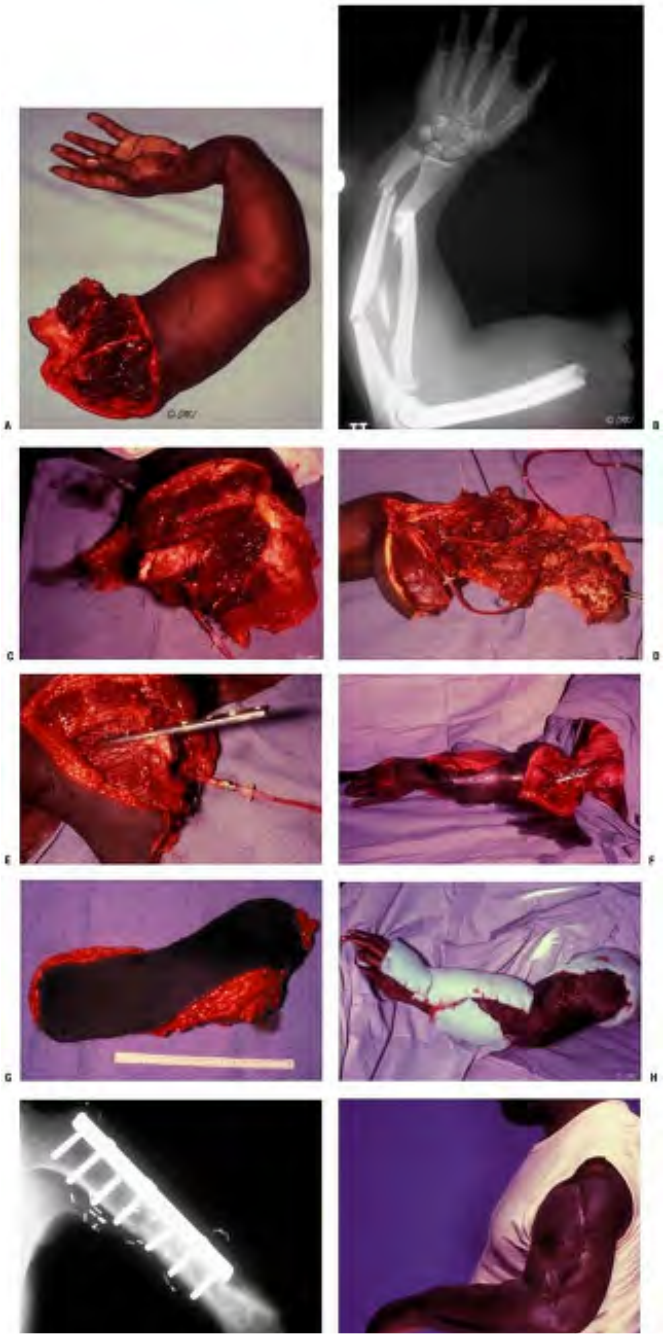

FIGURE 36-4 A: Amputated part demonstrated gross deformity at forearm level. B: Radiographic examination revealed amputation at proximal humerus level with segmental forearm fractures. C: Proximal part showed soft tissue defect with large area of bony exposure. D: Amputated part was perfused with 1 unit of heparinized arterial blood 2 hours before revascularization. E: While patient was getting ready for surgery, amputated limb was debrided and internal fixation partially readied for replantation. F: Completion of internal fixation, revascularization, and fasciotomy. G: Free tensor fascia lata flap for soft tissue coverage. H: Completion of initial surgery with free flap and skin grafting. I: Radiographic examination demonstrated healed humerus with plate in place. J: Final result after multiple surgical procedures including pectoralis major muscle transfer for elbow flexion. K: Final result of elbow extension. L: Final result of hand gripping. The patient has some hand gripping function but has very little intrinsic muscle function. Sensory recovery is protective only.

FIGURE 36-5 A: Right forearm amputated at mid-forearm level and left forearm amputated at proximal forearm level. B: Radiograph of amputated parts. C: Finger extension at final follow-up. D: Finger flexion at final follow-up.

Recommended Reading

Bandyk DF. Vascular injury associated with extremity trauma. *Clin Orthop Relat*

Editors: Moran, Steven L.; Cooney, William P.
Title: *Master Techniques in Orthopaedic Surgery: Soft Tissue Surgery, 1st Edition*

Copyright Â©2009 Lippincott Williams & Wilkins

Chapter 37

Replantation of Digits

Samir Mardini

Kuang-Te Chen

Fu-Chan Wei

Indications/Contraindications

Replantation is defined as reattachment and revascularization of a completely amputated part, whereas revascularization is defined as restoration of arterial inflow or venous outflow to or from a part that is incompletely amputated. In revascularization cases, it is of utmost importance that the true severity of the injury is not underestimated due to the grossly intact image of the injured finger. The introduction and improvements in microsurgical instrumentation and techniques have lead to successful revascularization and replantation of amputated parts as distal as the finger tip and pulp. Overall, the results of finger replantation are encouraging and have provided many amputation victims with extremely rewarding results.

Amputations of upper extremity digits are devastating for the patient who is initially stunned by the events of the trauma. Therefore, care is taken to discuss the situation in a gentle and thoughtful manner, considering that the patient has just lost a part of the upper extremity, and that this loss might be permanent if the reconstructive effort is not indicated or is unsuccessful. Generally accepted indications for finger replantation include amputations of the thumb in almost every situation, a single digit with the level of injury distal to the superficialis insertion, multiple digits, and any amputation in children (Table 37-1). The patient's desires have a strong influence on

the approach.

Success in replantation is based on the functional outcomes as judged by objective parameters such as strength, range of motion, and sensory function, and not on the survival of the replanted or revascularized parts alone; therefore, the potential outcome must be considered for each individual patient. The potential for a long operative procedure, immediate or late failure, the potential requirement of regional or distant tissue flaps to complete the reconstruction, and the possible need for multiple procedures even after survival of the replants should be included in the formula for decision making and discussed with the patient and his or her family. The location, mechanism of injury, number digits involved, level of injury, condition of the amputated part, condition of the hand, length of ischemia time, the patient's general health, age, and associated injuries, and the patient's desires dictate the course of action. One advantage of replanting any digit is prevention of neuroma formation. A thorough knowledge of the anatomy and adequate exposure of structures aids in attaining optimal results (Figs. 37-1 and 37-2).

The thumb is the most important single digit, and its replantation is the most rewarding since provision of a post for opposition, even with lack of motion at the metacarpophalangeal and interphalangeal joints, is significant. In avulsion cases, the defects of the vessels and nerves may be segmental and the use of vein grafts for the arteries and veins, and nerve grafts or vein grafts for the nerves, may be required. The long nerve stumps of the amputated thumb can be anastomosed to the dorsal sensory nerves in the first web space, to a transposed nerve from the index finger, or to the more proximally located nerve stumps using nerve grafts. Despite the fact that sensory return may only be protective in nature, the replantation of thumb avulsion injuries usually results in significant improvement in function.

Patients presenting with multiple digit amputations are candidates for replantation, and efforts are made to replant all fingers with potential for good functional outcomes. The surgeon uses his or her judgment in deciding the location where an amputated part should be replanted. Replanting a digit onto the stump of the thumb is the first priority, and the other amputated digits should be replanted with the idea that a better grip is achieved when the long, ring, and small fingers are replanted without

gaps between the digits, and better fine pinch is achieved when the index and long finger are replanted. Efforts are made to replant each amputated digit to its natural position; however, less severely damaged amputated digits may be better placed onto a stump that is less damaged and is located in a more functionally significant position. For example, in a case of multiple digit amputations with involvement of both the thumb and index finger, in which the amputated thumb has sustained irreparable damage, the index finger is replanted to the stump of the thumb. After the thumb position has been filled, the replantation should begin with the one most appropriate for the patient's daily activities and functional needs at work.

Table 37-1. Indications and Contraindications for Finger Replantation

Indications	Controversial Indications
Amputations of the thumb at any level Amputation of multiple digits Any amputation in children Single-digit injury in zone I	Single-digit amputation at a level proximal to the insertion of the flexor digitorum superficialis tendon Ring avulsion injuries Severe contamination
Contraindications	
Amputated parts that are severely crushed or damaged Multiple-level amputations Significant associated trauma and/or medical conditions	

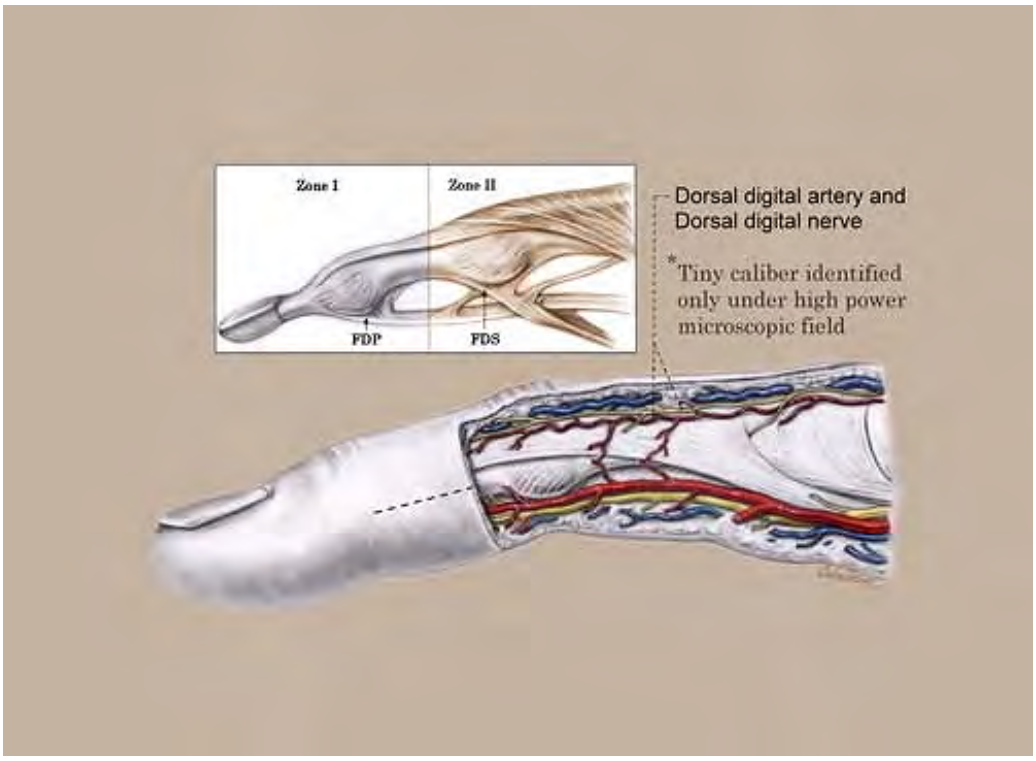

FIGURE 37-1 Lateral view of a finger demonstrating the location of arteries, veins, and nerves relevant to finger replantation. Also demonstrated are zones I and II as well as the level at which incisions are made for exploration of the neurovascular bundle.

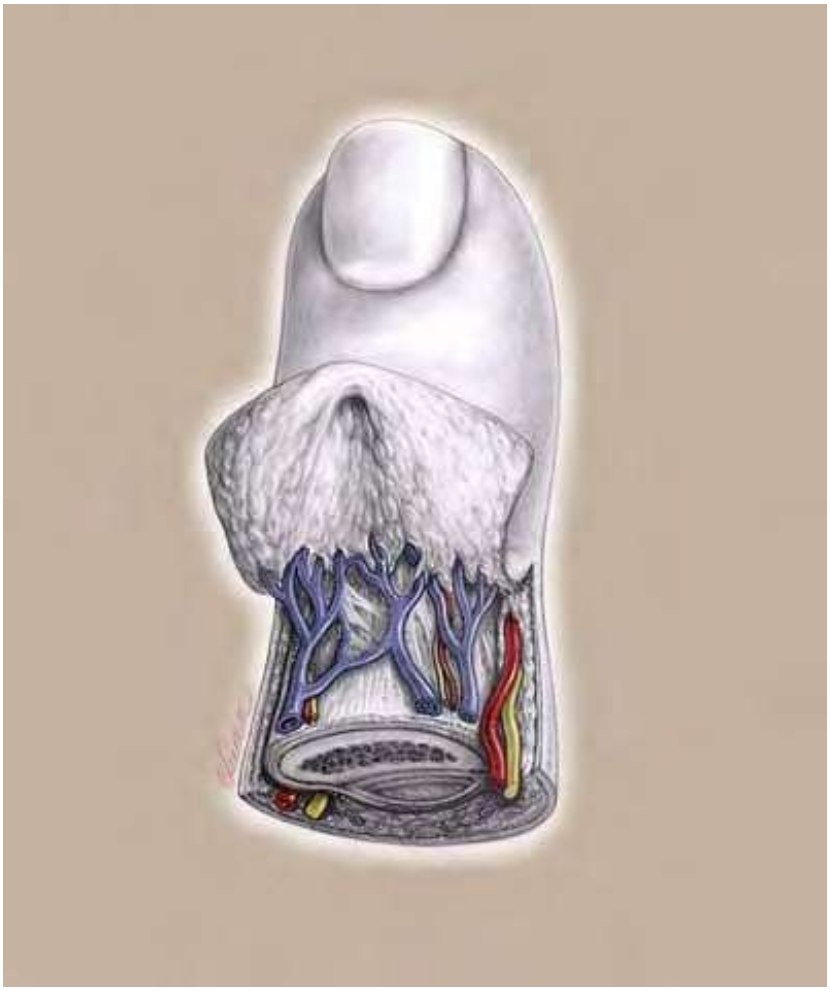

FIGURE 37-2 Mid-axial incisions and exposure of structures in the amputated part.

Single-digit amputations distal to the insertion of the flexor digitorum superficialis (FDS) tendon (zone I) yield good outcomes following replantation, with little need for secondary surgery, even if the distal interphalangeal (DIP) joint requires fusion. In some instances, venous outflow cannot be

reconstituted and postoperative leech therapy or heparinization and purposeful exsanguination can achieve digit survival. Four millimeters of dorsal skin proximal to the nail fold must be present for adequate vein size to be repaired. Tsai, McCabe, and

Maki described their technique of finger tip replantation, which used leech therapy and other methods of decongesting the finger when venous congestion developed postoperatively. These authors noted a 69% survival rate, a 75% rate of return of some form of two-point discrimination, and an average range of motion of the DIP joint of 56 degrees. Replantation of a single-digit amputation at a level proximal to the insertion of the flexor superficialis tendon and distal to the first annular pulley (zone II) often results in stiffness of the proximal and DIP joints and limited motion of the finger, with the replanted finger interfering with movement of the other fingers. Therefore, single-digit replantation may be only suitable for well-motivated individuals who are willing to undergo appropriate postoperative physical rehabilitation for a maximal functional recovery. Unless they are highly motivated, concerned about cosmetic outcome, or are children, patients are not usually replanted in this situation, as they often require secondary surgery and end up with disappointing outcomes.

Ring avulsion amputations present with segmental injuries of structures such as arteries, veins, and nerves; therefore, vein and nerve grafts are often necessary to bridge the gaps after debridement of unhealthy appearing neurovascular tissue. Otherwise, transposition of the neurovascular bundle from an adjacent digit (Fig. 37-3) or the use of arteriovenous shunting for inflow to the replanted digit can also yield successful outcomes. Most surgeons consider ring avulsion amputation patients to be poor candidates for replantation; however, replantation of an avulsed digit still can be attempted if the patient insists on replantation and he or she understands the likeliness of a suboptimal outcome.

The mechanism of injury plays an important role in decision making. Sharp, incisive injuries predict better functional outcomes than crush or avulsion injuries.

In elderly patients and patients with compromised medical conditions, the decision is based on the balance between the risk of undergoing the surgery, the functional outcome anticipated, and the significance of the replantation to the patient. Life-threatening injuries and significant medical conditions that preclude long anesthesia times are contraindications to digit replantation. Microsurgical replantation in children or infants is indicated and can yield good functional outcomes. In elderly patients, the surgeon should take into consideration the potential systemic insult from the anesthesia and operation, the arteriosclerosis of vessels, and the poorer potential of functional recovery.

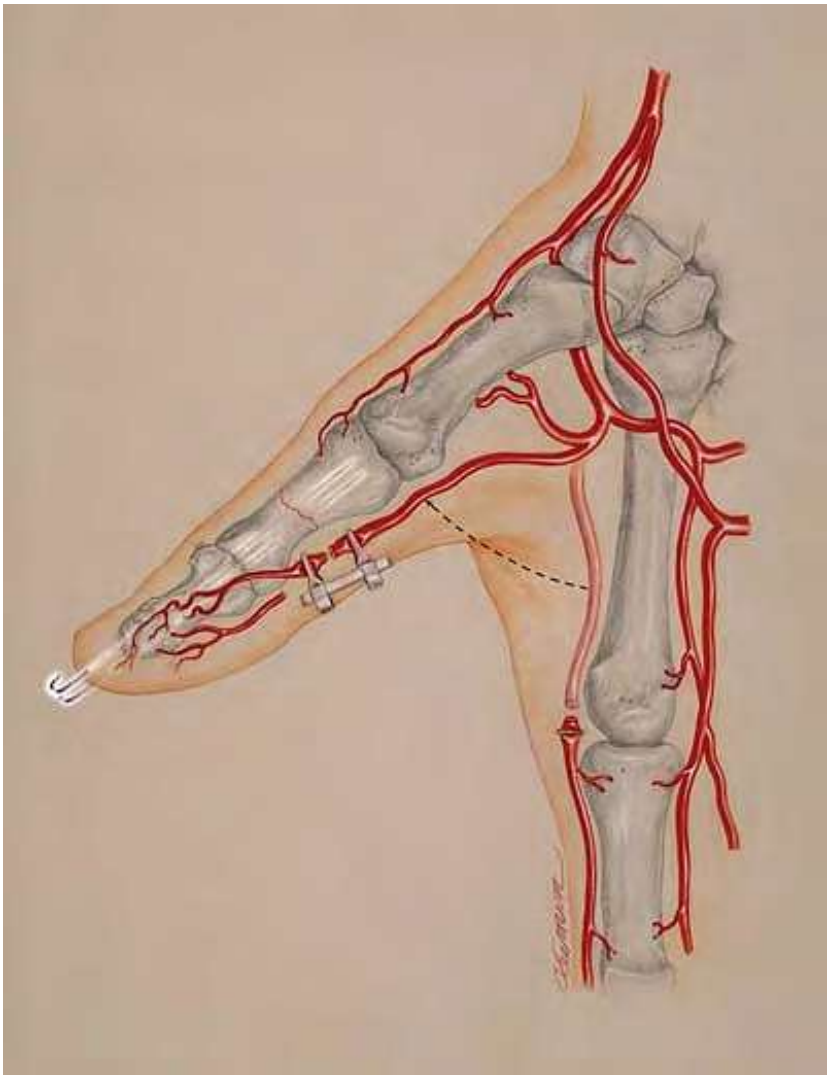

FIGURE 37-3 Transposition of an artery from an adjacent finger.

Warm ischemia times of less than 8 to 12 hours and cold ischemia times of less than 30 hours, measured from the time of amputation to the time of restoration of blood flow, are tolerated in digit replantation. Survival of replanted digits has been reported after as long as 42 hours of warm ischemia and 94 hours of cold ischemia. Immediate and proper cooling of the amputated part lengthens tissue survival.

A preoperative history of cigarette smoking does not seem to confer as negative an effect on the replantation as postoperative smoking. It has been shown to have significant negative effects on blood flow for up to 3 months. If the patient is unwilling to stop cigarette smoking, this could be considered a contraindication.

Amputated digits that are not commonly replanted are ones that are severely crushed or damaged, severely contaminated, have undergone multiple-level amputations, or are those of patients with multiple associated traumas or severe medical problems.

Preoperative Planning

A team approach is essential to the success of replantation. The transport team and the emergency room physicians should place the amputated part in appropriate dressings and arrange efficient transport to a center that has a replantation team. This team includes microvascular surgeons, anesthesiologists, operating room personnel and facilities, a microscope, and postoperative monitoring capabilities.

Minimal time should be wasted in obtaining unnecessary examinations. Many of the studies can be performed on the patient while the operating room is being prepared and, subsequently, the amputated finger is being prepared. If the operating room is ready and the patient is not, the finger can be taken to the operating room and prepared for replantation.

Obtaining a history and physical examination is necessary to aid in determining indications for surgery and avoiding postoperative complications and disappointing outcomes. Current events related to the injury are detailed, including the time and mechanism of injury, the location and conditions where the trauma took place, and the way the amputated digits or stumps were handled prior to admission. Information regarding the patient's general condition, including other injuries that occurred in relation to the current trauma and the patient's medical and surgical history, are obtained. The general examination of the patient is performed to exclude other major injuries. The amputated digit and the stump are examined in the emergency room. This examination is performed without digital blocks, which may cause injury to structures essential for successful replantation.

The amputated part is carefully preserved in a cool condition to minimize damage from ischemia. The amputated part is wrapped in gauze moistened with Ringer's

lactate or saline solution, and placed in a plastic bag or specimen cup, which is then placed on ice. The amputated part is never placed directly on or immersed in ice or ice water, as this can cause freezing and permanent damage to the tissues.

Radiographs of the amputated digit and hand are taken to evaluate the condition of the bones and the presence of foreign bodies. The patient is kept in a warm environment, provided with adequate hydration, tetanus prophylaxis, and broad-spectrum antibiotics that are appropriate for the mechanism of injury and condition of the wound.

A discussion with the patient and the patient's family should include the risks of anesthesia, possibility of long operating times, potential need for blood transfusions, requirement for postoperative hospitalization, possibility of further surgery to salvage the finger in case of vascular compromise, the likely need for secondary procedures such as tenolysis, and the definite need for extensive physical rehabilitation postoperatively to achieve optimal outcomes. Patients are informed of the possibility that revision of the stump and closure of the wound might be performed; that vein grafts, nerve grafts, skin grafts, local flaps, regional flaps, or free flaps might be necessary to complete the reconstruction; and that bone shortening is often necessary in the primary reconstruction. The patient is made aware that there is a chance of immediate or late failure of the reconstructive effort and that other options for rehabilitation, such as prosthetic devices, might need to be resorted to in the future. Finally, if the patient is a smoker, he or she is told that smoking can affect the outcome and that postoperative smoking for up to 3 months can be detrimental to the final outcome of the surgery. The patient is given clear and realistic expectations regarding the expected outcomes.

Appropriate consultations with medical specialists are obtained. If a psychiatric disorder is suspected, a psychiatric consultation is obtained which in some circumstances might affect the decision as to whether to proceed with the replantation.

If an operating room is available, the amputated part is taken there first, examined under magnification, and set up for replantation while the patient is being prepared and anesthetized. Replantation is performed as soon as possible to shorten the ischemic time.

Surgery

Patient Positioning

The patient is placed in a supine position with the arm abducted and placed over an arm table. A tourniquet is placed high up on the arm and on one of the lower extremities (in case vein, nerve, or skin grafts are required). Appropriate warming devices are placed to maintain normal body temperatures throughout the procedure.

General anesthesia is recommended, as replanting each digit requires between 2 to 4 hours to complete, and this usually becomes uncomfortable for the patient when an axillary block is performed. A brachial plexus block can be used to decrease the requirement for pain medication and decrease the incidence of vascular spasm in the operating room and postoperatively. Some authors use a brachial plexus block in addition to sedation in adults and older children with single-digit amputations in one extremity who are thought to be non-anxious and may tolerate 2 to 5 hours of immobility.

Preparation and cleansing of the stump and amputated part are performed using a povidone-iodine solution followed by copious irrigation with normal saline. A tourniquet is placed on the involved arm and one leg, and both are prepared and draped.

Technique

A two-team approach is used whenever possible, with one team working on the amputated part and the other working on the hand. The operative sequence for digit replantation varies slightly among surgeons and depends on the individual situation and ischemia time; however, our general routine is (1) identification of structures, (2) debridement, (3) bone shortening and fixation (Fig. 37-4), (4) repair of the extensor tendon, (5) repair of flexor tendons, (6) anastomosis of the artery, (7) coaptation of nerves, (8) anastomosis of veins, and (9) closure of the wounds.

Identification of Structures

During the dissection, the amputated part can be kept cool by placing it over a cooled apparatus. Bilateral mid-axial incisions are made, and the volar skin flap and dorsal

skin flap are elevated, folded back, and sutured in position (see Fig. 37-2). If one team is performing the surgery, the exploration begins by identifying the neurovascular bundles and other structures in the amputated part using an operating microscope. High magnification minimizes inadvertent injury to important structures and decreases the chances of vasospasm. The digital arteries, one or two dorsal veins, and the digital nerves are identified and tagged with 6-0 sutures. The extensor and flexor tendons are located, trimmed, and tagged. If there is no artery available in the amputated part for anastomosis as is sometimes the case with ring avulsion amputations, the options for replantation become more limited. The use of an afferent arteriovenous shunt for the revascularization of unfavorable amputations of fingertips and distal phalanges may be an option (Fig. 37-5), or the replantation effort is halted and a revision and closure of the stump is performed (12,28â€“30).

Debridement

All non-viable tissue is excised while being careful not to cause inadvertent injury to important structures including the pulleys, particularly the A2 and A4 pulleys.

Bone Shortening and Fixation

Without excessive periosteal stripping, the bone is trimmed to provide surfaces that allow for good bone contact during fixation. Bone shortening also allows for direct vessel or nerve repair without tension, minimizes the need for using vein grafts and nerve grafts, and permits a more relaxed skin closure as postoperative edema can create compression of the vessels. Preservation of 5 to 10 mm of bone near a joint helps increase the range of motion that can be achieved. In traction avulsion amputations, the bone may require further shortening (19,36). Amputations through a joint are commonly treated with arthrodesis at the time of replantation. If this is not performed in the initial setting, eventual use of a silastic arthroplasty can solve some problems related to joint stiffness but is insufficient when the collateral ligaments are not present or do not provide adequate stability.

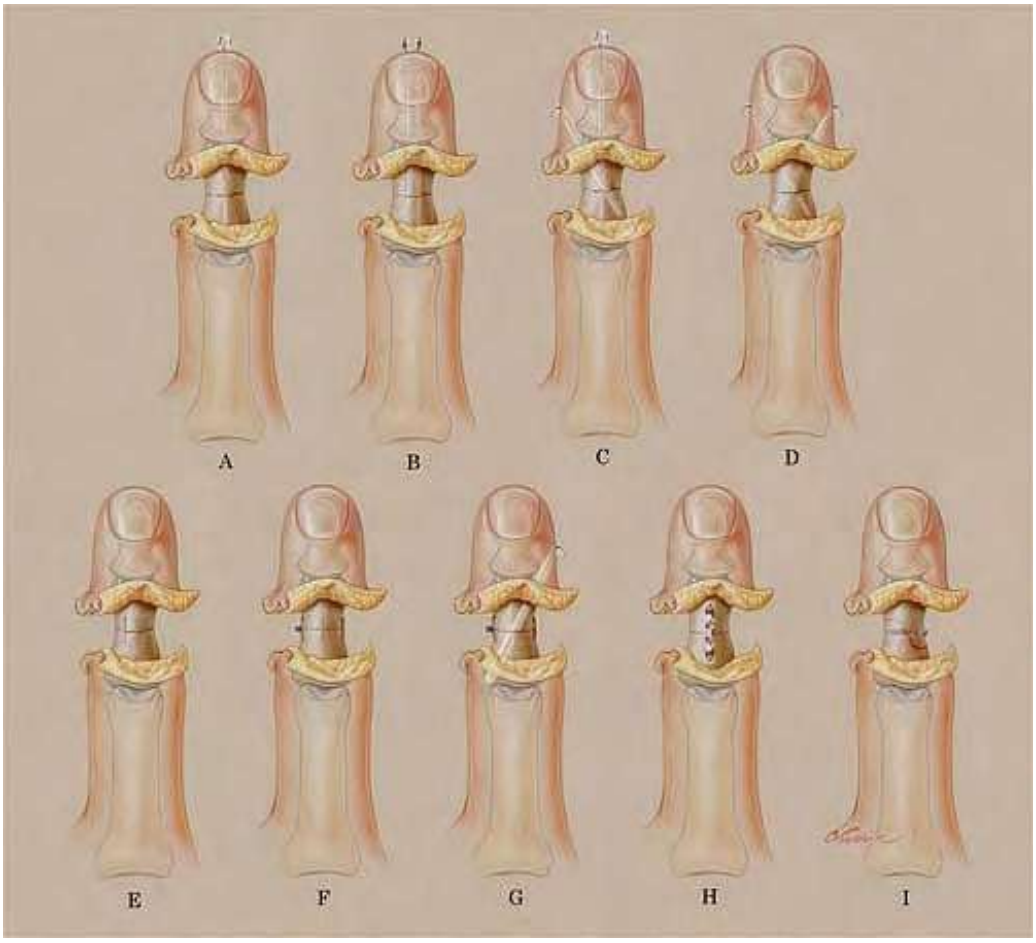

FIGURE 37-4 Different methods used for bone fixation. (*A*) One longitudinal Kirschner pin, (*B*) Two parallel Kirschner pins, (*C*) one longitudinal and one oblique Kirschner pins, (*D*) two oblique Kirschner pins, (*E*) two parallel interosseous wires, (*F*) two perpendicular interosseous wires, (*G*) one interosseous wire and one Kirschner pin, (*H*) plate fixation, (*I*) a screw for oblique fractures.

Bone fixation can be performed using crossed Kirschner wires (K-wires), two longitudinal K-wires, a single intramedullary K-wire, interosseous wiring, intramedullary screws, or external fixation devices (see Fig. 37-4). The type of fixation used depends on the stability of the fragments, the desire for early mobility, the compliance of the patient, and the preference of the surgeon. Plate fixation may be beneficial for fractures at the mid-shaft level as it allows for early mobility; however,

plate fixation requires more extensive soft tissue dissection and periosteal stripping, and requires more time to perform. Longitudinally placed intramedullary or crossed K-wires that do not cross non-involved joints are the most commonly used methods. For replantations distal to the DIP, one K-wire may suffice. The authors prefer the use of interosseous wires when the replantation is performed at the level of the proximal part of the proximal or middle phalanx. Interosseous wiring provides adequate stabilization for early mobility and allows for adjustment of the alignment on an outpatient basis if malalignment is noted within the first week postoperatively.

Repair of the Extensor Tendon

The extensor tendon is repaired using 4-0 medium-lasting sutures placed as horizontal mattresses or in a figure-of-eight fashion. Excessive and unnecessary manipulation, pinching, and grasping of tendons are avoided. In amputations at the proximal phalangeal level, the disrupted lateral band is repaired to achieve optimal extension function of the interphalangeal joints.

Repair of Flexor Tendons

The flexor tendons are repaired by applying a modified Kessler technique using 4-0 nonabsorbable sutures followed by a running circumferential suture using a 6-0 monofilament nonabsorbable suture. Both the FDS and flexor digitorum profundus (FDP) are repaired meticulously. In zone II, only the FDP tendon is repaired to minimize subsequent tendon adhesion.

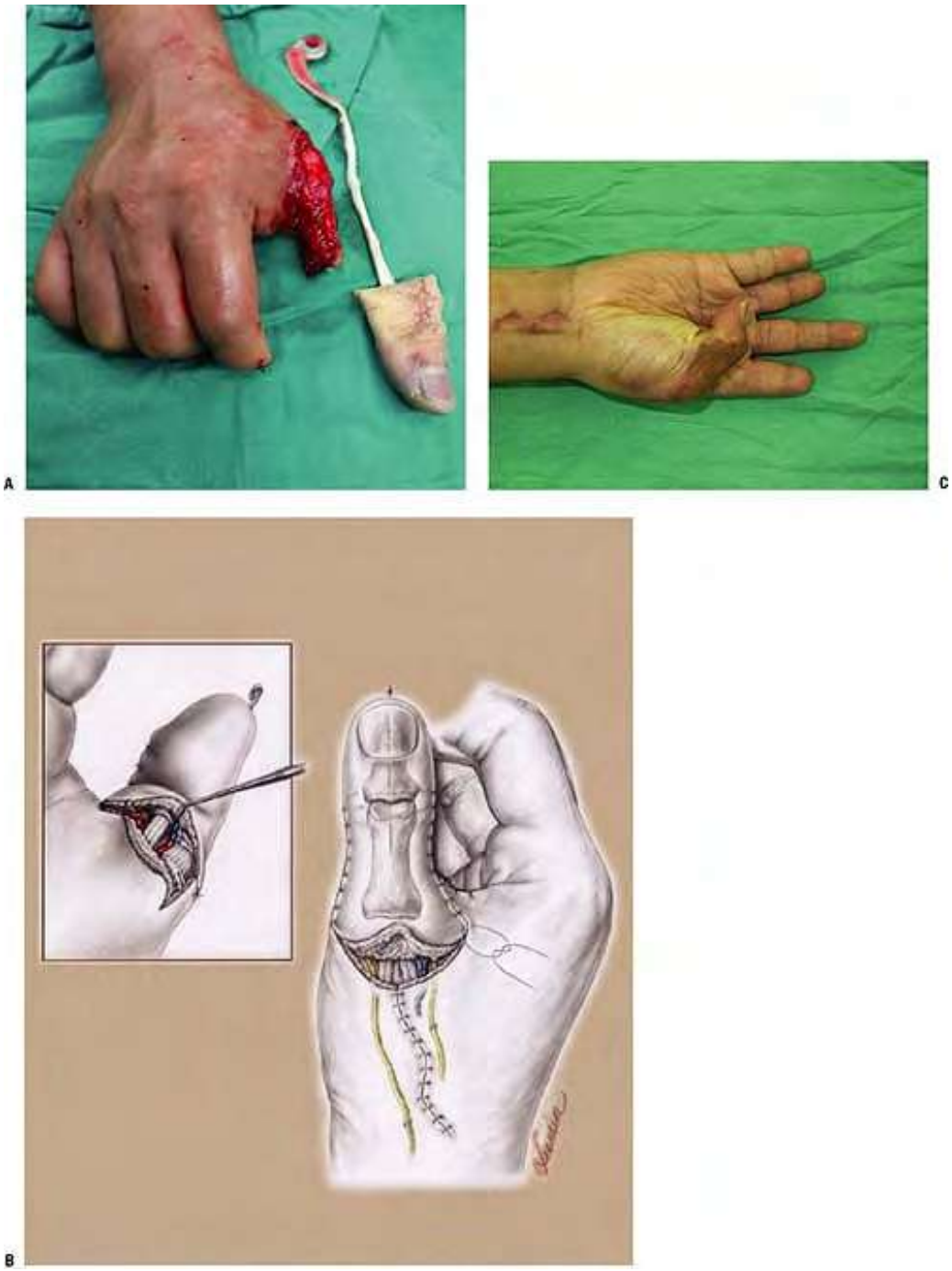

FIGURE 37-5 A: A 28-year-old man suffered from an avulsion amputation of his right thumb. The thumb was disarticulated at the interphalangeal joint and the skin envelope was avulsed proximally near the level of the metacarpophalangeal joint. Both digital arteries were avulsed from the distal amputee around the interphalangeal joint. The digital nerves were avulsed proximally near the level

of wrist. The flexor pollicis longus tendon was avulsed at the level of the tenomuscular junction. B: Illustration demonstrating intraoperative details. The recipient artery in the hand was anastomosed to one of the volar veins to create an arteriovenous fistula. One dorsal vein was used as the outflow of the finger (a vein graft was necessary to bridge the gap). The bone was fixed using one longitudinal Kirschner pin through the interphalangeal joint. Both nerves were coapted to the superficial branches of the superficial radial nerve at the level of the anatomic snuffbox. C: Six months after replantation of the thumb using the arteriovenous technique.

Anastomosis of the Artery

The artery on both the amputated part and the proximal stump are examined under microscopic visualization and are cut at a level that causes no apparent damage to the vessel wall. The presence of bruising or thrombus, or the appearance of a coiled artery, indicates the need to trim the vessel further. The presence of thrombus within the lumen, or separation of the vessel wall layers, indicates damage and requires the vessel to be trimmed further. The outflow from the proximal stump of the artery is checked and should be spurting before it is considered adequate for revascularization. Hydration of the patient and maintenance of a systolic blood pressure over 100 mm Hg are important to achieving this purpose. Adventitia stripping often aids in relieving and preventing vessel spasm. Vein grafts from the volar aspect of the forearm or the dorsal foot are used as conduits whenever the gap is too large to complete the anastomoses without tension.

Revascularization can be provided by mobilizing a healthy undamaged artery from the adjacent digit (see Fig. 37-3). This technique provides inflow from a vessel that is unaffected by trauma, and obviates the need for vein grafting with its inherent risks. This can be performed in conjunction with transfer of the digital nerve in cases of thumb replantation. Vein grafts, nevertheless, are reliable, readily available, and avoid the need for incisions and scars in adjacent non-injured fingers.

When the digital artery is not available on the amputated part, afferent arteriovenous shunting can be used to perfuse the digit. Fukui and Yabe used this technique and

found that a higher success rate was found when the inflow was from the artery of the stump to the volar veins rather than the dorsal veins. Anatomic studies by Moss and coworkers revealed that the direction of valves in the oblique communicating veins between the dorsal and volar veins favor flow from the volar veins to the dorsal veins. This technique can be successful in replantation of amputations as high as the interphalangeal joint (see Fig. 37-5).

If after release of the clamps the replant does not pink up, a waiting period up to 30 minutes is suggested. The vessels and the anastomosis are carefully inspected, adventitial stripping is performed, and the patency test is performed. Side branches, both proximal and distal to the anastomosis, can be tested to see if blood flow exists. Redo of the vascular anastomosis may become necessary because of either inadequate debridement of injured vessels or thrombosis at the anastomotic site due to inadequate microsurgical technique. Injection of high concentration heparin solution and/or thrombolytic agents is performed if all other factors are optimized and reperfusion seems to be inadequate.

Coaptation of Nerves

The nerve repair is performed by approximating the epineurium of both ends of the digital nerves using 10-0 or 11-0 sutures. Nerve grafts are used liberally to avoid any tension on the repair. Depending on the severity of the injury and outlook for wound coverage, secondary nerve grafting may be a better option to avoid exposure of this nonvascularized graft which may act as a source of infection and may result in poor function. The medial antebrachial cutaneous nerve and the sural nerve provide good sources of nerve fascicles for grafting purposes. In multiple-digit amputations, nerve grafts can be harvested from parts that are not planned for replantation. In avulsion cases of the thumb, sensory nerves from the avulsed finger can be attached to the dorsal sensory nerves of the first web space to restore protective sensibility.

Anastomosis of the Veins

The veins of the amputated part are inspected carefully after releasing the arterial clamps, and the ones with the most bleeding are anastomosed to the recipient veins in the hand. Dorsal veins are preferred for anastomosis. After separating the dorsal veins from the skin for a short distance and ligating a few branches, the length is usually

adequate to complete the anastomoses primarily.

At least two veins are anastomosed whenever possible. Matsuda and coworkers evaluated the correlation between the number of anastomosed veins and survival rate in finger replantation. Their conclusions were that the essential amount of venous anastomoses is one in zone II (lunula to distal interphalangeal joint), two in zone III, and one in zone IV (proximal phalangeal region).

In situations where perfusion of the flap is good and there are no available dorsal veins for anastomosis, volar veins may be used. If there are no available veins or no outflow from any veins after a long waiting period, a last resort is to inspect the contralateral digital artery. If retrograde flow is present from the contralateral artery, that artery may be anastomosed to a recipient vein which can serve as an outflow vessel for the finger.

When replanting finger tip amputations and no sizable vein is present, other options include removal of the nail to keep continuous bleeding from the raw nail bed, postoperative periodic digital massage, and/or use of medical leeches. All these methods are performed with or without systemically heparinizing the patient.

Vein gaps can be bridged with vein grafts from the volar forearm or dorsal foot. If soft tissue is also lacking in the area where a vein graft is needed, a venous flap can be used which will provide a vein graft and a soft tissue cover.

Skin Closure

Meticulous hemostasis is obtained prior to closure of the wound. Under microscopic visualization, free ends of disrupted veins and arteries are carefully ligated. If only one artery is used for a finger, the contralateral digital artery is ligated to avoid postoperative bleeding. The skin is loosely approximated with a few interrupted sutures. The vessels are checked to make sure there is no compression or kinking with placement or manipulation of the hand and finger in different positions. Although a skin graft can be placed directly over the pedicle, rotation of a local skin flap is preferred and the defect from the rotation is skin grafted. Antibiotic ointment, non-adherent dressings, and wet gauze are placed over the incisions. No significant dressings are placed between the fingers to avoid potential compression of the digital

vessels. Loosely placed circumferential dressings cover the hand.

Additional Procedures

Carpal tunnel release is performed for patients who have undergone avulsion amputations. The transverse metacarpal ligament is released, and the carpal tunnel is explored. In cases of severe avulsion amputations of the digits, forearm fasciotomies may be necessary.

Results

Survival of replanted fingers ranges from 70% to 95% in both children and adults, with many patients achieving extremely satisfying results (Figs. 37-6 and 37-7). Survival rates are reported to be lower in avulsion and crush injury cases. Re-exploration due to vascular compromise is necessary in fewer than 20% of replantations and carries a salvage rate of approximately 70%.

Replantations of single or multiple digits when the amputation is distal to the sublimis insertion are highly satisfactory, both in terms of function and appearance, even without motion at the DIP. Secondary surgeries in these cases are rarely necessary.

A replanted thumb results in satisfactory outcomes in most patients, regardless of the type of injury or level of amputation (see Fig. 37-6). Even if a replanted thumb has a limited range of motion, it still plays an important role as a post for opposition. Schlenker et al reviewed their experience with 64 thumb replantations and found that 50% of the patients had two-point discrimination of less than 10 mm and an average active range of motion after replantation of 35% of normal for the interphalangeal joint and 29% of normal for the metacarpophalangeal joint. Most patients were able to return to work after approximately 7 months.

Sensory reinnervation is extremely important for the fingers to achieve optimal function. Glickman and Mackinnon reviewed the recovery of sensibility in 367 fingers and 87 thumbs that were replanted successfully. Average static two-point discrimination (S2PD) was 9.3 mm in clean thumbs versus 12.1 mm in crush/avulsion thumb replantations. Average S2PD was 8 mm in incisive injuries and 15 mm in crush/avulsion finger replantations. Mean S2PD was 11 mm in thumb replantations and 12 mm in finger replantations. Sixty-one percent of the thumbs and 54% of the fingers regained S2PD that was useful. They found that the factors that influenced sensibility

were the age of the patient, the level and mechanism of injury, digital blood flow, cold intolerance, and postoperative sensory re-education. It was concluded that recovery of sensibility in the replanted digit is comparable to simple nerve repair and to nerve grafting techniques.

Cold intolerance is a significant problem following replantation. The exact mechanism of cold intolerance following digital nerve injury is not clear but may have a neurogenic or vascular origin. The more recovery of sensibility, the less cold intolerance seems to occur. Cold intolerance usually improves after approximately 2 years, but can persist for many more years.

Postoperative Management

The patient is transported to a unit equipped with a warming device and highly trained nursing staff capable of performing hourly checks and recognizing signs of vascular compromise (Table 37-2). Every check includes an inspection of finger color, capillary refill, skin Turgor, and temperature. Pinprick testing and evaluation of vascular flow by hand-held Doppler ultrasonography are performed when the finger displays clinical signs of circulatory compromise. Temperature monitoring

P.492
P.493
P.494

of replanted digits has proved a reliable indicator of vascular status, as replanted digits are located in a terminal site with a terminal vascular loop that will not be influenced to a significant degree by heat exchange from the underlying bed. In general, a drop in temperature greater than 2Â°C from that of the control digit, or below 30Â°C, indicates the possibility of circulatory compromise.

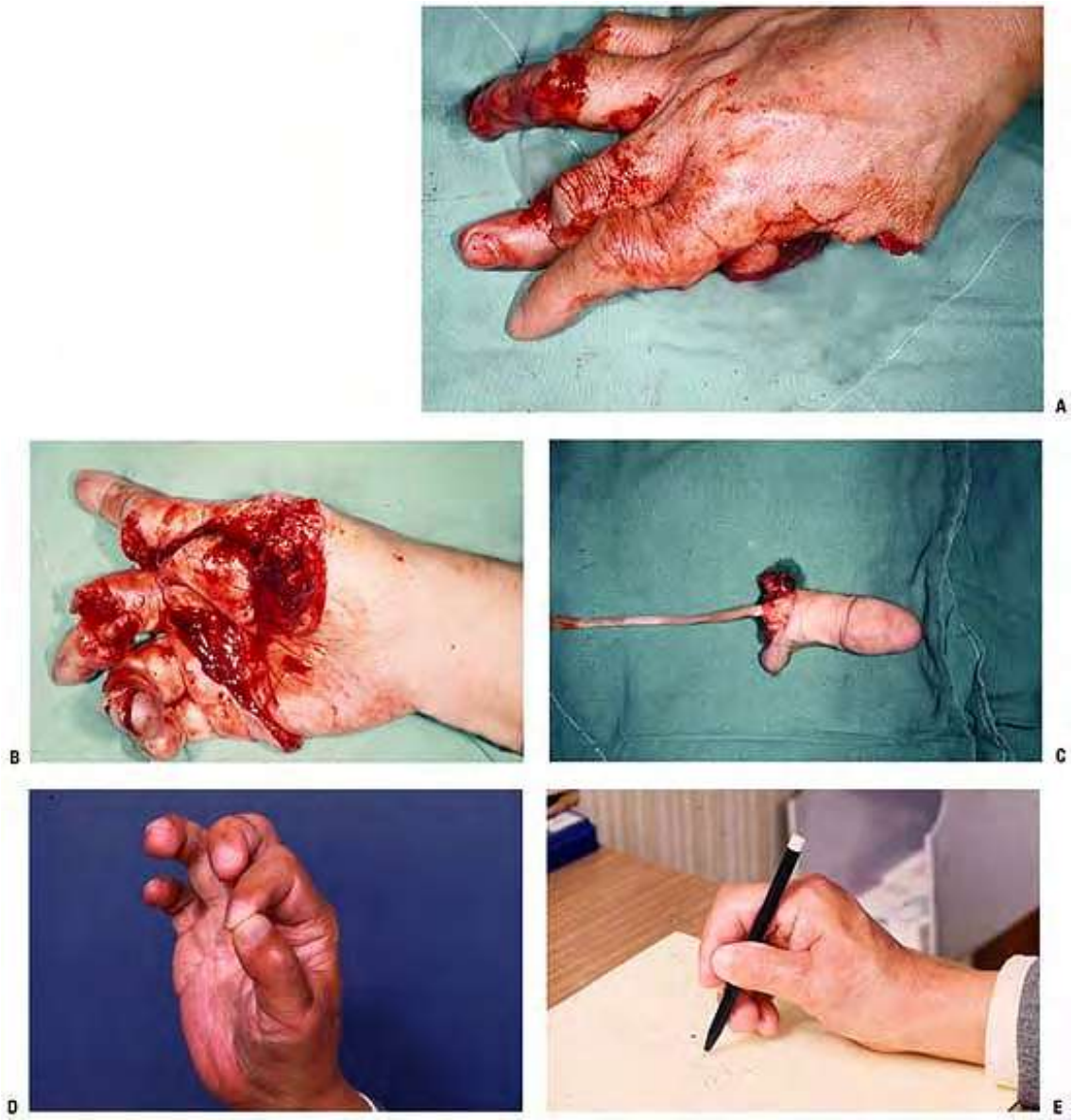

FIGURE 37-6 A: A 57-year-old man sustained a crush injury of his right hand by a rolling machine. B: The thumb was completely amputated and the index and middle fingers were incompletely amputated at the level of the proximal phalanx and middle phalanx, respectively. C: The thumb was avulsed at the level of the proximal phalanx and the flexor pollicis longus tendon was torn from its tenomuscular junction. D: Nine months follow-up demonstrating adequate positioning of the replants and good strength. E: At 9 months follow-up the patient is able to grasp a pen.

FIGURE 37-7 A: An 18 year-old man suffered from a sharp cutting injury by a machine saw. Volar view demonstrating the level of amputation of the three ulnar digits. B: This lateral view demonstrates the level of amputation of the long, ring, and little fingers. C: Dorsal view of the three amputated digits. D: At 9 months follow-up the patient has a good cascade and is able to make a fist. E: At 9 months follow-up, the patient is able to grip with good strength.

Table 37-2. Signs of Arterial and Venous Compromise in Finger Replantation

	Arterial Compromise	Venous Compromise
Finger color	Mottled, bluish, or pale	Bluish, cyanotic, or dusky
Capillary refill	Sluggish (>2 sec)	Brisker than normal
Tissue turgor	Flat, decreased turgor	Tense, increased turgor
Temperature	Cool (>2Â° difference compared to control)	Cool (>2Â° difference compared to control)
Pinprick test	Scant amount of dark blood or serum	Rapid bleeding of dark blood
Doppler signals	Absence of pulsatile arterial signals	Absence of continuous or venous augment signals

A pulse oximeter placed distally on the replanted digit can be used to monitor changes in oxygenation. Laser Doppler flowmetry and photoplethysmography are used at some centers, but the cost of the instruments and the need for interpretation of the data have not proven its benefit over traditional methods such as clinical monitoring and temperature monitoring.

Intensive monitoring of replanted digits is undertaken for 3 to 5 days post replantation. Although thrombosis of vessels can occur up to 2 weeks following the

replantation procedure, it is rare; and when it does occur, the chance of salvage is lower. This may be partly due to the delay in diagnosis in late thrombosis cases. In Betancourt's series of 71 replanted digits, the highest risk of critical thrombosis after replantation occurred in the first 3 days after surgery.

Patientsâ€™ activities are monitored closely for the first 3 to 5 days to ensure that injury to the replanted part does not occur. The arm is elevated above heart level to decrease swelling, pain, and venous congestion. The level of arm elevation should not be excessive as to decrease arterial pressure in the hand and digits. Urine output and electrolytes are monitored to ensure adequate hydration to avoid vasoconstriction. Intraoperative placement of a continuous axillary brachial plexus block catheter can be a source of postoperative pain relief and can provide a chemical sympathectomy which may be applied for 3 to 5 days postoperatively.

The use of anticoagulants is controversial; however, heparin is often used in replantation cases. The authors favor the use of low-molecular-weight dextran starting in the operating room at the time of completion of the first vascular anastomosis. Systemic heparinization is used in replantation of crushed or avulsed digits or in cases where intraoperative vascular thrombosis occurs and requires revision of an anastomosis. Some authors recommend the use of aspirin at 325 mg per day, with the first dose administered prior to the replantation.

A warming lamp placed in the region of the replant and warming the patient's room may have some benefit. All patients are advised to avoid caffeine intake and cigarette smoking for at least 2 to 3 months following the surgery.

Complications

Immediate complications include partial or complete loss of the replanted digit due to circulatory compromise, infection, and bleeding. The presence of infection can lead to vascular thrombosis. This can be partially avoided by performing an adequate debridement and administering proper antibiotics based on the level and type of contamination. Skin necrosis can also be seen in the early phase due to inadequate debridement or compromise of blood supply to a part of the replanted digit. This scenario requires debridement and closure with local skin flaps, skin grafts, or transfer of more distant flaps.

Vascular Compromise Post Replantation

Following the detection of a change in one of the parameters used to assess finger or thumb perfusion, the decision is made on whether to proceed with exploration. It is our practice to have an aggressive approach to re-exploration and attempts at salvage. If arterial occlusion is suspected, the patient is immediately taken back to the operating room for re-exploration. If venous compromise is suspected, several maneuvers may be attempted prior to re-exploration: the dressing and some of the sutures are removed, the limb is elevated, and the patient is placed on a systemic heparin drip. If the digit does not show signs of improvement within 20 to 30 minutes, it is highly likely that the veins draining the replanted finger have undergone thrombosis. In this situation, the decision is made by the surgeon whether to take the patient back to the operating room or pursue a nonsurgical approach. For venous outflow obstruction in distal replantation, management by using medical leech therapy or continuous bleeding can be performed with high success rates. The nail plate is removed, a stab incision in the paraungual area is made, and a heparinized saline drip is placed over the incision site to maintain external bleeding. The smaller the replanted digit, the higher the salvage rate using these methods. In Han's series of 144 finger tip replantations with venous outflow through the use of this technique, the salvage rate was 70% (101 of 144 digits). The average period of the salvage procedure was 7.6 days. The duration of bleeding was different for different types of injuries: guillotine injuries required 5.9 days, crush injuries 8.2 days, and avulsion injuries 8.0 days.

Medical leeches, *Hirudo medicinalis*, have been reported to be effective for the treatment of venous outflow obstruction. The leech secretes hirudin, a potent anticoagulant, which acts locally without significant systemic effects. The therapeutic period required is approximately 3 to 5 days. Significant blood loss and the risk of infection (7%â€“20%) are the main disadvantages of this modality. The organism involved in infections from leeches is *Aeromonas hydrophilia*, and a third-generation cephalosporin is prescribed as a prophylactic method to prevent infections.

On exploration, the artery and vein are inspected. Occasionally, a hematoma is the cause of problem; however, a venous or arterial occlusion is often found. The thrombus is evacuated and arterial inflow or venous outflow are re-established. Vein

grafts are used to reach healthy vessels when necessary. The steps are similar to the previously described procedure, including the occasional need to infuse high-concentration heparin solution and/or thrombolytic agents.

Late Complications and Secondary Procedures

The mechanism and level of injury, ischemia time, quality of the original repair, the postoperative rehabilitation program, and the effort put forth by the patient as well as the healing process all determine the final outcome. Depending on the combination of these factors, the resultant quality of finger function and the patient's lifestyle and motivation will determine whether the patient will seek secondary procedures to improve outcome. The majority of patients undergoing replantation (approximately 30% to 40% of patients who underwent finger replantation and up to 90% of patients with injuries in zone II) will require one or more secondary procedures. The physical therapist and surgeon follow the patient closely after the replantation and provide an aggressive and thorough physical rehabilitation program. Once the improvement has plateaued, usually at 6 months post-replantation, an evaluation of finger function and overall hand function will determine the need for secondary procedures. The patient's desires and level of function are of utmost importance in determining the need for surgery.

Frequent sequelae include tendon adhesions and joint stiffness. Avoiding transfixation of non-involved joints and performing a meticulous tendon repair decreases the incidence of these complications. Tenolysis as a secondary procedure is required in approximately 50% of finger replantations and up to 90% in patients with the level of amputation in zone II. Stiff proximal interphalangeal (PIP) and DIP joints are sometimes acceptable if the metacarpophalangeal (MCP) joint is mobile and allows for adequate grip. This is particularly true for the more ulnar digits. If the patient is able to use the thumb, index finger, and long finger to pinch objects adequately, he or she may be satisfied with the result and elect not to undergo further surgery. During the procedure of flexor tenolysis and/or tendon grafting, care is taken not to injure the repaired neurovascular bundle during the dissection. A detailed knowledge of the previous operative procedure can aid in avoiding injury to repaired structures. Pulleys are preserved, particularly the A2 and A4 pulleys. The tendon is exposed and inspected. If tendon rupture has occurred, tendon grafts with or without tendon rods are necessary in some patients. Ruptured or atrophied tendons are debrided. If the

tendon is of poor quality, particularly in zone II injuries, reconstruction of the pulleys and insertion of a tendon rod followed later by tendon reconstruction may be the procedure of choice. Tendon grafts can be used to reconstruct the tendons if the quality of tendon appears good and the pulleys are intact. Pulley reconstruction and tenolysis are usually not performed in the same procedure. Extensor tenolysis is performed through dorsal longitudinal incisions and usually yields good results. Intrinsic reconstruction and dorsal arthrolysis are worthwhile when active and passive flexion are present. Arthroplasty can yield excellent results and is indicated for some MP and PIP joints. Arthroplasty can be performed along with other procedures such as tenolysis, capsulotomies, and capsulorrhaphies. Arthroplasty is performed through the same volar incisions that are used for the flexor tenolysis, or dorsal incisions are made. Vascularized joint transfers are indicated for replacing an MP or PIP joint, a flaccid thumb, or a fused MP or PIP of the index or other digits when the global hand function is good. The second or third toe metatarsophalangeal, PIP, or DIP joints are used from the toes.

Certain procedures are not performed simultaneously, such as tenolysis with procedures that require finger or hand immobilization (i.e., nerve repair, osteotomies, joint fusions, vessel repair and bone grafts). Some procedures can be performed together, such as tenolyses and arthroplasties. Malunion or nonunion can occur in up to 20% of replanted digits and seems to occur more commonly with replantations fixed with K-wires. Neurolysis, nerve repairs, and nerve grafts are indicated when the Tinel sign fails to cross the site of repair for 3 months. Early repair is advocated. Nerve grafts are used if the nerve injury is on the same side as the vascular repair to avoid injury to

the vascular structures. It is important to find a healthy proximal stump for anastomosis. If not found, nerves from the dorsal radial sensory branches or the same or adjacent digits may be transferred.

Soft tissue coverage such as pulp transfers for thumb or opposable digits are sometimes necessary. Local flaps or free tissue transfer is sometimes necessary to cover open wounds or defects after contracture release of the flexor or dorsal side or the webspace. Perforator flaps and the lateral arm flap are good options for this.

Cold intolerance can be found in the majority of patients and most often resolves with

time. Delayed circulatory compromise may occur up to 2 to 3 weeks post replantation, and often is caused by cigarette smoking or infection. Severe infections after digit replantation are rare and are usually related to inadequate debridement during the initial surgery or poor perfusion to the digit. Severe complications are unusual in digit replantation. Compression syndromes at the forearm or wrist level can occur and are usually associated with severe avulsion injuries. Fasciotomy and carpal tunnel release may be required in those cases. Web space contracture can occur as a result of thenar muscle contracture and/or soft tissue contracture in the web space. If the initial replantation is successful, late problems are usually related to suboptimal function due to poor motion and can be corrected with secondary procedures. Other secondary procedures that may be required are re-amputation, rotational osteotomies, and bone grafts.

Recommended Reading

Axelrod TS, Buchler U. Severe complex injuries to the upper extremity: revascularization and replantation. *J Hand Surg Am.* 1991;16(4):574â€“584.

Baek SM, Kim SS. Successful digital replantation after 42 hours of warm ischemia. *J Reconstr Microsurg.* 1992;8(6):455â€“458.

Betancourt FM, Mah ET, McCabe SJ. Timing of critical thrombosis after replantation surgery of the digits. *J Reconstr Microsurg.* 1998;14(5):313â€“316.

Buncke GM, Buncke HJ, Kind GM, Buntic R. Replantation. In: Achauer BM, Eriksson E, Guyuron B, Coleman JJ, Russell RC, Vander Kolk CA, eds. *Plastic surgery indications, operations, and outcomes*. St. Louis: Mosby; 2000:2131â€“2147.

Buncke HJ, Alpert BS, Johnson-Giebink R. Digital replantation. *Surg Clin North Am.* 1981;61(2):383â€“394.

Buncke HJ, Whitney TM, Hill MK. Bony fixation in replantation. In: Buncke HJ, ed. *Microsurgery: transplantation-replantation. An atlas text.* Philadelphia: Lea &

Latha Ganti
Editor

Atlas of Emergency Medicine Procedures

Springer

1 Arterial Cannulation (Radial and Femoral)

Jeffrey Kile, Katrina John, and Amish Aghera

Arterial cannulation is frequently performed in the care of critically ill patients for purposes of both serial arterial blood gas sampling and continuous intra-arterial blood pressure monitoring. It also provides arterial access for less common procedures, including thrombolysis, embolization, angiography, and infusion of vasoactive drugs. This chapter discusses cannulation of the radial and femoral arteries—the two most common sites for indwelling arterial catheter placement.

1.1 Indications

- Continuous monitoring of blood pressure in acute illness or major surgery
- Serial sampling of arterial blood during resuscitation
- Inability to use noninvasive blood pressure monitoring (*e.g.*, burns, morbid obesity)
- Continuous infusion of vasoactive inotropes (*e.g.*, phentolamine for reversal of local anesthesia)
- Angiography
- Embolization

1.2 Contraindications

- Absolute
  - Circulatory compromise in the extremity
  - Third-degree burns of the extremity
  - Raynaud's syndrome
  - Thromboangiitis obliterans (Buerger's disease)
- Relative
  - Recent surgery in the extremity
  - Local skin infection
  - Abnormal coagulation
  - Insufficient collateral circulation
  - First- or second-degree burns of the extremity
  - Arteriosclerosis

1.3 Materials and Medications

- Radial artery cannulation: standard over-the-needle catheter assembly (Fig. 1.1)
  - Antimicrobial solution and swabs
  - Sterile gloves
  - Local anesthetic (1–2 % lidocaine without epinephrine)
  - Blunt needle
  - 25- or 27-gauge needle
  - Two 5 mL syringes
  - 4″×4″ gauze sponges
  - Standard over-the-needle catheter assembly
- Additional materials required for radial artery cannulation: over-the-needle catheter assembly with integrated guide wire
  - Over-the-needle catheter assembly with integrated guide wire
- Additional materials required for femoral artery cannulation: The Seldinger Technique
  - Introducer needle
  - Guide wire
  - Scalpel
  - Dilator
  - Arterial catheter

J. Kile, MBBS, PhD, MPH (✉) • K. John, MBBS
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: jeffrey.kile@gmail.com; trenjohn@me.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_1

Fig. 1.1 Materials and medications

1.3.1 Procedure: Radial Artery Cannulation—Standard Over-the-Needle Catheter Assembly

1. Ensure adequate collateral flow in the selected extremity using the Allen test (see details below).
2. Immobilize the extremity by dorsiflexing the wrist to approximately 60° over a small towel roll and taping the base of the fingers to an arm board or other flat, fixed surface (Fig. 1.2).
   - Rotation of the wrist may shift the artery from its normal anatomical position, complicating cannulation.
3. Locate vessel by palpation of arterial pulse using the second and third fingers of the gloved nondominant hand.
4. Sterilize overlying skin with antimicrobial solution.
5. Inject local anesthetic to raise a small (0.5 cm) wheal using 25- or 27-gauge needle, and direct needle through wheal to infiltrate skin superficial to the artery with additional local anesthetic.
   - Infiltration of the subcutaneous tissue with local anesthetic may also reduce vessel spasm during arterial puncture.
   - Injection of local anesthetic into the vessel may precipitate arrhythmia, so draw back on the plunger prior to infiltration to ensure the tip of the needle is not inside the vessel.
   - Injection of excessive anesthetic when raising a wheal may obscure palpation of the pulse.
6. Ensure proper function of needle-cannula assembly by checking that cannula advances smoothly over needle.
7. Connect a 5 mL syringe with the plunger removed to the over-the-needle catheter assembly.
   - Attachment of syringe improves control during cannulation.
8. Hold syringe connected to needle-cannula assembly like a pen with needle bevel facing upward.
9. Directing the needle at a 30° angle to the skin, puncture the skin through the anesthetic wheal immediately overlying the palpated artery, and advance needle slowly until tip enters arterial lumen, which is confirmed by visible arterial blood flow ("flashback") into the needle hub and syringe.
   - Avoid self-puncture by maintaining adequate distance between needle tip and index finger (Fig. 1.3).
10. Reduce the angle between needle and skin (by lowering the needle) and advance an additional 2 mm to ensure catheter tip (which sits approximately 2 mm behind the needle tip) has entered the lumen.
    - Advancing the needle too far (or failing to reduce angle between needle and skin) once initial flashback is visualized may result in piercing the back side (or "double puncture") of the artery wall, in which case visible blood flow will cease; if this occurs, slowly withdraw needle several millimeters until pulsatile blood flow reappears.
11. Stabilize position of introducer needle and advance catheter alone into artery over needle until hub of catheter is in contact with the skin; blood flow from catheter hub at this point indicates successful cannulation of the artery.
    - If difficulty is encountered at this step, catheter hub may be rotated slightly to facilitate advancement.
12. Remove the needle without dislodging catheter from artery.
13. Manually apply pressure to proximal aspect of artery to occlude blood flow from the catheter.
14. Attach desired extension tubing, injection cap, and stopcock to the catheter hub.
15. Secure the catheter hub to the skin using silk (2.0) or nylon (4.0) sutures as follows. Take a 0.5 cm bite of skin under the catheter hub with the suture needle, tie several knots in the suture without pinching the skin, then tie a second set of knots around the hub of the catheter firmly. If the catheter assembly contains an integrated suture wing for fixation, take a 0.5 cm bite of skin under suture wing with the suture needle, thread suture through wing perforation, and secure wing against the skin with several knots. If suture wing has two perforations, repeat this process to secure other half of wing to skin (Fig. 1.4).
16. Cover the catheter with an appropriate self-adhesive sterile dressing.
    - A small bead of antibiotic ointment applied to the puncture site prior to dressing reduces the likelihood of cutaneous wound infection.
17. Secure the tubing connected to the catheter with gauze and adhesive tape or other sterile dressing.
18. Ensure all connections extending from catheter are tight and well secured, as accidental disconnection may result in rapid exsanguination.

Fig. 1.2 Correct position of wrist prior to cannulation

Fig. 1.3 Puncture of radial artery with standard over-the-needle catheter assembly

Fig. 1.4 Radial arterial catheter secured to wrist

1.3.2 Procedure: Radial Artery Cannulation—Over-the-Needle Catheter Assembly with Integrated Guide Wire (Arrow and Other Brands)

Perform steps 1–6 from the "Radial Artery Cannulation: Standard Over-the-Needle Catheter Assembly," above, and then proceed with the steps below:

1. Remove protective cap from needle-cannula assembly and ensure proper function by sliding actuation lever along extension tubing to advance and retract guide wire through needle.
2. Retract guide wire as far back as possible using actuation lever to maximize visibility arterial blood flow ("flash-back") within introducer hub.

Perform steps 8–10 from the "Radial Artery Cannulation: Standard Over-the-Needle Catheter Assembly," above, and then proceed with the steps below (Fig. 1.5):

1. Hold needle stationery and slowly slide actuating lever forward to feed guide wire as far as possible into artery.
   - If resistance is met while feeding the guide wire, discontinue sliding actuating lever and withdraw entire unit from artery to prevent damage to guide wire or vessel wall
2. Advance entire assembly 1–2 mm further into vessel to ensure catheter tip (which sits approximately 2 mm behind the needle tip) has entered the lumen.
3. Stabilize clear introducer hub in position and advance catheter forward into artery over guide wire until hub of catheter is in contact with the skin.
   - If difficulty is encountered at this step, catheter hub may be rotated slightly to facilitate advancement.
4. Stabilize catheter in position and withdraw introducer needle, guide wire, and feed tube as a single unit; blood flow from catheter hub at this point indicates successful cannulation of the artery.

Perform steps 13–18 from the "Radial Artery Cannulation: Standard Over-the-Needle Catheter Assembly," above.

Fig. 1.5 Puncture of radial artery using over-the-needle catheter assembly with integrated guide wire

1.3.3 Procedure: Radial Artery Cannulation—The Allen Test

1. Occlude both radial and ulnar arteries of one extremity with digital pressure at wrist.
2. Instruct patient to repeatedly clench the fist tightly to exsanguinate the hand while occlusion of the arteries is maintained.
3. Without releasing digital pressure on arteries, instruct patient to extend fingers and observe palmar surface to confirm blanching of skin.
4. Release pressure on ulnar artery only and observe palmar surface for reperfusion (Fig. 1.6).
   - If reperfusion of the hand does not occur within 5–10 s, ulnar arterial blood flow may be compromised and radial artery cannulation should not be attempted. If reperfusion is brisk, repeat the test releasing pressure on radial artery only and observing palmar surface for reperfusion. If the return of rubor takes longer than 5–10 s, radial artery puncture should not be performed.

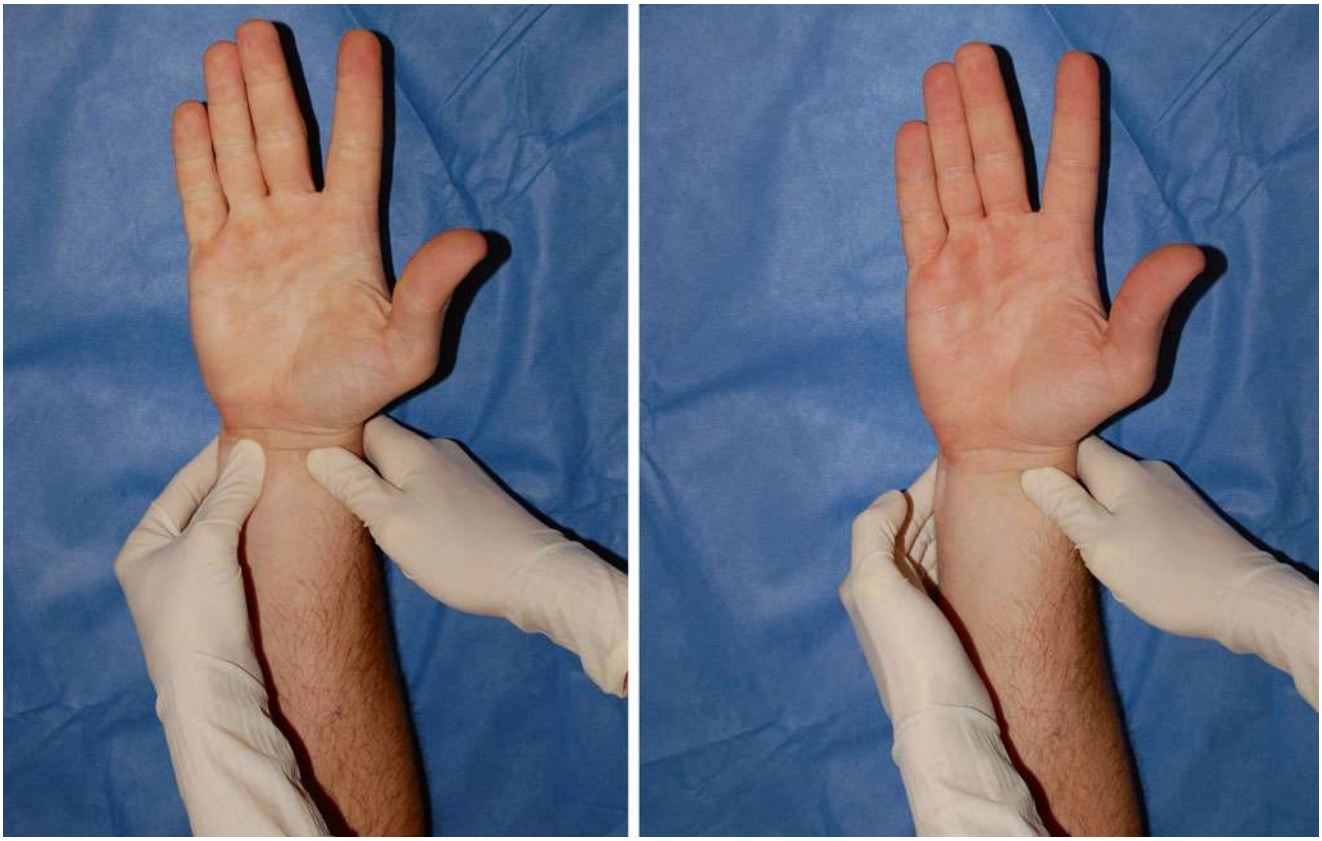

Fig. 1.6 The Allen Test

1.3.4 Procedure: Femoral Artery Cannulation—The Seldinger Technique

1. Place the patient in the supine position with the inguinal region adequately exposed.
2. Palpate the femoral pulse, located at midpoint between pubic symphysis and anterior superior iliac spine, using the second and third fingers of the gloved nondominant hand.

Perform steps 4 and 5 from the "Radial Artery Cannulation: Standard Over-the-Needle Catheter Assembly," above, and then proceed with the steps below (Fig. 1.7):

1. Attach 5 mL syringe to an introducing needle of bore sufficient to accommodate guide wire.
2. Hold syringe connected to introducing needle like a pen with needle bevel facing upward.
3. Directing the needle at a 45° angle to the skin in a cephalic direction, puncture the skin through the anesthetic wheal immediately overlying the palpated artery just distal to the inguinal ligament and advance needle slowly toward palpated artery until tip enters arterial lumen, which is confirmed by visible arterial blood flow ("flashback") into the needle hub and syringe.
   - Avoid self-puncture by maintaining adequate distance between needle tip and index finger.
   - Care must be taken to avoid trauma to the femoral nerve and vein bordering the femoral artery.
4. Hold needle stationery and remove syringe, taking care not to displace the intraluminal position of the needle tip.
   - Advancing the needle too far once initial flashback is visualized may result in piercing the back side (or "double puncture") of the artery wall, in which case visible blood flow will cease; if this occurs, slowly withdraw needle several millimeters until pulsatile blood flow reappears.
5. Occlude needle hub temporarily with gloved finger to prevent unnecessary blood loss and air embolism.
6. Thread blunt end of flexible guide wire smoothly into needle and gently into artery until at least one-quarter of guide wire is intravascular (Fig. 1.8).
   - If resistance is met while threading guide wire, remove wire from needle, reattach syringe, and aspirate blood to confirm continued intraluminal needle tip placement; if resistance is met while *removing* guide wire from needle, remove guide wire and needle from artery *as a single unit* to prevent shearing the guide wire off inside the vessel.
7. Holding the wire securely in place, remove the introducing needle.
8. Using a scalpel, make a small incision (approximately the width of the catheter to be used) through the dermis at the insertion site of the guide wire.
   - Avoid severing the guide wire by facing the sharp edge of the scalpel *away* from the guide wire.
9. While stabilizing the guide wire at its insertion site, thread the dilator over the free end of the guide wire until it is approximately one inch from the skin.
10. Grasp the free end of the guide wire protruding from the tail end of the dilator.
    - If it does not protrude from the tail end of the dilator, the guide wire must be removed sufficiently from the artery to be securely grasped; it must protrude visibly from the tail end of the dilator throughout the subsequent process of threading the dilator into the artery.
11. Holding the dilator firmly near its tip, thread the dilator over the wire into the skin with a back-and-forth twisting motion until it reaches the artery.
    - Only the skin tract should be dilated; dilation of the artery may result in excessive arterial injury and/or hemorrhage.
12. Holding the wire securely in place, remove the dilator.
13. While stabilizing the guide wire at its insertion site, thread the catheter over the free end of the guide wire until it nears the skin.
14. Grasping the guide wire where it protrudes from the tail end of the catheter, thread the catheter into the skin to its appropriate insertion length.
15. While stabilizing the catheter at its insertion site, slowly remove the guide wire.
    - If resistance is met while removing guide wire, remove guide wire and catheter from artery as a single unit to prevent shearing the guide wire off inside the vessel.
16. Secure the catheter to the skin using silk (2.0) or nylon (4.0) sutures. Take a 0.5 cm bite of skin with the suture needle. If the catheter assembly contains integrated "wings" for fixation, thread suture through the perforated wings and secure catheter against the skin with several knots. If no fixation device is included, tie several knots in the suture without pinching the skin, leaving both ends of the suture long. Using the loose ends of the suture, tie a second set of knots around the hub of the catheter, firmly, but without constricting its lumen.

Fig. 1.7 Anesthetic injection over the femoral artery

Fig. 1.8 Insertion of guide wire into femoral artery

1.4 Complications

- Hemorrhage
- Hematoma (at puncture site)
- Infection (at insertion site or systemic)
- Thrombosis
- Arteriovenous fistula
- Pseudoaneurysm formation
- Exsanguination (secondary to dislodgement of catheter)
- Cerebrovascular accident (CVA; secondary to air embolism)

1.5 Pearls and Pitfalls

- The shorter and stiffer the plastic tubing connected to the arterial cannula for blood pressure monitoring, the higher its frequency response and the accuracy of measurements.
- Use of an ultrasound probe can facilitate artery location and vessel cannulation.
- Puncture of the femoral artery proximal to the inguinal ligament, or distal to its bifurcation into superficial femoral and deep femoral arteries, may cause massive hemorrhage due to poor vessel compressibility in these regions; the artery should therefore be cannulated just distal to the inguinal ligament, where it is easily compressible against the femoral head if necessary.
- If difficulty is encountered when advancing an over-the-needle catheter into the artery, attach a 10 mL syringe containing 5 mL of sterile normal saline to the catheter hub, aspirate 1 or 2 mL of blood to confirm catheter tip placement within the vessel lumen, and then advance the catheter while gently injecting the saline-blood mixture; the jet of fluid momentarily dilates the lumen, aiding advancement of the catheter.
- An alternative approach to the over-the-needle catheter that will not fully advance is the use of a guide wire. After intraluminal placement of the cannula tip is confirmed by blood return, a guide wire is gently inserted through the catheter into the artery. The cannula is then passed along the guide wire until fully advanced. The guide wire employed must have a blunt, flexible tip to minimize the possibility of vessel wall trauma.
- Two potential consequences of arterial cannulation are vessel obstruction secondary to intravascular thrombosis and hemorrhage (the latter being the most common complication). Choice of puncture site is therefore essential. Due in part to their generous collateral blood flow, as well as ease of compressibility, the radial and femoral are the two most commonly cannulated arteries.
- Repeated puncture following unsuccessful cannulation increases the risk of arterial obstruction secondary to vessel wall damage and thrombosis.

- Double puncture of the cannulated artery by inadvertent over-insertion of the needle has not been shown to increase complications despite the additional trauma to the vessel walls.

Selected Reading

Anderson JS. Arterial cannulation: how to do it. Br J Hosp Med. 1997;57:497.

Gilchrist IC. Reducing collateral damage of the radial artery from catheterization. Catheter Cardiovasc Interv. 2010;76:677–8.

Lemaster CH, Agrawal AT, Hou P, Schuur JD. Systematic review of emergency department central venous and arterial catheter infection. Int J Emerg Med. 2010;3:409–23.

Mitchell JD, Welsby IJ. Techniques of arterial access. Surgery. 2004;22:3–4.

Wilson SR, Grunstein I, Hirvela ER, Price DD. Ultrasound-guided radial artery catheterization and the modified Allen's test. J Emerg Med. 2010;38:354–8.

2 Ultrasound-Guided Peripheral Intravenous Access

Coben Thorn and L. Connor Nickels

2.1 Indications

- Difficulty placing peripheral intravenous (PIV) using traditional methods of direct visualization and palpation
- To reduce needle sticks in a hypercoagulable patient

2.2 Contraindications

- Patient needs emergent central venous access

2.3 Materials and Medications

- Ultrasound machine
- High-frequency (5–8 MHZ) linear probe (Fig. 2.1)
- Ultrasound gel
- Minimum 1.5-in. needle length
- IV setup
- Skilled operator

Fig. 2.1 High-frequency (5–8 MHZ) linear probe

C. Thorn, MD
Department of Emergency Medicine, Bon Secours St. Francis Health System, Greenville, SC, USA
e-mail: cobenthorn@gmail.com

L.C. Nickels, MD, RDMS (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_2

2.4 Procedure

1. Scan the selected area to identify a target vessel for PIV cannulation.
   - Basilic (runs on the medial side of upper arm) (Fig. 2.2) and cephalic (runs on the lateral side of the upper arm) veins are good superficial veins that are generally not seen without ultrasound. The deep brachial vein is also an option; however, this is a deep vein that runs with the brachial artery, and there is higher chance for complications (Fig. 2.3).
2. Now the site should be prepared for IV insertion. It should be cleaned. An appropriate gauge long needle should be selected, and IV setup should be conveniently located for when access is obtained.
3. The ultrasound probe should be placed in the transverse plane (Fig. 2.4) to best visualize surrounding structures and vein. Alternatively, the probe can be placed longitudinally (Fig. 2.5) for better visualization of needle depth and slope.
   - Most practitioners seem to prefer the transverse approach.
4. The concept of the Pythagorean Theorem is used for accuracy. The needle should be inserted at a 45° angle to the skin and at an equal distance back from the probe as the approximate depth of the vessel vertically. The depth is given on the screen, usually at the bottom in centimeters. As soon as the needle has penetrated the skin the needle tip should be located by fanning the probe toward the needle until it is identified. The needle should then be advanced slowly always keeping the needle tip in view. Once directly on top of the vein, it should tent with pressure and then the needle should be inserted into the vein. Follow the needle in the vein as far as possible while keeping the tip of the needle in the center of the vein to make sure the catheter is securely in the vein and does not infiltrate.

Fig. 2.2 Basilic vein (BV) located medially when scanning proximally from the antecubital fossa

Fig. 2.4 Needle tip in vein seen in a transverse orientation

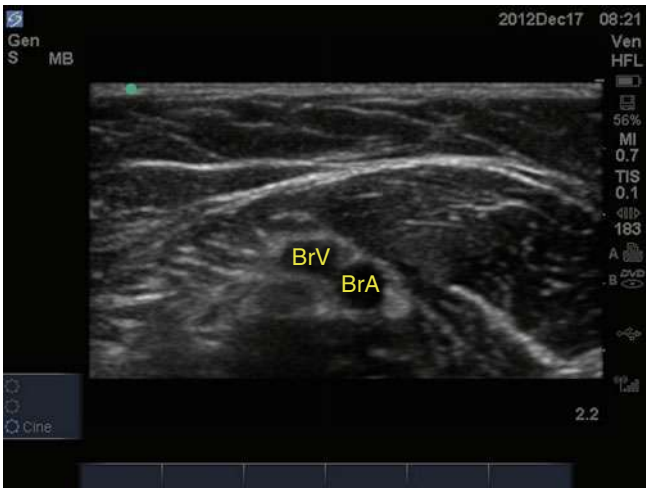

Fig. 2.3 Brachial artery (BrA) and vein (BrV). The less round, slightly compressed, anechoic structure on the left is the vein. The very circular, not compressed, anechoic structure to the right is the artery

Fig. 2.5 Needle tip in vein seen in a longitudinal orientation

2.5 Complications

Inadvertent puncture of artery. Veins should be thin walled, compressible, and have no pulsations.

2.6 Pearls and Pitfalls

- Mistaking the midshaft of the needle for the needle tip. If this occurs, the needle tip is actually deeper than expected. The ultrasound machine will plot a hyperechoic "dot" on the screen for the needle tip, as long as it crosses the ultrasound beam at any point. This same "dot" will appear whether the tip is directly centered under the beam or *any* segment of the needle shaft is intersecting the beam. This can be visually deceiving and makes this procedure difficult to grasp.
- Very slow movements of the needle and the probe are important for keeping the needle tip in view. Once the needle tip is identified, the probe should be fanned forward (away from operator) just slightly and then the needle advanced until the needle tip comes into view again. This is repeated until the needle is securely moved further into the vein.

Selected Reading

Brannam L, Blaivas M, Lyon M, et al. Emergency nurses' utilization of ultrasound guidance for placement of peripheral intravenous lines in difficult-access patients. Acad Emerg Med. 2004;11:1361–3.

Constantino TG, Parikh AK, Satz WA, et al. Ultrasonography-guided peripheral intravenous access versus traditional approaches in patients with difficult intravenous access. Ann Emerg Med. 2005;46(5):456–61.

Dargin JM, Rebholz CM, Lowenstein RA, et al. Ultrasonography-guided peripheral intravenous catheter survival in ED patients with difficult access. Am J Emerg Med. 2010;28(1):1–7.

Ma JO, Mateer JR, Blaivas M. Emerg ultrasound. 2nd ed. New York: McGraw-Hill Companies, Inc.; 2008.

Saul T, Rivera M, Lewiss R. Ultrasound image quality. ACEP News. 2011;4:24–5.

3 Central Venous Line Placement: Internal Jugular Vein, Subclavian Vein, and Femoral Vein

Kevin D. Ergle, Zachary B. Kramer, Jason Jones, and Rohit Pravin Patel

3.1 Indications

- Volume replacement
- Emergent venous access
- Administration of caustic medications: vasopressors, calcium chloride, hypertonic saline, high dose of potassium
- Dialysis catheter placement (hemodialysis)
- Nutritional support (total parenteral nutrition)
- Long-term antibiotics
- Chemotherapy
- Plasmapheresis
- Frequent or persistent blood draws or intravenous therapy when unable to establish peripheral access due to edema or other causes
- Jugular and subclavian: Central venous pressure monitoring, transvenous pacing wire introduction, pulmonary artery catheterization

3.2 Contraindications

- Absolute
  - Infection at site of insertion
  - Distorted anatomy/landmarks (prior surgery, radiation, or history of thrombus in the specified vein)
  - Subclavian only: Trauma to the ipsilateral clavicle, anterior proximal rib, subclavian or superior vena cava vessels
- Relative
  - Morbid obesity
  - COPD
  - Children less than 2 years (higher complication rates)
    - Coagulopathy (although ultrasound-guided internal jugular can be done in this situation)
  - Agitated or moving patient
  - Jugular only: Trauma to the ipsilateral clavicle, anterior proximal rib, subclavian or superior vena cava vessels
  - Jugular and subclavian: Inability to tolerate potential pneumothorax of the ipsilateral thoracic cage
    - Pneumothorax or hemothorax of the contralateral thorax
    - Patients receiving ventilatory support with high-end expiratory pressures (temporarily reduce the pressures)
  - Femoral only: Intra-abdominal (or retroperitoneal) hemorrhage

3.3 Materials and Medications

- Central venous catheter tray or bundle: single/double/triple/quadruple lumen, dialysis catheter, large-bore introducer (for transvenous pacing or pulmonary artery catheter kit)
- Sterile gloves
- Sterile drapes or towels
- Sterile gown
- Hat/hair cap and mask with eye protection
- Antiseptic solution with skin swabs (e.g., chlorhexidine)
- Sterile saline flushes (one 30 mL syringe or three 10 mL syringes)
- Lidocaine 1 %

K.D. Ergle, MD
Department of Internal Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: kevin.ergle@medicine.ufl.edu

Z.B. Kramer, MD
Emergency Medicine Resident, University of Florida College of Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: zkramer@ufl.edu

J. Jones, MD • R.P. Patel, MD (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: jasonjones@ufl.edu; rohitpatel@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_3

- Sterile gauze
- No. 11 blade scalpel
- Dressing (sterile waterproof transparent dressing or sterile 4×4 gauze with tape)
- Sterile biopatch
- Suture material with needle driver if needed
- Transducing line (optional)
- Sterile probe cover (if using ultrasound guidance)

3.4 Procedures

3.4.1 Internal Jugular Vein Access Procedure

1. Obtain informed consent if not emergently indicated procedure.
2. Obtain supplies and prepare the room, ensuring that all supplies are within operator reach prior to placing gown and commencing the procedure. Include a sterile ultrasound sheath on the sterile field if ultrasound is being used.
3. Raise bed to a comfortable height for the operator.
4. Place patient with head facing away from side of central line site (if using ultrasound, other positions may be preferred). Place patient in 15–20° Trendelenburg position to help fill the upper central veins and reduce the risk of air embolism.
5. Identify the anatomy. Palpate triangle made by the clavicle and sternal and clavicular heads of the sternocleidomastoid (SCM) muscle to identify the location of the internal jugular vein (Fig. 3.1). If using ultrasound guidance, identify optimal anatomical arrangement.
6. Wash your hands and wear sterile attire using aseptic technique, including cap, mask, gown, and sterile gloves.
7. Prepare the site from the clavicle to the ear and across the trachea with antiseptic solution. Allow the antiseptic (chlorhexidine or iodine) to fully dry.
8. Drape the site and patient with sterile towels and drapes included in most CVL bundles. Make sure to cover the whole area and bed.
9. Cover the ultrasound probe with a sterile sheath. This can be done solo or by holding the sterile ultrasound sheath and having an unsterile assistant hold the probe so that the probe can be covered by the sheath.
10. Prepare the kit by checking the guide wire and flushing the tubing and lines with saline included in the kit.
11. With a 25-gauge needle, use 1 % lidocaine to anesthetize the skin at the apex of the triangle made by the SCM and clavicle. Aspirate to make sure the operator is not in a vessel and make a superficial wheel for the insertion site.
12. Preferred method is with ultrasound guidance (see steps 13–17 and Sect. 3.6 for description of ultrasound guidance). If performing without ultrasound, palpate the carotid artery and insert needle lateral to the artery at the apex of the triangle formed by the SCM, aiming toward the ipsilateral nipple at an angle 30–45° above the horizontal plane (Fig. 3.1). Once blood returned, go to step 18.
13. Place sterile ultrasound gel over the insertion site. Use the ultrasound to identify vessel anatomy including internal jugular vein and carotid artery. Use the ultrasound probe to compress the vein, which is compressible as opposed to the carotid artery, which is not compressible (Fig. 3.2).
14. Prepare the insertion needle and syringe (if long and short needles are available, a short needle may be used to reduce posterior vein perforation) and prime the syringe by pulling back on the plunger prior to making the puncture.
15. Use the ultrasound probe to re-identify the patient's anatomy.
16. Ultrasound can be used in short axis or long axis (Fig. 3.3). Short axis is easier for novice operators due to increased ability to see the artery and vein but has a higher risk of posterior perforation if the needle tip is not visualized well. Once short axis of the vein is found, turning the probe 90° clockwise allows the operator to see the vein in long axis. The needle is better visualized in this view but technically more difficult and has less chance to penetrate the posterior wall. In patients with short necks, it may be difficult to obtain long-axis view and needle insertion in the limited space.
17. Insert the needle using the ultrasound guidance with dynamic approach preferred (see Sect. 3.6 for specifics). Make sure to aspirate while inserting the needle to identify when the venous access is obtained. The needle tip should be visualized through the whole process.
    - If using the static approach (see Sect. 3.6), insert needle lateral to carotid pulsation as this is where the vein anatomically is located. Standard method is to insert the needle as far back as the depth the vessel is visualized (e.g., if the vein is visualized 2 cm below skin surface, the needle should be inserted 2 cm behind the probe at a 45° angle).
    - If inserting the needle ~3 cm does not achieve access, gently withdraw the needle toward the surface of the skin while aspirating. Avoid withdrawing the needle completely from the skin. If needed, redirect the needle and advance until blood is aspirated. Cannulation of the vein often takes place while withdrawing the needle.
18. Hold the needle steady with your nondominant hand and remove the syringe, careful not to advance or withdraw the needle. You can place the base of your hand on the patient's chest to make your hand more stable during

this part of the procedure. Occlude the hub of the needle to prevent air embolus.

19. You may verify that you are in the vein by transducing pressure with a fluid column. The fluid should flow easily into the vein.
    - If the aspirated blood is pulsatile and moves up the column, withdraw the needle completely and apply pressure for 10–20 min while taking the patient out of Trendelenburg position (if nonemergent procedure).
20. Once it is verified that you are in the vein, insert the J-tip of the guide wire into the needle hub and advance into the vein. The J-tip can be straightened with a pinching motion (Fig. 3.4). Always keep one hand on the guide wire until it is removed from the patient. Monitor for arrhythmias as the guide wire is advanced toward the right atrium.
    - If the guide wire does not flow easily, remove the guide wire and reattach the syringe, checking for blood flow.
    - If arrhythmia occurs, slowly withdraw the guide wire until the patient's native rhythm returns.
    - Alternatively, the catheter/syringe found in most kits can be used as a bridge to guide wire placement. For the author, this has improved success when there is difficulty in wire placement. Use the same steps above with the catheter (Fig. 3.5) and when you have return of blood, advance the angiocath into the vein followed by insertion of the guide wire through the angiocath. This is especially useful in moving/agitated patients, patients who have collapsible veins due to hypovolemia, and patients who have abnormal anatomy and may have veins that take an abnormal angle shortly past the needle tip.
21. Remove the needle over the guide wire, making sure to always keep control of the guide wire.
22. Make an incision contiguous with the guide wire using a straight (No. 11) blade with the scalpel blade facing upward (away from the wire).
23. Advance the dilator over the guide wire in a twisting motion, keeping control of the guide wire.
    - The dilator only needs to go slightly beyond the anticipated depth of the patient's jugular vein. Do not advance the entire length of the dilator.
24. Withdraw the dilator and hold pressure over the wound site.
25. Advance the catheter over the guide wire while keeping control of the guide wire.
26. With the catheter inserted 10–12 in. from the skin insertion site, retract the guide wire until it comes out of the distal port. Maintain control of the guide wire and advance the catheter to the appropriate length. Usually catheters are inserted 15–16 cm from the right side and 18–20 cm from the left side (Fig. 3.6).
27. Flush each port of the catheter and check aspiration. If difficulty with aspiration or flushing, concern is raised for catheter malposition. The operator can change the depth slightly or twist the catheter and recheck.
28. At this time, an antibiotic ointment or biopatch may be applied to skin around the intersection with the lumen of the catheter. This step is based on local institutional guidelines.
29. Suture the line in place.
30. Enclose CVL site with sterile waterproof transparent dressing.
31. Confirm placement using chest X-ray (CXR). The tip of the catheter should be in the lower third of the superior vena cava (SVC) at the insertion of the SVC into the right atrium.

Fig. 3.1 Internal jugular blind approach; this would be the same location for probe placement if doing ultrasound guided

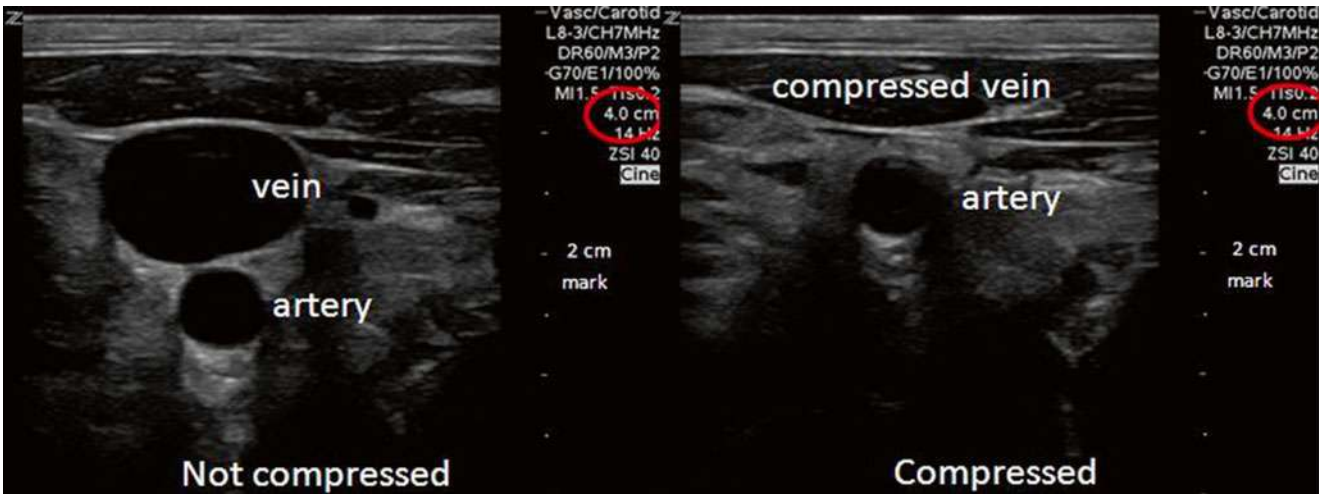

Fig. 3.2 Ultrasound showing internal jugular vein and artery with and without compression

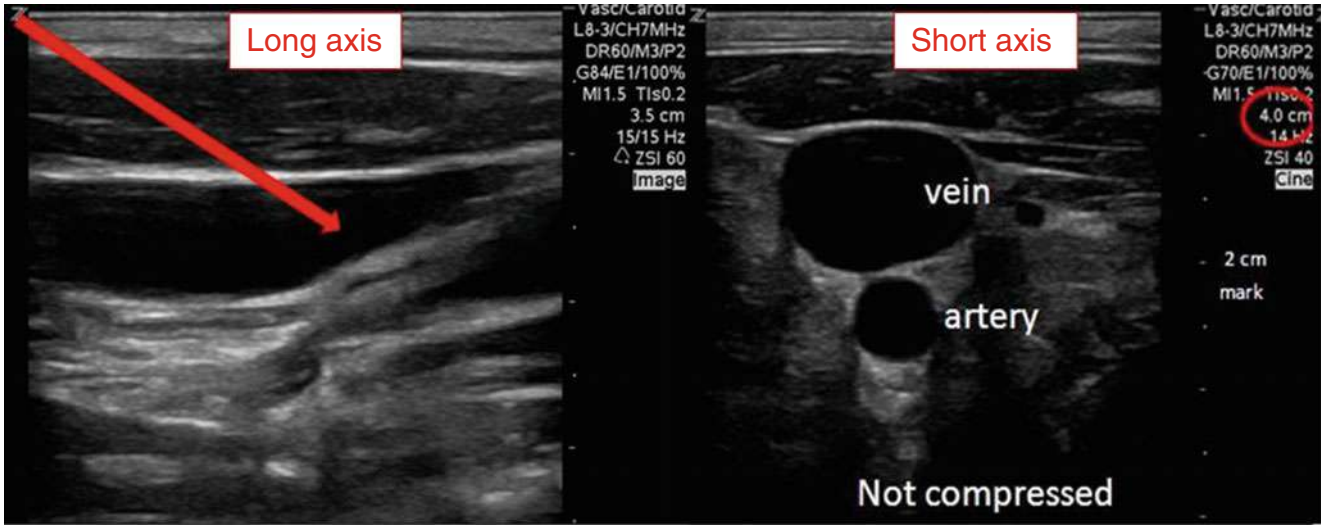

Fig. 3.3 Long- and short-axis views of the internal jugular vein

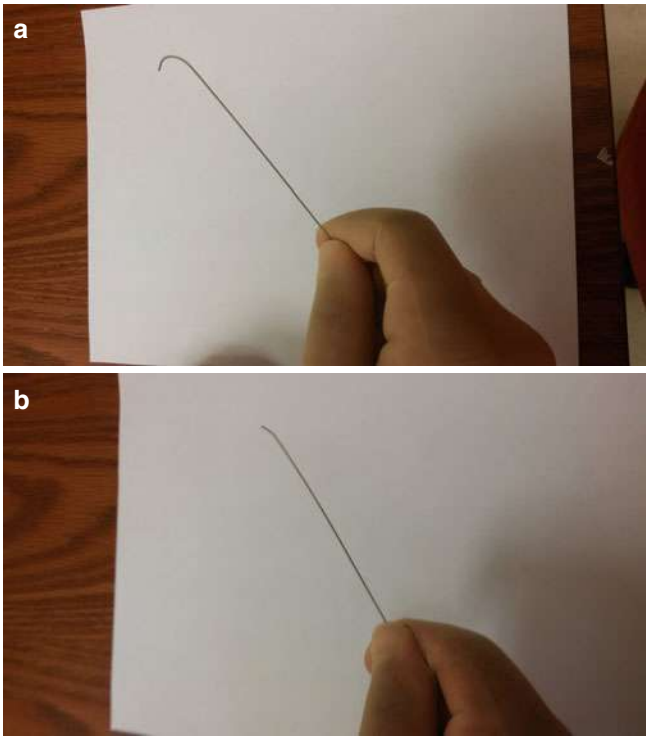

Fig. 3.4 (**a**, **b**) J-tip straightening using pinch/stretch method

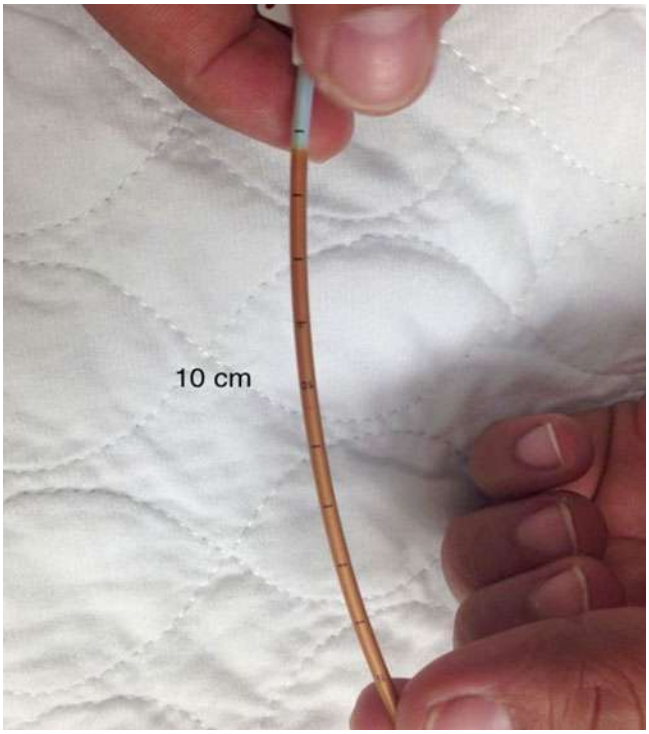

Fig. 3.6 Length: marking seen on typical central venous catheters; number indicates distance in centimeters from distal tip

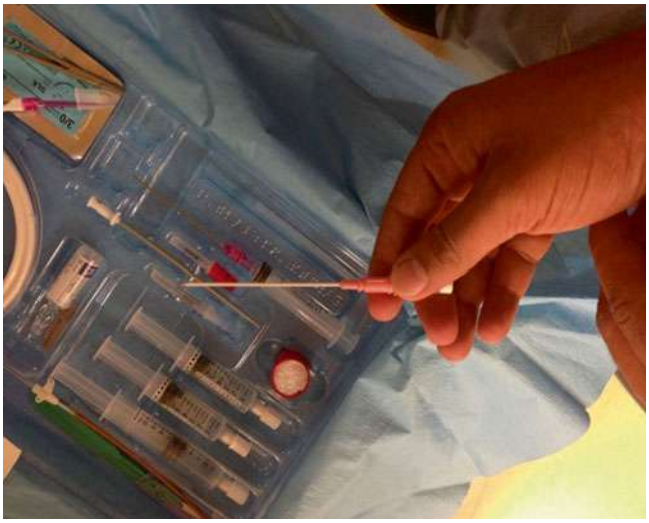

Fig. 3.5 Angiocath that can be used in difficult to cannulate/wire patients

3.4.2 Subclavian Vein Access Procedure

1. Obtain informed consent if not emergently indicated.
2. Raise bed to comfortable height for the operator.
3. Place patient in supine position and position so patient's head is at the top of the bed.
4. Place patient in 15–20° Trendelenburg position (if tolerated) to reduce risk of air embolism. Studies show this will also increase the size of the subclavian vein. Do NOT place towel between shoulder blades (arch shoulder back) as this has been shown to decrease vein diameter and affect reliability of accessibility. Keep shoulders at anatomical location (forward).
5. Prep area chosen from the anterior neck, clavicle, and upper chest (above nipple line) with chlorhexidine prep or iodine.
6. Open kit and place close to operator's dominant hand to allow for easy access. Diameter of catheter/kit used based on clinical situation:
   - Introducer or large bore if requiring large volumes of resuscitation
   - Triple lumen catheters for vasopressors
   - Introducers if anticipating pulmonary artery catheter or venous pacer
7. Operator should prepare with all aseptic techniques (e.g., hand washing) and maximal barrier precautions (e.g., sterile gowns, sterile gloves, caps, masks covering both mouth and nose, eye protection, and full-body patient drapes).
8. Once sterile and able to touch the inside of the CVL kit, the operator may want to retract the curved J-tip wire into the plastic loop sheath for easy directing into the introducer needle. The operator should also uncap all distal lumens and flush all ports with 3–5 cc of the sterile NS syringes to ensure no defects in the lumen of the catheter. Close all ports except the distal tip port (usually marked with words "distal tip") with the slide clamp.
9. Prep area chosen (right or left side) from the anterior neck, clavicle, and upper chest (above nipple line) with sterile chlorhexidine prep (this is the second cleaning).
10. Place full-body drape over patient with opening over selected side where needle will be inserted.
11. Needle insertion site options (Fig. 3.7):
    - One centimeter inferior to the junction of the middle and medial third of the clavicle
    - Just lateral to the midclavicular line, with the needle perpendicular along the inferior lateral clavicle
    - One fingerbreadth lateral to the angle of the clavicle
12. Anesthetize needle insertion site with 5–10 mL of 1 % lidocaine superficially (make sure to pull back on needle syringe to ensure operator is not in the vein or artery).
    - Never place equipment on a patient.
13. Prepare the needle and syringe by placing the long needle on the syringe. Make sure to break seal of syringe by pulling back on the plunger of the syringe prior to making incision with needle.
14. Turn patient's head to opposite of CVL placement and retract ipsilateral shoulder down to improve clavicle-vein relationship. The retraction of the arm can be done a few steps earlier and can be held in position using a person or tape/restraints.
15. Direct the insertion needle toward sternal notch in the coronal plane at an angle no greater than 10–15° while gently withdrawing the plunger of the syringe. Keep bevel of the needle facing up and in line with the numbers on the syringe until operator enters skin, then face bevel caudally to facilitate smooth progression of the guide wire down the vein toward the right atrium.
16. It helps to place nondominant hand (not holding the needle) on the sternal notch so operator can feel where sternal notch is and direct needle in that direction (Fig. 3.7).
17. NEVER increase the angle of the needle greater than 15° as pneumothorax may ensue.
18. Advance the needle under and along the inferior border of the clavicle, making sure the needle is virtually horizontal to the chest wall. Aim medially in the direction of the suprasternal notch, attempting to first aim for the clavicle then "walk" the needle below the clavicle.
19. Once under the clavicle, continue to advance the needle in a plane almost parallel to the skin approximately 2–3 cm until venous blood is freely aspirated into the syringe.
20. When venous blood is freely aspirated, disconnect the syringe from the needle, and immediately occlude the lumen to prevent air embolism and insert the guide wire. If the vein is difficult to locate, remove the introducer needle, flush it clean of clots, and try again. Change insertion sites after three unsuccessful passes with the introducer needle.
21. At this point, the hand holding the needle should be "set in stone." Use the patient's chest wall as a base to keep needle completely still as to not inadvertently advance or retract needle out of the vein.
22. Insert the guide wire through the needle into the vein with the J-tip directed caudally to improve successful placement into the subclavian vein.
    - Beware a return of red pulsatile blood. If this occurs, the wire is in an artery.
    - Beware aspirating air bubbles through the probing introducer needle. This indicates a pneumothorax.
23. Advance the wire until it is mostly in the vein or until arrythmia is seen on the cardiac monitor. Then, retract the wire 3–4 cm.
24. If the wire does not pass easily, remove the wire, reattach the syringe, and confirm that the needle is still in the

lumen of the vein before reattempting. The J-tip can be straightened with a pinching motion (Fig. 3.4).

- Alternatively, the catheter/syringe found in most kits can be used as a bridge to guide wire placement. For the author, this has improved success when there is difficulty in wire placement. Use the same steps above with the catheter (Fig. 3.5) and when you have return of blood, advance the angiocath into the vein followed by insertion of the guide wire through the angiocath. This is especially useful in moving/agitated patients, patients who have collapsible veins due to hypovolemia, and patients who have abnormal anatomy and may have veins that take an abnormal angle shortly past the needle tip.

25. Use the tip of the scalpel to make a small incision just against the needle to enlarge the catheter entry site for the dilator and catheter.
26. Holding the wire in place, withdraw the introducer needle and place in needle holder.
27. Thread the dilator over the wire and into the vein with a firm and gentle twisting motion while maintaining constant control of the wire. If a large-bore introducer is placed, the dilator/introducer goes in one step, after the introducer is inserted, hold the wire in place and remove the dilator.
28. If operator is having difficulty threading the dilator, the skin incision with the scalpel may have been too superficial or small. It may help to enlarge this incision to avoid having the dilator get caught on superficial skin or connective tissue.
29. It is helpful to have sterile gauze handy to apply pressure with the hand not holding the wire as the vein will now bleed profusely from around the wire secondary to dilation.
30. Thread the catheter until it is close to the skin insertion site. Then pull back on the guide wire until it shows outside of the distal port. Grasp the wire outside of the distal port and thread the catheter while holding onto the guide wire. Usually catheters are inserted 15–16 cm from the right side and 18–20 cm from the left side (Fig. 3.6)
31. Hold the catheter in place and remove the wire. After the wire is removed, occlude the open lumen.
32. Attach sterile saline syringe to the hub and aspirate blood. Take needed samples and then flush the line with saline and recap. Repeat this step with all lumens.
33. Place biopatch on skin around the intersection with the lumen of the catheter.
34. Suture the line in place.
35. Enclose CVL site with sterile waterproof transparent dressing.
36. Confirm placement using CXR. Tip of catheter should be in the lower third of the SVC at the insertion of SVC into right atrium (tip at right bronchiotracheal angle or up to 2.5 cm below bronchiotracheal angle).
    - Alternatively, ultrasound can be used for subclavian line access but only a few limited studies have confirmed this as to date so will not describe in detail (see below for typical ultrasound technique used). See references for more information.

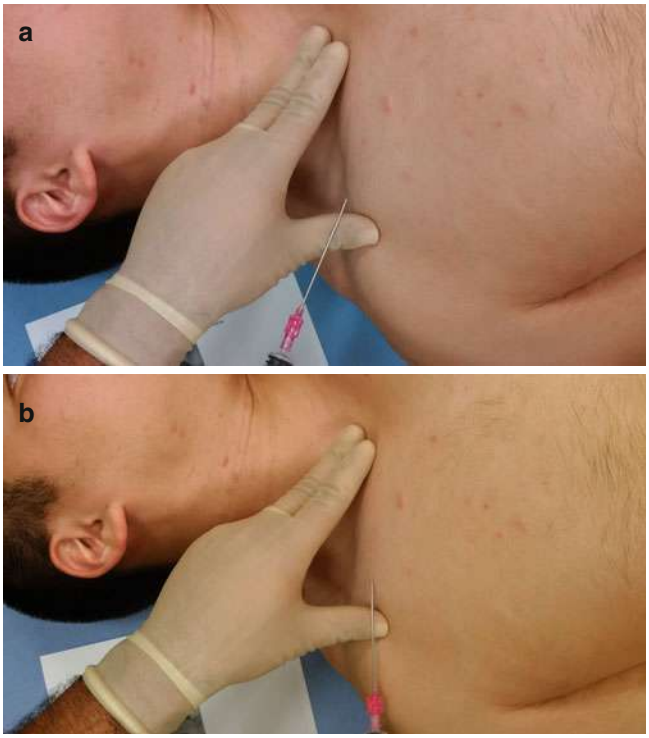

Fig. 3.7 Subclavian vein approach: wrong (**a**) and correct (**b**) angles to take when making skin puncture

3.4.2.1 Subclavian Vein Pearls and Pitfalls

- Inadequate landmark identification: operator should always palpate for landmarks and check anatomy prior to starting the procedure.
- Improper insertion position.
- Insertion of needle through periosteum.
  - Operator should NOT increase angle of needle to avoid the clavicle bone (this can cause a pneumothorax).
  - Operator should press on needle with downward pressure on chest wall to allow needle to maneuver under the clavicle without changing the angle of insertions of the needle.
- Taking too shallow a trajectory with needle.
- Aiming the needle too cephalad (aim for sternoclavicular junction).
- Failure to keep needle in place for wire passage: hand holding the needle should be planted on patient's chest for stabilization.

3.4.3 Femoral Vein Access Procedure

1. Palpate the patient's femoral artery below the inguinal ligament. This is usually found halfway between the anterior superior iliac spine (ASIS) and the midline of the symphysis pubis.
2. Trim overlying hair as necessary.
3. If ultrasound-guided approach is desired, use the linear probe (same as internal jugular) to detect the femoral vein at this location. The femoral vein will be easily compressible, while the femoral artery will be less compressible and pulsatile (Fig. 3.2).
4. Wash hands and use sterile technique to apply iodine or chlorhexidine solution (various forms available).
5. Open your femoral CVL kit and don cap, mask, sterile gown, and sterile gloves. Nonsterile assistants should wear a cap, mask, sterile gown, and sterile gloves. Flush all ports of your CVL kit with saline flushes and check for leaks or malfunction of catheter.
6. Under sterile technique, apply the drape over the area of insertion, and have an assistant extend the drape the length of the bed. Reapply sterile iodine or chlorhexidine at the site.
7. Anesthetize the skin overlying the femoral vein with lidocaine.
8. If ultrasound-guided approach is desired, have an assistant hold up the vascular ultrasound probe. Place your gloved hand through a sterile ultrasound sleeve and grasp the top of the ultrasound probe. Without breaking sterile technique, pull and invert the sterile ultrasound sleeve over the probe and cable. See Sect. 3.6 for detailed ultrasound-guided cannulation technique.
9. Insert the needle at a 45° angle, bevel down, directed superiorly, 1 cm medial to the palpable femoral artery pulse. Once the needle has broken the skin, aspirate by applying a small amount of continuous traction on the plunger of the attached syringe.
10. Advance smoothly and slowly until blood appears in the syringe. Stop once blood is aspirated. If femoral vein is not cannulated, withdraw your needle until just beneath the skin and redirect.
11. Hold the hub of the needle with thumb and forefinger to immobilize in place. Remove the syringe carefully. If blood appears arterial or pulsatile, remove the needle and hold pressure for 5–10 min.
    - Use the palm of your hand on the thigh to stabilize your hand. Not having your hand stable is a common mistake leading to needle movement out of the vein.
12. If blood appears venous (dark color, emerges as a continuous trickle, or transduced), cannulate the needle with the guide wire. Maintain a two-finger grip on the guide wire at all times. Advance the guide wire until approximately 15 cm remains.
    - Must keep handle of guide wire at ALL times and can be done through proper technique.
13. If guide wire does not advance easily, remove guide wire and reposition needle until blood aspirates easily. The J-tip can be straightened using a pinching motion if needed (Fig. 3.4)
    - Alternatively, the catheter/syringe found in most kits can be used as a bridge to guide wire placement. For the author, this has improved success when there is difficulty in wire placement. Use the same steps above with the catheter (Fig. 3.5) and when you have return of blood, advance the angiocath into the vein followed by insertion of the guide wire through the angiocath. This is especially useful in moving/agitated patients, patients who have collapsible veins due to hypovolemia, and patients who have abnormal anatomy and may have veins that take an abnormal angle shortly past the needle tip.
14. Using your scalpel, make a single 1/2 cm stab incision at the site of needle insertion to assist with dilator placement.
    - You can do the stab incision with or without the needle in place, but from experience, most novices have difficulty locating the correct stab location due to small amount of blood accumulation when the needle is taken out.
15. Remove your needle, carefully leaving guide wire in place. Apply dilator over guide wire and advance into the body with gentle pressure and a twisting motion in the same plane that you used to direct the needle.
16. Holding pressure at the insertion site with sterile gauze 4×4 pads, remove the dilator from the guide wire,

leaving the guide wire in place. Insert the central venous catheter over the guide wire until it fits snugly against the skin. Use the markings on the catheter to determine proper length placed (usually with femoral site you can "hub" the catheter) (Fig. 3.6).

17. Using a sterile saline flush, ensure that each lumen of the CVL draws blood easily and flushes easily. Carefully note any that do not and consider repositioning as needed. You can attempt moving catheter in or out a few centimeters or rotating the catheter and recheck. Apply caps to each open lumen of the CVL.
    - Remember to cover the introducer port if not using immediately. If you do not, it poses an infection and air embolism risk to the patient.
18. Suture the line in place.
19. Place an antibiotic biopatch or similar antimicrobial dressing.
20. Enclose CVL site with sterile waterproof transparent dressing.

3.4.3.1 Femoral Vein Pearls and Pitfalls

- Femoral central venous lines cannot accurately transduce central venous pressures.
- Asking the patient to perform a Valsalva maneuver has been shown to increase the width of the femoral vein by 1/3.
- The mnemonic NAVEL (Nerve Artery Vein Empty space Lymphatics) assists in remembering the order of femoral structures from lateral to medial.
- Traditionally, femoral venous lines were thought to have higher rates of infection than subclavian or internal jugular lines, but more recent analyses are challenging this belief.
  - Obesity is a more important risk factor for infection in femoral sites.

3.5 Complications

3.5.1 Jugular and Subclavian Complications

- Pneumothroax/Hemothorax
  - Prevention: Remove patient from ventilator before advancing the needle, choose the right side rather than left, and avoid multiple attempts when possible.
  - Management: Check postprocedure x-ray; if pneumothorax, arrange for thoracostomy depending on the size of the hemo-/pneumothorax.
- Catheter embolization
  - Prevention: Never withdraw a catheter past a needle bevel as this might shear off the catheter.
  - Management: X-ray the patient and contact specialist who can remove the embolized catheter.
- Arterial puncture: Hold compression if this occurs.
- Hematoma: Usually requires monitoring only.
- Thrombosis: This complication may lead to pulmonary embolism.
- Local site or systemic infection: Using maximal sterile precautions has been shown to greatly decrease rate of infection.
- Air embolism
  - May be caused by negative intrathoracic pressure when inspiration by the patient drawing air into an open line hub
  - Prevention: Be sure the line hubs are always occluded; placing the patient in the Trendelenburg position lowers the risk.
  - Management: The patient should be placed in Trendelenburg position with a left lateral decubitus tilt, which may prevent the movement of air into the right ventricle and onward into the left side of the heart. One hundred percent oxygen should be administered to speed the resorption of the air. If a catheter is located in the heart, aspiration of the air should be attempted.
- Dysrhythmias: Due to cardiac irritation by the wire or catheter tip. This can usually be terminated by simply withdrawing the line into the superior vena cava. One should always place a central venous catheter with cardiac monitoring.
- Lost guide wire: If the operator is not careful about maintaining control of the guide wire, it may be lost into the vein. This requires retrieval by interventional radiology or surgery and is an emergency.
- Catheter tip too deep: Check the postprocedure chest radiograph and pull the line back if the tip disappears into the cardiac silhouette.
- Catheter in the wrong vessel: Check the postprocedure chest radiograph for this complication; remove catheter and try again.
- Arterial puncture (subclavian only): The subclavian artery cannot be compressed; so, the subclavian approach should be avoided in anticoagulated patients.

3.5.2 Femoral Complications

- Arterial puncture: The femoral artery site can be compressed, so if punctured hold pressure.
- Hematoma: Usually requires monitoring only.
- Thrombosis: This complication may lead to pulmonary embolism.
- Catheter embolization.
  - Prevention: Never withdraw a catheter past a needle bevel which might shear off the catheter.
  - Management: X-ray the patient and contact specialist who can remove the embolized catheter.

- Lost guide wire: If the operator is not careful about maintaining control of the guide wire, it may be lost into the vein. This requires retrieval by interventional radiology or surgery and is an emergency.
- Local site or systemic infection: Using maximal sterile precautions has been shown to greatly decrease rate of infection.

3.6 Ultrasound Guided Cannulation: Tips for Each Approach

- Venous anatomy is best visualized using high-frequency (5–10 MHz) linear probe. Higher frequencies generate less penetration but better resolution.
- You can use the ultrasound to identify the location of the vessel prior to the procedure and utilize external landmarks during the procedure itself (static technique), or you can use the ultrasound to visualize cannulation of the vessel during the procedure (dynamic technique).
- Static view is advantageous in that the ultrasound transducer is not needed during the sterile portion of the procedure, but it does not allow for direct visualization of cannulation and guidance during the procedure.
- Dynamic view (preferred) allows for direct visualization during the procedure but requires more technique and requires use of transducer during the sterile portion of the procedure.
- The dynamic technique can be used in either a short-axis view, where a cross-sectional view of the vessel and needle is used, or a long-axis view, where a longitudinal view of vessel and needle is used (Fig. 3.3).
- The long-axis view allows for full visualization of the needle throughout the procedure and allows for better visualization and adjustment of needle depth. It is more difficult for lateral changes in positioning and tends to be more difficult technically.
  - Key in this view is that once a good section of vein is obtained, do not move probe to visualize the needle; move the needle into the ultrasound view by slightly adjusting trajectory.
- The short-axis view allows for lateral changes in position but is not as good at visualizing depth throughout the procedure, as visualization of the needle is in cross-sectional imaging. Perforation of the posterior wall is more common in this view.
- When using the short-axis view, remember to position the ultrasound probe such that the field of the ultrasound intersects the vessel (internal jugular, subclavian, femoral) at the anticipated site of insertion of the needle into the vein. Remember that the needle is only visualized as it intersects the plane of the ultrasound.
- When using the long-axis view, make sure to visualize the vessel with the ultrasound such that you can see the greatest diameter of the vessel along the entire length of the ultrasound probe. Keep the ultrasound steady during the procedure, and insert the needle at an angle at the lateral edge of the ultrasound probe. Using this technique, one can visualize the entire length of the needle.

3.7 Removing a Central Line

1. Place patient in supine or Trendelenburg position (for femoral removal can help decrease bleeding).
2. Remove suturing and dressing.
3. Jugular and subclavian: Have patient exhale and pull the line during the exhalation.
   - Exhalation increases intrathoracic pressure as compared to atmospheric pressure, thereby reducing the risk of air thromboembolism.
4. Hold pressure for approximately 1 min to stop bleeding.
5. Dress with a sterile dressing.
6. If central line-related infection is suspected, cut off the tip with sterile scissors and send for culture.

Selected Reading

Internal Jugular Vein Access

McGee DC, Gould MK. Preventing complications of central venous catheterization. N Engl J Med. 2003a;348(12):1123–33.

Mimoz O, et al. Chlorhexidine-based antiseptic solution vs alcohol-based providone-iodine for central venous catheter care. Arch Intern Med. 2007;167(19):2066–72.

Noble V, et al. Manual of emergency and critical care ultrasound. Cambridge: Cambridge University Press; 2007. p. 196–204.

Parry G. Trendelenburg position, head elevation and a midline position optimize right internal jugular vein diameter. Can J Anaesth. 2004;51(4):379.

Vesely T. Central venous catheter tip position: a continuing controversy. J Vasc Interv Radiol. 2003;14(5):527.

Subclavian Vein Access

Elliott TS, Faroqui MH, Armstrong RF, Hanson GC. Guidelines for good practice in central venous catheterization. Hospital Infection Society and the Research Unit of the Royal College of Physicians. J Hosp Infect. 1994;28(3):163–76.

Fortune JB, Feustel P. Effect of patient position on size and location of the subclavian vein for percutaneous puncture. Arch Surg. 2003;138(9):996–1000; discussion 1001.

Fragou M, et al. Real time ultrasound-guided subclavian vein cannulation versus the landmark method in critical care patients: a prospective randomized study. Crit Care Med. 2011;39(7):1–6.

Kilbourne MJ, Bochicchio GV, Scalea T, Xiao Y. Avoiding common technical errors in subclavian central venous catheter placement. J Am Coll Surg. 2009;208(1):104–9.

McGee DC, Gould MK. Preventing complications of central venous catheterization. N Engl J Med. 2003b;348(12):1123–33.

Femoral Vein Access

Dailey RH. Femoral vein cannulation: a review. J Emerg Med. 1985;2:367–72.

Lim T, Ryu H-G, et al. Effect of the bevel direction of puncture needle on success rate and complications during internal jugular vein catheterization. Crit Care Med. 2012;40(2):491–4.

McGee DC, Gould MK. Preventing complications of central venous catheterization. N Engl J Med. 2003c;348:1123–33.

Marik PE, Flemmer M, Harrison W. The risk of catheter-related bloodstream infection with femoral venous catheters as compared to subclavian and internal jugular venous catheters: a systemic review of the literature and meta-analysis. Crit Care Med. 2012;40:2479–85.

Swanson RS, Uhlig PN, Gross PL, et al. Emergency intravenous access through the femoral vein. Ann Emerg Med. 1984;13:244–7.

4 Pulmonary Artery Catheter

Rohit Pravin Patel and Marie-Carmelle Elie

4.1 Indications

- Prevention or treatment of multiorgan failure in high-risk patients
- Preoperative and postoperative management in high-risk patients with cardiac, pulmonary, or renal dysfunction
- Patients with anticipated large fluid shifts (sepsis, bleeding, burns, cirrhosis)
- Oliguria or hypotension not relieved by fluids
- Suspected cardiac event leading to shock
- For continuous SVO2 (central venous oxygenation) monitoring in shock
- To differentiate shock states
- For monitoring cardiac output in patients requiring high-positive end-expiratory pressure (>14 cm H20)
- Monitoring and management of complicated myocardial dysfunction or cardiogenic shock
- Congestive heart failure with poor response to afterload reduction and diuretic therapy
- Suspected tamponade or contusion from blunt chest injury
- Pulmonary hypertension with myocardial dysfunction
- Diagnosis of primary pulmonary hypertension
- Aspiration of air emboli
- Direct pulmonary artery administration of thrombolytic therapy

4.2 Contraindications

- Tricuspid or pulmonary valve mechanical prosthesis
- Right heart mass (thrombus or tumor)
- Tricuspid or pulmonary valve endocarditis
- Recurrent sepsis (catheter could serve as nidus for infection)
- Hypercoagulopathy (catheter could serve as site for thrombus formation)
- Patient known sensitivity to heparin (catheters with heparin coating)
- Electrocardiographic (ECG) monitoring encouraged in conditions of complete left bundle branch block (risk of complete heart block increased), Wolfe–Parkinson–White syndrome, and Ebstein's malformation (risk of tachyarrhythmias)

4.3 Materials

- Pulmonary artery or Swan–Ganz catheter
- Percutaneous sheath introducer and contamination shield
- Compatible cardiac output computer for measuring cardiac output by the bolus thermodilution method
- Injectate temperature sensing probe (bolus thermodilution method)
- Connecting cables
- Sterile flush system and pressure transducers
- Bedside ECG and pressure monitor system
- Appropriate ECG "slave" cables

4.3.1 Catheter Preparation

- Avoid forceful wiping or stretching of catheter to avoid injury to the thermistor wire circuitry; wiping the heparin coat may cause removal of the coating.
- In vivo calibration is required if in vitro calibration is not done; refer to the monitor operator's manual for detailed calibration instructions.
- Connect catheter's injectate and pressure-monitoring lumens to the flush system and pressure transducers; ensure all lines are free of air.
- Connect the thermistor to the monitor and confirm no fault messages appear.

R.P. Patel, MD • M.-C. Elie, MD, RDMS (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: rohitpatel@ufl.edu; elie@ufl.edu

© Springer Science+Business Media New York 2016

L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_4

4.4 Procedure

1. Maintain sterile precautions including sterile cap, mask, gown, gloves.
2. Place a central venous line introducer and verify placement of introducer with chest radiograph; it is also acceptable to obtain the chest radiograph after the insertion of the pulmonary artery catheter if no complication was suspected with the central venous line introducer.
3. Cleanse the skin and introducer thoroughly with chlorhexidine.
4. Have assistant open pulmonary artery catheter kit in sterile fashion.
5. Remove pulmonary artery catheter from kit and have assistant hook up all the ports to the transducers and make sure readings are accurate as the catheter is being manipulated.
6. Gently lift the distal portion of the catheter up from the silicone gripper; do not pull the balloon through the gripper to avoid damage.
7. Have assistant check the proximal and distal ports for patency by flushing with sterile saline. Also have the assistant check the patency of the balloon with the syringe provided in the kit (Fig. 4.1). Check for major asymmetry and for leaks (optional) by submerging in sterile saline or water. Deflate balloon prior to insertion. Carefully wave the distal catheter segment up and down to confirm electrical continuity by observing a pressure tracing on monitor. Ensure proper readings, no information is sometimes better than wrong information. Make sure each port transduces appropriately *prior to* insertion.
8. Familiarize yourself with the catheter line markings. Each thin line indicates 10 cm from the tip and thick line indicates 50 cm from the tip. These are used in combination to indicate length from tip (Fig. 4.2).
9. Place the sterile plastic sleeve (lock side toward patient, Fig. 4.3) over the catheter after flushing all ports to further protect the catheter during manipulation.
10. The distal end of the catheter is inserted into the introducer hub of the central venous line and threaded to the superior vena cava. The catheter must be placed at least 30 cm into the introducer for the balloon to clear the distal end of the introducer prior to inflation. At *no point* should the catheter be withdrawn with the balloon inflated; ensure the assistant has deflated the balloon prior to withdrawal. The balloon assists in directing the catheter through the vascular system using the directional blood flow.
11. At 20 cm, the balloon should be inflated and catheter advanced through right atrium, past the tricuspid valve into the right ventricle, then past the pulmonary valve to the pulmonary artery. The waveform and pressure readings can guide you through the various locations (Fig. 4.4).
12. Once in the pulmonary artery, the catheter should be carefully and slowly advanced to wedge position. The balloon can be deflated and pulmonary artery tracings should reappear. If a wedge is obtained with less than the maximum recommended volume, the catheter should be withdrawn to a position where full inflation volume produces a wedge tracing. Avoid prolonged times when obtaining wedge pressure (2 respiratory cycles or 10–15 s), especially in patients with pulmonary hypertension.
13. General guidelines for distance necessary at various points include: right atrium 20–25 cm, right ventricle 30–35 cm, and pulmonary artery 40–45 cm; catheter usually wedges at 50–55 cm. These are dependent on the starting location you are using to advance the catheter (subclavian, internal jugular, femoral).
14. Once the catheter is in correct position, it should be locked into place with the plastic sleeve tip onto the hub of the introducer.
15. Correct placement is confirmed with chest radiograph (Fig. 4.5).

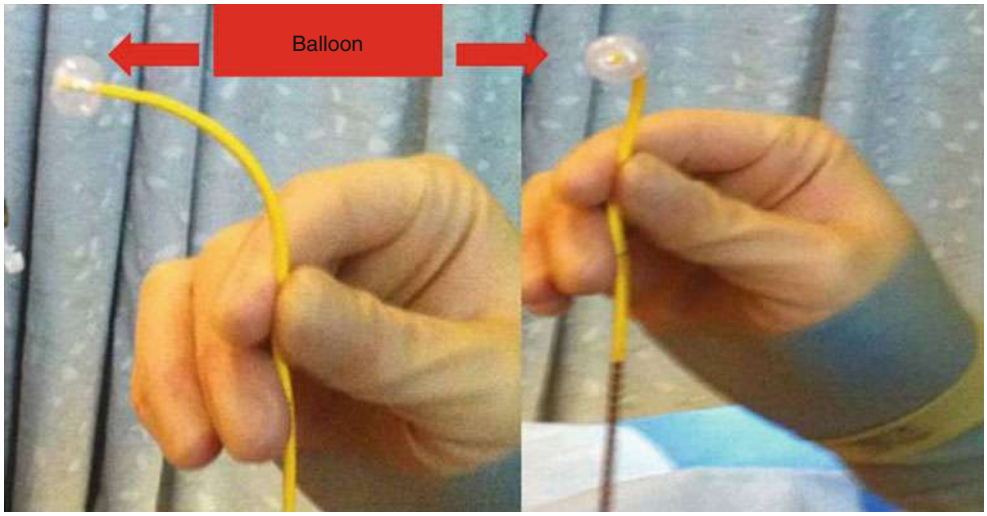

Fig. 4.1 Balloon inflation prior to insertion for evaluation of patency or leaks

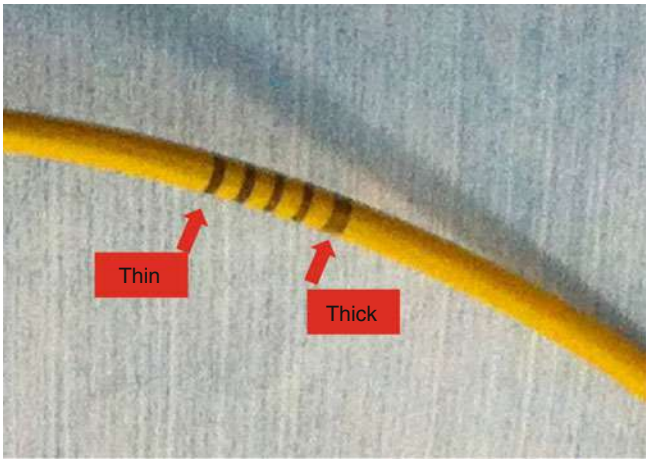

Fig. 4.2 Thick and thin markings found on the catheter representing length from distal tip

Fig. 4.3 Lock position on catheter for stabilization of catheter

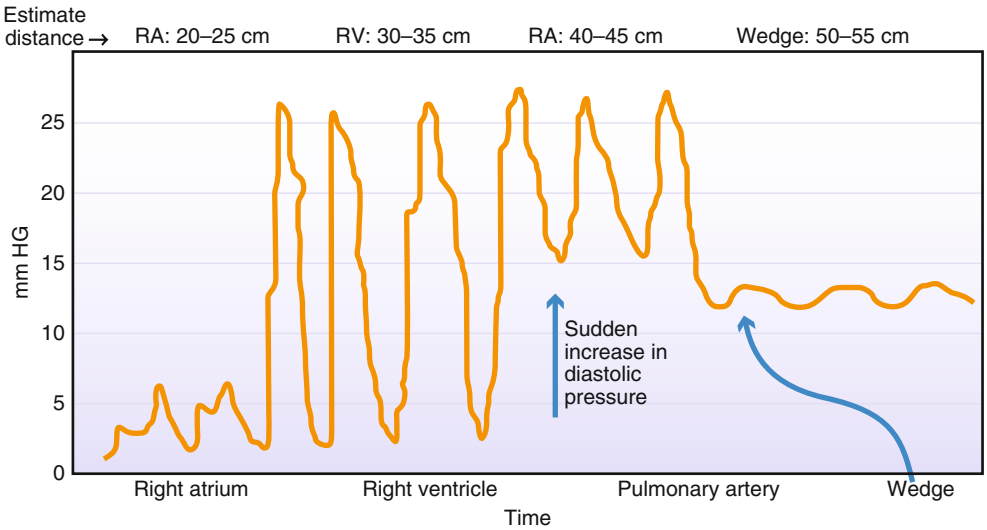

Fig. 4.4 Typical waveform seen at specific locations in the heart with associated estimated distances from catheter tip

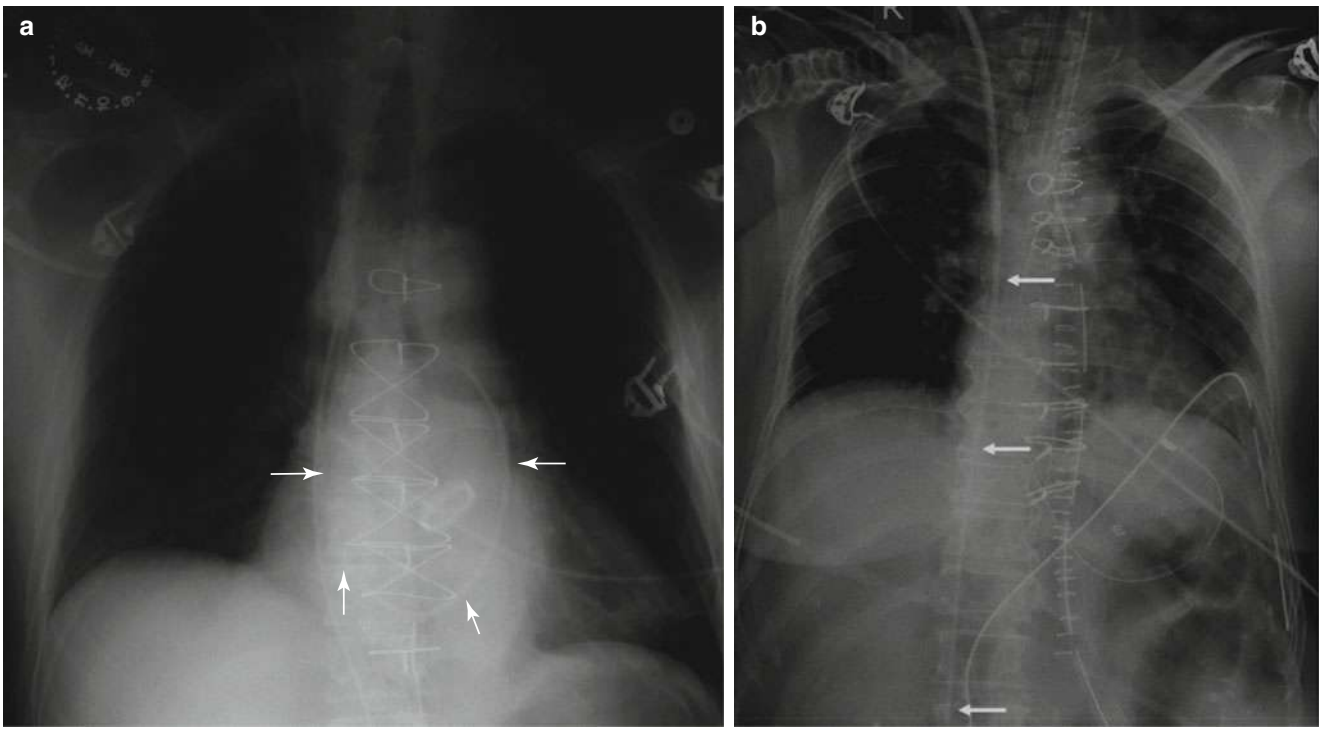

Fig. 4.5 Chest radiograph demonstrating correct (**a**) and incorrect (**b**) positioning of pulmonary artery catheter (*arrows*) (**a** Reproduced with permission from McGraw-Hill: Stead LG, et al. *First Aid for the Radiology Clerkship*. New York: McGraw-Hill; 2009; **b** Reproduced with permission from Wolters Kluwer: Jain SN. A pictorial essay: radiology of lines and tubes in the intensive care unit. *Indian J Radiol Imaging*. 2011;21(3):182–190)

4.5 Complications

- Arrhythmias: most are premature ventricular contractions that are self-limiting and resolve with advancement into pulmonary artery or withdrawal into atrium.
- Right bundle branch block: usually transient after positioning catheter into the pulmonary artery; if has already left bundle branch block may lead to complete heart block; should have temporary pacing equipment on standby.
- Knotting in the right ventricle (RV): risk increased in those with dilated cardiac chambers; a persistent RV tracing (15 cm beyond the point where initial RV tracing was observed) should alert you to this possibility.
- Pulmonary artery rupture: age >60 year, anticoagulation therapy, and presence of pulmonary hypertension increase risk of rupture; hemoptysis shortly after placement is indicative and management includes lateral decubitus positioning (bleeding side down), intubation with double lumen tube, and increasing positive end-expiratory pressure (PEEP).
- Infection.
- Pulmonary infarction: due to unintentional migration of distal tip.

4.6 Pearls and Pitfalls

- If the catheter requires stiffening during insertion, slowly perfuse the catheter with 5–10 mL of cold sterile solution as the catheter is advanced through a peripheral vessel.
- The incidence of complications increases significantly with periods of use longer than 72 h, so assess the need for the catheter on daily basis.
- Anticipate spontaneous catheter tip migration toward periphery of pulmonary bed; if a wedge tracing is observed when balloon is deflated, pull the catheter back.

Selected Reading

1. Edward Lifesciences. Pulmonary artery catheter Instruction manual. 2009. http://ht.edwards.com/resourcegallery/products/swan-ganz/pdfs/invasivehdmphysprincbook.pdf.
2. Edward Lifesciences. Pulmonary artery catheter Instruction manual. 2009. http://www.edwards.com/products/pacatheters/Pages/ThermodilutionCatheter.aspx.
3. Leatherman JW, Marini JJ. Clinical use of the pulmonary artery catheter. In: Hall JB, Schmidt GA, Wook LDH, editors. Principles of critical care. 3rd ed. New York: McGraw-Hill; 2005. p. 146–50.
4. Moran SE, Pei KY, Yu M. Hemodynamic monitoring: arterial and pulmonary artery catheters. In: Gabrielli A, Layon AJ, Yu M, editors. Civetta, Taylor, and Kirby's critical care. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2009.

Noninvasive Cardiac Monitoring: The Edwards Vigileo System

5

Dawood G. Dalaly and Rohit Pravin Patel

5.1 Indications

Cardiac output monitoring is indicated when trying to determine fluid responsiveness in patients. It assists in directing and assessing results of resuscitative efforts to ensure appropriate tissue perfusion. Although most catheters are systemically invasive, tools like the Vigileo (Edwards Lifesciences; Irvine, CA) are excellent noninvasive devices for determining values such as the stroke volume, stroke volume variation, stroke volume index, cardiac output, and cardiac index (Fig. 5.1).

D.G. Dalaly, DO
Department of Surgery,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: ddalaly@ufl.edu

R.P. Patel, MD (✉)
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: drpratik@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_5

Fig. 5.1 Typical monitoring devices and connecting devices from monitor to patient: (**a**) FloTrac monitor, (**b**) FloTrac red port connects to the heart monitor, (**c**) the FloTrac system connects to the arterial line on one end and the pressure bag/monitors on the other; monitor connections are color coded, (**d**) the green port connects to the green port from the FloTrac

5.2 Contraindications

There are no contraindications to noninvasive monitoring of the heart, although most systems will need to be connected to an arterial line catheter, therefore contraindications to arterial line placement must be determined. Current literature supports the use of noninvasive monitoring to those who are 100 % supported ventilation with tidal volumes ≥8 mL/kg. There is no support for use in those with spontaneous breaths or arrhythmia. Patients with significant variation in respiratory pattern may have results that are unreliable.

5.3 Definitions and Values (Table 5.1)

Table 5.1 Cardiac output monitoring definitions

Term	Reference range	Definition
Cardiac output	4–8 L/min	Volume of blood being pumped by the heart in 1 min
Cardiac index	2.5–4 L/min/m^2	Compares the amount of fluid being pumped by the heart with an individual's body surface area
Stroke volume	40–80 mL/beat	Volume of blood being pumped by the ventricle per beat
Stroke volume variation	10–15 %	Represents percentage of change between minimum and maximum stroke volumes and predictor of fluid responsiveness
Stroke volume index	33–47 mL/m^2/beat	Quantity of blood ejected from the heart per beat
Mixed venous saturation (SvO_2)	60–80 %	Percentage of oxygen bound to hemoglobin in blood returning to the right side of the heart; represents oxygen delivery and consumption at the tissue level. Usually obtained from pulmonary artery catheter
Central venous oxygen saturation ($ScvO_2$)	70 %	Surrogate marker for SvO_2, usually obtained from internal jugular or subclavian catheters

5.4 Materials

- FloTrac (Edwards Lifesciences, Irvine, CA, USA)
- Vigileo monitor
- Pressure bag

5.5 Procedure

1. Connect FloTrac to arterial line and distal FloTrac port to pressure bag. Inflate bag to 300 mmHg.
2. Connect green FloTrac cord to green Vigileo cord and red FloTrac cord to arterial line port on heart monitor.
3. Turn the port on the FloTrac off to the patient and zero on your heart monitor as you would with an arterial line. At the same time, you should have pressed the "Enter" knob on the Vigileo system, scrolled it to "zero arterial pressure," and pressed the knob again for that function.
4. You should now have an arterial waveform on your heart monitor and your cardiac values on the Vigileo system.

5.6 Pearls and Pitfalls

- Some components of the values obtained are not reliable on spontaneously breathing patients and it is critical to check the ventilator waveforms for these breaths.
- Patients without adequate tidal volumes (at least 8 mL/kg) will also have unreliable values.
- Stroke volume variation usually is more reliable when greater than 13 % and indicates fluid responsiveness rather than when it is less than 13 % (similar to a low central venous pressure (CVP) being more informative than normal or high CVP levels).

Selected Reading

1. Alarcon LH, Fink MP. Chapter 13. Physiologic monitoring of the surgical patient. In: Brunicardi FC, Andersen DK, Billiar TR, et al., editors. Schwartz's principles of surgery. 9th ed. New York: McGraw-Hill; 2010.
2. Holcroft JW, Anderson JT, Sena MJ. Chapter 12. Shock and acute pulmonary failure in surgical patients. In: Doherty GM, editor. CURRENT diagnosis & treatment: surgery. 13th ed. New York: McGraw-Hill; 2009. Available at http://www.accesssurgery.com/content.aspx?aID=5212482. Accessed 22 Aug 2012.
3. Edwards Critical Care Education. Available at http://www.edwards.com/education/Pages/cceducationmap.aspx. Accessed 11 Dec 2012.

6 Peripheral Venous Cutdown

Jeffrey Kile, Katrina John, and Amish Aghera

6.1 Indications

- Distorted anatomy of peripheral venous access sites
- Unavailability of cannulable veins (e.g., in hypovolemia, burn victim, traumatic anatomy, sclerosed veins, etc.)
- Emergency venous access for infusion/transfusion
- Unavailability of central venous access or less invasive means peripherally

6.2 Contraindications

- Absolute
  - Availability of less invasive or less time-consuming means of vascular access
  - Overlying infection, traumatic tissue, burn, etc., at cutdown site
  - Traumatic injury proximal to cutdown site
- Relative
  - Coagulation disorders

6.3 Materials and Medications (Fig. 6.1)

- Sterile gloves
- Antimicrobial solution and swabs
- 4″×4″ gauze sponges
- Local anesthetic (1 % lidocaine 5 mL)
- 5-mL syringe
- Blunt needle
- 25- or 27-gauge needle
- Scalpel
- Vein dilator/lifter
- Peripheral intravenous catheter
- Curved hemostat
- 0-0 silk sutures or 4.0 nylon sutures
- Iris scissors
- Intravenous infusion tubing
- Adhesive tape

J. Kile, MBBS, PhD, MPH (✉) • K. John, MBBS
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: jeffrey.kile@gmail.com; trenjohn@me.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_6

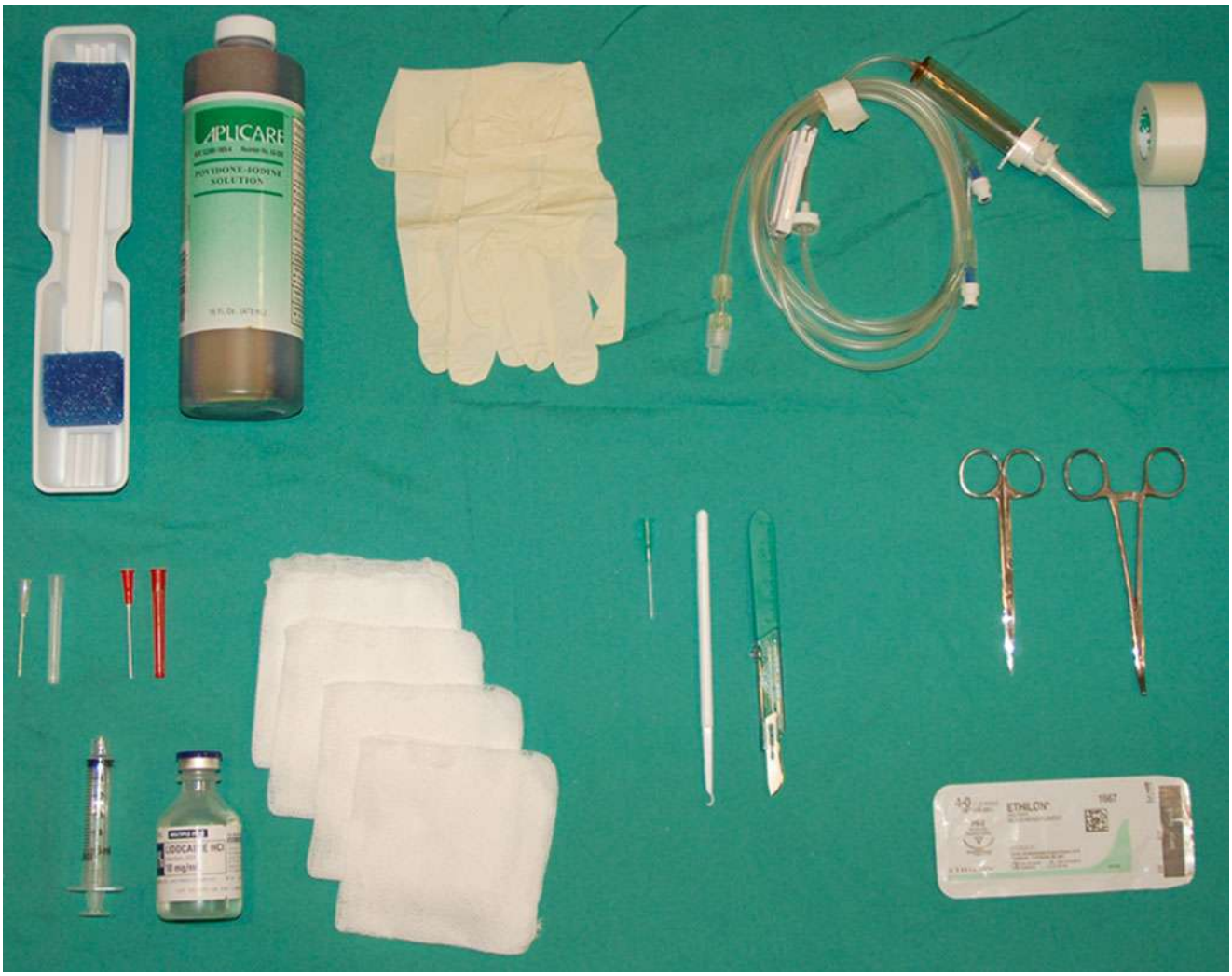

Fig. 6.1 Materials and medications

6.4 Choice of Vessel for Cutdown

- Greater saphenous vein: this vessel is the longest vein in the body, is predominantly subcutaneous, and is exposed with minimal blunt dissection just anterior to the medial malleolus at the ankle.
- Basilic vein: this vessel is reliably located 1–2 cm lateral to the medial epicondyle on the anterior aspect of the humeral region, typically catheterized just superior to the antecubital fossa approximately, and its diameter permits its localization relatively easily even in the hypotensive patient.
- Cephalic vein: this vessel runs anteromedially from the radial aspect of the wrist to the antecubital fossa, is superficial and large in diameter, and is most easily cannulated at the distal flexor crease in the antecubital fossa.

6.5 Procedure

6.5.1 Standard Venous Cutdown Technique

1. Apply antimicrobial solution liberally to the skin surrounding the incisional area.
2. Establish a sterile field by placing drapes around the incisional area.
3. Apply a tourniquet proximal to the planned cutdown site to maximize visualization of vein to be cannulated.
4. Inject local anesthetic to raise a small (0.5 cm) wheal using 25- or 27-gauge needle and then insert the tip of the needle through wheal to infiltrate skin superficial to the artery with approximately 4 mL of anesthetic.
   - Injection of local anesthetic into the vessel may precipitate arrhythmia, so draw back on the plunger prior to infiltration to ensure the tip of the needle is not inside the vessel.
5. Incise the skin with scalpel perpendicular to the course of the vein through all cutaneous layers until subcutaneous fat is visualized (Fig. 6.2).
   - Some practitioners prefer using a longitudinal incision to reduce the risk of transecting neurovascular structures, but this may not produce sufficient exposure of vein.
6. Using a curved hemostat or gloved finger, bluntly dissect the subcutaneous tissue to isolate and mobilize approximately 2–3 cm of the vein (Fig. 6.3).
   - A small self-retaining retractor or tissue spreader can be used in this step to improve visualization of vein if desired
7. Pass suture under the vein distal to the planned venous puncture site using hemostat to stabilize the vein and tie the suture over the vein (Fig. 6.4).
8. Pass a second suture under the vein proximal to the planned venous puncture site using hemostat (Fig. 6.5).
   - This step enables increased visualization, vessel control, and hemostasis during incision.
   - Leave the ends of both sutures long to facilitate maneuvering the vein.
9. Incise one-half to one-third of the diameter of the vein using a scalpel or iris scissors held at a 45° angle to the vessel (Fig. 6.6).
10. Grasping the proximal edge of the incision with a hemostat to apply counter traction (in a distal direction), insert the tip of the catheter into the venous incision (Fig. 6.7).
    - Do not force the catheter if it does not easily advance.
    - Catheter can be introduced directly through the skin incision or via skin puncture adjacent to the skin incision.
    - If the catheter lacks a tapered tip, cut the distal end of the cannula at a 45° angle to fashion a beveled tip.
11. Thread catheter into vein (Fig. 6.8).
12. Aspirate any air which may have entered the cannula during insertion.
13. Connect hub of catheter to intravenous tubing.
14. Tie the proximal suture around the vein just proximal to the venous incision, encircling both the vein and the intraluminal cannula with the suture.
15. Remove tourniquet.
16. Secure the catheter hub to the skin using nylon (4.0) sutures as follows. Take a 0.5 cm bite of skin under the catheter hub with the suture needle, tie several knots in the suture without pinching the skin, then tie a second set of knots around the hub of the catheter firmly.
17. Close the incision using nylon (4.0) sutures.
18. Dress the wound with appropriate self-adhesive sterile dressing or sterile gauze pads and adhesive tape.

Fig. 6.2 Incision of skin

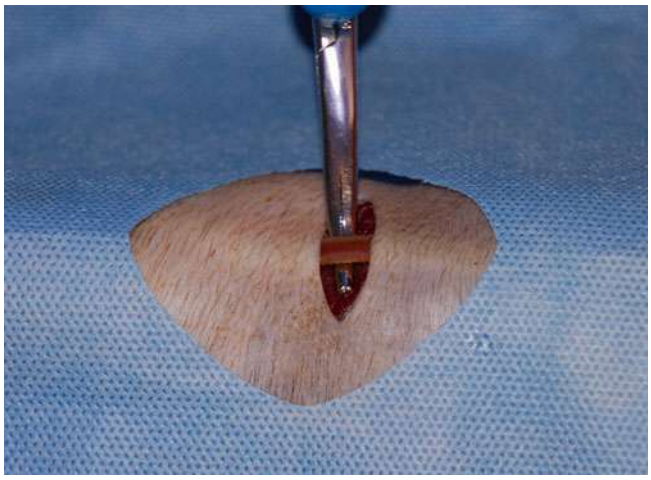

Fig. 6.3 Mobilization of vein

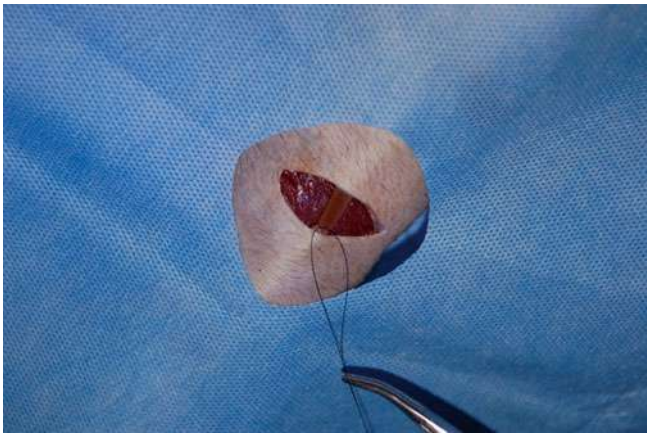

Fig. 6.4 Distal ligature tied around vein

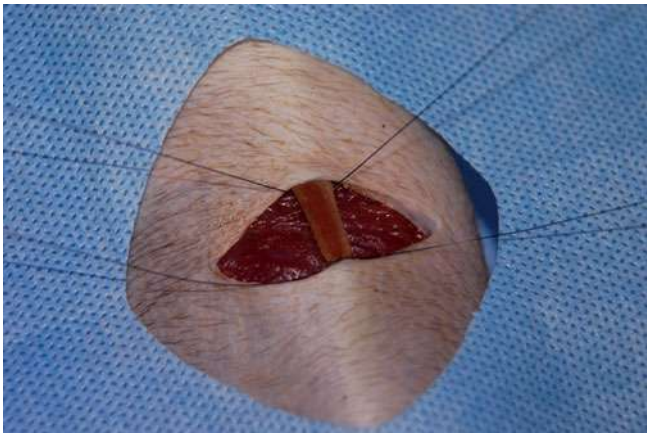

Fig. 6.5 Distal and proximal ligatures in place

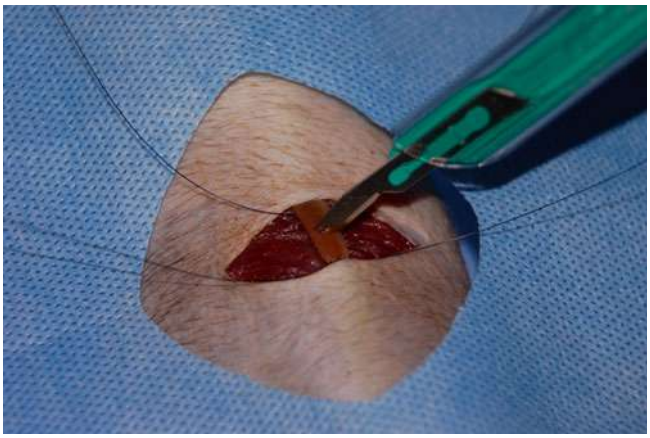

Fig. 6.6 Incision of vein

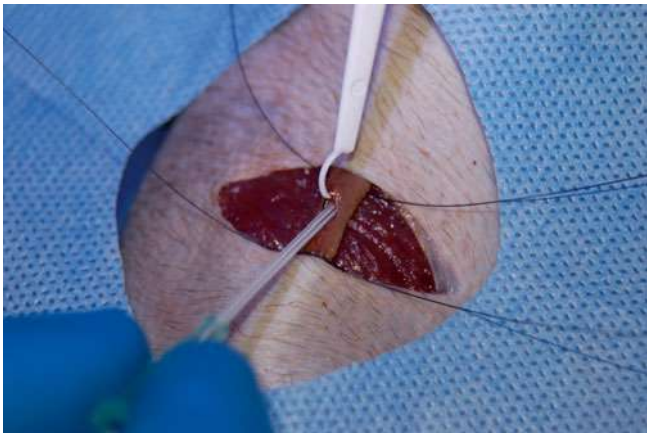

Fig. 6.7 Catheterization of vein

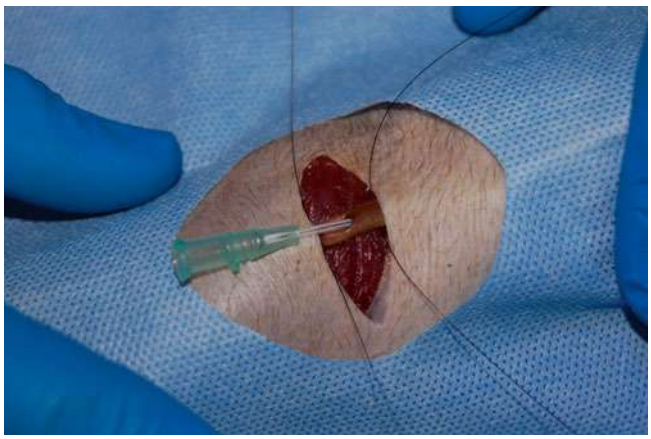

Fig. 6.8 Catheter threaded into vein

6.5.2 "Mini-Cutdown" Technique

(Perform steps 1–6 from the "Standard Venous Cutdown Technique," above, and then proceed with the steps below.)

1. Puncture the vein using a standard over-the-needle venous catheter.
   - Catheter can be introduced directly through the skin incision or via skin puncture adjacent to the skin incision.
2. Thread catheter into the vein over the needle.
3. Remove and discard the needle.
4. Aspirate any air which may have entered the catheter during insertion.
5. Connect catheter to intravenous tubing.

(Continue with steps 15–18 from the "Standard Venous Cutdown Technique," above.)

6.5.3 Modified/Guide Wire Technique

(Perform steps 1–6 from the "Standard Venous Cutdown Technique," above, and then proceed with the steps below.)

1. Insert the blunt end of the guide wire into the incised vein.
2. While stabilizing the guide wire at its insertion site, thread the dilator and sheath assembly over the free end of the guide wire until it is approximately one inch from the skin.
3. Grasp the free end of the guide wire protruding from the tail end of the assembly.
   - If it does not protrude from the tail end of the assembly, the guide wire must be removed sufficiently from the artery to be securely grasped. It must protrude visibly from the tail end of the dilator throughout the subsequent process of threading the dilator into the vein.
   - Never let go of the guide wire during this step, as insertion of the dilator and sheath assembly can otherwise push the guide wire completely into the vein.
4. Holding it firmly near its tip, thread the assembly over the wire into vessel with a gentle back-and-forth twisting motion.
5. Holding the sheath securely in the vein, remove and discard the dilator and guide wire.
6. Aspirate any air which may have entered the sheath during insertion.
7. Connect sheath to intravenous tubing.
8. Remove tourniquet.
9. Secure the sheath to the skin.
10. Close the incision using nylon (4.0) sutures.
11. Dress the wound with appropriate self-adhesive sterile dressing or sterile gauze pads and adhesive tape.

6.6 Complications

- Hematoma
- Infection
- Sepsis
- Phlebitis
- Embolization
- Wound dehiscence

6.7 Pearls and Pitfalls

- Fluids are infused most quickly via short, large-bore catheters.
- If the line is inserted for slow infusion of intravenous drugs, catheter lumen size is relatively insignificant.
- In larger children and adults, intravenous plastic tubing, small-bore pediatric feeding tubes, and Silastic catheters may be used as infusion catheters.
- Threading a 10-gauge intravenous catheter or intravenous tubing directly into the incised vein achieves excellent flow rates.
- If difficulty is encountered while threading the catheter into the incised vein, ensure an appropriately sized catheter has been used and that the vessel lumen has been correctly identified and that no false passage has been created in the adventitia.
- As compared to the standard venous cutdown technique, the mini-cutdown technique is easier and also preserves the vein, permitting repeated catheterization if necessary.
- As compared to the standard venous cutdown technique, the modified/guide wire technique reduces procedure time and increases the likelihood of vein salvage in the event of vessel transection.

Selected Reading

1. Chappell S, Vilke GM, Chan TC, et al. Peripheral venous cutdown. J Emerg Med. 2006;31:411–6.
2. Klofas E. A quicker saphenous vein cutdown and a better way to teach it. J Trauma. 1997;43:985–7.
3. McIntosh BB, Dulchavsky SA. Peripheral vascular cutdown. Crit Care Clin. 1992;8:807–18.
4. Nocito A, Wildi S, Rufibach K, et al. Randomized clinical trial comparing venous cutdown with the Seldinger technique for placement of implantable venous access ports. Br J Surg. 2009;96:1129–34.
5. Shockley LW, Butzier DJ. A modified wire-guided technique for venous cutdown access. Ann Emerg Med. 1990;19:393–5.

Part II

Airway and Vascular Procedures

7 Bag-Valve-Mask Ventilation

Braden Hexom and Tatiana Havryliuk

7.1 Indications

- Hypoxia
- Hypoventilation/apnea
- Rescue maneuver if failed intubation

7.2 Contraindications

- Absolute
  - Inability to ventilate due to lack of seal (thick beard, deforming facial trauma)
  - Inability to ventilate secondary to complete upper airway obstruction
  - Active, adequate spontaneous ventilation
- Relative
  - Full stomach (aspiration risk)
  - After induction and paralysis during rapid sequence intubation (aspiration risk)

7.3 Materials (Fig. 7.1)

- Bag valve mask (BVM) with reservoir
- Oxygen connector tubing
- Nasal pharyngeal airway/oral pharyngeal airway
- Lubricant jelly

Fig. 7.1 BVM supplies: bag, mask, oral airways, nasopharyngeal airways, lubricant

B. Hexom, MD (✉)
Department of Emergency Medicine, Mount Sinai Hospital, New York, NY, USA
e-mail: braden.hexom@mssm.edu

T. Havryliuk, MD
Department of Emergency Medicine, University of Colorado Denver, Denver, CO, USA
e-mail: tatiana.havryliuk@ucdenver.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_7

7.4 Procedure

1. Position patient in "sniffing" position.
2. Open the airway with chin-lift/head-tilt or jaw thrust maneuvers.
3. Place airway adjuncts to maintain airway patency. Use oral airway (Fig. 7.2) in unconscious patients. Use nasal airway (Fig. 7.3) in semiresponsive patients.
4. Attach oxygen tubing to high-flow oxygen (15 L/min).
5. Place appropriately sized mask on patient's face covering the nose and mouth.
   - For one-handed technique (Fig. 7.4), use nondominant hand to make a "C" with index finger and thumb on top of the mask and form an "E" with the rest of the fingers using them to pull up on the mandible (E–C technique). Use the dominant hand to provide bag ventilations.
   - For two-handed (Fig. 7.5), two-person technique (preferred), make two semicircles with index fingers and thumbs of both hands on top of the mask and use the rest of the fingers to pull up on the mandible.
6. Consider the Sellick maneuver (cricoid pressure) to compress the esophagus against the cervical vertebrae, preventing gastric insufflation.
7. Ventilate patient providing reduced tidal volume breaths (500 mL) at a rate of 10–12 breaths per minute.
8. Give each breath gently over 1–1.5 s to avoid high peak pressures, avoiding gastric insufflation.
9. Prepare for definitive airway as dictated by the clinical scenario.

Fig. 7.2 (**a**–**c**) Oral airway insertion

Fig. 7.3 (**a**, **b**) Nasal airway insertion

Fig. 7.4 (**a**–**c**) One-handed seal technique

Fig. 7.5 Two-handed seal technique: (**a**) two semicircles, (**b**) alternative thumbs method

7.5 Complications

- Stomach inflation may lead to vomiting and aspiration.
- Increased positive thoracic pressure may cause decreased preload, worsening cardiac output, and/or hypotension.
- Hypoventilation (inadequate O_2 tidal volume, airway patency, or mask seal).

7.6 Pearls and Pitfalls

- Pearls
  - Use jaw thrust to open mouth for patients with possible cervical spine injury.
  - Use airway adjuncts whenever available, especially if prolonged BVM ventilation is anticipated.
  - Use lubricant jelly to insert nasal airway; do not insert in patients with severe facial trauma.
  - Mask should be placed on patient's face prior to attaching to bag.
  - Apply pressure to the bony part of the mandible only to avoid soft-tissue injury.
  - Provide just enough tidal volume to see a chest rise and deliver each breath gently over 1–1.5 s to prevent gastric insufflation.
  - Ensure good seal:
    - Select appropriate mask size.
    - Choose two-handed technique over one-handed, if possible.
    - Keep well-fitting dentures in place if present (and remove before intubation).
    - Lift the mandible toward the mask (as opposed to pushing the mask onto the face).
    - Rock the mask on face until no leak is present.
    - Apply K–Y jelly to beard to improve the seal.
- Pitfalls
  - Overcompression during the Sellick maneuver can compress the cricoid ring, preventing ventilation.
  - Cricoid pressure (Sellick maneuver) is not recommended during cardiac arrest resuscitation.
  - Difficult BVM ventilation: MOANS
    - *M*ask seal
    - *O*besity/obstruction
    - *A*ge
    - *N*o teeth
    - *S*tiff

Selected Reading

ECC Committee, Subcommittees and Task Forces of the American Heart Association. American heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation. 2010;122:S685–705.

Joffe AM, Hetzel S, Liew EC. A two-handed jaw-thrust technique is superior to the one-handed "EC-clamp" technique for mask ventilation in the apneic unconscious person. Anesthesiology. 2010;113:873–9.

Roberts JR, Hedges JR. Clinical procedures in emergency medicine. Philadelphia: Saunders Elsevier; 2010.

Walls RM, Murphy MF. Manual of emergency airway management. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

8 Awake Orotracheal Intubation

Benjamin M. Mahon, Justin Bennett, and Lars K. Beattie

8.1 Indications

- Urgent but not emergent endotracheal intubation is required in a patient who:
  - Is awake
  - Is currently protecting his airway
  - Is not a candidate for a supraglottic airway (LMA)
  - The patient
    - Is predicted to have a difficult airway
    - Has structural abnormalities of the airway
    - Will not tolerate a period of apnea
    - May lose his airway (anaphylaxis, angioedema, traumatic airway)
- Patients requiring urgent but not emergent intubation in whom paralytics are contraindicated (i.e., allergies, myasthenic crisis)

8.2 Contraindications

- Absolute
  - Surgical airway indicated
  - Emergent crash airway needed
  - Obtunded patient
  - Allergies to medications needed (lidocaine, glycopyrrolate)
- Relative
  - Inability to maintain airway or tolerate secretions

8.3 Materials and Medications

- Suctioning equipment
- Intravenous access equipment, cardiac monitor, pulse oximetry, blood pressure cuff
- 4 % lidocaine solution
- 2 % viscous lidocaine jelly
- Nebulizer
- Mucosal atomization device
- 4×4 gauze
- Tongue depressor
- Glycopyrrolate/atropine
- Sedation: ketamine, propofol, Versed, and/or fentanyl
- Intubation equipment
- Backup emergency airway adjuncts
- Bag valve mask
- Laryngoscope, fiber optics, oral airway, etc.

8.4 Procedure

1. Preparation
   (a) Establish IV access.
   (b) Place the patient on a cardiac monitor with continuous pulse oximetry.
   (c) Keep backup RSI emergency airway medication and equipment at the bedside.
2. Administer 0.2–0.4 mg of intravenous glycopyrrolate (or a small dose (0.5–1 mg) of atropine to decrease secretions) 15 min prior to procedure.
3. Nebulize 2 mL of 4 % lidocaine with oxygen at 5 l O2 per minute to anesthetize the pharynx (Fig. 8.1).

B.M. Mahon, MD
Ponciana Medical Center,
Kissimmee, FL, USA
e-mail: drfaust2k9@aol.com

J. Bennett, MD
Department of Emergency Medicine, Wake Forest University
School of Medicine, Winston-Salem, NC, USA
e-mail: jubennet@wakehealth.edu

L.K. Beattie, MD, MS (✉)
Department of Emergency Medicine, University of Florida,
Gainesville, FL, USA
e-mail: lars.beattie@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_8

4. Use Yankauer suction (with the patient's assistance) to dry out the mouth as much as possible. Dabbing the tongue with gauze can assist in this step.
5. Continue preoxygenation.
6. Immediately after the nebulized solution is applied, give the patient a "lidocaine lollipop" (Fig. 8.2).
   (a) A 2 ml dollop of 2 % viscous lidocaine is to be placed on the end of a tongue depressor and is given to the patient to place in his mouth (like a lollipop).
   (b) Have the patient copiously gargle, then swallow the viscous lidocaine.
7. Using a mucosal atomizer, spray 2 ml of 4 % lidocaine in the posterior oropharynx and as far down toward the glottis as possible (Figs. 8.3, 8.4, and 8.5).
8. Sedation
   (a) It is feasible to proceed with the awake intubation in an un-sedated, wide awake but cooperative patient.
   (b) Sedation can be initiated using institutional preferences, but some options include:
      (i) Midazolam 2 mg IV
      (ii) Ketamine 1 mg/kg IV
      (iii) Propofol 1 mg/kg IV
      (iv) Ketofol (ketamine and propofol both at concentrations of 10 mg/ml, 5 ml of each mixed in a 10 cc syringe) titrated at 1–3 ml aliquots
   (c) More atomized lidocaine can be provided prior to endotracheal tube (ETT) passage, but one must be aware of the upper lidocaine dose for your patient.
   (d) Adequate anesthesia is confirmed by the absence of a gag reflex upon direct palpation (Fig. 8.6).
9. Intubation, induction, and gentle direct laryngoscopy can be performed at this point to place the ETT (Fig. 8.7).
   (a) Induction (if no prior sedation) and paralytic agents should be available to immediately administer after ETT placement.
   (b) Thorough discussions on intubation techniques can be found in other chapters in the atlas.

Fig. 8.1 Nebulization of 4 % lidocaine

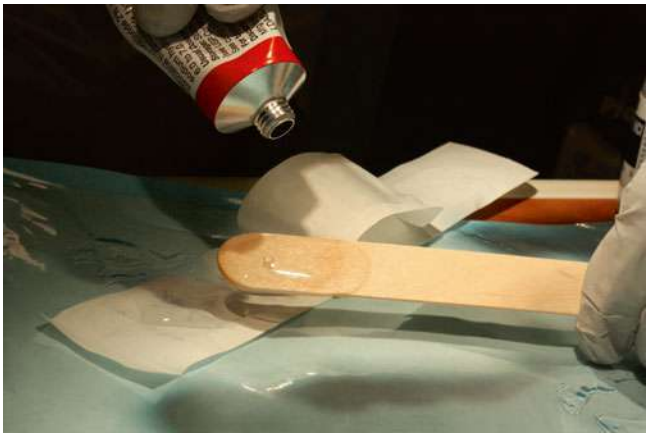

Fig. 8.2 Lidocaine lollipop

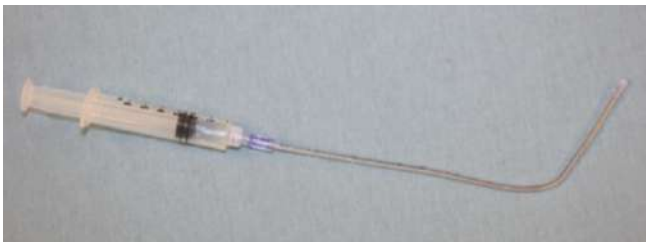

Fig. 8.3 Mucosal atomizer attached to syringe

Fig. 8.4 Pushing syringe plunger atomizes lidocaine

Fig. 8.5 Atomized lidocaine being administered to the posterior pharynx

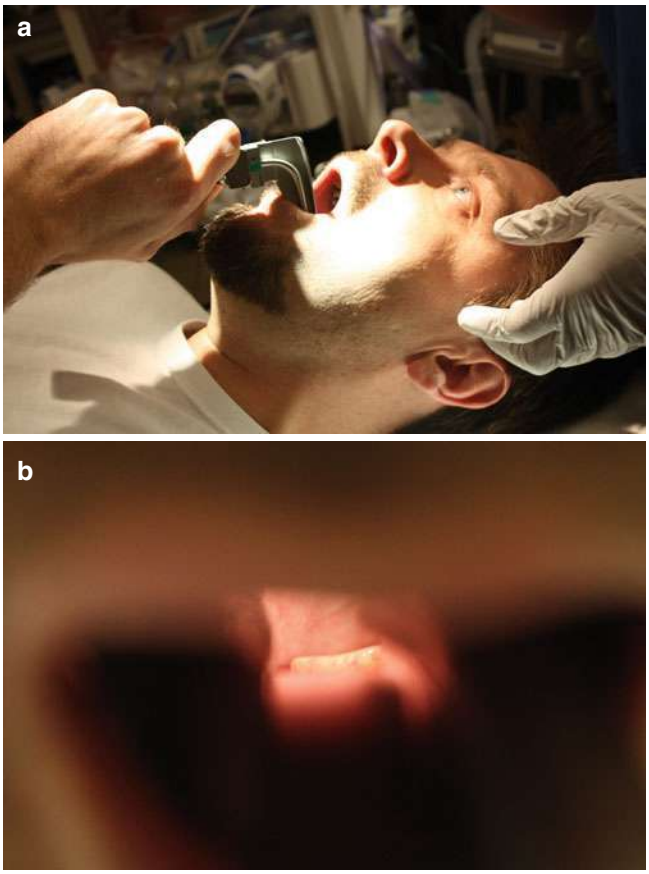

Fig. 8.6 (**a**, **b**) Adequately anesthetized awake patient with laryngoscopic view of epiglottis

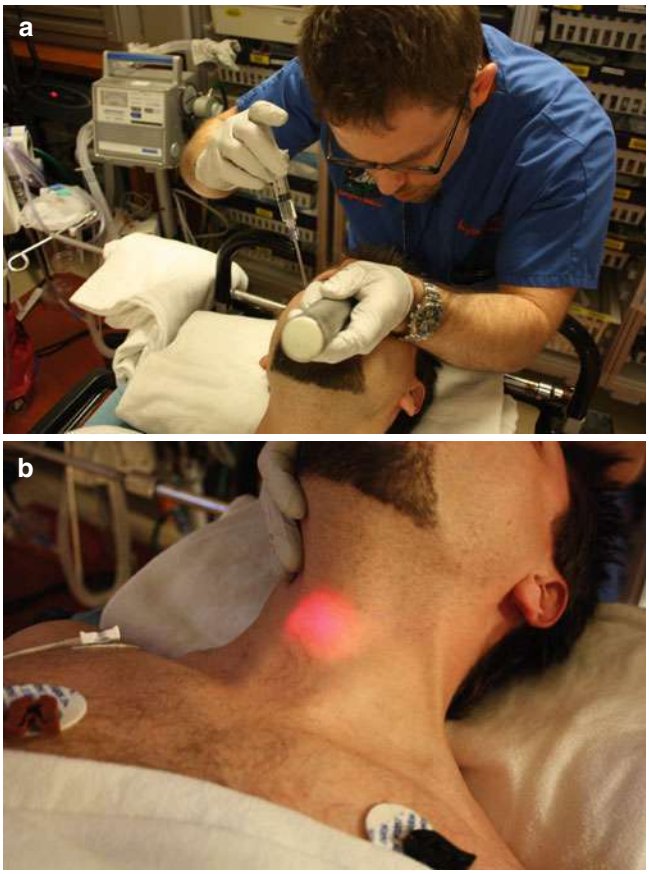

Fig. 8.7 (**a**) Final lidocaine atomization of deep structures and trachea, (**b**) intubation using a lighted stylet

8.5 Pearls and Pitfalls

- Pearls
  - Steps 3–7 should be done successively and as quickly as possible, to capitalize on the short half-life of lidocaine.
  - Simultaneous preoxygenation and anesthetization can be achieved by nebulizing the 4 % lidocaine through a face mask.
  - If the patient has been sedated, soft restraints may help prevent the patient from inadvertently grabbing the tube or your intubating equipment.
  - A nasotracheal intubation can be performed by simply anesthetizing the nares through which the ETT will be placed using lidocaine jelly and/or atomized lidocaine.
- Pitfalls
  - Failure to prepare all equipment beforehand may extend the procedure beyond the lidocaine half-life.
  - The toxic dose of lidocaine is 300 mg or 3–5 mg/kg. The dosages listed are intended for a 70 kg patient and yields a total dose of 280 mg (4 mg/kg). This dose may need to be decreased in smaller individuals.

Selected Reading

Agro F, Hung OR, Cataldo R, Carassiti M, Gherardi S. Lightwand intubation using the Trachlight: a brief review of current knowledge. Can J Anaesth. 2001;48(6):592–9.

Rhee KY, Lee JR, Kim J, Park S, Kwon WK, Han S. A comparison of lighted stylet (Surch-Lite) and direct laryngoscopic intubation in patients with high Mallampati scores. Anesth Analg. 2009;108(4): 1215–92.

Walls RM, Murphy MF. Manual of emergency airway management. 3rd ed. Philadelphia: Lippincott Williams and Wilkins, a Wolters Kluwer Business; 2008. Chap 11.

9 Rapid-Sequence Intubation

Ram A. Parekh

9.1 Indications

- Failure to oxygenate
- Failure to ventilate
- Unable to protect airway patency or reflexes
- Projected clinical course deterioration

9.2 Contraindications

- Absolute
  - Complete upper airway obstruction
  - Significant facial and airway trauma with loss of landmarks for orotracheal intubation
- Relative
  - Anticipated difficult intubation
    - Not an absolute contraindication.
    - Patient scenario requires a careful preintubation assessment and plan.
    - Consider an "awake" intubation.
    - Consider alternative airway adjuncts (e.g., extraglottic devices, video laryngoscopy, laryngeal mask airway [LMA]).
  - Induction or paralytic agent-specific contraindications given clinical circumstances
    - Caution: induction agents that lower blood pressure in hypotensive patients
    - Caution: succinylcholine in potentially hyperkalemic patients
  - Crash airway
    - Apneic, arrest, and periarrest situation

9.3 Materials and Medications

- Laryngoscope with appropriate blade (choice based on proceduralist's preference and patient anatomy) (Fig. 9.1)
- Intubating stylet
- Endotracheal tubes (ETTs)
- Syringe, 10 mL (to inflate ETT cuff)
- Surgilube
- Suction catheter
- Oral and nasal airways (Fig. 9.2)
- Ambu bag and mask attached to oxygen source
- Induction, pretreatment, and paralytic agents
- ETT confirmation device—EZ capnometry, quantitative end-tidal carbon dioxide concentration ($EtCO_2$) detection

R.A. Parekh, MD
Department of Emergency Medicine,
Icahn School of Medicine at Mount Sinai,
Elmhurst Hospital Center, Elmhurst, NY, USA
e-mail: ram.parekh@mssm.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_9

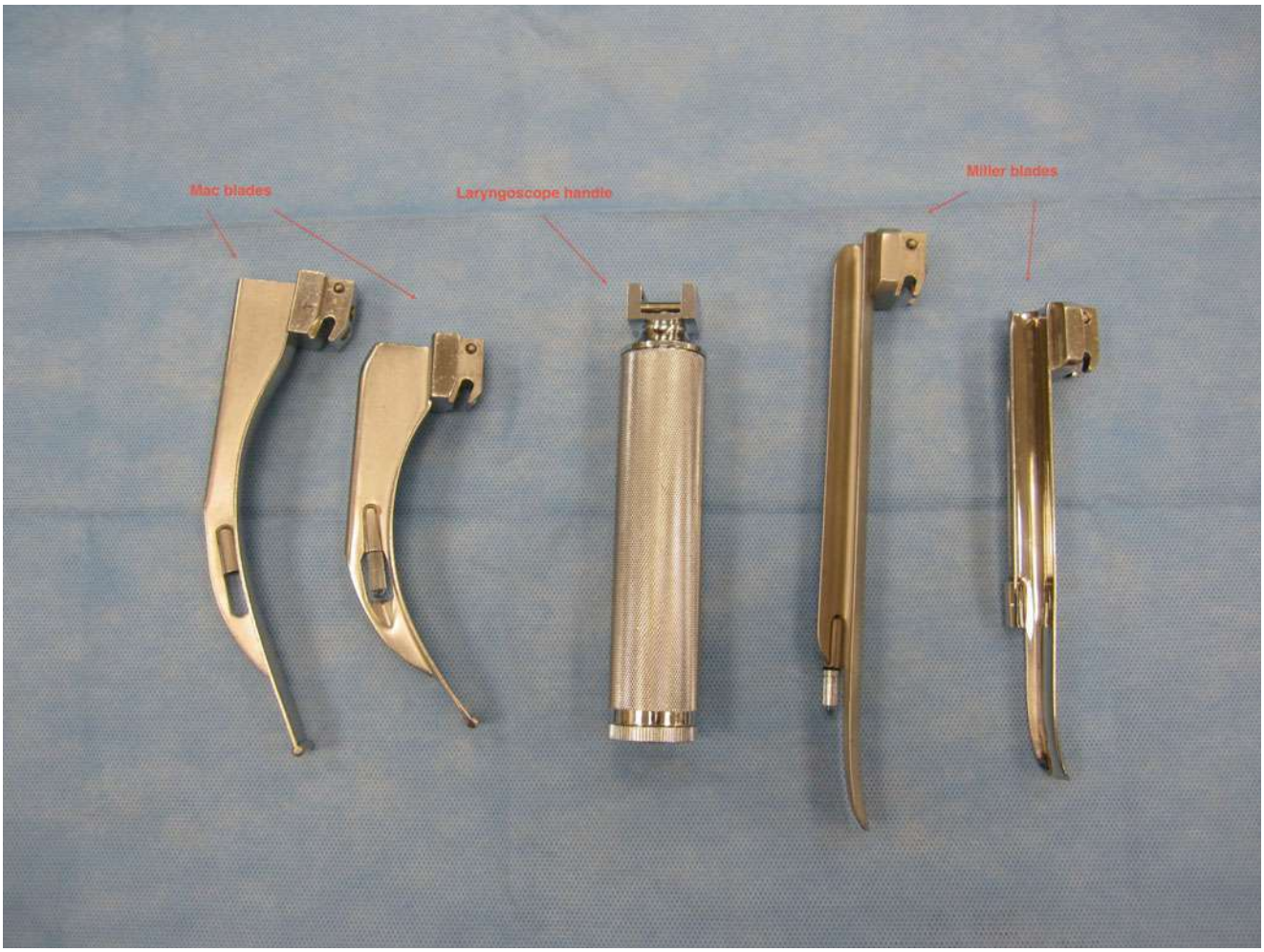

Fig. 9.1 Laryngoscope and blades

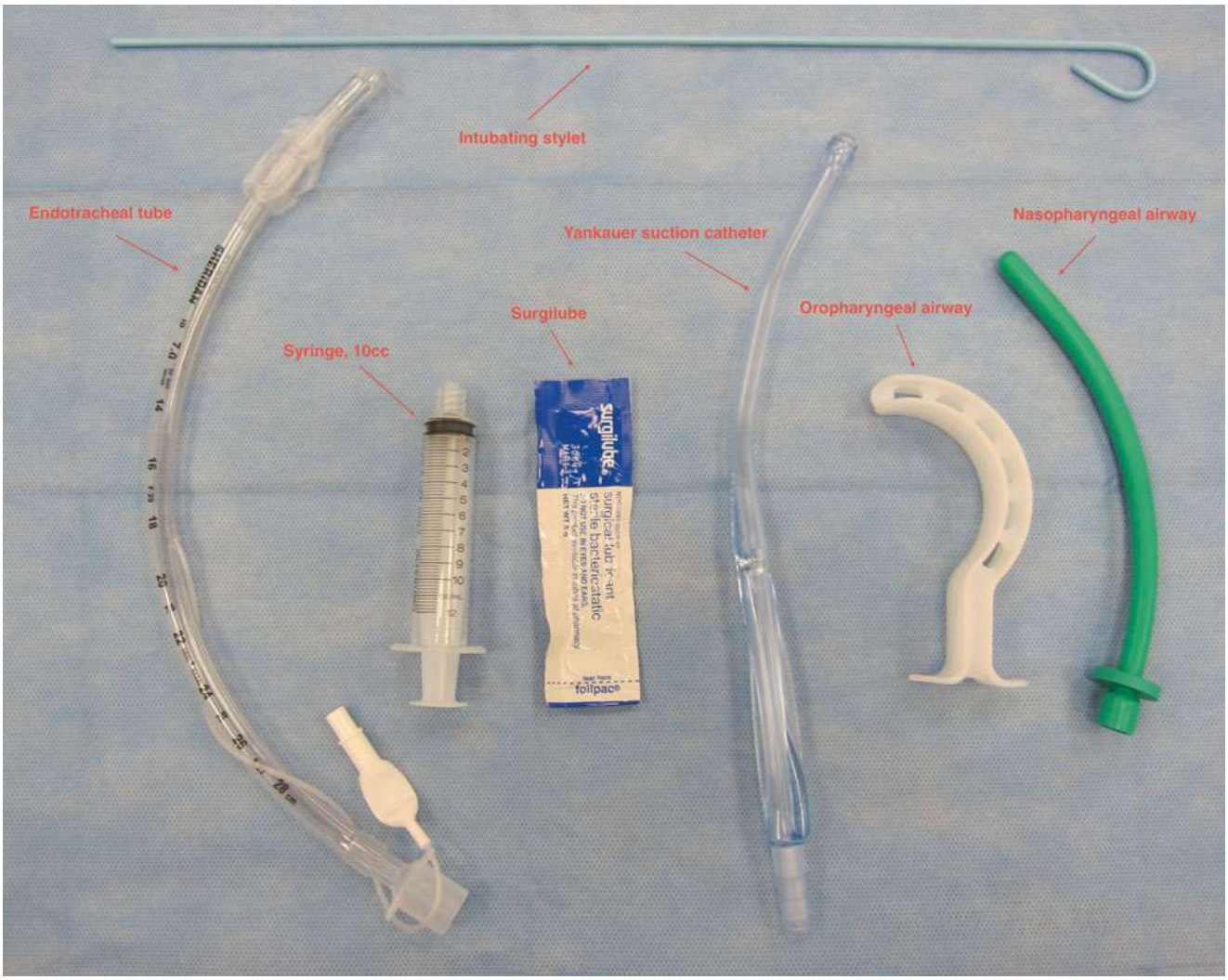

Fig. 9.2 Oral and nasal airways

9.4 Procedure

1. Preparation
   - Thoroughly assess patient for difficulty of intubation (Fig. 9.3).
   - Develop fallback plans for failed intubation attempt.
   - Establish at least one, but preferably two, secure intravenous (IV) lines.
   - Place on cardiac monitor with pulse oximetry, blood pressure monitoring, and continuous capnography.
   - Yankauer suction device attached to suction, suction on.
   - Pharmacological agents, drawn and labeled.
   - Laryngoscope and blades tested to ensure proper functioning of light source.
   - Desired ETT size, prepared for intubation:
     - Intubating stylet in position (tip at eye of ETT)
     - 10-mL syringe attached to ETT
     - ETT configured per proceduralist's preference (e.g., hockey stick, curved)
     - Cuff tested for air leak
2. Preoxygenation
   - Administer high-flow oxygen for 3–5 min (Fig. 9.4).
     - Nitrogen is exchanged for O_2 in the functional residual capacity of the lungs.
     - Establishes oxygen reservoir within lungs (primarily), blood, and body tissue.
     - Also known as nitrogen "washout."
   - This can be done using:
     - Non-rebreather masks—delivers 65–70 %
       - Difficult intubation not anticipated
     - Well-fitting bag-valve-masks (without positive-pressure ventilation)—delivers greater than 90 % oxygen (Fig. 9.5)
     - Noninvasive positive-pressure ventilation (NIPPV)—delivers 100 % oxygen
       - Consider NIPPV in high-risk patients with moderate to severe shunt physiology.
3. Pretreatment
   - Administer pharmacological agents to mitigate adverse physiological effects of intubation, induction, and paralysis, which may be undesirable in certain clinical circumstances (Table 9.1).
4. Induction and paralysis
   - Administer a rapidly acting induction agent to produce rapid loss of consciousness via IV push (Table 9.2).
   - Immediately follow induction agent with a neuromuscular-blocking agent via IV push (Table 9.3).
5. Positioning
   - If no cervical spine injury suspected, place the patient in the "sniff" position (Fig. 9.6):
     - Flex neck
     - Extend head
   - Ideally, the patient's pinna will be at the level of the sternum.
6. Direct laryngoscopy (see Chap. 10)
7. Proof of placement
   - Visualize ETT passing vocal cords
   - Confirm tube placement via $EtCO_2$:
     - Qualitative detection device—EZ Cap
     - Quantitative continuous $EtCO_2$ waveform on monitor (preferred) (Fig. 9.7)
   - Auscultation of breath sounds:
     - Lung fields bilaterally
     - Epigastric region (ensuring no breath sounds in the stomach)
8. Postintubation management
   - Secure ETT (Fig. 9.8)
   - Initiate mechanical ventilation.
   - Postintubation sedation and analgesia.
   - Postintubation chest x-ray.

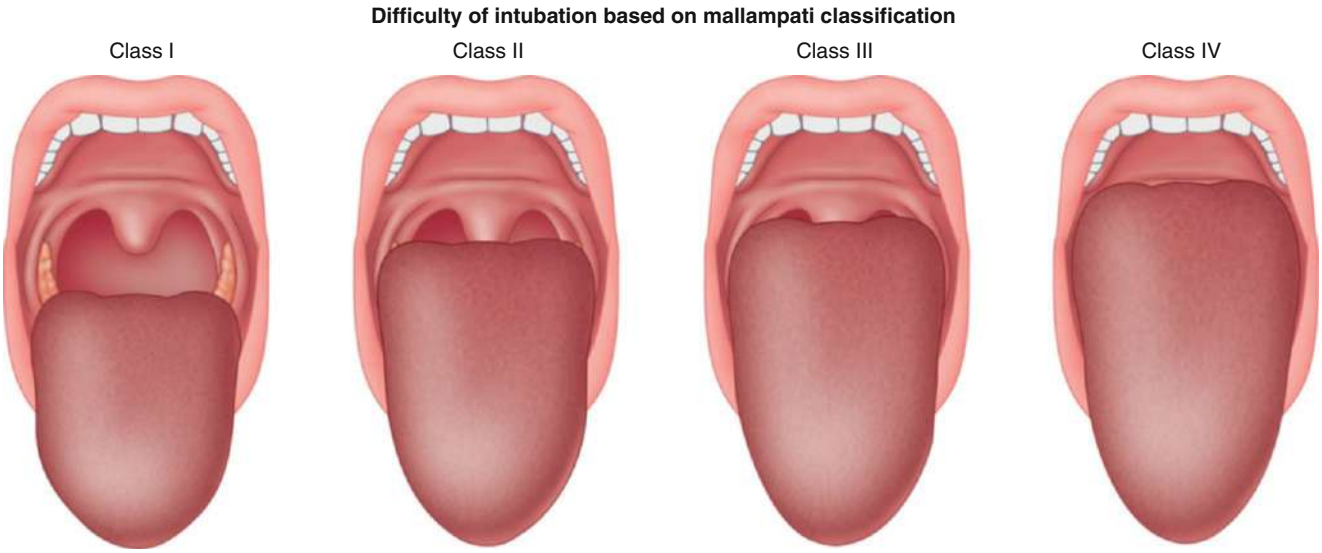

Fig. 9.3 Assess patient for difficulty of intubation

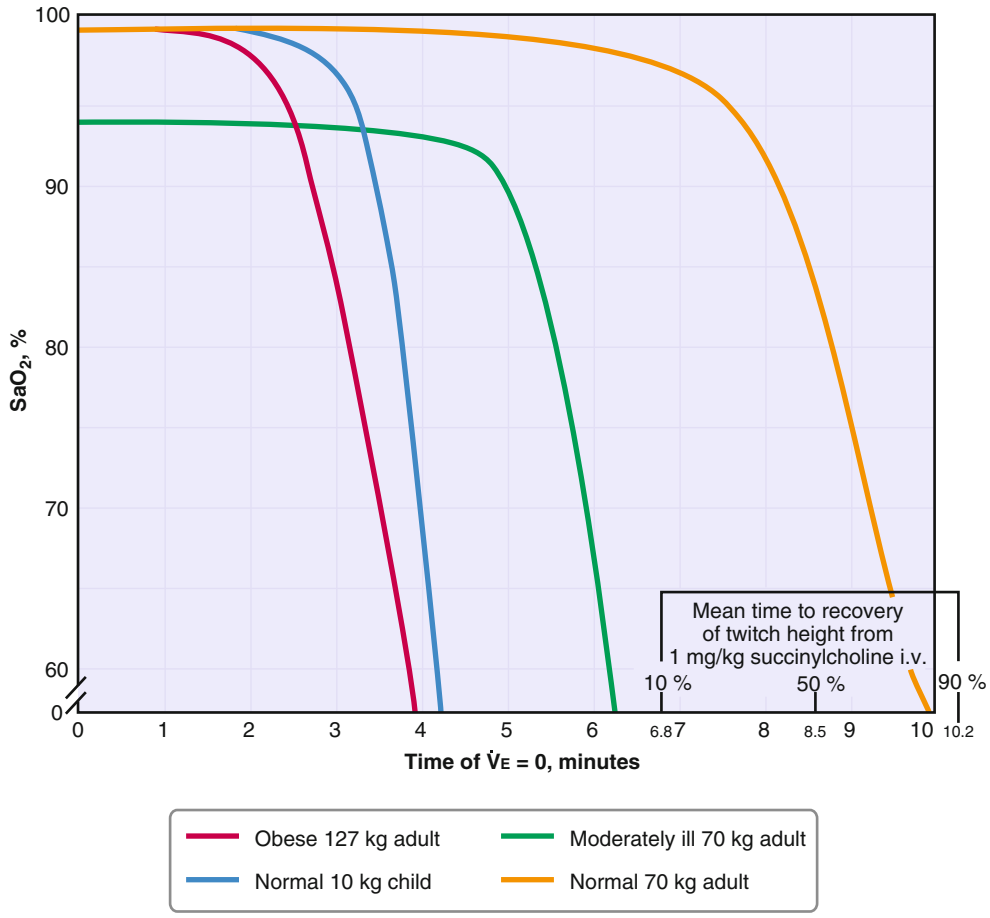

Fig. 9.4 *FaO2* fractional concentration of alveolar oxygen, *SaO2* arterial oxygen saturation, *Ve* expired volume per minute

Fig. 9.5 Preoxygenation materials

Table 9.1 Pretreatment: pharmacological agents used to mitigate adverse physiological effects of intubation, induction, and paralysis

Agent	Dose (IV)	Indication
Lidocaine	1.5 mg/kg Rapid push	Use in tight brain to attenuate ICP increase from laryngoscopy/ intubation; use in tight lungs to blunt bronchospastic response
Fentanyl	1–3 mcg/kg Slow push	Use in tight brain, tight heart, and tight vessels to blunt reflex sympathetic response to laryngoscopy

ICP intracranial pressure

Table 9.2 Induction: rapidly acting induction agents used to produce rapid loss of consciousness

Agent	Dose (IV) (mg/kg)	Onset (sec)	Duration (min)
Midazolam	0.2–0.3	60–90	15–30
Etomidate	0.3	15–45	3–12
Thiopental	3	<30	5–10
Ketamine	1.5–2.0	45–60	10–20
Propofol	1.5	15–45	5–10

Table 9.3 Paralysis: neuromuscular-blocking agents administered immediately after induction agent

Agent	Dose (IV)	Onset (sec)	Duration (min)
Succinylcholine	1.5 mg/kg	45	6–10
Rocuronium	1.0 mg/kg	60–75	40–60
Vecuronium	0.01 to prime, then 0.15 mg/kg	75–90	60–75

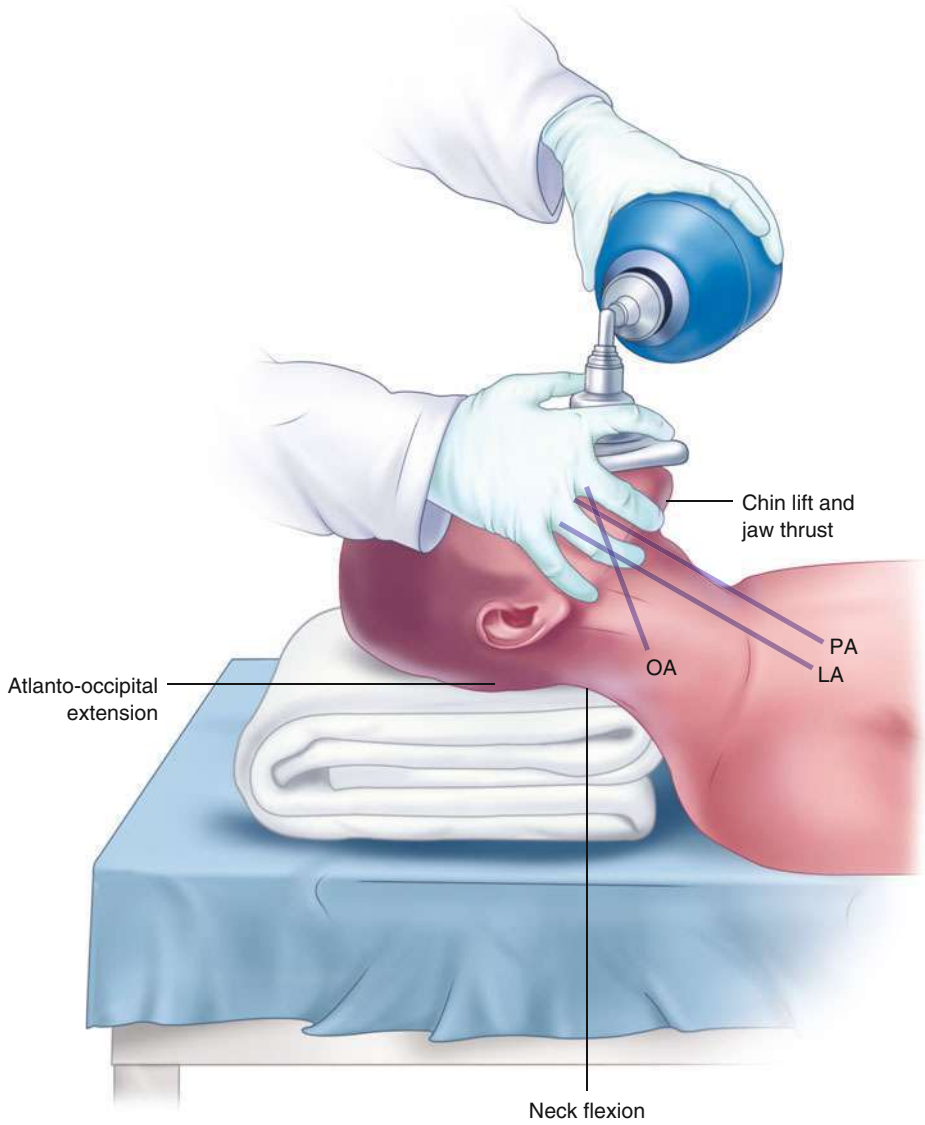

Fig. 9.6 Patient in the "sniff" position. *OA* oral axis, *LA* laryngeal axis, *PA* pharyngeal axis

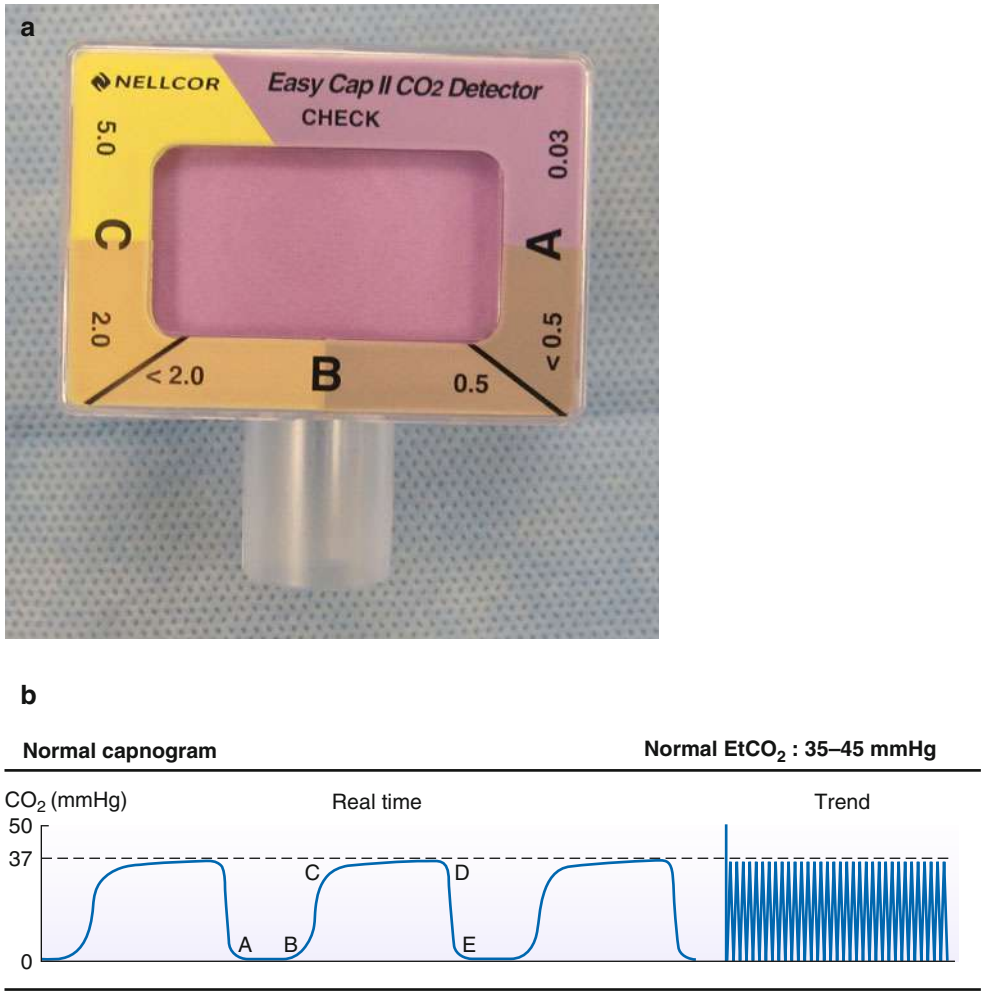

Fig. 9.7 $EtCO_2$, end-tidal carbon dioxide concentration: *a* Qualitative detection device—EZ Cap, *b* quantitative continuous $EtCO_2$ waveform on monitor

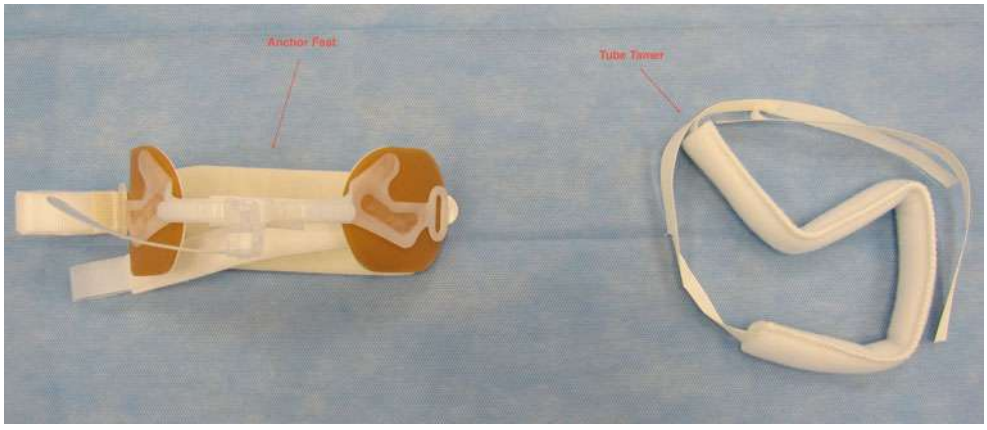

Fig. 9.8 Equipment used to secure ETT

9.5 Complications

- Esophageal intubation
- Right mainstem intubation
- Pneumothorax from laryngeal trauma
- Aspiration
- Dental trauma
- Vocal cord injury
- Hypotension
  - Induction agent
  - Decreased venous return from positive pressure
  - Pneumothorax
- Hyperkalemia (succinylcholine used in mildly hyperkalemic patient)
- Iatrogenically obstructed airway
- Failure to intubate

9.6 Pearls and Pitfalls

- Utilization of oral and nasal airways will greatly increase the ease of preoxygenation and reoxygenation.
- Inadequate preoxygenation will cause premature desaturation, decreasing time for intubation.
- Suboptimal patient positioning can prevent vocal cord visualization during direct laryngoscopy.
- Consider alternative airway devices before intubation to have a plan in place if a difficult airway is encountered.
- Underdosing induction or paralytic agent will prevent adequate time to perform procedure or create patient discomfort.
- Inappropriately sized McIntosh laryngoscopic blades.
  - Too small—more difficulty in moving tongue and epiglottis out of way for vocal cord visualization
  - Too large—easier to overshoot and go past vocal cords into the esophagus
- Excessive cricoid pressure (Sellick maneuver) may lead to disrupted laryngoscopic view and difficulty passing the ETT.
  - Disrupted view: readjust larynx using dominant hand to allow cord visualization.
  - Difficulty passing ETT: ask for release of some cricoid pressure to allow for ETT passage.
- Inadequate postintubation sedation and analgesia, especially when long-acting paralytics are used.
- Acidic gastric contents can cause CO_2 qualitative capnometry to change to yellow, falsely indicating tracheal placement of the ETT.

Selected Reading

Walls R, Murphy M. Manual of emergency airway management. Philadelphia: Lippincott Williams & Wilkins; 2008.

Weingart SD. Preoxygenation, reoxygenation, and delayed sequence intubation. J Emerg Med. 2011;40:661–7.

10 Direct Laryngoscopy

Bharath Chakravarthy and Weston Seipp

10.1 Indications

- Orotracheal intubation
  - Maintenance of oxygenation/ventilation
  - Airway protection
- Visualization of laryngeal anatomy
- Foreign body retrieval

10.2 Contraindications

- Absolute
  - None
- Relative
  - Presumed difficult airway
    - Anatomical limitations
      - Small oral opening (less than three of the patient's fingers)
      - Small mandible (hyomental distance less than three fingers)
      - Hyoid-thyroid distance (less than two fingers)
    - Clinical limitations
      - Patient with unstable cervical spine
      - Patient with multiple facial or neck trauma
      - Patient with history of tracheal stenosis, irradiation, or history of tracheal mass or surgery

B. Chakravarthy, MD, MPH (✉)
Department of Emergency Medicine,
University of California Irvine,
Orange, CA, USA
e-mail: bchakrav@uci.edu

W. Seipp, MD
Department of Emergency Medicine,
University of California Irvine Medical Center,
Orange, CA, USA

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_10

10.3 Materials and Medications (Fig. 10.1)

- Laryngoscope handle
- Laryngoscope blade with light
  - Macintosh blade ("Mac" or "curved blade")
  - Miller blade ("straight blade")
- Bag valve mask attached to 100 % O_2 source
- Endotracheal tube (ETT)
- 10-mL syringe
- Yankauer suction
- End-tidal CO_2 ($EtCO_2$) monitor (colorimetric or quantitative)
- McGill forceps (for foreign body retrieval)
- Postintubation chest radiograph

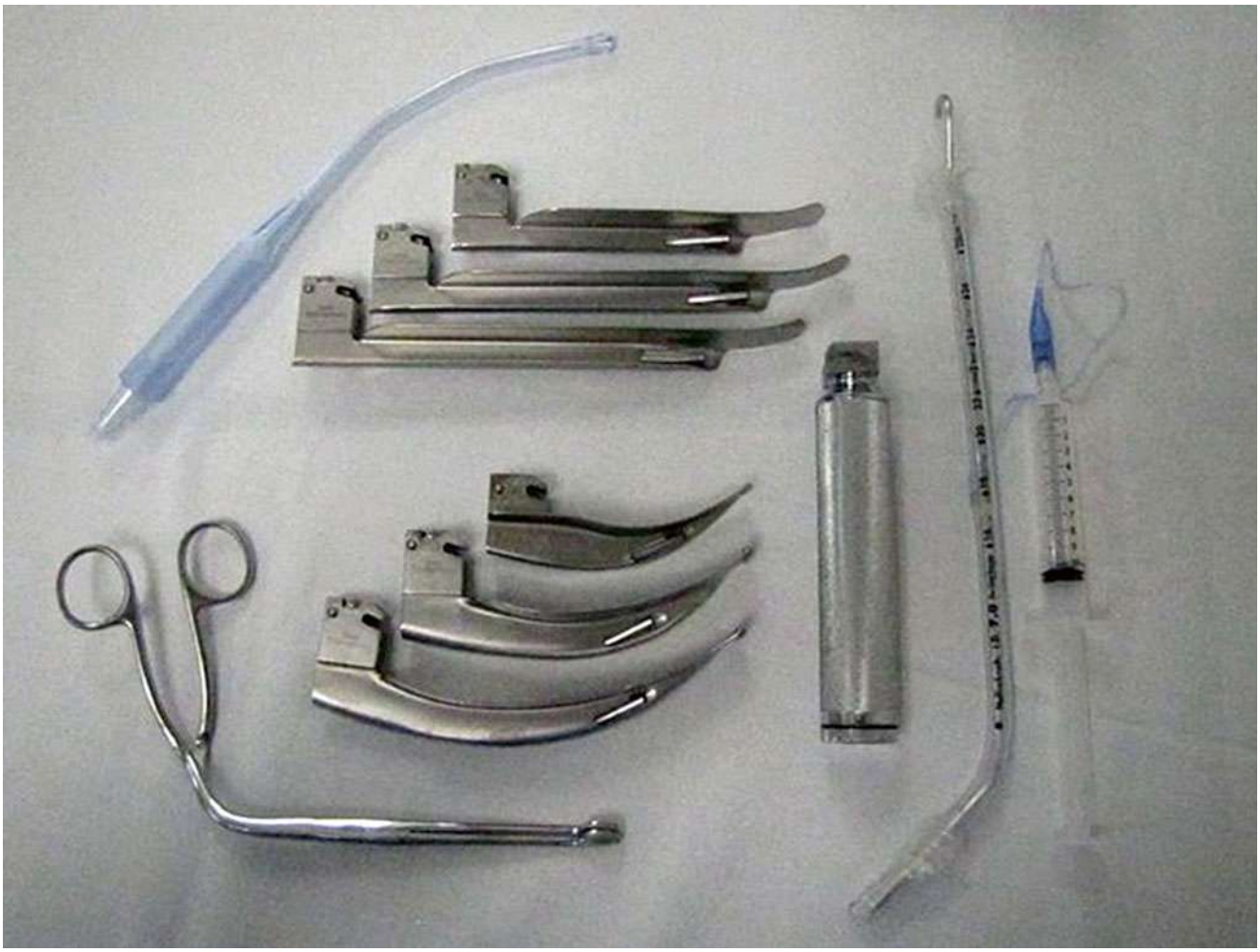

Fig. 10.1 Basic laryngoscopy supplies. *Clockwise from top left*: Yankauer suction, Miller Blades, endotracheal tube, 10-mL syringe, laryngoscope handle, Macintosh blades, and McGill forceps

10.4 Procedure

1. Check the laryngoscope handle and blades to ensure that the light is functioning.
2. Choose the appropriate blade based on patient size.
   (a) The Macintosh or Miller 3 size is appropriate for the majority of adults, and a 4 can be used for a larger body habitus.
   (b) The Macintosh blade is generally preferred in adults owing to increased space for ETT passage [1].
   (c) The Miller blade may be preferable in cases in which the patient has limited mouth opening (owing to its smaller vertical height), or in which the airway is particularly floppy (such as in infants and children) [1].
3. Position the patient (Fig. 10.2).
   (a) Raise the bed so that the patient's oral opening is at the level of the xiphoid process of the laryngoscopist.
   (b) The optimal laryngeal view is obtained in the neck flexion/head extension or "sniffing" position. To achieve this, place towels under the patient's occiput to raise it approximately 6–9 cm [2].
   (c) If patient is in cervical spine precautions, then an assistant must hold the cervical spine in midline immobilization throughout the laryngoscopy; elevation of the occiput is therefore contraindicated.
4. Provide 100 % O_2 via face mask to preoxygenate the patient before laryngoscopy.
5. After ensuring adequate anesthesia and neuromuscular blockade (if performing laryngoscopy for intubation), perform the scissor technique to open the patient's mouth and to lift the tongue base from the glottic opening.
6. Macintosh blade insertion (Fig. 10.3).
   (a) Insert the laryngoscope into the patient's mouth, starting from the right side, and slowly advance into the oropharynx, using the blade's vertical flange to "sweep" the tongue to the left and away from the glottic opening.
   (b) Advance the blade along the tongue toward the tongue base, until the epiglottis and posterior arytenoids are viewed.
   (c) In order to expose the cords, insert the Macintosh blade into the vallecula, which is the potential space anterior to the epiglottis and posterior to the tongue base. This will act as a fulcrum and raise the epiglottis, exposing the vocal cords.
   (d) To further expose the cords and/or expose the vallecula, exert force outward at a 45° angle to the patient. Do not "rock" the laryngoscope because this may cause injury to the teeth.
7. Miller blade insertion (Fig. 10.4).
   (a) Insert the blade into the right side of the mouth and slowly advance along tongue toward the tongue base. The Miller blade does not have a flange for isolating the tongue, and thus the Macintosh may be preferable in patients with large tongues.
   (b) Advance the blade along the right side of the tongue until the epiglottis and posterior arytenoids are visible.
   (c) In contrast to the Macintosh blade, the Miller blade is used to directly isolate the epiglottis and expose the vocal cords. Using the tip of the Miller, move the epiglottis anteriorly to expose the vocal cords.
   (d) As with the Macintosh, exert force outward at a 45° angle to the patient to increase the view of vocal cords. Do not "rock" the laryngoscope because this may cause injury to the teeth.
8. Assessing the glottic view (Fig. 10.5).
   (a) With the epiglottis either directly or indirectly lifted from the glottic opening, assess the Cormack-Lehane laryngeal view grade.
      (i) Grade I—view of entire laryngeal opening, including cords
      (ii) Grade II—view of posterior laryngeal cartilages
      (iii) Grade III—visualization of epiglottis only
      (iv) Grade IV—no structures visualized
   (b) A lower grade (higher quality, better) view is predictive of intubation success [3].
   (c) In the event of a higher grade view, the operator may request airway adjuncts, such as a bougie.
9. Improving the glottic view [4].
   (a) With the laryngoscope in the desired position, it is possible to improve the glottic view by exerting backward pressure on the thyroid cartilage either with the operator's right hand (bimanual laryngoscopy) or with an assistant applying BURP (backward-upward-rightward pressure) (Fig. 10.5a).
   (b) Backward pressure increases the vertical distance between the epiglottis and the posterior cartilages, thereby increasing the likelihood of vocal cord visualization.
10. Passing the ETT (Fig. 10.6).
   (a) With the optimal view of the cords obtained, pass the ETT from the right corner of the mouth through the vocal cords, to a depth of 21 cm at the incisors in females and 23 cm in males.
   (b) Inflate the ETT cuff with approximately 5 cc of air until the pilot balloon is firm to touch.
11. After completing intubation or after completion of laryngoscopy, slowly remove the blade from the mouth, taking care to avoid dental or lip trauma.
12. Attach the capnography device to the ETT tube to ensure $EtCO_2$ return.
   (a) Colorimetric devices will turn from purple to yellow in the presence of $EtCO_2$.
   (b) Quantitative devices will return a CO_2 waveform.
13. Attach the ETT to ventilator or bag-valve-mask connected to an oxygen source.

14. Auscultate breath sounds in both lung fields and ensure absence of breath sounds over the epigastrium (which could signify esophageal intubation).
15. Obtain a postintubation chest x-ray to ensure no right mainstem intubation or pneumothorax.

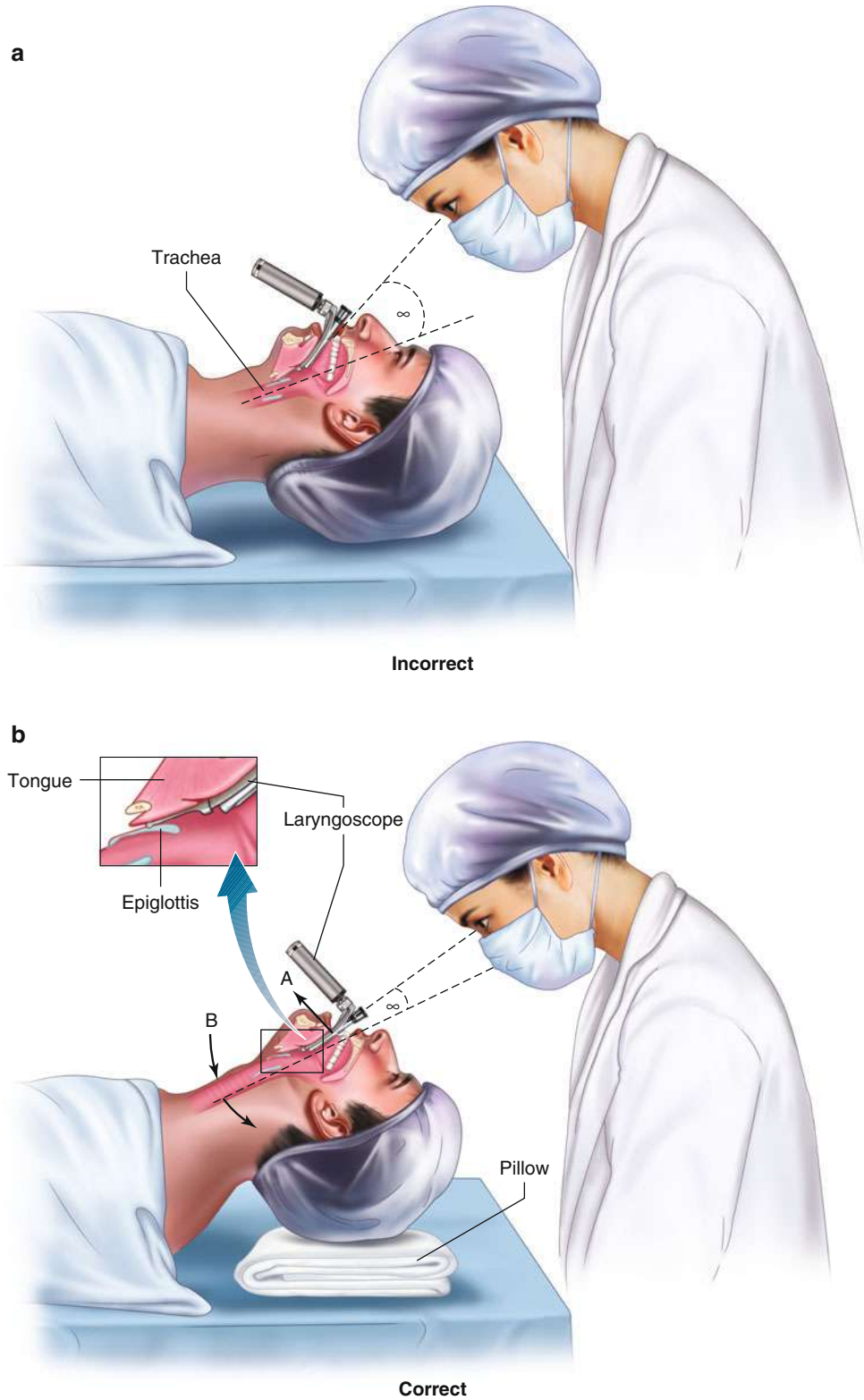

Fig. 10.2 Visualization axis and sniffing position: (**a**) The patient's occiput is not elevated and the neck is not in flexion, thereby creating a steep visual axis, (**b**) the occiput is correctly elevated 6–9 cm, placing the patient in sniffing position and allowing the visual axis to align with the airway axis

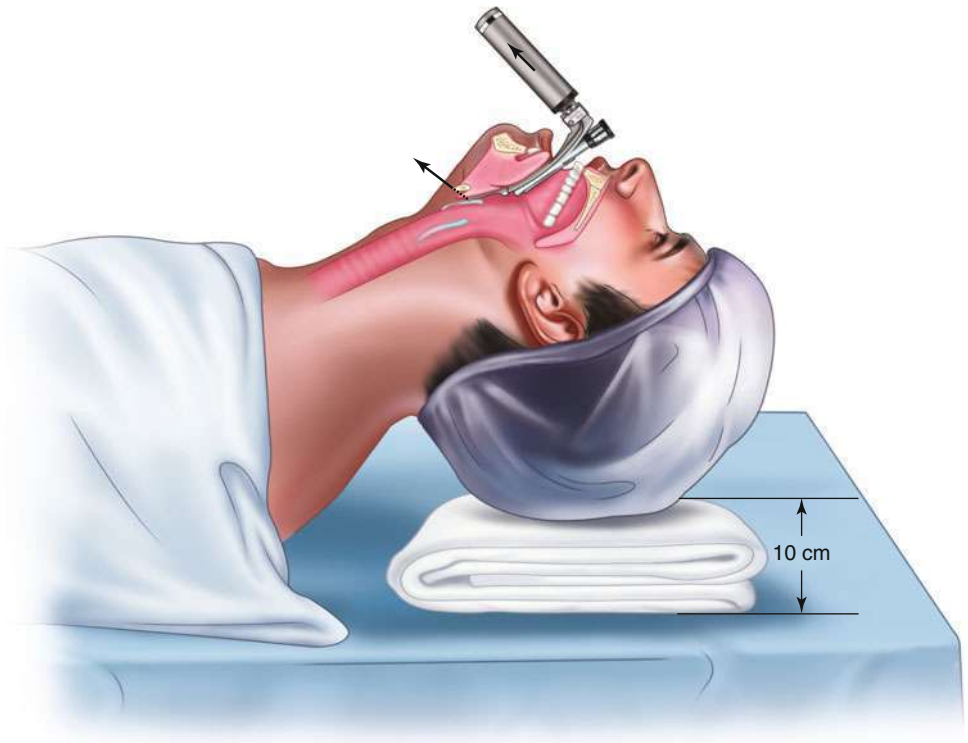

Fig. 10.3 Macintosh blade insertion. The blade is inserted into the vallecula, which raises the epiglottis and exposes the glottic opening

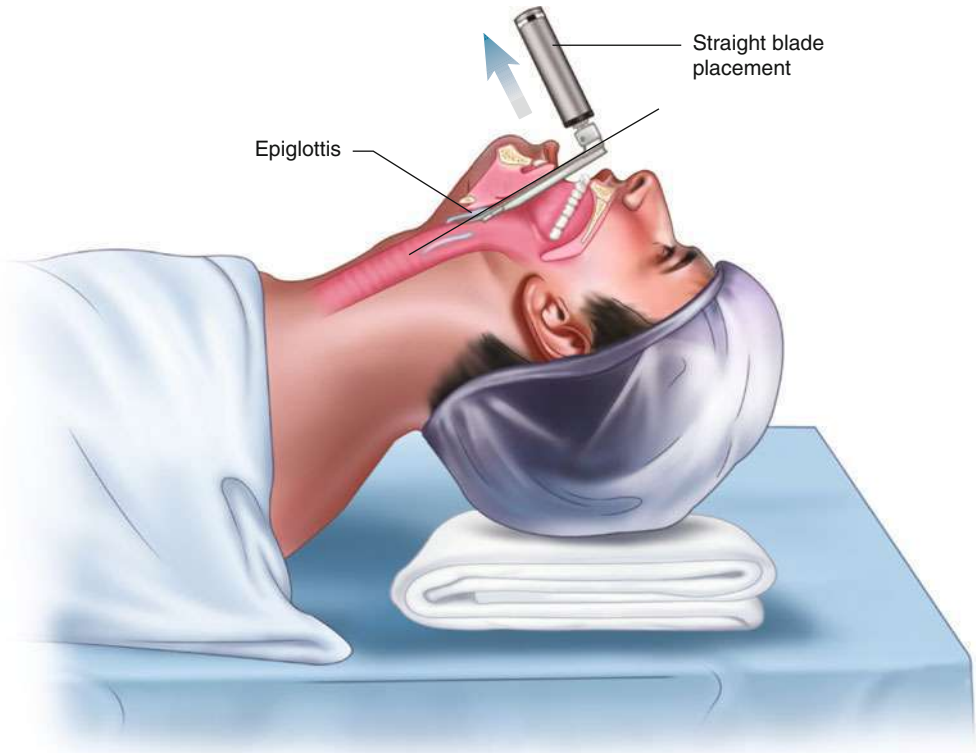

Fig. 10.4 Miller blade insertion. The blade is used to elevate the epiglottis directly, exposing the glottic opening

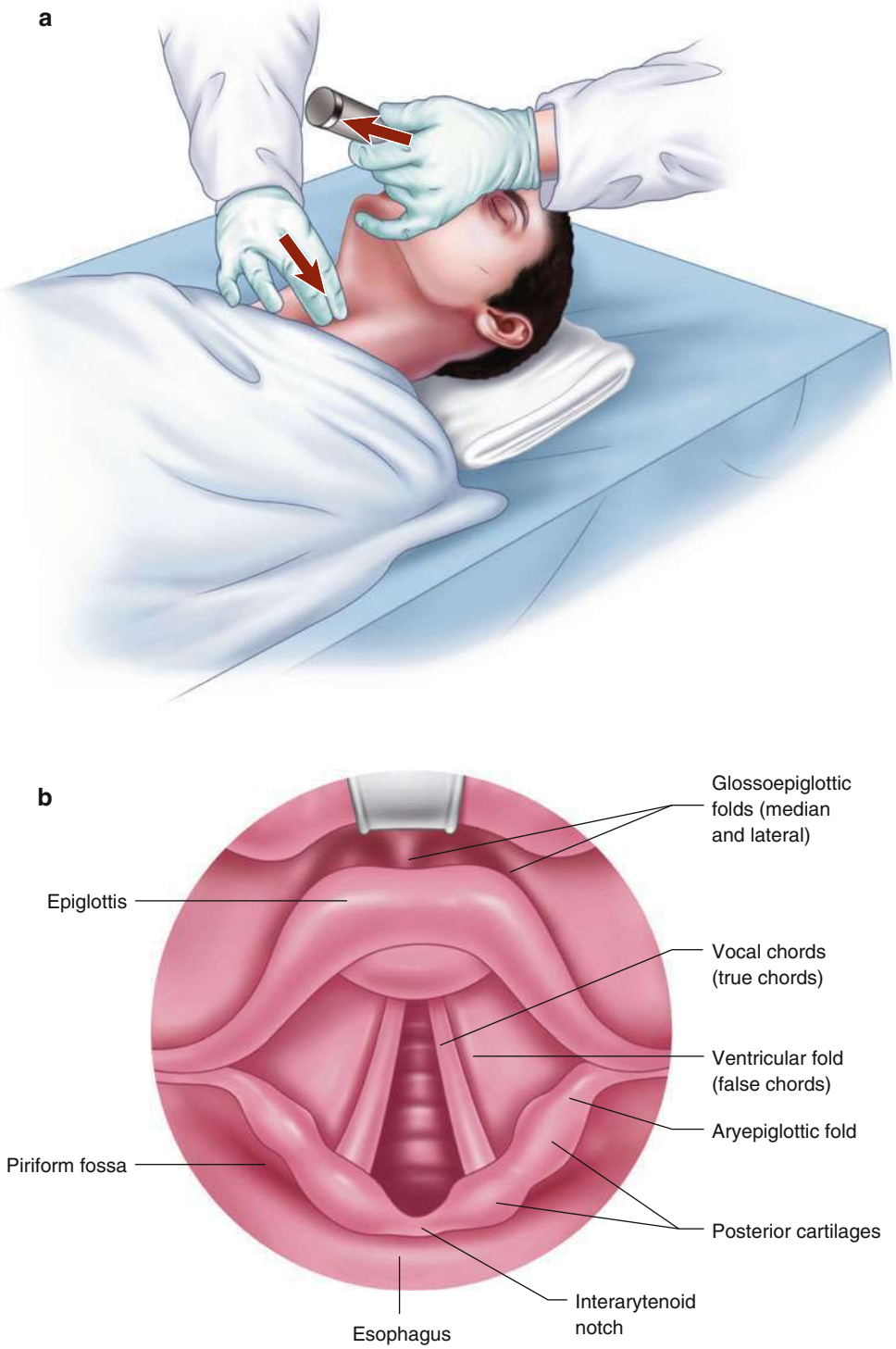

Fig. 10.5 (**a**) Bimanual laryngoscopy. The force on the neck is opposite the direction of lift by the laryngoscope, (**b**) laryngoscopy view, (**c**) Cormack and Lehane

Fig. 10.5 (continued)

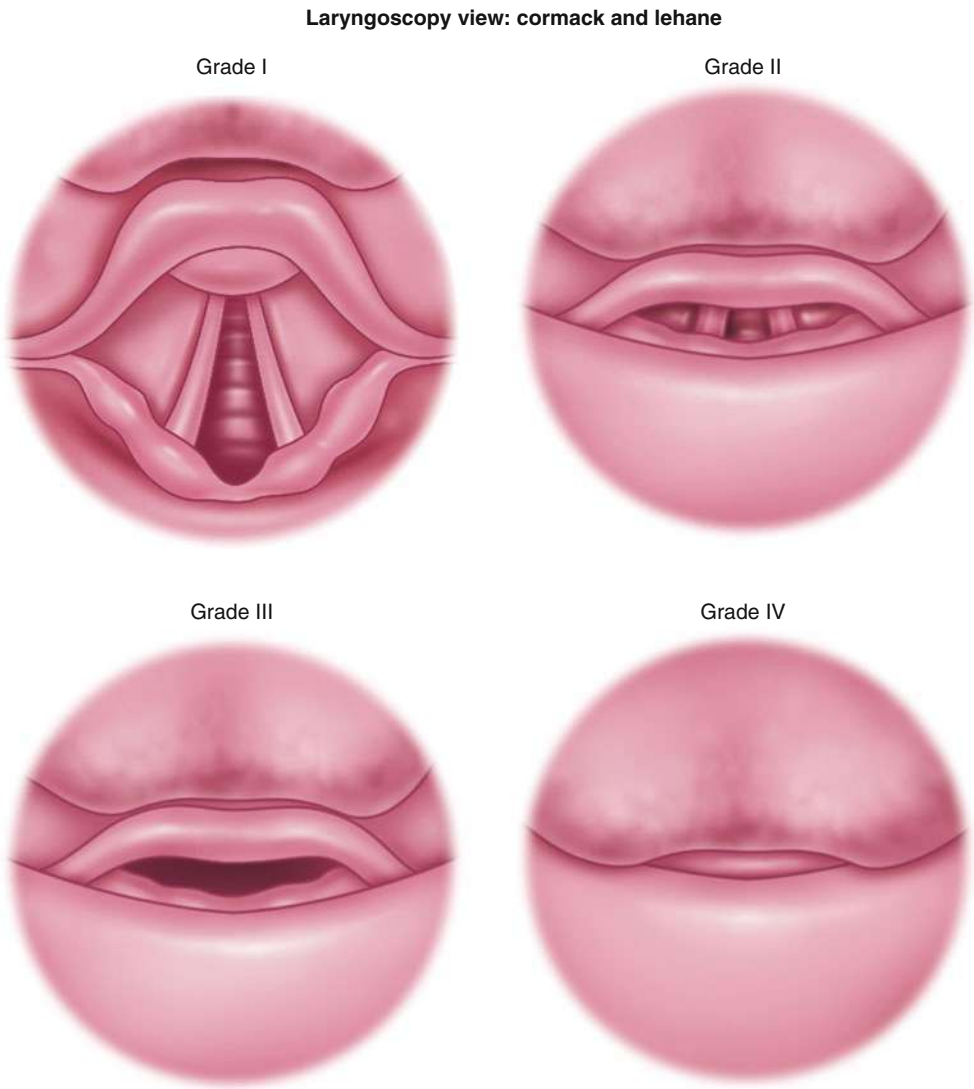

Fig. 10.6 (**a**) Insert the laryngoscope into the mouth, (**b**) sweep the tongue to the left, (**c**) advance the laryngoscope until the epiglottis is visible, and insert the blade into vallecula to expose the cords. When using the Miller blade, insert the blade until the epiglottis is seen. Slide the blade under the epiglottis and lift to expose the cords

10.5 Complications [3, 5]

- Common (1–4 %)
  - Esophageal intubation—*can be fatal if unrecognized*
  - Mainstem bronchus intubation
  - Tachycardia
  - Aspiration
  - Hypotension
- Uncommon (<1 %)
  - Dental/oral trauma
  - Oropharyngeal edema or bleeding
  - Laryngospasm
  - Dysrhythmia
  - Pneumothorax
  - Cardiac arrest

10.6 Pearls and Pitfalls

- Pearls
  - Positioning is of key importance—all patients with a stable cervical spine should be placed in the "sniffing" position to maximize view.
  - Consider the "ramping" position in obese patients with stable cervical spines—elevation of the head and shoulders allows redundant tissue to fall and gives an improved glottic view.
  - Always have suction readily available to remove blood, vomitus, or edema.
  - If structures are not readily visible, withdraw the blade gradually because it is common to insert the blade too deep.
- Pitfalls
  - "Rocking" the laryngoscope instead of lifting outward
  - Failure to recognize esophageal intubation
  - Failure to evaluate postintubation chest x-ray

References

1. Hagberg CA. Benumof's airway management. Maryland Heights, MO: Mosby; 2007. p. 363–5.
2. Park SH, Park HP, Jeon YT, Hwang JW, Kim JH, Bahk JH. A comparison of direct laryngoscopic views depending on pillow height. J Anesth. 2010;24:526–30.
3. Martin LD, Mhyre JM, Shanks AM, Tremper KK, Kheterpal S. 3,423 emergency tracheal intubations at a university hospital: airway outcomes and complications. Anesthesiology. 2011;114:42–8.
4. Levitan RM, Kinkle WC, Levin WJ, Everett WW. Laryngeal view during laryngoscopy: a randomized trial comparing cricoid pressure, backward-upward-rightward pressure, and bimanual laryngoscopy. Ann Emerg Med. 2006;47:548–55.
5. Walls RM, Brown CA 3rd, Bair AE, Pallin DJ, NEAR II Investigators. Emergency airway management: a multi-center report of 8937 emergency department intubations. J Emerg Med. 2011;41:347–54.

11 Laryngeal Mask Airway

Sohan Parekh

11.1 Indications

- Rescue device in a failed intubation
- Initial device in a predictably difficult airway
- Temporizing airway prior to definitive endotracheal intubation or surgical airway

11.2 Contraindications

- Absolute
  - Inadequate mouth opening
- Relative
  - Neck trauma/injury/radiation
  - High risk of aspiration

S. Parekh, MD
Department of Emergency Medicine,
University of Texas at Austin Dell Medical School,
Austin, TX, USA

Department of Emergency Medicine,
University Medical Center Brackenridge,
Austin, TX, USA
e-mail: sohan.parekh@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_11

11.3 Types (Fig. 11.1)

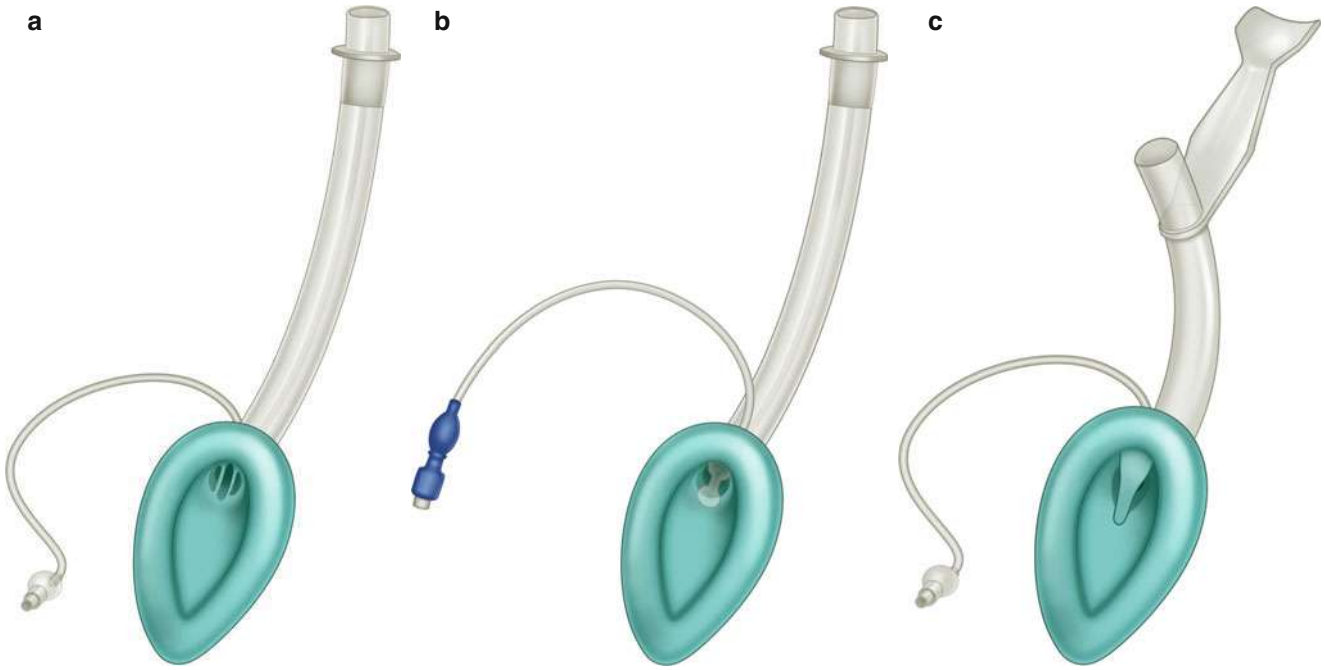

Fig. 11.1 Types of laryngeal mask airway: (**a**) LMA Unique, (**b**) LMA Classic Excel, (**c**) LMA Fastrach

11.4 Materials and Medications

- Appropriately sized laryngeal mask airway (LMA; LMA Unique/LMA Classic Excel/LMA Fastrach) and corresponding syringe (Table 11.1)
- Surgilube
- Bag valve mask
- Continuous end-tidal CO_2 ($EtCO_2$) or colorimetric $EtCO_2$ detector
- 8-mm or smaller endotracheal tube (ETT) (for Fastrach intubating LMA only)

Table 11.1 Laryngeal mask airway sizing

Size	Patient weight (kg)	Maximum cuff inflation volume (mL)	LMA product availability
1	<5	4	Unique
1½	5–10	7	Unique
2	10–20	10	Unique
2½	20–30	14	Unique
3	30–50	20	Unique, Classic Excel, Fastrach
4	50–70	30	Unique, Classic Excel, Fastrach
5	70–100	40	Unique, Classic Excel, Fastrach
6	>100	50	Unique

LMA laryngeal mask airway

11.5 Procedure

11.5.1 LMA Unique or Classic Excel

1. If using a reusable LMA Classic Excel, ensure that it is sterile and inspect it for any damage or wear.
2. Tightly deflate the cuff using a syringe such that it forms a spoon shape (Fig. 11.2).
3. Lubricate the posterior surface of the LMA with sterile lubricating jelly.
4. Stand behind the patient at the head of the bed as in direct laryngoscopy.
5. Place the patient's head in the sniffing position and ensure proper induction and paralysis.
6. Hold the LMA with the index finger of the dominant hand positioned at the juncture of the tube and cuff (Fig. 11.3).
7. Widely open the mouth with the nondominant hand and insert the LMA with the flattened tip flush with the palate.
   - Ensure that the tip of the device does not fold over during insertion.
8. Using the index finger, push the LMA along the curvature of the hard and soft palate (Fig. 11.4).
9. Continue to insert the LMA into the hypopharynx until resistance is felt. (At this point the tip of the LMA is in the esophagus.)
10. Stabilize the tube with the nondominant hand and remove index finger of the dominant hand from the LMA.
11. Inflate the cuff of the LMA to at least half of the maximum value using a syringe.
    - The LMA might move slightly outward during cuff inflation as the LMA positions itself in the hypopharynx.
12. Confirm placement and adequate gas exchange with continuous $EtCO_2$ capnography or colorimetry.

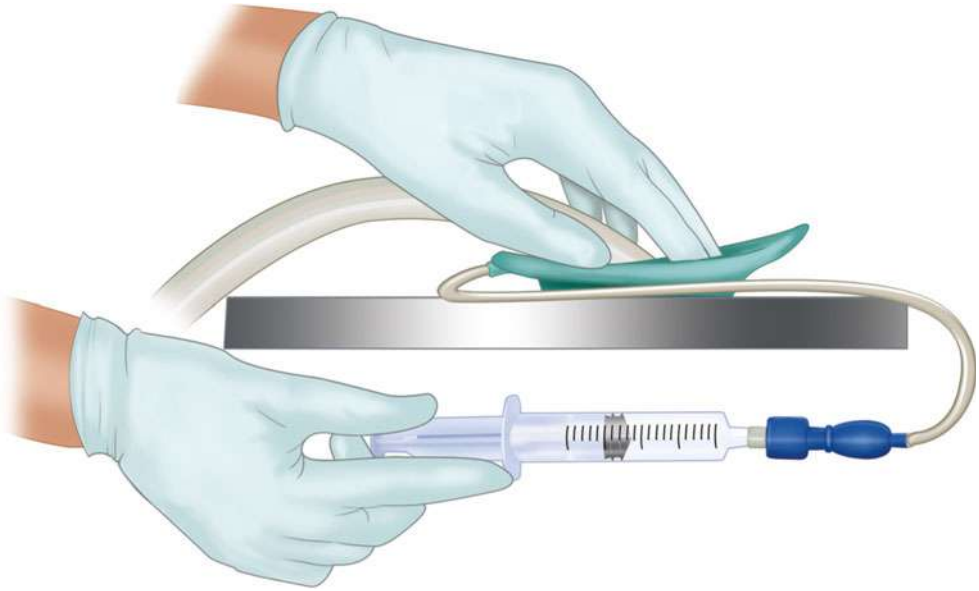

Fig. 11.2 Deflate the laryngeal mask airway (LMA) cuff using a syringe

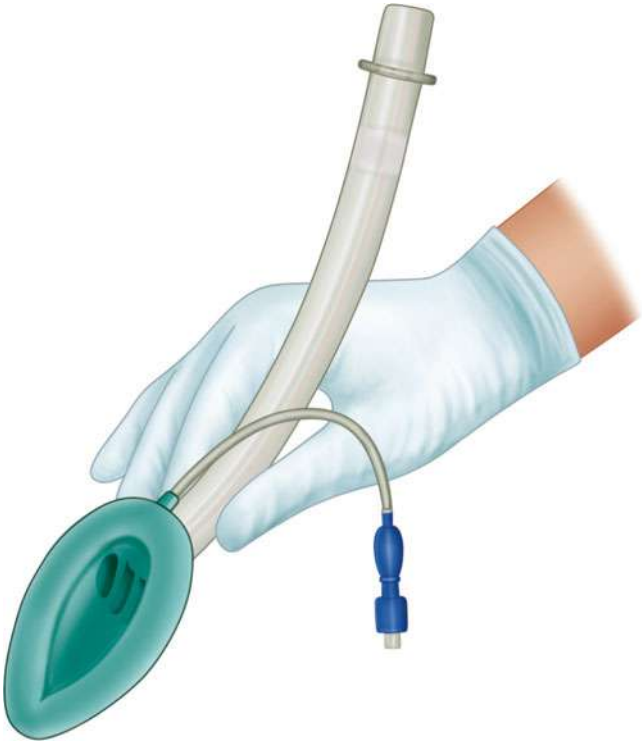

Fig. 11.3 Hold the LMA with the index finger positioned at the juncture of the tube and the cuff

Fig. 11.4 Use the index finger to guide the LMA along the hard and soft palate

11.5.2 LMA Fastrach

1. If using a reusable LMA Fastrach, ensure that it is sterile and inspect it for any damage or wear.
2. Tightly deflate the cuff using a syringe such that it forms a spoon shape.
3. Lubricate the posterior surface of the LMA with sterile lubricating jelly.
4. The LMA Fastrach may be inserted from any position with respect to the patient's head.
5. Position the patient's head in the neutral position. Do not extend the head.
6. Widely open the mouth with the nondominant hand.
7. Holding the handle of the LMA Fastrach, insert the device into the mouth, placing the deflated cuff flush with the superior palate.
   - Distribute the lubricant over the superior palate using a side-to-side motion to allow for easier insertion.
   - Ensure that the tip of the device does not fold over during insertion.
8. Using the handle, gently advance the LMA Fastrach directly into the oropharynx until the curved portion of tube comes into the contact with the patient's chin (Fig. 11.5).
9. At this point use the handle to rotationally advance the device further into the oropharynx following the natural curvature of the palate and posterior pharynx (Fig. 11.6).
   - Do not initiate any rotation until the tube is in contact with the patient's chin.
10. Once resistance is felt, inflate the cuff of the device to at least half of the maximum value using a syringe.
    - Note that the tube is directed slightly caudally when properly inserted.
    - Confirm placement and adequate gas exchange with $EtCO_2$ capnography or colorimetry.

Fig. 11.5 Using the handle, insert the LMA Fastrach such that the posterior surface is in contact with the superior palate

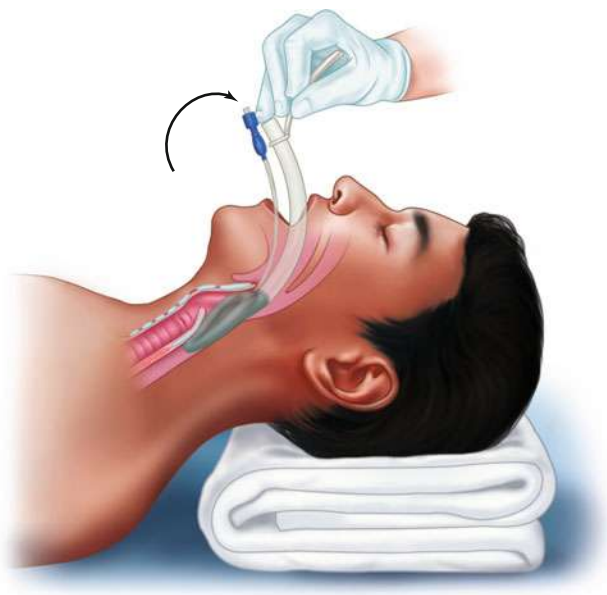

Fig. 11.6 Once the tube is in contact with the chin, use the handle to rotationally advance the device further into the oropharynx

11.5.3 Endotracheal Intubation through the LMA Fastrach

1. Ensure that the ETT will pass freely in the LMA.
2. Lubricate the cuff of the ETT.
3. Firmly hold the handle of the LMA Fastrach with the nondominant hand and insert the ETT to a depth of 15 cm (Fig. 11.7), which places the ETT tip at the point of emergency from LMA Fastrach.
   - Ensure that the tube does not pass beyond 15 cm at this point.
4. Using the handle of the LMA Fastrach, draw the device outward in order to displace the larynx slightly to accommodate insertion of the ETT (Fig. 11.8).
   - Use a lifting rather than a levering motion.
5. Carefully advance the ETT slightly further. If no resistance is felt, continue with insertion of the ETT (Fig. 11.9).
6. Confirm placement and adequate gas exchange with $EtCO_2$ capnography or colorimetry.
7. Once successful confirmation of intubation is established, deflate the cuff pressure on the LMA Fastrach.

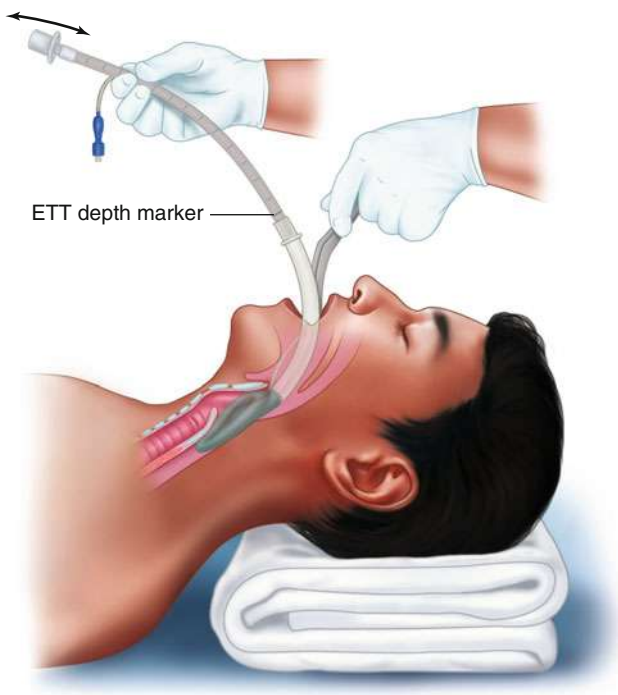

Fig. 11.7 While holding the handle of the LMA, insert the endotracheal tube (*ETT*) to the 15-cm mark

Fig. 11.8 Lift the handle outward to open the glottis for the ETT

Fig. 11.9 If no resistance is felt during further insertion of the ETT, the ETT may be fully advanced

11.5.4 Removing the LMA Fastrach after Successful Intubation

1. The LMA Fastrach need not be removed immediately, but if this is desired, first adequately oxygenate the patient and then disconnect the patient from the circuit.
2. Remove the airway connector from the proximal end of the ETT.
3. Ensure that the cuff of the LMA Fastrach is entirely deflated.
4. Stabilize the ETT with the nondominant hand, and using the dominant hand, gently ease the LMA Fastrach out by rotating the handle caudally (Fig. 11.10).
5. Once the tube of the LMA Fastrach reaches the proximal end of the ETT, use the stabilizer rod to maintain the position of the ETT while continuing to remove the LMA Fastrach using the handle (Fig. 11.11).
6. After the cuff of the LMA Fastrach has been fully removed from the oral cavity, release the stabilizer rod and ensure stability of the ETT by grasping it distally at the mouth with the nondominant hand (Fig. 11.12).
7. Continue to ease the LMA Fastrach out from around the ETT, ensuring that the pilot balloon and inflation line of the ETT cuff pass through the device (Fig. 11.13).
   - Take care not to rupture the pilot balloon or tear the inflation line of the ETT.
8. Replace the airway connector on the proximal end of the ETT and reconnect the patient to the circuit.

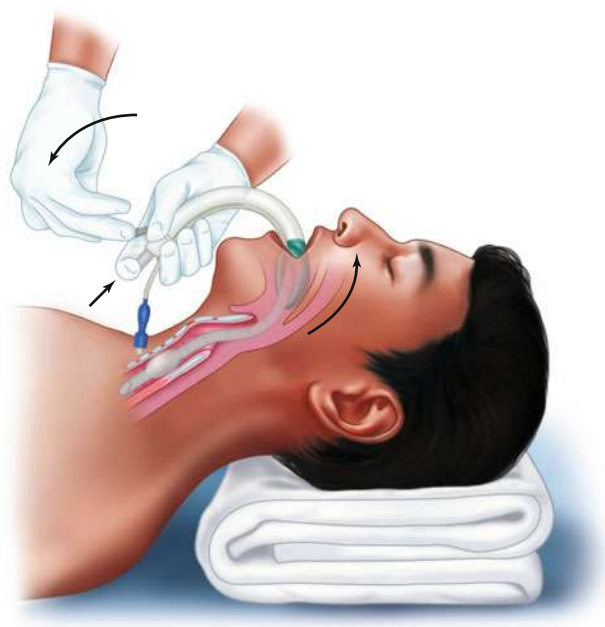

Fig. 11.10 Rotate the handle caudally to gently ease the LMA Fastrach out of the pharynx

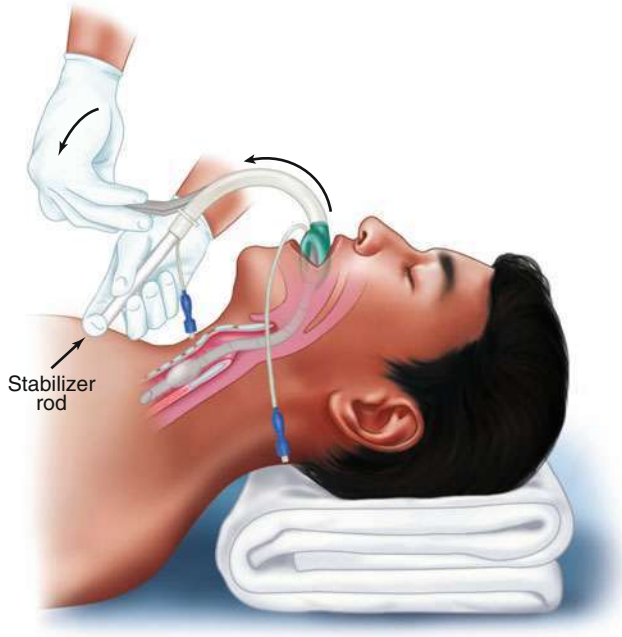

Fig. 11.11 Use the stabilizer rod to allow for further removal of the LMA Fastrach

Fig. 11.12 Once the cuff of the LMA Fastrach is out of the mouth, grasp the ETT distally and remove the LMA entirely

Fig. 11.13 Carefully pass the pilot balloon and inflation line of the ETT cuff through the tube of LMA Fastrach as it is removed

11.6 Complications

- Aspiration with resulting pneumonitis
- Ineffective seal resulting in insufficient ventilation
- Coughing, bucking, or breath holding
  - Ensure that the patient is adequately sedated.

11.7 Pearls and Pitfalls

- Cricoid pressure can push the tip of the LMA out of the esophagus and prevent optimal placement.

Selected Reading

Barata I. The laryngeal mask airway: prehospital and emergency department use. Emerg Med Clin North Am. 2008;24:1069–83.

LMA classic LMA flexible LMA classic single use LMA flexible single use instruction manual. Victoria: The Laryngeal Mask Company Limited; 2010.

LMA fastrach LMA fastrach single use instruction manual. Victoria: The Laryngeal Mask Company Limited; 2010.

Pollack CV. The laryngeal mask airway: a comprehensive review for the emergency physician. J Emerg Med. 2001;20:53–66.

Walls RW, Murphy MF, editors. Manual of emergency airway management. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

12 Combitube

Clint Masterson

12.1 Indications

- Need for ventilation and oxygenation in an unconscious, unresponsive, or paralyzed patient
- Rescue airway needed after failed intubation

12.2 Contraindications

- Absolute
  - Awake, responsive patient
  - Intact gag reflex
  - Known esophageal disease
  - Ingestion of caustic substances
  - Child (no Combitubes are made for children)
- Relative
  - D50 or naloxone about to be given
  - Facial trauma

12.3 Materials

- Combitube sized based upon height (Fig. 12.1)
  - >5 ft—size 41 French (cuff inflation 15 and 100 mL)
  - >4 ft to < 5.5 ft—size 37 French (cuff inflation 12 and 85 mL)

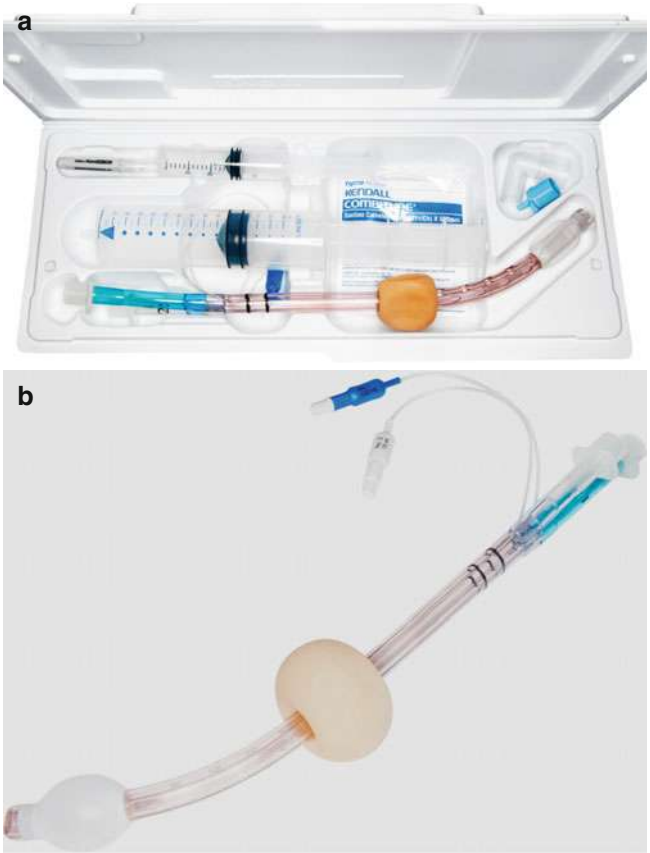

Fig. 12.1 (**a**, **b**) Combitube equipment

C. Masterson, MD
Department of Emergency Medicine,
Mayo Clinic Health System in Fairmont,
Fairmont, MN, USA
e-mail: clint@clintm.us

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_12

12.4 Procedure

1. Test both balloons and cuffs for leaks as one would an endotracheal tube (ETT).
2. Open up the airway.
   (a) Use a laryngoscope to move the tongue and open the oropharynx.
   OR
   (b) Use the left hand to elevate the chin, elevating the tongue and pharyngeal tissue.
3. Insert Combitube blindly into the oropharynx until the teeth lie between the two black bands on the proximal Combitube (Fig. 12.2).
4. Inflate the proximal blue cuff until air pressure is produced or the manufacturer-recommended pressure is reached.
   (a) 85 mL for 37 French Combitube
   (b) 100 mL for 41 French Combitube
5. Identify placement and attach to oxygen.
   (a) Ventilate through tube #1 (blue).
   (b) Auscultate the stomach and lungs.
      (i) If breath sounds are heard, the Combitube is in its more common esophageal location.
      (ii) Attach tube #1 to bag valve mask and O_2.
   (c) ONLY IF gurgling is present over the stomach when tube #1 is ventilated:
      (i) Ventilate through tube #2.
      (ii) If breath sounds are heard, the Combitube is in the less common tracheal location.
      (iii) Attach tube #2 to bag valve mask and O_2.
6. If no breath sounds are heard in either location:
   (a) Consider obstruction—Combitube may be obstructing the glottis or collapsing the trachea owing to deep proximal cuff inflation.
      (i) Deflate the cuffs.
      (ii) Withdraw 3 cm.
      (iii) Reinflate and start from step 4.
   (b) Consider equipment failure.
      (i) Check that balloons are maintaining pressure and intact.
   (c) Consider reinsertion.
7. Confirm placement with capnogram and pulse oximetry.
8. Secure the Combitube in position (Fig. 12.3).

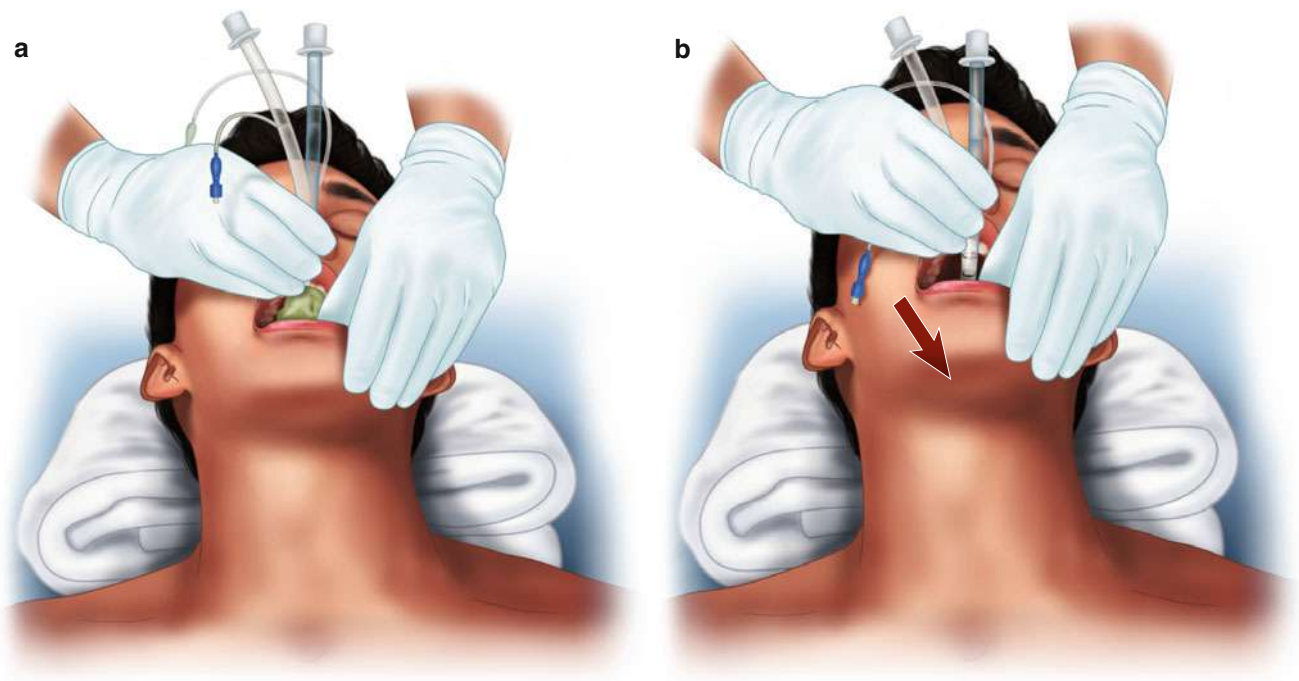

Fig. 12.2 (**a**) Insertion of Combitube. (**b**) Teeth should lie between the two black bands on the proximal Combitube

Fig. 12.3 Secure the Combitube in position

12.5 Pearls and Pitfalls

- Pearls
  - In an esophageal intubation situation, a suction tube may be threaded down using tube #2 to decompress the stomach.
  - The Combitube comes with an L-shaped piece that may also be attached to the end of tube #2 to deflect gastric contents away from practitioners.
- Pitfalls
  - After Combitube placement, a definitive airway should be placed when possible.
    - Gastric contents may aspirate despite placement of the Combitube.
    - Combitube should be considered a bridging airway device.
  - Combitubes are associated with a more pronounced hemodynamic stress response than ETTs or laryngeal mask airways (LMAs).
  - Balloon overinflation can lead to esophageal rupture (albeit rare).
  - Combitubes are associated with an increased incidence of sore throat, dysphagia, and upper airway hematomas than ETTs and LMAs.
  - Piriform sinus perforation.

Selected Reading

Agro F, Frass M, Beunmof JL, Krafft P. Current status of the Combitube: a review of the literature. J Clin Anesth. 2002;14:307–14.

Laurin E, Bair A. Devices for difficult airway management in adults. www.uptodate.com. Accessed 14 Mar 2014.

Liem EB. Combitube insertion. University of Florida Department of Anesthesiology, Center for Simulation, Advanced Learning and Technology, Virtual Anesthesia Machine Web site; 2006. http://vam.anest.ufl.edu/airwaydevice/combitube/index.html. Accessed 14 Mar 2014.

Walls R, Murphy M. Manual of emergency airway management. Philadelphia: Lippincott Williams & Wilkins; 2008.

13 Assessment of the Difficult Airway

Melinda W. Fernandez and Lars K. Beattie

13.1 Indications

- Should be performed on all patients that require airway management, conditions permitting
- Respiratory distress
- Airway protection

13.2 Materials and Medications

- None required

13.3 Procedure

1. Anticipating a difficult airway in emergency department patients is the first step to avoiding an unexpected surgical airway.
2. Two mnemonics can be applied quickly and easily to aid in an airway assessment: MOANS and LEMON.

13.3.1 Predicting Bag-Valve-Mask Difficulty

1. Use the mnemonic *MOANS* to assess for possible bag-valve-mask (BVM) difficulty.
   M—mask seal. Will you be able to get a good seal on the face? Predictors of difficulty include facial hair such as a beard, elderly, or emaciated owing to loss of muscle tone in the face.
   O—obesity. Body mass index (BMI) >30.
   A—age (>55 years). Loss of facial muscle tone can make bagging difficult.
   N—no teeth. Although being edentulous makes for an easier intubation, it makes bagging more difficult.
   S—stiff lungs. Acute or chronic lung disease can make a person difficult to bag. In the setting of Trauma, pulmonary contusion(s) and/or other direct lung injuries may increase BVM difficulty.

13.3.2 Predicting Difficult Laryngoscopy

1. Attempts should be made, if at all possible, to assess for a potentially difficult airway. This does not mean you cannot perform direct laryngoscopy if you are anticipating a difficult airway. It does, however, force you to consider all options and to have a solid backup plan in place with backup equipment readily available in the room.
2. Use the mnemonic *LEMON* to predict difficult direct laryngoscopy.
   L—look. A quick look at the patient will tell you a lot. Are there facial injuries; facial anomalies; obesity; short, thick neck; and small mouth or mandible?
   E—evaluate. Use the 3-3-2 rule to quickly assess for the strongest predictors of difficult laryngoscopy.
   - *3:* Open the patient's mouth and three vertically aligned fingers should fit between the incisors.
   - *3:* Three finger widths should fit along the length of the mandible from the mentum to the hyoid bone. Shorter or longer distances may make for a difficult intubation.
   - *2:* Thyromental distance should ideally be two fingers. Measure this from the hyoid to the thyroid.

M.W. Fernandez, MD
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: mindyfernandez@ufl.edu

L.K. Beattie, MD, MS (✉)
Department of Emergency Medicine, University of Florida, Gainesville, FL, USA
e-mail: lars.beattie@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_13

M—Mallampati classification (Fig. 13.1). If patient's condition and situation allow, have the patient open the mouth wide, stick out the tongue, and say "Ahh." Evaluate for visible structures.

- *Class I:* Tonsillar pillars and the entire uvula are visible.
- *Class II:* More than the base of the uvula is visible but no pillars are visible.
- *Class III:* Only the base of the uvula is visible.
- *Class IV:* No uvula or soft palate is visible. Only the hard palate is visible.
- These classifications correlate with the Cormack-Lehane grading system for laryngoscopic views. A Mallampati class I will correlate with a grade 1 view about 99 % of the time, whereas a Mallampati IV will be a grade 3 or 4 view all of the time and a rescue plan with backup equipment immediately available should always be in place [1, 2].

O—obstruction. Observe for anything that can get in the way (e.g., the tongue, dentures, blood, vomit, foreign body, edema, redundant tissue).

N—neck mobility. If patient's condition and situation allow, have the patient flex and extend the neck to evaluate mobility. Many patients in the emergency department have limited neck mobility. Examples include the trauma patient who arrives in cervical collar immobilization or a patient with degenerative or rheumatoid arthritis.

3. The *"6-D"* method is another assessment tool that can be used to predict difficult laryngoscopy and intubation. This method can be remembered by the fact that the word "difficult" begins with the letter "D":

Disproportion

- Increased tongue size in relation to pharyngeal size
- Airway swelling or trauma

Distortion

- Neck mass, hematoma, abscess, previous surgical airway, arthritic neck changes

Decreased thyromental distance

- Anterior larynx and decreased mandibular space.
- Look for a receding chin or greater than three fingerbreadths from the mentum to the hyoid bone.

Decreased inter-incisor gap

- Reduced mouth opening.
- Look for less than two to three fingerbreadths placed vertically in the patient's open mouth.

Decreased range of motion in any joints of the airway

- Limited head extension
- Previous neck radiation and/or surgery
- Neck contractures

Dental overbite

- Oversized, angled teeth disrupt the alignment of airway axes.
- Can decrease the interincisor gap.

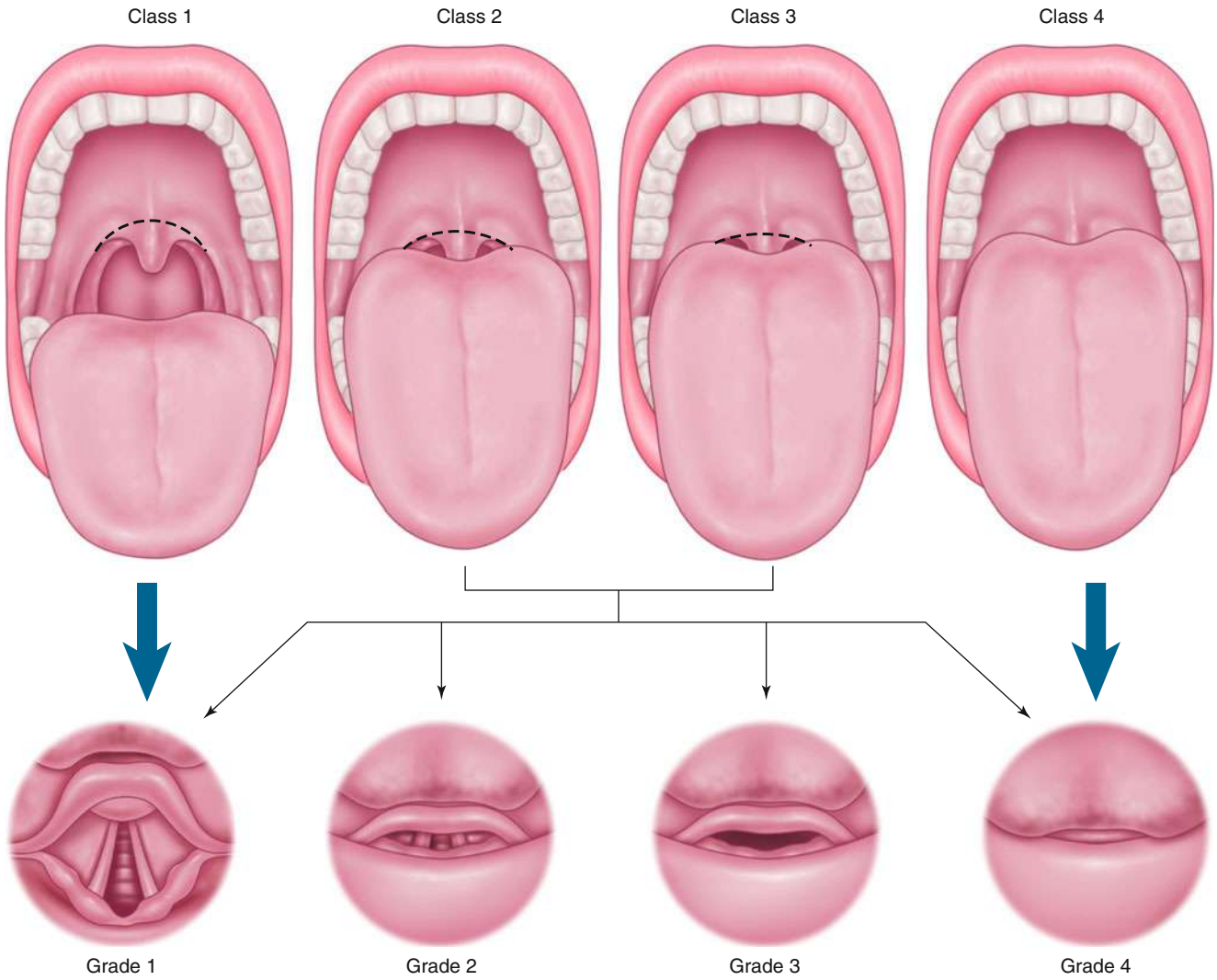

Fig. 13.1 Mallampati classification

13.3.3 Optimizing Laryngoscopy in the Obese Patient

- It is essential that emergency medicine physicians are able to successfully intubate the obese patient.
- Proper assessment and positioning will increase the success rate:
  - The goal is to ensure alignment of the oropharyngeal-pharyngeal-laryngeal (OA-PA-LA) airways by placing the patient in the head-elevated laryngoscopy position (Fig. 13.2a).
  - Align the external auditory meatus with the sternal notch along a horizontal line by positioning the patient on a "ramp."
  - The ramp can be created by stacking blankets/towels under the lower back ramping up to the neck and head (Fig. 13.2b).
- While the patient is in position on the ramp, the support is adjusted to minimize head flexion and allow for positioning in the sniffing position.
  - Because of the increased height, a step stool may be required to adequately visualize the airway from the head of the bed.

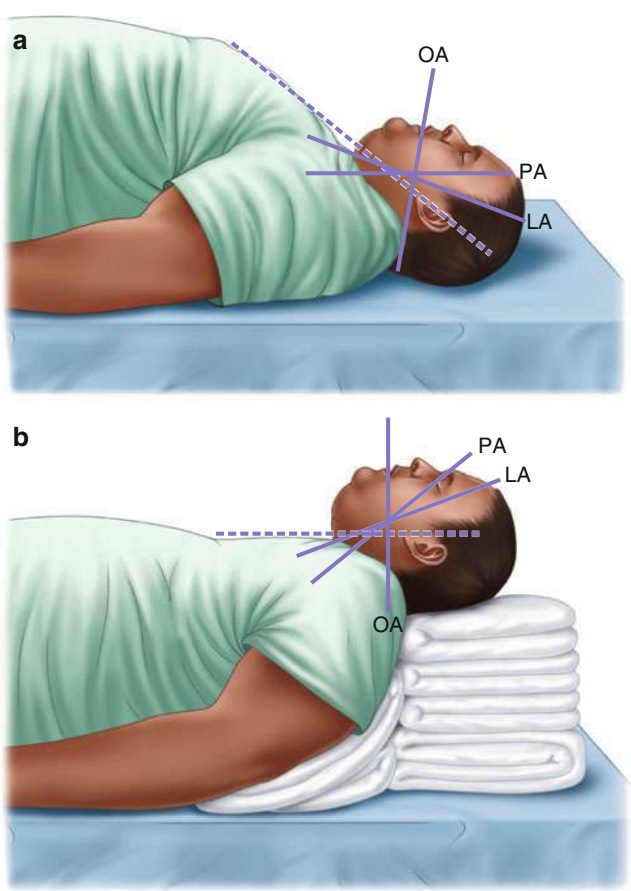

Fig. 13.2 (**a**, **b**) Ramping the obese patient will align the axes of the airway and allow easier direct laryngoscopy when viewed from the head of bed. (*LA* laryngeal airway, *OA* oropharyngeal airway, *PA* pharyngeal airway)

13.4 Pearls and Pitfalls

- Owing to time-sensitive patient care situations, emergency physicians are often not able to perform a thorough airway evaluation on every patient.
- With every airway that you manage and before pushing any drugs, always ask yourself:
  - Will I be able to ventilate this patient?
  - Will I be able to intubate this patient?
  - What is my difficult airway plan if I encounter trouble?
  - Will I be able to perform a surgical airway, if necessary?
- Be sure you have a solid backup plan A, B, and C before pushing any drugs.

References

1. Lee A, Fan LT, Gin T, Karmakar MK, Ngan Kee WD. A systematic review (meta-analysis) of the accuracy of the Mallampati tests to predict the difficult airway. Anesth Analg. 2006;102:1867–78.
2. Boschert S. Think L-E-M-O-N when assessing a difficult airway. ACEP News. Nov 2007.

Selected Reading

Murphy M. Bringing the larynx into view: a piece of the puzzle. Ann Emerg Med. 2003;41:338–41.
Rick J. Recognition and management of the difficult airway with special emphasis on the intubating LMA-Fastrach/whistle technique: a brief review with case reports. BUMC. 2005;18:220–7.
Roberts J, Hedges J. Clinical procedures in emergency medicine. 5th ed. Philadelphia: WB Saunders; 2009. p. 60–2.
Wilson W. Difficult intubation. In: Atlee J, editor. Complications in anesthesia. Philadelphia: WB Saunders; 1999. p. 138–47.

14 GlideScope

Sohan Parekh

14.1 Indications

- Initial device in a predictably difficult airway
- Rescue device in a failed intubation

14.2 Contraindications

- Absolute
  - Inadequate mouth opening
- Relative
  - Blood, vomit, or other secretions that can coat and obstruct the camera lens

14.3 Materials and Medications

- GlideScope video monitor with video cable (GlideScope Video Laryngoscope [GVL] system) (Fig. 14.1) or appropriate-size video baton (Cobalt System) (Fig. 14.2)
- Appropriate-size reusable video laryngoscope (GVL) or single-use laryngoscope blade (GVL Stat) (Table 14.1).
- Endotracheal tube (ETT)
- Malleable stylet or GlideRite rigid stylet
- 10 mL syringe
- End-tidal CO_2 ($EtCO_2$) capnography or colorimetry

S. Parekh, MD
Department of Emergency Medicine,
University of Texas at Austin Dell Medical School,
Austin, TX, USA

Department of Emergency Medicine,
University Medical Center Brackenridge,
Austin, TX, USA
e-mail: sohan.parekh@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_14

Fig. 14.1 GlideScope GVL system (With kind permission from Springer Science + Business Media: Noppens RR, Werner C, Piepho T. Indirekte Laryngoskopie. *Der Anaesthesist*. 2010;59(2):149–61)

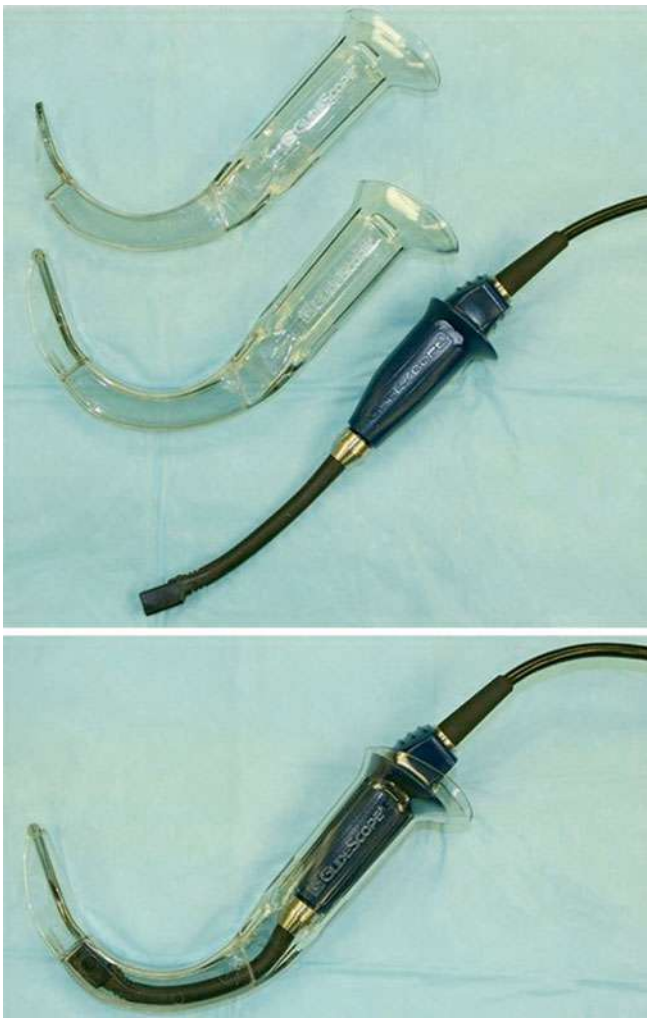

Fig. 14.2 GlideScope Cobalt system (With kind permission from Springer Science + Business Media: Jones PM, Turkstra TP, Armstrong KP, et al. Comparison of a single-use GlideScope® Cobalt videolaryngoscope with a conventional GlideScope® for orotracheal intubation. *Can J Anesthe/Journal canadien d'anesthésie*. 2010;57(1))

Table 14.1 GlideScope sizing

				Video Baton 1–2			Video Baton 3–4	
GVL 1	GVL 2	GVL 3	GVL 4	Stat 0	Stat 1	Stat 2	Stat 3	Stat 4
1.8–10 kg	10 kg—adult	40 kg—morbidly obese	40 kg—morbidly obese	<1.5 kg	1.5–3.6 kg	1.8–10 kg	10 kg—adult	40 kg—morbidly obese

14.4 Procedure

1. Insert the video cable (GVL system) or the video baton (Cobalt system) into the GlideScope video monitor (Fig. 14.3).
2. If using the GVL system, insert the distal end of the video cable into the port on the handle of the reusable video laryngoscope (GVL) (Fig. 14.4).
3. If using the Cobalt system, insert the video baton into the GVL Stat (Fig. 14.5).
   (a) Align the logo on the side of the video baton with the logo on the side of the single-use laryngoscope blade (GVL stat).
   (b) The video baton should slide smoothly and click into the place.
4. Turn on the GlideScope at for at least 30–120 s before use to fully activate the antifog mechanism.
5. Insert a stylet into the ETT. If using a malleable stylet, shape the curvature of the distal end of the tube to conform to the 60° curvature of the laryngoscope blade.
6. Firmly hold the laryngoscope handle in the left hand and ensure that an image can be clearly seen on the video monitor.
7. After ensuring adequate sedation and paralysis, open the mouth wide and insert the laryngoscope blade in the midline underneath the tongue (Fig. 14.6).

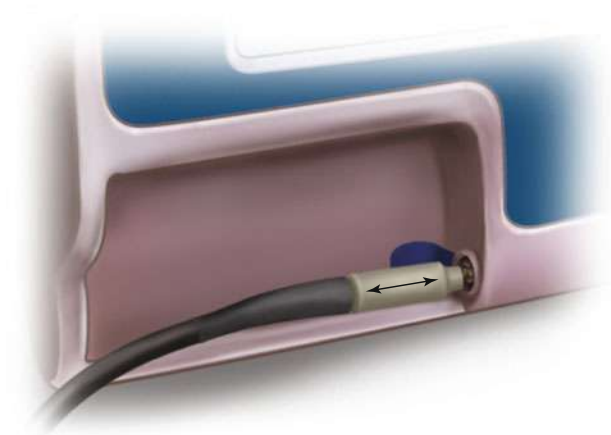

Fig. 14.3 Cable insertion into the video monitor

Fig. 14.4 Connect the distal end of the video cable the port on the handle of the GVL (GVL System)

Fig. 14.5 Slide the video baton into the GVL Stat, Cobalt System

Fig. 14.6 Insert the laryngoscope blade in the midline beneath the tongue (with kind permission from Springer Science + Business Media: Osborn IP, Kleinberger AJ, Gurudutt VV. Chapter 8: Airway emergencies and the difficult airway. In: Levine AI, Govindaraj S, DeMaria S, editors. *Anesthesiology and otolaryngology*. 2013)

8. Looking at the video monitor, advance the laryngoscope blade further into the oropharynx in order to obtain a view of the epiglottis.
   (a) Do not look directly into the oropharynx.
   (b) Movements and adjustments should be guided by the image on the video monitor.
9. Place the laryngoscope blade in the vallecula (analogous to a Macintosh blade) and apply a gentle backward tilt to expose the glottis.
10. In the event that a satisfactory glottic view cannot be obtained, the laryngoscope blade may be advanced and used like a Miller blade to lift the epiglottis out of the way.
11. Directing attention back toward the patient, insert the ETT into the mouth adjacent to the laryngoscope blade.
12. Guide the ETT toward the tip of the laryngoscope such that the end of the ETT emerges on the video monitor.
13. Looking at the video monitor, advance the ETT toward the glottis, and maneuver the tip of the tube between the vocal cords by rotating and altering the angle of the ETT.
    (a) If the ETT tip is posterior to the arytenoids:
        (i) Pull the ETT superiorly, rotate it over the left arytenoid, and gently twist the tube over the epiglottic aperture.
        (ii) Apply external laryngeal manipulation.
        (iii) Withdraw the blade to reduce tilting of the laryngeal axis and lessen the angle of introduction.
    (b) If the ETT abuts the false vocal cords, turn the ETT in the clockwise direction while withdrawing the stylet (Fig. 14.7).
14. Using the thumb, partially withdraw the stylet a few centimeters from the ETT.
    (a) The distal end of the tube should be free of the stylet.
    (b) An assistant can perform this task to allow for greater control and stability of the ETT.
15. Insert the ETT to the desired depth.
16. Fully remove the stylet, inflate the cuff of the ETT using a syringe, and confirm placement with $EtCO_2$, capnography, or colorimetry.

Fig. 14.7 Partially withdraw the stylet from the ETT to allow for passage through the vocal cords

14.5 Complications

- Dental injury
- Airway trauma

14.6 Pearls and Pitfalls

- Unlike conventional laryngoscopy, there is no need to displace the tongue.
- The greatest challenge when intubating with the GlideScope is maneuvering the ETT into the glottis aperture. Alternatives to the conventional technique are:
  - Make a 90° bend in the ETT just proximal to the cuff and insert it in the horizontal direction with the tip toward the right cheek. Once the tube is advanced past the flange of the laryngoscope, rotate it counterclockwise, at which point it should be pointed at the glottis. The tube can then be gently rotated into the glottis.
  - Consider inserting the laryngoscope slightly to the left of the midline upon initial insertion to allow greater space for advancement and maneuvering of the ETT.
- Do not overly lever the laryngoscope or use excessive lifting force after insertion into the vallecula. An adequate view of the glottis is generally easily obtained with minimal effort.
- Display settings can be adjusted using the menu button on the video monitor.

Selected Reading

Cho JE, Kil HK. A maneuver to facilitate endotracheal intubation using the GlideScope. Can J Anaesth. 2008;55:56–7.

GlideScope GVL and Cobalt user's manual & quick reference guide. Bothell: Verathon Inc; 2009–2011.

Kramer DC, Osborn IP. More maneuvers to facilitate tracheal intubation with the GlideScope. Can J Anaesth. 2006;53:737.

Lim HC. Utilization of a GlideScope videolaryngoscope for orotracheal intubations in different emergency airway management settings. Eur J Emerg Med. 2009;16:68–73.

Walls RW, Murphy MF, editors. Manual of emergency airway management. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

15 Endotracheal Tube Introducer (Bougie)

Joseph Rabinovich

15.1 Indications

- During orotracheal intubation, when only epiglottic visualization or partial glottic view is obtained during laryngoscopy.
- Particularly useful when neck mobility is limited, leading to inadequate visualization of the glottis (as in the case with cervical spine immobilization).
- When the glottic opening is narrowed either from pathological causes (burns, trauma, tumor, or other anatomical variation).
- When the direct view of the airway is very narrow, as with limited mouth opening or large tongue. In these scenarios, the endotracheal tube (ETT) can obstruct one's view of the cords during placement.

15.2 Contraindications

- When a failed airway occurs (three unsuccessful attempts at endotracheal intubation and inability to adequately oxygenate)
- When surgical airway is indicated (i.e., upper airway obstruction that prevents passage of the ETT via the orotracheal route)

15.3 Materials and Medications

- ETT introducer (bougie) (Fig. 15.1)
- Water-based lubricant
- Lubricated ETT 6 mm or larger *without* stylet (pediatric bougies are available that accommodate smaller ETTs)
- Standard orotracheal direct laryngoscopy (Miller or Macintosh blade) or video laryngoscopy setup
- Assistant

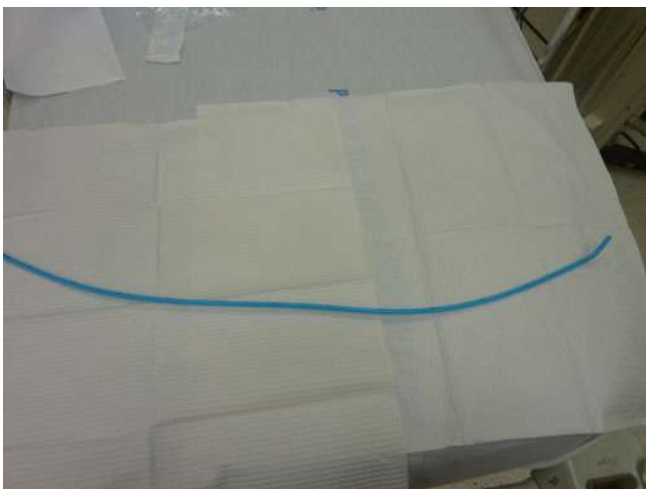

Fig. 15.1 Endotracheal tube introducer (bougie)

J. Rabinovich, MD
Department of Emergency Medicine,
Mount Sinai School of Medicine,
Elmhurst Hospital Center, Elmhurst, NY, USA
e-mail: joerab@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_15

15.4 Procedure

1. The operator first optimizes airway visualization (Fig. 15.2). (Bougie use should not be a substitute for poor technique.)
2. Without losing sight of the airway, the operator asks the assistant to hand him or her the bougie with the coudé tip facing up.
3. The operator directs the bougie tip underneath the epiglottis (Fig. 15.3).
4. Confirmation of placement can be done visually or by tactile sensation:
   (a) A ratchet-like sensation may be felt as the bougie tip is advanced into the airway and slides over the tracheal rings.
   (b) As the bougie is further advanced, the operator may feel the bougie rotate as it enters the bronchus and/or will get a "hold up," the most reliable sign that the bougie is in the trachea [1]. (The "hold-up" sign occurs when the Bougie encounters a terminal bronchus [typically at around 35 cm] and stops advancing.) [2]
5. Once placement is confirmed, the bougie needs to be partially withdrawn to about 25 cm at the lip line.
   (a) Some brands will have a thick black indicator line.
   (b) A sufficient amount of the bougie needs to extend out beyond the proximal end of the ETT.
6. While the operator holds the bougie in place, the assistant threads the ETT over the bougie (Fig. 15.4).
7. The operator now grasps the ETT in her or his right hand and advances it over the bougie.
8. Simultaneously, the assistant holds and stabilizes the proximal end of the bougie.
9. The ETT should be advanced to approximately 23 cm in males and 21 cm in females. The assistant removes the bougie as the operator holds the ETT in place (Fig. 15.5).
10. As the operator holds the ETT firmly in position, the assistant inflates the ETT balloon and withdraws the bougie.
11. Confirmation of proper ETT placement is achieved through traditional means (end-tidal CO_2 detection, auscultation of breath sounds).

Fig. 15.2 (**a**) Assistant hands operator bougie with coudé tip directed upward, while operator maintains his focus on the target. (**b**) Bougie being placed, parallel with line of sight, underneath epiglottis

Fig. 15.3 Bougie can be placed into the glottis using direct laryngoscopy or video assistance

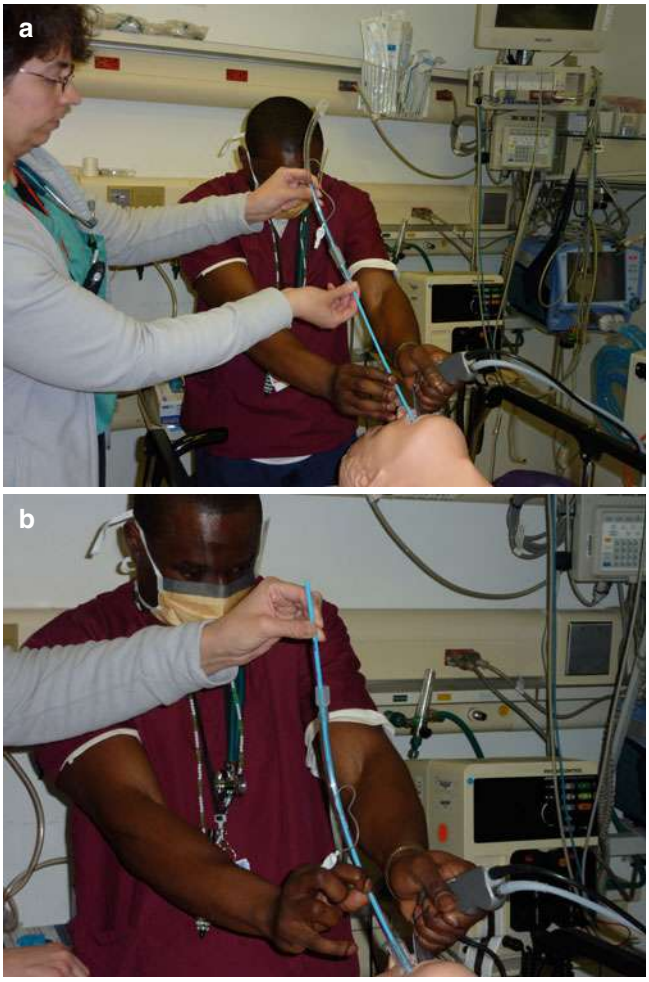

Fig. 15.4 (**a**) Assistant places ETT over bougie and operator withdraws bougie until it protrudes out the top of the ETT, (**b**) while assistant stabilizes the protruding portion of the bougie, the operator railroads the ETT into the airway. The operator continues to support the soft tissues with the laryngoscope blade to facilitate placement

Fig. 15.5 The assistant removes the bougie while the operator stabilizes the endotracheal tube

15.5 Complications

- Trauma to the esophagus, larynx, trachea, or bronchus, including perforation [3, 4].
- In general, complications are rare.

15.6 Pearls and Pitfalls

- The line of sight should be as parallel as possible to the axis of the bougie as it is being passed, allowing better eye-hand coordination. This augments more accurate placement of the bougie tip.
- Maintain the view of the airway as the assistant hands the operator the bougie.
- Keep the laryngoscope in place to support the soft tissues, as the endotracheal tube is slid over the bougie, to facilitate placement.
- If resistance is met during passage, withdraw the ETT slightly (~2 cm) and rotate the ETT counterclockwise one-quarter turn (90°) and reattempt passage.
  - This changes the position of the leading edge of the ETT, which may catch on the posterior laryngeal inlet [2].
  - By rotating the ETT, the leading edge now is anterior facing and is less likely to catch on the arytenoid cartilage along with other laryngeal inlet structures.
- If encountering resistance to ETT placement, consider releasing cricoid pressure (if used).
- Measurement markings on the bougie are aligned with the coudé tip. If not sure of orientation and the loose site of the tip, use the markings to properly orient the tip.

References

1. Kidd JF, Dyson A, Latto IP. Successful difficult intubation. Use of the gum elastic bougie. Anaesthesia. 1988;43:437–8.
2. Murphy MF, Hung OR, Law JA. Tracheal intubation: tricks of the trade. Emerg Med Clin North Am. 2008;26:1001–14.
3. Kadry M, Popat M. Pharyngeal wall perforation – an unusual complication of blind intubation with a gum-elastic bougie. Anaesthesia. 1999;54:404–5.
4. Smith BL. Haemopneumothorax following bougie-assisted tracheal intubation. Anaesthesia. 1994;49:91.

16 Lighted Stylet Intubation

Benjamin M. Mahon and Lars K. Beattie

When lighted stylet intubation is done correctly, the procedure can be very safe, with very little difference in outcome from that of primary laryngoscopy. (*Note*: *several lighted stylet devices*, *such as the TrachlightTM and Light WandTM*, *are no longer being manufactured*, *but these devices are still in use*.)

16.1 Indications

- Difficult/impossible direct laryngoscopy [1, 2]
  - Congenital abnormalities of airway
  - High Mallampati grade [3]
  - Dental appliances
- Failed direct laryngoscopy

16.2 Contraindications

- Absolute
  - Morbid obesity
  - Airway foreign body
  - Expanding neck mass
- Relative
  - Abnormal airway anatomy
  - Airway lesions (e.g., abscess, mass, epiglottitis) that change oropharyngeal anatomy
  - Acute care where concomitant resuscitation requires a well-lit room
  - Lack of familiarity or experience with procedure
  - "Can't oxygenate, can't ventilate" situation

16.3 Materials and Medications

- Intravenous (IV) access, O_2, and monitor
- Ambu bag with supplemental oxygen
- Suction (Yankauer and tubing)
- Lighted stylet (LS)
- Endotracheal tube (ETT) 2.5-mm larger than LS with 10-cc syringe
- Surgilube
- Intubation medications (this procedure may be performed as an awake or a rapid-sequence intubation)

B.M. Mahon, MD
Ponciana Medical Center,
Kissimmee, FL, USA
e-mail: drfaust2k9@aol.com

L.K. Beattie, MD, MS (✉)
Department of Emergency Medicine, University of Florida,
Gainesville, FL, USA
e-mail: lars.beattie@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_16

16.4 Procedure

1. Preoxygenate.
2. Positioning.
   (a) Sniffing position, pinna at the level of the sternal notch (Fig. 16.1).
   (b) Skip sniffing position if cervical spine injury is suspected.
3. LS-ETT unit preparation.
   (a) Insert the wire stylet into the device.
   (b) Check the LS light.
   (c) Lubricate the LS with K-Y Jelly.
   (d) Position the LS just distal to the Murphy eye.
   (e) Curve the LS to user preference at the line labeled "Bend Here."
4. Administer intubation medications.
5. Have an assistant to apply cricoid pressure.
6. Grasp and elevate the patient's jaw near the corner of the mouth with the operator's thumb, index, and middle fingers, elevating the tongue and epiglottis along with it.
7. Using the free hand, insert the LS-ETT unit into the oropharynx and advance (Fig. 16.2).
8. Use the midline glow in the neck to guide insertion of the LS-ETT (Fig. 16.3).
9. Bright light *below* the thyroid prominence indicates correct placement of the ETT tip.
10. Dim or blurred light or light at the thyroid prominence suggests incorrect positioning (Fig. 16.4).
11. If the transilluminated light is dim, off center, or not seen, esophageal positioning must be considered.
    (a) Withdraw the LS-ETT unit approximately 2–5 cm.
    (b) Reposition the patient's head and neck.
    (c) Reattempt according to steps 5–8.
12. Placement of the ETT (Fig. 16.5).
    (a) Hold the LS-ETT unit steady with one hand.
    (b) Check the depth of the ETT and adjust accordingly.
    (c) Release the LS latch that holds the ETT to the LS.
    (d) While holding the ETT in position, gently slide the LS out from the ETT.
    (e) Inflate the ETT balloon.
13. Confirm ETT placement (continuous end-tidal CO_2 [$EtCO_2$], colorimetric capnometry).
14. Secure the ETT.

Fig. 16.1 Sniffing position, pinna at the level of the sternal notch

Fig. 16.2 Grasp and elevate the patient's jaw near the corner of the mouth with the operator's thumb, index, and middle fingers, elevating the tongue and epiglottis along with it. Using the free hand, insert the LS-ETT unit into the oropharynx and advance

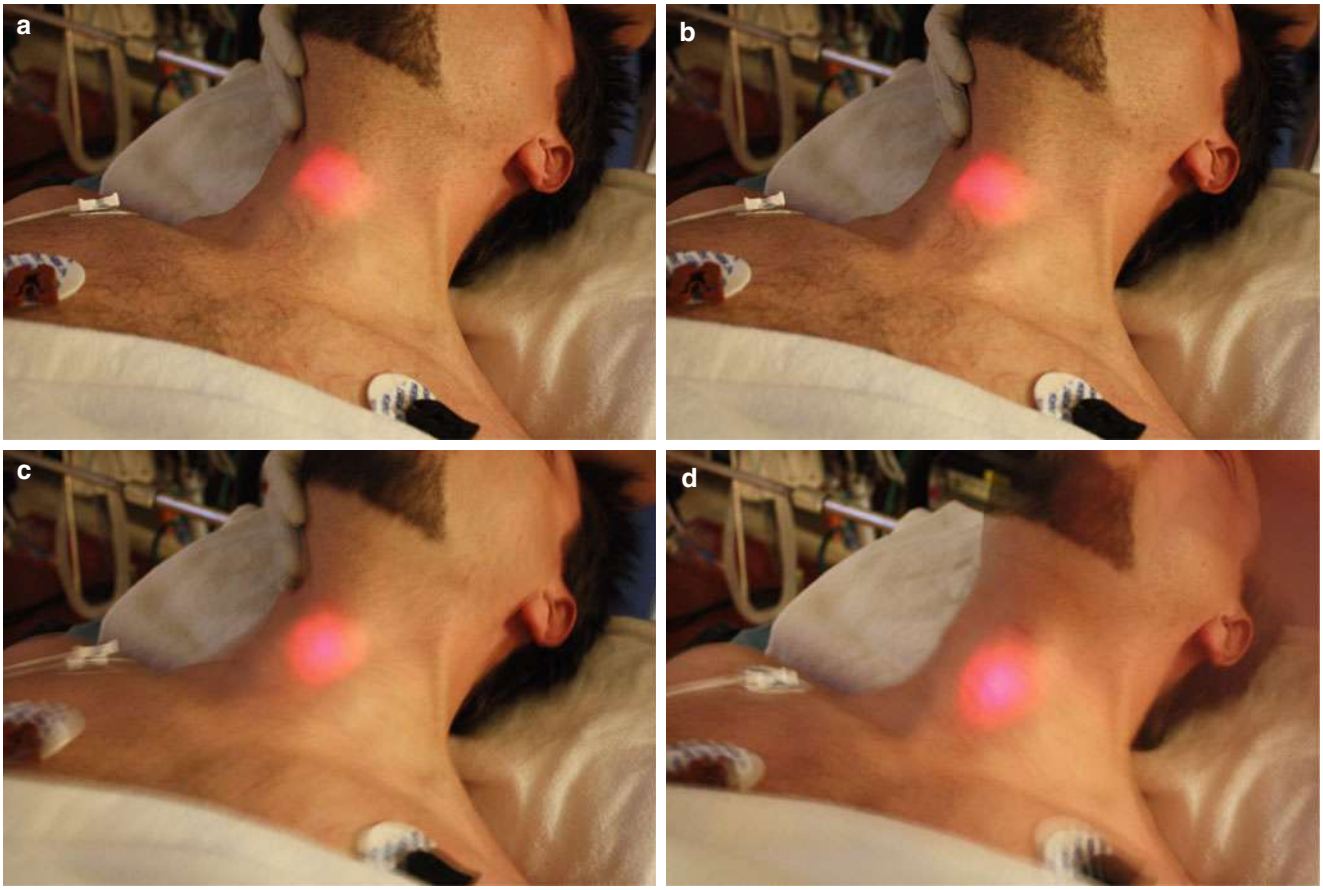

Fig. 16.3 (**a**–**c**) Use the midline glow in the neck to guide insertion of the LS-ETT. (**d**) Bright light *below* the thyroid prominence indicates correct placement of the ETT tip

Fig. 16.4 Dim or blurred light or light at the thyroid prominence suggests incorrect positioning

Fig. 16.5 (**a**) Hold the LS-ETT unit steady with one hand. (**b**) Check the depth of the ETT and adjust accordingly. (**c**) Release the LS latch that holds the ETT to the LS. (**d**) While holding the ETT in position, gently slide the LS out from the ETT

16.5 Pearls and Pitfalls

- Pearls
  - LS-ETT complex—Typically the classic "hockey-stick" shape with the 90° curve just proximal to the cuff is recommended [2].
  - Dimming the room lights will enhance transillumination.
  - Pulling the wire stylet out from the LS-ETT unit will make it more pliable and may facilitate its placement in the trachea and removal of the LS.
  - Some LS devices start to blink after 30 s to prevent bulb overheating.
  - The LS may be used with nasotracheal intubation, intubation through a laryngeal mask airway (LMA), or conventional laryngoscopy to enhance success.
- Pitfalls
  - LS intubation should not be used as an emergency airway alternative by a proceduralist unfamiliar with the technique:
    - It is technically complicated and more challenging than many other airway adjuncts in the standard difficult airway algorithm.
    - One study compared the use of four rescue airway devices in the difficult airway algorithm. A success rate of only 20 % was achieved with the Trachlight™ on the first attempt when in the hands of the novice physician when used as a rescue device in their difficult airway algorithm [4].
  - In very thin patients, transillumination may be visualized quite well even when the LS-ETT unit is in the esophagus:
    - When the unit is in the esophagus, typically the light it will be more diffuse.
    - When the unit is in the trachea, the transilluminated area will be well circumscribed.
  - In obese patients or patients with significant neck tissue, the transilluminated light from the LS-ETT unit may be dim despite correct positioning in the trachea.

References

1. Agro F, Hung OR, Cataldo R, Carassiti M, Gherardi S. Lightwand intubation using the Trachlight: a brief review of current knowledge. Can J Anaesth. 2001;48:592–9.
2. Davis L, Cook-Sather SD, Schreiner MS. Lighted stylet tracheal intubation: a review. Anesth Analg. 2000;90:745–56.
3. Rhee KY, Lee JR, Kim J, Park S, Kwon WK, Han S. A comparison of lighted stylet (Surch-Lite) and direct laryngoscopic intubation in patients with high Mallampati scores. Anesth Analg. 2009;108:1215–9.
4. Aikins NL, Ganesh R, Springmann KE, Lunn JJ, Solis-Keus J. Difficult airway management and the novice physician. J Emerg Trauma Shock. 2010;3:9–12.

Selected Reading

Langeron O, Birenbaum A, Amour J. Airway management in trauma. Minerva Anestesiol. 2009;75:307–11.

Walls RM, Murphy MF. Manual of emergency airway management. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. Chap. 11.

17 Fiber-Optic Stylet Intubation (Rigid and Semirigid)

Joseph Rabinovich

17.1 Indications

- For use in routine and predicted difficult oral intubations.
- Similar to a flexible fiber-optic scope with the specific advantages of:
  - Less setup time
  - Less time to perform the procedure
  - Appropriate for routine intubations (and easier to accumulate experience)
  - Rigid enough to lift up the epiglottis
  - Easier to navigate through tissue
  - Less susceptible to being obscured by blood and secretions
  - More durable, more portable, easier to clean, and less expensive
- Particularly useful when neck mobility or mouth opening is restricted.
- Advantageous in awake intubations because it can minimize tissue contact, resulting in less stimulation to the patient's airway and better tolerance.
- Certain stylets can be used to intubate through supraglottic airways such as laryngeal mask airways (LMAs).

17.2 Contraindications

- Complete upper airway obstruction where surgical airway is indicated
- Oral pharyngeal swelling requiring a nasotracheal or surgical approach
- Failed airway and unable to adequately maintain oxygenation

17.3 Relative Contraindications

- Large amounts of blood and secretions may obscure visualization of the airway and cords.
- Very distorted airways. Compared with flexible endoscopy, this device is less maneuverable.

17.4 Materials and Medications

- Endotracheal tube (ETT) 5.5 mm or greater (Fig. 17.1) (Pediatric stylets are also available.)
- Water-soluble lubricant
- Defogging agent
- Optional: Swivel adaptor and meconium aspirator

J. Rabinovich, MD
Department of Emergency Medicine,
Mount Sinai School of Medicine, Elmhurst Hospital Center,
Elmhurst, NY, USA
e-mail: joerab@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_17

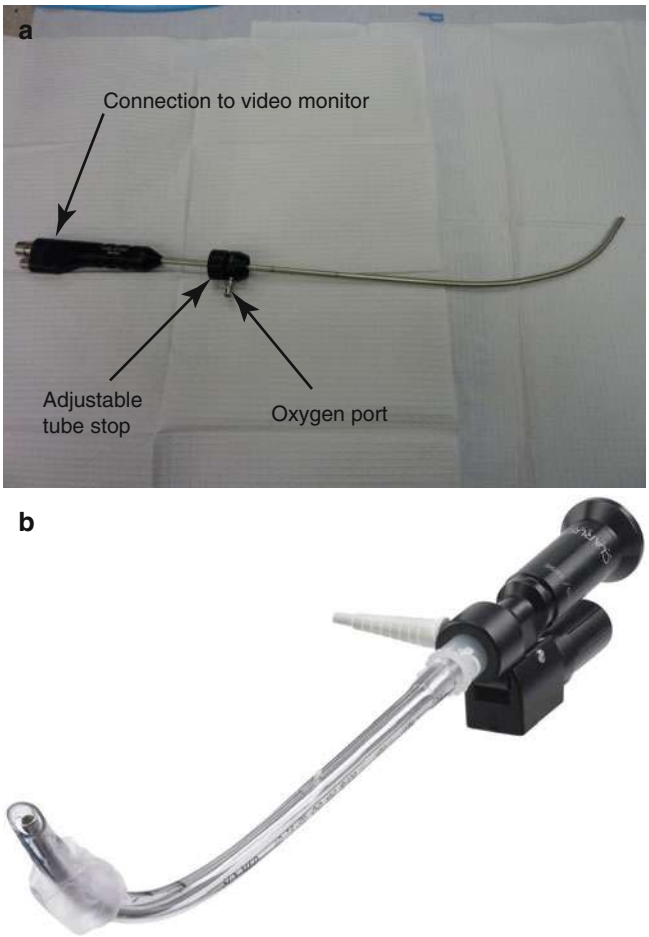

Fig. 17.1 (**a**) Bonfils rigid fiber-optic stylet, Karl Storz endoscopy, (**b**) Levitan FPS malleable fiber-optic stylet

17.5 Procedure

1. Use standard preparation for rapid sequence intubation (RSI) or for awake intubation.
2. Place the ETT over the stylet.
   (a) The ETT should extend slightly beyond the stylet tip.
   (b) If using the malleable stylet:
      (i) Without a laryngoscope: Bend tip to about 70° [1].
      (ii) With a laryngoscope: Bend tip to 35° [1].
   (c) Lubricate the tip of the ETT.
3. Depending on specific scope capability, connect oxygen tubing to the port on the scope.
   (a) This keeps secretions away from the tip while providing an oxygen source.
   (b) Keep flow less than 6 L/min [2].
4. Lens fogging prevention.
   (a) Warm the tip of the scope with the hand or immerse the tip in warm saline.
   (b) Apply the defogging agent.
   (c) Alternatively, chlorhexidine is an effective defogger.
5. Scope insertion.
   (a) Use the nondominant hand to pull the jaw forward while holding the tongue.
      (i) In an awake patient, have the patient protrude the tongue and the operator grasps it with 4×4 gauze. (Alternatively, a Macintosh laryngoscope blade can be used.)
      (ii) The goal is to move the base of the tongue off the posterior pharyngeal wall.
   (b) Initially position the scope horizontally and to the right of the patient's mouth.
   (c) Once the tip is in the oropharynx, reposition the scope vertically (Fig. 17.2a).
   (d) The scope tip should be in the midline or in the retromolar position (per scope design).
   (e) Position the scope and the tip of the ETT in front of the uvula.
   (f) Refer to the eyepiece or video screen to see if there is a clear image of the uvula (Fig. 17.2b).
   (g) Advance the scope very slowly to maintain a view of landmarks, avoiding tissue contact (Fig. 17.3).
6. Once the epiglottis is visualized:
   (a) Continue to advance slowly.
   (b) To get underneath the epiglottis, the tip of the scope may need to be moved posteriorly (by tilting the operator's hand slightly forward) (Fig. 17.4).

7. Once underneath the epiglottis, tilt the scope back, to advance into the more anterior directed airway.
   (a) Make sure the glottic opening is well centered on the screen to facilitate placement (Fig. 17.5).
   (b) If resistance is felt, operator may need to rotate the scope clockwise, or tip scope slightly forward, to prevent the ETT from abutting the anterior aspect of the trachea.
   (c) The ETT may need to be advanced off the rigid stylet to allow further advancement.
8. To remove scope:
   (a) Twist the proximal end of the ETT clockwise.
   (b) Stabilize the tube with the nondominant hand.
   (c) Use the dominant hand to pull the scope forward, following the curvature of the stylet (Fig. 17.6).
   (d) An assistant may be of use during this step.

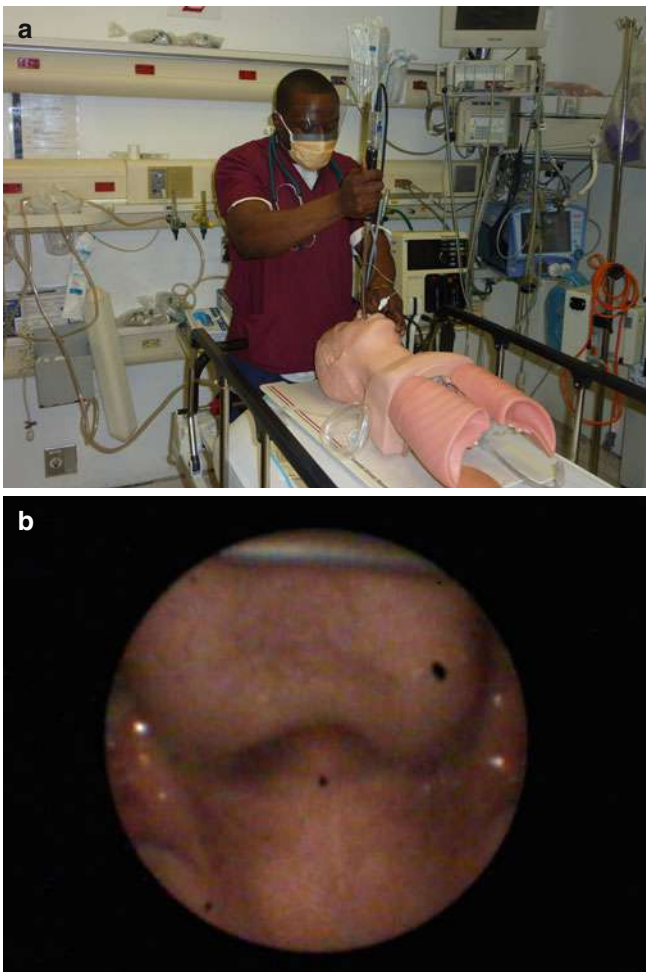

Fig. 17.2 (**a**) Initial placement of scope, under direct visualization; (**b**) tip is positioned in from the uvula

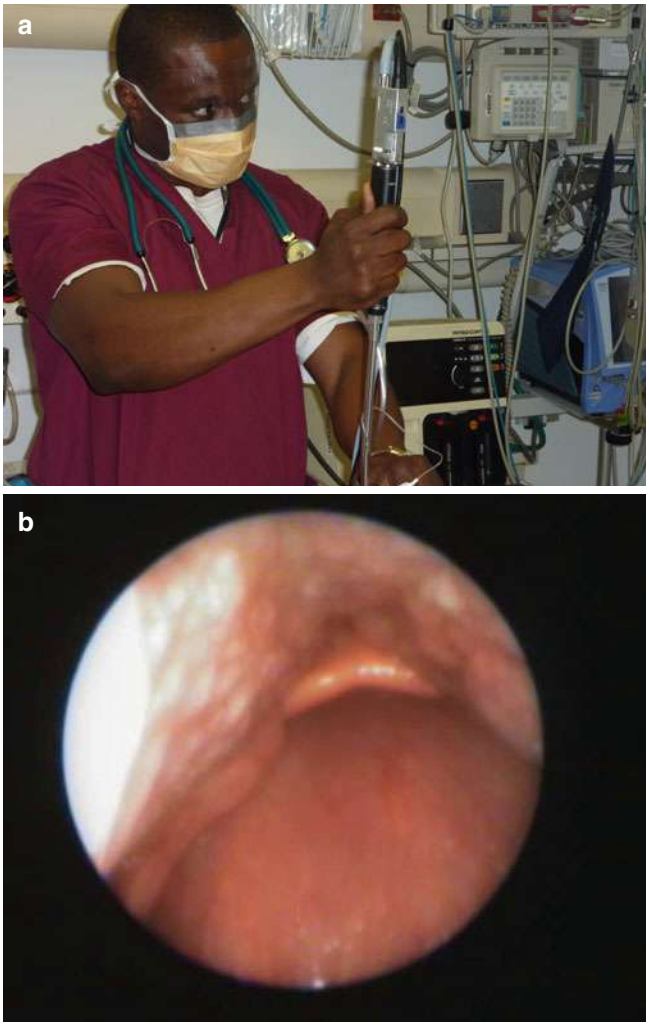

Fig. 17.3 (**a**) Using the video monitor, or through an eyepiece, the operator advances to the next landmark, (**b**) the epiglottis

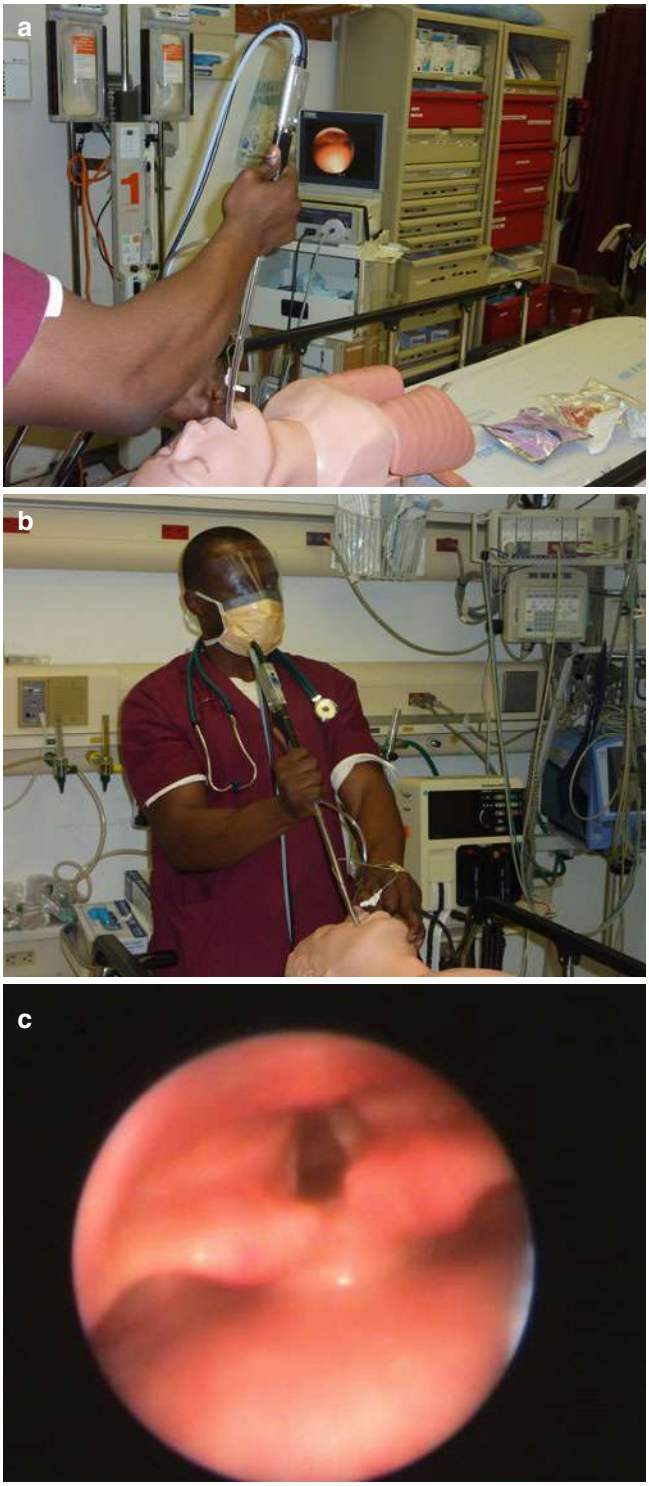

Fig. 17.4 (**a**) Operator tips scope forward to get underneath the epiglottis while advancing; (**b**) once under the epiglottis, the scope may need to be tipped back to advance to the glottic opening; (**c**) operator should try to keep the glottic opening in the center of the screen

Fig. 17.5 (**a**) Operator is advancing the scope, (**b**) through the glottic opening keeping the image centered

Fig. 17.6 To remove scope, operator must pull it forward while stabilizing the endotracheal tube

17.6 Pearls and Pitfalls

- Pearls
  - When the operator loses perspective or a clear view, withdraw the scope back to the point where identifiable structures are visualized and proceed.
  - The operator can suction through the scope by attaching a swivel adaptor and a meconium aspirator (Fig. 17.7) [3].
- Pitfalls
  - If the scope is advanced too quickly, orientation can be lost.
  - Structures that are too close to the scope will become blurred and unidentifiable.
  - If the scope tip abuts pharyngeal tissue, visualization can become blurred.
  - Flow greater than 6 L/min connected to the oxygen port may result in subcutaneous emphysema (single case report) [2].

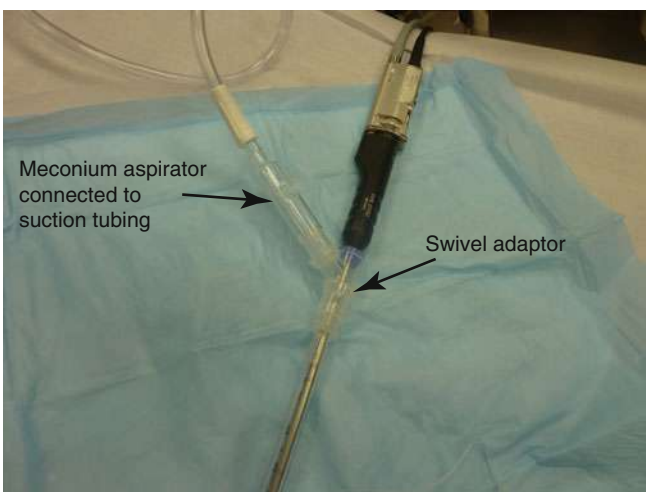

Fig. 17.7 By placing the stylet through the rubber valve of a swivel adaptor, which is then connected to suction via a neonatal meconium aspirator, the operator can now suction through the endotracheal tube

References

1. Levitan RM. Design rationale and intended use of a short optical stylet for routine fiberoptic augmentation of emergency laryngoscopy. Am J Emerg Med. 2006;24:490–5.
2. Hemmerling TM, Bracco D. Subcutaneous cervical and facial emphysema with the use of the Bonfils fiberscope and high-flow oxygen insufflation. Anesth Analg. 2008;106:260–2.
3. Weingart SD, Bhagwan SD. A novel set-up to allow suctioning during direct endotracheal and fiberscope intubation. J Clin Anesth. 2011;23:518–9.

18 Storz Video Laryngoscope

Joseph Rabinovich

18.1 Indications

- Orotracheal intubation for both routine and predicted difficult airways.
- Teaching traditional direct laryngoscopy to novice intubators.
- Ideal for unanticipated difficult airway with the option of intubating indirectly if an adequate direct view is unobtainable.
- An excellent tool when cervical spine precautions need to be taken: Because the video view of the airway is generated by a camera at the tip of laryngoscope blade, less manipulation is required for optimal glottic views.

18.2 Contraindications

- Absolute
  - When orotracheal intubation is contraindicated, e.g., for massive facial trauma, complete upper airway obstruction precluding orotracheal access to the airway
  - In a failed airway (three unsuccessful attempts with inability to maintain adequate oxygenation)
- Relative
  - Blood or copious secretions may prevent indirect viewing of the airway but does not always preclude the use of this device.

J. Rabinovich, MD
Department of Emergency Medicine,
Mount Sinai School of Medicine,
Elmhurst Hospital Center, Elmhurst, NY, USA
e-mail: joerab@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_18

18.3 Materials and Medications (Fig. 18.1)

- Standard materials and medications for endotracheal intubation. Operator should have a backup laryngoscope in case of equipment failure.
- Endotracheal tube (ETT) with or without stylet.
- Water-based lubricant.
- Antifogging agent (not required for C-Mac).

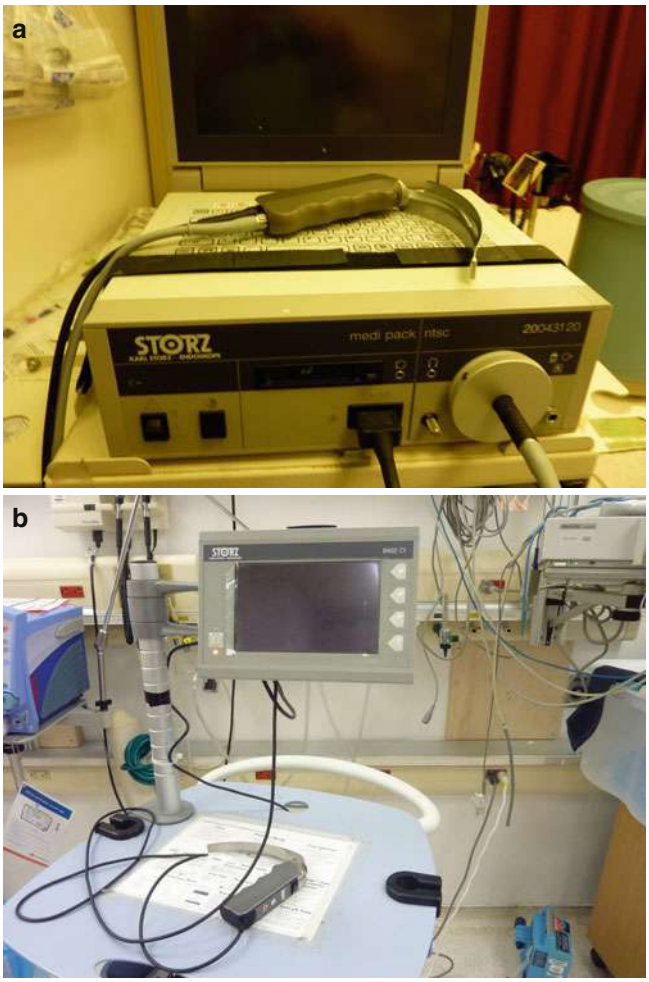

Fig. 18.1 (**a**) Storz video laryngoscope (older version), (**b**) Storz C-Mac (newer version)

18.4 Procedure

1. Standard preparation for orotracheal intubation. If there are no cervical spine precautions, then align the external auditory meatus with the sternal notch [1].
2. Apply antifogging drops to lens at tip of the blade, and/or hold the hand over the blade tip to warm it up to body temperature (older V-Mac model) (Fig. 18.2a).
3. Because blade geometry is the same as in standard laryngoscopes, the insertion technique is identical to that of standard laryngoscopy with a Macintosh blade (Fig. 18.2b).
4. Obtain the best direct view possible.
5. Consider the addition of the backward-upward-rightward pressure (BURP) maneuver [2].
6. Airway maneuvers may be performed by the operator or by the assistant using the video screen as a guide along with operator feedback (Fig. 18.3).
7. The operator has the option of intubating directly with adequate view or indirectly if visualization is improved.
8. Consider using an ETT introducer (bougie—see Chap. 15) if the view is inadequate.
9. Place the ETT, with or without a stylet, into the airway under direct or indirect visualization. If using a stylet, bend the distal end of the tube to approximately 35° as for a standard intubation (Fig. 18.4) [3].
10. Remove the stylet, inflate the cuff of the ETT using a syringe, and confirm placement with end-tidal CO_2, capnography, or colorimetry.

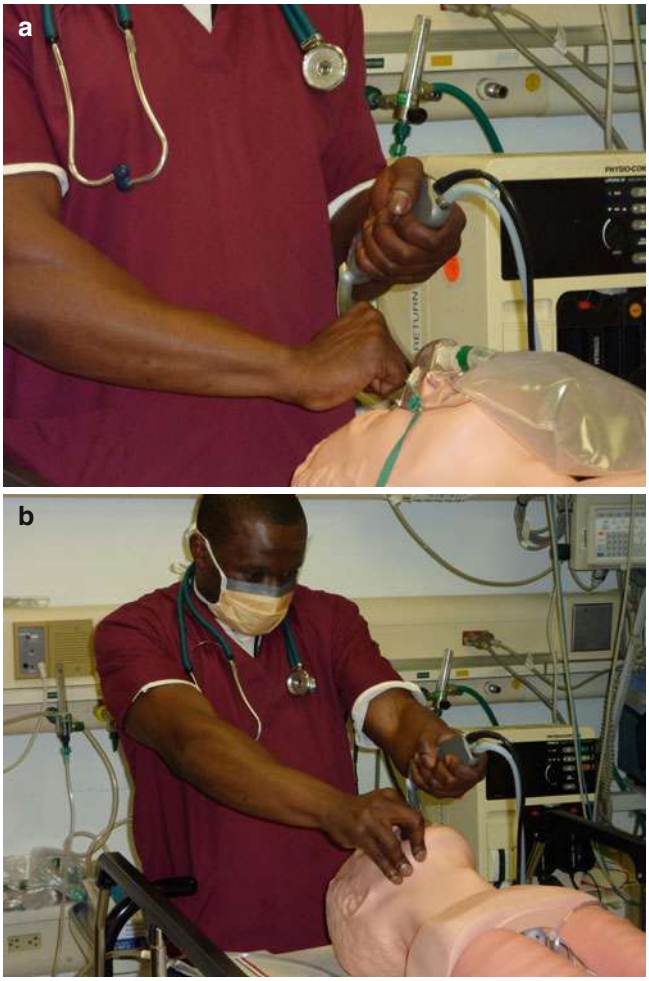

Fig. 18.2 (**a**) Operator warming the blade to prevent fogging, (**b**) laryngoscope blade insertion using the standard direct technique. Operator visually places the blade and optimizes the glottic view

Fig. 18.3 Operator has the choice of using direct visualization (preferred when a novice intubator is learning laryngoscopy) or the indirect video view of the glottis

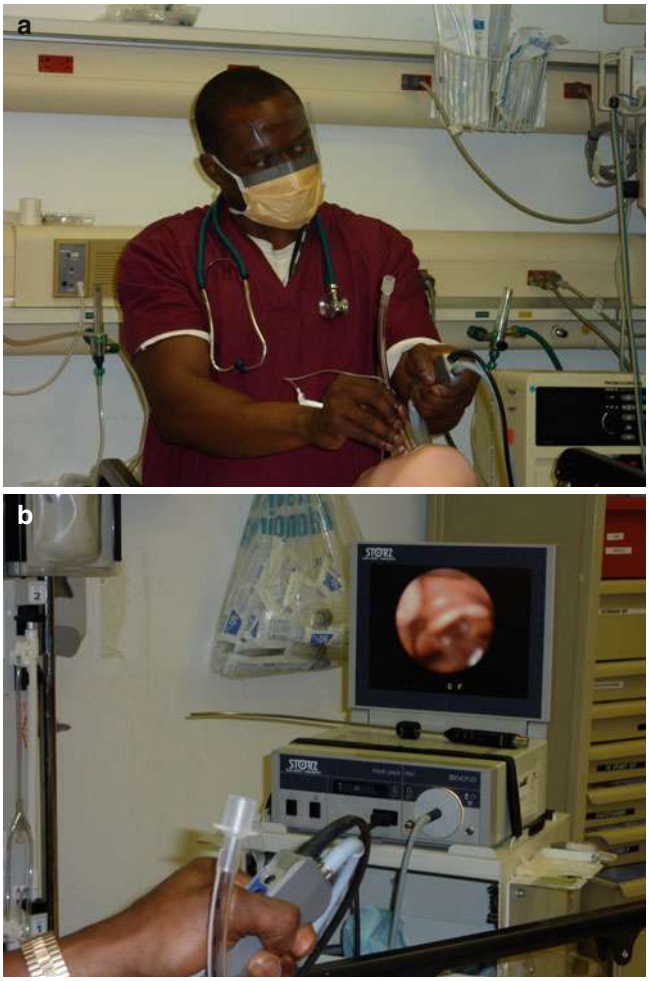

Fig. 18.4 (**a**) Operator initially places the endotracheal tube (ETT) into the oropharynx using direct visualization to avoid injury; (**b**) the tip of the ETT passing through the cords can be confirmed by direct visualization or by watching the video image

18.5 Complications (As with any Direct Laryngoscopy)

- Dental trauma
- Oropharyngeal trauma
- Vocal cord injury

18.6 Pearls and Pitfalls

- The initial placement of the laryngoscope blade and ETT should be done under direct visualization to avoid trauma to the oropharyngeal structures.
- As with direct laryngoscopy, the ETT should not be placed blindly, and the ETT must be seen to pass through the vocal cords to avoid placement in the esophagus.
- View can be obscured by secretions or fogging. If this occurs, the operator may need to remove the blade, wipe it down, and reinsert. The newer model, the C-Mac, is less likely to fog owing to design improvements.
- Observing the ETT pass through the vocal cords can sometimes be easier using the video image because the camera is angled to provide the most optimal view.
- Placement of the blade into the oropharynx can be awkward because the handle of laryngoscope is bulkier than a standard laryngoscope. Its handle is larger and has cables attached to its base. In patients with large anteroposterior diameter chests, the operator may need to rotate the laryngoscope handle toward the right corner of the mouth in order to introduce the blade into the oral cavity and then rotate it back to the proper position.
- These devices are ideal for teaching laryngoscopy. With same blade geometry, the technique is the same as with standard laryngoscopy. As the operator attempts intubation directly, the instructor can observe on the video screen and guide the student. The instructor will also be able to visually confirm that the ETT is entering the trachea.

References

1. Greenland KB, Edwards MJ, Hutton NJ, Challis VJ, Irwin MG, Sleigh JW. Changes in airway configuration with different head and neck positions using magnetic resonance imaging of normal airways: a new concept with possible clinical applications. Br J Anaesth. 2010;105:683–90.
2. Knill RL. Difficult laryngoscopy made easy with a “BURP.”. Can J Anaesth. 1993;40:798–9.
3. Levitan RM, Heitz JW, Sweeney M, Cooper RM. The complexities of tracheal intubation with direct laryngoscopy and alternative intubation devices. Ann Emerg Med. 2011;57:240–7.

Selected Reading

Brown 3rd CA, Bair AE, Pallin DJ, et al. Improved glottic exposure with the video Macintosh laryngoscope in adult emergency department tracheal intubations. Ann Emerg Med. 2010;56:83–8.

Niforopoulou P, Pantazopoulos I, Demestiha T, Koudouna E, Xanthos T. Video-laryngoscopes in the adult airway management: a topical review of the literature. Acta Anaesthesiol Scand. 2010;54:1050–61.

19 Cricothyroidotomy

Henry Young II, Shannon Toohey, Bharath Chakravarthy, and Lars K. Beattie

Up to seven intubation attempts in 1000 end up in a "can't intubate/can't ventilate" situation in the emergency department. These are considered failed airways that may require a surgical airway to maintain ventilation and oxygenation.

19.1 Indications

- Endotracheal tube (ETT) placement attempts unsuccessful
- Failed bag valve mask, laryngeal mask airway, or Combitube ventilation
- Severe facial trauma affecting the upper airway
- Severe oropharyngeal hemorrhage or profound emesis
- Obstruction (foreign body, mass, mass effect)

19.2 Contraindications

- Airway protection achievable using a less invasive strategy
- Tracheal transaction
- Pediatric patients younger than 8 years

19.3 Techniques

- Scalpel-bougie – minimalist
- Scalpel-Trousseau – standard

19.3.1 Scalpel-Bougie

19.3.1.1 Materials and Medications (Fig. 19.1)

- Betadine or chlorhexidine
- Scalpel #11 blade
- ETT (≥6 cm)
- Bougie
- Surgilube
- Bag valve mask

H. Young II, MD
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: hyoungii@ufl.edu

S. Toohey, MD
Department of Emergency Medicine, University of California Irvine Medical Center, Orange, CA, USA
e-mail: stoohey@uci.edu

B. Chakravarthy, MD, MPH
Department of Emergency Medicine, University of California Irvine, Orange, CA, USA
e-mail: bchakrav@uci.edu

L.K. Beattie, MD, MS (✉)
Department of Emergency Medicine, University of Florida, Gainesville, FL, USA
e-mail: lars.beattie@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_19

Fig. 19.1 *Right to left*, *top to bottom*: ETT (≥6 cm); bag valve mask; trauma shears; scalpel #11 blade; bougie

19.3.1.2 Procedure

1. Apply topical antiseptic.
2. Remove the 15-mm ETT ventilator connector from the ETT end.
3. Place copious Surgilube on the bougie and railroad over the end of the bougie.
4. Palpate the thyroid notch, cricothyroid membrane, and hyoid bone for orientation (Fig. 19.2a).
5. Stabilize the thyroid cartilage between the thumb and the middle finger of the nondominant hand.
6. Make a vertical skin incision (2–3 cm) over the cricothyroid membrane (Fig. 19.2b).
7. Use the index finger to palpate the cricothyroid membrane.
8. Turn the scalpel 90° and make a 1.5-cm horizontal incision through the lower half of the cricothyroid membrane (Fig. 19.3a).
9. With the scalpel still in the incision, turn it 90°, and insert the bougie into the incision, using the blade as a guide (Fig. 19.3b).
10. Advance the bougie caudally 5–6 cm. Stop if resistance is encountered.
11. Slide the ETT over the bougie into the incision (Fig. 19.4a).
12. Inflate the ETT cuff and ventilate the patient (Fig. 19.4b).
13. Verify the position of the ETT via auscultation, end-tidal CO_2 ($EtCO_2$), and chest radiograph.
14. Secure the ETT.

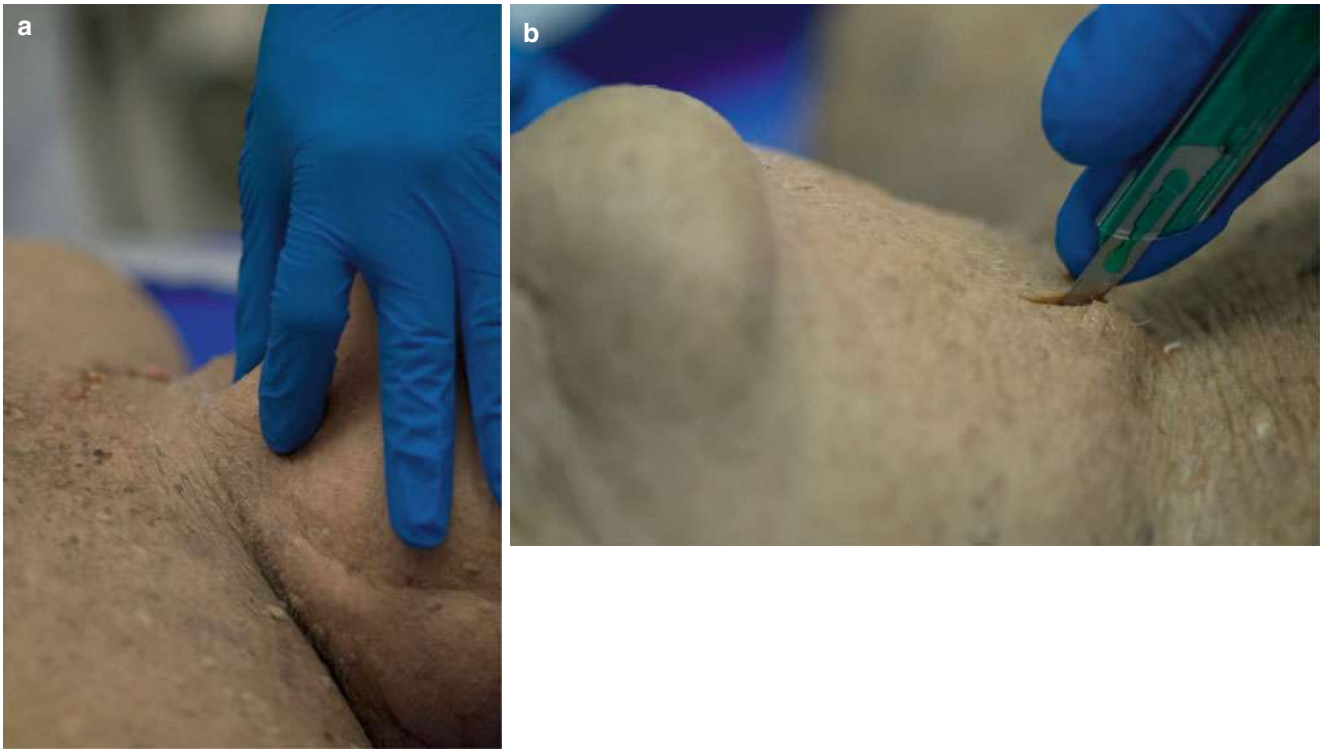

Fig. 19.2 (**a**) Palpate and stabilize the thyroid cartilage, (**b**) make a vertical 2–3 cm incision over the cricothyroid membrane

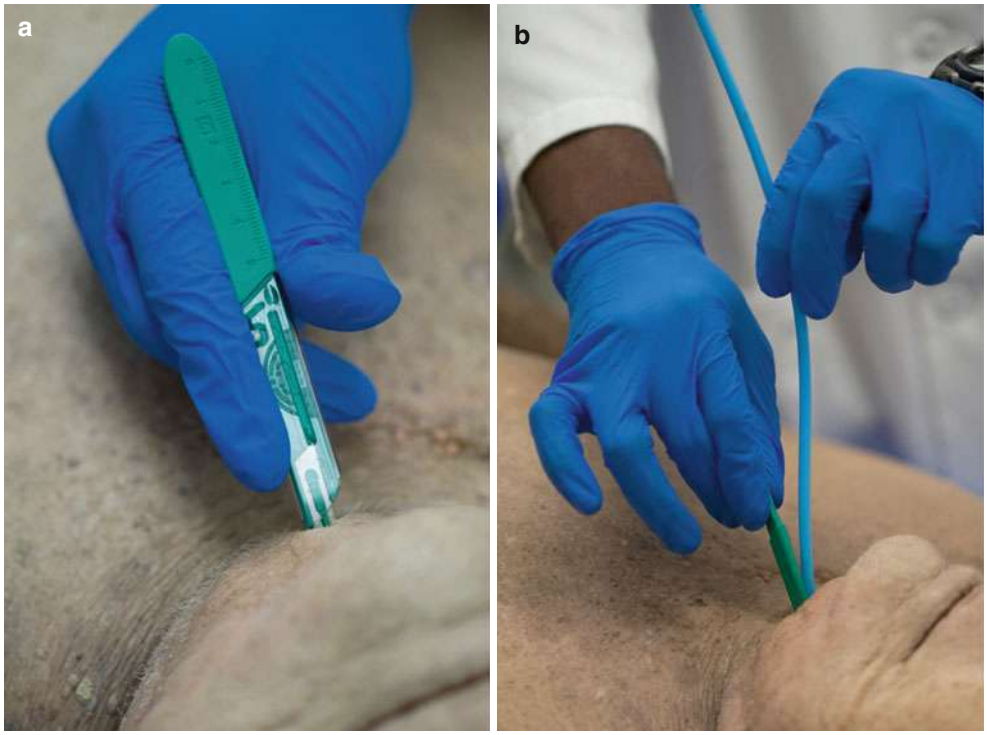

Fig. 19.3 (**a**) Make a horizontal incision into the cricothyroid membrane, (**b**) insert the bougie into the incision made in the cricothyroid membrane

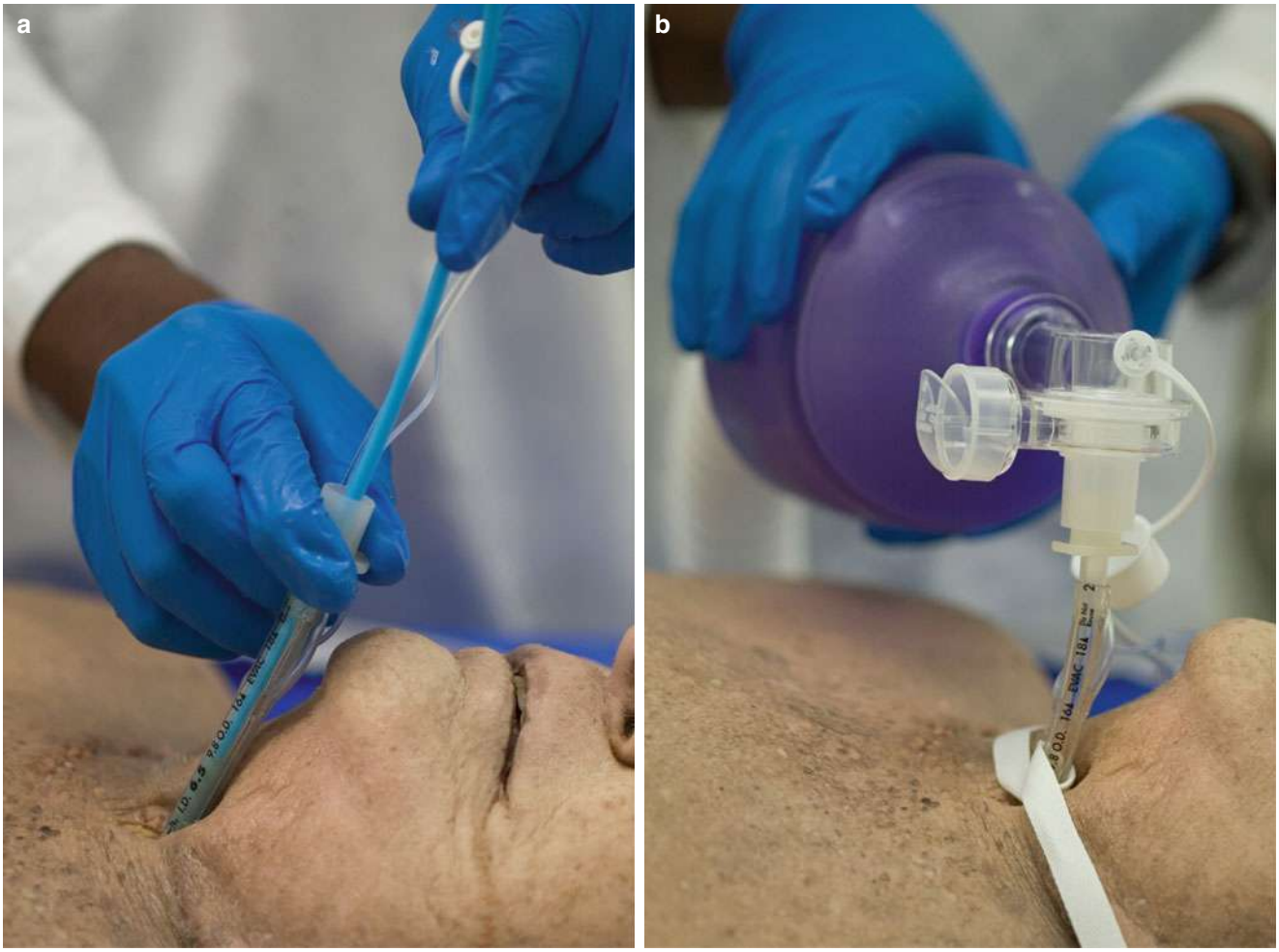

Fig. 19.4 (**a**) Slide the endotracheal tube (ETT) over the bougie into the trachea, (**b**) ventilate the patient

19.3.2 Scalpel-Trousseau

19.3.2.1 Materials and Medications (Fig. 19.5)

- Scalpel with #11 blade
- Tracheal hook
- Trousseau dilator
- Cuffed tracheostomy tube (TT) (6.5 or 7.0) or ETT (5.0, 5.5, or 6.0)
- Antiseptic preparation

Fig. 19.5 *Top to bottom*, *right to left*: Scalpel with #11 blade; tracheal hook; Trousseau dilator; cuffed tracheostomy tube (TT) (6.5 or 7.0) or ETT (5.0, 5.5, or 6.0); antiseptic preparation

19.3.2.2 Procedure

1. Apply topical antiseptic.
2. Palpate the thyroid notch, cricothyroid membrane, and hyoid bone for orientation.
3. Stabilize the thyroid between the thumb and the middle finger of the nondominant hand (Fig. 19.6a).
4. Make a vertical skin incision (2–3 cm) over the cricothyroid membrane (Fig. 19.6b).
5. Palpate with the index finger to verify the cricothyroid membrane location.
6. Use stabilization of the thyroid and palpation to maintain orientation of the anatomy.
7. Make a 1.5-cm horizontal incision through the lower half of the membrane (Fig. 19.7a).
8. Insert a tracheal hook into the incision, then rotate such that hook faces superiorly (Fig. 19.7b).
9. Withdraw at a 45° angle in a cephalad direction, applying gentle traction to the thyroid cartilage.
10. Place the Trousseau dilator into the incision transversely and open the membrane incision vertically (Fig. 19.8a).
11. Insert a cuffed ETT (5.0–6.0) or TT (6.5–7.0) into the incision between the prongs of the dilator in the horizontal access (Fig. 19.8b).
12. Rotate both the dilator and the ETT toward the head of the patient and then direct the tube downward into the trachea while removing the dilator.
13. Inflate the ETT cuff and ventilate the patient.
14. Verify the position of the ETT via auscultation, $EtCO_2$, and chest x-ray.
15. Once placement of the tube has been verified, the tracheal hook can be removed.
16. Secure the ETT.

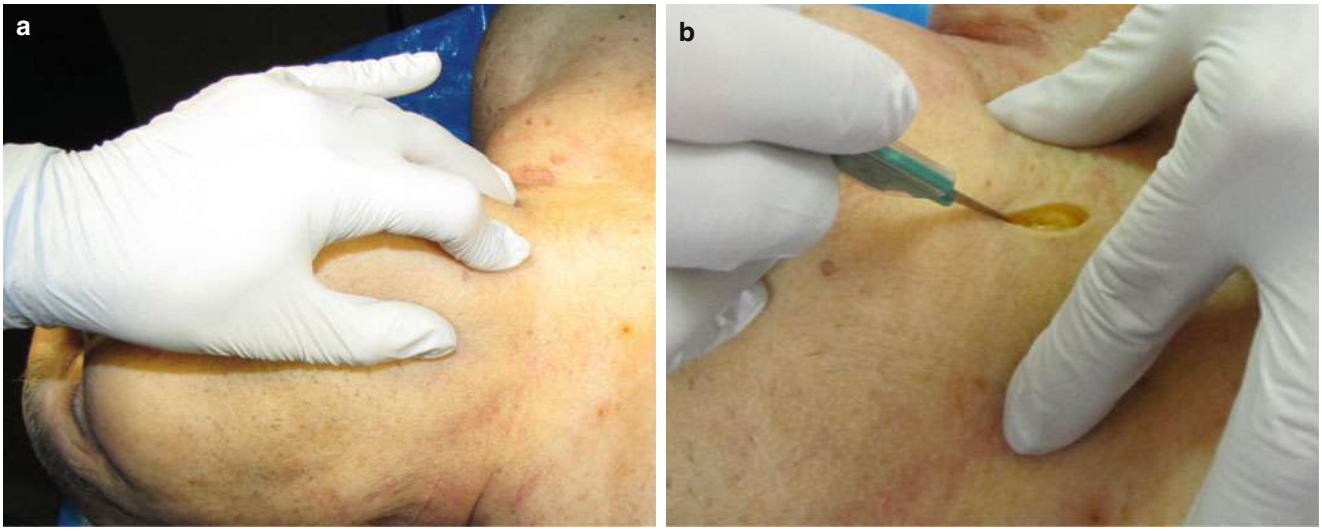

Fig. 19.6 (**a**) Stabilize the thyroid cartilage between the thumb and the middle finger of the nondominant hand; (**b**) make a vertical incision over the cricothyroid membrane

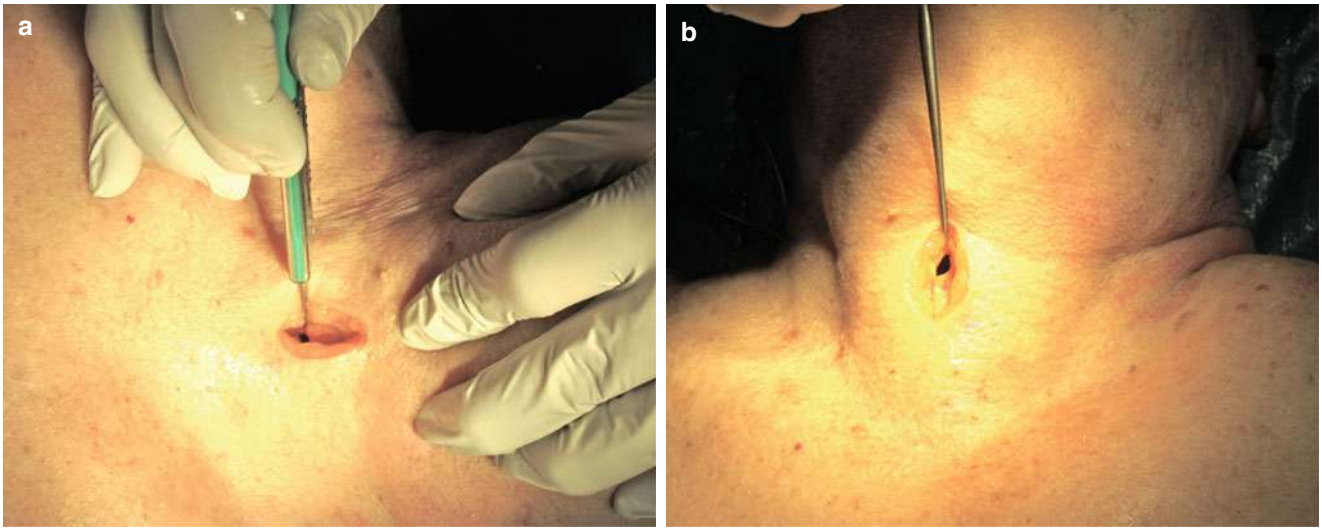

Fig. 19.7 (**a**) Make a horizontal incision in the lower half of the cricothyroid membrane. (**b**) Insert a tracheal hook and apply gentle traction superiorly at a 45° angle

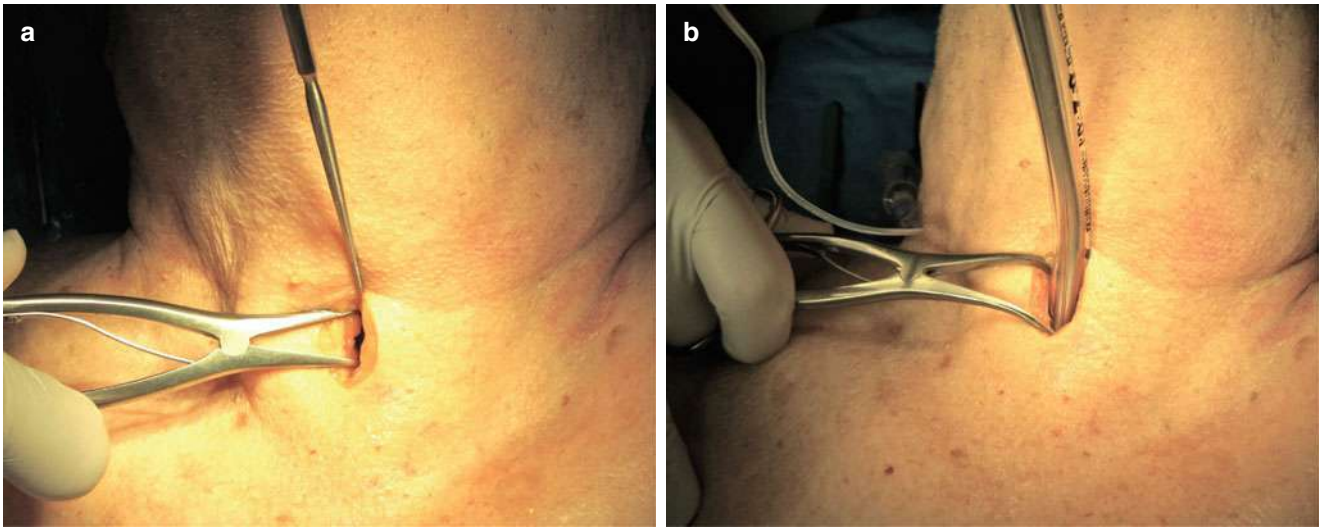

Fig. 19.8 (**a**) Insert the Trousseau dilator into the incision, and open the path for placement of the tracheostomy tube (TT) or the ETT. (**b**) Insert the TT or the ETT into the expanded incision between the Trousseau dilator prongs

19.4 Complications

- Bleeding
- ETT misplacement (false passage, through the thyrohyoid membrane, unintentional tracheostomy)
- Hoarseness, dysphonia, or vocal cord paralysis
- Subglottic or laryngeal stenosis
- Damage to thyroid cartilage, cricoid cartilage, or tracheal rings
- Perforated esophagus
- Infection
- Aspiration

19.5 Pearls

- Predictors of difficult cricothyrotomy: "SHORT" [1]
  - Prior *S*urgery or *S*car tissue
  - *H*ematoma
  - *O*bese
  - Prior *R*adiation
  - *T*umor/abscess
- The incision should cut through the skin and subcutaneous tissue down to the cricothyroid membrane and cartilages.
- Vertical incisions allow for extension in either direction if the cricothyroid membrane is above or below initial incision.
- Blood in the field may hinder visualization of the membrane, but the airway should be established before attempts to control any bleeding.
- Cricothyroid arteries are located cephalad to the cricothyroid membrane.
- Either a TT or an ETT can be used for the procedure.
  - ETTs are more ubiquitous.
  - TTs are easier to secure.
- If an ETT is used, a stylet can help direct placement.
- Cricothyroidotomy is preferred in the emergency setting over tracheostomy owing to the increased risks of bleeding, the mobility of the trachea, and the risk of lacerating the underlying thyroid gland [2].
- Ultrasound can be used to visualize landmarks in patients in whom landmarks are difficult to identify.

References

1. Walls RM, Murphy MF, editors. Manual of emergency airway management. 4th ed. Philadelphia: Wolters Kluwer; 2012.
2. Boon JM, Abrahams PH, Meiring JH, et al. Cricothyroidotomy: a clinical anatomy review. Clin Anat. 2004;17:478–86.

Selected Reading

DiGiacomo C, Neshat KK, Angus LD, et al. Emergency cricothyrotomy. Mil Med. 2003;168:541–4.

Hamilton PH, Kang JJ. Emergency airway management. Mt Sinai J Med. 1997;64:292–301.

Helm M, Gries A, Mutzbauer T. Surgical approach in difficult airway management. Best Pract Res Clin Anaesthesiol. 2005;19:623–40.

Sagarin MJ, Barton ED, Chng YM, Walls RM. Airway management by US and Canadian emergency medicine residents: a multicenter analysis of more than 6,000 endotracheal intubation attempts. Ann Emerg Med. 2005;46:328–36.

Walls RM. Cricothyroidotomy. Emerg Med Clin North Am. 1988;6:725–36.

20 Tracheostomy Tube Malfunction

Deena Bengiamin and Bharath Chakravarthy

Tracheostomy tube (TT) malfunction is the source of airway compromise in patients requiring these airway devices. TT malfunction may create an airway emergency, and the timely replacement of TTs is a challenging procedure in the most experienced hands.

20.1 Indications

- Cuff rupture
  - Can lead to dislodgment
- Dislodgment
  - Most common emergency department TT complication
  - Can lead to air passage obstruction
- Obstruction
  - Caused by blood or thick, dry secretions (formed in the absence of nasopharyngeal air humidification).
  - Dried secretions or blood can act as a one-way valve, allowing air in but restricting outward flow.

20.2 Materials and Medications

- Airway suction catheter
- Oxygen humidifier
- Bougie/nasogastric (NG) tube (12 French)
- Saline
- N-Acetyl-cysteine (NAC)
- Appropriately sized endotracheal tube (ETT) or TT

20.3 Procedure

1. Supply high-flow humidified oxygen through a bag valve mask (BVM), or a non-rebreather face mask.
2. Assess clinical indicators to determine TT problem.
   (a) Indicators of cuff rupture: Air leak with BVM and loose TT.
   (b) Indicators of dislodged TT (if TT is still in the stomal opening): Subcutaneous emphysema, crepitus, and diminished or absent breath sounds.
   (c) Indicators of an obstructed TT: ±stridor and diminished or absent breath sounds.
   (d) If time allows, chest radiograph, continuous capnography, and oxygen saturation can be helpful.
3. Obstruction
   (a) Remove the inner cannula and inspect for obstruction—clean if necessary.
   (b) If this fails, 5–10 mL of saline or NAC can be squirted directly down the TT to loosen secretions.
   (c) Suction thoroughly with a suction catheter (Fig. 20.1).
   (d) In refractory cases, the TT will need replacement.
4. Replacement
   (a) Ideally, the replacement tube should be of the same type and size as the original TT.
   (b) A smaller size TT or ETT can be helpful in settings of airway compromise.
   (c) A 6–7.5-cm ETT tube may be used if a TT is unavailable.
   (d) Remove the existing TT.
   (e) Hyperextend the patient's head and neck to maximize visualization of the stoma.

D. Bengiamin, MD
Department of Emergency Medicine, University of San Francisco Fresno, Fresno, CA, USA
e-mail: dbengiamin@fresno.ucsf.edu

B. Chakravarthy, MD, MPH (✉)
Department of Emergency Medicine, University of California Irvine, Orange, CA, USA
e-mail: bchakrav@uci.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_20

(f) Note: Careful inspection of the area is paramount because the thyroid isthmus may obscure visualization of the tracheal stoma.

(g) Techniques

(i) Direct insertion

1. As soon as possible, insert the new TT or ETT into the stoma to prevent stomal narrowing (Fig. 20.2).
2. Inflate the new TT/ETT cuff.

(ii) Bougie or NG tube

1. Lubricate the TT or ETT tube.
2. Lubricate a bougie or 12-French NG tube.
3. Insert the lubricated bougie or NG tube into the TT or ETT.
4. Insert the bougie or NG tube into stoma and advance into the trachea (Fig. 20.3a).
5. Direct the NG tube or bougie caudad toward the lower tracheobronchial tree.
6. Do not advance more than 7 cm.
7. If resistance is noted:
   (a) Either the operator has reached a terminal bronchiole or is in a false passage.
   (b) Do not force bougie/NG tube farther at this point.
   (c) Use clinical judgment (bougie/ETT depth, palpation) to determine likely placement.
8. Advance TT or NG tube over bougie or NG tube is in trachea (Fig. 20.3b).
9. After the TT is in place, remove the bougie/NG tube.

(iii) The fingertip technique

1. Insert a gloved forefinger into tracheal stoma (Fig. 20.4a).
2. Formulate a mental plan as to the direction and path of the stoma.
3. Then place a TT or an ETT into the stoma as the finger is withdrawn (Fig. 20.4b).

5. If placement of a TT or ETT is not possible through the stoma, consider endotracheal intubation.

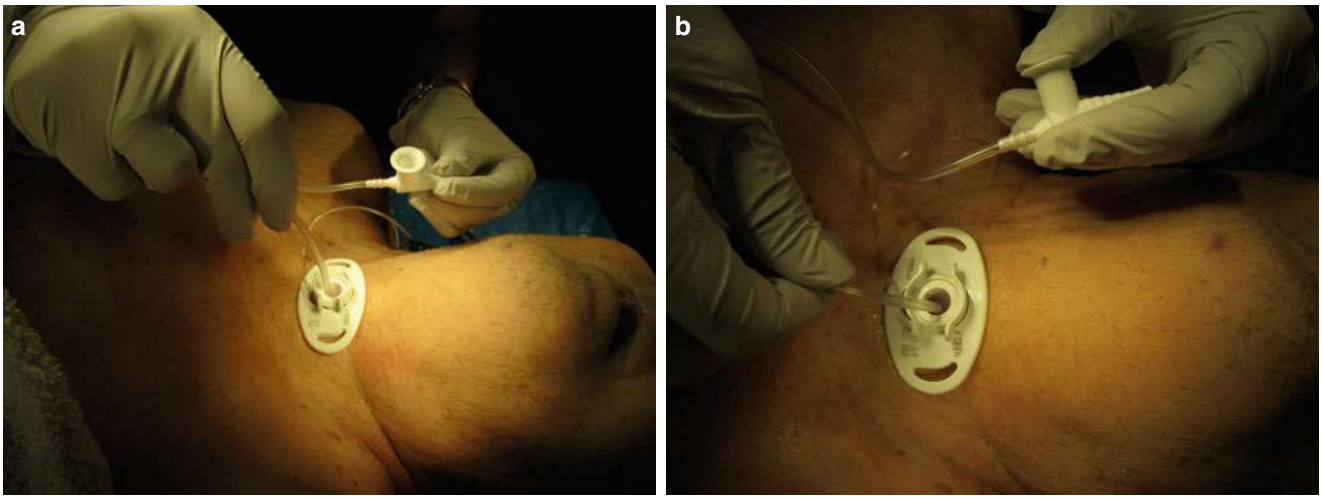

Fig. 20.1 (**a**) Insert the suction catheter to the appropriate depth, keeping the suction port open. (**b**) Slowly withdraw the catheter in a circular motion, keeping the suction port closed

Fig. 20.2 Insertion of the endotracheal tube (ETT) directly into the stoma

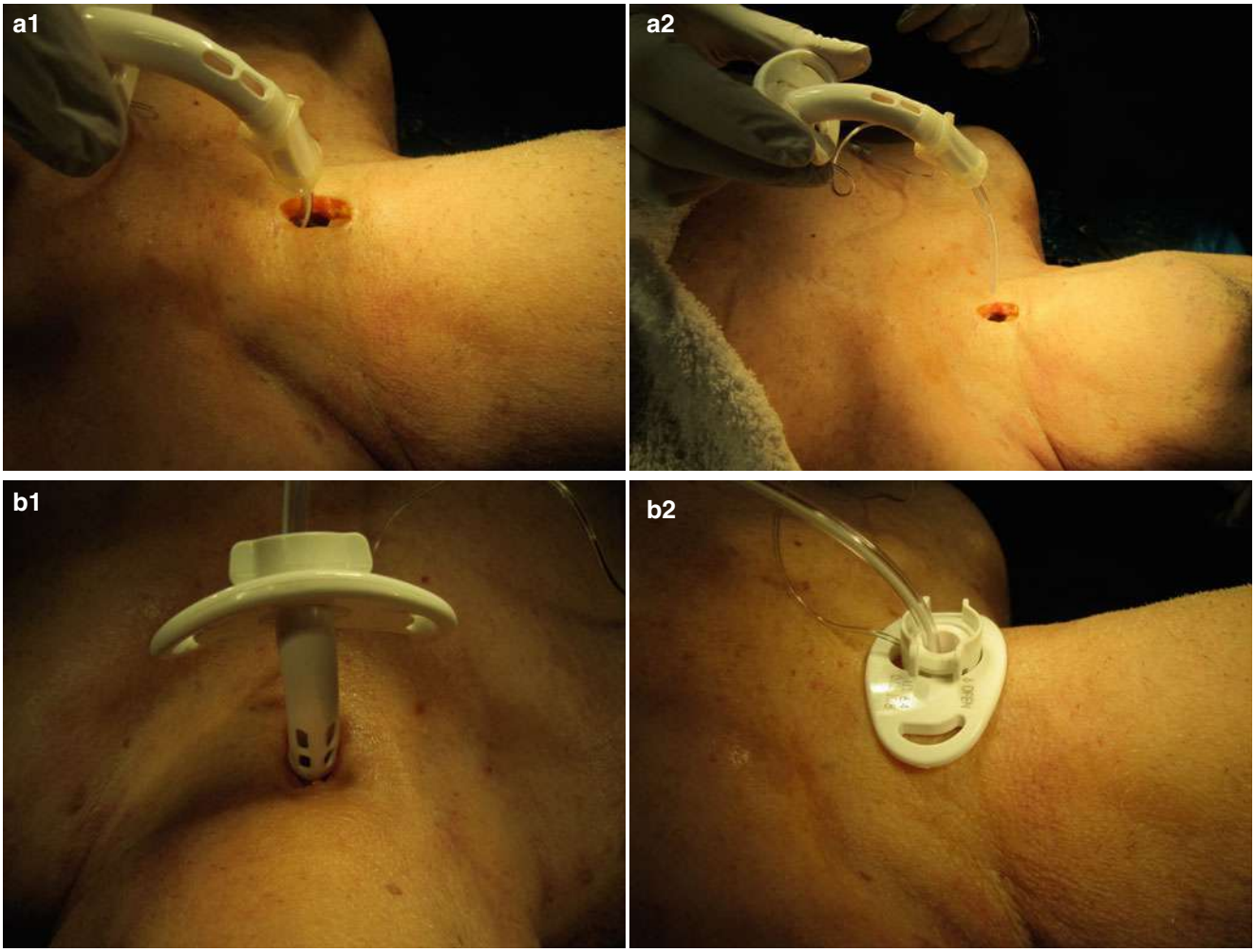

Fig. 20.3 (**a1, a2**) Insert the tracheostomy tube (TT) over the nasogastric (NG) tube or bougie, which is first inserted into the trachea. (**b1, b2**) Advance the TT over the NG tube or bougie, which serves as a guidewire

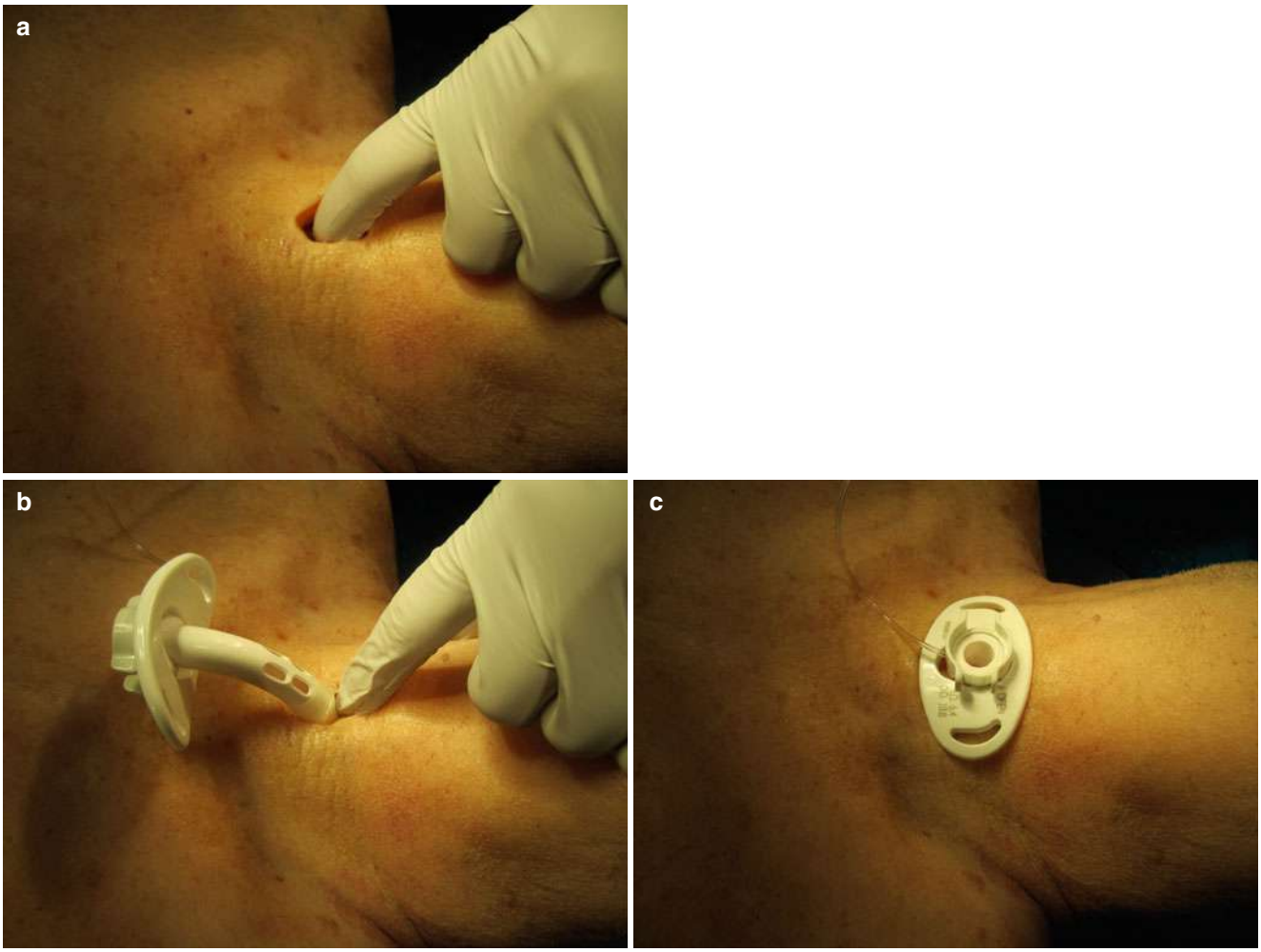

Fig. 20.4 (**a**) Insert a gloved finger into the stoma. (**b**) Insert the TT along the path while withdrawing the gloved finger. (**c**) TT in place

20.4 Pearls and Pitfalls

- Stomal closing
  - Stomal constriction begins as soon as the TT is removed or displaced.
  - Forceful attempts to replace a large TT may result in false passages and trauma.
- Tracheal stenosis
  - Constant TT cuff pressure may cause necrosis, ulceration, and granulation tissue formation, leading to tracheal narrowing.
  - Complicates TT replacement in the setting of dislodgment or obstruction.
  - Bougie use can be helpful to place a small(er) ETT until surgical dilation and/or resection can be achieved.
- Creation of a false lumen during recannulation
  - Subcutaneous emphysema will be an early indicator of this.
  - Early confirmation of correct placement of TT by confirming:
    - Continuous end-tidal CO_2 monitoring
    - Equal chest rise
    - Bilateral breath sounds
- Unrecognized trachea–innominate artery fistula (Fig. 20.5)
  - Usually occurs within 3–4 weeks of placement
  - Presentation: Bleeding around the tracheostomy tube (>10 mL) or massive hemoptysis
  - Requires
    - ETT cuff overinflation to compress the fistula.
    - Digital pressure on stoma may be helpful to tamponade the bleeding.
    - Place stomal ETT deep to bleeding fistula to protect airway.
    - Definitive surgical intervention in operating room.
  - Associated with high mortality
- Unrecognized tracheoesophageal fistula
  - Usually iatrogenic injury from TT placement or NG tube erosion
  - Presentation: Dyspnea, copious TT secretions, recurrent food aspiration, and gastric distention
  - Requires
    - Bronchoscopy or swallowing studies to confirm diagnosis
    - Surgical repair or stenting

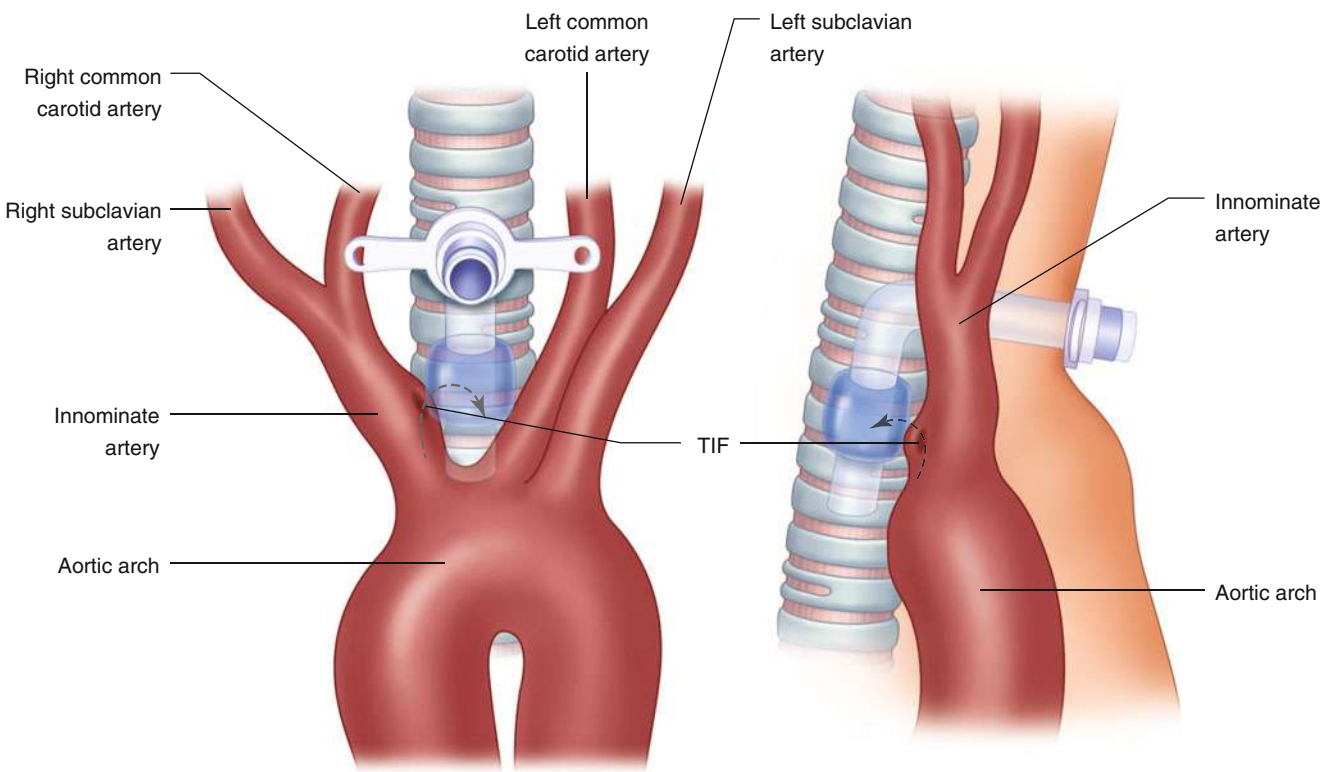

Fig. 20.5 Anatomical relationship between the trachea and the innominate artery. *TIF* trachea–innominate artery fistula

Selected Reading

De Leyn P, Bedert L, Delcroix M, et al. Tracheotomy: clinical review and guidelines. Eur J Cardiothorac Surg. 2007;32:412–21.

Dobiesx VA, Miller SA, Pitzele MJ. Complications of tracheostomies. In: Wolfson AB, Hendey GW, Ling LJ, Rosen CL, Scheider JJ, Sharieff GQ, editors. Harwood-Nuss' clinical practice of emergency medicine. 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2009.

Epstein SK. Late complications of tracheostomy. Respir Care. 2005;40:542–9.

Friedman M, Ibrahim H. The dislodged tracheostomy tube: "fingertip" technique. Oper Technol Otolaryngol. 2002;13:217–8.

Young JS, Brady WJ, Kesser B, Mullins D. A novel method for replacement of the dislodged tracheostomy tube: the nasogastric tube "guidewire" technique. J Emerg Med. 1996;14:205–8.

21 Percutaneous Transtracheal Jet Ventilation

Clint Masterson

21.1 Indications

- Failure to control the airway by other means
- As a temporary measure while preparing for definitive airway control
- Securing the airway in crash airways in infants and small children

21.2 Contraindications

- Absolute
  - Transection of the trachea below the cricothyroid membrane
- Relative
  - Inability to identify the cricothyroid landmarks
  - Anatomical distortion to the cricothyroid membrane
  - Supraglottic obstruction (preventing gas exhalation)

21.3 Materials and Medications (Fig. 21.1)

- Betadine, chlorhexidine, or similar skin sterilization solution
- 12- to 16-gauge angiocatheter or transtracheal jet ventilation (TTJV) purpose-specific catheter
- 10-mL syringe filled with 4 mL of normal saline, 2 % lidocaine, or viscous lidocaine
- Hand-operated regulator valve
- Attach oxygen supply.
  - Connect kit tubing to wall oxygen OR.
  - Connect 7–0 endotracheal connector to bag valve mask (BVM) attached to oxygen.

C. Masterson, MD
Department of Emergency Medicine, Mayo Clinic Health System in Fairmont, Fairmont, MN, USA
e-mail: clint@clintm.us

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_21

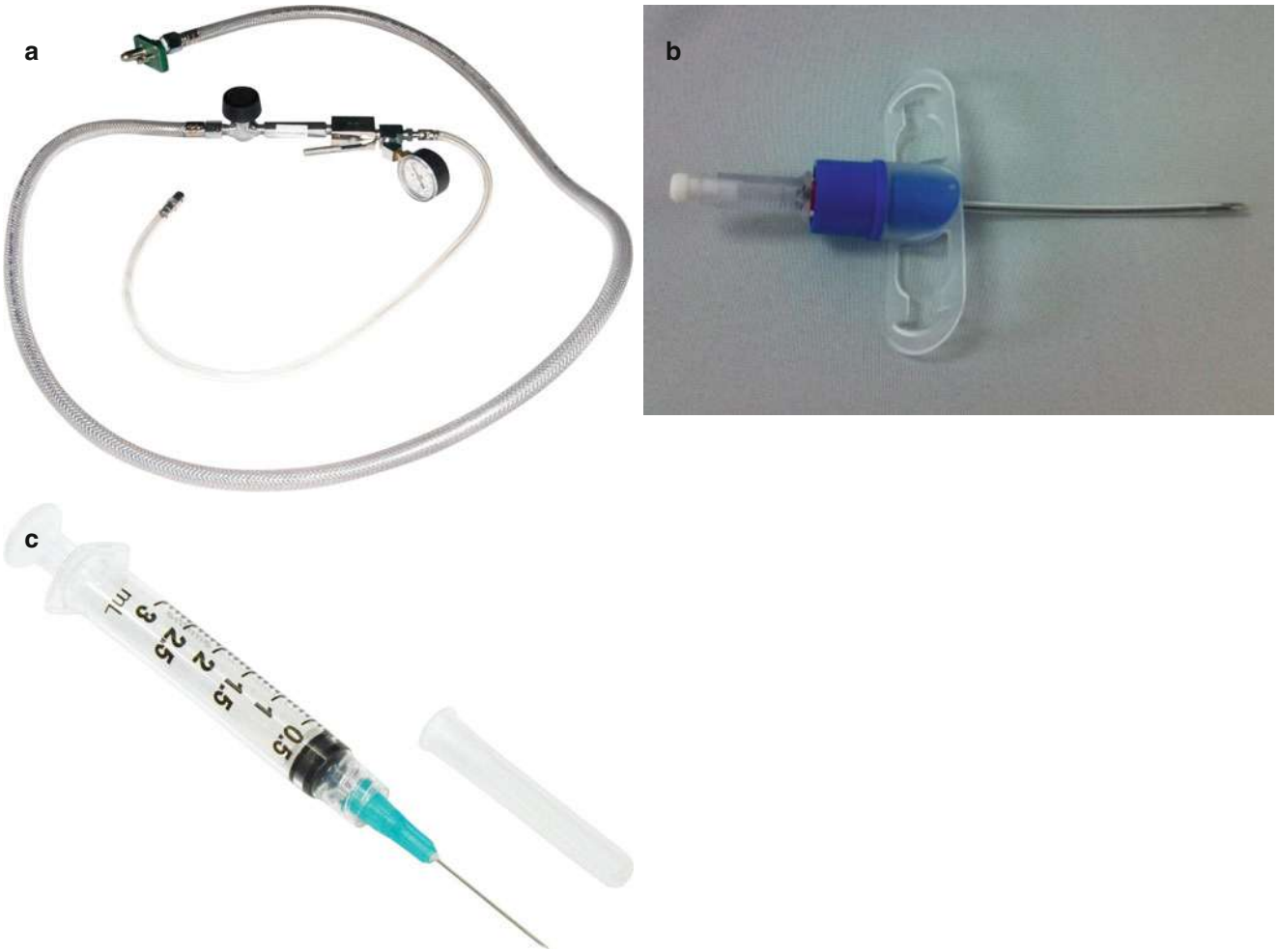

Fig. 21.1 (**a**) Tubing and regulator valve, (**b**) commercially available catheter, (**c**) 3-mL or 10-mL syringe

21.4 Procedure

1. Attach the tubing and the hand-operated regulator valve to wall oxygen (Fig. 21.2a), and place the distal end of the tubing near the patient in preparation for ventilation.
2. Adjust regulator to maximum pressure, 50 psi if possible (Fig. 21.2b).
3. Palpate the cricothyroid membrane just distal to the thyroid prominence (Fig. 21.3).
   (a) Sterilize the area with a suitable cleansing agent.
   (b) Use the thumb and index finger of the nondominant hand to stabilize the trachea for the procedure.
4. Attach the TTJV catheter (or angiocatheter) to the syringe (Fig. 21.4).
5. Advance the catheter through the cricothyroid membrane at a 30–45° caudal direction while aspirating with the syringe (Fig. 21.5).
6. Return of air confirms entry into the trachea.
7. If lidocaine is utilized, it can then be injected to prevent spasm during the procedure.
8. Fully advance the angiocatheter and secure it while the needle and syringe are withdrawn.
9. Remove the needle, secure it to the skin, and connect it to the regulator hose.
10. Secure the distal end of the oxygen tubing (distal to the hand-operated valve) to the catheter (Fig. 21.6).
11. If a BVM is used as the oxygen source:
    (a) Attach a 3-mL syringe to the angiocatheter.
    (b) Attach the BVM with the 7–0 endotracheal tube (ETT) connector to the end of the plungerless 3-mL syringe (Fig. 21.7).
12. Operate the valve 12–20 times a minute with long periods to allow gas exhalation and exchange (Fig. 21.8).
13. Preparations should be made for a definitive airway as soon as possible—preferably within 15 min.

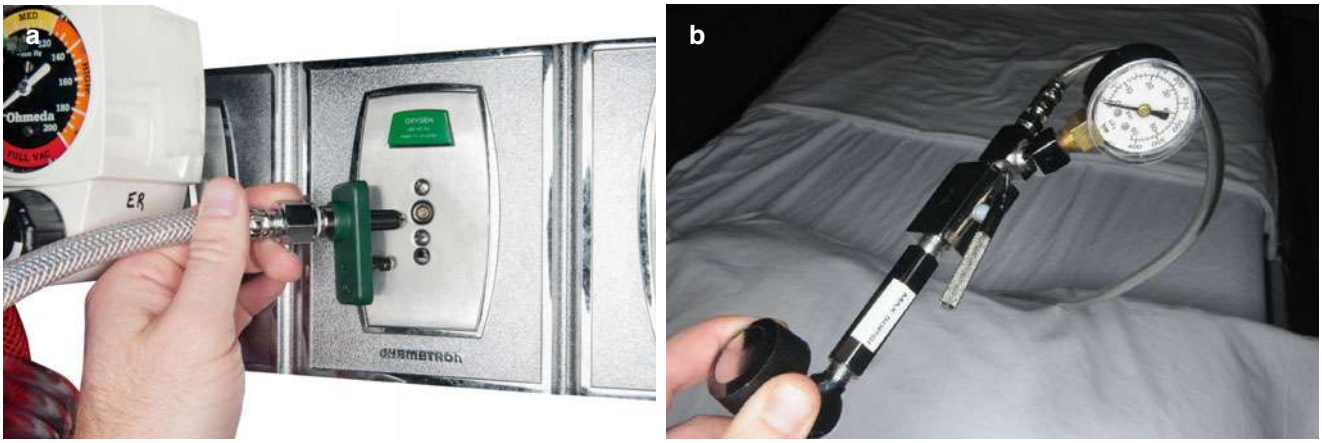

Fig. 21.2 (**a**) Attach tubing and the hand-operated regulator valve to wall oxygen, and (**b**) adjust regulator to maximum pressure (50 psi if possible)

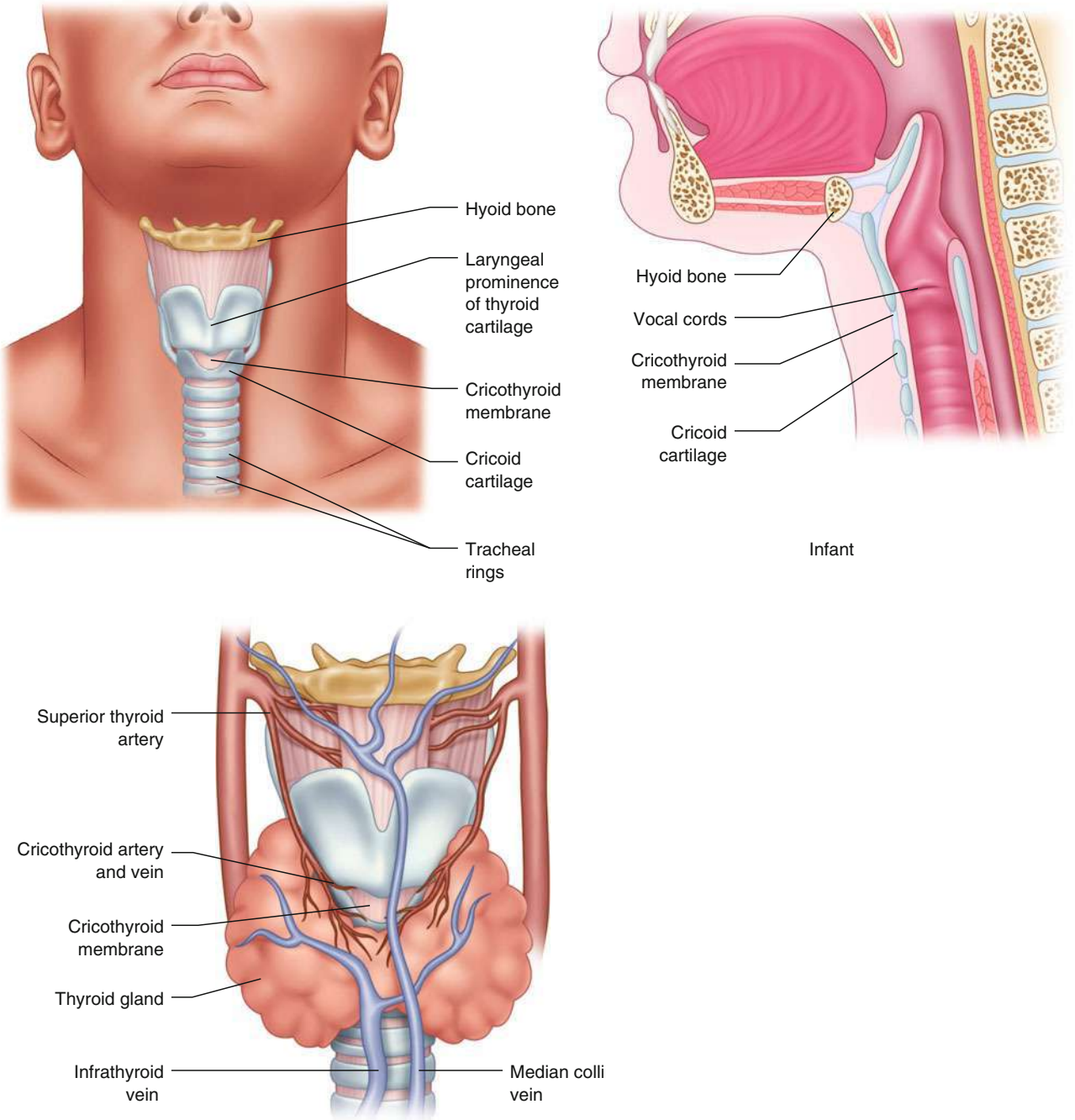

Fig. 21.3 Airway anatomy

Fig. 21.4 Attach the TTJV catheter to the syringe

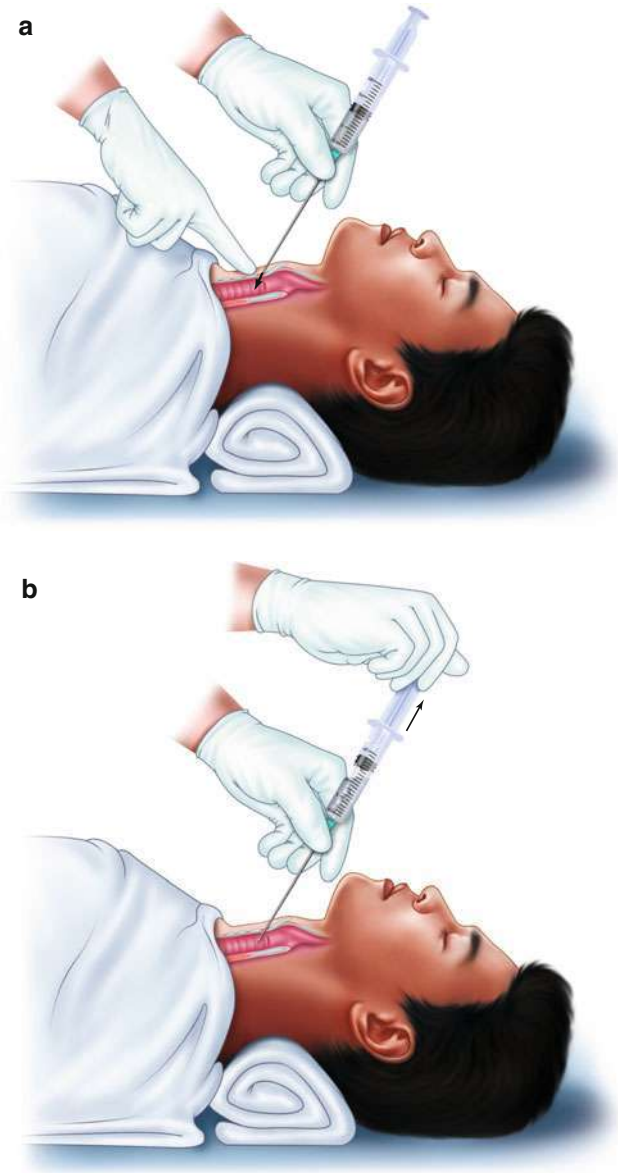

Fig. 21.5 (**a**) Advance the catheter through the cricothyroid membrane at a 30–45° caudal direction (**b**) while aspirating with the syringe

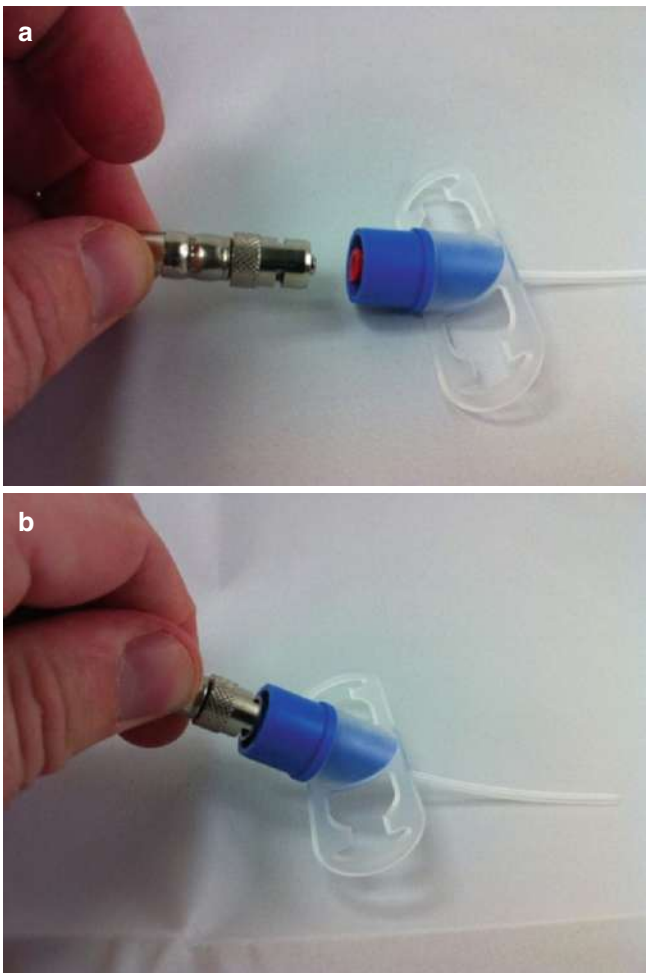

Fig. 21.6 Secure the distal end of the oxygen tubing to the catheter

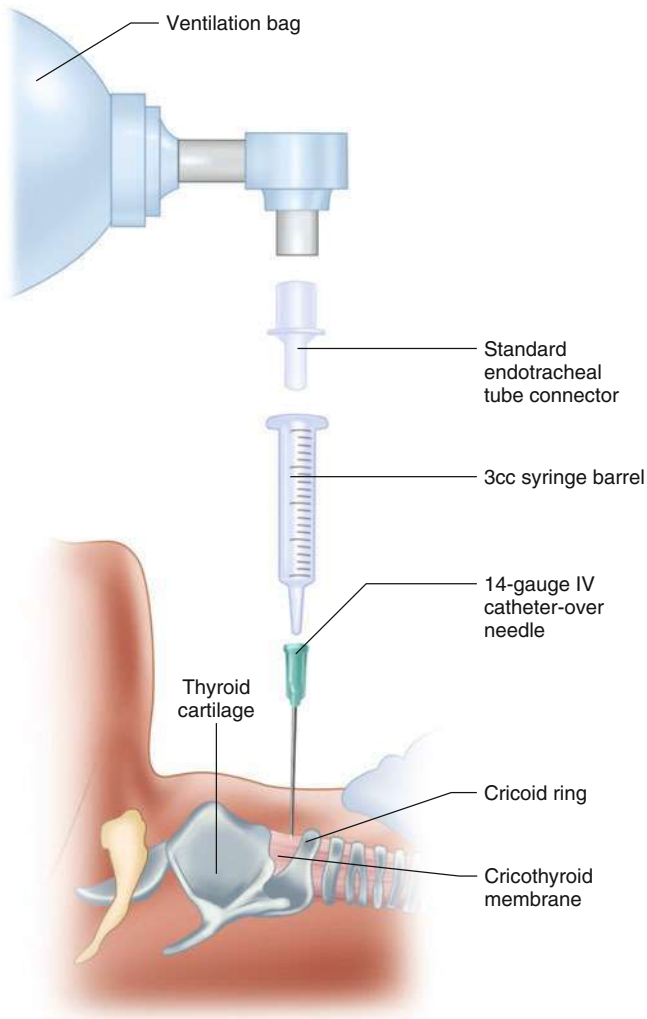

Fig. 21.7 Attach the BVM with the 7–0 endotracheal tube (ETT) connector to the end of the plungerless 3-mL syringe

Fig. 21.8 Operate the valve 12–20 times a minute with long periods to allow gas exhalation and exchange

21.5 Complications

- Pneumothorax
- Pneumomediastinum
- Subcutaneous emphysema
- Catheter kink or misplacement
- Hypercarbia and respiratory acidosis
  - Use of TTJV for prolonged periods of time without adequate ventilation will elevate CO_2.
- Barotrauma
- Coughing in conscious patients
- Aspiration
- Persistent stoma

21.6 Pearls and Pitfalls

- If the wall connector does not have a pressure regulator, it can still be used although the risk of barotrauma is greater. Use the endpoint of chest rise to determine the end of each ventilation burst in this case.
- Higher pressures and lack of supraglottic air exchange are risk factors for pneumothorax. If the supraglottic area is obstructed, a Y catheter can be attached to allow gas to escape before the next insufflation.
- TTJV may or may not allow sufficient gas exchange to prevent hypercarbia. Preparations should be made to obtain a definitive airway as soon as possible.
- Endotracheal intubation may be facilitated by the high pressures insufflated in the trachea, and a repeat attempt may be performed after the transtracheal ventilation is achieved.

Selected Reading

Patel R. Percutaneous transtracheal jet ventilation. A safe, quick and temporary way to provide oxygenation and ventilation when conventional methods are unsuccessful. Chest. 1999;116:1689–94.

Roberts JR, Hedges JR. Clinical procedures in emergency medicine. Philadelphia: Saunders Elsevier; 2010.

Tintinalli J. Tintinalli's emergency medicine: a comprehensive study guide. 7th ed. New York: McGraw Hill; 2010.

Walls R. Manual of emergency airway management. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

Part III

Thoracic Procedures

Needle Thoracostomy

22

Lucas McArthur and Christian Fromm

22.1 Indications

Needle decompression thoracostomy is a procedure used in the emergent treatment of a tension pneumothorax. Tension pneumothorax is a clinical diagnosis. Decompression treatment should not be delayed in order to obtain radiographic confirmation. The following scenarios illustrate some of the clinical signs that *may* be present in such patients:

- *Awake patient* with suspected or confirmed tension pneumothorax
  - Chest pain
  - Respiratory distress
  - Decreased breath sounds with hyperresonance and/or subcutaneous emphysema
  - Trachea deviated away from the side of the pneumothorax
  - Tachycardia
  - Falling pulse oximetry (SpO_2)
  - Shock
- *Ventilated patient* with suspected or confirmed pneumothorax (often insidious)
  - Increased resistance to ventilation
  - Hypotension
  - Elevated central venous pressure
  - Tachycardia
  - Decreased breath sounds with hyperresonance and/or subcutaneous emphysema
  - Trachea deviated away from the side of the pneumothorax
  - Falling SpO_2
  - Shock
- *Injured patient* (especially with penetrating chest trauma) with suspected or confirmed tension pneumothorax
  - In arrest
  - Unexplained hypotension
  - Apnea
  - Decreased breath sounds with hyperresonance and/or subcutaneous emphysema

22.1.1 Absolute Indications

- Patient in acute respiratory distress with rapid decompensation secondary to suspected or confirmed tension pneumothorax
- Injured patient in extremis with apnea, unexplained hypotension, or arrest

22.2 Contraindications

- No absolute contraindications.

22.3 Materials

- Large-bore needle/angiocatheter (minimum of 16 gauge)
- 10-mL syringe (optional)
- One-way valve (optional)
- Betadine (povidone-iodine) swab/chlorhexidine scrub
- Tape

L. McArthur, MD (✉)
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: LMcarthur@jfkhealth.org

C. Fromm, MD
Department of Emergency Medicine, Maimonides Medical Center, SUNY Downstate College of Medicine, New York, NY, USA
e-mail: cfromm@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_22

22.4 Procedure

1. Expose the anterior chest at the level of the second intercostal space on the affected side (Fig. 22.1). Alternatively, expose the chest wall at the level of the anterior axillary line in the fourth or fifth intercostal space on the affected side.
2. Cleanse the area with a Betadine swab or chlorhexidine scrub (Fig. 22.2).
3. Using a gloved hand, locate the second intercostal space at the midclavicular line.
   (a) The first rib is normally not felt.
   (b) The second rib is felt just below the clavicle.
   (c) The second intercostal space is the area between the second and the third ribs.
   Note: Alternatively, this procedure may also be performed on the midaxillary line in the fourth intercostal space of the affected side. The same general steps listed later are employed in this approach and care is taken to avoid the neurovascular bundles inferior to the fourth rib.
4. Insert the needle/angiocatheter perpendicular to the chest wall into the second intercostal space just above the superior edge of the third rib to avoid the intercostal neurovascular bundle (Fig. 22.3).
   (a) This step may be done with or without a syringe attached.
   (b) Local anesthesia is usually unnecessary but may be used if the patient is not in extremis.
5. Carefully walk the needle over the third rib and advance until the pleural space is entered.
   (a) Entry into the pleural space is accompanied by a "popping" sound or a sensation of "giving way."
6. If you are able to withdraw air with the syringe or hear a "hiss" of air escaping through the angiocatheter during expiration and inspiration, then placement is considered successful.
7. After removing the needle, secure the angiocatheter in place with tape (Fig. 22.4).
   Caution: Do not reinsert needle into the angiocatheter owing to the danger of sheering the angiocatheter.
8. Assess the patient and evaluate the effectiveness of the procedure.
   (a) The patient should exhibit immediate and obvious improvement in respiratory status including improved lung sounds and vital signs.
   (b) The procedure may be repeated if the patient is not improving.
   (c) Excess pleural air may be aspirated through the angiocatheter with a syringe.
9. Obtain a chest radiograph to confirm success.
   (a) Repeat in 6 h.
10. Because needle decompression is only a temporizing measure, tube thoracostomy (see Chap. 23) must be performed for definitive management of the pneumothorax.

Fig. 22.1 The preferred site for needle thoracostomy is the second intercostal space in the midclavicular line

Fig. 22.2 Prepare the skin with povidone-iodine or chlorhexidine

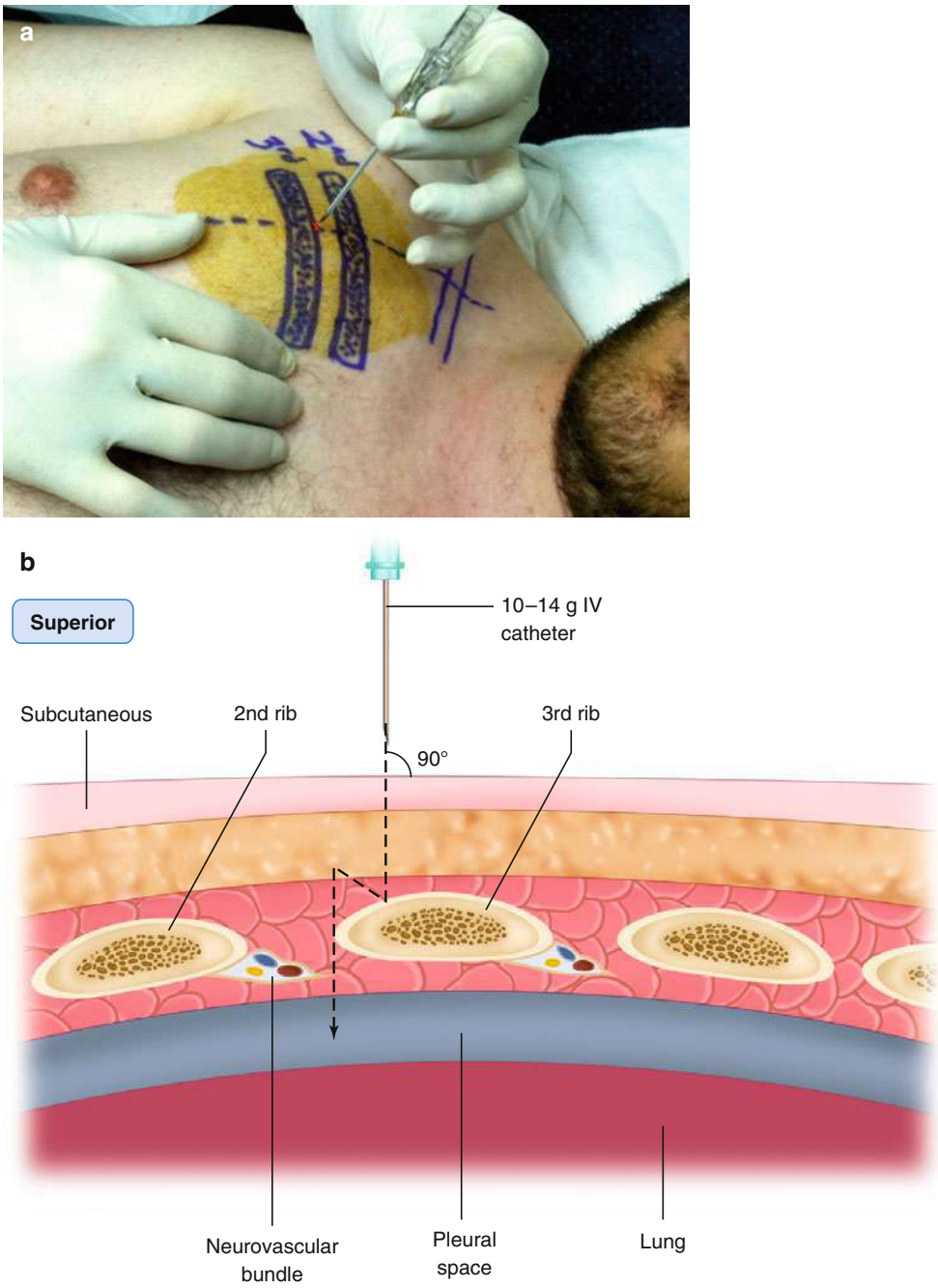

Fig. 22.3 (**a**) Insert the needle into the second intercostal space just above the superior edge of the third rib and (**b**) avoid the neurovascular bundle by approaching the skin with the needle perpendicular to the chest wall just above the superior edge of the third rib

Fig. 22.4 After removing the needle, secure the angiocatheter in place with tape

22.5 Complications

- Failure to resolve the tension pneumothorax.
  - Obese or muscular patients may require a longer needle and catheter to reach the pleural space or, alternatively, may require proceeding immediately to tube thoracostomy.
- Iatrogenic pneumothorax.
- Laceration of intercostal artery or nerve.
- Rapid re-expansion may result in the development of pulmonary edema.
- Infection.

22.6 Pearls and Pitfalls

- Use the sternum as a landmark to more easily locate the second and third ribs.
- The same procedure may also be done on the midaxillary line in the fourth intercostal space, which is usually landmarked with the nipple.
- Primary pneumothorax is unusual in those older than 40 years. Consider the presence of underlying disease in this population.

Acknowledgments We would like to thank Antonios Likourezos, MA, MPH, and Abraham Lederman for assisting with the photographs.

Selected Reading

Britten S, Palmer SH. Chest wall thickness may limit adequate drainage of tension pneumothorax by needle thoracocentesis. Emerg Med J. 1996;13:426–7.

Custalow CB. Color atlas of emergency department procedures. Philadelphia: Saunders; 2005.

Leigh-Smith S, Harris T. Tension pneumothorax—time for a re-think? Emerg Med J. 2005;22:8–16.

Roberts JR, Hedges JR. Clinical procedures in emergency medicine. 3rd ed. Philadelphia: Saunders; 1998.

Chest Tube Thoracostomy

23

Brandon R. Allen and Latha Ganti

23.1 Indications

- Spontaneous pneumothorax (large and/or symptomatic)
- Tension pneumothorax (or suspected)
- Iatrogenic pneumothorax
- Penetrating chest injuries
- Hemopneumothorax in acute trauma
- Patient in extremis with evidence of thoracic trauma
- Complicated parapneumonic effusions (empyema)
- Chylothorax/hemothorax
- Post-thoracic surgery
- Bronchopleural fistula

23.2 Contraindications

- Absolute
  - Emergent thoracotomy
- Relative
  - Coagulopathy
  - Pulmonary bullae
  - Pulmonary, pleural, or thoracic adhesions
  - Loculated pleural effusion or empyema
  - Skin infection over the chest tube insertion site

23.3 Materials and Medications

- Tube thoracostomy tray
  - #10 scalpel; 18-, 22-, and 25-gauge needles; 10-mL syringes; forceps; clamps; scissors; drape; abdominal pads; 0 or 1–0 silk suture; needle driver; curved clamp (Fig. 23.1a)
- Betadine (povidone-iodine) or other skin antiseptic preparation solution
- Lidocaine (1 % or 2 % with epinephrine)
- Appropriate chest tube size (approximate)
  - Adult male: 28–36 French
  - Adult female: 28 French
  - Child: 12–24 French
  - Infant: 12–16 French
  - Neonate: 10–12 French

Vaseline gauze
Chest drainage system (Fig. 23.1b)

B.R. Allen, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: brandonrallen@ufl.edu

L. Ganti, MD, MS, MBA (✉)
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_23

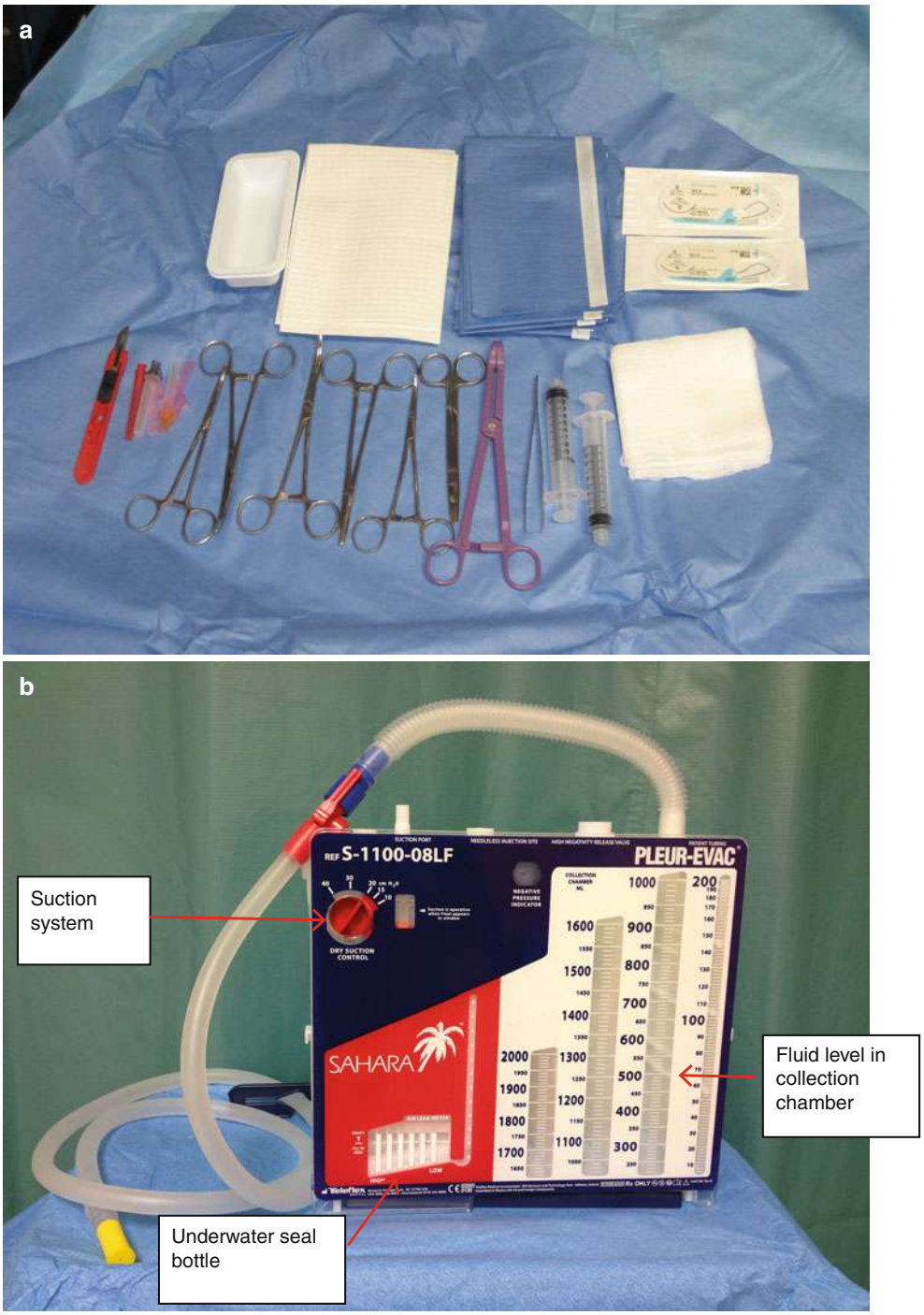

Fig. 23.1 (**a**) Tube thoracostomy tray, (**b**) chest drainage system

23.4 Procedure (Fig. 23.2)

1. Sterile skin preparation with sterile drape.
2. Anesthetize the appropriate area subcutaneously up to and including the rib periosteum with 5 mL of 1 % lidocaine with epinephrine (Fig. 23.2a).
3. Using a #10 or #11 blade, make an approximately 4-cm skin incision over the desired intercostal level of entry (most often the fourth or fifth intercostal space in the midaxillary line) (Fig. 23.2b, c).

   If the incision is placed *below* the fifth intercostal space, the risk of subdiaphragmatic placement into the abdominal space is increased.
4. Bluntly dissect with a hemostat or Kelly clamp through the subcutaneous tissue to the level of the intercostal muscles with intermittent opening of the dissection instrument during advancement (Fig. 23.2d, e).
5. Digitally palpate the selected intercostal space and the superior margin of the inferior rib (pay careful attention to avoid the neurovascular bundle lying inferiorly) (Fig. 23.2f).
   - If time permits, additional analgesia is recommended at this point of the procedure.
6. Guiding the closed Kelly clamp over the upper margin of the rib, enter the chest wall into the pleural cavity. (This will require some controlled force and a twisting motion.) Once the pleural space is entered, a rush of air or fluid should occur (Fig. 23.2g).

   Uncontrolled force and a lunging motion can result in penetration to the lung, heart, liver, or spleen.
7. Open the Kelly clamp while still inside the pleural space and then withdraw while the clamp is still open to enlarge the dissected tract of entry and allow easier passage of the thoracostomy tube (TT).
8. Explore the dissected tract with a sterile finger to appreciate lung tissue and possible adhesions.
9. To estimate the length the TT is to be inserted, measure the distance between the skin incision and the apex of the lung. If preferred, place a clamp over the tube at the estimated length (Fig. 23.2h).
10. Grasp the proximal end of the TT with the large Kelly clamp and pass the tube through the thoracic cavity along the previously dissected tract.
11. Release the Kelly clamp and continue to advance the tube posteriorly and superiorly.

    Make sure all of the fenestrated holes of the TT are within the thoracic cavity to prevent unnecessary manipulation and/or replacement of the TT.
12. Once the TT is in the desired position, connect the tube to the drainage device (Fig. 42.1b). Once connected, release the cross clamp on the distal end of the TT.
13. Secure the TT to the skin with 0 or 1–0 silk or nylon suture. A simple, interrupted suture above and below the TT with each stitch wrapped tightly around the TT is recommended.

    Incomplete security of the TT leads to dislodging of the tube with routine patient movements.
14. Apply petrolatum gauze over the skin closure surrounding the TT and then apply a support dressing with 4×4 gauze and adhesive tape (4 in.).
15. Obtain a chest radiograph to confirm placement of the TT.

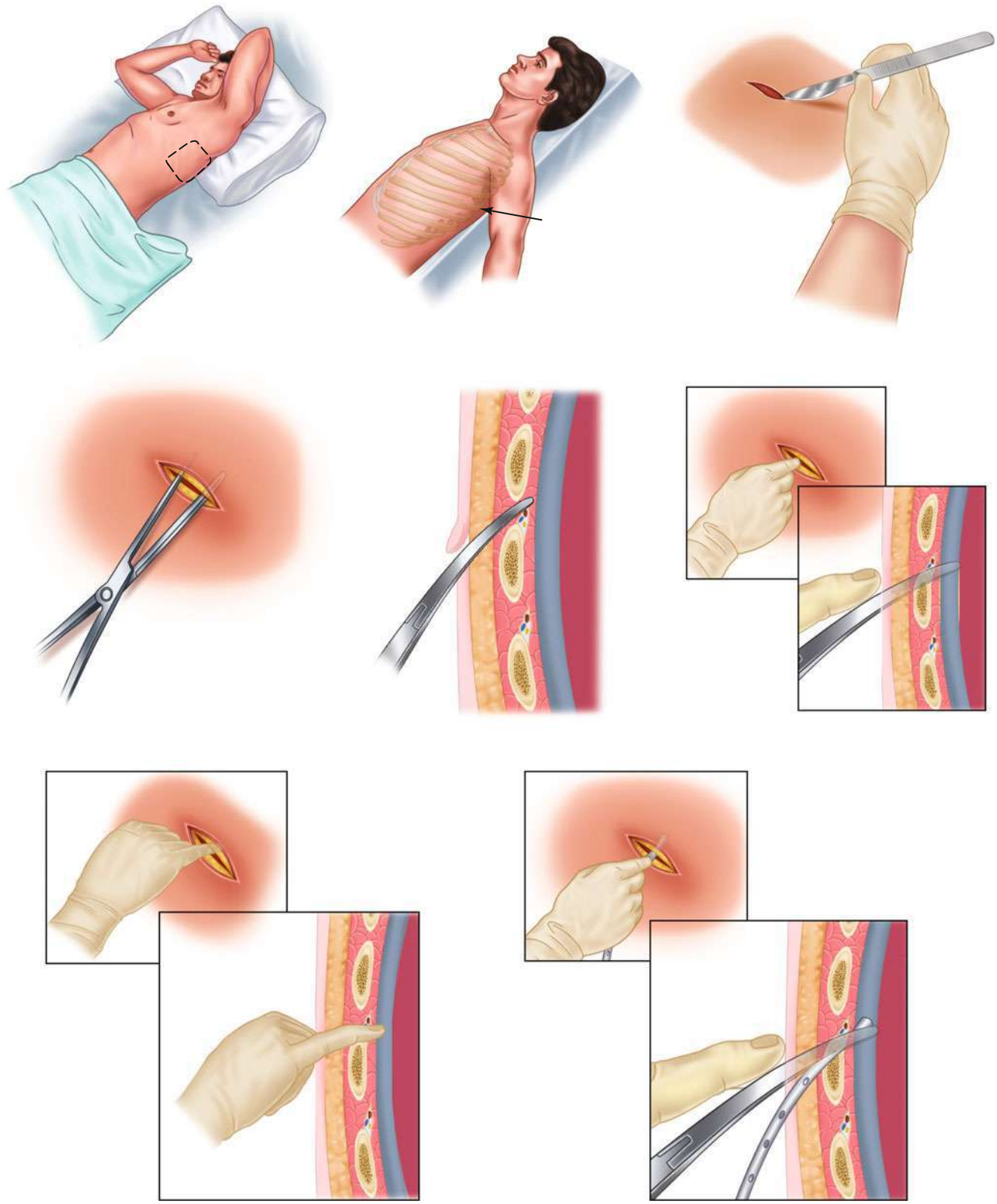

Fig. 23.2 Chest tube thoracostomy procedure

23.5 Complications

- Improper placement for pneumothorax
  - Reposition if:
    - Horizontal (over diaphragm)
    - Subcutaneous
    - Placed too far into the chest (against apical pleura)
  - Remove if:
    - Placed into the abdominal space
- Bleeding (local vs. hemothorax)
- Hemoperitoneum (liver or spleen injury)
- Tube dislodgment
- Empyema (TT introduces bacteria into the pleural space)
- Retained pneumothorax (may require second TT)
- Re-expansion pulmonary edema
- Subcutaneous emphysema

23.6 Pearls and Pitfalls

- Water seal acts as a one-way valve; if the system bubbles, there is an air leak.
- In the Pleur-evac® systems, there is an orange floater which, when static, means the desired suction pressure (usually 20 cmH_2O) has been reached.
- The negative pressure in the chest cavity equals the amount of water in water seal plus amount of suction.
- A chest tube can be removed when there is no air loss or blood for 24 h.
- When removing the tube, have the patient exhale and remove as quickly as possible.
- Leave petrolatum gauze in place for 48 h before changing it (allows wound to heal better).

Selected Reading

Ball CG, Lord J, Laupland KB, et al. Chest tube complications: how well are we training our residents? Can J Surg. 2007;50:450.

Collop NA, Kim S, Sahn SA. Analysis of tube thoracostomy performed by pulmonologists at a teaching hospital. Chest. 1997;112:709.

Dalbec DL, Krome RL. Thoracostomy. Emerg Med Clin North Am. 1986;4:441.

Miller KS, Sahn SA. Chest tubes. Indications, technique, management and complications. Chest. 1987;91:258–64.

Millikan JS, Moore EE, Steiner E, et al. Complications of tube thoracostomy for acute trauma. Am J Surg. 1980;140:738.

24 Thoracentesis

Lee Richard Donner and Michael Anana

24.1 Indications

- *Therapeutic thoracentesis* is performed to relieve dyspnea, hypoxia, or otherwise compromised respiratory function due to a large pleural effusion.
- *Diagnostic thoracentesis* is performed to aid in the diagnosis and workup of:
  - Pleural effusions of unknown cause
  - Unilateral pleural effusions
  - Pleural effusions originally determined to be due to heart failure but persisting after 3 days of diuresis

24.2 Contraindications

- Absolute
  - None
- Relative:
  - Coagulopathy, thrombocytopenia.
  - Small or loculated pleural effusion. These will increase the risk of missing the effusion and causing lung injury.
  - Positive-pressure ventilation.
  - Skin infection over the needle insertion site.

24.3 Materials and Medications

- Thoracentesis tray (commercially available kits generally include the items in the following list) (Fig. 24.1)
  - (1) Fenestrated drape
  - (1) 25-gauge × 1-in. needle
  - (1) 21-gauge × 1.5-in. needle
  - (1) 8-French catheter over 18-gauge needle
  - (1) Small plastic syringe, 5 mL
  - (1) Small plastic syringe, 10 mL
  - (1) Large plastic syringe, 50–60 mL
  - (1) Three-way stopcock
  - Specimen vials and caps
  - (1) Collection bag, 1500 mL, or vacuum container
  - (1) Tubing set
  - (1) Hemostat
  - Betadine (povidone-iodine) or other skin antiseptic preparing solution
  - 10-mL lidocaine 1 % without epinephrine

Fig. 24.1 Typical commercial thoracentesis tray

L.R. Donner, MD (✉)
Emergency Medicine Department, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: LeeDonnerMD@gmail.com

M. Anana, MD
Emergency Department, University Hospital, Rutgers New Jersey Medical School, Newark, NJ, USA
e-mail: ananami@njms.rutgers.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_24

24.4 Procedure

1. Place patient in seated position with arms at rest on a bedside table (Fig. 24.2).
2. The location and height of the pleural effusion are confirmed by physical examination. Auscultation of decreased or absent breath sounds, dullness to percussion, and decreased tactile fremitus are physical findings to confirm the location and height of the effusion (Fig. 24.3).
3. Determine and mark the site of needle insertion. This will be at the midscapular line one or two intercostal spaces below the maximum height of the effusion as determined by a combination of imaging and physical examination (Fig. 24.4).
   - *Do not* attempt thoracentesis inferior to the eighth intercostal space because respiratory cycle and anatomical variation place the diaphragm and intra-abdominal organs at risk.
4. Prepare the skin with the sterile skin preparation and sterile drape.
5. Anesthetize the appropriate area subcutaneously using a 25-gauge needle on a 10-mL syringe. Create a wheal in the skin and then infiltrate up to and including the periosteum of the rib inferior to the landmarked space with 5 mL of 1 % lidocaine.
   - Remember that the neurovascular bundle is found at the inferior border of each rib. Keep this in mind as you approach the rib and throughout the rest of the procedure to avoid injury to these structures (Fig. 24.5).
6. Using a 22-gauge needle, slowly walk the needle up and over the superior border of the rib. Continue to advance the needle along the superior border of the rib with the syringe withdrawn and infiltrating lidocaine intermittently along the way.
7. Pleural fluid will be aspirated once the pleural space is reached. Do not advance the needle further. Deposit 1–2 mL more lidocaine. A hemostat can be placed on the needle at the level of the skin to mark the depth of the pleural space and the needle can be removed.
8. Some commercial kits may come with an adjustable depth guard to be positioned at the determined depth. On an 18-gauge catheter-over-needle device, position the depth guard to the appropriate depth determined from the prior step. If a depth guard is not available, use the index finger and thumb on the catheter at the appropriate depth. With the 5-mL syringe attached, advance the device over the superior border of the rib while aspirating, expecting pleural fluid to return again at the determined depth (Fig. 24.6).
9. When the pleural space is reached, do not advance the needle further. Advance the catheter over the needle until the hub reaches the skin. Remove the needle during expiration and immediately cover the open hub with a gloved finger to prevent development of pneumothorax. Some kits provide catheters with one-way valves to prevent air entry.
10. Attach the 50–60-mL syringe to the catheter via the three-way stopcock. Pleural fluid can be drained and transferred to appropriate collection vials for diagnostic thoracentesis. A collection bag may be attached with tubing to the third port of a three-way stopcock for larger volume evacuation in the case of a therapeutic thoracentesis. Employ a syringe pump method to drain 50–60 mL of fluid at a time to the collection bag. Fill the syringe by withdrawing the plunger while the stopcock is closed to the bag. Then, close the stopcock to the patient and pump the contents of the syringe to the bag. Next, close the syringe closed to the bag, and repeat the cycle until the desired volume is drained. A vacuum container is an alternative that simply attaches via tubing to the stopcock (Fig. 24.7).
    - If using a three-way stopcock and a device that does not have a one-way valve on the catheter, be sure to always keep the stopcock closed to the patient unless withdrawing fluid in order to decrease the risk of pneumothorax.
11. When the desired amount of pleural fluid is obtained, remove the catheter during expiration and apply an occlusive dressing.

Fig. 24.2 Patient in upright, seated position

Fig. 24.3 Determining the location and height of the pleural effusion

Fig. 24.4 Marked site for needle insertion

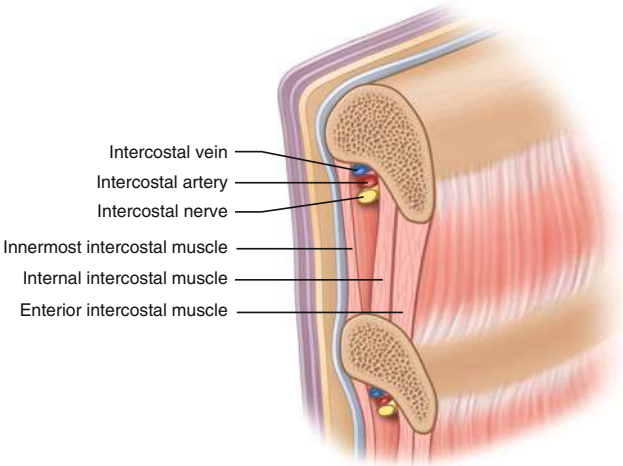

Fig. 24.5 The intercostal neurovascular bundle

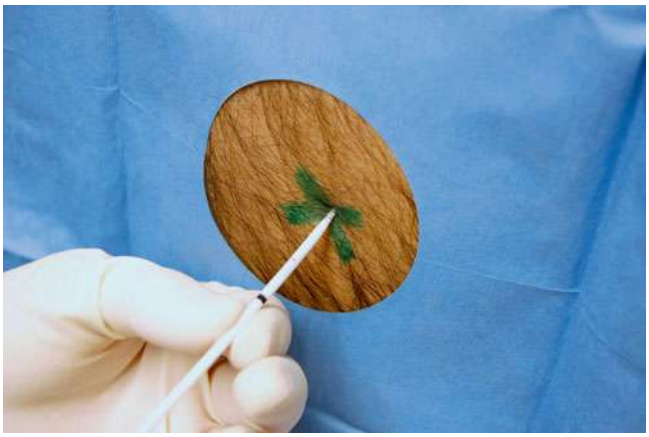

Fig. 24.6 Needle and catheter insertion, thumb and index finger at desired depth

Fig. 24.7 Pleural fluid collection via syringe pump method

24.5 Complications

- Pneumothorax
- Re-expansion pulmonary edema
- Hemothorax, hematoma
- Intra-abdominal organ injury
- Air embolism
- Empyema

24.6 Considerations

- If available, the use of bedside ultrasound is highly recommended because ultrasound guidance has been shown to substantially reduce the risk of pneumothorax. Before the procedure, the height, width, and depth of the effusion can be appreciated by scanning the chest and viewing the effusion through the intercostal spaces. The use of ultrasound aids in selecting the needle insertion site by:
  - Visualizing the distance the needle must pass to reach the parietal pleura
  - Confirming the thickness of the effusion in the site selected is at least a minimum of 1.5 cm
  - Providing the clinician with a view of the effusion and surrounding structures through the complete respiratory cycle

  With these items in mind, the needle insertion site can be selected with confidence and marked before beginning the procedure.
- In addition, the use of bedside ultrasound in real time will allow the clinician to visualize the needle as it passes toward and enters the pleural space. This use requires sterile probe covers.
- Re-expansion pulmonary edema is a rare but feared complication of thoracentesis. The cause is not fully understood. Historically it was thought that re-expansion pulmonary edema was caused by removing too large a volume of fluid from the pleural space (>1–1.5 L). Another theory is that re-expansion pulmonary edema is caused when great negative intrapleural pressures (<20 cmH_2O) are generated during the procedure. The low incidence of this complication has yielded inconclusive evidence.
  - In light of this, it is prudent to continue to limit the volume of pleural fluid removed to no more than 1–1.5 L. Pleural manometry is not widely available for use in the emergency department, but should also be considered if available to maintain intrapleural pressures from reaching more negative values.

24.7 Pearls and Pitfalls

- Bedside ultrasound reduces the risk of complications.
- Never attempt thoracentesis below the eighth intercostal space.
- Thoracentesis should not be performed on pleural effusions demonstrated to be less than 1–1.5 cm thick by ultrasound or on a lateral decubitus film.
- Positive-pressure ventilation mandates extreme care while performing thoracentesis because the lungs can be punctured during inflation. In addition, ventilated patients will not be able to sit upright and will need to have the procedure performed in lateral decubitus position with the effusion side down and the posterior axillary line as the needle insertion site. Alternatively, the patient can be placed supine with the head of the bed elevated to 45° and the midaxillary line as the needle insertion site.
  - Routine post-procedure radiographs are not necessary to exclude pneumothorax. Indications for post-procedure imaging include onset of chest pain during the procedure, persisting cough or chest discomfort after the procedure, air aspiration along any step of the procedure, or positive-pressure ventilation.
  - Pneumothorax rates are higher for inexperienced providers. Although an effusion compromising respiratory function is considered a clear indication for therapeutic thoracentesis, the performance of nonurgent diagnostic thoracentesis might best be delayed for those more practiced in the procedure.

Selected Reading

Dewitz A, Jones R, Goldstein J. Additional ultrasound-guided procedures. In: Ma OJ, Mateer JR, Blavias M, editors. Emergency ultrasound. 2nd ed. New York: McGraw-Hill; 2008. p. 546–50.

Feller-Kopman D, Berkowitz D, Boiselle P, Ernst A. Large-volume thoracentesis and the risk of reexpansion pulmonary edema. Ann Thorac Surg. 2007;84:1656–61.

Gordon CE, Feller-Kopman D, Balk EM, Smetana GW. Pneumothorax following thoracentesis: a systematic review and meta-analysis. Arch Intern Med. 2010;170:332–9.

Light RW. Pleural effusion. N Engl J Med. 2002;346:1971–7.

Thomsen TW, DeLaPena J, Setnik G. Videos in clinical medicine. Thoracentesis. N Engl J Med. 2006;355, e16.

25 Open Chest Wounds and Flail Chest

Jacob J. Glaser and Carlos J. Rodriguez

25.1 Background

Thoracic injuries are commonly associated with penetrating and blunt abdominal trauma and are implicated in 50–70 % of trauma deaths [1]. Cardiac tamponade, tension pneumothorax, massive hemothorax, airway obstruction, flail chest, and open pneumothorax represent the six immediately life-threatening injuries attributed to chest trauma [2]. Accordingly, they must be accurately identified and dealt with urgently.

Open pneumothorax ("sucking" chest wound) is seen in penetrating chest injuries. If the associated chest wound is greater than 2/3 the diameter of the trachea (generally anything greater than 1.5–2 cm), air can preferentially enter the intrapleural space, via the trachea, with each inspiration [3] (Fig. 25.1). This allows equilibration of pressure between the pleural space and the atmosphere, causing the lung to collapse and leading to profound hypoventilation and hypoxia.

Flail chest results from high-energy blunt, crushing chest trauma causing two or more fractures in two or more contiguous ribs. Classically, the fractures are lateral or sternal. Posterior rib fractures rarely cause flail physiology (Fig. 25.2). Flail chest has been reported to have mortality as high as 16 % [4]. This injury pattern is associated with a high incidence of underlying pneumothorax, hemothorax, pulmonary contusion, and chest wall instability. Mortality from flail chest is thought to be correlated with the degree of underlying pulmonary contusion and attendant hypoxia [2].

Both open pneumothorax and flail chest are immediately life threatening and require early appropriate management.

J.J. Glaser, MD (✉)
Combat Casualty Care Directorate, Naval Medical Research Unit, San Antonio, TX, USA
e-mail: Jacob.glaser1@gmail.com

C.J. Rodriguez, DO
Division of Trauma Surgery, Surgical Critical Care, Walter Reed National Military Medical Center, Bethesda, MD, USA
e-mail: Carlos.J.Rodriguez.mil@health.mil

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_25

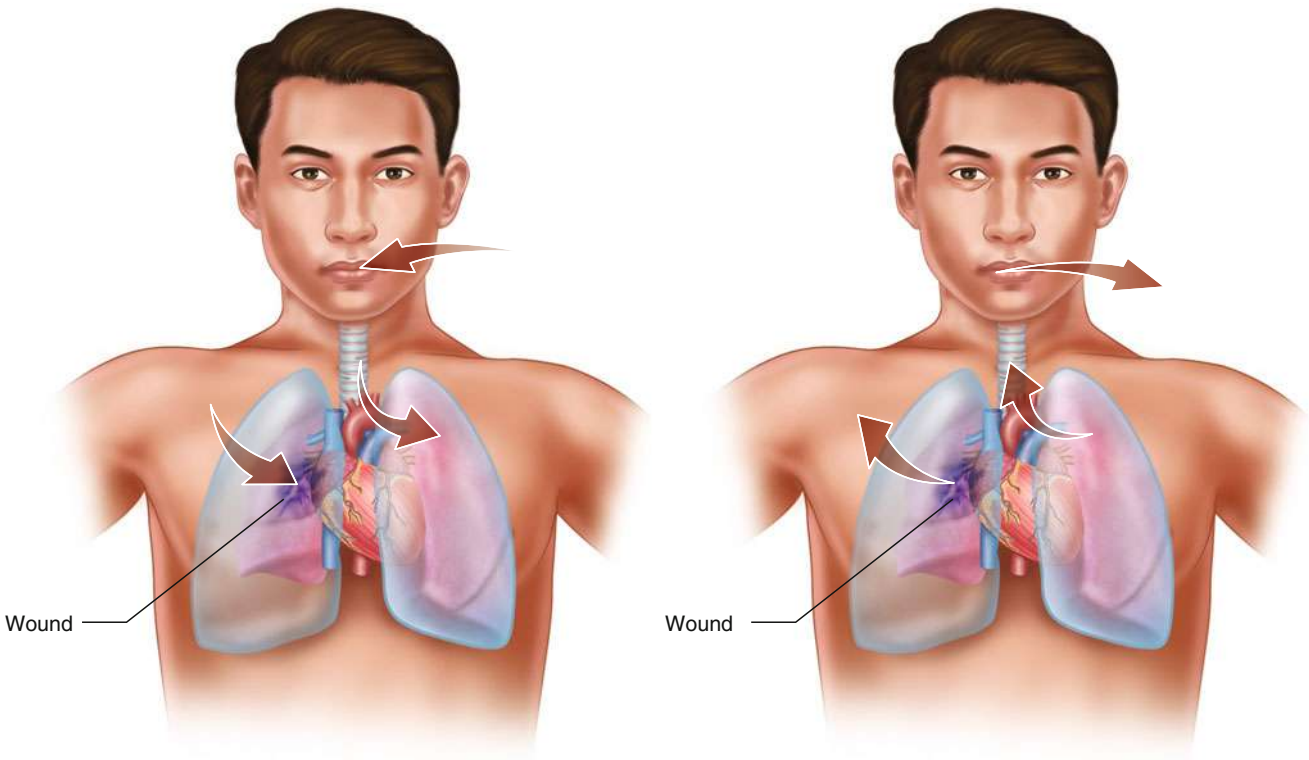

Fig. 25.1 Air preferentially will enter the chest via the wound, collapsing the lung on the affected side

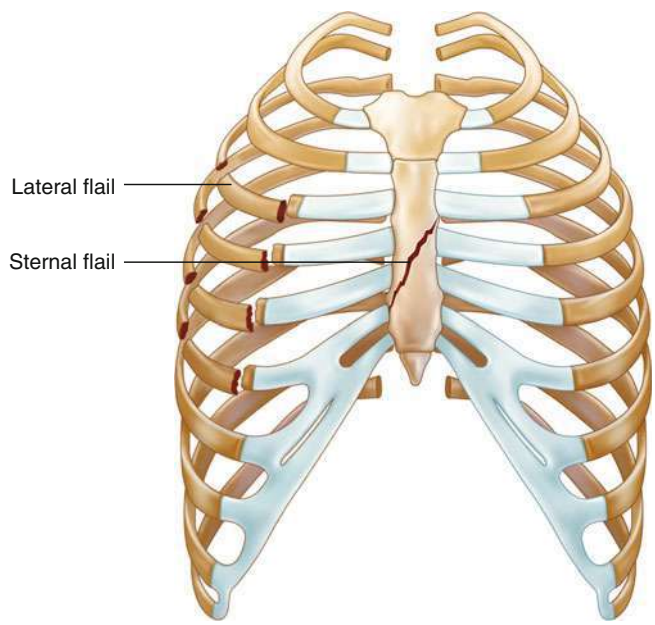

Fig. 25.2 Flail chest: two or more rib fractures in two or more segments. Lateral flail (most common) and sternal flail segments represented below

25.2 Initial Evaluation of Thoracic Trauma

- To the best extent possible, obtain good patient handover from the prehospital caregivers.
- Initial physical evaluation.
  - Appropriate attention should be given to the ABCs (airway, breathing, circulation) of ATLS (Advanced Trauma Life Support) management.
  - Evaluation and resuscitation are to be *concurrent* with diagnostic procedures and immediate interventions.
  - Maintain a high degree of suspicion for open chest wound in impalement injury and destructive penetrating trauma (blast injury or shotgun).
  - Maintain a high degree of suspicion for flail chest in high-energy direct impact trauma (motor vehicle crash, fall, crush injury).
- Administer high-flow O_2 with a non-rebreather mask.
  - If patient is in respiratory distress, is unstable, or has an obvious chest wall defect, consider early intubation to secure the airway.
  - Inspect the chest wall for occlusive dressings.
- Decompress the chest if tension physiology is present or suspected.
  - Immediate decompression of a suspected tension pneumothorax can be readily accomplished by removing any existing occlusive dressings *or*
  - Place a large-bore cannula over the rib, second intercostal space, or midclavicular line.
- Specific immediate management appropriate to open chest wound and flail chest (see later).
- Monitor continuous pulse oximetry and electrocardiogram.
- Initiate crystalloid resuscitation via large-bore intravenous (IV) access.
  - Early mobilization of blood products if ongoing hemorrhage or expectation of excessive blood loss.
  - Placement of resuscitative lines concurrent with management of respiratory parameters.
- Early surgical consultation for management of intrathoracic injuries and management of chest wall defect.

25.3 Open Pneumothorax ("Sucking Chest Wound")

- Immediate management requires attention to airway and respirations.
  - If in respiratory distress, *intubate*.
- Close the chest wall defect with an occlusive three-sided dressing.
  - This includes a valve mechanism that allows trapped air to escape, preventing tension (Fig. 25.3).
  - An IV bag cut to fit the wound and then taped on three sides can be useful in an emergent situation.
  - Commercial products are available and appropriate for smaller wounds, including the Asherman Chest Seal and HyFin Vent.
- A completely occlusive dressing may quickly convert an open chest wound into a tension pneumothorax [2, 3, 5] and therefore should *never* be done.
  - The patient and dressing must be serially checked to ensure that trapped air is allowed to escape.
  - If there is any doubt, immediately remove the dressing and replace it with an appropriate dressing.
- These maneuvers are a bridge to definitive care.
- When the timing is appropriate (i.e., time and resources are available), perform a formal tube thoracostomy and convert to a completely occlusive dressing over the wound.
  - Avoid placing the tube through the open wound.
- Once tube thoracostomy, placement of occlusive dressing, and the airway are secured, the pathophysiology of the open pneumothorax becomes inconsequential.
- Immediate consultation with surgery for definitive care of associated intrathoracic injuries is required.
  - The patient may need urgent thoracotomy to treat associated injuries.
  - Irrigation and debridement should take place in the operating room.
  - Depending on injury severity, the patient may need chest wall reconstruction.

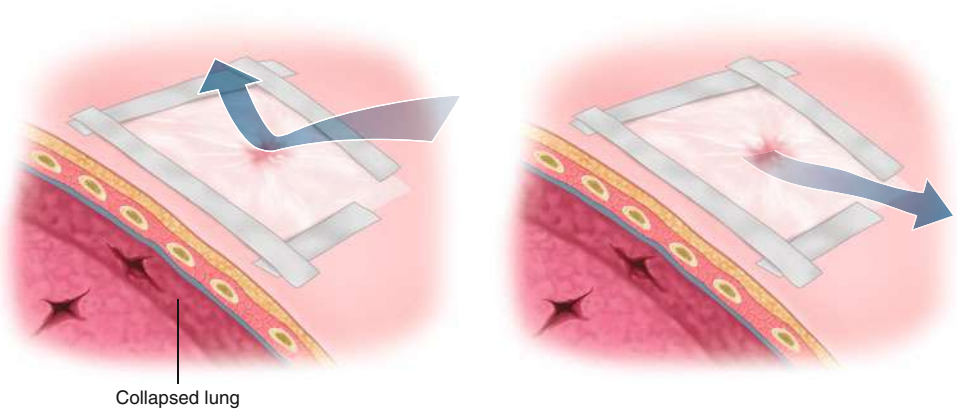

Fig. 25.3 Three-sided dressing, allowing a valve to decompress tension in the chest

25.4 Flail Chest

- Diagnosis is made from mechanism and examination, not radiographically.
  - With inspiration, the affected chest will move inward with negative pressure.
  - With expiration, the affected chest will move outward.
  - Patients who are intubated on positive pressure (and not spontaneously breathing) will often not show this paradoxical chest movement.
- Attention *must* be paid to presumed underlying blunt lung injury and contusion.
  - The degree of underlying contusion (not the flail segment itself) is directly related to the degree of hypoxia and associated morbidity and mortality [2, 3].
  - There should be a low threshold for intubation to manage respiratory distress, hypoxemia, or hemodynamic instability.
  - The patient is at high risk for hemothorax and pneumothorax requiring tube thoracostomy.
- Abdominal injuries may be present in up to 15 % of patients [2].
- After initial stabilization, treatment is supportive.
  - Intensive care unit admission for management of underlying pulmonary contusion.
  - Pain control in the form of epidural or regional block for excellent pain control.
  - Close attention to pulmonary toilet and lung re-expansion.
  - Patients require observation and treatment in a monitored setting until ensured that respiratory parameters and oxygenation are improving.
- Surgical stabilization of the chest wall is rarely performed.
  - Early surgical consultation for chest wall fixation in questionable cases is warranted.

References

1. LoCicero J, Mattox KL. Epidemiology of chest trauma. Surg Clin North Am. 1989;69:15–9.
2. Pietzman AB, Schwab CW, Yealy DM, editors. The trauma manual. 2nd ed. Philadelphia: Wolters Kluwer Health; 2000.
3. Weinberg JA, Croce MA. Chapter 33: Chest wall injury. In: Flint L, Meredith JW, Schwab CW, editors. Trauma: contemporary principles and therapy. 1st ed. Philadelphia: Lippincott Williams & Williams; 2007.
4. Clark GC, Schecter WP, Trunkey DD. Variables affecting outcome in blunt chest trauma: flail chest vs. pulmonary contusion. J Trauma. 1988;28:298–304.
5. Borden Institute Walter Reed Army Medical Center. Emergency war surgery. 3rd ed. Washington: Office of the Surgeon General U.S. Army, Borden Institute; 2004.

26 Emergent Resuscitative Thoracotomy, Open Cardiac Massage, and Aortic Occlusion

Kevin M. Jones and Jay Menaker

26.1 Indications

- Penetrating chest trauma with recent or immanent loss of vital signs
- Consider in blunt trauma with pericardial tamponade or exsanguination where aortic occlusion may provide proximal control

26.2 Contraindications

Absolute
- Prolonged cessation of vital signs
- Injury profile obviously incompatible with life
- Absence of surgical services to whom care can be transferred

Relative
- None

26.3 Materials and Medications (Fig. 26.1)

- Betadine (povidone-iodine) for rapid skin preparation
- #10 Scalpel
- Mayo or long Metzenbaum scissors
- Finochietto retractor (rib spreader)
- Long DeBakey or other tissue forceps (2)
- Satinsky vascular clamp and/or straight vascular clamp
- Long needle holders (2)
- Lebsche knife or sternal osteotome with hammer
- Lap sponges or gauze pads

K.M. Jones, MD (✉)
Trauma and Surgical Critical Care, R Adams Cowley Shock Trauma Center, University of Maryland Medical Center, Baltimore, MD, USA
e-mail: KJones12@umm.edu

J. Menaker, MD
Departments of Surgery and Emergency Medicine, R Adams Cowley Shock Trauma Center, University of Maryland Medical Center, Baltimore, MD, USA
e-mail: jmenaker@umm.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_26

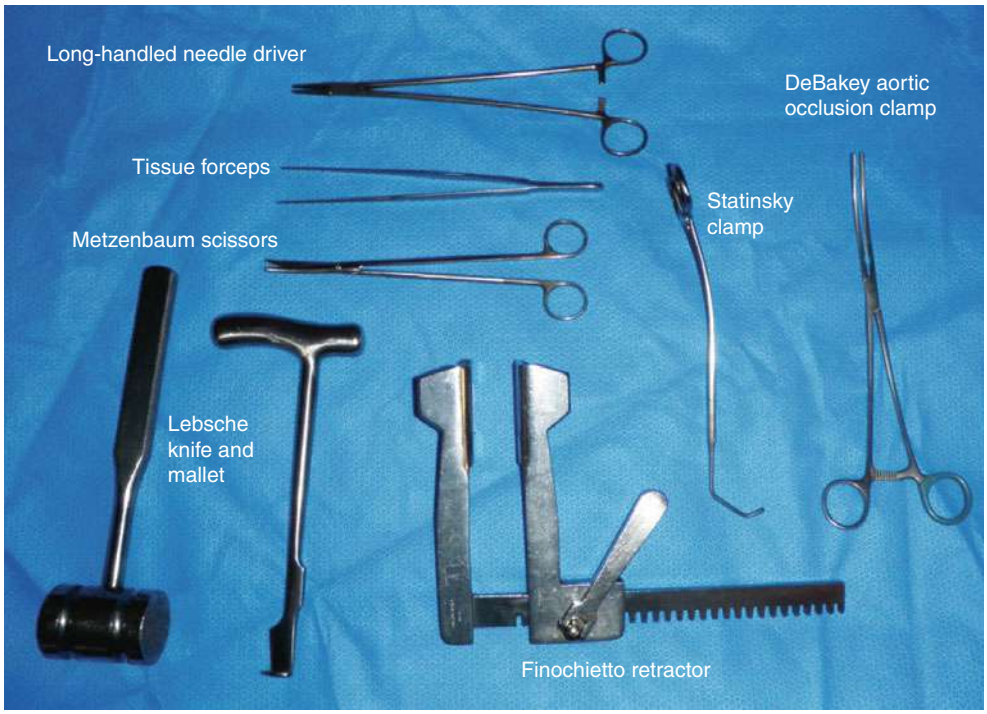

Fig. 26.1 Basic components of an emergency thoracotomy tray: Lebsche knife and mallet for crossing sternum (bone-cutting forceps or sternal osteotome would also suffice), Finochietto retractor, atraumatic vascular clamps (a Satinsky clamp and a DeBakey aortic occlusion clamp), long-handled needle driver, tissue forceps, and Metzenbaum scissors. *Not illustrated* Scalpel with #10 or #20 blade, Mayo scissors

26.4 Procedure

26.4.1 Resuscitative Thoracotomy and Open Cardiac Massage

1. Rapidly prepare the entire anterior and bilateral chest with Betadine.
2. Using a scalpel, incise the skin and subcutaneous tissue from just right of the sternum to the anterior margin of the left latissimus dorsi, following the curvature of the infra-mammary crease or the fourth or fifth intercostal space (Fig. 26.2).
   - This incision is often made too low on the patient's chest. It should extend across the sternum, not at the level of the xiphoid. Upward retraction of the breast may help provide access to the fourth or fifth intercostal space, where the incision should be located.
3. Bluntly enter the right pleural space through the fourth or fifth intercostal space.
4. Using scissors, cut the intercostal muscles, dividing between the fourth and the fifth ribs from the sternum to the posterior axillary line.
5. If better exposure to the heart is desired, some practitioners advocate extending the incision across the sternum using a Lebsche knife, sternal osteotome, or bone-cutting forceps at this time (Fig. 26.3).
   - If the thoracotomy incision extends to or through the sternum, tie off the internal mammary arteries before closing the chest should the patient be successfully resuscitated, because these will have been divided with the incision.
6. Insert a Finochietto retractor and retract the ribs in order to gain access to the left chest and expose the pericardium (Figs. 26.4 and 26.5).
   - Insert a Finochietto retractor with the rack and pinion bar down and lateral, as in Fig. 26.4, so as not to interfere with extension of the thoracotomy across the sternum into a clamshell maneuver if needed.
7. If massive left pleural hemorrhage is encountered, investigate and control the source at this time.
8. Using tissue forceps, raise a portion of the pericardium anterior to the phrenic nerve, and enter the pericardium using scissors (Fig. 26.6).
9. Widely open the pericardium with scissors, cutting in a cranial-caudal direction anterior to the phrenic nerve and deliver the heart (Fig. 26.7).
10. If hemopericardium is encountered, investigate for and initiate appropriate repair of identified cardiac injuries.
11. Initiate open cardiac massage by cupping the heart between the flattened palmar aspect of the fingers of both hands and rhythmically compressing the heart from apex to base, relaxing completely between compressions to allow filling (Fig. 26.8).

26.4.2 Aortic Occlusion

1. Expose the posterior aspect of the left mediastinum by having an assistant retract the left lung superomedially, dividing the inferior pulmonary ligament if necessary (Fig. 26.9).
2. Bluntly dissect the pleura separating the pleural and mediastinal space just anterior to the vertebral bodies, exposing the aorta (Fig. 26.9).
3. Completely encircle the aorta with the finger of the non-dominant hand (Fig. 26.10).
   - Differentiating the aorta from the esophagus when the patient is in a state of profound shock is very difficult. Having an assistant pass an orogastric tube may help distinguish the two. The aorta should be the most posterior structure, lying immediately on the anterior aspect of the vertebral bodies.
4. With the aorta completely encircled, place a vascular clamp across the aorta and verify by sight and feel that the complete vessel is occluded within the clamp (Fig. 26.11).

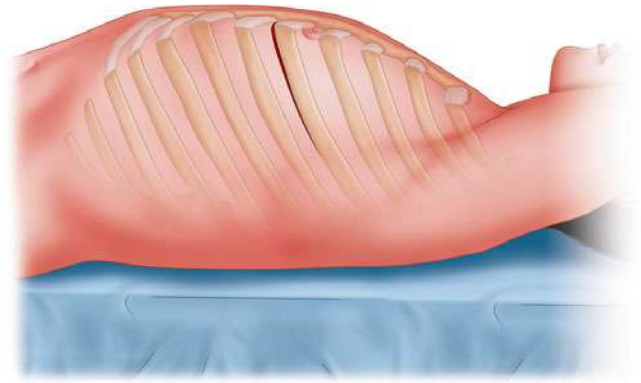

Fig. 26.2 Raise the left arm above the head. Make an incision along the left fourth intercostal space, just below the nipple in a male or at the inframamillary crease in a female

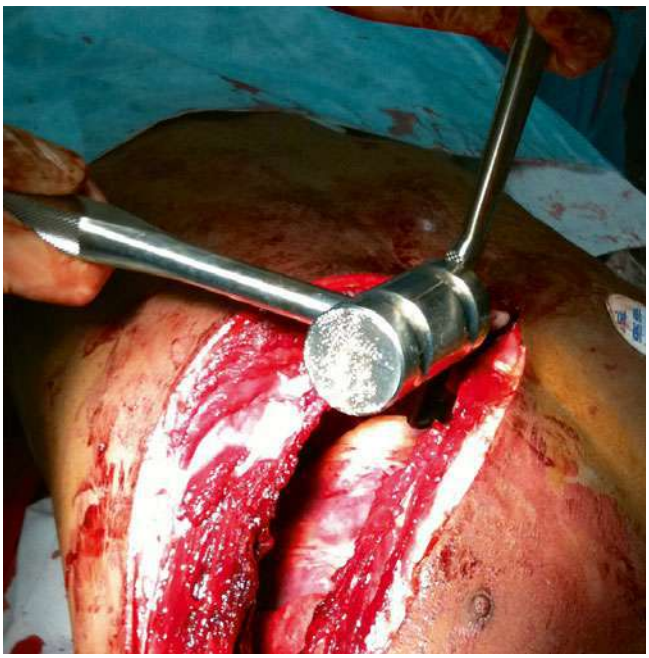

Fig. 26.3 Use a Lebsche knife to extend the incision across the sternum to improve exposure to the heart

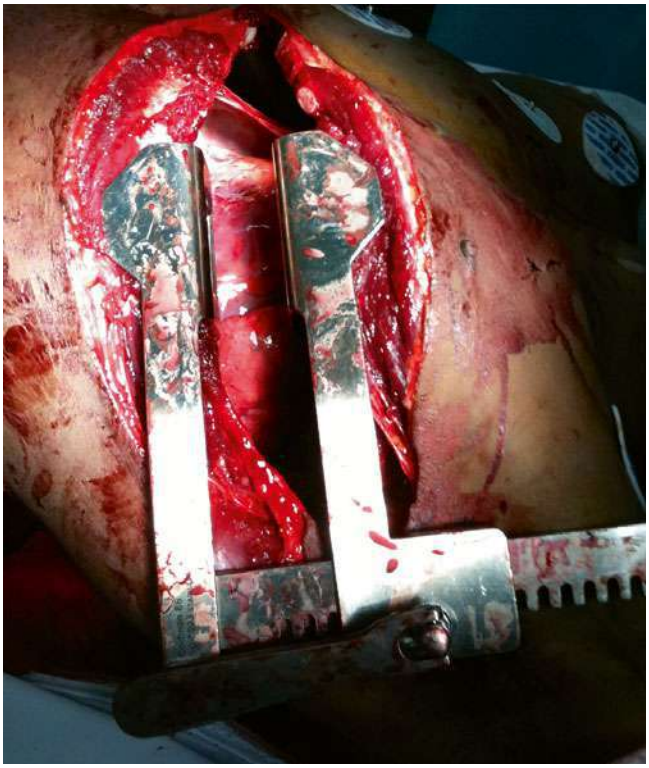

Fig. 26.4 Finochietto retractor placed through the fourth intercostal anterolateral incision. Note the rack and pinion bar placed posterolaterally, where it will not impede access to the midline

Fig. 26.5 Finochietto retractor extended, exposing the left pleural space and the pericardium. In this case the sternum has been divided as above, and the resultant window into the right pleural space is seen anterior to the pericardium. The lung is deflated in this postmortem picture but would be far more an obstacle in the actively ventilated patient. The clamp seen in the upper portion of the picture is reapproximating the pericardium, which has previously been divided, for the sake of illustration

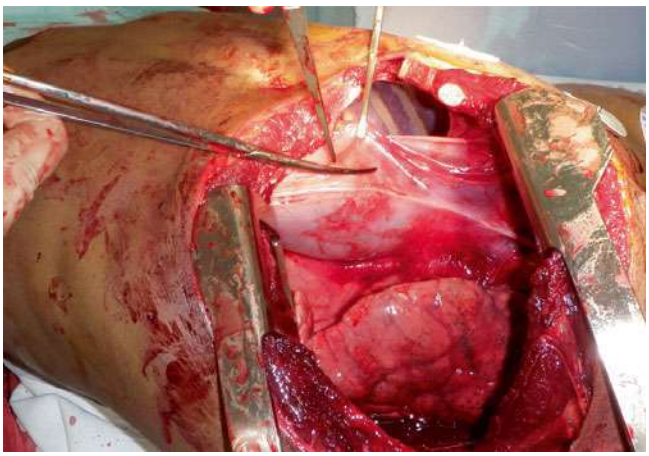

Fig. 26.6 The pericardium is lifted with tissue forceps and opened with a nick using scissors. The phrenic nerve is easily visualized running cranial-caudal just below the scissors in this picture

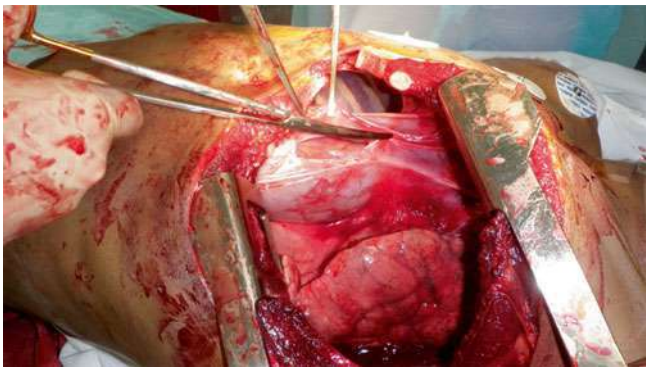

Fig. 26.7 Open the pericardium widely in the cephalad-caudad plane anterior to the phrenic nerve, taking care not to damage the phrenic nerve

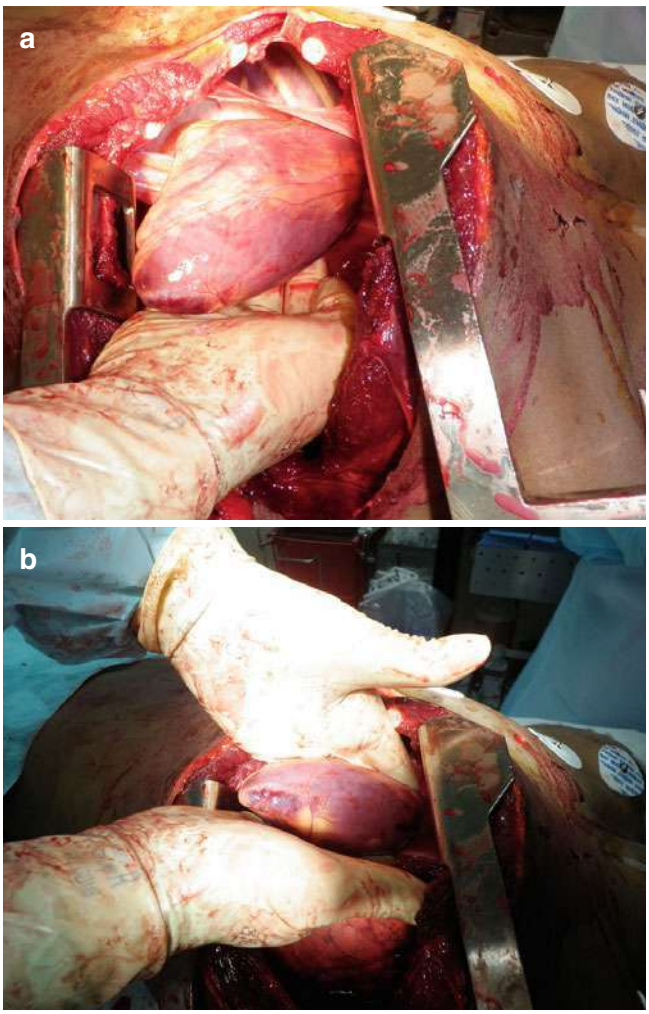

Fig. 26.8 (**a**) Deliver the heart from the pericardium, rapidly assess for cardiac injury requiring damage-control repair; (**b**) initiate open cardiac massage

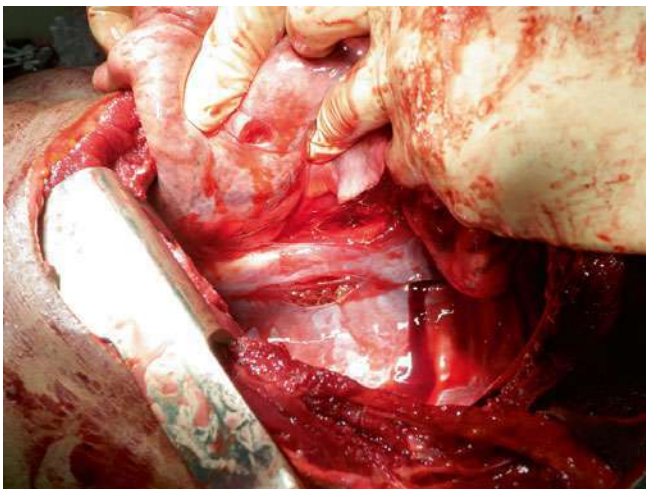

Fig. 26.9 The heart and lung are retracted superomedially, allowing visualization of the left posterior mediastinal pleura; the aorta lies just anterior to the vertebral bodies and has been previously isolated through the pleural interruptions seen here. The heart and lung are assertively retracted here for the benefit of illustration of the posterior mediastinum. Such retraction would completely occlude venous return

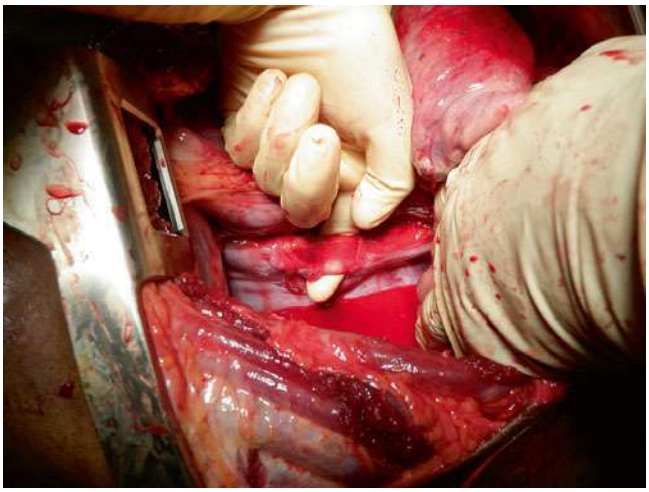

Fig. 26.10 After bluntly dissecting the mediastinal pleura, the aorta is looped using a finger of the nondominant hand

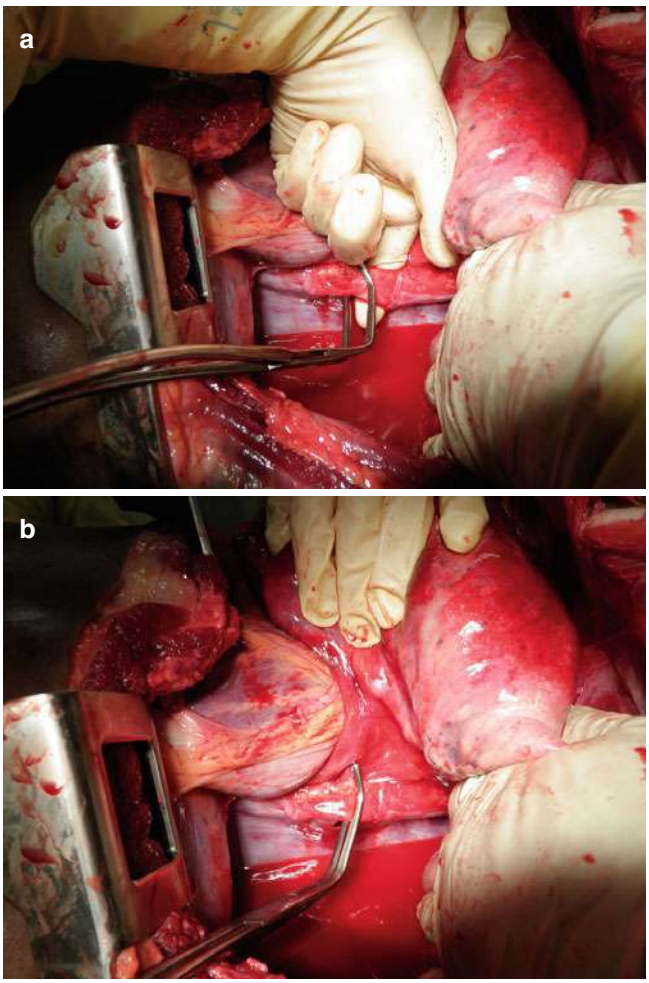

Fig. 26.11 (**a**) A vascular clamp (a Satinsky clamp is used here, although any large atraumatic vascular clamp can be used) is applied across the descending aorta, (**b**) followed by visual and tactile confirmation that the aorta is completely occluded

26.5 Complications

- Injury to care providers, by means of scalpel, needlestick, or sharp foreign body, is the principal concern.
- Post-emergency department thoracotomy infections are rare, even given the less than optimal sterile conditions.
- Damage to lung parenchyma during the initial incision is common and often leads to air leak in survivors.
- Neglect of the mammary arteries, often divided during emergent thoracotomy and not briskly bleeding in the shock state, will result in intrathoracic hemorrhage if not tied off.

Selected Reading

Bartlett RL. Resuscitative thoracotomy. In: Roberts JR, Hedges JR, editors. Clinical procedures in emergency medicine. 3rd ed. Philadelphia: WB Saunders; 1998.

Feliciano DV, Mattox KL. Indications, technique, and pitfalls of emergency center thoracotomy. Surg Rounds. 1981;4:32.

Siemans R, Polk Jr MC. Indications for thoracotomy following penetrating thoracic injury. J Trauma. 1977;17:493.

Wall Jr MJ, Huh J, Mattox KL. Indications and techniques of thoracotomy. In: Feliciano DV, Mattox KL, Moore EE, editors. Trauma. 6th ed. New York: McGraw Hill; 2008.

27 Lung Ultrasonography

Ali H. Dabaja, L. Connor Nickels, and Rohit Pravin Patel

27.1 Indications

- Evaluation of respiratory failure and insufficiency due to pneumothorax, pleural effusion, pulmonary edema, acute respiratory distress syndrome (ARDS), and alveolar consolidation (atelectasis, pneumonia, aspiration)
- Monitoring progress of diseases such as pulmonary edema and pneumothorax
- Procedural guidance during pleural fluid removal or pneumothorax treatment
- Procedural guidance for chest tube placement for complex pleural effusions, hemothorax, pneumothorax, and other pleural diseases

27.2 Contraindications

- None

27.3 Materials

- Ultrasound equipment/machine, ultrasonography gel
- Microconvex or phased-array cardiac probe for sufficient evaluation of lung artifact (use abdominal preset). Linear-array probes, or vascular probes, can be used for more detailed evaluation of the pleura, although this is insufficient for penetration to deeper structures.
- Sterile materials and equipment where appropriate

27.4 Preparations

- Clean ultrasound equipment and use sterile operating procedures and ultrasound conductive gel whenever adequate.
- Ultrasound positioning. Place ultrasound device where it is easily accessible and in adequate view.
- Proper environment. Minimize light interference. Close shades and lights where appropriate.
- Position the patient. This helps optimize and expedite the examination. Supine position should be adequate for critically ill patients. Upright positioning is best for all others, with arms abducted above the head or spaced away from the chest to allow space for the probe. Patients should be moved to the edge of the bed. Uncover appropriate areas, including anterior and lateral chest walls.
- Equipment use. Turn on equipment, enter patient data, and select proper probe (cardiac probe or abdominal probe with abdomen presets). Depth and gain should be adjusted accordingly, where depths in the range of 4–10 cm give proper evaluation of more superficial and deeper structures, respectively.

A.H. Dabaja, DO
Division of Critical Care, Department of Anesthesia, University of Florida Health, Gainesville, FL, USA
e-mail: ali.dabaja@gmail.com

L.C. Nickels, MD, RDMS (✉) • R.P. Patel, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; rohitpatel@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_27

27.5 Procedure

1. Probe position (Fig. 27.1). Start examination at the mid-clavicular level at the space between the second and the third ribs. Probe positioning should be perpendicular to the ribs (longitudinal positioning) with the ultrasound marker pointed cephalad. This should place the most superficial structures at the top, with deeper structures at the bottom of the monitor. Upon completion, probe positioning can be mapped to evaluate three or four additional areas, typically between the anterior and the posterior axillary lines. Lateral views with the probe should be most posterior, typically along the posterior axillary line, tracking caudad toward the diaphragm. PLAPS (posterolateral alveolar and/or pleural syndrome) pointed posteriorly should also be evaluated, specifically in supine patients. PLAPS is lateral to the scapula and typically requires that the patient be lifted off the bed from one side.
2. Identify the "bat sign" (Fig. 27.2). The initial view observed should be a window of lung flanked by two rib shadows. This view, termed the "bat sign," should now allow evaluation of the parietal and visceral pleura, seen most superficial as echogenic line (approximately 0.5 cm below start of rib shadows), subsequent lung sliding, and other findings such as "A" and "B" lines, as well as abnormal lung tissue.
3. Identify "A lines" (Fig. 27.3). A lines indicate air. These are multiple echogenic lines appearing horizontally in sequence deep to the pleural line. This artifact represents reverberations of the pleura and can be found in aerated lungs, which can be normal or abnormal (e.g., pulmonary embolism, chronic obstructive pulmonary disease [COPD]). The first true A line, denoted "A1," is found equidistant from the chest wall to the pleural line. Many other A lines might be seen and are denoted "A'" lines. Subsequent equidistant A lines are "A2," "A3," and so on.
4. Identify "B lines" (Fig. 27.4). These artifacts appear in well-aerated lung and are vertical echogenic lines (ray, flashlight, lung rockets) transmitted from the pleura to the deeper parts of the lung on the ultrasound monitor field. They are due to thickened interlobular septa and represent alveolar fluid surrounded by air. True B lines arise from the pleural line and shoot all the way down to the far lung fields, whereas "comet tails" are seen only close to the pleural and are sometimes referred to as "shimmering" or "glimmering" during movement of the pleural line. When multiple B lines are seen in a patient, it is sometimes referred to as "lung rockets" or "flashlights" because many rays are shooting from the pleura. Even though most of the time B lines represent pulmonary edema, they can be seen in other conditions such as aspiration, pulmonary fibrosis, acute respiratory distress syndrome (ARDS), and pneumonia.
5. Identify lung sliding (Fig. 27.5). Lung sliding identifies movement of a normal parietal-visceral interface. Patient breathing causes a rhythmic movement of parietal against the visceral pleura appearing as movement of the hyperechoic line. M-mode can be used to show a timed clip of this through a still image and should only be used as a method of reporting or saving for documentation purposes. A "lung point" is an area adjacent to lung sliding (parietal pleura) that is devoid of movement. This is highly specific for pneumothorax. If one suspects a pneumothorax, quantification of the size can be made by evaluation at the anterior lung points and movement toward the lateral wall of the chest. The more rib spaces found to have absent lung sliding, the larger the pneumothorax. Lung sliding (Fig. 27.5) can be evaluated with M-mode, which can help identify a normal parietal-visceral interface at that level. Obtain an adequate two-dimensional view ("bat sign") and press the "M-mode" option on the equipment. A normal interface appears as multiple hyperechoic lines, the pleura (termed "seashore"; Fig. 27.5a), followed by a sandlike pattern, the lung tissue. This pattern together is termed a "seashore sign." Air that disrupts the parietal-visceral interface, as found within a pneumothorax, is identified as horizontal repeating echogenic lines, similar to a barcode (Fig. 27.5b), and is termed "stratosphere sign." M-mode can be discontinued.
   - Lung sliding can be absent in conditions other than pneumothorax: apnea, right or left bronchial intubation, lung collapse (blebs), pneumonia, and pulmonary fibrosis.
6. Identify lung pulse. This appears as a shimmering of the pleural line due to cardiac activity. This is most apparent on the left side of the chest, closest to the heart. This helps to exclude pneumothorax as well.
7. Move posteriorly. Move the ultrasound probe laterally and posteriorly to the PLAPS point. The transducer can be directed toward the center of the patient's body in supine patients. Pleural effusions and consolidations are found in the dependent areas of the lung.
8. Move caudally. With the marker still pointing cephalad, move along the posterior axillary line in two or three additional rib spaces. Identification of pleural disease and other pathology requires multiple views and will aid in evaluating the extent of the disease. This will also allow for identification of boundaries of the lung, such as the diaphragm. Identification of the diaphragm is most critical to determine location of fluid.

9. Identify the diaphragm and liver and/or spleen (Fig. 27.6). Along the posterior axillary line, or the posterior chest wall, move the probe caudad to identify the diaphragm. This appears as an echogenic curvilinear structure, with the liver or spleen being subdiaphragmatic and typically of different echogenicity than the lung. Many times the diaphragm is very high in the supine critically ill patient. Massive edema and obesity may also degrade image quality in this location.
   - Always identify the diaphragm. Hypoechoic fluid surrounding the liver or spleen can appear as a pleural effusion and must not be mistaken as such. In addition, lung tissue may mimic hepatic tissue in certain diseases such as dense consolidations termed "hepatization" of the lung. Proper probe positioning, clear identification of the diaphragm, subdiaphragmatic structures, and lung are crucial. This is a common error in novice operators owing to the confusion of the hepatorenal or splenorenal recess for the diaphragm.
   - Identifying the diaphragm can be technically difficult depending on patient position, size, and clinical condition. It maybe useful to start below the diaphragm, first identifying the hepatorenal recess (liver and kidney view interface) and then moving cephalad until the lung and diaphragm are visualized. In addition as ribs change their orientation anatomically, the probe may need to be adjusted while still in the longitudinal axis. Moving the probe clockwise and counterclockwise may be of benefit to bring into view the lung, the diaphragm, and the subdiaphragmatic structures.
10. Identify pleural effusions (Fig. 27.6). Confirming the presence of pleural effusions requires identifying anechoic material between the pleura and the lung. This can be seen as lung movement in an undulating pattern, which typically is facilitated by cardiac activity and respirations. This is termed "jellyfish sign," where the lung flaps as it freely floats in the effusion. Floating debris can also confirm effusion, termed "plankton sign." It is also important to identify the depth of the chest wall to the pleural fluid in order to determine the best location/depth of needle insertion when attempting thoracentesis or chest tube insertion. The challenge is to find a safe path for needle insertion. The key when using ultrasound as guidance is that the angle of the needle/syringe assembly must duplicate the angle of the probe. The time between scan and needle insertion must also be minimized. Real-time guidance is not required. In patients who are obese or edematous, skin indentation during probe placement can result in underestimation of depth of needle insertion required and must be taken into account. Safety margin of thoracentesis is thought to be 10 mm of visceral-parietal distance. Dry taps may be due to loculations the blocking needle, needle plugs, patient movement from scan to tap, and poor angle selection.
    - Exudates, empyemas, and hemothoraces may appear more echogenic, unlike, for example, a transudative effusion that could be anechoic. Complex effusions can also appear as heterogeneous and echogenic. The consistency of the effusion can make identification technically difficult because this can limit lung motion. Sometimes the operator may think there is no effusion when there is an echo-dense effusion.
11. Identify consolidations (Fig. 27.7). Compressed lung appears with alveolar consolidation pattern (tissue-like sign). Alveolar consolidations are devoid typically of air and appear as tissue density; these can be atelectasis, pneumonia, aspiration, or other diseased lung. "Hepatization" of the lung is typical, where the images mimic liver tissue. Images may also have hyperechoic foci representing air bronchograms, which would indicate pneumonia. Probe location should be correlated with an anatomical lobular or segmental area.
    - Lung may slide into the effusion during the respiratory cycle and can be problematic during needle insertion, causing pneumothorax or abnormal wire placement during the performance of pigtail chest tube catheters. This is called a "curtain sign."
12. Sinusoid sign (Fig. 27.8). M-mode is placed in the center of the visible lung when a large amount of pleural fluid is seen. A sinusoid sign strengthens the operator's determination that pleural fluid is present and that the pleural fluid is not necessarily compromising lung dynamics. If the sinusoid sign is absent, it may indicate a "trapped" lung dynamic.
13. Assessment and clinical decision making. Upon completion of ultrasonography of bilateral lung fields, clinical decision-making tools may be of benefit, especially in undifferentiated respiratory failure. A protocol has been developed to organize the exam of a respiratory failure patient (on noninvasive or invasive ventilation only). The BLUE protocol assesses patients based on findings (e.g., A lines, B lines, lung sliding) of both lungs and incorporates them into an algorithm. With acceptable sensitivities and specificities, practitioners can diagnose pulmonary edema, pneumonia, pneumothorax, and COPD/asthma with the BLUE protocol.

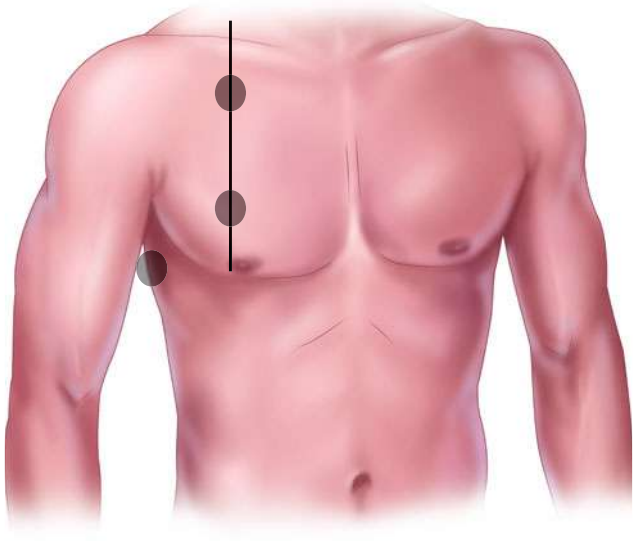

Fig. 27.1 *Black circles* indicate possible ultrasound view points. These are just guides, and one may view a few interspaces left/right and cephalad/caudad from each point to obtain better views

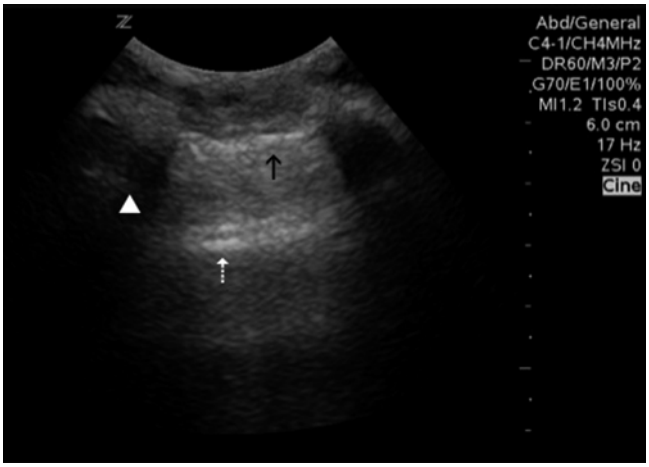

Fig. 27.2 Typical bat sign with rib shadows (*arrowhead*) that surround the pleura (*black arrow*) and lung tissue. Incidentally, A lines (*white arrow*) are seen

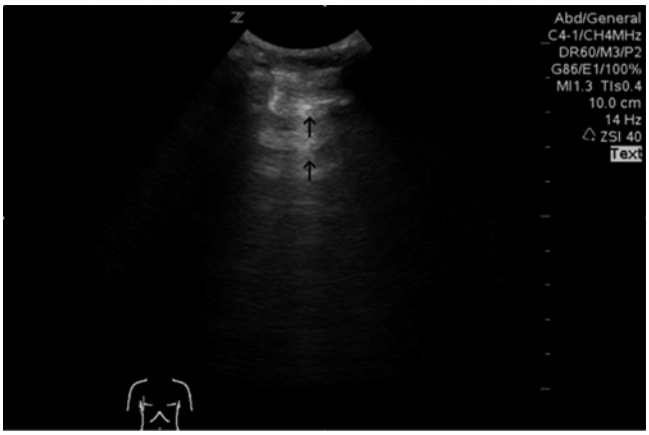

Fig. 27.3 Rib shadows can be seen, with multiple A lines (*arrows*)

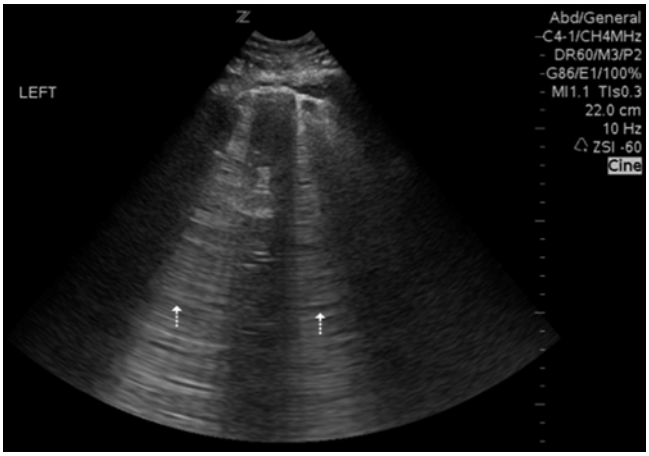

Fig. 27.4 Two B lines can be appreciated (*arrows*)

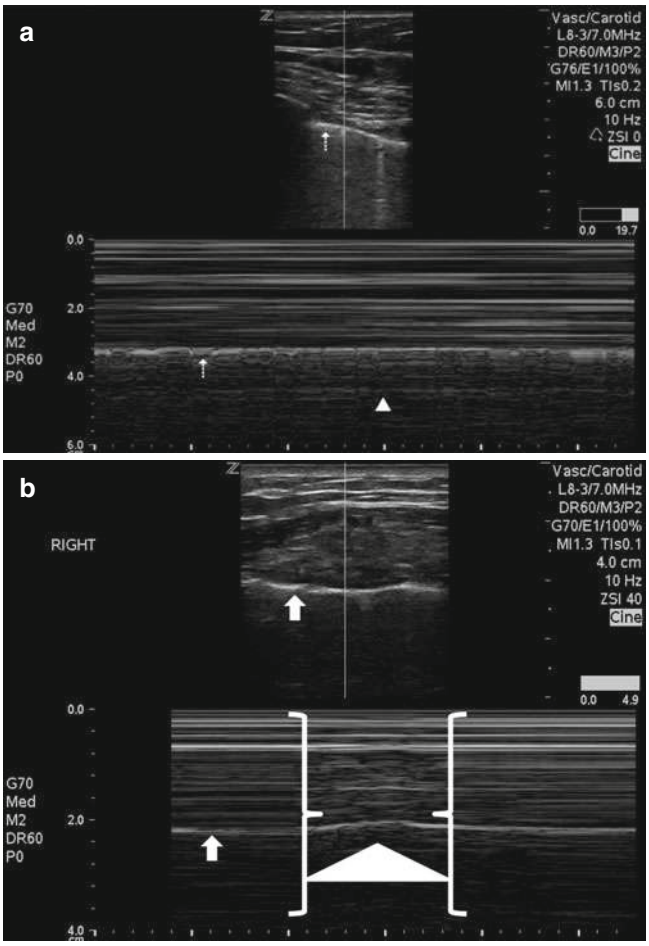

Fig. 27.5 (**a**) M-mode identifies the seashore sign, with the pleura (*arrows*) and the lung parenchyma (*arrowhead*) together producing an image similar to a seashore. (**b**) M-mode shows the stratosphere or barcode sign. *White arrows* indicate the pleura, and the *large triangle* shows an area of artifact from movement. Artifact from movement can be distinguished by a similar pattern change above and below the pleural line. On either side of the bracket is a continued horizontal line pattern typical of absent lung sliding

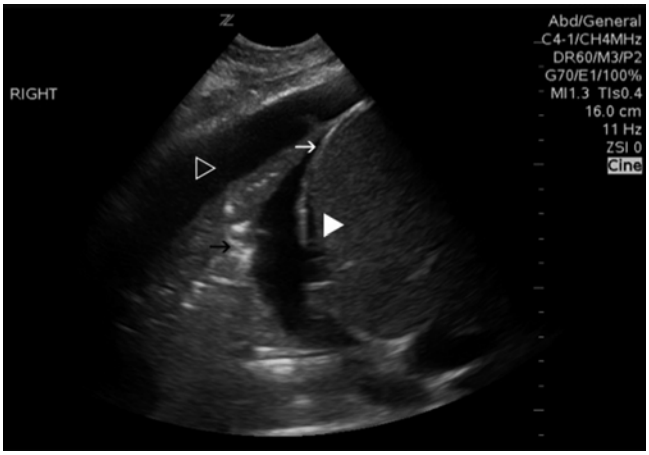

Fig. 27.6 Diaphragm (*white arrow*), liver (*white arrowhead*), lung (*black arrow*), and anechoic fluid, likely an effusion (*open arrowhead*). Incidentally, hyperechoic foci in the lung tissue, likely representing air bronchograms, can be seen

Fig. 27.7 Lung here appears as tissue density, likely representing the alveolar consolidation pattern seen in pneumonia, and is termed "hepatization" and "tissue-like sign"

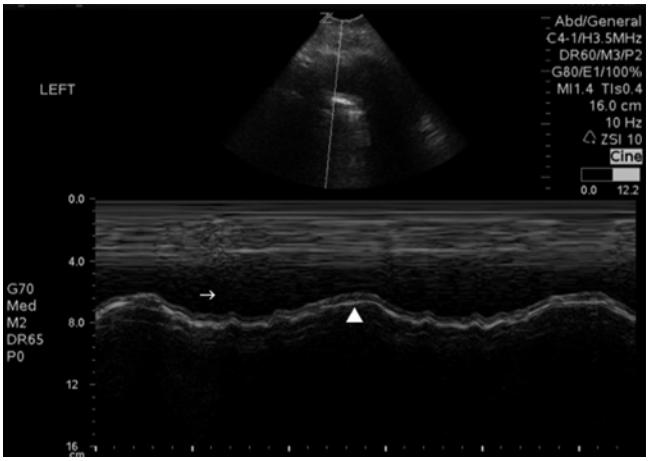

Fig. 27.8 In M-mode, identification of pleural effusion becomes evident with the sinusoidal sign (*arrowhead*) and identification of pleural fluid (*arrow*)

Selected Reading

Levitov A, Mayo PH, Slonim AD. Ultrasound evaluation of the lung. In: Critical care ultrasonography. New York: McGraw-Hill; 2009.

Lichtenstein DA, Menu Y. A bedside ultrasound sign ruling out pneumothorax in the critically ill. Lung sliding. Chest. 1995;108: 1345–8.

Lichtenstein DA, Meziere GA. A lung ultrasound sign allowing bedside distinction between pulmonary edema and COPD: the comet-tail artifact. Intensive Care Med. 1998;24:1331–4.

Lichtenstein DA, Meziere GA. Relevance of the lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. Chest. 2008;134:117–25.

Mayo PH, Doelken P. Pleural ultrasonography. Clin Chest Med. 2006;27:215–27.

Mayo PH, Goltz HR, Tafreshi M, Doelken P. Safety of ultrasound-guided thoracentesis in patients receiving mechanical ventilation. Chest. 2004;125:1059–62.

Part IV

Cardiac Procedures

28 Repair of Cardiac Injuries

Ronald Tesoriero

Whether performed by an emergency physician or a surgeon, the majority of cardiac repairs performed at emergency department thoracotomy will be temporary in nature and require further revision in the operating room. This chapter's main focus is the temporary control of cardiac injuries.

28.1 Indications

- Wounds to the heart in patients presenting with pulseless electrical activity (PEA) or asystole with evidence of cardiac tamponade
  - Penetrating wounds: <15 min of prehospital cardiopulmonary resuscitation (CPR)
  - Blunt wounds: <10 min of prehospital cardiopulmonary resuscitation (CPR)

28.2 Contraindications

- Absolute
  - Presenting rhythm of asystole and no evidence of pericardial tamponade on Focused Assessment with Sonography for Trauma (FAST)
- Relative
  - None

28.3 Materials and Medications

- Diagnostic ultrasound
- Emergency department thoracotomy tray (Fig. 28.1)
  - Sterile drapes, #10 scalpel, curved Mayo scissors, Finochietto retractor, Lebsche sternal knife and mallet, forceps, curved Metzenbaum scissors, surgical skin stapler, Foley catheter, clamps, needle driver, 2–0 and 3–0 polypropylene suture on MH or SH (noncutting) needle, Satinsky vascular clamps
- Betadine (povidone-iodine) or other skin antiseptic preparing solution
- Defibrillator and internal cardiac panels
- Epinephrine

R. Tesoriero, MD
Department of Surgical Critical Care,
University of Maryland School of Medicine,
R Adams Cowley Shock Trauma Center,
Baltimore, MD, USA
e-mail: rtesoriero@umm.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_28

Fig. 28.1 Emergency department thoracotomy tray. *Left to right*: Scalpel, curved Mayo scissors, Finochietto rib spreader, DeBakey forceps, Metzenbaum scissors, needle driver, Lebsche knife and mallet, Satinsky vascular clamp, aortic clamp, bone cutter

28.4 Procedure

1. Identify the pericardium anterior to the phrenic nerve, pinch it between the fingers, and enter it by making a nick with the Metzenbaum scissors. Then open the pericardium longitudinally anterior to the phrenic nerve.
2. If the heart is not contracting effectively (PEA, asystole, fibrillation), immediately begin internal cardiac massage.
   - Unless the injury is so large that it cannot be controlled, taking time out at this point to repair the injury may lead to significant acidosis and inability to reestablish a perfusing rhythm.
3. While continuing internal cardiac massage, identify the area of injury and attempt to control it with manual pressure; if a larger wound, use a Foley catheter to prevent continued bleeding.
4. If necessary owing to the location of injury, extend the incision to a bilateral anterolateral thoracotomy and transect the sternum with a Lebsche knife and mallet or heavy shears.
5. If, after several minutes of internal massage and appropriate red blood cell and plasma transfusion, the patient remains in PEA or asystole, irrigate the heart with warmed saline and administer intravenous (IV) or intracardiac epinephrine.
6. If the heart enters ventricular fibrillation or ventricular tachycardia, cardiovert with internal paddles applied directly to the heart with an energy between 10 and 30 J.
7. Once a perfusing rhythm is reestablished, or if on initial evaluation the wound is so large that it cannot be controlled and will require immediate repair before internal massage can be effective, proceed to cardiac repair.
8. Choose the simplest method that will allow control of the injury until definitive repair can be performed in the operating room.
   (a) Injuries to the atrium: The pliable nature of the atria will often allow placement of a Satinsky clamp for control followed by repair with a running 3–0 polypropylene suture (Fig. 28.2).
   (b) Small injuries to the ventricles: Control the injury with direct manual pressure and close with either interrupted 3–0 polypropylene suture or a surgical skin stapler (Figs. 28.3 and 28.4).
   (c) Medium to large injuries to the ventricle: Attempt to control the wound by placing a Foley catheter through it, then blow up the balloon and apply gentle traction. Either staple or suture the wound closed before deflating the balloon and removing the catheter (Fig. 28.5).
      - Place an occluding clamp on the open end of the catheter or blood loss will continue through the catheter
      - Avoid excessive traction on the Foley because it will pull through the ventricle and make the hole larger
   (d) Extensive or inaccessible injury: Perform temporary inflow occlusion to the heart by manually compressing the right atrium against the heart so that it cannot fill. The heart will likely immediately enter PEA, fibrillation, or asystole giving the physician a couple of minutes to gain control of the injury before the patient becomes unrecoverable.
   (e) Injuries in proximity to coronary vessels: To avoid compression of the vessel, perform a horizontal mattress suture that passes beneath the artery. Teflon pledgets may assist in the repair (Fig. 28.6).

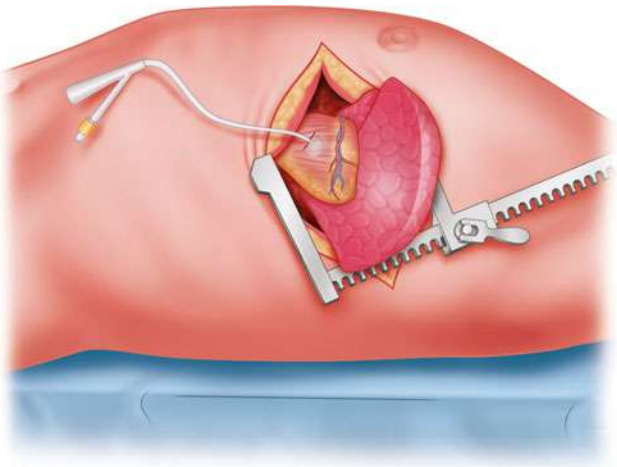

Fig. 28.2 Position the Finochietto retractor with the closed end toward the axilla. A Foley catheter may be used for initial control of large cardiac lacerations

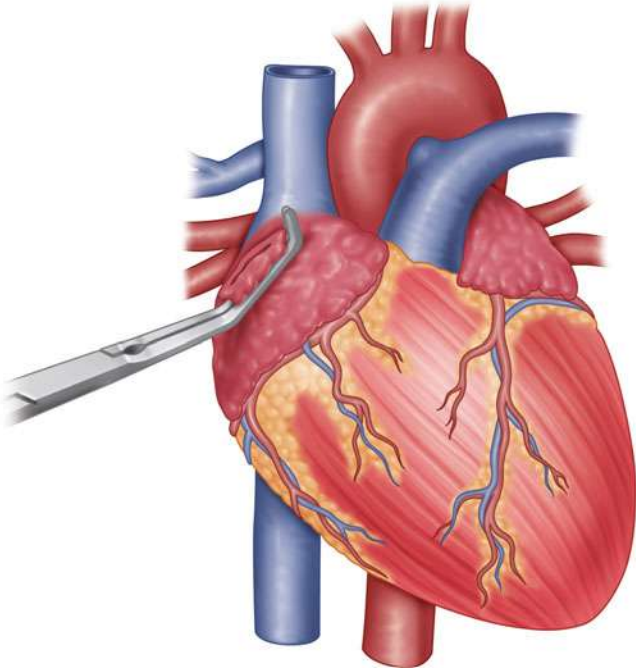

Fig. 28.3 Atrial injuries may be quickly controlled with a Satinsky clamp followed by repair with a 3–0 polypropylene suture

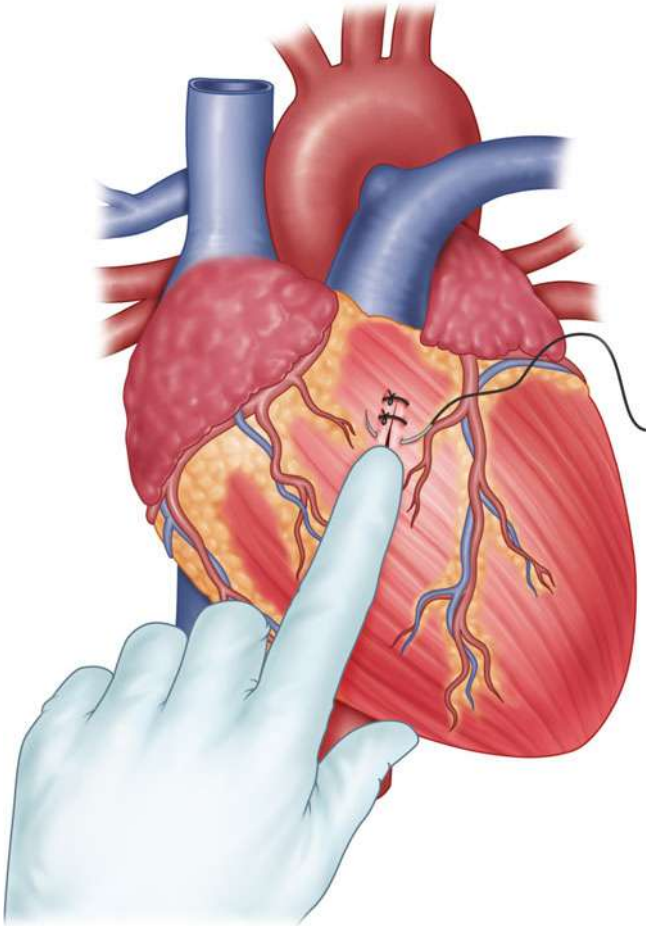

Fig. 28.4 Injuries may be controlled with direct manual pressure and closed with interrupted 3–0 polypropylene suture

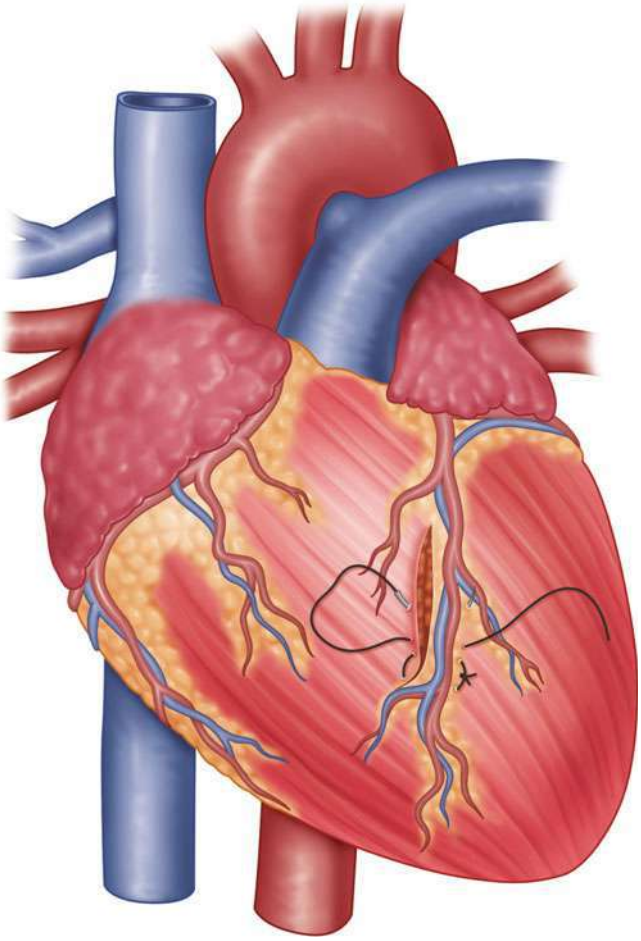

Fig. 28.6 Lacerations in close proximity to coronary vessels can be controlled with horizontal mattress sutures passed beneath the artery to avoid coronary arterial compression

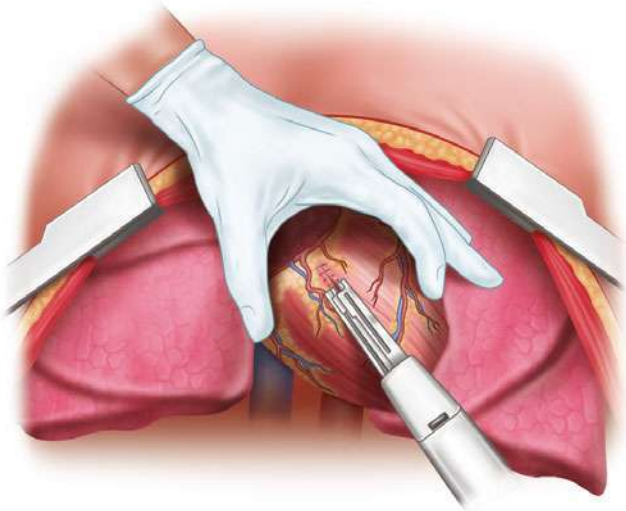

Fig. 28.5 A surgical skin stapler is a quick way to gain initial control of simple cardiac lacerations with minimal risk to the provider

28.5 Complications

- Cardiac and/or pulmonary laceration during entry into the chest
- Bleeding
- Delayed hemorrhage (failure to control internal mammary artery, disruption of sutures or staples)
- Infections: empyema and sternal infection
- Missed intracardiac valvular or septal injury (echocardiography should be performed after repair is complete)
- Air embolus

28.6 Pearls and Pitfalls

- Pearls
  - There may be more than one wound to the heart (especially with gunshot wounds). Look for them. However, if the wound is posterior and not bleeding with the heart in its natural position, it will be more prudent to leave the injury alone until the patient can be transported to the operating room. Elevating the heart can cause both inflow and outflow obstruction, leading to dysrhythmia that may be difficult to recover from.
  - To avoid sutures pulling through in patients with thin, edematous, or friable myocardium, consider horizontal mattress rather than simple sutures. These may be buttressed with Teflon pledgets for added security.
- Pitfalls
  - The myocardium tears easily. When tying sutures, take care to not tighten them too forcefully.

Selected Reading

Asensio JA, Trunkey DD. Current therapy of trauma and surgical critical care. Philadelphia: Mosby/Elsevier; 2008.

Feliciano DV, Mattox KL, Moore EE. Trauma. 6th ed. New York: McGraw-Hill Medical; 2008.

Hirshberg A, Mattox KL. Top knife: the art & craft of trauma surgery. Castle Hill Barns: TFM; 2005.

Moore EE, Knudson MM, Clay CB, et al. Defining limits of resuscitative emergency department thoracotomy: a contemporary Western Trauma Association perspective. J Trauma. 2011;70:334–9.

Wall MJ, Mattox KL, Chen CD, et al. Acute management of complex cardiac injuries. J Trauma. 1997;42:905–12.

29 Synchronized Electrical Cardioversion

Jason Jones, Ann Tsung, and Marie-Carmelle Elie

Tachycardia is defined as >100 beats per minute

29.1 Indications

- Non-sinus-rhythm tachycardias with a pulse including:
  - Atrial fibrillation
  - Atrial flutter
  - Monomorphic ventricular tachycardia (VT)
  - Refractory or unstable supraventricular tachycardia (SVT)
- Unstable signs and symptoms including acute coronary syndrome, decreased level of consciousness, chest pain, dyspnea, pulmonary edema, and hypotension

29.2 Contraindications

- Absolute
  - Ventricular fibrillation and pulseless or polymorphic (irregular) VT require unsynchronized electrical cardioversion (defibrillation), not synchronized cardioversion.
  - Known atrial thrombus.
  - Sinus tachycardia.
- Relative
  - Digitalis toxicity-related tachycardia
  - Atrial fibrillation of greater than 48 h duration without anticoagulation
  - Multifocal atrial tachycardia
  - Electrolyte abnormalities
  - Left atrial diameter greater than 4.5 cm
  - Patients with low probability of maintaining sinus rhythm and readily return to atrial fibrillation
  - Patients with sick sinus syndrome or sinoatrial blockage who will require a pacemaker for maintenance of stable rhythm

29.3 Materials and Medications

- Airway management equipment (laryngoscopes, endotracheal tubes)
- Cardiac monitoring, pulse oximetry, end-tidal CO_2 monitoring
- Cardioverter/defibrillator
- Sedation and analgesic medications

29.4 Procedure

1. Obtain a 12-lead electrocardiogram (ECG) and intravenous (IV) access.
2. If possible, correct underlying electrolyte abnormalities that may cause or contribute to the patient's arrhythmia.
3. Discuss risks, benefits, and alternatives (including pharmacological cardioversion) with the patient and obtain consent.
4. Prepare airway equipment and Advanced Cardiac Life Support (ACLS) code drugs.
5. Consider IV sedation (e.g., propofol, midazolam).
6. Provide IV analgesia (e.g., fentanyl, morphine).
7. Place defibrillator adhesive pads (8- to 12-cm diameter in adults) or paddles on the patient. Pediatric-sized pads/paddles should be used if the patient is less than 10 kg.

J. Jones, MD • M.-C. Elie, MD, RDMS (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: jasonjones@ufl.edu; elie@ufl.edu

A. Tsung, MD
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: TSUNGANN@UFL.EDU

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_29

8. The first paddle/pad is placed to the right of the sternum at the second/third intercostal space. The second paddle/pad can be placed in one of two equally efficacious positions:
   (a) Anterolateral position—left fourth/fifth intercostal space in the midaxillary line (Fig. 29.1)
   (b) Anteroposterior position—between the spine and the edge of the left scapula (Fig. 29.2)
9. Turn the defibrillator/cardioverter into synchronized mode—marker above R-waves will be present (Fig. 29.3).
10. Select the energy level to be delivered based on the underlying rhythm
   (a) Regular VT (with pulses)—Adults: 100 J (monophasic or biphasic), 200 J for subsequent shocks
   (b) Atrial fibrillation—120–200 J (biphasic), 200 J (monophasic), 360 J for subsequent shocks
   (c) Atrial flutter and paroxysmal SVT—50–100 J (biphasic), 100 J for subsequent shocks
   (d) Pediatric dosage (regular and pulsed VT or SVT)—0.5–1 J/kg, up to 2 J/kg for subsequent shocks (Fig. 29.4)
11. Announce that you are going to deliver the shock on the count of three, and ensure that everyone is clear of the patient.
12. Deliver the shock by pressing button marked "SHOCK."
   - If using paddles, apply firm pressure and keep paddles in place until shock is delivered.
13. Reassess the patient's pulse and cardiac rhythm.
14. Repeat with escalating energy in a stepwise fashion if cardioversion is unsuccessful.

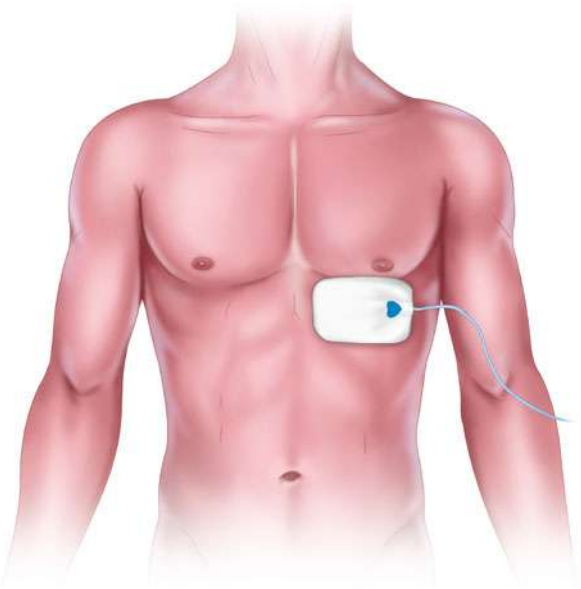

Fig. 29.1 Anterolateral pad placement

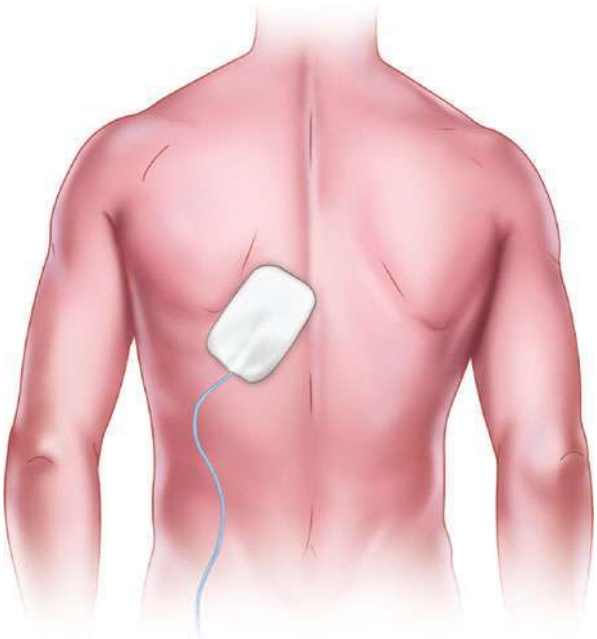

Fig. 29.2 Anteroposterior pad placement

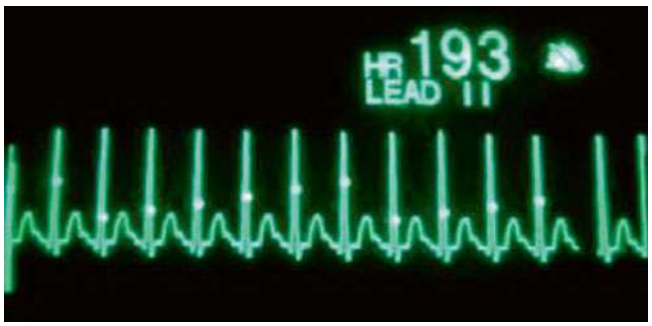

Fig. 29.3 Synchronized cardioversion—mark on R wave

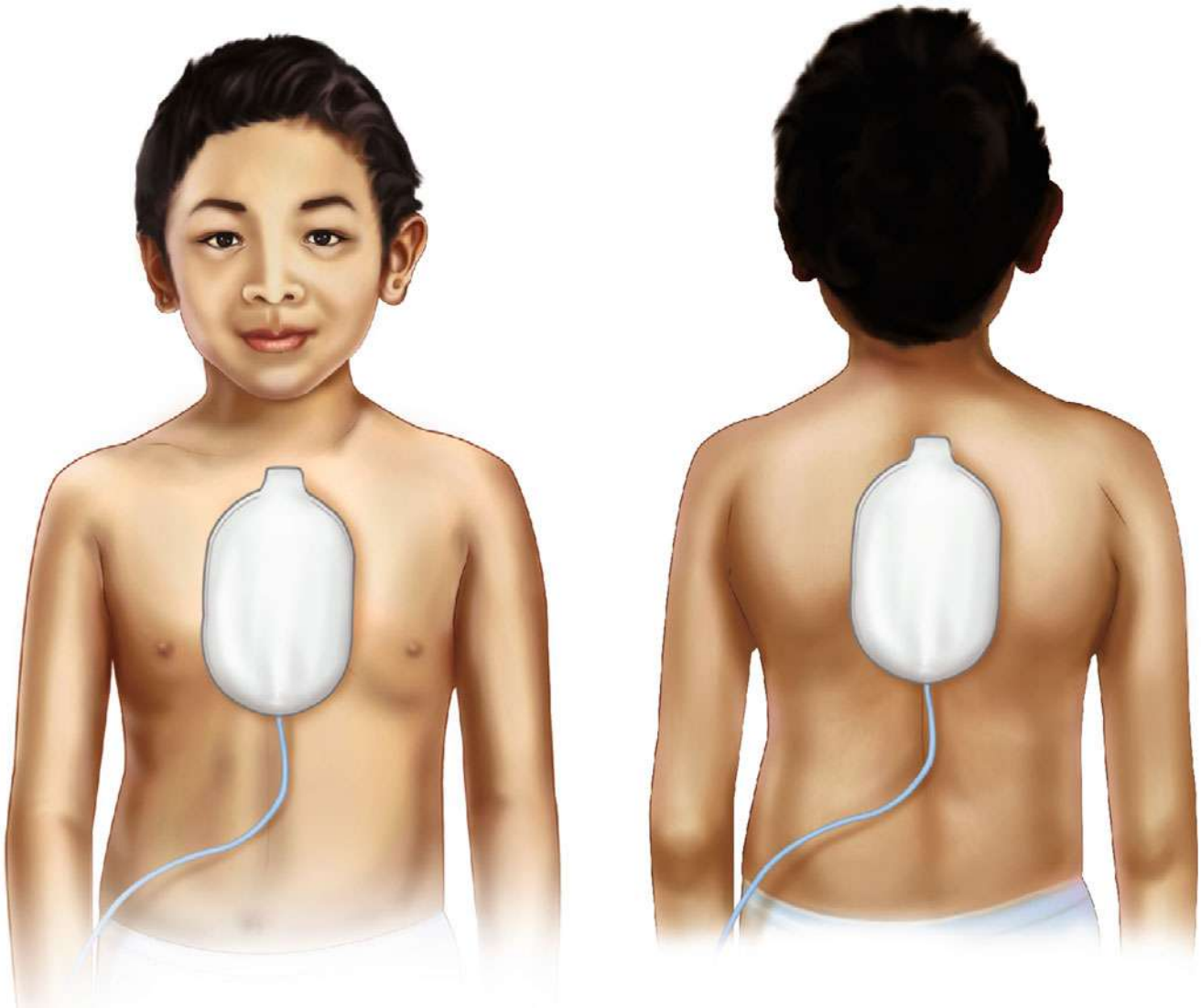

Fig. 29.4 Pediatric pad placement

29.5 Complications

- Superficial burns if there is inadequate gel.
- Induced arrhythmias (bradycardia in patients with previous inferior myocardial infarction, atrioventricular block, VT, ventricular fibrillation, asystole).
- Improperly synched cardioversion may rarely induce ventricular fibrillation.
  - Ectopy of the atria or ventricle in first 30 min after cardioversion
  - Atrial clot embolization in patients without adequate anticoagulation
- Apnea, hypoxia, hypercarbia, or hypotension may occur from sedation/analgesia.
- Medical professionals who incidentally touch the patient during shock delivery may be shocked or burned.
- Rarely, fire has occurred as a consequence of poor pad placement and a hyperoxygenated environment.

29.6 Pearls and Pitfalls

- Hirsute males should be shaved at the pad/paddle placement sites.
- Placing the cardioverter in "synchronized" mode avoids delivering a shock during the relative refractory segment, which could induce ventricular fibrillation.
- Cardioversion of pregnant patients is performed as in the general adult population.
- Keep pacemakers and implantable cardioverter-defibrillators at least 10 cm away from contact with paddles.
- If patient has implanted pacemaker, position pads so that they are not directly over the device.

Selected Reading

Gowda SA, Shah A, Steinberg JS. Cardioversion of atrial fibrillation. Progr Cardiovasc Dis. 2005;48:88–107.

Kleinman ME, Chameides L, Schexnayder SM, et al. Part 14: pediatric advanced life support. 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2010;122 Suppl 3:S876–908.

Link MS, Atkins DL, Passman RS, et al. Part 6: electrical therapies: automated external defibrillators, defibrillation, cardioversion, and pacing. 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2010;122 Suppl 3:S706–19.

Mayeaux EJ. The essentials guide to primary care procedures. Philadelphia: Lippincott Williams & Wilkins; 2012. p. 88–92.

Sirna SJ, Ferguson DW, Charbonnier F, et al. Factors affecting transthoracic impedance during electrical cardioversion. Am J Cardiol. 1988;62:1048–52.

Trohman RG, Parrillo JE. Direct current cardioversion: indications, techniques, and recent advances. Crit Care Med. 2000;28:170–3.

30 Unsynchronized Cardioversion (Defibrillation)

Matthew R. Tice, Zachary B. Kramer, and Marie-Carmelle Elie

Unsynchronized cardioversion or defibrillation is the delivery of a *high-energy* shock as soon as the button is pushed on defibrillator. This means it can be delivered anywhere in the cardiac cycle. By contrast, synchronized cardioversion (see Chap. 29) delivers a low-energy shock at the peak of the R wave in the cardiac (QRS) cycle.

30.1 Indications

- Ventricular fibrillation (VF)
- Pulseless ventricular tachycardia (VT)
- Cardiac arrest due to or resulting from VF

30.2 Contraindications

- Absolute
  - Conscious patient
  - Presence of a pulse
  - Pulseless electrical activity (PEA)
  - Asystole
  - Multifocal atrial tachycardia
  - Defibrillation without knowing the rhythm
  - A second defibrillation before 2 min (or five cycles) of CPR
  - Advanced Directive, Physician Order for Life-Sustaining Treatment (POLST) indicating no cardiopulmonary resuscitation (CPR) or do not resuscitate (DNR)
- Relative
  - Potential electrical catastrophe (explosive environment [i.e., operating rooms])
  - Dysrhythmias due to enhanced automaticity such as in digitalis toxicity and catecholamine-induced arrhythmia (because mechanism of tachycardia remains after the shock)
- Factors that are not contraindications
  - Pregnancy.
  - Chest trauma.
  - Automatic implantable cardiac defibrillators (AICDs).
  - The patient is on a wet or moist surface.
  - Piercings on the chest.

30.3 Materials and Medications

- Electrocardiogram (ECG) monitor/defibrillator.
- Self-adhesive defibrillation pads or defibrillation paddles (paddles may be more successful than self-adhesive pads, but they have more complications and pose more danger to operators).
- Conductive gel for defibrillation paddles (not ultra sound gel).
- ECG electrodes.
- Supplemental oxygen.
- Intubation equipment as needed.

M.R. Tice, MD • Z.B. Kramer, MD • M.-C. Elie, MD, RDMS (✉)
Department of Emergency Medicine,
University of Florida, Gainesville, FL, USA
e-mail: mtice@ufl.edu; kramer@ufl.edu; elie@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_30

30.4 Procedure

Defibrillation is an emergent maneuver and, when necessary, should be promptly performed in conjunction with or before administration of induction or sedative agents to facilitate intubation.

1. Assess the ABCs (airway, breathing, circulation).
   - Open the airway with a head tilt/chin lift (or jaw thrust in a suspected traumatic patient). If the patient is apneic, provide breaths with a bag-valve-mask (BVM) and observe chest rise.
   - Check for pulses. If absent, start CPR.
   - CPR should be initiated before any shock while getting all equipment ready for at least 2 min to provide adequate circulation to the brain and heart.
2. Wipe off the patient's chest if moist or wet.
3. Remove transdermal patches, jewelry, and piercings if possible.
4. Attach ECG electrodes to the patient.
   - The self-adhesive defibrillation pads or defibrillation paddles can be used as ECG electrodes to access the rhythm.
5. Paddles: With conductive gel applied to the metal surface, place one paddle on the patient's right chest, just below the clavicle, near the sternal border. The other should be on the left chest, midaxillary line above the fifth or sixth intercostal space (Fig. 30.1).
   - The long axis of the paddles should be perpendicular to the ribs to allow for better transduction of current through the chest.
   - Pads: Same placement as paddles except that pads can be placed in any orientation as long as they are in full contact with the chest.
   - If a lot of breast tissue is present, push the tissue to one side or lift it away and place the paddles or pads underneath.
   - An error in pad or paddle placement can distort the rhythm into looking like a rhythm that does not require defibrillation.
6. Place the ECG monitor/defibrillator into a mode to acquire a rhythm from the pads or paddles.
   - Steps 1–3 can be done simultaneously.
7. Stop CPR and assess the rhythm and pulse for no longer than 10 s.
8. If VF or pulseless VT is observed, then switch the defibrillator to charge mode.
9. Charge to 200 J.
   - Continue CPR while the defibrillator is being charged.
10. When the defibrillator indicates that it is charged, clearly order that everyone stop touching the patient and observe that there is no physical contact before defibrillating the patient.
    - If using the paddles, apply extra force to the chest through the paddles to deflate the lungs to allow for better defibrillation.
    - The operator should observe a muscle twitch during defibrillation.
11. Restart CPR for 2 min or five cycles.
    - Another operator may charge (but not fire) the defibrillator while CPR is being performed to expedite the time between pulse/rhythm check and the initiation of a shock.
12. After 2 min or five cycles of CPR, assess the rhythm and pulse and repeat steps 6–8 and give appropriate Advanced Cardiac Life Support (ACLS) medications.
13. If successful return of spontaneous circulation (ROSC) occurs, initiate the hypothermia protocol per hospital guidelines.

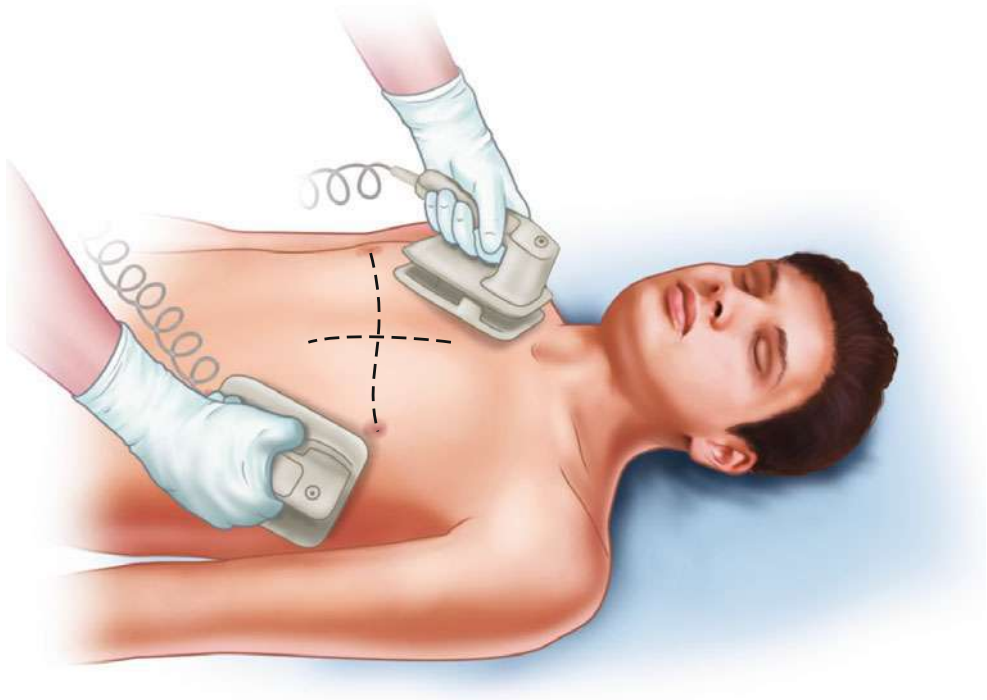

Fig. 30.1 Proper placement of defibrillation paddles

30.5 Complications

- Skin burns (most common and likely due to improper technique).
- Injury to cardiac tissue (myocardial necrosis secondary to burn).
  - ST segment elevation that lasts longer than 2 min usually indicates myocardial injury unrelated to the shock.
- Abnormal heart rhythms (usually benign like atrial, ventricular, and junctional premature beats).

30.6 Pearls and Pitfalls

- Pearls
  - Post-defibrillation cardiac dysrhythmias are more common following prolonged VF and higher-energy-level countershocks. Early defibrillation at the recommended energy levels minimizes this complication. Follow ACLS protocols to manage the resulting dysrhythmias.
  - The presence of liquid (body fluids, medications, or intravenous fluids) may cause electrical arcing thermal burns to the skin and soft tissue and produce ineffective defibrillation by allowing the current to pass across the trunk rather than transthoracically. To minimize this potential complication, ensure that any body fluids or liquids are wiped away from the skin before defibrillation attempts.
- Pitfalls
  - Myocardial and epicardial injury (not direct thermal injury) may result from the electrical current applied in defibrillation. Use the minimal recommended energy levels to minimize injury.
  - Electrical injuries to health-care providers can result if participants remain in contact with the patient during delivery of a countershock because they can serve as a ground for the current discharged.
    - This can be minimized by wearing gloves and using biphasic machines and electrode pads.
  - Fires that can result from sparks in the presence of nitroglycerin patches or ointment, flammable gases, or an oxygen-rich environment can also be a source of injury to the patient or health-care providers. Ensure "all clear" from the patient before delivery of shock to avoid these complications.

Selected Reading

Advance cardiovascular life support: provider manual. Dallas: American Heart Association; 2011.

Deakin CD, Sado DM, Petley GW, Clewlow F. Is the orientation of the apical defibrillation paddle of importance during manual external defibrillation? Resuscitation. 2003;56:15–8.

Pagan-Carlo LA, Spencer KT, Robertson CE, Dengler A, Birkitt C, Kerber RE. Transthoracic defibrillation: importance of avoiding electrode placement directly on the female breast. J Am Coll Cardiol. 1996;27:449–52.

Walker C, Callies F, Langenfeld H. Adverse effects of direct current cardioversion on cardiac pacemakers and electrodes: is external cardioversion contraindicated in patients with permanent pacing systems? Eur Soc Cardiol. 2004;6:165–8.

31 Transcutaneous Pacing

Nour Rifai and Christian Coletti

31.1 Indications

- Hemodynamically unstable (i.e., hypotension, pulmonary edema, chest pain, shortness of breath, or evidence of decreased cerebral perfusion) bradyarrhythmias refractory to medical therapies
- As a bridge to a transvenous or permanent pacemaker
- As an overdrive pacer in tachyarrhythmias
- Controversially, within the first 10 min of a witnessed asystolic cardiac arrest
- In children only with bradycardia associated with a known congenital cardiac defect or after cardiac surgery

31.2 Contraindications

- Absolute
  - None
- Relative
  - Bradyarrhythmia associated with hypothermia (ventricles are more prone to defibrillation-resistant fibrillation)
  - Prolonged cardiac arrest (>20 min)
  - Bradyarrhythmia in children (usually secondary to hypoxia or a respiratory issue)
  - Patient is unable to tolerate the procedure despite sedation and analgesia

N. Rifai, MBCHB • C. Coletti, MD (✉)
Department of Emergency Medicine,
Christiana Care Health System,
Newark, DE, USA
e-mail: nourifai@gmail.com; chris.coletti@gmail.com

31.3 Materials and Medications

- Pacemaker device (modern units offer combined pacer and defibrillator functions) (Fig. 31.1)
- One set of standard electrocardiogram (ECG) electrodes
- One set of pacer pads
- Code cart and airway equipment (prophylactically)
- Sedation and analgesia (typically a short-acting benzodiazepine and an opioid)
  - Midazolam: 0.2–0.10 mg/kg intravenous (IV) push and may repeat with 25 % of initial dose after 3–5 min. Do not exceed 2.5 mg/dose or a cumulative dose of 5 mg.
  - Fentanyl: 1–2 mcg/kg IV slow push over 1–2 min, may repeat dose in 30 min. (Fentanyl is the opioid of choice because it is less likely to exacerbate any hypotension.)

Fig. 31.1 Pacemaker device

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_31

31.4 Procedure

1. Time permitting, clean and dry the skin, and shave any excess hair off the chest.
2. Administer any appropriate sedation and analgesia.
3. Attach the ECG electrodes to both the input port of the pacemaker unit and the patient. On the patient, the white lead is placed just above the right clavicle, the black lead is just above the left clavicle, and the red lead is around the left midaxillary line.
4. Attach the pacer pads either in the anteroposterior or anterolateral positions as pictured (avoid placement over an implanted pacemaker or defibrillator; Fig. 31.2).
5. Turn the machine on and switch it to synchronous (or on-demand) mode.
   - Asynchronous (or fixed) mode fires impulses with no regard to the intrinsic cardiac cycle, increasing the likelihood of an R on T phenomenon, which could result in ventricular tachycardia or fibrillation.
   - Synchronous (or on-demand) mode will not fire an electrical impulse when a QRS complex is sensed; this is the preferred mode for transcutaneous pacing.
6. Set the desired heart rate: typically 60–80 beats/min to achieve adequate perfusion.
7. Select a lead on the pacemaker unit and then press Start.
8. Slowly increase the output current until electrical capture is denoted by a visible pacemaker spike, which will precede every QRS complex on the ECG monitor (Fig. 31.3).
   - Electrical capture is usually achieved between 50 and 100 mA.
   - If a patient is unconscious or truly deteriorating quickly or in cardiac arrest, it may be prudent to set the initial currents at maximum to ensure rapid capture and then decrease the current to just above that at which electrical capture was achieved.
9. After electrical capture is appreciated on the monitor, assess for mechanical capture by palpation of a pulse at a rate that corresponds to that which the machine is set at. An improved blood pressure or resolution of chest pain, shortness of breath, or altered mental status also suggests that the heart rate has improved and perfusion has been restored.
10. When pacing in overdrive for tachyarrhythmias, the pacer rate is set 20–60 beats/min faster than the detected tachycardic rate.
    - Bear in mind that rhythm acceleration or the induction of ventricular fibrillation is a possibility with pacing, hence the recommendation of having a code cart and airway equipment in the room at all times.

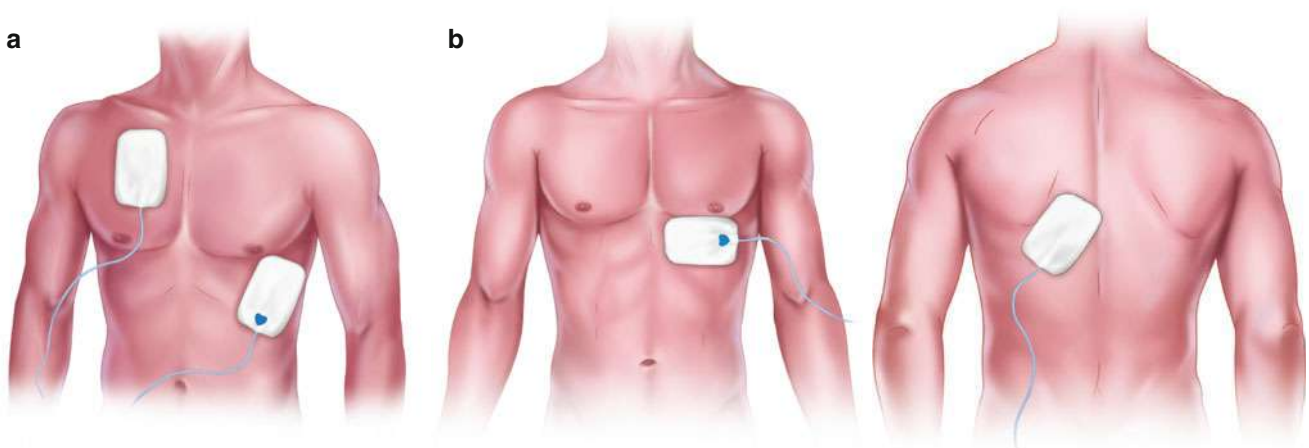

Fig. 31.2 Pacer pads attached in the anterolateral (**a**) or anteroposterior positions (**b**)

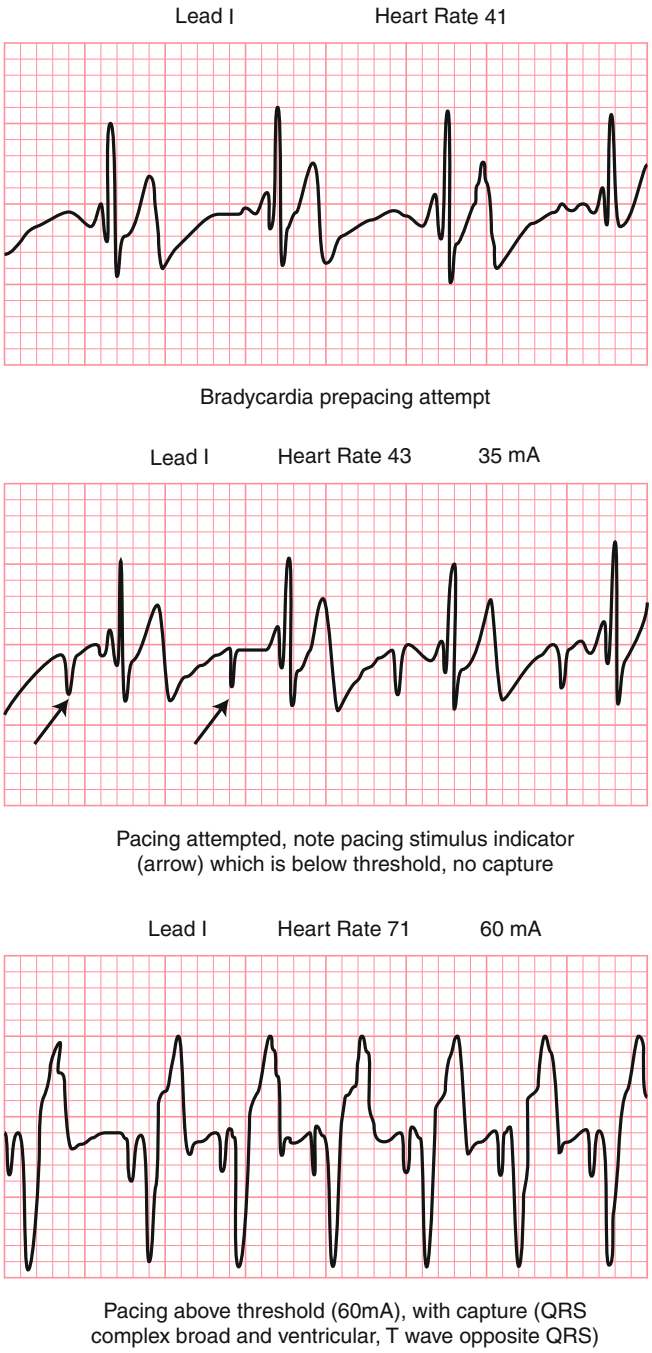

Fig. 31.3 A visible pacemaker spike will precede every QRS complex on the ECG monitor

31.5 Pearls and Pitfalls

- Pearls
  - Pacer pads can be prophylactically placed on bradycardic but stable patients in preparation for potential decompensation.
  - It may be prudent to discuss with a stable but bradycardic patient the option of trying the pacemaker not only to ensure that the machine and pads are working but also to gauge an idea of the output current that will be needed to electrically capture for quicker, more efficient, and potentially lifesaving pacing that may prove necessary.
  - When palpating for mechanical capture, the femoral pulse may be easier because carotid palpation may prove difficult with muscular contractions induced by the pacer impulses.
  - If impulses are not capturing, try slightly readjusting the pacer pads, making sure the leads are still on, the settings were not changed, the machine's battery did not die, or that the machine is plugged in.
  - If electrical capture is achieved but mechanical is not, increase the rate until adequate pulses are palpated.
  - Be careful to not miss an underlying ventricular fibrillation if the monitor is not blanked or dampened by an ECG screen.
  - Cardiopulmonary resuscitation (CPR) can be continued even while transcutaneous pacing is taking place because the pacer pads are insulated and the power delivered in each impulse is minimal, making the risk of injury to health-care workers very low.
- Pitfalls
  - Certain factors including a large body habitus, the presence of large pericardial effusions, scarring secondary to intrathoracic surgeries, or large amounts of intrathoracic air associated with obstructive pulmonary diseases may not only increase the threshold for capture but also may even cause a failure to capture, in which case transvenous pacing should immediately be considered.
  - Be aware that long-term pacing and pacing in children increases the likelihood for cutaneous and soft tissue damage.

Selected Reading

Bonow JS, Mann DL, Zipes DF, editors. Braunwald's heart disease: a textbook of cardiovascular medicine. 9th ed. Philadelphia: Saunders Elsevier; 2008.

Pfenninger JL, Fowler GC, editors. Pfenninger and Fowler's procedures for primary care. Philadelphia: Saunders Elsevier; 2010.

Roberts JR, Hedges JR, editors. Clinical procedures in emergency medicine. 5th ed. Philadelphia: Saunders Elsevier; 2010.

Tintinalli JE, Stapczynski JS, Cline DM, Ma OJ, Cydulka RK, Meckler GD, editors. Tintinalli's emergency medicine: a comprehensive study guide. 7th ed. New York: McGraw-Hill Medical; 2011.

32 Transvenous Cardiac Pacing

Katrina John, Jeffrey Kile, and Amish Aghera

32.1 Indications

Hemodynamic compromise in the presence of:

- Sinus node dysfunction
- Second- and third-degree heart block
- Atrial fibrillation with slow ventricular response
- New left bundle branch block (LBBB)
- Right bundle branch block (RBBB)
- Bifascicular block
- Alternating bundle branch block (BBB)
- Implanted pacemaker malfunction

32.2 Contraindications

- Absolute
  - Prosthetic tricuspid valve
- Relative
  - Bradycardia in the presence of severe hypothermia

K. John, MBBS • J. Kile, MBBS, PhD, MPH (✉)
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: trenjohn@me.com; jeffrey.kile@gmail.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_32

32.3 Materials and Medications (Fig. 32.1)

- Sterile gloves
- Sterile gown and drapes
- Face mask and surgical cap
- Two 10-mL syringes
- One 3-mL syringe
- Local anesthetic (1 or 2 % lidocaine without epinephrine)
- 22-gauge needle
- Povidone-iodine or chlorhexidine/isopropyl alcohol
- Several 4×4-in. gauze sponges
- Sterile dressing pack
- 4-0 nylon or silk sutures
- Scalpel (#11 blade)
- Needle holder
- Scissors
- Introducer needle
- Guidewire
- Dilator
- Introducer sheath
- Collapsed sterile extension sheath
- Balloon-tipped pacing catheter
- Pacing generator with working battery
- Insulated cable with alligator clamps or other suitable connectors
- Electrocardiogram (ECG)/cardiac monitor

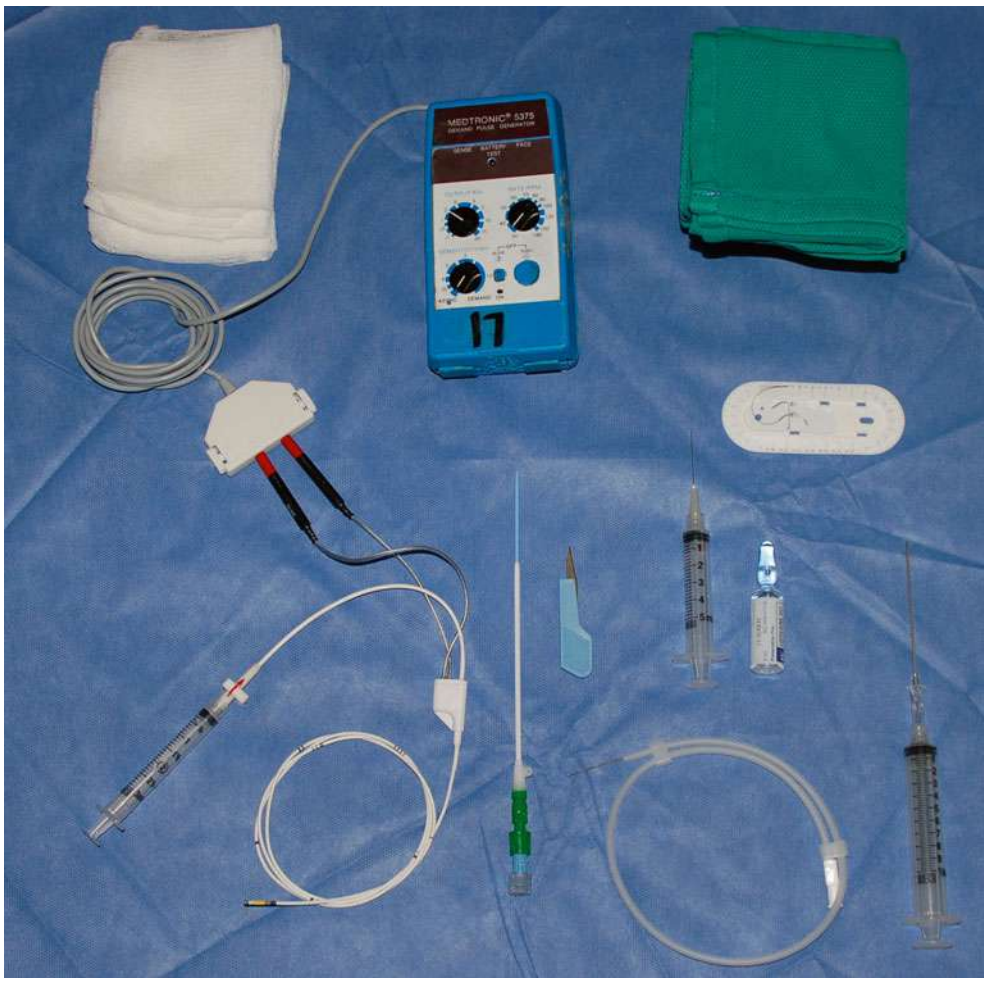

Fig. 32.1 Materials and medications

32.4 Procedure

1. Select appropriate catheterization site.
   - Four sites provide superior access to right atrium: brachial, femoral, internal jugular, and subclavian veins. The right internal jugular and left subclavian veins, respectively, are the most direct routes and, thus, the most commonly used in the emergency department setting.
2. If clinical situation permits, explain the procedure fully to the patient and obtain consent.
3. Sterilize the skin and apply sterile drapes.
4. Anesthetize the intended insertion site with lidocaine.
5. Prepare the pacing wire by inflating and deflating the balloon with 1–1.5 mL of air using a 3-mL syringe, and confirm that locking lever functions correctly to maintain balloon inflation (Fig. 32.2).
6. Using ultrasound guidance by means of a probe covered in a sterile sheath, insert the introducer needle into the vein while applying negative pressure to the attached 10-mL syringe (as with standard central line placement) (Fig. 32.3).
7. When flashback of blood is visualized in the syringe, remove the syringe, stabilizing the introducer needle firmly in place.
8. Pass the guidewire through the needle to a depth of 10–15 cm.
9. Holding the guidewire securely in place, remove the introducing needle.
10. Using a scalpel, make a small incision (approximately the width of the catheter to be used) through the dermis at the insertion site of the guidewire.
    - Avoid severing the guidewire by facing the sharp edge of the scalpel *away* from the guidewire.
11. Pass a dilator over the guidewire to make a tract in the skin and then remove the dilator.
12. Pass the introducer catheter over the guidewire until the hub is in contact with the skin and then remove the guidewire (Fig. 32.4).
13. Attach the collapsible sterile sheath to the hub of the introducer catheter.
14. Attach the positive and negative connectors of the pacing wire to their respective terminals on the cable(s) connected to the generator.
15. Set generator rate at 80 beats/min, output to 5 mA, and sensitivity to 3 mV.
16. Insert the free end of the pacing wire into introducer catheter with the balloon deflated and slowly advance the wire.
17. When the tip of the wire is within the superior vena cava, inflate the balloon, lock the valve to ensure that the balloon remains inflated, and continue to advance the pacing wire.
18. Closely watch the cardiac monitor for signs of capture (i.e., a wide QRS preceded by a pacing spike).
    - Markings on the pacing wire indicate its depth of insertion. At 20 cm, the wire should be in the right ventricle. If there is no capture observed by 25 cm, then the wire should be slowly withdrawn and advanced again.
19. Once capture is achieved, deflate the balloon.
20. Advance the wire a further 1–2 cm.
21. Coil the wire that remains outside the extended sterile sheath and suture it loosely but securely to the skin.
22. Suture the introducer hub to the skin.
23. Reduce generator output to zero and then increase it slowly to determine the minimum pacing threshold (i.e., the minimum voltage at which capture is achieved) and then set the output to twice this value.
24. Set the generator rate.
25. Stabilize the generator near the catheterization site.
    - For example, the generator can be attached to an intravenous (IV) fluid stand to the right of the patient's head if the pacing wire has been placed via the patient's right internal jugular vein.
26. Obtain a chest x-ray to confirm pacing wire position and absence of potential complications (e.g., pneumothorax) (Fig. 32.5).
27. Obtain a 12-lead ECG (Fig. 32.6).

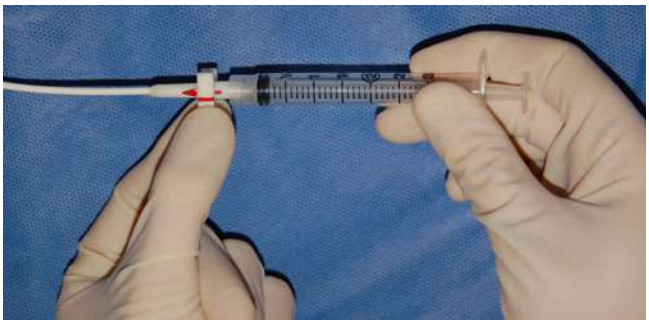

Fig. 32.2 Pacing wire balloon locking lever

Fig. 32.3 Insertion of introducer needle into internal jugular vein with ultrasound guidance

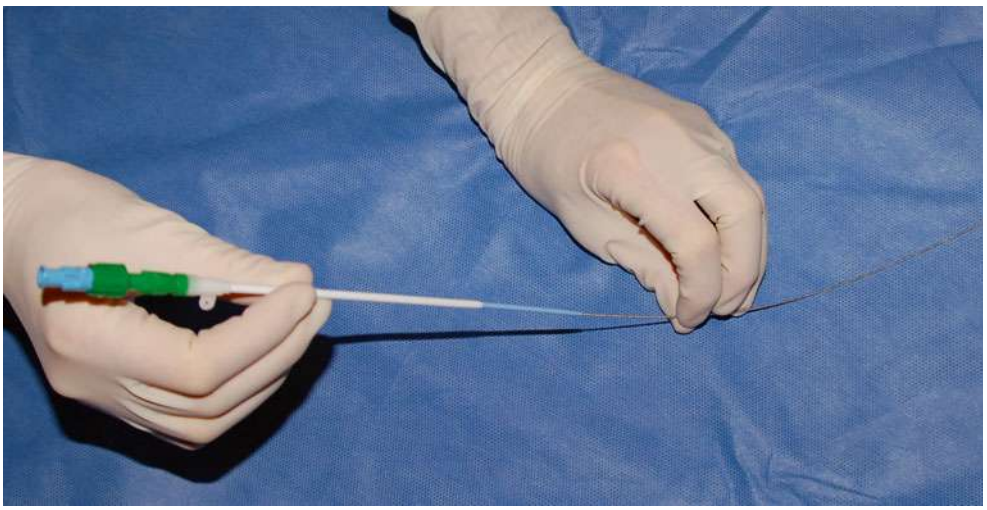

Fig. 32.4 Placement of introducer catheter over guidewire

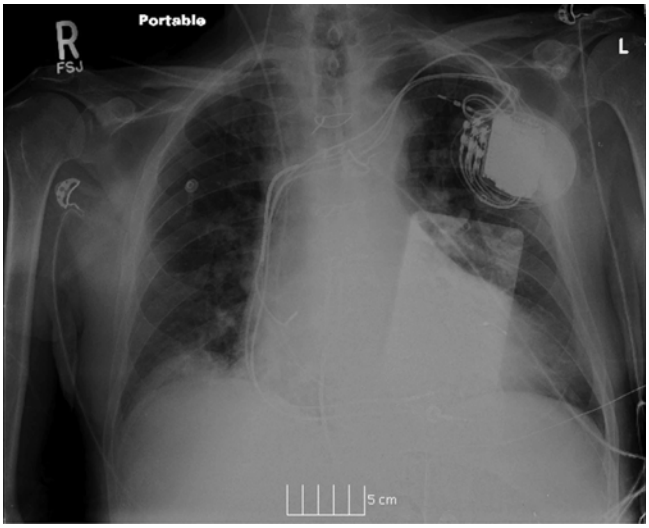

Fig. 32.5 X-ray displaying correct position of pacing wire in the right ventricle

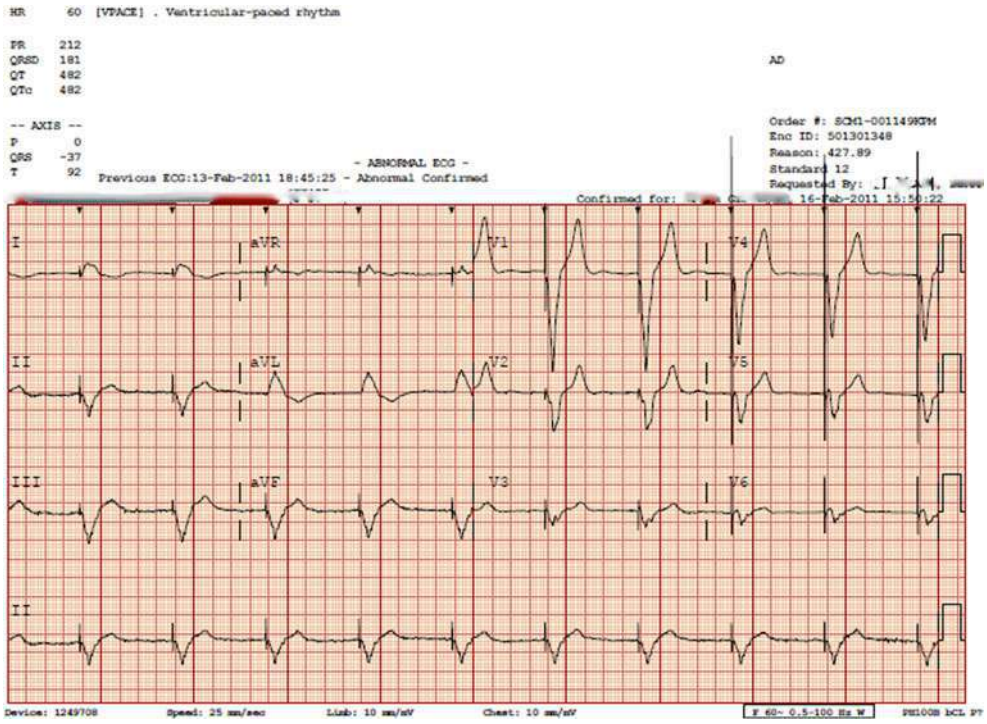

Fig. 32.6 Cardiac monitor displaying wide-complex QRS and pacing spike

32.5 Complications

- Arterial puncture
- Venous thrombosis
- Thrombophlebitis
- Pulmonary embolus
- Pneumothorax
- Fracture of guidewire with embolization
- Hemothorax
- Thoracic duct laceration
- Indwelling line infection (local or systemic)
- Dysrhythmias/premature ventricular contractions (PVCs)
- Insertion of pacing wire into coronary sinus/pulmonary artery
- Left (rather than right) ventricular pacing through atrial septal defect/ventricular septal defect (ASD/VSD)
- Septal perforation
- Ventricular perforation
- Entrapment/twisting of preexisting permanent pacing wires by the temporary pacing wire
- Balloon rupture
- Chordae tendineae rupture
- Pacing wire fractures
- Loss of capture owing to wire displacement or fracture
- Generator failure

32.6 Pearls and Pitfalls

- Confirm all necessary equipment is present and functional before beginning the procedure. If a dedicated transvenous pacing set is unavailable, individual components required (e.g., pacing catheter, connector cables, introducer catheter, guidewire) may need to be "cannibalized" from several different procedure kits in an emergency department.
- Become familiar with the locking valve for the pacing wire balloon before beginning the procedure. This small part is essential for inflating and deflating the balloon but is frequently difficult to adjust owing to stiffness (and all the more so if covered in blood).
- The ideal position for the pacing wire tip is against the diaphragmatic aspect of the right ventricle between its midpoint and its apex. This position is confirmed by development of an LBBB pattern with left axis deviation on the cardiac monitor. If difficulty is encountered achieving this position, the tip of the pacing wire can be moved to the right ventricular outflow tract, where the wire is less stable, but adequate pacing can still be achieved. In this latter position, the threshold will need to be kept relatively high and the monitor should display an LBBB pattern with an inferior axis.
- The use of balloon-tipped catheters enables the catheter tip to be directed with blood flow and its position (and successful capture) to be confirmed with a cardiac monitor. Nonballoon-tipped catheters, however, have typically required insertion with guidance of either constant ECG monitoring or fluoroscopy.

Selected Reading

Aguilera PA, Durham BA, Riley DA. Emergency transvenous cardiac pacing placement using ultrasound guidance. Ann Emerg Med. 2000;36(3):224–7.

Birkhahn RH, Gaeta TJ, Tloczkowski J, Mundy T, Sharma M, Bove J, Briggs WM. Emergency medicine-trained physicians are proficient in the insertion of transvenous pacemakers. Ann Emerg Med. 2004;43(4):469–74.

Dalsey WC, Syverud SA, Hedges JR. Emergency department use of transcutaneous pacing for cardiac arrests. Crit Care Med. 1985;13(5):399–401.

Harrigan RA, Chan TC, Moonblatt S, Vilke GM, Ufberg JW. Temporary transvenous pacemaker placement in the Emergency Department. J Emerg Med. 2007;32(1):105–11.

Silver MD, Goldschlager N. Temporary transvenous cardiac pacing in the critical care setting. Chest. 1988;93(3):607–13.

Pericardiocentesis (Optional: Ultrasound Guidance)

33

Katrina Skoog Nguyen, L. Connor Nickels, and Rohit Pravin Patel

33.1 Indications

- Treatment of hemodynamic compromise from cardiac tamponade
- To diagnose the cause or presence of a pericardial effusion

33.2 Contraindications

- Absolute
  - Aortic dissection
  - Need for immediate surgery for trauma patients
- Relative
  - Coagulopathy
  - Anticoagulant therapy
  - Thrombocytopenia

33.3 Materials and Medications

- Antiseptic (e.g., Chloraprep)
- 1 % lidocaine
- 25-gauge needle, 5/8-in. long
- 18-gauge catheter-type needle, 1½-nch long
- Syringes (10, 20, and 60 mL)
- Ultrasound (US) machine and cardiac/phased array probe
- Sterile US probe cover
- Cardiac monitor

33.4 Procedure

1. Identify the point of maximal effusion with US. Evaluate for hypoechoic or anechoic (dark) effusion around the heart, between the pericardial sac and the myocardium. A patient with hemodynamic compromise from a pericardial effusion or tamponade will have right ventricular collapse, septal bulging, and dilation of the inferior vena cava. Diastolic collapse of the right ventricular free wall can be absent in elevated right ventricular pressure and right ventricular hypertrophy or in right ventricular infarction.
2. Measure the distance from the skin surface to the effusion border to assess the expected needle depth.
3. Choose the needle trajectory based on the point of maximal effusion in the path with the fewest intervening structures. The most commonly used approaches are left parasternal, apical, and subxiphoid. For complex loculated posterior pericardial effusions, optional techniques such as transatrial and transbronchial may be performed by specialists. These types of loculated effusions can occur in autoimmune diseases, infective pericarditis, after cardiac surgery, and after radiotherapy.
4. Sterile preparation: prepare the skin of the entire lower xiphoid and epigastric area with antiseptic. Prepare the US transducer with a sterile sleeve.
5. Local anesthetic: if the patient is awake, anesthetize the skin and planned route of the needle.
6. Pericardial needle insertion: depends on approach used.

K.S. Nguyen, DO
Northwest Community Hospital, Arlington Heights, IL, USA
e-mail: katrinaskoog@hotmail.com;

L.C. Nickels, MD, RDMS (✉) • R.P. Patel, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; rohitpatel@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_33

33.4.1 Subxiphoid Approach (Fig. 33.1)

- The US transducer is placed just inferior to the xiphoid process and left costal margin.
- Insert the needle between the xiphoid process and the left costal margin at a 30–45° angle to the skin.
- Aim for the left shoulder.

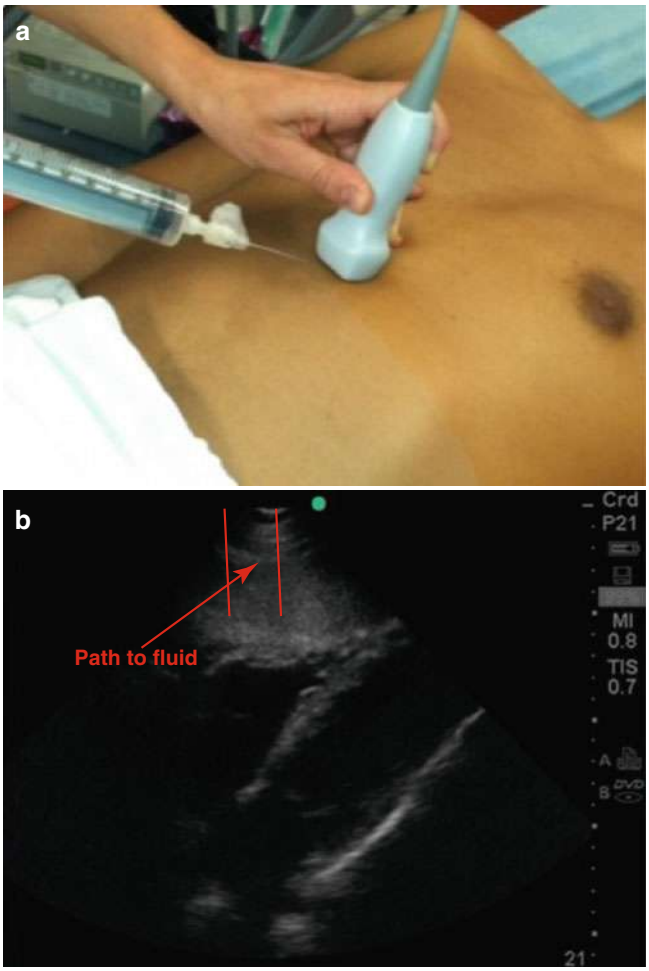

Fig. 33.1 Subcostal approach (**a**) and corresponding ultrasound image (**b**). *Red line* through liver to obtain fluid (no fluid present on ultrasound image)

33.4.2 Apical Approach (Fig. 33.2)

- The US transducer is placed at the patient's point of maximal impulse and aimed at the patient's right shoulder for a four-chamber view of the heart.
- Insert the needle in the fifth intercostal space 1 cm lateral to and below the apical beat, within the area of cardiac dullness.
- Aim for the right shoulder.
- A Mayo Clinic review showed, in 80 % of total effusions, that the distance to the effusion was least and the size was maximal in the apical approach [1].

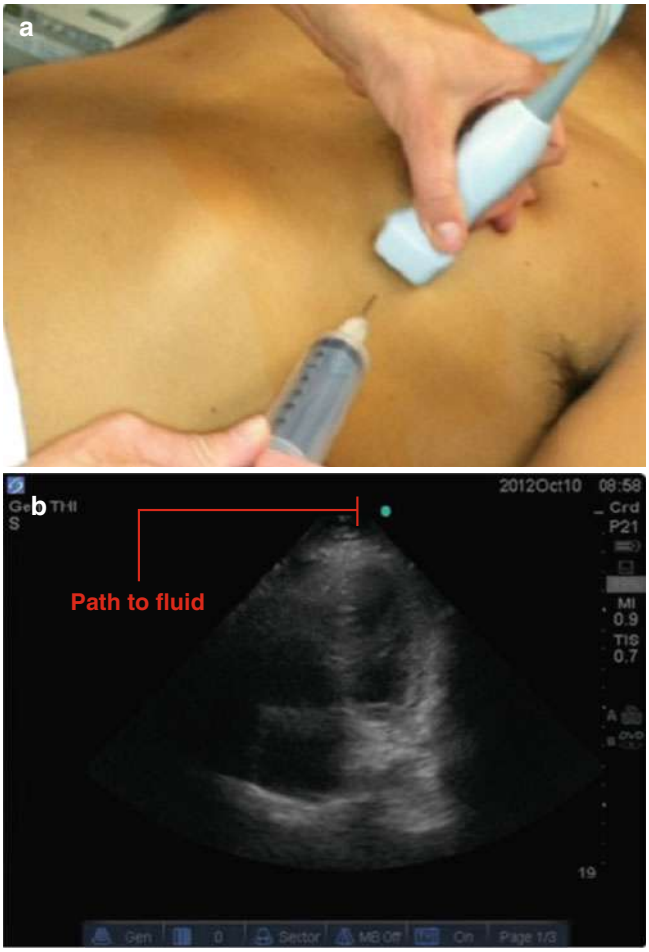

Fig. 33.2 Apical approach (**a**) and corresponding ultrasound image (**b**). *Red line* with minimal distance to fluid if present (no fluid present on ultrasound image)

33.4.3 Parasternal Long-Axis Approach

- The US transducer is placed obliquely on the left sternal border between the fourth and fifth ribs with the transducer indicator aimed at the right shoulder.
- Insert the needle perpendicular to the skin in the fifth intercostal space medial to the border of cardiac dullness.
  - Visualize and feel a giving way as the needle penetrates the pericardium. Removal of fluid confirms successful entry. Remove fluid with the goal of restoring hemodynamic stability. Aspiration of fluid should result in improvement in blood pressure and cardiac output.
  - Remove the catheter and apply a dressing. Optional: Place a pigtail catheter using the Seldinger technique for continued drainage.

33.5 Complications

- Blind techniques are associated with 20 % morbidity and 6 % mortality. The complication rate with US-guided approaches is less than 5 % [2].
- Any vital structure within reach of the pericardial needle has the potential for injury: pneumothorax; hemothorax; coronary vessel laceration; hemopericardium; heart chamber lacerations; intercostal vessel injury; dysrhythmias; ventricular tachycardia; puncture of the liver, diaphragm, or gastrointestinal tract; bacteremia; purulent pericarditis; air embolisms; and pleuropericardial fistulas.

33.6 Pearls and Pitfalls

- Pearls
  - A scoring index is available that can be obtained at initial presentation in patients without hemodynamic compromise that identifies those who require pericardial effusion drainage later in the course of treatment. The scoring index uses echocardiographic findings, etiological information, and the size of the effusion at end-diastole [3].
  - Take care to avoid the left internal mammary artery, which travels in a cephalad-to-caudad direction 3–5 cm lateral to the left sternal border [4].
- Pitfalls
  - Tamponade should always be considered in the differential diagnosis of shock or cardiac arrest owing to pulseless electrical activity (PEA). Complications of acute coronary syndrome, aortic dissection, and decompensations in chronic advanced constriction may also need treatment [5].
  - When cardiac compensation mechanisms are exhausted, small increases in pericardial volume can lead to an increase in ventricular diastolic pressure, systemic and pulmonary congestion, and decrease preload and cardiac output.
  - Use US-guidance rather than a blind or an electrocardiogram (ECG)-alone-guided approach for pericardiocentesis to significantly decrease the risk of injury to vital structures [4].
  - Do not delay patient transport to the operating room to perform pericardiocentesis on a traumatically induced effusion unless the patient is hemodynamically unstable and on the verge of cardiac arrest.

References

1. Tsang TS, Enriquez-Serano M, Freeman WK, Barnes ME, Sinak LJ, Gersh BJ, et al. Consecutive 1127 therapeutic echocardiography guided pericardiocentesis: clinical profile, practice patterns, and outcomes spanning 21 years. Mayo Clin Proc. 2002;77:429–36.
2. Guo K, Ding ZP, Tan J. Trans-pleural pericardiocentesis: revisiting an old technique. Catheter Cardiovasc Interv. 2011;78:815–8.
3. Halpern DG, Argulian E, Briasoulis A, Chaudhry F, Aziz EF, Herzog E. A novel pericardial effusion scoring index to guide decision for drainage. Crit Pathw Cardiol. 2012;11:85–9.
4. Seferović PM, Ristić AD, Imazio M, et al. Management strategies in pericardial emergencies. Herz. 2006;31:891–901.
5. Sagristà-Sauleda J, Mercé AS, Soler-Soler J. Diagnosis and management of pericardial effusion. World J Cardiol. 2011;3:135–43.

Ultrasound Evaluation of Pulmonary Embolism and Heart Strain

34

Giuliano De Portu, L. Connor Nickels, and Marie-Carmelle Elie

34.1 Indications

- Undifferentiated respiratory distress
- Unexplained hypotension
- Evaluation for cardiac tamponade, pericardial effusion

34.2 Contraindications

- Relative
  - Morbidly obese patients
  - Patients with chest wall deformities
  - Patient with subcutaneous emphysema, pneumopericardium
  - Combative or altered patients

34.3 Materials and Medications

- Portable ultrasound machine with appropriate probes (phased array)
- Ultrasonic gel for the probe
- Drapes for the patient (if the condition allows)

34.4 Procedure

1. Begin with the parasternal long-axis view. If possible have the patient turn to the left decubitus side (that helps "move" the heart closer the chest wall).
2. Find the phased array probe, select the "cardiac examination" on the machine, and make sure the orientation marker is pointed toward the right shoulder of the patient.
3. Place the probe along the left side of the sternum over the fourth to sixth intercostal space. This should produce the image shown in Fig. 34.1.
4. Measure the diameter of the ventricle during the end of diastole (normal values, 21 mm ± 1 mm; any measurement >25–30 mm is abnormal).
5. Using the aorta as a landmark, evaluate the structures starting with the pericardium (bright white line around the heart), making sure there is no fluid around it. (Fluid above the aorta indicates a pericardial effusion. Fluid below the aorta indicates a pleural effusion.) If a hypoechoic or anechoic stripe appears in the anterior side of the heart, it is most likely a fat pad. Fluid seen "all around" categorizes it as an effusion.
6. While keeping the probe in the same place, rotate the probe marker 90° clockwise toward the left shoulder to obtain the parasternal short-axis view (Fig. 34.2).

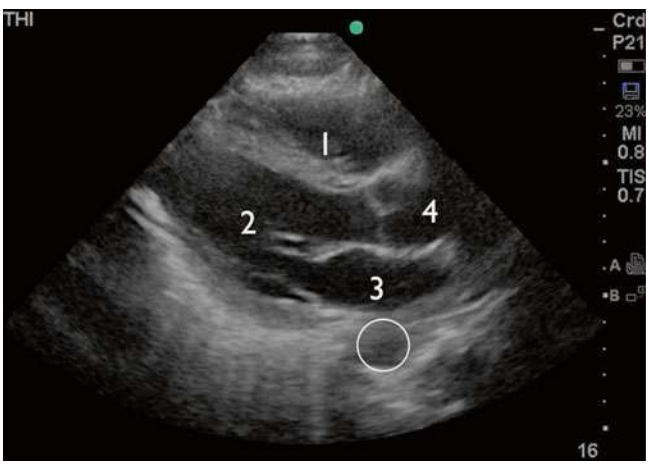

Fig. 34.1 Parasternal long-axis view (PSLA). *1* right ventricle, *2* left ventricle, *3* left atrium, *4* aortic outflow track; *circle* descending aorta (right atrium is not visualized in the PSLA)

G. De Portu, MD • L.C. Nickels, MD, RDMS
M.-C. Elie, MD, RDMS (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: gdeportu@ufl.edu; cnickels@ufl.edu; elie@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_34

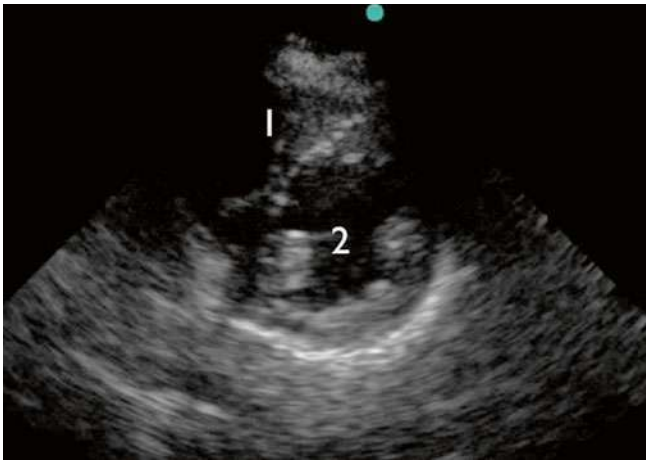

Fig. 34.2 Parasternal short-axis view (PSSA). *1* right ventricle, *2* left ventricle

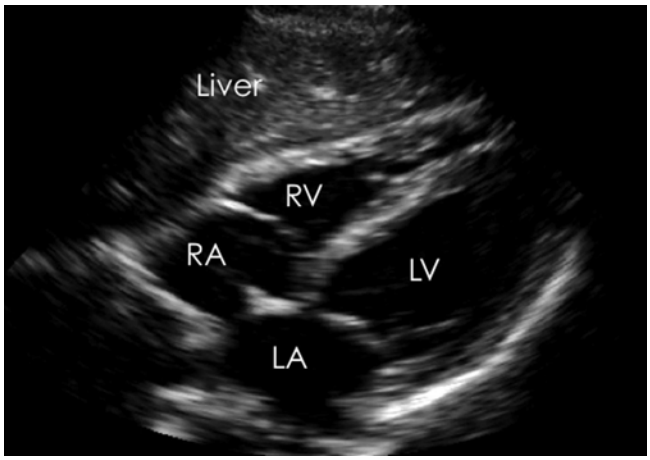

Fig. 34.4 Subcostal/subxiphoid view. *RV* right ventricle, *RA* right atrium, *LV* left ventricle, *LA* left atrium

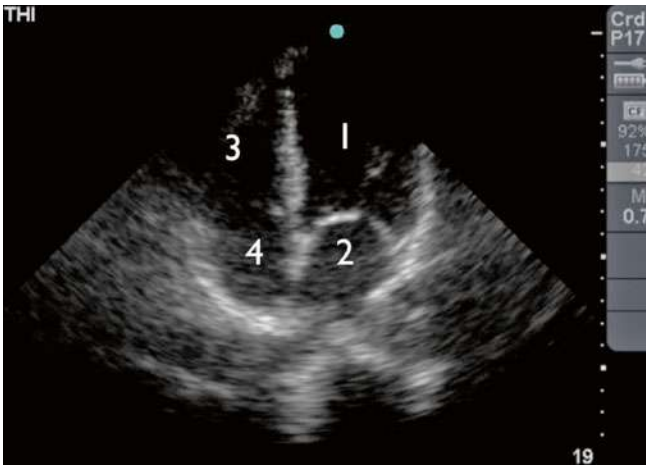

Fig. 34.3 Four-chamber view (4C). *1* left ventricle, *2* left atrium, *3* right ventricle, *4* right atrium

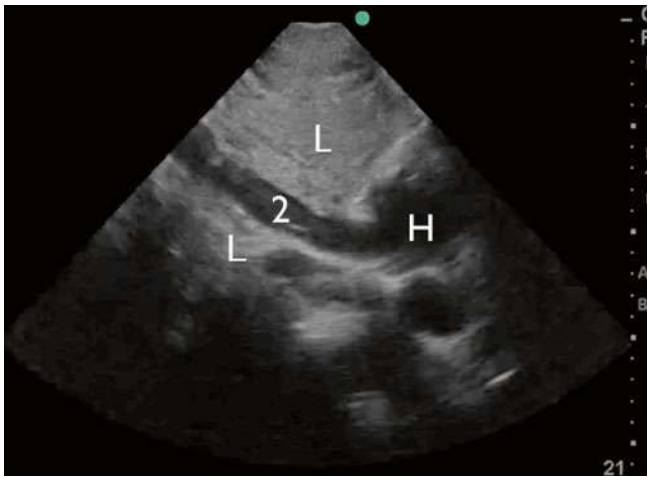

Fig. 34.5 Inferior vena cava. *L* liver, *2* inferior vena cava, *H* heart

7. The parasternal short-axis view will give information on the contractility of the heart. The right ventricle should be anterior and to the left and the left ventricle to the right. The normal position of the septum bows slightly toward the right ventricle.
8. Obtain a four-chamber view. Place the orientation marker to the patient's left.
9. Palpate for the point of maximum impulse (PMI) and place the transducer. All four chambers should appear in one view (Fig. 34.3).
10. Compare the sizes of the ventricles and note any difference. Notice also the interventricular septum. The normal right-to-left ventricular ratio is less than 0.5. (In an unstable patient, this is probably the most useful because both ventricles can be quickly visualized and the ratio compared.)
11. A subxiphoid approach is also possible. Make sure the orientation marker is toward the right side of the patient (Fig. 34.4).
12. Use the liver for orientation. The right side of the heart will be nearest to the liver (think that the liver is on the right of the body); again note for any differences in the size ratio.
13. Next take a look at the inferior vena cava (IVC) by placing the curved array transducer just inferior to the xiphoid in a longitudinal fashion. (Switch from the subxiphoid view to the IVC by rotating the probe counterclockwise until the IVC is seen.) Evaluate the IVC. During normal physiological inspiration, the drop in intrathoracic pressure "pulls" blood into the heart, thus decreasing the relative IVC size. If something is preventing venous return, such as a massive PE, collapse will not be as evident and the suspicion for PE increases (fluid overload and increased central venous pressure [CVP] will also account for this finding). Normal IVC diameter is 1.2–2.3 cm, and total collapse and greater than 50 % collapse are normally visualized. An increase in IVC size and less than 50 % or no change has been correlated with increased right atrial pressures (11 to >20 cm Hg) (Fig. 34.5).

34.5 Findings

A right heart that is "strained" or pumping against a higher resistance owing to a PE will show some or all of these changes:

1. Right ventricular dilation (Fig. 34.6).
2. Right ventricular hypokinesis (especially of the middle segment), McConnell's sign but normal motion of the apex.
3. Tricuspid regurgitation.
4. Abnormal septal motion: deviated toward the left ventricle (normally it relaxes during diastole toward the right ventricle); as pressure increases, the right ventricle will not empty properly and septal flattening can be seen.
5. Dilated IVC with little or complete loss of changes in diameter with respiration (variability); the IVC collapses less than 50 % during inspiration.

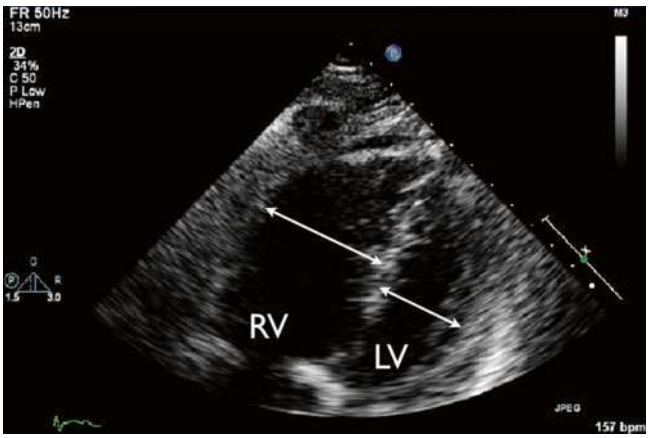

Fig. 34.6 Four-chamber view that shows right ventricular dilatation due to pulmonary embolism. Normally, the LV is greater than the RV during diastole, but in the case of increased pressure, the RV will be enlarged. *RV* right ventricle, *LV* left ventricle

34.6 Complications

- Ultrasound per se has been shown to cause no direct complications with proper use unless it is interfering with advanced airway or life-saving procedures.

34.7 Pearls and Pitfalls

- Pearls
  - Proper interpretation of the images is necessary because some normal anatomy could be confused with a positive finding. Make sure to properly identify the structures because a common mistake is to identify the aorta thinking it is the IVC.
  - Depth: increase to be sure to visualize all structures.
- Pitfalls
  - Ultrasound is user dependent, and the acquisition of images will vary with each user, creating the possibility of false-negatives if images not properly acquired.
  - Reversal of the orientation owing to transducer misplacement could "reverse" the anatomy and allow mistakenly identification of structures.
  - An enlarged right ventricular wall (>5 mm) is considered hypertrophied and would indicate a more chronic etiology of a right ventricular dilation.
  - Gain problems: adjust so the heart chambers are anechoic and the walls are echogenic.

Selected Reading

Kline JA. Thromboembolism In: Tintinalli J, Stapczynski J, Ma OJ, Cline D, Cydulka R, Meckler G, editors. Tintinalli's emergency medicine. 7th ed. New York: McGraw Hill; 2012.

McConnell MV, Solomon SD, Rayan ME, Come PC, Goldhaber SZ, Lee RT. Regional right ventricular dysfunction detected by echocardiography in acute pulmonary embolism. Am J Cardiol. 1996;78:469–73.

Reardon RF, Joing SA. Cardiac. In: Ma OJ, Mateer J, Blaivas M, editors. Emergency ultrasound. 2nd ed. New York: McGraw Hill Professional; 2007. p. 109–44.

Vanden Hoek TL, Morrison LJ, Shuster M, et al. Part 12: cardiac arrest in special situations: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2010;122(18 Suppl 3):S829–61.

35 Pacemaker Evaluation in the Emergency Department

Joseph D. Romano and Christian Coletti

35.1 Pacemaker Function

There are approximately 500,000 implanted cardiac pacemakers in the USA and another 100,000 are implanted each year [1]. In the emergency department, a physician may be faced with a pacemaker that is not functioning appropriately. To understand the ways in which a pacemaker can malfunction or lead to medical complications, first it is important to understand how pacemakers work when they do so appropriately.

A common pacemaker system is composed of a pulse generator and insulated wire leads that originate in the pulse generator and end within the myocardium. The pulse generator is implanted in the pectoral region. It contains circuitry and the battery and creates the electrical impulses that depolarize the myocardium. The leads leave the pulse generator and are directed to the heart by following the venous system. Leads may terminate in the right atrium, the interventricular septum of the right ventricle, or biventricular pacing; a third lead navigates the coronary sinus to the left ventricular wall (Figs. 35.1 and 35.2).

Early pacemakers produced impulses only at a set rate. Today, however, virtually all pacemakers have sensing leads that detect intrinsic activity and react with electrical impulses when the intrinsic intervals fall outside of a set time threshold. Some pacemakers are programmed to allow for rate changes based on physical activity. To simplify the classification of different pacemaker types, the North American Society of Pacing and Electrophysiology and British Pacing and Electrophysiology Group (NASPE/BPEG) developed a five-letter code to describe each pacemaker (Table 35.1) [2].

J.D. Romano, MD (✉) • C. Coletti, MD
Departments of Emergency Medicine and Internal Medicine,
Christiana Care Health System,
Newark, DE, USA
e-mail: jromano@christianacare.org; chris.coletti@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_35

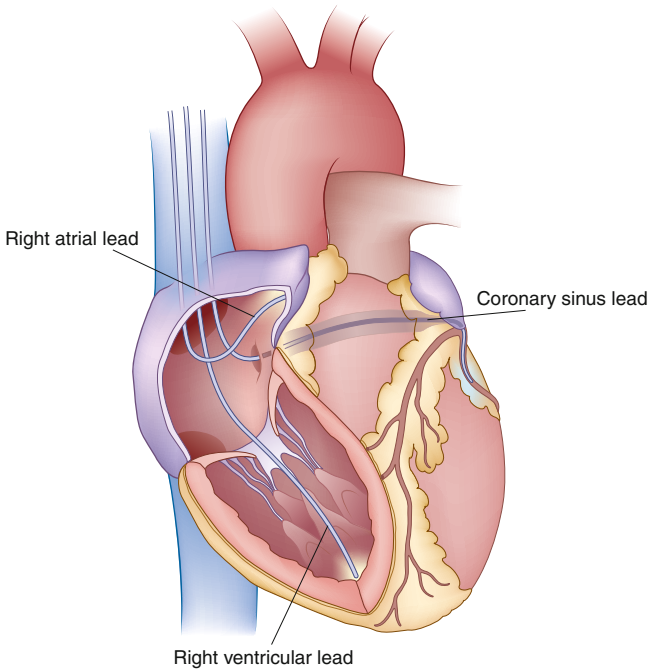

Fig. 35.1 Common pacemaker lead insertion sites

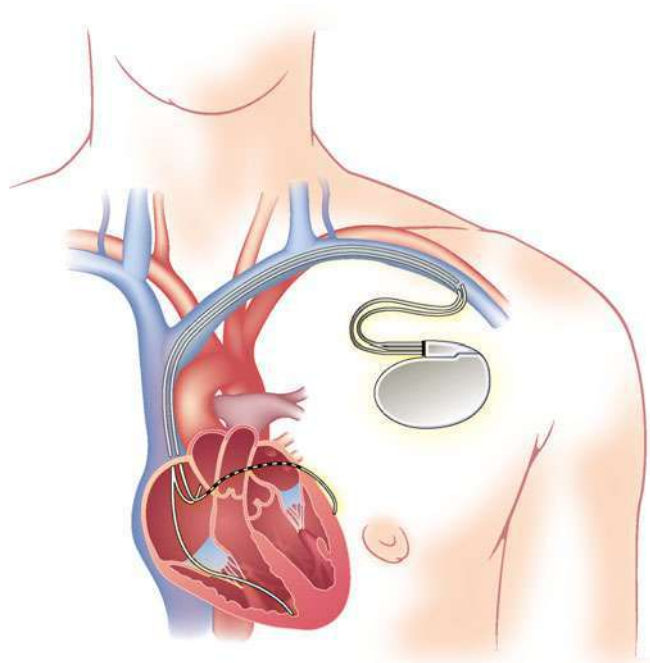

Fig. 35.2 Common generator location and path of pacemaker leads

Table 35.1 NASPE/BPEG five-letter pacemaker code

Paced chamber	Sensed chamber	Response to sensing	Rate modulation	Multisite pacing
A = atrium	A = atrium	T = triggered	R = rate modulation	A = atrium
V = ventricle	V = ventricle	I = inhibited	O = none	V = ventricle
D = dual	D = dual	D = dual		D = dual
O = none	O = none	O = none		O = none

Reproduced with permission from Bernstein et al. [2]

35.2 Common Pacemaker Codes [3]

- VVI: ventricular paced and sensed, inhibited by normal intrinsic pacing; used for patients with a need for pacing if bradycardia occurs
- VVIR: ventricular paced and sensed, inhibited by normal intrinsic pacing; adaptive to changes in intrinsic rate for physically active patients
- DDD: dual-chamber paced and sensed, inhibited by normal intrinsic atrial and ventricular electrical activity; used in third-degree atrioventricular (AV) block to allow for more physiological conduction
- DDDR: dual-chamber paced and sensed, inhibited by normal intrinsic atrial and ventricular electrical activity; adaptive to changes in intrinsic rate; used primarily in sinoatrial (SA) node dysfunction to closely mimic normal, adaptable heart conduction

35.3 Pacemaker Malfunction

Malfunctioning pacemakers can have complicated programming that is best altered after thorough "interrogation" by a trained electrophysiologist. This type of evaluation is beyond the scope of this chapter and, in many clinical settings, subspecialty support is not readily available. In the emergency department, it is vital to stabilize the patient and identify common pacemaker system issues. Pacemaker system malfunctions can be identified as a failure to sense, failure to pace, failure to capture, or pacing at an inappropriate rate [4]. Another common abnormality associated with pacing is the "pacemaker syndrome" [3].

- Failure to sense:
  - Oversensing: The sensor interprets external stimuli as a normal ventricular rate leading to inappropriate inhibition. This may be due to lead fracture, fibrosis of the lead tip, or lead dislodgment. Cross talk is present when the atrial stimulus is sensed on the ventricular lead causing inappropriate inhibition. Oversensing is rarely due to failure of the generator.
  - Undersensing: Present when there is constant pacing despite cardiac activity that has not exceeded threshold. This can be due to low-amplitude intrinsic cardiac activity, lead dislodgment, battery depletion, or metabolic abnormalities.
- Failure to pace:
  - Lack of pacing despite appropriate sensing of the intrinsic electrical activity or activity below the threshold rate.
  - Typically caused by lead fracture, battery depletion, or fibrosis of the lead tip.
  - It is rare for any specific part of the device to fail, but suspicion should be high if the patient had recent radiation therapy, electrocautery, defibrillation, electroshock therapy, magnetic resonance imaging (MRI), lithotripsy, or transcutaneous electrical nerve stimulation (TENS) treatments.
  - Rarely due to battery depletion.
- Failure to capture:
  - A lack of cardiac response despite appropriate sensing and subsequent pacer stimuli.
  - Commonly due to lead dislodgement, but can be due to myocardial perforation, lead fracture, fibrosis of the lead tip, poor lead placement, battery depletion, and antiarrhythmic medications.
  - Functional failure to capture occurs when pacer stimuli fall within the refractory period of previous depolarization.
- Pacing at an inappropriate rate:
  - Likely due to an endless-loop reentry tachycardia known as pacemaker-mediated tachycardia.
    - More common in DDD pacemakers.
    - Initiated by a premature ventricular stimulus, which is carried retrograde through the intrinsic conduction system to the atrioventricular node (AVN) and then the atria. This conduction is sensed by the atrial lead and causes triggering of pacing back in the ventricle. The ventricular depolarization is then sensed again in the atria, forming a loop. The intrinsic conduction system acts as a retrograde limb and the pacer circuit acts as the anterograde limb of a reentry tachycardia.
- The "pacemaker syndrome":
  - Constellation of symptoms found in 20 % of patients with pacemakers.
  - Symptoms include syncope, near-syncope, dizziness, fatigue, weakness, pain, shortness of breath, and cough.
  - Ventricular rates are poorly timed with atrial activity such that atrial contraction occurs against closed mitral and tricuspid valves.
  - Ventricles lose the benefit of the atrial kick, the atria enlarge, and signs and symptoms similar to congestive heart failure ensue.

35.4 Initial Evaluation of the Patient with a Pacemaker [4]

- History:
  - Symptoms of syncope, near-syncope, chest pain, palpitations, and irritation at the generator pocket
  - Brand and NASPE/BPEG code
  - Date of implantation
  - Location of generator pocket and any previous locations
  - Most recent electrophysiological interrogation
  - Medications that interfere with function, such as flecainide or lidocaine
  - Recent procedures, such as MRI, lithotripsy, or defibrillation
- Physical examination:
  - Check vital signs, listen for heart sounds (paradoxically split S2 is normal).
  - Inspect generator pocket; turning of the leads is associated with Twiddler's syndrome, in which a patient compulsively touches the skin around the device.
  - Look for jugular venous distension.
- Laboratory studies and electrocardiogram (ECG):
  - Obtain a chest x-ray to determine effusions, infiltrates, generator placement, and lead placement or fracture.
  - Electrolyte imbalances must be detected and corrected.
  - ECG: The following are examples of appropriate ECG patterns in patients based on type of pacing [3]:
    - VOO (asynchronous pacing): Regular pacer spikes lead to immediate QRS waveforms. Left bundle branch block (LBBB) is a normal finding in a right ventricular lead location.
    - VVI: Pacer spikes and an LBBB waveform should be seen if the intrinsic rate is below the threshold rate.
    - DDD: Various patterns are possible. If the intrinsic rhythm and intervals are normal, then no pacer spikes will be seen. If the atrial rate is slow and AV delay is normal, then an atrial spike will cause a P wave and a normal QRS. If AV delay is prolonged, then two spikes may be observed: a P wave and a QRS with an LBBB waveform.
- If malfunction is suspected:
  - Obtain intravenous (IV) access, place the patient on a heart monitor, and, if possible, consult cardiology for interrogation and reprogramming.
  - Obtain a ring magnet:
    - Positioning over the generator causes the pacing stimuli to revert to an asynchronous ventricular pacing mode at a set rate [5].

35.5 Interpreting the Type of Malfunction Based on ECG

The first step in evaluating the ECG of a malfunctioning pacemaker is to determine the presence of pacing spikes [6]:

- If pacing spikes are noted on the ECG, determine whether capture is present by ensuring that an appropriate waveform follows each pacer spike and that there is an associated pulse.
  - If capture is present, check the rate:
    - If the rate is appropriate, the pacer is functioning normally.
      - If the rate is slow, suspect oversensing.
      - If the rate is rapid, suspect undersensing or pacemaker-mediated tachycardia.
  - If capture is not present, consider metabolic effects or component failure.
- If pacing spikes are not noted on the ECG, determine whether the patient is in an intrinsic rhythm:
  - If the patient is in an intrinsic rhythm and the rate is appropriate, application of a ring magnet will cause pacer spikes to show up at a set rate. This is normal functioning.
    - If application of the magnet does not cause pacing, suspect mechanical failure.
  - If the patient is not in an intrinsic rhythm, place a magnet over the generator.
    - If magnet application causes pacing, consider oversensing.
    - If it does not, consider mechanical failure.
- If a patient is hemodynamically unstable and application of a magnet leads to stability, keep the magnet in place until the patient is able to have the pacemaker interrogated.

35.6 Management of Pacemaker-Mediated Tachycardia

- After ensuring adequate IV access, placing a heart monitor, and interpreting a baseline ECG, put a ring magnet over the generator [5].
  - If a normal rhythm results, keep the magnet on the chest.
  - If it does not change the rate, attempt isometric pectoral exercises by having the patient press the left hand against the right shoulder.
    - This is an attempt to overstimulate the pacemaker sensor and precipitate inhibition of pacer output.
    - If this is unsuccessful, consider transcutaneous pacing.
    - If transcutaneous pacing is unsuccessful, then the leads may require surgical adjustment or removal.

References

1. Ford-Martin PA, Spiwak AJ. Pacemakers. In: Gale encyclopedia of surgery: a guide for patients and caregivers. 2004. Encyclopedia.com: http://www.encyclopedia.com/doc/1G2-3406200337.html. Accessed 01 Apr 2011.
2. Bernstein A, Daubert J, Fletcher R, et al. The revised NASPE/BPEG generic code for antibradycardia, adaptive-rate, and multisite pacing. North American Society of Pacing and Electrophysiology/British Pacing and Electrophysiology Group. Pacing Clin Electrophysiol. 2002;25:260–4.
3. Marx JA, Hockberger RS, Walls RM, Adams J, editors. Rosen's emergency medicine. 7th ed. Philadelphia: Mosby; 2009.
4. Bonow RO, Mann DL, Zipes DP, Libby P, editors. Braunwald's heart disease: a textbook of cardiovascular medicine. 9th ed. Philadelphia: Saunders; 2011.
5. Roberts JR, Hedges JR, editors. Clinical procedures in emergency medicine. 4th ed. Philadelphia: Saunders; 2004.
6. Kaszala K, Huizer JF, Ellenbogen KA. Contemporary pacemakers: what the primary care physician needs to know [review]. Mayo Clin Proc. 2008;83:1170–86.

Part V

Spine Evaluation

36 Cervical Collar Placement

Justin Bennett and Lars K. Beattie

The cervical spine (C-spine) accounts for the majority of all spinal injuries. In the US prehospital setting, patients are often transported using cervical collars (C-collars) and rigid backboards in trauma and when spinal cord injuries are suspected. After airway, breathing, and circulation, every effort must be made to secure the C-spine.

There are two main subcategories of C-collars: the one-piece and the two-piece. One-piece C-collars include the Stifneck and the Ambu Perfit. Two-piece C-collars have posterior and anterior pieces, with the anterior piece usually the larger of the two. Examples of two-piece collars include the Aspen collar, the Philadelphia collar, and the Miami J collar. The basic features of a C-collar include:

- Adjustable circumference with fasteners (usually Velcro straps)
- Adjustable height with a locking device of different sizes
- Hooks for a nasal cannula
- Exposure of the anterior neck for pulse checks and advanced airway procedures
- Posterior access for cervical palpation
- Padding to protect the soft tissues of the neck

J. Bennett, MD
Department of Emergency Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA
e-mail: jubennet@wakehealth.edu

L.K. Beattie, MD, MS (✉)
Department of Emergency Medicine, University of Florida, Gainesville, FL, USA
e-mail: lars.beattie@ufl.edu

36.1 Indications

- Prehospital suspicion for spine trauma:
  - Emergency department patients or trauma patients who fail clinical rule-out criteria such as NEXUS (National Emergency X-Radiography Utilization Study) and Canadian C-spine rules (see Chap. 37)

36.2 Contraindications

- Absolute:
  - Cervical dislocation with fixed angulation
  - Impaled foreign object in the neck
  - Massive soft tissue swelling in the neck
- Relative:
  - Unsecured airway
  - Surgical airway
  - Vomiting
  - Mandible or soft tissue injuries with potential for airway compromise
  - Preexisting anatomical abnormalities

36.3 Materials and Medications

- Properly fitting C-collar
- Consider:
  - Head blocks, if needed for lateral stabilization
  - Towels or backboard pads for custom support
    - Under shoulders – pediatric patients
    - Under occiput – adults with poor C-spine mobility

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_36

36.4 Procedure (Aspen Collar)

1. Address airway, breathing, and circulation while maintaining in-line immobilization, before placing a C-collar.
2. Gather personnel:
   (a) One person is needed to apply the collar to an awake patient.
   (b) Two or more people may be required when a patient has an altered level of consciousness:
      (i) One to maintain in-line immobilization in the neutral position
      (ii) One to place the C-collar
3. While the C-spine is being held in neutral position, assess the airway before placing the C-collar:
   (a) Anticipate and prepare for airway compromise early to avoid a crash intubation.
   (b) Place airway if necessary.
4. Palpate and inspect the C-spine, head, and shoulders for evidence of trauma *before* placing the C-collar.
5. While maintaining the neutral position of the C-spine, place the C-collar:
   (a) Remove loose clothing, jewelry, and earrings that may cause soft tissue pressure wounds.
   (b) Begin with the piece of C-collar that fits under the occiput (Fig. 36.1).
   (c) Fold the Velcro straps behind the C-collar.
   (d) Gently hold back the hair (Fig. 36.2).
   (e) Slide the occipital section or piece of the C-collar behind the occiput (Fig. 36.3). (Use in-line C-spine stabilization in patients with an altered level of consciousness.)
   (f) Wrap (one-piece) or place (two-piece) the anterior section of the C-collar around the circumference of the patient's neck and snugly under the chin (Fig. 36.4).
   (g) Once the Velcro fasteners are in place, ensure that the height is properly adjusted on the C-collar to minimize C-spine mobility.
      (i) Most C-collars have height adjustments that utilize a locking device that requires releasing a locking mechanism by pulling out.

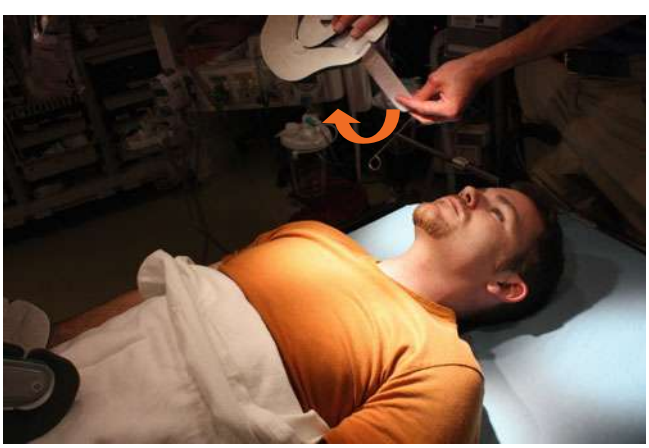

Fig. 36.1 Fold Velcro strap of C-collar posteriorly

Fig. 36.2 Slide occipital section of collar behind occiput flat against stretcher

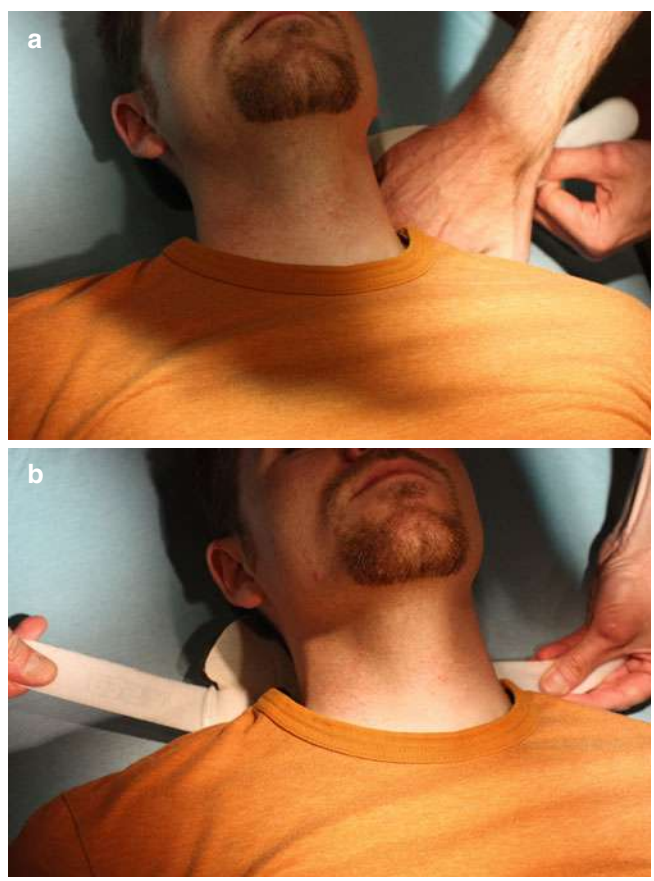

Fig. 36.3 (**a**) Use of hand to guide C-collar behind occiput; (**b**) unfold Velcro strap and position collar behind cervical spine

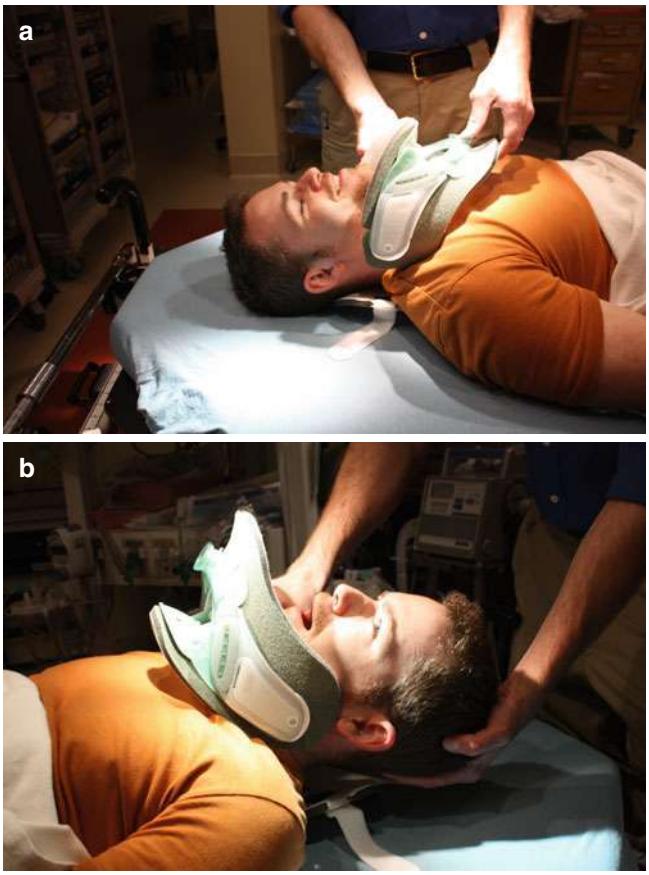

Fig. 36.4 (**a**) Correct orientation of anterior portion of C-collar; (**b**) incorrect orientation of anterior portion of C-collar may cause injury

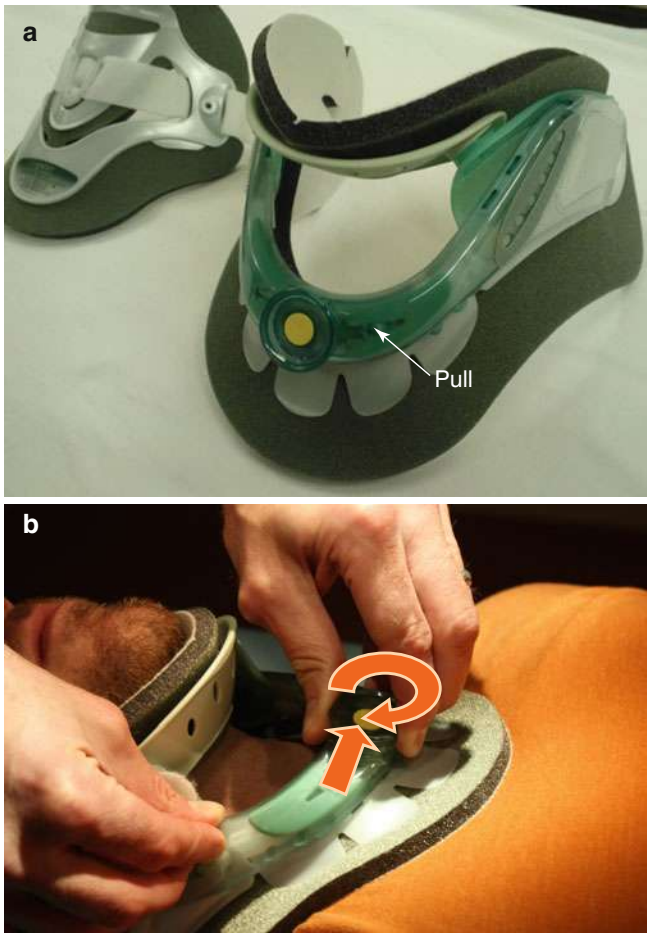

Fig. 36.5 Aspen collar is adjusted with a combination of (**a**) pulling out while (**b**) twisting a round knob at the sternal notch

(ii) The Aspen collar is adjusted with a combination of pulling out while twisting a round knob at the sternal notch (Fig. 36.5).

(iii) Pull out (away from patient) and twist for height adjustment of the Aspen two-piece collar.

(iv) Adjust the height of the Ambu Perfit one-piece collar by simultaneously pulling (away from the patient) the two locking pins out and adjusting the height of the collar before pushing the pins back in (toward the patient) to lock the collar at the desired height (Fig. 36.6).

6. Special cases:
   (a) Depending on the age of the patient, it may be necessary to place towels under the shoulder blades to keep the neck in a neutral position.
   (b) Children:
      (i) Head-to-body ratios are relatively larger than that of adults.
      (ii) Placement on a backboard may cause significant neck flexion because the occiput rests on a flat board.
      (iii) Towels can be placed under a child's shoulders to minimize flexion.

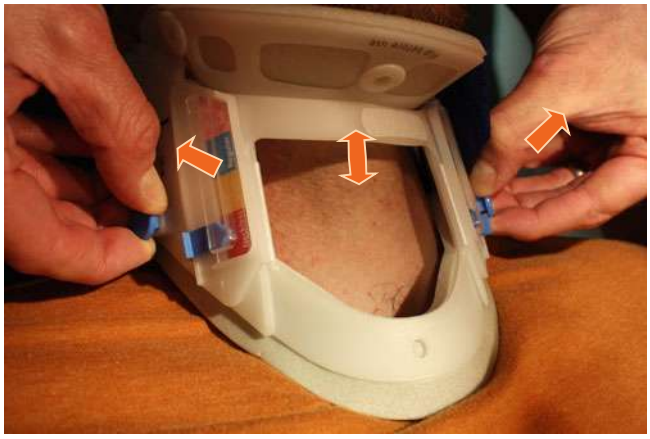

Fig. 36.6 Adjustment the height of the Ambu Perfit one-piece collar by simultaneously pulling (away from the patient) the two locking pins out and adjusting the height of the collar before pushing the pins back in (toward the patient) to lock the collar at the desired height

(c) Adults with excess soft tissue or degenerative changes that prevent C-spine straightening:
   (i) A towel or pad is placed behind the occiput to prevent hyperextension.

36.5 Pearls and Pitfalls

- Pearls
  - Airway, breathing, and circulation should be assessed before placement of the C-collar.
  - A high index of suspicion of C-spine injury is needed in intoxicated or comatose patients and the elderly.
  - Always remember to adjust the height and circumference of a C-collar for a snug fit.
- Pitfalls
  - Spinal immobilization increases the risk of aspiration in vomiting patients.
  - Failure to palpate and inspect the C-spine, head, and shoulders before C-collar placement may delay recognition of an impending airway emergency and conceal signs of critical injury needing rapid assessment and care.
  - Lateral C-spine motion is unsecured unless tape, rubber blocks, or towels are used for support.
  - Overlooking C-spine stabilization during C-collar placement leaves the C-spine at risk of further injury in patients with an altered level of consciousness.
  - C-collars that are too tight may decrease venous return from the head and may increase intracranial pressure.

Selected Reading

American College of Surgeons Committee on Trauma. Advanced trauma life support for doctors. 8th ed. Chicago: American College of Surgeons; 2004.

Hankins DG, Boggust A. Prehospital equipment and adjuncts. In: Tintinalli JE, Stapczynski JS, Cline DM, Ma OJ, Cydulka RK, Meckler GD, editors. Tintinalli's emergency medicine: a comprehensive study guide. 7th ed. New York: McGraw Hill; 2012.

Hoffman JR, Wolfson AB, Todd K, Mower WR. Selective cervical spine radiography in blunt trauma: methodology of the National Emergency X-Radiography Utilization Study (NEXUS). Ann Emerg Med. 1998;32:461–9.

Roberts JR, Hedges JR. Clinical procedures in emergency medicine. 5th ed. Philadelphia: Saunders Elsevier; 2009.

Stiell IG, Clement CM, McKnight RD, et al. The Canadian C-spine rule versus the NEXUS low-risk criteria in patients with trauma. N Engl J Med. 2003;349:2510–8.

Cervical Spine Clearance

37

Braden Hexom and Tatiana Havryliuk

37.1 Indications

- Assessment of the need for radiological imaging in trauma

37.2 Contraindications

- Known unstable cervical spine (C-spine) fracture
- Known unstable ligamentous injury
- Intoxication/altered mental status
- Presence of distracting injury

37.3 National Emergency X-Radiography Utilization Study (NEXUS) and Canadian C-Spine Rules (CCR) to Assess the Need for Imaging

- *The NEXUS clinical criteria*
  1. Tenderness at the posterior midline of the C-spine
  2. Focal neurological deficit
  3. Decreased level of alertness
  4. Evidence of intoxication
  5. Clinically apparent pain that might distract the patient from the pain of a C-spine injury
- The presence of any one of the above findings is considered to be clinical evidence that a patient is at increased risk for C-spine injury and requires radiographic evaluation.
- *CCR*
  For alert (Glasgow Coma Scale [GCS] 15) and stable trauma patients
  1. Any high-risk factor that mandates radiography? *YES → Radiography*
     - Age older than 65 years
     - Dangerous mechanism (Table 37.1)
     - Paresthesias in extremities

     ↓ *NO*
  2. Any low-risk factor that allows safe assessment of range of motion? *NO → Radiography*
     - Simple rear-end motor vehicle crash (MVC)
     - Sitting position in emergency department
     - Ambulatory at any time
     - Delayed onset of neck pain
     - Absence of midline C-spine tenderness

     ↓*YES*
  3. Able to actively rotate neck? *NO → Radiography*
     - 45° left and right

     ↓*ABLE*

     No radiography
- CCR found to have higher sensitivity (99.4 % vs. 90.7 %) and specificity (45.1 % vs. 36.8 %) and a lower rate of imaging (55.9 % vs. 66.6 %) than NEXUS criteria for C-spine clearance of low-risk alert trauma patients.

Table 37.1 Dangerous Mechanism

Fall from elevation ≥3 ft/5 stairs
Axial load to head (e.g., diving)
Motor vehicle crash at high speed (>100 km/h), rollover, ejection
Motorized recreational vehicles
Bicycle crash

B. Hexom, MD (✉)
Department of Emergency Medicine, Mount Sinai Hospital, New York, NY, USA
e-mail: braden.hexom@mssm.edu

T. Havryliuk, MD
Department of Emergency Medicine, University of Colorado Denver, Denver, CO, USA
e-mail: tatiana.havryliuk@ucdenver.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_37

37.4 Type of Imaging

- Bones
  - C-spine computed tomography (CT)—the new gold standard.
  - Plain X-rays—less sensitive than CT; in one study X-rays missed 45 % of injuries that were picked up by CT [1].
- Ligaments
  - C-spine magnetic resonance imaging (MRI)—consider for possible ligamentous injuries and for further evaluation of obtunded patients.
  - Flexion-extension X-rays—less sensitive than MRI for detection of ligamentous injuries in the acute phase. Delayed flexion-extension films are more sensitive for ligamentous injury than those done the day of injury, but less sensitive than MRI.

37.5 Materials

- Cervical collar (C-collar) (Fig. 37.1)

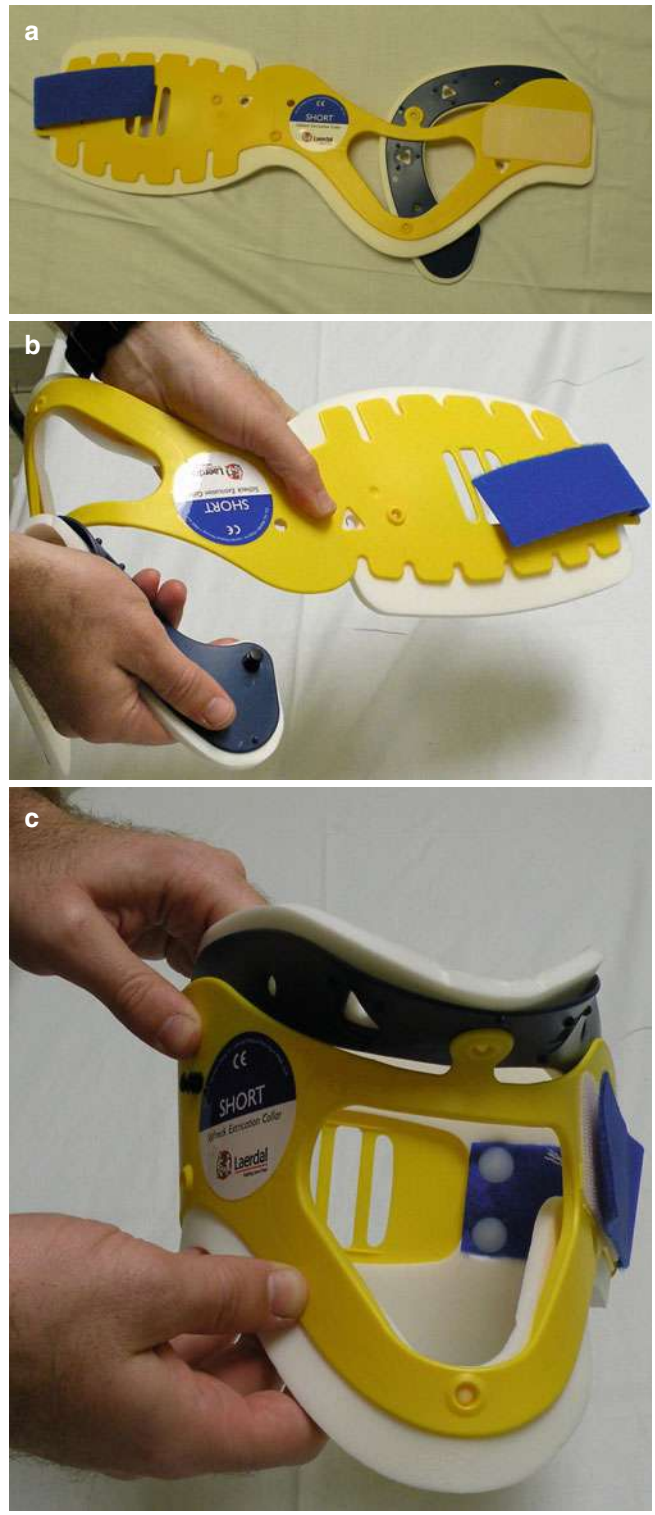

Fig. 37.1 Assembly of C-collar: Select appropriate size (**a**). Invert chinstrap and snap into place (**b**). Assembled C-collar (**c**)

37.6 Procedure (Fig. 37.2)

1. Apply C-collar to stabilize spine. Select appropriate size.
2. Perform a brief neurological examination and assess the patient's mental status; proceed only if both are normal.
3. Loosen the collar to palpate the midline while the patient holds his or her head still. If spinal tenderness exists replace the C-collar and proceed to imaging (Fig. 37.3).
4. Instruct the patient to rotate the neck 45° to each side and flex the neck. If the patient is pain free and with no neurological comprise, the C-collar may be removed and no imaging of the C-spine is required (Fig. 37.4).

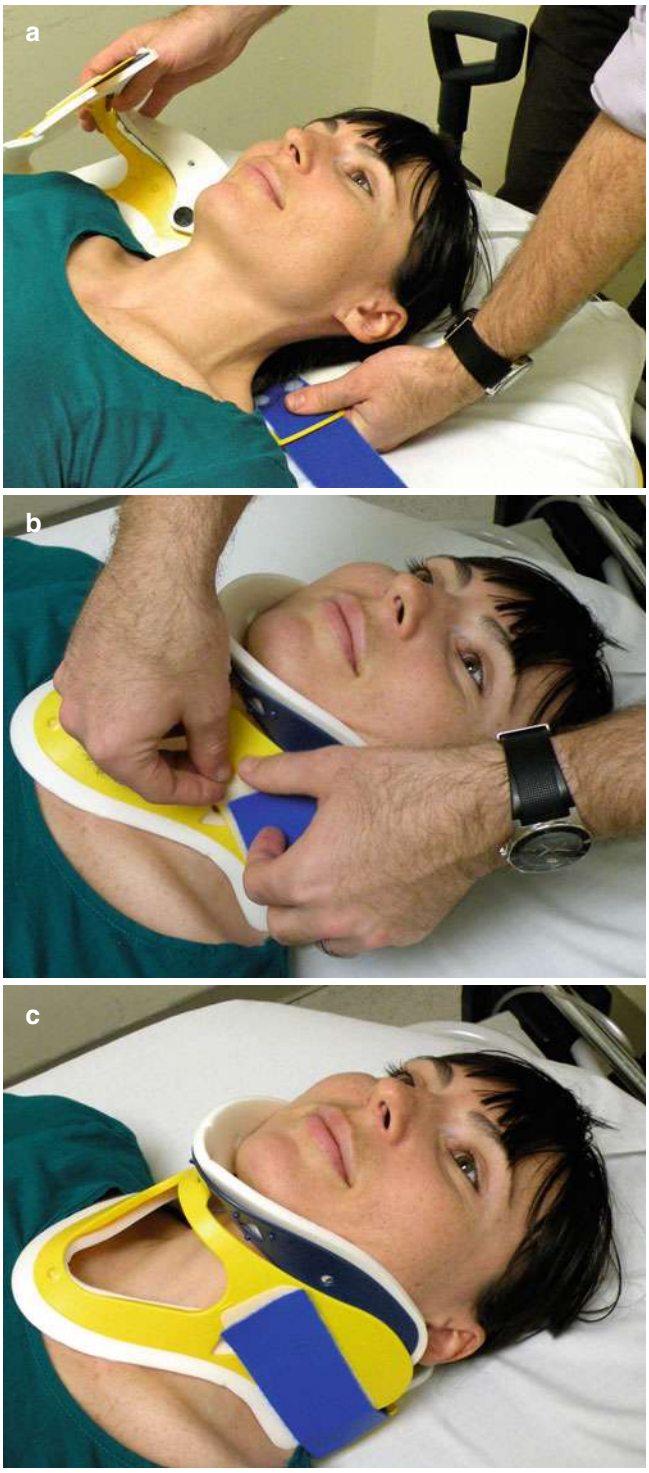

Fig. 37.2 Proper placement of C-collar: Slide C-collar under the neck while keeping neck immobilized (**a**). Secure the collar (**b**). Correct C-collar placement (**c**)

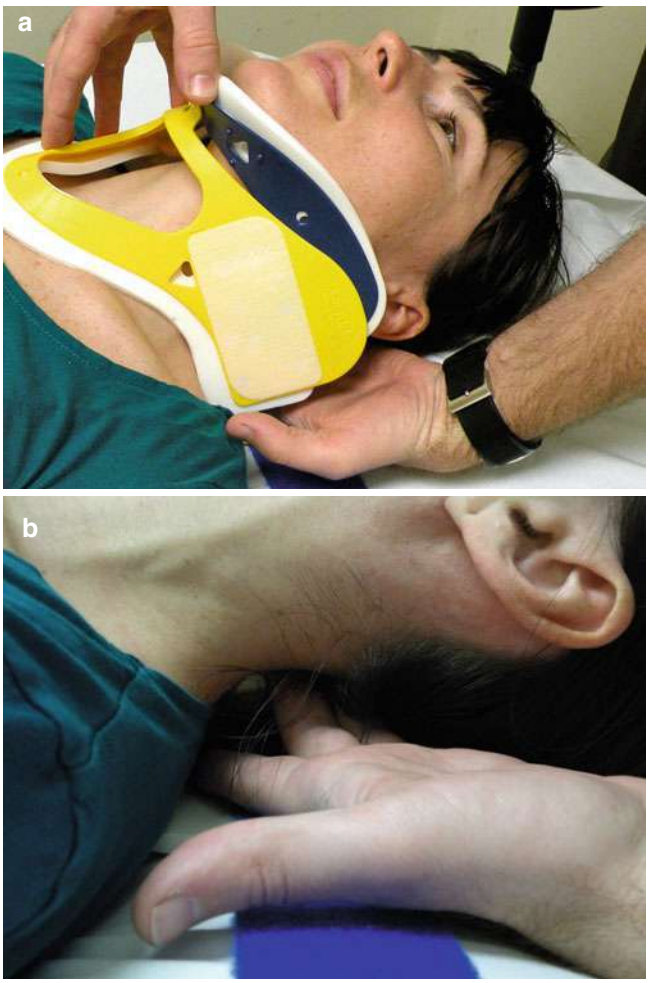

Fig. 37.3 Assessment of C spine tenderness: Loosen C-collar while keeping neck immobilized (**a**). Palpate midline of C spine (**b**)

Fig. 37.4 C-spine clearance: Instruct patient to rotate neck 45° each way (**a**). Instruct patient to flex neck (**b**)

37.7 Complications

- Missing clinically important C-spine injury
- Pressure ulcer from prolonged use of C-collar

37.8 Pearls and Pitfalls

- *Pearls*
  - Select appropriate size of C-collar.
  - Ensure adequate mental status because further imaging might be required in obtunded/intoxicated patients.
- *Pitfalls*
  - Avoid prolonged use of C-collar because this can lead to a pressure ulcer.

Reference

1. Schenarts PJ, Diaz J, Kaiser C, Carrillo Y, Eddy V, Morris Jr JA. Prospective comparison of admission computed tomographic scan and plain films of the upper cervical spine in trauma patients with altered mental status. J Trauma. 2001;51:663–8; discussion 668–9.

Selected Reading

Hoffman JR, Schriger DL, Mower W, Luo JS, Zucker M. Low-risk criteria for cervical-spine radiography in blunt trauma: a prospective study. Ann Emerg Med. 1992;21:1454–60.

Platzer P, Jaindl M, Thalhammer G, et al. Clearing the cervical spine in critically injured patients: a comprehensive C-spine protocol to avoid unnecessary delays in diagnosis. Eur Spine J. 2006;15:1801–10.

Stiell IG, Clement CM, McKnight RG, et al. The Canadian C-spine rule versus the NEXUS low-risk criteria in patients with trauma. N Engl J Med. 2003;349:2510–8.

Stiell IG, Wells GA, Vandemheen K, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. JAMA. 2001; 286:1841–8.

38 Log Roll

Justin Bennett and Lars K. Beattie

38.1 Indications

- Any patient arriving on a rigid backboard in the emergency department
- Assessment of posterior traumatic injuries
- Performed as early as possible after arrival to prevent skin breakdown from pressure ulcers

38.2 Contraindications

- Improperly fitted cervical collar (C-collar) or unsecured cervical spine (C-spine) before log roll
- Unsecured endotracheal tube before log-rolling intubated patients

38.3 Materials and Medications

- Personnel: three or four people
  - One to stabilize the C-spine
  - One or two to roll the patient
  - One to palpate the length of the spine
- Properly fitting hard C-collar
- Trauma shears for removing transport straps

38.4 Procedure

1. Ensure that the airway, breathing, and circulation (ABCs) are established before attempting to remove the patient from the backboard.
2. Gather personnel, at least three, but preferably four people. Larger patients will often require additional assistance.
3. For intubated patients an additional person will be needed to secure the endotracheal tube during the log roll.
4. Ensure the height is properly adjusted on the C-collar to minimize C-spine mobility (Fig. 38.1).
5. Position the stretcher at an ergonomic lifting position for the person responsible for stabilizing the C-spine.
6. Clothing and transport straps should be removed before the log roll to improve visualization of injuries.
7. The patient is asked to cross his or her hands over the chest.
8. C-spine stabilization (at head of bed)
   - Grasp the patient's trapezoids at the midclavicular line.
   - Secure the sides of the patient's head between the operator's forearms.
   - Stabilize the patient's head in neutral anatomical position relative to the body using the operator's forearms (Fig. 38.2).

J. Bennett, MD
Department of Emergency Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA
e-mail: jubennet@wakehealth.edu

L.K. Beattie, MD, MS (✉)
Department of Emergency Medicine, University of Florida, Gainesville, FL, USA
e-mail: lars.beattie@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_38

9. Thoracic and lumbar spine stabilization
   - One or two (preferably two) people should stand next to the patient on the side to which the patient will be rolled.
     - One person: Place hands over the patient's shoulder and hip.
     - Two people
       - First person places hands on the patient's shoulder and hip.
       - Second person places hands on the patient's hip and knees.
   - The decision to roll to the left or right side is determined by injury sites, to minimize injury exacerbation and pain, and to minimize risk of endotracheal tube dislodgment.

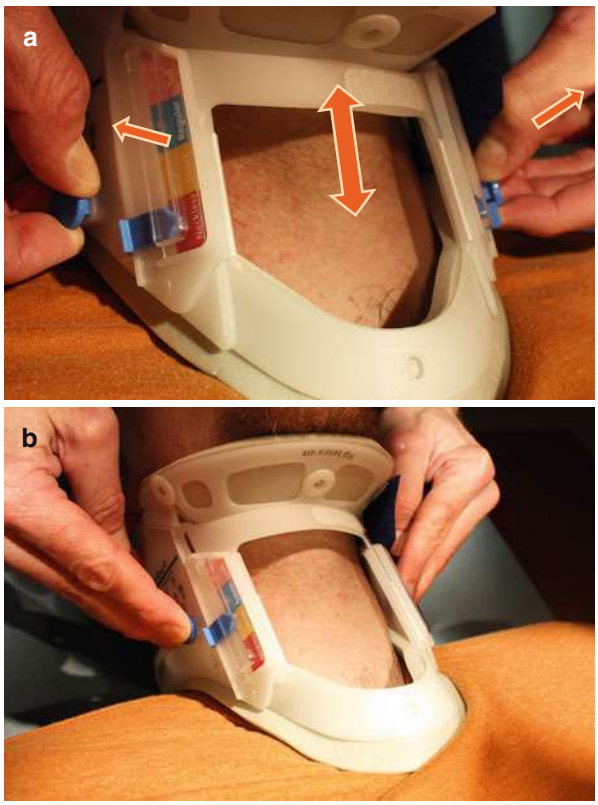

Fig. 38.1 (**a**) The height of the Ambu Perfit one-piece collar is adjusted by simultaneously pulling the two locking pins out (away from the patient) and adjusting the height of the collar, (**b**) then pushing the pins back in (toward the patient) to lock the collar at the desired height

Fig. 38.2 (**a**) Incorrect: Single-point stabilization does not keep head and C-spine in line with torso (**b**, **c**). Correct: The two-point stabilization technique keeps the head and C-spine in line with the torso during the log roll

10. The person at the head of the bed signals the initiation of a synchronized log roll when everyone is in position.
11. Attention should be directed at maintaining a neutral axis along the anatomical line of symmetry during the roll (Fig. 38.3).
12. While the patient is on her or his side, the rigid backboard should be removed and secured to prevent injury of caretakers.
13. The patient's entire posterior should be exposed and examined, taking care to note lacerations and obvious deformities.
14. The entire length of the spine is examined by inspection and then by palpation to assessed for tenderness, step-offs, and deformities (Fig. 38.4).
15. Before returning the patient back onto the stretcher, ensure that any debris, glass, lumps of clothing, or blankets are removed.

Fig. 38.3 (**a**) The interlocking hands of operators performing lateral log roll help maintain spine immobilization and minimize twisting. (**b**) Two-point cervical spine stabilization is maintained in neutral axis relative to the anatomical line of symmetry during the roll. Appropriate bed height adjustment will increase the ease of the procedure

Fig. 38.4 (**a**–**c**) After backboard has been removed, the entire length of the of the spine can be appropriately palpated and assessed for injury to the thoracic and lumbar spine

38.5 Pearls and Pitfalls

- Pearls
  - The log roll should be performed in unison to avoid segmental rotation.
  - The ABCs should be established before initiating the log roll, which is part of the secondary survey in the trauma evaluation.
  - A proper log roll begins with stabilization of the C-spine because it is the most freely mobile part of the spine and, therefore, most frequently injured.
  - Take the patient off the backboard as soon as possible to prevent formation of pressure sores.
- Pitfalls
  - Failure to establish a two-point stabilization between the patient's body and the head leaves the C-spine at risk of further injury during the log roll.
  - Patients should not be left on the backboard for spinal precautions.
  - Failure to ensure that the ABCs are established before attempting a log roll.

Selected Reading

American College of Surgeons Committee on Trauma. Advanced trauma life support for doctors. 8th ed. Chicago: American College of Surgeons; 2008. ISBN 1880696312.

Roberts JR, Hedges JR. Clinical procedures in emergency medicine. 5th ed. Philadelphia: Saunders Elsevier; 2009.

Part VI

Neurologic and Neurosurgical Procedures

39 Burr Hole Craniotomy

Latha Ganti

39.1 Indications

- Extradural hematoma (EDH) or subdural hematoma (SDH) with Glasgow Coma Scale (GCS) <8 and unavailability of timely neurosurgical intervention in the face of increased intracranial pressure
- Signs of increased intracranial pressure [1]
  - Deteriorating level of consciousness
  - Slowing of the pulse
  - Dilating pupils
  - Focal seizures
  - Hemiparesis
  - Extensor posturing of the limbs

39.2 Contraindications

- Absolute
  - GCS >8
  - Immediate availability of neurosurgeon
- Relative
  - Lack of imaging (in this case, decision is guided by neurologic findings and signs of increased intracranial pressure or impending herniation)

39.3 Materials and Medications

- Razor to shave area
- 2 % lidocaine with epinephrine to numb scalp
- 10 % povidone-iodine or chlorhexidine prep
- Light, suction, cautery, dressing tray
- Gelfoam
- Penrose drain
- 3-0 silk
- #10 scalpel blade and #3 handle
- Small self-retractors or rakes
- Drill and drill bits: can be manual (Fig. 39.1) or automatic stopping variety (Fig. 39.2)
- Bone wax or electric cautery apparatus
- Suction apparatus
- Saline irrigation (IV tubing connected to a saline bag with clamp set so flow is low, or a saline syringe flush)

Fig. 39.1 Hudson brace, burr, and perforator

L. Ganti, MD, MS, MBA
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_39

Fig. 39.2 Automatic stopping craniotomy drill set

39.4 Procedure

Preparation

1. Patient should be supine, intubated, with appropriate C-spine precautions
   - Hypertonic saline and/or mannitol can be considered for medical management of increased ICP if/as directed by neurosurgery
2. Determine location and depth for burr hole placement (Figs. 39.3 and 39.4):
   - Have CT images immediately available for viewing
   - Most common location for EDH or SDH is temporal
     - Temporal burr hole placement: 2 fingerbreadths above the ear, 2 fingerbreadths forward (Fig. 39.5)
     - Parietal burr hole placement: 2 fingerbreadths above the ear, 3 fingerbreadths backward (behind the ear)
     - Frontal burr hole placement: 3 fingerbreadths from midline, 3 fingerbreadths above hairline
   - Estimate depth of hematoma by counting the number of slices the hematoma is as seen on CT scan and multiplying by the slice thickness [2]
3. Shave hair around area of hematoma.
4. Prep skin with betadine and chlorhexidine.
5. Anesthetize scalp skin with lidocaine and epinephrine.

   Accessing hematoma
6. Make a vertical incision approximately 4 cm long down to bone.s
7. Use periosteal elevator or end of scalpel blade to scrape muscle and periosteum away from bone.
8. Place self-retaining retractors (or rakes) to keep field open (Fig. 39.6).
9. Using a drill with a clutch mechanism [2], drill through outer table (resistance), diploic space (no resistance), then inner table.
   - Burr hole needs to be placed over the *center* of the hematoma (Fig. 39.7).
10. Control bone bleeding with bone wax; control bleeding from veins and/or muscle with gelfoam sponge or cautery (diathermy).
    - Wound edges may also be cauterized or tamponaded with manual pressure +/− epinephrine gauze.
11. Once in the inner table, separate dura from bone
12. Enlarge opening by switching to a conical or cylindrical burr or use a rongeur

    Evacuating hematoma
13. EDH blood will be visible at this point and should be gently suctioned out.
14. SDH blood will be seen as a tenting of the dura and may be clotted: lift the dura with a hook or make an incision with a fresh scalpel through it to expose the clot and drain.

15. Irrigate with saline (this can be via a hand syringe or via IV tubing connected to a saline bag at low sped flow).
16. Repeat gentle suction.

    Closure

17. For temporal burr holes, ligate middle meningeal artery (if visible) or cauterize.
18. Place Penrose drain (Fig. 39.8) and close the dura first with a 3-0 absorbable suture, ensuring there is no dural leak (will lead to infection if present).
    - A watertight seal of the duraplasty is essential to minimize cerebrospinal fluid leakage.
19. Loosely suture scalp using 3-0 silk.

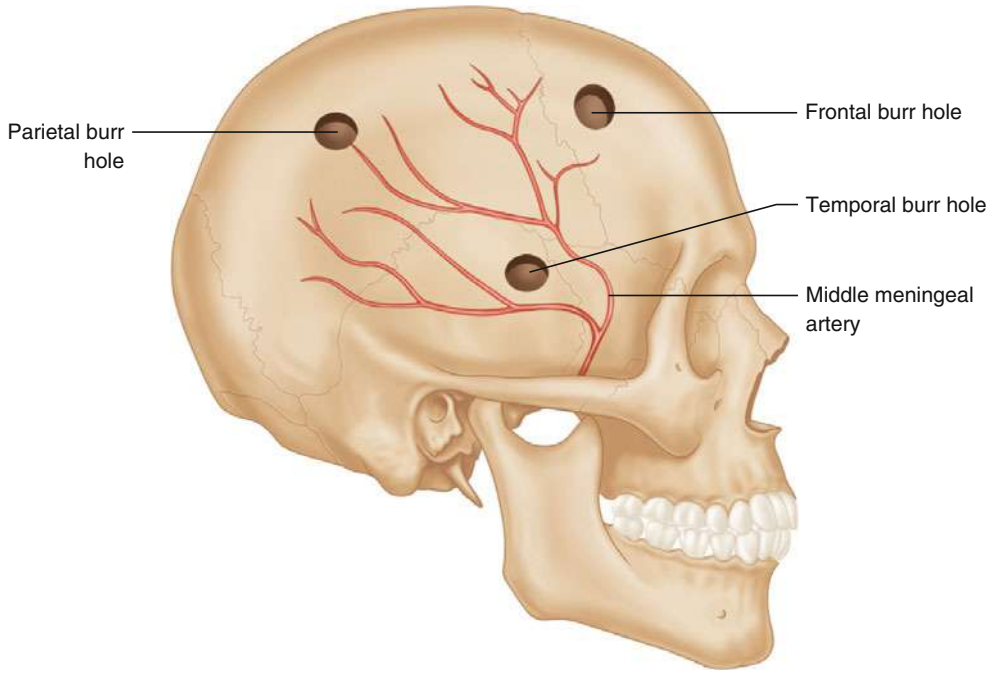

Fig. 39.3 Positions of burr hole placement

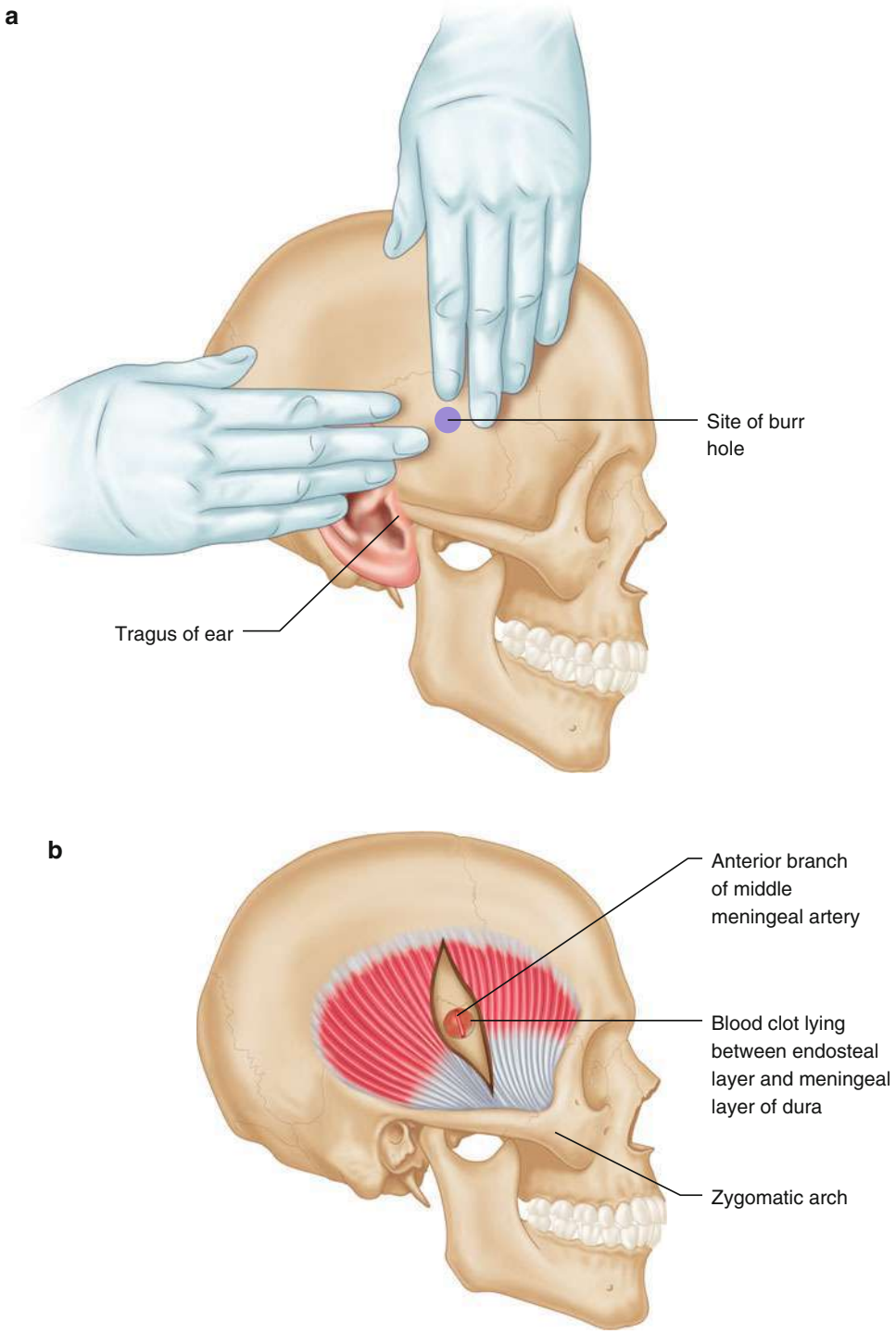

Fig. 39.4 (**a**, **b**) Anatomy for temporal burr hole placement (most common location)

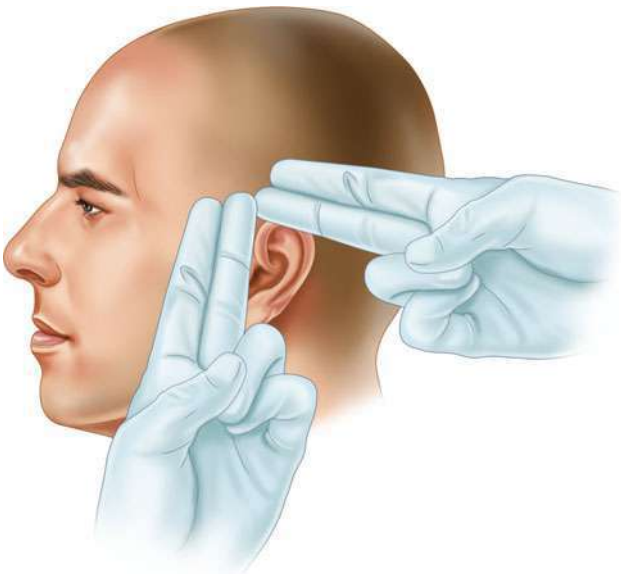

Fig. 39.5 Landmark for temporal burr hole above zygoma, make 5 cm incision over site

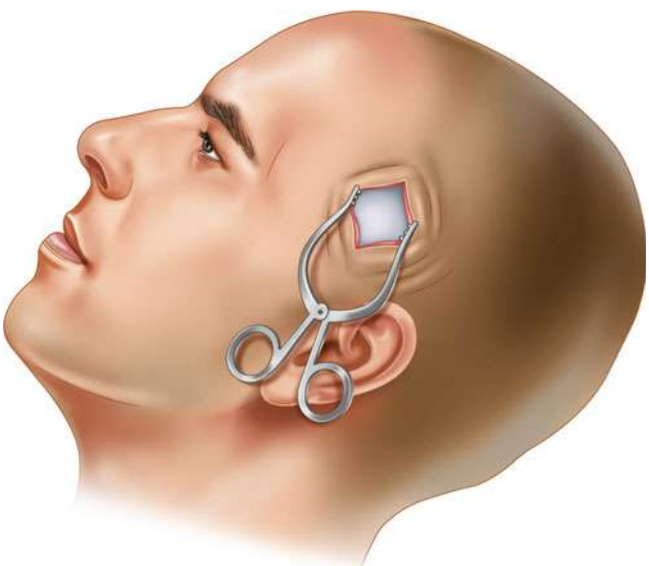

Fig. 39.6 After incision is made, use retractors (preferably self-retaining) to visualize field

Fig. 39.7 Use penetrator drill to get through outer table. Follow up with a burr to get through the rest of the layers. Can also use an automatic stop craniotomy drill

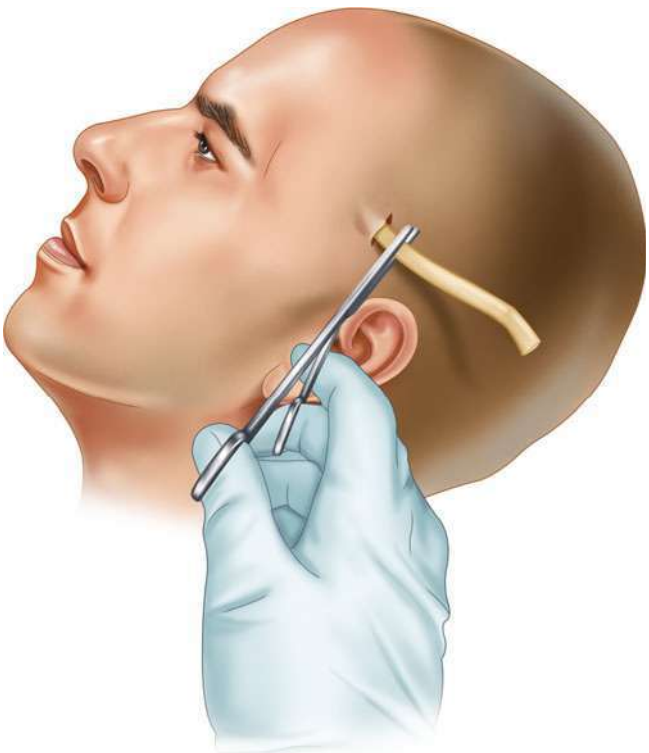

Fig. 39.8 Suture in Penrose drain and close the wound

39.5 Complications

- Brain laceration/perforation
- Temporal artery laceration
- Wrong location (minimized when viewing images during procedure)
- Facial artery laceration
- Meningitis
- Brain abscess

39.6 Pearls and Pitfalls

- Pearls
  - If brain herniates through the burr hole, there is likely a hematoma at another location.
  - Remember this is only a lifesaving procedure that is to be done when timely neurosurgical intervention is not possible. *Do not delay transfer* of the patient
- Pitfalls
  - Bone in temporal area is quite thin; if not using an automatic stop drill, be very careful not to go too far and perforate the brain.

References

1. Wilkinson DA, Skinner MF. The primary trauma care manual for trauma management in district and remote locations. World Health Organization (WHO). http://www.steinergraphics.com/surgical/006_17.6.html. Accessed 22 May 2014.
2. Wilson MH, Wise D, Davies G, Lockey D. Emergency burr holes: "how to do it.". Scand J Trauma Resusc Emerg Med. 2012;20:24. doi:10.1186/1757-7241-20-24.

Selected Reading

http://www.viewmedica.com/vm/pages/library/L_df8516c9#vm_A_ac54d3a1. Accessed 29 June 2014.

40 External Ventricular Drain Placement

Latha Ganti

An external ventricular drain (EVD; also known as a ventriculostomy tube) is placed via a burr hole in one of the ventricles of the brain (Fig. 40.1) to drain excess cerebrospinal fluid (CSF) that causes elevated intracranial pressure.

40.1 Indications

- Emergent need for intracranial pressure (ICP) monitoring and or management
- Enlarged ventricles on neuroimaging with Glasgow coma scale <12
- Subarachnoid hemorrhage with Hunt-Hess grade ≥3 (Table 40.1)
- Coma
- Obstructive hydrocephalus
- Intraventricular hemorrhage
- Signs of increased intracranial pressure

40.2 Contraindications

- Absolute
  - Immediate availability of a neurosurgeon who can do the procedure
- Relative
  - Coagulopathy
  - Scalp infection

L. Ganti, MD, MS, MBA
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

40.3 Materials and Medications

- Intubation equipment and medications for sedation as needed
- Sterile gloves, gown, mask
- Ruler
- Surgical marking pen
- 1–2 % lidocaine with epinephrine to numb scalp
- 5 cc syringe and needles to give anesthetic
- 10 % povidone-iodine solution and swabs
- Razor to shave area
- Fenestrated clear drape
- Scalpel #11 blade (for scalp) and #15 blade (for periosteum) with #3 handle (Fig. 40.2a)
- 4×4 sterile gauze
- Adson forceps
- Mosquito forceps
- Self-retaining eyebrow retractor (Fig. 40.2b)
- 3-0 nylon suture and needle holder and/or skin stapler
- Scissors
- A hand drill with variable chuck
- One or more drill bits with depth guards in 5/32″ (3.97 mm), 13/64″ (5.31 mm), and 1/4″ (6.35 mm) sizes
- A hex wrench for depth guard adjustment
- Ventricular catheter
- Primed ventricular drainage collection system

40.4 Procedure

Patient preparation

1. Patient is intubated and placed supine in neutral position with head of the bed elevated 30–45°.
2. Administer one dose of intravenous antibiotics that covers skin flora.
3. Make precise measurement of where the hole and incision will be made. Most commonly, EVDs are placed in

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_40

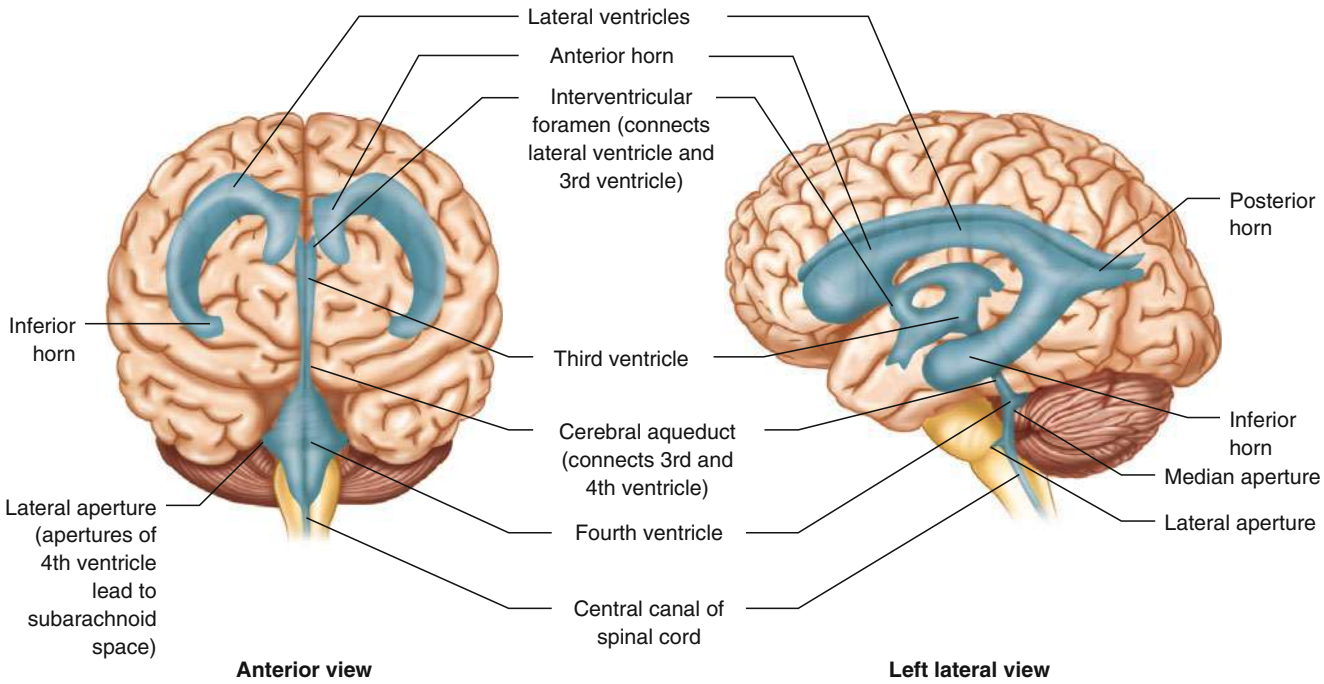

Fig. 40.1 Ventricles of the brain that contain CSF

Table 40.1 Hunt and Hess scale for subarachnoid hemorrhage [1]

Grade	Signs and symptoms
1	Alert and oriented, mild headache, slight or no nuchal rigidity
2	Alert and oriented, moderate to severe headache, nuchal rigidity, no neurologic deficit
3	Drowsiness, confusion, mild focal neurologic deficit
4	Stupor, moderate-severe hemiparesis
5	Coma, decerebrate posturing

the right frontal scalp, as this is the nondominant hemisphere in >95 % of the population.

4. Using a sterile skin/surgical marker, draw a line 11 cm back from nasion and then a point 3 cm to the right of that, which corresponds to the mid-pupillary line. This is called Kocher's point (Fig. 40.3).
5. Once Kocher's point is marked, shave skin and remove loose hairs so field is wide enough for subcutaneous catheter placement and tunneling.
6. Prep area with povidone-iodine.
7. Place sterile clear drape over field.
8. Prep area of incision once more.
9. Infiltrate scalp with 1–2 % lidocaine with epinephrine (the epinephrine acts as a hemostatic agent, keeping surgical field clean) (Fig. 40.4).
10. Make a 2 cm linear stab incision and extend incision to skull (Fig. 40.5).
11. Use eyebrow self-retaining retractors to hold skin edges back.
12. Drill burr hole with hand drill (Fig. 40.6).
13. Irrigate burr hole with sterile saline.
14. Put on a new pair of sterile gloves (prior to handling ventriculostomy catheter).
15. Remove ventricular tubing from sterile package. It has markings on it that are 1 cm apart.
16. Insert ventriculostomy catheter perpendicular to the skull at the point of insertion. Catheter is aimed at the ipsilateral medial canthus of the eye (anteroposterior plane) and tragus of ear (lateral plane).
17. Advance ventriculostomy tube 5–6 cm with stylet from outer skull table so it sits in the anterior horn of the lateral ventricle (Fig. 40.7).
18. Ensure ventriculostomy catheter is draining CSF.
19. Attach metal trocar to tip of ventriculostomy catheter and tunnel the trocar and catheter under the galea approximately 3–5 cm to the right of the original incision. Bring trocar out through separate stab incision in scalp.
20. Remove trocar, make sure ventriculostomy catheter is still draining CSF.
21. Place temporary cap on ventriculostomy catheter to prevent overdrainage of CSF.

22. Close the original incision with either sutures or staples.
23. Secure catheter to the scalp using staples.

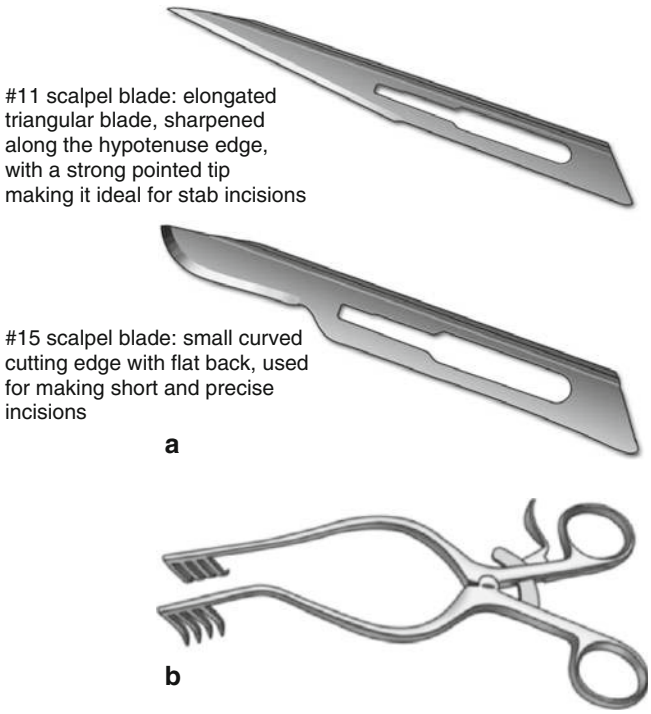

Fig. 40.2 (**a**) Surgical scalpel blades. (**b**) Self-retaining eyebrow retractors

24. Cover incision with sterile transparent dressing.
25. Suture stopcock to ventricular catheter to ensure it does not come off (take care not to make suture too tight or it will occlude ventriculostomy catheter).
26. Remove blind end of stopcock, attach ventricular catheter to ventricular drain system once it is primed.

    Priming, setup, and maintenance of ventricular drainage system

27. Prepare drainage system by priming system with sterile, preservative-free saline.
28. Place zero point of drainage system at midbrain (level of patient's ear tragus) or above midbrain at 15–20 cm of H_20 (target ICP set by neurointensivist and will change depending on patient's situation) (Fig. 40.8).
29. Attach drainage tubing.
30. Patients who require an EVD should be closely monitored by nurses trained and competent in assessment and management of the drain and in recognizing signs of increased ICP in the patient.
31. Assessment of the drainage system should be done a minimum of every 4 h, which includes inspecting the EVD from the insertion site along the entire drainage system, checking for cracks in the system or fluid leaking from the insertion site [2].

    Post-procedure

32. Obtain CT scan of brain to verify placement (Fig. 40.9).

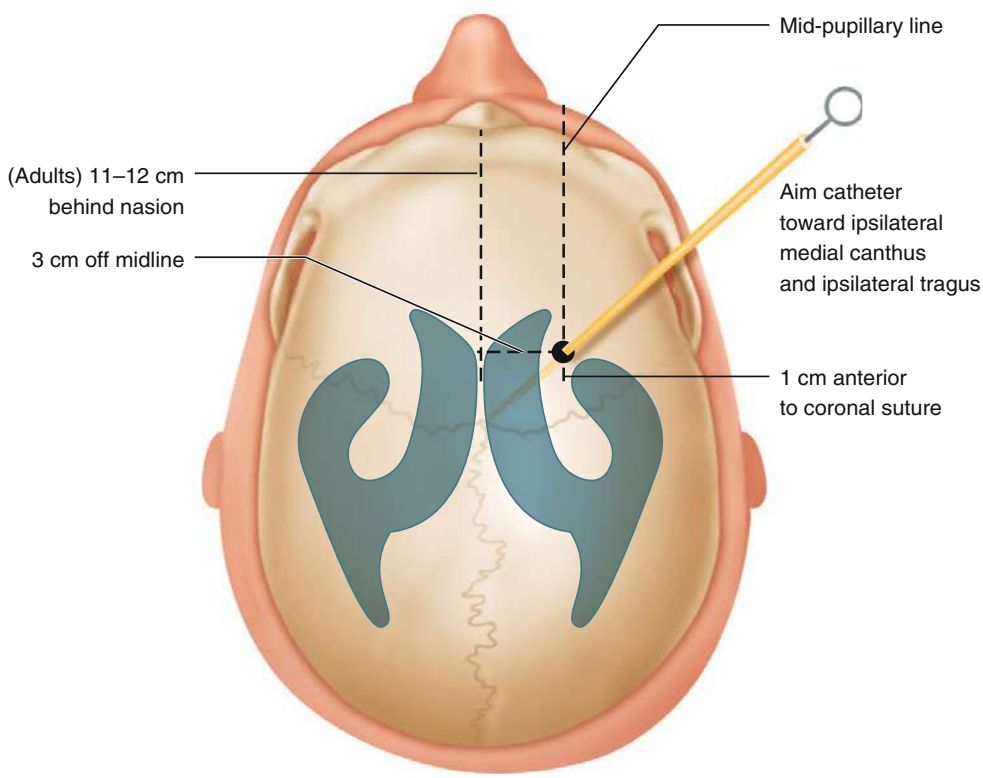

Fig. 40.3 Kocher's point for frontal EVD placement

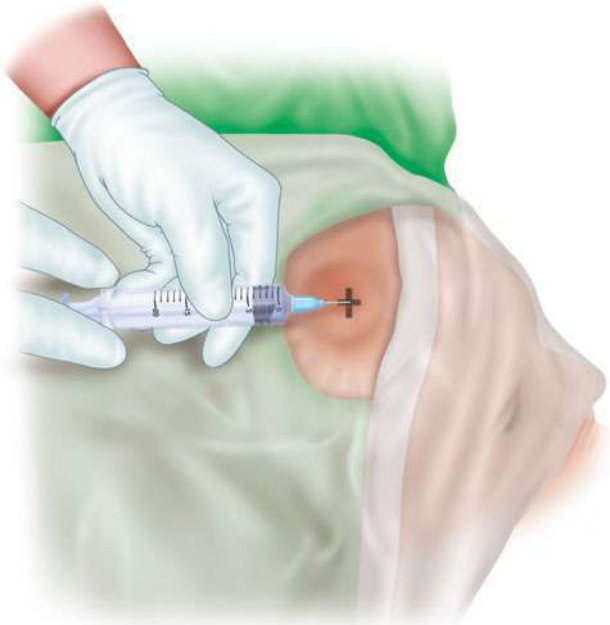

Fig. 40.4 Anesthetize skin

Fig. 40.6 Make burr hole

Fig. 40.5 Make stab incision

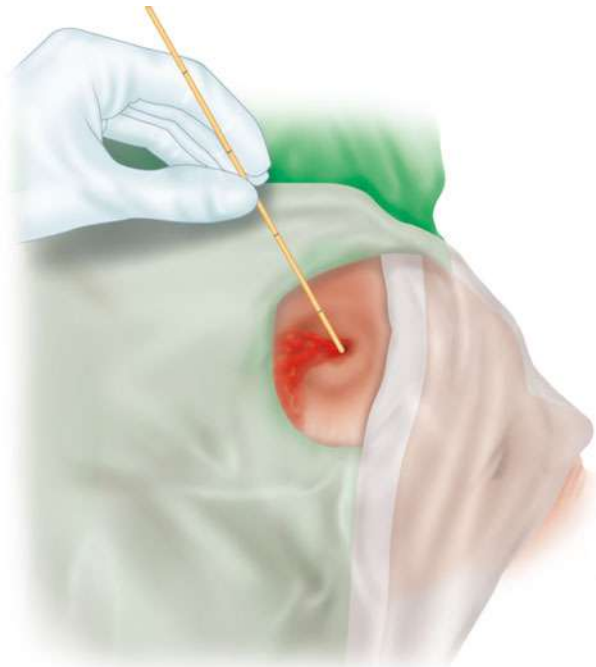

Fig. 40.7 Advance ventriculostomy catheter with stylet

40.5 Complications

- Hemorrhage
- Infection: meningitis, ventriculitis
- Tract hematoma
- Migration/dislodgement

40.6 Pearls and Pitfalls

- Pearls
  - EVD needs to be re-leveled every time the patient moves. Family and all visitors must be informed that any patient movement or change in elevation of head

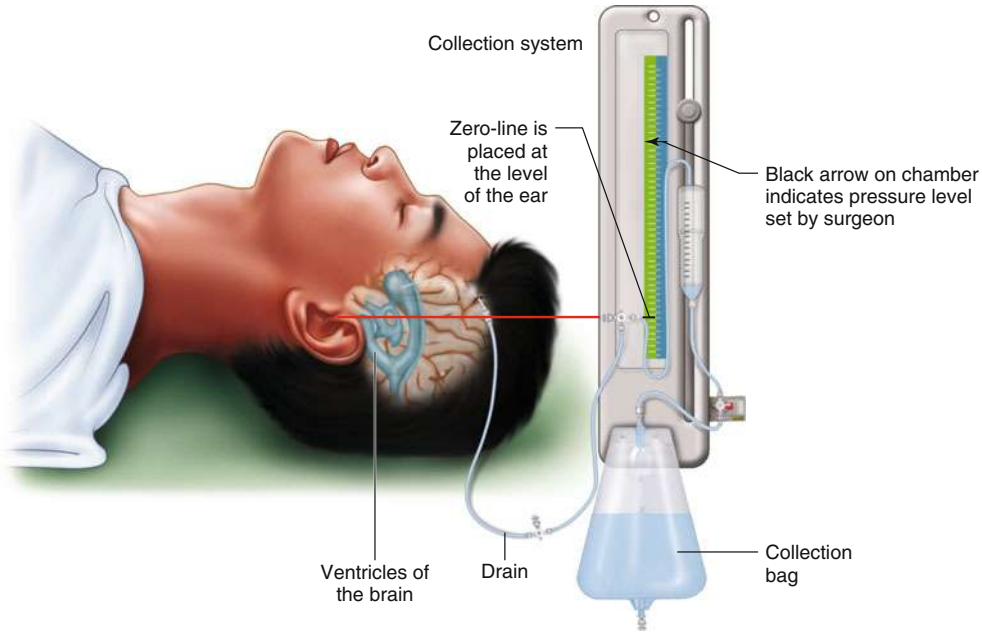

Fig. 40.8 Zero point of drainage system placed at midbrain (level of patient's ear tragus)

Fig. 40.9 CT scan demonstrating typical placement of ventriculostomy catheter in ipsilateral anterior horn of lateral ventricle [3] (Reproduced with permission from National Institutes of Health: Jaffe et al. [3])

of bed will require re-leveling of EVD in order to ensure it is a the appropriate level for CSF drainage.
  - CSF collection chamber should remain upright to prevent reflux/leakage.
  - EVD placement in the parieto-occipital scalp or frontotemporal or trans-sylvian locations is generally not done at the bedside due to higher rate of complications. This is performed in the operating room.
- Pitfalls
  - If the drain is leveled too low: CSF will drain too easily, which can lead to re-rupture of an unrepaired ruptured cerebral aneurysm or cause a subdural hematoma due to shearing of bridging veins.
  - If the drain is leveled too high, CSF will not drain or will not drain at desired rate, leading to hydrocephalus. The ICP will need to be higher in order for the CSF to drain.

References

1. Hunt WE, Hess RM. Surgical risk as related to time of intervention in the repair of intracranial aneurysms. J Neurosurg. 1968; 28(1):14–20.
2. Care of the patient undergoing intracranial pressure monitoring/external ventricular drainage or lumbar drainage. AANN clinical practice guideline series. 2011. http://www.aann.org/uploads/AANN11_ICPEVDnew.pdf. Accessed 27 Sept 2014.
3. Jaffe J, Melnychuk E, Muschelli J, et al. Ventricular catheter location and the clearance of intraventricular hemorrhage. Neurosurgery. 2012;70(5):1258–64.

Lumbar Puncture in Adults

41

Kevin Tench, L. Connor Nickels, and Rohit Pravin Patel

41.1 Indications

- Diagnostic
  - Evaluation for the possibility of a central nervous system (CNS) infection: viral, bacterial, and fungal meningitis and encephalitis
  - Evaluation for inflammatory processes: multiple sclerosis, Guillain-Barré syndrome
  - Evaluation for spontaneous subarachnoid hemorrhage (SAH)
  - Suspicion of CNS diseases: oncological and metabolic processes
- Therapeutic
  - Therapeutic reduction of cerebrospinal fluid (CSF) pressure
  - Procedures requiring lower body analgesia or anesthesia
  - Intrathecal antibiotic administration for some types of meningitis
  - Chemotherapy and methotrexate for some forms of leukemia and lymphomas

41.2 Contraindications

- Presence of infection in tissues at or around puncture site.
- Increased intracranial pressure (ICP) from a space-occupying lesion; patients with signs of cerebral herniation or with potential of increased ICP and focal neurological signs.
- Bleeding diathesis (thrombocytopenia, anticoagulant therapy, hemophilia); may increase risk of spinal hematoma, but level of coagulopathy that increases risk is unclear.
- Patients with cardiorespiratory compromise may worsen owing to position.

Patients with prior history of lumbar surgery, osteoarthritis, ankylosing spondylitis, kyphoscoliosis, or degenerative disk disease might have more success if lumbar puncture is performed by an interventional radiologist using imaging techniques and should be considered.

41.3 Materials and Medications (See Fig. 123.1)

- Spinal needle(s) with stylet
  - Adults: 3.5-in. 20-gauge needle; obese may require 5.0-in. 22- to 24-gauge needle
  - Children: 2.5-in. 22-gauge needle
  - Infants: 1.5-in. 22-gauge needle
- Three-way stopcock (optional: drainage catheter)
- Manometer (optional: extension tube for higher opening pressures)
- Specimen tubes (# may vary, but in general labeled 1–4, important to obtain from 1, 2, 3, 4 owing to cell count obtained from tubes 1 and 3)
- Local anesthetic (lidocaine 1 or 2 %), 5- to 10-mL syringe and needle (25-gauge) for local anesthesia
- Sterile drapes and gauze
- Mask, sterile gown, sterile gloves
- Antiseptic solution for skin preparation (Chloroprep or iodine)

K. Tench, MD
Department of Emergency Medicine, Banner Boswell Medical Center, Sun City, AZ, USA

L.C. Nickels, MD, RDMS (✉) • R.P. Patel, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; rohitpatel@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_41

41.4 Procedure

1. Positioning
   - Determined by practitioner preference or patient capability.
   - Options: lateral recumbent position, upright sitting position (Fig. 41.1).
   - Lateral recumbent position is preferred to obtain accurate opening pressure and to reduce the risk of post-puncture headache.
   - Both positions require the patient to arch the lower back toward the practitioner in order to open up the intervertebral spaces (obtain the "fetal position" or arch "like a cat").
   - Shoulders and hips should remain aligned during process.
2. Landmarks
   - Determined by palpation.
   - Draw a visual line between the superior aspects of the iliac crests that intersects the midline at the L4 interspace. The L3–4 and L4–5 spaces are preferred because these points are below the termination of the spinal cord.
   - Palpate the posterosuperior iliac crests with the midpoint of a visual line that connects the two crests representing the L4 spinous process.
   - Palpate the space between the L3–4 or the L4–5 spinous processes and mark where the needle will be placed.
3. Ultrasound guidance (optional)
   - Helpful in obese patients, patients with previous surgical scarring, or anyone in whom palpation of the spinous processes is not easily done.
   - Sonographic measurement of the dura mater strongly correlates with needle depth needed to obtain CSF.
   - Identify the spinal process in the short and long axis to determine the midline and the interspinous space.
   - Identify the interspinous ligament for estimation of the depth of needle insertion.

     Commonly only the spinous processes are well visualized, and the interspinous ligament, ligamentum flavum, and subarachnoid space are less clearly seen.
   - High-frequency (5–10 MHz) linear probe to best evaluate anatomy.
   - A marking pen can be used to create a cross-hair-type figure (Fig. 41.1).
   - After placing the patient as described, locate the midline at the lumbar spine in transverse and longitudinal orientations.
   - Bright echogenic structures with shadowing posteriorly identify the spinous processes.
   - Transverse probe positioning to identify midline (Fig. 41.2), and then longitudinal probe positioning to identify interspinous space (Fig. 41.3).
4. Sterile preparation
   - After positioning and palpating the appropriate landmarks, the practitioner should then dress in the appropriate protective gear: mask, gown, and sterile gloves.
   - After dressing, the practitioner can then sterilely prepare the patient.
   - Make sure the patient's back is completely exposed.
   - Clean the patient's back with an antiseptic solution (Chloroprep should be scrubbed in an up/down and side/side fashion; iodine in a circular motion starting from the center of the anticipated insertion point).
   - Apply sterile drapes with the puncture site exposed.

     This is an optimal time to make sure all equipment in a standard lumbar puncture tray connects properly and that the stopcock for opening pressure measurement is assembled. Make sure the stopcock is closed away from the patient so that CSF can flow from the patient to the manometer. If the assembly is done, it will decrease the amount of CSF lost after the puncture.
- Local anesthesia
  - 1 % Lidocaine or anesthetic cream topically before preparing skin
  - For injection, form a skin wheal over the insertion site.
  - Inject into the deep tissues below the wheal in all directions while only breaking the skin once.
  - Systemic sedatives and analgesics may also be used.
5. Needle insertion
   - Needle should be inserted in the midline between the L3–4 or the L4–5 spinous process, and the stylet should be firmly in place.
   - Initially parallel to the bed, but once into the subcutaneous tissue, the needle should be angled toward the umbilicus (slightly cephalad, 15°) with the bevel facing upward (Fig. 41.4). This sagittal plane orientation spreads rather than cuts the fibers of the dural sac, which run parallel to the spinal axis.
   - If properly positioned, the needle passes through the skin; subcutaneous tissue; supraspinous ligament; interspinous ligament between the spinous processes; ligamentum flavum; epidural space including the internal vertebral venous plexus, dura, and arachnoid; into the subarachnoid space and between the nerve roots of the cauda equina.
   - In most cases, a "pop" will be felt when the needle penetrates the ligamentum flavum, entering into the subarachnoid space; then intermittent withdrawal should be done in 2-mm intervals to assess for CSF flow.
   - If bone is encountered during insertion, the needle should be withdrawn partially without exiting the skin and readjusted to a different angle more cephalad.

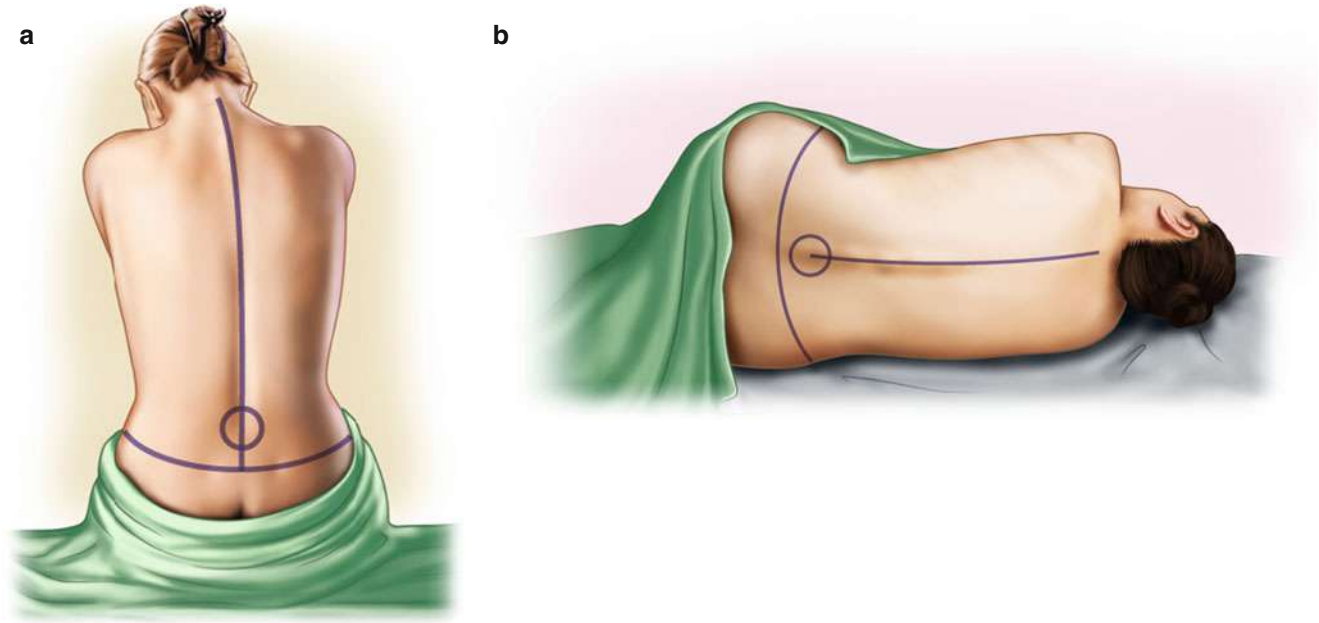

Fig. 41.1 Sitting position (**a**) left lateral decubitus position, (**b**) with general areas of insertion of needle

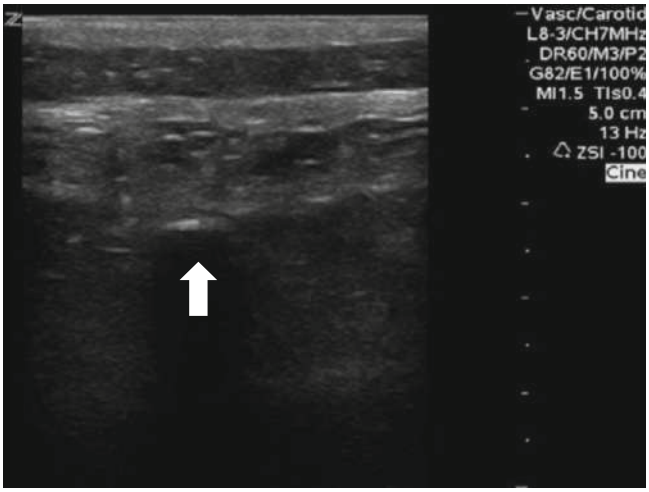

Fig. 41.2 Transverse view, *white arrow* indicates spinous process. Place in middle of ultrasound view to locate appropriate midline access point

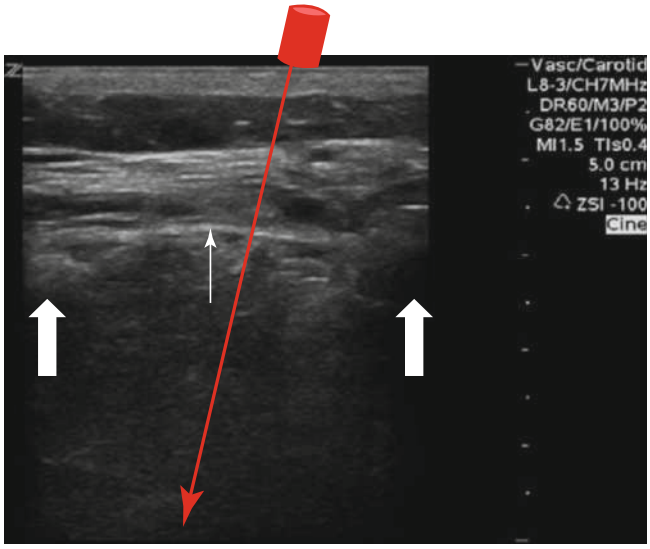

Fig. 41.3 Longitudinal view of lumbar spine. *Thick white arrows* indicate vertebral shadows; *thin white arrow* indicates supraspinous ligament. *Red arrow* indicates trajectory of needle

- If the tap is traumatic, CSF may be blood tinged but should clear as more is collected. If it does not clear, it may indicate intracranial hemorrhage or subarachnoid blood. Also in traumatic patients, clotting will be seen in the tubes; clotting does not occur in SAH owing to defibrinated blood being present in the CSF. Blood-tinged CSF can also be seen in herpes simplex virus (HSV) encephalitis.
- A dry tap is usually due to incorrect positioning and misdirection of needle, often due to a superior direction of the needle with obstruction by the lamina or spinous process of the superior or inferior vertebra. If the needle is too lateral, an inferior or superior articular process may be hit.

  If flow slows down, rotate the needle 90° because a nerve root may be obstructing the opening.

6. Opening pressure measurement
   - Must be performed in the lateral recumbent position. Although there are some conversion formulas from the sitting position, these are not standard of care.

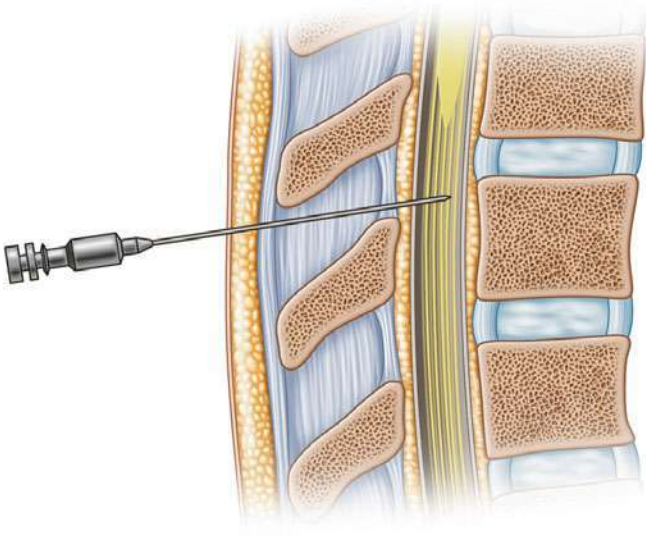

Fig. 41.4 Angle of insertion of needle, cephalad with bevel up

- Once the needle is in the subarachnoid space and CSF is flowing from the needle, the three-way stopcock should be attached to the needle and the manometer should be attached to the stopcock to take a measurement. Use the flexible tube to connect the manometer to the hub of the needle.
- Note the height of fluid in the manometer after it stops rising (normal opening pressure, <20 cm Hg); it may be possible to see pulsations from cardiac or respiratory motion.
  - Elevated CSF pressure is seen with meningeal inflammation, hydrocephalus, pseudotumor cerebri, SAH, and CHF.
  - Decreased CSF pressure is seen in leakage of CSF and severe dehydration.

7. Collecting CSF fluid
   - Collect at least 1–2 mL of CSF fluid in each tube, going from 1 to 4 and never aspirate because this can cause hemorrhage.
   - After collecting the fluid, replace the stylet and remove the needle, clean the skin, and place a bandage over the puncture site.
   - General recommendations
     - Tube 1: glucose, protein, protein electrophoresis
     - Tube 2: Gram stain, bacterial and viral cultures
     - Tube 3: cell count and differential
   - When ruling out SAH, cell count should be performed in tubes 1 and 3 or 1 and 4 to differentiate between SAH and traumatic tap.
     - Tube 4: Any special tests: myelin basic proteins, lactate, pyruvate, and smear on cell concentrates all depend on suspicion.

41.5 Complications

- Implantation of epidermoid tumors: from introducing skin plug into the subarachnoid space and can be avoided by using stylet when advancing.
- Postlumbar puncture headache: most common, occurring in 36.5 % of patients within 48 h
- CSF leak: causes headache when CSF leak through puncture site exceeds rate of production
- Bleeding: most common in patients with bleeding diathesis; may result in spinal cord compression
- Epidural hematoma
- Infection: local cellulitis, abscess (local or epidural), or meningitis
- Herniation syndromes: high risk can be identified by computed tomography but may not completely identify all patients with increased ICP
- Backache: local or referred pain
- Cardiorespiratory compromise

41.6 Pearls and Pitfalls

- Pearls
  - Positioning the patient is key to a successful procedure.
  - In adults the spinal cord may terminate higher than previously thought and it may be okay to go one interspace higher than recommended; but in infants owing to the differential in longitudinal growth of the spinal canal and cord, the spinal cord usually ends in L3. So in children the tap must go L4–5 or L5–S1.
  - Always keep the stylet in place until after the skin barrier is penetrated because this will avoid introduction of epidermoid tissue.
- Pitfalls
  - Postspinal headaches can be avoided with smaller needles and intravenous (IV) fluids.
    - Having the patient lie on the back for 1 h after the procedure has no change in incidence of headache.
    - Treatment consists of IV fluids initially, then caffeine, and, ultimately, if the headache persists, a blood patch.

Selected Reading

Boon JM, Abrahams PH, Meiring JH, Welch T. Lumbar puncture: anatomical review of a clinical skill. Clin Anat. 2004;17:544–53.

Ellenby MS, Tegtmeyer K, Lai S, Braner DA. Videos in clinical medicine. Lumbar puncture. N Engl J Med. 2006;355:e12.

Ferre RM, Sweeney TW, Strout TD. Ultrasound identification of landmarks preceding lumbar puncture: a pilot study. Emerg Med J. 2009;26:276–7.

Peterson MA, Abele J. Bedside ultrasound for difficult lumbar puncture. J Emerg Med. 2005;28:197–200.

42 Reflex Eye Movements (Doll's Eyes and Caloric Testing)

Thomas T. Nguyen, Tina Dulani, and Saadia Akhtar

42.1 Doll's Eyes (Oculocephalic Reflex Testing)

42.1.1 Indications: Doll's Eyes

- To assess brain stem function of a comatose patient
- To assess cerebral function in a comatose patient if brainstem function is intact

42.1.2 Contraindications: Doll's Eyes

- Absolute
  - Occult cervical spine injury; rule out radiographically and clinically
  - Basilar skull fracture
- Relative
  - Rheumatoid arthritis; increased risk of atlantoaxial subluxation resulting in spinal cord compression
  - Osteoporosis; increased risk of cervical spine injury
  - Cervical spine ankylosis; increased risk of cervical spine injury

42.1.3 Procedure: Doll's Eyes (Fig. 42.1)

1. Stand at the head of bed and grasp the patient's head with both hands.
2. Use the thumbs of both of hands to open the patient's eyelids.
3. Rapidly move the patient's head to one side and hold.
4. Simultaneously observe for the presence or absence of horizontal movements.
   (a) Head movement to one side should result in conjugate eye movement to the opposite side and then in spontaneous return of the eyes to the midline (normal test).
   (b) Normal oculocephalic reflex is the observation of conjugate eye movements to the opposite side of head turning. This indicates a functionally intact brainstem in a comatose patient.
   (c) Abnormal oculocephalic reflex: incomplete or absent horizontal eye movements. The eyes remain in the midline. This indicates impairment of the brainstem; caloric testing should be done if not contraindicated.
   (d) A partially abnormal oculocephalic reflex: conjugate eye movement opposite to head turning but does not return to the midline means the brainstem is intact but cerebrum function is not.
5. Repeat by rotating the head to the opposite side.
6. Vertical oculocephalic response can be tested by moving the patient's head up and down. A compensatory vertical eye movement should be observed.
   (a) This test is useful only if the horizontal oculocephalic reflex is negative. An intact vertical oculocephalic reflex with a negative horizontal oculocephalic reflex suggests a pontine lesion.
7. Document the observations.

42.1.4 Pearls and Pitfalls: Doll's Eyes

- Oculocephalic reflex may not be present in the first 10 days of life and is unreliable until 2 years of age.
- Do not attempt the doll's eye maneuver in patients with cervical spinal injuries.

T.T. Nguyen, MD (✉) • S. Akhtar, MD
Department of Emergency Medicine, Mount Sinai Beth Israel, New York, NY, USA
e-mail: tnguyen@chpnet.org

T. Dulani, MD
Department of Emergency Medicine, New York Methodist Hospital, Brooklyn, NY, USA
e-mail: tdulani@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_42

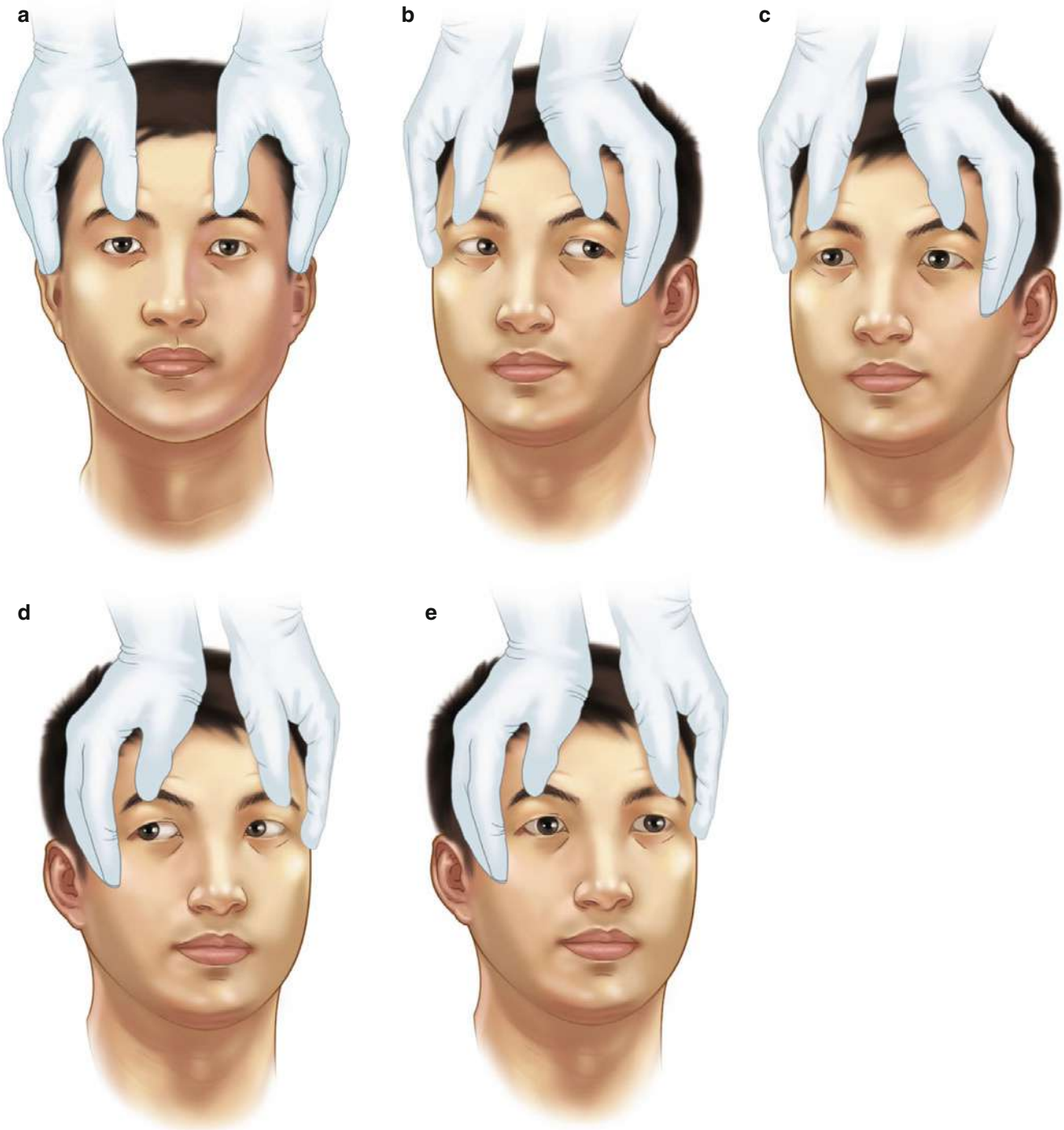

Fig. 42.1 The oculocephalic (doll's eyes) reflex in a patient with an intact brainstem. (**a**) Start with the head facing upright and grasp as depicted using both thumbs to keep the eyes open. (**b**) Rotate the head 90° to the right and the eyes deviate to the left (opposite side). (**c**) The eyes will spontaneously return to the midline. (**d**) Then rotate the head 180° to the left and the eyes should deviate to the right (opposite side). (**e**) The eyes will spontaneously return to the midline

- The doll's eye reflex can be absent or partial in patients with ocular muscle nerve palsy (e.g., cranial nerve [CN] 6).
- Make sure the patient does not have a neuromuscular blockade agent or other toxins present.
- A conscious person can suppress the doll's eye or oculocephalic reflex.

42.2 Caloric Reflex Testing (Vestibuloocular Reflex)

42.2.1 Indications: Caloric Reflex Testing

- In any comatose patient with abnormal doll's eye or if it cannot be performed
- To assess brainstem function of a comatose patient
- To assess asymmetrical function in the peripheral vestibular system

42.2.2 Contraindications: Caloric Reflex Testing

- Absolute
  - Perforated tympanic membrane
  - Presence of tympanostomy tubes
  - Basilar skull fracture, petrous bone fracture
  - Suspicion of cerebrospinal fluid (CSF) otorrhea

42.2.3 Materials and Medications: Caloric Reflex Testing

- Otoscope
- 60-mL syringe
- 16- to 18-gauge angiocatheter
- Thermometer
- Ice-cold water (30–33 °C)
- Warm water (44 °C)
- Emesis basin (to collect water)
- Towels or Chux

42.2.4 Procedure: Caloric Reflex Testing (Fig. 42.2)

1. Place the patient supine and elevate the head to 30° to bring the lateral semicircular canal into the upright position.
   (a) Careful otoscopic examination should be done before caloric testing to make sure there are no contraindications.
   (b) Remove any cerumen from the external auditory canal (EAC); the irrigation fluid should be able to reach the tympanic membrane.
2. Get a 60-mL syringe with an 18-gauge angiocatheter. Remove the plastic angiocatheter from the needle and discard the needle.
3. Place the angiocatheter on the 60-mL syringe and fill it with ice water (30 °C).
4. Place the angiocatheter in the EAC and squirt the water in over a period of 30–40 s. The water should freely enter and exit the auditory canal.
   (a) The stimulus depends upon the temperature of the water and not on water pressure.
   (b) Reflex horizontal movements may be delayed for up to 1 min after irrigation of the EAC.
5. Observe: Have an assistant hold the eyelids open with the head still and facing forward.
   (a) Normal test: Cold water induces slight deviation of the eyes to the side being irrigated followed by a latent period of approximately 20 s and then nystagmus to the opposite side (direction of the fast phase).
   (b) Abnormal test: Eyes do not deviate; this implies brainstem problems.
   (c) The quick phase may return in patients in a persistent vegetative state. Search for the slow, full eye deviation in response to caloric stimulation and not nystagmus when assessing a comatose patient.
6. Pause at least 5 min so the auditory canal can warm up.
   (a) The same EAC can be irrigated with warm water if the contralateral side cannot be tested for any reason.
7. Repeat steps 3–5 in the opposite ear.
   (a) Warm water testing can be used if there is no response to cold water testing. Use warm water (44 °C) and repeat testing, starting with one ear and then, after 5 min, with the other ear.
   (b) Warm water induces nystagmus to the same side (direction of the fast phase).
   (c) COWS: Cold opposite, warm same (referring to the compensatory fast phase of eye movement, the nystagmus).
8. Dry the patient off and reexamine the tympanic membranes to assess for testing-related injury.

42.2.5 Complications: Caloric Reflex Testing

- If contraindications are excluded, no significant complications are expected.
- Tympanic membrane injury or EAC injury can occur from the angiocatheter, forceful irrigation, or injury during cerumen removal.
- Potential complications from caloric testing include meningitis, otitis media, and vomiting.

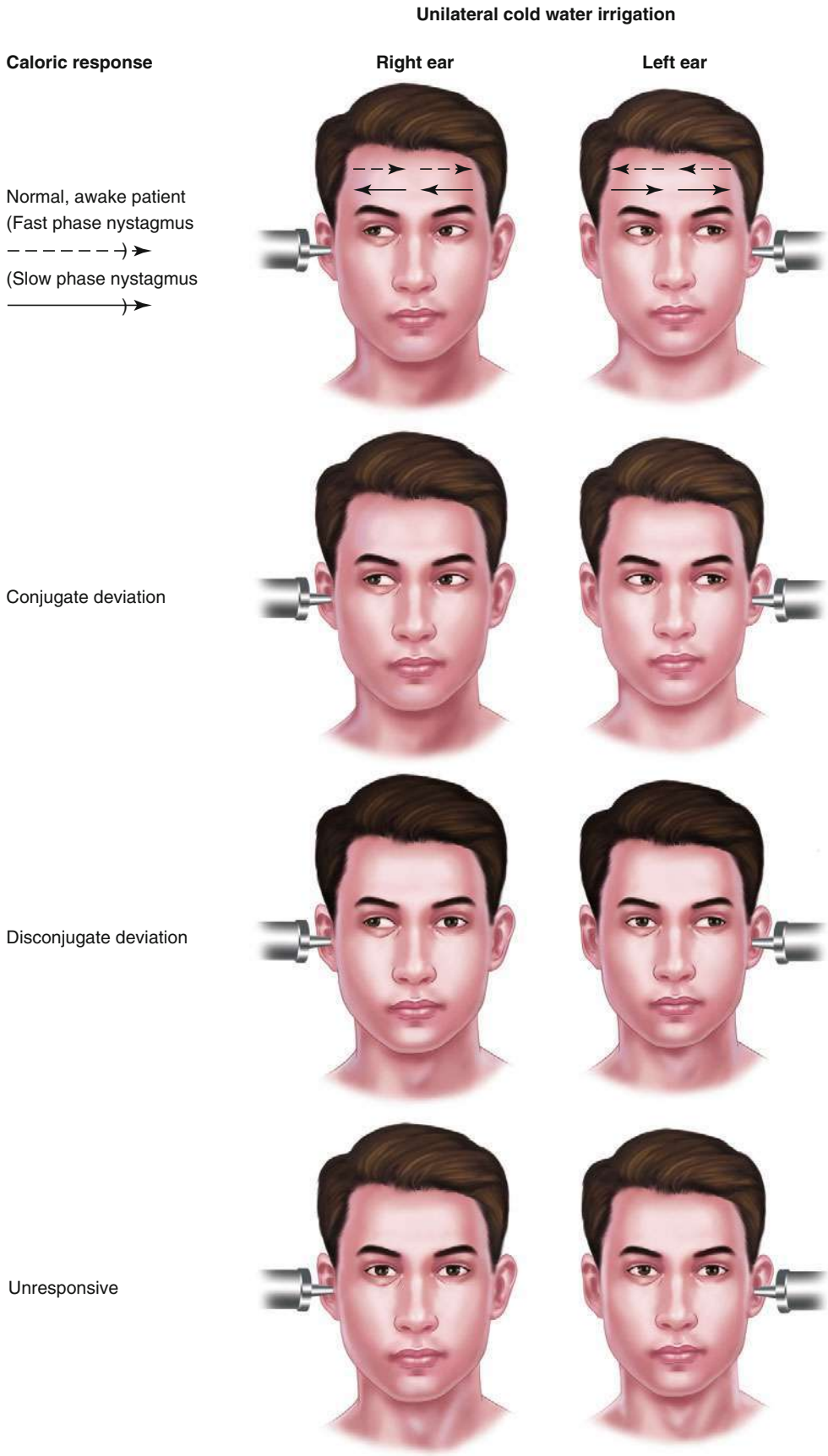

Fig. 42.2 Depiction of the vestibuloocular reflexes with unilateral cold-water irrigation

42.2.6 Pearls and Pitfalls: Caloric Reflex Testing

- A positive response to caloric testing indicates intact brainstem function.
- In an awake patient, excessive reflex vagal activity may occur (i.e., nausea, vomiting, dizziness).
- Absence of horizontal eye movement means comatose brainstem injury.
- Disconjugate or impaired reflex horizontal eye movements indicate impaired brainstem function at or below the level of the oculomotor nucleus.
- A comatose patient with intact full-reflex horizontal eye movements indicates that the lesion causing coma is in the cerebral hemisphere.

Selected Reading

Gomella LG. Central nervous system. In: Clinician's pocket reference. 11th ed. New York: McGraw-Hill; 2007.

McCann J, et al. Rapid assessment: a flowchart guide to signs and symptoms. Philadelphia: Lippincott William & Wilkins; 2004.

Reeve A, Swensen R. Auditory and vestibular function. In: Disorders of the nervous system: a primer. Hanover, NH: Dartmouth Medical School; On Line; 2009.

Reichman EF, Simon RR. Chapter 103. Reflex eye movements (caloric testing and doll's eyes). In: Emergency medicine procedures. New York: McGraw Hill Education; 2004.

Ropper AH, Samuels MA. Chapter 15. Deafness, dizziness, and disorders of equilibrium. Chapter 17. Coma and related disorders of consciousness. In: Adams and Victor's principles of neurology. 9th ed. New York: McGraw Hill; 2009.

Simon RP, Greenberg DA, Aminoff MJ. Chapter 3. Disorders of equilibrium. In: Clinical neurology. 7th ed. New York: McGraw Hill; 2009. http://www.accessmedicine.com.elibrary.einstein.yu.edu/content.aspx?aID=5146162.

43 Dix-Hallpike Maneuver

Rui Domingues and Muhammad Waseem

The Dix-Hallpike maneuver, also termed the "head-hanging positioning maneuver," is helpful in confirming the clinical suspicion of benign paroxysmal positional vertigo (BPPV). This maneuver provokes abnormal nystagmus, which is a characteristic feature of BPPV.

43.1 Indications

- BPPV is one of the most common types of vertigo.
  - The pathophysiology of BPPV, in brief, is believed to be due to free-moving densities (canaliths/otoliths) in the posterior semicircular canal; with head movement, the particles would alter the flow of the endolymph and cause the stimulation of the ampulla. The particles in the canal cause slow or even reversal of the movement of the cupula and create signals that do not correlate with the actual head movements, therefore causing the sensation of nystagmus.
  - This maneuver locates the cause of vertigo as either the inner ear or the brain; if the problem is in the ear, this maneuver helps localize which ear is affected.
- This maneuver is indicated for patients presenting with vertigo, which is evoked by a change in position and has no symptoms at rest.
- This maneuver is inexpensive, easily done, and part of the physical examination when a patient presents with the complaint of dizziness or vertigo.

R. Domingues, MD • M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: rbdomingues@hotmail.com; waseemm2001@hotmail.com

43.2 Contraindications

- Severe cervical spine disease.
- Unstable spinal injury.
- High-grade carotid stenosis.
- Unstable heart disease.
- Elderly patients may not tolerate this maneuver.
- There is no need to perform this test in the presence of nystagmus at rest.

43.3 Materials

- Examination table
- Flat cushion
- Frenzel goggles: These high-powered (+20 diopters) magnifying glasses can be placed on a patient during the performance of the maneuver and have shown to increase the sensitivity of the Dix-Hallpike maneuver by preventing the patient from visually fixating on an object, thereby preventing suppression of nystagmus. They are not required to perform the maneuver; they usually are used by specialists.

43.4 Procedure (Fig. 43.1)

1. Have the patient sit at the edge of a bed. The patient is instructed to maintain eye contact with the physician throughout the maneuver.
2. With the patient seated, the examiner will extend the neck, approximately 20°, and turn the head to one side, approximately 30–45°.
3. The examiner then assists the patient by lowering the patient quickly into a supine position, so that the head hangs over the edge of the bed or table, with the neck in a hyperextended position. A flat cushion can be placed

© Springer Science+Business Media New York 2016

L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_43

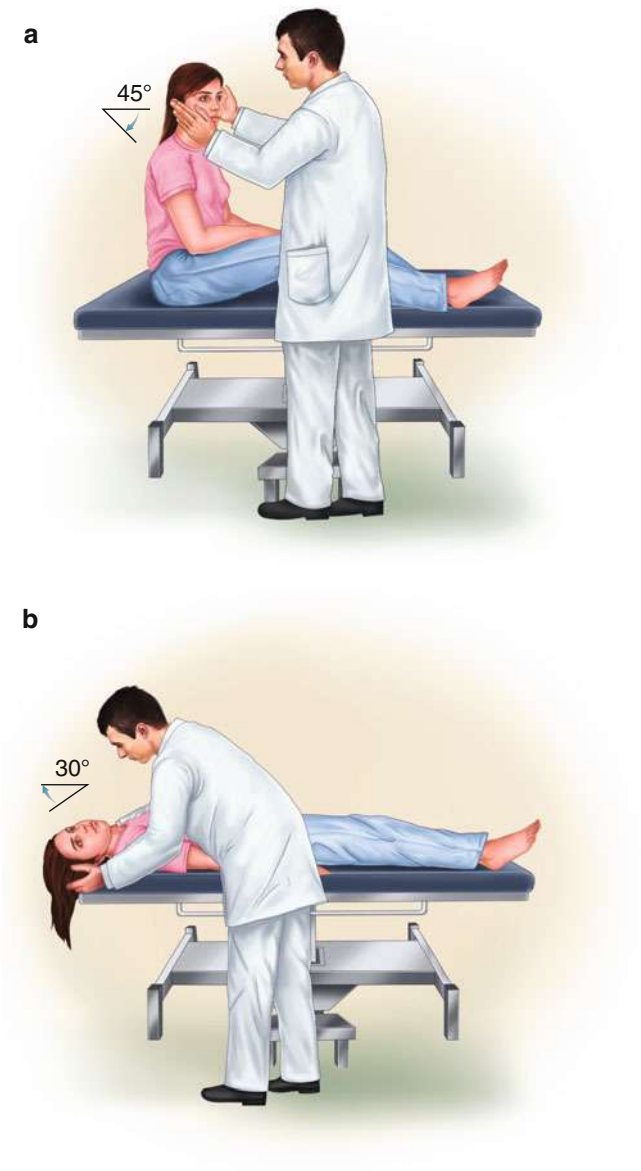

Fig. 43.1 (**a**) With the patient seated, the doctor will extend the neck, approximately 20°, and turn the head to one side, approximately 30–45°. (**b**) The physician then assists the patient by lowering the patient quickly into a supine position, so that the head hangs over the edge of the bed or table, with the neck in a hyperextended position

beneath the person's back in the shoulder blade area to assist with obtaining head extension.

4. This position is held, and the examiner observes for nystagmus for up to 60 s.
   (a) Nystagmus is a rapid, rhythmic movement of the eyes and usually appears after a brief period and lasts less than 30 s.
   (b) The direction of the nystagmus is usually up and twisted; therefore, the eyes will beat toward the ground.
   (c) If the patient becomes dizzy or the doctor sees nystagmus, the test is positive for the ear that is pointed to the floor.
5. The patient is then returned to the upright position and is again observed for nystagmus for 30 s.
6. If nystagmus is not provoked, the maneuver is repeated with the head turned to the opposite direction.
7. If nystagmus is provoked, the patient should have the maneuver repeated to the same side.
   (a) With each repetition, the intensity and duration of nystagmus should decrease.

43.5 Complications

- Vertigo
- Nausea

43.6 Pearls and Pitfalls

- The maneuver can be uncomfortable to the patient because it can cause vertigo and nausea.

Selected Reading

Brandt T, Steddin S, Daroff RB. Therapy for benign paroxysmal positioning vertigo, revisited. Neurology. 1994;44:796–800.

Buckingham RA. Anatomical and theoretical observations on otolith repositioning for benign paroxysmal positional vertigo. Laryngoscope. 1999;109:717–22.

Fife TD, Iverson DJ, Lempert T, et al. Practice parameter: therapies for benign paroxysmal positional vertigo (evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2008;70:69–74.

Halker RB, Barrs DM, Wellik KE, Wingerchuck DM, Demaerschalk BM. Establishing a diagnosis of benign paroxysmal positional vertigo through the Dix-Hallpike and side-lying maneuvers: a critically appraised topic. Neurologist. 2008;14:201–4.

Herdman SJ. Treatment of benign paroxysmal vertigo. Phys Ther. 1990;70:381–8.

Lanska DJ, Remler B. Benign paroxysmal positioning vertigo: classic descriptions, origins of the provocative positioning technique and conceptual developments. Neurology. 1997;48:1167–77.

Li JC, Meyers AD. Benign paroxysmal positional vertigo. http://emedicine.medscape.com/article/884261-overview. Accessed December 20, 2015.

Von Brevern M, Radtke A, Lezius F, et al. Epidemiology of benign paroxysmal positional vertigo: a population based study. J Neurol Neurosurg Psychiatry. 2007;78:710–5.

44 Epley Maneuver for Vertigo (Particle Repositioning or Canalith Repositioning Procedure)

Rui Domingues and Muhammad Waseem

During the Dix-Hallpike test (see Chap. 43), the particles move in the canal and trigger a burst of upbeat-torsional nystagmus. The Epley maneuver causes resolution of positional nystagmus. This maneuver is effective in about 80 % of patients with benign paroxysmal positional vertigo (BPPV).

44.1 Indications

- Performed to alleviate the symptoms of posterior canal BPPV

44.2 Contraindications

- Back or spine injuries or other problems
- Presence of detached retina

44.3 Materials

- Pillow or pad
- Padded table or bed

44.4 Procedure (Fig. 44.1)

1. The patient is positioned on the bed with the head slightly extended and turned 45° to the affected side; hold this position for 60 s.
   - The patient may hold the physician's arm for support.
2. The patient is then assisted by the physician to a supine position, with the head in extension, with a pillow or pad placed at the shoulder level allowing for extension; this position is held for a 60 s. This position will likely provoke transient dizziness and vertigo.
3. The head is then turned 90° to the opposite side; this position is held for 60 s. This position will likely provoke transient dizziness and vertigo.
4. The head is then turned through a further 90° while the patient rolls onto the unaffected side; this position is held for 60 s.
5. The patient then sits up slowly, by rolling up from lying on the unaffected side; this position is held for 60 s.
6. The patient should wait for 10–15 min before discharge.
7. Discharge instructions are then provided. Home instructions:
   - Instruct the patient to sleep in a semi-recumbent position for the next two nights; the patient should sleep at a 45° angle, which is most easily done by sleeping in a recliner or with several pillows arranged on a couch.
   - Sleep on the nonaffected side.
   - During the day, try to keep the head in a vertical position; no sudden head movements to the right, left, up, or down.
   - When men shave under their chins, they should bend their bodies forward in order to keep their heads vertical.
   - Do not go to the hairdresser or dentist because these require head movements.
   - Care should be taken when putting in eye drops because it requires head extension.
   - No vacuuming or mopping the floor.
   - Try not to wear clothing that needs to be pulled over the head. Try to avoid bending down to tie shoes.

R. Domingues, MD • M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: rbdomingues@hotmail.com; waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016

L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_44

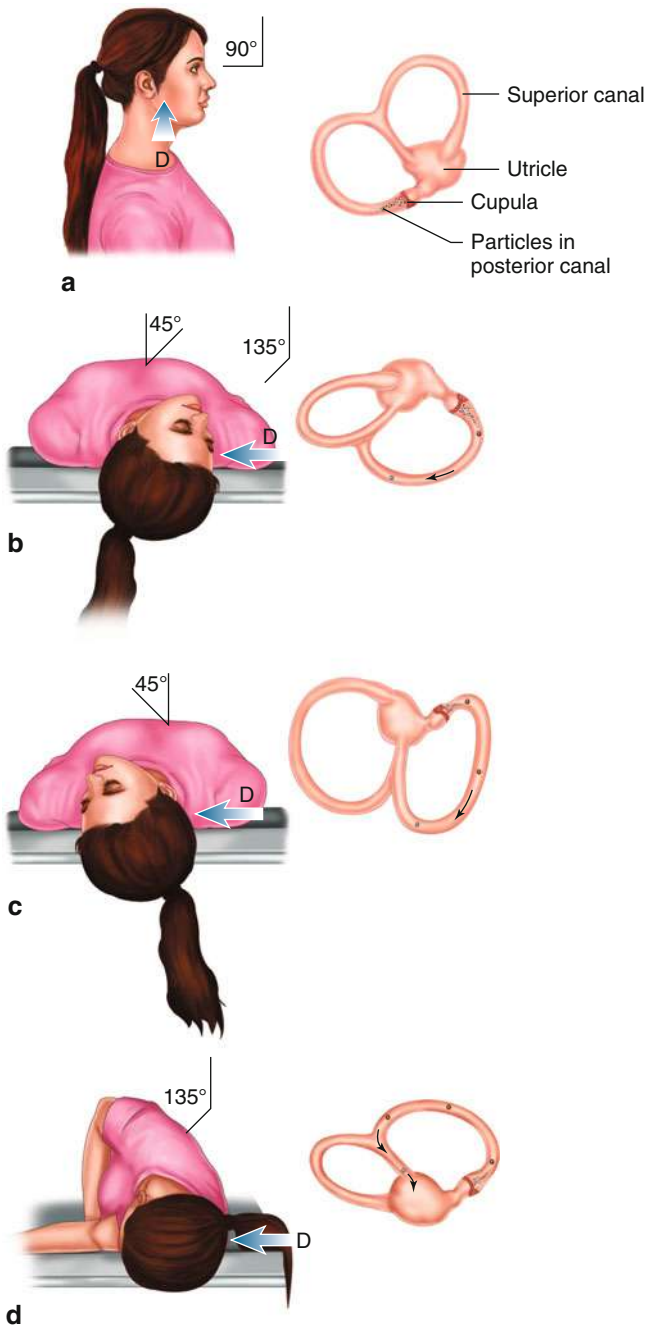

Fig. 44.1 (**a**) Patient seated on the table. (**b**) Patient in supine position, with the head turned 45° to the affected side. (**c**) Head then turned to the unaffected side (45°). (**d**) Head turned additional 90° on the unaffected side. *D* direction of view of labyrinth

44.5 Complications

When performing the previous maneuver, caution is advised should neurological symptoms occur. Occasionally such symptoms are caused by compression of the vertebral arteries; if it persists for a long period of time, a stroke can occur.

44.6 Pearls and Pitfalls

- Pearls
  - This maneuver should initially be performed by a trained therapist or medical physician.
  - It is best to perform the Epley maneuver before going to bed at night.
- Pitfalls
  - The patient should not drive herself or himself home after the procedure.
  - Avoid rapid changes in head position that might provoke BPPV.

Selected Reading

Harvey SA, Hain TC, Adamiec LC. Modified liberator maneuver: effective treatment for benign paroxysmal positional vertigo. Laryngoscope. 1994;104:1206–12.

Herdman SJ. Treatment of benign paroxysmal vertigo. Phys Ther. 1990;70:381–8.

Massoud EA, Ireland DJ. Post-treatment instructions in the nonsurgical management of benign paroxysmal positional vertigo. J Otolaryngol. 1996;25:121–5.

Parnes LS, Price-Jones RG. Particle repositioning maneuver for benign paroxysmal positional vertigo. Ann Otol Rhinol Laryngol. 1993;102:325–31.

Radtke A, Neuhausen H, von Brevern M, Lempert T. A modified Epley's procedure for self-treatment of benign paroxysmal positional vertigo. Neurology. 1999;53:1358–60.

Smouha EE. Time course of recovery after Epley maneuvers for benign paroxysmal positional vertigo. Laryngoscope. 1997; 107:187–91.

Welling DB, Barnes DE. Particle repositioning maneuver for benign paroxysmal positional vertigo. Laryngoscope. 1994;104: 946–9.

Clinical Brain Death Examination in Adults

45

Eric S. Papierniak, Hassan M. Alnuaimat, Tracy Timmons, and Deborah M. Stein

45.1 Indications

- Deep coma without any identifiable reversible causes
- Suspected brain death in a mechanically ventilated patient

45.2 Contraindications

- Metabolic derangements
- Acid-base disorders
- Electrolyte abnormalities sufficient to cause the coma
- Profound hypotension
- Hypothermia (core body temperature <36°C)
- Drug overdose or toxic exposure
- Locked-in syndrome

E.S. Papierniak (✉)
Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, University of Florida Health, Gainesville, FL, USA

Malcom Randall VA Medical Center, Gainesville, FL, USA
e-mail: eric.papierniak@medicine.ufl.edu

H.M. Alnuaimat, MD
Department of Medicine, University of Florida, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: Hassan.Alnuaimat@medicine.ufl.edu

T. Timmons, MD
Department of Trauma/Surgical Critical Care, R Adams Cowley Shock Trauma Center, University of Maryland Medical Center, Baltimore, MD, USA
e-mail: ttimmons1@umm.edu

D.M. Stein, MD, MPH
Department of Surgery, University of Maryland School of Medicine, R Adams Cowley Shock Trauma Center, University of Maryland Medical Center, Baltimore, MD, USA
e-mail: dstein@umm.edu

45.3 Materials and Medications

- A bright light
- Intravenous (IV) catheter
- 30-mL syringe
- 50-mL of ice water
- Endotracheal suction catheter
- Long swab or tongue depressor
- Gauze 4×4 or saline eye drops

45.4 Procedure

1. Evaluate the motor response to pain.
   - Apply pressure to the supraorbital nerve and nail beds. No motor response is consistent with brain death (Figs. 45.1 and 45.2).
2. Test for the absence of brainstem function.
   - Check pupillary response to bright light.
   - Check the oculocephalic reflex. With the eyelids held open, quickly turn the head to the side. The normal response is for the eyes to move in the opposite direction, maintaining the focus forward. Failure of the eyes to deviate during head rotation is consistent with brain death (Fig. 45.3).
   - Check the oculovestibular reflex with cold caloric testing. Elevate the head of the bed to 30°. Attach an IV catheter to the tip of a 30-mL syringe. Inject 50 mL of cold water into the external ear canal. Observe the pupils for 1 min for deviation toward the ear with the cold stimulus. Wait 5 min before testing the other side. No deviation of the pupils is consistent with brain death (Figs. 45.4 and 45.5).
   - Check for bulbar paralysis.
     - Check for a cough response to bronchial suctioning.

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_45

- Stimulate the posterior pharynx with a long swab or tongue depressor and observe for a gag.
- Check the corneal reflex by lightly touching the cornea with a cotton swab, corner of gauze, or drops of sterile saline and observe for blinking of the eyelids (Fig. 45.6).

3. Check for respiratory effort by performing an apnea test. Before starting the apnea test, the following conditions must be met:
   (a) The patient must not be hypothermic. The core temperature must be greater than 36°C.
   (b) The patient must be hemodynamically stable with a systolic blood pressure greater than 90 mmHg.
   (c) The arterial blood gas (ABG) must demonstrate a normal arterial partial pressure of carbon dioxide ($PaCO_2$). The arterial partial pressure of oxygen (PaO_2) may be elevated with preoxygenation to minimize the risk of hypoxemia prematurely ending the examination.
      - Monitor the patient with a pulse oximeter.
      - Disconnect the ventilator and deliver oxygen at 6 L/min by cannula into the endotracheal tube.
      - Alternately the patient can remain connected to the ventilator with no applied support, which allows for an in-line negative pressure monitor to be attached to the circuit. Note: Most ventilators are too sensitive to be used as the sole method of sensing respiratory effort as they can produce false-negative results.
      - Watch the chest and abdomen for respiratory motion, or monitor for negative inspiratory force on the pressure gauge. Check an ABG every 10 min and when the test ends.
      - If the $PaCO_2$ increases by 20 mmHg or the $PaCO_2$ is greater than 60 mmHg and there was no respiratory effort, the test is consistent with brain death.
      - The test must be aborted and the patient reconnected to mechanical ventilation if the patient becomes hypotensive, desaturates, or develops cardiac arrhythmias.
      - If there is respiratory effort, the patient must be reconnected to mechanical ventilation (Figs. 45.7 and 45.8).

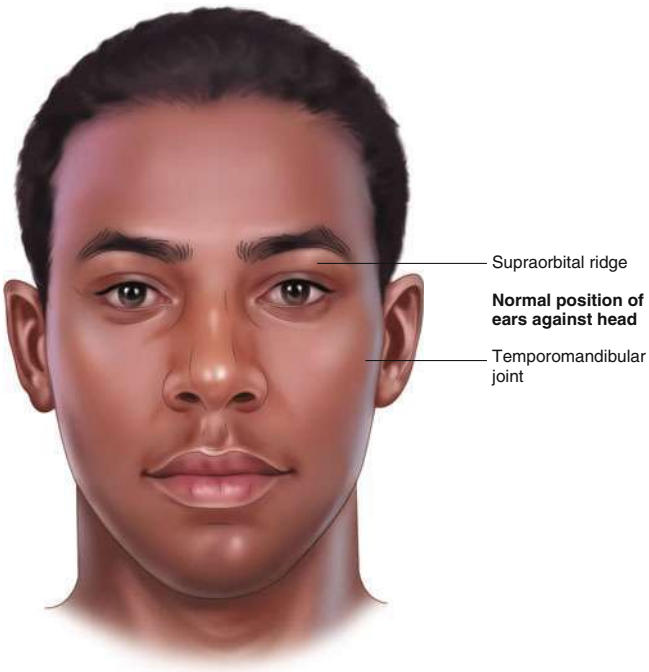

Fig. 45.1 Motor response to pain 1. Temporomandibular joint and supraorbital ridge test supraorbital nerve function

Fig. 45.2 Motor response to pain 2: Firm pressure can be applied to the nail beds with a penlight or similar instrument

a Normal reaction: Eyes move side to side when head is turned

b Abnormal reaction: Eyes remain in fixed position in skull when head is turned

Fig. 45.3 (**a**, **b**) Oculocephalic reflex or "Doll's eyes." Movement of the eyes away from the direction of head turning (i.e., to keep the gaze forward) indicates intact functioning

Fig. 45.4 Preparation of cold saline for testing of the oculovestibular reflex or "cold calorics"

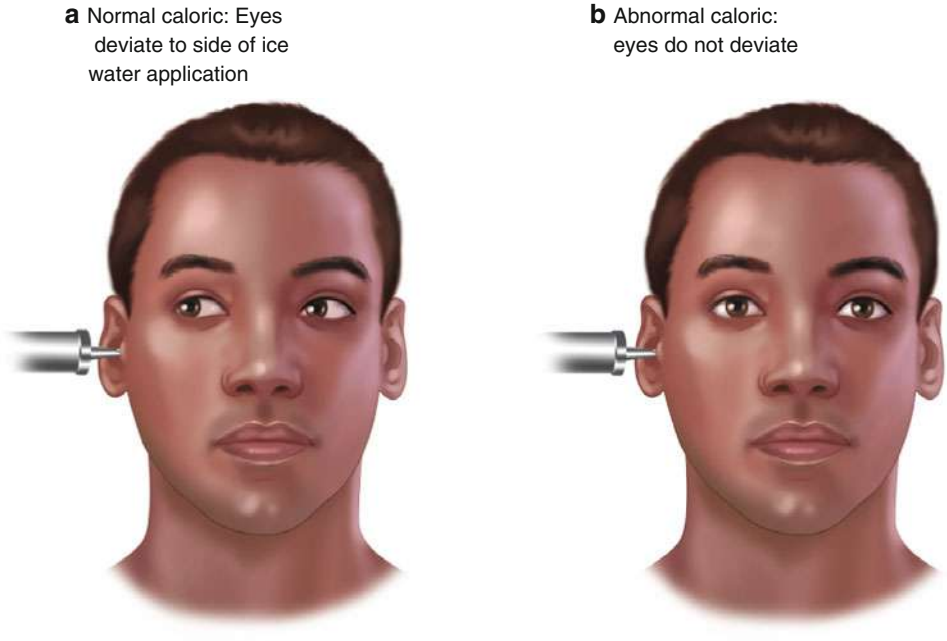

Fig. 45.5 (**a**, **b**) Oculovestibular reflex. Intact functioning is demonstrated by deviation of the gaze toward the side being tested

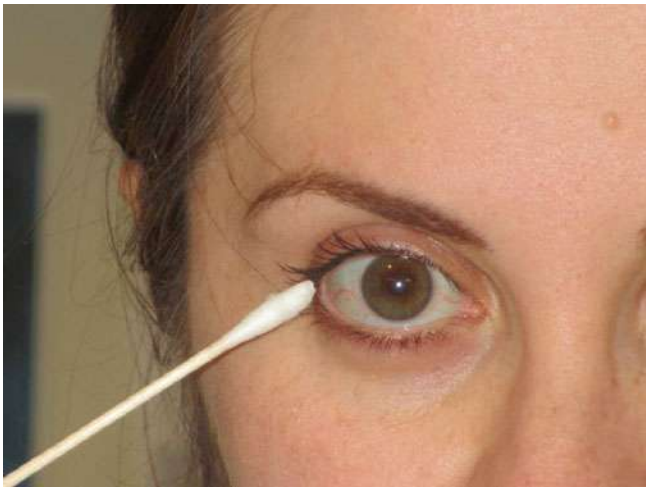

Fig. 45.6 Use of a cotton ball or unwound tip of a cotton swab for testing of the corneal reflex. A piece of sterile gauze ("4×4") is also commonly used

Fig. 45.8 Ventilator waveform in a patient without spontaneous respirations during the apnea test

Fig. 45.7 Ventilator waveform before apnea test

45.5 Pearls and Pitfalls

- Pearls
  - The clinical examination should be performed by someone with experience or training in brain death examinations. Local regulations as well as hospital policies should be reviewed in order to determine the requirements for brain death. Some locales require a repeat examination by a different practitioner.
  - A patient with a cervical spine injury causing quadriplegia may not feel nail bed pressure and may be unable to respond to pain with more than facial movements.
- Pitfalls
  - Pupillary constriction less than 3 mm is not consistent with brain death.
  - Do not test oculocephalic reflex (doll's eyes) on patients who may have a cervical spinal cord injury from trauma.
  - Do not test the oculovestibular reflex with cold caloric testing in patients with a ruptured tympanic membrane.
  - A large number of patients may demonstrate spinal reflex movements during brain death. These reflexes may be triggered by touch, noxious stimuli, or removal of the ventilator. These movements can include plantar flexion, upper extremity posturing, eye opening, and the "Lazarus sign" (the arms raise off the bed and cross). Although these movements can be disconcerting to the health care team and family, they do not preclude a diagnosis of brain death.

45.6 Controversies

- Repeat examinations
  - The American Academy of Neurology guidelines recommend repeating the brain death examination at 6 h. The guidelines acknowledge that this is an arbitrary interval. Recent evidence suggests that the second examination may be unnecessary.
- Confirmatory Testing
  - Confirmatory testing is an option if specific elements of the clinical examination cannot be performed. For example, a trauma patient with suspected cervical spinal cord injuries precluding the oculocephalic reflex test would be a candidate for confirmatory testing.
  - Conventional cerebral angiography will demonstrate no filling beyond the carotid bifurcation or within the circle of Willis. Electroencephalography demonstrates no electrical activity. Transcranial Doppler ultrasonography will demonstrate vascular resistance associated with elevated intracranial pressure. Technetium-99m hexamethylpropylene-amine oxime brain scans demonstrate the "hollow skull" sign or no uptake of isotope in the brain (Fig. 45.9).
  - Newer modalities of confirmatory testing include computed tomography (CT) angiography and magnetic resonance imaging/magnetic resonance angiography (MRI/MRA). These tests are being used in some hospitals; however, the recent review by the American Academy of Neurology finds evidence insufficient to recommend using these newer modalities to confirm brain death.

Fig. 45.9 "Empty skull sign" on nuclear medicine flow study confirms brain death

Selected Reading

A definition of irreversible coma: report of the Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death. JAMA. 1968;205:337–40.

Lustbader D, O'Hara D, Wijdicks EF, et al. Second brain death examination may negatively affect organ donation. Neurology. 2011;76:119–24.

Practice parameters for determining brain death in adults (summary statement). The quality standards subcommittee of the American Academy of Neurology. Neurology. 1995;45:1012–4.

Saposnik G, Maurino J, Saizar R, Bueri JA. Spontaneous and reflex movements in 107 patients with brain death. Am J Med. 2005;118:311–4.

Wijdicks EF. Current concepts: the diagnosis of brain death. N Engl J Med. 2001;344:1215–21.

Wijdicks EF, Varelas PN, Gronseth GS, Greer DM, American Academy of Neurology. Evidence-based guideline update: determining brain death in adults: report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2010;74:1911–8.

Part VII

Ophthalmic Procedures

46 Slit Lamp Examination

Bobby K. Desai

46.1 Indications

- A slit lamp magnifies structures of the eye (Fig. 46.1).
- Gives the operator a three-dimensional view of the area visualized.
- Used to delineate abnormalities that cannot be visualized by other means.
- Helpful in foreign body removal.

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_46

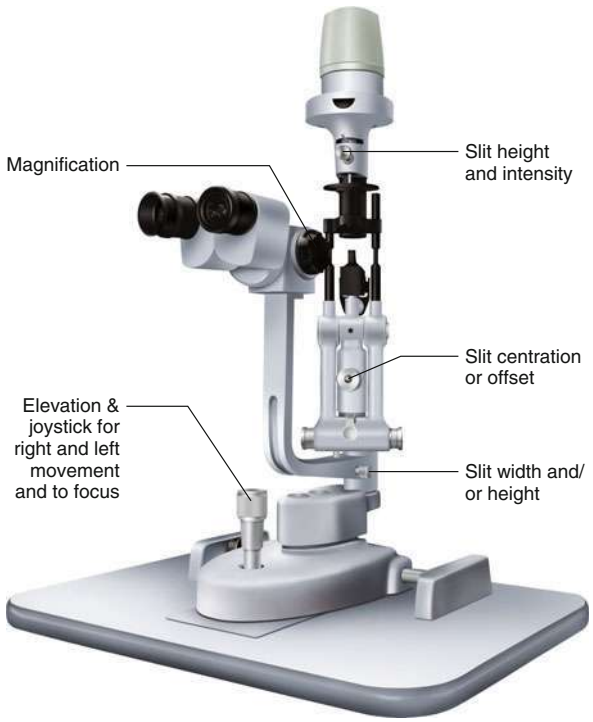

Fig. 46.1 A basic slit lamp

46.2 Contraindications

- Not to be used in patients who cannot tolerate an upright posture (e.g., trauma patients).
  - In these cases if a slit lamp examination is required, a portable slit lamp may be beneficial (Fig. 46.2).

Fig. 46.2 A portable slit lamp

46.3 Materials

- The slit lamp is an eye-specific binocular stereoscope specifically designed to examine the eye and its structures and gives the practitioner a three-dimensional assessment of the eye.
- The operator has full movement of the microscope in all directions and the apparatus can be locked in place if required.
- Its light source can be manipulated to change the characteristics of the beam from its intensity to the angle at which it projects (Fig. 46.3).
  - A low-power setting is typically used for normal examination.
  - A higher-intensity beam is used when evaluating the anterior chamber with a narrow slit beam.
- It has colored filters typically built within the slit lamp.
  - Cobalt blue: Used with fluorescein dye to evaluate for corneal abrasions (Fig. 46.4) and avulsions; the dye will collect where the corneal epithelium is absent. It results in a yellow glow or hue visible through the microscope.
  - Green filter: Used to increase contrast of blood vessels. They appear black and the filter is useful for the assessment of hemorrhage.
- The operator is able to adjust the magnification of the microscopic typically through dial controls.
  - Low magnifications are most helpful for general examination.
  - Higher magnifications are used for examination of a particular area in fine detail.

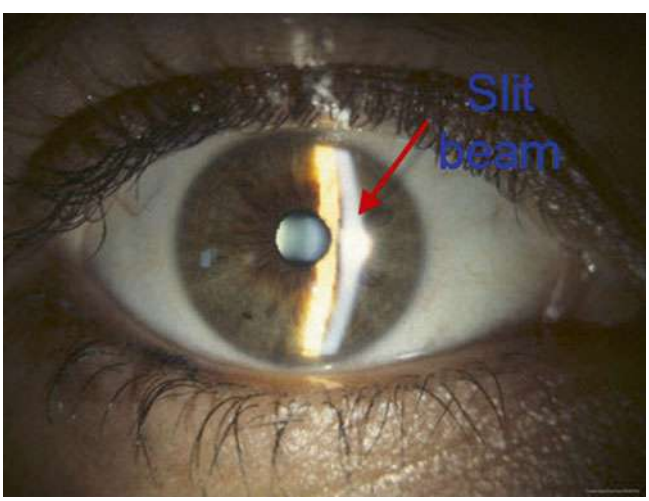

Fig. 46.3 The slit lamp beam and the reflection of the beam

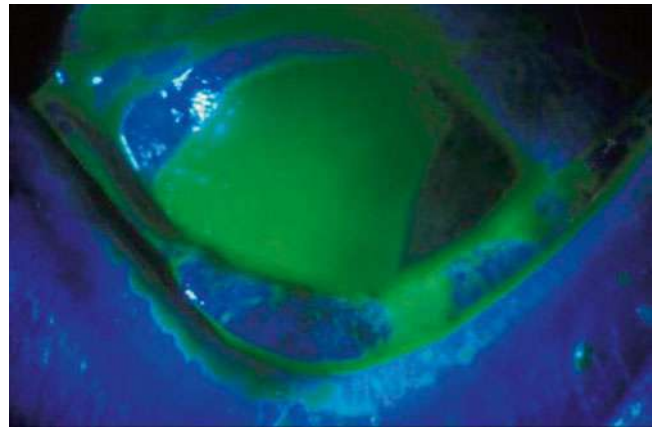

Fig. 46.4 Example of a corneal abrasion seen with a cobalt blue filter (With kind permission from Springer Science + Business Media: Das S, Chohan A, Snibson GR, Taylor HR. Capsicum spray injury of the eye. *International Ophthalmology*. 2005;26(4–5):171–3)

46.4 Procedure

1. As part of a complete eye examination, informed consent is not generally required.
2. An explanation of the procedure and reassurance are helpful.
3. Lock the slit lamp before positioning the patient because unintended movement may inadvertently damage the equipment or cause injury to the operator or patient.
4. The apparatus has a chin and a head rest. Ask the patient to place the chin and forehead in the respective areas (Fig. 46.5).
5. Adjust the height of the apparatus and the patient's chair for optimum comfort (Figs. 46.6 and 46.7). It may be necessary for children to stand during the examination.
6. The patient may place his or her hands on the sides of the table that the lamp is mounted on; this ensures stability of both the lamp and the patient.
7. Adjust the eyepieces of the microscope to fit the operator.
8. Turn on the slit lamp at its lowest setting to avoid a sudden power surge that could potentially damage the bulb.
9. Move the stage forward and narrow the beam and angle it at 45° to the patient. Aim the beam laterally so as not to cause the patient discomfort.
10. Focus the beam by manipulating the joystick to move the apparatus forward and backward so that the beam is clearly visible and its lines are sharp.
11. For each area of the eye to be examined, inspect the area thoroughly using the joystick to slowly manipulate the slit lamp across the eye in all directions, using the height adjustment of the joystick to slightly raise and lower the slit lamp as needed.
    - The operator may find that the slit lamp may move too freely; in that case she or he may find slightly tightening the locking nut of the C-arm may provide better and more precise control.

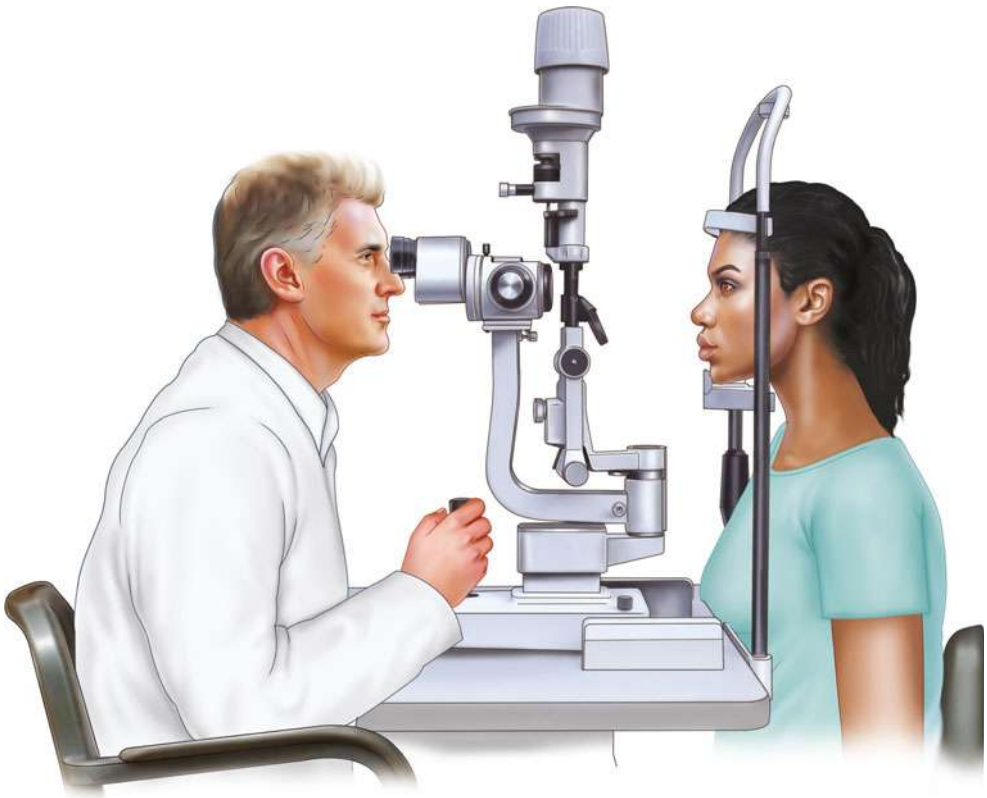

Fig. 46.5 Appropriate positioning for a slit lamp examination

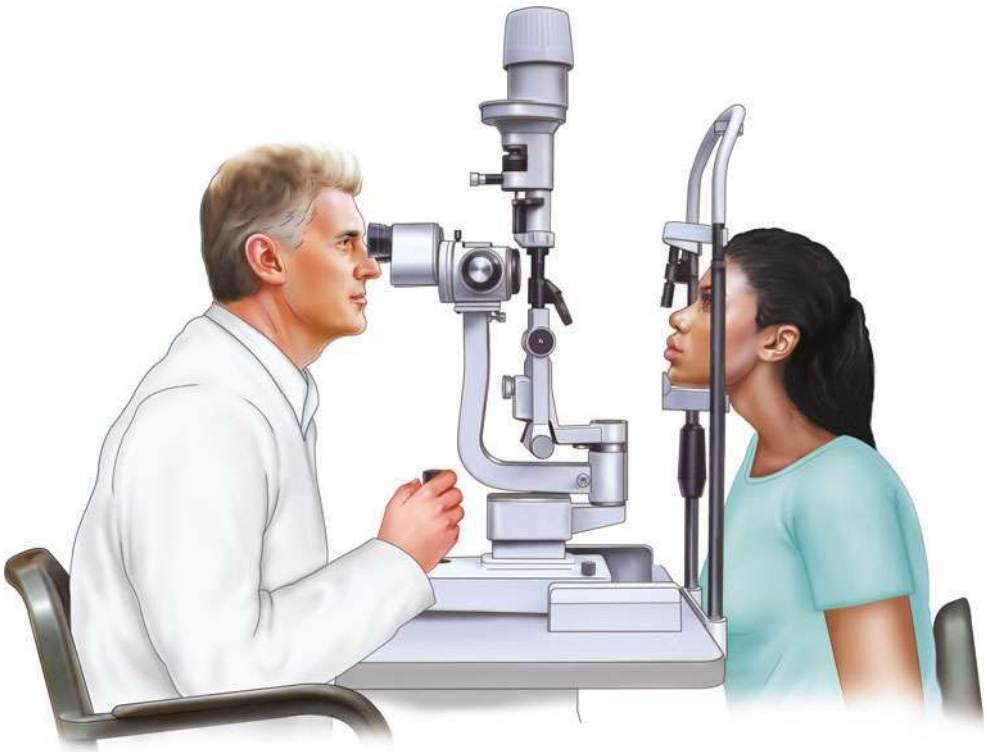

Fig. 46.6 The slit lamp is too high

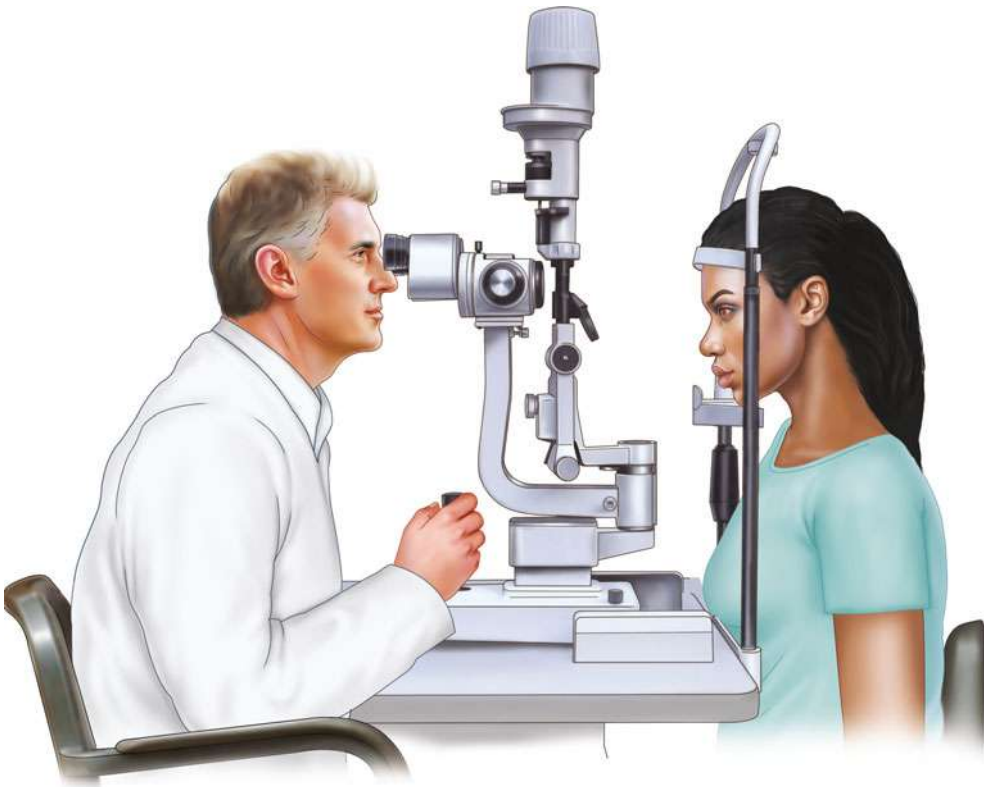

Fig. 46.7 The slit lamp is too low

46.5 Complications

There are no complications following routine slit lamp examination.

46.6 Pitfalls

- Ensure a working bulb.
- Appropriately set the microscope's oculars for the operator's pupillary distance.
- Ensure all dials and knobs are firmly set.
- Inadequate focusing
- Patient noncompliance

46.7 Fluorescein Testing

46.7.1 Indications and Uses

- Examination of the cornea under a cobalt blue filter on a slit lamp or Wood's lamp: The chemical fluoresces under blue light and appears green under these circumstances.
- Useful for detecting corneal abrasions, corneal foreign bodies, and infections of the eye.

46.7.2 Procedure

- Fluorescein is typically used after installation of a topical anesthetic, which provides patient comfort especially for those with significant pathology.
- The fluorescein strip comes packaged in single-use wrappers (Fig. 46.8).
- Hold the strip by the white end and wet the orange end *lightly* because heavily moistening the strip may cause a significant amount of dye to be present, obscuring the examination.
  - A dry strip may be used, but it may irritate the patient's eye especially if already sensitive.
- If this occurs, the clinician may use tissue paper to gently blot the excess solution away.
  - The exception to this is the performance of the *Seidel test*, which is used to assess the eye for potential perforation (Fig. 46.9).
  - The clinician will instill a large amount of dye into the eye by wetting the orange strip copiously.
  - The clinician will next examine the eye for a stream of fluid leaking from the ruptured globe.
  - This stream will fluoresce green or blue in distinction to the rest of the globe, which appears orange.
- The choice of solution to wet fluorescein strip is up to the clinician because saline, tap water, or the recently used anesthetic solution may be used safely.
- Place the now-wet orange end on the lower lid of the patient's eye.
- Ask the patient to blink several times to allow the solution to spread evenly.
- The clinician may use a Wood's lamp, penlight, or the cobalt blue filter on the slit lamp to examine the now-stained eye.
  - The slit lamp is preferable owing to the potential for missing small abrasions.

Fig. 46.8 A typical fluorescein strip package

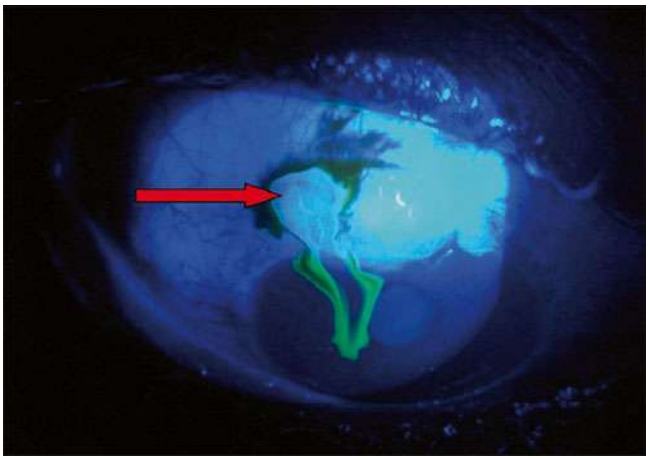

Fig. 46.9 A positive Seidel test (Reproduced with permission from: Lingam V, Panday M, George R, Shantha B. Management of complications in glaucoma surgery. *Indian J Ophthalmol*. 2011;59(Suppl1): S131–40)

46.7.3 Complications

- There is a theoretical risk of the development of superficial punctate keratitis from instillation of topical anesthetic before fluorescein testing.
- Discoloration of soft contact lenses.
- Potential for infection using premixed fluorescein solution.

46.7.4 Pitfalls

- Contact lens wearers should remove their lenses because the fluorescein will permanently stain the lens.
  - The wearer should not put the contacts back in for several hours.

Selected Reading

DuBois L. The slit lamp examination. In: DuBois L, Ledford JK, Daniels K, Campbell R, editors. Clinical skills for the ophthalmic examination: basic procedures. 2nd ed. Thorofare: Slack; 2006. p. 61–70.

Galor A, Jeng BH. Red eye for the internist: when to treat, when to refer. Cleve Clin J Med. 2008;75:137–44.

Lang GK. Ophthalmology: a short textbook. Stuttgart: Thieme; 2000.

Ledford JK, Sanders VN. The slit lamp primer. 2nd ed. Thorofare: Slack; 2006.

47 Eye Irrigation

Bobby K. Desai

47.1 Indications

- Chemical burns to the eye
- Removal of superficial foreign bodies

47.2 Contraindications

- For suspected globe perforation, extreme care must be taken to not exacerbate the injury.

47.3 Materials and Medications

- Irrigating device—the Morgan lens (Fig. 47.1)
- Saline solution or lactated Ringer's (preferably warmed)
- Topical anesthetic drops (Fig. 47.2)
- Basin to secure the extruded solution
- Intravenous (IV) tubing to attach the IV bag to the Morgan lens
- pH paper (Fig. 47.3)

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_47

Fig. 47.1 The Morgan lens and packaging

Fig. 47.2 Examples of topical anesthetics

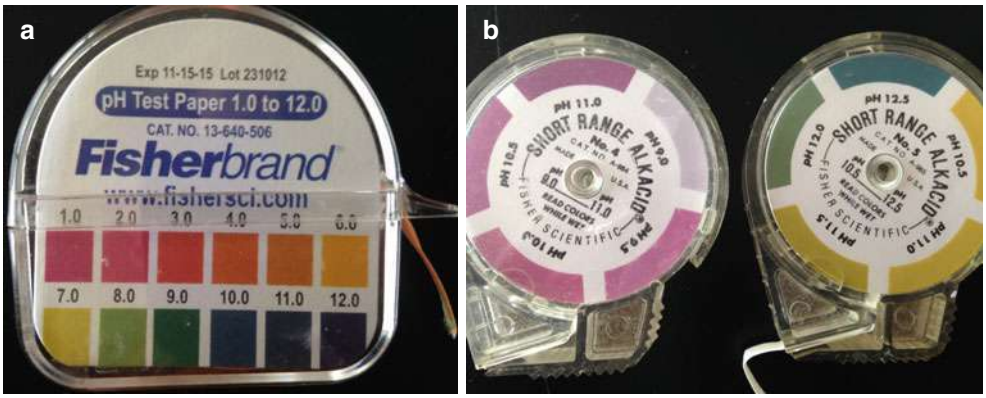

Fig. 47.3 (**a**, **b**) Examples of pH paper

47.4 Procedure

1. Informed consent is generally not required, although a thorough explanation to the patient is warranted.
2. Obtain pH of the eye *before* instilling anesthetic drops.
3. Anesthetize the eye(s) with topical anesthetic of choice by instilling drops within the lower lid and then asking the patient to blink several times in order to evenly distribute the solution.
4. Ensure there are no foreign bodies on the inside of the upper lids by inverting the upper lid.
   - Particulate foreign bodies may be removed with moistened cotton tip applicators.
5. After adequate anesthesia is ensured, place one end of the Morgan lens within the fornix of the upper lid (Fig. 47.4).
6. Next, gently retracting the lower lid will ensure smooth placement of the remaining portion of the Morgan lens (Fig. 47.5).
7. Using the end of the Morgan lens, screw in the prepared IV tubing (Fig. 47.6).
8. Attach the end of the IV tubing to the saline bag and place at height to allow for gravity to ensure a smooth flow of solution.
   - Continue to irrigate the eye until desired pH is obtained.
9. To remove the Morgan lens, use the opposite technique for insertion.

Fig. 47.4 Inserting the Morgan lens under the upper lid

Fig. 47.5 Inserting the Morgan lens under lower lid

Fig. 47.6 The Morgan lens in place, ready to be attached to IV tubing for saline irrigation

47.5 Complications

- Corneal abrasions may be caused by the Morgan lens and are treated in the usual fashion.
- Deep corneal injury may occur with inadequate irrigation.

47.6 Pearls and Pitfalls

- Note that alkali burns will require significant irrigation and more topical anesthesia may be required.
- Ophthalmological consultation may be required, especially for alkaline and hydrofluoric acid burns.

Selected Reading

Lang GK. Ophthalmology: a short textbook. Stuttgart: Thieme; 2000.
Rhee DJ, Pyfer MF, Rhee DM, editors. The Wills Eye manual: office and emergency room diagnosis and treatment of eye disease. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 1999.

48 Corneal Foreign Body Removal

Bobby K. Desai

48.1 Indications

- Presence of a corneal foreign body (Fig. 48.1)

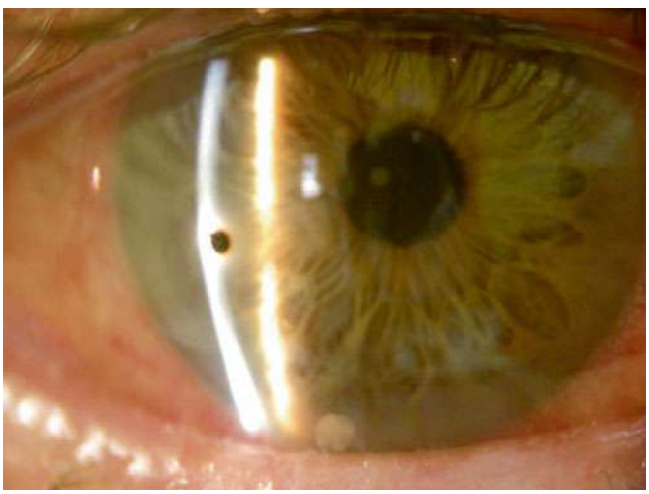

Fig. 48.1 Corneal foreign body (With kind permission from Springer Science+Business Media: Zuberbuhler B, Tuft S, Gartry D, Spokes D. Ocular Surface and Reconstructive Surgery. In: *Corneal Surgery*. 2013. 29–48)

48.2 Contraindications

- An uncooperative patient may require ophthalmological consultation as well as sedation.
- Suspected foreign bodies from high-velocity injuries must be referred to an ophthalmologist subsequent to initial evaluation.
- Any signs of globe penetration (e.g., hyphema) require emergent ophthalmological consultation.
- Overt globe rupture.
- Evidence of an inflammatory process such as iritis requires emergent ophthalmological consultation.
- Foreign bodies associated with corneal burns secondary to chemical exposure such as alkaline corneal burns will require emergent ophthalmological consultation.

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_48

48.3 Materials and Medications

- Eye chart for visual assessment
- Cotton tip applicator
- Slit lamp or other source of magnification
- 27-Gauge needle or ophthalmic foreign body needle (Fig. 48.2)
- Topical ophthalmic anesthetic
- Ophthalmic burr (Fig. 48.3)

Fig. 48.2 Foreign body needle

Fig. 48.3 Ophthalmic burr drill

48.4 Procedure

1. Informed consent may be required.
2. Visual acuity and a formal assessment of the eye should be done and documented.
3. Consider intraocular foreign body.
   - Computed tomography (CT) scan may be used to assess for the presence of an intraocular foreign body.
4. Eversion of the upper lid should be performed to evaluate for retained foreign body under the lid.
   - These usually can be removed by:
     - A moistened cotton tip applicator
     - Irrigation
5. Magnification is preferable, but some foreign bodies may be large enough to see with the naked eye.
6. Anesthetize the eye with topical anesthetic.
7. For multiple loose foreign bodies, removal may be attempted with irrigation after appropriate anesthesia.
8. Removal of the foreign body may be attempted with a moistened cotton swab.
   - Metallic foreign bodies recently embedded within the cornea may be difficult to remove in this fashion.
   - However, during the healing process, some metallic foreign bodies may be pushed closer to the surface during reepithelialization of the cornea. These potentially can be removed with this technique.
     - Care must be taken to avoid leaving a rust ring that could permanently stain the cornea.
9. A 27-gauge needle bent at a 90° angle may be used to remove foreign bodies by gently prying it out.
   - An advantage of using a needle is that any rust ring as well as any metallic foreign body can potentially be removed.
10. Technique: A slit lamp may be used to magnify the area.
    - See slit lamp chapter (Chap. 46) for details.
    - The patient should be informed about the procedure; formal informed consent may not be required.
    - Proper positioning is critical for success.
    - The patient's head should be fully forward and firmly placed against the head rest.
    - The patient's hands may be placed on the sides of the slit lamp stage for stabilization.
    - Apply topical anesthesia to the cornea.
11. Using the patient's cheek as a bolster potentially avoids significant movement of the needle if the patient unexpectedly moves.
    - Other modalities for stabilization include supporting the elbow by placing on a box or using towels.
12. Have the patient gaze at one point in the far distance.
    - Using the needle or burr as a scoop, gently manipulate the foreign body out.
      - Using a burr can allow for the total removal of any rust ring.
13. Once the foreign body is removed, patching is not required.
14. Consider application of topical antibiotics.
15. Tetanus prophylaxis should be given as for open other wounds.
16. Arrange follow-up with a primary care physician or preferably an ophthalmologist.

48.5 Complications

- Forceful attempts to dislodge a foreign body may result in corneal perforation.
- Incomplete removal of a foreign body.

Selected Reading

Lang GK. Ophthalmology: a short textbook. Stuttgart: Thieme; 2000.

Rhee DJ, Pyfer MF, Rhee DM, editors. The Wills eye manual: office and emergency room diagnosis and treatment of eye disease. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 1999.

Thomas SH, White BA. Foreign bodies. In: Marx J, Hockberger R, Walls R, editors. Rosen's emergency medicine: concepts and clinical practice. 7th ed. Philadelphia: Mosby; 2010. p. 715–32.

Ultrasound Evaluation of Retinal Detachment

49

Shalu S. Patel, L. Connor Nickels, and Rohit Pravin Patel

49.1 Indications

- To aid in the evaluation of vision loss or change in vision

49.2 Contraindications

- Relative: Suspicion of increased intraocular pressure
- Relative: Suspicion of globe rupture (see Chap. 50)

49.3 Materials and Medications

- Bedside ultrasound machine with high-frequency (7.5- to 10-MHz) linear transducer
- Transparent adhesive such as Tegaderm (optional)
- Sterile ultrasound gel
- Sterile gauze

49.4 Procedure

1. Position the patient supine with his or her eyes closed. The eyelid may be taped closed with a transparent adhesive if desired (Fig. 49.1).
2. Place a liberal amount of ultrasound gel over the eyelid.
3. Place the ultrasound probe over the eye in a transverse position.
4. Adjust the depth of the ultrasound so that the whole eye fills the screen.
5. Scan through the eye fully in the transverse and sagittal planes.
6. When viewing the ultrasound image, the normal eye is a circular hypoechoic structure (Fig. 49.2). The structures should be evaluated from anterior to posterior.
7. Identify the cornea. This is the first thin hypoechoic line in the anterior eye.
8. Continuing posteriorly, identify the anterior chamber, which is an anechoic area bordered by the cornea, iris, and lens. The iris is an echogenic linear structure, and the normal lens is anechoic.
9. Identify the vitreous chamber, the large anechoic region posterior to the lens.
10. Carefully evaluate the posterior globe. The normal retina cannot be distinguished from the other choroidal layers on ultrasound.
11. A detached retina will appear as a hyperechoic linear floating membrane in the posterior vitreous chamber (Fig. 49.3).
12. In the retrobulbar region, the optic nerve can be identified as a hypoechoic linear structure perpendicular to the globe (Fig. 49.3).

S.S. Patel, MD
Department of Emergency Medicine, Florida Hospital Tampa, Florida Hospital Carrollwood, Tampa, FL, USA
e-mail: shalu314@gmail.com

L.C. Nickels, MD, RDMS (✉) • R.P. Patel, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; rohitpatel@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_49

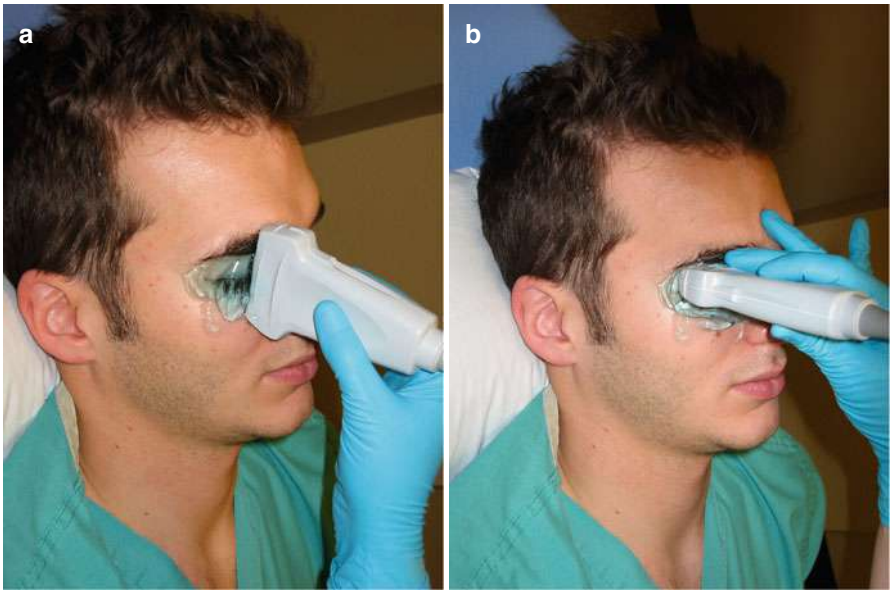

Fig. 49.1 Probe positioning with Tegaderm applied to orbit: (**a**) sagittal view, (**b**) axial view

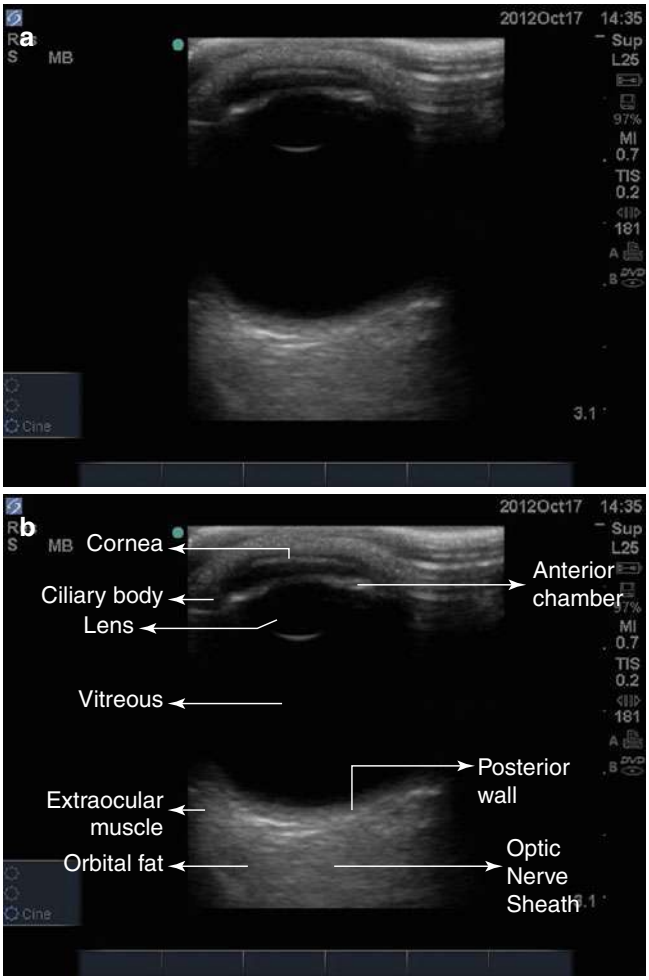

Fig. 49.2 Ocular ultrasound with normal anatomy findings ((**a**) without labels, (**b**) with anatomy labeled)

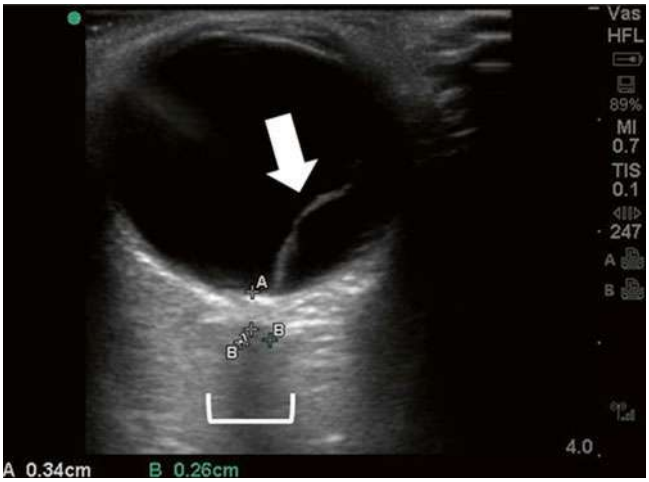

Fig. 49.3 Retinal detachment appears as a hyperechoic linear structure (*white arrow*). *White* bracket shows the optic nerve sheath shadow (see Chap. 50 for more information)

49.5 Complications

- Conjunctival infection
- Increased intraocular pressure (if too much pressure is applied)

49.6 Pearls and Pitfalls

- Pearls
  - Use a transparent adhesive such as Tegaderm to decrease the risk of conjunctival infection.
  - Use a liberal amount of gel to obtain the best images.
  - Use minimal pressure to obtain the best images.
  - False positives on ocular ultrasound may occur with disk edema or vitreous hemorrhage.
- Pitfalls
  - Subacute retinal detachments may be missed on ocular ultrasound.

Selected Reading

Blaivas M. Bedside emergency department ultrasonography in the evaluation of ocular pathology. Acad Emerg Med. 2000;7:947–50.

Blaivas M, Theodoro D, Sierzenski PR. Elevated intracranial pressure detected by bedside emergency ultrasonography of the optic nerve sheath. Acad Emerg Med. 2003;4:376–81.

Shinar Z, Chan L, Orlinksy M. Use of ocular ultrasound for the evaluation of retinal detachment. J Emerg Med. 2011;4:53–7.

Whitcomb MB. How to diagnose ocular abnormalities with ultrasound. AAEP Proc. 2002;48:272–5.

50 Ultrasonography in the Evaluation of Intraocular Pathology

Benjamin M. Mahon, Marie-Carmelle Elie, L. Connor Nickels, and Rohit Pravin Patel

50.1 Indications

- Eye trauma
- Loss of vision or decreased vision
- Head injury
- Ocular pain
- Suspected foreign body

50.2 Contraindications

- None

50.3 Materials

- Linear probe (7.5–10 MHz)
- Tegaderm or other protective eye covering
- Copious amounts of water-soluble transmission gel

50.4 Procedure

1. Begin by asking the patient to close her or his eye, and then apply the Tegaderm or other suitable eye cover over both eyes.
2. Apply a copious amount of transmission jelly over both eyes.
3. Gently apply the ultrasound probe to the outer eyelid. If the operator has used a sufficient enough amount of jelly, the probe itself should not need to be in actual contact with the eyelid. This becomes of critical importance in suspected globe trauma or intraocular foreign body, when excess manipulation of the intraocular contents could precipitate further damage or worsening of present damage. Ensure the pressure is gentle enough to prevent this, but sufficient to obtain good imaging (see Chap. 49).
4. Image the normal eye first, to become familiar with the patient's anatomy (see Chap. 49), before proceeding to the suspected abnormal eye.
5. Start with low gain.
6. Ask the patient to move the eye slowly from the left to the right. This is called kinetic echography and will provide information elaborated upon later. Have the patient keep her or his eye straight ahead, as the operator phases up and down and side to side, between the transverse and the sagittal planes.
7. Increase gain slowly while the patient is moving the eye. Shifting from low gain to higher gain during the examination will help to prevent overlooking certain disease processes. As a fluid-filled structure, the eye provides its own acoustic window, and imaging is generally easy.

50.5 Complications

- Through excessive pressure applied to the eye, as described previously, further disruption of the intraocular contents and worsening of preexisting injury can occur.

50.6 Pearls and Pitfalls

- Pitfalls
  - Applying too much pressure to the globe and disrupting already damaged intraocular contents
  - Failing to image with full spectrum of gain
  - Using an insufficient amount of gel

B.M. Mahon, MD
Ponciana Medical Center, Kissimmee, FL, USA
e-mail: faust2k9@ufl.edu

M.-C. Elie, MD, RDMS (✉) • L.C. Nickels, MD, RDMS
R.P. Patel, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: elie@ufl.edu; cnickels@ufl.edu; rohitpatel@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_50

50.7 Specific Disease Processes/ Pathology [1]

50.7.1 Vitreous Hemorrhage (Fig. 50.1)

- Best seen in high gain.
- Clinical: Very common pathology. Seen as "floaters" on the visual field. Can cause blindness if large.
- Ultrasound: Appearance may vary. Early on in the evolution of a hemorrhage, these can be seen as small dots or mobile linear opacities that float freely in the back of the posterior chamber with eye movement. Has been described as a "snow-storm appearance." As hemorrhage ages, the blood organizes and membranes form, sometimes layering inferiorly owing to gravity.

50.7.2 Vitreous Detachment

- Best seen with higher gain.
- Clinical: Common in older patients. Presents as "flashers." Vitreous gel loses its attachment from the internal membrane lining.
- Ultrasound: Seen as "swaying seaweed." There are layering opacifications. Also seen as "snow-globe appearance." Concurrent vitreous hemorrhage is also common.
- Note: A small percentage of patients also have a retinal tear, so ophthalmology follow-up is necessary.

50.7.3 Retinal Detachment (Figs. 50.2 and 50.3)

- Best seen with lower gain.
- Clinical: Acute loss of vision, usually painless, often preceded by "flashers" and/or "floaters" as the neurosensory component of the retina pulls away from the retinal pigment epithelium. *An ophthalmological emergency.* Can be caused by ocular trauma or no known precipitant at all.
- Ultrasound: Seen as a taut, hyperechoic, linear opacity that moves with the eye, attached to the optic disc, in the posterior region of the globe.

Fig. 50.1 Vitreous hemorrhage

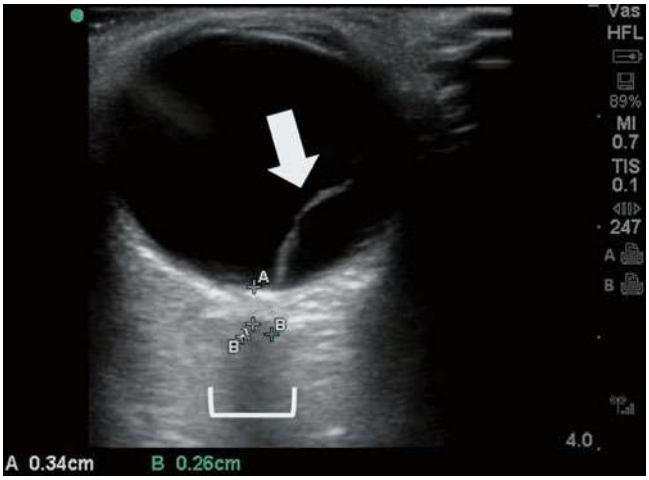

Fig. 50.2 Retinal detachment ultrasound image

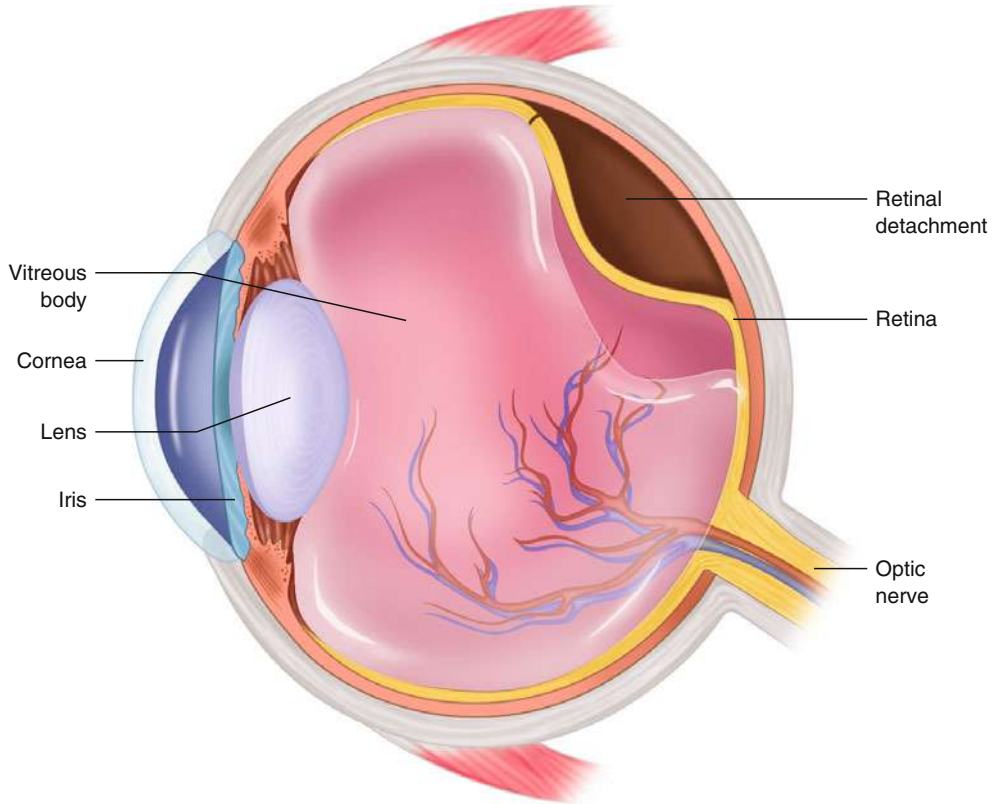

Fig. 50.3 Retinal detachment graphical representation

50.7.4 Intraocular Foreign Body (Fig. 50.4)

- Clinical: Usually a preceding episode of trauma, often from metalworking or landscaping.
- Ultrasound: Foreign bodies are seen as hyperechoic substances in the eye often with distal shadowing. Certain shadow patterns distal to the foreign body, as well as comet tails, can help to differentiate the type of foreign body material.

50.7.5 Globe Rupture (Fig. 50.5)

- Clinical: Usually the result of trauma. Frequently, the mechanism causing the injury has subsequently made the eye examination itself impossible, if edema, blepharospasm, hyphema, or distorted anatomy makes visualization of the posterior chamber too difficult. In these circumstances ultrasound can be immeasurably helpful.
- Ultrasound: Findings include decreased size of the anterior chamber, posterior chamber, or both, often with buckled sclera. Globe rupture should be suspected when the injury mechanism is present and any major distortion of the ocular anatomy is appreciated.

Fig. 50.4 Foreign bodies

Fig. 50.5 Globe rupture

50.7.6 Increased ICP

- Clinical: Increased ICP is encountered in a number of pathological processes, such as closed head injury from trauma, altered mental status, pseudotumor cerebri, space-occupying lesion, or any other process causing intracranial injury. Physical examination may be suggestive of an intracranial process, but often key examination findings such as papilledema are not appreciated, owing to both limitations from the patient and their mental status, as well as limitations in the skill of the ED doctor, or the ability to dilate the eyes sufficiently, to appreciate the finding.
- Ultrasound: The size of the *optic nerve sheath diameter* is an indirect measurement of ICP. It is visualized as a hypoechoic linear strip radiating posteriorly from the edge of the back of the globe. A normal diameter is 5.0 mm or less. This measurement is taken at 3 mm back from the posterior edge of the globe (Fig. 50.6). Take measurements bilaterally to compare. Two or more measurements should be taken and then averaged on each side.
- A diameter greater than 5.0 mm, at 3.0 mm distal from the posterior globe border, is strongly suggestive of increased ICP and warrants further evaluation and imaging [2].

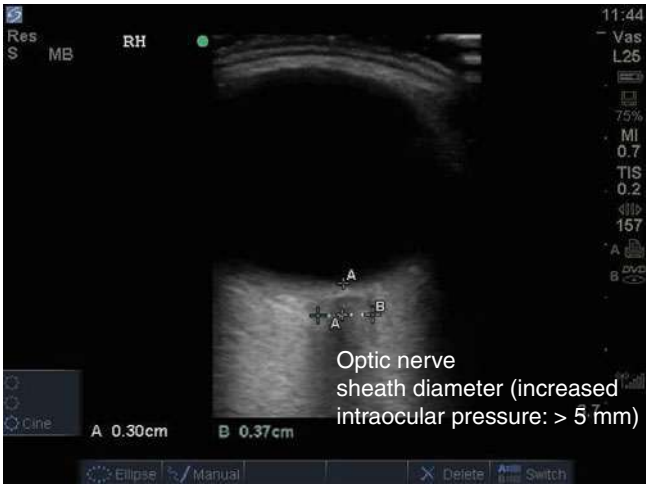

Fig. 50.6 Optic nerve sheath measurement

References

1. Blaivas M, Theodoro D, Sierzenski P. A study of bedside ocular ultrasonography in the emergency department. Acad Emerg Med. 2002;9:791–9.
2. Blaivas M, Theodoro D, Sierzenski P. Elevated intracranial pressure detected by bedside emergency ultrasonography of the optic nerve sheath. Acad Emerg Med. 2003;10:376–81.

Selected Reading

Mustafa M, Montgomery J, Atta H. A novel educational tool for teaching ocular ultrasound. Clin Ophthalmol. 2011;5:857–60.

Qureshi MA, Laghari K. Role of B-scan ultrasonography in preoperative cataract patients. Int J Health Sci (Qassim). 2010;4:31–7.

Shinar Z, Chan IL, Orlinsky M. Use of ocular ultrasound for the evaluation of retinal detachment. J Emerg Med. 2011;40:53–7.

Vodapalli H, Murthy SI, Jalali S, Ali MJ, Rani PK. Comparison of immersion ultrasonography, ultrasound biomicroscopy and anterior segment optical coherence tomography in the evaluation of traumatic phacoceles. Indian J Ophthalmol. 2012;60:63–5.

Whitcomb MB. How to diagnose ocular abnormalities with ultrasound. AAEP Proc. 2002;48:272–5.

Yoonessi R, Hussain A, Jang TB. Bedside ocular ultrasound for the detection of retinal detachment in the emergency department. Acad Emerg Med. 2010;17:913–7.

51 Tonometry

Bobby K. Desai

51.1 Indications

- To measure intraocular pressure (IOP)

51.2 Contraindications

- A relative contraindication is the presence of an active or indolent infection about the cornea or conjunctiva, unless a one-time-use device is used (e.g., a Tono-Pen XL).
- Recent trauma to the eye.
- Uncooperative patients because improper technique may cause damage to the eye.

51.3 Overview

- Tonometry is the measurement of IOP obtained by evaluating the resistance of the eye to indentation by a force applied to it.
- Can be obtained by several methods.
- Two methods are routinely used in an emergency department setting:
  - *Impression tonometry*: measures the indentation made by a plunger carrying a known amount of weight.
  - More weight can be added to the apparatus.
  - The more weight needed to indent the cornea results in a higher IOP reading.
    - The Schiøtz tonometer is the most commonly used apparatus to utilize this method (Fig. 51.1).
- *Electronic indentation tonometry*: does not exert pressure on the eye (Fig. 51.2).

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_51

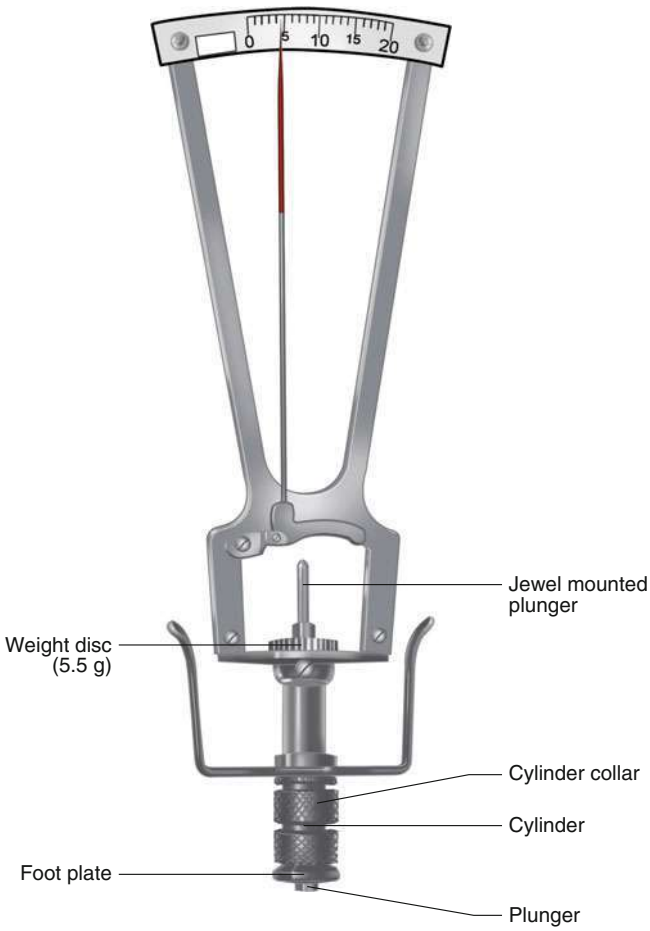

Fig. 51.1 Schiøtz tonometer

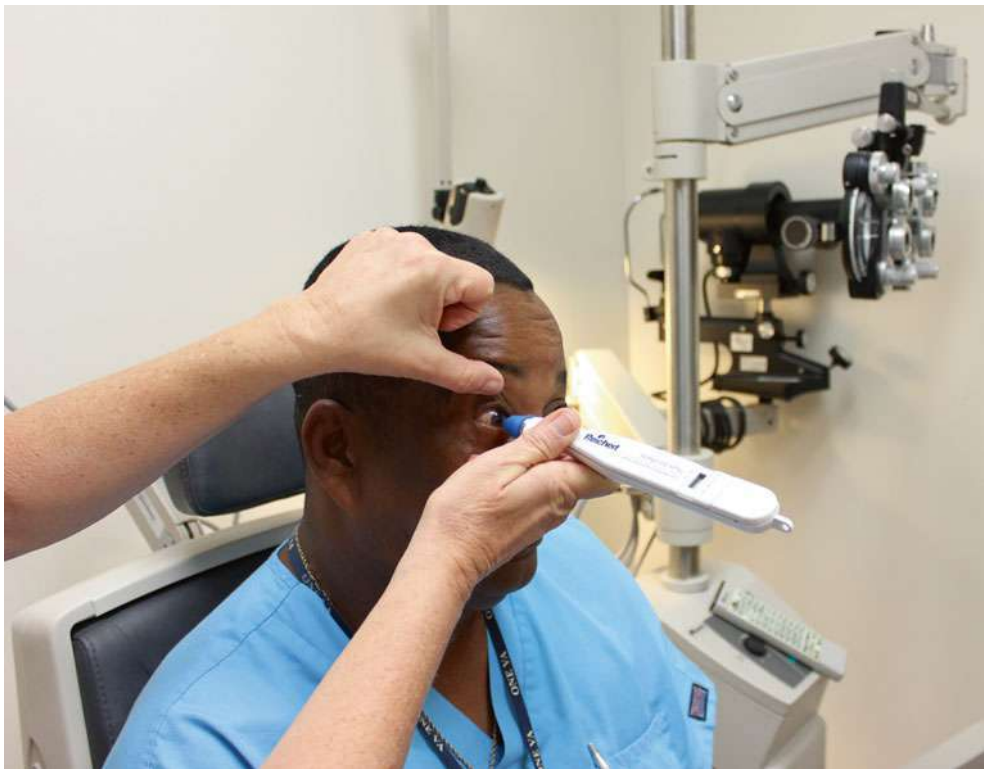

Fig. 51.2 Tono-Pen

51.4 Schiøtz Tonometry

51.4.1 Procedure: Schiøtz Tonometry (Fig. 51.3)

1. Carefully clean the apparatus between each patient.
   - Rubbing alcohol may be used.
   - Because it is metal, it can be autoclaved.
2. Calibrate the Schiøtz tonometer with the metal test block provided.
   - Test by placing the apparatus directly on the metal block.
   - The needle should be at "0" at the far end of the scale.
     – If not, loosen the screw at the base of the needle to rezero it.
     – Ensure that the needle is completely straight because any bend will produce an erroneous reading.
3. An explanation of the procedure is helpful because patient cooperation is critical for accurate results.
4. Anesthetize the eye with topical ophthalmic anesthetic of choice.
5. The patient should be in a recumbent position.
   - Have the patient focus on an area of the ceiling.
6. Hold the instrument with the aid of the curved arms at the side of the tonometer.
   - The operator can rest her or his hand on the patient's cheek or forehead to maintain stability.
7. Gently rest the tonometer on the patient's eye such that the instrument is centered on the eye and the instrument is completely vertical; no pressure should be exerted on the eye.
8. Note the scale reading.
9. Lift the Schiøtz directly off the cornea to avoid injury.
10. Using the table provided with the Schiøtz, the operator may convert the scale reading into the IOP.
    - The scale is inversely proportional to the actual IOP.
    - If the scale is low (i.e., high IOP), the additional weights provided with the instrument may be used and the patient retested.

51.4.2 Pitfalls

- Ensure the plunger is clean because it can transmit infection.
- False readings may be obtained without proper calibration.
- Placing pressure on the instrument will cause false readings.

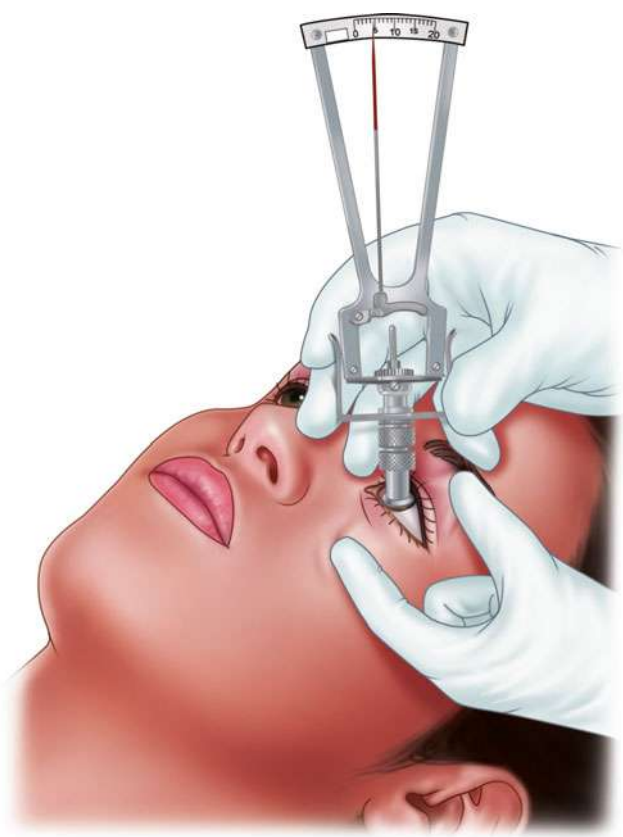

Fig. 51.3 Correct use of Schiøtz tonometer

51.5 Tono-Pen Tonometry

51.5.1 Overview

- Electronically measures IOP
  - Combines applanation and indentation tonometry
- Uses pressure-sensitive electronics to average four successive readings and displays the reading and a reliability factor digitally

51.5.2 Calibration of the Tono-Pen

- Should be performed once daily
- Hold the instrument with the tip down (Fig. 51.4).
- Press the black button twice.
  - The letters "CAL" will appear in the display.
- Press the black button again.
  - After a few seconds the word "UP" will appear.
- Rotate the instrument so that the tip points upward.
  - If "GOOD" appears, the Tono-Pen is ready for use.
  - If "BAD" appears, repeat the process until "GOOD" appears.
    - If "GOOD" does not appear, the device cannot be used.

Fig. 51.4 Tono-Pen calibration position

51.5.3 Procedure: Tono-Pen Tonometry

1. An explanation of the procedure is helpful because patient cooperation is critical for accurate results.
2. Anesthetize the eye with the topical ophthalmic anesthetic of choice.
3. Place a probe cover on the unit.
4. The patient should be in a comfortable position because the unit can be used in any position.
   - Have the patient focus on a specific area.
5. The best way to hold the instrument is similar to that of a pen.
   - Ensure that the digital readout is visible.
6. The operator can rest his or her hand on the patient's cheek or forehead to maintain stability.
7. Hold the unit perpendicular to the patient's cornea.
8. Press the black button only once.
   - If "ICALI" is seen, followed immediately by a single row of dashes [– – – –], it indicates that the Tono-Pen requires calibration before it will measure.
   - If "====" is seen and a "beep" is heard, the unit is ready.
   - Proceed with measurement.
9. Gently touch the Tono-Pen to the cornea and withdraw (Fig. 51.5). Repeat several times.
   - Indentation is not required.
10. The unit will "chirp" and a digital reading will be displayed if a valid reading is obtained.
11. After four valid readings are obtained, the average of these measurements as well as a single bar that signifies statistical reliability will appear on the readout.
    - After the final beep if the liquid crystal display (LCD) readout shows "----," not enough valid readings were obtained.
    - In this case, the measurement must be repeated.

51.5.4 Pitfalls

- Unsuccessful calibration mandates a repeat attempt at calibration.
- Loosen the Ocu-Film tip cover and repeat for multiple failed attempts.
- Press the reset button and reattempt calibration.
- Replace the battery if necessary.
- If all else fails, use another device or Schiøtz tonometer.

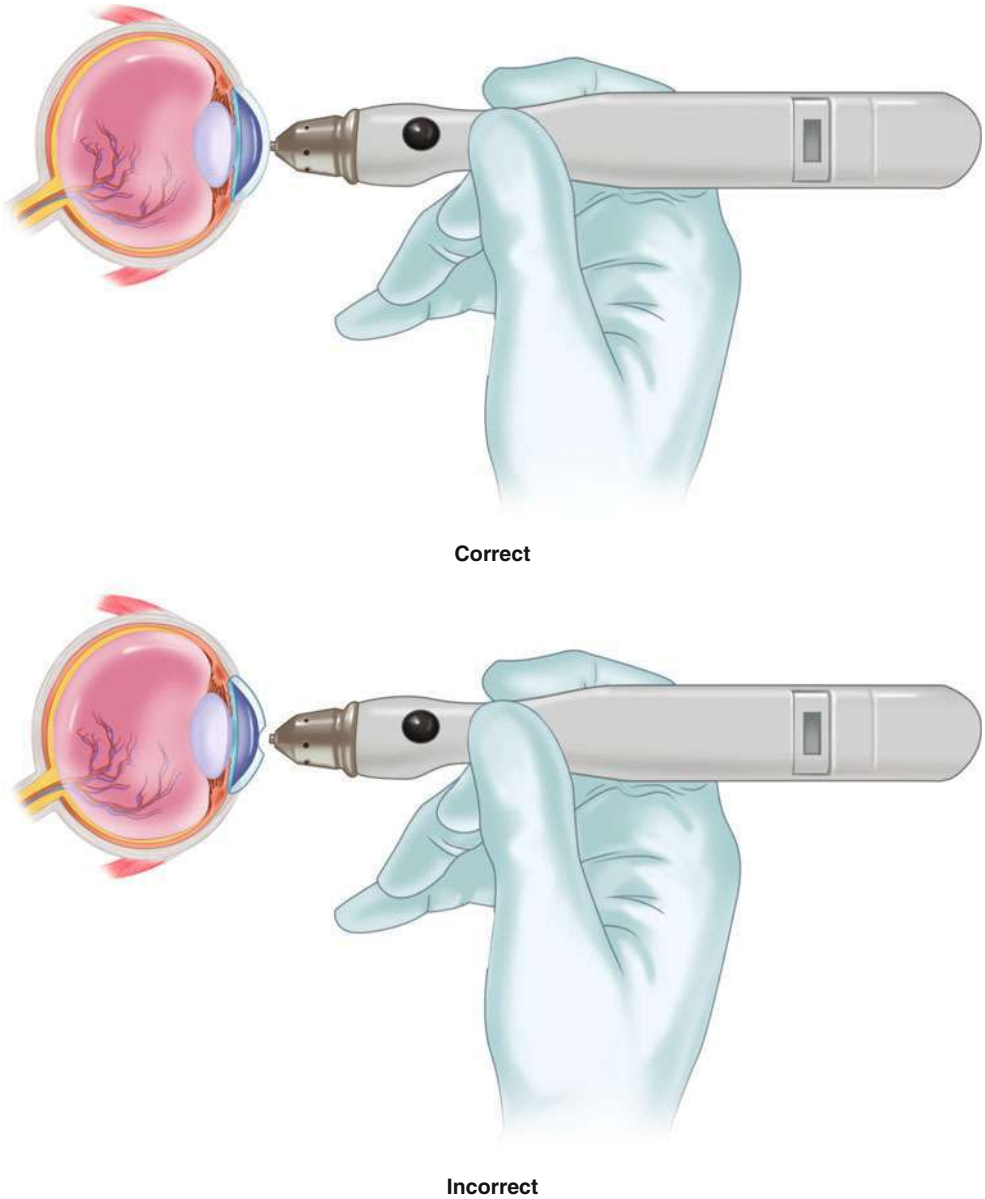

Fig. 51.5 Proper use of Tono-Pen

51.6 Complications

- Unusual if proper technique is used.
- If too much force is applied, a corneal abrasion may result.
- Infection when using an improperly sterilized Schiøtz tonometer.

Selected Reading

Lang GK. Ophthalmology: a short textbook. Stuttgart: Thieme; 2000.

Rhee DJ, Pyfer MF, Rhee DM, editors. The Wills eye manual: office and emergency room diagnosis and treatment of eye disease. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 1999.

Lateral Canthotomy

52

Benjamin M. Mahon and Bobby K. Desai

52.1 Indications

- Absolute indications: In the presence of presumed or confirmed retrobulbar hemorrhage
  - Acute visual loss
  - Intraocular pressure (IOP) greater than 40 mmHg (normal IOP is 10–20 mmHg)
  - Proptosis not amenable to retropulsion
- Relative indications: In the context of presumed or confirmed retrobulbar hemorrhage (Fig. 52.1)
  - Ophthalmoplegia
  - Cherry red macula
  - Profound eye pain
  - Afferent pupillary defect (Marcus Gunn pupil)
    - This defect is seen with the *swinging flashlight test.*
      - Shine a light into both eyes; a normal response is equal constriction of both pupils.
      - In those patients with an afferent papillary defect, when light is swung from an unaffected pupil to an affected pupil, the pupil will seem to paradoxically dilate, rather than constrict. The pupil on the nonaffected side will similarly dilate as light is shown into the affected eye. This results from injury to the afferent fibers of cranial nerve (CN) II on the affected side, while the efferent fibers, innervated by CN III, remain intact.
  - This procedure is most effective if performed as soon as possible because irreversible vision loss can occur secondary to ischemia in as little as 90 min.

Fig. 52.1 Retrobulbar hemorrhage

B.M. Mahon, MD
Ponciana Medical Center, Kissimmee, FL, USA

B.K. Desai, MD (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_52

52.2 Contraindications

- Absolute
  - Globe rupture
- Relative
  - Suspected globe rupture: Heralded by a number of signs including hyphema, irregularly shaped pupil, exposed uveal tissue, or profound restriction of extra-ocular movement

52.3 Materials and Medications

- Straight hemostat
- Suture or iris scissors
- Forceps
- 1–2 % lidocaine with epinephrine
- Sterile gloves
- Sterile towels/drapes
- 4×4 gauze
- Face shield/mask
- Large-bore needle to withdraw anesthetic, 25-gauge needle to infiltrate
- 10-cc syringe
- Normal saline

52.4 Procedure

1. Position the patient
   - While waiting for the procedure to begin, the patient should initially be upright to produce any decrease in IOP that can be provided.
   - At the start of the procedure, lay the patient supine.
   - A cooperative patient is absolutely necessary because even slight movements can cause devastating iatrogenic injury.
   - If the patient is anxious or conscious, sedation may be required.
   - In extreme cases of altered mental status or combative trauma patients, endotracheal intubation and mechanical ventilation may be required.
2. Gently irrigate the affected eye to remove any debris.
3. Inject approximately 1–2 mL of 1–2 % lidocaine with epinephrine into the lateral canthus of the affected eye (Fig. 52.2).
4. Using a curved hemostat, gently crimp the skin over the lateral corner of the patient's eye down to the orbital rim for 3 min to establish hemostasis and set the boundaries for the incision.
   - Using forceps, use the hemostats to pick up the skin just crushed.
   - Raise the skin with forceps and then use scissors to make a 1- to 2-cm incision from the lateral corner of the eye extending laterally outward (Fig. 52.3).
     - This incision can often sufficiently decrease IOP.
     - It is feasible to remeasure the IOP at this point, and if it is still greater than 40 mmHg, proceed to the next step.
   - Visualize the lateral canthal tendon by retracting the inferior orbital lid inferiorly.
   - Direct the scissors along the lateral side of the orbital rim away from the globe, and cut the lateral canthal tendon inferiorly (Fig. 52.4).
   - At this point, again measure IOP, and if greater than 40 mmHg, proceed to the next step.
   - Direct the scissors along the lateral side of the orbital rim away from the globe, this time directed superiorly, and cut the superior crux of the lateral canthal tendon.

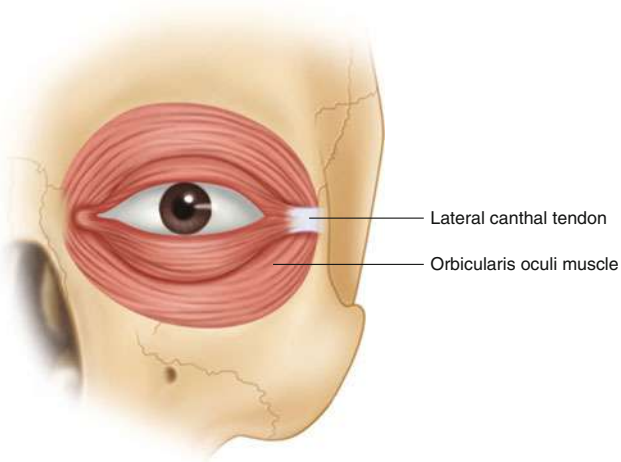

Fig. 52.2 Lateral canthal tendon

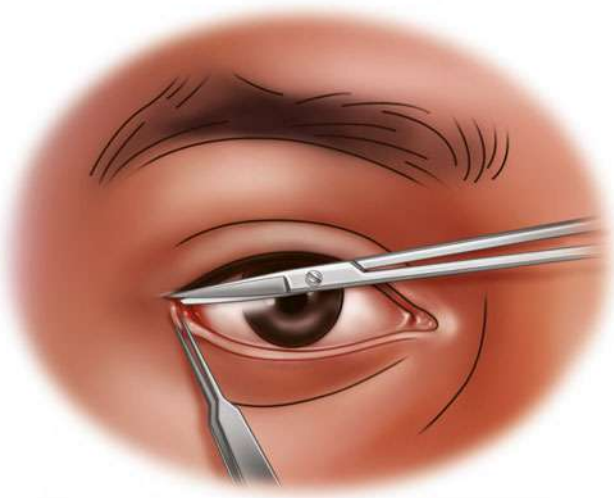

Fig. 52.3 Initial cut

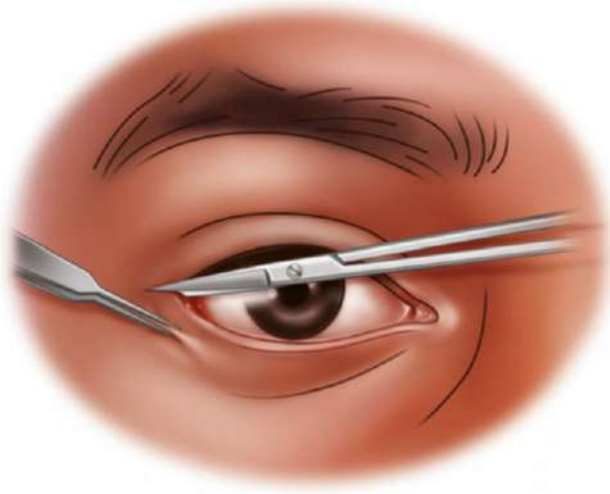

Fig. 52.4 Further cut to reduce intraocular pressure

52.5 Complications

- If one is making a superior incision, caution must be taken during this procedure to avoid injury to the lacrimal gland or artery.
- An additional complication of the procedure is ptosis due to iatrogenic injury to the levator aponeurosis, either partially or completely.
- The most obvious complication involves direct injury to the globe itself due to either operator error or poor patient control and immobility.
- Hemorrhage from inadequate hemostasis.
- Infection.

52.6 Pearls and Pitfalls

- Abrupt decrease in pain, decrease in IOP, and resolution of afferent pupillary defect will herald a successful procedure, assuming it was performed within an early enough time frame.
- Having an assistant present is beneficial.
  - In this case, the role of the assistant is to provide lateral retraction of the tissues to decrease likelihood of globe rupture.
- After cantholysis of the inferior component of the lateral canthal tendon, the lower lid will become lax as its attachment to the lateral wall is separated. This signifies a successful incision.
- Incisions made during this procedure generally heal without the need for suturing.
- Emergent ophthalmological consultation should ideally be sought before beginning this procedure.
- Emergent ophthalmological consultation is mandatory subsequent to the procedure.

Selected Reading

Goodall KL, Brahma A, Bates A, Leatherbarrow B. Lateral canthotomy and inferior cantholysis: an effective method of urgent orbital decompression for sight threatening acute retrobulbar haemorrhage. Injury. 1999;30:485–90.

McInnes G, Howes DW. Lateral canthotomy and cantholysis: a simple vision-saving procedure. CJEM. 2002;4:49–52.

Roberts JR, Hedges JR, editors. Clinical procedures in emergency medicine. 5th ed. Philadelphia: WB Saunders; 2009: Chap. 63.

Vassallo S, Hartstein M, Howard D, Stetz J. Traumatic retrobulbar hemorrhage: emergent decompression by lateral canthotomy and cantholysis. J Emerg Med. 2002;22:251–6.

Part VIII

Otorhinolaryngologic Procedures

53 Epistaxis Control

Benjamin M. Mahon and Bobby K. Desai

53.1 Etiology

- Anterior source: approximately 90 % of nosebleeds, usually Kiesselbach's plexus (Fig. 53.1)
  - Nose picking ("epistaxis digitorum")
  - Trauma
  - Infection
  - Nasal foreign body
  - Dry air
  - Atmospheric pressure alterations (e.g., increased altitude, lower arterial partial pressure of oxygen [PaO_2])
  - Allergies
  - Blood dyscrasias
  - Malignancy (e.g., leukemia, lymphoma)
- Posterior source: approximately 10 % of nosebleeds, usually sphenopalatine artery (Fig. 53.2)
  - Iatrogenic coagulopathy (e.g., warfarin, heparin, high-dose aspirin)
  - Blood dyscrasia
  - Liver failure
  - Renal failure
  - Malignancy
  - Older age

B.M. Mahon, MD
Ponciana Medical Center, Kissimmee, FL, USA

B.K. Desai, MD (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_53

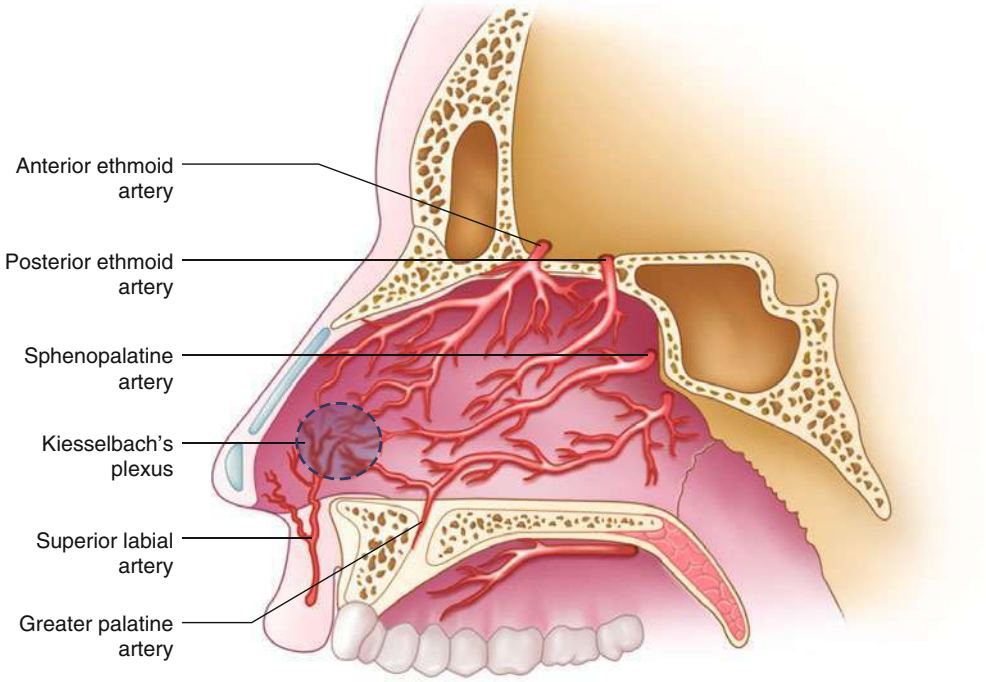

Fig. 53.1 Vascular supply to the nasal septum

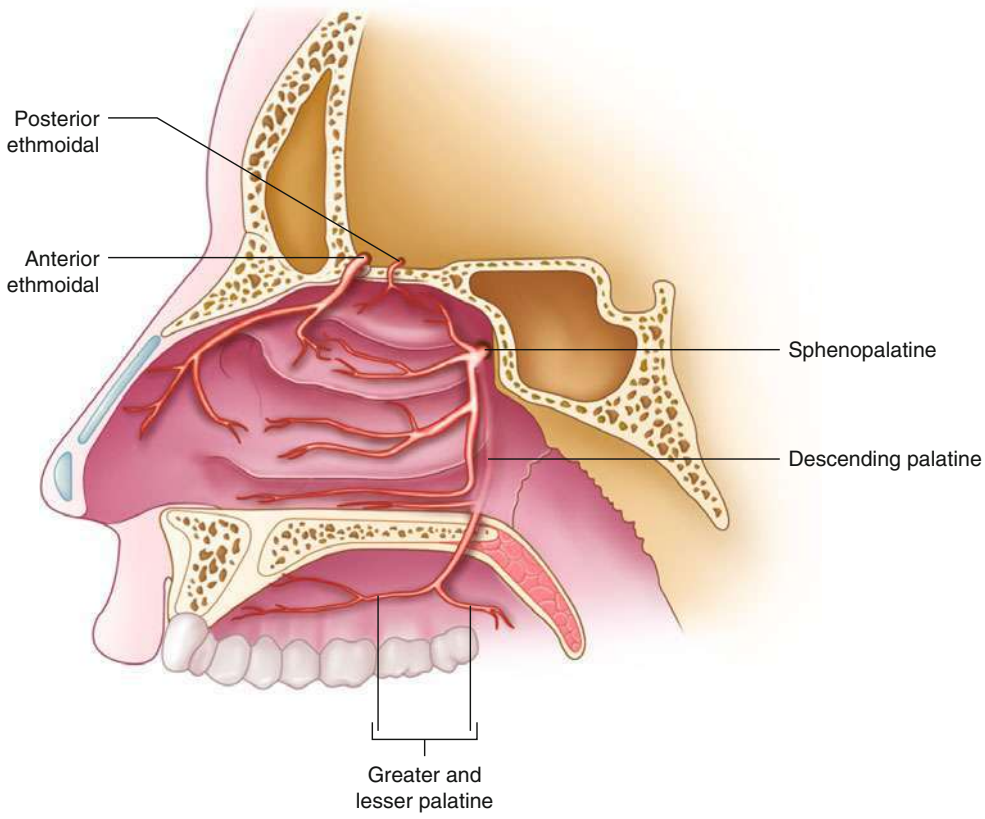

Fig. 53.2 Vascular supply of the lateral wall of the nose

53.2 Indications

- Acute/recurrent epistaxis

53.3 Contraindications

- Resolution of epistaxis episode before arrival
- Massive facial trauma

53.4 Materials and Medications

- Headlight with focused beam
- Gown
- Gloves
- Mask with visor
- Full-body drape or sterile towels
- Nasal packing material (any or all of the following, as needed [see later]): several cotton pledgets, 3–5 feet of ½-inch ribbon gauze (preferably enriched with petroleum jelly and bacitracin), anterior epistaxis balloon nasal pack
- Topical vasoconstrictor (e.g., 1 % pseudoephedrine or 1:1000 epinephrine)
- Topical anesthetic (e.g., 4 % lidocaine solution or 2 % tetracaine)
- Nasal speculum
- Bayonet forceps
- Anterior epistaxis balloon (e.g., Rhino Rocket)
- Any of the following commercial products, as needed (see later): Gelfoam, Surgicel, Crosseal, and FloSeal
- Foley catheter

53.5 Procedure: Anterior Epistaxis

- Initial preparation
  1. Assemble necessary equipment at the bedside.
  2. Have the patient seated upright, with head and neck in the sniffing position.
  3. Universal precautions are mandatory: gown, glove, and mask.
  4. Drape the patient with towels or a large body drape.
  5. If initial evaluation using the headlamp and nasal speculum with gentle spreading in a vertical fashion or a simple visual examination reveals the source (anterior vs posterior), proceed with the appropriate management pathway as determined later.
     - If clot obstructs visualization, it is recommended for the patient to gently blow her or his nose once to remove any easily friable clots, and then proceed with immediate inspection. If anterior nose bleeds, go to the next step.
- Management of anterior epistaxis:
  1. Initial step in the management of anterior nosebleeds involves pinching the patient's nostrils firmly for at least 10–15 min.
     - This can be done by the patient.
     - Simple pressure is very often enough to provide appropriate hemostasis in anterior nosebleeds. If it fails, proceed to the next step.
  2. A topical vasoconstrictor of choice (e.g., cocaine, oxymetazoline, 1 % pseudoephedrine) can be applied to the septum and lateral walls of the nose either topically with cotton-tipped applicators or by using soaked pledgets.
  3. Continue to apply firm pressure to the nares for 10–15 min, and recheck the nose for evidence of bleeding.
  4. If the bleeding site can be easily visualized, chemical cautery may be attempted with silver nitrate (Fig. 53.3).
     - This can be accomplished with the help of a nasal speculum and a headlight.
     - It is advised to cauterize both the site itself and 0.5 cm around the site, by holding the silver nitrate stick to the site for at least 20 s until the bleeding stops.
     - If bleeding then stops, cover the site with an absorbable gelatin foam or an oxidized cellulose such as Gelfoam or Surgicel or simply apply topical antibiotic to the site.
  5. If there is continued bleeding, the clinician may apply either two elongated cotton pledgets or a commercially available substitute (e.g., a Rhino Rocket or similar device [Fig. 53.4]) presoaked in a few milliliters or a 1:1 mixture of a topical vasoconstrictor and an anesthetic. Combinations can include the previously cited two agents, plus 2 % tetracaine or 4 % lidocaine.
  6. Universal instructions for the insertion of a commercially available anterior nasal pack include (Fig. 53.5):
     - Soak the fabric via the manufacturer's recommended solution if required.
     - Insert antibiotic ointment in the nares to facilitate insertion.
     - Insert the entire length of the balloon along the inferior surface of the nasal cavity.
     - The fluid within the nasal cavity should cause it to expand spontaneously while in the nares, after which the remaining exposed string can be taped to the cheek.
     - For those devices that require air instillation, read the manufacturer's instructions on how much air to instill within the device.

7. Another option is to provide an anterior nasal packing using a few feet of ½-inch ribbon gauze (Fig. 53.6).
8. It is preferable if this gauze is impregnated with petroleum jelly and is enriched with bacitracin ointment.
   – If it is not, bacitracin can be applied to the strip before its insertion into the nasal cavity.
9. To facilitate the insertion, bayonet forceps can be used to gently lay each successive strip on top of the one before it.
10. Begin inferiorly, laying each strip on top of the other, about three or four layers at a time, with the end points protruding from the opening of the nasal cavity.
11. Pack the strands firmly.
12. It may take several feet to pack the entire nasal cavity.
13. This may be both uncomfortable for the patient and time consuming for the provider, so ensure adequate time and copious topical anesthetic.
14. Parenteral analgesic or anxiolytics may be required for any of these techniques.
15. If none of these resources is available, another technique requires the use of two cotton pledgets gently inserted into the nostrils after vasoconstrictor/anesthetic topical solution has been sprayed into the nose either manually or with the addition of bayonet forceps, where they will remain for 20 min, supplemented by firm pressure to the nares as described previously.
   – If the bleeding has then stopped, they can be removed at this time. If it does not, it is advised to repeat the previous step a second time, this time with a longer period of pressure, before proceeding to the next step.
16. If all of these techniques fail to control bleeding, the clinician may apply a sealant spray or foam enriched with thrombin to enhance clotting.
   – Commercially available options include Crosseal or FloSeal.
17. Finally, if none of these techniques manage to halt what one is sure is only an anterior, albeit persistent, nosebleed, consult the otolaryngologist for further guidance.
   – Another consideration at this point may be that the uncontrollable anterior nosebleed is really a posterior nosebleed from an undetermined location.
   – It may be prudent to proceed to the algorithm for treatment of posterior epistaxis, even if a posterior bleed is not specifically confirmed.

53.5.1 Complications

- Using an inadequate amount of packing when inserting the ribbon gauze, the entire product may serve as a plug with potential aspiration risk, rather than as a tool for hemostasis.
- Otitis, sinusitis, and toxic shock syndrome.

53.5.2 Pearls and Pitfalls

- Pearls
   – Be sure to perform a reevaluation of both the posterior oropharynx and the posterior nasopharynx after insertion of any of the previously discussed devices, and inspect for continued bleeding/oozing to ensure hemostasis has been accomplished.
   – For those patients with nasal packs in place, consider the use of oral antibiotics to avoid subsequent sinusitis, otitis, or toxic shock syndrome.
      - Appropriate antibiotic choices include cephalexin, clindamycin, or amoxicillin–clavulanic acid.
      - The use of antibiotics for short-term packing has not been proven.
   – Consider admission for elderly patients or for those patients with potential airway complications such as those patients with chronic obstructive pulmonary disease (COPD) and for those at risk for aspiration.
   – Have the patient follow up with otolaryngology within 2–3 days for reevaluation and packing removal.
   – Provide adequate discharge instructions warning the patient against nose blowing, sneezing with the mouth closed, or any movements or actions that cause a Valsalva maneuver.
   – The patient should try to keep his or her head elevated for the next 24–48 h even while sleeping to prevent potential aspiration.
- Pitfalls
   – Discharging a patient without an adequate period of observation.
   – Failing to ensure appropriate follow-up.
   – Providing inadequate discharge instructions.
   – Mistaking an anterior nosebleed for a posterior nosebleed.
   – Leaving silver nitrite on exposed nasal mucosa can lead to iatrogenic nasal septal perforation.
   – Not informing the patient that minimal oozing of blood from the nose can be expected.

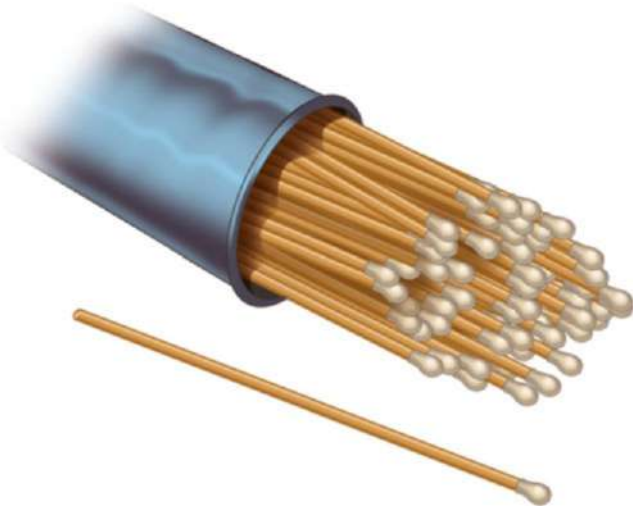

Fig. 53.3 Example of silver nitrate stick and packaging

Fig. 53.4 Example of commercial nasal pack

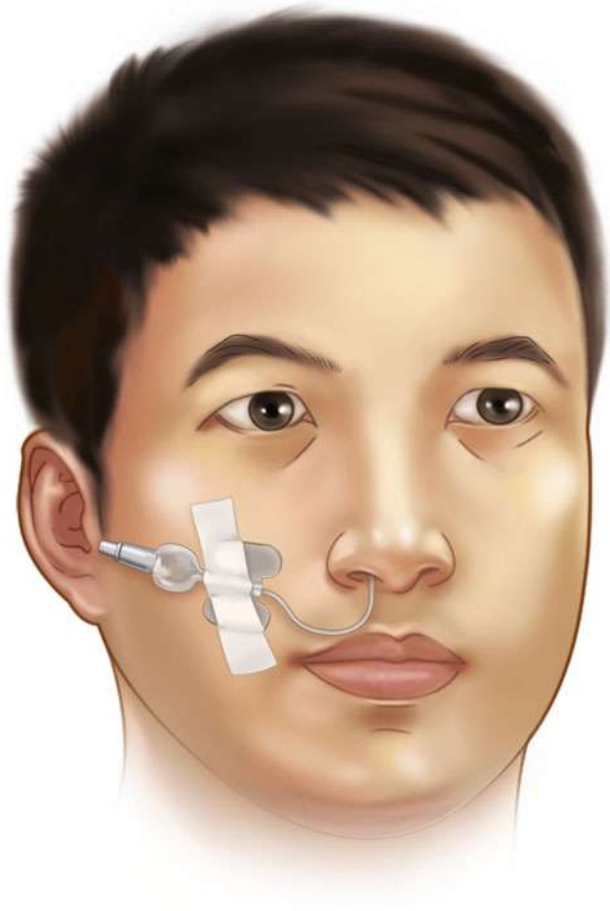

Fig. 53.5 Proper taping of a commercial anterior nasal pack

Fig. 53.6 Gauze method

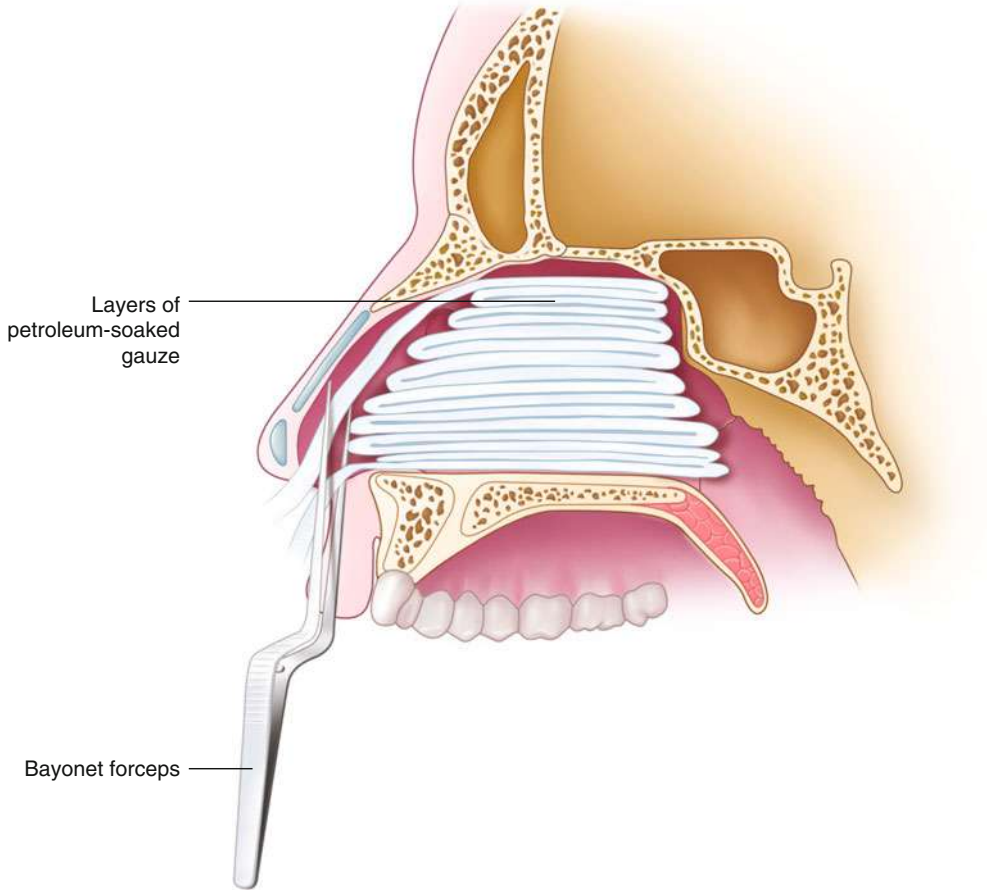

53.6 Procedure: Posterior Epistaxis

- Preliminary Steps
  1. Owing to the possibility of extreme exsanguination from a posterior source of epistaxis or airway obstruction from a large clot dislodgment or large volume of blood aspiration, the stability of the patient's airway, breathing, and circulation must be ensured first.
     - Appropriate interventions entirely depend on the presentation of the patient and therefore are not described in this chapter.
     - Patients with posterior epistaxis can lose a large volume of blood quickly, so this needs to be an initial consideration.
  2. Consider establishing intravenous access and placing the patient on a cardiac monitor.
  3. Because patients with posterior epistaxis tend to be either older, on various blood thinners, or to potentially have some type of coagulopathy, consider obtaining a complete blood count, prothrombin time, and partial thromboplastin time as part of the preliminary workup.
  4. Consider a blood type and screen.
- Management of posterior epistaxis:
  1. Inspect the nares as described in the previous section on anterior nosebleeds.
  2. Several of the techniques discussed may be attempted but may not be successful.
  3. Assemble necessary equipment at the bedside.
  4. Have the patient seated upright, with the head and neck in the sniffing position.
  5. Universal precautions including a gown, glove, and mask should be employed.
  6. Drape the patient with towels or a large body drape.
  7. Apply a combination of a topical vasoconstrictor and anesthetic to the nose.
     - An appropriate choice (e.g., 1 % pseudoephedrine or 1:1000 epinephrine in a 1:1 ratio with 4 % lidocaine or 2 % tetracaine)
     - Be aware, however, that there are situations in which the bleeding may be too brisk, and this step may need to be avoided owing to time constraints and the volume of bleeding. So it is not unheard of to be required to intervene in posterior epistaxis without this initial step.
  8. Otolaryngology should be consulted as early as possible in the workup of this condition, both for admission and in case the following techniques fail.
  9. An advised initial step for the treatment of posterior epistaxis is to proceed directly to insertion of a unilateral or, preferably, bilateral posterior nasal pack or balloon.
     - An example of these is the elongated version of the Rhino Rocket as described previously.
     - Insertion and securing of the product proceed essentially the same way, except that it is inserted deeper into the posterior oropharynx.
  10. Consider the use of a dedicated posterior balloon, such as a Nasostat or Epistat (Fig. 53.7).
     - These devices are double-balloon systems that are lubricated preferably with bacitracin (sterile jelly is also appropriate) and inserted into the posterior nasopharynx.
     - The posterior balloon is then inflated with the product-specific quantity of air.
     - The product is then withdrawn slightly to ensure that the posterior balloon is situated firmly in its desired location of the posterior nasal cavity.
     - The anterior balloon is then inflated with the device-specific quantity of normal saline, and the device is secured.
- If the previously discussed resources are not available, a clinician can insert an adult-caliber Foley catheter into the posterior nasopharynx, with inflation of the balloon, and gentle traction anteriorly to ensure the bulb is situated similarly to the Nasostat or Epistat, firmly in the posterior nasal cavity to tamponade off the bleeding (Fig. 53.8).
- If all of these techniques still fail to control bleeding, an otolaryngologist will likely need to intervene, performing techniques outside the scope of practice of the emergency physician, including electrocautery, submucosal lidocaine/epinephrine injections, or other surgical interventions.
- The final step in the management of this patient includes admission to a monitored bed.

53.6.1 Pearls and Pitfalls

- Pearls
  - Consult otolaryngology early in the workup as soon as a posterior bleed is identified.
  - Patients receiving packing will still need to be placed on antibiotics, such as a first-generation cephalosporin, clindamycin, or amoxicillin–clavulanic acid.
  - Perform a good physical examination concomitantly with the management of the bleed specifically geared toward assessing the patient's volume status (orthostatics, capillary refill, heart rate), because the patient may have significant volume loss depending on the duration and quantity of the bleed.
  - Take a good history including prior bleed, prior admission requirements owing to epistaxis, any anticoagulant use, blood dyscrasias, and so on.

- Pitfalls
  - Misidentifying a posterior bleed as an anterior bleed. This can be catastrophic if the patient is discharged. Reference the "Etiology" section earlier.
  - Discharging the patient.
  - Consulting otolaryngology too late.
  - Underestimating the volume of blood loss. It may be helpful to think of these patients as trauma patients in their initial management and to proceed down the standard primary survey/intervention/secondary survey/intervention approach, ensuring that the volume needs are identified early.
  - Infection.
  - Pressure necrosis of the septum.
  - Hypoxia.
  - Aspiration.
  - Arrhythmias.
  - Dysphagia.
  - Dislodgment of the pack.

Fig. 53.7 Example of posterior balloon

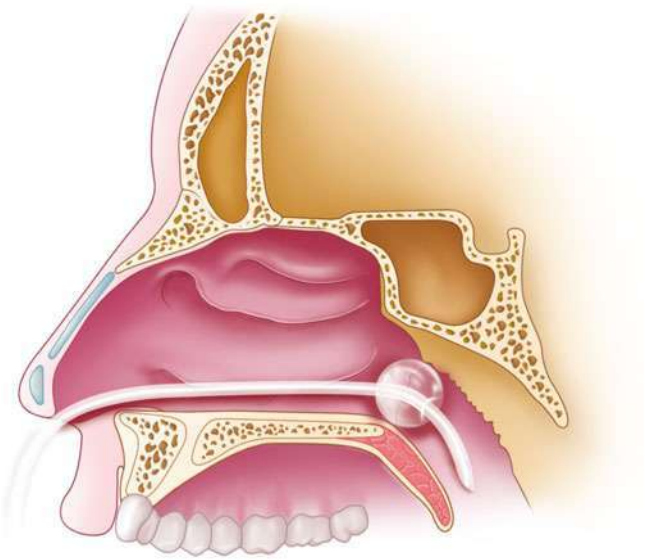

Fig. 53.8 Foley catheter in the posterior nasopharynx

Selected Reading

Buttaravoli P. Minor emergencies, splinters to fractures. 2nd ed. Philadelphia: Mosby; 2007.

Kucik CJ, Clenney T. Management of epistaxis. Am Fam Physician. 2005;71(2):305–11.

Marx JA, Hockberger RS, Walls RM, editors. Rosen's emergency medicine: concepts and clinical practice. 6th ed. Philadelphia: Mosby; 2006.

Pope LE, Hobbs CG. Epistaxis: an update on current management. Postgrad Med J. 2005;81(955):309–14.

Singer AJ, Blanda M, Cronin K, et al. Comparison of nasal tampons for the treatment of epistaxis in the emergency department: a randomized controlled trial. Ann Emerg Med. 2005;45:134–9.

Summers SM, Bey T. Chapter 239. Epistaxis, Nasal Fractures, and Rhinosinusitis. In: Tintinalli's Emergency Medicine, 7th ed. New York, NY: McGraw Hill; 2011.

Treatment of Septal Hematoma

54

Bobby K. Desai

54.1 Indications

- Septal hematomas occur after force applied to the nasal cartilage results in leakage of blood from the perichondrium (Figs. 54.1 and 54.2). They may be unilateral or bilateral.
  - Untreated, this hematoma can expand and mechanically obstruct the blood vessels that supply the nasal cartilage.

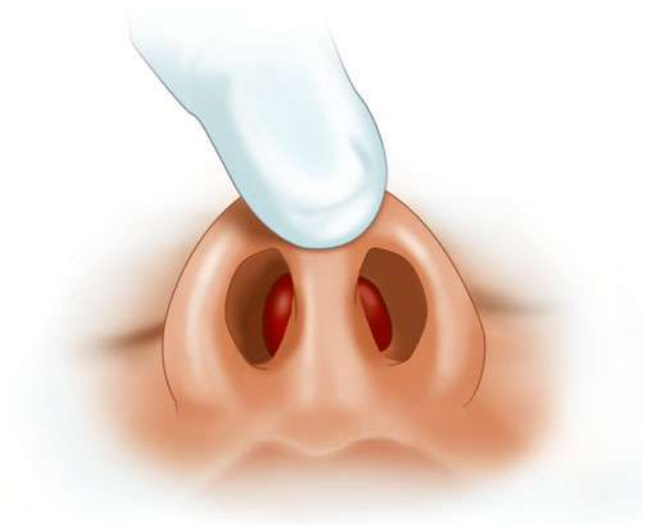

Fig. 54.1 Example of a septal hematoma

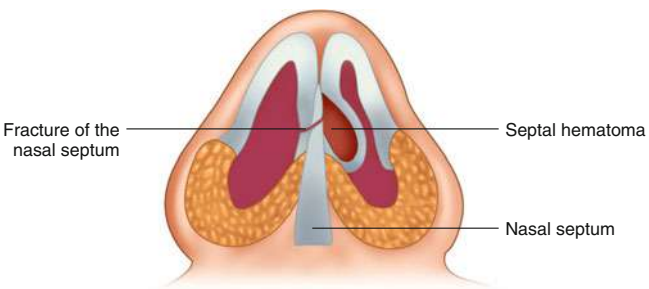

Fig. 54.2 Diagram of a septal hematoma

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_54

54.2 Contraindications

- In the setting of massive facial trauma, airway protection and consultation with otolaryngology or oral surgery may be required.

54.3 Materials and Medications

- 5-cc syringe
- 20-gauge needle (may use larger-gauge needle, but smaller ones may not allow for adequate suctioning).
- Topical antiseptic
- Sterile drapes, gown, eye protection, and gloves
- Light source—preferably a head lamp
- If using procedural sedation, specific medications of choice are needed as well as monitoring and resuscitation equipment
- Nasal speculum
- 4 % liquid cocaine or 4 % liquid lidocaine for anesthesia
- Topical vasoconstrictor of choice (oxymetazoline or phenylephrine)
- Scalpel (#15 blade)
- Suction setup
- Saline for irrigation
- Sterile gauze
- Iodoform packing or sterile rubber band
- Anterior nasal pack (Rhino Rocket)
- Intravenous antibiotics to cover *Staphylococcus aureus*, group A β-hemolytic streptococci, *Streptococcus pneumonia,* and *Haemophilus influenzae*

54.4 Procedure

- Preprocedure
  - Explain the procedure to the patient.
  - Informed consent may be required.
- Procedure—unilateral (Fig. 54.3):

1. Place the patient in a seated position if the procedure is to be done without procedural sedation.
   - If performing the procedure under procedural sedation, the patient may be placed in the supine position with appropriate monitoring and resuscitation equipment on standby.
2. Provide local anesthesia by applying the topical anesthetic of choice to cotton pledgets, wringing out the excess before insertion.
   - Remove after 10 min.
   - If using lidocaine, a topical vasoconstrictor may be required that may be mixed with the lidocaine or applied directly on the nasal mucosa before the procedure.
3. Using a nasal speculum, open the affected nares vertically as wide as possible. Horizontal use may obscure the hematoma.
4. Using a #15 blade, make a vertical incision at the hematoma extending the incision posteriorly at its base.
   - Take care not to make too deep an incision because septal perforation may occur.
5. Suction and gauze should be at the ready to remove as much blood as possible.
6. Pack the incision site with iodoform gauze or a sterile rubber band in order to keep the incision open to prevent accumulation of blood.
7. This packing can be held in place with an anterior nasal pack.
   - The pack should be placed in both nares to prevent deformity of the nasal septum.
8. Consider administration of a dose of intravenous antibiotics.
9. Send the patient home with a prescription for oral antibiotics.
10. Arrange follow-up for 24 h.

- Procedure—bilateral (Fig. 54.4):

1. If bilateral hematomas are present, otolaryngology consultation should be sought.
2. One method involves the incision of one side only through the nasal septum so that both may drain through one side.
   - Care must be taken NOT to penetrate the overlying mucosa of the other side.
3. The second option is to repeat the unilateral drainage procedure on the contralateral side.
4. Care must be taken to not have the incision exactly in the same location as the contralateral side owing to the risk of permanent septal perforation.

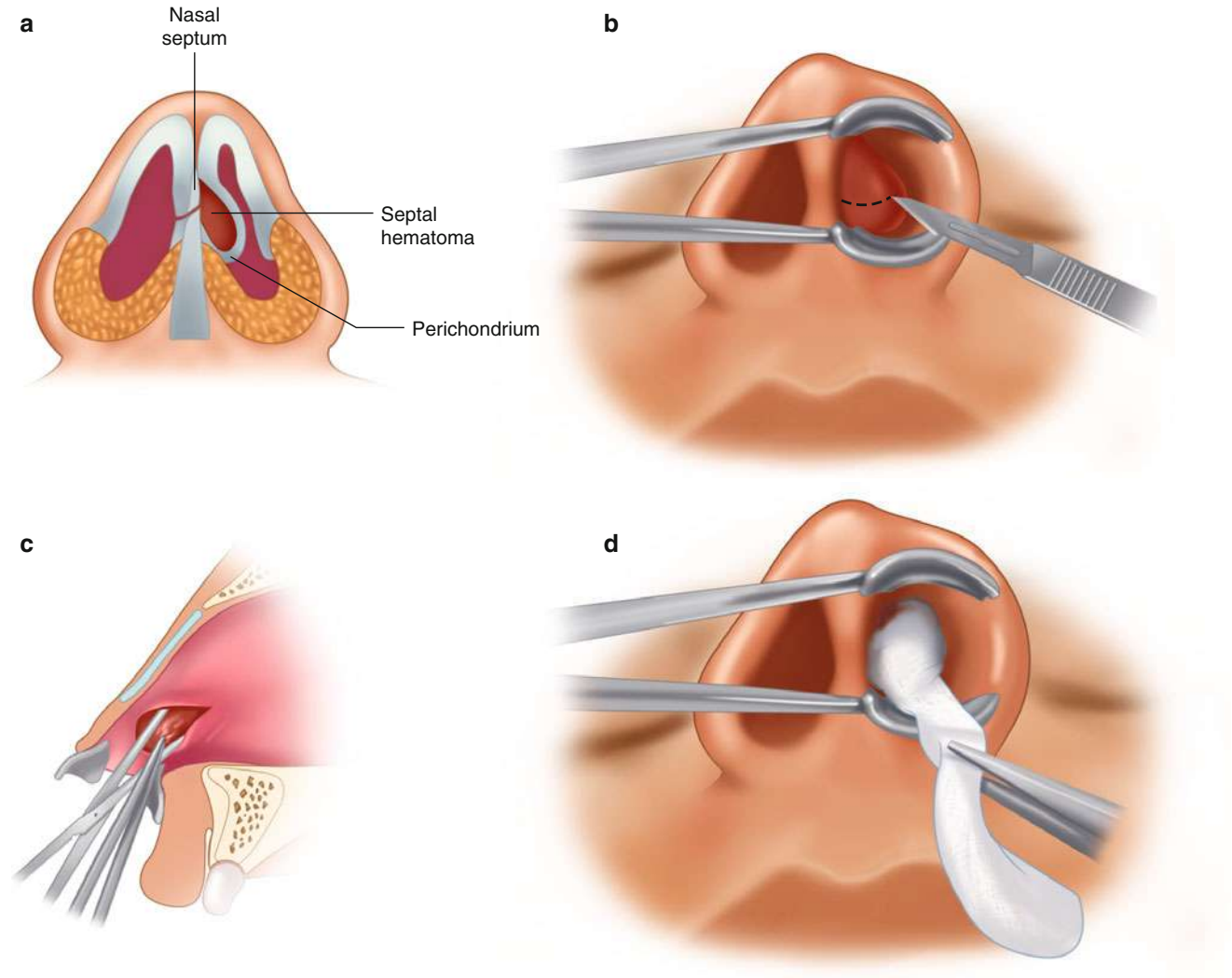

Fig. 54.3 (**a**–**d**) Technique for drainage of a unilateral septal hematoma

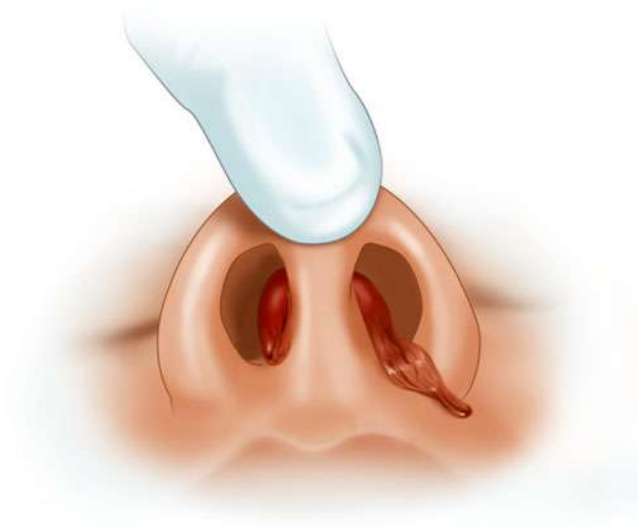

Fig. 54.4 Technique for drainage of a bilateral septal hematoma

54.5 Pitfalls and Complications

- Bleeding
- Infection: toxic shock syndrome
- Deformity of the nose
- Permanent septal perforation
- Inadequate drainage
- Reaccumulation of blood
- Development of septal abscess

Selected Reading

Ginsburg CG. Consultation with the specialist: nasal septal hematoma. Pediatr Rev. 1998;19:142–3.

Savage RR, Valvich C. Hematoma of the nasal septum. Pediatr Rev. 2006;27:478–9.

55 Nasal Foreign Body Removal

Bobby K. Desai

55.1 Indications

- Visualized foreign body (FB) on inspection
- Continued unilateral nasal discharge of unknown etiology
- Recurrent epistaxis

55.2 Contraindications: Need Urgent Referral

- Those cases in which the emergency physician is not confident of success.
- Combative patients who are not candidates for conscious sedation.
- Embedded FBs.
- Penetrating injuries with an FB.
- Button batteries that cannot be removed.
- Patients with bleeding diathesis.
- Respiratory distress.
- FBs superior and medial to the middle turbinate owing to risk of puncture to the cribriform plate.
- Chronic FBs may be difficult to visualize and these should be referred.

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_55

55.3 Materials and Medications

- 1 % lidocaine without epinephrine (solution) for anesthesia
- 2 % lidocaine jelly
- 0.5 % phenylephrine to vasoconstrict vessels
- Nasal speculum (Fig. 55.1)
- Headlamp or other direct lighting instrument
- Ambu bag (Fig. 55.2)
- Alligator forceps (Fig. 55.3)
- Curved hook (Fig. 55.4)
- Foley catheter
- Suction: schuknecht catheter (Fig. 55.5)
- Eye and biohazard protection for the clinician

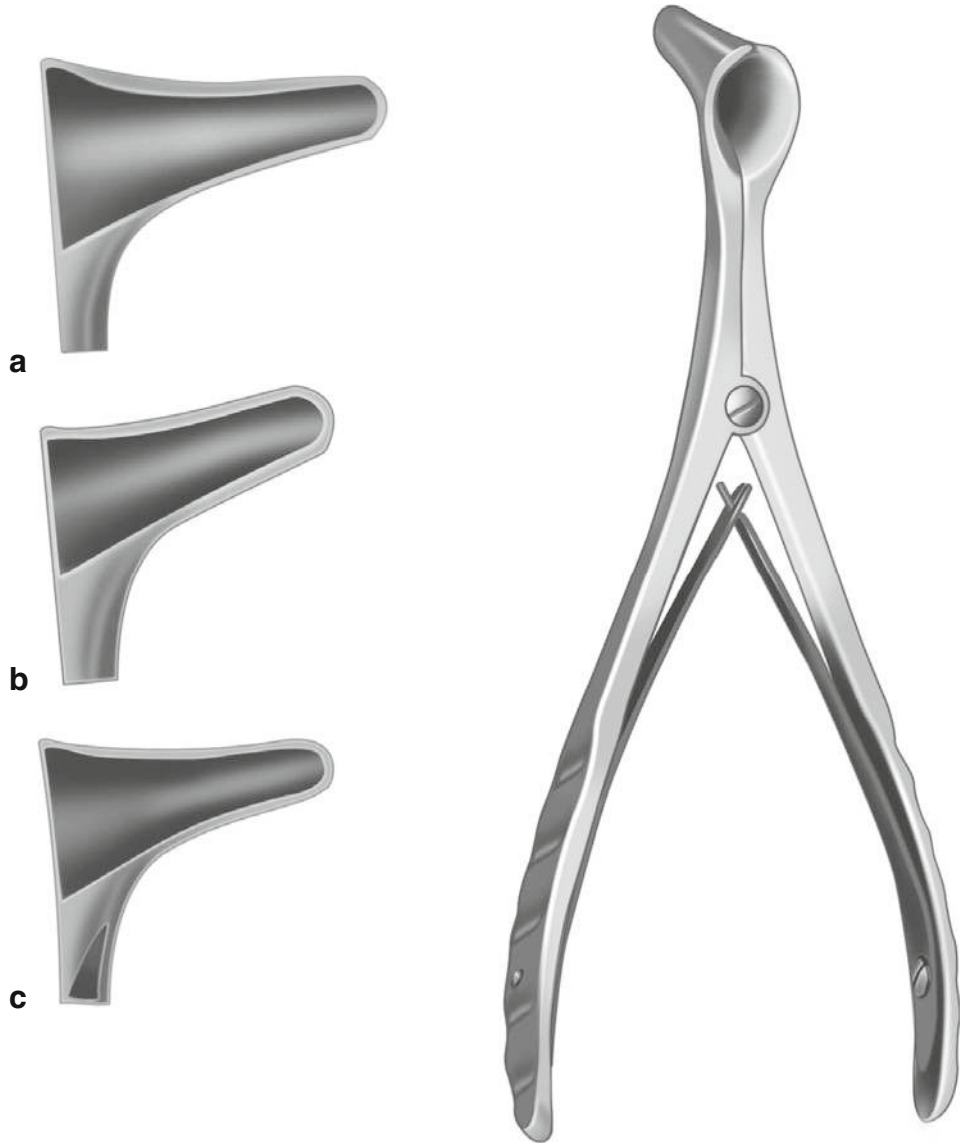

Fig. 55.1 (**a**–**c**) Example of nasal speculum

Fig. 55.2 Example of Ambu bags

Fig. 55.4 Example of a curved hook

Fig. 55.5 Example of a schuknecht suction tube

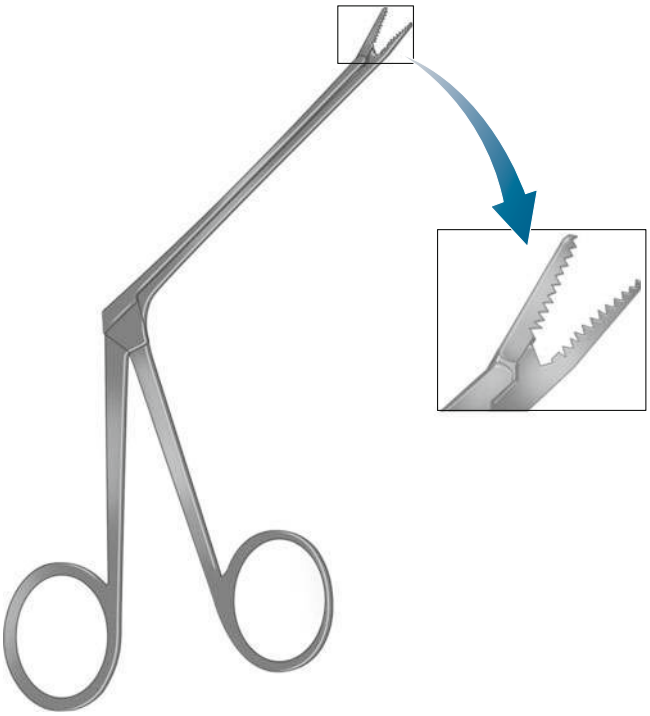

Fig. 55.3 Example of alligator forceps

55.4 Procedures (Based on the Method Chosen)

55.4.1 Alligator Forceps

- Perhaps the easiest method to use for easily visualized FBs.
- Care must be taken for those FBs that are easily broken apart (e.g., peas) to ensure that all of the matter has been removed.
- Using a nasal speculum, spread the nares as much as possible to maximize visualization.
- Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
- Attempt to grasp the FB with the forceps.
- After removal, ensure the complete removal of the object and ensure that no other FBs are present.

55.4.1.1 Pitfalls and Complications

- Pushing the object deeper into the nares.
- Inability to totally remove the FB.
- Attempted removal of an embedded FB may cause significant bleeding.

55.4.2 Curved Hook

- Used for nongraspable objects especially in the anterior nares.
- Using a nasal speculum, spread the nares as much as possible to maximize visualization.
- Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
- Pass the hook behind the object, or if the object has holes, the hook may be placed inside the hole to facilitate removal.
- Slowly withdraw the hook, thereby removing the FB.
- After removal, ensure the complete removal of the object and ensure that no other FBs are present.

55.4.2.1 Pitfalls and Complications

- This technique cannot be used if the object cannot be directly visualized because significant trauma may arise owing to the misplacement of the hook.
- Attempted removal of an embedded FB may cause significant bleeding.

55.4.3 Foley Catheter Removal

- Use a 5 or 6 French catheter.
- Anesthetize the nares with lidocaine with or without epinephrine and vasoconstrict the nasal vessels with phenylephrine.
- Ensure that the Foley balloon has no leaks.
- Place the patient in a supine position.
  - Consider procedural sedation for children and/or a papoose board.
- Lubricate the Foley tip and balloon with lidocaine jelly.
- Advance the tip past the object.
- Blow up the balloon with 2–3 mL of air and gently withdraw the catheter.
  - The amount of air may need to be adjusted depending on the size of the object to be withdrawn.
  - Pulling too quickly may cause a soft FB to break apart.

55.4.3.1 Pitfalls and Complications

- This technique may be used if the FB is not visible.
- In order for this technique to work, the catheter must be able to slide posterior to the object in order for the balloon to pull out the object.

55.4.4 Suction

- A small metal suction catheter—the schuknecht catheter—may be used.
- The suction is placed against the FB and slowly withdrawn.
- This technique works best for round smooth objects visualized in the anterior nares.

55.4.4.1 Pitfalls and Complications

- May not work for posteriorly displaced FBs.
- May not work for objects too tightly entrenched in the nares.

55.4.5 Nasal Positive Pressure

- This technique works best for round FBs occluding the nares.
- Place the patient in a supine position.
  - Consider a papoose board to secure the patient.
- Apply pressure to the contralateral nares to occlude it.
- Connect an Ambu bag to high-flow oxygen.
  - Use a facemask that covers only the mouth.
- Place the mask on the patient's mouth and allow it to expand by covering the thumbhole.
- If this pressure is insufficient to expel the object, the Ambu bag may be compressed to add pressure to the attempt.
- This may allow the object to be seen and grasped in the anterior nares.

55.4.5.1 Pitfalls and Complications

- Too much pressure could theoretically rupture the tympanic membranes.

Selected Reading

Backlin SA. Positive-pressure technique for nasal foreign body removal in children. Ann Emerg Med. 1995;25:554–5.

Chan TC, Ufberg J, Harrigan RA, et al. Nasal foreign body removal. J Emerg Med. 2004;26:441–5.

Kadish H. Ear and nose foreign bodies: it is all about the tools. Clin Pediatr. 2005;44:665–70.

56 Cerumen Removal

Bobby K. Desai

56.1 Indications

- Hearing loss
- Profound dizziness
- Pain

56.2 Contraindications

- Uncooperative patient
- Potential for foreign bodies
- Prior surgery to ear or mastoid
- History of middle or outer ear disease

56.3 Materials and Medications

- Ear speculum (Fig. 56.1).
- Eye and biohazard protection for the clinician
- Headlamp or other direct lighting instrument
- Cerumen-softening solutions
  - Carbamide peroxide
  - Mineral oil or other oil-based agents including almond or olive oil
  - Liquid docusate sodium
  - Acetic acid
  - Sodium bicarbonate
- Alligator forceps (Fig. 56.2)
- Curved hook or cerumen spoon (Figs. 56.3 and 56.4)
- Suction—Schuknecht catheter
- Irrigation setup
  - Ear syringe (Figs. 56.5 and 56.6)
  - 20- to 60-mL syringe with an 18-gauge angiocath attached
  - Commercial ear irrigation setup
  - Water (tap water may be used)
  - Basin or other collection device
  - Towels

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_56

Fig. 56.1 Example of an ear speculum

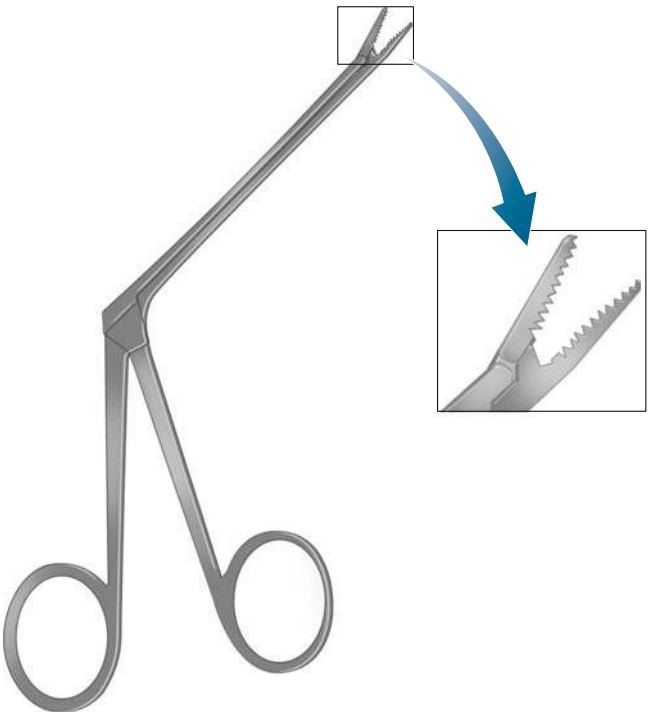

Fig. 56.2 Example of alligator forceps

Fig. 56.3 Examples of cerumen spoons

Fig. 56.4 Examples of ear curettes

Fig. 56.5 Example of a typical bulb syringe used in irrigation

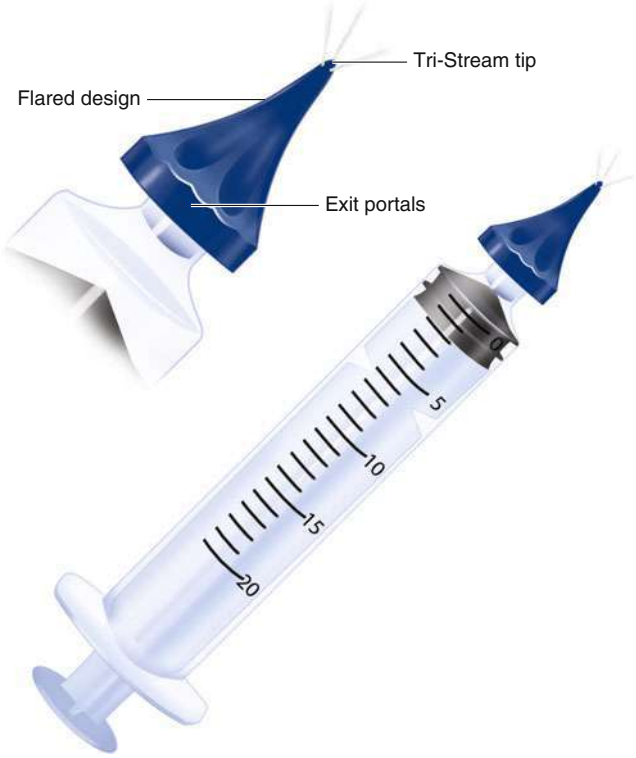

Fig. 56.6 Example of a specialized ear syringe

56.4 Procedure

1. Explain the procedure to the patient to facilitate cooperation.
   - Some discomfort may be involved, including a sensation of pressure and the cold, wet feeling from the water used.
   - Nausea, vomiting, or a sensation of vertigo may also be experienced.
     - Consider pretreatment of vertigo with agent of choice.
2. The patient may be seated or supine with the head slightly turned to the affected side to facilitate the collection of fluid.
3. Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
4. Pull the ear up and out to straighten the ear canal.
5. Before irrigation, if the cerumen is hard, alligator forceps or a cerumen spoon may be used under direct visualization to remove as much cerumen as possible.
6. Use a cerumen-softening agent before irrigation by instilling and leaving in for 30 min to maximize softening.
7. Place a collection device next to the patient's ear, or alternatively, have the patient hold a basin near the ear to collect any liquid.
8. Attach an angiocath to a 20- or 60-mL syringe, place the tip of the catheter within the ear canal, and aim toward the tympanic membrane.
9. Slowly inject the solution into the ear canal.
   - Care must be taken to limit the force used owing to the potential for tympanic membrane rupture.
   - The fluid can be injected in short bursts or all at once.
10. Direct the stream as much as possible toward the rear of the cerumen to be removed.
11. Continually reassess the ear canal to ensure that all cerumen has been removed.
    - If one large piece remains, use a cerumen hook to facilitate removal.
12. If the patient complains of severe sharp pain, stop the procedure and assess the tympanic membrane for rupture.
    - Consider referral or otolaryngological consultation at this point.
13. After successful removal, dry the ear canal with cotton-tipped applicators to prevent otitis externa.

56.5 Pitfalls and Complications

- Owing to instillation of room temperature water, a caloric response may occur including vertigo, nausea, and vomiting.
- Rupture of the tympanic membrane.
- Unsuccessful irrigation secondary to inadequate time given to softening agents before attempting removal.
- Bleeding of the external auditory canal.
- Otitis externa.
- Hearing loss.

Selected Reading

Blake P, Matthews R, Hornibrook J. When not to syringe an ear. N Z Med J. 1998;111:422.

McCarter D, Courtney AU, Pollart SM. Cerumen impaction. Am Fam Physician. 2007;75:1523–8.

Roland PS, Smith TL, Schwartz SR, et al. Clinical practice guideline: cerumen impaction. Otolaryngol Head Neck Surg. 2008;139:S1–21.

57 Ear Foreign Body Removal

Bobby K. Desai

57.1 Indications

- Visualized foreign body (FB) on inspection.
- Continued unilateral otorrhea of unknown etiology mandates a search.
- Recurrent bleeding.
- History of FB placement by patient.

57.2 Contraindications: Need Urgent Referral

- Those cases in which the emergency physician is not confident of success.
- Combative patients are not candidates for conscious sedation.
- Embedded FBs.
- Penetrating injuries with an FB.
- FBs close to the tympanic membrane.
- Consider referral of spherical FBs because these may be difficult to grasp.
- Chronic FBs may be difficult to visualize; these should be referred.
- Consider referral for those patients who have attempts made at other institutions.

57.3 Materials and Medications

- Ear speculum (Fig. 56.1)
- Eye and biohazard protection for the clinician
- Headlamp or other direct lighting instrument
- Alligator forceps (Fig. 56.2)
- Curved hook (Fig. 57.1) or cerumen spoon (Fig. 56.3)
- Suction—Schuknecht catheter (Fig. 57.2)
- Irrigation setup
  - Ear syringe (Fig. 57.3)
  - 20- to 60-mL syringe with an 18-gauge angiocath attached
  - Water (tap water may be used)
  - Basin or other collection device
  - Towels
- Cyanoacrylate (Fig. 57.4)

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_57

Fig. 57.1 Example of a curved hook

Fig. 57.2 (**a**) Example of a Schuknecht suction tube, (**b**) example of ear irrigation setup using a Waterpik system

Fig. 57.3 Example of a typical syringe used in irrigation

Fig. 57.4 Example of cyanoacrylate

57.4 Procedures (Based on the Method Chosen)

57.4.1 Alligator Forceps

- Ideal for easily visualized FBs, especially graspable FBs in the lateral third of the ear canal.
  - May have limited success with rounded objects.
- Care must be taken for those FBs that are easily broken apart (e.g., peas) to ensure that all of the matter has been removed.
  - May need irrigation subsequent to the attempt.
- Explain the procedure to the patient to facilitate cooperation.
- Pull the ear up and out to straighten the ear canal.
- Consider using a specialized otoscope with magnification (an operating otoscope) to maximize visualization.
- The patient may be seated or supine with the unaffected side turned to the gurney or examining chair.
- Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
- Attempt to grasp the FB with the forceps.
  - If using an ear speculum, attempt to bring the object as close as possible to the speculum in order to stabilize it and slowly remove both.
- After removal, ensure the complete removal of the object and ensure that no other FBs are present.

57.4.1.1 Pitfalls and Complications

- Pushing the object deeper into the auditory canal.
- Inability to totally remove the FB.
- Attempted removal of an embedded FB may cause significant bleeding.

57.4.2 Curved Hook

- Ideal for spherical objects in the lateral third of the auditory canal.
- Can be used for non-graspable objects especially in the lateral third of the auditory canal.
- Direct visualization is necessary to avoid trauma to the ear canal.
- Consider using a specialized otoscope with magnification (an operating otoscope) to maximize visualization.
- Explain the procedure to the patient to facilitate cooperation.
- The patient may be seated or supine with the unaffected side turned to the gurney or examining chair.
- Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
- Pull the ear up and out to straighten the ear canal.
- Pass the hook behind the object, or if the object has holes, place the hook inside the hole to facilitate removal.
  - In order for this technique to succeed, there must be sufficient area near the object to facilitate the hook going behind it.
- Slowly withdraw the hook, thereby removing the FB.
- After removal, ensure the complete removal of the object and ensure that no other FBs are present.

57.4.2.1 Pitfalls and Complications

- Pushing the object deeper into the auditory canal.
- This technique cannot be used if the object cannot be directly visualized because significant trauma may arise owing to the misplacement of the hook.
- Attempted removal of an embedded FB may cause significant bleeding.
- Do not use this technique if the object is close to the tympanic membrane owing to risk of rupture.

57.4.3 Suction

- A small metal suction catheter—the Schuknecht catheter—may be used.
- Ideal for hard rounded objects in the lateral third of the ear canal.
  - Does not work as well for those objects closer to the tympanic membrane owing to lack of suction and the narrow ear canal
- Direct visualization of the object is mandatory.
  - The catheter must be placed directly on the object for this technique to succeed.
- Explain the procedure to the patient to facilitate cooperation.
- The patient may be seated or supine with the unaffected side turned to the gurney or examining chair.
- Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
- Pull the ear up and out to straighten the ear canal.
- Place the suction against the FB and slowly withdraw.

57.4.3.1 Pitfalls and Complications

- Pushing the object deeper into the auditory canal
- May not work for objects too tightly entrenched in the ear canal

57.4.4 Irrigation

- Useful for small particulate matter unable to be grasped by forceps or too small for a hook to pull out.
  - Cannot be used for button batteries or vegetable matter.
    - Irrigation may cause an alkaline necrosis with the button battery.
    - Vegetable matter may expand, increasing the difficulty of removal.
- Explain the procedure to the patient to facilitate cooperation.
  - Some discomfort may be involved, including a sensation of pressure and a feeling cold wetness from the water used.
- The patient may be seated or supine with the head slightly turned to the affected side to facilitate collection of fluid.
- Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
- Pull the ear up and out to straighten the ear canal.
- Place a collection device next to the patient's ear, or alternatively, have the patient hold a basin.
- Attach an angiocath to a 20- or 60-mL syringe and place the tip of the catheter within the ear canal, or use a Waterpik system.
- Slowly inject the solution into the ear canal.
  - Care must be taken to limit the force used owing to the potential for tympanic membrane rupture.
- Direct the stream as much as possible toward the rear of the object(s) to be removed.
- Continually assess the ear canal to ensure that all foreign matter has been removed.
- If the patient complains of severe sharp pain, stop the procedure and assess the tympanic membrane for rupture.
  - Consider referral or otolaryngological consultation at this point.

57.4.4.1 Pitfalls and Complications

- Owing to instillation of room temperature water, a caloric response may occur including vertigo, nausea, and vomiting.
- Rupture of the tympanic membrane.

57.4.5 Cyanoacrylate

- Useful for objects that can be directly visualized.
- Ideal for hard rounded objects in the lateral third of the ear canal.
  - Does not work as well for those objects closer to the tympanic membrane owing to lack of visualization
- Direct visualization of the object is mandatory.
  - The stick with the cyanoacrylate must be placed directly on the object for this technique to succeed.
- Consider using a specialized otoscope with magnification (an operating otoscope) to maximize visualization.
- Explain the procedure to the patient to facilitate cooperation.
- The patient may be seated or supine with the unaffected side turned to the gurney or examining chair.
- Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
- Pull the ear up and out to straighten the ear canal.
- Place a small amount of cyanoacrylate on the end of a cerumen spoon or wooden end of a cotton-tipped swab.
- Using direct visualization, place the stick in the ear canal and touch the FB.
- Leave the stick on the object for approximately 30–60 s to facilitate drying of the glue.
- Remove the stick and the FB.

57.4.5.1 Pitfalls and Complications

- This technique cannot be used if the object cannot be seen.
- The object must be relatively smooth to facilitate placement of the glue.

Selected Reading

Hanson RM, Stephens M. Cyanoacrylate-assisted foreign body removal from the ear and nose in children. J Paediatr Child Health. 1994;30:77–8.

Kadish H. Ear and nose foreign bodies: it is all about the tools. Clin Pediatr. 2005;44:665–70.

58 Treatment of Auricular Hematoma

Bobby K. Desai

58.1 Indications

- Presence of auricular hematoma
  - These occur after significant force is applied to the ear.
  - This type of hematoma will separate the perichondrium from the cartilage.
    - Untreated, this hematoma prevents the development of new cartilage, subsequently deforming the auricle, causing *cauliflower ear* (Figs. 58.1 and 58.2).

Fig. 58.1 Example of a cauliflower ear

B.K. Desai, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_58

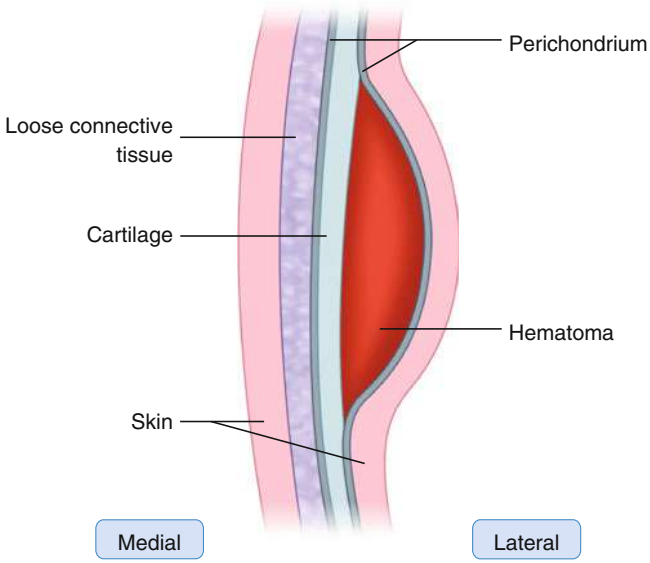

Fig. 58.2 Schematic of auricular hematoma

58.2 Contraindications

- Uncooperative patient.
- Evidence of infection mandates plastic surgery or otolaryngology consultation.
- Significant trauma to area around the ear.

58.3 Materials and Medications

- 5-cc syringe
- 20-gauge needle (may use larger-gauge needle, but smaller ones may not allow for adequate suctioning)
- Topical antiseptic
- Sterile drapes, gown, eye protection, and gloves
- 1 % lidocaine for anesthesia
- Scalpel (#11 or #15 blade)
- Suction setup
- Saline for irrigation
- Hemostat and forceps
- Penrose drain or plain gauze packing
- Xeroform gauze
- Laceration tray setup
  - Needle driver
  - 4–0 Nonabsorbable suture
- Antibiotic ointment
- Gauze bandages

58.4 Procedure (Needle Aspiration)

1. Explain the procedure to the patient.
2. Informed consent may be required.
3. The patient may be seated or supine with the unaffected side turned to the gurney or examining chair.
4. Stabilize the head with the nondominant hand, or for children, consider a papoose board to secure the patient.
   - Consider procedural sedation for children.
5. Carefully clean the area with an antiseptic solution of choice (povidone-iodine [Betadine] or chlorhexidine solution) and allow to dry.
6. Anesthetize the area where the hematoma is of greatest diameter with 1 % lidocaine.
7. Inject the hematoma itself to promote increased anesthesia.
8. The clinician should prepare for the procedure by practicing universal precautions.
9. Drape the patient's ear.
10. Attach a 20-gauge needle to a syringe.
    - A 5-cc syringe should be sufficient.
11. Attempt to aspirate the hematoma contents with the prepared syringe (Fig. 58.3).

12. "Milking" the hematoma may be required to fully remove the clot.
    - If the hematoma has been present for sufficient time for the blood to completely clot, needle aspiration will fail, and an alternative technique, described later, may be necessary to completely evacuate the hematoma.
13. Place antibiotic ointment around the injection site.
14. Loosely dress the area.
15. Refer to an otolaryngologist or a plastic surgeon for follow-up care.

If needle aspiration fails, a larger incision may be required using the following procedure:

1. Perform the previously discussed steps.
2. Consider a local field block of the greater auricular nerve.
3. Make an incision over the hematoma, following the anatomy of the helix (Fig. 58.4).
    - The incision should be 5–6 mm in length and curvilinear, following the concavity of the helix.
    - More than one incision may be required.
4. Remove the hematoma with hemostats or forceps.
    - Suctioning may be used as well.
5. Copiously irrigate the cavity with saline to make certain of complete removal of the hematoma.
6. Insert the Penrose drain or plain packing into the incision.
7. Antibiotic ointment may be applied to the incision site.
8. Place the Xeroform gauze over the incision site.
9. Apply a compression dressing to the ear (Fig. 58.5).
    - Place dry cotton into the ear canal to protect from possible drainage and subsequent infection (Fig. 58.6).
    - Mold Xeroform gauze in and around the contours of the helix (Fig. 58.7).
    - With the aid of an assistant, place gauze behind the pinna and over the Xeroform gauze already molded to the anterior portion of the ear (Fig. 58.8).
    - Place more fluffed gauze over the ear (Fig. 58.9).
    - While the assistant holds the anterior and posterior gauze in place, use a web roll (Kerlix) around the head to keep the gauze in place. Then wrap the head with an ACE wrap or stretchy bandage (Figs. 58.10 and 58.11).
    - These steps are critical to maintain equal compression of the ear, which will help prevent future deformity.
    - Alternatively, dental rolls can be secured with suture to the anterior and posterior site of the drained hematoma (Fig. 58.12).
10. Place the patient on prophylactic antibiotics.
11. Refer to an otolaryngologist or a plastic surgeon for follow-up care.

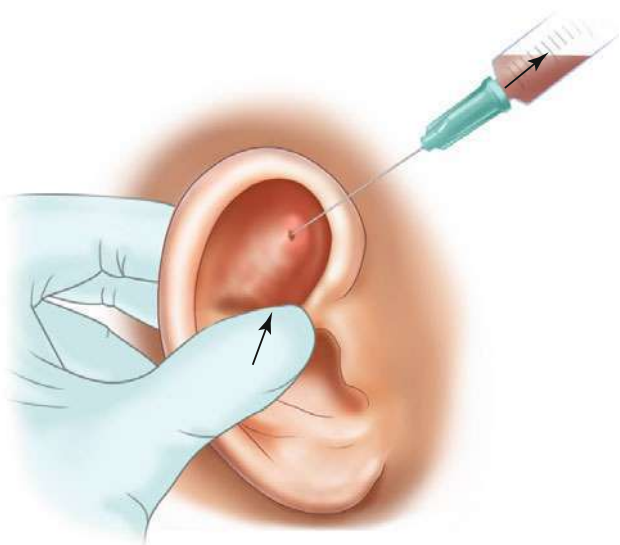

Fig. 58.3 Aspiration of a hematoma

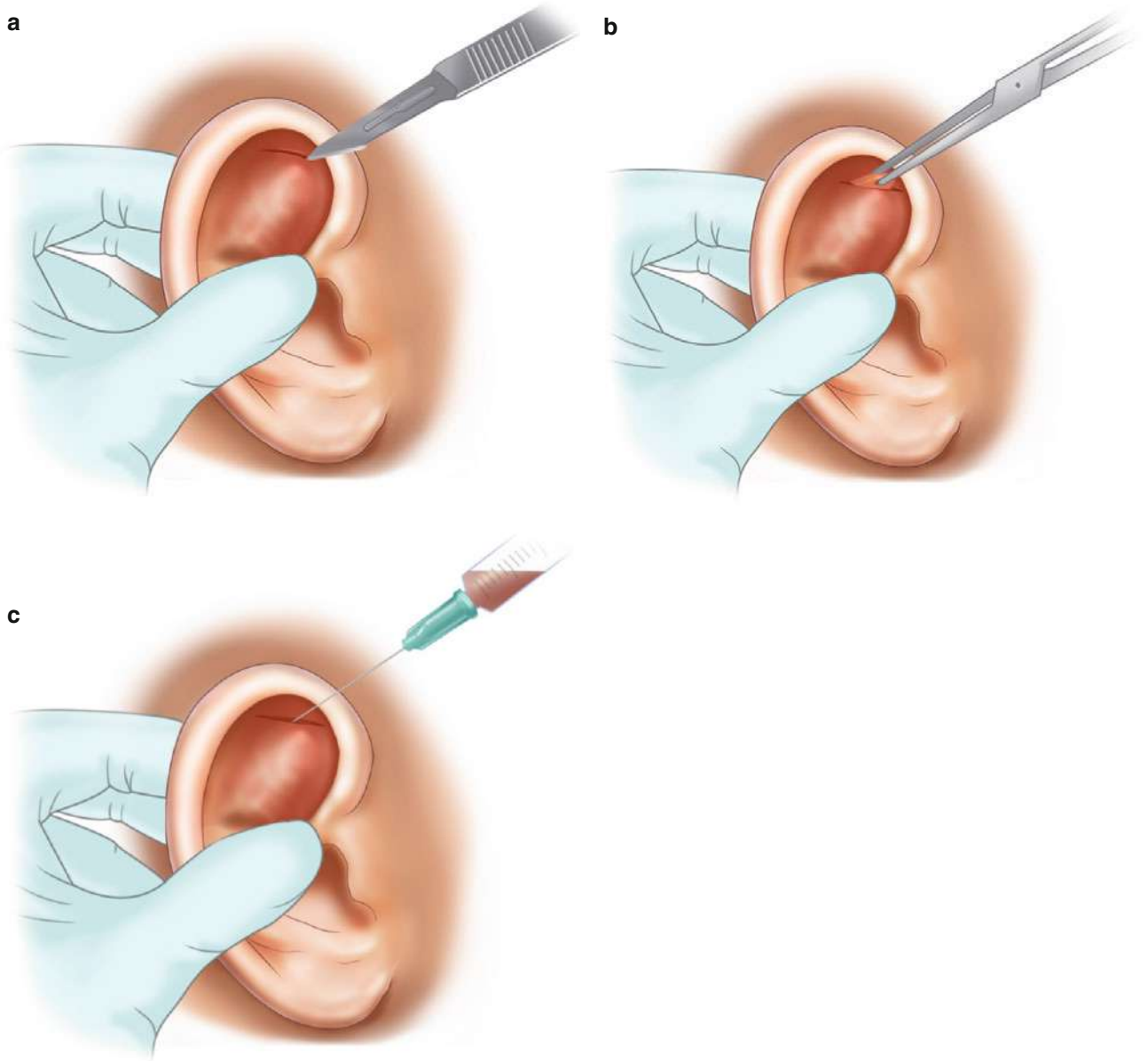

Fig. 58.4 Incision of an auricular hematoma. (**a**) Incision; (**b**) removal of clot with forceps; (**c**) aspiration of blood with syringe

Fig. 58.5 Materials needed for compression dressing

Fig. 58.6 Place cotton in the ear to prevent drainage into ear

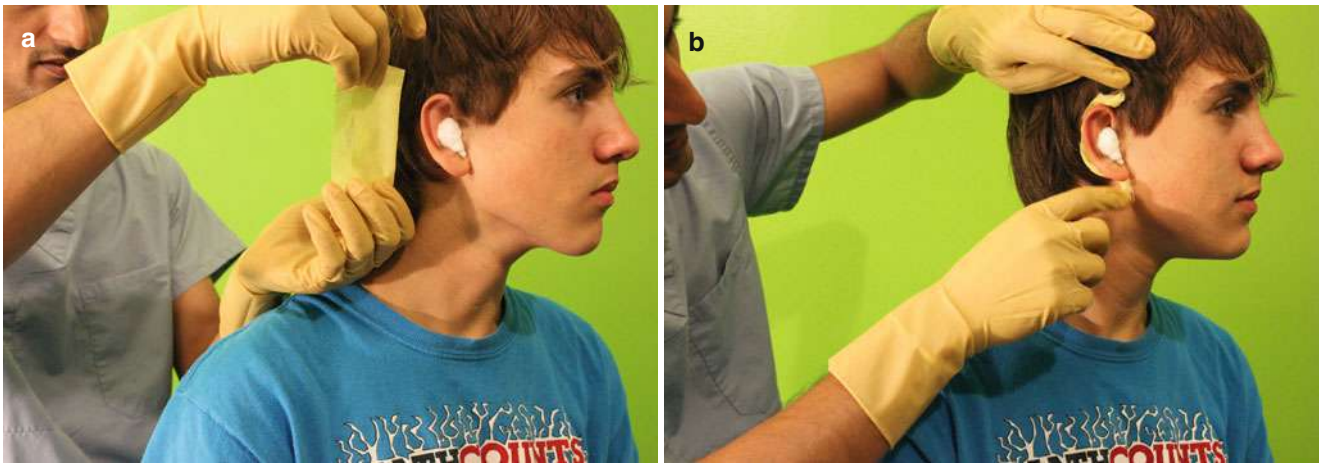

Fig. 58.7 (**a**, **b**) Mold Xeroform gauze around ear helix

Fig. 58.8 (**a**, **b**) Mold sterile gauze on top of Xeroform gauze around ear helix

Fig. 58.9 Pad outside of ear with 4×4 gauze pads

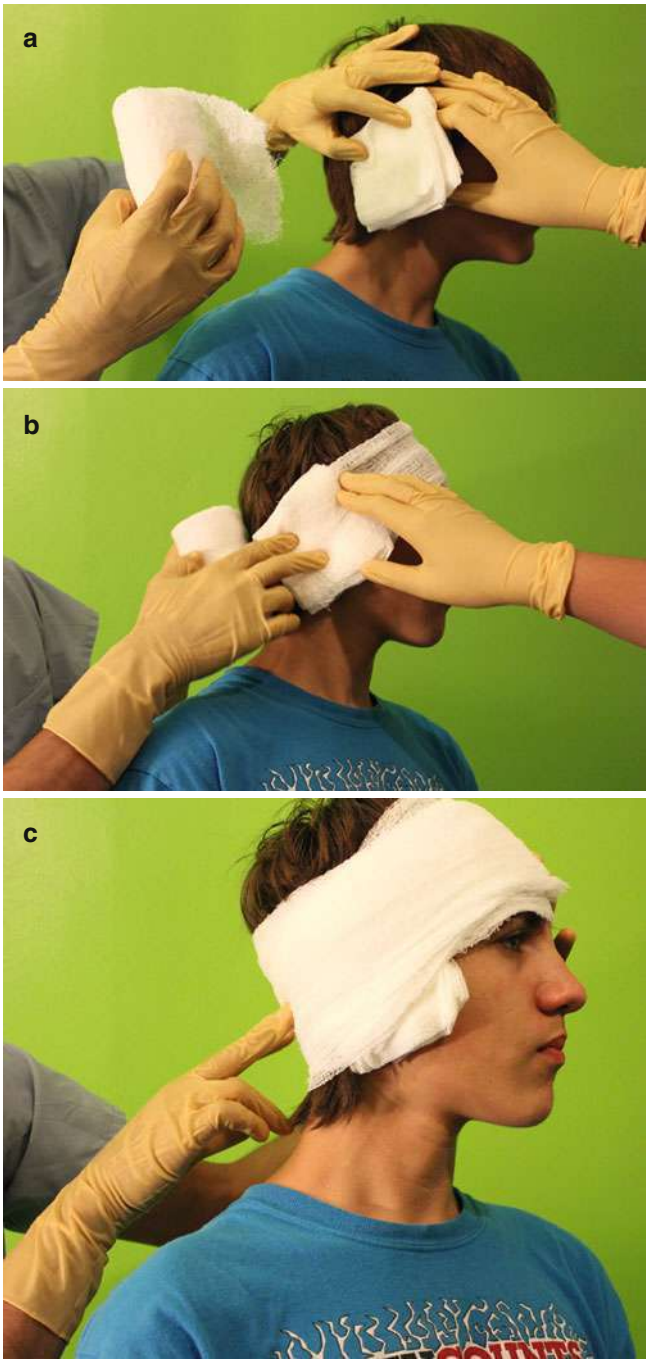

Fig. 58.10 (**a**–**c**) Wrap head with Kerlix (sterile gauze roll)

Fig. 58.11 (**a**–**d**) Wrap head with elastic bandage

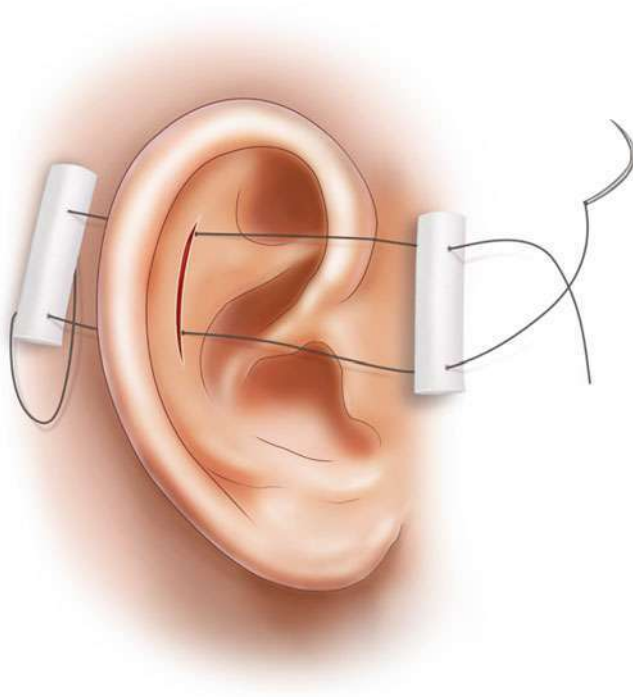

Fig. 58.12 Dental rolls secured to the ear

58.5 Pitfalls and Complications

- Bleeding
- Infection—perichondritis
- Deformity of the auricle
- Inadequate clot removal
- Hematomas present for several days warrant a plastic surgery or an otolaryngology consultation

Acknowledgement The authors would like to thank Thor Shiva Stead for serving as the subject in many of the photographs in this chapter.

Selected Reading

O'Donnell BP, Eliezri YD. The surgical treatment of traumatic hematoma of the auricle. Dermatol Surg. 1999;25:803–5.

Starck WJ, Kaltman AI. Current concepts in the surgical management of traumatic auricular hematoma. J Oral Maxillofac Surg. 1992;50:800–2.

Incision and Drainage of Peritonsillar Abscess

59

Melinda W. Fernandez and Bobby K. Desai

59.1 Indications

- Peritonsillar abscess

59.2 Contraindications

- Absolute
  - Malignancy
  - Vascular malformations
- Relative
  - Pediatric patient
  - Severe trismus
  - Uncooperative patient

59.3 Materials and Medications

- #11 or #15 scalpel
- 27-gauge 1.5-inch needle, 18- to 20-gauge 1.5-inch or longer needle
- 5-mL syringe, 10- to 20-mL syringe
- Tape
- Trauma shears
- Viscous lidocaine
- Lidocaine 1 % with epinephrine
- Laryngoscope with MAC 3 or 4 *or* tongue blade(s) and headlamp or other light source
- Suction setup with Frazier or tonsil suction tip

M.W. Fernandez, MD
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: mindyfernandez@ufl.edu

B.K. Desai, MD (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

59.4 Procedure: Aspiration

1. Informed consent may be required.
2. Raise head of bed to at least 60° and place a pillow or other support behind the patient's head.
3. Prepare an 18- to 20-gauge 1.5-inch needle on a 10- to 20-mL syringe with a needle guard.
   - This can be accomplished by using trauma shears to cut 1–1.5 cm off the distal plastic needle cover and replacing the cover over the needle.
   - This now creates a guard to prevent deep penetration into vascular structures (Fig. 59.1).
4. Topically anesthetize area with viscous lidocaine (alternatively, can use Cetacaine [benzocaine, tetracaine hydrochloride, and butamben] spray)
5. Using lidocaine 1 % with epinephrine and a 27-gauge 1.5-inch needle on a 5-mL syringe, infiltrate 1–2 mL into the area. Blanching should be apparent.
6. Palpate the oropharynx with the gloved finger to evaluate for fluctuance.
7. Assemble a MAC 3 or 4 intubation blade on a laryngoscope handle and open into the light-on position. Insert the blade into the patient's mouth, and advance the blade as far posteriorly as possible without inducing gagging. Have the patient hold the laryngoscope handle. Alternatively, use a tongue blade and head lamp or other light source, but the laryngoscope technique allows for an unobstructed view of the area and the weight of the handle helps hold the patient's mouth open.
8. Using the previously prepared needle guard on the long 18- to 20-gauge needle on a 12-mL syringe, insert the needle into the most fluctuant area (as determined from the previous examination), aspirating as the needle advances (Fig. 59.2). The most fluctuant area will usually be located in the superior pole of the tonsil.

- It is very important to be careful not to angle the needle laterally toward the carotid artery. Also remember, this is a peritonsillar abscess so do not aspirate the tonsil itself.

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_59

9. After aspiration, if there is a pus return, remove as much purulent drainage as possible. Typical drainage is between 2 and 6 mL. If a large amount of pus is obtained (>6 mL), incision and drainage (I&D) may be indicated. (See later for details on I&D.)
10. If the aspirate did not return pus, attempt aspiration again by moving the insertion site 1 cm inferiorly to the middle pole of the peritonsillar space. If there is still no pus aspirated, make a final attempt by moving again 1 cm inferiorly to the lower pole of the peritonsillar space.
11. Suction should be set up, readily available, and turned on. Use suction to prevent the patient from aspirating or swallowing any purulent drainage.
12. Expect a small amount of bleeding when the procedure is complete and the needle is removed.

Fig. 59.1 Guard to prevent deep penetration of vascular structures

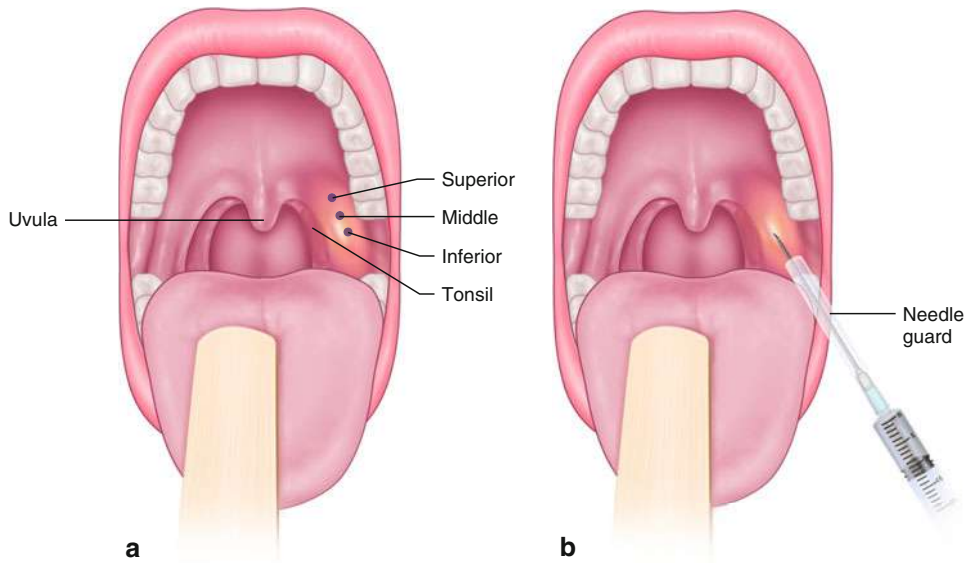

Fig. 59.2 Aspiration of a peritonsillar abscess: (**a**) Aspirate superior pole first. If no pus aspirated, move 1 cm inferior to the middle pole. If still no pus, make final attempt by moving 1 more cm inferiorly to the inferior pole. (**b**) Demonstrates use of needle guard

59.5 Procedure: Incision and Drainage

1. Raise the head of the bed and anesthetize the area as described previously.
2. Have suction set up with a Fraser or tonsil tip, available, and turned on.
3. Have the patient hold the laryngoscope handle after the blade has been inserted orally as described previously.
4. Fashion a blade guard on a #11 or #15 blade scalpel by taping over the blade with only 0.5 cm of the blade exposed at the tip. Incise the area that was previously aspirated with a small stab incision. Suction the area.
5. Insert a curved Kelly clamp into the incision and open gently to enter the abscess cavity and break up loculations (Fig. 59.3). Suction as necessary.
6. Have the patient gargle with a saline solution and expectorate.
7. Do not pack the abscess cavity.
8. Observe for 2–4 h for bleeding.

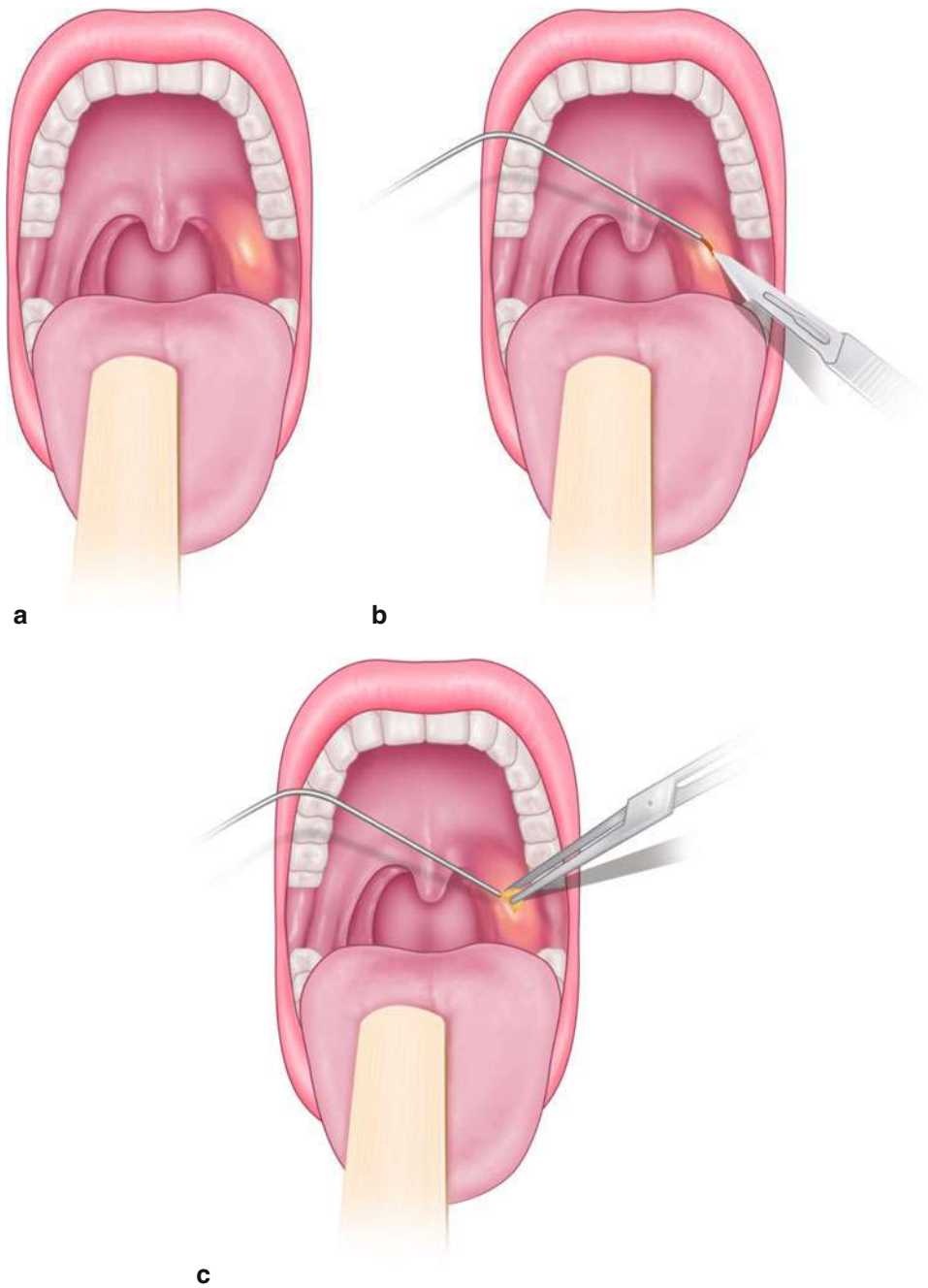

Fig. 59.3 Incision and drainage of peritonsillar abscess: (**a**) Gain access with appropriate lighting. (**b**) Fashion a blade guard on a #11 or #15 blade scalpel by taping over the blade with only 0.5 cm of the blade exposed at the tip. Incise the area that was previously aspirated with a small stab incision. (**c**) Insert a curved Kelly clamp into the incision and open gently to enter the abscess cavity and break up loculations. Do not pack!

59.6 Complications

- Aspiration or incision of the carotid artery
- Excessive bleeding
- Aspiration of purulent material
- Pain

59.7 Pearls and Pitfalls

- Pearls
  - Airway protection should always be considered. Intubation may be necessary for very large abscesses with airway compromise.
  - Bedsides ultrasound is a valuable tool in confirming the diagnosis of peritonsillar abscess. Perform using the intracavitary probe covered with a sterile glove or other appropriate probe cover and insert into the oropharynx. The abscess will appear as any other abscess: an encapsulated, hypoechoic structure. Doppler flow can also be used to locate the carotid artery relative to the abscess.
  - Empirical oral antibiotics after aspiration or I&D are considered usual care and should cover for group A streptococcus and oral anaerobes.
    - Amoxicillin/clavulanate or clindamycin is the most commonly used.
  - A single dose of high-dose steroids may be helpful in relieving symptoms.
  - Patients who cannot tolerate oral fluids, cannot take oral antibiotics, or who appear to have a toxic response should be admitted. Others may be discharged with 24-h follow-up.
- Pitfalls
  - The carotid artery is located approximately 2.5 cm posterolaterally to the tonsils. Take care that the needle is not inserted too far laterally or the risk of aspirating the carotid artery is increased.
  - Incision into the tonsil itself may cause excessive bleeding. Aspiration or incision into the tonsil will likely miss the abscess altogether and may result in misdiagnosis.
  - There is a 1–15 % failure and recurrence rate.

Selected Reading

Afarian H, Lin M. Tricks of the trade—say “ah!”—needle aspiration of peritonsillar abscess. ACEP News. 2008;27(5):38.

Braude DA, Shalit M. A novel approach to enhance visualization during drainage of peritonsillar abscess. J Emerg Med. 2008;35:297–8.

Galioto NJ. Peritonsillar abscess. Am Fam Physician. 2008;77:199–202.

Ozbek C, Aygenc E, Tuna EU, Selcuk A, Ozdem C. Use of steroids in the treatment of peritonsillar abscess. J Laryngol Otol. 2004;118:439–42.

Roberts J, Hedges J, editors. Clinical procedures in emergency medicine. 5th ed. Philadelphia: Saunders; 2009. p. 1184–9.

Vieira F, Allen SM, Stocks RM, Thompson JW. Deep neck infection. Otolaryngol Clin North Am. 2008;41:459–83.

Incision and Drainage of Sublingual Abscess

60

Melinda W. Fernandez and Bobby K. Desai

60.1 Indications

- Sublingual abscess (Fig. 60.1)

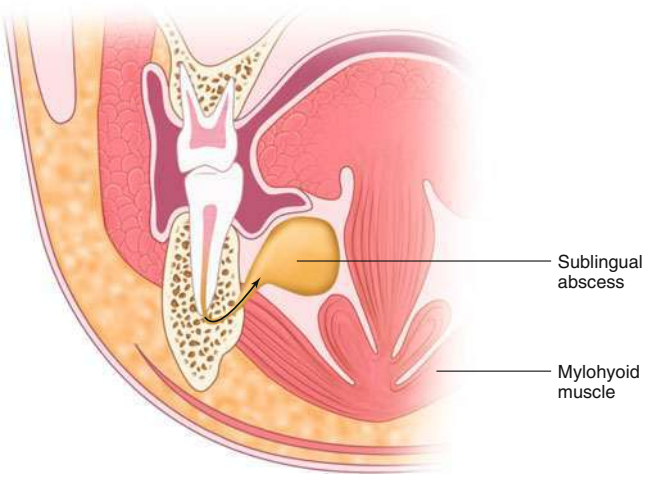

Fig. 60.1 Sublingual abscess

M.W. Fernandez, MD
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: mindyfernandez@ufl.edu

B.K. Desai, MD (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_60

60.2 Contraindications

- Relative
  - Pediatric patient
  - Severe trismus—may need sedation or drainage in the operating room
  - Uncooperative patient—may need sedation or drainage in the operating room
  - Coagulopathy, patients taking anticoagulants, or patients with a known bleeding disorder

60.3 Materials and Medications (Fig. 60.2)

- #11 scalpel
- Hemostats
- Penrose drain
- 4-0 silk suture
- 25- to 27-gauge 1.5-inch needle for injecting anesthetic
- 18- to 20-gauge needle for drawing up anesthetic
- 5- to 10-mL syringe or control syringe
- Viscous lidocaine (or other topical anesthetic)
- Lidocaine 1 % with epinephrine
- Light source
- Suction setup with tonsil suction tip
- Culture swab

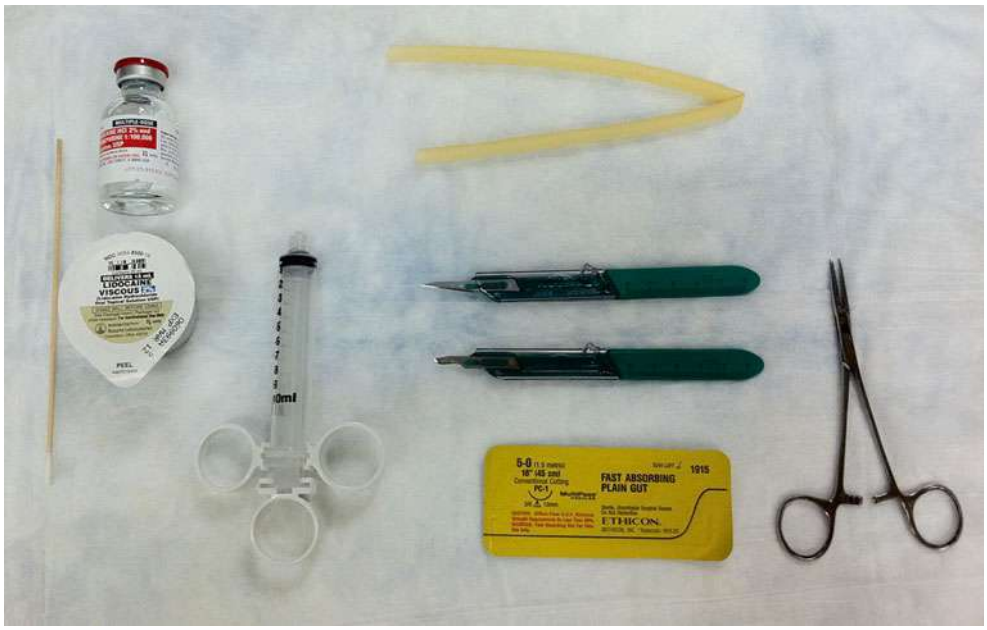

Fig. 60.2 Materials needed: #11 scalpel, hemostats, Penrose drain, 4-0 silk suture, 25–27-gauge 1.5-inch needle for injecting anesthetic, 18- to 20-gauge needle for drawing up anesthetic, 5- to 10-mL syringe or control syringe, viscous lidocaine (or other topical anesthetic), lidocaine 1 % with epinephrine, light source, suction setup with tonsil suction tip, culture swab (optional)

60.4 Procedure: Incision and Drainage

1. After explaining the procedure, risks, and benefits to the patient, put the head of the bed in the most comfortable working position for the clinician and the patient. Alternatively, the clinician may also elect to position the patient in an upright procedure chair that has multiple positions.
2. Apply viscous lidocaine topically with a cotton-tipped swab to the area to be injected with anesthetic. Leave in place for a minute or 2. Alternatively, spray the area with a topical anesthetic spray.
3. Have suction set up with a tonsil tip, available, and turned on.
4. Draw up the appropriate amount of lidocaine with epinephrine into a 5-mL syringe (control syringe if available).
5. Direct light source to area to be worked on. *Adequate lighting is essential.* Options include a headlamp or an overhead adjustable light.
6. Change needle to the 27-gauge needle and inject 1–2 mL of lidocaine with epinephrine using an inferior alveolar block. Alternatively, anesthetize the floor of the mouth around the most fluctuant area. Avoid injecting through infected tissue to avoid possible spread into deeper spaces.
7. Using the scalpel, make an intraoral stab incision superficially at the lowest point of the pus accumulation (Fig. 60.3). This will facilitate evacuation of pus under gravity. Have the suction catheter in the mouth and allow the purulent material to drain into the suction.
8. Obtain a specimen of the purulent fluid for culture and sensitivity.
9. Insert a hemostat into the incision to facilitate drainage, but do not open it up to avoid injury to neurovascular structures. Gently massage the soft tissue surrounding the abscess to assist drainage. Suction as necessary to avoid swallowing or aspirating the pus.
10. Once adequate drainage has been achieved, place a small Penrose drain (or other rubber-type drain) into the cavity, and stabilize on one side with a silk suture that goes through the drain and the mucosa (Fig. 60.4).
11. Have the patient rinse and spit with a half-strength peroxide solution followed by either water or saline.
12. Watch for signs of bleeding or upper airway symptoms.
13. Ensure that the patient can tolerate oral fluids before discharge.

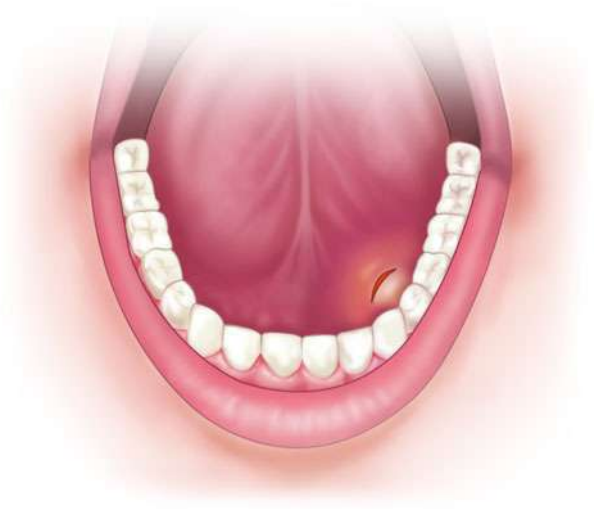

Fig. 60.3 Incise lowest portion of abscess

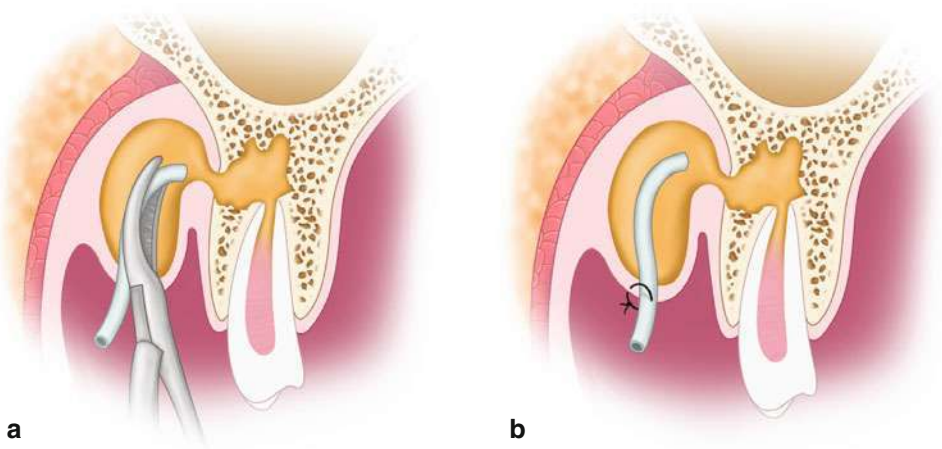

Fig. 60.4 (**a**) Using hemostats, place small Penrose drain into the abscess cavity. (**b**) Suture Penrose in place with a silk suture that goes through the drain and the mucosa

60.5 Complications (Generally Minimal)

- Excessive bleeding (apply pressure—to avoid potential nerve injury, do not cauterize or ligate unless absolutely necessary)
- Aspiration of purulent material (have the patient sit upright and use suction as the abscess is incised)
- Pain

60.6 Pearls and Pitfalls

- Pearls
  - Airway protection is paramount. Infections in this space can quickly spread to the submandibular region and can compromise the airway owing to swelling (Ludwig's angina).
    - Assess for elevation of the floor of the mouth and tongue, ability to lie supine, drooling, stridor, and restlessness.
    - If these signs are present, emergent surgical airway will be necessary before incision and drainage (I&D).
  - The sublingual space is bounded superiorly by the oral mucosa and inferiorly by the mylohyoid muscle.
  - Infected premolars and first molars drain into this space because the apices of their roots are located superior to the mylohyoid muscle.
  - The source molar should be extracted as soon as possible after I&D.
  - Empirical oral antibiotics after I&D should cover for group A streptococcus and oral anaerobes. Penicillin remains the drug of choice, but clindamycin or amoxicillin–clavulanic acid can be substituted.
- Pitfalls
  - The lingual artery, vein, and nerve are contained in the posterolateral area of the floor of the mouth, and the hypoglossal nerve is nearby. These must be avoided when performing I&D for sublingual abscesses.

Selected Reading

Flynn TR, Shanti RM, Levi MH, Adamo AK, Kraut RA, Trieger N. Severe odontogenic infections, part 1: prospective report. J Oral Maxillofac Surg. 2006;64:1093–103.

Reichman E, Simon R, editors. Emergency medicine procedures. New York: McGraw-Hill Education; 2003. p. 1342–5.

Roberts J, Hedges J, editors. Clinical procedures in emergency medicine. 5th ed. Philadelphia: Saunders; 2009. p. 1184–9.

Vieira F, Allen SM, Stocks RM, Thompson JW. Deep neck infection. Otolaryngol Clin North Am. 2008;41:459–83.

Incision and Drainage of Parotid Duct Abscess

61

Melinda W. Fernandez and Bobby K. Desai

61.1 Indications

- Parotid duct abscess

61.2 Contraindications

- Absolute
  - None
- Relative
  - Pediatric patient
  - Severe trismus—may need sedation or drainage in the operating room
  - Uncooperative patient—may need sedation or drainage in the operating room
  - Coagulopathy, patients taking anticoagulants, or patients with a known bleeding disorder

61.3 Materials and Medications (Fig. 61.1)

- #11 scalpel
- 4×4 gauze
- Hemostat
- Penrose drain or ¼-inch packing gauze
- Light source (headlamp or overhead light)
- Culture swab
- 4-0 silk suture
- 25- to 27-gauge needle, 1.5–2 inches long
- 18-gauge needle to withdraw anesthetic from vial
- 5-mL syringe
- Viscous lidocaine or other topical anesthetic
- Lidocaine with epinephrine
- Suction setup with Frazier or tonsil suction tip

M.W. Fernandez, MD
Department of Emergency Medicine,
University of Florida Health, Gainesville, FL, USA
e-mail: mindyfernandez@ufl.edu

B.K. Desai, MD (✉)
Department of Emergency Medicine, University of Florida Health
Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_61

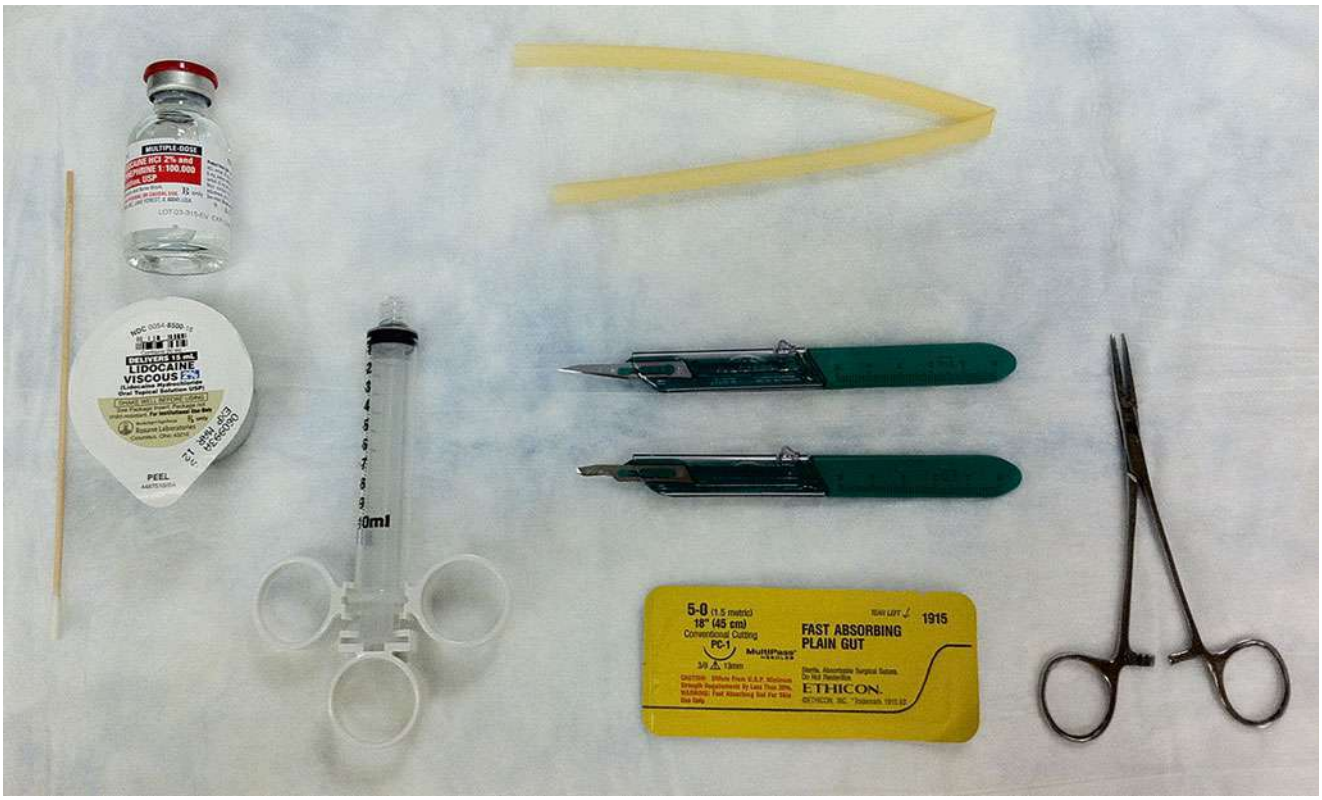

Fig. 61.1 Materials needed: #11 scalpel, hemostats, Penrose drain, 4-0 silk suture, 25- to 27-gauge 1.5-inch needle for injecting anesthetic, 18- to 20-gauge needle for drawing up anesthetic, 5- to 10-mL syringe or control syringe, viscous lidocaine (or other topical anesthetic), lidocaine 1 % with epinephrine, light source, suction setup with tonsil suction tip, culture swab (optional)

61.4 Procedure (Figs. 61.2 and 61.3)

1. After explaining the procedure, risks, and benefits to the patient, put the head of the bed in the most comfortable working position for the operator and the patient. Alternatively, position the patient in an upright procedure chair that has multiple positions.
2. Find the most fluctuant area. Apply viscous lidocaine topically with a cotton-tipped swab to the area to be injected with anesthetic. Leave in place for a minute or 2. Alternatively, spray the area with a topical anesthetic spray.
3. Have suction setup with a tonsil tip, available, and turned on.
4. Draw up the appropriate amount of lidocaine with epinephrine into a 5-mL syringe (control syringe if available).
5. Direct light source to area to be worked on. *Adequate lighting is essential.* Options include a headlamp or an overhead adjustable light.
6. Change needle to a 25- or 27-gauge needle. Inject 1–2 mL of lidocaine with epinephrine just beneath the mucosal surface. Avoid injecting through infected tissue to avoid possible spread into deeper spaces.
7. Using the scalpel, make an intraoral stab incision superficially into the area of greatest fluctuance (Fig. 61.4). Have the suction catheter in the patient's mouth and allow the purulent material to drain into the suction.
8. Obtain a specimen of the purulent fluid for culture and sensitivity.
9. Insert a hemostat into the incision to facilitate drainage, but do not open it up in order to avoid injury to neurovascular structures. Gently massage the soft tissue surrounding the abscess to assist drainage. Suction as necessary to avoid swallowing or aspirating the pus.
10. Have the patient rinse and spit with a half-strength peroxide solution.
11. Place a small Penrose drain (or other rubber-type drain) into the cavity, and stabilize on one side with a silk suture that goes through the drain and the mucosa. Alternatively, cut a strip of ¼- inch packing gauze and insert into the abscess cavity.
12. Have the patient rinse and spit with a half-strength peroxide solution followed by either water or saline.
13. Watch for signs of bleeding or upper airway symptoms.
14. Ensure that the patient can tolerate oral fluids before discharge.

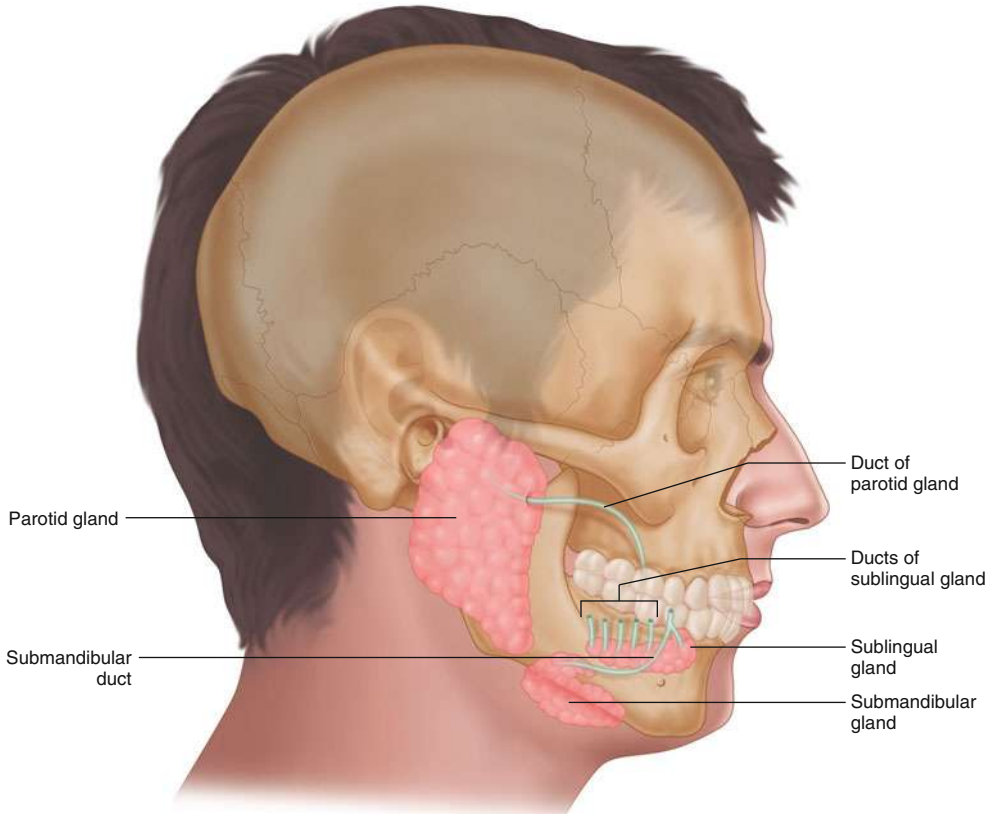

Fig. 61.2 Anatomy of the salivary glands

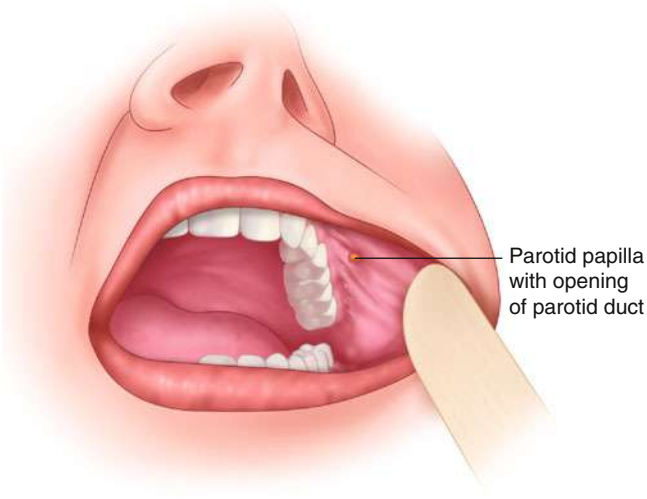

Fig. 61.3 Parotid papilla: opening of the parotid gland located adjacent to third upper molar

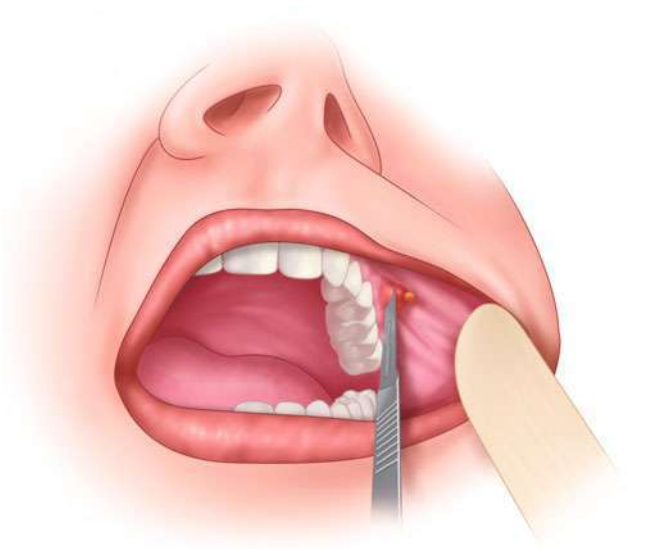

Fig. 61.4 Parotid abscess; make a superficial stab into most fluctuant area

61.5 Complications

- Complications are generally minimal but may include:
  - Excessive bleeding (apply pressure; to avoid potential nerve injury, do not cauterize or ligate unless absolutely necessary)
  - Aspiration of purulent material (have the patient sit upright and use suction as the abscess is incised)
  - Pain

61.6 Pearls and Pitfalls

- Airway protection should always be considered.
- Empirical oral antibiotics after incision and drainage are considered usual care and should cover for group A streptococcal and oral anaerobes. Penicillin remains the drug of choice, but clindamycin or amoxicillin–clavulanic acid can be substituted.
- Patients who cannot tolerate oral medications, cannot take oral antibiotics, or who appear to have a toxic response should be admitted. Pediatric patients should be admitted. Others may be discharged with 24-h follow-up.

Selected Reading

Reichman E, Simon R, editors. Emergency medicine procedures. New York: McGraw-Hill Education; 2003. p. 1346–9.

Roberts J, Hedges J, editors. Clinical procedures in emergency medicine. 5th ed. Philadelphia: Saunders; 2009. p. 1184–9.

Part IX

Dental Procedures

62 Techniques of Mandibular Anesthesia

Susana Perry, Joshua Perry, and Rosalia Rey

62.1 Inferior Alveolar Nerve Block

62.1.1 Nerves Anesthetized

- Inferior alveolar, branch of the posterior division of the mandibular nerve (V3, branch of the trigeminal nerve)
- Incisive
- Mental
- Lingual (usually)

62.1.2 Areas Anesthetized (Fig. 62.1)

- Mandibular teeth to midline
- Body of the mandible
- Buccal mucoperiosteum, mucous membrane anterior to the mandibular first molar
- Anterior two thirds of the tongue and floor of the mouth (via the lingual nerve)
- Lingual soft tissues and periosteum (via the lingual nerve)

S. Perry, DMD (✉)
Department of Pediatric Dentistry, University of Florida College of Dentistry, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: sperry@dental.ufl.edu

J. Perry, DMD
Department of Prosthodontics, University of Florida College of Dentistry, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: jperry@dental.ufl.edu

R. Rey, DDS
Department of Restorative Dental Sciences, University of Florida College of Dentistry, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: RREY@dental.ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_62

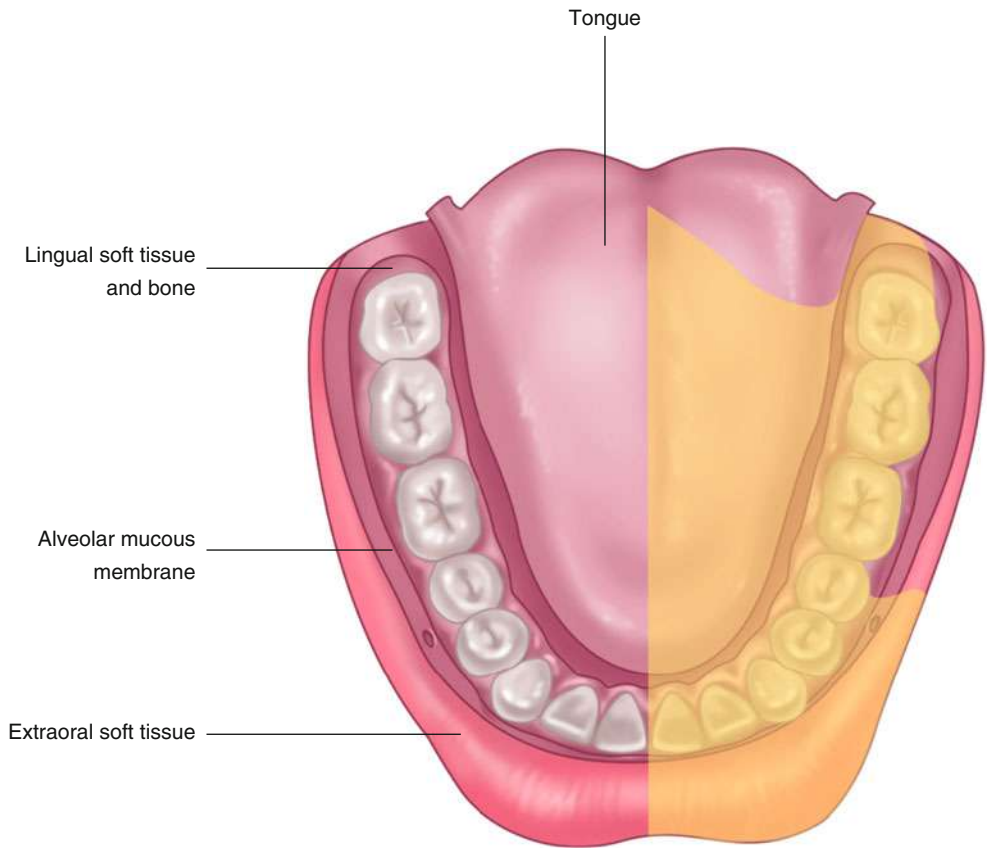

Fig. 62.1 Areas anesthetized with inferior mandibular nerve block

62.1.3 Indications

- When buccal soft tissue anesthesia is necessary
- When lingual soft tissue anesthesia is needed
- When performing procedures on multiple mandibular teeth in one quadrant

62.1.4 Contraindications

- Infection or acute inflammation in the area of injection
- Patients who might bite their lip or tongue (e.g., very young child or physically or mentally handicapped adult or child)
- In relation to local anesthetic use

Absolute

- Local anesthetic allergy
  - Avoid all local anesthetics in the same chemical class (e.g., esters).
- Bisulfite allergy
  - Avoid vasoconstrictor-containing local anesthetics.

Relative

- Atypical plasma cholinesterase
- Methemoglobinemia (idiopathic or congenital)
- Significant liver dysfunction (American Society of Anesthesiologists [ASA] III–IV)
- Significant kidney dysfunction (ASA III–IV)
- Significant cardiovascular disease (ASA III–IV)
  - Avoid high concentrations of vasoconstrictors.
  - Use local anesthetics with epinephrine concentrations of 1:200,000 or 1:100,000 or 3 % mepivacaine or 4 % prilocaine.
- Clinical hyperthyroidism (ASA III–IV)
  - Avoid high concentrations of vasoconstrictors.
  - Use local anesthetics with epinephrine concentrations of 1:200,000 or 1:100,000 or 3 % mepivacaine or 4 % prilocaine.

62.1.5 Materials and Medications

- Local anesthetic carpule (1.7–1.8 mL)
  - Mepivacaine 3 % (+epinephrine 1:100,000)
  - Articaine HCI 4 % (+epinephrine 1:100,000 or 1:200,000)
  - Lidocaine HCI 2 % (+ epinephrine 1:50,000 or 1:100,000) (Fig. 62.2)
  - Bupivacaine HCI 0.5 % + epinephrine 1:200,000
- Aspirating syringe (Fig. 62.3)
- Needle (Fig. 62.4)
  - Gauge refers to the lumen of the needle: The smaller the number, the greater the diameter of the lumen.
  - Needles are color coded by gauge: red = 25 gauge, yellow = 27 gauge, and blue = 30 gauge.
  - *Recommendations: For inferior alveolar nerve (IAN) block, it is best to use a 25-gauge long needle.*
- Mouth props
- Retractors

Fig. 62.2 Local anesthetic carpule (1.7–1.8 mL)

Fig. 62.3 Aspirating syringe

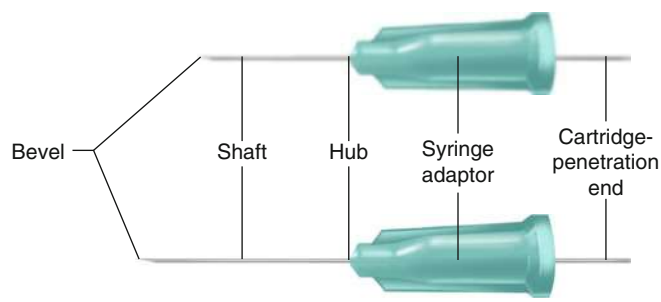

Fig. 62.4 Needle

62.1.6 Procedure

1. Target area: IAN as it passes downward toward the mandibular foramen.
2. Landmarks.
   (a) Coronoid notch
   (b) Pterygomandibular raphe
   (c) Occlusal plane of mandibular posterior teeth
3. Procedure.
   (a) Correct position for operator.
      (i) For a right IAN block, a right-handed administrator should sit at the 8 o'clock position facing the patient.
      (ii) For a left IAN block, a right-handed administrator should sit at the 10 o'clock position facing in the same direction as the patient.
   (b) Recommended to position the patient supine and with the mouth wide open.
   (c) Place thumb on the coronoid notch and index finger extraorally on the posterior border of the ramus in order to estimate the distance between these two points (Fig. 62.5).
      (i) The needle insertion should be three fourths of the anteroposterior distance from the coronoid notch to the deepest part of the pterygomandibular raphe.
   (d) Place the barrel of the syringe in the corner of the mouth on the contralateral side, usually corresponding to the premolars.
   (e) Slowly advance the needle until bony resistance is met.
      (i) For anxious or sensitive patients, a small volume of anesthetic may be deposited as the soft tissue is penetrated.
      (ii) Average depth of penetration to bony contact will be 20–25 mm, approximately two thirds to three fourths the length of a long needle.
      (iii) If the bone is contacted too soon (less than half the length of a long needle), the needle tip is usually located too far anteriorly (laterally) on the ramus. To correct:
         - Withdraw it slightly from the tissues and bring the syringe barrel anteriorly toward the lateral incisor or canine; reinsert to the proper depth.
      (iv) If the bone is not contacted, the needle tip is usually located too far posterior (medial). To correct:
         - Withdraw it slightly in tissue (leaving approximately one fourth its length in tissue), and reposition the syringe barrel more posteriorly (over the mandibular molars).
         - Continue the insertion until contact with the bone is made at an appropriate depth (20–25 mm).
   (f) Aspirate. If negative, slowly deposit 1.5 mL of anesthesia over 60 s.
   (g) Wait 3–5 min before commencing the dental procedure.
4. Precaution: *Do not deposit anesthesia if the bone is not contacted*. The needle tip may be resting within the parotid gland near the facial nerve (cranial nerve VII), and a transient paralysis of the facial never may occur if solution is deposited.

Fig. 62.5 Needle orientation for inferior alveolar nerve block

62.2 Buccal Nerve Block

62.2.1 Nerve Anesthetized

- Buccal nerve, a branch of the anterior division of the mandibular nerve

62.2.2 Area Anesthetized (Fig. 62.6)

- Soft tissues and periosteum buccal to the mandibular molars

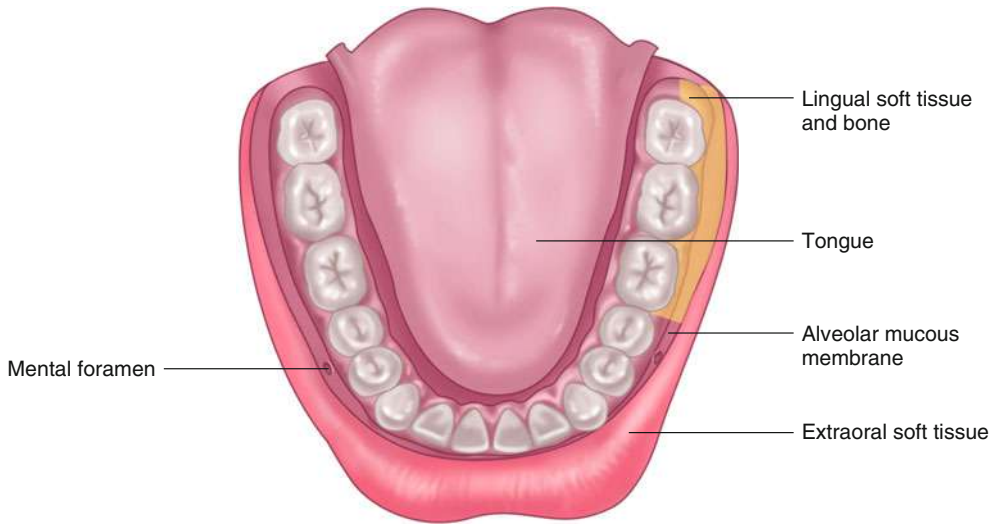

Fig. 62.6 Areas anesthetized with buccal nerve block

62.2.3 Procedure

1. A 25-gauge long needle is recommended.
2. Landmarks: mandibular molars and mucobuccal fold.
3. Orient the bevel of the needle *toward* the bone during injection.
4. Correct positioning.
   (a) For a right buccal nerve block, a right-handed administrator should sit at the 8 o'clock position directly facing the patient.
   (b) For a left buccal nerve block, a right-handed administrator should sit at the 10 o'clock facing in the same direction as the patient.
5. Procedure.
   (a) With the index finger, pull the buccal soft tissues in the area of injection laterally to allow for better visualization.
   (b) Align the syringe parallel to the occlusal plane of the teeth.
   (c) Penetrate the mucous membrane at the injection site, distal and buccal to the last molar (Fig. 62.7).
6. If tissue at the injection site becomes swollen, stop depositing solution.

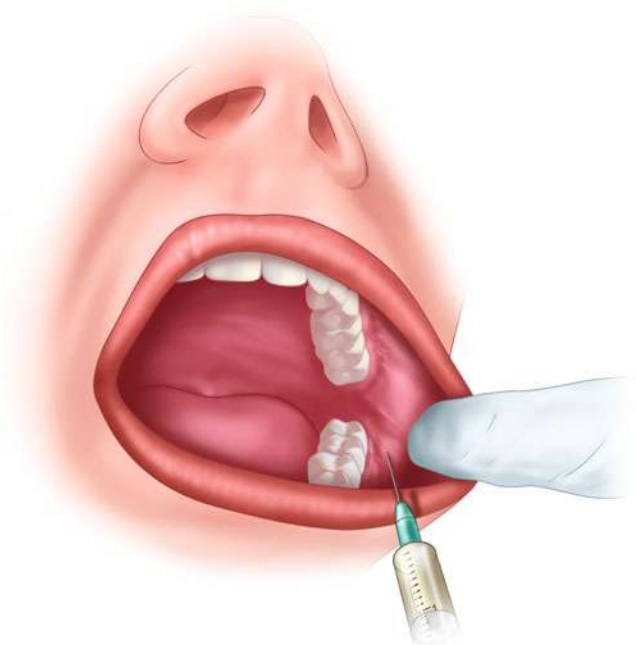

Fig. 62.7 Needle orientation for buccal nerve block

62.3 Mental Nerve Block

62.3.1 Nerve Anesthetized

- Mental nerve, a terminal branch of the inferior alveolar nerve

62.3.2 Area Anesthetized (Fig. 62.8)

- Buccal mucous membranes anterior to the foramen (around the second premolar) to the midline and skin of the lower lip

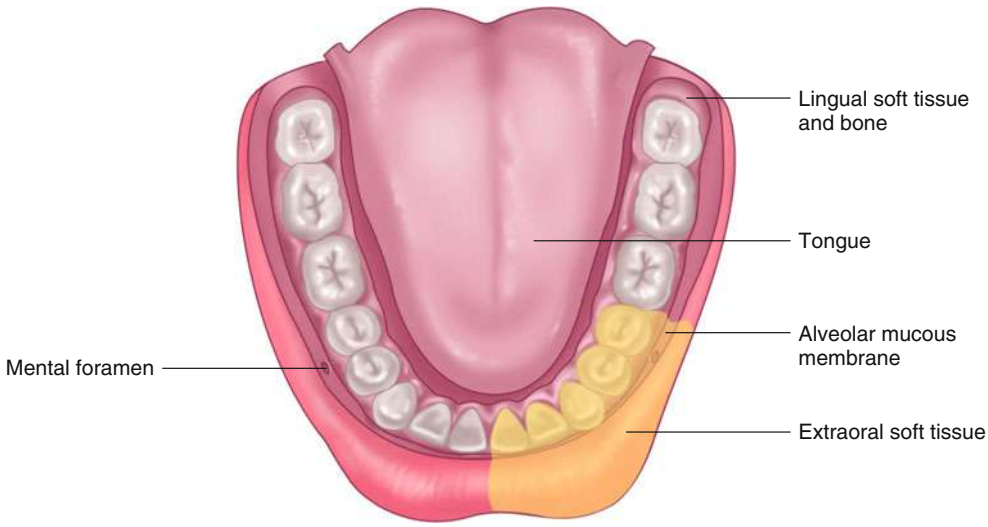

Fig. 62.8 Areas anesthetized with mental nerve block

62.3.3 Indications

- For buccal soft tissue anesthesia in procedures such as:
  - Soft tissue biopsies
  - Suturing of soft tissues

62.3.4 Procedure

1. Area of insertion: mucobuccal fold at or just anterior to the mental foramen.
2. Orientation of bevel should be *toward* the bone during injection.
3. Operator should sit in front of the patient so that the syringe is below the patient's line of sight.
4. Locate the mental foramen.
   (a) Place the index finger in the mucobuccal fold and press against the body of the mandible in the area of the first molar.
   (b) Move the finger anteriorly until the bone beneath the finger feels somewhat concave.
   (c) Mental foramen is usually found around the apex of the second premolar.
   (d) Orient the needle with the bevel directed toward the bone.
   (e) Penetrate the mucous membrane and advance needle slowly; penetration depth is usually 5–6 mm.
   (f) If aspiration is negative, deposit approximately one third of the cartridge over 20 s.
      (i) If the site balloons, stop the deposition of anesthetic and remove the syringe.

Selected Reading

Bennett CR. Monheim's local anesthesia and pain control in dental practice. 6th ed. St. Louis: Mosby; 1978.

Gow-Gates GAE. Mandibular conduction anesthesia: a new technique using extraoral landmarks. Oral Surg. 1973;36:321–8.

Jastak JT, Yagiela JA, Donaldson D. Local anesthesia of the oral cavity. Philadelphia: WB Saunders; 1995.

Malamed SF. The Gow-Gates mandibular block: evaluation after 4275 cases. Oral Surg. 1981;51:463.

Malamed SF. Handbook of local anesthesia. 5th ed. St. Louis: Mosby; 2004.

63 Reduction of Dislocated Temporomandibular Joint

Christopher J. Spencer and Geraldine Weinstein

63.1 Indications

- Open lock: associated with yawning, vomiting, or opening the mouth wide
- Open lock: associated with a dental procedure
- Open lock: associated with endoscopy
- Open lock: associated with oral intubation
- Time duration: acute to 3 weeks or less duration

63.2 Contraindications for Closed Reduction

- Absolute
  - Head trauma with fracture of the skull, maxilla, mandible, or mandibular
    Condyles
- Relative
  - Dislocation of 30 days or longer (will likely be unable to accomplish reduction without general anesthesia and/or open surgical approach)

63.3 Materials and Medications

- Local anesthetic syringe.
- Lidocaine 2 % 1–2 mL.
- 25- to 27-gauge needle (long or approximately 2 inches long).
- Betadine (povidone-iodine) or other skin antiseptic preparation.
- Gauze padding for thumbs.
- Consider a muscle relaxant.
- Consider conscious sedation.

C.J. Spencer, DDS • G. Weinstein, DDS (✉)
Department of Restorative Dental Sciences, University of Florida
College of Dentistry, Gainesville, FL, USA
e-mail: cspencer@dental.ufl.edu; gweinstein@dental.ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_63

63.4 Procedure

63.4.1 Manual Closed Reduction w*ithout* Local Anesthesia (Figs. 63.1)

1. Position the patient in an upright posture with the mandible at the physician's flexed elbow height (physician's comfortable position).
2. Place the thumbs on the mandibular molars with wrapping around the thumbs to protect from possible biting force once the mandible reduces.
3. Apply bilateral firm force in an inferior direction.
4. The mandible will move rapidly in an inferior and then a posterior direction as the condyles slide back over the height of their respective articular eminences.

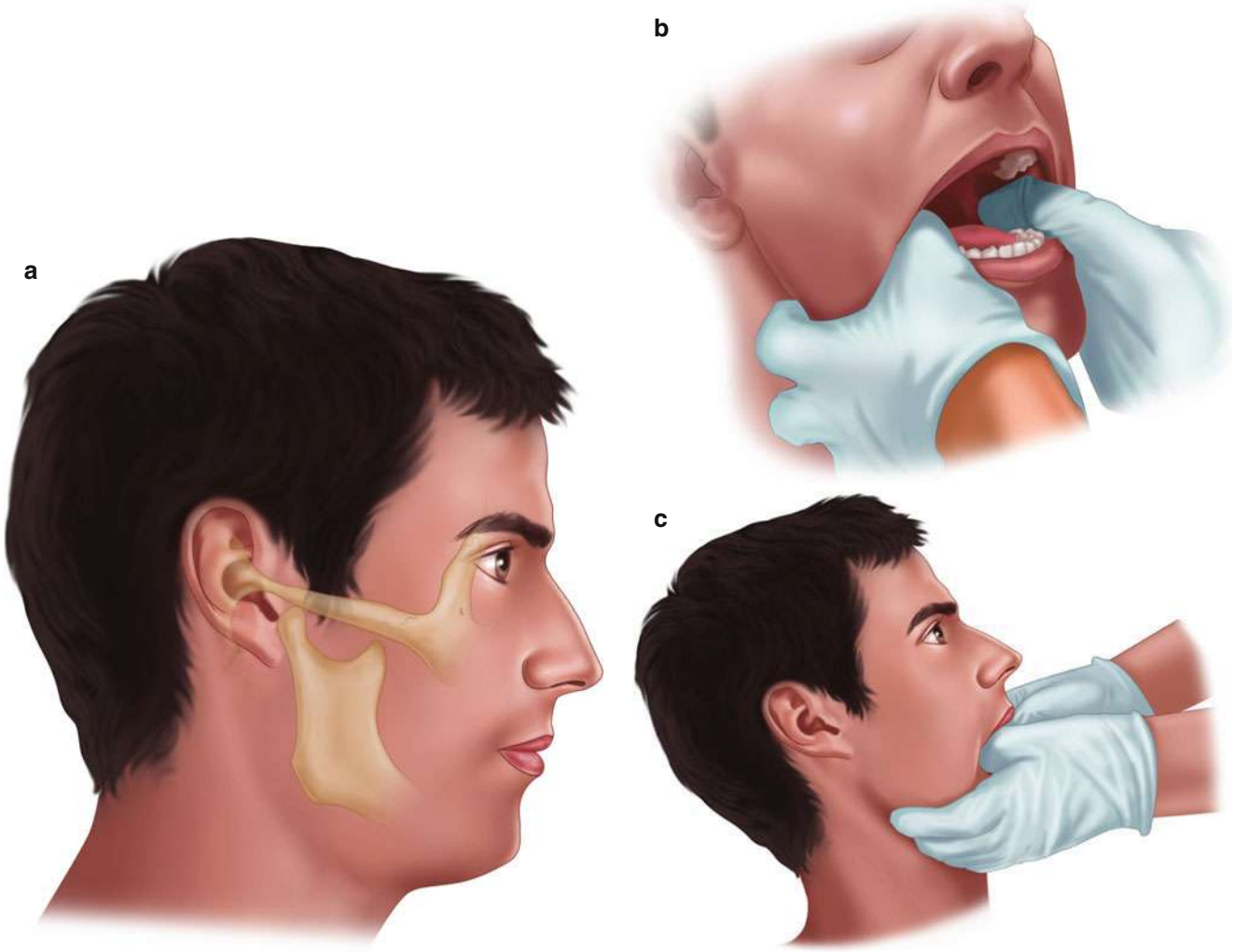

Fig. 63.1 Reduction in progress: (**a**) In a mandibular dislocation, the condyle will be anterior and superior to the articular eminence. (**b**) Position the thumbs on mandibular molars, and apply firm pressure in an inferior direction to distract the TM joint condyles so that they can reposition themselves into the glenoid fossa so that the TM joint can be reduced. (**c**) Lateral view of the distraction force with direction in an inferior direction to distract the condyle of the TM joint

63.4.2 Manual Closed Reduction *with* Local Anesthesia

- If the mandible will not respond to closed reduction with just thumb pressure, likely the masticatory muscles are contracting with sufficient force to prohibit the condyles from being sufficiently distracted owing to pain.
  1. Reduction of pain in the temporomandibular joint (TMJ) with local anesthesia.
  2. *Auriculotemporal block of V3:*
     - The auriculotemporal nerve that innervates the TMJ may be anesthetized inferior to the TMJ capsule. It can be accessed through the skin just anterior to the tragus.
     - With the patient's mouth wide open (it already is in this case), a triangular-shaped hollow will be evident inferior and posterior to the mandibular condyle. Insert the needle at a 20-degree anterior inclination, in the horizontal plane, at the level of the inferior border of the tragus of the ear (Fig. 63.2). The bevel of the needle should be anterior.
     - The needle should be inserted behind the (posterior) ramus and approximately 2 cm deep (aiming for the medial aspect of the posterior border of the ramus). If the posterior border of the ramus is contacted, the needle will need to be directed in a more posterior direction. Then deposit 1–2 mL of lidocaine 2 %.
  3. Then, as before, place the thumbs in a bilateral position on the patient's mandibular molars, and depress the mandible to distract the condyles in an inferior direction.
  4. Conscious sedation may be utilized if the reduction procedure has been arduous and stressful for the patient.

Fig. 63.2 Infiltration of cranial nerve V, the auriculotemporal branch (V3)

63.5 Complications

- Inability to reduce the condyles manually which may lead to more invasive procedures.
- If the condition is acute (≤24 h) and not associated with trauma, there are few if any significant complications or risks for this procedure.

63.6 Pearls

- The manual pressure required on the mandibular molars needs to be sustained and very firm.
- If both condyles are dislocated, it is likely beneficial to attempt one side at a time.
- The clinician needs to protect the thumbs from the impact of the patient's molars during the sudden successful reduction because the biting forces are significant in the molar region.

Selected Reading

Chan TC, Harrigan RA, Ufberg J, Vilke GM. Mandibular reduction. J Emerg Med. 2008;34:435.

Donlon WC, Truta MP, Eversole LR. A modified auriculotemporal nerve block for regional anesthesia of the temporomandibular joint. J Oral Maxillofac Surg. 1984;42:544.

Huang IY, Chen CM, Kao YH, Chen CM, Wu CW. Management of long-standing mandibular dislocation. Int J Oral Maxillofac Surg. 2011;40:810–4.

Prabhakar V, Singla S. Bilateral antersuperior dislocation of the intact mandibular condyles in the temporal fossa. Int J Oral Maxillofac Surg. 2011;40:640–3.

Thagarajah T, Mcculloch N, Thangarajah S, Stocker J. Bilateral temporomandibular joint dislocation in a 29-year-old man: a case report. J Med Case Rep. 2010;4:263.

Dry Socket (Alveolar Osteitis, Fibrinolytic Osteitis)

64

Michael A. Abraham, Amir Azari, Jennifer Westcott, and Franci Stavropoulos

64.1 Indications (Fig. 64.1)

- Definition: severe pain occurring 2–3 days after tooth extraction
- Recent tooth extraction, especially of a mandibular tooth or an impacted third molar
- Partially or completely visible bone socket
- Intense radiating pain (often to the ear)
- Fetid odor without suppuration
- Absence of swelling, lymphadenitis, or bacteremia
- Foreign bodies present in the extraction socket

Fig. 64.1 Clinical photograph suggestive of a dry socket. Clinical correlation is necessary

M.A. Abraham, DMD (✉)
United States Air Force, Dental Corps, Minot, ND, USA
e-mail: mikeabrahamdmd@gmail.com

A. Azari, DMD
Department of Oral and Maxillofacial Surgery, Oregon Health and Science University, Portland, OR, USA
e-mail: Azari@OHSU.edu

J. Westcott, DMD
Private Practice, Palm Beach Gardens, FL, USA
e-mail: Jwestcott561@gmail.com

F. Stavropoulos, DDS
Department of Dental Specialties – Oral and Maxillofacial Surgery, Gundersen Health System, La Crosse, WI, USA
e-mail: Mfstavro@gundersenhealth.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_64

64.2 Contraindications

- Absolute
  - Osteomyelitis
  - Jaw fracture
- Relative
  - Retained roots
  - Infection

64.3 Materials and Medications

- Warm saline or 0.12 % aqueous chlorhexidine solution
- 12-mL syringe with curved tip (Monoject® syringe)
- 25-gauge needle, syringe
- Local anesthetic, 2 % lidocaine 1:100,000 epinephrine
- Suction with small tip or gauze
- Socket dressing
  - Gelfoam or small gauze strips
- Socket medicament
  - Sultan dry socket paste® (guaiacol, balsam of Peru, eugenol, 1.6 % chlorobutanol), iodoform, or eugenol® (zinc oxide eugenol dental cement)
- Curved forceps

64.4 Procedure

1. Administer local anesthesia as necessary.
2. Remove any sutures closing the extraction site.
3. Irrigate the wound gently with warm saline or 0.12 % aqueous chlorhexidine.
4. Carefully suction or gently dry any excess saline; the socket area should be isolated from saliva by using gauze or cotton rolls.
5. Gently place iodoform-soaked gauze, Gelfoam soaked in eugenol, or Sultan dry socket paste in the extraction socket with forceps/Monoject syringe.
6. Rinse with saline and replace the dressing as needed for the first 2–3 days and every 2–3 days thereafter.
7. Remove the dressing, if it does not dissolve, without replacement once the pain has resolved.
8. Pain medication (nonsteroidal anti-inflammatory drugs [NSAIDS] or narcotics) should be prescribed if necessary.
9. Follow-up with dentist.

64.5 Complications

- Delayed healing
- Wound dehiscence

64.6 Pearls and Pitfalls

- *Pearls*
  - Wound irrigation may be so painful at the first visit that administration of a local anesthetic without a vasoconstrictor should be considered.
  - The patient should experience profound pain relief within minutes of placement of the soaked medicated dressing.
  - If a medicated dressing is necessary for more than 2 weeks, reevaluate for development of osteomyelitis.
  - "Dry socket" is not a progressive disease but may persist for 10–14 days whether treated or not; therapy is palliative.
  - Instruct the patient to avoid the following, which can cause changes of pressure in the mouth:
    - Smoking
    - Using a straw
    - Spitting
    - Drinking carbonated beverages (e.g., soda, seltzer water, beer)
- *Pitfalls*
  - Avoid over manipulating the socket because this will increase the amount of exposed bone and pain.

Selected Reading

Bloomquist D, Hooley J, Whitacre R. A self-instructional guide: surgical complications. 3rd ed. Seattle: Stroma; 1983.

Matocha DL. Postsurgical complications. Emerg Med Clin North Am. 2000;18:549–64.

Roberts G, Scully C, Shotts R. Dental emergencies. West J Med. 2001;175:51.

65 Postextraction Hemorrhage

Michael A. Abraham, Amir Azari, Jennifer Westcott, and Franci Stavropoulos

65.1 Indications

- Recent tooth extraction site, presenting with more than a slight oozing of blood
- Full evaluation indicating amount of blood loss, present physical condition, and reason for hemorrhage including coagulopathy or medication use

65.2 Contraindications

- Absolute
  - None
- Relative
  - None

65.3 Materials and Medications

- 2×2 gauze pad
- Saline
- 25-gauge needle, syringe
- Local anesthetic without vasoconstrictor—2 % lidocaine plain
- Gelfoam® (absorbable gelatin-compressed sponge) or oxidized cellulose
- Topical thrombin
- Suture kit with 3-0 chromic gut suture or 3-0 Vicryl® suture (synthetic absorbable sterile surgical suture composed of a copolymer made from 90 % glycolide and 10 % L-lactide)
- Hemostat

65.4 Procedure

1. Use suction and saline irrigation to gently rinse the affected area. If a "liver" clot is present, irrigate and remove it with suction.
2. Determine the source of hemorrhage without local anesthesia, if possible, because the use of local anesthetic with an added vasoconstrictor may obscure bleeding sites.
3. Moisten a folded 2×2 gauze pad with saline and place it directly onto the extraction site.
4. Instruct the patient to apply firm biting pressure, and observe for 1 h, changing gauze as necessary.
5. If bleeding persists, an intraoral nerve block should be performed.
   - Blocks are preferred to infiltrations; anesthetic with epinephrine infiltrated near the bleeding site will produce only temporary local hemostasis from vasoconstriction.
6. Gently curette the tooth extraction socket and remove areas of old blood clot or granulation tissue.
7. Check soft tissue for associated arterial bleeding.
   - If hemorrhage is localized to soft tissue, use pressure or tie off vessels.
8. Fold Gelfoam® into a small cylinder to fit into the extraction socket.
9. Place Gelfoam® with topical thrombin or Surgicel® (absorbable hemostat) into the socket and hold in

M.A. Abraham, DMD (✉)
United States Air Force, Dental Corps, Minot, ND, USA
e-mail: mikeabrahamdmd@gmail.com

A. Azari, DMD
Department of Oral and Maxillofacial Surgery, Oregon Health and Science University, Portland, OR, USA
e-mail: Azari@OHSU.edu

J. Westcott, DMD
Private Practice, Palm Beach Gardens, FL, USA
e-mail: Jwestcott561@gmail.com

F. Stavropoulos, DDS
Department of Dental Specialties – Oral and Maxillofacial Surgery, Gundersen Health System, La Crosse, WI, USA
e-mail: Mfstavro@gundersenhealth.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_65

position with a figure-of-eight stitch using 3-0 chromic gut suture or 3-0 Vicryl suture (Figs. 65.1 and 65.2).

10. Fold 2×2 gauze, moisten it with saline, and place it over the suture.
11. Instruct the patient to bite down with firm pressure for 30 min; repeat as necessary.
12. Follow-up with dentist.

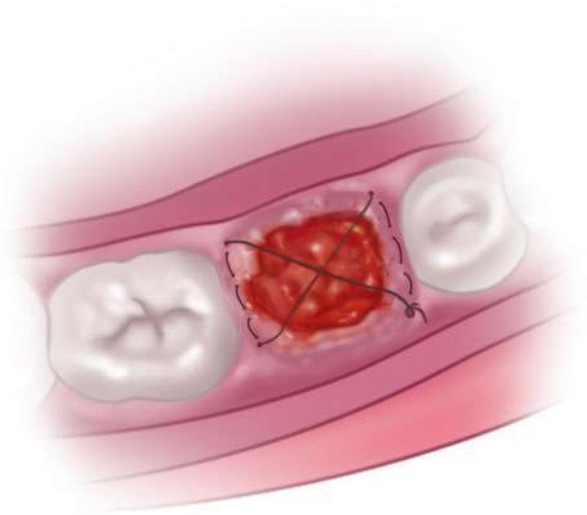

Fig. 65.1 A resorbable figure-of-eight suture placed over an extraction socket

Fig. 65.2 Gelfoam® (absorbable gelatin-compressed sponge) being placed into the extraction socket (Photograph courtesy of Dr. Michael Abraham)

65.5 Complications

- Continued bleeding due to lack of patient compliance.
- Continued bleeding due to coagulopathy or medication use.
- If bleeding persists and coagulopathy is identified, the administration of intravenous blood replacement products may be necessary.

65.6 Pearls and Pitfalls

- Pearls
  - Minor bleeding concerns may be addressed at home by instructing the patient to bite on the affected area with a tea bag for 30 min (tannic acid in tea is a vasoconstrictor).
  - It is normal for an extraction socket to ooze slight amounts of blood for 12–24 h; it is normal for patients to see some blood on their pillow after waking.
  - The patient should be instructed to avoid the following, which can cause changes of pressure in the mouth:
    - Smoking
    - Using a straw
    - Spitting
    - Drinking carbonated beverages (e.g., soda, seltzer water, beer)
- Pitfalls
  - Small amounts of blood mixed with saliva may deceptively appear as large amounts of blood.

Selected Reading

Bloomquist D, Hooley J, Whitacre R. A self-instructional guide: surgical complications. Seattle: Stroma; 1983. pp. 50–5.

Hupp JR, Ellis III E, Tucker MR, editors. Contemporary oral and maxillofacial surgery. 5th ed. St. Louis: Mosby Elsevier; 2008. p. 195–7.

Fractured Tooth

66

Geraldine Weinstein

66.1 Indications (Fig. 66.1)

- Temporary repair of an acute dental fracture until follow-up by a dentist can be secured.

Fig. 66.1 Examples of fractured teeth

66.2 Methods of Sustaining Dental Fracture

- Traumatic injury to the head/facial area
- Falling down
- Extensive tooth decay that has undermined the integrity of the tooth structure
- Biting down on something hard

G. Weinstein, DDS
Department of Restorative Dental Sciences,
University of Florida College of Dentistry,
Gainesville, FL, USA
e-mail: gweinstein@dental.ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_66

66.3 Four Types of Fractured Tooth (Fig. 66.2)

- *Type 1*: contained to the enamel of the tooth, asymptomatic, and can be treated easily with a composite filling by a dentist.
- *Type 2*: involves a fracture through the dentin layer of the tooth. The patient may experience some sensitivity to temperature changes and chewing. Depending on the severity, treatment may include a root canal and a restoration by a dentist.
- *Type 3*: involves the pulp of the tooth and will require endodontic treatment by a dentist.
- *Type 4*: a root fracture in the tooth that makes it nonrestorable and requiring extraction. It is diagnosed by means of a periapical radiograph taken in a dental office.

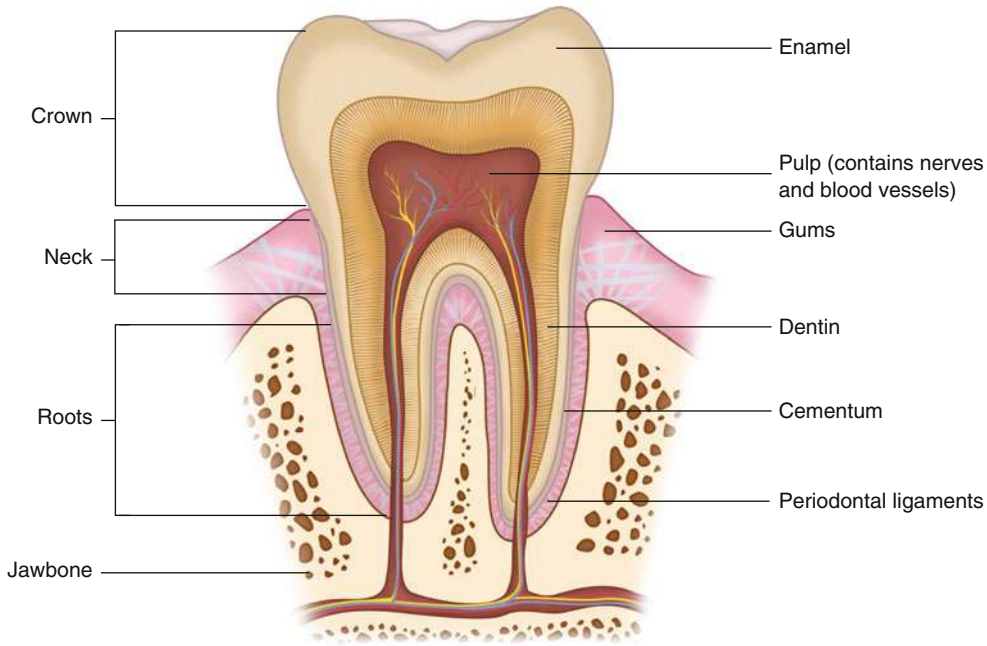

Fig. 66.2 Anatomy of a tooth

66.4 Contraindications

- *Relative*
  - Patients at high risk of aspiration owing to intoxication and altered mental status

66.5 Materials and Medications

- Warm saline or 0.12 % aqueous chlorhexidine solution for irrigation of tissues and tooth
- Irrigating syringe
- Suction with a small tip
- Gauze to control hemorrhage
- Resorbable sutures and local anesthetic as needed for soft tissue lacerations
- Temporary tooth restoration material like intermediate restorative material (IRM) and glass ionomer (like Fuji)

66.6 Procedure

1. Have the patient rinse the mouth out with warm water to clean out any debris. Apply an ice pack to the affected cheek to reduce swelling.
2. Anesthetize the area, either locally at the tooth apex or with a nerve block (mental or inferior alveolar).
3. Irrigate the area; assess the fracture intraorally and check for soft tissue lacerations. Use the gauze with finger pressure to control bleeding in the soft tissue or the tooth.
4. Check if the tooth or bony segment is mobile. If so, a referral to a dentist or oral surgeon is necessary as soon as possible for proper assessment.
5. Type II fractures (fracture limited to dentin layer): cover the exposed surface with a temporary dental cement. In a pinch, 2-octyl cyanoacrylate (Dermabond) is an acceptable secondary alternative (Fig. 66.3).
6. Type III (pulp involved)
   - Provide immediate dental follow-up and analgesics.
   - Initiate antibiotic coverage with penicillin or clindamycin.

Fig. 66.3 Example of temporary dental cement (Reproduced with permission from DenTek Oral Care, Inc.)

66.7 Complications

- Loss of a tooth
- Infection or abscess
- Aspiration of a segment or a whole tooth
- Cosmetic deformity

66.8 Pearls

- Be certain to perform a thorough intraoral examination, looking for tooth fragments or lacerations that may be hiding fragments.
- Dental blocks are very useful for pain control.
- If a tooth is not mobile and the *pulp is exposed*, immediate referral (within a few hours) to a dentist is necessary for extraction or endodontic (root canal) treatment of the tooth. Placement of a temporary-type restoration on this tooth is *not* recommended at this time because it may exacerbate symptoms. Prescribe pain medication and possibly antibiotics when the tooth's pulp is exposed and the patient is unable to see the dentist within 24 h.
- If the tooth is not mobile and *the pulp is not exposed*, a temporary restoration can be placed on the tooth and the patient referred to a dentist for treatment. The fractured part of the tooth should be saved in the event that it can be used. If temporary tooth restoration is unavailable in the emergency department, advise the patient that it is readily available at local pharmacies.
- ALL DENTAL FRACTURES, EXCEPT TYPE I, REQUIRE DENTAL FOLLOW-UP WITHIN 24 h.

Dental Avulsion Management

67

Laura Tucker and Abimbola O. Adewumi

67.1 Indications

- The tooth is completely displaced *out* of its socket, leading to severance of the neurovascular pulp supply and separation of the periodontal ligament (Fig. 67.1).
- Diagnosis
  - Clinically, the socket is found empty or filled with coagulum.
  - Imaging (occlusal, periapical, and lateral views of the affected tooth and surrounding area) (Fig. 67.2):
- Confirm vacuous socket.
- Ensure that the missing tooth is not intruded.
- Diagnose root fracture or alveolar fracture.

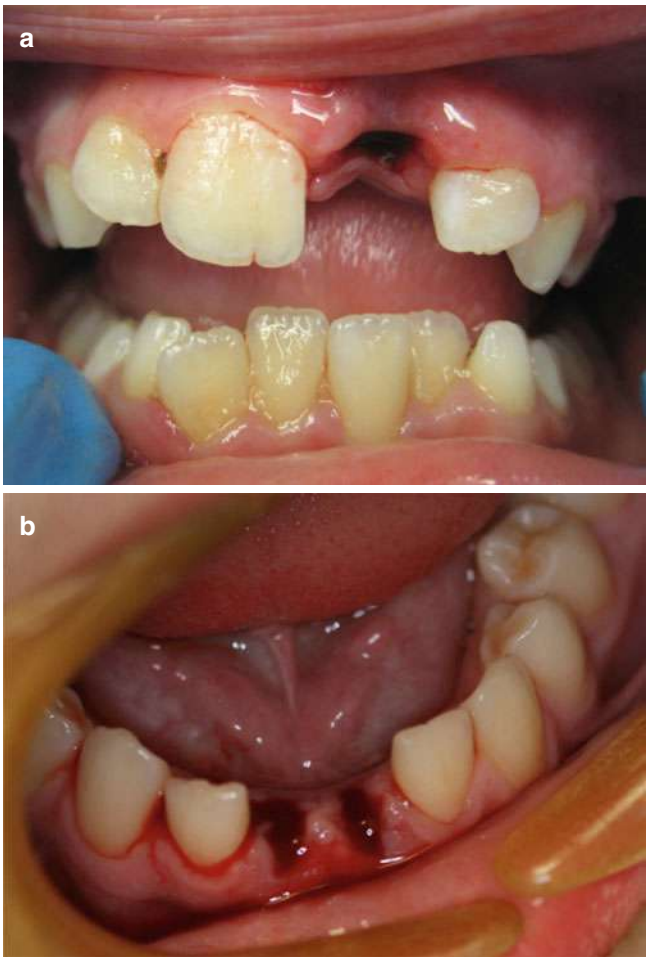

Fig. 67.1 (**a**, **b**) Empty socket following traumatic dental avulsion

L. Tucker, DDS • A.O. Adewumi, BDS, FDSR (✉)
Department of Pediatric Dentistry, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: lauragracesullivan@gmail.com; aadewumi@dental.ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_67

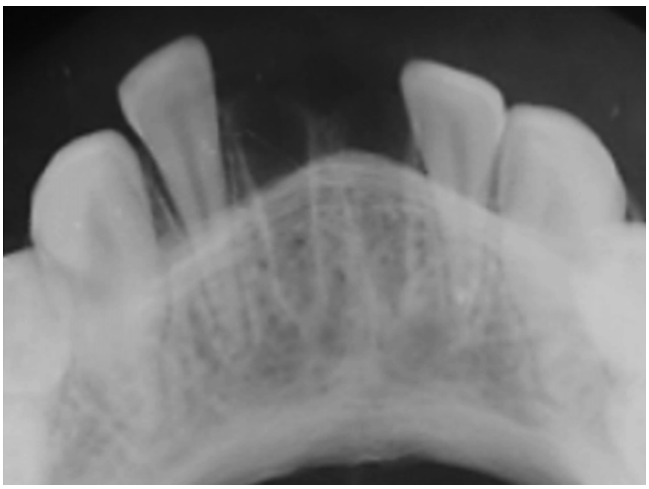

Fig. 67.2 Lower occlusal radiograph shows a complete avulsion of the mandibular right and left central incisors

67.2 Contraindications

- Absolute
  - Avulsed tooth is a primary tooth.
    - DO NOT REPLANT OR REPOSITION AVULSED PRIMARY TEETH.
    - Replantation of primary teeth increases the potential for damage to the developing permanent tooth owing to the increased frequency of pulpal necrosis.
- Relative
  - Fractured root (further intervention required before replantation)
  - Alveolar fracture (further intervention required before replantation)
  - Prolonged extraoral dry time and out of compatible solution (>1 h)
  - Immunocompromised host or congenital heart disease
  - Severe seizure disorder that may put tooth at risk for coming out while the airway is compromised
  - Patient with potential to lose airway reflexes

67.3 Materials and Medications

- Hank's Balanced Salt (Save-A-Tooth) solution or milk in which to preserve tooth until it can be replanted
  - Normal saline if neither of those is available
- Normal saline rinse
- 20- or 18-gauge cannula for gentle irrigation of the socket
- Absorbable suture for gingival lacerations, if present
- Flexible splint materials
  - Round dental wire
  - A flat pliable metal long enough to cover the affected tooth and the two teeth on other side (e.g., the metal nasal bridge from a respirator mask)
- Fixative
  - Dental adhesive
  - Dermabond or some other brand of cyanoacrylate for adhesive skin closures

67.4 Procedure

1. Be certain the tooth is a permanent one and not a primary tooth.
2. If not done by the patient, gently wash the tooth under water for approximately 10 s.
   - Be certain to hold the tooth by the crown, not the root (Fig. 67.3)
3. If the tooth cannot be replanted immediately, place the tooth in Hank's solution or milk.
4. If no such media is available, instruct the patient to hold the tooth inside his or her mouth between the cheek and the gums.
5. Gently replant the tooth, using digital pressure only into as anatomical a position as possible (Fig. 67.4).
   - Assess clinically for alignment.
   - Radiograph for confirmation.
6. Suture any gingival lacerations if present.
7. Apply a flexible splint, securing the affected tooth to the teeth on either side.
   - Consider using skin adhesive both to secure the tooth to its neighbors and, perhaps, to apply a makeshift splint until the patient can be seen by her or his dentist.
     - Towel dry the teeth as best as possible.
     - Apply skin adhesive (using the standard applicator) to the lateral edges of the avulsed tooth where it will make contact with its adjacent teeth.
     - If dental wire is available, apply adhesive to the buccal surfaces of the three teeth (the avulsed tooth central to the other two) and apply length of metal to the Fixodent (Fig. 67.5).
8. Systemic antibiotics with anaerobic coverage is empirical.
9. Ascertain tetanus status for the patient; update if uncertain.

Fig. 67.3 Proper way to hold an avulsed tooth

Fig. 67.4 Gentle replantation using digital pressure

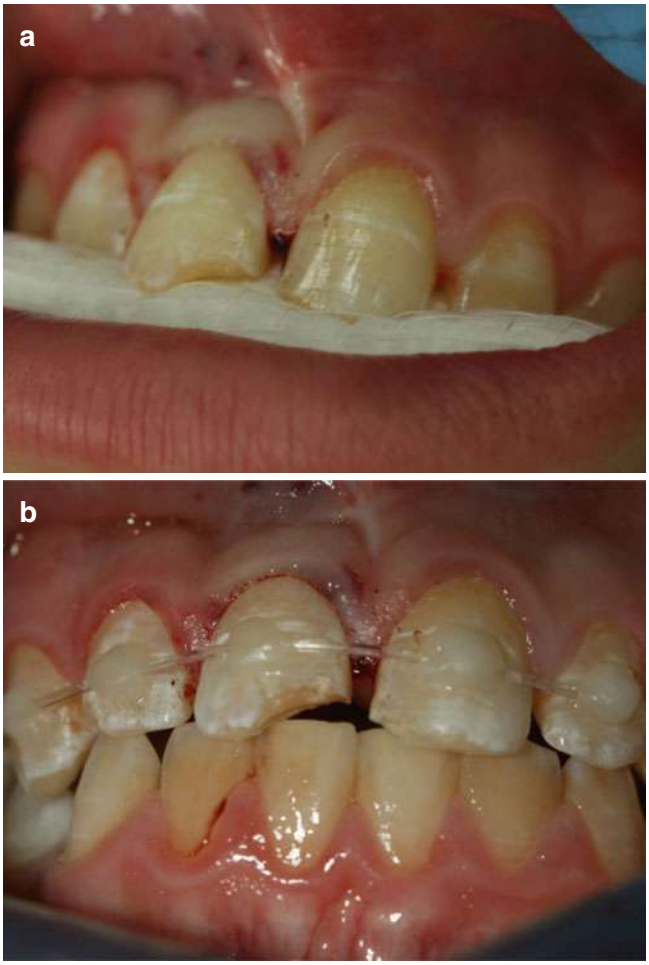

Fig. 67.5 (**a**, **b**) Splint stabilization

67.5 Complications

- Primary teeth
  - Dilaceration (bend) in the permanent tooth crown
  - Enamel defect of the lower permanent incisors as a result of avulsion of the preceding primary tooth
- Permanent teeth
  - Discoloration as a result of loss of vitality of the avulsed tooth.
  - Ankylosis of the alveolar ridge, leading to functional and aesthetic changes.
  - Replacement resorption occurs when the replanted tooth is slowly replaced with bone.
  - External inflammatory resorption is a progressive loss of tooth associated with destruction of adjacent alveolar bone.
  - Infection
  - Aspiration of an inadequately secured replanted tooth.

67.6 Pearls and Pitfalls

- Pearls
  - At the initial examination, make sure that all avulsed teeth are accounted for.
    - If not, a radiographic examination is necessary to ensure that the missing tooth is not completely intruded (pushed into the gum) or has sustained a root fracture with loss of the coronal fragment.
  - In children, always consider the likelihood of nonaccidental trauma (abuse).
  - Short-term and long-term dental follow-up cannot be emphasized enough.
- Pitfalls
  - If the avulsed tooth cannot be accounted for, aspiration is a possibility.
- Prognosis
  - Depends on extraoral dry time (length of time the tooth has been out of the mouth and not stored in an appropriate medium):
    - Ideally, tooth should be implanted within 5 min.
    - Extraoral dry time greater than 60 min has a poor prognosis for periodontal healing.
  - Depends on stage of root development of the avulsed tooth (Fig. 67.6)
- The more advanced the root development, the lower the probability of pulp healing and survival.

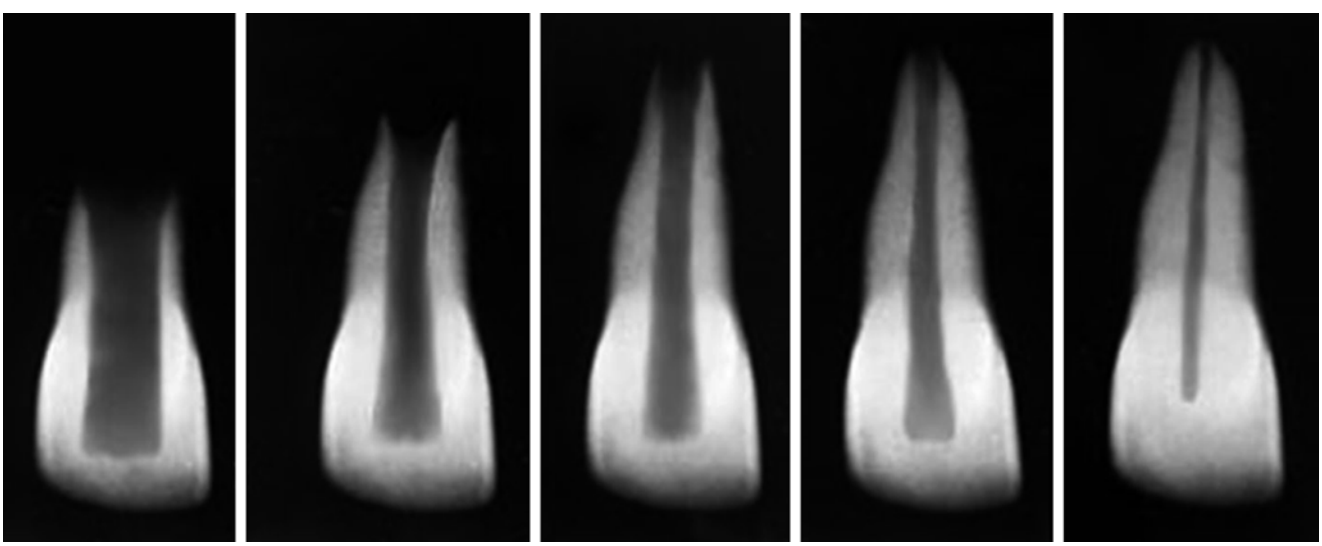

Fig. 67.6 The various stages of root development, from the less advanced (*open apex, left*) to the more advanced (*closed apex, right*)

Selected Reading

AAPD Council on Clinical Affairs. Guideline on management of acute dental trauma. AAPD reference manual. Chicago: American Academy of Pediatric Dentistry; 2010–2011. pp. 202–12.

Andreasen FM, Andreasen JO. Avulsions. In: Andreasen JO, Andreasen FM, Andersson L, editors. Textbook and color atlas of traumatic injuries to the teeth. 4th ed. Oxford: Blackwell; 2007. p. 444–88.

Andreasen JO, Jensen SS, Sae-Lim V. The role of antibiotics in preventing healing complications after traumatic dental injuries: a literature review. Endod Topic. 2006;14:80–92.

Finucane D, Kinirons MJ. External inflammatory and replacement resorption of luxated, and avulsed replanted permanent incisors: a review and case presentation. Dent Traumatol. 2003;19: 170–4.

Flores MT, Andersson L, Andreasen JO, et al. Guidelines for the management of traumatic dental injuries. II. Avulsion of permanent teeth. Dent Traumatol. 2007;23:130–6.

Hile LM, Linklater DR. Use of 2-octyl cyanoacrylate for the repair of a fractured tooth. Ann Emerg Med. 2006;47:424–6.

The dental trauma guide. Available at: www.dentaltraumaguide.org.

Part X

Gastrointestinal Procedures

68 Excision of Thrombosed External Hemorrhoid

Latha Ganti

68.1 Indications

- Acute pain within 72 h of thrombosis onset
- The thrombosis will be visible as a bluish-purplish painful mass in perianal area (Fig. 68.1).

Fig. 68.1 Bluish-purplish appearance of an external thrombosed hemorrhoid. The *ellipse* denotes the area of the elliptical incision to be made (Reproduced with permission from: Fargo and Latimer [3])

68.2 Contraindications

- Absolute
  - Severe coagulopathy
  - Hemodynamic instability
  - Concurrent internal hemorrhoid with rectal prolapse
  - Painless rectal mass (external hemorrhoids are almost always painful, so a painless mass is not a thrombosed external hemorrhoid; also, the whole point of doing this procedure is to provide pain relief).
- Relative
  - Allergy to local anesthetics
  - Perianal infection
  - Inflammatory bowel disease
  - Serious systemic illness or comorbidity that could increase risk of procedure

68.3 Materials and Medications

- Sterile gloves and drape
- Alcohol swabs or pads
- 10 % povidone-iodine or chlorhexidine prep
- 2 % lidocaine with epinephrine
- 5 ml syringe with 25 or 27 gauge and 18 gauge needles
- #11 or #15 scalpel blade and handle
- Direct light source
- Forceps
- Iris scissors
- 4×4 gauze pads
- Adhesive tape
- 3-0 absorbable suture
- 1/4 inch iodoform packing
- Silver nitrate sticks
- Sterile dressing

L. Ganti, MD, MS, MBA
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_68

68.4 Procedure

1. Patient should be in either prone, left lateral decubitus, or jackknife position (Fig. 68.2).
2. Place 2 vertically oriented pieces of tape down each buttock from lower back to upper thigh. Next, place a perpendicular (horizontal) strip spreading buttocks to either side, securing gurney (Fig. 68.3).
3. Place sterile drape over field and center direct lighting over field (Fig. 68.4).
4. Wipe area with alcohol.
5. Inject 1–2 cc of anesthetic into base of hemorrhoid.
6. Clean area with povidone-iodine or chlorhexidine prep.
7. Make an elliptical incision in the roof of the hemorrhoid, being careful to avoid anal sphincter muscle.
8. Remove blood clot(s); multiple clots are often present.
9. If profuse bleeding is a problem, consider cauterization with silver nitrate sticks.
10. Wound can be closed with figure-of-8 absorbable suture OR can be loosely packed with1/4 inch iodoform gauze if not suturing (do not suture wound closed with packing inside).
11. Cover wound with 4×4 gauze folded in half and taped into place (Fig. 68.5).

- Discharge medicines:
  - Antibiotics generally not necessary.
  - Prescribe ibuprofen and/or acetaminophen for analgesia. Avoid opiates, which are constipating.
  - Prescribe stool softeners, to be taken two to three times daily.
- Discharge instructions to patient:
  - Sitz baths 3–4 times daily, for 20 min, warm not hot water.
  - Packing should fall out spontaneously in 2 days.
  - Keep well hydrated.
  - Use gauze to protect underclothing from soilage/blood stains.
  - Return to ED if pain persists beyond 48 h.

Fig. 68.2 Jackknife position

Fig. 68.3 Taping of buttocks to maximize visualization of hemorrhoid

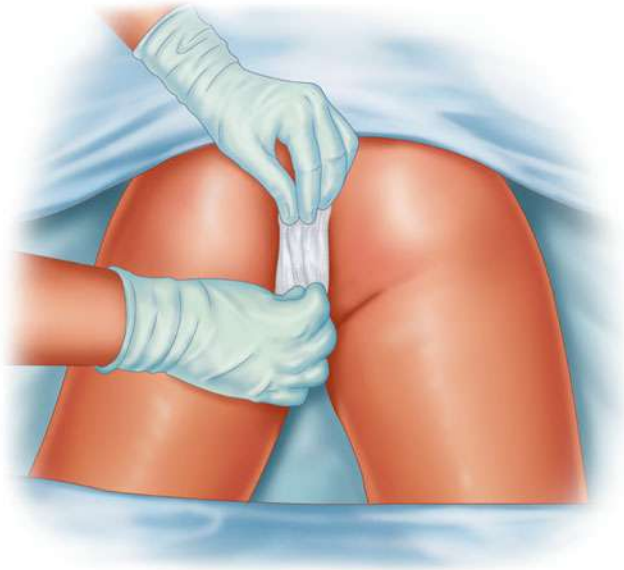

Fig. 68.5 Dress wound with sterile gauze

Fig. 68.4 Direct lighting over field

68.5 Complications

- Common
  - Bleeding: usually self-limited. Can apply cautery or figure-of-8 suture if not previously done
  - Pain: usually controlled with ibuprofen or acetaminophen
  - Perianal skin tag: benign
- Rare
  - Infection rate is 5 % [1].
  - Recurrence rate is 5–19 %, vs. 30 % for simple lancing [2].
  - Stricture and/or incontinence: prevented by avoiding underlying external anal sphincter muscle

68.6 Pearls and Pitfalls

- Pearls
  - Elliptical excision of the hemorrhoid results in much lower recurrence rate than simple lancing.
  - Risk factors for thrombosed external hemorrhoids include constipation, 2nd or 3rd trimester pregnancy, and traumatic vaginal delivery.
- Pitfalls
  - Excision of multiple hemorrhoids in circumferential fashion on all sides of the anal canal can cause anal stenosis.
  - Excision of a painless mass: if it is *painless*, it is *not* a thrombosed external hemorrhoid.

References

1. Lorber BW. Thrombosed external hemorrhoid excision. Medscape.com. www.emedicine.medscape.com/article/81039. Accessed 27 July 2014.
2. Rivadeneira DE. Outpatient and surgical procedures for hemorrhoids. UpToDate.com. http://www.uptodate.com/contents/outpatient-and-surgical-procedures-for-hemorrhoids. Accessed 27 July 2014.
3. Fargo MV, Latimer KM. Evaluation and management of common anorectal conditions. Am Fam Physician. 2012;85(6):624–30.

Selected Reading

Fargo MV, Latimer KM. Evaluation and management of common anorectal conditions. Am Fam Physician. 2012;85(6):624–30.

Jongen J, Bach S, Stübinger SH, Bock JU. Excision of thrombosed external hemorrhoid under local anesthesia: a retrospective evaluation of 340 patients. Dis Colon Rectum. 2003;46(9):1226–31.

69 Diagnostic Peritoneal Lavage

Latha Ganti

69.1 Indications (Table 69.1)

- Inability to perform FAST exam due to lack of equipment or operator
- Hemodynamically unstable patient in whom FAST exam is negative or equivocal

Table 69.1 Comparison parameters for DPL, FAST, and CT [1]

	DPL	FAST	CT
Speed	10–15 min	Fastest: <5 min	Variable
Repeatable	Yes, but rarely done	Yes, and frequently done	Yes, but not done often
Cost	$	$$	$$$
Invasive	Yes	No	No
Mobile	Yes	Yes	No
Advantages	Most sensitive for mesenteric and hollow viscus injuries	Highest specificity	Highly accurate but can be hampered by patient movement
Disadvantages	Misses retroperitoneal and diaphragm injuries	Hampered by subcutaneous or intra-abdominal air, obesity, and pelvic fractures Significant false-negative rate	Misses diaphragm, small bowel, and pancreatic injuries Small but significant risk of radiation-associated malignancy Cannot be done at bedside

L. Ganti, MD, MS, MBA
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_69

69.2 Contraindications

- Absolute
  - Indication for laparotomy already exists.
- Relative
  - 2nd or 3rd trimester pregnancy
  - Previous lower abdominal surgery
  - Inexperienced operator
  - Abdominal wall infection
  - Coagulopathy
  - Cirrhosis
  - Morbid obesity

69.3 Materials and Medications

- 10 % povidone iodine prep
- 1 % lidocaine with epinephrine
- Fenestrated drape
- #10 scalpel blade and scalpel holder
- Skin retractors
- Hemostats
- Diagnostic peritoneal lavage (DPL) catheter (standard peritoneal dialysis catheter)
- 10 cc syringe
- Warmed lactated Ringer's or normal saline solution
- Skin stapler
- Simple suture tray with suture material

69.4 Procedure

69.4.1 Patient Preparation

- Place patient in supine position.
- Ensure nasogastric and urethral catheter (Foley) are in place.
- Prep and drape the area from the umbilicus to the symphysis pubis.
- Anesthetize the skin using 1 % lidocaine with epinephrine in the midline where incision will be made (Fig. 69.1).

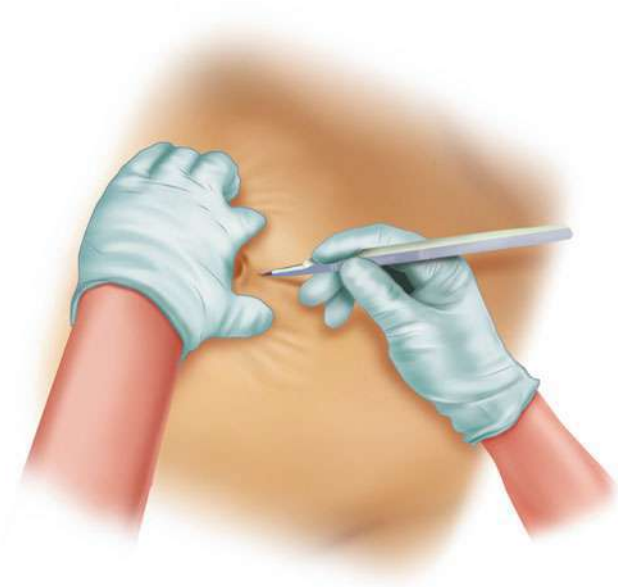

Fig. 69.1 Anesthetize skin where incision will be made (supraumbilical incision shown here, can also make infraumbilical incision)

69.4.2 Three DPL Techniques

- Semiopen (Seldinger) technique
  1. Using a #10 scalpel blade, make a 2 cm incision either superior or inferior to the umbilicus.
  2. Dissect subcutaneous fat until linea alba is exposed.
  3. Hold incision open with skin retractors (Fig. 69.2).
  4. Grasp fascia with hemostats on either side of midline.
  5. Insert 18 gauge needle at 45° angle toward pelvis (Fig. 69.3).
  6. First "pop" will be heard once fascia is penetrated.
  7. Second "pop" will be heard once peritoneum is traversed.
  8. Pass guidewire through needle into pelvis (should pass easily without resistance).
  9. Remove needle while keeping wire stable.
  10. Pass dilator over the wire through fascia, and remove (Fig. 69.4).
  11. Slip DPL catheter over guidewire aiming toward pelvis.
  12. Aspirate peritoneal contents with syringe; aspiration of blood is a positive DPL and means immediate laparotomy (can stop DPL procedure here).
  13. If no blood is immediately obvious, then connect the DPL catheter to a liter of warmed lactated Ringer's (LR) or normal saline (NS) solution for lavage (ensure setup has no one-way valves as solution and peritoneal fluid need to be able to freely mix).
  14. Place LR or NS bag on floor once it is almost empty (minimum 300–350 ml for adults or 10–15 ml/kg for children) and allow intra-abdominal fluid to return (Fig. 69.5).
  15. Send fluid for analysis (Table 69.2).
  16. Irrigate wound, and close skin only with staples or sutures.
- Open technique
  1. Make a 5 cm incision inferior to the umbilicus over linea alba and directly visualize peritoneal cavity.
  2. Both fascia (absorbable suture) and skin (nonabsorbable suture) need to be closed.
- Closed technique
  1. Access peritoneal cavity via percutaneous needle access.
  2. No surgical closure required.

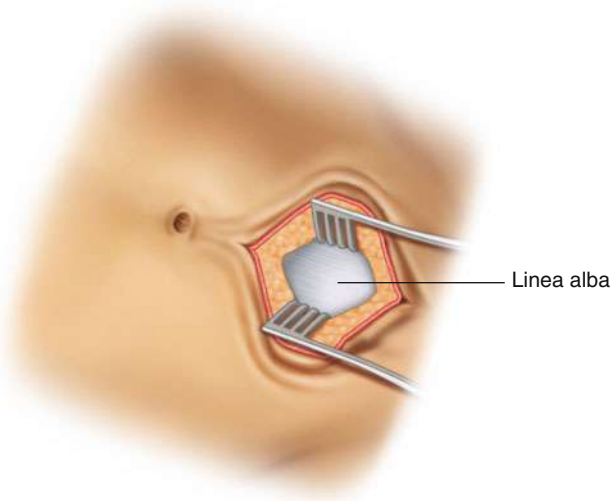

Fig. 69.2 Hold incision open with skin retractors

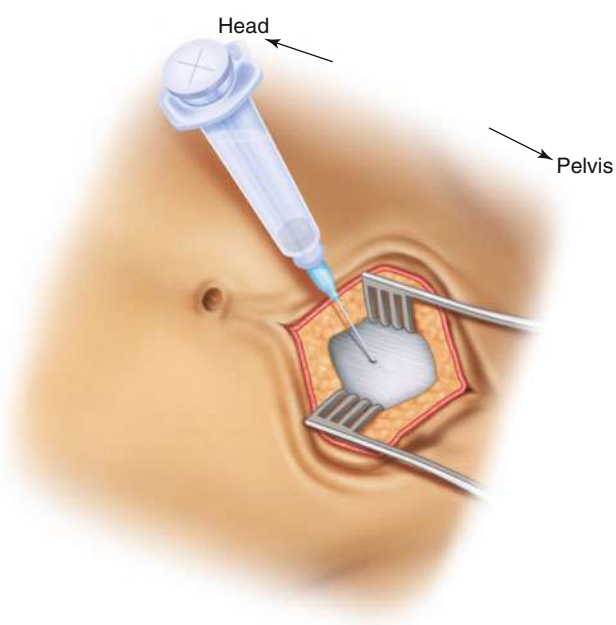

Fig. 69.3 Insert 18 gauge needle at 45° angle toward the pelvis

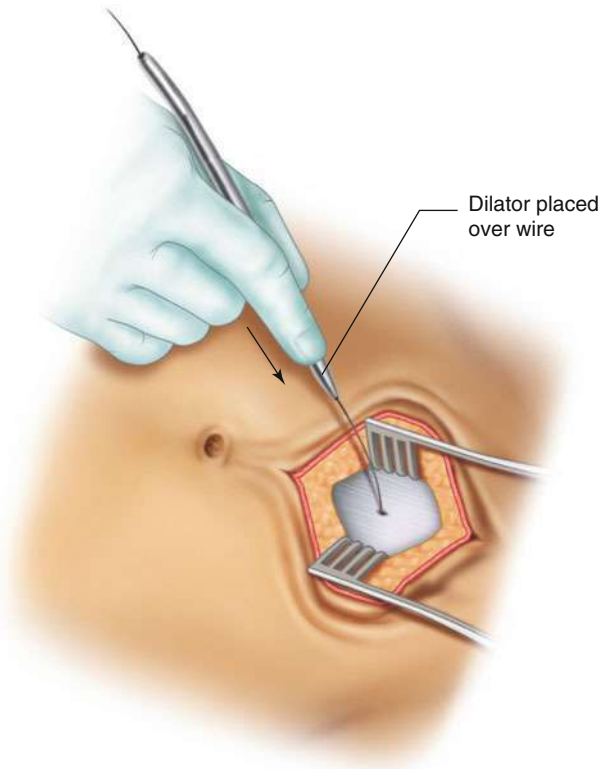

Fig. 69.4 Pass dilator over the wire through fascia and remove

Table 69.2 Diagnostic peritoneal lavage red blood cell criteria (per mm^3) [2]

	Positive	Indeterminate
Immediate gross return of blood via catheter	Any amount	
Immediate return of food particles/intestinal contents	Any amount	
Aspiration of blood	10 cc	
RBC in blunt trauma	100,000	20,000–100,000
RBC in penetrating trauma	10,000	5000–10,000
RBC in gunshot wound	5000	1000–5000
Amylase level (IU/L)	≥175	
Alkaline phosphatase level (IU/L)	≥3	
WBCs (per mm^3)	>500	250–500

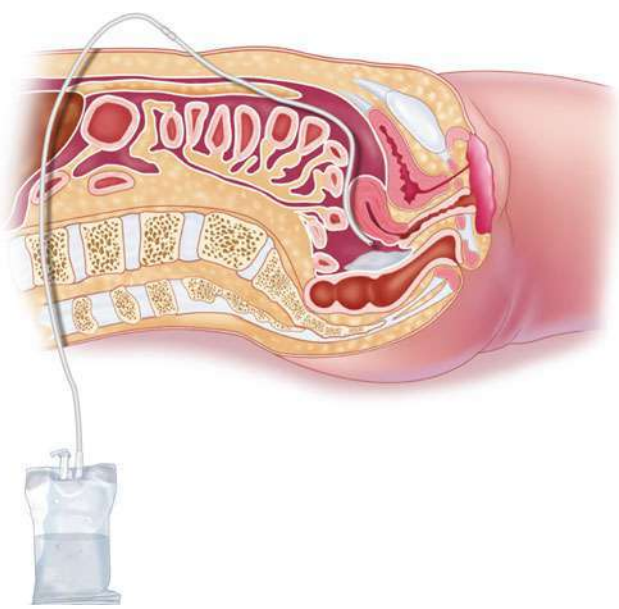

Fig. 69.5 Lavage

69.5 Complications

- Wound infection or dehiscence
- Intraperitoneal injury to organs or vessels (iatrogenic hemoperitoneum)
- Unnecessary laparotomy due to false-positive result from bleeding within rectus sheath or from site of incision
- Potential failure to recover lavage fluid due to:
  - Inadvertent placement of the catheter into the preperitoneal space
  - Compartmentalization of fluid by adhesions
  - Obstruction of fluid outflow (e.g., by omentum)
  - Fluid pooling in the intrathoracic cavity due to diaphragmatic injury
- Incisional hernia

69.6 Pearls and Pitfalls

- Pearls
  - When properly done, complication rate for DPL is low.
  - Prophylactic antibiotics are generally not indicated.
- Pitfalls
  - Inadequate decompression of stomach and urinary bladder increases the chance of injury to these organs; thus, nasogastric and Foley decompression is an important step in patient preparation.

References

1. Jagminas L. Diagnostic peritoneal lavage. Medscape.com. http://emedicine.medscape.com/article/82888-overview#a17. Accessed 28 Aug 2014.
2. Marx JA. Diagnostic peritoneal lavage. In: Ivatury RR, Cayten CG, editors. The textbook of penetrating trauma. Baltimore: Williams & Wilkins; 1996. p. 337.

Selected Reading

Whitehouse JS, Weigelt JA. Diagnostic peritoneal lavage: a review of indications, technique, and interpretation. Scand J Trauma Resusc Emerg Med. 2009;17:13.

70 Manual Reduction of Abdominal Hernia

Latha Ganti

An abdominal wall hernia is a protrusion of the intestine through an opening or area of weakness in the abdominal wall. See Table 70.1 for types of abdominal hernias and Fig. 70.1 for locations along the abdominal anatomy.

Table 70.1 Types of abdominal hernias

Type	Defect	Most commonly seen in	Notes
Inguinal	Intestine or bladder protrudes through abdominal wall or into inguinal canal in the groin	Men because of a natural weakness in this area	96 % of all groin hernias are inguinal; 4 % are femoral
Femoral	Intestine enters canal carrying femoral artery into the upper thigh	Women, especially those who are pregnant or obese	
Incisional	Intestine pushes through abdominal wall at the site of previous abdominal surgery	Elderly or overweight people who are inactive after abdominal surgery	
Umbilical	Part of the small intestine passes through abdominal wall near the navel	Newborns and obese women or those who have had many children	In children, not repaired until age five because often resolve on their own
Hiatal	Upper stomach squeezes through hiatus, an opening in the diaphragm through which the esophagus passes		

L. Ganti, MD, MS, MBA
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_70

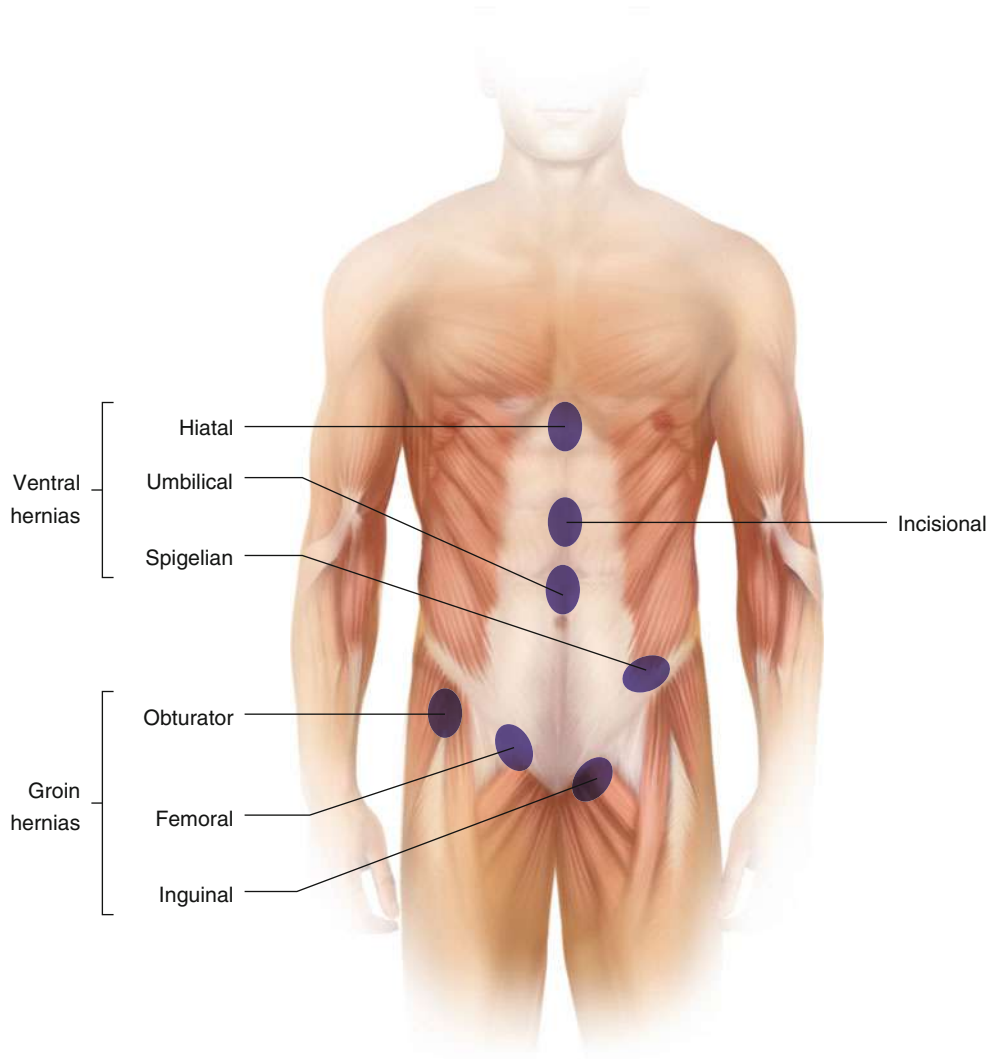

Fig. 70.1 Types of abdominal wall hernias

70.1 Indications

- Incarcerated hernia

70.2 Contraindications

- Absolute
  - Strangulated hernia (could result in placing dead bowel into abdominal cavity)
- Relative
  - Inability to get patient relaxed enough
  - Previous unsuccessful attempts

70.3 Materials and Medications

- Ice or cold compress
- Stretcher or gurney that can tilt to provide Trendelenburg position
- Moderate sedation drugs if providing moderate sedation
- Truss for post-procedure

70.4 Procedure

1. Patient positioning:
   - For abdominal hernia: place patient supine.
   - For groin hernia in adult: place in 20 ° of Trendelenburg.
   - For groin hernias in children: place in unilateral frog leg position (Fig. 70.2).
2. Apply ice or cold compress directly over hernia site to reduce swelling.
3. Administer opiate analgesia or moderate/procedural sedation.
4. Wait up to 30 min as hernia may reduce spontaneously after swelling has gone down and patient is relaxed.
5. Gently apply steady pressure distally on the tissue at the neck of the hernia with one hand and with other hand, guide hernia proximally through fascial defect. Too much pressure distally can cause hernia to balloon further, making manual reduction difficult. Takes up to 15–20 min. Do not rush procedure.
6. Once hernia is reduced, pain will improve.
7. An external support garment or truss (Fig. 70.3) can be helpful to hold reduced hernia in place and serve as a temporizing measure until surgical repair can be done.
8. Advise patient to schedule elective surgical repair.
9. If unable to reduce the hernia, obtain surgical consultation. Do not force repeated attempts.

Fig. 70.2 Frog leg position in child

Fig. 70.3 An example of a truss, or external support, that can be useful as a temporizing measure until definitive hernia repair can be done

70.5 Complications

- Pain.
- Inability to achieve manual reduction, leading to strangulation of the hernia.
- Strangulation can result in peritonitis and sepsis.
- Recurrence.
- Hydrocele.

70.6 Pearls and Pitfalls

- Pearls
  - Definitive treatment for a hernia is surgery (herniorrhaphy). Without surgery, hernias grow larger over time; they do not disappear. Small hernias are easier to fix and result in fewer complications.
  - The only hernia that can resolve on its own is an umbilical hernia in a child.
  - Trusses, bandages, and tape may provide some comfort but do not reduce risk of incarceration or strangulation.
  - Note that if a truss is worn, it should be in place after reduction of the hernia. Also, it can be impractical in hot climates.
- Pitfalls
  - When the constricting neck and the protrusion are both reduced into the abdomen together (known as a reduction *en masse*), without actually reducing the hernia itself, strangulation ensues even though it appears one has reduced the hernia.
  - If there is still considerable pain after the reduction, it is likely the reduction was not successful or that dead bowel has been reduced into the abdominal cavity.
  - Not recognizing strangulation leads to gangrenous bowel, peritonitis, and sepsis.

Selected Reading

Campanelli G, Canziani M, Frattini F, et al. Inguinal hernia: state of the art. Int J Surg. 2008;6 Suppl 1:S26–8.
Jenkins JT, O'Dwyer PJ. Inguinal hernias. BMJ. 2008;336(7638):269–72.
Moses S. Hernia reduction. 2014. http://www.fpnotebook.com/mobile/Surgery/GI/HrnRdctn.htm. Accessed Sept 15, 2014.

Extended Focused Assessment with Sonography for Trauma

71

Coben Thorn and L. Connor Nickels

71.1 Indications

- Blunt abdominal or chest trauma
- Penetrating abdominal or chest trauma
- Undifferentiated hypotension
- The "E" in EFAST refers to the "extended" ability to detect lung pathology such as a pneumothorax or hemothorax during the otherwise standard trauma FAST exam using the same equipment with or without an additional transducer probe.
- Specific findings that can be detected on extended focused assessment with sonography for trauma (EFAST):
  - Pericardial fluid
  - Pleural fluid
  - Free intraperitoneal fluid
  - Pneumothorax
- Free fluid appears as anechoic or black.

71.2 Contraindications

- Need for immediate operative intervention

C. Thorn, MD
Department of Emergency Medicine, Bon Secours St. Francis Health System, Greenville, SC, USA
e-mail: cobenthorn@gmail.com

L.C. Nickels, MD, RDMS (✉)
Department of Emergency Medicine,
University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu

71.3 Materials and Medications

- Ultrasound machine
- Probe(s): phased array probe (5 to 1 MHz) or curved array probe (5 to 2 MHz)
  - Phased array has a smaller footprint, allowing easier access between intercostal spaces (Fig. 71.1); however, curved array provides better resolution of images

Fig. 71.1 Phased array transducer (P17) with a small footprint that is helpful to fit between the ribs and can be used for focused assessment with sonography for trauma (FAST) examination

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_71

(Fig. 71.2). The linear array transducer (L38, 10–5 MHz) is good for lung images.

- Gel
- Skilled ultrasound operator
- ± Laboratory work, cardiac monitor, and two large-bore intravenous (IV) needles
  - All trauma alerts and unstable patients must have all of these.

Fig. 71.2 Curved array transducer (C60) with a larger footprint and better resolution for deeper imaging that can be used for FAST examination as well as lung examination

71.4 Procedure

1. Ultrasound machine in the abdominal preset.
2. Patient in the supine position.
3. Phased array or curved array probe for focused assessment with sonography for trauma (FAST) and linear array for lung.
4. Begin scanning the patient in a systematic fashion.
   - All the views should be scanned by thoroughly sweeping through the area in question in order to maximize the information obtained.
   - All views should be obtained in the same order every time.
   - Obtain all four views, five views if pneumothorax is included.

71.4.1 Subxiphoid Four-Chamber View (Fig. 71.3)

1. Examine for free pericardial fluid.
   - Anechoic (black) stripe seen between the myocardium and the pericardium
2. Probe is placed in the subxiphoid area.
3. Indicator is to the patient's right.
4. Probe is directed toward the patient's left shoulder.
5. Use a shallow angle in the head to feet direction.
6. Should adequately visualize the following:
   - Liver edge superficially
   - Right ventricle
   - Left ventricle
   - Right atrium
   - Left atrium
7. If unable to obtain this view, proceed to *parasternal long-axis view*:
   - Probe is placed perpendicular at the left parasternal border.
   - Third to fourth intercostal space.
   - Indicator is to the patient's right shoulder.
   - Coronal section through the heart's long axis should adequately visualize the following:
     – Right ventricle most superficially
     – Left ventricle
     – Mitral valve
     – Left atrium
     – Aortic valve
     – Aortic outflow tract

Fig. 71.3 (**a**) Image shows how to obtain the subxiphoid four-chamber view of the heart using the curved array transducer (C60) (Photograph courtesy of F. Eike Flach, MD).(**b**) Ultrasound image of four-chamber view of the heart (Used with permission from First aid for the emergency medicine clerkship 3rd Ed, McGraw Hill, 2011.) *RV* right ventricle, *LV* left ventricle, *RA* right atrium, *LA* left atrium

71.4.2 Right Upper Quadrant View (Fig. 71.4)

1. Examine for free fluid in all of the following areas:
   - Right intrathoracic space
     - Anechoic area above the diaphragm
   - Morison's pouch: hepatorenal space
     - Anechoic stripe between the liver and the kidney
   - Right paracolic gutter
     - Anechoic collection surrounding the inferior tip of the kidney
2. Probe is placed in the midaxillary line on the right.
3. Indicator is directed toward the patient's head.
4. Probe is in the coronal plane, angle can be aimed obliquely while scanning anterior to posterior.

Fig. 71.4 (**a**) Image shows how to obtain the right upper quadrant view of the FAST exam using the curved array transducer (C60). The probe is aimed slightly obliquely in the coronal plane to get a better view between the ribs (Photograph courtesy of F. Eike Flach, MD). (**b**) Ultrasonographic view of the liver. (**c**) Ultrasonographic view of the liver–kidney interface (Morrison's pouch). *L* liver, *K* kidney

71.4.3 Left Upper Quadrant View (Fig. 71.5)

1. Examine for free fluid in all of the following areas:
   - Left intrathoracic space
     - Anechoic area above the diaphragm
   - Subphrenic space
     - Anechoic stripe below the diaphragm and above the spleen
   - Splenorenal space
     - Anechoic stripe between the spleen and kidney
   - Left paracolic gutter
     - Anechoic collection surrounding the inferior tip of the kidney
2. Probe is placed in the midaxillary line on the left.
3. Indicator is directed toward the patient's head.
4. Probe in the coronal plane, angle can be aimed obliquely while scanning anterior to posterior.

71.4.4 Pelvic View (Figs. 71.6 and 71.7)

1. Examine for intraperitoneal free fluid in the pelvis:
   - Anterior pelvis, above the bladder
     - Anechoic fluid above the bladder
   - Posterior cul-de-sac (pouch of Douglas)
     - Anechoic fluid posterior to the bladder or uterus
2. Probe is placed above the pubic symphysis over the bladder.
3. Scan through in both planes:
   - Transverse plane (Fig. 71.6)
     - Indicator is to the patient's right.
     - Scan through the bladder in the head to feet direction.
   - Sagittal plane (Fig. 71.7)
     - Indicator is aimed to the patient's head.
     - Scan through the bladder in a right to left direction.

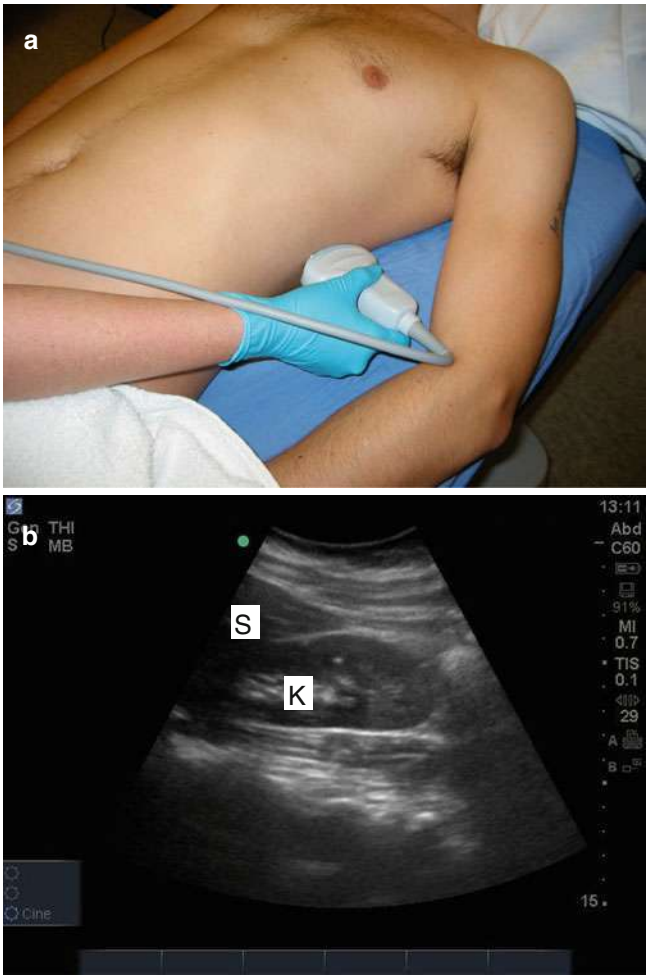

Fig. 71.5 (**a**) Image shows how to obtain the left upper quadrant view of the FAST examination using the curved array transducer (C60). Again, the probe is aimed slightly obliquely and is placed more superiorly in the midaxillary line (Photograph courtesy of F. Eike Flach, MD). (**b**) Ultrasonographic view of spleen–kidney interface. *S* spleen, *K* kidney

Fig. 71.6 Transverse pelvic view of the FAST examination using the phased array transducer (P17). With gentle force, the probe is pressed downward in order to look back behind the pubic symphysis and view the bladder (Photograph courtesy of F. Eike Flach, MD)

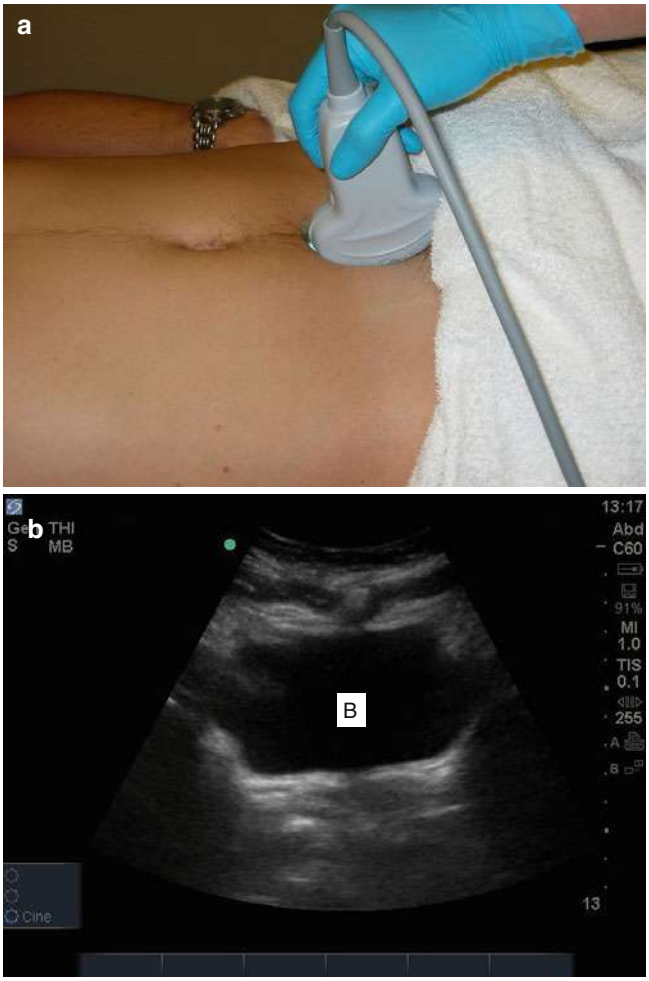

Fig. 71.7 (**a**) Sagittal pelvic view of the FAST exam using the curved array transducer (C60). With gentle force, the probe is pressed downward in order to look back behind the pubic symphysis and view the bladder (Photograph courtesy of F. Eike Flach, MD). (**b**) Ultrasonographic view of the bladder. *B* bladder

71.4.5 EFAST with Lung Views (Figs. 71.8 and 71.9)

1. Examine for pneumothorax:
   - Lung sliding
     - Absence: pneumothorax
     - Presence: normal lung
   - M mode tracing
     - Stratosphere sign: pneumothorax
     - Seashore sign: normal lung
2. Probe is placed on the anterior chest in the midaxillary line.
3. Level of the second to fourth intercostal spaces.
4. Sagittal position.
5. Center the probe over the pleural line between the ribs.
   - Find the rib and then slide the probe toward the head or feet to center the pleural line.
6. Observe for lung sliding.
7. Press M mode and move the line over the pleural line and press M mode again to get the tracing.
8. Examine multiple other areas anteriorly, moving distally, and in midaxillary line laterally, moving from superior to inferior.

Fig. 71.8 Right lung view of the extended focused assessment with sonography for trauma (EFAST) examination using the linear array transducer (L38). The probe is placed in the sagittal plane on the anterior chest in the midaxillary line approximately at the second intercostal space and centered over the pleural line (Photograph courtesy of F. Eike Flach, MD)

Fig. 71.9 Left lung view of the EFAST examination using the linear array transducer (L38) (Photograph courtesy of F. Eike Flach, MD)

71.5 Complications

- Overreliance on ultrasound to rule out abdominal injury:
  - FAST examinations do not detect retroperitoneal bleeding, solid organ injury, contained subcapsular hematomas, and bowel injuries.
- Not scanning through the object in question could lead to false-negative results.

71.6 Pearls and Pitfalls

- Always follow the ABCs (airway, breathing, circulation) first in any unstable patient.
- Always make sure the depth is set adequately.
  - Recommend starting deeper to make sure positive findings are not missed and then adjustments can be made from there.
- The curved array probe may be used throughout the entire EFAST for convenience if necessary.

71.6.1 Subxiphoid Four-Chamber View

- For larger body habitus, need to parallel the probe with the body in the subxiphoid area and use firm pressure to press the entire probe downward so as to look up under the xiphoid process at the heart.
- Moving the entire probe more to the patient's right in the subxiphoid area while still looking toward the left shoulder may improve visualization by using the liver as a window.
- Fat pad:
  - May be mistaken for pericardial fluid
  - Contains echoes and, therefore, is hypoechoic rather than anechoic
  - Should only be present anteriorly
    - Fluid should be gravity dependent, completely encircling the heart, and seen in multiple views.

71.6.2 Right Upper Quadrant View

- Normal artifacts of mirroring and loss of the spine are obscured when pleural fluid is present and, instead, the anechoic fluid is seen and there is loss of mirroring and continuation of the spine.

71.6.3 Left Upper Quadrant View

- Same as right upper quadrant view
- May be more difficult view to find than in right upper quadrant view for all of the following:
  - Spleen and kidney are more posterior and superior than in right upper quadrant view.
  - Spleen is smaller and less of a window for viewing.

71.6.4 Pelvic View

- Bowel can be mista0ken for free fluid or vice versa, but holding the probe still and observing can sometimes help distinguish the two.
  - Peristalsis will occur with bowel.
  - Internal echoes may be present in bowel.

71.6.5 Lung Views

- Ultrasound is more sensitive than a supine portable chest x-ray.
  - Apex anteriorly in midaxillary line.
- Rib
  - Hyperechoic horizontal line with a dense shadow posteriorly
  - Evenly spaced along the chest
- Pleural line
  - First hyperechoic line deep to the rib.
  - Actually includes the visceral and parietal pleura, but appears as one line.
  - Lung sliding is present in normal lung.
  - Comet tail artifact.
  - M mode tracing will be the same in normal lung and pneumothorax above the pleural line and different below the pleural line.
    - Seashore sign (Fig. 27.5a):
      - Appears as waves washing up on the shore.
      - Granular appearance represents movement.
    - Stratosphere sign (Fig. 27.5b):
      - Appears as straight lines
      - Bar code appearance

Selected Reading

Brunett P, Cameron P. Trauma in adults. In: Tintinalli J, Stapczynski J, Ma OJ, Cline D, Cydulka R, Meckler G, editors. Emergency medicine: a comprehensive study guide. 7th ed. New York: McGraw Hill; 2012. p. 1671–5.

Ma JO, Mateer JR, Blaivas M. Trauma. In: Emergency ultrasound. Course Materials; New York, NY: McGraw Hill; 2008. pp. 7–109.

Saul T, Rivera M, Lewiss R. Ultrasound image quality. ACEP News. 2011;4:24–5.

Nasogastric Tube Placement

72

David P. Nguyen, L. Connor Nickels, and Giuliano De Portu

72.1 Indications

- Evaluation of upper gastrointestinal (GI) bleeding (history of melena, bright red blood per rectum, or coffee-ground emesis)
  - Only in the cases in which frank blood is obtained, the sensitivity/specificity in detecting upper GI bleeding is poor.
  - It should not be used for diagnostic purposes. It is used to remove blood that is irritating the stomach and to determine whether bleeding is still occurring (lavage does not clear).
- Commonly used in decompression of the GI tract (partial/complete small bowel obstruction)
- Prevents aspiration and gastric dilation in intubated patients
- Used during gastric lavage and/or removal of toxins (activated charcoal) for acute overdose or poisonings

72.2 Contraindications

- Absolute
  - Facial trauma with possible cribriform plate fracture
- Relative
  - Severe coagulopathy (consider orogastric tube placement)
  - Esophageal strictures and alkali ingestions (possible esophageal perforation)
  - Esophageal varices (studies show that it is actually safe)

72.3 Materials and Medications

- For awake patients, consider pretreatment: lidocaine gel (2 % viscous)/nebulized lidocaine (4 or 10 %), vasoconstrictors (e.g., phenylephrine 0.5 %), and antiemetic (e.g., ondansetron 4 mg).
- 16- or 18-French sump tube lubricating jelly
- 50- or 60-mL syringe stethoscope

D.P. Nguyen, DO
Department of Emergency Medicine, Rush-Copley Medical Center, Aurora, IL, USA
e-mail: davidpnguyen@yahoo.com

L.C. Nickels, MD, RDMS (✉) • G. De Portu, MD
Department of Emergency Medicine,
University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; gdeportu@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_72

72.4 Procedure

- Preparation
  1. Awake patients, should receive antiemetics 15 min before procedure.
  2. Anesthetize both nares at least 5 min before placement.
     – Spray vasoconstrictor into both nares.
     – Inject about 5 mL of lidocaine gel along the floor of the nose.
     – Nebulized lidocaine via facemask also reduces both nasal and pharyngeal discomfort.
  3. Elevate the head of the bed to an upright position (when possible).
  4. Estimate tube insertion distance by measuring the tube from the xiphoid to the earlobe and then to the tip of the nose. Add 6 in. to this estimate and note the total distance. This helps with placement in the stomach and prevents esophageal placement or coiling in stomach. Mark the tube with markers or tape at the desired length.
  5. Lubricate the nasogastric (NG) tube.
- Insertion (Fig. 72.1)
  1. Always insert the tube gently into the nares along the floor of the nose under direct visualization. Always point inferiorly (do not point upward).
  2. If resistance is encountered, try to apply a small amount of pressure. STOP if unable to advance. Try the other side. It is necessary to prevent bleeding or dissecting the tissues.
  3. Have the patient flex his or her head forward when the tube is in the nasopharynx. This helps direct the tube toward the correct placement in the esophagus and not the trachea. Have the awake and cooperative patient sip water from a straw and swallow as the tube enters the oropharynx.
  4. Making the tube more rigid by placing it in cold water will help advance it because the "warmer" tube will tend to coil.
  5. Once the tube is in the esophagus, rapidly advance the tube into the stomach, taking into consideration the previously marked depth.
- Confirmation of tube placement
  1. Insufflate air into the end of the NG tube, via a 50- or 60-mL syringe, while auscultating for a rush of air (borborygmi) over the stomach.
  2. Aspiration of gastric contents (pH<4, there is >90 % gastric placement).
  3. The awake and cooperative patient should be able to talk, and if coughing or severe discomfort occurs, consider that esophageal or bronchial placement might have occurred.
  4. Radiographic evaluation:
     – "Gold standard" is to evaluate simple radiograph for position.
     – Consider in comatose patients.
- Secure the tube
  1. Tape the NG tube in place by taping both the tube and the nose. A butterfly bandage is typically used. Some companies produce a specific fixation for the tube.
  2. Secure the tube to where it does not press on the medial or lateral nostril (can lead to bleeding/necrosis).

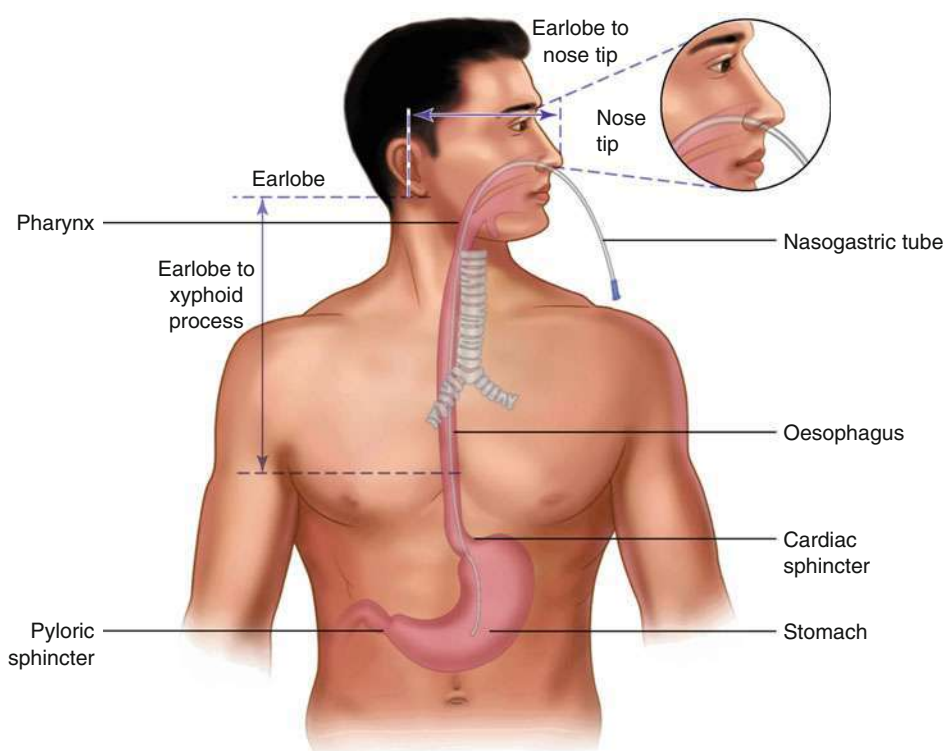

Fig. 72.1 NG tube placement

72.5 Complications

- Inability to pass the tube
- Bleeding
- Curling of the NG tube in the patient's mouth
- Pulmonary placement
- Nasal necrosis

72.6 Pearls

- Pearls
  - NG tube placement was ranked #1 as the most painful procedure in the emergency department so it is imperative to maintain patient's comfort by using anesthetics and even maybe intravenous anxiolytics.
  - Estimate the proper length of the tube before passage to avoid placing the tip of the tube in the esophagus or excessively coiling it in the stomach.
  - If leaving the tube for a prolonged period of time, make sure that the suction is set "intermittent" or "off" to prevent irritation to the gastric mucosa owing to direct pressure.

Selected Reading

Chun DH, Kim NY, Shin YS, Kim SH. A randomized, clinical trial of frozen versus standard nasogastric tube placement. World J Surg. 2009;33:1789–92.

Goff JS. Gastroesophageal varices: pathogenesis and therapy of acute bleeding. Gastroenterol Clin North Am. 1993;22:779.

Henneman PL. Gastrointestinal bleeding. In: Marx J, Hockberger R, Walls R, editors. Rosen's emergency medicine: concepts and clinical practice. 7th ed. Philadelphia: Mosby; 2010.

Tho PC, Mordiffi S, Ang E, Chen H. Implementation of the evidence review on best practice for confirming correct placement of nasogastric tube in patients in an acute care hospital. Int J Evid Based Healthc. 2011;9:51–60.

Esophageal Foreign Body Removal

73

David P. Nguyen, L. Connor Nickels, and Rohit Pravin Patel

73.1 Indications

- Patient presenting with any one or combination of the following:
  - Foreign body sensation in throat, neck, substernal chest, or epigastric area
  - Clear history of ingestion
  - Dysphagia
  - Airway compromise
  - Drooling
  - Inability to tolerate fluids
  - Inability to tolerate solids
  - Evidence of perforation
  - Active bleeding
- All unstable patients should have immediate airway management and urgent endoscopy.
- FBs lodged superior to the epiglottis may be retrieved by emergency physicians in an emergent situation, but generally, a consultant should be present, whether ear, nose, and throat, gastroenterology, or general surgery.

73.2 Contraindications

- Absolute
  - None
- Relative
  - Performing rapid sequence intubation (RSI) in a patient with an FB that could compromise the airway.
  - Generally, if the patient is breathing on their own, collaborate with a consultant on the best method to secure the airway (emergency department, intensive care unit, operating room).
  - Treating with glucagon repeatedly if it is inducing vomiting and/or not working.
  - Conservatively managing a patient who should otherwise undergo a procedure for removal.

73.3 Materials

- Esophagoscopy/endoscopy is the definitive diagnostic and therapeutic procedure for impacted esophageal FBs.
  - Generally, should not be performed by an emergency physician.
  - Devices used include forceps, baskets, polypectomy snares, and nets.
  - Endoscopic techniques include push into stomach, push plus fragmentation, pull with retrieval forceps, and pull with various items (basket, snare, nets).
- Foley catheter removal
  - Widely used technique for recently ingested single, smooth, blunt, and radiopaque objects
  - #12 to #16 French Foley catheter
  - Forceps (bayonet and Magill) of various sizes
  - Often done under fluoroscopic guidance
- Bougienage
  - A single, smooth object, such as a coin, lodged less than 24 h in a patient with no respiratory distress or esophageal disease can be advanced successfully into the stomach by using bougienage.
  - Dilator size is selected based on age:
    - 1–2 years: 28 French
    - 2–3 years: 32 French
    - 3–4 years: 36 French
    - 4–5 years: 38 French
    - Longer than 5 years: 40 French
- Relaxation of the lower esophageal sphincter (LES)
  - Some FBs lodged at the LES can be medically managed by relaxation of the LES.

D.P. Nguyen, DO
Department of Emergency Medicine, Rush-Copley Medical Center, Aurora, IL, USA
e-mail: davidpnguyen@yahoo.com

L.C. Nickels, MD, RDMS (✉) • R.P. Patel, MD
Department of Emergency Medicine,
University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; rohitpatel@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_73

- Most ingested FBs and impacted food boluses eventually pass spontaneously.
  - 1–2 mg of glucagon intravenously
  - 0.4–0.8 mg of nitroglycerin sublingually
  - 5–10 mg of nifedipine sublingually
  - Carbonated beverage

73.4 Procedure

- Push technique and push with fragmentation technique (generally performed by specialists)
  1. First accepted endoscopic method.
  2. Gentle pressure is applied with the tip of the endoscope on the esophageal food bolus after air insufflation.
  3. If pressure does not disimpact the bolus, fragmentation can be attempted but is generally avoided owing to unknown pathology behind the food bolus.
- Foley catheter removal
  1. Moderate sedation and nasopharyngeal topical anesthesia may be used.
  2. Place the patient in a head-down Trendelenburg position.
  3. Check for symmetrical balloon inflation of the Foley catheter.
  4. Under fluoroscopy, visually pass the catheter distal to the FB.
  5. Fill the balloon slowly with 3–5 mL of saline or contrast agent.
  6. Using steady, gentle traction, withdraw the catheter with the balloon inflated distal to the FB.
  7. Grasp the object with fingers, forceps, or clamp once it is visualized in the oropharynx.
- Bougienage
  1. Topical anesthesia is recommended.
  2. Blind esophageal bougienage resembles placement of an orogastric tube.
  3. Place the patient in a sitting position.
  4. Pass a well-lubricated, appropriately sized bougie posteriorly along the roof of the mouth, following the natural curve of the soft palate caudally to the hypopharynx.
  5. Encourage the patient to swallow (to help pass the dilator through the cricopharyngeus muscle).
  6. Ask the patient to phonate to help exclude accidental laryngeal intubation.
  7. Once past the cricopharyngeus muscle, extend the head to aid the bougie in passing distally to the stomach.
  8. Post-procedure radiograph is used to confirm passage into the stomach.
- Relaxation of the LES
  1. Premedicate with an antiemetic, such as ondansetron.
  2. Administer 1–2 mg of glucagon intravenously (0.02–0.03 mg/kg in children, not to exceed 0.5 mg) with the patient in a sitting position over 1–2 min.
  3. Carbonated beverages given after glucagon ingestion have shown to have higher success rates.
  4. An alternative is to use either sublingual nitroglycerin (1–2 0.4 mg tabs) or 5–10 mg of nifedipine to relieve LES tone.
  5. This procedure does not work in patients with structural abnormalities.

73.5 Complications

- Esophageal FBs may cause esophageal pressure leading to edema, necrosis, infection, laceration, and/or perforation.
- Be cognizant of time (risk of complications is higher the longer the FB is left in place) and treatment side effects (i.e., do not continue to give patient water or glucagon if these induce vomiting).
- Aspiration and perforation during procedures listed previously.
- Late complications: esophageal stricture, abscess, mediastinitis, tracheoesophageal fistula, vascular injuries, pneumothorax, pericarditis, aspiration pneumonia, and vocal cord paralysis.

73.6 Pearls

- Esophageal foreign bodies can be lodged in the upper (proximal), middle, or lower (distal) one third:
  - Proximal: cervical web, Zenker's diverticulum
  - Middle: eosinophilic esophagitis, cancer, radiation structure, spastic dysmotility
  - Distal: peptic stricture, eosinophilic esophagitis, cancer, achalasia, esophageal diverticula, spastic dysmotility
- Because food bolus impactions are generally associated with pathology, follow-up evaluation for these abnormalities should be considered.
- Esophagus foreign bodies should not be allowed to remain in the esophagus beyond 24 h from presentation.
- Button/disc batteries in esophagus (emergent removal)
  - Considered an emergency, because liquefaction necrosis and perforation can occur rapidly.
  - Most common ingestions are hearing aid batteries.
  - If in stomach, and patient is a symptomatic, can wait up to 24 h.
- Sharp objects (emergent removal)

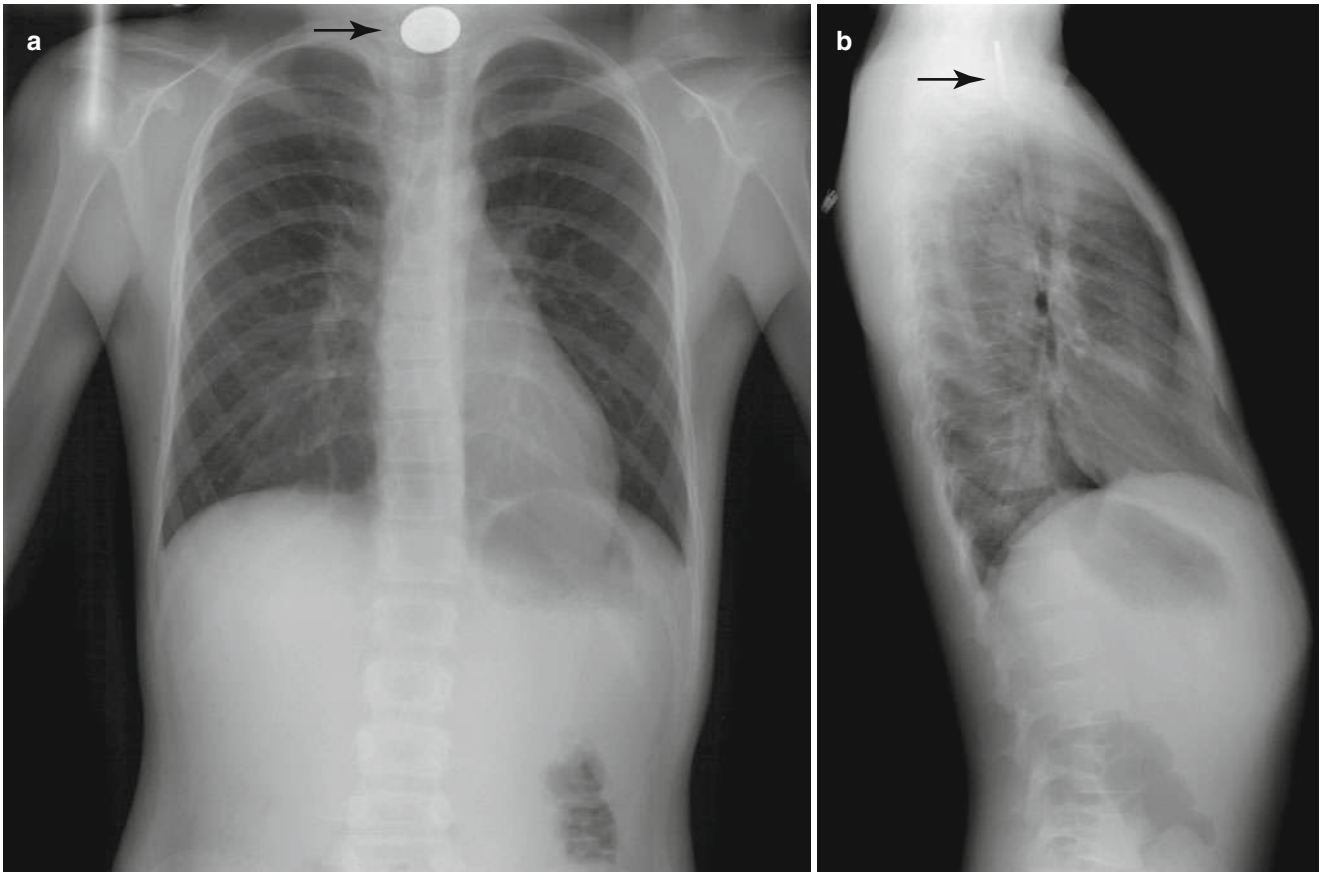

Fig. 73.1 (**a**) Anteroposterior (AP) and (**b**) lateral views demonstrating a coin in the esophagus. A coin in the trachea would present in the opposite manner—the coin would be seen on edge in the AP view and flat on the lateral view (Reproduced with permission from McGraw-Hill: Stead LG, et al. *First Aid for the Pediatrics Clerkship*. New York: McGraw-Hill, 2010)

  - Cause the majority of complications (~35 %) with esophageal FBs.
  - Direct visualization with endoscopy is the only appropriate removal technique.
- Magnets (urgent removal)
  - Can cause necrosis and fistula formation due to the way they adhere to the mucosa
- Esophageal coins (remove within 24 h) (Fig. 73.1)
  - Up to 80 % of coins at the LES will pass spontaneously within 24 h without interventions. The watchful waiting approach is used only in patients with single coins and who are asymptomatic.
  - Common complications of these procedures include mild bleeding, lip laceration, bradycardia with Foley catheter insertion, and teeth injuries.
  - Some protocols include RSI as part of the management process and should be considered if lifesaving.

Selected Reading

ASGE Standards of Practice Committee. Management of ingested foreign bodies and food impactions. Gastrointest Endosc. 2011; 73:1085.

Bhargava R, Brown L. Esophageal coin removal by emergency physicians: a continuous quality improvement project incorporating rapid sequence intubation. CJEM. 2011;13:28–33.

Conway WC, Sugawa C, Ono H, Lucas CE. Upper GI foreign body: an adult emergency hospital experience. Surg Endosc. 2007; 21:455–60.

Katsinelos P, Kountouras J, Paroutoglou G, et al. Endoscopic techniques and management of foreign body ingestion and food bolus impaction in the upper gastrointestinal tract: a retrospective analysis of 139 cases. J Clin Gastroenterol. 2006;40:784–9.

Activated Charcoal

74

Deylin I. Negron Smida and Judith K. Lucas

74.1 Indications

- Single-dose activated charcoal (AC) (Fig. 74.1)
  - Does not meet criteria for gastric emptying.
  - Gastric emptying may be too harmful.
  - Ingestion of toxic xenobiotic is known to be adsorbed by AC.
  - Ingestion occurred with a time frame amenable to adsorption by AC, or clinical factors are present that suggest that not all of the xenobiotic had already been systemically absorbed.
  - Ingestion of extended- or sustained-release formulations.
- Multiple-dose activated charcoal therapy (MDAC)
  - Life-threatening ingestion of:
    - Carbamazepine
    - Phenobarbital
    - Quinine
    - Theophylline
    - Dapsone
  - Life-threatening ingestion of another xenobiotic that undergoes enterohepatic recirculation and is adsorbed to AC
  - Ingestion of a significant amount of a slowly released xenobiotic
  - Ingestion of a xenobiotic known to form concretions or bezoars, such as aspirin

Fig. 74.1 Activated charcoal (AC)

D.I.N. Smida, MD
Department of Emergency Medicine, University of Pittsburgh Medical Center, Saint Margaret Hospital, Pittsburgh, PA, USA
e-mail: Deylin.negron@gmail.com

J.K. Lucas, MD (✉)
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_74

74.2 Contraindications

- Absolute
  - Gastric perforation
  - Gastrointestinal ileus, obstruction, or diminished peristalsis
  - Nonintubated patients with the potential of losing protective airway reflexes
  - Intestinal obstruction
  - Ingestion of:
    - Corrosives
    - Petroleum distillates
- Relative
  - Altered or decreased level of consciousness unless intubated.
  - Vomiting.
  - Xenobiotic has limited toxicity at almost any dose.
  - Dose ingested is less than the dose expected to produce significant illness.
  - Presentation many hours after ingestion.
  - Minimal signs or symptoms of poisoning.
  - Ingested xenobiotic has a highly efficient antidote.
  - Administration of charcoal may increase the risk of aspiration (i.e., hydrocarbons).

74.3 Materials and Medications

- Nasogastric (NG) tube/orogastric (OG) tube (Fig. 74.2)
- Baby bottle with split nipple (designed for drinking of slurry solutions, such as thickened formulas) or sippy cup without the valve
- Absorbent pad
- Basin
- Water-soluble lubricant
- Tubing connected to suction device
- Flavored syrup

Fig. 74.2 Drinking AC (by cup, sippy cup, or bottle) is effective, but it may also be infused via nasogastric or orogastric tube

74.4 Procedure

- Single-dose administration
  - Adult
    - Can be taken via a cup and straw (drunk) if the patient is cooperative.
    - The optimal dose of AC is unknown.
    - 50–100 g/dose (1 g/kg), administered at a rate no less than 12.5 g/h or its equivalent.
    - If vomiting is anticipated, an intravenous antiemetic is recommended.
  - Children
    - 1 g/kg or 10:1 ratio of AC to drug ingested.
    - After massive ingestion, give 2 g/kg.
    - Many children will drink the suspension from a bottle or sippy cup, especially if it is mixed with juice or flavored syrup (e.g., chocolate or strawberry).
- MDAC
  - Adults: 0.5–1 g/kg every 2–4 h for 24–48 h
  - Children (<12 y old): 0.25–0.5 g/kg every 2–4 h or rate of 0.2 g/kg/h for 24–48 h
- Mixtures
  - Ready to drink
  - Powder form
    - Add eight parts water to the selected powdered form.
    - Gatorade or juices can also be used to help hide the flavor and texture.
    - In children, the AC can be mixed with cold chocolate or some other flavored syrup, which also hides the flavor.
  - Shake liquid suspension well for 1 min.
  - If the patient vomits, the dose should be repeated. Smaller, more frequent, dosing may be better tolerated, and an antiemetic may be needed.

74.5 Complications

- Aspiration pneumonitis
- Transient constipation
- Intestinal bezoars
- Bowel obstruction
- Diarrhea, dehydration, hypermagnesemia, and hypernatremia with coadministered cathartics or MDAC
- Vomiting
- Corneal abrasion if spilled in the eyes

74.6 Pearls and Pitfalls

- Pearls
  - If an OG or NG tube is used, time should be allowed for the last dose to pass through the stomach before the tube is removed. Suctioning the tube before removal may prevent subsequent AC aspiration.
  - With children, the colder and sweeter the solution and if the color is camouflaged (cup with a lid or a sippy cup), there will be increased success at oral administration (vs NG/OG).
- Pitfalls
  - No evidence-based literature supports the assertion that AC changes clinical outcome.
  - Xenobiotics and AC adsorption (Table 74.1).
  - Incorrect application (e.g., into the lungs) results in pulmonary aspiration, which can be fatal if unrecognized.
    - Incorrect placement of NG/OG tube into trachea.
    - Administration of AC to a patient with an ileus (e.g., in anticholinergic overdoses).
  - No specific contraindication for AC in pregnant women; however, diarrhea or hypernatremia in the mother may adversely affect the fetus.

Table 74.1 Absorption of xenobiotics by AC

Good absorption	Poor absorption
Acetaminophen	Alkali
Bupropion	Chlorpropamide
Caffeine	Doxepin
Carbamazepine	Ethanol or other alcohols
Chlordecone	Ethylene glycol
Dapsone	Fluoride
Digitoxin	Heavy metals
Nadolol	Imipramine
Phenobarbital	Inorganic salts
Phenylbutazone	Iron
Phenytoin	Lithium
Salicylate	Methotrexate
Theophylline	Mineral acids
	Potassium
	Tobramycin
	Valproate sodium
	Vancomycin

Selected Reading

American Academy of Clinical Toxicology; European Association of Poisons Centres and Clinical Toxicologists. Position statement and practice guidelines on the use of multi-dose activated charcoal in the treatment of acute poisoning. J Toxicol Clin Toxicol. 1999;37:731–51.

Chyka PA, Seger D. American Academy of Clinical Toxicology; European Association of Poisons Centres and Clinical Toxicologists. Position statement: single-dose activated charcoal. J Toxicol Clin Toxicol. 1997;35:721–41.

Gude A, Hoegberg LCG. Techniques to prevent gastrointestinal absorption. In: Nelson LS, Lewin NA, Howland MA, et al., editors. Goldfrank's toxicologic emergencies. 9th ed. New York: McGraw-Hill; 2011. p. 93–7,431.

Lie D. Use of activated charcoal in drug overdose. Medscape family medicine. 25 Mar 2004. www.medscape.com/viewarticle/471331

Olson KR. Emergency evaluation and treatment. In: Olson KR, Anderson IB, Benowitz NL, et al., editors. Poisoning and drug overdose. 5th ed. New York: McGraw-Hill; 2007. p. 1–56.

Gastric Lavage

75

Deylin I. Negron Smida and Judith K. Lucas

75.1 Indications

- Recent ingestion (<30–60 min).
- Life-threatening exposure where there is a high suspicion that a xenobiotic is still present in the stomach and evacuation is expected to contribute to an improved outcome (e.g., iron, tricyclic antidepressants).
- Ingested agent is not absorbed with activated charcoal (e.g., pesticides, hydrocarbons, iron, alcohols, lithium, and solvents).
- Activated charcoal is unavailable.
- Ingestion exceeds adsorptive capacity of initial activated charcoal dosing (e.g., >100 mg/kg of pills).
- Ingestion of an agent likely to form a durable mass or bezoars after overdose.

75.2 Contraindications

- Vomiting
- Unintubated patients with potential to lose airway protective reflexes
- Ingestion of a xenobiotic with aspiration potential (e.g., hydrocarbon) without intubation
- Ingestion of caustic substances (alkali or acidic)
- Ingestion of sharp metals
- Ingestion of a foreign body (e.g., drug packet)
- Risk for hemorrhagic gastrointestinal perforation
- Ingestion of xenobiotic in a form known to be too large to fit into the lumen of the orogastric tube
- Nontoxic ingestions

D.I.N. Smida, MD
Department of Emergency Medicine, University of Pittsburgh Medical Center, Saint Margaret Hospital, Pittsburgh, PA, USA
e-mail: Deylin.negron@gmail.com

J.K. Lucas, MD (✉)
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

75.3 Materials and Medications

- Orogastric tube (Ewald tube or the Tum-E-Vac) (Fig. 75.1)
  - Adults and adolescents: 36–40 French
  - Children: 22–28 French
- Pen or tape to mark the length of the tube
- Water-soluble lubricant
- Suction
- Emesis basin
- Absorbent pad
- Catheter-tip syringe with 2 mL water/saline to check position of the tube
- Room temperature irrigation fluid
- Bite block or oral airway to prevent patients from biting down on the tube

Fig. 75.1 Materials needed for gastric lavage include a large-bore nasogastric tube, a 60-cc non-Luer-Lok syringe, and a solution, typically normal saline, for lavage

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_75

75.4 Procedure

1. If there is potential airway compromise, endotracheal or nasotracheal intubation should precede orogastric lavage.
2. Place an oral airway or a bite block to prevent biting of the endotracheal tube if the patient recovers consciousness or has convulsions during the procedure.
3. Ensure suction apparatus is available and functioning.
4. Place the patient in an upright-seated position if awake and alert.
5. Place patient in the left lateral decubitus position if obtunded.
6. Before insertion, the proper length of tubing to be passed should be measured from the mouth, back to the ear, and down anterior to the chest and abdomen, beyond the point where any side ports on the tube would be beyond the level of the estimated lower esophageal sphincter (Figs. 75.2 and 75.3).
7. If the patient is still awake, insert the gastric tube to the level of the glottis, and encourage the patient to swallow.
8. Pass the tube to the stomach.
   (a) Coughing, airflow, or fog from the tube raises the concern for inadvertent tracheal positioning.
9. After the tube is inserted, it is essential to confirm that the distal end of the tube is in the stomach, by "popping" 5–10 mL of air into the tube while someone is listening with a stethoscope over the stomach.

Fig. 75.3 Measuring correct placement of the tube. Place the distal tip over the stomach and wrap the tube up behind the ear (usually the right ear because the tubes generally pass easier through the right nares) and around the nares. The black line or centimeter mark at the level of the nares is the point of insertion when passage stops

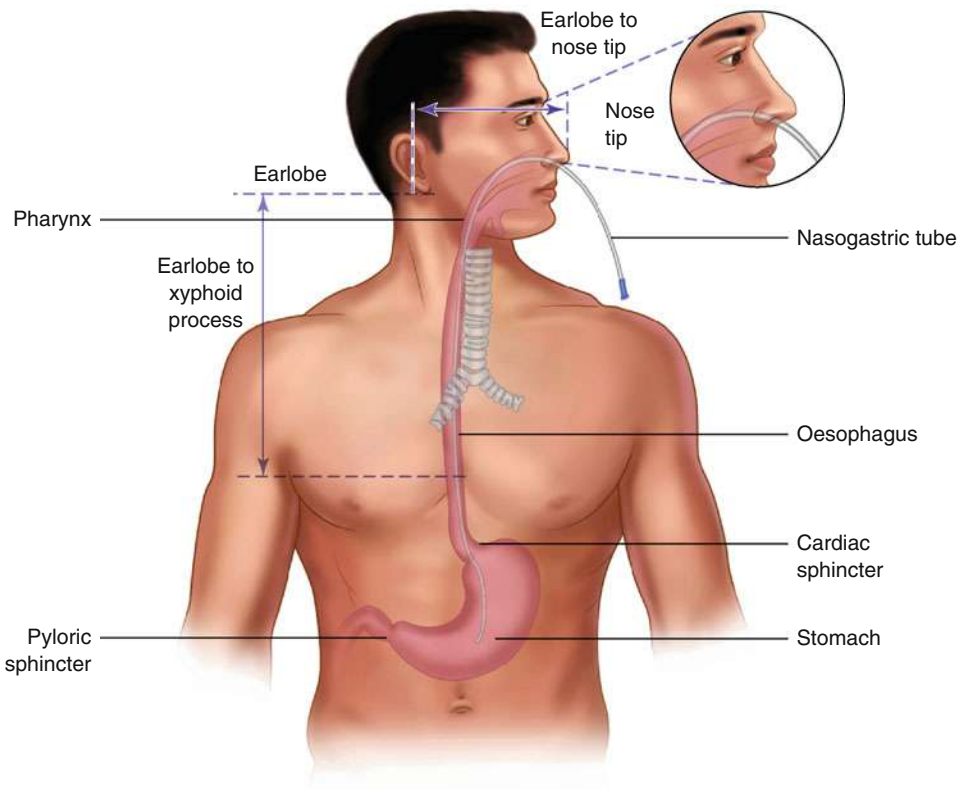

Fig. 75.2 Diagram illustrates appropriate placement of the lavage tube

10. In adults, 250-mL aliquots of a room temperature saline lavage solution are instilled via a funnel or lavage syringe. In children, aliquots should be 10–15 mL/kg to a maximum of 250 mL and suctioned back out of the tube attached to low to moderate continuous wall suction. Instillation of lavage solution and suction is repeated (Fig. 75.4).
11. Orogastric lavage should continue for at least several liters in an adult and/or at least 0.5–1 L in a child if the return is free of debris or until no particulate matter returns and the effluent lavage solution is clear.
12. Those caring for the patient must remain protected at all times, using goggles, mask, gown, and gloves. If the ingested poison is toxic via pulmonary or skin absorption, isolate the ingestant immediately in a self-contained wall suction unit.
13. Any material still in the stomach should be withdrawn, and immediate instillation of the activated charcoal should be considered for large ingestions of xenobiotics known to be adsorbed by activated charcoal.

75.5 Complications

- Vomiting
- Esophageal tears or perforation after orogastric tube insertion
- Inadvertent tracheal intubation and/or airway trauma
- Aspiration pneumonitis

75.6 Pearls and Pitfalls

- Pearls
  - You must use a large-bore orogastric tube for maximal efficacy.
  - The left lateral decubitus position is recommended because the pylorus points upward in this orientation. This position theoretically helps prevent the xenobiotic from passing through the pylorus during the procedure.
- Pitfalls
  - Large drug packets, adherent masses of pills, and plant and mushroom fragments will not pass through a 40-French lavage tube.

Fig. 75.4 Lavage in progress

Selected Reading

Gude A, Hoegberg LCG. Techniques to prevent gastrointestinal absorption. In: Nelson LS, Lewin NA, Howland MA, et al., editors. Goldfrank's toxicologic emergencies. 8th ed. New York: McGraw-Hill; 2006. p. 91–3.

Olson KR. Poisoning & drug overdose. In: Olson KR, Anderson IB, Benowitz NL, et al., editors. Emergency evaluation and treatment. 6th ed. New York: McGraw-Hill; 2012.

Smilktein MJ. Techniques used to prevent gastrointestinal absorption of toxic compounds. In: Nelson LS, Lewin NA, Howland MA, et al., editors. Goldfrank's toxicologic emergencies. 7th ed. New York: McGraw-Hill; 2002. p. 46–8.

76 Whole-Bowel Irrigation

Judith K. Lucas

76.1 Indications

- Whole-bowel irrigation (WBI) should not be used routinely in the management of the poisoned patient (because there is no clinical proof it will change clinical outcome).
- Ingestion of significant amount of medications.
  - Not adsorbed by activated charcoal
    - Lead, lithium, arsenic, and zinc
    - Substantial amounts of iron (high morbidity and no other effective method to gastrointestinal decontamination)
  - Sustained-release medications or enteric-coated drugs
  - Disk batteries distal to the pylorus
  - Whole transdermal patches (fentanyl, clonidine, nicotine)
  - Drug concretions
  - Ingested packets of illicit drugs

76.2 Contraindications

- Absolute
  - Bowel obstruction
  - Bowel perforation
  - Ileus
  - Hemodynamic instability
  - Compromised or unprotected airway
  - Intractable vomiting
- Relative
  - Concurrent or recent administration of activated charcoal (may decrease the effectiveness of activated charcoal)

76.3 Materials and Medications

- Topical anesthesia, although not mandatory, will reduce the pain of nasogastric (NG) tube placement.
  - 10 % lidocaine spray
  - Lidocaine gel
- Small-bore (12-French) NG tube (Fig. 76.1).
- Tape for securing the NG tube.
- Reservoir or feeding bag used for NG tube feedings (Fig. 76.2).
- Intravenous pole.
- Bedside commode or toilet (Fig. 76.3).
- Polyethylene glycol-electrolyte solution (PEG-ES) (Fig. 76.4).
- Antiemetic.
  - No absolute indication for prophylactic use
  - May be helpful if vomiting ensues during infusion
    - Metoclopramide
      - Antiemetic
      - *Increases gastric motility*

J.K. Lucas, MD
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_76

Fig. 76.1 Nasogastric (NG) tube. Typically, the infusion of the lavage solution is too rapid to be taken orally, so an NG tube can be placed. Since the irrigation solution is of low viscosity, a small-bore NG tube should be used for comfort

Fig. 76.2 Bag from which the lavage solution will drain; it is similar to the bags used for gastrostomy tube feeding

Fig. 76.3 Almost always, the patient will need to be seated on or very near a portable commode, as once the irrigation solution starts to move through the bowels, defecation will occur rapidly

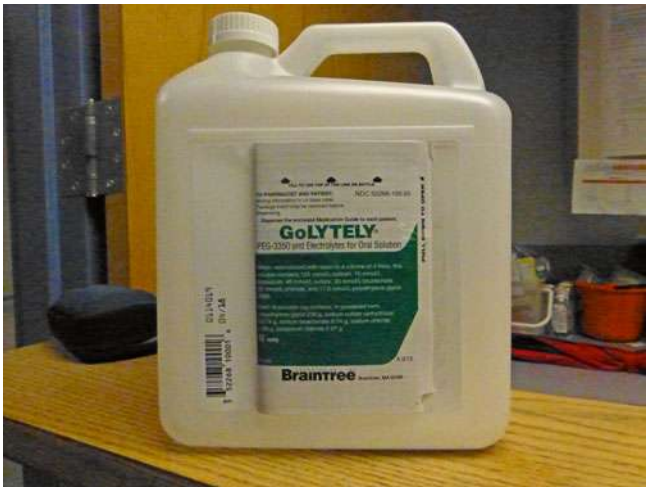

Fig. 76.4 Example brands of intestinal irrigation electrolyte solutions

76.4 Procedure

1. An NG tube is required because most patients will not drink the PEG-ES at the necessary rate.
2. Place a small-bore (12-French) NG tube to a sufficient distance that the tip lies in the central portion of the stomach.
3. Confirm NG placement with a radiograph.
4. Attach the tube to the reservoir bag of PEG-ES and hang from an elevated site (an extended intravenous pole).
5. The patient should be seated in an upright position.
   - Promotes settling of the intoxicant in the distal portion of the stomach
   - Decreases the likelihood of vomiting
6. Dosing:
   - Children 9 months to 6 years: 500 mL/h
   - Children 6–12 years: 1,000 mL/h
   - Adolescents/adults: 1,500–2,000 mL/h
7. Collect effluent.
8. Continue infusion.
   - Until the rectal effluent is the same color as the influent (i.e., clear), usually between 4 and 6 h.
   - You may continue beyond clear effluent if clinical evidence indicates ongoing effectiveness:
     – Continued pill fragments or drug packets are present in the effluent.
     – Radiographic evidence that pills, pharmacobezoars, or packets are still present.

76.5 Complications

- Nausea, vomiting, and bloating
- Misplacement of the NG tube
- Esophageal perforation owing to NG tube placement
- Aspiration pneumonitis in the unprotected airway

76.6 Pearls

- Overall, WBI is probably more effective than gastric lavage, but probably less effective than activated charcoal in preventing poison absorption (when the intoxicant can be adsorbed to charcoal).
- Vomiting.
  – Usually secondary to the ingestant (i.e., emetogenic toxins, such as iron)
  – May be due to rate of infusion
    - Slow rate by 50 % for 30–60 min.
    - Then return to original rate.
- If resistance is encountered during NG tube placement, do not force passage. Remove and redirect.

Selected Reading

Bailey B. To decontaminate or not to decontaminate? the balance between potential risks and foreseeable benefits. Clin Pediatr Emerg Med. 2008;9:17–23.

Hanhan UA. The poisoned child in the pediatric intensive care unit. Pediatr Clin North Am. 2008;55:669–86. xi.

Lheureux P, Tenenbein M. Position paper: whole bowel irrigation. American Academy of Clinical Toxicology/European Association of Poison Centres and Clinical Toxicologists. J Toxicol Clin Toxicol. 2004;42:843–54.

Othong R. Whole-bowel irrigation. MedScape Reference: drugs, diseases, and procedures. Updated: Aug 2011

Postuma R. Whole bowel irrigation in pediatric patients. J Pediatr Surg. 1982;17:350–2.

Sengstaken-Blakemore Tube

77

Thomas T. Nguyen, Etan Eitches,
and Stephanie Wetmore-Nguyen

77.1 Indications

- Life-threatening esophageal variceal bleed refractory to endoscopy and medical therapy
- Life-threatening esophageal variceal bleed refractory to medical therapy in the absence of possible endoscopy

77.2 Contraindications

- Absolute
  - Known esophageal rupture
  - Unable to intubate or maintain airway
- Relative
  - History of prior esophageal trauma or strictures
  - Recent surgery of the gastroesophageal junction
  - Resolved or resolving variceal bleeding

77.3 Materials and Medications

- Sengstaken-Blakemore (SB) tube (Fig. 77.1)
- 60-mL syringe with catheter tip (Fig. 77.2)
- Sphygmomanometer (Fig. 77.3) or cuffalator (Fig. 77.4)
- Y-Tube connector (Fig. 77.5) or 3-way stop-valve connector (Fig. 77.6)
- Vacuum suction device and tubing (Fig. 77.7)
- Tube clamps (4) (Fig. 77.8)
- Lubricant (water soluble)
- Lidocaine (Xylocaine) spray or gel
- Anchoring device such as a football helmet or catcher's mask (Fig. 77.9)
- Cup of water and straw if the patient is awake
- Scissors (Fig. 77.10)
- Intubation equipment
- Sterile water

T.T. Nguyen, MD (✉)
Department of Emergency Medicine,
Mount Sinai Beth Israel, New York, NY, USA
e-mail: tnguyen@chpnet.org

E. Eitches, MD
Department of Emergency Medicine,
Beth Israel Medical Center, New York, NY, USA
e-mail: eeitches@gmail.com

S. Wetmore-Nguyen, MD
Department of Emergency Medicine,
New York-Presbyterian Hospital/Columbia University Medical Center,
New York, NY, USA
e-mail: slw9004@nyp.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_77

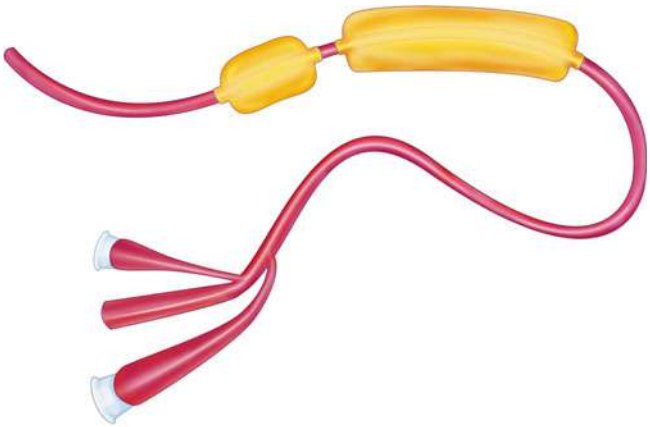

Fig. 77.1 Sengstaken-Blakemore (SB) tube

Fig. 77.2 Syringe

Fig. 77.3 Sphygmomanometer

Fig. 77.4 Cuffalator

Fig. 77.5 Y-Tube connector

Fig. 77.6 3-way stop-valve tube connector

Fig. 77.7 Suction device and tubing

Fig. 77.8 Tube clamps

Fig. 77.9 Helmet traction setup

Fig. 77.10 Scissors

77.4 Procedure

1. Sedate and/or intubate the patient for adequate control of the patient during the procedure.
2. Ensure that the SB tube balloons are functional by inflating and deflating the balloons to ensure the absence of leaks.
3. Perform gastric lavage and irrigate the stomach with copious amount of sterile water.
4. Coat the distal and proximal portions of the SB tube with a thin layer of lubricating jelly or lidocaine gel. Spray the nasal passage with lidocaine spray.
5. Pass the SB tube via the nasogastric (NG) or the orogastric route (in intubated patients) to the 50-cm line. You may confirm placement with x-ray.
6. Inflate the gastric balloon to 200 mL of air and clamp the tube.
7. Apply gentle traction of 1–2 lb (0.4–2 kg) of force until it is felt that the gastric balloon has lodged at the gastroesophageal junction (Fig. 77.11).
8. Secure the tube to an anchor (e.g., football helmet or catcher's mask) placed on the patient's head.
9. Aspirate and lavage the gastric aspiration port. If it is clear of blood, do not inflate the esophageal balloon.
10. If it is not clear of blood, connect the esophageal tube of the SB tube to the sphygmomanometer/cuffalator using the 3-way stop-valve device (Fig. 77.12). You may use the Y-Tube connector instead.
11. Inflate the esophageal balloon to the lowest pressure determined to stop bleeding, typically 20–45 mmHg. Clamp the balloon.
12. Place an NG tube until it is felt overlying the top of the esophageal balloon of the SB tube. Check for further proximal esophageal bleed through aspiration and gentle lavage. Attach this NG tube to intermittent section to aid in the clearance of secretions.
13. Obtain a portable radiograph to confirm the position of the SB tube.
14. The esophageal tube should be at the lowest pressure that prevents bleeding and kept inflated for 24 h or until other definitive treatment is obtained.

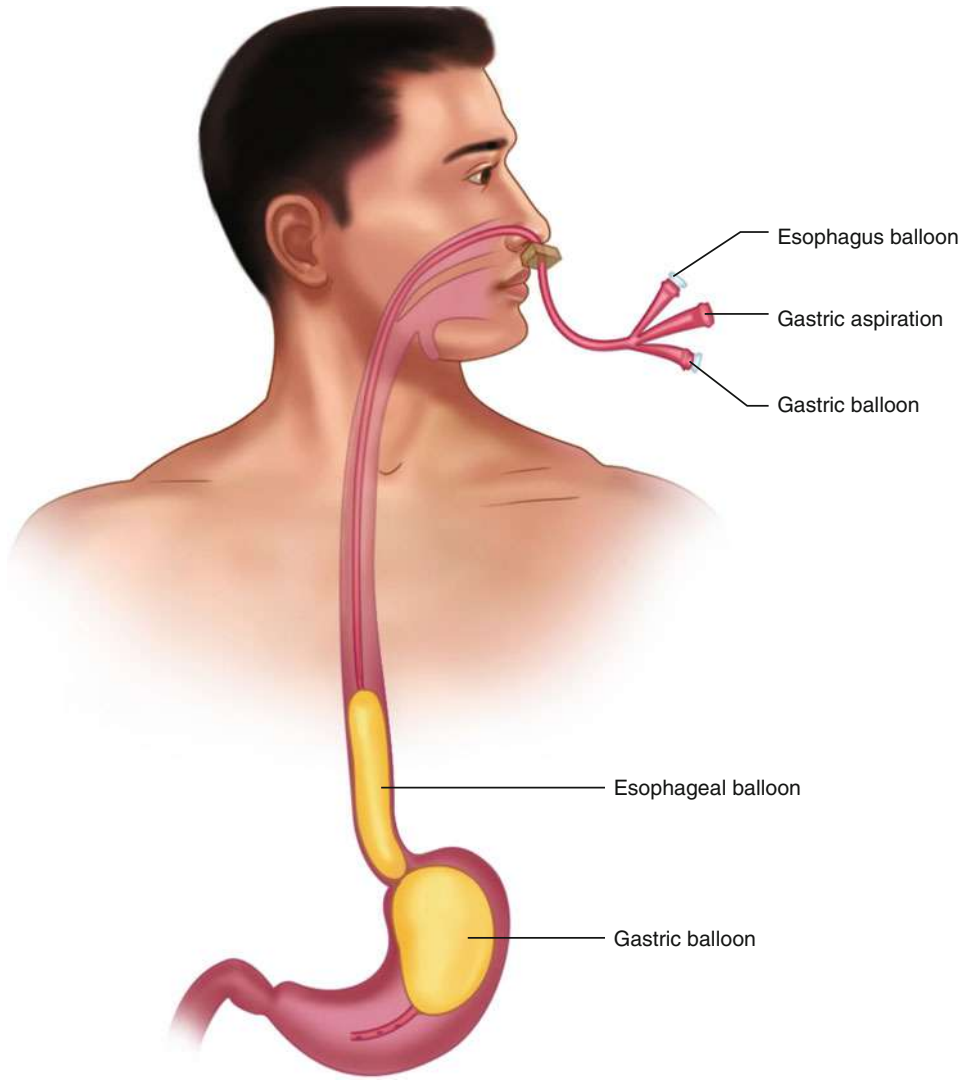

Fig. 77.11 Proper placement of the balloons

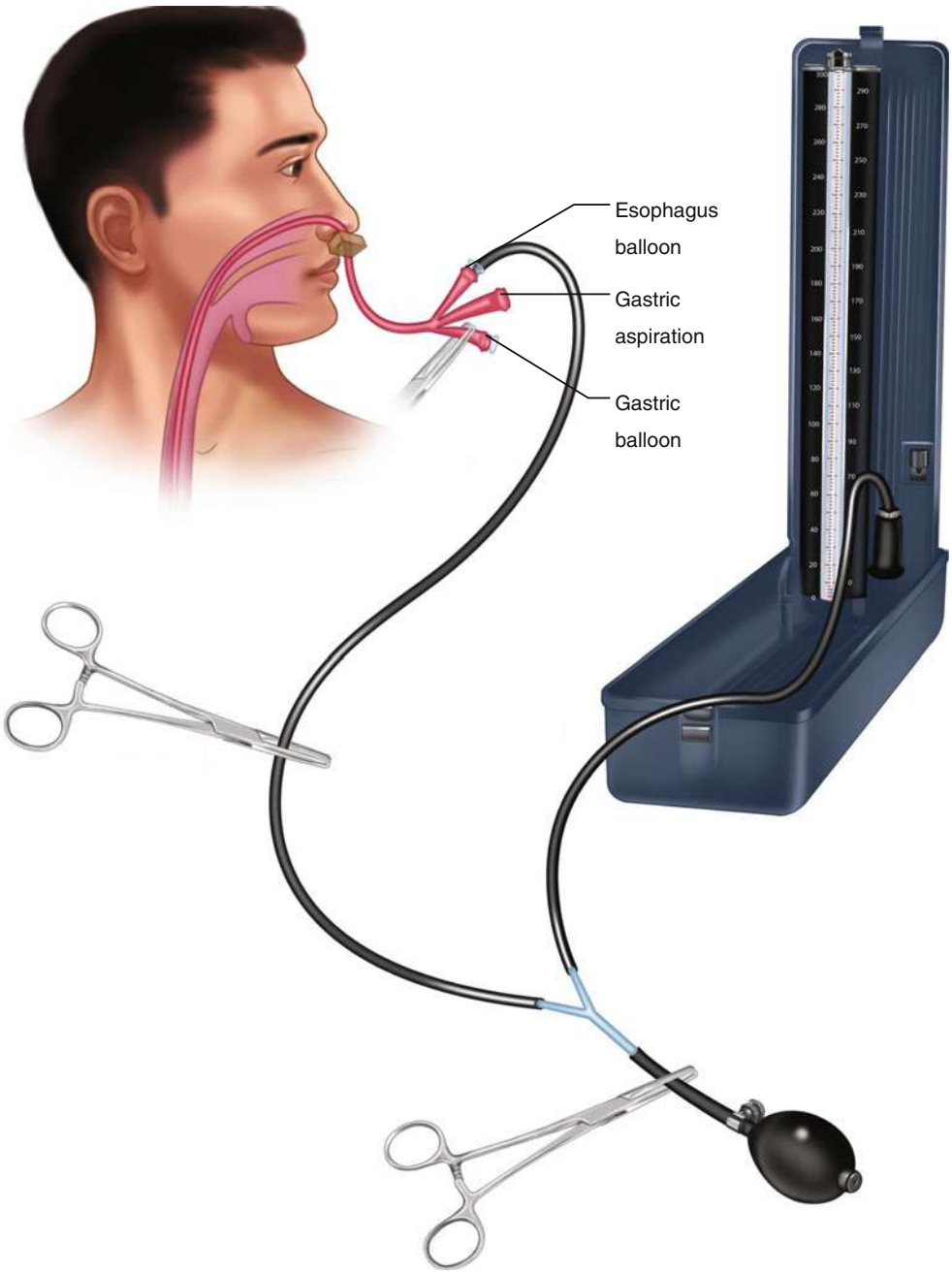

Fig. 77.12 Connection of the sphygmomanometer to the esophagus balloon port

77.5 Complications

- Esophageal rupture occurs owing to esophageal erosion and necrosis owing to a balloon tamponade effect on tissue perfusion or overzealous balloon inflation.
- Airway obstruction owing to gastric tube deflation or failure, allowing esophageal tube to move up and occlude airway. Keep scissors near the patient to cut the SB tube lumens and remove the tube as necessary.
- Regurgitation and aspiration pneumonia from failure to adequately suction oropharyngeal secretions.

77.6 Pearls and Pitfalls

- Pearls
  - The esophageal balloon should not be inflated if the gastric balloon alone stops the bleeding.
  - Never inflate the esophageal balloon without the inflating the gastric balloon first. This will prevent it from slipping proximally into the oropharynx and obstructing the airway.
  - Nausea, vomiting, or aspiration is highly likely to occur. Use antiemetics and lavage the stomach before the procedure.
  - Intubate if there is airway compromise or risk of aspiration into the lungs.
  - Inflate the esophageal balloon only to the minimum pressure necessary to stop the variceal bleeding.
  - Using a catcher's mask may be more practical and comfortable for the recumbent patient.
- Pitfalls
  - The SB tube may induce hiccups.

Selected Reading

Bauer J, Kreel I, Kark A. The use of the Sengstaken-Blakemore tube for immediate control of bleeding esophageal varices. Ann Surg. 1974;179:273–7.

Henneman PL. Gastrointestinal bleeding. In: Rosen P, Barkin RM, editors. Emergency medicine. 6th ed. St. Louis: Mosby; 1998.

Remonda G, Morachioli N, Petruzzelli C. The use of the Sengstaken-Blakemore tube for immediate control of bleeding esophageal varices. Ann Osp Maria Vittoria Torino. 1981;24:115–20.

Sengstaken RW, Blakemore AH. Balloon tamponage for the control of hemorrhage from esophageal varices. Ann Surg. 1950;131:781–9.

Treger R, Graham T, Dea S. Sengstaken-Blakemore tube. Available at http://emedicine.medscape.com/article/81020-overview#a01. Accessed 18 May 2014.

78 Gastrostomy Tube Placement

Nathaniel Lisenbee and Latha Ganti

There are several types of gastrostomy tubes and related procedure variations:

- Procedure variations:
  - Open gastrostomy tube (G tube)
  - Percutaneous endoscopic gastrostomy (PEG) tube
  - Laparoscopic G tube
- Tube types:
  - PEG tube (Fig. 78.1)
  - Malecot tube (Fig. 78.2)
  - Balloon G tube (Fig. 78.3)
  - Low-profile G tube (nonobturated, button) (Fig. 78.4)
  - Low-profile G tube (obturated, button) (Fig. 78.5)

N. Lisenbee, MD
U. S. Air Force Medical Center Keesler, Biloxi, MI, USA
e-mail: nate4085@ufl.edu

L. Ganti, MD (✉)
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_78

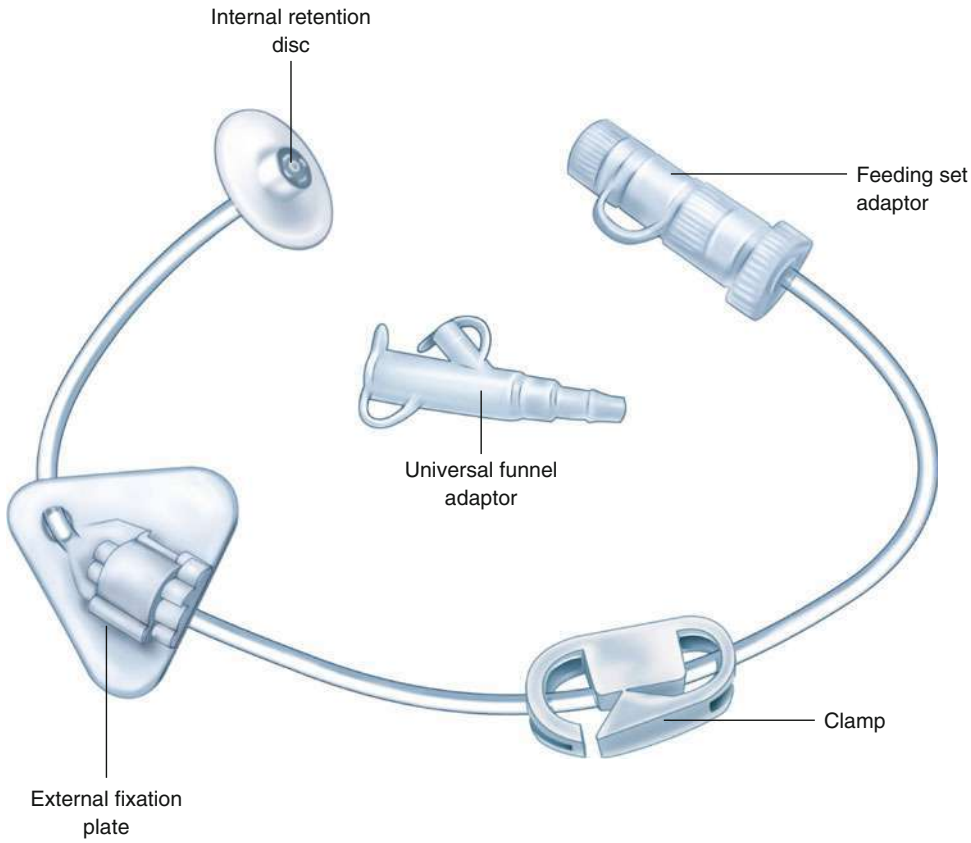

Fig. 78.1 Percutaneous endoscopic gastrostomy (PEG) tube

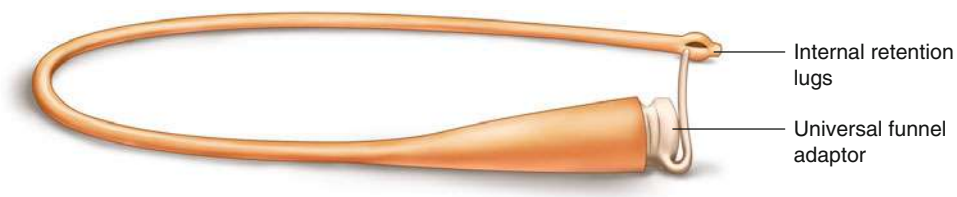

Fig. 78.2 Malecot tube

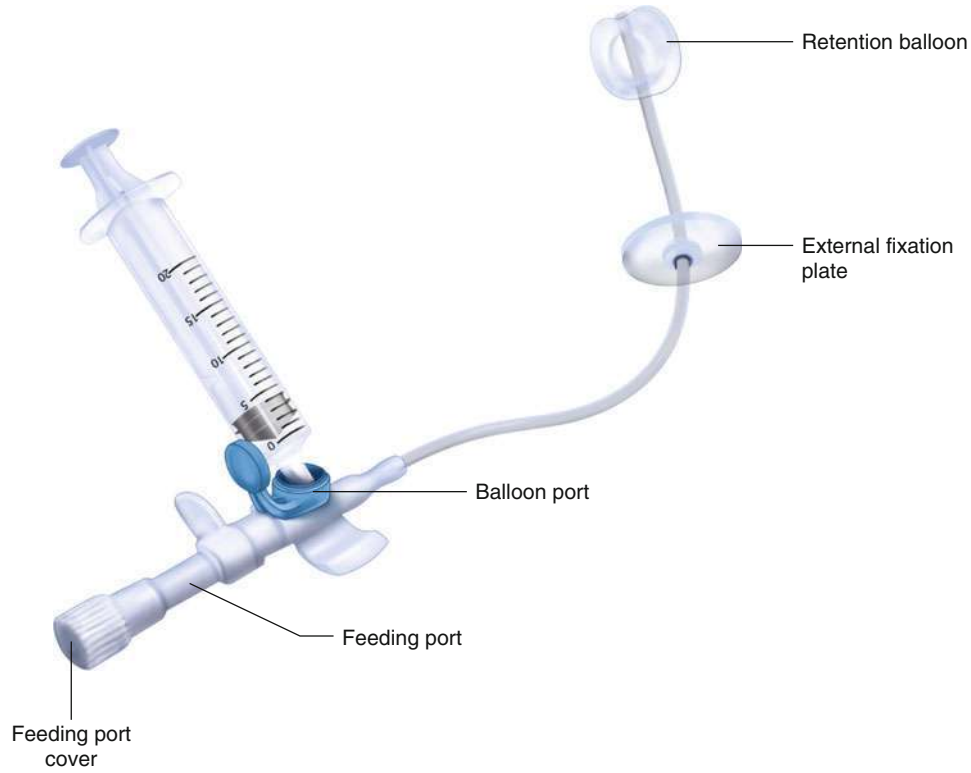

Fig. 78.3 Balloon gastrostomy tube

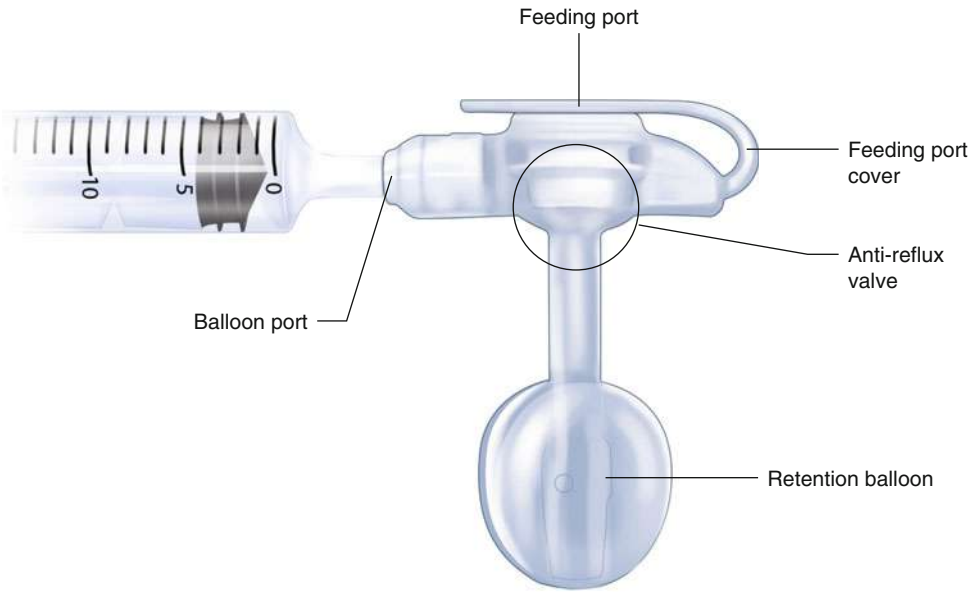

Fig. 78.4 Low-profile gastrostomy tube (nonobturated, button)

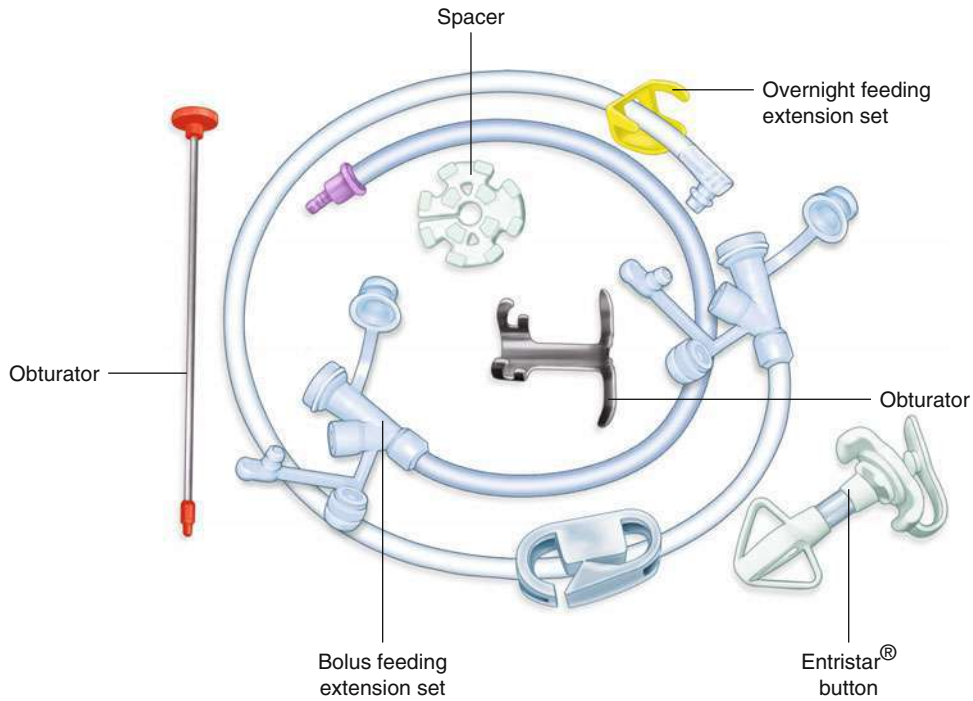

Fig. 78.5 Low-profile gastrostomy tube (obturated, button)

78.1 Indications

- Inability to swallow
  - Neurological deficit
  - Head trauma
  - Facial burns
  - Decreased mental status
- Need for gastric decompression
  - Gastric outflow obstruction
  - Small bowel obstruction (SBO), ileus, or volvulus
  - Intra-abdominal malignancy

78.2 Contraindications

- Absolute
  - Peritonitis
  - Ascites
- Relative
  - Hemodynamic instability
  - Coagulopathy
  - Abdominal wall infection at surgical site
  - History of gastric resection
  - Portal hypertension
  - Gastric varices

78.3 Materials and Medications

- Materials
  - Gloves
  - Lubricant
  - G tube (commercial kit) or Foley catheter
  - Sterile saline
  - External bolster
  - Multiple syringes
  - Suture material
  - Needle driver
  - Scissors
  - Stethoscope
- Medications
  - PEG tube placement
    - Moderate sedation (e.g., propofol, midazolam, fentanyl)
    - Local anesthesia (e.g., lidocaine, bupivacaine)
  - Open or laparoscopic G tube placement (not an emergency department procedure)
    - General anesthesia

78.4 Procedure (PEG Tube Replacement)

1. If the G tube is only partially removed upon patient presentation, the tube must first be removed.
2. Remove the G tube by deflating the balloon and pull gently on the tube while applying pressure to the abdominal wall at the surgical site.
3. The G tube should slide out easily with gentle traction, and the procedure should be discontinued if it does not.
4. Initially, it is important to assess the tract to determine the size and potential need for dilation.
5. If necessary for tube passage, the tract can be dilated with a cotton-tipped applicator or hemostat; however, be sure to dilate gently because it is possible to create a false tract.
6. Once the tract has been assessed, obtain the appropriate tube for replacement.
7. Initially, attempt replacement of the patient's G tube with an identical tube.
8. If an identical tube is not available, attempt placing a small Foley catheter in the tract to ensure that it stays open.
9. To initiate placement, place lubricant on the tube and carefully advance the tube into the tract.
10. Once the tube is in place, secure an air-filled syringe to the tip of the tube and insufflate a small amount of air into the stomach while auscultating to confirm passage of air into the stomach.
11. Secondary procedural confirmation can be performed by aspirating gastric contents from the tube.
12. Once placement is verified, do not forget to inflate the balloon with saline and then pull the tube backward until it abuts the inside of the stomach wall (Fig. 78.6).
13. Finally, it is very important to secure the external portion of the tube to prevent the tube from being lost into the stomach.
14. Commercial G tubes are accompanied by a bolster made specifically for the specific type of G tube in order to provide security of tube placement.

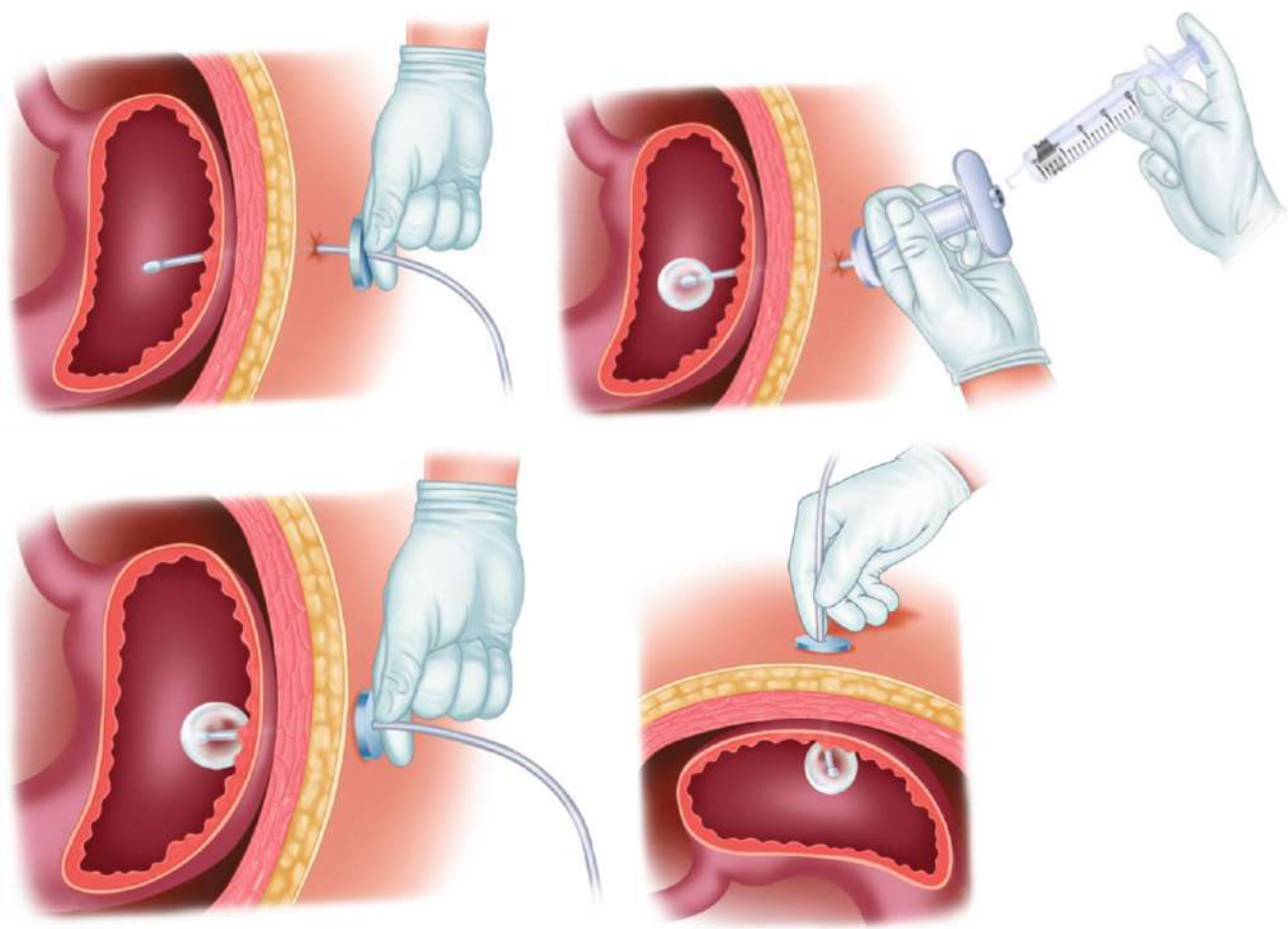

Fig. 78.6 Placement of low-profile gastrostomy tube

15. If a Foley catheter is used to maintain tract patency, an external bolster must be created using the following steps:
    - Trim a 2- to 3-in. portion from the tip of the catheter.
    - Cut two small holes just above each other on the 2- to 3-in. trimmed catheter portion.
    - Slide the external portion of the Foley catheter through both holes of the bolster.
    - Slide the bolster down the Foley catheter to the abdominal wall.
    - Secure the two ends of the bolster to the skin in order to maintain patency of the Foley catheter.
    - An interventional radiologist may also be contacted to advance the tube over a wire under fluoroscopic guidance.
16. Radiographic confirmation.
    - Typically, G tube placement is confirmed by injecting approximately 20 mL of Gastrografin® (diatrizoate meglumine, diatrizoate sodium) water-soluble contrast into the G tube, followed shortly by an abdominal x-ray.
    - Proper placement will result in an abdominal radiograph showing contrast outlining the stomach (Fig. 78.7).

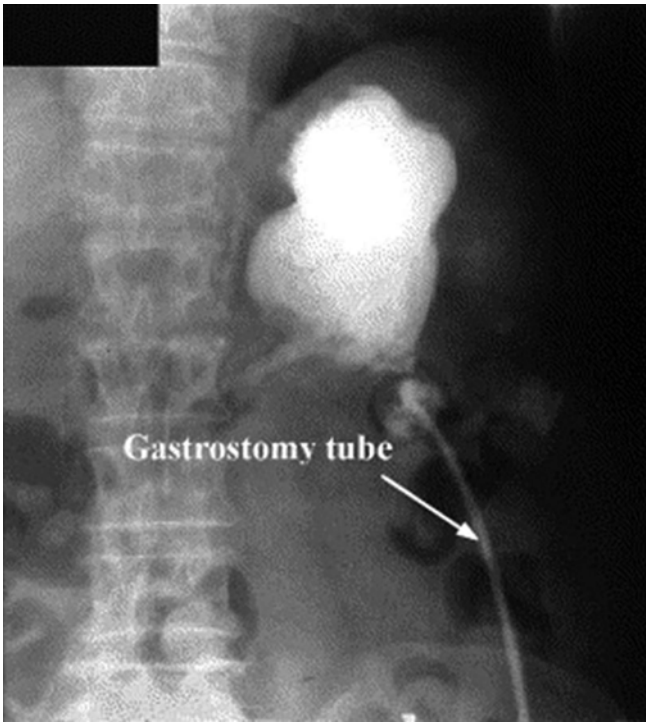

Fig. 78.7 An anteroposterior abdominal radiograph after PEG tube placement and injection of 25 mL of gastrografin. The tube can be seen projecting up the left side of the abdomen, and contrast medium appears to enter the stomach lumen. The balloon is not visualized in the stomach (Reproduced with permission from: Burke DT, El Shami A, Heinle E, and Pina BD. Comparison of gastrostomy tube replacement verification using air insufflation versus gastrografin. Archives of physical medicine and rehabilitation. 2006;87(11): 1530–3.)

78.5 Pearls and Pitfalls

- Pearls
  - Even if a patient's G tube is only partially removed upon presentation to the emergency department, the tube likely needs complete removal and replacement.
  - Most G tubes are easily removed at the bedside; however, some are not able to be safely removed in the emergency department setting. Thus, if the tube does not withdraw easily, attempt to contact the proceduralist who placed the tube in order to inquire about the best method for removal.
- Pitfalls
  - Never use barium contrast when confirming G tube placement radiographically because barium can cause significant intra-abdominal damage if accidentally injected in the intraperitoneal cavity.

78.6 Complications

- Aspiration
- Surgical site infection
- Bleeding
- Pneumoperitoneum
- Accidental perforation of the colon or small bowel
- Tube dislodgment
- Peritonitis

Selected Reading

Arora G. Medscape: percutaneous endoscopic gastrostomy tube placement. Retrieved 3 Jan 2013, from http://emedicine.medscape.com/article/149665-overview#a09.

Gauderer MW, Ponsky JL, Izant Jr RJ. Gastrostomy without laparotomy: a percutaneous endoscopic technique. J Pediatr Surg. 1980; 15:872–5.

Great Ormand street hospital for children website. Retrieved 3 Jan 2013, from www.gosh.nhs.uk/EasySiteWeb/GatewayLink.aspx?alId=102263.

Sarani B. Percutaneous endoscopic gastrostomy tube placement. In: Falter F, editor. Bedside procedures in the ICU. Philadelphia: Springer; 2012. p. 113–22.

Tawa Jr NE, Fischer JE. Metabolism in surgical patients. In: Townsend Jr CM, Beauchamp RD, Evers BM, Mattox KL, editors. Sabiston textbook of surgery. 18th ed. Philadelphia: Elsevier; 2007.

79 Paracentesis

Shalu S. Patel and Bobby K. Desai

79.1 Indications

- Diagnosis of infection in ascites
- Diagnosis of malignant ascites
- Diagnosis of hemoperitoneum in traumas
- Relief of abdominal pressure/pain or respiratory compromise secondary to ascites

79.2 Contraindications

- Severe coagulopathy
  - Prothrombin time (PT) >21 s
  - International normalized ratio (INR) >1.6
  - Platelets <50,000/mm^3
- Skin infection over the needle insertion site
- Acute abdomen that requires surgery
- Pregnancy
- Distended bowel
- Intra-abdominal adhesions

79.3 Materials and Medications

- 18- to 22-gauge 1.5- to 3.5-in. needle or angiocatheter, 25-gauge needle
- Lidocaine 1 or 2 % (10 mL)
- Syringes
  - 10 mL (1), 50 mL (2)
- 1-L vacuum bottle (4) (if therapeutic tap)
- Thoracentesis kit tubing or any high-pressure connection tubing (if therapeutic tap)
- Sterile gloves
- Surgical pen (recommended)
- Povidone-iodine (Betadine) or other skin antiseptic
- Sterile drape
- Sterile gauze (4×4)
- Band-Aid
- Bedside ultrasound (recommended)

79.4 Procedure (Fig. 79.1)

1. Position the patient supine. If possible, adjust the head of the bed to make a 45° angle to help the fluid accumulate in the pocket. Sometimes, it may also be beneficial to have the patient lie recumbent toward the site of drainage.
2. Scan the abdomen with an ultrasound to determine whether there is a pocket of fluid that can be drained. This also allows the physician to see how far the needle needs to be inserted and how deep it can be placed without risking injury to the bowel (Fig. 79.2).
3. Mark the optimal needle insertion site with a surgical pen.
4. Prepare the skin and drape in a sterile fashion.
5. Using lidocaine, anesthetize the appropriate area subcutaneously and then continue to insert the needle, and inject anesthetic through the deeper tissues until ascitic fluid can be drawn back.
6. Withdraw the needle.
7. When ready for the paracentesis, stretch the skin caudad and insert the needle or angiocatheter (connected to a syringe) while aspirating. Then, release the skin and continue to insert the needle or angiocatheter through the peritoneal wall until fluid is retrieved. This will create a "Z-track" that will decrease leakage of peritoneal fluid through the skin.

S.S. Patel, MD
Department of Emergency Medicine, Florida Hospital Tampa, Florida Hospital Carrollwood, Tampa, FL, USA
e-mail: shalu314@gmail.com

B.K. Desai, MD (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_79

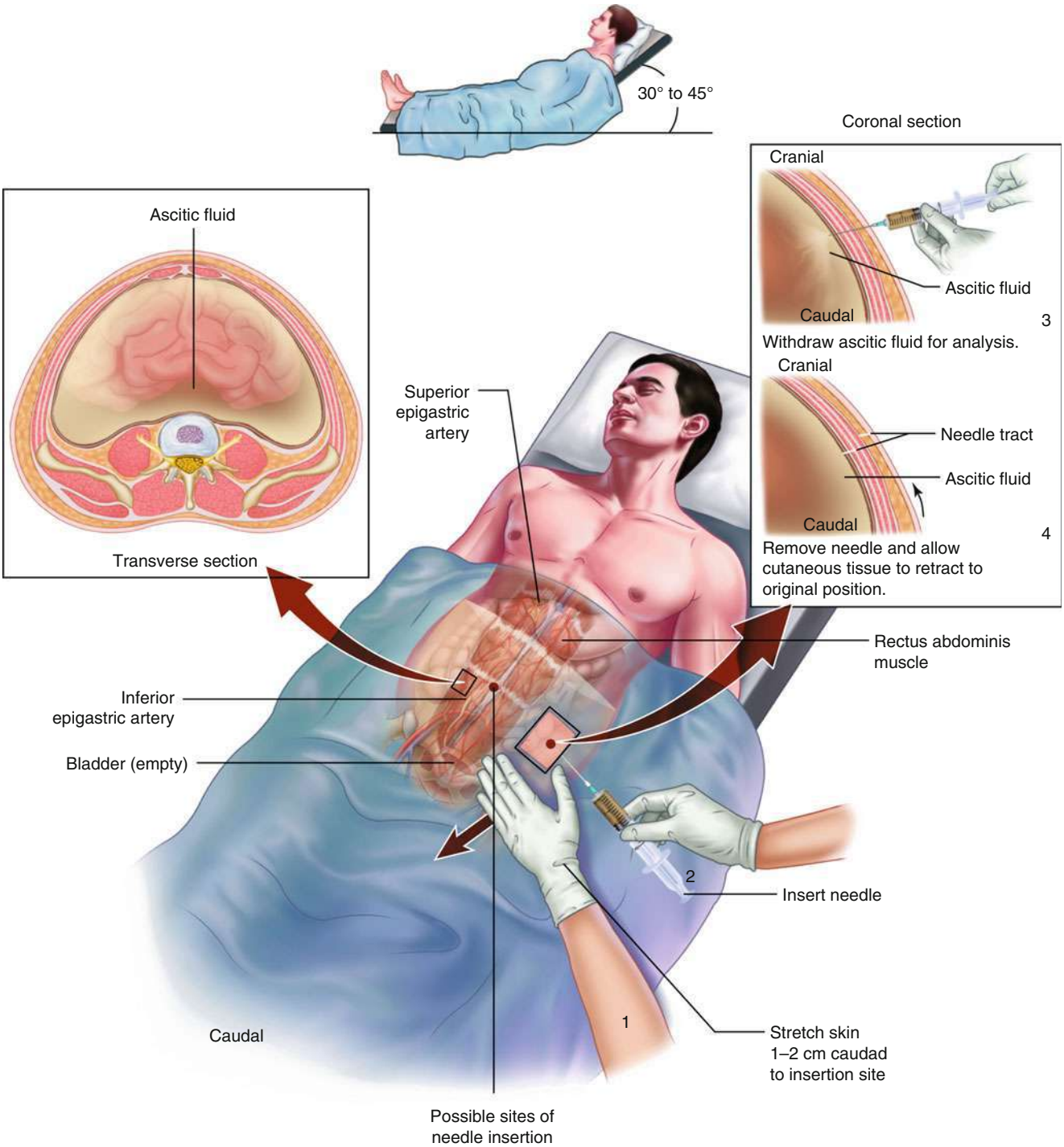

Fig. 79.1 Paracentesis procedure

8. Once fluid is retrieved, push in the catheter and remove the needle portion (if used) or hold the needle steady.
9. Aspirate from the catheter to ensure that it is in the appropriate location.
10. If fluid easily is aspirated, unscrew the syringe and connect a 50-mL syringe to the needle or catheter and fill it with fluid. This may be done twice. Alternatively, if the procedure is done for therapeutic purposes, attach the tubing that is already connected to the vacuum bottle to the catheter and allow the vacuum to withdraw fluid into the collection bottles.
11. If fluid cannot be aspirated easily, the catheter can be repositioned further in the pocket or turned by 45° sequentially as needed.

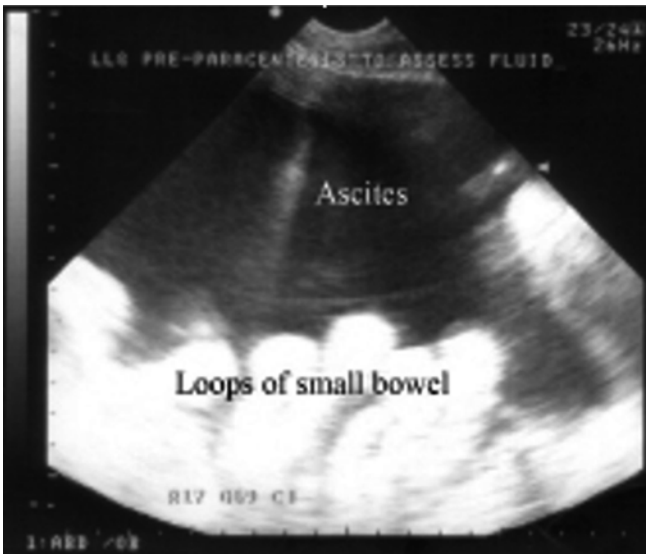

Fig. 79.2 Ultrasound to determine whether there is a pocket of fluid that can be drained

12. Once the fluid is aspirated, pull out the needle or angiocatheter and hold pressure with gauze. Bleeding should be minimal.
13. Place a Band-Aid or other dressing over the site.
14. Send the fluid to the laboratory. Generally, laboratory analyses include protein, albumin, specific gravity, glucose, bilirubin, amylase, lipase, triglyceride, lactate dehydrogenase (LDH), cell count and differential, culture and sensitivity (C&S), Gram stain, acid-fast bacillus (AFB), fungal culture, cytology, and pH.

79.5 Complications

- Persistent leakage from the needle insertion site
- Abdominal wall hematoma
- Bowel perforation
- Introduction of infection
- Hypotension (after a large-volume paracentesis)
- Dilutional hyponatremia
- Hepatorenal syndrome
- Bleeding
- Postparacentesis circulatory dysfunction

79.6 Pearls and Pitfalls

- Pearls
  - The preferred site of entry is in the midline of the abdomen, below the umbilicus.
  - The serum-ascites albumin gradient (SAAG) can be used to identify the cause of the ascites. It is calculated by subtracting the albumin concentration in the ascites from the albumin concentration in the serum. A high gradient (>1.1 g/dL) suggests portal hypertension, whereas a low gradient (<1.1 g/dL) suggests other causes.
  - Postparacentesis circulatory dysfunction (PPCD) occurs secondary to hypovolemia after large-volume paracentesis (>4 L) in cirrhotic patients. It is associated with worsening hyponatremia, renal dysfunction, shorter time to ascites recurrence, and increased mortality. Prevention of PPCD has been demonstrated with the administration of 6–8 g of albumin per liter of ascites removed.
- Pitfalls
  - Polymorphonuclear lymphocyte (PMN) count greater than 250/mm^3 is diagnostic of spontaneous bacterial peritonitis.

Selected Reading

Ginès P, Cárdenas A, Arroyo V, Rodés J. Management of cirrhosis and ascites. N Engl J Med. 2004;350:1646–54.

Ginès P, Tito L, Arroyo V, et al. Randomized comparative study of therapeutic paracentesis with and without intravenous albumin in cirrhosis. Gastroenterology. 1988;94:1493–502.

Ruiz del Arbol L, Monescillo A, Jimenéz W, et al. Paracentesis-induced circulatory dysfunction: mechanism and effect on hepatic hemodynamics in cirrhosis. Gastroenterology. 1997;113:579–86.

Runyon BA. Paracentesis of ascitic fluid. A safe procedure. Arch Intern Med. 1986;146:2259–61.

Wong CL, Holroyd-Leduc J, Thorpe KE, Straus SE. Does this patient have bacterial peritonitis or portal hypertension? How do I perform a paracentesis and analyze the results? JAMA. 2008;299:1166–78.

Anal Fissure Management

80

David P. Nguyen, L. Connor Nickels, and Giuliano De Portu

80.1 Indications

- An anal fissure is a small ulcer of the mucosa at the anal verge (Fig. 80.1).
- It is the most common cause of intense sudden rectal bleeding.
- Posterior midline anal fissures are the most common type (90 %).
  - Mostly found in young adults (30–50 y) but can occur at any age.
  - Usually associated with constipation (firm, large-caliber, painful bowel movements) or chronic diarrhea.
  - Most uncomplicated fissures resolve in 3–4 weeks.
- Can be extremely painful, during and after defecation.
- Classified as acute or chronic.
- Now believed to be caused by reduced anal blood flow in the posterior midline, anal sphincter hypertonia, and thus mucosal ischemia.

D.P. Nguyen, DO
Department of Emergency Medicine, Rush-Copley Medical Center, Aurora, IL, USA
e-mail: davidpnguyen@yahoo.com

L.C. Nickels, MD, RDMS (✉) • G. De Portu, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; gdeportu@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_80

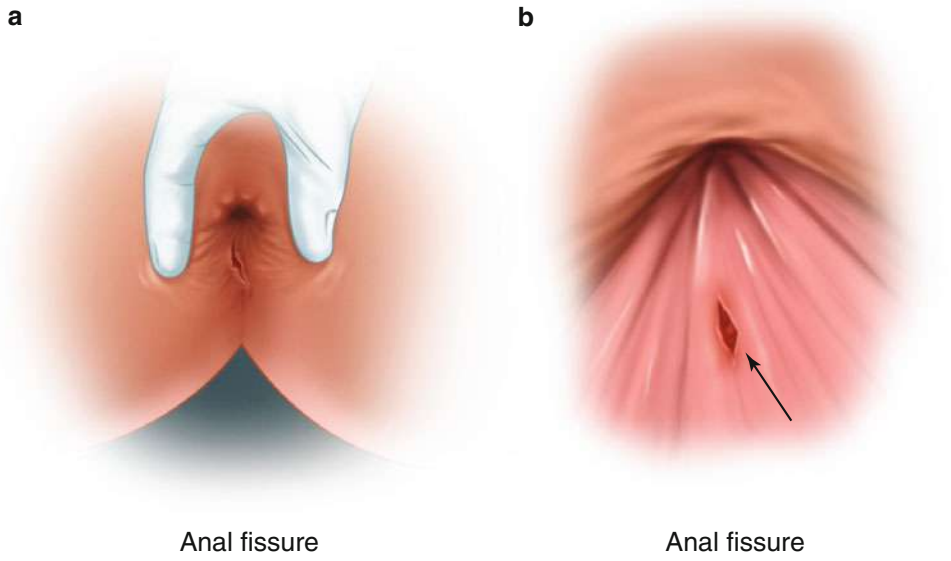

Fig. 80.1 (**a**, **b**) Anal fissures

80.2 Contraindications

- Digital rectal examination should be avoided unless the diagnosis is in doubt.
- Surgical procedures are generally reserved for when medical management has failed after 1–3 months of treatment.

80.3 Materials and Medications

- Standard precautions barrier protection for the provider.
- Good light source.
- Optional emergency department treatments:
  - Topical anesthetic/preparation (Anusol [pramoxine hydrochloride; zinc oxide] with cortisone).
  - Nitroglycerin (0.2 %) or nifedipine gel (2 %) is second-line therapy (relaxes muscles and promotes blood flow).

80.4 Procedure

1. In a private, calm environment, gently spread the buttocks for complete visual inspection.
   - This may cause an increase in the patient's pain and spasming.
   - If a fissure is clearly identified, stop here.
2. Apply topical anesthetic/preparation for symptomatic relief (optional, as the physician may want to just start with the treatments that follow).
3. Discharge the patient with conservative therapy management.
4. In acute anal fissures (onset of 3–6 weeks), medical management is indicated along with dietary modifications (WASH regimen [warm baths, analgesia, stool softeners, high-fiber diet]):
   - Warm sitz baths.
     - Usually 20 min soaking each time
     - Recommended after every bowel movement
     - At least twice per day if not having regular bowel movements
   - High-fiber diet with fiber supplements.
   - Increase fluid intake.
   - May add stool softeners, if needed.
   - If chronic or the previous regimen has been exhausted, one of the following may be considered:
     - 0.2–0.4 % nitroglycerin cream applied to anal area
       - May cause headache
       - Recommend wearing a glove to prevent absorption through digital skin
     - Calcium channel blockers:
       - Topical nifedipine
       - 2 % diltiazem cream
     - Botulinum toxin A injection:
       - Controversial; may have poorer success rates than surgery
5. Provide surgical referral for nonhealing wounds.
   - Lateral internal sphincterotomy is the surgical procedure of choice.

80.5 Complications

- Infection
- Abscess
- Bleeding
- Chronic fissure formation
- Postsurgical fecal incontinence

80.6 Pearls and Pitfalls

- Pearls
  - Multiple or recurrent fissures are associated with Crohn's disease, tuberculosis, syphilis, human immunodeficiency virus (HIV), and malignancy.
- Pitfalls
  - Suspect child abuse if an anal fissure is found in a child.

Selected Reading

Feldman M, Friedman LS, Brandt LJ, editors. Sleisenger and Fordtran's gastrointestinal and liver disease. 9th ed. Philadelphia: Saunders; 2010.

Marx J, Hockberger R, Walls R, editors. Rosen's emergency medicine. 7th ed. Philadelphia: Mosby; 2010. Chapter 94: Disorders of the anorectum

Oztürk H, Onen A, Dokucu AI, Otçu S, Yağmur Y, Yucesan S. Management of anorectal injuries in children: an eighteen-year experience. Eur J Pediatr Surg. 2003;13:249–55.

Townsend CM Jr, Beauchamp RD, Evers BM, Mattox KL, editors. Sabiston textbook of surgery. 19th ed. Philadelphia: Saunders; 2012. Chapter 53: Anus

Part XI

Genitourinary Procedures

81 Bladder Catheterization

Maritza A. Plaza-Verduin and Judith K. Lucas

81.1 Indications

- Obtaining a sterile urine specimen
- Preventing or relieving urinary retention
- Close monitoring of urine output for fluid balance with an indwelling urinary catheter
- Urgent cystourethrography
- Child with contusion or burns to the perineum and at risk for meatal swelling and obstruction to urine outflow
- Temporary measure to relieve lower urinary tract obstruction
- Neurogenic bladder
- Anesthesia-induced and/or surgery-induced urinary retention has occurred.

81.2 Contraindications

- Absolute
  - Potential urethral injury from trauma
    - Pelvic fractures
    - Known trauma to the urethra
    - Blood at the meatus
- Relative
  - Recent genitourinary surgery (consult with a urologist before placing a catheter)

81.3 Materials and Medications

- Bladder catheterization kit:
  - Sterile gloves
  - Sterile drapes
  - Povidone-iodine (Betadine) solution
  - Cotton sponges or applicators for sterilizing solution
  - Lubricant
  - Specimen collection cup
  - Catheter
    - 5-French feeding tube for neonates
    - 8-French catheters for Infants
    - 10- to 12-French catheters in older children
- Local anesthetic (if desired—2 % lidocaine hydrochloride jelly)
- Absorbent pad

M.A. Plaza-Verduin, MD
Pediatric Division, Department of Emergency Medicine, Arnold Palmer Hospital for Children, Orlando, FL, USA
e-mail: mari.plazav@gmail.com

J.K. Lucas, MD (✉)
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_81

81.4 Procedure

1. Inspect the urinary catheterization tray for all the appropriate materials.
2. Place the patient supine with an absorbent pad under the buttocks.
   (a) Girls should be placed in the frog-leg position (Fig. 81.1).
3. Before sterilizing the field, locate the urethral opening.
4. Remove any powder, ointments, or medicated creams the child might have on the perineum.
5. If needed, apply anesthetic to the area.
   (a) Soak a cotton ball with anesthetic (2 % lidocaine hydrochloride jelly) and hold over the urethra opening for 2 min.
   (b) 0.5–2.0 mL of anesthetic can also be injected into the urethra.
6. Sterilize the area and place drapes appropriately, exposing the genitalia.
7. Catheterization of males
   (a) If uncircumcised, gently retract foreskin, *if possible*, for cleaning and visualization of the meatus.
   (b) Hold the penis using the nondominant hand at a 90° angle from the body (Fig. 81.2).
   (c) Lubricate the catheter tip.
   (d) Insert the lubricated catheter into the meatal opening and advance it while applying gentle traction to the penis from the base of the penis.
   (e) If resistance is met, maintain gentle pressure with the catheter.
      - *Do not attempt to force the catheter that could create a false tract or traumatic fistula.*
   (f) Advance the catheter until urine is obtained, approximately inserting the catheter to just beyond the penile length.
   (g) Once completed, gently withdraw the catheter.
   (h) Clean the area, wiping away the Betadine solution.
   (i) If uncircumcised, pull the foreskin over the glans to avoid paraphimosis.
8. Catheterization of females
   (a) Sterilization of the area should occur from anterior to posterior.
   (b) Have an assistant hold the labia majora apart.
      (i) If no assistant is available, use the nondominant hand to hold the labia apart.
         - Holding the labia majora with a gentle outward, lateral, and upward traction will help visualize the meatus (Fig. 81.3).

Fig. 81.1 Infant held in the frog-leg position for catheterization

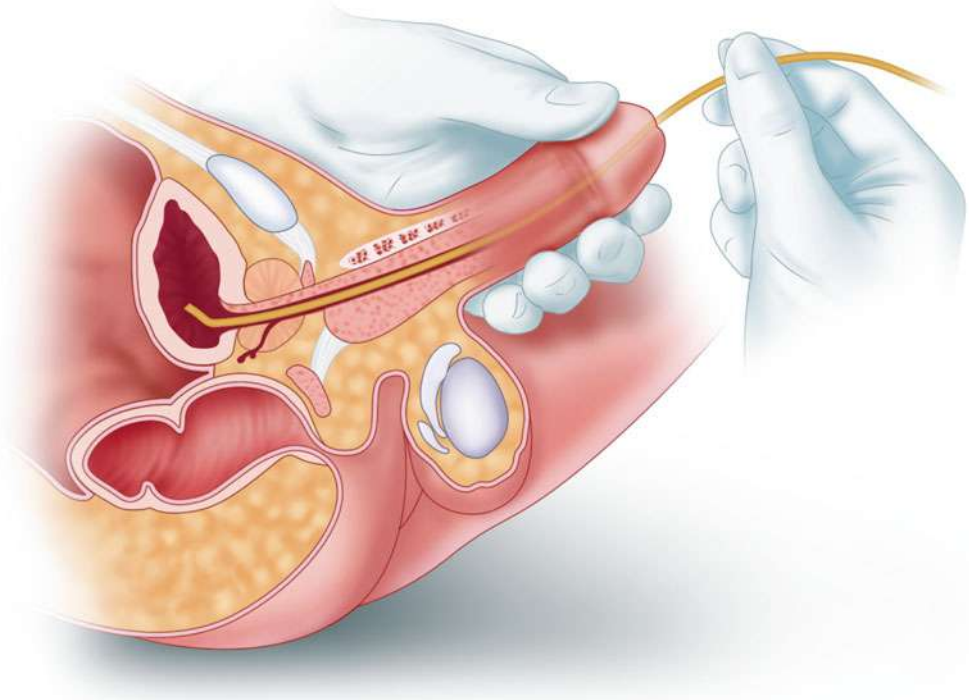

Fig. 81.2 Bladder catheterization of a male; penis should be held perpendicular to the body

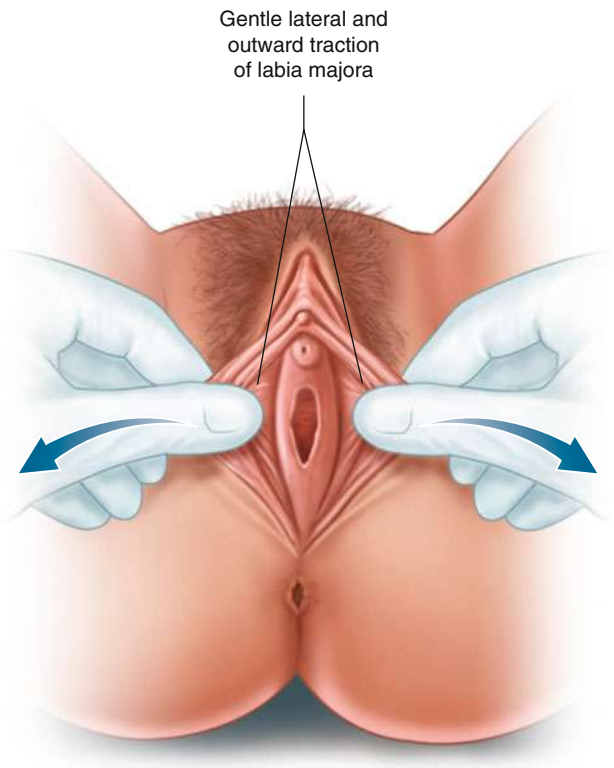

Fig. 81.3 Positioning of labia for better visualization of the meatus

- Downward displacement of the cephalad aspect of the vaginal introital fold with a cotton-tipped applicator can help visualize the urethral meatus (Fig. 81.4).

(c) Lubricate the catheter tip.

(d) Insert the lubricated catheter into the meatal opening. Advance slowly until urine is obtained (Fig. 81.5).

(e) Once completed, gently withdraw the catheter.

(f) Clean area, wiping away the Betadine solution.

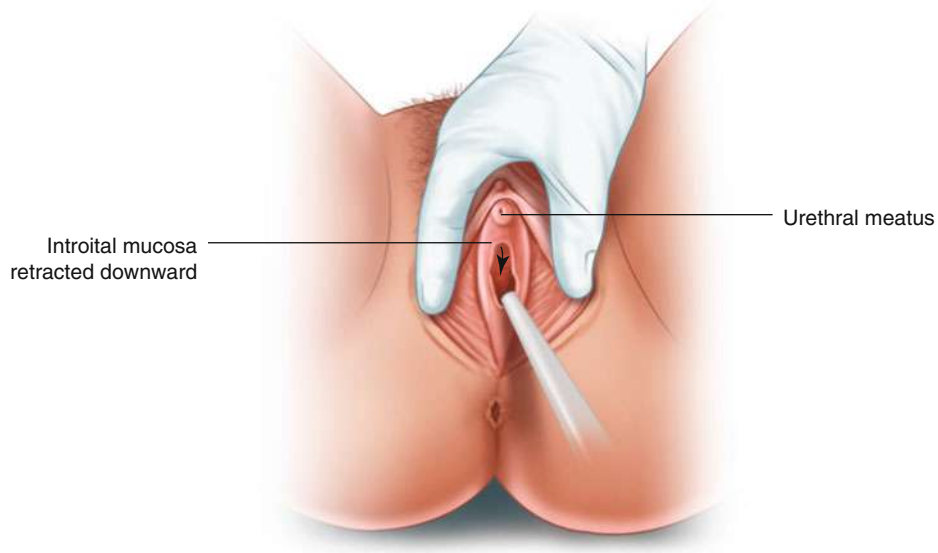

Fig. 81.4 Better visualization of the meatus is achieved with downward displacement of the introital mucosa

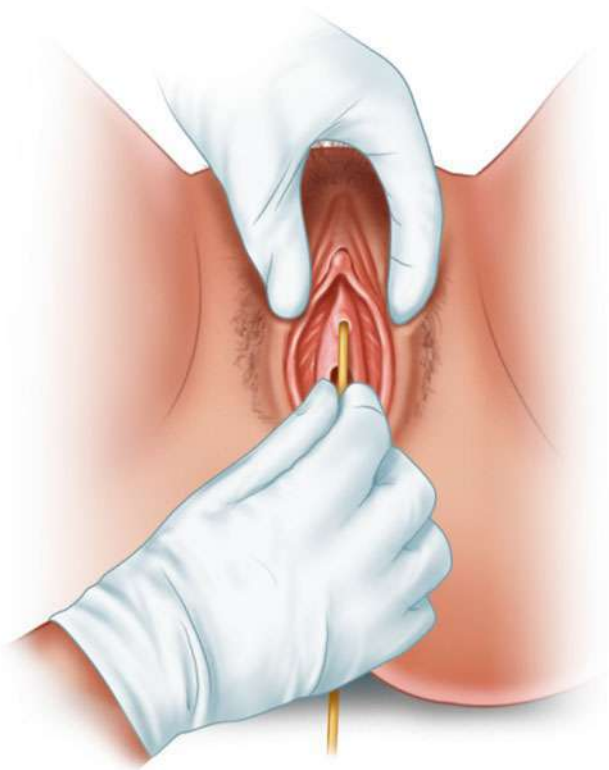

Fig. 81.5 Bladder catheterization of a female

81.5 Complications

- Urethral or bladder injury
- Infection if sterile field not maintained
- Paraphimosis owing to failure to restore a retracted foreskin to its normal position

81.6 Pearls and Pitfalls

- Pearls
  - It is not necessary to fully retract a foreskin. This only causes trauma and increases the likelihood of paraphimosis. As the infant/boy ages, the foreskin will loosen and the naturally occurring adhesions will spontaneously release.
  - The urethral meatus in an infant female is usually tucked just above the redundant hymen (as opposed to the more anteriorly located meatus in the adult woman) and often looks like a dimple or small blind pouch.
  - Using viscous lidocaine in lieu of, or blended with, lubricant anesthetizes the meatus and urethra as the catheter passes.
  - Have the specimen cup at the ready because, at times, once the (usually cold) Betadine or antiseptic solution is applied, the infant often releases the urine and it can be caught, literally, midstream.
  - In the uncircumcised male, be certain to return the foreskin over the glans to avoid paraphimoses.
- Pitfalls
  - If catheterizing a child in search of infection, send a urine culture regardless of the urinalysis results because the younger infants can have false-negative urinalysis and still have positive cultures.

Selected Reading

American Academy of Pediatrics, Subcommittee on urinary tract infection, steering committee on quality improvement and management. Urinary tract infection: clinical practice guideline for the diagnosis and management of the initial UTI in febrile infants and children 2 to 24 months. Pediatrics. 2011;128:595–609.

Beno S, Schwab S. Bladder catheterization. In: King C, Henretig FM, editors. Textbook of pediatric emergency procedures. 2nd ed. New York: Lippincott Williams & Wilkins; 2008.

Cheng YW, Wong SN. Diagnosing symptomatic urinary tract infection in infants by catheter urine culture. J Paediatr Child Health. 2005;41:437–40.

Gerard LL, Cooper CS, Duethman KS, et al. Effectiveness of lidocaine lubricant for discomfort during pediatric urethral catheterization. J Urol. 2003;170:564–7.

Kozer E, Rosenbloom E, Goldman D, et al. Pain in infants who are younger than 2 months during suprapubic aspiration and transurethral bladder catheterization: a randomized, controlled study. Pediatrics. 2006;118:e51–6.

82 Pelvic Examination and Wet Preparation

Nauman W. Rashid, Elaine B. Josephson, and Muhammad Waseem

82.1 Indications

- Lower abdominal or pelvic pain
- Vaginal bleeding or discharge
- Cancer screening
- Pregnancy
- Exposure to sexually transmitted disease
- Sexual assault

82.2 Contraindications

- Physical or mental disability
- Recent gynecological surgery
- Third-trimester pregnancy with bleeding
- Premenstrual females (may not be indicated in adolescents, who are not sexually active, unless there is discharge, bleeding, suspicion for abuse, or a foreign body)
- If a speculum examination is necessary, examination under general anesthesia should be considered.

N.W. Rashid, MD
Department of Emergency Medicine, WellStar Kennestone Hospital, Marietta, GA, USA
e-mail: naumanrashid@hotmail.com

E.B. Josephson, MD
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, Weill Cornell Medical College of Cornell University, Bronx, NY, New York
e-mail: elaine.Josephson@nychhc.org

M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: waseemm2001@hotmail.com

82.3 Materials and Medications

- Examination table with stirrups (Fig. 82.1)
- Reliable light source
- Appropriately sized speculum (Fig. 82.2)
- Endocervical brush or spatula
- Culture swab for gonorrhea and chlamydia
- Large cotton swabs for vaginal discharge or bleeding (Fig. 82.2)
- pH paper
- Saline and potassium hydroxide dropper bottles for wet preparations
- Lubricating gel
- Disposable gloves (Fig. 82.2)
- Small stool or chair for examiner

82.4 Procedure

1. Obtain permission from patient before beginning examination.
2. Chaperone should be present (medical staff member).
3. Make sure the examination table is clean and appropriately draped.
4. Have the patient in a loose-fitting gown.
5. Place the patient on the examination table in the lithotomy position with both feet in the stirrups and have the patient's pelvis as close to the edge of the table as possible.
6. Turn on the light source and adjust for optimum illumination. Put on the disposable gloves.
7. Communicate the procedure well to the patient.
8. Begin the examination with inspection and palpation of the abdomen.
9. Examine the external genitalia. Evaluate the skin, labia minora and majora, clitoris, urethral meatus, vaginal

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_82

Fig. 82.1 Examination table with stirrups

Fig. 82.2 Gloves, speculum, and swabs

canal, and Bartholin glands (Fig. 82.3). Look for skin abnormalities, lesions, masses, rashes, excoriation, abscesses, discharge, bleeding, or trauma. Palpate for tenderness.

10. Lubricate the appropriate-size speculum (mostly medium size). Insert the speculum through the vaginal opening with gentle downward pressure. The speculum should advance without any resistance until the cervical os is visualized.
11. Inspect the vaginal walls for any lesions or masses (Fig. 82.4). The cervical os is inspected to see if it is open or closed. Cervical cultures for gonorrhea and chlamydia are obtained with a cotton swab and sent for microbiology.
12. A sample of the discharge or bleeding is taken with a large cotton swab. The color, odor, and amount should be noted. The pH of the vaginal discharge can be evaluated. Normal pH is less than 4.5. An elevated pH indicates an infection (Table 82.1).

Fig. 82.3 Female external genitalia

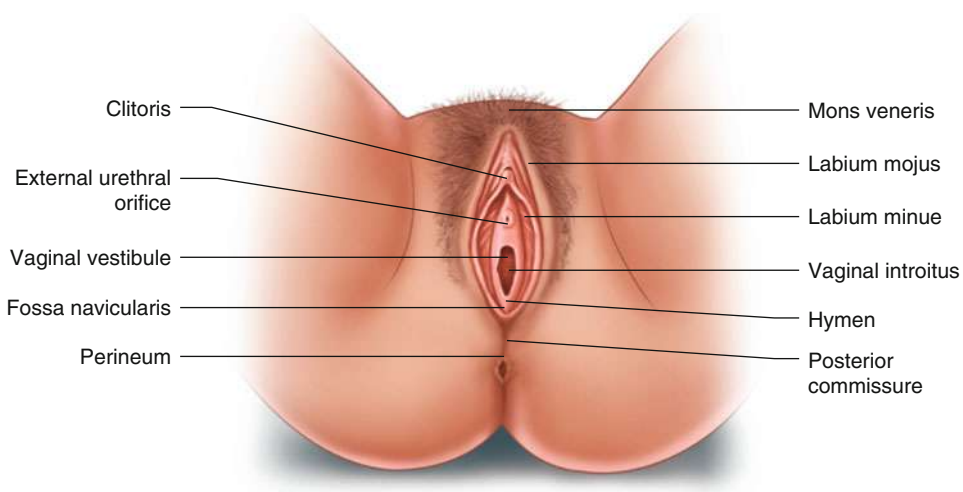

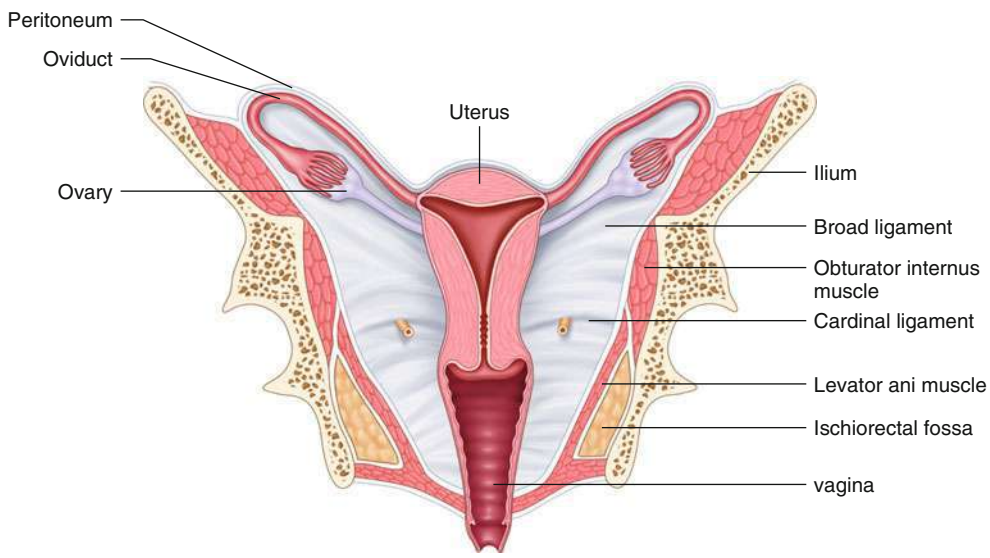

Fig. 82.4 Female internal genitalia

Table 82.1 Wet preparation interpretation

Organism	Preparation	pH	Microscope	Cervix	Appearance of discharge
Bacterial vaginosis	Saline	>4.5	Clue cells	Redness	Thin, milky, fishy odor
Trichomoniasis	Saline	>4.5	Motile flagella	Strawberry red	Yellow-green, foamy
Yeast	Potassium hydroxide	3.8–4.5	Budding yeast pseudohyphae	Normal	White, cottage cheese

13. Next, a bimanual examination should be performed (Fig. 82.5). Lubricating gel is applied to the nondominant gloved hand and the index and middle finger are inserted into the vagina until the cervix is felt. The other hand is placed on the abdomen to palpate the uterus and ovaries. Pressure is applied to the abdomen while the vaginal hand is elevated upward.
14. The cervix is palpated to elicit any cervical motion tenderness. The uterus is palpated and the size, position, and mobility are noted. The adnexa are examined for masses and tenderness. If a mass is palpated, the size, mobility, consistency, and tenderness are noted.
15. The final part of the pelvic examination is the rectovaginal examination. Lubricate the index and middle fingers

of the left hand. Place the index finger in the vagina and the middle finger in the rectum. Palpate for any fistulas or masses. With the finger, also palpate the uterosacral ligaments, the broad ligaments, and the pelvic side walls. The finger is then gently removed and any feces are inspected for mucous or occult blood.

16. A wet preparation is made by obtaining a sample of the vaginal discharge and placing it in a vial mixed with saline solution. A drop of the solution is placed on a microscopic slide and examined under high magnification for the presence of clue cells (Fig. 82.6) diagnostic for bacterial vaginosis and trichomonads (Fig. 82.7) diagnostic for trichomoniasis. For yeast, two drops of the solution is mixed with two drops of potassium hydroxide. Presence of hyphae is diagnostic of candida (yeast) species (Fig. 82.8)

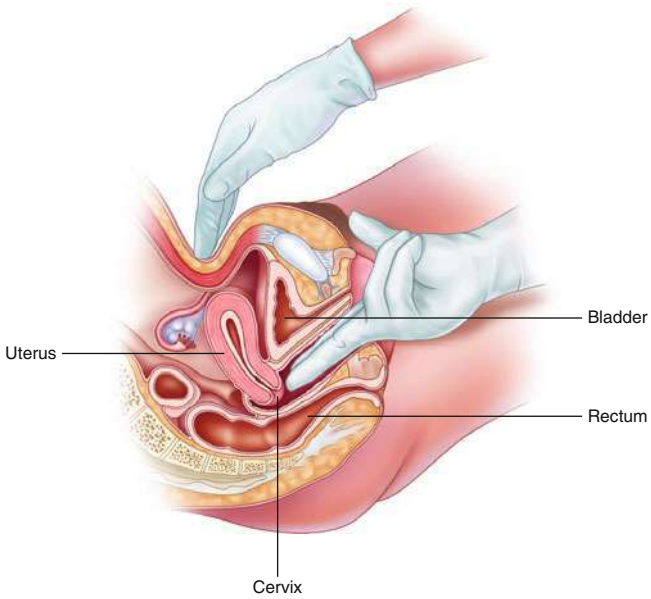

Fig. 82.5 Bimanual pelvic examination

82.5 Complications

- Urinary tract infection
- Vaginal bleeding
- Cramping

82.6 Pearls and Pitfalls

- Pearls
  - Good communication is essential to ensure the patient is comfortable and not anxious.
  - The chaperone should be a medical staff member.
  - Do not skip the pelvic examination if the patient is menstruating.
- Pitfalls
  - Do not forget to perform a complete abdominal exam along with the pelvic exam to rule out any GI etiology.
  - In older females (>50), perform the DRE for a stool occult sample as a possible source of bleeding.

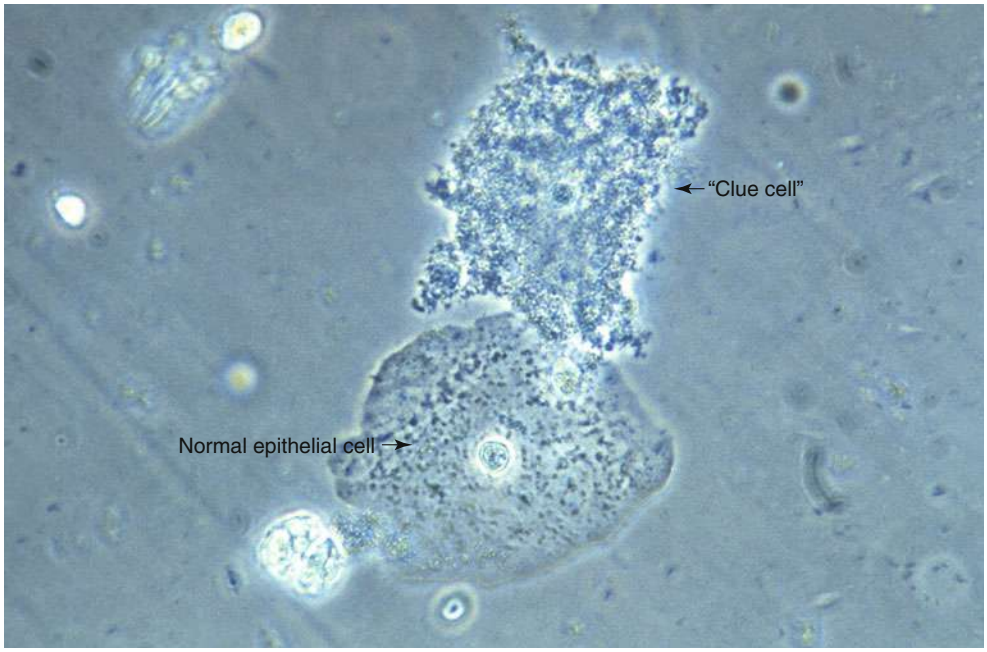

Fig. 82.6 Photomicrograph of a vaginal smear specimen depicting two epithelial cells, a normal cell, and an epithelial cell with its exterior covered by bacteria giving the cell a roughened, stippled appearance known as a "clue cell" (From the CDC Public Health Image library)

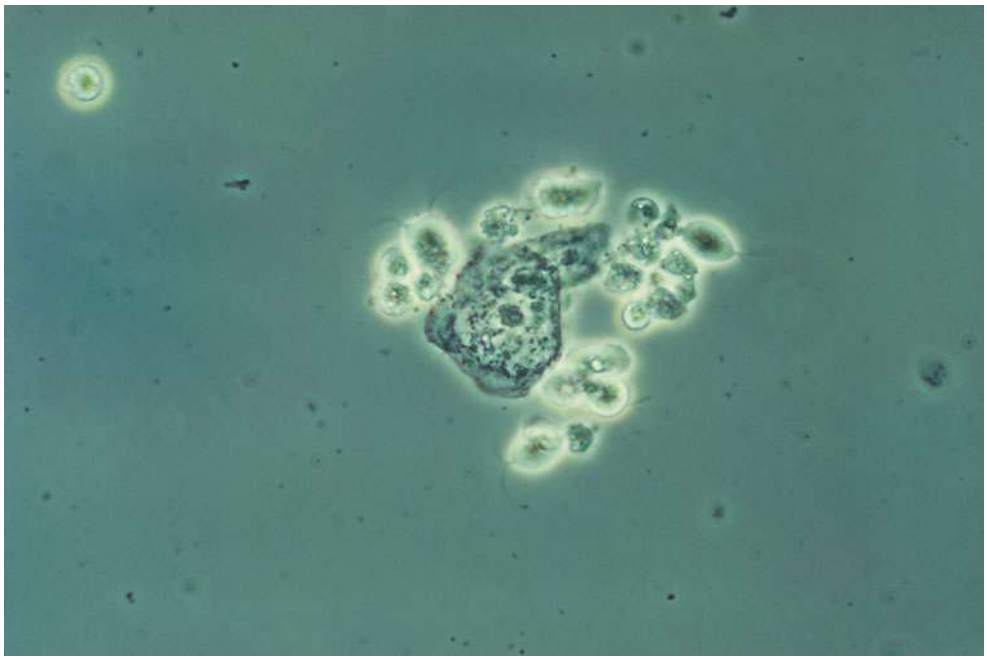

Fig. 82.7 Photomicrograph of trichomonads in wet mount prepared witvh physiological saline (From the CDC Public Health Image library)

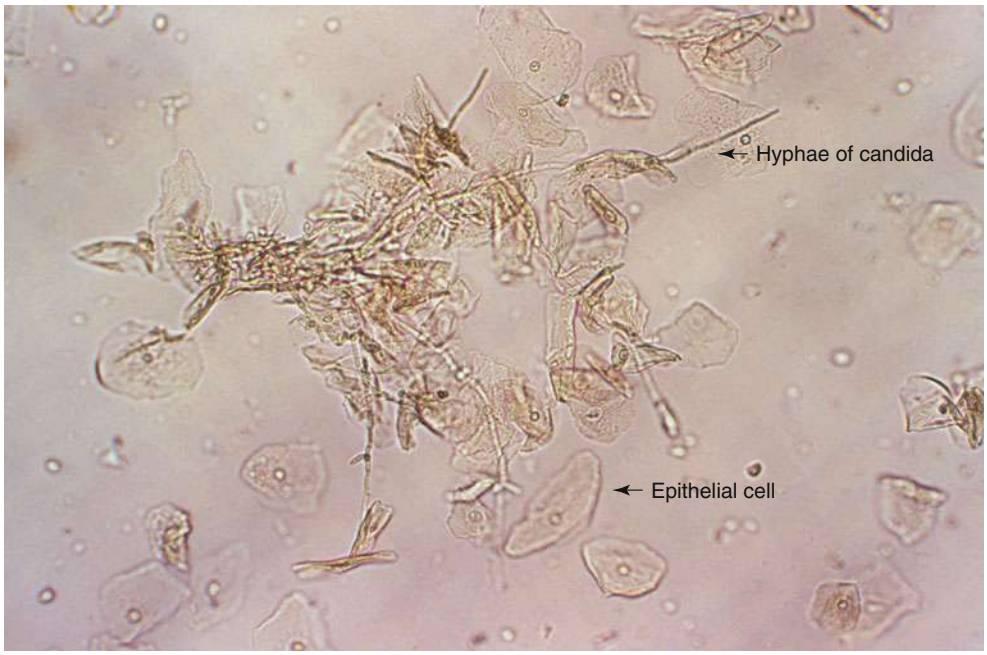

Fig. 82.8 *Candida albicans* from vaginal wet prep (From the CDC Public Health Image library)

Selected Reading

Brown J, Fleming R, Aristzabel J, Gishta R. Does pelvic exam in the emergency department add useful information? West J Emerg Med. 2011;12:208–12.

Butler J, Barton D, Shepherd J, Reynolds K, Kehoe S. Gynaecological examinations. Good not bad medicine. BMJ. 2011;342:d1760.

Carr SE, Carmody D. Outcomes of teaching medical students core skills for women's health: the pelvic examination educational program. Am J Obstet Gynecol. 2004;190:1382.

Katz VL, Lentz G, Lobo RA, Gershenson D, editors. Comprehensive gynecology. 5th ed. Philadelphia: Mosby; 2007.

Tiemstra J, Chico P, Pela E. Genitourinary infections after a routine pelvic exam. J Am Board Fam Med. 2011;24:296–303.

83 Bartholin Gland Abscess/Cysts Drainage

Holly H. Charleton, Marylin Otero, Diane F. Giorgi, and Joseph A. Tyndall

83.1 Indications

- Bartholin gland cysts larger than 1 cm or painful
- Bartholin gland abscess

83.2 Contraindications

- Absolute
  - None
- Relative
  - Recurrent/complex abscess requiring general anesthesia in operating room
  - Coagulopathy

83.3 Procedure Types

- Incision and drainage (I&D)
- Iodoform packing
- Word catheter
- Jacobi ring catheter
- Silver nitrate stick ablation (after I&D)

H.H. Charleton, MD
Emergency Department, The Brooklyn Hospital Center, Brooklyn, NY, USA
e-mail: Hcmd08@gmail.com

M. Otero, MD
Department of Emergency Medicine, Franklin Hospital, Valley Stream, NY, USA
e-mail: momd42@hotmail.com

D.F. Giorgi, MD
Department of Emergency Medicine, Mount Sinai Queens, New York, NY, USA
e-mail: dgiorgi@earthlink.net

J.A. Tyndall, MD (✉)
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: tyndall@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_83

83.4 Materials

- Sterile gloves
- Sterile skin preparatory solution/swabs and drapes
- Lidocaine 1 % (local anesthesia); may give oral or intravenous sedatives or analgesics
- Needles 25 or 27 gauge with 3-mL syringe for lidocaine injection
- Scalpel #11 for incision
- Culture swab
- Hemostat or suture kit
- Word catheter (if choosing that method) (Fig. 83.1)
- Jacobi ring catheter (if choosing that method) (Fig. 83.2)
- 3-mL syringe with saline for inflation of Word catheter balloon
- Silver nitrate stick (if choosing that method)
- Gauze pads for bleeding and effluents
- Iodoform packing

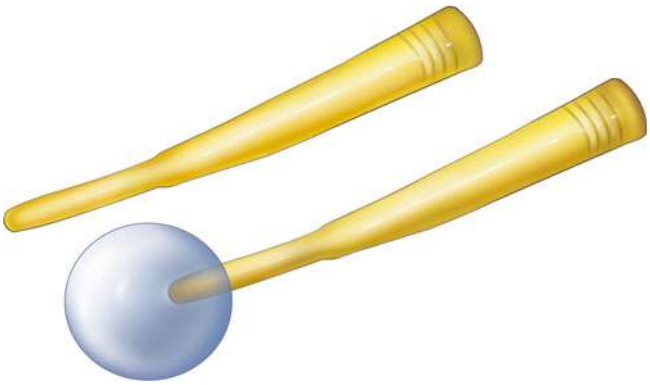

Fig. 83.1 Word catheter (inflated/deflated)

Fig. 83.2 Jacobi ring catheter

83.5 Procedures

83.5.1 Incision and Drainage

1. Obtain informed consent from the patient.
2. Place the patient in the lithotomy position.
3. Prepare the cyst/abscess and the surrounding area with sterilizing fluid/swabs and drape the area, leaving the cyst/abscess accessible for the procedure (Fig. 83.3)
4. May need to hold traction to the labia to fully expose the cyst/abscess (Fig. 83.4).
5. Inject 1–4 mL of lidocaine at the planned site of incision (Fig. 83.5).
6. Hold one side of the cyst/abscess with a forceps or hemostat to maintain traction while incising.
7. With a #11 scalpel, make an incision approximately 0.5–1 cm and 1.5 cm deep in the introitus or behind the hymnal ring to prevent vulvar scarring. The incision should be made through the fluctuant area of the abscess on the mucosal surface. The incision should be linear and large enough to fit the Word catheter (if using that method).
8. Drain the cyst/abscess completely, using the hemostat to break up the loculations.
9. Culture the abscess with a swab and send to microbiology.
10. All previous steps should be done regardless of the procedure type to follow.

Fig. 83.3 Cyst/abscess accessible for the procedure

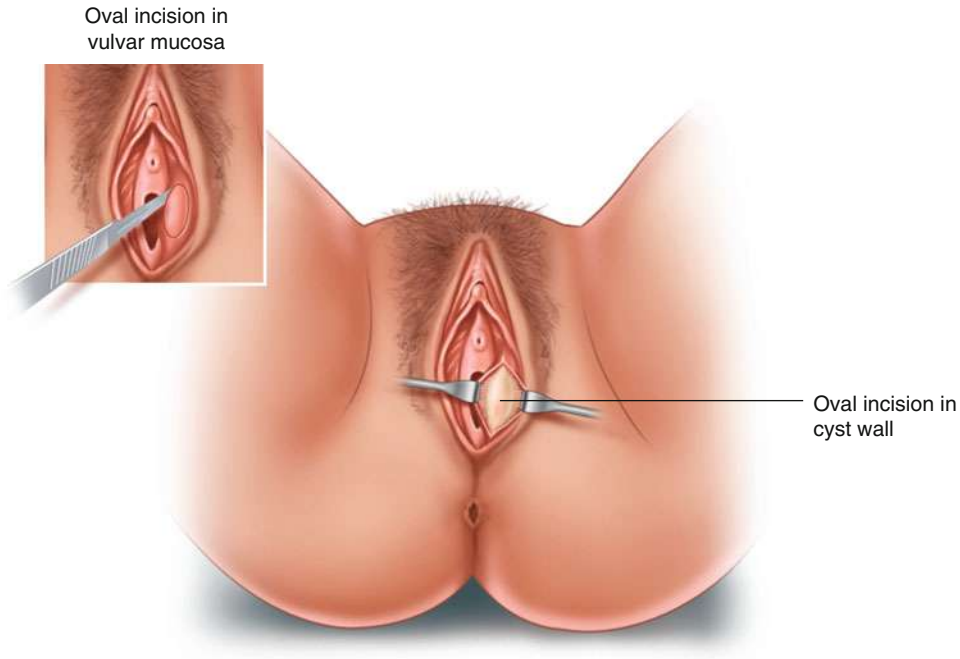

Fig. 83.4 May need to hold traction to the labia to fully expose the cyst/abscess

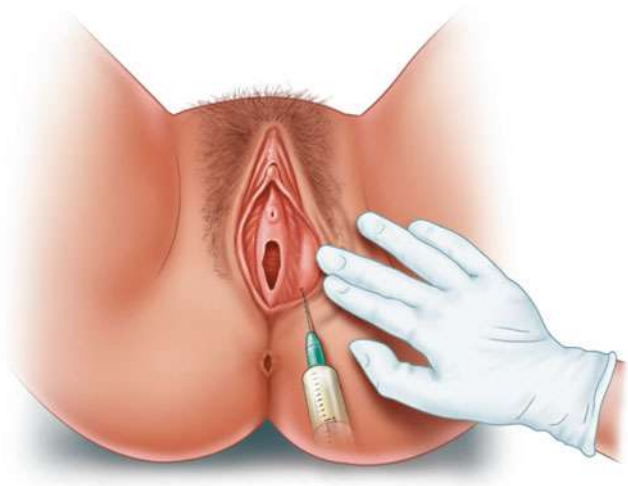

Fig. 83.5 Inject 1–4 mL of lidocaine at the planned site of incision

83.6 Iodoform Packing

1. Pack with iodoform packing, grasping the end of the packing material with a hemostat and inserting deeply within the cavity.
2. The cavity should be filled with the packing material and a small piece left exposed (for ease of retrieval during follow-up).
3. Clean the site and cover with gauze.

83.6.1 Word Catheter

- After culturing the abscess
  1. Place the Word catheter into the incision site as deep as possible (if the incision is too large, the catheter will slip out) (Fig. 83.6).
  2. Inflate the balloon of the Word catheter with 2–3 mL of saline or water injected into the hub with a needle and syringe (Fig. 83.7).
  3. Tuck the end of the Word catheter into the vagina for comfort purposes.
  4. The catheter should remain in place for 4 weeks to allow for epithelialization of the tract.

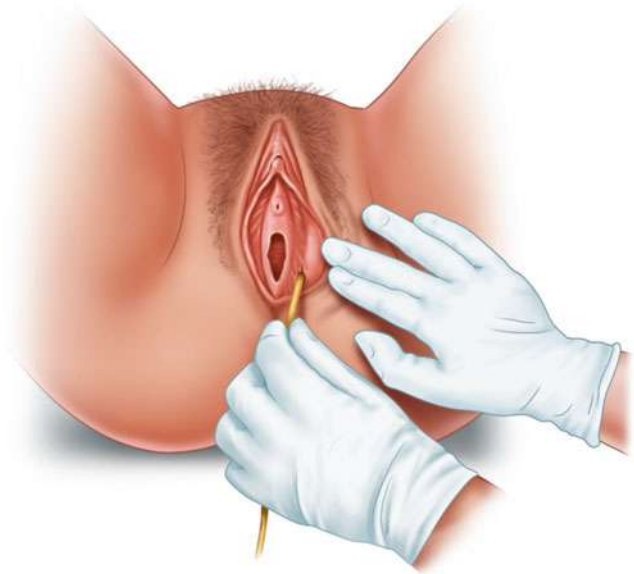

Fig. 83.6 Place the Word catheter into the incision site as deep as possible

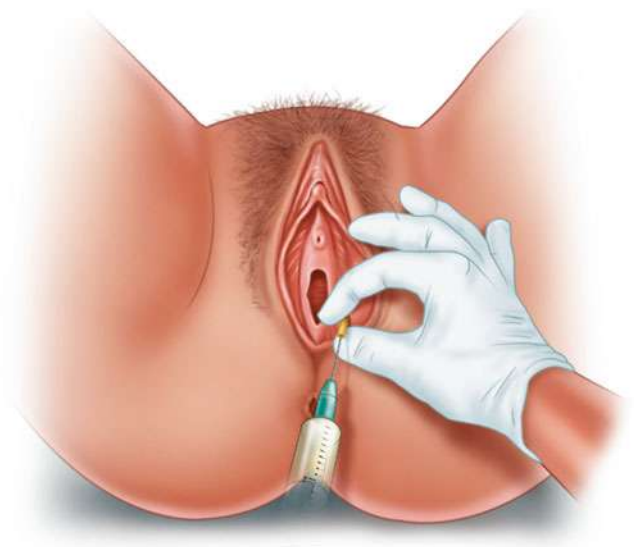

Fig. 83.7 Inflate the balloon of the Word catheter with 2–3 mL of saline or water injected into the hub with a needle and syringe

83.6.2 Jacobi Ring Catheter

1. Grasp one end of the Jacobi ring with a hemostat and pass it through the initial incision site.
2. At this time, use a hemostat to break loculations and culture the material for microbiology.
3. Pull the Jacobi ring through the abscess cavity (be careful not to pull the suture out of the catheter) and make a second incision to pull the catheter out through it.
4. The two ends of the catheter are then tied, forming a ring.

83.6.3 Nitrate Stick

- After sending the culture
  1. Take the silver nitrate stick and place it deep within the cyst/abscess cavity (Fig. 83.8).
  2. The patient is instructed to return in 48 h for removal of the remaining nitrate material and necrotic tissue and wound cleaning.
  3. The patient should be warned of side effects including pain, chemical burns of nearby tissue, edema, discharge, and scarring.

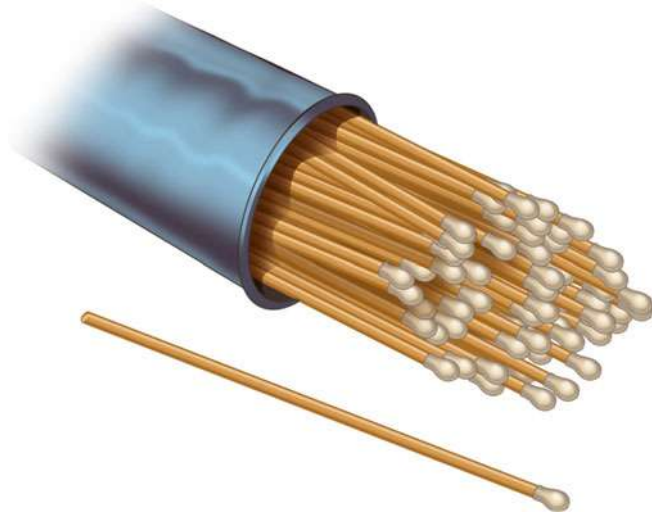

Fig. 83.8 Silver nitrate sticks

83.7 Aftercare

- Follow-up care is required after each of these procedures.
- A high-risk patient (e.g., diabetic) may need to be covered with broad-spectrum antibiotics.
- Pregnant women are also considered high risk and should be given antibiotics and followed closely.
- The patient should be instructed to remain on pelvic rest (nothing in vagina) while the catheter is in place, wear a pad owing to discharge, use sitz baths and analgesics for pain control, and follow up 1–2 days after the procedure.

Although a few other methods are available for draining Bartholin cyst/abscess including marsupialization, carbon dioxide laser vaporization, and excision, these methods are typically done by gynecologists and rarely, if ever, performed by an emergency department physician.

Selected Reading

Chen KT, Robert LB, Falk SJ. Disorders of Bartholin's gland. Available at: WWW.Uptodate.com.

Chen KT, Robert LB, Falk SJ. Word catheter placement for treatment of Bartholin's cysts and abscesses. Available at: WWW.Uptodate.com.

Gennis P, Li SF, Provataris J, et al. Jacobi ring catheter treatment of Bartholin's abscesses. Am J Emerg Med. 2005;23:414–5.

Vitaly KA, Mosquera C. Novel technique for management of Bartholin gland cysts and abscesses. J Emerg Med. 2008;36:388–90.

Wechter ME, Wu JM, Marzano D, Haefner H. Management of Bartholin duct cyst and abscesses: a systematic review. Obstet Gynecol Surv. 2009;64:395–404.

88.3 Materials and Medications (Fig. 88.1)

- 1 % Xylocaine with or without 1:200,000 epinephrine, 0.25 % bupivacaine solution. 8.4 % (1 mL/mL) sodium bicarbonate (optional)
- 18-, 25-, or 27-gauge needles, syringes up to 10 mL
- Sterile and nonsterile gloves, face shield
- Alcohol pads and povidone-iodine swabs

Fig. 88.1 Local anesthesia materials

Regional Anesthesia (Nerve Blocks)

89

Derek Ailes and Muhammad Waseem

89.1 Indications

- Repair of wounds where preserving anatomical landmarks or having precise anatomical alignment is important (e.g., vermillion border of lip)
- Pain control in dislocation or fracture reductions
- Incision and drainage of abscesses
- Burn and wound care
- Extensive or multiple lacerations (reduces total amount of local anesthetic needed)
- Foreign body removal

89.2 Contraindications

- Allergic to local anesthetic (see Chap. 88)
- History of coagulopathy or bleeding disorder
- Injection through infected tissue

D. Ailes, MD • M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: derekailes@yahoo.com; waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_89

89.3 Materials and Medications (Fig. 89.1)

- Povidone-iodine, alcohol swabs
- Sterile gloves and drapes
- Local anesthetic solution (e.g., lidocaine, Marcaine with or without epinephrine)
- 18-gauge, 20- to 30-gauge needles 2 in. in length
- 22- to 24-gauge spinal needles
- Syringes up to 60 mL

Fig. 89.1 Materials

89.4 Procedure: General Block

1. Obtain consent after explaining the risks of procedure including temporary paresthesias and expected duration of block. Perform neurological examination before procedure, documenting any preexisting deficits.
2. Position the patient comfortably, preferably supine, anticipating vasovagal response.
3. Identify landmarks for the block. Clean area, prepare with povidone-iodine, and surround with sterile drapes.
4. A small skin wheal of local anesthetic may be placed at the site of needle entry before block.
5. Insert the needle into the site while aspirating to ensure it is not in a vessel.
6. If paresthesia is elicited, withdraw the needle slightly allowing paresthesia to improve and inject.
7. Wait 5–15 min for the block to reach full effect.
8. Test for sharp sensation in the anesthetized area and document.

89.5 Complications

- Infection
- Hemorrhage
- Hematoma
- Allergic reaction
- Systemic toxicity (exceeded maximum dose or inadvertent injection into vasculature)
- Paresthesias, pain
- Intraneural injection causing ischemia
- Intra-arterial injection of epinephrine causing vasospasm and tissue ischemia

89.6 Pearls and Pitfalls

- Pearls
  - It is important to aspirate before injecting anesthesia when performing regional anesthesia because, unlike local techniques, the needle is deeper and in proximity to larger vessels.
  - Shooting pain and/or paresthesias occurs when the needle contacts the nerve. When this happens, withdraw the needle 1 mm and wait for the paresthesia to resolve before injecting.
  - Injury can occur to a limb or digit if the patient manipulates it before the anesthesia wears off. The patient should be cautioned not to use the affected area until motor and sensation return. If an extensive block was done, monitor the patient in the emergency department until return of baseline neurological function.
- Pitfalls
  - It is traditionally taught to avoid epinephrine-containing solutions in blocks in end-artery blocks (e.g., digital blocks). However, evidence for any vascular insufficiency and necrosis as a result is lacking in standard commercially available lidocaine with epinephrine preparations. They should, however, be avoided in patients with peripheral artery disease [1, 2].

89.7 Selected Specific Blocks

- Contraindications and materials are the same as general block.

89.7.1 Facial Blocks: Trigeminal Nerve (Fig. 89.2) [2, 3]

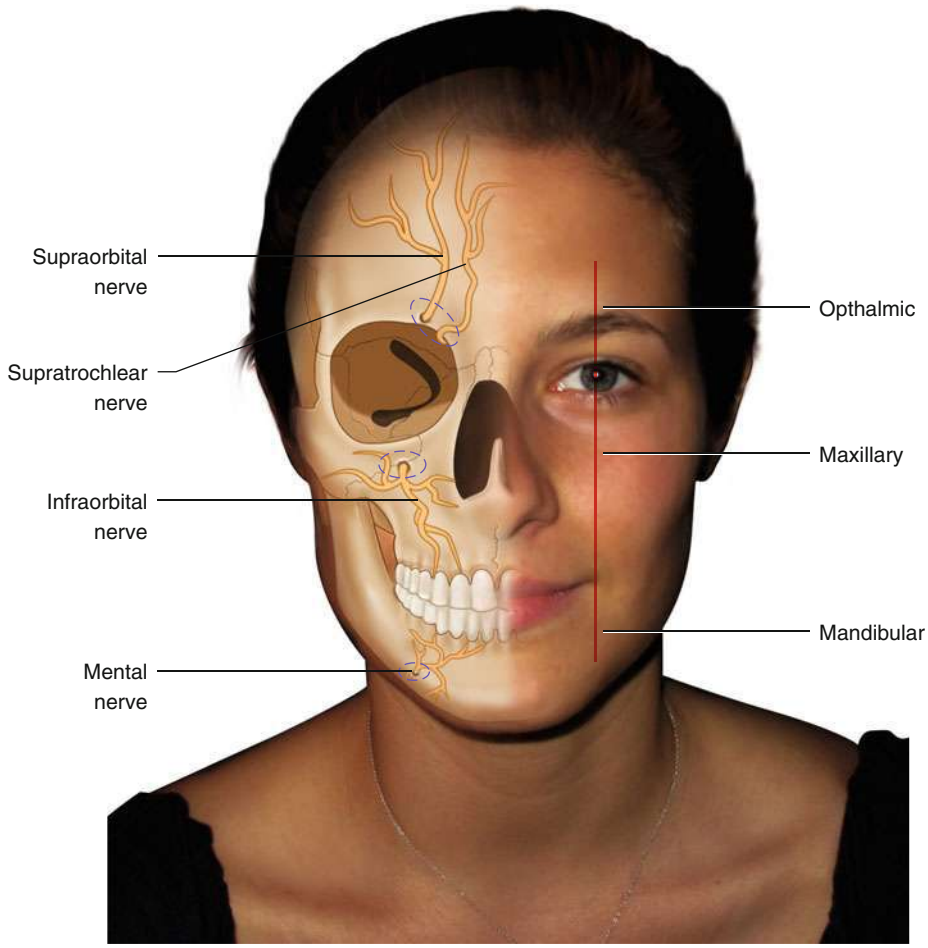

Fig. 89.2 Vertical plane through the midposition of the pupil shows the position of the supraorbital foramen, infraorbital foremen, and mental foramen

89.7.1.1 Supraorbital (Fig. 89.3) and Supratrochlear Nerve Block

- Indications
  - To anesthetize the forehead from the orbital ridge to the vertex of the scalp. The supraorbital nerve emerges from the supraorbital foramen/notch and is a branch of the ophthalmic division of the trigeminal nerve. The supratrochlear nerve also is a branch of the ophthalmic division of the trigeminal nerve and exits through the superior medial aspect of the orbit.
- Procedure
  1. Inject local anesthetic solution over the midline of the forehead at eyebrow level.
  2. Inject a 25- or 27-gauge needle through skin wheal aimed laterally while injecting 3–5 mL local anesthetic subcutaneously. Stop infiltrating when the needle reaches the midline of the orbit.

Fig. 89.3 Supraorbital nerve block

89.7.1.2 Infraorbital Nerve Block

- Indications
  - To anesthetize the medial cheek, upper lip, philtrum, skin between the lips and the nose, and nasal ala. The infraorbital nerve emerges from the infraorbital foramen and is a branch of the maxillary division of the trigeminal nerve. Anesthesia to the infraorbital nerve will also provide anesthesia to its terminus, the superior alveolar nerves.
- Procedure: Extraoral Approach (Fig. 89.4)
  1. Palpate the inferior orbital foramen in its midline position. The infraorbital nerve is often tender on palpation as it exits the foramen.
  2. Inject a 25- or 27-gauge needle just above the infraorbital foramen injecting 1–2 mL of local anesthetic.
  3. Take care not to inject into the foramen because there is an increased risk of intraneural injection.
  4. Hold a finger on the inferior orbital rim to avoid ballooning of the lower eyelid with injection.
  5. Intraoral approach is possible and preferred because it is less painful.

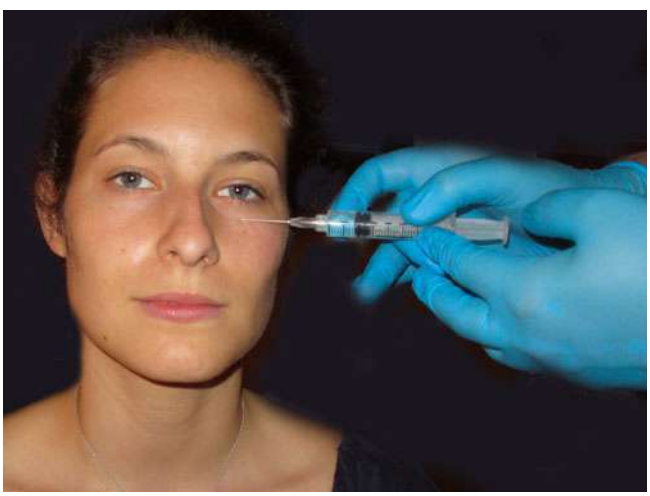

Fig. 89.4 Extraoral approach to the infraorbital nerve block

- Procedure: Intraoral Approach (Fig. 89.5) [3]
  1. Apply topical benzocaine or lidocaine gel to the point of insertion, which is the height of the mucobuccal fold over the first premolar, which is the site of insertion. Wipe off after 1–3 min.
  2. Palpate with the finger of the noninjecting hand over the inferior border of the inferior orbital rim. Retract the lip with the noninjecting hand.
  3. Using a long 25- to 27-gauge needle, with the bevel toward the bone, advance the needle at the insertion site toward the infraorbital foramen. Once the target is reached, aspirate and inject 1 mL of local anesthetic.
  4. Exert pressure on the foramen for 1 min after injection to force the anesthetic through the infraorbital foramen.
  5. If the needle is difficult to advance and the patient experiences pain on insertion, redirect the needle laterally and advance.
  6. If analgesia is attained for the lip but not the eyelid, the analgesia was placed inferior to foramen, and if analgesia is attained for the eyelid but not the lip, placement was superior to the foramen.

Fig. 89.5 Intraoral approach to the infraorbital nerve block

89.7.1.3 Mental Nerve Block

- Indications
  - To anesthetize the lower lip and chin and is especially useful in laceration repair at those sites. The mental nerve emerges from the mental foramen and is a branch of the mandibular division of the trigeminal nerve. Mental foramen lies in the vertical plane with the mid-point of the pupil and sits in the middle of the body of the mandible.
- Procedure: Extraoral Approach (Fig. 89.6)
  1. Inject local anesthetic solution over the identified location of the mental foramen, creating a skin wheal.
  2. Advance a 25- or 27-gauge needle through the skin wheel until the mandible is contacted, injecting 1–2 mL of local anesthetic.
  3. Intraoral approach is possible and preferred because it is less painful.
- Procedure: Intraoral Approach (Fig. 89.7) [3]
  1. Apply topical benzocaine or lidocaine gel to the point of insertion, which is the mucobuccal fold between the apices of the first and the second premolars. Wipe off after 1–3 min.
  2. Insert a 25- to 27-gauge needle, with the bevel toward the mandible, aimed toward the mental foramen.
  3. After advancing one-third the depth of the mandible and contacting the mandible, inject 1–2 mL of local anesthetic.
  4. By pressing firmly on the mental foramen for 2–3 min after the mental foramen has been blocked, an incisive nerve block is also created. This is useful if anesthesia to the lower anterior teeth is also desired.

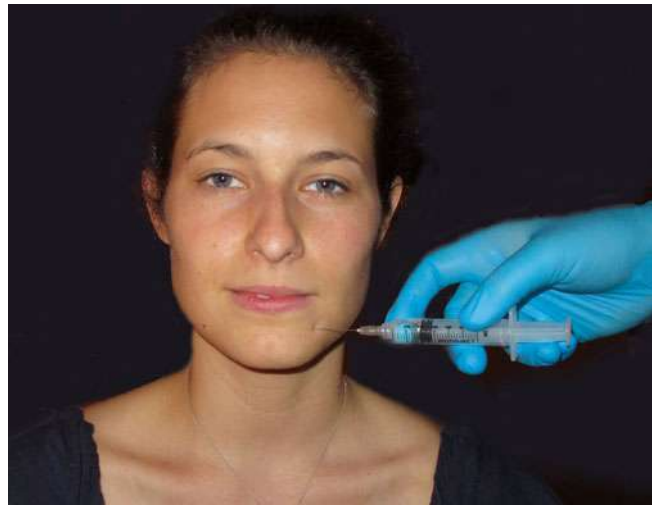

Fig. 89.6 Extraoral approach to the mental nerve block

Fig. 89.7 Intraoral approach to the mental nerve block

89.7.2 External Ear Block (Fig. 89.8) [4]

- Indications
  - To anesthetize the entire external ear, excluding the external auditory canal and the concha
  - Especially useful in large lacerations of the ear and surrounding skin, hematoma evacuations, or incision and drainage of abscess.
- Procedure: Auricular Ring Block (Fig. 89.9) [4]
  1. Using a 25- to 27-gauge needle, insert the needle just inferior to the earlobe directing it toward the tragus.
  2. Aspirate and advance the needle superiorly subcutaneously injecting 3–4 mL of local anesthetic (Fig. 89.9, #1).
  3. Withdraw the needle without fully removing it and redirect it posterosuperiorly along the inferior posterior auricular sulcus, aspirating and injecting as before (Fig. 89.9, #2).
  4. Remove the needle and insert it just superior to the point of helix insertion into the scalp.
  5. Advance the needle and aspirate and inject in the direction of the tragus. Inject into the subcutaneous tissue while avoiding the ear cartilage (Fig. 89.9, #3).
  6. Withdraw and redirect the needle posteriorly and inferiorly toward the skin behind the ear, injecting as before (Fig. 89.9,#4).
  7. Beware of inadvertent cannulation of the superficial temporal artery, which crosses the zygomatic arch and crosses medial to the ear. If the artery is violated, it requires 20–30 min application of firm pressure.

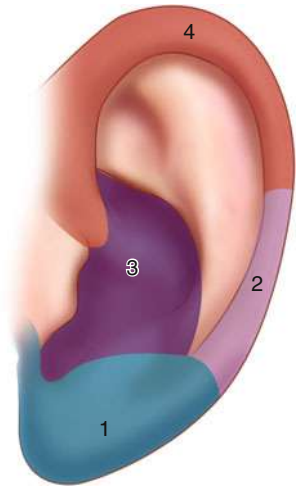

Fig. 89.8 Auricular block anesthetizes four nerves that innervate the auricle. *1* Great auricular nerve, *2* lesser occipital nerve, *3* auricular branch of vagus nerve, *4* auriculotemporal nerve

Fig. 89.9 Auricular ring block technique

89.7.3 Wrist Block (Fig. 89.10) [3, 5, 6]

- Indications
  - To anesthetize the hand in preparation for laceration repair, fracture or dislocation reduction, or pain relief

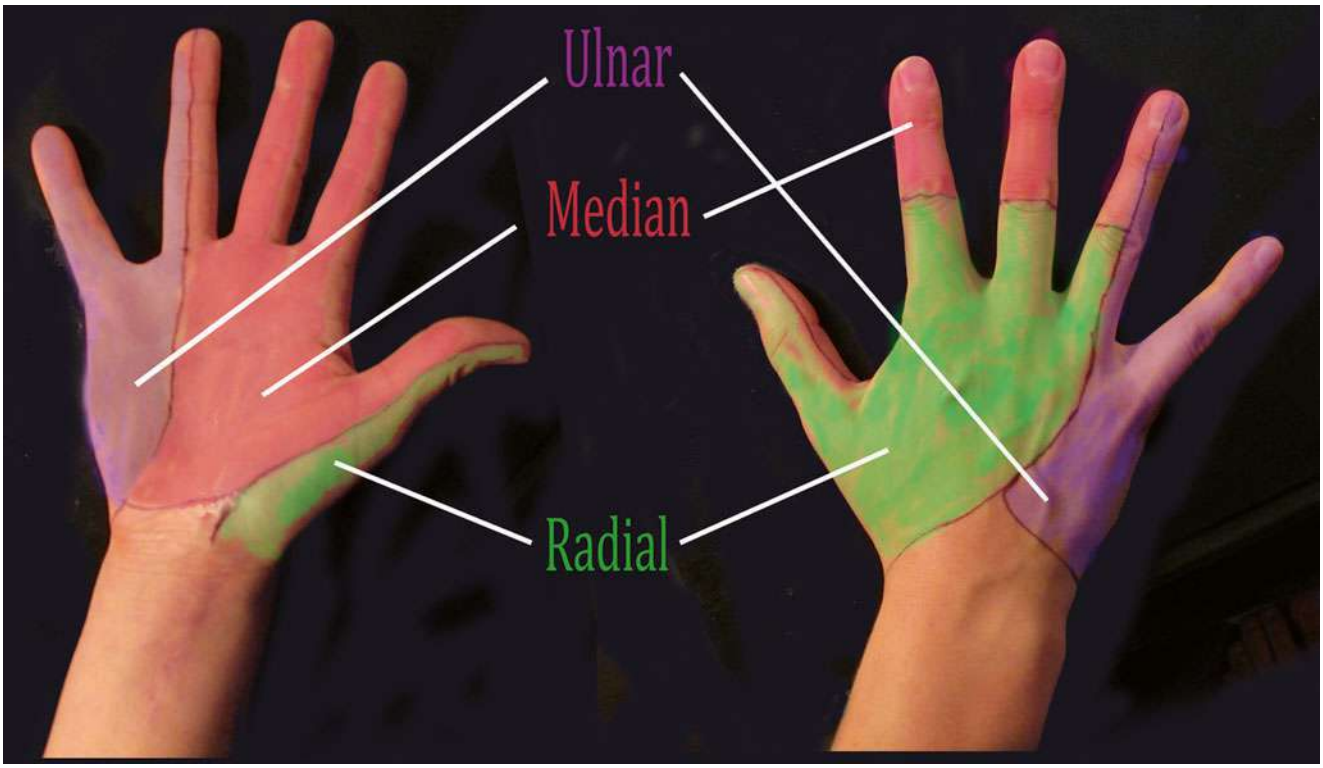

Fig. 89.10 Nerve distribution in the hand

89.7.3.1 Wrist Block: Median Nerve (Fig. 89.11)

- Procedure
  1. Position the patient supine with the palmar surface of the hand face up.
  2. Have the patient make a fist and slightly flex the wrist so that the palmaris longus and flexor carpi radialis tendons become prominent.
  3. Create a skin wheal of local anesthetic between the two tendons between the proximal skin crease and the distal skin crease at the wrist. Alternatively, anesthetic can be injected in line with the ulnar styloid process at the proximal skin crease.
  4. The palmaris longus tendon is absent normally in 10–20 % of the population. In this case, inject over the midpoint of the proximal skin crease at the level of the styloid process.

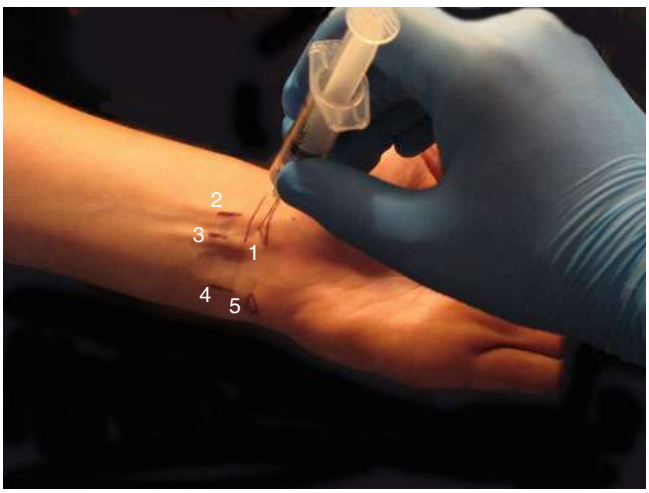

Fig. 89.11 Median nerve block at the wrist: *1* proximal and distal wrist creases radial artery, *2* flexor carpi radialis tendon, *3* palmaris longus tendon, *4* ulnar artery, *5* styloid process

89.7.3.2 Wrist Block: Radial Nerve (Figs. 89.12 and 89.13) [3, 7]

- Procedure
  1. To block the radial nerve, block multiple peripheral branches on the dorsal and the radial aspects of the lateral wrist.
  2. A field block in and around the anatomical snuffbox may be performed, requiring roughly 5–6 mL of local anesthetic. This anesthetizes the terminal branches of the radius arising from the forearm.
  3. Position the patient's palm face up and inject a skin wheal to the area 1 mm lateral to the radial pulse and in line with the proximal wrist crease. Inject 2 mL of local anesthesia with a 25- to 27-gauge needle. This anesthetizes the terminal trunk of the radial nerve.

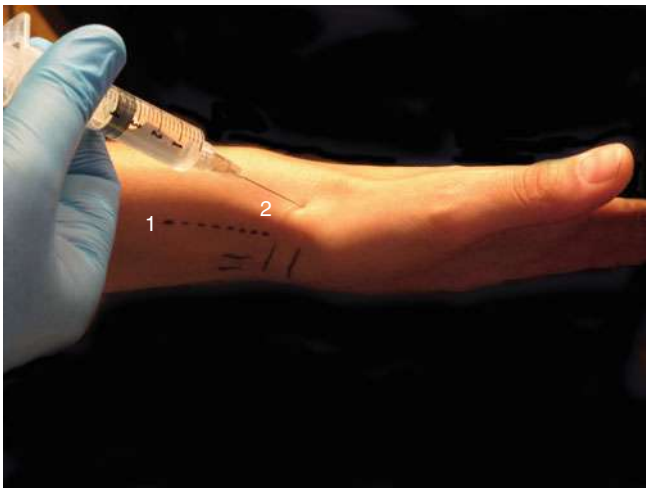

Fig. 89.12 Radial nerve block at the wrist: *1* radial artery, *2* anatomical snuffbox

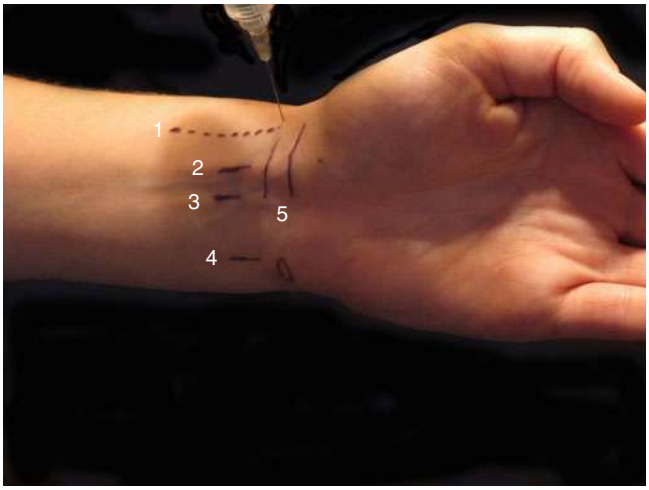

Fig. 89.13 Radial nerve block at the wrist: *1* radial artery, *2* flexor carpi radialis tendon, *3* palmaris longus tendon, *4* ulnar artery, *5* proximal and distal wrist creases

89.7.3.3 Wrist Block: Ulnar Nerve (Fig. 89.14)

- Procedure
  1. Position the patient with the hand face up.
  2. Palpate the styloid process of the ulna and pisiform and find the ulnar artery pulse.
  3. Create a skin wheal of local anesthetic between lateral to the ulnar artery and medial to the flexor carpi ulnaris tendon just proximal to the styloid process. This is at the level of the proximal wrist crease.

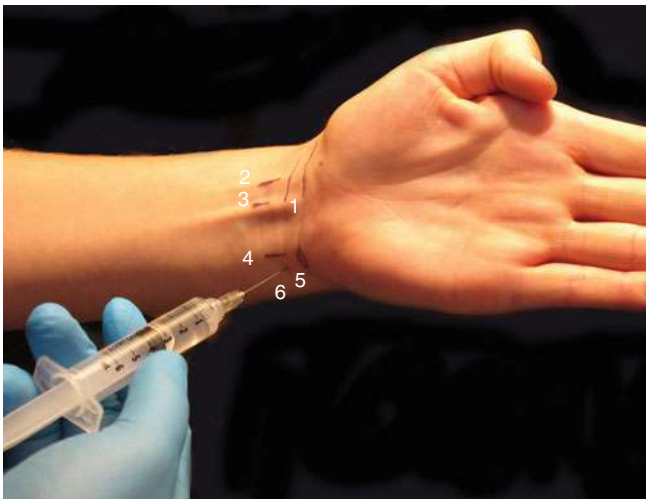

Fig. 89.14 Ulnar nerve block at the wrist: *1* proximal and distal wrist creases, *2* flexor carpi radialis tendon, *3* palmaris longus tendon, *4* ulnar artery, *5* styloid process, *6* flexor carpi ulnaris

89.7.4 Digital Nerve Blocks: Ring, Web Space, and Tendon Sheath (Fig. 89.15) [3, 8, 9]

- Indications
  - To anesthetize the digits in preparation for laceration repair, nail bed repair, joint reduction, or pain relief

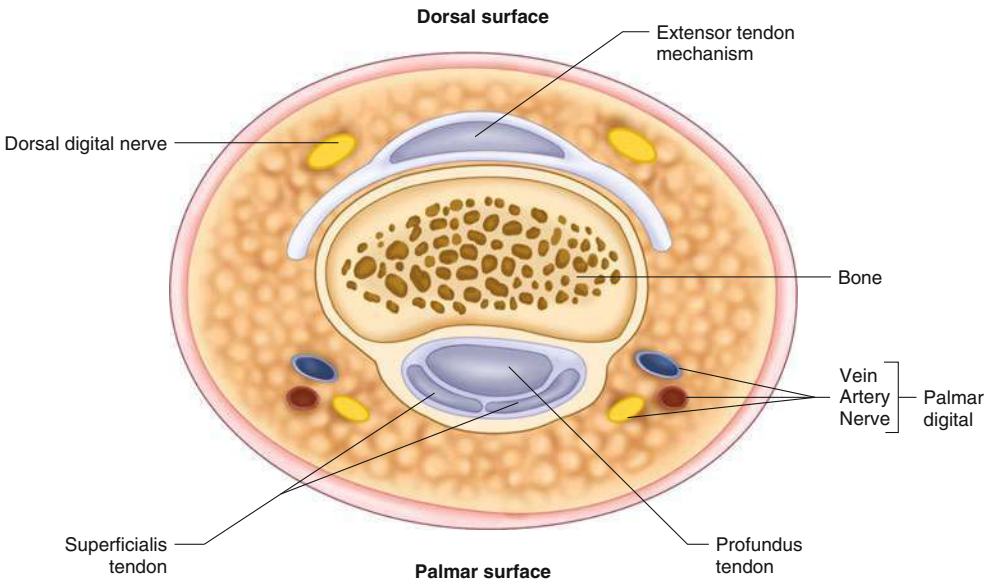

Fig. 89.15 Cross section of finger

89.7.4.1 Ring Block (Fig. 89.16) [3]

- Procedure
  1. Insert a 25-gauge needle on the dorsal surface of the proximal phalanx of the digit to be anesthetized. Inject 1 mL along the dorsal surface and withdraw the needle.
  2. Reinsert the needle again perpendicular to the last injection and running on the lateral surface of the phalanx. Inject 1–1.5 mL of local anesthetic to just past the phalanx base.
  3. Repeat the injection in the same fashion on the medial aspect of the phalanx.
  4. Do not inject more than 5 mL into a digit.
  5. Toe blocks are similar to finger ring blocks, except that the great toe requires plantar surface injection as well, owing to its unique nerve supply.

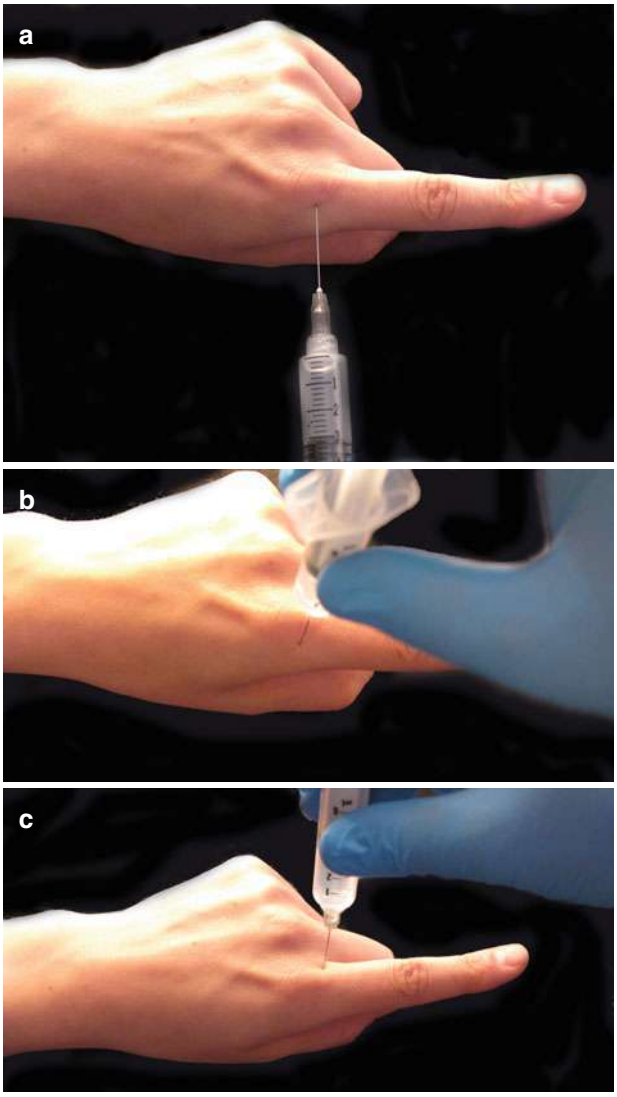

Fig. 89.16 Dorsal surface injection of digital nerve ring block. Digital nerve block is performed by injecting onto the (**a**) dorsal, (**b**) lateral, and (**c**) medial surfaces of the proximal phalanx

89.7.4.2 Web Space Digital Block (Fig. 89.17) [9]

- Procedure
  1. Have the patient abduct the fingers.
  2. Palpate the metacarpophalangeal joint and then insert a 25- to 27-gauge needle into the lateral web space subcutaneously, directing it dorsally. Aspirate then inject 1 mL of local anesthetic.
  3. Withdraw the needle but before the exiting skin, redirect toward the palmar aspect until the tip is next to the metacarpophalangeal joint, and inject 1 mL of local anesthetic.
  4. Repeat the procedure on the medial web space of the digit. Each digit blocked requires injection on both the lateral and the medial web spaces.

Fig. 89.17 Web space approach to the digital block, requiring injection on the medial and lateral web spaces for a blocked digit

89.7.4.3 Intrathecal Digital Block: Flexor Tendon Sheath (Fig. 89.18) [8]

- Procedure
  1. Inject anesthetic directly into the flexor tendon sheath. Palpate on the palmar surface over and proximal to the metacarpophalangeal joint. Gentle flexion of digit may better reveal the sheath. Have the patient abduct the fingers.
  2. Insert a 25-gauge needle at a 45° angle to the skin and along the long axis of the digit directly into the flexor tendon sheath at the level of the distal skin crease.
  3. Inject 2 mL of local anesthetic. The anesthetic should flow freely if it is in the sheath. If it does not, it is likely in the tendon and should be withdrawn slightly.
  4. Contraindications to intrathecal block are local infection and preexisting flexor tendon injury.
  5. Risk of tenosynovitis; sterilize the skin before introducing the needle.
  6. If laceration has involved the tendon, anesthetic may leak from the wound.

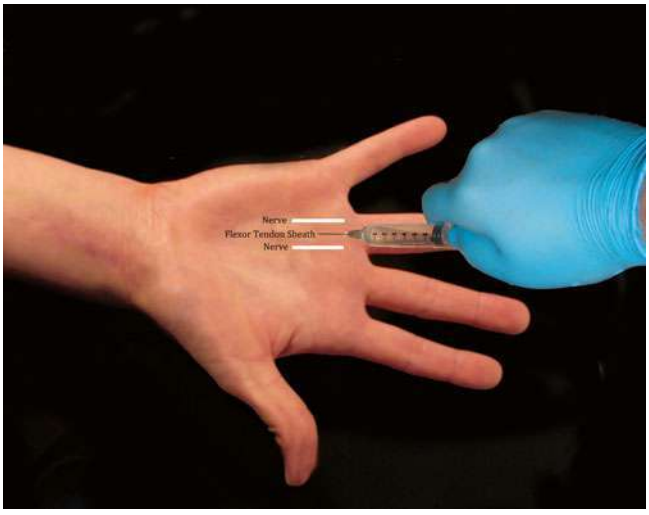

Fig. 89.18 Intrathecal/tendon sheath approach to the digital block

Acknowledgment The authors would like to thank Katy Howard for serving as the subject in many of the photographs in this chapter.

References

1. Muck AE, Bebarta VS, Borys DJ, Morgan DL. Six years of epinephrine digital injections: absence of significant local or systemic effects. Ann Emerg Med. 2010;56:270–4.
2. Waterbrook AL, Germann AC, Southall JC. Is epinephrine harmful when used with anesthetics for digital blocks? Ann Emerg Med. 2007;50:472–5.
3. Reichman EF, Simon RR, editors. Emergency medicine procedures. New York: McGraw-Hill Medical; 2004. p. 961–3.
4. Benko K. Fixing faces painlessly: facial anesthesia in emergency medicine. Available at: Emergency Medicine Practice (ebmedicine.net). 2009;11.
5. Rosh AJ. Ear anesthesia. Medscape reference. Available at: http://emedicine.medscape.com/article/82698-overview#a15.
6. Butterworth JF. Atlas of procedures in anesthesia and critical care. Philadelphia: WB Saunders; 1992. p. 160–4.
7. Brown DL. Atlas of regional anesthesia. Philadelphia: WB Saunders; 1992. p. 52.
8. Morrison WG. Transthecal digital block. Arch Emerg Med. 1993;10:35–8.
9. Mueller J, Davenport M. Digital nerve block (web space and tendon sheath). New York: McGraw-Hill's Access Emergency Medicine. Available at: http://www.accessemergencymedicine.com/videosPDF/DigitalNerveBlock.pdf.

90 Wound Management and Hemostasis

Rich Teitell and Muhammad Waseem

90.1 Indications for Primary Wound Repair

- Wounds that have been sustained less than 8 h before presentation may be closed primarily.
- If there is suspicion of broken glass or the presence of a foreign body, the patient should receive an x-ray before primary wound repair. Of note, x-ray has poor sensitivity for radiolucent foreign bodies such as wood and plastic.
- Wounds that are free of foreign bodies or show no gross evidence of infection or dead tissue may be closed primarily.

90.2 Contraindications for Primary Wound Repair

- History must address the factors that predispose a wound to greater risks of infection and not implement primary wound closure in such wounds that have had a duration greater than 8–12 h or contamination with saliva, stool, or foreign matter and a wound that has been sustained by blunt or crush mechanism.
- When wounds are highly contaminated with debris or if devitalized tissue is present, the wound should be left open for 3–4 days and then reevaluated.
- After a proper physical examination has been performed that assesses distal pulse and sensory and motor function of the surrounding region and distal extremity, if any of these are compromised, appropriate orthopedic and/or vascular consultation should be sought before wound closure, especially in the setting of a suspicion for open fracture.
- Consider delayed closure in wounds over joints.
- Wounds that have a significant loss of skin should be left to heal by secondary intention.

90.3 Procedure for Wound Closure and Hemostasis

1. Wound should be thoroughly prepared with Betadine (10 % povidone with 1 % iodine), which has a rapid onset with greatest antimicrobial effect.
2. Skin surrounding the wound should be covered with sterile drapes (Fig. 90.1) so as to reduce additional contamination within the open wound.
3. The most important determinant for preventing wound infection is adequate high-pressure irrigation.
4. Normal saline is the best solution for irrigation. Use pressures above 7 psi and an 18-gauge needle with a syringe of 30 mL or more; this gauge needle will achieve pressure greater than 7 psi.
5. It has been suggested that tap water irrigation may be just as effective. Because this approach will reduce costs, it can be considered.
6. Sharp debridement should be performed when foreign matter is found in an open wound that does not easily rinse out with saline. In this circumstance, consider wound closure by secondary intention or delayed primary closure.
7. At this point, wounds that do not meet criteria for primary closure must be left open, covered with a sterile dressing, and reevaluated in 3–4 days for consideration of delayed primary closure versus healing by secondary intention.
8. Wounds that do meet criteria as previously discussed for primary closure should be closed with either sutures or staples.
9. Staples are the preferred method of closure of scalp wounds and can be used under low wound tension.

R. Teitell, MD
Department of Emergency Medicine, Waterbury Hospital, Waterbury, CT, USA
e-mail: rjt0013@gmail.com

M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_90

10. The angle of the stapler is critical and must be positioned perpendicular to the intended area of delivery. A two-person approach can be utilized for optimal placement of the staple, using forceps to approximate the edges of the tissue.
11. Synthetic or monofilament sutures have a lower rate of infection than braided sutures (for closure techniques with sutures, see Chap. 92).
12. Dermabond should be utilized only for the most superficial linear lacerations and confers the benefit of efficiency of time and less discomfort to the patient. However, it should be mentioned that Dermabond does not offer any statistically significant difference in cosmesis when compared with suturing or other adhesives (see Chap. 93).
13. When a wound presents with deep underlying tissue involvement and direct bleeding is evident, direct pressure may be used to achieve hemostasis within a wound. Surgicel may also be applied with pressure to achieve control of small capillary and venous oozing (Fig. 90.2).
14. When a persistent bleed is visible within a wound that does not respond to direct pressure, a figure-of-eight suture (Fig. 90.3) may be used.
15. If a bleeding vessel is visible within a wound, an alternate method of hemostasis may be achieved by utilizing a hemostat to clamp the vessel and then tying an absorbable suture (Fig. 90.4).

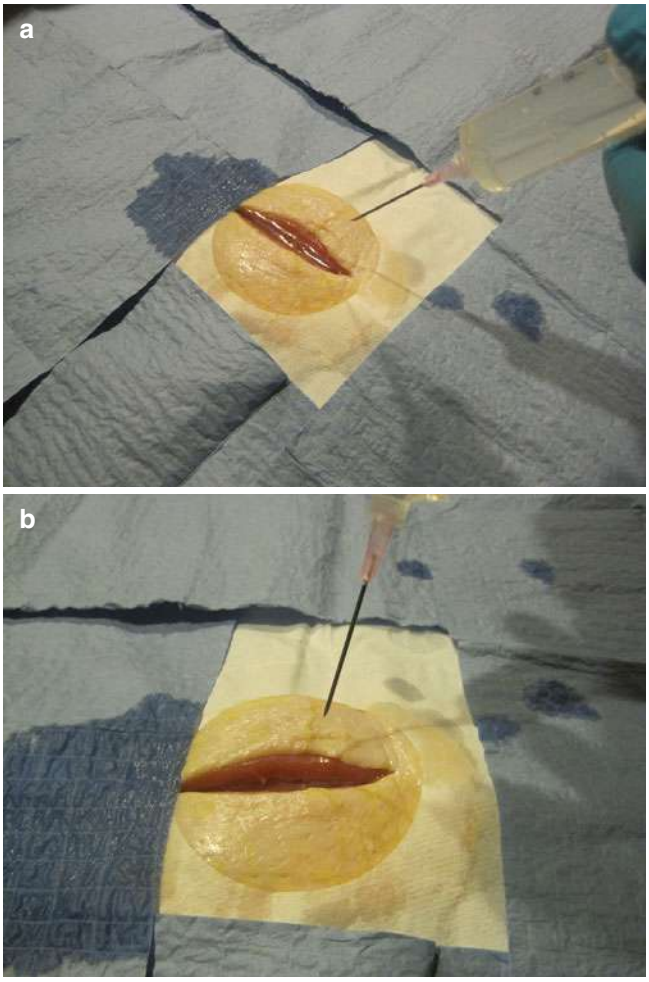

Fig. 90.1 (**a**) High-pressure irrigation using an 18-gauge needle (**b**)

Fig. 90.2 Direct pressure with or without Surgicel may be used to achieve hemostasis within a wound and also to achieve control of small capillary and venous oozing

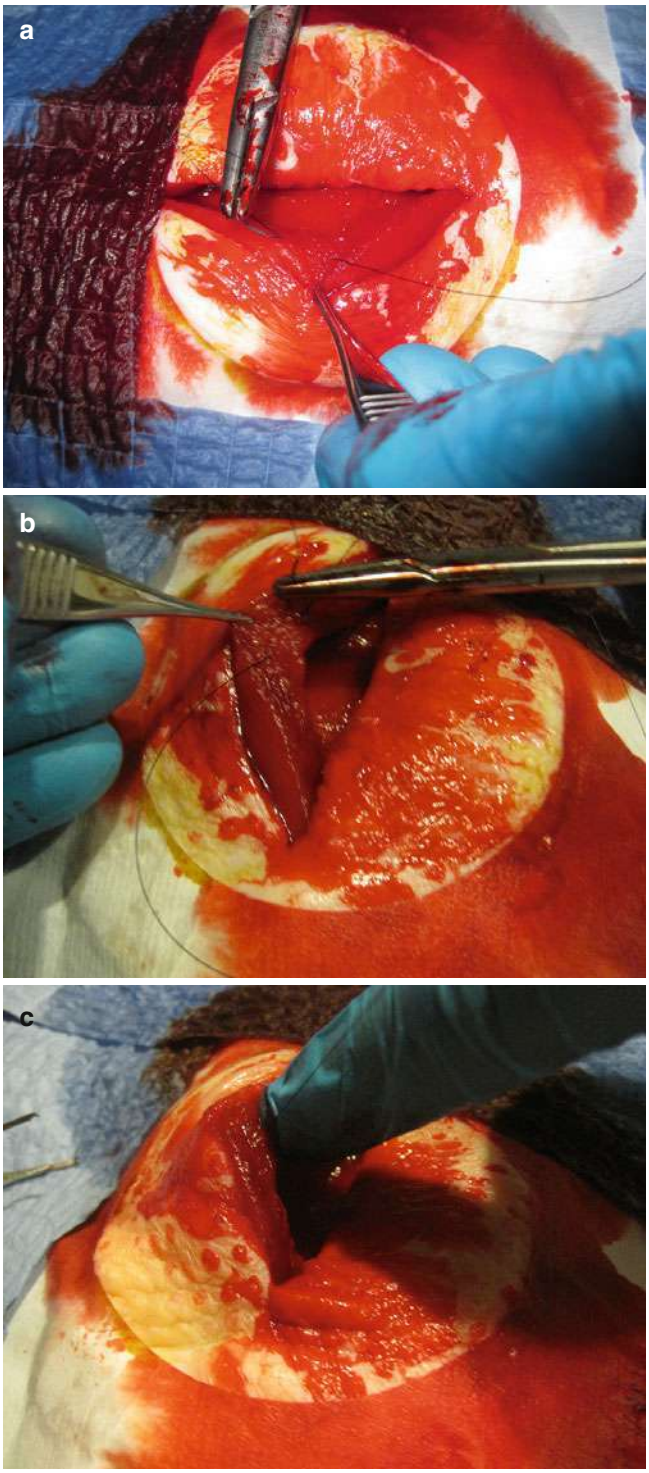

Fig. 90.3 (**a**) The location of the bleeding vessel in the subcutaneous tissue is identified and an absorbable suture is threaded through the region. (**b**) In a diagonal direction, a second suture is placed, which when tied will form the shape of an eight. (**c**) Once tied, the suture forms a figure eight in the tissue and hemostasis is achieved

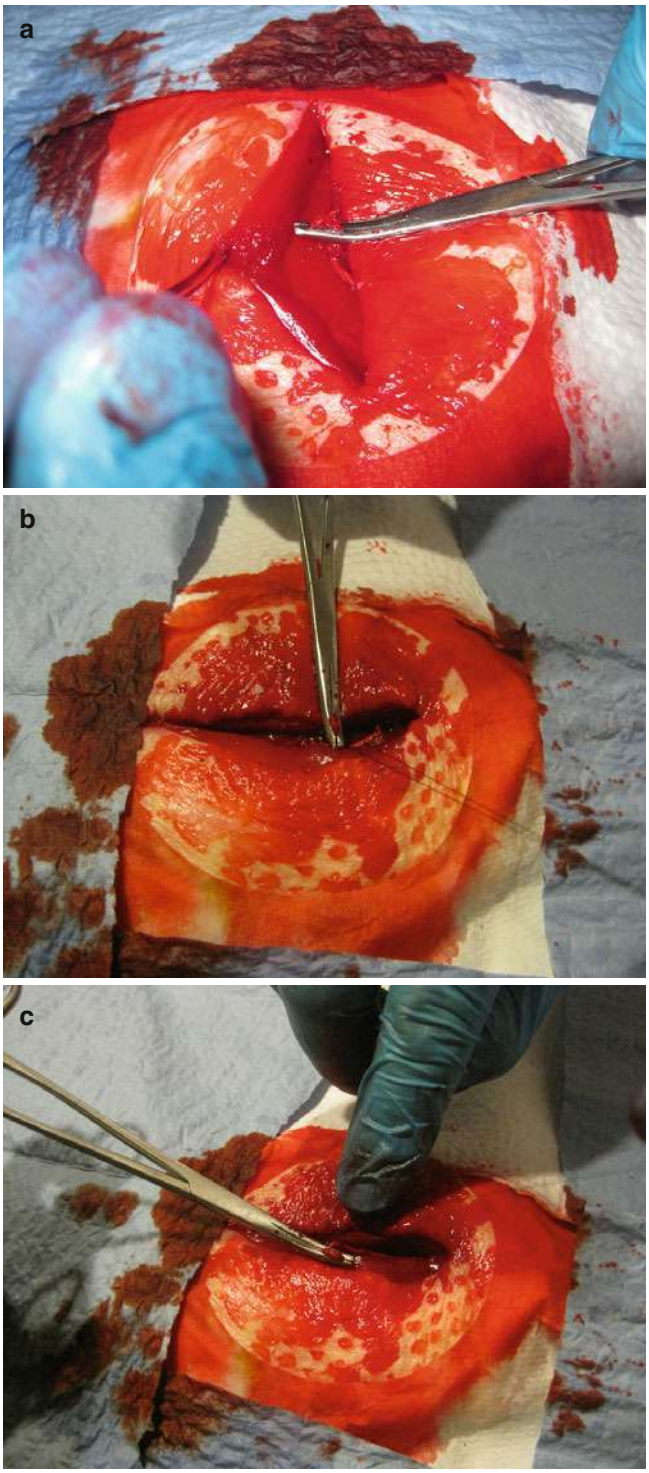

Fig. 90.4 (**a**) The culprit vessel is identified and clamped with a hemostat. (**b**) An absorbable suture is wrapped around the vessel inferior to the hemostat. (**c**) Once the first tie is placed being driven down with index finger, the hemostat may be removed and the knot can be completed for a total of three or four passes

90.4 Complications

- Hematoma formation
- Infection
- Scar formation

90.5 Pearls and Pitfalls

- Pearls
  - Antibiotic coverage should be employed judiciously depending on the wound and location.
  - Wounds to the face that are not heavily contaminated generally do not require antibiotics because the face and scalp have a significant vascular supply.
  - Patients with diabetes have higher incidences of wound infection, and the physician should consider prophylactic antibiotics in such individuals.
  - Wounds that are clean-contaminated and sutured in the emergency department should have a nonadherent dressing applied for 48 h until epithelialization has occurred. After that, the dressing may be removed, and soap and water gentle washing should commence to keep area clean. Only small amounts of bacitracin should be applied to wound in the interim period before sutures are removed.
- Pitfalls
  - Wounds located on the extremities have greater incidences of infection than those on the trunk or head/face.

Selected Reading

Blankenship RB, Baker T. Imaging modalities in wounds and superficial skin infections. Emerg Med Clin North Am. 2007;25:223–34.

Dimick AR. Delayed wound closure: indications and techniques. Ann Emerg Med. 1988;17:1303–4.

Edlich RF, Rodeheaver GT, Morgan RF, Berman DE, Thacker JG. Principles of emergency wound management. Ann Emerg Med. 1988;17(12):1284–302.

Farion K, Osmond MH, Hartling L, et al. Tissue adhesives for traumatic lacerations in children and adults. Cochrane Database Syst Rev. 2002;(3):CD003326.

Hall S. A review of the effect of tap water versus normal saline on infection rates in acute traumatic wounds. J Wound Care. 2007;16:38–41.

Marx J, Hockberger R, Walls R, editors. Rosen's emergency medicine: concepts and clinical practice. 6th ed. Philadelphia: Mosby-Elsevier; 2006. p. 702–3.

Ritchie AJ, Rocke LG. Staples versus sutures in the closure of scalp wounds: a prospective, double-blind, randomized trial. Injury. 1989;20:217–8.

Singer AJ, Hollander JE, Quinn JV. Evaluation and management of traumatic lacerations. N Engl J Med. 1997;337:1142–8.

Burn Care

91

Thomas Parry and Jeffrey Pepin

Burns result from exposure to heat, caustic chemicals, electricity, or radiation. Damage to the natural barrier provided by skin results in rapid fluid losses and risk of infection. Permanent scarring is a common long-term complication. They may also be complicated by sensory deficits due to loss of nerve connections and limb loss due to circulatory compromise.

91.1 Burn Description (Fig. 91.1)

- First degree or superficial: burn that remains confined to the epidermis (e.g., sunburn)
- Second degree or partial thickness: burn that extends into the dermis (e.g., blistering scald burns)
- Third degree or full thickness: burn involving the entire depth of the dermis and epidermal appendages

Fig. 91.1 Tea scald over the chest and shoulder of a child showing heterogeneity of burn depth. *D* deep (second or third degree), *I* intermediate (second degree), *S* superficial (first degree) (Reproduced from: Enoch S, Roshan A, Shah M. Emergency and early management of burns and scalds. *BMJ*. 2009;338:b1037, with permission from BMJ Publishing Group Ltd.)

T. Parry, MD (✉)
Department of Emergency Medicine, Lincoln Hospital and Mental Health Center, New York, NY, USA
e-mail: Thomas.parry@nychhc.org

J. Pepin, MD
Department of Emergency Medicine, University of Minnesota Medical Center Fairview, Minneapolis, MN, USA
e-mail: medicone76@hotmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_91

91.2 Indications

- Superficial, deep second degree, and third degree.
- First-degree burns generally need only supportive care.

91.3 Materials and Medications

- Appropriate analgesia
- Antibiotic ointment
- Silver sulfadiazine (avoid on face)
- Sulfamylon for ears
- Alternative agents include bacitracin or polymyxin B ointments
- Cleansing solution such as chlorhexidine prep and water
- Basin
- Petroleum gauze
- Several 4×4 gauze pads
- Rolled gauze
- Tape
- Gloves

Airway, breathing, and circulation and cervical spine should be immediately assessed before any burn wound management. It is important to remember that any patient suspected of an inhalation injury or carbon monoxide poisoning should receive 100 % humidified oxygen until carboxyhemoglobin returns to normal. Inhalation injuries may not present themselves until after fluid resuscitation has already been started, so early intubation is recommended in these cases.

91.4 Sizing

Total body surface area (TBSA) of body parts is estimated by multiples of 9 (Rule of Nines)

- Adults
  - Head and neck: 9
  - Arms: 9 each
  - Legs: 18 each
  - Trunk: 18 front and 18 back
  - Perineum and palms: 1
- Infants/children
  - Head and neck: 18
  - Arms: 9 each
  - Legs: 14 each
  - Trunk: 18 front and 18 back

A second way to estimate TBSA in smaller burns is by using the palm surface area: using the patient's palm as a guide, each palmar surface equals 1 %.

A third way to estimate TBSA is via the use of the Lund and Browder chart (Fig. 91.2).

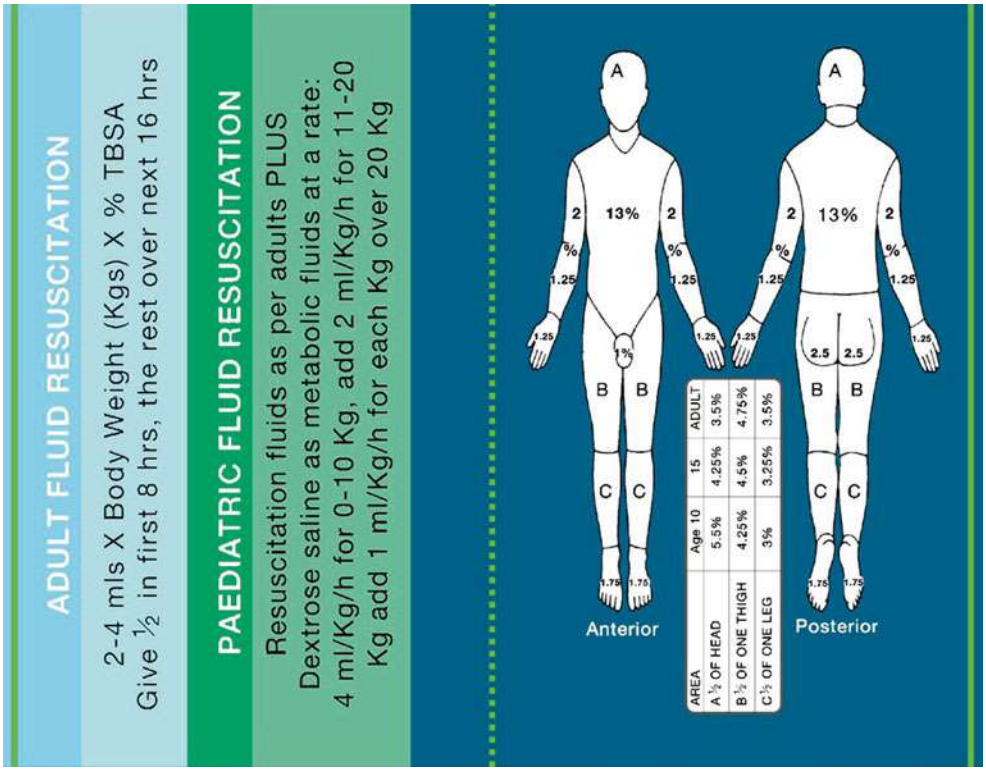

Fig. 91.2 Lund Browder chart for estimating total body surface area of burns, with suggested fluid resuscitation guidelines (Reprinted from Malic CC, Karoo RO, Austin O, Phipps A. Resuscitation burn card–a useful tool for burn injury assessment. *Burns*. 2007;33(2)195–9, with permission from Elsevier)

91.5 Fluid Resuscitation

Because fluid resuscitation is absolutely essential in the early aspects of burn care, the Parkland formula has been used to estimate fluid requirements in burn patients. The patient's weight in kilograms is multiplied by the percent body surface area (BSA) involved; this number is multiplied by 4 mL of lactated Ringer's solution. Half of this amount is given during the first 8 h and the remaining amount is given over the next 16 h of resuscitation. The goal is to keep urine output approximately 0.5–1.0 mL/kg/h.

In pediatric cases, the Galveston formula may be used instead: 5000 mL/m^2 BSA burned plus 2000 mL/m^2. TBSA of 5 % dextrose in lactated Ringer's solution intravenously over the first 24 h, half in the first 8 h, and the other half over the next 16 h.

91.6 Procedure

1. Provide appropriate analgesia.
2. Clean the wound with antiseptic solution and water (if it is a dry chemical burn, make sure to brush off as much chemical before using copious amounts of water to clean the wound).
3. Debride any loose skin or foreign debris using a dry 4×4 or rolled gauze.
4. Apply antibiotic ointment or cream to the burn.
5. Apply petroleum gauze in a single layer just over the affected skin.
6. Use loose 4×4 gauzes and "fluff" them to make a thick layer of padding to place over the petroleum gauze.
7. Either wrap the entire area with a rolled gauze or tape a small layer of 4×4 gauze over the "fluffed" layer of 4×4 gauze.

91.7 Complications

- Wound infection
- Nonhealing wound requiring skin graft (deep second-degree and third-degree burns)
- Compartment syndrome (circumferential burns may require escharotomy)
- Rhabdomyolysis

91.8 Pearls and Pitfalls

- Pearls
  - Determination of the depth of burns on initial presentation is difficult (especially when covered with petroleum). A good rule of thumb is if it blanches and/or hurts, it is a partial thickness burn.
  - First-degree burns are not included in burn size estimations for fluid resuscitation calculation.
  - Burn size determines fluid requirements and transfer decision.
  - Burn of greater than 20 % TBSA should receive intravenous fluids.
  - Patients who have inhalation injury, greater than 10 % TBSA, high-voltage electrical injury, or chemical burns should be referred to burn unit.
  - Fluid resuscitation should be adjusted according to physiological response such as urine output (30–50 mL/h in adults and 1 mL/kg/h in children).
  - Assure tetanus is up to date.
- Pitfalls
  - Keep the patient warm in the first few hours. There is no need to apply ice.
  - All jewelry and rings should be removed.
  - Prophylactic antibiotics are not recommended.
  - Silver sulfadiazine should be avoided in facial burns because of the risk of staining of the skin.

Blister care is a very controversial topic. Current research suggests that it may be beneficial to keep the blister intact unless it appears to be tense or over a joint. Most blisters will rupture in 2–4 days. Ruptured blisters should be debrided with all the extra skin removed. Most burn units will scrub everything off once they receive a patient.

Wounds should be kept clean to prevent an environment that will increase the chances of infection. Wrap in saline-soaked sterile gauze prior to transfer.

Dressing changes should be done daily with all previously applied antibiotic ointment removed before a reapplication of new ointment. It is important to provide analgesic 30 min before a dressing change.

91.9 Admission Criteria

- Partial-thickness burns of noncritical areas not including the eyes, ears, face, hands, feet, or perineum that total a BSA of 10–20 % in adults
- Partial-thickness burns of noncritical areas involving 5–10 % of BSA in children younger than 10 years
- Suspicious of non-accidental trauma
- Patients unable to care for wounds in outpatient settings
- Prompt referral to a burn specialist is required in the following cases:
  - Partial-thickness and full-thickness burns greater than 10 % of the TBSA in patients younger than 10 years or older than 50 years of age.
  - Partial-thickness and full-thickness burns greater than 20 % of the TBSA in other age groups.

- Partial-thickness and full-thickness burns involving the face, eyes, ears, hands, feet, genitalia, or perineum or the skin overlying major joints.
- Full-thickness burns greater than 5 % TBSA in any age group.
- Electrical burns, including lightning injury: significant volumes of tissue beneath the surface may be injured and result in acute renal failure and other complications.
- Significant chemical burns.
- Inhalation injury.
- Burn injury in patients with preexisting illness that could complicate management, prolong recovery, or affect mortality.
- Any burn patient in whom concomitant trauma poses an increased risk of morbidity or mortality may be treated initially in a trauma center until stable before transfer to a burn center.
- Children with burns seen in hospitals without qualified personnel or equipment for their care should be transferred to a burn center with these capabilities.
- Burn injury in patients who will require special social and emotional or long-term rehabilitative support, including cases involving suspected child abuse and neglect.

Selected Reading

Bezuhly M, Fish JS. Acute burn care. Plast Reconstr Surg. 2012;130(2):349e–58.

Rex S. Burn injuries. Curr Opin Crit Care. 2012;18(6):671–6.

Wasiak J, Cleland H, Campbell F, Spinks A. Dressings for superficial and partial thickness burns. Cochrane Database Syst Rev. 2013;(3):CD002106.

Wound Closure

92

Oliver Michael Berrett, Jeffrey Joseph Harroch, Karlene Hosford, and Muhammad Waseem

92.1 Indications

- Open wound of skin or mucosal tissues
- Purpose
  - Preserve function
  - Control bleeding
  - Promote healing
  - Cosmesis

92.2 Contraindications

- Wounds caused by animal or human bites
- Contaminated, infected, or puncture wounds
- Complex wounds (may require operating room)

92.3 Methods

- Suture placement
- Tissue adhesives
- Adhesive tapes
- Staples

92.4 Preparation

- Obtain a complete history of injury
  - Mechanism
  - Time since injury
  - Tetanus status
  - Comorbidities
- Examine the extent of the wound, remove any contaminants, debride devitalized tissue, inspect for foreign bodies. Obtain a radiograph if foreign body is suspected.
- Copious irrigation with normal saline.
- Clean around the wound with a povidone-iodine solution.
- Drape the area in sterile manner.
- Inject the anesthetic agent into the wound edges using a 25- or 27-gauge needle. The most common agent is lidocaine 2 % with or without epinephrine 1 % (maximum dose is 3 mg/kg without epinephrine and 5 mg/kg with epinephrine).
- Close the wound using the appropriate technique (see later).
- Dress the wound.
- Update tetanus and diphtheria vaccination if needed.

O.M. Berrett, MD • K. Hosford, MD • M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: omberrett@gmail.com; krlhos@aol.com; waseemm2001@hotmail.com

J.J. Harroch, MD
Department of Emergency Medicine, University of Miami Miller School of Medicine, University of Miami Hospital, Miami, FL, USA
e-mail: jharroch@med.miami.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_92

92.5 Suture Repair

92.5.1 Materials and Medications

- Commercial kits commonly contain all the following except the suture material (Fig. 92.1)
  - Povidone-iodine solution
  - Normal saline for irrigation
  - 5- to 12-mL syringe with a 25-gauge needle
  - Anesthetic agent
  - Needle holder
  - Pickups
  - Suture scissors
  - Suturing material
  - Sterile drape or sheet, gloves, gauze

Fig. 92.1 Suture kit

92.5.2 General Guidelines

- Local anesthetic lidocaine 1 % or lidocaine 1 % with epinephrine.
- Minimize direct use of instruments on the tissues.
- Wound edges should be everted to maximize healing and cosmetic effect. This is achieved by inserting the needle at 90° to the skin.
- Sutures should be evenly spaced, placed 1–3 mm apart and 2 mm from the wound edge.
- Optimal tension is achieved by tying the sutures so the edges lightly approximate.

92.5.3 Suture Material

- Nonabsorbable
  - Silk: for specialty use, reactive and weak
  - Nylon (Ethilon), polypropylene (Prolene): good strength, good overall material for cutaneous wounds
  - Polypropylene: good strength, difficult to use
  - Require removal at a specified time
- Absorbable
  - Undergo rapid degradation in tissues, losing their tensile strength within 60 days
  - Indication: buried suture to reduce wound edge tension
  - Vicryl: subcutaneous placement, mucous membranes
  - Chromic: use for intraoral lacerations
  - Removal not required

92.5.4 Suture Size

- 0.0–2.0: thick material for large wounds, trunk
- 0.3–0.4: used on medium-sized wounds, extremities, scalp
- 0.5–0.6: fine sutures, used on facial wounds

92.5.5 Suture Techniques

- Simple interrupted sutures (Fig. 92.2)
  - Most common method
  - Position needle 2 mm from the wound edge at a 90° angle. Enter the needle into the skin and arc through the wound edge and into the opposing edge at the same level, exiting the skin on the opposing side 2 mm from wound edge, tie.
- Deep dermal suture (Fig. 92.3)
  - Also known as buried sutures, used to minimize tension in a wound.
  - Use absorbable suture material.
  - The needle is placed at the base of the wound wall and arched upward, exiting the ipsilateral wall more superficially. The needle is then directed across to the opposing wound wall at the same level and directed downward, exiting deep, tie. Note: The knot will be deep in the wound.
- Simple running suture (Fig. 92.4)
  - This method provides rapid closure of long and relatively linear lacerations.
  - Place the initial suture in the same manner as a simple suture, tie, cut the free strand, and leave the needle attached. Reintroduce the needle into the skin on the opposite side so the suture crosses the wound superficially at a 65° angle. The needle is then inserted perpendicular to the skin, emerging on the opposite side about 3 mm from the wound edge. Repeat without tying until closure is complete, maintaining appropriate tension. When the last stitch is placed, leave a loose loop of suture on one side so that both ends can be tied together.
- Vertical mattress suture (Fig. 92.5)
  - Provides the benefits of both simple and deep techniques. For use on deeper, gaping wounds and wounds over high-tension areas such as joints.
  - Position the needle 1 cm from skin edge, at a 90° angle, drive a deep arc perpendicular through the wound, exiting the skin on the opposing side the same distance from the wound edge. Next, reinsert the needle on the ipsilateral side 2 mm from the edge, emerging on the opposing side and approximate, tie.
- Horizontal mattress suture (Fig. 92.6)
  - For large wounds with tension.
  - Place the initial suture in the same manner as a simple suture, only do not tie. Reposition the needle on the ipsilateral side, horizontally 5 mm to the side of the exit at the same distance from the wound edge and drive through to the other side, tie. Note: The tie lies parallel to the wound.
- Half-buried horizontal mattress suture (Fig. 92.7)
  - Used on wounds with skin flap.
  - On one side, drive the needle percutaneously, and then pass horizontally through the dermal tissue of the tip of flap, finally passing into the dermis of the opposing edge and exiting the skin, tie.

92.5.6 Recommendations

- Face
  - 5.0–6.0 nylon.
  - Remove after 3–5 days.
- Scalp
  - 2.0–3.0 nylon or staple.
  - Remove at 8–10 days.

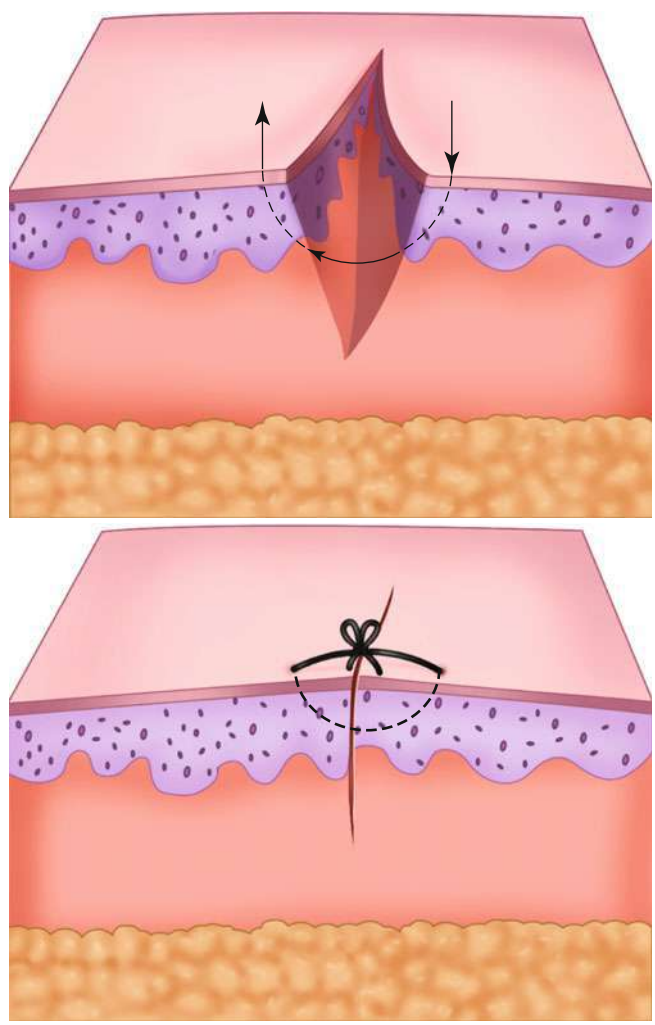

Fig. 92.2 Simple suture

- Hand
  - 4.0–5.0 nylon, consider vertical or horizontal mattress.
  - 5.0–6.0 Monocryl for nail bed.
  - Remove at 10–14 days, use for a longer time if directly over the joint.
- Extremity
  - Nonmobile skin: 3.0–4.0 nylon, remove at 8–10 days
  - Over joint: 3.0–4.0 nylon, remove at 10–14 days
- Trunk
  - Anterior trunk: 3.0–4.0 nylon, remove at 8–10 days.
  - Posterior trunk: 2.0–3.0 nylon, remove at 10–14 days.
  - Consider staples.
- Oral mucosa
  - Thin mucosa: 4.0 Vicryl will absorb; duration, 5–7 days
  - Tongue: 3.0 Vicryl will absorb; duration, 5–7 days

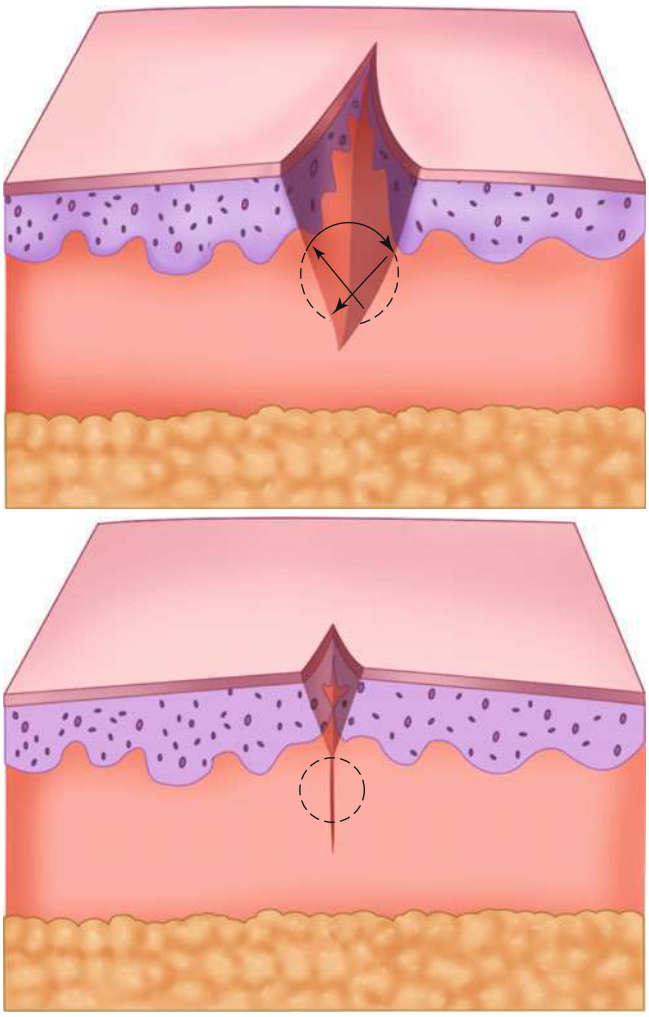

Fig. 92.3 Deep suture

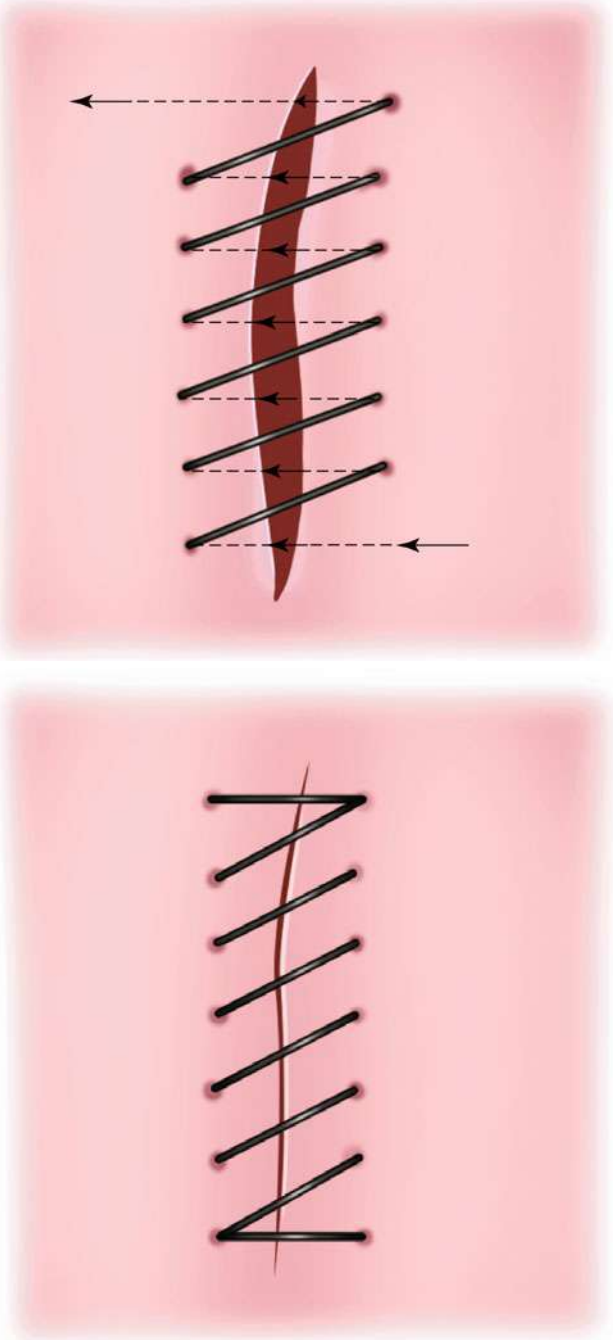

Fig. 92.4 Running suture

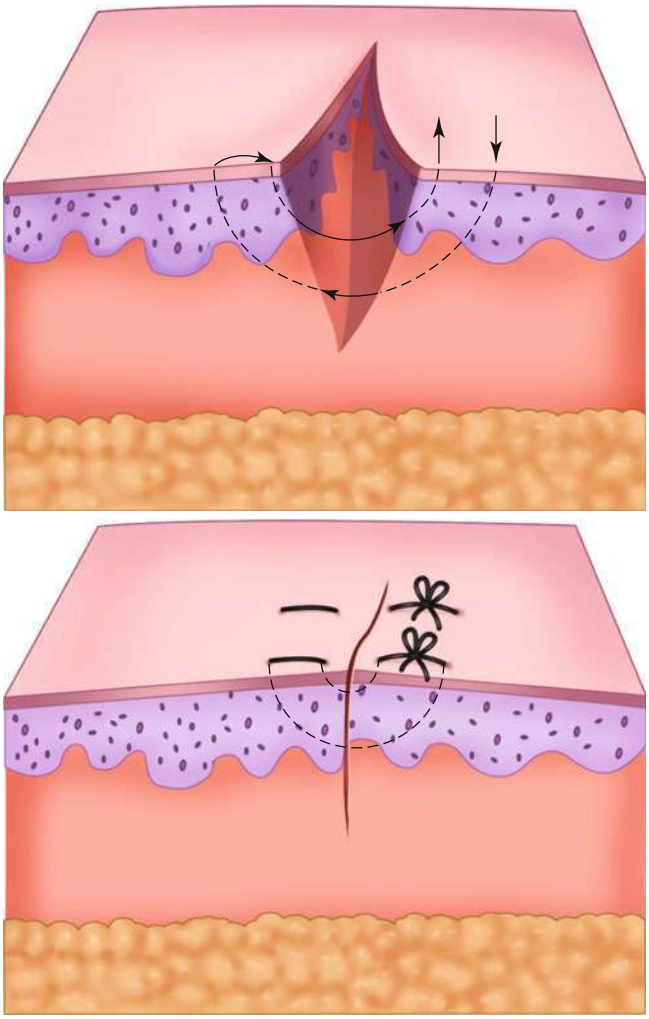

Fig. 92.5 Vertical mattress suture

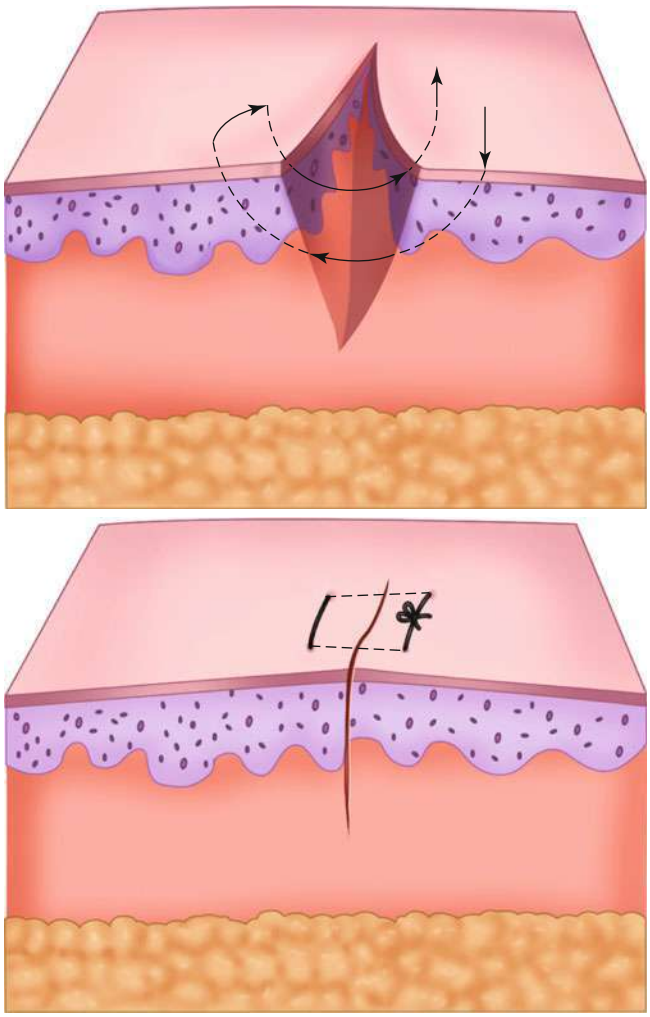

Fig. 92.6 Horizontal mattress suture

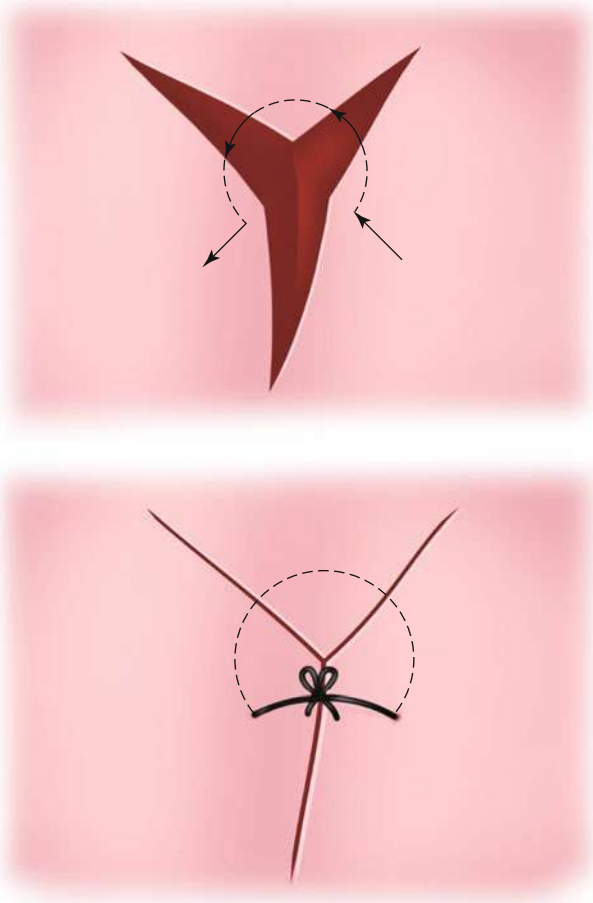

Fig. 92.7 Half-buried horizontal mattress suture

92.6 Alternative Methods of Wound Closure

92.6.1 Cyanoacrylate Tissue Adhesives

See Chap. 93.

92.6.2 Adhesive Tape

- Advantages
  - Rapid and painless application
  - Inexpensive
  - Good cosmetic result
- Disadvantages
  - Minimal strength
- Contraindications
  - Allergy to product
- Precautions
  - For use with wounds under little tension
- Procedure
  - Thoroughly clean the wound as described previously.
  - Approximate the wound edges.
  - Apply adhesive tape directly over the wound with 2–3 mm of space between strips.
  - An adjunct adhesive, such as benzoin, may be used to improve durability.

92.6.3 Staple Closure (Fig. 92.8)

- Advantages
  - Rapid
  - Inexpensive
- Disadvantages
  - Minimal strength
- Contraindications
  - Wounds on face, hands, or feet
- Procedure
  - Thoroughly clean the wound as described previously.
  - Anesthetize the wound edges.
  - Approximate the wound edges (may require an additional set of hands).
  - Apply staples.

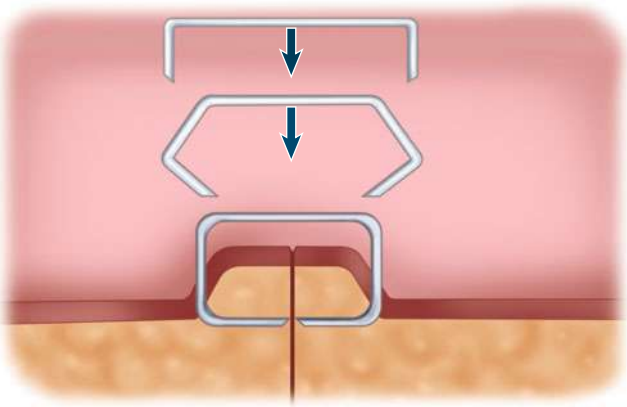

Fig. 92.8 Staple

92.7 Pearls and Pitfalls

- Pearls
  - Sutures should not remain in place longer than the recommended period. If the wound is not completely healed, sutures should be removed and an adhesive tape may be used for edge approximation.
  - Remember to verify tetanus status.
  - Antibiotics should be used judiciously.
- Pitfalls
  - Wound infection.
  - Incomplete healing may lead to wound separation.
  - Tissue reaction to suture or adhesive materials.
  - Allergy to anesthetic agent.

Selected Reading

http://apps.med.buffalo.edu/procedures/repairoflacerations.asp?p=17. Accessed 19 May 2014.

Singer AJ, Hollander JE. Methods for wound closure. In: Ma OJ, Cline DH, Tintinalli JE, Kelen GD, Stapczynski JS, editors. Emergency medicine manual. 6th ed. New York: McGraw Hill; 2003: Chap. 13, Fig. 13–14.

Singer AJ, Hollander JE, editors. Laceration and acute wounds: an evidence-based guide. Philadelphia: FA Davis; 2003. p. 122.

University of Connecticut Health Center, suturing 101. fitsweb.uchc.edu/suturing101. Accessed 19 May 2014.

Zuber TJ. The mattress sutures: vertical, horizontal and corner stitch. Am Fam Physician. 2002;66:2231–6.

93 Wound Closure with Tissue Adhesive

Pratik S. Patel and Latha Ganti

93.1 Indications

- Small superficial skin incisions or laceration repairs which require 5.0 or smaller-diameter sutures.

93.2 Contraindications

- Absolute
  - Large irregular/stellate lacerations
  - Infected/contaminated wounds
  - Animal/human bites
  - Puncture wounds
  - Crush wounds
  - Skin ulcers
  - Mucous membranes and mucocutaneous junctions
  - Axillae and perineum (owing to high moisture)
- Relative
  - Wounds on extremities (unless kept dry)
  - Joints (unless kept immobilized with a splint)

93.3 Materials and Medications

- Dermabond (2-octyl cyanoacrylate)
- Betadine (povidone-iodine) solution
- 0.9 % normal saline solution
- 20-mL sterile syringe
- Sterile gloves
- Dry 4×4 gauze

93.4 Optional Materials

- Topical anesthesia
- 1:1000 epinephrine solution
- Forceps
- Bacitracin ointment or sterile petroleum jelly ointment
- Gown and protective eyeglasses
- Splint

P.S. Patel, MD
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: drpratik@gmail.com

L. Ganti, MD, MS, MBA (✉)
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_93

93.5 Procedure

1. Have the patient rest comfortably on a chair or bed.
2. Use universal precaution measures: sterile gloves (gown and eye-screen, if necessary for wound irrigation).
3. Wash the wound with 0.9 % normal saline irrigation.
4. Use a topical anesthetic such as LET (lidocaine, epinephrine, tetracaine) or EMLA (eutectic mixture of local anesthetics) cream (lidocaine and prilocaine), or a 1:1000 epinephrine solution soaked into gauze can be used to achieve hemostasis in a bleeding wound.
5. Approximate the edges of the wound with fingers. Toothed forceps or other skin approximation devices may be used as an adjunct. Apply bacitracin ointment or Vaseline to the tips of the forceps and wipe off the excess to prevent sticking of Dermabond glue to the forceps.
6. Crush Dermabond vial between the thumb and the finger while in the inverted position until the adhesive is seen at the applicator tip (Fig. 93.1).
7. Squeeze gently until a drop of adhesive forms at the applicator tip.
8. Gently brush the adhesive at the applicator tip over the approximated wound edges (Fig. 93.2). (Do not force or press the applicator tip over the wound.)
9. Cover the entire wound with single coat of adhesive.
10. Hold the wound edges for 30 s to 1 min until it dries.
11. Apply two or three more coats of adhesive over and around the wound in a circular or oval movement to provide extra strength.
12. Wipe off extra adhesive in the surrounding skin with gauze if needed.
13. Apply a splint (optional) to provide wound stability over the joints.
14. Recommend the patient to keep the wound dry for 4–5 days. Patients may shower but should be instructed to pat dry instead of rubbing a towel over the skin.
15. No topical antibiotic ointment is required before or after application of Dermabond.

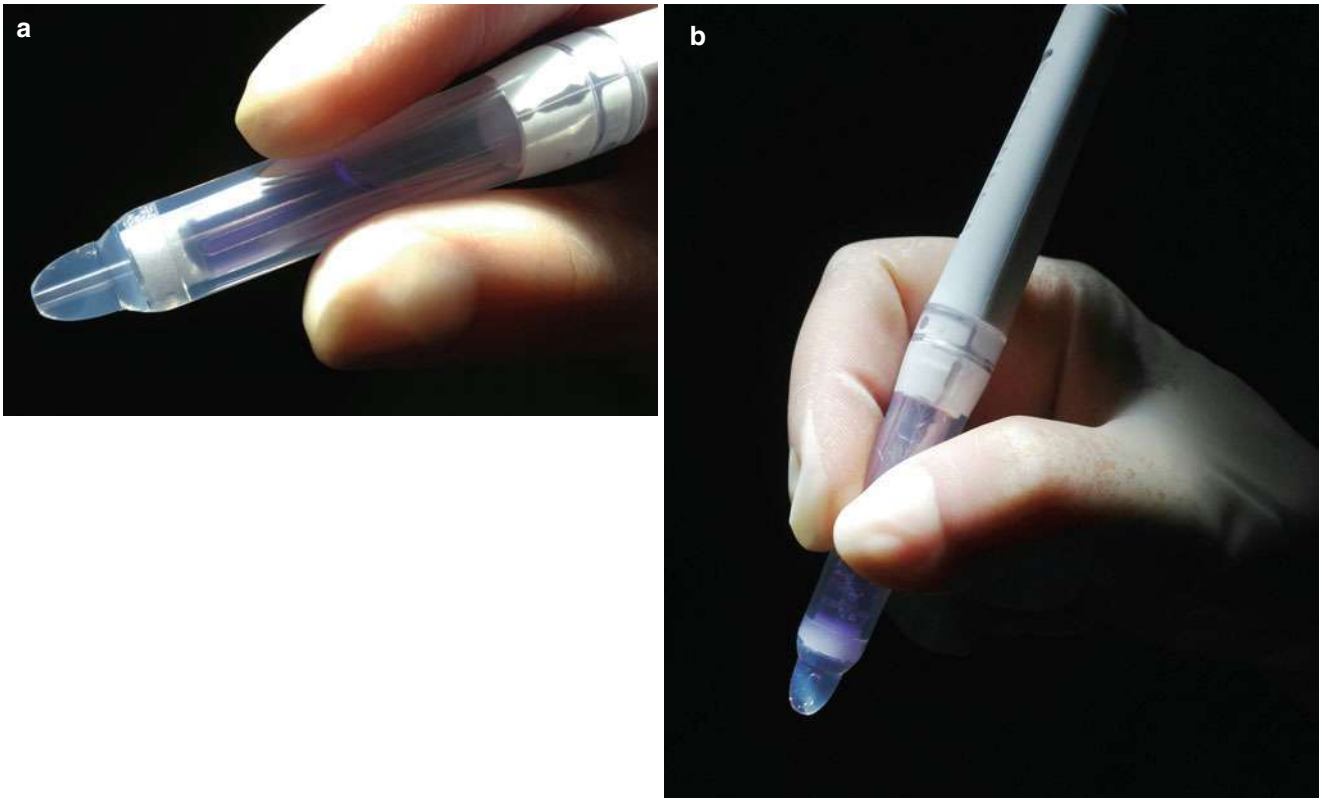

Fig. 93.1 (**a**) Before the tip has been crushed. (**b**) The purple Dermabond is in the tip

Fig. 93.2 Closing the wound: use tip of Dermabond pen to "paint" over the laceration

93.6 Complications

- Wound dehiscence
- Wound infection

93.7 Pearls and Pitfalls

- Pearls
  - Advantages of tissue adhesives for wound repair compared with sutures include faster repair time, better acceptance by patients (especially children), water-resistant covering, and no need for a second visit to remove sutures (sloughs off in 5–10 days).
- Pitfalls
  - Dermabond is a super adhesive. Take care not to have the glove, finger, drape, gauze, or instrument inadvertently stuck to the wound or the patient by having a bacitracin or petroleum jelly coating around the wound, on gloved fingers, and on forceps as needed.

Selected Reading

Bruns TB, Worthington JM. Using tissue adhesive for wound repair: a practical guide to Dermabond. Am Fam Physician. 2000;61:1383–8.

Farion K, Osmond MH, Hartling L, et al. Tissue adhesives for traumatic lacerations in children and adults. Cochrane Database Syst Rev. 2002;(3):CD003326.

94 Fishhook Removal

Judith K. Lucas

94.1 Indications

- Removal of a fishhook from nonvital structures (Fig. 94.1)

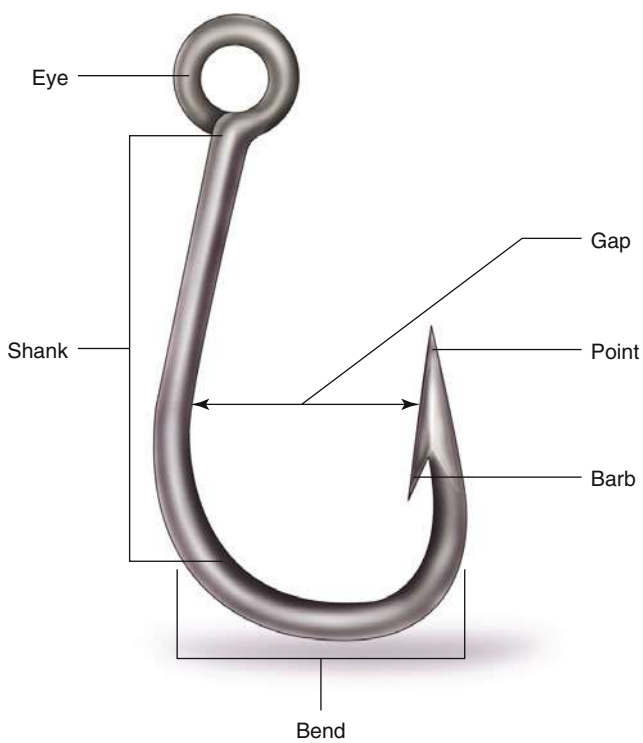

Fig. 94.1 Anatomy of a fishhook

94.2 Contraindications

- Removal of hooks located near/in eyes or eyelids, embedded near or within neurovascular structures, or embedded within vital structures such as peritoneum, testicle, or urethra.
  - Fishhooks in these areas require specialist consultation.

94.3 Materials and Medications

- Antiseptic cleansing solutions
  - Betadine (povidone-iodine)
  - Chloraprep
- Local anesthetic
  - 1 % lidocaine, with or without epinephrine
- Needle drivers or pliers
- 18- or 20-gauge needle
- 3-0 silk suture or umbilical tape
- Wire cutters
- Protective eyewear

J.K. Lucas, MD
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_94

94.4 Procedures

- Retrograde technique (Fig. 94.2): simplest, least traumatic, but least successful; good for small to medium hooks, superficially embedded hooks, and hooks with no barbs or a single barb.
  1. Detach extra hooks, line, or foreign materials (e.g., worms, fish, debris).
  2. Cleanse the puncture site and surrounding tissue with antiseptic solution.
  3. Infiltrate the entry site and surrounding area with local anesthetic.
  4. Apply downward pressure on the shank or shaft of the hook, near the eye, thus disengaging the barb from the tissue.
  5. Back the hook out of the skin along the path of entry.
- String and yank technique (Fig. 94.3): Good for small to medium hooks. Often requires no anesthesia. Good for areas of deep soft tissue penetration. Cannot be used on parts of the body that are not fixed, such as earlobes.
  1. Detach extra hooks, line, or foreign materials (e.g., worms, fish, debris).
  2. Cleanse the penetration site and surrounding area with antiseptic wash.
  3. Consider infiltrating the entry site with local anesthetic.
  4. Wrap 3-0 silk suture, fishing line, or umbilical tape several times around the bend of the hook (at the point of greatest curvature).
  5. The loose ends of the string need be held tightly. Sometimes the loose ends can be more firmly held if wrapped around a pencil or tongue depressor.
  6. The skin around the entry site should be well stabilized, while simultaneously depressing the shank, close to the eye of the hook.
  7. While stabilizing the skin and applying downward pressure on the shaft, quickly and firmly yank the string, in a parallel line to the shaft. Be certain downward pressure is applied along the shank.
  8. The hook will "fly" out quickly. Be certain to wear eye protection.
- Needle cover technique (Fig. 94.4): good method to remove large hooks with a single barb, especially if superficially embedded
  1. Detach extra hooks, line, or foreign materials (e.g., worms, fish, debris).
  2. Cleanse the penetration site and surrounding area with antiseptic wash.
  3. Infiltrate the entry site and surrounding area with local anesthetic.
  4. Advance an 18-gauge needle along the entrance wound of the hook. Pass the needle parallel to the shank. The bevel should be pointed downward, toward the barb and point.
  5. The needle is advanced until it disengages the barb, entrapping it within the needle lumen.
  6. The hook and needle are advanced just enough to disengage the barb.
  7. The hook and needle are withdrawn along the track of the wound.
- Advance and cut technique (Fig. 94.5). Almost always successful but does cause additional tissue trauma.
  1. Detach extra hooks, line, or foreign materials (e.g., worms, fish, debris).
  2. Cleanse the penetration site and surrounding area with antiseptic wash.
  3. Infiltrate the entry site and surrounding area with local anesthetic.
  4. If the hook has a single barb:
     - Grip the hook on the shank, near the bend, with either a needle driver or pliers.
     - Push the hook through along its natural trajectory, until the point and barb pass completely through the skin.
     - Clip the point proximal to the barb. Be certain to wear eye protection because the point can fly off in an unpredictable direction.
     - Withdraw the hook out back along its entry path.
  5. If the hook has multiple barbs (Fig. 94.6):
     - Grip the hook on the shank, near the hook's bend with either a needle driver or pliers.
     - Push the hook through along its natural trajectory, until the point and barb pass completely through the skin.
     - Clip the eye of the hook with wire cutters.
     - Grasp the point and withdraw the hook through the skin forward along its natural course.

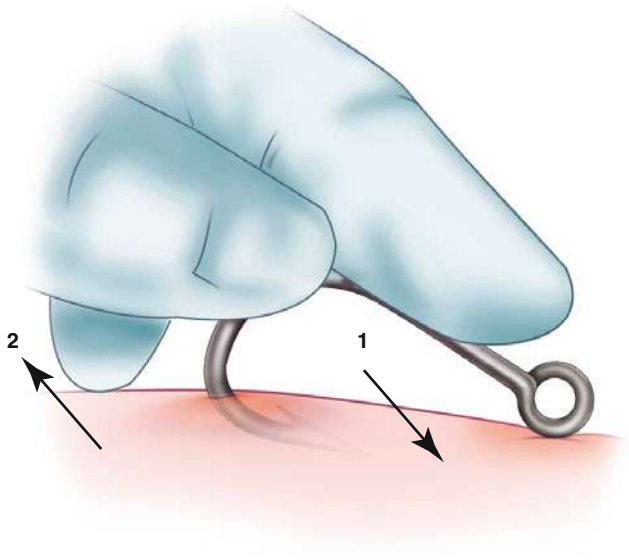

Fig. 94.2 Retrograde technique. *1* Push down along whole shank and slightly forward to release the barb. *2* Pull back and out along hook's entry path, applying steady downward pressure

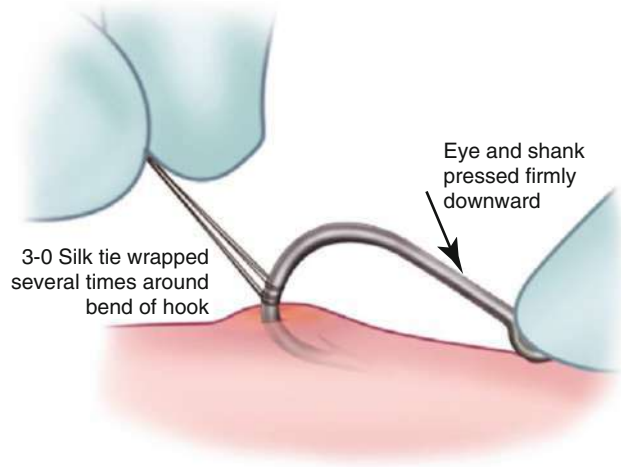

Fig. 94.3 String and yank technique

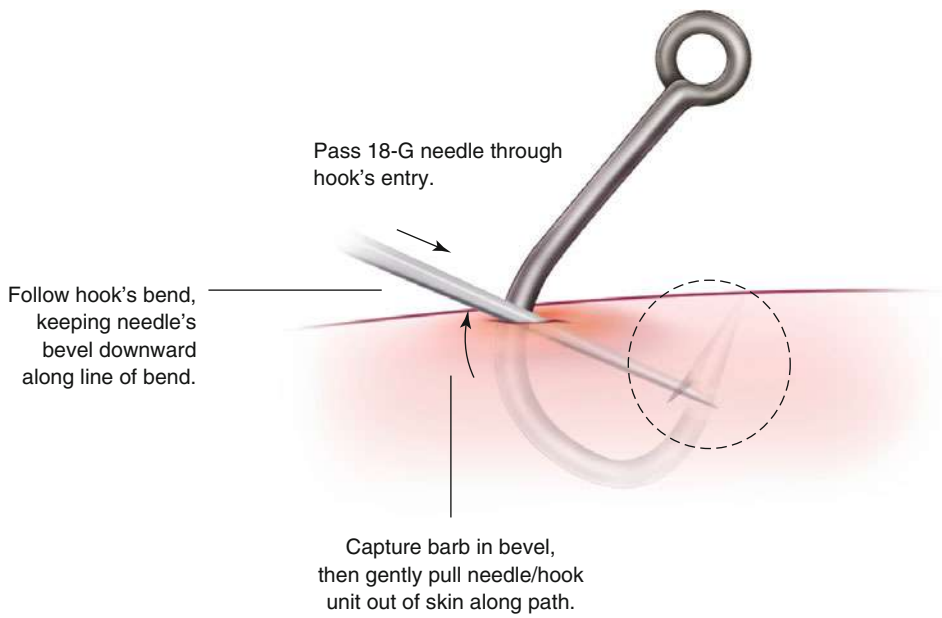

Fig. 94.4 Needle technique

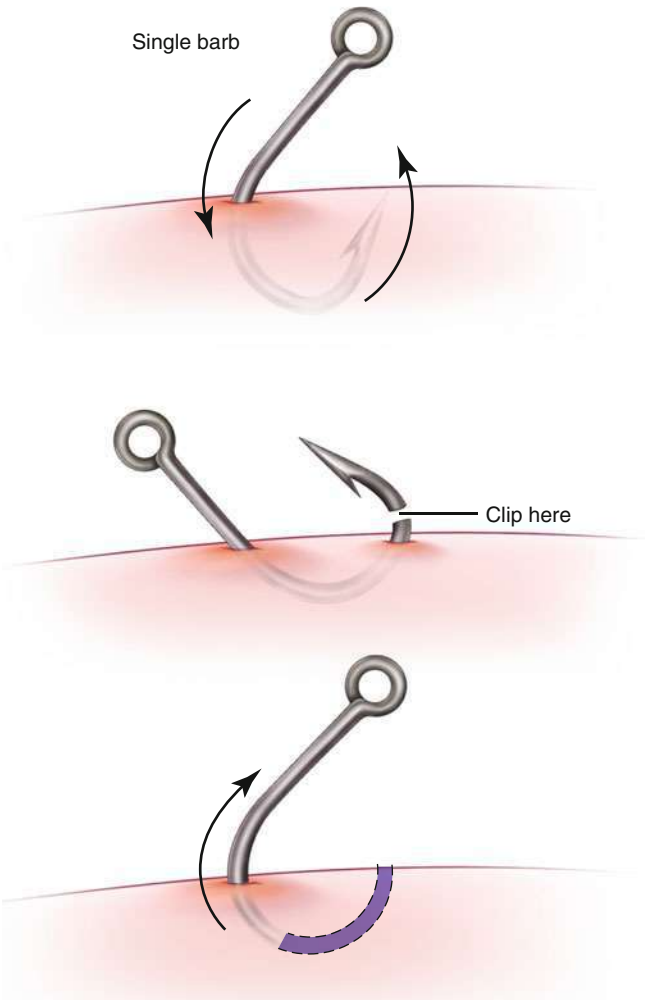

Fig. 94.5 Advance and cut technique

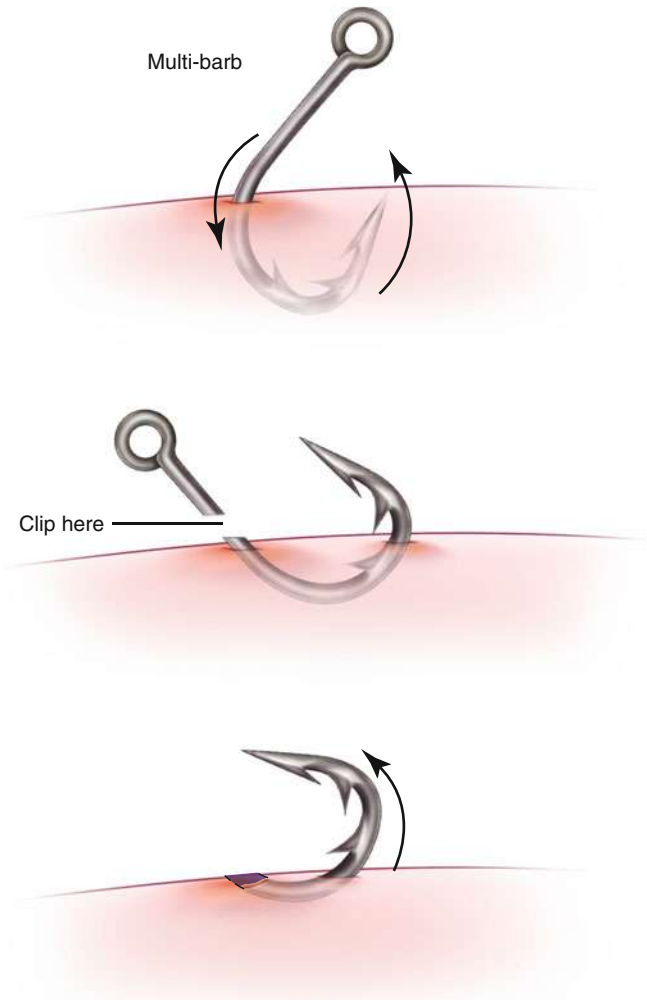

Fig. 94.6 Advance and cut a hook with multiple barbs present

94.5 Postremoval Wound Care

- Immune competent, without peripheral vascular disease:
  1. Explore the wound for possible foreign bodies.
  2. Irrigate or scrub the wound copiously with soapy solution.
  3. Apply antibiotic ointment and sterile dressing.
  4. Tdap (or Dtap) should be given to anyone in whom the last tetanus booster has been longer than 5 years.
  5. Wound check in 24–48 h with care provider.
- Immunocompromised or a patient with peripheral vascular disease:
  1. As previously.
  2. Give strong consideration to treating prophylactically with antibiotics, choosing a fluoroquinolone, third-generation cephalosporin, or aminoglycoside.

94.6 Complications

- Infection
- Retained foreign body
- Injury to neurovascular structures, if method of removal is not carefully selected
- Injury to provider if adequate (eye) protection not used

94.7 Pearls and Pitfalls

- Pearls
  - Begin with either the retrograde method or the string-yank method, because these result in less tissue damage and are the easiest to perform, although they have the lowest rate of success.
  - Eye protection is imperative, especially if utilizing the string-yank method or clipping any portion of the hook (eye or point) because the retraction of the hook, or its parts, is generally at a high velocity and travels an unpredictable path.
  - When trying the advance and cut technique, stop immediately if, when advancing the hook, impasse, or resistance is met because this may indicate bone or neurovascular structures are blocking the natural path of the hook.
  - Fishhooks that embed into or near the eye or lids should be covered with a metal patch or cup and the patient should be sent (immediately) for ophthalmologic consultation.
  - Close follow-up is imperative to watch for signs of infection.
- Pitfalls
  - When utilizing the advance and cut technique, do not cut anything until certain there is another portion of the hook on which to grasp.

Selected Reading

Bothner J. Fish-hook removal techniques. Available at: www.UpToDate.com. Literature review version 19.2, May 2011. Topic last updated: Sep 2010. Accessed 23 June 2014.

Gammons M, Jackson E. Fishhook removal. Am Fam Physician. 2001;63:2231–7.

Wakeman K. Fishhook removal. Available at: www.fishgame.com. Texas Fish and Game, LLC. July 2003. Accessed 23 June 2014.

Tick Removal

95

David N. Smith and Judith K. Lucas

95.1 Indications

- Tick attachment to the skin

95.2 Contraindications

- None

95.3 Materials and Medications

- Gloves
- Skin disinfectant (commercially available product, such as Chloraprep, isopropyl alcohol, or Betadine [povidone-iodine])
- Fine-toothed forceps

95.4 Procedure

1. Comfortably position the patient with the tick site exposed.
2. Grasp the tick as close to the skin surface as possible (e.g., grasp the mouth parts).
3. Gently pull upward with steady, nontwisting, even traction (Figs. 95.1 and 95.2).
4. After removal, clean the bite area and apply antibiotic ointment.

Fig. 95.1 Place the forceps as close to the mouth of the tick as possible, hold firmly, and pull straight up with steady gentle traction

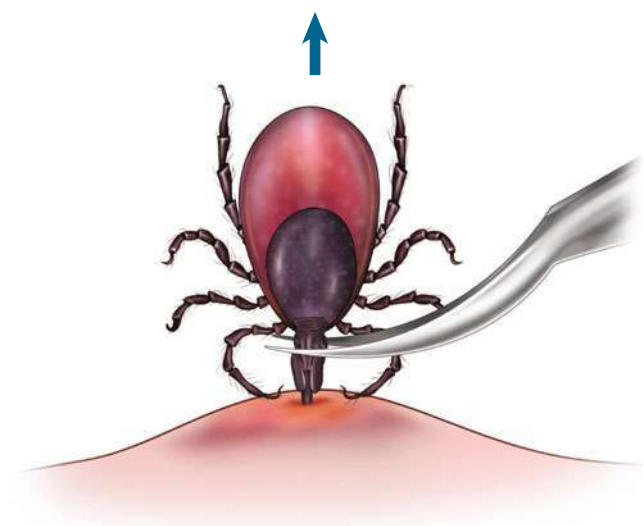

Fig. 95.2 Steady upward traction

D.N. Smith, MD
Department of Pediatrics, University of Alabama at Birmingham, Children's of Alabama, Birmingham, AL, USA
e-mail: dsmith@peds.uab.edu

J.K. Lucas, MD (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_95

95.5 Complications

- Multiple diseases including:
  - Lyme disease
  - Human granulocytic and monocytic ehrlichiosis
  - Babesiosis
  - Relapsing fever
  - Rocky Mountain spotted fever
  - Colorado tick fever
  - Tularemia
  - Q fever
  - Tick paralysis
- Secondary infection (methicillin-resistant *Staphylococcus aureus* [MRSA] and group A streptococcus).
- Scratching can lead to lichenification.
- Rare cases of alopecia when tick located in the scalp.

95.6 Pearls and Pitfalls

- Pearls
  - Do not twist or jerk the tick out (may cause breakage of mouth parts; they may remain in the skin).
  - Do not squeeze the body of the tick.
  - Do not use a hot match, gasoline, or other noxious stimulus for removal (causes irritation of tick and release of internal contents).
  - Lyme disease transmission increases significantly after 24–48 h of attachment, so early removal is the key.
  - Patients should be monitored for up to 30 days for signs of tick-borne diseases, including erythema migrans (bull's-eye rash) indicating Lyme disease (Fig. 95.3).
- Prophylactic antibiotic treatment, a single dose of doxycycline, is used only in patients with an identified *Ixodes scapularis* tick that has been attached for longer than 36 h and if treatment can start within 72 h of tick removal.
  - Serological testing for Lyme disease is not indicated for a reported tick bite.
- Pitfalls
  - Use of lidocaine subcutaneously can irritate the tick and cause it to regurgitate its stomach contents, increasing the risk of disease transfer.

Fig. 95.3 Erythema migrans (the bull's-eye rash), the rash typically associated with Lyme disease. Often the rash is so pale as to go unnoticed

Selected Reading

Centers for Disease Control and Prevention. Tick removal. In: Ticks. 2014. http://www.cdc.gov/ticks/removing_a_tick.html. Accessed 23 Feb 2014.

Needham GR. Evaluation of five popular methods for tick removal. Pediatrics. 1985;75:997–1002.

Sexton DJ. Evaluation of a tick bite for possible Lyme disease. Up To Date; 2010. http://uptodate.com.

Sloan S. Background. In Tick removal. 2014. http://emedicine.medscape.com/article/1413603-overview. Accessed 23 Feb 2014.

96 Subungal Hematoma Drainage

Pratik S. Patel and Latha Ganti

96.1 Causes

- Crushing injury to the finger or toe
- Ill-fitted shoes/inadequate space for toes

96.2 Indications

- Pain that is not tolerable with nail edges intact

96.3 Contraindications

- Relative
  - Disrupted nail edges
  - Tolerable pain that can be managed conservatively
  - Skin infections around toe/finger
  - Bleeding disorder

96.4 Materials and Medications

- 18-gauge needle
- Betadine (povidone-iodine) solution
- Gloves
- Topical antibiotic

96.4.1 Optional materials

- Electrocautery tool
- Paper clip or sewing needle and sterilizing flame
- Finger splint
- Lidocaine (1 or 2 % with epinephrine)

P.S. Patel, MD
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: drpratik@gmail.com

L. Ganti, MD, MS, MBA (✉)
Professor of Emergency Medicine, University of Central Florida, Orlando, FL, USA

Director, SE Specialty Care Centers of Innovation, Orlando Veterans Affairs Medical Center, Orlando, FL, USA
e-mail: lathagantimd@gmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_96

96.5 Procedure

1. Have the patient rest the affected finger comfortably on a flat surface.
2. Use universal precaution measures: gloves, gown, and eye screen.
3. Sterile skin preparation with Betadine.
4. Optional (for complicated subungal hematoma): use 1 or 2 % lidocaine with epinephrine for providing local digital block.
5. Using the thumb and index finger gently twist the 18-gauge needle with light pressure over the base of the nailbed or in the center of the hematoma until no resistance is felt. Do not apply any further pressure in order to avoid nailbed damage. Nail penetration is confirmed by return of dark blood from the hole. This procedure is known as trephination (Figs. 96.1 and 96.2).
6. Apply light pressure around the tip of the finger and hematoma to facilitate drainage. In case of continuous bleeding elevate the digit and apply firm and continuous pressure over the nail with a gauze piece.
7. Apply topical antibiotic ointment (e.g., bacitracin) over the puncture hole.
8. Apply a gauze dressing or a bandage over the wound site/fingernail.
9. Apply a finger splint (optional) to provide additional comfort.
10. Recommend the patient to keep the finger or toe dry (avoid soaking) and elevated for 1–2 days.

Fig. 96.1 Subungal hematoma of right index finger

Fig. 96.2 Positioning of needle to perform subungal hematoma drainage

96.6 Complications (Rare)

- Infection
- Injury to nail bed or underlying bone if too much pressure is applied with puncture
- Onycholysis when there is incomplete drainage in rare cases

96.7 Pearls and Pitfalls

- Pearls
  - Multiple holes may be necessary for appropriate drainage of the hematoma.
  - When using electrocautery for hole puncture, execute high caution with acrylic nails because they are flammable.
  - When using a paper clip or regular needle, make sure to sterilize the tip over a flame.
  - Take a radiograph of the finger whenever necessary to rule out phalangeal fracture.
  - Check for avulsion of the extensor tendon.
  - Inform the patient that the existing nail may fall off and will regenerate in few months if the nailbed is intact.
  - Systemic or oral antibiotics are not recommended.
- Pitfalls
  - Development of a dark color change over the nailbed without any history of trauma should raise suspicion of a tumor and should be evaluated accordingly.
  - Do not remove the nail to evaluate injury to nailbed.

Selected Reading

Brown RE. Acute nail bed injuries. Hand Clin. 2002;18:561–75.

Dean B, Becker G, Little C. The management of the acute traumatic subungual haematoma: a systematic review. Hand Surg. 2012;17:151–4.

97 Incision and Drainage of Abscess

Nicholas D. Caputo, Karlene Hosford, and Muhammad Waseem

97.1 Indications

- Abscess greater than 5 mm in diameter and in accessible areas (e.g., axilla, extremities, trunk)

97.2 Contraindications

- Absolute
  - Absence of fluctuation
  - Large, deep, and complicated (multiloculated) abscesses
  - Location
    - Perianal
    - Mastoid
- Relative
  - Location
    - Face (e.g., nose, nasolabial fold)
    - Palms
  - Coagulopathy
  - Recurrent pilonidal cysts (may mandate operative excision)
  - Area of cosmetic importance where aspiration may be preferred

97.3 Materials and Medications

- Incision and drainage tray (Fig. 97.1)
  - Drape
  - Betadine (povidone-iodine) swabs
  - 1 % lidocaine
  - 18- and 27-gauge needles
  - 12-mL syringes, gauze pads
  - #11 scalpel, mosquito clamps (hemostat)
  - Iodoform packing of appropriate size
- Ultrasound machine (Fig. 97.2)

N.D. Caputo, MD, MSc
Emergency Department Critical Care, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: Ncaputo.md@gmail.com

K. Hosford, MD • M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: krlhos@aol.com; waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_97

Fig. 97.1 Supplies necessary for incision and drainage

Fig. 97.2 Bedside SonoSite ultrasound

97.4 Procedure

1. Ultrasound (optional) may be helpful when abscess is suspected in the absence of fluctuation. Using the vascular probe (7 mHz), confirm the clinical suspicion of abscess and check the depth and width of the abscess (Fig. 97.3).
2. Sterile skin preparation with Betadine swab and sterile drape.
3. Anesthetize the appropriate area subcutaneously with 5 mL of 1 % lidocaine by inserting the 27-gauge needle at an acute angle into the intradermal space (Fig. 97.4a, b).
4. Using a #11 blade, make an approximately 1- to 2-cm skin incision over the desired area parallel to the Langer lines. The incision must approach into the abscess cavity (Fig. 97.4c).
   - Some physicians still advocate the technique of making a cruciate incision. This may leave a larger scar and should be discussed with patient before doing so because of cosmetic consequences.
5. Allow for spontaneous drainage. After resolution of drainage, you may express more pus with gentle downward pressure.
6. Using the hemostat, enter the incision to break any suspected loculations. This should be done with the clamps closed and curved part down. The clamps should then be opened and removed slowly (Fig. 97.4d).
7. After clearing the remaining loculations, the wound should be packed.
   - Evidence suggests that packing the abscesses does not prevent recurrence; however, this is still practiced. If packing the wound, follow the next step.
8. Take the iodoform packing with the hemostat, and place the packing into the incision site until no further packing will fit.
9. Cut the packing leaving a tail out of the incision site (Fig. 97.4e).
10. Apply dressing with 4×4 gauze and adhesive tape (2 in.).

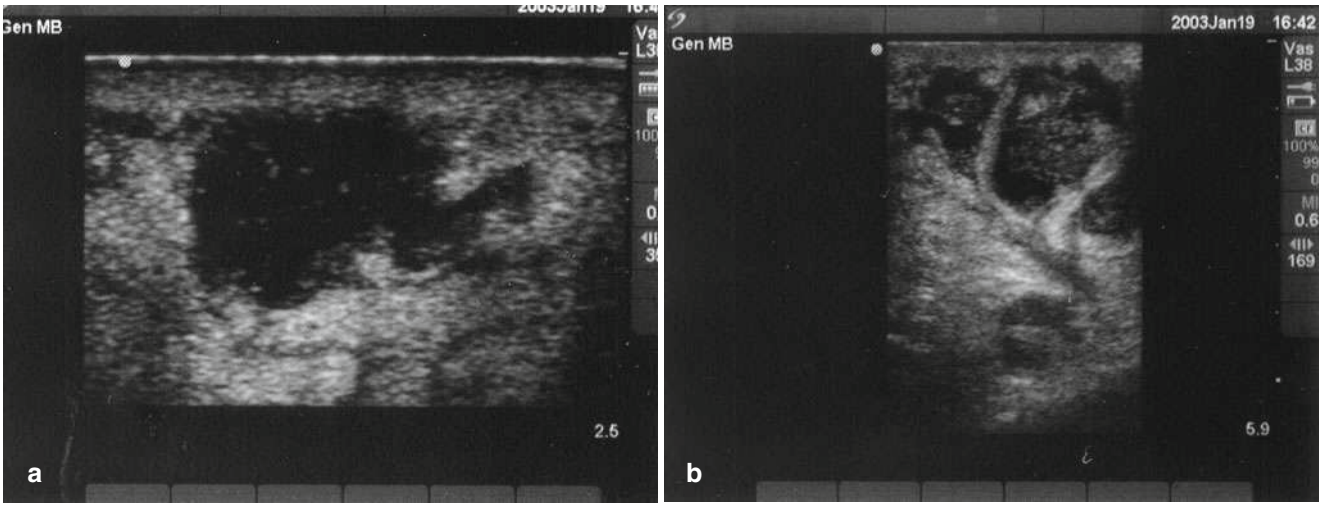

Fig. 97.3 (**a**) Example of an abscess as viewed on bedside ultrasound. (**b**) A multiloculated abscess

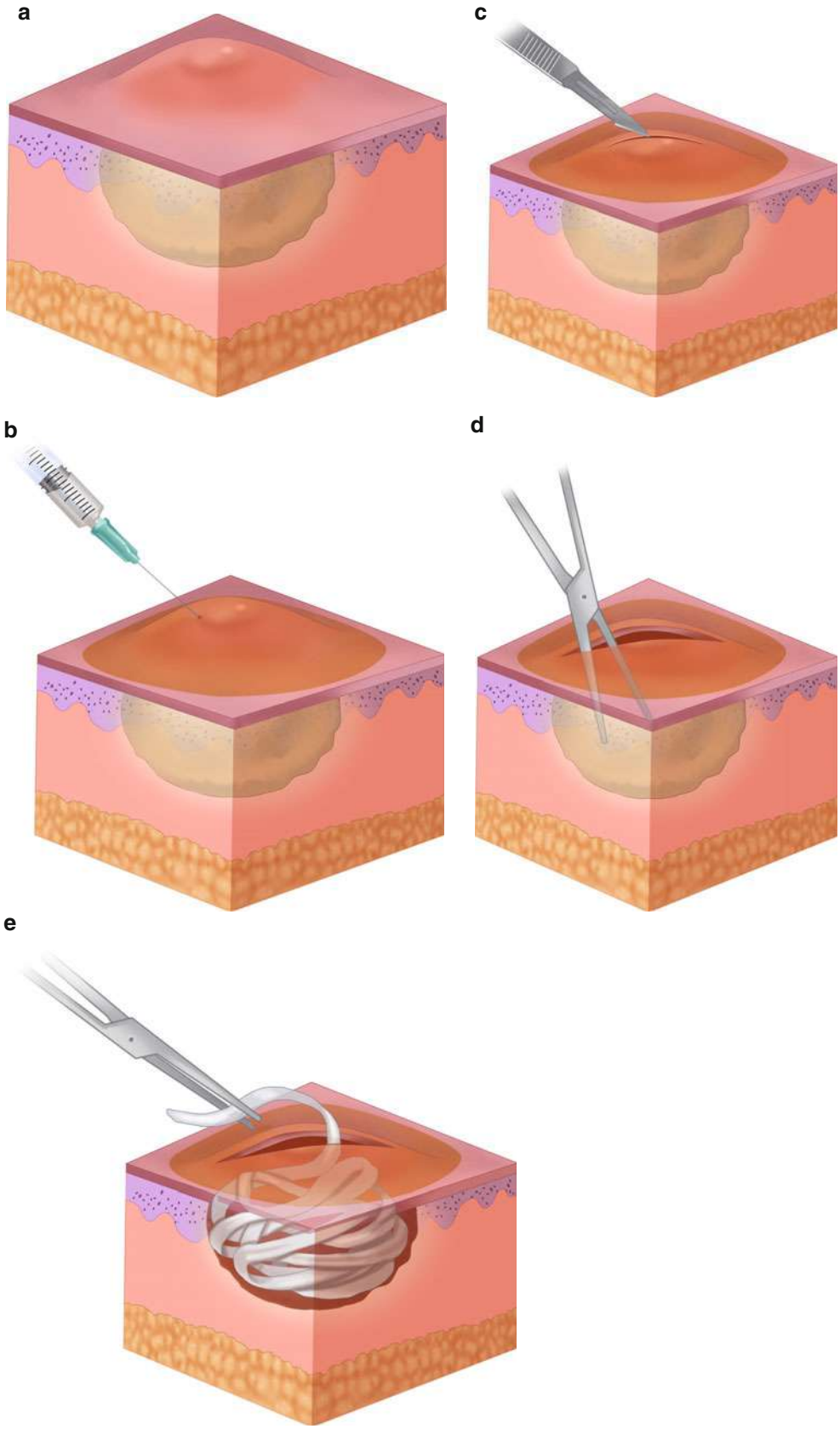

Fig. 97.4 (**a**) Abscess with overlying erythema. (**b**) Lidocaine injection in the superficial layer. (**c**) Linear incision with #11 blade. (**d**) Expression of purulent material and breaking of loculations with clamps. (**e**) Optional placement of packing

97.5 Complications

- Recurrence of abscess
- Progression of cellulitis
- Neurovascular injury to adjacent structures

97.6 Pearls

Antibiotic coverage is a controversial topic. Methicillin-resistant *Staphylococcus aureus* (MRSA) is a concern not only in the immunocompromised and diabetic patients. *S. aureus* has been detected in up to 51 % of patients with abscesses. Of these isolates, approximately 75 % were MRSA. Bactrim (trimethoprim/sulfamethoxazole) should be utilized for all prophylactic measures.

Selected Reading

Barnes SM, Milsom PL. Abscess: an open and shut case. Arch Emerg Med. 1988;5:200–5.

Burney RE. Incision and drainage procedures: soft tissue abscesses in the emergency service. Emerg Med Clin North Am. 1986;4:527–42.

Duong M, Markwell S, Peter J, Barenkamp S. Randomized, controlled trial of antibiotics in the management of community-acquired skin abscesses in the pediatric patient. Ann Emerg Med. 2010;55:401–7.

Fitch MT, Manthey DE, McGinnis HD, Nicks BA, Pariyadath M. Abscess incision and drainage. N Engl J Med. 2007;357:e20.

Frazee BW, Lynn J, Charlebois ED, Lambert L, Lowery D, Perdreau-Remington F. High prevalence of methicillin-resistant *Staphylococcus aureus* in emergency department skin and soft tissue infections. Ann Emerg Med. 2005;45:311–20.

Hankin A, Everett WW. Are antibiotics necessary after incision and drainage of a cutaneous abscess? Ann Emerg Med. 2007;50:49–51.

O'Malley GF, Dominici P, Giraldo P, et al. Routine packing of simple cutaneous abscesses is painful and probably unnecessary. Acad Emerg Med. 2009;16:470–3.

Part XIII

Orthopedic Procedures

98 Splinting

Christopher H. Stahmer and Muhammad Waseem

98.1 Indications

- Need for immobilization for fracture, dislocation, or soft tissue injury
- Suspicion for occult injury of an extremity
- Immobilization for pain management

98.2 Contraindications

- Absolute
  - Open fracture (requires operative intervention)
- Relative
  - Infection
  - Compartment syndrome

98.3 Materials and Medications (Fig. 98.1)

- Plaster of Paris
- Fast drying: 5–8 min to set
- Extra fast drying: 2–4 min to set
  - Variety of widths depending upon splint of choice:
  - Splints may take up to 2 days to dry and achieve maximum strength.
- Prefabricated splinting materials
  - Plaster OCL® (Orthopedic Casting Laboratories)
    - 10–20 sheets of plaster with padding and cover
    - Faster setup time but less customizable
  - Fiberglass splints
    - Cure rapidly
    - Less messy
    - Less moldable
    - Stronger and lighter
- Stockinette.
  - Protects the skin.
  - Variety of sizes available.
- Soft wrap (Webril™).
  - Provides padding.
  - Five to six layers depending on anticipated swelling.
    - Too much padding reduces the stability of the splint.
  - Use extra padding over bony prominences.
  - Pad between digits for splinting of digits.
  - Avoid wrinkles, which generate pressure points.
  - Do not wrap circumferentially.
    - Increased risk of ischemia.
- Ace wraps.
  - Variety of sizes depending.
    - Larger widths over legs.
    - Narrow widths around fingers and joints.
      - Avoid bunching by using narrow widths at joints.
- Water
  - Warm water and splint sets more quickly but increases the risk of burns.
    - Splint drying is an exothermic or heat-releasing reaction.
  - Hot water leaves less time to mold the splint.

C.H. Stahmer, MD • M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: stahmer@gmail.com; waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_98

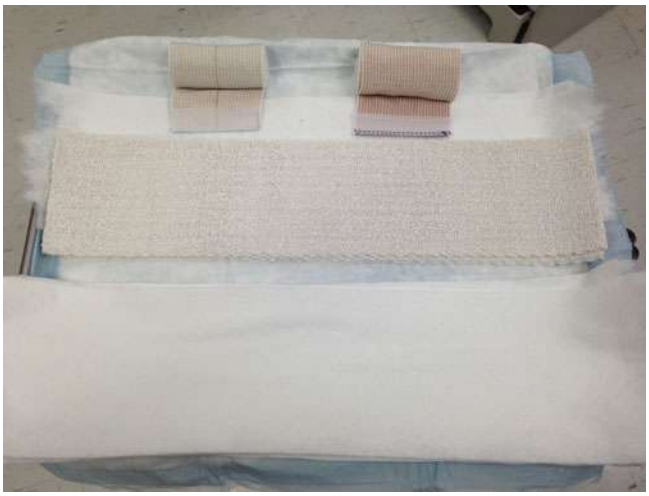

Fig. 98.1 Posterior splint materials: plaster of Paris, Ace wraps, soft roll. Note there are two layers of soft wrap: the inner layer to face the patient (eight layers) and the outer layer to pad the exterior (two layers)

98.4 Procedure

1. Completely expose and examine the afflicted body part for tissue, vascular, or neurological injury.
   - Address respective injuries before proceeding.
2. Lay out all splinting materials before initiating procedure.
   - Layer plaster of Paris.
     - Upper extremity: 8–10 layers.
     - Lower extremity: 12–15 layers.
     - Up to 20 for a large person.
     - More layers of plaster of Paris increase the risk of burn and the weight of the splint.
3. Administer appropriate anesthesia.
   - Conscious sedation
   - Hematoma block
   - Intra-articular injection
   - Intravenous pain medication
   - Oral pain medication
4. Hang fractures as indicated for improved success of reduction to relax muscles before reduction attempt.
5. Reduce afflicted extremity.
6. While maintaining reduction, apply respective splint.
7. Apply in the following order for plaster of Paris splint.
   - Stockinette (not necessary).
   - Soft wrap.
   - Select appropriate layers of plaster of Paris.
   - Prepare plaster of Paris to create splint:
     - Layer plaster with no overlap.
     - Submerge completely into water.
     - Crumple into ball without letting go of the ends of the splint.
     - Release the lower end of the splint while holding the top tightly together.
     - Run fingers in a "squeegee" manner from top to bottom to smooth the splint (Fig. 98.2).
       - This also removes excess water.
       - Repeat until the splint is smooth and free of dripping water.
     - Apply soft wrap layers to the splint.
       - Apply thicker layer to the patient's body.
       - Apply two or three layers of soft wrap to the exterior of plaster of Paris for padding and to facilitate drying.
     - Apply Ace wrap to hold the splint and assist in contouring the splint to the patient's extremity (Fig. 98.3).
       - Applying the Ace wrap too tightly may cause ischemia. Observe the patient after splinting for 30 min for tingling, burning, pain, or discomfort.
     - Mold the splint without making indentations with the fingertips (Fig. 98.4).
       - An indentation may cause a pressure point, which may result in an ulcer.

Fig. 98.2 Hold the top of the saturated plaster securely with one hand while removing excess water with the other hand

Fig. 98.3 Apply Ace wrap to hold the splint and assist in contouring the splint to the patient's extremity

Fig. 98.4 Hold the splint in a neutral anatomical position while taking care not to make indentations with the fingertips

  - Allow the splint to cure while the practitioner maintains the appropriate position. This will take approximately 5 min depending upon water temperature and splint thickness.

98.5 Complications

- Ischemia may result in compartment syndrome.
  - Advise the patient to unwrap the splint for the following indications.
    - Increasing pain.
    - Discoloration of fingers, toes, or the splinted extremity.
    - Loss of sensation of splinted extremity.
- Burns
  - Plaster drying releases heat.
  - Increased risk with limited layers of padding.
  - If pain is troubling the patient, remove the splint and add more padding.
- Pressure sores
  - Apply ample padding.
  - Smooth all wrinkles.
  - Instruct the patient to return for increased discomfort.
- Infection
  - Clean and débride all devitalized tissue before application.
  - Requires close follow-up to reevaluate wounds.

Selected Reading

Fitch MT, Nicks BA, Pariyadath M, McGinnis HD, Manthey DE. Basic splinting techniques. N Engl J Med. 2008;359:e32.

Marx JA, Hockberger R, Walls R, editors. Rosen's emergency medicine: concepts and clinical practice. 7th ed. Philadelphia: Mosby; 2010.

Simon R, Sherman S, Koenigsknecht S. Emergency orthopedics—the extremities. New York: McGraw-Hill; 2007.

Ulnar Gutter Splint

99

Jeffrey Kile, Katrina John, and Amish Aghera

99.1 Indications

- Fractures and soft tissue injuries of the ring or little finger
- Fractures of the neck, shaft, or base of the fourth or fifth metacarpal

99.2 Contraindications

- Relative
  - Evidence of compartment syndrome or any neurovascular compromise

99.3 Materials (Fig. 99.1)

- Splint roll (plaster of Paris, prefabricated foamcore, or fiberglass)
- Stockinette
- Cotton padding, such as simple cotton sheet wadding or a newer alternative such as Webril (Curity), Specialist (Johnson & Johnson), and so on
- Elastic bandages, such as Ace (3M) or similar
- Shears
- Adhesive tape
- Gloves

J. Kile, MBBS, PhD, MPH (✉) • K. John, MBBS
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: jeffrey.kile@gmail.com; trenjohn@me.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_99

Fig. 99.1 Materials and medications

99.4 Procedure

1. Select the appropriate diameter stockinette according to the size of the forearm.
   - Generally, 3-in. diameter stockinette is used for the ulnar gutter splint.
2. Cut the stockinette to a length approximately 20 cm longer than the total length of the desired splint.
3. Apply the stockinette to the injured limb so that the stockinette extends approximately 10 cm beyond the region to be splinted proximally and distally.
4. Cut the distal aspect of the stockinette to free the thumb and also the index and middle finger (Fig. 99.2).
5. Insert cotton padding cut to the appropriate size between the ring and the little fingers to prevent skin degradation (Fig. 99.3).
6. Wrap cotton padding circumferentially around the entire region to be splinted, with each turn overlapping the previous turn by approximately 25 % of its width.
   - Generally, 3-in.-wide padding is used for the ulnar gutter splint.
   - Apply padding to an approximate thickness of 1 cm, with extra padding placed over the bony prominences of the wrist and carpometacarpal joints.
   - Extend the padding at least an inch beyond the desired splint length so that it may later be folded back over the jagged ends of the plaster or fiberglass splint roll (Fig. 99.4).
7. Unwrap and prepare an appropriate length splint.
   - The width of splinting material used should be approximately one-half the circumference of the extremity measured at the wrist, such that it is wide enough to extend approximately halfway around the distal forearm once applied.
   - Layer the splint eight sheets thick if using plaster roll and six sheets thick if using fiberglass roll.
   - It is best to begin with a generous length of splinting material because both plaster and fiberglass splint rolls shrink slightly once moistened. If the wet splint turns out too long, the ends can be either folded back or cut with shears before hardening.
8. Apply the dry splinting material to the ulnar aspect of the extremity from the midforearm to just beyond the distal interphalangeal joint of the little finger and cut the splint to approximately 20 cm longer than the desired length.
9. Submerge the dry splint in water until bubbling stops.
10. Remove the splint from water, place on a hard flat surface, and smooth out the excess water to ensure no wrinkles are present.
11. Apply to the splint to the limb over the cotton padding.
    - If the splint is too long at this stage, the ends may be folded back or cut with shears to the proper length.
12 Once positioned properly, fold the side of the splint up around the ulnar aspect of the forearm and hand to form a gutter (Fig. 99.5).
13. Fold the underlying cotton padding and stockinette back over the ends of the splint, which both protects the skin and holds the splint in place.
14. Secure the loose end(s) of the elastic bandage with adhesive tape.
    - Avoid using the metal clips often packaged with elastic bandages because these can become displaced and embed in the skin.
15. Manipulate the forearm and hand into the "neutral" position: (1) the wrist in slight extension (10–20°), (2) the metacarpophalangeal joints in 50° of flexion, and (3) the proximal and distal interphalangeal joints of the ring and little finger in slight flexion (10–15°) (Fig. 99.6).
    - When splinting a boxer's (i.e., metacarpal neck) fracture, the metacarpophalangeal joint should be flexed to 90°.
16. Mold the splint to the contour of the extremity (Fig. 99.7).
    - Use only the palms of the hands when molding because the fingertips may cause indentations, resulting in excessive skin pressure.
17. Wrap the extremity with an elastic bandage in a distal-to-proximal direction to secure the splint in place (Fig. 99.8).
18. Ensure the extremity is neurovascularly intact.
19. Instruct the patient to loosen the elastic bandage if it feels too tight.

Fig. 99.2 Stockinette applied to the forearm and hand

Fig. 99.3 Cotton padding applied to the ring and little fingers

Fig. 99.4 Cotton padding applied to the forearm and hand

Fig. 99.5 Splint folded around the ulnar aspect of the forearm and hand

Fig. 99.6 Forearm and hand in the neutral position (without splint)

Fig. 99.7 Splint molded to maintain the forearm and hand in the neutral position

Fig. 99.8 Splint secured in place with an elastic bandage

99.5 Complications

- Neurovascular or other soft tissue compromise (if the splint/bandage is too tight)
- Soft tissue degradation (if the splint is left in place too long)

99.6 Pearls and Pitfalls

- Pearls
  - Simple plaster of Paris splints are inexpensive and allow a thoroughly customizable fit but can be damaged by water and require more time to set and more clean up than prefabricated foamcore or fiberglass splint rolls.
  - Simple fiberglass splints set quickly, are not damaged by water, are stronger and lighter than simple plaster and prefabricated splints, and offer a fully customized fit, but are not applied as quickly as prefabricated splints.
  - Prefabricated fiberglass splints are quickly applied, require virtually no clean up, and are not damaged by water, but are relatively expensive and provide a somewhat less customizable fit than simple splint rolls.
  - Stockinette protects the skin and, when folded back over the ends of the plaster or fiberglass splint roll, holds the wet splint in place before the elastic bandage is placed, and provides a padded rim with a professional appearance.
- Pitfalls
  - Avoid wrinkling the cotton padding applied between the stockinette and the splinting material because, once under the pressure of the elastic bandage, wrinkles can cause unnecessary skin pressure.
  - Avoid using more or fewer layers of splinting material than recommended. Additional layers can result in excessive heat during the setting process and a splint that is too heavy, whereas insufficient layers can result in a splint that is too weak.

Selected Reading

Harrison BP, Hilliard MW. Emergency department evaluation and treatment of hand injuries. Emerg Med Clin North Am. 1999;17:793–822.

Henry MH. Fractures of the proximal phalanx and metacarpals in the hand: preferred methods of stabilization. J Am Acad Orthop Surg. 2008;16:586–95.

Margic K. External fixation of closed metacarpal and phalangeal fractures of digits. A prospective study of one hundred consecutive patients. J Hand Surg Br. 2006;31:30–40.

McMahon PJ, Woods DA, Burge PD. Initial treatment of closed metacarpal fractures. A controlled comparison of compression glove and splintage. J Hand Surg Br. 1994;19:597–600.

Viegas SF, Tencer A, Woodard P, Williams CR. Functional bracing of fractures of the second through fifth metacarpals. J Hand Surg Am. 1987;12:139–43.

Shoulder Dislocation Reduction Techniques

100

Katrina Skoog Nguyen, L. Connor Nickels, and Rohit Pravin Patel

100.1 Indications

- Subjective history of new-onset dislocation or recurrent dislocations combined with clinical assessment consistent with shoulder dislocation
  - Anterior dislocations (96 %)
    - Typical mechanism of injury being indirect, with combination of abduction, extension, and external rotation. Rarely, the etiology is a direct blow to the posterior shoulder.
    - Prominent acromion with a palpable drop off below the acromion and subclavicular region fullness is consistent with anterior shoulder dislocation.
  - Posterior dislocations (4 %)
    - Mechanism of injury is indirect with a combination of internal rotation, adduction, and flexion. Precipitating events include seizure, electrical shock, and falls.
    - More subtle presentation. Patient will maintain arm locked in internal rotation and adduction; he or she cannot externally rotate. Shoulder is flattened anteriorly and rounded posteriorly.
    - Ultrasound can be used to prevent missed or delayed diagnosis (Figs. 100.1 and 100.2 show probe positioning and a diagram of abnormal ultrasound anatomy).
  - Inferior dislocations (luxation erecta)
    - Arm will be held fixed in overhead position.
- Radiographs reveal shoulder dislocation.
- Ultrasound can be used to identify the nature of the dislocation (anterior or posterior) and can be determined by the position of the humeral head relative to the transducer and glenoid. Although at this point, it should not replace radiographs owing to missed fractures. Advantages may include less radiation (decreased need for postreducation x-rays) and re-sedation if reduction is not complete.

K. Skoog Nguyen, DO
Northwest Community Hospital, Arlington Heights, IL, USA
e-mail: katrinaskoog@hotmail.com

L.C. Nickels, MD, RDMS (✉) • R.P. Patel, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; rohitpatel@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_100

Fig. 100.1 (**a**, **b**) Ultrasound image of normal shoulder anatomy. "Dot fits the Dot" means when looking at the ultrasound machine from the sonographer's standpoint, the side of the probe marker corresponds to the side the marker is on the screen. This ensures when doing procedures, the direction of needle correction is the same as the orientation of the probe (Images courtesy of Dr. Rohit Patel)

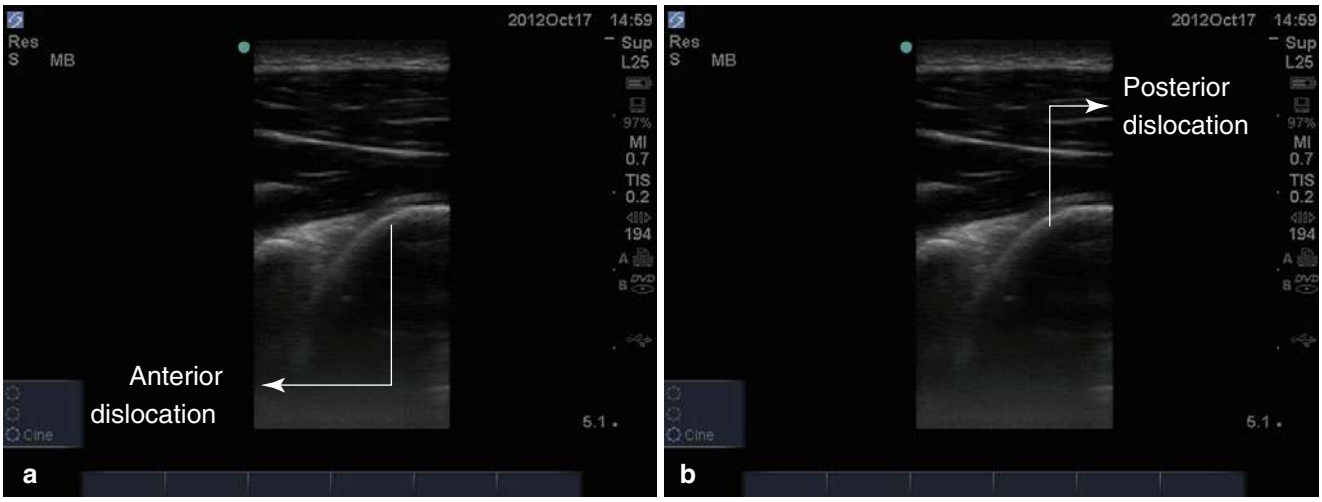

Fig. 100.2 Anterior (**a**) and posterior (**b**) dislocations (Images courtesy of Dr. Rohit Patel)

100.2 Contraindications

- Associated fracture
  - This warrants orthopedic evaluation.
- Associated neurovascular deficit
  - May attempt reduction once but avoid multiple attempts.

100.3 Materials and Medications

- 1 % lidocaine, with syringe and needle and povidone-iodine prep if administering local anesthesia
- Moderate sedation medications if administering moderate sedation
- Bed sheet for traction-countertraction method
- Dangling weight for Stimson maneuver

100.4 Procedure

- Physical examination
  - Compare affected with unaffected shoulder.
  - Perform a complete neurovascular examination: test axillary, radial, ulnar, and median nerves for sensory deficit and motor function.
- Radiographs
  - Always obtain before attempting reduction for assessment of possible fracture and type and position of dislocation.
  - Obtain three views: anteroposterior, scapular Y, and axillary lateral views.
  - Anterior dislocations: humeral head appears anterior to the glenoid fossa on lateral or Y views.
  - Posterior dislocations: on anteroposterior view (vacant glenoid sign, 6-mm sign, lightbulb sign; on lateral or Y view: humeral head appears posterior to glenoid fossa).
- Pain management and sedation
  - Decide whether to use intra-articular lidocaine versus procedural sedation and analgesia.
  - For intra-articular lidocaine
    - Use 10–20 mL of 1 % lidocaine.
    - Attach a 1.5-in., 20-gauge needle.
    - Prepare the shoulder with povidone-iodine.
    - Insert the needle lateral to the acromion process and 2 cm inferiorly into the sulcus.
    - After withdrawing to ensure that the needle is not in a vessel, inject 10–20 mL lidocaine into the joint.
- Reduction techniques: it is important for the emergency department physician to be familiar with several different techniques. The following techniques are presented:

100.4.1 Stimson Maneuver (Fig. 100.3)

1. Patient is placed prone with 2.5–5 kg of weight hanging from the wrist.
2. Reduction may be facilitated by traction and external rotation of the arm.
3. A success rate of 96 % has been reported using the combined prone position, hanging weights, intravenous drug therapy, and scapular manipulation.

- Advantage: can be performed by one person only.
- Disadvantages: requires time to gather materials; the danger involved in the patient falling off the stretcher, requiring staff to monitor the patient.

Fig. 100.3 Stimson maneuver

100.4.2 Scapular Manipulation Technique (Fig. 100.4)

1. Place the patient in the prone position with the affected arm hanging downward.
2. Apply traction down on the arm.
3. Locate the inferior tip of the scapula. Simultaneously push the inferior tip of the scapula medially toward the spine and use the other hand push the superior scapula laterally.

- Advantages: high success rate, greater than 90 %; very safe to perform.
- Disadvantages: it requires the patient to assume the prone position; may require another person to perform traction.

Fig. 100.4 Scapular manipulation method

100.4.3 External Rotation Method (Fig. 100.5)

1. Place the patient in the supine position with the affected arm adducted directly next to the patient's side with the elbow flexed to 90°.
2. The operator uses one hand to direct downward traction on the affected arm while maintaining it next to the patient's side.
3. The operator uses the other hand to hold the patient's wrist and guide the arm into slow external rotation.
4. Reduction usually takes place between 70° and 110° of external rotation.

- Advantages: requires no strength by operator; well tolerated by patients.
- Disadvantage: patient may have persistent dislocation during procedure, requiring operator to make adjustments.

Fig. 100.5 (**a**–**d**) Kocher technique: external rotation method

100.4.4 Milch Technique (Fig. 100.6)

1. Technique looks as though one is reaching up to grab an apple from a tree.
2. Abduct the injured arm up to the overhead position.
3. Once in the overhead position, apply gentle vertical traction with external rotation.
4. An adjustment may need to be made if the reduction does not occur easily; push the humeral head upward into the glenoid fossa.

- Advantages: lack of complications; patient tolerance
- Disadvantage: variable success rate reported: 70–90 %

100.4.5 Spaso Technique (Fig. 100.7)

1. Place the patient in the supine position.
2. Operator grasps the affected arm at the wrist and lifts the straight arm directly upward while applying longitudinal traction.
3. Apply external rotation.

- Advantages: single operator, high level of success
- Disadvantage: may require more time to allow the shoulder muscles to relax

Fig. 100.6 (**a**, **b**) Milch technique

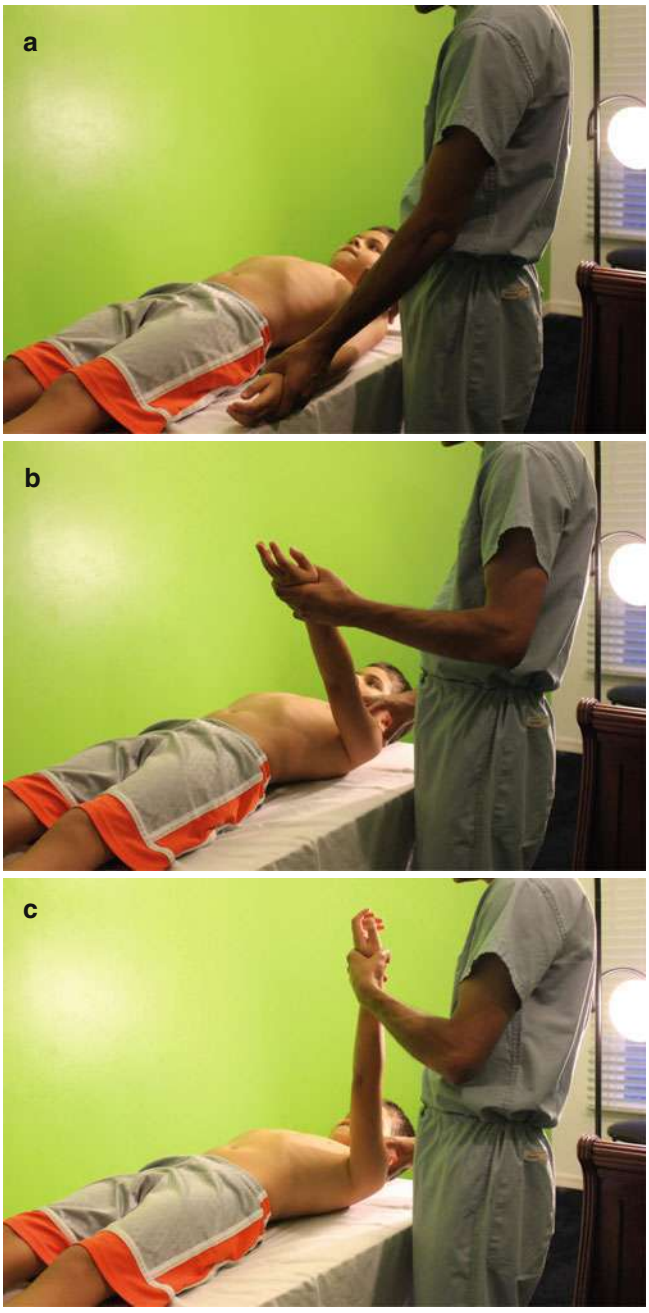

Fig. 100.7 (**a**–**c**) Spaso technique (Photographs courtesy of Dr. Pratik S. Patel)

100.4.6 Traction-Countertraction Technique (Fig. 100.8)

1. With the patient is sitting up, have an assistant wrap a sheet around the upper chest and under the axilla of the affected shoulder. Have the assistant wrap the sheet behind her or his back. Now have the patient lay supine.
2. Wrap another sheet around the flexed elbow of the affected arm and behind the operator's back.
3. Both the operator and the assistant lean back, applying gentle traction.

- Advantage: many older physicians are familiar with this method and, therefore, have a high degree of success.
- Disadvantages: requires two people; may cause skin tears on elderly patients.

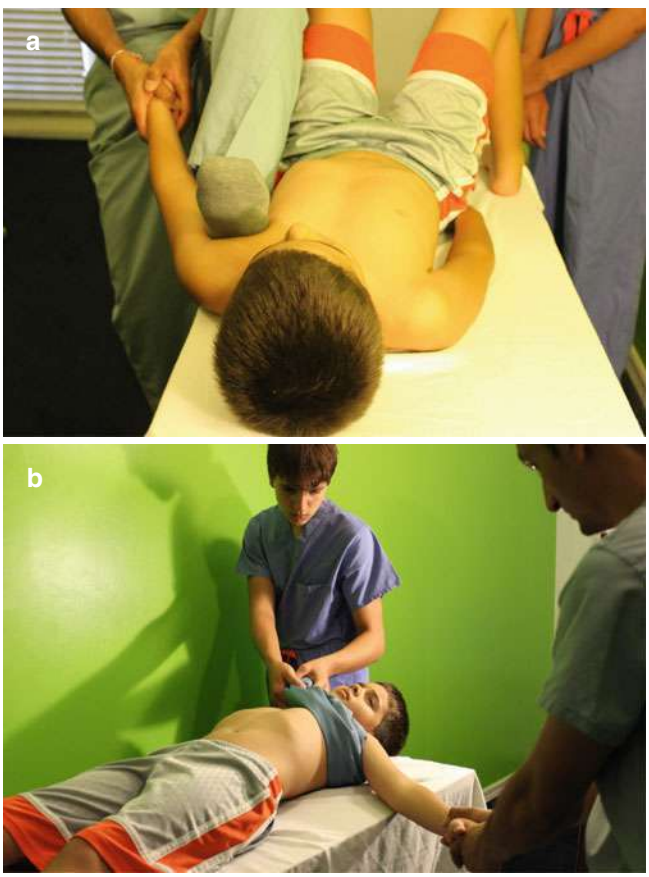

Fig. 100.8 (**a**, **b**) Hippocrates method/traction-countertraction method

100.4.7 Posterior Shoulder Dislocation Reduction

1. Give adequate premedication.
2. Place the patient supine and apply lateral traction on the proximal humerus.
3. Have an assistant apply anterior pressure to the posteriorly located humeral head.

- Advantage: logical methods for reduction
- Disadvantages: require sufficient premedication because often posterior dislocations present late; may require open reduction

100.4.8 Postreduction

- Obtain postreduction x-rays. There is some literature on using ultrasound to confirm adequate reduction, which allows repetitive assessments throughout procedure, as well as reduce radiation (see Fig. 100.2 for ultrasound of anterior and posterior dislocations).
- Do a postreduction neurovascular examination.
- Sling and swath or shoulder immobilizer for 2–3 weeks.
- Orthopedic follow-up in 1 week.

100.5 Complications

- Fractures
- Adhesive capsulitis, or frozen shoulder; especially a concern in the elderly with prolonged immobilization in sling
- Brachial plexus injury, especially of the axillary nerve
- Vascular laceration, most commonly of the axillary artery
- Rotator cuff tears

100.6 Pearls

- It is imperative to document the pre- and postreduction neurovascular status in the medical record.
- If unsure whether the reduction was successful, attempt to place the palm of the injured extremity on the contralateral shoulder. This is a good sign the reduction was successful.

Acknowledgment The authors would like to thank Karthik Stead for serving as the subject in many of the photographs in this chapter.

Selected Reading

Beck S, Chilstrom M. Point-of-care ultrasound diagnosis and treatment of posterior shoulder dislocation. Am J Emerg Med. 2013;31:449.e3–5.

Blakeley CJ, Spencer O, Newman-Saunders T, Hashemi K. A novel use of portable ultrasound in the management of shoulder dislocation. Emerg Med J. 2009;26:662–3.

Dala-Ali B, Penna M, McConnell J, Vanhegan I, Cobiella C. Management of acute anterior shoulder dislocation. Br J Sports Med. 2014;48(16):1209–15.

Simão MN, Noqueira-Barbosa MH, Muqlia VF, Barbieri CH. Anterior shoulder instability: correlation between magnetic resonance arthrography, ultrasound arthrography, and intraoperative findings. Ultrasound Med Biol. 2012;38:551–60.

Yuen CK, Chung TS, Mok KL, Kan PG, Wong YT. Dynamic ultrasonographic sign for posterior shoulder dislocation. Emerg Radiol. 2011;18:47–51.

Elbow Dislocation Reduction

101

Katrina John, Jeffrey Kile, and Amish Aghera

101.1 Indications

- Any dislocation of the elbow joint. Direction of the dislocation (i.e., anterior, posterior, lateral and divergent radius, and ulnar dislocations) is determined by the position of the ulna relative to the joint space (Fig. 101.1).

101.2 Contraindications

- Relative
  - Compound fracture dislocation

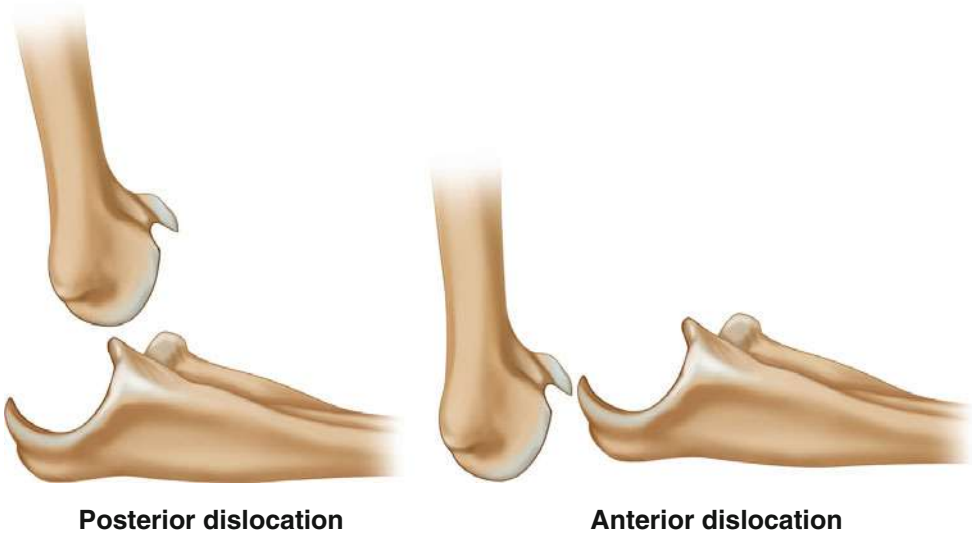

Fig. 101.1 Anatomical depiction

K. John, MBBS • J. Kile, MBBS, PhD, MPH (✉)
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: trenjohn@me.com; jeffrey.kile@gmail.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_101

101.3 Materials and Medications

- Parenteral sedation and analgesia medications
- Local anesthetic for local and intra-articular anesthesia
- Splinting material
- Stockinette
- Padding
- Elastic bandage
- Tape
- Sling

101.4 Procedure for Posterior Dislocations

1. Obtain a true lateral and anteroposterior radiographs of the affected elbow.
2. Ensure adequate sedation and analgesia.
3. Consider intra-articular analgesia.
4. Check the neurovascular status of affected extremity.
5. Follow a selected method for reduction as detailed later.
6. Following successful reduction gently flex the elbow to ensure full range of motion.
7. Place a long-arm posterior splint with the elbow in at least 90° flexion and secure the arm in a regular sling.
8. Check neurovascular status.
9. Obtain a postreduction radiograph of the elbow.

101.4.1 Method A (Fig. 101.2)

1. Position the patient on a stretcher in the supine position.
2. Apply steady traction at the supinated distal forearm keeping the elbow slightly flexed, while an assistant applies countertraction to the midhumerus with both hands.

101.4.2 Method B (Fig. 101.3)

1. Position the patient on a stretcher in the supine position.
2. Extend the affected extremity over the edge of the stretcher.
3. Apply traction to the supinated forearm slightly flexed at the elbow, while an assistant holds the distal humerus with both hands and uses thumbs to apply pressure to the olecranon as if pushing it away from the humerus.

101.4.3 Method C (Fig. 101.4)

1. Position the patient on a stretcher in the prone position.
2. Hang the affected extremity over the side of the stretcher toward the floor.
3. Apply downward traction to the pronated distal forearm and with the other hand just above the patient's antecubital fossa lift the humerus toward you.

Fig. 101.2 Posterior method A

Fig. 101.3 Posterior method B

Fig. 101.4 Posterior method C

101.5 Procedure for Anterior Dislocations (Fig. 101.5)

1. Follow pre- and postprocedure steps as documented for the posterior dislocation.
2. Position the patient on a stretcher in the supine position.
3. With one hand, apply traction to the supinated distal forearm with the elbow extended, while an assistant applies countertraction with both hands around the distal humerus.
4. With the other hand apply downward and backward pressure over the proximal forearm just below the antecubital fossa.

Fig. 101.5 Anterior elbow

101.6 Procedure for Radial Head Subluxations (See Also Chap. 127)

1. This procedure can normally be performed without any sedation or parenteral analgesia.
2. Position the patient, most commonly a child aged 1–3 years, facing forward on the caretaker's lap.
3. Hold the flexed elbow of the affected extremity placing your thumb firmly over the radial head.
4. With the other hand, take the child's hand and wrist, and in one continuous movement, hyperpronate and flex the forearm (Figs. 101.6 and 101.7).
5. Another method is to supinate and flex the forearm instead of hyperpronating it (Fig. 101.8).
6. Leave the room, encourage the caretaker to engage the child with distracting activities and reexamine the child in 10–20 min, at which stage, if reduction was successful, the child should be using the extremity normally again.
7. No postreduction radiograph or immobilization is required.

Fig. 101.6 Subluxation hyperpronated

Fig. 101.7 Subluxation hyperpronated and flexed

Fig. 101.8 Subluxation supinated

101.7 Complications

- Concomitant fractures
- Vascular injury, most commonly to the brachial artery
- Median nerve injury/entrapment
- Recurrent dislocation—rare

101.8 Pearls and Pitfalls

- Pearls
  - A true lateral radiograph is necessary to accurately detect and identify elbow fractures, dislocations, and soft tissue abnormalities (i.e., the fat pad sign). It is obtained with the patient's elbow in 90° flexion, in neutral rotation with the thumb pointing up and the arm and forearm resting on the radiograph cassette and the beam nearly perpendicular to the cassette. On a true lateral, the "hourglass" or "figure-of-eight" formation at the distal humerus should be clearly visible, and the rings of the capitellum and trochlea should be concentric.
  - During nursemaid elbow reduction, provide age appropriate distractions to divert the child's attention and minimize resistance.
- Pitfalls
  - On the pre- and postreduction radiographs, search for commonly associated fractures of the distal humerus, radial head, and coronoid process.
  - Inability to range the elbow after apparent reduction indicates possible trapped fracture fragments and the need for operative intervention.
  - Vascular or open injuries are common with anterior dislocations, and early orthopedic consultation is advised.
  - Ninety percent of simple elbow dislocations are posterior, and this injury is rarely associated with vascular injury. However, it does occur, and vascular evaluation after every reduction is good clinical practice.

Selected Reading

Jain K, Shashi Kumar Y, Mruthyunjaya, Ravishankar R, Nair AV. Posterior dislocation of elbow with brachial artery injury. J Emerg Trauma Shock. 2010;3:308.

Kuhn MA, Ross G. Acute elbow dislocations. Orthop Clin North Am. 2008;39:155–61.

McDonald J, Witelaw C, Goldsmith LJ. Radial head subluxation. Comparing two methods of reduction. Acad Emerg Med. 1999;6:715.

Sheps DM, Hildebrand KA, Boorman RS. Simple dislocations of the elbow: evaluation and treatment. Hand Clin. 2004;20:389–404.

Villarin Jr LA, Belk KE, Freid R. Emergency department evaluation and treatment of elbow and forearm injuries. Emerg Med Clin North Am. 1999;17:843–58.

Distal Interphalangeal Joint Reduction

102

Justin Chen and Muhammad Waseem

Distal interphalangeal (DIP) joint dislocation is rare. It occurs when an axial force is applied to the distal phalanx (Fig. 102.1).

Fig. 102.1 Distal interphalangeal joint dislocation

102.1 Indications

- DIP joint reduction is performed to alleviate functional and anatomical derangements resulting from DIP joint dislocation, commonly dorsal, from axial compression.

102.2 Contraindications

- Absolute
  - Absence of radiographic confirmation (anteroposterior, true lateral, and oblique) of simple DIP joint dislocation, especially in pediatric cases
- Relative
  - Open joint dislocation, associated fracture, or entrapped volar plate
  - Digital neurovascular compromise

102.3 Materials and Medications

- Latex-free gloves
- Local anesthetic: 2 % lidocaine without epinephrine, 1.5 % mepivacaine, 0.5 % ropivacaine, or 0.5 % bupivacaine
- 25-gauge × 1.5-in. needle (can substitute with 27 or 30 gauge)
- Small plastic syringe, 10 mL
- Padded, malleable, aluminum digital splint

J. Chen, MD, MSc
Department of Emergency Medicine, North Shore University Hospital, Manhasset, NY, USA
e-mail: chen.justin@gmail.com

M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_102

102.4 Procedure

1. Place the patient in the seated position with the arms at rest on a bedside table or supported by an assistant.
2. Pronate the patient's hand, remove rings if present, and rest on a flat surface.
3. Insert a 25-gauge needle at the dorsolateral aspect of the base of the finger to form a wheal to reduce patient discomfort.
4. Advance the needle and direct anteriorly toward the phalangeal base.
5. Inject 0.5–1 mL of local anesthetic as the needle is withdrawn 1–2 mm from the point of bone contact.
6. Inject an additional 1 mL of local anesthetic continuously as the needle is withdrawn.
7. The injection should never render the tissue tense nor be circumferential.
8. Hyperextend the DIP joint while applying longitudinal traction, followed by immediate joint flexion at the base of the distal phalanx.
9. Place finger(s) in an aluminum digital dorsal splint in slight flexion for 2 weeks (Fig. 102.2).
10. Postreduction radiograph is recommended for confirmation.

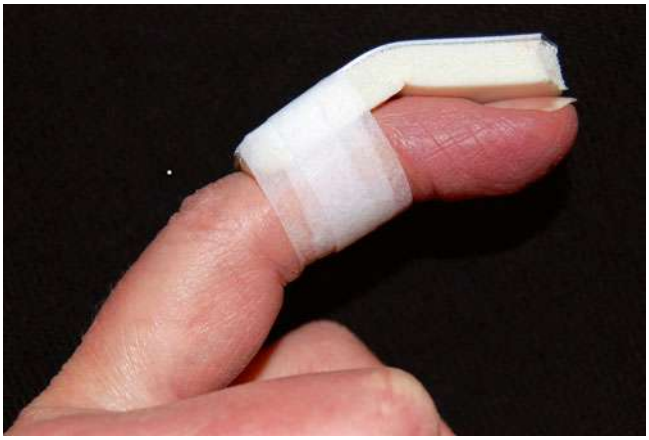

Fig. 102.2 Padded aluminum splint applied to block the DIP joint in flexion but allow further flexion, which encourages active flexion of that joint when the PIP joint flexes (Reproduced with permission from: HandLab Clinical Pearls Feb 2011, No 12. www.handlab.com)

102.5 Complications

- Irreducible dislocations
- Stiffness
- Recurrent dislocation
- Extensor lag in joints with residual subluxation
- Associated with dorsal joint prominences, swan-neck/boutonnière deformity, and degenerative arthritis

102.6 Pearls and Pitfalls

- Pearls
  - Lidocaine without epinephrine is preferred owing to the risk of vasoconstriction of the digital vessels with epinephrine.
  - Regardless of the mechanism of trauma, all joints (DIP, proximal interphalangeal, metacarpophalangeal) should be assessed for instability.
  - Joint dislocations involving volar plate entrapment may require surgical repair (open reduction internal fixation) for successful reduction.
- Pitfalls
  - Irreducible DIP joint dislocations may be due to entrapment of an avulsion fracture, the profundus tendon, or the volar plate.

Selected Reading

Calfee RP, Sommerkamp TG. Fracture-dislocation about the finger joints [review]. J Hand Surg Am. 2009;34:1140–7.

Knoop KJ. Atlas of emergency medicine. 3rd ed. New York: McGraw-Hill Professional; 2010.

Simon RR, Sherman SC, Sharieff GQ. Emergency orthopedics. 6th ed. New York: McGraw-Hill Medical; 2011.

Stone CK, Humphries RL. Current diagnosis & treatment emergency medicine. 6th ed. New York: McGraw-Hill; 2008.

Tintinalli JE, Stapczynski JS, Ma OJ, Cline D, Cydulka R, Meckler G, editors. Tintinalli's emergency medicine: a comprehensive study guide. 7th ed. New York: McGraw-Hill; 2012.

103 Hip Dislocation Reduction

Katrina John, Jeffrey Kile, and Amish Aghera

103.1 Indications

Displacement of the femoral head in relation to the acetabulum without concomitant femoral neck, head, or acetabulum fractures:

- Posterior hip dislocations make up 80–90 % of cases.
- Anterior hip dislocations make up 10–15 % of cases. These are classified into obturator, pubic, iliac, central, or inferior types. Central dislocations are associated with comminuted acetabulum fractures, and inferior dislocations are a rare occurrence normally occurring in children younger than 7 years of age.
- Prosthetic hip dislocations

103.2 Contraindications

- Absolute
  - Femoral neck fracture: attempted reduction may increase the displacement of the fracture and increase the probability of avascular necrosis.
- Relative
  - Fractures in other parts of the affected lower extremity: these may limit the pressure that can be applied necessary for traction during reduction.

103.3 Materials and Medications

- Parenteral sedation and analgesia medications
- Sheet or belt to fix the pelvis to the stretcher
- Knee immobilizer
- Abduction pillow

103.4 Procedure

1. Check the neurovascular status of the affected extremity.
2. Obtain anteroposterior (AP) views of the pelvis and lateral views of the hip.
3. Ensure adequate parenteral sedation and analgesia.
4. Decide upon a technique, as detailed later, and position the patient accordingly.
5. Once the hip has been successfully reduced, test the joint for stability by moving it gently thought its range of motion.
6. Place a knee immobilizer and an abduction pillow between the knees.
7. Check the neurovascular status.
8. Obtain repeat AP films of the pelvis.

K. John, MBBS • J. Kile, MBBS, PhD, MPH (✉)
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: trenjohn@me.com; jeffrey.kile@gmail.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_103

103.4.1 Stimson Maneuver

1. Place the patient prone on the stretcher with the affected extremity hanging over the edge and the hip flexed to 90°.
2. Flex the knee and the foot to 90°.
3. Apply downward pressure to the area just distal to the popliteal fossa with a hand (Fig. 103.1) or knee (Fig. 103.2) while using the opposite hand to internally and externally rotate the hip at the ankle.
4. Have an assistant simultaneously manipulate the displaced femoral head into position with both hands, applying downward pressure over the affected buttock (Fig. 103.3).

Fig. 103.1 Stimson maneuver with hand

Fig. 103.2 Stimson maneuver with knee

Fig. 103.3 Manipulation of the femoral head

103.4.2 Allis Maneuver

1. Position the patient supine on the stretcher.
2. The operator should stand on the stretcher to achieve maximum leverage or have the patient on a backboard on the ground.
3. Have an assistant apply downward pressure to both iliac crests.
4. Apply constant, gentle upward traction in line with the deformity while maneuvering the hip to 90° flexion and through internal and external rotation (Fig. 103.4).
5. Have a second assistant provide lateral traction to the midthigh.
6. Once the femoral head has cleared the outer lip of the acetabulum, continue traction while keeping the hip in external rotation and gently abducting and extending the hip (Fig. 103.5).

Fig. 103.4 Allis flexion

Fig. 103.5 Allis extension

103.4.3 Whistler Technique

1. Position the patient supine on the stretcher with the knee and hip flexed to 45°.
2. Have an assistant stabilize the pelvis with downward pressure on both iliac crests.
3. Stand on the side of the affected extremity and place one arm under the knee, resting the hand on the flexed knee of the unaffected extremity.
4. Secure the ankle of the affected extremity with the other hand and elevate the shoulder of the opposite arm, providing upward traction at the distal thigh and a strong fulcrum to reduce the dislocation (Fig. 103.6).
5. Internal and external rotation can be achieved with the opposite hand at the ipsilateral ankle.

103.4.4 Captain Morgan Technique

1. Position the patient supine on the stretcher with the knee and hip flexed to 90°.
2. Stabilize and fix the pelvis with a sheet tied securely over the pelvis and under the stretcher.
3. Standing on the side of the affected extremity, the operator's foot should be resting perpendicular on the stretcher with the knee placed under the patient's knee.
4. With the opposite hand, apply downward pressure to the ankle and provide a sustained upward force to the patient's thigh by elevation of the knee through plantar flexion of the toes and upward pressure of the other hand placed behind the patient's knee.
5. Internal and external rotation can be applied simultaneously if necessary by gently twisting the ankle (Fig. 103.7).

Fig. 103.6 Whistler technique

Fig. 103.7 Captain Morgan technique

103.5 Complications

- Sciatic nerve injury
- Avascular necrosis of the femoral head due to delay in adequate reduction
- Inability to perform reduction due to occult fractures and fracture fragments, incarceration of the joint capsule, or associated tendons
- Unstable or irreducible dislocations
- Traumatic arthritis and joint instability

103.6 Pearls and Pitfalls

- Pearls
  - On AP radiograph, posterior dislocations can be more easily detected by the presence of a smaller femoral head compared with the unaffected side and poor visualization of the lesser trochanter.
  - On AP radiographs, anterior dislocations can be detected by a larger femoral head and a clear lesser trochanter seen in profile alongside the femoral shaft.
  - Pay close attention to the femoral vessels and the sciatic nerve. Injury to the sciatic nerve most commonly affects the common peroneal branch, therefore causing weakness in great toe extension and foot dorsiflexion. Sensation may also be reduced over the dorsum of the foot.
  - Check the femoral head is intact and clearly in the acetabulum and for intact Shenton lines, symmetrical intra-articular spaces, and clear outlines of the lesser trochanters.
  - For any of the techniques requiring stabilization of the pelvis, an alternative is to fix it to the stretcher using a sheet or belt.
  - To overcome the powerful muscles that oppose successful reduction, it is important to provide adequate muscle relaxation and steady, prolonged traction.
  - An assistant should stand on the floor behind to support the operator if standing on a stretcher.
- Pitfalls
  - Owing to the force necessary to dislocate a native hip, this injury should serve as a red flag to the physician to consider other potentially life- or limb-threatening occult injuries.
  - Hip dislocation is a true orthopedic emergency and must be treated without delay. Delay in reduction, especially greater than 6 h, results in increased incidence of avascular necrosis of the femoral head and sciatic nerve injury.
  - Review imaging carefully because associated fractures of the femoral head, neck, and acetabulum are often present.
  - It is recommended that anterior dislocations be reduced by orthopedic surgeons under general anesthetic in the operating room. These are often more complicated and difficult to reduce, and failure at closed reduction in the operating room can be followed by an open procedure.
  - Multiple attempts at reduction should not be performed in the emergency department because these are unlikely to be successful and will only delay definitive management and lead to an increased risk of complications.

Selected Reading

Hendey GW, Avila A. The captain Morgan technique for the reduction of the dislocated hip. Ann Emerg Med. 2011;58:536–40.

Newton EJ, Love J. Emergency department management of selected orthopedic injuries. Emerg Med Clin North Am. 2007; 25:763–93.

Nordt WE. Maneuvers for reducing dislocated hips. Clin Orthop Relat Res. 1999;360:160–4.

Rupp JD, Schneider LW. Injuries to the hip joint in frontal motor-vehicle crashes: biomechanical and real-world perspectives. Orthop Clin North Am. 2004;35:493–504.

Walden PD, Hamer JR. Whistler technique used to reduce traumatic dislocation of the hip in the emergency department setting. J Emerg Med. 1999;17:441–4.

Knee Dislocation Reduction

104

Katrina John, Jeffrey Kile, and Amish Aghera

104.1 Indications

- Dislocation of the knee/fibular head/patella

104.2 Contraindications

- Absolute
  - None
- Relative
  - Immediate availability of orthopedic consultation

104.3 Materials and Medications

- Parenteral sedation and analgesia medications
- Knee immobilizer or splinting materials

104.4 Procedure

104.4.1 Knee (Femur/Tibia) Dislocation Reduction

1. Assess neurovascular function.
2. Pretreat the patient with sedation or analgesia as appropriate.
3. Position the patient supine with the affected leg fully extended.
4. Instruct an assistant to stand near the patient's hip and, facing the patient's affected knee, grasp the distal femur firmly with both hands to fix it in place.
5. Stand near the patient's foot and, facing the patient's affected knee, grasp the distal tibia and apply straight traction in a distal direction.
   - Longitudinal traction-countertraction alone, as described previously, will usually reduce the dislocation. If reduction does not occur, proceed with the following steps.
6. While applying straight traction in a distal direction to the tibia with the dominant hand, with the nondominant hand:
   (a) Anterior dislocation: push the proximal tibia in a posterior direction (Fig. 104.1)
   (b) Posterior dislocation: lift the proximal tibia in an anterior direction (Fig. 104.2)
   (c) Lateral dislocation: push the proximal tibia in a medial direction (Fig. 104.3)
   (d) Medial dislocation: push the proximal tibia in a lateral direction (Fig. 104.4)
   (e) Rotary dislocation: rotate the proximal tibia into proper linear alignment with the femoral condyles (Fig. 104.5)
   - Reduction may be facilitated by the use of two assistants rather than just one. The second assistant grasps the distal tibia and applies straight traction in a distal direction, freeing the operator to manipulate the proximal tibia as described previously using both hands.
7. After reduction, reassess neurovascular function and, if available, obtain angiography.
8. Immobilize the knee in 15° of flexion in a knee immobilizer or long-leg posterior splint.

K. John, MBBS • J. Kile, MBBS, PhD, MPH (✉)
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: trenjohn@me.com; jeffrey.kile@gmail.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_104

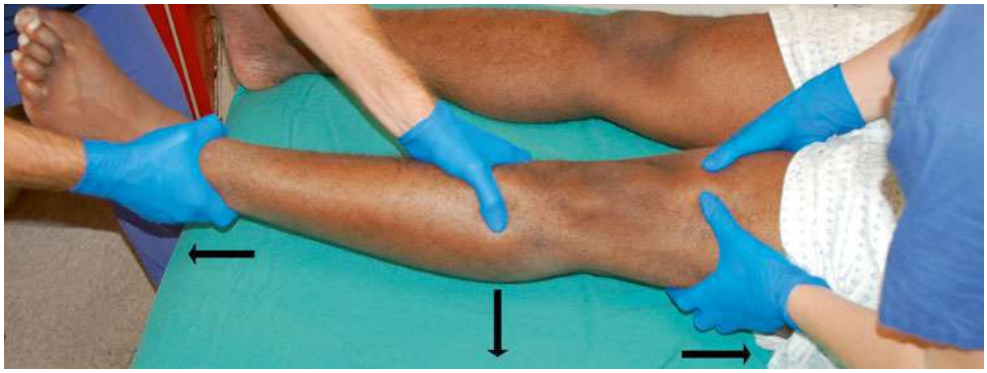

Fig. 104.1 Anterior dislocation of the knee: the proximal tibia is pushed in a posterior direction. The *arrows* indicate the direction in which force should be applied by the operator during reduction of dislocation

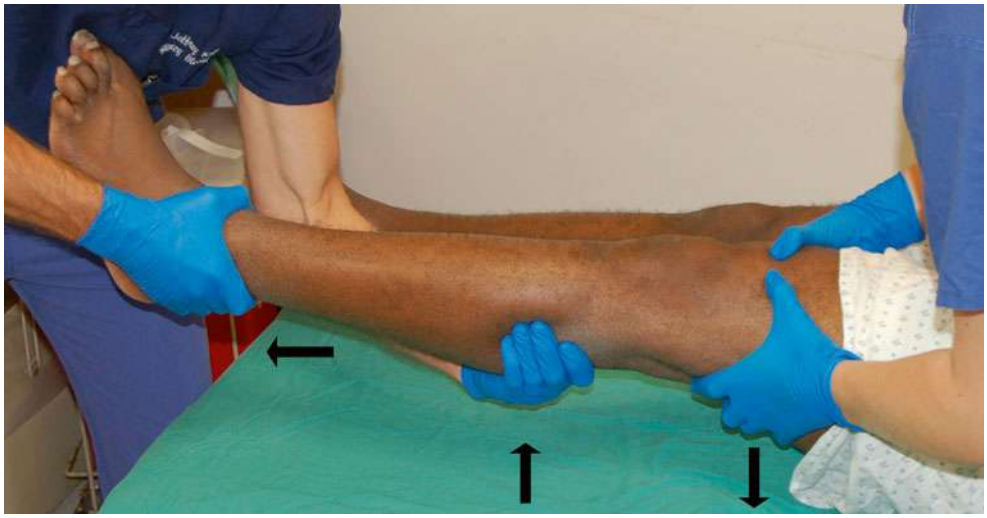

Fig. 104.2 Posterior dislocation of the knee: the proximal tibia is pushed in an anterior direction. The *arrows* indicate the direction in which force should be applied by the operator during reduction of dislocation

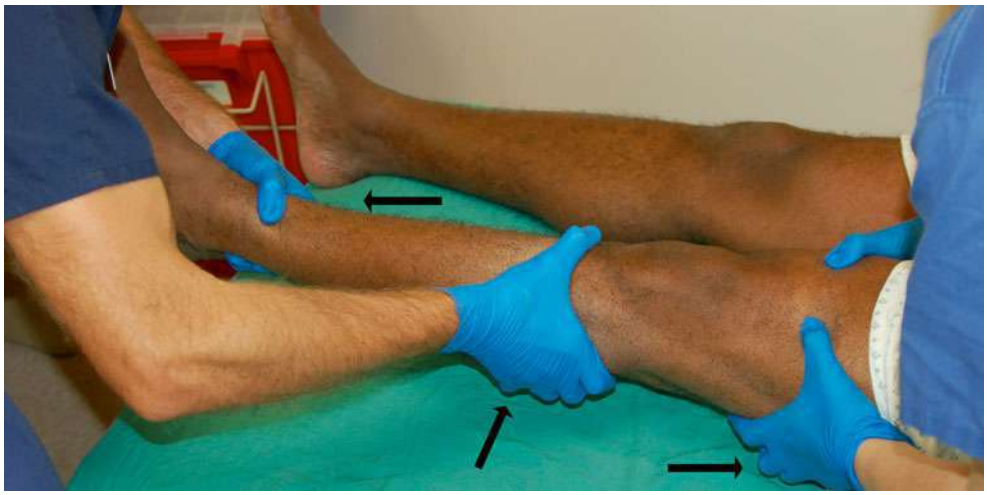

Fig. 104.3 Lateral dislocation of the knee: the proximal tibia is pushed in a medial direction. The *arrows* indicate the direction in which force should be applied by the operator during reduction of dislocation

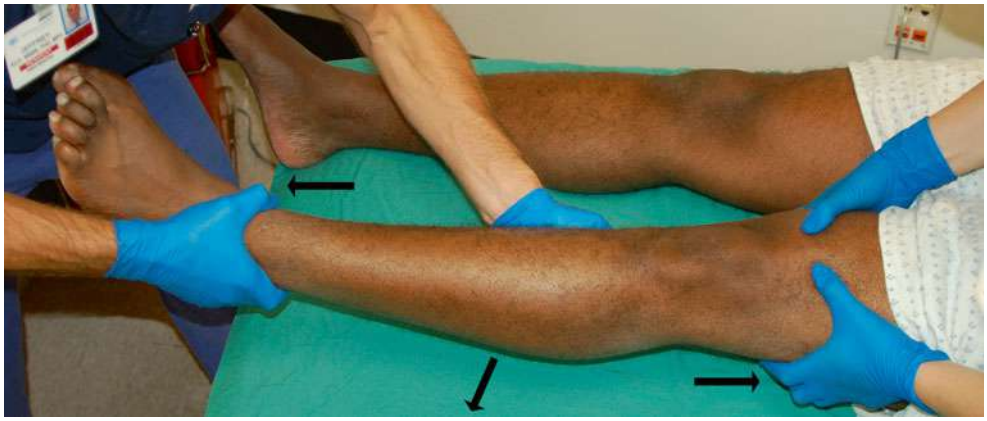

Fig. 104.4 Medial dislocation of the knee: the proximal tibia is pushed in a lateral direction. The *arrows* indicate the direction in which force should be applied by the operator during reduvction of dislocation

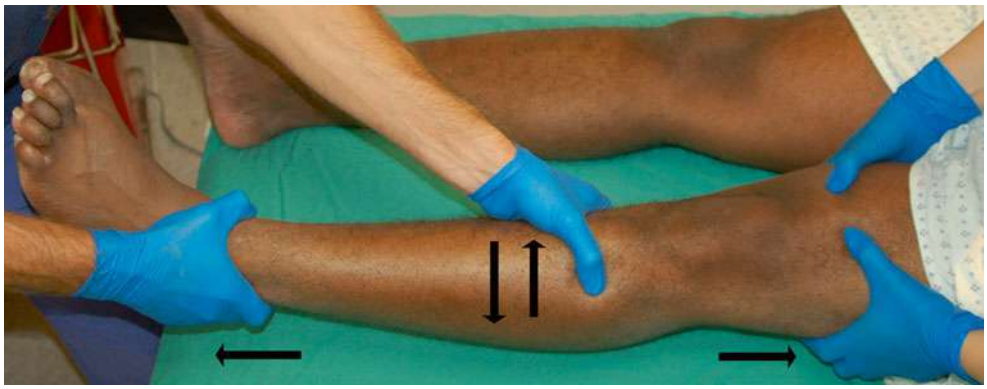

Fig. 104.5 Rotary dislocation of the knee: the proximal tibia is rotated into proper alignment with the femoral condyles. The *arrows* indicate the direction in which force should be applied by the operator during reduction of dislocation

104.4.2 Fibular Head Dislocation Reduction

1. Assess neurovascular function.
2. Pretreat the patient with sedation or analgesia as appropriate.
3. Position the patient supine.
4. Flex the knee to 90° to relax the biceps femoris tendon.
5. Instruct an assistant to stand near the patient's hip and, facing the patient's affected knee, grasp the distal femur firmly with both hands to fix it in place.
6. Stand near the patient's foot and, facing the patient's affected knee, grasp the distal tibia and apply straight traction in a distal direction with the dominant hand and with the nondominant hand.
   (a) Anterior dislocation: push the fibular head in a posterior direction (Fig. 104.6)
   (b) Posterior dislocation: push the fibular head in an anterior direction (Fig. 104.7)
      - Reduction may be facilitated by the use of two assistants rather than just one. If a second assistant is available, instruct the second assistant to stand near the patient's foot and, facing the patient's affected knee, grasp the distal tibia and apply straight traction in a distal direction. This enables the operator to grasp and move the proximal fibula as described previously using both hands.
      - Reduction is often signified by a palpable and audible click as the fibula snaps back into position.
7. After reduction, reassess neurovascular function and, if available, obtain angiography.
   - After reduction, patients should receive orthopedic referral, avoid weight-bearing for the first 2 weeks, and then gradually increase weight-bearing over the next 6 weeks.
   - Typically, immobilization is not required following reduction of an i.solated fibular head dislocation.

Fig. 104.6 Anterior dislocation of the fibular head: the fibular head is pushed in a posterior direction. The *arrow* indicate the direction in which force should be applied by the operator during reduction of dislocation

Fig. 104.7 Posterior dislocation of the fibular head: the fibular head is pushed in an anterior direction. The *arrow* indicate the direction in which force should be applied by the operator during reduction of dislocation

104.4.3 Lateral Patellar Dislocation Reduction

1. Pretreat the patient with sedation or analgesia as appropriate.
2. Stand at the side of the affected knee and, facing the knee, grasp the distal tibia and slowly extend the knee with one hand, and with the other hand simultaneously apply gentle pressure to the patella in a medial direction.

- The lateral edge of the patella may be lifted slightly to facilitate its travel over the femoral condyle during reduction (Fig. 104.8).
- After reduction, the knee should be immobilized in full extension in a knee immobilizer or long-leg posterior splint, and the patient should receive orthopedic referral, avoid weight-bearing for the first 2 weeks, and then gradually increase weight-bearing over the next 6 weeks.

Fig. 104.8 Lateral dislocation of the patella: the patella is pushed in a medial direction. The *arrow* indicate the direction in which force should be applied by the operator during reduction of dislocation

104.5 Complications

104.5.1 Knee (Femur/Tibia) Dislocations

- Distal ischemia (even requiring amputation)
- Degenerative arthritis
- Joint instability due to ligamentous injury

104.5.2 Fibular Head Dislocations

- Peroneal nerve injury
- Fibular head instability/subluxation
- Degenerative arthritis

104.5.3 Patellar Dislocations

- Failure of reduction
- Degenerative arthritis
- Recurrent dislocation/subluxation

104.6 Pearls and Pitfalls

104.6.1 Knee (Femur/Tibia) Dislocations

- Pearls
  - Dislocations of the knee are described in terms of the tibia's position in relation to the femur.
  - All knee dislocations require orthopedic evaluation at the earliest possible opportunity.
  - Owing to the frequency of associated popliteal artery and peroneal nerve injury, a neurovascular examination should be performed before and after any attempts at reduction or manipulation of the knee.
  - Dislocations of the knee should be reduced as soon as possible, particularly if distal neurovascular compromise exists.
  - Operative ligamentous repair is often required approximately 2 weeks postreduction (once acute swelling has resolved) to achieve the maximum functional recovery.
- Pitfalls
  - If the knee hyperextends more than 30° when the horizontal leg is lifted by the foot, the knee is considered severely unstable. This is likely due to a previous dislocation, and thus, the knee should be evaluated for the neurovascular complications of dislocation.
  - Because the joint capsule is commonly disrupted during knee dislocation, synovial fluid may diffuse into the surrounding tissue, such that an effusion is not always present.
  - A posterolateral dislocation may be irreducible because the medial femoral condyle traps the medial capsule within the joint.

104.6.2 Fibular Head Dislocations

- Pearls
  - Fibular head dislocations are usually anterolateral, but these do not result in neurovascular compromise.
  - A knee joint effusion is usually not seen in a fibular head dislocation because the tibiofibular ligaments are contained within a separate synovium.
  - Anterior dislocations typically result from a fall on the flexed, adducted leg, often combined with ankle inversion.
  - Flexion of the knee relaxes the fibular collateral ligament, reducing the stability of the tibiofibular joint.
  - Superior dislocation is accompanied by interosseus membrane damage and proximal displacement of the lateral malleolus.
- Pitfalls
  - Posterior fibular head dislocations usually result from direct trauma to the flexed knee and may be accompanied by peroneal nerve injury.

104.6.3 Patellar Dislocations

- Pearls
  - Patellar dislocation occurs most frequently among adolescents.
  - Patellar dislocation typically occurs in the setting of external rotation combined with a strong valgus force and quadriceps contraction.
  - Patellar dislocations are described in terms of the patellar relationship to the normal knee joint.
  - The most common patellar dislocations are lateral.
  - If a spontaneous reduction has occurred, a knee effusion and tenderness along the medial aspect of the patella are likely to be present on examination, and the patellar apprehension test will be positive.
  - To perform the patellar apprehension test, flex the knee to 30° and push the patella laterally. If the patient senses an impending redislocation, the test is considered positive.

  - Isolated lateral patellar dislocations do not usually require hospitalization, but orthopedic follow-up is recommended owing to the likelihood of persistent instability.
  - Intracondylar and superior dislocations require surgical reduction.
  - Patients with an isolated patellar dislocation typically present with the knee in 20–30° of flexion and the patella displaced laterally.
- Pitfalls
  - Dislocations tend to be recurrent, particularly in patients with patellofemoral anatomical abnormalities.

Selected Reading

Martinez D, Sweatman K, Thompson EC. Popliteal artery injury associated with knee dislocations. Am Surg. 2001;67:165–7.

Peskun CJ, Levy BA, Fanelli GC, et al. Diagnosis and management of knee dislocations. Phys Sportsmed. 2010;38:101–11.

Rihn JA, Groff YJ, Harner CD, Cha PS. The acutely dislocated knee: evaluation and management. J Am Acad Orthop Surg. 2004;12:334–46.

Roberts DM, Stallard TC. Emergency department evaluation and treatment of knee and leg injuries. Emerg Med Clin North Am. 2000;18:67–84. v–vi.

Wascher DC, Dvirnak PC, DeCoster TA. Knee dislocation: initial assessment and implications for treatment. J Orthop Trauma. 1997;11:525–9.

105 Ankle Dislocation Reduction

Katrina John, Jeffrey Kile, and Amish Aghera

105.1 Indications

- Dislocation of the ankle joint. This is defined by the articulation of the talus with the mortise that is formed by the distal tibia and fibula. Dislocations can be posterior, anterior, superior, or lateral and are classified by the position of the talus in relation to the tibial mortise.

105.2 Contraindications

- Relative
  - Open dislocations where there is no evidence of acute neurovascular compromise are better managed definitively in the operating room to avoid further contamination.

105.3 Materials and Medications

- Parenteral sedation and analgesia medications
- Local anesthetic for local and intra-articular anesthesia
- Splinting material
- Stockinette
- Padding
- Elastic bandage
- Tape
- Sheet

105.4 Procedure

1. Check the neurovascular status of the affected foot and ankle.
2. If there is no evidence of critical neurovascular compromise, obtain a lateral and an anteroposterior radiograph of the affected ankle.
3. Ensure adequate parenteral sedation and analgesia to maximize success and limit pain and suffering.
4. Position the patient on a stretcher with the knee flexed at 90° over a folded pillow or rolled-up sheet or with the lower leg and knee hanging over the edge of the stretcher.

K. John, MBBS • J. Kile, MBBS, PhD, MPH (✉)
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: trenjohn@me.com; jeffrey.kile@gmail.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_105

105.4.1 Posterior Dislocations

1. Hold the heel in one hand and pull with longitudinal traction.
2. With the other hand, hold the top of the foot and gently plantarflex it downward, while an assistant provides countertraction at the back of the midcalf (Fig. 105.1).
3. Continue longitudinal traction at the heel and countertraction at the calf.
4. Dorsiflex the foot while another assistant applies downward pressure to the distal anterior leg (Fig. 105.2).
5. Examine foot for restoration of normal anatomy and for any new lacerations or defects to the skin.
6. Recheck neurovascular integrity.
7. Place the leg in a sugar-tong splint with the foot at 90°.
8. Recheck neurovascular integrity.

Fig. 105.1 Plantarflexion with longitudinal heel traction

Fig. 105.2 Dorsiflexion with longitudinal heel traction

105.4.2 Anterior Dislocations

1. Hold the heel in one hand and pull with longitudinal traction.
2. With the other hand, hold the top of the foot and dorsiflex, while an assistant provides countertraction at the back of the midcalf (Fig. 105.3).
3. Continue longitudinal traction at the heel and countertraction at the calf.
4. Keeping the foot at 90° to the leg, hold the foot firmly and push the foot downward toward the floor while another assistant applies upward pressure to the distal posterior leg (Fig. 105.4).
5. Examine the foot for restoration of normal anatomy and for any new lacerations or defects to the skin.
6. Recheck neurovascular integrity.
7. Place the leg in a sugar-tong splint with the foot at 90°.
8. Recheck neurovascular integrity.

Fig. 105.3 Dorsiflexion with longitudinal heel traction

Fig. 105.4 Downward movement of foot (toward the floor) with longitudinal heel traction

105.5 Complications

- Compound fractures
- Neurovascular injury
- Skin and soft tissue damage
- Compartment syndrome

105.6 Pearls and Pitfalls

- Pearls
  - The ankle rarely dislocates without associated fractures.
- Pitfalls
  - Ankle dislocation is an orthopedic emergency, and reduction should not be delayed by imaging if there is evidence of neurovascular impairment. Complications that are exacerbated by delay in management include concomitant fractures, gross deformity of the ankle, severe stretching and tenting of the skin with resultant skin blisters, skin necrosis, and possible conversion to a compound fracture.
  - Be sure to check the radiograph carefully for commonly associated fractures notably of the malleoli.

Selected Reading

Collins DN, Temple SD. Open joint injuries: classification and treatment. Clin Orthop. 1989;243:48.

Hamilton WC. Injuries of the ankle and foot. Emerg Med Clin North Am. 1984;2:361.

Kelly PJ, Peterson FP. Compound dislocations of the ankle without fractures. Am J Surg. 1986;103:170.

Simon RR, Sherman SC, Koenigsknecht SJ, editors. Emergency orthopedics—the extremities. 5th ed. New York: McGraw-Hill; 2007. p. 264.

Wedmore IS, Charette J. Emergency department evaluation and treatment of ankle and foot injuries. Emerg Med Clin North Am. 2000;18:85.

106 Arthrocentesis

Shalu S. Patel and Bobby K. Desai

106.1 Indications

- Diagnosis of septic joint
- Diagnosis of traumatic effusion
- Diagnosis of inflammatory effusion
- Diagnosis of crystal-induced arthritis
- Therapeutic relief of pain from effusion

106.2 Contraindications

- Severe coagulopathy
- Skin infection over the needle insertion site
- Joint prosthesis
- Patients with bacteremia or sepsis (except to diagnose a septic joint)

106.3 Materials and Medications

- Betadine (povidone-iodine) or other skin antiseptic
- Sterile gloves
- Sterile towels
- Lidocaine 1 % or 2 % (5 mL) or other anesthetic of choice
- 18- to 22-gauge needle, 25-gauge needle
- Syringes (5 mL, 5–50 mL)
- Sterile gauze (4×4)
- Band-aid

106.4 Procedure

1. Informed consent may be required.
2. Position the patient appropriately. The joint should be placed in slight flexion.
3. Palpate the joint and identify anatomical landmarks.
   (a) For knee arthrocentesis, the needle should be inserted at the midpoint of either the medial or the lateral side of the patella (Fig. 106.1).
   (b) For acromioclavicular (AC) joint arthrocentesis, the needle should be inserted at the superior surface of the AC joint (Fig. 106.2).
   (c) For glenohumeral joint arthrocentesis, there are two approaches.
      (i) In the anterior approach, the needle is inserted into the groove lateral to the coracoid process (Fig. 106.3).
      (ii) In the posterior approach (preferred), the needle is inserted below the posterior border of the acromion process and lateral to the border of the scapula (Fig. 106.4).
4. Prepare the skin and drape in a sterile fashion.
5. Using lidocaine (drawn up in 5-mL syringe), anesthetize the skin with the 25-gauge needle.
6. Secure the 18- to 22-gauge needle on the 5- to 50-mL syringe (depending on the size of the joint) and insert it into the skin.
7. Advance the needle slowly into the joint space while aspirating until joint fluid can easily be withdrawn. While inserting the needle into the joint space, avoid scraping the needle against the bone.

S.S. Patel, MD
Department of Emergency Medicine, Florida Hospital Tampa, Florida Hospital Carrollwood, Tampa, FL, USA

B.K. Desai, MD (✉)
Department of Emergency Medicine, University of Florida Health Shands Hospita, Gainesville, FL, USA
e-mail: bdesai@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_106

8. If fluid cannot be aspirated easily, the catheter can be repositioned further in the joint space or turned by 45° sequentially as needed.
9. Once the joint fluid is aspirated, pull out the needle and hold pressure with gauze. Bleeding should be minimal.
10. Place a band-aid or other dressing over the site.
11. Send the synovial fluid to the laboratory. Generally, laboratory analyses may include crystals, protein, glucose, cell count and differential, culture and sensitivity, and Gram stain.

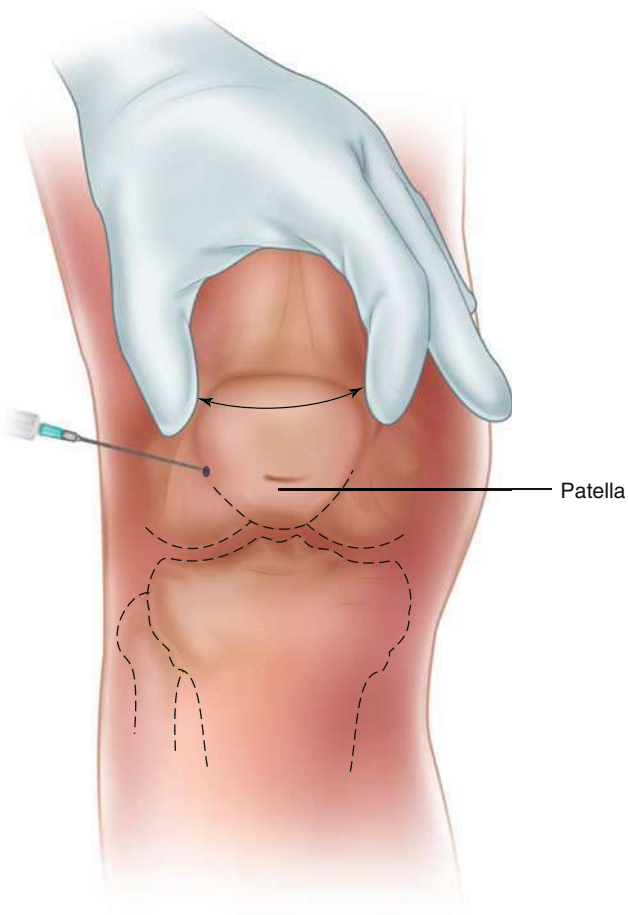

Fig. 106.1 Knee arthrocentesis

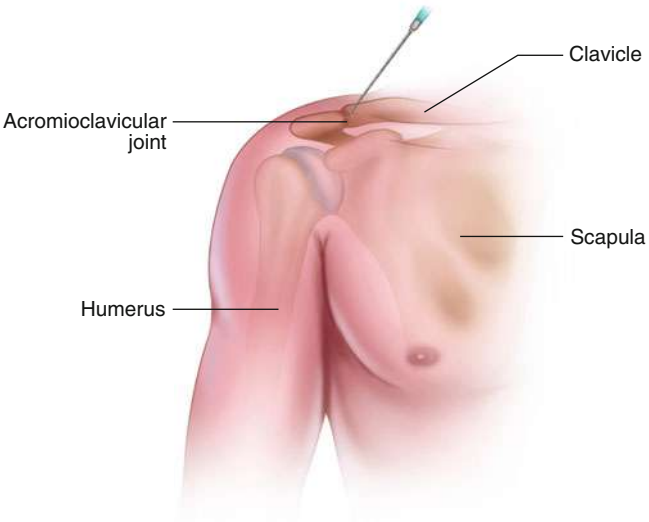

Fig. 106.2 Acromioclavicular joint arthrocentesis

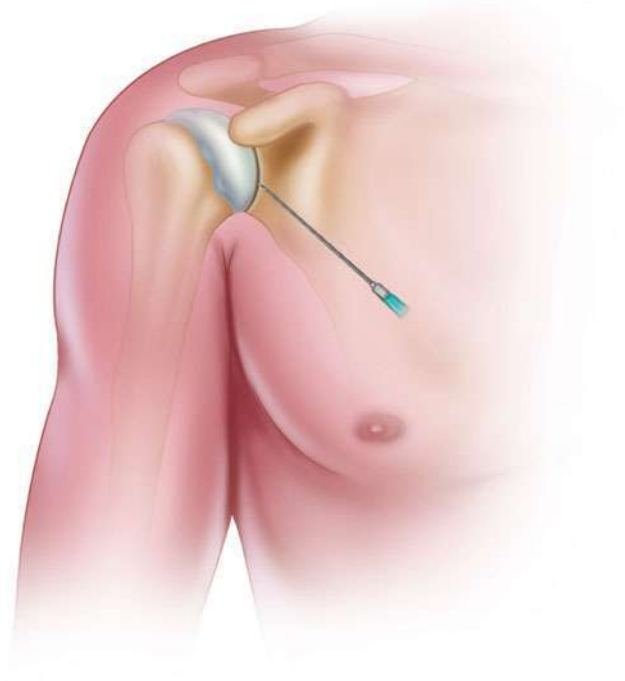

Fig. 106.3 Glenohumeral joint arthrocentesis: anterior approach

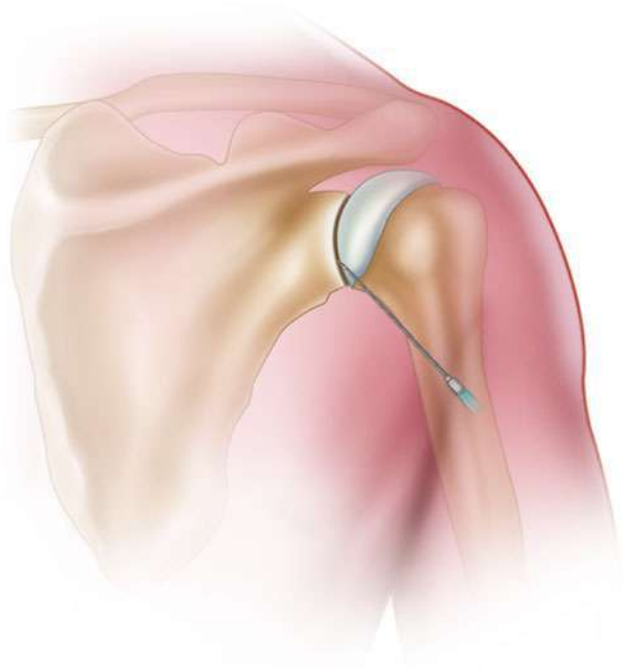

Fig. 106.4 Glenohumeral joint arthrocentesis: posterior approach

106.5 Complications

- Introduction of infection
- Bleeding

106.6 Pearls and Pitfalls

- The preferred site of entry is over the extensor surface of the joint. This will reduce the risk of damage to tendons, ligaments, and blood vessels.
- When assessing synovial fluid, the Rule of Twos may be used to differentiate among normal, inflammatory, and septic fluid. Normal synovial fluid has less than 200 white blood cells (WBCs)/mm^3. Noninflammatory synovial fluid has 200–2000 WBCs/mm^3. Inflammatory synovial fluid has greater than 2000 WBCs/mm^3 (but <50,000 WBCs/mm^3). Septic synovial fluid has greater than 75,000 WBCs/mm^3.
- Only septic synovial fluid will have a positive Gram stain and culture.

Selected Reading

Biundo JJ, Roberts N, Deodhar A. Regional musculoskeletal complaints. In: Stone JH, editor. A clinician's pearls and myths in rheumatology. New York: Springer Science; 2009. p. 433–4.

Parrillo SJ, Fisher J. Arthrocentesis. In: Roberts JR, Hedges J, editors. Clinical procedures in emergency medicine. 4th ed. Philadelphia: Saunders; 2004. p. 1042–57.

Self WH, Wang EE, Vozenilek JA, del Castillo J, Pettineo C, Benedict L. Dynamic emergency medicine. Arthrocentesis. Acad Emerg Med. 2008;15:298.

Thomsen TW, Shen S, Shaffer RW, Setnik GS. Arthrocentesis of the knee. N Engl J Med. 2006;354:e19.

107 Intra-articular Injection

Bharat Kothakota and Muhammad Waseem

107.1 Indications

- Aspiration of fluid (arthrocentesis)
  - For diagnosis: to rule out infection
  - To relieve pressure from large, painful joint effusion
- Injection of joints with inflammatory arthritis
  - Tendinitis
  - Bursitis
  - Rheumatoid arthritis (RA)
- Injection of joints with osteoarthritis (OA)
  - Injection of large weight-bearing joints
  - Injection of small joints of hands
- Intra-articular anesthetic
  - Shoulder reduction
  - Ankle impingement syndrome

107.2 Contraindications

- Cellulitis
- Bacteremia
- Fracture

107.3 Materials and Medications

- Glucocorticoid
  - Duration of effect inversely proportional to solubility
    - Less soluble → longer acting
  - Choice of steroid is the personal preference of the physician
  - Methylprednisone (Depo-Medrol) and triamcinolone acetonide (Kenalog)
    - Less likely to induce postinjection flare
  - Kenalog and triamcinolone hexacetonide (Aristospan)
    - Longest-acting agents
- Local anesthetic
  - 1 % Lidocaine
- Needle
  - 18- to 22-gauge used for knee, ankle, hip, elbow, and shoulder
  - 25-gauge or smaller used for smaller joints (interphalangeal)
- Syringe

B. Kothakota, MD • M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: bharatkothakota@yahoo.com; waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016

L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_107

107.4 Procedure

1. Selecting an injection approach.
   - Knee
     - Lateral approach: 1 cm inferior to the patella laterally (Fig. 107.1)
     - Medial approach: 1 cm inferior to the patella medially (Fig. 107.2)
   - Ankle
     - Lateral approach: just inferior to the lateral malleolus (Fig. 107.3)
     - Medial approach
       - Plantar flex the foot.
       - Angle the needle cephalad to pass between the medial malleolus and the tibialis anterior tendon (Fig. 107.4).
   - Shoulder
     - Posterior approach
       - Insert the needle 1 cm inferior and 1 cm medial to the posterolateral corner of the acromion.
       - Direct the needle anterior and medial toward the coracoid process (Fig. 107.5).
2. Skin preparation.
   - Make three separate concentric outward spirals with iodine disinfectant
   - Scrub with cyclohexidine preparation
3. Mark the injection site by impressing the skin with a hard object.
   - Sterile end of needle sheath
   - Ball point pen with tip retracted
4. Local anesthesia.
   - 1 % Lidocaine injected into the skin and subcutaneous tissue
   - Short burst of ethyl chloride spray before iodine preparation
   - Mixing lidocaine with glucocorticoid preparation
5. Always aspirate joint fluid before injecting the corticosteroids.
   - Use a 1.5-inch 18-gauge needle for aspiration.
   - Confirm that the needle is in the joint space.
   - Reduced effusion size before injection can improve outcomes.
   - Compress the opposite side of the joint to aid in aspiration (Fig. 107.6).
6. After aspirating, change the syringe.
   - Use a sterile hemostat or hand to stabilize the needle within the joint space (Fig. 107.7).
   - Avoid injecting corticosteroids if the aspirate appears purulent.
7. Injection of medication.
   - Can use the same needle used for aspiration.
   - Insert needle 0.75–1.25 inch in depth for injection.
8. Remove the needle, wipe the iodine solution clean, and apply the bandage.
9. Postinjection care.
   - First 48 h: bedrest versus minimize walking
   - Next 2–3 weeks: crutches or cane

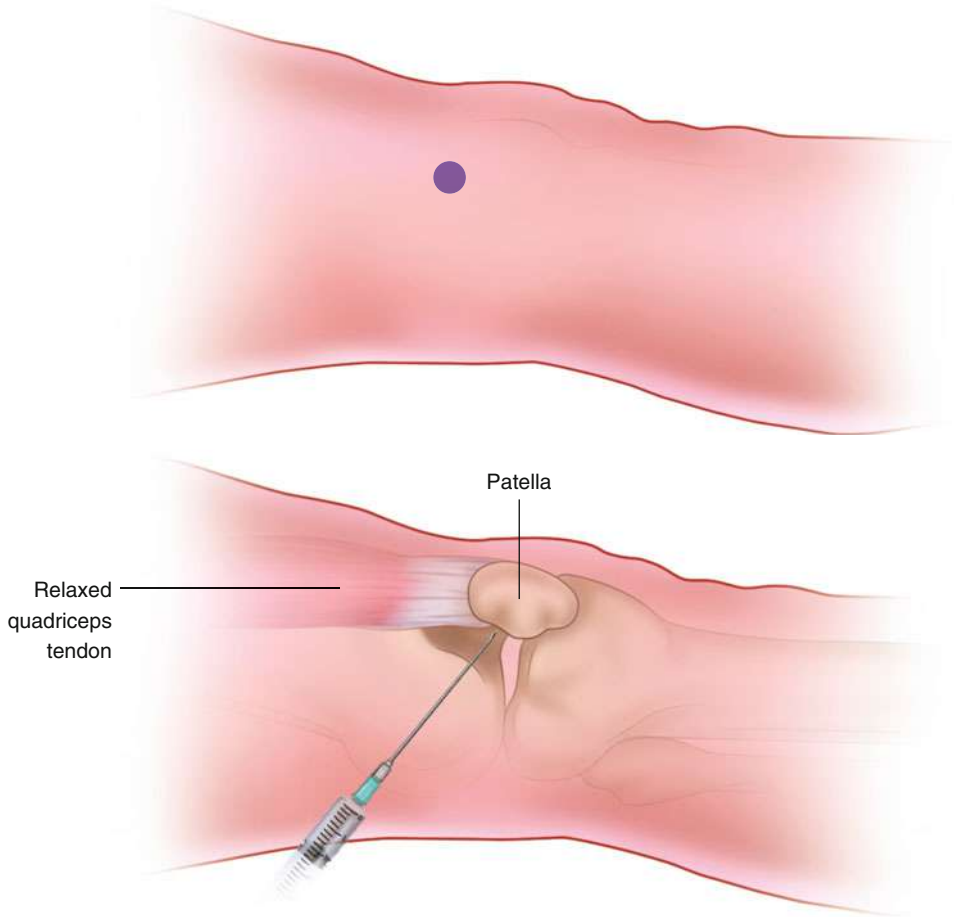

Fig. 107.1 Knee arthrocentesis, lateral approach

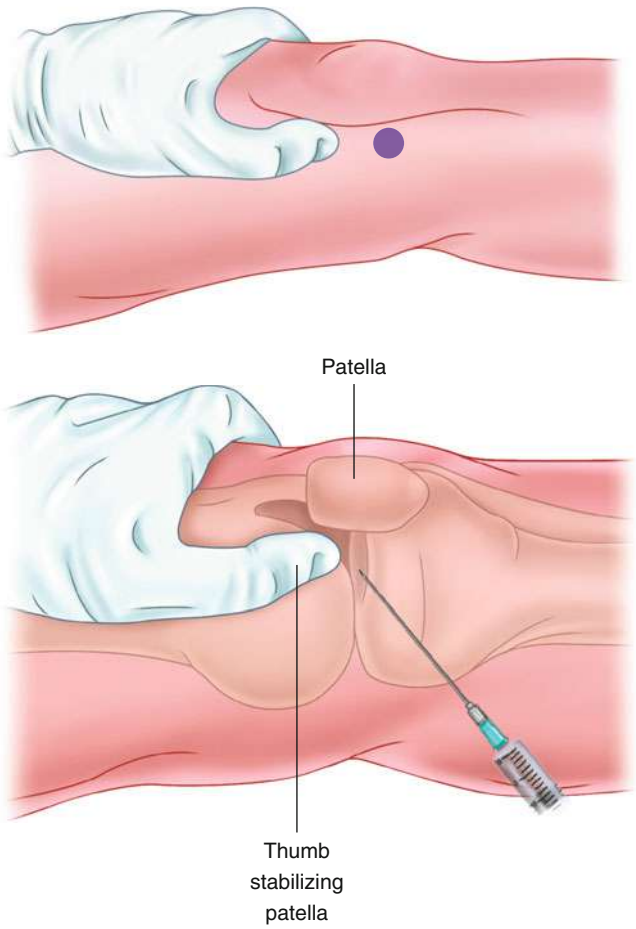

Fig. 107.2 Knee arthrocentesis, medial approach

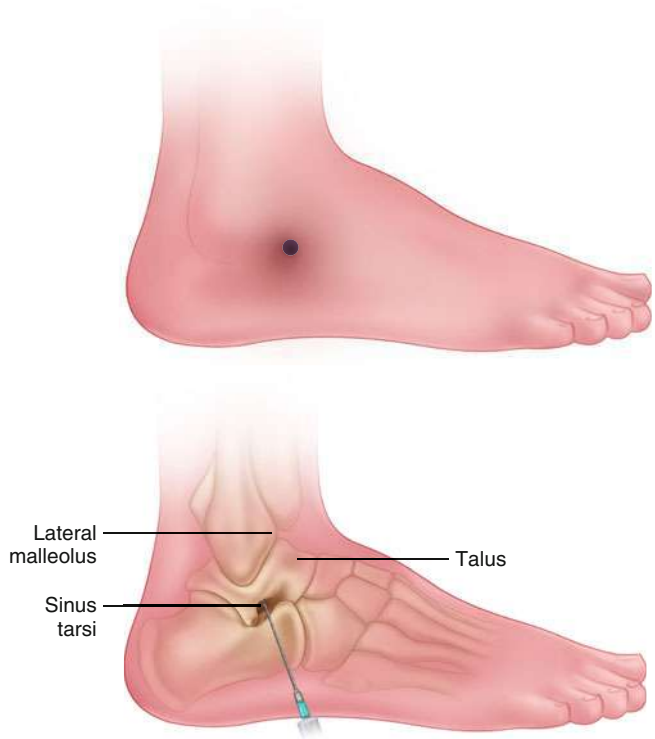

Fig. 107.3 Ankle arthrocentesis, lateral approach

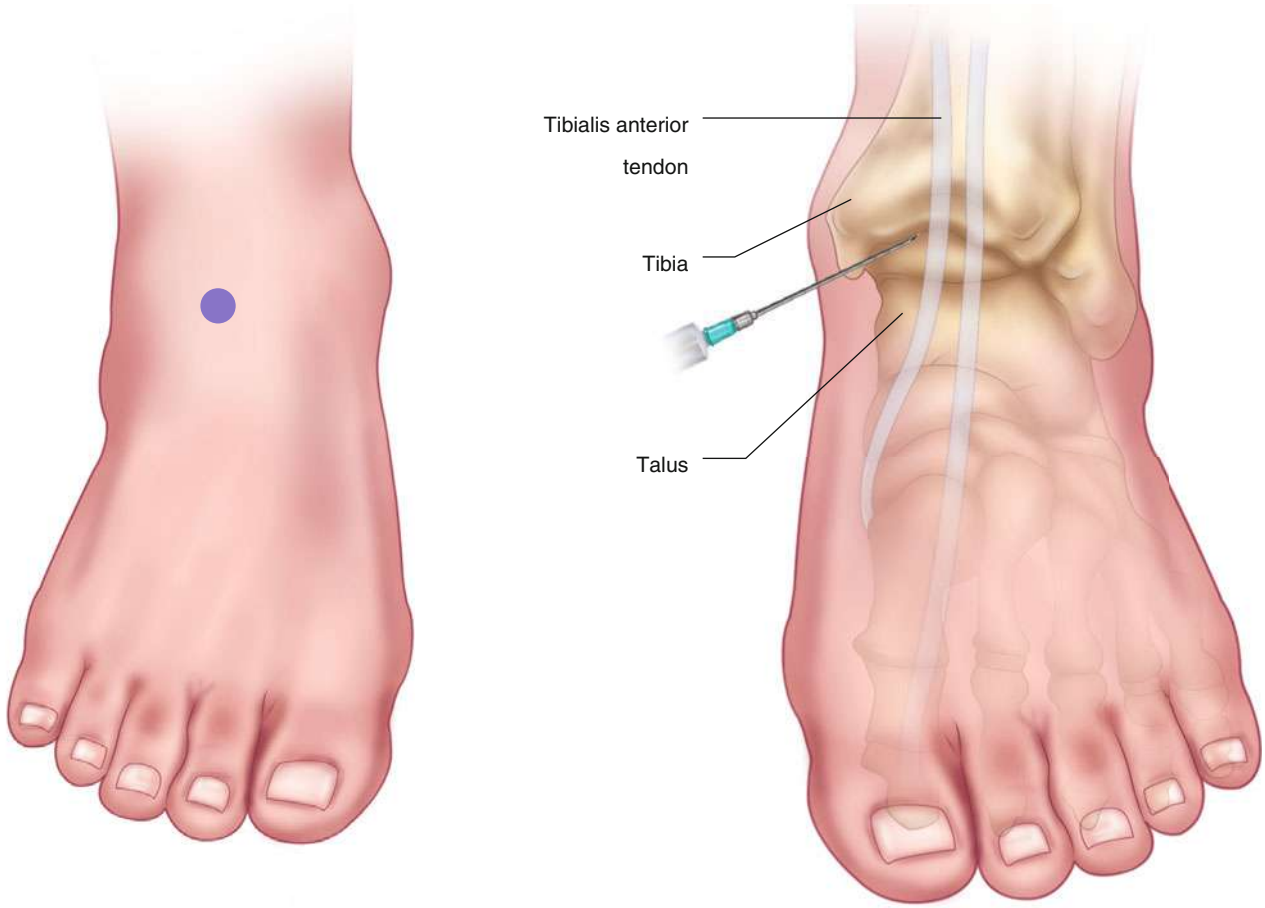

Fig. 107.4 Ankle arthrocentesis, medial approach

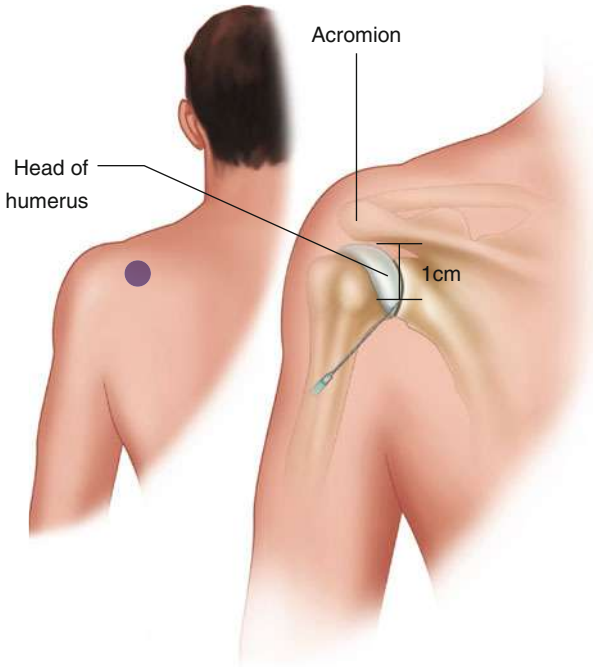

Fig. 107.5 Shoulder arthrocentesis, posterior approach

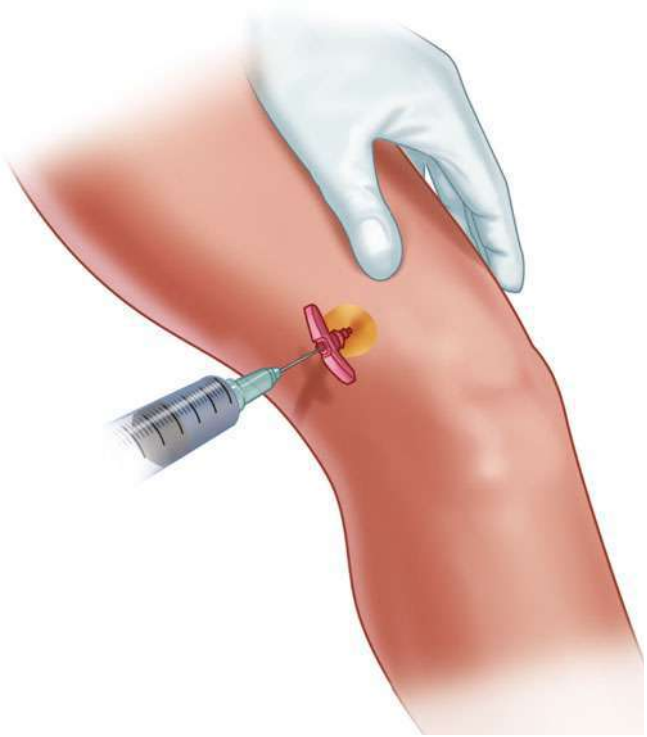

Fig. 107.6 Compress opposite side of joint to aid in aspiration

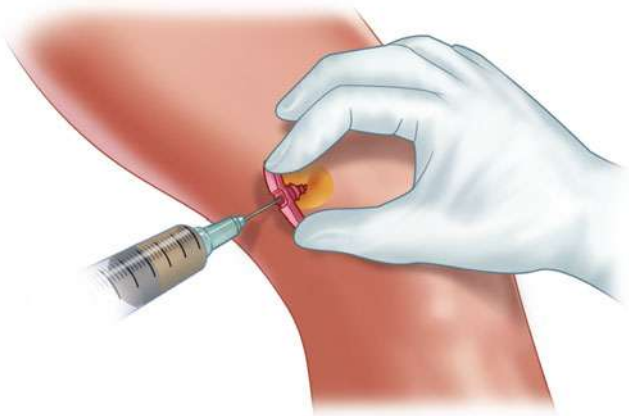

Fig. 107.7 Use hemostat or hand to stabilize needle within joint space

107.5 Complications

- Local postinjection flare
  - Irritation of the synovium by steroid microcrystals
  - Can be confused for infection
  - Occurs and resolves within 48 h after injection
  - Treat with ice and appropriate analgesics.
- Iatrogenic joint infection
  - Suspect if it begins later than, or lasts longer than, flare
  - Increasing pattern of pain
  - Fever, malaise, redness, or drainage around injection site
  - *Staphylococcus aureus* most common
- Subcutaneous atrophy and depigmentation
  - Leakage of corticosteroids into soft tissues
- Aspiration of blood
  - Indicative of trauma or bleeding disorder (hemophilia)
- Systemic absorption
  - Water-soluble preparations
  - Dose dependent
  - Injection into multiple joints
  - Transient hyperglycemia in diabetic patients
  - Avascular necrosis of the femoral head

107.6 Pearls

- Mixing lidocaine with glucocorticoids
  - Reduces pain caused by injection of steroids into joint space
  - Less likely to cause soft tissue atrophy and tendon rupture
  - Immediate relief from anesthetic indicates proper injection
- Limiting intra-articular glucocorticoid injections per joint
  - OA
    - Four injections per lifetime of the joint
    - Injections reduce the rate of accelerated degeneration in joints
  - RA
    - Limit of one injection per month
    - No evidence of glucocorticoid-induced cartilage loss

Selected Reading

Aponte EM, Schraga ED. Joint reduction, shoulder dislocation, anterior. Available at: http://emedicine.medscape.com/article/109130-overview#a08.

Cianflocco AJ. Intra-articular injections of the knee: a step-by-step guide. J Fam Pract. 2011;60(Suppl):S48–9. Available at: http://www.jfponline.com/pages.asp?id=10062.

Lavelle W, Lavelle ED, Lavelle L. Intra-articular injections. Anesthesiol Clin. 2007;25:835–62.

Molis MA, Young CC. Ankle impingement syndrome. Available at: http://emedicine.medscape.com/article/85311-overview.

Neustadt DH. Intra-articular injections for osteoarthritis of the knee. Cleve Clin J Med. 2006;73:897–911.

Roberts WN. Intraarticular and soft tissue injections: what agents(s) to inject and how frequently? Available at: www.uptodate.com.

Roberts WN. Joint aspiration or injection in adults: techniques and indications. Available at: www.uptodate.com.

108 Sugar-Tong Splint

Katrina John, Jeffrey Kile, and Amish Aghera

108.1 Indications

- Fractures to the wrist or forearm
- To prevent motion at the wrist and elbow
- To prevent supination and pronation

108.2 Contraindications

- Relative
  - Evidence of compartment syndrome or any neurovascular compromise

108.3 Materials (Fig. 108.1)

- Stockinette
- Padding
- Splint material: fiberglass/plaster of Paris or prefabricated splint rolls, 2, 3, 4 inches depending on age and body habitus
- Trauma shears/scissors
- Elastic bandage
- Tape
- Container with water
- Gloves, eyemask, sheet
- Sling

K. John, MBBS • J. Kile, MBBS, PhD, MPH (✉)
Department of Emergency Medicine, Eisenhower Medical Center, Rancho Mirage, CA, USA
e-mail: trenjohn@me.com; jeffrey.kile@gmail.com

A. Aghera, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: aaghera@maimonidesmed.org

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_108

Fig. 108.1 Equipment

108.4 Procedure

1. Ensure the skin of the affected extremity is clean, dry, and intact.
2. Cover the patient with a sheet or gown to protect the patient's clothing and the surrounding area.
3. Position the patient's arm abducted at 90° at the shoulder and internally rotated with the elbow flexed at 90° (Fig. 108.2).
4. Measure the splinting material by running a single layer from the metacarpal heads of the dorsum of the hand along the extensor surface of the forearm over the elbow and humeral condyles and back down the flexor surface of the forearm to the palmar aspect of the hand to the metacarpal heads.
5. If using non-prefabricated splint rolls, lay the measured piece of splinting material out on a flat surface and multiply the layers to the same length, 6–8 layers for fiberglass and 10–12 layers for plaster.
6. Measure the stockinette from the finger tips to the midhumerus and cut a hole for the thumb.
7. Place the stockinette on the arm (Fig. 108.3).
8. Use a 3- to 4-inch padding roll to apply several layers of circumferential padding extending from the metacarpal heads to the midhumerus below the level of the stockinette (Fig. 108.4).
9. Wet the already prepared and measured splinting material and remove the excess water.
10. Ensure the forearm is in the aforementioned position, and apply the splinting material from the metacarpal heads of the dorsum of the hand along the extensor surface of the forearm over the elbow and humeral condyles and back down the flexor surface of the forearm to the palmar aspect of the hand to the metacarpal heads (Fig. 108.5).
11. Fold each end of the stockinette down over the padding and splinting material.
12. An extra layer of padding can be added at this stage.
13. Secure the entire splint with two elastic bandages/ace wrap and apply tape to ensure the bandages stay in place (Fig. 108.6).
14. Place the arm in a sling.

Fig. 108.2 Arm positioning

Fig. 108.3 Fiberglass being measured and stockinette on patient

Fig. 108.4 (**a**) Application of padding over the stockinette (partial), (**b**) Application of padding over the stockinette (complete)

Fig. 108.5 (**a**) Application of splinting material over padding (anteromedial view), (**b**) Application of splinting material over padding (anterolateral view)

Fig. 108.6 Application of elastic bandage over splinting material

108.5 Complications

- Compartment syndrome

108.6 Pearls and Pitfalls

- Pearls
  - This procedure is best done with an assistant to hold the extremity in the desired position and to prevent the splint slipping as it is secured.
  - Having the patient in the illustrated position enables the practitioner to use gravity to hold the splint in the correct position while securing it; especially helpful if an assistant is not available.
- Pitfalls
  - If the splint is too short, it fails to immobilize the wrist.
  - If the splint is too long, it will cause reduced motion and stiffness at the metacarpophalangeal joints and swelling of the fingers due to immobility.

Selected Reading

Bong MR, Egol KA, Leibman M, Koval KJ. A comparison of immediate postreduction splinting constructs for controlling initial displacement of fractures of the distal radius: a prospective randomized study of long-arm versus short-arm splinting. J Hand Surg Am. 2006;31:766–70.

Denes AE, Goding R, Tamborlane J, Schwartz E. Maintenance of reduction of pediatric distal radius fractures with a sugar-tong splint. Am J Orthop (Belle Mead NJ). 2007;36:68–70.

Gartland JJ. The sugar tong splint. Am J Orthop. 1963;5:131.

McGeorge DD, Stilwell JH. The sugar tong splint: a reliable method of arm splintage in the child. J Hand Surg Br. 1989;14:357.

Simon RR, Koenigsknecht SJ, editors. Emergency orthopedics—the extremities. 3rd ed. New York: McGraw Hill; 1995.

Part XIV

Obstetric Procedures

109 Fetal Heart Rate Monitoring

Nathaniel Lisenbee and Joseph A. Tyndall

109.1 Indications

- Fetal heart rate (FHR) monitoring is important because it provides basic patterns that can be correlated to the acid–base status, circulatory volume, and oxygenation status of the fetus through brainstem detection and subsequent cardiac response. It has numerous indications during the antepartum and intrapartum stages [1].
- Antepartum indications include:
  - Nonstress test (consists of monitoring FHR in conjunction with fetal movements)
  - Contraction stress test (consists of monitoring FHR during contractions, which are induced pharmacologically)
  - Biophysical profile (BPP; consists of a nonstress test with an additional ultrasound)
- Intrapartum indications include monitoring FHR during:
  - Uterine contractions
  - Pain medications/anesthetic administration to the mother during labor
  - Procedures performed during labor
  - Second stage of labor
  - High-risk pregnancies, which can be defined by a number of conditions including [2, 3]:
    - Maternal diabetes, asthma, preeclampsia/eclampsia
    - Multiple gestations
    - Intrauterine growth restriction
    - Premature rupture of membranes
    - Lack of prenatal care

N. Lisenbee, MD
U. S. Air Force Medical Center Keesler,
Biloxi, MI, USA
e-mail: nate4085@ufl.edu

J.A. Tyndall, MD, MPH (✉)
Department of Emergency Medicine, University of Florida Health,
Gainesville, FL, USA
e-mail: tyndall@ufl.edu

109.2 Contraindications

- Contraindications for internal FHR monitoring
  - Presence of placenta previa
  - Lack of ability to identify the portion of the fetal body where device application is being considered
  - Active herpes, active hepatitis, or human immunodeficiency virus (HIV) in the mother
- Contraindications for external FHR monitoring
  - None

109.3 Methods

- Two methods for FHR monitoring:
  - Auscultation monitoring
    - Defined as auscultating FHR every 15 min in the first stage of labor and auscultating every 5 min in the second stage
    - Does not provide strips with information on FHR variability or the shape of FHR accelerations and decelerations
- Electronic FHR monitoring
  - Allows for real-time continuous monitoring of FHR activity
  - Provides strips with information on FHR variability or the shape of FHR accelerations and decelerations
  - Can be performed by Doppler ultrasound or internal fetal electrocardiography (ECG)

When comparing the two methods for electronic FHR monitoring, both are equally as reliable in most settings. Thus, external monitoring is the preferred method because it is noninvasive. However, in instances in which external monitoring becomes difficult owing to poor quality or technical difficulties, invasive monitoring is indicated.

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_109

109.4 Equipment and Procedures

Multiple methods exist for electronic FHR monitoring [4]. The most commonly used are external monitoring by Doppler ultrasound and internal monitoring by fetal ECG.

109.4.1 Doppler Ultrasound is a Noninvasive Method to Monitor FHR (Fig. 109.1)

- Equipment
  - Electronic FHR monitor
  - Contraction monitor sensor with belt
  - FHR sensor with belt (consists of ultrasound transducer and ultrasound sensor)
  - Ultrasound coupling gel
- Procedure
  1. Place the patient in a supine position.
  2. Palpate the fetal anatomy through the maternal abdomen to find the approximate location of the fetal heart.
  3. Place ultrasound coupling gel on the maternal abdomen at the sight of suspected fetal cardiac activity.
  4. Place the transducer probe on gel and locate the fetal heart tones.
  5. Once the fetal heart tones are located, secure the FHR sensor to the maternal abdomen with the attached belt.
  6. Place the contraction monitor sensor near the fundus in order to monitor uterine contractions.
  7. Attach the FHR sensor and contraction monitor to the electronic FHR monitor to obtain printouts of FHR and uterine contractions.

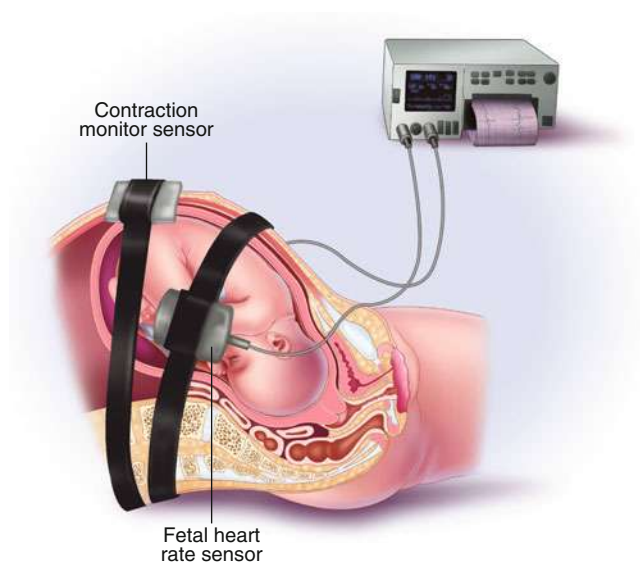

Fig. 109.1 External fetal heart rate monitoring

109.4.2 Internal Fetal ECG is an Invasive Method to Monitor FHR and is Used Only in the Intrapartum Period (Fig. 109.2)

- Equipment
  - Fetal scalp monitoring electrode
  - Leg plate electrode
  - Sterile vaginal lubricant
  - Electronic FHR monitor
- Procedure (Fig. 109.3)
  1. Place the patient in a dorsal lithotomy position.
  2. Sterilize the perineal area.
  3. Perform a bimanual vaginal examination to identify the presenting fetal head. (Note: rupture of membranes must occur before scalp electrode placement.)
  4. Place the spiral electrode guide tube on the fetal scalp and advance the electrode until it contacts the scalp.
  5. Rotate the drive tube clockwise approximately one rotation while maintaining pressure on the guide tube and drive tube.
  6. Release the electrode locking device by pressing together the arms on the drive tube grip.
  7. Carefully slide the drive and guide tubes off the electrode wires while holding the locking device open.
  8. Attach the leg plate to the inner thigh of the mother as a means to eliminate electrical interference.
  9. Attach the spiral electrode wires to the color-coded leg plate, which is then connected to the electronic fetal monitor.
  10. Do not forget to sterilize the area of electrode placement after delivery is completed and the scalp electrode is removed.

When comparing the two methods, both are equally reliable in most settings. Thus, external monitoring is the preferred method because it is noninvasive. However, in instances in which external monitoring becomes difficult owing to poor quality or technical difficulties, invasive monitoring is indicated.

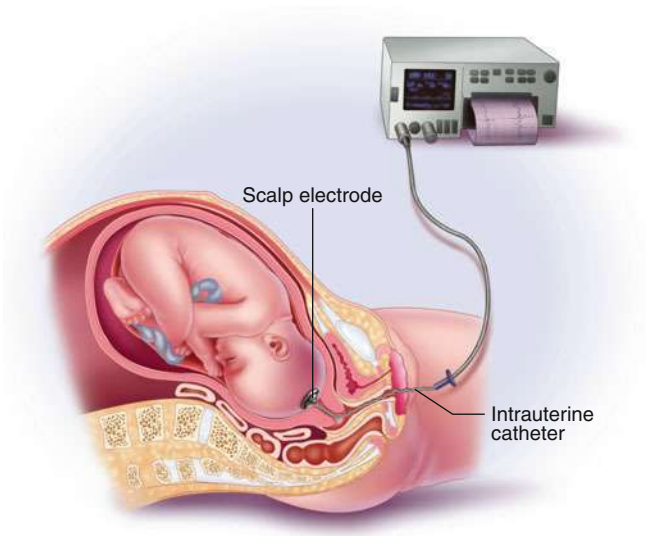

Fig. 109.2 Internal fetal heart rate monitoring

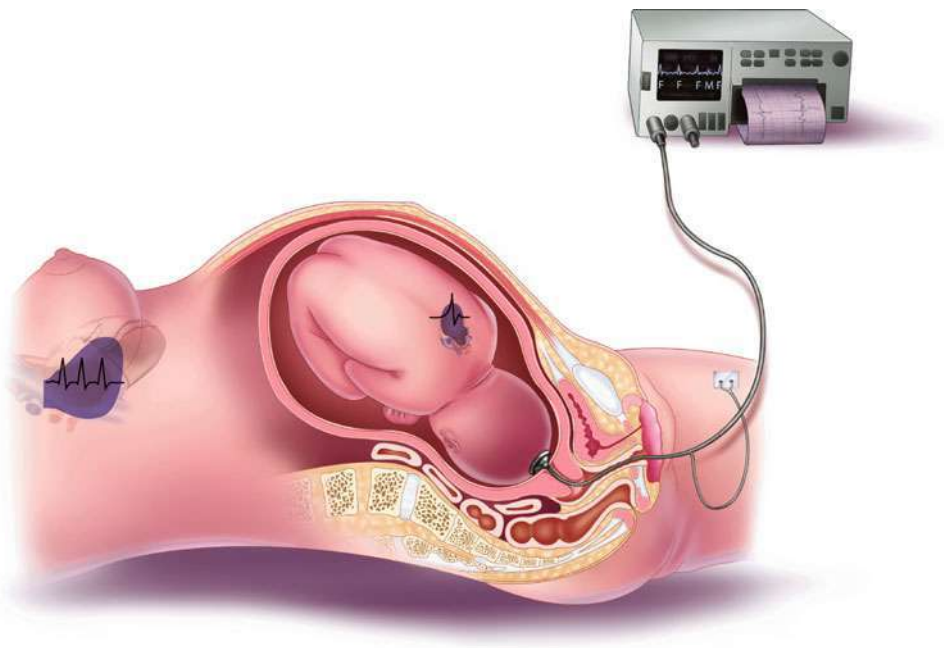

Fig. 109.3 Internal fetal heart rate monitoring

109.5 Complications

- Complications of external FHR monitoring
  - Confusing maternal aortic pulsations with FHR
  - Inability to locate FHR
- Complications of internal FHR monitoring
  - Fetal or maternal hemorrhage, fetal infection (usually scalp abscess at the site of insertion)
  - Uterine perforation
  - Subsequent fetal infection due to the invasive nature of the procedure

References

1. Hobel CJ. Intrapartum clinical assessment of fetal distress. Am J Obstet Gynecol. 1971;110:336–42.
2. Byrd JE. Intrapartum electronic fetal heart rate monitoring (EFM) and amnioinfusion. In: Advanced life support in obstetrics course syllabus. Kansas City: American Academy of Family Physicians; 1996. p. 97–106.
3. Queenan JT, Hobbins JC, Spong CY. Protocols for high-risk pregnancies. New York: Wiley; 2010. Retrieved 16 Jan 2012, from http://lib.myilibrary.com?ID=268955.
4. External and internal heart rate monitoring of the fetus. New Haven: Yale Medical Group; 2012. Retrieved from http://www.yalemedicalgroup.org/stw/.

Selected Reading

Alfirevic Z, Devane D, Gyte GM. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane Database Syst Rev. 2006;(3):CD006066.

American College of Obstetricians and Gynecologists. Fetal heart rate patterns: monitoring, interpretation, and management, ACOG technical bulletin, vol. 207. Washington, DC: ACOG; 1995.

Cunningham FG, Leveno KJ, Bloom SL, Hauth JC, Rouse DJ, Spong CY. Chapter 18: Intrapartum assessment. In: Williams obstetrics. 23rd ed. New York: McGraw Hill; 2010.

Freeman RK. Problems with intrapartum fetal heart rate monitoring interpretation and patient management. Obstet Gynecol. 2002;100:813.

Gomella LG, Haist SA. Chapter 13: Bedside procedures. In: Clinician's pocket reference. 11th ed. Columbus: McGraw-Hill; 2007.

Sweha A, Hacker TW, Nuovo J. Interpretation of the electronic fetal heart rate during labor. Am Fam Physician. 1999;59:2487–500.

Ultrasonography for Ectopic Pregnancy

110

L. Connor Nickels

An ectopic pregnancy is a pregnancy occurring outside of the uterine cavity (fundus) (Fig. 110.1).

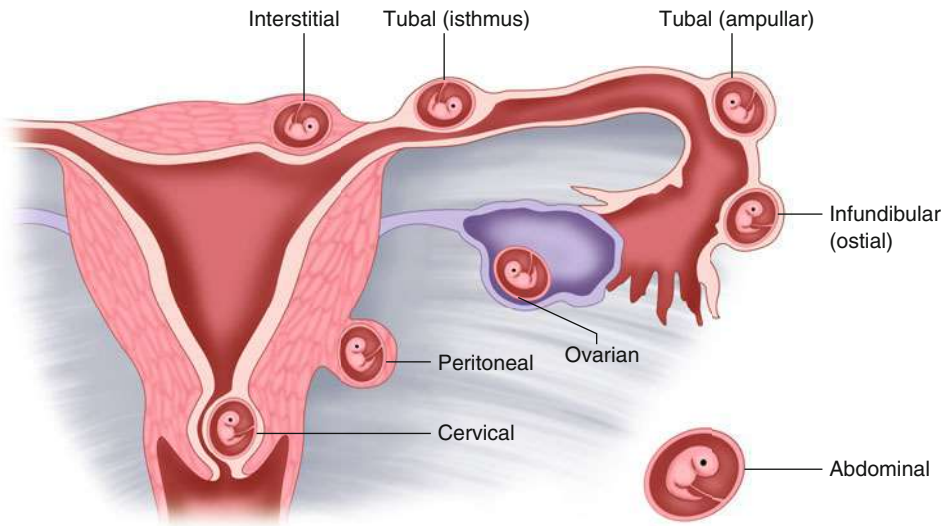

Fig. 110.1 Ectopic pregnancy diagram: pregnancy occurring outside of the uterine cavity

L.C. Nickels, MD, RDMS
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: cnickels@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_110

110.1 Indications

- Patient in the first trimester of pregnancy with any combination of the following:
  - Vaginal bleeding.
  - Acute pelvic pain.
  - Hypotension or shock.
  - Dizziness or syncope.
  - Positive pregnancy test.
  - Adnexal mass.
  - Cervical tenderness.
  - Abnormal rise in the serum human chorionic gonadotropin (hCG).
  - No specific sign or symptom is absolute; therefore, the index of suspicion must be high.
- Risk factors for ectopic pregnancy
  - Pelvic inflammatory disease (PID)
  - Previous or current intrauterine device
  - Infertility treatment
  - Previous tubal surgery
  - Tubal ligation
  - Advanced maternal age
  - Previous ectopic pregnancy
- Usually a sonographic diagnosis
- Sonographic signs of an ectopic can include any of the following:
  - Gestational sac seen in one of the following:
    - Adnexa with any of the following:
      - Yolk sac (Figs. 110.2 and 110.3)
      - Fetal pole with or without cardiac activity
      - Both of these
    - Low position in the cervix (cervical ectopic)
    - Seemingly in the uterus, but off to one side and with minimal surrounding myometrium (interstitial ectopic)
    - Within the peritoneal cavity, outside of the tubes (abdominal ectopic)
  - Pseudogestational sac seen in the uterus
    - Uterine enlargement or decidual reaction (single outline only) in the endometrium without a gestational sac
  - Other unidentifiable adnexa mass
  - Free fluid in the pelvis or other gravity-dependent area (i.e., Morison's pouch in the right upper quadrant in a supine or reverse Trendelenburg patient) (Fig. 110.4)
    - Small free fluid: tracks less than one third of the posterior cul-de-sac
    - Moderate free fluid: tracks less than two third of the posterior cul-de-sac
    - Large free fluid: tracks greater than two third of the posterior cul-de-sac
    - Right upper quadrant free fluid: 100 % predictability for ectopic
  - Empty uterus with serum hCG >1000 mIU/mL
- Patient in the first trimester of pregnancy with a serum hCG at or above the discriminatory zone (i.e., 1000 mIU/mL) without a sonographically normal gestational sac visualized within the uterus has an ectopic pregnancy until proven otherwise and should have an obstetrics consult in the emergency department.
  - Discriminatory zone may differ depending on the reference, but typically serum hCG between 1000 and 2000 mIU/mL.
  - For this chapter, we have used serum hCG greater than 1000 mIU/mL [1].
- Yolk sac should be first sign of definitive intrauterine pregnancy for emergency physicians because the decidual reaction and gestational sac are not 100 % accurate.
- May be treated surgically or medically depending on findings.
  - Decision to be made by obstetrics consultant

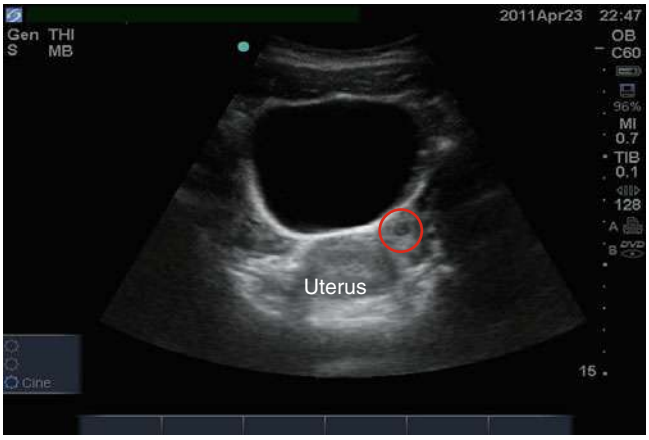

Fig. 110.2 Transabdominal transverse image of the uterus shows an ectopic pregnancy in the left adnexa (red circle) with the yolk sac present (Photo courtesy of L. Connor Nickels, MD, RDMS)

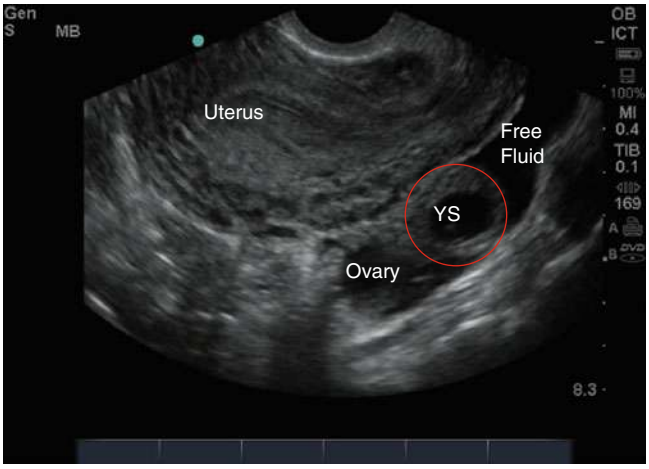

Fig. 110.3 Transvaginal sagittal image of an empty uterus with an ectopic pregnancy (in red circle) noted just posterior in the adnexa, adjacent to the ovary. The gestational sac contains a yolk sac and there is free fluid surrounding the ectopic, concerning for rupture (Photo courtesy of L. Connor Nickels, MD, RDMS). YS–yolk sac

Fig. 110.4 Transvaginal sagittal image of an empty uterus and large free fluid in the posterior cul-de-sac in a patient with a presumed ectopic pregnancy (Photo courtesy of L. Connor Nickels, MD, RDMS)

110.2 Contraindications

- Treating ectopic pregnancy medically when there is a fetal pole with fetal cardiac activity present
- Failing to consult obstetrics/gynecology when the patient has a serum hCG above the discriminatory zone and no sonographic findings to diagnose intrauterine pregnancy

110.3 Materials and Medications

- Ultrasound machine
- Probes: transabdominal and transvaginal
- Gel
- Skilled ultrasound operator
- Endocavitary probe covers
- Pelvic setup (speculum and cultures)
- Cardiac monitor, two large-bore intravenous needles
- Laboratory work: serum quantitative hCG, hemoglobin, group and Rh, type and screen
  - May add additional laboratory tests depending on the stability and symptoms of the patient

110.4 Procedure

- Ultrasound machine in obstetrics preset
- Transabdominal
  1. Place the patient in the supine position.
  2. Ideally, the bladder will be full for a good acoustic window.
  3. Using a curvilinear probe, 3.5–5.0 MHz.
  4. Begin scanning the patient in a sagittal position to identify the uterus as it lies in position with the bladder. Scan through the uterus completely in this plane, looking for signs of intrauterine pregnancy.
  5. Imaging in the transverse plane should be done in the same fashion.
  6. Any signs of an intrauterine pregnancy should be clearly identified and measured.
     - Gestational sac diameter
     - Yolk sac diameter
     - Fetal pole with crown rump diameter
       - If present, fetal cardiac activity should be recorded by using M mode to obtain a tracing and measure the fetal heart rate.
  7. Although sometimes limited transabdominally, an attempt to identify the adnexa should be performed in both planes bilaterally.
  8. Any abnormalities identified should be noted.
     - Free fluid surrounding the uterus (anterior or posterior cul-de-sac) or ovaries
     - On pelvic imaging or FAST examination
     - Intrauterine contents not consistent with an intrauterine pregnancy and/or not clearly, centrally visualized within the fundus of the uterus
     - Masses or contained fluid collections outside the uterus
     - Yolk sac or fetal pole with or without fetal cardiac activity seen outside the uterus
  9. Transvaginal scanning should be performed if a definitive intrauterine pregnancy is not identified on transabdominal imaging.
- Transvaginal
  1. Place the patient in the lithotomy position.
  2. Bladder is preferably empty.
  3. Use a transvaginal probe, 5–7.5 MHz.
  4. Repeat same steps as transabdominal imaging previously.
- The procedure is the same for all pelvic ultrasounds including transabdominal and transvaginal imaging because this should be performed in a systematic fashion so as to not miss pertinent findings. Therefore, the findings may change, but the examination remains the same.
- Yolk sac should be first sign of definitive intrauterine pregnancy for emergency physicians because the decidual reaction and gestational sac are not 100 % accurate.

110.5 Complications

- Bleeding: internal and/or external
- Maternal death if ectopic ruptures
  - Nine percent of pregnancy-related deaths
  - Leading cause of maternal death in the first trimester
- Sterility if tube(s) are damaged or surgically removed

110.6 Pearls and Pitfalls

- Pearls
  - Following the algorithm despite the patient being asymptomatic can help avoid missed ectopic pregnancies.
  - Cervical ectopic pregnancy can be difficult to distinguish from a spontaneous abortion. If the patient is aborting, the ultrasound findings should change quickly and the patient should have vaginal bleeding.
  - Double decidual reaction versus pseudogestational sac can be a very subtle distinction and should not be made by emergency physicians; hence, the statement made earlier, requiring a yolk sac as the earliest definitive sign of an intrauterine pregnancy.

  - Fibroids, bicornuate uterus, and eccentrically located normal pregnancy can all appear similar to a cornual pregnancy.
  - The majority of times when the patient is pregnant, has an empty uterus on sonographic imaging, and has vaginal bleeding, the final diagnosis is still unknown because it could still be an early normal pregnancy or an ectopic pregnancy.
- Pitfalls
  - Failing to obtain a pregnancy test in all reproductive-age women who have not undergone a hysterectomy
  - Failing to identify subtle signs of ectopic pregnancies

Reference

1. Ma OJ, Mateer JR, Blaivas M, editors. Emergency ultrasound. 2nd ed. New York: McGraw-Hill Professional; 2007.

Selected Reading

Gabbe SG, Niebyl JR, Galan HL, et al., editors. Obstetrics: normal and problem pregnancies. 5th ed. Philadelphia: Churchill Livingstone; 2007.

Marx J, Hockberger R, Walls R, editors. Rosen's emergency medicine: concepts and clinical practice. 7th ed. Philadelphia: Mosby; 2010.

Sanders RC, Winter T, editors. Clinical sonography: a practical guide. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2007.

Stead LG, Behera SR. Ectopic pregnancy. J Emerg Med. 2007;32(2):205–6.

Ultrasonography for Hydatidiform Mole

111

L. Connor Nickels

Definition: molar pregnancy = hydatidiform mole = anomalous growth of trophoblastic tissue

- Complete: 46,XX or 46,XY
  - Completely paternal in origin
  - Contains no fetal tissue
  - Most recognizable by clinical symptoms
- Incomplete (partial mole): 69,XXX or 69,XXY
  - Maternal and paternal in origin
  - Contains fetal tissue
  - More subtle clinical presentation

111.1 Indications

- Clinical presentation of any combination of the following:
  - Vaginal bleeding
  - With or without vomiting, persistent hyperemesis gravidarum
  - High blood pressure
  - Uterine size large for dates
- Serum human chorionic gonadotropin (hCG) very elevated (more than would be consistent for dates)
  - Usually greater than 100,000 mIU/mL

111.2 Contraindications

- Conservative management

L.C. Nickels, MD, RDMS
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: cnickels@ufl.edu

111.3 Materials and Medications

- Ultrasound machine
- Probes: curvilinear (abdominal) and endocavitary (transvaginal)
- Gel
- Skilled ultrasound operator
- Endocavitary probe covers
- Pelvic setup (speculum and cultures)
- Lab work: serum quantitative hCG, hemoglobin, group and Rh
  - May add additional laboratory tests depending on the stability and symptoms of the patient

111.4 Procedure

- Ultrasound machine in obstetrics preset
- Transabdominal imaging
  1. Place the patient in the supine position.
  2. Ideally the bladder will be full for a good acoustic window.
  3. Using a curvilinear probe, 3.5–5.0 MHz, begin scanning the patient in a sagittal position to identify the uterus as it lies in position with the bladder. Scan through the uterus completely in this plane, looking for signs of intrauterine pregnancy.
  4. Rotate the probe to a transverse position (counterclockwise), with the probe indicator toward the patient's right. Scan through the uterus completely in this plane, looking for signs of intrauterine pregnancy.
  5. Any signs of an intrauterine pregnancy should be clearly identified and measured.
  6. Sonographic findings of a molar pregnancy include:
     - Uterus filled with heterogeneous material.
       - Echogenic material interspersed with anechoic areas known as the "snowstorm" is the most common appearance of a mole (Fig. 111.1).

© Springer Science+Business Media New York 2016

L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_111

  - Can be confused with missed abortion or fibroid, underscoring the importance of the serum hCG because this will be markedly elevated, which would make the alternative diagnoses less likely.
- Theca lutein cysts are present in the adnexa.
  - Multiple cysts that occur with trophoblastic disease, multiple pregnancies (e.g., twins, triplets, quadruplets), and induced ovulation.

7. Although sometimes limited transabdominally, an attempt to identify the adnexa should be performed in both planes bilaterally.
8. Any abnormalities should be noted.
9. Transvaginal scanning should be performed if a definitive intrauterine pregnancy is not identified on transabdominal imaging.

- Transvaginal imaging
  1. Place the patient in the lithotomy position.
  2. Bladder is preferably empty.
  3. Use an endocavitary probe, 5–7.5 MHz.
  4. Repeat the same steps as for transabdominal imaging previously discussed (#2).
  5. After scanning through the uterus in sagittal and transverse planes, the adnexa should be scanned in the same format bilaterally, looking for any abnormalities, as mentioned previously.
  6. Again, any abnormalities should be noted:
     - Free fluid surrounding the uterus (anterior or posterior cul-de-sac) and/or ovaries
     - Masses or contained fluid collections outside the uterus
     - Yolk sac or fetal pole with or without fetal cardiac activity seen outside the uterus
- The procedure is the same for all pelvic ultrasounds, transabdominal and transvaginal imaging, because this should be performed in a systematic fashion so as to not miss pertinent findings. Therefore, the findings may change, but the examination remains the same.
- Yolk sac should be the first sign of definitive intrauterine pregnancy for emergency physicians because the decidual reaction and gestational sac are not 100 % accurate.

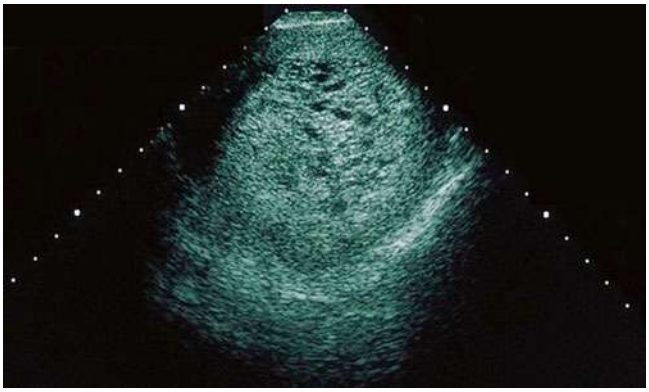

Fig. 111.1 Uterus with "cluster of grapes" represents a molar pregnancy (With kind permission from Springer Science + Business Media: Swisher E, Greer B, Montz FJ, Stenchever M. Chapter 14. In: *Atlas of Clinical Gynecology*. Vol. 4. 2002)

111.5 Complications

- Multiple complications can occur and include the following:
  - Invasive mole: when a hydatidiform mole recurs after a dilation and curettage and, subsequently, invades the muscle of the uterus.
    - Molar tissue is extremely vascular and necessitates an ultrasound with color flow to visualize any residual invasive tissue.
  - Choriocarcinoma: molar tissue develops into an aggressive malignancy, metastasizing early throughout the body.
    - In this scenario, the tumor takes on a very cystic appearance with an echogenic rim. When this is seen, one should attempt to view the liver as well for heterogeneous appearance consistent with metastasis or consider other imaging tests (more comprehensive ultrasound, computed tomography, or magnetic resonance imaging) as indicated to further assess concerns for malignancy.
    - Extremely sensitive to chemotherapy, but most favorable results are seen when diagnosed and treated early.

111.6 Pearls and Pitfalls

- Pearls
  - An obstetrics consult must be obtained in the emergency department when this diagnosis is made.
    - Patient must have a dilation and curettage to evacuate the mole.
    - Serum hCG must be followed to less than detectable levels because they have a high rate of recurrence (invasive mole) and potential for malignancy (choriocarcinoma).
  - May occur with intrauterine or ectopic pregnancies or after spontaneous abortions or full-term pregnancies
  - Qualitative and quantitative β-hCG results may be falsely negative in the setting of a molar pregnancy secondary to the "high-dose hook effect" found in sandwich immunoassays in which there is an overabundance of antigen. This error can be corrected by diluting the urine or serum sample and repeating the test.
- Pitfalls
  - Many pitfalls with potentially life-threatening consequences surround this diagnosis. Meticulous observation of the entire picture, including clinical presentation, laboratory results, and ultrasound findings, will help to avoid these outcomes.

Selected Reading

Gabbe SG, Niebyl JR, Galan HL, et al., editors. Obstetrics: normal and problem pregnancies. 5th ed. New York: Churchill Livingstone; 2007.

Hunter CL, Ladde J. Molar pregnancy with false negative â-hCG urine in the emergency department. West J Emerg Med. 2011;12:213–5.

Lentz GM, Lobo RA, Gershenson DM, et al., editors. Comprehensive gynecology. 6th ed. Philadelphia: Mosby; 2012.

Ma OJ, Mateer JR, Blaivas M. Emergency ultrasound. 2nd ed. New York: McGraw Hill Professional; 2008.

Sanders RC, Winter T, editors. Clinical sonography: a practical guide. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2007.

Ultrasonography for Blighted Ovum (Anembryonic Gestation)

112

Katrina Skoog Nguyen and L. Connor Nickels

112.1 Indications

- A blighted ovum or anembryonic gestation is a pregnancy in which the embryo never develops in the gestational sac (Fig. 112.1).
- Presents as first-trimester vaginal bleeding and/or pelvic pain or cramping
- Criteria for defining an anembryonic pregnancy:
  - Based, in part, on mean gestational diameter (MGD), which is an average of three orthogonal measurements of the gestational sac
    - MGD greater than or equal to 25mm with no embryo
    - 5 MHz or less
      - MGD greater than or equal to 20 mm without a yolk sac
  - May have abnormally low sac position
- References differ in their measurements depending on the approach, transabdominal versus transvaginal, and the frequency of the probe (e.g., 5 MHz vs. 6.5 MHz or greater).

Fig. 112.1 Low-lying empty gestational sac consistent with blighted ovum (Photo courtesy of L. Connor Nickels, MD, RDMS)

K.S. Nguyen, DO • L.C. Nickels, MD, RDMS (✉)
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: katrinaskoog@hotmail.com; cnickels@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_112

112.2 Contraindications

- Misinterpreting images as blighted ovum when any of the following could potentially be present:
  - Early intrauterine pregnancy
  - Pseudogestational sac
  - Molar pregnancy
- Discussions with the patient should always involve an obstetrics consult or referral for definitive decisions and treatment.

112.3 Materials and Medications

- Ultrasound machine
- Probes: transabdominal and transvaginal
- Ultrasound gel
- Endocavitary probe covers
- Skilled ultrasound operator
- Laboratory tests: serum quantitative human chorionic gonadotropin (hCG), hemoglobin, group, and Rh
  - May add additional laboratory tests depending on the stability and symptoms of the patient

112.4 Procedure

- Ultrasound machine in obstetrics preset
- Transabdominal
  1. Place the patient in the supine position.
  2. Ideally the bladder will be full for a good acoustic window.
  3. Use a curvilinear probe, 3.5–5.0 MHz.
  4. Begin scanning the patient in a sagittal position to identify the uterus as it lies in position with the bladder. Scan through the uterus completely in this plane, looking for signs of intrauterine pregnancy.
  5. Imaging in the transverse plane should be done in the same fashion.
  6. Any signs of an intrauterine pregnancy should be clearly identified and measured.
     - In this case, gestational sac and any other intrauterine contents
  7. Although sometimes limited transabdominally, an attempt to identify the adnexa should be performed in both planes bilaterally.
  8. Any abnormalities should be noted.
     - Free fluid surrounding the uterus (anterior or posterior cul-de-sac) or ovaries
     - Masses or contained fluid collections outside the uterus
     - Yolk sac or fetal pole with or without fetal cardiac activity seen outside the uterus
  9. Transvaginal scanning should be performed if a definitive intrauterine pregnancy is not identified on transabdominal imaging.
- Transvaginal
  1. Place the patient in the lithotomy position.
  2. Bladder is preferably empty.
  3. Use a transvaginal probe, 5–7.5 MHz.
  4. Repeat the same steps as for transabdominal imaging previously discussed (#2).
- The procedure is the same for all pelvic ultrasounds including transabdominal and transvaginal imaging because this should be performed in a systematic fashion so as to not miss pertinent findings. Therefore, the findings may change, but the examination remains the same.
- Yolk sac should be first sign of definitive intrauterine pregnancy for emergency physicians because the decidual reaction and gestational sac are not 100 % accurate.

112.5 Complications

- Bleeding
- Retained products of conception

112.6 Pearls and Pitfalls

- Pearls
  - See "Contraindications." If following an algorithm, a blighted ovum would fall into indeterminate category and obstetrics consult is recommended in the emergency department.
- Pitfalls
  - It is imperative to inform all pregnant patients that obstetrical ultrasonography performed by emergency physicians is limited and not used to detect fetal health and/or anatomy.

Selected Reading

Cosby K, Kendall J, editors. Practical guide to emergency ultrasound. Philadelphia: Lippincott Williams & Wilkins; 2006.

Doubilet PM, Benson CB, Bourne T, Blaivas M. Diagnostic criteria for nonviable pregnancy early in the first trimester. N Engl J Med. 2013;369(15):1443–51.

Ma OJ, Mateer JR, Blaivas M, editors. Emergency ultrasound. 2nd ed. New York: McGraw Hill Professional; 2007.

Morin L, Van den Hof MC. Ultrasound evaluation of first trimester pregnancy complications. J Obstet Gynaecol Can. 2005;161:581–5.

Sanders RC, Winter T, editors. Clinical sonography: a practical guide. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2007.

Shah K, Mason C, editors. Essential emergency procedures. Philadelphia: Lippincott Williams & Wilkins; 2008.

Ultrasonography for Threatened, Incomplete, or Compete Abortion

113

L. Connor Nickels and Giuliano De Portu

Key Points

Threatened abortion: closed os, no passage of products of conception (POC).
Inevitable abortion: open os, no POC.
Incomplete abortion: open os, passage of POC.
Complete abortion: closed os, empty uterus (all POC passed).
Missed abortion: fetus has died and the uterus has failed to enlarge any further.

113.1 Indications

- Vaginal bleeding in the setting of early intrauterine pregnancy

113.2 Contraindications

- There are no contraindications to ultrasonography.

113.3 Materials and Medications

- Ultrasound machine
- Probes: curvilinear (transabdominal) and endocavitary (transvaginal)
- Gel
- Endocavitary probe covers
- Towels
- Skilled ultrasound operator
- Pelvic setup (speculum and cultures)
- Cardiac monitor and intravenous access if bleeding is significant
- Laboratory tests: serum human chorionic gonadotropin (hCG), hemoglobin, group and Rh, type, and screen

L.C. Nickels, MD, RDMS (✉) • G. De Portu, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; gdeportu@ufl.edu

113.4 Procedure

The procedure is the same for all pelvic ultrasounds including transabdominal and transvaginal imaging because this should be performed in a systematic fashion so as to not miss pertinent findings. Therefore, the findings may change, but the examination remains the same.

1. Place the patient in the supine position.
2. Ultrasound machine in obstetrics preset with abdominal probe and gel.
3. Begin scanning the patient in a transabdominal sagittal plane to identify the uterus as it lies in position with the bladder. Scan through the uterus completely in this plane looking for signs of intrauterine pregnancy (gestational sac, yolk sac, fetal pole, fetal cardiac activity).
   - For our purposes as emergency physicians, a yolk sac (at the least) must be seen within the uterus to definitively report an intrauterine pregnancy (Fig. 113.1). The decidual reaction and/or presence of gestational sac is not 100 % accurate.
4. Transabdominal imaging in the transverse plane should be done in the same fashion.
5. If an intrauterine pregnancy is clearly identified on transabdominal scanning, appropriate measurements can be taken to estimate dates:
   - Gestational sac diameter.
   - Yolk sac diameter or crown rump length (CRL).
   - If present, fetal cardiac activity should be recorded by using M mode to obtain a tracing and measure the fetal heart rate (Fig. 113.2).
6. For a complete examination, the adnexa should be scanned in an attempt to identify the ovaries bilaterally.
7. Any abnormalities identified should be noted:
   - Free fluid surrounding the uterus (anterior or posterior cul-de-sac) or ovaries
     - On pelvic imaging or focused assessment with sonography in trauma (FAST) examination

© Springer Science+Business Media New York 2016

L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_113

- Intrauterine contents not consistent with an intrauterine pregnancy and/or not clearly, centrally visualized within the fundus of the uterus
- Any masses or contained fluid collections noted outside the uterus
- Yolk sac or fetal pole with or without fetal cardiac activity seen outside the uterus

8. Transvaginal scanning with the endocavitary probe should be performed if a definitive intrauterine pregnancy is not identified on transabdominal imaging.
9. Transvaginal imaging is performed in the sagittal and transverse planes to identify an intrauterine pregnancy, with bladder preferably empty.
10. Measurements should be taken of any intrauterine findings.
    - Same as #5.
    - CRL >7 mm on transvaginal ultrasound should have fetal cardiac activity.
11. Each adnexa should be scanned in the sagittal and transverse planes.
12. Any abnormalities identified should be noted.
    - Same as #7.
13. The uterus will be empty in the case of a complete abortion (Figs. 113.3 and 113.4).

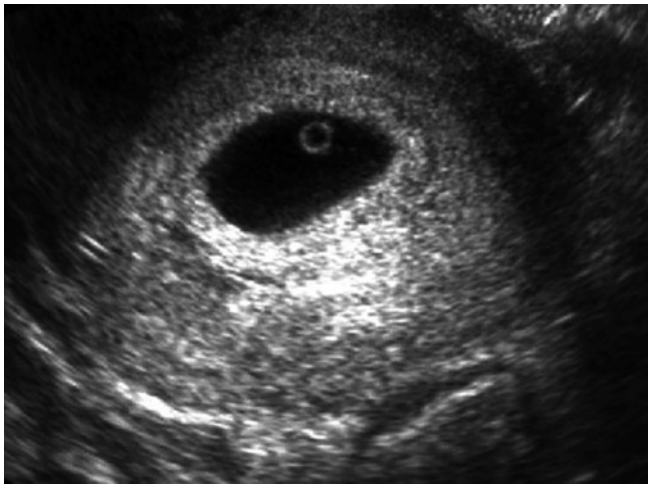

Fig. 113.1 Yolk sac present within the gestational sac with double decidual reaction (With kind permission from Springer Science+Business Media: Buja LM, Chandrasekhar C. Chapter 7: pathology of the breast and female genital tract. In: Krueger GRF, Buja LM, eds. *Atlas of Anatomic Pathology with Imaging*. 2013)

Fig. 113.3 Transvaginal sagittal image of an empty uterus (Photo courtesy of L. Connor Nickels, MD, RDMS)

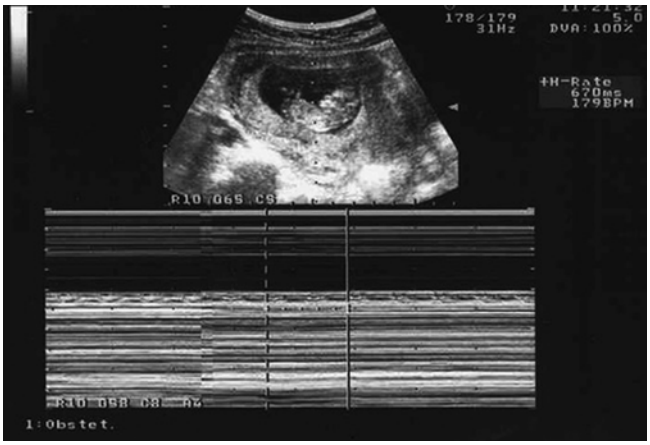

Fig. 113.2 Fetal cardiac activity recorded using M mode to obtain a tracing and measure the fetal heart rate (With kind permission from Springer Science+Business Media: Hanprasertpong T, Phupong V. First trimester embryonic/fetal heart rate in normal pregnant women. *Archives of Gynecology and Obstetrics*. 2006;274(5))

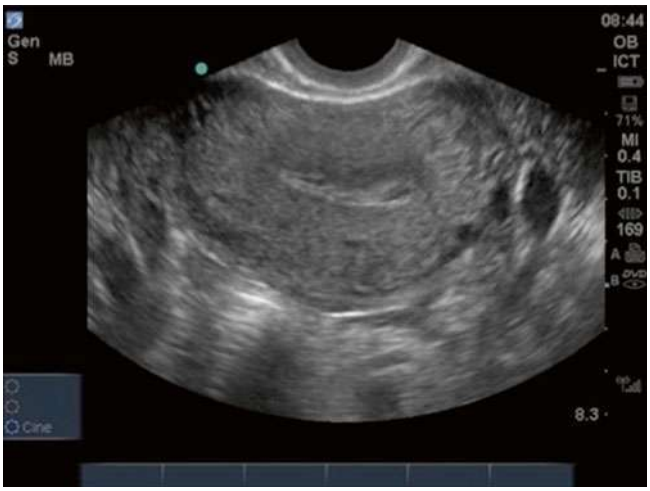

Fig. 113.4 Transvaginal transverse image of an empty uterus (Photo courtesy of L. Connor Nickels, MD, RDMS)

113.5 Complications

- Vaginal bleeding
- Miscarriage
- Rh isoimmunization

113.6 Pearls and Pitfalls

- Pearls
  - Threatened abortion:
    - Threatened abortion is not visible sonographically. It is the presumed diagnosis when a patient presents with vaginal bleeding in the first 20 weeks of pregnancy, an intrauterine pregnancy is found with dates corresponding to the patient's dates, and the cervix is closed.
    - Fetal cardiac activity suggests a better prognosis than those without.
    - Fetal heart rate less than 120 bpm suggests impending fetal death (only 6 % survival).
  - Completed abortion:
    - There are only three scenarios in which this diagnosis can confidently be made in the emergency department:
      1. Intact gestation is passed and identified in the emergency room.
      2. Ultrasound shows an empty uterus in the setting of a prior known intrauterine pregnancy.
         - Small internal echoes within the uterus may represent blood rather than retained products of conception, but this determination should be made in consultation with obstetrics.
      3. Negative pregnancy test result in the setting of a prior known intrauterine pregnancy.
    - In all other cases, the quantitative β-hCG must be followed until it is less than 2 mIU/mL.
- Pitfalls
  - Treating threatened abortion as fetal demise or abortion – many of these pregnancies actually go on to completion with a normal fetus.
  - If there is any concern for alternative diagnoses in any of the abortion types (i.e., ectopic gestation), obstetrics consult should be obtained.
  - Failing to order a quantitative serum hCG to rule out other possibilities (e.g., molar pregnancy) and to ensure that it is trending downward (if others to compare).
  - Failing to establish follow-up for an incomplete abortion to ensure appropriate management. The patient may eventually require a dilation and curettage if products do not pass naturally.

Selected Reading

Doubilet PM, Benson CB, Bourne T, Blaivas M. Diagnostic criteria for nonviable pregnancy early in the first trimester. N Engl J Med. 2013;369(15):1443–51.

Gabbe SG, Niebyl JR, Galan HL, et al., editors. Obstetrics: normal and problem pregnancies. 5th ed. Orlando: Churchill Livingstone; 2007.

Ma OJ, Mateer JR, Blaivas M, editors. Emergency ultrasound. 2nd ed. New York: McGraw-Hill Professional; 2007.

Marx J, Hockberger R, Walls R, editors. Rosen's emergency medicine. 7th ed. Philadelphia: Mosby; 2010.

Sanders RC, Winter T, editors. Clinical sonography: a practical guide. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2007.

Ultrasonography for Placenta Previa

114

L. Connor Nickels and Giuliano De Portu

Placenta previa is a condition in which the placenta covers the internal cervical os. It is the leading cause of antepartum hemorrhage. There are four grades of the condition:

- Complete: Placenta completely covers the internal os.
- Partial: Placenta partially covers the internal os.
- Marginal : Lower margin of placenta reaches the internal os but does not cover it (within 3 cm).
- Low lying: Lower margin of placenta is located in the lower uterine segment but does not reach the internal os (Fig. 114.1).

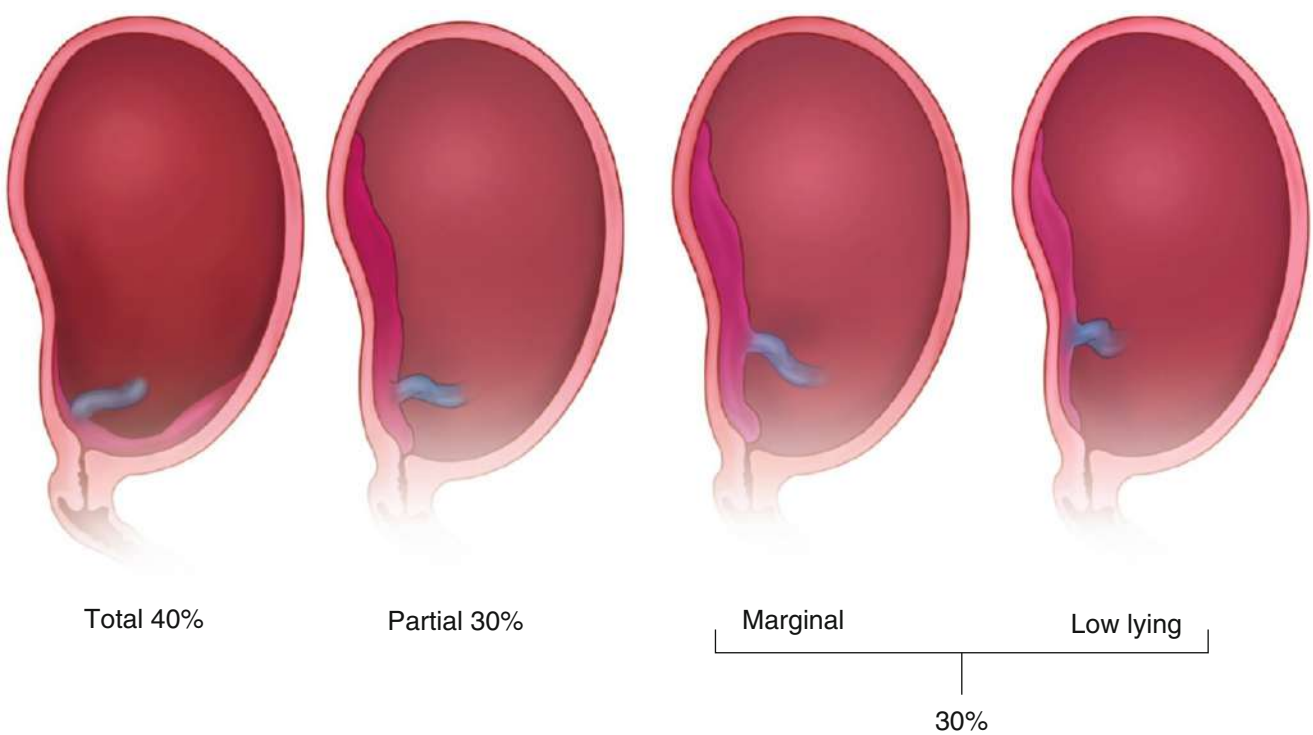

Fig. 114.1 Varying degrees of placenta previa

L.C. Nickels, MD, RDMS (✉) • G. De Portu, MD
Department of Emergency Medicine, University of Florida Health Shands Hospital, Gainesville, FL, USA
e-mail: cnickels@ufl.edu; gdeportu@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_114

114.1 Indications

- A patient in the second half of pregnancy who presents to the emergency room with bright red, painless vaginal bleeding OR premature labor should be evaluated for this condition.

114.2 Contraindications

- Absolute:
  - There are no absolute contraindications to transabdominal ultrasonography.
- Relative:
  - Transvaginal ultrasound should be carried out with caution, so as to avoid hemorrhage.

114.3 Materials and Medications

- Ultrasound machine
- Probes: curvilinear or phased array, endocavitary
- Sterile gel
- Endocavitary probe covers
- Towels
- Sterile speculum (see previously)
- Cardiac monitor, two large-bore intravenous needles, laboratory tests

114.4 Procedure

1. Place the patient in the supine position.
2. Ultrasound machine with abdominal probe and gel.
3. Begin scanning the patient in a transabdominal sagittal position (Fig. 114.2) to determine the placental position and whether it is lying in the lower uterine segment.
4. The patient should then be scanned in a transverse fashion to further evaluate the exact position of the placenta.
5. If the placenta appears to lie in the lower uterine segment, the patient should be scanned in the oblique plane as well.
6. The bladder should initially be full for best visualization. However, if the placenta appears to reside low or lie over the internal os, the scanning should be repeated after the patient has voided.
   - An overdistended bladder may create the appearance of a placenta previa. The anterior wall of the uterus is compressed against the posterior wall by the distended bladder, shortening the distance between the placenta and the internal os (Fig. 114.3).
7. The following ultrasonographic findings exclude placenta previa [1]:
   - Direct apposition of the presenting part of the fetus and the cervix without space for interposed tissue
   - Presence of amniotic fluid between the presenting part of the fetus and the cervix, without the presence of placental tissue
   - Distance of greater than 2 cm between the inferior aspect of the placenta
   - Indirect visualization of the internal cervical os
8. If placenta previa cannot be ruled out with transabdominal ultrasound, the patient should then be scanned transvaginally with the endocavitary probe because this is more sensitive for diagnosing placenta previa.
9. A sterile speculum examination should be performed before transvaginal scanning to assess the cervix and ensure there are no presenting parts or bulging membranes. Transvaginal imaging is contraindicated if the patient has ruptured or bulging membranes.
10. The probes should be swapped out and a sterile cover placed on the endocavitary probe with gel inside the cover and sterile gel on the outside of the cover.
11. The endocavitary probe should be inserted into the vaginal canal, ensuring that caution is taken to stay off the cervix and distal to it, keeping the cervix in view on the screen.
12. The patient should be scanned in both the sagittal and the transverse planes to assess the inferior margin of the placenta.

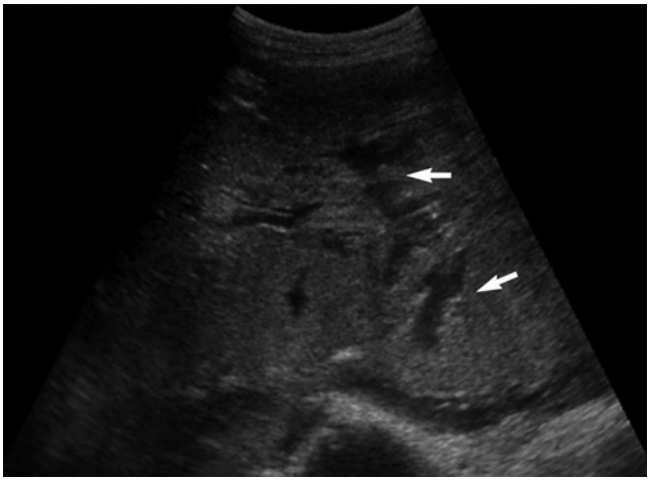

Fig. 114.2 Transabdominal ultrasound of placenta previa. Note loss of the decidual interface between the placenta and the myometrium on the lower part of the uterus and multiple intraplacental lacunae (*arrows*) (Reproduced with permission from *J Korean Med Sci*. 2010; 25(4):651–5)

13. If the inferior margin appears to be located near the internal os, then the distance should be measured.
    - Steps 8 through 13 should only be performed by the emergency physician if he or she feels confident in this ultrasound skill and there is no obstetrics available. If obstetrics is available, these steps should be performed in conjunction with them or by them to ensure the best outcome for the patient.

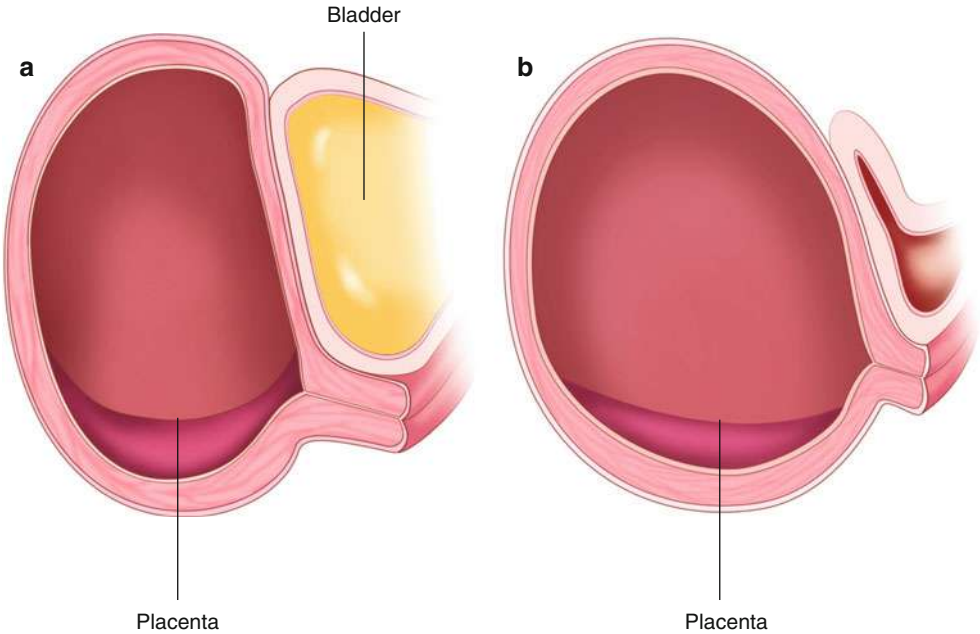

Fig. 114.3 (**a**) Overdistended bladder creating the appearance of placenta previa, (**b**) empty bladder showing the more accurate position of the placenta

114.5 Complications

- Bleeding:
  - May range from self-limited to life-threatening hemorrhage
- Maternal and/or fetal distress or death

114.6 Pearls and Pitfalls

- Pearls
  - Risk factors include:
    - Multiparity
    - Multiple pregnancy
    - Advanced maternal age
    - Previous placenta previa
    - Cigarette smoking
    - Cocaine abuse
    - Hypertension
    - Previous cesarean delivery/uterine surgeries
  - Myometrium contraction: can mimic placenta previa by temporarily displacing the placenta in the lower uterine segment.
  - If ultrasound capabilities are not available, but the patient is in the second half of pregnancy and having vaginal bleeding, do not perform a digital cervical examination.
- Pitfalls
  - Digital examination should be avoided because this may precipitate life-threatening hemorrhage and/or death.
  - Always consult obstetrics as soon as possible if this is suspected or known and the patient is symptomatic. Not consulting obstetrics could be detrimental to the mother and fetus.
  - Gentle sterile speculum examination should be done only if an obstetrician is not available. This is to ensure the bleeding is coming from the cervix. If placenta previa is suspected or known and obstetrics are available, then abdominal ultrasound evaluation alone should be sufficient for the examination of the patient.

Reference

1. American College of Radiology. Role of imaging in second and third trimester bleeding. In: ACR Appropriateness Criteri. Reston: American College of Radiology; 2001.

Selected Reading

Gabbe SG, Niebyl JR, Galan HL, et al., editors. Obstetrics: normal and problem pregnancies. 5th ed. Orlando: Churchill Livingstone; 2007.

Ma OJ, Mateer J, Blaivas M, editors. Emergency ultrasound. 2nd ed. New York: McGraw Hill Professional; 2007.

Marx J, Hockberger R, Walls R, editors. Rosen's emergency medicine: concepts and clinical practice. 7th ed. Philadelphia: Mosby; 2010.

Sanders RC, Winter T, editors. Clinical sonography: a practical guide. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2007.

Vaginal Delivery

115

Umarfarook Javed Mirza, Christopher Shields, and Muhammad Waseem

115.1 Indications

- Inevitable delivery of the fetus (cervix fully dilated at 10 cm and fully effaced 1 mm with crowning of the head)

115.2 Contraindications

- Absolute:
  - Indications for emergent cesarean section (C-section)
  - Prolapsed cord
  - Prior C-section with classic vertical incision
  - Placenta previa – complete/partial
  - Breech presentation – footling
- Relative:
  - Placenta previa – marginal/low lying
  - Breech presentation – complete/incomplete/frank

115.3 Materials and Medications (Fig. 115.1)

- 4 Crile clamps or Kelly clamps
- 1 Mayo scissor curved and 1 Mayo scissor straight
- 2 sponge forceps
- 2 towel clamps
- 1 Mayo-Hegar needle holder
- 1 mouse-tooth forceps
- Suction bulb
- Betadine (povidone-iodine)
- Umbilical cord clamp (Fig. 115.2)
- Incubator warmer (Fig. 115.3)
- Most importantly, help.

U.J. Mirza, DO
Department of Emergency Medicine,
Baylor University Medical Center, Dallas, TX, USA
e-mail: Umarfarook.Mirza@gmail.com

C. Shields, MD • M. Waseem, MD (✉)
Department of Emergency Medicine,
Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: cpatrickshields@hotmail.com; waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_115

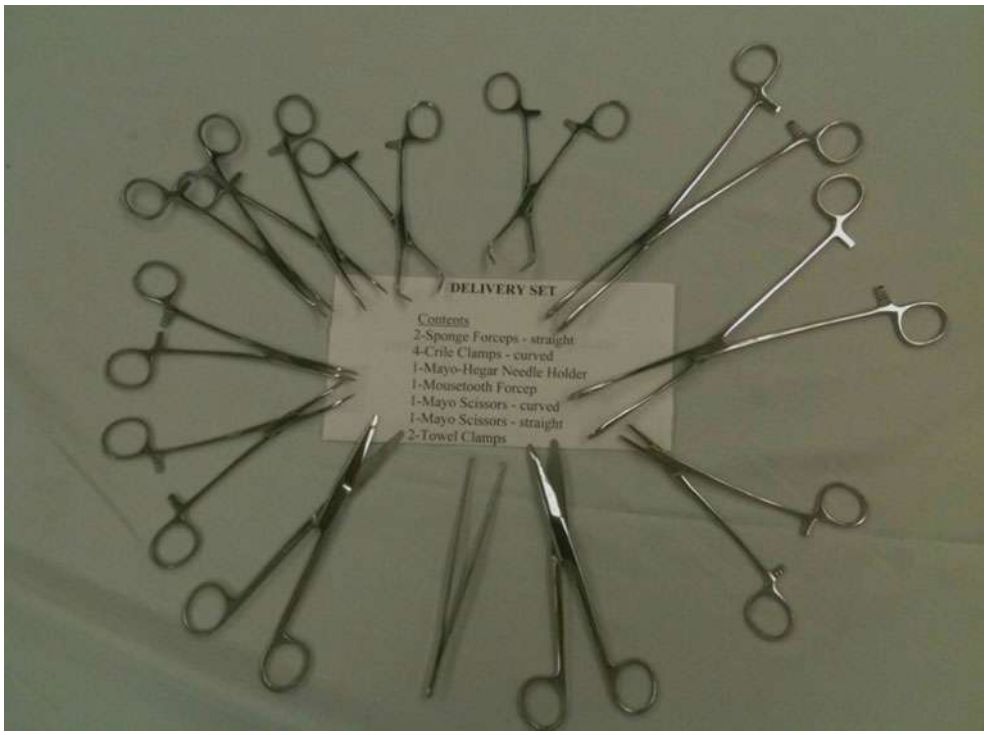

Fig. 115.1 Equipment: delivery set

Fig. 115.2 Equipment: umbilical clamp

Fig. 115.3 Equipment: incubator warmer

115.4 Procedure

1. Determine the fetal presentation (breech or cephalic) with Leopold's maneuvers or bedside ultrasound.
2. If breech, follow the breech pathway.
3. Perform a vaginal examination.
4. Check for cord prolapse.
5. Check effacement.
6. Check dilation.
7. Prepare the site.
8. Position the mother in the extreme lithotomy position.
9. Deliver the head in a controlled manner.
10. Check to see if the cord is wrapped around the neck. If present, follow cord presentation pathway.
11. Deliver the anterior shoulder with downward traction of the head.
12. Check for shoulder dystocia. If present, follow the McRobert-Rubin maneuvers (see Chap. 116).
13. Deliver the posterior shoulder by gently pulling the trunk upward.
14. Suction the airway.
15. Clamp the cord 5 cm from the umbilicus in two places and cut.
16. Obtain Apgar score of the baby. Initiate resuscitation if required.
17. Keep the baby warm in the incubator.
18. Clamp the cord again closer to the vaginal opening.
19. Use this clamp to help deliver the placenta. Use gentle controlled traction on the clamp with one hand while placing the other hand suprapubically to push the uterus upward.
20. If the placenta does not deliver easily, stop and wait a few minutes to allow it to come away from the uterine wall naturally and then try again.
21. Placenta must be fully intact. If not, check for retained intrauterine placental products, and manually remove them; retained placenta may cause postpartum hemorrhage.
    - Massage the fundus externally to prevent uterine atony, but only after all products have been removed.
22. Repair any vaginal or cervical lacerations.

115.5 Complications

- Postpartum hemorrhage due to uterine atony or retained products
- Uterine inversion
- Rectal or urethral injuries
- Shoulder dystocia
- Meconium aspiration

115.6 Pearls and Pitfalls

- Pearls
  - Postpartum hemorrhage has a high incidence during the first hour postpartum.
  - Be aware of possible second delivery.
  - Oxytocin or Methergine (methylergonovine maleate) may be used after delivery of the placenta; both help increase uterine contractions and decrease postpartum hemorrhage.
- Pitfalls
  - Rushing delivery of the head/shoulders can lead to trauma to the mother/baby.
  - Rushing delivery of the placenta, when the placenta is not ready to detach, may cause placental or cord tearing or uterine inversion. Remember: it may take up to 20 min for the placenta to detach from the uterus.

Selected Reading

Desai S, Henderson SO, Mallon WK. Labor and delivery and their complications. In: Marx J, Hockberger R, Walls R, editors. Rosen's emergency medicine: concepts and clinical practice. 7th ed. Philadelphia: Mosby; 2010.

Liao JB, Buhimschi CS, Norwitz ER. Normal labor: mechanism and duration. Obstet Gynecol Clin North Am. 2005;32:145–64.

Norwitz ER, Robinson JN, Repke JT. Labor and delivery. In: Gabbe SG, Niebyl JR, Simpson JL, editors. Obstetrics: normal and problem pregnancies. 4th ed. Philadelphia: Saunders; 2001. p. 353–94.

116 Shoulder Dystocia Management

Irina Fox Brennan and Joseph A. Tyndall

116.1 Indications

- Retracting fetal head ("turtle sign")
- Failure of anterior shoulder delivery following gentle downward traction
- Difficult face-chin delivery
- Failure of shoulders to descend

116.2 Contraindications

- Mother with diabetes and fetus larger than 4500 g
- Mother without diabetes and fetus larger than 5000 g
- History of shoulder dystocia with prior birth
- Factors that signal the need for cesarean section:
  - Complete/partial placenta previa
  - Cord prolapse
  - Brow/face presentation
  - Non-reassuring fetal heart rate
  - Extensive uterine surgery and previous classic uterine incision

116.3 Materials and Medications

- No specialized materials/equipment required.
- Team coordination is of utmost importance.

I.F. Brennan, MD • J.A. Tyndall, MD, MPH (✉)
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: ibrennan@ufl.edu; tyndall@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_116

116.4 Procedure

1. *Accurate documentation*: Designate a team member to document the progress of labor, the position and rotation of the infant's head, the time of delivery for the head and body, the presence of episiotomy, anesthesia requirements, the duration of shoulder dystocia, whether the shoulder is anterior or posterior at the time of delivery, the onset/duration/results of maneuver(s) performed, and Apgar scores.

   The HELPERR mnemonic is commonly employed:
2. *Help recruitment*: Involve anesthesiology, pediatric resuscitation, and obstetric/gynecology colleagues either as part of advance preparations or via activation of protocol during labor.
3. *Episiotomy consideration*: Not required but might be of benefit if more space is needed for rotation maneuvers.
   - Episiotomy alone will not release the impacted shoulder.
4. *Legs positioned for the McRoberts maneuver*: Position the woman's thighs onto her abdomen by abducting and flexing her hips. Have two assistants help her, one on either side. This increases the functional size of the bony pelvis by rotating the pubic symphysis toward the mother's head, hence aiding delivery (Fig. 116.1).

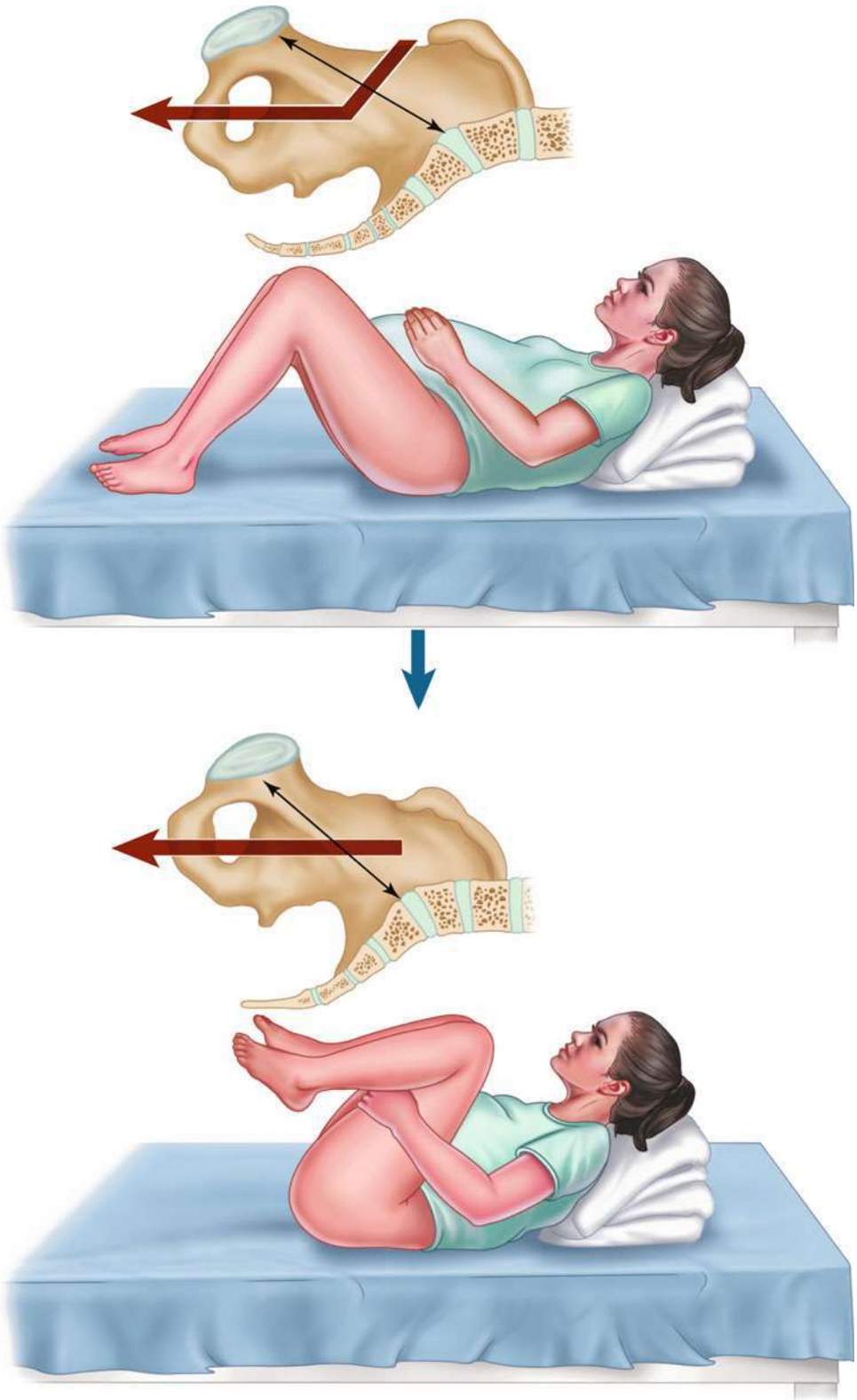

Fig. 116.1 McRoberts maneuver: hyperflexion of the maternal thighs against the abdomen

5. *Pressure suprapubically* (*Rubin I maneuver*): While continuing downward traction, have a colleague place his or her hand suprapubically on the fetal anterior shoulder and apply pressure down and laterally on its posterior aspect in 30-second increments. This allows the fetal shoulders to enter the pelvis in an oblique fashion (pelvic inlet is widest in the transverse plane) (Fig. 116.2).
6. *Enter maneuver* (*internal rotation*): Rotate the anterior shoulder into the oblique plane and under the maternal symphysis.
   - It may be necessary to push the fetus slightly up into the pelvis in order to successfully perform this maneuver.
     (a) *Rubin II maneuver*: Insert one hand vaginally behind the fetal anterior shoulder and rotate the shoulder toward the chest. This will reduce the diameter of the fetal shoulder girdle and facilitate delivery (Fig. 116.3).
     (b) *The Woods corkscrew maneuver*: Place one hand at the front of posterior shoulder and push upward gently. This can be combined with the Rubin II maneuver (Fig. 116.4).
     (c) *The reverse Woods corkscrew maneuver*: Place one hand behind the fetal posterior shoulder and rotate in the direction opposite to that of the Woods corkscrew maneuver to adduct the shoulder.
7. *Remove the posterior arm*: Flex the fetal elbow and deliver the forearm in a sweeping motion over the fetal anterior chest wall. This allows the fetus to drop into the sacral hollow, freeing the anterior shoulder impaction.
   - Avoid grasping and pulling on the fetal arm directly because it may fracture the fetal humerus.
8. *Roll the patient* (*Gaskin maneuver*): Position the patient on all fours (this acts as an upside-down McRoberts maneuver). Continue gentle traction. The turning itself, as well as gravity, will often dislodge the impacted shoulder.

   Last-resort maneuvers:
9. *Deliberate fracture of the fetal clavicle*: Apply direct pressure upward in the middle of the clavicle to reduce the shoulder-to-shoulder distance.

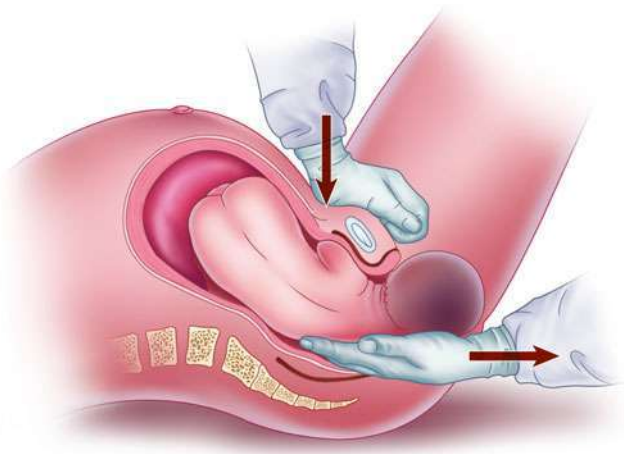

Fig. 116.2 Ruben I maneuver: suprapubic pressure is directed at the anterior shoulder

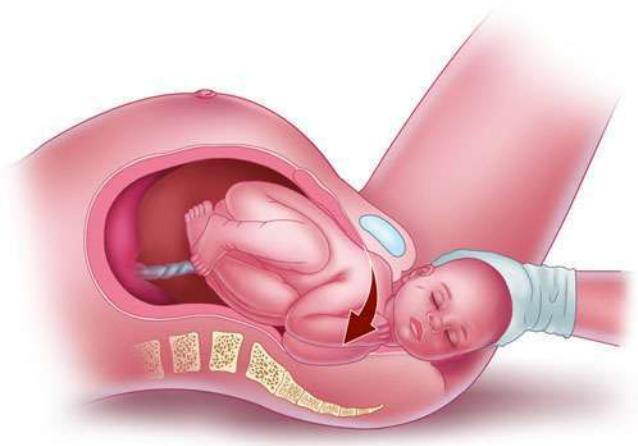

Fig. 116.3 Ruben II maneuver: pressure is applied to the most accessible part of the fetal shoulder and rotated toward the chest

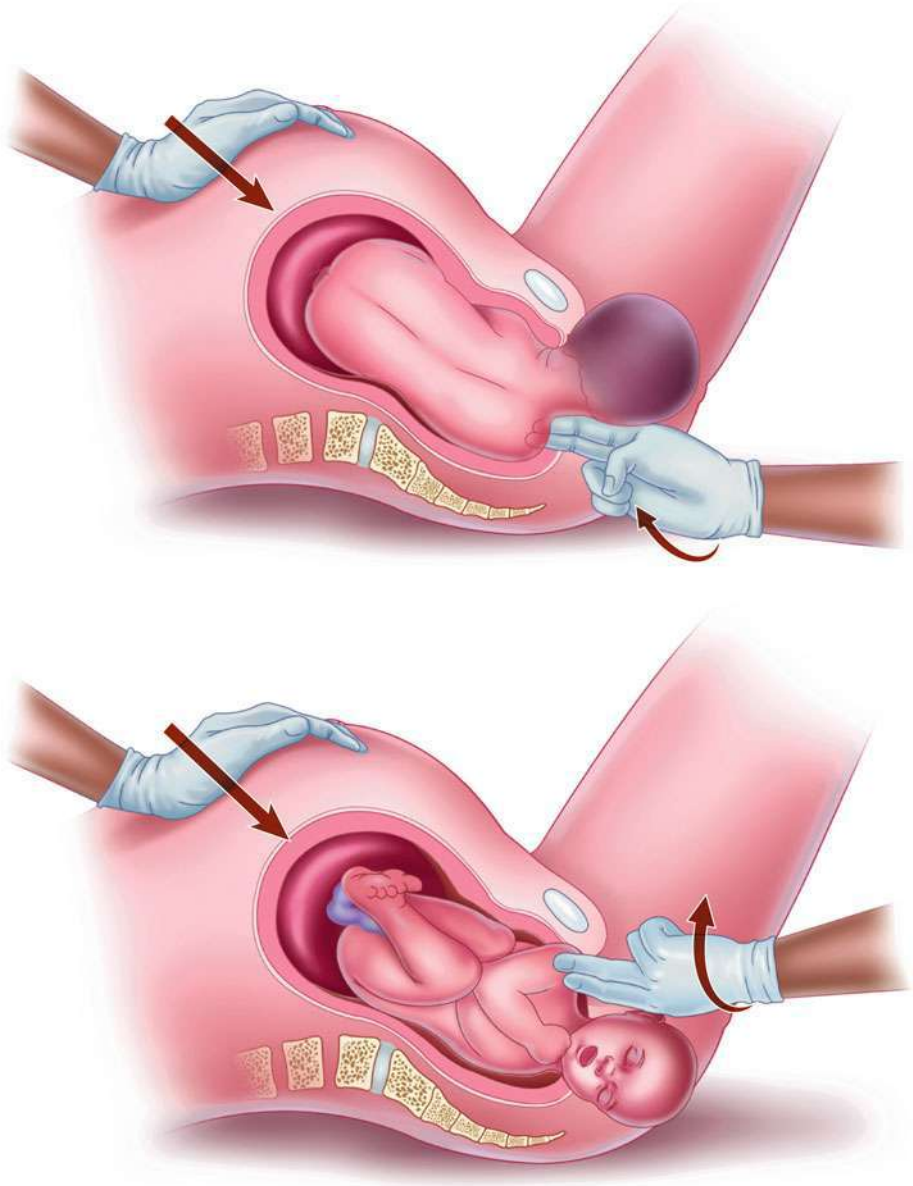

Fig. 116.4 The Woods corkscrew maneuver: pressure is applied to the clavicle of the posterior arm, enabling rotation and dislodgement of the anterior shoulder

10. *Zavanelli maneuver*: Rotate the fetal head into the direct occiput anterior position, flex, and push back into the birth canal. Continue holding upward pressure until cesarean delivery. Ensure the operating and anesthesiology teams are present (Fig. 116.5).
    - Do not perform this maneuver if the nuchal cord has been clamped or cut.
11. *Symphysiotomy*: Place the patient in the exaggerated lithotomy position, insert a transurethral catheter, displace the urethra laterally, and separate the cephalad portion of fibrous cartilage of the symphysis pubis under local anesthesia with either a scalpel blade or Kelly clamp.
    - This technique is associated with significant maternal morbidity, including bladder neck injury and infection. It is truly the last resort and should be used only when all other methods have failed and cesarean delivery is unavailable.

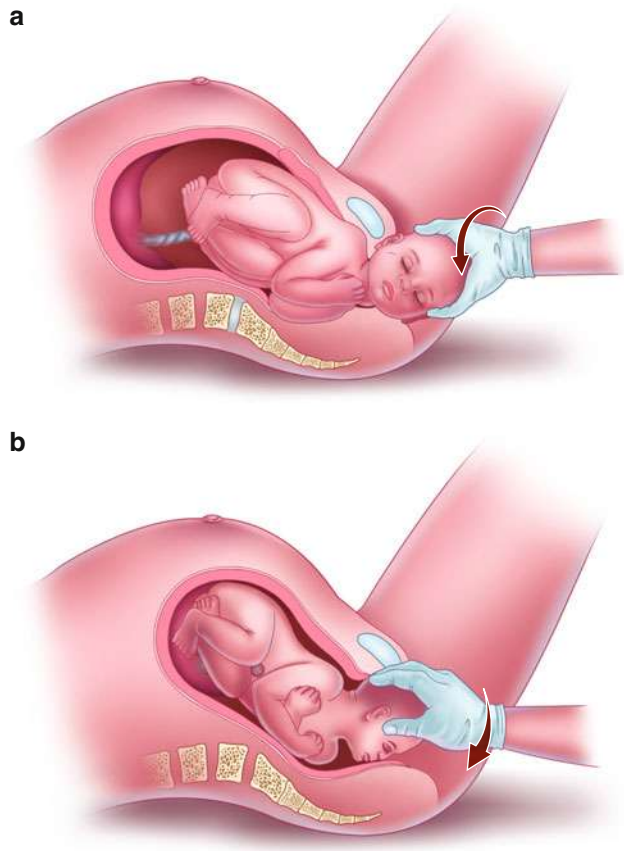

Fig. 116.5 Zavanelli maneuver: the fetal head is rotated into the direct occiput anterior position, flexed, and pushed back into the birth canal. (**a**) Rotation of the head to the occiput anterior position. (**b**) Replace head with constant firm pressure on occiput with palm of hand flexing head

116.5 Complications

- Maternal:
  - Postpartum hemorrhage
  - Fourth-degree laceration/soft tissue damage to cervix and vagina
  - Uterine rupture
  - Bladder atony
  - Symphyseal separation with or without femoral neuropathy (transient)
  - Sacroiliac dislocation
- Fetal:
  - Brachial plexus palsies (Erb, Klumpke): most resolve within 6–12 months; permanent injury occurs in less than 10 % of patients
  - Clavicular and/or humeral fractures
  - Ischemia/hypoxia/asphyxia due to umbilical cord compression with or without neurological damage
  - Significant fetal acidosis, with pH drop of 0.04 U/min between delivery of the head and trunk
  - Death

116.6 Pearls and Pitfalls

- Pearls
  - Many cases of shoulder dystocia and birth trauma are encountered in the absence of risk factors and in non-macrosomic infants. Shoulder dystocia is largely neither predictable nor preventable.
  - Presence of risk factors for shoulder dystocia should prompt advance preparation in anticipation of a difficult delivery. Risk factors include maternal diabetes, obesity, multiparity, advanced maternal age, prolonged pregnancy, macrosomia (fetal weight >4,500–5,000 g), male fetal gender, and prior birth complicated by shoulder dystocia.
- Pitfalls
  - Do not cut and clamp the nuchal cord if at all possible in order to avoid fetal hypoxia and hypotension should shoulder dystocia arise.
  - Labor induction or prophylactic cesarean delivery of macrosomic fetuses has not decreased rates of shoulder dystocia.
  - Avoid fundal pressure in shoulder dystocia because it has been shown to increase the risk of permanent neurological damage and uterine rupture.

Selected Reading

Acker DB. A shoulder dystocia intervention form. Obstet Gynecol. 1991;78:150–1.

Athukorala C, Middleton P, Crowther CA. Intrapartum interventions for preventing shoulder dystocia. Cochrane Database Syst Rev. 2006;4:CD005543.

Baxley EG, Gobbo RW. Shoulder dystocia. Am Fam Physician. 2004;69:1707–14.

Chauhan SP, Gherman R, Hendrix NW, Bingham JM, Hayes E. Shoulder dystocia: comparison of the ACOG practice bulletin with another national guideline. Am J Perinatol. 2010;27:129–36.

Gherman RB, Chauhan S, Ouzounian JG, Lerner H, Gonik B, Goodwin TM. Shoulder dystocia: the unpreventable obstetric emergency with empiric management guidelines. Am J Obstet Gynecol. 2006;195:657–72.

Gottlieb AG, Galan HL. Shoulder dystocia: an update. Obstet Gynecol Clin North Am. 2007;34:501–31.

Hoffman MK, Bailit JL, Branch DW, et al. Consortium on safe labor. A comparison of obstetric maneuvers for the acute management of shoulder dystocia. Obstet Gynecol. 2011;117:1272–8.

Patterson DA, Winslow M, Matus CD. Spontaneous vaginal delivery. Am Fam Physician. 2008;78:336–41.

Sokol RJ, Blackwell SC, American College of Obstetricians and Gynecologists. Committee on Practice Bulletins–Gynecology. ACOG practice bulletin no. 40: shoulder dystocia. November 2002 (replaces practice pattern no. 7, October 1997). Int J Gynaecol Obstet. 2003;80:87–92.

117 Breech Delivery in the Emergency Department

Kristin Stegeman, Sapnalaxmi Amin, Anton A. Wray, and Joseph A. Tyndall

Breech presentation: the buttocks enter the pelvis before the head (Fig. 117.1)

- Frank (extended):
  - Hips flexed and knees extended, buttocks presenting
  - Most common
- Incomplete/complete (flexed):
  - One or both hips and knees flexed, buttocks presenting
- Footling:
  - One or both hips and knees extended, foot presenting

The emergent delivery of a breech baby is one of the most challenging situations for an emergency physician.

Fig. 117.1 Variations of the breech presentation

K. Stegeman, MD • A.A. Wray, MD (✉)
Department of Emergency Medicine, The Brooklyn Hospital Center, New York, NY, USA
e-mail: kristin.stegeman@gmail.com; anw9071@nyp.org

S. Amin, MD
Department of Family Medicine/Urgent Care, Bayside Urgent Care Center, Clearwater, FL, USA
e-mail: sapna_amin59@yahoo.com

J.A. Tyndall, MD, MPH
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: tyndall@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_117

117.1 Indications

- Inevitable delivery of fetus with complete or frank breech presentation
- Absence of vertex presentation

117.2 Contraindications

- Footling presentation (increased risk of cord prolapse and entrapment of the after-coming head).
- There is adequate time to transfer the mother safely to labor and delivery in the knee/chest position (administer subcutaneous terbutaline before transfer).

117.3 Materials and Medications

- Supplemental oxygen for mother
- Ultrasound machine
- Piper forceps
- Sterile towels/gloves
- Betadine (povidone-iodine) or another antiseptic preparing solution
- Sterile lubricant (Surgilube)
- Instruments to cut the umbilical cord/perform episiotomy/resuscitate baby (Kelly clamps/scissors/#10 scalpel/bulb suction)
- Available emergency department staff
- Pediatric, obstetrics, and anesthesia practitioners

117.4 Procedure

1. Assess the health of the mother and the baby (vitals/physical examination/history).
2. Listen to the fetal heart with stethoscope or Doppler ultrasound.
   - Should be 120–160 beats/min.
3. Identify the type of presentation by bedside ultrasound or by digital exam.
4. Perform a sterile digital examination to confirm the position of the baby and the stage of labor.
   - Membranes should not be artificially ruptured. The amniotic sac will help to dilate the cervix, lubricate the canal, and protect the umbilical cord from compression.
5. If footling presentation: await OB/general surgery for emergency C-section.
6. If frank/complete/incomplete: instruct the patient to push when the cervix is completely dilated.
7. When the breech has descended to the perineum, consider performing an episiotomy if more space is needed. Cleanse the perineum with antiseptic and sterile lubricant beforehand.
8. Allow the baby to extrude to the umbilicus with maternal efforts alone. Do not exert traction before this time.
9. If frank: deliver the posterior leg by gently guiding the sacrum anteriorly, grasping the thigh and flexing the leg at the knee.
10. Deliver the anterior leg in a similar manner while guiding the sacrum posteriorly.
11. If incomplete: deliver the extended leg as #8 or #9 as appropriate.
12. Delivery continues as it would for a complete presentation.
13. Wrap the legs/buttocks in a clean towel to decrease trauma (create grip).
14. Grasp the upper legs with the index fingers holding the anterior iliac crests. Place the thumbs on the sacrum (Fig. 117.2).
15. Apply gentle traction as the mother pushes until the scapulae and axillae are visible.
16. If there is difficulty delivering the shoulders, deliver the posterior shoulder by rotating the trunk 90° and applying gentle downward traction to rotate the shoulder anteriorly.
17. Rotate the baby 180° so as to deliver the anterior arm in a similar manner.
18. If the arms do not spontaneously deliver, a finger can be hooked over the shoulders to bring the arm down while rotating the trunk as above (Fig. 117.3).
19. Use the McRoberts position to increase the diameter of the pelvis (Fig. 117.4).

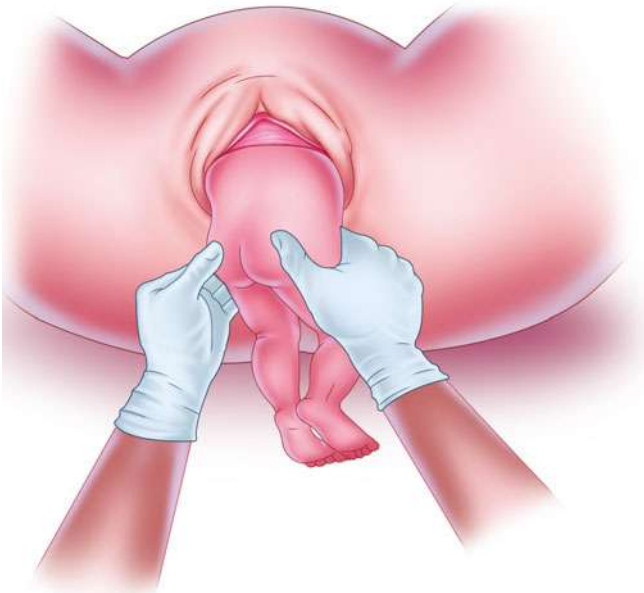

Fig. 117.2 Correct placement of hands on sacrum

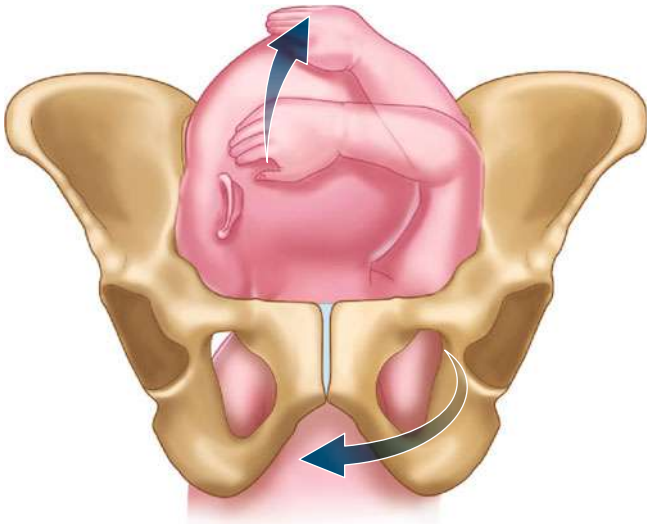

Fig. 117.3 Rotation toward maternal symphysis pubis to avoid nuchal arm

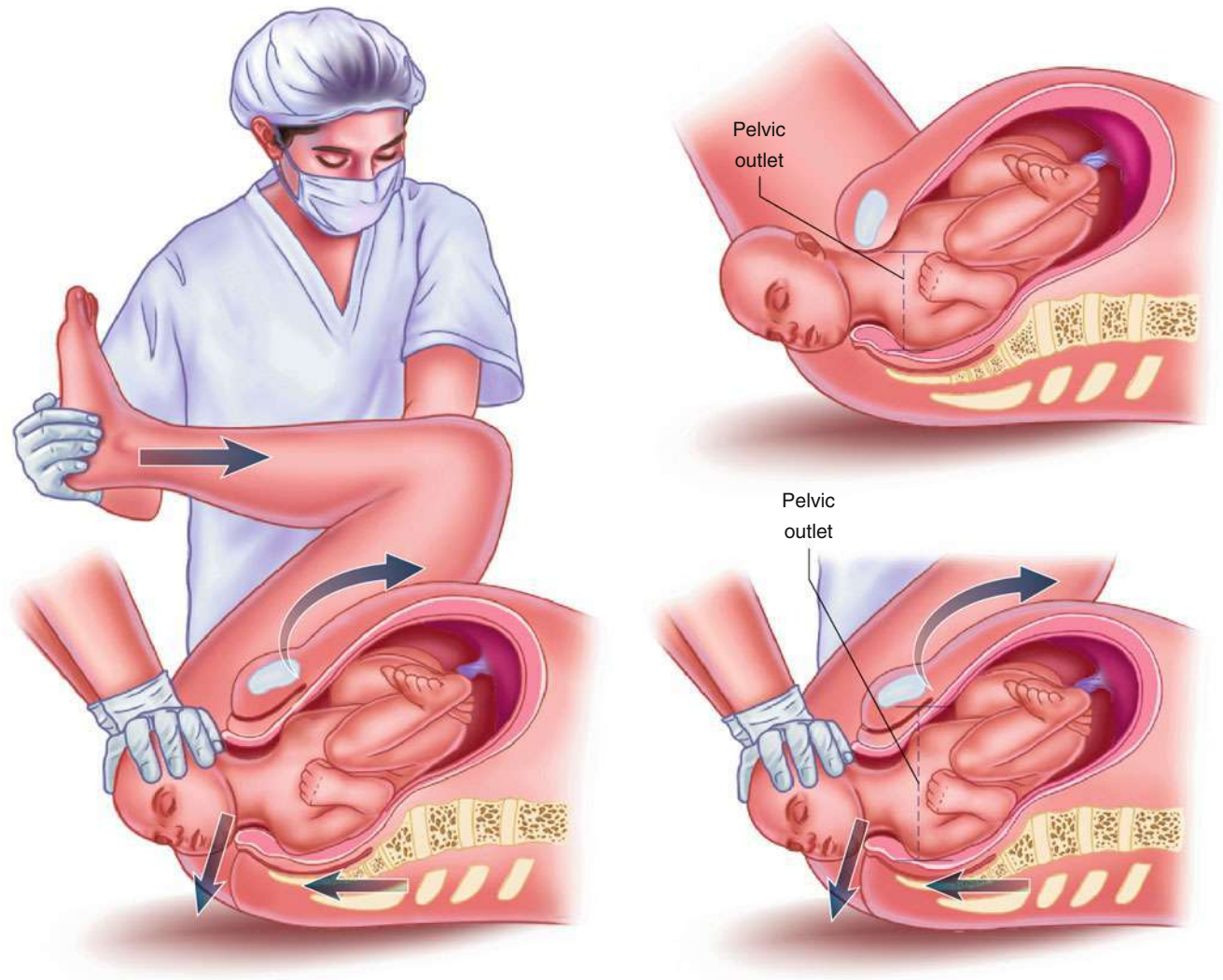

Fig. 117.4 The McRoberts maneuver in vertex (head-first) presentation

20. Maintain the baby in the same plane as the vagina (support the body with the forearm) and place the second and fourth fingers over the maxilla of the baby. Place the middle finger in the mouth or on the chin and the other hand on the upper back/occiput (Fig. 117.5).
    - Avoid extreme elevation of the fetus to prevent hyperextension and cervical spine injury. Piper forceps may also be used to promote flexion.
21. Deliver the head in the flexed position.
    - If the baby descends with the neck and abdomen facing anteriorly: Grasp the shoulders posteriorly with two fingers of one hand while the other hand flexes the abdomen and the baby's feet are brought upward (Fig. 117.6).
    - If the baby's neck remains extended: Leave the baby hanging (weight = traction). When the hairline appears under the symphysis, grab the baby by the feet and elevate upward (Fig. 117.7).
22. Clamp and cut the umbilical cord (collect arterial and venous samples for pH).
23. Suction the baby's mouth and nose, and resuscitate as indicated.
24. Deliver the maternal placenta.
25. Repair tears or episiotomy made during delivery.

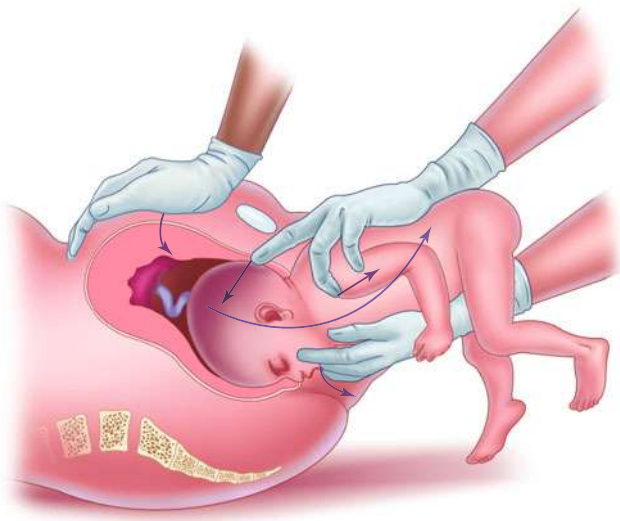

Fig. 117.5 Delivery of head in flexed position

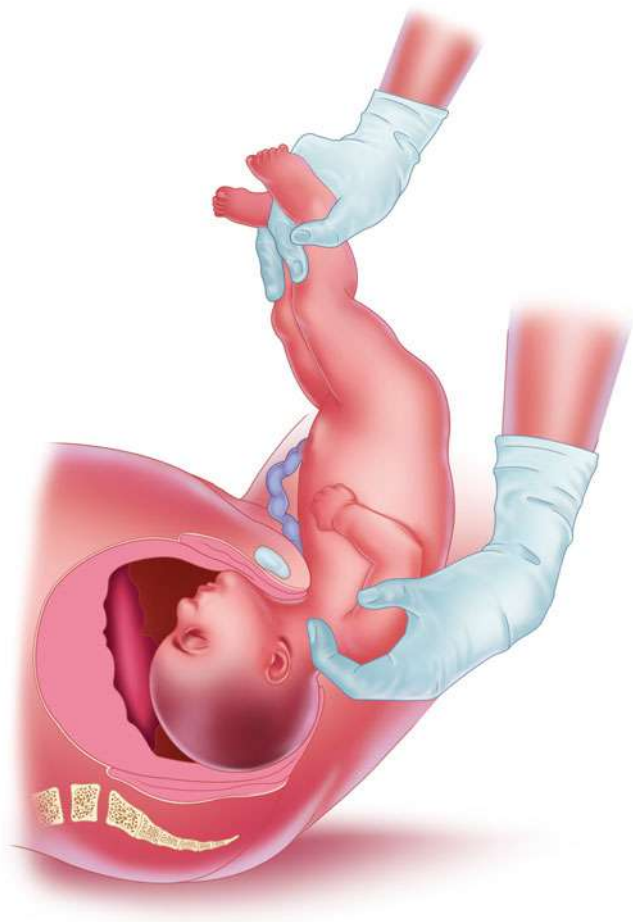

Fig. 117.6 Correct hand placement if abdomen facing anterior

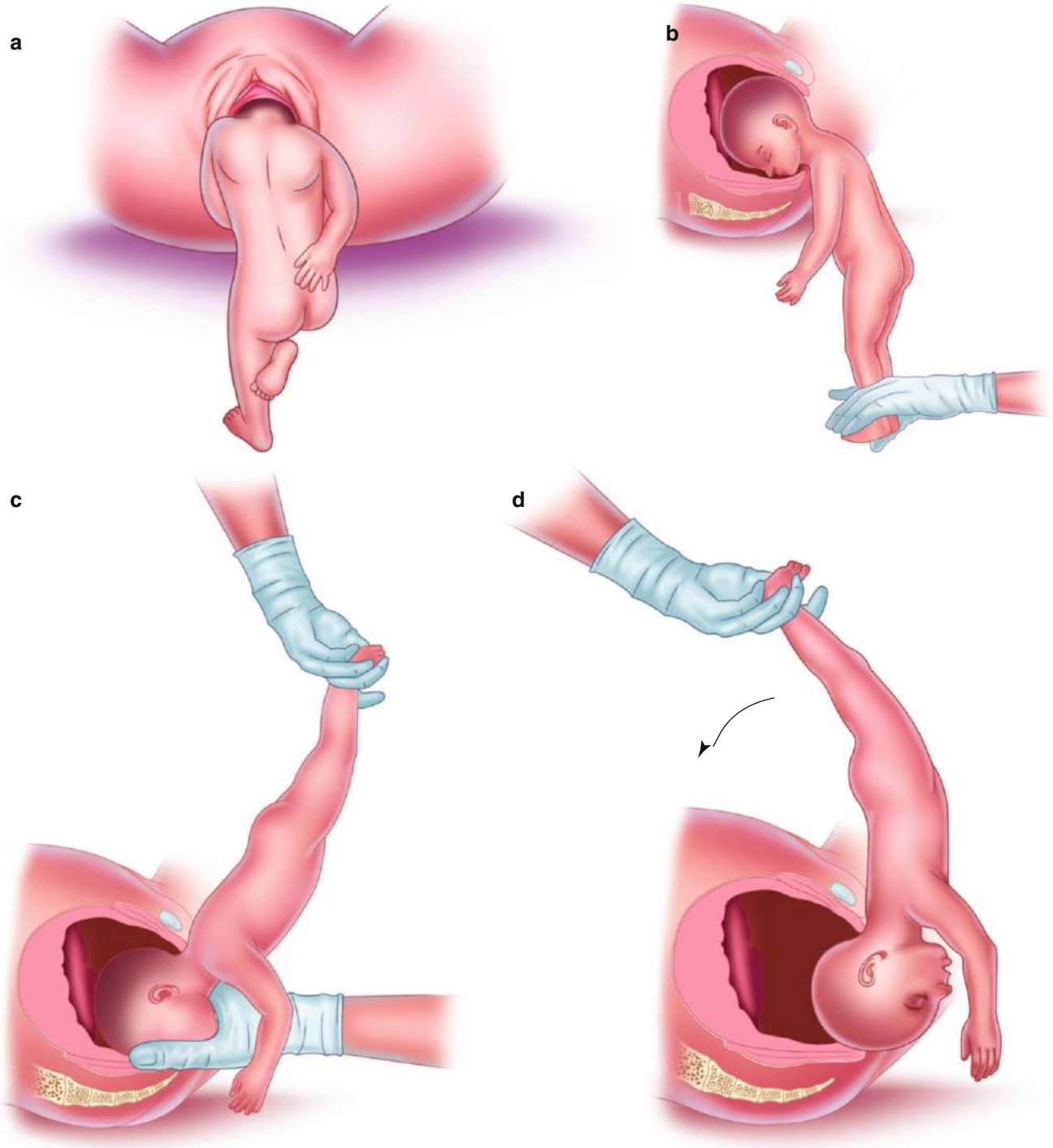

Fig. 117.7 The Burns-Marshall maneuver. (**a**) Allow baby to hang until you can see the hair at the nape of his neck; (**b**) swing the baby's head clear of the birth canal; (**c**) left hand guards and slips the perineum over fetal mouth; suction baby's air passage to clear mucus; (**d**) hold baby's feet

117.5 Complications

- Umbilical cord prolapse
- Brachial plexus injury (from the nuchal arm)
- Fetal head entrapment
- Cervical spine injury (from hyperextension of the neck)
- Birth asphyxia

117.6 Pearls and Pitfalls

- Pearls
  - Allow the uterine contractions to help deliver the baby.
- Pitfalls
  - Do not rush the delivery or use too much force. This can increase the risk of trauma to the baby and mother.
  - Beware of nuchal arm. To avoid brachial plexus injury: Rotate the face of baby toward the maternal symphysis pubis. This will reduce the tension keeping the arm around the back of the fetal head (Fig. 117.3).

Selected Reading

Auerbach PS. Gynecologic and obstetric emergencies. In: Wilderness medicine. 6th ed. Philadelphia: Elsevier; 2012.

Buckley RG, Knoop KJ. Gynecologic and obstetric conditions. In: Knoop KJ, Stack LB, Storrow AB, Thurman RJ, editors. The atlas of emergency medicine. 3rd ed. New York: McGraw Hill; 2010.

Cunningham FG, Leveno KJ, Bloom SL, Hauth JC, Rouse DJ, Spong CY. Breech presentation and delivery. In: Williams obstetrics. 23rd ed. New York: McGraw Hill; 2010.

Kish K, Collea JV. Chapter 21. Malpresentation & cord prolapse. In: DeCherney AH, Nathan L, editors. Current diagnosis & treatment: obstetrics & gynecology. 10th ed. New York: McGraw Hill; 2007.

Kotaska A, Menticoglou S, Gagnon R. Vaginal delivery of breech presentation. Int J Gynecol Obstet. 2009;107:169–76.

Probst BD. Emergency childbirth. In: Roberts JR, Hedges JR, editors. Clinical procedures in emergency medicine. 5th ed. Philadelphia: Elsevier; 2010.

Management of Primary Postpartum Hemorrhage

118

Megan Kwasniak, Anton A. Wray, and Joseph A. Tyndall

Postpartum hemorrhage (PPH) is defined as ≥500 ml blood loss within 24 h of vaginal delivery or 1000 ml loss within 24 h of cesarean section. It is the leading cause of maternal mortality worldwide.

118.1 Indications

- Excessive vaginal bleeding with or without pain and/or hemodynamic instability within 24 h of delivery

118.2 Contraindications

- There are no absolute contraindications to the management of PPH.

118.3 Materials and Medications

- Sterile technique
- Good lighting
- Sponge forceps
- Gauze
- Towels
- IV fluids
- Type and screen/crossmatch of blood
- Absorbable suture with curved needle
- Needle holder
- Tooth forceps

118.4 Procedure

- Standard resuscitation measures: Place IV, O_2, monitor.
- Assessment and treatment should occur simultaneously.
- All techniques should be performed under strict sterile conditions.

118.4.1 Uterine Exam

1. Assess by placing the hand on the uterine fundus and checking its size and firmness.
2. High, soft, or boggy uterus implicates retained placenta or uterine atony. Start external fundal massage.
3. Non-palpable uterus implicates uterine inversion.

118.4.2 Vaginal Exam

1. Keeping one hand on the abdomen, gently examine the vaginal canal using the other sterilely gloved hand.
2. Gently scoop out any clots/retained placenta that are easily removable.
3. Look for traumatic sources of bleeding from the perineum, vaginal walls, and cervical lacerations.
4. Gauze-wrapped ring forceps can be used to assist direct visualization and clearing of clots.

M. Kwasniak, MD
Emergency Department, The Brooklyn Hospital Center, New York, NY, USA
e-mail: megan.kwasniak@gmail.com

A.A. Wray, MD (✉)
Department of Emergency Medicine, The Brooklyn Hospital Center, New York, NY, USA
e-mail: anw9071@nyp.org

J.A. Tyndall, MD, MPH
Department of Emergency Medicine, University of Florida Health, Gainesville, FL, USA
e-mail: tyndall@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_118

118.4.3 If Uterine Atony Suspected

1. Continue external fundal massage with the abdominal hand. Make a fist with the vaginal hand and start bimanual massage. Raise the uterus from the pelvis and pivot it anteriorly, compressing it between the external hand and the internal fist. This maneuver will result in expression of any clots present and decrease uterine bleeding via direct compression (Fig. 118.1).
2. Administer oxytocin (20–40 units in 1 L of normal saline or lactated Ringer's solution intravenously [IV]; alternatively give 10 units intramyometrially with a spinal needle).
3. Methylergonovine may also be used at this time if oxytocin fails to reduce uterine bleeding (100 or 125 mcg IV or intramyometrially; alternatively 200 or 250 mcg may be given intramuscularly).
4. Continue massage until bleeding slows and the uterus becomes more firm. This can take 15–30 mins.

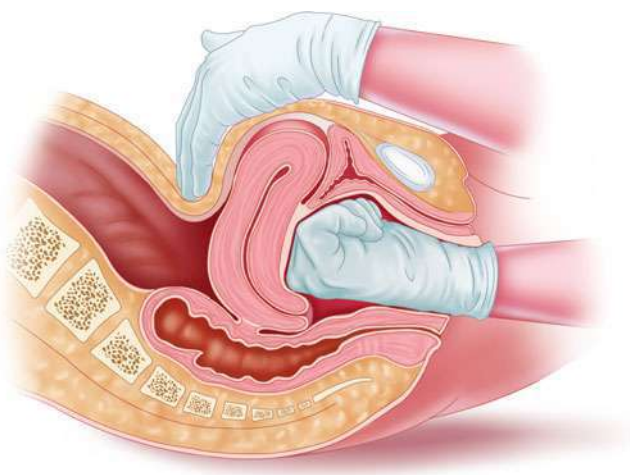

Fig. 118.1 Intrauterine massage for uterine atony

118.4.4 If Retained Placenta Suspected

1. Manual removal of the placenta or any of its retained tissue should be facilitated with sedation or additional analgesia.
2. Keeping the thumb and fingers together in a teardrop shape and using sterile technique as described previously, insert the hand through the vaginal canal and the cervix into the lower uterine segment.
3. Keep the other hand on the lower abdomen to continue gentle yet firm upward pressure and massage.
4. Find the placental edge within the uterus, grasp it gently, detach it from the uterine wall, and withdraw the hand from the patient.
5. Repeat the maneuver as necessary; remove any additional clots from the uterus and continue bimanual massage until it becomes firm and bleeding decreases.
6. If the entire placenta has been removed this way, an assistant should be available to inspect it for completeness and any torn vessels.
7. If the bleeding is particularly severe and placenta accreta is strongly suspected, pack the uterus with gauze and prepare the patient for urgent surgical intervention.

118.4.5 If Uterine Inversion Suspected

1. Using sterile technique, insert a fist through the vaginal canal and push the inverted fundus back through the cervical canal with pressure directed toward the umbilicus.
2. If the uterus has contracted, tocolytics such as IV magnesium sulfate or terbutaline can be used to aid myometrial relaxation. Alternatively, provided the patient is not overly hypotensive, nitroglycerin 50–100 mcg IV may be administered to relax the myometrium and facilitate return to normal uterine position.
3. If manual replacement is ineffective, hydrostatic reduction may be attempted. Warm fluids are run into the upper vagina under high pressure while occluding the introitus.

118.4.6 Trauma: Genital and Perineal Lacerations

1. For significant cervical lacerations: use absorbable sutures with a continuous interlocking stitch technique to close (Fig. 118.2).
2. For vaginal wound repair: place the initial and final stitch above the apices of the lacerations and grab a good amount of tissue with the needle. Small bites can lead to ongoing bleeding and hematoma formation.
3. Observe the repaired lacerations for any additional bleeding after the torn edges have been sutured.
   - Apply additional pressure to any site that continues to ooze blood; gauze-wrapped ring forceps may be used for this purpose if necessary.

If none of the above causes are apparent, consider underlying coagulopathies and treat appropriately. This may require administration of fresh-frozen plasma, platelets, or clotting factors as indicated.

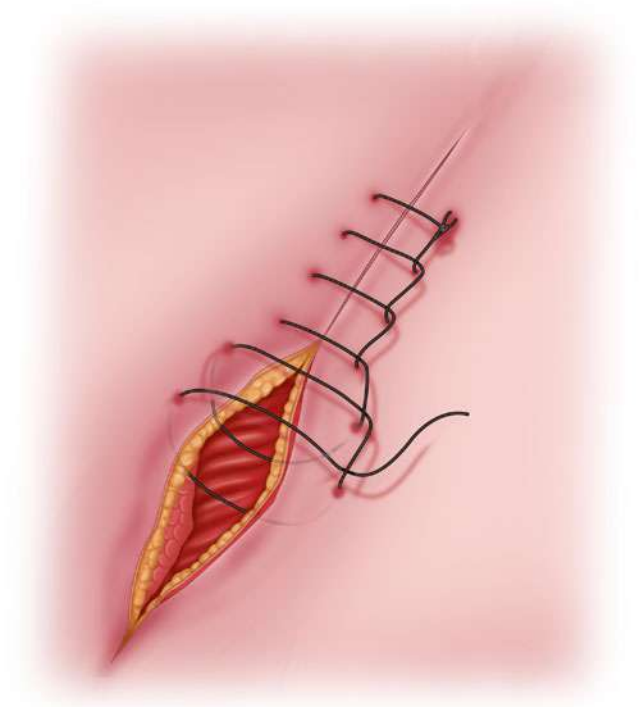

Fig. 118.2 Continuous interlocking stitch: perineal lacerations repair

118.5 Complications

- Uterine perforation and scarring
- Urinary and genital tract trauma and injury
- Genitourinary and genitointestinal fistula
- Pelvic hematoma
- Genital vascular injury
- Infection and sepsis
- Disseminated intravascular coagulation (DIC)
- Maternal death

118.6 Pearls and Pitfalls

- Pearls
  - Causes of PPH can be divided into the "5 Ts":
    - Tone: uterine atony, occurring within the first 4 h after delivery.
    - *T*issue: retention of the placenta, especially placenta accreta and its fragments, more common at extreme preterm deliveries.
    - *T*rauma: injury to the uterus, cervix, and perineal structures after delivery of a large fetus, use of forceps and/or vacuum, frequent vaginal manipulation during delivery, and episiotomy procedures.
    - *T*hrombosis: intrinsic or acquired coagulation disorders, including idiopathic thrombocytic purpura (ITP); hemolysis, elevated liver enzymes, and low platelets (HELLP) syndrome; and disseminated intravascular coagulation (DIC), as well as preexisting conditions such as von Willebrand disease.
    - *T*raction: inversion of the uterus during placental delivery secondary to excessive traction on the umbilical cord. The uterine fundus can be within the endometrial cavity, in the cervical canal, or outside the external os and within the vaginal canal.
  - The administration of broad-spectrum antibiotics should be strongly considered following any manual removal, exploration, or instrumentation of the uterus and the genital tract.
  - Bedside ultrasonography can be very helpful for identifying uterine abnormalities, retained placental tissue, free fluid in the pelvis, and/or intrauterine hematoma.
  - Risk factors for PPH include prolonged active phase of labor, previous PPH, multiple pregnancy, and history of a bleeding disorder.
- Pitfalls
  - Failure to recognize and treat PPH early increases morbidity and mortality.
  - Underestimating the potential blood loss of PPH. The gravid uterus at term has a blood flow of 600 ml/min (non-gravid: 60 ml/h).

Selected Reading

Anderson J, Duncan E. Prevention and management of postpartum hemorrhage. Am Fam Physician. 2007;75:875–82.

Druelinger L. Postpartum emergencies. Emerg Med Clin North Am. 1994;12:219–37.

Leduc D, Senikas V, Lalonde AB, et al. Active management of the third stage of labour: prevention and treatment of postpartum hemorrhage. J Obstet Gynaecol Can. 2009;31:980–93.

Miller S, Lester F, Hensleigh P. Prevention and treatment of postpartum hemorrhage: new advances for low-resource settings. J Midwifery Womens Health. 2004;49:283–92.

Sheiner E, Sarid L, Levy A, Seidman DS, Hallak M. Obstetric risk factors and outcome of pregnancies complicated with early postpartum hemorrhage: a population-based study. J Matern Fetal Neonatal Med. 2005;18:149–54.

Tessier V, Pierre F. Risk factors of postpartum hemorrhage during labor and clinical and pharmacological prevention. J Gynecol Obstet Biol Reprod (Paris). 2004;33(8 suppl):4S29–56.

World Health Organization. World health report 2005: make every mother and child count. Available at: http://www.who.int/whr/2005/whr2005_en.pdf.

119 Perimortem Cesarean Section

Jordana J. Haber, Elaine B. Josephson, and Muhammad Waseem

119.1 Indications

- Maternal arrest with a viable fetus (gestation >24 weeks)

119.2 Contraindications

- Stable mother
- Fetus less than 24 weeks' gestation
- Extreme fetal prematurity
- Maternal hypoxia longer than 15 min

119.3 Materials and Medications

- Cesarean section instrument tray if available
  - #10 or #11 scalpel blade, scissors, bladder retractor, 2 large retractors, gauze sponges, hemostats, suction, forceps, and straight and curved clamps
- Skin antiseptic preparing solution, such as Betadine (povidone-iodine)
- Silk suture with needle driver or skin stapler
- Sterile drapes
- Sterile gloves
- Obstetrical pack (See Chap. 115)
  - Bulb syringe and umbilical cord clamp
- Clean blanket or towels for delivery
- Neonatal resuscitation equipment

Owing to the rarity of this procedure in the emergency department, it is unlikely to have a prepared cesarean section tray available. In this case, a thorocotomy or thorocostomy tray combined with an obstetrical pack would contain all the supplies needed. At a minimum, a scalpel and an obstetrical pack are necessary.

J.J. Haber, MD
Department of Emergency Medicine, Maimonides Medical Center, New York, NY, USA
e-mail: Jordana.haber@gmail.com

E.B. Josephson, MD
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, Weill Cornell Medical College of Cornell University, Bronx, New York, NY, USA
e-mail: elaine.Josephson@nychhc.org

M. Waseem, MD (✉)
Department of Emergency Medicine, Lincoln Medical and Mental Health Center, New York, NY, USA
e-mail: waseemm2001@hotmail.com

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_119

119.4 Procedure

1. Prepare skin with antiseptic solution and a sterile drape.
2. Insert a Foley catheter to empty the bladder.
3. Continue cardiopulmonary resuscitation until delivery.
4. Obtain emergent obstetrician and neonatologist consult if available, but do not delay procedure.
5. Using a #10 or #11 blade, make a vertical midline incision beginning 4–5 cm below the xiphoid process and extend the incision to the pubic symphysis (Fig. 119.1).
6. Incise through the subcutaneous fat no further than the rectus sheath.
7. Lift the rectus sheath with a toothed forceps and make an incision with scissors to expose the uterus (Fig. 119.2).
8. With forceps and scissors, lift and incise the peritoneal membrane in the midline.
9. Identify and lift the bowel and cover it with saline-soaked gauze.
10. Retract the rectus sheath and bladder with a bladder retractor or, if not available, use saline-soaked gauze or a towel.
11. Create a 2- to 4-cm midline vertical opening in the uterus.
12. Place a finger in the opening directed caudally to protect the fetus while making a superior incision through the uterine wall. Once complete, repeat this step in the inferior direction.
13. Use a clamp to rupture the amniotic membranes. Immediately deliver the fetus and clamp the umbilical cord.
14. Expulse the head by placing a hand between the pubic symphysis landmark and the fetal occiput. Then, gently flex the fetus while simultaneously moving the head superiorly and anteriorly until delivery (Fig. 119.3).
15. Suction the mouth and nose with a bulb syringe immediately.
16. Deliver the shoulders, followed by the torso and extremities. Secure the umbilical cord with a hemostat or umbilical cord clamp 10 cm distal to the fetus and a second clamp 2 cm distal to this clamp. With scissors, incise the umbilical cord between the two clamps.
17. Immediately begin resuscitation of the infant (Fig. 119.4).
18. If the patient is still alive or regains vital signs, prepare to deliver the placenta. Begin with an oxytocin infusion at 20 U in 1 L at 10 mL/h. Apply cautious traction to the umbilical cord until the placenta separates from the uterus (Fig. 119.5).
19. Following delivery, the uterus should be closed using two layers with either 2-0 or 1-0 suture. In the case of maternal death, skin staples or a running stitch is an acceptable method of skin closure.

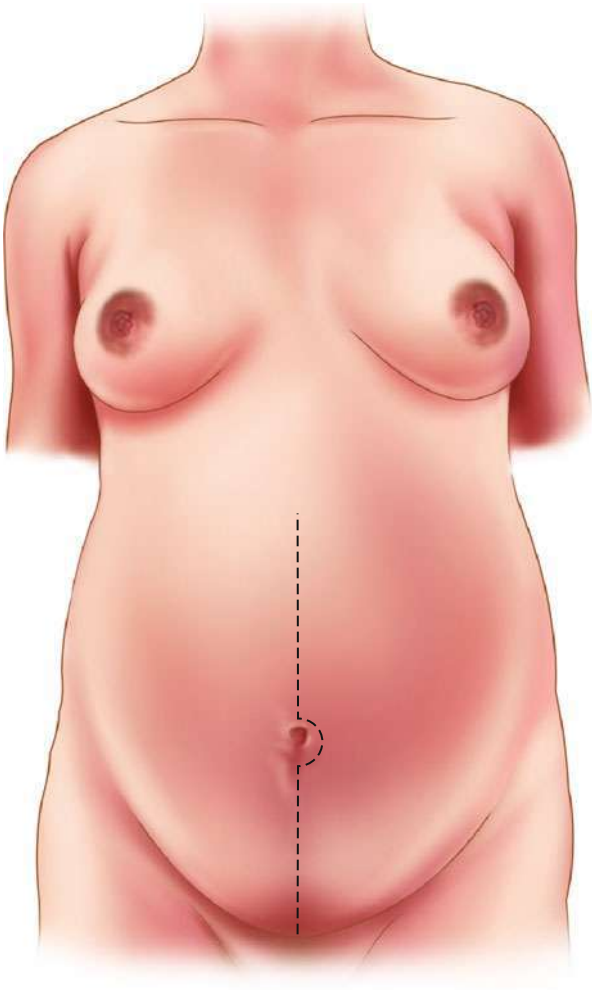

Fig. 119.1 Vertical incision

Fig. 119.2 Exposing rectus sheath

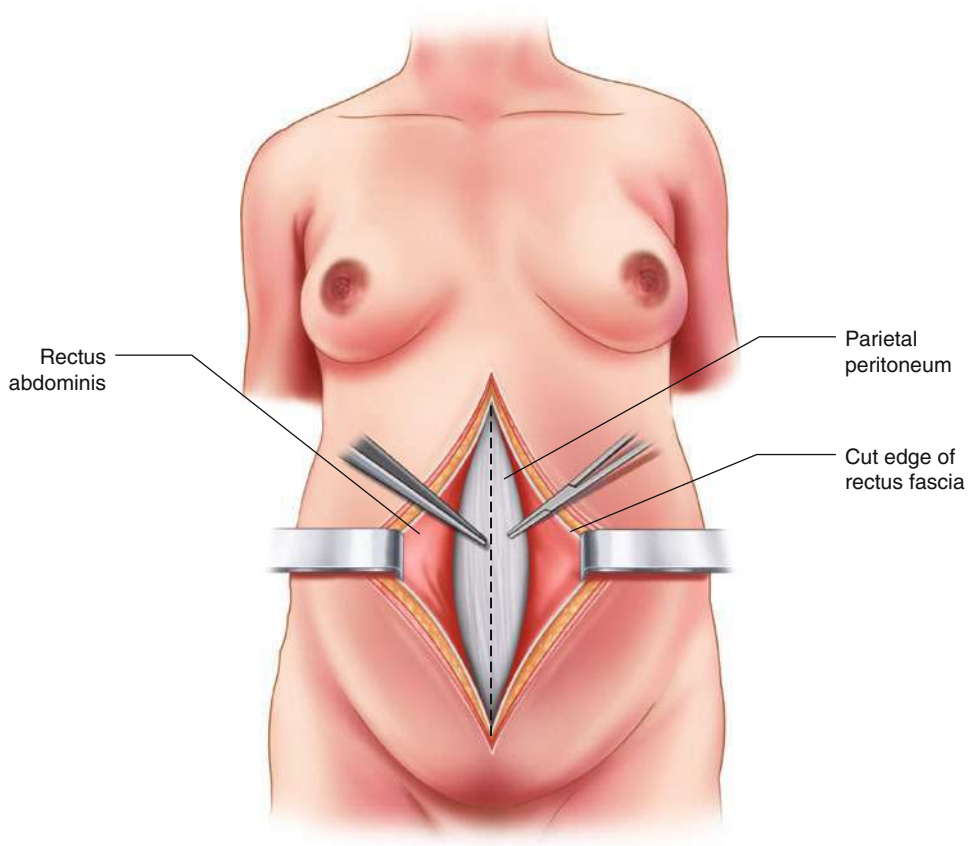

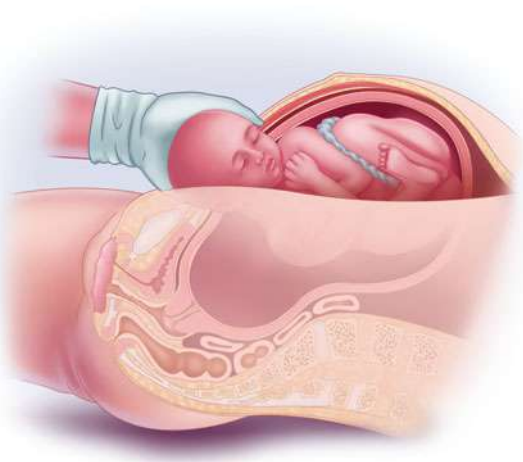

Fig. 119.3 Delivery of the fetus

Fig. 119.4 Suctioning newborn as part of resuscitation

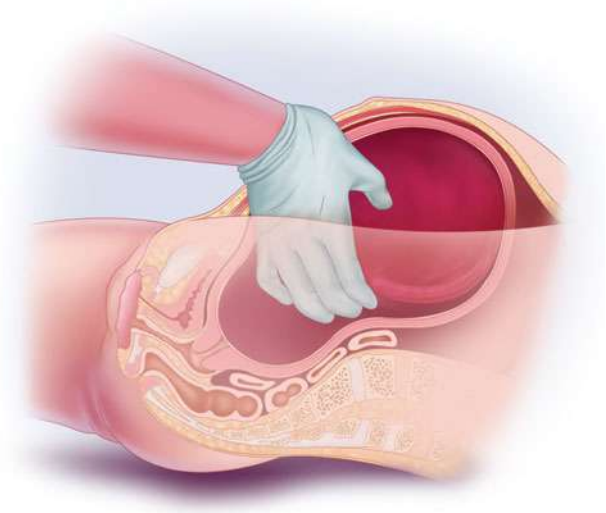

Fig. 119.5 Delivery of the placenta

119.5 Complications

- Maternal sepsis
- Maternal visceral injury
- Maternal hemorrhage
- Maternal death
- Fetal injuries and laceration
- Fetal sepsis

119.6 Pearls and Pitfalls

- Pearls
  - Perimortem cesarean section, although rarely performed, should be considered in any maternal arrest when the fetus is greater than 24 weeks' gestation.
  - In addition to saving the life of the fetus, this procedure may aid in resuscitation of the mother. Emptying of the uterus may improve thoracic compliance and, therefore, improve maternal ventilation.
- Pitfalls
  - The decision to perform an emergency cesarean section must be made early. There is a higher chance of survival if performed no more than 5 min after the onset of maternal cardiac arrest.

Selected Reading

Doan-Wiggins L. Emergency childbirth. In: Roberts JR, Hedges JR, editors. Clinical procedures in emergency medicine. 4th ed. Philadelphia: Saunders; 2004. p. 1117–43.

Flippin A, Hendricks S. Perimortem cesarean section. In: Reichman EF, Simon RS, editors. Emergency medicine procedures. New York: McGraw Hill Medical; 2004. p. 1070–8.

Gianopoulos JG. Emergency complications of labor and delivery. Emerg Med Clin North Am. 1994;12:201–17.

Jeejeebhoy FM, Zelop CM, Windrim R, Carvalho JC, Dorian P, Morrison LJ. Management of cardiac arrest in pregnancy: a systematic review. Resuscitation. 2011;82:801–19.

Whitten M, Irvine LM. Postmortem and perimortem caesarean section: what are the indications? J R Soc Med. 2000;93:6–9.

Part XV

Common Pediatric Procedures

Peripheral Venous Catheterization

120

David N. Smith and Judith K. Lucas

120.1 Indications (See Also Chap. 2)

- Fluid resuscitation
- Medication administration
- Blood draws

120.2 Contraindications

- Relative
  - Avoid catheterizing areas of trauma in which extravasation of fluid is possible (e.g., burns, open wounds, or severe edema in tissue).
  - Avoid catheterizing in an area of local infection for risk of inoculating the circulation with bacteria (e.g., cellulitis).
- Absolute
  - None

120.3 Materials and Medications

- Gloves
- Skin disinfectant (isopropyl alcohol, chlorhexidine, or Betadine [povidone-iodine])
- Appropriate-sized catheter (18- to 24-gauge [IV]) (Fig. 120.1)
  - Large child: 18 to 20 gauge
  - Infant or small child: 22 to 24 gauge
- Tourniquet
- Sterile 2×2 gauze
- Appropriate-sized Tegaderm transparent dressing
- Adhesive tape
- IV bag with solution set (tubing flushed and ready) or saline lock
- Sharps container

D.N. Smith, MD
Department of Pediatrics, University of Alabama at Birmingham,
Children's of Alabama, Birmingham, AL, USA
e-mail: dsmith@peds.uab.edu

J.K. Lucas, MD (✉)
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_120

Fig. 120.1 24-Gauge angiocatheter (*yellow*), 18-gauge angiocatheter (*red*), 20-gauge angiocatheter (*blue*)

120.4 Procedure

1. Comfortably position the patient with the site exposed.
2. Assemble the equipment and don a pair of (nonlatex) examination gloves.
3. Apply the tourniquet to the extremity above the site to be catheterized (Fig. 120.2).
4. Visualize and palpate the vein.
5. Cleanse the site with a disinfectant swab using an expanding circular motion.
6. Prepare and inspect the catheter and flush the tubing; be certain that the stylet and catheter separate easily, then fit again into the notch, aligning the bevel with the hub.
7. Stabilize the vein and apply countertension to the skin, being careful not to touch the cleansed area.
8. Insert the stylet through the skin and then reduce the angle while advancing through the vein (Fig. 120.3).
9. Observe for "flashback" as blood slowly fills the flashback chamber.
10. Advance the needle approximately 1–2 mm further into the vein, depending on the gauge and age of the patient, to ensure that the catheter is within the vein.
11. Slowly advance the catheter into the vein while keeping tension on the vein and skin (Fig. 120.4).
12. While advancing the catheter, be certain to hold the stylet portion with the thumb and forefinger, so as to avoid advancing the needle portion into and through the opposite side of the vessel, thus "blowing" the vein.
13. When the catheter is advanced about halfway, slowly withdraw and remove the stylet while simultaneously continuing to advance the catheter to its hub.
14. Attach a 3-mL non-Luer-Lok syringe to the hub.
15. Remove the tourniquet.
16. Gently attempt to aspirate blood. The blood should be free flowing.
17. Secure the catheter by either placing a transparent occlusive dressing (e.g., Tegaderm) over the lower half of the catheter hub or taping over the catheter hub in a cruciate

fashion, taking care not to cover the IV tubing connection (Fig. 120.5).

18. Remove the cover from the end of the IV tubing and insert the IV tubing into the hub of the catheter (the tubing must have been flushed with IV solution before connecting with the catheter hub: the unit from the solution bag/bottle through the catheter must be air free).
19. Open up the IV roller clamp and observe for drips forming in the drip chamber.
20. Place a piece of tape over the catheter hub then make a small (kink-free) loop in the IV tubing and place a second piece of tape over the first piece to secure the loop (Fig. 120.6).
21. Place a third piece of tape over the IV tubing above the site.
22. Ensure that the IV is properly secured and infusing properly.
23. Ensure that all "sharps" are placed in the sharps container.

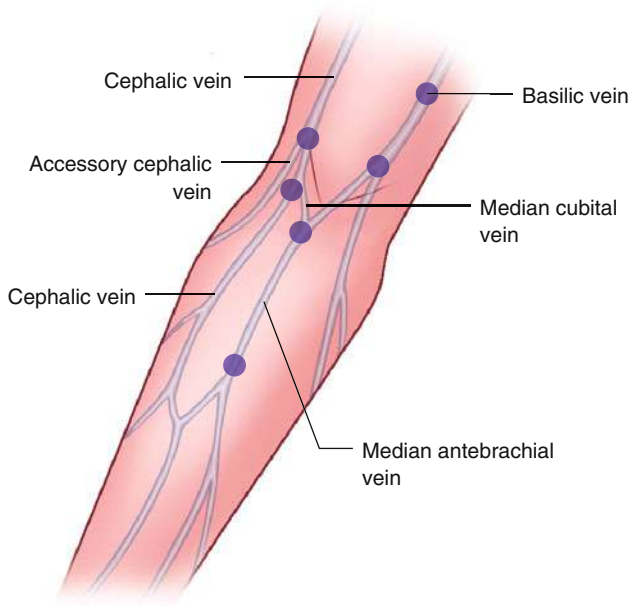

Fig. 120.2 Anatomy of the volar surface antecubital fossa and forearm

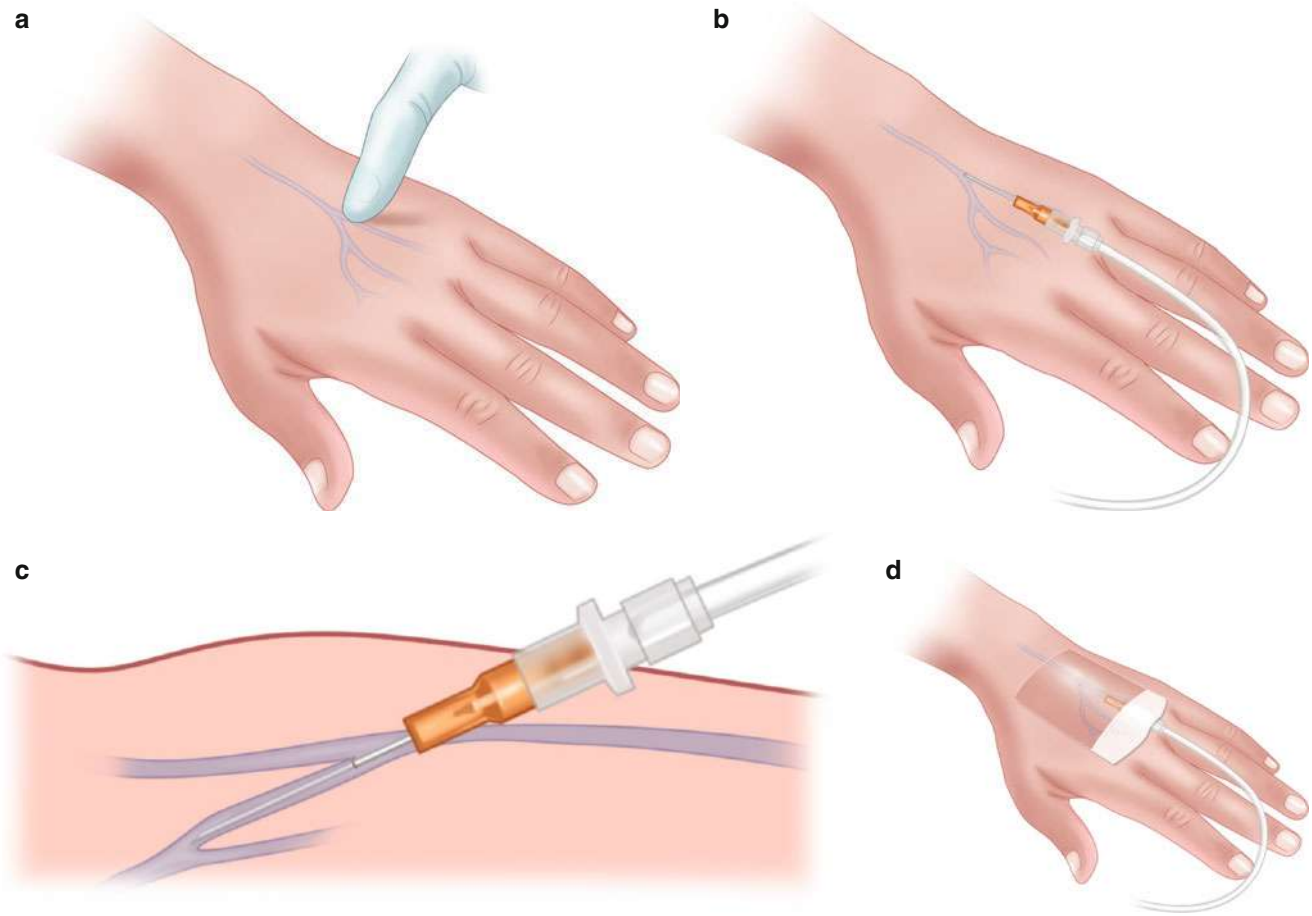

Fig. 120.3 (**a**–**d**) After applying the tourniquet, palpate a vein, as straight as possible, and ideally without many "knots" (i.e., valves)

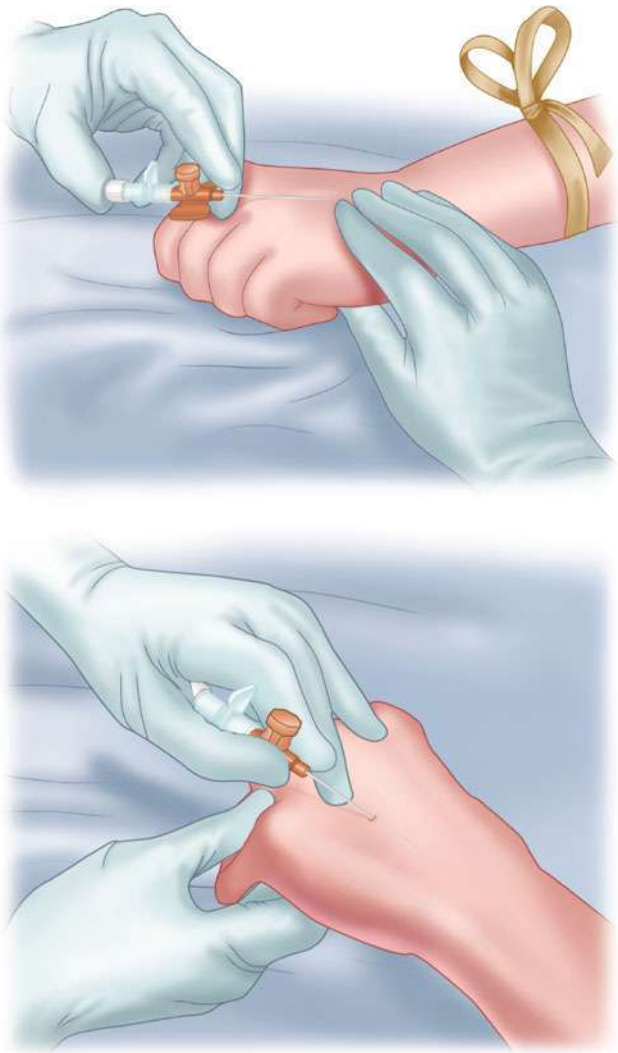

Fig. 120.4 With the tourniquet "up," and the vein distended, apply traction to the skin, pierce the skin, and pass the catheter tip (into the vein) until blood return is noted in the catheter hub

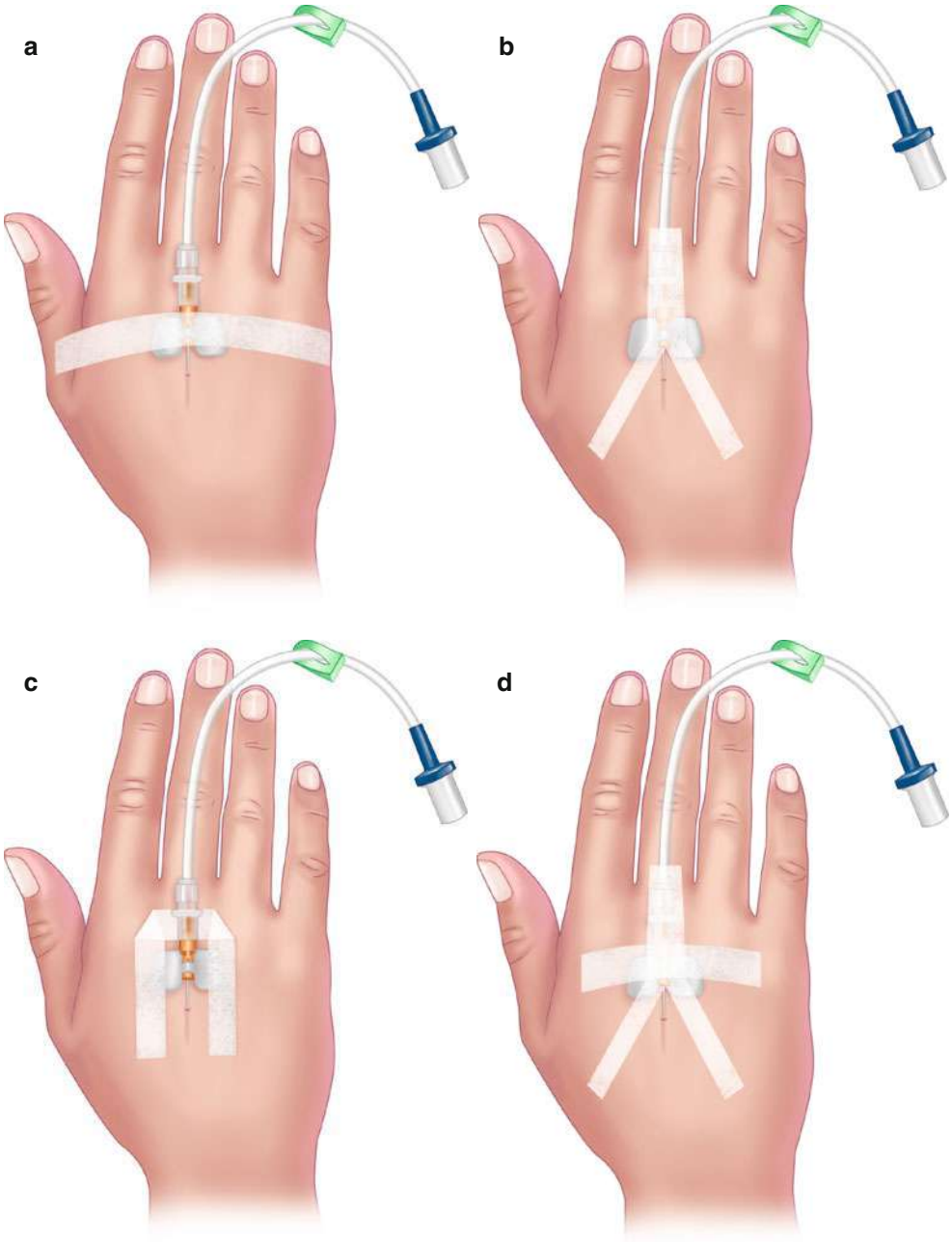

Fig. 120.5 (**a**–**d**) Securing the intravenous line utilizing the cruciate taping style

Fig. 120.5 (continued)

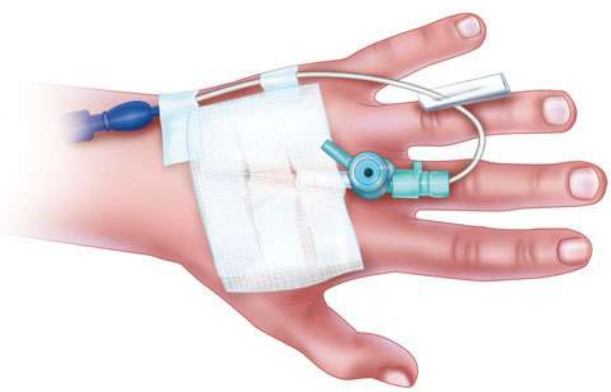

Fig. 120.6 Commercially available hub stabilizer, minimizing the need for excessive tape

120.5 Pearls and Pitfalls

- Pearls
  - Start catheter attempts distal in the extremities and move proximally with each subsequent attempt.
  - The use of ultrasound or a light source in infants can aid in location of the vessel and placement of the line.
  - In an emergent situation, in which fluids or medications are needed quickly, intraosseous access (see Chap. 122) can be obtained if venous catheterization fails.
- Pitfalls
  - The use of lidocaine subcutaneously can improve patient comfort, but it does disrupt anatomical landmarks.

Selected Reading

Bailey P. Vascular (venous) access for pediatric resuscitation and other pediatric emergencies. UpToDate. http://www.uptodate.com/contents/vascular-venous-access-for-pediatric-resuscitation-and-other-pediatric-emergencies. Accessed 23 June 2014.

Department of Emergency Medicine, University of Ottawa. Peripheral intravenous access. 2003. http://www.med.uottawa.ca/procedures/iv/. Accessed 23 June 2014.

Kost S. Ultrasound-guided vascular (venous) access. UpToDate. http://www.uptodate.com/contents/principles-of-ultrasound-guided-venous-access. Accessed 23 June 2014.

Nursing Resource Administration. Medical procedure: insertion of peripheral IV line. http://nursing-resource.com/iv-insertion/. Accessed 23 June 2014.

Umbilical Venous Catheters (Insertion and Removal)

121

Judith K. Lucas

121.1 Indications

- Temporary vascular access for infants up to roughly 10 days of life (between 7 and 14 days) with shock or cardiopulmonary failure
- Emergency vascular access in this age group, when peripheral intravenous (IV) access cannot be rapidly obtained
- Preferred vascular access in infants less than 1000 g

121.2 Contraindications

- Omphalitis
- Omphalocele
- Necrotizing enterocolitis
- Peritonitis

121.3 Materials and Medications

- Anesthetic: not necessary; procedure is painless
- Soft ties to restrain infant's extremities
- Sterile gloves/gowns
- Antiseptic solution
- Sterile towels/drapes
- 3.5- (infants <1500 g) or 5-French (>1500 g) umbilical venous catheter
- 5-French feeding tube
- Three-way stopcock
- 10-mL syringe with heparinized saline flush (1 U/mL)
- Umbilical tape or 3-0 silk on a cutting needle
- Non-toothed forceps
- Small hemostats (2)
- #11 scalpel and blade
- Scissors
- Graph depicting length of catheter insertion, if placing umbilical venous catheter (UVC) above the diaphragm in either very small infants or infants for whom measurement of central venous pressure (CVP) is indicated (Fig. 121.1)

J.K. Lucas, MD
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_121

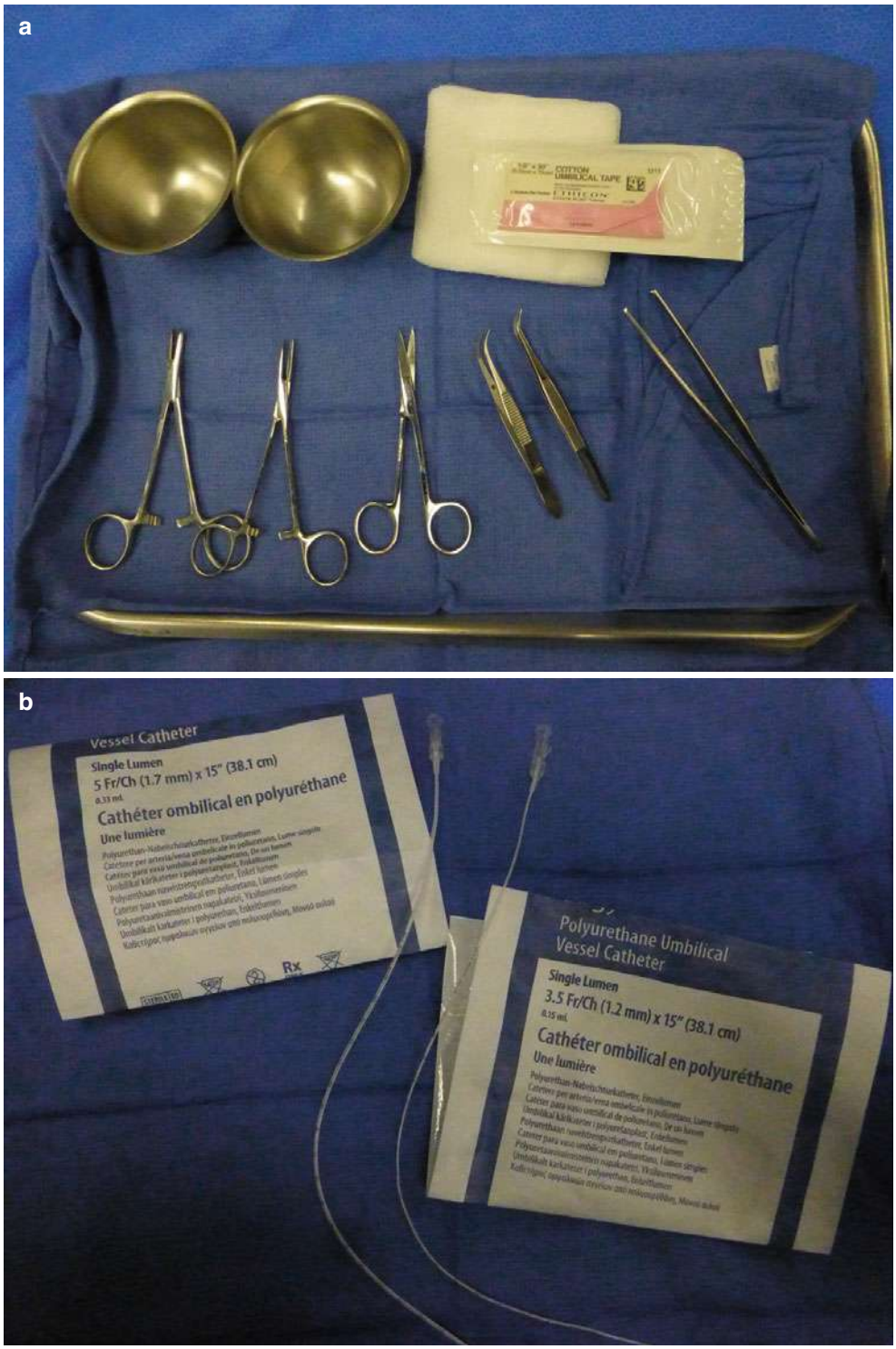

Fig. 121.1 (**a**) Instrumentation suggested for umbilical venous catheter placement. (**b**) Umbilical venous catheters (in an emergency, 5 F feeding tube is an acceptable alternative)

121.4 Procedure (Insertion)

1. Place the infant under a radiant warmer.
2. Using soft ties, restrain the infant's extremities.
3. Scrub the umbilicus and surrounding abdomen with antiseptic solution.
4. Drape the umbilicus and area in sterile manner (leave the infant's head exposed).
5. Tie a loose loop with the umbilical tape around the base of the cord OR run the 3-0 silk through the skin *of the cord* in a purse-string fashion.
   - This will be used later to anchor the line after placement and to provide hemostasis should the line accidentally be pulled out and bleeding ensue.
6. Using the scalpel blade, cut the umbilical cord horizontally approximately 2 cm above the junction between the cord and the skin.
7. Identify the umbilical vessels
   - The vein is thinner walled, larger in diameter, and somewhat floppy appearing, relative to the umbilical arteries, and typically located at the 12 o'clock position.
   - The arteries are smaller, thick walled, and paired (a single umbilical artery often signifies the presence of a congenital malformation/syndrome) and located at the 4 and 8 o'clock positions (Fig. 121.2).
8. Place the stopcock on the receiving end of the umbilical catheter or the 5-French feeding tube and flush with heparinized saline solution, and then close the stopcock.
   - It is imperative that there is no air in the catheter.
9. Introduce a closed smooth-surfaced forceps into the lumen of the umbilical vein and allow the forceps to separate, allowing the vein to dilate (Fig. 121.3).
10. Insert the catheter (or feeding tube) into the lumen and gently advance, directing the catheter toward the right shoulder (Fig. 121.4).
11. Advance the catheter only until good blood flow is noted and then another 1 to 2 cm (this should be a total of only 4–5 cm in a term infant).
    - *Do not force the advancement.*
    - At this level the tip of the catheter should still be inferior to the liver (Fig. 121.5).
12. Tighten the umbilical tie or the purse-string suture.
13. Secure the catheter with a tape bridge (Figs. 121.6, 121.7, and 121.8).
14. Although if placed only up to point of blood return, the catheter tip should be below the liver; it is best to get an abdominal x-ray to confirm; many solutions are caustic to the liver and can result in complications.

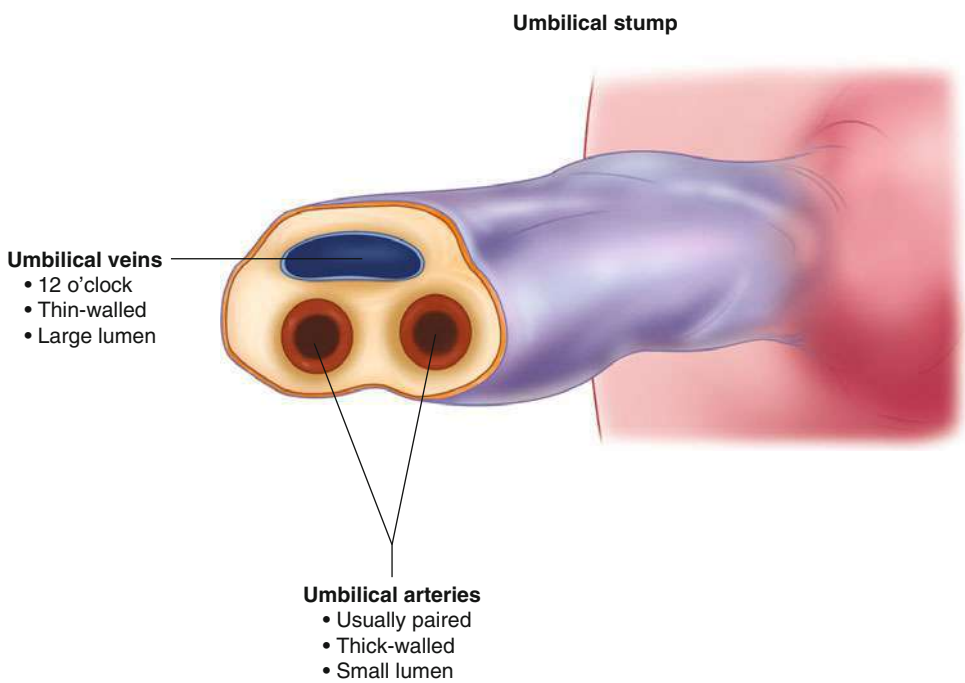

Fig. 121.2 Anatomy of the umbilical cord when cut transversely approximately 2 cm from abdominal wall

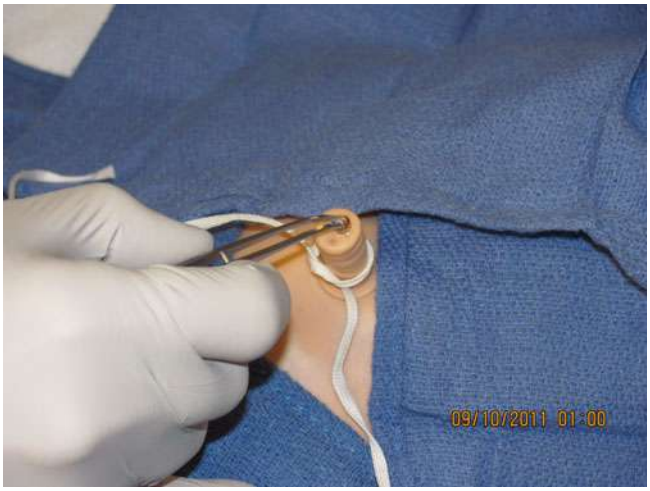

Fig. 121.3 Sterilely draped umbilicus, with umbilical tape loosely tied at base of umbilical cord. This tie will be cinched and secured once the catheter is placed, but meanwhile, can assist with homeostasis should the catheter be inadvertently dislodged during placement. Gently dilate the umbilical vein with a smooth toothed forceps. Insert the distal most couple of mm of a closed forceps into umbilical vein and relax, so the forceps tips smoothly separate

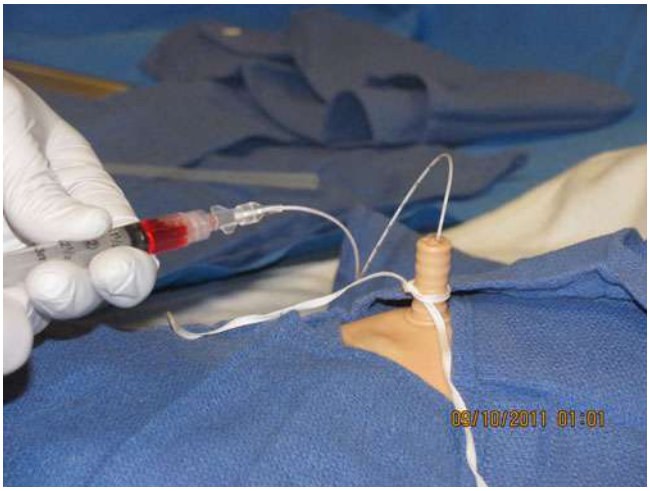

Fig. 121.5 Once there is easy blood return, pass the catheter an additional 1–2 cm. The catheter should still be inferior to the liver at this point

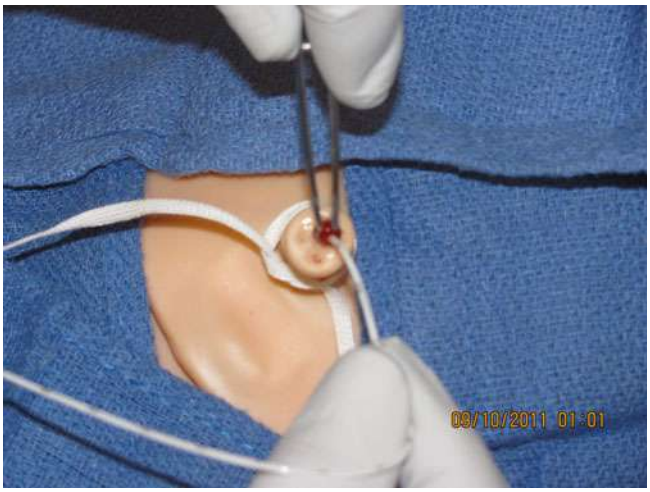

Fig. 121.4 Gently pass the umbilical venous catheter (or 5 F feeding tube) until there is easy blood return

Fig. 121.6 Creating an umbilical catheter tape bridge: the uprights

Fig. 121.7 Umbilical tape catheter bridge: the "cross bar" secures the catheter without applying tape to the umbilical stump

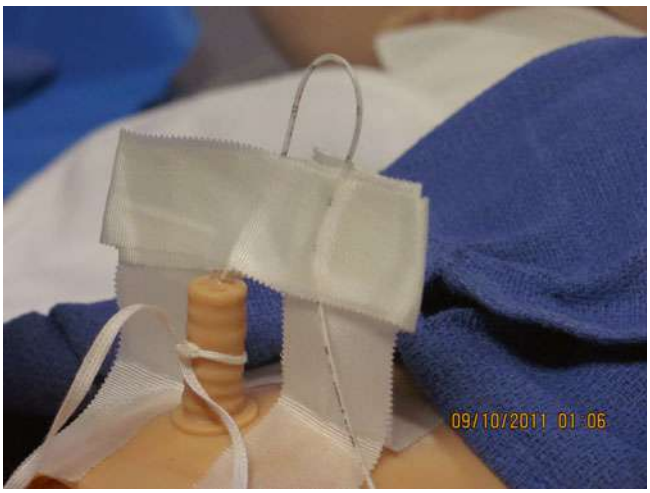

Fig. 121.8 Fold the catheter over to create a "U." Be careful not to kink the catheter, and secure with a second taped cross bar to lend additional security to the catheter and prevent displacement

121.5 Procedure (Removal)

1. The UVC should be removed as soon as adequate peripheral venous access is obtained (unless in an infant estimated to weigh <1000 g).
2. Turn infusions off.
3. Be certain the stopcock is closed to the infant.
4. It is imperative that there be no air in the catheter before withdrawal (if air is present and infant takes inspiration, the negative pressure generated can pull significant amount of air into the central vasculature).
5. Remove the securing tape from the infant.
6. Withdraw the catheter gradually as a single maneuver.

121.6 Complications

- Infection
- Bleeding due to disconnection of tubing (*always use Luer-Lok connections*) or perforation of vessels
- Arterial injury by accidental perforation
- Hepatic injury and necrosis if the catheter sits within a portal vein
- Thrombosis
- Air embolus
- Dysrhythmia or pericardial tamponade or perforation if catheter is advanced too far

121.7 Pearls and Pitfalls

- Pearls
  - The umbilical vein is 2–3 cm long before it widens into the umbilical recess, just before intersecting with the left portal vein and the ductus venosus.
  - Be certain to include the length of the umbilical stump in any calculations for placement.
  - If using the calculating graph, measure from the right shoulder to the umbilicus (Fig. 121.9).
- If choosing the above-diaphragm site for placement, cork-screwing the catheter clockwise while passing the catheter will encourage it to pass through the ductus venosus.
  - A kidney, ureter, and bladder study is mandatory in high UVC placement.
- Pitfalls
  - When preparing with Betadine (povidone-iodine), be certain to wipe away and remove any pooled Betadine along the infant's side because this will cool the infant initially; as the Betadine is warmed, it becomes highly irritating to newborn skin.

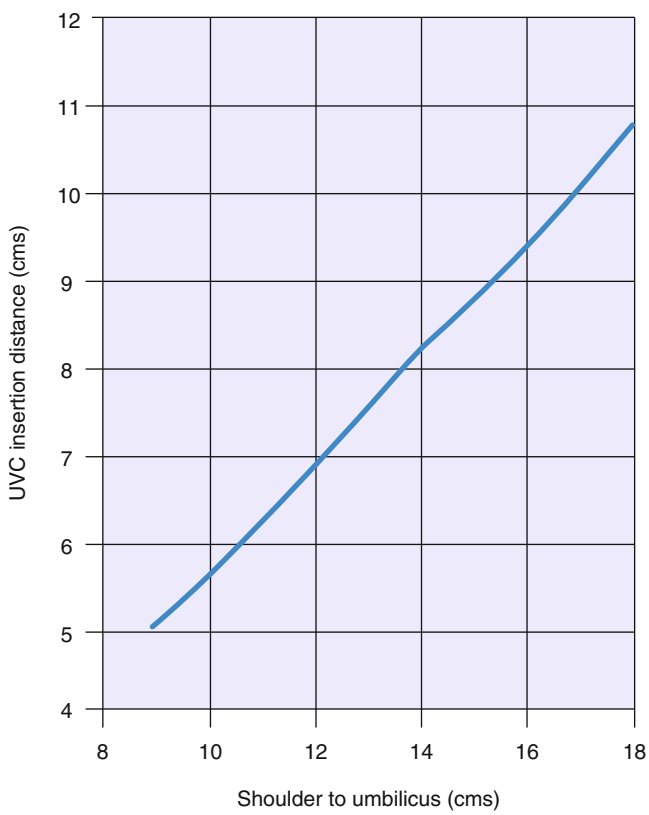

Fig. 121.9 Graph for estimation of insertion distance of the catheter; be certain to add the length (cm) of the umbilical stump to the insertion distance

Selected Reading

Sudbury, Jones, Bartlett. Emergency vascular access. In: APLS: the pediatric emergency medicine resource. 5th ed., p. 741.

Magnan JP. Umbilical Vein Catheterization. http://emedicine.medscape.com/article/80469-overview.

Schlesinger AE, Braverman RM, DiPietre MA. Neonates and umbilical venous catheters: normal appearance, anomalous positions, complications, and potential aid to diagnosis. Presented at the 2001 annual meeting of the American Roentgen Ray Society, Seattle, Apr 2001.

Intraosseous Access

122

Judith K. Lucas

122.1 Indications

- During cardiac arrest: Failure to attain vascular access after three peripheral intravenous attempts or 90 s, whichever comes first
- Inability to gain vascular access in pediatric patients presenting in shock due to hemorrhage (trauma), sepsis, profound dehydration, or cardiac failure

122.2 Contraindications

- Absolute
  - Fracture of the long bone considered for intraosseous (IO) access
- Relative
  - Previous IO access attempt in the same long bone
  - Cellulitis over the insertion site
  - Inferior vena cava injury (circulatory access proximal to the injury site is preferred).
  - Osteogenesis imperfecta

122.3 Materials and Medications

- Preparation materials, such as an antiseptic solution and sterile drapes, IF the patient's stability offers the time.
- Lidocaine without epinephrine: If it becomes necessary to place an IO in a conscious patient, use lidocaine to anesthetize the skin to the bony cortex; then once accessed, infiltrate 2–3 mL into the marrow to alleviate some of the pain of medications infusing through the marrow.
- IO needle (a few examples shown) (Figs. 122.1, 122.2, and 122.3).

Fig. 122.1 Jamshidi disposable sternal/iliac aspiration needle (Jamshidi, Cardinal Health Dublin, OH)

J.K. Lucas, MD
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_122

Fig. 122.2 Cook intraosseous needle (Cook Critical Care, Bloomington, IN)

Fig. 122.3 EZ-IO (Vida-Care, San Antonio, TX)

122.4 Procedure

- Most common site: proximal tibia
  1. Patient should be supine, with the intended leg slightly externally rotated and flexed at the hip. Flex the knee about 90°.
  2. Place a towel underneath the knee and the proximal lower leg.
  3. Palpate the tibial tuberosity. Then move fingers 2 cm distal to the tuberosity (1–2 fingerbreadths) and 2 cm medial. This area is consistently flat and is distal to the growth plate (Figs. 122.4 and 122.5).
  4. If the patient is stable, create a sterile field focused on the insertion site.
  5. Insertion site should be cleaned with an antiseptic solution.
  6. Pass the needle through the skin and subcutaneous tissue to the cortex.
  7. Once bone is reached, stabilize the IO with the thumb and first fingers adjacent to where the needle penetrates the skin.
  8. Apply steady and firm pressure downward and with a twisting motion (back and forth) with the palm on the end of the device. Although inserting the needle perpendicularly to the bone is totally acceptable, ideally angling the needle caudally slightly (~15°) will avoid the growth plate (Fig. 122.6).
  9. As the cortex is fully penetrated indicating entrance into the marrow space, one *may* feel an ease in resistance. Do not advance the needle further. The needle should be able to stand up without support.
  10. Remove the inner trocar (Fig. 122.7).
  11. Attempt to aspirate marrow/blood. Not being able to do so DOES NOT mean inadequate or inaccurate placement. Instead, if marrow cannot be aspirated, affix a 10-mL syringe with normal saline and attempt to infiltrate. Resistance to flow should be minimal, and one should not appreciate either extravasation around the insertion site, coolness, or tissue expansion posterior to the site (indicating the fluid is passing through the IO).

12. Once placement is confirmed, connect an intravenous line, using a three-way stop-cock.
13. Secure the needle with tape and gauze (Fig. 122.8).

- Alternative sites
  - Distal tibia
    - The landmarks are the medial aspect of the tibia (the flat portion), 2 fingerbreadths proximal to the medial malleolus.
    - Again, externally rotate and abduct the hips, with the knee flexed about 60° (as with the proximal tibia).
    - Angle the needle toward the knee (cephalad) about 10–15° to avoid the distal tibia growth plate.
    - The remainder of the IO insertion is identical to the procedure for proximal tibia placement.
  - Distal femur
    - Slightly flex and externally rotate the hip.
    - Flex the knee enough that the quadriceps muscle group is relaxed.
    - The landmarks for the distal femur are the anterior thigh, midline, about 3 fingerbreadths proximal to the medial and lateral condyles.
    - The IO is inserted perpendicular to the bone (because there is no growth plate in the distal femur other than the condyles).
    - The remainder of the procedure is identical to the procedure for placement in the proximal tibia.
  - Proximal humerus (Fig. 122.9)
    - The patient should be supine, with the shoulder, upper arm, and elbow as close to the body as possible, yet still on the bed. The elbow should be flexed at 90°, with the forearm and palm resting on the patient's abdomen.
    - The provider slides his or her thumb up the anterior shaft of the humerus toward the shoulder until the greater humerus tubercle is palpated, which identifies the surgical neck of the humerus.
    - The insertion site is perpendicular to the humerus, approximately 1–2 fingerbreadths proximal to the tubercle.
    - The remainder of the insertion is identical to that for the proximal tibia.
  - Sternum (not usually recommended for small children).
    - Requires a special IO needle and system.
    - Method of placement is specific to the insertion system chosen.
    - Manubrium is the desired site (as opposed to the body of the sternum, which would interfere with cardiopulmonary resuscitation).
    - Risks specific to sternal placement include pneumothorax, mediastinitis, and great vessel injury.

Fig. 122.4 Step 1. Identify the tibial tuberosity

Fig. 122.5 Step 2. One to 2 fingerbreadths below the tuberosity and medially, to the flat aspect of the proximal tibia

Fig. 122.6 Step 3. Angle the device slightly caudal to avoid growth plate

Fig. 122.7 Step 4. Uncap the top of the device and remove the trocar

Fig. 122.8 Step 5. Stabilize the intraosseous cannula

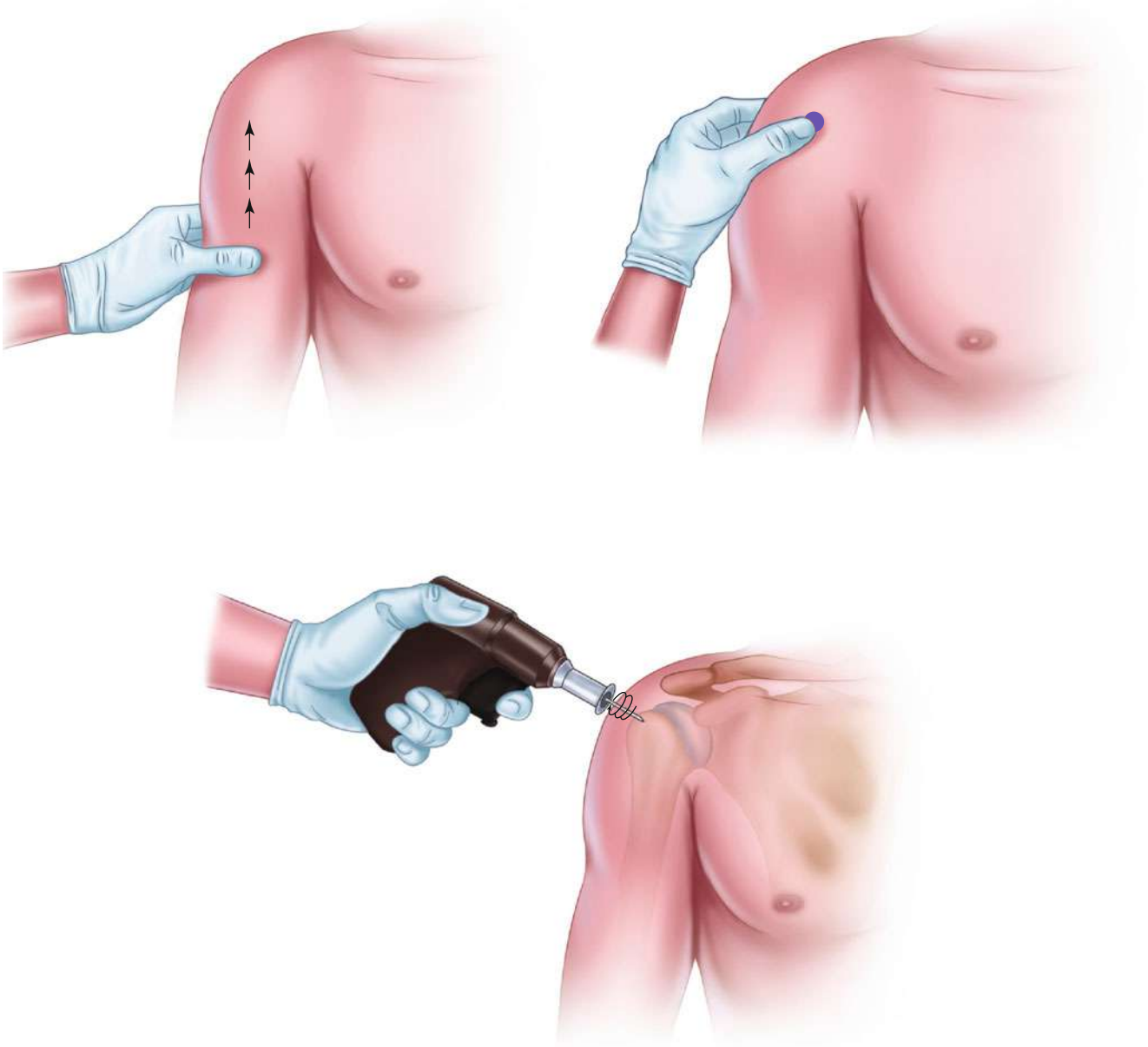

Fig. 122.9 IO insertion into proximal humerus site

122.5 Complications

- Extravasation of fluid.
  - Occurs as a result of a misplaced IO, either not completely in the marrow space anteriorly or through the cortex posteriorly.
  - Tissue necrosis.
- Compartment syndrome if extravasation is not recognized.
- Fracture and growth plate injury.
- Infection and osteomyelitis are rare complications if sterile technique is used.
- Fat embolism, considered rare and only in adult patients.

122.6 Pearls and Pitfalls

- Be certain to avoid placing a hand beneath the IO site during placement in order to prevent possible IO penetration into the provider's tissue.
- Do not rely on the sensation of a "pop" to determine appropriate penetration into the cortex, especially in infants.
- Likewise, one may not be able to aspirate bone marrow or blood, even with an appropriately placed IO. DO NOT pull out the IO if marrow cannot be aspirated. Rather, gently but firmly infiltrate 10 mL normal saline.
- Blood gases, body chemistries, and blood typing can be obtained from IO samples, but the sample CANNOT be used for hematocrit determination.

Selected Reading

Bohn D. Intraosseous vascular access: from archives to the ABC. Crit Care Med. 1999;27:1053.

DeCaen AR, Reis A, Bhutta A. Vascular access and drug therapy in pediatric resuscitation. Pediatr Clin North Am. 2008;55:909–27.

Deitch K. Intraosseous infusion. In: Roberts JR, Hedges JR, editors. Clinical procedures in emergency medicine. 5th ed. Philadelphia: Saunders; 2009. Chap. 25.

EMS World. Intraosseous infusion: not just for kids anymore. Posted Jan. 12, 2011. Updated from Mar 2005. EMSWorld.com. Cygnus Business Media Site.

Halm B, Yamamoto LG. Comparing ease of intraosseous needle placement: Jamshidi versus Cook. Am J Emerg Med. 1998;16:420.

123 Lumbar Puncture in Pediatrics

Maritza A. Plaza-Verduin and Judith K. Lucas

123.1 Indications

- Evaluation of cerebrospinal fluid (CSF) for infection or malignancy
- Measurement of opening pressure
- Treatment of pseudotumor cerebri
- Diagnosis of central nervous system (CNS) metastases
- Instillation of intrathecal chemotherapy
- Injection of radiopaque dye for spinal cord imaging

123.2 Contraindications

- Increased intracranial pressure
- Bleeding diathesis (platelet count <50,000)
- Overlying skin infection near the area of puncture site
- Spinal cord trauma or spinal cord compression
- Signs of progressive cerebral herniation
- Condition of the patient (e.g., unstable airway, potentially dangerous breathing problem, severe circulatory instability) that could cause an abrupt decompensation
- Known spinal cord deformity

123.3 Materials and Medications

- Lumbar puncture (LP) tray (Fig. 123.1)
  - Sterile drapes
  - Betadine (povidone-iodine) swabs or tray to pour Betadine
  - Sterile sponges for preparing the puncture site
  - Sterile 3-mL syringe with needle for lidocaine injection
  - Sterile collecting tubes (4)
  - Sterile spinal needle with stylet (size depending on age of patient)
    - Premature infant: 22 gauge or smaller, 1.5 inch
    - Neonate to 2 years: 22 gauge, 1.5 inch
    - 2–12 years: 22 gauge, 2.5 inch
    - Older than 12 years: 20 or 22 gauge, 3.5 inch
  - Pressure manometer column with a three-way stopcock
- Betadine solution
- Sterile gloves
- Mask
- Lidocaine (1–2 % without epinephrine)
- 4 % Lidocaine cream (LMX-4) or lidocaine and prilocaine mixture (EMLA [eutectic mixture of local anesthetics])

M.A. Plaza-Verduin, MD
Pediatric Division, Department of Emergency Medicine, Arnold Palmer Hospital for Children, Orlando, FL, USA
e-mail: mari.plazav@gmail.com

J.K. Lucas, MD (✉)
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_123

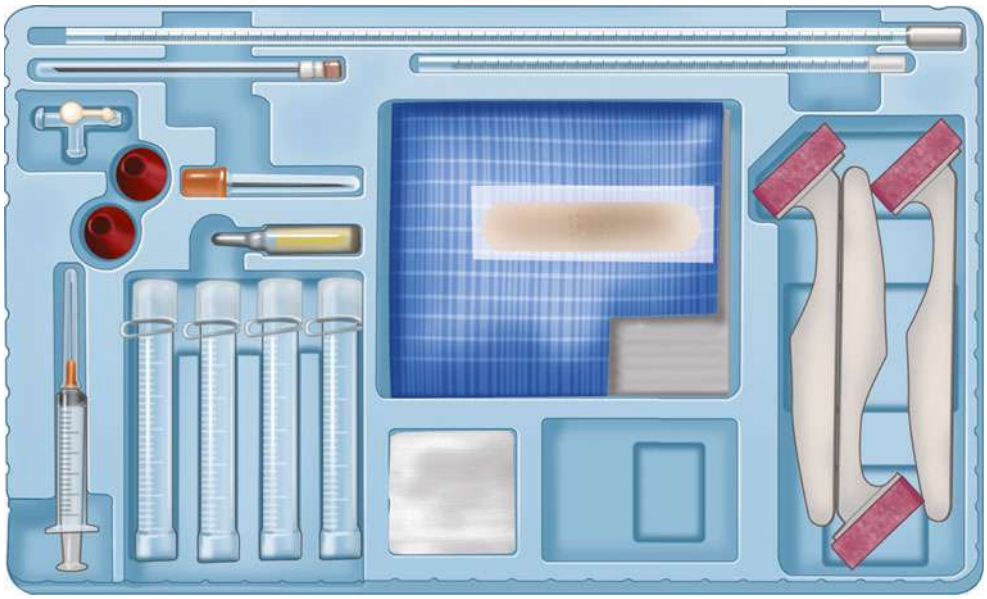

Fig. 123.1 Lumbar puncture (LP) tray

123.4 Procedure

1. Position the child (Figs. 123.2 and 123.3).
   (a) Can be positioned in either the sitting or the lateral recumbent position with the hips, knees, and neck flexed.
   (b) For small infants or any patient with any degree of cardiorespiratory compromise, keep close monitoring of heart rate, respirations, and oxygen saturation while in the flexed position.
2. Palpate the top of the iliac crest and draw an imaginary line connecting the two across the back, which should cross the midline just above the fourth lumbar spine (Fig. 123.4).
3. Palpate L3–4 or L4–5 space along this line.
4. Place EMLA or LMX-4 on the area at this time and allow some time for anesthesia to occur (can take up to 30 min to be effective).
5. Prepare the skin in sterile fashion with Betadine solution using enlarging circles that begin at the puncture site.
6. Drape the patient with sterile towels, exposing the puncture site.
   (a) If an infant, do so conservatively to be able to monitor the infant during the procedure.
7. Locate the intervertebral space (L3–4 or L4–5) once again.
   (a) Make a small mark on the chosen intervertebral space with a fingernail depression or the plastic cap of the spinal needle.
8. If desired, or if anesthetic cream was not previously used or more is needed, apply a small wheal of lidocaine at the desired puncture site using a 25-gauge needle.
9. Insert the spinal needle in the intervertebral space.
   (a) Puncture the skin in the midline just caudal to the palpated spinous process.
   (b) Bevel should be positioned so that the dura mater is pierced parallel to its fibers (which will reduce the likelihood of CSF leakage).
      (i) If in the lateral recumbent position, the bevel of the needle should be positioned horizontally.
      (ii) If in the sitting position, the bevel of the needle should be positioned vertically.
      (iii) Angle the needle slightly cephalad toward the umbilicus and parallel to the bed if the patient is in the lateral recumbent position or slightly caudal (perpendicular to the skin) if the patient is in the sitting position.
10. Advance the needle several millimeters at a time and withdraw the stylet frequently to check for CSF flow.
   (a) Advance the needle until a loss of resistance is felt or approximately 1–2 cm.
      (i) In the infants, whose dura are not so thick and a "pop" or give may be unnoticeable, after passing between the spiny processes and approximately 2 cm through the skin, remove the stylet frequently to check for CSF return. This will allow you to avoid passing through the subarachnoid space.
      (ii) After the change in resistance occurs, if resistance is met again, pull back gently on the needle to reposition in the subarachnoid space and remove the stylet to check for CSF fluid.
11. Once CSF is free flowing, attach a pressure manometer to the needle hub via the three-way stopcock (Fig. 123.5).
   (a) Make sure to hold the spinal needle with one hand in place while attaching the manometer to prevent movement of the needle.
   (b) Measure the CSF pressure once the CSF reaches the highest level in the manometer column.
   (c) CSF pressures will be best obtained in the lateral recumbent position with the neck and legs extended.
   (d) Normal CSF pressure is 5–20-cm H_2O with the neck and legs extended, 10–28-cm H_2O with the neck and legs flexed.
   (e) Have an assistant hold the top of the pressure manometer while the spinal needle is supported at the connection of the manometer and stopcock.
12. Remove the manometer and collect the CSF into sterile tubes.
   (a) Make sure to keep the needle in place when removing the manometer.
   (b) Continue to drain fluid into the collecting tubes, approximately 1 mL per tube.
   (c) The tubes should be labeled for specific studies, depending on the order of collection.
      (i) First tube for gram stain and culture.
      (ii) Second tube for quantitative glucose and protein.
      (iii) Third or fourth tube for cell count and differential.
      (iv) Leftover tube for any additional studies that may be needed.
      (v) When the LP is to assess for possible subarachnoid blood, the first and fourth tubes are sent for cell count.
13. Replace the stylet and remove the spinal needle.
14. Cleanse the puncture area to avoid staining with the Betadine solution.
15. Apply a sterile dressing to the puncture site.
16. Patients 4 years and older should remain in the supine position, with the head elevated no more than the height of a pillow (for comfort), for at least an hour to avoid spinal tap headaches.
17. Likewise, after the LP, giving a fluid bolus of normal saline (10–20 mL/kg in children; 1 L in non-volume-sensitive adults) will also assist with avoidance of spinal tap headaches.

Fig. 123.2 Curling up for an LP

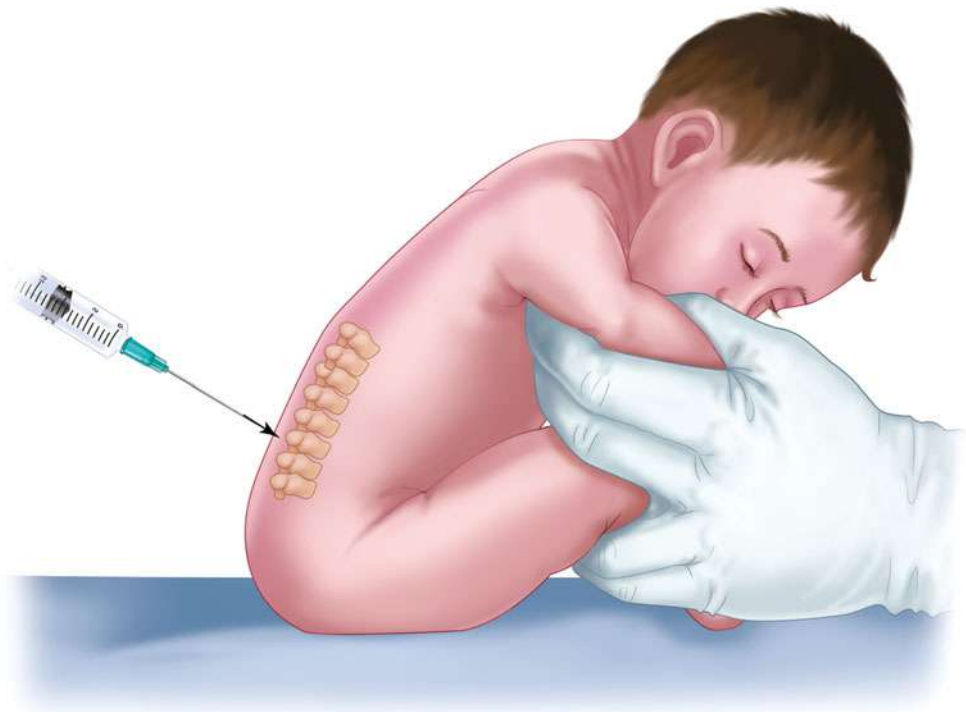

Fig. 123.3 Having the LP

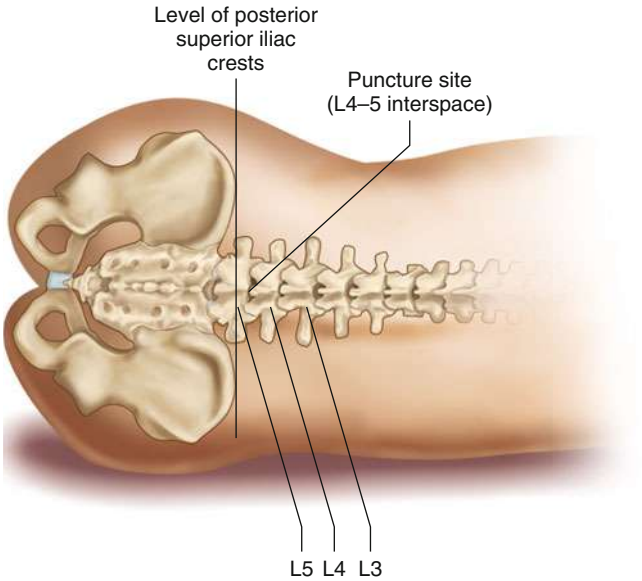

Fig. 123.4 Anatomy of the lumbar spine showing the sites for dural puncture

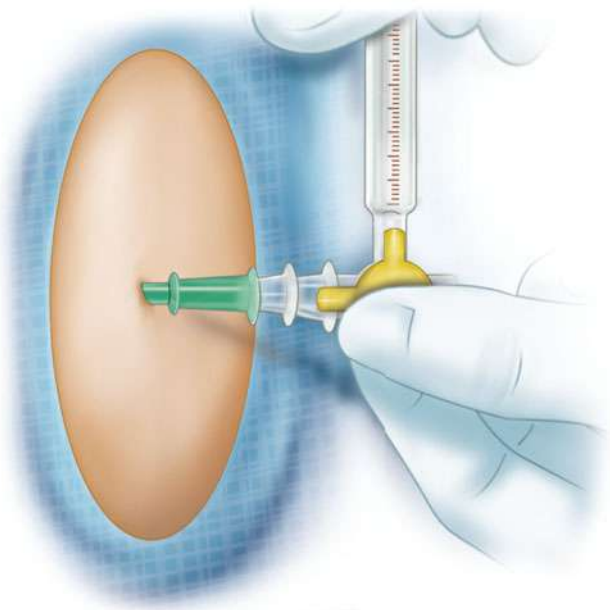

Fig. 123.5 Pressure manometer attached to the needle hub via the three-way stopcock

123.5 Complications

- Minor
  - Localized back pain
  - Transient paresthesia during procedure
  - Post-LP headache
- Major
  - Severe back pain associated with neurological signs (may be subdural or epidural spinal hematoma)
  - LP-induced meningitis
  - Cerebral herniation
  - Acquired epidermoid tumor
  - Damage to adjacent structures (disk herniation, retroperitoneal abscess, spinal cord hematoma)

123.6 Pearls and Pitfalls

- Pearls
  - Success of the LP depends on the positioning of the patient.
    - Goal of positioning is to stretch the ligamentum flavum and increase the interlaminar spaces.
    - In the recumbent position, the shoulder and hips should be perpendicular to the bed, keeping the spinal cord straight, with no rotation.
    - Sitting position is useful with older, cooperative patients or with very young infants who are unlikely to struggle and may have increased respiratory distress in the lateral recumbent position.
  - Administration of an anxiolytic (e.g., midazolam) can be used to facilitate the procedure in an older child. In some cases, procedural sedation may be required.
  - Placing the tip of the thumb on the spinous process just above the space being entered can ensure good alignment of the needle.
  - In older patients a "pop" may often be felt as a change in resistance occurs once the dura is penetrated; however, in infants and neonates, that "pop" may be exceedingly subtle or not palpable.
  - If no CSF returns, attempt the following options:
    - Ensure the needle is in the appropriate position, withdrawing the needle slowly if necessary.
    - Rotate the needle 90°.
    - If an infant, have the assistant massage the anterior fontanel to help facilitate CSF flow.
    - If the procedure continues without CSF, withdraw the needle to just under the skin and redirect it.
    - If the procedure continues without CSF, withdraw the needle and insert a new needle with stylet at an alternate site.

    - If these steps do not yield CSF, the infant may be dehydrated, not allowing for adequate CSF flow.
      - Give the patient a bolus and reattempt later.
      - Attempt putting the infant in a sitting position to increase flow.
- Pitfalls
  - Traumatic LP
    - Bloody CSF fluid, which usually clears as the CSF drains if the needle is in the correct space.
    - Occurs with improper technique (inserting the needle too far to one side into an epidural venous plexus or through the subarachnoid space into or adjacent to the vertebral body).
    - Can occur with proper technique as well.
    - If fluid does not clear and clots form in the tubes, LP should be reattempted at a different site.
  - Failed LP attempts despite proper procedure and positioning
    - Ultrasound can be used to visualize the area and determine the reason for failure or the likelihood of success with future attempts.

Selected Reading

Coley BD, Shiels WE, Hogan MJ. Diagnostic and interventional ultrasonography in neonatal and infant lumbar puncture. Pediatr Radiol. 2001;31:399–402.

Cronan KM, Wiley JF. Lumbar puncture. In: King C, Henretig FM, editors. Textbook of pediatric emergency procedures. 2nd ed. New York: Lippincott Williams & Wilkins; 2008.

Ebinger F, Kosel C, Pietz J, Rating D. Headache and backache after lumbar puncture in children and adolescents: a prospective study. Pediatrics. 2004;113:1588–92.

Friedman AG, Mulhern RK, Fairclough D, et al. Midazolam premedication for pediatric bone marrow aspiration and lumbar puncture. Med Pediatr Oncol. 1991;19:499–504.

Partin WR. Emergency procedures. In: Stone CK, Humphries RL, editors. Current diagnosis & treatment: emergency medicine. 6th ed. New York: McGraw-Hill; 2007.

126 Use of the Broselow Tape

Judith K. Lucas

126.1 Indications

- To determine the equipment size and the medication doses during a pediatric resuscitation, without having to take the time to perform calculations

126.2 Contraindications

- Premature infant or newborn whose heel, while the infant is fully extended, does not fall at least into the white area (corresponding to 3, 4, and 5 kg).
- The length of the child exceeds the distal end of the green area (36 kg).

126.3 Materials

- Broselow Pediatric Emergency Tape (Armstrong Medical Industries, Wilshire, IL) (Figs. 126.1 and 126.2)

Fig. 126.1 Broselow tape, proximal end (at patient's head)

Fig. 126.2 Broselow tape, distal end (the tape will accommodate a 36-kg child, at most)

J.K. Lucas, MD
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_126

126.4 Procedure (Figs. 126.3, 126.4, and 126.5)

- Place the infant or child on the bed in the supine position.
- Extend the Broselow Pediatric Emergency Tape next to the patient, placing the red line at the top of the patient's head.
- With the patient fully extended, especially through the hips and knees, align the bottom of the heel, while the ankle is flexed and toes upward, with the color on the tape adjacent to the heel.
- The color on the tape corresponds to the patient's weight.
- The side of the tape with weights noted along the bottom has medication doses for resuscitation, rapid-sequence intubation (RSI), and paralytics (Fig. 126.6).
- The top half of the other side of the tape has the appropriate dosing for managing seizures, overdoses, elevating intracranial pressure, and fluids. The bottom half of this side has the appropriate equipment sizes, such as endotracheal tubes, nasogastric tubes, chest tubes, and other equipment, that correspond to the patient's weight (Fig. 126.7).

Fig. 126.3 Placement of the child on the tape

Fig. 126.5 Placement of the child on the tape

Fig. 126.4 Placement of the child on the tape

Fig. 126.6 Front of tape. Weight groupings are noted at the bottom. This is the side with resuscitation and RSI medications

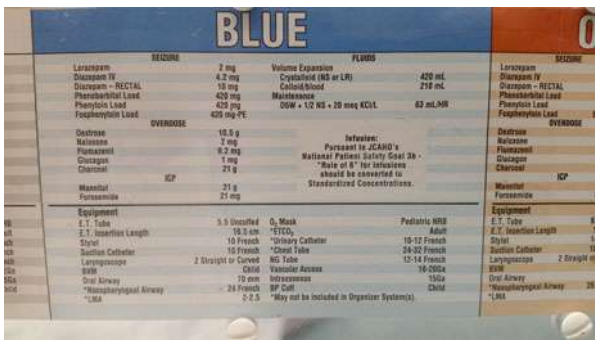

Fig. 126.7 Reverse side of tape. This side has medications for critical interventions aside from the primary survey, as well as equipment sizes

126.5 Pearls and Pitfalls

- Pearls
  - Be certain that the top of the head is at the red line and that the patient's body is fully extended. Infants tend to lay flexed at the hips and knees.
  - Determining the appropriate color code depends on noting the bottom of the heel while the ankle is flexed and the toes are pointing upward.
- Pitfalls
  - Beware of the child who is unusually heavy compared with other children of the same height.

Selected Reading

Luten R. Error and time delay in pediatric trauma resuscitation: addressing the problem with color-coded resuscitation aids. Surg Clin North Am. 2002;82:303–14, vi.

Luten R, Broselow J. Standardization of product concentration in emergency dosing: a response to Fineberg and Arendt. Ann Emerg Med. 2008;52:477–8.

Rosenberg M, Greenberger S, Rawal A, Latimer-Pierson J, Thundiyil J. Comparison of Broselow tape measurements versus estimations of pediatric weights. Am J Emerg Med. 2011;29:482–8.

Nursemaid's Elbow

127

Judith K. Lucas

127.1 Definition

- An injury commonly seen in children between 6 months and late preteens, although generally seen between 1 and 3 years
- A subluxation of the radial head, usually resulting from sudden, longitudinal traction on an extended arm with the wrist pronated
- Often occurs when a parent/caregiver is holding the child by the hand while walking and suddenly pulls the child away from a dangerous situation (Fig. 127.1)

Fig. 127.1 Note that the toddler prefers to hold her right arm pronated and somewhat flexed at the elbow. When asked where it hurts, she may as often point to her wrist as to her elbow. There will be no soft tissue swelling anywhere along the upper extremity

J.K. Lucas, MD
Department of Emergency Medicine,
University of Florida Health Shands Hospital,
Gainesville, FL, USA
e-mail: judithklucas@ufl.edu

© Springer Science+Business Media New York 2016
L. Ganti (ed.), *Atlas of Emergency Medicine Procedures*, DOI 10.1007/978-1-4939-2507-0_127

127.2 Clinical Diagnosis

- History of pulling-type injury.
- Child presents with arm held slightly flexed at elbow and pronated.
- Patient may point to elbow or wrist as source of pain, but both areas are without any swelling and are nontender on palpation.
- Elbow can be flexed and extended, but the forearm cannot be supinated.
- Radiography is not helpful.

127.3 Indications

- Clinical presence of subluxed radial head

127.4 Contraindications

- Absolute
  - Radiographic evidence of elbow or forearm fracture
  - Swelling or pain about the elbow, forearm, or wrist
- Relative
  - Unknown history or witnessing of pulling-type injury

127.5 Materials and Medications

- None

127.6 Procedure

127.6.1 Method A: Superpronation (preferred method) (Fig. 127.2)

1. Explain to the caretaker that the reduction will cause the child *very brief* discomfort.
2. Seat the child on the lap of a parent/caregiver or an assistant, facing the operator and holding the child in such a way as to hold the child's humerus against her or his side.
3. The person performing the procedure holds the elbow at approximately 90° and grasps the elbow, with the physician's thumb over the region of the radial head (this is done in order to be able to palpate the reduction "clunk" or "click").
4. The physician then holds the patient's wrist firmly and rapidly hyperpronates the forearm. A palpable or audible "click" signifies successful reduction but may not be appreciable.
5. The child may cry for a few minutes after reduction; however, the provider should leave the bedside for 5–10 min and instruct the parents to allow the child to simply play, without focusing on the affected arm.
6. After approximately 10 min, return to the bedside and reevaluate the child, who should have full use of her or his arm by then.

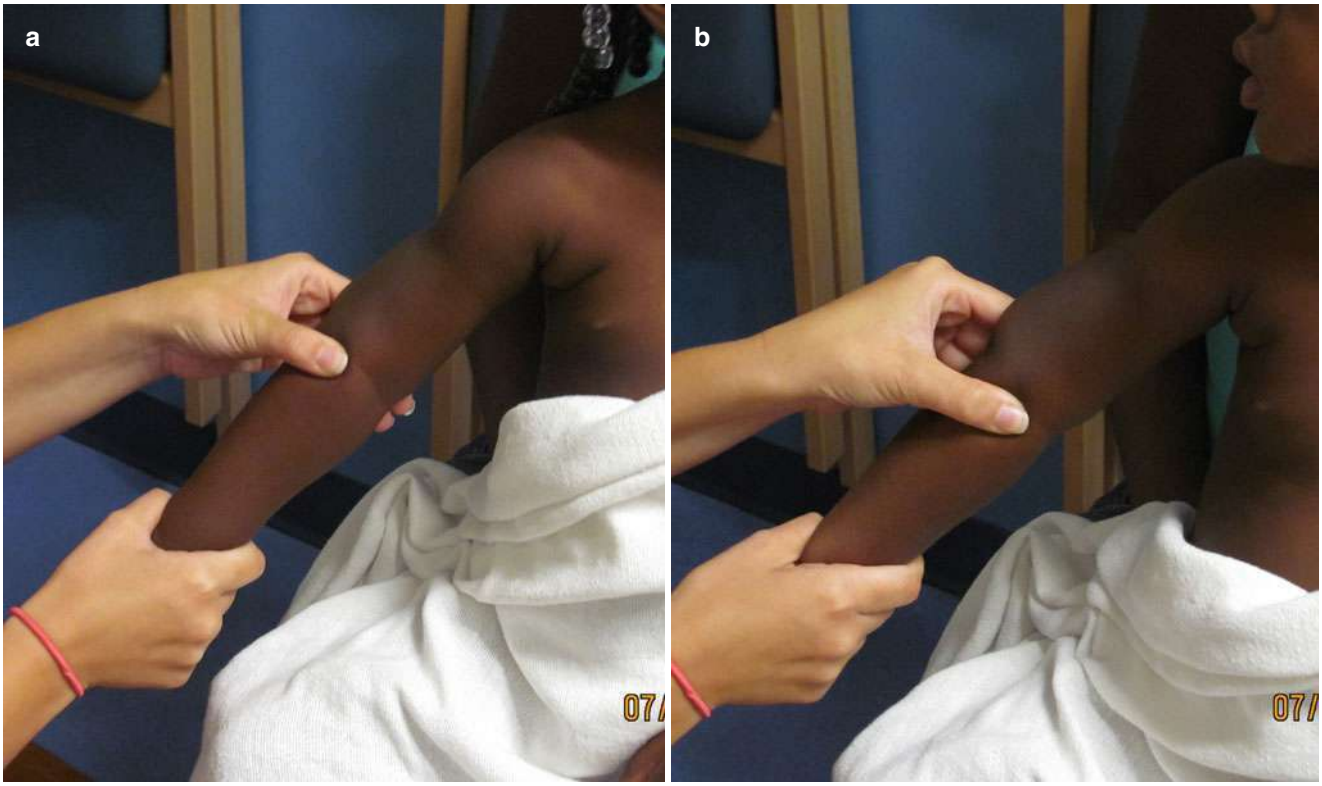

Fig. 127.2 Pronation: (**a**) The examiner's left hand holds the patient's radial head and medial condyle to better appreciate palpable reduction of the subluxation, (**b**) The examiner's right hand gently, but firmly, hyperpronates the patient's forearm from the level of the wrist. Within this maneuver, the examiner should feel or sense a "pop"

127.6.2 Method B: Supination (Fig. 127.3)

1. Follow the first three steps in Method A.
2. While holding the patient's wrist firmly, the operator steadily supinates the patient's forearm completely, followed by flexing the elbow, bringing the patient's wrist up to the shoulder.
3. Return to the fifth and sixth steps in Method A.

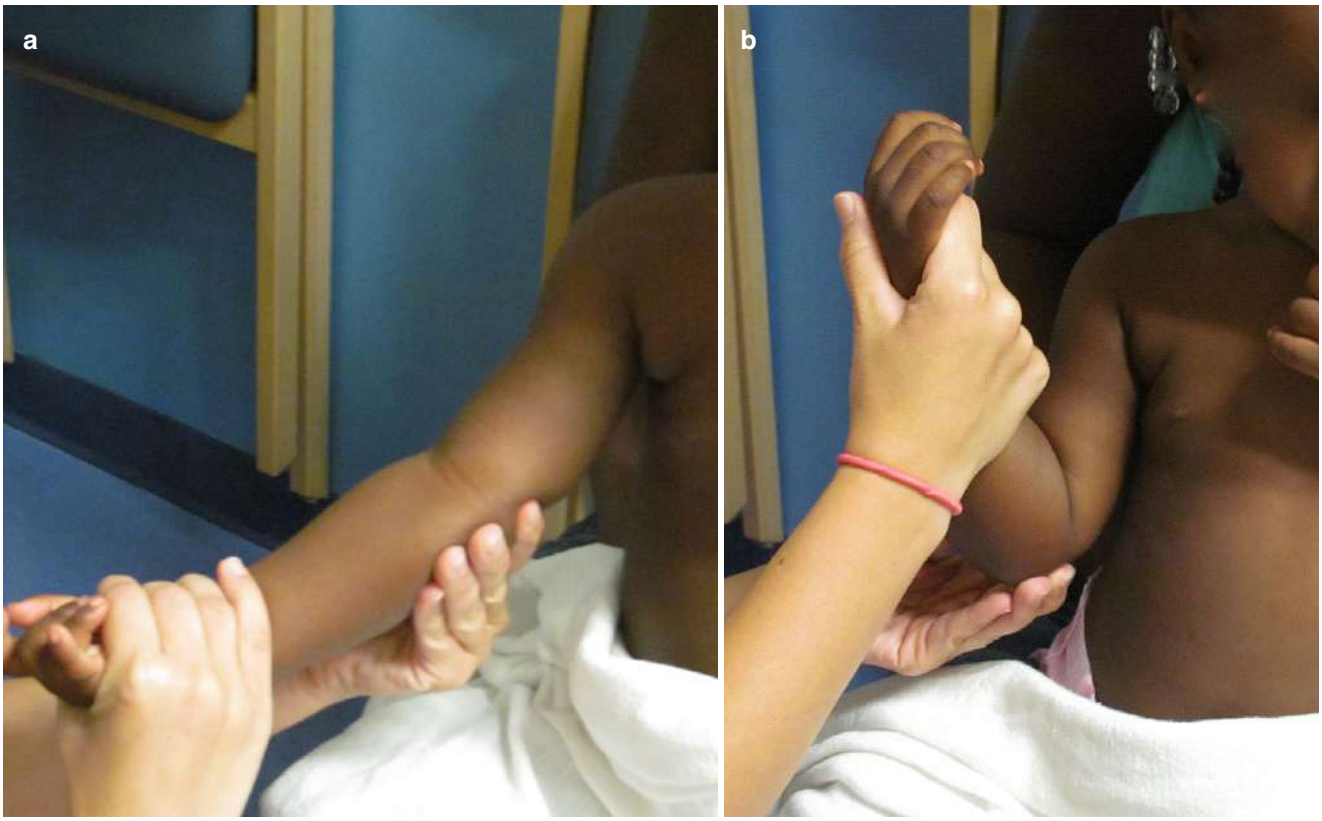

Fig. 127.3 Supination: (**a**) With the examiner's left hand in the same position as for the pronation technique, (**b**) the examiner supinates the forearm, from the level of the wrist, and fully flexes at the elbow. In either of these two maneuvers, the examiner should feel or sense a "pop"

WAR SURGERY

WORKING WITH LIMITED RESOURCES IN ARMED CONFLICT AND OTHER SITUATIONS OF VIOLENCE

VOLUME 1

ICRC

WAR SURGERY
WORKING WITH LIMITED RESOURCES IN ARMED CONFLICT AND OTHER SITUATIONS OF VIOLENCE
VOLUME 1

0973/002 03 2009 3,000

ICRC

ICRC

REFERENCE

WAR SURGERY

WORKING WITH LIMITED RESOURCES IN ARMED CONFLICT AND OTHER SITUATIONS OF VIOLENCE

VOLUME 1

C. Giannou M. Baldan

ICRC

International Committee of the Red Cross
19 Avenue de la Paix
1202 Geneva, Switzerland
T + 41 22 734 60 01 F + 41 22 733 20 57
E-mail: shop.gva@icrc.org
www.icrc.org
© ICRC, March 2009

Cover photos: A. Bakkour/ICRC; A. Bakkour/ICRC; V. Louis/ICRC.

PREFACE

In 1863, a small group of Swiss citizens founded the International Committee of Geneva for the Relief of Wounded Soldiers. A year later, an international diplomatic conference negotiated the first Geneva Convention for the Amelioration of the Condition of Wounded Soldiers in the Field, which to this day is one of the cornerstones of international humanitarian law, and gave the International Committee of the Red Cross its definitive name.

International humanitarian law – the law of war – has developed since, and the mandate, role and activities of the ICRC have expanded to include both protection and assistance for all the victims of armed conflict and other situations of violence. Assistance and relief programmes are now run according to a public health approach and aim to be holistic regarding human needs while respecting the dignity of each and every one.

War surgery – the care of the wounded in armed conflict and other situations of violence – remains a pillar of the ICRC's identity. Over the years, ICRC hospital teams have unfortunately been witness to a great deal of physical and mental suffering in this world. Through caring for the sick and wounded in so many different conflict zones, the ICRC and its partners in the International Red Cross and Red Crescent Movement have developed an expertise they are keen to share, and created a pool of human resources ever willing and prepared to help alleviate some of the suffering.

This new manual presents some of this expertise, gained at great human cost, in the hope that one day it will no longer be required.

Jakob Kellenberger

President
International Committee of the Red Cross

TABLE OF CONTENTS

INTRODUCTION 9

Chapter 1 SPECIAL CHARACTERISTICS OF SURGERY IN TIMES OF CONFLICT 17
1.1 Differences between surgery in times of conflict and civilian practice 19
1.2 How war surgery differs 20
1.3 "Surgeries" for victims of war 25
1.4 Differences between military and non-military war surgery: the ICRC approach 27
Annex 1. A ICRC criteria for introducing a new technology 31

Chapter 2 APPLICABLE INTERNATIONAL HUMANITARIAN LAW 33
2.1 Historical introduction 35
2.2 International humanitarian law: basic principles 36
2.3 The distinctive emblems 38
2.4 The International Red Cross and Red Crescent Movement and its Fundamental Principles 39
2.5 Rights and duties of medical personnel according to IHL 39
2.6 Responsibility of States 41
2.7 Reality check: some people do not follow the rules 43
2.8 The neutrality of a National Red Cross/Red Crescent Society 44
2.9 The role and mandate of the ICRC in situations of armed conflict 45
Annex 2. A The distinctive emblems 48
Annex 2. B The International Red Cross and Red Crescent Movement 50

Chapter 3 MECHANISMS OF INJURY DURING ARMED CONFLICT 53
3.1 The various mechanisms of injury 55
3.2 Ballistics 59
3.3 Terminal ballistics 63
3.4 Wound ballistics 70
3.5 Wound dynamics and the patient 78

Chapter 4 RED CROSS WOUND SCORE AND CLASSIFICATION SYSTEM 81
4.1 Applications of the RCWS and classification system 83
4.2 Principles of the Red Cross Wound Score 84
4.3 Grading and typing of wounds 88
4.4 Wound classification 88
4.5 Clinical examples 89
4.6 Conclusions 91

Chapter 5 THE EPIDEMIOLOGY OF THE VICTIMS OF WAR 93
5.1 Introduction: purpose and objectives 95
5.2 Public health effects of armed conflict 95
5.3 Epidemiology for the war surgeon 98
5.4 General questions of methodology 101
5.5 Aetiology of injury 104
5.6 Anatomic distribution of wounds 106
5.7 Fatal injuries 110
5.8 The lethality of context: delay to treatment 114
5.9 Hospital mortality 117
5.10 ICRC statistical analysis of hospital workload 118
5.11 Conclusions: lessons to be gained from a study of epidemiology 121
Annex 5. A ICRC surgical database 123
Annex 5. B Setting up a surgical database for the war-wounded 125

Chapter 6 THE CHAIN OF CASUALTY CARE 129
6.1 The links: what kind of care, and where? 131
6.2 Surgical hospital treating the war-wounded 133
6.3 Transport 134
6.4 Forward projection of resources 135
6.5 The reality: common war scenarios 136
6.6 Conflict preparedness and implementation 137
6.7 The pyramid of ICRC surgical programmes 138
Annex 6. A Initial assessment of a surgical hospital treating the war-wounded 140
Annex 6. B Strategic assessment of a conflict scenario 146
Annex 6. C Humanitarian intervention for the wounded and sick: typical settings 148

Chapter 7 FIRST AID IN ARMED CONFLICT 153
7.1 First aid: its crucial importance 155
7.2 First aid in the chain of casualty care 156
7.3 First aiders: an important human resource 157
7.4 Essential elements of the first-aid approach and techniques 157
7.5 Setting up a first-aid post 159
7.6 ICRC involvement in first-aid programmes 161
7.7 Debates, controversies and misunderstandings 161

Chapter 8 HOSPITAL EMERGENCY ROOM CARE 169
8.1 ABCDE priorities 171
8.2 Initial examination 171
8.3 Airway 172
8.4 Breathing and ventilation 175
8.5 Circulation 177
8.6 Blood transfusion where supplies are limited 181
8.7 Disability 184
8.8 Environment/exposure 185
8.9 Complete examination 185

Chapter 9 HOSPITAL TRIAGE OF MASS CASUALTIES 189
9.1 Introduction 191
9.2 Setting priorities: the ICRC triage system 193
9.3 How to perform triage 196
9.4 Triage documentation 198
9.5 Emergency plan for mass casualties: disaster triage plan 199
9.6 Personnel 200
9.7 Space 202
9.8 Equipment and supplies 203
9.9 Infrastructure 204
9.10 Services 204
9.11 Training 204
9.12 Communication 204
9.13 Security 205
9.14 Summary of triage theory and philosophy: sorting by priority 206
Annex 9. A Sample triage card 207
Annex 9. B Hospital emergency plan for a mass influx of wounded 208

Chapter 10 SURGICAL MANAGEMENT OF WAR WOUNDS 211
10.1 Introduction 213
10.2 Complete examination 214
10.3 Preparation of the patient 215
10.4 Examination of the wound 216
10.5 Surgical treatment 216
10.6 Retained bullets and fragments 221
10.7 Final look and haemostasis 222
10.8 Wound excision: the exceptions 223
10.9 Leaving the wound open: the exceptions 224
10.10 Dressings 225
10.11 Anti-tetanus, antibiotics, and analgesia 226
10.12 Post-operative care 226

Chapter 11 DELAYED PRIMARY CLOSURE AND SKIN GRAFTING 229
11.1 Delayed primary closure 231
11.2 Skin grafting 233
11.3 Full-thickness grafts 238
11.4 Healing by secondary intention 240

Chapter 12 NEGLECTED OR MISMANAGED WOUNDS 243
12.1 General considerations 245
12.2 Chronic sepsis: the role of biofilm 247
12.3 Surgical excision 248
12.4 Antibiotics 250
12.5 To close or not to close? 251

Chapter 13 INFECTIONS IN WAR WOUNDS 253
13.1 Contamination and infection 255
13.2 Major bacterial contaminants in war wounds 256
13.3 Major clinical infections of war wounds 257
13.4 Antibiotics 263
13.5 Neglected or mismanaged wounds 265
Annex 13. A ICRC Antibiotic protocol 266

Chapter 14 RETAINED BULLETS AND FRAGMENTS 269
14.1 The surgeon and the foreign body 271
14.2 Early indications for removal 271
14.3 Late indications 274
14.4 Technique for the removal of a projectile 275

Chapter 15 BURN INJURIES 277
15.1 Introduction 279
15.2 Pathology 279
15.3 Burn management 282
15.4 Burns presenting late 286
15.5 Nutrition 286
15.6 Care of the burn wound 286
15.7 Closure of the burn wound 289
15.8 Scar management 293
15.9 Electrical burns 294
15.10 Chemical burns 294
Annex 15. A Nutrition in major burns: calculating nutritional requirements 297

Chapter 16 LOCAL COLD INJURIES 299
16.1 Physiology of thermal regulation 301
16.2 Types of local cold injuries 301
16.3 Management 302

Chapter 17 ANAESTHESIA AND ANALGESIA IN WAR SURGERY 305
17.1 Introduction 307
17.2 Anaesthesia methods 308
17.3 Local and regional anaesthesia 309
17.4 Dissociative anaesthesia with ketamine 310
17.5 Post-operative pain management 312
Annex 17. A ICRC pain management protocols 314

Chapter 18 DAMAGE CONTROL SURGERY AND HYPOTHERMIA, ACIDOSIS, AND COAGULOPATHY 319
18.1 Resuscitative surgery and damage control surgery 321
18.2 Hypothermia, acidosis, and coagulopathy 322

ACRONYMS 329

SELECTED BIBLIOGRAPHY 333

INTRODUCTION

Our common goal is to protect and assist the victims of armed conflict and to preserve their dignity. This book is dedicated to the victims of situations which, in a better world, would not exist.

Facing the challenges

One night while on duty Dr X, an experienced surgeon working in an ICRC field hospital in the midst of a civil war, performed a craniotomy on one patient injured by a bomb, an amputation following an anti-personnel landmine injury on another, and a laparotomy after a gunshot wound on the third; not to mention the emergency Caesarean section that arrived, as always, at the most inopportune time, after midnight. She was the only surgeon available that night. This was common practice at the time, and not very much has changed in the last thirty-odd years.

Standard peacetime health services are already limited or lacking in many low-income countries, and faced with the added burden of weapon-wounded they are quickly overwhelmed. A precarious healthcare system is one of the first victims of armed conflict: the disruption of supply lines, the destruction of premises and the flight of medical personnel are all too common.

The lack of adequate resources is not limited to diagnostic and therapeutic technologies. Above all there is a dearth of human resources. Surgeons trained to practise in multidisciplinary teams find themselves alone to face the entire surgical workload and deal with subspecialties with which they have, at best, only a passing acquaintance. Reverting to the philosophy, so common 50 years ago, of the multidisciplinary single surgeon who has to "do it all" is not an easy task.

ICRC teams usually include only one or two surgeons. They are generalists, able to treat all kinds of injuries from simple soft-tissue wounds to penetrating abdominal and head injuries and complicated fractures. They must also provide emergency non-trauma surgical and obstetric care for the civilian population in the area. Ideally, they should be very general surgeons with a broad approach and wide experience.

The principles of war surgery have been known for centuries but need to be relearned by each new generation of surgeons and in every new war.

This commonly held view remains true to this day. Whether performed by military or by civilian surgeons, war surgery has its particular characteristics that are due to the special nature of the context of armed conflict, its limitations and dangers, and the particular physio-pathology of high-energy, penetrating missile and blast wounds. The care of weapon-wounded patients follows accepted surgical standards, but is performed under extreme conditions, which is why the management of a gunshot wound due to criminal violence in a civilian context cannot be easily extrapolated to surgery in a situation of armed conflict.

Where only limited resources are available the surgeon must accept that he cannot fully utilize his capacities and expertise.

Working with limited resources means that the limits of surgery that can be performed are not the expertise of the surgeon, but rather the level of anaesthesia and post-operative nursing care, and the availability of diagnostic and therapeutic equipment.

Limited resources, even in peacetime, may lead to the death of patients who would have survived had more sophisticated means been available. This is often the case in remote, and not so remote, hospitals in low-income countries; a situation exacerbated during armed conflict.

When the principles of triage are applied, saving "life and limb" for the greatest number, with the least possible expenditure of time and resources, often takes precedence.

Indeed, these characteristics mean that war surgery is very different from that practised in peacetime, when most operations are elective and most trauma is blunt, and the surgeon concentrates on doing everything he possibly can – using the full range of resources necessary – for each and every patient.

Furthermore, the work of medical staff in a situation of armed conflict is governed by a special set of rules, in addition to standard medical ethics: international humanitarian law, or the law of war. This is yet another specificity of this type of surgical care and is important for the security of patients and medical personnel alike living and working under dangerous circumstances.

International humanitarian law – the law of war – complements medical ethics in times of armed conflict or other situations of violence.

The ICRC's experience

The ICRC has been providing medical care for the war-wounded ever since its inception, for example during the Franco-Prussian War (1870). The 1970s and 80s, however, saw a tremendous increase in the already considerable humanitarian activities for the victims of war, armed conflict, and other situations of violence. These included relief efforts for refugees, internally displaced persons, and the affected resident population, and medical care for the sick and wounded. In addition, many new organizations were founded and, together with United Nations agencies, they deployed renewed efforts to respond to these humanitarian challenges.

The ICRC embarked on large-scale programmes to provide surgical care for the victims of war. Several independently-run ICRC hospitals were established and surgical staff recruited from various national Red Cross/Red Crescent Societies and Switzerland. A large number of enthusiastic and idealistic surgical staff set off on humanitarian missions. The surgeons were well-trained and experienced, but their training and experience were largely confined to sophisticated hospital facilities in industrialized countries. They faced a steep learning curve.

The ICRC also faced a steep learning curve and has acquired considerable expertise in caring for the victims of conflict in settings where the healthcare system is severely disrupted. This know-how derives from three different, but related programmes in various countries afflicted by armed conflict and other situations of violence around the world.

1. Independent ICRC-run hospitals.
2. Support to local hospitals through the short-term presence of expatriate surgical teams, with a strong focus on training and capacity building; the provision of supplies and equipment; the renovation of infrastructure and water and sanitation facilities; and financial incentives and salaries for local staff when necessary.
3. Organization of war surgery seminars, which provide opportunities for colleagues to exchange experiences and expertise.

This three-pronged approach has enabled the ICRC to develop basic clinical protocols and procedures for surgical techniques appropriate to the management of weapon-wounded patients in situations of limited resources and precarious circumstances. In addition, over the last thirty years, the ICRC has trained and maintained a pool of experienced hospital personnel who do not have to re-invent everything with each new conflict.

However, thanks to more widespread educational opportunities in recent years, there has been a sharp increase in the number of surgeons in conflict-afflicted countries. This has allowed the ICRC to switch the emphasis of its programmes from one of substitution – i.e. independent ICRC hospitals – for a poorly-functioning or non-existent hospital system to one that focuses on support and training of medical personnel in the niceties of the management of patients wounded by the weapons of war.

As part of its training programmes, the ICRC has co-organized more than 120 war surgery seminars – over a dozen a year – in the last decade. On these occasions much expertise and many ideas have been exchanged between surgeons from a number of countries – with a varying degree of experience in war trauma – and ICRC surgeons. We have all learned from these discussions, and some of the lessons are reflected in the contents of this new book.

However, in several contexts the ICRC has continued to provide direct surgical services in a neutral and impartial manner. This form of substitution contributes a fundamental element to the protection of the victims and the medical mission in situations where these humanitarian principles have been sorely tried.

While a number of manuals on war surgery have been published, they are mostly produced by and for the armies of industrialized countries. Their operational norms usually require large investments in means and personnel: helicopter evacuation of patients; well-trained field medics and stretcher-bearers; sophisticated technology; multidisciplinary teams of specialist surgeons, anaesthetists, and nurses. ICRC surgical staff refer to these manuals as references but the conditions and means described therein are seldom met in present-day theatres of armed conflict. Many of their "lessons" are inadequate – or even irrelevant – to humanitarian war surgery or the functioning of public hospitals in many countries working with limited resources.

ICRC surgical care aims to be economical, non-specialized, based on solid scientific principles, and offer good returns given the constraints. The clinical protocols and surgical techniques described in this manual are the standard procedures used among the ICRC's pool of experienced surgeons.

Putting pen to paper

To meet the challenge posed by these conditions, our predecessors in the surgical department of the Medical Division of the ICRC edited a basic reference manual for surgeons embarking on their first humanitarian mission: *Surgery for Victims of War*.

The first three editions of this book have been extensively distributed and received wide acclaim the world over from surgeons who faced the challenge of treating war-wounded patients for the first time. The general surgeon in an isolated rural hospital has perhaps drawn the greatest benefit from it.

Originally intended as a fourth edition of *Surgery for Victims of War* to address the particular needs and specific requests many of our colleagues raised during ICRC-led seminars, as well as to reflect the developments in ICRC surgical practice, it soon became clear that a new book would better serve this purpose. This book now comprises a significant amount of new material to be presented in two volumes, while maintaining the basic reasoning of the original manual.

This first volume is devoted to broad topics, with a number of entirely new chapters of a more general character whose contents are relevant not only to surgeons but also to those responsible for the organization and coordination of surgical programmes in times of armed conflict and other situations of violence. It presents the characteristics of surgical care for the victims of war, in particular the epidemiological, organizational and logistic aspects, drawing on the experience of ICRC medical staff and other colleagues. The second volume will deal with weapon-related trauma to specific body systems.

The surgical techniques presented herein share many fundamental ideas with more sophisticated medical services. However, they also derive from tried-and-tested improvisation and the use of very simple methods of treatment aiming to use technological means as appropriate as possible to the prevailing conditions of limited infrastructure, equipment, and human resources.

The explanations of techniques are geared to the level of knowledge and practice of *general surgeons in a rural hospital*. These surgeons are often the first to see patients wounded in conflict and they know that, under the circumstances, referral to more sophisticated facilities – far away in an inaccessible capital city – is impractical or impossible. This book attempts to provide surgeons who do not have specialized training with basic advice about the treatment of various weapon wounds, describing the types of operations that have proved successful in ICRC and other comparable practice.

The manual and other basic ICRC surgical publications and documents of general interest are included in the DVD attached to this volume. The disk also contains several downloadable files – forms and checklists for example – that can be used in everyday practice and adapted by the reader.

Unless stated otherwise masculine nouns and pronouns do not refer exclusively to men, for the manual is gender neutral. Any use of trade or brand names is for illustrative purposes only and does not imply any endorsement by the ICRC.

We hope that civilian and military surgeons, as well as Red Cross/Red Crescent surgeons, facing the challenge of treating the victims of armed conflict and other situations of violence for the first time under precarious and, at times dangerous, circumstances will find this book useful.

Philippa Parker	Christos Giannou	Marco Baldan
Head of Health Unit ICRC Assistance Division	Former ICRC Head Surgeon	ICRC Head Surgeon

Acknowledgements

This manual is based on *Surgery for Victims of War*, first published by the ICRC in 1988 and edited by Daniel Dufour, Michael Owen-Smith, and G. Frank Stening. The authors included:
Bernard Betrancourt, Switzerland
Daniel Dufour, Switzerland
Ora Friberg, Finland
Soeren Kromann Jensen, Denmark
Antero Lounavaara, Finland
Michael Owen-Smith, United Kingdom
Jorma Salmela, Finland
Erkki Silvonen, Finland
G. Frank Stening, Australia
Björn Zetterström, Sweden
and was illustrated by Penelope L. Zylstra (Australia). Many of her line drawings have been included in the present book.

The second edition (1990) was revised by Robin Gray (United Kingdom) and the third (1998) by Åsa Molde (Sweden).

We owe a debt of gratitude to their pioneering work, and their clear and simple approach which has served as a model.

The present publication has benefited from the input of many colleagues with much experience within the ICRC and outside it. Critical comments and valuable advice throughout were provided by:
Ken Barrand, United Kingdom
Franco De Simone, Italy
Herman Du Plessis, South Africa
Jacques Goosen, South Africa
Åsa Molde, Sweden
Valery Sasin, Byelorussia
Harald Veen, Netherlands
Gunter Wimhoefer, Germany

Beat Kneubuehl (Switzerland) acted as the scientific adviser on ballistics and Sylvain Vité (Switzerland), an ICRC jurist, provided technical expertise on international humanitarian law and revised the appropriate sections. Massey Beveridge (Canada) served as technical adviser on burns and skin grafting and made significant contributions to the relevant chapters.

The chapter on the Red Cross Wound Score is largely based on the revised edition of the ICRC brochure by Robin Coupland (United Kingdom), who also provided essential comments and advice on ballistics and epidemiology, and has moreover played an important role through his many other pertinent publications. Holger Schmidt (Germany) and Eric Bernes (France) gave advice on first aid and emergency room trauma care. Haide Beckmann (Germany) and Thomas Walker (Switzerland) made contributions to the chapter on anaesthesia and Dieter Jacobi (Germany) provided comments for the chapter on chronic infections.

The ICRC Master Surgeons Workshop, held in Geneva in March 2002, revised the Red Cross Wound Score, ICRC triage categories, and established the ICRC antibiotic protocol. Participants included:
Marco Baldan, Italy
Massey Beveridge, Canada
Christos Giannou, Greece-Canada
François Irmay, Switzerland
Dieter Jacobi, Germany
Ben Mak, Netherlands
Valery Sasin, Belarus
Jukka Siegberg, Finland
Harald Veen, Netherlands
Gunter Wimhoefer, Germany

In addition, the ICRC Senior Anaesthetists Workshop, held in Geneva in November 2002, was the basis for the chapter on anaesthesia and established the ICRC pain management protocol. Its participants included:
Sunao Asai, Japan
Haide Beckmann, Germany
Lisa Bennett, Australia
Jeanne Frossard, United Kingdom
Christiane Gerber, Switzerland
Christos Giannou, Greece-Canada
Tuula Kangas-Saarela, Finland
Svante Linden, Sweden
Peter Mahoney, United Kingdom
Barthélémy Merzouga, Switzerland
Pascal Ollé, France
Erkki Saarela, Finland
Bernadette Sterckx, France
Vladislav Virago, Belarus
Eric Vreede, Netherlands
Jolanda Walker, Switzerland

Both workshops also helped define the ICRC criteria for the adoption of new technologies, the level of laboratory expertise required, and general strategies for ICRC surgical programmes.

Christiane de Charmant handled the editing of the final text and was responsible for the production while Pierre Gudel provided the graphic design. Their contribution is greatly appreciated.

The authors are ICRC staff members and no outside financial or material support has been received for the publication of this book.

Chapter 1

SPECIAL CHARACTERISTICS OF SURGERY IN TIMES OF CONFLICT

1. SPECIAL CHARACTERISTICS OF SURGERY IN TIMES OF CONFLICT

1.1 Differences between surgery in times of conflict and civilian practice 19

1.2 How war surgery differs 20
1.2.1 IHL: protection of non-combatants and those *"hors de combat"* and the rights and obligations of medical personnel 20
1.2.2 Specific epidemiology of war wounds 21
1.2.3 Predominance of emergency surgery 21
1.2.4 Surgery in a limited technical environment 21
1.2.5 Surgery in a hostile, violent environment 21
1.2.6 Mass casualties involving the principles of triage 22
1.2.7 Triage and surgery in successive echelons 22
1.2.8 Outcome of hospital patient care depends on the efficiency of pre-hospital echelons 23
1.2.9 Specific wound pathology: bullets, bombs, blast weapons and non-conventional weapons 24
1.2.10 Specific techniques appropriate to the context and pathology 24
1.2.11 Increased prevalence of endemic disease 25

1.3 "Surgeries" for victims of war 25

1.4 Differences between military and non-military war surgery: the ICRC approach 27
1.4.1 Military-civilian cooperation 27
1.4.2 Constraints: security 27
1.4.3 Constraints: logistics 28
1.4.4 Constraints: hospital equipment 28
1.4.5 Constraints: blood transfusion 29
1.4.6 Constraints: geography and climate 29
1.4.7 Constraints: culture shock 29
1.4.8 Constraints: the human factor 30

ANNEX 1. A ICRC criteria for introducing a new technology 31

1.1 Differences between surgery in times of conflict and civilian practice

The differences between civilian and war trauma are manifold – as are the differences between the experience of the ICRC and those of conventional military medical services.

Most surgeons today, the world over, derive their trauma training from road-traffic accidents and much that applies to the management of casualties in civilian settings will also apply to the situation of armed conflict: war surgery follows classical surgical standards. However, the generation of surgeons who mostly had to deal with accidents among agricultural or industrial workers were well aware of the dangers of gas gangrene and tetanus and the necessity of good wound excision and delayed primary closure. It was relatively easy for the surgeon to shift from this "septic" civilian surgery to war wounds. This is no longer the case for many surgeons trained today; laparoscopy, radioscopic embolization, and unreamed intramedullary nails will not get you very far when facing a landmine injury of the abdomen or a machine-gun wound to the thigh. Early specialist training of surgeons and sophisticated modern technology benefit many patients in a peacetime environment, but can be an obstacle to the practice of surgery during armed conflict.

War wounds are different. The extent of tissue destruction and contamination seen in war injuries is nothing like what is seen in everyday trauma practice. Working conditions during war are radically different from those prevailing in peacetime. Resources are limited and surgeons are often obliged to improvise or make compromises in their management decisions. Their aim should be to bring the best care possible to their patients under the circumstances, not the best care as described in the academic literature.

War surgery is a surgery of mass casualties. The logic of war triage has little to do with the routine emergency department triage of a major civilian trauma centre: war triage has a "leave to die in dignity" category unheard of in everyday civilian practice.

War surgery involves the staged surgical management of the wounded, often at different echelons of care and provided by different surgeons, especially in a military context. Even in a humanitarian context, such as ICRC surgical programmes, several surgeons deployed on short missions may participate in the treatment of a single patient. In everyday civilian practice on the other hand, the same surgeon assumes responsibility for the entire surgical management of his or her patient. While modern civilian practice often involves a "multidisciplinary approach", war surgery often demands a "multi-surgeon" approach.

> "All the circumstances of war surgery thus do violence to civilian concepts of traumatic surgery."
>
> Michael E. DeBakey[1]

These and other challenges mean that practitioners facing surgery for the victims of war for the first time will have to change their mindset, i.e. their "professional mental software".

1 DeBakey M E. Military surgery in World War II – a backward glance and a forward look. *NEJM* 1947; **236**: 341 – 350. Michael E. DeBakey (1908 – 2008), an American surgeon of Lebanese origin, was a pioneer of modern cardio-vascular surgery. His treatise on the management of vascular trauma constitutes a basic reference. He invented the concept of Mobile Army Surgical Hospitals (MASH) for the US armed forces in Korea.

1.2 How war surgery differs

> War surgery is the management of an "epidemic of trauma" in a series of steps: echelons.
>
> N. I. Pirogov[2]

A number of special features characterize the practice of surgery in time of war.[3]

1. Special rules: international humanitarian law (IHL), i.e. the protection of the sick and wounded, and the rights and obligations of medical personnel.
2. Specific epidemiology of war wounds.
3. Predominance of emergency surgery.
4. Surgery in a limited technical environment.
5. Surgery in a hostile, violent environment: constraints of the tactical situation.
6. Mass casualties involving the principles of triage.
7. Triage and surgery in successive echelons of casualty care.
8. Outcome of hospital patient care is a function of the efficiency of pre-hospital echelons.
9. Specific wound pathology: bullets, bombs, blast, non-conventional weapons.
10. Specific techniques appropriate to the context and pathology.
11. Increased prevalence of endemic disease.

1.2.1 IHL: protection of non-combatants and those "*hors de combat*" and the rights and obligations of medical personnel

The Geneva Conventions of 1949 and their Additional Protocols of 1977 define categories of individuals who, by virtue of these treaties, are protected during armed conflict. These include non-combatants, combatants who no longer participate in hostilities – "*hors de combat*" – either through sickness, injury, shipwreck, or by becoming prisoners of war; and those who care for the sick and injured, namely medical and religious personnel. The latter two use the protective emblem of the red cross, red crescent or red crystal to mark the means and facilities that take care of the sick and wounded. According to the law, all these categories of protected individuals are immune from attack, as long as they do not take an active part in hostilities. International humanitarian law – the law of war – offers particular rights to medical personnel, but also ascribes obligations to them.

All health professionals are bound by medical ethics, in times of peace and of war. These are not replaced but are complemented by international humanitarian law. Compliance with these laws may create particular ethical dilemmas and security problems, and the military hierarchy does not always understand the demands of medical ethics. Civilian health staff may face particularly problematic and dangerous situations during a civil war where their own community is involved in the conflict. Chapter 2, Applicable international humanitarian law, explains the main principles and rules governing the rights and obligations of medical staff in times of armed conflict.

ICRC

Figure 1.1
Soldiers "*hors de combat*": POWs.

Japanese Red Cross Society

Figure 1.2
Soldiers "*hors de combat*": the wounded.

ICRC

Figure 1.3
Those who care for the sick and wounded.

2 Nikolai Ivanovich Pirogov (1810 – 81): Russian anatomist and surgeon, professor at the Academy of Military Medicine, St Petersburg. Founded modern field surgery during the Crimean War (1854); devised the plaster cast and the mass use of anaesthesia on the battlefield and wrote a reference field surgery manual. Representative of the Russian Red Cross sent to inspect the hospitals on both sides during the Franco-Prussian War in 1870.

3 List modified and adapted from bibliographic sources.

1.2.2 Specific epidemiology of war wounds

The nature of warfare – on land, at sea or in the air – will create a particular epidemiology of the wounded. The nature of weapons, protective body armour, and any delay in transport will affect the anatomic distribution of injuries and their severity. The understanding of these epidemiological factors will have important consequences in terms of preparation and allocation of resources: i.e. standardized supplies and specialized personnel (see Chapter 5).

1.2.3 Predominance of emergency surgery

War surgery primarily consists of emergency surgery, especially during early tactical field care. Sophisticated techniques or reconstructive procedures have no place here, except well after combat and in a distant referral hospital (see Chapters 6 and 8).

1.2.4 Surgery in a limited technical environment

The environment in times of war is bleak and harsh. The limits of surgical work are largely determined by the logistic difficulties attending the supply of remote and dangerous areas and the lack of maintenance, repair and spare parts. There is seldom enough technical support staff to ensure that infrastructure functions correctly.

V. Louis / ICRC

Figure 1.4
Surgery in a limited technical environment.

Despite lavish outlays for field hospitals by the armies of modern industrialized countries, equipment limitations in tactical situations are well recognized. The lack of sophisticated diagnostic equipment rather than the surgeon's technical capabilities and expertise is often what determines what can be done. What is essential must be differentiated from what is "nice to have".

1.2.5 Surgery in a hostile, violent environment

The adverse conditions of a tactical situation may put the safety of patients and medical personnel at risk, and thereby create less than optimal working conditions. Dangerous evacuation routes may compromise transport and result in delays. Security must be ensured for patients and staff, by selecting suitable sites for first-aid posts and hospitals. Health facilities and ambulances should be clearly marked with the protective emblem of the red cross, red crescent or red crystal, according to the dictates of international humanitarian law.

Not all combatants are disciplined and well trained. For anyone caught up in armed conflict, there is an all too familiar syndrome seen amongst young fighting men who are under the influence of a "toxic cocktail": testosterone, adrenaline, alcohol and cannabis (and, at times, "other substances" as well).

T. A. Voeten / ICRC

Figure 1.5
Working in a hostile environment.

The setting of surgical work can change rapidly, and surgeons must be prepared to adapt to various difficult and austere environments. With few doctors and limited staff available, and wounded people pouring in, hospital facilities are easily overwhelmed. Doctors and nurses also get tired and sick, and are sometimes frightened. The physical and mental strain of working in new and strange circumstances and at times dangerous conditions means that they may not perform as well as usual.

R. Bigler / ICRC

Figure 1.6
Mass casualties: the principles of triage.

1.2.6 Mass casualties involving the principles of triage

Much has been written about the carnage of the major wars of the twentieth century, and the role of triage in the management of thousands of casualties resulting from a single battle. These lessons are still relevant in contemporary conflicts. The logic followed must be to "do the best for the most" and not "everything for everyone". This involves the most important change in the professional mindset of the surgeon.

Triage decisions are amongst the most difficult in all medical practice, possibly creating ethical dilemmas. There may also be, at times, a conflict between medical criteria and tactical ones of military necessity that require some form of compromise. Persons performing triage must be prepared to accept these necessary compromises while maintaining their personal medical integrity (see Chapter 9).

1.2.7 Triage and surgery in successive echelons

The sick and wounded are evacuated and transferred along a chain of casualty care. The principles of triage are applied at every stage in this chain. The initial surgery must not compromise the performance of later, definitive surgery. The prognosis will be much better if the wounded are evacuated rapidly to a higher echelon hospital structure. A surgeon in the field must understand the system and know what will happen to the patient at the next echelon of care and what is required of him at his echelon (see Chapter 6).

War surgery requires a logic of phased wound management. Treatment takes place in five phases often, although not necessarily, in five different places. This is the classical set-up in military planning. It involves high costs, including the means of transportation and the discipline required of such organization. Modern concepts can be integrated into this approach, such as forward surgical teams performing damage-control surgery. The five phases are as follows.

1. First aid at the point of wounding: self- or "buddy" treatment, or that provided by a field medical officer or first aider.
2. First medical treatment: vital emergency measures including the beginning of resuscitation, generally at the first-aid or dressing post, clearing station or collection point.
3. First surgical treatment: debridement/wound excision, without primary closure, at the first echelon hospital.
4. Definitive treatment: delayed primary closure of wounds and surgical treatment according to traditional principles in a referral hospital. Physiotherapy and convalescence.
5. Reconstructive surgery and rehabilitation: specialized surgery with multiple reconstructive procedures and the fitting of prostheses when required.

The rapid turnover of medical personnel treating numerous patients at different points in a chain of casualty care creates a necessity for standard protocols, which cannot be left to the discretion or whim of the individual surgeon. One cannot individualize treatment for each single patient in a tactical situation; different surgeons will operate on the same patient at the different echelons. In civilian practice, by contrast, the same surgeon manages the surgical care of the individual patient throughout that patient's course of treatment; even in a multidisciplinary approach, the same team is involved.

These five echelons for the management of the war-wounded do not always exist in a civilian or humanitarian context, where all five phases may well occur in the same medical facility – as is in fact common practice for ICRC hospital teams. However, in the humanitarian context, there is also a rapid turnover of personnel. Standardized protocols are the only means to ensure continuity of treatment and to organize surgical and nursing care on a sound footing; the protocols cannot be changed with every change of team surgeon.

Old lessons for new surgeons

Save life and limb, sacrifice limb to save life, prevent infection, and render the casualty transportable to the next echelon.

Heroic surgery will never replace good surgery.

1.2.8 Outcome of hospital patient care depends on the efficiency of pre-hospital echelons

First aid provided at the point of wounding – or at the safest place near the battlefield – and rapid evacuation are of vital importance; mortality and morbidity increase with delay. If first aid is inadequate or unavailable and the evacuation chain is long, then the outcome will be decided by nature. However, hospital mortality decreases as evacuation time increases; with very long delays, the severely injured die long before reaching hospital and surgeons spend most of their time dealing with the septic complications of the survivors (see Chapters 5 and 7).

Old lessons for new surgeons

It is more important to provide proper first aid and render the injured fit for transport, than to give early but inadequate treatment, with insufficient means, and insufficient knowledge.

T. A. Voeten / ICRC

Figure 1.7
Inadequate pre-hospital care.

1.2.9 Specific wound pathology: bullets, bombs, blast, and non-conventional weapons

War wounds are qualitatively different from trauma seen in civilian practice: all are dirty and contaminated. Projectiles may cause massive destruction of soft tissues, bones, and important organs. Infection is the great danger and the rules of septic surgery apply (see Chapters 3 and 13).

Old lessons for new surgeons

War wounds are dirty and contaminated from the moment of injury. The rules of septic surgery apply.

Nothing in routine civilian practice is similar to the polytrauma of multiple shell fragments, traumatic amputation due to an anti-personnel landmine, or the devastating effects of a high-kinetic energy transfer from a military rifle bullet. Again, a change in the professional mindset appropriate to a modern trauma centre is required in order to adjust to the management of wounds in times of armed conflict. Surgeons practising in a low-income country, dealing with much septic pathology, will find this change much easier to handle than those whose usual work involves sophisticated technology with abundant good nursing care in an aseptic environment.

1.2.10 Specific techniques appropriate to the context and pathology

The care of numerous patients, treated in many locations by different surgeons, in austere conditions demands simplicity, security, and speed of surgical procedures. The need for speed in dealing with mass casualties with inadequate numbers of staff should not lead to confusion and disorder. Phased wound care imposes standards and a systematic approach: doing the least amount of surgical work for the greatest results, saving "life and limb", and then sending the patient on to the next stage in the chain of casualty care. The follow-up of surgical procedures by the individual surgeon in the field is difficult if not impossible, and precludes personal and individualistic techniques. Standard protocols, as mentioned, are necessary.

The majority of wounds will be to the extremities, and the objective is to treat them so that they heal as quickly as possible without becoming infected. Sepsis, potentially fatal (tetanus, gas gangrene, haemolytic septicaemia), is the greatest danger for survivors. As aforesaid: the rules of septic surgery apply.

The basic principles of management of war wounds comprise the following steps.

1. Early and thorough wound excision and irrigation.
2. No unnecessary dressing changes.
3. Delayed primary closure.
4. Antibiotics as an adjuvant.
5. Anti-tetanus vaccine and immunoglobulin if necessary.
6. No internal bone fixation.[4]
7. Early physiotherapy.

4 No internal bone fixation is the rule, at least in the "acute phase". Recent experience has shown the possibility of using internal fixation once the soft tissues have healed in the absence of any infection, but only in skilled hands and with excellent conditions of hygiene and nursing care. This is not, however, standard policy in ICRC practice where no internal fixation whatsoever is used. The risk, and probability, of abuse of internal fixation precludes its availability.

Old lessons for new surgeons

The best antibiotic is good surgery.

Correct surgery gives the patient the best chance of survival with a good quality of life, and shortens the stay in hospital. High-quality physiotherapy is required to ensure early mobilization after surgery and a good functional result. Treatment is not complete, however, until the patient is rehabilitated; prosthetic workshops are needed to fit amputees and provide other suitable devices, such as orthoses, crutches or wheelchairs.

1.2.11 Increased prevalence of endemic disease

Until World War I, more soldiers died of disease than of their wounds. Battle-injury attrition was usually around 20%, and disease four times more common among soldiers. Non-battle injury attrition rates remain very high even today; infectious and communicable diseases differ according to the geography and climate, but psychological disorders and vehicle accidents are universal.

The destruction, disruption and disorganization that accompany armed conflict often cause the public health system to be among the very first to suffer. The humanitarian consequences translate into reduced availability of basic public health requirements such as water, food, shelter, etc., for the civilian population and often overwhelm the capacities of the public health system. This compounds the challenge of caring for the civilian population in a conflict zone: i.e. residents, internally displaced persons, and refugees fleeing to or from neighbouring countries (see Chapter 5).

1.3 "Surgeries" for victims of war

There is more than one type of war surgery. While the needs of the wounded are the same, the means and resources available to meet those needs vary widely from country to country and situation to situation, giving rise to different approaches to war surgery. The management of the war-wounded as performed by the military medical services of an industrialized country is not the same as that of a public rural hospital in a low-income country. Although the principles of wound management are the same in both cases, the diagnostic and therapeutic possibilities are very different. The latter should be *appropriate* to the technological, financial, and human resources at hand. Obviously many of these same constraints apply to the practice of civilian, everyday trauma and elective surgery around the world.

At least four major scenarios for the surgical management of victims of war in contemporary armed conflicts can be described.

1. Conventional army of an industrialized society, with a high level of public financing and where the military attempt to provide the same level of surgical care that is practised in civilian life. Rapid evacuation and transfer of patients to specialized facilities is routine. Safe access to adequate medical care for the sick and wounded is perceived as a right; and the duty to provide it incumbent on the armed forces.

2. Developing country with an emergent economy where, at least in the capital and other major cities, a high-level of specialized surgical care including enough trained personnel is available, even if this is not the case in rural areas. Evacuation and transfer of patients to specialized facilities is possible, if sometimes difficult. Safe access to adequate medical care is a goal to be achieved.

3. Poor country, with limited financial and human resources. A few major surgical centres may exist in the capital city but provincial and rural hospitals are largely staffed by young general surgeons or general practitioners with some surgical experience. Supplies, budgets, and human resources are chronically inadequate or even absent. Evacuation and transfer of patients are difficult or impossible. Safe access to adequate medical care is rarely guaranteed.

4. Non-State actors, guerrilla groups, populations without safe access to public structures. Field surgery is practised by a few trained doctors and nurses, because there is no alternative. Safe access to the victims by health professionals, and the victims' access to medical care, is impossible or rare, problematic, and always a challenge.

H. Du Plessis / South African Military Health Service, U. Pretoria

Figure 1.8

Surgery for victims of armed conflict in a modern hospital.

F. McDougall / ICRC

Figure 1.9

Another type of surgery for victims of armed conflict.

1.4 Differences between military and non-military war surgery: the ICRC approach

Non-military war surgery is practised by civilian medical structures (health ministry, missionary and private hospitals) and those of the ICRC or other humanitarian agencies. This section details the ICRC's experience and explains its approach.

The aims of war surgery for the ICRC are to protect the sick and wounded and help maintain their dignity by ensuring access to adequate care; save "life and limb"; minimize residual disability and assist amputees. Besides the direct victims, the ICRC also attempts to support the health system by assisting local medical colleagues to maintain the infrastructure and human resources necessary to resume functioning after the end of the conflict, thus providing the civilian population with at least the basics of healthcare. ICRC assistance to local structures may involve construction and renovation of health facilities, water and sanitation work, food supplements for patients and staff, equipment, supplies and basic salaries. Training programmes for doctors and nurses may be included. In addition, the ICRC may at times set up its own independent hospitals with expatriate staff supplemented by local personnel (see Chapter 6).

1.4.1 Military-civilian cooperation

Armed forces deployed in the field have a specific military mission. Their medical component aims primarily to give support to their soldiers in this mission, i.e. to achieve their tactical and strategic plans. The military may have many of the same "assistance and reconstruction" aims as civilian organizations, but medical criteria in a military context often must take second place to the tactical and strategic demands of military and political necessity.

The ICRC is a neutral, impartial and independent, totally humanitarian, institution. It promotes adherence to international humanitarian law and aims to protect and assist the victims of conflict – all the victims on all sides. Any cooperation with armies in the field of battle that compromises the perception of that neutrality, impartiality or independence can only compromise the humanitarian work of the ICRC, as well as that of other organizations and agencies.

The ICRC insists on maintaining its independence, and the independence of a "humanitarian space" separate from the "assistance and reconstruction" work of armies in the field. Many humanitarian organizations share the same opinion and approach.

1.4.2 Constraints: security

The ICRC often has little or no control over casualty evacuation because of security constraints. In many countries the wounded are transported by private means: taxis, donkeys, oxcarts or on foot. In some contexts, the ICRC has been able to set up first-aid posts, or assist a National Red Cross or Red Crescent Society in doing so. One extraordinary example was a 16-year medical evacuation system by fixed-wing aircraft run by the ICRC and United Nations' Operation Lifeline Sudan to assist victims of the conflict in southern Sudan: it transported over 30,000 sick and wounded patients to an ICRC hospital in northern Kenya. Even with aircraft available, the logistics and distances involved regularly resulted in a one- to three-week delay in evacuation.

While the armed forces deploy the necessary means to protect their health facilities from the "lethal chaos of the battlefield",[5] the ICRC depends on the red cross emblem and negotiations with all belligerent forces for its physical protection. The ICRC has no guns to protect itself and is just as dependent on local authorities and leaders as public hospitals are. The ICRC relies on the limits imposed by international humanitarian law and the discipline of combatants, and its diplomatic negotiating ability. Other humanitarian organizations who work in war zones face similar constraints.

5 Butler F. Tactical Combat Casualty Care: combining good medicine with good tactics. *J Trauma* 2003; **54 (Suppl.)**: S 2 – 3.

1.4.3 Constraints: logistics

A hostile environment presents more than security risks. Remote areas with dangerous routes and extreme climates pose numerous logistic problems for the delivery of supplies and the maintenance of basic infrastructure, for both hospital and living quarters. The military often have lift, delivery and transport capacities that civilian institutions lack. Although the armed forces also have their logistic limits, they are of a different order of magnitude to those of the health ministry, non-governmental organizations, or the ICRC.

V. Mboah / ICRC

Figure 1.10

Challenging environment.

1.4.4 Constraints: hospital equipment

These limitations also affect hospital equipment. The military have lift constraints because they must also transport arms and munitions. For the ICRC, equipment limitations mean resorting to appropriate technology and mastering the tasks of maintenance, repair and availability of spare parts. This is particularly important when working in remote areas in a poor country with training programmes for local colleagues. The aim is to avoid creating a technological dependency that the permanent local staff will not be able to cope with when the conflict is over, and the ICRC leaves the hospital and the country. To this end, the ICRC has developed criteria for the introduction of any new technology into its standard list of medicines and equipment for hospital programmes (see Annexe 1. A: ICRC criteria for introducing a new technology).

K. Barrand / ICRC

Figure 1.11

Limited surgical equipment.

A standard list is a basic, limited set of essential supplies, both medical and non-medical, which corresponds to an appropriate standard of care. All items on a standard list should always be available from a central store or reliable supplier. Standardization provides a simple framework, within which resources can be used to maximum effect, promotes continuity in patient care, helps to simplify staff training, and makes it easier to introduce new and inexperienced staff into the system. The ICRC and the International Federation of Red Cross and Red Crescent Societies have established an Emergency Items Catalogue[6] with predetermined sets and kits that run the gamut of assistance programmes.

ICRC experience shows that it is possible to perform good quality surgery with basic technology such as simple X-ray, and electronic monitoring in the theatre and post-operative ward limited to a pulse oxymeter. An ICRC laboratory set-up is also basic; there is no capacity for bacteriological culture and sensitivity and no blood components.

K. Barrand / ICRC

Figure 1.12
Basic supplies.

1.4.5 Constraints: blood transfusion

Blood for transfusion is often difficult to obtain in some countries because of cultural and religious restrictions and beliefs. With the steady increase in HIV infection around the world, testing must be adequate and indications for transfusion strictly limited. In some areas of the world, giving blood should probably be totally avoided. The use of blood should be restricted to vital needs and to patients with a good chance of survival, following the principles of triage. As in many, if not most, provincial and rural hospitals around the world, ICRC practice uses whole blood for transfusion, as fresh as possible, and usually donated by a relative.

1.4.6 Constraints: geography and climate

The geographic context may be important in terms of disease and the added burden it represents. The wounded may suffer from other illnesses, such as tuberculosis, malaria, typhoid and intestinal worms, as well as from malnutrition. In countries with chronic malaria infestation, there is often a peak of fever post-operatively. The surgeon must therefore try to acquire some basic knowledge of the diseases specific to the area and their treatment. Local healthcare workers are usually familiar with these conditions and more expert in their treatment than expatriate staff. These pathologies may also affect and represent a danger for expatriate staff.

1.4.7 Constraints: culture shock

Cultural constraints are another challenge that may add to the frustrations of medical work in a combat zone. In some societies, amputations and laparotomies can only be performed with the consent of the *family* of the patient. After a discussion in which the clear advantages are explained, the final decision must be left to the family. This procedure, which respects local cultural and social behaviour and norms, has to be followed and accepted even though it may be considered as a limitation and constraint by surgical and nursing staff. It is particularly difficult for committed medical personnel to see young people die because permission for necessary surgery has been denied.

In many societies, it is common for a relative to stay with a hospitalized patient, helping with nursing duties related to hygiene and feeding, and providing psychological support. This tradition must be accepted.

Adaptation to the cultural, social, and geographic context is essential.

6 See Selected bibliography.

1.4.8 Constraints: the human factor

> "War surgery is a surgery of complications, performed by doctors who are often ill-trained or without surgical training. It is surgery replete with adaptations and improvisations to replace that which is missing, a surgery of surprises that new means and methods of combat reveal."[7]

This classic quotation from the Swiss army's war surgery manual describes the situation for the new military surgeon and, as is more and more often the case in contemporary conflicts, for civilian surgeons faced with the victims of armed conflict for the first time. New ICRC surgeons coming from the National Red Cross or Red Crescent Society of a rich industrialized country face the same challenges: a constantly changing field of battle with new limits and constraints that come as a surprise. One must plan for alternatives and always keep an open mind to new options.

If military war surgery is the management of an "epidemic of trauma" in a series of echelons, this is not always the prevailing situation in non-military circumstances. Unlike a military field hospital, an ICRC hospital is responsible for all levels of medical care. It often acts as a first-aid post, field hospital, base hospital, and referral centre all in one. The "multi-surgeon" military approach to treatment in echelons gives way to a more traditional one of attending to the entire surgical history of a patient. However, since deployment of ICRC surgeons is usually for a short period of time (three months on average), a number of surgeons may be involved in caring for a single patient. Continuity of care is essential.

The modern military may "project forward" technical skills by deploying field surgical teams close to the battlefield. The aim is to perform critical surgery, often damage control surgery, as soon as possible after injury in an attempt to save lives, thereby reducing the number killed in action. The ICRC has also deployed field surgical teams – Somalia in 1992, southern Sudan in 2000, and Darfur in 2005 – but with a different aim: the protection of non-combatants and wounded fighters who no longer participate in hostilities and who would otherwise have no access to surgical care. The performance of this medical act and the protection of access to surgical care are based on the essential principles of the ICRC as a neutral and impartial humanitarian actor.

The surgeon must be able to adapt to the conditions of field surgery where "somewhat clean with soap and water" replaces a "sterile" environment and "favourite" surgical instruments are not available on the standard list. Furthermore, living conditions may resemble camping out in the bush and everyone in the team (4 members: surgeon, anaesthetist, theatre nurse and post-operative nurse) participates in the preparation of food and accommodation.

The following qualities are particularly needed by personnel working in ICRC hospitals or field surgical teams:
- professionalism;
- sound judgement and common sense;
- adaptability.

Local skills and material improvisation in some countries may bring to the attention of the surgeon efficient, cheap, and useful ways of treatment: mashed papaya for burns or autoclaved banana leaves as a non-adherent dressing, for instance. Expatriate personnel must show themselves capable of learning "new old tricks" and adapting to the circumstances. War surgery is a challenge and the work trying. Medical staff must be prepared – physically and mentally – for frustrations, fatigue and long hours, and being a witness to the results of "man's inhumanity to man".

> War is deleterious to your health.

7 War Surgery Commission of the Federal Military Department. *Chirurgie de guerre (Aide-mémoire 59.24 f) [War Surgery (a primer)]*. Bern: Swiss Army, 1970 and 1986.

ANNEX 1. A ICRC criteria for introducing a new technology

1. Needs assessment
What is the added value of this new technology? Are requested instruments "essential", "important" or "nice to have", or even "superfluous" and a luxury?

2. Maintenance requirements
What are the extra burdens for the daily maintenance of such equipment?

3. Ease of repair
Are specialized technicians necessary, and available?

4. Availability of spare parts
Is there a reliable local supplier?

5. Cost
Expense alone is not a criterion if the equipment is necessary. However, it is taken into account along with the other factors in a total cost/benefit analysis.

6. Competency required to use the technology in question
Is the expertise widely mastered and available or does it correspond to the particular practice of an individual doctor or nurse?

7. Continuity of the competency required
Can successive surgical teams use the equipment or does it depend on the expertise of a limited number of people?

8. Presence of the technology in the country
To all intents and purposes, the ICRC shall not be the first to introduce a new technology to a country; some local practice or competency must already exist.

9. Professionalism and ethical concerns
The supply of equipment and instruments must at all times meet demanding standards of professionalism in surgical care and address possible ethical concerns. (e.g. in Europe, following the outbreak of bovine spongio-encephalopathy – so-called "mad-cow disease" – catgut suture material has been banned by the European Union and Switzerland. It would not be ethical for the ICRC to continue to provide such sutures in its assistance programmes in other parts of the world, thereby creating a lower safety standard than in Europe.) Quality control of medicines and equipment is becoming a major worldwide problem as identified by the World Health Organization.

10. Sustainability
Only if the technologies can be sustained after ICRC withdrawal is it worth considering their utilization.

Chapter 2

APPLICABLE INTERNATIONAL HUMANITARIAN LAW

2. APPLICABLE INTERNATIONAL HUMANITARIAN LAW

2.1 Historical introduction 35

2.2 International humanitarian law: basic principles 36
2.2.1 Principles underlying IHL: the "law of war" 37
2.2.2 IHL and human rights law 38

2.3 The distinctive emblems 38

2.4 The International Red Cross and Red Crescent Movement and its Fundamental Principles 39

2.5 Rights and duties of medical personnel according to IHL 39

2.6 Responsibility of States 41

2.7 Reality check: some people do not follow the rules 43

2.8 The neutrality of a National Red Cross/Red Crescent Society 44

2.9 The role and mandate of the ICRC in situations of armed conflict 45
2.9.1 Health services: assistance to the war-wounded and sick 46

ANNEX 2. A The distinctive emblems 48

ANNEX 2. B The International Red Cross and Red Crescent Movement 50

2.1 Historical introduction

> "You may not be interested in war. War is interested in you."
>
> Leon Trotsky

One of the specificities of the war-wounded and those who care for them is their relationship to international humanitarian law, mainly the Geneva Conventions and their Additional Protocols.

In 19th century Europe, large-scale battles caused real carnage on the battlefields. Soldiers were regarded as cannon fodder and almost no medical services were available. Solferino, a town in northern Italy, was the site of one of these awful battles in 1859: in 16 hours 40,000 people were killed or wounded. Many wounded soldiers were left to die on the battlefield, although many of them could have been saved had relief been available. The medical services of the armies were too limited to respond to such a number of casualties – there were more veterinarians for the horses than doctors for the soldiers! Wounded soldiers lay agonizing for days without help.

Henry Dunant, a Swiss businessman travelling in the area, was struck by this drama. Out of compassion, he spontaneously organized help with the women from the neighbouring villages. Relief was given to all wounded soldiers without discrimination, regardless of their nationality. Others had performed humanitarian deeds on battlefields before; Dunant's genius lies in having taken things a lot further. Back in his hometown, Geneva, still traumatized, he wrote a book, *A Memory of Solferino*, published in 1862, in which he not only related the horrible battle and the suffering of the soldiers, but also launched an appeal around a vision and two key ideas.

Carlo Bossoli, Museo Nazionale del Risorgimento, Turin

Figure 2.1

The battle of Solferino, 1859.

The vision was to provide neutral medical care for wounded soldiers in the field, and the ideas to accomplish this vision were the following.

- On the one hand, to create a relief society in every country which, already in peacetime, would train volunteers who could assist the armed forces' medical services, doctors and nurses ready to care for wounded soldiers in the event of war – this later gave birth to National Red Cross and Red Crescent Societies.
- On the other hand, to negotiate an international agreement that would grant protection and assistance to the sick and wounded and the medical services which cared for them, thus guaranteeing access to the wounded. This idea became the first Geneva Convention, the basis for modern international humanitarian law (IHL). Every human society throughout history has had rules about the conduct of warfare: this was the first concerted attempt to standardize and institutionalize, on an international basis, the law of war. IHL is also referred to as the "law of war" or the "law of armed conflict".

Five citizens of Geneva, including Dunant, members of a charitable association, created the "International Committee for the Relief of Wounded Soldiers" in 1863 in response to Dunant's book. This Geneva Committee persuaded the Swiss government to convene a diplomatic conference in 1864, in order to formalize the protection of medical services on the battlefield through an internationally recognized treaty: twelve governments participated. The result was the Geneva Convention of 22 August, 1864, for the Amelioration of the Condition of the Wounded in Armies in the Field. The Geneva Committee became the International Committee of the Red Cross, and the emblem of the red cross was adopted as a symbol of protection of the medical services that cared for the sick and wounded – the symbols of the red crescent, the red sun and lion, and red crystal were introduced later. Not only was this first Geneva Convention a decisive step towards formalizing the laws of armed conflict, but it also made it the duty of signatory States to create military medical services to care for their own wounded. Soldiers were no longer to be simply regarded as cannon fodder.

2.2 International humanitarian law: basic principles

Throughout history, humanity has known war. All human societies have developed customary rules that regulate how wars are fought. Over 500 cartels, codes of conduct, covenants and other texts designed to regulate hostilities had been recorded before the advent of contemporary humanitarian law. The first laws of war were proclaimed by major civilizations several millennia before our era: "I establish these laws to prevent the strong from oppressing the weak" declared Hammurabi, King of Babylon.

Just as there is no society of any sort that does not have its own set of rules, so there has never been a war that did not have some vague or precise rules covering the outbreak and end of hostilities, as well as how they were to be conducted.

The Geneva Conventions (GCs) as they stand today are the result of a long process. Over the years, the First Geneva Convention has been extended to meet the changing needs of modern warfare. Three other Conventions were adopted, one after the other, covering victims other than wounded soldiers: shipwrecked sailors, prisoners of war, and civilian populations. The four Geneva Conventions of 1949 and their two Additional Protocols of 1977, in particular, as well as other written treaties and customary law, constitute international humanitarian law (IHL): i.e. what is and what is not permitted during international and non-international conflicts. All States in the world have now ratified the Geneva Conventions, which means that they are bound by these legal instruments.

The four Geneva Conventions are mainly designed to regulate the behaviour of combatants and to protect persons who are not (or are no longer) taking part in hostilities in armed conflicts between States (international conflicts).

I. Convention for the amelioration of the condition of the wounded and sick in armed forces in the field (GC I: Revision of the 1864 Convention).

II. Convention for the amelioration of the condition of wounded, sick and shipwrecked members of armed forces at sea (GC II: revision of the 1899 Convention).

III. Convention relative to the treatment of prisoners of war (GC III: revision of the 1929 Convention).

IV. Convention relative to the protection of civilian persons in time of war (GC IV: new Convention adopted in 1949).

Article 3 common to the four Geneva Conventions of 1949 also provides basic rules applicable in armed conflicts "not of an international character occurring in the territory of one of the High Contracting Parties".

In response to new needs in contemporary armed conflicts, the Conventions were developed and supplemented in 1977 by two further agreements: the Additional Protocols I and II relating to the protection, respectively, of victims of international and internal armed conflicts. A third Additional Protocol was adopted in 2005 to introduce a new protective emblem called the red crystal in addition to the red cross and red crescent emblems.

IHL regulates humanitarian issues in times of war; it aims to define the rights and obligations of the parties to a conflict in the conduct of warfare and to protect persons who are not (or are no longer) taking part in hostilities (civilians, wounded or sick soldiers, prisoners of war). These persons must be respected, protected and treated humanely by all parties. All the wounded and sick must also receive the medical care required by their condition. Those who provide such care must be respected and protected, as long as they are involved in their humanitarian tasks.

The various treaties defining IHL have become intricate, but the underlying message is simple. The human dignity of all individuals must be respected at all times without any kind of discrimination. Everything that *can* be done *should* be done to alleviate the suffering of those "*hors de combat*": persons who take no direct part in the conflict or who have been put out of action by sickness, injury, or captivity.

ICRC

Figure 2.2

I. Amelioration of the condition of the wounded and sick in armed forces in the field.

ICRC

Figure 2.3

II. Amelioration of the condition of wounded, sick and shipwrecked members of armed forces at sea.

ICRC

Figure 2.4

III. Treatment of prisoners of war.

ICRC

Figure 2.5

IV. Protection of civilian persons in time of war.

International humanitarian law: main legal sources

- Customary law
- Four Geneva Conventions of 1949 (194 States party to the Conventions)
- Two Additional Protocols of 1977
- Third Additional Protocol of 2005
- The Hague Conventions of 1899 and 1907
- United Nations Convention on Conventional Weapons of 1980
- Mine Ban Convention of 1997
- Cluster Munitions Convention of 2008

2.2.1 Principles underlying IHL: the "law of war"

- The human dignity of all individuals must be respected at all times.
- Persons no longer involved in the fighting (sick, wounded and shipwrecked combatants, and prisoners of war) and those who do not take a direct part in hostilities (civilians) are entitled to respect for their lives and physical and moral integrity. They shall in all circumstances be protected and treated humanely without any adverse distinction.
- The wounded and sick shall be collected and cared for.
- Captured combatants and civilians under the authority of an adverse party are entitled to respect for their lives, dignity, personal rights and convictions. It is forbidden to kill or injure an enemy who surrenders.
- Everyone shall be entitled to benefit from fundamental judicial guarantees. No one shall be held responsible for an act he has not committed. No one shall be subjected to physical or mental torture, corporal punishment or cruel and degrading treatment. Hostage taking is prohibited.
- The choice of methods and means of warfare is not unlimited, and must be proportional to the military goals sought. It is prohibited to use weapons and methods of warfare that cause unnecessary losses, superfluous injury or unnecessary suffering.
- Attacks shall distinguish between the civilian population and combatants and between civilian objects and military objectives. Accordingly, operations shall be directed only against military objectives.
- Medical personnel have rights and obligations in times of armed conflict.

British Museum

Figure 2.6

Soldiers blinded by chemical weapons during World War I: an example of means of warfare that cause superfluous injury and unnecessary suffering.

2.2.2 IHL and human rights law

IHL applies in situations of armed conflict, whereas human rights, or at least some of them, protect the individual at all times, in war and peace alike. Some human rights treaties permit governments to derogate from certain rights in situations of public emergency and if strict conditions are fulfilled. However, certain fundamental human rights must be respected in all circumstances and may never be waived, regardless of the emergency: in particular the right to life, the prohibition of torture and inhuman punishment or treatment, slavery and servitude, and the principle of legality and non-retroactivity of the law. No derogations are permitted under IHL because it was conceived for emergency situations, namely armed conflict.

IHL mainly aims to protect people who do not or are no longer taking part in hostilities. The rules embodied in IHL impose duties on all parties to a conflict, including non-governmental groups. Human rights, being tailored primarily for peacetime, apply to everyone. Their principal goal is to protect individuals from arbitrary behaviour by their own governments. Human rights law does not deal with the conduct of hostilities.

2.3 The distinctive emblems

The distinctive emblems of the red cross, red crescent, and red crystal are meant to mark certain medical and religious personnel and equipment which must be respected and protected during armed conflict (protective use). They also serve to show that persons or objects are linked to the International Red Cross and Red Crescent Movement including in situations other than armed conflict (indicative use). Their use is strictly defined (see Annex 2. A: The distinctive emblems).

Use of the emblems is restricted to the following entities.

1. Medical staff and services of the armed forces are the very first persons authorized to use the distinctive emblem, as a sign of protection.
2. Volunteers of a National Red Cross or Red Crescent Society, when duly authorized by the competent State Authority.
3. Staff of the International Committee of the Red Cross and the International Federation of Red Cross and Red Crescent Societies.

Persons and objects displaying the emblems must not be attacked, but on the contrary must be respected, protected, and encouraged in their work.

V. Louis / ICRC

Figure 2.7
Armed forces' medical services.

B. Fjortoft / ICRC

Figure 2.8
National Society volunteers.

B. Heger / ICRC

Figure 2.9
ICRC, International Federation and National Society staff.

2.4 The International Red Cross and Red Crescent Movement and its Fundamental Principles

The International Committee of the Red Cross and the International Federation of Red Cross and Red Crescent Societies, together with the National Red Cross and Red Crescent Societies, form the International Red Cross and Red Crescent Movement.

Fundamental Principles of the Red Cross and Red Crescent Movement[1]

Humanity

Impartiality

Neutrality

Independence

Voluntary service

Unity

Universality

In peacetime, as in time of conflict, it is the duty of the armed forces and the National Society of every country to disseminate the rights and duties enshrined in IHL. This is necessary not only for these rules to become known, understood, accepted and respected by combatants, but also so that the whole population understands and supports its Red Cross or Red Crescent Society for it to become more efficient, for the benefit of all, in times of conflict and in times of peace. Part of this dissemination includes the respect due to the red cross, red crescent and red crystal emblems as a sign of protection.

2.5 Rights and duties of medical personnel according to IHL

IHL provides medical personnel with rights in times of armed conflict, but also assigns duties to them. The duties incumbent on them are directly linked to the rights of the protected persons placed in their care. These provisions are a case-specific refinement of the basic rights and duties defined by medical ethics and the Hippocratic oath. Medical personnel are bound by medical ethics and IHL to treat patients solely on the basis of need and without regard for their nationality, race and class, religious or political beliefs. These rights and duties have been defined to enable medical personnel to perform the humanitarian task entrusted to them, born of the desire to alleviate human suffering by helping and tending the wounded and the sick, which is the very purpose of the medical mission.

The Geneva Conventions define medical personnel as:

- persons assigned by a party to the conflict, whether on a permanent or temporary basis, exclusively to medical purposes (search, collection, transportation, diagnosis and treatment of the wounded and sick, and for the prevention of disease) – this includes doctors, nurses, orderlies, first aiders, and stretcher-bearers;
- persons assigned by a party to the conflict, whether on a permanent or temporary basis, exclusively to the administration or operation of medical units or medical transport – this includes administrators, drivers, cooks, mechanics, etc.

The term "medical personnel" is thus not confined to the narrow meaning of the word. All personnel required to ensure the adequate treatment of the wounded and sick are covered by the protection granted to them, as long as they form an integral part of the medical service.

1 See Annex 2. B: The Fundamental Principles of the International Red Cross and Red Crescent Movement.

Medical units, whether military or civilian, include what is set up for medical purposes:
- all buildings or installations (hospitals, clinics, first-aid posts, field hospitals, tents, etc.),
- blood transfusion and preventive medicine centres,
- medical and pharmaceutical depots and stores.

They may be fixed or mobile, permanent or temporary.

Medical equipment includes:
- stretchers,
- medical and surgical appliances and instruments,
- medical supplies, dressings, etc.

Medical transport can be organized by land, water, and air:
- ambulances, lorries or trucks,
- hospital ships, rescue craft,
- medical aircraft, etc.

Medical personnel must not be the object of attack, and must be allowed to care for the sick and wounded unhindered. They should wear the distinctive emblem of the red cross, red crescent or red crystal on a white ground and carry an identity card. Military medics may bear arms only for their own defence and to defend the wounded and sick from pillage. However, they must not use arms to prevent capture of their patients, units or themselves by the enemy; by doing so they lose their protected status.

If medical personnel fall under the control of enemy troops, they shall be allowed to continue their duties towards the wounded and sick. They shall not be compelled to perform acts contrary to medical ethics, or refrain from performing acts required by medical ethics. A patient remains a patient, and medical personnel have a responsibility to provide care according to their training and the means available. If captured medical personnel are not indispensable to the care of other prisoners, they should be repatriated. Those retained shall not be considered prisoners of war and their work shall be facilitated.

In occupied territory, civilian medical personnel should be allowed to continue providing adequate medical services for the civilian population.

The civilian population shall respect the wounded and sick, even if they belong to the enemy, and shall commit no act of violence against them. Civilians are permitted to collect and care for the wounded and sick of whatever nationality, and shall not be penalized for doing so. On the contrary, they must be aided in this work.

Medical units enjoy protected status so long as they are not used to commit acts harmful to the enemy, such as sheltering able-bodied combatants, storing arms or ammunition, or being used as military observation posts. Otherwise, their protection ceases and they become legitimate military targets. This is why strict controls must be established to safeguard the protected status of medical units and transport.

To enhance the protection of medical units and medical transport, they should be clearly marked with the red cross, red crescent or red crystal emblem, of the largest size possible. The emblem is the visible sign of protection conferred by the Geneva Conventions and their Additional Protocols.

2.6 Responsibility of States

As with any international agreement, the governments of States have a distinct responsibility whenever they become party to a treaty.

- The Geneva Conventions and their Additional Protocols are both contracts with other States and commitments towards humanity by which governments have agreed to certain rules that regulate the conduct of armed hostilities and the protection of persons who are not (or are no longer) taking part in hostilities, what is known as "law in war", or *jus in bello*. Its provisions apply to the warring parties regardless of the reasons for the conflict and whether or not the cause upheld by either party is just.
- On becoming party to the four Geneva Conventions of 1949, States agree to respect and ensure respect for their obligations in all circumstances (common Art. 1).
- Reciprocity is not supposed to be a prerequisite to a country adhering to and applying these rules, but it exists in reality and practice. It is in everyone's interest that everyone apply the law of his or her own accord.
- States are responsible for protecting those who do not participate or are no longer participating in hostilities.
- During peacetime, States are to promote IHL and train the members of their armed forces to respect IHL.
- At all times, States must take all measures necessary for the prevention and repression of any misuse or abuse of the distinctive emblems.
- On becoming party to the Geneva Conventions, States undertake to enact any legislation necessary to punish persons guilty of grave breaches of them, and are also bound to prosecute in their own courts any person suspected of having done so, or to hand over that person to another State for judgment. In other words, perpetrators of grave breaches, i.e. war criminals, must be prosecuted at all times and in all places, and States are responsible for ensuring that this is done.
- A number of independent mechanisms have been set up, each within its mandate and competency, to promote better compliance with IHL, including the ICRC, the International Fact-finding Commission, regional and United Nations bodies, and the International Criminal Court (Rome Statute of 1998). Their adequacy and sufficiency in meeting the task remains a challenge to humanity.
- Generally speaking, a State's criminal laws apply only to crimes committed on its territory or by its own nationals. International humanitarian law goes further in

ICRC

Figure 2.10

All too often the red cross emblem is used to indicate any health service without regard to its privileged legal status, which confers protection.

that it requires States to seek out and punish any person who has committed a grave breach, irrespective of his nationality or the place where the offence was committed. This principle of universal jurisdiction is essential to guarantee that grave breaches are effectively repressed. Such prosecutions may be brought either by the national courts of the various States or by an international authority. In this connection, the International Criminal Tribunals for the former Yugoslavia and Rwanda were set up by the United Nations Security Council in 1993 and 1994, respectively, to try those accused of war crimes committed during the conflicts in those countries.

- Lastly, the international community has created a permanent International Criminal Court (Rome Statute of 1998), which is competent to try war crimes, crimes against humanity, and genocide.
- The United Nations Security Council is the main UN body responsible for the maintenance of international peace and security. To this end, it may decide on measures, including coercive ones, against any State threatening or breaching international peace (Chapters VI and VII of the United Nations Charter). The rules that outlaw war, with some exceptions, are know as "the law on the use of force" or the *jus ad bellum*. This is different from the *jus in bello*, or IHL, which is the law applicable in times of armed conflict.
- "New" or "anarchic" conflicts lead to the partial, and sometimes total, weakening or breakdown of State structures. In such situations, armed groups take advantage of the political vacuum in an attempt to grab power. It is not because a State's structures have been weakened or are non-existent that there is a legal vacuum with regard to international law. The law, and its obligations, remain.

Admittedly, humanitarian rules are harder to apply in these types of conflict. The lack of discipline among belligerents, the arming of the civilian population as weapons flood the territory, and the increasingly blurred distinction between fighters and civilians often cause confrontations to take an extremely brutal turn – in which there is little room for respect for IHL.

As a result, this is the type of situation in which particular efforts are needed to make people aware of humanitarian law. Better knowledge of the rules of law will not solve the underlying problem which led to the conflict, but it is likely to attenuate its deadlier consequences.

Insofar as a "right – or even a duty – to intervene" is tantamount to justifying armed intervention undertaken for humanitarian reasons, this is a matter not for humanitarian law but for the rules on the legality of the use of armed force in international relations, i.e. of *jus ad bellum*, according to the Charter of the United Nations.

R. Sidler / ICRC

Figure 2.11

This ICRC car was targeted by "uncontrolled elements".

2.7 Reality check: some people do not follow the rules

Given what the law dictates, what is the reality on the battlefield? Whether in conflict situations or in peacetime, and whether national or international law is applicable, laws are violated and crimes committed. There are many examples of violations of IHL: surrounding a military objective with medical units so that it will not be targeted; hiding weapons in a hospital; transporting able-bodied combatants in an ambulance; using an aircraft displaying the emblem for reconnaissance missions; not respecting the non-combatant status of sick and wounded soldiers – on the contrary, all too often their survival is seen as an invitation by some to "finish the job", and atrocities are the result. All too often hospitals and medical personnel are the target of attack, or are prevented from doing their duty of caring for the sick and wounded, which is perceived by some as giving "succour and aid to the enemy". All these violations have three things in common: they seriously weaken the system of protection embodied in IHL, they divert people and objects displaying the red cross, red crescent or red crystal from their humanitarian purpose, and they put lives in danger by fostering mistrust.

T. Pizer / ICRC

Figure 2.12

Unfortunately, hospitals are not immune to attack: this photo shows a blatant contravention of IHL.

Red Cross and Red Crescent personnel are not immune to this lack of respect for IHL. Mussolini's air force bombed a Swedish Red Cross ambulance during the invasion of Abyssinia on 30 December 1935, and 28 people were killed and 50 wounded. More recently, masked gunmen entered the residence of the ICRC hospital in the village of Novye Atagi, Chechnya (southern Russia), in December 1996 and assassinated five nurses and the construction engineer in cold blood. The authors of this text have all too often been threatened, prevented from reaching and treating the victims, or seen their hospital attacked or bombed. The world over, colleagues, both civilian and military, are time and again made to suffer "simply" because they want to fulfil their humanitarian, ethical, and legal duty to care for their patients.

Humanitarian action is based on a fundamentally "optimistic philosophy", wrote Jean Pictet.[2] But this optimism in no way detracts from the "philosophy's" realism. It is aware that humanitarian work is difficult. Its greatest enemies may well be neither weapons nor disaster, but selfishness, indifference and discouragement. One should not despair, however. Rather one should see this as an incentive to redouble efforts to educate those involved in armed conflict. Medical ethics go hand in hand with IHL and the Fundamental Principle of impartiality: the sick and wounded are entitled to treatment, whatever their origin or political persuasion.

2 Jean Pictet (1914 – 2002) was Director-General and Honorary Vice-President of the ICRC. He was responsible for the preparatory works that led to the 1949 revision of the Geneva Conventions and the Additional Protocols of 1977, and his writings formed the basis of the Fundamental Charter of the International Red Cross and Red Crescent Movement, adopted in 1965.

All medical personnel should not only comply with medical ethics and IHL but should also disseminate their meaning to military and civilian authorities, ordinary soldiers and citizens.

Medical practitioners have a particularly important part to play: they have responsibilities with regard to individual patients but also towards society at large. They must convince combatants of the need to respect IHL, to refrain from attacking the medical facilities and personnel of the enemy and allow them to perform their humanitarian task. While the underlying philosophy of the Movement is *not* rooted in the principle of reciprocity, in practice medical professionals can only expect to benefit from the "protection" offered by the Geneva Conventions – in a general atmosphere of respect for the Fundamental Principles – if their colleagues practising on the "other" side also benefit from the same protection. They must inform those who bear weapons of their obligations and trust that colleagues in contact with the opposing side are doing the same. This has proved true in the past, and soldiers have often respected the principle of protection afforded to the sick and wounded and medical and religious personnel.

2.8 The neutrality of a National Red Cross/ Red Crescent Society

Another major problem faced during armed conflicts of a non-international character, that is to say internal conflicts, is the question of neutrality, especially of a National Red Cross or Red Crescent Society. The requirement of non-discrimination is of particular concern to Red Cross/Red Crescent Societies, it is in fact a condition for their recognition. They must be open to all those who wish to become members and must permit all social, political and religious groups to be represented; this representativity is the guarantee of the Societies' ability to engage in exclusively humanitarian activities and to resist all partisan considerations.

It is admittedly not always an easy task to apply the principle of neutrality, not least because everyone has personal convictions. When tension mounts and passions are aroused, every member of the Red Cross or Red Crescent must exercise great self-control and refrain from expressing his/her opinions in the discharge of his duties. Volunteers are not asked to be neutral – everyone is entitled to an opinion – but to act neutrally. That is an important distinction. The next difficulty is the fact that the parties to the conflict often take a dim view of neutral behaviour. In countries where an internal conflict is taking place, the armed forces fail to understand why the National Society does not condemn the activities of those they regard as "bandits", much less why it wants to provide assistance to any of their number no longer able to fight. As for the opposition, they are critical of the Society's connections with the authorities.

Anyone trying to work on both sides to help non-combatants is considered at best naïve, at worst a traitor. The extremely polarized nature of many struggles is such that *not taking a stand* is a hostile act in itself. This is why the Red Cross and Red Crescent's neutrality and impartiality must be explained. As one National Society first aider put it: "The best argument I have is to tell one of the parties to the conflict that if I take its side and ignore the victims on the other side, I will never again be able to bring help to its own wounded members".

It would appear, indeed, that the principles of neutrality and impartiality are ideals to be attained, inner qualities that are rarely inborn but that most often require one to overcome one's instincts. They demand from members of the Red Cross and Red Crescent Societies arduous and sustained efforts to overcome their own prejudices and preferences in order to be able to perform the purest act of impartiality, which is to give more help to the adversary who is the victim of great misfortune than to the friend whose suffering is less severe, or to care for the more severely wounded, even if guilty, before assisting the innocent whose injuries are slighter.

2.9 The role and mandate of the ICRC in situations of armed conflict

Established in 1863, the International Committee of the Red Cross (ICRC) is an impartial, neutral, and independent organization whose exclusively humanitarian mission is to protect the lives and dignity of the victims of war and internal violence and to provide them with assistance. The ICRC endeavours to prevent suffering by promoting and strengthening international humanitarian law and universal humanitarian principles.

The ICRC in a nutshell

Neutral, impartial and independent humanitarian organization.

Works in situations of armed conflict.

Entrusted with an international mandate under the Geneva Conventions.

Protects and assists victims of conflict.

Part of the International Red Cross and Red Crescent Movement.

The ICRC has been named the "guardian and promoter" of IHL by the States party to the Geneva Conventions. It works during armed conflict to protect and assist the victims: the wounded and sick, prisoners of war and other detainees, and the civilian population.

The ICRC does not take sides or determine right or wrong in a conflict. The neutrality of the ICRC, however, is not a widely appreciated principle. There are many who express indignation at its neutrality in the mistaken belief that neutrality betokens lack of commitment and courage. The ICRC, for its part, has great difficulty in getting the parties to a conflict to understand that the only thing it must grant equally to each is its willingness to serve, and that in other respects its activities are proportional to the needs, and consequently unequal when distress is greater on one side than on the other.

At the beginning of hostilities, the ICRC:
- reminds the belligerents of their obligations according to IHL;
- offers its services as a neutral intermediary for the accomplishment of humanitarian activities;
- makes an assessment of the needs;
- acts in favour of victims according to the needs.

The protection role the ICRC plays consists in acting in defence of people who are not, or no longer, taking part in combat: the wounded or sick, detainees or prisoners of war, and civilians, including the inhabitants of territories administered or occupied by a hostile power. The ICRC approaches the competent authorities to ensure that these victims receive humane treatment. The ICRC has the right to have access to prisoners of war (Third Geneva Convention), and detained persons covered by the Fourth Geneva Convention, including the right to visit detention camps.

The ICRC has an official policy of discretion. Only when it observes grave and repeated breaches of IHL and when its confidential representations have been in vain and it considers that the only means of helping the victims is to ask for the support of the international community, does it make public representations. This sometimes takes the form of an appeal to the States party to the Geneva Conventions, whose responsibility it is to respect and *ensure* respect for IHL. Such initiatives are nevertheless the exception rather than the rule.

ICRC activities

1. Protection during armed conflict:
   - protection of civilians confronted with the violence of armed conflict;
   - protection and care of the wounded, sick, and shipwrecked;
   - protection of prisoners of war and other detainees (registration and visits);
   - as a neutral intermediary, facilitation and accompaniment of the release of prisoners of war;
   - re-establishment of family links through Red Cross messages;
   - re-establishment of family links through the tracing agency for the missing and dead.
2. Assistance to victims of armed conflict – public health approach (shelter, water, sanitation, food, preventive and curative medical care):
   - economic security;
   - water and habitat;
   - health services (including assistance to the war-wounded and sick).
3. Preventive action:
   - promotion of international humanitarian law (IHL);
   - advice and assistance with national measures to implement IHL;
   - mine risk education programmes to raise awareness among people exposed to the dangers posed by mines and unexploded ordnance.

ICRC delegates must be willing to talk to all those who are responsible for violations of IHL and human rights. They cannot pass judgment on them publicly, but must speak to them on behalf of those to whom speech is denied, and who have no one else to turn to. They often do so at considerable risk to their own personal safety and their words may fall on deaf ears, but if this policy of refraining from public denunciation makes it possible to alleviate the suffering of just one man, woman or child, that is ample recompense.

As a neutral and independent institution, the ICRC is authorized by the Geneva Conventions and their Additional Protocols to administer first aid and other care in the field to victims of armed conflict. Military authorities are required to permit the ICRC to collect and care for the wounded or sick of whatever nationality, even in invaded or occupied territories. The ICRC can offer its services to the parties, in particular in the medical realm to establish neutral or hospital zones, set up hospitals for the sick and wounded, assist existing hospitals, and provide rehabilitation for amputees, especially victims of anti-personnel landmines.

The ICRC helps to organize, or provide directly, relief to victims of armed conflict. These relief supplies cover the most essential needs, such as food, clean drinking water, shelter, clothing, and medical care.

Figure 2.13
The assistance pyramid: public health analysis of population needs.

2.9.1 Health services: assistance to the war-wounded and sick

Although proper medical care of sick and wounded soldiers and civilians in times of armed conflict is taken for granted by many governments today, poverty may compromise a government's efforts to provide such care. The ICRC maintains a capacity to aid States whose authorities show a willingness to assume their responsibilities towards their own soldiers and the civilian population.

The ICRC has deployed many assistance and training programmes to this effect. One example of an offer of services by the ICRC to two countries involved in a conflict was the case of the 1998 – 2000 war between Ethiopia and Eritrea. The ICRC provided assistance to help both governments to develop the programmes described below.

ICRC EXPERIENCE

The ICRC carried out the following programmes in Eritrea and Ethiopia from 1998 to 2001.

Ethiopia

- Training in advanced first aid, triage and evacuation for the war-wounded.
- War surgery seminars.
- Specialist surgical training for the war-wounded: thoracovascular, orthopaedic and neurosurgery.
- Maxillo-facial reconstructive surgery training programme.
- Short-term ICRC surgical team presence in an emergency situation.
- Support to the Ethiopian Red Cross Society ambulance service.
- Surgical material and equipment.
- Visits to thousands of POWs and their repatriation after cessation of hostilities.
- Repatriation of thousands of interned enemy aliens.

Eritrea

- Training in advanced trauma nursing care for ambulance staff.
- Emergency Room Trauma Course for the war-wounded.
- War surgery seminar.
- Short-term ICRC surgical team presence in an emergency situation.
- Intensive care nursing training for the war-wounded.
- Development of a physiotherapy training curriculum at the Institute of Nursing Care.
- Support to the Red Cross Society of Eritrea ambulance service.
- Surgical material and equipment.
- Visits to thousands of POWs and their repatriation after cessation of hostilities.
- Repatriation of thousands of interned enemy aliens.
- Assistance to thousands of internally displaced persons (food, shelter, water, sanitation, medical care).

While the ICRC is authorized to care for the wounded and sick during armed conflict, it is also involved in post-conflict reconstruction, indeed in development assistance, and, sometimes, the ICRC achieves the right balance.

ICRC EXPERIENCE

"We want to thank the doctors and nurses of the ICRC. Thank you for coming. And, thank you for leaving."

Rui Paolo [3]

3 Rui Paolo, Director of Hospital Services, Ministry of Public Health, Dili, Timor Leste, June 2001, on the occasion of the handover of the administration of the Dili General Hospital from the ICRC to the Ministry of Public Health.

ANNEX 2. A The distinctive emblems

The distinctive emblems of the red cross, red crescent and red crystal on a white ground are the symbols of impartial humanitarian work and do not represent any particular religious belief. They provide protection for military medical services and relief workers in armed conflicts. Moreover, they are also used for identification purposes by National Societies of the Red Cross and Red Crescent Movement in each country.

In the event of armed conflict, the distinctive emblems are a visible sign of the protection conferred by international humanitarian law upon medical personnel and equipment. The persons and objects displaying them must not be attacked, but on the contrary must be respected and protected.

Their use as a *protective sign* during armed conflict is authorized exclusively for:

- medical units, transport and personnel, as well as religious personnel, of the armed forces;
- civilian medical units**,** transport, and personnel, as well as civilian religious personnel, that have received special permission by the competent authorities to use the emblem;
- medical units, transport and personnel that a Red Cross or Red Crescent National Society has put at the disposal of the medical services of the armed forces.

Persons and buildings/structures/objects displaying the emblem must not be attacked, damaged or prevented from operating, but, on the contrary, must be respected and protected, even if, for the moment, they are not caring for or housing either wounded or sick people. The perfidious use of the emblems is explicitly prohibited.

In order to ensure effective protection during wartime, the *indicative* use of the emblem must be *strictly controlled* and only used for:

- Red Cross or Red Crescent National Societies, in order to indicate that persons or goods have a connection with the Society in question (it must be small in size to avoid any confusion with the emblem used as a protective symbol);
- exceptionally, ambulances and first-aid stations exclusively assigned to the purpose of giving free treatment to the wounded and sick, with the authorization of a National Society.

The International Committee of the Red Cross and the International Federation of Red Cross and Red Crescent Societies are authorized to use the emblem for all their activities and at all times.

Misuse of the emblem is a serious problem. In peacetime, hospitals, clinics, doctors' offices, pharmacies, non-governmental organizations and commercial companies tend to use the emblem in order to benefit from its reputation, although they are not entitled to do so. This abuse clearly weakens the protective value of the emblem in wartime.

Any case of misuse of the emblem should be reported to the relevant Red Cross or Red Crescent National Society, the ICRC, or the International Federation of Red Cross and Red Crescent Societies.

Parties to the Geneva Conventions shall take the precautions necessary for the prevention or repression of any abuse of the distinctive emblems.

During peacetime staff and volunteers of the International Red Cross and Red Crescent Movement, through their behaviour, activities and awareness-raising efforts, seek to ensure that the protective value of the distinctive emblems is well known to the military and the general public.

Please note:

On 8 December 2005, a Diplomatic Conference adopted Protocol III additional to the Geneva Conventions, which recognizes an additional distinctive emblem. The "Third Protocol emblem", also known as the red crystal, is composed of a red frame in the shape of a square on edge on a white background. According to Protocol III, all distinctive emblems enjoy equal status.[4] The conditions for use of and respect for the Third Protocol emblem are identical to those for the distinctive emblems established by the Geneva Conventions and, where applicable, the 1977 Additional Protocols.

J. Perez / International Federation

4 While no longer in use, the red lion and sun on a white background is still recognized by the Geneva Conventions.

Annex 2. B The International Red Cross and Red Crescent Movement

The International Committee of the Red Cross and the International Federation of Red Cross and Red Crescent Societies, together with the National Red Cross and Red Crescent Societies, form the International Red Cross and Red Crescent Movement.

Born of the compassion felt by a Swiss citizen, Henry Dunant, at the sight of the dead and wounded abandoned on the battlefield of Solferino, the International Red Cross and Red Crescent Movement believes that its first duty is to render wars that cannot be avoided less inhuman, and to alleviate the suffering that they cause. Its aim has been to bring a little bit of humanity into the horrors of war. The Movement was founded as a result of a conflict and for conflicts, with the aim of helping people in distress on the battlefield.

The International Red Cross and Red Crescent Movement endeavours, in its international and national capacity, to prevent and alleviate human suffering wherever it may be found. Its purpose is to protect life and health and to ensure respect for human beings. It promotes mutual understanding, friendship, cooperation and lasting peace amongst all peoples.

The humanitarian ideas of the Movement are reflected in the seven Fundamental Principles that guide the activities of all its components at all times: humanity, impartiality, neutrality, independence, voluntary service, unity and universality.

Components of the International Red Cross and Red Crescent Movement

The International Committee of the Red Cross (ICRC)
The International Committee of the Red Cross, created in Geneva, Switzerland in 1863, is the founding body of the Red Cross/Red Crescent Movement. It is an independent humanitarian organization. As a neutral intermediary and on the basis of the Geneva Conventions or the customary law of nations, which grant it the right of initiative, it endeavours to protect and assist the military and civilian victims of international and non-international armed conflict and internal disturbances and tension.

The countries of the world have entrusted to the ICRC the roles of promoter and custodian of international humanitarian law, and the task of working for their development and worldwide dissemination.

The functions of the ICRC are defined in its own statutes, in those of the Red Cross/ Red Crescent Movement, and also in the international treaties known as the Geneva Conventions of 1949 and their Additional Protocols of 1977.

The International Federation of Red Cross and Red Crescent Societies
The International Federation is the permanent liaison body between National Red Cross and Red Crescent Societies. It attempts to prevent and alleviate human suffering through the promotion of activities by National Societies and so contribute to peace. The International Federation encourages the creation and assists the development of National Societies in providing services to the community.

The International Federation organizes and coordinates international relief efforts for victims of natural disasters and promotes the adoption of national disaster-preparedness plans. First aid is an important part of both everyday community service and disaster preparedness.

The National Red Cross/Red Crescent Society
There are 186 National Red Cross/Red Crescent Societies around the world, with more currently being created. From the outset, the goal of the Red Cross/Red Crescent Movement has been to create relief societies which, in their capacity of *auxiliaries to their countries' armed forces' medical services*, would be called upon to tend wounded or sick soldiers. The establishment of such Societies was consistent with the aims of the original Geneva Convention of 1864.

The National Societies' activities have developed steadily and diversified over the years. The Societies' concern, at first directed only at members of the armed forces, now extends to protection and assistance of all those, whether military or civilian, who are the victims of conflicts, in close cooperation with the ICRC, and of natural catastrophes, in close cooperation with the International Federation of Red Cross and Red Crescent Societies. National Societies have also developed many activities in peacetime as *auxiliaries to the public authorities* to alleviate suffering, improve health, and prevent disease.

National Society membership is open to everyone and services are provided on the sole criterion of need. National Societies must fulfil stringent conditions to achieve recognition by the ICRC and obtain International Federation membership. Among these conditions: respect for the Fundamental Principles, and recognition by their home government as a voluntary aid society, auxiliary to the public authorities.

Each National Red Cross/Red Crescent Society has its own particular characteristics, and its activities cover a wide range depending on the needs of the country and on the Society's operational capabilities. The one activity that all National Societies share is the provision and teaching of first aid.

The Fundamental Principles of the International Red Cross and Red Crescent Movement

Humanity
The International Red Cross and Red Crescent Movement, born of a desire to bring assistance without discrimination to the wounded on the battlefield, endeavours – in its international and national capacity – to prevent and alleviate human suffering wherever it may be found. Its purpose is to protect life and health and to ensure respect for the human being. It promotes mutual understanding, friendship, cooperation and lasting peace amongst all peoples.

Impartiality
It makes no discrimination as to nationality, race, religious beliefs, class or political opinions. It endeavours to relieve the suffering of individuals, being guided solely by their needs, and to give priority to the most urgent cases of distress.

Neutrality
In order to continue to enjoy the confidence of all, the Movement may not take sides in hostilities or engage at any time in controversies of a political, racial, religious or ideological nature.

Independence
The Movement is independent. The National Societies, while auxiliaries in the humanitarian services of their governments and subject to the laws of their respective countries, must always maintain their autonomy so that they may be able at all times to act in accordance with the principles of the Movement.

Voluntary service
It is a voluntary relief movement, not prompted in any manner by the desire for gain.

Unity
There can be only one Red Cross or one Red Crescent Society in any one country. It must be open to all. It must carry out its humanitarian work throughout its territory.

Universality
The International Red Cross and Red Crescent Movement, in which all Societies have equal status and share equal responsibilities and duties in helping each other, is worldwide.

Humanity and Impartiality express the objectives of the Movement.

Neutrality and Independence ensure access to those in need of help.

Voluntary service, Unity and Universality enable the International Red Cross and Red Crescent Movement to work effectively all over the world.

Chapter 3

MECHANISMS OF INJURY DURING ARMED CONFLICT[1]

1 Acknowledgement: the ballistics portion of this chapter is largely based on the work of Doctor of Forensic Science Beat Kneubuehl, Institute of Forensic Medicine, University of Bern, in collaboration with Armasuisse, Centre for Military and Civilian Systems, Science and Technology, Ballistics Laboratory, Thun, Switzerland. His cooperation and collaboration with ICRC surgeons over the years have permitted a whole generation of war surgeons to gain an important insight into the wounding potential of weapons. We hope that this knowledge has permitted better surgical management of the victims of armed conflict and other situations of violence where the weapons of war are used.

3. MECHANISMS OF INJURY DURING ARMED CONFLICT

3.1	The various mechanisms of injury	55
3.1.1	Blunt injury in war	55
3.1.2	Penetrating wounds caused by weapons of war	55
3.1.3	Anti-personnel landmine wounds	57
3.1.4	Explosive blast injuries	58
3.1.5	Burns	59
3.1.6	Non-conventional weapons	59
3.2	Ballistics	59
3.2.1	Introduction	59
3.2.2	Internal ballistics	60
3.2.3	External ballistics	62
3.2.4	Summary	62
3.3	Terminal ballistics	63
3.3.1	Role of kinetic energy	63
3.3.2	Laboratory experiments	63
3.3.3	Non-deforming rifle bullet: full metal jacket (FMJ) military bullet	64
3.3.4	Deforming and fragmenting rifle bullets: dum-dum	66
3.3.5	Handgun bullets	68
3.3.6	Fragment projectiles	69
3.4	Wound ballistics	70
3.4.1	Laboratory studies *versus* human body	70
3.4.2	Projectile–tissue interaction	70
3.4.3	Tissue factors	71
3.4.4	Pathological description of the permanent wound cavity	72
3.4.5	Clinical applications	73
3.4.6	Sonic shock wave in tissues	76
3.4.7	Pressure wave in tissues and blood vessels	76
3.4.8	Fragment wounds	77
3.5	Wound dynamics and the patient	78

3.1 The various mechanisms of injury

The surgeon caring for the war-wounded will face a wide variety of injuries. During armed conflict, all the usual trauma of peacetime continues unabated and natural catastrophes occur as well. Armed conflict itself provokes blunt injuries and burns, and trauma that are specific to weapons and the circumstances of warfare. More specifically, combat involves penetrating and blast injuries; these will be the major focus of this book.

Modern armed conflict causes specific penetrating and blast injuries, as well as blunt and burn trauma.

3.1.1 Blunt injury in war

Blunt trauma is common in war. Severe injury due to blunt trauma may be more difficult to detect than in cases of penetrating trauma, and X-ray diagnosis and other more sophisticated technologies are more valuable for patient assessment.

Road traffic crashes
Military vehicles often drive at high speed over difficult terrain where safe roads do not exist. In addition, the environment of any accident may be hostile (presence of enemy forces, minefields, etc.).

Collapsed buildings and falls
The collapse of bombed buildings will cause blunt and crush injuries to anyone inside. People may fall from destroyed buildings or a balcony.

Explosions and anti-tank mines
A bomb or shell explosion can throw people against objects. A car, bus or lorry carrying passengers may drive over an anti-tank mine. The explosion of the mine overturns or destroys the vehicle and people are thrown out and onto the ground, thus suffering blunt injuries.

Beatings
Mistreatment of prisoners, officials, "suspected sympathizers" or other civilians is, alas, all too common.

3.1.2 Penetrating wounds caused by weapons of war

A moving projectile has kinetic energy. When the projectile enters the human body, it releases energy to the tissues thus causing a wound. There are two types of wounding projectiles: bullets and metallic fragments (or splinters, shards, and shrapnel)[2].

Fragment wounds
Exploding bombs, shells, rockets and grenades, submunitions (cluster bomblets) and some landmines, produce metal fragments from the weapon casing (primary projectiles). In the past these fragments were usually irregular in size and shape. In many modern weapons, however, the inner lining of the casing is scored, creating weak points that break off easily on explosion. This generates a controlled fragmentation of large numbers of pre-formed fragments that are regular in size and shape, and usually less than 1 g in weight. In other, usually improvised bombs and explosive devices, nails, bolts, steel pellets or other metallic debris surround the explosive material.

T. A. Voeten / ICRC

Figure 3.1
Different types of shells and explosives.

2 The word "shrapnel" is derived from the name of Major-General Henry Shrapnel (1761–1842), an English officer, who designed a new type of artillery shell. The term originally referred only to the metal balls dispersed when a shrapnel shell bursts, but is often used to describe metal fragments intentionally included in explosive devices. For shells, bombs or other munitions, the technical term for these particles is fragments, splinters or shards, fragments being the preferred name in scientific documents and the term that is used in this manual.

Fragments are shot off at a very high speed, which decreases rapidly with the distance travelled because of their non-aerodynamic character. The farther the victim is from the explosion, the more superficial the wounds. At very close range, multiple high-energy fragments combined with the blast effect create mutilating injuries and are often fatal.

Explosions may cause stones or bricks to break up, shatter glass panes, or the force of the blast wind may transport other debris, also producing penetrating fragments (secondary projectiles), as described in Section 3.1.4.

Fragment wounds are usually multiple and the wound tract is always widest at the entry. There may or may not be an exit wound.

Bullet wounds

Handguns and military assault rifles shoot bullets at high speed. Bullet wounds seen in the injured are usually single with a small entry; multiple bullet injuries are more likely to kill. There may be no exit wound but, if there is, the size is variable. The amount of tissue damage depends on a number of factors.

According to international humanitarian law – the Third Hague Convention of 1899 – all bullets used during armed conflict are supposed to remain intact. They are designed to wound, not kill. They have a full metal jacket (usually copper or plated steel) and should be so manufactured as to prevent any fragmenting when they hit the human body. Because of various ballistic effects, some bullets do break up into fragments in the body.

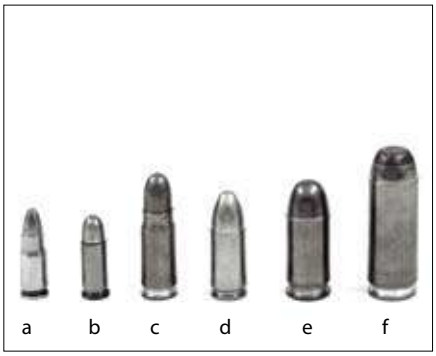

B. Kneubuehl

Figure 3.2.1

Examples of pistol ammunition:

a. 5.45 x 19 mm

b. 6.35 mm Browning

c. 7.63 mm Mauser

d. 9 mm Luger

e. 45 mm automatic

f. 50 mm AE (Action Express) semi-jacketed bullet

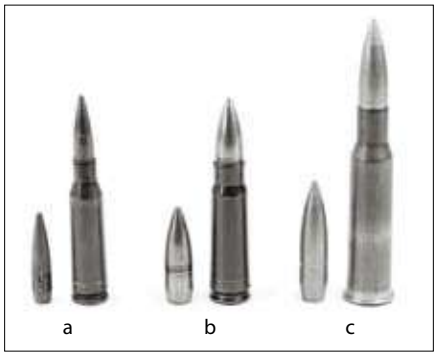

B. Kneubuehl

Figure 3.2.2

Examples of military rifle ammunition:

a. 5.45 x 39 mm Kalashnikov

b. 7.62 x 39 mm Kalashnikov

c. 7.62 x 54R Dragunov

B. Kneubuehl

Figure 3.2.3

Semi-automatic pistol 9 mm Luger (SIG-Sauer P 228).

B. Kneubuehl

Figure 3.2.4

7.62 x 39 mm AK-47 Kalashnikov military assault rifle.

Cut wounds, "*arme blanche*"

Apart from the modern soldier's bayonet, knives, machetes or *pangas*[3] may be used by combatants in "traditional" warfare in some societies. These produce incised or puncture wounds.

3.1.3 Anti-personnel landmine wounds

Anti-personnel landmines (APM) come in two basic types: blast mines with a pressure plate that explode when someone steps on them, and fragmentation mines which explode when a tripwire is touched. Unexploded ordnance is often left on the battlefield and has similar injuring effects as fragmentation mines (the lethal remnants of war consisting of cluster bomblets, bombs and shells that did not explode when fired). These weapons continue to kill and injure large numbers of civilians long after the end of hostilities and have widespread humanitarian and economic repercussions.

ICRC

Figure 3.3.1

Blast mine.

ICRC

Figure 3.3.2

Fragmentation mine.

ICRC

Figure 3.3.3

PMF "Butterfly mine".

Patterns of injury

Anti-personnel mines cause three distinct patterns of injury according to the blast effect or the production of fragments.

Pattern 1

A person steps on the pressure plate of a blast mine. The explosion and local primary blast effect cause a traumatic amputation or severe injury of the contact foot and leg. There may be wounds to the other leg, genitals, abdomen or pelvis, and the contralateral arm. The severity of the wound depends on the amount of explosive in the mine compared to the body mass (Figure 3.4).

Pattern 2

A person touches the tripwire attached to a fragmentation mine, which provokes its detonation. Fragmentation mines cause the same injuries as other fragmentation devices, such as bombs or grenades, and the severity of the injury depends on the distance of the victim from the explosion.

Pattern 3

A person handles a mine: setting a mine; trying to clear a mine; a child playing with a mine. The explosion causes severe injury to the hand and arm, and frequent injuries to the face and eyes or chest.

For further information on anti-personnel landmine injuries, see Volume 2.

R. Coupland / ICRC

Figure 3.4

"Umbrella" effect of blast mine: superficial tissues suffer less damage than deep-tissue planes. The wounds are all severely contaminated by mud, grass, and pieces of the shoe driven deep into the tissues at the moment of explosion.

3 In some countries where a "traditional" form of warfare is still practised, a large, heavy knife – the machete or *panga* – is often used as a weapon. The overhead motion strikes the victim on the head, neck or shoulder.

3.1.4 Explosive blast injuries

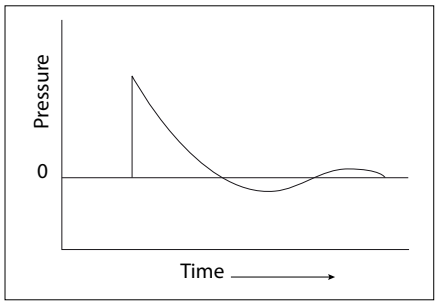

Figure 3.5

Blast wave overpressure followed by negative pressure trough.

The detonation of high-energy explosives creates a travelling pressure blast wave in the air, (or in water for underwater explosions). The wave causes rapid and large changes in the outside atmospheric pressure: the positive pressure shock wave is followed by a negative pressure phase (Figure 3.5). Immediately behind the pressure wave is a mass movement of air: the blast wind.

As the blast wave passes any unprotected person it affects all parts of the body, especially those normally containing air. A victim of blast may not have any external injury. A single, large explosion may injure many people at the same time. Explosions in a closed space (building, bus, etc.) are more fatal than those out in the open.

Categorization of blast injuries

Blast injuries are commonly categorized into four types.

Primary

These are direct pressure effects. Rupture of the tympanic membrane is the most common injury. Rupture of the lung alveoli and their capillaries ("blast lung") is the most lethal injury amongst survivors.

Secondary

These injuries include missile wounds. Fragments may arise from the bomb casing or contents (primary missiles). Home-made bombs (improvised explosive devices, known as IED) may be packed with nuts and bolts, screws and ball bearings. In addition, the blast wind may mobilize various objects (secondary missiles) that then cause penetrating wounds.

Tertiary

These effects are directly due to the blast wind. It can cause total body disintegration in the immediate vicinity of the explosion; traumatic amputations and evisceration further away. The wind can make buildings collapse or throw people against objects. Trauma may be blunt, crush or penetrating.

Quaternary

Miscellaneous types of harm due to burns, asphyxia from carbon monoxide or toxic gases, or the inhalation of dust, smoke or contaminants.

The various injuries caused by major blasts cover a whole spectrum of trauma and many patients suffer several injuries from a variety of effects: i.e. multiple injuries from one single weapon system.

Other situations of blast injuries

The blast wave travels more rapidly and much further in water than in air. Blast injuries in water occur at greater distances and can be more severe. Moreover, underwater explosions tend to cause pure primary blast injury. A "fuel air explosive" – the liquid explosive material is dispersed in the air like an aerosol and then ignited – also tends to bring about pure primary blast injury and quaternary effects due to the consumption of all oxygen in the nearby air.

The blast effect of a marine mine exploding below an icy surface, or the "deck slap" of a ship hit by a torpedo, produces a shock wave that can severely fracture the bones of anyone on deck or inside the ship. Similarly, some anti-tank mines send a blast wave through the floor of the vehicle causing closed fractures of the foot and leg. The foot appears like a "bag of bones" inside intact skin, which was described in World War I as "*pied de mine*": a mine foot. Anti-personnel blast mines have a local blast effect, vaporizing the tissues of the contact foot, as described previously.

For further information on blast injuries, see Volume 2.

3.1.5 Burns

A large explosion may cause flash burns or the carbonization of tissues. Bombing may start secondary fires in buildings and an anti-tank mine may ignite the petrol tank of a vehicle. Burns are common amongst the crews of tanks, ships and aircraft hit by missiles. Some types of anti-personnel blast mines provoke burning as well as traumatic amputation of the limb.

Certain weapons cause specific burns: napalm and phosphorus bombs, magnesium flares.

3.1.6 Non-conventional weapons

International humanitarian law prohibits the use of chemical and biological weapons. Nonetheless, many countries have stocks of them. If not used during actual combat, they may nevertheless be released when storage warehouses are bombed.

Radioactive agents have come into greater use recently, such as depleted uranium in anti-tank shells. Residual radioactive shells and debris constitute a long-term hazard. A bomb surrounded by radioactive material – the so-called "dirty bomb" – is not a nuclear bomb. The explosion is caused by conventional means but radioactive material is spread out over a wide area. The bombing of nuclear medicine and other laboratory facilities, or nuclear power plants, may also release radioactive material into the atmosphere.

The reader should refer to standard military texts and those of the Organization for the Prohibition of Chemical Weapons (OPCW), and relevant documents of the World Health Organization for further details. Nuclear weapons shall not be discussed here.

3.2 Ballistics

3.2.1 Introduction

Blast and penetrating missile wounds occur in times of armed conflict and whenever the weapons of war are used in peacetime. Weapons cause specific but variable patterns of injury. While standard surgical techniques will suffice to treat simple wounds, the management of war wounds produced by high-energy weapons is based on an understanding of the mechanisms by which projectiles cause injury: wound ballistics. Only by understanding certain physical phenomena can the surgeon appreciate the different varieties of wounds seen in armed conflict and the difference between these wounds and the trauma that is seen in everyday civilian practice.

Although the study of ballistics may be "interesting" in its own right, the clinician does not always know what weapon inflicted the injury; and *never* knows the energy available at point of impact. One can only estimate the transfer of energy in the tissues from the extent of tissue damage.

The study of ballistics gives us an understanding of the basic mechanisms at work during wounding. The importance of this knowledge lies in the fact that projectile injuries should be neither under- nor over-treated. The clinical assessment of the actual wound is the most important factor determining management, and an understanding of ballistics allows the surgeon to better understand the pathology and assess the injuries that he sees, rather than explain every wound and determine specific treatment. "Treat the wound, not the weapon"[4] is not a vain concept.

Basic definitions
Ballistics is the part of the science of mechanics that studies the motion and behaviour of a projectile, and its effects on a target.

Internal ballistics
Internal ballistics deals with the processes that occur inside a gun barrel when a shot is fired: gas pressures on combustion of the propellant powder, the energy and heat released, and the course of the projectile in the barrel, are just some of its concerns.

4 Lindsey D. The idolatry of velocity, or lies, damn lies, and ballistics. *J Trauma*. 1980; **20**: 1068 –1069.

External ballistics

External ballistics describes the projectile trajectory once it has left the barrel. The influences affecting the flight include gravity, air resistance and crosswind deflection, stability of the projectile (spin and yaw), as well as any contact prior to the projectile reaching the target, known as ricochet.

Terminal ballistics

Terminal ballistics describes what happens when the projectile hits the target, as well as any counter-effects produced by the target on the projectile. If the target is biological tissues, terminal ballistics is called *wound ballistics* and describes the effects on the tissues.

3.2.2 Internal ballistics

Bullets: fundamental concepts

Figure 3.6 shows the main components of a bullet cartridge. The primer is struck by a mechanism in the gun to produce a small detonation and flame that sets off the powder in the case. This causes a very rapid burn with the production of a large volume of rapidly expanding gas, which pushes the bullet out of the barrel of the gun. The velocity with which the bullet leaves the barrel is called the muzzle velocity.

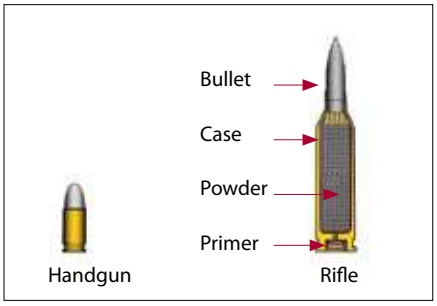

B. Kneubuehl

Figure 3.6

Main parts of a bullet cartridge.

Bullet construction

Bullets are classified according to a number of parameters; one is their manufactured composition (Figure 3.7). They vary according to their calibre and mass.

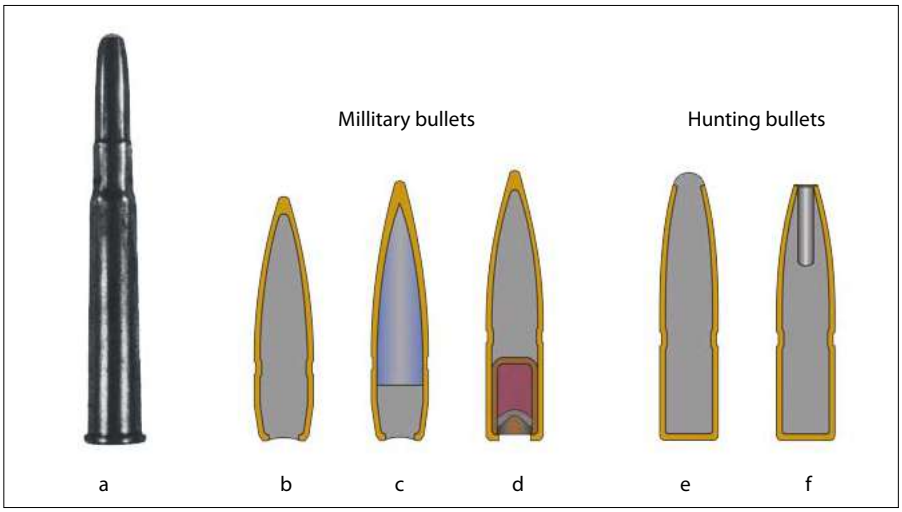

B. Kneubuehl

Figure 3.7

Bullets according to construction.

a. 303 British Mark II produced in Dum Dum, India, 1896 (see footnote number 6).

b. Full metal jacket (FMJ): military-issue bullets have a soft core (lead) with a full metal jacket of copper. This standard has been established by various conventions and treaties of IHL, especially the Third Hague Convention of 1899.

c. Hard core bullet: the lead has been replaced by steel or tungsten for better penetrating power.

d. Tracer bullet: contains a pyrotechnical substance at the tail-end that burns up in flight and illuminates the trajectory. Used to identify and pinpoint the target.

e. Semi-jacketed bullet (SJ): part of the point is not jacketed, thus exposing the lead inside. This is supposed to be used only for hunting; its use by combatants during armed conflict is illegal.

f. Semi-jacketed hollow-point (SJ-HP): also a hunting bullet with a hollow nose tip; also illegal during armed conflict.

Apart from the hunting bullets mentioned in Figure 3.7, special rifles and their ammunition exist for hunting: shotguns whose shells give off multiple lead or steel shots (pellets) available in diameters from 2 to 9 mm (Figure 3.8).

Hunting bullets are not supposed to be used by combatants in armed conflict. These bullets are constructed in such a way as to kill rather than wound. The idea is to be more "humane" when hunting wild animals and to have a "rapid and clean kill". Nonetheless, the surgeon will be faced with wounds caused by such weapons either because of accidents, criminal activity, or because combatants have illegally used them during combat.

The international legal limits placed on bullets used by the military do not apply to domestic use during peacetime, and the civilian surgeon may at times face more devastating gunshot wounds than the military surgeon on the battlefield.

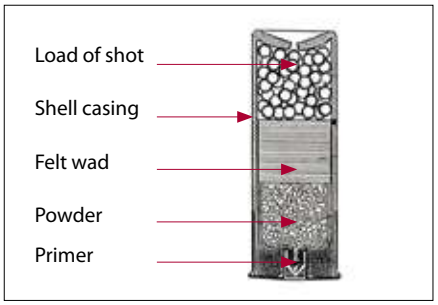

Figure 3.8
Shotgun shell and load of shot.

Muzzle velocities

Firearms are classically divided into two categories: high-velocity (rifles) and low-velocity (handguns).The latter fire relatively heavy bullets at low muzzle velocities of 150 – 200 m/s. A typical military assault rifle fires a smaller bullet at 700 – 950 m/s. This, however, says nothing about the actual velocity when the bullet reaches the target.

9 mm Luger pistol	350 m/s
.38 Special handgun	260 m/s
.44 Magnum handgun	440 m/s
7.62 x 51 mm NATO rifle	830 m/s
5.56 x 45 mm M16 rifle	960 m/s
7.62 x 39 mm AK-47 Kalashnikov rifle	720 m/s
5.45 x 39 mm AK-74 Kalashnikov rifle	900 m/s
12 gauge shotgun	420 m/s

Table 3.1 Some examples of muzzle velocity from different firearms.

The gun barrel

A bullet is a long projectile that only flies with stability if a high spin around its longitudinal axis is attained, thus giving it a gyroscopic effect. To achieve this spin, gun barrels are constructed with interior spiral grooves (twists). This is a "rifled" barrel, used in all handguns and high-velocity rifles (Figure 3.9).

The barrels of shotguns are smooth on the inside; they are not "rifled". This limits their accuracy and distance of fire (Figure 3.10).

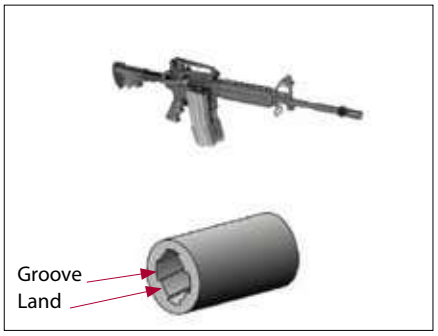

Figure 3.9
Cross section of a rifled gun barrel:
5.56 x 45 mm M-16 A4 military assault rifle.
The grooves and lands (the elevated sections between the grooves) are shown.

Design of firing

Another classification of firearms is according to their design of firing. A weapon operated with a single hand is a "handgun". If the barrel and cartridge chamber are in one piece, this is a "pistol". If several chambers rotate behind the barrel, it is called a "revolver". If both hands are required to operate the weapon, it is referred to as a "rifle" or "shotgun".

Firing capacity defines how individual shots are fired. For a single-shot weapon, every shot is loaded individually. A repeating weapon includes a magazine that holds a series of cartridges which load manually one after the other. If the loading movement is repeated automatically after every shot, but every shot is fired individually, this is a semi-automatic weapon. If the automatic loading of cartridges allows several shots to be fired with one pull of the trigger, it is known as an automatic weapon.

In modern military usage, most weapons are automatic rifles and sub-machine guns or semi-automatic pistols.

Figure 3.10
Typical hunting shotgun with smooth barrel.

3.2.3 External ballistics

Flight in air

Once shot off, several variables affecting the projectile in flight tend to influence its stability, accuracy of targeting, or velocity. The most important variables include the following.

- Spin around its own longitudinal axis, giving it gyroscopic stability.
- Distance travelled and the effect of gravity, including the shooting angle – is the bullet shot upwards or is the trajectory flat?
- Air drag: friction is responsible for 10 % of total area drag, while 90 % is due to pressure effects; faster bullets are retarded proportionately more – rifle bullets are aerodynamically shaped to decrease air drag; this is not the case with shotgun cartridges or metal fragments.
- Crosswind deflection.
- Raindrops hitting the bullet.
- Bullet hitting an obstacle before hitting the target.

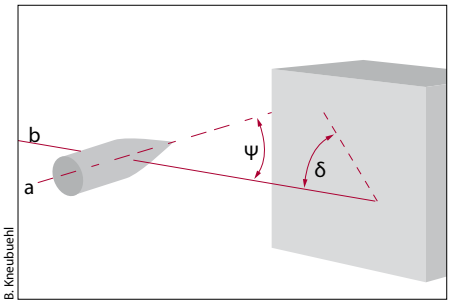

B. Kneubuehl

Figure 3.11

Angle of incidence and angle of impact:

a. longitudinal axis of bullet
b. direction of flight

Ψ. angle of incidence
δ. angle of impact

Yaw

A rifle bullet in flight does not travel in a simple straightforward motion. Because it is a spinning gyroscope, the bullet "wobbles" and undergoes some very complicated movements (nutation, precession), the most important being yaw: the nose of the bullet moves up and down, away from the line of flight, and describes an angle with the target surface at impact (Figure 3.11). This influences the wounding potential of the bullet, as an important degree of yaw will destabilize the motion of the bullet inside the target.

Ricochet

A bullet may hit an obstacle during its flight. This may be a branch of a tree, a belt buckle, a concrete wall or the ground, or a soldier's helmet or flak jacket. This causes a ricochet; the bullet is given a small "push" that destabilizes it (Figure 3.12). There is an increase in the deviation of the nose of the bullet (yaw); the bullet may even tumble, rotating around itself. Such a large yaw angle at point of impact, as mentioned above, will have important consequences for the wounding potential of the bullet in the target.

In addition, if the "push" is great enough, the bullet can be deformed or even broken up into fragments before hitting the target.

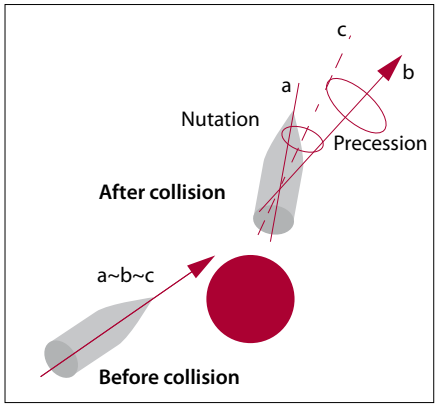

B. Kneubuehl

Figure 3.12

Ricochet: effect on a bullet after collision with an obstacle.

3.2.4 Summary

There are, therefore, quite a number of variables involved in determining the characteristics of a projectile *before* it hits the target and these will influence its behaviour in the target and the efficiency in creating wounds. They include:

- velocity at moment of impact and any residual velocity if there is an exit;
- mass, form and shape of the projectile, bullet construction;
- type of weapon, handgun or rifle;
- stability of the projectile in flight;
- any yaw at moment of impact.

3.3 Terminal ballistics

3.3.1 Role of kinetic energy

Any moving object has kinetic energy, whether it is a hand-wielded knife or club, a bullet fired from a gun, or a fragment given off by the explosion of a shell. This energy is described by the well-known formula:

$$E_K = \frac{1}{2} m v^2$$

This defines the total kinetic energy that the object possesses, but does not define the kinetic energy expended when the projectile passes into and through a target. For a bullet or fragment that does not change its mass, this is calculated from the difference in velocity between entry and exit.

$$E_{K\,EXP} = \frac{m\,(v_1^2 - v_2^2)}{2}$$

If there is no exit, then $v_2 = 0$, and all the kinetic energy has been transferred. If the bullet breaks up into fragments, then the mass (m) also changes, and $E_{K\,EXP}$ is modified.

Total kinetic energy is the potential for causing damage; transferred kinetic energy is the capacity to cause damage. Actual tissue damage, however, depends on the efficiency of this energy transfer, which is determined by many other variables, as will be shown.

Weapons according to E_K
Yet another classification of weapons is based on the amount of kinetic energy available for transfer.

- Low energy: knife or hand-energized missiles.
- Medium energy: handguns.
- High energy: military or hunting rifles, with a muzzle velocity above 600 m/s.

For metal fragments given off by an explosion the initial velocity is very high; this decreases rapidly with distance. The wounding potential depends on the mass of the fragment and the distance of the victim from the weapon.

3.3.2 Laboratory experiments

Many researchers have performed different experiments to describe the effect of projectiles on tissues. Targets have included human cadavers, various animals (pigs, dogs, goats) and tissue simulants.

Tissue simulants are specially-prepared blocks of gelatine or glycerine soap, which have a density and/or viscosity close to that of muscle tissue. Soap is plastic and any deformation remains unchanged, representing the maximum effects. Gelatine is elastic and deformations disappear almost entirely; they are studied with high-speed cameras. If the stress on gelatine exceeds the limit of its elasticity then it will crack and tear, showing fracture lines.

Compared to human cadavers and animals, tissue simulants have the advantage of allowing repetitions of the experiment, changing only one variable at a time. These laboratory experiments are all approximations, however, of what happens in a live human being.

Surgeons working for the ICRC have collaborated over many years with the ballistics laboratory of the Swiss Federal Department of Defence.[5] This laboratory uses gelatine and glycerine soap for its ballistics experiments. Its results have been confirmed by comparison with clinical cases treated by ICRC surgeons in various war zones throughout the world.

5 See footnote 1.

Following these laboratory studies, five categories describe the behaviour of projectiles: non-deforming and deforming bullets, be they from rifles or handguns, and fragments.

3.3.3 Non-deforming rifle bullet: full metal jacket (FMJ) military bullet

When a standard military bullet hits a soft object while in stable flight, it produces a "shooting channel" that presents three distinctive phases: narrow channel; primary temporary cavity; and terminal narrow channel (Figure 3.13).

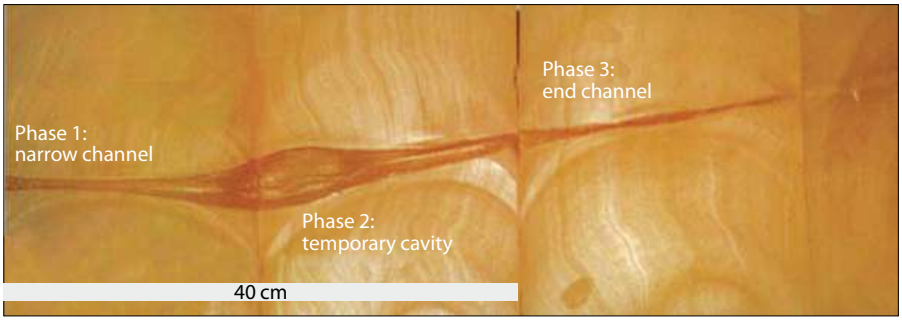

B. Kneubuehl

Figure 3.13
Non-deforming rifle bullet (FMJ) in soap.

Phase 1
Straight narrow channel with a diameter about 1.5 times the calibre of the bullet. The greater the velocity, the wider the channel. Different types of bullets have a different length of narrow channel, typically 15 – 25 cm.

Phase 2
The channel opens into the "primary temporary cavity". The reported diameter of this cavity is anywhere between 10 and 15 times the calibre of the bullet.

Figure 3.14 shows the motion of the bullet in gelatine. It starts to yaw, and tumbles, turning a full 270°, and then advances with its tail end forward. The whole lateral surface of the bullet comes into contact with the medium, which greatly decelerates the bullet and subjects it to an important stress.

B. Kneubuehl

Figure 3.14
The bullet tumbles in the gelatine or soap: it rotates 270° around a transverse axis that is perpendicular to the long axis. (Graphic demonstration of bullet superimposed on soap blocks. The proportions between bullet and trajectory are exaggerated for the sake of clarity.)

The slowing of the bullet involves a release of kinetic energy, which, in turn, causes the rapid and massive displacement of gelatine radially outwards, producing a cavity *behind* the projectile; there is a slight delay in the production of the cavity because of the inertia of the gelatine mass. There is an almost complete vacuum in this cavity that quickly sucks in air from the entry hole, and from the exit hole if there is one. The cavity collapses after a few milliseconds, only to reform again, with a smaller volume. The cavitation continues until all the transferred energy is used up: the cavity pulsates! In water or gelatine, there are up to 7 – 8 pulsations, in biological tissues usually 3 – 4.

The diameter of this cavity depends upon the elastic properties of the medium, as well as the amount of kinetic energy transferred. Fissures radiating from the shooting channel indicate that the shearing effect of the cavity has exceeded the elasticity of the gelatine.

Phase 3

The tumbling slows down and the bullet continues in a sideways position at a considerably reduced speed. In some cases, a narrow straight channel is observed; in others, the tumbling seems to continue but backwards, the bullet again assuming a lateral position, and a second cavity occurs. This does not reach the size of the primary temporary cavity. The bullet then creeps forward and finally stops, always with the tail end facing forwards.

In an elastic medium such as glycerine, what remains in the shooting channel at the end of the process, and all temporary effects, is called the "permanent channel".

References to these basic definitions of phases of the shooting channel will be made throughout this chapter.

Please note:

These three phases are found with all FMJ rifle bullets, yet each bullet has a specific shooting channel. The narrow channel of the 7.62 mm AK-47 is long (15 – 20 cm), while the 5.45 mm AK-74 produces a narrow channel of less than 5 cm before cavitation begins.

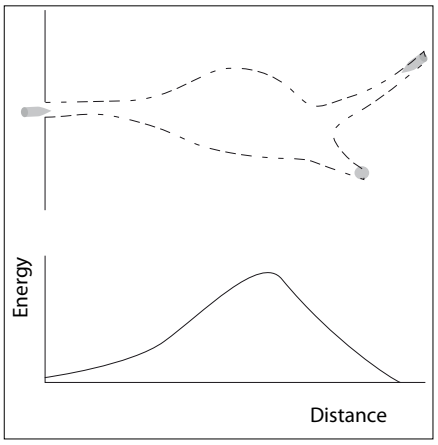

R. Coupland / ICRC

Figure 3.15

The sketch shows the position of the bullet and the extent of the shooting channel at different phases. The graph represents the transfer of kinetic energy along the bullet track: bullet fragmentation occurs at the peak of energy transfer.

Tumbling – yaw – in the target medium

Given a long enough shooting channel, all FMJ rifle bullets tumble. How early the tumbling motion begins determines the length of the narrow channel, and the onset of cavitation. This depends on the stability of the bullet – yaw – at the point of impact. The less stable the bullet in flight, the greater the yaw, which rapidly brings a larger bullet surface into contact with the medium, leading to early tumbling and a short narrow channel. How early FMJ bullets tumble is also a function of their manufacture: mass, centre of gravity and shooting distance, etc.

Fragmentation of FMJ bullets

It is during phase 2 – the temporary cavity – that certain bullets deform or even break apart because of the enormous stress exerted on them. This occurs when the bullet-medium interface is at its maximum, the cavity is widest, and the transfer of kinetic energy is highest (Figures 3.15 and 3.16). If the projectile acts on the target medium, this is a good example of the medium acting on the projectile. This fragmentation occurs only at short shooting distances, up to 30 – 100 m, depending on the bullet's construction and stability.

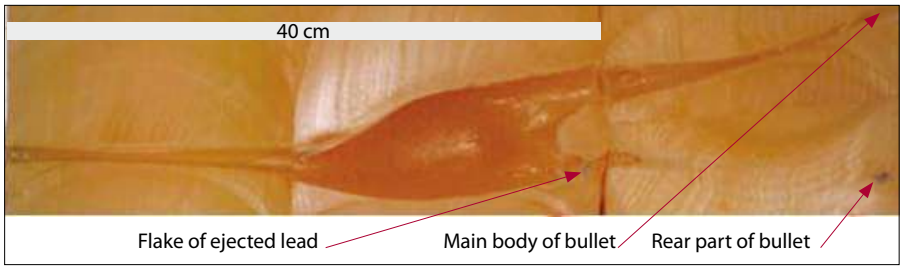

B. Kneubuehl

Figure 3.16

Fragmentation of FMJ bullet during phase 2, i.e. temporary cavitation.

The bullet flattens on its sides, bends in the middle and, finally, the jacket splits and the lead inside spills out (Figure 3.17). This gives the "shower of lead" picture often seen on X-rays (Figures 3.35, 4.5 and 10.5). If the bullet breaks, the smaller fragment typically deviates downwards. The fragmentation depends on the construction of the bullet and the velocity; with impact velocities below 600 m/s, no FMJ deforms or breaks apart.

If the bullet does fragment then the primary temporary cavity is larger than is the case with an intact bullet. This represents a much larger transfer of kinetic energy and has important clinical repercussions.

ICRC

Figure 3.17

Jacket split apart, exposing the lead within.

3.3.4 Deforming and fragmenting rifle bullets: dum-dum[6]

Some bullets (e.g. hunting bullets) are constructed to always deform or break up into fragments: hollow-nosed, semi-jacketed (SJ), soft-pointed, etc. (Figure 3.7 e and f). These are usually grouped under the term "dum-dum" bullets and are illegal for military use according to international law.

Deforming bullets are made to change their shape readily ("mushrooming"), thus increasing their cross section area, but without losing mass. The spent bullet weighs as much as the original. They are used mainly in handguns with a muzzle velocity below 450 m/s, and are available for domestic use (special police forces *and* criminals). Fragmenting bullets, on the other hand, break up and lose mass, and create a "wall" of particles, thus increasing the total effective area of cross section. They are used for hunting.

Figure 3.18 demonstrates a semi-jacketed rifle bullet mushrooming immediately on impact with a soft medium. The increased cross section causes a greater bullet-medium interaction; the bullet slows down quickly, releasing kinetic energy very early. The narrow channel almost completely disappears and the temporary cavity occurs right after impact. At first, the cavity is almost cylindrical; it then decreases conically.

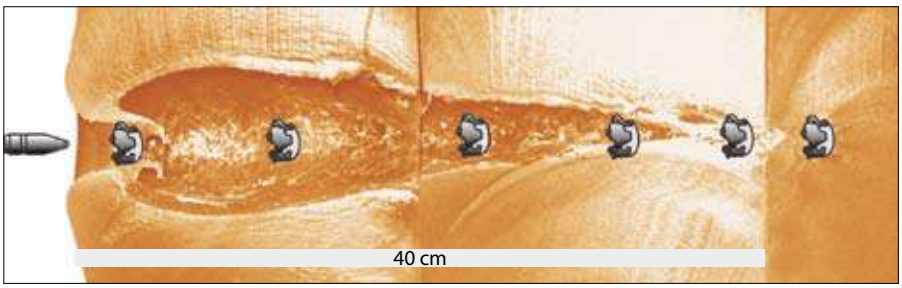

B. Kneubuehl

Figure 3.18

Deforming semi-jacketed rifle bullet (SJ) in soap. The bullet "mushrooms" just after impact, then continues in a linear trajectory. (Graphic demonstration of bullet superimposed on soap blocks.)

The principal difference between a full metal jacket and semi-jacketed bullet is at what depth of penetration the maximum energy transfer occurs in the shooting channel. The volume of the cavities is the same in both examples in Figure 3.19 indicating equal transfer of kinetic energy.

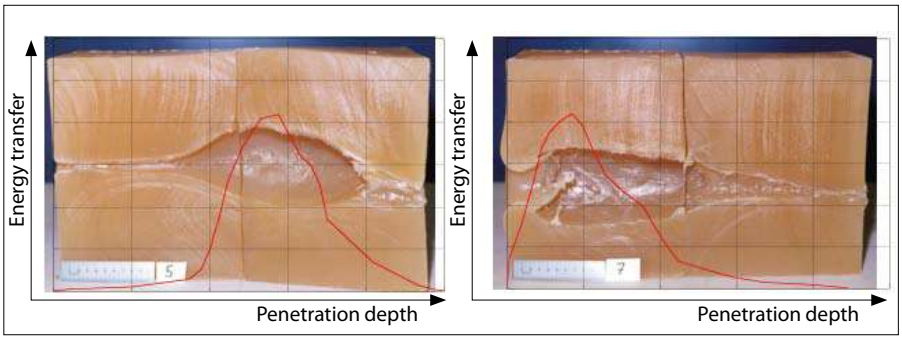

B. Kneubuehl

Figure 3.19

Soap blocks showing comparison between ballistic profiles of FMJ and SJ bullets – the transfer of kinetic energy occurs far sooner with the SJ bullet.

6 The Indian Ordnance Department of the British Army developed a bullet for its colonial troops in 1897 at an ammunition factory in the town of Dum Dum, north-east of Calcutta (Kolkata) because the previous one was judged to be ineffectual and caused insignificant injuries. The bullet had a round nose with a full metal jacket of copper-nickel covering a lead core, with the exception of 1 mm at its tip where the lead was bare. The bullet was used against the Afridi Afghans in 1897 – 98, and again against the Mahdi Army at Omdurman, Sudan, in 1898 with devastating effect. The bullet was judged to be "inhumane" according to the Declaration of St Petersburg of 1868 and was banned by the Hague Convention of 1899. Projectiles causing unnecessary injuries ("*maux superflus*") were not allowed. As an example, the Conventions mentioned projectiles whose metal jacket did not fully cover the lead core. Since then, all projectiles possessing the same qualities (deformation by expanding or flattening easily, fragmentation) have received the collective name of dum-dum.

The same effect can be demonstrated when using a synthetic bone simulant encased in gelatine (Figure 3.20).

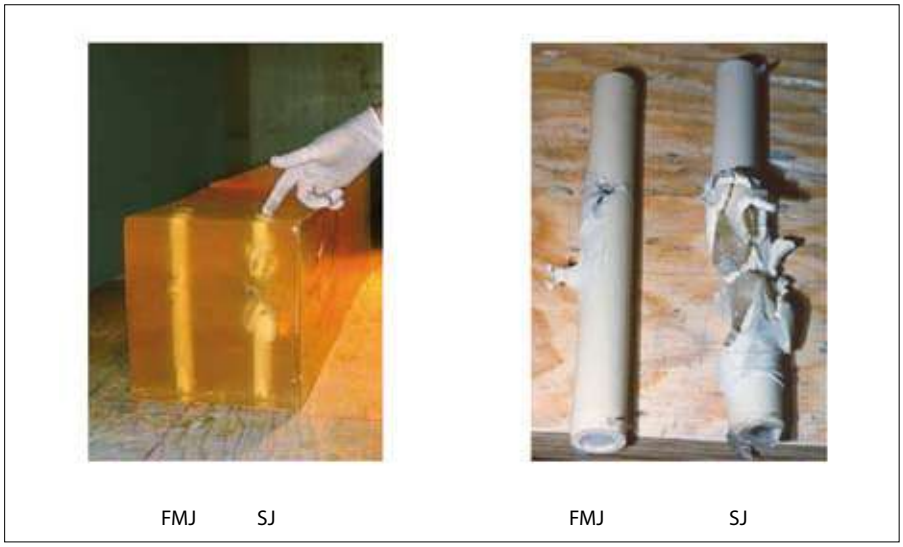

B. Kneubuehl

Figure 3.20

Comparison of FMJ bullet versus SJ bullet: synthetic bone at a shallow depth of gelatine. The FMJ bullet breaks the bone simulant during the narrow channel phase; the shooting channel is practically the same with or without the bone simulant. At the same depth, the SJ bullet shatters the bone simulant completely.

Ricochet effect

When a FMJ bullet strikes an obstacle before hitting the target it is destabilized. After impact, there is almost no narrow channel and the shooting channel resembles what occurs with a deforming or dum-dum bullet (Figure 3.21). This phenomenon has important clinical consequences.

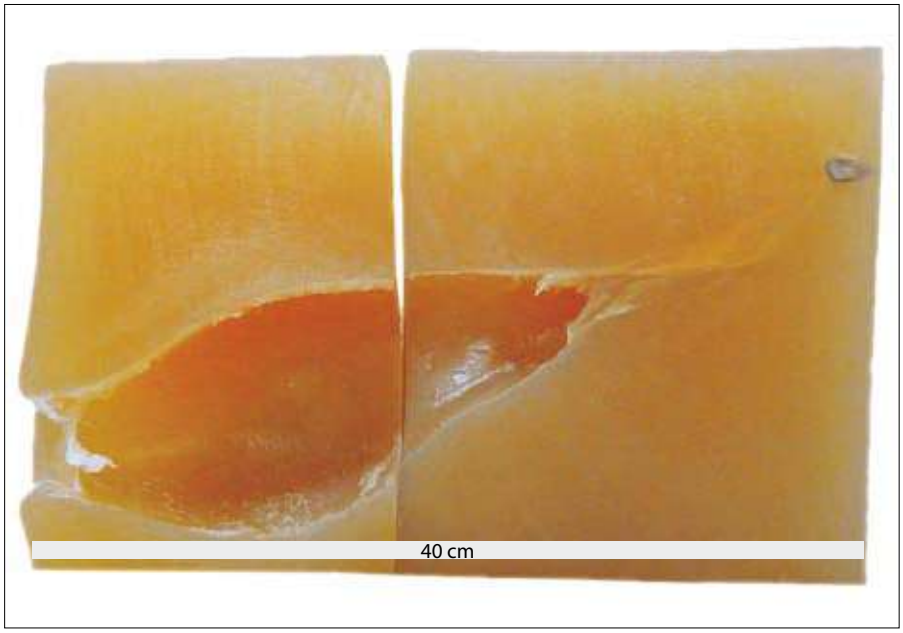

B. Kneubuehl

Figure 3.21

FMJ rifle bullet after ricochet effect in soap. The large impact angle after ricochet destabilizes the bullet, which tumbles easily and early in the shooting channel. Note that the cavitation occurs almost immediately on impact, similar to an SJ bullet.

Please note:

Jacket construction is not the only decisive factor in bullet behaviour be it full or semi-jacketed. It is quite possible to manufacture a bullet that will fragment at high velocities, deform at medium velocities, and retain a stable shape at low velocities.

3.3.5 Handgun bullets

Handgun bullets may be much heavier than rifle bullets.

Non-deforming bullet

The shooting channel of a non-deforming bullet shows little yaw and no tumbling; the bullet remains with the tip pointing forward with deep penetration (Figure 3.22). Temporary cavitation is long and narrow.

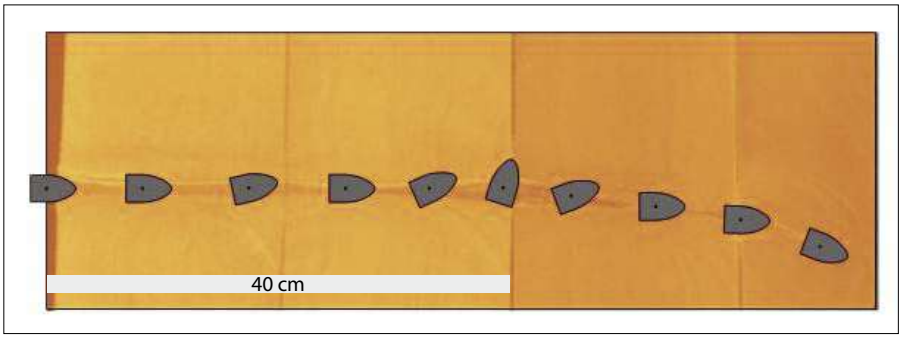

B. Kneubuehl

Figure 3.22

Standard military pistol bullet, full metal jacket in soap, no tumbling of the bullet. (Graphic demonstration of bullet superimposed on soap blocks.)

Deforming bullet

A soft point, deforming pistol bullet – as used by special police forces – mushrooms at entry (Figure 3.23). The large cross section area causes an abrupt reduction in velocity and transfer of kinetic energy with an immediate, and large, temporary cavity.

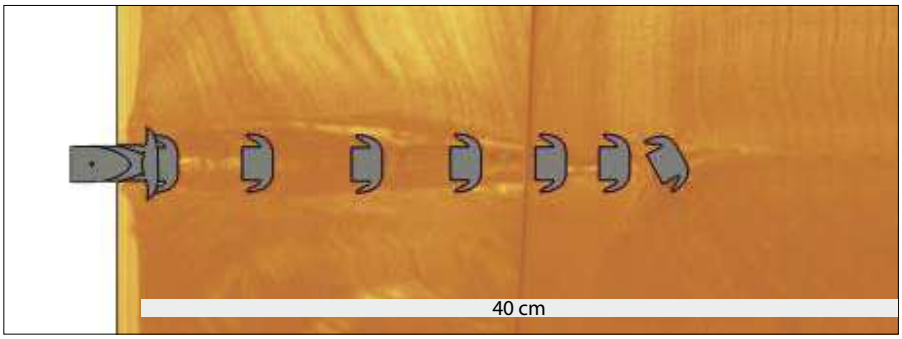

B. Kneubuehl

Figure 3.23

Deforming handgun bullet in soap: mushrooming effect. (Graphic demonstration of bullet superimposed on soap blocks.)

3.3.6 Fragment projectiles

Fragments given off from the explosion of a bomb, rocket or grenade are not aerodynamic; they have an irregular shape. Their velocity decreases rapidly with distance owing to air drag. They have an unstable flight path with an irregular rotation around an indefinite axis. On impact, the biggest cross section comes into contact with the target surface, immediately transferring a maximum of kinetic energy. No yaw or tumbling occurs in the target.

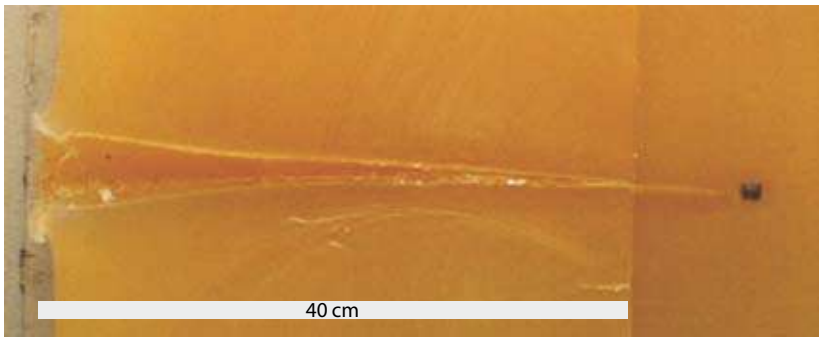

B. Kneubuehl

Figure 3.24

Fragment profile in soap: the widest point is at entry, the cavity is cone shaped.

The shooting channel always has the largest cavity diameter at the entry, and is larger than the diameter of the fragment. The cavity then decreases constantly in a cone-like fashion (Figure 3.24).

The depth of penetration of the fragment will depend on its kinetic energy, but with a particular relationship between velocity and mass. Figure 3.25 shows two fragments with the same kinetic energy; the volumes of the cones are equal.

Thus, a light but rapid fragment tends to transfer most of its energy shortly after impact on the target; a heavy and slow fragment penetrates further and dissipates its energy along the longer shooting channel.

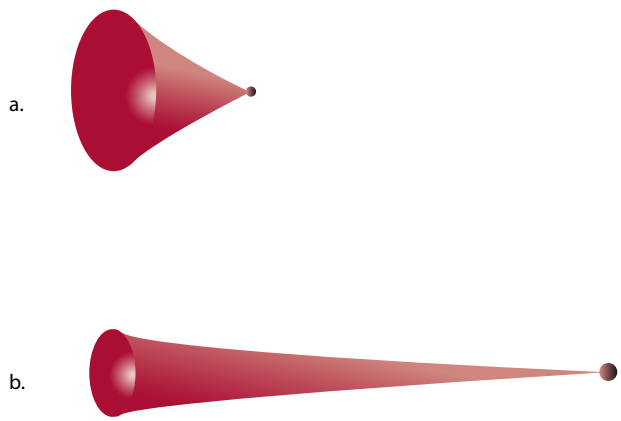

B. Kneubuehl

Figure 3.25

Two fragments with the same kinetic energy – note the difference in energy deposition along the track, demonstrated by the difference in the cavities:

a. lightweight and fast fragment;
b. heavy and slow fragment.

3.4 Wound ballistics

Moving projectiles cause wounds by a transfer of kinetic energy to the body that destroys, disrupts, and deforms tissue. One studies wound ballistics in order to understand the mechanisms that produce this injury.

Potential for wounding is one thing; the *actual* wound facing the surgeon is another.

3.4.1 Laboratory studies *versus* human body

The experiments using tissue simulants as described above help one to understand the processes at work. Laboratory models are, nonetheless, only an approximation and show only the physical part of the process. The tissue structure of the human body is far too complex for a laboratory model to duplicate injuries. Tissue simulants have the disadvantage of being homogeneous media; living tissue is not, as described later. The laboratory results must be compared to clinical cases for confirmation and this has been the ICRC procedure in conjunction with the Armasuisse laboratory in Thun, Switzerland.

In the real world of clinical practice there are so many variables involved that one cannot be *predictive*; the surgeon cannot say with certainty that such and such a bullet will always result in a specific wound. One can however be *descriptive* and, after a good clinical examination, the surgeon can better understand the pathology that he sees, how it was created, and what sort of surgical management is required. The type of wound, its anatomic location, and the amount of actual tissue disruption are the determinant clinical factors.

3.4.2 Projectile–tissue interaction

When a bullet strikes a human body there is a projectile–tissue interaction that results in tissue damage, with a reciprocal influence of the tissues on the bullet. This interaction depends on a number of factors, which all result in the transfer of kinetic energy from the projectile to the tissues.

This transfer of kinetic energy either compresses, cuts or shears tissues, causing crush, laceration or stretch. The local energy transfer, at every point along the bullet track, is more important than the total quantity of energy transferred in producing tissue damage.

Tissue disruption is due to compression, cutting and shear: crush, laceration and stretch.

A bullet striking a human body demonstrates the *same three phases* as in the laboratory simulants – if the bullet track is long enough. In the laboratory experiments with glycerine, the "permanent channel" was defined as what remains of the shooting channel at the end of the process and all temporary effects. In biological tissues, the "permanent wound cavity" of the shooting channel is the definitive tissue damage after all temporary effects have been taken into account. This is the wound channel that the surgeon sees and is the final result of the crush and stretch of the tissues.[7]

Crush and laceration injury

A projectile causes physical compression and forcing apart of tissues along the shooting channel – crushing and laceration. This is the immediate physical effect of a penetrating foreign body; the bullet damages tissues with which it comes into direct contact, cutting through tissues as would a knife. This tissue damage is permanent and found in the final wound. This is the principal effect caused by low- and medium-energy weapons, such as handgun bullets.

7 Much confusion has been created in surgical literature by authors calling the immediate crush damage of the phase 1 narrow shooting channel the "permanent cavity", as distinct from the "temporary cavity" where stretch damage occurs. In this manual, in accordance with the Armasuisse laboratory, the permanent wound cavity is the channel that remains at the end of the wounding process and constitutes the sum of crush and stretch injury.

At higher energies, when a bullet tumbles or deforms, a larger area of tissue is exposed to its effective cross section and suffers crush. The track left behind by crush is not equal along its length, but increases with bullet yaw in the tissues.

Stretch injury

Tissues have an elastic tensile strength that resists stretching. When a certain limit is reached, capillaries are ruptured and contusion of the tissues occurs. Beyond a critical limit, the tissues themselves are torn, just as laboratory gelatine shows fracture lines. The tissue damage from stretch may be permanent or only temporary.

Tissue stretch occurs during cavitation, which takes place in all projectile wounds, whatever the energy, type or motion of the projectile, and at all points along the entire shooting channel. There is even a minor cavitation effect during the phase 1 narrow channel.

The volume of the cavity is determined by the amount of energy dissipated and the tissue elasticity-resistance. This cavitation stretch performs work on tissues that have already been injured by crush, thus adding to the local immediate damage. In low- or medium-energy projectiles wounds, it is minimal.

When a bullet tumbles (or deforms or fragments), the release of kinetic energy is much greater and is superimposed on bigger tissue crush; the large phase 2 temporary cavity is the result: a momentary massive displacement of tissue in all directions away from the track of the bullet.

As in gelatine, the cavity pulsates: an elastic acceleration followed by a deceleration of the surrounding tissues: a shearing action. The near vacuum in the cavity sucks in air, contaminants and foreign bodies (textile strands from the clothing, dust, etc.), and bacteria, through the entry and any exit wound.

In high-energy wounds the *volume* of the temporary cavity can be up to 25 times that of the permanent cavity left behind, and anywhere between 10 and 15 times the *diameter* of the bullet. Although this volume is proportional to the kinetic energy dissipated, and will affect the amount of tissue damage, other factors related to the tissues can be even more important in determining the actual extent of injury.

3.4.3 Tissue factors

Resistance to crush, laceration and stretch differs widely among different types of tissues and according to anatomic structures. Tissue elasticity and heterogeneity are important factors in determining much of the bullet–tissue interaction.

Tissue elasticity

Elastic tissues tolerate stretch well, but may still suffer severe crush. The lung and skin have excellent tolerance, and leave relatively little residual damage after stretch. Skeletal muscle and the empty bowel wall have good tolerance. The brain, liver, spleen, and kidney are non-elastic, and shatter when stretched. Fluid-filled organs (heart, full bladder, full stomach and intestines) react badly owing to the incompressibility of the fluid contents; they may even "explode".

Nerves and tendons are mobile and blood vessels elastic; they are usually pushed out of the way by the cavitation.

Cortical bone is dense and rigid and resists stretch. However, if the cavitation accelerates a large enough muscle mass with sufficient energy the bone is bent beyond its tensile strength and breaks; it may even be shattered violently, especially the diaphysis of a long bone. This phenomenon is an example of a fracture without a direct hit by a bullet. The same mechanism takes place during fracture from a blunt injury, except that the energy transfer bending the bone occurs from the outside.

Tissue heterogeneity

Apart from these considerations of tissue elasticity, there is the particular local anatomy to consider. A block of gelatine approximates skeletal muscle in its elasticity and density, but it is homogeneous. A human limb or the face are a mixture of rigid

and elastic tissues comprising fascial compartments enclosing muscles, tendons and ligaments, large blood vessels and nerves, and bones. Not only does each anatomic element have a different elastic tensile strength, but the mixture of elements has its own interactive properties. The various interactions of projectile, soft-tissue organs and bony fragments can become very complicated.

Bound together and boundaries

How tightly tissues are bound together, and to nearby structures such as fascia of varying thickness, also affects the amount of permanent damage a cavity stretch can leave behind. If one side of a structure is fixed and the other side has free movement, a shearing force develops. Fascial planes may also serve as channels for the dissipation of energy – path of least resistance – to more remote tissues.

Boundaries limiting the pulsation of a temporary cavity can be very significant. They include not only fascia and aponeuroses, but also hollow, fluid-filled organs: the brain in the rigid cranium, the heart, the full urinary bladder. An empty stomach will be simply transfixed by a stable bullet. The same bullet will cause a full stomach to "explode".

A graphic laboratory example of boundary effect is found in Figures 3.26.1 and 3.26.2. The two pictures show the results with a high-energy bullet cavity. The apple literally explodes after passage of the bullet. Note that the cavitation effect *always follows* the passage of the bullet; it is not simultaneous.

Harold & Esther Edgerton Foundation, courtesy of Palm Press, Inc.

Harold & Esther Edgerton Foundation, courtesy of Palm Press, Inc.

Figures 3.26.1 and 3.26.2

Demonstration of boundary effect on the temporary cavitation due to a high-energy rifle bullet: the apple literally explodes after the passage of the bullet.

Clinically, these effects can create apparent paradoxes. For example, a heavy and slow bullet can cause a more severe wound in highly elastic tissue, such as lung, than a lighter, faster bullet with more kinetic energy. The heavier, slower bullet produces more crush. The faster, lighter one uses up more energy in creating the temporary cavity, which leaves little residual damage. However, the faster and lighter bullet will create more damage in less elastic tissue (liver, brain), which cannot accommodate stretch as well.

Transfer of kinetic energy does not affect all tissues to the same extent and in the same way.

3.4.4 Pathological description of the permanent wound cavity

The permanent cavity of the final wound that the surgeon sees is the result of a mixture of crushing, laceration and stretching of the tissues. Most of the damage in the majority of ballistic wounds is due to direct crush and laceration.

A number of macroscopic and microscopic changes occur. Severe vasoconstriction blanches the skin around the entry wound for 3 – 4 hours owing to the stretch. This is followed by a hyperaemic reaction that lasts up to 72 hours.[8]

8 Fackler ML, Breteau DVM et al. Open wound drainage versus wound excision in treating the modern assault rifle wound. *Surgery* 1989; **105**: 576 – 584.

In the wound of skeletal muscle itself, three histological zones have been described (Figure 3.27).[9]

1. Crush zone of the wound channel filled with torn and necrotic tissues: 2 – 4 times the diameter of the bullet.
2. Contusion zone of muscle adjacent to the missile tract: the thickness of this zone is variable and averages about 0.5 cm; tissue damage is irregular and uneven.
3. Concussion zone of variable distance with congestion and extravasation of blood: the stretch is not great enough to tear tissues, but sufficient to injure capillaries. The demarcation between the contusion and concussion zones is not always well defined.

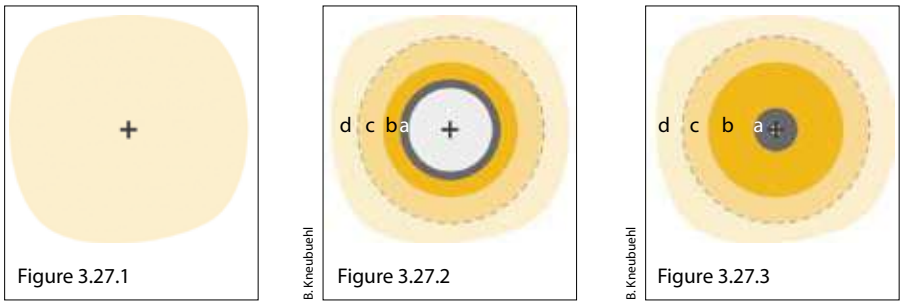

Figure 3.27

Schematic drawing of histopathological changes in the shooting channel:

3.27.1 geometric shooting channel
3.27.2 phase of maximum temporary cavitation
3.27.3 final wound channel

a: zone of crushed tissues
b: contusion zone
c: concussion zone
d: unaffected tissues

These histological changes are more severe with high kinetic-energy transfer – tumbling or deformation of the bullet – and do not change notably over 72 hours. The tissue damage is irregular along the wound track and how much of the damage is irreversible is not immediately apparent. This has obvious consequences for wound management and the extent of surgery necessary (see Chapter 10).

Bacteria are also seen within the permanent cavity. The sucking action of cavitation aspirates bacterial flora from outside, and projectiles are not sterile; the heat of firing is not high enough for long enough to sterilize the bullet.

3.4.5 Clinical applications

The external appearance of a bullet wound can be deceptive. Tiny entrance and exit holes can be associated with extensive internal damage.

The length of the shooting channel in the body, the existence of an exit wound, and the characteristics of the particular structures traversed, will all greatly affect the final wound caused by a high-energy bullet that the surgeon sees.

The position of the phase 2 temporary cavity along the wound track is clinically very relevant. Figure 3.28 demonstrates the trajectory of an FMJ rifle bullet in a laboratory soap block with a simple entry wound and three possible exit wounds.

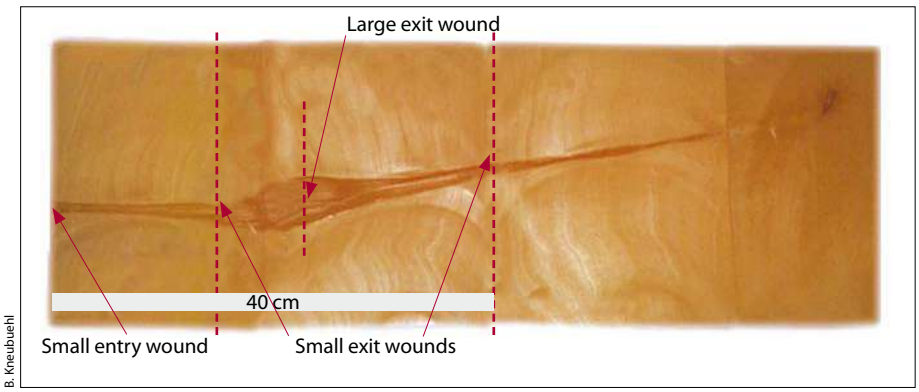

Figure 3.28

Exit wounds may occur before, during or after cavitation, depending on the length of the shooting channel in the body.

9 Wang Z, Feng JX, Liu YQ. Pathomorphological observation of gunshot wounds. *Acta Chir Scan* 1982; **508**: 185 – 189.

The injured body part may not be long enough for temporary cavitation to occur. Figures 3.29.1 and 3.29.2 show small entry and exit wounds along the narrow channel of phase 1, with little intervening tissue damage.

R. Coupland / ICRC

Figure 3.29.1

Small entry and exit wounds, through-and-through injury.

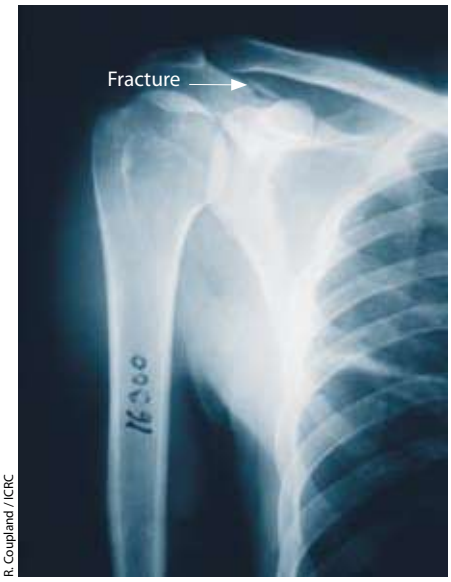

R. Coupland / ICRC

Figure 3.29.2

The radiograph shows a small drill-hole fracture of the acromion: phase 1 narrow-channel injury only.

The wound is large when the exit takes place *during* cavitation (Figures 3.30.1 – 3.30.3).

R. Coupland / ICRC

Figure 3.30.1

Gunshot wound to the thigh; the small entry is medial and the large exit on the lateral aspect.

R. Coupland / ICRC

Figure 3.30.2

There is a severe comminuted fracture of the femur and fragmentation of the bullet.

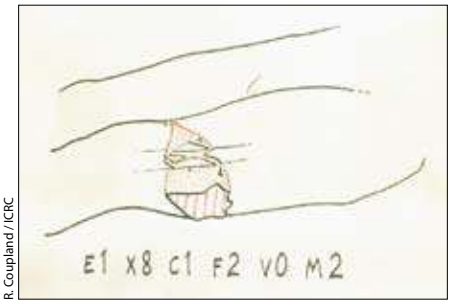

R. Coupland / ICRC

Figure 3.30.3

The exit occurred during cavitation. The fragmentation of the bullet is a tell-tale sign of severe tissue damage. Wound Score: Grade 3, Type F (see Chapter 4).

When the small exit wound occurs *after* cavitation, the *intervening* damage can be severe (Figures 3.31.1 and 3.31.2).

R. Coupland / ICRC

Figure 3.31.1

Small entry and exit wounds with severe damage in between.

R. Coupland / ICRC

Figure 3.31.2

The same wound after excision and partial healing.

With a handgun bullet, the cavitation effects are minimal and the final wound channel is almost entirely due to crush whatever the length of the shooting channel (Figure 3.32.1 and 3.32.2).

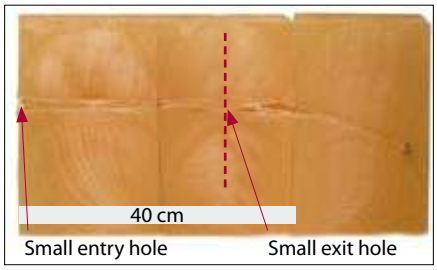

B. Kneubuehl

Figure 3.32.1

Small entry and exit wounds from an FMJ handgun bullet: demonstration in a soap block.

B. Kneubuehl

Figure 3.32.2

Handgun bullet wound to the thigh.

The case of bone

The direct impact of a bullet on bone is different according to where this occurs in the shooting channel. Three clinical situations arise, corresponding to the phases of the shooting channel. In the first case, the stable bullet of phase 1 causes a small "drill-hole" fracture. A small cavitation occurs, and the bone then collapses on itself. The final hole is smaller than the calibre of the bullet (Figure 3.29.2).

In the second case, the bone is broken into multiple fragments owing to a higher release of kinetic energy from the larger cross section of the tumbling bullet during phase 2 (Figure 3.33). Each bone fragment creates its own local crush damage; the muscle is cut up in between the fragments. Afterwards, the temporary cavitation acts on this shredded muscle with lowered tensile strength; the cavity is greater and the final remaining wound channel is larger. The bone fragments always remain within the temporary cavity; they do not produce a new second wound outside the cavity. Detached pieces of muscle and the bone fragments are found inside the final wound, which is usually very severe. This is of obvious clinical relevance to the management of these wounds.

The damage that occurs to bone during the phase 3 shooting channel depends on how much kinetic energy remains in the bullet. If it is a very small amount, the bone will stop the bullet without fracturing.

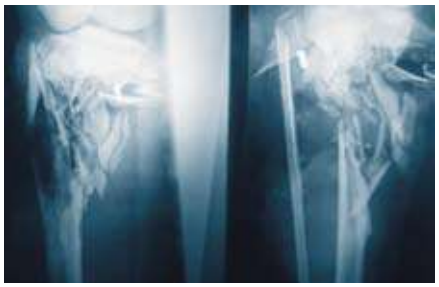

ICRC

Figure 3.33

Severely comminuted fracture of the tibia.

Ricochet

As seen in the tissue simulants, a ricochet FMJ bullet creates a wound that resembles that of a soft-nosed, dum-dum bullet: early dissipation of energy through crush and stretch (Figure 3.34). This phenomenon can be significant for a soldier wearing body armour. If the armour is pierced by a bullet, the wound may be more severe than if he had not been wearing the protective device.

R. Coupland / ICRC

Figure 3.34

The wound of a ricochet FMJ bullet resembles that of a dum-dum bullet. The head of the humerus has literally exploded.

<u>Please note</u>:

People are not injured in the anatomic position. The situation of a muscle group may alter with the patient's position so that a track is obscured, lengthened or shortened.

Disruption of the bullet and secondary fragments

At impact velocities above 700 m/s, standard FMJ bullets tend to fragment in tissues at short shooting distances (< 30 – 100 m). Some of these fragmented pieces make their own track of crushed tissue. Tissue disruption from bullet fragmentation is then synergistic: the tissues are multiply perforated by the metallic fragments, thus weakening the tissue's cohesiveness, before being subjected to cavitation stretch. The injury is severe. The same phenomenon occurs with shotgun pellets; each pellet creates its own crush track.

Clinically, an X-ray showing a "shower of lead" from a fragmented bullet should alert the surgeon to the likelihood of a large transfer of kinetic energy to the tissues and consequent severe damage (Figures 3.35, 4.5 and 10.5).

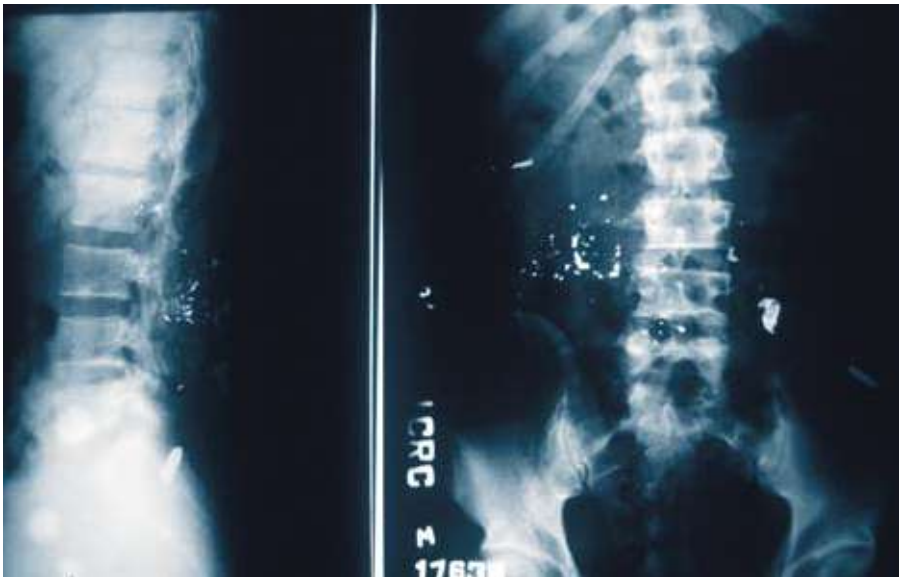

ICRC

Figure 3.35
Bullet fragmentation: a "shower of lead".

Other secondary missiles are objects to which the bullet imparts sufficient motion through the transfer of kinetic energy. These may include foreign bodies such as a belt buckle, a small stone, the metallic contents of a pocket, or the contents of flak jackets – or autologous bodies including the teeth, dental fillings or dentures, and bone fragments (once an ICRC surgeon found a piece of a shattered mandible embedded in a patient's neck).

3.4.6 Sonic shock wave in tissues

A projectile in flight is accompanied by a series of waves propagated at the speed of sound (330 m/s in air). When the projectile strikes a person, this sonic wave is propagated throughout the body at the speed of sound in tissue (4 times the speed of sound in air).

This sonic shock wave may have high amplitude but is of very short duration, not enough to move or injure tissue. However, researchers have noted microscopic cell changes as well as stimulation of peripheral nerves, if the pressure generated reaches a certain threshold. The nerve stimulation occurs immediately, while the cell damage is apparent only after 6 hours. Occasional, short-lived neurapraxia appears to be the only clinically relevant damage.

3.4.7 Pressure wave in tissues and blood vessels

This pressure wave is part of the phenomenon of the temporary cavitation and should *not be confused* with the sonic shock wave. The outer boundary of the cavity consists of a "front" of compressed tissues, which creates a pressure wave at the limit of tissue compression. The wave develops fully only after the cavitation and decreases with distance. It is measured in milliseconds (1,000 times longer than the shock wave). The pressure wave can thus disrupt capillaries or provoke their thrombosis, rupture a full bowel or the liver, and cause retinal detachment and fractures, at a distance from the cavity.

Additionally, blood vessels contained in tissues are squeezed and emptied suddenly with tissue compression. This creates a pressure wave in the column of blood that is propagated away from the site of cavitation. The clinical result of this pressure wave along the course of a blood vessel can be thrombosis or dissection of the intima or muscularis.

3.4.8 Fragment wounds

Their non-aerodynamic shape means that fragments rapidly lose speed in air. If initial velocities are reported to be up to 2,000 m/s, the impact velocity in survivors is usually much lower. If the person is very close to the exploding device, the penetration is deep. If very far away, the casualty is only "peppered" superficially with multiple fragments.

Fragments do not tumble in the tissues, as was demonstrated in the tissue simulants. Thus, most tissue damage is due to crush. At the end of the trajectory, the sharp edges of an irregular fragment will cut tissues. By contrast, a bullet tends to push aside tissues at the end of its track. The wound profile resembles a *cone of tissue destruction*, with the largest diameter at the entry since most energy is dissipated at the surface. The diameter of the entry wound can be anywhere between 2 and 10 times the calibre of the fragment, depending on the impact velocity, mass and shape (Figures 3.36.1 – 3.36.3).

Large, slow fragments penetrate and crush tissues more; small, fast ones result in more stretch. This means that a large and slow fragment tends to cause the same type of wound whatever the tissues, while the wounds of small and fast fragments vary according to tissue elasticity. Nonetheless, in all fragment wounds the extent of tissue damage is always larger than the fragment.

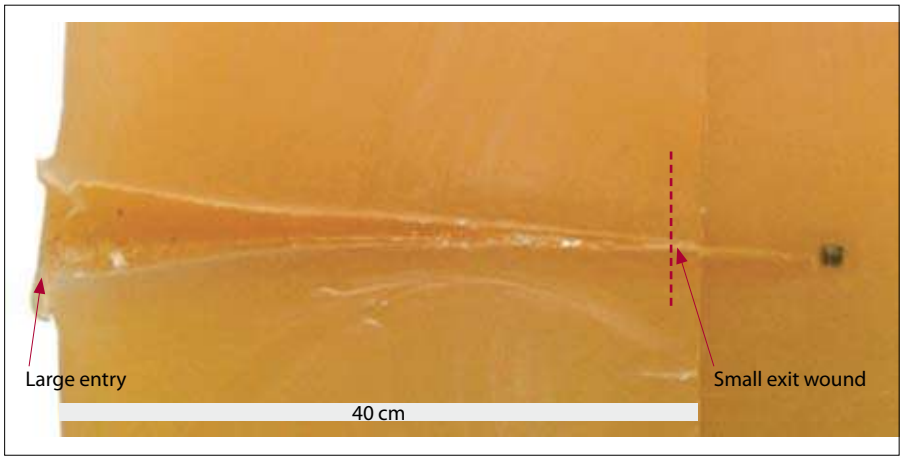

B. Kneubuehl

Figure 3.36.1

Soap block demonstration of fragment injury.

R. Coupland / ICRC

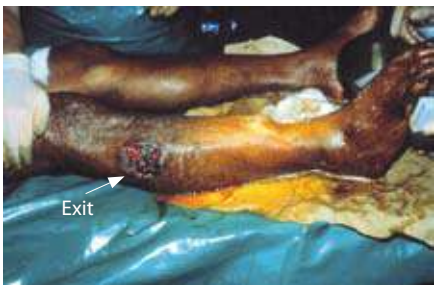

R. Coupland / ICRC

Figures 3.36.2 and 3.36.3

Fragment wound with entry larger than exit.

Wound channel: pathological description

The sharp and irregular surfaces of fragments carry pieces of skin and clothing material into the wounds. As with gunshot wounds, fragment wounds show different zones of tissue injury. The cavity contains necrotic tissues surrounded by a zone of muscle fibre fragmentation with haemorrhage both within and between fibres, further surrounded by a region of acute inflammatory changes and oedema.

The uneven distribution of tissue damage in the concussion and contusion zones has been described as "jumping" or a "mosaic".[10] This is probably because energy has been transferred both along (crush effect) and across (shear and stretch effect) muscle fibres.

10 Wang ZG, Tang CG, Chen XY, Shi TZ. Early pathomorphologic characteristics of the wound track caused by fragments. *J Trauma* 1988; **28 (1Suppl.)**: S89 – S95.

3.5 Wound dynamics and the patient

The wound has a life history, from injury to complete healing. As we have seen, tissue damage depends on a number of physical factors. However, the physical effects of energy transfer do not tell us all the pathological and physiological consequences of the act of wounding.

The tissues in and around the wound undergo reversible and irreversible pathological changes, as well as inflammatory reactions. It can be extremely difficult to diagnose injured tissue that will heal from that which is nonviable and will not heal (see Chapter 10).

Total kinetic energy is the *potential* to cause damage; transferred kinetic energy is the *capacity* to cause damage. Actual tissue damage depends on the *efficiency* of this energy transfer.

The most valuable information gained from this discussion is that small missiles may cause small and minor wounds, or large and serious wounds; and a small entry wound can be associated with large internal damage. Nothing can replace a good clinical examination of the patient and the wound. The extent of the crush and stretch is well represented by the grades of the Red Cross Wound Score (see Chapter 4).

The injurious effects on the patient, however, involve even more than the local pathology. As with all trauma and many diseases, the physiological and psychological state of the victim must be taken into consideration as well. Fit young military personnel, well-trained for their role in warfare and mentally prepared for being wounded, and wounding and killing others, are not the same as civilians. Psychological status cannot be calculated by ballistics. Only these factors can explain the many anecdotes of a person being shot – even several times – and yet continuing to advance or fight in combat.

Chapter 4

RED CROSS WOUND SCORE AND CLASSIFICATION SYSTEM

4. RED CROSS WOUND SCORE AND CLASSIFICATION SYSTEM

4.1 Applications of the RCWS and classification system 83
4.1.1 Assessment of and communication about war wounds in a standardized scheme 83
4.1.2 Establishing a scientific approach to war surgery 83
4.1.3 Surgical and hospital audit 83
4.1.4 Wound information from the field 83

4.2 Principles of the Red Cross Wound Score 84
4.2.1 Examples 86
4.2.2 Notes on wound scoring 87

4.3 Grading and typing of wounds 88
4.3.1 Wound Grade according to amount of tissue damage 88
4.3.2 Typing wounds according to the injured tissues 88

4.4 Wound classification 88

4.5 Clinical examples 89

4.6 Conclusions 91

4.1 Applications of the RCWS and classification system

Surgeons from civilian practice usually have no previous experience of managing war wounds. Even experienced surgeons often do not find it possible to tell the true extent of tissue damage. As noted in Chapter 3, ballistics studies show that there is not a uniform pattern or degree of wounding. Preparation for war surgery involves an understanding of the translation of the kinetic energy of a wounding projectile into tissue damage, i.e. the actual wounds that the surgeon must deal with.

The severity of such wounds depends on the degree of tissue damage and the structure(s) that may have been injured: thus, the clinical significance of a wound depends on its size and site. The Red Cross Wound Score (RCWS) and classification system[1] is based on the features of the wound itself, and not on the weaponry or the presumed velocity or kinetic energy of the missile.

Any wound classification system will aid the surgeon if it helps to assess the severity of the injury, influences surgical management, predicts outcome, and provides for an accurate database that can be used in comparative studies. The RCWS goes a fair way to fulfilling these criteria.

4.1.1 Assessment of and communication about war wounds in a standardized scheme

The Wound Score is a useful clinical tool to communicate the severity of the wound to staff and colleagues without having to remove the dressings. It is also an element to take into consideration for the purposes of triage.

4.1.2 Establishing a scientific approach to war surgery

The RCWS permits comparison of treatments and prognoses of similar wounds. An analogy can be made with the TNM (tumour, node, metastasis) staging of different cancers, allowing comparison of treatment regimes for a similar pathology. In the case of a gunshot wound to the thigh, treatment and prognosis will differ according to the amount of tissue damage, degree of bone comminution and whether there is injury to the femoral vessels. Other wound scoring systems are primarily devised for blunt trauma. The presence of a penetrating wound in these systems often indicates a "serious" wound, with little further nuance.

4.1.3 Surgical and hospital audit

Wound Scores can be used to evaluate quality of care when combined with information on length of hospital stay, number of operations performed per patient or units of blood used, morbidity and mortality. An example relating to the adequacy of primary wound surgery is the number and cause of deaths associated with non-vital wounds, or the number of operations performed per patient for each Wound Grade.

4.1.4 Wound information from the field

ICRC surgical teams treat thousands of war-wounded every year. Analysis of a larger number of scored wounds will eventually clarify the relationship between experimental laboratory wound ballistics and the clinical management of war wounds. Worthy of note is the fact that information gained in the field served as a scientific basis for the campaign to ban anti-personnel landmines, thus promoting new standards in international humanitarian law.

1 This chapter is largely based on the brochure *The Red Cross Wound Classification, Revised Edition*, by Dr R. M. Coupland. The ICRC Master Surgeons Workshop held in Geneva in 2002 revised the original Wound Score (see Introduction).

R. Coupland / ICRC

Figure 4.1

The width of two fingers is approximately the length of a military rifle bullet.

F. Herkert / ICRC

Figure 4.2.1

Gunshot wound: F1 fracture of the ulna.

R. Coupland / ICRC

Figure 4.2.2

Gunshot wound: F2 fracture of the femur.

R. Coupland / ICRC

Figure 4.2.3

Gunshot wound: F 1 fracture (clinically insignificant comminution) of the fibula.

4.2 Principles of the Red Cross Wound Score

Wounds are given a six-figure Score according to several parameters. The Scores can then be translated into a classification system:

- grading of the wound according to the severity of tissue injury and
- typing of the wound according to the tissue structures injured.

The RCWS is intended for quick and easy use in the field, its simplicity especially useful under conditions of stress. It is a simple clinical system that requires no extra equipment or sophisticated procedures. The time required for scoring wounds can be measured in seconds.

E	entry wound in centimetres	
X	exit wound in centimetres (X = 0 if no exit wound)	
C	cavity	Can the cavity of the wound take two fingers before surgical excision?
		C 0 = no
		C 1 = yes
F	fracture	Are any bones fractured?
		F 0 = No fracture
		F 1 = Simple fracture, hole or insignificant comminution
		F 2 = Clinically significant comminution
V	vital structure	Are dura, pleura, peritoneum, or major peripheral vessels injured?
		V 0 = no vital structure injured
		V N = (neurological) penetration of the dura of the brain or spinal cord
		V T = (thorax or trachea) penetration of the pleura or of the larynx/ trachea in the neck
		V A = (abdomen) penetration of the peritoneum
		V H = (haemorrhage) injury to a major peripheral blood vessel down to brachial or popliteal arteries, or carotid artery in the neck
M	metallic body	Are bullets or fragments visible on X-ray?
		M 0 = no
		M 1 = yes, one metallic body
		M 2 = yes, multiple metallic bodies

Table 4.1 Parameters of Wound Score.

E (entry)	centimetres
X (exit)	centimetres
C (cavity)	C 0, C 1
F (fracture)	F 0, F 1, F 2
V (vital structure)	V 0, VN, VT, VA, VH
M (metallic bodies)	M 0, M 1, M 2

Table 4.2 Total scheme of Wound Score.

Wounds are scored on the patient's admission form after operation or after initial examination if there is no surgery.

ICRC EXPERIENCE

An ICRC admission sheet with Wound Score recorded.

ICRC

NAME: A. Victim
NUMBER: 16838
COMING FROM: The Borderk.
MALE FEMALE AGE: 40

DATE: 4.3.90 TIME: 15.00
GSW MI: SHELL: BOMB: BURNS: OTHER:
TIME SINCE INJURY: 4 hours
GENERAL CONDITION: OK
PULSE 90 BP: 110 RESP: 25 TEMP: N
ANTIBIOTICS: Penicillin 5 mega
ATS/ANATOXAL: ✓

◆ MEDICAL ASSESSMENT

GSW (L) Thigh
Femur
Pulse & Sensation – normal

MEDICAL ◆ ORDERS

Hb 12.5
Hct
X match

IV fluids 1L N. Saline
NPO from 8.00 am.

TRIAGE: I Serious II Secondary III Superficial IV Supportive

◆ OPERATION NOTE

4.3.90

Excision GSW (L) thigh
Large normal
Dead muscle & bone fragments excised

– Saline wash
– Dry Bulky dressing
Traction pin.

◆ POST OPERATIVE INSTRUCTIONS

Antibiotics: Penicillin 5 mega qid
to stop 48 hours
Position Physio drains traction
4 kg traction
By mouth Food Fluids Nil
Other
Next in OT: 9.3.90 JPC.

◆ PENETRATING WOUND SCORE ◆

E	1	X	8	C	1	F	2	V	0	M	2
E		X		C		F		V		M	

◆ OTHER INFORMATION

4.2.1 Examples

The following two sketches (Figures 4.3 and 4.4) present various wounds due to weapons, and their assessment according to the ICRC's wound scoring system.

Figure 4.3

Example of Wound Scores.
a. Simple bullet track.
b. Track produced by bullet with temporary cavitation at exit.
c. Simple track with involvement of a vital structure (artery).
d. Low-energy transfer wound with simple fracture.
e. High-energy transfer wound by fragment with comminuted fracture.

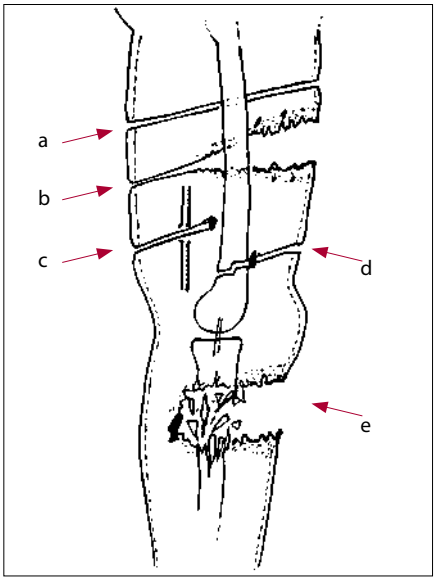

ICRC

	E	X	C	F	V	M
Wound (a)	1 ?	2	0	0	0	0
Wound (b)	1	4	1	0	0	0
Wound (c)	1	0	0	0	H	1
Wound (d)	1	0	0	1	0	1
Wound (e)	6	0	1	2	0	1

Figure 4.4

Further example of Wound Scores.
a. Through-and-through bullet wound of soft tissue with central cavitation.
b. Through-and-through bullet wound with central cavitation and comminuted fracture.
c. High-energy transfer bullet wound with cavitation and comminuted fracture.

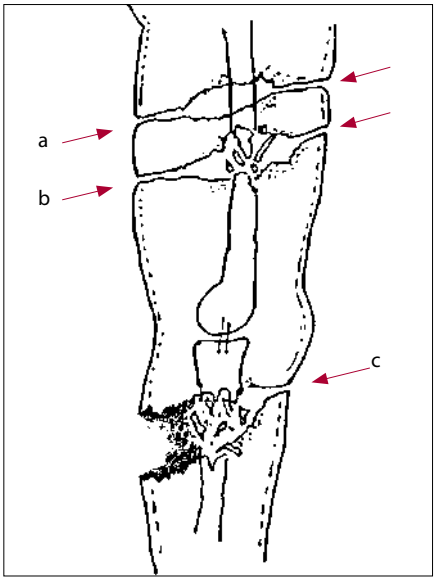

ICRC

	E	X	C	F	V	M
Wound (a)	1 ?	1	1	0	0	0
Wound (b)	1 ?	1	1	2	0	0
Wound (c)	1	6	1	2	0	2

4.2.2 Notes on wound scoring

1. When it is impossible to differentiate between the entry and the exit wounds, put a (?) between the E and X Scores.
2. With multiple wounds, only the two most serious are scored.
3. If a wound cannot be scored (unclassifiable), U/C is written on the Score. This applies to a minority of wounds.
4. When one missile causes two separate wounds (e.g. through the arm and into the chest), the 2 separate Scores are joined by a bracket.
5. Include penetrating wounds only, not superficial tangential skin wounds; e.g. skin wound 20 cm long and 1 cm wide but without penetration deep to the fascia.
6. Estimation of the cavity: the width of two fingers is approximately the length of a bullet from a military assault rifle (Figure 4.1). If the permanent cavity of a wound admits two fingers before surgical excision, then something other than crush by a bullet travelling sidelong must have taken place, i.e. stretch and laceration by cavitation. A C1 wound is likely to present significant tissue damage whatever the cause.
7. Fracture severity: it is inevitable that some wounds fall between F1 and F2, but for simplicity this is not accurately defined here. A further refinement is given in Volume 2. An example of a clinically insignificant comminution (F1) is a wound with a comminuted fibula but with an intact tibia (Figure 4.2.3).
8. A vital injury implies a more dangerous wound involving a surgical task in addition to simple wound management (e.g. craniotomy, pleural drainage or thoracotomy, laparotomy). VH includes the popliteal and brachial vessels but not those more distal. The outcome of injuries to the head, thorax or abdomen, or causing massive peripheral haemorrhage, is only partly determined by the clinical size of the wounds as determined by the RCWS (see below).
9. Metallic fragments: note the difference between an intact bullet (M1) and a fragmented bullet (M2), as shown in Figure 4.5. If the full metal jacket of a bullet has been disrupted and the lead interior has leaked out as fragments, this indicates a severe stress on the bullet and a large transfer of kinetic energy to the tissues (Figures 3.35 and 10.5).

   Note also the difference between multiple metallic fragments from a grenade or shell (Figure 4.6) – which are common and do not necessarily represent a large transfer of kinetic energy – and a disrupted and fragmented bullet, which does. If there are numerous metallic fragments, do not count them; only mark M = 2.
10. If X-rays are not available the RCWS is still valid. The F Score is judged clinically and the M parameter omitted, or included if fragments are found during surgical excision. X-rays should not be performed just to fill in the M and F Scores.
11. Traumatic amputation of limb (Figure 4.7): this injury is often seen with anti-personnel blast mines. The open end of the amputated limb equals a combined entry-exit, whose diameter represents the sum of E+X. It is equivalent to a cavity (C = 1), while the vaporized missing part of the limb qualifies as a severe fracture (F = 2). The level of traumatic amputation (above or below the knee or elbow) will determine if the Vital structure Score is 0 or H.

ICRC

Figure 4.5

Disrupted and fragmented bullet: M2.

ICRC

Figure 4.6

Multiple shell fragments: M2.

K. Barrand / ICRC

Figure 4.7

Traumatic amputation of the forearm.

E	X	C	F	V	M
20	?	1	2	0	0

4.3 Grading and typing of wounds

Once scored, the wound can be *graded* according to severity (E, X, C and F), and *typed* according to tissue type (F and V).

4.3.1 Wound Grade according to amount of tissue damage

The wound should be *graded* according to severity.

Grade 1
E + X is less than 10 cm with Scores C 0 and F 0 or F 1.
(Low energy transfer.)

Grade 2
E + X is less than 10 cm with Scores C 1 or F 2.
(High energy transfer.)

Grade 3

E + X is 10 cm or more, with C 1 or F 2.
(Massive energy transfer.)

These Grades represent the outcome of a simple clinical assessment that corresponds to the effective transfer of kinetic energy of projectiles to body tissues. Large wounds are more serious and require greater resources; this is particularly true of wounds to the limbs.

4.3.2 Typing wounds according to the injured tissues

Once *scored*, the wound can be *typed* according to tissue structures.

Type ST
Soft-tissue wounds: F 0 and V 0.

Type F
Wounds with fractures: F1 or F 2, and V 0.

Type V
Vital wounds putting the patient's life at risk: F 0 and V = N, T, A or H.

Type VF
Wounds with fractures and involving vital structures putting life or limb at risk: F 1 or F 2 and V = N, T, A or H.

4.4 Wound classification

Combining Grades and Types gives rise to a classification system divided into 12 categories.

	Grade 1	Grade 2	Grade 3
Type ST	1 ST Small, simple wound	2 ST Medium soft-tissue wound	3 ST Large soft-tissue wound
Type F	1 F Simple fracture	2 F Important fracture	3 F Massive comminution
Type V	1 V Small wound threatening life	2 V Medium wound threatening life	3 V Large wound threatening life
Type VF	1 VF Small wound threatening limb and/or life	2 VF Important wound threatening life and/or limb	3 VF Large wound threatening life and/or limb

Table 4.3 Grade and tissue Type categories.

These categories help determine surgical management for a number of wounds, though by no means all (see Chapters 10 and 12).

Further refinement of the fracture type is possible, especially for assessing the extent of bone defect. This can be useful in a specialized study of war wounds with fractures (see Volume 2).

4.5 Clinical examples

Epidemiological studies from the ICRC surgical database of over 32,000 patients have shown that the Red Cross Wound Score gives excellent prognostic results for surgical workload and morbidity due to war wounds. As demonstrated in the statistical analysis in Chapter 5, the Wound Grade is highly relevant to the number of operations per patient. This is particularly the case for wounds to the extremities. Vital Wound Scores to the extremities (V = H) also give good measures for mortality and the amputation rate.

One of the avowed weaknesses of the Classification concerns the prognostic mortality of vital central wounds. It is in the anatomic nature of the brain, heart and great vessels that even a very small wound from a projectile with low kinetic energy can be lethal if a vital centre is injured. The RCWS does classify such wounds as "potentially" lethal; that is why, by definition, they are categorized as "vital" injuries. However, a Grade 1 wound can be just as lethal as a Grade 3 wound depending on what particular structure is hit: medulla oblongata or temporal cortex; lung parenchyma or ascending aorta.

What is certain is that for most Grade 3 wounds to the head, thorax and abdomen, pre-hospital mortality is bound to be high. Without autopsies of those killed in action (KIA), analysis can only be made of the relatively few "survivors" who reach hospital; for them, wound Grade is not significant with respect to mortality (Tables 4.4 – 4.6).

Thus, the outcome of injuries to the head, thorax and abdomen are not only, nor even mainly, determined by the clinical size of the wound as defined by the RCWS. The various factors affecting the mortality from vital wounds; e.g. for the abdomen, number of organs injured, degree of faecal contamination, intraoperative blood loss, delay in operative treatment, etc., will be discussed in Volume 2.

	Number	Deaths	% mortality
Grade 1	75	14	18.7
Grade 2	70	15	21.4
Grade 3	9	3	33.3

Table 4.4 Mortality from head and neck injuries (VN) : N = 154. (ICRC Kabul, 1990-92)

	Number	Deaths	% mortality
Grade 1	82	4	4.9
Grade 2	41	2	4.9
Grade 3	3	0	0

Table 4.5 Mortality from chest injuries (VT) : N = 126. (ICRC Kabul, 1990-92)

	Number	Deaths	% mortality
Grade 1	120	9	7.5
Grade 2	70	11	15.7
Grade 3	5	0	0

Table 4.6 Mortality from abdominal injuries (VA) : N = 195. (ICRC Kabul, 1990-92)

Nonetheless, as explained in Chapter 5, a distinction should be made in patient databases between superficial and penetrating wounds to the head, thorax and abdomen to allow for a better analysis of results of treatment. Using the RCWS enables this.

R. Coupland / ICRC

Figure 4.8.1

Two fragment wounds to the thigh. The surgeon is assessing the larger wound cavity.

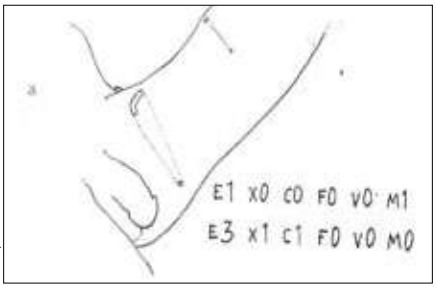

R. Coupland / ICRC

Figure 4.8.2

Both wounds are Type ST. The smaller wound is Grade 1, and the larger Grade 2.

R. Coupland / ICRC

Figure 4.9.1

Through-and-through gunshot wound to the knee.

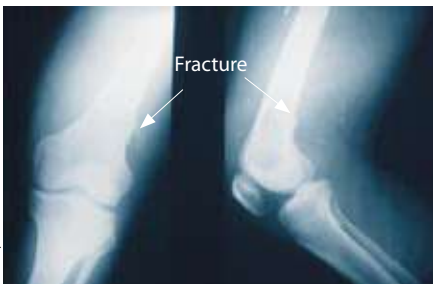

R. Coupland / ICRC

Figure 4.9.2

There is a small fracture above the lateral condyle.

R. Coupland / ICRC

Figure 4.9.3

The popliteal artery has been injured.

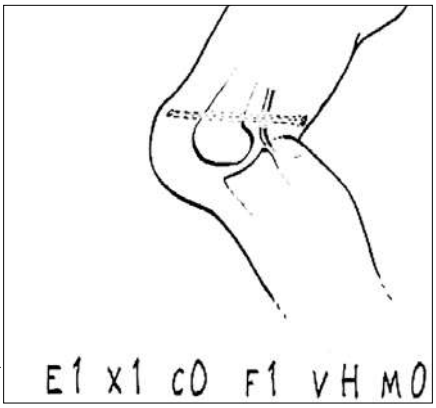

R. Coupland / ICRC

Figure 4.9.4

The wound is Type V(H)F, Grade 1.

R. Coupland / ICRC

Figure 4.10.1

Gunshot wound to the abdomen.

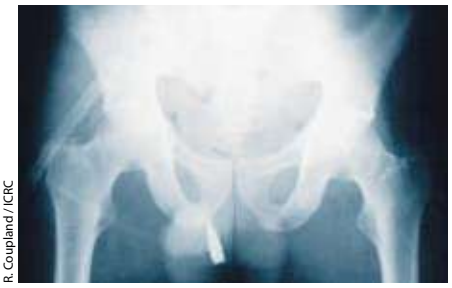

R. Coupland / ICRC

Figure 4.10.2

The bullet is lodged in the scrotum. There is a small fracture of the right pubic bone.

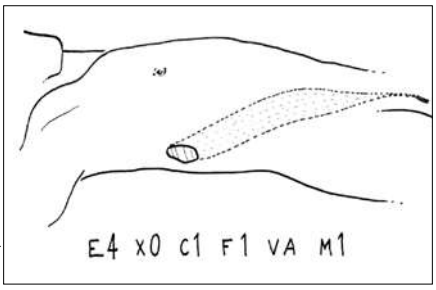

R. Coupland / ICRC

Figure 4.10.3

This is a Grade 2 wound, Type V(A)F.

R. Coupland / ICRC

Figure 4.11.1

A bullet has grazed the left buttock and then entered the right buttock at the anal cleft. The exit is located laterally.

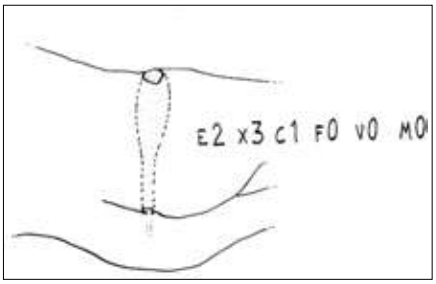

R. Coupland / ICRC

Figure 4.11.2

This is a Grade 2, Type ST wound.

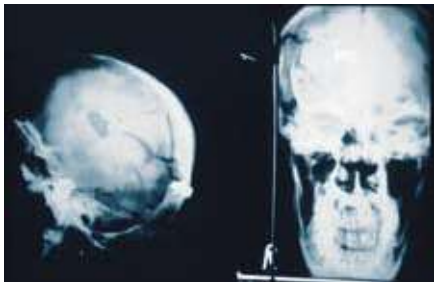

R. Coupland / ICRC

Figure 4.12.1

Fragment wound to the head with entry right parietal. Note the severe linear fractures.

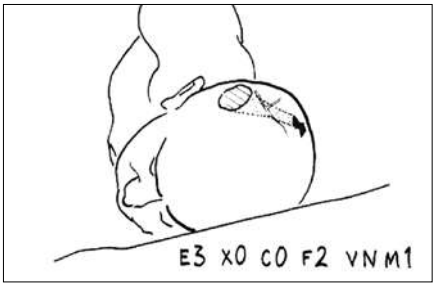

R. Coupland / ICRC

Figure 4.12.2

The cone of tissue destruction extends to the occipital lobe. Grade 2, Type V(N)F.

4.6 Conclusions

The Red Cross Wound Score enables medical staff to translate war wounds into surgical lesions rather than weaponry phenomena. It refines the heterogeneity of wounds and helps define them according to their clinical significance, and it is a good indicator of tissue damage due to the transfer of kinetic energy in penetrating wounds.

The limitations of the scoring system are recognized; complete accuracy cannot be obtained. It should be emphasized that the RCWS is for rapid use under adverse conditions, including extreme stress, and uses no additional equipment or sophisticated technology. The simplicity and advantages of the RCWS outweigh any disadvantages that may occasionally be introduced by observer error. As with any classification system, the more experience gained in using the RCWS, the more accurate the results and the fewer the observer errors.

Chapter 5

THE EPIDEMIOLOGY OF THE VICTIMS OF WAR

5. THE EPIDEMIOLOGY OF THE VICTIMS OF WAR

5.1 Introduction: purpose and objectives 95
5.1.1 The reader and this manual 95

5.2 Public health effects of armed conflict 95
5.2.1 Historical perspective 95
5.2.2 The public heath effects of social disruption 96
5.2.3 The burden of war-wounded civilians 97
5.2.4 Methodology 98

5.3 Epidemiology for the war surgeon 98
5.3.1 Preparation of the surgeon 98
5.3.2 Preparation of the institution 99
5.3.3 Surgical audit: monitoring 99
5.3.4 The surgeon and IHL 100
5.3.5 The surgeon and the specialized literature 100

5.4 General questions of methodology 101
5.4.1 Large and small numbers 101
5.4.2 Problems of data collection 101
5.4.3 Important first questions: who is counting whom? 102
5.4.4 Some definitions: the wounded and the dead 103
5.4.5 Clinical and operational consequences and IHL 104

5.5 Aetiology of injury 104
5.5.1 Warfare scenarios 104
5.5.2 Definitions of weapon-systems: methodology 105
5.5.3 ICRC statistics 106

5.6 Anatomic distribution of wounds 106
5.6.1 Body exposure 106
5.6.2 Definitions and methodology 107
5.6.3 Historical results 108
5.6.4 ICRC results 108
5.6.5 Primary tissue injury 109

5.7 Fatal injuries 110
5.7.1 Site 110
5.7.2 Trimodal distribution of trauma mortality 111
5.7.3 Ratio of dead to survivors 113
5.7.4 Lethality of weapons 113
5.7.5 Conclusions of clinical relevance 114

5.8 The lethality of context: delay to treatment 114
5.8.1 Historical developments 114
5.8.2 Forward projection of resources 115
5.8.3 Urban warfare: hospitals on front lines 115
5.8.4 The paradox of early treatment: changing mortality ratios and rates 115

5.9 Hospital mortality 117
5.9.1 Historical considerations 117
5.9.2 Hospital mortality *versus* post-operative mortality 117
5.9.3 ICRC hospital mortality 117

5.10 ICRC statistical analysis of hospital workload 118
5.10.1 Surgical audit: methodology 118
5.10.2 Number of operations per patient: all patients 118
5.10.3 Number of operations according to delay in evacuation 119
5.10.4 Number of operations according to Wound Grade 119
5.10.5 Number of operations according to Wound Grade and delay in evacuation 120
5.10.6 Number of operations according to weapon 120

5.11 Conclusions: lessons to be gained from a study of epidemiology 121
5.11.1 Setting up a surgical database for the war-wounded 122

ANNEX 5. A ICRC surgical database 123

ANNEX 5. B Setting up a surgical database for the war-wounded 125

5.1 Introduction: purpose and objectives

Epidemiological studies are standard practice in medicine. Doctors need to understand the characteristics of their patient population and the relationships between particular circumstances and the outcome of treatment. These elements help to establish the high risk factors.

As regards the victims of war, the total effects of conflict must be analysed, the impact on civilian populations as well as the military, the direct effects of war trauma and the indirect public health consequences. These help the ICRC and other organizations determine their response in terms of assistance (see Chapter 2).

5.1.1 The reader and this manual

Most readers of this manual will be civilian surgeons who are unacquainted with the practice of war surgery and its literature. Even among military surgeons many may have only a superficial acquaintance with these topics. This chapter on the epidemiology of victims of war presents a certain approach to the understanding of the many factors involved in patient care and outcome.

First the public health repercussions of armed conflict will be dealt with briefly, then the purely surgical aspects will be addressed in more detail. Questions of general methodology and definitions will begin the surgical study, and will be followed by a presentation of historical and ICRC experience and results. Relevant clinical conclusions will be emphasized throughout, as will recurring problems of methodology.

For the purpose of this discussion the authors refer directly only to the ICRC's experience in armed conflict and the importance of epidemiological studies in determining ICRC surgical activities. In addition, reference is made to the statistics in standard and well-accepted scholarly books and journal articles for historical results that have come to influence clinical practice (see Selected bibliography).

5.2 Public health effects of armed conflict

5.2.1 Historical perspective

It is commonly said that contemporary warfare causes more civilian victims than military, and that the public health effects are greater than those caused by trauma. In all major conflicts over the last one hundred years, with few exceptions (World War I), civilians have become a direct target and accounted for one-half to two-thirds of the casualties.

> "Throughout history there have been many small local wars, often involving wholesale starvation and occasionally the elimination of whole populations or tribes."
>
> G.W. Odling-Smee[1]

The heavy civilian toll in contemporary conflicts is therefore not new. Nevertheless, in certain recent conflicts – whether revolutionary wars of national liberation, civil wars, territorial and border disputes, or peasant uprisings – the destabilization of political, social, and economic infrastructures, the destruction of cultural symbols and institutions, and the psychological terrorizing of civilians have become explicit political and military goals. A number of reports have indicated that the indirect effects of social disruption have caused two to fifteen times more civilian deaths

1 Odling-Smee GW. Ibo civilian casualties in the Nigerian civil war. *BMJ* 1970; **2**: 592 – 596.

than war wounds have, and, in some cases, civilians may account for up to 90 % of the victims.

> "Today's armed conflicts are essentially wars on public health."
>
> Rémi Russbach, former Chief Medical Officer at the ICRC

5.2.2 The public heath effects of social disruption

Social disruption includes population displacement – either internal or refugee – pauperization, the breakdown of nutrition and sanitary conditions, the lack of safe drinking water, and the disorganization of medical systems (Table 5.1).

Direct effects	Indirect effects
Death	Economic pressures and disruption
Disability	Reduced food production and distribution Family disruption (orphans, abandoned children)
Destruction of health services	Refugees Psychological stress
Disruption of health programmes	Effects on housing, water supply, and sewage disposal
Psychological stress	Economic pressures on those caring for the war-disabled
Illness	Environmental (landmines, deforestation)

Table 5.1 Effects of political violence on health and health systems.[2]

Pathologies and the pathology of the health system

Complex emergencies during armed conflict in poor countries tend to cause death by endemic communicable disease and malnutrition. In more developed countries, increased mortality rates due to trauma and chronic disease dominate. The lack of access to care, either because of poverty, the disorganization of the medical system, or for reasons of security, aggravates the problems.

To pre-conflict poverty and fragile health services are added the problems of the collapse of medical supply and distribution, the targeting of hospitals and clinics in violation of medical neutrality, and the neglect or abandonment of preventive health programmes. In addition, health professionals are often amongst the first people to flee a conflict zone. Having to provide for war trauma also means a greater investment in curative services to the disadvantage of the preventive. The entire health system becomes organized around the conflict and there is an imbalance between the needs of victims and the available public health services (Figure 5.1).

Moreover, all too often, increased military spending in a country at war is to the detriment of social and health investments. Another recent development is the creation of a war economy controlled by criminalized military networks. The extraction of diamonds and other precious stones, oil, timber, and various minerals, the trafficking of narcotics, and sheer banditry, take a further and severe social toll.

A public health approach shows the full range of the humanitarian consequences of the use of anti-personnel landmines, especially during the post-conflict period. In addition to death and disability, there are the long-term socio-economic effects of disrupted economic activity through the loss of farmland, livestock, water resources, and industrial capacity because of the presence of landmines.

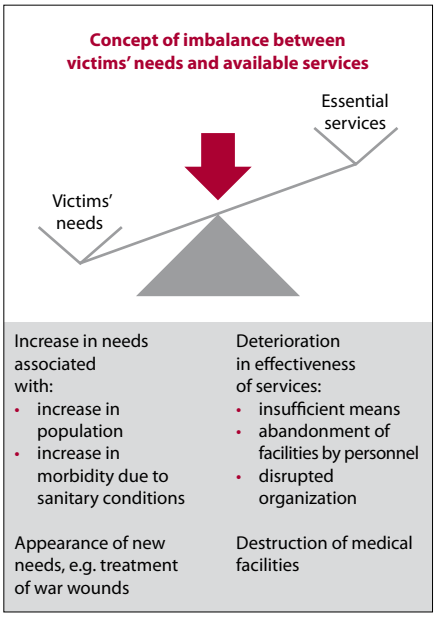

P. Perrin / ICRC

Figure 5.1

Victims' needs and available services during armed conflict. [3]

2 Zwi A, Ugalde A. Towards an epidemiology of political violence in the third world. *Soc Sci Med* 1989: **28**: 633 – 642. Cited in Lautze S, et al., 2004.

3 From Perrin P. *War and Public Health: A Handbook*. Geneva: ICRC; 1996.

Landmines and other UXO (unexploded ordnance) are not the only weapons left over after war. Demobilized combatants with guns are not often successfully reintegrated into social and economic life: criminal violence replaces political violence; peace combined with high criminality offers little respite from war. And the social and economic costs continue to cast a shadow over everyday life.

In addition, one must count the social costs of deliberate physical, psychological, and sexual abuse as methods of warfare. Torture and rape have profound and long-term consequences.

Public health effects on the military

Even among classical armies of industrialized countries, disease caused more deaths than combat until World War II. Mortality does not tell the whole story, however. Serious disease continues to affect modern armies, still accounting for high rates of non-availability of personnel for military duty. Of course, certain natural environments (jungle, high mountains) increase the incidence of non-battle pathology.

5.2.3 The burden of war-wounded civilians

Although trauma may account for only a small number of all war-related deaths in a given civilian population, the relative proportions of mortality and morbidity due to disease and trauma vary over time and between different war zones; refugee camps are different again.

In some conflicts, the burden of war trauma is greater than the public health consequences. This was the case in the wars in the former Yugoslavia (1991 – 99) in Lebanon (1975 – 90), and Rwanda (1994), among others. One must also take into consideration the relatively small population of certain countries; a limited number of deaths may represent a proportionately high death toll.

In some conflicts, the burden of the war-wounded is greater than the consequences on public health.

The type of combat may place civilians at greater risk and war casualties may overwhelm civilian medical services, even in situations where the public health effects are greater than the direct effects of trauma (Biafra 1967 – 70, Uganda 1987, Democratic Republic of the Congo since 1997 to the time of writing).

Mortality data, however, do not reflect the extent and severity of war injuries. Morbidity and disability rates due to injuries are testimony to the long-term socio-economic burdens.

Demographics of the victims

Historical military casualty statistics concern young, fit and healthy men. In the past, women were not close to combat in most classical armies; more recently, however, various revolutionary guerrilla groups have had female combatants.

When a civilian population finds itself in the midst of war, the demographic profile of casualties is closer to its population pyramid. This has clinical consequences for the treatment of endemic diseases and chronic pathologies among the wounded.

J-P. Di Silvestro / ICRC

Figure 5.2

The number of wounded and dead by direct trauma may at times far exceed the effects on public health.

5.2.4 Methodology

Public health monitoring and data collection are notoriously difficult during the chaotic and overwhelming conditions of complex emergencies. Missing and displaced persons, constraints of time, lack of access to populations, and dangerous security situations, all create formidable problems for a limited number of qualified personnel to conduct proper studies. Furthermore, the distinction between civilian and military during a civil war is not always obvious.

The ratio of military to civilian wounded and sick may be very sensitive political and military information, easily exploited for propaganda purposes by the various antagonists. In an attempt to prevent this, should ICRC delegates determine that civilians are being expressly targeted by combatants, the organization undertakes specific *démarches* regarding the conduct of hostilities with the authorities concerned through its traditional confidential procedures.

Many authors have recently written about the public health effects of armed conflict and complex emergencies. The ICRC first published its standard *War and Public Health* in 1996 and, together with the World Health Organization and a number of universities around the world, it organizes a dozen H.E.L.P.[4] courses on the management of humanitarian aid every year.

5.3 Epidemiology for the war surgeon

What should one look for in a study of epidemiology? What information will help the individual surgeon managing war wounds for the first time? What needs to be taken into consideration to set up an efficient system for the care of the war-wounded? Is it possible to determine the "normal" levels of morbidity and mortality during armed conflict?

To respond adequately to the challenge requires preparation, on the part of the individual surgeon as well as the institution's, and a system for monitoring the results of patient management.

5.3.1 Preparation of the surgeon

The surgeon facing war-wounded casualties for the first time will ask a number of questions: "What wounds should I expect? Which injuries kill? What will the surgical workload be like?"

In many conflicts, civilian surgeons, albeit without previous military experience, become heavily involved in the management of the war-wounded. Yet, epidemiological studies clearly show that combat injuries differ from those in civilian practice: different aetiology and pathology; multiple-cause injuries; delayed access to care and austere working conditions requiring different philosophies of treatment.

> Epidemiological studies show that the management of the war-wounded is different from civilian trauma. The war surgeon must develop a new clinical mindset.

Various indices have been devised to define injury patterns, which differ according to the type of conflict and the nature of the weapons used: the lethality of wounding agents; the anatomic distribution of wounds; delay to treatment; and pre-hospital and post-operative mortality are among the most important. Knowledge of these helps determine the high-risk factors affecting the outcome of surgical treatment and what set of specialist skills are needed: orthopaedic surgery, vascular, visceral, etc.

4 H.E.L.P.: Health Emergencies in Large Populations. The reader is directed to the publications in the Selected bibliography for further study.

These patterns also have an impact on hospital activities and surgical workload and influence the standardization of protocols for patient management.

5.3.2 Preparation of the institution

The institution involved – military sanitary services, public health ministry, ICRC or any other humanitarian agency – must also be prepared.

The standardization of protocols and knowledge of the workload permits the establishment of ready-made kits of equipment, medicines and supplies that are easily deployed in an emergency situation as a rapid response to well-known needs. This facilitates the setting up of a surgical hospital to treat the victims of war.[5]

Preparation also implies choosing the correct profile of surgeon required, and training him or her to understand the context, the pathology, and the functioning of the institution in which they will be working, as well as the prevailing clinical protocols.

Knowing what kind of patient comes through the hospital door helps in planning a response: allocation of medical resources and personnel and the training of surgeons.

5.3.3 Surgical audit: monitoring

Monitoring of the outcome of patient management in a context of of armed conflict involves a surgical audit using an epidemiological approach: demographics, types of wounds, anatomic distribution, mechanism of injury, time since injury, post-operative mortality, number of operations and blood transfusions per patient, and morbidity, etc. There are several reasons for this.

- The adequacy of surgical expertise is important: the "fog of war" and the confusion and stress of a situation of armed conflict are not an excuse for bad surgery. Professional standards must be maintained, even when resources are limited.
- The adequacy of the clinical protocols in use must be put to the test: changes must be made in response to poor or unexpected outcomes. Many surgeons will be seeing war pathology for the first time, under conditions that do not resemble their routine work circumstances. At times, a strict and rigid application of protocols is required; on other occasions, there will be a need for improvisation and adaptation.
- The adequacy of the chain of casualty care must also be put to the test: hospital as well as pre-hospital data helps to assess the efficiency of first-aid measures in the field and that of the evacuation system. Surgeons, civilian or military, may be involved in training or helping set up pre-hospital programmes so that the injured reach them in better condition and in a timely fashion, which makes their surgery easier and ensures a better outcome for the patient.
- The role of surgeons in the development and application of international humanitarian law.

5 The *Emergency Items Catalogue of the International Movement of the Red Cross and Red Crescent* has a listing of such kits for the establishment of a surgical field hospital and for the treatment of 100 war-wounded hospitalized and operated on.

5.3.4 The surgeon and IHL

Surgeons working for the ICRC believe that health professionals have a duty to gather data "to prevent and not just to treat" with regard to the law and norms of behaviour in the conduct of hostilities. They are not alone in this belief.

Health professionals were instrumental in gathering the epidemiological data from clinical studies to make the humanitarian argument for the banning of blinding laser weapons and anti-personnel landmines.[6]

In the field, ICRC delegations in areas of armed conflict monitor the compliance of belligerents with IHL. This is an essential element of the mandate of the ICRC to protect and assist the victims of armed conflict and promote respect for the law of war. Epidemiological studies, in both hospital and pre-hospital settings, can help in determining certain abuses of IHL. For example, mortality statistics can help elucidate the occurrence of executions of prisoners of war being passed off as killed in combat. The ICRC then undertakes a series of confidential *démarches* with the offending party in order to promote observance of humanitarian norms.

Other actors may do likewise, in compliance with their humanitarian responsibilities.[7] The ICRC warns, nonetheless, against the possibility of exploitation and manipulation of statistical epidemiological studies for political purposes.

5.3.5 The surgeon and the specialized literature

"Statistics can be made to say whatever you want them to say" is a well-worn cliché. It is therefore important for the surgeon reading the vast literature on war surgery to understand what is being said, what the shortcomings are, and what some of the traps and pitfalls are.

Most of the war surgery literature consists of large retrospective studies written by military surgeons of industrialized countries; some of the exceptions are included in the Selected bibliography at the end of this volume. Most surgeons are not well trained in epidemiology or statistical methods, and civilian surgeons are usually not knowledgeable about military terminology. Thus, the surgeon facing the new experience of dealing with the war-wounded who wishes to read up on the topic will often find a bewildering vocabulary and methodology.

Differences

Added to the differences between war and civilian trauma are the disparities between the experience of ICRC interventions (and other humanitarian agencies) and ministry of public health structures on the one hand, and classical military medical services on the other. The rest of this chapter deals with some of these differences and disparities.

6 Respectively, Protocol on Blinding Laser Weapons 1995 (Protocol IV to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects,1980), and the Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction, 1997, known as the Ottawa Convention.

7 Burnham G, Lafta R, Doocy S, Roberts L. Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey. *Lancet* 2006; **368**: 1421 – 1429.; Dudley HAF, Knight RJ, McNeur JC, Rosengarten DS. Civilian battle casualties in South Vietnam. *Br J Surg* 1968; **55**: 332 – 340.

5.4 General questions of methodology

There are a number of problems and difficulties relating to how epidemiological studies have been conducted. The surgeon reading the literature should be aware of them.

5.4.1 Large and small numbers

First it must be clear which phenomenon is being studied. Comparisons of data between a single battle and longer wars may not be valid; a small cohort may not be representative. Most clinical protocols take as a basis large numbers in big wars. On the other hard, the individual surgeon will face individual casualties from particular battles. Thus, the experience with small cohorts may help prepare the surgeon to face a novel situation or specific injuries.

5.4.2 Problems of data collection

Trauma registries are notoriously difficult to maintain during war. Fatigue and lack of time, a shortage of trained staff, and the personal danger that medical staff must often confront complicate the keeping of statistics. The administrative personnel and time and effort required to keep up full documentation and archives can appear to be a luxury when faced with the stress of caring for battle casualties. In an article on American casualties during the war in Viet Nam, based on data obtained from a statistical record filled out by physicians in Viet Nam at the time of discharge or transfer, the author made the following telling comment.

> " Many of the physicians who filled out these forms recognized that they were frequently completed under stress, by command, and without enthusiasm, by physicians who considered it another thankless task, and therefore the accuracy of some of the data may be questionable…"
>
> R. M. Hardaway III [8]

Besides the problems of data collection mentioned with respect to public health statistics and the displacement of populations, other discrepancies and difficulties appear when dealing with a retrospective analysis of hospital-based data.

- The quantity and quality of data from different hospitals are irregular.
- Initial admission data, operating theatre observations, and ward medical records for individual patients do not always match.
- Hospital records are often incomplete, or contain clerical errors.
- The management of casualties at different points in a military chain of evacuation compounds reporting problems, particularly with respect to patient outcome.
- Access to medical care for the wounded – especially civilians or combatants in a civil war or during irregular guerrilla warfare – is not always available or must be done "in secret".
- Patients may disappear from the hospital before completing their treatment because of personal security concerns.
- Families do not always bring their dead for registration.

As mentioned before, one must also always consider the political and military sensitivity of trauma information; ICRC health personnel have on occasion been accused of being "spies" because they were asking a hospital director for admission statistics.

8 Hardaway RM III. Viet Nam Wound Analysis. *J Trauma* 1978; **18**: 635 – 643.

5.4.3 Important first questions: who is counting whom?

Foremost is the very important problem of different definitions of epidemiological categories. Some military definitions have changed over the last century or been replaced by other terms. Articles written by civilian authors have often invented their own epidemiological categories and definitions. Who exactly composes the target population under study is different from one article to the next, although all deal with the "wounded" and during the same war. Large retrospective studies are fraught with such traps.

There are important first questions the reader must ask of all war surgery studies: who is counting whom and where? When is a wounded person counted as a wounded person? Which patients are included in the nominator and denominator in any formula or equation?

Important first questions:

Who is counting whom?
When is a wound a wound?
Who are the nominator and denominator?

Who is counting?
Different hospitals – military, public, and other (mission hospitals, non-governmental organizations, ICRC) – all have their own objectives, mandates and routines. They all collect statistics; but often for very different reasons, and very different statistics (see Annex 5. A: ICRC Surgical Database).

A well-organized classical army can tell you how many of their soldiers were killed or disabled; pensions and allowances must be paid and the necessary administrative structures to do so exist. Guerrilla forces and civilian medical facilities are incapable of fully recording such statistics, with rare exceptions.

Who is being counted and where?
Are casualties calculated at the point of wounding, at first-aid posts, first echelon or referral hospitals? The total number of wounded will depend on which level of the evacuation chain is counting, and what happens to them after injury.

The wounded that arrive at hospitals – where most studies are made – are only a sample of the casualties; they do not represent the total reality of war.

When is a wounded person a wounded person?
Many studies in the military literature, even when dealing with the same war, define the wounded differently. One example: for World War II, the United States Surgeon General's Statistical Health Report indicated 724,000 wounded and 228,000 battle-related deaths among American troops; a fatality rate of 23.9 %. The Adjutant General's Report gave the figure of 593,000 battle injuries and 235,000 deaths; a rate of 28.4 %. The latter report excluded the lightly wounded.[9] Retrospective analyses do not always indicate which figure is being used.

Studies do not always make clear if all superficial and light injuries are excluded or not from the calculations. To state the obvious: a scalp laceration is not the same as an open wound to the brain. The reader should note that non-penetrating head, chest, and abdominal injuries are not always clearly differentiated from penetrating ones. The use of the Red Cross Wound Score permits this distinction (see Chapter 4).

Similarly, studies of the anatomic distribution of wounds, the wounding agent, and delay in evacuation do not always state specifically whether the authors are dealing with the survivors only, or with total combat casualties comprising the dead as well as the wounded survivors.

9 Carey ME. Learning from traditional combat mortality and morbidity data used in the evaluation of combat medical care. *Mil Med* 1987; **152**: 6 –12.

5.4.4 Some definitions: the wounded and the dead

Table 5.2 includes some basic epidemiological definitions intended to standardize reporting by the armed forces of the United States of America.

Killed in action (KIA)	Those dying from battle injuries before reaching a medical facility. Comparable to civilian dead on arrival (DOA).
Wounded in action (WIA)	Those wounded who survive to reach a medical facility staffed by a physician capable of trauma life support, whether pre-hospital or hospital.
Died of wounds (DOW)	WIA but dying from wounds later on.
Returned to duty (RTD) or Carded for Record Only (CRO)	Wounded but not hospitalized and returned to active duty within 72 hours. May include patients who are registered in a pre-hospital setting: carded for record only but not evacuated.
Hospital mortality	Died post-operatively in a surgical facility.

Table 5.2 Standardized epidemiological definitions as used by USA armed forces.[10]

ICRC EXPERIENCE

In Kisangani, a city of 600,000 inhabitants in the remote east of the Democratic Republic of the Congo, major fighting broke out in June 2000. Hostilities lasted seven days. One week later, visiting ICRC delegates recorded that 4 hospitals and 62 clinics had registered 1,691 wounded; 369 were still hospitalized (over 90 % civilians, most military casualties had been evacuated). The remaining casualties were under ambulatory care. It was impossible to determine the number killed. Almost a month later, a total of 2,393 casualties had been registered and 248 were still hospitalized. The vast majority of these patients were so superficially wounded that, had they been living in a rural environment far from a health facility, they would never have presented. These are the civilian equivalent of RTD and CRO.

In nine and half months in Beirut, in 1976, an ICRC field hospital treated 1,289 war-wounded on an outpatient basis and another 696 in hospital. Similarly, three months of urban warfare in Monrovia, Liberia in 2003 saw 2,588 patients triaged by an ICRC surgical team, but only 1,015 were admitted to hospital.

These examples tell us a great deal about the sociology of medical care during armed conflict and the effects on statistical results. In urban warfare, all civilian city-dwellers who are injured, even with very minor wounds, will seek medical treatment and be registered as war-wounded. Everybody goes to the hospital! Their reasons for presenting at a hospital or clinic are many: the medical sophistication of an urban population used to seeking curative care; the need for reassurance and being in a "safe" haven; hopes of financial compensation for all those injured; and the search for simple psychological support for a traumatized civilian population.

Although the presence in hospitals and clinics of so many superficially wounded patients may not represent the true extent of the surgical workload, the burden on the emergency department (triage and first aid), nursing care (wound dressings), and non-clinical work (administration, registration, laundry and kitchen) can be overwhelming. Statistical analysis is both difficult and tedious under such circumstances (e.g. Monrovia) or almost impossible (e.g. Kisangani).

10 Holcomb JB, Stansbury LG, Champion HR, Wade C, Bellamy RF. Understanding combat casualty care statistics. *J Trauma* 2006; **60**: 397 – 401.

5.4.5 Clinical and operational consequences and IHL

A number of clinical and operational elements are significant as regards IHL.

1. For armies, the total number of casualties – soldiers removed from active participation in combat or support – and the overall lethality of the battlefield as represented by the case fatality rate (CFR) are important in terms of medical planning and allocation of fighting resources.

   For civilian health structures, the total number of casualties is only one measure of the social and economic burden of war and the humanitarian consequences of the disorganization of society.

2. Killed in action (KIA) helps to determine the lethality of weapons and the type of combat. IHL concerns itself with the lethality of weapons systems and the conduct of hostilities.

3. The KIA and died of wounds (DOW) help determine the efficiency of the evacuation chain, which is dependent on access to the wounded, appropriate pre-hospital care (field triage and initial treatment), and transport logistics (evacuation time and access to a hospital for the wounded). Access of health personnel to the wounded and access of the latter to medical facilities are supposed to be guaranteed, in accordance with the protection conferred upon these persons by IHL.

4. The adequacy of medical treatment depends on the efficiency of the evacuation system, the wounding mechanism and pathology (burns, blast, penetrating wounds, anti-personnel mines, etc.), and hospital competency. DOW and hospital mortality are the main indicators. Again, IHL is interested in the medical consequences of weapons systems.

5. These various indices of mortality are important. However, they say nothing about the severity of injury in survivors – morbidity and disability – or surgical workload that may have a greater long-term effect on civilian populations and health structures. The number of operations per patient, infection rates, amputation rates, and length of hospitalization better define the severity of injury and hospital burden of war-wounded. The humanitarian consequences of such effects are of direct concern to IHL.

5.5 Aetiology of injury

5.5.1 Warfare scenarios

The type of warfare greatly influences the kinds of wounds the surgeon sees and the anatomic distribution of injuries, which obviously affects hospital workload. The widespread use of anti-personnel blast mines in guerrilla and counter-insurgency warfare results in many traumatic amputations of the legs. Close combat with assault rifles causes many gunshot wounds with severe single injuries; shelling and bombing from a distance produce multiple fragment wounds, many of which are superficial. Major contemporary wars have seen a shift from gunshot wounds to fragment injuries in the majority of casualties.

In intercommunal strife, combatants in some societies engage in "traditional" warfare using machetes and *pangas*; slash wounds to the head, neck and shoulders predominate and, secondarily, injuries to the forearms when the victim attempts to protect him or herself. Elsewhere, the ready availability of small arms means that "traditional" combat and cattle raiding are now undertaken with the use of AK-47 Kalashnikov rifles.

Weapons systems available to belligerents vary. Classical armies of industrialized countries use air bombardment as well as artillery and armour, and their infantry is often fitted out with personal protective equipment. Classical armies of low-income

countries tend to rely on infantry and artillery. Rural guerrilla formations depend on ambush with personal firearms and landmines; urban guerrilla warfare is largely street fighting with rifles, rocket-propelled grenades and, sometimes, short-range mortars. The weapons systems available to combatants can also change during the course of a war.

The intensity of fighting, being on the offensive or defensive, losing or winning a given battle, the size of the unit actually engaged, and the percentage of soldiers really exposed to combat, all affect various combat casualty statistics.

Relevant clinical consequences

The numbers and relative proportion of civilian and military wounded and dead depend to a large extent on combat strategy and operational tactics, geographical site of fighting, access to medical care, and who is winning or losing.

5.5.2 Definitions of weapon-systems: methodology

Some confusion arises in surgical literature because of differing categories and definitions for wounding agents and mechanism of injury: there is no universally accepted standard.

"Landmines" include both anti-tank (ATM) and anti-personnel (APM); the two are usually not differentiated. ATM may be the agent of wounding, but the mechanism of injury can be blast, blunt, burn, or penetrating fragment. APM may be either a blast or a fragmentation device. The severity of the injury differs with the mechanism and aetiology, but this cannot always be determined from the categories in a study.

The category of "fragment wounds" often groups together injuries from shells, bombs, and grenades, together with APM. The differences are important in terms of surgical management, however. Small, superficial wounds from preformed grenade fragments often do not require surgery; this is never the case for APM fragments (see Chapter 10). Table 5.3 gives an indication of the distribution of wounds according to projectile, from a number of conflicts.

	Bullets %	Fragments %	Other %
World War I (Western Allies)	39	61	
World War II (Western Allies)	10	85	5
Korea (USA)	7	92	1
Indochina (France)	62	38	
Algeria (France)	71	23	
Borneo (UK)	90	9	1
Lebanon: Beirut 1975 – 86	49	36	14
Northern Ireland	55	22	20
Falkland Islands / Malvinas (UK)	32	56	12
Thailand 1981	38	20	42 (APM)
Lebanon 1982 (Israel)	12	53	35
Lebanon: Bourj el-Barajneh refugee camp 1986 – 87	20	60	20
Eritrea 1988 – 91	33	63	2.2 (APM)
Bosnia-Herzegovina: Sarajevo 1992 – 96 (France)	59	37	4 (APM)
Bosnia-Herzegovina 1993 (Bosnia-Herzegovina)	49	46	5
Croatia 1991 – 93 (Croatia)	25	70	6 (APM)
Yugoslavia 1991 – 92 (Yugoslavia)	41	2	52 (APM)
Somalia: Mogadishu 1992 (US helicopter)	55	31	14

Table 5.3 Aetiological distribution of the war-wounded: generally accepted historical examples. Anti-personnel landmines included in some. The country name in brackets indicates the source of information, see Selected bibliography.

5.5.3 ICRC statistics

ICRC surgical teams have worked in many different wars and combat scenarios. Table 5.4 shows the mechanism of injury in different types of warfare.

Hospital	N =	Fragments %	Bullets %	Mines %
Butare	40	8	92	–
Kabul	6,244	52	29	19
Kao-i-Dang	1,067	22	16	63
Kandahar	1,159	24	50	26
Novye Atagi	186	44	35	22
Peshawar	4,340	42	23	35
Quetta	6,570	33	39	28
Lokichokio	12,196	10	87	2
Monrovia JFK Memorial	867	38	62	–

Table 5.4 Aetiology of war-wounded in a number of ICRC hospitals.

Extremes of the distribution of wounding mechanisms are to be found in these very different conflicts, and reflected in the statistics held in the various hospitals (Figure 5.3). Southern Sudan was the scene of a rural guerrilla war in semi-desert, sahelian landscapes, and reed swamps (Lokichokio hospital in northern Kenya) where most injuries were gunshot wounds (GSW). Fighting in Monrovia, Liberia, was irregular urban warfare (JFK Memorial Hospital). Patients coming to Kao-i-Dang hospital (Thailand) had been injured in rural guerrilla warfare waged in the forests and jungles of Cambodia with widespread use of anti-personnel landmines. Needless to say, the types of pathology encountered in these three hospitals differed widely.

Figure 5.3
Different combat tactics produce very different distributions of wounding agents.

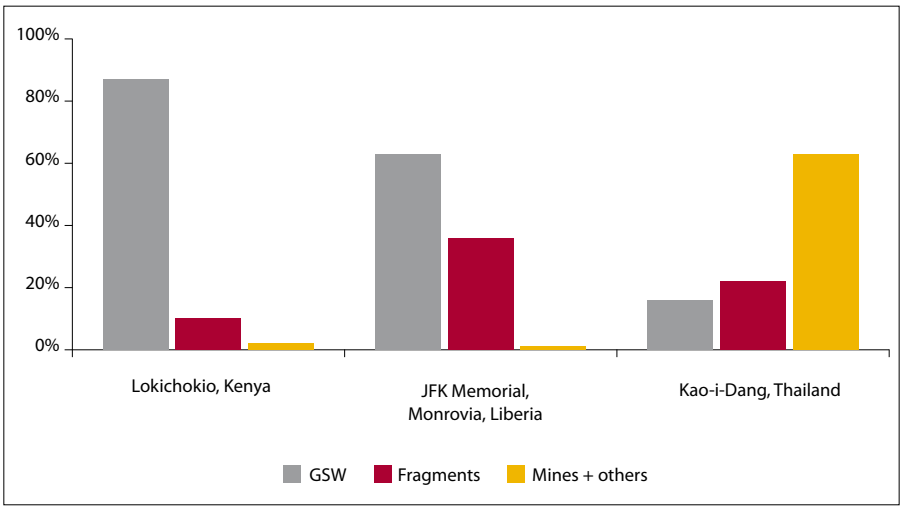

5.6 Anatomic distribution of wounds

The aetiology of wounds and their anatomic distribution change constantly depending on the development of new weapons systems, personal protective material (body armour), and the nature of combat. One constant, however, is the preponderance of wounds to the limbs amongst survivors. This distribution is important in determining surgical workload.

5.6.1 Body exposure

Body exposure to injury differs according to the type of weapon:

- anti-personnel blast mines strike the legs (Pattern 1 APM injury);
- manipulation of mines injures the upper limbs, face and chest (Pattern 3 APM injury) as described in Chapter 3;
- fragmentation devices spray the whole body with multiple hits in a random fashion; and, secondarily, according to the type of combat and military activity:

- aimed sniper fire targets the head or chest;
- trench warfare particularly exposes the head, resulting in the many mutilating facial wounds of World War I, known as "*gueule cassée*";
- purposeful observation and lookout activities of the soldier also increase the risk to the head and neck.

On the other hand, improved helmets and body armour, covering the chest and upper abdomen, effectively modify the relative exposure of anatomic regions.

Apart from aimed sniper fire, projectiles cause injuries in a random manner. Traditionally, combat exposure to injury has been estimated using as a basis the percentages of body surface area applied to assess burns, but taking into account operational necessities for soldiers (Table 5.5). It should be noted that combat exposure of the head and limbs is over-represented when compared to burns.

	Head	Neck	Thorax	Abdomen	Limbs
Body exposure in combat %	12		16	11	61
Surface area (burns) %	9	1	18	18	54

Table 5.5 Percentage body surface area versus surface area exposed in combat.[11]

5.6.2 Definitions and methodology

The anatomic distribution of wounds has been remarkably consistent over the last century. Reporting, however, has been remarkably inconsistent. Some studies cite only the primary site of injury if there is more than one; others include the category "multiple". Some count the "wounds" and not the "wounded", so that sites of injury exceed the number of patients. Many studies do not clearly define the counting method; in some reports only survivors are counted, while for others total combat casualties – KIA and WIA – are included. Again, the methodology is not always specified.

Anatomic regions are not standardized either. Different studies use different definitions for anatomic regions; there is again no universal standard. Some report only the "torso"; "pelvis and buttocks" are under a separate category from "abdomen" in some, and included in others. To be truly accurate, head, face, and neck categories should be clearly separated, which they rarely are. These latter wounds represent very different clinical problems – traumatic brain injury, asphyxia, haemorrhage – and degrees of lethality.

Anatomic distribution furthermore only indicates a potential threat; crude statistics often include superficial wounds, which is not a true indication of hospital workload and threat to life and limb. As has been mentioned above, it would be preferable to separate out penetrating from non-penetrating wounds of vital regions: head, thorax, and abdomen. The Red Cross Wound Score and Classification System attempts to address some of these problems in a simplified manner.

11 Carey ME, 1987.

5.6.3 Historical results

The reader going through the literature will find very different figures for the same war, depending on the source and the methodology. This can make for frustrating reading. Nonetheless, some historical approximations are presented in Table 5.6. Wounds to the limbs predominate, from 50 % to 79 %.

Conflict	Head & Neck %	Thorax %	Abdomen %	Limbs %	Other & multiple %
World War I (Western Allies)	17	4	2	70	7
World War II (Western Allies)	4	8	4	79	9
World War II (USSR)	19	9	5	67	–
Korea (USA)	17	7	7	67	2
Viet Nam (USA)	14	7	5	74	–
Borneo (UK)	12	12	20	56	–
Northern Ireland	20	15	15	50	–
Arab-Israeli war 1973 (Israel)	13	5	7	40	31
Thailand 1981	10	12	4	66	8
Falkland Islands / Malvinas (UK)	16	15	10	59	–
Lebanon 1982 (Israel)	14	5	5	41	34
Lebanon: Bourj el-Barajneh refugee camp 1986 – 87	12	16	18	54	–
Gulf War 1991 (UK)	6	12	11	71	(32)*
Gulf War 1991 (USA)	11	8	7	56	(18)**
Afghanistan (USSR)	16	12	11	61	–
Chechnya 1995 (Russia)	24	9	4	63	–
Somalia: Mogadishu 1992 (USA)	20	8	5	65	(2)**
Croatia 1991 – 93 (Croatia)	15	11	4	69	1
Yugoslavia 1991 – 92 (Yugoslavia)	21	9	8	62	(23)**
Croatia 1991 (Yugoslavia)	12	15	8	65	–
Bosnia-Herzegovina 1992 (Yugoslavia)	14	15	9	62	–
Bosnia-Herzegovina 1993 (BH)	19	16	11	53	–
Bosnia-Herzegovina: Sarajevo 1992 – 96 (France)	11	11	14	61	3
Bosnia-Herzegovina: Sarajevo market bombing 1995	13	13	12	62	–
Eritrea 1988 – 91	20	9	6	63	2
ICRC surgical database	12.5	7.2	7.8	65.5	7
Overall average	~15	~10	~7	~65	

* Buttock and back wounds – all multiple fragment injuries – as a separate figure.

** Multiple wounds.

Table 5.6 Anatomic distribution of major wounds; some statistics include both dead and survivors, others include minor injuries. The country name in brackets indicates the source of information, see Selected bibliography.

5.6.4 ICRC results

The experience of ICRC surgical teams in various conflicts closely mirrors that of surgeons in other wars, particularly in relation to the type of conflict (Table 5.7). Many patients have multiple wounds; the distribution is given for injuries and not for patients, hence more wounds than patients. Only one wound is counted per anatomic region, however.

Pelvis and buttocks, back and soft tissue are given separately. Head, face and neck are not differentiated further.

Hospital	Number	Head & Neck %	Thorax %	Abdomen %	Pelvis & Buttock %	Back & soft tissue %	Upper limb %	Lower limb %	Limbs combined %
Kabul	8,804	15	9	10	4	3	24	35	59
Kao-i-Dang	1,660	15	8	7	4	3	24	39	63
Peshawar	6,840	18	8	6	5	3	25	35	60
Kandahar	1,396	11	9	11	3	2	24	40	64
Quetta	9,373	15	9	8	5	3	24	36	60
Butare	45	16	7	2	–	2	31	42	73
Novye Atagi	210	10	3	7	2	3	26	50	76
Lokichokio	14,203	7	8	3	7	3	29	44	73
Monrovia	904	14	13	4	4	–	21	43	64

Table 5.7 Anatomic distribution of wounds in various ICRC hospitals serving victims of differing types of combat.

5.6.5 Primary tissue injury

The anatomic distribution of wounds and the analysis of primary tissue injury permit the assessment of the surgical workload. In both, wounds to the limbs predominate; soft-tissue wounds and fractures of long bones constitute the majority of the injuries in all published series. The WDMET[12] statistics from Viet Nam are eloquent (Table 5.8); as are the figures from a French military field hospital in Sarajevo, 1992 – 1996, (Table 5.9). These analyses take into account non-penetrating wounds to vital regions.

Soft tissue (especially of limbs)	47 %
Extremities (long bone fractures)	26 %
Abdomen	8 %
Chest	4 %
Neck	2 %
Face	6 %
Head	2 %
Multiple	5 %

Table 5.8 Distribution of wounds according to primary tissue injury (USA).[13]

Soft tissue (especially of limbs)	56 %
Extremities (long bone fractures)	22 %
Abdomen	14.5 %
Chest	11 %
Face & Neck	6 %
Head	6 %
Peripheral vessels	6 %
Multiple	6 %

Table 5.9 Distribution of wounds according to primary tissue injury (French military field hospital in Sarajevo).[14]

ICRC results

In the hospitals included in the ICRC database, where evacuation was difficult for numerous patients and many with minor soft-tissue wounds did not bother to come to hospital, soft-tissue injuries still represented 36 % of wounds; fractures of the extremities 46 %; and vital central wounds 20 %.

Various other ICRC reports confirm these findings (Tables 5.10 and 5.11).

Soft tissue	33 %
Bones and joints	33 %
Vascular	11 %
Peripheral nerves	11 %
Abdomen hollow organs	17 %
Abdomen solid organs	9 %

Urogenital	5 %
Chest	9 %
Brain and spinal cord	3 %
Maxillo-facial	3 %
Eye, internal ear	2 %
Other	1 %

Table 5.10 Distribution of wounds according to tissue injury, hospitalized patients in the ICRC field hospital, Beirut 1976 (N = 696).[15]

12 In the military literature, the Wound Data and Munitions Effectiveness Team (WDMET) of the United States set a gold standard for data collection. A large number of administrative staff meticulously collected information on 7,989 patients during the war in Viet Nam from 1967 to 1969 in a systematic way. The analysis was published by: Bellamy RF. Combat trauma overview. In: Sajtchuk R, Grande CM, eds. *Textbook of Military Medicine, Anesthesia and Perioperative Care of the Combat Casualty.* Falls Church, VA: Office of the Surgeon General, United States Army; 1995: 1 – 42. This report is often cited in review articles.

13 Champion HR, Bellamy RF, Roberts P, and Leppäniemi A. A profile of combat injury. *J Trauma* 2003; **54 (Suppl.)**: S13 – S19.

14 Versier G, Le Marec C, Rouffi J. Quatre ans de chirurgie de guerre au GMC de Sarajevo (juillet 1992 à août 1996) *Médecine et armées* 1998; **26**: 213 – 218.

15 Translated from: Kjaergaard J. Les blessés de guerre de l'hôpital de campagne du CICR à Beyrouth en 1976. *Schweiz Z Milit Med* 1978; **55**: 1 – 23.

Parts injured	Total % (N = 1,033)	Bullet wounds % (n = 231)	Shell fragment wounds % (n = 508)	Mine injuries % (n = 294)
Soft tissue	73	67	75	70
Bone	39	52	20	63
Thorax	7	7.5	9	4
Abdomen	11	10.5	14	7.5
Brain	2.5	–	5	1
Other	4	2.5	4	4

Table 5.11 Tissue injuries according to wounding weapon, Kao-i-Dang ICRC hospital 1984 – 85. [16]

The clinically important point for the surgeon is the very large hospital workload represented by soft-tissue and orthopaedic injuries.

5.7 Fatal injuries

5.7.1 Site

It is not a simple matter to determine cause of death and site of lethal injury. Multiple injuries tend to have a synergetic effect and it can be impossible to define which of several wounds was the direct cause of death. Furthermore, many fatal war injuries involve total body disruption or severe mutilation.

A formal complete autopsy for every combat death is onerous even for the armed forces of a wealthy industrialized country, and this has rarely been common practice. Three examples of the anatomic distribution of fatal wounds are given in Table 5.12.

	US dead in World War II[17]	US dead in Viet Nam (WDMET)[18]		Israeli dead in Lebanon 1992[19]	
Head	42 %	37 %	46 %	9 %	34 %
Neck		6 %		3 %	
Face		3 %		22 %	
Chest	30 %	24 %		45 %	
Abdomen	12 %	9 %			
Multiple	–	17 %		–	
Limbs	13 %	3 %		21 %	
Soft tissue	–	1 %		–	

Table 5.12 Anatomic distribution of fatal wounds.

Central wounds predominate, as would be expected; particularly to the head, face, neck and thorax.

The great majority of war wounds are to the limbs.
The most lethal wounds are to the head and chest.

16 Trouwborst A, Weber BK, Dufour D. Medical statistics of battlefield casualties. *Injury* 1987; **18**: 96 – 99.

17 Garfield RM, Neugut AI. Epidemiologic analysis of warfare. *JAMA* 1991; **266**: 688 – 692.

18 Champion HR, et al., 2003.

19 Gofrit ON, Kovalski N, Leibovici D, Shemer J, O'Hana A, Shapira SC. Accurate anatomical location of war injuries: analysis of the Lebanon war fatal casualties and the proposition of new principles for the design of military personal armour system. *Injury* 1996; **27**: 577 – 581.

5.7.2 Trimodal distribution of trauma mortality

Mortality in civilian trauma was classically described by D. Trunkey in 1983[20] to fall into three categories: immediate death (50 %), early death (30 %), and late death (20 %), see Figure 5.4.

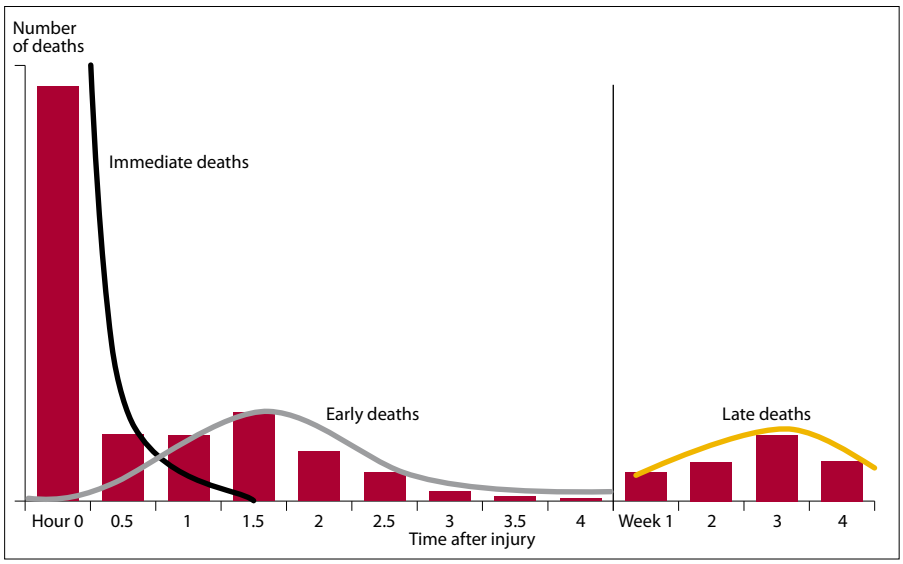

Figure 5.4
Trimodal distribution of trauma deaths.

Peak 1: immediate deaths
Most deaths occur at the moment of injury or within minutes. These are due to overwhelming trauma incompatible with life (severe brain injury, massive haemorrhage).

In war, in addition to severe injuries to the head and torso (heart, liver, and major blood vessels), some casualties suffer total body disruption or incineration through burns. It is estimated that 70 % of deaths occur within five minutes, and little to nothing can be done for these patients, which represent about 17 – 20 % of the severely wounded.

Peak 2: early deaths
These occur after the first minutes until the first few hours after injury. This peak gave rise to the idea of the "golden hour"; if certain measures were implemented rapidly, the patient could be saved.

In war trauma, there are three major causes of early death:
- continuing exsanguinating haemorrhage;
- inadequate/impaired airway due to survivable penetrating head injury;
- compromised breathing due to tension pneumothorax.

Many of these early deaths are avoidable within this "golden hour" if adequate first-aid measures in the field are implemented in time.

Peak 3: late deaths
These occur a few days to weeks after injury and are the result of subsequent complications of infection, multiple organ failure and coagulopathy, and uncontrollable increased intra-cranial pressure due to post-injury cerebral oedema (closed injury).

Good and early first-aid measures can decrease infectious and other complications in the setting of armed conflict where wounds are dirty and contaminated from the very beginning. Inadequate early care affects morbidity (sepsis, disability) as well as mortality.

The understanding of this trimodal distribution of deaths stimulated an improvement in emergency medical services and patient transport in civilian settings. Rapid evacuation times and early advanced life support in many industrialized countries, in a bid to reach the injured during the "golden hour", have reduced these three peaks to a bimodal profile: immediate deaths and late.

20 Trunkey DD. Trauma. *Sci Am* 1983; **249**: 220 – 227.

Applicability to the military war-wounded

The military have found that the trimodal distribution is relevant to armed conflict. They have tended to underline three categories of patients corresponding to what can be done under various operational scenarios in the field. This has an important influence on field triage categories.

1. Unsalvageable injuries: i.e. KIA for whom nothing can be done (17 – 20 %).
2. Severe but potentially survivable injury (10 – 15 %).
3. Moderate to minor injuries (65 – 70 %).

One result of improved pre-hospital care and evacuation is that some casualties become DOW (early deaths) instead of being KIA (immediate deaths) while a few others are saved. The most important field measures undertaken include stopping haemorrhage from controllable bleeding and maintaining a patent airway and breathing by simple procedures.

ICRC EXPERIENCE

Not all penetrating head injuries are so severe as to be incompatible with life. However, precarious pre-hospital care and long and difficult evacuation routes can result in large numbers of survivors of head injuries (WIA) succumbing later (DOW) because of inadequate control of the airway, resulting in asphyxia or vomiting and aspiration.

This was the case during one recent war in Africa involving a "classical" army. Large numbers of patients with salvageable head injuries were dying during a three-day evacuation in the back of lorries travelling over dirt roads in the bush. There was no possibility of monitoring endotracheal intubation.

An ICRC surgeon advised performing a tracheotomy in a front-line field hospital before evacuation, the only way to ensure an adequate airway under the circumstances in these comatose patients. The mortality rate for these cases was halved by this simple procedure.

Applicability to civilian war-wounded

At the shelling of the Markale market place in Sarajevo, August 28, 1995, only minutes away from two referral hospitals, 104 persons were injured of whom 42 ultimately died; a fatality rate of 40.8 %.[21]

Twenty-three died instantly and another ten were dead on arrival at hospital (79 % of deaths were first peak). Five patients succumbed during surgery (second peak 12 %) and another four a week later (third peak 10 %) as shown in Figure 5.5. The trimodal distribution of deaths appears to have been preserved.

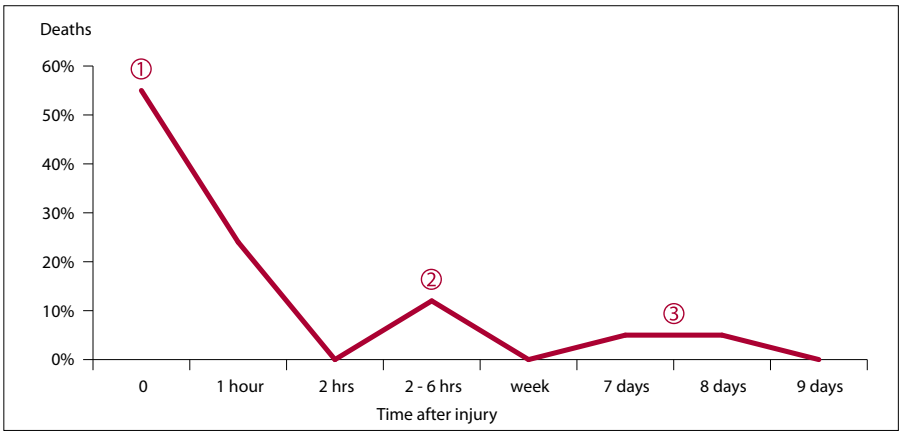

Figure 5.5

Trimodal distribution of deaths after shelling of Sarajevo's Markale market.

Furthermore, in remote rural areas with difficult evacuation routes, both armies and civilian populations will probably experience the trimodal distribution of deaths, as originally described. Trauma survival under such circumstances is geography specific.

21 Suljevic I, Surkovic I. Medical aspects of the mass-scale civilian casualties at Sarajevo Markale Market on August 28, 1995: triage, resuscitation, and treatment. *Croat Med* J 2002; **43**: 209 – 212.

5.7.3 Ratio of dead to survivors

Many authors have noted that the ratio of dead to survivors in modern conflicts tends to be about 1:4 over the long term. This corresponds to our threshold fatality rate of about 20 – 25 %.

$$\frac{\textbf{Dead = KIA plus DOW}}{\textbf{Survivors = WIA minus DOW}} = \textbf{1:4}$$

Many factors can distort the specific results under given circumstances, including:

- the inclusion of minor wounds in the calculations – the famous methodological problem;
- the tactical situation (note the 40 % fatality rate of the Markale market place bombing mentioned above; burn injuries in tank crews or on ships; a successful surprise ambush, etc.);
- the lethality of particular weapons systems (anti-personnel landmines, napalm, etc.);
- any delay in medical evacuation; and
- the execution of wounded prisoners in contravention of the laws of war.

Improvement in personal protective equipment (body armour), better pre-hospital care, and the early availability of more aggressive surgery and enhanced intensive care facilities has changed the ratio in recent conflicts. For US armed forces in Iraq and Afghanistan, the dead to survivor ratio has become 1:8.[22] Effective body armour has decreased fatal wounds to the chest and abdomen, but increased wounds to the head and neck in survivors. No other armed forces use such improved body armour as extensively and the US experience is not easily translated to others, but it sets a standard nonetheless.

5.7.4 Lethality of weapons

If the KIA equals approximately 20 % and the DOW 5 %, this represents the total lethality of weapons during combat on land. Sea and air warfare are different.

It has long been recognized that different weapons systems have different lethality. In general, and over a large number of casualties the following percentages appear.

- Military rifle bullets: 30 – 40 % lethality; or one death in every 3 – 4 wounded.
- Randomly-formed fragments: 20 % for shells and 10 % for grenades.
- Preformed fragments: 15 % for shells and 5 % for grenades.
- Blast injury has a fatality of about 22 %.

Tactical considerations for distinct individual battles can change these figures. A well-laid ambush with small arms can easily result in more than 40 % of a small patrol being killed.

The particular case of anti-personnel landmines

All studies indicate the heavy surgical and nursing burden of injuries due to anti-personnel landmines, especially blast mines. Specific types of APM – bounding mines that jump up one metre in the air before exploding – invariably kill their victim: lethality lies close to 100 %.

Public health studies and ICRC surveys in poor countries where APM have been widely used in rural areas with little or no organized evacuation system and limited surgical facilities (Mozambique, Somalia, Cambodia, Afghanistan and Angola) indicate a case fatality rate of well over 50 % for traumatic amputations due to APM.

It is difficult to exclude the non-weapon factors in these studies and mortality statistics. Suffice it to say that the lethality of weapons includes their actual use in field conditions, and their total socio-economic and humanitarian effects. This is important in terms of IHL and was one of the significant factors in determining countries to negotiate the 1997 Ottawa Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-personnel Mines and on their Destruction.

22 Holcomb JB, et al., 2006.

5.7.5 Conclusions of clinical relevance

From this analysis and further extrapolations a number of conclusions can be drawn.

1. Wounds to the head and torso are the most lethal; they represent most of the mortality. The great majority of survivors have injuries to the limbs; they represent the majority of surgical workload and morbidity.
2. Despite the fact that the head represents only 9 % of exposed body surface area, it accounts for a disproportionate number (25 %) of all battle casualties. The lethality of penetrating head wounds is approximately 75 % and they account for slightly less than 50 % of combat deaths (KIA + DOW), and about 8 % of survivors.
3. Head wounds kill either through devastating brain injury or through asphyxiation of the comatose patient who would otherwise survive the injury.
4. Death through closed head injury is relatively greater in the civilian setting than during armed conflict.
5. Uncontrolled exsanguinating haemorrhage accounts for another 50 % or so of deaths. Most of these (80 %) present central injuries to the thorax and/or abdomen, which require surgical facilities for control. The lethality of chest wounds is about 70 %.
6. The other 20 % of haemorrhage deaths are due to bleeding from peripheral, compressible blood vessels (half in the neck and half in the limbs). Thus, about 10 % of all deaths are due to haemorrhage from the limbs.
7. Death rates through haemorrhagic shock are higher in armed conflict than in the civilian trauma setting.
8. Little can be done medically for most immediate deaths and there is a threshold percentage of killed-in-action no matter how sophisticated the medical services are.

Conclusions for improvement of patient outcome

1. Some immediate and early deaths can be prevented by personal protective armour covering the torso.
2. Many early deaths can be prevented by simple measures to:
   - control limb haemorrhage;
   - relieve airway obstruction, especially in comatose patients after head injury;
   - relieve tension pneumothorax.
3. Some early deaths due to haemorrhage, particularly of the abdomen, can be prevented by rapid evacuation to a surgical facility.

5.8 The lethality of context: delay to treatment

5.8.1 Historical developments

Possibly one of the most important developments in combat casualty care over the last half-century has been improved pre-hospital care and the timely evacuation of the wounded to a surgical facility. Lasting days during much of World War I, evacuations took an average of 10.5 hours during World War II; the use of helicopters by American troops in Korea lowered this to 6.3 hours, and an average of 2.8 hours in Viet Nam. Transfer of Israeli wounded during the 1982 war in Lebanon took an average of 2.3 hours. The use of helicopters has radically changed evacuation and pre-hospital care for the armies of industrialized countries, but requires air superiority to be truly effective. It has also revolutionized equivalent civilian trauma systems. This resource is seldom available in a low-income country.

5.8.2 Forward projection of resources

Delayed evacuation may still result from operational contingencies, tactical situations, and difficult geography. Consequently, many armies have projected forward their surgical capacities close to zones of actual combat in an attempt to reach wounded soldiers earlier and avoid the mortality and morbidity related to any delay. The aim is primarily to reduce the number of "early" deaths.

Soviet deployment of "Special Surgical Teams" in forward locations in Afghanistan shortened time to reach surgery: 31 % of wounded were in surgery within 1 hour and another 39 % within 2 hours. Cumulatively, 92 % were in surgery within 6 hours with this system. This compares with an overall average during the war of 88 % operated within 12 hours, before and after the establishment, of the forward teams.

During fighting in Croatia in 1991, a Yugoslav mobile field hospital was located 5 – 10 km from the front: 61 % of wounded were evacuated within 30 minutes and another 22 % between 30 and 60 minutes of being injured.

US troops in Afghanistan and Iraq have also seconded "Forward Surgical Teams" and "Forward Resuscitative Surgery Suites" to the field; surgical care routinely begins within 1 – 4 hours following injury. Preliminary reports give a mean time of one and a half hours for evacuation of USA military wounded in Iraq to a forward surgical facility.

5.8.3 Urban warfare: hospitals on front lines

During urban warfare, fighting can take place literally in front of surgical facilities; more than once has a patient been wounded at the front door of a hospital. This was the case during much of the fighting in Beirut in the Lebanese civil war, often with evacuation times of only minutes. A French medical team in Sarajevo in 1992 – 96 (UNPROFOR – IFOR), serving military and civilian casualties, had evacuation times of 15 – 45 minutes.

Comparable times were encountered by ICRC hospital teams in Kabul in 1992 and Monrovia, Liberia, in 2003. ICRC and Somali Red Crescent surgical teams working in the Keysaney Hospital in North Mogadishu have witnessed similar short evacuation times from 1992 up until the present time of writing.

Although the vast majority of wounded patients arrive at the hospital within minutes under these circumstances, a few suffer considerable delay. Without an organized pre-hospital system and given an absolute absence of ambulances, civilians are often cut off from care facilities during street fighting and have to wait hours or even days for evacuation.

5.8.4 The paradox of early treatment: changing mortality ratios and rates

There is an apparent paradox of early evacuation and treatment: an increase in the absolute numbers who survive, but also an increase in the DOW and hospital mortality rates. Medics reach a number of former KIA in time; more severely injured casualties enter the evacuation chain earlier, but since natural automatic triage that weeds out the more severely injured does not come into play, a higher percentage of injured now die of their wounds after they have received care.

For US armed forces who died in World War II and Viet Nam, 88 % were KIA and 12 % DOW. In the conflicts in Iraq and in Afghanistan, the figures are 77 % and 23 %.[23]

23 Holcomb JB, et al., 2006.

This effect has also been observed in hospital mortality rates at ICRC facilities (Tables 5.13 and 5.14).

Time until arrival in hospital	Patients (n =)	Died	Mortality
< 6 hours	3,114	172	5.5 %
6 – 24 hours	3,588	141	3.9 %
24 – 72 hours	1,668	46	2.8 %
> 72 hours	2,430	55	2.3 %

Table 5.13 Mortality according to evacuation time to ICRC hospitals: period from January 1991 to July 1993.

Longer evacuation times allow for "natural triage" to come into play. The more seriously wounded die before they reach hospital when evacuation is delayed.

Time until arrival in hospital	Patients (n =)	Died	Mortality
< 6 hours	79	5	6.3 %
6 – 24 hours	704	21	3.0 %
24 – 72 hours	210	5	2.4 %
> 72 hours	134	2	1.5 %

Table 5.14 Mortality according to evacuation time: patients transferred from ICRC first-aid posts to Peshawar ICRC hospital 1990 – 91.[24]

A new ratio: critical area to extremity wounds

A statistical method has been used to try to overcome this "bias" of logistical progress in order to better evaluate the effect of "natural triage" due to a lack of appropriate surgical care and evacuation facilities in low-income countries. In such conditions pre-hospital mortality cannot be properly determined either. The ratio of wounds in critical areas (trunk, head and neck) to wounds of the extremities is calculated among survivors: CA : Ext.

$$\frac{\textbf{Critical areas (head, face, neck, thorax, abdomen, pelvis)}}{\textbf{Non-critical areas (limbs, back, soft tissue injuries of trunk)}} = \textbf{CA : EXT}$$

CA = critical area, EXT = extremities and other non-critical areas.

Obviously, the wearing of body armour and the inclusion in statistics of minor or superficial, non-penetrating wounds to the head or trunk will distort the anatomic distribution of wounds, and the ratio.

In most wars, with evacuation counted in hours, the ratio is approximately 0.5. In urban areas and other situations of rapid evacuation, it tends towards 1. With increasing delay of evacuation to over 24 hours, and even days and weeks, the ratio falls.

Thus, studies of guerrilla and counter-insurgency warfare in difficult geographical terrain give much lower ratios indicating that the most seriously injured patients (critical area wounds) died before treatment (Table 5.15).

Conflict	Ratio
Thailand counter-insurgency	0.39
Eritrean war of independence	0.26
Ugandan guerrilla war	0.21
Afghan mujaheddin	0.07
ICRC Lokichokio hospital: southern Sudan guerrilla war	0.33

Table 5.15 Ratio of critical area to limb wounds in various insurgency conflicts.[25]

24 Korver AJH. Outcome of war-injured patients treated at first aid posts of the International Committee of the Red Cross. *Injury* 1994; **25**: 25 – 30.

25 Adapted from Bhatnagar MK, Smith GS. Trauma in the Afghan guerrilla war: effects of lack of access to care. *Surgery* 1989; **105**: 699 – 705.

5.9 Hospital mortality

5.9.1 Historical considerations

As mentioned, rates of killed in action for armies have remained relatively stable over the last half-century at about 20 – 25 %. Hospital mortality, however, has diminished greatly with medical progress (safe anaesthesia, blood transfusion, better understanding of the physiology of shock, antibiotics, and more aggressive surgical techniques) as shown in Table 5.16.

Conflict	Hospital mortality rate %
Crimean War 1854 – 55 (UK casualties)	16.7
American Civil War 1861 – 65 (Union casualties)	14.1
Boer War 1899 – 1901 (UK casualties)	8.6
World War I 1917 – 18 (USA casualties)	7.6
World War II (USA casualties)	4.5
American – Viet Nam War (USA casualties)	2.5

Table 5.16 Hospital mortality: examples from the past.

Hospital mortality has come to be an indicator of the efficiency of the management system, bearing in mind what has already been said about the paradox of early evacuation of the very severely injured. In calculating these figures, however, one must take into account the percentage of truly "vital" injuries and exclude or specify the superficial ones.

5.9.2 Hospital mortality *versus* post-operative mortality

A number of considerations must be taken into account if one wishes to use hospital mortality as an indication of the efficiency of patient management systems. In armed conflict, surgeons often face mass casualties. Some patients will be triaged into the "expectant" category, and given supportive treatment only, to die without pain and with dignity (see Chapter 9). These patients are recorded as DOW and are often included in hospital mortality figures.

Other patients die shortly after arrival, or "on-table", in an extreme attempt to save life. With a slightly longer evacuation time, many such patients would be KIA, dying before reaching hospital, as mentioned previously. Nonetheless, they are registered as DOW and are again included in hospital mortality figures.

Then there are the truly post-operative deaths; some from irreversible shock, others from irretrievable brain injury and, others still, from surgical complications, mostly sepsis, and additional medical pathologies.

Gross statistics of hospital mortality do not always take into consideration these very different categories of patients.

5.9.3 ICRC hospital mortality

Austere working conditions, often a limited number of professional staff and at times a perilous security situation, all make work in ICRC hospitals resemble that commonly seen in public hospitals in a poor country. Military medical services from industrialized countries may also face these constraints, but they tend to be of a different order of magnitude.

Post-operative mortality in ICRC hospitals varies from 2.2 % in Quetta and 3.1 % in Peshawar to 4.2 % in Kao-i-Dang, 4.8 % in Kabul, and 6.1 % in the Monrovia battle, where evacuation times were extremely short.

5.10 ICRC statistical analysis of hospital workload

5.10.1 Surgical audit: methodology

Without control over the pre-hospital phase and the known inadequacies of data collection, the mortality rate is not a credible statistic in ICRC surgical programmes and hospitals. The same is true of most public civilian structures. Far more substantial is the workload and morbidity as represented by the number of operations and blood transfusions per patient and the duration of hospitalization.

For statistical purposes, most patients, if well managed, undergo two operations: wound excision and delayed primary closure. There are a number of factors that explain this:

- the ICRC does not have a series of operative echelons of treatment as do the military; all operations are performed in the same hospital (see Chapters 1 and 6);
- rarely do ICRC surgeons perform planned serial debridements of war wounds, requiring more operations (see Chapter 10);
- some patients require a third operation for skin grafting to close the wound, or even more if a burn complicates a penetrating wound (see Chapter 11);
- others call for a single operation: (craniotomy, chest tube drainage, or laparotomy);
- some patients call for no operation, if they have been admitted solely for observation – expectant triage category, paraplegic, etc. (see Chapter 9);
- small superficial wounds are often treated conservatively with dressings and antibiotics, with the exception of those caused by anti-personnel landmines.

These "normal" single and multiple operations tend to balance each other out. (Pure burn injuries are not included in this discussion.)

Most patients require two operations: debridement and delayed primary closure. More than two usually denotes a complication, typically sepsis.

Performing more than two operations per patient usually denotes a complication, typically sepsis, and is a good indication of morbidity. Thus, as an appropriate approximation of workload and correct surgical management, the ICRC uses the average of two operations per patient in large studies. These premises form the basis for the following discussion.

5.10.2 Number of operations per patient: all patients

The number of operations per patient for all patients with the relevant information in the ICRC database is given in Table 5.17.

None	9 %
1 operation	16 %
2 operations	41 %
≤ 2 operations	66 %
3 operations	14 %
≥ 4 operations	20 %

Table 5.17 Number of operations per patient (N = 16,172).

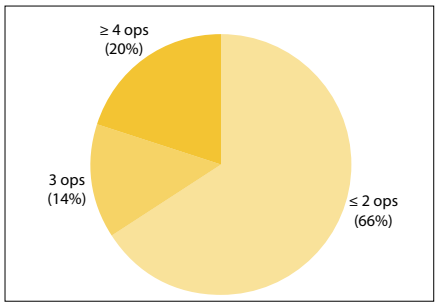

Figure 5.6

Number of operations per patient, simplified (N = 16,172).

To simplify, the number of operations are described as two or less; three; or four and more (Figure 5.6).

Fully two-thirds of all patients in the database had two operations or less, indicating a minimum of morbidity and good surgical management. There are a number of factors that influence this surgical workload: delay in evacuation, severity of the wound, type of wound, and wounding mechanism. A brief analysis of some of these factors is given in the following sections.

5.10.3 Number of operations according to delay in evacuation

Table 5.18 shows the results for the number of operations per patient according to evacuation time to the ICRC hospital in Kabul during a period of major urban combat.

Evacuation time	N = 6,140	None	1 op	2 ops	≤ 2 ops	3 ops	≥ 4 ops
< 6 hrs	3,214	7 %	30 %	47 %	84 %	7 %	8 %
6 – 24 hrs	1,606	7 %	23 %	51 %	81 %	9 %	11 %
24 – 72 hrs	605	6 %	24 %	50 %	80 %	7 %	12 %
> 72 hrs	715	9 %	26 %	42 %	77 %	9 %	14 %

Table 5.18 Number of operations per patient according to time of evacuation to ICRC Kabul hospital 1990 – 92.

These figures are confirmed if one examines the entire database (Table 5.19).

Evacuation time	N = 16,172	≤ 2 ops	3 ops	≥ 4 ops
< 6 hrs	2,409	81 %	9 %	10 %
6 – 24 hrs	3,727	70 %	13 %	17 %
24 – 72 hrs	2,785	69 %	13 %	17 %
> 72 hrs	7,251	71 %	12 %	17 %

Table 5.19 Number of operations per patient according to time of evacuation to ICRC hospitals 1990 – 99.

While hospital mortality is influenced by delay in evacuation, morbidity as determined by number of operations per patient apparently is not. The critical limit would appear to be 6 hours delay – the category "6 – 12 hours since injury" does not exist in the ICRC database, and this is a field for future study. However, before rushing to such a conclusion, another factor must be analysed: wound severity.

5.10.4 Number of operations according to Wound Grade

War wounds are categorized into one of three Grades of increasing severity according to the Red Cross Wound Score and classification system (see Chapter 4). The distribution in the ICRC surgical database is as follows.

- Grade 1: 42 %.
- Grade 2 : 37 %.
- Grade 3 : 21 %.

If one looks at the number of operations according to Wound Grade, an important difference appears (Table 5.20 and Figures 5.7.1 – 5.7.3).

Severity of wound		Number of operations per patient					
	Patients (N =16,172)	None %	1 op %	2 ops %	≤ 2 ops %	3 ops %	≥ 4 ops %
Grade 1	n = 6,729	16	23	45	84	9	7
Grade 2	n = 5,974	4	12	44	61	16	23
Grade 3	n = 3,469	3	7	30	40	19	41

Table 5.20 Number of operations per patient according to Red Cross Wound Grade.

The Grade of the RCWS shows a distinct difference in the number of operations required. Surgical workload is definitely influenced by the severity of injury, as would be expected, and the RCWS is capable of determining this.

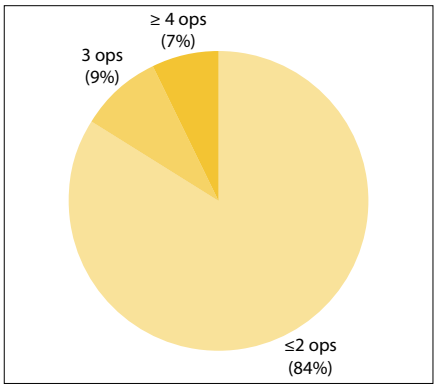

Figure 5.7.1

Number of operations per patient for Grade 1 wounds (N = 6,729).

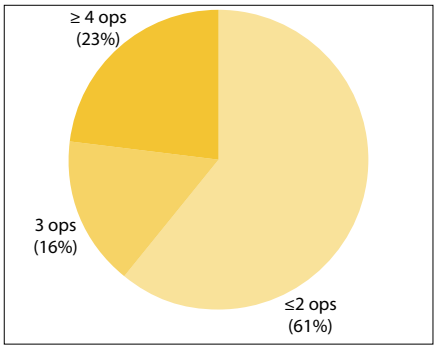

Figure 5.7.2

Number of operations per patient for Grade 2 wounds (N = 5,974).

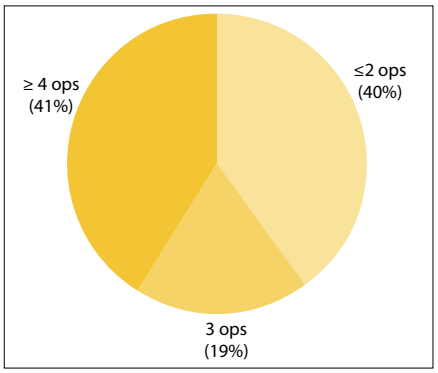

Figure 5.7.3

Number of operations per patient for Grade 3 wounds (N = 3,469).

5.10.5 Number of operations according to Wound Grade and delay in evacuation

If one now analyses the number of operations by combining the RCWS Wound Grade and the delay in evacuation the results are telling (Table 5.21).

Grade and evacuation time	Number of patients	≤ 2 operations	3 operations	≥ 4 operations
Grade 1	N = 6,729			
< 6 hrs	1,124	93 %	5 %	3 %
6 – 24 hrs	1,694	84 %	9 %	7 %
24 – 72 hrs	1,182	82 %	11 %	7 %
> 72 hrs	2,729	82 %	10 %	9 %
Grade 2	N = 5,974			
< 6 hrs	788	77 %	11 %	12 %
6 – 24 hrs	1,186	62 %	16 %	22 %
24 – 72 hrs	1,110	58 %	17 %	25 %
> 72 hrs	2,890	56 %	18 %	26 %
Grade 3	N = 3,469			
< 6 hrs	497	47 %	17 %	35 %
6 – 24 hrs	847	37 %	19 %	44 %
24 – 72 hrs	493	39 %	19 %	42 %
> 72 hrs	1,632	40 %	20 %	40 %

Table 5.21 Number of operations per patient according to Red Cross Wound Grade and time since injury.

It would thus appear that the Red Cross Wound Score better represents morbidity and surgical workload than the delay in evacuation to hospital alone, but a combination of the two is more noteworthy. Grade 3 wounds tend to be very severe and delay does not appear to make a difference. Many such patients will simply not survive to reach hospital.

5.10.6 Number of operations according to weapon

The results of an analysis of the ICRC hospitals treating a variety of weapons-induced injury (Kabul, Kandahar, Kao-i-Dang, Novye Atagi, Peshawar, and Quetta), are presented in Table 5.22.

No distinction is made in the ICRC database between anti-personnel and anti-tank mines, or unexploded ordnance. Furthermore, some wounds classified as being caused by fragments may well have been from fragmentation anti-personnel mines. Hospital staff have only the account of the patient to go by and, naturally, many patients are ignorant of weapons systems; they only know of "bombs" or "guns".

Weapon	None	1 op	2 ops	≤ 2 ops	3 ops	≥ 4 ops
Mines N = 5,236	9 %	15 %	38 %	62 %	14 %	24 %
GSW N = 5,984	9 %	22 %	44 %	75 %	12 %	13 %
Fragment N = 7,674	11 %	24 %	44 %	80 %	9 %	11 %

Table 5.22 Number of operations per patient according to weapon.

It is obvious that mines imply a much greater workload and morbidity than gunshot (GSW) or fragment wounds (Table 5.22 and Figure 5.8).

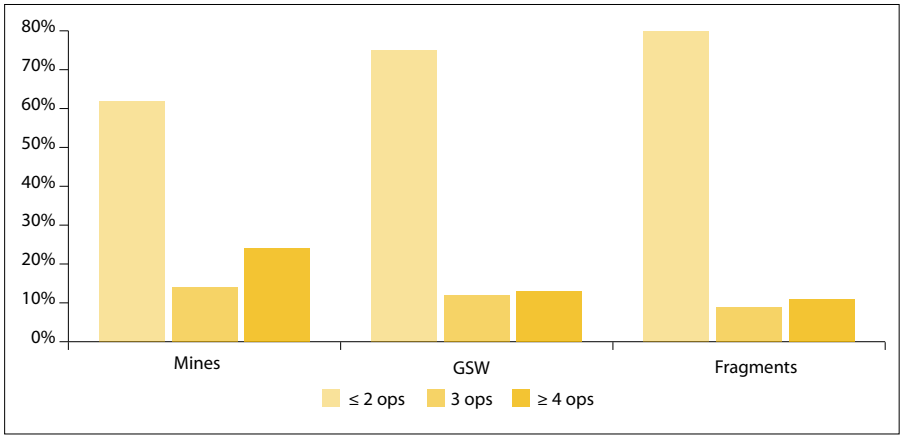

Figure 5.8

Number of operations per patient according to weapon, simplified.

Although the lethality of weapons is important, the total effects of surgical workload, morbidity and suffering, and the socio-economic consequences must also be taken into account. It is not for nothing that poison gas, bacteriological weapons, blinding laser weapons, anti-personnel mines and cluster submunitions have been proscribed and banned by international treaty.

5.11 Conclusions: lessons to be gained from a study of epidemiology

From this brief overview of the epidemiology of the victims of war, a few conclusions affecting clinical work and humanitarian action may be drawn.

1. For a civilian population in a poor country, the public health effects of war are usually greater than the effects of direct trauma. In some conflicts, war injuries carry a greater burden and post-traumatic morbidity may have more profound long-term effects than mortality. This is especially the case in a post-conflict situation with widespread contamination by anti-personnel mines whose socio-economic repercussions are long-lasting.

2. The outcome of war surgery in a civilian context is influenced by:
   - the type of injury according to wounding agent;
   - the severity of the injury;
   - the general condition of the patient (malnutrition, chronic disease, concomitant endemic disease such as malaria, etc.);
   - early and appropriate first aid;
   - the time needed for transport to the hospital;
   - the quality of hospital treatment (resuscitation, surgery, post-operative care, physiotherapy and rehabilitation);
   - possibility of evacuation to a better-equipped hospital with more experienced staff.

3. In a civilian context, and especially in a poor country, pre-hospital care is the phase most open to improvement. Much can be done to prevent death and morbidity with early and efficient first aid and life support.

   An efficient first-aid and evacuation system can prevent deaths from controllable haemorrhage and compromised airway. As pre-hospital care improves, the "killed in action" or "dead on arrival" may diminish slightly, only to see the "dead of wounds" and hospital mortality rise; more casualties are saved, but the rates become distorted. This bias should be taken into account when judging the adequacy of care.

   Long and difficult evacuation of the wounded results in an automatic "natural triage" of the most severely injured. Hospital mortality rates decrease as a result.

4. Up to 40 – 50 % of the civilians wounded during urban fighting do not need hospitalization. First-aid measures – plus a simple oral antibiotic and analgesic – are all that are required. Their presence constitutes a supplementary burden on hospital management. However, discharging them from hospital can prove problematic because of security concerns, socio-economic factors and patients' fears.
5. Wounds to the head and torso are the most lethal. Injuries to the soft tissues and the limbs constitute the majority of surgical workload.
6. Gunshot wounds carry a heavier burden than injuries due to fragments; but burns and anti-personnel landmine injuries represent the heaviest hospital workload and morbidity.
7. The Red Cross Wound Score allows a good appreciation of the severity of war wounds, and the surgical workload that they represent.

5.11.1 Setting up a surgical database for the war-wounded

Various military medical services have their own categories and charts for data collection. The ICRC offers civilian health facilities an example of categories that could be placed in a simple spreadsheet (Excel ®) to allow for a collection of relevant data for further study (see Annex 5. B: Setting up a surgical database for the war-wounded).

In addition, a downloadable file form of Annex 5. B (Setting up a surgical database for the war-wounded) is available on the accompanying DVD (Annex 5. C). It can be modified by the user.

ANNEX 5. A ICRC surgical database

The ICRC set up a centralized wound database and trauma registry in 1990, originally designed to give the organization an indication of the surgical workload of its independent hospitals (i.e. hospitals that were established and run under ICRC administration). All war-wounded patients admitted to ICRC hospitals have routinely had a data form filled out on their death or discharge. Patients are not asked whether they are combatants or civilians.

Age and sex, the cause and anatomic site of injury, and the time lapsed between injury and admission are recorded for each patient. Projectile injuries are scored according to the Red Cross Wound Classification System described in Chapter 4. The emphasis is on surgical workload as described by severity of injuries, number of operations per patient, number of blood transfusions required, and length of hospitalization.

These independent ICRC hospitals have served the victims of a number of armed conflicts of very different types of combat. As at 31 December 2007, the database contained the records of 32,285 war-wounded patients. Not all are complete, however.

Hospital	Period of activity	Nature of combat
Kao-i-Dang Hospital, Thailand	1979 – 92	Jungle guerrilla warfare in Cambodia
Peshawar Hospital, Pakistan	1981 – 93	Mountain guerrilla warfare in Afghanistan
Quetta Hospital, Pakistan	1983 – 96	Mountain guerrilla warfare in Afghanistan
Karteh-Seh Hospital, Kabul, Afghanistan	1988 – 92	Irregular warfare, mostly urban
Mirweis Hospital, Kandahar, Afghanistan	1996 – 2001	Irregular warfare, mostly urban
Butare Hospital, Rwanda	1995	Irregular guerrilla warfare, mostly rural
Novye Atagi Hospital, Chechnya, Russian Federation	1996	Irregular guerrilla warfare, rural and urban
Lopiding Hospital, Lokichokio, Kenya	1987 – 2006	Rural guerrilla war in semi-desert, sahel and reed swamps of southern Sudan

The above-mentioned conflicts were different in nature. Time to treatment often included extremes – minutes to weeks. Logistic difficulties over great distances meant that the ICRC could rarely organize efficient and timely evacuation of the victims. Notable exceptions were the first-aid posts set up near the Afghan border with Pakistan and close to Kabul. The air medical evacuation programme for southern Sudan, coordinated with the United Nations Operation Lifeline Sudan, saw over 20,000 patients transferred to the ICRC hospital in Lokichokio, northern Kenya, but the distances and delays in notification of patients nevertheless seldom allowed timely evacuation.

The ICRC has also set up other hospitals, and its surgical teams have worked in local public hospitals, but their patients are not included in the database. In addition to the surgical database, other major sources of data include the Somali Red Crescent-run Keysaney Hospital in Mogadishu (1992 – present), and the JFK Memorial Hospital, Monrovia, Liberia (2001 – 2004), operated jointly by the ICRC and the hospital's Board of Governors.

When compiling ICRC statistics from the database, the limits of a retrospective analysis of prospectively collected data taken exclusively from hospital admissions must be recognized. The usual problems have been observed:

- clerical errors,
- missing patient files,
- incomplete patient files (not all categories were filled in for all patients),
- confusion as to the definition of classification categories,
- lack of continuity due to the constant turnover of hospital personnel,
- lack of properly trained administrative support staff to maintain a database.

Nonetheless, ICRC statistics are offered throughout this manual as an example of non-military experience in different combat zones as an approximation of the reality of the battlefield, especially for non-combatant populations.

ANNEX 5. B Setting up a surgical database for the war-wounded

The following categories can be recorded in an electronic database (Approach ®) or listed in the *columns* of a simple spreadsheet (e.g. Excel ®) and patients registered in the *rows*. Lessons have been learnt from shortcomings in the ICRC surgical database: this version has been modified accordingly.

Administrative data:

- hospital (if the database contains patients from more than one hospital)
- patient hospital number
- database number
- date of admission
- date of discharge
- number of days hospitalized
- re-admission of patient for the same injury?
- age
- sex

Time since injury:

- hours (or < 6 ; 6 – 12; 12 – 24)
- days (or 24 – 72 hrs ; >72 hrs)
- weeks

Weapon causing injury:

- rifle, gun
- bomb, shell, mortar, grenade
- anti-personnel mine (APM)
- anti-tank mine (ATM)
- unexploded ordnance (UXO), including cluster bombs
- "arme blanche": bayonet, machete, *panga*

Mechanism of injury:

- gunshot (GSW)
- fragment
- blast
- blunt
- burn

Please note:

A bomb, artillery shell or anti-tank mine may give off penetrating fragments, cause blast or burn injury and, by destroying a vehicle or building, also cause blunt injury. One weapon is registered; several mechanisms of injury are recorded. The same can be said of anti-personnel mines.

Clinical data:

- blood pressure on admission
- number of operations
- number of anaesthesias
- number of units of blood transfused
- outcome: healed, complications, death (including cause)

Anatomic data:

Site of injury: if more than one, each injury is given a number and the appropriate number put in the column for the anatomic region. Only one injury per anatomic region is noted.

- head
- face
- neck
- thorax
- abdomen
- pelvis, buttocks
- back and soft tissue of torso
- upper limb left (to determine bilateral limb injuries)
- upper limb right
- lower limb left
- lower limb right

Red Cross Wound Score:

This should be entered for the two most important wounds, more if so desired. The number 1 wound should correspond to the number 1 anatomic region; wound 2 for number 2 anatomic region.

- Wound 1: Entry
- Wound 1: Exit
- Wound 1: Cavity
- Wound 1: Fracture
- Wound 1: Vital injury
- Wound 1: Metallic fragment
- Wound 1: Grade
- Wound 1: Type
- Wound 2: Entry
- Wound 2: Exit
- Wound 2: Cavity
- Wound 2: Fracture
- Wound 2: Vital injury
- Wound 2: Metallic fragment
- Wound 2: Grade
- Wound 2: Type

Major operation:

- craniotomy
- thoracotomy
- chest tube
- laparotomy
- peripheral vascular repair
- amputation above-elbow left
- amputation above-elbow right
- amputation below-elbow left
- amputation below-elbow right
- amputation above-knee left
- amputation above-knee right
- amputation below-knee left
- amputation below-knee right

Comments:

Please note:

This sample database is available as a downloadable spreadsheet file (Excel ®) on the accompanying DVD.

Chapter 6

THE CHAIN OF CASUALTY CARE

6. THE CHAIN OF CASUALTY CARE

6.1 The links: what kind of care, and where? 131
6.1.1 Protection: IHL 131
6.1.2 Levels and locations of medical care 132

6.2 Surgical hospital treating the war-wounded 133
6.2.1 Hospital A: rural hospital = basic trauma services 133
6.2.2 Hospital B: provincial hospital = advanced trauma services 133
6.2.3 Hospital C: major city hospital = comprehensive trauma services 133
6.2.4 Hospital assessment 133

6.3 Transport 134
6.3.1 Command, control and communications: coordination 134

6.4 Forward projection of resources 135

6.5 The reality: common war scenarios 136

6.6 Conflict preparedness and implementation 137

6.7 The pyramid of ICRC surgical programmes 138
6.7.1 Checklist 139

ANNEX 6. A Initial assessment of a surgical hospital treating the war-wounded 140

ANNEX 6. B Strategic assessment of a conflict scenario 146

ANNEX 6. C Humanitarian intervention for the wounded and sick: typical settings 148

6.1 The links: what kind of care, and where?

Modern armed conflict embraces classical war between conventional armies, urban combat between militias, and isolated and sporadic but fierce guerrilla attacks in remote rural areas. It may involve mass conflict or chronic irregular, low-intensity combat, or individual terrorist attacks. Civilians often represent the majority of victims (see Chapter 5). Field situations vary considerably, but basic medical problems for the wounded are universal. A system must be developed and adapted to deliver the best possible care in a timely manner, under all circumstances.

Modern armed conflict takes place in rural areas and cities.

Civilian populations are increasingly at risk.

Landmines and unexploded cluster munitions continue to cause victims after the conflict has ended.

Field situations vary; medical needs are universal.

The system of evacuation and treatment of the sick and wounded – the chain of casualty care – like any chain, is as strong as its weakest link. It is measured not in kilometres but in hours and days. The setting up of such a system requires planning ahead of time: an assessment of the tactical circumstances must be made; an analysis of the physical limitations and human resources carried out; and the resulting plan implemented.

Wounded patients are transferred along a chain of medical care, beginning with simple "life- and limb-saving" procedures and continuing to ever-greater levels of sophistication. The principle of echelons of care in a military system has been described in Chapter 1. In civilian practice patients also tend to follow a chain of evacuation and treatment although, in many countries, this is not very efficient.

6.1.1 Protection: IHL

Protection and prevention from further injury come first and foremost in patient care.

Article 3 common to the four Geneva Conventions and article 7 of Additional Protocol II:

The wounded and sick shall be collected and cared for.

The wounded and sick have the *right* to be cared for and to have access to appropriate medical care. The red cross, red crescent and red crystal emblems symbolize the legal protection afforded the casualty and the medical personnel, who have the right and the obligation to care for the wounded and sick. The protective quality of these emblems is a function of the training and degree of discipline of the combatant forces, and their adherence to internationally accepted norms of behaviour on the battlefield: the Geneva Conventions and their Additional Protocols. This protection begins with the first aider in the field and continues through all the levels of treatment.

Special measures must be taken to prevent further injury to any casualties and to protect them from the elements (see Chapter 7).

6.1.2 Levels and locations of medical care

The following list defines the generic places where the wounded receive different levels of medical care in the multiple-phase treatment of war wounds.

1. On the spot: first aid.
2. Collection point: first aid with/without resuscitation.
3. Intermediate stage: resuscitation with/without emergency surgery.
4. Surgical hospital: primary surgical treatment.
5. Specialized centre: definitive surgical treatment including reconstructive procedures; physiotherapy and rehabilitation, both physical and psychological.
6. Transport system for transfer from one echelon to another.

Treatment on the spot may be self- or "buddy"-treatment, or practised by a military medic or first aider. First aid starts at the point of wounding, but can be given anywhere and everywhere along the evacuation line to the place of definitive treatment. First aid is the only possible treatment on the battlefield.

The setting up of a collection point or clearing station may be a spontaneous response; alternatively, an established first-aid post (FAP), dispensary, or primary health centre might serve this purpose. Some FAP will be staffed by military stretcher-bearers and medics, others by Red Cross/Red Crescent volunteers, or other civilian personnel. The closer to the battlefield, the more prominent will be the role of military medical services. In purely civilian rural settings, village health workers, nurses or medical assistants are often the only health professionals available. Collecting the wounded at one specific location allows for the proper organization of field triage and their efficient evacuation. Besides basic first aid and trauma life support, resuscitative measures may be begun here.

A health centre or rural hospital might represent the intermediate stage, where more sophisticated resuscitation and emergency surgery are available. The military might establish a forward surgical unit.

B. Fjortoft / ICRC

Figure 6.1
National Society first aiders in the field.

J-P. Kolly / ICRC

Figure 6.2
First-aid post on the Afghan-Pakistani border.

J-J. Kurz / ICRC

Figure 6.3
Transfer of patients to the ICRC hospital in Peshawar, Pakistan.

6.2 Surgical hospital treating the war-wounded

The level of sophistication of hospital care will depend on the degree of socio-economic development within the country *before* conflict breaks out. War will then usually degrade the functioning of any pre-existing hospital. Efficient evacuation of the wounded to deficient surgical facilities is not an effective chain of casualty care.

Levels of hospital competency differ from country to country and between geographic regions. Three basic levels can be described.

M. Bleich / ICRC

Figure 6.4
ICRC field hospital, Lokichokio, Kenya.

6.2.1 Hospital A: rural hospital = basic trauma services

A district hospital (Africa) or primary health centre (South Asia) without full-time specialist doctors, especially without a fully-trained general surgeon, is a typical example. These hospitals are usually staffed by general practitioners or medical assistants with some surgical training and equipped with a minimum of proper surgical facilities. Front-line hospitals with "field surgeons" exist in some military medical services and are the equivalent. Basic resuscitation and some simple but critical operations are usually performed here.

6.2.2 Hospital B: provincial hospital = advanced trauma services

A regional hospital (Africa), district hospital (South Asia), or general hospital (Latin America) have a full-time general surgeon and proper operating theatre and sterilization facilities. Most life-threatening conditions can be dealt with adequately. The military equivalent is usually a forward field hospital specializing in damage control and resuscitative surgery.

6.2.3 Hospital C: major city hospital = comprehensive trauma services

This is a referral structure offering a broad range of specialties and sub-specialties and usually represents the highest level of care in the country or administrative region. It is often a teaching or university hospital. In military terms, this is a referral structure far from the front lines.

6.2.4 Hospital assessment

ICRC programmes aim to maintain, and upgrade if possible, the competencies of these different levels of hospital care when necessary. The ICRC has developed an analytical tool for assessing the quality of hospital management and surgical work prior to providing assistance. This tool comprises a number of factors to be considered. There are factors external to the hospital, involving an analysis of the national and provincial health systems, and internal factors: hospital infrastructure, administration and functional organization, departments, available resources, medical personnel (number and expertise), non-medical support services, finances (see Annex 6. A: Initial assessment of a surgical hospital treating the war-wounded).

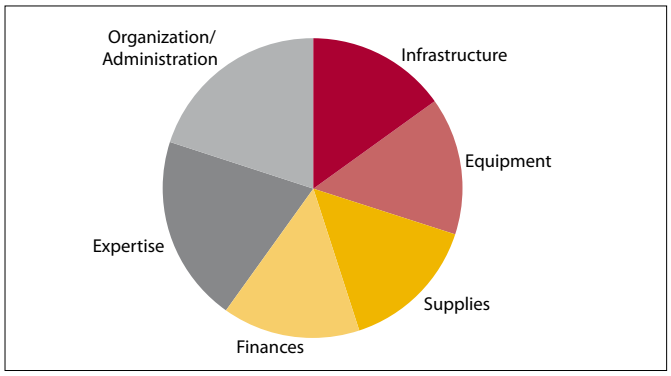

Figure 6.5

Assessment of a surgical hospital caring for the war-wounded.

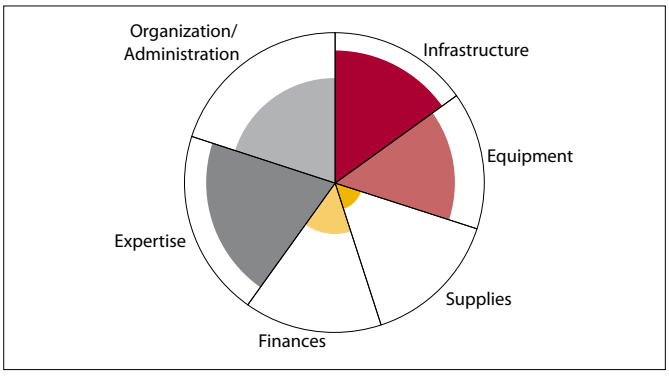

Figure 6.6

Assessment results of a typical hospital in a low-income country disorganized by war.

T. Gassmann / ICRC

Figure 6.7.1

National Society volunteers transporting patients by boat.

L. Petridis / ICRC

Figure 6.7.2

Airplane medical evacuation from southern Sudan to ICRC Lokichokio hospital.

T. Gassmann / ICRC

Figure 6.7.3

National Society "Zero Mobile" horse-drawn carriage ambulance.

The pie-charts above show the various factors that affect the functioning of a hospital dealing with the added burden of war-wounded patients compounded by the constraints of a weakened health system, and help identify dysfunctional areas. An additional factor is the emotionally stressful presence of casualties who are related to, or friends of, the medical staff.

6.3 Transport

Transport of some sort provides the connection between the different echelons of the chain of casualty care. Moving a patient has a price: transportation is a trauma in itself. It uses up additional resources and involves security risks ("mortality of the ambulance ride"), and perhaps even exposure to military activity. These extra costs must be weighed up against the likely benefits of moving the wounded. In many contexts, the availability of transport for the sick and wounded is almost a "luxury".

Moving wounded people is difficult, always takes longer than expected, adds to the trauma, and is often dangerous.

6.3.1 Command, control and communications: coordination

A chain of command is necessary to make the various echelons function correctly. A central command or dispatch centre will assume overall coordination (e.g. decisions about transfer/evacuation destinations, engagement of resources, etc.), and be responsible for contacts with related command levels of various authorities (e.g. armed forces, police, headquarters of the National Red Cross/Red Crescent Society, civil defence and national rescue service, etc.).

Circulation of information between the different levels is assured by some means of telecommunication – radio and mobile phones – if possible, or by other means (e.g. foot messengers) if not. Mobile telephone systems have a tendency to stop functioning – be turned off – during crises and armed conflict. The efficiency of the command and communication systems depends on strict observance of established procedures.

6.4 Forward projection of resources

Bringing more than basic first aid closer to the collection point is called "forward projection of resources". The availability of advanced procedures closer to the battlefield has many advantages. It allows quicker access to "life- and limb-saving" emergency measures, thus decreasing both mortality and morbidity. The need for potentially dangerous transport is reduced. The projection of resources applies particularly to treatment at the intermediate stage, but can be applied to any echelon in the chain.

> Forward projection of resources allows quicker access to life- and limb-saving measures.
>
> However, certain conditions must be met.

A number of factors limit the possibility to project resources.

1. Security (essential).
2. Human resources and expertise (essential).
3. Equipment (appropriate technology).
4. Supplies (appropriate).
5. Infrastructure (minimum requirements).
6. Possibility of onward evacuation.

When a front-line hospital is coming under bombardment, putting both patients and personnel at risk, there is little point in attempting more than first aid if evacuation to another facility is possible. On the other hand, a forward collection point may provide advanced treatment. Only a minimum of equipment is necessary for well-trained medics to perform endotracheal intubation or place a chest tube, and then quickly evacuate the patient. A secure building with adequate resources may be used as an intermediate stage where a forward surgical team may perform damage control and resuscitative surgical procedures.

All of these possibilities will depend on meeting the criteria defined above; but especially on *security and human expertise*. Infrastructure, equipment and supplies must all meet minimum requirements, and be appropriate to the task under the prevailing conditions.

> The most important factors determining the forward projection of medical care for the war-wounded are security and human expertise.

The appropriate choice of procedures to undertake for the injured outside a formal hospital setting will depend on the factors mentioned above, and will vary from country to country, and even from region to region within the same country.

The organization of any chain of casualty care for the war-wounded – military or civilian – should rely on a large dose of common sense to determine what is practical and what can realistically be accomplished to assure the best results for the greatest number, while at the same time guaranteeing the safety of the injured and the health workers. What exactly can be done for the injured outside a formal hospital setting will depend on the particular circumstances and means available. There is no dogma to be followed blindly; situations are different and improvisation and adaptation the key to success.

	On the spot	Collection point	Intermediate stage
Who ?	Family, friends, the community[1] Community health workers First aiders (Red Cross or Red Crescent, military stretcher-bearers and medics, combatants, etc.) Health professionals	Health professionals First aiders (Red Cross or Red Crescent, military stretcher-bearers and medics, combatants, etc.)	General practitioners, emergency room staff, other medical and surgical professionals Field surgical team
Where ?	In a combat zone, at the front lines	Spontaneously chosen (e.g. the shade of a tree) First-aid post Dispensary, primary health care centre	First-aid post, dispensary, primary healthcare centre, outpatient clinic Rural hospital Forward surgical hospital
What ?	Life-saving first-aid measures *The only appropriate care on the spot*	Collection of casualties Evaluation of their condition Advanced emergency care and/or stabilization Evacuation planning Routine care (fever, diarrhoea, scabies, etc.) and ambulatory care (pneumonia, non-combat trauma, etc.)	Advanced emergency care Forward life-saving surgery Occasional hospital care, although uncomplicated and requiring few days of observation Routine care (fever, diarrhoea, scabies, etc.) and ambulatory care (pneumonia, non-combat trauma, etc.)

Table 6.1 What can be done at which level?

6.5 The reality: common war scenarios

As mentioned in Chapter 1, there is more than one "surgery" for the care of the war-wounded. The exact number of levels of care and the path followed by casualties are determined on a case-by-case basis according to the sophistication of care and logistics available. In some armies or countries, the organization may be so efficient that a wounded soldier may expect to receive treatment virtually as sophisticated as the care available to him in peacetime.

In developing countries, however, the healthcare system might already be weak before the conflict and may almost cease to function because of it. Water and electricity supplies can be unreliable, trained staff often flee the area, drugs and disposable equipment cannot be replaced, budgets and salaries are not paid, and buildings are destroyed. The quality of hospital care is badly affected.

Every combination of levels of care is possible:

- the wounded from a modern industrialized army are transferred by helicopter from the scene of injury directly to a surgical hospital;
- patients in low-income countries reach health facilities on foot or by oxcart, donkey, private motor vehicle, taxi or lorry;
- during urban warfare, family and neighbours transport a casualty directly to the emergency room of a surgical hospital, which then serves as a collection point;
- in the rural areas of a low-income country, the hospital is the only health structure functioning and serves as all levels in one – forward referral of patients to a more sophisticated level of care is not available;
- rural health centres may flourish while surgical hospitals may exist only in large cities;
- a collection point or intermediate stage in a secure building is upgraded with the necessary equipment and personnel to serve as a forward surgical hospital, with or without possibility of forward referral.

In some contexts, the ICRC has resorted to deploying a Field Surgical Team that is not hospital based. This mobile team "goes to the wounded" rather than having the wounded come to a hospital; the chain of casualty care is turned upside-down. This

1 In armed conflicts, under international humanitarian law, civilians are permitted to collect and care for the wounded and sick of whatever nationality, and shall not be penalized for doing so. On the contrary, they must be aided in this work. Furthermore, IHL prescribes that the civilian population shall respect the wounded and sick, even if they belong to the enemy, and shall commit no act of violence against them.

has been necessary where casualties have not had access to medical care for reasons of personal security (Figures 1.4 and 6.8).

Techniques will also differ according to context. For the military, there is a balance between the needs of the wounded soldier and the necessities of combat. Some armed forces may teach the use of the self-applied tourniquet, which purportedly allows a wounded soldier to continue shooting. This logic does not apply to civilian institutions, and the ICRC's first-aid manual[2] proscribes the use of a tourniquet in the field except in very rare circumstances.

V. Louis / ICRC

Figure 6.8
The ICRC Field Surgical Team in Darfur.

6.6 Conflict preparedness and implementation

Every country should have a disaster plan. Part of emergency preparedness is the capacity to respond to a situation of armed conflict or internal troubles, as well as natural disasters. This is normal procedure for most armed forces. Ministries of public health and National Red Cross/Red Crescent Societies usually also have a disaster plan, which should be integrated into the national emergency preparedness programme.

> The objective of planning is to ensure that wounded people get the right care, at the right place, and at the right time.

Those who have to face the challenges of armed conflict should understand how to proceed in setting up a chain of casualty care. The best possible outcome for wounded people can only be achieved with proper planning and training. Plans must be realistic, flexible and reviewed regularly. Should a country be caught unawares at the outbreak of hostilities without having a pre-prepared plan, then the planning process must be accelerated and the necessary analysis take place immediately.

All plans begin with a *strategic assessment* of the possible conflict scenarios. What may happen and where? What needs are foreseen? What resources exist? (see Annex 6. B: Strategic assessment of a conflict scenario).

An *analysis of the assessment* will then answer the questions of what should be done, where, and by whom to improve the care for the wounded.

The reality check asks a few basic questions. Are the recommendations compatible with the context? Are they relevant? Practical? Reality checks are important; they influence planning, care, and training, and make certain that academic theory does not override simple pragmatism.

The organization of a chain of casualty care and in particular the assignment of resources (physical and human) to the different echelons must take into consideration a number of factors:

- the nature of the conflict, the tactical and geographic circumstances, and security concerns;
- the size of the casualty load;
- the quality of the casualty load, i.e. the types of injuries;
- the relative and proportionate number of human resources (surgical capacity depends on the number of teams working *and* the rate of arrival of casualties);
- the quality of human resources, i.e. expertise;
- logistics and supplies;
- the infrastructure.

2 Giannou C, Bernes E. *First Aid in Armed Conflicts and Other Situations of Violence*. ICRC: Geneva; 2006.

Some typical situations of armed conflict and internal violence can be described. Model scenarios are then based on a consideration of the above-mentioned factors and the analytical tool of strategic assessment (see Annex 6. C: Humanitarian intervention for the wounded and sick: typical settings).

6.7 The pyramid of ICRC surgical programmes

ICRC delegates are often called upon to help organize a chain of casualty care. At times this will be in support of existing governmental or non-governmental structures. On occasion, the ICRC is required or asked to establish its own independently-run hospitals. The main reasons for doing so are either a question of protection – of the patients or the medical mission – or because of a dire shortage of local human resources.

Unlike a military field hospital, an independent ICRC hospital is often responsible for all levels of medical care. It acts as a first-aid post, field hospital, base hospital, and referral centre. In this, an ICRC hospital resembles many provincial hospitals in non-industrialized countries ravaged by war. The level of care is a combination of advanced and comprehensive trauma services. Only general surgeons are present, and no further referral of patients to a specialized centre is possible.

An independent ICRC hospital comprises all echelons in one structure.

In both instances – support to national structures or establishment of an ICRC hospital – a number of factors are taken into consideration to ensure the neutrality and independence of ICRC activities on the one hand, and the quality and professionalism of care on the other. The same logic applies to other humanitarian agencies. These factors are summarized in the following pyramid and checklist (Figure 6.9 and Section 6.7.1).

Figure 6.9
The ICRC surgical pyramid.

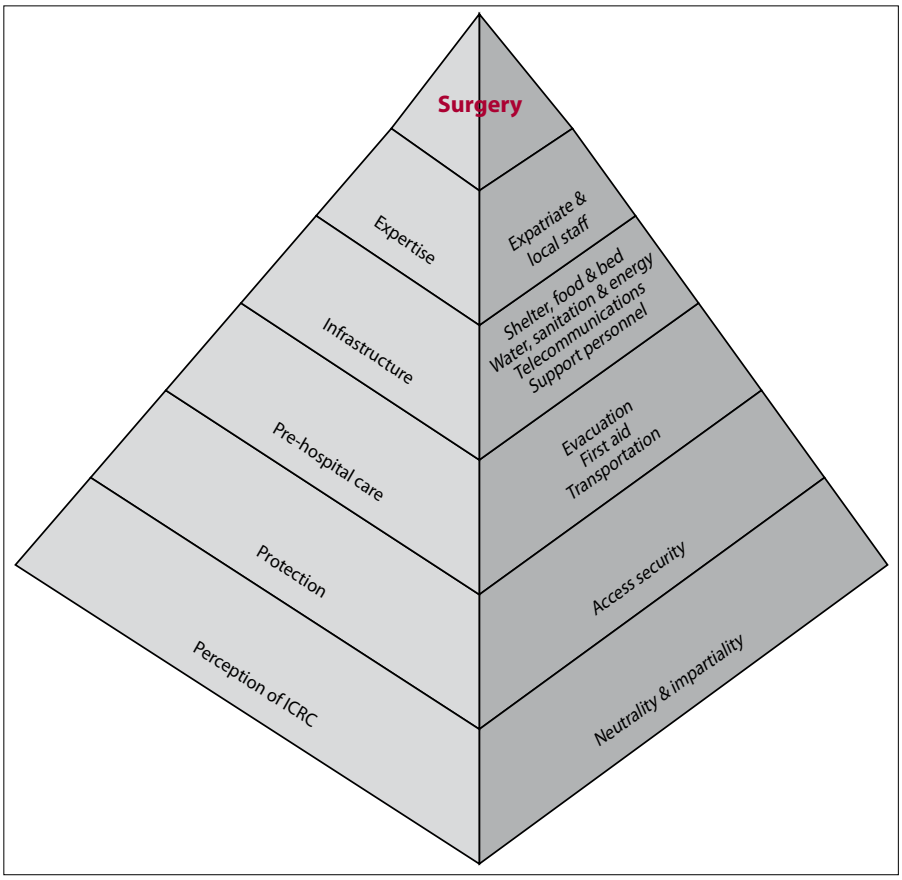

6.7.1 Checklist

Political and operational considerations

- Perception of neutrality and impartiality, i.e. ICRC image.
- Acceptability of ICRC by political factions, the military, and the population.
- Possibility for negotiations: availability and access to interlocutors (health, defence, foreign affairs, and interior ministries, faction chiefs); contacts with owners of facilities to be rented.
- Role of the hospital in the protection of hospitalized patients.
- Role of the hospital in protecting the medical mission by providing access to neutral medical care, across enemy lines if necessary.
- Any possible competition with other health structures, private or public, i.e. "medical market".

Security

- Hospital and staff security concerns owing to the conflict or possible developments in the fighting:
  - location and environment, i.e. distance from hostilities or military targets;
  - type of building, i.e. number of storeys, ground floor, underground cellar or bomb shelter, tents.
- Incidence of banditry, hostage-taking, etc.
- Patients' safety, inside the hospital and upon discharge.

Access

- Distance and evacuation time.
- Availability of means of transportation: roads, vehicles, air evacuation.
- Safety of roads and checkpoints: military activity, political affiliation, banditry.
- Possibility of setting up a system of first-aid posts.
- Logistics: medical supply, fuel, food (local, regional, from ICRC headquarters).

Infrastructure

- Pre-existing hospital structure.
- Building capable of being transformed into a hospital (school, factory); structural integrity, possibility of expansion.
- Use of tents, prefabricated and/or temporary structures.
- Water and sanitation, electricity.
- Facilities: kitchen, laundry, residence for personnel.
- Warehouse.

Local and expatriate personnel

- Availability and number of local and expatriate Red Cross/Red Crescent medical personnel.
- Recruitment of new national staff and the question of neutrality.
- Professional competence and level of education.
- Language problems: need for translators.
- Availability of support personnel: ICRC delegates, medical and general administration, water and sanitation engineer, builder, mechanic, electrician, etc.

ANNEX 6. A Initial assessment of a surgical hospital treating the war-wounded

This form should be regarded as a guide, a kind of checklist, to help the health professional responsible for the assessment of the hospital remember certain key points.

Its purpose is to have a quick overview and understanding of the functioning of the hospital; to identify its capacity, limits and insufficiencies. It should allow for rapid and adequate decisions about the kind of *support* the hospital might need.

The form can be used *in toto* or in part according to the objectives of the assessment. It consists of questions deliberately left open to allow the health professional in charge to decide to enquire further or not according to the results expected. It is more important to try to get a correct understanding of how the hospital is working rather than to have exact figures.

The assessment is divided into six sections

General

Management and administration (including non-medical support services)

Medical support services

Clinical services *(only the surgical component is dealt with in this annex)*

Further comments

Conclusion

General

Name of the hospital: Town: Country:

Assessment done by:

Date:

Interlocutors:

1. Type (MoPH, private, military, missionary, NGO, others):
2. Catchment population:
3. Assistance from others than the authority in charge:
4. Level of reference (rural, district, regional):
5. If rural or district hospital, number of primary facilities served (first-aid posts, dispensaries, health centres):
6. Possibilities for further referral:
7. Transport system for patients (in and out):
8. Reputation of the hospital (indicate source of information):
9. Bed capacity, effective number of beds (breakdown by department):
10. Present bed occupancy:
11. Activities including specialities (surgery, medicine, paediatric, obstetric, specialized services, etc.):
12. Security (Is the area safe? Is the hospital secured? i.e. clearly marked, fenced, watchmen present, absence of arms inside the hospital compound?):
13. Endemic diseases and epidemic risk in the region:

Management & administration

I General management

1. Set-up (management team/board):
2. How are decisions taken and implemented?

II Human resources management

1. Who is in charge?
2. Does the staff receive salary/incentives?
3. Total number of personnel/breakdown by function (MD, medical assistants, nurses, students, etc.):
4. Is there a roster system in place in the hospital?

III Financial management

1. Management of finance (Is there a budget? How is the hospital financed?):
2. Is there any cost participation, "cost-recovery system"? Do the destitute have access to care?

IV Statistics

1. Management of statistics and reporting:
2. Are statistics available?
3. Is there an annual report?
4. Are there people specifically in charge of collecting data?

V **Infrastructure & utilities** (general condition of):

1. Wall and roof:
2. Water (running water, wells, safety of water supply, etc.):
3. Sanitation (type of toilets, etc.):
4. Electricity and/or generator (number of hours per day, fuel supply, etc.):
5. Heating/fans/air-conditioning:
6. Maintenance team (number, composition, etc.). Is there a maintenance schedule?
7. Is there a functioning workshop?

VI Waste disposal

1. Waste management systems (including toxics such as X-ray developer/fixator, etc.):
2. Incinerator (type, condition, etc.):

VII Non-medical support services

1. Kitchen (staff, nutritionist, origin of food, number of meals served per day, special diets, etc.):
2. Laundry (staff, washing by hand, machine, supplies, etc.):
3. Tailor (staff, supplies, etc.):
4. Cleaning and hygiene (system, staff, supplies, etc.):
5. Morgue (infrastructure, management, etc.):

Medical support services

I Pharmacy

1. Pharmacy staff and management:
2. Is there a standard list of medicines?
3. Are stock cards used?
4. Where do the drugs and medical equipment come from (regular supplier, local market, donations, etc.)?
5. Is there a reliable system of communication between the pharmacy and the wards (request forms, delivery forms, etc.)?
6. Did the pharmacy run out of basic drugs last month (penicillin, anti-malaria, paracetamol, ORS)?
7. What are the storage conditions (air-conditioning, refrigerator, etc.)?
8. Is medical equipment regularly maintained and serviced?

II Laboratory

1. Laboratory staff and management:
2. Tests available (haematology, chemistry, parasitology, bacteriology, serology, etc.):
3. Source of supplies:
4. Is there a reliable system of communication between the laboratory and the wards (request and results forms)?
5. Quality of the working relationship between the clinical and laboratory staff:

III Blood transfusion

1. Staff and management:
2. Policy of blood sampling and transfusion: HIV/AIDS policy?
3. Indications for blood transfusion/average number of requests:
4. How are the blood units kept? Is there a functioning refrigerator to store the blood?
5. Testing process and quality control:

IV Imaging (X-ray & ultrasound)

1. Staff and management:
2. Average number of X-rays per day:
3. Type and quality of machine:
4. Are there guidelines for the prescription of X-rays?
5. Is more sophisticated imaging equipment available?

V Other diagnostic services

1. ECG, EEG, etc.:

Clinical services

I Outpatient department (OPD)

1. Role of the OPD (consultation, follow-up of patients, admission, emergency):
2. Are there specialized OPDs?
3. Are there criteria to admit the patient to the OPD?
4. Is there a register with data about all patients seen every day?
5. Average number of cases seen every day (medicine, paediatric, surgery, obstetric, etc.):
6. Personnel in charge (MD, medical assistants, nurses):
7. Is there a clear roster?
8. Opening days and hours:
9. Main pathologies:
10. Access to laboratory and imaging:

II Admission/emergency department

1. Number of beds:
2. Is there a team on duty 24 hours a day; composition of the team?
3. Is there an on-call system in place?
4. Is there an admission book or regular procedure for admitting and registering patients?
5. Is there a regular procedure to send patients to the appropriate wards or to the OT?
6. Number and type of emergencies per day:
7. Are basic supplies and equipment available?

III Operating theatre (OT) and anaesthesia

1. Staff and roster:
2. Hygiene of the OT:
3. Is there an accurate operation book? If yes, number of surgical operations in the last month:
4. What kind of surgery is performed?
5. What kind of instruments and sets are available (laparotomy, caesarean section, debridement, skeletal traction sets, etc.)?
6. Number of operating rooms, tables:
7. Surgical linen (availability and source of provision):
8. Functioning surgical equipment (lamps, suction, diathermy, oxygen, etc.):
9. Source of materials and consumables:

IV Sterilization

1. Staff and roster:
2. Equipment (autoclaves, dry ovens):
3. Protocols in place?

V Anaesthesia

1. Staff and roster (MD and/or anaesthetist nurses):
2. Is a laparotomy performed safely with full muscle relaxation (including endotracheal intubation) by a trained anaesthetist?

3. Common anaesthesia (gas, ketamine, spinal, local):
4. Type of anaesthesia machines:
5. Availability of other equipment (pulse oxymeters, oxygen supply, etc.):

VI Nursing care

1. Is there 24-hour nursing supervision in the wards?
2. Are the patient records complete?
3. Is the nursing handover book used properly?
4. Are drugs administered on time?
5. Is a laparotomy performed safely with the patient supervised (vital signs) for 24 hours post-operatively in a room with light, and where he or she receives intravenous fluids and antibiotics?
6. What do dressings look like (clean, smelly, etc.)?
7. Are bedsores a problem?
8. Are relatives involved in patient care?

VII Frequently asked questions

1. Availability of mosquito nets for all beds:
2. Is there an admission book or regular procedure for admitting and registering patients in the ward? If yes, number of admissions to the ward in the last month?
3. Is there a person in charge of the admissions who controls a system whereby the patients are assessed and then go to the OT or receive treatment?
4. Are new admissions systematically seen by a senior surgeon/MD and within what timeframe?
5. Are there regular rounds in the wards and/or regular meetings to discuss the cases?
6. Are the diagnosis and treatment clearly formulated in the patients' files and the treatment copied onto the patients' charts?

VIII Surgical care

1. Main pathologies present in the wards (fractures, burns, chest, abdomen, etc.):
2. Management of the ward/hygiene:
3. Human resources (number, composition, roster):
4. Infrastructure and beds:
5. Is a laparotomy performed safely – patients seen a few days after operation with healing wound and eating normally?
6. Can five or more laparotomies be performed in 24 hours under good conditions including anaesthesia? If not, why?
7. What type of orthopaedic treatment is present in the surgical wards (POP, skeletal traction, external or internal fixation)?
8. What do the wounds present in the ward look like (clean, dirty, smelly, pus)?

IX Physiotherapy unit

1. Are patients walking on crutches in the wards? If not, why?
2. Management of physiotherapy department:
3. Human resources:

Further comments

1. Particularities of the context:
2. Do you have any requests for the ICRC (logic and rationality of the request)?

Conclusion

1. First general impression (cleanliness and hygiene, staff present, presence of patients):
2. Main positive findings:
3. Main negative findings:
4. Capacity to cope with mass influx of wounded:
5. Emergency/contingency plan:
6. Proposals:
7. Next step:

Please note:

This sample form may be photocopied and is also available as a downloadable document on the accompanying DVD.

ANNEX 6. B Strategic assessment of a conflict scenario

The main appraisals in a strategic assessment of a conflict situation in order to determine some of the factors influencing the chain of casualty care are the following:

1. Geography:
   a. topography of the conflict area
   b. routes of communication and transportation
   c. distribution of medical facilities available and their safety
2. Where is the fighting taking place? Safe areas, dangerous areas?
3. Where do patients come from?
4. How many wounded are there?
5. Who are the wounded?
   a. trained soldiers in a regular army
   b. guerrilla fighters, militiamen
   c. civilians
6. Who performs first aid, if anyone?
   a. National Red Cross/Red Crescent Society volunteers
   b. military services
   c. trained civilians (RC/RC community-based first aid, MoPH)
   d. untrained civilians
   e. non-governmental organizations
7. Assessment of efficiency of first-aid system
8. How are the wounded transported from point of wounding to hospital?
   a. private means
   b. public transportation
   c. ambulance service
   d. military services: air, land, etc.
9. Assessment of efficiency of evacuation system
10. Which hospitals receive the wounded ?
11. Assessment of work performed in hospitals and their capacity to receive and treat patients (see Annex 6. A: Initial assessment of a surgical hospital treating the war-wounded).

12. Which other actors are present in the area?
    a. National Red Cross/Red Crescent Society
    b. ministry of public health
    c. military medical services
    d. national non-governmental organizations
    e. missionary clinics or hospitals
    f. foreign non-governmental organizations

On the basis of the above assessment, the military or civilian health services or National Red Cross/Red Crescent Society should take action to fill in any gaps by creating the following facilities.

1. System of first-aid posts.

2. Transport system.

3. Reliable surgical units: central, regional, zonal, local; alternative sites for hospitals; convalescence houses.

The above-mentioned institutions have several means of intervention to face the challenge of establishing an efficient chain of casualty care.

1. Negotiation with the various belligerents to make sure international humanitarian law is respected in order to ensure that:
    a. medical personnel has access to the wounded and sick;
    b. the wounded and sick can reach medical care;
    c. first-aid and health personnel and structures are safe.

2. Support to existing health structures, which may take the form of infrastructure renovation, equipment, medical supplies, or re-enforcement with human resources.

3. Mobilization of local infrastructure and human resources to improve the chain of casualty care or project forward medical care for the wounded.

4. Mobilization of international agencies to supplement national efforts.

Please note:

Points 1 – 12 of this sample form are available as a downloadable document on the accompanying DVD.

ANNEX 6. C Humanitarian intervention for the wounded and sick: typical settings

Many factors affect the deployment of humanitarian medical teams: this Annex helps to analyse a number of them. The terminology used is an operational description and has no legal significance.

Possible scenarios

1. Military activity, natural disaster or major accident? (Is health infrastructure intact?)
2. Military context: classical front lines, guerrilla war, internal troubles, post-conflict (particularly the presence of landmines, cluster bombs and other unexploded ordnance)
3. Urban or rural setting?
4. Industrially-developed or low-income country: funds available?
5. Human resources: trained doctors, nurses, and first aiders in sufficient, limited numbers, or very few?

Three typical general settings for a humanitarian intervention exist.

1. **Optimal situation:**
   adequate access to medical care in spite of the violence and conflict.
2. **Austere situation:**
   poverty before the conflict has already jeopardized access to care.
3. **Dire situation:**
   very poor access to care because of the violence and conflict, sometimes compounded with pre-existing poverty.

	OPTIMAL	AUSTERE	DIRE
Location	Urban in a developed country	Poor rural area	Developing country Major destruction
Duration of trouble	Single, isolated event (e.g. act of terrorism)	Ongoing low-intensity fighting (e.g. guerrilla warfare)	Continuous heavy fighting and/or bombardment
Casualty flow	Small irregular numbers (compared to the population of the city)	Discontinuous/ intermittent flow, including massive influx	Continuous but unpredictable flow, including massive influx
Infrastructure (roads, ambulance service, health facilities)	Intact and functioning	Poor or irregular (e.g. few good roads, hospitals lacking supplies)	Severely damaged or dysfunctional (roads damaged, debris in streets, hospitals looted, etc.)
Communications	Good	Poor to moderate Irregular	None or poor
Personnel	Adequate (in number and skills)	Variable	Minimum available or complete lack
Materials and supplies	Adequate (in quantity and quality)	Irregular and inadequate	Irregular supply or non-existent
Environment	Good (daytime, good weather)	Harsh	Bleak (night, cold, heat, etc.)
Evacuation	Safe and short	Predictable but long and arduous	Uncertain or unknown
Destination of transfers	Known and available	Known but variable	Unknown or absent

Types of situations of armed conflict and violence and their effects on humanitarian medical work[3]

Example	International armed conflict	Internal armed conflict/ guerrilla warfare	Civil disturbance/revolt	Widespread banditry and other crime
Description	A straightforward war between country X, and its allies, and country Y, and its allies	Intense fighting within one country	Unpredictable; hit and run skirmishes, often with a vested interest in the continuation of anarchy	May coincide with any of the other situations
Combatants/ Fighters	Easily visible, and wearing distinctive uniforms	Not all wear uniforms Government forces opposing well-organized military groups	Armed individuals, gangs, bandits, and militia groups	Purely self-interested individuals or groups
Front lines	Well known	May not exist, or may change very quickly	Linked to constantly shifting alliances among or between forces	At the fringes of the conflict, combatants ready to take advantage of circumstances
Chain of command	Structured and with available points of contact	Tenuous points of contact on the opposing sides.	Not clear, and variable from one faction to another (often based on an individual leader surrounded by a small group and supported by part of the population)	Traditional and local leader, personalized (e.g. street gangs)
Respect for IHL	Parties aware of their obligations and try to meet them	A certain degree of respect	Very little, with a complete breakdown of law and order	IHL unknown or no regard for, or both
Humanitarian tasks	Classic	Classic	Extremely difficult	?
Risk level	Low	Growing and less predictable	Very high, perhaps bordering on the unacceptable	Very real and very dangerous threat
Obstacles to medical work	Few, if any	More restrictions, negotiations, controls and delays, etc.	Severely restricted ability to intervene – vehicles, radios, goods, etc. are extremely attractive to the warring parties	Numerous: greatest caution required

Model scenarios

In contemporary armed conflict, health staff and structures are called upon to function in various scenarios: typical ones include the following.

1. Safe urban setting

- Urban, developed environment
- Single, isolated event
- Casualty numbers relatively small compared to population of city
- Infrastructure intact: roads, emergency vehicles
- Health infrastructure intact, sophisticated hospitals
- Short evacuation time: route is secure
- Good communications
- Personnel: adequate number and quality of trained health staff
- Materials adequate
- Environment good: weather, daytime

3 The terminology used is an operational description, not a legal one.

- Final destination of the wounded is known

2. Unsafe urban setting

- Low-income country: under-developed or destroyed urban setting
- Continuing danger: street fighting and bombardment in city
- Continuing and unpredictable casualty flow including massive influx of wounded
- Poor infrastructure: potholed roads, debris in streets
- Disrupted health infrastructure: hospitals damaged or looted
- Availability and length of evacuation uncertain or unknown
- No or poor communications
- Minimum health personnel available
- Material re-supply uncertain, irregular, or non-existent
- Environment poor: cold, wet, dark
- Final destination of the wounded not always obvious

3. Unsafe rural setting

- Low-income country: under-developed rural area neglected in peacetime
- Constant danger: ongoing combat, landmines
- Continuing and unpredictable casualty flow
- Poor infrastructure: badly maintained or no roads
- Poor health infrastructure: few health posts, even fewer district hospitals
- Availability and length of evacuation uncertain, long and arduous
- No or poor communications
- Minimum health personnel available
- Material re-supply uncertain, irregular, or non-existent
- Environment poor: extreme cold or heat, rainy season and dry season
- Final destination of the wounded not always obvious

4. Safe but austere setting

- Low-income country
- Continuing danger: ongoing low-intensity warfare
- Discontinuous casualty flow; includes irregular mass evacuations
- Poor infrastructure: few good roads and few vehicles
- Minimum of health infrastructure: some rural clinics or health centres, fewer district hospitals
- Evacuation predictable but long and arduous
- Poor to moderate communications
- Minimum to moderate number of health personnel available
- Minimum material re-supply
- Environment harsh
- Final destination of the wounded: distant, but known

Chapter 7

FIRST AID IN ARMED CONFLICT

7. FIRST AID IN ARMED CONFLICT

7.1 First aid: its crucial importance 155

7.2 First aid in the chain of casualty care 156

7.3 First aiders: an important human resource 157

7.4 Essential elements of the first-aid approach and techniques 157
7.4.1 Security first and always 157
7.4.2 Basic actions 158

7.5 Setting up a first-aid post 159
7.5.1 Location 159
7.5.2 Infrastructure 159
7.5.3 Staffing, equipment and supplies 160
7.5.4 Organization 160

7.6 ICRC involvement in first-aid programmes 161

7.7 Debates, controversies and misunderstandings 161
7.7.1 "Scoop-and-run" *versus* "stay-and-treat" 161
7.7.2 Mechanism of injury and the problem of the cervical spine 162
7.7.3 ABCDE or C-ABCDE? 163
7.7.4 The tourniquet: when and how? 163
7.7.5 Resuscitation or rehydration? 164
7.7.6 Supplemental oxygen in the field? 164
7.7.7 Disability: the AVPU system 165
7.7.8 Evacuation: a risk to take 165

7.1 First aid: its crucial importance

Optimal management of the injured involves a continuum of care from the point of wounding to the surgical hospital: the chain of casualty care. The surgeon wishes to receive patients who arrive in good condition and well-stabilized, in a timely manner, and according to priority for treatment. To understand how this is achieved, or not, there are a number of things that need to be known about first aid. In addition, the surgeon working in a conflict area may well be called upon to participate in the training of first aiders to achieve a more efficient chain of casualty care.

First aid is the initial assistance given to an injured or sick person until the condition of the person is stabilized or remedied, or professional medical help is made available. How first aid is applied varies according to the security environment, the number and condition of the wounded in a particular location, the resources that can be mobilized for care, transport, and access to surgical hospitals and the latter's capacity to receive and treat patients.

Since its foundation, the International Red Cross and Red Crescent Movement has been, and continues to be, predominantly associated with the provision of first aid. It pioneered the concept of immediate response by first aiders and local communities to the consequences of war, disasters and epidemics.

The goals of a first aider in action are:
- to preserve life through the support of vital functions;
- to intervene securely and safely;
- to limit the effect of injury and to prevent further injury;
- to prevent complications and disability;
- to alleviate suffering and to provide psychological/moral support;
- to promote recovery;
- to ensure proper handover of the injured and sick to healthcare professionals when needed.

First aiders can also help mobilize their community to prepare for and respond to emergencies that happen in everyday life and during crises such as armed conflict.

Experience has shown that one of the most important factors determining the outcome of the management of the war-wounded is the pre-hospital phase. First aid saves lives and decreases morbidity. It also results in easier surgery, thus decreasing the hospital surgical burden. Furthermore, up to 40 – 60 % of the civilians and soldiers wounded during armed conflict do not require hospitalization. First aid measures – plus a simple oral antibiotic and analgesic – constitute all the treatment they need. Military terminology calls them "returned to duty" (see Chapter 5).

> First aid, begun early, saves lives and can prevent many complications and much disability.

The provision of first aid is one of the fundamental responsibilities of military medical services, National Red Cross and Red Crescent Societies and, increasingly in contemporary armed conflict, of medical staff in both urban and rural public hospitals. The essential role of local communities in providing assistance on the spot should not be forgotten, as Henry Dunant, founder of the Red Cross and inspiration behind the Geneva Conventions, witnessed after the battle of Solferino on 24 June 1859.[1]

1 See Henry Dunant. *A Memory of Solferino*. ICRC: Geneva; 1986.

Consequently, basic first aid, initial and refresher training, should be supported among:
- the general population;
- soldiers and members of security forces;
- community healthcare workers;
- military and other medical staff (nurses, doctors and surgeons).

Commanders do not appreciate seeing their troop levels decrease when healthy uninjured soldiers are involved in transferring their own wounded because first-aid services are inadequate in the field. The fighting capacity is then doubly reduced.

Advanced competencies can be added to the training curriculum of those specially dedicated to providing first-aid services in the field, such as military medics and the first-aid teams of the National Red Cross and Red Crescent Societies, as a forward projection of resources.

7.2 First aid in the chain of casualty care

First aid starts at the point of wounding, but can be given anywhere and everywhere along the chain of casualty care to the place of definitive treatment.

Point of wounding
On-the-spot first aid, performed in the field and often on the actual battlefield: this may be self- or "buddy-" first aid if combatants have received the proper training prior to deployment. Otherwise, it is practised by a military medic, civilian or Red Cross/Red Crescent first aider.

Collection point
It is common practice and convenient to bring all the wounded to one spot, depending on the tactical situation, to evaluate their condition, start first aid if it has not yet been given, and stabilize those for whom lifesaving measures have already been undertaken; and then to decide who needs to be evacuated for further treatment according to triage priorities. This is best served by a first-aid post (FAP).

Evacuation
The decision to transport a casualty should be assessed carefully considering the dangers and difficulties inherent in situations of armed conflict. Whatever method of transport is used along the chain of casualty care, first-aid measures should be maintained throughout.

Hospital emergency room
In the rural areas of a poor country and during urban warfare, the first site where any care is available is often the emergency reception of an established hospital. Even if there is an efficient emergency transport service, family and neighbours often do not wait for it and prefer to transport the wounded directly to the hospital, whose emergency room (ER) then serves as a first-aid post.

The degree of development and sophistication of the emergency transport and emergency medical services in a given country will determine what level of first aid and triage are performed "in the field" and what occurs only in a healthcare facility.

First aid can be performed everywhere and anywhere along the chain of casualty care.

Even in armed conflicts, ordinary life goes on. There is no let-up in road traffic crashes, accidents, or diseases among the local population, as well as combatants. The efficient work of first aiders is as necessary as usual.

7.3 First aiders: an important human resource

Usually first aiders are organized in teams with proper supervision and equipment, such as those of the Red Cross or Red Crescent or of the military. Civilian and Red Cross/Red Crescent first aiders are an essential part of the healthcare team given that they are members of the local community, reflect its characteristics, and are well accepted by society. They take on many roles from the front lines to the healthcare facility; their availability and versatility are respected.

Therefore, it is important to respect their knowledge, and appreciate their courage and dedication. They have important rights and duties according to international humanitarian law, and should be trained accordingly. They should also be trained in the methods and practice of triage. Not only is it important to welcome them when they bring in a casualty but also to give them feedback on the measures they have taken and on the evolution of the casualties they have already brought in, to ensure a proper handover of the patient and prepare for future work. The surgeon plays a vital role in this communication with first aiders.

First aiders are often volunteers and an essential part of the healthcare team, and should be appreciated accordingly.

The presence of first aiders before, during and after emergencies helps rekindle the humanitarian spirit of individuals and communities, inspiring tolerance and ultimately building healthier and safer living environments.

7.4 Essential elements of the first-aid approach and techniques

7.4.1 Security first and always

One of the specific characteristics of armed conflict, as mentioned, is the particularly hazardous situations encountered, more dangerous than during natural catastrophes or industrial accidents. The situation is made even worse because of the will of combatants to continue fighting and inflict harm after the initial damage has been done, and the increasing unwillingness of many combatants to recognize and obey the rules of armed conflict.

First aiders are exposed and at great risk, or are at times even targeted in an ambush or a gun battle. All too often, when a bomb explodes or is dropped there is a rush of first-aid rescuers to the site; a second bomb goes off a short time later causing far more casualties than the first. Furthermore, the work space may be limited by an excited and angry crowd of bystanders and the casualties' friends, relatives and comrades, who may threaten first aiders.

The point of wounding is, by definition, a very dangerous and chaotic place.

This entails practical consequences for the first aider and imposes necessary constraints for the safety of patient and first aider alike. A wounded first aider needs help and cannot help others. The question of first-aid care under fire – for military personnel – will depend on the doctrine and training of each army.

As previously stated, the sick and wounded benefit from protection and the first aider has rights and responsibilities according to international humanitarian law. An opportunity to provide care more safely may be created by negotiations or a ceasefire. The military may choose to remove the threat or provide security by armed force. This does not mean first aiders need not take security measures prior to engagement in the field. Circumstances determine how and when care can be provided with acceptable risk.

The benefits of caring for the wounded must always be weighed against the risk of first aiders themselves being wounded.

An absence or limitation of transportation and the impossibility of deploying personnel in the field to perform first aid because of political or security constraints – all too often the case in the circumstances where the ICRC works – will cause some patients who might otherwise have survived to die, and surviving patients to have established infections and other complications which lower their subsequent chances of quality survival. The challenge is how to deploy and organize the necessary first-aid personnel, material, and equipment in the field. For the military, this is also a challenge but has a different connotation. Tactical considerations may mean that soldiers must "win" the firefight before treating or evacuating the wounded.

7.4.2 Basic actions

The basics of a routine first-aid deployment apply, while at the same time safety and security *must be* ensured.

Preventing further injury: removing casualties from danger
A wounded person is more likely to be injured again or even killed, especially if the person cannot take measures of self-protection, such as seeking cover from shooting or bombing. The prevention of further injury thus involves removing casualties from the point of wounding, and putting patients and medical personnel in as safe a place as circumstances allow. Furthermore, conscious and walking wounded must be separated from the others as a matter of scene security management, as well as triage.

Preventing aggravation of the injury: sheltering the casualties
Shelter offers some protection from further injury in a combat environment, and also provides a shield from the elements. Exposure to the sun and heat, or cold and rain, are deleterious to the wellbeing and stabilization of the injured. It is also easier and more efficient to deliver medical care when working under more comfortable physical conditions.

Performing the standard sequence of basic life support

A = Airway

B = Breathing

C = Circulation

D = Disability (neurological status)

E = Environment and Exposure

1. ABCDE sequence leading to immediate life-saving measures.
2. Complete examination leading to stabilization measures.
3. Additional measures:
   - keeping the casualty warm;
   - ensuring proper rehydration;
   - providing psychological support.
4. Monitoring of the casualty's condition and the efficiency of the measures taken.

Protection of the casualty, prevention of further injury, and lifesaving and stabilization measures are the fundamental responsibilities of the first aider in the field.

The analysis of the trimodal distribution of deaths in Chapter 5 demonstrated that first-aid measures in the first instance should specifically target those injuries that are severe but potentially survivable. Thus, the primary objectives of immediate life-saving measures are the following.

1. Establish and maintain an adequate open airway.
2. Maintain breathing.
3. Maintain circulation by controlling peripheral haemorrhage and reducing or preventing shock.

In the second instance, first-aid measures target the great majority of casualties: those who suffer fractures and soft-tissue injuries, which may cause disability.

The patient must be exposed in order to perform a correct and complete examination. Cultural and religious constraints, as well as the tactical situation, may impose limits as to how much a casualty may be undressed in the field.

All injured persons who have bled lose body heat, even in the tropics. It is important to avoid hypothermia, which can have disastrous results later in terms of coagulopathy (see Chapter 18). Casualties should be covered with a simple sheet or blanket, depending on the prevailing climate, without forgetting to put something *under* the patient since much body heat is dissipated into the ground.

National and institutional guidelines will establish the protocols for the level of treatment (i.v. fluids, antibiotics, analgesia, endotracheal intubation, etc.) that are implemented in any first-aid programme.

Simple life-saving procedures in the field – immediate life-saving measures – take precedence over more complex techniques and are based on the well-known priorities of the ABCDE system.

7.5 Setting up a first-aid post

The establishment and organization of a first-aid post should rely on a large dose of common sense to determine what is practical and realistic in a given situation, and will depend on how long it will operate (from a few minutes to a few days or weeks).

7.5.1 Location

Choosing a site for an FAP should follow certain rules. It should be placed in a secure position, far enough from the fighting not to be in danger, yet near enough to enable the rapid transfer of casualties to the post. For operational and security reasons, its location should be indicated as soon as possible to the dispatch or command centre of the chain of casualty care. Its presence should be made known to the local population and combatants, who are the potential beneficiaries of its services. A prominently displayed distinctive emblem (a red cross, red crescent or red crystal) large enough to be seen from all directions and from as far as possible helps to identify the FAP and provide protection under IHL.

Security and protection of casualties and first aiders are the primary considerations for setting up a first-aid post.

7.5.2 Infrastructure

An FAP is a functional unit: it can therefore be set up in a makeshift fashion under a tent, in a school, or any available building; or in an already existing dispensary or primary healthcare centre. Certain minimum requirements should be fulfilled: appropriate shelter against the elements; adequate size to accommodate casualties on stretchers; easy access for the "walking wounded" (e.g. avoid long stairways); and adequate in-and-out access for ambulances, and ample parking space.

ICRC

Figure 7.1.1
First-aid post: formal.

T. Gassmann / ICRC

Figure 7.1.2
First-aid post: ad hoc.

7.5.3 Staffing, equipment and supplies

The level of technical expertise of the personnel in an FAP will depend on the circumstances and standards of the country. Anyone from a first aider to a nurse, general practitioner or even a surgeon may be found working in an FAP. This allows for the "forward projection" of care for casualties (see Chapter 6). Equipment and supplies must meet minimum standards and be adequate for basic trauma care, and adapted to the personnel's skills and local standards. Telecommunication equipment, if any, should be properly installed and placed in a reserved area.

Please note:

The Emergency Items Catalogue of the International Red Cross and Red Crescent Movement contains the description of a standard first-aid and triage post, equipped for experienced nursing staff and/or a general practitioner (see Selected bibliography).

7.5.4 Organization

The premises of the FAP should be organized and the staff prepared for dealing with a mass influx of casualties. The principles and practice of triage should be well understood by all present through simulation exercises (see Chapter 9).

If the FAP stays open for a certain length of time, and the facilities allow for it, the following areas should be pre-arranged:
- admission area at the entrance to register and triage casualties;
- holding area to care for and monitor casualties awaiting evacuation;
- waiting area for relatives, friends and comrades;
- temporary mortuary;
- storage area for equipment and supplies;
- storage area for weapons removed from the wounded;
- rest area for personnel, and personal hygiene facilities.

ICRC EXPERIENCE

Occasionally, the ICRC has had to set up an ad hoc FAP where access to surgical facilities was insecure. During the long-running civil war in Sierra Leone, fighting broke out several times in the capital city, Freetown. During one such episode in 1998, the ICRC delegation was turned into a first-aid post: the parking lot became the reception and triage department and the various offices became wards and intensive care areas. Over a period of 10 days, 6 nurses and 3 Red Cross volunteers treated 244 war casualties and 228 sick patients, while a surgical team from Médecins sans Frontières (MSF) operated in the general hospital. Urban fighting often made it dangerous for patients to reach the hospital and they regularly stayed in the improvised FAP for 2 or 3 days prior to evacuation.

In addition to basic first-aid treatment, intravenous fluid resuscitation, tetanus immunoglobulin, penicillin, and analgesia were given. Many patients survived 2 - 4 days, before being transferred to the hospital, and surgery. The MSF surgical team appreciated receiving patients who had dressed wounds and were "well hydrated", ready for surgery.

Security, however, was a major concern. Looting of the ICRC delegation by armed men occurred twice and several of the 4,500 civilians, who had taken refuge in the delegation premises, effectively turning it into a displaced persons camp, were injured or killed by stray bullets. The emblem of the red cross offers only limited protection under such circumstances.

7.6 ICRC involvement in first-aid programmes

The ICRC supports first-aid programmes and activities in contexts of armed conflict and other situations of violence. These programmes include:

- deployment of medical staff to operate first-aid posts;
- training where no other organization has access (e.g. for irregular guerrilla groups);
- helping to adapt routine first-aid programmes of National Red Cross and Red Crescent Societies and other organizations to the context of conflict preparedness and response;
- development of strategic, management and planning capacities of national and local institutions and team leaders;
- provision of materials and equipment;
- assistance in the organization of an emergency care system;
- production of standards and reference documents (see Selected bibliography).

7.7 Debates, controversies and misunderstandings

Detailed first-aid techniques are described in the manuals listed in the Selected bibliography. The rest of this chapter deals only with a few topics that may give rise to controversy or misunderstanding or are specific to the scenario of armed conflict.

7.7.1 "Scoop-and-run" *versus* "stay-and-treat"

Like the rest of the chain of casualty care, the organization of first aid requires planning and training. A strategy for the efficient care and evacuation of the wounded must be established. Two well-known strategies are described by the expressions "scoop-and-run" and "stay-and-treat". These derive largely from the civilian practice of emergency medical services and may not be particularly relevant in a context of armed conflict.

In a peacetime environment, there is no impediment to patient transfer other than infrastructure and geography. The efficiency of evacuation is determined by the availability and coordination of transport, either by air (fixed wing with

airports or helicopters and helipads), or by land (proper roads, necessary vehicles, etc.); and the distances involved. The context of armed conflict is not as simple.

Scoop-and-run implies basic first-aid measures, then taking the patient *immediately* to a surgical hospital. This is appropriate if the distances are short, reliable transport is available, and the nearby hospital is adequately supplied and staffed. Keeping patients on the spot until patient stabilization prior to evacuation – stay and treat – is correct if distances are great, infrastructure and transport are less than adequate, and proper surgical facilities are more difficult to access.

In a military context, the factors of ongoing combat and enemy activity must be added to those present during peacetime. Here, scoop-and-run implies air superiority if helicopters are used, safe and controlled roads for transport over land, and, in both situations, surgical facilities at a reasonable distance. These conditions are fulfilled for only a few armies in the world, and even then only under certain tactical situations. More common is the establishment of a collection point or clearing station: that is to say "stay-and-treat". How much "treatment" occurs depends on the forward projection of resources. The performance of more advanced techniques in the field has been discussed in Chapter 6.

7.7.2 Mechanism of injury and the problem of the cervical spine

The first aider, or other medical personnel, must immediately determine whether the casualty has suffered blunt or penetrating trauma. Blunt trauma above the level of the clavicles or causing unconsciousness requires immediate observation and immobilization of the cervical spine, while still giving priority to the airway. A penetrating wound to the brain causing unconsciousness requires no special care of the cervical spine.

In the case of a penetrating missile wound to the neck hitting the cervical vertebrae, any damage to the spinal cord has already been done. Unstable cervical spine injury from a projectile results in permanent cord injury. The reported mortality is over 95 %. The first aider should be careful while handling the casualty, but cannot prevent damage that is already irreversible. A combination of blunt and penetrating neck injuries should be managed in the same way as blunt ones.

M. Dalla Torre / ICRC

Figure 7.2

Gunshot wound to the mandible: no danger to the cervical spine.

ICRC EXPERIENCE

Here are two practical examples taken from ICRC experience to demonstrate the difference between blunt and penetrating trauma:

A person wounded in a road-traffic crash with a fractured mandible and bleeding from the mouth that obstructs the airway also requires care of the cervical spine in case there is hidden injury. A bullet wound to the mandible – with bone fragments, haematoma, drooling saliva, and damaged soft tissue of the floor of the mouth – does not. If the patient is conscious, he will maintain a particular position – head thrust forward and looking down, blood and saliva flowing out – in order to maintain as patent an airway as possible. There is no use in placing a protective cervical collar on this patient and, if the first aider attempts to place the casualty in a reclining position, the wounded person will resist.

A fall may result in both loss of consciousness and injury to the cervical spine. Loss of consciousness due to a bullet wound to the head, however, does not require any specific precautions regarding the cervical spine.

7.7.3 ABCDE or C-ABCDE?

In statistical terms, severe peripheral haemorrhage is the main lethal yet salvageable injury in a context of armed conflict (see Chapter 5). Various studies show that far more soldiers die from catastrophic peripheral haemorrhage than from airway obstruction or respiratory inadequacy. Thus, the "C" which stands for Catastrophic bleeding, may be put first, simply because it is more common.

Once catastrophic external haemorrhage, which is quite obvious, has been excluded – the great majority of cases – the first aider should resort immediately to the standard routine of ABCDE.

7.7.4 The tourniquet: when and how?

The tourniquet is said to save lives, but endanger limbs. It is difficult to apply correctly and to maintain and supervise, and open to many complications if widely used when not necessary. Furthermore, the prolonged application of an improvised tourniquet is usually not effective and causes tissue damage by itself (Figures 7.3.1 and 7.3.2). There is little or no use for a tourniquet in civilian practice for the war-wounded. It may be applied *as a last resort*, when all else has failed, to temporarily control bleeding while packing the wound and applying a compressive bandage. It should then be removed.

E. Winiger / ICRC

Figure 7.3.1

Improvised tourniquet: how effective?

R. Coupland / ICRC

Figure 7.3.2

Improvised tourniquet: how harmful? Here a tourniquet was applied for more than six hours, resulting in a very high above the knee amputation.

By far the most common and most effective technique for a first aider to control peripheral haemorrhage is direct pressure over the wound completed by a compressive bandage; or, if insufficient, pressure over the proximal artery in the axilla or groin, tamponade of the wound with dressings, and then a compressive bandage.

The military might continue to use a tourniquet under certain tactical circumstances: to free the hands of the single medic when dealing with mass casualties; while performing first aid under direct fire, and especially at night. Some armies equip their soldiers with a pre-manufactured self-applicable tourniquet applied with one hand to allow a wounded soldier to continue participating in combat. The ICRC declines to pronounce on these tactical considerations, but warns against an all too easy abuse of the tourniquet as has been seen in many wars in the past.

The use of a tourniquet for crush injuries (people trapped under the rubble of a collapsed building whether due to an earthquake or bombardment) is justifiable, and due precautions should be taken to ensure rehydration/resuscitation and thereby avoid the complications of the otherwise inevitable myoglobinaemia.

For further information on crush injury, see Volume 2.

New locally-active haemostatic powders and sponges to promote clotting in traumatic wounds are being tested by some armies; however direct pressure must still be applied. The ICRC has no experience with these and, like most others, eagerly awaits the published results of these field trials. Cost and availability will be factors to be taken into consideration for their use worldwide.

Peripheral haemorrhage can be controlled by external pressure and a compressive bandage.

7.7.5 Resuscitation or rehydration?

The first aider can only implement simple procedures in the field. The administration of intravenous fluids requires a certain degree of medical knowledge and specific equipment that are beyond the normal scope and competencies of a first aider. He can be involved in the monitoring of a perfusion, however.

The subject of pre-hospital i.v. fluids has given rise to considerable controversy. Maintaining adequate tissue perfusion and oxygenation is balanced off against the dilution of clotting factors and the dislodging of existing clots when the blood pressure goes up ("popping the clot"). The currently recommended "hypotensive resuscitation" involves giving fluids to a just palpable radial pulse – equivalent to a systolic blood pressure of 90 mm Hg – and no more (see Chapter 8).

A great deal of this controversy concerns efficient civilian trauma systems where evacuation times are short, less than 2 hours. It is probable that only the war-wounded showing obvious clinical signs of shock require pre-hospital i.v. fluids. In addition, in many contemporary conflicts transfer to a hospital takes much longer – days or even weeks – where "popping the clot" is not really relevant. Rehydration – as part of resuscitation – may be a more appropriate term under these circumstances.

If the wounded patient is conscious and not suffering from head trauma, the first aider can rehydrate with oral fluids (e.g. oral rehydration salts), especially if evacuation is delayed and transport long. This procedure is probably not employed nearly enough. Oral rehydration can probably be given with no detriment to the patient even in cases of abdominal trauma, and especially for severe burns.

7.7.6 Supplemental oxygen in the field?

Warning!

Oxygen cylinders are the equivalent of a bomb if hit by a bullet or piece of shrapnel. In addition to the danger they represent, the cylinders are heavy and must be replaced, lasting only a short time at high flows. Furthermore, their replenishment is complex and requires special factory capacity.

Oxygen cylinders must be ruled out in the event of any deployment to a dangerous area. This is now standard ICRC policy.

Palestine Red Crescent Society

Palestine Red Crescent Society

Figures 7.4.1 and 7.4.2

Ambulance destroyed by the detonation of an oxygen cylinder which has gone through the roof like a rocket. The photos show the ambulance and the cylinder on the ground.

Depending on security conditions, the collection point or intermediate station may have oxygen available. An oxygen concentrator (requiring an electrical supply) is preferable to compressed cylinders.

7.7.7 Disability: the AVPU system

Disability refers to the neurological status of the casualty: the state of consciousness and any paralysis due to injury to the spinal cord.

The Glasgow Coma Scale is the hospital standard. However, a simpler determination of consciousness is the AVPU system and it may be easier for first aiders to use in the field. It easily translates into the Glasgow system when the patient arrives at the hospital and gives the surgeon a good appreciation of the evolution of the patient's condition.

Alert	The casualty is awake, lucid, speaks normally and is responsive to the environment (e.g. eyes open spontaneously as you approach).
Voice responsive	The casualty is able to respond in a meaningful way when spoken to.
Pain responsive	The casualty does not respond to questions but moves or cries out in response to a painful stimulus (pinching the neck muscles, earlobe or nipple; rubbing the supra-orbital margin, or the angle of the mandible).
Unresponsive	The casualty does not respond to any stimuli.

Other conditions may influence the level of consciousness: hypoxia, shock, medication, and narcotics and alcohol (the latter all too often consumed to excess during times of armed conflict).

In the field, securing the airway is the first step to take when facing a casualty whose consciousness is compromised.

Examination of the spinal cord and vertebral column
This involves two phases: determining the movement of the limbs – comparing the two sides – and palpation of the bony prominences of the spine. Palpation of every vertebra – like "fingers playing on piano keys" – aims to identify any induration or deformity. If there is any danger to the spinal cord, then the four-person log-roll technique should be used to move the patient onto a stretcher. The cervical spine can be immobilized initially by manually controlling the head before a neck collar/ Minerva splint is placed. Immobilization of the cervical spine must not increase the risk of airway obstruction, however.

The importance of the mechanism of injury, blunt or penetrating, has already been mentioned. In all cases of suspected injury to the vertebral column, proper handling of the patient and immobilization techniques are of the utmost importance; the injury may already be there, but the first aider should act in such a way that he does not worsen the condition.

7.7.8 Evacuation: a risk to take

Transportation of the wounded is always difficult, always takes longer than expected, adds to the trauma, and can be dangerous to the patient and those transporting the casualty: the famous "mortality of the ambulance ride". In situations of armed conflict, the danger of the ambulance ride is complicated by the risks of ongoing combat.

Figures 7.5.1 and 7.5.2
Different means of patient evacuation: modern and traditional.

L. Berlemont / ICRC

Y. Muller / ICRC

Speed is less important than safety of transport: ambulance crashes are notorious. All these parameters, aggravation of the patient's condition, use of resources and security, must be weighed against the likely benefits of moving the wounded.

In an urban environment, it is normal for a first aider to bypass the neighbourhood clinic if the injury is serious, and to evacuate to a clinic for a minor wound in order not to overburden the hospitals. Unfortunately, city-dwellers often spontaneously rush patients to hospital no matter how slight the injury. Evacuation and transport may be difficult and dangerous during urban fighting even if distances are very short and transfer may be delayed while waiting for a calm period or ceasefire.

There is also the confusion that reigns during street fighting and the highly-charged emotional atmosphere to contend with. Lack of discipline to perform good pre-hospital triage – and the "pressure" of bystanders – can result in chaos in the receiving hospital. The first wave of ambulances transports the dead and badly mutilated. The second wave of ambulances carries people who are shouting, hysterical, and frightened: the lucid and superficially injured. The third wave evacuates the seriously wounded who truly require emergency treatment; these patients lie still – they are usually haemorrhaging – without shouting to bring attention to themselves.

In remote rural areas, if transportation takes many hours or even days, it is logical to project forward competencies by training local providers in more advanced skills. The principles of triage apply to the choice of which patients to evacuate first (see Chapter 9). Note that there is an important *difference* in the priority to treat and the priority to evacuate, especially if evacuation is prolonged. The recognition of injuries that are not survivable helps to spare the victim and would-be rescuers the agony and frustration of unsuccessful efforts to reach a higher level of care, and to provide better care for those who can survive.

Delay in evacuation will contribute to an increase in pre-hospital mortality; the more severely injured suffer "nature's triage". As was seen in Chapter 5, longer and more difficult evacuation sorts out central injuries, with a consequent decline in hospital mortality: only those patients with a good chance of survival reach the hospital. This remains a major problem in the care of the war-wounded.

Chapter 8

HOSPITAL EMERGENCY ROOM CARE

8. HOSPITAL EMERGENCY ROOM CARE

8.1 ABCDE priorities 171

8.2 Initial examination 171

8.3 Airway 172
8.3.1 Opening the airway 173
8.3.2 Definitive patent airway 173
8.3.3 Definitive patent airway: endotracheal intubation 173
8.3.4 Definitive patent airway: surgical airway 174

8.4 Breathing and ventilation 175
8.4.1 Assisted ventilation 176

8.5 Circulation 177
8.5.1 Pericardial tamponade and pericardiocentesis 177
8.5.2 Haemorrhagic shock 178
8.5.3 Fluid replacement 179
8.5.4 Hypotensive resuscitation 181
8.5.5 Adjuvant therapy 181

8.6 Blood transfusion where supplies are limited 181
8.6.1 Clinical use of blood in ICRC practice 182
8.6.2 Two unit rule 182
8.6.3 Fresh whole blood 183
8.6.4 Autotransfusion 183
8.6.5 Total blood requirements: the ICRC's experience 183

8.7 Disability 184

8.8 Environment/exposure 185
8.8.1 "Fatal triad" of coagulopathy, acidosis, and hypothermia 185

8.9 Complete examination 185
8.9.1 Complementary diagnostic examinations and monitoring 187

8.1 ABCDE priorities

The framework for casualty management in a hospital emergency room (ER) is a continuation of the basic life support procedures of first aid. The logic is the same; the diagnostic and treatment means available are simply more advanced.

Whether or not the casualty arriving at a hospital has received first aid in the field, he/she should undergo the full ABCDE examination. The patient's condition may have changed during transport; important wounds may have been missed because of the confusion in the field; first-aid competence may have been limited or entirely absent. Efforts should again first concentrate on lifesaving treatment for asphyxia and shock, the most common remediable causes of death. The "golden hour" begins at the site of trauma, *not* on arrival at the emergency room.

Priorities: the **ABCDE** scheme of thinking

Airway before Breathing, before Circulation.

1. Assess:
Initial examination: **A**irway, **B**reathing, **C**irculation, **D**isability, **E**nvironment and **E**xposure. Triage in a mass casualty situation: see Chapter 9.

2. Act:
Emergency resuscitation: act on life-threatening conditions.

3. Assess:
Complete examination: head-to-toe palpation, front and back and sides.

4. Act:
Definitive treatment, surgical or not: stabilization.

5. Assess and Act:
Treatment or evacuation of the casualty, according to triage priority, to a higher echelon hospital for specialized care if necessary.

Please note:

In certain countries a very definite separation must be made for emergency examination and treatment of male and female patients. This can be problematic in a mass casualty situation; nonetheless, the emergency department must be organized accordingly.

8.2 Initial examination

The initial examination and emergency resuscitation are carried out simultaneously. The receiving ER doctor must automatically ask a series of questions.

1. Is the patient dead or alive?
2. Is the casualty conscious or not?
3. What is the mechanism of injury: penetrating or blunt?
4. What are the life-threatening conditions, if any, according to the ABCDE scheme?

The natural reflex when faced with a wounded person is to look at the bleeding first. It must be understood that the greatest threat to life is an inadequate intake of air. Airway takes precedence over breathing and over circulation. With practice, the airway and breathing can be secured quickly enough to then allow the doctor to deal with any visible haemorrhage. (The situation in a hospital emergency room is different from one of "catastrophic haemorrhage" in the field, as discussed in Chapter 7 on first aid.)

The ER doctor must learn to use the ABCDE scheme of thinking in a circular fashion. Determining if the patient is dead or alive already involves a rapid ABCDE

examination. The dead have no air intake (A), no ventilation of the lungs (B), and no pulse (C); the pupils are dilated and do not react to light and there is no reaction to painful stimulus (D); there is no movement of the limbs and the body feels cold (E).

When mastered, the ABCDE scheme of thinking allows the doctor to answer all the above questions in one integrated process.

Most casualties are conscious, anxious or frightened, and in pain, and will tell you what happened and where "it hurts": they are alive, conscious, and the act of speaking shows that the airway is free. These signs of life may seem obvious. Nonetheless, one should make a mental checklist and practice the routine of asking the series of initial questions in order to establish an organized and coherent method of examination.

As mentioned in the previous chapter on first aid, blunt trauma above the level of the clavicles requires adequate care of the cervical spine, but not at the expense of a life-threatening airway problem. Simple measures include: manual stabilization in a neutral position with the head maintained in the axial line; semi-rigid collar; sandbags; taping; specialized backboard.

The initial examination should include the diagnosis of any life-threatening A, B, or C problems and the number, location and extent of *obvious* wounds. These two elements together are important for establishing triage priorities (see Chapter 9). A more complete examination to locate *all* wounds should follow the initial ABCDE examination.

It is important to determine the time since injury, estimate blood loss since injury, and the presence or absence of any allergies. In civilian casualties, past or pre-existing diseases and medications for chronic conditions should be taken into account. Young healthy combatants usually do not present such medical problems.

In a hospital environment the casualty should be completely undressed to allow for a correct examination – again, certain cultural or religious constraints may apply. Vital signs should be taken, i.v. lines placed and blood samples withdrawn simultaneously for blood grouping and cross-matching, and a haematocrit or haemoglobin estimation to determine a baseline for future comparison. Other tests, such as electrolytes, blood gases, glucose, etc. may be performed depending on local standards and laboratory capacity.

8.3 Airway

Airway obstruction is an emergency that requires an immediate response. Some injuries cause an immediate problem; others may result in delayed impairment of the airway.

The following list presents the common conditions that cause airway problems.

Head injury with decreased consciousness:
Apart from the danger of aspiration of vomitus, blood, broken teeth, bone or foreign bodies, the tongue and epiglottis may fall back and block the airway. Furthermore, the initially conscious casualty may slowly lose consciousness after some time.

Maxillo-facial injuries:
Even if the patient can breathe adequately in the beginning, the development of oedema of the tongue, floor of the mouth and pharynx will obstruct air intake after several hours.

Penetrating wounds to the larynx or upper trachea:
If large enough, they will create a "traumatic tracheotomy".

Blunt injury to the larynx (by a rifle butt):
This may crush the cartilage resulting in collapse of the airway.

Compressive haematoma in the neck:
This may accumulate slowly compressing the hypopharynx or larynx from the outside.

Burns to the face and neck or inhalation burns of the larynx and trachea:
Whether due to ordinary fire and smoke or specific irritating chemical agents, these require close observation for delayed obstruction or respiratory failure due to oedema.

Please note:

Inhalation of chemical agents requires special precautions for decontamination of the patient and of any equipment coming into contact with him. Not only is this an essential part of the treatment, but it is also necessary for the protection of the hospital staff and other patients.

The conditions causing delayed airway obstruction may develop slowly and there is a danger that they will not be recognized in time, especially during triage of mass casualties with inadequate supervision.

8.3.1 Opening the airway

As in first aid, the standard treatment for an actual or potential obstruction applies.

Basic steps for securing the airway

1. Open the mouth.
2. Displace the tongue forward.
3. Remove any blood or debris from the mouth and oropharynx.
4. Maintain the airway patent.

To open and clean the mouth the standard manoeuvres are jaw thrust or chin lift: displace the tongue forward, and then finger-sweep the mouth while protecting the finger, aided by mechanical suction if available. These techniques will cause some movement in the cervical spine. Manual in-line axial stabilization of the head during the manoeuvres will reduce such movement. Patency of the airway, nonetheless, takes precedence.

To maintain the airway patent a number of devices can be used:
- oropharyngeal airway (Guedel tube);
- nasopharyngeal airway;
- laryngeal mask airway;
- combitube (double lumen tube inserted blindly into the trachea and oesophagus).

All of these methods keep the airway open but cannot ensure protection against vomiting and aspiration of gastric contents.

During the reception of mass casualties, the overwhelmed ER personnel may have to revert to a simpler method as a temporary procedure. Under these conditions , the lateral recovery position (lateral security, post-tonsillectomy or semi-prone position) will be the position of choice for most patients with an airway at risk until more secure procedures can be achieved. The lateral recovery position affords relative protection against vomiting and aspiration.

8.3.2 Definitive patent airway

The following conditions require a *definitively* secured airway:
- apnoea or cardiac arrest, whatever the cause (with assisted ventilation);
- Glasgow Coma Scale score of less than or equal to 8, which is equal to being unresponsive to pain;
- continued seizures and convulsions (with assisted ventilation);
- unstable fractures of maxilla or mandible (usually bilateral fractures of the mandible, or instability of the whole mid-face at the level of the glabella and zygomata);
- large flail chest segment (with assisted ventilation);
- respiratory failure (with assisted ventilation);
- moderate to severe facial or oropharyngeal burn.

8.3.3 Definitive patent airway: endotracheal intubation

The best and simplest technique to maintain a patent airway with protection from aspiration is endotracheal intubation, through either a nasal or oral approach. Deeply unconscious patients can usually be intubated easily. Others may be restless, irritable,

uncooperative and hypoxic. Intubation under these circumstances usually requires sedation. Various agents given intravenously will allow rapid intubation without struggling and without compromising the patient's hypoxic condition (diazepam, pentothal, propofol, ketamine). Cardiac arrest, whatever the cause, and severe haemorrhagic shock (Class IV, see below),require emergency endotracheal intubation. The alternative to endotracheal intubation is a surgical airway.

8.3.4 Definitive patent airway: surgical airway

The need for a surgical airway should be identified *early* and performed *quickly*. This may be the primary airway (maxillo-facial injuries, wounds to the neck involving the larynx or pharynx or haematoma accumulation, etc.) or following failure of endotracheal intubation. A surgical airway is also beneficial where there are no facilities for mechanical ventilation.

Cricothyroidotomy is preferable to tracheostomy, which can be a difficult procedure under emergency conditions and be associated with profuse haemorrhage.

Cricothyroidotomy

This is a quick, safe, and relatively bloodless procedure (Figures 8.1.1 – 8.1.4). A horizontal incision is made in the skin and extended through the cricothyroid membrane. The handle of the scalpel is inserted and turned 90° to hold the membrane open until a small tracheostomy tube can be inserted.

Figures 8.1.1 – 8.1.4
Cricothyroidotomy.

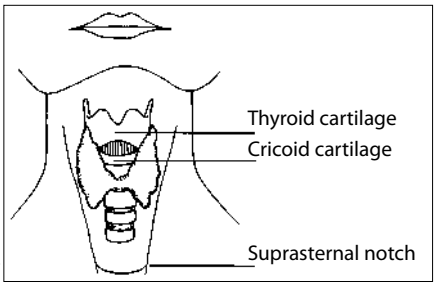

P. Zylstra / ICRC

Figure 8.1.1
Surgical landmarks: the patient's neck should be placed in extension with a pad beneath the shoulders. The thyroid and cricoid cartilages are identified by finger palpation, the cricothyroid membrane is then identified as the depression in between them.

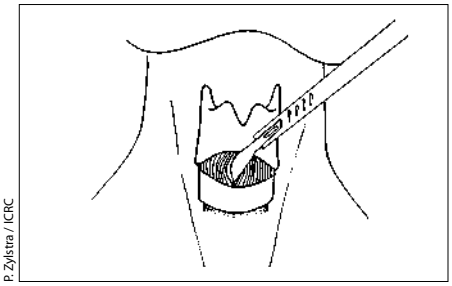

P. Zylstra / ICRC

Figure 8.1.2
A horizontal skin incision is made over the cricothyroid membrane. The wound is spread apart using the thumb and index finger.

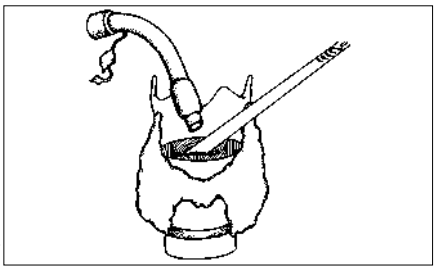

P. Zylstra / ICRC

Figure 8.1.3
The incision is carried down through the membrane and widened by insertion of the scalpel handle, which is then rotated through 90°.

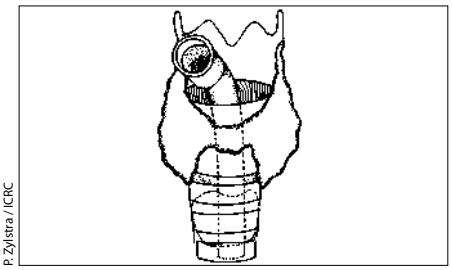

P. Zylstra / ICRC

Figure 8.1.4
A tracheostomy tube is placed through the opening and secured. The entire procedure should take no more than 30 seconds.

Tracheostomy

Tracheostomy should be an elective procedure. The only specific indication for an emergency tracheostomy in missile wounds is *direct laryngeal injury*, thus transforming a traumatic tracheostomy into a surgical one. The urgency of the problem will determine which technique is the safest and most appropriate for ensuring a secure airway.

8.4 Breathing and ventilation

The cause of respiratory distress must be found and treated. Patients with head injuries often require intubation and ventilation to support respiration, as is the case for quadriplegia, blast lung, chemical injuries and inhalation of smoke fumes. Previous disease may also impair ventilation in an injured patient.

Clinical examination may reveal a chest injury that decreases respiration including:
- flail segment of the chest;
- open pneumothorax or sucking wound to the chest;
- tension pneumothorax and haemopneumothorax.

Flail segment of the chest
This condition should be treated initially by bandaging or adhesive plaster strapping, good analgesia and positioning of the patient. More severe and complicated cases may require a chest tube and intubation with mechanical ventilation. It is usually the underlying lung trauma that causes the greatest difficulty in treatment.

For further details on flail segment of the chest, see Volume 2.

Sucking wound
A sucking wound requires a three-sided occlusive dressing in the ER. The patient then goes to theatre for debridement and closure of the wound of the chest wall and placement of an intercostal drain.

Please note:

There is a risk of transforming an open wound into a closed tension pneumothorax if all four sides of a dressing are firmly closed.

Tension pneumothorax
The diagnosis of a tension pneumothorax can be difficult on purely clinical grounds, except in the case of obvious respiratory distress; in this case no time should be wasted taking an X-ray (Figure 8.2). The condition requires an immediate wide-bore i.v. cannula to be inserted into the second or third intercostal space in the midclavicular line (needle thoracocentesis) attached to an improvised Heimlich one-way flutter valve as a temporary measure (Figure 8.3). A positive finding is indicated by the sudden rush of air as the needle enters the pleural cavity. The cannula should be long enough (8 cm) to fully penetrate the muscles of the chest wall (average thickness 4 – 6 cm). An intercostal chest tube (midclavicular or 5th intercostal space at the midaxillary line) should replace the needle thoracocentesis as soon as possible.

H. Nasreddine / ICRC

Figure 8.2
Tension pneumothorax with respiratory distress: the only thing wrong with the X-ray film is that it was taken. The diagnosis should have been a clinical one.

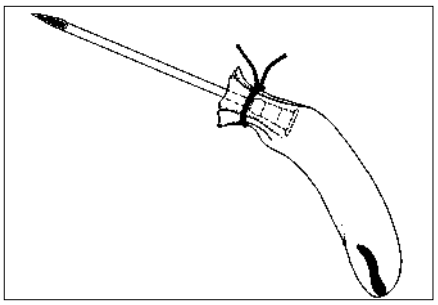

P. Zylstra / ICRC

Figure 8.3
Improvised Heimlich flutter valve: a wide bore cannula is inserted at the upper edge of a rib. A finger from a surgical glove, with a 1 cm-long incision in the end, is tied around the cannula.

Needle thoracocentesis may fail, however, even in the presence of a tension pneumothorax. A negative trial should not necessarily rule out its existence. A simple finger thoracostomy in the 5th intercostal space at the midaxillary line is a good alternative. A positive result, again, is indicated by the hiss of air rushing out; a chest tube is placed immediately.

H. Nasreddine / ICRC

Figure 8.4.1

Simple, but massive, pneumothorax. Patient lying comfortably and breathing without effort. X-ray films are justified.

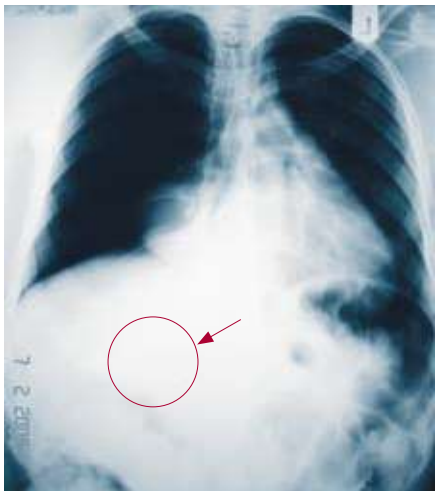

H. Nasreddine / ICRC

Figure 8.4.2

Antero-posterior radiograph showing large right-sided pneumothorax. The arrow marks the bullet.

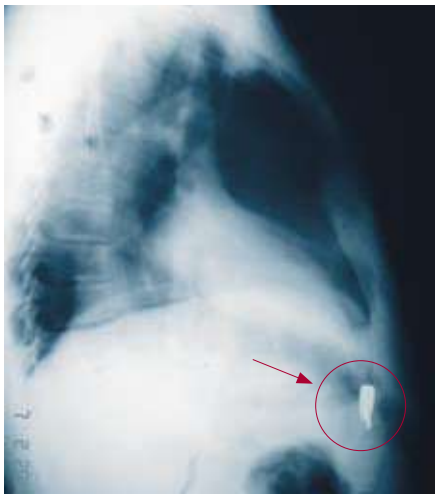

H. Nasreddine / ICRC

Figure 8.4.3

Lateral film.

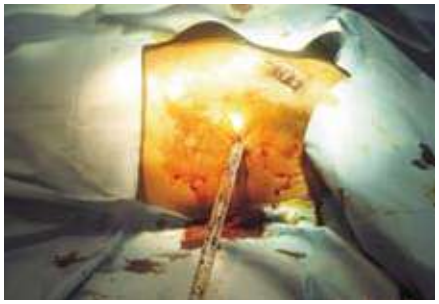

H. Nasreddine / ICRC

Figure 8.5

Intercostal tube.

Other, less extreme presentations of tension pneumothorax are more common than is realized. Table 8.1 shows the important signs and symptoms for making the diagnosis. These pertain to the awake casualty and differ from those in the sedated and ventilated patient. It is acceptable to perform an X-ray in patients whose presentation may be suspicious but who are not in respiratory distress to confirm the diagnosis. The patient should be accompanied, however, by someone capable of performing an emergency thoracocentesis in case of rapid decompensation in his condition.

Reliable and early findings	Ipsilateral side	Contralateral side	Pre-terminal decompensation	Inconsistent findings
Pleuritic chest pain (universal)	Hyper-expansion of the chest	Increased mobility of the chest	Decreasing respiratory rate	Tracheal deviation to contralateral side
Air hunger (universal)	Decreased mobility of the chest		Hypotension	Distended neck veins
Respiratory distress (universal)	Hyper-resonance on percussion		Much decreased SpO_2	Surgical emphysema of chest wall
Tachypnoea	Decreased breath sounds on auscultation		Decreasing level of consciousness	
Tachycardia	Added breathing sounds – crackles, wheezes			
Decreasing SpO_2				
Agitation				

Table 8.1 Diagnosis of tension pneumothorax in awake patients.[1]

Haemothorax

A haemothorax should be drained using a wide bore chest tube. This procedure can be lifesaving. When clinical signs of haemothorax exist, an intercostal tube should be inserted *before X-rays are taken*. Chest tubes are usually placed under local anaesthesia. If there is a wound to excise, ketamine anaesthesia would be more appropriate.

For information on insertion of chest tube – thoracostomy – see Volume 2.

8.4.1 Assisted ventilation

After intubation, assistance with respiration may be required. Common pathologies requiring such assistance include:

- head injury;
- large flail segment;
- blast injury of lungs;
- inhalation of toxic gases or smoke or flash burn to the bronchial tree;
- aspiration pneumonitis; and
- other medical causes of respiratory insufficiency.

Assisted ventilation may be manual:

- mouth-to-mouth or -nose for infants (use a compress barrier);
- mouth-to-mask;
- bag-valve-mask;
- bag-valve-endotracheal tube or surgical airway;

or mechanical, provided by a ventilator.

In hospital surroundings it is possible to administer supplemental oxygen, either by a central oxygen supply, compressed cylinders, or an oxygen extractor/concentrator.

The ICRC usually operates in situations of limited resources, and mechanical ventilators are not standard equipment. Few patients can be maintained for any length of time using manual ventilation by nurses and doctors. Furthermore, in a mass casualty situation the principles of triage will apply and most patients requiring assisted ventilation will be Category IV; therefore, it will not be performed (see Chapter 9).

1 Adapted from: Leigh-Smith S, Harris T. Tension pneumothorax – time for a re-think? *Emerg Med J* 2005; **22**: 8 – 16.

8.5 Circulation

The main circulatory problem encountered in the war-wounded is hypovolaemic shock, and is usually due to haemorrhage or burns. In addition to any blood loss, big soft-tissue wounds sequestrate a great deal of tissue oedema with further loss of plasma and circulating volume. Dehydration complicates any previous fluid loss if evacuation is long and delayed.

Neurogenic, anaphylactic, and cardiogenic shock may also occur; septic shock is a late complication. Injury to the spinal cord causing quadri- or paraplegia will result in neurogenic shock due to a disproportion between blood volume and dilated vascular bed. Allergy to antibiotics must always be kept in mind. Traumatic cardiogenic shock is due to direct injury to the heart that is not immediately lethal (e.g. small shrapnel wound causing traumatic myocardial infarction and/or pericardial tamponade).

8.5.1 Pericardial tamponade and pericardiocentesis

In the rare instance of constrictive *haemopericardium* with cardiac tamponade due to a penetrating missile wound to the heart, pericardiocentesis may be required if there is acute decompensation, to *gain time* until emergency thoracotomy can be performed.

The procedure for pericardiocentesis is as follows.

1. A 20 ml syringe is attached to a long over-the-needle cannula (or alternatively a spinal anaesthesia needle).
2. The skin is punctured 1 – 2 cm to the left of the xiphisternal junction, at a 45° angle, and the needle advanced through fascia and muscle.
3. (The inner trocar is now removed if using a spinal needle – but not if using the needle of the cannula).
4. The needle is gently advanced while aiming at the tip of the left scapula and applying continuous suction on the syringe.
5. When the needle tip enters the pericardial sac, blood will immediately appear in the syringe. The needle is withdrawn if using a cannula, which is then advanced into the sac.
6. As much blood as possible is aspirated. If the tamponade has been relieved, the patient's condition improves immediately; if the right ventricle has been entered, there is no change.
7. At the end of aspiration, the synthetic over-the-needle cannula can be left in place with a 3-way stopcock valve. (The spinal needle is gently withdrawn, millimetre-by-millimetre, if this technique is being used).
8. Should the tamponade return quickly, the stopcock can be re-opened and aspiration repeated (or the spinal needle tap is repeated).
9. The patient should *immediately* be taken to theatre!

An emergency room thoracotomy should not be performed when the operating theatre is just down the corridor: it is as pointless as it is dangerous in *most* hospitals around the world.

8.5.2 Haemorrhagic shock

Types of haemorrhage

Haemorrhage may be arterial, venous or capillary and blood loss may be:

- peripheral and obvious (blood clot the size of a fist or an open wound the size of a hand represents 500 ml);
- peripheral and hidden:
  - closed fractures of long bones (tibia = 500 ml, femur = 1.5 litres);
  - open wound with small entry that is blocked by a piece of torn muscle;
- central (chest, abdomen, pelvis and retroperitoneum) – significant bleeding into the chest should have been diagnosed during the breathing phase of the initial examination; a massive haemothorax can represent 2 – 3 litres of blood, a severe fracture of the pelvis 3 litres.

Penetrating trauma causing central bleeding into a body cavity requires surgery of some sort, and severe internal haemorrhage may require urgent operation as part of the resuscitation process. Such patients have a high priority for surgery.

Direct pressure and elevation of the limb will stop most cases of peripheral venous and capillary haemorrhage. Fractures should be splinted.

If a bleeding blood vessel can be *seen* in the depths of a wound, *and only if it is clearly visible*, direct control may be obtained with haemostatic artery forceps.

Blind clamping must never be attempted.

Otherwise, severe peripheral bleeding can usually be stopped by packing the wound. Digital pressure is applied to the artery proximal to the wound (pressure point), while the wound is carefully packed with small quantities of gauze compress at first, followed by a more bulky dressing, and finally a firm compressive elastic bandage in order to apply even pressure to achieve haemostasis. Alternatively, and if available, a pneumatic tourniquet can be used to temporarily control severe arterial haemorrhage while bringing the patient to the operating theatre.

Once a wound has been packed and bleeding has been arrested, the dressing (or tourniquet) should not be removed until the patient has been resuscitated and is in the operating theatre, with blood ready for transfusion if available. The surgical team should be ready to arrest bleeding by rapid intervention to obtain access to and control of the major blood vessels.

In the same manner of thinking, implanted bullets and fragments, or other foreign bodies in the wound, should only be removed as part of the operative procedure.

Beware of removing packs from missile wounds, as the resultant bleeding can be torrential and difficult to control.

The body's response and classes of shock

Immediately after haemorrhage, the body initiates a number of homeostatic circulatory changes that aim to stop the bleeding as well as compensate and preserve perfusion of the vital organs. The circulating blood volume represents 7 – 8 % of body weight in the adult (5 – 5.6 litres in the 70-kg male or 70 ml/kg body weight) and 9 % in children (80 ml/kg body weight).

Haemorrhage and the shock response are traditionally graded into 4 classes according to the volume of *acute blood loss*.

Class I:

Up to 15 % blood volume lost (750 ml or less). Mild tachycardia is the only clinical sign since the body's normal homeostatic mechanisms are capable of fully compensating the loss.

Class II:
15 – 30 % blood volume lost (750 – 1,500 ml). Definite tachycardia; slight decrease in systolic blood pressure with rise of diastolic pressure (decreased pulse pressure); refilling of blanched capillary bed of fingers delayed; restlessness or anxiety.

Class III:
30 – 40 % blood volume lost (1,500 – 2,000 ml). Marked tachycardia; tachypnoea; hypotension; small urine volume; classical picture of shock. Compensatory mechanisms start to fail.

Class IV:
> 40 % blood volume lost (> 2,000 ml). Full classic symptoms of shock are present: cold, clammy and pale skin; irritability, aggressiveness and confusion leading to loss of consciousness with loss of more than 50 % circulating volume.

Class	I up to 750 ml (<15 % loss)	II 750 – 1,500 ml (15 – 30 % loss)	III 1,500 – 2,000 ml (30 – 40 % loss)	IV >2,000 ml (>40 % loss)
Pulse	<100/min Full and bounding	100 – 120/min Full	120 –140/min Weak	>140/min Thready
Systolic blood pressure	120 Normal	90 – 120	< 90 Radial not felt	< 60 Carotid not felt
Pulse pressure	Normal	Narrowed	Greatly decreased	Absent
Capillary refill	Normal	Delayed	Delayed	Absent
Respiratory rate	14 – 20/min Normal	20 – 30/min Mild tachypnoea	>30/min Marked tachypnoea	>35/min Marked tachypnoea
Urine output	>30 ml/hr	20 – 30 ml/hr	5 – 20 ml/hr	Negligible
Mental function	Lucid/ thirsty/ slightly anxious	Anxious/ frightened/ irritable	Hostile/ irritable/ confused	Confused/ lethargic/ unresponsive
Physiological status	Fully compensated	Peripheral vasoconstriction	Compensation fails, classical clinical picture	Immediately life-threatening

Table 8.2 Signs and symptoms of haemorrhagic shock according to Class.

8.5.3 Fluid replacement

The important point is to maintain adequate tissue perfusion until the haemorrhage is brought under control. Sufficient blood pressure to maintain this tissue perfusion is judged to be a systolic pressure of 90 mm Hg, equivalent to a palpable radial pulse.

The great majority of war-wounded patients are relatively healthy young adults who suffer injuries to the extremities, and in whom the amount of blood loss is not lethal. They are haemodynamically stable (Class I blood loss) and the usefulness of oral fluids for their resuscitation – in the field and in hospital – is probably underestimated. Head, thoracic and abdominal injuries, however, even with Class I blood loss, should always have an intravenous line placed.

In massive injury, with obviously gross blood loss, several wide-bore intravenous lines should be established. A venous cut-down may be required if shock is profound. Available sites include: median basilic or cephalic in the arm, greater saphenous in the groin, and the distal saphenous vein at the medial malleolus. A venous cut-down should be used for less than 24 hours, or until the patient is filled up with fluids and another i.v. line can be put in. The site of the wounds will influence where i.v. lines are placed.

Modern practice has been to give a fluid challenge and monitor the response. The challenge has usually been with 2 l of isotonic crystalloid solution[2] in 30 minutes, or 3 ml for every 1 ml of estimated blood lost as a general rule of thumb. (Except for very large estimates: 3,000 ml blood loss would require 9,000 ml Ringers lactate!). The fluid challenge for a child is calculated as 20 ml/kg body weight for the initial bolus of Ringers lactate.

2 Ringers lactate is the preferred resuscitation fluid in ICRC practice.

Monitoring of clinical response

The estimation of blood loss is a very approximative measure, however. Rather than relying on this to guide resuscitation efforts, medics should observe the signs and symptoms of the *clinical response* to determine continuing fluid requirements. These include:

- pulse,
- systolic blood pressure,
- pulse pressure difference between systolic and diastolic,
- capillary refill,
- urine output, and
- mental status.

The most important simple parameter for the adequacy of fluid replacement is *urine output*: a urine flow of 0.5 – 1 ml/kg body weight/hour, and even greater in case of crush syndrome, should be aimed for.

The clinical response to a fluid challenge can be summarized as follows.

Rapid and stable response:

The pulse falls below 100, the systolic blood pressure rises above 100, and the pulse pressure widens. Urine output is good. These measurements remain stable. No further fluid resuscitation is required; the i.v. line is maintained open, however. This is the usual result with Class II and some Class III blood loss.

Transient unstable response:

An initial positive response of the pulse, blood pressure, and pulse pressure is followed by a return to subnormal values. Urine output remains low. A further challenge with Ringers lactate or a plasma expander (dextran 70, colloid, etc.) is called for. A return to normal values indicates compensated Class III blood loss, but early surgery is still required. Continuing subnormal shock values indicate continuing blood loss; the patient should be prepared for emergency surgery.

Please note:

Not more than two units of colloid or dextran should be given in any 24-hour period; besides being more expensive, colloids are less effective than Ringers lactate in diffusing into the interstitial space and can help provoke coagulation defects and interfere with cross-matching.

No response:

The patient remains in shock, indicating Class IV blood loss of more than 40 % blood volume, and requires emergency surgery – surgery as part of resuscitation – or is triaged to Category IV in a mass casualty situation (see Chapter 9);

or

other, less common pathologies that become manifest with time have been overlooked. The airway and breathing should be re-assessed to diagnose a possible cardiac tamponade, tension pneumothorax, or myocardial injury. Neurogenic shock and acute gastric dilatation should not be overlooked.

Time since injury should also be taken into account to evaluate the evolution of the shock state. If grade IV shock is present less than 1 hour after injury, emergency operation is needed for resuscitation. If grade IV shock has developed in 4 hours, resuscitation is required before operation.

Warning!

Beware of resuscitation injury: “shock lung” or acute respiratory distress syndrome in patients who receive massive crystalloid infusions that overburden the heart and lungs.

8.5.4 Hypotensive resuscitation

Giving all shocked patients a standard fluid challenge of 2 litres or more has been called into question. A rapid return to normal blood pressure *before* bleeding is controlled can be associated with hydraulic disruption of an effective blood clot ("popping the clot"), a dilution of clotting factors, and a lowering of blood viscosity that causes a decrease in resistance to flow around an incomplete clot. All of these factors may cause bleeding that has temporarily stopped to start again, especially if it is a non-compressible central haemorrhage. Thus, *moderate resuscitation* may be better than aggressive resuscitation; raising the systolic blood pressure to only 90 mm Hg rather to over 100. This is known as "hypotensive resuscitation" and should be considered in Class III and IV blood loss. It should *not* be used in head trauma where any hypotension is deleterious, or in the very young or old.

Therefore, in cases of massive haemorrhage, one must strive for a balance between death from shock after primary exsanguination or an increase in bleeding by prevention of clot formation or dislodging of formed and effective blood clot with secondary haemorrhage. In these extreme cases, surgery to stop haemorrhage is part of resuscitation: this should *never* be forgotten. Hypotensive resuscitation is widely used by ICRC surgeons today for patients who arrive quickly at a hospital.

However, if evacuation is delayed (more than 6 – 12 hours), rebleeding is less likely. In addition, delayed evacuation will result in more profound shock because of plasma loss due to post-traumatic tissue oedema and general dehydration rather than pure blood loss. Excessive sweating, vomiting or diarrhoea, rough handling of the casualty during transport, and a lack of splinting of fractured limbs will aggravate the total fluid and electrolyte loss. More aggressive resuscitation is permissible under these circumstances and the patient should be fully hydrated prior to surgery.

8.5.5 Adjuvant therapy

Never give cold fluids by rapid intravenous infusion.

No effort should be spared to prevent hypothermia in the shocked patient. Intravenous fluids should be warmed (see Section 8.8.1 and Chapter 18).

Oxygen and small doses of i.v. analgesia are equally important. The best is morphine (5 mg i.v., repeated every 10 minutes as necessary). Morphine should *not* be given if there is any question of head injury or respiratory depression. A good alternative in this case would be tramadol. Analgesia in shock should be given intravenously only.

A naso-gastric tube to prevent gastric dilatation should not be overlooked.

Steroids or a "vitamin C cocktail" should *not* be administered in cases of haemorrhagic shock.

Some clinical experiments have taken place using hypertonic saline for resuscitation. ICRC surgical teams have no experience in this and cannot comment.

8.6 Blood transfusion where supplies are limited

Where blood supplies are scarce, what should the role of blood transfusion be? This scenario is a far cry from optimal conditions in which there are relatively few limits to blood or component administration, yet it is common.

The aim of blood transfusion is to save life or to prevent significant morbidity, and *not* to restore a normal haemoglobin level. Blood is a rare and expensive commodity with serious risks attached to its administration and it should therefore be used with caution. The decision to give blood to a particular patient should be based primarily on the clinical state correlated with the laboratory findings, balanced against the risks and the shortage of supply. A better understanding of the physiology of oxygen

transport, the shortage of donors and the increased risk of viral infections has led to a policy of accepting lower levels of haemoglobin than before, without overly detrimental effects to the patient. One should avoid a transfusion for reasons such as "accelerating the patient's recovery", increasing patient comfort, or providing a supplement to correct anaemia. (Certain medical conditions, such as small children with severe malaria are an exception to this last rule.)

ICRC policy is to screen donated blood for hepatitis A, B and C, HIV, syphilis, and – in non-endemic areas – malaria. In countries where malaria is endemic, it is not uncommon for trauma patients to have a malarial attack 2 to 3 days post-operatively even without a blood transfusion; this is treated when it occurs. Where donors are scare, they are not excluded because of malaria. A blood transfusion can never be 100% safe.

The ICRC seeks blood donors in the community. Family, friends and clan members are usually the source. In certain countries, the National Red Cross/Red Crescent Society plays an important role in the collection of blood.

Culture and tradition in some societies may render blood collection extremely difficult; consequently, blood for transfusion is often in very short supply. Accordingly, and in keeping with the principles of triage, a maximum number of units of blood per intervention is set in ICRC practice. Today, this is usually about 4 units, and this figure should only be exceeded in cases of anti-personnel landmine injury with traumatic amputation and severe burn patients undergoing skin grafting. (Early tangential debridement with immediate skin grafting is not practised by ICRC teams, partly because of the risk of major blood loss.)

ICRC practice has been to transfuse whole blood, as fresh as possible. Otherwise, whole blood is stored with CPD-A anticoagulant (citrate – phosphate – dextrose – adenine). Blood components are not available in ICRC surgical programmes as is the case in many rural public hospitals. In practice, this has not been considered a handicap.

8.6.1 Clinical use of blood in ICRC practice

- Blood should not be given during resuscitation until bleeding has been controlled.
- Crystalloids and/or colloids are administered first: if the patient remains haemodynamically unstable *and* the haemoglobin is less than 6 g/dl, blood is administered. Haemoglobin less than 6 but in a stable patient is *not* an indication for transfusion. (However, there is a threshold haematocrit value of 5 – 10%, below which continued crystalloids or colloids can provoke cardiac arrest by "washout anaemia".[3])
- In the non-resuscitation scenario (preparing the patient for a second operation), pre-operative transfusions are not administered where blood supplies are low, unless the patient is anaemic *and* symptomatic.
- The blood supply and stocks determine whether a particular patient will receive blood: the principles of triage apply.
- Consider autotransfusion where appropriate (see Section 8.6.4).

If massive transfusion of stored blood is necessary, every second unit should be supplemented with one ampoule of sodium bicarbonate (44.3 mEq) and one ampoule of calcium chloride (10 g) by a separate i.v. line. As with crystalloid fluids, blood should be warmed to body temperature to avoid increasing hypothermia. This can be achieved through the use of locally-made water baths or the body heat of staff members.

8.6.2 Two unit rule

It has been traditional practice to administer not less than 2 units of blood to any patient who requires transfusion. In the ICRC context, this rule is not always appropriate since extreme shortage of blood is so common. It is sometimes more apt to prescribe just one

3 Takaori M, Safar P. Treatment of massive hemorrhage with colloid and crystalloid solutions. *JAMA* 1967; **199**: 297 – 302. Cited in Barkana Y, Stein M, et al. Prehospital blood transfusion in prolonged evacuation. *J Trauma* 1999; **46**: 176 – 180 and Shoemaker WC, Peitzman AB, Bellamy R, et al. Resuscitation for severe hemorrhage. *Crit Care Med* 1996; **24(2Suppl.)**: S12 – S23.

unit to certain symptomatic patients as this may improve their condition sufficiently, thus allowing supplies of this scarce resource to be kept for other patients in need. This is particularly true for patients in danger of "washout anaemia". World Health Organization (WHO) guidelines also accept one unit of blood in case of symptomatic anaemia.

8.6.3 Fresh whole blood

This is best when screened and given within one hour of collection. Fresh whole blood is especially reserved for:

- massive haemorrhage;
- coagulopathy;
- septic shock;
- non-trauma pathologies:
  - snake bite with haemolysis
  - amniotic fluid embolism.

8.6.4 Autotransfusion

When faced with patients suffering massive haemorrhage and inadequate blood supplies, ICRC surgical teams have practised recuperation of shed blood and autotransfusion. Haemothorax and haemoperitoneum from the spleen, liver or ruptured ectopic pregnancy are the most common indications.

For further information on autotransfusion in acute haemorrhage, see Volume 2.

8.6.5 Total blood requirements: the ICRC's experience[4]

ICRC colleagues studied total blood requirements for 4,770 patients in two hospitals treating the wounded from the war in Afghanistan during a six-month period in 1990 – 91. No differentiation was made for blood transfusion given pre-operatively, peri-operatively, or immediately post-operatively: all were considered part of patient resuscitation given the difficulties and delay in obtaining blood. Guidelines at the time allocated a maximum of 6 units per patient (although this was exceeded in some cases) and a limiting haemoglobin level of 8 g/dl. Autotransfusion was not used in this series.

A comparison was made of the number of units transfused according to time since injury for all wounded patients and separately for those with central injuries (head, neck, chest, and abdomen), and according to the cause of injury.

Need for transfusion was greatest for patients arriving less than six hours after injury, steadily decreasing to those arriving after 72 hours. This was true for all patients. Surprisingly, patients with central injuries required less blood, on average, than those with peripheral wounds.

The most remarkable results were associated with the cause of injury. Blood requirement for anti-personnel landmine injuries far exceeded that for bullets or fragments, (Table 8.3) and was the main reason for the high blood use in peripheral injuries.

	Mine	Fragment	Gunshot	Burn
Patients	787	2,577	1,016	111
Percentage transfused	27.9 %	13.1 %	15.0 %	18.0 %
Average transfusion units/ patient transfused	3.7	2.6	2.7	4.1
Average units/100 patients	103.2	34.1	40.5	73.8

Table 8.3 Blood given to patients related to cause of injury: 279 patients were categorized under "other" injuries.

4 Eshaya-Chauvin B, Coupland RM. Transfusion requirements for the management of war injured: the experience of the International Committee of the Red Cross. *Br J Anaesth* 1992; **68**: 221 – 223.

Consequently, ICRC recommendations for a hospital undertaking total surgical care of the wounded from conventional warfare were amended as follows:

- for every 100 war-wounded patients, the blood bank should expect to provide 45 units as a baseline;
- baseline should be increased to 60 units, if the majority of patients are admitted within 6 hours of injury;
- baseline should be increased to 100 units, if anti-personnel landmines are widely used in combat;
- burn patients require greater transfusion capacities, even in the absence of early tangential excision with immediate grafting (not practised in ICRC hospitals);
- long evacuation lines with much delay and no use of anti-personnel landmines in combat may negate the need for a blood bank, and blood for transfusion should be collected on an individual patient basis.

These recommendations may have little relevance to a modern industrialized army with very early evacuation and resuscitation but may be quite relevant in conditions of limited resources.

Please note:

ICRC guidelines have become even more stringent since: 4 unit limit and 6 g/dl haemoglobin level *with* unstable haemodynamics.

8.7 Disability

Any neurological deficit must be determined, whether central or peripheral. As mentioned previously, if the mechanism of injury is blunt trauma above the level of the clavicles then the cervical spine must be cared for in the classical manner.

According to the nursing expertise available, the AVPU system may be used or one may proceed immediately to the Glasgow Coma Scale (GCS) for the determination of the level of consciousness and any traumatic brain injury (Table 8.4). Although the GCS was originally intended for closed trauma to the head, and there are certain shortcomings when dealing with penetrating head trauma, nonetheless, it has been ICRC practice to use it.

	Response	Score
Eye opening	Spontaneous	4
	To verbal command	3
	To pain	2
	No response	1
Verbal response	Orientated and conversing	5
	Disorientated	4
	Inappropriate words	3
	Incomprehensible sounds	2
	No response	1
Motor response	Obeys verbal command	6
	Localizes pain	5
	Flexion-withdrawal	4
	Flexion-abnormal	3
	Extension	2
	No response	1

Table 8.4 Glasgow Coma Scale: the best response for each variable is scored. Maximum score is 15, minimum is 3.

A GCS < 8 indicates severe head injury and airway protection is mandatory, by intubation or cricothyroidotomy/tracheostomy.

Disability examination includes the entire vertebral column/spinal cord: presence of paraplegia, level, etc. Palpation of the vertebrae, one by one, for tenderness,

induration and crepitus is more important than deformity, which may be masked by haematoma. A rectal examination to determine sphincter tonus is important for prognosis!

Proper stabilization measures should be undertaken and a urinary catheter placed. Neurogenic shock is common in paraplegic patients: fluid resuscitation and an i.v. vasopressor are often required.

For the management of paraplegic patients, see Volume 2.

8.8 Environment/exposure

Hypothermia must be avoided at all costs – rapid, yet thorough examination of the patient – and treated aggressively. With a core body temperature of 37° C, an ambient temperature of 32 – 34° C is considered neutral. Less than this and the body loses heat to the environment. After examination, the patient should be kept covered, even in a tropical climate.

8.8.1 "Fatal triad" of coagulopathy, acidosis, and hypothermia

Recently, the importance of the potentially "fatal triad" of hypothermia, acidosis, and coagulopathy in trauma patients has become recognized. Hypothermia (core temperature less than 35° C) is probably the most potent factor in causing the vicious cycle of this syndrome.

Every effort should be made to preserve heat in an injured patient, since rewarming consumes far more energy than maintaining normothermia. Warm inhaled O_2, warm i.v. fluids, and external rewarming to a maximum of 40 – 42° C should be the first step. More aggressive measures of "central reheating" such as rectal enema and gastric, bladder and peritoneal lavage (at 37° C) can be used.

Post-traumatic coagulopathy is a well-recognized danger in patients suffering severe shock and submitted to aggressive resuscitation, especially if they have received massive amounts of i.v. fluids and blood transfusions with stored blood. It results from a combination of factors. Fresh whole blood transfusion is very useful in the absence of blood components, and probably in their presence as well (see Chapter 18).

8.9 Complete examination

At this stage it is even more important than in the pre-hospital setting to undress the patient and perform a thorough examination, from head to toe, front and back and sides. In some societies, this may contravene certain cultural and religious traditions (male doctor examining a female patient). Compromises must be found.

In the more accommodating atmosphere of a hospital emergency room, a systematic approach should be used to thoroughly examine the scalp and head (mouth, nose, and ears), neck, thorax, abdomen, perineum (scrotum and urethra, rectum, and vagina), the back of the trunk and buttocks, and the extremities. The peripheral pulses, temperature and capillary refill are compared on both sides. Motor function of the main peripheral nerves is tested. The aim is to have a complete assessment of all injuries and a more accurate assessment of the organ-specific damage.

The complete examination is best described as a thorough palpation. The entry wound may be exceedingly small and not seen. This is especially the case with fragment wounds to the head or perineum where body hair matted with blood can easily hide the wound (Figure 8.6). The entry wound must be identified by close palpation. Remember also that contusion/erythema can be better felt rather than seen in dark-skinned people.

M. Baldan / ICRC

Figure 8.6

Small temporo-zygomatic entry wound hidden by hair.

One should attempt to identify the likely path through the body of the projectile. This may involve any structure between the entry and exit wounds, or the position of the projectile on X-ray if no exit exists. Remember that wounds to the chest, buttock, thigh or perineum may involve the abdominal cavity (Figures 8.7.1 – 8.7.3).

H. Nasreddine / ICRC

Figure 8.7.1

Bullet wound to the pelvis: the entry is located on the right side. A defunctioning colostomy has been performed.

H. Nasreddine / ICRC

Figure 8.7.2

Exit wound in the left buttock.

H. Nasreddine / ICRC

Figure 8.7.3

Any structure between entry and exit should be identified: here blood is seen in the rectum.

A simple outline drawing of the body on the admission chart (homunculus), front and back, is useful for recording all injuries.

Dressings on the limbs should not be removed if the casualty is haemodynamically unstable. Only once resuscitation has begun and the patient's condition is under control is it safe to examine wounds to the extremities, preferably in the operating theatre. Fractures should nonetheless be immobilized if this has not already been done in the field.

The ABCDE paradigm should be monitored for any change in condition of the patient. Resuscitation and stabilization are continued, while complementary examinations are performed. The extent of the latter will depend on the level of sophistication and competency of the particular hospital.

A basic complement is a plain X-ray, one body cavity above and below any entry or exit wound. If no exit wound exists and no projectile is evident, further radiographs should be taken to locate its position. It may be difficult to differentiate a radio-opaque bullet from a normal anatomic radio-opacity such as the heart shadow (see Chapter 10 and Figures 8.4.2 and 14.9.1).

The complete and definitive examination will involve each specialized body system. The signs, symptoms, and treatment will be described in the relevant chapters of Volume 2.

8.9.1 Complementary diagnostic examinations and monitoring

Electrocardiogram (ECG) monitoring is not routinely available in ICRC surgical hospitals, neither are computerized tomography (CT) scans, angiography, sonography, Doppler blood flow, or arterial blood gases. The use of central venous pressure lines in most circumstances where the ICRC works carries too high a risk of septicaemia. Diagnostic peritoneal lavage for abdominal injury is also not practised routinely.

The minimal ICRC standards for emergency war surgery under precarious circumstances in hospital settings with limited resources include:

- plain X-ray;
- pulse oxymeter;
- haemoglobin;
- haematocrit;
- total and differential white blood cell count;
- platelet count;
- coagulation time;
- bleeding time;
- fasting blood sugar;
- blood smear for malaria (other blood parasites where applicable);
- sickle-cell test (where applicable);
- urine analysis: dipsticks, pregnancy tests;
- blood grouping, testing and cross-matching.

If more general surgery, as well as internal medicine and paediatrics, is widely practised in the ICRC hospital, then more elaborate laboratory analyses are added. In more precarious situations, as experienced by field surgical teams, none of the above is usually available.

Chapter 9

HOSPITAL TRIAGE OF MASS CASUALTIES

9. HOSPITAL TRIAGE OF MASS CASUALTIES

9.1 Introduction 191
9.1.1 The logic of triage 191
9.1.2 Where to perform triage 192
9.1.3 A delicate balance 192

9.2 Setting priorities: the ICRC triage system 193
9.2.1 The ICRC triage categories 194
9.2.2 Notes on triage categories 195

9.3 How to perform triage 196
9.3.1 "Sift" 196
9.3.2 "Sort" 197
9.3.3 Avoid undertriage and overtriage 197
9.3.4 Avoid confusion and disagreements within the team 198

9.4 Triage documentation 198

9.5 Emergency plan for mass casualties: disaster triage plan 199
9.5.1 Planning 199
9.5.2 The team 200

9.6 Personnel 200
9.6.1 Triage team leader 200
9.6.2 Triage officer 200
9.6.3 Head nurse 201
9.6.4 The team 201
9.6.5 Triage groups/nursing teams 201
9.6.6 Surgeons and theatre personnel 202
9.6.7 Rest and relief 202

9.7 Space 202

9.8 Equipment and supplies 203

9.9 Infrastructure 204

9.10 Services 204

9.11 Training 204

9.12 Communication 204

9.13 Security 205

9.14 Summary of triage theory and philosophy: sorting by priority 206
9.14.1 Triage system: a simple emergency plan organizing the personnel, space, infrastructure, equipment, and supplies 206
9.14.2 Emergency hospital disaster triage plans differ and no two triage scenarios are the same 206
9.14.3 "Best for most" policy 206

ANNEX 9. A Sample triage card 207

ANNEX 9. B Hospital emergency plan for a mass influx of wounded 208

9.1 Introduction

9.1.1 The logic of triage

In daily, routine practice surgeons face patients one by one. They use all of the means, equipment, and supplies at hand to do the most they can for every patient; they try to do everything they possibly can for every single individual. The priority is high-intensity care for the sickest.

In a single mass-casualty incident means may be stretched to the limit, but one can still manage to do the best possible for all patients. With a massive influx of wounded, however, the hospital is overwhelmed; resources cannot meet the needs of all the casualties. This is when the logic of mass casualty triage comes into play. It is no longer possible to do everything for everyone. Medics must try to do what they can: the greatest good – but not necessarily everything – for the greatest number. This is another example of the change in professional "software" – the mindset – that is called for in surgery for the victims of war.

A change of mindset: from "everything for everyone" to "best for most".

The French term *triage*, which means to sort into groups according to quality, has been applied to the sorting of battle casualties since the time of Napoleon's surgeon of the Imperial Guard, Baron Dominique-Jean Larrey (circa 1812), who recognized the need to categorize wounded soldiers during an actual battle according to their priority for treatment.

The experience of two world wars during the twentieth century, when tens of thousands were wounded during a single battle, demonstrated the importance of sorting casualties and determining priorities for evacuation and treatment. This concept of triage has now been extended to natural disasters and single mass-casualty incidents (terrorist bombings, industrial accidents, fires in social institutions, etc.) with considerable benefit, and its use is now generally accepted worldwide.

Natural disaster triage, however, is not the same as war triage. An important natural disaster is a single mass-casualty event that overwhelms hospital services; the wounded suffer injuries at one time, but there is then a respite. The same can be said for other mass-casualty incidents. Armed conflict on the other hand can involve a continuing mass arrival of freshly wounded victims that may go on for weeks or months on end; there may be no respite until the end of hostilities. Nonetheless, many of the fundamental concepts underlying war triage apply to the disaster scenario.

Establishing priorities for treatment amongst many casualties is the most difficult decision to be taken in all of medical practice. Patients with very severe wounds whose management consumes many resources and much time, and who have little chance of survival despite treatment, may have to wait or may only receive minimal care in the interest of being able to intervene effectively for the others. One must often choose to treat first the patients for whom the smallest surgery gives the biggest results; i.e. the wounded who, though seriously injured, stand a fair chance of "good survival".

The aim in a mass casualty situation is to do the "best for most".

Some armies have practised "inverted" or "reverse" triage: the least injured were treated first so as to make them able to return more quickly to the battlefield! It should be noted that this practice contradicts the principles of international humanitarian law.

The practice of triage is not dogmatic. It is not a series of rules, but a logical approach and philosophy that must be adapted to each and every particular situation.

9.1.2 Where to perform triage

As mentioned in Chapter 1, one of the special characteristics of surgery in times of war is the staged management of patients in a chain of casualty care. The principles of triage are applied at every stage.

Triage takes place at every echelon of the chain of casualty care, including during evacuation.

The organization of a casualty collection point and effective triage permits the orderly evacuation of the wounded, the most efficient use of stretchers, ambulances or other transport, and the optimal use of medical and surgical personnel available. With the necessary training, rapid assessment of casualties can be carried out not only by doctors, but also by nurses, paramedical staff and first-aid workers. It should be possible to implement triage at *any point* along the chain of casualty care, in order to set *priorities* for first aid and evacuation of the wounded to the next echelon of treatment.

More efficient lines of evacuation and shorter duration of transport, coupled with better pre-hospital care, ensure that a greater number of the severely injured will survive to reach a surgical hospital. The number of KIA decreases; however, the DOW rate often increases (see Chapter 5).

In many low-income countries, however, civilian casualties often arrive at the hospital without having received first aid and without the help of an ambulance service. In this case the less seriously injured usually arrive first. The hospital then becomes the first triage station.

Whatever the prior treatment or sorting, however, triage must be carried out again when casualties reach the hospital; the condition of patients changes and the priorities in the hospital may be different from those in the field. All patients arriving at the hospital during an influx of casualties go through the triage process.[1]

9.1.3 A delicate balance

To set priorities for patient management a number of factors must be taken into consideration to define the needs on the one hand, and the resources available on the other. The practice of triage is a fine balance between these two.

Triage is a dynamic equilibrium between needs and resources:

- needs = number of wounded and types of wounds;
- resources = facilities at hand and number of competent personnel available.

Needs

- How many patients are arriving? 10, 50 or 100?
- Are they all suffering from penetrating injuries? Or are there many burn cases amongst them?

The number of wounded and the different pathologies should be taken into account in the total workload. Penetrating wounds will require a great deal of surgery. Burn patients require relatively little immediate surgery, but a great deal of nursing care.

1 For field triage, see Giannou C, Bernes E. *First Aid in Armed Conflicts and Other Situations of Violence*. Geneva: ICRC; 2006.

Resources

- How many surgeons and anaesthetists?
- How many operating tables?
- How many instrument boxes, and what is the efficiency of the sterilization system?
- How many beds? etc.

A hospital may have three fully-equipped operating theatres, but if there is only one surgeon then only one patient can be operated at a time. If there are three surgeons and three anaesthetists, but only one theatre, then the team must improvise extra operating facilities, if there are sufficient boxes of surgical instruments.

ICRC EXPERIENCE

During fighting in Kabul in 1992 with an influx of mass casualties, four surgical teams worked in the ICRC hospital, supplemented by Afghan general practitioners with some surgical experience. There were two operating theatres in different buildings with 4 tables that could function at the same time. One team operated in one building and one in the other; the third team performed triage; and the fourth was resting!

All these factors – needs and resources – must be balanced out. No two triage situations are the same. It is this constantly changing equilibrium between needs and resources that will determine the priority for treatment among all the patients who reach the hospital. One cannot be dogmatic about which patient to operate first, since no two situations are the same. Only a full understanding of the logic and philosophy of triage will help to establish efficient priorities that will result in the best treatment possible for the greatest number of victims.

9.2 Setting priorities: the ICRC triage system

There are a number of triage category systems used throughout the world today. Some are more sophisticated than others and depend on injury severity scores and physiological parameters. There are two important factors to keep in mind when deciding upon a system to use in a hospital.

1. It should be kept as simple as possible – a mass influx of wounded always creates confusion, tension and anxiety.
2. All members of the hospital team should understand the system being used.

The ICRC uses a triage category system based upon several factors.

- Physiological appreciation of life-threatening conditions according to the ABCDE protocol – in practice, severe haemorrhage will constitute the great majority of life-threatening injuries during armed conflict (see Chapters 5 and 8).
- Anatomic factors, including the Red Cross Wound Score for penetrating injuries – in general, vital injuries (head, neck, thorax, abdomen and major peripheral vessels) will have priority, but the grade of injury must also be taken into consideration (see Chapter 4).
- Mechanism of injury – especially anti-personnel landmine injuries and burns (see Chapter 3).
- Time since injury – important for determining degree of shock and response to resuscitation as well as limb ischaemia (see Chapter 5).
- Epidemiology of the war-wounded – knowledge of the relative numbers of severe and superficial wounds (see Chapter 5).

9.2.1 The ICRC triage categories[2]

Category I: Serious wounds – resuscitation and immediate surgery

Patients who need lifesaving surgery and have a good chance of recovery. Some examples include:

- airway – injuries or burns to the face and neck requiring tracheostomy;
- breathing – tension pneumothorax, major haemothorax;
- circulation – internal haemorrhage, wounds to major peripheral blood vessels, traumatic amputation.

Category II: Second priority wounds – can wait for surgery

Patients who require surgery but not on an urgent basis. In practice this applies to many casualties, including:

- penetrating abdominal injury in haemodynamically stable patient, most probably injury to hollow organs only;
- penetrating head injuries with GCS > 8, providing definitive airway control can be maintained – if a surgical airway is necessary, then Category I for tracheostomy only;[3]
- most compound fractures: in practice, a large number of the wounded;
- major soft-tissue wounds: in practice, a large number of the wounded.

Category III: Superficial wounds – ambulatory management

Patients who do not require hospitalization and/or surgery because their wounds are so minor that they can be managed on an ambulatory basis.

They are often called the "walking-wounded". In practice, this is a very large group including those presenting superficial wounds managed under local anaesthesia in the emergency room or with simple first-aid measures.

Category IV: Severe wounds – supportive treatment

Patients with injuries so severe that they are unlikely to survive or would have a very poor quality of survival. These include the moribund, or patients with multiple major injuries whose management could be considered wasteful of scarce resources in a mass casualty situation in terms of operative time and blood. Examples include:

- penetrating head wound with GCS < 8;
- quadriplegia;
- burns > 50 % body surface area;
- major blood loss and no blood available.

2 The system of triage categories employed by the ICRC was revised by the Master Surgeons Workshop held in Geneva in 2002 (see Introduction).

3 Please note: ICRC hospitals are *not* equipped with ventilators and close monitoring of intubated patients is not always possible. A surgical airway avoids many problems, and should replace an endotracheal tube if the patient remains intubated for several days in any case.

9.2.2 Notes on triage categories

The number of casualties who require urgent treatment may exceed the surgical capacity available. A second triage within Category I is then necessary (see Section 9.3.2). On the other hand, when evacuation time to the hospital is longer than 12 hours, few patients may fall into Category I.

Many surgeons believe that all penetrating brain wounds are a Category I emergency; others believe that all penetrating brain wounds are hopeless Category IV. The use of the GCS helps to differentiate those patients who will probably survive, even with some delay in operation (Category II in the experience of ICRC surgical teams) from those whose condition is definitely a Category IV. This, evidently, provided the airway is patent.

The Category III superficially wounded may be so numerous, frightened, in pain and excited, that their uncontrolled presence in the emergency room/triage area creates great confusion. In urban contexts, they tend to be the very first evacuated to hospital, to the detriment of the more seriously wounded.

The categories are not rigid; patients waiting for surgery may change category and a single patient may fall into two categories, such as one with severe maxillo-facial injury who requires an immediate tracheostomy and basic haemostasis (Category I), while the debridement and primary reconstruction of the patient's face, which may take many hours, can wait (Category II).

Please note:

There is one exception to these categories: when an excited, and often drunk, combatant holds a gun to your head and demands that you treat his wounded comrade first. This patient immediately becomes top priority.

ICRC EXPERIENCE

Two examples of the epidemiology of triage patients recorded in recent years.

Fighting in Kisangani, Democratic Republic of the Congo, June 2000

Four hospitals and 62 clinics registered a total of 2,393 wounded. Only 25 % required hospital care, and even less a surgical operation. The vast majority were Category III and could have been treated as ambulatory cases; many were hospitalized for several days nonetheless.

Combat in Monrovia, Liberia, June – August 2003

2,567 patients were triaged in the JFK Memorial Hospital, but only 1,015 were deemed to require hospital admission (40 %), only 718 of whom (71 %) underwent operation. Of those admitted, some did not require any operation (paraplegia), while for many others there was simply not enough time or human resources available. Many patients who were originally placed in Category II – and given antibiotics, dressings and analgesia – eventually became Category III, with and without sequelae.

The equilibrium – between needs and resources – is a dynamic one; change is constant, no two triage situations are the same.

9.3 How to perform triage

Triage is a dynamic process: it requires a continuous reassessment of patients.

Patients may change triage category: their condition may deteriorate, or improve over time and with pre-operative resuscitation. As a result, a continuous reassessment of patients is absolutely necessary.

Triage is a multiple-step process: "sift and sort", then re-examine, re-examine, re-examine. "Sift" involves placing the patient in a general Category; "sort" then decides priority within that Category.

9.3.1 "Sift"

On reception of the casualties in the hospital, the triage officer must perform a rapid examination of each patient – a maximum of 30 seconds – checking the whole body, including the back. Field dressings should be changed by the nursing team as part of the examination; except for obviously large and severe wounds whose dressings are removed only in theatre. Severity of injury and probability of survival are the keys to decision-making. Clinical experience with the war-wounded – not training or reading books – is the most important preparation for this task.

To sift, the triage officer first looks for life-threatening conditions based on:
- airway, breathing, circulation;
- important physiological clues (mental status, character and rate of pulse, ease of respiration);
- anatomic site of wounds (head, chest, abdomen);
- severity of obvious wounds according to the RCWS (amputation of limb, etc.).

Based on clinical judgement alone, a first decision is made, to "sift" the casualty into a Category. The Category does *not* depend on how many patients arrive with serious wounds.

Every wounded patient must be examined *and* categorized immediately.

The idea is to quickly recognize the two extremes of injury, which, together, comprise a very large number of casualties: the very minor (Category III) and the very serious (Category IV); and to identify and separate out the dead. As little time as possible is spent on these Category III and IV patients and they must quickly be removed from the triage area to specially designated areas.

The triage officer concentrates on the critically and seriously injured (Categories I and II):
- patients who require resuscitation and immediate surgery as part of the resuscitation process;
- those who need continuing resuscitation; and
- patients who will tolerate some delay before receiving surgical attention.

The triage officer should not perform *treatment*, with one exception: if the patient is not breathing, the mouth should be opened and cleaned and the person put in the lateral recovery position, before the next patient is dealt with.

Å. Molde / ICRC

Figure 9.1
Triage officer assisted by an administrative clerk.

An "administrative team" consisting of a nurse and a clerk supports the triage officer for purposes of documentation, including all personal details of each patient, the securing of their identity papers and valuables, and the signing of any necessary consent forms. Other supporting personnel present at the entrance or in the triage area include those responsible for hospital security and crowd control.

Initial treatment is started by a dedicated team, and patients should be transferred as quickly as possible out of the triage area to theatre or the ward designated for their triage Category in order to make room for new arrivals. Resuscitation, as necessary, is continued in these assigned areas.

The use of X-rays should be *limited*; it is seldom essential under these circumstances and the triage officer should not wait for an X-ray to decide on the triage Category.

An identifying tabard or armband stating "triage officer" is useful; if, inevitably, the triage officer must move away from the triage area this can be handed over to the deputy/replacement.

9.3.2 "Sort"

After "sifting" casualties into the general Categories, a second examination is performed: "sorting". No two triage situations are equal and, therefore, according to the number of competent personnel available, the triage officer or a second physician-in-charge continues the reassessment of patients within the Category I cohort. "Sorting" decides which of the Category I patients to send to theatre first: *priority amongst the priorities*. The others, while waiting, remain under close observation and continued resuscitation.

This re-examination may reveal patients with an "emergent" condition who should receive priority over the stable: e.g. one of two Category I patients with penetrating abdominal wounds becomes haemodynamically unstable because of continuing internal haemorrhage, while the other has a stable pulse and blood pressure because bleeding has stopped.

The surgical teams should begin operating on Category I patients immediately. Since the number of casualties cannot be foreseen, it is not possible to wait for all patients to arrive and be triaged before deciding which should be taken to the operating theatre first.

When in doubt about priority amongst Category I patients, put in chest tubes and send in a laparotomy; get going!

The triage officer must keep in close contact with the operating theatre, to reassess continuously the priorities of the operating list.

The time that patients are waiting for surgery can be profitably used for completing X-rays and laboratory analyses, if necessary, but should not interfere with resuscitation efforts.

The same logic applies to the patients in the other Categories. Either the triage officer or the person in charge of the department re-examines the patients in Category II. Priorities are set for the operating list among the Category II patients.

Reassessments may result in changing a patient's Category. The condition of a patient originally triaged into Category III because he had no *obviously* serious wound may suddenly deteriorate and become Category I or II. If a second-in-charge is continuing the reassessments, a system must be available to transfer patients from one Category area to another. Referring back to the triage officer for every decision will only result in overburdening him.

Usually the hardest part of triage is having to accept that some patients may only receive analgesics and be removed to a quiet place where they can die in comfort and with dignity. After treatment of priority patients has been completed, a repeat examination of Category IV patients, if still alive, may allow the triage officer to contemplate surgery for them at this stage.

9.3.3 Avoid undertriage and overtriage

In undertriage the assessment underestimates the severity of injury and the patient is not given sufficient priority. Overtriage overestimates the injury, and a patient is assigned to a higher Category than necessary. This will divert resources from the truly seriously injured and overburden the critical care services. Repeated re-examinations will correct these errors.

It may be necessary to transfer minimal care patients to other sites for convalescence, even outside the surgical hospital. If this is done, care should be taken to identify patients for follow-up treatment and administration of medication, and to exercise some control over their whereabouts. Documentation is important: patients can easily get lost.

9.3.4 Avoid confusion and disagreements within the team

Triage decisions must be respected.

There is no time or place for disagreements during a triage of mass casualties. The decisions of the triage officer must be "dictatorial". The post-triage evaluation session is the place for "democratic" discussion and constructive criticism.

In spite of training, practising and planning, the unexpected will happen. The hospital team – led by the team leader, triage officer, head nurse – will at times have to improvise and invent new protocols and procedures to meet new circumstances. One should not be dogmatic, but rather understand the philosophy and logic of the triage process, and adapt to new situations in accordance with that logic.

9.4 Triage documentation

Good records are essential and no effort should be spared to record important aspects of the wounds, treatment, and the patient's triage category.

Each casualty should be appropriately identified, numbered, and assigned a medical chart. Large plastic bags, labelled with the patient's number, are used for clothing; smaller labelled plastic bags are used to collect patients' valuables. They are stored separately and the valuables put in a safe place.

Some system must be devised for indicating the patient's triage Category. This could be colour-coded tags tied around a hand or foot, or hung around the neck. These are easily removed and changed should the patient's triage Category change. Writing an "indelible" number on the forehead or chest only creates confusion with a change in Category.

The patient's medical chart should include basic information and be in telegraphic style: clear, concise, yet complete. It should include *at least* the following information:
- name, age, sex, time of injury, cause of injury, any first aid given;
- time of admission to hospital;
- vital signs: BP, pulse, respiration rate, neurological status;
- diagnosis: using a diagram (homunculus) is very useful;
- triage Category;
- complete pre-operative orders.

See Annex 9. A: Sample triage card.

This basic information is particularly important if patients are being transferred to another facility. A list of admitted or treated patients is necessary so that people who come looking for their relatives or friends may be informed. The local authorities may require information about the number of admissions and deaths. The media must be dealt with. These issues are addressed by the triage team leader, or replacement.

ICRC EXPERIENCE

During the post-triage evaluation, it was decided to use a plastic sleeve to hold the triage cards in order to protect them from various body fluids.

ICRC

Figure 9.2
State of the admission card after triage.

9.5 Emergency plan for mass casualties: disaster triage plan

Any hospital treating the war-wounded must be prepared to receive large numbers of casualties. Preparation means planning and training.

9.5.1 Planning

The triage tent in Figure 9.3 shows a number of key points:
- there is enough space to move around;
- the small lightweight beds/stretchers that are easily moved and inexpensive;
- small carts for emergency medical supplies;
- i.v. fluids hanging on a rope strung across the room, for flexibility of patient placement;
- a patient carried on a stretcher by dedicated bearers;
- relatively few staff members present, getting on with their allotted tasks in what appears to be a calm atmosphere.

This scene illustrates proper organization, which requires planning of the space, infrastructure, equipment, supplies and personnel.

ICRC

Figure 9.3
Inside a triage tent.

Figure 9.4 shows the premises surrounding a triage tent:
- the site is temporary;
- space is available to receive ambulances or to expand the facilities;
- the structure is lightweight and easy to set up or take down;
- the structure is inexpensive;
- there is a water tower nearby;
- there is a rubbish bin outside.

This scene illustrates proper organization, which involves planning of the space and infrastructure.

ICRC

Figure 9.4
Outside a triage tent.

A triage system cannot be organized ad hoc: it requires planning and training.

The hospital team must be prepared for any kind of crisis: every hospital should have a disaster/triage plan (see Annex 9. B: Hospital emergency plan for mass influx of wounded). The aim of the plan is to organize:
- personnel;
- space;
- equipment;
- supplies (medical and non-medical);
- infrastructure (water, fuel for electricity generators, etc.);

- services (laundry, kitchen and cafeteria, etc.);
- training of the hospital team;
- communications;
- security.

9.5.2 The team

The hospital team should hold a series of meetings, to discuss the organization of the disaster/triage plan. Everyone working in the hospital should be aware of the plan and their respective role during a crisis. The plan should be posted so that everyone is well acquainted with it.

The plan should be put into operation as soon as notice is given of the expected arrival of mass casualties. It should include the mechanism for deciding who declares the emergency and under what conditions to implement the plan. The ordinary operating list and other routine activities should be suspended until the situation is resolved. This kind of organization does not require money or special technology; only time, effort, discipline and motivation. Any disaster plan should be an extension of the normal hospital routines, and the roles allocated to individual staff remain as close as possible to their familiar daily work. The plan should not be a complete change in the system. Introducing many new procedures will only add to the confusion.

A simple emergency plan: personnel, space, infrastructure, equipment, supplies = system.

9.6 Personnel

There are three key leadership functions in a disaster triage scenario: triage team leader, triage officer, head nurse.

9.6.1 Triage team leader

The triage team leader is the coordinator. He is usually designated to announce the onset of the hospital triage plan; he then coordinates the work of the different units and services, and makes sure that all departments are informed. The triage team leader maintains an overview of the situation, including a constant reassessment to determine the need for additional staff, supplies, and ward areas. In addition, he must be aware of events outside the hospital, maintaining contact with relevant authorities in order to anticipate any new arrival of casualties due to continuing combat.

9.6.2 Triage officer

The triage officer performs the actual clinical triage, assigning a category to each and every patient on entry. There has been much discussion about who should perform triage: surgeon or anaesthetist? Again, there is no strict rule. Each hospital team must decide according to its circumstances.

The logic of triage demands that the most experienced and respected person willing and able to take on the responsibility should do so. This person must know how to organize the emergency room/triage area and have a good understanding of the functioning and capacity of the hospital.

Even more importantly, the hospital team must be able to live with the decisions taken by the triage officer. Staff members, relatives, and military commanders might try to influence the triage decisions; nevertheless, these should be made on purely medical grounds. This can be especially difficult in a public hospital where friends and family members of the hospital staff may be among the victims. Deciding priority of treatment solely on the basis of medical need and resources available when the injured include close relatives of the personnel may be a heart-wrenching decision.

The hospital staff must be able to continue their work and live with the triage decisions for the rest of their lives. As a consequence, the experience of the triage officer and the respect with which the personnel hold the triage officer must be beyond reproach.

Triage Officer

No task in medicine requires greater understanding, skill, and judgement than the categorization of casualties and the establishment of priorities for treatment.

9.6.3 Head nurse

The head nurse organizes the nursing and paramedical personnel (laboratory, pharmacy, etc.) and non-medical support staff (trolley orderlies/stretcher-bearers, kitchen, laundry, cleaning, etc.). This is largely a coordination function. Any clinical role or supervision by the head nurse will depend on the particular circumstances of the hospital involved.

9.6.4 The team

Whether one person fills more than one leadership function will depend on the availability of competent personnel. The triage team leader may be the same as the triage officer in a small hospital; in a larger facility, this coordination function is probably best assigned to someone else, for instance an administrative officer or the head nurse. In a very large hospital, three different people should assume the three functions.

ICRC EXPERIENCE

What is the reality in the field?

In many small rural hospitals, the only surgeon is also the director. This one person often becomes the triage team leader and triage officer. After performing triage, the surgeon and the only anaesthesia nurse go to the operating theatre. A new influx of casualties arrives. Who then performs triage? Who continues the necessary coordination tasks? Either a general practitioner or the head nurse. This should be decided upon beforehand and be part of the hospital's disaster plan.

The ICRC supports the Somali Red Crescent Society in running the Keysaney Hospital in north Mogadishu, where triage has been routinely practised from 1992 until the time of writing. The triage officer in the emergency room is an "experienced female nurse, because she is the only one who can organize things".

9.6.5 Triage groups/nursing teams

Such teams should be formed for the triage area, with responsibility for the following tasks.

1. Setting up intravenous lines and taking blood for grouping and cross-matching.
2. Administration of tetanus prophylaxis, antibiotics, analgesics and other medication as prescribed.
3. Dressing wounds and splinting fractures.
4. Bladder catheterization, if indicated.
5. Arranging for an orderly flow of categorized casualties to the operating theatre or to designated areas for continuing resuscitation, delayed, or minimal treatment.

The disaster plan may be nominative, designating by name the triage officer and triage groups: who is in charge of setting up i.v. lines, dressing wounds, giving analgesics and antibiotics, etc.; *or* it may designate the function (ER doctor 1, ER nurse 3) regardless of who is on the roster that day (Figure 9.5). This depends on the staffing system, and availability of personnel, of the particular hospital.

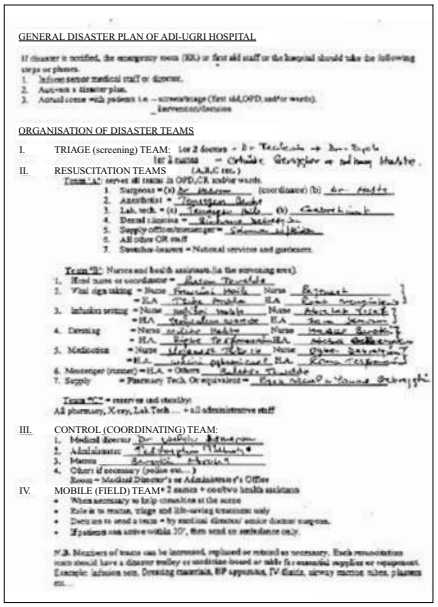
GENERAL DISASTER PLAN OF ADI-UGRI HOSPITAL

ORGANISATION OF DISASTER TEAMS

I. TRIAGE (screening) TEAM:

II. RESUSCITATION TEAMS

III. CONTROL (COORDINATING) TEAM:

IV. MOBILE (FIELD) TEAM

M. Ghebrehiwet / MoPH Eritrea

Figure 9.5

Example of a nominative disaster/triage plan in a small rural hospital.

9.6.6 Surgeons and theatre personnel

These members of staff should have prepared the operating theatre and be ready and waiting for the casualties. The surgeon in theatre will see patients he or she has not previously examined and who may not have a name or signed consent form. Outside the mass casualty situation this would be regarded as negligent practice, and so another "mental adjustment" is required.

9.6.7 Rest and relief

Early consideration should be given to the *rest and relief* of staff. Triage is not an everyday routine activity. The shift system may need to be changed in order to deal with the increased workload: 2 twelve-hour shifts instead of 3 eight-hour ones, for example.

During a crisis the hospital team will be under severe emotional and physical stress. All staff members will need to pace themselves in order to perform efficiently and maintain a professional attitude. As mentioned before, during armed conflict there may be a mass arrival of new casualties every day for weeks on end. It is not possible to foresee how long the influx will last.

ICRC EXPERIENCE

At the beginning of the fighting in Monrovia in 2003, the surgical team worked 48 hours non-stop, without sleep: they were not able to continue operating. Subsequently, a strict 18-hour operating schedule was instituted: the team was able to maintain this schedule for three months.

A triage situation is tiring for everyone. Some, out of a misplaced sense of professional duty, find it difficult to accept that they *need* to rest; rest must be insisted upon. Working for extended periods without sleep only results in hospital staff no longer being able to adequately treat the new patients who continue to arrive.

9.7 Space

During an influx of mass casualties the various hospital departments must be re-arranged according to the pre-decided plan. Besides the original hospital site, any alternative sites (building, underground shelter, etc.) must be included in the plan, should the hospital require evacuation for reasons of security. The equivalent in cases of natural disaster is the destruction of the hospital premises and/or its access routes (earthquake, landslide, tsunami, etc.).

R. Aburabi / ICRC

Figure 9.6.1
Alternative triage department in an unused building: empty.

R. Aburabi / ICRC

Figure 9.6.2
The same premises: full.

The usual emergency reception/admissions room may not be large enough to accommodate a large influx of wounded. A large and extendable area capable of serving as a triage department should be pre-selected. The area should be cleared of all inpatients, and be large enough to permit easy movement of casualties and medical personnel.

In certain contexts, for religious and cultural reasons, it may be necessary to have separate male and female areas for incoming mass casualties. This must be taken into account during planning.

Once triaged, patients should be transferred out of the triage department to the operating theatre or designated wards.

Different units or wards should be identified for the different triage categories. Category I patients, requiring resuscitation and immediate surgery, might be put in a pre-operative intensive care unit near the operating theatre while waiting for an operating table to free up. Resuscitation with i.v. fluids can continue here under close supervision. The Category II cohort of patients who require surgery but not on an urgent basis can be grouped in a special ward where they can receive treatment (antibiotics, analgesics, i.v. fluids, dressings) under observation while awaiting surgery.

Category III patients suffering superficial wounds could be sent to the outpatient department or an area for convalescence outside the surgical hospital. This group is composed of many patients who are lucid, frightened (because of the shooting and shelling), panicky, and often in pain. Organization and basic staffing of this space are important in order to identify and isolate such patients, give them rapid treatment, and then discharge them.

Finally, a quiet, secluded room should be made available for Category IV patients with very severe wounds who should be left to die peacefully with dignity. An i.v. line should be put up and analgesics given if appropriate.

Some provisions should be made for the well-regulated visit of friends and relatives to all seriously-injured patients. This comes under the necessary security measures to be implemented (see Section 9.13).

9.8 Equipment and supplies

A large number of stretchers or trolleys are needed at the hospital entrance during triage, to accommodate the casualties left by incoming ambulances. Blankets and sheets are required in the triage area, as are lines on which to hang infusions. Complete sets of supplies for triage should be prepared and stored in boxes or trunks which can easily be carried to the triage area from an accessible storage space.

Triage boxes should include:
- disposable latex or plastic gloves;
- venepuncture equipment;
- i.v. fluids;
- dressings, bandages, scissors;
- catheters, naso-gastric tubes;
- drugs which are likely to be needed.

Needless to say, these triage boxes should be inspected regularly for the expiration date of their components. The drugs will probably need to be stored separately. Stocks should include the appropriate antibiotics and analgesics, as well as tetanus toxoid and anti-tetanus serum.

A reasonable stock of items that will be required for the management of extra cases in the wards should be maintained. There should be adequate numbers of bed blocks for elevation of the foot of the bed, supplies of i.v. bottle holders, and supplies of dressings, plaster of Paris, splints and traction apparatus.

Again, depending on the particular circumstances of the hospital, the pharmacy may hold designated emergency triage stocks or not. The re-supply of the hospital may be disrupted because combat has interrupted the ordinary lines of supply.

Standard documentation charts and patient folders, *each with a unique number*, should be prepared and maintained in readiness. Each folder should include a triage/admission form, a fluid balance chart, and laboratory and X-ray request forms.

9.9 Infrastructure

Plans must be made to ensure adequate supplies of water, sufficient electricity, proper sanitation, and the disposal of waste. This may include special reserves of fuel for electric generators. Spare parts are also important to keep in stock; things tend to break down in the middle of an emergency situation.

The designation of roles, responsibilities and tasks is not limited to the medical personnel. Technicians and maintenance workers to run the generators and ensure the water supply must be put on a special shift system so as to be available at all times.

9.10 Services

Hospital personnel, patients and their relatives, and volunteers must all eat. Hospital linen must be washed and theatre linen re-sterilized. The kitchen, cafeteria and laundry staff and facilities must all be included in the disaster plan. Relatives are a great nuisance in the triage area but their energies can be harnessed for the general good. They may be directed to give blood and engaged as volunteer stretcher-bearers, carriers of water, cleaners and kitchen staff, etc.

9.11 Training

The hospital team should regularly practise different triage scenarios, on its own and as part of any national disaster or conflict-preparedness plan. Volunteers from the National Red Cross or Red Crescent Society, and their first aiders, may be mobilized to work in the hospital and/or play the role of the injured.

Clinical protocols and guidelines for triage and patient management must be standardized and understood by all doctors and nurses. This helps to avoid confusion and disagreements under tense and tiring circumstances.

After every triage incident and when feasible, a general meeting of the staff should be held to discuss what went well and what went wrong. This evaluation permits the fine-tuning of the triage plan, to improve it for the next mass influx.

Some individuals or even the whole team may have found the experience very stressful. An open, frank discussion about what happened and why certain decisions were made can be very beneficial, and even therapeutic.

9.12 Communication

If the hospital is part of an integrated health system, then a means of coordination and communication with other health facilities may make it possible to transfer the wounded from the overwhelmed hospital to another which has received comparatively few patients. Alternatively, other health facilities may be able to provide help in the form of additional personnel.

The plan should include the means to contact staff who are off duty, bearing in mind that if combat is occurring in the area, hospital personnel may have difficulty in getting to work. Mobile telephone systems tend to stop functioning (or are stopped by certain authorities) during urban fighting or troubles.

In an age of immediate mass communication contact with the media is inevitable, and what is happening in the hospital can be easily exploited for political purposes. The patients, hospital and staff are all protected under IHL. Obviously, military and governmental civilian hospitals are not "politically" neutral and cannot be expected to be. However, hospital personnel must act and speak publicly in accordance with their "medical" neutrality and impartiality. An "official spokesperson" for the hospital should be duly selected.

9.13 Security

Last, but certainly not least, is the safety and security of the hospital premises, patients and staff. When armed conflict results in mass casualties every wounded person transported to the hospital is accompanied, as a general rule, by two to four friends, relatives, comrades-in-arms, or bystanders who have helped in the evacuation or transport. Curious onlookers may try to enter the hospital as well. The civilian population may be in a state of panic and consider the hospital a safe place. This well-known phenomenon is called the "convergence reaction". People's fear and excitement add to the confusion and danger, especially if many in the crowd happen to be armed. Limiting the number of people entering the hospital reduces the confusion considerably.

One major factor which creates the convergence reaction and confusion is the lack of proper pre-hospital triage, as is often seen in an urban context. As mentioned in Chapter 7, the waves of ambulances transporting casualties evacuate the dead and badly mutilated first, the superficially injured, who are shouting to call attention to themselves, second, and the seriously wounded who require emergency surgery last.

Security must be ensured by having guards posted at the gate of the hospital. Only wounded persons, possibly accompanied by a close relative according to local cultural traditions, should be allowed to enter. All weapons must be left outside the gate.

Access to the triage area should be blocked off, and a guard posted here as well to keep other patients and onlookers from mingling with new casualties.

The large numbers of relatives and friends who inevitably rush to the hospital must be prevented from crowding the premises and impeding the work of the medical staff. A system for well-regulated visits by patients' friends and relatives must be set up to reduce the possibility of "friction" occurring. Under certain circumstances, board and lodging close to the hospital compound must also be organized and provided.

Whether the guards are armed or not will depend on the circumstances of the country concerned. IHL does not prohibit armed guards if their purpose is to maintain order and protect patients and personnel.

In extreme circumstances, the hospital can organize a "security lock"[4] at the hospital or triage area entrance. This forced channel allows for better crowd control, especially of armed men.

ICRC EXPERIENCE

The ICRC hospital in Kabul, Afghanistan, used commercial, steel containers with doors cut into the ends at the hospital entrance in 1992 to create a "tunnel" that allowed the filtering of all who entered. The width of the door was calculated to allow for the passage of a carried stretcher only.

The ICRC – Somali Red Crescent Hospital in Mogadishu, Somalia, in 1992 was located in a converted prison. A first set of guarded gates led to a triage area with a large tent. A second set of gates separated the triage tent from the hospital compound itself, permitting controlled entry to the hospital.

The ICRC hospital team working in the JFK Memorial Hospital in Monrovia, Liberia, in 2003 set up a so-called "triage tent" just outside the fence and gate entrance to the hospital. No medical triage was actually performed there, it was more of a "disarmament tent". The tent provided a space for disarming any combatants, stripping the wounded of their soiled clothing, weapons and ammunition, and performing basic first aid. The wounded were then carried into the triage department – unarmed – on stretchers where the clinical triage was performed.

R. Aburabi / ICRC

Figure 9.7.1
Monrovia "triage tent" in front of the hospital.

R. Aburabi / ICRC

Figure 9.7.2
"Triage tent" seen from the outside.

R. Aburabi / ICRC

Figure 9.7.3
Volunteers inside "triage tent" before the action.

R. Aburabi / ICRC

Figure 9.7.4
Volunteers after the action.

4 A lock is a short section of a canal with gates at each end which can be opened or closed to change the water level, used for raising and lowering boats going through the canal. This permits control of the movement between areas at different water levels. By analogy, a security lock permits the control of people's movements between different hospital areas.

9.14 Summary of triage theory and philosophy: sorting by priority

The triage process has three components.

1. Clinical assessment to determine which patients take priority for the limited surgical time and resources available.
2. The organization and management involved in admitting large numbers of wounded to the hospital.
3. Re-assessment of the functioning of triage and its adaptation to the number of incoming wounded.

9.14.1 Triage system: a simple emergency plan organizing the personnel, space, infrastructure, equipment, and supplies

The sudden arrival of large numbers of casualties may occur at any time. Prior planning and training prevents poor performance. Unless a plan exists for the reception and triage of mass casualties, chaos will result. Hospital staff should be prepared, however, to improvise when faced with a new evolving situation.

Evaluation of hospital capacity is essential in emergency planning.

An emergency plan does not cost money; organization does not cost money. They cost time, effort, discipline and motivation.

9.14.2 Emergency hospital disaster triage plans differ and no two triage scenarios are the same

Hospital teams must regularly practise receiving a mass influx of wounded, resulting from armed conflict or natural disaster. Simulation exercises should be organized covering a variety of scenarios, and adapted to the particular circumstances of the hospital involved.

9.14.3 "Best for most" policy

Priority patients are those with a fair chance of "good survival" with the least amount of surgical work.

Triage is essential to put some order into a chaotic situation.

However good the disaster plan and extensive the training, a mass influx of casualties is *always* stressful and attended by confusion. Flexibility and adaptability of the hospital team are important. Triage is not a series of rules. It has a logic and philosophy that must be adapted to each particular situation. Triage is not a simple science; it is an art.

Please note:

Further practical information on the organization of hospital management in a triage situation is available in the ICRC manual *Hospitals for War-Wounded: A Practical Guide for Setting up and Running a Surgical Hospital in an Area of Armed Conflict* (see Selected bibliography).

In a military context, all standard war surgery manuals written by and for armed forces deal with the organization and implementation of triage under military constraints.

ANNEX 9. A Sample triage card

Triage Card No: _____

Name: _____

Coming from: _____ Male / Female / Age _____

Date: _____ Time: _____

GSW: ☐ Mine: ☐ Fragment: ☐ Blast: ☐ Burn: ☐ Other: _____

Time since injury: _____

General condition:

Pulse: _____ BP: _____ Resp. rate: _____ Consciousness: _____

Wound assessment:

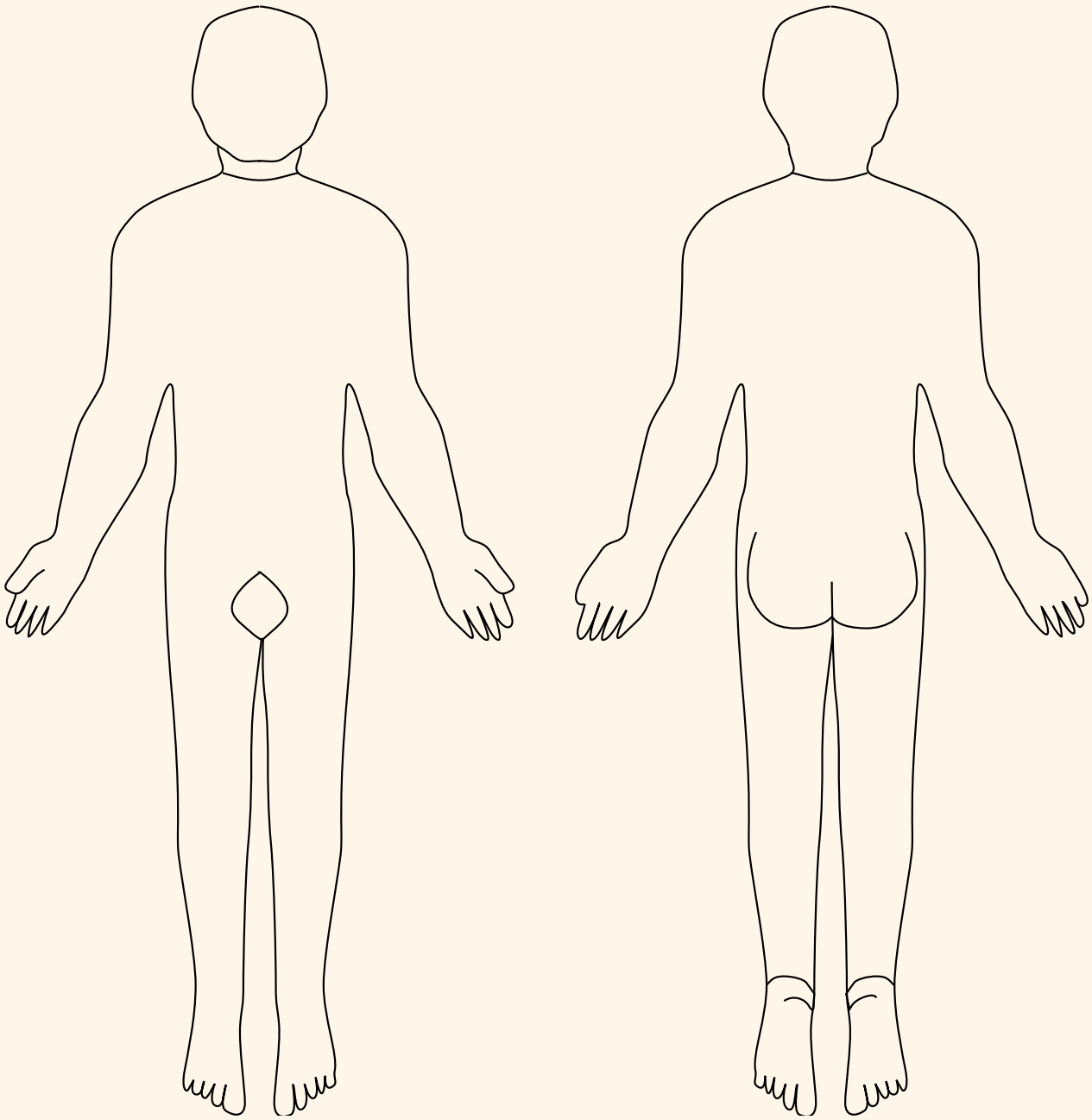

Triage Category:

I | II | III | IV

ANNEX 9. B Hospital emergency plan for a mass influx of wounded

HOSPITAL EMERGENCY PLAN FOR A MASS INFLUX OF WOUNDED										
PHASE	Influx of wounded (number of patients)	Secrity + Porters/ stretcher - bearers	TTL / TO / HN	Triage Area (ER + OPD + Physio room)	OT	Wards	Administration	Laboratory X-ray	Kitchen Laundry Tailor	Transport
1	**1 – 10**	Routine	HN informed OT informed	Routine	Stop routine Inform wards	Routine No more elective patients sent to OT	Routine	Routine	Routine	Routine
2	**11 – 20**	Extra guards at patients' entrance	TTL informed TO: INITIATE TRIAGE HN: Inform wards, OT, administration	Request triage material and staff	Keep staff on duty Check supplies	Update bed situation Inform HN/TTL	"	"	"	Standby ambulance
3	**21 – 30**	Visitors leave Visiting hours suspended	Call in on-call nurses Ensure that all staff have breaks	Call in extra staff Physio staff on standby to help in triage or in wards	Extra surgical team and staff standby	"	Keep pharmacy staff on duty	"	Prepare food and drinks for OT staff if needed	"
4	**31 – 40**	"	HN: Reassess bed situation TTL/:Monitor OT situation	"	"	Assess need for additional staff Inform HN	Keep purchaser and storekeeper on duty	Call in extra staff if needed	Laundry: priority is OT linen	Standby bus for staff transport
5	**41 – 50**	Call in extra porters/ stretcher-bearers	TTL/HN: REASSESS Define needs for staff, beds, supplies	"	Check supplies Open reserve linen stocks	"	Monitor needs for staff	"	Open emergency food stocks	"
6	**> 50**	"	REVIEW AGAIN	"	"	"	Provide resting area and sleep accommodation for staff on duty	"	Provide food and drinks for all staff on request if needed	"

ER: emergency reception room

HN: Head Nurse

OPD: outpatient department

OT: operating theatre + sterilization

TO: Triage officer

TTL: Triage Team Leader

Chapter 10

SURGICAL MANAGEMENT OF WAR WOUNDS

10. SURGICAL MANAGEMENT OF WAR WOUNDS

10.1 Introduction 213

10.2 Complete examination 214

10.3 Preparation of the patient 215

10.4 Examination of the wound 216

10.5 Surgical treatment 216
10.5.1 Technique of wound debridement 217
10.5.2 Skin 218
10.5.3 Subcutaneous fat 218
10.5.4 Fascia and aponeurosis 218
10.5.5 Muscle 219
10.5.6 Haematoma 220
10.5.7 Bone and periosteum 220
10.5.8 Arteries, nerves and tendons 220

10.6 Retained bullets and fragments 221

10.7 Final look and haemostasis 222

10.8 Wound excision: the exceptions 223
10.8.1 Management of minor Grade 1 wounds 223
10.8.2 Serial debridement 223

10.9 Leaving the wound open: the exceptions 224
10.9.1 Head, neck, scalp and genitals 224
10.9.2 Soft tissues of the chest (sucking chest wound) 224
10.9.3 Soft tissues of the abdominal wall 224
10.9.4 Hand 224
10.9.5 Joints 224
10.9.6 Blood vessels 224

10.10 Dressings 225
10.10.1 The exceptions 226

10.11 Anti-tetanus, antibiotics, and analgesia 226

10.12 Post-operative care 226

10.1 Introduction

There are many factors that determine the final outcome of the management of war-wounded patients:

- the actual injury – the clinical significance of the wound is a function of the severity of tissue damage and the anatomic structures involved i.e. size and site;
- the general condition of the patient – nutritional status, dehydration, concomitant diseases, host resistance, etc. ;
- pre-hospital care: protection, shelter, first aid, triage, evacuation time;
- resuscitation, especially in the presence of haemorrhagic shock;
- hospital triage;
- surgery;
- post-operative nursing care;
- physiotherapy and rehabilitation.

However, the single most important act that the surgeon performs for the great majority of casualties is wound excision or debridement.[1]

The basic principles of trauma wound management have been known for a considerable period of time. Ibn Sinna commented on this (Figure 10.1), as did Alexander Fleming one thousand years later.

Å. Molde / ICRC

Figure 10.1

"Damaged tissues must be removed in time." Ibn Sinna, *Qanun fi al-Tib*.[2]

> "The severity of these [war] wound infections is merely the result of the very extensive destruction of the tissues by the projectile, thus furnishing an admirable culture medium for the bacteria out of reach of the natural protective forces of the body, and if it were possible for the surgeon to remove completely this dead tissue I am quite sure the infections would sink into insignificance."
>
> Alexander Fleming[3]

K. Barrand / ICRC

Figure 10.2.1

Wound sutured primarily to "close the hole", without debridement.

Often, ICRC health personnel have witnessed inexperienced medical staff in a low-income country dealing with severely war-wounded patients for the first time. Their initial reaction is to try to stop the bleeding by "closing the holes". If suture material is available, they stitch the wounds shut without excising dead tissue; if not, they plug the wounds with gauze compresses. Soon, infection develops. This is treated by a frequent change of dressing and antibiotic anarchy, wasting already insufficient hospital supplies. The end of the story may be a slow recovery with a major disability, or, more frequently, major complications often leading to the death of the patient accompanied by a deep sense of frustration among the medical staff.

K. Barrand / ICRC

Figure 10.2.2

Sutures removed: pus pours out.

The basic management of war wounds is an art based on sound scientific principles. A good knowledge of wound ballistics helps the surgeon assess more accurately the extent of tissue damage and of the necessary surgery; it will not explain every wound or prescribe specific treatment for every patient. In only a few cases will knowledge of the specific causative weapon be of any clinical use.

Wound ballistics teaches us that the cavitation effect of projectiles conveys pathogenic organisms, pieces of dirty skin and clothing and dust into the depths of the wound. Anti-personnel blast mines drive pieces of the shoe or bones of the foot,

1 "*Débridement*" is a French term originally used to mean the removal of a constriction, or to unbridle, and thus to promote drainage and relieve tension by incision. The English word "excision" means to cut away or cut out. A confusion of terminology occurred at the Inter-Allied Surgical Conference for the Study of War Wounds held in Paris in 1917, and the historic French expression came to mean a wound excision – the removal of dead tissue – for English-speaking surgeons.
A further misunderstanding may be caused by the fact that excision, when used with reference to cancer surgery, tends to mean "*excision en bloc*" with a wide margin of healthy, surrounding tissue. In this book, the expressions "excision" and "debridement" are used interchangeably and are understood to mean the cutting away of dead or devitalized tissue.

2 Ibn Sinna – Avicenna (980 – 1036 CE), Persian physician and philosopher, author of *Qanun fi al-Tib* (The Laws of Medicine).

3 Alexander Fleming (1881 – 1955), British physician, discoverer of penicillin.
Fleming A. On the bacteriology of septic wounds. *Lancet* 1915; **186**: 638 – 643.

gravel, soil, leaves and grass, and pieces of the weapon into the proximal tissues. Although the wounds are contaminated, infection does not set in for the first 6 – 8 hours. Ideally, therefore, war wounds should be debrided within six hours; this is not often the case.

ICRC

Figure 10.3

Anti-personnel landmine injury: war wounds are dirty and contaminated.

Old lessons for new surgeons

War wounds are dirty and contaminated, from the moment of injury.

Whatever the weapon, a thorough examination of the patient and the wound(s) is essential.

Examine the patient:
- initial examination and resuscitation;
- complete examination to identify all open wounds and any closed injuries.

Examine the wounds and damaged organs:
- to operate or not to operate;
- to establish priorities if more than one surgical procedure is required;
- to plan the operation.

Please note:

The treatment of mismanaged and neglected wounds is dealt with in Chapter 12.

10.2 Complete examination

All wounds involve soft tissues, and many of them will be complicated by damage to other structures. *War wounds are often multiple* and the *pathology often multiple*: a bomb explosion can simultaneously cause primary blast injury, penetrating metallic fragments, blunt trauma and burns. Close and complete examination of the patient is important to determine the site and size of *all* wounds present, in order to determine which wounds require surgery and to best plan the sequence of operations.

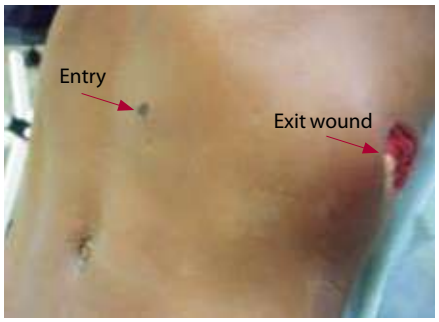

H. Nasreddine / ICRC

Figure 10.4

Entry and exit thoraco-abdominal wounds: think anatomy!

As mentioned in Chapter 8 the surgeon should try to identify the path of any projectile. The wound may perforate the body part through-and-through, or the projectile may be retained without an exit. Entry and exit wounds may be at the same or different levels. The surgeon should visualize the anatomic structures that lie along the probable projectile trajectory, which may include a vital cavity: brain, thorax, or abdomen. The surgeon must think "anatomy".

The examination may reveal a tense, tender swelling in a through-and-through wound to the calf, which indicates haematoma and severe muscle damage; by contrast, the same exit and entry wounds with soft and relaxed tissues signify minimum injury. The surgeon must think "pathology".

Fractures and peripheral vascular and nerve damage should be identified. If there has been delay in evacuation, a *bruit* or vascular murmur may be heard and a "thrill" felt, signifying a pseudoaneurysm or arteriovenous fistula. Peripheral motor function and sensation are tested. Neurapraxia is more common than a severed nerve.

Most wounds involving vital organs should already have been identified during the initial examination for airway, breathing and circulation. Nonetheless, the full-body palpation of a complete examination may reveal the small entry wound of a tiny fragment that has penetrated the meninges, pleura or peritoneum without immediately affecting the vital functions.

A full assessment may require X-rays as well – with the exception of through-and-through soft-tissue wounds, which do not require them. One body region above and below any entry or exit wound should be filmed. Deformation or fragmentation of a

bullet, showing the characteristic "shower of lead particles", is a good indication of severe tissue damage. (Figures 10.5, 3.35 and 4.5). Many foreign materials, it should be noted, are *not* radio-opaque: pieces of shoe or clothing, mud, leaves and grass, plastic fragments from some landmines, for instance. On the other hand, it may be difficult to differenciate a projectile from the normal radio-opacity of some anatomic features (Figures 8.4.2 and 14.9.1). Major bone comminution also implies severe soft-tissue damage. Having radiography helps, but is not an absolute necessity for diagnosing fractures.[4]

Note that the presence of air in tissues on X-ray does not necessarily mean the gas of clostridial myositis. The temporary cavity associated with high velocity missiles often leaves behind palpable and radiographic intrafascial and intramuscular air in healthy tissue some distance from the wound. This usually indicates severe tissue damage. The diagnosis of gas gangrene is a clinical one and the radiograph is not pathognomonic: *vide* Figures 10.6 and 13.2.

More sophisticated diagnostic means can be used, of course, depending on the equipment and professional expertise of the hospital in question.

R. Coupland / ICRC

Figure 10.5

Massive comminution of humerus showing characteristic "shower of lead particles" after bullet fragmentation.

ICRC

Figure 10.6

X-ray showing air in tissues in a patient not suffering from gas gangrene.

10.3 Preparation of the patient

Not only are war wounds dirty and contaminated but battlefields are dirty places as well. The wounded do not have access to basic sanitary facilities and every precaution should be taken to meet fundamental hygiene standards. The majority of casualties have wounds to the limbs and are haemodynamically stable. All stable patients should be put through a warm shower upon admission; dressings are changed as necessary for examination and triage. Only critical cases go straight to the operating theatre.

In the operating theatre, the surgeon should prepare the patient's position – in coordination with the anaesthetist and nursing staff – for the sequence of the operation before starting. A pneumatic tourniquet is very useful for limb wounds, especially before removal of a field dressing on an obviously serious injury.

In the case of multiple soft-tissue wounds, those on the posterior aspect of the trunk and limbs should be dealt with *before* those on the anterior aspect, unless there is clinical evidence of a vascular injury. Many wounds to the back have been *"forgotten"* after much work on anteriorly situated wounds, and it is easier for the anaesthetist to finish off the operation with the patient supine.

Under anaesthesia, dressings and splints are carefully removed. The skin over a large surrounding area, including the whole circumference of the limb or torso, is cleansed with soap and water and a brush, shaved, dried, and then painted with povidone iodine. The wound is irrigated copiously.

Sterile sheets are put in place. Drapes with holes should only be used for the smallest and most superficial wounds. Most wounds will have to be extended and the operative field enlarged as a result.

The wound is then thoroughly irrigated again to remove surface dirt and debris. Under optimal conditions, sterile normal saline should be used. Under conditions of limited resources, potable water from the tap can be resorted to: "If you can drink it, you can put it into a dirty wound". Groundwater may be treated with sodium hypochlorite to reach a concentration of 0.025 % (5 ml of bleach in one litre water) if necessary.

Only in extreme cases of impending asphyxia or exsanguinating haemorrhage should the basic rules of hygiene and sterile technique be broken.

ICRC

Figure 10.7

Washing an anti-personnel landmine injury prior to surgery.

4 Many ICRC surgical teams have had to operate in emergency situations with *no* X-ray equipment.

ICRC

Figure 10.8
Finger exploration of the wound.

10.4 Examination of the wound

Following the initial careful examination of the patient, full assessment of the wound(s) may require finger exploration in theatre under anaesthesia. In this era of HIV and Hepatitis B and C, the surgeon should take great care not to suffer injury from sharp bony edges.

Old lessons for new surgeons

The best instrument for wound exploration is the surgeon's (protected) finger.

The surgical work required to treat a wound depends on its site and size: i.e. the location and amount of tissue damage. Knowledge of ballistics allows only for an understanding of wounding possibilities, but knowing that certain weapons have greater wounding potential does not help the surgeon in this assessment. With the exception of anti-personnel blast mines, even knowing which weapon was used is often of no practical use. Many wounds are the result of full metal jacket bullets that ricochet, creating large injuries similar to a dum-dum bullet (see Chapter 3).

The most significant component of tissue damage is the permanent wound cavity, yet energy loss by the projectile and tissue disruption along the track is not uniform, resulting in a "mosaic of tissue damage". Injury can also occur well beyond this, owing to stretch and shear during temporary cavitation: late ischaemic necrosis in the wall of hollow organs; or intimal detachment and vascular thrombosis of vessels that still look intact from the outside. The time between injury and treatment and the possible beginning of sepsis, must also be taken into account.

War wounds are never identical, even if caused by the same weapon.

There is an art to the assessment of tissue damage that is learnt largely through experience. It is important neither to under-treat – causing sepsis and even death – nor to over-treat – resulting in excess loss of normal tissue and increased disability.

This art does have a scientific basis, nonetheless. We have seen that the Red Cross Wound Score classifies penetrating weapon wounds into different categories according to the degree of tissue damage (Grade) and the structures injured (Type). This Score helps determine the extent of surgery necessary. Careful examination is important because not all wounds require surgical excision.

10.5 Surgical treatment

The surgical treatment of a major wound is performed in two stages:

- *wound debridement*, leaving the lesion wide open, without any suturing of skin or deep structures;
- *delayed primary closure* 4 – 5 days later.

The surgical treatment of most war wounds is a staged process involving two main procedures, the first being wound debridement or excision. The resulting wound is left unsutured. The open wound then undergoes delayed primary closure (DPC) after 4 – 7 days, once the exudative phase of trauma inflammation has subsided and the proliferative phase begun. In practice, DPC is usually performed after 4 – 5 days. Closure may be by simple direct suture or require sophisticated techniques of grafting and reconstruction (see Chapter 11).

10.5.1 Technique of wound debridement

Basic principles of wound debridement

1. Stop haemorrhage.
2. Make adequate skin incisions and fasciotomies.
3. Remove dead and severely contaminated tissues to prevent/control infection.
4. Leave the wound open – unsutured.
5. Re-establish physiological function.
6. Handle gently and treat the tissues with respect, as always.

Control of haemorrhage is the first priority. In a large wound with heavy bleeding, blind clamping in the depths of a blood-filled cavity is mentioned only to be condemned. Direct local pressure should be exerted while proximal and distal control of the vessels by standard surgical exposure is performed.

For management of vascular injuries, see Volume 2.

After haemorrhage, the greatest danger to the patient is sepsis. Surgical infection requires a culture medium, which in the case of weapon wounds is a mixture of dead muscle, haematoma, bone fragments, dirty skin, foreign material (pieces of cloth or shoe, mud, gravel, leaves, dust, the missile, etc.), and, at times, extrinsic bone (bone fragments which come from the wound in another part of the body or another person and whose penetration, as secondary missiles, then creates a new wound). Injury caused by extrinsic bone fragments is seen with bomb explosions and APM.

Wound debridement is the process whereby this dead and damaged tissue, grossly contaminated with bacteria and debris, is completely cut away. This leaves an area of healthy tissue with a good blood supply and capable of combating residual surface contamination, provided the wound is not closed. Excess excision of healthy tissue, however, will increase the potential for deformity and disability.

Good vascularisation and oxygenation of the wound requires relief of tension within the wound and good drainage of the inflammatory exudate. This is provided by adequate incision of the skin and fascia, and leaving the wound unsutured.

Wound debridement involves incision and excision.

A basic instrument set is all that is needed in the vast majority of cases: scalpel, Metzenbaum (tissue) and Mayo (suture) scissors, toothed dissection forceps, bone curette, six haemostats, and retractors. Diathermy is not necessary; absorbable ligature material is preferable.

For the young surgeon, or one without experience of war wounds, it is best to excise the wound layer-by-anatomic-layer, proceeding from superficial tissues to deeper ones in order to best visualize the anatomy and pathology. At all times the surgeon should anticipate the anatomic structures present. Tissue damage, haematoma, and oedema can easily *alter the visible anatomy*, and camouflage important structures.

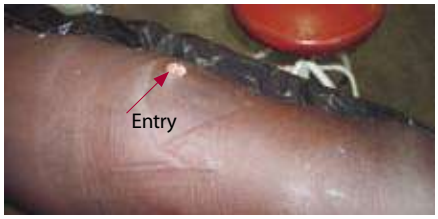

F. Jamet / ICRC

Figure 10.9.1

Patient X: entry wound on anterior aspect of thigh.

F. Jamet / ICRC

Figure 10.9.2

Patient X: longitudinal extension of the skin incision.

F. Jamet / ICRC

Figure 10.9.3

Patient X: opening of the fascia throughout the length of the skin incision. Note the contused and necrotic muscles.

10.5.2 Skin

Skin is elastic, with a good blood supply, very resistant to damage and remarkably viable. It should be treated conservatively. Only skin that is grossly pulped should be cut away. Usually no more than 2 – 3 mm of the skin edge need be removed at both entry and exit sites.

This is followed by as generous an incision as necessary of healthy skin in order to gain access to the depths of the wound (Figure 10.9.2). Small entry and exit wounds may hide considerable internal injury. The most common error is to attempt to make a wound excision through a small entry or exit wound. In limbs, the incision should be made in the longitudinal axis, but not over subcutaneous bone, and at flexion creases it should curve in the usual way.

The commonest mistake is to attempt a wound excision through small entry or exit holes or to make incisions too short.

This extension of the skin wound not only permits better visualization, but also allows for the proper decompression of deeper tissues and their subsequent drainage.

10.5.3 Subcutaneous fat

Subcutaneous fat has a poor blood supply and is sticky, easily holding heavy contamination. This layer should be excised generously, 2 – 3 cm all the way around the original traumatic wound.

10.5.4 Fascia and aponeurosis

Shredded fascia should also be trimmed. Large amounts of damaged muscle may lie underneath a small hole in the fascia; therefore, the muscular compartment should be opened up by a large incision of the deep fascia parallel to the muscle fibres along the entire length of the skin incision (Figure 10.9.3). This essential step allows wide and deep retraction to expose the depths of the wound. It may be necessary to add transverse cuts to the deep fascia to improve access.

Post-traumatic oedema of the wound can easily cause compartment syndrome, compromising the local circulation with resultant necrosis of the muscles. The fascial incision is left open to allow oedematous and congested muscle to swell without tension, so as to avoid interference with the blood supply and to promote drainage of the inflammatory exudate and haematoma.

After haemorrhage control, the relief of tissue tension is the most important step in wound debridement.

Deep to the fascia, the gloved finger is the best and gentlest probe to follow the track and estimate the extent of damage. Again, particular attention should be paid to the sharp edges of any fractured bone.

Fasciotomy

Compartment syndrome can occur in any fascial space, but is seen most commonly in the lower leg. Great care should be taken when dealing with any penetrating wound below the knee, with or without tibial fracture.

If there is any suspicion of compartment syndrome, decompression must be performed without delay.

For information on the technique of fasciotomy of the lower leg, see Volume 2.

10.5.5 Muscle

Dead muscle is the *ideal medium* for the development of *clostridial* infection leading to gas gangrene or tetanus as well as for the growth of many other bacteria. The track of the missile through the muscles must be opened up, layer by layer, to be properly visualized. It is vital that all grossly contaminated, obviously necrotic and detached muscle lining the track be excised.

All dead or heavily contaminated muscle must be removed.

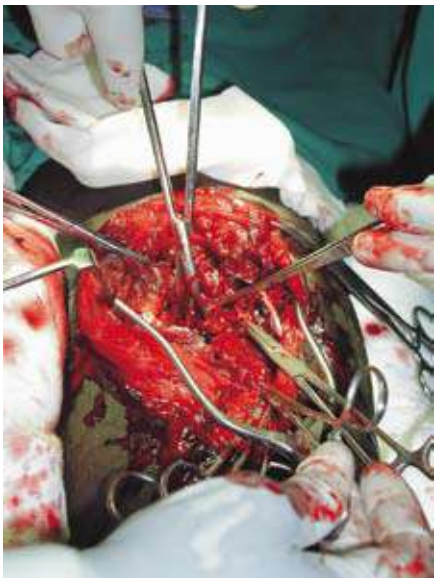

F. Jamet / ICRC

Figure 10.9.4

Patient X: wound cavity completely opened and excised.

Warning!

Muscles that are completely transected contract *away* from the wound cavity. The retracted muscle bellies must be found for inspection and excision. During wound exploration, division of uninjured muscle in a transverse direction should be avoided.

However, not all injured muscle will undergo necrosis. How to differentiate between muscle that is injured but will heal from that which is not viable? Classically, reference is made to the four Cs of muscle viability:

- colour,
- consistency,
- contractility, and
- capillary bleeding.

All muscle that is not healthy and red, that does not contract when pinched or bleed when cut, must be excised until healthy, contractile, bleeding muscle is found.

However, confusion may arise because of certain pathological changes described below.

- As mentioned in Chapter 3, wound ballistics studies have shown that there is an intense but transient vasoconstriction lasting several hours, followed by the reactive local hyperaemia of the inflammatory response. Consequently, it is not always necessary to "cut until it bleeds". The surgeon must therefore take into account the time since wounding.
- However, minimal excision is often followed by necrosis of the muscle when the wound is looked at several days later. What is irreversible is not always immediately apparent. The wound evolves and these changes demonstrate the different stages in the "life-history" of the wound[5] and should be well understood by the surgeon.
- The surgeon should also understand that shock or the use of a tourniquet can change the colour or the capacity of muscle to bleed and that contractility is affected by hypothermia and paralysing anaesthetic agents.

The criteria of the 4 Cs are very subjective and a function of the surgeon's experience, but are nonetheless the best clinical measure available. The colour and consistency – texture – of the muscle should be noted. To test for contractility, 2 cm^3 lumps of muscle should be picked up and pinched with a forceps and, if they do not contract, excised using scissors or knife. Note if the cut edge of the muscle lump bleeds. Cutting away lumps larger than 2 cm^3 may result in the inadvertent removal of healthy tissue. Meticulous technique is the key.

Intermuscular fascia should be excised if blood-stained and contaminated.

5 Pearson W. Important principles in the drainage and treatment of wounds. *Lancet* 1917; **189**: 445 – 450.

10.5.6 Haematoma

The presence of a large haematoma generally implies that a major vessel has been damaged. Dislodging the haematoma can result in sudden heavy blood loss. It is wise to be prepared for vascular control before a haematoma is evacuated. If working under a pneumatic tourniquet, the anatomical structures must be clearly identified.

10.5.7 Bone and periosteum

The Haversian vascular system of bone is a fragile one. Fragments of bone with no attachment to periosteum or muscle are already sequestrated and should be discarded, but any bone still attached should be retained. Exposed medullary bone should be curetted back to firm marrow. Any bone left *in situ* must be cleaned of dead muscle and foreign material, dirty bone ends are trimmed by a bone-nibbling forceps (*rongeur*). *Bone defect* is not important at this stage, the *wound* is of paramount importance, and every attempt should be made to avoid infection, whose development will only lead to greater bone loss. The danger of non-union from the removal of multiple bone fragments is overrated.

H. Nasreddine / ICRC

Figure 10.10.1
Patient Y: gunshot wound to arm and chest.

H. Nasreddine / ICRC

Figure 10.10.2
Patient Y: wound track opened up.

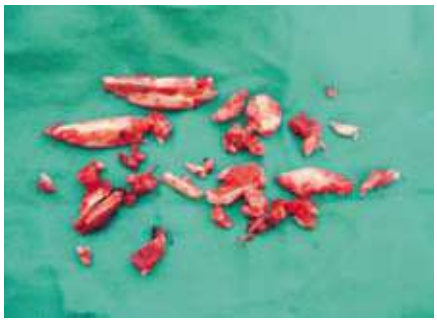

H. Nasreddine / ICRC

Figure 10.10.3
Patient Y: detached bone fragments removed.

Periosteum, on the other hand, is resilient and has a rich blood supply: it plays the predominant role in new bone formation. Its debridement should be conservative and restricted to obviously dirtied and contaminated edges.

ICRC practice is to use the simplest form of bone immobilization at the first debridement; this is usually a plaster of Paris (POP) posterior slab or some form of skeletal traction. Seldom is external fixation required at this stage and internal fixation is proscribed in ICRC practice. More definitive bone immobilization is decided during delayed primary closure of the wound.

For information on indications for different techniques of fracture immobilization and repair of bone defect, see Volume 2.

10.5.8 Arteries, nerves and tendons

As aforesaid, bleeding should be controlled if a major artery to the limb is damaged and either immediately repaired or replaced by a saphenous vein graft or temporary stent if a limb is to survive. The surgeon should pay particular attention to the possibility of a vascular injury near severely comminuted fractures with multiple fragments.

All nerves must be preserved as far as possible. Large nerves are resistant to section, although they may suffer neurapraxia. They are often the only structure left traversing a wound cavity. If they are found to be damaged, the site and degree of damage should be recorded. A non-absorbable suture may be placed in the proximal and distal ends and pulled together to facilitate their identification at a future operation. Exploration of the implicated nerve should be attempted during wound debridement *only* if it does not involve opening up healthy tissue planes.

Tendons should be trimmed and only grossly destroyed fibres removed. Severed tendons, if important and requiring later repair, should be marked with a non-absorbable suture as with nerves.

No attempt should be made at primary tendon or nerve repair, as success is unlikely in these grossly contaminated wounds. The failure of immediate repair will only make subsequent efforts that much more difficult. Also, repairs take time and energy at the "wrong time of the day"; they are best dealt with as planned elective procedures. Nerves and tendons should nonetheless be protected from extended exposure, by covering them with muscle or skin flaps or wet dressings.

For techniques on artery, nerve, and tendon repair, see Volume 2.

10.6 Retained bullets and fragments

Obviously, if the surgeon comes across a projectile during wound debridement then it should be removed, but healthy tissue should not be dissected in an attempt to find one. Otherwise, there are two conditions that require *immediate removal* of bullets and fragments, and these are related to specific proven risks and complications.

1. Projectile situated in a synovial joint – the piece of metal will cause pain, disability and progressive destruction of the joint cartilage though a mechanical action, and possible toxicity if made of lead: it should be removed as part of the wound debridement of the acute injury (Figures 10.11.1 and 10.11.2).

2. Risk that the projectile may cause erosion of an important structure (usually a major blood vessel) with the possibility of major haemorrhage or embolization (Figures 10.12.1 and 10.12.2). Should the surgeon suspect a pseudoaneurysm or an arterio-venous fistula then an operation to treat these pathologies involves the removal of the foreign body.

   Whether this removal is performed as an acute or planned elective procedure will depend on the exact anatomic location and endangered structure, the haemodynamic stability of the patient, the availability of diagnostic and operative equipment, and especially the *expertise of the surgeon*. The risks of a major procedure (removal of a bullet in the mediastinum, brain, etc.) whose morbidity may be quite high (especially in inexperienced hands) must be weighed against the benefits where the overall incidence of complications is low: for further indications and details, see Chapter 14.

ICRC

ICRC

Figures 10.11.1 and 10.11.2

Retained bullet in the hip joint.

F. Irmay / ICRC

F. Irmay / ICRC

Figures 10.12.1 and 10.12.2

Retained bullet in the upper mediastinum.

10.7 Final look and haemostasis

The edges of the wound should be retracted and blood clot, dirt and missile fragments removed from the sides and depths. Gentle and copious irrigation under low pressure, preferably with normal saline at this point, will wash out any residual debris and clot and dilute any bacterial load. A plastic bottle with holes cut into the top squeezed manually with both hands provides sufficient pressure; depending on the size of the wound cavity, one to three litres of saline is used. Very large and complicated fractures may require more until the wound "looks" clean.

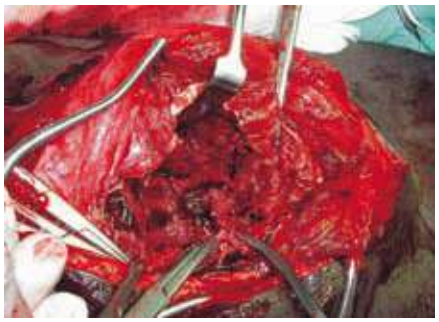

F. Jamet / ICRC

Figure 10.13

Patient X: final appearance of the wound cavity.

All the structures in the wound cavity should now be visualized and identified (Figure 10.13). The surgeon should carefully explore the wound with the finger to identify any foreign bodies or unexpected extensions of the wound.

- *Do not* open fresh planes in healthy tissue.
- *Do not* explore unnecessarily for metallic fragments.
- Remove fragments of clothing, in-driven dirt and vegetation.

F. Jamet / ICRC

Figure 10.14

Patient X: wound is left open.

After removal of the pneumatic tourniquet, bleeding should be controlled by pressure with gauze and by fine absorbable ligatures. Electric cautery (diathermy) is best avoided as it leaves dead burnt tissue behind, which is more harmful than the foreign body of an absorbable knot.

The wound should be left wide open (Figure 10.14). There is no use in "putting in a few sutures" to partially close the wound, "just to bring the edges together a bit". This will defeat the purpose of allowing for extensive wound decompression and drainage and, after the development of reactive oedema, a loose stitch becomes tight. In addition, although the wound may look clean, it is not sterile. Bacteria and microscopic debris still exist in the wound and will only be expelled with the post-traumatic inflammatory exudate if drainage is adequate.

Should a drain be put in place? If the wound is relatively shallow and left wide open, there is no need for a drain. If there are deep pockets in the wound, which cannot be fully opened up because of anatomical constraints, then a soft penrose or corrugated rubber drain may be required. Counter-drainage through a stab wound in a dependent part might be more useful.

These procedures are *not new*. They are the rules of all septic surgery to be found in all standard surgical reference texts. "Never close infected wounds. Do not close contaminated wounds or clean wounds that are more than 6 hours old. Systematically perform wound toilet and debridement and saline irrigation until the wound is completely clean. Perform delayed primary closure as a second procedure."

Old rules for new surgeons

In war wounds, the rules of septic surgery apply.

10.8 Wound excision: the exceptions

10.8.1 Management of minor Grade 1 wounds

Many Grade 1 soft-tissue wounds, according to the Red Cross Wound Score, can be treated conservatively. Examples include:

- perforating bullet wound with small entry and exit (narrow channel wound) without swelling of intervening tissues (haematoma/oedema) or other signs of injury to important structures (Figure 3.29.1);
- multiple superficial wounds due to "peppering" with tiny fragments of obviously low velocity and low kinetic energy (e.g. hand grenade) as shown in Figure 10.15.1.

Some superficial Grade 1 soft-tissue wounds will require only simple local wound toilet. This will suffice with many low-energy small fragment wounds where the potential culture medium is so small that the body's normal defence mechanisms can deal with it. They can be cleansed with soap, water, and a disinfectant and a simple dry dressing applied; the small wound is left open to heal by secondary intention. This is especially the case if antibiotics can be administered early (see Chapter 13). This ICRC experience has been confirmed by others.[6]

Excision of entry and exit wounds under local anaesthesia may be necessary to promote free drainage in other superficial Grade 1 wounds. This may be accompanied by syringe injection of normal saline for irrigation of the wound track with/without a drain. Some such wounds, however, will require a full surgical exploration and excision, particularly in the one well-known instance where the type of weapon is of utmost importance: anti-personnel blast mines. Even small puncture wounds from landmines can be loaded with mud and grass or the mine casing, all of which must be removed (Figure 10.16).

Heterogeneity of war wounds: there is no single treatment that applies to all wounds. Wounds must be considered by Grade and Type.

R. Coupland / ICRC

Figure 10.15.1

Superficial "peppering" by grenade fragments.

R. Coupland / ICRC

Figure 10.15.2

There is no penetration of the joint and no vascular injury – the wounds require simple wound toilet.

ICRC

Figure 10.16

All APM fragment wounds – large and small – require debridement.

10.8.2 Serial debridement

For some large wounds the line of demarcation between dead tissue and damaged, but viable, tissue is not clear. The life-history of a wound is such that apparently clean and living tissue may become necrotic after a few days, especially if there has been a delay between injury and debridement; and if the surgeon is not experienced in this type of surgery. The idea here is to debride obviously dead tissue; otherwise excise conservatively, and then re-examine the wound in theatre after 48 hours.

This method is called a "serial debridement", i.e. a wound excision over several sessions, and it should be *planned* as such. The surgeon should make a conscious decision that, "in this patient, because I am not certain of the viability of the remaining tissues and excessive excision of normal tissue would cause deformity or compromise function, I shall perform a second debridement later" (Figure 10.17).

In armies with very efficient means of evacuation and sufficient human resources, serial debridement may be the preferred method of treatment and takes place at different hospitals with different operating surgeons along the chain of casualty care.

Serial debridement cannot be considered a standard method of treatment in a situation of mass casualties and poor or no evacuation capacity. It requires sufficient personnel and resources to handle the number of wounded and to allow for a second look and follow-up. More often, the surgical workload is so great, or the tactical situation such, that serial debridement is not possible. The surgeon must treat each wound excision as a definitive one. In this instance, if in doubt about tissue viability, it is wiser to excise.

6 Bowyer GW, Cooper GJ, Rice P. Small fragment wounds: biophysics and pathophysiology. *J Trauma* 1996; **40 (Suppl.)**: S159 – S164.

ICRC

Figure 10.17

Serial debridement of a large wound: the line of demarcation of necrotic tissue has now become apparent.

The practice of serial debridement should not be confused with an incomplete or failed wound excision. In the latter case, the patient returns to theatre for delayed primary closure after five days and the wound is found to be infected with remaining necrotic tissue. It is not ready for suture and requires re-debridement.

10.9 Leaving the wound open: the exceptions

As is usually the case in surgery, there are exceptions where wounds may, or even should, be closed.

10.9.1 Head, neck, scalp and genitals

The excellent blood supply and minimal soft tissue of these structures usually allows for immediate primary closure after wound excision. Only in the presence of severe contamination, or if in doubt, might it be wiser to leave these wounds open.

M. Baldan / ICRC

Figure 10.18

Panga wound to the head.

In maxillo-facial wounds, the oral mucosa is an exception in all respects and every attempt should be made to close it primarily.

Machete or *panga* wounds, especially to the face or scalp, are *not* incised wounds, but rather a mixture of crush and laceration. Dirty connecting tissue is found to underlie an overhanging edge of skin and galea aponeurotica. If less than 6 hours since wounding, immediate primary closure, after full debridement, with a subcutaneous drain is permissible. If more than 6 hours old, it is better to leave the wound open for delayed primary closure after 2 – 4 days.

In penetrating brain injury, it is best to close the dura. This can rarely be sutured directly, but the problem can easily be overcome with a patch of pericranium or aponeurosis. After debridement of the scalp wound, the skin should be closed either directly or with a rotation flap.

10.9.2 Soft tissues of the chest (sucking chest wound)

These wounds must be debrided, but healthy muscle and pleura should be closed to preserve a functional serous cavity. The skin and subcutaneous tissue should be left open and a chest tube inserted.

10.9.3 Soft tissues of the abdominal wall

As with the chest, the wound should be excised and every effort made to secure peritoneal closure. Furthermore, if the development of abdominal compartment syndrome is suspected, temporary abdominal closure is preferable (Bogotá bag, etc).

For further details on abdominal compartment syndrome see Volume 2.

10.9.4 Hand

Excision should be very conservative and all viable tissue preserved to simplify reconstruction and improve the functional result. These wounds should be left open for DPC after 2 – 4 days; however, tendons and nerves should be covered by healthy tissue, through rotation flaps if necessary. Small wounds may be closed primarily.

10.9.5 Joints

Synovial membranes should be closed; if this is not possible the capsule alone should be sutured. Little harm seems to be done if the synovium cannot be closed securely. The skin and muscle should be left open.

10.9.6 Blood vessels

Blood vessels that have been repaired primarily or by vein graft should be covered by viable muscle if possible. The skin should be left open.

For specifics of each anatomic region, see Volume 2.

10.10 Dressings

Once the wound has been adequately excised, it should be covered with a bulky absorbent dressing made of dry fluffed-up gauze re-enforced with a layer of absorbent cotton wool. This is held in place with a *loose* crepe bandage or non circumferential adhesive tape. A tight bandage wrapped around the limb and soaked with exudate that dries will have a tourniquet effect. The gauze compresses should *not* be packed tightly in the wound; this will only impede drainage. The aim is to draw inflammatory exudate out of the wound and into the dressing. Exposed tendons and joint capsules may be covered with saline-soaked compresses.

ICRC

Figures 10.19.1 – 10.19.4

Big bulky dressing of fluffed gauze covered by an elastic bandage.

ICRC

ICRC

ICRC

Vaseline gauze should not be used, and the wound should not be "packed" in any way with gauze that will form a plug and prevent the free outflow of fluid.

The dressing should not be removed until taken down in the operating theatre under anaesthesia at the time of delayed primary closure. Ward dressings are an invitation to nosocomial infection. The surgeon should resist the temptation to change the wound dressing to "have a look at how it is doing". Each dressing change constitutes trauma to the healing granulation tissue and exposes it to cross-infection. A good look at the patient instead will suffice: the wound is doing quite well if the patient is smiling, eating, and sitting comfortably in bed.

If the dressing and bandage have become soaked with exudate, either over-dress with absorbent cotton, or take down the bandage and wet cotton and replace without disturbing the gauze compress that is in direct contact with the wound. The state of the dressing is not an indication of the state of the wound.

Wounds awaiting DPC should not have dressings changed until formal closure.

10.10.1 The exceptions

- Continuing *haemorrhage* requires immediate re-exploration; as do vascular changes indicating ischaemia.
- Obvious *signs and symptoms of infection:* fever, toxicity, excessive pain and tenderness, warmth, redness or shiny surface in dark-skinned people, oedema and induration, or a moist wound dressing with an offensive smell. These indicate the need for further surgical excision, which should be attended to in the operating theatre, not by changing the dressing in the ward.

All wounds awaiting DPC develop a "sour" odour after several days: the "good-bad smell" of ammoniacal products from the breakdown of serum proteins. The infected wound has a characteristic offensive odour: the "bad-bad smell".

10.11 Anti-tetanus, antibiotics, and analgesia

All patients should receive prophylaxis against tetanus. Penicillin, 5 mega-units 6-hourly i.v. should also be given from the time of admission. This should be followed by oral penicillin, 500 mg 6-hourly for a total of 5 days (see Chapter 13).

For proper rest of the injured part, and to make the patient ready for physiotherapy, good analgesia should be given (see Annex 17. A: ICRC pain management protocols).

10.12 Post-operative care

It goes without saying that proper post-operative nursing is crucial. ICRC experience has shown that the most important factor limiting the sophistication of surgical procedures performed in an ICRC hospital is not the technical expertise of the surgeon, but rather the level of post-operative nursing care. This should not be underestimated in the context of a poor, war-ravaged country.

In all cases where there is an *extensive* soft-tissue wound, even in the absence of a fracture, the entire limb should be immobilized to provide rest. This can be achieved by a plaster of Paris (POP) back slab.

The normal catabolic response to trauma must be overcome and patients offered a good nutritious diet. In a poor country, many wounded patients will arrive at hospital suffering from malnutrition. This will only compromise the body's ability to heal and resist infection.

The functional result of ultimate wound healing depends to a great extent on proper physiotherapy to retain muscle mass and joint mobility, and should be instituted early as part of the healing process.

The principles of good wound management

1. Adequate wound excision: removal of necrotic tissue, contaminated debris, organic foreign material, and blood clots.
2. Adequate wound drainage: fascial decompression, wound left open without any suturing, large bulky absorbent dressing.
3. Haemostasis.
4. Limb immobilization until healing of soft tissues.
5. Tetanus prophylaxis, antibiotics, and analgesia.
6. Nutrition.
7. Nursing and physiotherapy: mobilization of the patient.
8. No unnecessary dressing changes.
9. Delayed primary closure (4 – 5 days).

Chapter 11

DELAYED PRIMARY CLOSURE AND SKIN GRAFTING

11. DELAYED PRIMARY CLOSURE AND SKIN GRAFTING

11.1	Delayed primary closure	231
11.1.1	Methods of delayed primary closure	232
11.1.2	Dead space	232
11.1.3	Wound care	232
11.2	Skin grafting	233
11.2.1	Types of autologous skin grafts	233
11.2.2	Requisites for skin grafting	233
11.2.3	Partial thickness grafts	233
11.2.4	Reverdin pinch grafts	235
11.2.5	Application of grafts and graft meshing	235
11.2.6	Graft take	236
11.2.7	Graft care	237
11.2.8	Dressing of donor sites	237
11.3	Full-thickness grafts	238
11.3.1	Donor sites	238
11.3.2	Technique	239
11.4	Healing by secondary intention	240

11.1 Delayed primary closure

Delayed primary closure (DPC) is wound closure performed *four to seven days after injury*, which corresponds to the fibroblastic phase of wound healing. In the practice of ICRC surgical teams, the standard has been 4 – 5 days. The timing of wound closure is important; this is still defined as healing by *primary intention*.

Attempts to close wounds before they are clean must be avoided; but DPC is seldom possible later than 8 days after wound excision because of fibrosis. At this stage, healing by *secondary intention* will set in.

> No wound should be closed if there is persistent contamination or infection.

For DPC, the patient is taken to theatre and the wound opened and inspected under anaesthesia. The dressing over a clean wound is dry and greenish-black in colour, and has the ammoniacal odour ("good-bad smell") of degraded serum proteins. The muscle is adherent to the gauze and, as the dressing is gently peeled away, the muscle contracts and bleeds slightly. The surface is a bright red and oozes blood. This wound is ready for DPC. Figures 11.1.1 – 11.1.5 show an example of the full sequence of wound managment with closure by skin graft.

R. Coupland / ICRC

Figure 11.1.1
Very dirty wound.

R. Coupland / ICRC

Figure 11.1.2
After debridement.

R. Coupland / ICRC

Figure 11.1.3
Five days later, removal of dressing – note the dried haemoserous discharge. As the dressing comes off, the muscle contracts and bleeds.

R. Coupland / ICRC

Figure 11.1.4
Clean wound ready for DPC; in this case by split-skin grafting owing to the large area of skin loss.

R. Coupland / ICRC

Figure 11.1.5
Wound covered by early split-skin graft.

If the wound is infected, the dressing slides off with no resistance whatsoever because there is a film of pus between it and the wound surface, which may contain areas of necrotic tissue as well as pus. The surface is a dull- or greyish-red and does not bleed. The dressing gives off the "bad-bad smell" of wound sepsis.

Such an infected wound, or one containing residual contamination or devitalized tissue, requires further surgical excision – redebridement – and is then left open again. Closure is postponed. This failed attempt at DPC is not to be confused with "serial debridement" (see Chapter 10).

H. Nasreddine / ICRC

Figure 11.2
Direct suture as DPC.

11.1.1 Methods of delayed primary closure

This is usually accomplished by direct suture: simple approximation of the deep structures and skin with minimal mobilization of the skin edges, and *without tension* (Figure 11.2). If there is tension in the suture line, the skin edges will become necrotic and the wound will break down. Small wounds may be closed using adhesive tape to approximate the edges.

The wound should not be closed under tension.

Significant tissue loss prevents approximation of the deep structures and skin. Rotation skin flaps might be useful in some anatomic locations. If bone is exposed, a musculocutaneous flap may be necessary. Larger areas require skin grafting (Figure 11.1.5).

Adequate primary surgical excision is vital for uncomplicated wound closure by DPC.

11.1.2 Dead space

As with all wounds, the obliteration of dead space is an accepted principle of DPC management. This may be technically difficult when there has been extensive loss of deep tissue. Absorbable sutures are used to approximate deep structures, but, here again, undue tension should be avoided as it will only result in local tissue ischaemia that decreases local resistance to the development of infection and impedes healing. It is not necessary to suture fascia or subcutaneous tissues.

Drains should be *avoided* whenever possible in DPC. Drains may act as retrograde conduits for skin bacterial flora and contaminants to gain entrance into the wound, and they impair the resistance of tissue to infection; a drain constitutes an additional foreign body in the wound.

Drains should never be placed in wounds without dead space; they should only be used for those with significant dead space and where there is associated oozing of blood. Within 24 hours, the oozing has usually stopped and the drain should be removed. If used, a drain should be placed dependently and, if possible, be of the suction catheter variety. Otherwise, a simple soft penrose drain will suffice.

Avoid drains whenever possible in DPC. If used, remove within 24 hours.

11.1.3 Wound care

After DPC, the wound should be covered with some layers of dry gauze which can be left until the time of suture removal. If signs of infection develop after closure, the patient is taken back to theatre where the wound is inspected, sutures removed, and the wound re-opened for drainage and redebridement if necessary. Avoid performing this procedure in the ward, where proper inspection and drainage are difficult, as it only favours the development of cross-infection.

11.2 Skin grafting

If the wound cannot be closed by direct suture or rotational flaps because of skin loss, a skin graft can be applied, sometimes combined with partial direct suture.

11.2.1 Types of autologous skin grafts

Free skin grafts may be partial-thickness (epidermis and part of the dermis) or full-thickness (including all the dermis). Partial-thickness grafts, also called split-skin grafts (SSG), vary in thickness depending upon how much of the dermis is incorporated. The thicker the graft, the less it will contract and the more it will resemble normal skin in colour and texture; but the risk of the graft failing is greater. Conversely, thinner grafts are hardy and take more readily, but contract and distort more than thick ones and have a poorer functional and aesthetic result.

Thin partial-thickness grafts
These are used for large areas and when the development of a contracture or quality of skin cover is not important.

Thick partial-thickness grafts
These are used when skin quality is important, such as over flexion creases where contractures should be avoided. However, the recipient area must be very healthy with good vascularity.

Full-thickness grafts
These are best for a good cosmetic result in the face, or to cover a sensitive functional area on the hands or fingers.

11.2.2 Requisites for skin grafting

Skin grafts can cover any wound with enough vascularity to produce granulation tissue. This does not imply that fully mature granulation tissue must be present prior to skin grafting, and early grafting for DPC, when the wound is first opened on the fifth day post-debridement, is often used (Figure 11.1.5). Early SSG has advantages and disadvantages: the wound is closed preventing any infection from developing, which is particularly important in large wounds; but the quality of resultant skin is less and there is greater tissue fibrosis and contraction. Most full-thickness grafts are also performed early.

On the other hand, a deep wound, or one over a flexion crease, may be dressed while waiting for more granulation-tissue to grow in, and then grafted at a later date. However, excessive granulation tissue should be scraped clear prior to graft application. A limb with a wound over a flexion crease should be splinted in the extension position with a plaster back slab to avoid contracture while waiting for closure.

Tissues which cannot take skin grafts include avascular areas, such as hyaline cartilage, exposed tendon without paratenon, and exposed bone cortex without periosteum. These cases will require some form of skin or musculocutaneous flap for closure. In the case of exposed bone cortex, an alternative may be the drilling of multiple small holes in the cortex to let granulation tissue grow through from the inner cancellous bone.

11.2.3 Partial thickness grafts

Split-skin grafts (thin and thick) are taken from an area which can produce a broad area of skin. The common donor sites are the thighs, back, and arms and forearms (Figure 11.3).

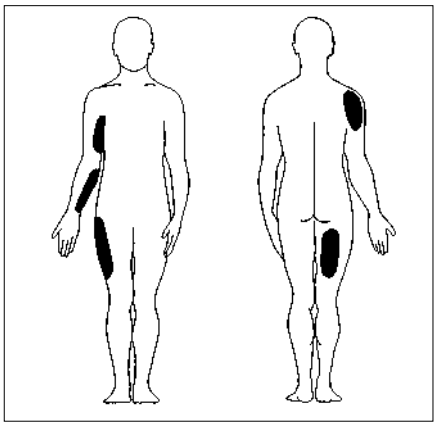

P. Zylstra / ICRC

Figure 11.3
Donor sites for partial-thickness grafts.

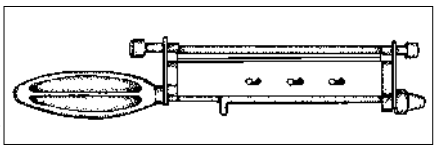

P. Zylstra / ICRC

Figure 11.4

A dermatome: various models, with disposable blades, are available – the thickness of the split skin graft harvested is controlled by the screw at the right end of the instrument; the setting is then locked using the screw at the left end.

Split-skin grafts should be taken using a dermatome, such as a Humby knife (Figure 11.4). A free hand knife, such as a De Silva knife which incorporates a razor blade, or scalpel, can be used if a dermatome is not available or if the area to be harvested is small.

The principles for the use of all hand knives are the same. The blades are disposable and are fitted into the knife. A calibration control is adjusted to set the depth of split skin required. Once this has been adjusted, a locking nut is tightened at the opposite end of the knife. It is important to check the gap between the blade and the knife before use, by holding the instrument to the light and visually checking the gap. The gap should be even throughout its length. Electric and pneumatic knives are also available but are not standard in ICRC practice.

The gap between the knife and the blade must be visually checked before use.

Harvesting the SSG

The area of skin to be removed is washed with soap and water and shaved if necessary; povidone iodine is applied. Intradermal saline or, better still, dilute adrenaline solution (1:500,000), is injected into the donor area prior to cutting, to facilitate harvesting and diminish local blood loss. The donor site, the cutting edge of the knife, and the edge of a skin board are greased with vaseline-gauze.

An assistant applies an ungreased skin board to the distal end of the donor site and pulls in such a fashion as to tense the skin tight. The assistant's other hand grasps the underside of the donor site (thigh or arm) to flatten the area from which the graft is being taken (Figure 11.5.1).

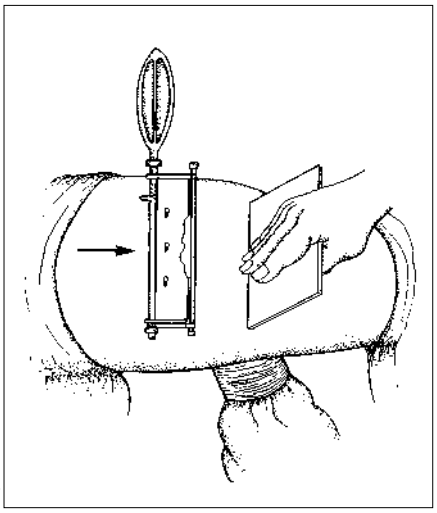

P. Zylstra / ICRC

Figure 11.5.1

Harvesting a split skin graft from the medial side of the thigh – note the assistant's left hand flattening the donor site by exerting upward pressure on the undersurface of the thigh.

E. Winiger / ICRC

Figure 11.5.2

Surgeon's wooden board applying counter-traction.

The surgeon places the greased board about 4 – 5 cm in front of the ungreased board, and exerts counter traction to increase the skin tension (Figure 11.5.2). The knife goes into the gap between the two boards at an angle of 30° and even, gentle, side-to-side cutting movements of about 2 cm are made. Continuous movement of the knife is essential. The surgeon advances the greased board proximally while exerting *minimal effort* to advance the knife. The tendency is to cut too quickly, or exert too much pressure at too high an angle; the blade will not shave the skin off but cut deeply into the subcutaneous tissue. When the desired amount of skin has been raised, the wrist is supinated and the knife cuts through. The harvested skin is now placed on a saline-soaked gauze swab, undersurface upwards, and stored until applied to the recipient area.

11.2.4 Reverdin pinch grafts

These can be used to cover large areas of defect and where the recipient area is unfavourable, such as muscles moving in different planes. A full take is not expected and the gaps between the grafts will fill in as the skin spreads outwards and coalesces. The cosmetic result is poor.

As with SSG, an intradermal infiltration of local anaesthetic with adrenaline is injected into the donor area. This helps to avoid cutting the grafts too deeply and is often all the anaesthesia required. The grafts are cut using a scalpel blade and can vary in size up to about 2 cm^2 (Figure 11.6). The harvested skin is handled as described previously until it is applied to the recipient area.

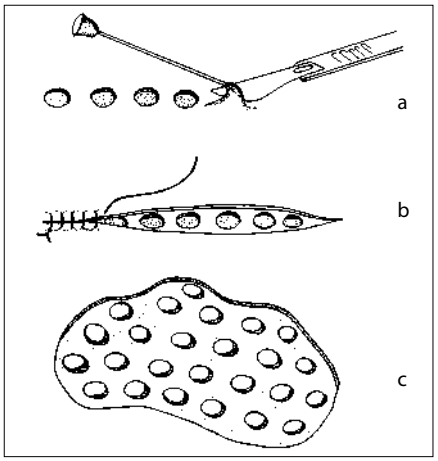

P. Zylstra / ICRC

Figure 11.6

Reverdin pinch grafts:

a. The skin is picked up with a needle and round patches of thick SSG, 1 – 2 cm in diameter, are cut with a scalpel.

b. If the patches are taken in a line, the area can be excised and the wound closed by primary suture.

c. The patches are spread over the receptor site with gaps of a few millimetres between each piece.

11.2.5 Application of grafts and graft meshing

The recipient bed is prepared by washing with saline and scraping clean any exuberant granulation tissue.

V. Sasin / ICRC

Figure 11.7

Recipient site prepared to receive SSG.

The harvested partial-thickness skin is laid across the recipient bed and cut so that it is slightly larger than the site. This allows for the cut undersurface to be in complete contact with the recipient bed and for later graft contraction. Often a graft is placed on a wound that is deeper than the thickness of the graft. In these cases the edge of the graft at the base of the wound must be carefully approximated to include the entire height of the defect.

The graft should be meshed to allow for drainage of serum and blood; this is important to avoid the graft "floating" and provide for close contact between it and the recipient bed. Meshing is accomplished by placing the harvested skin on a surface, preferably the wooden board, undersurface facing upwards, and making numerous parallel incisions through the graft with a No. 15 scalpel blade. Optimally, the ratio of gaps to skin should be 3:1 thus allowing the SSG to cover an area three times its original size (Figure 11.8). Over a period of 10 – 14 days, the skin grows across the interstices and complete healing is obtained.

Graft meshing is particularly useful in cases where there is insufficient autograft to completely cover raw areas, such as large wounds or burns (see Figure 11.1 for a clinical example). Mechanical meshing machines are available – and expensive – but are not standard ICRC equipment except in hospitals with a large number of patients suffering from burns. A good improvisation is a sharp circular pizza cutter, modified with notches ground into its circumference, and used to mesh skin against the wooden board.

H. Nasreddine / ICRC

Figure 11.8

Meshing of a skin graft.

Grafts can often be kept in place by a suitable dressing and elastic bandaging. If graft stability cannot be ensured by the covering dressing, it should be sutured in place (Figure 11.9). One technique is a continuous 3/0 non-absorbable suture run around the edge of the graft to provide adequate contact. Another is to use several anchor sutures about the circumference, left long and used as tie-overs across the dressing; especially useful for small grafts on hollow areas (Figure 11.11). The excess skin of the graft can either be trimmed once the graft has been sutured in place, or cut 10 – 14 days later.

V. Sasin / ICRC

Figure 11.9
SSG sutured in place.

Once the graft has been sutured or placed on the recipient bed, it should be covered with a sheet of vaseline-gauze followed by a gauze compress or cotton-wool pledgets soaked with saline. These dressings should be pressed into the concavities of the defect to ensure that there is complete apposition between the graft and the bed. The dressing should then be covered with a thick layer of cotton wool held in place by a crepe bandage.

In case of consistent oozing from the recipient bed, consider delaying the application of the graft for 24 – 48 hours. The harvested skin can be stored for up to three weeks in an ordinary domestic refrigerator at 4° C without compromising the success rate. It is placed on a moist piece of gauze with the undersurface facing upwards and then put in a sterile jar containing normal saline. Jars should be stored with an airtight lid and *carefully labelled* with the patient's identification and the date of harvesting.

Delayed application of split skin grafts is often the preferred method.

11.2.6 Graft take

Grafts adhere initially by a thin layer of fibrin and are fed by "plasmatic imbibition" from the recipient bed granulation tissue until in-growth of capillaries occurs by about the fourth post-operative day. The new capillaries link with the graft and the fibrin clot is transformed into fibrous tissue. For these reasons, there are three major factors determining the success rate of split-skin grafts.

1. A vascular recipient bed which is *free of pathogenic bacteria*. Normal bacterial flora will not necessarily interfere with graft survival. The most common pathogen causing graft failure is beta-haemolytic *Streptococcus pyogenes*, probably through its fibrinolytic enzymes. ICRC protocol suggests prophylaxis with penicillin for five days, to which this bacterium remains sensitive.

2. The *vascularity* of the graft itself. Thin split-skin grafts have a higher concentration of capillaries on their undersurface than thick grafts. For this reason, a thin split-graft has a higher chance of survival than a thick one.
3. The maintenance of *close contact* between the graft and the recipient bed. If the graft is under tension, if blood or serum collects between the surfaces, or if there is movement of the graft on its bed, satisfactory contact cannot be maintained. Thus, meshing of the graft and tight contact between the graft and recipient site are important; grafts spanning joints should be splinted to prevent graft loss from shearing forces.

Grafts should be harvested under sterile conditions and placed on a clean, well-vascularized recipient bed. It is imperative that there be tight contact between the graft and its bed.

11.2.7 Graft care

Grafts must be cared for by experienced personnel. A graft may be *saved* if a collection of serum or haematoma is evacuated early. Great care must be taken not to *lose* a good graft through clumsy or careless removal of the dressing.

Grafts require careful nursing care.

In ICRC practice, the graft is first inspected after 48 – 72 hours; the dressing is carefully removed using two pairs of tissue forceps: one holding the graft and keeping it applied to its bed, the other removing the dressing. Care must be taken not to tear away the graft. If the graft is healthy and adherent to the recipient base, another vaseline-gauze and compress are bandaged in place and no further dressing is necessary for another 10 days, until the time of stitch removal.

If a haematoma or seroma is found, the fluid should be expressed through the meshed gaps using gentle pressure applied through the forceps and a pledget of soaked cotton-wool. Another firm dressing of vaseline-gauze is applied. These grafts are then inspected on a daily basis until firmly adherent to the recipient base.

Any small pus collections should be debrided with scissors to prevent spread of infection. Intra-epidermolysis is a phenomenon where the outer layer of the graft is lost while viable epithelial cells remain at the base. Dark blisters can be seen and may be debrided but great care should be taken not to disturb underlying graft that may still be viable.

On the other hand, any graft which is dead or floating on a "sea of pus" should be removed and the wound cleaned with normal saline. (If Pseudomonas is the culprit – presence of bluish-green pus – a dilute vinegar solution is useful.) If the raw area of the wound is more than 2 x 2 square cm, it should be re-grafted when clean. If smaller, it can be left to heal by secondary intention.

At ten days, a healthy graft should be firmly adherent and can be left exposed and submitted to a simple daily inspection.

11.2.8 Dressing of donor sites

Split-thickness donor sites can bleed briskly, and in children can result in significant blood loss. As mentioned above, the donor site should be infiltrated intradermally with a 1:500,000 solution of adrenaline or a local anaesthetic with adrenaline to decrease bleeding. They can also be extremely painful, especially when dressings become firmly adherent.

The following regimen minimizes discomfort and complications.

1. After harvesting the SSG, a dry dressing should be applied immediately, kept in place if necessary with an elastic bandage, and left alone while the graft is being applied; by which time haemostasis has usually been achieved.
2. The dressing should then be removed. If there is persistent bleeding, a compress wetted with the dilute adrenalin solution should be applied and direct pressure exerted for a few minutes.
3. Next, the donor site should be covered with a firm pressure dressing of vaseline-gauze, dry compresses and elastic bandage; it should be kept from falling down the limb by adhesive strapping. The dressings should be left *in situ* for 10 –14 days, unless there are signs of underlying infection.
4. A membrane dressing, such as Opsite ®, can be used if available. If fluid accumulates beneath the membrane, it can be evacuated by pricking with a needle and pressing. The perforation site should be covered with a smaller piece of membrane. If the dressing falls off it can be reapplied. Generally, this stays in place for the desired period of about ten days.

11.3 Full-thickness grafts

Full-thickness skin grafts consist of the epidermis and entire dermis. These grafts are used principally for the *head and neck* and *hands and feet* to provide thicker, better-quality cover.

Full-thickness grafts have a number of advantages apart from better texture and colour. They allow for the transfer of hair-bearing skin, as all the dermal adnexal structures are intact, and they contract less than split-skin grafts. The main disadvantages are a lower survival rate and their limited size.

The recipient bed must be in optimal condition to take a full-thickness graft; haemostasis, in particular, must be perfect. There should be meticulous suturing and approximation between the graft and the recipient edge. Generally, a full-thickness graft is small because the plasma exudate of the recipient bed is barely sufficient to feed and oxygenate the graft, and capillaries from the recipient bed and edge and graft undersurface must grow in quickly to assure survival.

11.3.1 Donor sites

The best and most readily available donor sites are the supraclavicular fossa, in front of and behind the ear, and the flexor surface of the forearm or the groin creases. In older persons, the cheek or the neck in the line of a skin crease may also be used. The donor site is sutured primarily following harvesting of the graft.

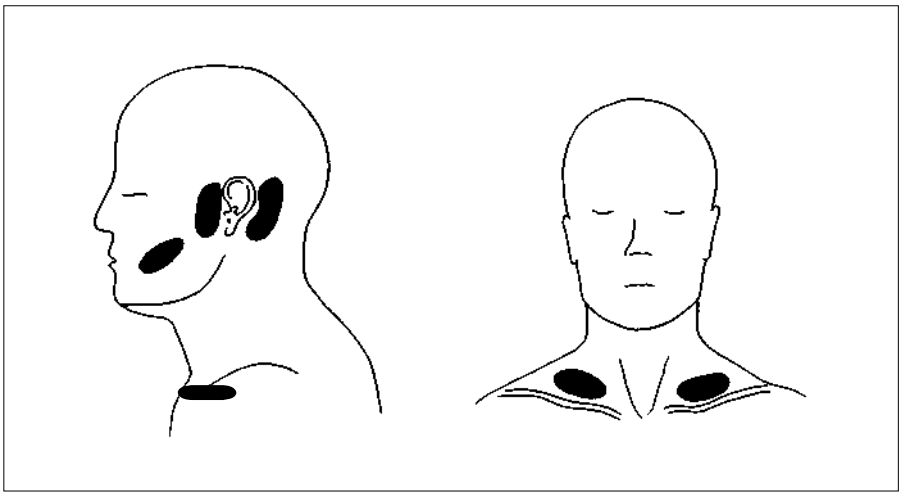

P. Zylstra / ICRC

Figure 11.10
Donor sites for full-thickness skin grafts.

11.3.2 Technique

The graft can be taken under local anaesthesia. The donor and recipient areas are cleansed in a sterile fashion and a template of cut gauze compress made to fit the defect exactly. The template is placed on the donor site and an outline drawn around its circumference. The donor area should be injected with local anaesthetic *with* adrenaline and the edge of the recipient area with local anaesthetic *without* adrenaline; both are left for five minutes before incision. The full-thickness skin should then be excised exactly.

The graft must be *thoroughly defatted* before application. This is best achieved by placing the moist graft, epidermis-side down, on the finger tip and scrupulously excising the fat using fine scissors. The graft is left unmeshed.

The graft is then sutured in place immediately and meticulously. Several interrupted nylon sutures are inserted close to the edge and left long for tie-overs (Figure 11.11). Small interrupted or continuous fine nylon sutures (5/0 if available) are placed around the circumference.

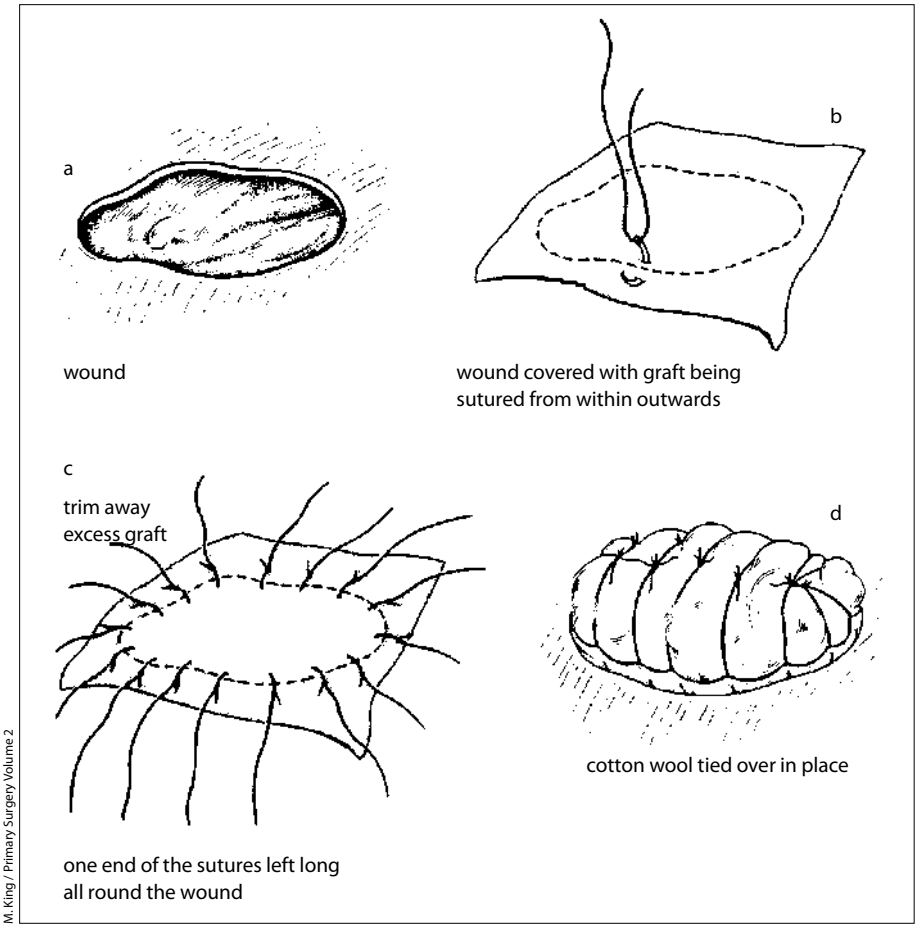

M. King / Primary Surgery Volume 2

Figure 11.11
Tie-over method of securing skin grafts.

Face grafts may be left open and covered with antibiotic ointment. They may be gently rolled with gauze or poked with a needle in the first 48 hours to express small serum collections.

On the hands or over a flexion crease, once the graft has been sutured in position a bolus of dressing is fashioned with vaseline-gauze and moist cotton-wool pledgets to hold the graft closely in contact with the recipient bed. The tie-over sutures are brought together over the dressing, which should be snug, but not so tight as to strangle the graft.

The dressing should be left in place for 7 – 10 days and then removed. At this point, in both light- or dark-skinned people, the graft may be any colour from pinkish or pinkish-white to bronze to black. Even a graft that has a blackish appearance on inspection can survive. No judgement on graft survival should be passed until a month from the time of application.

11.4 Healing by secondary intention

Some small wounds are difficult to suture without tension, or without very wide mobilization of skin flaps, because their surrounding edges are fibrotic. There is little benefit in another surgical operation, even skin-grafting. These wounds are best left to granulate and heal by secondary intention (Figure 12.10).

A change of dressing and gentle washing with normal saline every 4 – 5 days is usually sufficient: every dressing is a trauma to a healing wound. Daily dressings are not necessary if the wound is dry. Some traditional local dressings (sugar, honey, etc.) have an antibacterial effect and promote the formation of granulation tissue and are useful if the wound is deep. They have been used successfully by ICRC surgical teams.

Chapter 12

NEGLECTED OR MISMANAGED WOUNDS

12. NEGLECTED OR MISMANAGED WOUNDS

12.1 General considerations 245

12.2 Chronic sepsis: the role of biofilm 247

12.3 Surgical excision 248
12.3.1 Soft tissues 249
12.3.2 Bone 249
12.3.3 Irrigation 250

12.4 Antibiotics 250

12.5 To close or not to close? 251

12.1 General considerations

As described in Chapter 10, in many war zones today, characterized by irregular bush warfare, rebellions and insurrections, neglected and mismanaged wounds are amongst the commonest injuries that the surgeon sees. First aid is not available, doctors and nurses are scarce, and health services have been disrupted by poverty and conflict. Distances are long and the terrain difficult with little in the way of organized transport. Many patients reaching hospital have wounds which were inflicted more than 24 hours earlier, and some have wounds that are many days, or even weeks, old. Even if injured patients reach hospital fairly quickly, the sheer number of casualties often exceeds the surgical capacity or expertise available. The result is an inordinate delay in their treatment or frank mismanagement.

ICRC EXPERIENCE

In the ICRC hospital of Lokichokio in northern Kenya serving the victims of the conflict in southern Sudan, 12,264 war-wounded patients were registered between 1991 and 2006 in the ICRC surgical database. Evacuation was by ICRC and United Nations airplanes and few, if any, received pre-hospital treatment of any sort. Of these patients, 84 % reached the hospital more than 72 hours after injury. ICRC staff have seen similar conditions of delay in evacuation to a surgical facility in Somalia, the Democratic Republic of the Congo, Nepal and elsewhere, and have seen the same phenomenon with crush wounds after earthquakes in low-income countries.

H. Nasreddine / ICRC

Figure 12.1.1

Patient A: The gunshot wound to the left iliac region was sutured 5 days previously. The wound to the right thigh underwent insufficient wound debridement.

Before reaching a proper surgical hospital, some patients receive no treatment at all; others a cursory dressing, while yet others have undergone an inadequate wound excision. Some wounds will have been sutured, generally without any form of excision.

All sutured wounds, no matter how clean they appear, must have their sutures removed and opened up for drainage. As mentioned previously, the simple rules of septic surgery apply. Never close infected or contaminated wounds primarily.

Remove all sutures from previously sutured wounds.

After delay, some Grade 1 minor soft-tissue wounds are found to be healing spontaneously. Most wounds, however, will be inflamed or frankly infected with a degree of chronic sepsis, and some will be grossly putrefying. This is often seen in mismanaged wounds. Tetanus, gas gangrene, and invasive haemolytic streptococcal infection are ever present dangers (see Chapter 13). These wounds require aggressive excision.

H. Nasreddine / ICRC

Figure 12.1.2

The sutured wound is infected; a pearl of pus is evident at the lateral edge.

H. Nasreddine / ICRC

Figure 12.1.3

Sutures have been removed revealing abundant pus.

ICRC

Figure 12.2.1

Patient B: primary suture was performed – note the tension in the wound due to tissue oedema and infection. Some sutures have been removed.

ICRC

Figure 12.2.2

All stitches have been removed – the skin edges are ischaemic and necrotic and the subcutaneous tissues oedematous.

ICRC

Figure 12.2.3

After redebridement – the wound is now larger than the original injury.

H. Nasreddine / ICRC

Figure 12.3

Neglected gunshot wound to the knee with frank sepsis.

H. Nasreddine / ICRC

H. Nasreddine / ICRC

Figures 12.4.1 and 12.4.2

Neglected wound to the scrotum with gangrenous tissue.

Please note:

Many neglected wounds are infested with maggots. There is a long historical literature about "maggot debridement therapy", especially for chronic wounds, and a number of surgeons around the world practise it today. Suffice it to say that, for cultural and psychological reasons, most patients do not accept such a method. ICRC experience in this domain, although often positive, can only be described as anecdotal.

E. Dykes / ICRC

Figure 12.5

Maggot infestation of a wound. (The colouring is due to gentian violet.)

Please note:

There is much clinical and experimental evidence to back the claim that the early administration of antibiotics, penicillin in particular, can delay the advent of serious infection in war wounds. It is ICRC practice in first-aid posts to begin penicillin as soon as possible. However, adequate first-aid measures, as described in Chapter 7, are seldom available in many contemporary theatres of war and it is to this context that the present chapter is devoted.

12.2 Chronic sepsis: the role of biofilm

The chronic pyogenic sepsis of a neglected wound has its own particular pathology and bacteriology, and life history. Bacteria exist in discrete colonies only in a laboratory. The natural state of bacteria through natural selection is to become attached to surfaces, especially to inorganic or dead matter, such as sequestrated bone and cartilage. In chronic infection, bacteria secrete a glycopolysaccharide biofilm; this is the "slime" one feels on the rocks in a river. This biofilm protects the bacteria and prevents not only antibiotics, but macrophages, leukocytes and antibodies from attacking them. It must be actively broken up in order to eradicate the infection.

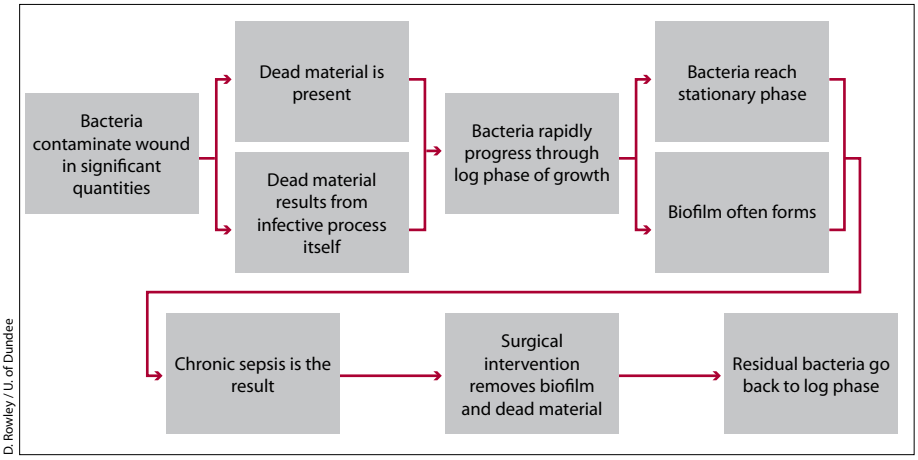

D. Rowley / U. of Dundee

Figure 12.6

Schema of the chronic infective process.[1]

The biofilm is secreted during the stationary phase of bacterial life. Physical disruption of this biofilm and surgical removal of the dead tissues are required to push the bacteria back into the rapid multiplication of the log phase when they are most susceptible to antibiotics and the body's natural defence mechanisms.

1 Rowley D.I., University of Dundee, Course on chronic bone infection, Seminar on the Management of War Wounds; ICRC: March 18, 2005; Geneva.

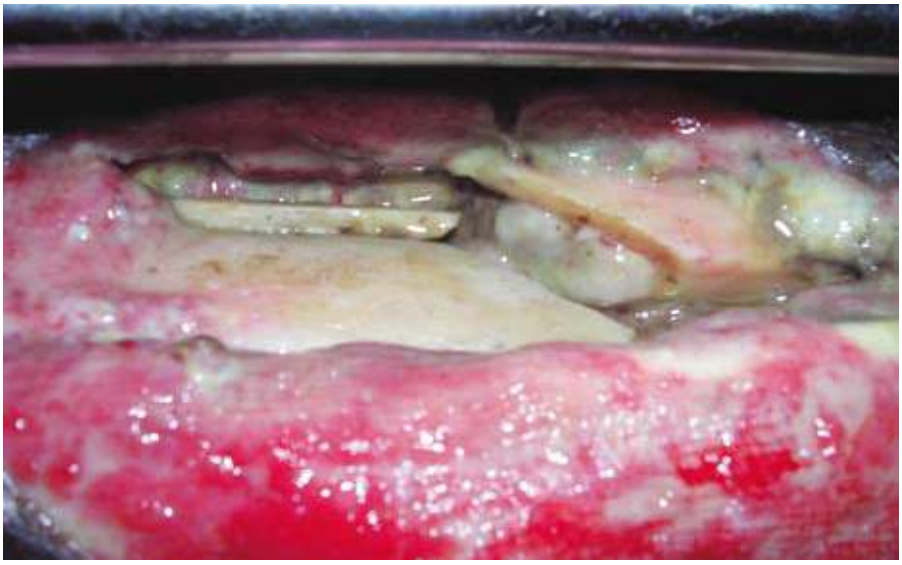

E. Dykes / ICRC

Figure 12.7.1

Neglected infected wound showing a layer of biofilm pus.

M. Dalla Torre / ICRC

Figure 12.7.2

Neglected wound of an amputation stump, without biofilm layer.

R. Coupland / ICRC

Figure 12.7.3

Neglected wound to the arm with necrotic tissue, but dry.

12.3 Surgical excision

Surgical excision is a more difficult procedure in these neglected and mismanaged wounds. The line of demarcation between viable and non-viable tissue, especially in oedematous muscle and fascia, is less obvious, and the zone of post-traumatic inflammatory hyperaemia is compounded by that of infective inflammation and the presence of the biofilm. The extent of adequate excision is more difficult to assess since there is not only a "mosaic" of ballistic tissue destruction in the permanent cavity itself but also that due to the septic process. Festering wounds that are days old have regions of infection mixed with regions of fibrous tissue healing.

The wound often resembles a cavity filled with detached soft tissue, bone fragments or the ends of fractured long bones, foreign debris, and fibrous tissue, covered with a coat of pus (Figures 12.7.1 and 12.8.2). Access may be difficult because of areas of wound contracture due to the tough fibrous tissue.

The principles of surgery remain the same, however. The skin and deep fascia must be widely re-incised and the wound cavity laid open for proper visualization and drainage (drainage following the same principle as for a simple abscess). The excision is directed towards the *removal* of all non-viable and heavily contaminated tissue and foreign debris, and the *physical disruption* of the biofilm.

Because surgical excision is more difficult in these wounds, there is a higher incidence of persistent infection. In such circumstances, several redebridements may become necessary. Depending on the experience of the surgeon, these are the patients who might most benefit from staged serial debridements.

12.3.1 Soft tissues

If present, all sutures must be removed and debridement performed as if no previous wound toilet had been done.

All macroscopic contamination should be excised. Skin and subcutaneous fat must be cut back until they ooze blood. Grossly infected fascia is usually shredded and dull grey, while healthy fascia is a glistening white. The muscle compartments are in even greater need of decompression than those of a fresh wound.

Contractility is the best indication of muscle viability. On the other hand, the surgeon must learn to distinguish between bleeding from small vessels in partially necrotic tissue, and capillary oozing from healthy but inflamed tissue. Experience brings an increase in the amount of tissue excised. Vessels and nerves should be left intact on the other hand, because they rarely turn septic.

Operative blood loss is usually considerable because of the inflammatory oedema and hyperaemia.

It is even more important *not* to open up healthy tissue planes in an attempt to remove any projectile. This will only help to spread the infection.

12.3.2 Bone

If there is an underlying fracture in these septic wounds, it usually contains *detached and necrotic* bone fragments that provide the perfect "surface" for bacterial adhesion. It is vital that these be found and removed. A plain X-ray will assist in their identification (Figure 12.9.2).

F. Jamet / ICRC

Figure 12.8.1

Gunshot wound 7 days after inadequate debridement.

F. Jamet / ICRC

Figure 12.8.2

Obvious sequestra presenting at the wound surface – note the presence of a pus-filled biofilm covering bone and soft tissue.

F. Jamet / ICRC

Figure 12.8.3

Wound 7 weeks after debridement.

Totally detached bone fragments are often trapped inside fibrous tissue and their localization and identification can be extremely difficult and frustrating.

- A dead bone fragment is pearly white; living bone is dull greyish-white with red capillary specks.
- If scratched with a bone curette, dead bone will give off a high and hollow resonance; living bone sounds low and dull.
- Another telltale sign can be observed when the bone fragment is grasped with forceps and a motion of pronation and supination imparted. If muscle or periosteum is holding the fragment in place, these structures will move with the pronation and supination. If the fragment is detached and held by fibrous tissue only, the movement will break the fibrous adhesion and the bone fragment will come free.

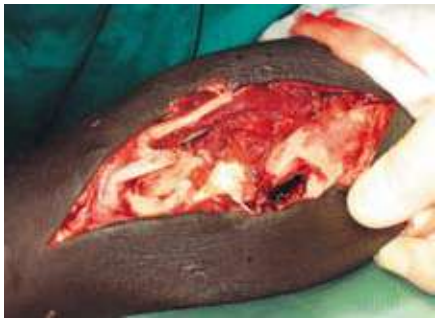

H. Nasreddine / ICRC

Figures 12.9.1 and 12.9.2

All loose bone fragments must be removed.

H. Nasreddine / ICRC

H. Nasreddine / ICRC

Figure 12.9.3

Bone fragments on the right, excised fibrous tissue on the left.

Once all loose bone fragments have been removed, the wound cavity is then curetted and a careful finger exploration performed to confirm a smooth surface: sharp bone edges may injure the exploring finger! All detached and dead bone fragments must be removed; they are already sequestra. Sometimes, several attempts may be necessary to eliminate them all, i.e. several redebridements.

All detached bone fragments are sequestra.

12.3.3 Irrigation

The wound is now copiously irrigated and dressed as for routine surgical excision. The ICRC has tested irrigation with a high-pressure pulsatile lavage and daily showering of the open cavity in these neglected septic wounds. Although both techniques appeared to clear up the biofilm and afford a better assessment of tissue viability, the results were inconclusive. Both involve a great deal of time, effort, and nursing care, and are difficult to maintain if a large number of patients are under treatment with no possibility of transfer to other less overstretched facilities.

12.4 Antibiotics

Penicillin and metronidazole are given, according to the antibiotic protocol (see Chapter 13); and gentamycin is added if there are signs of active spreading inflammation. No topical antibiotics or antiseptics are used in ICRC practice.

Bacterial culture should be done, if available. Good bacterial culture and sensitivity in a forward hospital is far more diffucult to accomplish than is usually realized. Note that clinical response does not always follow laboratory sensitivities, or lack thereof. Not only do "wild" bacteria not live in colonies, but *in vitro* sensitivity does not always reflect *in vivo* response, and the surface or discharge flora does not always represent the bacteria present *inside* the tissues.

Nonetheless, the basic rules of septic surgery still apply. Infection requires good drainage – as with an abscess – and the elimination of the bacterial culture medium, i.e. the dead tissue. Antibiotics will only be effective once the biofilm has been disrupted and the bacteria made susceptible to their action.

Persistent or recurrent wound infection requires re-excision, not ward dressings.

12.5 To close or not to close?

The primary phase of wound healing begins at the moment of wounding. It is well established by the time of presentation if the wound is a few days old. Thus, several days after surgical excision, many wounds are past the time of healing by primary intention. There is already much fibrous tissue present (Figure 12.9.3) and even more by the time the wound is clean and ready for closure. If secondary suture is attempted there is usually considerable tension on the wound edges with a high incidence of necrosis and breakdown.

The majority of these wounds are unsuitable for delayed suture and require skin grafting or rotation skin flaps for closure; or, if small, should be left to granulate and heal by secondary intention (Figure 12.10).

Most delayed or neglected wounds will be unsuitable for delayed primary closure.

It is in healing by secondary intention that several traditional local wound treatments may be of value, as mentioned in Chapter 11. ICRC surgeons and nursing staff have used dressings of honey or sugar, or hypertonic saline (add salt to normal saline until it no longer dissolves), which promote granulation tissue formation and have antibacterial properties. It must be emphasized that these topical treatments are an *adjunct to*, and *not a replacement for*, adequate surgery.

F. Jamet / ICRC

Figure 12.10
Healing by secondary intention.

Many patients with severe wound infection or frank putrefaction are malnourished, anaemic and dehydrated. As a result wound healing is poor and special measures should be taken to address these concerns.

For a discussion of post-traumatic osteomyelitis, see Volume 2.

Chapter 13

INFECTIONS IN WAR WOUNDS

13. INFECTIONS IN WAR WOUNDS

13.1 Contamination and infection 255

13.2 Major bacterial contaminants in war wounds 256

13.3 Major clinical infections of war wounds 257
13.3.1 Definitions 257
13.3.2 Gas gangrene 258
13.3.3 Tetanus 260
13.3.4 Invasive streptococcal infection 262
13.3.5 Necrotizing soft-tissue infections 263

13.4 Antibiotics 263
13.4.1 Antibiotic prophylaxis in the primary treatment of war wounds 264
13.4.2 Cephalosporins and others 265

13.5 Neglected or mismanaged wounds 265

ANNEX 13. A ICRC Antibiotic protocol 266

13.1 Contamination and infection

All war wounds are grossly contaminated by bacteria. Bullets and fragments are not sterilized on discharge and, at the entry site, bacteria are carried into the tract by the contaminated projectile itself. Furthermore, the negative pressure of the temporary cavity at both entrance and exit wounds also sucks in bacteria.

War wounds are dirty and contaminated from the moment of injury.

Other contaminants found in war wounds include clothing fragments, dust, miscellaneous organic matter (mud, grass, leaves, etc.), and wood and glass splinters from bombed buildings. Studies in Korea and Viet Nam showed that different varieties of soil are associated with different organisms – which also vary according to weather and season – with varying likelihood of producing severe infection (mud and agricultural soil contaminated with animal or human excrement are apparently the worst). In addition, the chemical composition of some soils can inactivate antibodies and impair leukocyte phagocytosis. Various home-made landmines have contained animal manure contaminant.

It is important, however, to distinguish between:
- simple contamination of a wound;
- local sepsis; and
- invasive infection.

In a contaminated wound, bacteria proliferate in dead tissues but no invasion of viable tissue takes place until the number of bacteria reaches a threshold of 10^6/gram of tissue. The presence of crushed tissues, dirt, and other foreign materials lowers the threshold of infection and invasion, by providing an environment where bacteria can grow and replicate without the patient's immune system being able to eliminate them. There are a number of *microbe-related risk factors* involved: virulence, production of enzymes and toxins, etc.

The body attempts to wall off the mass of necrotic muscle and bacteria with a fibrin barrier. Without adequate drainage, this will form an abscess; with proper drainage the necrotic mass will be expelled after 10 days.[1] In both cases, the body's natural defence mechanisms come into play to isolate the contaminants and the dead tissues.

When the local defence mechanisms can no longer cope, invasive deep tissue and systemic infection occurs. Patient immunity and resistance can also be weakened by malnutrition and chronic disease, especially HIV/AIDS. The local defence mechanisms and general resistance constitute the *host-related risk factors*.

Six hours appears to be the critical period after contamination. The efficacy of surgery and antibiotics to prevent infection decreases with time from injury, while the bacterial load increases exponentially. As with the differential grading of wounds according to the Red Cross Wound Score described in Chapter 4, a good understanding of the pathology involved is important in determining the proper and adequate procedure to follow, and the usefulness of antibiotics to *supplement* excision, drainage, and the body's natural resistance to infection. Uncomplicated soft-tissue Grade 1 wounds, as we have seen, can be treated with simple wound toilet; an antibiotic treatment helps if begun within 6 hours. Early administration of antibiotics in more severe wounds seems to inhibit bacterial growth and limit local invasion, at least temporarily. Any delay beyond six hours, however, greatly increases the risk of infection.

1 Fackler ML, Breteau JPL, Courbil LJ, Taxit R, Glas J, Fievet JP. Open wound drainage versus wound excision in treating the modern assault rifle wound. *Surgery* 1989; **105**: 576 – 584.

13.2 Major bacterial contaminants in war wounds

The relative preponderance of various bacterial species differs according to geography, topology of the terrain, and climate: agricultural land, urban environment, jungle, or desert; summer or winter. The presence of community-acquired resistance in "normal" body flora also differs from region to region, and with the passing of time.

Nonetheless, certain constants remain, and the following list gives a general indication of the common pathogens involved in infections of war wounds.

Gram-positive pyogenic cocci:
Staphylococcus aureus and β-haemolytic streptococcus. These are a normal part of the skin flora. In recent years, community-acquired methicillin-resistant *Staphylococcus aureus* (MRSA) has become more of a problem in certain countries.

Gram-negative bacilli:
Escherichia, Proteus, Klebsiella, Pseudomonas, and Bacterioides are found in the gastro-intestinal tract. *Acinetobacter baumannii* is to be found in the soil and on the skin and has recently become important in causing nosocomial infections.

Gram-positive bacilli:
The Clostridia species are spore-bearing obligatory anaerobes. They are widely distributed in the environment, especially in agriculturally fertile soils, and are carried in the faeces. They are the cause of gas gangrene and tetanus.

Wounds to the upper thigh and perineum are particularly at risk from contamination by faecal organisms (Clostridia and Gram-negative bacilli), even without bowel perforation. Bowel injury demonstrably increases the risk of sepsis.

War wounds have a "bacteriological life history": contaminating bacteria change with time.

Different infections occur early or later in the wound's history.

The important point to note about this varied bacterial flora is that wound contamination is not static. That war wounds have an evolving "bacteriological life history" from the time of injury has been known since World War I and the studies made by Alexander Fleming.[2]

Many studies since have demonstrated that wound contamination – and any resulting infection – evolves with time. At the instant of wounding, there is polymicrobial contamination with particular risk from Clostridia and β-haemolytic streptococcus. Afterwards, self-contamination from skin and gastro-intestinal flora occurs; finally, hospital-acquired nosocomial bacteria may step in.

This concept is essential to understanding the roles of hygiene and environmental control measures, adequate surgery, and the use of antibiotics: i.e. the *operation-related risk factors*.[3] The widespread availability of antibiotics – and their abuse – during the last fifty years has further complicated the bacteriology of wounds because of the hospital selection of resistant strains. Whether infection follows contamination, and what type of infection develops, depends on a number of factors.

The major threats faced by the war-wounded are the development of gas gangrene, tetanus, and invasive streptococcal bacteraemia, which can all be fatal.

2 Fleming A. On the bacteriology of septic wounds. *Lancet* 1915; **186**: 638 – 643.

3 Rubin RH. Surgical wound infection: Epidemiology, pathogenesis, diagnosis and management. *BMC Infect Dis* 2006; **6**: 171. Available at: http://www.biomedcentral.com/1471 – 2334/6/171.

13.3 Major clinical infections of war wounds

The polymicrobial nature of wound contamination can result in a variety of infections. Several specific lethal infections are notorious in war wounds, especially in neglected or mismanaged wounds, and particularly if they have been stitched closed without proper excision. Common pyogenic wound infection runs through a spectrum of disease from the minor (superficial surgical site infection) to the severe (organ or space surgical site infection) and systemic (septicaemia).

13.3.1 Definitions

Simple contamination

The surgical literature classifies wounds as clean, clean-contaminated, contaminated, and infected or dirty. War wounds are considered contaminated *and* dirty.

Cellulitis/local abscess formation

Bacteria begin to spread to tissues that are contiguous to the wound, in the zone of contusion (see Chapter 3). Little systemic toxicity exists. Superficial erythema and seropurulent discharge occur; some necrotic tissue is evident, but this is related to the original injury and not to the infection. Eventually a mass of necrotic tissue and bacteria forms; if drainage is inadequate (small skin wounds), an abscess develops. In a large wound, the mass will be walled off and then expelled.

Myositis/deep tissue infection

Bacteria spread into muscles and tissues further than the permanent cavity and the zone of contusion and invade the zone of concussion and beyond. Systemic symptoms become more pronounced and, with certain infections, come to dominate the clinical picture. Overwhelming infections usually present with the classical signs of sepsis (fever, tachycardia, etc.); but they may also present without these signs, particularly in immune-compromised patients (HIV/AIDS, chronic disease, malnutrition).

In the past, deep-tissue infections were usually caused by Clostridia and invasive β-haemolytic streptococcus and accounted for the greatest amount of tissue damage, and the most severe systemic signs and symptoms. The availability of penicillin has radically changed this clinical picture.[4]

Different types of invasive tissue infection have been described:

- pyogenic deep tissue infection – this is usually a mixed infection, Gram-positive and -negative organisms, often including non-clostridial anaerobes;
- gas gangrene: this infection is still seen in neglected and mismanaged wounds;
- tetanus: poverty and the disorganizing effects of war disrupt immunization programmes in many countries – this infection is still, all too often, a major killer;
- invasive streptococcal infection – also still seen in neglected and mismanaged wounds;
- bone infection: when deep-tissue infection extends to sequestrated bone fragments chronic sepsis tends to develop with the production of a "biofilm" – this is often seen in neglected and mismanaged wounds (see Chapter 12);
- necrotizing fasciitis, synergistic gangrene and anaerobic cellulitis.

4 Polhemus ME, Kester KE. Infections. In: Tsokos GC, Atkins JL, eds. *Combat Medicine: Basic and Clinical Research in Military, Trauma, and Emergency Medicine*. Totowa, New Jersey: Humana Press; 2003: 149 – 173.

13.3.2 Gas gangrene

Gas gangrene as a peculiar complication of wounds has been known since the time of Hippocrates. Its occurrence in war wounds throughout history as well as in other trauma has been widely described mainly because of its dramatic manifestations and very high associated mortality. Delay to surgery, as may occur during war or in a civilian disaster with mass casualties, greatly increases the risk of gas gangrene.

Gas gangrene is a rapidly-spreading oedematous myonecrosis occurring characteristically in severe wounds in muscles contaminated with pathogenic obligatory anaerobes, particularly *Clostridium perfringens*. Almost every case presents a mixed bacterial flora; aerobes use up available oxygen and promote the anaerobic environment required by Clostridia.

In large muscle wounds there will always be areas of ischaemia and the potential for gas gangrene. But gas gangrene can also develop even when the trauma is not very severe. If the wound is deep, contains necrotic tissue, and is *isolated from the surface*, an anaerobic environment is formed and it is possible for a clostridial infection to become established. This deep anaerobic environment is seen particularly in wounds sutured closed without debridement. The presence of foreign material in wounds, especially soil, will increase the risk of developing gas gangrene. The prolonged application of tourniquets or tight plasters and fascial compartment syndrome also present high risks. The infection is more common in wounds to the lower limb and perineum than to the upper limb.

The features of the disease result, first, from the local action of the organisms on muscle glucose, producing acid and gas, and on muscle protein, causing digestion. Second, the organisms produce soluble, very potent toxins, which diffuse into the tissues causing further tissue destruction and profound toxaemia. The muscle breakdown products are very toxic in their own right. The combination of breakdown products and specific toxins causes the relentless and profound toxaemia which, if untreated, will inevitably lead to death.

Early and adequate surgery will prevent gas gangrene in war wounds.

The *incubation period* of gas gangrene is usually short, almost always less than three days, and in the majority of cases less than 24 hours long; it may even be as short as 6 hours. Incubation periods of up to six weeks have occasionally occurred.

Typically, gas gangrene manifests itself with the sudden appearance of increasing pain in the region of the wound. The *sudden* onset of pain, sometimes so sudden as to suggest a vascular catastrophe, always suggests the possibility of gas gangrene in a wounded person. Soon afterwards the limb becomes oedematous and is associated with drainage of thin serous or serosanguinous exudate, which may become gelatinous. The pulse rate rises markedly but the patient's temperature is rarely higher than 38° C. As classically described, clinical deterioration takes place quickly and within several hours the patient becomes anxious and frightened, or may even be euphoric, and exhibits all the signs of severe sepsis.

The skin is tense, pale, often marbled with blue, and rather colder than normal. In untreated cases, local bronzing of the skin becomes more diffuse, greenish yellow areas appear in which blebs may form and become filled with dark red fluid, and patches of cutaneous gangrene may occur. The skin *may*, however, appear normal, even when lying over massive gangrene.

R. Coupland / ICRC

Figure 13.1.1

Gas gangrene of the leg following the application of a tourniquet.

J. Goosen / South African Military Health Service, U. Witwatersrand

Figure 13.1.2

Gas gangrene of the arm.

Rapidly, swelling oedema and toxaemia increase, the serous discharge becomes more profuse and a peculiar sweetish smell may be present. The smell is variable and not pathognomonic.

Gas is usually produced at this stage and is partly responsible for the swelling of the affected part (Figure 13.2). It is produced in and between the muscle fibres, follows the fascial planes, and eventually escapes into the subcutaneous tissues under pressure through holes in the fascia, spreading rapidly beyond the confines of the infected area. Crepitus may be felt under the skin if the part is palpated. The extent of the spread of gas is not equivalent to the extent of muscle necrosis.

Infection spreads up and down the muscle from the site of the wound but has little tendency to spread to other healthy muscles; the bacteria is a *strict anaerobe*. Even in well-established gas gangrene, the bloodstream is rarely invaded by Clostridia until immediately before death.

The muscle changes are usually only seen at operation. In the early stages, they consist of little more than swelling and pallor. Later, the colour alters to a lustreless pinkish-grey, then to brick red, and finally to a dark green-purple colour.

E. Dykes / ICRC

Figure 13.2

X-ray showing gas collections between muscle groups.

Management

All patients with missile wounds should receive prophylactic antibiotics that are effective against Clostridia, such as penicillin, metronidazole, or erythromycin. But the antibiotics can only reach tissues with a blood supply and good tissue perfusion. Antibiotic concentrations in the dead muscle in the depths of an anaerobic environment are ineffective. Surgery as early as possible is the goal to be attained.

The basis of management of established gas gangrene is *thorough* excision of dead tissue, which might necessitate *urgent amputation*. The limit of excision is *necrotic muscle*, not the extent of tissues containing gas. Tissues must be cut away until healthy red, bleeding, and contracting aerobic muscle is reached.

Immediate surgical intervention is essential in the management of gas gangrene.

It is often too late for surgery when a patient presents with full-blown gas gangrene because of the severe toxaemia. Nonetheless, supportive measures including oxygen, fluids, and fresh whole blood should be instituted. Appropriate antibiotics should be continued.

13.3.3 Tetanus

Tetanus is always a risk with any penetrating wound. The risk is greater for severely contaminated wounds, especially in small, deep, punctuate-type ones. Pyogenic infection in the depth of a narrow track can create the necessary anaerobic environment. The incubation period is 3 to 21 days, but can be as short as 1 day to as long as several months.

E. Dykes / ICRC

Figure 13.3
Badly infected wound giving rise to tetanus.

Virtually total protection can be obtained by active immunization with a full course of tetanus toxoid injections.

Å. Molde / ICRC

Figure 13.4
Risus sardonicus in a patient suffering from tetanus.

Clostridium tetani is a strict anaerobe. The organism produces an extremely potent toxin called tetanospasmin that spreads along peripheral nerves to the spinal cord and brainstem. The toxin affects the motor end-plate by inhibiting cholinesterase, resulting in a build-up of acetylcholine and tonic muscle spam. There is, in addition, hyperexcitability of lower motor-neurones causing muscular rigidity and dysfunction of activity in antagonistic muscles that cause unopposed reflex activity, giving rise to the typical spastic phenomena of tetanus. Early signs are neck rigidity and *trismus* (stiffness of the jaw) with difficulty in swallowing, followed by *risus sardonicus* (grimacing smile). Profound autonomic nervous system dysfunction is also present in severe cases with unstable pulse rate, blood pressure, and temperature, which are difficult to treat. Once the toxin is fixed to the nerves, antitoxin can not longer neutralize it.

The infection may affect a single muscle group or be more generalized. Three clinical degrees are described:
- mild = no generalized spasms;
- moderate = generalized spasms on provocation;
- severe = spontaneous generalized spasms with opisthotonus.

The muscular spasms are quite painful, and can last a short time or go on for several weeks. The greatest danger is asphyxiation due to laryngeal spasm or aspiration. Frequently, there is a high temperature with profuse sweating that requires careful fluid replacement. The patient remains conscious throughout.

Prophylaxis
The best precaution against tetanus is active immunization. However, many patients in developing countries where conflict has further disrupted public health programmes are not immunized and therefore the risk of tetanus in war wounds is great. In some countries, ICRC practice is to proceed with active and passive immunization for *all* patients.

Early and adequate excision of dead tissue, and leaving wounds open, is vital to prevent its development. This is particularly important in small, deep, punctuate-like wounds. Penicillin and metronidazole are the drugs of choice.

Tetanus prophylaxis for war wounds

All patients whatever their immunization status:

1. tetanus toxoid vaccine 0.5 ml i.m. (5 LF units) – this is a booster dose in a previously vaccinated patient;
2. penicillin;
3. thorough excision of the wound.

Non-immunized patients or those in whom immunization status is in doubt, addition of:

4. human tetanus immunoglobulin 500 IU i.m. (adults) or 250 IU (children under 15 years of age) – also known as human anti-tetanus serum;
5. tetanus toxoid 0.5 ml i.m. to be repeated at four weeks and again at six months.

Please note:

Vaccine and immunoglobulin should be administered through separate syringes and at separate sites.

Treatment of established tetanus

Treatment of established tetanus should follow the steps listed below.

1. Extensive debridement of the wound. Ketamine anaesthesia should be avoided if possible as patients waking from this form of anaesthesia are often restless and have hallucinations and this may stimulate spasms. Spinal anaesthesia or a regional nerve block are excellent methods.
2. Antibiotics: crystalline penicillin G (5 MIU. i.v. QID) and/or metronidazole (500 mg i.v. TID). Erythromycin, tetracycline and chloramphenicol are also active against Clostridia, in case of allergy to penicillin.
3. Anti-tetanus human immunoglobulin (3,000 – 10,000 IU): a single large dose is given intravenously and as soon as possible. The actual dose ordered is dependent on the severity of the disease and the age of the patient. It is diluted in 20ml of normal saline and given slowly over a period of 15 minutes. This can be given intramuscularly (undiluted) in the proximal part of the wounded limb if the type of immunoglobulin is not recommended for intravenous use. An intrathecal injection of 250 IU can be given e.g. with spinal anaesthesia for debridement of the lower part of the body.

   In some regions, human immunoglobulin is in short supply or not available at all. In these circumstances, one must rely on equine anti-tetanus serum. A test dose must be administered before giving the full dose (20 000 IU).
4. Control of the spasms: the patient should be nursed in an environment as devoid of stimulation as possible in a dark, quiet room. For severe cases, and in the best of circumstances, the patient should be in an intensive care unit, anaesthetized and paralysed, with mechanical ventilation. Many hospitals lack these facilities; control is based on sedation.

Intravenous diazepam, either intermittently or continuously, is used with doses up to 20 mg or more per hour depending on the duration and intensity of the spasms. If diazepam alone does not control the spasms, chlorpromazine may be added, or thiopental (pentothal) may be required. This latter is given preferably by continuous infusion, which has been found to control the spasms more effectively with lower total dosages than if given intermittently.

Warning!
Oversedation can be as fatal as the disease.

Recently, ICRC teams have used magnesium sulphate in high doses (40 mg/kg over 30 minutes, followed by i.v. infusion of 1 – 3 g/hour) with encouraging results. This regime allows lower doses of diazepam and thiopental to be used, thus avoiding oversedation: this line of treatment is still under investigation.[5]

5. Airway management is of the utmost importance to prevent aspiration. If laryngospasms persist tracheostomy may be needed, and the decision to perform one should be taken early. Care of the tracheostomy tube is instituted with particular attention to how much this stimulates spasms, and adapting the sedation accordingly. Chest physiotherapy is required to prevent respiratory complications.
6. Fluid replacement: there may be excessive fluid loss due to sweating – this needs to be carefully monitored along with urinary output. A Foley urinary catheter should be placed.
7. Nutrition: if the spasms continue for some days, a feeding gastrostomy or jejunostomy may be required as a nasogastric tube may provoke spasms and become obstructed. Again, it is important to decide to perform this procedure early. These patients require a high calorie diet.
8. Excellent nursing care is fundamental. The patient's position must be changed four-hourly. Extra sedation may be required prior to this, as movement can provoke spasms. Patient hygiene – skin, oral and bowels – is essential.

Please note:

Clinical tetanus does *not* confer normal immunity. The patient therefore needs to complete immunization after recovering from the disease.

13.3.4 Invasive streptococcal infection

Beta-haemolytic streptococcus produces powerful toxins that trigger off a rapid progression of disease from local to systemic infection. It begins with a local cellulitis with severe pain out of proportion with physical findings, and advances rapidly to fever, tachycardia, disorientation and delirium. The wound is malodorous with large quantities of thin, blood-stained seropurulent discharge. The muscles suffer from a moist oedema, and turn from pale, to bright red, then dark purple-brown. Progression can occur in a matter of hours. It is still seen in mismanaged wounds, and especially if there is a delay in treatment.

Wound debridement, drainage, and massive doses of penicillin are the basis of management, along with supportive measures. Blood transfusion is usually necessary.

ICRC

Figure 13.5

Invasive β-haemolytic streptococcus infection of the abdominal wall.

5 Thwaites CL, Yen LM, Loan HT, Thuy TTD, Thwaites GE, Stepniewska K, Soni N, White NJ, Farrar JJ. Magnesium sulphate for treatment of severe tetanus: a randomised controlled trial. *Lancet* 2006; **368**: 1436 – 1443.

13.3.5 Necrotizing soft-tissue infections

Various non-clostridial spreading gangrenous infections have been described in the surgical literature: necrotizing fasciitis, synergistic gangrene, anaerobic cellulitis, etc. There is necrosis of the skin, subcutaneous tissues and fascia, but not of the muscles. It can rapidly progress to systemic toxicity and even death. Spreading gangrene can follow trauma or thoracic and abdominal surgery (postoperative synergistic gangrene), and affect persons suffering from a general disease, such as diabetes, or malnutrition; alcoholics are particularly at risk. Infection usually affects the torso, perineum or extremities. The bacterial flora is mixed causing synergy amongst various organisms, anaerobic and aerobic, particularly microaerophilic nonhaemolytic streptococci in association with haemolytic staphylococci.

In no case are healthy muscles involved, and the clinical picture is less dramatic than in gas gangrene. Nonetheless, the systemic toxicity is life-threatening. There is severe pain disproportionate to local findings, crepitus can often be felt and soft-tissue air is present on plain X-rays. The skin is discoloured (blue, purple or black) with blistering leading to haemorrhagic bullae and induration. The necrosis can spread widely causing a large soft tissue defect. The diagnosis is primarily a clinical one, and the earlier treatment is instituted the better the result.

Management includes:

- aggressive excision of necrotic tissue and the relief of tension, which may require serial debridements – amputation may be necessary in cases of necrotizing infections of the limbs – and defunctioning colostomy in cases of infection of the perineum with faecal contamination;
- leaving the wound open for drainage;
- triple intravenous antibiotics: penicillin, gentamycin and metronidazole;
- fluid resuscitation, blood transfusion and other supportive treatment.

Reconstructive surgery to cover the defect should only be considered once the patient's condition has stabilized and the infection is fully eradicated.

H. Nasreddine / ICRC

H. Nasreddine / ICRC

Figure 13.6.1 and 13.6.2

Necrotizing fasciitis with large skin loss and tissue defect.

13.4 Antibiotics

It is important to distinguish between the prophylactic use of antibiotics and their role in the therapy of established sepsis. The surgeon must also remember the life history of wounds and their changing bacterial flora.

Prophylaxis is aimed at preventing a specific infection. It is impossible to find any reasonable cocktail of antibiotics that would be effective against the entire "polymicrobial cesspool" that can contaminate a war wound. Nor is it desirable to do so. Such a practice would constitute a simple abuse of antibiotics and contribute to the development of bacterial resistance.

Old lessons for new surgeons

The best antibiotic is good surgery.

The surgeon must realize that trust in the efficiency of antibiotics will never replace good surgery. Good surgery involves good diagnosis, good clinical decision-making, and good holistic management of the patient. Sometimes "good surgery" means knowing when *not* to operate. As we have seen, uncomplicated Grade 1 wounds and many civilian gunshot wounds can be treated conservatively and expectantly.

War wounds, however, are usually understood to be different from civilian weapon trauma; not always because of the weapon, but because of the environment. The battlefield is a dirty and contaminated place and the danger of invasive infection is always present, even in minor wounds; mass casualties cannot always be followed up correctly; there is often a long delay between injury and treatment; proper hygiene and nutrition cannot always be maintained; and immunization is not always universal.

Under such circumstances, the basis of prevention of primary infection remains complete wound excision and good drainage, respect in the handling of the tissues, and leaving the wound open for delayed primary closure. Antibiotics are only an *adjunct* to good surgical practice, and cannot replace poor surgery.

A case can be made for prophylaxis against the second series of infections, those acquired from the flora of the patient (skin, respiratory and gastrointestinal tracts) if the occurrence of such infections becomes an important clinical problem in the functioning of a given hospital, and is proven by proper bacteriological studies. Infections caused by bacteria with multiple antibiotic resistance and by opportunistic organisms, such as *Pseudomonas aeruginosa*, have become more prevalent with the uncontrolled use of wide-spectrum antibiotics. Again, good surgery and proper hygiene measures and environmental control cannot be replaced by antibiotics, and their use should be regarded as an adjunct only.

The prevention of nosocomial infection is another story. Proper clinical procedures and hygiene are the correct means of prevention: frequent hand washing; no unnecessary change of dressings on the ward; isolation of infected patients; adequate sterilization; proper cleaning of the hospital premises, etc. These alone can and will stop nosocomial infection. The use of antibiotics to *supplement* such measures will depend on the *virulence* of the particular bacteria involved.

A well-functioning bacteriology laboratory is important if antibiotic use is to be anything more than a "shotgun" approach or an "educated guess". Wound cultures are notorious for not predicting subsequent infections or infecting pathogens. In the absence of correct collection of specimens, aerobic and anaerobic, and proper bacterial culture and sensitivity techniques, an educated guess is the best that can be hoped for in most cases. Accurate bacteriological capability is a far more difficult endeavour in a forward hospital than most people realize.

13.4.1 Antibiotic prophylaxis in the primary treatment of war wounds

Antibiotics do not reach the source of infection in a missile wound – i.e. the culture medium of dead tissue, debris and foreign material – they only affect the contusion and concussion zones around the wound. However, early administration of antibiotics seems to inhibit bacterial growth and delay invasive infection. Antibiotics especially help prevent spread to the blood stream. The benefit of antibiotics to prevent invasive infections in wounds that received no other treatment for hours to days after injury has been demonstrated,[6] and ICRC clinical experience confirms this. The ICRC recommends that penicillin treatment be begun in the pre-hospital setting if at all possible.

As mentioned, historically, the biggest killers of surviving war-wounded casualties have been primary infection by β-haemolytic streptococcus and Clostridia. This is a relatively narrow spectrum of bacteria, for which penicillin remains the best antibiotic.

Prophylaxis for war wounds can only be assured for Clostridia and β-haemolytic streptococcus: penicillin is the drug of choice.

6 Mellor SG, Cooper GJ, Bowyer GW. Efficacy of delayed administration of benzylpenicillin in the control of infection in penetrating soft-tissue injuries in war. *J Trauma* 1996; **40 (3 Suppl.)**: S128 – S134.

Single-dose or the administration of prophylactic antibiotics limited to 24 hours are indicated under optimal conditions of rapid evacuation, early pre-hospital first aid, and adequate infrastructural hygiene. In the ICRC context of limited resources, often less than ideal environmental control, and delayed evacuation, antibiotics are usually given for 5 days, until delayed primary closure. This period of 5 days may be regarded as an exaggeration by some colleagues, but is considered a reasonable combination of a prophylactic *and* therapeutic regime by ICRC practitioners and has proved its worth in clinical experience (see Annex 13. A: ICRC antibiotic protocol).

Topical antibiotics and the washing of wounds with antibiotic solutions are not recommended.

13.4.2 Cephalosporins and others

The ICRC does not routinely use cephalosporins or other sophisticated antibiotics in a surgical setting. In many areas where the ICRC works, typhoid fever and shigellosis, amongst other infections, are endemic diseases. Resistance to first-line antibiotics for these fatal diseases is becoming more and more of a problem in clinical practice. Often, cephalosporins are the only recourse left available. The judicious use of antibiotics is an important element to keep in mind when dealing with the surgery of the victims of war, where proper surgical, nursing, and hygiene measures should be the basis of prevention and treatment of infection.

13.5 Neglected or mismanaged wounds

The particular, although commonly encountered, condition of neglected or mismanaged wounds has been dealt with in Chapter 12. As mentioned above, these wounds are particularly at risk of developing gas gangrene, tetanus, and deeply invasive sepsis. According to the ICRC antibiotic protocol, penicillin and metronidazole should be administered; if there are systemic signs of pyogenic infection, gentamycin is added.

Antibiotics are important in the treatment of these diseases, but do not replace the essentials of septic surgery: wide excision, good drainage, copious lavage, and leaving the wound wide open.

ANNEX 13. A ICRC Antibiotic protocol

Injury	Antibiotic	Remarks
Minor soft-tissue wounds, uncomplicated Grade 1	Penicillin-V tablets 500 mg QID for five days	Anti-tetanus measures for all weapon - injured patients
Compound fractures Traumatic amputations Major soft-tissue wounds (Grades 2 and 3)	Penicillin-G 5 MIU i.v. QID for 48 hours Follow with penicillin-V tablets 500 mg QID until DPC	Continue penicillin-V for five days if closure is performed with a split skin graft If redebridement is performed instead of DPC: stop antibiotic unless there are signs of systemic infection or active local inflammation – in latter case, add (metronidazole 500 mg i.v. TID) and gentamycin 80 mg i.v. TID
Compound fractures or major soft-tissue wounds with delay of more than 72 hours Anti-personnel landmine injuries to limbs whatever the delay	Penicillin-G 5 MIU i.v. QID and metronidazole 500 mg i.v. TID for 48 hours Follow with penicillin-V tablets 500 mg QID and metronidazole tablets 500 mg TID until DPC	
Haemothorax	Ampicillin 1 gm i.v. QID for 48 hours, followed by amoxycillin tablets 500 mg QID	Continue until two days after removal of the chest tube
Penetrating cranio-cerebral wounds	Penicillin-G 5 MIU i.v. QID and chloramphenicol 1 gm i.v. TID for at least 72 hours	Continue i.v. or orally according to patient's condition for a total of 10 days
Brain abscess	Same regime as for penetrating cranio-cerebral wounds plus metronidazole 500 mg i.v. TID	Continue i.v. or orally according to patient's condition for total of 10 days
Penetrating eye injuries	Penicillin-G 5 MIU i.v. QID and chloramphenicol 1 gm i.v. TID for 48 hours	Continue i.v. or orally according to patient's condition for total of 10 days Local instillation of antibiotic eye-drops
Maxillo-facial wounds	Ampicillin 1 gm i.v. QID and metronidazole 500 mg i.v. TID for 48 hours	Continue i.v. or orally according to patient's condition for total of 5 days
Abdominal wounds: 1. solid organs only: liver, spleen, kidney; or isolated bladder injury 2. stomach, small intestines 3. colon, rectum, anus	 Penicillin-G 5 MIU i.v. QID Ampicillin 1 gm i.v. QID and metronidazole 500 mg i.v. TID Ampicillin 1 gm i.v. QID and metronidazole 500 mg i.v. TID and gentamycin 80 mg i.v. TID	Continue for 3 days depending on drainage

MIU = million international units.

QID = quater in die (4 x day).

TID = ter in die (3 x day).

Please note:

This protocol was established by the ICRC Master Surgeons Workshop held in Geneva in 2002 (see Introduction).

Chapter 14

RETAINED BULLETS AND FRAGMENTS

14. RETAINED BULLETS AND FRAGMENTS

14.1 The surgeon and the foreign body 271

14.2 Early indications for removal 271

14.3 Late indications 274

14.4 Technique for the removal of a projectile 275

14.1 The surgeon and the foreign body

There are hundreds of thousands of perfectly healthy people walking around the present and past war zones of the world who have retained projectiles in their body. These metallic foreign bodies often catch the attention of patients and relatives, who impute to them any pains and disabilities they suffer from and insist upon their removal, even after the wound has completely healed.

It can be very difficult, and frustrating, for the surgeon to point out to an insistent patient that a projectile is dangerous when moving but is no longer so when it has come to a stop. The incidence of late infection is low (2 – 3 %)[1] and vascular embolization rare and more common with fragments than bullets. The risk of embolization – "migrating bullet" – is more common in published clinical cases in the specialized surgical literature than it is in real life. One American study from Viet Nam reported a rate of 0.3 % in 7,500 cases with arterial injuries; and rarer still in the overall wound population.[2]

The frustration is directly related to the number of patients presenting with such requests, who can at times be counted literally in the hundreds. The problem tends to be psychological and cultural rather than surgical. How can one convince a young combatant that the operation to remove a bullet may cause more damage than the bullet has accomplished? Such a foreign body in the midst of a muscle mass, where there are no pain receptors, is not painful and the surgeon must find a way to overcome the patient's subjective fixation; perhaps by making the bullet's presence a matter of "pride".

The experience of the ICRC shows that many of these operations to remove a projectile are risky, useless, consume time and materials, and result in new surgical trauma with its possible complications, often without identification and removal of the foreign body. There are few surgical endeavours more frustrating for an unprepared young surgeon than to spend two hours searching ever more frantically for an "easy" fragment or bullet, and finally not finding it.

14.2 Early indications for removal

However, as stated previously, there are indications for removal of bullets and fragments, which may be divided into early and late. The most important early indications have been dealt with in Chapter 10: as an integral part of the primary operation or to address the possibility of erosion of an important structure (Figures 10.11, 10.12 and 14.1).

As mentioned in the latter case, whether the removal is an immediate or planned procedure will depend on a number of factors, especially the experience of the surgeon and the possible morbidity and lethality of the operation itself compared to the relatively low incidence of serious complications. The decision is not always a simple one and the reader may consider what should be done in the following examples (Figures 14.2 – 14.8).

M. Baldan / ICRC

Figures 14.1.1 and 14.1.2
Bullet lodged in full-term pregnant uterus.

M. Baldan / ICRC

1 Rhee JM, Marin R. The management of retained bullets in the limbs. *Injury* 1997; **28**: 23 – 38.
2 Rich NM, Collins GJ, Andersen CA, McDonald PT, Kozloff L, Ricotta JJ. Missile emboli. *J Trauma* 1978: **18**: 236 – 239.

M. Baldan / ICRC

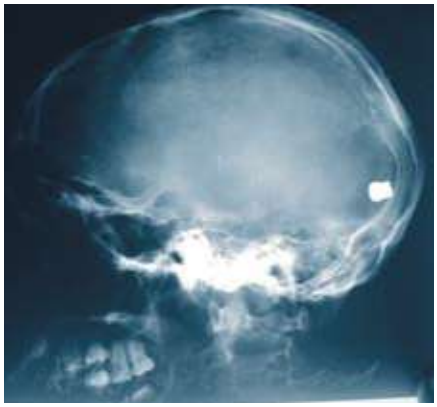

M. Baldan / ICRC

Figures 14.2.1 and 14.2.2

A fragment entered the forehead and is retained in the occipital region.

ICRC

ICRC

ICRC

Figures 14.3.1 – 14.3.3

A fragment has entered the parieto-occipital area.

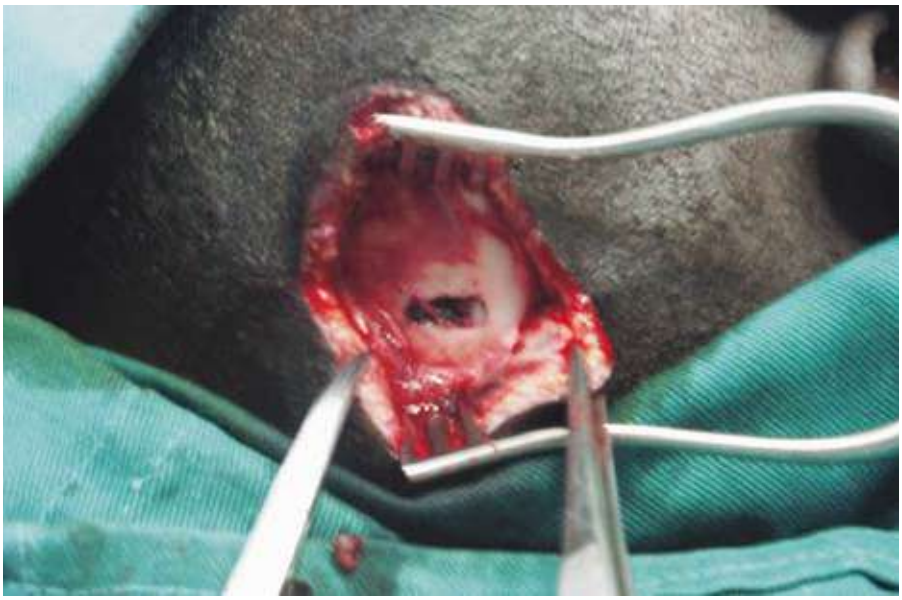

ICRC

Figures 14.3.4

The fragment is evident under the bone.

R. Coupland / ICRC

Figure 14.4.1

Nurse pointing out site of fragment entry.

R. Coupland / ICRC

Figure 14.4.2

Radiograph showing a fragment retained in the neck.

M. Baldan / ICRC

Figure 14.5

Bullet in the axilla: the vessels and nerves are intact.

M. Baldan / ICRC

Figure 14.6

Fractured humerus with bullet retained in the subscapular muscles.

M. Baldan / ICRC

M. Baldan / ICRC

Figures 14.7.1 and 14.7.2

Extra-synovial bullet near the hip.

F. Herkert / ICRC

Figure 14.8

Fragmented bullet in extraperitoneal muscle mass.

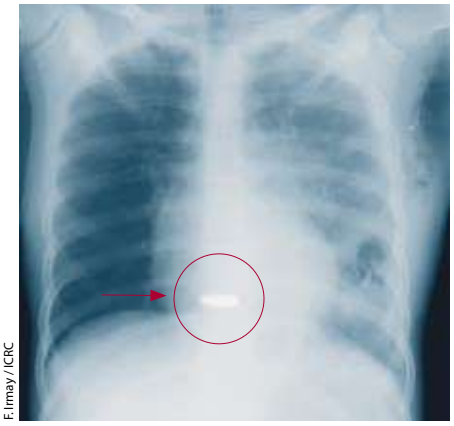

F. Irmay / ICRC

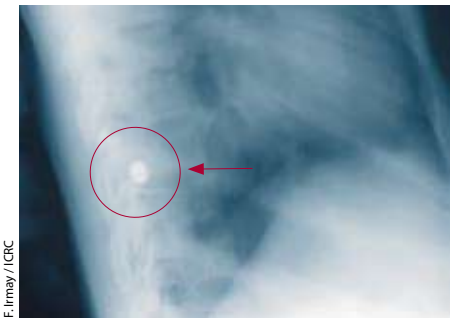

F. Irmay / ICRC

Figures 14.9.1 and 14.9.2

Bullet retained in vertebral column, the patient is paraplegic.

ICRC

Figure 14.10

Metallic fragment located in a pressure point: sole of the foot.

The following scenarios should be added to the early indications already mentioned.

- A small fragment in a heart chamber – the concomitant pericardial tamponade is an urgent condition, the removal of a retained fragment is not – or a fragment obviously within the lumen of an important vessel. The possibility, and danger, of embolization is high in these particular instances. Operation should *only* be undertaken, however, if the surgeon *and* anaesthetist are experienced and the necessary diagnostic and operative equipment are available.
- A projectile located in the spinal cord: emergency laminectomy and removal should be considered *only* if there is clear progressive neurological deficit and radiological evidence of spinal cord compression by the foreign body *and* the surgeon is experienced in this field. A definitive diagnosis of paraplegia, indicating that the spinal cord is irrecoverably sectioned, means that it is already too late.
- Small metallic fragment penetrating the eye: removal is indicated if located in the anterior chamber, again, *only* if the surgeon is experienced, *and* magnification, proper instruments and suture material are available.

14.3 Late indications

Late indications are largely the result of the complications listed below.

Sepsis

If the projectile, along with surrounding contaminants, is acting as a nidus for infection: an abscess or sinus. The removal is performed as a planned operation after the necessary diagnostic procedures (X-ray, sinugram, etc.).

Pain

If located superficially, especially if over a pressure point (palm of hands, sole of foot, elbow, etc.), the projectile will cause true pain. The removal may be performed as an elective procedure after healing from the acute trauma.

Pain

If the foreign body is impinging on a nerve and causing radicular pain or paraesthesia. As above, the removal is performed as a planned operation.

ICRC

Figure 14.11.1

Bullet in antero-lateral compartment of the leg pressing on common peroneal nerve.

ICRC

Figure 14.11.2

Bullet not pressing on a nerve.

Lead toxicity

This is extremely rare, usually involves a synovial joint or intervertebral disc, and surgery should be contemplated only if there is documented elevation of serum lead level (above 10 micrograms/dl in children and 40 micrograms/dl in adults). The response to an EDTA (ethylene diamine tetra-acetic acid) challenge is a useful diagnostic test. Standard chelation therapy is instituted (EDTA, dimecaprol, d-penicillamine, succimer) and the surgical removal of the metal should not be performed before serum lead levels have been reduced to avoid acute lead poisoning.[3] This complication does not arise with an intact, full metal jacket bullet where no lead filling has spilled out.

We shall not deal with wounds not related to armed conflict where the bullet may be required as part of a legal and forensic investigation. National law of the country concerned determines what exactly can be done while not subjecting the patient to unnecessary harm and suffering.

14.4 Technique for the removal of a projectile

For those patients requiring removal of a deeply-seated metallic fragment or bullet, proper localization prior to surgery is of the utmost importance. Most rural or provincial hospitals do not have fluoroscopy or an image-intensifier in the operating theatre enabling the surgeon to operate under direct vision.

A simple stereotactic technique to assist localization involves taping a series of radio-opaque objects (paper clips, injection needles, Kirschner wires) to the body part, both anterior and lateral surfaces, to act as guides. Plain X-rays with antero-posterior and lateral views are taken. The radio-opaque object is then removed and its position on the skin marked with a felt pen.

By looking at the two X-ray views, the surgeon can estimate the relative distance of the bullet from the radio-opaque objects in the axes both across and in the depth of the body part – e.g. half-way between the second and third paper clips in the A-P view and one-third of the way between the first and second paper clips in the lateral view. The idea is to use two-dimensional X-rays to extrapolate into three dimensions. The films should be taken on the morning of the operation and the patient should be undressed (combatants often carry bullets in their pockets).

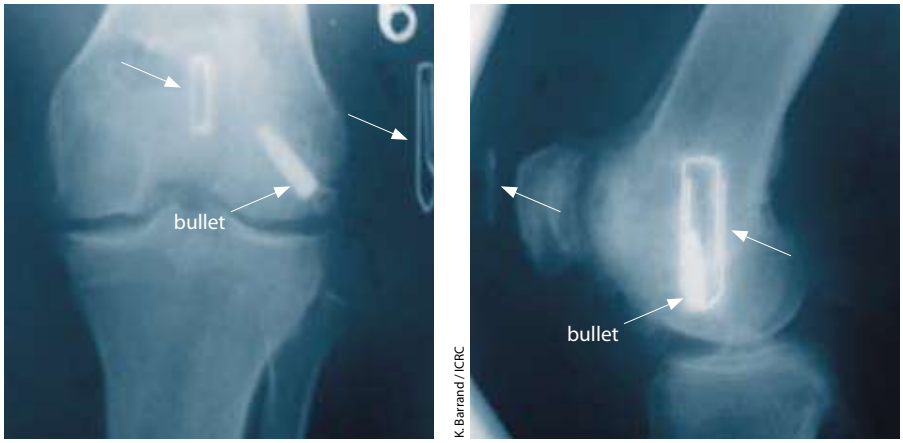

K. Barrand / ICRC

K. Barrand / ICRC

Figure 14.12.1 and 14.12.2

Stereotactic localization of metallic foreign body. Safety-pin is anterior and paper clip lateral.

In due course, the body will form tough avascular scar tissue around the metallic fragment: a foreign body granuloma. This includes other foreign matter and dirt, and should be excised together with the projectile.

3 Linden MA, Manton WI, Stewart RM, Thal ER, Feit H. Lead poisoning from retained bullets: Pathogenesis, diagnosis, and management. *Ann Surgery* 1982; **195**: 305 – 313.

Chapter 15

BURN INJURIES

15. BURN INJURIES

15.1 Introduction 279

15.2 Pathology 279
15.2.1 Burn depth 279
15.2.2 Pathophysiological changes 280
15.2.3 Types of burns 281
15.2.4 Extent of burns 281

15.3 Burn management 282
15.3.1 First aid 282
15.3.2 Resuscitation 282
15.3.3 Initial fluid replacement therapy 284
15.3.4 Monitoring resuscitation 285
15.3.5 After 48 hours 285

15.4 Burns presenting late 286

15.5 Nutrition 286

15.6 Care of the burn wound 286
15.6.1 Initial wound management 287
15.6.2 Local care 288

15.7 Closure of the burn wound 289
15.7.1 Mechanical cleansing and debridement 290
15.7.2 Surgery 290

15.8 Scar management 293

15.9 Electrical burns 294

15.10 Chemical burns 294
15.10.1 Acid and alkali burns 294
15.10.2 Phosphorus burns 294
15.10.3 Napalm injuries 295
15.10.4 Magnesium 296
15.10.5 Vesicant chemical weapons 296

ANNEX 15. A Nutrition in major burns: calculating nutritional requirements 297

15.1 Introduction

Burns are a common event in a war environment. Flame weapons, explosive blast and ignition of combustible materials all create burn hazards. The causal agent may be thermal, chemical, electrical, or linked to radiation. Each cause has specific consequences that may require specific assistance.

A serious burn injury is a painful and life-threatening event requiring an inordinate amount of hospital resources and nursing care. The most immediate threat to life is an impaired airway due to inhalation of hot air and smoke with consequent oedema; the most important subsequent threats to life are from hypovolaemic shock and infection, followed by the complex pathophysiological effects which continue after injury. It is associated with many complications, prolonged morbidity, multiple operations, and great demands on equipment, materials, medical and nursing time. The long-term sequelae, physical, cosmetic and psychological, deeply affect the morale of both patients and staff. Modern burn centres have made great progress in the successful treatment of major burns, but such facilities are never available in situations of limited resources.

Nevertheless, the principles of treatment are the same, and the aim is to do the best one possibly can under the austere circumstances, where stark reality imposes its frustrating limitations. Surgery will have the greatest impact on young people with small but potentially crippling injuries such as bilateral hand burns. Patients with burns covering 40 – 50 % of the total body surface will rarely survive under field conditions and are best treated by giving sufficient fluids to relieve thirst, and generous amounts of analgesia. In a triage situation of mass casualties they are considered Category IV: i.e. given supportive treatment only.

15.2 Pathology

15.2.1 Burn depth

Burns involve varying amounts of injury to the skin, partial or full thickness, and classically correspond to three degrees of burn depth of increasing severity (Figure 15.1).

First degree burns:
First degree burns, or superficial burns, are painful, red, and have no blisters. They will heal on their own.

Second degree burns:
These may be superficial or deep partial-thickness burns. They are invariably blistered and the blister floor is usually pink or mottled red in appearance and the surface is moist. The less they blanch with pressure, the deeper they are. They are painful and have some sensation to pinprick. Hairs will be difficult to pluck out if the base of the follicle is alive. Most will eventually heal by a combination of re-epithelialization and wound contraction, but often with a seriously disabling scar contracture. Some will benefit from appropriate skin grafting.

Third degree burns:
These refer to full-thickness burns. Destruction of the full thickness of the skin gives a charred, leathery or waxy appearance. Any surviving hairs pull out easily. The burns are usually dry and have no sensation. Injury may extend to the muscles and deeper tissues. They are normally caused by flames, immersion in very hot liquids, electric current or chemicals. Smaller full-thickness burns will eventually heal by contraction but this inevitably causes severe deformity and loss of function. They are best treated by skin grafting.

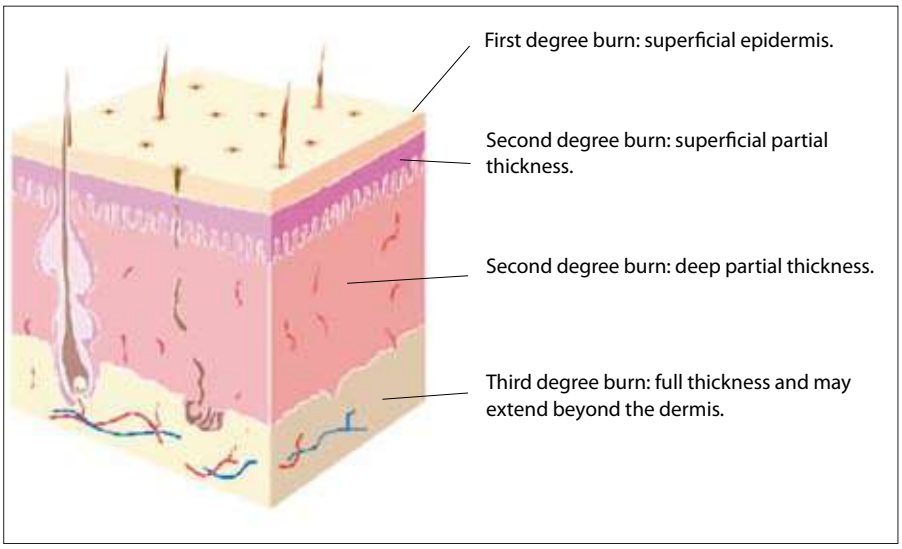

ICRC

Figure 15.1

Histology of the skin and degree of burn depth.

E. Dykes / ICRC

Figure 15.2

Different regions of a burn wound have different depths of damage: central eschar of full-thickness burn surrounded by areas of partial-thickness burn.

Different regions of a burn wound have different depths of damage. A burn is basically an ischaemic three dimensional wound with:

- a zone of coagulation – a central area of irreversible skin death that creates the eschar;
- a zone of stasis – a middle layer of damaged but viable tissue with a significant inflammatory reaction, but an early stasis of local blood flow;
- a zone of hyperaemia – a deep and peripheral area that looks like cellulitis, but is only hyperaemic.

Appropriate fluid resuscitation preserves cells in the zone of stasis, but subsequent infection or drying of the wound can rapidly enlarge the injury, both peripherally and in depth.

Superficial burn wounds are painful; deep burns are anaesthetic.

15.2.2 Pathophysiological changes

The most important pathophysiological change from thermal injury is increase in capillary permeability, which, if resuscitation is successful, returns to normal within 24 – 48 hours. Plasma water and proteins up to a molecular weight of 350,000 are freely exchanged between the intra- and extravascular compartments of the extracellular space. A strongly negative interstitial fluid pressure develops that creates a strong "suction" in the burnt tissue. If the affected area is very large, these changes become widespread with the loss of important amounts of fluid from the vascular compartment.

This loss is most marked in the region of the burn wound and accounts for the local oedema, which peaks at 6 – 12 hours post-burn. However, with aggressive crystalloid resuscitation in large wounds (>25 – 30% of the total body surface area or TBSA), the hypoproteinaemia causes a systemic disorder resulting in *generalized oedema* of non-injured skin and internal tissues, above all swelling of the larynx, and resultant obstruction to the airway, and abdominal compartment syndrome.

For further details on abdominal compartment syndrome, see Volume 2.

There is an associated rapid rise in the haematocrit which, together with polymerization of some plasma proteins, causes a marked increase in blood viscosity. The immediate danger linked to the sequestration of fluids in the extravascular space is hypovolaemic shock and, with haemoconcentration, acute tubular necrosis and renal failure. The loss of skin cover, and its function as a regulator of temperature, means that hypothermia, with its complication of coagulopathy, is an ever-present danger (see Chapter 18).

15.2.3 Types of burns

Flame and scald burns are by far the most common. Flame burns are usually deep and appear so at presentation, whereas scald burns may appear much less severe at first; experienced burn surgeons usually withhold prognostication until they have inspected them on the third day. Contact flame burns are typically very deep at the centre, which should be taken into consideration if they are to be operated on.

Electrical burns fall into two distinct categories. Flash burns occur when a person causes a short circuit and resulting electrical flash, but no current travels through the body. These may be treated as regular thermal injuries. High voltage (>1,000 volts) electrical conduction injuries – the current travels through the body and is characterized by the "can't let go" phenomenon – are "iceberg injuries" as they usually present with small cutaneous wounds and severe deep-tissue damage.

Chemical burns are caused by particular agents: acids, alkalis, and specific compounds (napalm, phosphorus, vesicants, etc.), with their individual characteristics.

15.2.4 Extent of burns

The sequestration of large amounts of fluid and plasma proteins in the extravascular space is a function primarily of the extent of burnt tissues. It is thus important to estimate the total body surface area that has been burnt. The depth of the burn must also be taken into account; only second and third degree burns are included in the estimation of TBSA.

The simplest calculation is best done using the "Rule of Nines" (Figure 15.3). The size of the *patient's hand* (including palm and fingers) is about 1 % of TBSA.

The head and neck of a child under one year of age represent about 18 % and the lower extremity 14 % of TBSA (Figure 15.4). With age these figures gradually change to adult proportions.

Although it is difficult to grade the severity of burn wounds, the following provides a rule of thumb.

Minor:
- second degree less than 15 % TBSA;
- third degree less than 3 % TBSA.

Moderate:
- second degree 15 – 25 % TBSA;
- third degree less than 10 % TBSA.

Major:
- second degree more than 25 % TBSA;
- third degree more than 10 % TBSA.

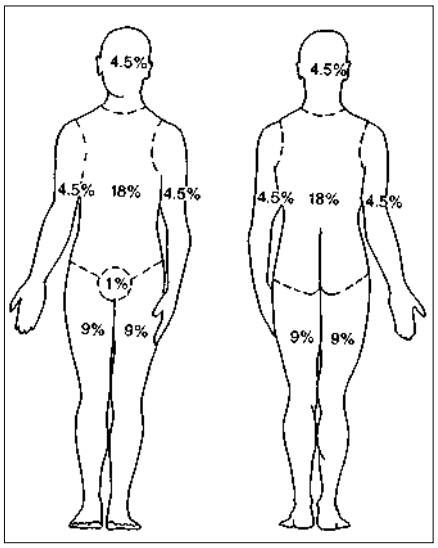

P. Zylstra / ICRC

Figure 15.3

"Homunculus" diagram showing the "Rule of Nines" to assess burnt surface area in an adult.

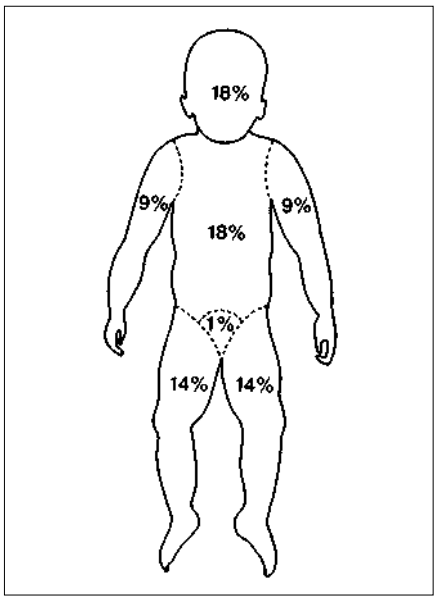

P. Zylstra / ICRC

Figure 15.4

Assessment of burnt surface area in a child "homunculus".

15.3 Burn management

The management of burns follows a series of standard measures as with any wound or trauma.

1. First aid.
2. Resuscitation:
   - airway;
   - breathing;
   - circulation / fluid replacement therapy.
3. Analgesia.
4. Prophylactic antibiotics.
5. Tetanus prophylaxis.
6. Nutrition.
7. Hypothermia.
8. Wound management.

15.3.1 First aid

Rescuers should first ensure that the scene is safe and take suitable precautions if un-ignited fuel, explosives, electricity, or chemical agents are present. The patient should be removed to safety and clean fresh air, and the vital signs assessed. If available, oxygen should be provided if there is any indication of smoke inhalation.

The burn should then be cooled with water or wet towels (for 20 minutes), and covered to reduce pain. The patient should not, however, be left wrapped in cold wet material so long as to provoke hypothermia. After cooling the *burn*, keep the *patient* warm.

If evacuation to the hospital is delayed and the airway is not at risk, the patient should be encouraged to drink plenty of small quantities of fluids often and regularly, and the colour and volume of the urine monitored.

15.3.2 Resuscitation

The following elements should be determined:

- nature of the causal agent – flame, scald, contact, electrical conduction, electrical flash, chemical;
- possible aggravating factors – additional trauma, smoke inhalation (fire in an enclosed space equals smoke inhalation);
- time since injury – fluid resuscitation is calculated from the *time of burn*, not the time the patient arrives at hospital.

As with all injured persons, examination begins with the ABCDE sequence. Deep burns to the face, neck or anterior chest cause laryngeal oedema, which increases dramatically with fluid resuscitation. Inhalation of smoke, hot gases or chemical products contributes to this swelling. However, critical laryngeal oedema can occur in *any* deep burn to this important area. The patient should be checked for burnt hair in the nostrils or soot in the nose, mouth or sputum.

The airway must be secured and kept open, preferably by tracheostomy. This should be performed *before* the airway is compromised since it will be very difficult once oedema has closed the airway and cutting through oedematous tissues can be bloody and complicated.

Please note:

Because of this oedema even the tracheostomy hole will disappear in the depths of the tissues. Rather than an ordinary tracheostomy tube, an endotracheal tube should be inserted in the tracheal opening.

True smoke inhalation induces carbon monoxide poisoning and a chemical pneumonitis related to the inhalation of hot toxic gasses. Carbon monoxide poisoning should be assumed in any person found unconscious at the scene of a fire, and is treated by administering as high a concentration of oxygen as possible for 6 hours.

Increased fluid requirements early in resuscitation suggest severe inhalation injury, which may not appear on a chest X-ray until the second or third day. Inhalation injuries increase the 24-hour fluid requirements by 1 – 2 ml per kg of body weight per percentage of TBSA (1 – 2ml/kg/%), about 50 % increase. Hypoxia and hypercapnoea despite maximal oxygen supplementation or ventilation are ominous signs. Serious cases of smoke inhalation will not usually survive without mechanical ventilation. This can be difficult to organize in a setting offering limited resources.

The sequestration of large amounts of fluid and plasma proteins in the extravascular space produces hypovolaemic shock.

Please note:

Only second and third degree burns should be included in the estimation of the percentage of TBSA. Therefore, the patient must be completely exposed and the extent and depth of the burn carefully estimated using the "Rule of Nines". The patient should be weighed and the "homunculus" diagram used to assess and mark the extent of the injury. Particular attention should be paid to circumferential burns which may require escharotomy.

The natural tendency is to *overestimate the extent* of the burn. In controlled series this has been shown to be up to 25 %. A good idea is to first calculate the burnt area and then the area which is not burnt; the total should add up to 100 %. Another natural tendency is to *underestimate the depth* of the burn; periodic re-examination will help to determine this better.

Patients with burns of moderate and major severity should get a Foley catheter placed in the bladder to monitor the hourly urine volume, which is the most important single measurement for adequacy of resuscitation. A nasogastric tube should also be passed and, if there is no acute gastric dilatation, enteral feeds can be started within the first 24 hours. Early feeding through the nasogastric tube and appropriate acid suppression (antacids, H2 blockers) prevent acute haemorrhagic gastritis, which is usually fatal. Where maintaining an i.v. line is difficult, fluid resuscitation may be administered via the nasogastric tube or even orally in smaller burns. This may be particularly useful in small children.

Adequate analgesia (intravenous narcotic) is necessary at all stages of burn management. ICRC practice is to give penicillin for the first five days to prevent invasive haemolytic streptococcal infection; any other infection is treated ad hoc. Tetanus prophylaxis must be instituted as required. Other associated injuries (penetrating wounds, fractures, etc.) must be diagnosed and treated in parallel with the burn wound.

15.3.3 Initial fluid replacement therapy

Given adequate initial resuscitation with crystalloids, the *capillary integrity* is mostly restored within 18 – 24 hours following the burn. At this point, colloid can be given and will stay within the vascular compartment, increasing the plasma volume. Cardiac output will respond to fluid replacement long before blood and plasma volumes return to normal and a first, mild diuresis begins at about 12 hours after fluid therapy. Red cell life is reduced and although replacement of red cell mass is not necessary for the first 48 hours, for major burns blood transfusion will probably be necessary after this time.

Most minor burns presenting less than 15 % TBSA second degree damage do not require formal fluid resuscitation and can be treated with oral fluids, and patients can be managed on an outpatient basis. (Some surgeons prefer to hospitalize even a 3 % full-thickness burn, especially to the face, hands or feet.) Moderate and major burns require hospitalization and intravenous fluids. The ICRC uses the modified Brooke/ Parkland formula[1] for fluid resuscitation.

There are three phases in resuscitation therapy.

1. First 24 hours following the infliction of the burn wound (not from the start of treatment).
2. Second 24 hours.
3. After 48 hours.

The first 24 hours

Ringer's lactate is the fluid of choice. Administration should be divided into three 8-hour periods.

Ringer's lactate solution 2 – 4 ml/kg/% burn = total volume for first 24 hours:

- first eight hours: one half the volume
- second eight hours: one quarter the volume
- third eight hours: one quarter the volume

Urine output should be 0.5 ml/kg/hour.

Fluid administration should start at the low end of the formula (2 ml/kg/%; for children start at 3 ml) and the hourly urine output monitored. There is a *non-linear* relationship in burn patients between the infusion rate and the urine flow. Normal urine flow is described as 0.5 – 1.5 ml/kg/hour. The *lower* limit should be aimed for, and any increase above this should call for a lowering of the infusion rate to avoid over-resuscitation.

Apparently, there is a natural "oedema-limiting mechanism" related to the amount of fluid that is easily mobilized from plasma into the burnt areas. Giving more fluid to maintain tissue perfusion and urine flow above this level may close down this mechanism and lead to more tissue fluid sequestration (tissue is a "fluid addict"[2]); giving even more intravenous fluid does not lead to improvement.

If the urinary output is low and does not respond to increased infusion in the second eight-hour period, the third eight-hour period Ringer's lactate should be replaced with colloid, plasma or 5 % albumin if available. If the vascular compartment is well-filled, however, but the patient is not making urine, the kidneys are likely to be failing and may respond to furosemide or mannitol.

1 Brooke formula: 2 ml/kg/% TBSA for first 24 hours; Parkland formula: 4 ml/kg/% TBSA for first 24 hours.

2 Sjöberg F. (Department of Anaesthesia and Intensive Care, University Hospital, Linköping, Sweden.) Monitoring of Resuscitation Adequacy. Proceedings of the International Society for Burn Injuries, 42nd World Congress of the International Society of Surgery; 2007 Aug 26 – 30; Montreal.

Other clinical vital signs must be closely watched, particularly the peripheral circulation, the patient's general condition, such as consciousness, restlessness, nausea or vomiting, and the haematocrit.

Constant monitoring is required and, after 12 hours, the patient's condition and fluid requirements should be reviewed and recalculated.

Please note:

Not only is there a natural tendency to overestimate the TBSA but, in much contemporary clinical practice, it has been discovered that over-resuscitation has become more common and more of a problem than under-resuscitation. The traditional fear of renal failure has led many clinicians to administer an excess of fluids. This has been called "fluid creep"[3] or "resuscitation morbidity" and manifests itself most commonly as pulmonary oedema and, later, abdominal compartment syndrome, delayed wound healing, increased susceptibility to infection, and multiple organ failure. It should be remembered that the fluid resuscitation formula is only a guideline, and actual amounts of i.v. fluids should be adapted accordingly.

The second 24 hours

During this phase, expansion of plasma volume can be achieved.

Plasma should be administered: 0.3 – 0.5 ml/kg/%/day.

If plasma is safe and *available*, theoretically it is best given at this stage. Albumin 5 % run at 50 ml/hour for 2 – 3 days is an expensive alternative, but the evidence to support either strategy is very slim. Otherwise, and in ICRC practice, Ringer's lactate should continue to be given at the rate of one-quarter of the first day's volume; intravenous fluid should be adjusted to urine output and enteral feeding increased as tolerated. This will provide not only essential nutrition but free water to cover evaporative losses from the burn wound.

15.3.4 Monitoring resuscitation

Clinical evaluation is particularly important in the absence of sophisticated means and laboratory measures. A clear sensorium, good tissue perfusion, good pulse, and an adequate urinary output are all signs of good progress. Less and less fluid is needed at the end of resuscitation in order to maintain urine volume. A flow sheet should be used to monitor vital signs and fluid input/output. The patient should be serially weighed if possible.

15.3.5 After 48 hours

Mobilization of burn wound oedema causes an expansion of the blood volume and results in massive diuresis, high cardiac output, tachycardia and anaemia. The better the fluid management in the first period – avoiding over-resuscitation – the less pronounced these clinical signs and the more stable the patient will be.

If available, plasma or albumin should be given to maintain serum albumin of 20 g/l and blood to maintain haemoglobin greater than 70 g/l. Deep burns cause greater anaemia. Fresh whole blood is best. Potassium, calcium, magnesium and phosphate are generally excreted in large amounts at this time, and should be supplemented where possible.

3 Pruit BA Jr. Fluid and electrolyte replacement in the burned patient. *Surg Clin N Am* 1978; **48**: 1291 – 1312.

15.4 Burns presenting late

Often, patients with severe burns present late. Those presenting after a delay, but within the first 24 hours, should receive fluid resuscitation and an attempt should be made to infuse most of the calculated volume by the end of 24 hours from the time of burning.

E. Dykes / ICRC

Figure 15.5
Old flame burn to the leg.

Patients presenting after 24 hours will probably need some fluid but the amount should be guided primarily by clinical assessment of hydration and renal function. Patients who have survived the first 72 hours without renal failure have compensated the losses themselves (usually with oral fluids); they may require some rehydration but infection of the burn wound is the main concern. For those presenting weeks after the burn, infection is compounded by poor nutritional status, anaemia, and hypoproteinaemia. In these cases steps should be taken to control infection (debride grossly infected necrotic tissue) and to improve nutrition before any definitive surgery is undertaken. Early creation of a feeding gastrostomy should be considered.

15.5 Nutrition

Catabolism is particularly elevated in burn patients, especially those losing large amounts of protein through open wounds, and healing will require an important increase in caloric and protein intake for a lengthy period. Early enteral feeding is very important in maintaining gut function (reducing gastroparesis), and preventing complications. Major burn patients may require more than twice their normal protein and calorie intake until their wounds have closed. Nutritional requirements may be easily calculated for the individual patient (see Annex 15. A: Nutrition in major burns).

Blenderized enteral feeding solutions can readily be made from commonly available local foods and can be delivered by naso-gastric tube, gastrostomy or jejeunostomy. Especially when dealing with burn wounds that are weeks to months old, the patient's nutritional status should be evaluated and improved before any attempt is made at skin grafting or both the graft and the donor site risk delayed or failed healing.

15.6 Care of the burn wound

After successful resuscitation of the patient has been achieved, the next greatest threats to life that must be addressed are the burn wound itself and the complications of sepsis.

The aim of treatment is to achieve healing of the burn by the following measures.

1. Control of bacterial colonization by removal of all dead tissue.
2. Prevention of accumulation of purulent fluid and debris.
3. Prevention of secondary bacterial contamination.
4. Maintenance of an environment which promotes wound healing.
5. Avoidance of techniques or treatment which will damage the healing burn wound.

Most of the morbidity and mortality associated with major burns are the result of infection. All the methods for cleaning the wound, excising dead tissue and treating the burn are directed towards the control of "burn wound sepsis".

Untreated, eschar desiccates and sloughs off by means of bacterial enzymatic invasion of the plane between viable and non-viable tissue. Full-thickness burns will require skin grafting to heal properly; otherwise, healing will be by fibrous tissue contraction alone with chronic open wounds and disabling scar contracture.

In partial-thickness burns, viable areas of dermis underlie the necrotic tissue and if sufficient epidermal cells remain at the base of skin appendages, such as sweat glands and hair follicles, re-epithelialization will gradually take place if given the right conditions.

Infection will transform a partial-thickness burn into a complete-thickness one. Because of the complete or partial ischaemia associated with burn wounds, systemic antibiotics may not reach the site of bacterial colonization. Local treatment, both mechanical and antimicrobial, are the basic elements of wound care.

15.6.1 Initial wound management

Initial wound toilet should be undertaken after resuscitation has begun and proceed in parallel. Once the patient's condition has stabilized, attention can be turned to more *definitive* measures.

Any constricting agents (rings, wrist-watch, jewellery, etc.) should have been removed on admission. The patient should be sedated and the burn wound gently washed with soap and water. Freely running, clean water, without excess pressure but a regular flow and at a temperature comfortable for the patient, is best. This will cool the burn, decrease pain, and remove superficial debris and adherent clothing.

Small intact blisters may be left *undisturbed* but large, bloody, or pus-filled blisters, and those which interfere with joint movement, should be *unroofed and debrided*. Large burn wounds may be more easily cleaned by placing the patient in a shower. Baths should be avoided because of the attending logistical difficulties and the danger of cross-infection under field conditions. The routine daily immersion of burn patients in filthy tubs of cold water is to be condemned.

Particular attention should be paid to circumferential deep burns. During the first 48 hours the increasing oedema of the tissues and the thick unyielding burn eschar may result in a tourniquet effect. Circumferential burns of the chest will restrict breathing and those of the limbs will cause peripheral ischaemia which can lead to amputation. This disaster is easily averted.

Escharotomy is the process of incising the eschar down to the subcutaneous fat to allow relief of the constriction.

Escharotomy should be performed with a sharp knife or electric diathermy through the burnt skin into the subcutaneous fat.

Escharotomy incisions are placed in the mid-lateral and mid-medial lines of the affected extremity and should extend up to, but never include, unburnt skin. A "T-cut" at each end of the escharotomy allows tissue to expand without causing an acute constriction at the end of the incision. On the hands, the mid-lateral incisions coming down each side of the forearm should be swung onto the dorsum, and branched to make a single escharotomy incision down the mid-lateral line of each finger on its less-used side. (i.e. ulnar side of thumb, index and ring finger; radial side of middle and little finger).

A thoracic escharotomy begins at the midclavicular lines, continues along the anterior axillary folds down to the costal margin, and across the epigastrium to the xyphoid process.

Please note:

Few patients who truly need a thoracic escharotomy will survive without mechanical ventilation.

ICRC

Figure 15.6

Burn to the face with intact blisters.

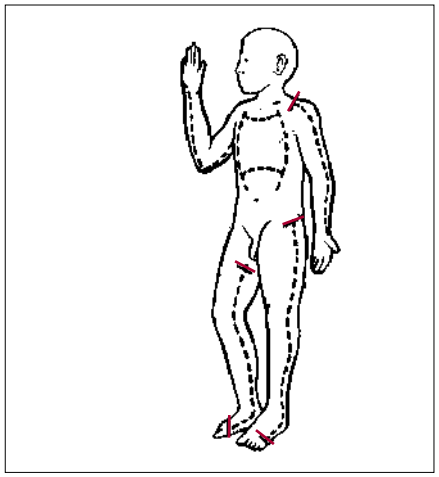

M. King / Primary Surgery Volume 2

Figure 15.7.1

Sites for escharotomy incisions.

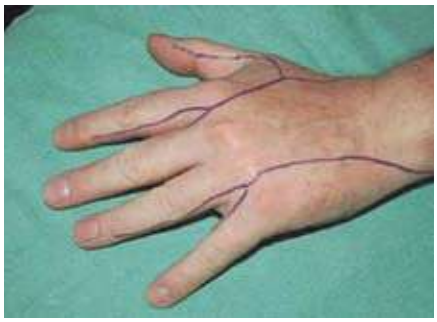

M. Beveridge / ICRC

Figure 15.7.2

Placing for escharotomy incisions on the hand.

M. Baldan / ICRC

Figure 15.7.3

Escharotomy incision of the arm.

Though full-thickness burns are usually anaesthetic, the escharotomy should be done with some form of anaesthesia as the edges can be very painful and the cut goes into the subcutaneous fat. Ketamine is ideal.

Particular attention should be paid to burns with underlying fractures or burns extending deep into the fascia; these can create a compartment syndrome. Formal fasciotomy, with excision of the aponeurosis, may be necessary in addition to escharotomy.

Severe perineal burns may necessitate faecal diversion.

15.6.2 Local care

Care of burn wounds involves an extraordinary amount of nursing care. The method used depends on the depth, extent, and location of the burn. ICRC surgical teams use occlusive dressings – and its plastic bag modification – and open therapy; both with a topical antibacterial agent.

K. Yagi / U. of Khartoum

Figure 15.8
Paraffin gauze and occlusive dressings.

K. Yagi / U. of Khartoum

Figure 15.9
Plastic bag modification of occlusive dressing.

Many antibacterial compounds can be used. Silver sulphadiazine (Flamazine®) and silver nitrate solution are unique in that they can penetrate the wound surface and reach the bacteria beneath it; therefore, they are best for infected full-thickness burns. Antibiotic ointment (polymyxin/bacitracin or similar) and paraffin gauze is another excellent dressing, particularly for partial-thickness burns with no eschar. Honey and ghee dressings can be used where expensive compounds such as silver sulfadiazine are scarce. Equal parts of honey and either ghee (clarified butter), or oil, are mixed together and poured over open sheets of gauze in a pan. The honey provides a hyperosmotic environment that discourages bacterial growth and the ghee or oil keeps the gauze from adhering. Silver nitrate solution (0.5 %) is effective but oxidizes easily, staining everything it contacts black. Other agents used around the world include gentian violet, which desiccates the burn, tea which tans the burn like leather, papaya and banana leaf, boiled potato skins, fermented fish sauce, and Amazonian frog skin. In some instances local methods may be preferable to imported products.

Occlusive dressings
Bulky sterile dressings relieve pain and are comfortable for the patient, and protect the burn wound from infection. They absorb serum and exudate and foster a moist healing environment by keeping the injured part immobile and warm; they contain antibiotics able to penetrate the dead eschar (i.e. silver sulfadiazine ointment).

The dressings have three components: an inner layer of a liberal application of silver sulphadiazine covered by fine-mesh gauze or paraffin gauze; a middle layer of large gauze-wrapped absorbent cotton pads, which absorb the exudate and protect the wound; and an outer layer of bandages that secures the dressing.

If the dressing becomes soaked through, the outer layers should be changed or bacterial contamination will occur by capillarity. Dressings should be changed every day or two under adequate analgesia, and the old silver sulphadiazine washed off in the shower. The wounds should be inspected and gently cleaned at dressing time, and morsels of dead eschar removed with scissors and forceps.

Occlusive dressings are best for small burn areas, especially on the limbs, or when hygienic conditions are less than optimal.

The plastic bag or surgical glove method
This method is used for burns to the hands and feet. After cleaning the burn and applying silver sulphadiazine directly to the wound with a spatula or the gloved hand, a plastic bag is used as a glove or a sock, tied around the wrist or ankle. It must not be too tight. The limb should be kept elevated to reduce swelling. The burnt area will be kept moist and movement of the joints, both passive and active, is encouraged. A surgical glove may also be used instead of a plastic bag; it allows for more mobility during physiotherapy.

Open therapy

Where resources are inadequate to allow for occlusive dressings, this is the method of choice – as well as being the standard method for burns of the face and perineum – but requires clean and isolated surroundings, and the ambient temperature must be warm; hypothermia is to be avoided at all costs.

The patient is placed on clean sheets and the burnt area left completely exposed. Silver sulphadiazine ointment is applied to the burn wound liberally with a sterile gloved hand and repeated twice daily or as required. If the room is cold, the patient may be covered with a clean sheet and blanket draped over a frame to prevent contact with the wound. The entire bed should be kept under overhanging mosquito netting.

The *advantages* of this method of treatment include simplicity of examination of the burn wound and ease of nursing care. It also allows early mobilization by physiotherapy.

The *disadvantages* are pain, odour, and desiccation of the wound, delayed eschar separation and hypothermia. Frequent showering is required to wash away exudate and fragments of softened eschar. The bed linen must be changed regularly; it is readily dirtied by the burn exudate. Local customs and religious practice may limit the use of this "open" method.

Facial burns are best treated open with frequent gentle cleansing and application of moist warm saline gauze soaks, interspersed with application of topical antibiotic ointment (e.g. polymyxin/bacitracin). Beards and facial hair growing through the burn should be shaved at least every two days to prevent accumulation of exudate which may harbour infection. Antibiotic eye ointment should be applied frequently to the conjunctiva if the lids are burnt and retracted, to prevent keratitis and corneal ulceration. The lids should only very rarely be sutured together (blepharoplasty) as the sutures almost inevitably pull out causing yet more damage to both lid and globe.

ICRC

Figure 15.10

Open burn therapy with frame.

ICRC

Figure 15.11

Healing burn to the face.

15.7 Closure of the burn wound

The preparation of the burn wound and its subsequent closure are the two main steps in the surgical management of burns. The type of surgery required depends upon the skill and training of the surgeon, the specific burn injury, and the facilities available to support the treatment, especially the availability of blood for transfusion. Like all other injuries treated under difficult circumstances, great judgement is required to select the technical procedure that is appropriate.

Partial-thickness burns will generate new epithelium on their own if infection is prevented. The wounds should be inspected carefully as they progress. In superficial partial-thickness burns (and donor skin graft sites) epithelial cells grow out from around tiny skin appendages giving a typical "leopard spot" appearance in those with pigmented skin and, on close inspection, a slightly dull, silvery coat of epithelial cells may be seen growing atop the dermis (Figure 15.12.1). Tiny white pearls of epidermis herald re-growth and healing; areas of raspberry-red granulating dermis or fat do not have sufficient epidermal cells to heal (Figure 15.12.2). A "nicely granulating" burn wound is not a good thing unless there is a plan to graft the wound.

ICRC

Figure 15.12.1

Partial-thickness burn well on the way to healing.

E. Dykes / ICRC

Figure 15.12.2

Granulating burn: no healing. The pale colour of the granulations indicates that the patient is anaemic.

For full-thickness burns, the eschar may be removed completely in a single procedure or in a number of stages. The aim of treatment is to prepare the wound for eventual closure and to prevent colonization by bacteria and fungi.

15.7.1 Mechanical cleansing and debridement

Thorough cleaning and removal of debris and fragments of eschar are performed at each dressing session. Gentle washing and sharp debridement of dead skin fragments should be combined with thorough irrigation with water. The surface can then be cleansed with a mild disinfectant (dilute hypochlorite solution, detergent soap) and thoroughly washed again with water. Silver sulphadiazine is reapplied. Where resources are scarce, the cycle of dressing and plucking can be continued until the wound is completely free of eschar. This gives partial thickness burns the optimum chance to re-epithelialize and minimizes the area that will require grafting. Such patients need large amounts of nursing time, dressing material, and will probably require blood transfusion even without surgery.

15.7.2 Surgery

The most difficult decisions in burn care involve the timing and extent of surgery. Judicious staging of burn surgery is highly important and, where possible, should take into account the patient's livelihood. Donor and graft areas should be matched in advance and the process broken down into manageable parts. Attention should be paid to how the body parts will be positioned for the operation: if the plan is to graft the arm, it should be seen to before the hand as the latter will be needed to hold up the arm for surgery.

The hands, feet, and joint surfaces are considered priority areas in order to re-establish function; early grafting of these parts must be balanced against the metabolic advantages of closing larger areas on the limbs and torso. At least two weeks should pass before grafting of facial burns is considered, as even quite deep burns may go on to heal on their own. Eyelids are the first priority here.

Tangential excision

The removal of the full thickness of the burn wound is performed in a single session. Early tangential excision and immediate grafting of burns reduces mortality, morbidity, suffering and hospital stay while improving functional and cosmetic results, but it requires considerable resources and is *impracticable* for areas larger than 10 % TBSA outside specialized burns centres. It is bloody surgery.

Most field surgeons should be *conservative* with regard to this technique. It is, however, recommended in ICRC practice for small burnt areas, especially on the face, hands and feet, and over joint surfaces.

In tangential excision of the eschar, the superficial layers of the burnt tissue are progressively shaved away by knife, dermatome or cutting diathermy until viable tissue is reached. Traditionally, this is determined by the appearance of copious punctuate bleeding. It is accompanied by significant blood loss and is the most important limiting factor of this technique.

Blood loss can be reduced by Esmarch bandage exsanguination of the limb and a tourniquet, and subcutaneous infiltration of a dilute adrenaline solution (1:500,000); in the face, lidocaine with adrenaline can be used. The subcutaneous infiltration of some type of fluid (saline, dilute adrenaline solution, or local anaesthetic), causes a local tumescence that facilitates the excision. With adrenaline solution, as dead burn eschar is sliced away, the surgeon should look for a level with visible pearly white dermis, or glistening yellow fat, and no capillary thrombosis. After excision, larger bleeders should be cauterized and the wound wrapped for ten minutes with gauze soaked in adrenaline solution. The wrapping should be removed, and the process repeated until there is no active bleeding, before the skin graft is applied.

When tangential excision is successful, only dead tissue is removed. It can be difficult to judge how much to take away in order to leave a viable layer that will take an immediate skin graft.

Face

The skin of the face, particularly the beard areas on males, is very thick and well populated with deep epidermal cells that will provide for re-epithelialization if given time. If there is any doubt about the depth of burn on the face it is best to wait two weeks before tangential excision.

As mentioned, face burns are treated by the open technique with warm wet gauze soaks, gentle cleaning and shaving every second day, and the application of a local antibiotic ointment. Severe face burns will require scraping and cleaning under general anaesthesia to properly evaluate which areas are healing and which will eventually need grafting. Gauze soaked in an adrenaline-saline solution (1:33,000) should be used and pressure applied to control bleeding. When clean, a thin coat of antibiotic ointment should be applied and the previous dressing routine resumed while waiting to decide whether to excise and graft.

Excision of a small deep burn to the face can be performed under local lidocaine with adrenaline anaesthesia; larger areas require general anaesthesia, but simultaneous subcutaneous infiltration with a dilute adrenaline solution will render the excision easier and less bloody.

Hands, feet, and joint surfaces

Tangential excision of the hands, feet and joint surfaces can be performed at three days onwards, once the patient is well resuscitated.

Many severe hand burns will benefit from early escharotomy since deep partial-thickness or full-thickness burns heal with extremely disabling contractures; operating these early should be considered and good, thick skin for the grafts saved. Typically people clench their fists when they sustain a burn so the palmar skin extending up to the mid-lateral lines of the fingers is usually preserved, or burnt much less deeply than the dorsum, and rarely needs grafting. Therefore most burnt hands and fingers will only require grafting of the dorsum. If the escharotomy was performed accurately along the edges of the full-thickness burns along the mid-lateral lines of the fingers, this will mark the extent of excision necessary.

Proper preparation is the key to good excision. The areas to be excised should be carefully marked with ink or gentian violet. The hand and forearm should be exsanguinated by five-minute elevation and application of a rubber Esmarch bandage, beginning with the hand and progressing proximally; a pneumatic tourniquet is then applied. (Correct use of a tourniquet is painful; the operation should be performed under general anaesthesia.) Normal saline or a weak adrenaline solution should be infiltrated subcutaneously over the dorsum of the hand. The edges of the areas to be excised should be scored with a #15 blade.

Tangential excision should be performed using a small dermatome or scalpel, preserving viable dermis where possible and being very careful not to damage tendon sheaths. The hand should be wrapped in adrenaline-soaked gauze and the tourniquet briefly released. The tourniquet should then be re-inflated for ten minutes to allow natural haemostasis, then removed. The hand should be unwrapped and diathermy used to cauterize remaining bleeders. Wrapping in adrenaline gauze and cauterizing of bleeders may need to be repeated several times to ensure perfect haemostasis prior to application of the skin grafts. Sheets of relatively thick split-thickness skin should be used. Grafts should be carefully tailored over the dorsum of the hand and fingers, and sutured in place. Each finger should be covered in paraffin gauze dressing, then wrapped independently in gauze, taking care to leave the tips of the fingers exposed in order to assess perfusion.

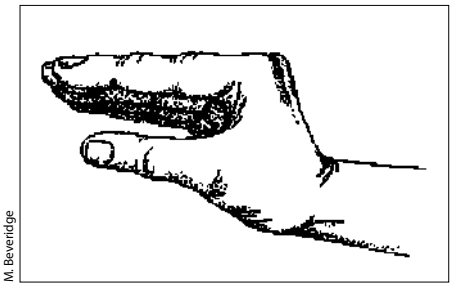

M. Beveridge

Figure 15.13

The "safe position" for splinting a hand.

Finally, the hand should be splinted in the "safe position" (Figure 15.13) with the wrist dorsiflexed 30°, the metacarpal-phalangeal joint flexed as close to 90° as possible, the fingers splayed apart and the inter-phalangeal joints straight. The initial dressing should be left for 5 – 7 days and then unwrapped carefully. The burn should then be re-dressed daily with paraffin gauze and re-splinted. Physiotherapy and mobilization should begin as soon as the grafts are solid. A hand should never be immobilized longer than ten days. K-wires placed through the fingers may help in bad cases. The same general principles apply to the feet and over joint surfaces.

Early tangential excision gives the speediest recovery and best functional and cosmetic results when it is practicable, but should *never* be undertaken for large areas unless the resulting blood loss can be properly managed.

Skin grafting

The hands, feet, and joint surfaces are priority areas for skin grafting, with or without tangential excision. The anterior chest and neck take priority over the belly and buttocks. The back has very thick skin, and so burns to it may be observed for some time while waiting to see if they will heal on their own.

E. Winiger / ICRC

Figure 15.14.1

Meshed skin graft: recipient site.

Allowing burns to granulate and demarcate under dressings (2 – 6 weeks) is prudent practice where resources are scarce. This means accepting inevitable protein loss from open wounds, possible infection, delayed healing, and chronic anaemia; supplementary measures must therefore be taken to reduce these effects.

To prepare for grafting, the jelly-like granulation tissue must be scraped away with the back of a scalpel handle before the skin graft is carefully secured and dressed. The advantage of delayed grafting is that often a much smaller area need eventually be grafted.

E. Winiger / ICRC

Figure 15.14.2

SSG after grafting.

Grafting on burns is time-consuming and adequate time should be allocated for these operations. The staging of the surgery should be carefully planned and just one limb or body area operated at a time. In general, the thinner the graft, the better its chances of "take", and the thicker the graft, the better the functional and cosmetic result (see Chapter 11 for details on skin grafting). A limb or digit that has to be amputated should be regarded as a prime source of donor skin. Harvesting of grafts in children whose skin may be extremely thin should be performed with *great care*. If there is no solid layer of dermis left behind, the donor site will not heal. Except on the face, hands and feet, skin grafts should be meshed to allow serum to drain through rather than lift the graft off the bed of the recipient site.

Old wounds or sites where previous grafts have failed are notorious for difficulty in achieving an adequate graft take. Good nutrition and meticulous preparation of the surface are the keys to success. Early excision of grossly dead and infected tissue, topical and systemic antimicrobial therapy, and aggressive nutritional supplementation should precede any attempt at skin grafting. The best course may be to graft critical areas while leaving some of the larger, less functional areas to granulate.

The surface of the debrided burn wound often has a superficial slime of exudate and bacterial contamination. Dressings with a supersaturated saline solution (add salt to normal saline until it no longer dissolves), changed frequently for a couple of days, will produce a clean, bright red granulating surface ready for grafting.

The occlusive dressing which is applied after skin grafting plays a great part in the survival of the graft. It must be put on with great care. The dressing must hold the graft closely applied to the recipient site for the first few days for capillaries to grow in. Any blood or serum that floats the graft away from its bed will cause it to fail.

The life of a patient with a severe burn is in danger until the dead tissue has been excised and the defect closed with a healthy skin graft.

15.8 Scar management

Burns always result in scars and the burn graft itself is best thought of as a scar. One of the most disastrous consequences of burns is the severe scar contracture that may render life horrendous for the victim afterwards. Management of the burn scar begins before grafting, during local care of the burn wound. Rigorous splinting – using plaster of Paris slabs – and stretching routines should be implemented to prevent contracture of major joints: the limb should be splinted *against* the force of contracture. If a joint is equally burnt on both sides, it should be splinted in extension. Airplane-strut type splints should be used for burns to the axilla. Patients must be given adequate analgesia for the daily passive stretching exercises.

Much, if not most of the functional benefit of burn grafting depends on assiduous splinting and stretching of the tissues afterwards to manage the process of burn scar contraction. This can be particularly active in children – a perfect operation may result in recurrence of a dreadful scar if the tissues are not managed properly over the next 6 – 12 months.

All burn grafts crossing joints should be splinted at operation with plaster slabs. Later, when the grafts have taken, a plaster slab covered in tube bandage and fitted correctly makes an excellent, reusable splint that can be worn at night and removed for therapy during the day. Dedicated staff and adequate analgesia are essential to active and passive stretching of burn scars; if this hurts too much the patient will simply not comply. Burns crossing joints should be splinted and stretched even if they have not been grafted, as this will reduce the degree of contraction as the scar heals.

Pressure garments are important to optimum scar management and their use results in much softer and more pliable scars. Where they are not available, elastic bandages and a variety of tight-fitting stretchy commercial clothing may help. Water-based skin cream will help with itching as will antihistamines. Burn scar massage is also helpful where traditional practitioners are present.

ICRC

Figures 15.15.1 and 15.15.2

Post-burn scar contractures.

ICRC

15.9 Electrical burns

As mentioned, there are two distinct categories. With flash burns, the patient typically presents with fairly deep burns to the face and one or both hands and forearms. Treatment is as with regular thermal injuries.

High voltage (>1,000 volts) electrical conduction injuries have small cutaneous entry and exit wounds, which extend deep into the muscles causing myonecrosis. The rhabdomyolysis has a *systemic* effect, with myoglobinaemia and myoglobinuria leading to acute tubular necrosis; and a *local* effect, i.e. compartment syndrome.

The patient should receive Ringers' lactate, with 50 mEq sodium bicarbonate per litre, in sufficient volume to maintain urine output at 0.5 ml/kg. If the urine is dark or bloody, or urine output ceases, the vascular space should be well filled and a bolus of 20 % mannitol given (1g/kg); furosemide may be added as well.

Any suspicious compartments should be released promptly by generous full-length fasciotomy, including carpal tunnel release in the forearm. Dead muscle should be debrided conservatively, and numerous returns to the operating theatre may be needed (serial debridement).

Numerous complications attend electrical conduction injury including arrhythmias (ECG monitoring is strongly recommended), cervical fractures associated with the jolt of electricity, intestinal perforation and a variety of unusual neurological sequelae.

15.10 Chemical burns

A number of chemical substances cause specific burns. The presence of any of them on an injured person poses a danger to first aiders, hospital staff, and other patients. Careful removal of contaminated clothing and proper decontamination measures must be undertaken – of the patient and of any equipment used – and specific protocols followed to protect the medical personnel treating the patient.

The wounding chemical agent constitutes a danger to first aiders, hospital staff, and other patients. Proper protective measures must be taken.

15.10.1 Acid and alkali burns

In general, acids cause coagulative necrosis of the skin, and strong alkalis liquefactive necrosis and burrowing of the chemical deep into the tissues. "Acid violence" involving the throwing of concentrated sulphuric acid into the face is an increasingly common phenomenon and the chemical is readily available wherever there is a motor vehicle garage. The acid burn should be washed with very large volumes of water and the eyes thoroughly irrigated. Strong alkalis may come in dry form (NaOH crystals); any remaining material must be brushed off the patient before washing with water.

After this decontamination, the treatment of chemical burns follows the same sequence as the treatment of thermal burns. Acid attacks usually involve the face and typically cause extremely disfiguring injuries which are very difficult to reconstruct.

15.10.2 Phosphorus burns

Some modern anti-personnel weapons contain white phosphorus. This element *ignites on contact with air*, and fragments of phosphorus will be scattered throughout any wounds; it is lipid soluble and sticks to the subcutaneous fat. The burn is deep and painful and the phosphorus continues to burn as long

as it is in contact with oxygen or until all the phosphorus is consumed, all the way down to the bone. Local treatment is more urgent than with conventional burns because of the aggressive nature of phosphorus. Much of the injury in an individual patient, however, results from the ignition of clothing, which causes a conventional burn.

Contaminated clothing must be removed immediately, care being taken not to contaminate the staff attending to the casualty. Visible, smoking particles can be removed with a spatula or knife, and should be placed in a basin of water to exclude them from the air. Phosphorus burn wounds must then be isolated from oxygen by being kept wet through liberal soaking with water, by covering with wet dressings, or by placing the injured part in a basin of water. On no account must they be allowed to dry out.

When surgical treatment is available, the idea is to identify and remove the remaining phosphorus particles. The wet wound can be irrigated with a neutralizing agent. A freshly prepared solution of 1% copper sulphate combines with the phosphorus to form black copper sulphide, which impedes violent oxidation and identifies the particles. The black particles can then be removed with forceps and placed in a basin of water. The solution must be very dilute, the palest blue colour, since its absorption can cause haemolysis and acute renal failure. If used, the copper sulphate solution must be washed away immediately. Or, if copper sulphate solution is not available, the operating theatre lights may be put out; any remaining particles will glow with phosphorescence in the dark and can be carefully picked out with forceps and placed in a basin of water.

Care must be taken not to allow the wound and the phosphorus to dry out and re-ignite in theatre; appropriate, non-flammable anaesthetic agents should be used. In all cases, the wound should then be excised and dressed as usual.

Phosphorus may provoke hypocalcaemia and hyperphosphataemia; intravenous calcium should be given. Absorbed phosphorus can be toxic to many organs:
- central nervous system – delirium, psychosis, convulsions, coma;
- gastro-intestinal tract – abdominal colics, melaena;
- liver – hepatomegaly, jaundice;
- kidneys – proteinuria, acute tubular necrosis;
- blood – thrombocytopenia, hypoprothrombinaemia;
- myocardium – ventricular arrhythmia, myocarditis.

15.10.3 Napalm injuries

Napalm is jellied petrol, an intensely flammable agent that clings to the clothing and skin while still burning and causes serious, deep and extensive burns. Its incomplete combustion of the oxygen in the air around the victim provokes an acute rise in carbon monoxide that can lead to a loss of consciousness and even death. The intense heat and benzene fumes easily cause inhalation burns.

Napalm burns are invariably *full thickness*, with coagulation of muscles and other deep tissues. Nephrotoxicity is a serious complication of the rhabdomyolysis, and the mortality may be high in proportion to total body surface area involved. A full-thickness burn of only 10% of the body surface area may result in renal failure. The patient should be kept well hydrated and in alkalosis; mannitol may be necessary to protect renal function.

First-aid treatment includes extinguishing the burning napalm by smothering it, i.e. excluding the oxygen. Unlike phosphorus, napalm does not then reignite in contact with the air. Surgical treatment involves removal of the napalm with a stick, spatula or knife. The wound is then excised deep to any remaining contaminant, care being taken to avoid contact ("no-touch" technique), and dressed as usual.

15.10.4 Magnesium

Flares released from aircraft to avoid heat-seeking missiles use magnesium, which gives off great heat when lighted. These may reach the ground and cause fires and injuries. The intense heat causes a full-thickness burn. The wound should be excised deep to the contaminant using the "no-touch" technique. Some reports from ICRC surgical teams (Kabul, Afghanistan) have made mention of toxic side effects due to absorption of magnesium, similar to phosphorus; this has not been confirmed.

15.10.5 Vesicant chemical weapons

Unlike other agents causing specific burns, chemical weapons have been banned by international treaties.[4] However, some States retain stockpiles, which may either be used militarily or released into the air if the storage facilities are bombed. Certain chemicals have a potential dual function: they can be used in weapons and are widely employed for civilian purposes (the disinfection of public water supplies in the case of chlorine).

Traditional chemical weapons are either neurotoxic or vesicant (blistering); the latter cause burns to the skin and inhalation injury. Vesicant agents (mustard gas, lewisite, phosgene) cause skin burns similar to flame burns.

The patient's clothing must be removed and properly destroyed. The patient is then decontaminated with abundant soap and water. Care must be taken not to contaminate hospital personnel, equipment, and other patients with the chemical agent. Correct decontamination protocols include the use of protective clothing and equipment (mask, gloves, boots, etc.) by first aiders or hospital staff.

Once decontamination has been performed, the wounds are dealt with in the traditional manner; however, a "no-touch" technique should be used during wound debridement and removed tissues disposed of with care. The patient's respiratory function must be closely monitored. Inhalation of chemical fumes will burn the respiratory mucosa with the development of acute respiratory distress syndrome (ARDS). Assisted ventilation may be required.

4 1925 Geneva Protocol for the prohibition of the use in war of asphyxiating, poisonous or other gases and of bacteriological methods of warfare and the 1993 Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction.

ANNEX 15. A Nutrition in major burns: calculating nutritional requirements

Calorie requirements = Basal energy expenditure x stress factor x activity factor
The *basal energy expenditure* is calculated as follows:

[66 + (14 x weight in kg) + (5 x height in cm) – (6.8 x age in years)]

The *stress factor* for minor procedures is 1.3; for skeletal trauma, 1.35; for major sepsis 1.6; and for major burns 2.1.

The *activity factor* is 1.2 for those in bed and 1.3 for those who are mobilizing. Women require about 4 % less than men for equal body size and age.

Example
For a 25-year-old male weighing 60 kg and 170 cm tall, in bed with a major burn, the calorie requirements =

[66 + (14 x 60) + (5 x 170) – (6.8 x 25)] x 2.1 x 1.2 = 3,997 kcal/day

Protein, glucose and fat requirements
Daily *protein* requirement for acute burns is 2 g/kg in adults and 3 g/kg in children. Protein provides about 4 kcal/g (120 g and 480 kcal in the example given above).

Daily *glucose* requirement is about 6 g/kg/day in burns.
Glucose provides 4 kcal/g (360 g and 1,440 kcal in the example).

The difference between the calculated energy requirement (3,997 kcal) and that provided by protein and glucose should be made up with fat.

Daily *fat* requirement = 3,997 kcal – 480 – 1,440 = 2,077 Kcal
Each gram of fat provides 9 kcal, therefore, 2,077 ÷ 9 = 231 g of fat

The larger the volume and the higher the fat concentration the more likely the patient is to develop diarrhoea. In an adult with a major burn, 3 litres of feeds per day is a reasonable target; therefore for the patient in the example, a "cocktail" containing 40 g of protein, 120 g of glucose and 80 g of fat per litre should be prepared.

Making a high-energy enteral feeding solution for burn patients

Ingredients	Glucose	Protein	Fat	kcal
Skimmed milk powder 110 g (244 ml)	44 g	40 g		385
Edible oil 80 g (80 ml)			80 g	720
Sugar 50 g (50 ml)	50 g			200
1 Banana (15 mEq potassium)	25 g			110
Add: Salt 3 g Calcium containing antacid 3 tablets Multivitamin tablet 1 daily Ferrous sulphate + folic acid tablets Codeine 30 – 60 mg per litre provides analgesia and reduces diarrhoea Eggs contain 15 g of protein each: beware of salmonellosis from raw eggs Supplement tube feeds with cooked eggs fed by mouth when possible				
Boiled and filtered water to make 1,000 ml of solution			Total 1,415 kcal per litre	

Make a paste of milk powder with a little water; add sugar, salt, crushed tablets and oil. Slowly add more water while mixing well; add mashed banana and mix thoroughly (using a blender if possible). Filter the mix through a gauze compress and refrigerate. Irrigate the feeding tube regularly with water to keep it from blocking. Use within 24 hours.

Chapter 16

LOCAL COLD INJURIES

16. LOCAL COLD INJURIES

16.1 Physiology of thermal regulation 301

16.2 Types of local cold injuries 301
16.2.1 Non-freezing injury 301
16.2.2 Freezing injury 301
16.2.3 Local signs and symptoms 301

16.3 Management 302
16.3.1 First aid and transport 302
16.3.2 Hospital treatment 302
16.3.3 After care 303

16.1 Physiology of thermal regulation

The significance of prevention and treatment of local cold injuries during armed conflict cannot be overemphasized. Although most commonly seen in arctic and subarctic climates, cold injuries can occur whenever the combination of cold, wet, and immobility exists. High altitudes, even in tropical or temperate regions, can experience cold weather. Wind is an aggravating factor in all circumstances.

Normal body temperature is maintained through a balance between heat production and heat loss, and is regulated by a hypothalamic "thermostat". At least 95 % of the heat produced by the metabolism of viscera and muscles is normally lost to the environment by conduction, convection, radiation, and evaporation, largely through the skin and lungs; the head and neck account for 20 – 30 %. The skin primarily dissipates heat by regulating its blood flow, which may vary from 50 ml/min to 7,000 ml/min.

In a cold environment, the *core temperature* (that is, the temperature of the vitally important visceral organs) is conserved by decreasing heat loss through peripheral vasoconstriction, and by increasing the production of heat by involuntary muscle contractions (shivering). If the heat loss exceeds the heat-producing resources of the body, the core temperature begins to fall and *hypothermia* develops. Wetness and wind increase the effect of cold in inducing loss of body heat.

In peripheral tissues subjected to low temperatures, wetness, wind, and contact with a cold surface (metal), local cold injury may result from vasomotor and/or cellular effects, including intracellular ice crystal formation.

O. Litvin / ICRC

Figure 16.1
Patient with "trench foot" affecting both feet.

16.2 Types of local cold injuries

Local cold injuries can occur at temperatures above or below freezing point and are classified as non-freezing or freezing injuries.

16.2.1 Non-freezing injury

This is also known as "immersion foot" or "trench foot", and is seen in prolonged exposure to cold ambient temperatures above freezing, with high humidity and immobilization, as often witnessed during the trench warfare of World War I, hence the name. Prolonged wetness and cooling of the feet, as can occur in a jungle or rice paddy, can also provoke an "immersion foot" injury. Such injuries are diagnosed and treated like other cold injuries, except that the feet should not be immersed in warm or hot water.

16.2.2 Freezing injury

This type of injury, also known as "frostbite", may be superficial, in which only skin and subcutaneous fat are involved, or deep, reaching structures such as muscle.

Frostbite occurs in the extremities and exposed body parts: nose, ears, etc. In the early stages of injury, it is not easy to differentiate between superficial and deep wounds.

16.2.3 Local signs and symptoms

These include:
- paraesthesia;
- numbness and insensitivity to pin prick;
- pallor (wax-white or mottled blue discoloration);
- impaired movement leading to paralysis;
- firmness of the body part;
- oedematous swelling (particularly in non-freezing injuries) followed by blister formation after 24 – 36 hours.

16.3 Management

16.3.1 First aid and transport

When a body part is frozen solid for hours it runs a lesser risk of tissue loss than when it is exposed to a cycle of thawing, refreezing, and thawing. A person can continue to walk on a frozen foot while it remains frozen; but will be unable to walk after thawing because of the pain and swelling. This should be kept in mind *before* pre-hospital treatment is attempted; it might be preferable to transfer the patient to adequate shelter and facilities rather than begin treatment with inadequate means.

The patient should be moved to a sheltered place as soon as possible. Boots and socks should be removed, avoiding trauma to the skin. As with burns, constricting agents such as rings should also be removed.

As some degree of central hypothermia generally coexists with local cold injury, the general body temperature should be raised by hot drinks, blankets, or skin-to-skin contact. The frozen extremity should be *prevented* from thawing before normal core temperature is reached.

Once hypothermia has been corrected, any available form of heat transfer (skin-to-skin contact, foot-in-axilla, hand-to-nose) should be utilized to warm the body part. Rapid rewarming in warm water (40 – 42°C, or just tolerable to elbow immersion) should only be used when it is certain that refreezing can be avoided.

Analgesia is essential, as thawing of frostbite is extremely painful.

Do not:

- rub or massage injured tissue;
- apply ointments or other topical medication;
- break blisters;
- rewarm by exposure to fire, radiant heat or excessively hot water.

16.3.2 Hospital treatment

Central hypothermia should be dealt with *first* through external rewarming by blankets and warm water bath. In severe cases (core body temperature below 30° C) combined or not with frostbite, central warming has a priority over peripheral warming because of the risk of "after-drop". This represents a situation where the core temperature tends to decrease during peripheral heating of the body. Rewarming of the limbs causes a local vasodilatation with shunting of cold stagnated blood towards the central part of the body; arrhythmias and cardiac arrest may result. Central core warming measures include warm i.v. fluids, rectal enema, and bladder, gastric and peritoneal lavage at 37° C. A low-reading thermometer is necessary to satisfactorily monitor core temperature (see Chapter 18).

Hypothermia takes precedence over local cold injury.

General rewarming should precede local rewarming.

Once hypothermia has been sufficiently corrected, attention may be given to the local cold injury. Superficial injuries may be rewarmed rapidly in 40 – 42° C water. For deep injuries, if the limbs are still frozen or cold and vasoconstricted, they should be rewarmed with dry heat at 37 – 39° C. Even if only one limb is affected, both should be rewarmed together until the nail beds become pink. The patient should receive 100 % oxygen, warmed and humidified, during the warming process.

If the patient is comatose, it should be borne in mind that the patient is not dead until he is *warm* (33° C core temperature) and dead!

Aspirin still appears to be a most useful drug for pain relief and possibly for preventing tissue loss. Pethidine may be added as necessary. Amitriptyline is the preferred analgesic for immersion injury. Tetanus prophylaxis and penicillin should be given. Smoking is contraindicated.

Heparin, anticoagulants, corticosteroids, antihistamines and intravenous dextran have all been shown to be of little benefit. Disagreement exists on the effect of sympathectomy.

16.3.3 After care

Once rewarming is complete, little more can be done to alter the course of events.

The basic treatement is conservative wound care.

Good nursing care and physiotherapy is the essence of conservative treatment. The extremities should be kept on sterile sheets under cradles. Sterile cotton pledgets are placed between toes or fingers. Warm povidone iodine soaks twice daily help prevent superficial infection. As blebs appear, precautions are taken to avoid their rupture: they should not be allowed to dry up. The limb should be placed so as to avoid pressure on the injured part as much as possible. A good functional result will be helped by maintaining active exercises and elevating the affected limb.

It is difficult to predict the extent of eventual tissue loss from frostbite during the first weeks after injury. Tissue loss, however, is generally smaller than expected. It is therefore important to wait until the damaged tissue becomes necrotic and mummified, with a clear line of demarcation and spontaneous amputation of fingers or toes. As with burns, escharotomy of circumferential injuries, and even fasciotomy, might be necessary.

Wait until a clear line of demarcation appears between necrotic and viable tissue.

The *surgical treatment* of local cold injury is to avoid excision of tissues unless secondary infection intervenes. The natural process is allowed to happen: "frozen in January, amputation in July".

Chapter 17

ANAESTHESIA AND ANALGESIA IN WAR SURGERY[1]

1 Much of this chapter is based on the report of the Senior Anaesthetists Workshop held in Geneva, November, 2002 (see Introduction).

17. ANAESTHESIA AND ANALGESIA IN WAR SURGERY

17.1 Introduction 307

17.2 Anaesthesia methods 308

17.3 Local and regional anaesthesia 309

17.4 Dissociative anaesthesia with ketamine 310
17.4.1 General considerations 310
17.4.2 Ketamine i.m. and i.v. bolus 311
17.4.3 Ketamine infusion anaesthesia 311
17.4.4 Ketamine analgesia 312

17.5 Post-operative pain management 312
17.5.1 General guidelines 312
17.5.2 Pain scoring systems 313

ANNEX 17. A ICRC pain management protocols 314

17.1 Introduction

Standard anaesthetic practices, as utilized in trauma care, should be followed. However, providing safe and effective anaesthesia under circumstances of limited resources is probably the most challenging task of hospital work. Many limitations must be accepted owing to security concerns, lack of infrastructure, and difficult logistics.

> A reminder to surgeons: operations are "big" or "small".
> All anaesthesia is potentially fatal.

This chapter is not intended for anaesthetists, but rather presents what surgeons should know about anaesthesia when working under precarious circumstances. There are several reasons for this.

1. Surgeons perform "big" operations and "small" operations. There is no such thing as a "small" anaesthesia. Every anaesthesia is potentially fatal. The limits of operational activity in the theatre are not those of the surgeon's expertise. They are determined far more by the level of competency and sophistication of the anaesthesia department. The anaesthetist will tell the surgeon what can be done; not the other way around. The surgeon must understand, and accept, this limit. There is only one other limiting factor that is as important as anaesthesia, if not more so, in determining the level of sophistication of surgery to be performed: postoperative nursing care.

2. In wartime, a surgeon may sometimes find himself without an anaesthetist. In this case, the war surgeon must know how to administer safe and adequate anaesthesia in order to perform the most important and basic of lifesaving procedures. There is much that a surgeon alone can do in terms of anaesthesia. Good knowledge of local anaesthesia techniques – practical, cheap and safe – can be very helpful when resources are limited.

3. Moreover, in many countries practising surgeons far outnumber the anaesthetists, and this is likely to remain the case in the foreseeable future. Anaesthesia under these circumstances is usually administered by a nurse or technician under the "guidance", and medical responsibility, of the surgeon, who must understand the important indications and contraindications of the various anaesthetic techniques. Needless to say, the surgeon must know about the possible complications of the anaesthesia used and how to counter them.

It would be presumptuous to try to explain the safe administration of anaesthesia in a half-dozen pages. This account comprises the basic precepts that a surgeon should be familiar with. For a full description of appropriate anaesthetic techniques under circumstances of limited resources, practicable by a surgeon or other medical officer, the reader is referred to the excellent texts listed in the Selected bibliography.

Figures 17.1.1 – 17.1.4
Standard ICRC equipment for anaesthesia.

R. Aburabi / ICRC

Figure 17.1.1
Typical operating theatre. Note the inhalation anaesthesia machine and oxygen concentrator.

M. Baldan / ICRC

Figure 17.1.2
Pulse oxymeter for monitoring; laryngoscope and endotracheal tube, bag for manual ventilation.

C. Gerber / ICRC

Figure 17.1.3
Foot-driven suction pump.

M. Baldan / ICRC

Figure 17.1.4
Electric suction pump.

ICRC EXPERIENCE

The points listed below have been found to be particularly important by ICRC anaesthetists and may be useful to other humanitarian organizations or foreign staff operating in a war environment in a new and strange context.

- The presence of a good interpreter, preferably with a medical background, is essential for communication with patients.
- Unfamiliarity with anaesthetic equipment and local conditions increases the likelihood of complications. The presence of an assistant knowledgeable about local standards is essential.
- The assistance of a second person – doctor, nurse – preferably with experience in anaesthesia, is mandatory for all anaesthetic procedures.
- Assessing the time the patient last ate is difficult, especially with breastfed babies.
- Many of the wounded, especially in hot climates, are severely dehydrated. It is important to correct any hypovolaemia before surgery.
- Routine premedication is not necessary.
- Arrangements for blood transfusion may be complicated (negotiating with family or clan members) and require local blood-collecting procedures to be activated. Likely blood requirements should be ordered early.

17.2 Anaesthesia methods

Local and regional anaesthesia techniques are excellent and can be used in many patients, although they tend to be underestimated and, therefore, underused: a below-knee amputation or Caesarean section can be performed under local anaesthesia, for example. They are safe in avoiding vomiting and aspiration, especially when the time of the last meal cannot be determined accurately.

Oxygen, nitrous oxide and other volatile anaesthetics may be in short supply. In the context of actual combat, oxygen cylinders are to be banned. Not only are the logistics for refilling cylinders and transporting them difficult and dangerous, but the presence of an oxygen cylinder is the equivalent of a bomb. Oxygen concentrators – with a pulse oxymeter for control – are standard ICRC equipment for these circumstances; however, they require an electrical supply.

Ketamine is the anaesthetic of choice in ICRC practice for major surgery. Equipment for the use of ketamine is minimal – not even an oxygen concentrator is necessary – and muscle relaxants can be added to provide full general anaesthesia. The further the hospital from the actual conflict zone, the more sophisticated the equipment and supplies that may be provided, and therefore the techniques employed.

The following anaesthetic agents and their characteristics should be well known to surgeons.

17.3 Local and regional anaesthesia

The clinically significant point in all forms of local and regional anaesthesia is to allow sufficient time for the anaesthetic to work. The most common mistake is to make a local anaesthetic infiltration, and then incise immediately.

	Technique	Drug of choice	Recommended volume	Remarks
Surface anaesthesia	Ophthalmic	0.4 % oxybuprocaine		If not available, 2 % or 4 % lidocaine (xylocaine) drops
	Mucosa	2 % lidocaine gel or 5 % lidocaine nebuliser spray		
Local infiltration anaesthesia	In general	1 % lidocaine with adrenaline	40 ml	If a volume greater than 40 ml is required, dilute with an equal volume of isotonic saline
	Fingers, toes, ears or penis	1 % lidocaine without adrenaline	20 ml	Danger of ischaemic gangrene if adrenaline is used
Nerve block	Digital block	2 % lidocaine without adrenaline	2 – 4 ml per digit	
	Axilla: brachial plexus block	2 % lidocaine with adrenaline or 0.5 % bupivacaine with adrenaline	30 – 40 ml	Where early movement is desired Provides prolonged sensory analgesia
	Intercostal block (flail chest)	0.5 % bupivacaine with adrenaline	2 ml for each nerve	Prolonged analgesia required
Regional intravenous anaesthesia		0.5 % lidocaine without adrenaline	40 ml	
Spinal anaesthesia	Subarachnoid block	0.5 % bupivacaine in 5 % dextrose		Hyperbaric, single dose ampoule
	Epidural	0.5 % bupivacaine	7 – 30 ml according to level of anaesthesia	Requires appropriate conditions of hygiene Not to be used routinely

Table 17.1 Anaesthetic agents of choice for local and regional anaesthesia.

Doses above are for adults. As a "rule of thumb", the maximum dose of lidocaine without adrenaline is 3 mg/kg (200 mg for an adult); and double with adrenaline: 6 mg/kg.

17.4 Dissociative anaesthesia with ketamine

The main components of safe and adequate general anaesthesia are:
- unconsciousness (hypnosis),
- analgesia,
- amnesia,
- immobility/muscle relaxation.

Ketamine is the anaesthetic of choice for major war surgery in conditions of limited resources.

17.4.1 General considerations

Ketamine is very safe, and can be given intramuscularly, or as an intravenous bolus or perfusion.

M. Baldan / ICRC

Figure 17.2
Drugs for general ketamine anaesthesia.

It causes amnesia, analgesia, and a state of dissociation. The patient feels detached from the environment, but his reflexes are intact, particularly the laryngo-pharyngeal reflexes that protect the airway. Opening of the eyes, shouting, and limb movements are frequent and *normal*; the patient is nevertheless anaesthetized. It may cause hallucinations and should be combined with a sedative such as a benzodiazepine (diazepam is inexpensive but burns on injection; midazolam is not irritant but is costly).

Ketamine increases the cardiac output and the arterial blood pressure and, therefore, is particularly useful in patients with haemorrhagic shock. Ketamine dilates the bronchi, but also increases bronchial secretions and saliva. Atropine should be administered first to control saliva and bronchial secretions. There are important contraindications to the use of atropine: marked tachycardia, hypertension, valvular stenosis, hyperthyroidism, or fever. Ketamine also increases the muscular tone, which is why it makes a laparotomy, and especially the closure of the abdomen, more difficult if used without myorelaxants.

There is no absolute contraindication for ketamine anaesthesia. Relative contraindications include:
- psychiatric disease,
- mitral and aortic stenosis,
- untreated hyperthyroidism,
- eclampsia,
- epilepsy,
- operation on the eye – this is due to nystagmus, which causes the globe to move and makes operating difficult.

The question of the use of ketamine in patients with a head injury or increased intracranial pressure has been controversial, although based on very few and very early reports in the 1970s. There is no longer any contraindication for its use in these patients consistent with evidence-based medicine, and ICRC standard practice includes craniotomy performed under ketamine anaesthesia (see Selected bibliography).

The *advantages* of ketamine are the following:
- easy and rapid to administer with quick onset of action;
- safe;
- provides anaesthesia, amnesia, and analgesia;
- circulation is stimulated;
- respiration remains normal (very slow intravenous injection);
- protective airway reflexes remain intact in the majority of patients – vomiting may however occur: staff should be ready to suck clean the mouth and secure the airway;
- maintains cerebral blood flow;
- particularly suitable for infants.

17.4.2 Ketamine i.m. and i.v. bolus

A simple form of ketamine administration is represented by both the intramuscular and the bolus intravenous routes. Table 17.2 compares the two.

	Ketamine i.m.	Ketamine i.v. bolus
Indications	Short duration operation (10 – 20 minutes) Anaesthesia in children (injected while the mother is holding the infant) Repeated injections for change of dressings in burn patients with poor i.v. access	Short duration operation (10 – 20 minutes) Induction of anaesthesia
Premedication	Preferable if time allows In children diazepam may be given orally or rectally	Preferable if time allows
Technique	Ketamine and atropine can be mixed in the same syringe Diazepam should be given in a separate syringe	Place an intravenous line and inject atropine Inject diluted diazepam solution very slowly over 3 minutes, until the patient becomes drowsy Inject ketamine very slowly over one minute (rapid injection can provoke respiratory arrest)
Latent period	5 – 10 minutes	Painful stimuli (skin incision) can be tolerated after one minute
Continuing anaesthesia: second dose	Second injection can be given if necessary before the effect of the first wears off	After 10 – 15 minutes, the patient begins to feel pain and reacts to stimuli with movements and speech – not to be confused with normal movements under ketamine Administer another i.v. bolus: one-third to one-half the initial dose

Table 17.2 Comparison of ketamine anaesthesia administered intramuscularly versus intravenous bolus.

In special circumstances (repeated ketamine anaesthesias with "ketamine-resistance" and/or disturbing limb movements), one or more of the following may be added according to the reaction of the individual patient:

- 50 – 100 mg tramadol or 5 –10 mg morphine,
- 10 – 25 mg promethazine,
- 50 – 100 mg thiopental.

17.4.3 Ketamine infusion anaesthesia

This is the preferred technique in ICRC practice. It is not only more economical of ketamine, but allows for a longer operation without re-injections. It can be used either after i.v. bolus induction of anaesthesia or as an induction method on its own.

A solution is made of ketamine in normal saline and placed in a different vein from that used for fluid replacement. The rate of infusion is titrated against the patient's response, both for induction and as maintenance of anaesthesia.

Ketamine infusion anaesthesia can be combined with muscle relaxants and endotracheal intubation. This is standard ICRC procedure where muscle relaxation is necessary (abdominal or thoracic surgery). As mentioned previously, the absence of mechanical ventilators means that the paralysed patient must be bagged manually.

ICRC

Figure 17.3
Ketamine infusion.

ICRC EXPERIENCE

The Senior Anaesthetists Workshop recommended that, in exceptional circumstances (extreme emergency with a very small ICRC surgical team, no training of local staff intended, equipment to be withdrawn afterwards), the ICRC deploy small portable ventilators "to free the hands" of the anaesthetist. These ventilators must be able to function with an oxygen concentrator (there must be a regular supply of electricity) rather than with compressed air or oxygen. They were sent to Banda Aceh, Indonesia, after the 2006 tsunami hit this conflict-affected area because of an "epidemic" of tetanus cases encountered there.

Hallucinatory side effects do occur with ketamine, and once encountered are likely to be repeated on other occasions in the same patient. The frequency of hallucinations tends to differ with different population groups, and bizarre reactions may occur in heavy users of alcohol or other drugs. In addition, such patients do not reach the required state of general anaesthesia unless a neuroleptic and opioid are added.

The incidence and severity of hallucinations can be lessened by using diazepam 10 mg i.v. at least 5 minutes before induction and another 10 mg i.m. at the end of the procedure. After the operation, patients often sleep for hours. They should be in a calm and quiet place. On awakening, they respond to orders "show your tongue" or "raise your head". Sometimes these patients are very loud, they shout or sing, and move a great deal. This is normal and not dangerous for them.

17.4.4 Ketamine analgesia

The analgesic effect of ketamine can be profitably used for a number of conditions. One noteworthy example is for repeated change of dressings in burn patients. For analgesia, a lower dose than that given for intramuscular anaesthesia is administered to those whose venous access is compromised, as is often the case with major burns (see Annex 17. A: ICRC pain management protocols).

17.5 Post-operative pain management

Good post-traumatic or post-operative pain relief not only helps to alleviate suffering, but also allows for rapid mobilization of the patient and early physiotherapy which help attain as good a functional result as possible.

17.5.1 General guidelines

1. Analgesia should be given *before* the pain starts.
2. Combination therapy is recommended. Analgesic drugs have a better effect in combination than alone; e.g. paracetamol and a NSAID have additive actions because of their different sites of action.
3. Local anaesthetic infiltration or blocks are used whenever appropriate in conjunction with other forms of analgesia.
4. Injectable analgesics act more quickly and are more effective if administered as an i.v. bolus, titrated until the desired effect has been achieved. This is especially true in case of hypovolaemia and shock where the peripheral circulation is decreased and, therefore, the intramuscular or subcutaneous routes are unreliable.
5. The choice of drugs is made with the staff and facilities in mind; e.g. opioids may be undesirable if proper monitoring is unavailable.
6. It is advisable to institute a system of pain scoring, especially for post-operative pain relief.

17.5.2 Pain scoring systems

Pain intensity can be measured in various ways and the use of a system of measurement is recommended. The choice of scoring system depends on a number of factors, including cultural issues and the level of literacy and numeracy of patients and the nursing staff. In all of them it is the patient who assesses the intensity of pain.

Common scoring systems include the following.

Verbal score
The pain intensity is measured using words:
- none
- mild
- moderate
- severe
- extreme

Number scale
Numbers from 0 to 10 should be used:
- 0 means no pain and 10 the worst pain imaginable.

Visual analogue score
This is analogous to the number scale. The pain intensity is measured on a 10 cm line. The left end indicates "no pain" and the right-hand side means "the worst pain imaginable". The distance in cm from the left edge to the point the patient indicates is the pain score.

The availability of particular drugs for pain relief in a given country is often subject to importation and distribution restrictions. The widespread abuse of opioids in particular has created many problems in the proper management of patients. Annex 17. A includes various alternatives that may be used according to the availability of various medications. These must of course be adapted to specific situations.

ANNEX 17. A ICRC pain management protocols

Analgesic drugs

Modern analgesics can be very sophisticated, and expensive. The ICRC has the following simple analgesic drugs available for each of the three WHO levels:

Level 1:
Paracetamol tablets/syrup, aspirin tablets/injection, the non-steroidal anti-inflammatories (NSAID) diclofenac injection and ibuprofen tablets;

Level 2:
Tramadol injection/tablets;

Level 3:
Morphine injection, pethidine injection, pentazocine injection/tablets.

Opioids

1. When level 3 opioids are used in a hospital, naloxone must also be immediately available.
2. Morphine remains the gold standard for pain relief, intra-operatively as well as pre- and post-operatively.
3. The use of opioids – morphine and pethidine, but not tramadol – may be severely restricted because of the risk of respiratory depression. ICRC surgical teams use opioids only if nursing staff are sufficient in numbers and trained in patient monitoring, as well as in recognition and treatment of respiratory depression. In practice this will usually mean that post-operative opioids are only used in the recovery ward or intensive care unit (ICU).
4. Do not mix opioids (e.g. tramadol and morphine) until an appropriate time has lapsed.
5. In the emergency room, operating theatre and recovery area/ICU, the preferred route for opioid administration is intravenous. If they are used in the ward, the oral or subcutaneous routes are preferred.
6. The fear of addiction when using opioids for pain relief, although a generally held belief, is exaggerated; if a patient is in pain and needs an opioid, the development of addiction is extremely rare. The fear of addiction must not stop good pain control.
7. Sedation score: when using level 3 opioids the level of sedation should be monitored.
   0 = none: patient awake and alert
   1 = mild: occasionally drowsy, but easy to rouse
   2 = moderate: frequently drowsy, but easy to rouse
   3 = severe: drowsy and difficult to rouse
   S = sleep: normal sleep and easy to rouse

Monitoring, diagnosis, and treatment of respiratory depression
When using level 3 opioids the patient must be monitored regularly for:
- blood pressure
- pulse
- temperature
- respiratory rate
- sedation score
- pain score

All findings must be documented.

Respiratory depression

1. Diagnosis
   Respiratory depression is diagnosed by:
   – sedation score of 3: this is the earliest and most reliable sign;
   – respiratory rate 8 or less: this comes later than sedation;
   – decreasing pO_2, as shown by pulse oximetry: this is a late sign especially if the patient is also being given oxygen.

2. Treatment:
   – oxygen;
   – respiratory assistance with bag and mask if necessary;
   – naloxone i.v. in 50 mcg increments until clinical signs improve.

Remember that the duration of action of naloxone is shorter than morphine and therefore it may need to be repeated; alternatively, use a continuous infusion of 1 – 5 mcg/kg/hr.

Never must a patient be discharged to the ward with a sedation level of 3 or more, or a respiratory rate of 8 or less, or if respiratory depression has occurred.

Ketamine

A low dose of ketamine is a good alternative analgesic where opioids are not available or for high-risk patients. Give repeated doses 0.1 – 0.3 mg/kg i.v., titrated until the desired analgesia is achieved; or an i.m. bolus of 2 – 3 mg/kg. Ketamine in a low dose does not require routine atropine and diazepam as adjuvant therapy.

Pre-hospital care

For pre-hospital first aid, often through National Red Cross/Red Crescent Societies, ICRC distributes only:

- paracetamol tablets/syrup;
- tramadol injection (does not provoke respiratory depression).

Emergency room

The following are available in the emergency room:

- paracetamol
- diclofenac injection
- tramadol injection

In a hospital with adequate nursing care, the following are also provided:

- morphine injection
  – adult: 1 – 3 mg i.v. titrated
  – child: 0.05 mg/kg i.v. titrated
- low-dose ketamine (see above).

Operating theatre

All analgesic drugs are available and their intra-operative use is greatly encouraged. The choice of drug depends to a large extent on the quality of post-operative monitoring.

1. Opioids:
   note the level of post-operative monitoring.

2. NSAID:
   it is advisable to administer an NSAID before the end of surgery.

3. Ketamine:
   0.1 – 0.3 mg/kg boluses as analgesia can be used if opioids are unavailable.

4. Local and regional anaesthesia:
   The use of local anaesthetic infiltration or local and regional blocks is greatly encouraged.

Post-operative pain control

The level of pain control needed depends on the psychology of the patient, the type of surgery, and the time since surgery. The basic principles of post-operative analgesia are the following.

1. Give analgesia regularly and not on an as necessary basis (PRN).
2. Do not wait for the pain to be felt, but start immediately once the patient has regained consciousness. This also means that analgesia should be started before spinal anaesthesia has worn off.
3. Start with combination treatment and decrease doses in subsequent days.
4. Review post-operative analgesia regularly.
5. Use local anaesthetic infiltration or blocks as much as possible.

MILD PAIN

Paracetamol
+
Local infiltration or block

MODERATE PAIN

Paracetamol
+
NSAID
+
Local infiltration or block

SEVERE PAIN

Paracetamol
+
NSAID
+
Opioid
+
Local infiltration or block

Analgesic doses

ADULT			
PARACETAMOL	oral	1 g QID Max. 4 g/day	
IBUPROFEN	oral	400 mg TID/QID Max. 2.4 g/day	Caution in cases of asthma and renal impairment Max. 72 hours
DICLOFENAC	i.v./i.m.	75 mg BID Max. 150 mg/day	Caution in cases of asthma and renal impairment Max. 72 hours
TRAMADOL	oral/i.v.	50 – 100 mg 4 hourly Max. 600 mg/day	
PETHIDINE	i.m.	50 – 150 mg 3 hourly	
	i.v.	10 mg increments	Titrate to effect
MORPHINE	subcutaneous / i.m.	5 –15 mg 4 hourly	
	i.v.	2 mg increments	Titrate to effect
NALOXONE	i.v.	50 mcg increments	Repeat until clinical signs improve

PAEDIATRIC			
PARACETAMOL	INFANTS (0 – 12 months)		
	oral / suppository if available	Loading dose: 15 mg/kg Maintenance: 10 – 15 mg/kg QID Max. dose: 60 mg/kg/day	
	CHILDREN		
	oral / suppository if available	Loading dose: 20 – 30 mg/kg Maintenance: 20 mg /kg QID Max dose: 90 mg/kg/day	
IBUPROFEN	oral	20 mg/kg/day in 3 – 4 divided doses Max. single dose: 200 mg Max. daily dose: 800 mg	Not under 6 months (immature kidneys) Caution in cases of asthma and renal impairment Max. 72 hours
DICLOFENAC	i.m.	1 mg/kg TID Max. single dose: 50 mg Max. daily dose: 150 mg	Not under 6 months (immature kidneys) Caution in cases of asthma and renal impairment Max. 72 hours
TRAMADOL	Not recommended, but commonly used in some European countries for children under 1 year		
PETHIDINE	i.m.	1 mg/kg 4 hourly	
	i.v.	0.25 – 0.5 mg/kg increments	Titrate to effect
MORPHINE	subcutaneous / i.m.	0.05 – 0.1 mg/kg 4 hourly	
	i.v.	0.05 mg/kg increments	Titrate to effect
NALOXONE	i.v.	4 mcg/kg	Repeat until clinical signs improve

BID: *bis in die* (twice a day)

QID: *quater in die* (four times a day)

TID: *tri in die* (three times a day)

Chapter 18

DAMAGE CONTROL SURGERY AND HYPOTHERMIA, ACIDOSIS, AND COAGULOPATHY

18. DAMAGE CONTROL SURGERY AND HYPOTHERMIA, ACIDOSIS, AND COAGULOPATHY

18.1 Resuscitative surgery and damage control surgery 321
18.1.1 Damage control: three-stage protocol 321

18.2 Hypothermia, acidosis, and coagulopathy 322
18.2.1 Hypothermia 323
18.2.2 Acidosis 326
18.2.3 Acute coagulopathy associated with trauma 326

18.1 Resuscitative surgery and damage control surgery

Resuscitative surgery has long been practised, and re-invented by many surgeons beginning with Pringle in 1908 and Halstead in 1913. Today, this is especially the case in critical situations where blood for transfusion is in short supply. The surgeon controls massive haemorrhage with temporary measures and then stops the operation. The patient is taken to the ward to continue resuscitation and, 24 – 48 hours later when haemodynamically stable, is re-operated for the definitive procedure.

ICRC EXPERIENCE

Afghan colleagues working with ICRC surgical teams in the Jalalabad Teaching Hospital in 1993 developed a simple protocol for dealing with patients severely injured by anti-personnel landmines. Many patients had suffered a traumatic amputation of one leg and severe injuries to the other and succumbed, exsanguinated, by the time the debridement of the second leg had been performed.

Our Afghan colleagues then decided to divide the surgical procedure into two parts: the first operation was on the traumatic amputation, the other leg was simply washed and dressed, and the operation ended. Fresh whole blood was sought from family members and the patient fully resuscitated and given penicillin. Forty-eight hours later, another operation to debride the second leg was performed. This was Afghan "damage control surgery", locally invented, to face the situation of a lack of blood for transfusion.

Shortly after a war surgery seminar in Kinshasa, two young Congolese general practitioners belonging to the military medical services, with limited surgical expertise, told the ICRC surgeon of one of their experiences and asked a simple question. While working in a field hospital in the bush, they received a comrade with a bullet injury to the abdomen. Evacuation to another hospital was impossible. They operated and found a severe wound to the liver, and had no blood for transfusion.

"We couldn't stop the bleeding, and not knowing what to do, we packed the liver and stopped the operation", they recounted. They then managed to persuade some other soldiers to donate several units of blood, and re-operated the patient after 48 hours; the patient survived. "Did we do the right thing?" they asked. The ICRC surgeon replied that they had discovered on their own what was now standard practice in many parts of the world when faced with exsanguinating haemorrhage.

18.1.1 Damage control: three-stage protocol

Recently, it has become clear that fixing the tissues and organs during a long operation is not sufficient if certain physiological limits are thereby exceeded.

In response, a three-stage protocol for "damage control surgery"[1] has been defined to preclude the risk of pushing the patient beyond those limits into the "lethal triad" of hypothermia, metabolic acidosis, and coagulopathy that follows much major trauma.

First stage:
Operation, as short as possible, to control haemorrhage and contamination: do the least to overcome life-threatening conditions.

Second stage:
Resuscitation, to stabilize the patient by correction of shock, hypothermia, acidosis, and coagulopathy.

Third stage:
Operation, for definitive repair.

1 "Damage control" is originally a United States Navy term denoting "the capacity of a ship to absorb damage and maintain mission integrity". In surgery, we mean the procedures necessary to maintain the physiological capacity to maintain life's vital functions.

The success of this formalized protocol depends almost entirely on the early recognition that the injury is so severe as to require a curtailment of the initial operation; and, especially, the second-stage capacity to correct certain physiological parameters in the intensive care unit (ICU).

Damage control surgery is obviously restricted to the very severely injured, a small minority. It is a much individualized form of management requiring many resources, and may not be compatible with a situation of triage of mass casualties. The patient who qualifies for damage control in a civilian context, or in a military one with a small number of casualties, will in many cases be triaged into the expectant Category IV in a war scenario.

In much of the world, sophisticated ICUs and blood components are simply not available and a full-scale damage-control approach is not compatible with limited resources. However, basic principles apply and simple measures can and must be adapted to try to prevent and correct the syndrome of the lethal triad.

In the two examples cited above, the time between operations – used to obtain fresh whole blood from family and friends and "stabilize" the patient – helped to fight hypothermia, acidosis, and coagulopathy without knowing it. Attention was paid only to the haemodynamics and the comfort of the patient (keeping him warm), but this was sufficient to overcome the triad. What started as a "resuscitative surgery" approach became a "damage control" one unknowingly.

ICRC

Figure 18.1

Typical patient who would benefit from a damage control approach: eventration and a shattered liver.

R. Gray / ICRC

Figure 18.2

Another candidate for damage control surgery: APM injury with penetrating wounds to the abdomen, thorax, left thigh and hand, and the face.

ICRC

Figure 18.3

Patient with APM blast mine injury: traumatic amputation of the left foot, penetrating injuries and burns to both legs, the perineum and genitalia, with penetration of the abdomen.

18.2 Hypothermia, acidosis, and coagulopathy

The effects of hypothermia in trauma patients have long been underestimated. It afflicts the injured even in tropical climates. While metabolic acidosis due to shock – the anaerobic metabolism resulting from reduced tissue perfusion and hypoxia – and coagulopathy are better known, the combined triad is far more prevalent than is often realized and its effects easily fatal. These three elements compound and reinforce each other, leading to a self-sustaining vicious circle. Haemorrhagic shock may not be corrected until hypothermia is reversed, and hypothermia can provoke the other two elements of the triad and is probably the most important factor initiating the vicious circle.

Early diagnosis is essential but simple preventive measures, that should be initiated during first aid and evacuation and continued in the hospital, are often not taken. Even if the context does not allow for staged damage control surgery, a great deal can be done with adequate but appropriate means to prevent or fight off the effects of the triad.

18.2.1 Hypothermia

Physiology

The body's temperature is kept in a homeostatic balance between heat production and heat loss. There are many non-traumatic causes and factors that can lead to hypothermia; cold injuries were discussed in Chapter 16. In trauma patients, haemorrhagic shock decreases tissue perfusion and metabolism and thus heat production. This is often compounded by exposure of the casualty to the elements, particularly in a context of armed conflict.

> It should be taken as a general rule that every injured person will lose body heat, even in a tropical climate.

Hypothermia is determined by central body – core – temperature, taken *per rectum*. A normal medical thermometer is useless; a special low-reading one beginning at 30° C is necessary. Classically, a core temperature below 35° C denotes hypothermia and medical classifications reach profound states of 25° C or less (immersion in a cold lake, hypothalamic disorders, drug abuse, etc.). Some patients suffering from severe hypothermia do survive. There are also reported beneficial effects of controlled therapeutic hypothermia in certain critical pathologies, but this topic shall not be considered here.

Pathology

It is very rare for patients suffering trauma *and* uncontrolled hypothermia lower than 32° C to survive. This observation is not new, although the threshold has changed.

> "It has been said that any wounded man in a state of stupor whose temperature sinks below 36° C will die."
>
> E. Delorme, 1915[2]

In modern practice, the temperature of 32° C is recognized as the crucial cut-off point. All enzyme systems in the body are temperature dependent and, therefore, all organ systems will tend to fail at this low core temperature, especially if stress is accentuated by trauma and shock.

Thus, medical classification systems of hypothermia with wide ranges of temperature are not suitable for the trauma patient. A more appropriate classification system has been proposed (Table 18.1), and takes into account the profound effects of trauma and haemorrhage with concomitant acidosis and coagulopathy. In the trauma patient, any temperature below 36° C should be considered hypothermic and it is particularly important to distinguish between Stage I and Stage II. A damage-control approach, with abbreviated first operation as mentioned above, is indicated at 34° C, now considered the critical clinical temperature.

General medical classification		Trauma classification	
Mild	35° – 32° C	I	36° – 35° C
		II	34° – 32° C
Moderate	32° – 28° C	III	32° – 28° C
Severe	28°– 20° C	IV	< 28° C
Profound	< 20° C		

Table 18.1 Classification systems of hypothermia.[3]

2 Delorme E. *War Surgery*. [Translated by Méric H.] London: H.K. Lewis; 1915. [WWW Virtual Library, The Medical Front WWI website] Available at: http://www.vlib.us/medical/delorme/delorme.htm.

3 Adapted from Kirkpatrick AW, Chun R, Brown R, Simons RK. Hypothermia and the trauma patient. *Can J Surg* 1999; **42**: 333 – 343.

The *clinical effects* of hypothermia are multiple and mimic intense sympathetic stimulation:
- shivering – the patient attempts to produce body heat by these muscular contractions but this leads to increased oxygen consumption and tissue hypoxia;
- hypoventilation – leads to postoperative hypoxaemia, with further tissue hypoxia;
- peripheral vasoconstriction – to shunt blood to the central organs, with tissue hypoxia;
- decreased oxygen liberation in the tissues (O_2 dissociation curve shifted to the left);
- decreased metabolism.

Furthermore, the tissue hypoperfusion and hypoxia of shock in the trauma patient also increase anaerobic metabolism. All of which result in a mixed metabolic and respiratory *acidosis*.

Enzyme failure results in compromised platelet function; platelet-fibrin clots are not sustained; and fibrinolytic activity is increased. There is dramatic prolongation of prothrombin time (PT) and partial thromboplastin time (PTT).

Please note:

Prolonged PT and PTT may be overlooked because the testing is performed at 37° C rather than at the patient's actual temperature.

There is in addition a rapid drop in the number of platelets and increased blood viscosity following shock. The end result is a disseminated intravascular *coagulopathy* with a marked bleeding tendency.

Other physiological effects include:
- decreased immune response;
- hyporeflexia and prolonged action of neuro-muscular blockers;
- reduced cardiac output, depressed contractility and bradycardia;
- arrhythmias, including atrial and ventricular fibrillation, which begin at 30° C;
- impaired hepatic and renal function;
- impaired cerebral blood flow with decreasing consciousness.

Clinical picture

Early signs include shivering and tremors; the patient first complains of cold then becomes confused. There is cyanosis and the axilla and groin (normally warm spots) feel cold to the touch because of intense vasoconstriction. Tachycardia and tachypnoea (with hypoventilation) occur. All are signs of sympathetic stimulation. Making the diagnosis of hypothermia is a challenge because *most of the early signs* are the normal physiological response to the stress of trauma and haemorrhage; the surgeon should be particularly attentive to the presence of shivering and tremors. It is also difficult to judge clinically when the patient passes from Stage I to II without close monitoring of the rectal temperature.

As the patient goes into Stage III, there is a general slowing down of all the vital functions:
- loss of shivering and hyporeflexia;
- respiratory and circulatory depression;
- decreased urine output;
- increasing acidosis;
- prolonged clotting time;
- stupor; and,
- with myocardial ischaemia, the onset of attacks of atrial fibrillation.

With even greater hypothermia the patient becomes unconscious; bradycardia and bradypnoea are pronounced; urine output fails; and ventricular fibrillation resistant to defibrillation occurs. The patient may appear dead: no palpable pulse or heart beat and dilated pupils. Re-warming should continue until a core temperature of at least 33° C before pronouncing death.

High risk cases

At high risk of developing post-traumatic hypothermia are patients who have:

- been severely injured and are submitted to a long delay in evacuation to hospital;
- remained trapped under rubble (earthquake, collapsed buildings due to bombing, etc.), and thus been exposed to the environment without being able to take protective measures for a considerable period of time;
- sustained major burns;
- suffered haemorrhagic shock treated with large quantities of room temperature intravenous perfusions or cold transfusions – banked blood is stored at 4° C and its administration acts as a heat sink that withdraws large amounts of body heat;
- been subjected to long, extensive laparotomy or thoracotomy in air-conditioned operating theatres – the surgeon and anaesthetist are comfortable, but the patient is dying;
- suffer from chronic illnesses involving metabolism, alcoholism, and drug abuse; and
- the very young and very old, in general.

Please note:

These high risk factors are partly intrinsic to the patient and partly iatrogenic.

Management

Treatment begins with active measures of prevention during first aid and evacuation: the casualty must be kept warm! The patient must be sheltered from cold and wind, relieved of wet clothing, and covered with a dry blanket or sheet, even in a tropical climate.

In hospital, active preventive measures must continue: the emergency room should be kept thermo-neutral (28° C for an adult) and, after the patient has been undressed, examined, and resuscitative measures begun, covered with a blanket. For the severely injured, further environmental measures should continue. In the operating theatre, air-conditioning should be turned off (the OT should be heated if necessary), and prolonged contact with wet drapes avoided. The impact of the latter should not be underestimated: the patient starts off dry and warm and ends up wet and hypothermic.

A patient shivering on the operating table is a clear warning sign!

Under muscle relaxant anaesthesia, a patient can no longer shiver and is particularly at risk. Since 20 – 30 % of body heat is lost through the head and neck, keeping these body parts warm is a priority (e.g. wrapping the head in a towel and plastic bag intra-operatively). The rest of the body, apart from the operative site, can equally be wrapped in dry towels and secured with plastic bags to preserve heat. Warm saline should be used for peritoneal or pleural lavage. Operating time should be kept as short as possible; an open abdomen or thorax inevitably entails loss of body heat which is just as critical at this stage as blood loss.

Administered oxygen should be humidified – a heat and moisture exchange filter (HME) should be used, if available – and intravenous fluids and blood transfusions warmed. Fluids and blood can be heated by being placed in a tub of hot water, comfortable to an elbow touch.

All of these simple measures, which do not require sophisticated equipment, should be standard procedure in all surgical patients, especially those suffering from trauma.

Not only will these simple measures prevent hypothermia, but their early implementation will help treat the condition.

If a damage-control approach is being used, or even if a major trauma case presents signs of hypothermia postoperatively, resuscitation should include active internal "central heating" of the patient by:

- gastric, colonic and bladder lavage with warm water (37 – 39° C); and
- continued perfusion of similarly warm i.v. fluids.

More sophisticated technologies exist – including extracorporeal circulation, pleural and peritoneal lavage in the ICU – but the vast majority of patients will be well served by these simple and inexpensive preventive and active measures.

A patient in Stage I hypothermia may have definitive surgery completed. A Stage II patient should have damage control surgery only. If Stages III or IV have been reached, damage control surgery may be slightly postponed in order to warm the patient somewhat before opening the peritoneum or pleura, but this will depend on the particular circumstance and haemodynamic stability of the individual patient. Again: rare is the haemorrhaging trauma patient with a core temperature below 32° C who survives, however good the surgery.

Type of measures	Procedures	Applicability
Standard measures	Passive external warming (warm environment, blankets and covers) + warmed intravenous fluids + warmed and humidified oxygen	Standard measures are applicable for all Stages and at all times
Active external warming	Electric blankets Environmental heater	Particularly suitable for ICU/ emergency room
Active internal warming intraoperatively	Irrigation of pleural/peritoneal cavity with warm saline during surgery	Should be routine surgical procedure
Active internal warming postoperatively	Warm gastric, colonic, and bladder lavage Warm pleural/peritoneal lavage if feasible	Particularly suitable for ICU

Table 18.2 Summary of management of hypothermia in the trauma patient.

18.2.2 Acidosis

As mentioned, shock will lead to tissue hypoperfusion and hypoxia with a subsequent rise in anaerobic metabolism. This acidosis is compounded by the effects of hypothermia. Full resuscitation and haemodynamic stability in a warm patient, assuring good tissue perfusion, is the best antidote. Intravenous sodium bicarbonate is risky and requires sophisticated monitoring.

18.2.3 Acute coagulopathy associated with trauma

Post-traumatic coagulopathy is also more common than realized and results from a combination of intrinsic and extrinsic factors. Amongst the intrinsic ones, the most important include:

- extensive tissue injury, which releases tissue factors resulting in consumption coagulopathy (with reduced platelet count), excessive fibrinolysis, and activation of inflammatory cascade;
- activation of the C-reactive protein inflammatory cascade independently of the release of tissue factors and entirely due to the shocked state and tissue hypoperfusion;
- haemodilution resulting from the mobilization of extravascular fluids as a homeostatic response to shock;
- decrease in the total and ionized calcium concentrations; and
- the effects of progressive hypothermia and acidosis.

The severity of the coagulopathy seems to parallel the severity of injury and shock.

Important factors resulting in post-traumatic coagulopathy:
- tissue damage
- shock: hypoperfusion and homeostatic haemodilution
- hypothermia
- acidosis
- inflammation
- therapeutic haemodilution with i.v. fluid resuscitation

There are also a number of extrinsic factors. Coagulopathy itself is often exacerbated by clinical practice in a "bloody vicious cycle".[4] Haemorrhagic shock is aggressively treated with large volumes of intravenous fluids (not warmed), leading to further haemodilution, and massive transfusion of stored blood or packed RBCs (not warmed) in the attempt to maintain blood pressure (see Chapter 8); all of which adversely affect the coagulation cascade and only produce more bleeding. Bleeding and coagulation times remain prolonged despite adequate replacement of blood, plasma and platelets if the patient remains hypothermic.

It must also be remembered that the anticoagulant preservative solution and refrigeration of stored blood or its components has profound effects on the physiological properties of transfused blood with time, the most important being greatly reduced oxygen availability to the tissues through decreased 2,3-diphosphoglycerate (2,3-DPG) and loss of red cell viability because of decreased adenosine triphosphate (ATP). These effects do not help overcome early tissue hypoxia in spite of the increase in haemoglobin. In addition, the glucose in stored blood is slowly metabolized with production of lactate and fall in the pH, exacerbating the acidosis.

Warm whole blood, as fresh as possible, is probably the best treatment; and, as often repeated in this manual, in conditions of limited resources is often all that is available. With numerous potential donors (family and friends), it may be possible to build up a small reserve.

Whole blood is screened and best given within 1 hour of collection. In ICRC practice, this freshest of blood is reserved for:
- massive haemorrhage, with incipient coagulopathy and hypothermia;
- septic shock;
- snake bites with haemolysis;
- amniotic fluid embolism.

Intravenous calcium should be administered, separately and at least one ampoule added per two units of blood transfused.

Coagulopathy also attends many patients with severe head injury, and a damage control approach is applicable to injuries to most body systems. Volume 2 will include relevant observations in the discussion of war injuries to the different anatomic regions.

4 Kashuk J, Moore EE, Milikan JS, Moore JB. Major abdominal vascular trauma – a unified approach. *J Trauma* 1982; **22**: 672 – 679.

ACRONYMS

ABCDE	Airway, Breathing, Circulation, Disability, Environment/Exposure
APM	Anti-personnel mine
ARDS	Acute respiratory distress syndrome
ATM	Anti-tank mine
ATP	Adenosine triphosphate
AVPU	Alert, Voice responsive, Pain responsive, Unresponsive
BID	*Bid in die* (twice a day)
CFR	Case fatality rate
CPD-A	Citrate – phosphate – dextrose – adenine
CRO	Carded for record only
CT	Computerized tomography
DOA	Dead on arrival
DOW	Died of wounds
DPC	Delayed primary closure
2,3-DPG	2,3-diphosphoglycerate
E_K	Kinetic energy
$E_{K\,EXP}$	Kinetic energy actually expended, transferred or dissipated
ECG	Electrocardiogram
EDTA	Ethylene diamine tetra-acetic acid
EEG	Electroencephalogram
ER	Emergency (reception) room
FAP	First-aid post
FMJ	Full metal jacket
FST	Field surgical team
GC	Geneva Convention
GCS	Glasgow coma scale
GSW	Gunshot wound

H.E.L.P.	Health emergencies in large populations
HIV/AIDS	Human immunodeficiency virus / Acquired immunodeficiency syndrome
HME	Heat and moisture exchange (filter)
HN	Head nurse
ICRC	International Committee of the Red Cross
ICU	Intensive care unit
IED	Improvised explosive device
IHL	International humanitarian law
i.m.	Intramuscular
i.v.	Intravenous
IU	International unit
KIA	Killed in action
LF	Low frequency
mEq	Milliequivalents
MIU	Million international units
MoPH	Ministry of Public Health
MRSA	Methicillin-resistant *Staphylococcus aureus*
MSF	Médecins sans frontières (Doctors Without Borders)
NATO	North Atlantic Treaty Organization
NGO	Non-governmental organization
NSAID	Non-steroidal anti-inflammatory drug
OPCW	Organization for the Prohibition of Chemical Weapons
OPD	Outpatient department
ORS	Oral rehydration salts
OT	Operating theatre
POP	Plaster of Paris
POW	Prisoner of war
PRN	*Pro re nata* (whenever necessary)
PT	Prothrombin time
PTT	Partial thromboplastin time
QID	*Quater in die* (4 times per day)
RBC	Red blood cell
RC/RC	Red Cross/ Red Crescent
RCWS	Red Cross Wound Score
RTD	Returned to duty
SJ	Semi-jacketed
SSG	Split-skin grafts
TBSA	Total body surface area

TID	*Ter in die* (3 times per day)
TNM	Tumour, node, metastasis
TO	Triage officer
TTL	Triage team leader
UN	United Nations
UNPROFOR – IFOR	United Nations Protection Force – (NATO) Implementation Force
UXO	Unexploded ordnance
WDMET	Wound Data and Munitions Effectiveness Team
WHO	World Health Organization
WIA	Wounded in action

SELECTED BIBLIOGRAPHY

General texts

Please note:

The publications listed under general texts serve as basic references throughout this manual.

Military publications

Kirby NG, Blackburn G, eds. *Field Surgery Pocket Book*. London: Her Majesty's Stationery Office; 1981.

Roberts P, ed. *The British Military Surgery Pocket Book. (AC 12552)* Camberley, Surrey, UK: Department of Military Surgery, Army Medical Directorate; 2003.

Bowen TE, Bellamy RF, eds. *Emergency War Surgery NATO Handbook Second United States Revision*. Washington, DC: US Department of Defense; 1988.

Lounsbury DE, Brengman M, Belamy RF, eds. *Emergency War Surgery Third United States Revision*. Washington, DC: Borden Institute, US Department of Defense; 2004. Available at: http://www.bordeninstitute.army.mil/other_pub/ews.html.

Nessen SC, Lounsbury DE, Hetz SP, eds. *War Surgery in Afghanistan and Iraq: A Series of Case Studies, 2003 – 2007*. Washington, DC: Office of the Surgeon General, Borden Institute, US Department of Defense; 2008.

Pons J, ed. *Memento de chirurgie de guerre [War Surgery Primer]*. Paris: Ecole d'application du Service de santé pour l'armée de terre, ORA éditions; 1984.

War Surgery Commission of the Federal Military Department. *Chirurgie de guerre (Aide-mémoire 59.24 f) [War Surgery (A Primer)]*. Bern: Swiss Army; 1970 and 1996.

Civilian publications

Courbil L-J, ed. *Chirurgie d'urgence en situation précaire [Emergency Surgery under Precarious Circumstances]*. Paris: Editions Pradel; 1996.

Geelhoed GW, ed. *Surgery and Healing in the Developing World*. Georgetown, TX: Landes Bioscience; 2005.

Husum H, Ang SC, Fosse E. *War Surgery: Field Manual*. Penang, Malaysia: Third World Network; 1995.

Loefler I. Africa – Surgery in an unstable environment. Weary Dunlop Memorial Lecture. *Aust NZ J Surg* 2004; **74**: 1120 – 1122.

Mahoney PF, Ryan JM, Brooks AJ, Schwab CW, eds. *Ballistic Trauma: A Practical Guide 2nd Edition*. London: Springer-Verlag; 2005.

Trauma surgery references

Boffard KD. *Manual of Definitive Surgical Trauma Care 2nd Edition*. London: International Association for Trauma Surgery and Intensive Care, Hodder/Arnold; 2007.

Botha AB, Brooks A, Loosemore T, eds. *Definitive Surgical Trauma Skills Manual*. London: Royal College of Surgeons of England; 2002.

Hirshberg A, Mattox KL. *Top Knife: The Art and Craft of Trauma Surgery*. Shrewsbury, UK: tfm Publishing Ltd; 2005.

Integrated Management on Emergency and Essential Surgical Care. *E-Learning tool Kit* [CD-ROM]. Geneva: World Health Organization; 2005.

King M, ed. *Primary Surgery, Volume Two: Trauma*. Oxford: Oxford University Press; 1987. Available at: http://www.primary-surgery.org/ps/vol2/html/index.html. Updated May 2, 2008.

Trauma.org. (Dedicated international trauma-interest web site.) [Web site http://www.trauma.org].

Chapter 1

Emergency Items Catalogue of the International Movement of the Red Cross and Red Crescent. Geneva: International Committee of the Red Cross and International Federation of Red Cross and Red Crescent Societies; 2004. [CD-ROM] [online] Available at: http://www.icrc.org/emergency-items.

Bowyer GW. War surgery and the International Committee of the Red Cross: a historical perspective. *Int J Orthop Trauma* 1996; **6**: 62 – 65.

Morris DS. Surgeons and the International Committee of the Red Cross. *Aust NZ J Surg* 1992; **62**: 170 – 172.

Mulli J-C. Activités chirurgicales en zone de guerre: l'expérience du Comité international de la Croix-Rouge [Surgical activities in a war zone: the experience of the International Committee of the Red Cross]. *Bulletin Medicus Mundi* 1995; **57**: 42 – 54.

Russbach R. Les unités chirurgicales du Comité international de la Croix-Rouge: le personnel, le matériel, les coûts [Surgical units of the International Committee of the Red Cross: the personnel, the equipment, the costs]. *Médecine et Hygiène* 1991; **49**: 2629 – 2632.

Vassallo DJ. The International Red Cross and Red Crescent Movement and lessons from its experience of war surgery. *J R Army Med Corps* 1994; **140**: 146 – 154.

Chapter 2

Basic references

Dunant H. *A Memory of Solferino*. Geneva: ICRC; 1950.

The Geneva Conventions of August 12, 1949. Geneva: ICRC; 1995.

Pictet J, ed. *Commentary on the Geneva Conventions I-IV of 12 August 1949*, 4 vol. Geneva: ICRC; 1952 – 1960.

Protocols Additional to the Geneva Conventions of 12 August 1949, revised Edition. Geneva: ICRC; 1996.

Sandoz Y, ed. *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*. Geneva: ICRC; 1987.

David E. *Principes de droit des conflits armés 4e éd [Principles of the Law of Armed Conflicts 4th Edition]*. Brussels: Bruylant; 2008.

Gasser H-P. International Humanitarian Law: An Introduction. In: Haug H, ed. *Humanity for All*. Geneva: Henry Dunant Institute; 1993.

Kolb R. *Ius in bello, Le droit international des conflits armés [Ius in Bello, The International Law of Armed Conflicts]*. Brussels: Bruylant; 2003.

Palwankar U. *Symposium on Humanitarian Action and Peace-Keeping Operations, Geneva, June 22 – 24, 1994*. Geneva: ICRC; 1994.

IHL and the medical profession

Annas GJ. Military medical ethics – physician first, last, always. *N Engl J Med* 2008; **359**: 1087 – 1090.

Baccino-Astrada A. *Manual on the Rights and Duties of Medical Personnel in Armed Conflicts*. Geneva: International Committee of the Red Cross, League of Red Cross and Red Crescent Societies; 1982. (Out of print, under revision.)

British Medical Association. *Medicine Betrayed*. London: Zed Books; 1992.

The Medical Profession and the Effects of Weapons. *Symposium: Montreux, Switzerland, 1996*. Geneva: Media Natura, Geneva Foundation to Protect Health in War, International Committee of the Red Cross; 1996.

Coupland RM. Weapons intended to blind. [Editorial]. *Lancet* 1994; **344**: 1649 – 1650.

Coupland RM. Wounds, weapons and the doctor. *Schweiz Z Milit Med* 1995; **72**: 33 – 35.

Coupland RM. Abhorrent weapons and "superfluous injury or unnecessary suffering": from field surgery to law. *BMJ* 1997; **315**: 1450 – 1452.

Coupland RM. "Non-lethal" weapons: precipitating a new arms race. Medicine must guard against its knowledge being used for weapon development. [Editorial]. *BMJ* 1997; **315**: 72.

Coupland R, Herby P. Review of the legality of weapons: a new approach The SIrUS Project. *International Review of the Red Cross* 1999; **835**: 583 – 592.

Giannou C. The Mine Information System: The principal factors determining the severity of landmine infestation. Presented at the signing of the Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-personnel Mines and on their Destruction; 1997 Sept; Ottawa, Canada.

Chapter 3

General references

Sellier KG, Kneubuehl BP. *Wound Ballistics and the Scientific Background*. Amsterdam: Elsevier; 1994.

Kneubuehl BP, Coupland RM, Rothschild MA, Thali MJ. *Wundballistik, Grundlagen und Anwendungen [Wound Ballistics, Basics and Applications]*. Berlin, Heidelberg, New York: Springer-Verlag; 2008. (Available in German only.)

Neuenschwander J, Coupland R, Kneubuehl B, Baumberger V. *Wound Ballistics: An introduction for health, legal, forensic, military and law enforcement professionals*. [Brochure and film on DVD]. Geneva: ICRC; 2008.

Journal articles

Bowyer GW, Cooper JG, Rice P. Small fragment wounds: biophysics and pathophysiology. *J Trauma* 1996; **40 (3Suppl.)**: S159 – S164.

Brismar B, Bergenwald L. The terrorist bomb explosion of Bologna, Italy, 1980: an analysis of the effects and injuries sustained. *J Trauma* 1982; **22**: 216 – 220.

Cheng XM, Liu YQ, Guo RF, Lian WK, Wang DT. Analysis of wound ballistics in 2,414 cases of battle casualties. *Journal of Trauma (China)* 1990; **6 (Suppl.)**: S169 – S172.

Cooper GJ, Ryan JM. Interaction of penetrating missiles with tissues: some common misapprehensions and implications for wound management. *Br J Surg* 1990; **77**: 606 – 610.

Coupland RM. Clinical and legal significance of fragmentation of bullets in relation to size of wounds: retrospective analysis. *BMJ* 1999; **319**: 403 – 406.

Coupland RM, Hoikka V, Sjoeklint OG, Cuenod P, Cauderay GC, Doswald-Beck L. Assessment of bullet disruption in armed conflicts. *Lancet* 1992; **339**: 35 – 37.

DePalma RG, Burris DG, Champion HR, Hodgson MJ. Blast injuries. *N Engl J Med* 2005; **352**: 1335 – 1342.

Fackler ML, Malinowski JA. The wound profile: a visual method for quantifying gunshot wound components. *J Trauma* 1985; **25**: 522 – 529.

Fackler ML, Bellamy RF, Malinowski JA. The wound profile: illustration of the missile-tissue interaction. *J Trauma* 1988; **28 (1Suppl.)**: S21 – S29.

Fackler ML. Wound ballistics: a review of common misconceptions. *JAMA* 1988; **259**: 2730 – 2736.

Farjo LA, Miclau T. Ballistics and mechanisms of tissue wounding. *Injury* 1997; **28 (3Suppl.)**: C12 – C17.

Hayda R, Harris RM, Bass CD. Blast injury research: modelling injury effects of landmines, bullets, and bombs. *Clin Orthop Relat Res* 2004; **422**: 97 – 108.

Hollerman JJ, Fackler ML, Coldwell DM, Ben-Menachem Y. Gunshot Wounds: 1. Bullets, ballistics, and mechanisms of injury *Am J Roentgenol* 1990; **155**: 685 – 690.

Houdelette P. Notions de balistique lésionnelle concernant les armes déflagrantes légères antipersonnel [A primer on wound ballistics concerning anti-personnel explosive small arms]. *Médecine et armées* 1997; **25**: 261 – 264.

Hull JB, Cooper GJ. Pattern and mechanism of traumatic amputation by explosive blast. *J Trauma* 1996; **40 (3Suppl)**: S198 – S205.

Katz E, Ofek B, Adler J, Abramowitz HB, Krausz MM. Primary blast injury after a bomb explosion in a civilian bus. *Ann Surg* 1989; **209**: 484 – 488.

Leibovici D, Gofrit ON, Stein M, Shapira SC, Noga Y, Heruti RJ, Shemer J. Blast injuries: bus versus open-air bombings: a comparative study of injuries in survivors of open-air versus confined-space explosions. *J Trauma* 1996; **41**: 1030 – 1035.

Lindsey D. The idolatry of velocity, or lies, damn lies, and ballistics. *J Trauma* 1980; **20**: 1068 – 1069.

Peters CE, Sebourn CL, Crowder HL. Wound ballistics of unstable projectiles. Part I: projectile yaw growth and retardation. *J Trauma* 1996; **40 (3Suppl)**: S10 – S15.

Peters CE, Sebourn CL. Wound ballistics of unstable projectiles. Part II: Temporary cavity formation and tissue damage. *J Trauma* 1996; **40 (3Suppl)**: S16 – S21.

Santucci RA, Chang Y-J. Ballistics for physicians: myths about wound ballistics and gunshot injuries. *J Urol* 2004; **171**: 1408 – 1414.

Sebourn CL, Peters CE. Flight dynamics of spin-stabilized projectiles and the relationship to wound ballistics. *J Trauma* 1996; **40 (3Suppl)**: S22 – S26.

Vail S. The study of wound ballistics is based on a significant amount of science and a tremendous amount of art [Defense Review Web site]. May 14, 2006. Available at: http://www.defensereview.com/modules.php?name=News&file=article&sid=875.

Volgas DA, Stannard JP, Alonso JE. Ballistics: a primer for the surgeon. *Injury* 2005; **36**: 373 – 379.

Volgas DA, Stannard JP, Alonso JE. Current orthopaedic treatment of ballistic injuries. *Injury* 2005; **36**: 380 – 386.

Wang ZG, Feng JX, Liu YQ. Pathomorphological observations of gunshot wounds. *Acta Chir Scand* 1982; **508 (Suppl.)**: S185 – S195.

Wang ZG, Tang CG, Chen XY, Shi TZ. Early pathomorphological characteristics of the wound track caused by fragments. *J Trauma* 1988; **28 (1Suppl.)**: S89 – S95.

Chapter 4

Bowyer GW, Stewart MPM, Ryan JM. Gulf war wounds: application of the Red Cross Wound Classification. *Injury* 1993; **24**: 597 – 600.

Bowyer GW. Afghan war wounded: application of the Red Cross Wound Classification. *J Trauma* 1995; **38**: 64 – 67.

Coupland RM. The Red Cross classification of war wounds: the EXCFVM scoring system. *World J Surg* 1992; **16**: 910 – 917.

Coupland RM. *The Red Cross Wound Classification, Revised Edition.* Geneva: ICRC; 2005.

Giannou CP. Penetrating missile injuries during asymmetric warfare in the 2003 Gulf conflict. [Correspondence]. *Br J Surg* 2005; **92**: 1047 – 1048.

Hinsley DE, Rosell PAE, Rowlands TK, Clasper JC. Penetrating missile injuries during asymmetric warfare in the 2003 Gulf conflict. *Br J Surg* 2005; **92**: 637 – 642.

Rosell PAE, Clasper JC. Ballistic fractures: the limited value of existing classifications. *Injury* 2005; **36**: 369 – 372.

Rowley DI. *War Wounds with Fractures: A Guide to Surgical Management.* Geneva: ICRC; 1996.

Savic J, Cernak I, Jevtic M, Todoric M. Glucose as an adjunct triage tool to the Red Cross Wound Classification. *J Trauma* 1996; **40 (3Suppl.)**: S144 – S147.

Stewart MPM, Kinninmonth A. Shotgun wounds of the limbs *Injury* 1993; **24**: 667 – 670.

Vassalo D, McAdam G. Modification to the Red Cross Wound Classification. *Injury* 1995; **26**: 131 – 132.

Chapter 5

Public health

Aboutanos MB, Baker SP. Wartime civilian injuries: epidemiology and intervention strategies. *J Trauma* 1997; **43**: 719 – 726.

Burnham G, Lafta R, Doocy S, Roberts L. Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey. *Lancet* 2006; **368**: 1421 – 1429.

Coupland RM. The effects of weapons on health. *Lancet* 1996; **347**: 450 – 451.

Coupland RM, Meddings DR. Mortality associated with use of weapons in armed conflicts, wartime atrocities, and civilian mass shootings: literature review. *BMJ* 1999; **319**: 407 – 410.

Coupland RM, Samnegaard HO. Effect of type and transfer of conventional weapons on civilian injuries: retrospective analysis of prospective data from Red Cross hospitals. *BMJ* 1999; **319**: 410 – 412.

Giannou C. Antipersonnel landmines: facts, fictions, and priorities. *BMJ* 1997; **315**: 1453 – 1454.

Guah-Sapir D, van Panhuis WG. The importance of conflict-related mortality in civilian populations. *Lancet* 2003; **361**: 2126 – 2128.

Horton R. Croatia and Bosnia: The imprints of war – 1. Consequences. *Lancet* 1999; **353**: 2139 – 2144.

Ityavyar DA, Ogba LO. Violence, conflict and health in Africa. *Soc Sci Med* 1989; **28**: 649 – 657.

Jeffries SJ. Antipersonnel mines: who are the victims? *J Accid Emerg Med* 1996; **13**: 343 – 346.

de Jong K, Mulhem M, Ford N, van der Kam S, Kleber R. The trauma of war in Sierra Leone. *Lancet* 2000; **355**: 2067 – 2068.

Lautze S, Leaning J, Raven-Roberts A, Kent R, Mazurana D. Assistance, protection, and governance networks in complex emergencies. *Lancet* 2004; **364**: 2134 – 2141.

Meddings DR. Weapons injuries during and after periods of conflict: retrospective analysis. *BMJ* 1997; **310**: 1417 – 1420.

Meddings DR, O'Connor SM. Circumstances around weapon injury in Cambodia after departure of a peacekeeping force: prospective cohort study. *BMJ* 1999; **319**: 412 – 415.

Meddings DR. Civilians and war: a review and historical overview of the involvement of non-combatant populations in conflict situations. *Med Confl Surviv* 2001; **17**: 6 – 16.

Perrin P. *War and Public Health: A Handbook*. Geneva: ICRC; 1996.

Perrin P. *H.E.L.P. – Health Emergencies in Large Populations: Public Health Course in the Management of Humanitarian Aid*. Geneva: ICRC; 2001.

Spiegel PB, Salama P. War and mortality in Kosovo, 1998 – 99: an epidemiological testimony. *Lancet* 2000; **355**: 2204 – 2209.

Taback N, Coupland R. Towards collation and modelling of the global cost of armed violence on civilians. *Med Confl Surviv* 2005; **21**: 19 – 27.

The Sphere Project: Humanitarian Charter and Minimum Standards in Disaster Response. Geneva: The Sphere Project; 2004. Available at: http://www.sphereproject.org.

Toole MJ, Galson S, Brady W. Are war and public health compatible? *Lancet* 1993; **341**: 1193 – 1196.

Udwadia TE. Surgical care for the poor: a personal Indian perspective. *Indian J Surg* 2003; **65**: 504 – 509.

Civilian war surgery literature

Bhatnagar MK, Smith GS. Trauma in the Afghan guerrilla war: effects of lack of access to care. *Surgery* 1989; **105**: 699 – 705.

Cutting PA, Agha R. Surgery in a Palestinian refugee camp. *Injury* 1992; **23**: 405 – 409.

De Wind CM. War injuries treated under primitive circumstances: experiences in an Ugandan mission hospital. *Ann R Coll Surg Engl* 1987; **69**: 193 – 195.

Dudley HAF, Knight RJ, McNeur JC, Rosengarten DS. Civilian battle casualties in South Vietnam. *Br J Surg* 1968; **55**: 332 – 340.

Fosse E, Husum H, Giannou C. The siege of Tripoli 1983: war surgery in Lebanon. *J Trauma* 1988; **28**: 660 – 663.

Fosse E, Husum H. Surgery in Afghanistan: a light model for field surgery during war. *Injury* 1992; **23**: 401 – 404.

Nassoura Z, Hajj H, Dajani O, Jabbour N, Ismail M, Tarazi T, Khoury G, Najjar F. Trauma management in a war zone: the Lebanese war experience. *J Trauma* 1991; **31**: 1596 – 1599.

Odling-Smee GW. Ibo civilian casualties in the Nigerian civil war. *BMJ* 1970; **2**: 592 – 596.

Rukovansjki M. Spinal cord injuries caused by missile weapons in the Croatian war. *J Trauma* 1996; **40 (3Suppl.)**: S189 – S192.

Suljevic I, Surkovic I. Medical aspects of the mass-scale civilian casualties at Sarajevo Markale Market on August 28, 1995: triage, resuscitation, and treatment. *Croat Med J* 2002; **43**: 209 – 212.

Military war surgery literature

Acosta JA, Hatzigeorgiou C, Smith LS. Developing a trauma registry in a forward deployed military hospital: preliminary report. *J Trauma* 2006; **61**: 256 – 260.

Bellamy RF. Combat trauma overview. In: Sajtchuk R, Grande CM, eds. *Textbook of Military Medicine, Anesthesia and Perioperative Care of the Combat Casualty*. Falls Church, VA: Office of the Surgeon General, United States Army; 1995: 1 – 42.

Bilski TR, Baker BC, Grove JR, Hinks RP, Harrison MJ, Sabra JP, Temerlin SM, Rhee P. Battlefield casualties treated at Camp Rhino, Afghanistan: lessons learned. *J Trauma* 2003; **54**: 814 – 822.

Burkle FM Jr, Newland C, Meister SJ, Blood CG. Emergency medicine in the Persian Gulf War – Part 3: battlefield casualties. *Ann Emerg Med* 1994; **23**: 755 – 760.

Carey ME. Learning from traditional combat mortality and morbidity data used in the evaluation of combat medical care. *Mil Med* 1987; **152**: 6 – 12.

Chambers LW, Green DJ, Gillingham BL, Sample K, Rhee P, Brown C, Brethauer S, Nelson T, Narine N, Baker B, Bohman HR. The experience of the US Marine Corps' Surgical Shock Trauma Platoon with 417 operative combat casualties during a 12 month period of Operation Iraqi Freedom. *J Trauma* 2006; **60**: 1155 – 1164.

Champion HR, Bellamy RF, Roberts P, Leppäniemi A. A profile of combat injury *J Trauma* 2003; **54 (5Suppl.)**: S13 – S19.

Danon YL, Nili E, Dolev E. Primary treatment of battle casualties in the Lebanon war, 1982. *Is J Med Sci* 1984; **20**: 300 – 302.

Eastridge BJ, Jenkins D, Flaherty S, Schiller H, Holcomb JB. Trauma system development in a theater of war: experiences from Operation Iraqi Freedom and Operation Enduring Freedom. *J Trauma* 2006; **61**: 1366 – 1373.

Fekadu T. *Mass Casualty Management Under Unique War Situation: Inside Eritrea's War for Independance*. Asmara, Eritrea: self published; 2002.

Garfield RM, Neugut AI. Epidemiologic analysis of warfare. *JAMA* 1991; **266**: 688 – 692.

Gofrit ON, Kovalski N, Leibovici D, Shemer J, O'Hana A, Shapira SC. Accurate anatomical location of war injuries: analysis of the Lebanon war fatal casualties and the proposition of new principles for the design of military personal armour system. *Injury* 1996; **27**: 577 – 581.

Gofrit ON, Leibovici D, Shapira SC, Shemer J, Stein M, Michaelson M. The trimodal death distribution of trauma victims: military experience from the Lebanon war. *Mil Med* 1997; **162**: 24 – 26.

Grau LW, Jorgensen WA. Handling the wounded in a counter-guerrilla war: the Soviet/Russian experience in Afghanistan and Chechnya. *U.S. Army Medical Dept Journal* 1998; **Jan/Feb**: 2 – 10.

Hardaway RM III. Viet Nam wound analysis. *J Trauma* 1978; **18**: 635 – 643.

Holcomb JB. Current perspective on combat casualty care. The 2004 Fitts Lecture. *J Trauma* 2005; **59**: 990 – 1002.

Holcomb JB, Stansbury LG, Champion HR, Wade C, Bellamy RF. Understanding combat casualty care statistics. *J Trauma* 2006; **60**: 397 – 401.

Jackson DS, Batty CG, Ryan JM, McGregor WSP. The Falklands war: army field surgical experience. *Ann R Coll Surg Engl* 1983; **65**: 281 – 285.

Jevtic M, Petrovic M, Ignjatovic D, Ilijevski N, Misovic S, Kronja G, Stankovic N. Treatment of wounded in the combat zone. *J Trauma* 1996; **40 (3Suppl.)**: S173 – S176.

Johnson DE, Panijayanond P, Lumjiak S, Crum JW, Boonkrapu P. Epidemiology of combat casualties in Thailand. *J Trauma* 1981; **21**: 486 – 488.

Jones EL, Peters AF, Gasior RM. Early management of battle casualties in Vietnam. *Arch Surg* 1968; **97**: 1 – 15.

Mabry RL, Holcomb JB, Baker AM, Cloonan CC, Uhorchak JM, Perkins DE, Canfield AJ, Hagmann JH. United States Army Rangers in Somalia: an analysis of combat casualties on an urban battlefield. *J Trauma* 2000; **49**: 515 – 529.

Spalding TJW, Stewart MPM, Tulloch DN, Stephens KM. Penetrating missile injuries in the Gulf war 1991. *Br J Surg* 1991; **78**: 1102 – 1104.

Van Rooyen MJ, Sloan EP, Radvany AE, Peric T, Kulis B, Tabak P. The incidence and outcome of penetrating and blunt trauma in central Bosnia: the Nova Bila Hospital for war wounded. *J Trauma* 1995; **38**: 863 – 866.

Versier G, Le Marec C, Rouffi J. Quatre ans de chirurgie de guerre au GMC de Sarajevo (juillet 1992 à août 1996) [Four years of war surgery at the French surgical facility in Sarajevo (July 1992 – August 1996)]. *Médecine et armées* 1998; **26**: 213 – 218.

Vojvodic V. Management of war casualties in the Military Medical Academy (Belgrade) during combat operations in 1991/1992: an overview. *J Trauma* 1996; **40 (3Suppl.)**: S180 – S182.

ICRC war surgery articles

Bowyer GW. Management of small fragment wounds: experience from the Afghan border. *J Trauma* 1996; **40 (3Suppl.)**: S170 – S172.

Coupland RM. Hand grenade injuries among civilians. *JAMA* 1993; **270**: 624 – 626.

Coupland RM. Epidemiological approach to surgical management of the casualties of war. *BMJ* 1994; **308**: 1693 – 1696.

Kjaergaard J. Les blessés de guerre de l'hôpital de campagne du CICR à Beyrouth en 1976 [War wounded in the ICRC field hospital in Beirut 1976]. *Schweiz Z Milit Med* 1978; **55**: 1 – 23.

Korver AJH. Outcome of war-injured patients treated at first aid posts of the International Committee of the Red Cross. *Injury* 1994; **25**: 25 – 30.

Morris D, Sugrue W, McKenzie E. On the border of Afghanistan with the International Committee of the Red Cross. *NZ Med J* 1985; **98**: 750 – 752.

Scott-Findlay J, Smith FS. A Timor experience in war and civilian trauma. *Med J Aust* 1976; **2**: 90 – 92.

Trouwborst A, Weber BK, Dufour D. Medical statistics of battlefield casualties. *Injury* 1987; **18**: 96 – 99.

Chapter 6

Hayward-Karlsson J, Jeffery S, Kerr A, Schmidt H. *Hospitals for War-Wounded: A Practical Guide for Setting up and Running a Surgical Hospital in an Area of Armed Conflict*. Geneva: ICRC; 1998.

Hayward-Karlsson J. Hospital and System Assessment. In: Mahoney PF, Ryan JM, Brooks AJ, Schwab CW, eds. *Ballistic Trauma: A Practical Guide 2nd Edition*. London: Springer-Verlag; 2005: 513 – 526.

Pruit BA. Combat casualty care and surgical progress. *Ann Surg* 2006; **243**: 715 – 729.

Chapter 7

Butler FK Jr. Tactical combat casualty care: combining good medicine with good tactics. *J Trauma* 2003; **54 (Suppl.)**: S2 – S3.

Coupland RM. Epidemiological approach to surgical management of the casualties of war. *BMJ* 1994; **308**: 1693 – 1696.

Coupland RM, Molde Å, Navein J. *Care in the Field for Victims of Weapons of War: A Report from the Workshop Organized by the ICRC on Pre-Hospital Care for War and Mine-Injured*. Geneva: ICRC; 2001.

Dubick MA, Atkins JL. Small-volume fluid resuscitation for the far-forward combat environment: current concepts. *J Trauma* 2003; **54 (Suppl.)**: S43 – S45.

Giannou C, Bernes E. *First Aid in Armed Conflicts and Other Situations of Violence*. Geneva: ICRC; 2006.

Guidelines on First Aid and HIV/AIDS. Geneva: International Committee of the Red Cross /International Federation of Red Cross and Red Crescent Societies; 2001.

Husum H. Effects of early prehospital life support to war injured: the battle of Jalalabad, Afghanistan. *Prehosp Disaster Med* 1999; **14**: 75 – 80.

Husum H, Gilbert M, Twisborg T. *Save Lives, Save Limbs*. Tromsoe, Norway: Third World Network, Tromsoe Mine Victim Resource Centre; 2000.

Husum H, Gilbert M, Wisborg T, Heng YV, Murad M. Rural prehospital trauma systems improve trauma outcome in low-income countries: a prospective study from North Iraq and Cambodia. *J Trauma* 2003; **54**: 1188 – 1196.

Korver AJH. Outcome of war-injured patients treated at first aid posts of the International Committee of the Red Cross. *Injury* 1994; **25**: 25 – 30.

Mabry R, McManus JG. Prehospital advances in the management of severe penetrating trauma. *Crit Care Med* 2008; **36 (Suppl.)**: S258 – S266.

Roberts DL. *Staying Alive: Safety and Security Guidelines for Humanitarian Volunteers in Conflict Areas 2nd Edition*. Geneva: ICRC; 2006.

Royal Defence Medical College and the Royal Centre for Defence Medicine. *Battlefield Advanced Trauma Life Support (Incorporating Battlefield Resuscitation Techniques and Skills)*. London: British Armed Forces, UK Minister of Defence; 2003.

Wisborg T, Murad MK, Edvardsen O, Husum H. Prehospital trauma system in a low-income country: system maturation and adaptation during 8 years. *J Trauma* 2008; **64**: 1342 – 1348.

Cervical spine immobilization

Arishita GI, Vayer JS, Bellamy RF. Cervical spine immobilization of penetrating neck wounds in a hostile environment. *J Trauma* 1989; **29**: 332 – 337.

Barkana Y, Stein M, Scope A, Maor R, Abramovich Y, Friedman Z, Knoller N. Prehospital stabilization of the cervical spine for penetrating injuries of the neck – is it necessary? *Injury* 2000; **31**: 305 – 309.

Rhee P, Kuncir EJ, Johnson L, Brown C, Velmahos G, Martin M, Wang D, Salim A, Doucet J, Kennedy S, Demetriades D. Cervical spine injury is highly dependent on the mechanism of injury following blunt and penetrating assault. *J Trauma* 2006; **61**: 1166 – 1170.

Tourniquet use

Brodie S, Hodgetts TJ, Ollerton J, McLeod J, Lambert P, Mahoney P. Tourniquet use in combat trauma. *J R Army Med Corps* 2008; **153**: 310 – 313.

Lakstein D, Blumenfeld A, Sokolov T, Lin G, Bssorai R, Lynn M, Ben-Abraham R. Tourniquets for hemorrhage control on the battlefield: a 4-year accumulated experience. *J Trauma* 2003; **54 (Suppl.)**: S221 – S225.

Navein J, Coupland R, Dunn R. The tourniquet controversy. *J Trauma* 2003; **54 (Suppl.)**: S219 – S220.

Welling DR, Burris DG, Hutton JE, Minken SL, Rich NM. A balanced approach to tourniquet use: lessons learned and relearned. *J Am Coll Surg* 2006; **203**: 106 – 115.

Chapter 8

Trauma training

Driscoll P, Skinner D, Earlam R, eds. *ABC of Major Trauma 3rd Edition*. London: BMJ Books; 2001.

American College of Surgeons, Committee on Trauma. *Advanced Trauma Life Support for Doctors, Student Course Manual (ATLS) 7th Edition*. Chicago, IL: American College of Surgeons; 2004.

Kortbeek JB, Al Turki SA, Ali J, et al. Advanced Trauma Life Support, 8th Edition. The evidence for change. *J Trauma* 2008; **64**: 1638 – 1650.

International Association for Trauma Surgery and Intensive Care (IATSIC). *Definitive Surgical Trauma Care Course (DSTC)*. Various country and regional sessions.

Trauma training in low-income countries: some examples

Aboutanos MB, Rodas EB, Aboutanos SZ, Mora FE, Wolf LG, Duane TM, Malhotra AK, Ivatury RR. Trauma education and care in the jungle of Ecuador, where there is no advanced trauma life support. *J Trauma* 2007; **62**: 714 – 719.

Basic Emergency Skills in Trauma (B.E.S.T.). Manila: Philippine College of Surgeons.

National Trauma Management Course (NTMC). Academy of Traumatology of India and Trauma Unit, Johannesburg Hospital and Faculty of Health Sciences, University of Witwatersrand. Available at: http://www.indiatrauma.org.

Emergency Room Trauma Course (ERTC). Geneva: ICRC.

Pneumothorax

Leigh-Smith S, Harris T. Tension pneumothorax – time for a re-think? *Emerg Med J* 2005; **22**: 8 – 16.

Shock resuscitation

Alam HB, Koustova E, Rhee P. Combat casualty care research: from bench to the battlefield. *World J Surg* 2005; **29 (Suppl.)**: S7 – S11.

Bickell WH, Wall MJ Jr, Pepe PE, Martin RR, Ginger VF, Allen MK, Mattox KL. Immediate versus delayed fluid resuscitation for hypotensive patients with penetrating torso injuries. *N Engl J Med* 1994; **331**: 1105 – 1109.

Champion HR. Combat fluid resuscitation: introduction and overview of conferences. *J Trauma* 2003; **54 (Suppl.)**: S7 – S12.

Moore FA, McKinley BA, Moore EE. The next generation in shock resuscitation. *Lancet* 2004; **363**: 1988 – 1996.

Rhee P, Alam HB, Ling GSF. Hemorrhagic Shock and Resuscitation. In: Tsokos GC, Atkins JL, eds. *Combat Medicine: Basic and Clinical Research in Military, Trauma, and Emergency Medicine*. Totowa, NJ: Humana Press; 2003: 177 – 218.

Rhee P, Koustova E, Alam HB. Searching for the optimal resuscitation method: recommendations for the initial fluid resuscitation of combat casualties. *J Trauma* 2003; **54 (Suppl.)**: S52 – S62.

Rushing GD, Britt LD. Reperfusion injury after hemorrhage. *Ann Surg* 2008; **247**: 929 – 937.

Shoemaker WC, Peitzman AB, Bellamy R, Bellomo R, Bruttig SP, Capone A, Dubick M, Kramer GC, McKenzie JE, Pepe PE, Safar P, Schlichtig R, Severinghaus JW, Tisherman SA, Wiklund L. Resuscitation from severe hemorrhage. [Symposium article] *Crit Care Med* 1996; **24 (Suppl.)**: S12 – S23.

Blood transfusion

Eshaya-Chauvin B, Coupland RM. Transfusion requirements for the management of war injured: the experience of the International Committee of the Red Cross. *Br J Anaesth* 1992; **68**: 221 – 223.

Kiebooms L, Rouvillois A, Jones T. *ICRC Blood Transfusion Guidelines*. Geneva: ICRC; 2004.

Kauvar DS, Holcomb JB, Norris GC, Hess JR. Fresh whole blood transfusion: a controversial military practice. *J Trauma* 2006; **61**: 181 – 184.

Key NS, Negrier C. Transfusion medicine 3: Coagulation factor concentrates: past, present, and future. *Lancet* 2007; **370**: 439 – 448.

Klein HG, Spahn DR, Carson JL. Transfusion medicine 1: Red blood cell transfusion in clinical practice. *Lancet* 2007; **370**: 415 – 426.

Repine TB, Perkins JG, Kauvar DS, Blackborne L. The use of fresh whole blood in massive transfusion. *J Trauma* 2006; **60 (6Suppl.)**: S59 – S69.

Spinella PC, Perkins JG, Grathwohl KW, Repine T, Beekley AC, Sebesta J, Jenkins D, Azarow K, Holcomb JB. Fresh whole blood transfusions in coalition military, foreign national, and enemy combatant patients during Operation Iraqi Freedom at a U.S. combat support hospital. *World J Surg* 2008; **32**: 2 – 6.

Stroncek DF, Rebulla P. Transfusion medicine 2: Platelet transfusions. *Lancet* 2007; **370**: 427 – 438.

World Health Organization. *WHO Guidelines for Blood Transfusion* [WHO Web site]. Available at: http://www.who.int/bloodsafety/clinical_use/en/Handbook_EN.pdf and http://www.who.int/bloodsafety/clinical_use/en/WHO_BLS_98.2_EN.pdf.

Chapter 9

Further practical information for the organisation of hospital management in a triage situation is available in: Hayward-Karlsson J, Jeffrey S, Kerr A, Schmidt H. *Hospitals for War Wounded*. Geneva: ICRC; 1998.

Please note:

In a military context, all standard war surgery manuals written by and for armed forces deal with the organization and implementation of triage under military constraints.

Almogy G, Belzberg H, Mintz Y, Pikarsky AK, Zamir G, Rivkind AI. Suicide bombing attacks: update and modifications to the protocol. *Ann Surg* 2004; **239**: 295 – 303.

Barbera JA, Macintyre AG. *Janes's Mass Casualty Handbook: Hospital Emergency Preparedness and Response*. Coulsdon, Surrey, UK: Jane's Information Group; 2003.

Burkle FM, Orebaugh S, Barendse BR. Emergency medicine in the Persian Gulf war – part 1: preparations for triage and combat casualty care. *Ann Emerg Med* 1994; **23**: 742 – 747.

Coupland RM, Parker PJ, Gray RC. Triage of war wounded: the experience of the International Committee of the Red Cross. *Injury* 1992; **23**: 507 – 510.

Frykberg ER. Medical management of disasters and mass casualties from terrorist bombings: how can we cope? *J Trauma* 2002; **53**: 201 – 212.

Frykberg ER. Principles of mass casualty management following terrorist disasters. [Editorial]. *Ann Surg* 2004; **239**: 319 – 321.

Gertsch P. Assessment of hospital workload in war surgery. *Br J Surg* 1987; **74**: 831 – 833.

Gray RC. War surgery and triage. *Postgrad Doctor Mid East*. 1993; **16**: 150 – 157.

Hirshberg A, Scott BG, Granchi T, Wall MJ Jr, Mattox KL, Stein M. How does casualty load affect trauma care in urban bombing incidents? A quantitative analysis. *J Trauma* 2005; **58**: 686 – 695.

Hogan DE, Lairet JR: Triage. In: Hogan DE, Burstein JL, eds. *Disaster Medicine 2nd Edition*. Philadephia, PA: Lippincott Williams & Wilkins; 2007: 12 – 28.

Iserson KV, Moskop JC: Triage in medicine, part I: concept, history, and types. *Ann Emerg Med* 2007; **49**: 275 – 281.

Iserson KV, Moskop JC: Triage in medicine, part II: underlying values and principles. *Ann Emerg Med* 2007; **49**: 282 – 287.

Kennedy K, Aghababian RV, Gans L, Lewis CP. Triage: techniques and applications in decision making. *Ann Emerg Med* 1996; **28**: 136 – 144.

Rignault D, Wherry D. Lessons from the past worth remembering: Larrey and triage. *J Trauma* 1999; **1**: 86 – 89.

Rochat C-H. Le triage chirurgical en conditions de guerre (Kaboul 1989) [Surgical triage under war conditions (Kabul 1989)]. *Médicine Militaire* 1992; **69**: 35 – 38.

Ryan JM, Sibson J, Howell G. Assessing injury severity during general war: will the military triage system meet future needs? *J R Army Med Corps* 1990; **136**: 27 – 35.

Suljevic I, Surkovic I. Medical aspects of the mass-scale civilian casualties at Sarajevo Markale Market on August 28, 1995: triage, resuscitation, and treatment. *Croat Med J* 2002; **43**: 209 – 212.

Torkki M, Koljonen V, Sillanpää K, Tukiainen E, Pyörälä S, Kemppainen E, Kalske J, Arajärvi E, Keränen U, Hirvensalo E. Triage in a bomb disaster with 166 casualties. *Eur J Trauma* 2006; **32**: 374 – 380.

World Medical Association. Statement on Medical Ethics in the Event of Disasters. Adopted by the 46th WMA General Assembly,1994 Sept.; Stockholm: Sweden. Available at: http://www.wma.net/e/policy/d7.htm.

Chapters 10 and 11

Anglen JO, Gainor BJ, Simpson WA, Christensen G. The use of detergent irrigation for musculoskeletal wounds. *Int Orthop* 2003; **27**: 40 – 46.

Baldan M, Giannou CP. Basic surgical management of war wounds: the ICRC experience. *East Cent Afr J Surg* 2003; **8**: 35 – 38.

Bewes P. The Management of Wounds in Developing Countries. Health Development e-TALC [CD – ROM]. January, 2004. No. 5. [TALC: Teaching-aids At Low Cost Web site]. Available at: http://www.talcuk.org.

Bowyer GW, Cooper GJ, Rice P. Small fragment wounds: biophysics and pathophysiology. *J Trauma* 1996; **40 (3Suppl.)**: S159 – S164.

Bowyer GW. Management of small fragment wounds in modern warfare: a return to Hunterian principles. *Ann R Coll Surg Engl* 1997; **79**: 175 – 182.

Coupland RM. Technical aspects of war wound excision. *Br J Surg* 1989; **76**: 663 – 667.

Fackler ML, Breteau JPL, Courbil LJ, Taxit R, Glas J, Fievet JP. Open wound drainage versus wound excision in treating the modern assault rifle wound. *Surgery* 1989; **105**: 576 – 584.

Trunkey DD. Comments on the article by Fackler et al. [Editorial] *Surgery* 1989; **105**: 693 – 694.

Gray RC. Surgery of war and disaster. *Trop Doct* 1991; **21 (Suppl.)**: S56 – S60.

Gray RC. *War Wounds: Basic Surgical Management*. Geneva: ICRC; 1994.

Hamer ML, Robson MC, Krizek TJ, Southwick WO. Quantitative bacterial analysis of comparative wound irrigations. *Ann Surg* 1975; **181**: 819 – 822.

Molan PC. The evidence supporting the use of honey as a wound dressing. *Int J Low Extrem Wounds* 2006; **5**: 40 – 54.

Molde Å. Victims of war: surgical principles must not be forgotten (again)! *Acta Orthop Scand* 1998; **281 (Suppl.)**: 54 – 57.

Rautio J, Paavolainen P. Delayed treatment of complicated fractures in war wounded. *Injury* 1987; **18**: 238 – 240.

Rautio J, Paavolainen P. Afghan war wounded: experience with 200 cases. *J Trauma* 1988; **28**: 523 – 525.

Rochat C-H, Graber P, Ursprung T. Traitement des plaies par projectiles en conditions de guerre (Afghanistan). Suture primaire différée ou suture secondaire? [Treatment of projectile wounds under conditions of war (Afghanistan). Delayed primary suture or secondary suture?]. *Médecine Militaire* 1986; **1**: 20 – 22.

Rowley DI. The management of war wounds involving bone. *J Bone Joint Surg Br* 1996; **78**: 706 – 709.

Strada G, Coupland RM, Gray RC. Surgery for the victims of war: the experience of the International Committee of the Red Cross. *J Emerg Surg* 1991; **14**: 126 – 130.

Svoboda SJ, Owens BD, Gooden HA, Melvin ML, Baer DG, Wenke JC. Irrigation with potable water versus normal saline in a contaminated musculoskeletal wound model. *J Trauma* 2008; **64**: 1357 – 1359.

Verbeke JH. Initial treatment of war casualties in a field hospital. *Acta Anaesth Belg* 1987; **38**: 261 – 265.

Vermeulen H, Ubbink DT, Goossens A, de Vos R, Legemate DA. Systematic review of dressings and topical agents for surgical wounds healing by secondary intention. *Br J Surg* 2005; **92**: 665 – 672.

Chapter 12

Bhaskar SN, Cutright DE, Hunsuck EE, Gross A. Pulsating water jet devices in debridement of combat wounds. *Mil Med* 1971; **136**: 264 – 266.

Bhatnagar MK, Smith GS. Trauma in the Afghan guerrilla war: effects of lack of access to care. *Surgery* 1989; **105**: 699 – 705.

Coupland RM, Howell P. An experience of war surgery and wounds presenting after 3 days on the border of Afghanistan. *Injury* 1988; **19**: 259 – 262.

Craig G. Treating the Afghan war wounded. *J Roy Soc Med* 1993; **86**: 404 – 405.

Gross A, Cutright DE, Bhaskar SN. Effectiveness of pulsating water jet lavage in treatment of contaminated crushed wounds. *Am J Surg* 1972; **124**: 373 – 377.

Mellor SG, Cooper GJ, Bowyer GW. Efficacy of delayed administration of benzylpenicillin in the control of infection in penetrating soft tissue injuries in war. *J Trauma* 1996; **40 (3Suppl.)**: S128 – S134.

Morris DS, Sugrue WJ, McKenzie E. On the border of Afghanistan with the International Committee of the Red Cross. *NZ Med J* 1985; **98**: 750 – 752.

Rowley DI. War surgery in an African conflict. *Scott Med J* 1997; **42**: 163 – 164.

Biofilm

Evans LV, ed. *Biofilms: Recent Advances in their Study and Control*. Amsterdam: Harwood Academic Press; 2000.

Fletcher M, ed. *Bacterial Adhesion: Molecular and Ecological Diversity*. New York: John Wiley & Sons; 1996.

Monroe D. Looking for chinks in the armor of bacterial biofilms [Public Library of Science Web site]. Available at: http://www.plos.org. *PLoS Bio*. 2007; **5**: e307.

Wilson M, ed. *Bacterial Adhesion to Host Tissues: Mechanisms and Consequences*. Cambridge: Cambridge University Press; 2002.

Chapter 13

Dahlgren B, Berlin R, Brandberg A, Rybeck B, Seeman T. Bacteriological findings in the first 12 hours following experimental missile trauma. *Acta Chir Scand* 1981; **147**: 513 – 518.

Dahlgren B, Berlin R, Brandberg A, Rybeck B, Schantz B, Seeman T. Effect of benzylpenicillin on wound infection rate and on the extent of devitalized tissue twelve hours after infliction of experimental missile trauma. *Acta Chir Scand* 1982; **148**: 107 – 112.

Fleming A. On the bacteriology of septic wounds. *Lancet* 1915; **186**: 638 – 643.

Lindberg RB, Wetzler TF, Marshall JD, Newton A, Strawitz JG, Howard JM. The bacterial flora of battle wounds at the time of primary debridement. *Ann Surg* 1955; **141**: 369 – 374.

Mellor SG, Cooper GJ, Bowyer GW. Efficacy of delayed administration of benzylpenicillin in the control of infection in penetrating soft tissue injuries in war. *J Trauma* 1996; **40 (3Suppl.)**: S128 – S134.

Miclau T, Farjo LA. The antibiotic treatment of gunshot wounds. *Injury* 1997; **28 (3Suppl.)**: C1 – C5.

Munoz-Price LS, Weinstein RA. Acinetobacter infection. *N Engl J Med* 2008; **358**: 1271 – 1281.

Murray CK, Roop SA, Hospenthal DR, Dooley DP, Wenner K, Hammock J, Taufen N, Gourdine E. Bacteriology of war wounds at the time of injury. *Mil Med* 2006; **171**: 826 – 829.

Murray CK, Hospenthal DR, eds. Prevention and management of combat-related infections: clinical practice guidelines consensus conference. *J Trauma* 2008; **64 (3Suppl.)**: S207 – S286.

Petersen K, Riddle MS, Danko JR, Blazes DL, Hayden R, Tasker SA, Dunne JR. Trauma-related infections in battlefield casualties from Iraq. *Ann Surg* 2007; **245**: 803 – 811.

Polhemus ME, Kester KE. Infections. In: Tsokos GC, Atkins JL, eds. *Combat Medicine: Basic and Clinical Research in Military, Trauma, and Emergency Medicine*. Totowa, NJ: Humana Press; 2003: 149 – 173.

Rubin RH. Surgical wound infection: epidemiology, pathogenesis, diagnosis and management. *BMC Infect Dis* 2006; **6**: 171. Available at: http://www.biomedcentral.com/1471-2334/6/171.

Sebeny PJ, Riddle MS, Petersen K. *Acinetobacter baumannii* skin and soft-tissue infection associated with war trauma. *Clin Infect Dis* 2008; **47**: 444 – 449.

Tian HM, Deng G, Huang MJ, Tian F, Suang G, Liu YQ. Quantitative bacteriological study of the wound track. *J. Trauma* 1988; **28 (Suppl.)**: S215 – S216.

Tian HM, Huang MJ, Liu YQ, Wang ZG. Primary bacterial contamination of wound track. *Acta Chir Scand* 1982; **508 (Suppl.)**: S265 – S269.

Simchen E, Sacks T. Infection in war wounds: experience during the 1973 October war in Israel. *Ann Surg* 1975; **182**: 754 –761.

Tetanus

Oladiran I, Meier DE, Ojelade AA, OlaOlorun DA, Adeniran A, Tarpley JL. Tetanus: continuing problem in the developing world. *World J Surg* 2002; **26**: 1282 – 1285.

Thwaites CL, Yen LM, Loan HT, Thuy TTD, Thwaites GE, Stepniewska K, Soni N, White NJ, Farrar JJ. Magnesium sulphate for treatment of severe tetanus: a randomised controlled trial. *Lancet* 2006; **368**: 1436 – 1443.

Necrotising fasciitis

Angoules AG, Kontakis G, Drakoulakis E, Vrentzos G, Granick MS, Giannoudis PV. Necrotising fasciitis of upper and lower limb: a systematic review. *Injury* 2007; **38 (Suppl.)**: C18 – C25.

Hasham S, Matteucci P, Stanley PRW, Hart NB. Necrotising fasciitis: clinical review. *BMJ* 2005; **330**; 830 – 833.

Chapter 14

Baldan M, Giannou CP, Sasin V, Morino GF. Metallic foreign bodies after war injuries: should we remove them? The ICRC experience. *East C Afr J Surg* 2004; **9**: 31 – 34.

Linden MA, Manton WI, Stewart RM, Thal ER, Feit H. Lead poisoning from retained bullets: pathogenesis, diagnosis, and management. *Ann Surg* 1982; **195**: 305 – 313.

Rhee JM, Martin R. The management of retained bullets in the limbs. *Injury* 1997; **28 (3Suppl.)**: C23 – C28.

Rich NM, Collins GJ, Andersen CA, McDonald PT, Kozloff L, Ricotta JJ. Missile emboli. *J Trauma* 1978; **18**: 236 – 239.

Chapter 15

Arturson G: Pathophysiology of the burn wound and pharmacological treatment. The Rudi Hermans Lecture, 1995. *Burns* 1996; **22**: 255 – 274.

Cartotto R, Musgrave M, Beveridge M, Fish J, Gomez M. Minimizing blood loss in burn surgery. *J Trauma* 2000; **49**: 1034 – 1039.

Hettiaratchy S, Dziewulski P: ABC of burns: pathophysiology and types of burns. *BMJ* 2004; **328**: 1427 – 1429.

Lindahl OA, Zdolsekz J, Sjöberg F, Ängquist K-A. Human postburn oedema measured with the impression method. *Burns* 1993; **19**: 479 – 484.

Lund T, Onarheim H, Reed RK. Pathogenesis of edema formation in burn injuries. *World J Surg* 1992; **16**: 2 – 9.

Pruit BA Jr. Fluid and electrolyte replacement in the burned patient. *Surg Clin N Am* 1978; **48**: 1291 – 1312.

Sheridan RL. Burns. *Crit Care Med* 2002; **30 (Suppl.)**: S500 – S514.

Thomas SJ, Kramer GC, Herndon DN. Burns: military options and tactical solutions. *J Trauma* 2003; **54 (5Suppl.)**: S207 – S218.

Zdolsek HJ, Lindahl OA, Ängquist K-A, Sjöberg F. Non-invasive assessment of intercompartmental fluids in burn victims. *Burns* 1998; **24**: 233 – 240.

Chapter 16

Britt LD, Dascombe WH, Rodriguez A. New horizons in management of hypothermia and frostbite injury. *Surg Clin North Am* 1991; **71**: 345 – 370.

Chapter 17

Bion JF. An anaesthetist in a camp for Cambodian refugees. *Anaesthesia* 1983; **38**: 798 – 801.

Bion JF. Infusion analgesia for acute war injuries: a comparison of pentazocine and ketamine. *Br J Acc Surg* 1984; **39**: 560 – 564.

Eshaya-Chauvin B, Nyffenegger E. Anesthésie pour blessés de guerre: étude rétrospective [Anaesthesia for war-wounded: retrospective study]. *Revue Médicale de la Suisse Romande*. 1990; **110**: 429 – 432.

Husum H, Heger T, Sundet M. Postinjury malaria: a study of trauma victims in Cambodia. *J Trauma* 2002; **52**: 259 – 266.

King M, ed. *Primary Anaesthesia*. Oxford: Oxford University Press; 1986.

Korver AJH. Relation between fever and outcome in injured victims of an internal armed conflict: the experience in a war surgery hospital of the International Committee of the Red Cross. *Milit Med* 1996; **161**: 658 – 660.

Leppäniemi AK. Where there is no anaesthetist. *Br J Surg* 1991; **78**: 245 – 246.

Pesonen P. Pulse oximetry during ketamine anaesthesia in war conditions. *Can J Anaesth* 1991; **38**: 592 – 594.

Vreede E, Lasalle X, Rosseel P. *Field Anaesthesia: Basic Practice: A Guide for Anaesthetists*. Paris: Médecins sans Frontières; 2001.

Ketamine anaesthesia in head trauma

Bourgoin A, Albanese J, Wereszczynski N, Charbit M, Vialet R, Martin C. Safety of sedation with ketamine in severe head injury patients: comparison with sulfentanyl. *Crit Care Med* 2003; **31**: 711 – 717.

Gofrit ON Leibovici D, Shemer J, Henig A, Shapira SC. Ketamine in the field: the use of ketamine for induction of anesthesia before intubation of injured patients in the field. *Injury* 1997; **28**: 41 – 43.

Green SM, Clem KJ, Rothrock SG. Ketamine safety profile in the developing world: survey of practitioners. *Acad Emerg Med* 1996; **3**: 598 – 604.

Himmelseher S, Durieux ME. Revising a dogma: ketamine for patients with neurological injury? *Anesth Analg* 2005; **101**: 524 – 534.

Ketcham DW. Where there is no anaesthesiologist: the many uses of ketamine. *Trop Doct* 1990; **20**: 163 – 166.

Sehdev RS, Symmons DAD, Kindl K. Ketamine for rapid sequence induction in patients with head injury in the emergency department. *Emerg Med Austr* 2006; **18**: 37 – 44.

Tighe SQM, Rudland S. Anesthesia in northern Iraq: an audit from a field hospital. *Mil Med* 1994; **159**: 86 – 90.

Trouwborst A, Weber BK, Dufour D. Medical statistics of battlefield casualties. *Injury* 1987; **18**: 96 – 99.

Chapter 18

Damage control surgery

Burch JM, Ortiz VB, Richardson RJ, Martin RR, Mattox KL, Jordan GL Jr. Abbreviated laparotomy and planned reoperation for critically injured patients. *Ann Surg* 1992; **215**: 476 – 483.

Damage Control Surgery. *Surg Clin North Am* 1997; **77**: 753 – 952.

Hirshberg A, Mattox KL. Planned reoperation for severe trauma. *Ann Surg* 1995; **222**: 3 – 8.

Kashuk JL, Moore EE, Millikan JS, Moore JB. Major abdominal vascular trauma: a unified approach. *J Trauma* 1982; **22**: 672 – 679.

Moore EE. Staged laparotomy for the hypothermia, acidosis, coagulopathy syndrome. *Am J Surg* 1996; **172**: 405 – 410.

Parker PJ. Damage control surgery and casualty evacuation: techniques for surgeons, lessons for military medical planners. *J R Army Med Corps* 2006; **152**: 202 – 211.

Rotondo MF, Schwab CW, McGonigal MD, Phillips GR, Fruchterman TM, Kauder DR, Latenser BA, Angood PB: "Damage Control": an approach for improved survival with exsanguinating penetrating abdominal injury *J Trauma* 1993; **35**: 375 – 382.

Sharpiro MB, Jenkins DH, Schwab CW, Rotondo MF. Damage control: collective review. *J Trauma* 2000; **49**: 969 – 978.

Stone HH, Strom PR, Mullins RJ. Management of the major coagulopathy with onset during laparotomy. *Ann Surg* 1983; **197**: 532 – 535.

Hypothermia and coagulopathy

Bernabei AF, Levison MA, Bender JS. The effects of hypothermia and injury severity on blood loss during trauma laparotomy. *J Trauma* 1992; **33**: 835 – 839.

Brohi K, Singh J, Heron M, Coats T. Acute traumatic coagulopathy. *J Trauma* 2003; **54**: 1127 – 1130.

Brohi K, Cohen MJ, Ganter MT, Matthay MA, Mackersie RC, Pittet J-F. Acute traumatic coagulopathy: initiated by hypoperfusion: modulated through the protein C pathway? *Ann Surg* 2007; **245**: 812 – 818.

Brohi K, Cohen MJ, Ganter MT, Schultz MJ, Levi M, Mackersie RC, Pittet J-F. Acute coagulopathy of trauma: hypoperfusion induces systemic anticoagulation and hyperfibrinolysis. *J Trauma* 2008; **64**: 1211 – 1217.

Cosgriff N, Moore EE, Sauaia A, Kenny-Moynihan M, Burch JM, Galloway B. Predicting life-threatening coagulopathy in the massively transfused trauma patient: hypothermia and acidosis revisited. *J Trauma* 1997; **42**: 857 – 862.

Gentilello LM, Jurkovich GJ, Stark MS, Hassantash SA, O'Keefe GE. Is hypothermia in the victim of major trauma protective or harmful? A randomized, prospective study. *Ann Surg* 1997; **226**: 439 – 449.

Gregory JS, Flanebaum L, Townsend MC, Cloutier CT, Jonasson O. Incidence and timing of hypothermia in trauma patients undergoing operations. *J Trauma* 1991; **31**: 795 – 800.

Grosso SM, Keenan JO. Whole blood transfusion for exsanguinating coagulopathy in a U.S. field surgical hospital in postwar Kosovo. *J Trauma* 2000; **49**: 145 – 148.

Gubler KD, Gentilello LM, Hassantash SA, Maier RV. The impact of hypothermia on dilutional coagulopathy. *J Trauma* 1994; **36**: 847 – 851.

Hess JR, Lawson JH. The coagulopathy of trauma versus disseminated intravascular coagulation. *J Trauma* 2006; **60 (6Suppl.)**: S12 – S19.

Holcomb JB, Jenkins D, Rhee P, Johannigman J, Mahoney P, Mehta S, Cox ED, Gehrke MJ, Beilman GJ, Schreiber M, Flaherty SF, Grathwohl KW, Spinella PC, Perkins JG, Beekley AC, McMullin NR, Park MS, Gonzalez EA, Wade CE, Dubick MA, Schwab CW, Moore FA, Champion HR, Hoyt DB, Hess JR. Damage control resuscitation: directly addressing the early coagulopathy of trauma. *J Trauma* 2007; **62**: 307 – 310.

Jurkovich GJ, Greiser WB, Luterman A, Curreri PW. Hypothermia in trauma victims: an ominous predictor of survival. *J Trauma* 1987; **27**: 1019 – 1024.

Kirkman E, Watts S, Hodgetts T, Mahoney P, Rawlinson S, Midwinter M. A proactive approach to the coagulopathy of trauma: the rationale and guidelines for treatment. *J R Army Med Corps* 2008; **153**: 302 – 306.

Kirkpatrick AW, Chun R, Brown R, Simons RK. Hypothermia and the trauma patient. *Can J Surg* 1999; **42**: 333 – 343.

Luna GK, Maier RV, Pavlin EG, Anardi D, Copass MK, Oreskovich MR. Incidence and effect of hypothermia in seriously injured patients. *J Trauma* 1987; **27**: 1014 – 1018.

MacLeod JB, Lynn M, McKenney MG, Cohn SM, Murtha M. Early coagulopathy predicts mortality in trauma. *J Trauma* 2003; **55**: 39 – 44.

Niles SE, McLaughlin DF, Perkins JG, Wade CE, Li Y, Spinella PC, Holcomb JB. Increased mortality associated with the early coagulopathy of trauma in combat casualties. *J Trauma* 2008; **64**: 1459 – 1465.

Seekamp A, van Griensven M, Hildebrandt F, Wahlers T, Tscherne H. Adenosine-triphosphate in trauma-related and elective hypothermia. *J Trauma* 1999; **47**: 673 – 683.

Shafi S, Elliott AC, Gentilello L. Is hypothermia simply a marker of shock and injury severity or an independent risk factor for mortality in trauma patients? Analysis of a large national trauma registry. *J Trauma* 2005; **59**: 1081 – 1085.

Tisherman SA. Hypothermia and injury. *Curr Opin Crit Care* 2004; **10**: 512 – 519.

Watts DD, Trask A, Soeken K, Perdue P, Dols S, Kaufmann C. Hypothermic coagulopathy in trauma: effect of varying levels of hypothermia on enzyme speed, platelet function, and fibrinolytic activity. *J Trauma* 1998; **44**: 846 – 854.

MISSION

The International Committee of the Red Cross (ICRC) is an impartial, neutral and independent organization whose exclusively humanitarian mission is to protect the lives and dignity of victims of armed conflict and other situations of violence and to provide them with assistance.

The ICRC also endeavours to prevent suffering by promoting and strengthening humanitarian law and universal humanitarian principles.

Established in 1863, the ICRC is at the origin of the Geneva Conventions and the International Red Cross and Red Crescent Movement. It directs and coordinates the international activities conducted by the Movement in armed conflicts and other situations of violence.

WAR SURGERY

WORKING WITH LIMITED RESOURCES IN ARMED CONFLICT AND OTHER SITUATIONS OF VIOLENCE

VOLUME 1

ICRC

WAR SURGERY
WORKING WITH LIMITED RESOURCES IN ARMED CONFLICT AND OTHER SITUATIONS OF VIOLENCE
VOLUME 1

0973/002 03 2009 3,000

ICRC

ICRC

REFERENCE

S Devaji Rao

General Surgical Emergencies

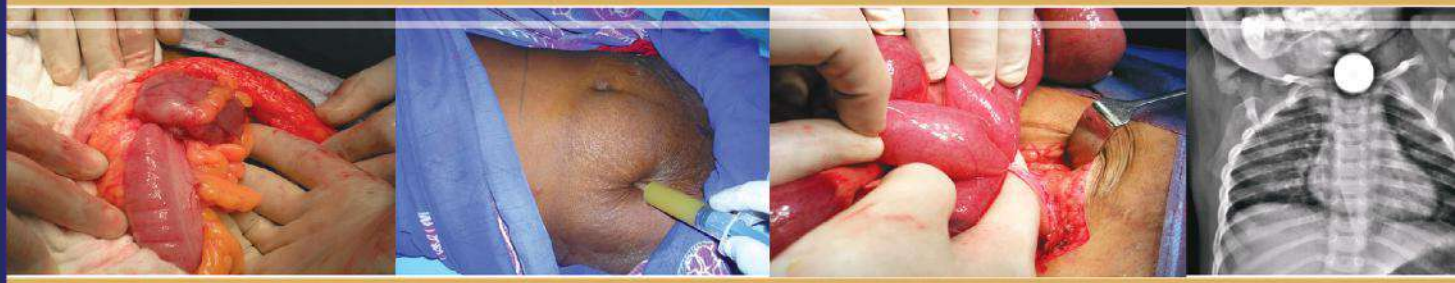

JAYPEE

General Surgical Emergencies

General Surgical Emergencies

S Devaji Rao MS MNAMS FICS
Teaching Faculty
National Board of Examinations
Senior Consultant in General Surgery
Surgical Gastroenterology and Surgical Oncology
St Isabel's Hospital, Mylapore, Chennai, Tamil Nadu, India
Chennai Meenakshi Multispecialty Hospital Ltd, Chennai, Tamil Nadu, India

JAYPEE BROTHERS MEDICAL PUBLISHERS (P) LTD.
New Delhi • Panama City • London • Dhaka • Kathmandu

Jaypee Brothers Medical Publishers (P) Ltd.

Headquarters

Jaypee Brothers Medical Publishers (P) Ltd.
4838/24, Ansari Road, Daryaganj
New Delhi 110 002, India
Phone: +91-11-43574357
Fax: +91-11-43574314
Email: jaypee@jaypeebrothers.com

Overseas Offices

J.P. Medical Ltd.
83 Victoria Street, London
SW1H 0HW (UK)
Phone: +44-2031708910
Fax: +02-03-0086180
Email: info@jpmedpub.com

Jaypee-Highlights Medical Publishers Inc.
City of Knowledge, Bld. 237, Clayton
Panama City, Panama
Phone: +507-301-0496
Fax: +507-301-0499
Email: cservice@jphmedical.com

Jaypee Brothers Medical Publishers (P) Ltd.
17/1-B Babar Road, Block-B, Shaymali
Mohammadpur, Dhaka-1207
Bangladesh
Mobile: +08801912003485
Email: jaypeedhaka@gmail.com

Jaypee Brothers Medical Publishers (P) Ltd.
Shorakhute, Kathmandu
Nepal
Phone: +00977-9841528578
Email: jaypee.nepal@gmail.com

Website: www.jaypeebrothers.com
Website: www.jaypeedigital.com

© 2012, Jaypee Brothers Medical Publishers

All rights reserved. No part of this book may be reproduced in any form or by any means without the prior permission of the publisher.

Inquiries for bulk sales may be solicited at: jaypee@jaypeebrothers.com

This book has been published in good faith that the contents provided by the author contained herein are original, and is intended for educational purposes only. While every effort is made to ensure accuracy of information, the publisher and the author specifically disclaim any damage, liability, or loss incurred, directly or indirectly, from the use or application of any of the contents of this work. If not specifically stated, all figures and tables are courtesy of the author. Where appropriate, the readers should consult with a specialist or contact the manufacturer of the drug or device.

General Surgical Emergencies

First Edition : **2012**

ISBN 978-93-5025-961-0

Printed at

Dedicated to

My teachers

My parents
Uma Bai and Siva Rao

My beloved wife
Kalpana

My daughters
Bhavna and Kirthana

Preface

Many textbooks are available dealing with the details of emergency surgical conditions. Not many are available dealing exclusively with the emergency conditions, especially those related to surgery. Not that every patient coming to the casualty at odd hours with an acute symptom has a life-threatening emergency, but it becomes necessary to comfort the patient at the first instant, followed by a quick diagnosis of the clinical condition. It is also imperative that the impending emergency should be identified by the clinician, so that a catastrophe is avoided. The classical example of such confusing situation is the acute pancreatitis. This can present with a variety of symptoms which will not give a lead towards diagnosis, but the whole clinical picture establishes after some time, and takes the patient through a very tough morbid situation, and may also end in death.

This concise book has been designed in such a way that the common emergency conditions are detailed to the extent so that surgical students or the casualty doctors or young surgeons may not miss the diagnosis. It cannot replace the many well-established textbooks, but it gives sufficient information for a clinician to manage the emergency.

Color photographs are useful add-ons to these chapters, which will make the reader remember the information for a long time.

Every new doctor who is resident in the casualty and the intensive care unit will face with a variety of clinical problems, and I hope the handbook will come handy.

I also hope that I have hit the required details at the right level for the young surgeons.

S Devaji Rao

Acknowledgments

I sincerely thank Professor SM Balaji, Professor J Cornelius, Dr Gayathri, Dr M Kanagavel, Dr Kirthana Rao, Professor N Mohan, Dr R Narasimhan, Professor MG Rajamanickam, Professor R Rajaraman, Professor N Sekar, Dr K Sridhar, Professor V Srinivasan, Dr V Thulasiraman, Dr Usha Dorairajan, Professor PS Venkatesh Rao and Dr S Vijayaraghavan, who contributed their clinical photographs, from their personal collections. My special thanks to my friends Dr Mani Veeraraghavan for the endoscopy photographs and Dr V Ganesan for the ultrasound photographs. Most of the CT and MRI pictures were procured from Bharat Scans, Chennai, Tamil Nadu, India, and I sincerely thank Dr R Emmanuel for providing them.

My special thanks to my daughters Dr Bhavna for going through the script and making corrections, and Dr Kirthana for drawing the diagrams.

My very special thanks goes to my wife Kalpana for the tolerance during the preparation of this manual.

Contents

SECTION ONE: INTRODUCTION

1. Introduction **3**

SECTION TWO: ASSESSMENT

2. Assessment of Surgical Emergencies **7**
- ❖ Make the Patient Lie Down Comfortably 7
- ❖ Elicit a Quick History 8
  - • Make a Thorough Clinical Examination 8
  - • Come to a Quick Working Clinical Diagnosis 10
  - • Ask for Essential Meaningful Investigations 10
  - • Admit the Patient where Thought to be Required 11
  - • Put the Patient in Intensive Care Whenever Required 11
  - • Collect Investigation Reports at the Earliest 11
  - • Start Emergency Treatment 11
  - • Ask for Expert Opinions 12

SECTION THREE: CRITICAL CARE

3. Critically Ill Patient and Critical Care **15**
- ❖ Cardiac Support 15
- ❖ Respiratory Support 15
- ❖ Inotropic Support 16

4. Shock **17**
- ❖ Hypovolemic Shock 17
  - • Treatment of Hypovolemic Shock 18
- ❖ Septic Shock 18
  - • Symptoms and Signs 18
  - • Pathogenesis 18
  - • Treatment 19
- ❖ Anaphylactic Shock 19
  - • Incidence and Etiology 19
  - • Pathogenesis 19
  - • Symptoms and Signs 19
  - • Treatment 20
- ❖ Cardiogenic Shock 20
- ❖ Neurogenic Shock 20
  - • Pathogenesis 20
  - • Treatment 20

5. Acute Respiratory Distress Syndrome **21**
  - • Incidence and Etiology 21
  - • Pathogenesis 21
  - • Clinical Presentation 21
  - • Relevant Investigations 21
  - • Treatment 22

SECTION FOUR: TRAUMA

6. Polytrauma **25**
- ❖ Introduction 25
- ❖ Preparation 25
  - • Prehospital Communication 25

❖ Trauma Reception Team 26
• Receiving the Patient 27
• Primary Survey and Resuscitation 27
• Airway and Cervical Spine Control 27
• Breathing 28
• Circulation and Hemorrhage Control 29
• Recognition and Assessment of Hypovolemia 29
• Fluid Resuscitation 30
• Disability 30
• Exposure 30
❖ Secondary Survey 30
❖ Clinical Evaluation 31
• Eliciting History 31
• Physical Examination 31
• Scalp 31
• Neurological State 31
• Base of Skull 31
• Neck 32
• Eyes 32
• Face 32
• Thorax 32
• Abdomen 33
• Extremities 33
❖ Definitive Care 33

7. Head Injuries 34
❖ Brain Injury 34
❖ Fractures of Skull 36
• Symptoms and Signs 36
❖ CSF Fistulae 39
❖ Scalp Injuries 39

8. Facial Injuries 41
❖ Fractures of Midfacial Skeleton 41
❖ Fractures of Mandible 42
• Clinical Presentation 43
• Relevant Investigations 45
• Treatment 45

9. Spine and Spinal Cord Injuries 46
❖ Injuries to Bony and Ligamentous Spine 46
• Incidence and Etiology 46
• Symptoms 46
• Signs 47
• Relevant Investigations 47
• Treatment 48
❖ Injuries of the Spinal Cord 48
• Incidence and Etiology 48
• Symptoms and Signs 48
• Relevant Investigations 49
• Treatment 49

10. Thoracic Injuries 50
❖ Rib Fractures 50
• Incidence and Etiology 50
• Symptoms 50
• Signs 51
• Relevant Investigations 51
• Treatment 51
❖ Flail Chest 53
• Incidence and Etiology 53
• Symptom 54
• Sign 54
• Relevant Investigations 54
• Treatment 54
❖ Sternal Fracture 54
• Incidence and Etiology 54
• Symptom 54
• Sign 55
• Relevant Investigations 55
• Treatment 55
❖ Pneumothorax 55
• Incidence and Etiology 55

- Symptoms 56
- Signs 56
- Relevant Investigations 56
- Treatment 58

❖ Surgical Emphysema 58
- Incidence and Etiology 58
- Pathogenesis 59
- Symptoms and Signs 59
- Relevant Investigations 59
- Treatment 59

❖ Hemothorax 59
- Incidence and Etiology 59
- Symptoms 60
- Signs 60
- Relevant Investigations 60
- Treatment 61

❖ Pulmonary Contusion/laceration 61
- Incidence and Etiology 61
- Symptoms and Signs 61
- Relevant Investigations 61
- Treatment 62

❖ Injuries of Thoracic Aorta 62
- Incidence and Etiology 62
- Symptom 63
- Signs 63
- Relevant Investigations 63
- Treatment 64

❖ Injuries of Myocardium 64
- Incidence and Etiology 64
- Symptoms 64
- Signs 64
- Relevant Investigations 65
- Treatment 65

❖ Esophageal Injuries 65
- Incidence and Etiology 65
- Symptoms 65
- Signs 65
- Relevant Investigations 65
- Treatment 66

❖ Diaphragmatic Injuries 66
- Incidence and Etiology 66
- Symptoms and Signs 67
- Relevant Investigations 67
- Treatment 67

11. Abdominal Injuries 68

❖ Closed Injuries 68

❖ Open Injuries 69
- Symptoms and Signs 69

❖ Injuries of Liver 69
- Incidence and Etiology 69
- Symptoms and Signs 70
- Relevant Investigations 70
- Treatment 70

❖ Injuries of Spleen 72
- Incidence and Etiology 72
- Symptoms 72
- Signs 72
- Relevant Investigations 73
- Treatment 74

❖ Injuries of Mesentery 74
- Incidence and Etiology 74
- Symptoms 74
- Sign 75
- Relevant Investigations 75
- Treatment 75

❖ Injuries of Duodenum 75
- Incidence and Etiology 75
- Symptoms and Signs 75
- Relevant Investigations 75
- Treatment 76

❖ Injuries of Small Intestine 76
- Incidence and Etiology 76

- Clinical Presentation 76
- Relevant Investigations 76
- Treatment 77

❖ Injuries of Large Intestine 77
- Incidence and Etiology 77
- Symptoms 77
- Signs 77
- Relevant Investigations 77
- Treatment 77

❖ Abdominal Compartment Syndrome 78
- Incidence and Etiology 78
- Pathogenesis 78
- Symptom 78
- Signs 78
- Relevant Investigations 78
- Treatment 78

12. Urological Injuries 79

❖ Renal Injuries 79
- Incidence and Etiology 79
- Symptom 79
- Signs 79
- Relevant Investigations 80
- Treatment 80

❖ Ureteric Injuries 80
- Incidence and Etiology 80
- Symptom and Sign 81
- Relevant Investigation 82
- Treatment 82

❖ Urinary Bladder Injuries 82
- Incidence and Etiology 82
- Symptoms 83
- Signs 83
- Relevant Investigations 83
- Treatment 84

❖ Urethral Injuries 84
- Incidence and Etiology 84
- Symptom 84
- Sign 84
- Relevant Investigation 84
- Treatment 85

13. Male Genital Injuries 86

❖ Injuries of Penile Skin 86
- Incidence and Etiology 86
- Symptom 86
- Sign 86
- Relevant Investigation 86
- Treatment 86

❖ Foreign Bodies in Penile Urethra 87
- Incidence and Etiology 87
- Symptoms and Signs 87
- Relevant Investigation 88
- Treatment 88

❖ Scrotal Laceration 88
- Incidence and Etiology 88
- Symptom 89
- Signs 89
- Relevant Investigation 89
- Treatment 89

14. Female Genital Injuries 90

❖ Coital Injuries 90
- Incidence and Etiology 90
- Symptom 90
- Signs 90
- Relevant Investigations 90
- Treatment 90

❖ Injuries during Childbirth 91
- Incidence and Etiology 91
- Clinical Presentation 91
- Relevant Investigations 91
- Treatment 91

15. Hand Injuries 92
- Incidence and Etiology 92

- Clinical Presentation 92
- Relevant Investigations 96
- Treatment 96

SECTION FIVE: ORGANS AND SYSTEMS

16. Oral Cavity 99

- Acute Odontogenic Infections 99
  - Incidence and Etiology 99
  - Pathogenesis 99
  - Symptoms and Signs 100
  - Relevant Investigations 100
  - Treatment 100

17. Head and Neck 101

- Acute Suppurative Sialadenitis 101
  - Incidence and Etiology 101
  - Clinical Presentation 101
  - Relevant Investigation 103
  - Treatment 103
- Ludwig's Angina 103
  - Incidence and Etiology 103
  - Pathogenesis 103
  - Symptoms 103
  - Signs 103
  - Relevant Investigation 103
  - Treatment 104

18. Thorax 105

- Acute Mediastinitis 105
  - Incidence and Etiology 105
  - Clinical Presentation 105
  - Relevant Investigations 105
  - Treatment 106
- Acute Pleuritis 106
  - Incidence and Etiology 106
  - Symptom 106
  - Sign 106
  - Relevant Investigations 107
  - Treatment 107
- Pleural Effusion 107
  - Incidence and Etiology 107
  - Symptom 107
  - Signs 107
  - Relevant Investigations 108
  - Treatment 109
- Acute Empyema Thoracis 109
  - Incidence and Etiology 109
  - Symptom 110
  - Signs 110
  - Relevant Investigations 110
  - Treatment 110
- Spontaneous Pneumothorax 110
  - Incidence and Etiology 110
  - Symptom 110
  - Sign 111
  - Relevant Investigations 111
  - Treatment 111
- Foreign Bodies in the Respiratory Tract 112
  - Incidence and Etiology 112
  - Pathogenesis 112
  - Symptoms 112
  - Signs 112
  - Relevant Investigations 112
  - Treatment 113
- Suppurative or Aspiration Pneumonia 113
  - Incidence and Etiology 113
  - Pathogenesis 113
  - Symptoms 114
  - Sign 114
  - Relevant Investigation 114
  - Treatment 114
- Hospital Acquired Pneumonia 115
  - Incidence and Etiology 115

- Symptoms 115
- Signs 115
- Relevant Investigations 115
- Treatment 115

■ Acute Lung Abscess 115
- Incidence and Etiology 115
- Symptoms 116
- Sign 116
- Relevant Investigations 116
- Treatment 116

■ Pulmonary Embolism 116
- Incidence and Etiology 116
- Symptoms and Signs 116
- Relevant Investigations 117
- Treatment 117

19. Breast 118

■ Breast Hematoma 118
- Incidence and Etiology 118
- Symptoms 118
- Sign 118
- Relevant Investigations 118
- Treatment 118

■ Acute Breast Abscess 119
- Incidence and Etiology 119
- Symptoms 119
- Signs 119
- Relevant Investigations 119
- Treatment 119

20. Spine 120

■ Degenerative Diseases of Disk and Facet Joints 120
- Incidence and Etiology 120
- Symptoms and Signs 120
- Relevant Investigations 121
- Treatment 121

■ Spondylolisthesis 122
- Incidence and Etiology 122
- Symptom 122
- Signs 122
- Relevant Investigations 123
- Treatment 123

21. Gastroenterology 124

■ Acute Abdomen 124
- Introduction 124
- Pain 125
- Vomiting 128

❖ Acute Upper Abdominal Pain 130
- Abdomen 131
- Relevant Investigations 133
- Treatment Plan 134

❖ Acute Lower Abdominal Pain 134
- Treatment Plan 138

■ Acute Liver Abscess 138
- Incidence and Etiology 138
- Pathogenesis 138
- Symptoms 138
- Signs 139
- Differential Diagnosis 139
- Relevant Investigations 139
- Treatment 140

■ Acute Cholecystitis 140
- Incidence and Etiology 140
- Pathogenesis 140
- Complications 140
- Symptoms 140
- Signs 141
- Differential Diagnosis 141
- Relevant Investigations 141
- Treatment 142

■ Acute Cholangitis 143
- Incidence and Etiology 143

- Pathogenesis 143
- Symptoms 143
- Signs 144
- Relevant Investigations 144
- Treatment 146

■ Primary Sclerosing Cholangitis 146
- Incidence and Etiology 146
- Clinical Presentation 146
- Relevant Investigations 146
- Treatment 147

■ Splenic Abscess 148
- Incidence and Etiology 148
- Pathogenesis 148
- Symptoms 148
- Sign 148
- Relevant Investigation 148
- Treatment 149

■ Acute Hemorrhagic Pancreatitis 149
- Incidence and Etiology 149
- Pathogenesis 150
- Symptoms 150
- Signs 150
- Relevant Investigations 151
- Treatment 153

■ Acute Appendicitis 154
- Incidence and Etiology 154
- Pathogenesis 154
- Symptoms 154
- Signs 155
- Relevant Investigations 157
- Treatment 159

■ Acute Mesenteric Lymphadenitis 159
- Incidence and Etiology 159
- Pathogenesis 159
- Symptoms 159
- Signs 159
- Relevant Investigations 159
- Treatment 160

■ Acute Colonic Diverticulitis 160
- Incidence and Etiology 160
- Pathogenesis 160
- Symptoms 161
- Signs 161
- Relevant Investigations 161
- Treatment 162

■ Acute Meckel's Diverticulitis 162
- Incidence and Etiology 162
- Symptom 163
- Sign 163
- Differential Diagnosis 163
- Relevant Investigations 163
- Treatment 164

■ Acute Solitary Cecal Diverticulitis 164
- Incidence and Etiology 164
- Symptom 164
- Sign 164
- Relevant Investigations 164
- Treatment 164

■ Acute Ulcerative Colitis 164
- Incidence and Etiology 164
- Symptoms 165
- Signs 165
- Relevant Investigations 165
- Treatment 165

■ Acute Intestinal Obstruction 166
- Symptoms 166
- Signs 167
- Relevant Investigations 167
- Treatment 168

■ Gallstone Ileus 168
- Incidence and Etiology 168
- Symptoms 169

- Signs 169
- Relevant Investigations 169
- Treatment 169

■ Acute Intussusception 170
- Incidence and Etiology 170
- Symptoms 171
- Signs 172
- Relevant Investigations 172
- Treatment 173

■ Swallowed Foreign Bodies 173
- Incidence and Etiology 173
- Symptoms 173
- Signs 173
- Relevant Investigations 173
- Treatment 173

■ Perforated Bowel Pathologies 174
- Differential Diagnosis 175
- Relevant Investigations 175
- Treatment 177

■ Intestinal Strictures 177
- Incidence and Etiology 177
- Symptoms 177
- Signs 177
- Relevant Investigations 177
- Treatment 177

■ Bands and Adhesions 178
- Incidence and Etiology 178
- Symptoms and Signs 178
- Relevant Investigations 178
- Treatment 178

■ Enteroliths/Food Bolus 179
- Incidence and Etiology 179
- Symptoms and Signs 179
- Relevant Investigations 179
- Treatment 179

■ Volvulus 179
❖ Sigmoid Volvulus 179
- Incidence and Etiology 179
- Pathogenesis 179
- Symptoms 179
- Signs 180
- Relevant Investigations 180
- Treatment 180

❖ Cecal Volvulus 180
- Incidence and Etiology 180
- Clinical Presentation 181
- Relevant Investigations 181
- Treatment 181

❖ Midgut Volvulus 181
- Incidence and Etiology 181
- Symptom 181
- Signs 181
- Relevant Investigation 181
- Treatment 181

❖ Gastric Volvulus 182
- Incidence and Etiology 182
- Symptoms 182
- Sign 182
- Relevant Investigation 182
- Treatment 182

■ Intestinal Obstruction due to Herniae (Internal and External) 182
- Incidence and Etiology 182
- Symptoms 183
- Sign 183
- Relevant Investigation 183
- Treatment 183

■ Paralytic Ileus 183
- Incidence and Etiology 183
- Symptoms 184

- Sign 184
- Relevant Investigations 184
- Treatment 185

■ Torsion of Mesenteric Cyst 185
- Incidence and Etiology 185
- Pathogenesis 186
- Symptom 186
- Signs 186
- Relevant Investigation 186
- Treatment 186

■ Torsion of Omentum 186
- Incidence and Etiology 186
- Pathogenesis 186
- Symptoms 187
- Signs 187
- Relevant Investigation 187
- Treatment 187

■ Colics 187
- Incidence and Etiology 187
- Clinical Features 188
- Symptoms 188
- Signs 189
- Differential Diagnosis by Clinical History and Examination 190
- Relevant Investigations 190
- Radiology 190
- Treatment Plan 190

■ Gastrointestinal Hemorrhage 190
- Definitions 190
- Types of Gastrointestinal Hemorrhages 191
- Hematemesis and Melena 192
- Melena 192
- Hematochezia 192
- Eliciting History 193
- Hematochezia 193
- Past History 194
- Family History 194
- Clinical Examination 194
- Abdomen 195
- Differential Diagnosis by Clinical History and Examination 195
- Relevant Investigations 196
- Treatment 198
- Surgical Treatment 198

22. Anorectum 199

■ Acute Anal Fissure 199
- Incidence and Etiology 199
- Pathogenesis 199
- Symptoms 199
- Signs 200
- Relevant Investigation 201
- Treatment 201

■ Anorectal Abscess 201
- Incidence and Etiology 201
- Pathogenesis 202
- Symptoms 202
- Signs 202
- Relevant Investigation 203
- Treatment 203

■ Hemorrhoids 203
- Incidence and Etiology 203
- Symptoms 203
- Signs 204
- Relevant Investigation 205
- Treatment 205

■ Perianal Hematoma 206
- Incidence and Etiology 206
- Pathogenesis 206
- Symptoms 206
- Sign 206

- Relevant Investigation 207
- Treatment 207

■ Prolapse of Rectum 207
- Incidence and Etiology 207
- Pathogenesis 207
- Symptoms 207
- Signs 207
- Relevant Investigation 208
- Treatment 208

23. Vascular System 209

■ Acute Limb Ischemia 209
- Incidence and Etiology 209
- Symptoms 210
- Signs 210
- Relevant Investigations 211

■ Acute Intestinal Ischemia 211
- Incidence and Etiology 211
- Clinical Presentation 211
- Relevant Investigations 212
- Treatment 212

■ Leaking or Dissecting Aortic Aneurysm 212
- Incidence and Etiology 212
- Clinical Presentation 212
- Relevant Investigation 213
- Treatment 213

24. Urology 214

■ Acute Retention of Urine 214
- Definition 214
- Incidence and Etiology 214
- Symptoms 214
- Signs 214
- Relevant Investigations 215
- Treatment 215

■ Hematuria 215
- Definition 215
- Incidence and Etiology 215
- Symptoms 216
- Signs 216
- Relevant Investigations 216
- Treatment 217

■ Renal Colic 217
- Relevant Investigations 218
- Treatment 218

■ Ureteric Colic 218
- Symptom 218
- Relevant Investigations 218
- Treatment 219

■ Acute Urethritis 220
- Incidence and Etiology 220
- Symptoms 220
- Sign 220
- Relevant Investigations 220
- Treatment 220

■ Acute Prostatitis 221
- Incidence and Etiology 221
- Symptom 221
- Sign 221
- Relevant Investigation 221
- Treatment 221

■ Acute Prostatic Abscess 221
- Incidence and Etiology 221
- Symptom 221
- Sign 222
- Relevant Investigations 222
- Treatment 222

25. Male Genitalia 223

■ Acute Scrotal Pain 223
- Definition 223
- Causes of Acute Scrotal Pain 223
- Symptoms 224
- Past History 224
- Signs 224

- On Palpation 225
- Differential Diagnosis by Clinical History and Examination 225
- Relevant Investigations 225
- Treatment 226

■ Torsion of Testis 226
- Incidence and Etiology 226
- Pathogenesis 226
- Symptoms 226
- Signs 227
- Relevant Investigation 227

■ Torsion of Appendages of Testis 228
- Incidence and Etiology 228
- Pathogenesis 228
- Symptom 228
- Signs 228
- Relevant Investigation 228
- Treatment 228

■ Acute Epididymo-orchitis 229
- Incidence and Etiology 229
- Pathogenesis 229
- Symptom 229
- Signs 229
- Relevant Investigations 229
- Treatment 229

■ Traumatic Orchitis 230
- Incidence and Etiology 230
- Clinical Presentation 230
- Relevant Investigation 230
- Treatment 230

■ Hematocele 230
- Definition 230
- Incidence and Etiology 230
- Symptoms 230
- Signs 231
- Relevant Investigation 231
- Treatment 231

■ Pyocele 231
- Definition 231
- Incidence and Etiology 232
- Symptom 232
- Signs 232
- Relevant Investigations 232
- Treatment 232

■ Idiopathic Scrotal Edema 232
- Incidence and Etiology 232
- Pathogenesis 232
- Symptom 232
- Signs 233
- Relevant Investigations 233
- Treatment 233

■ Acute Scrotal Abscess 233
- Incidence and Etiology 233
- Symptoms 233
- Sign 233
- Relevant Investigations 233
- Treatment 234

■ Fournier's Gangrene 234
- Incidence and Etiology 234
- Symptoms 235
- Signs 235
- Relevant Investigation 235
- Treatment 235

■ Acute Filarial Scrotum 236
- Incidence and Etiology 236
- Clinical Presentation 236
- Relevant Investigation 236
- Treatment 236

■ Fracture Penis 237
- Incidence and Etiology 237
- Symptoms and Signs 237

- Relevant Investigations 238
- Treatment 238

■ Paraphimosis 238
- Incidence and Etiology 238
- Symptoms 238
- Sign 238
- Relevant Investigation 238
- Treatment 238

■ Priapism 239
- Incidence and Etiology 239
- Clinical Presentation 240
- Relevant Investigation 240
- Treatment 240

26. Hernias 241

■ Complicated Hernias 241
- Definition 241
- Anatomy of the Hernial Sac 242
- Complications 242
- Symptoms 243
- Signs 244
- Relevant Investigations 244
- Treatment 244

27. Gynecology 245

■ Acute Torsion of Ovarian Cyst 245
- Incidence and Etiology 245
- Pathogenesis 245
- Symptom 245
- Signs 245
- Relevant Investigations 245
- Treatment 246

■ Acute Salpingitis 246
- Incidence and Etiology 246
- Pathogenesis 246
- Symptoms 246
- Signs 247
- Relevant Investigations 247
- Treatment 247

■ Rupture of Lutein Cyst 247
- Incidence and Etiology 247
- Clinical Presentation 247
- Relevant Investigations 247
- Treatment 247

28. Pediatrics 248

■ Acute Intussusception 248
- Incidence and Etiology 248
- Pathogenesis 248
- Symptoms 248
- Signs 249
- Relevant Investigations 249
- Treatment 249

■ Congenital Pyloric Stenosis 250
- Incidence and Etiology 250
- Pathogenesis 250
- Symptoms 250
- Signs 250
- Relevant Investigations 250
- Treatment 250

■ Necrotizing Enterocolitis 250
- Incidence and Etiology 250
- Pathogenesis 251
- Symptoms 251
- Signs 251
- Relevant Investigations 251
- Treatment 251

■ Tracheoesophageal Fistula 251
- Incidence and Etiology 251
- Pathogenesis 251
- Symptoms 252
- Signs 252
- Relevant Investigations 252
- Treatment 252

29. Lymphatic System **253**
- Acute Lymphangitis 253
  - Incidence and Etiology 253
  - Clinical Presentation 253
  - Relevant Investigations 254
  - Treatment 254
- Acute Viral and Bacterial Lymphadenitis 254
  - Incidence and Etiology 254
  - Symptoms 254
  - Signs 254
  - Relevant Investigations 255
  - Treatment 255
- Acute Filarial Lymphangitis and Lymphadenitis 255
  - Incidence and Etiology 255
  - Symptoms 255
  - Signs 255
  - Relevant Investigation 255
  - Treatment 256

30. Skin and Subcutaneous Tissues **257**
- Hematoma 257
  - Incidence and Etiology 257
  - Symptom 257
  - Sign 257
  - Relevant Investigations 257
  - Treatment 257
- Erysipelas 258
  - Incidence and Etiology 258
  - Symptoms 258
  - Investigations 259
  - Treatment 259
- Furuncle 259
  - Incidence and Etiology 259
  - Symptom 259
  - Signs 260
  - Relevant Investigations 260
  - Treatment 260
- Cellulitis 260
  - Incidence and Etiology 260
  - Pathogenesis 260
  - Symptoms 261
  - Signs 261
  - Relevant Investigations 261
  - Treatment 261
- Acute Pyogenic Abscess 261
  - Incidence and Etiology 261
  - Symptoms 262
  - Sign 262
  - Relevant Investigations 262
  - Treatment 262
- Carbuncle 262
  - Incidence and Etiology 262
  - Symptom 263
  - Signs 263
  - Relevant Investigations 263
  - Treatment 263
- Burns 263
  - Definition 263
  - Pathogenesis 264
  - Clinical Features 264
  - Metabolic Effects of Burns 267
  - ❖ Physical Examination 268
    - Preliminaries Before Examination 268
  - ❖ Inspection 271
    - General Examination 272
    - Systemic Examination 272
    - Relevant Investigations 272
    - Treatment 273
- Necrotizing Fasciitis 274
  - Incidence and Etiology 274
  - Pathogenesis 274

- Symptom 274
- Sign 274
- Relevant Investigations 274
- Treatment 274

SECTION SIX: SPECIAL PROCEDURES AND SITUATIONS

31. Emergency Surgical Procedures 277

- Pericardial Aspiration 277
  - Materials Required 277
  - Procedure 277
  - Complications 278
- Intubation of Trachea 279
  - Materials Required 279
  - Procedure 279
  - Complication of Endotracheal Intubation 280
- Percutaneous Tracheostomy 280
  - Advantages of Tracheostomy 280
  - Indications 280
  - Contraindications 281
  - Materials Needed for Percutaneous Tracheostomy 281
  - Procedure 281
  - Complications 282
- Cricothyroidotomy/Minitracheostomy 282
  - Indications 282
  - Contraindications 283
  - Materials Required 283
  - Procedure 284
  - Complications 284
- Insertion of Chest Drains 284
  - Indications 284
  - Type of Drain 285
  - Site of Drain 285
  - Materials Required 286
  - Procedures 286
- Passing a Nasogastric Tube 287
  - Indications 287
  - Contraindications 289
  - Materials Required 289
  - Procedure 289
- Peritoneal Tap 289
  - Materials Required 289
  - Procedure 291

32. Preparing the Patient for Emergency Surgery 292

- General Measures 292
- Investigations Support 292
- Situations which Need Special Attention 293

33. Antibiotics 294

- Antibiotics and Emergency Surgery 294
  - Which Emergency Surgical Procedure should have Antibiotic Prophylaxis ? 294

34. Death 296

- Breaking the News 296

Index 299

Section I

Introduction

1. Introduction

1 Introduction

Emergencies are defined as the clinical situations which require urgent care. Not necessarily they should have a life-threatening pathology, but even a simple innocuous pathology can create an emergency situation, such as a passing calculus in the ureter can cause such a pain which will upset the patient, the relatives and friends, and warrant emergency care and also confuse the attending doctor till he makes a final diagnosis.

Clinical diagnosis of such emergencies may be difficult in some patients forcing the clinician to run through all the investigations to make the final diagnosis. A well informed medical student or a young surgeon will take it right, will not upset himself, comfort the patient and go through the clinical examination and administer the right treatment.

In some emergencies like hemorrhagic conditions, history taking cannot be done in detail due to paucity of time, and the treatment should get initiated at the earliest, probably in the intensive care unit. Failure to recognize and manage such critical illnesses, may lead to pathophysiological changes leading to multiple organ dysfunction. In such a situation, intensive care plays a very important role. The role of intensive care includes:

- Resuscitation and stabilization
- Optimization to prevent organ failure
- Decision regarding complex surgery
- Support of failing organs
- Recognition of futility.

In well developed hospitals, the critical care is managed at two levels, namely, intensive care units (ICU) and high dependency units (HDU), only the availability of ventilators in the ICU differentiating the both. In the ICU, some patients may require one patient – one nurse combination.

Critically ill patients can be classified according to the level of medical and nursing care needed.

- Level 0: Patients who can be managed by ward based care
- Level 1: Patients who can be managed by ward based care with advise or support from the critical care team
- Level 2: Patients requiring more advanced levels of monitoring or intervention, such as management of single failing organ system
- Level 3: Patients requiring advanced respiratory support and management of multiple organ failure.

For a good outcome, the first step is to make a proper diagnosis, followed by the necessary investigations and sensible management.

Section II

Assessment

2. Assessment of Surgical Emergencies

2

Assessment of Surgical Emergencies

As soon as the patient is wheeled into the hospital or the casualty, it is required to determine whether the patient's condition warrants emergency care or not. Though there are parameters which help to determine this, many times, experience matters. For example, a patient with acute pancreatitis may not look very sick after an episode of acute abdominal pain. The following measures will help determine the seriousness of the patient.

1. Make the patient lie down comfortably.
2. Elicit a quick history.
3. Make a thorough clinical examination.
4. Come to a quick working clinical diagnosis.
5. Ask for essential meaningful investigations.
6. Admit the patient where thought to be required.
7. Put the patient in intensive care whenever required.
8. Collect investigation reports at the earliest.
9. Start emergency treatment.
10. Ask for expert opinions.

MAKE THE PATIENT LIE DOWN COMFORTABLY

It is necessary to make the patient lie down, as a walk in patient may collapse suddenly (e.g. septic shock, hemorrhagic shock).

ELICIT A QUICK HISTORY

A good history will establish the diagnosis in majority of cases.

However, when the patient is too sick, a quick questioning is only possible. They are:

- A – allergies
- M – medications
- P – previous medical history, pregnancy
- L – Last meal
- E – Events leading to admission.

If possible, and if there is enough time, full history should be elicited in the following:

- Chief complaint
- History of present illness
- Previous history of present complaint
- Past medical history
- History of drug intake and allergies
- Social and personal history
- Family history
- History of immunization.

Make a Thorough Clinical Examination

General Examination

- Skin
  - Pallor (e.g. anemia)
  - Yellow discoloration (e.g. jaundice)
  - Bluish discoloration (e.g. cyanosis)
  - Pigmentation (e.g. von Recklinghausen's disease)
  - Eruptions (e.g. macule, papule, vesicle, bulla, etc.)
  - Nodules (e.g. Heberden's nodes of osteoarthrosis)
  - Visible veins (e.g. visible neck veins of cardiac overload, caput medusae of portal hypertension).
- Eyes
  - Orbital region
    - Prominence of eyeballs (e.g. exophthalmos of thyrotoxicosis)
    - Sunken eyeballs (e.g. dehydration)

- ❖ Eyelids
  - Puffiness (e.g. trauma, renal pathology)
  - Color (e.g. black eye of head injuries)
- ❖ Conjunctivae
  - Color (e.g. pale in anemics)
- ❖ Sclera
  - Color (e.g. yellow in jaundiced patients)
- ❖ Cornea and iris
  - Opacities and ulcers (e.g. trauma, infections)
- ❖ Pupils
  - Size, shape, symmetry and reaction to light (e.g. changes in cranial nerve lesions)

- Oral cavity
  - ❖ Lips
    - Color and texture (e.g. cracked lips of exposure to extreme cold)
    - Pigmentation (e.g. telangiectasia)
  - ❖ Gums
    - Color (e.g. blue line of lead poisoning)
    - Texture (e.g. inflammation due to gingivitis)
    - Pigmentation (e.g. hemorrhages due to leukemia)
  - ❖ Teeth
    - Color and pigmentation (e.g. mottled yellow color of fluorosis)
    - Number and health status of the teeth (e.g. Hutchinson's teeth of congenital syphilis)
  - ❖ Tongue
    - Color (e.g. blue color of central cyanosis)
    - Texture (e.g. smooth tongue of vitamin B_{12} deficiency)
    - Ulcer (e.g. aphthous ulcers)
    - Altered movements (e.g. hypoglossal nerve palsy)
  - ❖ Hard palate and soft palate
    - Color (e.g. anemia)
    - Cleft (e.g. cleft palate)
    - Ulcer (e.g. malignancy).

Halitosis (Bad breath) may be due to poor oral hygiene or even oral pathologies, e.g. oral cancers or bronchopulmonary pathologies like lung abscesses.

- Hands and feet
  - ❖ Hands of the patient should be examined carefully, like
    - Shape of the hands (e.g. characteristic shape of Dupuytren's contracture, tetany)
    - State of the joints (e.g. deformed joints of rheumatoid arthritis)
    - Shape and color and deformity fingers (e.g. clubbing or nicotine staining of chronic smokers)
    - Nails (e.g. koilonychia of iron-deficiency anemia)
    - Abnormal movements of fingers (e.g. tremors of thyrotoxicosis)
  - ❖ Feet of the patient are generally tucked under the bedclothes and their examination should not be forgotten. Examination may reveal
    - Edema, unilateral (e.g. filarial leg) and bilateral (e.g. hypoproteinemia)
    - Skin changes (e.g. pigmentation or ulcers due to ischemia)
    - Ulcers (e.g. trophic ulcers)
- Look for (jaundice, anemia, cyanosis, clubbing, oedema, lymphadenopathy – JACCOL)
- Vitals (pulse, respiration, blood pressure)
- Peripheral examination (warm/cool, capillary refill)
- Jugular venous pressure
- Mental test score or GCS (Glasgow Coma Scale), if the patient appears confused or impaired consciousness).

Systemic Examination

- Site of interest
- Other systems.

Come to a Quick Working Clinical Diagnosis

- A quick working clinical diagnosis should come to the mind of the clinical examiner
- He should also keep the various differential diagnosis in mind, so that diagnosis can be revised at anytime, without bias.

Ask for Essential Meaningful Investigations

- It is essential to ask only for meaningful investigations
- If investigation packages are available in the hospital, it is worthwhile to ask for it which may include the required investigations also
- If asking for this package will delay the results, it is appropriate to ask for the relevant ones first, followed by the rest.

The blood tests in general should include:

- Complete blood count (CBC)
- Urea, creatinine and electrolytes
- Blood sugar (random)
- Serum amylase, lipase
- Liver function tests
- Coagulation profile
- Blood grouping and Rh typing
- Arterial blood gas analysis (ABG).

The radiological tests should include:

- Chest X-ray (erect)
- Abdominal X-ray (supine and erect)
- Ultrasound scan
- CT.

Admit the Patient where Thought to be Required

No time should be wasted in deciding where to admit the patient. The best bed available is the best choice.

Put the Patient in Intensive Care Whenever Required

When in doubt about the seriousness, admit the patient in the intensive care.

Collect Investigation Reports at the Earliest

In the commotion during the emergency, it is not unusual to forget to collect the reports. The clinician should stay composed, and should remember to collect all the reports, which will help in the management.

Start Emergency Treatment

- It is always wise to start an intravenous line for control of lifeline. The choice of fluid depends on the situation (e.g. normal saline in diabetics, glucose solutions in hypertensives, plasma expanders in shocked patients)
- Antibiotics may be needed if infective pathology is suspected
- Oxygen by nasal mask is a safe additive
- Urethral catheterization and hourly output determination is necessary.

- Start emergency treatment if required even in the casualty, and history taking can be followed later
- While examining a patient in emergency, site of interest should be examined first to save time so that appropriate treatment is started.

Ask for Expert Opinions

- It is always necessary to obtain expert opinions to manage the situation
- Multiple opinions may be necessary, but it should not hurt the sentiments of the experts involved
- Intravenous fluids (rate and nature) and antibiotics will vary according to the laboratory reports and revised diagnosis
- Colloids may be necessary and should be adequately given at appropriate times
- The average daily water requirements for adult are 30 to 35 ml/kg + 500 ml/day/degree of pyrexia (above 37°C)
- Normal adult requirements for electrolytes are 1 mmol/kg each of Na, K, Cl
- Normal fluid requirements for children depend on weight
  - <10 kg : 100 ml/kg/day
  - 10 to 20 kg : 1 liter + 50 ml/kg/day
  - >20 kg : 1.5 liters + 25 ml/kg/day.

Section III

Critical Care

3. Critically Ill Patient and Critical Care
4. Shock
5. Acute Respiratory Distress Syndrome

3

Critically Ill Patient and Critical Care

A critically ill patient requires high level monitoring, especially when circulatory/respiratory support is required, or if more than one organ is failing. The intensive care is started early enough to be able to reverse the condition.

CARDIAC SUPPORT

It depends on the Hb, perfusion pressure, ventilation and gas exchange. For this to be achieved, the following are required:

- Arterial lines (beat to beat BP indication, easy access for arterial blood sampling)
- Central venous pressure monitoring (measures intravascular volume, and the right heart preload)
- Pulmonary artery catheterization (measures cardiac output, ensures optimal fluid resuscitation and helps in the use of vasoactive drugs).

> CVP monitoring indicates only the right heart filling or preload, is not a reliable measure of left ventricular preload. Pulmonary artery catheterization is useful for monitoring the left ventricular filling.

RESPIRATORY SUPPORT

Respiratory support is needed when respiratory failure occurs, due to inadequate exchange of oxygen or CO_2 to meet metabolic needs, which is determined by the lack of improvement with oxygen therapy or the patient is tiring with an increasing pCO_2. The methods by which respiratory support can be given are:

- Continuous positive airways pressure (CPAP) in a spontaneously breathing patient. This is indicated when:
  - The patient is tiring with rising pCO_2
  - The patient is unable to maintain their own airway
  - Noninvasive ventilation is contraindicated
  - Need for endotracheal suction
- Noninvasive positive pressure ventilation (NIPPV) is an alternative via a face mask.

INOTROPIC SUPPORT

This is needed when the patient is not able to maintain the blood pressure and urine output, by normal crystalloid support. Different inotropic agents are used in various circumstances:

- Noradrenaline: Dose 0.01 to 0.4 mcg/kg/min IV infusion — acts on α_1-receptor causing vasoconstriction, in sepsis increases the renal blood flow and enhance urine production
- Adrenaline: Dose 0.01 to 0.30 mcg/kg/min IV infusion
  - In low doses: Acts on β-receptors (causes increase in heart rate and contraction)
  - In high doses: Acts on α-receptors (causes increased peripheral resistance)
- Dobutamine: Dose 1 to 25 mcg/kg/min via central vein—predominantly acts on β_1 (increases the heart rate and force of contraction) and also on β_2- and α_1-receptors (decreases peripheral and pulmonary vascular resistance)
- Dopamine: Acts directly on α, β and dopaminergic receptors and indirectly by releasing noradrenaline
  - Low dose: 0.5 to 2.5 mcg/kg/min (renal dose—increases renal and mesenteric blood flow. ↑ renal flow results in ↑ GFR and ↑ renal sodium excretion)
  - Moderate dose: 2.5 to 10 mcg/kg/min (cardiac dose—stimulates β_1-receptors causing ↑ myocardial contractility, stroke volume and cardiac output)
  - High dose: >10 mcg/kg/min (cardiac plus—stimulates α-receptors causing ↑ SVR, ↓ renal blood flow and ↑ potential for arrhythmias
- Dopexamine: Dose 0.25 to 0.5 mcg/kg/min—a synthetic analog of dopamine with β_1 activity with no α activity.

4

Shock

Shock is defined as inadequate organ perfusion and tissue oxygenation.

It is classified as:

- Hypovolemic shock
- Septic shock
- Anaphylactic shock
- Cardiogenic shock
- Neurogenic shock.

HYPOVOLEMIC SHOCK

This occurs due to loss of intravascular volume (blood or fluid), which results in activation of sympathetic nervous system, which causes tachycardia and vasoconstriction of skin, muscle and GI system, so that blood is directed to vital organs, e.g. brain, heart, etc.

The vasoconstriction of renal and splanchnic circulation causes renal failure, GI sloughing and hemorrhage. When shock persists, the perfusion of brain suffers causing confusion and aggression. The hyperventilation causes respiratory alkalosis, which is overtaken by metabolic acidosis due to poor tissue perfusion and anaerobic metabolism. The hypovolemic shock is classified depending on various factors (Table 4.1).

Table 4.1: Classification of hypovolemic shock

Class	*Blood loss*	*Pulse rate (per min)*	*Blood pressure*	*Urine output (ml/hr)*	*Resp rate (per min)*	*Consciousness level*	*Treatment*
I	<15% (up to 750 ml)	60–100	No change	>30	<20	No change	Crystalloid/colloid
II	15–30% (750 – 1500 ml)	>100	Pulse pressure reduced	20–30	20–30	Anxious	Crystalloid/colloid
III	30–40% (1500 – 2000 ml)	>120	BP fall	5–15	30–40	Confused	Blood
IV	>40% (>2000 ml)	>140	BP significant fall	Anuria	>40	Lethargic	Blood

Treatment of Hypovolemic Shock

- Control of hemorrhage
- Restoration of fluid volume.

- A fluid challenge of 1 to 2 liters should be given for adults and 20 ml/kg for children
- Rapid restoration of pulse and BP indicates a loss of <20 percent
- If the response is transient in spite of resuscitation, it is beter to intervene surgically.

SEPTIC SHOCK

Shock occurring due to severe sepsis called systemic inflammatory response syndrome (SIRS).

Symptoms and Signs

- Heart rate (>90/min)
- Temperature (>38 or 36)
- Repiratory rate (>20/min)
- Leukocyte count (>12000/cmm)

  Note: In the absence of infection, SIRS can be said to exist when two of the above criteria.

Pathogenesis

This results due to severe infection and inflammation, mediated by acute phace cytokines, which have a generalized effect distant to the site of original insult. Leukocytes adhere to endothelial cells via adhesion molecules leading to changes in vascular permeability and edema.

Treatment

- Resuscitation
- Identification of the source of sepsis
- Treatment of focus of sepsis (antibiotics/drainage of pus)
- Critical care management
  - Fluids
  - Oxygen administration
  - Vasopressors
  - Steroids
  - Activated protein C
  - Hemofiltration.

ANAPHYLACTIC SHOCK

Incidence and Etiology

This is a hypersensitivity reaction.

The usual causes are:

- Drugs
- Blood transfusion
- Radiological contrast.

Pathogenesis

It is mediated by immunoglobulin E (IgE).

Symptoms and Signs

- Urticarial rash
- Wheeze due to bronchospasm
- Tachycardia
- Fainting
- Rhonchi
- Hypotension.

Treatment

- Stop the cause if found
- Airway management and oxygen
- IV access and fluids
- Adrenaline (1 ml in 1:1000 IM)
- Antihistamine (IV)
- Bronchodilators.

CARDIOGENIC SHOCK

Occurs due to pump failure, commonly after myocardial infarction.

NEUROGENIC SHOCK

Occurs after spinal cord injury.

Pathogenesis

Occurs due to disruption of autonomic nervous system control over vasoconstriction, which results in a decrease in peripheral vascular resistance and blood pressure, with resultant bradycardia. Temperature control may also be lost.

Treatment

- Ventilation
- Fluid management
- Inotrope support.

5

Acute Respiratory Distress Syndrome

Incidence and Etiology

- Acute respiratory distress syndrome (ARDS) indicates the acute diffuse pulmonary inflammatory response to either direct or indirect insults from extrapulmonary pathology
  - Direct—via airway or injury to chest (e.g. aspiration, toxic gases, pneumonia)
  - Indirect—blood-borne insults (e.g. sepsis, polytrauma, severe burns, drugs)
- Frequently associated with multiple organ (kidney, liver, intestines) dysfunction.

Pathogenesis

- Pathogenesis is complex
- There is transudation of fluid in the lungs, thickening of alveolar capillaries resulting in ventilation/perfusion mismatch. The pulmonary artery wedge pressure remains normal, distinguishing it from pulmonary edema.

Clinical Presentation

They are not specific, but share many things with other pulmonary pathologies.

Relevant Investigations

- Blood gas analysis (PaO_2 / FiO_2 of less than 200 mm Hg)
- Chest X-ray shows bilateral diffuse infiltrates
- Pulmonary artery wedge pressure (less than 15 mm Hg).

Treatment

Current treatment is supportive and no specific therapy exists to modulate the sequence of events of ARDS:

- Monitoring
  - Monitoring of all vitals
- Ventilatory management
  - Mechanical ventilation to permit adequate oxygen uptake
- Nonventilatory management
  - Treatment of underlying risk factors
  - Enteral feeding
  - Maintenance of hemodynamic stability and cardiac output.

Section IV

Trauma

6. Polytrauma
7. Head Injuries
8. Facial Injuries
9. Spine and Spinal Cord Injuries
10. Thoracic Injuries
11. Abdominal Injuries
12. Urological Injuries
13. Male Genital Injuries
14. Female Genital Injuries
15. Hand Injuries

6

Polytrauma

INTRODUCTION

Management of a patient with polytrauma depends on proper and systematic clinical evaluation, which identifies immediate and potentially life-threatening conditions before the limb threatening ones, but does not omit the latter. This usually starts with a call from ambulance control.

The first couple of hours following injury are critical, as the patient is very vulnerable. This critical time period is called 'golden hour', which is usually spent in the place of accident or in the emergency department, making this period extremely crucial.

The trauma life support program consists of:

- Preparation
- Primary survey and resuscitation
- Secondary survey
- Continuous monitoring and evaluation.

PREPARATION

Prehospital Communication

A warning from the ambulance control ideally from the scene of accident provides essential information so that receiving personnel are ready to receive the patient of trauma, which provides assessment and treatment without delay. The essential prehospital information are:

- Nature of injury
- Mechanism of injury

- Number, age and sex of the casualties
- Consciousness level of the casualties
- Airway, breathing and circulatory status of casualties
- Treatment provided
- Estimated time of arrival
- Treatment options required.

TRAUMA RECEPTION TEAM

The make-up of this team varies between hospitals depending upon the resources and the time of the day. The trauma team should consist of:

- Trained medical personnel
- Supportive trained paramedical personnel
- A team leader who should assign specific tasks to each person.

The trauma reception team undertakes the following tasks:

Team Leader

- Coordinates the tasks assigned to the team members
- Questions ambulance personnel
- Assimilates clinical findings
- Determines investigations in order of priority
- Liaises with relatives and provide information to trauma team
- Liaises with specialists who are called.

Team Members

- Manage the airway
- Clear the secretions and intubate if necessary
- Manage circulation
- Establish infusion line quickly
- Takes blood for investigations
- Supportive measures
- Connect to monitors
- Urinary catheterization
- Jugular vein catheterization where necessary.

Receiving the Patient

Receiving and transferring the patient (Fig. 6.1) is very crucial and requires five people to do the job, which should be a well practiced procedure in order to protect the spinal cord if it is intact, and to prevent further injury if it is already compromised. During the transfer, the patient's head and neck are stabilized by one member of the team, three others lift from the side and the fifth member replaces the ambulance trolley with the resuscitation trolley.

Primary Survey and Resuscitation

The activities listed below are carried out simultaneously if there are enough personnel, if not should be done in alphabetical order (ABCDE).

- Airway and cervical spine control
- Breathing
- Circulation and hemorrhage control
- Disability
- Exposure.

Airway and Cervical Spine Control

A cervical spine injury should be assumed if the patient has been the victim of significant blunt trauma or if the mechanism of injury indicates that the cervical region may have been damaged. One member of the team needs manually to immobilize the cervical spine while talking to the patient which also assesses the airway.

FIG. 6.1: Patient with polytrauma

- If the patient is able to give a logical answer in a normal voice, the airway is assumed to be patent and the brain adequately perfused.
- If the patient gives an impaired or fails to reply, the airway could be obstructed, and immediate measures should be taken:
  - A simple chin lift will help in relieving the obstructing soft tissue usually the tongue
  - Saliva, blood, vomitus, tooth or other foreign bodies should be removed
  - Suction through rigid suction tube is necessary to remove secretions
  - In patients who vomit and regurgitate, head end of the bed should be dropped 20 degrees, allowing the secretions to drip down and facilitate removal by suction
  - A nasogastric tube may be inserted to aspirate the stomach contents and prevent further vomiting.

When it is established that the airway is clear and patent, 100 percent oxygen is provided via mask or by endotracheal intubation. Pulse oxymeter is connected to maintain good SaO_2, if needed with ventilatory support.

The neck should be examined for the following five signs which could indicate the presence of immediately life-threatening thoracic conditions (Table 6.1).

When below given signs in Table 6.1 are checked, the neck may be immobilized with appropriate collar if the patient is not restless, in a restless patient, semi-rigid collar is accepted.

All the multiple injured patients, particularly those who have injuries above the clavicle or a change in level of consciousness, should be treated as though they have a cervical spine injury, until it is ruled out.

Breathing

The clinical examination of chest consists of:

- Inspection
  - Marks and wounds
  - Respiratory rate
  - Inspiratory effort
  - Symmetry of chest movements

Table 6.1: Signs of life-threatening thoracic conditions

	Signs	*Conditions*
1.	Swellings and wounds	Vascular and airway injury
2.	Distended neck veins	Cardiac tamponade, tension pneumothorax
3.	Tracheal deviation	Tension pneumothorax
4.	Subcutaneous emphysema	Pneumomediastinum
5.	Laryngeal crepitus	Fracture of laryngeal cartilage

- Percussion
  - Assess ventilation at the periphery
- Auscultation
  - Assess defects in air movement.

> The respiratory rate and effort are sensitive indicators of underlying lung pathology, and requires to be monitored and recorded at frequent intervals.

Circulation and Hemorrhage Control

Bleeding can result due to:
- Fractures of long bones
- Vascular injuries
- Soft tissue injuries

Application of pressure is the best way of management of hemorrhage, as tourniquets increase intraluminal pressure of the vessels, distal ischemia and tissue necrosis. Application of tourniquet is undertaken in select situations, and when used the time of application of tourniquet has to be noted so that neighboring soft tissue is not jeopardized.

Recognition and Assessment of Hypovolemia

The assessment of hypovolemia has to be done in a systematic manner:
- Skin – color, clamminess and capillary refilling
- Vital signs – heart rate, blood pressure and pulse volume
- Consciousness level

> Isolated determinations of above parameters is unreliable as various organs try compensate at early stages, especially at extremes of age.

Common sites of occult bleeding are:
- Chest
- Abdomen and retroperitoneum
- Pelvis
- Long bone fractures
- External bleed into splints and dressings.

Fluid Resuscitation

Once any overt bleeding is controlled, it is necessary to maintain the circulatory volume. This is done by administering a warm crystalloid followed by blood transfusions, maintaining the radial pulse and blood pressure. This administration of fluids is done by peripheral venous cannulation, and when not possible done through central venous catheterization.

- Vital signs return to normal after less than 2 liters of fluid are administered, when the lost blood is less than 20 percent of the blood volume
- Transient responders who are actively bleeding or recommence bleeding during the resuscitation, and the improved vital signs deteriorate indicating loss of over 30 percent of the blood volume
- Little or no response indicates that the loss is more than 40 percent or no hypovolemia.

Disability

Disabilities when occur in a patient with trauma shows seriousness. It can occur in:

- Hypoxia
- Hypovolemia
- Hypoglycemia
- Increased intracranial pressure.

Exposure

The patient's clothes have to be removed by cutting through the seams so that there is minimal patient movement. All clothes are removed only after adequate intravenous access is established, as a rapid removal of tight trousers can precipitate sudden hypotension due to the loss of the tamponade effect in a hypovolemic patient.

Once stripped, trauma victims should be kept warm with blankets when not being examined. Now the patient is rolled on and the spine examined from base of skull to the coccyx, with a rectal examination.

What to look for during rectal examination in a trauma victim:

- Is the sphincter tone present?
- Is the rectal wall breached?
- Is the prostate in a normal position?
- Is there blood on the examiner's finger?

SECONDARY SURVEY

Once the immediately life-threatening conditions are recognized and treated, the whole of the patient is assessed. This requires head to toe, front to back assessment along with detailed medical history and appropriate investigations.

If the patient deteriorates at any stage, the airway, breathing and circulation must be immediately re-assessed as described in the primary survey.

CLINICAL EVALUATION

Eliciting History

The history is elicited quickly and in concise manner, concentrating on:

- Nature of injury
- Mechanism of injury
- Treatment provided
- Medical history.

Physical Examination

The physical examination should be done in a systematic manner.

Scalp

The scalp is examined for:

- Lacerations
- Swellings
- Depressions

The examination is carried out by inspection and palpation (by running the fingers on the scalp).

- It should be remembered to examine the occipital region, which is usually missed out
- Blind probing should be avoided as it may further damage the underlying structures
- In children, bleeding from scalp lacerations can cause hypotension and efforts to stop bleeding (application of pressure, applying self-retaining retractors) have to be taken immediately.

Neurological State

The neurological state of the patient by the Glasgow Coma Score, the papillary responses and the presence of lateralizing signs should be assessed. This examination should be done frequently so that any deterioration is detected early.

Base of Skull

Since the skull base lies along a diagonal line running from the mastoid to the eye, the signs of a fracture in this region are also found along this line.

Signs of a base of skull fracture

- Bruising over the mastoid (Battle's sign)
- Panda's eyes
- Blood and CSF rhinorrhea
- Blood and CSF otorrhea
- Hemotympanum
- Scleral hemorrhage with no posterior margin
- Subhyoid hemorrhage.

> - Battle's sign and Panda's eyes appear after about 12 to 36 hours, and is not a reliable sign in a resuscitation room.
> - CSF rhinorrhea may be missed as it is invariably mixed with blood.

Neck

If any deformity is found, it is necessary to splint the neck preferably with a collar.

Eyes

- Hemorrhage
- Foreign bodies (including contact lenses)
- Papillary response and corneal reflexes (in unconscious patient)
- Visual acuity (in conscious patient).

Face

The face is examined by systematic inspection and palpation, and look for:

- Soft tissue injuries
- Fractures of midface
- Fractures of mandible
- Missing or lost teeth.

> - Midface fractures may be associated with fractures of base of skull
> - Mandibular fractures can cause airway obstruction due to loss of stability of tongue.

Thorax

The thorax is examined for

- Soft tissue injuries
- Fractures of clavicles and ribs
- Crepitus (e.g. surgical emphysema)
- Movements of chest (e.g. flail chest).

Abdomen

The abdomen is examined in a very systematic manner, without forgetting to examine the pelvis and perineum. Percussion is an ideal way to determine the injuries of liver and spleen. In penetrating injuries, if bowel is exposed, it should be covered immediately with sterile cloth.

Urine output measurement is a good indicator in a shocked patient, and this requires catheterization for accurate measurement. Transurethral catheterization without any difficulty may indicate that is no severe urethral disruption. If the patient is not able to urinate, and catheterization is not possible, severe urethral injury should be suspected, and suprapubic catheterization should be done. The signs of urethral injury in a male are:

- Bruising around the scrotum
- Blood at the urethral meatus
- High-riding prostate.

Extremities

The limbs are examined in the traditional manner of inspection, palpation and active and passive movement.

- The fractures should be assessed and the blood loss estimated
- Distal loss of arterial pulsation may be due to profound shock or arterial injury.

DEFINITIVE CARE

Once the patient is adequately assessed and successfully resuscitated, he is moved to the next stage for definitive care, but they are vulnerable during transfers as monitoring becomes difficult during this time.

Before transfer, it is necessary to:

- Close an open wound with sterile pads
- Correct the deformity and splint the limb.

Note:

- Intra-abdominal bleed should be suspected if the patient is hemodynamically unstable, especially if the lower six ribs are fractured or there are marks on the abdominal surface
- In unexplained hypotension, pelvic bone injuries should be suspected. Compression or distraction test is useful
- Clinical examination may not be reliable in neurologically unstable patients, and investigations like ultrasound and CT or MRI may be needed. Diagnostic peritoneal lavage may be relied upon in certain situations like massive hemoperitoneum.

7

Head Injuries

Head injuries are classified into:

- Brain injury
- Fractures of skull
- CSF fistula
- Injuries of cranial nerves
- Vascular injuries (rare)
- Scalp injuries.

BRAIN INJURY

For the purpose of understanding the pathophysiology, brain injury is subdivided into:

- Primary brain injury and
- Secondary brain injury

Primary brain injury is further classified into:

- Concussion brain
- Diffuse axonal injury
- Primary brainstem injury
- Contusions and lacerations of brain

These injuries occur depending on the:

- Severity of impact
- Direction of impact force

- Movement of head:
  - *Type of injury:* closed or penetrating injury. Mild force causes ***concussion*** of brain characterized by transient loss of consciousness, Post traumatic amnesia, confused state and patient recovers completely
  - Severe force causes ***Diffuse Axonal Injury*** characterized by prolonged unconsciousness and neurologic deficits
  - More severe force causes ***Primary Brain stem Injury***. With extremely severe impact force, fatal injury occurs with death at the accident spot itself
  - To-and-fro movement of head shakes various parts of the brain within the skull causing different injuries: gray matter moves over white matter; subcortical white matter moves over basal ganglia; brain surface hits against the rough floor of cranial fossa floor and sharp edges of falx and tentorium leading to ***contusions of brain***. When the pia-arachnoid is torn it is termed ***laceration***.
- With impact force/acceleration-deceleration force, brain moves within the cranial cavity—cerebrum moves over brainstem leading to deformation of neuronal-synaptic membranes at central reticular core.

Secondary Brain Injury

- Hypoxic-Ischemic injury to brain because of impaired breathing, aspiration, airway obstruction, chest injury, cervical spine injury, hypotension, etc.
- Intracranial hematomas (viz., acute extradural hematoma, acute subdural hematoma, acute intracerebral hematoma) or expanding contusions and brain edema can cause brain shifts and increased intracranial pressure
- Metabolic abnormalities of glucose, sodium, acid-base, etc.

Acute extradural hematoma Vs Acute subdural hematoma:

- Acute extradural hematoma occurs at the site of impact and hence the scalp injury or fracture skull will give clue to the location of extradural hematoma
- Bleeding is usually from the middle meningeal branches or from fracture edges
- Acute subdural hematoma is 6 to 8 times more common than extradural hematoma and is often due to acceleration and deceleration injury
- Subdural hematoma can be contracoup, i.e. its location can be diagonally opposite to the site of scalp injury. The bleeding is from the cortical veins
- Due to associated severe primary brain injury, prognosis is in general poor compared to isolated extradural hematoma.

FRACTURES OF SKULL

- Skull fractures are classified into:
  - ***Fractures of skull vault*** which can be:
    - Linear fracture (Fissured fracture)
    - Depressed fracture
  - ***Fractures of skull base*** which can involve:
    - Anterior cranial fossa
    - Middle cranial fossa
    - Posterior cranial fossa
- They can be also be classified into:
  - Simple (or closed) fractures
  - Compound (or open) fractures
    (depending on the absence or presence of communication to atmospheric air through breach of skin or air containing cavities in the skull)
- Fractures of skull vault will indicate the location of underlying extradural hematoma when present
- Compound depressed fractures are potential source for spread of infection leading to brain abscess, meningitis, epidural /subdural abscess
- Depressed fractures of vault can lead to direct brain contusions and hematomas.

Skull fracture indicates that the force of injury was severe but does not indicate the severity of brain injury.

Symptoms and Signs

- Fractures of anterior cranial fossa (Fig. 7.1) can cause:
  - Ecchymosis of the upper eyelids appearing after about 24 hours
  - Subconjuctival hemorrhage in the upper eyelid in anterior skull base fractures is pointed towards the cornea and posterior limit will not be seen
  - In contrast in "Black Eye" where the injury is directly over orbit, the ecchymoses occur in both upper and lower eyelids and appear within minutes of injury
  - The subconjuctival hemorrhages are actually conjuctival in plane and will move with conjunctiva
- Cranial nerve palsies occur in fractures of the base of skull across or near the cranial nerve exit foraminae:
  - Fracture of temporal bone can explain 7th and 8th nerve palsies
  - Fracture of superior orbital fissure can explain spread of injuring force close to 3rd, 4th and 6th nerves
  - Fractures near jugular foramen can cause lower cranial nerve palsies
- CSF rhinorrhea occurs in fractures involving cribriform plate or paranasal sinuses

FIG. 7.1: Fracture of anterior cranial fossa

- CSF otorrhea occurs in fractures of temporal bone Eustachian tube CSF rhinorrhea. CSF leaks have the risk of meningitis
- Pulsatile proptosis in head injury can be due to carotico-cavernous fistula or more commonly comminuted fracture of orbital roof with normal pulsation of brain being transmitted to orbit. In carotico-cavernous fistula bruit may be present over orbit.

Note: Cranial nerve palsies can occur without skull fracture, e.g. olfactory nerve injury or optic nerve injury.

Relevant Investigations

X-rays (AP, lateral and oblique views) and CT of skull (Figs 7.2A to D) are diagnostic.

Treatment

- Closed depressed fractures over forehead may need elevation for cosmetic purposes
- Compound fractures will need debridement to prevent brain abscess.

FIG. 7.2A: CT—Fracture of frontal bone

FIG. 7.2B: CT—Fracture

FIG. 7.2C: CT—Fracture of frontal sinus

FIG. 7.2D: CT—Subdural hematoma

CSF FISTULAE

Incidence and Etiology

Leak of CSF is caused by:

- ❖ Fractures involving cribriform plate or paranasal sinuses can lead to CSF rhinorrhea
- ❖ Fractures of temporal bone can lead to CSF otorrhea and through Eustachian tube CSF rhinorrhea.

Complications: Meningitis

Symptoms and Signs

Watery discharge is pathognomonic of CSF leak.

Relevant Investigations

CT of paranasal sinuses after intrathecal contrast is useful.

Treatment

- Lumbar pucture and CSF drainage intermittently for 10 – 20 days
- Surgical repair is necessary if leak persists.

SCALP INJURIES

Scalp injuries indicate the sites of impact force.

They are of three types:

1. Contusions
2. Lacerations
3. Hematomas
   - ❖ Subpericranial—Confined to an area of one cranial bone and fixed
   - ❖ Subaponeurotic or subgaleal—More diffuse and extend beyond the margins of bones
   - ❖ Subcutaneous—Superficial and moves with the scalp.

- Fractures beneath scalp injury in unconscious patients suggest the possibility of underlying extradural hematoma
- Scalp injury over occipital region may give clue to posterior fossa hemorrhage
- The center of scalp hematomas may liquefy in the center after a few days and often give a false impression of depressed fracture to the palpating finger.

Symptoms and Signs

- Bleeding from an open wound
- Swelling in the head with or without external injury to the scalp.

Treatment

Immediate suturing of the wound, as the vessels are prevented from normal contraction by fixation of their walls to fibrous stroma of scalp.

Note: When there is scalp laceration, before wound closure, depressed fracture has to be excluded clinically under aseptic precautions.

8

Facial Injuries

FRACTURES OF MIDFACIAL SKELETON

- The fracture of facial skeleton is usually due to direct injury
- Types of fractures of midface are:
  - Floor of the orbit (Blowout fracture)
  - Zygomatic
  - Nasal
  - Maxillary [LeFort classification (Fig. 8.1)]
    - LeFort I (low level fracture): This is horizontal fracture above the level of the nasal floor (Guerin's fracture, floating fracture)

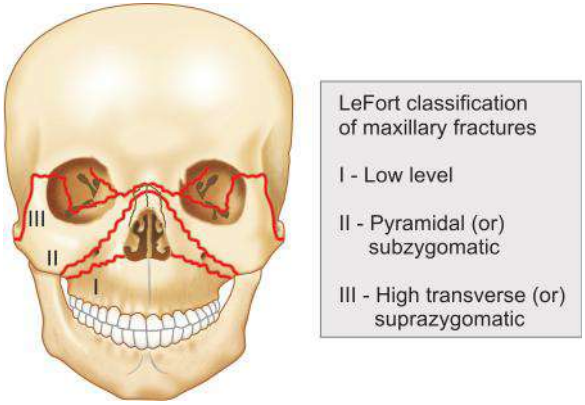

FIG. 8.1: Fractures of maxilla

- LeFort II (pyramidal or subzygomatic fracture): The fracture runs from the thin middle area of the nasal bones down either side crossing the maxillary processes into the medial wall of each orbit
- LeFort III (high transverse or suprazygomatic fracture): The fracture runs from near the frontonasal suture transversely backwards, parallel with the base of the skull and involves the full depth of the ethmoid, including the cribriform plate.

FRACTURES OF MANDIBLE

Fractures of the mandible (Fig. 8.2) can be divided according to the anatomical location into 8 types:

1. Dentoalveolar
2. Condylar
3. Coronoid
4. Ramus
5. Angle
6. Body
7. Parasymphysis
8. Symphysis

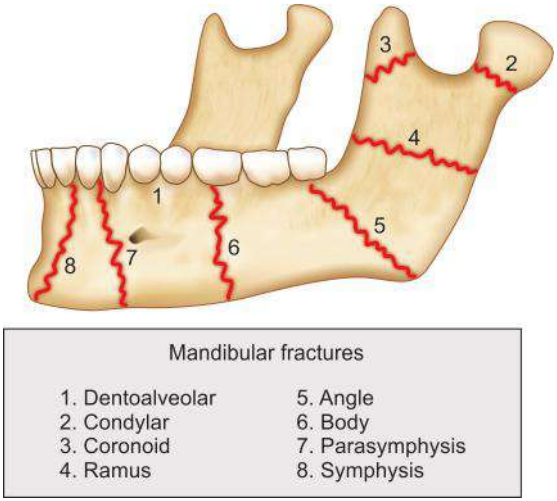

FIG. 8.2: Fractures of mandible

The mandibular fractures can be:

- Unilateral
- Bilateral
- Multiple
- Comminuted.

Clinical Presentation

History

- History of injury and hearing or feeling of a bone crack
- The nature of impact (direct violence, indirect violence or excessive muscular contraction) should be determined.

Symptoms

- Pain and loss of function
- Diplopia and enophthalmos are present in LeFort III type
- But loss of function may not be a feature of impacted fracture.

FIG. 8.3A: Ecchymosis and nasal bleed in injury of midface

FIG. 8.3B: Subconjunctival hemorrhage

Signs

- Tenderness of the bone on applying pressure
- Swelling and ecchymosis (Fig. 8.3A)
- Nasal bleed (Fig. 8.3A) and CSF rhinorrhea are common in associated base of skull fractures
- Subconjuctival hemorrhage occurs in orbital bone fractures (Fig. 8.3B)
- Blebs on the skin overlying the fractured site
- Deformity: This is elicited by inspection, palpation and measurement. This is the most important sign of a fracture. The deformity in case of a flat bone is either depression or elevation and in case of long bone this is angular, lateral, longitudinal or rotatory. Hard palate is deformed in Le Fort I fracture (Fig. 8.4)
- Dental alignment is lost in displaced fractures (Fig. 8.5).

FIG. 8.4: LeFort I fracture

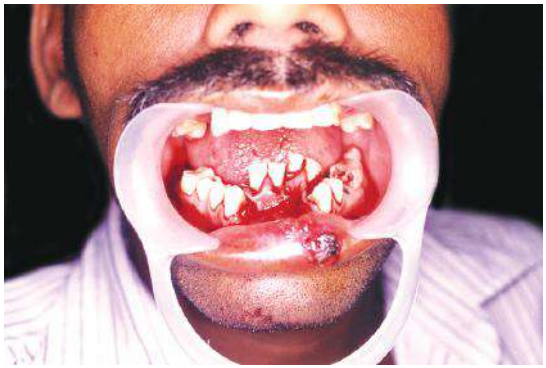

FIG. 8.5: Displaced mandibular fracture

FIG. 8.6A: X-ray—Fracture of ramus of mandible

FIG. 8.6B: X-ray—Fracture of condyle of mandible

FIG. 8.6C: X-ray—Zygomatic fracture

FIG. 8.7: CT—Fracture of orbital bones

- Abnormal mobility: This is present in case of impacted fractures
- Crepitus: Also present in case of impacted fractures and should be elicited gently
- Loss of function: For example, if the maxilla or the mandible is fractured, the patient may not be able to masticate properly.

Relevant Investigations

X-rays (Figs 8.6A to C) and CT (Fig. 8.7) are diagnostic.

Treatment

- Undisplaced fractures need conservative management
- Displaced fractures require surgical treatment, fixation with plates and screws (craniomaxillofascial plating) and interdental wiring, with the aim to restore precise anatomical alignment.

9

Spine and Spinal Cord Injuries

INJURIES TO BONY AND LIGAMENTOUS SPINE

Incidence and Etiology

- When an overwhelming force strikes and compresses the spine, the bone crumbles resulting in fracture of spine
- The type of the vertebral fracture depends on the direction of injuring force and the curvature of spine at the time of injury
- With increasing force of injury, two or three columns of spine get injured leading to instability and dislocation
- Generally compression forces cause fracture while distraction forces cause ligament injury
- The injuring forces can cause injury to spinal cord/root.

Symptoms

- When the spinal cord is not affected, pain on movement, stiffness and tenderness are the symptoms
- When the spinal cord is affected, neurological deficit occurs, depending on the level and completeness of damage to the cord
  - Limb paralysis like paraplegia or quadriplegia may be the presenting symptoms
  - Respiratory, circulatory and urinary bladder dysfunctions may be superadded.

Signs

Clinical examination is to determine the level and type of injury

- External injuries give clue to the mechanism of injury (e.g. chin injury may indicate hyperextension injury, occipital injury may indicate flexion-distraction injury, vertex injuries may indicate vertebral compression).

Relevant Investigations

X-rays (Figs 9.1A and B), CT (Fig. 9.2) and MRI are essential to assess the extent of injury.

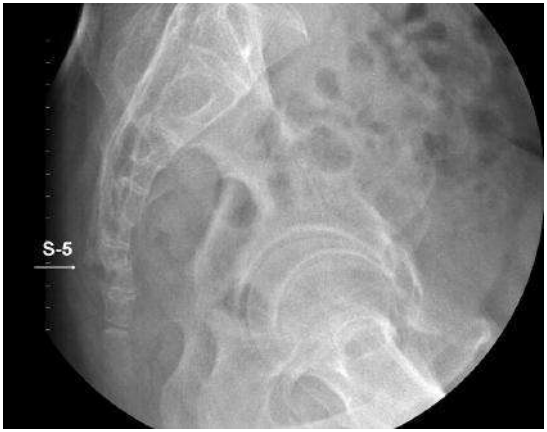

FIG. 9.1A: X-ray—Fracture of 5th sacral vertebra

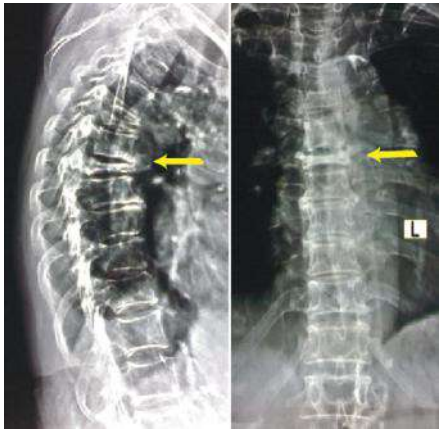

FIG. 9.1B: X-ray—Compression fracture vertebra

FIG. 9.2: CT cervical spine—Fracture body of C3 vertebra

Treatment

- Rest to the injured area may be sufficient with analgesic support. Many of them recover completely
- Surgical decompression and stabilization are mandatory for incomplete cord injuries. Though, the procedures are useful for bony stability in complete cord injuries, the neurologic recovery is poor.

INJURIES OF THE SPINAL CORD

Incidence and Etiology

- During the injuries of the spinal cord, the neurons suffer primary damage on impact and are prone to secondary injuries from hypoxia, hypotension, hematoma and swelling
- When the damage is severe, spinal cord function at and below the level of the lesion is abolished
- Spinal cord injuries are usually associated with injuries of vertebral column, usually fracture or fracture-dislocation
- The mechanisms of injury to the spinal cord /root are by:
  - Primary injuries: The primary injuring force itself can cause violent movement and stretching or disruption of cord with or without abnormal bony displacements
  - Secondary injuries: The fractured segments of vertebrae, extruded disk, infolded ligaments and intraspinal hematomas can cause secondary compression of cord or roots causing neurologic deficit.

Symptoms and Signs

- Paralysis occurs below the level of the lesion, e.g. thoracic cord injury leads to paraplegia and cervical cord injury leads to quadriplegia
- Flaccid paralysis with no sensation indicates complete lesion with a poor prognosis
- Signs suggestive of spinal cord injury are given in Table 9.1
- In an unconscious patient, some symptoms suggest spinal cord injury (Box 9.1).

Table 9.1: Spinal cord injuries and their signs

Complete cord injury	Total paralysis with loss of sensation below level of injury
Cord hemisection (Brown-Sequard syndrome)	Ipsilateral paralysis with contralateral loss of sensation below level of lesion
Central cord syndrome	Greater motor loss in the upper limbs than the lower limbs Variable sensory loss below level of lesion
Anterior cord syndrome	Paralysis, loss of pain/temperature sensation below level of lesion Proprioception and vibration preserved

Box 9.1: Signs suggestive of spinal cord injury in an unconscious patient

- Diaphragmatic pattern of breathing
- Unexplained hypotension
- Flaccid paralysis
- Reduced anal tone
- Urinary retention/Priapism

Relevant Investigations

X-rays and CT are useful in diagnosis.

Treatment

- The treatment should focus on maintaining stability of the vertebral column either by external or internal fixation
- Recovery is variable and rehabilitation should start early for better results
  - Establish IV access and give good volume load to support blood pressure
  - Vasopressors may be needed to maintain circulation
  - Tracheal intubation and assisted ventilation may be needed to support ventilation
  - Nasogastric intubation may be needed for gastric decompression
  - Urethral catheterization is needed.

10

Thoracic Injuries

RIB FRACTURES

Incidence and Etiology

- This constitutes the most common chest injury
- Minor fractures are those confined to one or two ribs
- Mechanism of injury
  - Upper rib injuries involve major energy transfer and are often associated with injuries to major vessels, brachial plexus and tracheobronchial tree
  - Fractures of lower ribs are frequently associated with liver and splenic injuries
  - Rib fractures in the elderly can occur after relatively low energy transfers as the bones are of low density and chest wall compliance is poor.

> Fractures of brittle ribs of elderly patients cause very little underlying injuries, whereas, flexible rib injuries of younger individuals cause severe injury without obvious fractures.

Symptoms

Severe pain on deep inspiration and coughing, poor inspiratory effort, and progressive atelectasis and pneumonia due to underlying lung contusion.

Signs

- Crepitus, and bony tenderness
- Skin bruises should give the suspicion
- The hallmarks of rib fractures are intense pain, poor inspiratory effort, and progressive atelectasis and pneumonia due to underlying lung contusion.

> - Apical rib fractures are associated with injury to great vessels
> - Mid zone rib fractures are associated with pulmonary contusion
> - Basal rib fractures are associated with abdominal visceral injuries (liver, spleen).

Relevant Investigations

- Chest X-ray shows the site and number of fractures (Figs 10.1A and B), underlying pleural and lung injuries
- Chest CT (Figs 10.2A and B) gives clearer view of fractures.

Treatment

Centers around pain management:

- Oral and parenteral analgesics
- Intercostal nerve blocks
- Epidural analgesia especially in elderly or patients undergoing abdominal surgeries.

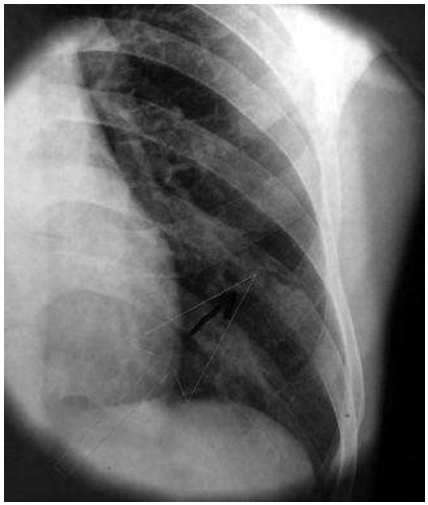

FIG. 10.1A: X-ray—Fracture rib

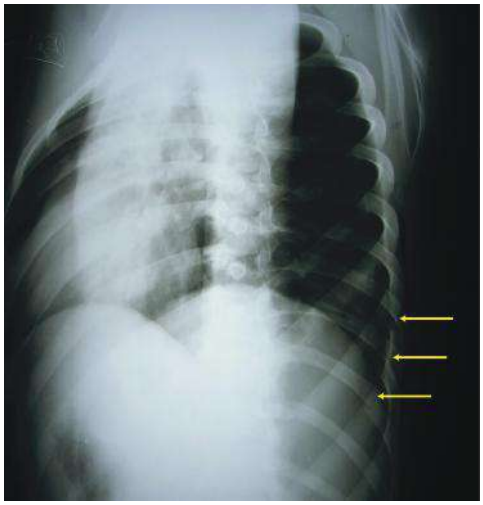

FIG. 10.1B: X-ray—Fractures of lower ribs

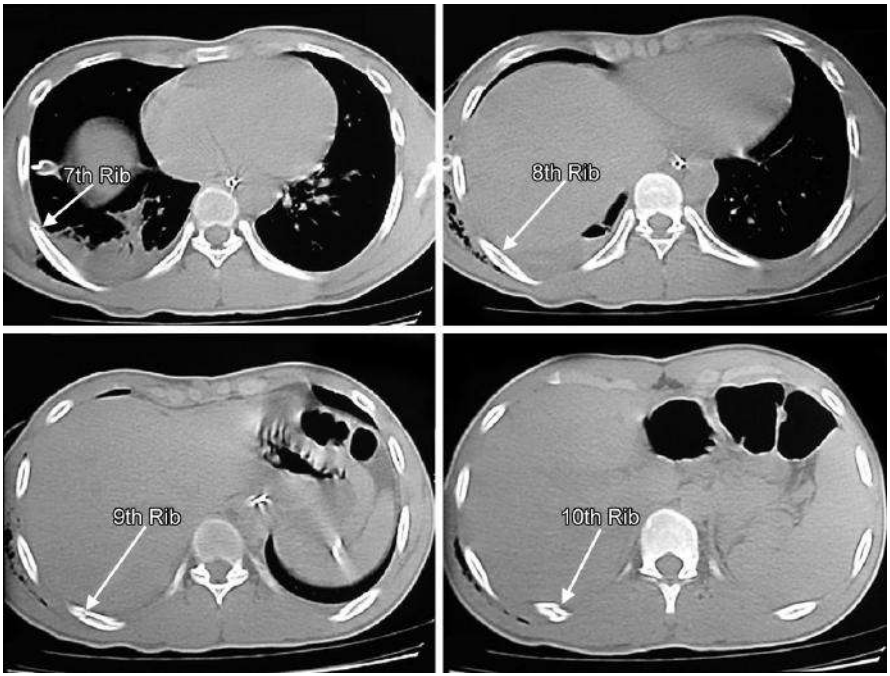

FIG. 10.2A: Chest CT—Fractures of 7, 8, 9 and 10th ribs

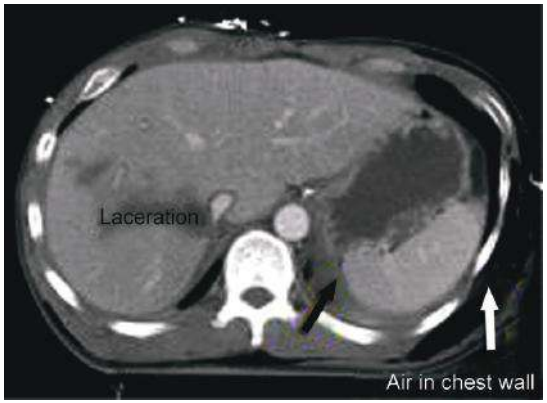

FIG. 10.2B: CT—Laceration of liver and perisplenic collection

FLAIL CHEST

Incidence and Etiology

- When three or more ribs are fractured, each in more than one place, producing a free floating section of the chest wall with or without separation of the costochondral junction it is called a flail chest (Fig. 10.3)
- The flail segment interferes with the ventilatory function by ineffective chest wall motion (paradoxical movement) i.e. movement inward with inspiration and outward with expiration, producing pain and splinting and thereby a fall in tidal volume, hypoxia and hypercarbia.

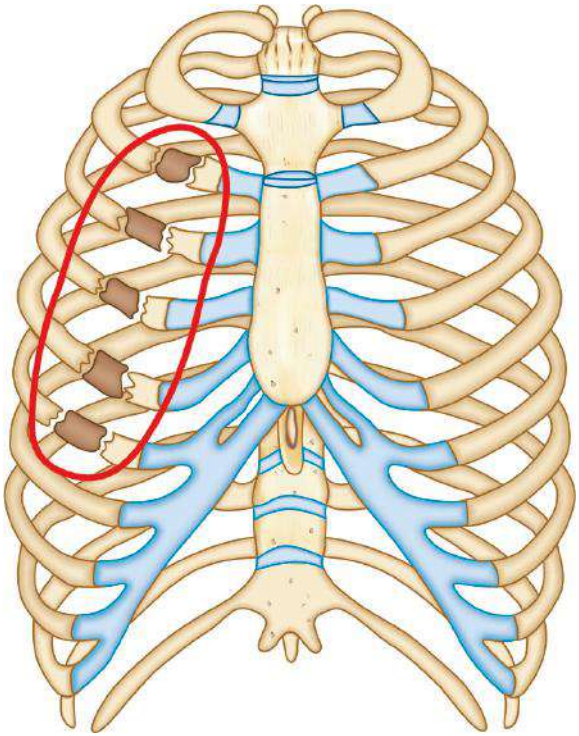

FIG. 10.3: Flail chest

- Other causes for flail chest are:
  1. Traumatic disruption of ligaments and cartilages of ribs—not seen on X-ray.
  2. Destruction of ribs from malignant disease, e.g. multiple myeloma.
  3. Metabolic disease—osteitis fibrosa cystica.
  4. Nonclosure of median sternotomy wound.

Symptom

Dyspnea.

Sign

Paradoxical respiration and hypoxia.

Relevant Investigations

- Chest X-ray—to assess fracture, lung injury, hemopneumothorax
- Arterial blood gas analysis—to aid treatment of respiratory insufficiency (ventilation perfusion mismatch).

Treatment

Treatment of flail chest is shown in Table 10.1.

Table 10.1: Treatment of flail chest

Segment of flail	*Respiratory distress*	*Respiratory function*	*Treatment*
Small	No	Good	Pain relief and observation
Moderate	Severe	Moderate	Mechanical ventilation and analgesics
Large	Severe	Bad	Chest wall reconstruction and mechanical stabilization

STERNAL FRACTURE

Incidence and Etiology

- Occurs mostly at the manubriosternal junction and is associated with very high velocity trauma
- Injury to aorta, esophagus, bronchi, myocardium and spine need to be kept in mind.

Symptom

Severe pain over the anterior chest wall.

Sign

Crepitus over the fracture site is characteristic.

Relevant Investigations

Chest X-ray lateral view and CT (Fig. 10.4) demonstrate the fracture.

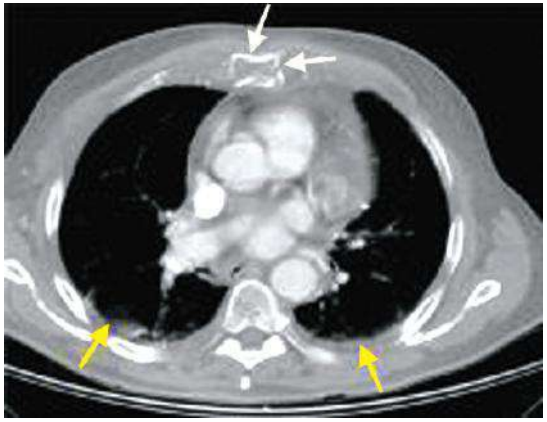

FIG. 10.4: CT—Fracture sternum

Treatment

- Sternal fractures can be managed conservatively with pain relief
- Rarely, in case of persistent chest wall instability, fixation may be necessary.

PNEUMOTHORAX

Incidence and Etiology

- Defined as air in the pleural cavity
- The types (Fig. 10.5) are:
  - *Closed pneumothorax:* Air in the pleural cavity and has no external communication (e.g. rupture of emphysematous bulla) or from outside
  - *Open pneumothorax:* Air in the pleural cavity has external communication (e.g. penetrating chest wall injury or rib fracture)

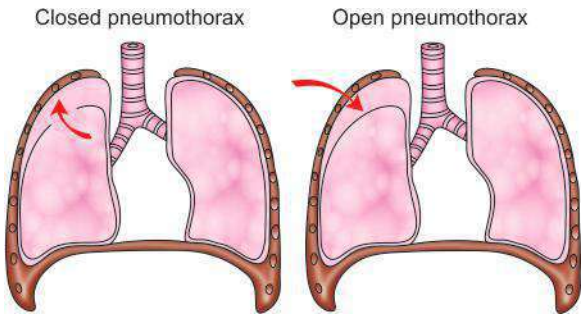

FIG. 10.5: Types of pneumothorax

- *Tension pneumothorax:* Continued entry of air into the pleural cavity, increasing the intrapleural pressure above the atmospheric pressure, which results in the shift of the mediastinum away from the side of injury.

Symptoms

Chest pain, dyspnea and tachycardia.

Signs

- On examination the neck veins are distended, the trachea and apex beat are shifted away from the side of tension, breath sounds become distant or absent on the side of tension, due to the presence of air between the chest wall and the lung substance
- The chest on the affected side is more resonant (DD—hemothorax, hydrothorax) on percussion.

Relevant Investigations

- Chest X-ray (Figs 10.6A to C) is conclusive, with shift of mediastinal structures away from the side of the pathology, with air shadow lateral to the lung parenchyma on the side of the pathology
- CT (Figs 10.7A and B) is diagnostic.

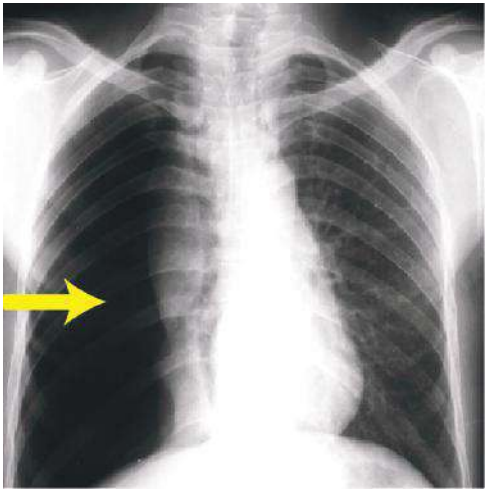

FIG. 10.6A: Chest X-ray—Right pneumothorax

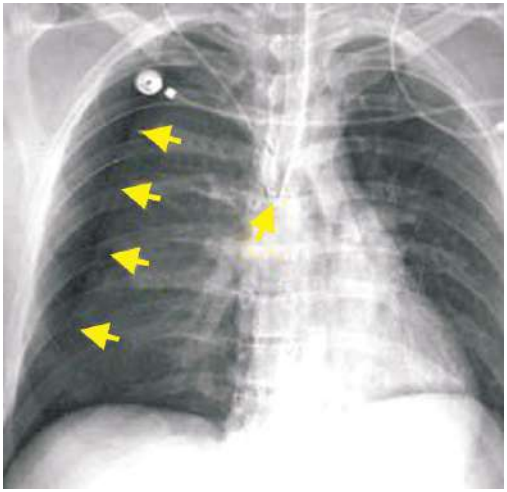

FIG. 10.6B: Chest X-ray—Right pneumothorax in patient on ventilator

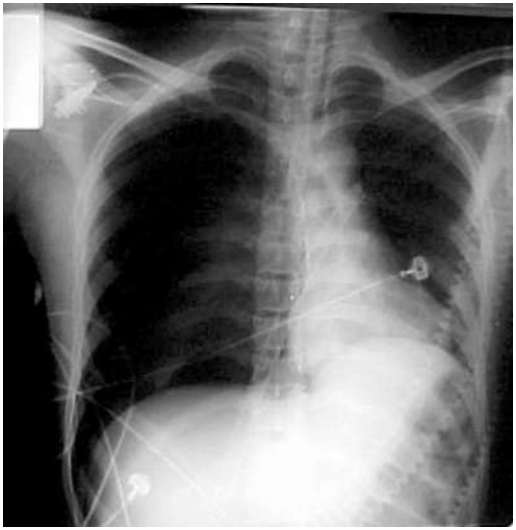

FIG. 10.6C: Chest X-ray—Tension pneumothorax

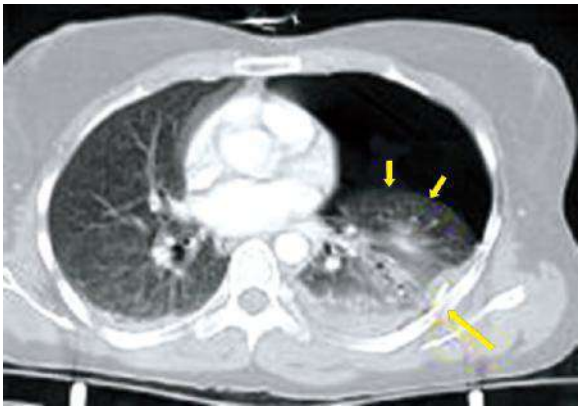

FIG. 10.7A: CT—Left pneumothorax with fractures of ribs

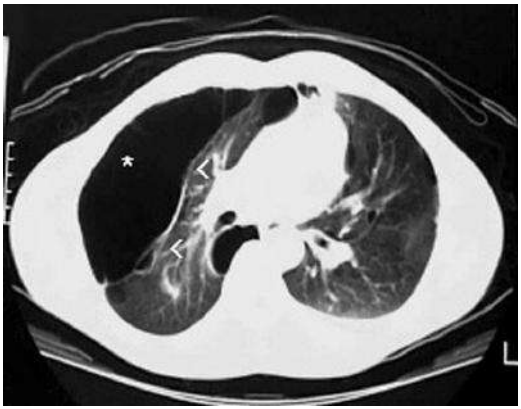

FIG. 10.7B: CT—Right tension pneumothroax

Treatment

- *Open pneumothorax:* The external wound is closed with a tape to convert it into a closed variety, supported by intercostal drainage
- *Closed and tension pneumothorax:* Simple aspiration of air from the pleural space followed by tube thoracostomy.

- Large chest wall wounds more than 75 percent the diameter of trachea allow preferential air entry through the chest wall. Any attempt to ventilate leads to movement of air in and out of the defect. No ventilation is achieved, and severe respiratory compromise occurs
- All traumatic pneumothoraces should be drained
- Massive air leaks may require bronchoscopy to exclude bronchial rupture
- Bronchial rupture should be suspected in the presence of deceleration injury, mediastinal widening, hemoptysis, first rib and clavicular fractures.

SURGICAL EMPHYSEMA

Incidence and Etiology

Defined as air in the subcutaneous tissues due to the air entry from the injured lung or external injuries like fractured rib.

Pathogenesis

The entrapped air gradually spreads along the fascial planes into the neck, mediastinum but, rarely down into the scrotum producing, a pneumoscrotum.

Symptoms and Signs

Subcutaneous tissues appear swollen and crepitus is a classical finding.

Relevant Investigations

- X-ray is diagnostic and reveals the air shadow in the subcutaneous plane, and also the underlying cause (e.g. rib fracture)
- CT (Fig. 10.8) is diagnostic.

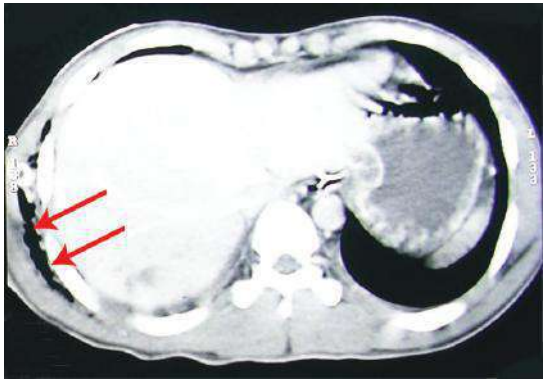

FIG. 10.8: CT—Right sided surgical emphysema with rib fracture

Treatment

- Small emphysema resolves spontaneously
- Hemodynamic instability warrants surgical intervention
- Treating the underlying cause.

HEMOTHORAX

Incidence and Etiology

Defined as blood in the pleural space, which is usually due to external (blunt or penetrating) trauma.

Symptoms

Chest pain, dyspnea and tachycardia.

Signs

- Tachycardia
- Neck veins are distended
- Trachea and apex beat are shifted away from the side of tension
- The chest on the affected side is less resonant or dull (DD—pneumothorax) on percussion
- Breath sounds become distant or absent on the side of lesion, due to the presence of blood between the chest wall and the lung substance
- Hypotension and shock may be evident depending on the amount of blood loss (bleeding from lung parenchyma is usually small but those from the intercostals and internal mammary arteries may be large).

Relevant Investigations

- Chest X-ray (Fig. 10.9A) is conclusive, with shift of mediastinal structures away from the side of the pathology, with haziness with obliteration of costophrenic angle
- CT (Fig. 10.9B) is diagnostic.

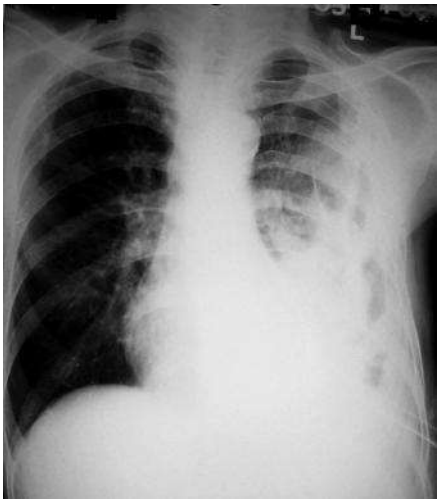

FIG. 10.9A: Chest X-ray—Left hemothorax

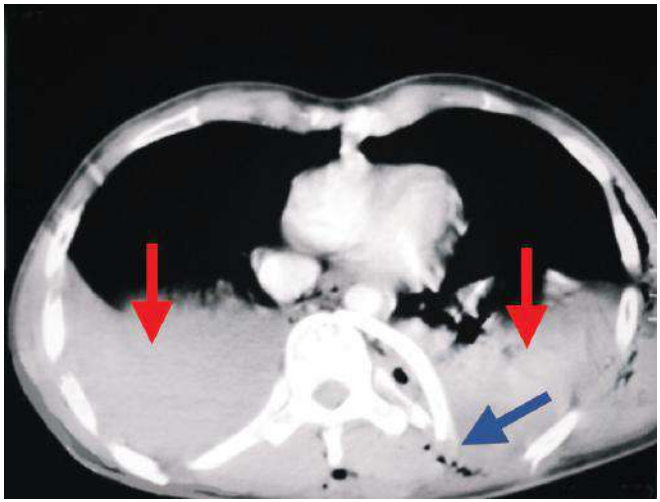

FIG. 10.9B: CT chest—Bilateral hemothorax (Red arrows) with left 10th rib fracture (Blue arrow)

Treatment

- Small collections of blood may be aspirated under aseptic conditions
- Large collections need intercostal drainage
- Thoracotomy is needed for massive bleeds (initial bleed of more than 1 – 1.5 liters), or continued bleeds of more than 200–300 ml/hr, to control the source of bleeding.

- Early drainage is the key to success in the management of hemothorax
- Once clot gets established, thoracotomy is needed
- While draining hemothorax, it is advisable to use large drains (28 – 32 F)
- Initial drainage of >600 ml or continued drainage of > 150 ml/hr will need thoracic surgical referral.

PULMONARY CONTUSION/LACERATION

Incidence and Etiology

- While contusions of the lungs are produced by blunt chest injury with hemorrhage and edema in the lung parenchyma, lung lacerations are due to penetrating injuries
- There may be associated injury to larger airways:
  - Blunt injury usually produces injuries within 2.5 cm of the carina
  - Penetrating injuries may be at any level.
- Alveolar microhemorrhages are responsible for the poor ventilatory status
- Tracheobronchial injuries may coexist.

Complications: Mediastinal emphysema in case of major airway injury and hemopneumothorax in case of peripheral bronchial injuries.

Symptoms and Signs

- Dyspnea, tachycardia and chest pain
- Small lacerations produce no symptoms.

Relevant Investigations

- Chest X-ray is diagnostic, which shows vague opacification in the injured area (usually within 1-2 hours of injury)
- CT (Fig. 10.10) and MRI are conclusive
- Bronchoscopy is needed to evaluate tracheobronchial injuries.

FIG. 10.10: CT chest—Bilateral pulmonary contusion

Treatment

- Most lacerations of the lung resolve spontaneously and need no treatment
- Large lacerations may require resuscitation and tube thoracostomy
- Bronchial injuries involving more than 1/3rd circumference of the bronchus require surgery
- Mechanical ventilation is warranted when there is profound V/Q mismatch.

INJURIES OF THORACIC AORTA

Incidence and Etiology

- Thoracic aorta is vulnerable for injuries at its three fixed sites:
  - Annulus
  - Ligamentum arteriosum—isthmus
  - Aortic hiatus
- The most common sites being:
  - The ascending aorta proximal to innominate artery
  - The descending aorta at the point beyond the origin of left subclavian artery (ligamentum arteriosum)
- Mechanism of injury:
  - Penetrating: More common and may involve other great vessels
  - Blunt: Rapid deceleration produces shear at the fixed portions of the aorta, or by anteroposterior compression.
- Almost 80 percent die on the way to the hospital.

Symptom

Severe respiratory distress.

Signs

- Features of massive hemothorax
- Signs of shock may be evident
- Neurological signs of cord compression may be present.

> Injuries distal to innominate artery may show pseudocoarctation syndrome (upper extremity hypertension and hypotension and low pulse in lower limbs).

Relevant Investigations

- Chest X-ray and CT (Fig. 10.11) are diagnostic with the following findings:
  - Widening of superior mediastinum
  - Depression of left main stem bronchus
  - Loss of aortic knob
  - Massive hemopnuemothorax
  - Look for associated 1st rib fracture, flail chest, sternal fracture and fracture of thoracic spine.
- Arteriography is useful for definitive diagnosis of location and extent of injury.

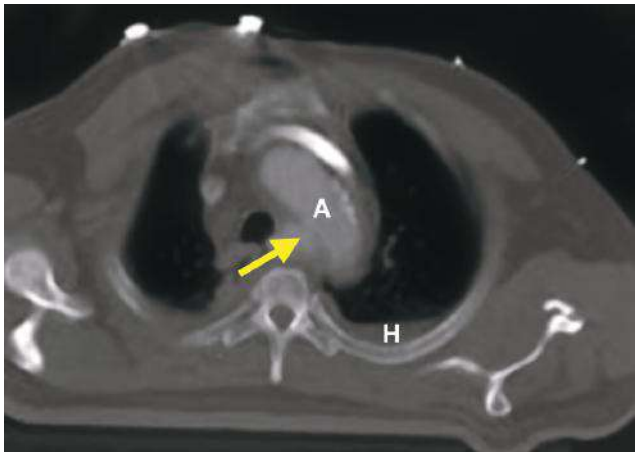

FIG. 10.11: CT chest—Aortic rupture with upper rib fractures

Treatment

- Resuscitation is the primary and emergent treatment
- Resection of damaged segment of aorta and repair/interposition grafts is curative.

INJURIES OF MYOCARDIUM

Incidence and Etiology

- Myocardial injuries are caused by:
  - ❖ Penetrating injuries (e.g. gunshot or stab injuries) where the outcome of injury depends on size of pericardial defect
  - ❖ Blunt injuries
    - Rupture into the pericardium producing pericardial tamponade
    - Myocardial contusion
    - Arrhythmias.
- Commotio cordis is the condition of sudden cardiac death or near sudden cardiac death after blunt, low-impact chest wall trauma in the absence of structural cardiac abnormality. Ventricular fibrillation is the most commonly reported induced arrhythmia in commotio cordis.
- Blunt impact injury to the chest with a baseball is the most common mechanism and does not result solely from the force of a blow as it is not seen in association with any rib or sternal fracture. It is largely the result of the exquisite timing of the blow during a narrow window within the repolarization phase of the cardiac cycle, 15 to 30 msec prior to the peak of the T wave. Survival rates for commotio cordis are low, even with prompt CPR and defibrillation.

Symptoms

Dyspnea and cyanosis.

Signs

- Examination reveals, distension of jugular veins, hypotension and narrowing pulse pressure and distant heart sounds (Beck's triad)
- The jugular venous distension raises paradoxically on deep inspiration (Kussmaul's sign), because the increased venous return cannot be accommodated within the constricted heart
- Signs of shock may be evident
- Pulsus paradoxus is a cardinal sign (drop in systolic BP >10 mm during inspiration due to CO_2 absorption).

Relevant Investigations

- Chest X-ray is contributory with cardiomegaly
- ECG may remain normal for over 36 hours but exhibit features of ischemia, dysrhythmia later
- ECHO is done to assess regional wall motion abnormalities
- CPK-MB values show elevation (check values at admission, 24–48 hours).

Treatment

- Cardiac monitoring and resuscitation are important
- Cardiac tamponade warrants pericardiocentesis/subxiphoid pericardial window
- Thoracotomy is done to create an opening of pericardial sac.

ESOPHAGEAL INJURIES

Incidence and Etiology

- Esophageal injuries are caused by:
  - *Penetrating injury* may occur at any level and are should be suspected when the injury crosses the midline (e.g. in sword swallowers as circus act), during esophagoscopy
  - *Blunt injury:* Usually following severe blow to the sternum or epigastrium. The common site of injury is at the lower 1/3rd esophagus.

Symptoms

Fever, dyspnea (due to mediastinitis or mediastinal emphysema) or tachypnea.

Signs

Features of surgical emphysema (spread of mediastinal emphysema to neck, face and chest wall) may supervene, with signs of hypoxia about 3–4 days later.

Relevant Investigations

- Chest X-ray may reveal
  - Pneumomediastinum
  - Air in the prevertebral space
  - Left pleural effusion
  - Hemo or pneumothorax in the absence of rib fracture.

FIG. 10.12: Gastrograffin swallow—Leaking dye in esophageal perforation (penetrating injury)

- Gastrograffin swallow may show the leak (Fig. 10.12)
- Esophagoscopy may show the injury
- Intercostal drainage shows particulate food matter, and show air leak during both phases of respiration.

Treatment

- Intercostal drainage is mandatory
- Early operative repair of the esophageal tear is necessary.

DIAPHRAGMATIC INJURIES

Incidence and Etiology

- Injuries of diaphragm are caused by:
  - Blunt injuries produce large radial tears and herniation of abdominal viscera into the chest
  - Penetrating injuries are small initially and enlarge over a period of time.

Symptoms and Signs

Diagnosis is difficult unless the tears are large and allows herniation of abdominal contents into the chest cavity—dyspnea and tachycardia.

Relevant Investigations

- Chest X-ray is contributory and may show:
  - Bowel loops in the chest (in herniation of small bowel)
  - Double shadow overlying the diaphragm
  - Nasogastric tube in the chest—in patients with gastrothorax.
- Contrast studies and CT of chest are conclusive.

Treatment

Surgical repair with or without mesh is mandatory.

11

Abdominal Injuries

Abdominal injuries can be of two varieties. They are:

1. Closed injuries (caused by blunt injuries)
2. Open injuries (caused by penetrating or stab injuries).

CLOSED INJURIES

- Caused by a blunt force exerted suddenly on the abdomen, such as:
  - Fall from a height
  - Blow with a fist
  - Injuries inflicted by heavy weapons like crowbars, poles, sticks
  - Run over injuries in road traffic accidents
  - Hitting against the steering wheel during sudden deceleration or braking of a speeding vehicle (in the absence of seat belt).
- They may lead to compression of intra-abdominal organs against the vertebral column causing rupture of:
  - Solid organs
  - Hollow organs
  - Mesentery
  - Detachment of gut from the mesentery
  - Contusion of abdominal wall.

OPEN INJURIES

- Caused by any sharp instrument like knife, flying objects like bullets, missiles, pieces of wood or glass.
- The incriminating agents enter the abdominal cavity taking with them some infection resulting in peritonitis.
- The points of entry and exit of the agent will indicate the direction and the possible organs injured.
  - Hollow organs may perforate and infect the peritoneal cavity (due to contamination by the contents of the organ—feces, urine, intestinal or gastric contents), or bleed (either into the organ or outside or both).
  - Solid organs cause hemorrhage (inside the organ or outside or both), and large bleeds cause shock and sometimes death.

Symptoms and Signs

- General features:
  - *Symptoms:* Weakness, clouding of thoughts and speech, air hunger, restlessness, decreased or absent urine output
  - *Signs:* Pallor, tachycardia, hypotension, subnormal temperature.
- Local features: Depend on the organ involved and the severity of damage.

INJURIES OF LIVER

Incidence and Etiology

The liver ranks high on the list of intra-abdominal organs involved by injury:

- Blunt injuries are more common than the penetrating injuries, due to increase in motor traffic moving at high speeds, and are associated with fracture of lower ribs on the right side. The dome of the liver is involved with anterior-posterior tears, more on the right lobe (7:1).
- Spontaneous rupture of liver is seen in:
  - Primary carcinoma in adults
  - Trauma during birth in children (postmature babies) being delivered per vaginum.

Liver injuries (Fig. 11.1) are classified into:

- Transcapsular (blood and bile will seep into the peritoneal cavity)
- Subcapsular (collection of blood between the capsule and the liver parenchyma mostly on the superior surface of liver)
- Central (interruption of liver parenchyma leading to intrahepatic hematoma, abscess and hematobilia).

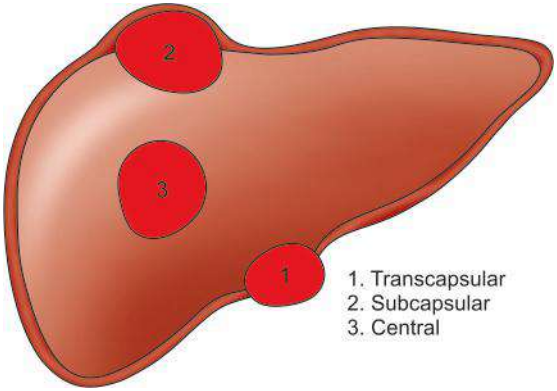

FIG. 11.1: Types of liver injuries

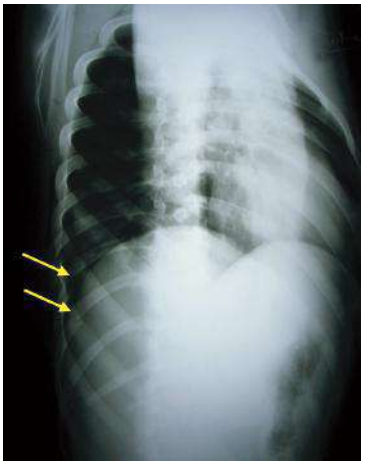

FIG. 11.2: X-ray showing lower rib fracture

Symptoms and Signs

- *Transcapsular:* Symptoms and signs of shock and peritoneal irritation (pain right upper abdomen with reference to the right shoulder, guarding and rigidity of the right hypochondrium, absent bowel sounds, shifting dullness)
- *Subcapsular:* Local tenderness and increase in area of liver dullness
- *Central:* Signs of shock may be present with hemobilia and hematemesis.

Relevant Investigations

- Plain chest X-ray (Fig. 11.2) will demonstrate fracture of lower ribs on the right side
- Plain X-ray abdomen may show haziness in the area of the liver with elevation of right dome of diaphragm
- CT (Figs 11.3A and B) and MRI are useful in localizing the damaged areas of liver and collections of blood or bile
- Peritoneal tap may be useful in identifying bile leaks
- Colloidal gold Au^{198} or Technitium-sulfur colloid Tc^{99m} scans are useful during active bleeding.

Treatment

- Correction of shock
- No surgical intervention is needed for small injuries

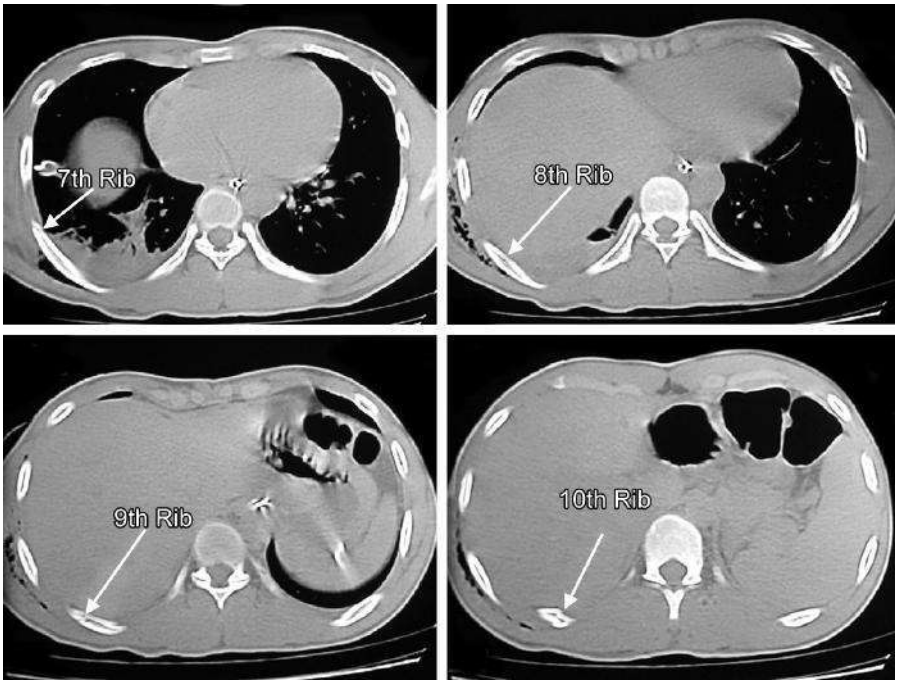

FIG. 11.3A: CT—Right lower rib fractures

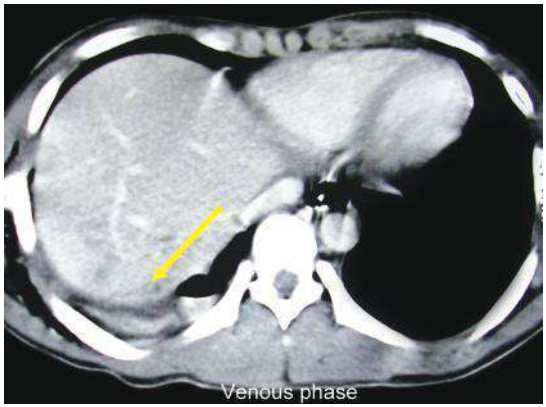

FIG. 11.3B: CT—Liver injury with hemoperitoneum

- Early surgical intervention, sometimes amounting to hepatectomy, is needed for large injuries associated with vascular or biliary tract injuries.

- Post-traumatic liver hemorrhage is amenable to radiological intervention (embolization).

INJURIES OF SPLEEN

Incidence and Etiology

- Spleen is the intra-abdominal organ most frequently injured by blunt trauma, usually by thoraco abdominal injuries associated with fractures of left lower ribs.
- The causes of splenic injuries are:
  - Blunt injuries of lower chest and abdomen (e.g. automobile accidents—may be associated with rib fractures, lung injuries, fracture of spine, intra-abdominal organs)
  - Transabdominal and transthoracic penetrating injuries (e.g. gunshot or knife injuries—may be associated with injuries to left kidney, colon, pancreas, vascular structures of mesentery)
  - Operative injuries (e.g. during operations on adjacent viscera—gastrectomy)
  - Spontaneous rupture or minor trauma (e.g. enlarged spleen of malaria).
- Splenic injuries may be:
  - Linear or stellate lacerations (due to excessive traction on the capsule during operations)
  - Capsular tears (due to excessive traction on the capsule during operations)
  - Subcapsular hematoma (due to excessive traction on the capsule during operations)
  - Puncture wounds (caused by penetrating injuries)
  - Intrasplenic hematomas (caused by penetrating injuries)
  - Rupture of parenchyma (due to extreme friability and vascularity of the organ).

Symptoms

Symptoms and signs of shock and peritoneal irritation.

Signs

- Pain left upper abdomen with reference to the left shoulder (Kher's sign)
- Guarding and rigidity of the left hypochondrium
- Absent bowel sounds
- Shifting dullness

- A mass of percussible area of fixed dullness in the left hypochondrium (Ballance's sign)
- Lacerations may be seen on the left lower chest.

- The clinical presentation of splenic injury is of three types:
  1. *Acute:* Immediate presentation of symptoms
  2. *Delayed:* A quiescent period of 7 days to less than 2 weeks, between the injury and intraperitoneal bleeding (Latent period of Baudet), probably related to a temporary tamponade of a minor laceration or the presence of a slowly enlarging subcapsular hematoma which eventually ruptures
  3. *Occult:* Mild symptoms caused by rupture of organized intrasplenic or parasplenic hematoma.

Relevant Investigations

- Plain chest X-ray chest will demonstrate fracture of ribs on the left side
- Plain X-ray abdomen may show:
  - Haziness in the area of the spleen
  - Elevated immobile left dome of diaphragm
  - Medial displacement of gastric shadow with indentation
  - Obliteration of psoas shadow
  - Widening of space between splenic flexure and the preperitoneal pad of fat.
- CT (Figs 11.4A to C) and MRI are useful in localizing the damaged areas of spleen and collections of blood
- Peritoneal tap may be useful in identifying blood in the peritoneal cavity.

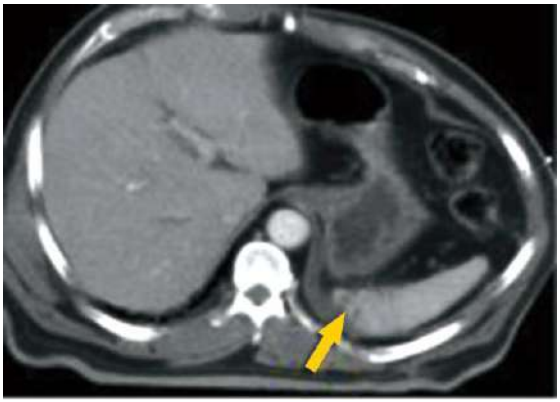

FIG. 11.4A: CT—Splenic injury with lower rib fracture

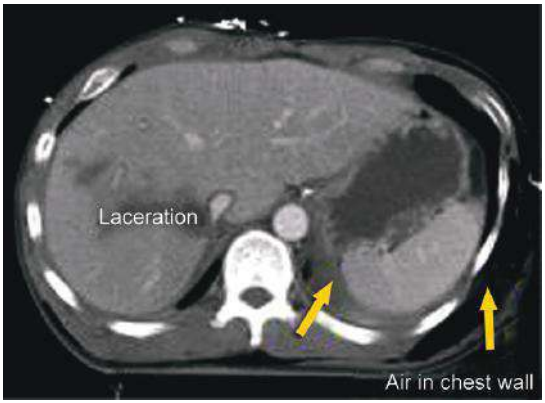

FIG. 11.4B: CT—Laceration of liver and perisplenic collection

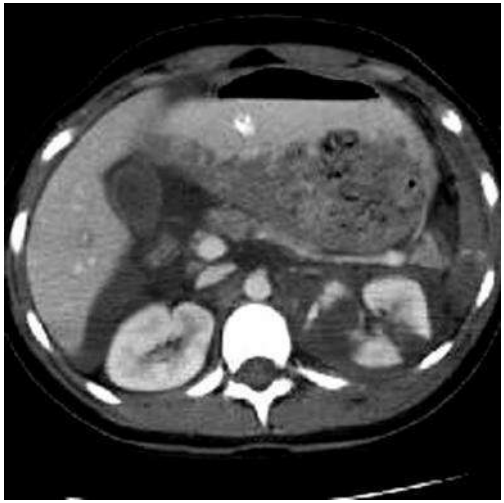

FIG. 11.4C: CT—Combined renal and splenic injuries of automobile accident

Treatment

- Once the diagnosis is made the operation should not be delayed
- Splenectomy is the recommended treatment, regardless of the type and the extent of the injury.

- Following splenectomy, to prevent the life-threatening infections (pneumococci), long-term prophylaxis with penicillin (minimum of 2 years), with immunization against *Pneumococcus*, *Meningococcus* and *Hemophilus influenzae* is required.

INJURIES OF MESENTERY

Incidence and Etiology

- Mesentery consists of its arteries and veins and they may be injured by either penetrating or non-penetrating abdominal trauma
- In most cases, associated organ injuries are found
- Isolated injury to mesenteric vessels is rare.

Symptoms

Depending on the size of the vessel lacerated, the rapidity of bleeding and associated organ injury, the patient will present with signs of shock, abdominal pain and distension.

Sign

On examination, tenderness is marked.

> When the bleeding occurs within the layers of mesentery, the clinical signs evolve slowly and the viability of the bowel is threatened.

Relevant Investigations

- Plain X-rays may show air under the diaphragm when the bowel is perforated by the injury
- Peritoneal tap may be useful in establishing the bleeding.

Treatment

- Repair of the torn mesentery is required
- Resection and anastomosis is required for nonviable intestine.

INJURIES OF DUODENUM

Incidence and Etiology

- Duodenum can be injured both by penetrating and nonpenetrating abdominal trauma
- The duodenum can rupture:
  - Intraperitoneally and cause immediate chemical irritation of the peritoneum due to the highly alkaline duodenal content
  - Retroperitoneally (more common with blunt trauma such as steering wheel injuries).

Symptoms and Signs

- Intraperitoneal rupture:
  - Abdominal pain, fever and distension with vomiting
  - Examination shows marked tenderness in the upper abdomen and later signs of generalized peritonitis.
- Retroperitoneal rupture:
  - Pain in the epigastrium and back, with pronounced vomiting
  - Testicular pain is a common feature of retroperitoneal rupture of duodenum.

Relevant Investigations

- Plain X-ray of abdomen may show air under the domes of diaphragm (intraperitoneal rupture) or large accumulation of air above the right kidney (retroperitoneal rupture). Diagnostic accuracy can be increased

by injecting air in the Levine's tube to increase the air collections. Water soluble dye injections in the tube can make the diagnosis more precise
- CT with contrast is conclusive
- Paracentesis may show bile-stained fluid, if the rupture is intraperitoneal.

Treatment

- Simple suturing may be adequate in many cases
- Supplementary gastroenterostomy is required for large tears
- Rarely, even a pancreatoduodenectomy may be necessary for extensive trauma involving the periampullary region.

INJURIES OF SMALL INTESTINE

Incidence and Etiology

- Injuries to the small intestine are more common than injuries to the duodenum and large intestine, the usual mechanism being the blunt trauma crushing the bowel against the vertebral column, more commonly the duodenojejunal flexure and the ileocecal junction, the fixed parts
- Blunt injuries cause slow necrosis of bowel and leak occurs late
- Signs and symptoms develop late—2 to 3 days later, depending on the size of the damage and leak of contents
- Penetrating injuries can also cause small bowel trauma, but less commonly, probably due to its sliding away from a knife because of its great mobility
- Associated mesenteric tears are common.

Clinical Presentation

- Abdominal pain, distension and vomiting
- Tenderness and guarding are pronounced around the damaged bowel and the patient may point it (Pointing sign).

Relevant Investigations

- Plain X-rays may show air under the domes of the diaphragm
- Paracentesis will show bile-stained fluid.

Treatment

- Simple suturing is done for simple tears
- Resections are required for large tears with nonviable bowel
- Peritoneal toileting is mandatory under cover of antibiotics.

INJURIES OF LARGE INTESTINE

Incidence and Etiology

- Large bowel injuries can be caused by penetrating and nonpenetrating injuries
- Ruptures may be:
  - Extraperitoneal for ascending and descending colon injuries, whereas
  - Intraperitoneal, when the injuries are of transverse and sigmoid colons
- Patient develops septic complications quickly as the large fluid leak is fecal and infected
- Delayed presentation is not uncommon as in small bowel injuries

Symptoms

- Abdominal pain, vomiting and distension
- High grade fever occurs due to fecal contamination.

Signs

Signs of peritonitis.

Relevant Investigations

- Plain X-rays may show air under the domes of the diaphragm (intraperitoneal ruptures)
- Paracentesis will show feculent fluid.

Treatment

- Early laparotomy is required
- Closure of tears with proximal diversion is necessary
- Peritoneal toileting is mandatory under cover of broad-spectrum antibiotics.

ABDOMINAL COMPARTMENT SYNDROME

Incidence and Etiology

Occurs when the intra-abdominal pressure rises above the venous pressure.

This can be caused by:

- Blood
- Free gas
- Free fluid
- Tissue or splanchnic edema
- Organomegaly.

Pathogenesis

Occurs due to underperfusion of intra-abdominal organs (e.g. gut, kidneys and liver).

Symptom

Falling urinary output.

Signs

Tense and quiet abdomen.

Relevant Investigations

Measurement of intra-abdominal pressure by connecting the urinary catheter to a pressure transducer (>20 mm Hg suggests abdominal compartment syndrome).

Treatment

Surgical exploration

- Patients may bleed torrentially when the abdomen is opened followed by hemodynamic instability, following visceral reperfusion
- Assisted ventilation may be required for many days.

12

Urological Injuries

RENAL INJURIES

Incidence and Etiology

- Kidneys are the most common organs to get injured in the urological system
- Injuries to the kidneys may be caused by:
  - Closed blunt abdominal trauma such as road traffic accidents, during active sports and may be associated with lower rib fractures
  - Penetrating injuries by bullets, knives are also causes of renal trauma
- Renal trauma is of three types:
  - Minor trauma (e.g. renal contusion, subcapsular hematoma)
  - Major trauma (e.g. deep cortical lacerations, extravasation of urine)
  - Renal vascular trauma (e.g. avulsion of renal pedicle).

Symptom

Hematuria is the most common symptom.

Signs

- Bruising over the ribs posteriorly or evidence of penetrating injury
- Tenderness and guarding in the loin and expanding mass
- Signs of shock may be seen in major trauma.

Relevant Investigations

- Plain X-rays may show fractures of lower ribs
- CT is the investigation of choice and accurate assessment is possible (Figs 12.1 and 12.2)
- Renal arteriography may be needed in cases where CT is not contributory, especially with renal vessel injuries.

FIG. 12.1: CT—Injury of right kidney

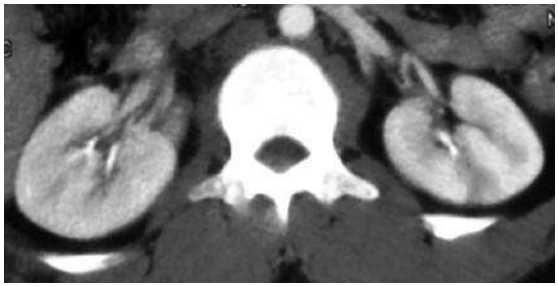

FIG. 12.2: Contrast enhanced CT—Left renal injury

Treatment

- Any patient with renal injury should rest in bed and all urine samples should be examined for blood
- Surgical exploration is warranted in closed injuries, when there is retroperitoneal bleeding, expanding loin mass, urinary extravasation, evidence of nonviable renal parenchyma and renal pedicle injuries
- Exploration of abdomen is required for penetrating injuries for assessment of other injuries and kidney should only be explored when the condition warrants.

URETERIC INJURIES

Incidence and Etiology

- Most common cause of ureteric injuries is during abdominal and pelvic operations
- Urinary leak occurs around the 5th postoperative day
- The operations associated with ureteric injuries are given in Table 12.1.

Table 12.1: Causes of ureteric injuries		
Operations		
Gynecology	*General surgery*	*Urology*
Hysterectomy (abdominal or vaginal) – open or laparoscopic	Sigmoid colectomy	Excision of bladder diverticula
Ovarian cystectomy	Abdominoperineal resection	Ureterolithotomy
Anterior colporrhaphy	Surgery for aortic aneurysm	Ureteroscopy

Nature of Injuries

- Complete ligation of one or both ureters, leads to increase in intraureteric pressure, kidneys stop secreting urine, with resultant dilatation of pelvis and atrophy of kidney
- When the obstruction is incomplete (inclusion in a stitch), secretion continues to be normal, hydronephrosis and infection occur
- When the ureter is divided or suffers crushing injury, urine leaks into the retroperitoneum or peritoneal cavity which may result in a urinary fistula
- The injury may be recognized at the time of surgery
- If bilateral ligation is not recognized during surgery, it manifests as uremia.

Symptom and Sign

The clinical presentation varies according to the nature of injury (Table 12.2).

Table 12.2: Ureteric injuries and their clinical presentations		
	Nature of injury	*Clinical presentation*
Symptoms	Bilateral ligation	Anuria in the immediate postoperative period
	Unilateral ligation	No symptoms/loin pain/fever
Signs	Division	Urine leak through the drain/wound/vagina
	Retroperitoneal urine leak	Abdominal distension and paralytic ileus
	Intraperitoneal urine leak	Abdominal distension and paralytic ileus with peritonitis and sepsis

FIG. 12.3: IVU—Leakage of contrast from damaged ureter

Relevant Investigation

IVU (Fig. 12.3) or contrast enhanced CT will demonstrate the nature of obstruction and urinary leak.

Treatment

- If injury is recognized during surgery, it should be rectified immediately:
  - Spatulation and end-to-end anastomosis is done, when the length is not lost
  - Incomplete damages to the ureter may be treated by stenting and spontaneous healing may be expected.
  - Surgical corrections (ureteroureterostomy, ureteroneocystostomy) are necessary to restore the anatomy, before renal function is compromised.

URINARY BLADDER INJURIES

Incidence and Etiology

- Urinary bladder is partially extraperitoneal and is also closely covered by the parietal peritoneum, and because of this anatomy, it can rupture:
  - Extraperitoneally (80%)
  - Intraperitoneally (20%)

15

Hand Injuries

Incidence and Etiology

- Hand injuries form nearly one-third of all injuries to human body
- They can be:
  - Open
  - Closed
- Acute injuries can be classified depending on the nature of force and on contamination as tidy, untidy and indeterminable injuries:
  - *Tidy injuries:* Clean cuts and incised wounds, usually resulting from sharp force
  - *Untidy injuries:* Ragged skin wounds, may be with associated fractures, usually resulting from crush or avulsion force
  - *Indeterminable injuries:* Injuries where it is difficult to determine the extent of injury and the viability of tissues, usually caused by very severe force
- Injuries to the vessels, nerves, bones and joints may be present in any of the above injuries.

Complications: Post-traumatic deformities like contractures, neurovascular deficit.

Clinical Presentation

The patient presents to the surgeon either in the acute stage or late.

- In acute injuries (Figs 15.1A to C), the pain is excruciating, and assessment is difficult, unless pain is alleviated by analgesics. When patients arrive late, infection gets superadded, and result in post-traumatic sequelae.

FIG. 15.1A: Traumatic amputation of tip of middle finger

FIG. 15.1B: Traumatic amputation of entire index finger

FIG. 15.1C: Penetrating injury of three fingers

- *Nail injury:* Avulsion of nail (complete or partial) (Fig. 15.2A)
- *Skin damage:* Loss of skin, especially degloving
- *Nerve damage:* Sensory or motor loss depending on the nerve affected
- *Damage to arteries:* Suspected by profuse hemorrhage (incomplete tears), gangrene (complete injuries) (Figs 15.2B to D)
- *Damage to tendon sheaths and tendons:* Lack in active movement
- *Damage to bones and joints:* Pain, deformity, tenderness and loss of movement.

FIG. 15.2A: Avulsion of nails by crush injury

FIG. 15.2B: Traumatic gangrene of middle finger

FIG. 15.2C: Gangrene of fingertip following crush injury

FIG. 15.2D: Traumatic gangrene of little finger

In acute injuries, for assessment of damage, surgical exploration may be necessary when in doubt and it should be done meticulously. The structures should be considered divided until otherwise proved.

- In late post-traumatic cases detailed examination of movements and sensation can be done with ease, and damage assessed completely.
  - *Skin injuries:* Necrosis of skin, may manifest as raw area, when very late, can present with contracture due to secondary healing.
  - *Vascular injuries:* Necrosed skin and deeper tissues

FIG. 15.3A: Boutonnière deformity of little finger

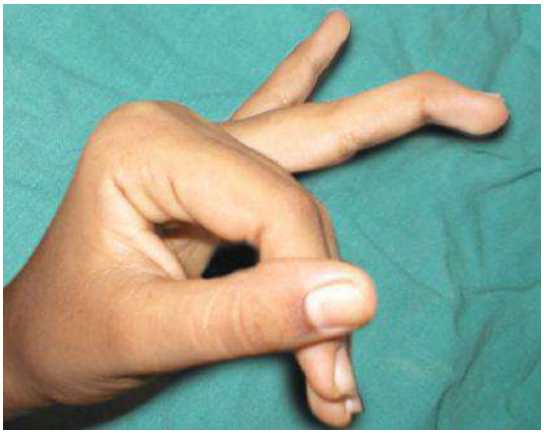

FIG. 15.3B: Mallet finger

FIG. 15.3C: Swan-neck deformity

- *Extensor tendon injuries:* Dropped finger (if injury is at dorsum or proximal level), Boutonnière deformity (when injury is at proximal finger level) (Fig. 15.3A) and Mallet finger (when injury is at distal finger level) (Fig. 15.3B). Neglected or untreated Mallet fingers may progress to Swan-neck deformity (Fig. 15.3C).

- *Flexor tendon injuries:* Lack of movement of PIP joint suggest flexor digitorum sperficialis injury and DIP joint suggests flexor digitorum profundus injury
- *Nerve damage:* Wasting of muscle groups, lack of sweating and trophic changes
- *Bone or joint injuries:* Fractures with or without dislocations of joints.

Relevant Investigations

- X-rays (in various views) are necessary to assess bone injuries
- Doppler studies are required to assess vascular damage.

Treatment

- Bleeding can be controlled by elevating the limb or by applying pressure directly over the site with pads
- In acute injuries:
  - *Skin injuries:* Primary repair wherever possible
  - *Vascular injuries:* Primary microvascular repair
  - *Bone and joint injuries:* Proper debridement and splinting, reduction of dislocations and immobilization
  - *Nerve injuries:* Primary repair under magnification where possible
  - *Tendon injuries:* Primary repair in clean wounds.
- In late post-traumatic cases:
  - *Skin injuries:* Wounds healing by secondary intention cause severe fibrosis and contractures, and may need release and repair and skin grafting in some cases
  - *Vascular injuries:* Grafts wherever required
  - *Bone and joint injuries:* Malunion or nonunion should be treated accordingly
  - *Nerve injuries:* Nerve grafts
  - *Tendon injuries:* Tendon transfers or grafts.

Note: Dirty wounds should be debrided well and converted into a clean wound and further treatment is carried out.

Section V

Organs and Systems

16. Oral Cavity
    - Acute Odontogenic Infections
17. Head and Neck
    - Acute Suppurative Sialadenitis
    - Ludwig's Angina
18. Thorax
    - Acute Mediastinitis
    - Acute Pleuritis
    - Pleural Effusion
    - Acute Empyema Thoracis
    - Spontaneous Pneumothorax
    - Foreign Bodies in the Respiratory Tract
    - Suppurative or Aspiration Pneumonia
    - Hospital Acquired Pneumonia
    - Acute Lung Abscess
    - Pulmonary Embolism
19. Breast
    - Breast Hematoma
    - Acute Breast Abscess
20. Spine
    - Degenerative Diseases of Disk and Facet Joints
    - Spondylolisthesis
21. Gastroenterology
    - Acute Abdomen
      - Acute upper abdominal pain
      - Acute lower abdominal pain
    - Acute Liver Abscess
    - Acute Cholecystitis
    - Acute Cholangitis
    - Primary Sclerosing Cholangitis
    - Splenic Abscess
    - Acute Hemorrhagic Pancreatitis
    - Acute Appendicitis
    - Acute Mesenteric Lymphadenitis
    - Acute Colonic Diverticulitis
    - Acute Meckel's Diverticulitis
    - Acute Solitary Cecal Diverticulitis
    - Acute Ulcerative Colitis
    - Acute Intestinal Obstruction
    - Gallstone Ileus
    - Acute Intussusception
    - Swallowed Foreign Bodies
    - Perforated Bowel Pathologies

- Intestinal Strictures
- Bands and Adhesions
- Enteroliths/Food Bolus
- Volvulus
  - Sigmoid volvulus
  - Cecal volvulus
  - Midgut volvulus
  - Gastric volvulus
- Intestinal Obstruction due to Herniae (Internal and External)
- Paralytic Ileus
- Torsion of Mesenteric Cyst
- Torsion of Omentum
- Colics
- Gastrointestinal Hemorrhage

22. Anorectum
    - Acute Anal Fissure
    - Anorectal Abscess
    - Hemorrhoids
    - Perianal Hematoma
    - Prolapse of Rectum

23. Vascular System
    - Acute Limb Ischemia
    - Acute Intestinal Ischemia
    - Leaking or Dissecting Aortic Aneurysm

24. Urology
    - Acute Retention of Urine
    - Hematuria
    - Renal Colic
    - Ureteric Colic
    - Acute Urethritis
    - Acute Prostatitis
    - Acute Prostatic Abscess

25. Male Genitalia
    - Acute Scrotal Pain
    - Torsion of Testis
    - Torsion of Appendages of Testis
    - Acute Epididymo-orchitis
    - Traumatic Orchitis
    - Hematocele
    - Pyocele
    - Idiopathic Scrotal Edema
    - Acute Scrotal Abscess
    - Fournier's Gangrene
    - Acute Filarial Scrotum
    - Fracture Penis
    - Paraphimosis
    - Priapism

26. Hernias
    - Complicated Hernias

27. Gynecology
    - Acute Torsion of Ovarian Cyst
    - Acute Salpingitis
    - Rupture of Lutein Cyst

28. Pediatrics
    - Acute Intussusception
    - Congenital Pyloric Stenosis
    - Necrotizing Enterocolitis
    - Tracheoesophageal Fistula

29. Lymphatic System
    - Acute Lymphangitis
    - Acute Viral and Bacterial Lymphadenitis
    - Acute Filarial Lymphangitis and Lymphadenitis

30. Skin and Subcutaneous Tissues
    - Hematoma
    - Erysipelas
    - Furuncle
    - Cellulitis
    - Acute Pyogenic Abscess
    - Carbuncle
    - Burns
    - Necrotizing Fasciitis

16

Oral Cavity

ACUTE ODONTOGENIC INFECTIONS

Incidence and Etiology

Acute odontogenic infections present in three ways:

i. Abscess
ii. Cellulitis
iii. Fulminating infections.

The odontogenic abscesses present in many ways:

- Acute periapical abscess (cause – Infective necrosis of pulp)
- Acute dentoalveolar abscess
- Acute periodontal abscess
- Acute pericoronal abscess.

Causative organisms

- Staphylococci
- Streptococci
- Bacteroides.

Pathogenesis

The infection follows the path of least resistance, the deciding factor being the fascial planes:

- Mandibular 2nd and 3rd molar infections perforate the lingual cortex and spread to submandibular space, as the roots of these teeth lie below the mylohyoid line

- Mandibular premolars and first molars, involve the sublingual space, as the roots of these teeth lie above the mylohyoid line.

Spread of Infection

- *Acute periapical abscess (cause–Infective necrosis of pulp):* Organisms from infected pulp invade periapical tissue through apical foramina
- *Acute dentoalveolar abscess:* Continuation of periapical abscess
- *Acute periodontal abscess:* Arises in the periodontal membrane adjacent to a periodontal pocket
- *Acute pericoronal abscess:* Arises around the crown of impacted or erupting teeth.

Symptoms and Signs

- *Acute periapical abscess:* Severe throbbing pain in the affected tooth which may be carious
- *Acute dentoalveolar abscess:* Severe pain, submucosal and or extraoral swelling
- *Acute periodontal abscess:* Dull pain, rarely severe, pus discharge via gingival pocket
- *Acute pericoronal abscess:* Dull continuous pain, swelling of gingiva around the crown with or without pus discharge.

Relevant Investigations

Diagnosis is more clinical. X-rays are not diagnostic.

Treatment

- *Medical:* Hydration, soft diet, analgesics, mouthwashes, broad-spectrum antibiotics
- *Surgery:* Extraction of affected tooth or incision and drainage of abscess when possible
- *Conservative:* Root canal treatment for acute periapical abscess.

17

Head and Neck

ACUTE SUPPURATIVE SIALADENITIS

Incidence and Etiology

- Occurs generally in:
  - People with poor oral hygiene
  - Postoperative patients
  - Dehydration and reduced or obstructed salivary flow
  - By ascending infection from oral flora, usually *Streptococcus viridans* or pneumococci
- Parotid gland is usually involved
- Submandibular gland may also be involved.

Clinical Presentation

- Acute parotitis:
  - Painful unilateral parotid swelling with trismus, fever and tachycardia
  - On examination, the parotid is diffusely enlarged as a brawny swelling (Fig. 17.1) and extremely tender with purulent discharge through the Stenson's duct.
- Acute submandibular sialadenitis:
  - Painful swelling of the submandibular gland
  - On examination, the gland is enlarged and tender (Fig. 17.2) purulent discharge through the Wharton's duct
  - Obstructing calculus may be seen or felt in the floor of the mouth (Fig. 17.3).

FIG. 17.1: Acute suppurative parotitis

FIG. 17.2: Submandibular salivary gland abscess

FIG. 17.3: Calculus in the Wharton's duct (*Courtesy:* Dr Jacinth Cornelius)

Relevant Investigation

No special investigation is required.

Treatment

- Broad-spectrum antibiotics are necessary
- If abscess is formed, external drainage is necessary without waiting for fluctuation to establish.

LUDWIG'S ANGINA

Incidence and Etiology

- Usually due to dental infections
- The infection is a mixture of aerobic and anaerobic organisms.

Pathogenesis

It is cellulitis involving the sublingual and submandibular spaces beneath the deep cervical fascia.

Complication: Edema of glottis.

Symptoms

- Severe pain and swelling of neck (Fig. 17.4)
- May cause respiratory embarrassment.

Signs

- Severely swollen neck
- Marked tenderness.

Relevant Investigation

Culture of pus and identifying the organism.

Oral cavity examination is mandatory.

FIG. 17.4: Ludwig's angina

Treatment

- Broad-spectrum antibiotics
- Incision and drainage, if abscesses are found in the dental planes, with or without dental extractions.

18

Thorax

ACUTE MEDIASTINITIS

Incidence and Etiology

Acute suppurative mediastinitis occurs due to varied causes. They are:

- Injury to cervical or thoracic esophagus
- Extension of cervical infections
- Secondary to osteomyelitis of ribs and sternum
- Extension of retroperitoneal infections
- Suppuration of mediastinal lymph nodes.

Clinical Presentation

- Chest pain, fever and chills are commonly present
- Symptoms related to primary cause are usually present.

Relevant Investigations

- X-ray chest (Fig. 18.1) shows mediastinal widening
- CT chest is diagnostic, especially of the primary cause.

FIG. 18.1: X-ray—Superior mediastinitis

Treatment

- Broad-spectrum antibiotics are necessary
- Immediate surgery with adequate mediastinal drainage for esophageal injuries.

ACUTE PLEURITIS

Incidence and Etiology

- A simple term denoting the pleural involvement of any disease giving rise to pleuritic pain or evidence of pleural friction
- Common feature of pleural invasion by pulmonary tuberculosis or bronchogenic carcinoma.

Symptom

Pain characteristically inspirational.

Sign

Chest movement may be restricted locally with an audible pleural rub.

Relevant Investigations

X-ray chest is useful in identifying the underlying cause.

Treatment

Analgesics are required to treat pleuritic pain with treatment focused towards the primary cause.

PLEURAL EFFUSION

Incidence and Etiology

Denotes accumulation of serous fluid in the pleural space, which results due to:

- Increased hydrostatic pressure or decreased osmotic pressure (transudative effusion). The causes are:
  - Cardiac failure
  - Hepatic failure
  - Renal failure
- Increased microvascular permeability due to disease of pleura or adjacent lung (exudative effusion). The causes are:
  - Infections (e.g. bacterial, tuberculous, fungal, parasitic, viral)
  - Collagen vascular disease (e.g. rheumatoid, lupus)
  - Malignancy (e.g. mesothelioma, lung cancer, metastases)
  - Pulmonary embolism
  - Abdominal disease (e.g. pancreatitis, subphrenic abscess)
- The effusion can be:
  - Unilateral (e.g. tuberculosis, malignancy)
  - Bilateral (e.g. cardiac failure, hypoproteinemia).

Symptom

Breathlessness is the most common symptom.

Signs

- Reduced chest wall movement on the affected side, dullness on percussion and reduced or absent breath sounds and vocal resonance
- Large effusions shift the trachea to the opposite side.

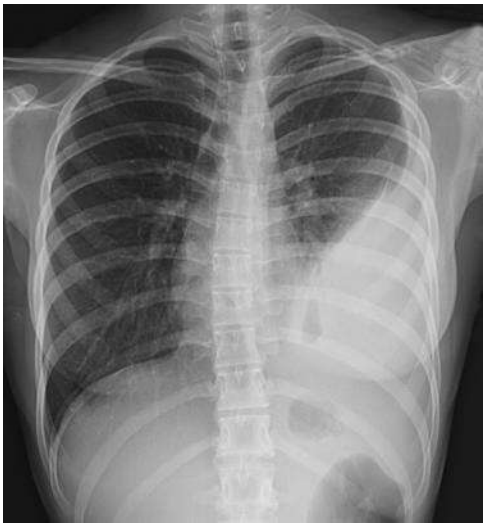

FIG. 18.2A: X-ray chest—Left pleural effusion

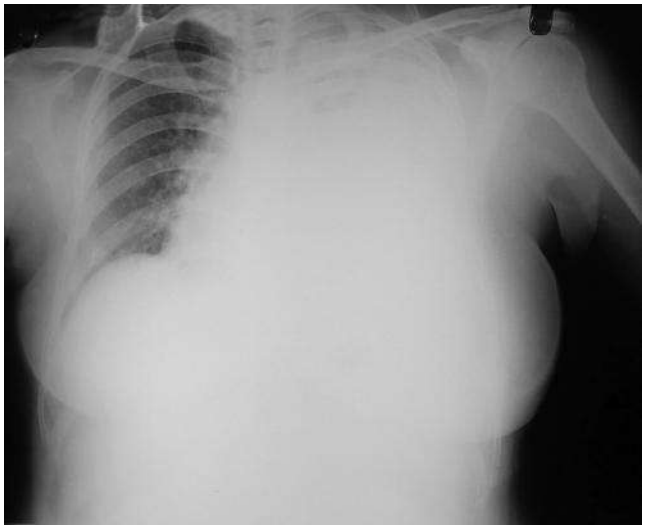

FIG. 18.2B: X-ray chest—Massive pleural effusion left side

Relevant Investigations

- Chest X-ray shows a dense uniform opacity in the lower and lateral parts of hemithorax, shading off above and medially into translucent lung (more than 400 ml is required to blunt the costodiaphragmatic angle) (Figs 18.2A and B)
- US is very useful in differentiating the effusion and tumor
- CT (Fig. 18.3) is diagnostic
- Aspiration of pleural fluid is necessary for:
  - Nature (blood – malignancy, embolus; milky – chyle; pus – empyema; straw color – tuberculosis and transudates)
  - Microbiology – culture sensitivity
  - Cytology
  - Biochemistry (Total proteins >3 g% - exudate, pleural LDH/serum LDH ratio >0.6 – exudates, high levels of triglycerides >110 mg% - chylothorax, amylase – pancreatitis, malignant effusions)
- Bronchoscopy biopsy, thoracoscopy and biopsy of enlarged lymph nodes may be needed to identify the primary cause
- Pleural biopsy yields good results.

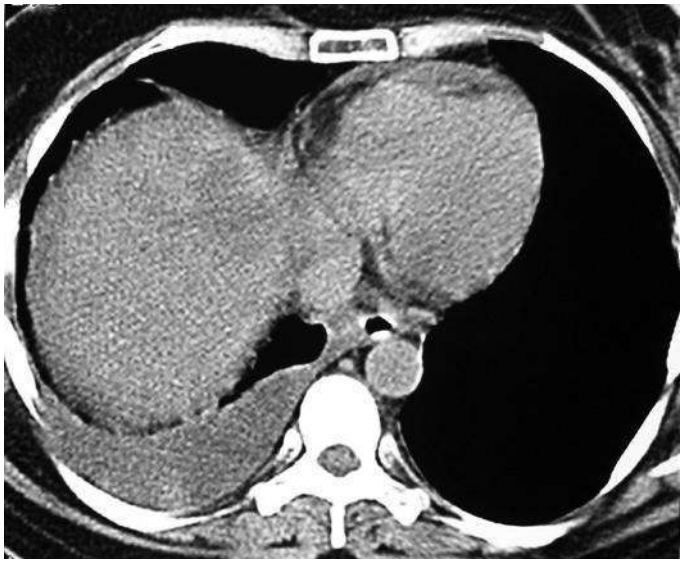

FIG. 18.3: CT—Right sided pleural effusion

Treatment

Aspiration of pleural fluid may be necessary to relieve breathlessness, but treatment of underlying cause is necessary.

ACUTE EMPYEMA THORACIS

Incidence and Etiology

- Empyema indicates pus in the pleural space, which may involve a part (loculated) or the whole of it
- Always secondary to infection in a neighboring structure, usually the lung (bacterial pneumonia and tuberculosis)
- Other causes being, infection of hemothorax or rupture of subphrenic abscess or liver abscess.

Complications: Bronchopleural fistula, empyema necessitans, (empyema burrowing through the chest wall to present as an abscess externally), pleurocutaneous fistula, pericarditis, mediastinal abscess.

Symptom

High intermittent pyrexia, with rigors, sweating, malaise and weight loss.

Signs

- Pleural pain, breathlessness and cough with purulent sputum
- Clinical signs are those of pleural effusion.

Relevant Investigations

- Chest X-ray is diagnostic
- US and CT give invaluable information
- Aspiration of pus is confirmatory
- Isolation of organism from sputum or pleural aspirate is useful in treatment.

Treatment

- General: Supportive respiratory care, physiotherapy
- Drainage of pleural space (thoracentesis or close drainage) and irrigation with normal saline is necessary to clear the pleural space of the pus, under cover of appropriate antibiotics
- Underlying cause like pneumonia and tuberculosis need active treatment.

SPONTANEOUS PNEUMOTHORAX

Incidence and Etiology

- Pneumothorax is presence of air in the pleural cavity
- This is a sudden event
- Spontaneous pneumothorax can be:
  - Primary—without any obvious evidence of pulmonary pathology (e.g. rupture of small emphysematous bulla)
  - Secondary—due to underlying pulmonary pathology (e.g. COPD, tuberculosis).

Symptom

Sudden unilateral chest pain or breathlessness.

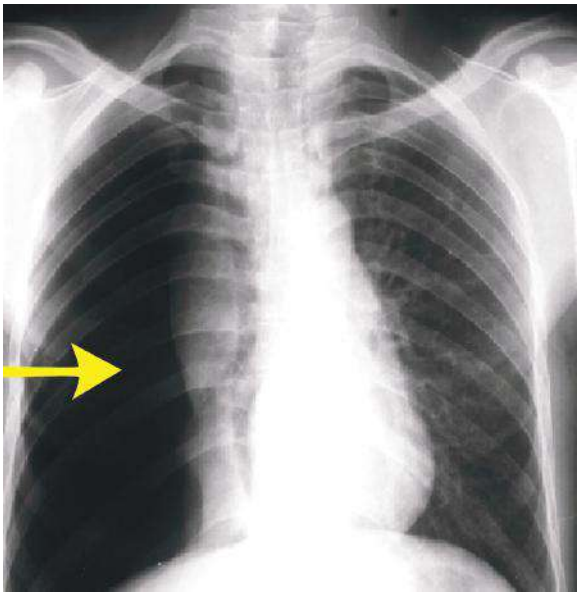

FIG. 18.4: Chest X-ray—Right pneumothorax

Sign

Small pneumothorax shows no clinical signs, whereas, large ones show decreased movement of chest wall, hyper-resonant percussion note and decreased or absent breath sounds.

Relevant Investigations

- Chest X-ray (Fig. 18.4) shows sharply defined edge of the deflated lung with complete translucency between the line and the chest wall with no lung markings
- CT is useful in defining the underlying pathology.

Treatment

- Percutaneous needle aspiration of air is necessary for full lung expansion
- Intercostal tube drainage with underwater seal may be needed [See Chapter 31 (Insertion of Chest Drains)]
- Smoking should be completely avoided and the underlying pathology treated.

FOREIGN BODIES IN THE RESPIRATORY TRACT

Incidence and Etiology

- Foreign bodies get lodged in the trachea and bronchi accidentally, and are more common in:
  - Children than in adults
  - The bronchus than the trachea
  - The right bronchus than the left, as the right bronchus is a direct continuation of trachea and is wider than the left bronchus.
- In children the objects are in the form of seeds, buttons, cells, etc. and in adults, they are dentures, loose teeth, etc.
- In adults, accidental lodging when they are under the influence of alcohol.

Pathogenesis

- Foreign bodies act like a valve and cause symptoms (Table 18.1).

Symptoms

Dyspnea, cough, stridor, cyanosis and fever.

Signs

- Rhonchi
- Reduced breath sounds.

Relevant Investigations

- X-rays reveal the radio-opaque foreign bodies, and collapsed lungs
- Bronchoscopy is diagnostic (Figs 18.5 and 18.6).

Table 18.1: Pathogenesis of foreign body obstruction in lower respiratory tract

Mechanism	*Pathophysiology*
Stop valve	Foreign body causes total obstruction and does not allow air entry (ingress) or allow air and secretions to escape (egress). The lobe of lung may collapse and consolidate
Bypass valve	When the foreign body is small or has a hole, it allows ingress and egress of air
Ball valve	During inspiration, the bronchi dilate allowing air to enter (ingress), but does not allow air to escape (egress) as the bronchi constrict during expiration. The trapped air caused pneumothorax or emphysema

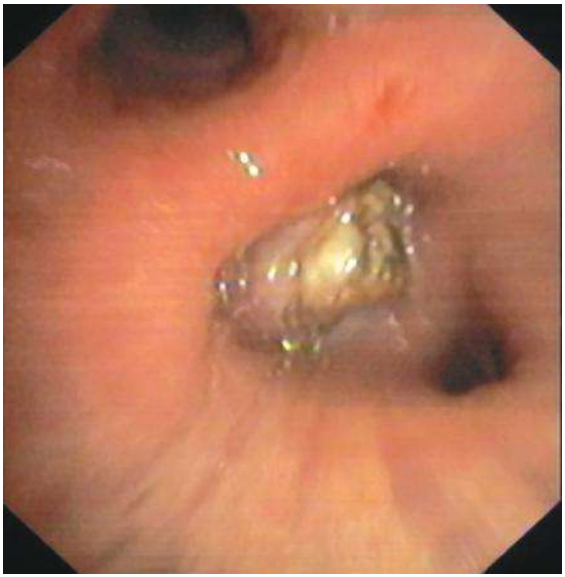

FIG. 18.5: Bronchoscopy—Tooth in the bronchus (*Courtesy*: Dr R Narasimhan and Dr Gayathri)

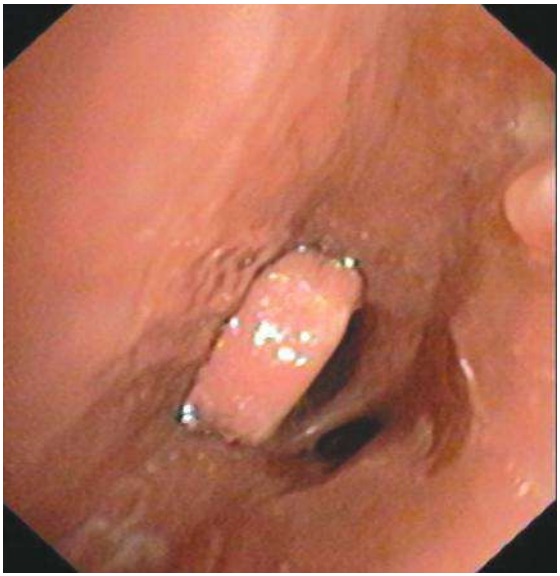

FIG. 18.6: Bronchoscopy—Tablet in the bronchus (*Courtesy*: Dr R Narasimhan and Dr Gayathri)

Treatment

Bronchoscopic removal (using a rigid bronchoscope) is curative.

SUPPURATIVE OR ASPIRATION PNEUMONIA

Incidence and Etiology

- Suppurative pneumonia or pneumonic consolidation denotes destruction of lung parenchyma by inflammatory process with microabscess formation on histology
- *Staphylococcus aureus* and *Klebsiella pneumoniae* infections cause suppuration
- Suppuration has an insidious onset whereas aspiration is acute.

Pathogenesis

- Inhalation of septic material during endotracheal anesthesia or by aspiration of gastric contents
- Aspiration may lead to severe acute respiratory distress syndome (ARDS).

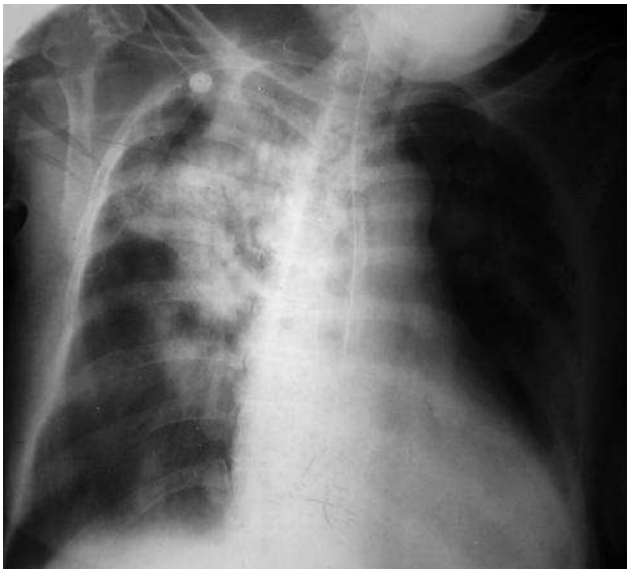

FIG. 18.7: X-ray—Aspiration pneumonia right lung

Symptoms

- Productive cough (fetid or blood stained)
- Pleural pain may be present
- High remittent pyrexia is common.

Sign

Sign of consolidation, with pleural rub.

Relevant Investigation

Chest X-ray (Fig. 18.7) shows homogeneous lobar or segmental opacity. A cavity with fluid level may indicate an abscess.

Treatment

- Broad-spectrum antibiotics form the mainstay of treatment
- Surgical intervention may be required for abscess, which does not respond to medical therapy.

HOSPITAL ACQUIRED PNEUMONIA

Incidence and Etiology

- Refers to a new episode of pneumonia 2 to 3 days after admission to hospital.
- The predisposing factors for the development of pneumonia are:
  - Immunocompromised state (e.g. corticosteroid treatment, malignancy, AIDS)
  - Reduced cough reflex (e.g. postoperative)
  - Aspiration of gastric contents (e.g. vomiting, severe reflux, nasogastric intubation)
  - Endotracheal intubation (e.g. mechanical ventilation)
  - Bacteremia (e.g. sepsis).

Symptoms

- Cough with purulent expectoration is the predominant symptom
- Breathlessness will appear soon.

Signs

- Cyanosis
- Crepitations are heard on auscultation.

Relevant Investigations

- Leukocytosis is present
- Chest X-ray will show mottled opacities in both lung fields.

Treatment

- Broad-spectrum intravenous antibiotics are necessary
- Physiotherapy is mandatory in immobile patients.

ACUTE LUNG ABSCESS

Incidence and Etiology

- A suppurative focus within the lung associated with necrosis
- The bacteria responsible for a lung abscess may reach the lung through various routes. They are:
  - Aspiration down the tracheobronchial tree (e.g. bronchiectasis)
  - By blood (e.g. septicemia)
  - Trauma (e.g. penetrating trauma)
  - Extension of adjacent suppurative focus (e.g. liver abscess).

- Other causes are:
  - Tumors of bronchus (e.g. bronchogenic carcinoma)
  - Suppuration of hematoma (e.g. penetrating trauma).

Symptoms

- Appear acutely ill
- Cough with foul smelling purulent expectoration
- Hemoptysis may occur
- Fever
- Breathlessless
- Chest pain.

Sign

Crepitations may be heard.

Relevant Investigations

X-ray and CT are diagnostic.

Treatment

- Antimicrobial therapy resolves most abscesses
- Surgical intervention (lobectomy or segmentectomy), is reserved for those which do not respond.

PULMONARY EMBOLISM

Incidence and Etiology

Majority of pulmonary emboli result from deep venous thrombosis of lower limbs, and they can be acute (minor and massive) or chronic.

Symptoms and Signs

The symptoms and signs are tabulated in Table 18.2.

Table 18.2: Symptoms and signs of pulmonary embolism

Type	*Symptoms*	*Signs*
Acute minor	Shortness of breath and hemoptysis – pleurisy	Pleural rub and signs of pleural effusion
Acute massive	Central chest pain, apprehension, low cardiac output and syncope	Sinus tachycardia, hypotension and peripheral vasoconstriction
Chronic	Exertional dyspnea, syncope and chest pain over months and years	Pulmonary hypertension, loud pulmonary component of second heart sound and a right ventricular heave.

Relevant Investigations

Chest radiographs, ECG, arterial blood gases, ventilation-perfusion lung scanning and pulmonary angiography are useful investigations.

Treatment

- *General:* Opiates to relieve pain and distress, resuscitation by external cardiac massage, with oxygen support
- Anticoagulation is necessary atleast for 5 days
- Thrombolytic therapy is used in acute massive types.

19

Breast

BREAST HEMATOMA

Incidence and Etiology

Associated with history of trauma such as seatbelt injury during a road traffic accident, or following a violent contraction of pectoralis muscles responding to a blow.

Symptoms

- Pain in the breast
- History of trauma.

Sign

Presence of bruise over the breast may be the contributory finding for diagnosis.

Relevant Investigations

Fine needle aspiration cytology (FNAC) and mammography are needed to rule out malignancy.

Treatment

Exploration and histopathological confirmation.

FIG. 19.1: Acute breast abscess

ACUTE BREAST ABSCESS

Incidence and Etiology

Collection of pus resulting from mastitis.

Symptoms

- Starts with dull ache, which proceeds on to throbbing pain
- Systemic manifestations like fever and malaise may develop.

Signs

- Superficial abscesses may show as fluctuant tender lumps
- Deep abscesses may show as severe cellulitis with edema of breast (Fig. 19.1) without fluctuation.

Relevant Investigations

- Polymorphonuclear leukocytosis is generally present
- Diagnostic aspiration may confirm the presence of pus.

Treatment

- Surgical drainage gives full relief
- Spontaneous or inadequate drainage may result in the formation of antibioma, chronic abscess, mammary fistula with purulent or sero sanguinous discharge.

20

Spine

DEGENERATIVE DISEASES OF DISK AND FACET JOINTS

Incidence and Etiology

Degenerative diseases of the spine are very common. They affect the lumbar, cervical and thoracic regions in decreasing order of frequency.

- Lumbar degenerative disease:
  - Prolapse of intervertebral disk: Disks between L5/S1 and L4/L5 are most often affected. Part of the gelatinous nucleus pulposus protrudes through the rent in the annulus fibrosus at its weakest part, which is posterolateral and compress the nerve roots (compressive radiculopathy) (Figs 20.1A to D). Occasionally, a large central disk prolapse at L5/S1 compresses the cauda equina (Cauda equina syndrome).
  - Facet joint disease: Degenerative changes cause root compression.
  - Lumbar canal stenosis: Increasing wear and tear, bony overgrowth and ligamentous hypertrophy can cause narrowing of the spinal canal.
- Cervical degenerative disease: Pathogenesis is the same as in lumbar disease.

Cauda equina syndrome consists of back pain, bilateral sciatica and urinary retention.

Symptoms and Signs

- Lumbar degenerative disease: Back pain is the constant feature, but varies according to the compression of the root:

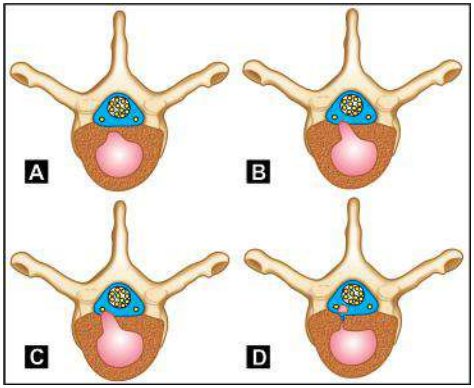

FIGS 20.1A to D: Stages of prolapse of intervertebral disk: (A) Torn annulus fibrosus with disk bulge; (B) Extrusion of nucleus pulposus protrusion not pressing the nerve; (C) Extrusion of nucleus pulposus protrusion pressing the nerve; (D) Sequestration of disk

FIG. 20.2: MRI—Prolapse of L4-L5 intervertebral disk

- ❖ Prolapse of intervertebral disk: Back pain, radiating to
  - Inner thigh (L4 root compression—L3/L4 prolapse)
  - Outer aspect of the leg to the big toe (L5 root compression—L4/L5 prolapse)
  - The sole of foot (S1 root—L5/S1 prolapse), is the presenting symptom.

  Acute radiating pain and inability to move, aggravated by coughing and sneezing is the symptom of acute disk prolapse. Straight leg raising test is positive on the affected side.
- ❖ Facet joint disease presents with pain similar to the synovial joint pain (more after a period of rest like early mornings and becomes less after movement)
- ❖ Lumbar canal stenosis causes numbness and weakness of lower limbs on walking.

- Cervical degenerative disease: The presenting symptom is neck pain related to movements with radiculopathy.

Relevant Investigations

CT and MRI (Fig. 20.2) are diagnostic.

Treatment

- Acute attacks require rest and analgesics, followed by physiotherapy
- Chronic, persistent or progressive symptoms require surgery to remove the prolapsed disk or decompressive laminectomy.

SPONDYLOLISTHESIS

Incidence and Etiology

- This term signifies slipping of one vertebra over the lower, which is usually forwards, but may be backwards. In the normal spine, forward displacement of a vertebral body is prevented by engagement of the articular processes with those of the segment next below it, and this check mechanism is weak in spondylolisthesis
- The displacement is most common in the 5th lumbar vertebra, and the displacement may increase slowly over years, and reach a severe degree of spondyloptosis.
- The predisposing factors are:
  - Spondylolysis (a defect in the pars interarticularis of the neural arch)
  - Osteoarthritis of the posterior (facet) joints (degenerative)
  - Congenital malformation of the articular processes (rare).

Symptom

Chronic backache with or without sciatica.

Signs

- A step above the sacral crest
- Lumbar vertebral bodies may be felt per abdomen due to its forward displacement
- Straight leg raising (SLR) test may be positive.

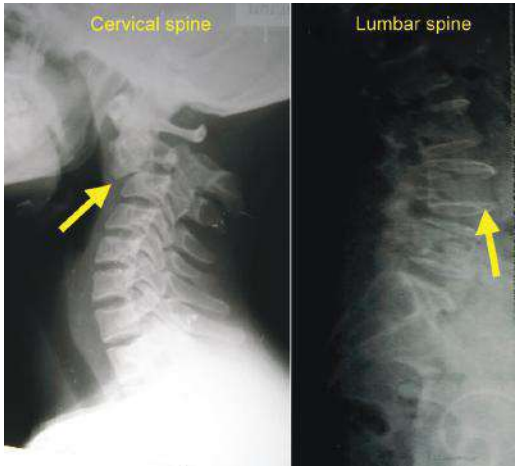

FIG. 20.3: X-rays—Spondylolisthesis

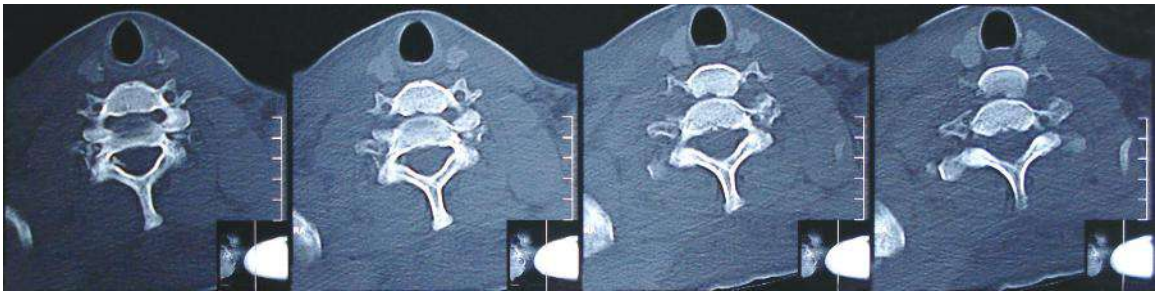

FIG. 20.4: CT—Spondylolisthesis (*Courtesy*: Dr V Thulasiraman)

FIG. 20.5: MRI spine—C4-C5 spondylolisthesis

Relevant Investigations

- X-rays (Fig. 20.3) and CT (Fig. 20.4) are diagnostic
- MRI (Fig. 20.5) is used to assess root compression.

Treatment

- Asymptomatic patients require no treatment
- Mild cases require surgical corset
- Surgery is justified only when the disability is severe.

21

Gastroenterology

ACUTE ABDOMEN

Introduction

Acute pain often denotes the presence of a disease process or injury, which needs to be treated with elimination of cause. It is a signal of ongoing or impending tissue damage. Acute abdominal pain denotes similar intra-abdominal organ disease or pathology, which warrants emergent care.

The main visceral pain receptors in the abdomen respond to mechanical and chemical stimuli.

- *Mechanical stimuli*: Stretch, distension, contraction, compression and torsion
- *Chemical stimuli*: Bradykinin, substance P, serotonin and prostaglandins. These receptors are located on the serosal surfaces, within the mesentery and within the walls of hollow viscera
- Gut related visceral pain is usually perceived in the midline because it is a midline structure in an embryo and has bilateral symmetric innervations, except for pains originating from the gallbladder and the ascending and descending colon. Pain from other intra-abdominal organs tends to be unilateral.
  - *Pain at epigastrium*: Diseases of the foregut (abdominal esophagus, stomach and proximal half of second part of duodenum and their offshoots like liver, gallbladder, pancreas and spleen) (e.g. gastric and duodenal ulcers)
  - *Pain at the umbilical region*: Diseases of midgut (distal half of second part of duodenum, small bowel, colon up to the proximal 2/3 of transverse colon) (e.g. intestinal tuberculosis)
  - *Pain at the hypogastrium*: Diseases of hindgut (distal 1/3 of transverse colon to the anorectal junction) (e.g. colorectal and urinary bladder pathologies)

The abdomen is divided into 10 arbitrary regions for convenience of understanding and localizing. The two lateral vertical planes pass from the costal margin close to the tip of the ninth cartilage above to the femoral artery below. The horizontal plane (the subcostal plane) connects the lowest points on the costal margins and the interiliac plane connects the tubercles of the iliac crests (Fig. 21.1). These divisions will help in localizing the diseases.

Pain

Nature of Pain

- *Sudden onset pain*: Pain is sudden in otherwise healthy and asymptomatic persons. It increases in a very short time (e.g. pain due to perforations of duodenal ulcer and appendicitis)
- *Sudden onset with pain-free intervals*: The pain of quick onset can reach a peak making the patient writhe in pain and buckle up, and also quickly recede to absolutely pain-free period, only to recur again (e.g. colics)
- *Dull continuous pain increasing in severity*: A persistent pain may increase in severity over a period of time (e.g. a dull pain of subacute appendicitis in the right iliac fossa may become severe when the appendicitis becomes severe and acute due to superadded infection and inflammation)
- *Burning pain*: The pain may be of burning nature occurring suddenly (e.g. pain in acid peptic disease)
- *Constant or continuous pain*: Persistent pain without variation in intensity (e.g. peritonitis)
- *Agonizing pain*: Very severe pain which upsets the morale of the patient (e.g. pancreatitis, torsion of pedicled organ)

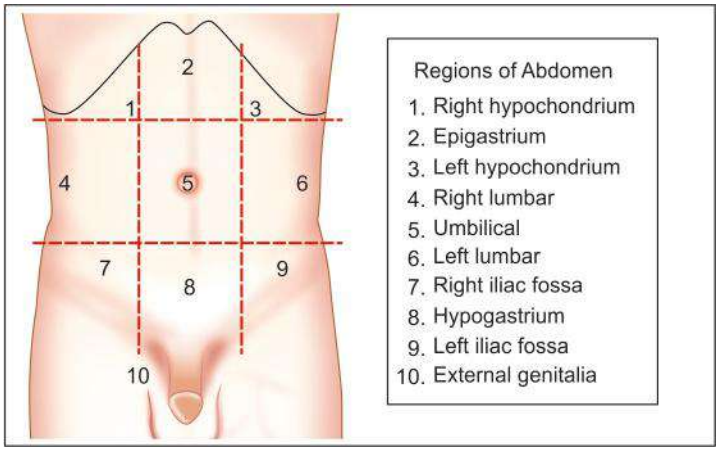

FIG. 21.1: Regions of abdomen

- *Throbbing pain*: Continuous pain throbbing in nature (e.g. acute cholecystitis—due to inflammation being inside closed confines of a structure).

Shift, Migration and Referred Pain

- ***Shifting pain***: Shifting pain is defined as the pain, which originates in one region and shifts totally to another region, without any pain at the point of origin (e.g. pain in acute appendicitis originates at the umbilical region and shifts to right iliac fossa later)
- ***Migrating pain***: Migrating pain is that pain which originates at one region and shifts to another region, but does continue to stay at a lesser degree at the point of origin (e.g. pain in perforated duodenal ulcer is right hypochondrial to start with, and spreads to the right iliac fossa due to the flow of the gastric contents down the right paracolic gutter)
- ***Referred pain***: The pathology and the area of the pain are different, since the organ of pathology and the area of pain share the same nerve supply. The examples are given in Figure 21.2

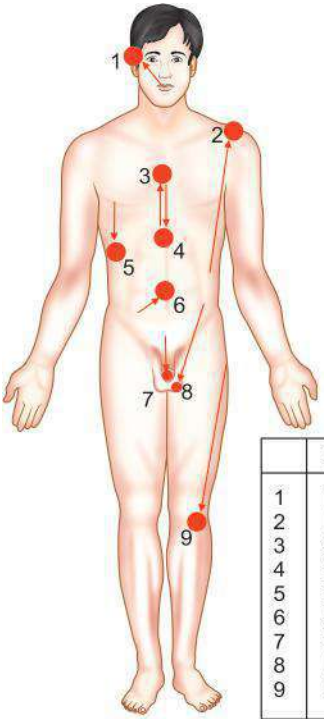

	Pathology	Pain
1	Tongue	Ear
2	Diaphragm	Shoulder
3	Stomach	Chest
4	Heart	Epigastrium
5	Pleura	Hypochondrium
6	Appendix	Umbilicus
7	Bladder	Penis
8	Ureter	Testis
9	Hip	Knee

FIG. 21.2: Referred pain

- The perception of visceral pain corresponds to the spinal segments where the visceral afferent nerve fibers enter the spinal cord. Table 21.1 shows some common spinal segments where visceral pain is perceived.

Effects of Movements, Food or Habits

The abdominal pain may be aggravated by certain factors like movements, food or habits. Some examples are given in Table 21.2.

The pain may get relieved by certain factors. Some examples are given in Table 21.3.

Table 21.1: Visceral pain and dermatomal perceptions

Organ of pathology	*Site of pain*	*Dermatome*
Stomach	Epigastrium	T5-T10
Small bowel	Umbilicus	T9-T10
Large bowel up to splenic flexure	Umbilicus	T11-L1
Large bowel from splenic flexure	Hypogastrium	L1-L2
Gallbladder	Epigastrium, scapular region	T7-T9
Pancreas	Epigastrium	T6-T10
Ureter	Loin to groin	Genitofemoral nerve (L1-L2)
Testis and ovary	Umbilicus	T10-T11

Note: Segmental nerve supply mentioned here is sympathetic supply of the viscus. Parasympathetic supply is from the vagus nerve, excepting for the hindgut and the urinary bladder, which is from the sacral segments.

Table 21.2: Aggravating factors of abdominal pain

Aggravating factor	*Pathology*
Physical movements like jolting	Appendicitis, peritonitis
Deep inspiration	Pleurisy
Intake of fatty food	Cholecystitis
Intake of spicy food, alcohol	Acid peptic disease
Lying supine	Pancreatitis
Drugs—Analgesics, NSAIDs	Acid peptic disease

Table 21.3: Relieving factors of abdominal pain

Relieving factor	*Pathology*
Vomiting	Peptic ulcer pain
Intake of bland food	Peptic ulcer pain
Local pressure	Colicky pain
Leaning forward	Pancreatitis
Drugs—antacids, H_2 blockers	Acid peptic disease

Vomiting

Vomiting is a very common feature associated with pain in acute abdominal emergencies. The patient should be asked about the following details, as each has its own significance. They are:

- *Character of vomiting*
  - *Projectile*: Involuntary projectile ejection of large quantities of vomitus (e.g. high intestinal obstruction)
  - *Regurgitative*: Effortless involuntary regurgitation of intestinal contents (e.g. peritonitis due to perforation)
- *Frequency of vomiting*
  - *Constant*: Persistent vomiting even in the absence of food intake (e.g. acute intestinal obstruction, acute pancreatitis)
  - *Periodical*: Vomiting with some periodicity or following food intake indicates bowel obstruction (e.g. acute peptic ulcer, gastric outlet obstruction)
- *Nature of vomitus*
  - Coffee ground vomitus (Brown to dark brown colour)—gastric contents with altered blood (e.g., bleeding duodenal ulcer)
  - Bloody (red in colour) (e.g. bleeding oesophageal varices)
  - Faeculent (yellowish green and foul smelling) (e.g. gastric contents followed by duodenal and intestinal contents in intestinal obstruction)
- *Quantity of vomitus*
  - Large quantities indicate distal bowel obstruction
  - Small quantities indicate gastric outlet obstruction
- *Relationship with pain*: The pain may precede, accompany or follow abdominal pain. The examples are given in Table 21.4.

Table 21.4: Relationship of vomiting with abdominal pain

Relationship of vomiting with abdominal pain	*Pathology*
Pain preceding vomiting	Acute appendicitis, acute peptic ulcer, biliary and renal colic
Pain and vomiting occurring together	High intestinal obstruction
Vomiting occurs much later than the pain	Low intestinal obstruction
Vomiting as a late feature or absent	Large intestinal obstruction

- Abdominal pain due to acute peptic ulcer may get relieved by vomiting, but gives temporary relief in colics
- Vomiting is not a constant feature in acute appendicitis as the stomach gets empty after one bout, but nausea persists
- In peritonitis, vomiting is absent but appears at a later stage.

Alteration in Bowel Habits

- ***Obstipation***: Inability to pass feces and flatus (absolute constipation) (e.g. acute intestinal obstruction)
- ***Constipation***: Inability to pass stools but able to pass flatus (due to solidified fecal mass impacted at the anal verge) (e.g. acute fissure in ano)
- ***Tenesmus***: A desire to evacuate but results in passage of flatus only (e.g. acute appendicitis)
- ***Passage of blood stained stools***: Red currant jelly stools in acute intussusception, mesenteric vascular obstruction, mesenteric thrombosis, malignancy
- ***Diarrhea***: Passage of loose stools—rarely seen in illnesses requiring emergency surgery (e.g. ulcerative colitis, Crohn's disease, acute enteritis).

Urinary Symptoms

Patients presenting with abdominal pain may have associated urinary symptoms:

- ***Frequency in micturition***: Patients presenting with renal or ureteric colic, may have associated urinary infections along with urolithiasis. Frequency is one of the common symptoms
- ***Strangury***: Frequent passing of urine with excruciating pain. (e.g. impacted stones in the urinary tract, pelvic or retrocecal appendicitis). Patient succeeds in passing a small quantity of blood stained urine
- ***Hematuria***: Passing blood in the urine (e.g. stones in the urinary tract, rarely retrocecal or pelvic appendix lying close to the ureter).

ACUTE UPPER ABDOMINAL PAIN

Causes

Right Hypochondrium

- Acute cholecystitis
- Acute cholangitis
- Acute hepatitis
- Acute hyperacidity
- Perforated duodenal ulcer

Epigastrium

- Acute hyperacidity
- Acute pancreatitis
- Perforated duodenal ulcer
- Acute hepatitis (left lobe)

Left Hypochondrium

- Acute pancreatitis
- Acute hyperacidity
- Splenic infarct

Eliciting History

1. *Nature of pain*
   - Continuous (e.g. acute pancreatitis)
   - Episodic (e.g. acute hyperacidity)
   - Colicky (e.g. biliary colic).
2. *Location of pain*
   - Epigastric pain (e.g. acute hyperacidity, acute pancreatitis, acute colitis)
   - Left hypochondrial pain (e.g. acute hyperacidity, left renal colic)
   - Right hypochondrial pain (e.g. acute cholecystitis, acute hepatitis, right renal colic).

Renal colics can present as upper abdominal pain in their respective sides.

3. *Association of vomiting*: Presence of vomiting is not a very reliable symptom to narrow down the diagnosis, as it can be present with any severe painful pathology in the upper abdomen.
4. *Association of fever*: Fever indicates infective pathology (e.g. acute cholecystitis, acute pancreatitis, acute colitis, perforated duodenal ulcer).

5. *Association of jaundice*: Jaundice may be present with acute cholangitis, acute hepatitis or acute cholecystitis.
6. *Association of loose stools*: Association of loose stools may indicate colitis or rarely acute pancreatitis.
7. *Radiation*: Radiation to right scapula or shoulder is common with acute cholecystitis, perforated duodenal ulcer due to irritation of diaphragm.
8. *Aggravating factors*
   - Food—in acute hyperacidity
   - Lying supine—in acute pancreatitis
   - Deep breathing—in acute cholecystitis.
9. *Relieving factors*: Leaning forward while sitting—acute pancreatitis.
10. *Referred pain*: In some pathologies, the pathology and the area of the pain are different, since both of them share the same nerve supply.

Past History

- History of pain (e.g. acute on chronic cholecystitis)
- Previous surgery (e.g. cholecystectomy will rule out cholecystitis from consideration).

Family History

Gallstones

Clinical Examination

General

- Breath for fetor (e.g. alcoholic hepatitis, acute pancreatitis)
- Conjunctiva for anemia, jaundice
- Tongue for anemia
- Neck for lymphadenopathy
- Hands for signs of liver failure (e.g. clubbing, palmar erythema, liver flap, etc).

Abdomen

Inspection

Distension:

- Generalized (e.g. perforated duodenal ulcer or gallbladder with peritonitis)
- Right upper abdominal (e.g. hepatomegaly)
- Epigastric (e.g. left lobar hepatomegaly, carcinoma stomach)
- Left upper abdominal (e.g. splenomegaly)
- Scars, swellings and sinuses.

Palpation

Tenderness:

- All quadrants—generalized peritonitis
- Right upper quadrant (e.g. acute hepatitis, acute cholecystitis, acute hyperacidity)
- Epigastric (e.g. acute gastritis, acute hepatitis)
- Left upper quadrant (e.g. acute gastritis, acute pancreatitis).

Lump:

- Right upper quadrant (e.g. hepatomegaly, distended gallbladder)
- Epigastric (e.g. carcinoma stomach, left lobar hepatomegaly)
- Left upper quadrant (e.g. carcinoma stomach, splenomegaly).

Percussion

Percuss the liver for:

- Its enlargement (e.g. acute hepatitis)
- Obliteration of liver dullness (e.g. perforated duodenal ulcer).

Auscultation

- Absence of bowel sounds indicates paralytic ileus (e.g. perforated peritonitis)
- Exaggerated bowel sounds may indicate obstruction of small bowel (e.g. intestinal colic)
- Normal bowel sounds indicate that there is no gross infection of the peritoneum.

Examination of

- Groins (e.g. obstructed hernia)
- Genitalia (e.g. obstructed hernia)

Differential Diagnosis by Clinical History and Examination

- *Right hypochondrial pain*
  - Without fever and but local tenderness (e.g. acute cholecystitis, acute hepatitis, acute hyperacidity)
  - With fever and local tenderness
    - + Distension (e.g. acute cholecystitis)
    - Obliteration of liver dullness (e.g. perforated duodenal ulcer)
    - + Hepatomegaly (e.g. acute liver abscess)
  - With vomiting
    - Local tenderness (e.g. acute cholecystitis, acute hyperacidity)
    - Local tenderness with obliteration of liver dullness (e.g. perforated duodenal ulcer)
  - With diarrhea and local tenderness (e.g. acute colitis)

- ❖ With jaundice
  - Local tenderness + hepatomegaly (e.g. acute hepatitis, acute cholecystitis, choledocholithiasis, cholangitis, acute liver abscess)
  - Local tenderness, fever, +/– abdominal lump (e.g. mucocele gallbladder, choledochal cyst)
- ❖ Colicky in nature +/– local tenderness (e.g. biliary colic, right renal colic)

- *Epigastric pain*
  - ❖ Without fever but with local tenderness (e.g. acute hyperacidity, acute pancreatitis, acute hepatitis—left lobe)
  - ❖ With fever and local tenderness + distension (e.g. acute pancreatitis, perforated duodenal or gastric ulcer)
  - ❖ With vomiting and
    - Local tenderness (e.g. acute hyperacidity, acute pancreatitis)
    - Local tenderness and obliteration of liver dullness (e.g. perforated ulcer)
  - ❖ With diarrhea and local tenderness (e.g. acute colitis)
  - ❖ With jaundice and local tenderness
  - ❖ + tender hepatomegaly (e.g. left lobar hepatitis, left lobar liver abscess
  - ❖ Nontender hepatomegaly (e.g. metastatic liver)
- *Left hypochondrial pain*
  - ❖ Without fever but with local tenderness (e.g. acute gastritis, acute pancreatitis)
  - ❖ With fever and local tenderness + /– splenomegaly (e.g. acute pancreatitis, splenic infarct)
  - ❖ With diarrhea and local tenderness (e.g. acute colitis)
  - ❖ Colicky in nature + local tenderness (e.g. left renal colic).

- Perforated bowel can give rise to generalized abdominal distension due to generalized peritonitis
- Obstructed groin hernia is one of the common causes of intestinal obstruction causing generalized distension.

Relevant Investigations

Hematology

- Leukocytosis in infective pathologies (e.g. acute cholecystitis, perforated dudodenal ulcer, perforated cholecystitis, acute pancreatitis)
- Raised ESR in all infective pathologies.

Radiology

- *Plain X-ray abdomen*: Gas under the diaphragm (e.g. perforated hollow viscus)
- *Ultrasonography:* Radiopaque shadows in right upper abdomen (e.g. gallstones, renal stones)

Treatment Plan

- *Nonperforated pathologies:* Medical management
- *Perforated pathologies:* Early surgical management
- Diagnosis not clear and not responding to medical management—exploratory laparotomy.

ACUTE LOWER ABDOMINAL PAIN

Causes

Right Iliac Fossa

- Acute appendicitis
- Perforated appendicitis
- Acute mesenteric adenitis
- Ureteric colic
- Acute Meckel's diverticulitis
- Acute cecal diverticulitis
- Torsion of cyst of right ovary
- Pelvic inflammatory disease
- Incarcerated right inguinal hernia
- Ruptured ectopic gestation
- Right ureteric colic.

Hypogastrium

- Acute cystitis
- Acute congestive dysmenorrhea
- Uterine fibroid.

Left Iliac Fossa

- Acute diverticulitis
- Ureteric colic
- Torsion of cyst of left ovary
- Pelvic inflammatory disease
- Incarcerated left inguinal hernia
- Ruptured ectopic gestation
- Left ureteric colic.

Diabetic ketoacidosis is one of the important metabolic causes of acute lower abdominal pain.

Eliciting History

Nature of pain

- Continuous (e.g. acute appendicitis)
- Episodic (e.g. acute cystitis, ruptured ectopic gestation)
- Colicky (e.g. appendicular colic, ureteric colic, dysmenorrhea).

Association of nausea and vomiting

Presence of vomiting is not a very reliable symptom to narrow down the diagnosis, as it can be present with any severe painful pathology in the lower abdomen. Nausea is a predominant symptom of acute appendicitis, whereas vomiting is a common symptom of ureteric colic.

Association of fever

Fever indicates infective pathology (e.g. acute appendicitis, acute cystitis, acute colitis, perforated appendicitis).

Association of loose stools

Association of loose stools may indicate colitis, acute diverticulitis.

Radiation

Radiation to pain to external genitalia occurs with ureteric colic.

Past History

- History of pain (e.g. acute on chronic appendicitis, ureteric colic)
- Previous surgery (e.g. appendicectomy will rule out appendicitis from consideration).

Personal History

Menstrual irregularities

- ❖ Dysmenorrhea (e.g. congestive dysmenorrhea)
- ❖ Irregularities (e.g. pelvic inflammatory diseases)
- ❖ Amenorrhea (e.g. ruptured ectopic gestation).

Family History

- Diverticulosis
- Colonic malignancy.

Clinical Examination

General

- Conjunctiva for anemia (e.g. ruptured ectopic gestation)
- Tongue for anemia (e.g. ruptured ectopic gestation)
- Neck for lymphadenopathy (e.g. mesenteric adenitis as a part of tuberculosis).

Abdomen

- ***Inspection***

  Distension:
  - Generalized (e.g. perforated appendicitis with generalized peritonitis)
  - Right lower abdominal (e.g. ruptured appendicitis, torsion of right ovary or its cyst)
  - Hypogastric (e.g., distended urinary bladder in cystitis, enlarged uterus)
  - Left lower abdominal (e.g. torsion of left ovary or its cyst)

  Scars, swellings and sinuses
- ***Palpation***

  Tenderness:
  - All quadrants – generalized peritonitis
  - Right lower quadrant (e.g. acute appendicitis, acute mesenteric adenitis)
  - Hypogastric (e.g. acute cystitis)
  - Left lower quadrant (e.g. acute colitis, acute diverticulitis)

  Lump:
  - Right lower quadrant (e.g. appendicular abscess, mesenteric adenitis, right ovarian cyst)
  - Hypogastric (e.g. distended urinary bladder, uterine fibroids)
  - Left lower quadrant (e.g. carcinoma colon, left ovarian cyst)
- ***Percussion***

  Percuss the liver for:
  - Its enlargement (e.g. associated metastases liver)
  - Obliteration of liver dullness (e.g. perforated appendicitis and diverticulitis)
- ***Auscultation***
  - Absence of bowel sounds indicates paralytic ileus (e.g. perforation and peritonitis)
  - Exaggerated bowel sounds may indicate obstruction of small bowel (e.g. intestinal colic)
  - Normal bowel sounds indicate that there is no gross infection of the peritoneum

Differential Diagnosis by Clinical History and Examination

- *Pain in the right iliac fossa with*
  - Nausea/vomiting and fever and local tenderness (e.g. acute catarrhal/perforated appendicitis, acute typhlitis, acute Meckel's diverticulitis, acute mesenteric adenitis)
  - Vomiting and
    - +/– Local tenderness (colicy pain—e.g. right ureteric colic, appendicular colic, radiating pain—right ureteric colic)
    - Menstrual irregularities and palpable lump (e.g. right ovarian cyst torsion)
    - Amenorrhea with local tenderness and anemia (e.g. ruptured ectopic gestation)
    - Abdominal lump (e.g. ileocecal tuberculosis, Crohn's disease, cecal malignancy, right ovarian malignancy, right ovarian cyst)
  - Diarrhea or dysentery and local tenderness (e.g. acute amebic typhlitis, cecal diverticulitis)
- *Pain in the left iliac fossa with*
  - Vomiting
  - +/– Local tenderness (e.g. left ureteric colic)
    - Menstrual irregularities and palpable lump (e.g. left ovarian cyst torsion)
    - Amenorrhea with local and cervical tenderness and anemia (e.g. ruptured ectopic gestation)
    - Abdominal lump (e.g., colonic malignancy, left ovarian malignancy, left ovarian cyst)
    - Diarrhea or dysentery with or without local tenderness (e.g. acute colitis, acute diverticulitis)
- *Hypogastric pain with*
  - Urinary symptoms with
    - Local tenderness (e.g. acute cystitis)
    - Distended urinary bladder (e.g. acute distension of bladder—bladder outlet obstruction)
  - Menstrual irregularities
    - +/– With or without local tenderness (e.g. congestive dysmenorrhea, pelvic inflammation)
    - Enlarged uterus (e.g. uterine fibroids, endometrial malignancy)
- *Acute lower abdominal pain (any or all quadrants of lower abdomen) with*
  - Menstrual irregularities and local tenderness (e.g. pelvic inflammatory disease)
  - Vague symptoms and signs (e.g. metabolic causes—diabetes mellitus).

Relevant Investigations

Hematology

- Reduced hematocrit (e.g. ruptured ectopic gestation, colonic malignancy)
- Leukocytosis in infective pathologies (e.g. acute appendicitis, perforated appendicitis, and diverticulitis, dudodenal ulcer)
- Raised ESR in all infective pathologies.

Radiology

- Plain X-ray abdomen
  - Gas under the diaphragm (e.g. perforated appendicitis and diverticulitis)
  - Opaque abdomen (e.g. peritonitis)
  - Radiopaque shadows (e.g. ureteric stone)
- Ultrasonography
  - Cystic swelling (e.g. ovarian cyst torsion)
  - Collection of pus (e.g. appendicular abscess, paracolic abscess)
  - Radiopaque shadows in the lines of ureters (e.g. ureteric calculus) and bladder region (e.g. vesical calculus)
  - Dilated pelvicalyceal system or ureters (e.g. obstructed ureteric stone).

Treatment Plan

- Nonperforated pathologies—medical management (except acute appendicitis)
- Perforated pathologies—early surgical management
- Diagnosis not clear and not responding to medical management—exploratory laparotomy or diagnostic laparoscopy.

ACUTE LIVER ABSCESS

Incidence and Etiology

- *Amebic liver abscess:* Amebic infections causing hepatitis and dysentery may resolve into a liver abscess (usually in alcoholics)
- *Pyogenic liver abscess:* Septicemic patients may present with pyemic abscesses in the liver.

Pathogenesis

- *Amebic liver abscess:* Amebic infections originating in the colon as amebic colitis, travel through the portal blood to reach the liver, forming an abscess more commonly in the right lobe of the liver
- *Pyogenic liver abscess:* Septicemic patients' infected blood reach the liver through the systemic and portal circulation, forming multiple abscesses in both lobes of the liver.

Symptoms

- Constitutional symptoms with high-grade fever, tachycardia and sometime shock.
- Patients with amebic etiology may give history of preceding diarrhea or dysentery.

Signs

- Intercostal tenderness is pathognomonic of a right lobar liver abscess.

Differential Diagnosis

- Lower thoracic lesions (basal pleurisy, pneumonia, and lung abscess), which irritate the diaphragm
- Acute cholecystitis
- Acute hepatitis.

Relevant Investigations

- Plain X-ray of abdomen or chest X-ray will reveal elevation of the right dome of diaphragm (Fig. 21.3). Right pleural effusion is common
- Fluoroscopy will show reduced mobility of the right dome
- US and CT (Fig. 21.4) are diagnostic
- Aspiration is confirmatory
- Stool examination is routine
- Isolation of organism in culture is required
- Serologic test (for amebiasis) is positive in majority of cases.

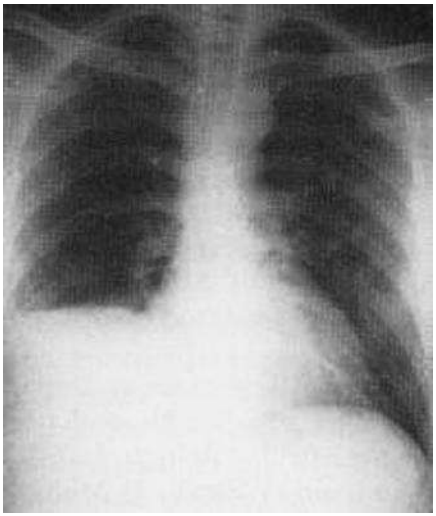

FIG. 21.3: Elevated right dome of diaphragm in amebic liver abscess

FIG. 21.4: CT—Amebic liver abscess right lobe with subphrenic collection

Treatment

- Antiamebic treatment is administered for small abscess and hepatitis of amebic etiology
- Drainage under US guidance with broad-spectrum antibiotics and antiamebic drugs is required for large abscess
- Growth of pyogenic organisms in culture of pus requires appropriate treatment.

ACUTE CHOLECYSTITIS

Incidence and Etiology

- Common in fat, flatulent, fertile, female of fifty (Five Fs)
- Ninety-five percent people with acute cholecystitis have gallstones.

Pathogenesis

- Inflammation of gallbladder occurs due to detergent action of bile (chemical cholecystitis), and infection by bacteria supervenes (commonly enteric organisms)
- Obstruction caused by a calculus in the cystic duct.

Complications

- Acute obstruction of the cystic duct causes distension and the bile can be replaced by mucus (mucocele) or pus (empyema)
- Ongoing inflammation causes gangrene and perforation of gallbladder (common in diabetics).

Symptoms

- Severe colicky pain (biliary colic) in the right hypochondrium, radiating to the inferior angle of the right scapula and the right shoulder
- Pain may be associated with vomiting.
- Fever and jaundice* are associated when there is associated cholangitis
- Symptom complex of pain, jaundice and fever with chills is called "Charcot's triad".

*Jaundice may occur due to the obstruction of the common bile duct caused by an impacted stone in the cystic duct (Mirizzi's syndrome type 1)

Signs

- Tenderness at the tip of the right ninth costal cartilage (Murphy's sign)
- Hyperesthesia between the right 9th and 11th ribs posteriorly (Boas's sign)
- There may be associated guarding and rigidity in the right hypochondrium
- Palpable mass in the RUQ (inflamed and distended gallbladder).

Differential Diagnosis

- Acute appendicitis
- Acute right pyelonephritis
- Duodenal ulcer perforation
- Right basal pleurisy
- Myocardial infarction.

Relevant Investigations

- Elevated leukocyte count
- Deranged LFT (due to inflammation and edema of biliary tract)
- Plain X-ray abdomen (About 10% of gallstones are visible)
- US (Figs 21.5A and B) is diagnostic
- Rarely CT scan (Figs 21.6A to C) may be required
- Isotope scans are contributory.

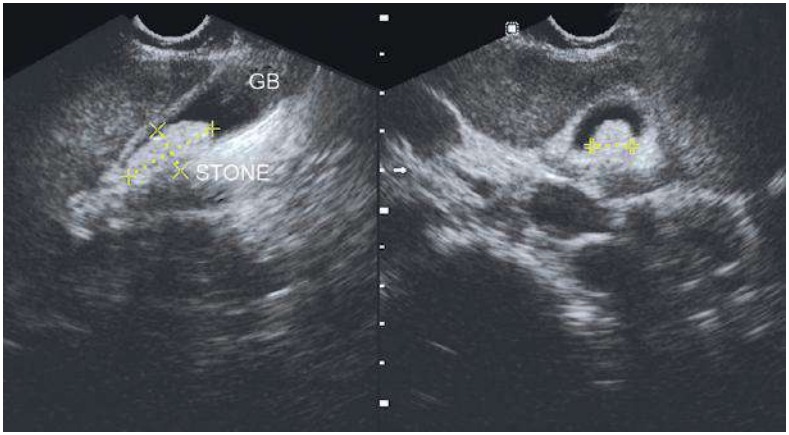

FIG. 21.5A: US—Stone in the cystic duct of gallbladder (*Courtesy:* Dr V Ganesan)

FIG. 21.5B: US—Acute edematous cholecystitis

FIG. 21.6A: CT—Stone in the gallbladder

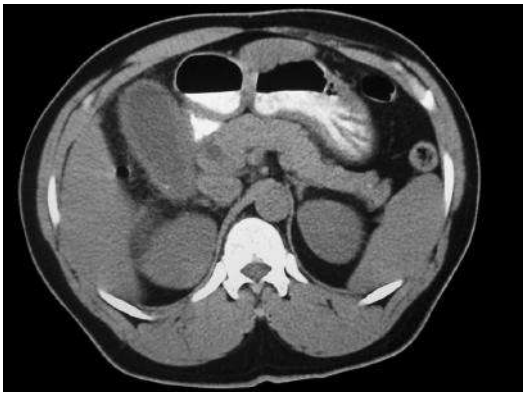

FIG. 21.6B: CT—Edematous gallbladder of acute cholecystitis

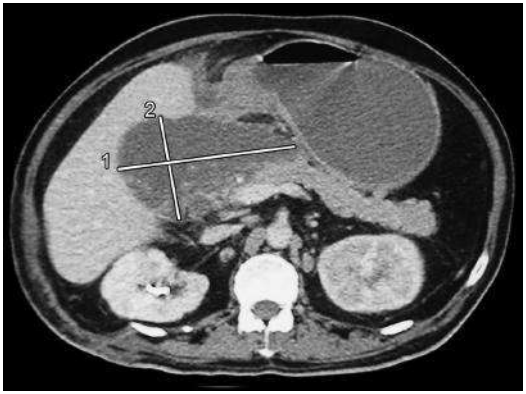

FIG. 21.6C: CT—Ruptured empyema of gallbladder

Treatment

- Acute cholecystitis is treated conservatively with antibiotics followed by elective cholecystectomy (Laparoscopic or open) **
- Common duct stones need to be removed by ERCP, sphincterotomy and basketing/stenting or by open choledochotomy, to relieve cholangitis and jaundice followed by cholecystectomy later.

** Laparoscopic cholecystectomy can be performed within 72 hours of onset of symptoms, with no higher conversion or complication rate, than a delayed operation.

ACUTE CHOLANGITIS

Incidence and Etiology

- Approximately 10 percent of patients with gallstones have associated common bile duct stones (choledocholithiasis)
- Obstructive lesions in the common bile duct can cause cholangitis (Fig. 21.7). They are:
  - Stricture
  - Tumor
  - Foreign body (e.g. stent)
  - Acute pancreatitis
  - ERCP

Pathogenesis

Obstructing lesion causes bile stasis, and this obstructed flow of bile encourages multiplication of bacteria, which have entered the biliary tract from the intestines through the sphincter of Oddi, causing ascending cholangitis.

Symptoms

RUQ pain, fever with rigors and jaundice (Charcot's triad).

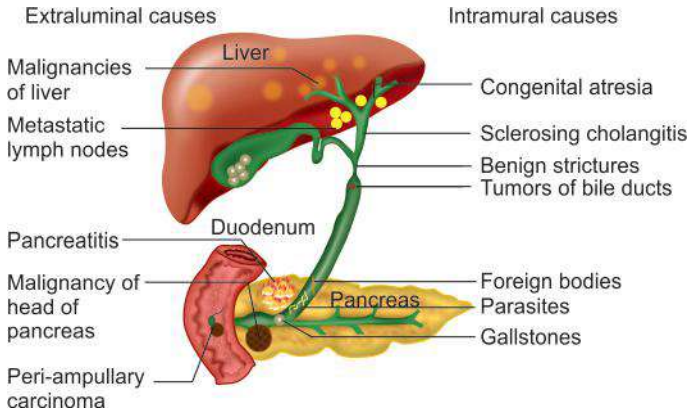

FIG. 21.7: Causes of mechanical obstruction of biliary system

Signs

- Pyrexia, tachycardia, tachypnea
- Hypotension in shocked in individuals
- Jaundice
- Tenderness and guarding in RUQ.

Relevant Investigations

- Elevated leukocyte count
- Conjugated bilirubinemia
- Deranged LFT (due to inflammation and edema of biliary tract)
- Plain X-ray abdomen (About 10% of gallstones are visible)
- US and CT may reveal the dilated intra-hepatic radicals
- ERCP (Fig. 21.8) and MRCP (Fig. 21.9) will reveal the level and severity of stricture
- PTC (Fig. 21.10) is useful in impassable strictures
- MDCT with MPR technique (Fig. 21.11) has sensitivity of about 90 to 95 percent in localizing the causes of biliary obstruction.

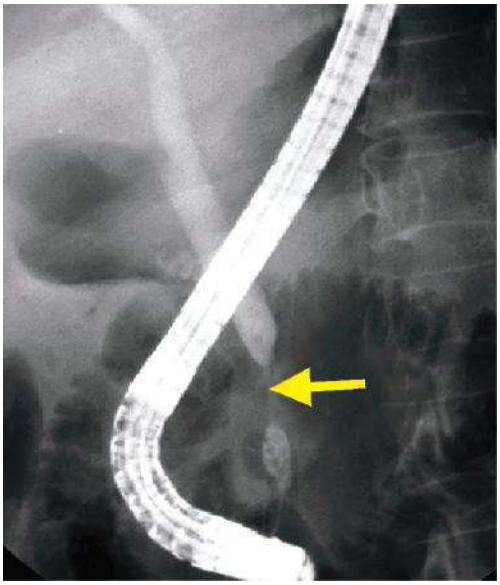

FIG. 21.8: ERCP—Benign stricture of CBD

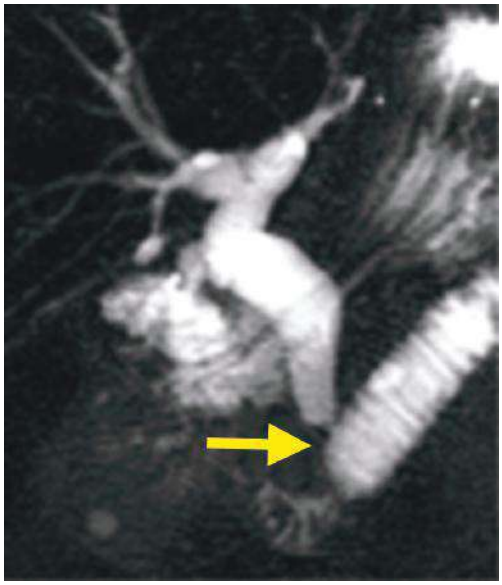

FIG. 21.9: MRCP—Benign stricture of CBD

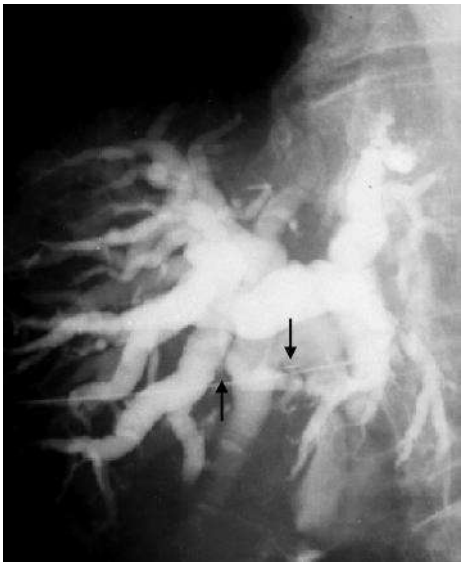

FIG. 21.10: PTC—Grossly dilated intrahepatic bile ducts due to CBD obstruction

FIG. 21.11: MDCT—Cholangiocarcinoma

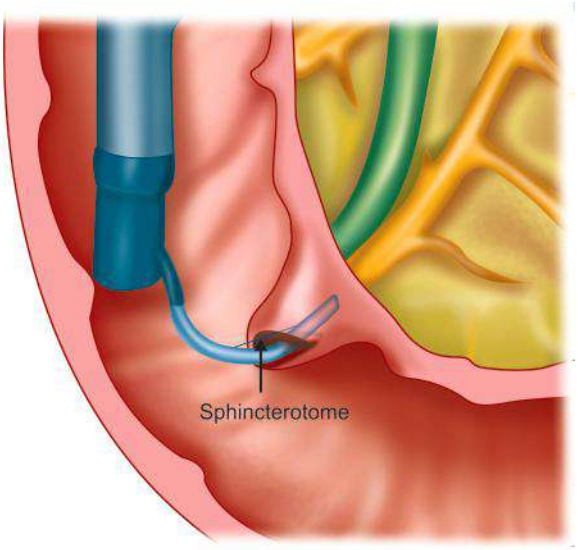

FIG. 21.12: Sphincterotomy

Treatment

- Inpatient care/Intensive care
- Analgesics
- Broad spectrum antibiotics
- Fluid resuscitation
- Emergency ERCP, sphincterotomy and basketing/stenting (Fig. 21.12) to relieve cholangitis and jaundice followed by cholecystectomy.
- Surgical treatment is required for impassable strictures
  - Resectable strictures can be resected with primary anastomosis
  - Strictures of the retropancreatic region require, choledocho-duodenostomy
  - Strictures of the common hepatic duct, require hepatico-jejunostomy.

Acute suppurative cholangitis is uncommon, but serious enough and if left untreated, has 100 percent mortality.

PRIMARY SCLEROSING CHOLANGITIS

Incidence and Etiology

- Rare condition of autoimmune origin
- Results in progressive fibrosis of the biliary system
- Causes luminal narrowing and progressive obstructive jaundice and secondary cirrhosis
- Strongly associated with inflammatory bowel disease, ulcerative colitis (75% of patients with PSC have inflammatory bowel disease and 7.5% with ulcerative colitis have PSC)
- Risk of malignant transformation (cholangiocarcinoma) is 15 percent.

Clinical Presentation

- Progressive obstructive jaundice
- Low-grade fever with chills, sweats.

Relevant Investigations

- CT (Fig. 21.13) and ERCP (Fig. 21.14) are diagnostic, classically shows diffuse stricturing and beading involving both intra and extrahepatic bile ducts, but indistinguishable from cholangiocarcinoma
- Liver biopsy may show the characteristic lesion of concentric fibrosis around small bile ducts, termed 'onion skin' fibrosis
- Peripheral antineutrophil cytoplasmic antibody (pANCA) is detected.

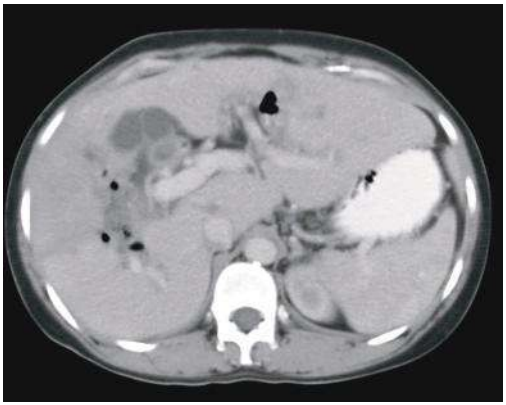

FIG. 21.13: CT—Sclerosing cholangitis

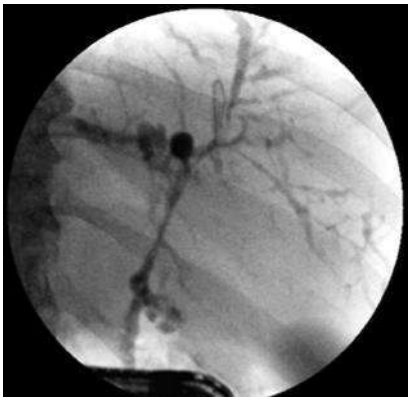

FIG. 21.14: ERCP—Sclerosing cholangitis

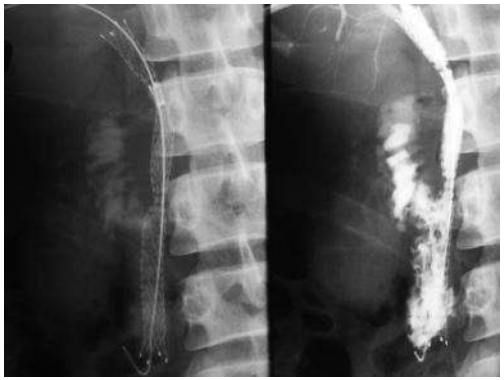

FIG. 21.15: Biliary stenting with endoprosthesis for cholangiocarcinoma

Treatment

- May settle spontaneously
- May respond to antibiotics, and UDCA
- Stenting of the biliary tree (Fig. 21.15) is very useful
- Liver transplantation is used widely for advanced disease.

SPLENIC ABSCESS

Incidence and Etiology

- Cyst of the spleen is a rare condition
- They can be:
  - Congenital
  - Parasitic (hydatid).

Pathogenesis

Severe systemic infections cause splenic abscesses which are usually multiple.

Symptoms

- Constitutional symptoms (e.g. high-grade fever)
- Previous history of severe intra-abdominal sepsis may be present.

Sign

Tender splenomegaly may be present.

Relevant Investigation

US and CT (Fig. 21.16) are diagnostic.

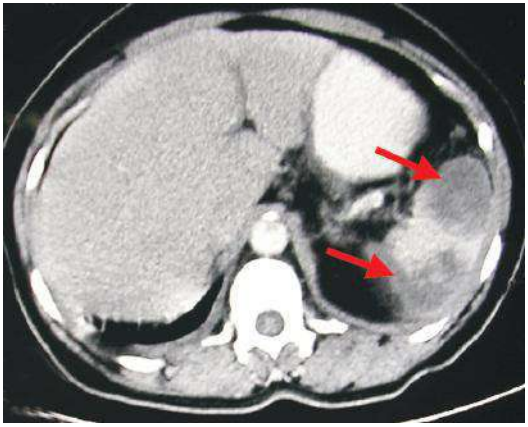

FIG. 21.16: CT abdomen—Splenic abscesses

Treatment

- Broad spectrum antibiotics
- Splenectomy is done for large spleen with abscesses.

ACUTE HEMORRHAGIC PANCREATITIS

Incidence and Etiology

- Common in the third decade
- Classification:
  - *Mild type:* Pancreatitis with minimal or no organ dysfunction and has a self-limiting course with uneventful recovery
  - *Severe type:* Pancreatitis with multiorgan failure and/or local complications like pseudocyst, necrosis or abscess
- Common causes are (nemonic: I get smashed)
  - I - Idiopathic
  - G - Gallstone (in the ampulla of Vater)
  - E - Ethanol
  - T- Trauma
  - S - Steroids
  - M - Mumps (paramyxovirus, Epstein-Barr virus, cytomegalovirus)
  - A - Autoimmune (Polyarteritis nodosa, Systemic lupus erythematosus)
  - S - Scorpion sting (Tityus Trinitatis), Snake bite
  - H - (Hypercalcemia, hyperlipidemia hypertriglyceridemia and hypothermia)
  - E - (Endoscopic Retrograde Cholangiopancreatography—ERCP)
  - D - Drugs (Steroids and Sulfonamides, Azathioprine, NSAIDs, Diuretics—SAND)
- Less common causes are:
  - Pancreas divisum
  - Long common duct
  - Carcinoma of the head of pancreas, and other cancer
  - Ascaris blocking pancreatic outflow
  - Chinese liver fluke
  - Ischemia from bypass surgery
  - Fatty necrosis
  - Pregnancy

- Infections other than mumps, including varicella zoster
- Repeated marathon running
- Cystic fibrosis.

- Most common causes of pancreatitis, by demography are as follows:
  - *Western countries:* Chronic alcoholism and gallstones
  - *Eastern countries:* Gallstones.

Pathogenesis

- In acute pancreatitis, there is edema, hemorrhage and necrosis of the organ partly due to autodigestion. The inflamed pancreas may return to normal, but may recur, and likely to occur under two circumstances:
  - The initiating cause has not been removed (gallstones, alcohol consumption)
  - Major pancreatic necrosis, resulting in chronic pancreatitis or stricture of main pancreatic duct.

Complications

- Systemic complications:
  - Respiratory failure
  - Renal failure
  - Metabolic abnormalities
  - Coagulation disorders
  - Multiple organ failure.
- Local complications:
  - Pancreatic necrosis
  - Infection of pancreatic necrosis
  - Fungal infections
  - Hemorrhage
  - Pancreatic pseudocyst, pancreatic fistula, pancreatic abscess.

Symptoms

- Very severe, unbearable constant epigastric pain radiating to the back, relieved by sitting and bending forwards
- Nausea and vomiting are marked, frequent and persistent.

Signs

- Shock and cyanosis are marked
- Elevated temperature, tachycardia, tachypnea

- Epigastric tenderness but guarding and rigidity are not marked
- Retroperitoneal hemorrhage may lead to characteristic discoloration:
  - ❖ Around the umbilicus (Cullen's sign) (Fig. 21.17)
  - ❖ Of the flanks (Grey Turner's sign) (Fig. 21.18)

  (These signs are not pathognomonic of pancreatitis, as they can be seen with ruptured ectopic pregnancy as well).

Differential Diagnosis

- Perforated peptic ulcer
- Leaking aortic aneurysm
- Myocardial infarction
- Acute cholecystitis
- Acute hyperacidity

Relevant Investigations

- Elevation of serum amylase over 400 Somogyi units is indicative and more than 1000 Somogyi units is diagnostic (It usually rises 2 to 12 hours from the onset of symptoms, and normalizes within 48 to 72 hours)
- Serum lipase levels are elevated (It rises 4 to 8 hours from the onset of symptoms and normalizes within 7 to 14 days).

FIG. 21.17: Cullen's sign

FIG. 21.18: Grey Turner's sign

Serum amylase may be normal (in 10% of cases) for cases of acute on chronic pancreatitis (depleted acinar cell mass) and hypertriglyceridemia. Reasons for false positive elevated serum amylase include salivary gland disease (elevated salivary amylase) and macroamylasemia. If the lipase level is about 2.5 to 3 times that of Amylase, it is an indication of pancreatitis due to alcohol

- Plain X-rays of abdomen show characteristic features:
  - ❖ 'Sentinel loop sign' – duodenum – which represents a focal dilated jejunal loop in the left upper quadrant
  - ❖ 'Cut off sign' – Transverse colon (Fig. 21.19) - Inflammatory exudate of acute pancreatitis extends into the phrenicocolic ligament directly spreading through the lateral attachment of the transverse mesocolon causing functional spasm and/or mechanical narrowing of the splenic flexure at the level where the colon returns to the retroperitoneum.

  (Absence of gas under the diaphragm eliminates the diagnosis of perforated duodenal ulcer)
- US may not be very useful at all times (Fig. 21.20)
- Contrast enhanced computed tomography (CECT) is very useful in assessing the size of the pancreas (Fig. 21.21), and also in determining the causes like the biliary or pancreatic duct calculi (Fig. 21.22).

Predicting the severity of an attack of acute pancreatitis is made on Glasgow or Imrie criteria (simplified Ranson criteria) Box 21.1.

Presence of three or more criteria reached before or at 48 hours of an attack predicts a severe attack and two or less predicts a mild attack.

APACHE II score (Acute Physiology and Chronic Health Evaluation) can be applied at anytime but it is cumbersome as it requires 15 different or biochemical criteria.

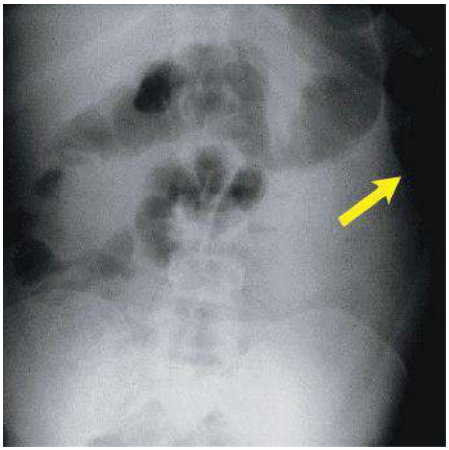

FIG. 21.19: X-ray—Colon cut-off sign

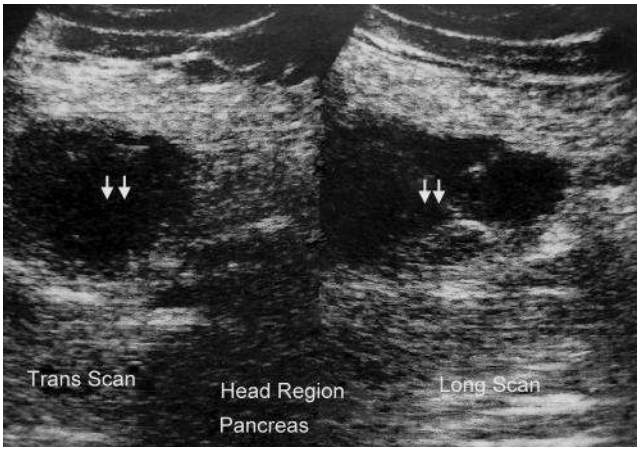

FIG. 21.20: US—Acute pancreatitis

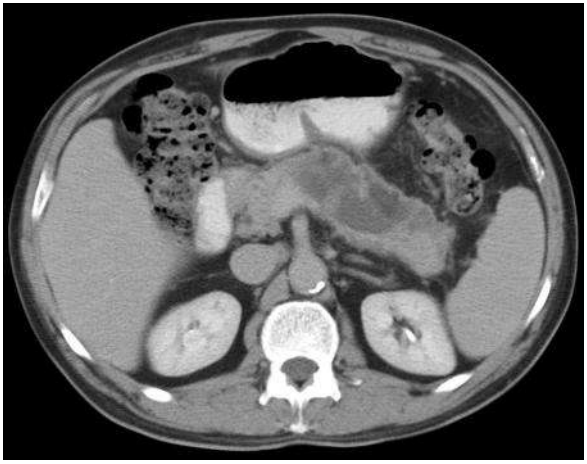

FIG. 21.21: CT—Enlarged and edematous pancreas - Acute pancreas

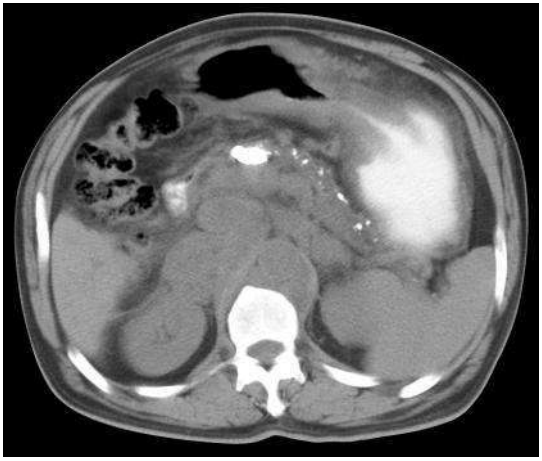

FIG. 21.22: CT—Pancreatic calculi

Box 21.1: Criteria to assess the severity of acute pancreatitis

P	Arterial PaO_2 <9 kpa
A	Albumin <32 g/L
N	Urea Nitrogen >10 mmol/L
C	Calcium <2 mmol/L
R	Raised white cell count >16 mmol/L
E	Enzyme – lactate dehydrogenase >600 mmol/L
A	Age >55 years
S	Sugar glucose >10 mmol/L

Treatment

- Initial management is conservative in intensive care unit:
  - Continuous arterial and CVP monitoring
  - Assisted ventilation if required
  - Inotropic support

- ❖ Enteral feeding
- ❖ Parenteral feeding if required
- ❖ Hemodialysis of renal failure if warranted

- Endoscopic treatment:
  - ❖ ERCP sphincterotomy and extraction of stones followed by laparoscopic cholecystectomy
- Pancreatic necrosectomy is the treatment of choice with questionable outcome in a grave situation.

ACUTE APPENDICITIS

Incidence and Etiology

- Most common acute abdominal condition, occurring at any age
- Uncommon only below the age of two
- Most commonly seen between the age of 18 and 35.
- The two varieties of appendicitis are:
  1. Catarrhal appendicitis
  2. Obstructive appendicitis.

Pathogenesis

- *Catarrhal appendicitis:* Occurs due to acute inflammation of the appendix, which produces edema and even gangrene due to vascular involvement in inflammatory process
- *Obstructive appendicitis:* Caused by obstruction of its lumen by worms, fecoliths (Fig. 21.23) or hypertrophied lymphoid follicles. The appendix itself may be filled with pus (Fig. 21.24).

Complications

- When the adjacent tissues and omentum wall off the appendix or its perforation, it forms a mass called 'Appendicular mass'.
- When there is suppuration, it forms an abscess named 'Appendicular abscess', which may burst into peritoneal cavity to produce severe peritonitis and even death.

Symptoms

- *Pain:* A dull continuous ache starting at the umbilical region (visceral pain) and then localizing at the right iliac fossa (parietal pain) - catarrhal variety. Obstructive appendicitis presents with colicky pain (appendicular colic) in the right lower abdomen
- Nausea, vomiting and anorexia are usually present and are diagnostic

FIG. 21.23: Appendix with fecoliths

FIG. 21.24: Acute suppurative appendicitis (note the pus in the cup)

- Fever is the last to develop
  (Pain, vomiting and fever in appendicitis is called Murphy's syndrome).

Signs

- Hyperesthesia over Sherren's triangle
- Tenderness over McBurney's point
- Guarding and rigidity in the right iliac fossa
- Rovsing's sign (pain in the right iliac fossa on application of pressure in the left iliac fossa) may be elicited
- Tender mass may be felt (appendicular mass/abscess)
- Dullness on percussion (if mass already formed).

The signs and symptoms of appendicitis vary according to the position of the appendix (Fig. 21.25). They are given in Table 21.5.

Special Situations

- In children:
  - Constitutional symptoms like fever and tachycardia are more predominant
  - Use the child's hand itself for palpation, and if there is tenderness in the McBurney's point, the child will withdraw its hand
  - Appendicular mass is rare as the omentum is small in size and does not reach the appendix.

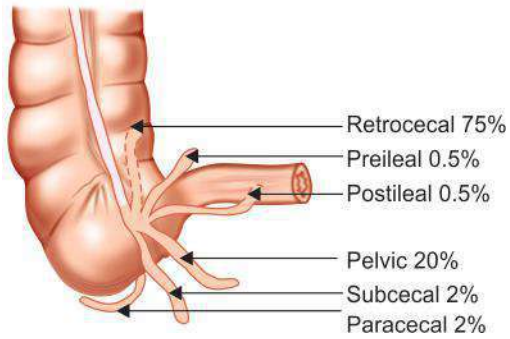

FIG. 21.25: Various positions of appendix

Table 21.5: Signs and symptoms of acute appendicitis related to the position of the appendix

Symptoms and signs	*Retrocecal and paracecal appendicitis*	*Pelvic and subcecal appendicitis*	*Pre and post-ileal appendicitis*
Pain	Right flank and back lateral to sacrospinalis muscle	Right iliac fossa	Right iliac fossa
Diarrhea	Absent	Absent	May be present
Tenderness and guarding	Not marked	Absent	Present
Positive test	Cope's psoas test Baldwing's test	Obturator test	Nil specific
Tenderness in rectal examination	Absent	Present	May be present

- In the elderly:
  - Guarding and rigidity are not pronounced as the abdominal musculature is weak
  - Incidence of gangrene is more as there is associated atherosclerosis
  - Peritonitis supervenes early.
- In pregnant women:
  - The point of tenderness is shifted up, as the appendix itself is pushed up by the enlarged gravid uterus
  - Pyelitis and cystitis of pregnancy adds to the difficulties in diagnosis of appendicitis
  - Accidental hemorrhage mimics acute appendicitis.

Cardinal signs and symptoms of acute appendicitis are:

- Periumbilical pain shifting to the right iliac fossa (very important)
- Association of nausea
- History of similar episodes in the past
- Tenderness at McBurney's point
- Guarding and rigidity in the right iliac fossa.

Differential diagnosis

- Right ureteric colic (most common)
- Right ovarian pathology
- Acute mesenteric adenitis
- Acute cholecystitis.

The differentiating features of acute appendicitis and right ureteric colic are given in Table 21.6.

Relevant Investigations

- US (Figs 21.26A and B) may be contributory. The immobile swollen appendix with free fluid in the right iliac fossa may be imaged. Mildly swollen appendix is generally not seen in an ultrasound scan. The scan helps to eliminate other lesions like the ureteric calculus, ovarian pathology, which can be imaged by US
- CT is useful in identifying inflamed appendix and (Fig. 21.27) appendicular mass. The signs of appendicitis are tabulated in Table 21.7.

Table 21.6: Differentiating features of acute appendicitis and right ureteric colic

Sign and symptom	*Pathology*	
	Acute appendicitis	*Right ureteric colic*
Pattern of pain	Periumbilical pain shifting to right iliac fossa	Loin to groin radiation on the right side
Nature of pain	Dull and continuous	Very severe and colicky with painfree intervals
Onset of pain	Slow and continuous	Sudden onset
Relationship to body movements	Aggravated	Not related to movement
Urinary symptoms	Absent	May be present
Rebound tenderness	May be present	Absent
Plain X-ray of abdomen	Nonspecific	Ureteric calculus

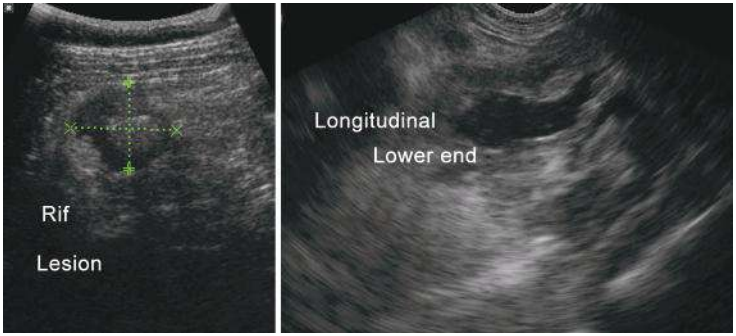

FIG. 21.26A: US—Inflamed appendix (*Courtesy*: Dr V Ganesan)

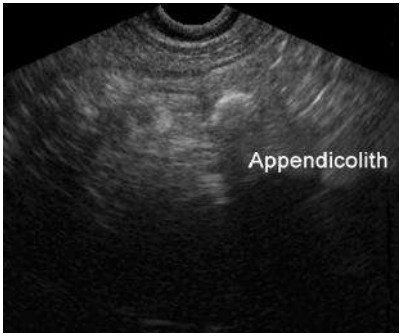

FIG. 21.26B: US—Appendicolith (*Courtesy*: Dr V Ganesan)

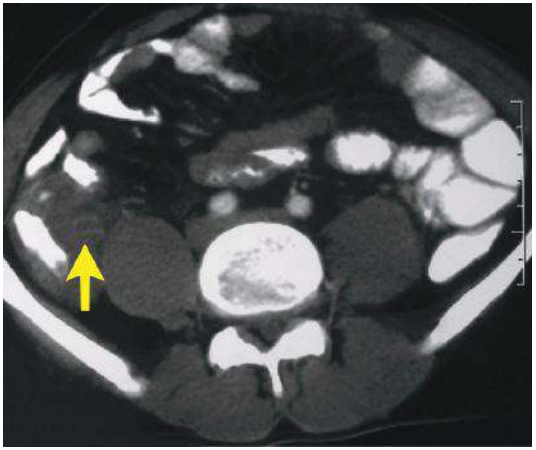

FIG. 21.27: CT—Appendicular mass

Table 21.7: CT findings associated with acute appendicitis

Appendiceal signs	*Periappendiceal signs*
Appendix >6 mm in anteroposterior diameter	Increased fat attenuation (stranding) in the right lower quadrant
Failure of appendix to fill with oral contrast or gas to its tip	Cecal wall thickening
Enhancement of appendix with IV contrast	Phlegmon in the right lower quadrant
Appendicolith	Abscess or extraluminal gas
	Fluid in the right lower quadrant or pelvis

Treatment

- Emergency appendicectomy (Laparoscopic or open) is the treatment of choice
- Conservative management is adopted for appendicular mass, subacute or chronic appendicitis (Oschner Scherren regime), followed by appendicectomy at a later date (interval appendicectomy)
- Appendicular abscess requires drainage, followed by interval appendicectomy.

ACUTE MESENTERIC LYMPHADENITIS

Incidence and Etiology

- Seen commonly in the pediatric age group
- Acute infection and inflammation of the mesenteric lymph nodes mostly of the ileocolic group (Fig. 21.28).

Pathogenesis

Commonly caused by viral infections.

Symptoms

- Periumbilical pain associated with high grade fever
- Vomiting is rare.

Signs

- The point of tenderness is usually in the right iliac fossa in the supine position, and shifts to the left side, if the patient is made to lie on the left side and vice versa ("Klein's sign")
- Guarding is not predominantly present
- In thin children, enlarged lymph nodes may be felt.

Differential Diagnosis

- Tubercular infection
- Acute appendicitis.

Relevant Investigations

- Clinical suspicion is important
- US (Fig. 21.29) and CT may show large lymph node swellings.

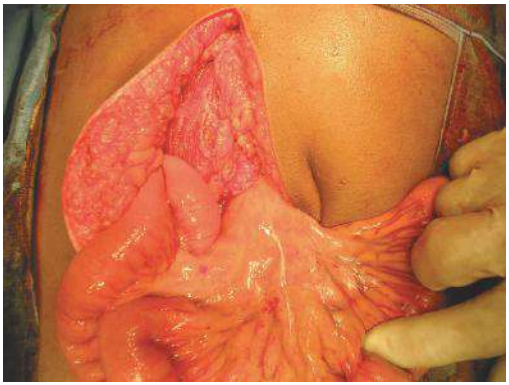

FIG. 21.28: Mesenteric lymphadenitis

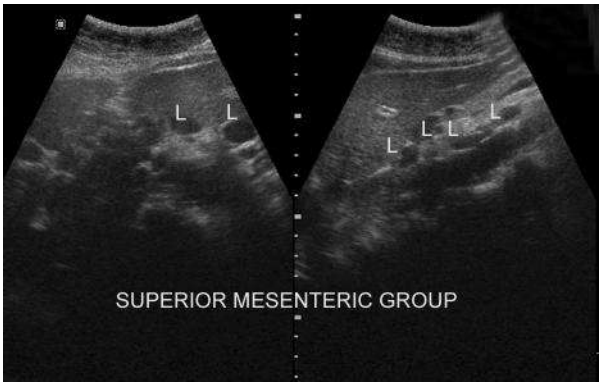

FIG. 21.29: US—Mesenteric lymphadenitis
(*Courtesy*: Dr V Ganesan)

Treatment

Medical management with antibiotics and supportive measures.

ACUTE COLONIC DIVERTICULITIS

Incidence and Etiology

- Common in the West
- Disease of the elderly
- Sigmoid colon is most commonly affected.

Pathogenesis

- Acute inflammation of the diverticulum of large bowel (colonic diverticulitis)
- They are thought to arise from increased pressure in the colonic lumen, occurs at weak areas between the taeniae where vessels perforate through the submucosal layer.

Complications

- Bleeding
- Perforation
- Peritonitis
- Abscess.

FIG. 21.30: Barium enema— Diverticulitis of colon

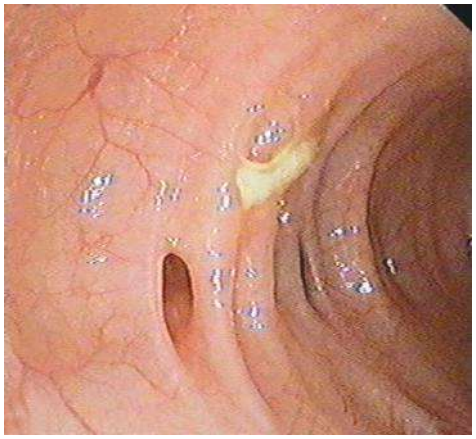

FIG. 21.31: Colonoscopy—Acute diverticulitis (*Courtesy*: Dr Mani Veeraraghavan)

The severity of diverticulitis is classified by Modified Hinchey classification

- Stage 1 – pericolic abscess
- Stage 2a – distant abscess amenable to percutaneous drainage
- Stage 2b – complex abscess associated with/without fistula
- Stage 3 – generalized purulent peritonitis
- Stage 4 – fecal peritonitis

Symptoms

- Colonic diverticulitis presents with colicky pain in the left flank
- When perforated, it may form an abscess in the paracolic region and present with high grade fever and a palpable tender lump.

Signs

- Left iliac fossa tenderness
- Tender mass (abscess) may be palpable.

Relevant Investigations

- Double contrast barium enema (Fig. 21.30) is informative
- Colonoscopy (Fig. 21.31) is diagnostic
- US and CT (Fig. 21.32) are useful in diagnosing abscesses.

FIG. 21.32: CT—Diverticulitis of sigmoid colon

Treatment

- Uncomplicated diverticulitis needs to be treated with antibiotics
- Perforated diverticulitis with or without abscess formation needs surgical intervention.

ACUTE MECKEL'S DIVERTICULITIS

Incidence and Etiology

- Meckel's diverticulum* - An embryological remnant (Fig. 21.33) present as a free diverticulum with a wide mouth, about 25 to 30 cm from the ileocecal junction; may contain ectopic gastric mucosa (reason not known)
- Occurrence follows rough rule of 2s:
  - 2 percent of the population
  - 2 feet from the ileo-cecal junction
  - 2 inches in length
  - 2 times more common in males than in females.

Complications: Perforation and peritonitis.

* Meckel's diverticulum is often found incidentally during laparotomy and remains asymptomatic in majority of individuals.

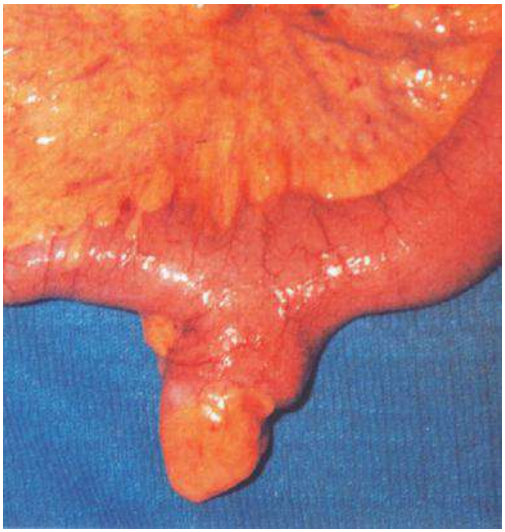

FIG. 21.33: Meckel's diverticulum

Symptom

Right iliac fossa pain.

Sign

Tenderness right iliac fossa (McBurney's point).

> Meckel's diverticulitis may present with peptic ulceration, lower GI hemorrhage, perforation, intussusception and intestinal obstruction.

Differential Diagnosis

- Acute appendicitis
- Right ureteric colic
- Right ovarian pathology in women.

Relevant Investigations

No investigation is useful.

Treatment

- Acute diverticulitis warrants diverticulectomy
- Perforation and peritonitis need appropriate management.

ACUTE SOLITARY CECAL DIVERTICULITIS

Incidence and Etiology

- Solitary diverticulum is common in the cecum
- Inflammation of the diverticulum presents like acute appendicitis and mislead the examiner.

Symptom

Pain in the right iliac fossa (similar to acute appendicitis).

Sign

A lump may be felt in the right iliac fossa.

> Clinically it mimics acute appendicitis, and many times, cecal diverticulitis is identified in second surgery, after appendicectomy.

Relevant Investigations

US and CT may be useful.

Treatment

Diverticulectomy or right hemicolectomy is the treatment of choice.

ACUTE ULCERATIVE COLITIS

Incidence and Etiology

- A chronic inflammatory bowel disease with relapses and remissions
- Acute inflammation is one of its clinical presentations.

Complications: Perforation and peritonitis.

FIG. 21.34: X-ray—Toxic megacolon

Symptoms

- Incessant diarrhea, mixed with blood, mucus and pus with constitutional symptoms
- They present an emaciated appearance.

Signs

Deep ulcers involving the entire colon is called *Toxic megacolon*, presenting as grossly distended abdomen.

Relevant Investigations

- Plain X-ray abdomen (Fig. 21.34) is diagnostic
- Colonoscopy is essential for diagnosis.

Treatment

- Uncomplicated cases are treated medically
- Perforation and peritonitis need appropriate treatment.

ACUTE INTESTINAL OBSTRUCTION

- Intestinal obstructions may be dynamic or adynamic.
  - ❖ Dynamic obstruction is a physical or mechanical obstruction of the intestinal lumen due to various causes. They are:
    - Intramural causes (e.g. fecal impaction, worms)
    - Mural causes (e.g. strictures or stenosis due to tuberculosis, malignancies)
    - Extramural causes (e.g. adhesions, obstructed hernia)
  - ❖ Adynamic obstruction of the bowel is due to its neural (autonomic) paralysis, commonly seen after abdominal surgery in the immediate postoperative period, peritonitis or any other cause like fractures, tight bandages, etc.
- Intestinal obstruction may be at different levels. They are:
  - ❖ Small intestinal (high or low)
  - ❖ Large intestinal (colonic or sigmoid)
- Causes of intestinal obstruction are different for different age groups (Table 21.8).

Symptoms

1. Sudden episodic colicky abdominal pain
2. Vomiting
3. Constipation
4. Abdominal distension

(The symptoms vary according to the level of obstruction)

- *Abdominal pain:* It is sudden and squeezing, and the patient doubles up. It is felt in the umbilical region, sometimes accompanied by the appearance of a contracting loop. There may be pain free intervals. Colonic pain presents in the hypogastrium

Table 21.8: Causes of intestinal obstruction in different age groups

Newborn	*Infants*	*Adolescents*	*Adults*	*Elderly*	*Rare causes*
Duodenal atresia	Helminths	Bands	Postoperative adhesions	Growth	Enteroliths
Pyloric stenosis	Intussusception	Intussusception	Intussusception	Intussusception	Foreign bodies
Meconium ileus		Meckel's diverticulum	Volvulus	Obstructed or strangulated hernia	Gallstones
Hirschsprung's disease		Obstructed or strangulated hernia	Obstructed or strangulated hernia		Trichobezoar
			Growth		Phytobezoar

- *Vomiting:* Vomiting is predominant in high obstructions. The vomitus consists of gastric contents, followed by the duodenal and lastly the intestinal, depending on the level of obstruction. In the late stages, the vomitus becomes feculent – ominous sign. Vomiting by itself is a late sign of chronic intestinal obstruction
- *Constipation:* The patient evacuates his bowel (contents distal to obstruction) once or twice, and constipation becomes a noticeable feature after 24 hours

Diarrhea can be a feature in certain situations like intussusception (red currant jelly stools), Richter's hernia, adynamic obstruction caused by mesenteric vascular occlusion, pelvic abscess, etc.

- *Abdominal distension:* Common feature of intestinal obstruction. Distension is
  - centrally located in small bowel obstruction (ladder pattern),
  - more on the flanks when distal colon is obstructed (asymmetrical)
  - more on the left flank in sigmoid volvulus
- *Dehydration:* When vomiting is pronounced as in high level obstructions, dehydration is a presenting feature.

Early dehydration and less abdominal distension suggests duodenal or jejunal obstruction whereas, late dehydration and more abdominal distension suggests distal ileal obstruction. Vomiting and dehydration are usually not present in isolated acute colonic obstruction.

Signs

- *General:* Pulse rate and blood pressure are maintained at normal levels in the initial stages. As dehydration becomes prominent, tachycardia and hypotension result
- *Abdomen:* Bowel sounds are not heard as obstruction worsens.

The summary of signs and symptoms related to intestinal obstruction are given in Table 21.9.

Relevant Investigations

- Plain X-rays of abdomen in the erect posture will reveal multiple air fluid levels (Fig. 21.35) and colonic obstruction may show distended colon also (Fig. 21.36).

Table 21.9: Signs and symptoms related to intestinal obstruction

Level of obstruction	*Signs and symptoms*					
	Duration of colic	*Pain free interval*	*Vomiting*	*Distension of abdomen*	*Constipation*	*Dehydration*
High	Short	Short	More	Minimal	Not constant	Severe
Low	Long	Long	Less	More	Late feature	Mild-to-moderate

(Findings of clinical examination and treatment of individual diseases are discussed at appropriate headings in the following pages)

FIG. 21.35: X-ray—Multiple air fluid levels of small bowel obstruction

FIG. 21.36: X-ray—Distention of large bowel

Treatment

- Inpatient/intensive care
- Nil by mouth
- Intravenous fluids
- Correction of electrolytes
- Nasogastric decompression
- Urinary catheterization of better monitoring
- Identifying and treating the cause.*

* If adhesions are the likely cause, a trial of conservative management is justified.

GALLSTONE ILEUS

Incidence and Etiology

Pathogenesis

- Gallstone enters the bowel through a perforated gallbladder (postcholecystitis) adherent to the small bowel (cholecysto-enteric fistula)
- When the stone is big and reaches the ileocecal junction, it causes small bowel obstruction.

Symptoms

Previous history of vague attacks of right upper quadrant pain, suggesting frequent cholecystitis.

Signs

Features of intestinal obstruction.

Relevant Investigations

- Plain X-ray abdomen (Fig. 21.37) almost always shows air in the biliary tree as bowel gas passes through the cholecystoenteric fistula. A gallstone may also be seen in the right lower quadrant, if it is radiopaque
- CT (Fig. 21.38) is more informative. Rarely, a gallstone may also be seen in the small bowel, and also in the gallbladder if there are many.

Treatment

- During laparotomy:
  - Simple crushing of stone with finger from outside the bowel may be enough if it is soft
  - Simple surgical removal through an enterotomy is required if the stone is hard and big
  - Cholecystectomy must be performed with closure of fistula.

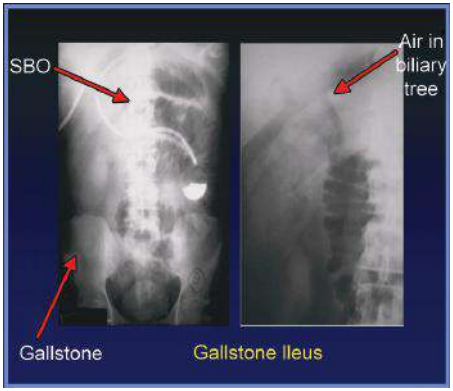

FIG. 21.37: X-ray—Gallstone ileus

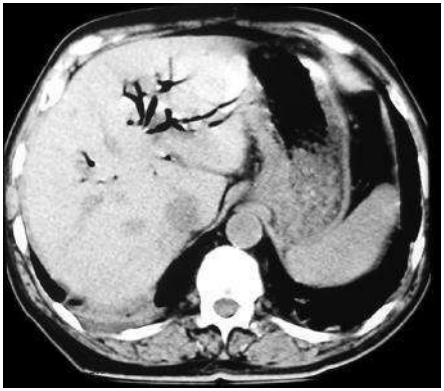

FIG. 21.38: Air in the biliary system

ACUTE INTUSSUSCEPTION

Incidence and Etiology

Intussusception is the invagination of a segment of bowel into the distal adjacent loop (proximal into the distal) (Fig. 21.39).

In Children

- Two per thousand infants are affected with male preponderance, commonly affecting the age group of 3 months to 1 year
- Commonly, it is secondary to an enlarged Peyer's patch due to viral or bacterial infections
- The other less common causes are:
  - Meckel's diverticulum
  - Duplication cyst in the bowel wall
  - Polyp.

In Adults

- Intussusception of small bowel is always secondary to a polypoid lesion, a lipoma (Figs 21.40A and B)
- In large bowel, it is due to a malignant polypoid lesion (Figs 21.41A and B).

FIG. 21.39: Ileoileal intussusception

FIG. 21.40A: Ileoileal intussusception

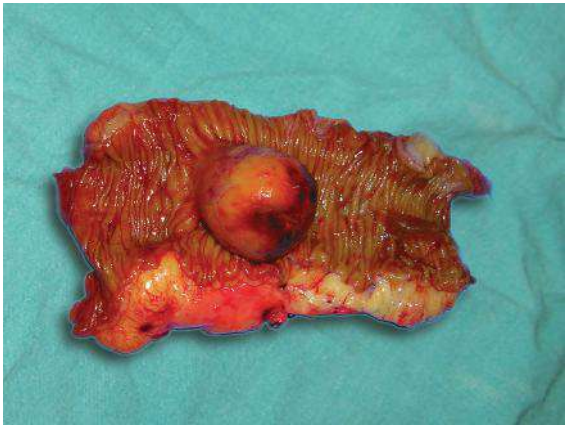

FIG. 21.40B: Lipoma the cause of ileoileal intussusception

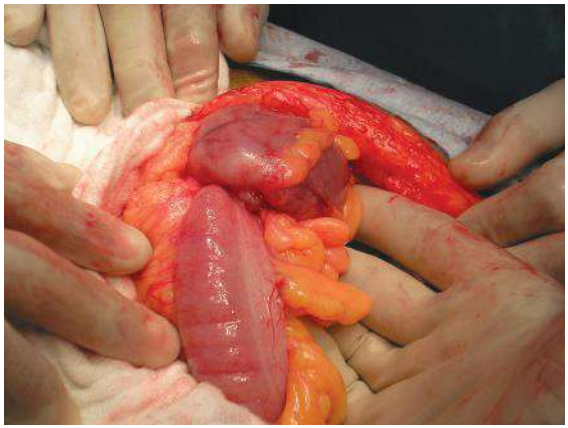

FIG. 21.41A: Colocolic intussusception

FIG. 21.41B: Malignant growth of colon cause for intussusception

Complications: When the mesentery is drawn between the loops, it may result in vascular compromise, which may lead to strangulation, gangrene and perforation.

Symptoms

- In children, there may a history of preceding gastroenteritis following a change in diet (weaning from milk to solid food)

- Severe acute colicky pain, with abdominal distension
- Passing of frequent semisolid stools with bright red blood may be predominant (red currant jelly).

Signs

- During the attacks of pain, a sausage-shaped mass may be felt, which appears during the time of colic and disappears after the colic disappears. The right iliac fossa is empty – Sign de Dance
- Rectal examination may reveal bloodstain on the examining finger (red-currant jelly)
- Colo-rectal intussusception may be felt by the examining finger on rectal examination, or it may even present through anus, resembling a rectal prolapse.

Relevant Investigations

- Plain X-ray abdomen – soft tissue shadow in the region of transverse colon with empty distal colon. Multiple air fluid levels may be seen when obstruction predominates
- Barium enema may show a filling defect called pincer-shaped filling defect (caused by the intussusceptum with the intussuscipiens)
- Colonoscopy can identify, colonic intussusceptions (Fig. 21.42A)
- US and CT (Figs 21.42B and C) will reveal the intussuscepting mass (pseudokidney appearance).

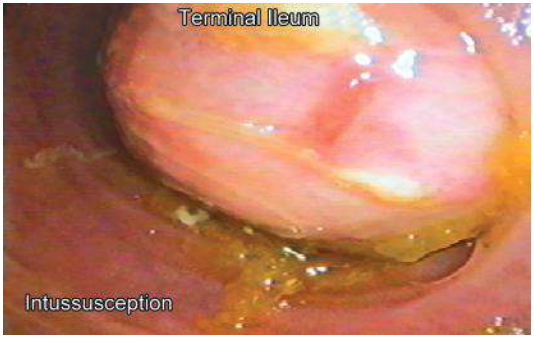

FIG. 21.42A: Intussusception of small bowel (*Courtesy*: Dr Mani Veeraraghavan)

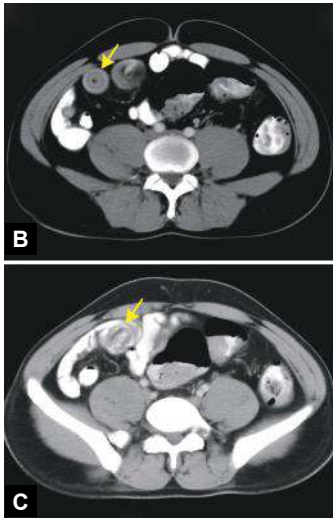

FIGS 21.42B and C: CT—Intussusception

Treatment

- Barium enema and colonoscopy, by themselves may reduce the colonic intussusception
- Laparotomy is required to reduce the small bowel intussusception, and treat the cause appropriately
- Bowel resections may be needed if the bowel segment is strangulated, and nonviable
- Perforation and peritonitis need appropriate treatment.

SWALLOWED FOREIGN BODIES

Incidence and Etiology

- Common in children and psychotic individuals
- Most foreign bodies pass through GIT without any difficulty.

Symptoms

- Abdominal pain may be present
- Vomiting may supervene.

Signs

- Signs of intestinal obstruction may appear
- Chest signs may appear.

Examine the oral cavity, oropharynx or nasopharynx for impacted foreign body, before asking for radiographs.

Relevant Investigations

- Plain X-ray of throat (Fig. 21.43), chest and abdomen may identify the foreign body
- Bronchoscopy/Gastroscopy may be required.

- By the time the patient is seen by the clinician, the foreign body might be expelled through feces, if it is small and blunt
- Since button cells have the tendency to erode through the bowel wall, it is better to monitor its position by serial radiography, and removed if stuck in a place.

Treatment

- Blunt objects reaching the stomach will pass without difficulty
- Sharp foreign bodies impacted in the GIT should be removed (endoscopically or by open surgery)
- Catastrophic bleed, obstruction and perforation peritonitis need laparotomy.

FIG. 21.43: Foreign body larynx (*Courtesy:* Bharat scans)

PERFORATED BOWEL PATHOLOGIES

There are a variety of intra-abdominal pathologies, which are serious emergencies, and also threaten life, if not recognized early and treated. Such conditions are called "acute abdominal catastrophes". Most of these conditions are a result of infection of peritoneum called 'peritonitis', caused by the leak of infected intestinal contents into the peritoneal cavity. Since the peritoneum is a semi-permeable membrane, it allows the spread of infection systemically into the blood to result in septicemia, a fatal condition.

Peritonitis consists of three stages. They are:

1. *Stage of peritonism:* Occurs due to the chemical irritation of the peritoneum caused by the leaked contents of the bowel, and the patient feels sudden excruciating pain. Constitutional symptoms are not prominent, though tenderness and guarding can be elicited
2. *Stage of reaction:* The contents of the bowel are neutralized by the peritoneal exudates, and pain becomes less. Muscle guarding may be present, and shifting dullness may be elicited indicating the presence of free fluid in the abdomen
3. *Stage of spreading peritonitis:* This is an ominous stage, as there are signs of severe infection. They exhibit classic facies, called "Facies Hippocratica" – sunken eyes, hollow cheeks and anxious face. Severe tachycardia (fast thready pulse) is present. Abdomen shows board like rigidity, indicating generalized peritonitis. Death becomes imminent if not attended to, surgically.

The signs and symptoms of perforations caused by various parts of the bowel have certain indicative findings. They are given in the Table 21.10.

Table 21.10: Signs and symptoms of perforations caused by various diseases of the bowel

Signs and symptoms	*Perforation of*						
	*Duodenal or gastric ulcer (benign/ malignant)**	*Appendicitis***	*Cholecystitis*	*Typhoid/ Tubercular ulcer*	*Ulcerative colitis*	*Diverticular disease*	*Colonic malignancy*
Previous history	Frequent use of antacids	Nil or pain in the right iliac fossa in the past (treated or untreated)	Gallstones or chronic cholecystitis	Prolonged fever (high grade in typhoid and low grade in tuberculosis)	Repeated attacks of diarrhea and abdominal pain	Repeated attacks of constipation and abdominal pain	Constipation
History of drug intake	NSAIDs	Nil	Nil	Treatment for typhoid fever	Medical management	Medical management	Laxatives
Area of abdominal pain	Upper abdominal	Umbilical to start and then in the right iliac fossa	Right hypo-chondrial	Umbilical or lower abdominal	Flanks and hypogastric	Umbilical or flanks	Flank on the side of lesion
Gastro-intestinal bleeds	Upper	Nil	Nil	Lower	Lower	Lower	Lower
Treatment	Surgical						

* Leaks of gastric contents due to perforated gastric ulcer (posteriorly placed), into the lesser sac may mask abdominal symptoms
** Clinical presentation of perforated Meckel's diverticulitis is the same as that of perforated acute appendicitis.

Differential Diagnosis

- Myocardial infarction
- Diaphragmatic irritation caused by lower lobar lung lesions
- Acute pancreatitis
- Ruptured or dissecting aortic aneurysm.

Relevant Investigations

- Plain X-rays of abdomen in the erect posture is confirmatory (gas under the diaphragm) in most cases (Figs 21.44 and 21.45) and peritonitis gives the classic 'ground glass appearance' (Fig. 21.46)
- CT (Fig. 21.47) is useful.

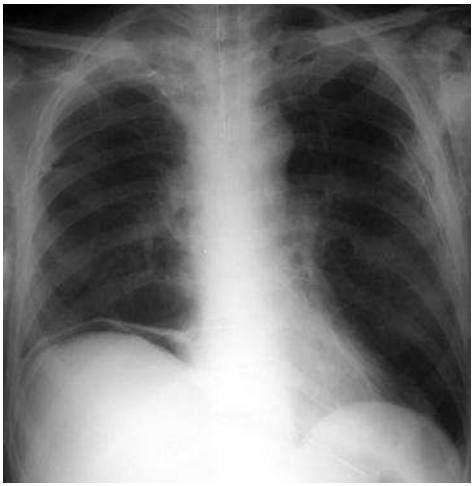

FIG. 21.44: X-ray—Air under the right dome of diaphragm-duodenal ulcer perforation

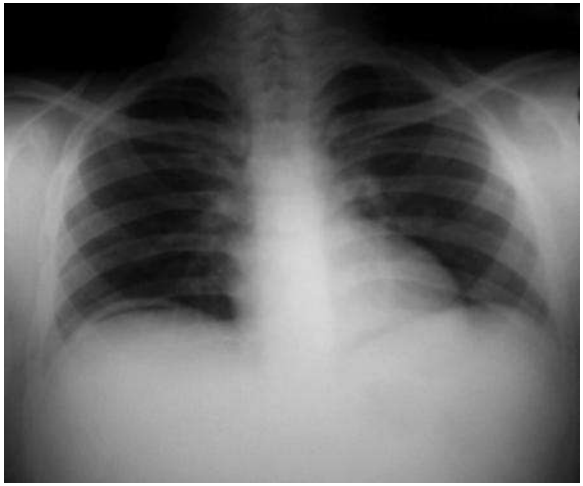

FIG. 21.45: Minimal air under right dome of diaphragm - appendicular perforation

FIG. 21.46: X-ray abdomen - Ground glass appearance of peritonitis

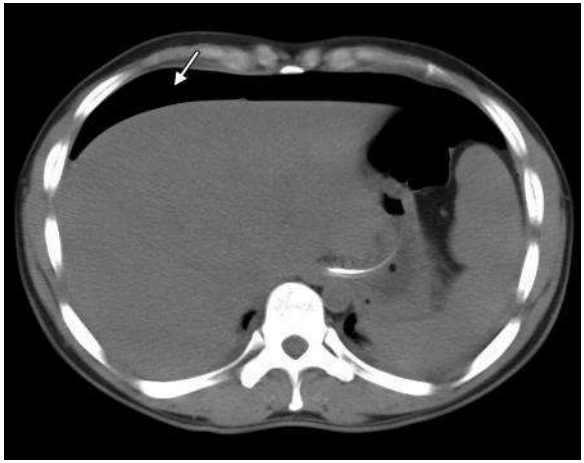

FIG. 21.47: CT—Pneumoperitoneum

Treatment

Treatment is always surgical:

- Closure of bowel perforation if it is simple and of benign cause
- Cholecystectomy (gallbladder perforations)/ appendicectomy (appendicular perforations)/appropriate bowel resections (benign or malignant bowel perforations)
- Peritoneal toileting is mandatory and should be complete.

INTESTINAL STRICTURES

Incidence and Etiology

- Obstructions can be caused by strictures due to:
  - Tuberculosis (healing lesions)
  - Malignant lesions
- Postoperative strictures are not uncommon.

Symptoms

- Symptoms of subacute or acute intestinal obstruction
- History of weight loss, low-grade pyrexia, anemia and vague abdominal pain may be present.

Signs

- Clinical examination may show a mass in the right iliac fossa (differential diagnosis—Crohn's disease)
- Ascites may be present.

Relevant Investigations

- Plain X-ray abdomen (erect) shows air fluid levels
- X-ray chest may reveal a primary tubercular lesion.

Treatment

- Laparotomy is needed for acute obstructions
- Stricturoplasty or bypass procedures or resections are done for tubercular strictures
- Radical resections or bypass procedures are done for malignant strictures.

BANDS AND ADHESIONS

Incidence and Etiology

- Bands may be congenital, but adhesions are post surgical
- The adhesions may be like a small band or generalized, flimsy or dense, single (Figs 21.48 and 21.49) or multiple.

Symptoms and Signs

Features of subacute or acute intestinal obstruction, months or years after surgery.

Relevant Investigations

- Plain radiographs of abdomen are useful in diagnosis
- Diagnostic laparoscopy is conclusive.

Treatment

- Release of adhesions by laparoscopy or by open surgery
- Open surgery and release of adherent bowel loops for dense adhesions obstructing the bowel
- It is better to manage the acute obstructions by trial medical management with gastric suction and intravenous fluids

  Note: Surgery has the disadvantage of recurrence, especially in generalized adhesions.

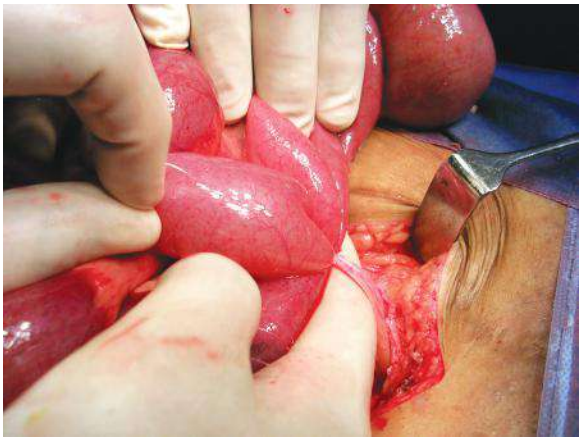

FIG. 21.48: Band in the pelvis causing internal rotation of small bowel

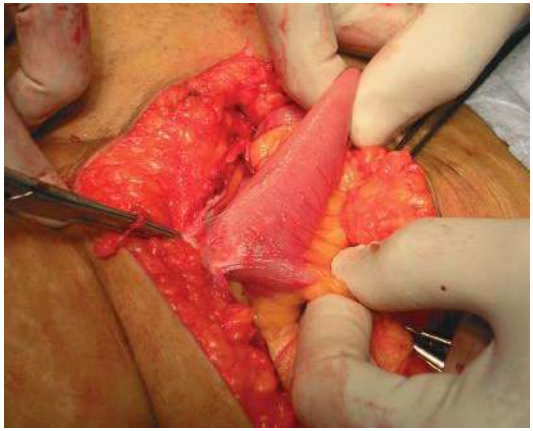

FIG. 21.49: Adhesions of small bowel to previous surgery scar

ENTEROLITHS/FOOD BOLUS

Incidence and Etiology

Obstructions may be caused by enteroliths or food bolus, resulting from poor chewing in an edentulous patient, high consumption of high fiber (e.g. orange pith), usually at a pre-existing narrowing due to tuberculosis, Crohn's disease and surgery.

Symptoms and Signs

Symptoms of subacute or acute intestinal obstruction.

Relevant Investigations

Plain X-ray abdomen (erect) shows air fluid levels. Diagnosis is more clinical and is difficult.

Treatment

Surgery is indicated for acute obstructions, and the bolus or enterolith can be milked into the large intestine and rarely it is necessary to open the bowel to remove it.

VOLVULUS

- Volvulus is defined as a twist of the bowel around its mesenteric axis
- It is more common in the large bowel (commonly the sigmoid colon) than in the small bowel
- Rotation of more than 180 degrees may result in strangulation.

SIGMOID VOLVULUS

Incidence and Etiology

Disease of the middle aged and elderly.

Pathogenesis

Rotation of the sigmoid around its axis occurs when its mesentery is unusually long.

Symptoms

- Sudden severe pain, frequently when straining to pass stool. The patient retches and develops hiccoughs
- The patient may give history of attacks of abdominal pain with constipation, relieved by passing watery stools and large volumes of flatus.

FIG. 21.50: X-ray – Sigmoid volvulus

Signs

- Abdomen rapidly distends, disproportionate to the duration of pain, and the distension is confined more to the left flank
- Rectum is empty on examination.

Relevant Investigations

Plain radiograph of the abdomen will reveal a distended sigmoid, coffee bean appearance – Freeman Dahl sign (Fig. 21.50) (Convergence of three white lines towards the base of the pedicle).

Treatment

- Untwisting of the volvulus, and fixing the colon to parietal peritoneum to prevent recurrence
- Sigmoidectomy is the treatment of choice in long redundant sigmoid colon.

CECAL VOLVULUS

Incidence and Etiology

- This occurs in those whose entire right colon has a mesentery continuous with that of the small bowel, and the cecum does not lie in the right iliac fossa
- More common in women and during pregnancy.

Clinical Presentation

- A distended, tense palpable resonant mass in the umbilical region, with an empty right iliac fossa
- Features of distal small bowel obstruction.

Relevant Investigations

Plain radiograph of the abdomen is diagnostic (Cecal bubble is seen).

Treatment

Untwisting of the volvulus with or without resection of the segment of the bowel and cecopexy is required.

MIDGUT VOLVULUS

Incidence and Etiology

- In children, this occurs due to malrotation and failure of fixation of midgut, usually before one year, rarely in neonates
- In adults, a loop of bowel rotates around a point of adhesion (to the abdominal wall or to an adjacent viscera).

Symptom

Severe abdominal pain.

Signs

Signs of intestinal obstruction.

Relevant Investigation

Contrast study shows 'Cork-screw effect'.

Treatment

Emergency laparotomy and correction of the defect with fixing of the bowel.

FIG. 21.53: Step-ladder pattern of paralytic ileus

Treatment

- Correction of electrolyte imbalances
- Treatment of retroperitoneal or intraperitoneal causes like hemorrhage and sepsis.

TORSION OF MESENTERIC CYST

Incidence and Etiology

- Cystic lesions of mesentery are more common than solid tumors (2:1)
- Cystic tumors of mesentery are:
  - Chylous cyst, serous cyst (developmental) – more common
  - Lymphangioma (lymphatic tissue) – more common
  - Traumatic cyst (trauma)
  - Enteric cyst, dermoid (embryonic rests)
- Majority of cysts are benign (except lymphangiosarcoma, malignant teratoma).

Pathogenesis

Torsion occurs more by its weight, and mobile nature of mesentery encourages this torsion.

Symptom

Sudden acute abdominal pain.

Signs

- Central abdominal tenderness with guarding and rigidity
- A cystic lump may be felt in the central abdomen.

Relevant Investigation

US and CT of abdomen are diagnostic.

Treatment

Emergency laparotomy is needed. The cyst should be removed.

TORSION OF OMENTUM

Incidence and Etiology

- More common in 4th to 5th decades
- Equal sex incidence
- Causes of omental torsion are:
  - *Primary:* Always unipolar, cause is unknown
  - *Secondary:* Usually bipolar, associated with adhesions (intra-abdominal inflammation—tuberculosis, postsurgical adhesions, internal and external herniae).

Pathogenesis

- The omentum twists on its long axis to an extent causing vascular obstruction
- May vary from mild vascular obstruction producing edema to complete strangulation leading to infarction and gangrene
- The situations to cause torsion are:
  - Redundant and mobile segment
  - Fixed point around which the segment can twist

- The precipitating factors for torsion are:
  - Anatomic variations (bifid omentum, accessory omentum, large and bulky omentum)
  - Venous redundancy
  - More common in right side (due to increased bulk and mobility).

Symptoms

- Acute abdominal pain usually is localized to right lower quadrant
- Movement increases the pain
- Nausea and vomiting are common.

Signs

- Tenderness, rigidity and guarding of abdomen
- A vague tender mass may be felt in the upper abdomen.

Differential Diagnosis

- Acute cholecystitis
- Acute appendicitis
- Torsion of right ovary.

Relevant Investigation

US and CT of abdomen are diagnostic.

Treatment

- Emergency laparotomy is needed.
  - Twisted omentum needs to be excised with release of adhesions
  - Associated hernia needs repair.

COLICS

Incidence and Etiology

Colic is defined as a sudden squeezing or griping pain lasting for about 3 to 5 minutes with pain free intervals.

- Nausea, vomiting and retching are common accompaniments
- The cause of a colic is partial obstruction of a tubular structure due to varied causes.

Clinical Features (Table 21.12)

Table 21.12: Clinical features of various colics

Clinical feature	*Biliary colic*	*Renal colic*	*Colics* *Ureteric colic*	*Intestinal colic*	*Appendicular colic*
Incidence	Fat, fertile, flatulent, female of fifty	Young age	Younger age	Any age	Any age
Etiology	Gallstones	Renal calculus	Ureteric calculus, papilloma, clot, stricture	Parasitic infestations (younger age), strictures (middle age), malignancy (old age)	Fecoliths, Worms
Nature of pain	Right hypochondrial pain, referred to right scapula or shoulder	Severe pain in the lumbar region radiating towards umbilicus	Severe colicky pain radiating from loin to groin, testis or thigh	Colicky pain in the umbilical region	Colicky pain in the right iliac fossa
Associated symptom	Dyspepsia	Urinary symptoms	Urinary symptoms	Constipation or diarrhea	Repeated attacks of dull pain in right iliac fossa

Symptoms

- Severe griping pain with pain free intervals
- Location of pain is definite (Table 21.13)
- Radiation (Fig. 21.54) may be present:
  - Radiation to right scapula or shoulder is common with biliary colic
  - Radiation to external genitalia and groins (e.g. ureteric colic)
- Vomiting is usually present

Table 21.13: Location of pain related to regions of abdomen

Right hypochondrium Biliary colic Right renal colic		*Left hypochondrium* Left renal colic
Right lumbar region Right renal colic	*Umbilical region* Intestinal colic	*Left lumbar region* Left renal colic
Right iliac fossa Appendicular colic Right ureteric colic	*Hypogastrium* Uterine colic Urethral colic	*Left iliac fossa* Left ureteric colic

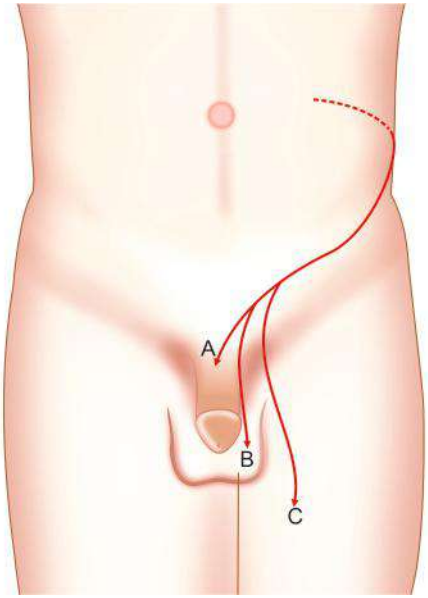

FIG. 21.54: Ureteric colic: (A) Base of penis; (B) Genitalia; (C) Thigh

- Fever may be present if infection is associated
- Jaundice may be present if cholangitis is associated
- Loose stools may be present if colitis is associated
- History of similar episodes in the past
- Family history may present regarding biliary and urinary calculi.

Signs

- Tenderness in the region of pathology
- Mass may be felt proximal to the level obstruction:
  - Right upper quadrant (e.g. distended gallbladder)
  - Right lumbar (e.g. obstructed and enlarged kidney as in PUJ obstruction)
  - Umbilical (e.g. distended bowel)
  - Hypogastric (e.g. distended uterus in dysmenorrhea).

Differential Diagnosis by Clinical History and Examination

- Biliary colic, jaundice, palpable gallbladder (e.g. cholangitis with mucocele, gallbladder choledochal cyst)
- Renal colic, urinary symptoms, palpable kidney (e.g. obstruction at PUJ)
- Ureteric colic, urinary symptoms, tenderness in iliac fossa (e.g. obstruction of ureter)
- Intestinal colic, exaggerated bowel sounds (e.g. obstruction of small bowel)
- Appendicular colic, tenderness right iliac fossa (e.g. obstructive appendicitis)
- Uterine colic, dysmenorrhea, palpable uterus (e.g. congestive dysmenorrhea).

Relevant Investigations

Hematology

- Leukocytosis in infective pathologies (e.g. acute cholecystitis)
- Raised ESR in all infective pathologies.

Radiology

- *Plain X-ray abdomen:* Gas filled loops of bowel (e.g. acute intussusceptions, acute intestinal obstruction)
- *Ultrasonography:* Radiopaque shadows in the abdomen (e.g. renal stones, ureteric stones).

Treatment Plan

- Medical management will suffice in most instances
- Obstructive pathologies may require surgery or removal to relieve the cause of obstruction
- Repeated attacks of colic will require evaluation and management.

GASTROINTESTINAL HEMORRHAGE

Definitions

- Hematemesis—vomiting blood, which may be fresh or partly altered.
- Melena—passing altered blood (meaning digested blood) per rectum.
- Hematochezia—passing blood per rectum bright in color
- Occult bleeding—bleeding not visible to naked eye, but microscopic.

Types of Gastrointestinal Hemorrhages

The blood loss due to gastrointestinal hemorrhage is divided into three types:

1. Mild—less than 500 ml.
2. Moderate—500 to 1500 ml.
3. Severe—more than 1500 ml.

- Gastrointestinal hemorrhages, when they are in large quantities, called major hemorrhages present either as vomiting of blood or passage of blood per rectum, but usually they coexist
- Many times, melena is the sole clinical presentation of upper gastrointestinal bleed.

While evaluating a GI hemorrhage, the following pathologies are to be kept in mind (Fig. 21.55)

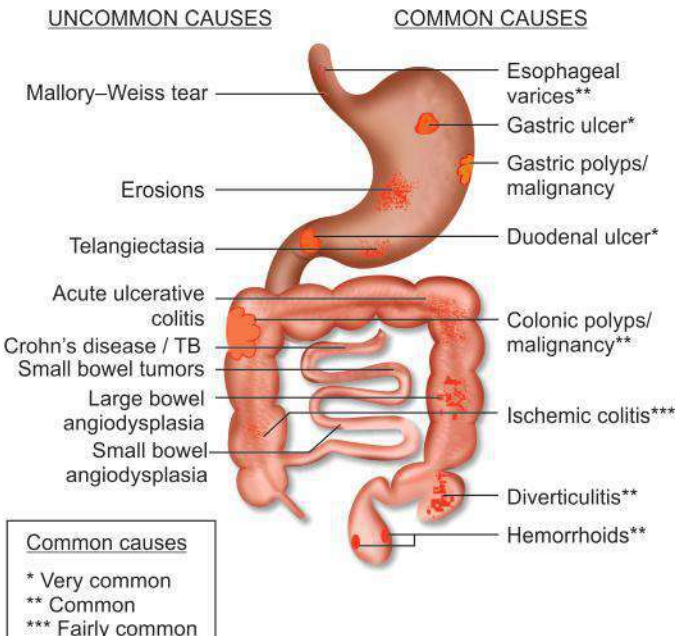

FIG. 21.55: Causes of gastrointestinal bleeding

Hematemesis and Melena

Esophagus

- Reflux esophagitis
- Esophageal varices
- Mallory–Weiss tears
- Corrosive poisoning.

Stomach and Duodenum

- Acid peptic disease
- Gastric polyps
- Gastric lymphoma
- Carcinoma stomach
- Corrosive poisoning.

Melena

Small Bowel

- Angiodysplasia
- Diverticulitis
- Radiation enteritis
- Infections and inflammations
- Ischemic disease
- Intussusception
- Richter's hernia
- Benign tumors
- Malignant tumors.

Hematochezia

Large Bowel

- Angiodysplasia
- Diverticulitis
- Radiation colitis
- Infections and inflammations

- Ischemic disease
- Inflammatory bowel disease
- Benign polyps
- Malignant tumors.

Rectum and Anus

- Polyps
- Malignant tumors
- Hemorrhoids, fissures.

Lower gastrointestinal hemorrhages do not present with hematemesis.

Eliciting History

Hematemesis

1. *Nature of bleed*
   - Frank blood (e.g. esophageal varices)
   - Altered blood (e.g. peptic ulcer, gastric malignancy).
2. *Duration of hematemesis*
   - Short duration (e.g. acute hyperacidity)
   - Recurrent attacks (e.g. chronic duodenal ulcer, gastric malignancy).
3. *Association of rectal bleed*
   - Melena (e.g. upper GI bleed)
   - Hematochezia (e.g. lower GI bleed).
4. *Association of abdominal pain*
   - Absence of abdominal pain (e.g. esophageal varices)
   - Presence of abdominal pain (e.g. chronic duodenal ulcer, gastric malignancy).
5. *Association of fever:* Fever – usually low grade (e.g. GI malignancy).
6. *Association of jaundice*: Jaundice may be present (e.g. periampullary carcinoma, hepatic failure).

Hematochezia

1. *Duration of bleed*
   - Short duration (e.g. hemorrhoids, inflammatory bowel disease, colonic malignancy)
   - Recurrent attacks and long duration (e.g. hemorrhoids, inflammatory bowel disease).

2. *Association of abdominal pain*
   - Absence of abdominal pain (e.g. hemorrhoids, malignancy)
   - Presence of abdominal pain (e.g. inflammatory bowel disease).
3. *Association of fever*: Fever (e.g. inflammatory bowel disease)
4. *Association of jaundice*
   Jaundice may be present (e.g. portal hypertension, metastatic colonic malignancy)
5. *Association of bowel disturbances*
   - Constipation (e.g. colorectal malignancies)
   - Diarrhea (e.g. inflammatory bowel disease).

Past History

- History of pain (e.g. chronic duodenal ulcer, diverticulitis)
- History of bleeds (e.g. chronic duodenal ulcer, diverticulitis, hemorrhoids)
- Previous surgery (e.g. recurrent malignancy of GIT).

Family History

- Familial polyposis
- Colonic malignancy.

Clinical Examination

General

- Vital signs (e.g. shock)
- Breath for fetor (e.g. hepatic failure)
- Conjunctiva for anemia (e.g. GI malignancy), jaundice (e.g. hepatic failure)
- Tongue for anemia (e.g. GI malignancy)
- Neck for lymphadenopathy (e.g. GI malignancy)
- Hands for clubbing, palmar erythema, liver flap, etc. (e.g. liver failure).

- Mild hemorrhages are rarely associated with systemic signs. When the blood loss approaches about 40 percent of the blood volume, shock ensues
- The volume of blood loss either by vomiting or through rectum, is not a very reliable measure, as large amounts stay in the bowel.

Abdomen

- **Inspection**

  Distension
  - Generalized (e.g. ascites of liver failure, malignant ascites)
  - Right upper abdominal (e.g. hepatomegaly)
  - Epigastric (e.g. left lobar hepatomegaly, carcinoma stomach)
  - Left upper abdominal (e.g. splenomegaly).

  Scars, swellings and sinuses

- **Palpation**

  Tenderness
  - Right upper quadrant (e.g. hepatic failure)
  - Epigastric (e.g. chronic duodenal ulcer)
  - Left upper quadrant (e.g. chronic duodenal ulcer).

  Lump
  - Right upper quadrant (e.g. hepatomegaly, distended gallbladder)
  - Epigastric (e.g. carcinoma stomach, left lobar hepatomegaly)
  - Left upper quadrant (e.g. carcinoma stomach, splenomegaly).
- **Percussion**

  Percuss the liver for
  - Its enlargement (e.g. chronic hepatitis, metastatic liver).
- **Auscultation**
  - Bowel sounds are usually normal
  - Exaggerated bowel sounds may indicate obstruction of small bowel (e.g. intestinal colic).

Differential Diagnosis by Clinical History and Examination

- Hematemesis and melena (e.g. pathologies of upper GI tract proximal to ligament of Treitz)
  - With upper abdominal pain, epigastric tenderness (e.g. bleeding gastric/duodenal ulcer)
  - Without abdominal pain and
    - No clinical findings (e.g. esophageal varices, vascular malformations, upper GI malignancy)
    - Abdominal lump (e.g. upper GI malignancy).
  - Without abdominal pain and
    - Increasing jaundice, abdominal lump, + / – hepatomegaly (e.g. upper GI malignancy + / – liver metastases)
    - Fluctuating jaundice, palpable GB, + / – hepatomegaly (e.g. periampullary carcinoma + / – liver metastases).

- Melena (e.g. pathologies of lower GI tract distal to ligament of Treitz)
  - With abdominal pain
    - Epigastric tenderness (e.g. bleeding duodenal/gastric ulcer)
    - Umbilical tenderness (e.g. inflammatory bowel disease)
    - Flank tenderness (e.g. ulcerative colitis)
    - Palpable lump (e.g. proximal colonic malignancy).
  - Without abdominal pain
    - No clinical findings (e.g. esophageal varices, vascular malformations)
    - Abdominal lump (e.g. upper GI malignancy).
- Hematochezia (e.g. pathologies of lower GI tract in the colorectum)
  - With pain during defecation, positive inspection findings (e.g. acute fissure in ano)
  - Without perianal pain
    - Positive proctoscopy findings (e.g. hemorrhoids, rectal pathologies)
    - Negative proctoscopy findings (e.g. colonic pathologies)
  - With abdominal pain + / – fever, local tenderness (e.g. inflammatory bowel disease, diverticulitis)
  - With colicky pain, abdominal lump (e.g. obstructing colonic malignancy)
  - With jaundice + / – abdominal pain, + /– abdominal lump, + / – hepatomegaly (e.g. colonic malignancy with liver metastases).

Whatever be the external visible loss, signs of hypovolemia should be watched for.

Relevant Investigations

Hematology

- Hemoglobin and PCV for anemia
- Total and differential leukocyte count (e.g. tuberculosis, ulcerative colitis)
- Thrombocytopenia (e.g. hypersplenism)
- ESR may be raised in infections and malignancies
- Coagulation profile (e.g. bleeding disorders).

Coagulation profile should be assessed in all cases of GI bleed, as bleed by itself the sole presentation of coagulation disorders.

Liver Function Tests

- Serum bilirubin (e.g. raised levels indicate jaundice)
- Serum transaminases (e.g. raised levels indicate liver cell disease)
- Serum alkaline phosphatase (e.g. raised levels may indicate biliary obstruction due to hepatic metastases).

- Serum proteins (e.g. reduced levels indicate undernutrition)
- Prothrombin time is prolonged in liver diseases and needs to be corrected by administration of vitamin K, in the management of gastrointestinal bleeds.

Stool Examination

Examination of stool for occult blood is required in cases of occult bleeding from GIT.

> Occult bleeding from GIT is common, but the bleed should be atleast 10 ml to identify by examination.

Radiology

- Chest X-ray (e.g. aspiration pneumonitis, mediastinal widening and hilar lymphadenopathy of esophageal malignancy)
- Contrast studies of bowel
  - ❖ Barium swallow (e.g. esophageal malignancy)
  - ❖ Barium meal (e.g. gastric malignancy, periampullary carcinoma)
  - ❖ Barium enema (e.g. colonic malignancy)
- CT/MRI scan
  - ❖ Chest (e.g. esophageal malignancy, and paraesophageal pathology like lymph nodes, pulmonary secondaries)
  - ❖ Abdomen (e.g. cirrhosis of liver, malignant deposits of liver, lymph node metastases of GI malignancies, intestinal tuberculosis)
- Ultrasonography of abdomen (e.g. malignant deposits, dilated portal vein and biliary radicals)
- Magnetic resonance cholangiopancreatography (MRCP) (e.g. periampullary carcinoma).

Endoscopy

- Upper gastrointestinal endoscopy (e.g. gastroesophageal reflux disease, esophageal tears, gastric malignancies, periampullary carcinoma)
- Lower gastrointestinal endoscopy (e.g. polyps, tumors of rectum and colon)

> - Endoscopy has the advantage of obtaining tissue for histopathology
> - Endoscopy may be used as a therapeutic tool at the same sitting (e.g. sclerotherapy in bleeding esophageal varices and endoclipping of bleeding vessel)
> - Endoscopy should be performed at the earliest opportunity and after adequate resuscitation.

Radioisotope Studies

Radioisotope scanning using the patient's own labeled red blood cells can be useful in small bowel bleeds, especially those from angiodysplasia.

Selective Arteriography

Selective arteriography (superior mesenteric arteriography) may be useful in determining the small bowel bleeds, particularly those from angiodysplasia.

Capsule Endoscopy

Swallowing a small capsule with a video camera, and recording the images of the lumen of small bowel gives tremendous information of bleeding from small bowel. Histopathology cannot be obtained by capsule endoscopy.

Histopathology

Biopsy through endoscopy (upper and lower) is confirmative.

Treatment

Medical Management

- Large bore vascular access and correction of hypovolemia and hematocrit
- Correction of coagulopathy
- Emergency endoscopy
  - For nonvariceal bleeding
    - IV PPI (80 mg 6 hrly)
    - Nasogastric suction and gastric lavage
  - For variceal bleeding
    - Insertion of Sengstaken-Blakemore tube for compression of varices
    - Administration of vasopressin (upto 20 units SC or slow IV).
    - Variceal injection or banding
    - Transhepatic intravenous portosystemic shunt (TIPSS) for uncontrolled bleeding.

Surgery is needed for persistent uncontrollable GI bleeding.

Patient requiring more than six units of blood for resuscitation for acute bleed, will require surgical management.

Surgical Treatment

Surgery is indicated in recurrent esophageal variceal bleeding, complicated inflammatory bowel disease, polyps, benign and malignant tumors of GIT and hemorrhoids.

- Rapid changes occur in the tissues distal to the occlusion and the limb survival depends on the presence of collaterals
- Skeletal muscle and nerve tissue are the most sensitive to hypoxia and suffer damage early.

Complications: Compartment syndrome, acute renal failure (acidosis), muscle contracture (Volkmann's ischemic contracture).

Symptoms

- Acute ischemia produces classical symptoms (6 Ps)
  - *Pain:* Usually excruciating, sudden in onset and continuous
  - *Pallor:* Affected limb appears pale due to lack of blood
  - *Pulselessness:* Pulses distal to the block are absent
  - *Paresthesia:* Due to nerve damage
  - *Paralysis:* Sensory motor deficit due to ischemic nerve damage secondary to severe ischemia (a late sign)
  - *Poikilothermia:* The affected limb appears cold.

Signs

- Affected muscle gets swollen, tense and is very tender, due to persistent ischemia
- On palpation, muscle gives a characteristic rubbery feel (can be mistaken for DVT)
- Later skin changes like mottling (Fig. 23.1) and blisters start and ultimately the limb becomes gangrenous (Fig. 23.2).

FIG. 23.1: Mottling of ischemic limb (*Courtesy*: Dr N Sekar)

FIG. 23.2: Gangrene of acute ischemia of left lower limb (*Courtesy*: Dr N Sekar)

Relevant Investigations

All investigations should be done without delay:

- Duplex scan is done to confirm diagnosis in early cases
- Angiogram for treatment planning
- Tests for thrombophilia
- Hyperkalemia when renal shut down ensues.

Treatment

- Acute arterial occlusion is a surgical emergency. Usually irreversible damage occurs by 4 to 6 hours. Hence, revascularization should be done within 6 to 8 hours for limb salvage (Golden hour).
  - Intravenous heparin should be administered as soon as the diagnosis of acute ischemia of the limb is made. Best results are seen in those who undergo revascularization within 6 to 8 hours after the onset of ischemia
  - Embolism is best managed by embolectomy done through the brachial artery at the elbow or the femoral artery at the groin
  - Acute arterial thrombosis can be managed by surgical bypass or by catheter directed thrombolytic therapy. Frequently both modalities may be required to achieve revascularization
  - Fasciotomy will be required to avoid muscle damage in compartment syndrome caused by delayed revascularization
  - Early amputation for limbs with irreversible ischemia.

Note: Those who present late should be taken up for surgery without imaging and intraoperative angiogram should be done to reduce time delay.

ACUTE INTESTINAL ISCHEMIA

Incidence and Etiology

Intestinal ischemia occurs usually due to occlusion of mesenteric artery caused by:

- Thrombosis on an atheromatous plaque (e.g. origin of superior mesenteric artery)
- Embolus following atrial fibrillation/myocardial infarction/detached atheromatous plaque.

Clinical Presentation

- Severe acute abdominal pain with copious vomiting
- Very fast deterioration of health

- Hematemesis and melena also occur in some cases
- The clinical signs are disproportionate to the symptoms, and rarely an area of tenderness may be felt near the infarcted bowel.

Relevant Investigations

- ECG, plain X-ray and US abdomen are useful
- Selective angiography is informative.

> • Systemic inflammatory response, oliguria, persistent acidosis, raised serum amylase and bloody diarrhea may all occur but none is specific.

Treatment

- Conservative management to stabilize the patient
- Laparotomy and resection of nonviable bowel will be necessary, if the patient is stable
- Balloon angioplasty or bypass grafts may be feasible in select cases.

Note: Outcome is guarded in most cases.

- Intestinal ischemia should be considered as a possible cause of deterioration, particularly in the elderly with pre-existing vascular disease or vasculitis.

LEAKING OR DISSECTING AORTIC ANEURYSM

Incidence and Etiology

- Aneurysm of aorta (> 4 cm diameter – normal 1.5 to 2.5 cm) is the disease of elderly, and rupture is the deadly complication, with a mortality of more than 80 percent
- The rupture can occur into the retroperitoneum or into the peritoneal cavity.

Clinical Presentation

- Acute severe upper abdominal pain or in the lower chest
- Radiation to the back mimics acute pancreatitis
- Aneurysm may be felt per abdomen with guarding and rigidity
- Lower limb pulses are feeble or absent
- Dissecting aneurysms present with retrosternal pain, as the pathology starts at the aortic arch.

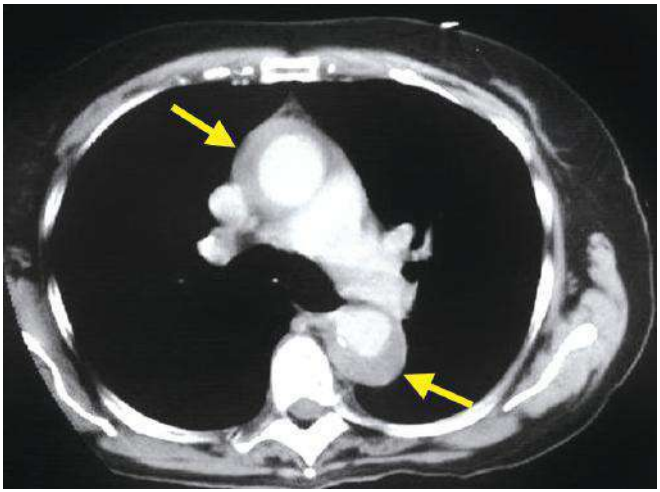

FIG. 23.3: CT chest—Dissecting aneurysm of thoracic aorta

Relevant Investigation

US and CT of the abdomen are conclusive (Fig. 23.3).

Treatment

Emergency surgical intervention is necessary, along with correction of shock.

- Only a third with rupture of aneurysm live to reach the hospital.

24

Urology

ACUTE RETENTION OF URINE

Definition

Defined as inability to micturate even when the bladder is full.

Incidence and Etiology

- This occurs when the sphincter is unable to relax or with proximal urethral obstruction like in prostatic enlargement
- Acute retention is of short duration and the bladder distension is painful
  The causes of acute retention of urine are given in Table 24.1.

Symptoms

- Severe lower abdominal pain
- Inability to void urine
- Symptoms relating to underlying cause (e.g. LUTS).

Signs

- Palpable distended tender urinary bladder
- Rectal examination: Anal sphincter spasm, enlarged prostate.

Table 24.1: Causes of acute retention of urine

Males	*Females*	*Both sexes*
Prostatic enlargement	Multiple sclerosis	Clot retention
Urethral stricture	Urethral stenosis	Calculus at bladder neck or urethra
Tight Phimosis / Meatal stenosis	Retroverted gravid uterus	Rupture of urethra
Acute urethritis	Cervical fibroid	Spinal anesthesia
Postoperative	Bladder neck hypertrophy	Neurogenic
		Fecal impaction
		Anal pain (post hemorrhoidectomy)
		Drugs (antihistamines, antihypertensives, anticholinergics, antidepressants)

Relevant Investigations

- Hemogram
- Urea/creatinine
- PSA levels may be high
- US may confirm the cause of obstruction.

Treatment

- Immediate urethral catheterization
- Urine examination for sugar and infective organisms
- Treatment of the cause.

HEMATURIA

Definition

Passing of blood in the urine is called hematuria.

- Frank hematuria: Presence of frank blood or blood clots in the urine
- Microscopic hematuria: Presence of blood discovered by microscopy or dipstick.

Incidence and Etiology

- Whatever the cause, hematuria is usually episodic:
  - Hematuria (painless) (e.g. tumors of urological system)
  - Hematuria with dysuria (e.g. urological infections)
  - The causes of hematuria are given in Table 24.2.

Table 24.2: Causes of hematuria

Etiology	*Diseases*				
	Kidneys	*Ureters*	*Bladder*	*Prostate*	*Urethra*
Congenital	Polycystic kidney				
	AV malformation				
Traumatic	Trauma		Trauma		Trauma
Inflammatory	Glomerulonephritis		Cystitis	Prostatitis	Acute urethritis
	Tuberculosis		Schistosomiasis		
Neoplastic	Malignancy	Urothelial tumors	Malignancy	Malignancy	Malignancy
Others	Calculi	Calculi	Calculi	Benign enlargement	Calculi
Nonurological	Coagulation and hematological disorders				

Note: Microscopic hematuria may represent a significant lesion in the urinary tract and should be taken seriously, though in about 5 percent of the cases, no cause can be found.

Symptoms

- Passing dark colored/blood stained urine
- Pain may or may not be present.

Signs

- Clinical signs may vary (palpable kidney in large tumors, palpable bladder in bladder obstruction)
- Anemia (in chronic hematuria or large hematuria)
- Enlarged prostate.

Relevant Investigations

- Hematocrit may be low
- Renal function may be impaired
- Coagulation profile may be altered
- Urine examination for infection (routine and C/S)
- Ultrasonography
- Intravenous urography (IVU)
- Cystoscopy
- CT/MRI may be needed while evaluating malignancies.

Treatment

- Primary treatment requires immediate urethral catheterization
- When catheterization is not possible, suprapubic cystostomy is needed
- Definitive treatment depends on the underlying cause.

RENAL COLIC

Caused by distension of renal capsule and pelvis or stretch of the peritoneum by large renal swellings:

- Nature of pain
  - ❖ Pain localized to renal angle (posterior space between the 12th rib and the sacrospinalis muscle) (Fig. 24.1) (e.g. acute pyelonephritis, renal tuberculosis)—described by putting the hand on the waist with his fingers on the renal angle and thumb pointing the umbilicus (Fig. 24.2)
  - ❖ Pain radiating to umbilicus (e.g. renal tuberculosis)
- Intensity of pain
  - ❖ Severe (e.g. acute pyelonephritis renal stones)
  - ❖ Dull and persistent (e.g. polycystic disease and malignancy)
  - ❖ Acute obstruction of the upper ureter can give a severe colicky pain in the loin called "Renal colic".

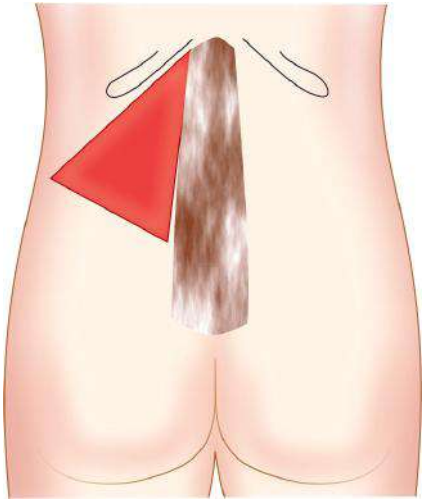

FIG. 24.1: Renal angle

FIG. 24.2: Description of renal pain

Relevant Investigations

- Hematocrit may be low
- Renal function may be impaired
- Urine examination for infection (routine and C/S)
- Ultrasonography will show stones or proximal dilatation
- IVU will show stones, or filling defects (stones, tumors) or proximal dilatations
- Cystoscopy is routine
- CT/MRI may be needed while evaluating malignancies.

Treatment

- Emergency treatment requires antispasmodics to relieve pain, antiemetics for vomiting
- Definitive treatment depends on the underlying cause.

URETERIC COLIC

Caused by obstruction of the ureter commonly by a stone or a clot, and may correspond to the level of obstruction:

- Nature of pain
  - When the stone is lodged in the upper ureter, the pain radiates to the testicle (T11-12)
  - When the stone is lodged in the mid ureter, the pain may be in the McBurney's point on the right side and simulate diverticulitis on the left side (T12-L1)
  - When the stone is lodged in the distal ureter, the pain resembles vesicular pain or may radiate to the genitalia or inner side of thigh (L1-L2)
- Intensity of pain
  - Dull pain in the side of abdomen (e.g. chronic obstruction of ureter with calculi)
  - Severe colicky pain—(loin pain radiating to the groin, genitalia or inner thigh—genitofemoral nerve) in acute obstructions of ureter called "Ureteric colic" (Fig. 24.3).

Symptom

Severe pain as described above.

Relevant Investigations

- Hematocrit may be low
- Renal function may be impaired

27

Gynecology

ACUTE TORSION OF OVARIAN CYST

Incidence and Etiology

It can occur at any age, may follow major physical exertion, which can initiate the torsion.

Pathogenesis

Due to the large size and vulnerability to rotate over a narrow pedicle, torsion occurs.

Symptom

Severe pain in the lower abdomen, with signs of shock.

Signs

- Guarding and rigidity may be present.
- A tender lump may be palpable, whose lower border cannot be made out.

Relevant Investigations

Ultrasonography (US) and Computed tomography (CT) or Magnetic resonance imaging (MRI) of abdomen (Fig. 27.1) are diagnostic.

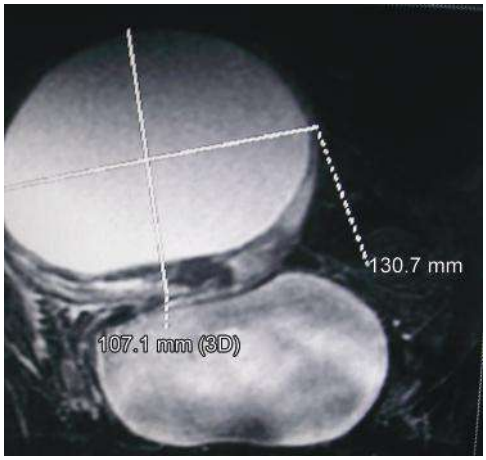

FIG. 27.1: MRI—Torsion of ovarian cyst

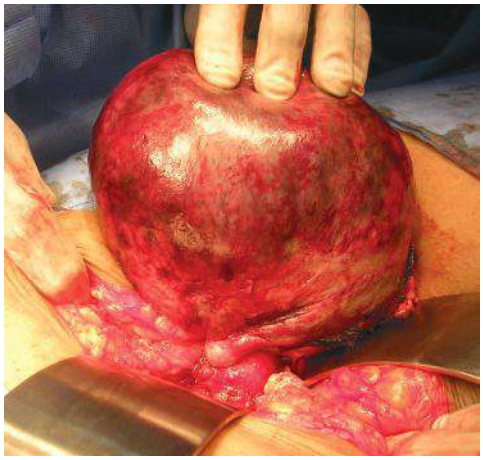

FIG. 27.2: Torsion of ovarian cyst

Treatment

Emergency laparotomy (Fig. 27.2) is needed. The cyst should be removed.

ACUTE SALPINGITIS

Incidence and Etiology

- A disease of the sexually active female, sometimes after sexually transmitted diseases, can confuse with acute appendicitis
- Usually associated with menstrual period, abortion or puerperium
- Common infective agents are *N. gonorrhoeae*, coliforms and chlamydiae.

Pathogenesis

Infective organism gains entry through the vagina to infect the salpinx.

Symptoms

Suprapubic pain, with or without white discharge per vagina (leukorrhea).

Signs

- Localized hypogastric tenderness
- Vaginal examination may reveal tender cervix and a bulky uterus.

Differential Diagnosis

Acute appendicitis is to be thought of in right sided salpingitis.

Relevant Investigations

Ultrasonography (US) is useful in diagnosing mass of the right ovary and the bulky uterus.

Treatment

Treatment with antibiotics.

RUPTURE OF LUTEIN CYST

Incidence and Etiology

A common condition in young females.

Clinical Presentation

- Acute abdominal pain associated with vomiting (right sided pathology may be confused with acute appendicitis)
- On examination:
  - A firm mass may be felt adjacent to the midline or in one of the iliac fossae
  - Guarding and rigidity may be present
- Vaginal examination may reveal a tender mass.

Relevant Investigations

Ultrasonography (US) of the abdomen is diagnostic.

Treatment

Emergency surgical intervention is mandatory.

28

Pediatrics

ACUTE INTUSSUSCEPTION

Incidence and Etiology

- Two per 1000 infants are affected with male preponderance, commonly affecting the age group of 2 months to 2 years
- Commonly, it is secondary to an enlarged Peyer's patch due to viral or bacterial infections (Fig. 21.39)
- The other less common causes are:
  - Meckel's diverticulum
  - Duplication cyst in the bowel wall
  - Polyp.

Pathogenesis

- Intussusception is the invagination of a segment of bowel into the distal adjacent loop (proximal into the distal)
- When the mesentery is drawn between the loops, it may result in vascular compromise, which may lead to strangulation, gangrene and perforation.

Symptoms

- In children, there may a history of preceding gastroenteritis following a change in diet (weaning from milk to solid food)

- Severe acute colicky pain, with abdominal distension
- Passing of frequent semisolid stools with bright red blood may be predominant (red recurrant jelly).

Signs

- During the attacks of pain, a sausage shaped mass may be felt, which appears during the time of colic and disappears after the colic disappears. The right iliac fossa is empty—Sign de Dance
- Rectal examination may reveal bloodstain on the examining finger (red-currant jelly)
- Colorectal intussusception may be felt by the examining finger on rectal examination, or it may even present through anus, resembling a rectal prolapse.

Relevant Investigations

- Plain X-ray abdomen: Soft tissue shadow in the region of transverse colon with empty distal colon. Multiple air fluid levels may be seen when obstruction predominates
- Barium enema may show a filling defect called pincer shaped filling defect (caused by the intussusceptum with the intussuscipient)
- Colonoscopy can identify, colonic intussusceptions
- US and CT (Fig. 28.1) will reveal the intussuscepting mass (pseudokidney appearance).

Treatment

- Resuscitation
- Hydrostatic or pneumatic radiological reduction

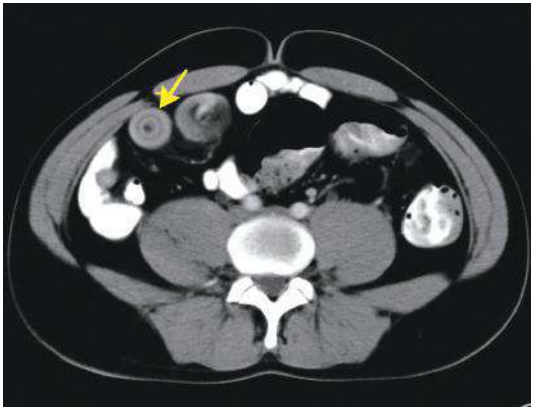

FIG. 28.1: CT—Intussusception

- Barium enema and colonoscopy, by themselves may reduce the colonic intussusception
- Laparotomy is required to reduce the small bowel intussusception, and treat the cause appropriately
- Bowel resections may be needed if the bowel segment is strangulated, and nonviable
- Perforation and peritonitis need appropriate treatment.

CONGENITAL PYLORIC STENOSIS

Incidence and Etiology

Four in 4000, with male predominance, most common in first born child.

Pathogenesis

Congenital inherited condition resulting in hypertrophy of the circular muscle fibers of pylorus.

Symptoms

- Nonbilious projectile vomiting
- Failure to thrive despite hunger.

Signs

- Visible gastric peristalsis
- Palpable mass in abdomen deep to right rectus in the transpyloric plane.

Relevant Investigations

- Serum electrolytes to be checked
- US is diagnostic.

Treatment

- Resuscitation
- Pyloromyotomy when the child is stable.

NECROTIZING ENTEROCOLITIS

Incidence and Etiology

Common in newborns.

Pathogenesis

Thought to be due to ischemia of large bowel wall with translocation of luminal bacteria resulting in systemic sepsis.

Symptoms

- Fever
- Bilious vomiting
- Bloody diarrhea
- Abdominal distension.

Signs

- Abdominal distension
- Absence of bowel sounds (may indicate perforation and peritonitis).

Relevant Investigations

- Leukocytosis
- Abdominal X-ray may show thickened dilated bowel wall containing intramural gas.

Treatment

- Fluid resuscitation
- Broad spectrum antibiotics
- Surgical resection of necrotic bowel may be needed.

TRACHEOESOPHAGEAL FISTULA

Incidence and Etiology

This fistulous communication can occur in newborn infants.

Pathogenesis

It is a developmental abnormality.

FIG. 28.2: Barium swallow—Tracheoesophageal fistula

Symptoms

Cough and fever - due to the entry of swallowed food into the respiratory tract.

Signs

Rales and rhonchi.

Relevant Investigations

Barium swallow (Fig. 28.2) allows the contrast to enter the trachea and bronchi, and is diagnostic.

Treatment

- Surgical closure of a congenital fistula is usually successful
- Malignant fistulae are difficult to handle and death becomes inevitable due to pulmonary infection.

29

Lymphatic System

ACUTE LYMPHANGITIS

Incidence and Etiology

In the tropics, the main causes are:

- Filariasis
- Tuberculosis.

Clinical Presentation

- Swelling of the limbs and genitalia associated with fever and rigor
- On examination, the limbs are swollen and erythematous with weeping eruptions (Fig. 29.1)
- Secondary infections can lead to localized patchy gangrene (Fig. 29.2).

FIG. 29.1: Acute lymphangitis

FIG. 29.2: Gangrenous patches in acute lymphangitis

Relevant Investigations

Peripheral smear for microfilaria.

Treatment

Antifilarial treatment and compression bandages.

ACUTE VIRAL AND BACTERIAL LYMPHADENITIS

Incidence and Etiology

- Caused by viruses (commonly rhinovirus), bacteria
- Bacterial infections are secondary to oral cavity infections
- Infections may result into abscesses
- Common in the neck.

Symptoms

- Painful swelling (commonly in the neck, rarely in the axilla and inguinal region)
- Fever may be present.

Signs

- Single or multiple tender soft to firm swellings with signs of inflammation on the skin (Figs 29.3 and 29.4)
- Examination of the drainage area is essential to assess the primary cause e.g. acute pharyngitis or dental infections causing secondary enlargements of neck glands.

FIG. 29.3: Acute viral lymphadenitis of neck

FIG. 29.4: Acute bacterial lymphadenitis of neck

Relevant Investigations

- No specific investigation is required in short lived cases
- Biopsy may be needed to differentiate persistent swellings from other causes
- Diagnostic aspiration is useful in abscesses.

Treatment

- No treatment is required for short lived cases (viral) as they resolve spontaneously
- Bacterial lymphadenitis requires broad spectrum antibiotics
- Incision and drainage for abscesses.

ACUTE FILARIAL LYMPHANGITIS AND LYMPHADENITIS

Incidence and Etiology

- Caused by thread like, parasitic filarial worms Wuchereria bancrofti or Brugia malayi
- Worms lodge in the lymphatic system, and live for 4 to 6 years producing millions of immature microfilariae that circulate in the blood
- Transmitted by mosquitoes *Culex fatigans* (develop and reach the infective stage in 7 to 21 days and migrate to biting mouth parts)
- Causes severe inflammation of the lymphatics and they settle down in the lymph nodes, commonly in the inguinal group.

Symptoms

- High-grade fever with rigor
- Swelling of the lower limbs and groin swellings.

Signs

- Inguinal lymphadenopathy (Fig. 29.5)
- Presentation with swelling of scrotum (hydrocele) (Fig. 29.6) is common.

Relevant Investigation

Demonstration of microfilaria in the peripheral blood smear (Fig. 29.7) is diagnostic.

FIG. 30.2: Erysipelas

Investigations

Isolation of the organism in culture.

Treatment

Appropriate antibiotics cure the lesion.

FURUNCLE

Incidence and Etiology

- Infection of the hair follicle by *Staphylococcus aureus*
- Furuncle may be a source of systemic sepsis, in diabetics
- Cavernous sinus thrombosis is rare but a serious complication of furuncle on the face above the line drawn from the angle of the mouth to the tragus of the ear.

Symptom

A painful swelling at the hair root (Fig. 30.3).

FIG. 30.3: Furuncle

Signs

- Indurated swelling containing pus in due course
- Draining lymph nodes may be involved
- Cellulitis is seen in immunocompromised individuals and diabetics.

Relevant Investigations

Culture of the pus and identifying the incriminating organism is necessary.

Treatment

- The abscess may burst spontaneously
- Some may need surgical drainage.

CELLULITIS

Incidence and Etiology

- Spreading inflammation of the subcutaneous and fascial tissues
- Commonly due to *Streptococcus pyogenes*
- Diabetics are mostly affected.

Pathogenesis

Organism gains entry through the broken skin due to trauma, and infection spreads along the subcutaneous tissue planes.

FIG. 30.4: Cellulitis foot

Symptoms

- Affected part appears grossly swollen, painful and red (Fig. 30.4)
- Constitutional symptoms like fever and toxemia are common.

Signs

- It is very tender
- The lymphatics may appear inflamed and appear as red streaks and lymph nodes may be enlarged and tender
- Abscesses may form in the subcutaneous plane and skin may undergo avascular necrosis and become gangrenous.

Relevant Investigations

Organisms should be isolated for culture from discharge if any.

Treatment

Appropriate antibiotics are necessary to control infection.

ACUTE PYOGENIC ABSCESS

Incidence and Etiology

- Abscess is defined as collection of pus
- Incomplete resolution of cellulitis and lymphadenitis are the causes
- When this occurs as a part of pyemia, they are multiple in number.

FIG. 30.5: Acute abscess leg

Symptoms

- A painful swelling (Fig. 30.5)
- Constitutional symptoms like fever, rigor and toxemia will be predominant.

Sign

Tender swelling, may show fluctuation, if the pus is liquid.

Relevant Investigations

- No special investigation is required, if it is solitary
- Isolation of organism in blood by culture if suspected to be part of pyemia
- Diagnostic aspiration may be helpful in localizing deep seated abscess.

Treatment

Incision and drainage of pus will be curative under cover of appropriate antibiotics.

CARBUNCLE

Incidence and Etiology

- An infective gangrene of subcutaneous tissue
- Caused by *Staphylococcus aureus*
- Commonly seen in diabetics.

Pathogenesis

- Thermal burns: Heat denatures cellular proteins by coagulation necrosis. The damage is directly related to the intensity of heat and the duration of contact of the incriminating agent. Usually the damage is partial thickness or full thickness
- Electrical burns: Electricity (high and low voltage) causes deep tissue destruction both at the point of entry and at the point of exit. Muscle tissue destruction is an integral part but it cannot be assessed accurately in the initial stages
- Radiation burns: Radiation causes full thickness dermal injury due to the deep penetration of ionizing radiation
- Chemical burns: Chemicals cause cell necrosis and the damage depends on the concentration of the chemical and the duration of contact with the skin. The damage ceases only when the agent is chemically expended, and majority of the times the damage is full thickness.

Clinical Features

The burn wounds are classified into four degrees based on the clinical features (Table 30.2). The clinical photographs are shown in Figures 30.7 to 30.17.

Table 30.2: Classification of burn wounds and their clinical features

Degree of burn	*Layer of skin involved*	*Clinical presentation*			*Tests*		
		Pain	*Color*	*Blisters*	*Touch*	*Pressure*	*Capillary filling*
First degree (superficial)	Epidermis	Severe	Erythema	Absent	Present	Present	Present
Second degree (partial thickness)	Superficial (superficial dermis)	Severe	Erythema/ pallor	Present	Present	Present	Present
	Deep (deep reticular dermis)	Moderate	Pallor	Usually absent	Absent	Present	Present
Third degree burn (full thickness)	Epidermis and dermis	Absent	Pallor	Absent	Absent	Absent	Absent
Fourth degree burns	Skin and subcutaneous structures	Absent	Pallor	Absent	Absent	Absent	Absent

Note: In any patient of burn injury, a combination of degrees occurs. This is important while treating a patient of burns, and repeated assessment and scrutiny of the wounds is necessary.

FIG. 30.7: Superficial burns

FIG. 30.8: Superficial partial thickness burns

FIG. 30.9: Superficial partial thickness burns

FIG. 30.10: Second degree burns with blister formation

FIG. 30.11: Mixed second and third degree burns

FIG. 30.12: Deep partial thickness burns (third degree)

FIG. 30.13: Deep burns of the forearm

FIG. 30.14: Acid burns of the face

FIG. 30.15: Chemical (tar) burns—Thermal contact

FIG. 30.16: Electrical burns

FIG. 30.17: Thermal burns caused by hyperthermic reaction of plaster of Paris

The depth of burns varies depending on the incriminating agent and its duration of contact with the skin (Table 30.3).

Table 30.3: Burns based on its etiology

Agent	*Exposure time*	*Depth of burns*	*Appearance*	*Pain*
Hot liquids	Short exposure	Superficial dermal	Wet, pink, blisters	Severe
	Long exposure	Deep dermal	Wet, red, dark	Minimal
Flame	Flash exposure	Partial thickness	Wet pink blisters	Severe
	Direct contact	Full thickness	Dry, white, waxy or brown, black leathery	Minimal
Chemicals	Direct contact	Full thickness	Light brown to light gray	Severe

Metabolic Effects of Burns

Burn injuries cause severe damage to the body structures and its metabolism. They are shown in Tables 30.4 and 30.5.

Table 30.4: Metabolic disturbances associated with burns (general effects)

Category	*Period*	*Clinical condition*	*Reason*	*Clinical presentation*
General effects	Immediate	Hypovolemic shock	Extracellular accumulation of water and excessive evaporation from burnt area	Tachycardia, tachypnea, hypotension, prerenal uremia, oliguria
		Hemoconcentration	Loss of water and plasma from burnt area	
		Dehydration	Excessive evaporation of water from burnt area	Tachycardia, tachypnea, hypotension, prerenal uremia oliguria
		Electrolyte imbalance	Retention of sodium and excessive excretion of potassium	Altered sensorium, uremia, oliguria
		Asphyxia and cyanosis	Due to burn of respiratory passages	Difficulty in breathing
	Late	Septicemia (multiorgan failure)	Systemic spread of infection	High grade fever, tachycardia, tachypnea, hypotension oliguria, death
		Acute renal failure	Fluid loss and electrolyte imbalance	Altered sensorium, uremia, oliguria or anuria
		Stress ulcers	Reason not known	Hematemesis and melena

Table 30.5: Metabolic disturbances associated with burns (late effects)

Category	*Period*	*Clinical condition*	*Reason*	*Clinical presentation*
Local effects	Immediate	Tissue damage	Direct injury of cells	Extensive wounds
		Edema	Excessive accumulation of fluid in the extracellular spaces	Swollen areas surrounding or below the burnt areas
		Water and electrolyte loss	Excessive loss from the burnt area from large wounds	Large dry wounds
		Local sepsis	Bacterial growth on the burnt area	Purulent discharge or slough formation, fever with rigors
	Late	Delayed healing	Inadequate and improper treatment, diabetes mellitus	Chronic wound with unhealthy granulation (Fig. 30.18A)
		Keloid formation	Severe scarring	Elevated thick irregular skin lesion (Fig. 30.18B)
		Marjolin's ulcer	Low grade epidermoid carcinoma arising from the epithelium covering the scar tissue after burns	A very slow growing painless ulcer, on the scar without lymphatic spread, resembling squamous cell carcinoma, but the edge is not raised and floor does not contain florid granulation tissue (Fig. 30.19)
		Contractures	Fibrosis of burnt areas around neck, joints	Contracted areas with restricted mobility (Fig. 30.20)
		Disability	Pain and restricted mobility due to contractures	Extensive fibrotic tissues resulting in debility (Figs 30.21A to D)
		Strictures of tubular organs	Annular healing by fibrosis of tubular structures	Dysphagia, if esophagus is involved and dyspnea, if trachea is involved

PHYSICAL EXAMINATION

Clinical examination of a patient of burns should be done quickly but in detail.

Preliminaries Before Examination

- The patient should be weighed using a trolley weighing machine (determination of weight is necessary to calculate the fluid requirement)
- The patient should be sedated well after ensuring the airway patency, and if required intubated
- The patient should be completely undressed.
- Two intravenous portals should be established for fluid resuscitation and medication.

FIG. 30.18A: Infected burns wound

FIG. 30.18B: Keloid formation after burns
(*Courtesy*: Dr K Sridhar)

FIG. 30.19: Marjolin's ulcer chest wall
(*Courtesy*: Dr R Rajaraman)

FIG. 30.20: Postburn contracture of neck
(*Courtesy*: Dr K Sridhar)

FIG. 30.21A: Postburn contractures (*Courtesy*: Dr K Sridhar)

FIG. 30.21B: Postburn scarring resulting in syndactyly (*Courtesy:* Dr K Sridhar)

FIG. 30.21C: Postburns scarring resulting in disability (*Courtesy:* Dr K Sridhar)

FIG. 30.21D: Postburns contracture of axillary fold (*Courtesy:* Dr K Sridhar)

Circumferential burn injuries (around the chest and limbs) especially for the third degree burns, escharotomy will have to be performed or otherwise, respiratory embarrassment and ischemic necrosis of limbs may occur.

INSPECTION

The patient should be examined head to foot, with special attention to the areas like the axillae, groins and the perineum. The extent of damage both by area and depth should be assessed quickly and precisely, as this helps in the determination of fluid replacement.

I. *Extent of area of burns:* This is calculated in two ways:
   i. For small, multiple and scattered areas: 'Rule of hand' is used; determination of surface area of burns, using the patient's own hand (with adducted fingers), which is equal to 1 percent
   ii. For large areas: 'Wallace's rule of nine is used (Fig. 30.22A). For children, the rule is modified (Fig. 30.22B), as the surface of area of the head is bigger relative to the other parts of the body.

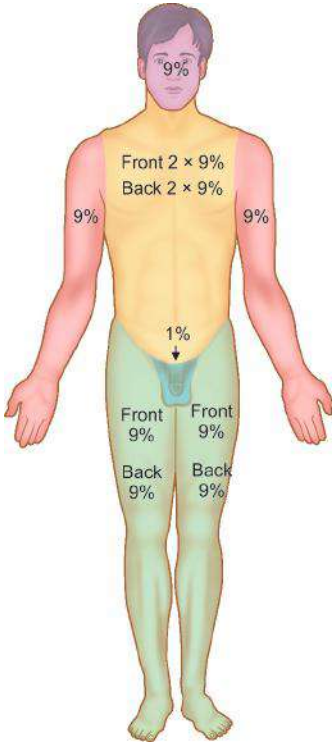

FIG. 30.22A: Wallace's rule of nine for adults

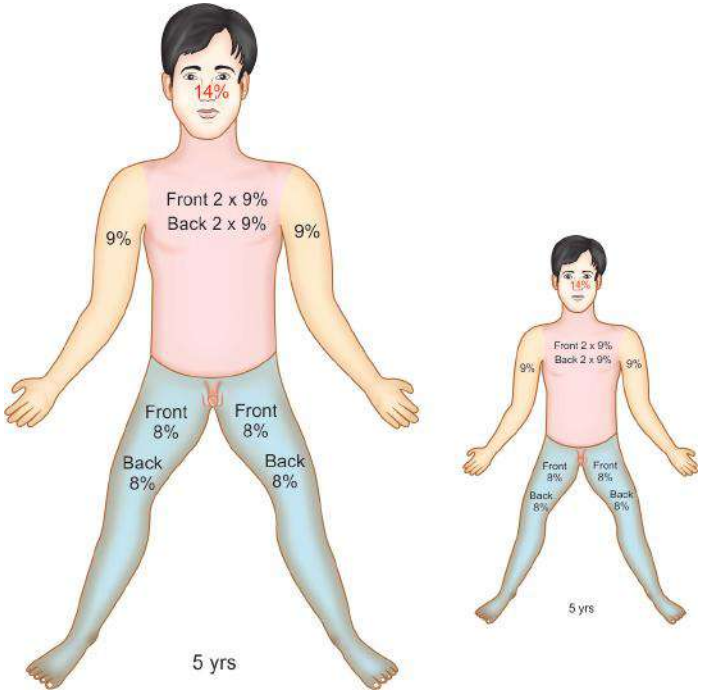

FIG. 30.22B: Modified rule of nine for children

II. *Depth of tissue damage:* This can be done by inspecting the burnt area, and reasonable conclusions may be reached.

It should be remembered that various degrees of tissue damage coexist in the same patient.

- Cutaneous sensation is tested by pricking the burnt area (sterile needle test) with a sterile needle. Presence of sensation is diagnostic of partial thickness burns, but it is not pathognomonic.

Plucking the hair from the burnt area can show the depth. In full thickness burns, the hair can be plucked easily, whereas in partial thickness burns it is not.

General Examination

- *Examination of sensorium:* Usually the sensorium remains normal and not willing to answer may indicate a psychiatric problem (e.g. depression or withdrawal). Inability to answer may indicate state of unconsciousness or altered sensorium (e.g. shock, hyponatremia, uremia)
- *Examination of face:*
  - *Eyes:* Sunken eyes indicate dehydration and undernutrition. Examination of conjunctivae will reveal anemia.
  - *Hairy areas:* Surging of hair in eyebrow, moustache or over forehead will indicate respiratory burns
- *Examination of skin:* The skin should be examined for dehydration and undernutrition. Generalized edema may be seen in severely burnt patients.
- *Examination of tongue:* Dry tongue indicates dehydration
- *Recording of vital signs:* Tachycardia may indicate infection or hypovolemia. Elevated temperature may indicate associated infections and toxemia (usually seen in the second week). Hyperventilation may indicate hydration, hypovolemia and is predominant in respiratory burns. Hypotension may indicate hypovolemia.

Systemic Examination

- *Examination of oral cavity:* A systematic examination of oral cavity is essential (e.g. chemical burns, burns due to inhalation of gases)
- *Examination of chest:* Auscultation—Adventitious sounds may indicate aspiration pneumonitis or respiratory infections.

Relevant Investigations

- Hematology
  - Hemoglobin for anemia
  - Total and differential leukocyte count, e.g. infections

- ESR may be raised in infections and malignancies
  Note: Hemoglobin levels initially remain normal due to hemoconcentration and serial determinations are necessary on subsequent days.

- Biochemistry
  - *Blood sugar:* Required especially in diabetics. Repeated determination is required in patients on hyperalimentation
  - *Liver function tests:* Reduced levels of serum proteins indicate loss of proteins through burn wound and will help in the replacement with blood or plasma
  - *Renal function tests:* Uremia and creatininemia indicate deranged renal status due to dehydration and hypovolemia
  - *Serum electrolytes:* Determination of serum levels of electrolytes is important in the management of burns, as electrolyte disturbances are commonly found in burns, due to evaporation during injury and later through wound.
- Blood grouping and crossmatching: It is important as the requirement of blood may be immediate.
- Radiology
  - Chest X-ray is useful (e.g. aspiration pneumonitis, inhalation injuries of lungs)
  - Contrast studies
    - Barium swallow may be required at later stages to rule out esophageal strictures
    - Bronchogram also may be required to rule out strictures of tracheobronchial tree.
- Endoscopy
  - Esophagoscopy and bronchoscopy may be required in some cases to assess the esophageal and bronchial injuries respectively, especially when injuries occur due to inhalation of toxic fumes
  - Gastroscopy may be needed to diagnose stress ulcers of stomach.
- *Electrocardiogram (ECG):* To rule out disturbances of heart especially in electrical burns, and may have to be repeated frequently.

Treatment

Resuscitation

- Basic principles of the burns victim are the same as for any other patient.
- Problems are specifically related to thermal injuries to airways, large fluid losses and potential for infection:
  - Cover burn areas with sterile drapes or plastic film to reduce infection and fluid loss
  - Give humidified oxygen by mask
  - Endotracheal intubation is required for airway injuries
  - Monitor hematocrit and electrolytes
  - Blood may be required to maintain hematocrit

- Good IV access is needed for fluid administration
  - IV analgesia
  - IV fluids

Examples of fluid regimens for resuscitation of burn victims	
Mount Vernon formula	*Parkland formula*
4.5% albumin	Ringer lactate
Volume (ml) = 0.5 × wt (kg) × % burn	Volume (ml) = 4 × wt (kg) × % burn
Over six consecutive periods of 4, 4, 4, 6, 6, and 12 hours each	Given over 24 hours

NECROTIZING FASCIITIS

Incidence and Etiology

- A rapidly spreading soft tissue infection, which can affect any part of the body
- It follows, usually a minor trauma but the exact etiology is not fully understood
- Causative organisms—Mixture of gram-negative and anaerobic organisms (bacteroides and *Clostridium* sp. anaerobic streptococci) and/or group A *streptococcus*.

Pathogenesis

The infection spreads along the fascial planes causing necrosis of skin and subcutaneous tissues. Muscle layers are usually spared. The infection may spread rapidly and can be fatal in few hours.

Symptom

Painful discolored area.

Sign

The affected area is discolored and indurated.

Relevant Investigations

- Isolation of organism in pus by culture
- Determination of blood sugar levels is mandatory.

Treatment

- Control of diabetes is necessary
- Emergency excision of slough and necrotic tissue under cover of appropriate antibiotics is curative.

Section VI

Special Procedures and Situations

31. Emergency Surgical Procedures
    - Pericardial Aspiration
    - Intubation of Trachea
    - Percutaneous Tracheostomy
    - Cricothyroidotomy/ Minitracheostomy
    - Insertion of Chest Drains
    - Passing a Nasogastric Tube
    - Peritoneal Tap
32. Preparing the Patient for Emergency Surgery
33. Antibiotics
    - Antibiotics and Emergency Surgery
34. Death

31

Emergency Surgical Procedures

PERICARDIAL ASPIRATION

Pericardial aspiration (Pericardiocentesis) is indicated in:

- Cardiac tamponade
- Large pericardial effusion
- For diagnostic pericardial fluid.

Materials Required

- Sterile gowns and gloves
- 10 ml syringe with needle
- Pericardial aspiration kit
- Sutures
- Securing tapes.

Procedure

- Position: Supine with 20 degrees head-up
- Establish IV access
- Adequate sedation
- With full aseptic technique, infiltrate the point of needle entry (just below and left of xiphisternum, between the xiphisternum and left costal margin)
- Introduce the needle directed to the left shoulder, and keep aspirating (Fig. 31.1)

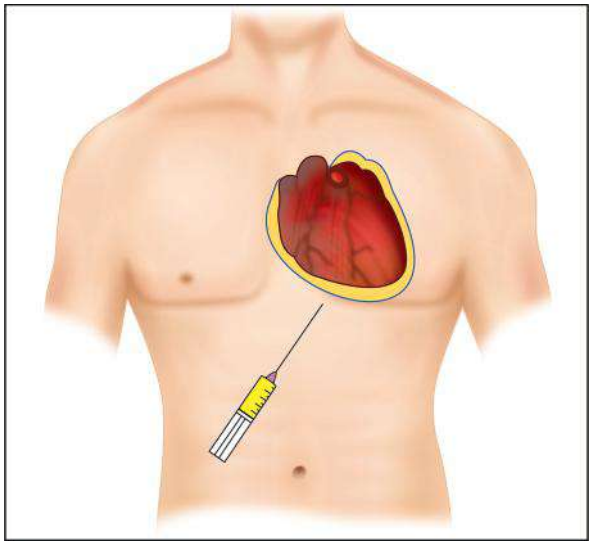

FIG. 31.1: Pericardial aspiration

- Once a straw colored fluid is drawn, keep the needle in that position, pass the guidewire and withdraw the needle
- Pass the catheter over the guidewire into the pericardial space and attach a 3-way tap
- Using a 50 ml syringe, aspirate the effusion or attach to a closed drainage system
- Suture the drain in place and strap.

Complications

- Pneumothorax
- Ventricular tachycardia
- Myocardial puncture
- Damage to coronary arteries.

- Small pericardial effusions not causing hemodynamic instability do not require pericardiocentesis
- Repeat chest X-ray and ECG

INTUBATION OF TRACHEA

Though this procedure is performed by the anesthetists it is better for the surgeons and surgical students to know, as this pertains to life-saving situation. They fall into three groups for which it is required.

- Relieving airway obstruction (e.g. tumors, head and neck trauma, surgery, airway edema)
- Protection of airway from aspiration (e.g. obtunded conscious level, impaired cough impulses)
- Facilitation of ventilation of airways (e.g. anesthesia and surgery, multiple organ failure, major trauma and brain injury).

Materials Required

- Self-inflating bags (Ambu bag)
- Face mask
- Oral/nasal airways
- Suction apparatus and suction catheters
- Laryngoscopes
- Endotracheal tubes (or various sizes)
- Sterile lubricant
- Syringe
- Anesthetic drugs/muscle relaxants
- Emergency drugs (atropine, adrenaline).

Procedure

- Oxygenate the patient well with 100 percent oxygen for about 3 to 4 minutes, which will wash out the nitrogen and fill the functional residual capacity with oxygen, thereby increasing the safety margin
- Keep the head in position (neck flexed, atlantoaxial joint extended on a firm pillow – 'sniffing the morning air' position)
- Give sedative/muscle relaxant as appropriate
- Hold the laryngoscope in the left hand and insert the blade into the right of the mouth sweeping the tongue under it. As the blade reaches the base of the tongue, the epiglottis is seen. Apply traction to gently draw the epiglottis forward exposing the V-shaped glottis behind
- Pass the endotracheal tube between the vocal cords so that cuff is distal to them
- Withdraw the laryngoscope gently
- Inflate the cuff of the endotracheal tube with air so that the tube snugly fits into the trachea
- Check the position of tube by auscultation and by observing the chest movements.

Complication of Endotracheal Intubation

Immediate

- Hypoxia (due to repeated attempts)
- Misplacement of tube
- Obstruction of airway
- Trauma to airway, teeth.

Late

- Accidental displacement of tube
- Ventilator associated pneumonia
- Injury to vocal cords
- Tracheal stenosis.

Tracheal intubation is to done only if the student is familiar with this and has done with a senior in the past.

PERCUTANEOUS TRACHEOSTOMY

Conventional tracheostomy was performed only after patients had been intubated for about 10 – 14 days for fear of laryngeal and subglottic injury. Percutaneous tracheostomy has come into practice as the benefits are many.

Advantages of Tracheostomy

- More comfortable than the oro/nasotracheal tubes
- Need for muscle relaxants is less
- Switching from IPPV to T piece is easy without extubation
- Effective endotracheal suction can be done
- Speech is possible with cuff deflation or speaking tube.

Indications

- Actual or impending airway obstruction
- Difficult intubation
- Need for prolonged ventilator support.

Contraindications

- Distorted/abnormal anatomy
- Significant hemodynamic instability.

Materials Needed for Percutaneous Tracheostomy

- Skin disinfectant
- Local anesthetic
- Syringe
- 10 ml normal saline
- Basic surgical instruments
- Suture
- Securing tapes
- Emergency drugs
- Percutaneous tracheostomy kit (Fig. 31.2).

Procedure

- Position: supine with neck extended over a pillow
- Palpate the cricothyroid membrane and sternal notch

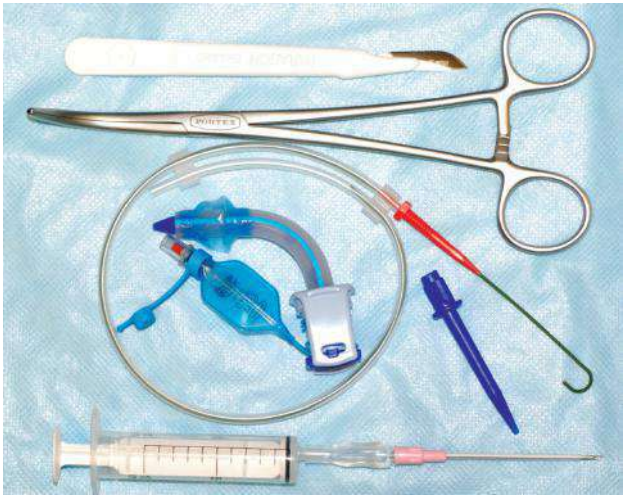

FIG. 31.2: Percutaneous tracheostomy kit

- Infiltrate with local anesthetic
- Make a 2 cm transverse incision over the membrane at the midline
- Puncture the trachea (between the first and third tracheal rings) with the introducer needle
- Confirm its position with a syringe aspiration (air/mucus)
- Pass the guidewire through the needle and remove the needle
- Dilate the trachea
- Insert the endotracheal tube
- Remove the guidewire
- Suck out all the secretions and blood clots in the trachea
- Inflate the cuff
- Place two stay sutures through the wings of tracheostomy tube, to prevent slippage, and secure with tapes
- Check the position of the tracheostomy tube (symmetrical chest expansion, X-ray).

Complications

Early

- Bleeding
- Tube misplacement
- Mucus plugging

Late

- Tracheal stenosis
- Tracheoesophageal fistula.

CRICOTHYROIDOTOMY/MINITRACHEOSTOMY

Cricothyroidotomy is an emergency procedure to access the airway, when measures like intubation have failed. It involves the insertion of a small tube through the cricothyroid membrane for attaching to the ventilator.

Minitracheostomy involved the use of a small bored and noncuffed bore, with an internal diameter of 4 mm. Kits are available for both procedures.

Indications

- Cannot intubate
- Cannot ventilate
- Severe midfacial trauma

- Cervical spine injuries
- Chemical inhalation injuries.

Contraindications

- Inability to identify landmarks (cricothyroid membrane)
- Underlying anatomical abnormality (tumor)
- Acute laryngeal disease
- Tracheal transection
- Small children.

Materials Required

- Skin disinfectant
- Sterile drape
- Syringe with local anesthetic
- Cricothyroidotomy (Fig. 31.3)/Minitracheostomy kit
- Suture
- Securing tapes.

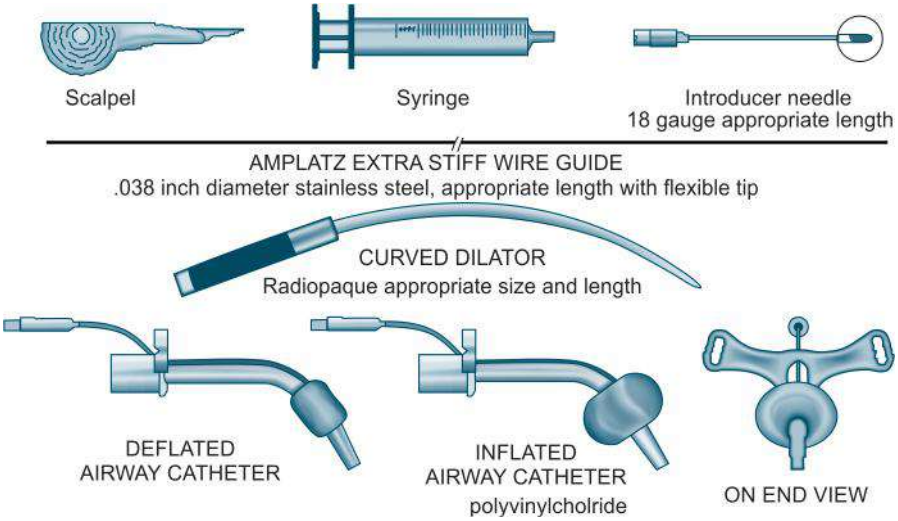

FIG. 31.3: Cricothyroidotomy kit

Procedure

- Palpate the cricothyroid membrane
- Clean the skin
- Infiltrate over the cricothyroid membrane
- Make a skin incision
- Introduce needle into trachea and aspirate air
- Pass guidewire and remove needle
- Pass the introducer over the guidewire and slide the cricothyroidotomy/minitracheostomy tube off the introducer
- Remove the introducer and guidewire together, leaving the tube *in situ*
- Suck any blood or mucus
- Place two stay sutures through the wings of tracheostomy tube, to prevent slippage, and secure with tapes
- Check the position of the tracheostomy tube (symmetrical chest expansion, X-ray).

Complications

Early

- Bleeding
- Tube misplacement
- Mucus plugging

Late

- Tracheal stenosis
- Tracheoesophageal fistula.

INSERTION OF CHEST DRAINS

Insertion of a chest drain is required as an emergency treatment of life-threatening tension pneumothorax. A large bore tube preferably a 14-gauge cannula is used for this purpose.

Indications

Draining of:

- Pneumothorax
- Hemothorax
- Pleural effusion

- Empyema
- Chylothorax.

Type of Drain

Chest drains are of two types:

1. Large bore tubes
2. Seldinger drains.

Site of Drain

This is partly decided by the position of the collection clinically and radiographically.

- For draining the pneumothorax, the tube is placed in the 2nd intercostal space in the midclavicular line (Fig. 31.4).
- For draining the fluid collections, the tube is placed in the 5th intercostals space just anterior to the midaxillary line (Fig. 31.4).

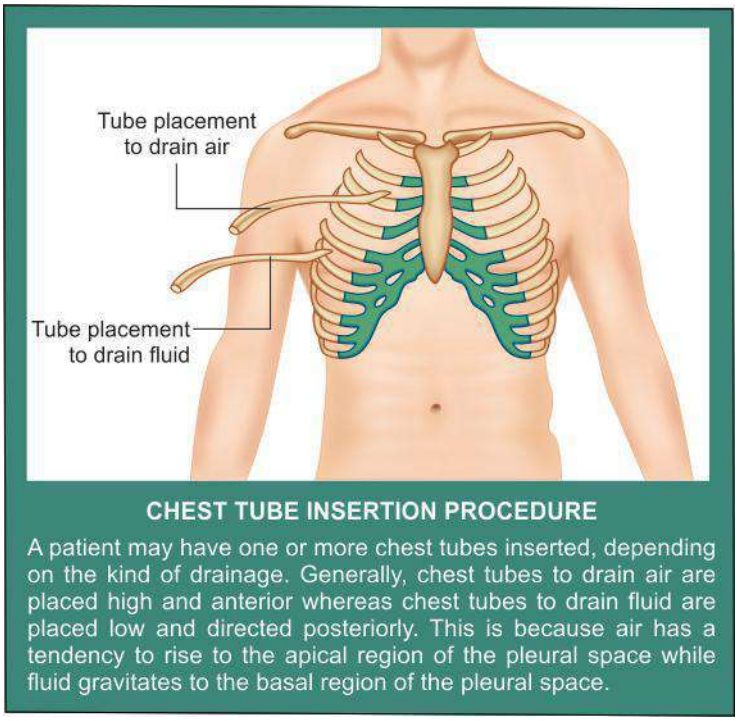

FIG. 31.4: Placing the chest drains

Materials Required

- Skin disinfectant
- Skin drapes
- Syringe
- Local anesthetic
- Basic instruments
- Chest drain
- Silk sutures
- Adhesive tapes
- Underwater seal.

Procedures

Through Thoracostomy (Figs 31.5A to F)

- Position: supine with arm lifted, with a pillow behind the back
- Clean the field with sterile preparation
- Infiltrate the local anesthetic at the chosen site
- Make a small skin incision
- The intercostal muscles and the pleura are dissected out and plunged with artery forceps
- The chest drain is inserted into the pleural cavity
- The drain is connected to the underwater drain (Fig. 31.6)
- The drain is secured with sutures to the skin
- Adhesive strapping is done
- Order for a X-ray and the position checked.

Seldinger Technique

- Position: supine with arm lifted, with a pillow behind the back
- Clean the field with sterile preparation
- Infiltrate the local anesthetic at the chosen site
- Use the chest drain kit (Fig. 31.7)
- Advance the needle through chest until blood/fluid/air is aspirated
- Pass chest tube/pigtail catheter over guidewire
- Attach to the underwater seal
- Secure well and dress.

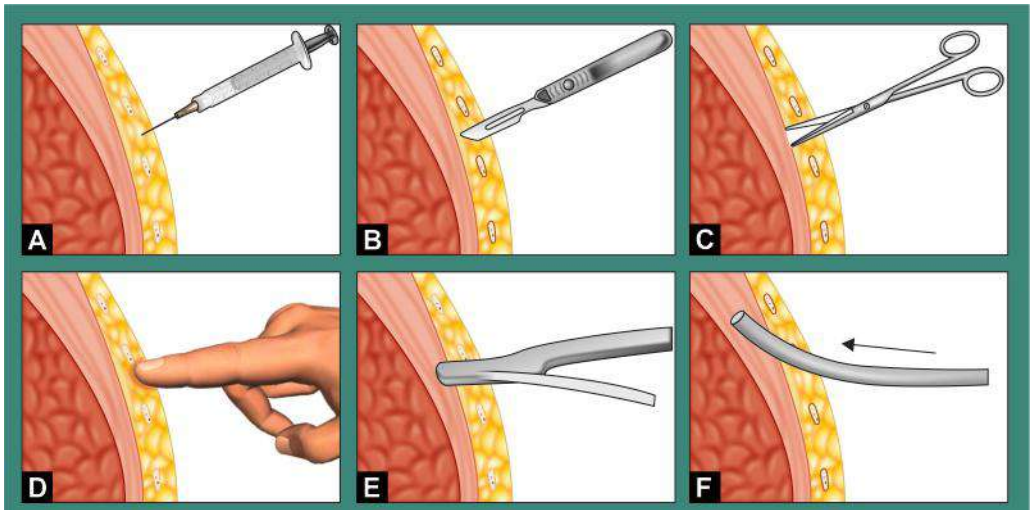

CHEST TUBE INSERTION PROCEDURE

The distal end of the chest tube is clamped and, using the clamp as a guide, inserted into the incisional site (E above). At this time, the patient should be encouraged to take a deep breath; this will displace the diaphragm downward, minimizing the risk of its injury. The clamp is removed and the tube is then advanced into the pleural space and directed anteriorly or posteriorly depending on the material being drained (F above).

FIGS 31.5A to F: Method of placing the chest drain

- Never clamp the chest drain, as clamping may produce a tension pneumothorax
- While shifting the patient, the underwater drain should be kept below the level of chest

PASSING A NASOGASTRIC TUBE

Nasogastric tube is generally used to decompress the stomach and also for feeding.

Indications

- To deflate the stomach
- To aspirate gastric contents
- To provide enteral nutrition.

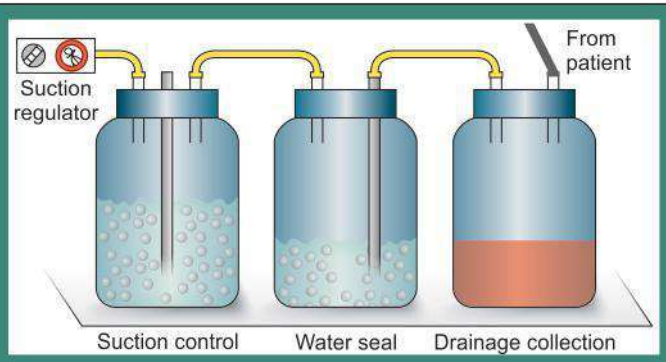

THREE-BOTTLE SYSTEM

When suction is turned on, air and fluid are pulled out of the pleural space and into the drainage collection bottle. Suction is applied through the entire system until it reaches the pressure that will draw atmospheric air in through the open tube of the suction control bottle. When the incoming atmospheric air reaches the lower end of the tube, it bubbles into the bottle. At the point, the desired suction level will be maintained as any increase in suction will just draw in more atmospheric air.

FIG. 31.6: Three chamber underwater drain

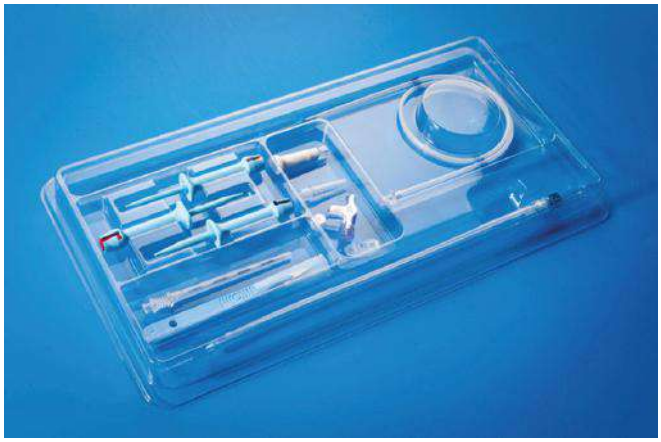

FIG. 31.7: Seldinger chest drain kit

Contraindications

Severe coagulopathy.

Materials Required

- Gloves and mask
- Nasogastric tube
- Lubricating jelly
- Laryngoscope (needed for unconscious patients).

Procedure (Fig. 31.8)

- Position: supine and head neutral or semirecumbent
- Lubricate the nasogastric (NG) tube well
- Keeping in alignment with the long axis of the patient, introduce the tube into the nostril
- If the patient is cooperative, ask him to swallow
- The tube generally flows down along with the swallowing movement
- Confirm the tip of NG tube in the stomach by auscultation or aspiration of contents
- Secure NG tube in position with adhesive tape.

PERITONEAL TAP

Peritoneal tap is required to obtain samples for diagnostic purposes and also relieve the tension of a tense ascites.

Materials Required

- Skin disinfectant
- Sterile drapes
- 10 ml syringe with local anesthetic
- Pig tail drain
- Sutures
- Adhesive strapping
- Ultrasound.

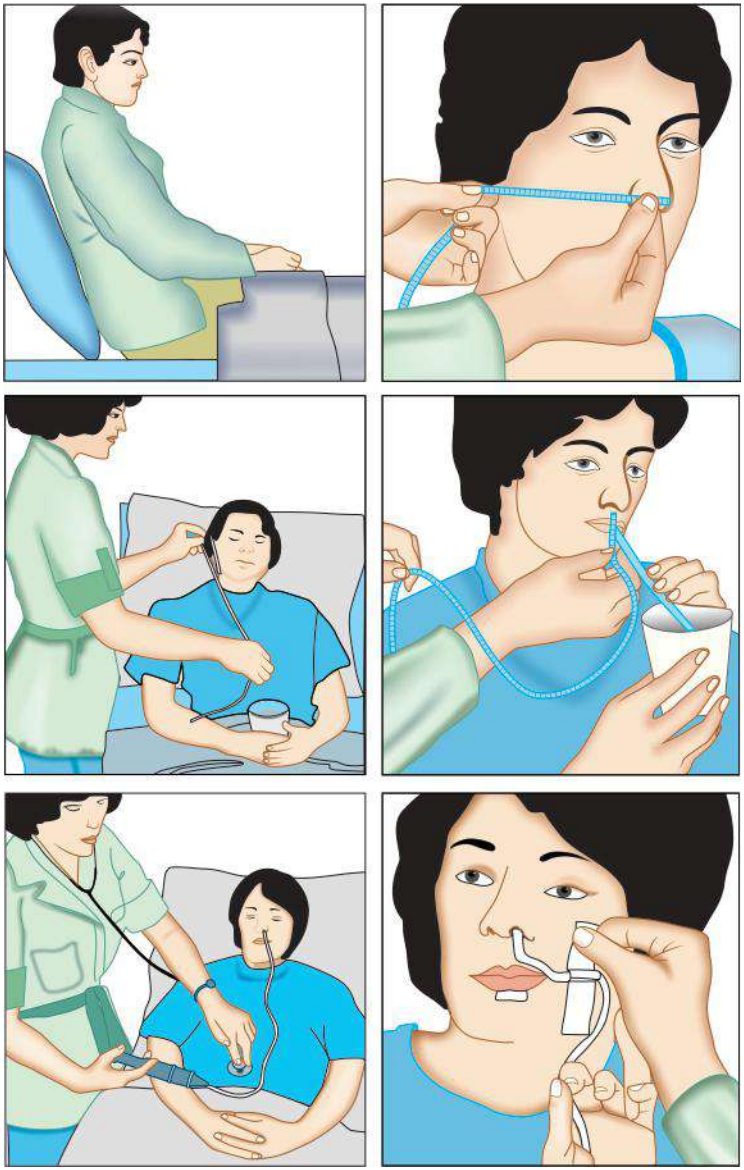

FIG. 31.8: Nasogastric tube insertion

FIG. 31.9: Peritoneal tap

Procedure

- Clean the skin with disinfectant
- Confirm the presence of fluid by percussion or by ultrasound
- At the chosen site, inject the local anesthetic
- Introduce the syringe and aspirate fluid to confirm its presence (Fig. 31.9)
- Introduce pig tail drain using Seldinger technique
- Connect the drain to collecting bottle
- Strap the drain.

32

Preparing the Patient for Emergency Surgery

Any patient for surgery needs optimization, especially so the patients for emergency surgery, as the parameters are never normal. This optimization becomes necessary to reduce the perioperative complications. The optimization requires:

- Fluid and electrolyte corrections
- Correction of hematological and metabolic disturbances.

General Measures

- Nil by mouth
- Good intravenous access
- Appropriate intravenous fluid replacement
- Appropriate antibiotics
- Adequate analgesia
- Measures for DVT prophylaxis
- NG tube insertion where stomach needs to be decompressed
- Urinary catheterization
- Central venous line access
- Informed consent.

Investigations Support

- Complete blood count (CBC)
- Urea and electrolytes
- Electrocardiogram (ECG) and Chest X-ray (CXR).

Situations which Need Special Attention

- Uncontrolled glycemic status
- Ketoacidosis
- Cardiovascular complications
- Altered conscious level
- Jaundiced patient with coagulation abnormalities
- Patient on steroids
- Patient on anticoagulants.

33

Antibiotics

ANTIBIOTICS AND EMERGENCY SURGERY

Antibiotics become necessary to prevent an infection or control an infection. Prevention becomes necessary so that sepsis related complications, which carry high morbidity and mortality are averted.

Prophylactic antibiotics are administered within 1 hour before the start of the surgical procedure. Second dose becomes necessary only when contamination is encountered or the blood loss exceeds 1 liter, since it is shown that antibiotic clearance is slower in patients undergoing surgery. Postoperative administration is only required when there is heavy contamination.

Which Emergency Surgical Procedure should have Antibiotic Prophylaxis ?

Antibiotic prophylaxis is recommended for the following emergency surgical procedures

- Colorectal surgery
- Appendicectomy
- Hernia surgery (use of mesh)
- Open biliary surgery
- Lower limb revascularization surgery.

The surgeries are classified based on the contamination at the time of surgery, and the infection rates vary differently (Table 33.1).

Surgical site infections (SSIs) are defined as infective complication occurring within 30 days of surgical procedure. The infection is caused by various organisms and some antibiotics are effective against these organisms (Table 33.2).

Table 33.1: Classification of surgeries related to contamination

Classification of case	*Type of surgery*	*Wound infection rate*
Clean	Surgery in noninflamed tissue, and bowel not opened	<2%
Clean contaminated	Surgery of hollow visus other than colon, with minimal contamination	10%
Contaminated	Surgery through a hollow viscus with gross spillage	20%
Dirty	Fecal peritonitis, traumatic wound contaminated for >4 hrs, frank pus	40%

Table 33.2: Relationship of surgical procedures with potential organisms and sensitive antibiotics

Surgical procedures	*Organism*	*Antibiotics*
General surgical procedures	*Staphylococcus aureus*	Flucloxacillin, ciprofloxacin, clindamycin
	Beta – hemolytic streptococci	Ciprofloxacin, clindamycin
Upper GIT	Enteric Gram-negative bacilli	Cefuroxime, co-amoxyclav, gentamycin
Lower GIT	Enteric Gram-negative bacilli	Cefuroxime, co-amoxyclav, gentamycin
	Anaerobes	Metronidazole
Procedures using prosthesis	Coagulase + staphylococci	Vancomycin + removal of prosthesis

Table 33.3: Antibiotics related to Gram staining

Gram stain	*Antibiotic*
Gram-negative rods	Cefuroxime + / – gentamycin +/– metronidazole
Gram-positive cocci (Streptococci)	Amoxicillin
Gram-positive cocci (Staphylococci)	Flucloxacillin (Vancomycin / Teicoplain)
Gram-negative cocci	Ceftriaxone

Prior to full culture reports becoming available, a Gram stain may provide information to the likely organism. Empiric antibiotics can be used on Gram stain with fair accuracy (Table 33.3).

- Infection prevention and control measures must be adhered to
- Sepsis is usually bacterial in origin, though fungi and viruses should be considered
- Obtaining appropriate microbiology samples will help to use the right antibiotic
- Choice of antibiotic should be carefully considered.

34

Death

A death is certified only when the following are documented:

- Absent respiratory rate for 3 minutes
- Absent pulse and heart sounds for 1 minute
- Fixed dilated pupils.

Breaking the News

Breaking a bad news especially of the impending death or after death is very tricky. This will upset the family which inherently is stressed. Well informed family will accept but ill informed or less informed will not take it very easy. To overcome this situation, the clinician should appear:

- Competent
- Honest
- Clear.

It is better to choose a suitable environment for discussion with the close relatives. Crowding of all family members and relatives should be politely discouraged. While discussing this it is better to:

- Review the case record in full
- Prepare the anticipated questions with your colleagues
- Keep a colleague or nurse alongwith
- Understand the current level of understanding
- Explain in detail

AUDEL™ THE BOOKS THAT WORK™

CARPENTER'S AND BUILDER'S MATH, PLANS & SPECIFICATIONS

ALL NEW 7TH EDITION

Mark Richard Miller & Rex Miller

Audel™

Carpenter's and Builder's Math, Plans, and Specifications

All New 7th Edition

Mark Richard Miller
Rex Miller

WILEY
Wiley Publishing, Inc.

TLFeBOOK

Vice President and Executive Group Publisher: Richard Swadley
Vice President and Publisher: Joseph B. Wikert
Executive Editor: Carol Long
Editorial Manager: Kathryn A. Malm
Development Editor: Kevin Shafer
Production Editor: Angela Smith
Text Design & Composition: TechBooks

Copyright © 2005 by Wiley Publishing, Inc., Indianapolis, Indiana. All rights reserved.

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Legal Department, Wiley Publishing, Inc., 10475 Crosspoint Blvd., Indianapolis, IN 46256, (317) 572-3447, fax (317) 572-4355, e-mail: brandreview@wiley.com.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Trademarks: Wiley, the Wiley Publishing logo, Audel, and The Books That Work are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates. All other trademarks are the property of their respective owners. Wiley Publishing, Inc., is not associated with any product or vendor mentioned in this book.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Control Number: 2004110825

eISBN: 0-7645-7981-9

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents

Acknowledgments **xiii**

About the Authors **xv**

Introduction **xvii**

Chapter 1 Nails, Screws, Bolts, and Other Fasteners **1**
Nails 1
The Penny System 3
Nail Characteristics 3
Selecting Nail Size 12
Driving Nails 13
Screws 16
Material 17
Dimensions of Screws 18
Shape of the Head 18
How to Drive a Wood Screw 18
Strength of Wood Screws 21
Lag Screws 23
Bolts 23
Manufacture of Bolts 23
Kinds of Bolts 27
Proportions and Strength of Bolts 27
Fasteners for Plaster or Drywall 33
Expansion Anchors 33
Hollow Wall Screw Anchors 34
Toggle Bolts 36
Hanger Bolts, Dowel Screws, and Toggle Studs 37
Framing Fasteners 39
Summary 39
Review Questions 40

Chapter 2 Wood as a Building Material **41**
Growth and Structure of Wood 41
Softwoods and Hardwoods 41

Lumber Conversion 43
Seasoning of Wood 44
Moisture Content 47
Density 49
Strength 51
Deadwood 53
Virgin and Second Growth 54
Time of Cutting Timber 54
Air-Dried and Kiln-Dried Wood 54
Sapwood Versus Heartwood 55
Defects and Deterioration 55
Decay of Wood 57
Nuclear Radiation 63
Weathering 63
Insect Damage 65
Summary 65
Review Questions 65

Chapter 3 Lumber, Plywood, and Other Wood Products 67
Standard Sizes of Lumber 67
Softwood Lumber Grades for Construction 68
Stress-Graded Lumber 68
Nonstress-Graded Lumber 69
Appearance Lumber 70
Plywood 77
Types of Plywood 79
Grades of Plywood 79
Particle Board 79
Lumber Distribution 82
Retail Yard Inventory 82
Important Purchase Considerations 86
Engineered Lumber 87
Summary 88
Review Questions 88

Chapter 4 Strength of Timbers 89
Tension 89
Compression 90
Working Stresses for Columns 91
Shearing Stresses 92
Horizontal Shears 94
Transverse or Bending Stress 96
Stiffness 96
Modulus of Elasticity 100
Beams 100
Allowable Loads on Wood Beams 100
Breaking Loads on Wood Beams 102
Distributed Load 102
Cantilever Beams 102
Wind Loads on Roofs 104
Definitions 106
Summary 108
Review Questions 109

Chapter 5 Mathematics for Carpenters and Builders 111
Arithmetic 111
Arithmetic Alphabet 111
Notation and Numeration 111
Roman Notation 113
Definitions 113
Signs of Operation 115
Use of the Signs of Operation 116
Fractions 117
Decimals 122
Compound Numbers 126
Ratios 127
Proportion 127
Rule of Three 128
Percentage 128
Powers of Numbers (Involution) 129
Roots of Numbers (Evolution) 129

Measures 131
Linear Measure 132
Square Measure 133
Cubic Measure 133
Dry Measure 135
Board or Lumber Measure 136
Liquid Measure 137
Measures of Weight 137
Time Measure 139
Circular Measure 139
The Metric System 140
Geometry 142
Lines 142
Angles 142
Plane Figures 143
Solids 146
Geometrical Problems 146
Mensuration 167
Trigonometry 182
Trigonometric Functions 182
Functions of Numbers 197
Summary 197
Review Questions 197

Chapter 6 Surveying 199
The Level 199
Construction of the Wye Level 199
Lines of the Level 201
Adjustments of the Wye Level 202
Leveling Rod 205
Methods of Leveling 206
Directions for Using Level 209
Trigonometric Leveling 211
The Transit 212
Construction of the Transit 212
Lines of a Transit 218
Adjustments of the Transit 219

	Adjustments of the Compass	221
	Instructions for Using the Transit	222
	Gradienter	224
	Care of Instruments	224
	The Stadia	225
	Other Devices	226
	Laser Levels	227
	Summary	232
	Review Questions	232
Chapter 7	**The Design Process**	**233**
	Design Considerations	233
	An Example of Design	236
	Conference with Builder	236
	Changes Agreed On	238
	Summary	240
	Review Questions	241
Chapter 8	**Building Specifications**	**243**
	Example of Specifications	243
	Introduction	243
	Height of Ceilings	243
	Interpretation of Drawings	244
	Conditions	244
	Mason's Work	244
	Foundation	245
	Chimneys	245
	Mortar	245
	Installing Drywall	245
	Tiling	245
	Other Floors	246
	Coping	246
	Timber	246
	Framing	246
	Spacing and Bridging	247
	Partitions	247
	Lumber	247

	Sheathing and Sheathing Paper	247
	Exterior Finish	247
	Shingling	247
	Flashing	248
	Flooring	248
	Window Frames	248
	Sash	248
	Screens	248
	Glazing	248
	Blinds	248
	Door Frames	248
	Doors	249
	Interior Trim	249
	Stairs	249
	Mantels	249
	Pantry Cabinets	250
	Closet Shelving	250
	Plumbing	250
	Electric Wiring	250
	Water Pipes	250
	Kitchen Sink	251
	Wash Trays	251
	Bathroom Fixtures	251
	Painting	251
	Condition of Bids	252
	Summary	252
	Review Questions	252
Chapter 9	**Architectural Drawings**	**255**
	Use of Drawings	255
	Reading Drawings	255
	Projected Views	255
	Top View or Plan	256
	Right-End View (Elevation)	256
	Left-End View (Elevation)	258
	Sections	258

Directions of View for Sectional Views 259
The Scale 260
Drawing Development 262
Graphic Symbols 267
Summary 278
Review Questions 279

Chapter 10 Building Styles Explored 281
The Two-Story New England Colonial House 281
A House of Modern Architecture 284
The Contemporary House 285
Utility Pole-Type Building 287
Summary 291
Review Questions 292

Appendix Construction Hardware 293

Index 349

Chapter 2

Wood as a Building Material

Wood is the most versatile, most useful building material, and a general knowledge of the physical characteristics of various woods used in building is important for carpenters and builders.

Growth and Structure of Wood

Wood, like all plant material, is made up of cells (or *fibers*) that, when magnified, have an appearance similar to (though less regular than) that of the common honeycomb. The walls of the honeycomb correspond to the walls of the fibers, and the cavities in the honeycomb correspond to the hollow or open spaces of the fibers.

Softwoods and Hardwoods

All lumber is divided as a matter of convenience into two great groups: *softwoods* and *hardwoods*. The softwoods in general are the coniferous or cone-bearing trees (such as the various pines, spruces, hemlocks, firs, and cedar). The hardwoods are the non–cone-bearing trees (such as the maple, oak, and poplar). These terms are used as a matter of custom, for not all so-called softwoods are soft, nor are all so-called hardwoods necessarily hard. In fact, such softwoods as long-leaf southern pine and Douglas fir are much harder than poplar, basswood, and so on, which are called hardwoods.

Other (and perhaps more accurate) terms often used for these two groups are the needle-bearing trees (softwoods) and the broad-leaved trees (hardwoods). In general, the softwoods are more commonly used for structural purposes such as for joists, studs, girders, and posts, whereas the hardwoods are more likely to be used for interior finish, flooring, and furniture. The softwoods are also used for interior finish and, in many cases, for floors, but are not often used for furniture.

A tree consists of the following:

- *Outer bark*—The bark is living and growing only at the cambium layer. In most trees, the bark continually sloughs away.
- *Inner bark*—In some trees (notably hickories and basswood) there are long tough fibers (called *bast fibers*) in the inner bark. In other trees (such as the beech), these bast fibers are absent.
- *Cambium layer*—This can be only one cell thick. Only these cells are living and growing.

- *Medullary rays or wood fibers*—These run radially from the center to the bark.
- *Annual rings*—These are layers of wood.
- *Pith*—This is at the very center.

Around the pith, the wood substance is arranged in approximately concentric rings (see Figure 2-1). The part nearest the pith is usually darker than the parts nearest the bark and is called the *heartwood*. The cells in the heartwood are dead. Nearer the bark is the *sapwood*, where the cells carry or store nutrients but are not living (see Figure 2-2).

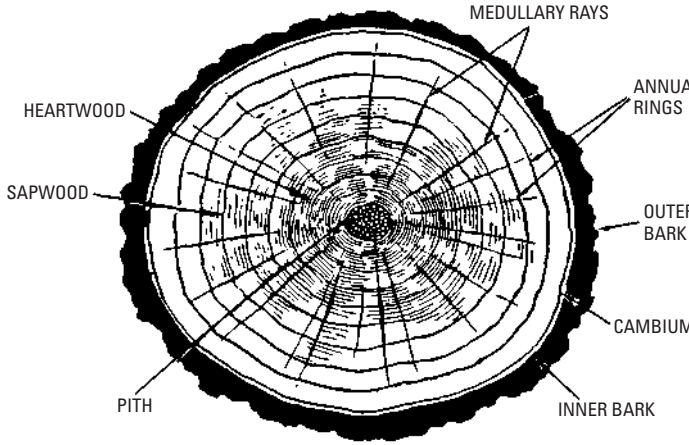

Figure 2-1 Cross-section of an oak nine years old, shows pith, concentric rings comprising the woody part, the cambium layer, and the bark. The tree grows in concentric rings (or layers) with one layer added each year. The rings are also called annual rings.

As winter approaches, all growth ceases. Each annual ring is separate and, in most cases, distinct. The leaves of deciduous trees (or trees that shed their leaves) and the leaves of some of the conifers (such as cypress and larch) fall, and the sap in the tree may freeze hard. The tree is dormant but not dead. With the warm days of the next spring, growth starts again strongly, and the cycle is repeated. The width of the annual rings varies greatly, from 30 to 40 or more per inch in some slow-growing species, to as few as 3 or 4 per inch in some of the quick-growing softwoods.

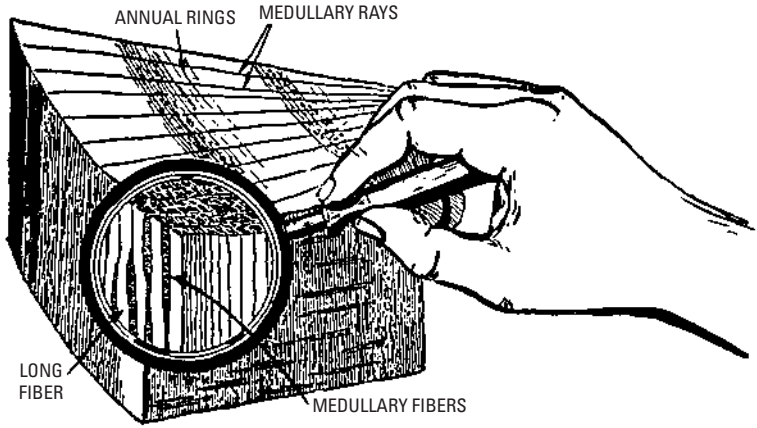

Figure 2-2 A piece of wood magnified to show its structure. The wood is made up of long, slender cells called fibers, which usually lie parallel to the pith. The length of these cells is often 100 times their diameter. Transversely, bands of other cells, elongated but much shorter, serve to carry sap and nutrients across the trunk radially. In addition, in the hardwoods, long vessels or tubes (often several feet long) carry liquids up the tree. There are no sap-carrying vessels in the softwoods, but spaces between the cells may be filled with resins.

Lumber Conversion

When logs are taken to the mill, they may be cut in a variety of ways. One way of cutting is *quartersawing* (see Figure 2-3). Here, each log is ripped into quarters, as shown in the figure. Quartersawing is rarely done this way, though, because only a few wide boards are yielded. There is too much waste. More often, wood is *rift-sawed*. The log is started as shown in Figure 2-4 and is plain-sawed until a good figure (pattern) shows, then turned over and rift-sawed. This way, there is less waste, and the boards are wide.

The result is *vertical-grain*, or *rift-grain* lumber. Vertical-grain lumber shrinks less in width. It is often used in door stiles and rails because it is less likely to warp.

The plain sawed stock is simply flat-sawn out of the log (see Figure 2-5). This results in *flat-grain* or *side-grain* lumber that is used where shrinkage and warping are less critical.

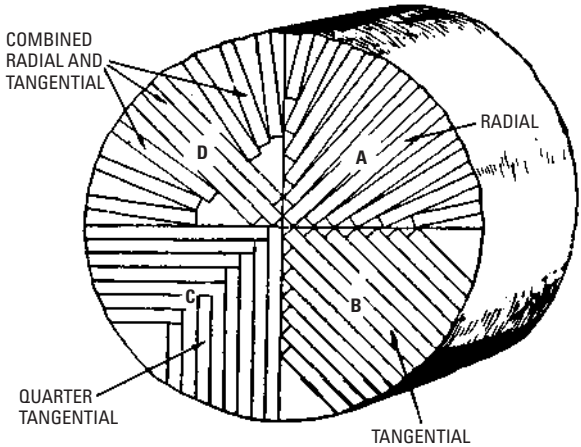

Figure 2-3 Methods of quartersawing. These are rarely used because waste is extensive.

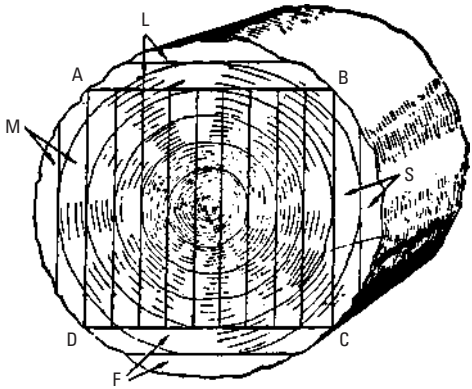

Figure 2-4 Plain or bastard sawing, sometimes called flat or slash sawing. The log is first squared by removing boards M, S, L, and F, giving the rectangular section ABCD. This is necessary to obtain a flat surface on the log.

Seasoning of Wood

Well-developed techniques have been established for removing the large amounts of moisture normally present in green wood. Seasoning is essentially a drying process, but for uses that require them, seasoning includes equalizing and conditioning treatments to improve

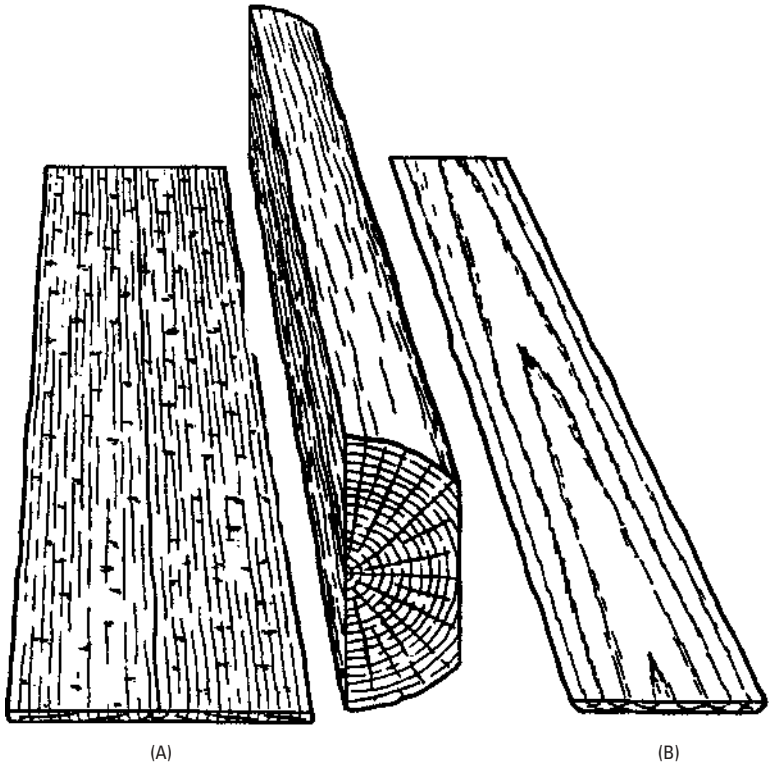

Figure 2-5 (A) Quartersawed and (B) plain-sawed boards cut from a log. *(Courtesy Forest Product Lab)*

moisture uniformity and relieve residual stresses and sets. Careful techniques are necessary, especially during the drying phase, to protect the wood from stain and decay and from excessive drying stresses that cause defects and degrade. The established seasoning methods are air-drying and kiln drying.

Drying has the following effects:

- Reduces the weight of wood, with a resulting decrease in shipping costs
- Reduces or eliminates shrinkage, checking, and warping in service
- Increases strength and nail holding power; decreases susceptibility to infection by blue stain and other fungi

- Reduces chance of attack by insects
- Improves the capacity of wood to take preservative and fire-retardant treatment and to hold paint

It is common practice at most softwood sawmills to kiln-dry all upper-grade lumber intended for finish, flooring, and cut stock. Lower-grade boards are often air-dried. Dimension lumber is air-dried or kiln-dried, although some mills ship certain species without seasoning. Timbers are generally not held long enough to be considered seasoned, but some drying may take place between sawing and shipment, or while they are held at a wholesale or distributing yard. Sawmills cutting hardwoods commonly classify the lumber for size and grade at the time of sawing. Some mills send all freshly sawed stock to the air-drying yard or an accelerated air-drying operation. Others kiln-dry directly from the green condition. Air-dried stock is kiln-dried at the sawmill, at a custom drying operation during transit, or at the remanufacturing plant before being made up into such finished products as furniture, cabinet work, interior finish, and flooring.

Air-drying is not a complete drying process, except as preparation for uses for which the recommended moisture content is not more than 5 percent below that of the air-dry stock. Even when air-drying conditions are mild, air-dry stock used without kiln-drying may have some residual stress and set that can cause distortions after nonuniform surfacing or machining. On the other hand, rapid air-drying accomplished by low relative humidities produces a large amount of set that will assist in reducing warp during final kiln drying. Rapid surface drying also greatly decreases the incidence of chemical and sticker stain, blue stain, and decay.

Air-drying is an economical seasoning method when carried out in a well-designed yard or shed, with proper piling practices (see Figure 2-6) and in favorable drying weather. In cold or humid weather, air-drying is slow and cannot readily reduce wood moisture to levels suitable for rapid kiln-drying or for use.

In kiln drying, higher temperatures and fast air circulation are used to increase the drying rate considerably. Average moisture content can be reduced to any desired value. Specific schedules are used to control the temperature and humidity in accordance with the moisture and stress situation within the wood, thus minimizing shrinkage-caused defects. For some purposes, equalizing and conditioning treatments are used to improve moisture content uniformity and relieve stresses and set at the end of drying, so the material will not warp when resawed or machined to smaller sizes or irregular

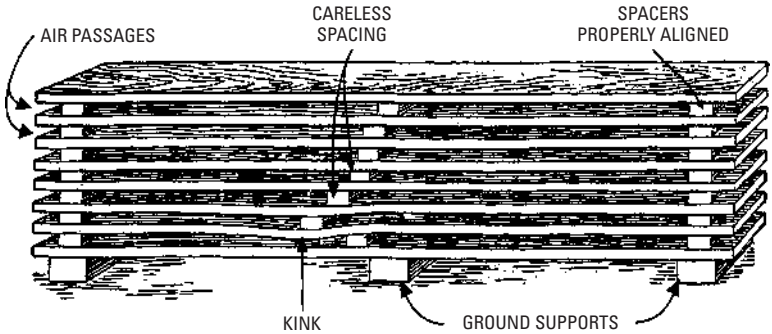

Figure 2-6 Horizontal stack of lumber for air-drying. One end of the pile should be a little higher. This is so that rainwater that falls on top or drives into the pile will drain. Each layer should be separated by three or four spacers so that the air will have free access to both sides of each board. The lowest layer should be well elevated above the ground to protect it from dampness. It takes from one to three years to thoroughly season lumber, depending upon the character of the wood, climatic conditions, and so on. When the spacers are carelessly placed so that they do not lie over each other, the weight must be supported by the board, which especially in the case of a high pile, is considerable, and will in time cause the board to sag, resulting in a permanent kink.

shapes. Further advantages of kiln drying are the setting of pitch in resinous woods, the killing of staining or decay fungi or insects in the wood, and reductions in weight greater than those achieved by air-drying. At the end of kiln-drying, moisture-monitoring equipment is sometimes used to sort out moist stock for redrying and to ensure that the material ready for shipment meets moisture content specifications.

Temperatures of ordinary kiln drying generally are between 110°F and 180°F. Elevated-temperature (180°F to 212°F) and high-temperature (above 212°F) kilns are becoming increasingly common, although some strength loss is possible with higher temperatures.

Moisture Content

While the tree is living, both the cells and cell walls are filled with water to an extent. As soon as the tree is cut, the water within the cells (or *free water* as it is called) begins to evaporate. This process

continues until practically all of the free water has left the wood. When this stage is reached, the wood is said to be at the fiber-saturation point (that is, what water is contained is mainly in the cell or fiber walls).

Except in a few species, there is no change in size during this preliminary drying process, and, therefore, no shrinkage during the evaporation of the free water. Shrinkage begins only when water begins to leave the cell walls themselves. What causes shrinkage and other changes in wood is not fully understood. However, it is thought that as water leaves the cell walls, they contract, becoming harder and denser, thereby causing a general reduction in size of the piece of wood. If the specimen is placed in an oven that is maintained at 212°F (the temperature of boiling water), the water will evaporate and the specimen will continue to lose weight for a time. Finally, a point is reached at which the weight remains substantially constant. This is another way of saying that all of the water in the cells and cell walls has been driven off. The piece is then said to be *oven dry*.

If it is now taken out of the oven and allowed to remain in the open air, it will gradually take on weight, because of the absorption of moisture from the air. As when placed in the oven, a point is reached at which the weight of the wood in contact with the air remains more or less constant. Careful tests, however, show that it does not remain exactly constant—it will take on and give off water as the moisture in the atmosphere increases or decreases. Thus, a piece of wood will contain more water during the humid, moist summer months than in the colder, drier winter months. When the piece is in this condition, it is in equilibrium with the air and is said to be *air dry*.

The amount of water contained by wood in the green condition varies greatly, not only with the species but also in the same species and even in the same tree, according to the position in the tree. However, as a general average, at the fiber-saturation point, most woods contain from 23 to 30 percent water as compared with the oven-dry weight of the wood. When air dry, most woods contain from 12 to 15 percent moisture.

As the wood dries from the green state (which is that of the freshly cut tree) to the fiber-saturation point, except in a few species, there is no change other than that of weight. It has already been pointed out that as the moisture dries out of the cell walls, in addition to the decrease in weight, shrinkage results in a definite decrease in size (see Figure 2-7). It has been found, however, that there is little or no decrease along the length of the grain, and that the decrease is at right angles to the grain.

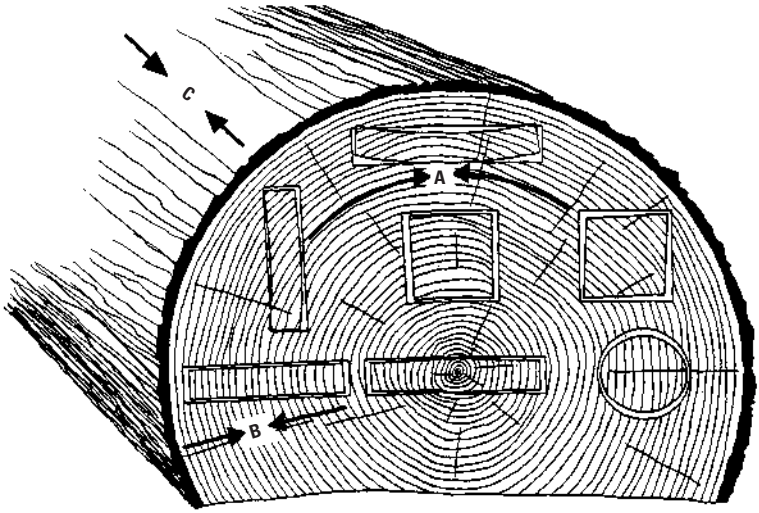

Figure 2-7 Greatest shrinkage is in the direction of the annual rings (A). Characteristic shrinkage and distortion of flat boards, squares, and rounds as affected by the direction of the annual rings. Tangential shrinkage (A) is about twice as great as radial (B). There is little or no longitudinal (C) shrinkage. *(Courtesy Forest Product Lab)*

This is an important consideration to be remembered when framing a building. For example, a stud in a wall will not shrink appreciably in length, whereas it will shrink somewhat in both the 2-inch and the 4-inch way. In like manner, a joist, if it is green when put in place, will change in depth as it seasons in the building. These principles of shrinkage also explain why an edge-grain or quarter-sawed floor is less likely to open up than a flat-grain floor.

Density

The tree undergoes a considerable impetus early every spring and grows very rapidly for a short time. Large amounts of water are carried through the cells to the rapidly growing branches and leaves at the top of the tree. This water passes upward mainly in the outer layers of the tree. The result is that the cells next to the bark (which are formed during the period of rapid growth) have thin walls and large passages. Later on, during the summer, the rate of growth slows and the demand for water is less. The cells that are formed during the summer have much thicker walls and much smaller pores. Thus,

a year's growth forms two types of wood: the *springwood* (characterized by softness and openness of grain) and the *summerwood* (characterized by hardness and closeness, or density, of grain). The springwood and summer wood growth for one year is called an *annual ring*.

Density and Strength

There is one ring for each year of growth. This development of springwood and summerwood is a marked characteristic of practically all woods that grow in a temperate climate. It is evident in such trees as the yellow pines and firs and less so in the white pines, maple, and the like. Careful examination will reveal this annual ring, however, in practically all species. It follows, therefore, that a tree in which the dense summerwood predominates is stronger than one in which the soft spring wood predominates. This is a point that should be kept in mind when selecting material for important members such as girders and posts carrying heavy loads. The strength of wood of the same species varies markedly with the density. For example, Douglas fir or southern pine, carefully selected for density, is one-sixth stronger than lumber of the same species and knot limitations in which the springwood predominates. Trees having approximately one-third or more of cross-sectional area in summerwood fulfill one of the requirements for structural timbers.

Estimating Density

It must be remembered that the small cells or fibers that make up the wood structure are hollow. Wood substance itself has a specific gravity of about 1.5. That means it will sink in water. It is stated that wood substance of all species is practically of the same density. Strength of wood depends upon its density and varies with its density. The actual dry weight of lumber is a good criterion of its strength, although weight cannot always be relied upon as a basis for determining strength, because other important factors frequently must be considered in a specific piece of wood.

The hardness of wood is also another factor that assists in estimating the strength of wood. A test sometimes used is cutting across the grain. This test cannot be utilized in the commercial grading of lumber because moisture content will affect the hardness and because hardness, thus measured, cannot be adequately defined. The annual rings found in practically all species are an important consideration in estimating density, although the annual rings indicate different conditions in different species. In ring-porous hardwoods and in the conifers, where the contrast between springwood and summerwood

is definite, the proportion of hard summerwood is an indication of the strength of the individual piece of wood. The amount of summerwood, however, cannot always be relied upon as an indication of strength because summerwood itself varies in density. When cut across the grain with a knife, the density of summerwood may be estimated based on hardness, color, and luster.

In conifers, annual rings of average width indicate denser material or a larger proportion of summerwood than in wood with either wide or narrow rings. In some old conifers of virgin growth in which the more recent annual rings are narrow, the wood is less dense than where there has been normal growth. On the other hand, in young trees where the growth has not been impeded by other trees, the rings are wider and in consequence the wood less dense. These facts may account for the belief that all second-growth timber and all sapwood are weak. In analyzing wood for density, the contrast between summerwood and springwood should be pronounced.

Oak, ash, hickory, and other ring-porous hardwoods in general rank high in strength when the annual rings are wide. In this respect, they contrast with conifers. These species have more summerwood than springwood as the rings become wider. For this reason, oak, hickory, ash, and elm of second growth are considered superior because of fast growth and increase in proportion of summerwood. These conditions do not always exist, however, because exceptions occur, especially in ash and oak, where, although the summerwood is about normal, it may not be dense or strong. Very narrow rings in ring-porous hardwoods are likely to indicate weak and brash material composed largely of spring wood with big pores. Maple, birch, beech, and other diffuse-porous hardwoods in general show no definite relationship between the width of rings and density, except that usually narrow rings indicate brash wood.

Strength

Wood, when used in ordinary structures, is called upon to have three types of strength: tension, compression, and shear.

Tension

Tension is the technical term for a pulling stress. For example, if two people are having a tug of war with a rope, the rope is in tension. The tensile strength of wood, especially of the structural grades, is very high.

Compression

If, however, the people at opposite ends of a 2 × 4 are trying to push each other over, the timber is in *compression*. Tension and compression represent, therefore, exactly opposite forces.

Shear

If two or three planks are placed one upon the other between two blocks, and a person were to stand in the middle, the planks would bend (see Figure 2-8). It will be noted that at the outer ends the boards tend to slip past each other.

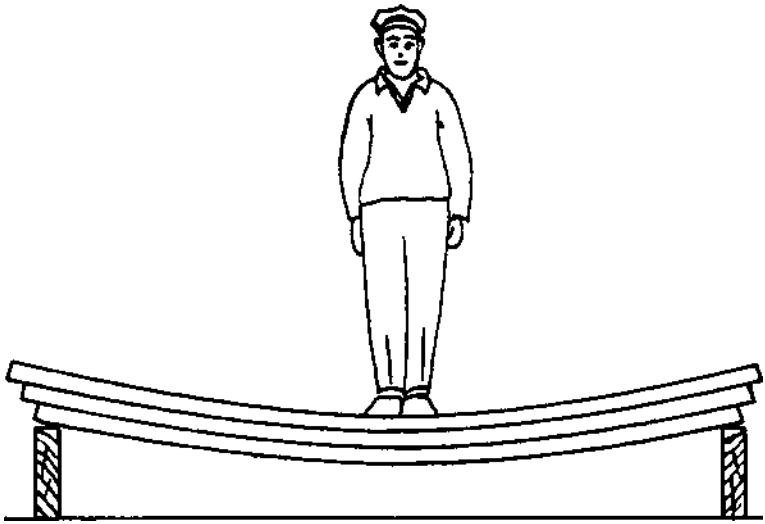

Figure 2-8 Illustrating shear in lumber.

If the planks were securely spiked through from top to bottom, the slipping would be in a great measure prevented and the boards would act more as one piece of wood. In very solid timber, there is the same tendency for the various parts of the piece to slip past each other. This tendency is called *horizontal shear*. A defect (such as a check, which runs horizontally through a piece of a timber and tends to separate the upper from the lower part) is a weakness in shear.

Suppose that the planks were spiked through at the center of span only (that is, halfway between the blocks). Such spikes would not increase the stiffness of the planks. It is clear, therefore, that there is no horizontal shear near the center of the span (see Figure 2-9), and that the shear increases as one approaches either end of the beam. This will explain why, as most carpenters have doubtlessly observed, steel stirrups are used in concrete beams (weak in shear), why there is usually none near the center, and why they are put closer and closer together near the ends of the beams.

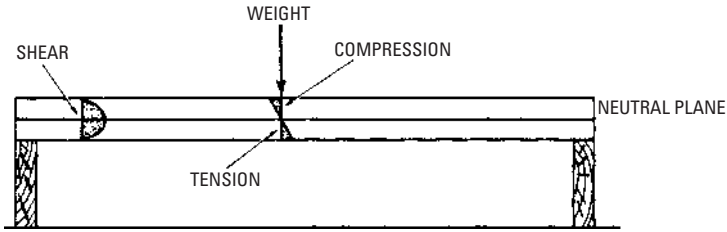

Figure 2-9 Illustrating the different proportions of tension, compression, and shear.

For all practical purposes, the compressive strength of wood may be considered to equal its tensile strength. It has been extremely difficult to make any direct measurements of the tensile strength of wood. In an experiment designed to ascertain the tensile strength of a specimen of wood, a 4-inch × 4-inch piece was selected. A portion about a foot in length near the center was carefully cut down on all four sides until it was exactly 3/4-inch square. The test specimen was placed in a machine that gripped the 4-inch × 4-inch ends securely and a pull was exerted. The specimen did not pull apart. The 3/4-inch-square section held and actually pulled out of the end of the 4-inch × 4-inch, leaving a 3/4-inch square hole. This is an excellent illustration of how a piece may fail from shear rather than tension, the shear in this case being insufficient to prevent the 3/4-inch-square piece from pulling out.

Deadwood

Because in some instances persons are prejudiced against the use of timber cut from dead trees, it is customary for individuals to specify that only timber cut from live trees will be accepted. It is true, however, that when sound trees that are dead are sawed into lumber and the weathered or charred outside is cut away, the resulting lumber cannot be distinguished from that coming from live trees except insofar as the lumber from dead trees may be somewhat seasoned at the time it is sawed. It must be remembered that the heartwood of a living tree is fully matured and that in the sapwood only a small portion of the cells are in a living condition. Consequently, most of the wood cut from trees is already dead, even when the tree itself is considered alive.

For structural purposes, it may be said that lumber cut from fire- or insect-killed trees is just as good as any other lumber, unless the wood has been subjected to further decay or insect attack.

Virgin and Second Growth

Occasionally an order calls for lumber of either virgin growth or second growth. The terms, however, are without significance, as an individual cannot tell one type from the other when it is delivered.

The *virgin growth* (which is also called *old growth* or *first growth*) refers to timber that grows in the forest along with many other trees, and, therefore, has suffered the consequence of the fight for sunlight and moisture.

The *second growth* is considered as that timber that grows up with less of the competition for sunlight and moisture that characterizes first-growth timber.

Because of environment, the virgin growth is usually thought of as wood of slow-growing type, whereas the second growth is considered as of relatively rapid growth, evidenced by wider annual rings. In such hardwoods as ash, hickory, elm, and oak, the wider annual rings are supposed to indicate stronger and tougher wood, whereas in the conifers such as pine and fir, this condition is supposed to result in a weaker and brasher wood. For this reason, where the strength and toughness are desired, the second growth is preferred among hardwoods, and virgin growth is desired in conifers. This is because of the following:

- The variety of conditions under which both virgin and second growth timbers grow
- Virgin growth may have the characteristics of second growth
- Second growth may have the characteristics of virgin timber

It is advisable in judging the strength of wood to rely upon its density and rate of growth, rather than upon its being either virgin or second growth.

Time of Cutting Timber

The time when timber is cut has very little to do with its durability or other desirable properties if, after it is cut, it is cared for properly. Timber cut in the late spring, however, or early summer is more likely to be attacked by insects and fungi. In addition, seasoning will proceed much more rapidly during the summer months and, therefore, will result in checking, unless the lumber is shaded from the intense sunlight. There is practically no difference in the moisture content in green lumber cut during the summer or winter.

Air-Dried and Kiln-Dried Wood

There is a prevailing misapprehension that air-dried lumber is stronger or better than kiln-dried lumber. Exhaustive tests have conclusively shown that good kiln-drying and good air-drying have

exactly the same results on the strength of the wood. Wood increases in strength with the elimination of moisture content. This may account for the claim that kiln-dried lumber is stronger than air-dried lumber. This has little significance because, in use, wood will come to practically the same moisture content whether it has been kiln-dried or air-dried.

The same kiln-drying process cannot be applied to all species of wood. Consequently, it must be remembered that lack of certain strength properties in wood may be because of improper kiln-drying. Similar damage also may result from air seasoning under unsuitable conditions.

Sapwood Versus Heartwood

The belief is common that in some species the heartwood is stronger than the sapwood, and that the reverse is the case in such species as hickory and ash. Tests have shown conclusively that neither is the case, sapwood is not necessarily stronger than heartwood or heartwood stronger than sapwood, but that density (rather than other factors) makes the difference in strength. In trees that are mature, the sapwood is frequently weaker, whereas in young trees the sapwood may be stronger. Density, proportion of springwood and summerwood, then must be the basis of consideration of strength rather than whether the wood is sapwood or heartwood. Under unfavorable conditions, the sapwood of most species is more subject to decay than the heartwood.

The cells in the heartwood of some species are filled with various oils, tannins, and other substances (called *extratives*) that make these timbers rot-resistant. There is practically no difference in the strength of heartwood and sapwood if they weigh the same. In most species, only the sapwood can be readily impregnated with preservatives, a procedure used when the wood will be in contact with the ground.

Defects and Deterioration

The defects found in manufactured lumber have several causes. Those defects found in the natural log include the following:

- Shakes
- Knots
- Pitchpockets

Those caused by deterioration include the following:

- Rot
- Dote

Those caused by imperfect manufacture include the following:

- Imperfect machining
- Wane
- Machine burn
- Checks and splits from imperfect drying

Heart shakes (see Figure 2-10) are radial cracks that are wider at the pith of the tree than at the outer end. This defect is more commonly found in old trees rather than in young vigorous saplings. It occurs frequently in hemlock.

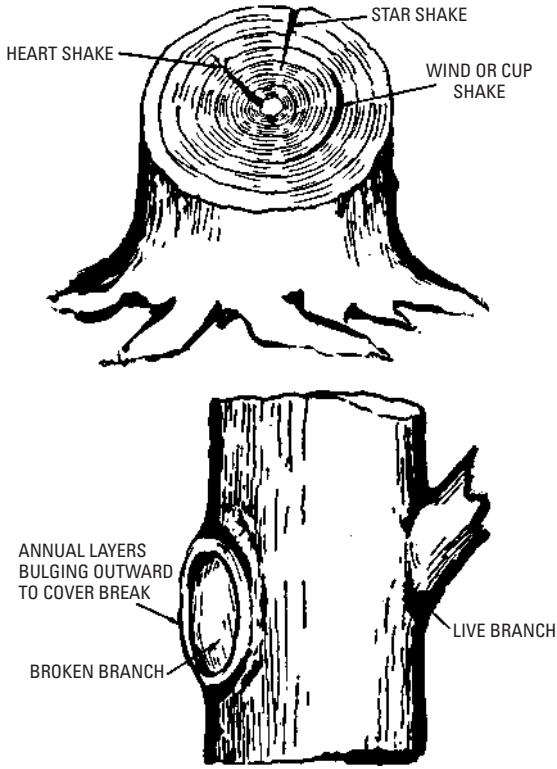

Figure 2-10 The causes of some lumber defects are found in the natural log.

A *wind shake* or *cup shake* is a crack following the line of the porous part of the annual rings and is curved by a separation of

the annual rings (see Figure 2-10). A wind shake may extend for a considerable distance up the trunk. Other explanations for wind shakes are expansion of the sapwood and wrenching from high wind (hence the name). Brown ash is especially susceptible to wind shake. Wind shakes cause cup checks in lumber (see Figure 2-11A). A *star shake* resembles a wind shake but differs from it in that the crack extends across the center of the trunk without any appearance of decay at that point; it is larger at the outside of the tree. Heart and star shakes cause splits in lumber (see Figure 2-11B).

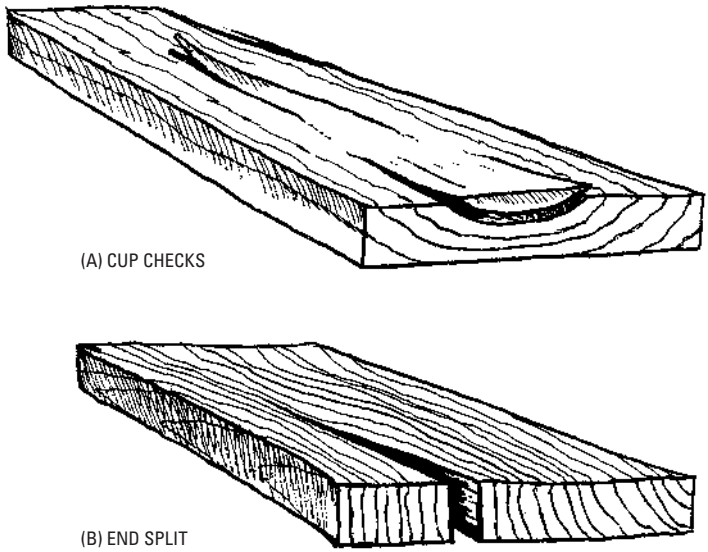

Figure 2-11 Lumber defects caused by defects in the log. (A) Cup checks are caused by wind shakes. (B) End splits are caused by star and heart shakes. *(Courtesy Practical Restoration Reports)*

Figure 2-12 shows some common causes of black or loose knots. Figures 2-13 and 2-14 show some results of improper drying practices.

Decay of Wood

Decay of lumber is the result of one cause and one cause only: the work of certain low-order plants called *fungi*. All of these organisms require water, air, and temperatures well above freezing to live, grow, and multiply. Consequently, wood that is kept dry, or that is dried quickly after wetting, will not decay.

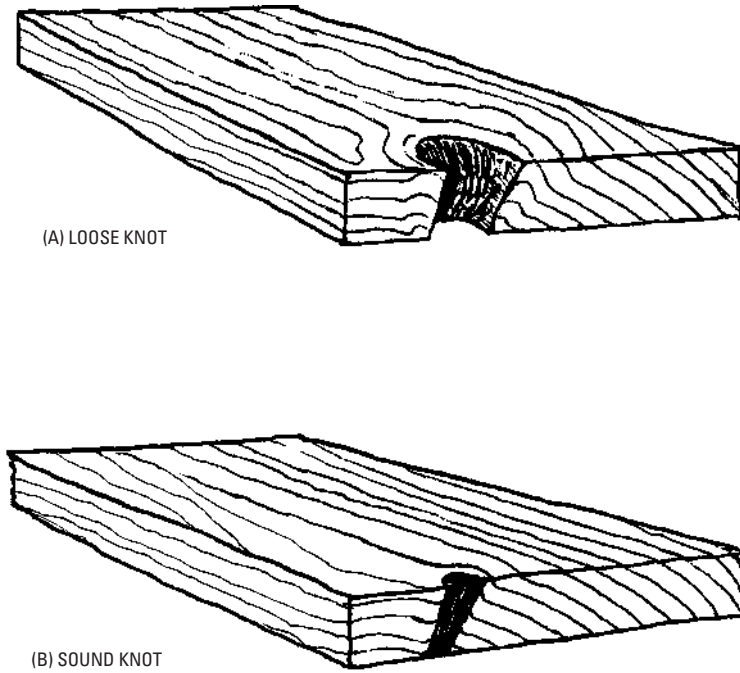

Figure 2-12 Black or loose knots (A) are caused by broken branches in the log. Bulging layers of growth cause a large swirl of cross grain at the surface of the board surrounding the knot. Tight red knots (B) are caused by a live branch in the log. The knot will not fall out and grain is relatively straight around the knot. *(Courtesy Practical Restoration Reports)*

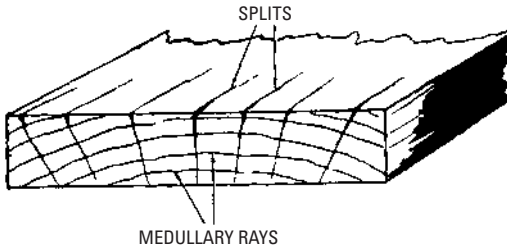

Figure 2-13 A board with splits along the medullary rays. This condition is caused by too-rapid kiln-drying.

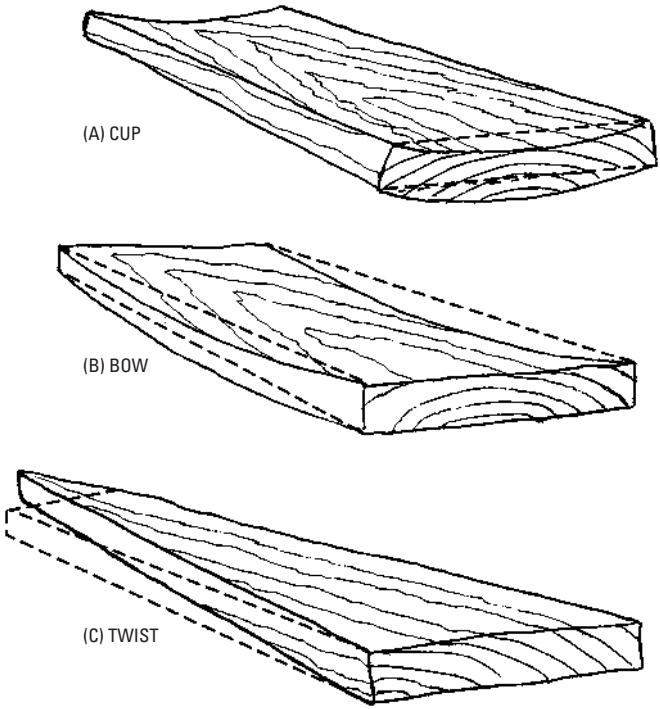

Figure 2-14 Warping is frequently caused by improper drying practices. Here the flatness of boards is distorted by irregular shrinkage. *(Courtesy Practical Restoration Reports)*

Further, if it is kept continuously submerged in water even for long periods, it is not decayed significantly by the common decay fungi regardless of the wood species or the presence of sapwood. Bacteria and certain soft-rot fungi can attack submerged wood, but the resulting deterioration is very slow. A large proportion of wood in use is kept so dry at all times that it lasts indefinitely. Moisture and temperature (which vary greatly with local conditions) are the principal factors affecting rate of decay. When exposed to conditions that favor decay, wood deteriorates more rapidly in warm, humid areas than in cool or dry areas. High altitudes, as a rule, are less favorable to decay than low altitudes because the average temperatures are lower and the growing seasons for fungi that cause decay are shorter.

The heartwoods of some common native species of wood have varying degrees of natural decay resistance. Untreated sapwood of substantially all species has low resistance to decay and usually has a short service life under decay-producing conditions. The decay resistance of heartwood is greatly affected by differences in the preservative qualities of the wood extractives, the attacking fungus, and the conditions of exposure. Considerable difference in service life may be obtained from pieces of wood cut from the same species, or even from the same tree, and used under apparently similar conditions. There are further complications because, in a few species, such as the spruces and the true firs (not Douglas fir), heartwood and sapwood are so similar in color that they cannot be easily distinguished. Marketable sizes of some species, such as southern pine and baldcypress, are becoming largely second growth. They contain a high percentage of sapwood.

Precise ratings of the decay resistance of heartwood of different species are not possible because of differences within species and the variety of service conditions to which wood is exposed. However, broad groupings of many of the native species, based on service records, laboratory tests, and general experience, are helpful in choosing heartwood for use under conditions favorable to decay (see Table 2-1). The extent of variations in decay resistance of individual trees or wood samples of a species is much greater for most of the more resistant species than for the slightly or nonresistant species.

Where decay hazards exist, heartwood of species in the resistant or very resistant category generally gives satisfactory service. However, heartwood of species in the other two categories will usually require some form of preservative treatment. For mild decay conditions, a simple preservative treatment (such as a short soak in preservative after all cutting and boring operations are complete) will be adequate for wood low in decay resistance. For more severe decay hazards, pressure treatments are often required. Even the very decay-resistant species may require preservative treatment for important structural or other uses where failure would endanger life or require expensive repairs.

Wood products sometimes are treated with preservative or fire-retarding salts (usually in water solution) to impart resistance to decay or fire. Such products generally are kiln-dried after treatment. Mechanical properties are essentially unchanged by preservative treatment.

Properties are affected to some extent by the combined effects of fire-retardant chemicals, treatment methods, and kiln-drying. A

Table 2-1 Grouping of Some Domestic Woods According to Heartwood Decay

Resistant or Very Resistant	*Moderately Resistant*	*Slightly or Nonresistant*
Baldcypress (old growth)*	Baldcypress (young growth)*	Alder
Catalpa	Douglas fir	Ashes
Cedars	Honey locust	Aspens
Cherry, black	Larch, western	Basswood
Chestnut	Oak, swamp chestnut	Beech
Cypress, Arizona	Pine, eastern white*	Birches
Junipers	Southern pine:	Buckeye
Locust, black†	Longleaf*	Butternut
Mesquite	Slash*	Cottonwood
Mulberry, red†	Tamarack	Elms
Oak:		Hackberry
Bur		Hemlocks
Chestnut		Hickories
Gambel		Magnolia
Oregon white		Maples
Post		Oak (red and black species)
White		Pines (other than long-leaf, slash, and eastern white)
Osage orange†		Poplars
Redwood		Spruces
Sassafras		Sweetgum
Walnut, black		True firs (western and eastern)
Yew, Pacific†		Willows
		Yellow poplar

*The southern and eastern pines and baldcypress are now largely second growth with a large proportion of sapwood. Consequently, substantial quantities of heartwood lumber of these species are not available.
†These woods have exceptionally high decay resistance.

variety of fire-retardant treatments have been studied. Collectively the studies indicate modulus of rupture, work to maximum load, and toughness are reduced by varying amounts depending on species and type of fire retardant. Work to maximum load and toughness are most affected, with reductions of as much as 45 percent. A reduction in modulus of rupture of as much as 20 percent has been observed. A design reduction of 10 percent is frequently used. Stiffness is not appreciably affected by fire-retardant treatments.

Wood is also sometimes impregnated with monomers (such as methyl methacrylate) that are subsequently polymerized. Many of the properties of the resulting composite are better. Generally, this is a result of filling the void spaces in the wood structure with plastic. The polymerization process and both the chemical nature and quantity of monomers are variables that influence composite properties.

Molding and Staining Fungi

Molding and *staining fungi* do not seriously affect most mechanical properties of wood because they feed upon substance within the structural cell wall rather than on the structural wall itself. Specific gravity may be reduced by from 1 to 2 percent, whereas most of the strength properties are reduced by a comparable or only slightly greater extent. Toughness or shock resistance, however, may be reduced by up to 30 percent. The duration of infection and the species of fungi involved are important factors in determining the extent of weakening.

Although molds and stains themselves often do not have a major effect on the strength of wood products, conditions that favor the development of these organisms are likewise ideal for the growth of wood-destroying (decay) fungi, which can greatly reduce mechanical properties.

Fungal Decay

Unlike the molding and staining fungi, the *wood-destroying (decay) fungi* seriously reduce strength. Even what appears to be sound wood adjacent to obviously decayed parts may contain hard-to-detect, early (incipient) decay that is decidedly weakening, especially in shock resistance.

Not all wood-destroying fungi affect wood in the same way. The fungi that cause an easily recognized pitting of the wood, for example, may be less injurious to strength than those that, in the early stages, give a slight discoloration of the wood as the only visible effect.

No method is known for estimating the amount of reduction in strength from the appearance of decayed wood. Therefore, when strength is an important consideration, the safe procedure is to discard every piece that contains even a small amount of decay. An exception may be pieces in which decay occurs in a knot but does not extend into the surrounding wood.

Blue Stain

In the sapwood of many species of both softwoods and hardwoods, there often develops a bluish-black discoloration known as *blue stain*. It does not indicate an early stage of decay, nor does it have any practicable effect on the strength of the wood. Blue stain is caused by a fungus growth in unseasoned lumber. Although objectionable where appearance is of importance, as in unpainted trim, blue stain need cause no concern for framing lumber. Precautions should be taken, however, to make sure that no decay fungus is present with the blue stain.

Nuclear Radiation

Very large doses of gamma rays or neutrons can cause substantial degradation of wood. In general, irradiation with gamma rays in doses up to about 1 megarad has little effect on the strength properties of wood. As dosage increases above 1 megarad, tensile strength parallel to grain and toughness decrease. At a dosage of 300 megarads, tensile strength is reduced about 90 percent. Gamma rays also affect compressive strength parallel to grain above 1 megarad, but strength losses with further dosage are less than for tensile strength. Only about one-third of the compressive strength is lost when the total dose is 300 megarads. Effects of gamma rays on bending and shear strength are intermediate between the effects on tensile and compressive strength.

Weathering

Without protective treatment, freshly cut wood exposed to the weather changes materially in color. Other changes due to weathering include warping, loss of some surface fibers, and surface roughening and checking. The effects of weathering on wood may be desirable or undesirable, depending on the requirements for the particular wood product. The time required to reach the fully weathered appearance depends on the severity of the exposure to sun and rain. Once weathered, wood remains nearly unaltered in appearance.

The color of wood is affected very soon on exposure to weather. With continued exposure, all woods turn gray. However, only the wood at or near the exposed surfaces is noticeably affected. This very

thin gray layer is composed chiefly of partially degraded cellulose fibers and microorganisms. Further weathering causes fibers to be lost from the surface, but the process is so slow that only about 1/4 inch is lost in a century.

In the weathering process, chemical degradation is influenced greatly by the wavelength of light. The most severe effects are produced by exposure to ultraviolet light. As cycles of wetting and drying take place, most woods develop physical changes, such as checks or cracks that are easily visible. Moderate to low-density woods acquire fewer checks than do high-density woods. Vertical-grain boards check less than flat-grain boards.

Because of weathering, boards tend to warp (particularly cup) and pull out their fastenings. The cupping tendency varies with the density, width, and thickness of a board. The greater the density and the greater the width in proportion to the thickness, the greater is the tendency to cup. Warping also is more pronounced in flat-grain boards than in vertical-grain boards (see Figure 2-15). For best cup resistance, the width of a board should not exceed eight times its thickness.

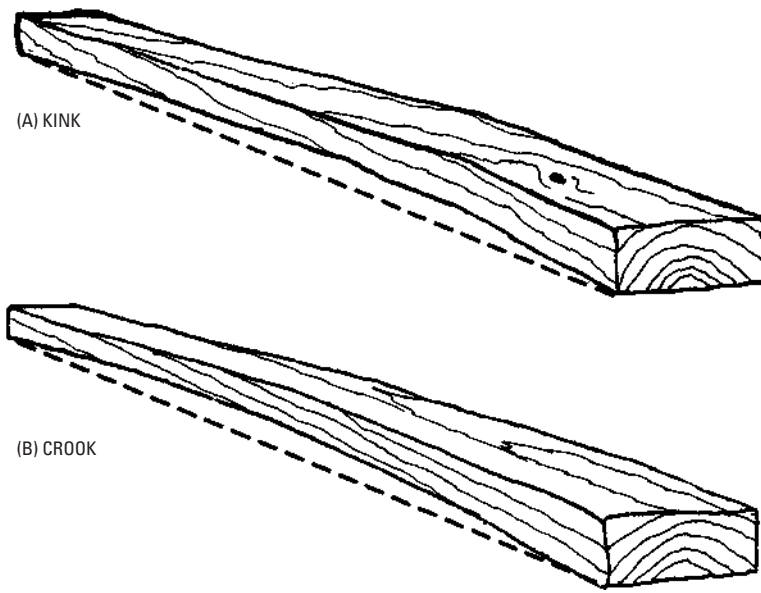

Figure 2-15 Warping can also affect straightness of a board.
(Courtesy Practical Restoration Reports)

Biological attack of a wood surface by microorganisms is recognized as a contributing factor to color changes. When weathered wood has an unsightly dark gray and blotchy appearance, it is caused by dark-colored fungal spores and mycelium on the wood surface. The formation of a clean, light gray, silvery sheen on weathered wood occurs most frequently where microorganism growth is inhibited by a hot, arid climate or where there is a salt atmosphere, such as in coastal regions.

The contact of fasteners and other metallic products with the weathering wood surface is a source of color, often undesirable if a natural color is desired.

Insect Damage

Insect damage may occur in standing trees, logs, and unseasoned or seasoned lumber. Damage in the standing tree is difficult to control, but otherwise insect damage can be largely eliminated by proper control methods.

Insect holes are generally classified as *pinholes*, *grub holes*, and *powder postholes*. The powder post larvae, by their irregular burrows, may destroy most of the interior of a piece, although the surface shows only small holes, and the strength of the piece may be reduced virtually to zero.

No method is known for estimating the amount of reduction in strength from the appearance of insect-damaged wood, and, when strength is an important consideration, the safe procedure is to eliminate pieces containing insect holes.

Summary

Wood is the most versatile building material. Softwoods and hardwoods grow as trees with a fibrous cellular structure. Logs are converted into lumber and seasoned by drying. The physical characteristics of wood affect its performance and specific uses. Defects in lumber relate to defects in the logs it was cut from and how the lumber was handled during drying and storage. After wood is installed in a building, it can deteriorate by weathering, decay, or insects unless protected.

Review Questions

1. Describe the physical structure of wood with a drawing. Label the parts.
2. What are the characteristics of softwoods and hardwoods? How is each type of wood used in building?

3. What is the difference between vertical-grain wood and flat-grain wood? How is each produced?
4. Why does wood have to be seasoned? How does the moisture content and size of wood change during seasoning?
5. How does shear affect the strength of wood?
6. What substances make some woods decay-resistant?
7. How do defects in logs relate to defects in lumber? Give some examples.
8. What causes decay in wood? How can you stop decay?
9. How does weathering affect the color of wood?
10. How long would it take for a 1-inch thick board to weather away by fiber loss?
11. Would vertical-grain or flat-grain boards hold up better on an outdoor deck? Why?

Chapter 3

Lumber, Plywood, and Other Wood Products

The basic construction material in carpentry is lumber. There are many kinds of lumber varying greatly in structural characteristics. Here, we deal with the lumber common to construction carpentry (see Figure 3-1).

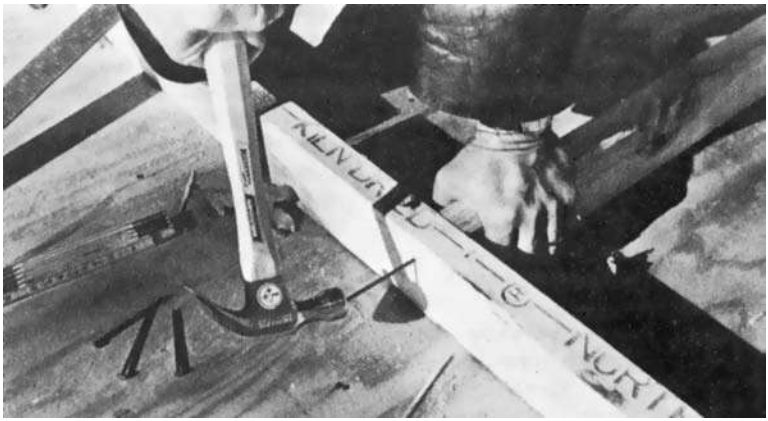

Figure 3-1 Construction lumber. Note the classification kiln-dried on one member. This stock will be used inside a structure.

(Courtesy of Vaughn & Bushnell)

Standard Sizes of Lumber

Lumber is usually sawed into standard lengths, widths, and thicknesses. This permits uniformity in planning structures and in ordering material (see Table 3-1). Standards have been established for dimension differences between *nominal size* and the *standard size*. It is important that these dimension differences be taken into consideration when planning a structure. A good example of the dimension difference may be illustrated by the common 2 × 4. As may be seen in the table, the familiar quoted size (2 × 4) refers to a rough or nominal dimension, but the actual standard size to which the lumber is dressed is $1^1/_2$ inches × $3^1/_2$ inches.

Table 3-1 Your Guide to New Sizes of Lumber

Standard (What You Order)	*Nominal (What You Get)*	
	*Dry or Seasoned**	*Green or Unseasoned*†
1 × 4	$^{3}/_{4} \times 3^{1}/_{2}$	$^{25}/_{32} \times 3^{9}/_{16}$
1 × 6	$^{3}/_{4} \times 5^{1}/_{2}$	$^{25}/_{32} \times 5^{5}/_{8}$
1 × 8	$^{3}/_{4} \times 7^{1}/_{4}$	$^{25}/_{32} \times 7^{1}/_{2}$
1 × 10	$^{3}/_{4} \times 9^{1}/_{4}$	$^{25}/_{32} \times 9^{1}/_{2}$
1 × 12	$^{3}/_{4} \times 11^{1}/_{4}$	$^{25}/_{32} \times 11^{1}/_{2}$
2 × 4	$1^{1}/_{2} \times 3^{1}/_{2}$	$1^{9}/_{16} \times 3^{9}/_{16}$
2 × 6	$1^{1}/_{2} \times 5^{1}/_{2}$	$1^{9}/_{16} \times 5^{5}/_{8}$
2 × 8	$1^{1}/_{2} \times 7^{1}/_{4}$	$1^{9}/_{16} \times 7^{1}/_{2}$
2 × 10	$1^{1}/_{2} \times 9^{1}/_{4}$	$1^{9}/_{16} \times 9^{1}/_{2}$
2 × 12	$1^{1}/_{2} \times 11^{1}/_{4}$	$1^{9}/_{16} \times 11^{1}/_{2}$
4 × 4	$3^{1}/_{2} \times 3^{1}/_{2}$	$3^{9}/_{16} \times 3^{9}/_{16}$
4 × 6	$3^{1}/_{2} \times 5^{1}/_{2}$	$3^{9}/_{16} \times 5^{5}/_{8}$
4 × 8	$3^{1}/_{2} \times 7^{1}/_{4}$	$3^{9}/_{16} \times 7^{1}/_{2}$
4 × 10	$3^{1}/_{2} \times 9^{1}/_{4}$	$3^{9}/_{16} \times 9^{1}/_{2}$
4 × 12	$3^{1}/_{2} \times 11^{1}/_{4}$	$3^{9}/_{16} \times 11^{1}/_{2}$

*19 percent moisture content or under.
†More than 19 percent moisture content.

Softwood Lumber Grades for Construction

The grading requirements of construction lumber are related specifically to the major construction uses intended, and little or no further grading occurs once the piece leaves the sawmill. Construction lumber can be placed in three general categories—stress-graded, nonstress-graded, and appearance lumber. *Stress-graded* and *nonstress-graded lumber* are employed where the structural integrity of the piece is the primary requirement. *Appearance lumber* encompasses those lumber products in which appearance is of primary importance (structural integrity, although sometimes important, is a secondary feature).

Stress-Graded Lumber

Almost all softwood lumber nominally 2 to 4 inches thick is stress-graded. Lumber of any species and size, as it is sawed from the log, is quite variable in its mechanical properties. Pieces may differ in strength by several hundred percent. For simplicity and economy in

use, pieces of lumber of similar mechanical properties can be placed in a single class called a stress-grade.

Visual grading is the oldest stress-grading method. It is based on the premise that mechanical properties of lumber differ from mechanical properties of clear wood because of characteristics that can be seen and judged by eye. These visual characteristics are used to sort the lumber into stress grades (see Table 3-2). The following are major visual sorting criteria:

- Density
- Decay
- Heartwood and sapwood
- Slope of grain
- Knots
- Shake
- Checks and splits
- Wane
- Pitch pockets

Nonstress-Graded Lumber

Traditionally, much of the lumber intended for general building purposes with little or no remanufacture has not been assigned allowable properties (stress-graded). This category of lumber has been referred to as *yard lumber*. However, the assignment of allowable

Table 3-2 Visual Grades Described in the National Grading Rule*

Lumber Classification	*Grade Name*
Light framing (2 to 4 inches thick, 4 inches wide)†	Construction Standard Utility
Structural light framing (2 to 4 inches thick, 2 to 4 inches wide)	Select structural, 1, 2, 3
Studs (2 to 4 inches thick, 2 to 4 inches wide)	Stud
Structural joists and planks (2 to 4 inches thick, 6 inches and wide)	Select structural, 1, 2, 3
Appearance framing (2 to 4 inches thick, 2 to 4 inches wide)	Appearance

*Sizes shown are nominal.
†Widths narrower than 4 inches may have different strength ratio.

properties to an increasing number of former yard items has diluted the meaning of the term yard lumber.

In nonstress-graded structural lumber, the section properties (shape, size) of the pieces combine with the visual grade requirements to provide the degree of structural integrity intended. Typical nonstress-graded items include boards, lath, battens, crossarms, planks, and foundation stock.

Boards (sometimes referred to as *commons*) are one of the more important nonstress-graded products. Common grades of boards are suitable for construction and general utility purposes. They are separated into three to five different grades depending upon the species and lumber manufacturing association involved. Grades may be described by number (No. 1, No. 2) or by descriptive terms (Construction, Standard).

Since there are differences in the inherent properties of the various species and in corresponding names, the grades for different species are not always interchangeable in use. First-grade boards are usually graded primarily for serviceability, but appearance is also considered. This grade is used for such purposes as siding, cornice, shelving, and paneling. Features such as knots and knotholes are permitted to be larger and more frequent as the grade level becomes lower. Second- and third-grade boards are often used together for such purposes as subfloors, roof and wall sheathing, and rough concrete work. Fourth-grade boards are not selected for appearance but for adequate strength. They are used for roof and wall sheathing, subfloor, and rough concrete formwork.

Grading provisions for other nonstress-graded products vary by species, product, and grading association. Lath, for example, is available generally in two grades, No. 1 and No. 2. One grade of batten is listed in one grade rule and six in another.

Appearance Lumber

Appearance lumber often is nonstress-graded. However, it forms a separate category because of the distinct importance of appearance in the grading process. This category of construction lumber includes most lumber worked to a pattern. Secondary manufacture on these items is usually restricted to onsite fitting, such as cutting to length and mitering. There is an increasing trend toward prefinishing many items. The appearance category of lumber includes trim, siding, flooring, ceiling, paneling, casing, base, stepping, and finish boards. Finish boards are commonly used for shelving and built-in cabinetwork.

Most appearance lumber grades are described by letters and combinations of letters (B&BTR, C&BTR, D), as shown in Table 3-3. Appearance grades are also often known as Select grades. Descriptive terms such as prime and clear are applied to a limited number of species. The specification FG (flat grain), VG (vertical grain), or MG (mixed grain) is offered as a purchase option for some appearance lumber products. In cedar and redwood (where there is a pronounced difference in color between heartwood and sapwood, and where heartwood has high natural resistance to decay), grades of heartwood are denoted as heart. In some species and products, two (or, at most, three) grades are available. A typical example is casing and base in the grades of C&BTR and D in some species and in B&BTR, C, C&BTR, and D in other species. Although several grades may be described in grade rules, often fewer are offered on the retail market.

Table 3-3 Standard Lumber Abbreviations

Abbreviation	*Meaning*
AD	Air-dried
ALS	American Lumber Standard
AST	Antistain treated. At ship tackle (western softwoods)
AV or avg	Average
AW&L	All widths and lengths
B1S	See EB1S, CB1S, and E&CB1S
B2S	See EB2S, CB2S, and E&CB2S
B&B, B&BTR	B and Better
B&S	Beams and stringers
BD	Board
BD FT	Board feet
BDL	Bundle
BEV	Bevel or beveled
BH	Boxed heart
BM	Board measure
BSND	Bright sapwood no defect
BTR	Better
c	Allowable stress in compression in pounds per square inch
CB	Center beaded

(continued)

Table 3-3 *(continued)*

Abbreviation	*Meaning*
CB1S	Center bead on one side
CB2S	Center bead on two sides
cft or cu. ft.	Cubic foot or feet
CG2E	Center groove on two edges
CLG	Ceiling
CLR	Clear
CM	Center matched
Com	Common
CSG	Casing
CV	Center V
CV1S	Center V on one side
CV2S	Center V on two sides
DBClg	Double beaded ceiling (E&CB1S)
DB Part	Double beaded partition (E&CB2S)
DET	Double end trimmed
DF	Douglas-fir
DIM	Dimension
DKG	Decking
D/S,DS, D/Sdg	Drop siding
D1S,D2S	See S1S and S2S
D&M	Dressed and matched
D&CM	Dressed and center matched
D&SM	Dressed and standard matched
D2S&CM	Dressed two sides and center matched
D2S&SM	Dressed two sides and standard matched
E	Edge
EB1S	Edge bead one side
EB2S, SB2S	Edge bead on two sides
EE	Eased edges
EG	Edge (vertical or rift) grain
EM	End matched
EV1S, SV1S	Edge V one side
EV2S, SV2S.	Edge V two sides
E&CB1S	Edge and center bead one side
E&CB2S, DB2S, BC&2S	Edge and center bead two sides

(continued)

Table 3-3 (*continued*)

Abbreviation	*Meaning*
E&CV1S, DV1S, V&CV1S	Edge and center V one side
E&CV2S, DV2S, V&CV2S	Edge and center V two sides
f	Allowable stress in bending in pounds per square inch
FA	Facial area
FAS	Firsts and Seconds
FBM, Ft. BM	Feet board measure
FG	Flat or slash grain
FJ	Finger joint. End-jointed lumber using a finger joint configuration
FLG, Fig	Flooring
FOHC	Free of heart center
FOK	Free of knots
FT, ft	Foot or feet
FT.SM	Feet surface measure
G	Girth
GM	Grade marked
G/R	Grooved roofing
HB,H.B.	Hollow back
HEM	Hemlock
Hrt	Heart
H&M	Hit and miss
H or M	Hit or miss
IN, in.	Inch or inches
J&P	Joists and planks
JTD	Jointed
KD	Kiln-dried
LBR, Lbr	Lumber
LGR	Longer
LGTH	Length
Ltt, Lt	Lineal foot or feet
UN, Un	Lineal
LL	Longleaf
LNG, Lng	Lining
M	Thousand

(continued)

Table 3-3 (*continued*)

Abbreviation	*Meaning*
MBM, MBF, M.BM	Thousand (feet) board measure
MC, M.C.	Moisture content
MG	Medium grain or mixed grain
MLDG, Mldg	Molding
Mft	Thousand feet
MSR	Machine stress rated
N	Nosed
NBM	Net board measure
No.	Number
N1E or N2E	Nosed one or two edges
Ord	Order
PAD	Partially air dry
PART, Part	Partition
PAT, Pat	Pattern
Pes.	Pieces
PE	Plain end
PET	Precision end trimmed
P&T	Posts and timbers
P1S,P2S	See S1S and S2S
RDM	Random
REG, Reg	Regular
Rfg.	Roofing
RGH, Rgh	Rough
R/L, RL	Random lengths
R/W,RW	Random widths
RES	Resawn
SB1S	Single bead one side
SDG, Sdg	Siding
S-DRY	Surfaced dry. Lumber 19 percent moisture content or less per American Lumber Standard for softwood
SE	Square edge
SEL, Sel	Select or select grade
SE&S	Square edge and sound
SG	Slash or flat grain

(continued)

Table 3-3 *(continued)*

Abbreviation	*Meaning*
S-GRN	Surfaced green. Lumber unseasoned, in excess of 19 percent moisture content per American Lumber Standard for softwood
SGSSND	Sapwood, gum spots and streaks, no defect
SIT. SPR	Sitka spruce
S/L,SL, S/Lap	Shiplap
STD.M	Standard matched
SM	Surface measure
Specs	Specifications
SO	Square
SORS	Squares
SR	Stress-rated
STD, Std	Standard
Std. lgths.	Standard lengths
SSND	Sap stain no defect (stained)
STK	Stock
STPG	Stepping
STR, STRUCT	Structural
SYP	Southern yellow pine
S&E	Side and edge (surfaced on)
S1E	Surfaced one edge
S2E	Surfaced two edges
S1S	Surfaced one side
S2S	Surfaced two sides
S4S	Surfaced four sides
S1S&CM	Surfaced one side and center matched
S2S&CM	Surfaced two sides and center matched
S4S&CS	Surfaced four sides and calking seam
S1S1E	Surfaced one side, one edge
S1S2E	Surfaced one side, two edges
S2S1E	Surfaced two sides, one edge
S2S&SL	Surfaced two sides and shiplapped
S2S&SM	Surfaced two sides and standard matched
t	Allowable stress in tension in pounds per square inch
TBR	Timber
T&G	Tongued and grooved

(continued)

Table 3-3 (*continued*)

Abbreviation	*Meaning*
VG	Vertical (edge) grain
V1S	See EV1S, CV1S, and E&CV1S
V2S	See EV2S, CV2S, and E&CV2S
WCH	West Coast hemlock
WDR, wdr	Wider
WHAD	Worm holes a defect
WHND	Worm holes no defect
WT	Weight
WTH	Width
WRD	Western redcedar
yp	Yellow pine

Courtesy Forest Products Laboratory

Grade B&BTR allows a few small imperfections, mainly in the form of minor skips in manufacture, small checks or stains caused by seasoning, and (depending on the species) small pitch areas, pin knots, or the like. Since appearance grades emphasize the quality of one face, the reverse side may be lower in quality. In construction, grade C&BTR is the grade combination most commonly available. It is used for high-quality interior and exterior trim, paneling, and cabinetwork, especially where these are to receive a natural finish. It is the principal grade used for flooring in homes, offices, and public buildings. In industrial uses, it meets the special requirements for large-sized, practically clear stock.

The number and size of imperfections permitted increase as the grades drop from B&BTR to D and E. Appearance grades are not uniform across species and products, however, and official grade rules must be used for detailed reference. Grade C is used for many of the same purposes as B&BTR, often where the best paint finish is desired. Grade D allows larger and more numerous surface imperfections that do not detract from the appearance of the finish when painted. Grade D is used in finish construction for many of the same uses as Grade C. It is also adaptable to industrial uses requiring short-length clear lumber.

Select Lumber

Select lumber is of good appearance and finished or dressed. See Table 3-4 for grade names, descriptions, and uses.

Table 3-4 Lumber Grades

Grade	*Description*	*Use*
Select:		
A	High quality, practically clear.	Suitable for natural finishes.
B	High quality, generally clear with a few minor defects.	Suitable for natural finishes.
C	Several minor defects.	Adapted to high-quality paint finish.
D	A few major defects.	Suitable for paint finishes.
Common:		
No. 1	Sound and tight-knotted.	Use without waste.
No. 2	Less restricted in quality than No. 1.	Framing, sheathing, structural forms where strain or stress not excessive.
No. 3	Permits some waste with defects larger than in No. 2	Footings, guardrails, rough subflooring.
No. 4	Permits waste, low quality, with decay and holes.	Sheathing, subfloors, roof boards in the cheaper types of construction.

Common Lumber

Common lumber is suitable for general construction and utility purposes and is identified by the grade names shown in Table 3-4.

Plywood

Plywood is a glued wood panel made up of relatively thin layers, or plies, with the grain of adjacent layers at an angle, usually 90° (see Figure 3-2). The usual constructions have an odd number of plies. The outside plies are called *faces* (or *face* and *back* plies), the inner plies are called *cores* (or *centers)*, and the plies immediately below the face and back are called *crossbands*. The core may be veneer, lumber, or particleboard. The plies may vary as to number, thickness, species, and grade of wood.

As compared with solid wood, the chief advantages of plywood are its having properties along the length nearly equal to properties along the width of the panel, its greater resistance to splitting, and its form, which permits many useful applications where large sheets

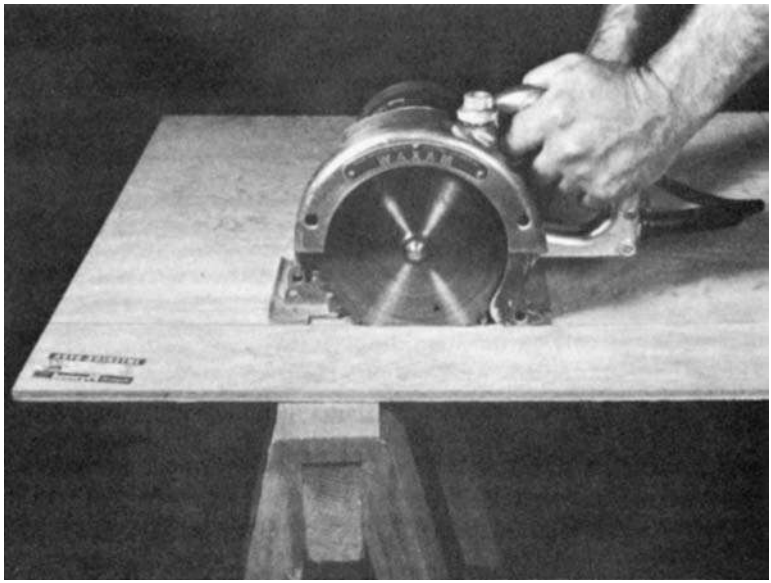

Figure 3-2 The multilayer construction of plywood gives large thin panels greater strength and stability than solid wood. *(Courtesy of the American Plywood Assn.)*

are desirable. Use of plywood may result in improved utilization of wood, because it covers large areas with a minimum amount of wood fiber. This is because it is permissible to use plywood thinner than sawn lumber in some applications.

The properties of plywood depend on the quality of the different layers of veneer, the order of layer placement in the panel, the glue used, and the control of gluing conditions in the gluing process. The grade of the panel depends on the quality of the veneers used, particularly of the face and back. The type of the panel depends on the glue joint, particularly its water resistance. Generally, face veneers with figured grain that are used in panels where appearance is important have numerous short, or otherwise deformed, wood fibers. These may significantly reduce strength and stiffness of the panels. On the other hand, face veneers and other plies may contain certain sizes and distributions of knots, splits, or growth characteristics that have no undesirable effects on strength properties for specific uses. Such uses include structural applications, such as sheathing for walls, roofs, or floors.

Types of Plywood

Broadly speaking, two classes of plywood are available: hardwood and softwood. In general, softwood plywood is intended for construction use and hardwood plywood for uses where appearance is important.

Originally, most softwood plywood was made of Douglas fir, but western hemlock, larch, white fir, ponderosa pine, redwood, southern pine, and other species are now used.

Most softwood plywood used in the United States is produced domestically, and U.S. manufacturers export some material. The bulk of softwood plywood is used where strength, stiffness, and construction convenience are more important than appearance. Some grades of softwood plywood are made with faces selected primarily for appearance and are used either with clear natural finishes or with pigmented finishes.

Hardwood plywood is made of many different species, both in the United States and overseas. More than half of all hardwood panels used in the United States are imported. Hardwood plywood is normally used where appearance is more important than strength. Most of the production is intended for interior or protected uses, although a very small proportion is made with glues suitable for exterior service. A significant portion of all hardwood plywood is available completely finished.

Plywood of thin, crossbanded veneers is very resistant to splitting and, therefore, nails and screws can be placed close together and close to the edges of panels.

Grades of Plywood

Plywood is graded according to defects on each surface and the type of glue used. Exterior or Interior plywood refers to the type of glue used to bond plies. Each face of the plywood has a letter grade—A, B, C, or D. Grade A means the face has no defects—it's perfect. Grade B means there are some defects (perhaps an area has a small patch). Grade C allows checks (splits) and small knotholes. Grade D allows large knotholes.

The commonly available grades are AC Interior, AD Interior, AC Exterior, and CDX, which is used for sheathing.

Particle Board

The group of materials generally classified as wood-base fiber and particle panel materials includes such familiar products as insulation boards, hardboards, particleboards, and laminated paperboards. In some instances, they are known by such proprietary names as

Masonite, Celotex, Insulite, and Beaver board or, in the instance of particleboards, by the kind of particle used, such as flakeboard, chipboard, or oriented strand board (OSB).

These panel materials are all reconstituted wood (or some other lignocellulose like bagasse) in that the wood is first reduced to small fractions and then put back together by special forms of manufacture into panels of relatively large size and moderate thickness. These board or panel materials in final form retain some of the properties of the original wood, but, because of the manufacturing methods, gain new and different properties from those of the wood. Because they are manufactured, they can be and are tailored to satisfy a use, need, or group of needs.

Speaking in general terms, the wood-base panel materials are manufactured by the following methods:

- Changing wood substance, essentially to fibers and then interfelting them together again into the panel material classed as building fiberboard.
- Strictly mechanical means of cutting or breaking wood into small discrete particles and then, with a synthetic resin adhesive or other suitable binder, bonding them together again in the presence of heat and pressure. These latter products are appropriately called *particleboards*.

Building fiberboards, then, are made essentially of fiberlike components of wood that are interfelted together in the reconstitution and are characterized by a bond produced by that interfelting. They are frequently classified as fibrous-felted board products. At certain densities under controlled conditions of hot pressing, rebonding of the lignin effects a further bond in the panel product produced. Binding agents and other materials may be added during manufacture to increase strength, increase resistance to fire, moisture, or decay, or to improve some other property. Among the materials added are rosin, alum, asphalt, paraffin, synthetic and natural resins, preservative and fire-resistant chemicals, and drying oils.

Particleboards are manufactured from small components of wood that are glued together with a thermosetting synthetic resin or equivalent binder. Wax sizing is added to all commercially produced particleboard to improve water resistance. Other additives may be introduced during manufacture to improve some property or provide added resistance to fire, insects such as termites, or decay. Particleboard is among the newest of the wood-base panel materials.

It has become a successful and economical panel product because of the availability and economy of thermosetting synthetic resins, which permit blends of wood particles and the synthetic resin to be consolidated and the resin set (cured) in a press that is heated.

Thermosetting resins used are primarily urea-formaldehyde and phenol-formaldehyde. Urea-formaldehyde is lowest in cost and is the binder used in greatest quantity for particleboard intended for interior or other nonsevere exposures. Where moderate water or heat resistance is required, melamine-urea-formaldehyde resin blends are being used. For severe exposures, such as exteriors or where some heat resistance is required, phenolics are generally used.

The kinds of wood particles used in the manufacture of particleboard range from specially cut flakes an inch or more in length (parallel to the grain of the wood) and only a few hundredths of an inch thick to fine particles approaching fibers or flour in size. The synthetic resin solids are usually between 5 and 10 percent by weight of the dry wood furnish. These resins are set by heat as the wood particle-resin blend is compressed in flat-platen presses.

As floor underlayment, particleboard provides (1) the leveling, (2) the thickness of construction required to bring the final floor to elevation, and (3) the indentation-resistant smooth surface necessary as the base for resilient finish floors of linoleum, rubber, vinyl tile, and sheet material. Particleboard for this use is produced in 4- by 8-foot panels commonly $^1/_4$, $^3/_8$, or $^5/_8$ inch thick. Separate use specifications cover particleboard floor underlayment. In addition, all manufacturers of particleboard floor underlayment provide individual application instructions and guarantees because of the importance of proper application and the interaction effects of joists, subfloor, underlayment, adhesives, and finish flooring. Particleboard underlayment is sold under a certified quality program where established grade marks clearly identify the use, quality, grade, and originating mill.

Other uses for particleboard have special requirements, as for phenol-formaldehyde, a more durable adhesive, in the board. Particleboard for siding, combined siding sheathing, and use as soffit linings and ceilings for carports, porches, and the like requires this more durable adhesive. For these uses, type 2 medium-density board is required. In addition, such agencies as the Federal Housing Administration have established requirements for particleboard for such use. The satisfactory performance of particleboard in exterior exposure depends not only on the manufacture and kind of adhesive used but also on the protection afforded by the finish. Manufacturers

recognize the importance by providing both paint-primed panels and those completely finished with liquid paint systems or factory-applied plastic films.

Lumber Distribution

Large primary manufacturers and wholesale organizations set up distribution yards in lumber-consuming areas to more effectively distribute both hardwood and softwood products. Retail yards draw inventory from distribution yards and, in wood-producing areas, from local lumber producers.

Retail Yard Inventory

The small retail yards throughout the United States carry softwoods required for ordinary construction purposes and small stocks of one or two hardwoods in the grades suitable for finishing or cabinetwork. Special orders must be made for other hardwoods. Trim items such as moulding in either softwood or hardwood are available cut to size and standard pattern. Cabinets are usually made by millwork plants ready for installation. Many common styles and sizes are carried or cataloged by the modern retail yard. Hardwood flooring is available to the buyer only in standard patterns. Some retail yards may carry specialty stress grades of lumber such as structural light framing for truss rafter fabrication.

Some lumber grades and sizes serve a variety of construction needs. Some species or species groups are available at the retail level only in grade groups. Typical are house-framing grades, such as joist and plank, which is often sold as No. 2 and Better (2&BTR). The percentage of each grade in a grouping is part of the purchase agreement between the primary lumber manufacturer and the wholesaler. However, this ratio may be altered at the retail level by sorting. Where grade grouping is the practice, a requirement for a specific grade, such as No. 1, at the retail level will require sorting or special purchase. Grade grouping occurs for reasons of tradition and of efficiency in distribution.

Another important factor in retail yard inventory is that not all grades, sizes, and species described by the grade rules are produced, and not all those produced are distributed uniformly to all marketing areas. Regional consumer interest, building code requirements, and transportation costs influence distribution patterns. Often small retail yards will stock only a limited number of species and grades. Large yards, on the other hand, may cater to particular construction industry needs and carry drier dimension grades along with clears, finish, and decking. The effect of these variable retail practices is

that the grades, sizes, and species outlined in the grade rules must be examined to determine what actually is available. A brief description of lumber products commonly carried by retail yards follows.

Stress-Graded Lumber for Construction

Dimension is the principal stress-graded lumber item available in a retail yard. It is primarily framing lumber for joists, rafters, and studs. Strength, stiffness, and uniformity of size are essential requirements. Dimension is stocked in all yards, frequently in only one or two of the general-purpose construction woods (such as pine, fir, hemlock, or spruce):

- The 2 × 6, 2 × 8, and 2 × 10 dimensions are found in grades of Select Structural, No. 1, No. 2, and No. 3. They may also be found in combinations of No. 2 & BTR or possibly No. 3 & BTR.
- The 2 × 4 grades available would normally be Construction and Standard. They are sold as Standard and Better (STD&BTR), Utility and Better (UTIL&BTR), or Stud, in lengths of 10 feet and shorter.

Dimensional lumber is often found in nominal 2-, 4-, 6-, 8-, 10-, or 12-inch widths and 8- to 18-foot lengths in multiples of 2 feet. Dimensional lumber that is formed by structural end-jointing procedures may be found. Dimensional lumber thicker than 2 inches and longer than 18 feet is not available in large quantity.

Other stress-graded products generally present are posts and timbers, with some beams and stringers also possibly in stock. Typical stress grades in these products are Select Structural and No. 1 Structural in Douglas fir and No. 1SR and No. 2SR in southern pine.

Nonstress-Graded Lumber for Construction

Boards are the most common nonstress-graded general-purpose construction lumber in the retail yard. Boards are stocked in one or more species, usually in nominal 1-inch thickness. Standard nominal widths are 2, 3, 4, 6, 8, 10, and 12 inches. Grades most generally available in retail yards are No. 1, No. 2, and No. 3 (or Construction, Standard, and Utility). These will often be combined in grade groups. Boards are sold square-ended, dressed and matched (tongued and grooved), or with a ship-lapped joint. Boards formed by end jointing of shorter sections may form an appreciable portion of the inventory.

Appearance Lumber

Completion of a construction project usually depends on a variety of lumber items available in finished or semifinished form. The following items often may be stocked in only a few species, in only a few finishes, or in limited sizes depending on the yards:

- *Finish*—Finish boards usually are available in a local yard in one or two species, principally in grade C&BTR. Redwood and cedar have different grade designations. Grades such as Clear Heart, A, or B are used in cedar; Clear All Heart, Clear, and Select are typical redwood grades. Finish boards are usually a nominal 1-inch thick, dressed two sides to $^{3}/_{4}$ inch. The widths usually stocked are nominal 2 to 12 inches in even-numbered inches.
- *Siding*—Siding, as the name implies, is intended specifically to cover exterior walls. Beveled siding is ordinarily stocked only in white pine, ponderosa pine, western red cedar, cypress, or redwood. *Drop siding* (also known as *rustic siding* or *barn siding*) is usually stocked in the same species as beveled siding. Siding may be stocked as B&BTR or C&BTR, except in cedar (where Clear, A, and B may be available) and redwood (where Clear All Heart and Clear will be found). Vertical grain (VG) is sometimes a part of the grade designation. Drop siding sometimes is stocked also in sound knotted C and D grades of southern pine, Douglas fir, and hemlock. Drop siding may be dressed, matched, or ship lapped.
- *Flooring*—Flooring is made chiefly from hardwoods such as oak and maple, and the harder softwood species, such as Douglas fir, western larch, and southern pine. Often, at least one softwood and one hardwood are stocked. Flooring is usually nominal 1-inch thick dressed to $^{25}/_{32}$ inch, and 3- and 4-inch nominal width. Thicker flooring is available for heavy-duty floors in hardwoods and softwoods. Thinner flooring is available in hardwoods, especially for recovering old floors. Vertical and flat grain (also called quartersawed and plain-sawed) flooring is manufactured from both softwoods and hardwoods. Vertical-grained flooring shrinks and swells less than flat-grained flooring, is more uniform in texture, wears more uniformly, and the joints do not open as much.

  Softwood flooring is usually available in B and Better grade, C Select, or D Select. The chief grades in maple are Clear

No. 1 and No. 2. The grades in quartersawed oak are Clear and Select, and in plain-sawed, Clear, Select, and No. 1 Common. Quartersawed hardwood flooring has the same advantages as vertical-grained softwood flooring. In addition, the silver or flaked grain of quartersawed flooring is frequently preferred to the figure of plain-sawed flooring. Beech, birch, and walnut and mahogany (for fancy parquet flooring) are also occasionally used.

- *Casing and Base*—Casing and base are standard items in the more important softwoods and are stocked by most yards in at least one species. The chief grade, B and Better, is designed to meet the requirements of interior trim for dwellings. Many casing and base patterns are dressed to $^{11}/_{16} \times 2^{1}/_{4}$. Other sizes used include $^{9}/_{16} \times 3$, $3^{1}/_{4}$, and $3^{1}/_{2}$. Hardwoods for the same purposes (such as oak and birch) may be carried in stock in the retail yard or may be obtained on special order.
- *Shingles and Shakes*—Shingles usually available are sawn from western red cedar, northern white cedar, and redwood. The shingle grades are: Western red cedar, No. 1, No. 2, No. 3; northern white cedar, Extra, Clear, 2nd Clear, Clear Wall, Utility; redwood, No. 1, No. 2 VG, and No. 2 MG.

  Shingles that are all heartwood give greater resistance to decay than do shingles that contain sapwood. *Edge-grained shingles* are less likely to warp than *flat-grained shingles*; *thick-butted shingles* less likely than *thin shingles*; and *narrow shingles* less likely than *wide shingles*. The standard thicknesses of shingles are described as $^{4}/_{2}$, $^{5}/_{2}$, $^{1}/_{4}$, and $^{5}/_{2}$ (four shingles to 2 inches of butt thickness, five shingles to $2^{1}/_{4}$ inches of butt thickness, and five shingles to 2 inches of butt thickness). Lengths may be 16, 18, or 24 inches. Random widths and specified widths (dimension shingles) are available in western red cedar, redwood, and cypress.

  Shingles are usually packed four bundles to the square. A square of shingles will cover 100 square feet of roof area when the shingles are applied at standard weather exposures.

  Shakes are handsplit, or handsplit and resawn from western red cedar. Shakes are of a single grade and must be 100 percent clear, graded from the split face in the case of handsplit and resawn material. Handsplit shakes are graded from the best face. Shakes must be 100 percent heartwood free of bark and sapwood. The standard thickness of shakes ranges from

$^3/_8$ to $1^1/_4$ inches. Lengths are 18 and 24 inches, and a 15-inch Starter-Finish Course length.

Important Purchase Considerations

The following are some of the points to consider when ordering lumber or timbers:

- *Quantity*—Feet, board measure, number of pieces of definite size and length. Consider that the board measure depends on the thickness and width nomenclature used and that the interpretation of this must be clearly delineated. In other words, nominal or actual, pattern size, and so forth must be considered.
- *Size*—Thickness in inches—nominal and actual if surfaced on faces. Width in inches—nominal and actual if surfaced on edges. Length in feet—may be nominal average length, limiting length, or a single uniform length. Often a trade designation, random length, is used to denote a nonspecified assortment of lengths. Note that such an assortment should contain critical lengths as well as a range. The limits allowed in making the assortment random can be established at the time of purchase.
- *Grade*—As indicated in grading rules of lumber manufacturing associations, some grade combinations (B&BTR) are official grades; other (STD&BTR light framing, for example) are grade combinations and subject to purchase agreement. A typical assortment is 75 percent Construction and 25 percent Standard, sold under the label STD&BTR. In softwood, each piece of such lumber typically is stamped with its grade, a name or number identifying the producing mill, the dryness at the time of surfacing, and a symbol identifying the inspection agency supervising the grading inspection. The grade designation stamped on a piece indicates the quality at the time the piece was graded. Subsequent exposure to unfavorable storage conditions, improper drying, or careless handling may cause the material to fall below its original grade.

  Note that working or rerunning a graded product to a pattern may result in changing or invalidating the original grade. The purchase specification should be clear regarding regrading or acceptance of worked lumber. In softwood lumber, grades for dry lumber generally are determined after kiln drying and surfacing. This practice is not general for hardwood factory

lumber, however, where the grade is generally based on grade and size prior to kiln drying.

- *Species or groupings of wood*—Douglas fir, cypress, hemlock, fir. Some species have been grouped for marketing convenience. Others are traded under a variety of names. Be sure the species or species group is correct and clearly depicted on the purchase specification.
- *Product*—Flooring, siding, timbers, boards, and so forth. Nomenclature varies by species, region, and grading association. To be certain the nomenclature is correct for the product, refer to the grading rule by number and paragraph.
- *Condition of seasoning*—Air dry, kiln dry, and so on. Softwood lumber dried to 19 percent moisture content or less (S-DRY) is defined as dry by the American Lumber Standard. Other degrees of dryness are partially air-dried (PAD), green (S-GRN), and 15 percent maximum (KD in southern pine). There are several specified levels of moisture content for redwood. If the moisture requirement is critical, the levels and determination of moisture content must be specified.
- *Surfacing and working*—Rough (unplaned), dressed (surfaced), or patterned stock. Specify condition. If surfaced, indicate S4S, S1S1E, and so on. If patterned, list pattern number with reference to the appropriate grade rules.
- *Grading Rules*—Official grading agency name, product identification, paragraph number or page number or both, date of rules or official rule volume (rule No. 16, for example).
- *Manufacturer*—Most lumber products are sold without reference to a specific manufacturer. If proprietary names or quality features of a manufacturer are required, this must be stipulated clearly on the purchase agreement.

Engineered Lumber

Some lumber has been engineered to produce a variety of improvements for use in construction of homes and office and industrial buildings. There is glued laminated lumber that is called glulam. This engineered lumber is composed of wood laminations (lams). They are bonded together with adhesives. The grain of the lams runs parallel to the length of the glulam member. Lams are $1^3/_8$ inches thick for southern pine species and $1^1/_2$ inches for western wood species. Laminated boards up to 2 inches thick may also be used to form glulam timbers. Glulam widths are standardized at

$3^1/_8$ inches, $3^1/_2$ inches, $5^1/_8$ inches, $5^1/_2$ inches, and $6^3/_4$ inches. However, almost any width can be custom made.

Glulam dates back to the early 1900s when the first patents for glulam were obtained in Switzerland and Germany. One of the first glulam structures in the United States was built for the USDA Forest Products Laboratories in 1934. Glulam manufacture has evolved to make glulam one of the most widely used engineered lumber products in the construction industry. In the past, glulam was mainly associated with heavy timber construction, such as industrial roofing systems, bridges, and marine piers. Glulam timbers are used in residential construction as well as in commercial industrial construction.

Summary

Lumber is the basic construction material used in construction. Wood is a highly variable product so lumber is standardized by sizes and grades. The three classes of grading are stress-graded, nonstress-graded, and appearance. Plywood is made up of layers of wood veneer in wide long panels that have several advantages over solid wood lumber. Particleboard is made of small chunks of wood compressed into a large panel. Understanding lumber and wood product distribution and inventories leads to more successful and economic construction projects.

Review Questions

1. Why is lumber standardized by size and grade?
2. What grades are used for structural light framing?
3. What grade of lumber would you use for kitchen cabinet construction?
4. What are the advantages of plywood over solid lumber?
5. What grades and species of select softwood lumber are available at your local retail lumberyard?

Chapter 4

Strength of Timbers

The various mechanical properties of woods have been investigated by exhaustive testing in many laboratories, most notably the Forest Products Laboratory of the U.S. Department of Agriculture in Madison, Wisconsin. In addition, much research has been done in civil and agricultural engineering laboratories at many state universities.

Tension

A tension test is made as indicated in Figure 4-1. Although modern testing machines are by no means as simple as the apparatus shown, it serves well to show how such tests are made. The specimen is placed in the machine, gripped at each end, and the load is progressively increased until the material reaches its failure point. Usually, the elongation is not of any significance unless it is desired to determine the modulus of elasticity. The allowable fiber stresses indicated by modern stress grading (such as 1200f or 1450f) apply equally to extreme fiber stress in bending and to tensile stresses.

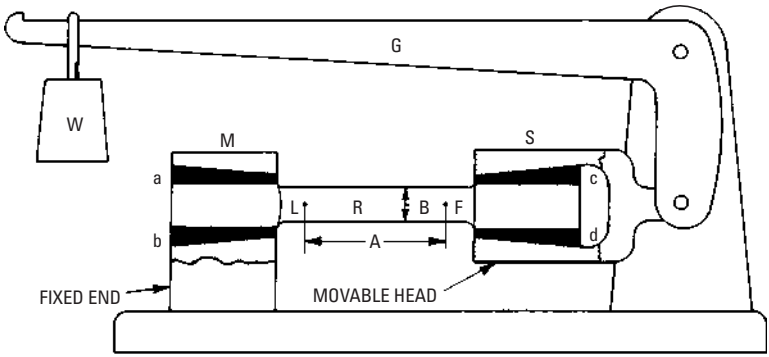

Figure 4-1 The tension test. The specimen R is placed in the wedge grips a, b, c, and d, thus applying tension between the fixed end and the movable head of the machine. The movable head is connected to the scale lever G on which the weight W slides. This arrangement is similar to an ordinary weighing scale. Two center marks (L and F) are punched on the specimen at a standard distance (A) apart. When testing, the pull on the specimen is gradually increased by moving W to the left. Dimensions A and B are then measured after each load increase.

Example A truss member is subjected to a tensile force of 50,000 pounds. What size timber of No. 1 dense yellow pine (1600f grade) will be required?

This is the calculation:

$$\frac{5000}{1600} = 31.25 \text{ square inches cross section}$$

Therefore, either a standard-dressed 4 inches × 10 inches ($3^1/_2$ inches × $9^1/_4$ inches) or a 6 inches × 6 inches ($5^1/_2$ inches × $5^1/_2$ inches) should be adequate.

Compression

A column supporting a load that tends to crush it is said to be in *compression*. Allowable compression parallel to the grain in stress-graded lumber is usually slightly less than the allowable bending and tensile stresses. For No. 1 dense timbers, 1600f grade, the compression stress is 1500 pounds per square inch, and for No. 2 dense, 1200f grade, it is 900 pounds per square inch. The builder need have no other concern than to see that the specified stresses are not exceeded. For loads applied across the grain, these stresses are much less. For the two grades mentioned, the cross-grain stress is 455 pounds per square inch for both grades.

When making a compression test (see Figure 4-2), a prepared specimen is placed between two plates, and a measured load is

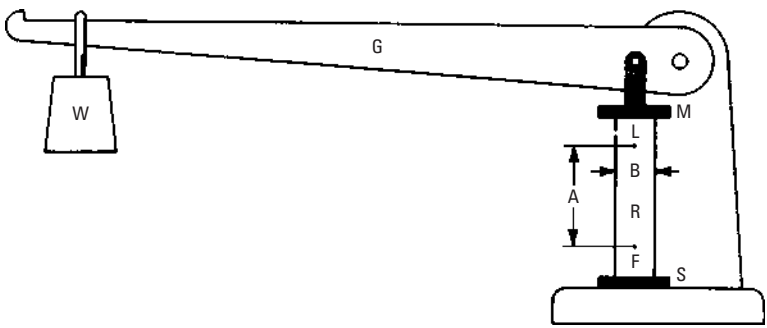

Figure 4-2 The compression test. The specimen R is placed between two plates (M and S), and a compression stress of any desired intensity is applied by moving the weight W on the lever G. As the load is gradually increased, the changes in dimensions A and B are noted, and a final result can be obtained that will indicate the amount of compression that the specimen can withstand.

applied. The load is increased progressively until the failure point is reached. Actually, modern wood-compression testing is not quite so simple. An instrument called a *compressometer* is pinned to the side of the specimen with sharp-pointed screws at points corresponding to points L and F in Figure 4-2. It is fitted with a series of levers that are connected to a dial gauge, usually reading to .0001 inch. The stem of the dial gauge is depressed until it shows a positive reading and compression in the specimen releases a part of the gauge. In this way, a sudden failure will release the gauge entirely, and jamming and ruining of the instrument is avoided. Usually, at least one of the plates through which pressure is applied is a cast iron hemisphere, thereby ensuring an evenly distributed pressure over the entire cross-section of the specimen.

Working Stresses for Columns

The amount of allowable loads on wood columns has been (and continues to be) a subject that is open to some discussion. The matter is complicated by the fact that columns of different lengths and diameters do not behave in the same manner under loadings. For the rather short column, failure will be caused by the actual crushing of the fibers of the wood, and the full compressive strength of the wood may thus be utilized. For a slightly longer column, failure may be caused by a sort of diagonal shearing action. For a long column, it will probably fail by bending sidewise and breaking. No one method can be adapted exactly for calculating allowable loads on columns of all lengths and slenderness ratios.

Column formulas are numerous. Some are based on empirical data or the results of actual testing. Only one, the *Euler* (pronounced *oiler*) *formula* seems to be based on purely mathematical calculations. It is of German origin, and, in its original form, it is so cumbersome that few designers in the U.S. care to use it. It assumes that a column will fail by *bending and breaking*. This is ensured only if the column is long and comparatively slender. However, a *modified* Euler formula is greatly favored by present-day timber designers. It is written as follows:

$$\text{Allowable load per square inch of cross section} = \frac{.3E}{(l/d)^2}$$

where the following is true:

E = modulus of elasticity of the timber used

l = unsupported length of the column (in inches)

d = least side, or diameter of a round column

This formula is applicable only when the results of its use do not indicate a higher stress than the maximum allowable unit stresses (as defined by the stress grade of the timber used). The ratio l/d in this equation is sometimes called the *slenderness ratio*. As an example of the use of this formula, consider the following problem.

How much of a load may safely be imposed on a 6-inch × 6-inch dressed yellow pine column that is 12 feet long?

The slenderness ratio, or l/d, of 6-inch × 6-inch columns 12 feet long is $^{144}/_{6} = 24$, and the modulus of elasticity of almost all good yellow pine is 1,760,000 pounds per square inch. This is the calculation:

$$\frac{0.3 \times 1{,}760{,}000}{24 \times 24} = 917 \text{ pounds per square inch}$$

The cross-sectional area of standard 6-inch × 6-inch timbers is 30.25 square inches. Therefore, the total load allowable on the column will then be:

$$917 \times 30.25 = 27{,}739 \text{ pounds}$$

An important point with respect to timber in compression is that the ends should be cut exactly square so that there will be a full bearing surface. Otherwise the timber will be subjected at the ends to more than the working stress (see Figure 4-3).

Shearing Stresses

Shearing stresses in wood are dangerous only in the direction parallel to the grain. It is almost impossible to shear the material across the grain until the specimen has been *crushed*. The crushing strength, then, and not the shearing resistance of the wood, will govern the maximum stress that can be applied to the wood.

The standard shearing test for woods with the grain was developed by the American Society for Testing Materials (ASTM), and the procedure is standardized. The specimens are standard dimensions, shaped as shown in Figure 4-4 in pairs, one with the notch at right angles to that in the other. The results of the tests on the two specimens are averaged, but usually they vary only slightly. The blocks are tested in a special shear tool that is loaded in a universal testing machine, with the load applied at a rate of .024 inch per minute.

The results of testing different types of woods vary widely. To provide for this lack of uniformity, the shear allowances in stress-rated timbers may contain a reduction factor (safety factor) of as much as

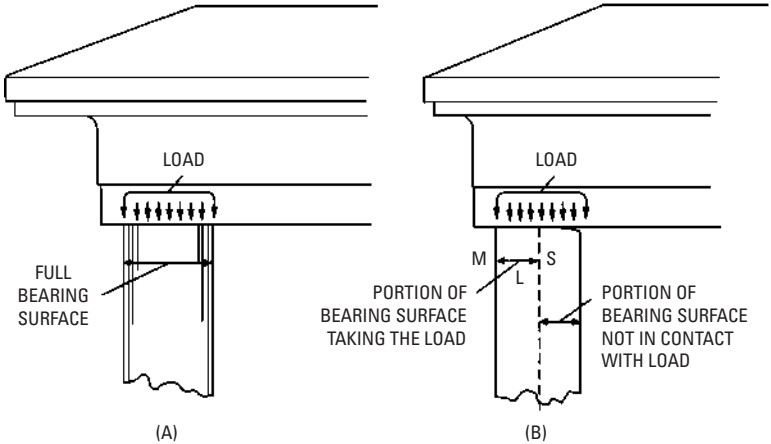

Figure 4-3 (A) Good and (B) poor column bearing surfaces illustrate the importance of squaring columns accurately when cutting them. The entire top of the column must be in contact with the load member so that the pressure per square inch of cross section on the column will correspond to the allowable working pressure for which the column was designed. If the portion MS of the bearing surface in contact is only half the entire surface, then the stress applied on the top of the column will be twice that of full contact, as shown by (A).

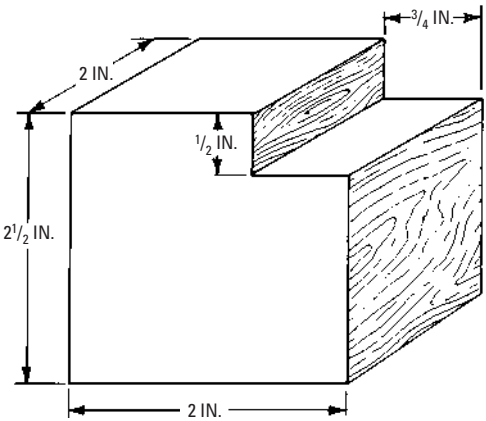

Figure 4-4 The ASTM standard wood-shear test specimens.

10, or it may be as little as 2 or 3. This is necessary because a piece that is below average in shearing resistance may appear anywhere.

The distinctions *across the grain* and *with the grain* should be carefully noted. Wet or green wood, in general, shears approximately one-fifth to one-half as easily as dry wood. A surface parallel to the rings (tangent) shears more easily than one parallel to the medullary rays. The lighter conifers and hardwoods offer less resistance than the heavier kinds, but the best pine shears one-third to one-half more readily than oak or hickory, thereby indicating that great shearing strength is characteristic of tough woods.

Horizontal Shears

The types of shears discussed in the preceding section are called *external* shears because they are caused by forces that originate outside the body of the material, and for the most part, they are evident and readily provided for (see Figure 4-5). In addition, in every loaded beam, a system of *internal* stresses is set up. This may be explained by observing that every shearing force results from *two* forces that are *unbalanced*. They do not meet at the same point, and a *stress couple* is set up that can be met and held motionless only by another couple that acts in opposition to it.

This is illustrated in Figure 4-6. It represents a particle of indefinite size, but possibly infinitely small, that has been extracted from the body of a beam that has been stressed by bending. In Figure 4-6A, the particle has been subjected to a vertical shearing force, or pair of forces, since there can be no action without an opposing and equal reaction. The particle is *unstable*. Since the two forces do not meet at a common point, the particle tends to revolve in a counterclockwise direction.

In Figure 4-6B, a pair of horizontal forces is supplied. These forces represent another stress couple that balances or neutralizes the vertical shearing forces. They are known as the *internal horizontal shears*. A reasonable deduction, therefore, is that at any point on the beam, there exist internal horizontal shears equal in intensity to the external vertical shears. The horizontal shears are dangerous in wood beams. That is because timbers have a low resistance to shearing with the grain.

Figure 4-6C shows how the four forces may be resolved into a single pair of concurrent forces that *do* meet at a common point. These are not shearing forces. They are, however, the result of shearing forces. When the material has approximately the same resistance to shears in all directions (such as concrete or steel), this is what occurs. Each particle tends to elongate in one diagonal

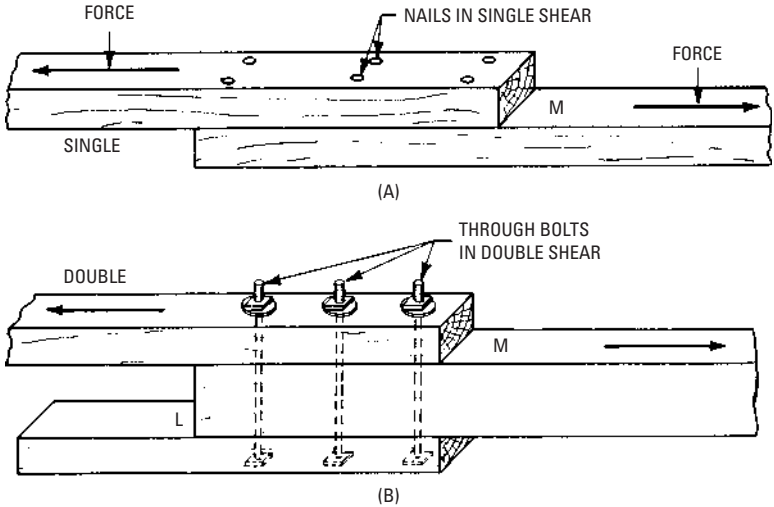

Figure 4-5 Examples of single and double shear. (A) The nails are in single shear. This is the assumption made as a rule when calculating the strength of nailed joints. However, nails may be placed in double shear, similar to the bolts in (B), if they are long enough to almost penetrate all three members. The nails in double shear will safely carry twice the load that could safely be placed on nails in single shear. However, most nailed joints are designed as if the nails were in single shear, though they may actually be in double shear. In (B), the bolts are truly in double shear, but the joint is not twice as strong as a plain lapped joint if the center member is not at least as thick as the combined thicknesses of the outer members. The outer members are usually both the same thickness.

and forms the diagonal tension that is so dangerous in reinforced concrete beams. In fact, it is the most dangerous stress in such beams, and to resist it, elaborate web reinforcing and bent-up bars are provided at points where this stress is highest (usually near the ends of the beams). In I-beams, diagonal tension results in buckling or wrinkling of the relatively thin webs. In wood beams, the forces are not resolved in this way, since wood is strong enough to resist the vertical components, but the horizontal components tend to split the beam (usually at or near the ends near the center of the height). Wood beams that are season-checked at the ends (as many are) are low in resistance to such stresses.

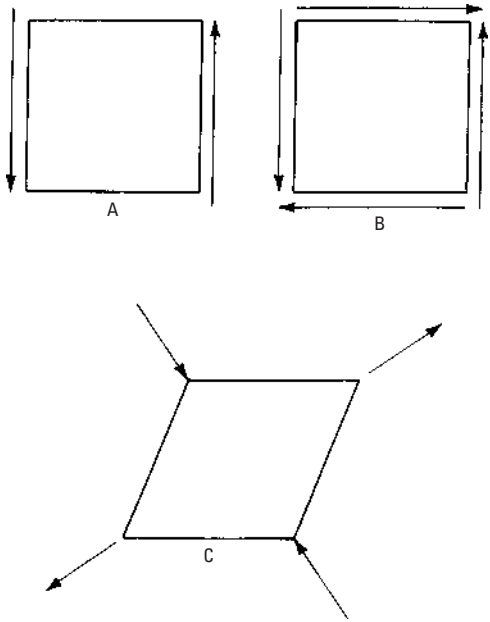

Figure 4-6 The actions of vertical and horizontal shears.

If the depth of a wood beam is greater than one-tenth to one-twelfth of its span, horizontal shears (and not bending strength) often govern its ability to carry loads. Shears are usually not dangerous in wood beams unless they are relatively deep and heavily loaded.

Transverse or Bending Stress

This is the kind of stress present on numerous building timbers (such as girders, joists, or rafters) that causes a deflection or bending between the points of support. What takes place when these or similar members are subjected to bending stress is discussed in the section titled *Beams* later in this chapter.

Stiffness

By definition, stiffness is that quality possessed by a beam or other timber to resist the action of a bending force. The action of the bending force tends to change a beam from a straight to a curved form (that is, a *deflection* takes place). When a load is applied, the beam (originally assumed straight and horizontal) sags or bends downward between the supports. The amount of downward

movement measured at a point midway between the supports is the amount of deflection. The action of beams subjected to bending forces is described as follows.

If a load of 100 pounds placed in the middle of a stick that is 2 inches × 2 inches and 4 feet long, supported at both ends, bends or deflects this stick one-eighth of an inch (in the middle), then 200 pounds will bend it about $^1/_4$ inch, 300 pounds, $^3/_8$ inch, and so on, the deflection varying directly as the load. This is in accordance with Hooke's Law that states that stress is proportional to strain. Soon, however, a point is reached where an additional 100 pounds adds more than $^1/_8$ inch to the deflection—the limit of elasticity has been exceeded.

Taking another piece from the straight-grained and perfectly clear plank of the same depth and width but 8 feet long, the load of 100 pounds will deflect it by approximately 1 inch. Doubling the length reduces the stiffness eightfold. Stiffness, then, decreases as the cube of the length.

If AB in Figure 4-7 is a piece of wood, and D is the deflection produced by a weight or load, then

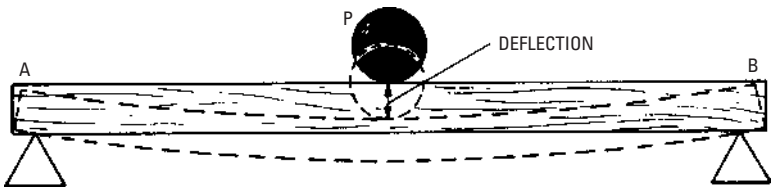

Figure 4-7 A simple beam, loaded at the middle and supported at both ends, is used to illustrate the term deflection.

$$\text{deflection } (D) = \frac{Pl^3}{48EI}$$

where the following is true:

P = the load, concentrated at the center of the span (in pounds)
l = the length of the span (in feet)
E = the modulus of elasticity of the material
I = the moment of inertia (for rectangular beams =

$$\frac{bd^3}{12}$$

where b = width and d = height).

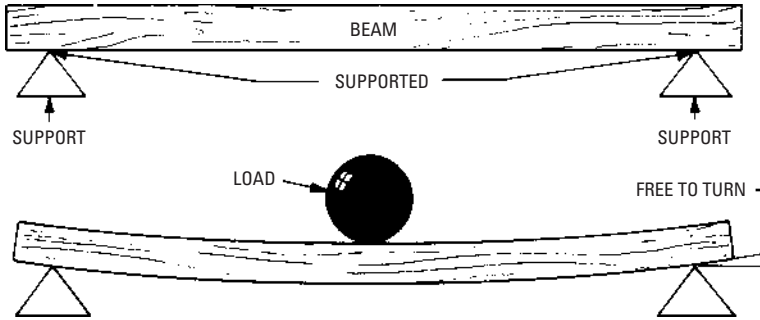

Figure 4-8 A beam supported at the ends. The ends of the beam are free to follow any deflection, thus offering no resistance and rendering the beam less stiff than when the ends are fixed.

The following rules (in conjunction with Figure 4-8, Figure 4-9, and Figure 4-10) define the stiffness and strength of practically all types of wood beams:

- For beams with a rectangular cross-section, equal depths, and equal spans, their load-carrying capacity varies *directly* as their widths.
- For beams with equal widths and equal spans, their strength varies directly as the square of their depth.

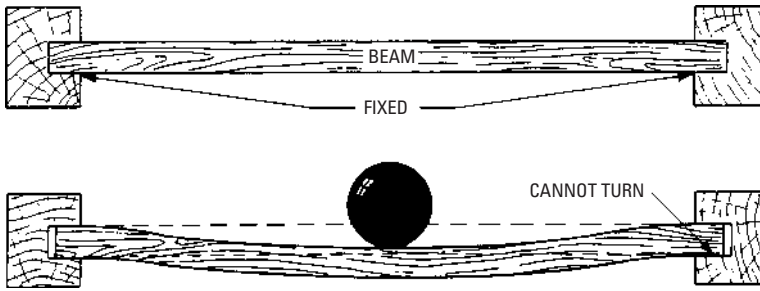

Figure 4-9 A beam supported at the ends. The ends are gripped or embedded in some unyielding substance so that they cannot turn or follow the deflection of the beam under an applied load. The beam then deflects in a compound curve, thus adding extensively to its stiffness. Therefore, a beam with fixed ends will deflect less than one with supported ends.

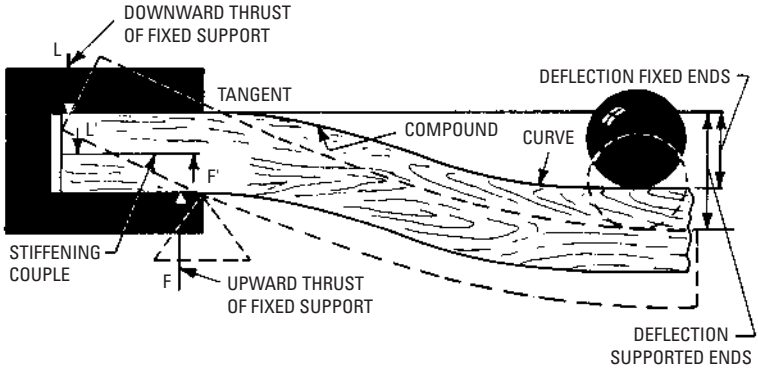

Figure 4-10 One end of a beam illustrating the stiffening effect of fixed ends as compared with supported ends. When the ends are fixed, the deflection of the beam will be resisted by an upward thrust, indicated by F, and a downward thrust, L. These thrusts form a resisting or stiffening couple that holds the portion of the beam embedded in the bearing in a horizontal position, thereby causing the beam to deflect in a compound curve, which increases its stiffness. The dotted lines show the excess deflection for the same load if the beam were simply supported at the ends.

- If depths and widths are the same, strengths vary *inversely* as the lengths of the spans.
- Their stiffness, or resistance to deflection, will vary *inversely* as the *cubes* of their spans, other factors being equal.
- Their stiffness will vary directly as the *cubes* of their depths, other factors being equal.
- Other factors being equal, stiffness will vary *directly* as their widths.
- If a beam is split horizontally, and the two halves are laid side by side, they will carry only *one-half* as much loading as the original beam.

These relations are not strictly true for I-beams because of their irregular shapes, but they are approximately true for all types of beams. It is usually most economical with materials to use as deep a beam as can conveniently be employed. Note that double 2-inch × 4-inch trimmers over window or door heads, if set edge up, are 8 times as strong and 32 times as stiff as when placed flatwise.

Both strength and stiffness are greater in dry timber than in green or wet wood of the same species. A piece of long-leaf yellow pine is 30 percent to 50 percent stronger, and 30 percent stiffer, when in an air-dry condition than when green. In general, both strength and stiffness are proportional to densities, or dry weights, although this is not always true. Edge-grain pieces are usually stronger and stiffer because the tangent to the rings runs horizontally, but not appreciably so. There is little or no difference in the sapwood and heartwood of the same species if the densities are the same. The tool handle of red heartwood is as serviceable as the handle of white sapwood, although white sapwood handles are still called premium grade.

Modulus of Elasticity

Since it is desirable (and, for many purposes, essential) to know beforehand that a given piece with a given load will bend only by a given amount, the stiffness of wood is usually stated in a uniform manner under the term *modulus* (measure) *of elasticity*. For good grades of Douglas fir and yellow pine that are stress-rated, the modulus of elasticity is 1,760,000 pounds per square inch.

Beams

A beam is a single structural member (usually horizontal or nearly so) that carries a load or loads over a given space. At their supports, beams may be:

- *Freely supported*—This merely means the beams are resting on their bearings.
- *Restrained, or partially fixed at their bearings*—Although some designers choose to consider such restraint in their designs, the actual degree of restraint can never be accurately determined, so restrained beams are more often considered as being freely supported.
- *Fixed at their supports*—In wood beams, this condition is rarely found. In steel frames, it is not unknown. In reinforced concrete frames, it is quite common. Attempts to fix the ends of wood beams are rarely permanent. Building the ends of a beam, *any* beam, into a wall or casting it into concrete for a short distance does not fix the beams at their supports.

Allowable Loads on Wood Beams

The allowable loads on freely supported wood beams of any species are readily calculated if the timbers are stress-rated and the allowable

fiber-stress is known. For a beam with a loading that is evenly distributed along its span, this is the equation to use:

$$W = \frac{f \times b \times d^2}{9 \times L}$$

where the following is true:

W = allowable evenly distributed loading, in pounds
f = allowable fiber stress in pounds per square inch
b = width of the beam, in inches
d = depth of the beam, in inches
L = length of the span, in feet

Figure 4-11 defines the dimensions of a beam—b, width; d, depth; and L, length.

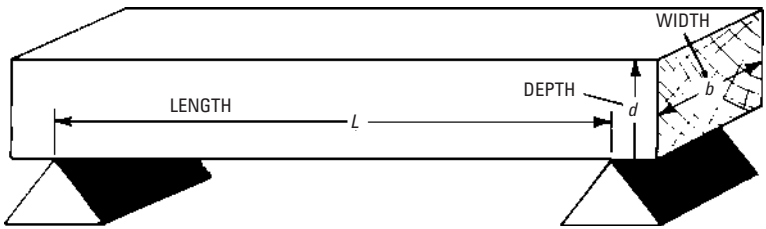

Figure 4-11 A beam resting on knife-edge supports illustrates the terms length, width, and depth.

Example What will be the maximum allowable load on a beam whose nominal size is 6 inches × 12 inches and whose actual size is $5^1/_2$ inches × $11^1/_2$ inches, with an 18-foot span? The timber is to be 1500f stress-rated. This is the calculation:

$$\frac{1500 \times 5.5 \times 11.5 \times 11.5}{9 \times 18} = 6735 \text{ pounds}$$

If the loading is to be concentrated at the center of the span, use one-half the load as calculated by the formula given. Therefore, for the same timber as calculated, with the same span, the allowable concentrated load will be

$$\frac{6735}{2} = 3367 \text{ pounds}$$

Breaking Loads on Wood Beams

Breaking loads are of no interest to the builder because the term is meaningless unless some explanation is made, since wood is extremely sensitive to the *duration* of loads. The load that would break a beam over a long period of time, for example, 10 years, will be only approximately $^{9}/_{16}$ of the load that would break it in a few minutes. The stresses specified in stress-rating lumber and timbers recognize this phenomenon. It is presumed that the full design loading will not be applied for more than 10 years during the life of the structure, and the time may be either cumulatively intermittent or continuous. It is also presumed that 90 percent of the full design loading may safely be applied for the full life of the structure. These presumptions make the use of stress-rated lumber quite conservative.

Example What is the safe working load, concentrated at the center of the span, for a full-size 6-inch × 10-inch white oak timber with a 12-foot span, stress-rated 1900f, if it is laid flatwise (see Figure 4-12A)? If it is set edge up (see Figure 4-12B)?

This is the calculation for the timber laid flatwise:

$$\frac{1900 \times 10 \times 6 \times 6}{9 \times 12} = 6333$$

$$\frac{6333}{2} = 3167 \text{ pounds}$$

For the timber set edge up:

$$\frac{1900 \times 6 \times 10 \times 10}{9 \times 12} = 10{,}555$$

$$\frac{10{,}555}{2} = 5278 \text{ pounds}$$

Distributed Load

Instead of placing the load in the middle of a beam, as in the example just given, the load may be regarded as being *distributed* (that is, the beam is uniformly loaded, as shown in Figure 4-13). Although this type of loading is actually a series of concentrated loads, it is often found in actual practice, and it is usually considered as being *uniformly* distributed.

Cantilever Beams

A *cantilever beam* (see Figure 4-14 and Figure 4-15) is firmly fixed at one end, or it is freely supported but with the end running some distance to a support above it. The loading may be either

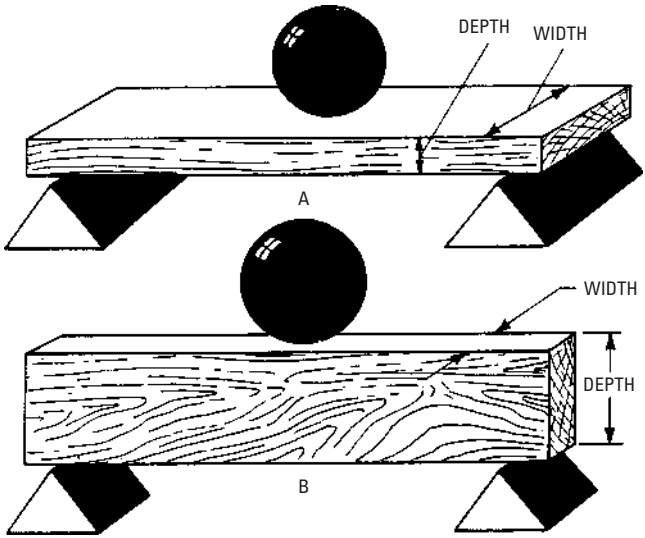

Figure 4-12 The terms width and depth of a beam depend on the position of the beam. In (A) the broad side is the width of the beam, whereas in (B), which is the same beam turned over 90°, the narrow side is the width.

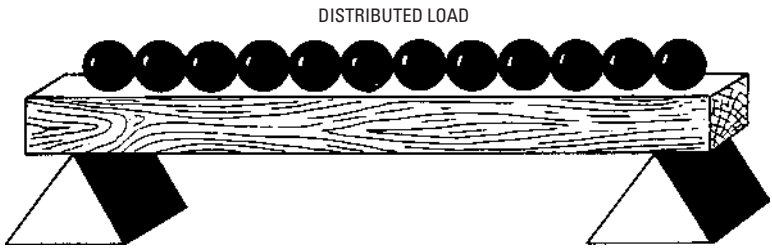

Figure 4-13 A distributed load is indicated by the iron balls equally spaced along the beam between the supports.

distributed or concentrated at any point on the span. All beams that project beyond a support and that carry a load at the free end are classed as cantilevers. Stresses set up in the beam external to the outside support are the same as when the beam is rigidly fixed at the support.

Figure 4-16 illustrates the comparison between the working loads of variously supported beams, contrasting middle and distributed

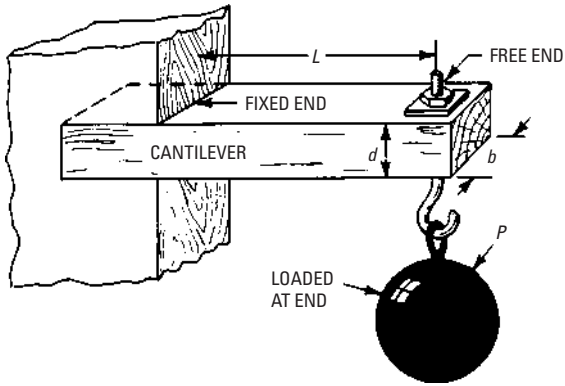

Figure 4-14 A simple cantilever beam that is fixed at one end and free at the other; this beam is supporting a concentrated load at its free end.

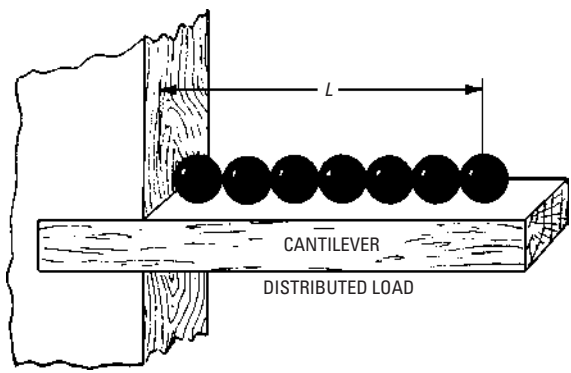

Figure 4-15 A cantilever beam with an equally distributed load.

loads of beams supported at both ends with equal loads on cantilever beams. In Figure 4-16, P= the concentrated loads, W= the evenly distributed load, f= the fiber stress, b= the width of the beam, d= the depth of the beam, and L= the length of the span.

Wind Loads on Roofs

Snow on roofs (if the roofs are of a slope that is low enough to permit the snow to lie in place) will weigh approximately 8 pounds per square foot per foot of depth when freshly fallen. When wet, its

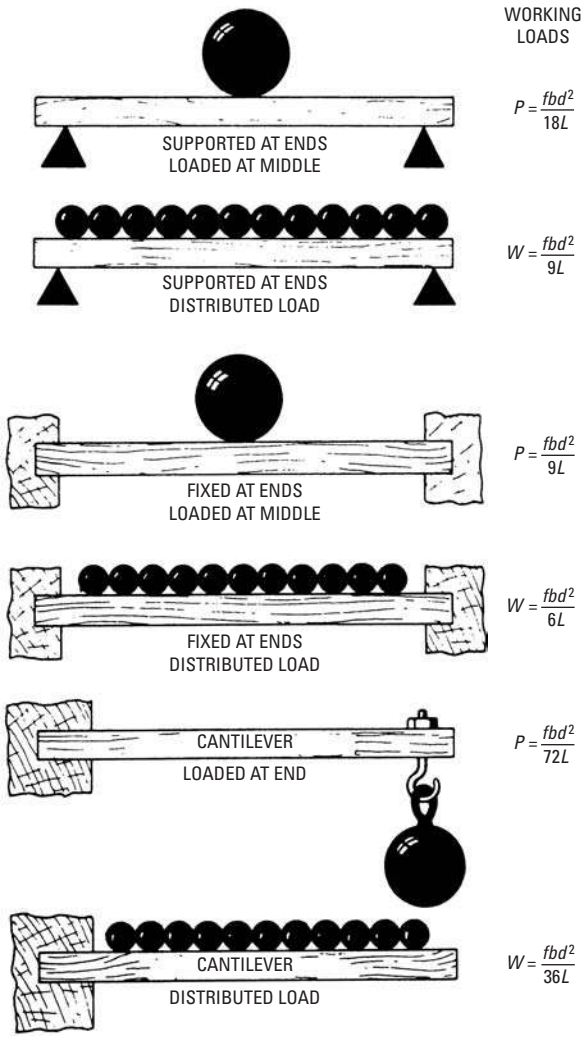

Figure 4-16 The working loads for beams with different types of loads and different modes of support.

weight will vary with the water content. A commonly used figure is 10 pounds per cubic foot. In calculating snow load, use the depth of the heaviest snow on record multiplied by 10, but never less than 20 pounds per square foot anywhere. If a roof is designed for much

less than 20 pounds per square foot of live load, it will not safely support workers.

In the past, there was considerable confusion regarding wind loads against sloping roofs, and some rather weird formulas were evolved to calculate their imaginary intensities (although no one had ever seen a roof blow *in*, and many persons have seen roofs blow *off*). In other words, wind loads are *negative* (that is, they create *suction*). This is now well recognized, and insurance companies are well informed on the subject.

In regions where hurricanes are not common, the insurance companies recommend that an uplift allowance of 30 pounds per square foot be provided for, acting at right angles to the slope of the roof. This will include the slight *positive* pressure commonly found inside buildings during windstorms. If eaves are wide and overhanging, it is recommended that a gross lifting force of 45 pounds per square foot be provided for. If the weight of the roof itself is appreciable, it may be deducted from the gross uplift allowance.

It should be recognized that a maximum snow load plus a maximum wind load could hardly occur at the same time. Thus, it cannot be considered compensating.

Definitions

In the shop and in the field, the fitness of any species of wood for a given purpose depends on various properties. When treating the strength, stiffness, hardness, and other properties of wood, many technical terms are used. For an understanding of these terms, the following definitions are worth noting:

- *Bending forces*—Forces that act on some members of a structural frame. They tend to deform them by flexure.
- *Brittleness*—Breaking easily and suddenly, usually with a comparatively smooth fracture. The opposite of *toughness* (sometimes incorrectly called *brashness*) that refers more to brittleness. Old and extremely dry wood is inclined to brashness. Green or wet wood is tougher, though not as strong in most cases.
- *Compression*—The effect of forces that tend to reduce or shorten the dimensions of a member.
- *Deformation*—A change of shape or dimension. Disfigurement, such as the elongation of a structural member under tension.

- *Ductile*—A term not applicable to wood. It is the property of a metal that allows it to be hammered thin or drawn into wires.
- *Elastic limit*—The greatest stress that a substance can withstand and still recover completely when the force or strain is removed.
- *Factor of safety*—The ratio between the stress at failure and allowable design stress. If the stress at failure in a bending beam is 4350 pounds per square inch (psi), and the timber is stress-graded at 1450f (allowable fiber stress in psi) the safety factor is $^{4350}/_{1450}$, or 3.
- *Force*—In common parlance, a *pull* or a *push* that would change the state of a body at rest, or would change the course of a body in motion (Newton's Law).
- *Load*—Pressure acting on a surface, usually caused by the action of gravity.
- *Member*—A part of a structure (such as a column, beam, or brace) that usually is subjected to compression, tension, shear, or bending.
- *Modulus, or coefficient, of elasticity*—Stress, tension, or compression, divided by the elongation or contraction per unit of length. Inside the elastic limit, the modulus of elasticity is approximately constant for most materials. In wood, it varies greatly with different species, with moisture content, and even in pieces sawed from the same log. In other words, it is subject to considerable natural variation.
- *Modulus of rupture*—The calculated fixed stress in a beam at the point of rupture. Since the elastic limit will have been passed at this point, it is not a true fiber stress, but it is a definite quantity, and the personal factor is not involved when obtaining it.
- *Permanent set*—When a member (either metal or wood) is stressed beyond its elastic limit, or subjected to stresses that may be far below its elastic limit for extremely long periods, it may take on permanent deformation. Permanent set in timbers does not mean or imply that the timbers have been weakened.
- *Resilience*—Synonym for *elasticity*. The property that enables a substance to spring back when a deforming force is removed.
- *Shear*—The effect of forces (external or internal) that causes bodies or parts of bodies to slide past each other.

- *Strain*—Alterations in the form of a member caused by forces acting on the member.
- *Strength*—The power to resist forces (which may be tensile, compressive, or shearing) without breaking or yielding.
- *Stress*—Distributed forces, such as pounds per square inch or tons per square foot. Within the elastic limit of materials, *stress* is approximately proportional to *strain*. This statement is called Hooke's Law.
- *Tenacity*—Synonym for *tensile strength*. The power to resist tearing apart.
- *Tension*—A force that tends to tear a body apart or elongate it.
- *Toughness*—Strong but flexible. Not brittle. Nearly the same as *tenacity*.
- *Ultimate strength*—The stress developed just before failure is evident.
- *Yield point*—This property is not evident in timbers. In steels, it occurs after the elastic limit has been passed. In materials that show no defined yield point, it may be arbitrarily assumed or defined as the stress where a permanent set occurs.

As an example of the uses for these terms, the tie rod in a truss resists being pulled apart because of its *tensile strength*. The *stress* thus applied *strains* the rod, *deforming* or *elongating* it. It is *stretched* and a *contraction* of the area of its cross-section results. If the *load* is not sufficient to *stress* the material past its *elastic limit*, the rod will return to its original length when the load is removed, depending on the duration of the load. If the *load* is heavy enough to stretch the rod past its *elastic limit*, it will not return to its original length when the *load* is removed, but it will remain *permanently set* if it is not pulled in two. The *elastic limit* is reached when *elongation* becomes proportionally greater than the *loading*. If the *load* is increased to the point where the rod breaks, or where its *tenacity* is overcome, it is *ruptured*. The rod itself is called a *member* of the truss.

Summary

In most shop work and in the field, the fitness of any species of wood for a given purpose depends on various properties. When treating the strength, stiffness, hardness, and other properties of wood, many factors must be considered. Stiffness is that quality possessed by a beam or other timber to resist the action of a bending force.

A beam is a single structural member (usually horizontal) that carries a load over a given space. The allowable loads on freely supported wood beams of any size can be calculated if the timber is stress-rated and the allowable fiber-stress is known. In many designs, the load is distributed uniformly over the length of the beam. Although this type of loading is actually a series of concentrated loads, it is often found in actual practice and is usually considered as being uniformly distributed.

Review Questions

1. What is meant by brittleness, bending forces, compression, and elastic limit?
2. What is meant by working stresses?
3. Explain shearing stresses.
4. Why is it important in some cases to distribute the load on a beam between the supports?
5. What is a cantilever beam? Explain its purpose.
6. What is a truss?
7. Where is compression experienced in construction?
8. When are shearing stresses in wood dangerous?
9. What are external shears?
10. What type of tension is so dangerous in reinforced concrete beams?

Chapter 5

Mathematics for Carpenters and Builders

An elementary knowledge of mathematics is essential to the carpenter to solve successfully the numerous problems encountered in almost any branch of carpentry. The branches of mathematics of which the carpenter should possess at least an elementary knowledge are the following:

- Arithmetic
- Geometry
- Trigonometry

Such knowledge will be found very useful, especially in making up estimates, solving steel square problems, and so forth.

Arithmetic

By definition *arithmetic* is the science of numbers and the art of reaching results by their use (see Figure 5-1).

Arithmetic Alphabet

In arithmetic figures are used to represent quantities or magnitudes, as follows:

Cipher	*one*	*two*	*three*	*four*	*five*	*six*	*seven*	*eight*	*nine*
0	1	2	3	4	5	6	7	8	9

A number is one or a collection of these figures to represent a definite quantity or magnitude as 1, 21, 517, 43,988, and so forth.

There are various kinds of numbers: simple, compound, integer, abstract, concrete, odd, even, prime, composite, and so forth.

Notation and Numeration

By definition, *notation* in arithmetic is *the writing down of figures to express a number,* and numeration is *the reading of the number or collection of figures already written*. By means of the ten figures given above any number can be expressed.

Figures have two values: simple and local. The *simple value* of a figure is its value when standing in units' place. The *local value* of a figure is the value that arises from its location.

Notation Basis—1—unit Arithmetic Alphabet 0 1 2 3 4 5 6 7 8 9	
Increased *By tens* 1, 10, 100, 1000, etc. By varying scales 1 oz 1 lb 1 cwt 1 pt 1 qt 1 gal 1 in. 1 ft 1 yd	***Diminished*** *By tens* 1, .1, .01, .001, etc. By varying scales $^1/_4$ $^6/_7$ $^1/_3$ oz $^{11}/_{23}$ etc. $^1/_6$ lb $^9/_8$ oz $^3/_8$ cwt etc.
According to the Four Ground Rules	
Addition Subtraction	Multiplication Division
By involution (powers)	By evolution (roots)
Relations Expressed by	
Ratios 2 : 3 5 : 6 8 : 9 etc. Proportion (equality of ratios) 2 : 3 : : 4 : 6 etc.	
Practical Applications Percentage, interest, profit and loss, reduction of weights and measures, measuring, etc.	

Figure 5-1 Scheme of arithmetic.

When one of the figures stands by itself, it is called a *unit*. If two of them stand together, the right-hand one is still called a unit, and the left-hand one is called *tens*. Thus, 79 is a collection of 9 units and 7 sets of ten units each, or of 9 units and 70 units, or of 79 units and is read as "seventy-nine."

If three of them stand together, then the left-hand one is called *hundreds*. Thus 279 is read "two hundred seventy-nine."

To express larger numbers other orders of units are formed. The figure in the fourth place denotes *thousands*, and in the fifth place denotes *ten thousands*. These are called *units of the fifth order*. The sixth place denotes hundred thousands, the seventh place denotes millions, and so forth.

The French method (which is the same as that used in the United States) of writing and reading large numbers is shown in Figure 5-2. This system is called Arabic notation and is the system in ordinary everyday use.

Names of periods	Billions	Millions	Thousands	Units		Thousandths
Order of Units	Hundred-billions Ten-billions Billions	Hundred-millions Ten-millions Millions	Hundred-thousands Ten-thousands Thousands	Hundreds Tens Units	Decimal point	Tenths Hundredths Thousandths
	8 7 6,	5 4 3,	2 0 1,	2 8 2,	•	4 8 9

Figure 5-2 Numeration.

Roman Notation

This system is occasionally used for chapter headings, corner stones, and so forth. The method of expressing numbers is by letters, as shown in Table 5-1.

In the Roman notation, when any character is placed at the right-hand side of a larger numeral, its value is added to that of such numeral: VI is V + I; XV is X + V; MD, is M + D; and so on. I, X, and rarely C are also placed at the left-hand side of other and larger numerals, and when so situated, their value is subtracted from such numerals: IV is V − I; XC is C − X; and so on. Formerly, the smaller figure was sometimes repeated in such a position twice, with its value being in such cases subtracted from the larger: IIX is X − II; XXC is C − XX; and so on. The custom of using the Roman notation for chapter numbers, year of copyright, sections, and so forth continues!

Definitions

Following are some important definitions worth noting:

- *Arithmetic*—The art of calculating by using numbers.
- *Number*—A total, amount, or aggregate of units. By counting the units, we arrive at a certain number (such as *two* horses or *five* dozen).
- *Unit*—May mean a single article, but often it means a definite group adopted as a standard of measurement (such as *dozen, ton, foot, bushel*, or *mile)*. Most commonly used units are standardized and are defined and fixed by law.

Table 5-1 Roman Numeral Values

Roman Numeral	*Value*
I	1
II	2
III	3
IV	4
V	5
VI	6
VII	7
VIII	8
IX	9
X	10
XI	11
XII	12
XIII	13
XIV	14
XV	15
XVI	16
XVII	17
XVIII	18
XIX	19
XX	20
XXX	30
XL	40
L	50
LX	60
LXX	70
LXXX	80
XC	90
C	100
D	500
M	1000
$\overline{X}$	10,000
$\overline{M}$	1 million

- *Concrete number*—A number applied to some particular unit (such as *ten* nails, *two* dozen eggs, *six* miles).
- *Abstract number*—One that is *not* applied to any object or group (such as simply *two, four, ten)*.
- *Notation*—The art of expressing numbers by figures or letters. Our system of notation is the Arabic notation. The Roman notation uses letters (such as V for 5, X for 10).
- *Cardinal numbers*—Numbers used in simple counting or in reply to the question "How many?" Any number may be a cardinal number.
- *Ordinal numbers*—Indicate succession or order of arrangement (such as *first, second, tenth)*.
- *Integer*, or *integral number*—A whole number, not a fraction or part.
- *Even number*—Any number that can be exactly divided by 2 (such as 4, 16, 96, 102).
- *Odd number*—Any number that is *not* exactly divisible by 2 (such as 3, 15, 49, 103).
- *Factor of a number*—A whole number that may be exactly divided into the number. For example, 3 is a factor of 27, 13 is a factor of 91.
- *Prime number*—A number that has no factors other than itself and 1. Thus, 3, 5, 7, and 23 are prime numbers.
- *Composite number*—A number that has factors other than itself and 1 (such as 8, 49, and 100).
- *Multiple of a number*—A number that is exactly divisible by a given number. For example, 91 is a multiple of 7, 12 is a multiple of 3.
- *Digit*—Any number from 1 to 9, and usually 0.

Signs of Operation

Following are some important signs of operation worth noting:

- The *sign of addition* is +, and it is read "plus" or "add." Thus, 7 + 3 is read "seven plus three." The numbers may be taken in *any order* when adding (7 + 3 is the same as 3 + 7).
- The *sign of subtraction* is –, and it is read "minus." A series of subtractions *must* be taken in the order written (11 – 7 is *not* the same as 7 – 11).

- The *sign of multiplication* is $\times$, and it is read "times" or "multiplied by." The numbers may be taken in *any order* when multiplying (4×7 is the same as 7×4).
- The sign of division is $\div$, and it is read "divided by." A series of divisions *must* be taken in the order written $(100 \div 2) \div 10 = 5$.
- The *sign of equality* is $=$, and it is read "equals" or "is equal to." The expressions on each side of an equality sign must be numerically the same. The complete expression is called an *equation*.

Use of the Signs of Operation

Following are some examples of the use of mathematical signs of operation.

Example *Use of the sign of addition*. A builder when building a house buys 1762 board feet of lumber from one yard, 2176 board feet from another, and 276 board feet from another. How many board feet did he buy?

The problem:

$$1762 + 2176 + 276 = ?$$

The solution:

$$\begin{array}{r} 1762 \\ 2176 \\ \underline{+\ 276} \\ 4214 \end{array} \text{ board feet}$$

Note how the numbers are aligned to permit addition. The units are all aligned on the right, then the tens, then the hundreds, then the thousands, each in the proper column.

Example *Use of the sign of subtraction*. A carpenter bought 300 pounds of nails for a job, and he had 28 pounds left when he finished. How many pounds of nails did he use to complete the job?

The problem:

$$300 - 28 = ?$$

The solution:

$$\begin{array}{r} 300 \\ \underline{-\ 28} \\ 272 \end{array} \text{ pounds}$$

Note that the numbers must be aligned as they were for addition.

Example *Use of the sign of division*. A carpenter's pickup truck gets an average of 17 miles per gallon of gasoline. How many gallons of gasoline would be required for him to travel 2040 miles in his truck?

The problem:

$$2040 \div 17 = ?$$

The solution:

```
      120
17 | 2040
     17
      34
      34
       00
```

Example *Use of the sign of equality*. A road contractor finds that he can lay the same amount of paving in 12 days using a 6-man crew that he can lay in 8 days using a 9-man crew. Express this statement as an equation.

$$12 \times 6 = 9 \times 8 \text{ or } 72 = 72$$

Fractions

A fraction indicates that a number or a unit has been divided into a certain number of equal parts, and shows how many of these parts are to be considered. Two forms of fractions are in common usage: the *decimal* (which is expressed in the tenths system) and the *common fraction*. The common fraction is written by using two numbers, one written over or alongside the other with a line between them. The lower (or second) number, called the *denominator*, indicates the number of parts into which the unit has been divided, and the upper (or first) number, called the *numerator*, indicates the number of parts to be considered. In the fraction $^2/_3$, the denominator shows that the unit is divided into 3 parts, whereas the numerator indicates that 2 parts are being considered (see Figure 5-3).

If the quantity indicated by the fraction is less than 1 (such as $^1/_2$, $^3/_4$, or $^5/_6$), it is called a *proper fraction*. If the quantity indicated by the fraction is equal to or greater than 1 (such as $^3/_3$, $^5/_4$, or $^7/_6$), it is called an *improper fraction*. When a whole number and a proper fraction are combined (such as $2^1/_4$, $6^1/_2$), it is called a *mixed number*.

Addition of Fractions

Fractions cannot be added without first reducing them to a common denominator.

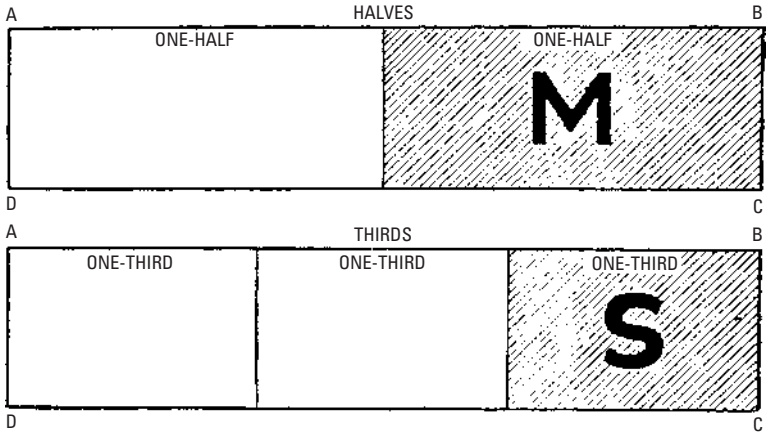

Figure 5-3 Graphic representations of fractional parts. The figures show a rectangle ABCD, representing a unit divided into two equal parts or halves and into three equal parts or thirds. The shaded section M (or one half) is larger than the shaded section S (or one third).

Example Add $^3/_4 + {}^2/_9 + {}^2/_3 + {}^7/_{12}$

To find the common denominator, place the denominators in a row, separated by dashes. Divide them by a prime number that will divide into at least two of them without leaving a remainder, and bring down the denominators with the dividends that did not contain the divisor without a remainder. Repeat this process as often as possible until there are no two numbers remaining that can be divided by the same number. Then, multiply the divisors and the remainders together, and the result will be the smallest common denominator.

The solution:

$$\begin{array}{l} 2\ \ )\ \underline{4-9-3-12} \\ 3\ \ )\ \underline{2-9-3-6} \\ 2\ \ )\ \underline{2-3-1-2} \\ \quad\ \ 1-3-1-1 \end{array}$$

The common denominator will then be $2 \times 3 \times 2 \times 3 = 36$. The fractions are then reduced to the common denominator of 36 by multiplying the numerator and the denominator by the same number that will produce 36 in the denominator, as follows:

$$^{3}/_{4} = {}^{3}/_{4} \times {}^{9}/_{9} = {}^{27}/_{36}$$

$$^{2}/_{9} = {}^{2}/_{9} \times {}^{4}/_{4} = {}^{8}/_{36}$$

$$^{2}/_{3} = {}^{2}/_{3} \times {}^{12}/_{12} = {}^{24}/_{36}$$

$$^{7}/_{12} = {}^{7}/_{12} \times {}^{3}/_{3} = {}^{21}/_{36}$$

The sum of the fractions is as follows:

$$\frac{27 + 8 + 24 + 21}{36} = {}^{80}/_{36} \text{ or } {}^{20}/_{9}$$

Multiplication of Common Fractions

A fraction may be multiplied by a whole number by multiplying the numerator of the fraction by that number.

Example If $^{3}/_{4}$ of a keg of nails is used for siding a garage, how many kegs of nails will be used when siding eight similar garages?

The solution:

$$8 \times {}^{3}/_{4} = {}^{24}/_{4}$$

A fraction may be simplified by dividing both the numerator and the denominator by the same number, and its value will not be affected.

Example Divide both the numerator and the denominator of the improper fraction $^{24}/_{4}$ by 4.

The result will be $^{6}/_{1}$ or 6. Therefore, six kegs of nails will be required to put the siding on the eight garages.

Fractions may be multiplied by fractions by multiplying their numerators together and their denominators together.

Example Multiply

$$^{2}/_{5} \times {}^{1}/_{4} \times {}^{5}/_{12} = \frac{2 \times 1 \times 5}{5 \times 4 \times 12} = \frac{10}{240} = \frac{1}{24}$$

Multiplication of Fractions by Cancellation

This may readily be done because any factor below the line may be divided by any factor above the line, and any factor above the line may be divided by any factor below the line, without altering the overall value of the expression.

Example The problem:

$$^{12}/_{30} \times {}^{14}/_{56} \times {}^{10}/_{24}$$

Now write it in this form:

$$\frac{\begin{array}{c}1 \times 1 \times 1\\ \cancel{12} \times \cancel{14} \times \cancel{10}\end{array}}{\begin{array}{c}\cancel{30} \times \cancel{56} \times \cancel{24}\\ 3 \times 4 \times 2\end{array}}$$

The 30 below the line may be divided by the 10 above the line with the result of 3. The 56 below the line may be divided by the 14 above the line with the result of 4. The 24 below the line may be divided by the 12 above the line with the result of 2. The result of the cancellation, then, is as follows:

$$\frac{1}{3 \times 4 \times 2} = \frac{1}{24}$$

Division of Fractions

Fractions may be divided by whole numbers by dividing the numerator by that number or by multiplying the denominator by that number.

Example

$$\frac{7}{8} \div 7 = \frac{1}{8}$$

$$\frac{7}{8} \div 7 = \frac{7}{56} = \frac{1}{8}$$

Fractions may be divided by fractions by inverting the divisor and multiplying.

Example

$$\frac{7}{8} \div \frac{2}{7} = \frac{7}{8} \times \frac{7}{2} = \frac{49}{16}$$

which is the mixed number $3^1/_{16}$.

Subtraction of Fractions

Fractions cannot be subtracted from fractions without first reducing them to a common denominator, as is done for the addition of fractions.

Example The problem:

$$^5/_6 - {}^{13}/_{16}$$

Finding the least common denominator,

$$\begin{array}{r|l} 2 & 16 - 6 \\ \hline & 8 - 3 \end{array} \qquad 2 \times 8 \times 3 = 48$$

$$\frac{13}{16} = \frac{13}{16} \times \frac{3}{3} = \frac{39}{48}$$

$$\frac{5}{6} = \frac{5}{6} \times \frac{8}{8} = \frac{40}{48}$$

$$\frac{40 - 39}{48} = \frac{1}{48}$$

To subtract a mixed number from another mixed number, it is usually most convenient to reduce both numbers to improper fractions and then proceed as shown in the last example.

To subtract a mixed number from a whole number, borrow 1 from the *minuend* (or upper number), and reduce the 1 to an improper denominator of the fraction in the *subtrahend* (or lower number), thereby reducing the whole number by 1. Then make the subtraction in the normal manner.

Example The problem:

$$14 - 6^{7}/_{8}$$

The solution:

$$14 - 6^{7}/_{8} = 13^{8}/_{8} - 6^{7}/_{8} = 7^{1}/_{8}$$

Applications of Cancellation

There are countless applications where this method will save appreciable time and work, but care and thought must be given to the proper arrangement of the fractional expression if there are many factors. Also, it must be remembered that if addition or subtraction signs appear, cancellation *may not* be used.

Example A circular saw has 75 teeth with a 1-inch spacing between each tooth. To do satisfactory work, the rim of the saw should travel at approximately 9000 feet per minute. How many revolutions should this saw make per minute? (Hint: 75 inches = $^{75}/_{12}$ feet.) The solution:

$$\frac{\overset{120}{\cancel{9000}} \times 12}{\cancel{75}} = 120 \times 12 = 1440 \text{ rpm}$$

Example If you go to a bank and borrow $1000 to purchase a truck, how much will the interest be, at 6 percent per annum (a theoretical example only), on the money you borrow for 1 year and 3 months? (Hint: 1 year 3 months = 15 months.)

The solution:

$$\frac{\overset{5}{\cancel{6}} \times \overset{\overset{1}{\cancel{10}}}{\cancel{1000}} \times 15}{\underset{\underset{2}{\cancel{12}}}{\cancel{100}}} = 5 \times 15 = \$75$$

Example If eight men in fifteen 8-hour days can throw 1000 cubic yards of gravel into wheelbarrows, how many men will be required to throw 2000 cubic yards of gravel into wheelbarrows in 20 days of 6 hours each?
The solution:

$$\frac{\overset{3}{\cancel{15}} \times \overset{2}{\cancel{8}} \times \overset{4}{\cancel{8}} \times \overset{2}{\cancel{2000}}}{\underset{\underset{4}{\cancel{20}}}{\cancel{6}} \times \underset{2}{\cancel{1000}}} = 2 \times 4 \times 2 = 16 \text{ men}$$

Example A building that is 30 feet × 30 feet with a 10-foot ceiling contains approximately 700 pounds of air. What will be the weight of the air in a room 120 feet long, 90 feet wide, and 16 feet high?
The solution:

$$\frac{\overset{4}{\cancel{120}} \times \overset{3}{\cancel{90}} \times 16 \times \overset{70}{\cancel{700}}}{\cancel{30} \times \cancel{30} \times \cancel{10}} = 4 \times 3 \times 16 \times 70 = 13{,}440 \text{ pounds}$$

Decimals

Decimal means numbering that proceeds by *tens*, and *decimal fractions* (usually simply called *decimals*) are formed when a unit is divided into ten parts. When decimals are written, the point where the numbers start is called the *decimal point*. To the left of the decimal point, the numbers read in the regular manner—units, tens, hundreds, thousands, ten thousands, hundred thousands, millions, and so forth. To the right of the decimal point, the figures are fractional, reading, from the point, tenths, hundredths, thousandths, ten thousandths, hundred thousandths, millionths, and so forth. (see Figure 5-4).

The common fraction $^{6}/_{10}$ can be expressed decimally as .6, and the fraction $^{105}/_{1000}$ can be written as .105. The mixed number $106\,^{6}/_{100}$ may be expressed decimally as 106.06. The decimal .6 is read "six-tenths," the same as the common fraction $^{6}/_{10}$, and the decimal 106.06 is read "one hundred six *and* six-hundredths," the same as the mixed number.

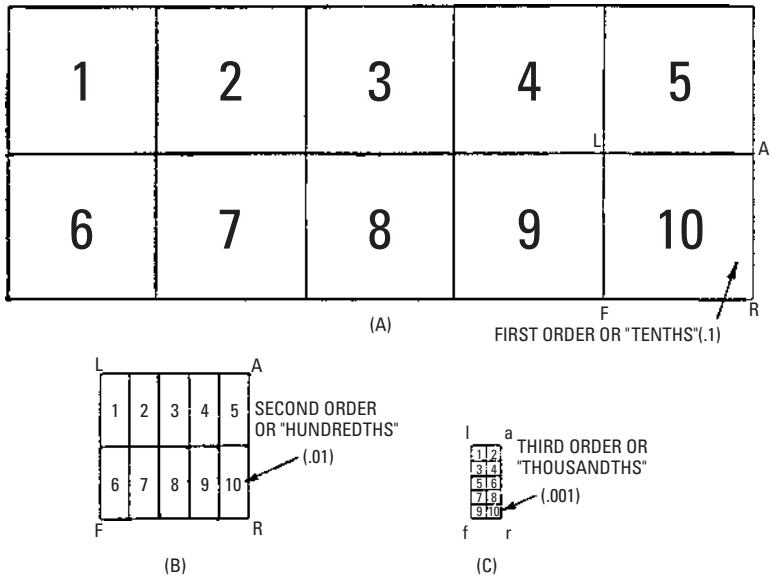

Figure 5-4 Graphic representation of decimal fractions. (A) A unit divided into ten parts—1st order or tens. (B) One of the tens as LARF, divided into ten parts—2nd order or hundredths. (C) One of the hundredths as larf, divided into ten parts—3rd order or thousandths. Similarly the process of division may be continued indefinitely.

It is not necessary to place a zero before the decimal point, as 0.06, but it is sometimes convenient when it is necessary to align a column of decimals for addition. The decimal points must *always* be aligned for addition, as shown in the following example:

$$\begin{array}{r} 0.6 \\ 6.29 \\ +10.72 \\ \hline 17.61 \end{array}$$

The position of the decimal point in the sum is established directly under the column of decimal points above the line.

Note

The number 327 is *not* read "three hundred *and* twenty-seven," but "three hundred twenty-seven." However, the decimal 300.27 is read

"three hundred *and* twenty-seven hundredths." The decimal .327 is read "three hundred twenty-seven thousandths," whereas the decimal 300.027 is read "three hundred *and* twenty-seven thousandths."

Reduction of Common Fractions to Decimals

Divide the numerator by the denominator, adding zeros and carrying the division to as many decimal places as are necessary or desirable (see Figure 5-5).

Example Reduce the common fraction $^{21}/_{32}$ to a decimal.

```
      0.65625
32 | 21.00000
     192
      180
      160
       200
       192
         80
         64
         160
         160
          00
```

Count off as many decimal places in the quotient as those in the dividend *exceed* those in the divisor. The quotient is .65625.

Subtraction of Decimals

Align the decimal points in the minuend and the subtrahend as shown for addition, and proceed as explained in subtraction of whole numbers. The decimal point in the remainder is placed in exact alignment with those decimal points above the line.

Example The problem:

$$167.02 - 27.267$$

The solution:

```
  167.020
– 27.267
  139.753
```

Note that it is necessary to add a zero to the decimal 167.02 in order to make the subtraction, but this does not change its value ($^{20}/_{1000}$ is the same value as $^{2}/_{100}$).

					1/64	.015625
				1/32		.03125
					3/64	.046875
			1/16			.0625
					5/64	.078125
				3/32		.09375
					7/64	.109375
		1/8				.125
					9/64	.140625
				5/32		.15625
					11/64	.171875
			3/16			.1875
					13/64	.203125
				7/32		.21875
					15/64	.234875
	1/4					.250
					17/64	.265625
				9/32		.28125
					19/64	.296875
			5/16			.3125
					21/64	.328125
				11/32		.34375
					23/64	.359375
		3/8				.375
					25/64	.390625
				13/32		.40625
					27/64	.421875
			7/16			.4375
					29/64	.453125
				15/32		.46875
					31/64	.484375
1/2						.500
					33/64	.515625
				17/32		.53125
					35/64	.546875
			9/16			.5625
					37/64	.578125
				19/32		.59375
					39/64	.609375
		5/8				.625
					41/64	.640625
				21/32		.65625
					43/64	.671875
			11/16			.6875
					45/64	.703125
				23/32		.71875
					47/64	.734375
	3/4					.750
					49/64	.765625
				25/32		.78125
					51/64	.796875
			13/16			.8125
					53/64	.828125
				27/32		.84375
					55/64	.859375
		7/8				.875
					57/64	.890625
				29/32		.90625
					59/64	.921875
			15/16			.9375
					61/64	.953125
				31/32		.96875
					63/64	.984375
						1.0000

Figure 5-5 Fractions of an inch and decimal equivalents.

Multiplication of Decimals

Proceed as in multiplication of whole numbers, and count off as many decimal places in the product as there are in *both* the multiplier and multiplicand.

Example The problem:

$$1.76 \times 0.06$$

The solution:

$$\begin{array}{r} 1.76 \\ \underline{\times\ 0.06} \\ 0.1056 \end{array}$$

Compound Numbers

A compound number expresses units of two or more denominations of the same kind, such as 5 yards, 1 foot, and 4 inches. The process of changing the denomination in which a quantity is expressed without changing its value is called *reduction*. Thus, 1 yard and 2 inches = 38 inches, 25 inches = 2 feet and 1 inch, and so forth, are examples of reduction. Reduction problems occur and are explained with the various measures and weights.

Reduction Descending

To reduce a compound number to a lower denomination, multiply the largest units in the given number by the number of units in the next lower denomination, and add to the product the units of that denomination in the given number. Continue this process until the original number is reduced as far as desired. For an explanation of this rule, see the following example.

Example The problem: Reduce the quantity 6 yards, 2 feet, 7 inches to inches.

The solution:

	6	yards
Multiply.....	$\times$ 3	
	18	feet
Add.....	+ 2	
	20	feet
Multiply.....	$\times$ 12	
	240	inches
Add.....	+ 7	
Total.....	247	inches

Reduction Ascending

To reduce a number of small units to units of larger denominations, divide the number by the number of units in a unit of the next higher denomination. The quotient is in the higher denomination and the remainder, if any, is in the lower. Continue this process until the number is reduced as far as is desired.

Example The problem: Reduce 378 inches to a quantity of yards, feet, and inches.

The solution:

```
12 ) 378
     3)31 feet, 6 inches remainder
       10 yards, 1 foot remainder
```

Therefore, 378 inches = 10 yards, 1 foot, 6 inches.

Ratios

By definition, a *ratio* is the relation of one number to another as obtained by dividing the first number by the second. Thus, the ratio of 2 to 4 is expressed as 2 : 4. The symbol : is read "to" in the case of a ratio and "is to" in the case of a proportion. It is equivalent to "divided by." Hence:

$$2 : 4 = {}^{1}/_{2}$$

The first term of a ratio is the *antecedent*, and the second term is the *consequent*, thus:

antecedent		consequent
2	:	4

Since a ratio is essentially a fraction, it follows that if both terms are multiplied or divided by the same number, the value of the ratio is not altered. Thus:

$$2 : 4 = 2 \times 2 : 4 \times 2 = 2 \div 2 : 4 \div 2$$

Two quantities of different kinds cannot form the terms of a ratio. Thus, no ratio can exist between \$5 and 1 day, but a ratio can exist between \$5 and \$2 or between 1 day and 10 days.

Proportion

When two ratios are equal, the four terms form a *proportion*. A proportion is, therefore, expressed by using the sign = or : : between two ratios. Thus:

(expressed) 4 : 8 :: 2 : 4

(read) 4 is to 8 as 2 is to 4

The same proportion is also expressed as follows:

$$^4/_8 = {}^2/_4$$

The first and last terms of a proportion are called the *extremes*, and the middle terms are called the *means*. Thus:

$$4 : 8 :: 2 : 4$$

The product of the extremes equals the product of the means. Thus, in proportion

$$4 : 8 = 2 : 4$$

$$4 \times 4 = 8 \times 2$$

Since the equation is not altered by dividing both sides by the same number, the value of any term can be obtained as follows:

$$\frac{4 \times 4}{4} = \frac{8 \times 2}{4}$$

$$4 = 2 \times 2 = 4$$

Rule of Three

When three terms of a proportion are given, the method of finding the fourth term is called the rule of three.

Example The problem: If 5 bundles of shingles cost \$100, what will 25 bundles cost?

The solution: Let X represent the unknown term in the proportion, and, remembering that each ratio must be made up of like quantities:

$$5 \text{ bundles} : 25 \text{ bundles} = 100\ (\$) : X\ (\$)$$

Multiply the extremes by the means:

$$5 \times X = 25 \times 100$$

$$X = \frac{25 \times 100}{5} = \$500$$

Percentage

By definition, *percentage* means the rate per 100, or the proportion in 100 parts. Therefore $^1/_{100}$ of a number is called 1 percent, $^2/_{100}$ is 2 percent, and so forth. The symbol % is read as "percent." Thus 1%, 2%, and so forth. Carefully note the following explanation with respect to the symbol %. The notation 5% means $^5/_{100}$, which, when reduced to a decimal (as is necessary when making a calculation),

becomes .05, but .05% means $^{.05}/_{100}$, which, when reduced to a decimal, becomes .0005 (that is, $^{5}/_{100}$ of 1%).

If the decimal has more than two places, the figures that follow the hundredths place signify parts of 1%.

Example The problem: If the list price of shingles is $90 per 1000, what is the net cost for 1000 shingles with a 5 percent discount for cash?

The solution: Reduce % rate to a decimal.

$$5\% = {}^{5}/_{100} = .05$$

Multiply decimal by list price.

$$\$90 \times .05 = \$4.50$$

Subtract product obtained from list price.

$$\$90 - \$4.50 = \$85.50$$

Powers of Numbers (Involution)

The word *involution* means the multiplication of a quantity by itself any number of times, and a *power* is the product arising from this multiplication. Involution, then, is the process of raising a number to a given power. The *square* of a number is its second power, the *cube* is its third power, and so forth. Thus:

$$\text{square of } 2 = 2 \times 2 = 4$$

$$\text{cube of } 2 = 2 \times 2 \times 2 = 8$$

The power to which a number is raised is indicated by a small superior figure called an *exponent*. Thus, in Figure 5-6, the exponent indicates the number of times the number, or root, has been multiplied by itself.

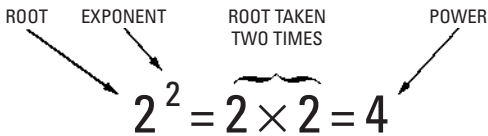

Figure 5-6 The root, exponent, and power of a number.

Roots of Numbers (Evolution)

The word *evolution* means the operation of extracting a root. The root is a factor that is repeated to produce a power. Thus, in the equation $2 \times 2 \times 2 = 8$, 2 is the root from which the power 8

is produced. This number is indicated by the symbol $\sqrt{\ }$, called the *radical sign*, which, when placed over a number, means that the root of that number is to be extracted. Thus:

$\sqrt{4}$ *means that the square root of 4 is to be extracted*

The *index* of the root is a small figure that is placed over the radical sign that denotes what root is to be taken. Thus, $\sqrt[3]{9}$ indicates the cube root of 9 and $\sqrt[4]{16}$ indicates the extraction of the fourth root of 16. When there is no index given, the radical sign alone always means the *square root* is to be extracted from the number under the radical sign.

Sometimes the number under the radical sign is to be raised to a power before extracting the root, as follows:

$$\sqrt[3]{4^3} = \sqrt[3]{4 \times 4 \times 4} = \sqrt[3]{64}$$

Example The problem: Extract the square root of 186,624.

The solution:

$$\begin{array}{rl} & \sqrt{18'66'24} \qquad 432 \\ & \quad 16 \\ 83 & \sqrt{266} \\ & \quad\ \ 249 \\ 862 & \sqrt{1724} \\ & \quad\ \ \underline{1724} \end{array}$$

From the decimal point, count off the given number into periods of two places each. Begin with the last period counted off (18). The largest square that can be divided into 18 is 4. Put this down in the quotient, and put the square (16) under the 18. Write down the remainder (2), and bring down the next period (66). Multiply 4 (in the quotient) by 2 for the first number of the next divisor; 8 goes into 26 three times. Place 3 after 4 in the quotient and also after 8 in the divisor. Multiply the 83 by 3, placing the product 249 under 266, and subtract, obtaining the remainder 17. Bring down the last period (24), and proceed as before, obtaining 432 as the square root of 186,624.

Extracting the cube root of a number is a more complicated (though similar) process, as indicated by the following procedure:

1. Separate the number into groups of three figures each, beginning at the decimal point.

2. Find the greatest cube that can be divided into the left-hand group, and write its root for the first figure of the required root.
3. Cube this root, subtract the result from the left-hand group, and annex the next group to the remainder for a dividend.
4. For a partial divisor, take three times the square of the root already found, considered as hundreds, and divide the dividend by it. The quotient (or the *quotient diminished*) will be (or be close to) the second figure of the root.
5. To this *partial divisor*, add three times the product of the first figure of the root, considered as tens, by the second figure, and to this add the square of the second figure. This sum will be the *complete divisor*.
6. Multiply the complete divisor by the second figure of the root, subtract the product from the dividend, and annex the next group to the remainder for a new dividend.
7. Proceed in this manner until all the groups have been annexed. The result will be the cube root required, as shown in the following example.

Example The problem: Extract the cube root of the number 50,653.

The solution:

```
                 ∛50′653.  ( 37
                  27
2700              23 653
 630              23 653
  49
3379
```

Therefore, the cube root of 50,653 is 37.

Measures

To *measure* is the act or process of determining the extent, quantity, degree, capacity, dimension, volume, and so forth, of a substance by comparing it with some fixed standard, which is usually fixed by law. A measure may relate to any of these standards. There are many kinds of measures, and practically all of them are standard, but standards vary in different countries. The measures mentioned in this text are all U.S. standards unless designated otherwise. The study of measurements is sometimes called *mensuration*.

Among the many kinds of measures are the following:

- *Linear*—Measures of length
- *Square*—Used to measure areas
- *Cubic*—Used to measure volume, or volumetric contents
- *Weight*—Many systems of weights are standard
- *Time*—Almost standardized all over the world
- *Circular* or *angular*—The same all over the world

Linear Measure

Table 5-2 shows linear measurement (long) equivalents.

Table 5-2 Long Measure

Measure	*Equivalent*	*Equivalent*
12 inches	1 foot	
3 feet	1 yard	36 inches
$5^1/_2$ yards	1 rod	$16^1/_2$ feet
40 rods	1 furlong	660 feet
8 furlongs	1 mile	5280 feet
3 miles	1 league (land)	

The furlong is practically never used, except at racetracks and in some athletic events.

Table 5-3 shows land survey measurement equivalents.

Table 5-3 Land Surveyor's Measure

Measure	*Equivalent*	*Equivalent*
7.92 inches	1 link	
100 links	1 chain	66 feet
10 chains	1 furlong	660 feet
80 chains	1 mile	5280 feet

The use of the surveyor's chain, or Gunter's chain, was abandoned in the late 1800s and was superseded by the steel tape, which is much more accurate. The chain (meaning 66 feet) is still used by the U.S. General Land Office, however, when surveying very old deeds. The standard surveyor's tape is often called, from habit, a *chain*. It is

100 feet long and is graduated in feet except for the last foot, which is divided into tenths and hundredths of a foot.

Table 5-4 shows nautical measurement equivalents.

Table 5-4 Nautical Measure (U.S. Navy)

Measure	*Equivalent*
6 feet	1 fathom
120 fathoms	1 cable length
The International Nautical Mile*	6076.1033 feet
3 nautical miles	1 marine league

*Adopted in 1954.

The knot is a measure of speed, not of length, and is equivalent to 1 nautical mile per hour. A speed of 16 knots is equal to 16 nautical miles per hour.

Square Measure

Square measure is used to measure areas. In most (but not all) cases, linear units are used to measure the two dimensions, length and width, and their product is the area in square units. Expressed as an equation:

$$\text{length} \times \text{width} = \text{area}$$

The two dimensions, length and width, must be measured in the same units, but any unit of linear measurement may be used. If inches are multiplied by inches, the result will be in square inches; if feet are multiplied by feet, the result will be in square feet, and so forth (see Figure 5-7).

For the small areas commonly found in everyday life (such as tabletops or shelves), the unit most commonly used is the square inch. Plywood and lumber are commonly sold by the square foot. Carpets and other floor coverings and materials and ceilings are measured in square yards. The carpenter measures roofing by the square of 10 × 10 feet, or 100 square feet. Tracts of land are usually measured in acres or, for large areas, in square miles.

Table 5-5 shows square measure equivalents.

Cubic Measure

Cubic measure is used to determine or appraise volumes. Three dimensions are involved (length, width, and height) and their product is volume. Expressed as an equation:

$$\text{length} \times \text{width} \times \text{height} = \text{volume}$$

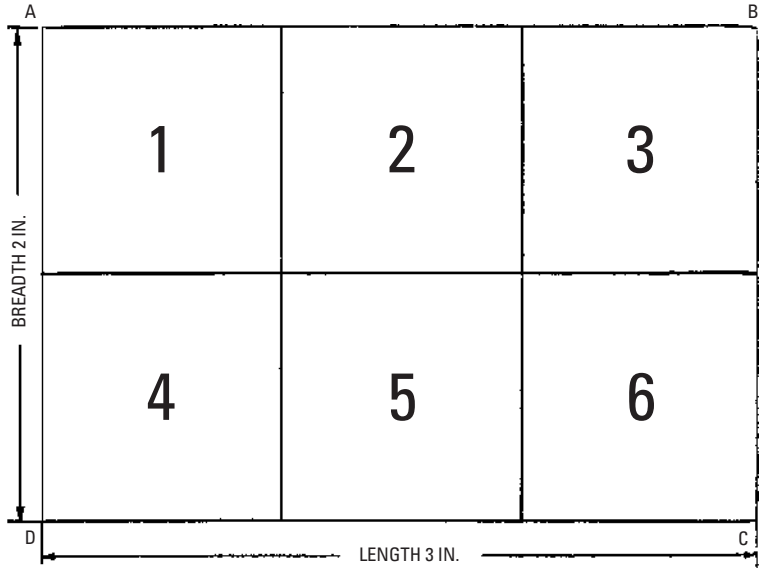

Figure 5-7 Square measure. If the rectangle ABCD measures 2 inches on one side and 3 inches on the other, and lines are drawn at each inch division, then each of the small squares will have an area of 1 square inch and the area of the rectangle will be area ABCD = breadth × length = 2 × 3 = 6 square inches.

Table 5-5 Square Measure

Measure	*Equivalent*	*Equivalent*
144 square inches	1 square foot	
9 square feet	1 square yard	
$30^1/_4$ square yards	1 square rod	272.25 square feet
160 square rods	1 acre	4840 square yards or 43,560 square feet
640 acres	1 square mile	3,097,600 square yards
36 square miles	1 township	

As with square measure, the usual linear units (inches, feet, and yards) are ordinarily used to measure these three dimensions. Most small measurements of capacity (such as small shipping cases or small cabinets) are measured in cubic inches. The contents of

buildings, their cubage, are ordinarily expressed in cubic feet. Earthwork (either excavated and loose, or in place) is expressed in cubic yards (see Figure 5-8).

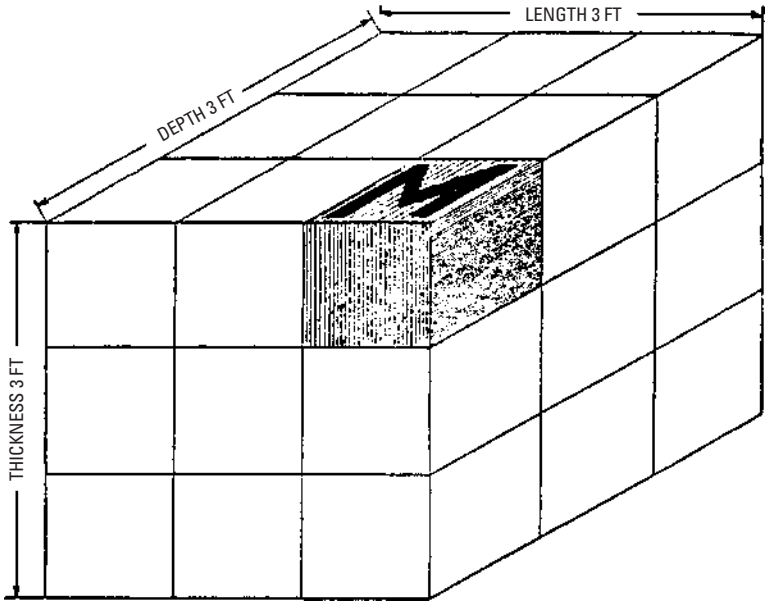

Figure 5-8 Cubic measure. If each side of the cube measures 3 feet and it is cut as indicated by the lines, each little cube as M, will have each of its sides 1 foot long and will contain 1 × 1 × 1 = 1 cubic foot. Accordingly, the large cube will contain 3 × 3 × 3 = 27 cubic feet or 1 cubic yard.

Table 5-6 shows cubic measure of volume equivalents.

Table 5-6 Cubic Measures of Volume

Measure	*Equivalent*
1728 cubic inches	1 cubic foot
27 cubic feet	1 cubic yard

Dry Measure

Quantities of loose, granular materials (such as grains, some fruits, and certain vegetables) are measured in arbitrary units that, in turn,

are defined by means of cubic measures of volume, usually in cubic inches. Their value is sometimes fixed by law.

Table 5-7 shows units of dry measure equivalents for the United States and Table 5-8 shows units of dry measure equivalents for Great Britain and Canada.

Table 5-7 Dry Measure (United States)

Measure	*Equivalent*	*Equivalent*
2 pints	1 quart	67.2 cubic inches
8 quarts	1 peck	537.61 cubic inches
4 pecks	1 bushel	2150.42 cubic inches

Table 5-8 Dry Measure (British and Canadian)

Measure	*Equivalent*	*Equivalent*
1 gallon	.5 peck	277.42 cubic inches
4 pecks	1 bushel	2219.23 cubic inches

The British dry quart is not often used. It is equal to 69.35 cubic inches, or 1.032 U.S. dry quarts.

The weight, rather than the volume, of grains is the standard fixed by the U.S. government (Table 5-9).

Table 5-9 Grain Measure (U.S.)

Measure	*Equivalent*
1 bushel of wheat	60 pounds
1 bushel of barley	48 pounds
1 bushel of oats	32 pounds
1 bushel of rye	56 pounds
1 bushel of corn (shelled)	56 pounds

Board or Lumber Measure

Timbers and logs are measured in *board* or *lumber measure*. The board foot is 1 foot wide, 1 foot long, and 1 inch thick, thereby containing 144 cubic inches. In the retail market, all lumber that is less than 1 inch thick is called one inch. At the sawmills, the full sizes govern the thickness of the saw kerfs; usually about $^1/_4$ inch is allowed for and accounted as sawing loss. Actual finished (dressed)

sizes of common lumber and the dimension and timbers for pine are as follows:

- The standard dressed thickness of 1-inch boards is $^3/_4$ inch.
- The standard thickness of 2-inch dimension boards is $1^1/_2$ inches.
- The standard dressed widths of lumber 2 inches thick and less are $^1/_2$ inch less for widths under 8 inches and $^3/_4$ inch less for 8-inch widths and wider.
- The standard dressed widths and thicknesses for lumber and timbers are $^1/_2$ inch less both ways under 8 inches wide and$^3/_4$ inch for 8-inch widths and over. So, a 2-inch × 8-inch board would be $1^1/_2$ inches × $7^1/_4$ inches. A 2-inch × 10-inch board would be $1^1/_2$ inches × $9^1/_4$ inches.

Liquid Measure

Liquid measure is used to measure various liquids such as oils, liquors, molasses, and water. Table 5-10 shows liquid measure equivalents.

Table 5-10 Liquid Measure

Measure	*Equivalent*	*Equivalent*
4 gills	1 pint	28.875 cubic inches
2 pints	1 quart	57.75 cubic inches
4 quarts	1 gallon	231 cubic inches

There is no legal standard barrel in the United States. By custom, a barrel of water is understood to be $31^1/_2$ gallons. The British barrel is generally 36 Imperial gallons. Crude oil is often disposed of at the wells in barrels of 50 gallons, whereas refined oils are marketed in barrels of 48 gallons. Owing to this lack of uniformity, it is safest to specify "barrels of 50 gallons," or something of that nature, to avoid misunderstanding. The barrel is sometimes used as a dry measure unit of varying value. For Portland cement, 4 bags = 1 barrel = 4 cubic feet = 376 pounds.

Measures of Weight

The simplest definition of *weight* is the force with which a body is attracted toward the earth. It is a quantity of heaviness. The three systems (or standards) of weights used in the United States are:

- *Avoirdupois*—Used for almost all ordinary purposes (see Table 5-11)

Table 5-11 Avoirdupois Weights

	Measure	*Equivalent*
U.S.	16 drams	1 ounce
	16 ounces	1 pound
	100 pounds	1 hundredweight
	20 hundredweights	1 ton
	Measure	Equivalent
England	14 pounds	1 stone
	112 pounds	1 hundredweight
	20 hundredweight	1 ton
	2240 pounds	

Note: The 2240-pound ton is sometimes used in the United States for weighing coal at the mines and at Customs houses for evaluating shipments from England.

- *Troy*—Used in weighing precious metals and jewels (see Table 5-12)
- *Apothecaries*—Used by pharmacists when compounding drugs (see Table 5-13)

Table 5-12 Troy Weights

Measure	*Equivalent*
3.086 grains	1 carat
24 grains	1 pennyweight
20 pennyweights	1 ounce
12 ounces	1 pound

Table 5-13 Apothecaries Weights

Measure	*Equivalent*
20 grains	1 scruple
3 scruples	1 dram
8 drams	1 ounce
12 ounces	1 pound

This standard of weights is fast becoming obsolete, although pharmacists must be familiar with it. Manufacturing pharmacists and chemists are rapidly changing to the metric weights, using the metric *gram* as a basis instead of the apothecaries' scruple (1 scruple = 1.296 grams).

Time Measure

Time is defined as measurable duration (see Table 5-14). It is the period during which an action or process continues. The basis (or standard) used in our ordinary determination of time is the *mean solar day*, beginning and ending at mean midnight. The word mean as used here simply means average. The direct ray of the sun does not move in an exact and uniform path around the equator.

Table 5-14 Time Measure

Measure	*Equivalent*
60 seconds	1 minute
60 minutes	1 hour
24 hours	1 day
7 days	1 week
30 days (commonly)	1 month
365 days	1 year
10 years	1 decade
100 years	1 century
1000 years	1 millennium

The length of an *astronomical* year is 365 days, 5 hours, 48 minutes, and 45.51 seconds, or approximately 365 $^1/_4$ days. This makes it necessary to add 1 day every 4 years, thus making the leap year 366 days.

Circular Measure

This measure is used in astronomy, land surveying, navigation, and in measuring angles of all kinds. Circles of all sizes are divisible into degrees, minutes, and seconds (see Table 5-15). Note that a degree is *not* a measurement of length. It is $^1/_{360}$ of the circumference of a circle with any radius. With widespread use of calculators and the need for accuracy in missiles, the degree has now been divided

Table 5-15 Circular Measure

Measure	*Equivalent*
60 seconds	1 minute
60 minutes	1 degree
360 degrees	1 circle

in decimal form rather than minutes and seconds. Most calculators can carry the degree out to 6 places after the decimal point

The Metric System

The base, or fundamental, unit in the metric system is the *meter*. The meter is defined as the distance between two scribed marks on a standard bar made of platinum-iridium kept in the vaults of the International Bureau of Weights and Measures, near Paris, France. Of course, many other standard meter bars have been made from the measurement on this bar. It is permissible and official to use this measurement in the United States, and, in fact, the yard, the basis for the English system of measurement, has been defined as exactly $^{3600}/_{3937}$ meter, or 1 meter = 39.37 inches.

The advantage (and immeasurably greater convenience) of the metric system over the English system of units lies in the fact that it is expressed in tenths, thereby readily allowing the use of decimals. However, the American public is accustomed to the English units, and as recent experience indicates, the system should continue for a long time. The metric system is, of course, in common use all over the world with the exception of some English-speaking countries. The meter is used like the yard to measure cloth and short distances.

Units of other denominations are named by prefixing to the word meter the Latin numerals for the lower denominations and the Greek numerals for the higher denominations, as shown in Table 5-16.

Table 5-16 Denominations

Lower Denomination		*Higher Denomination*	
Greek	***Equivalent***	***Greek***	***Equivalent***
Deci	$^{1}/_{10}$	Deka	10
Centi	$^{1}/_{100}$	Hecto	100
Milli	$^{1}/_{1000}$	Kilo	1000
Micro	$^{1}/_{1,000,000}$	Myria	10,000
		Mega	1,000,000

Therefore, 1 decimeter = $^{1}/_{10}$ of a meter, 1 millimeter = $^{1}/_{1000}$ of a meter, 1 kilometer = 1000 meters, and so forth. From this explanation of the metric prefixes, the linear equivalents shown in Table 5-17 can easily be understood.

Table 5-17 Metric Table of Linear Measure

Metric Denomination		*Meter*	*U.S. Value*
	1 millimeter	.001	.0394 inches
10 millimeters	1 centimeter	.01	.3937 inches
10 centimeters	1 decimeter	.1	3.937 inches
10 decimeters	1 meter	1.	39.3707 inches 3.28 feet
10 meters	1 dekameter	10.	32.809 feet
10 dekameters	1 hectometer	100.	328.09 feet
10 hectometers	1 kilometer	1000.	.62138 miles
10 kilometers	1 myriameter	10,000.	6.2138 miles

The kilometer is commonly used for measuring long distances. The square meter (see Table 5-18) is the unit used for measuring ordinary surfaces, such as flooring or ceilings.

Table 5-18 Metric Table of Square Measure

Measure	*Equivalent*	*Equivalent*
100 square millimeters (mm^2)	1 square centimeter	0.15 + square inch
100 square centimeters (cm^2)	1 square decimeter	15.5 + square inches
100 square decimeters (dm^2)	1 square meter (m^2)	1.196 + square yards

The acre is the unit of land measure and is defined as a square whose side is 10 meters, equal to a square dekameter, or 119.6 square yards (see Table 5-19).

Table 5-19 Metric Table of Land Measure

Measure	*Equivalent*	*Equivalent*
1 centiare (ca)	1 square meter	1.196 square yards
100 centiares (ca)	1 acre	119.6 square yards
100 ares (A)	1 hectare	2.471 acres
100 hectares (ha)	1 square kilometer	0.3861 square miles

The cubic meter is the unit used for measuring ordinary solids, such as excavations or embankments (see Table 5-20).

Table 5-20 Metric Table of Cubic Measure

Measure	*Equivalent*	*Equivalent*
1000 cubic millimeters (mm^3)	1 cubic centimeter	0.061 + cubic inches
1000 cubic centimeters (cm^3)	1 cubic decimeter	61.026 + cubic inches
1000 cubic decimeters (dm^3)	1 cubic meter	35.316 + cubic feet

The liter is the unit of capacity, both of liquid and of dry measures, and is equivalent to a vessel whose volume is equal to a cube whose edge is $^1/_{10}$ of a meter, equal to 1.0567 quarts liquid measure, and 0.9081 quart dry measure (see Table 5-21).

The hectoliter is the unit used for measuring liquids, grain, fruit, and roots in large quantities. The gram is the unit of weight equal to the weight of a cube of distilled water, the edge of which is $^1/_{100}$ of a meter, and is equal to 15.432 troy grains (see Table 5-22).

Geometry

By definition, *geometry* is that branch of mathematics that deals with space and figures in space. In other words, it is the science of the mutual relations of points, lines, angles, surfaces, and solids that are considered as having no properties except those arising from extension and difference of situation.

Lines

The two kinds of lines are straight and curved. A *straight line* is the shortest distance between two points. A *curved line* is one that changes its direction at every point. Two lines are said to be parallel when they have the same direction. A horizontal line is one parallel to the horizon or surface of the Earth. A line is perpendicular with another line when they are at right angles to each other. These definitions are illustrated in Figure 5-9.

Angles

An *angle* is the difference in direction between two lines proceeding from the same point (called the *vertex*). Angles are said to be *right* (90 degrees) when formed by two perpendicular lines (see

Table 5-21 Metric Table of Capacity

10 milliliters (ml.)	= 1 centiliter	= .0338 fluid ounce
10 centiliters (cl.)	= 1 deciliter	= .1025 cubic inch
10 deciliters (dl.)	= 1 liter	= 1.0567 liquid quart
10 liters (l.)	= 1 dekaliter	= 2.64 gallons
10 dekaliters (dl.)	= 1 hectoliter	= 26.418 gallons
10 hectoliters (hl.)	= 1 kiloliter	= 264.18 gallons
10 kiloliters (kl.)	= 1 myrialiter (ml.)	
1 myrialiter	= 10 cubic meters = 283.72 + bushels	= 2641.7 + gallons
1 kiloliter	= 1 cubic meter = 28.372 + bushels	= 264.17 gallons
1 hectoliter	= $^1/_{10}$ cubic meter = 2.8372 + bushels	= 26.417 gallons
1 decaliter	= 10 cubic decimeters = 9.08 quarts	= 2.6417 gallons
1 liter	= 1 cubic decimeter = .908 quart	= 1.0567 quart liquid
1 deciliter	= $^1/_{10}$ cubic decimeter = 6.1022 cubic inches	= .845 gallons
1 milliliter	= 10 cubic centimeters = .6102 cubic inches	= .338 fluid ounces
1 centiliter	= 1 cubic centimeter = .061 cubic inches	= .27 fluid dram

Figure 5-10A), *acute* (less than 90 degrees) when less than a right angle (see Figure 5-10B), and *obtuse* (more than 90 degrees) when greater than a right angle (see Figure 5-10C). All angles except right (or 90-degree) angles are called *oblique angles*.

Angles are usually measured in degrees (circular measure) (see Figure 5-10D). The *complement* of an angle is the difference between 90 degrees and the angle. The *supplement* of the angle is the difference between the angle and 180 degrees.

Plane Figures

The term *plane figures* means a plane surface bounded by straight or curved lines, and a *plane* (or *plane surface*) is one in which any straight line joining any two points lies wholly in the surface. Figure 5-11 defines a plane surface. There is a great variety of plane

Table 5-22 Metric Table of Weight Measure

Measure	*Equivalent*	*Equivalent*
10 milligrams (mg)	1 centigram	0.15432 + grains troy
10 centigrams (cg)	1 decigram	1.54324 + grains troy
10 decigrams (dg)	1 gram	15.43248 + grains troy
10 grams (g)	1 dekagram	0.35273 + ounce avoirdupois
10 dekagrams (Dg)	1 hectogram	3.52739 + ounces avoirdupois
10 hectograms (hg)	1 kilogram	2.20462 + pounds avoirdupois
10 kilograms (kg)	1 myriagram	22.04621 + pounds avoirdupois
10 myriagrams (Mg)	1 quintal	220.46212 + pounds avoirdupois
10 quintals	1 ton	2204.62125 + pounds avoirdupois

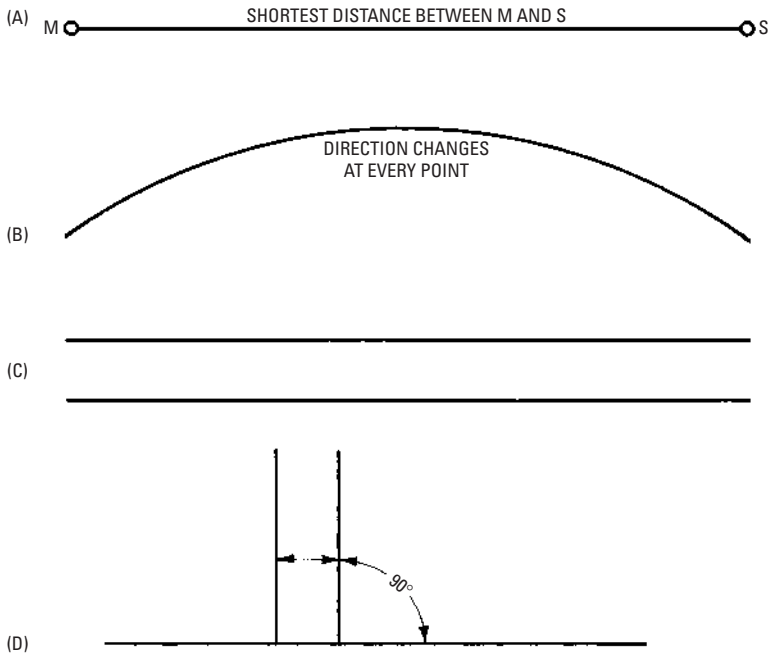

Figure 5-9 Various lines: (A) straight, (B) curved, (C) parallel, and (D) perpendicular.

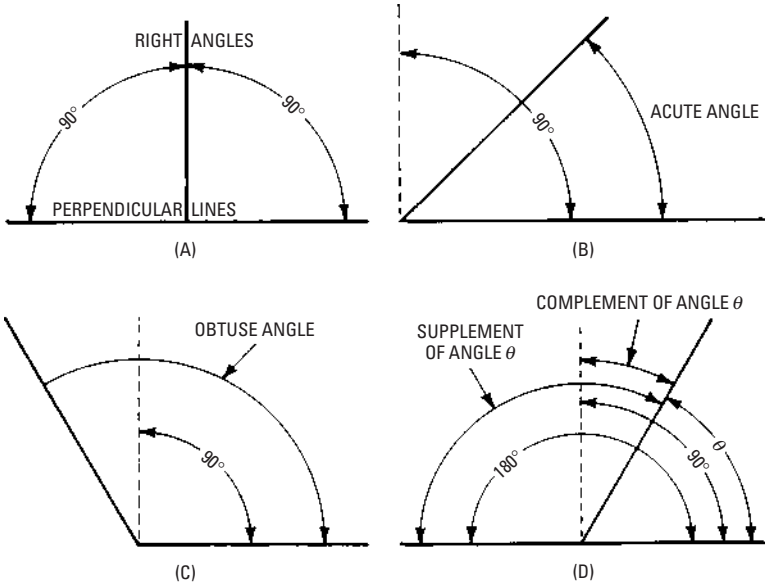

Figure 5-10 Various angles: (A) right, (B) acute, (C) obtuse, and (D) complement and supplement of an angle.

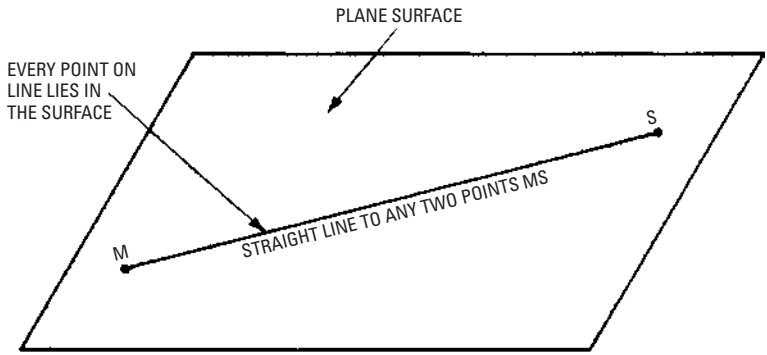

Figure 5-11 A plane surface means that every point on a straight line joining any two points in the surface lies in the surface.

figures, which are known as *polygons* when their sides are straight lines. The sum of the sides is called the *perimeter*. A regular polygon has all its sides and angles equal. Plane figures of three sides are known as triangles (see Figure 5-12), and plane figures of four sides are quadrilaterals. Figure 5-13 shows examples of these. Various plane figures are formed by curved sides and are known as circles, ellipses, and so forth, as shown in Figure 5-14. Figure 5-15 details the structure of the quadrilateral.

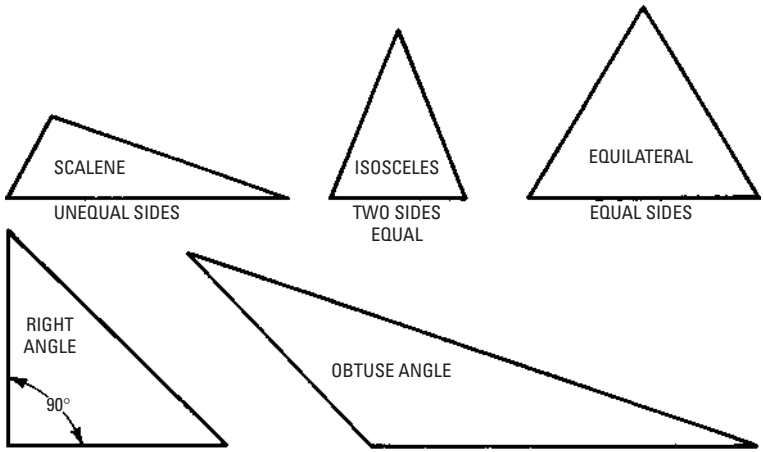

Figure 5-12 Various triangles. A triangle is a polygon having three sides and three angles.

Solids

Solids have three dimensions—length, width, and thickness. The bounding planes are called the *faces*, and the intersections are called the *edges*. A prism (see Figure 5-16) is a solid whose ends consist of equal and parallel polygons, and whose sides are *parallelograms*. The *altitude* of a prism is the perpendicular distance of its opposite sides or bases. A *parallelopipedon* is a prism that is bounded by six parallelograms; the opposite parallelograms are parallel and equal. A *cube* is a parallelopipedon whose faces are equal. One important solid is the *cylinder*, which is a body bounded by a uniformly curved surface and having its ends equal and forming parallel circles (see Figure 5-17). There are numerous other solids having curved surfaces (such as *cones* and *spheres*).

Geometrical Problems

The following problems illustrate the method in which various geometrical figures are constructed, and they should be solved by the

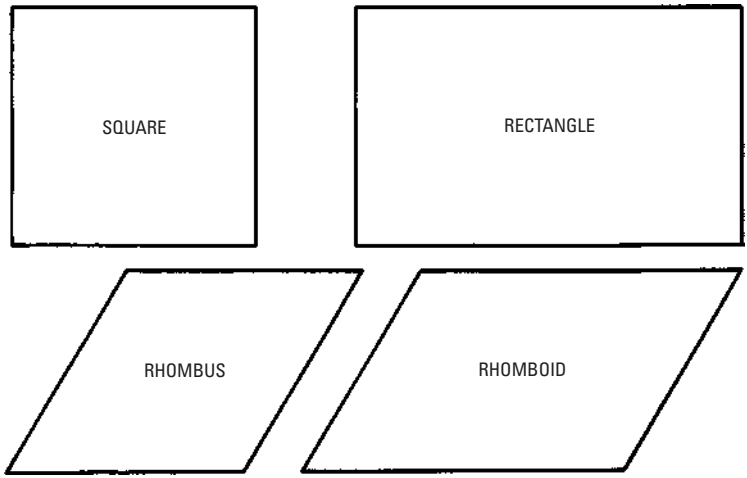

Figure 5-13 Various quadrilaterals. All opposite sides of a quadrilateral are equal.

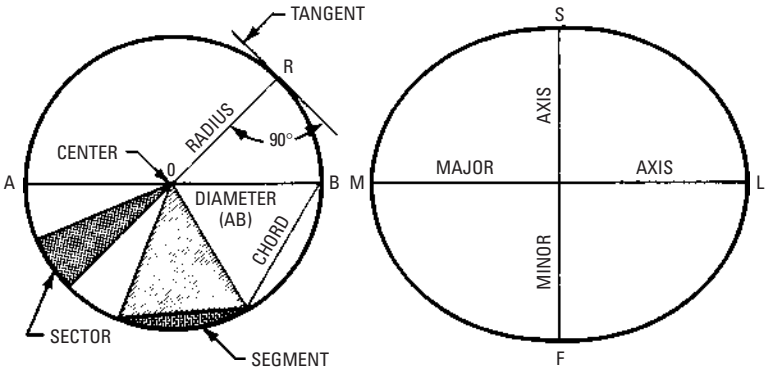

Figure 5-14 Curved figures. A circle is a plane figure bounded by a uniformly curved line, every point of which is equidistant from the center point O. OR is a radius, and AB is a diameter. An ellipse is a curved figure enclosed by a curved line that is such that the sum of the distances between any point on the circumference and the two foci is invariable. ML is the major axis, and SF is the minor axis.

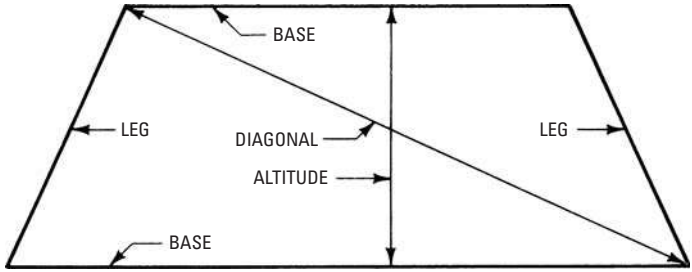

Figure 5-15 The parallel sides of a quadrilateral (four-sided polygon) are the bases. The distance between the bases is the altitude, and a line joining two opposite vertices is a diagonal.

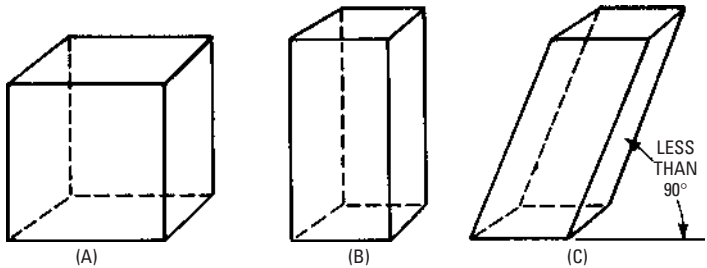

Figure 5-16 Various prisms: (A) cube, or equilateral parallelopipedon; (B) parallelopipedon; and (C) oblique parallelopipedon.

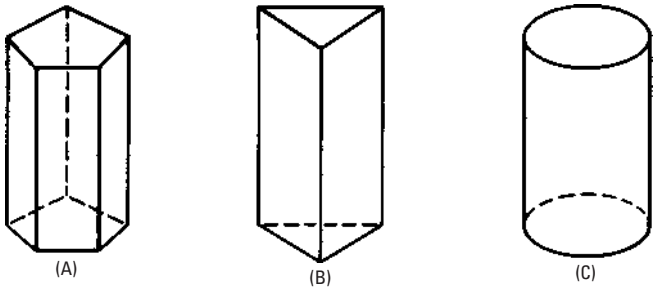

Figure 5-17 Various solids: (A) pentagonal prism, (B) triangular prism, and (C) cylinder.

use of pencil, dividers, compass, and scale. Many of these problems are commonly encountered in carpentry with layout work. Therefore, experience in working them out will be of value to carpenters and woodworkers.

Problem 1

To bisect (or divide into two equal parts) a straight line or arc of a circle.

In Figure 5-18, from the ends *A* and *B*, as centers, describe arcs cutting each other at *C* and *D*, and draw line *CD*, which cuts the line at *E*, or the arc at *F*.

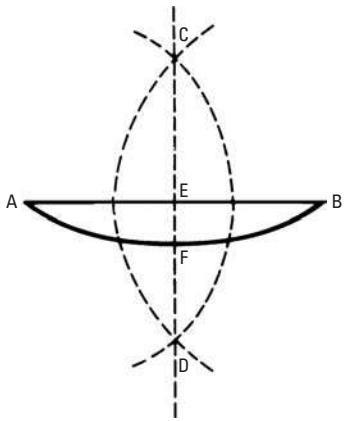

Figure 5-18 To bisect a straight line or arc of a circle.

Problem 2

To draw a perpendicular to a straight line, or a radial line to an arc.

The line *CD* is perpendicular to *AB*; also, the line *CD* is radial to the arc *AB* (see Figure 5-18).

Problem 3

To erect a perpendicular to a straight line from a given point in that line.

In Figure 5-19, with any radius from any given point *A*, in the line *BC* describe arcs cutting the line at *B* and *C*. Next, with a longer radius describe arcs with *B* and *C* as centers, intersecting at *D*, and draw the perpendicular *DA*.

Second Method

In Figure 5-20, from any point *F* above *BC*, describe a circle passing through the given point *A* and cutting the given line at *D*. Draw *DF*, and extend it to cut the circle at *E*. Draw the perpendicular *AE*.

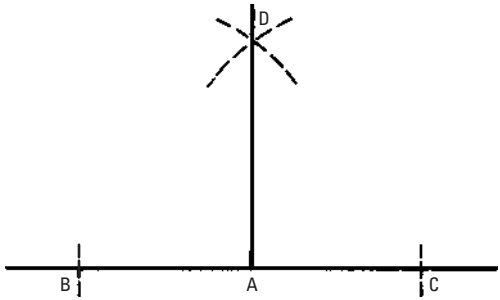

Figure 5-19 To erect a perpendicular to a straight line from a given point on that line.

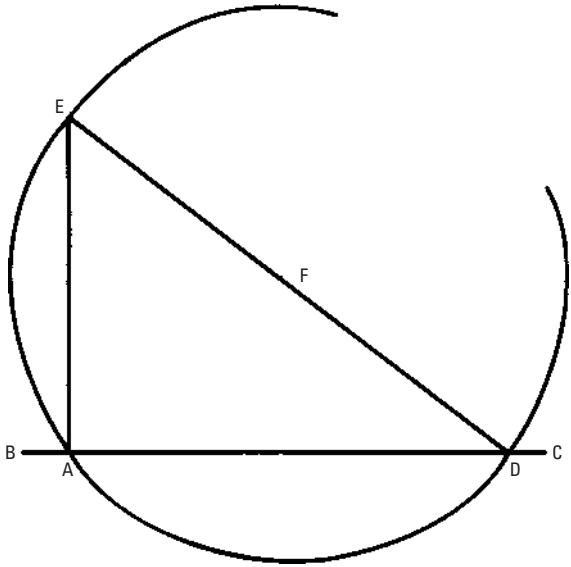

Figure 5-20 To erect a perpendicular to a straight line from a given point on that line, second method.

Third Method (Boat Builders' Layout Method)

In Figure 5-21, let *MS* be the given line and *A* be the given point. From *A*, measure off a distance *AB* (4 feet). With centers *A* and *B* and radii of 3 and 5 feet, respectively, describe arcs *L* and *F* intersecting at *C*. Draw a line through *A* and *C*, which will be the

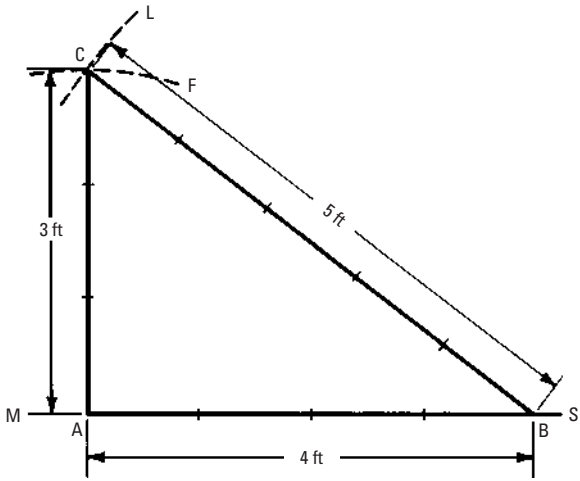

Figure 5-21 To erect a perpendicular to a straight line from a given point on that line, third method.

perpendicular required. This method is used extensively by carpenters when squaring the corners of buildings, but they ordinarily use multiples of 3, 4, and 5 (such as 6, 8, and 10, or 12, 16, and 20).

Fourth Method

In Figure 5-22, from *A*, describe an arc *EC*, and from *E* with the same radius describe the arc *AC*, cutting the other at *C*. Through *C*,

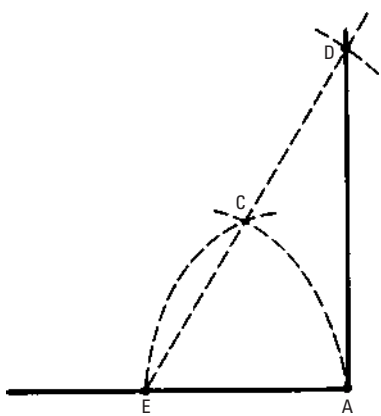

Figure 5-22 To erect a perpendicular to a straight line from a given point on that line, fourth method.

draw a line *ECD*. Lay off *CD* equal to *CE*, and through *D*, draw the perpendicular *AD*. The triangle produced is exactly 60 degrees at *E*, 30 degrees at *D*, and 90 degrees at *A*. The hypotenuse *ED* is exactly twice the length of the base *EA*.

Problem 4

To erect a perpendicular to a straight line from any point outside the line.

In Figure 5-23, from the point *A*, with a sufficient radius cut the given line at *F* and *G*, and from these points describe arcs cutting at *E*. Draw the perpendicular *AE*.

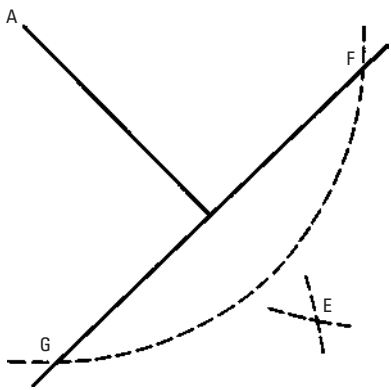

Figure 5-23 To erect a perpendicular to a straight line from any point outside the line.

Second Method

In Figure 5-24, from any two points *B* and *C* at some distance apart in the given line and with the radii *BA* and *CA*, respectively, describe arcs cutting at *A* and *D*. Draw the perpendicular *AD*.

Problem 5

To draw a line parallel to a given line through a given point.

In Figure 5-25, with *C* as the center, describe an arc tangent to the given line *AB*. The radius will then equal the distance from the given point to the given line. Take a point *B* on the given line remote from *C*, and describe an arc. Draw a line through *C*, tangent to this arc at *D*, and it will be parallel to the given line *AB*.

Second Method

In Figure 5-26, from *A*, the given point, describe the arc *FD*, cutting the given line at *F*; from *F*, with the same radius, describe the arc

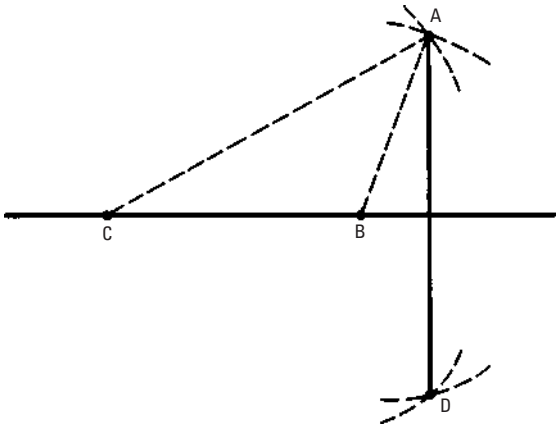

Figure 5-24 To erect a perpendicular to a straight line from any point outside the line, second method.

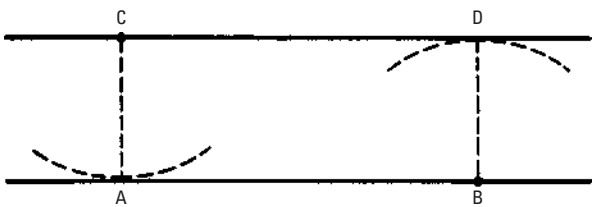

Figure 5-25 To draw a line parallel to a given line through a given point.

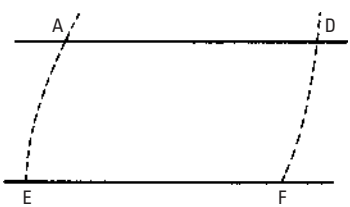

Figure 5-26 To draw a line parallel to a given line through a given point, second method.

EA, and lay off *FD* equal to *EA*. Draw the parallel line through the points *AD*.

Problem 6

To divide a line into a number of equal parts.

In Figure 5-27, assuming line *AB* is to be divided into five equal parts, draw a diagonal line *AC* of five units in length. Join *BC* at 5 and through the points 1, 2, 3, and 4. Draw lines 1*L*, 2*a*, and so forth, parallel to *BC*. *AC* will then be divided into five equal parts, *AL, La, ar, rf*, and *fB*.

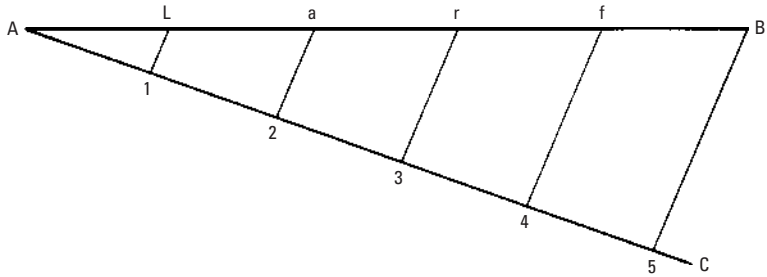

Figure 5-27 To divide a line into a number of equal parts.

Problem 7

To draw an angle equal to a given angle on a straight line.

In Figure 5-28, let *A* be the given angle, and *FG* the line. With any radius from the points *A* and *F*, describe arcs *DE* and *IH* cutting the sides of angle *A* and line *FG*. Lay off arc *IH* equal to arc *DE*, and draw line *FH*. Angle *F* is then equal to *A*, as required

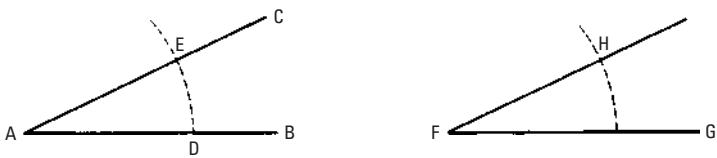

Figure 5-28 To draw an angle equal to a given angle on a straight line.

Problem 8

To bisect an angle

In Figure 5-29, let *ACB* be the angle. With the center of the angle at *C*, describe an arc cutting the sides at *A* and *B*. Using *A* and *B* as centers, describe arcs that intersect at *D*. A line through *C* and *D* will divide the angle into two equal parts.

Problem 9

To find the center of a circle.

In Figure 5-30, draw any chord *MS*. With *M* and *S* as centers, and with any radius, describe arcs *LF* and *L′F′*, and draw a line through their intersection, giving a diameter *AB*. Applying the same construction with centers *A* and *B*, describe arcs *ef* and *e′f′*. A line drawn through the intersections of these arcs will cut line *AB* at *O*, the center of the circle.

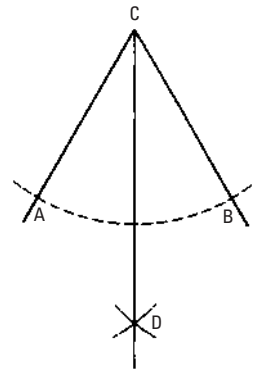

Figure 5-29 To bisect an angle.

Problem 10

To describe an arc of a circle with a given radius through two given points.

In Figure 5-31, take the given points *A* and *B* as centers, and, with the given radius, describe arcs that intersect at *C*. From *C*, with the same radius, describe an arc *AB*, as required.

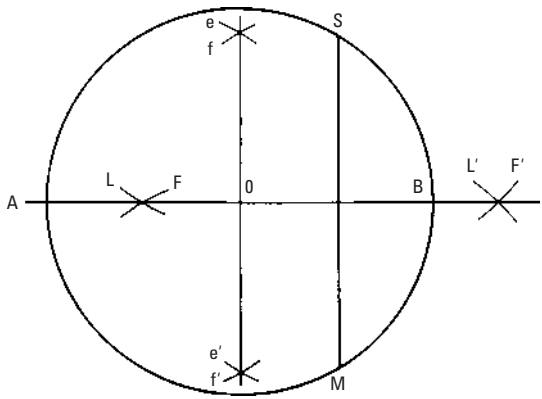

Figure 5-30 Find the center of a circle.

Second Method

In Figure 5-32, for a circle or an arc, select three points *ABC* in the circumference that are well apart. With the same given radius, describe arcs from these three points that intersect each other, and draw two lines, *DE* and *FG*, through their intersections. The point where these lines intersect is the center of the circle or arc.

Problem 11

To describe a circle passing through three given points.

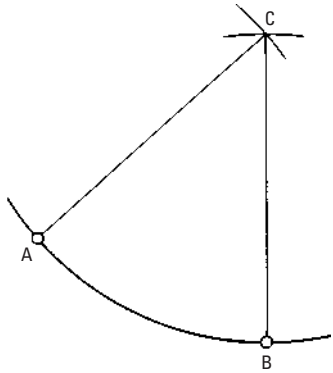

Figure 5-31 To describe an arc of a circle with a given radius through two given points.

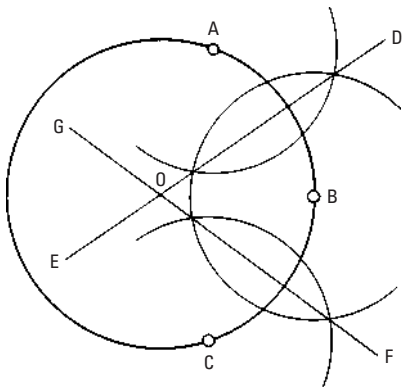

Figure 5-32 To describe an arc of a circle with a given radius through two given points, second method.

In Figure 5-32, let *A*, *B*, and *C* be the given points, and proceed as in Problem 10 to find the center *O* from which the circle may be described. This problem is useful in such work as laying out an object of large diameter (such as an arch) when the span and rise are given.

Problem 12

To draw a tangent to a circle from a given point in the circumference.

In Figure 5-33, from *A*, lay off equal segments *AB* and *AD*. Join line *BD*, and draw line *AE* parallel to *BD* for the tangent.

Problem 13

To draw tangents to a circle from points outside the circle.

In Figure 5-34, from *A*, and with the radius *AC*, describe an arc *BCD*. From *C*, with a radius equal to the diameter of the circle,

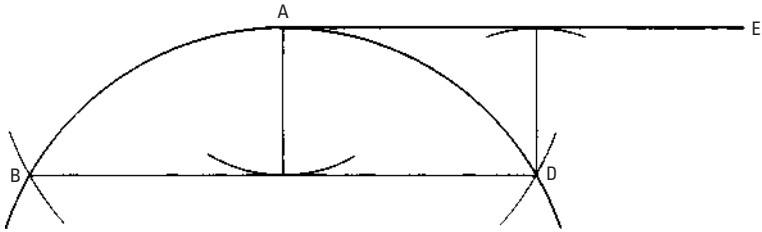

Figure 5-33 To draw a tangent to a circle from a given point in the circumference.

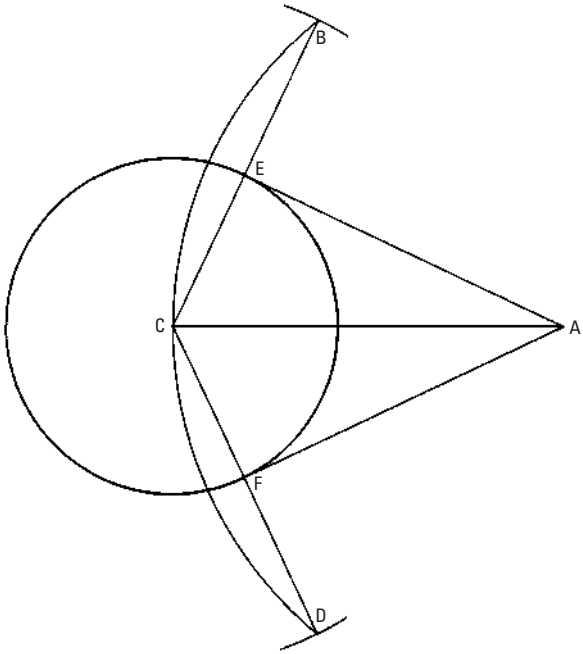

Figure 5-34 To draw tangents to a circle from points outside the circle.

intersect the arc at *BD*. Join *BC* and *CD*, which intersect the circle at *E* and *F*, and draw the tangents *AE* and *AF*.

Problem 14

To describe a series of circles tangent to two inclined lines and tangent to each other.

In Figure 5-35, bisect the inclination of the given lines *AB* and *CD* by the line *NO*. From a point *P* in this line, draw the perpendicular *PB* to the line *AB*, and on *P*, describe the circle *BD*, touching the lines and the centerline at *E*. From *E*, draw *EF* perpendicular to the center line intersecting *AB* at *F*, and from *F*, describe an arc *EG* intersecting *AB* at *G*. Draw *GH* parallel to *BP*, thus producing *H*, the center of the next circle, to be described with the radius *HE*, and so on for the next circle *IN*.

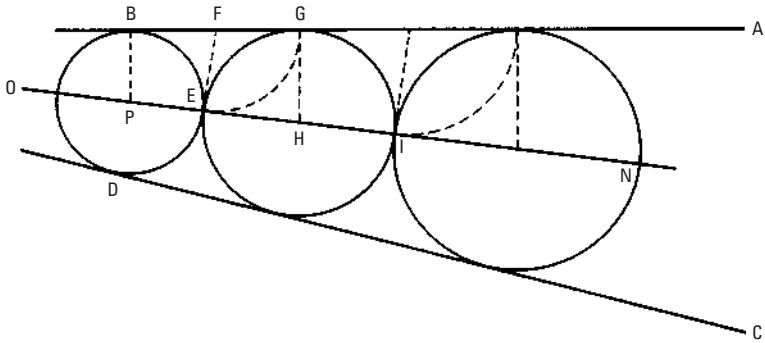

Figure 5-35 To describe a series of circles tangent to two inclined lines and tangent to each other.

Problem 15

To construct an equilateral triangle on a given base.

In Figure 5-36, with *A* and *B* as centers and a radius equal to *AB*, describe arcs *l* and *f*. At their intersection *C*, draw lines *CA* and *CB*, which are the sides of the required triangle. If the sides are to be unequal, the process is the same, taking as the radii the lengths of the two sides to be drawn.

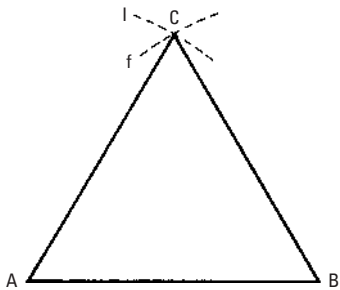

Figure 5-36 To construct an equilateral triangle on a given base.

Problem 16

To construct a square on a given base.

In Figure 5-37, with end points *A* and *B* of the base as centers and a radius equal to *AB*, describe arcs that intersect at *C*; on *C*, describe arcs that intersect the others at *D* and *E*, and on *D* and *E*, intersect these arcs *F* and *G*. Draw *AE* and *BG*, and join the intersections *HI* to form the square *AHIB*.

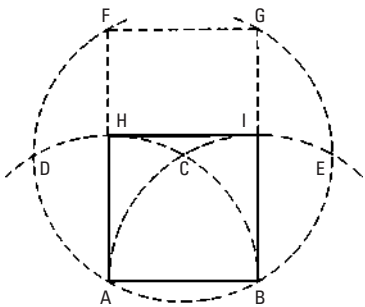

Figure 5-37 To construct a square or a rectangle on a given base.

Problem 17

To construct a rectangle on a given base.

In Figure 5-37, let *AB* be the given base. Erect a perpendicular at *A* and at *B* that is equal to the altitude of the rectangle, and join their ends *F* and *G* by line *FG*. *AFGB* is the rectangle required.

Problem 18

To construct a parallelogram given the sides and an angle.

In Figure 5-38, draw side *DE* equal to the given length *A*, and lay off the other side *DF*, equal to the other length *B*, thus forming the given angle *C*. From *E*, with *DF* as the radius, describe an arc, and from *F*, with the radius *DE*, intersect the arc at *G*. Draw *FG* and *EG*. The remaining sides may also be drawn as parallels to *DE* and *DF*.

Problem 19

To draw a circle around a triangle.

In Figure 5-39, bisect two sides *AB* and *AC* of the triangle at *E* and *F*, and from these points draw perpendiculars intersecting at *K*. From *K*, with radius *KA* or *KC*, describe the circle *ABC*.

Problem 20

To circumscribe and inscribe a circle about a square.

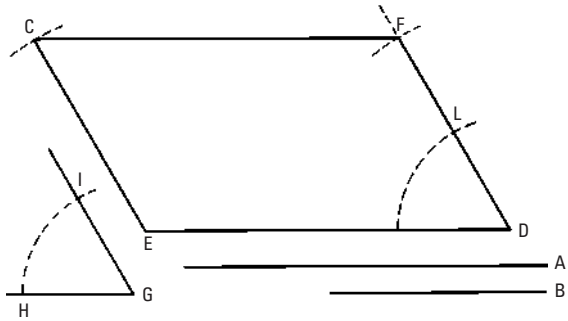

Figure 5-38 To construct a parallelogram given the sides and an angle.

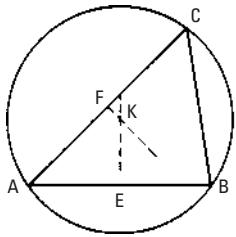

Figure 5-39 To draw a circle around a triangle.

In Figure 5-40, draw the diagonals *AB* and *CD* intersecting at *E*. With a radius *EA*, circumscribe the circle. To inscribe a circle, draw a perpendicular from the center (as just found) to one side of the square, as line *OM*. With radius *OM*, inscribe the circle.

Problem 21

To circumscribe a square around a circle.

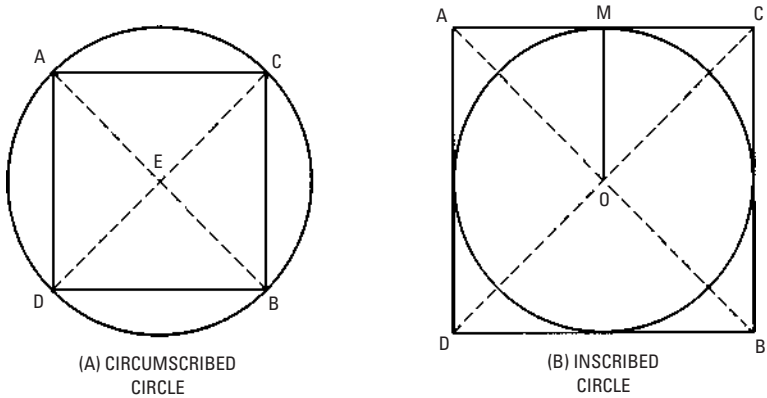

Figure 5-40 (A) To circumscribe a circle around a square. (B) To inscribe a circle inside a square.

In Figure 5-41, draw diameters *MS* and *LF* at right angles to each other. At points *M*, *L*, *S*, and *F*, where these diameters intersect the circle, draw tangents (that is, lines perpendicular to the diameters), obtaining the sides of the circumscribed square *ABCD*.

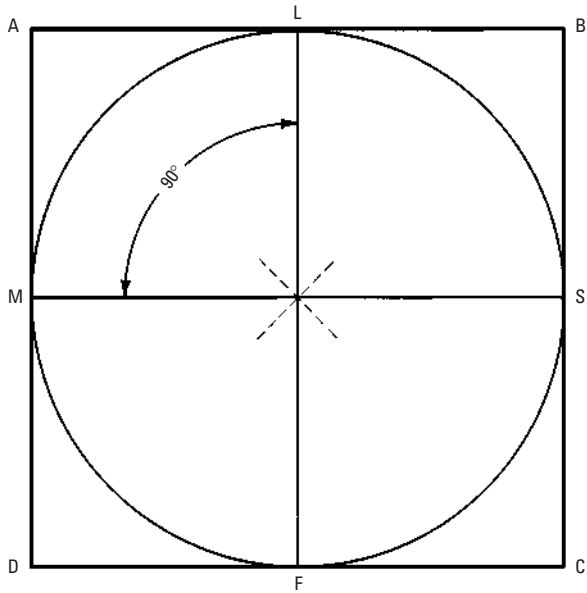

Figure 5-41 To circumscribe a square about a circle.

Problem 22

To inscribe a circle in a triangle.

In Figure 5-42, bisect two angles *A* and *C* of the triangle with lines that intersect at *D*. From *D*, draw a perpendicular *DE* to any side. With *DE* as the radius, describe a circle.

Problem 23

To inscribe a pentagon in a circle.

In Figure 5-43, draw two diameters *AC* and *BD* at right angles intersecting at *O*. Bisect *AO* at *E*, and from *E*, with radius *EB*, *AC* at *F*; from *B*, with radius *BF*. Intersect the circumference at *G* and *H*, and with the same radius, step round the circle to *I* and *K*; join the points thus found to form the pentagon *BGIKH*.

Problem 24

To inscribe a five-pointed star in a circle.

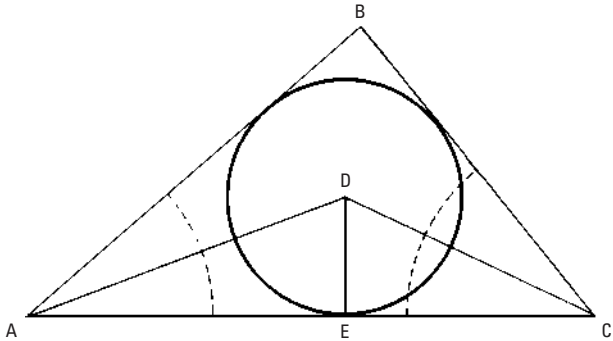

Figure 5-42 To inscribe a circle in a triangle.

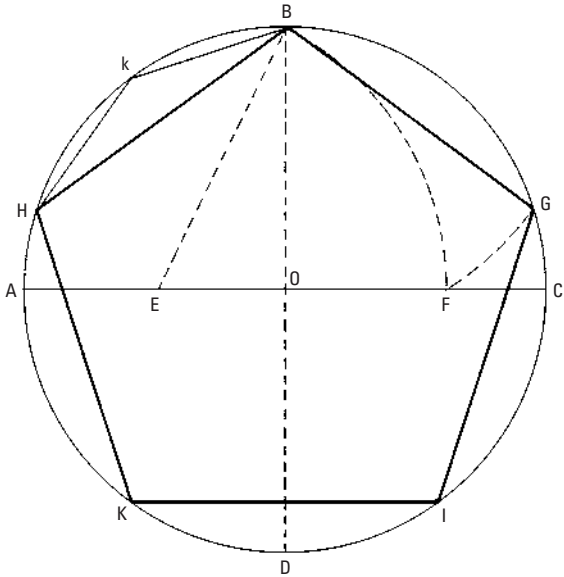

Figure 5-43 To inscribe a pentagon in a circle.

In Figure 5-44, proceed as explained for the inscribed pentagon in Problem 23. Then, connect point *B* with points *K* and *I*, point *H* with points *G* and *I*, and so forth. The star is mathematically correct.

Problem 25

To construct a hexagon from a given straight line

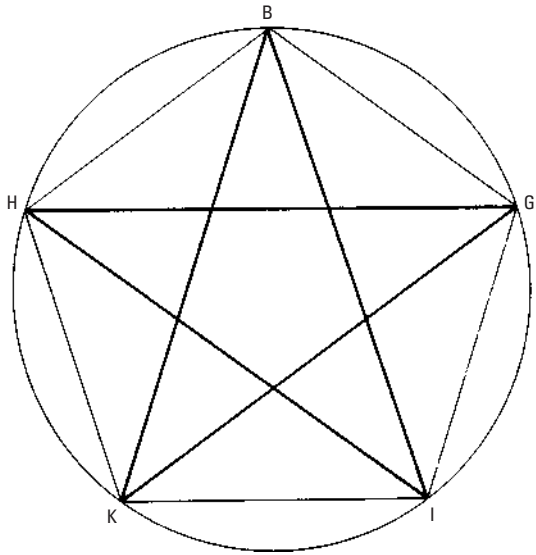

Figure 5-44 To inscribe a five-pointed star in a circle.

In Figure 5-45, from *A* and *B*, the ends of the given line, describe arcs intersecting at *g*. From *g*, with the radius *gA*, describe a circle. With the same radius, lay off arcs *AG*, *GF*, *BD*, and *DE*. Join the points thus found to form the hexagon.

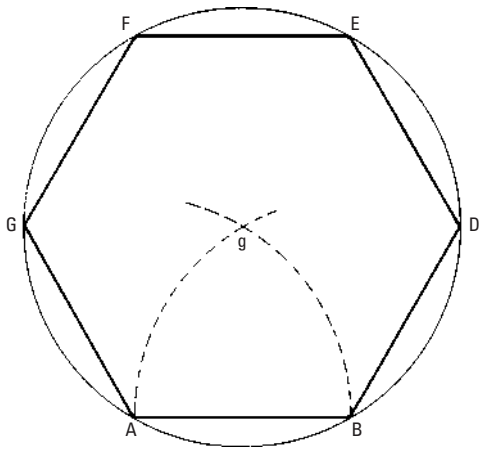

Figure 5-45 To construct a hexagon from a given straight line.

Problem 26

To inscribe a hexagon in a circle.

In Figure 5-46, draw a diameter *ACB*. From *A* and *B*, as centers with the radius of the circle *AC*, intersect the circumference at *D*, *E*, *F*, and *G*, and draw lines *AD*, *DE*, and so forth, to form the hexagon. The points *D*, *E*, and so forth, may also be found by stepping off the radius (with the dividers) six times around the circle.

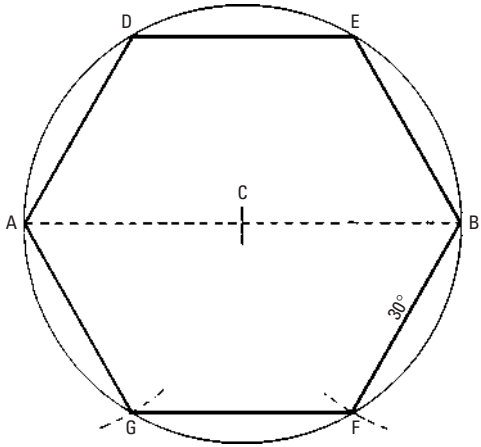

Figure 5-46 To inscribe a hexagon in a circle.

Problem 27

To describe an octagon on a given straight line.

In Figure 5-47, extend the given line *AB* both ways. Now, draw perpendiculars *AE* and *BF*. Bisect the external angles *A* and *B* by using lines *AH* and *BC*. These are made equal to line *AB*. Draw *CD* and *HG* parallel to *AE* and equal to line *AB*. Draw *CD* and *HG* parallel to *AE* and equal to line *AB*. With *G* and *D* as centers, and with the radius equal to *AB*, intersect the perpendiculars at *E* and *F*, and draw line *EF* to complete the hexagon.

Problem 28

To inscribe an octagon in a square.

In Figure 5-48, draw the diagonals of the square intersecting at *e*. From the corners *A*, *B*, *C*, and *D*, with *Ae* as the radius, describe arcs intersecting the sides of the square at *g*, *h*, and so forth, and join the points found to complete the octagon.

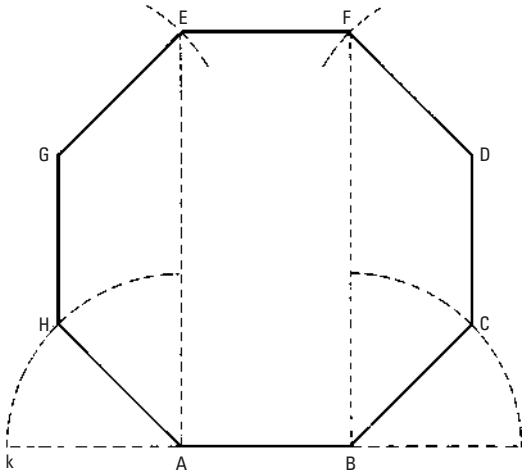

Figure 5-47 To describe an octagon on a given straight line.

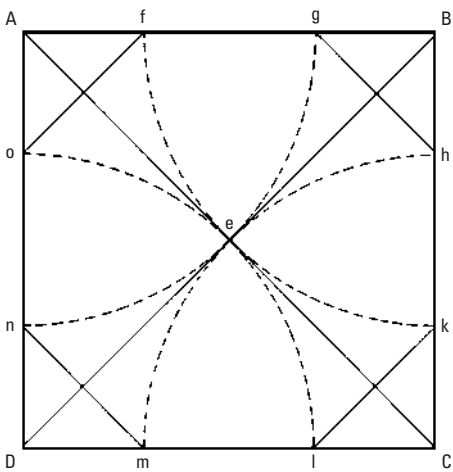

Figure 5-48 To inscribe an octagon in a square.

Problem 29

To inscribe an octagon in a circle.

In Figure 5-49, draw two diameters *AC* and *BD* at right angles. Bisect the arcs *AB*, *BC*, and so forth, at *e*, *f*, and so forth, to form the octagon.

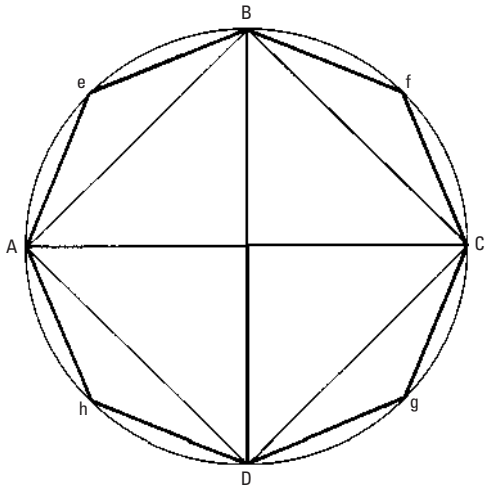

Figure 5-49 To inscribe an octagon in a circle.

Problem 30

To circumscribe an octagon about a circle.

In Figure 5-50, describe a square about the given circle *AB*. Draw perpendiculars *h*, *k*, and so forth, to the diagonals, touching the circle, to form the octagon. The points *h*, *k*, and so forth, may be found by cutting the sides from the corners.

Problem 31

To describe an ellipse when the two axes are given.

In Figure 5-51, draw the major and minor axes *AB* and *CD*, respectively, at right angles intersecting at *E*. On *C*, with *AE* as the radius, intersect the axis *AB* at *F* and *G*, the *foci*. Insert pins through the axis at *F* and *G*, and loop a thread or cord on them equal in length to the axis *AB*, so that when stretched, it reaches extremity *C* of the *conjugate axis*, as shown in dotted lines. Place a pencil inside the cord, as at *H*, and, by guiding the pencil in this manner, describe the ellipse.

Second Method

Along the edge of a piece of paper, mark off a distance *ac* equal to *AC*, one-half the major axis, and from the same point a distance *ab* equal to *CD*, one-half the minor axis, as shown in Figure 5-52. Place the paper to bring point *b* on the line *AB*, or major axis, and point *c* on the line *DE*, or minor axis. Lay off the position of point *a*.

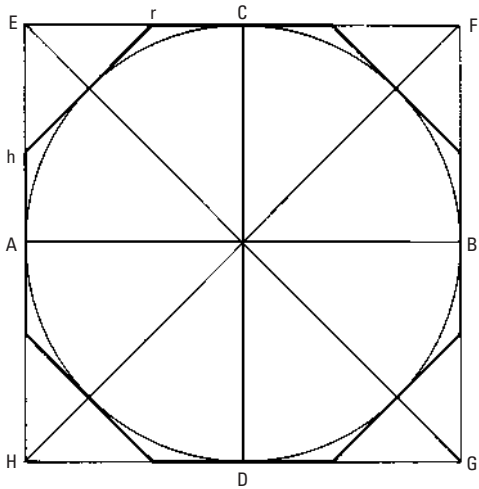

Figure 5-50 To circumscribe an octagon about a circle.

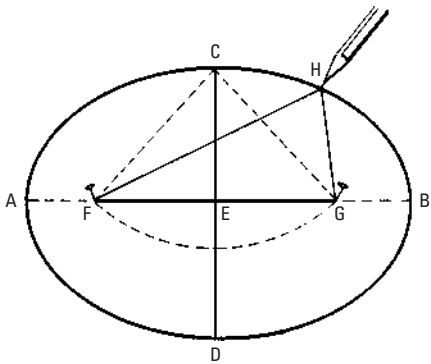

Figure 5-51 To describe an ellipse when the two axes are given.

By shifting the paper so that point *b* travels on the major axis and point *c* travels on the minor axis, any number of points in the curve may be found through which the curve may be traced.

Mensuration

As mentioned earlier, mensuration is the act, art, or process of measuring. It is the branch of mathematics that deals with finding the length of lines, the area of surfaces, and the volume of solids. Therefore, the problems that follow will be divided into three groups as follows:

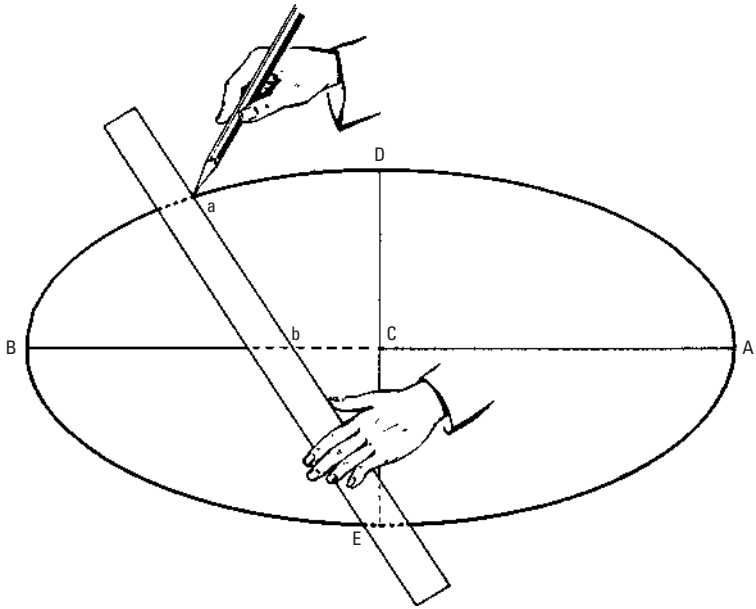

Figure 5-52 To describe an ellipse given the two axes, second method.

- Measurement of *lines*, one dimension (length)
- Measurement of *surfaces* (areas), two dimensions (length and width)
- Measurement of *solids* (volumes), three dimensions (length, width, and thickness)

Measurement of Lines—Length

Problem 1

To find the length of any side of a right triangle given the other two sides.

Rule: The length of the hypotenuse equals the square root of the sum of the squares of the two legs. The length of either leg equals the square root of the difference of the square of the hypotenuse and the square of the other leg.

Example The two legs of a right triangle measure 3 feet and 4 feet. Find the length of the hypotenuse. If the lengths of the hypotenuse and one leg are 5 feet and 4 feet, respectively, what is the length of the other leg?

In Figure 5-53A,

$$AB = \sqrt{3^2 + 4^2} = \sqrt{25} = 5 \text{ feet}$$

In Figure 5-53B,

$$BC = \sqrt{5^2 - 3^2} = \sqrt{25 - 9} = \sqrt{16} = 4 \text{ feet}$$

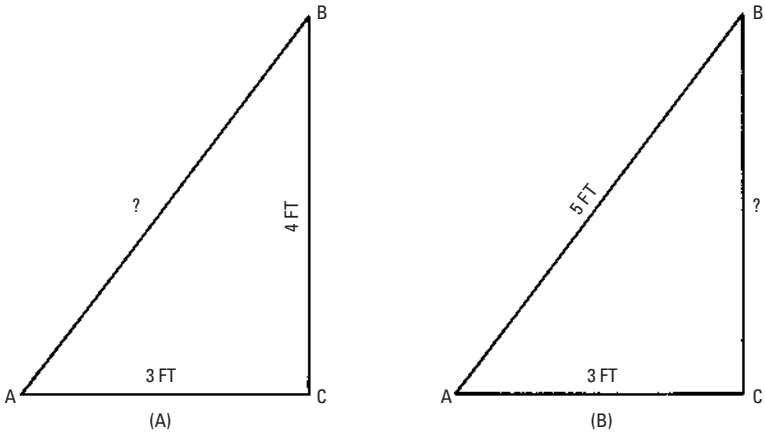

Figure 5-53 To find the length of any side of a right triangle given the other two sides.

Problem 2

To find the length of the circumference of a circle.

Rule: Multiply the diameter by 3.1416.

Example What length of molding strip is required for a circular window that is 5 feet in diameter?

$$5 \times 3.1416 = 15.7 \text{ feet}$$

Since the carpenter does not ordinarily measure feet in tenths, .7 should be reduced to inches. It corresponds to $8^1/_2$ inches from Table 5-23. That is, the length of molding required is 15 feet $8^1/_2$ inches.

Problem 3

To find the length of the arc of a circle.

Rule: Arc = .017453 × radius × central angle.

Example If the radius of a circle is 2 feet, what is the length of a 60° arc?

Table 5-23 Decimals of a Foot and Inches

Inch	*0 inch*	*1 inch*	*2 inches*	*3 inches*	*4 inches*	*5 inches*	*6 inches*	*7 inches*	*8 inches*	*9 inches*	*10 inches*	*11 inches*
0	0.0000	0.0833	0.1677	0.2500	0.3333	0.4167	0.5000	0.5833	0.6667	0.7500	0.8333	0.9167
1–16	0.0052	0.0885	0.1719	0.2552	0.3385	0.4219	0.5052	0.5885	0.6719	0.7552	0.8385	0.9219
1–8	0.0104	0.0937	0.1771	0.2604	0.3437	0.4271	0.5104	0.5937	0.6771	0.7604	0.8437	0.9271
3–16	0.0156	0.0990	0.1823	0.2656	0.3490	0.4323	0.5156	0.5990	0.6823	0.7656	0.8490	0.9323
1–4	0.0208	0.1042	0.1875	0.2708	0.3542	0.4375	0.5208	0.6042	0.6875	0.7708	0.8542	0.9375
5–16	0.0260	0.1094	0.1927	0.2760	0.3594	0.4427	0.5260	0.6094	0.6927	0.7760	0.8594	0.9427
3–8	0.0312	0.1146	0.1979	0.2812	0.3646	0.4479	0.5312	0.6146	0.6979	0.7812	0.8646	0.9479
7–16	0.0365	0.1198	0.2031	0.2865	0.3698	0.4531	0.5365	0.6198	0.7031	0.7865	0.8698	0.9531
1–2	0.0417	0.1250	0.2083	0.2917	0.3750	0.4583	0.5417	0.6250	0.7083	0.7917	0.8750	0.9583
9–16	0.0469	0.1302	0.2135	0.2969	0.3802	0.4635	0.5469	0.6302	0.7135	0.7969	0.8802	0.9635
5–8	0.0521	0.1354	0.2188	0.3021	0.3854	0.4688	0.5521	0.6354	0.7188	0.8021	0.8854	0.9688
11–16	0.0573	0.1406	0.2240	0.3073	0.3906	0.4740	0.5573	0.6406	0.7240	0.8073	0.8906	0.9740
3–4	0.0625	0.1458	0.2292	0.3125	0.3958	0.4792	0.5625	0.6458	0.7292	0.8125	0.8958	0.9792
13–16	0.0677	0.1510	0.2344	0.3177	0.4010	0.4844	0.5677	0.6510	0.7344	0.8177	0.9010	0.9844
7–8	0.0729	0.1562	0.2396	0.3229	0.4062	0.4896	0.5729	0.6562	0.7396	0.8229	0.9062	0.9896
15–16	0.0781	0.1615	0.2448	0.3281	0.4115	0.4948	0.5781	0.6615	0.7448	0.8281	0.9115	0.9948

Solution:

$$2 \times .017453 \times 60 = 2.094,$$

or approximately 2 feet $1^1/_8$ inches

Problem 4

To find the rise of an arc.

Rule: Rise of an arc =

$$\sqrt{(4 \times \text{radius}^2) - \text{length}}$$

Example If the radius of a circle is 2 feet, what is the rise at the center of a 2-foot chord?

Solution:

$$^1/_2\sqrt{(4 \times 2^2) - 2} = {}^1/_2\sqrt{14} = 1.87 \text{ feet}$$
$$= 1 \text{ feet } 10^1/_2 \text{ inches}$$

Measurement of Surfaces—Area

Problem 5

To find the area of a square.

Rule: Multiply the base by the height.

Example What is the area of a square whose side is 5 feet (see Figure 5-54)?

$$5 \times 5 = 25 \text{ square feet}$$

Problem 6

To find the area of a rectangle.

Rule: Multiply the base by the height (that is, width by length).

Example What is the floor area of a porch 5 feet wide and 12 feet long (see Figure 5-55)?

$$5 \times 12 = 60 \text{ square feet}$$

Problem 7

To find the area of a parallelogram.

Rule: Multiply the base by the perpendicular height.

Example What is the area of a 5-foot × 12-foot parallelogram (see Figure 5-56)?

$$5 \times 12 = 60 \text{ square feet}$$

Problem 8

To find the area of a triangle (see Figure 5-57)

Rule: Multiply the base by one-half the altitude.

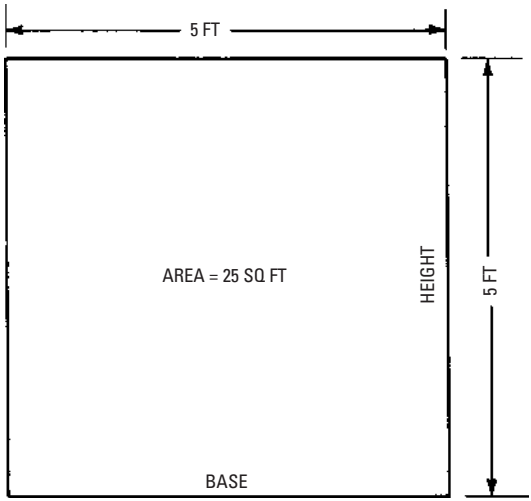

Figure 5-54 To find the area of a square.

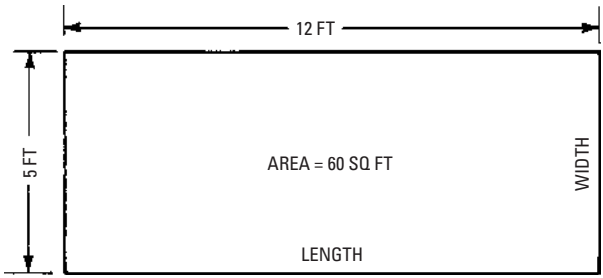

Figure 5-55 To find the area of a rectangle.

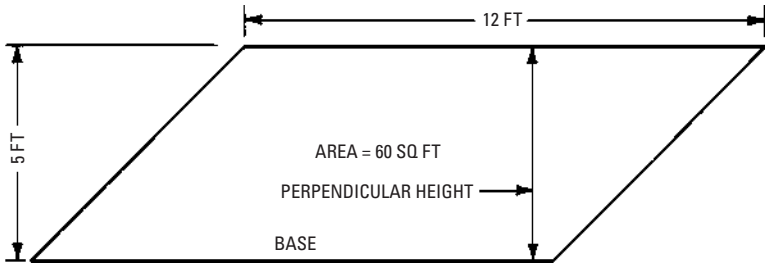

Figure 5-56 To find the area of a parallelogram.

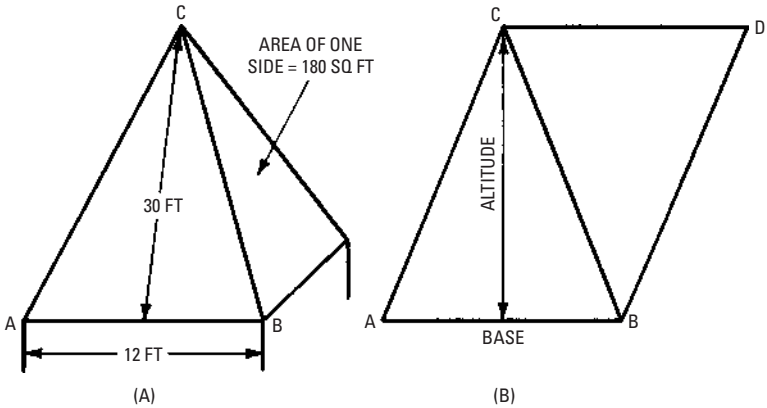

Figure 5-57 To find the area of a triangle (equal to ½ area of parallelogram ABDC).

Example How many square feet of sheathing are required to cover a church steeple having four triangular sides?

Problem 9

To find the area of a trapezoid.

Rule: Multiply one-half the sum of the two parallel sides by the perpendicular distance between them.

Example What is the area of the trapezoid shown in Figure 5-58?

LA and *FR* are the parallel sides, and *MS* is the perpendicular distance between them. Therefore,

$$\text{area} = {}^1/_2\,(LA + FR) \times MS$$

$$\text{area} = {}^1/_2\,(8 + 12) \times 6 = 60 \text{ square feet}$$

Problem 10

To find the area of a trapezium.

Rule: Draw a diagonal, dividing the figure into triangles. Measure the diagonal and the altitudes, and find the area of the triangles. The sum of these areas is then the area of the trapezium.

Example What is the area of the trapezium shown in Figure 5-59? (Draw diagonal *LR* and altitudes *AM* and *FS*.)

$$\text{area of triangle } ALR = {}^1/_2\,(12 \times 9) = 54 \text{ square feet}$$

$$\text{area of triangle } LRF = {}^1/_2\,(12 \times 6) = 36 \text{ square feet}$$

$$\text{area of trapezium } LARF = ALR + LRF = 36 + 54$$

$$= 90 \text{ square feet}$$

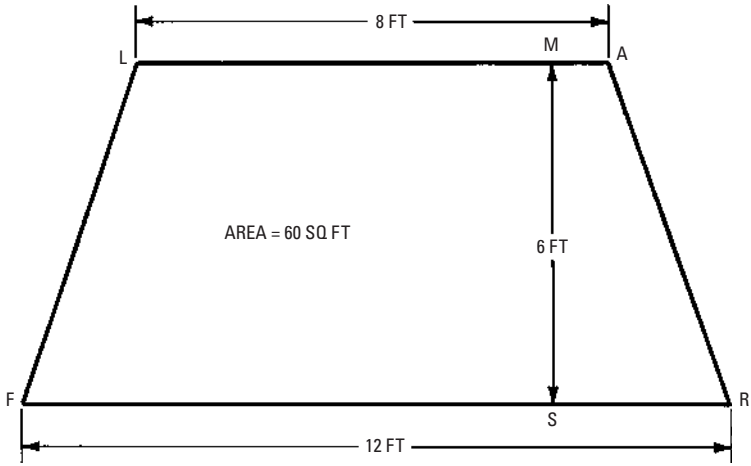

Figure 5-58 To find the area of a trapezoid.

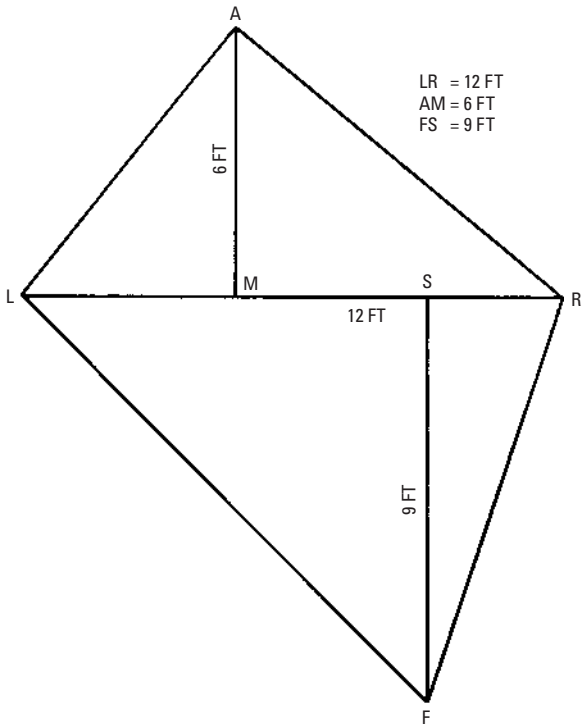

Figure 5-59 To find the area of a trapezium.

Problem 11

To find the area of any irregular polygon.

Rule: Draw diagonals, dividing the figure into triangles, and find the sum of the areas of these triangles.

Problem 12

To find the area of any regular polygon, such as shown in Figure 5-60, when the length of only one side is given

Rule: Multiply the square of the sides by the figure for "area when side = 1" opposite the particular polygon in Table 5-24.

Example What is the area of an octagon (8-sided polygon) whose sides are 4 feet in length?

In Table 5-24 under 8 find 4.828. Multiply this by the square of one side.

$$4.828 \times 4^2 = 77.25 \text{ square feet}$$

Problem 13

To find the area of a circle (see Figure 5-61).

Rule: Multiply the square of the diameter by 0.7854.

Example How many square feet of floor surface are there in a 10-foot circular floor?

$$10^2 \times 0.7854 = 78.54 \text{ square feet}$$

Problem 14

To find the area of a sector of a circle.

Rule: Multiply the arc of the sector by one-half the radius.

Example How much tin is required to cover a 60° section of a 10-foot circular deck?

$$\text{length of } 60° \text{ arc} = {}^{60}/_{30} \text{ of } 3.1416 \times 10 = 5.24 \text{ feet}$$

$$\text{tin required for } 60° \text{ sector} = 5.24 \times {}^{1}/_{2} \times 5$$
$$= 13.1 \text{ square feet}$$

Problem 15

To find the area of a segment of a circle.

Rule: Find the area of the sector that has the same arc, and also find the area of the triangle formed by the radii and chord. Take the sum of these areas if the segment is greater than 180°. Take the difference if the segment is less than 180°.

Problem 16

To find the area of a ring.

Rule: Take the difference between the areas of the two circles.

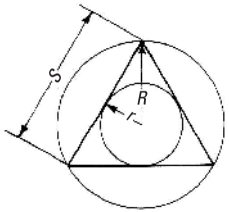

(A)

Steel square cut to miter sides at 30°

$R = 2r$, or $.577\ S$

AREA $= .433\ S^2$
$= 1.299\ R^2$
$= 5.196\ r^2$
$= \frac{(R+r)}{2}\ S$

(A) Equilateral triangle (3 sides).

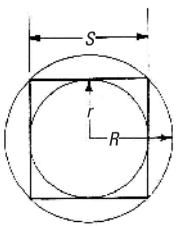

(B)

Steel square cut to miter sides at 45°

$R = .707\ S$, or $1.414\ r$

AREA $= S^2$
$= 2R^2$
$= 4r^2$

(B) Square (4 sides).

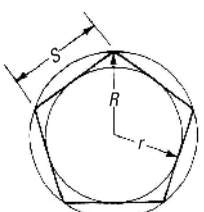

(C)

Steel square cut to miter sides at 36°

$R = .851\ S$, or $1.236\ r$

AREA $= 1.72\ S^2$
$= 2.378\ R^2$
$= 3.633\ r^2$

(C) Pentagon (5 sides).

Steel square cut to miter sides at 30°

$R = S$, or $1.155\ r$

AREA $= 2.598\ S^2$
$= 2.598\ R^2$
$= 3.464\ r^2$

(D) Hexagon (6 sides).

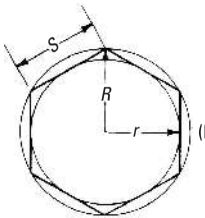

(D)

Steel square cut to miter corners at 25° 43'

$R = 1.152\ S$, or $1.11\ r$

AREA $= 3.634\ S^2$
$= 2.736\ R^2$
$= 3.371\ r^2$

(E) Heptagon (7 sides).

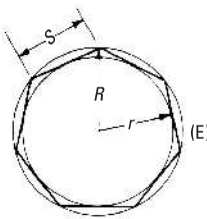

(E)

Steel square cut to miter corners at 22° 30'

$R = 1.307\ S$, or $1.082\ r$

AREA $= 4.828\ S^2$
$= 2.828\ R^2$
$= 3.314\ r^2$

(F) Octagon (8 sides).

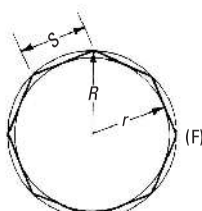

(F)

Figure 5-60 Regular polygons.

Table 5-24 Regular Polygons

Number of sides	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Area when side = 1	0.433	1.0	1.721	2.598	3.634	4.828	6.181	7.694	9.366	11.196

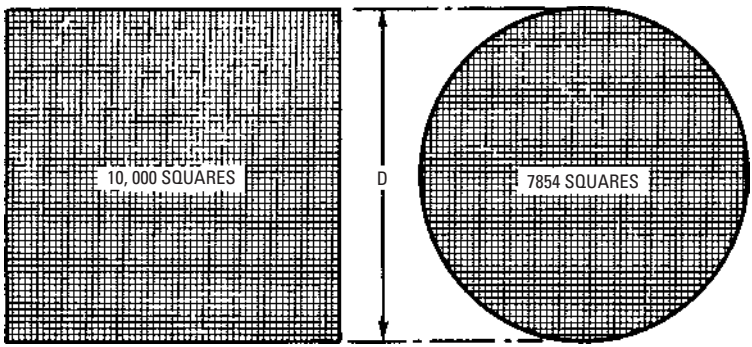

Figure 5-61 The decimal 0.7854 is used to find the area of a circle. If a square is divided into 10,000 equal parts (small squares), then a circle with a diameter D equal to one side of the large square will contain 7854 small squares. Therefore, if the area of the large square is 1 square inch, then the area of the circle will be $^{7854}/_{10,000}$, or 0.7854 square inch.

Problem 17

To find the area of an ellipse.

Rule: Multiply the product of the two diameters by 0.7854.

Example What is the area of an ellipse whose two diameters are 10 inches and 6 inches?

$$10 \times 6 \times 0.7854 = 47.12 \text{ square inches}$$

Problem 18

To find the circular area of a cylinder.

Rule: Multiple 3.1416 by the diameter and by the height.

Example How many square feet of lumber are required for the sides of a cylindrical tank (see Figure 5-62) that is 8 feet in diameter and 12 feet high? How many 4-inch × 12-feet pieces will be required?

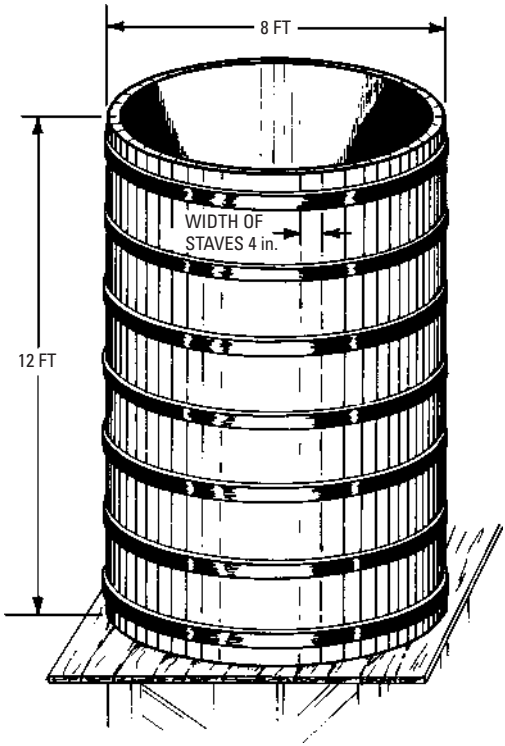

Figure 5-62 To find the area of a cylinder.

$$\text{cylindrical surface} = 3.1416 \times 8 \times 12 = 302 \text{ square feet}$$

$$\text{circumference of tank} = 3.1416 \times 8 = 25.1 \text{ feet}$$

$$\text{number of } 4'' \times 12' \text{ pieces} = \frac{25.1 \times 12}{4} = 25.1 \times 3 = 75.3$$

Problem 19

To find the area of a cone (see Figure 5-63).

Rule: Multiply 3.1416 by the diameter of the base and by one-half the slant height.

Example A conical spire with a base 10 feet in diameter and an altitude of 20 feet is to be covered. Find the area of the surface to be covered.

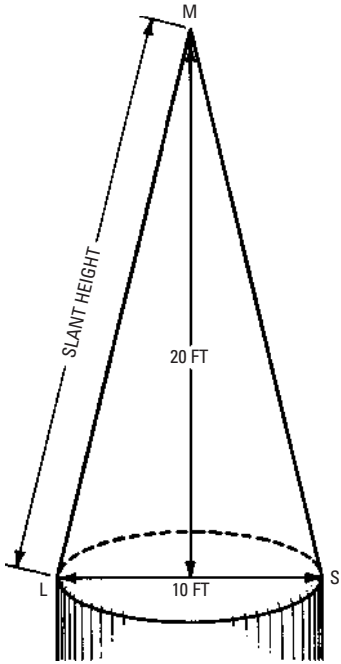

Figure 5-63 To find the surface area of a cone.

$$\text{slant height} = \sqrt{5^2 + 20^2} = \sqrt{425} = 20.62 \text{ feet}$$

$$\text{circumference of base} = 3.1416 \times 10 = 31.416 \text{ feet}$$

$$\text{area of conical surface} = 31.416 \times {}^1/_2 \times 20.62$$
$$= 324 \text{ square feet}$$

Problem 20

To find the area of the frustum of a cone (see Figure 5-64)

Rule: Multiply one-half the slant height by the sum of the circumference.

Example A tank is 12 feet in diameter at the base, 10 feet at the top, and 8 feet high. What is the area of the slant surface?

$$\text{circumference of 10-foot diameter} = 3.1416 \times 10$$
$$= 31.416 \text{ feet}$$

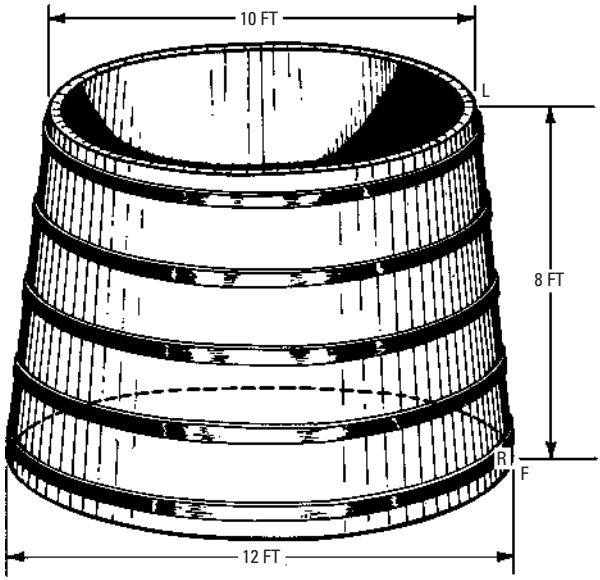

Figure 5-64 To find the area of the frustum of a cone.

$$\text{circumference of 12-foot diameter} = 3.1416 \times 12$$
$$= 37.7 \text{ feet}$$

$$\text{sum of circumferences} = 69.1 \text{ feet}$$

$$\text{slant height} = \sqrt{1^2 + 8^2} = \sqrt{65} = 8.12$$

$$\text{slant surface} = \text{sum of circumferences} \times {}^1/_2 \text{ slant height}$$

$$\text{slant surface} = 69.1 \times {}^1/_2 \times 8.12 = 280 \text{ square feet}$$

Measurement of Solids—Volume

Problem 21

To find the volume of a rectangular solid.

Rule: Multiply the length, width, and thickness together.

Example What is the volume of a 4-inch × 8-inch × 12-foot timber? (Before applying the rule, reduce all dimensions to feet.)

$$4 \text{ inches} = {}^1/_3 \text{ foot}$$

$$8 \text{ inches} = {}^2/_3 \text{ foot}$$

$$\text{volume of timber} = {}^1/_3 \times {}^2/_3 \times 12 = 2.67 \text{ cubic feet}$$

If the timber were a piece of oak weighing 48 pounds per cubic foot, the total weight would be calculated as follows:

$$48 \times 2.67 = 128 \text{ pounds}$$

Problem 22

To find the volume of a rectangular wedge.

Rule: Find the area of one of the triangular ends, and multiply the area by the distance between the ends.

Example An attic has the shape of a rectangular wedge. What volume storage capacity would there be for the proportions shown in Figure 5-65? In the illustration, the boundary of the attic is *LARFMS*.

$$\text{Area of triangular end } MLA = 20 \times {}^{10}/_{2} = 100 \text{ square feet}$$

$$\text{Volume of attic} = 100 \times 40 = 4000 \text{ cubic feet.}$$

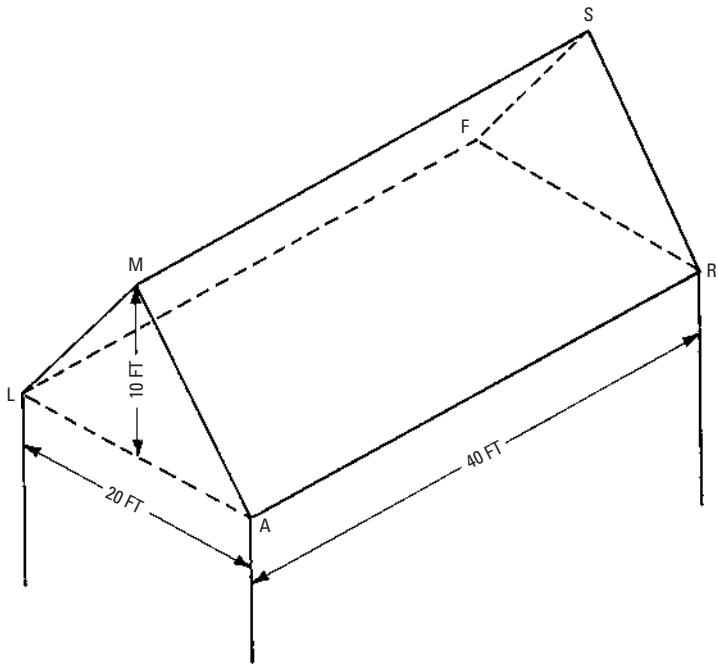

Figure 5-65 To find the volume of a rectangular wedge.

Trigonometry

Trigonometry is that branch of mathematics that deals with the relations that exist between the sides and angles of triangles, and more especially with those methods of calculating the required parts of triangles from given parts. The only branch of trigonometry useful to the carpenter and builder is *plane trigonometry*, where the lines in the triangles are straight and where they all lie in the same plane.

The six elements (or parts) in every triangle are three sides and three angles. The sum of the three angles, no matter what the lengths of the sides, will always be equal to 180 degrees.

When any three of the six parts are given, provided one or more of them are sides, the other three are calculable. The angles are measured in circular measure: in degrees (°), minutes (′), and seconds (″). The term *degree* has no numerical value. In trigonometry it simply means $^1/_{360}$ of a circle, nothing more.

To the student of trigonometry, any two radii that divide a circle into anything more than 0° or less than 360° form an angle.

The first 90° division is called the *first quadrant*. Angles in this quadrant are the *acute angles* (see Figure 5-66A) mentioned earlier in this chapter. Angles from 90° to 180° are in the *second quadrant*. These are the *obtuse angles* (see Figure 5-66B) mentioned. Angles from 180° to 270° lie in the *third quadrant*, and angles from 270° to 360° lie in the *fourth quadrant*. Figure 5-67 shows these four quadrants. Only angles in the first and second quadrants, from 0° to 180°, will be discussed in this section. Note that a straight line may be considered as an angle of 180°. Trigonometry is actually based on geometry, but it makes use of many algebraic operations that can be used by carpenters and builders.

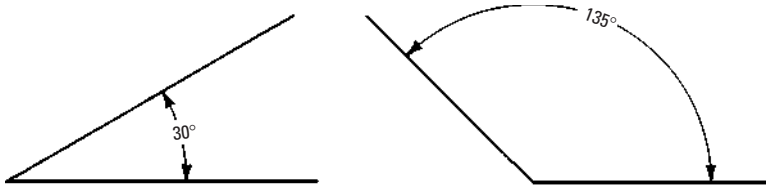

(A) An angle in the first quadrant is an acute angle.

(B) An angle in the second quadrant is an obtuse angle.

Figure 5-66 Acute and obtuse angles.

Trigonometric Functions

In mathematics, a *function* means a quantity that necessarily changes because of a change in another number with which it is connected

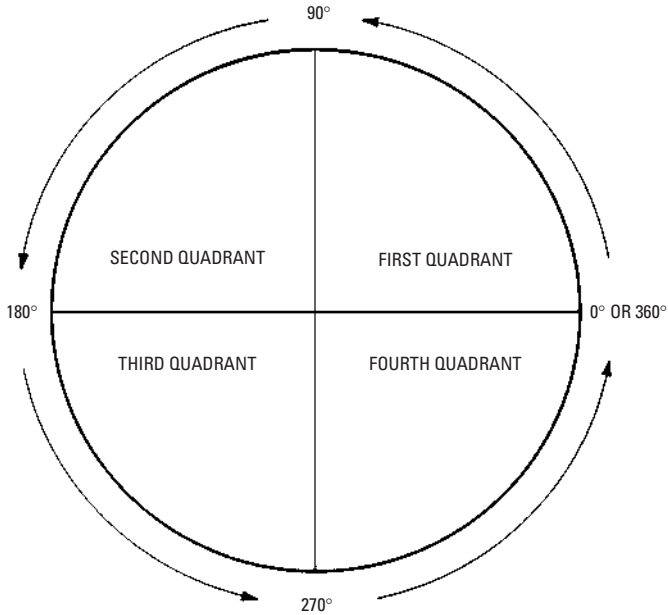

Figure 5-67 The four quadrants of a circle.

in some way. In trigonometry, it is probably less confusing to call the trigonometric functions simply *ratios*, which they truly are.

Refer to Figure 5-68 for an explanation of the following. There are six trigonometric ratios commonly used:

$$\text{Sine of angle } A = \frac{\text{opposite side}}{\text{hypotenuse}} \text{ or } \frac{BC}{AB}$$

$$\text{Cosine of angle } A = \frac{\text{adjacent side}}{\text{hypotenuse}} \text{ or } \frac{AC}{AB}$$

$$\text{Tangent of angle } A = \frac{\text{opposite side}}{\text{adjacent side}} \text{ or } \frac{BC}{AC}$$

$$\text{Cotangent of angle } A = \frac{\text{adjacent side}}{\text{opposite side}} \text{ or } \frac{AC}{BC}$$

$$\text{Secant of angle } A = \frac{\text{hypotenuse}}{\text{adjacent side}} \text{ or } \frac{AB}{AC}$$

$$\text{Cosecant of angle } A = \frac{\text{hypotenuse}}{\text{opposite side}} \text{ or } \frac{AB}{BC}$$

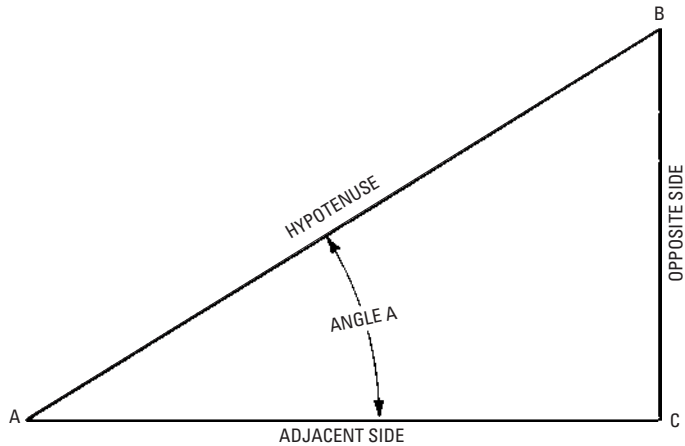

Figure 5-68 A right triangle illustrates the application of trigonometric ratios that are commonly used.

Note that these last three functions are only *reciprocals* of the sine, cosine, and tangent, respectively, or

$$\text{cosecant} = \frac{1}{\text{sine}}$$

$$\text{secant} = \frac{1}{\text{cosine}}$$

$$\text{cotangent} = \frac{1}{\text{tangent}}$$

If a proposition calls for multiplication by the sine of an angle, the same result will be obtained by dividing by the cosecant. It is convenient to do this in many calculations.

It is impossible in a discussion of this type to give a comprehensive table of the trigonometric ratios, although an adequate (but limited) number of trigonometric functions is presented in Table 5-25. Those who would like to follow up the information given here are advised to obtain a book of five- or six-place tables.

As an example of how trigonometric ratios are used to solve one of the carpenter's most common problems (determining the length of rafters given the rise and run), refer to Figure 5-69. The slope of the roof, in degrees, may be determined by dividing the opposite side, 12 feet, by the adjacent side, 18 feet. This is the tangent of the

Table 5-25 Natural Trigonometric Functions

Degree	*Sine*	*Cosine*	*Tangent*	*Secant*
0	0.00000	1.0000	0.00000	1.0999
1	0.01745	0.9998	0.01745	1.0001
2	0.03490	0.9994	0.03492	1.0006
3	0.05234	0.9986	0.05241	1.0014
4	0.06976	0.9976	0.06993	1.0024
5	0.08716	0.9962	0.08749	1.0038
6	0.10453	0.9945	0.10510	1.0055
7	0.12187	0.9925	0.12278	1.0075
8	0.1392	0.9903	0.1405	1.0098
9	0.1564	0.9877	0.1584	1.0125
10	0.1736	0.9848	0.1763	1.0154
11	0.1908	0.9816	0.1944	1.0187
12	0.2079	0.9781	0.2126	1.0223
13	0.2250	0.9744	0.2309	1.0263
14	0.2419	0.9703	0.2493	1.0306
15	0.2588	0.9659	0.2679	1.0353
16	0.2756	0.9613	0.2867	1.0403
17	0.2924	0.9563	0.3057	1.0457
18	0.3090	0.9511	0.3249	1.0515
19	0.3256	0.9455	0.3443	1.0576
20	0.3420	0.9397	0.3640	1.0642
21	0.3584	0.9336	0.3839	1.0711
22	0.3746	0.9272	0.4040	1.0785
23	0.3907	0.9205	0.4245	1.0864
24	0.4067	0.9135	0.4452	1.0946
25	0.4226	0.9063	0.4663	1.1034
26	0.4384	0.8988	0.4877	1.1126
27	0.4540	0.8910	0.5095	1.1223
28	0.4695	0.8829	0.5317	1.1326
29	0.4848	0.8746	0.5543	1.1433
30	0.5000	0.8660	0.5774	1.1547
31	0.5150	0.8572	0.6009	1.1663
32	0.5299	0.8480	0.6249	1.1792
33	0.5446	0.8387	0.6494	1.1924
34	0.5592	0.8290	0.6745	1.2062
35	0.5736	0.8192	0.7002	1.2208
36	0.5878	0.8090	0.7265	1.2361

(continued)

Table 5-25 (*continued*)

Degree	*Sine*	*Cosine*	*Tangent*	*Secant*
37	0.6018	0.7986	0.7536	1.2521
38	0.6157	0.7880	0.7813	1.2690
39	0.6293	0.7771	0.8098	1.2867
40	0.6428	0.7660	0.8391	1.3054
41	0.6561	0.7547	0.8693	1.3250
42	0.6691	0.7431	0.9004	1.3456
43	0.6820	0.7314	0.9325	1.3673
44	0.6947	0.7193	0.9657	1.3902
45	0.7071	0.7071	1.0000	1.4142
46	0.7193	0.6947	1.0355	1.4395
47	0.7314	0.6820	1.0724	1.4663
48	0.7431	0.6691	1.1106	1.4945
49	0.7547	0.6561	1.1504	1.5242
50	0.7660	0.6428	1.1918	1.5557
51	0.7771	0.6293	1.2349	1.5890
52	0.7880	0.6157	1.2799	1.6243
53	0.7986	0.6018	1.3270	1.6616
54	0.8090	0.5878	1.3764	1.7013
55	0.8192	0.5736	1.4281	1.7434
56	0.8290	0.5592	1.4826	1.7883
57	0.8387	0.5446	1.5399	1.8361
58	0.8480	0.5299	1.6003	1.8871
59	0.8572	0.5150	1.6643	1.9416
60	0.8660	0.5000	1.7321	2.0000
61	0.8746	0.4848	1.8040	2.0627
62	0.8829	0.4695	1.8807	2.1300
63	0.8910	0.4540	1.9626	2.2027
64	0.8988	0.4384	2.0503	2.2812
65	0.9063	0.4226	2.1445	2.3662
66	0.9135	0.4067	2.2460	2.4586
67	0.9205	0.3907	2.3559	2.5598
68	0.9272	0.3746	2.4751	2.6695
69	0.9336	0.3584	2.6051	2.7904
70	0.9397	0.3420	2.7475	2.9238
71	0.9455	0.3256	2.9042	3.0715
72	0.9511	0.3090	3.0777	3.2361

(continued)

Table 5-25 (*continued*)

Degree	*Sine*	*Cosine*	*Tangent*	*Secant*
73	0.9563	0.2924	3.2709	3.4203
74	0.9613	0.2756	3.4874	3.6279
75	0.9659	0.2588	3.7321	3.8637
76	0.9703	0.2419	4.0108	4.1336
77	0.9744	0.2250	4.3315	4.4454
78	0.9781	0.2079	4.7046	4.8097
79	0.9816	0.1908	5.1446	5.2408
80	0.9848	0.1736	5.6713	5.7588
81	0.9877	0.1564	6.3138	6.3924
82	0.9903	0.1392	7.1154	7.1853
83	0.9925	0.12187	8.1443	8.2055
84	0.9945	0.10453	9.5144	9.5668
85	0.9962	0.08716	11.4301	11.474
86	0.9976	0.06976	14.3007	14.335
87	0.9986	0.05234	19.0811	19.107
88	0.9994	0.03490	28.6363	28.654
89	0.9998	0.01745	57.2900	57.299
90	1.0000	Inf.	Inf.	Inf.
	–	–	–	–

angle A and is equal to $^{12}/_{18}$ or .6667. From Table 5-25, angle A is determined to be 33° 42′. The length of the rafter may be determined by the ratio:

$$\text{secant} = \frac{\text{hypotenuse}}{\text{adjacent side}}$$

or the hypotenuse (the length of the rafter) is equal to the following:

$$\text{Secant } 33°42' \times \text{adjacent side}$$

The secant of 33°42′ is equal to 1.2020. Therefore, the calculation for the length of the rafter is the following:

$$1.2020 \times 18 = 21.64 \text{ feet} = 21 \text{ feet } 7^{3}/_{16} \text{ inches}$$

Since the opposite side is known to be 12 feet, the calculation could just as easily be made by using the cosecant function.

Table 5-26 shows the slopes (in degrees) for all regular roof pitches. These pitches range from 12 × 1 to 12 × 12, and the three

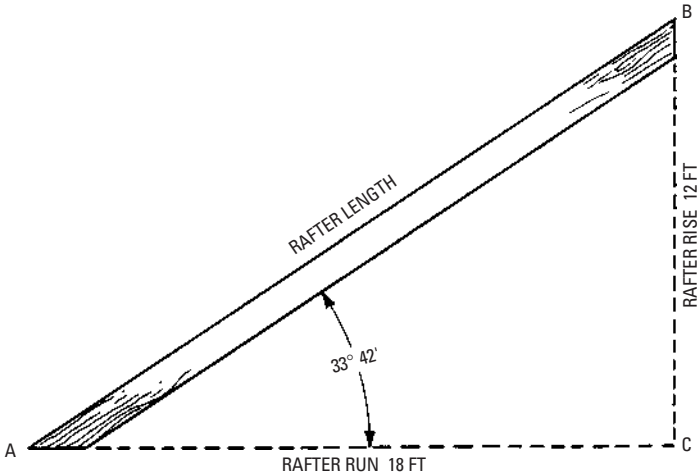

Figure 5-69 Trigonometric ratios may be used to determine the lengths of rafters for a roof.

Table 5-26 Roof Pitches in Degrees and Minutes (Measured from the Horizontal)

Pitch	*Sine*	*Cosine*	*Tangent*
$12 \times 1 = 4° 46'$	0.083098	0.996541	0.083386
$12 \times 2 = 9° 28'$	0.164474	0.986381	0.166745
$12 \times 3 = 14° 2'$	0.242486	0.970155	0.249946
$12 \times 4 = 18° 26'$	0.316201	0.948692	0.333302
$12 \times 5 = 22° 37'$	0.384564	0.923098	0.416601
$12 \times 6 = 26° 34'$	0.444635	0.895712	0.496404
$12 \times 7 = 30° 15'$	0.503774	0.863836	0.583183
$12 \times 8 = 33° 41'$	0.554602	0.832115	0.666497
$12 \times 9 = 36° 53'$	0.600188	0.799859	0.750366
$12 \times 10 = 39° 46'$	0.639663	0.768656	0.832183
$12 \times 11 = 42° 31'$	0.675805	0.737081	0.916866
$12 \times 12 = 45° 00'$	0.707107	0.707107	1.000000

main trigonometric ratios (sine, cosine, and tangent) are provided for each pitch.

Other typical examples of how trigonometric ratios can aid carpenters are shown in the following problems.

Problem 1

A grillwork consisting of radial and vertical members is to be built in a semicircular opening with a radius of 6 feet, as shown in Figure 5-70. Find the lengths of the vertical pieces MS and LF.

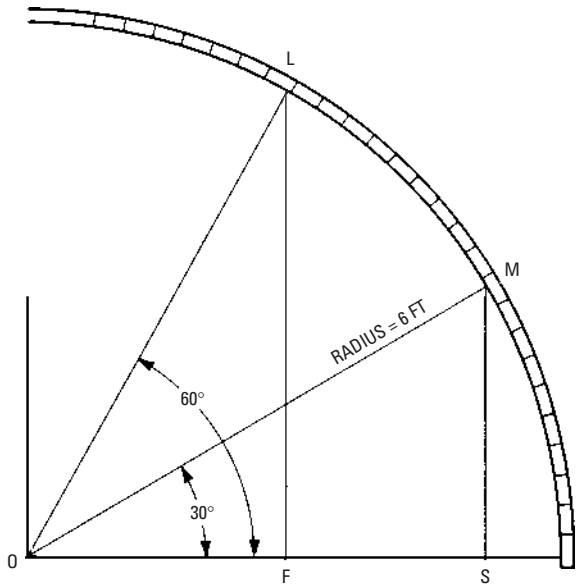

Figure 5-70 The method of finding the length of vertical pieces in grillwork with the aid of trigonometric relations.

For triangle OMS, the hypotenuse is known to be 6 feet, and angle O is 30°. Line MS is the opposite side of the triangle, and

$$\text{opposite side} = 30°, \text{ or}$$

$$\frac{\text{opposite side}}{\text{hypotenuse}} = \text{sine } 30° \times \text{hypotenuse}$$

$$\text{sine } 30° = 0.500$$

This is the calculation:

$$\text{opposite side} = 0.500 \times 6 = 3 \text{ feet}$$

For triangle *OLF*, the hypotenuse is 6 feet, and angle *O* is 60°. Line *LF* is the opposite side, and

$$\frac{\text{opposite side}}{\text{hypotenuse}} = 60°,$$

$$\text{or opposite side} = 60° \times \text{hypotenuse.}$$

$$\text{sine } 60° = 0.866$$

This is the calculation:

$$\text{opposite side} = 0.866 \times 6 = 5.196 \text{ feet} = 5 \text{ feet } 2^3/_8 \text{ inches}$$

Problem 2

When laying out the grillwork in Figure 5-70, how far must the members LF and MS be spaced from the center O to be vertical?

The hypotenuse is known to be 6 feet and the length of adjacent side *OF* is to be found.

$$\frac{\text{adjacent side}}{\text{hypotenuse}} = \cos 60°, \text{ or}$$

$$\text{adjacent side} = \cos 60° \times \text{hypotenuse.}$$

$$\cos 60° = 0.500$$

This is the calculation:

$$\text{adjacent side } OF = .500 \times 6 = 3 \text{ feet}$$

For the length of the adjacent side OS

$$\frac{\text{adjacent side}}{\text{hypotenuse}} = \cos 30°, \text{ or}$$

$$\text{adjacent side} = \cos 30° \times \text{hypotenuse.}$$

$$\cos 30° = 0.866$$

This is the calculation:

$$\text{Adjacent side } OS = .866 \times 6 = 5.196 \text{ feet}$$
$$= 5 \text{ feet } 2^3/_8 \text{ inches}$$

Problem 3

A bridge is to be constructed from the top of a building to an opening in the roof of an adjacent building, as in Figure 5-71. If the rise OF to the point of entry L is 15 feet and the pitch of the roof is $^1/_2$*, what length beams FL are required?*

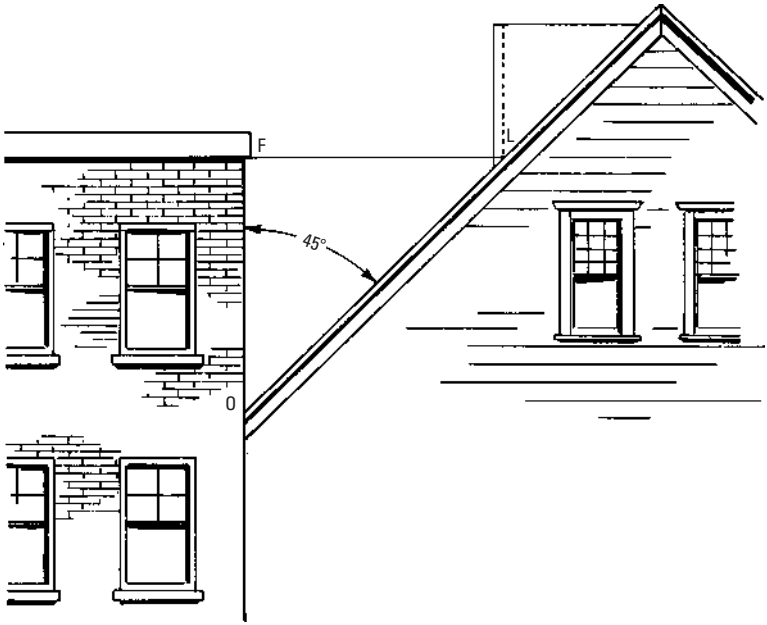

Figure 5-71 The method of finding the distance from one side of a building to a given point on an adjacent roof by employing trigonometric relations.

From Table 5-26, $^1/_2$ pitch, or 12 inches × 12 inches, is 45°. The adjacent side *OF* is known to be 15 feet. The required length of the opposite side = adjacent side × tan 45°. This is the calculation:

$$\text{opposite side } FL = 15 \times 1.00 = 15 \text{ feet}$$

Problem 4

When estimating the amount of roofing material necessary to cover the side of the roof from O to L in Figure 5-71, what is the distance from O to L?

It is required to find the hypotenuse with the adjacent side and included angle being given.

$$\frac{\text{hypotenuse}}{\text{adjacent side}} = \text{secant angle 0, or}$$

$$\text{hypotenuse} = 1.4142 \times 15 = 21.213 \text{ feet}$$

$$= 21 \text{ feet } 2^9/_{16} \text{ inches.}$$

Table 5-27 Functions of Numbers

							No. = Diameter	
No.	*Square*	*Cube*	*Square Root*	*Cubic Root*	*Logarithm*	*1000 × Reciprocal*	*Circum.*	*Area*
1	1	1	1.0000	1.0000	0.00000	1000.000	3.142	0.7854
2	4	8	1.4142	1.2599	0.30103	500.000	6.283	3.1416
3	9	27	1.7321	1.4422	0.47712	333.333	9.425	7.0686
4	16	64	2.0000	1.5874	0.60206	250.000	12.566	12.5664
5	25	125	2.2361	1.7100	0.69897	200.000	15.708	19.6350
6	36	216	2.4495	1.8171	0.77815	166.667	18.850	28.2743
7	49	343	2.6458	1.9129	0.84510	142.857	21.991	38.4845
8	64	512	2.8284	2.0000	0.90309	125.000	25.133	50.2655
9	81	729	3.0000	2.0801	0.95424	111.111	28.274	63.6173
10	100	1000	3.1623	2.1544	1.00000	100.000	31.416	78.5398
11	121	1331	3.3166	2.2240	1.04139	90.9091	34.558	95.0332
12	144	1728	3.4641	2.2894	1.07918	83.3333	37.699	113.097
13	169	2197	3.6056	2.3513	1.11394	76.9231	40.841	132.732
14	196	2744	3.7417	2.4101	1.14613	71.4286	43.982	153.938
15	225	3375	3.8730	2.4662	1.17609	66.6667	47.124	176.715
16	256	4096	4.0000	2.5198	1.20412	62.5000	50.265	201.062
17	289	4913	4.1231	2.5713	1.23045	58.8235	53.407	226.980
18	324	5832	4.2426	2.6207	1.25527	55.5556	56.549	254.469
19	361	6859	4.3589	2.6684	1.27875	52.6316	59.690	283.529
20	400	8000	4.4721	2.7144	1.30103	50.0000	62.832	314.159

21	441	9261	4.5826	2.7589	1.32222	47.6190	65.973	346.361
22	484	10648	4.6904	2.8020	1.34242	45.4545	69.115	380.133
23	529	12167	4.7958	2.8439	1.36173	43.4783	72.257	415.476
24	576	13824	4.8990	2.8845	1.38021	41.6667	75.398	452.389
25	625	15625	5.0000	2.9240	1.39794	40.0000	78.540	490.874
26	676	17576	5.0990	2.9625	1.41497	38.4615	81.681	530.929
27	729	19683	5.1962	3.0000	1.43136	37.0370	84.823	572.555
28	784	21952	5.2915	3.0366	1.44716	35.7143	87.965	615.752
29	841	24389	5.3852	3.0723	1.46240	34.4828	91.106	660.520
30	900	27000	5.4772	3.1072	1.47712	33.3333	94.248	706.858
31	961	29791	5.5678	3.1414	1.49136	32.2581	97.389	754.768
32	1024	32768	5.6569	3.1748	1.50515	31.2500	100.531	804.248
33	1089	35937	5.7446	3.2075	1.51851	30.3030	103.673	855.299
34	1156	39304	5.8310	3.2396	1.53148	29.4118	106.814	907.920
35	1225	42875	5.9161	3.2711	1.54407	28.5714	109.956	962.113
36	1296	46656	6.0000	3.3019	1.55630	27.7778	113.097	1017.88
37	1369	50653	6.0828	3.3322	1.56820	27.0270	116.239	1075.21
38	1444	54872	6.1644	3.3620	1.57978	26.3158	119.381	1134.11
39	1521	59319	6.2450	3.3912	1.59106	25.6410	122.522	1194.59
40	1600	64000	6.3246	3.4200	1.60206	25.0000	125.66	1256.64
41	1681	68921	6.4031	3.4482	1.61278	24.3902	128.81	1320.25
42	1764	74088	6.4807	3.4760	1.62325	23.8095	131.95	1385.44
43	1849	79507	6.5574	3.5034	1.63347	23.2558	135.09	1452.20
44	1936	85184	6.6332	3.5303	1.64345	22.7273	138.23	1520.53

(continued)

Table 5-27 (*continued*)

							No. = Diameter	
No.	*Square*	*Cube*	*Square Root*	*Cubic Root*	*Logarithm*	*1000 × Reciprocal*	*Circum.*	*Area*
45	2025	91125	6.7082	3.5569	1.65321	22.2222	141.37	1590.43
46	2116	97336	6.7823	3.5830	1.66276	21.7391	144.51	1661.90
47	2209	103823	6.8557	3.6088	1.67210	21.2766	147.65	1734.94
48	2304	110592	6.9282	3.6342	1.68124	20.8333	150.80	1809.56
49	2401	117649	7.0000	3.6593	1.69020	20.4082	153.94	1885.74
50	2500	125000	7.0711	3.6840	1.69897	20.0000	157.08	1963.50
51	2601	132651	7.1414	3.7084	1.70757	19.6078	160.22	2042.82
52	2704	140608	7.2111	3.7325	1.71600	19.2308	163.36	2123.72
53	2809	148877	7.2801	3.7563	1.72428	18.8679	166.50	2206.18
54	2916	157464	7.3485	3.7798	1.73239	18.5185	169.65	2290.22
55	3025	166375	7.4162	3.8030	1.74036	18.1818	172.79	2375.83
56	3136	175616	7.4833	3.8259	1.74819	17.8571	175.93	2463.01
57	3249	185193	7.5498	3.8485	1.75587	17.5439	179.07	2551.76
58	3364	195112	7.6158	3.8709	1.76343	17.2414	182.21	2642.08
59	3481	205379	7.6811	3.8930	1.77085	16.9492	185.35	2733.97
60	3600	216000	7.7460	3.9149	1.77815	16.6667	188.50	2827.43
61	3721	226981	7.8102	3.9365	1.78533	16.3934	191.64	2922.47
62	3844	238328	7.8740	3.9579	1.79239	16.1290	194.78	3019.07
63	3969	250047	7.9373	3.9791	1.79934	15.8730	197.92	3117.25
64	4096	262144	8.0000	4.0000	1.80618	15.6250	201.06	3216.99

65	4225	274625	8.0623	4.0207	1.81291	15.3846	204.20	3318.31
66	4356	287496	8.1240	4.0412	1.81954	15.1515	207.35	3421.19
67	4489	300763	8.1854	4.0615	1.82607	14.9254	210.49	3525.65
68	4624	314432	8.2462	4.0817	1.83251	14.7059	213.63	3631.68
69	4761	328509	8.3066	4.1016	1.83885	14.4928	216.77	3739.28
70	4900	343000	8.3666	4.1213	1.84510	14.2857	219.91	3848.45
71	5041	357911	8.4261	4.1408	1.85126	14.0845	223.05	3959.19
72	5184	373248	8.4853	4.1602	1.85733	13.8889	226.19	4071.50
73	5329	389017	8.5440	4.1793	1.86332	13.6986	229.34	4185.39
74	5476	405224	8.6023	4.1983	1.86923	13.5135	232.48	4300.84
75	5625	421875	8.6603	4.2172	1.87506	13.3333	235.62	4417.86
76	5776	438976	8.7178	4.2358	1.88081	13.1579	238.76	4536.46
77	5929	456533	8.7750	4.2543	1.88649	12.9870	241.90	4656.63
78	6084	474552	8.8318	4.2727	1.89209	12.8205	245.04	4778.36
79	6241	493039	8.8882	4.2908	1.89763	12.6582	248.19	4901.67
80	6400	512000	8.9443	4.3089	1.90309	12.5000	251.33	5026.55
81	6561	531441	9.0000	4.3267	1.90849	12.3457	254.47	5153.00
82	6724	551368	9.0554	4.3445	1.91381	12.1951	257.61	5281.02
83	6889	571787	9.1104	4.3621	1.91908	12.0482	260.75	5410.61
84	7056	592704	9.1652	4.3795	1.92428	11.9048	263.89	5541.77
85	7225	614125	9.2195	4.3968	1.92942	11.7647	267.04	5674.50
86	7396	636056	9.2736	4.4140	1.93450	11.6279	270.18	5808.80
87	7569	658503	9.3274	4.4310	1.93952	11.4943	273.32	5944.68

(continued)

Table 5-27 (*continued*)

							No. = Diameter	
No.	*Square*	*Cube*	*Square Root*	*Cubic Root*	*Logarithm*	*1000 × Reciprocal*	*Circum.*	*Area*
88	7744	681472	9.3808	4.4480	1.94448	11.3636	276.46	6082.12
89	7921	704969	9.4340	4.4647	1.94939	11.2360	279.60	6221.14
90	8100	729000	9.4868	4.4814	1.95424	11.1111	282.74	6361.73
91	8281	753571	9.5394	4.4979	1.95904	10.9890	285.88	6503.88
92	8464	778688	9.5917	4.5144	1.96379	10.8696	289.03	6647.61
93	8649	804357	9.6437	4.5307	1.96848	10.7527	292.17	6792.91
94	8836	830584	9.6954	4.5468	1.97313	10.6383	295.31	6939.78
95	9025	857375	9.7468	4.5629	1.97772	10.5263	298.45	7088.22
96	9216	884736	9.7980	4.5789	1.98227	10.4167	301.59	7238.23
97	9409	912673	9.8489	4.5947	1.98677	10.3093	304.73	7389.81
98	9604	941192	9.8995	4.6104	1.99123	10.2041	307.88	7542.96
99	9801	970299	9.9499	4.6261	1.99564	10.1010	311.02	7697.69

Functions of Numbers

Table 5-27 shows the functions of numbers.

Summary

All the sciences are based on arithmetic and the ability to use it. Arithmetic is the art of calculating by using numbers. A number is a total amount (or aggregate) of units. By computing the units, we arrive at a certain number or total. Similarly, a unit means a single article, often a definite group adopted as a standard of measurement (such as dozen, ton, foot, bushel, or mile).

Fractions indicate that a number or unit has been divided into a certain number of equal parts, and shows how many of these parts are to be considered. Two forms of fractions are in common usage: the decimal and the common fraction. The common fraction is written by using two numbers, one written over or alongside the other with a line between them, the lower (or second) number being called the denominator, and the upper (or first) number being called the numerator.

Geometry is a branch of mathematics that deals with space and figures in space. It is the science of the mutual relations of points, lines, angles, surfaces, and solids, which are considered as having no properties except those arising from extension and difference of situation. There are two kinds of lines—straight and curved. A straight line is the shortest distance between two points. A curved line is one that changes its direction at every point.

Trigonometry is the branch of mathematics that deals with the relations that exist between the sides and angles of triangles, especially the methods of calculating the required parts of triangles from given parts. There are six elements, or parts, in every triangle: three sides and three angles. The sum of the three angles, no matter what the lengths of the sides, will always be equal to 180 degrees.

Review Questions

1. What is the definition of arithmetic?
2. What are even numbers?
3. What are odd numbers?
4. What are fractions? How are they used?
5. What is trigonometry and how is it used in carpentry?

Chapter 6

Surveying

By definition, *surveying* means the art or science of determining the area and configuration of portions of the surface of the Earth. The two general divisions of surveying may be classified with respect to the nature of the measurements taken:

- Leveling
- Measurement of angles (transit work)

In surveying, *leveling* is the operation of determining the comparative levels of different points of land for the purpose of laying out a grade or building site, and so on, by sighting through a leveling instrument at one point to a leveling staff at another point (see Figure 6-1).

The Level

This instrument (see Figure 6-2) is employed to determine the difference in elevation between points. A common form is known as the *wye level*, so-called because its shape resembles the letter Y. It consists of a telescope mounted on two supports that from their shape are called Ys. The crossbar supporting the telescope is attached to a vertical spindle that allows it to be turned in a horizontal plane. Directly beneath the telescope and attached parallel to it is a spirit level by means of which the line of collimation of the telescope may be rendered horizontal. The *line of collimation* is the line that would connect the intersection of the crosshairs with the optical center of the objective.

Construction of the Wye Level

In construction, a circular plate is screwed to a tripod, and to this is attached a similar plate parallel to the first and connected with it by a ball-and socket joint. Four screws (sometimes only three), called *foot* or *plate screws*, hold these plates apart by resting on the lower one and passing through the other. A vertical spindle in the center of the plates supports a rod, bar, or beam, and is used to revolve the instrument. The beam is horizontal and carries at its ends two vertical standards or supports of equal size terminated by two forks of the general form of the letter Y. The inside of the Ys is Y-shaped, with an open bottom to prevent an accumulation of dirt. The top of the Ys may be closed by semicircular straps or bridles (called *clips*) that are hinged on one side and pinned on the other. The pins are tapered to permit fastening of the telescope. It is *never* clamped tightly.

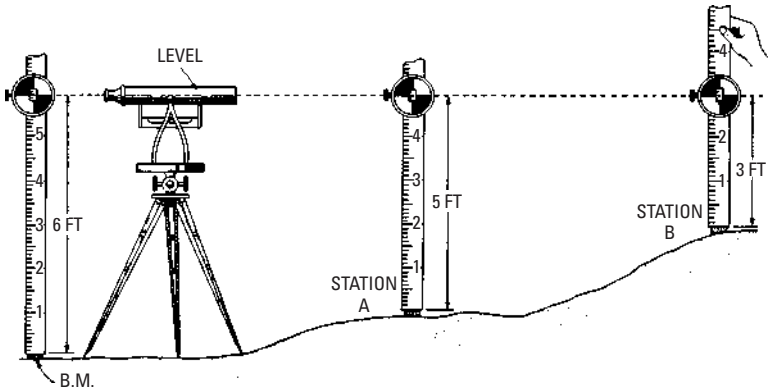

Figure 6-1 The terms backsight and foresight do not necessarily mean backward and forward. Readings taken on a point of known elevation (such as a benchmark or a turning point) are called backsights, whereas readings taken on a point of unknown elevation are termed foresights. In the illustration, the backsight (to the benchmark) is + 6 feet 0 inches, and the foresight to station A is −5 feet 0 inches. The difference in the two elevations is 6 feet −5 feet = 1 foot. If the reading at station A had been greater than the 6-foot HI (height of instrument), the calculation would have shown a negative result, thereby indicating that station A was lower than the benchmark. The elevation at station B is calculated in the same manner: 6 − 3 = 3 feet. Therefore, the elevations from any number of points can be obtained in the same way, if they can be seen from the same position of the instrument. If they cannot, a new HI must be used.

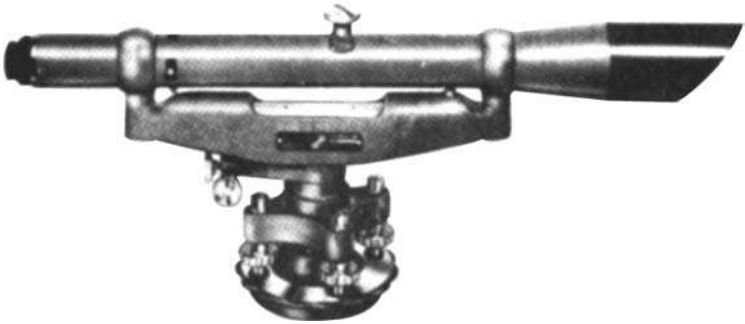

Figure 6-2 A typical builder's level. *(Courtesy David White Instruments, Division of Realist, Inc.)*

The tops of the Ys and the corresponding clips are called the *rings* or *collars,* and should be of equal diameter. A *telescope* is placed on the rings that support a spirit level. A *clamping screw* just above the upper plate serves to secure the instrument in any position desired. A *tangent screw* (also above the upper plate) provides slow motion (or *vernier*) adjustment to the instrument.

The Telescope

The *objective*, or *object glass* (so-called because it faces the object looked at) is a compound lens that is made to correct spherical and chromatic aberrations of single lenses. It gathers light and forms an image at a point in the tube where crosshairs are placed. The *ocular* (or eye) *piece* is also a compound lens through which the operator looks to see a magnified view of the image. In the best precision instruments (often foreign-made), the image is often *inverted*. A good instrument worker quickly becomes accustomed to the inverted image, but most American-made instruments have an *erecting* image that shows the object right side up. Tangent screws may be used to give motion to the tubes carrying the objective and ocular.

The Crosshairs

These are made of platinum-drawn wires or spider's threads attached to a ring within the telescope at the spot where the image is formed. The ring is secured by four capstan-headed screws that pass through the telescope tube. There are commonly two hairs, one horizontal and the other vertical, with their intersection in the axis of the telescope.

Bubble Level

The spirit level attached to the telescope can be raised vertically by means of *altitude screws* at the rear end, and it may be moved laterally to a limited extent by means of *azimuth screws* at the forward end.

The Supports

These form the Ys and are supported by the bar to that they are fastened by two nuts, one above and one below. These nuts may be moved to provide an adjustment (to move the scope in a horizontal direction).

Lines of the Level

The three principal lines of a level are as follows:

- *Vertical axis*—This passes through the center of the spindle.
- *Bubble line*—The metallic supports of the spirit level are equal, and the tangent at their top or bottom is horizontal

when the bubble is centered. This tangent is the bubble line.

- *Line of collimation*—The line that would connect the intersection of the crosshairs with the optical center of the objective is the line of collimation.

The following relations must be obtained:

- The bubble line and the line of collimation must be parallel.
- The plane described by the bubble line should be horizontal, that is, perpendicular to the vertical axis.

These conditions are generally satisfied in a new level, but exposure and use may alter these relations. Therefore, occasional adjustment of the instrument may be necessary.

Adjustments of the Wye Level

Levels and transits are expensive, precision instruments and should be treated as such. Although a passable job of leveling may be done by a relatively inexperienced worker, it is questionable if a major job of adjusting should be attempted by a novice. A perfect job of adjustment is difficult, even for an experienced adjuster, and there are few instruments in perfect adjustment. For precision work, the adjustment should be checked constantly. The first relation given earlier cannot be established directly. However, it does require several adjustments.

First Adjustment

Collimating the instrument means making the line of collimation parallel to the bottom element of the collars

1. Clamp the instrument, and unclip the collars.
2. Sight at some distant point, a point that is distinct.
3. Bring the horizontal crosshair on that point.
4. Carefully turn the telescope in the collars by one-half a revolution around its axis, and sight again. If the horizontal crosshair is still on the sighted point, the telescope is collimated with regard to that crosshair. If it is off the point, bring it halfway back by means of the capstan-headed screws and the rest of the way by the plate screws.
5. Repeat the operation over another point.
6. Collimate it with regard to the other crosshair.
7. Leave the screws at a snug bearing.

Second Adjustment
This is where you set the bubble line in a plane with the bottom element of the collars.

1. Unclip the telescope, and clamp the instrument over a pair of plate screws.
2. Center the bubble by means of the plate screws.
3. Carefully (and very slowly) turn the telescope in the collars in a small arc to the right, then to the left. If the bubble moves from center, bring it back by means of the azimuth or side screws.

Third Adjustment
Setting the bubble line parallel to the bottom element of the collars.

1. Unclip the telescope, and clamp the instrument over a pair of plate screws.
2. Center the bubble by means of the plate screws.
3. Carefully take the telescope up, replacing it carefully in the Ys in the opposite direction (that is, the objective sighting in the direction where the eyepiece originally was). If the bubble has moved, bring it back halfway by means of the altitude or foot screws of the spirit level and the rest of the way by the plate screws.
4. Repeat in another direction until the adjustment is satisfactory.

The second relation is established by making the bubble line stay in the center of the graduation during a complete revolution of the instrument around its spindle.

Fourth Adjustment
Now, make the axis of the instrument (not of the telescope) vertical.

1. Pin the clips.
2. Clamp and center the bubble over a pair of plate screws.
3. Reverse the telescope over the same pair of plate screws.
4. Bring the bubble halfway back (if it has moved) by means of the plate screws.

Fifth Adjustment
Again, make the bubble remain centered during a full revolution of the instrument.

1. Center the bubble.

2. Revolve the instrument horizontally by a one-half revolution.
3. If the bubble moves, correct it halfway by means of the support screws (at the foot of the Ys). If the rings become worn and unequal, use the two-peg method of the dumpy level.

Dumpy Level Adjustment

Figure 6-3 shows the *dumpy level* (so-called because of its compactness). It is used mostly in England, although it is used to some extent in the United States because of the better stability of its adjustments over the wye level. The dumpy level differs from the wye level mainly in that the telescope of the dumpy level is permanently attached to the supports or uprights, but these uprights are adjustable. The two-peg adjustment method is as follows:

1. Drive two stakes (pegs) several hundred feet apart.
2. Set the instrument approximately halfway between them.
3. Level up and sight the rod that is held in succession on each stake. The difference in the readings is the true difference of the

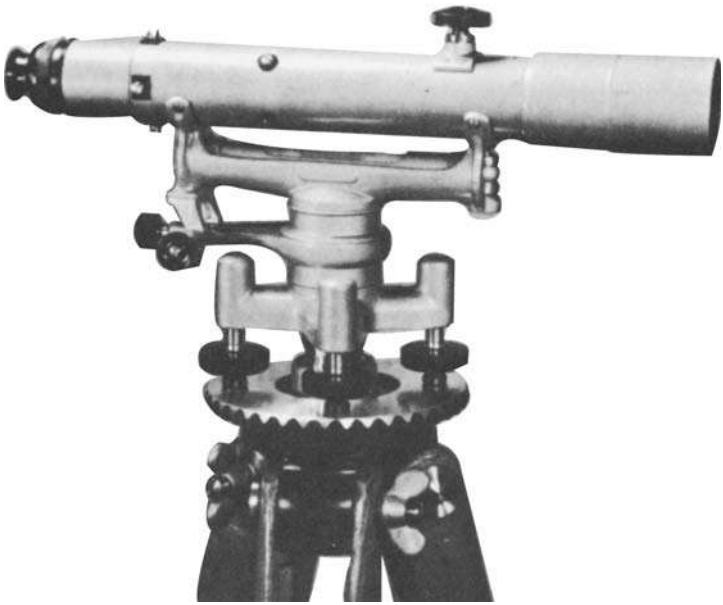

Figure 6-3 A dumpy level.

elevation of the stakes, even if the instrument is not in proper adjustment.

4. Test the instrument.
5. Set it near one of the stakes (the highest one, for instance). Level up and sight the rod held on the other stake.
6. Subtract the height of the instrument from the reading. The difference should be equal to the difference of elevation of the stakes as previously found. If these differences are not equal, set the target halfway between these readings, sight on it, and center the bubble by means of the altitude screws.
7. Repeat the operation until satisfaction is obtained.

Centering the Objective and Ocular

These adjustments are made permanently by the manufacturer. Usually, four screws hold the tubes carrying the glasses. Their heads pass through the outside tube where, after permanent adjustment, they are covered by a metallic ring.

Parallax

This is the apparent motion of the crosshairs on the object sighted when the eye is moved slightly. It shows the imperfect focusing of the ocular over the crosshairs. To correct this condition, hold a white surface (such as that of a piece of paper) slightly in front of the objective, and move the ocular tube in and out until the crosshairs are perfectly defined.

Leveling Rod

This instrument, used in leveling, is usually $6^1/_2$ feet high, graduated to hundredths of a foot and provided with a sliding target. The rod is made in two parts, arranged so that its length can be extended to 12 feet. Precision rods are of one-piece construction and have no target. Builders' rods may be graduated in feet, inches, and eighths of an inch, with a vernier reading in 64ths of an inch. A sliding disc (called a *target*) is provided with a vernier for extremely accurate work, reading to thousandths of a foot.

In use, the rod is held in a vertical position with its lower end resting on the desired point of elevation. The target is then moved up and down until its center coincides with the crosshairs in the telescope of the level. The reading of the elevation is made from the rod on a line corresponding with the centerline of the target. There are various kinds of rods. Some are designed to be read by the rodman, while others can be read through the telescope of the level.

Methods of Leveling

The simplest type of leveling is to find the difference in level between two points that are visible from a third point, the difference in level being less than the length of the leveling rod (see Figure 6-4).

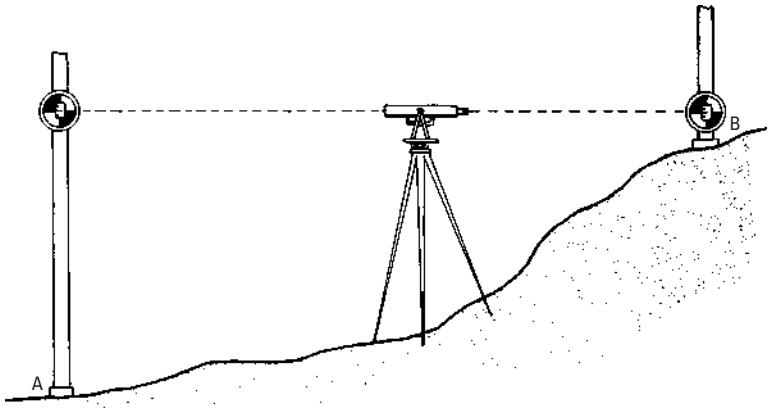

Figure 6-4 Leveling between two points whose difference in level is less than the length of the rod.

1. Set up and level the instrument at some point approximately halfway between the two points.
2. Have a worker hold the rod vertically on one of the points, and move the target up and down until its center coincides with the crosshairs of the level.
3. Take a reading. This is the HI, or height of the instrument above the benchmark *A*.
4. Turn the telescope on its spindle, have the rod held on the other point, and take a similar reading at *B*. The difference in level is equal to the difference in the readings.
5. If the difference in level is greater than the length of the rod, use the method shown in Figure 6-5. Divide the distance between the two points into sections of such length that the difference in level between the dividing points *A*, *B*, and *C* (called *stations*) are less than the length of the rod. Set up and level between points *A* and *B*.
6. Measure the distance *Aa* that is called *backsight*. Then, reverse the telescope, and take reading *Bb* that is called *foresight*.

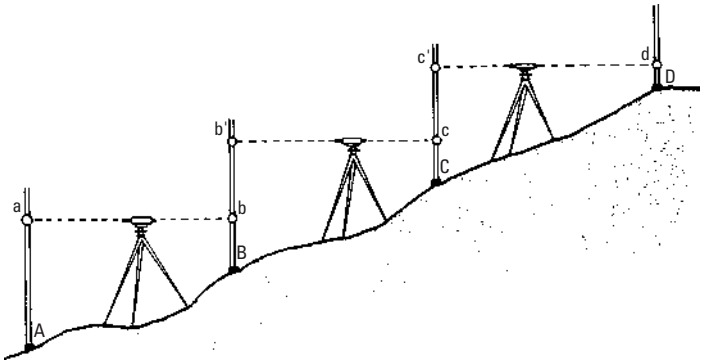

Figure 6-5 Leveling between points whose difference in level is greater than the length of the rod.

7. Next, set up and level between B and C, and take readings Bb' and Cc.
8. Repeat the operation between C and D, taking readings Cc' and Dd. The difference in level between stations A and D is equal to the sum of the differences between the intermediate stations. That is, this difference equals $(Aa - Bb) + (Bb' - Cc) - (Cc' - Dd)$, or $(Aa + Bb' + Cc') - (Bb + Cc + Dd)$.

Usually, you'll have to find the relative elevations of several points, as in grading work, in which case it is necessary to keep more elaborate notes and to measure distances between the stations. The method employed for this type of leveling is shown in Figure 6-6, and the field notes are recorded as shown in Table 6-1.

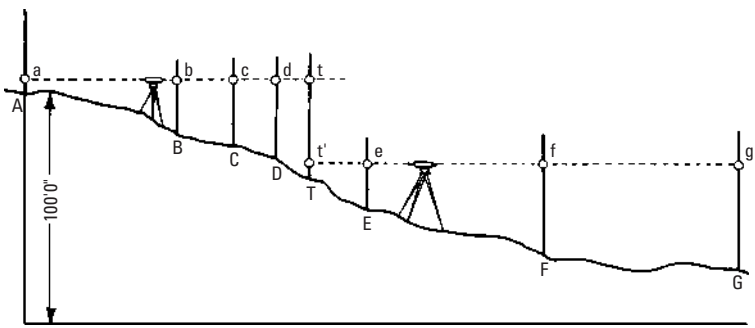

Figure 6-6 Finding the relative elevations of several points in rough terrain.

Table 6-1 Field Notes (Corresponding to the Operations Illustrated in Figure 6-6)

Station	*Distance*	*Backsight*	*Height of Instrument*	*Foresight*	*Elevation*	*Remarks*
A	0	4.2	104.2	.....	100.0	Bench mark, top of hydrant. Turning point
B	100	.....	.....	10.1	94.1	
C	60	.....	.....	7.3	96.9	
D	50	.....	.....	5.8	98.4	
T	.....	4.1	99.1	9.2	95.1	
E	70	6.8	.....	.....	92.3	
F	110	.....	.....	9.5	89.6	
G	80	.....	.....	11.5	87.6	

1. Assume a datum or reference line below the elevation of the lowest station, and refer all elevations to this line. Start at some permanently fixed point (such as a mark on a building or the top of a hydrant). This is called a *benchmark*.
2. Let A in Figure 6-6 be the benchmark, and assume a datum line 100 feet below the level of A.
3. Start with the instrument between A and B, and take a backsight on A. The distance Aa is found to be 4.2 feet, which, when added to 100 feet, gives the height of the instrument.
4. Next, take foresights on B, C, and D, and record these readings in the proper column. Readings Bb, Cc, and so on, subtracted from the height of the instrument, will give the elevations at B, C, and so on. This is done, and the results are recorded in the proper column of field notes.
5. The ground falls away so rapidly beyond D that it is necessary to set up the level farther along and, therefore, establish a new height of instrument. This is done by holding the rod at some convenient point (such as at T) called the *turning point*, and taking a foresight, which measures the distance Tt (9.2 feet).
6. The level is then set up in its second position between E and F, and a backsight is taken on the rod in the same position, which gives the distance Tt' (4.1 feet).
7. The distance $t't$ then equals $9.2 - 4.1 = 5.1$ feet, and this is subtracted from the previous height of instrument, thus giving the new HI, that is $104.2 - 5.1 = 99.1$ feet.

8. A backsight is now taken on *E*, and foresights are taken on *F* and *G*. These are recorded in the proper columns, and the elevations are found by subtracting these distances from the new *HI*.
9. The horizontal distances between the stations are measured with a tape and recorded in the second column.
10. When plotting a cross-section from notes kept in this manner, the datum line is drawn first, and perpendiculars are erected at points corresponding to the different stations. The proper elevations are then indicated on these vertical lines, and a contour line is drawn through the points so marked.

Directions for Using Level

Note carefully the following mode of procedure in leveling:

1. Center the bubble over one pair of plate screws, then over the other pair. Plate screws should have a snug bearing. When looking at the bubble or at the crosshairs, the eyes should look naturally (that is, without strain). Try to observe with both eyes open.
2. Adjust the eyepiece to the crosshairs for parallax.
3. Turn the instrument toward the target. It is better to level up facing the target.
4. Look again at the bubble.
5. Sight the target, and have it set right by motions according to a prearranged code with the worker holding the rod.
6. Look again at the bubble.
7. Read the rod or direct the target from the intersection of the crosshairs only.
8. Approve the target when sure.
9. Have the height of the target called out by the worker holding the rod.
10. Enter this height in the field book.
11. Quickly, make needed calculations.
12. Motion the worker holding the rod to a new station or to stay for a turning point and backsight, and move yourself to another position.

The following additional hints will also be useful:

- *Guarding against the sun*—Draw the telescope shade, or use an umbrella or a hat.

- *Length of sights*—Avoid sights too short and too long (250 to 350 feet should be the limit of the sights).
- *Equal sights*—The length of the backsight should practically be the same as the length of the foresight. This may be approximated by pacing, or by sighting with the stadia crosshairs in the telescope.
- *Long sights*—When sights longer than the maximum allowable in one direction only are unavoidable, correction should be made for curvature.
- *Leveling up or down a steep slope*—The leveler, after some practice, will place the instrument to take a reading near the top or the bottom of the rod (as the case may be), thus gaining vertical distance, but this produces unequal sights. The leveler may also follow a zig-zag course.
- *Leveling across a large body of water*
  - *A running stream*—Drive a stake to the water surface on each side of the stream and in a direction normal to the flow, although the line may not run so. Take a foresight reading on the first, a backsight reading on the second, and continue to and along the line. The elevations of the two stakes may be assumed equal.
  - *Across a pond*—If a pond or lake is too wide to ensure a good sighting across, use essentially the same method as for a stream. Drive stakes on each side and to the water surface. Take a foresight reading on the first and a backsight reading on the second.
- *Across a wall*—Take a foresight reading on the rod set on a stake, driven to the natural surface on the first side of the wall. Measure the height of the wall above the stake, and enter it as a backsight reading. Drive a stake to the natural surface on the second side of the wall. Measure the height of the wall on that side above the stake. Enter it as a foresight reading. Set the rod on the stake. Take a sight on it (which will be a backsight reading). Continue using this method until the leveling has been completed.
- *In underbrush*—If it cannot be cut down on the line of sight, find a high place or provide one by piling logs, rocks, and so on, to set the instrument on.
- *Through swamp*—Push the legs of the tripod down as far as possible. The leveler lies on his or her side. Two workers may

be necessary at the level. If the ground is still unsafe, drive stakes or piles to support the instrument.

- *Elevations taken at road crossings*—Take elevations both ways for some distance.
- *Elevations taken at river crossings*—Take elevations of high-water marks and flood marks, with the dates of it. Question the local residents for these dates. Also, ask for dates and data of extreme low water.
- *Proper length of sights*—This will depend on the distance at which the rod appears distinct and on the precision required. Under ordinary conditions, sights should not exceed 300 feet where elevations are required to the nearest .01 foot, and even at a much shorter distance, the boiling of the air may prevent a precision reading of this degree.
- *Correction for refraction and earth curvature*—A level line is a curved line at which every point is perpendicular to the direction of gravity, and the line of sight of a leveling instrument is tangent to this curve. This makes it necessary to consider this curve in some leveling operations. If reasonable care is used to make the lengths of backsights and foresights approximately equal, this aberration is self-correcting. However, in extremely long lines, it is approximately 2 inches in one-half mile, or about $^2/_3 d^2$, in feet, where d is equal to the distance, in miles. This correction is usually combined with that for refraction. The combined correction is $547d^2$, and it is *negative*.

Trigonometric Leveling

Finding the difference in elevation of two points by means of the horizontal distance between them and the vertical angle is called *trigonometric leveling*. It is used chiefly in determining the elevation of triangulation stations and in obtaining the elevation of a plane-table station from any visible triangulation point of known elevation.

In triangulation work, the vertical angles are usually measured at the same time the horizontal angles are measured to obtain the elevations of triangulation points as well as their horizontal positions. The vertical angle is measured to some definite point on the signal whose height above the center mark of the station was determined when the signal was erected. The height of the instrument above its station should be measured and recorded.

In the most exact work, the angles are measured with a special vertical circle instrument. In less precise work, an ordinary Theodolite, whose vertical arc reads by verniers to 30 seconds or to 20 seconds, may be used, but with such instruments, only single readings can be made. The best results with such an instrument are obtained by taking the average of several independent readings, one-half of which are taken with the telescope direct and the other half with the telescope inverted. In every case, the *index correction* (or reading of the vertical arc when the telescope is level) must be recorded.

The Transit

This instrument is designed and used for measuring both horizontal and vertical angles. It consists of a telescope mounted in standards that are attached to a horizontal plate (called the *limb*). Inside the limb, and concentric with it, is another plate (called the *vernier plate*). The lower plate or limb turns on a vertical spindle or axis that fits into a socket in the tripod head. By means of a clamp and a tangent screw, it may be fastened in any position and made to move slowly through a small arc.

The circumference of this plate is usually graduated in divisions of either one-half or one-third of one degree, and in the common form of transit, these divisions are numbered from one point on the limb in both directions around to the opposite point, which is 180 degrees. The graduation is generally concealed beneath the plate above it, except at the verniers. This upper plate is the vernier plate that turns on a spindle fitted into a socket in the lower plate. It is also provided with a clamp by means of which it can be held in any position and with a tangent screw by which it can be turned through a small arc. A *vernier* is a device for reading smaller divisions on the scales than could otherwise be read.

The transit is generally provided with a compass so that the bearing of any given line with the magnetic meridian may be determined if desired. It also has a spirit level attached to the telescope, so that it may be brought to a horizontal position and made to serve as a level. Figure 6-7 shows a typical transit.

Construction of the Transit

Figure 6-8 shows the general features of the transit construction. The following sections describe these features.

Figure 6-7 Typical transit.

Parallel Plates

There are two *plates*, one upper and one lower. The lower plate *A* is generally formed with two parts. The outside part is a flat ring and is screwed to the tripod head. The inside part is another flat ring of a diameter larger than the opening in the outside part and has a central dome, *C*, that is perforated on the top. The inside part is movable and rests on the under side of the outside part. The upper plate *B* is generally made in the form of a central nut, with four arms at right angles (or three at 120°). The upper plate carries an inverted conical shell, the lower portion of which passes through the perforation in the dome of the inside part of the lower plate, where it expands into a spherical shape and thus forms a ball joint with the lower plate. This spherical member is perforated in the center to allow the passage of a plumb-bob string.

Foot Screws

The two plates are connected by four-foot (sometimes only three) screws *F* in order to clamp the lower and upper plates (making them fast with each other and with the inverted shell) and to serve

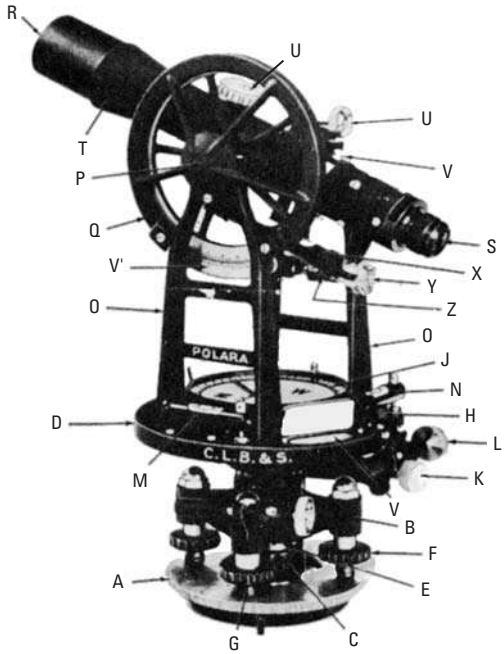

Figure 6-8 The transit. In the illustration, A represents the lower plate; B, the upper plate; C, the central dome; D, the divided limb; E, the spindle; F, the foot screws; G, foot-screw cups; H, the vernier plate; J, the compass circle; K, the clamp-screw vernier plate to divided limb; L, the tangent screw; M and N, spirit levels; O, standards or supports; P, the horizontal shaft; Q, the vertical arc; R, the objective; S, the ocular; T, the telescope; U, racks and pinions; V, the adjustable cross-hair ring; v, the divided-limb vernier; v′, the vertical-arc vernier; X, the spirit level; Y, the gradienter; and Z, the scaled index.

in leveling the instrument. The screws pass through the ends of the arms of the upper plate. They are surmounted by dust caps. There the feet fit into small cups, *G*, that rest on the top surface of the lower plate to avoid wear.

Shifting Center

Since these cups, as well as the central part of the lower plate, may be moved (after slightly loosening the foot screws), a slight motion

may be given to the instrument to better set it over a given point of the ground. This arrangement is called a *shifting center*.

Outer Spindle

A second conical shell fits and may revolve in the conical shell attached to the upper plate. It is the outer spindle, and it carries projections to form attachments with the other parts of the transit.

Divided Limb

The upper portion of the outer spindle terminates in a horizontal disc of plate *D*, the limb of which is divided into 360°, subdivided into one-half, one-third, or one-quarter of one degree. Every ten degrees are numbered, either from 0° to 360° or from 0° to 180°, either way. The degree marks are a little longer than the subdivisions, and every fifth degree has a mark slightly longer yet.

Lower Motion

The outer spindle and the divided limb are also called the *lower motion*.

Inner Spindle

A solid inverted cone fits into the outer spindle and may revolve in it. It is the *inner spindle*, and, like the outer one, it is provided with some projections for similar purposes.

Vernier Plate

The upper portion of the inner spindle projects farther than the divided limb and also carries a horizontal disc, H, that moves in a plane parallel to the divided limb (that it covers), except for two rectangular openings in opposite directions through which the divisions of the limb may be seen. These openings each carry a vernier *v* by means of which the subdivisions of degrees are again divided. Some verniers read to 1 minute, others to 0.5 minute, and some to 10 seconds. To facilitate the reading of the vernier, the openings are sometimes fitted with a reflector and a magnifying glass.

Upper Motion

The inner spindle and vernier plate *H* are also called the *upper motion*. The vernier plate carries a compass circle, shown at *J*.

Compass Circle

This consists of a circular box, the bottom of which carries at its center a sharp pivot of hard metal (hard steel or iridium) on which a magnetic needle approximately 5 inches long is balanced by an agate cup fixed in the middle of its length (see Figure 6-9). The needle

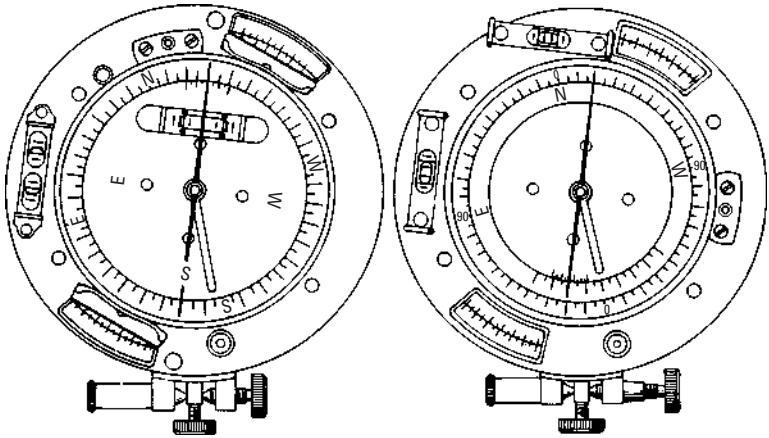

Figure 6-9 Two typical compass boxes. The compass circles are graduated to one-half degrees and numbered in quadrants. The variation plate is provided so that the magnetic declination may be set accurately. The cardinal points shift with the graduated circle. *(Courtesy C. L. Berger & Sons, Inc.)*

is strongly magnetized. Its north end is distinguished by color or ornamentation, and its balance is regulated by a small coil of fine wire wound around one arm that can be shifted. The limb that is formed by the edge of the sides of the box is divided into 360° with half degrees shown. They are numbered from two zeros marked at the ends of a diameter to 90° right and left. The bottom of the box is marked with two rectangular diameters corresponding to the graduations 0° and 90° of the vernier and two other diameters at 45° to the first.

The forward end of the diameter marked 0° is designated by the letter N, and the rear end is designated by the letter S, corresponding to north and south. The ends of the transverse diameter marked 90° are designated by the letters E on the left and W on the right, corresponding to east and west. Note that this designation is the reverse of the standard mariner's compass. Since the telescope is fixed to sight from south to north, the compass indicates the direction of the sighting. When set to an ordinary surveyor's compass, the forward end of the frame carries a vernier and a tangent screw to read fractions smaller than one-half degree.

Controlling Clamps
A screw K permits clamping of the vernier plate H to the divided limb D. Another screw attached to the upper plate permits clamping of the divided limb to the upper plate.

Tangent Screws
One *tangent* (or slow motion) *screw* L accompanies each clamp screw. It is used to complete the clamping at the exact spot where the clamp is to be made.

Spirit Levels
The *spirit levels* are attached to the vernier plate—one level M in front (north point of the box), the other N on the side—thus forming an angle of 90°.

Standards
The vernier plate carries two vertical *standards* or supports O, which are shaped like an inverted V and placed one on each side. The center of their legs is just opposite the 90° graduation of the compass box. They are made equal.

Horizontal Axis
The standards carry between and on the top of them a movable *horizontal axis* P.

Vertical Code
To the horizontal axis is attached, by means of a clamp screw, a vertical circle or arc, Q, that is divided like the horizontal circle in its vertical motion just touches a circular vernier v' carried by the left standard together with a slow-motion screw.

Telescope
In the middle of the horizontal axis and perpendicular to it is attached a *telescope* T of a description similar to that of the engineer's level, with an objective R and an ocular S, racks and pinions U for their motions, and an adjustable crosshairs ring V, with ordinary and stadia hairs.

Telescope Level
An adjustable spirit level X is also attached to the under part of the telescope, as in the engineer's level. This permits the transit to also be used as a leveling instrument, if necessary.

Motions of the Telescope
The telescope can function over the full range of the horizon and can measure any horizontal angle. In addition, since the telescope is on a

horizontal axis endowed with free motion, it may move in a vertical plane carrying with it the vertical arc, and it can, therefore, measure vertical angles. In the horizontal motion, the vertical crosshair of the telescope is brought exactly on the point sighted by means of the slow-motion screw *L*, carried by the vernier plate *H*. In the vertical motion, the horizontal crosshair of the telescope is brought exactly on the point sighted by means of the slow-motion screw carried on the inside of the left-hand support and by moving the vertical circle.

Lines of a Transit

The following are the principal lines of a transit:

- Vertical axis
- Horizontal axis
- Plate level line
- Attached level line
- Line of collimation

Vertical Axis

This vertical line passes through the center of the spindle *E* (see Figure 6-8).

Horizontal Axis

The axis *P* (see Figure 6-8) of the shaft by which the telescope rests on the supports; it must be made horizontal.

Plate Level Line

The top or bottom lines of the plate level case *N* (see Figure 6-8). These are level when the bubble is centered.

Attached Level Line

The level line of the bubble level *X* (see Figure 6-8) attached to the telescope. It is employed only when the instrument is used as an engineer's level.

Line of Collimation

The line determined by the optical center of the objective and the intersection of the crosshairs.

Relations Between the Lines of a Transit

The following relations must be obtained:

- The plate levels must be perpendicular to the vertical axis.
- The line of collimation must be perpendicular to the horizontal axis.

- The horizontal axis must be perpendicular to the vertical axis.
- The attached level line and the line of collimation must be parallel.
- The zero of the vertical circle must correspond to the zero of the vernier when the telescope is horizontal.

Adjustments of the Transit

The following sections describe the necessary adjustments of the transit.

First Adjustment

Make the axis of the spindle vertical and the planes of the plates perpendicular to it.

1. Set one level over a pair of plate screws. The other level will thus be set over the other pair.
2. Level up both levels by means of the plate screws.
3. Turn the vernier plate around by a one-half revolution. If the bubbles remain centered during the motion, the vernier plate is in adjustment. If they have moved, bring them halfway back by means of the adjusting screws and the rest of the way by means of the foot screws.
4. Repeat the operation, and determine if the bubbles remain centered when revolving the divided circle. If they do not, the plates are not parallel, and the transit must be sent to the manufacturer for repairs.

Second Adjustment

Collimate the telescope.

1. Set up the transit in the center of open and practically level ground. Carefully level the instrument.
2. Drive a stake or pin approximately 200 or 300 feet away.
3. Measure the distance.
4. Take a sight on that point, and clamp the plates.
5. Revolve the telescope vertically (in altitude) by one-half a revolution, thus reversing the line of sight.
6. Measure in the new direction the same distance as first measured, and drive a pin.
7. Unclamp and revolve the vernier plate by one-half a horizontal revolution.

8. Sight again at the first point and clamp.
9. Again, revolve the telescope vertically by one-half a revolution. If the line of sight falls on the pin, the telescope is collimated. If not, drive a new pin on the last sight at the same distance as before, and drive another pin at one-fourth the distance between the first pin and the second.
10. Move the vertical crosshair by means of the capstan-headed screw and an adjusting pin, until the intersection of the crosshairs covers the last pin set.
11. Repeat the operation to be certain of collimation.

Third Adjustment

Adjust the horizontal axis so that the line of collimation will move in a vertical plane.

1. Level up carefully and sight on a high, well-defined point (such as a corner of a chimney) and clamp.
2. Slowly move the telescope down until it sights the ground, and drive a pin there.
3. Unclamp. Revolve the vernier plate one-half of a revolution, and revolve the telescope vertically one-half of a revolution, thereby reversing the line of sight.
4. Look again at the high point and clamp.
5. Slowly move the telescope down until it sights the ground. If the intersection of the crosshairs covers the pin, the horizontal axis is in adjustment. If not, correct halfway by means of a support-adjusting screw and the rest of the way by means of the plate screws.
6. Repeat the operation, and verify the adjustment.

Fourth Adjustment

Make the line of collimation horizontal when the bubble of the attached level is centered.

1. Drive two stakes 300 to 400 feet apart, and set up the instrument approximately halfway between these stakes.
2. Level up and take readings on the rod held successively on the two stakes. The difference between the readings is the difference of elevation of the stakes.
3. Next, set the transit over one of the stakes, level up, and take a reading of the rod held on the other stake.

4. Measure the height of the instrument. The difference between this and the last rod reading should equal the difference of elevation as previously determined. If it does not, correct the error halfway by means of the attached level-adjusting screw.
5. Repeat the operation, and verify the adjustment.

Fifth Adjustment

Make the vernier of the vertical circle read zero when the bubble of the attached level is centered.

1. Level up the instrument.
2. Sight on a well-defined point, and take note of the reading on the vertical circle.
3. Turn the vernier plate one-half of a revolution, and also turn the telescope vertically one-half of a revolution.
4. Again, sight on the same point.
5. Read and record the reading on the vertical circle.

One-half the difference of the two readings is the index error. The error may be corrected by moving either the vernier or the vertical circle, or the error may be noted and applied as a correction to all measurements of vertical angles.

Adjustments of the Compass

It will sometimes be necessary to adjust the compass. When an adjustment is required, it may be accomplished by using the following procedure.

First Adjustment

Straighten the needle.

1. Examine to see if the ends of the needle are set on opposite divisions. If not, fix the pivot so that they are.
2. Revolve the box by one-half of a revolution. If the needle does not set on opposite divisions, bend both ends by one-half the difference.

Second Adjustment

Place the pivot in the center of the plate. If the needle is straight, move the pivot until the needle sets on opposite divisions at points such as 0°, 45°, and 90°.

Instructions for Using the Transit

The transit requires various adjustments, as explained in the preceding section. To center the transit over a stake, rest one leg of the tripod on the ground, then grasp the other legs and place the instrument as nearly over the stake as possible. Then attach the plumb bob, and center it accurately by means of the shifting head. Avoid having the plates too much out of level because this will result in unnecessary straining of the leveling screws and plates.

Once the instrument has been centered over the stake, level it up by the spirit levels on the horizontal plate. To do this, turn the instrument on its vertical axis until the bubble tubes are parallel to a pair of diagonally opposite plate screws. Then, stand facing the instrument and grasp the screws between the thumb and forefinger; turn the thumbscrew in the direction the bubble must move. When adjusting the screws, turn both thumbscrews in or out, never in the same direction. Adjusting one level will disturb the other, but each must be adjusted alternately until both bubbles remain constant.

Figure 6-10 shows the method of measuring a horizontal angle. The process of laying-off a given angle is similar to that of measuring

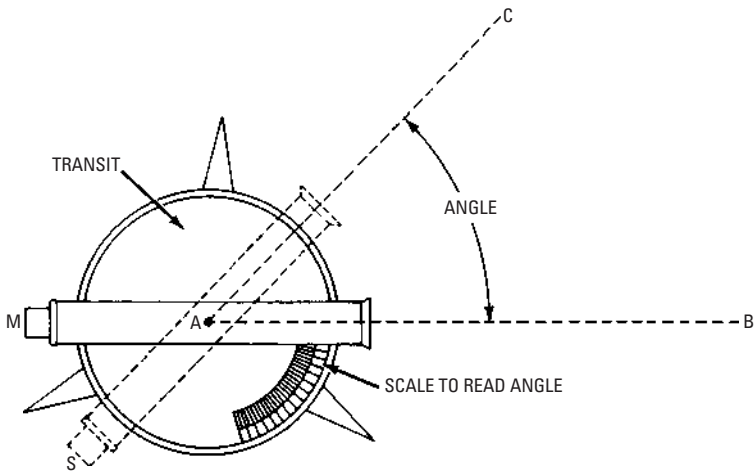

Figure 6-10 The general principle of transit work. The transit is placed over the apex A of the angle CAB that is to be measured. The telescope is sighted to stake B (position M), and a reading is taken. It is then turned horizontally and sighted to stake C (position S), and another reading is taken. The difference between these readings gives angle CAB.

the angle. The transit is set up at the vertex of the angle, the vernier is clamped at zero, and the telescope is pointed at the target, thereby marking the direction of the fixed line. The limb is now clamped, the vernier is unclamped, and the vernier plate is turned through the desired angle and clamped. A stake should now be driven in line with the vertical crosshair in the telescope, thus establishing the two sides of the angle.

When laying out the foundations of buildings, a corner stake is first located by measurement. Then the direction of one of the walls is laid out by driving a second stake. This direction may be determined by local conditions (such as the shape of the lot or the relation to other buildings). If the building is to be an extension to (or in line with) another building, the direction can be obtained by sighting along the building wall and driving two stakes in line with it. If it is to make a given angle with another building, this angle can be laid off as shown in Figure 6-11.

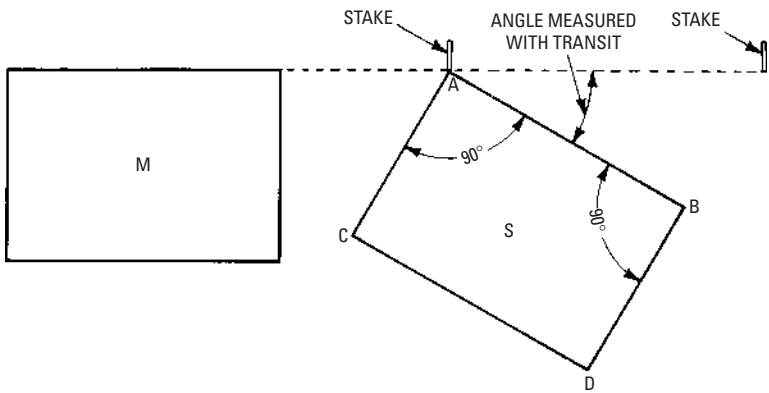

Figure 6-11 The method of laying out a new building (S) at a given angle with an old building (M). After the corner and the direction of one wall are determined, a right angle may be laid off (if the building is rectangular), thus locating two of the sides (AB and AC). The length of side AB is then measured, thereby locating corner B. The transit is set up at B, and line BD is laid off at right angles to AB. AC and BD are then laid off by the proper length, and the four corners of the building are thus located. If the building had not been rectangular, the proper angles could have been laid off instead of right angles.

Gradienter

Some transits carry a device called a *gradienter* (*Y* in Figure 6-8) that is attached to the horizontal axis by means of a clamp screw and is inside of the right-hand support. It is designed and employed for the determination of grades and distances and consists of an arm in the shape of an inverted Y with curved branches. To the extremities of this arm are attached an encased spiral spring and a nut through which moves a micrometer screw with a graduated head that revolves in front of a scaled index (*Z* in Figure 6-8) that is also carried by the arm. The ends of the screw and the spring are on opposite sides of a shoulder that is carried by the right-hand support. The head is divided into tenths and hundredths, and every revolution moves it in front of the scale by one division so that the scale gives the number of turns of the screw, and the graduated head gives the fraction of a turn.

In grading, if one revolution of the screw moves the crosshair a space of 1 foot on a rod held 100 feet away, the scope indicated by the telescope is 1 percent. To establish a grade, level up the telescope, clamp the arm of the gradienter, and turn the micrometer screw by as many divisions as are required in the grade. For example, to set the gradienter at 2.35, move the head two complete turns plus 35 subdivisions. Measure the height of the telescope from the ground. Set the rod at that height. Then hold the rod at any point on the line, raising it until the target is bisected by the crosshairs. The foot of the rod will then be on the grade.

Care of Instruments

With proper care, the usefulness of an instrument can be preserved for many years. Therefore, the following suggestions on the care of instruments should be noted:

- The lenses of the telescope, particularly the object glass, should not be removed, since this will disturb the adjustment. If it is necessary to clean them, great care should be taken, and only soft, clean linen should be used.
- To retain the sensitivity of the compass needle, the delicate point on which it swings must be carefully guarded, and the instrument should not be carried without the needle being locked. When the needle is lowered, it should be brought gently on the center pin.
- The object slide seldom needs to be removed. When removal is necessary, the slide should be carefully protected from dust.

Do not grease or oil the slide too freely; only a thin lubricant film is necessary. Any surplus of oil should be removed with a clean wiper.

- The centers, subject to considerable wear, require frequent lubrication. After a thorough cleaning, they should be carefully oiled with fine watch oil. All of the adjusting screws should be brought to a fine bearing, but they should never be tightened to such a degree that a strain is applied to the different parts. If this is done, the adjustment will be unreliable.
- When the instrument is carried on the tripod, all clamps should be tightened to prevent unnecessary wear on the centers.

The Stadia

This is a device that is used for measuring distances, and it consists essentially of two extra parallel hairs in addition to the ordinary crosshairs of the transit or a level telescope (see Figure 6-12). The stadia hairs may be adjustable, or they may be fixed permanently on the diaphragm.

When using the stadia, distances are measured by observing through the telescope of a transit the space (on a graduated rod) included between two horizontal hairs (called *stadia hairs*). If the rod is held at different distances from the instrument, different intervals on the rod are included between the stadia hairs. The spaces

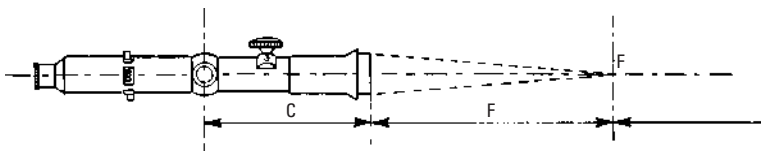

Figure 6-12 The principle of stadia operation. The fixed stadia hairs are set so that they will intercept 1 foot on a rod at a distance of 100 feet. Since the image of the crosshairs is projected to a point beyond the telescope objective equal to its focal length, the rays of light converge at that point, and measurements must begin from there. Therefore, a constant must be added to all stadia readings equal to the focal length of the object lens, plus the distance from the face of the objective to the center of the instrument. This constant is the factor F + C. For transit telescopes, it is equal to approximately 1 foot.

on the rod are proportional to the distances from the instrument to the rod so that the intercepted space is a measure of the distance to the rod. This method of measurement furnishes a rapid means of measuring distances when filling in details of topographic and hydrographic surveys.

Most transits, all plane-table alidades, and some precision leveling instruments are fitted with stadia hairs. Stadia surveying has the advantage in that the intervening country does not have to be taped, and it provides a means of measuring inaccessible distances (such as across water and up steep hills and bluffs). It is well adapted to preliminary surveys for highways and railroads because the errors tend to be compensating rather than cumulative, but it should not be used for short distances (such as farms and city lots). In sights of 200 to 400 feet, it is possible to read a rod to the nearest hundredth of a foot that represents 1 foot in distance. At 600 to 1200 feet, it is possible to read to the nearest hundredth of a yard that represents 3 feet in distance. This is the precision to be expected in stadia measurements.

The rod used is preferably a one-piece stadia rod (see Figure 6-13F), but any standard leveling rod (except builders rods graduated in inches and sixteenths) may be used. When leveling with an instrument that is equipped with stadia hairs, care should be taken not to confuse the center leveling crosshair with either of the two outside stadia hairs. It has been done. Although it is by no means obsolete, stadia surveying has been superseded by aerial photography. Neither is a substitute for careful taping.

Other Devices

For the vast majority of surveying purposes, the transit and levels described are more than sufficient. Three other devices can be even more useful, however.

The *Theodolite* (see Figure 6-14) is a transit that is more expensive but more accurate than a standard transit, and the Theodolite has more capacity. If you are doing readings over 500 feet, you may want to look into this.

The *automatic level* (see Figure 6-15) is also a tool used where great precision is required. Another device, the electronic distance measurer (EDM), is good when great distances must be read. It is a tool used on the sea and in other places (such as measuring from a point to a mountain) where there is no real point of reference. The electronic distance measurer and the Theodolite are often used together (see Figure 6-16). (By the way, no one seems to know the origin of the term Theodolite.)

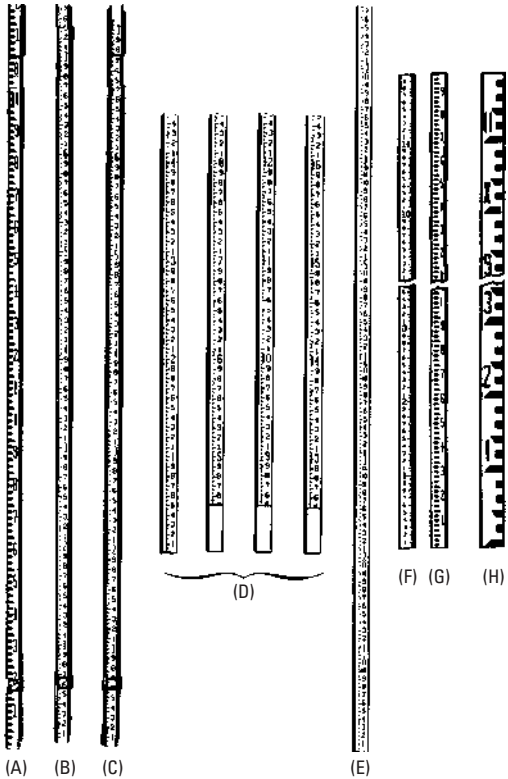

Figure 6-13 Various popular leveling rods: (A) is the Philadelphia rod, in English graduations; (B) the California rod; (C) the Philadelphia rod, in metric graduations; (D) the Chicago rod; (E) the architect's rod; (F) the stadia rod, in English graduations; (G) the stadia rod, in metric graduations; and (H) the broad stadia rod. *(Courtesy Eugene Dietzgen Co.)*

Laser Levels

The need for plumb walls and leveling moldings (as well as various other points straight and level) is paramount in house building. It is difficult in some locations to establish a reference point to check for level windows, doors, and roofs, as well as ceilings and steps.

The laser level (see Figure 6-17) has eliminated much of this trouble in house building. This simple, easy-to-use tool is accurate to within $^1/_8$ of an inch in 150 feet, and it has become less expensive recently so that even the do-it-yourselfers can rent or buy one.

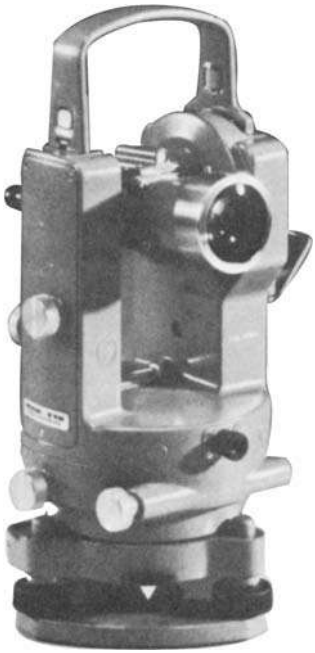

Figure 6-14 Wild T-16 Theodolite. *(Courtesy of Wild-Heerbrugg)*

Figure 6-15 NA2 automatic level. *(Courtesy of Wild-Heerbrugg)*

Figure 6-16 Typical EDM/Theodolite setup. *(Courtesy of Wild Heerbrugg)*

The laser level can generate a vertical reference plane for positioning a wall partition or for setting up forms (see Figure 6-18). It can produce accurate height gaging and alignment of ceilings, moldings, horizontal planes, and can accurately locate doorways, windows, and thresholds for precision framing and finishing (see Figure 6-19). The laser level can aid in leveling floors, both indoors and out. It can be used to check stairs, slopes, and drains. The laser beam is easy to use and accurate in locating markings for roof pitches, and it works well in hard-to-reach situations (see Figure 6-20). The laser beam is generated by two AAA alkaline batteries that will operate for up to 16 hours.

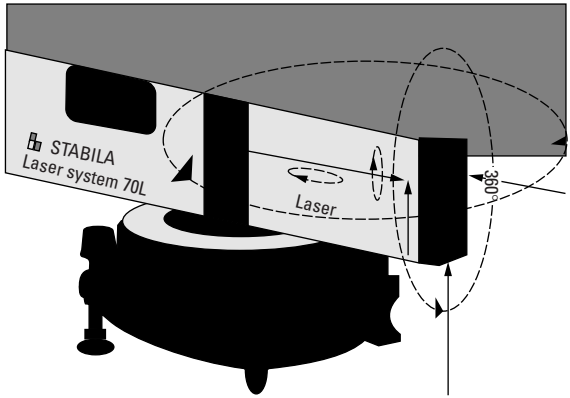

Figure 6-17 Laser spirit level moves 360° horizontally and 360° vertically with the optional lens attachment. It sets up quickly and simply with only two knobs to adjust. *(Courtesy of Stabila)*

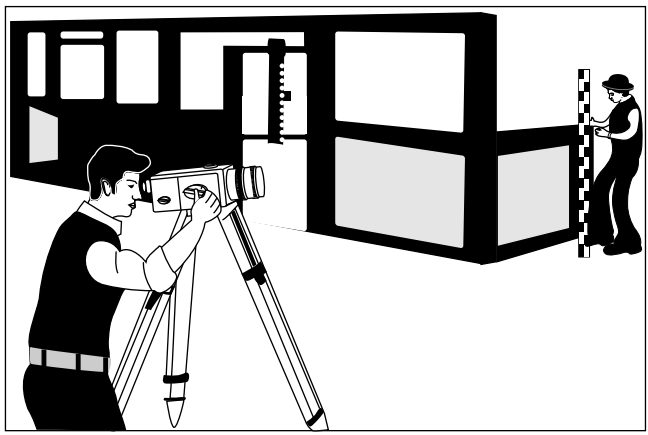

Figure 6-18 The laser level can be used to align ceilings, moldings, and horizontal planes. It produces accurate locations for doorways, windows, and thresholds for precision framing and finishing. *(Courtesy of Stabila)*

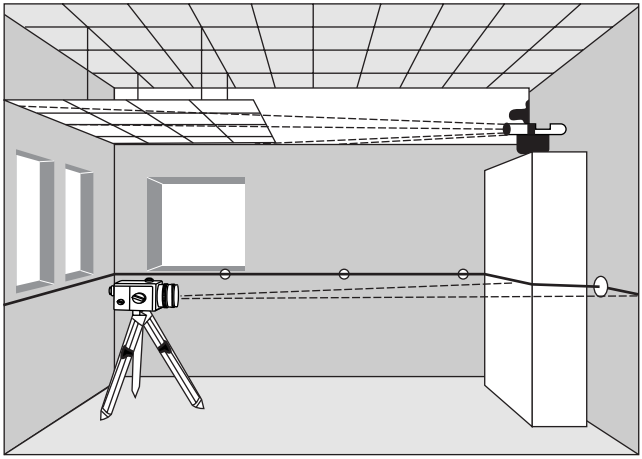

Figure 6-19 The laser level can be used for indoor or outdoor leveling of floors, stairs, slopes, drains, and ceilings, and moldings around the room. *(Courtesy of Stabila)*

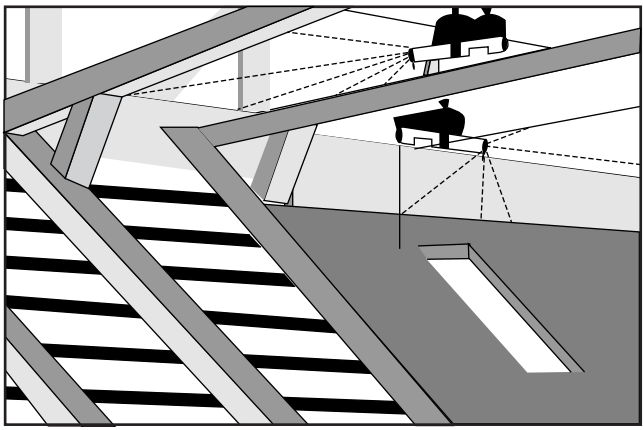

Figure 6-20 The laser beam is used to provide easy and accurate location markings on pitches and in hard-to-reach situations. *(Courtesy of Stabila)*

The combination laser and spirit level quickly and accurately lays out squares and measures plumb. No protective eyewear is needed. The laser operates on a wavelength of 635 nm and can have an extended range up to 250 feet.

Summary

Leveling, in surveying, is the operation of determining the comparative levels of different points of land for the purpose of laying out a grade or building site by sighting through a leveling instrument at one point to a leveling staff at another point.

The transit is designed and used for measuring both horizontal and vertical angles. It is a telescope mounted in standards attached to a horizontal plate, called the limb. The transit is generally provided with a compass so that the bearing of any given line with the magnetic meridian may be determined. It also has a spirit level attached to the telescope so that it may be brought to a horizontal position and made to serve as a level.

A stadia is a device used for measuring distance. It consists essentially of two extra parallel hairs in addition to the ordinary crosshairs that are used in transit or level telescopes. When using the stadia, distances are measured by observing through the telescope of a transit.

The laser level is rather inexpensive and easily used. It can be utilized in the squaring of steps, locating doors, windows, and many other squaring and leveling operations.

Review Questions

1. What are the three lines of the level?
2. What is a transit?
3. What are the fine lines of a transit?
4. Name the various leveling rods used for measuring distance.
5. Explain the terms *backsight* and *foresight* when used in leveling.
6. What is a stadia?
7. Where is the laser level most useful?
8. What is leveling in terms of surveying?
9. Why is it important to have square corners and level floors in a building?
10. How can a laser actually establish a useable level reference?

Chapter 7

The Design Process

The first consultation the designer has with the owner results in the owner setting forth what he or she wants or requires. This may include such things as the number of rooms, their sizes, the style of house preferred, the kind of materials he or she wants to use, and probably that all-important factor—how much money he or she will need, or how much is available. The first consultations may result in the designer making a dozen or more sketches.

It is the designer's function to guide the client in such matters. Usually, the owner has no skill or aptitude for building a house. Later consultations may settle such things as lighting, heating, perhaps the quality and style of plumbing fixtures, and the make or quality of such appurtenances as air-conditioning, laundry, and dishwashing equipment. The designer usually makes sketches (probably freehand) for the owner's approval.

In some instances, the purchaser comes into the office and consults with the designer or designee. The owner will decide on small changes in the chosen house plan. The designer just makes the suggested or agreed-upon changes and then draws up the blueprints. The owner picks out wall colors, kitchen cabinets, appliances, toilet-bowl designs, and many other small details. This may take a number of visits to a builder's office, where all the samples of tile, carpet, brick, and other design features are on display. If there is a model house available, it is much easier for most people to pick out the one they want and see what changes they desire.

Design Considerations

Following are some of the particulars that should be settled before final design work:

- Owners may greatly prefer a house with a basement. However, perhaps for reasons of economy, they may be willing to accept a house on a concrete slab. On the other hand, if the house is in Texas or some other location with soil problems, it will have to be a slab house. The designer should be able to give advice as to the advantages and disadvantages of the two systems along with the comparative cost of construction.
- The owner's mind may be set on a heating plant in the basement. However, in the event that a house with no basement is agreed on, the owner should understand that the heating plant must be in a utility room or in the attic.

- The owner may insist that there be a well-equipped laundry in the basement. The person who does the laundry may have some ideas on this subject.
- The owner will probably insist that there be adequate closets for each bedroom. The owner may have some ideas as to exactly what "adequate" consists of.
- The owner will want the kitchen to be convenient, with or without a garbage-disposal unit, probably with a dishwasher in a convenient location, possibly with room for a home freezer. Although saving steps is important, sufficient room in the kitchen may be more important. Few cooks appreciate a small kitchen—it greatly inhibits style.
- The owner may want a dining room, and not just dining space in one end of the living room, and may think that a counter or bar and no partition between the kitchen and dining room is not sufficiently odor-resisting when cooking fish or corned beef and cabbage.
- The owner (or perhaps the designer) may have some ideas concerning privacy in the home. There may be some objections to so-called window walls that are supposed to "bring the outdoors indoors," and a preference may be expressed for baseball-proof walls instead. Perhaps the idea of keeping large areas of glass clean is not appealing, and the owner would be quite content to leave the "outdoors" outdoors, if there is plenty of living space indoors.
- Then, there is the constant problem of sound resistance. Modern homes are often noisy, with air-conditioning, forced air heating, laundry, and dishwashing equipment, attic fans, kitchen fans, bathroom fans, radio and television, and many other noise-generating sources. It is the designer's duty and obligation to see that such noises are isolated insofar as is possible. Partitions should be noise-resistant.

The designer must understand these and many other problems. They are with us and will be with us for many years to come. The owner probably does not (possibly cannot) understand how to handle such problems. The designer can, and should. It is the designer's function to guide the owner's ideas or simple notions so that the home environment will be satisfactory as far as the owner's means will permit. The designer will probably be blamed for any serious discrepancy, no matter if the owner *did* insist on it.

It is not only desirable, but also necessary that the designer be familiar with the dimensions of the equipment, furniture, and other appurtenances found in the home. While it is rarely necessary that they be accurately detailed, space must be allotted for each one of them, and space is expensive.

All modern building is governed by codes of some sort. In all government-financed homes, the government's minimum standards are strictly enforced, and city codes are often much more restrictive. Electrical codes occasionally seem to be unreasonable, but the building designer must be governed by them. Plumbing is often seriously skimped when no one is watching. Many states have plumbing codes. However, they do not have the force of law unless augmented and enforced by local authority. There is a *National Plumbing Code*, and, although it is advisory, it is in line with good practice. It should be used where there are no local codes. State and local boards of health may make it quite difficult for the designer of an inadequate plumbing system if the occasion arises.

For economy, kitchens and bathrooms should be placed back-to-back. A 3-inch copper soil pipe will fit into a partition of 2 × 4 studs, whereas a 4-inch cast-iron soil pipe won't.

In perhaps most cases, the designer's duty is done when he or she prepares and delivers the drawings for a job. In some cases, he or she contracts to inspect the work at stated intervals, to ensure that the work is satisfactorily done.

The designer should allow for the following thicknesses of walls in drawings:

- *Standard wood outside wall*—$^3/_4$-inch plywood or insulating board sheathing, $3^1/_2$-inch studs, $^1/_2$-inch sheetrock inside, $4^1/_2$ inches under the siding
- *Inside partitions*—$3^1/_2$-inch studs, sheetrock $^1/_2$ inch (both sides), $4^1/_2$ inches
- *Sound-resistant staggered-stud partitions*—$3^1/_2$-inch studs staggered 2 inches, gypsum lath and plaster (both sides) $7^1/_4$ inches
- *Single-width brick veneer*—$^3/_4$-inch sheathing, $3^1/_2$-inch studs, $^1/_2$-inch sheetrock, $9^1/_2$ inches
- *Concrete blocks (plastered against the masonry)*—8-inch, $8^1/_2$ inches
- *Concrete blocks (with furring)*—8-inch, $^3/_4$-inch furring, $^1/_2$-inch sheetrock, $9^1/_4$ inches

- *Cavity masonry wall*—$3^3/_4$-inch brick, $2^1/_2$-inch air space, 4-inch concrete blocks, plaster on masonry, $10^7/_8$ inches
- *Chimneys*—minimum wall thickness, $3^3/_4$-inch brick liners, outside dimensions, $8^1/_2$ inches × $8^1/_2$ inches, $8^1/_2$ inches × 13 inches, $8^1/_2$ inches × 18 inches, 13 inches × 13 inches
- *Ceiling heights*—first floor, clear minimum, 7 feet 6 inches; basement, 6 feet 9 inches clear
- *Stair wells*—3 feet 2 inches × 9 feet, or clear headroom 7 feet above nosing of treads, vertically

An Example of Design

As an example of design, the series of figures presented in this section show the development of architectural drawings beginning with stock plans or plans that appear in newspapers and magazines from time to time. They give a prospective owner a good starting point, saving much time and study.

Certain things, however, should be kept in mind. The magazine drawings were doubtlessly prepared by a registered architect. While the publishing of a design may imply that the architect has given approval that the design can be copied, it is best to get the written consent of the original designer, or at least to secure this consent from the publisher. A registered architect's plans are protected from being copied, and court decisions have ruled that minor changes (regardless of how many of them there are) do not release the copier from liability. Purchasing a set of stock plans includes permission to build from the plans at least once. You might have to buy a second set to build the same design again. Determine where you stand legally before copying designs of any sort. It may save you from a costly and embarrassing situation.

Let us assume that we start with magazine plans (Figures 7-1 and 7-2). These two plans were prepared by a capable and experienced architect who has given us a practical and logical arrangement of a house plan, with stairs, doors, windows, closets, and so on. The architect has also indicated a proper and reasonable size for the various rooms. Any proposed changes or additions that the owner may desire may be taken up with the architect or builder, and may be easily whipped into form on the drawing board.

Conference with Builder

Let us now assume that the owner prefers to deal only with a carpenter or builder of his or her acquaintance, and asks the carpenter

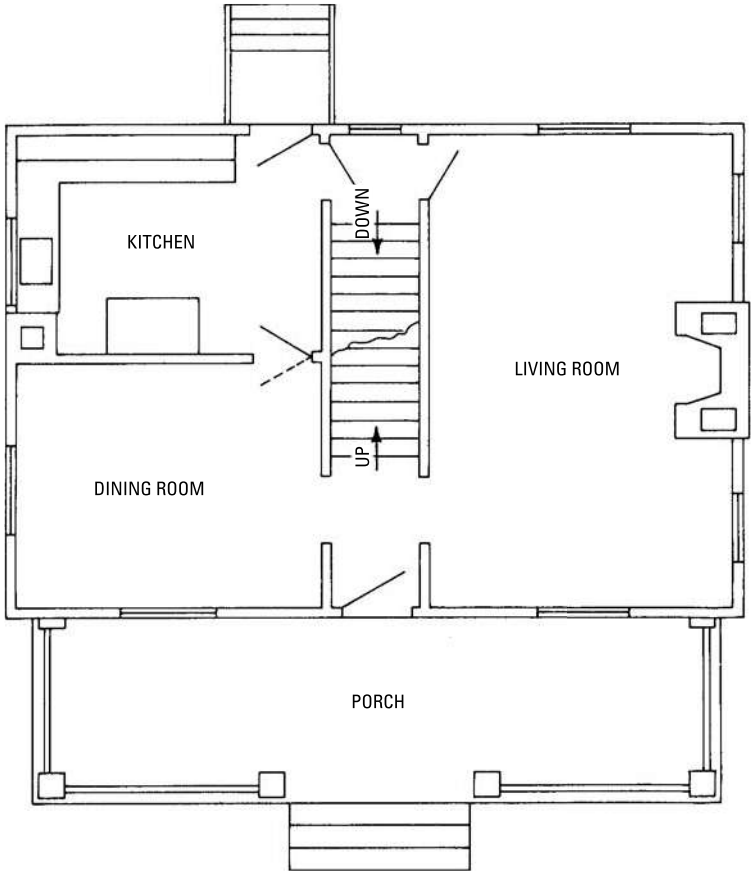

Figure 7-1 First-floor plan from a magazine.

or builder to prepare the drawings and give an estimate on construction. We will also assume that while the builder understands his or her own business from start to finish, he or she does not pretend to be a designer. A designer is, therefore, called in for advice on the preparation of the drawings. Having decided on the main features of the plans and elevations, we will start with the first-floor plan of the house, since that is always the controlling form factor. Then, following the easy and usual method, we may trace the other plans (second floor and basement) over it.

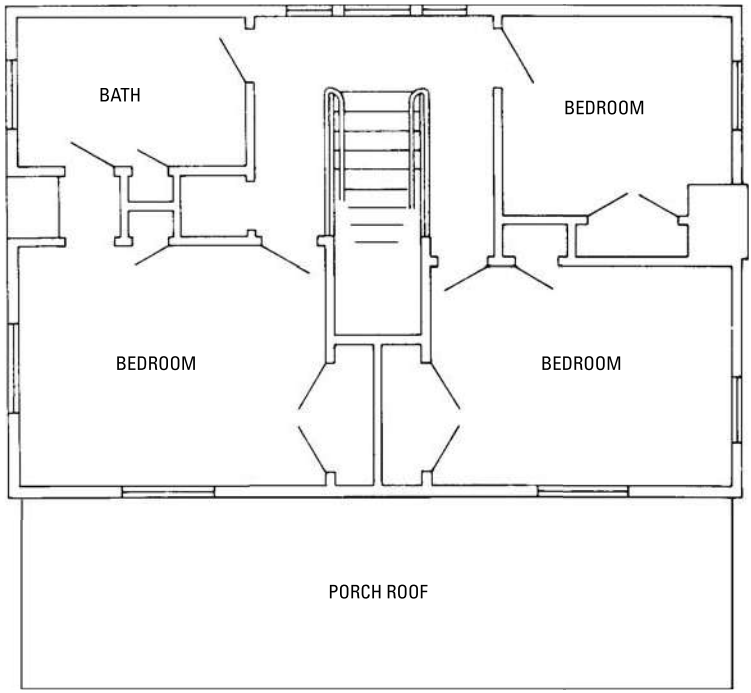

Figure 7-2 Second-floor plan from a magazine.

Changes Agreed On

After a careful study of the first-floor magazine plan, it is decided first to place the large chimney and fireplace inside the house against the stairs to conserve as much heat as possible, since considerable heat is necessarily lost from an outside wall. Next, omit the kitchen chimney, since a gas range is to be used. Third, move the large porch to the end of the house, since it provides more privacy, and change the window to a door. Fourth, provide the paved entrance platform with a settee at the front entrance. Fifth, provide a coat closet at the front entrance. Sixth, enlarge the dining room, taking the space from the kitchen. Seventh, install a first-floor lavatory in the small new wing at the rear. These, along with other minor changes, such as turning the back steps, adding two small closets near the fireplace, and substituting round for square porch columns, create a much more convenient and valuable first-floor plan. The principal second-floor changes are made to save cost and involve a

lowering of the front eaves line and decreasing the size of the front rooms.

Figure 7-3 shows the modified first-floor plan after the changes have been agreed on, and Figure 7-4 illustrates the completed drawing of the first floor, with all the specifications and dimensions given in full detail.

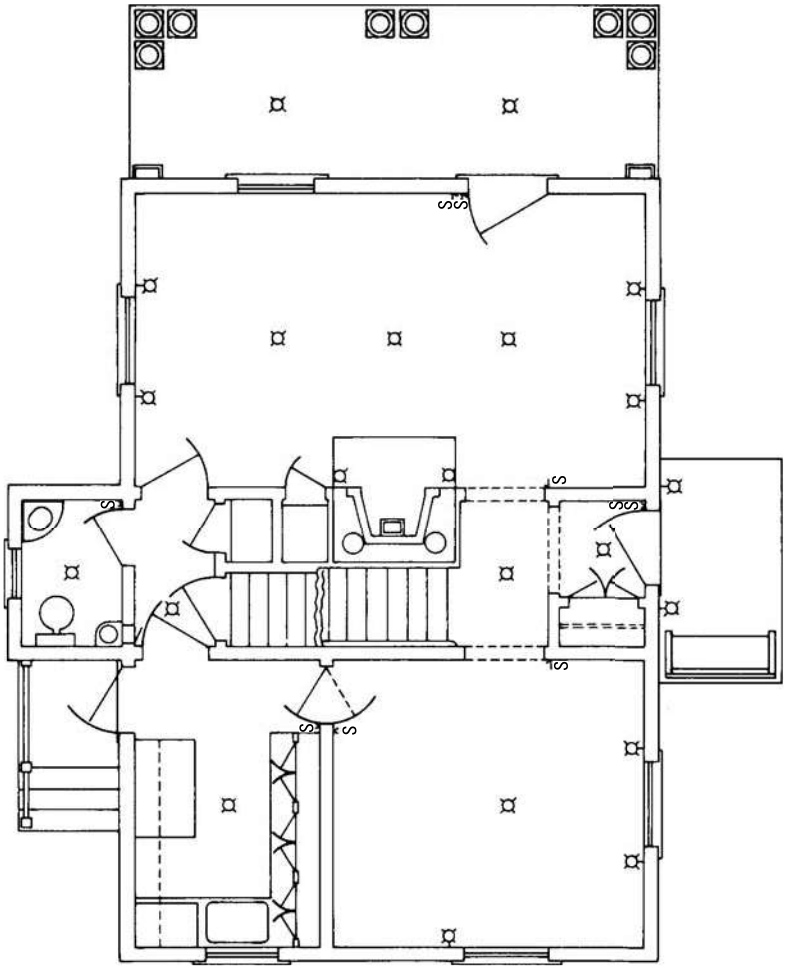

Figure 7-3 The modified first-floor plan.

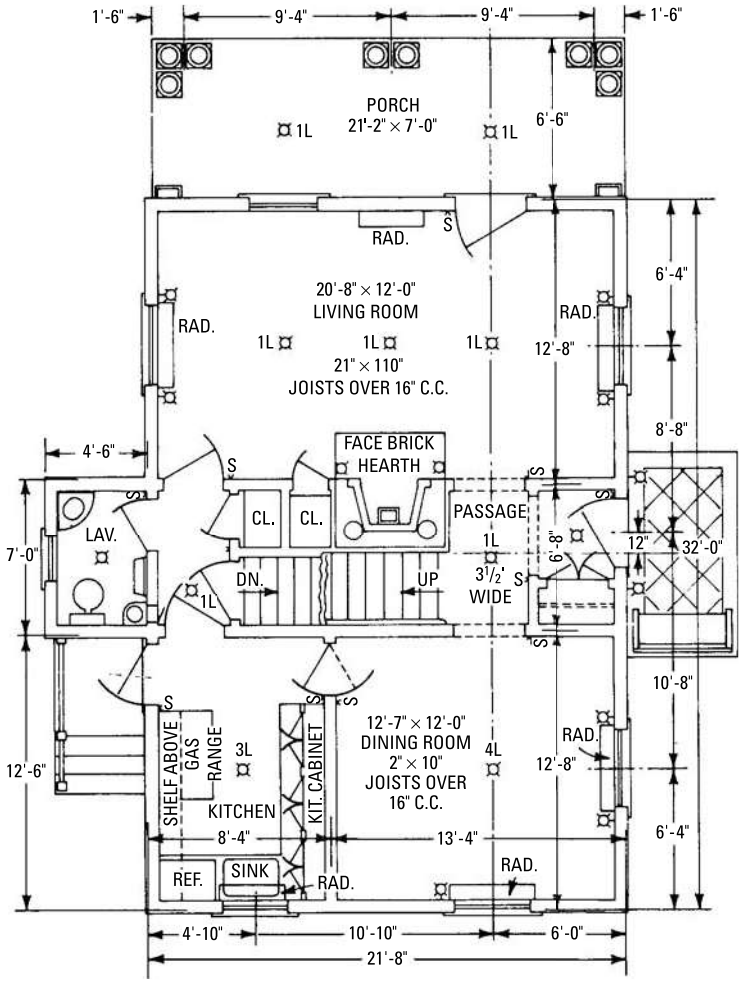

Figure 7-4 The completed drawing of the first floor with all dimensions and specifications in detail.

Summary

In architectural design, the designer usually makes several sketches (probably freehand) for the owner's approval. Among the many particulars that should be settled before the designer actually begins work are: will there be a basement (which determines the location of

the heating plant); how many closets in each bedroom; will there be kitchen conveniences (such as a garbage-disposer unit, dishwasher, and room for a home freezer).

These are just a few of the problems for the designer. The owner probably does not know how to understand and handle such problems. The designer should know how to handle these problems and guide the owner so that the home environment will be more than satisfactory when the work is completed.

It is, of course, necessary that a designer know the dimensions of the equipment, furniture, and other appurtenances found in a home. Electrical codes occasionally seem to be unreasonable, but the building designer must follow them. Plumbing is often seriously skimped, unless local authorities have some kind of a ruling enforcing certain regulations.

Many times the future owner prefers to deal only with a carpenter or builder, and asks him or her to prepare the drawings and give an estimate on construction. The builder then calls the designer in for advice on the preparation of the drawing.

Review Questions

1. What must a designer know from the client before drawing up plans for a house?
2. Is it generally a good idea to change floor plans after construction has started?
3. Is the kind of material used in the construction important to the designer?
4. Are wall thickness, plumbing, and electrical wiring considered when drawing up plans?
5. Why are some localities limited to slab houses?
6. What is the difference in price of a slab foundation and a basement?
7. Where is the air-conditioning and heating placed if the house is built on a slab?
8. What are some of the concerns of someone who is trying to have a house built?
9. What is the purpose of a plumbing code?
10. What is the purpose of an electrical code?
11. Why can't you use a copy of a blueprint found in a magazine to build a house?

Chapter 8

Building Specifications

By definition, a *specification* is a specific and complete statement detailing the nature and construction of the item to which it relates. As applied to the building trades, specifications describe briefly (yet exactly) each item in a list of features and materials required to complete a contract for building an entire project.

Great care should be used when reading specifications in order to avoid misunderstandings and disputes. Each item entering into the construction is defined and described with precision so that there can be no chance of misunderstanding or double interpretation.

Example of Specifications

Specifications refer to the contract form of which they are a part. This saves repetition of statements with regard to liability of contractor, owner, and so on. The following is an example of specifications.

Introduction

Specifications for a frame dwelling to be built for Mr./Ms. _____ of _____ in the county of _____ and state of _____ on lot number _____ on the _____ side of _____ street in the city or borough of _____, county of, _____, state of _____. These specifications relate to a set of attached drawings consisting of:

1. Plan of foundation and basement.
2. Plan of first floor.
3. Plan of second floor.
4. Plan of roof.
5. Four elevations (eight sheets in all).

Detailed working drawings furnished as required. All drawings are a part herewith, and are to be considered as such with any contract that may be made.

Height of Ceilings

The following dimensions for these heights are:

- Basement—7 feet 2 inches clear of finish.
- First-story—8 feet 6 inches between timbers.
- Second-story—7 feet 6 inches between timbers.

Interpretation of Drawings

For arrangement of floors, general finish, and measurements, reference must be made to the drawings. However, should any difference appear between the scale measurements and the figures, or between the wording of the specifications and the lettering on the drawings, the specification shall in all cases take precedence. If any errors that are not explained either by reference to the drawings or specifications become apparent, the contractor shall refer them to the architect for correction before proceeding with the work.

Conditions

The contractor must see that all the work on the said building is performed in a thorough, professional, and substantial manner by competent workers and must also furnish all materials (the best of their respective kinds), labor, implements, transportation, and so on, if not otherwise specified.

All painted parts of the exterior must have a prime coat of paint as fast as it is ready. The succeeding coats must not be applied within 3 days of the former, and then not in wet or freezing weather or other conditions specified by the paint manufacturer.

The contractor must protect all work while the building is in his or her hands, remove all superfluous materials or rubbish, and not obstruct the grounds around the foundation for grading and filling in as soon as the building is up. Figures are to take precedence over scale measurements.

Mason's Work

Excavate to the length, breadth, and depth required for the foundations, as shown on the architect's drawings. The topsoil is to be removed and placed in a separate pile from the other excavated materials—25 feet away from the excavation where directed. Also, excavate for a septic tank and overflow 75 feet from the foundation, as will be directed, containing 28 cubic yards to be built of concrete, with baffles, drains, and so on, all of which is set out in a special plan for same. The septic tank may be included in the plumbing subcontract. The main tank must be waterproof, although the overflow need not be. The tank shall be connected to the house at a point below the lowest fixture and below the frost line with a uniform declination of not less than 12 to 20 inches, and will have no running U trap. The drain is to be made of 6-inch socket-jointed Transite tile placed in tight cement joints from a point 4 feet outside the foundation.

Foundation

Foundations and footings, as shown on the plans, are to be made of poured concrete with 8-inch walls and 16-inch footings.

Chimneys

Build two chimneys, as shown, of the same size and shape. Use an approved hard-red-pressed brick for all exposed parts of the outside chimney and for topping out. The fireplaces in the parlor, dining room, and bedroom are to be faced with the same brick (smooth inside the fireplaces), with 8 inches on the sides and 24 inches at the top. Buff the joints, and straighten the arches on $3^1/_2$-inch × $3^1/_2$-inch angle irons, unexposed. Use firebrick for the backs. Lay the bricks in an approved refractory mortar (no fire clay). Spring trimmer arches for the hearths are to be laid with the same selected brick. All flues are to have tile linings, approved chimney pots, and clean outs.

Mortar

All mortar for brickwork is to be grade specified by building code regulations.

Installing Drywall

All walls, partitions, and ceilings, and all studded and furred places in all stories, are to be covered with $^1/_2$-inch-thick Sheetrock, which will have joints covered with tape and joint compound. Two coats of compound will be applied where wallpaper is used; three coats where paint is to be used. Panels shall be 4 feet × 8 feet and will be installed horizontally on walls, across the framing members on ceilings. The panels are to be secured with $1^3/_8$-inch blued ring-shank nails, and the pieces staggered. Inside corners of the paneling shall be covered with inside-corner molding, with nails placed 7 inches apart. Outside corners shall also be covered with metal molding. All joints will be smooth to the touch.

Tiling

The floors of bathrooms will be tiled with 3-inch × 3-inch octagonal and 1-inch-square vitrified tiling, colors to be selected. The side walls will be tiled 4 feet high of plain white glazed $2^1/_2$-inch × 4-inch molded base and nosing, with a narrow tinted stripe at the top of the sanitary base and under the nosing. The floors will be properly prepared by the carpenter by setting the rough floor $^1/_2$ inch below the top of the floor beams. All tiling will be set in adhesive recommended by the manufacturer, and the floors will be finished flush with the wood-finish floors.

Other Floors

There will be a concrete floor in the furnace room, in the shop, and in the area from the west end turning east to the cross wall, as shown in the plans. All floors will be $3^5/_8$ inches minimum, with 6-inch × 6-inch No. 11 reinforcing mesh. The kitchen hearth will be built in the same manner.

Coping

There will be 4-inch caps of blue stone on all piers showing them, edged on four sides, 3 inches larger than the piers. Cope area walls, which are to be 8 inches, with 2-inch × 10-inch blue stone where circular, fitted to radius. *No* patching of stone will be permitted.

Timber

All timber will be thoroughly seasoned, No. 1 common pine, square, straight, and free from any imperfection that will impair its durability or strength. *No* individual piece is to have moisture content of more than 19 percent. The architect will check this.

Framing

The framing will be as indicated on the drawings. Headers over openings will be the sizes indicated in detail. *No* header with checked moisture content of more than 15 percent will be acceptable. Frame so that sheathing will be flush with the foundation wall. All moldings are to be miter-spliced and mitered at angles. *No* butt ends will be showing in the finish.

Timber sizes will be as follows:

Design Element	***Timber Size***
Sills	2-inches × 8-inches
Girders	10-inch standard I-beams (25.4 pounds)
Corners	4 inches × 6 inches backed with 2 inches × 4 inches or built up
Main plate	4 inches × 4 inches (2 inches × 4 inches doubled)
Rafter plate	4 inches × 4 inches (2 inches × 4 inches doubled)
Studding (general)	2 inches × 4 inches
Closet studding	2 inches × 3 inches
Main rafters	2 inches × 6 inches
Dormer rafters	2 inches × 4 inches
Ridge boards	$1^1/_4$ inches × 8 inches
First floor joists	2 inches × 10 inches
Second floor joists	2 inches × 8 inches
Second story ceiling beams	2 inches × 6 inches

Spacing and Bridging

All studding, floor, and ceiling joists placed on 16-inch centers. In every span of flooring exceeding 10 feet, there will be a row of 1-inch × 2-inch bridging or 2-inch × 3-inch double nailed at each end. Rafters will be placed on 24-inch centers.

Partitions

All partitions are to be set plumb, well braced, and nailed. Studs at all angles and openings are to be doubled, and extra block is to be set at door openings for base nailing. All partitions that are not supported below are to be firmly trussed and braced. Ceilings to all closets will be furred down to within 12 inches of the door head except in closets more than 2 feet deep. There will be trued $^7/_8$-inch grounds at top of base and around all openings.

Lumber

All outside finish-lumber will be clear white pine unless otherwise specified. All exterior finish lumber is to be free from large or loose knots and will also be clear and thoroughly dry.

Sheathing and Sheathing Paper

Cover all the exterior walls with $^3/_4$-inch plywood sheathing nailed to each stud with 8d nails. With joints cut on studs or backed for end nailing, cover with Tyvek, which is to be well lapped, extending under all trim and around all corners to make a complete and tight job.

Exterior Finish

Windows, door casings, cornices, corner boards, water table, brackets, band courses, and so on are to be made to the detail furnished in the drawings. The stock moldings that are to be used are numbered on the drawings. The first story is to be covered with the best-grade cedar lap bevel siding, laid at $4^1/_2$ inches to the weather. The second story and gables are to be covered with 18-inch hand-split and resawn shakes, laid at $8^1/_2$ inches to the weather. Use hot-dipped galvanized nails, whose length will be approved by the architect. Window casings will be laid $2^1/_2$ inches to the weather, and the front door frames and casement windows will be according to detail shown in the plan.

Shingling

Cover all roofs with 8-inch × 16-inch Pennsylvania blue slate, laid 7 inches to the weather. All hips and other parts that require it are to be made secure against leaks by the proper use of slaters' cement and proper flashings. An ornamental galvanized-iron ridge crest will be placed on the main ridge. See details on drawings for this crest.

Flashing
Flash around chimneys, over all doors and windows, heads exposed to the weather, and where roofs join walls with 16-ounce sheet copper. Do the same in all valleys and wherever required to secure a tight job. Each side of a valley is to have a water check turned up 1 inch in the metal.

Flooring
First and second stories are to have double floors, with a subfloor of $^{5}/_{8}$-inch CDX plywood. The first- and second-story finish flooring is to be oak tongue-and-groove strip flooring, which is to be thoroughly seasoned and blind-nailed over building paper. There will be no joints in the main hall and only one joint in the run of boards in other rooms of the first floor. The second-story floors are to be cleaned and sandpapered to a smooth finish for the painter. Oak thresholds are to be set to all outside doorways, and hard rubber-tip doorstops are to be located behind all doors that open against a wall.

Window Frames
These are to be made of seasoned white pine.

Sash
All sash and frames are to be made by the Johnson Corporation and are to be of kiln-dried, vinyl-sheathed white pine. The numbers are given on the drawings.

Screens
All windows that open are to be fitted with bronze- or copper-wire window screens.

Glazing
All sash and outside doors, where indicated, are to be glazed with Johnson insulated windows or their equivalent. All hall doors are to be glazed with French plate. The plate in the Dutch door will be beveled. The basement sash is to be glazed with a single-strength glass.

Blinds
All windows, where indicated, are to be provided with an approved type of blind that will be $1^{1}/_{8}$ inches thick and made of the best grade of seasoned white pine. All blinds will move freely after painting. The blinds are to be hung on approved cast-iron blind hangers.

Door Frames
All inside door frames in finished parts of the house, first and second stories, are to be made of white pine $^{25}/_{32}$-inch thick, set plumb and

true, and blocked in four places on each side. Outside doorframes are to be rabbeted for doors. All frames are to be flush with the plaster finish.

Doors

Unless otherwise specified, all inside doors are to be made of slab-type birch veneer with hollow cores and will be $1^3/_8$ inches thick. Outside doors are to be $1^3/_4$-inches thick and will be made of solid-core, slab-type birch veneer. The front doors are to be of the design shown in the drawings. Hang all doors throughout with loose-joint ball-tip butts of sufficient size to throw them clear of the architraves. Doors are to have three $3^1/_2$-inch × $3^1/_2$-inch butts on $1^3/_8$-inch doors and three 4-inch × 4-inch butts on $1^3/_4$-inch doors. A hardware schedule will be furnished. Hang both double-swing butlery doors on double-acting brass spring hinges. Furnish all nails, except those used for inside work, galvanized and all other hardware that will be necessary for the completion of the work in the proper manner.

Interior Trim

For the basement, the interior trim is to be selected cypress or redwood. For the first and second floors, the trim is to be unselected birch. There will be a 4-foot 6-inch paneled wainscoting in the dining room, first floor hall, and up stairway. This panel will be made of $^1/_4$-inch birch plywood, with trim as shown in the details. There will be a 5-inch cabinet plate shelf in the dining room, the bottom member of which will be a picture molding. This shelf will match the door head trim.

Stairs

The main staircase is to be made of unselected birch. The stringers and treads are to be $1^1/_8$ inches thick, as shown in details. The risers are to be $^3/_4$ inch thick. The risers and treads are to be housed into the wall stringer and return-nosed over the outside string. The rails are to be 3-inch × 3-inch molded, with ramps as shown in the details. Balusters will be $1^5/_8$ inches, taper turned, three to a thread, and proportionately more for increased widths. Newels and column newels are to be as shown in the details. The run on the first flight is to be $10^1/_4$ inches from face to face of the rise with 12-inch treads. The second flight and basement stairs are to be $9^3/_4$ inches of run with $11^1/_2$-inch treads. For the basement stairs, cut 2-inch × 12-inch boards for the stringers and 2-inch × 10-inch yellow pine for the treads.

Mantels

There will be two mantels where indicated on the drawings. The contractor will figure them to cost $2000 each complete, including

linings and face and hearth tile. This amount will be allowed the owner to use at his or her option in the selection of it. The entire cost is to be figured in the contract price, including the setting of the mantels by the contractor.

Pantry Cabinets

There will be a cabinet, where indicated, with three glass doors, above the draining board. This cabinet will be 10 inches deep inside and will contain three shelves. The wall cabinets are to be constructed as shown in the details.

Closet Shelving

The trim on the inside of the cabinets is to be plain. There is to be an average of 10 feet of 12-inch shelving to a closet, with 6-inch clothes strips and 1 dozen clothes hooks, japanned. The kitchen closet and the closet under the kitchen stairs are to have suitable shelving and sufficient pothooks and other fixtures. There will be 25 feet of shelving in the shop closet. There will also be 1 dozen clothes hooks under the basement front stairs.

Plumbing

All necessary materials for completing the plumbing installation, as hereafter set forth, in a correct and sanitary manner are to be included in the general contract. The state plumbing code shall be strictly followed.

Electric Wiring

No. 12 Romex sheathed cable is to be installed under and subject to the requirements and regulations of the *National Electrical Code* and all state, county, and municipal codes. The locations for all electrical outlets will be shown in the drawings.

Water Pipes

Water is to be brought from the street main into the house through $^3/_4$-inch copper tubing, or plastic if preferred. Copper water tubing is to be used on all straight-line work. Place a hose-bib cock on the main at a point against the house for hose purposes, both front and rear, with a stop and waste cock in the basement.

Complete all necessary digging for the laying of sewer and water pipes to the house. No trenches are to be less than 36 inches below the grade at any point. The pipes are to enter the house in the basement at a suitable point for intersection with the inside piping system. The dirt is to be well rammed over the pipe in the trenches as it is refilled. The house sewer is to be installed by the plumber, and

it will consist of a perfect 4-inch glazed socket-jointed tile pipe to a point exactly 4 feet outside the foundation wall. The soil pipe is to be 4-inch cast iron. All water pipes must have a gradual fall from the fixtures that they supply, and they must open at their lowest point for drainage purposes. All the cast-iron waste pipe will be furnished with the necessary fittings.

Furnish and install one 60-gallon gas-fired water heater of an approved type and manufacture. This water heater is to be supplied with water through a $^3/_4$-inch copper tube. Place a shutoff cock on the supply pipe. Take hot water from the water heater to and over the kitchen sink and to all other fixtures, except toilets, with $^1/_2$-inch copper tubing. The supply to the toilets will be through $^3/_8$-inch copper tubing. Cold water to all fixtures will be through separate pipelines. There must be no depressions in any pipe, and hot water must be kept rising from the boiler head.

Kitchen Sink

Furnish and install a cast-iron enameled kitchen sink, with a garbage disposer of (*specify make and model*). Furnish and install an automatic dishwasher (*specify make and model*) where indicated in drawings.

Wash Trays

Provide and install where indicated on drawings, one two-part stone tub on galvanized iron legs. Supply the tub with hot and cold water through $^1/_2$-inch copper tube and brass faucets, one for cold water, threaded $1^1/_2$ inches waste, with traps, plugs, and chains complete. The waste drain is to be connected with the soil pipe through a 2-inch copper pipe.

Bathroom Fixtures

All bathroom fixtures are to be (*specify name of manufacturer*) make, as listed in their catalog. The basement-bathroom water closet will be (*name and model*); lavoratory, (*name and model*); bathtub, 5-foot 6-inch (*name and model*). The other toilets are to be (*name and model*). The two second-story baths are to be (*name and model*), 18 inches × 27 inches. Before any wall finishing is done, all supply pipes must be proven perfectly tight by a satisfactory test, and they must be left perfect at the completion of the test.

Painting

The entire exterior woodwork, except shingles, is to be painted with three coats of (*manufacturer's name*) best grade of ready-mixed paints, thinned as necessary and as specified by the manufacturer.

The color of paint will be plain white, except for the shutters and shakes, which will be green and will be given two coats of (*manufacturer's name*) shake and shingle finish. The second-story floors will be sanded smooth and given three coats of (*manufacturer's name*) polyurethane varnish. The first-story floors are to be left bare for wall-to-wall carpeting or vinyl tiling. All interior birch trim is to be finished natural, with one coat of white shellac and two coats of (*manufacturer's name*) of flat varnish. The type and color of the interior wall paint for all interior walls will be specified by the owner.

Condition of Bids

The owner reserves the right to accept or reject any or all bids.

Summary

A specification is a statement containing a detailed description or enumeration of particulars, as of the terms of a contract, and details of construction usually not shown in an architectural drawing. Great care should be used when reading specifications to avoid misunderstandings and disputes. Each item entering into the construction is defined and described with such precision that there can be no chance of misunderstanding or double interpretation.

Framing will be indicated on the drawings. Specifications as to type of timber used (such as No. 1 common yellow pine, square, straight, and free from any imperfections that will impair its longevity) will be indicated. Headers over window and door openings will be indicated as for size and installation procedure.

Interior trim, including stair casing, will be shown. Moldings used throughout the house will indicate miter angles and type of lumber used. Ceiling height (as well as plastered or drywall construction) will be specified. Most plastered walls will be specified as to type of finish coat (which, in most cases, is at the option of the owner) and will also specify that exposed corners be protected with metal corner beads.

All necessary materials for completing the plumbing installation, as set forth in the specifications, shall meet all state and local regulations. Electrical wiring will be listed and meet the requirements and regulations of the *National Electrical Code,* or state and local codes.

Review Questions

1. Why should great care be used when reading specifications?
2. What information is included in the specifications on framing a house?

3. What code should be followed when plumbing a house?
4. Can a specification on a dwelling be changed or altered? If so, how?
5. What code should be followed when electric wiring is installed?
6. Why should the purchaser of a house not only read, but also study all specifications prepared by the builder?
7. What should the clearance of the ceiling be in the basement?
8. What is an elevation? Where is the word found?
9. What is Tyvek? Why would you want it on a house?
10. What is the height of the second-story ceiling (between timbers)? Is this higher or lower than that of the first story?
11. What does the architect place in the specifications sheet?
12. Who determines the number of bathrooms in a house?

Chapter 9

Architectural Drawings

Drawings provide designers with a practical method of communicating their ideas to the carpenters and builders. Drawings are a type of shorthand that not only reflects, but also fixes on paper, the ideas of the designers. They help avoid possible misinterpretation.

Use of Drawings

Contractors and estimators should retain prints on jobs that they complete so that they might be used to compare future jobs. They should also keep at least two sets of prints on a job in progress. One set is the working-drawings for the tradespeople. On the second set, the lead carpenter adds the changes made as the structure is built, to be turned over to the owner to update the original drawings issued. The final drawings will identify any changes that were made during construction. The owner may change the original drawings to reflect the changes, so that the drawings will be fully up-to-date to assist with maintenance, troubleshooting, and design of additions to the original building.

Reading Drawings

In the drawings (sometimes called *prints* or *plans*), various elevations are shown (such as the front, side, and sectional). A plan is a horizontal view of an object. An *elevation* is a vertical view of an object.

Projected Views

The various views are projected on imaginary projection planes, similar to the projection of a picture on glass. To illustrate the first (or front) view, place a clear pane of glass in front of the object with the glass parallel to the surface of the object being projected. Figure 9-1 shows a simple building with a shed roof. In front of the building is the pane of glass marked *V*, representing a vertical plane.

When an observer looks through the glass directly at the front of the object from a considerable distance, the observer will see only one side, in this case the side marked *ABCD*. The rays of light falling upon the object are reflected into the eyes of the observer, and in this manner, the observer *sees* the object. The pane of glass (vertical plane) is placed so that the rays of light from the object will pass through the glass in *straight parallel lines* to the eyes of the observer.

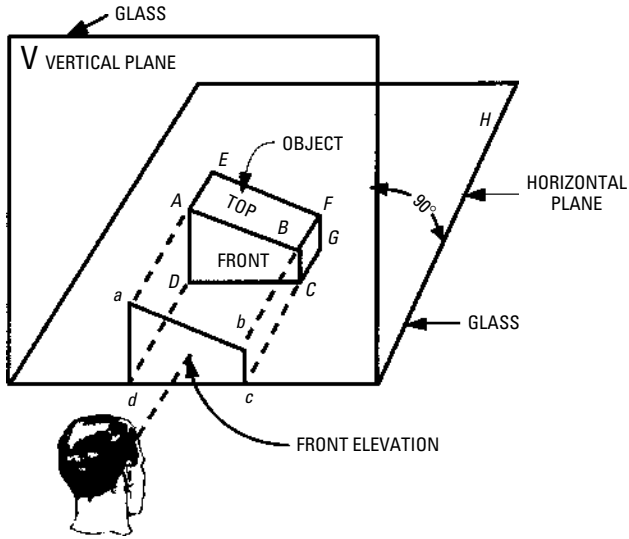

Figure 9-1 Rectangular object resting on a horizontal plane and facing a vertical plane.

The rays of light from points *ABCD* of the building pass through the glass at points *a*, *b*, *c*, and *d*. If these points (*a*, *b*, *c*, and *d*) are connected by lines, a view of the object as seen from the front is obtained, which is called *front elevation*.

The front elevation is identical in shape and size with the front side *ABCD* of the object (that is, $ab = AB$; $bc = BC$, and so on; angle dab = angle DAB; angle abc = angle ABC, and so on).

Top View or Plan

For this view, place a pane of glass in a horizontal position above the building that is resting on the horizontal plane (see Figure 9-2). Now, look at the object directly from above. Note that the rays of light from corners *AEFB* of the top pass through the glass at points *aefb*. If these points *aefb* are connected by lines, a view of the object as seen from the top is obtained, which is called the *top view*, or preferably *plan*.

Right-End View (Elevation)

A pane of glass is placed to the right of the building in a vertical position and parallel to the right side *BFGC* of the building (see Figure 9-3). Here, the pane of glass is marked *P*, which means *profile plane*, or plane from a side projection. Looking at the building directly from

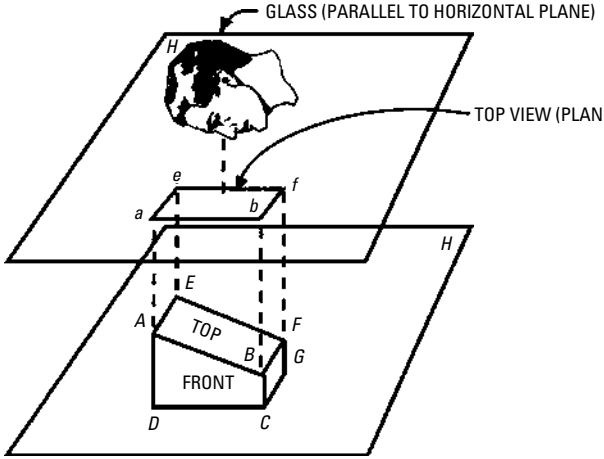

Figure 9-2 Projection of the top of an object to obtain top view.

the right side (as in position *S*), the rays of light from corners *BFGC* of the upper left-hand side (from points *AE*) pass through the glass at points *bfgc* and *ae*. If these points are connected by lines, a view of the object as seen from the right side is obtained, which is called the *right-side view*, or preferably *right-end elevation*.

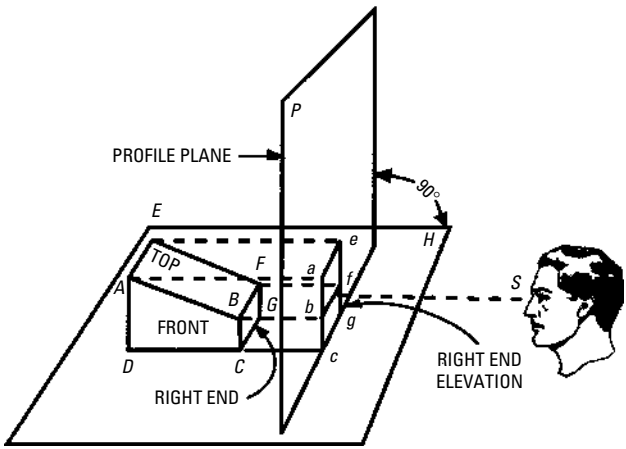

Figure 9-3 Projection of the right end of an object illustrating right profile plane.

The shape of the object is such that the entire visible surface does not lie in a plane parallel to the projection plane. The points *A* and *E*, though located at the other end of the object, are visible and accordingly form part of the right-end view. Figure *aefb* does not show the top in its true size because it is projected obliquely instead of at 90°. An *oblique projection* makes an object appear smaller than its real size.

Left-End View (Elevation)

With a pane of glass shifted to the left side of the object (see Figure 9-4), and the building viewed directly from the *left* side (as position *S*), the rays of light from corners *ADHE* of the left side pass through

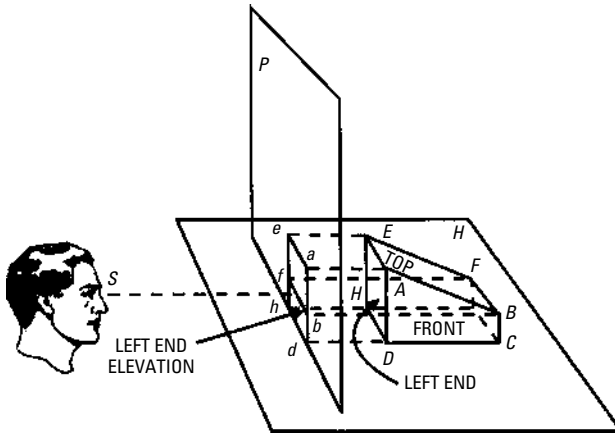

Figure 9-4 Projection of the left end of an object illustrating left profile plane.

the glass at points *adhe*. If lines connecting these points are drawn on the glass, a left side view of the object is obtained. However, the edge *FB* at the other end is invisible. It is shown by a dotted line connecting *f* and *b* projected from *F* and *B*. The completed drawing is then called a *left-side view*, or preferably a *left-end elevation*.

Sections

Most buildings are so complex they cannot be clearly represented by a plan and elevation alone. In such a case, the parts that do not appear properly in these drawings are better represented by a *section*, or *sectional views*. A cross-section is a drawing of a building showing that part cut by a plane (see Figure 9-5).

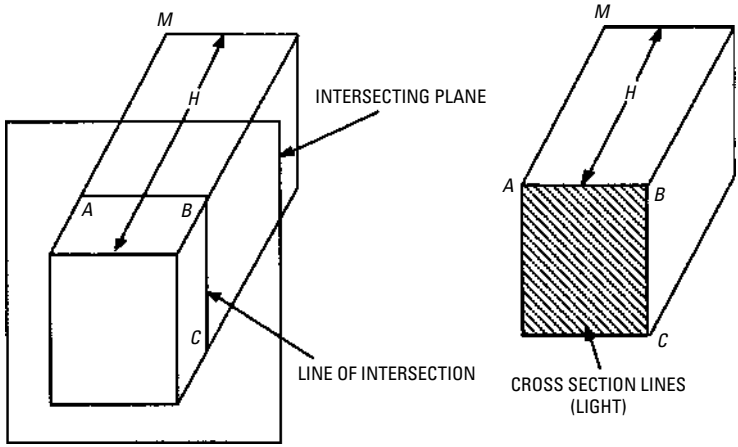

Figure 9-5 Pictorial view of a building illustrating cross-section.

Directions of View for Sectional Views

For an unsymmetrical object, it is important to know the direction in which the sectional view is viewed. This is indicated by arrows at the end of the line representing the cutting plane (see Figure 9-6).

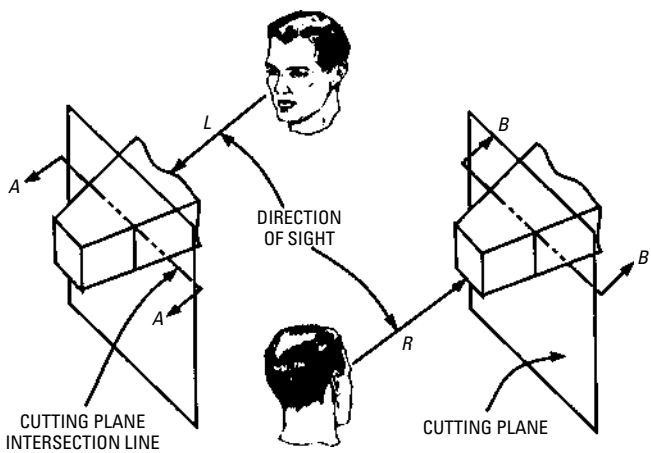

Figure 9-6 Direction in that a sectional view is viewed is indicated by the direction of arrows *AA*, at the ends of the cutting plane intersection line.

The arrows *AA* indicate that the object is viewed in the direction of point *L* (toward the smaller end of the object), and the arrows *BB* (in the direction of *R*) toward the larger end.

The Scale

Scale is the ratio between the actual size of the object and the size that it will be drawn. The scale is usually expressed on a drawing as full size, half size, quarter size, and so on, or it might be expressed as 1 inch = 1 foot, 1 inch = 100 feet, 1 inch = 1000 feet, or any other proportion that might be necessary to use. The scale is printed on the drawing.

On a full-size drawing, the object and drawing are of the same size. When the drawing is marked half size, the object is twice the size of the drawing. Thus, the drawing of an object is shown full size, half size, and quarter size (see Figure 9-7). If the building's height is

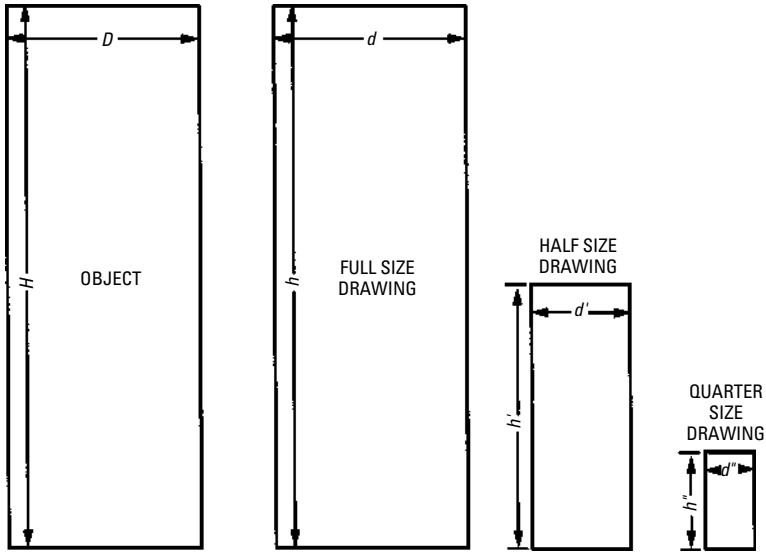

Figure 9-7 A drawing of an object to different scales: full size, half size, and quarter size.

represented by H and its diameter by D, then these dimensions will be the same for the full-size drawing. That is, $H = h$; $D = d$. For the half-size drawing, $h' = {}^1/_2\ H$; $d = {}^1/_2\ D$. Similarly, for the quarter-size drawing, $h'' = {}^1/_4\ H$; $d'' = {}^1/_4\ D$.

From this it is seen that when the length of any edge on the drawing is made the same as the length of the corresponding edge on the object, the drawing is marked *full size* (sometimes *actual size*). If the length of any line on the drawing is half the length of the corresponding line on the object, the drawing is *half size*.

The scale is smaller than the building. In using prints, the size is important, since prints are used in the field, and if they are too large, they may not be easily handled.

In the case of a building, it would be impossible to have a print as large as the building. Thus, it is necessary to cut the print down in size. This necessitates the use of a scale. In the drawing of a building the building, designers usually express scale as 1 inch = 1 foot, $^1/_2$ inch = 1 foot, and so on. This would indicate that 1 inch on the drawing would be equal to 1 foot on the actual structure, or $^1/_2$ inch on the drawing would equal 1 foot on the actual structure, respectively. The architect uses the architect's scale. This is laid out in inches ($^1/_8$, $^1/_4$, $^1/_2$, and so on).

To lay off a distance of 2 feet 6 inches, place the $^3/_4$ scale with division 2 at the given point *A*; then the zero division on the scale will be at a distance of 2 feet (see Figure 9-8). Because the end space is divided into twelfths, each division represents one inch on the $^3/_4$ scale. Therefore, measuring off six divisions indicates that *AB* = 2 feet 6 inches. Notice the difference in actual length of this measurement on the 1 inch = 1 foot scale (see Figure 9-9).

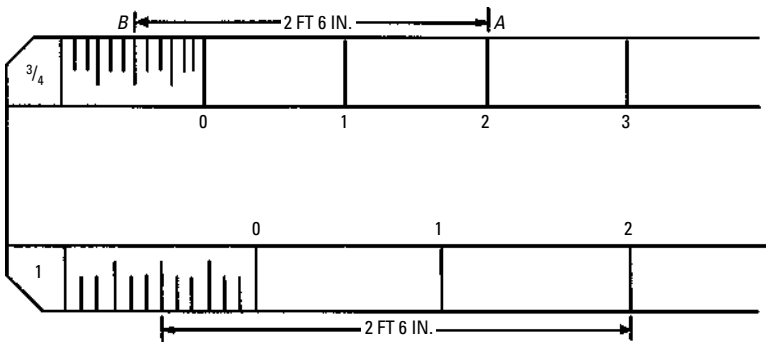

Figure 9-8 A typical architect's scale with detail showing $^3/_4$ and 1 inch to the foot.

The scale on the original would not apply to a reproduction made by photocopying. The scale on the original (see Figure 9-10) should

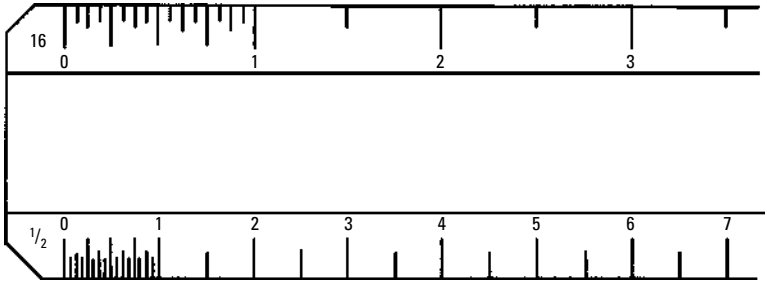

Figure 9-9 A typical architect's scale with detail with inch divisions instead of inches to foot.

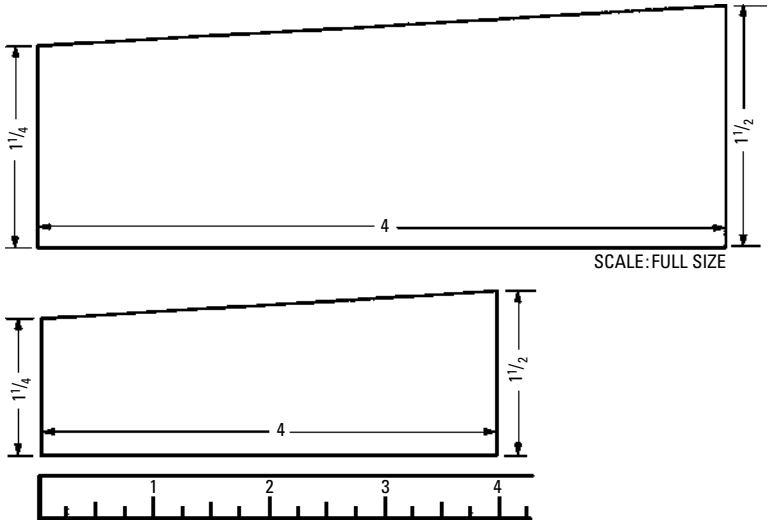

Figure 9-10 Graphic reproduction scale.

be crossed out and a *graphic scale of proportions* corresponding to the reproduction added.

Drawing Development

Knowledge of how architectural drawings are developed will help the carpenter and builder read drawings with more complete understanding. Architectural drawings are a means of transferring the thoughts of the designer to the builders and carpenters whose responsibility it is to construct the building. Graphic symbols are used to locate specific features and where they are to be placed.

The different parts or drawings that are necessary to show the structure (such as the mechanical and electrical installations) are all shown graphically. Therefore, you must become familiar with these symbols, not only those of one particular trade, but those of all the trades. This is necessary so that complete coordination may be reached between the various trades. In the construction of a building, time and money may be saved by representatives of the different trades going over the plans and laying out the pattern to be followed. No trade can work independently of the others. If independence is attempted, confusion is created and some work must be done over to make all parts of the scheme fit together. During construction, the general contractor, the plumber, the steel workers, and the electrical and mechanical contractors must lay out the work together and determine from the drawings who installs what, where, and when.

On the job, the architect may have a representative present to assist in coordinating the work and in making decisions that may be required. In designing the building, the owner or builder will often draw a rough sketch, after which he or she sits down with the architect to discuss what the owner will need and want in a building design. The requirements are noted as to space, machinery, electrical loads, numbers of persons that will occupy the building and what the future requirements might be. The owner will sometimes have a rough sketch of ideas (see Figures 9-11 and 9-12). These need not be drawn to scale or with any degree of accuracy. They are merely ideas of what the owner might want. In this discussion, no elaborate plans are given. Simple plans are used, as they show the intent and fall into line with more detailed plans.

Notice that there are no details shown, merely a sketch of the spaces to be enclosed. After the sketch is drawn, the architect and owner can sit down and discuss details, at which time, no doubt, another freehand sketch will be drawn with more details. When a tentative solution is reached, the architect will make a preliminary drawing (see Figure 9-13).

When the owner has signed construction papers, the architect will start drawing up the final plans and all details. There are many preliminary things to do, such as surveying the land to see how much excavating will be required. The location of the property lines and the general drainage plan for the immediate vicinity must be considered, and the water, sewer, gas, telephone, and power lines that exist must be considered. Local regulations regarding types of construction permitted, setbacks, and so on, all must be taken into consideration.

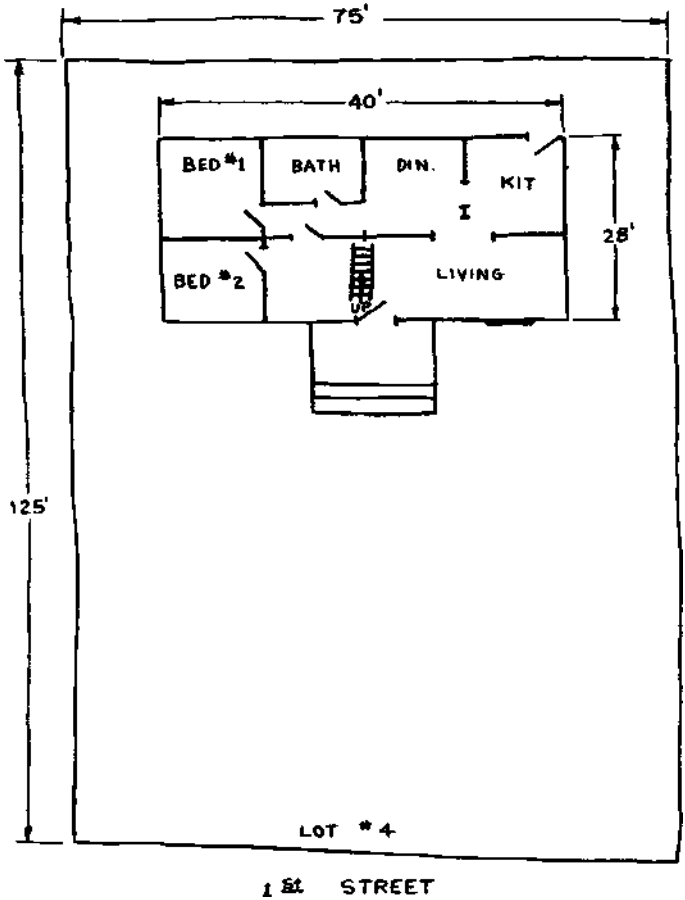

Figure 9-11 Owner's sketch of the first floor of a residence.

The final drawings are sent to the plan checkers of the inspection departments having jurisdiction. Here, they are checked to see that they conform to local codes. Corrections are noted, or the plans are approved. Most specifications that accompany plans put the burden of following applicable local codes on the builders and contractors. When questions arise (such as an electrical or mechanical problem), the contractor involved takes these problems up with the architect or an assistant who, in turn, takes them to the engineer who has performed the design work.

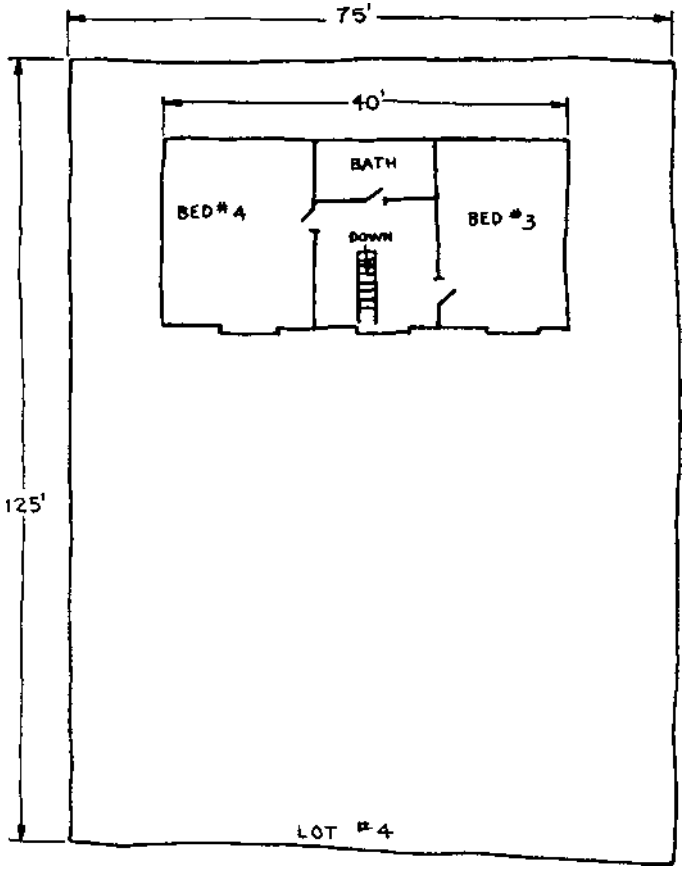

Figure 9-12 Owner's sketch of the second floor of a residence.

There are various methods of bidding on plans. Sometimes, the general contractor gives the entire bid, and then calls for bids from subcontractors. This method has some advantages in that the general contractor is responsible for the entire job. At other times, the general contractor and each subcontractor bid their parts separately. This method has an advantage. The general contractor's percentage is removed from the subcontractors' bidding, and the owner has more control over who gets the bids. The details of the architect's sketches for the preliminary part are not important to the mechanic. Therefore, they will not be covered in this book. What the mechanic is interested in is how to perform part of the work.

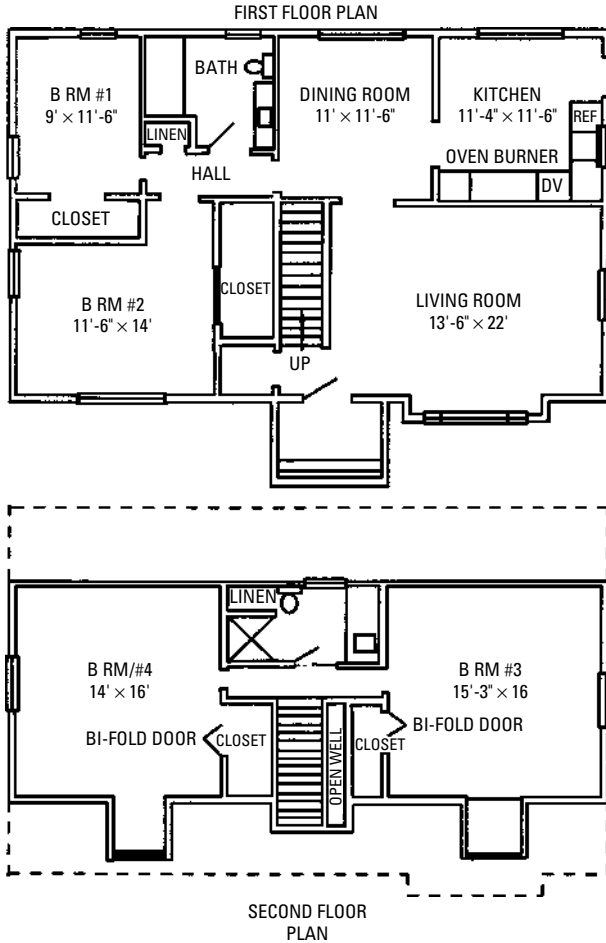

Figure 9-13 Architect's drawing (preliminary), without dimensions, of the owner's idea.

Typical floor plans of the first and second floors are shown in Figures 9-14 and 9-15. Figures 9-16 and 9-17 show typical floor plans of a first and second floor, with dimensions added. Figure 9-18 is a typical plan of a basement, with dimensions added. Figure 9-19 illustrates a typical detail of the basement wall, footings, and floor. Figure 9-20 shows a typical detail of the floor joists, brick veneering, and so on as they attach to the foundation.

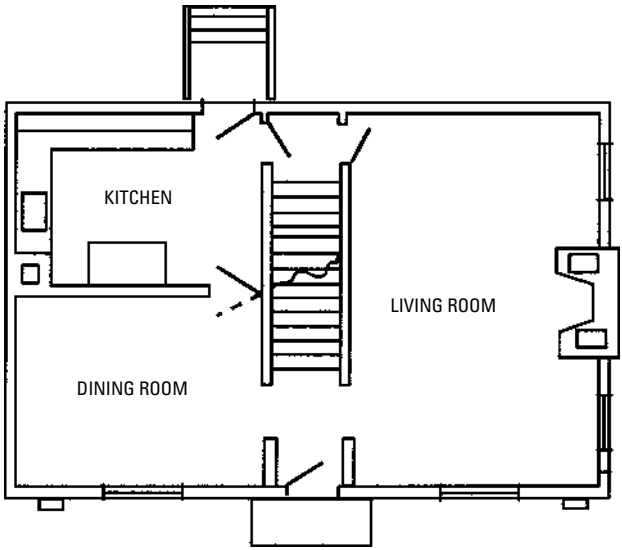

Figure 9-14 A typical first-floor plan.

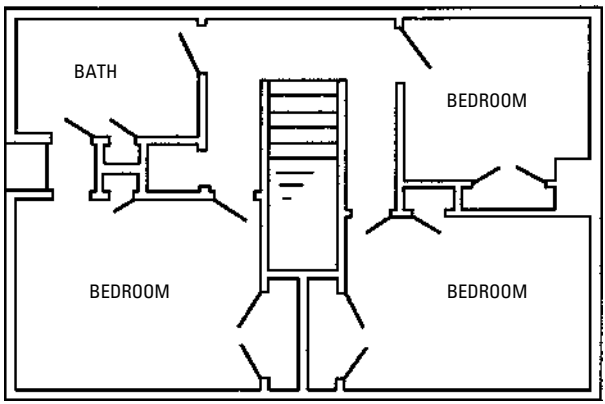

Figure 9-15 A typical second-floor plan.

Graphic Symbols

In architectural drawings, a form of shorthand is used to illustrate what is to be installed and at what point or location the installation will be in the building. These are commonly known as *graphic symbols*.

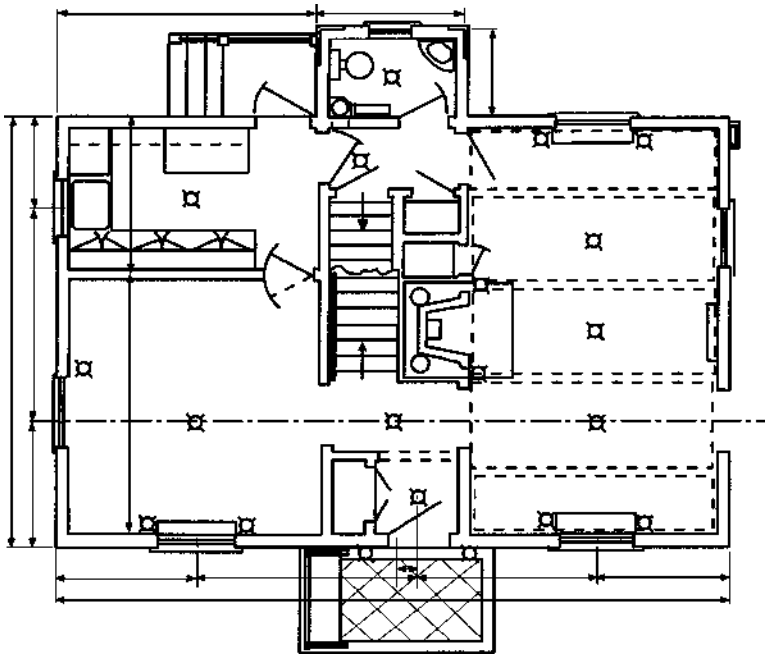

Figure 9-16 A typical first-floor plan with dimension lines.

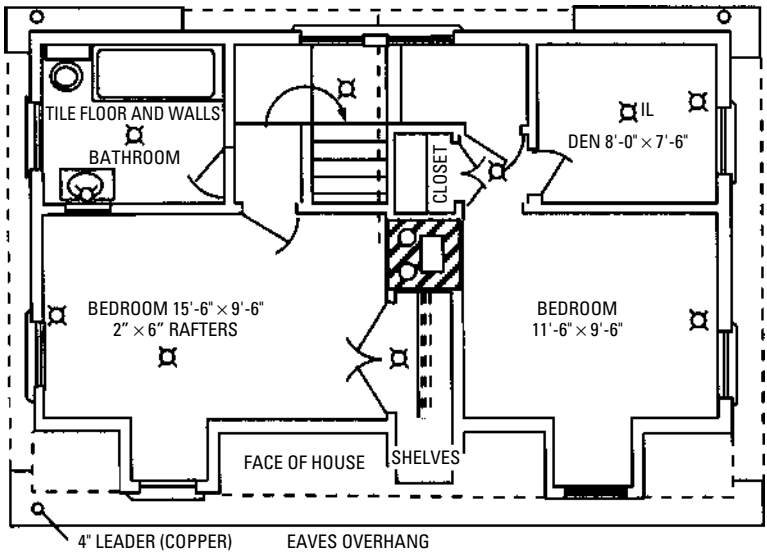

Figure 9-17 A typical second-floor plan with dimensions.

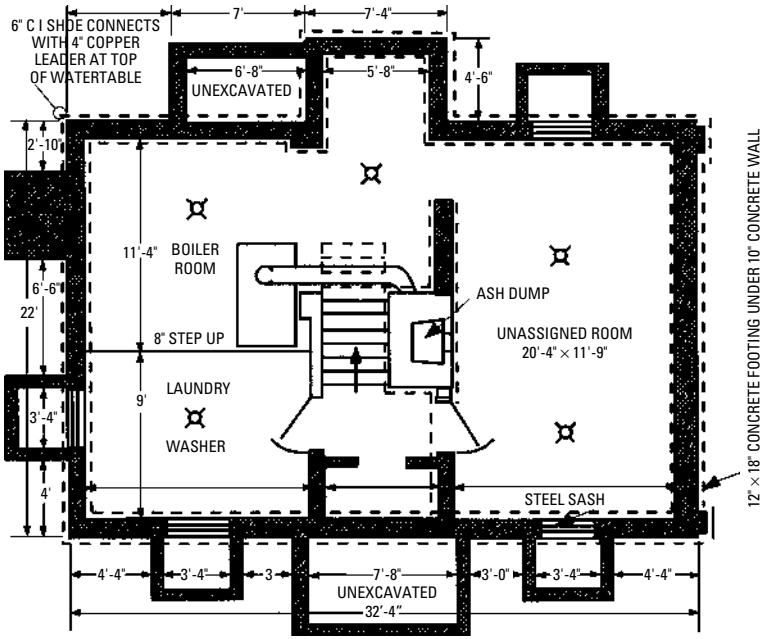

Figure 9-18 A typical basement plan with dimensions.

Symbols are used in drawings to represent various parts and systems. You must also become familiar with these. Each trade has its own symbols, and the workers of each trade should learn to recognize the symbols of all the other trades. For example, the electrician should understand the plumber's symbols. The plumber should understand the carpenter's symbols, and so forth. In this way, each craftsperson will know what obstacles may be encountered in the work, and will be better prepared to cope with them.

In this section, the symbols shown are standards for the construction industry. However, you will find that some designers or individual institutions will deviate from these standards. Where this is done, a legend showing what the symbols mean should be added to the drawings.

A drawing consists of many different kinds of lines, each having its own purpose. Certain characteristic lines are used to convey different ideas, and the drafting practice has been rather well standardized as to the use of lines to avoid confusion in reading drawings. A good working drawing is as simple as possible, using only

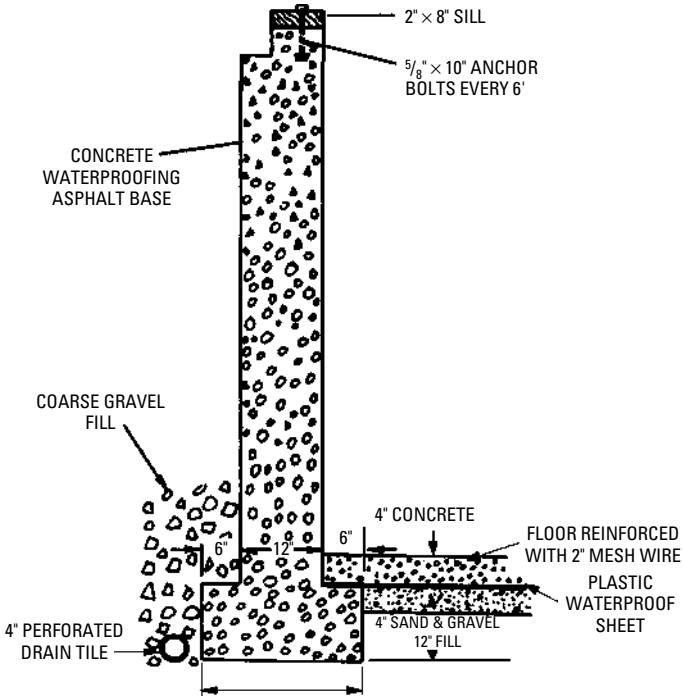

Figure 9-19 Detail of basement walls, footings, and floor.

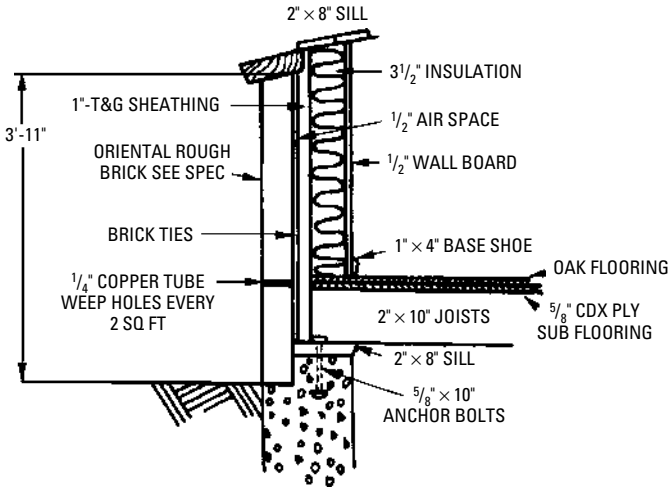

Figure 9-20 Detail of floor, brick veneering, and so on.

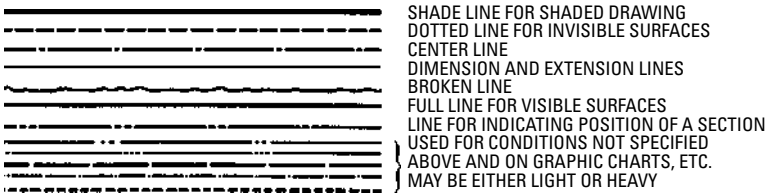

Figure 9-21 Various lines used in drawing.

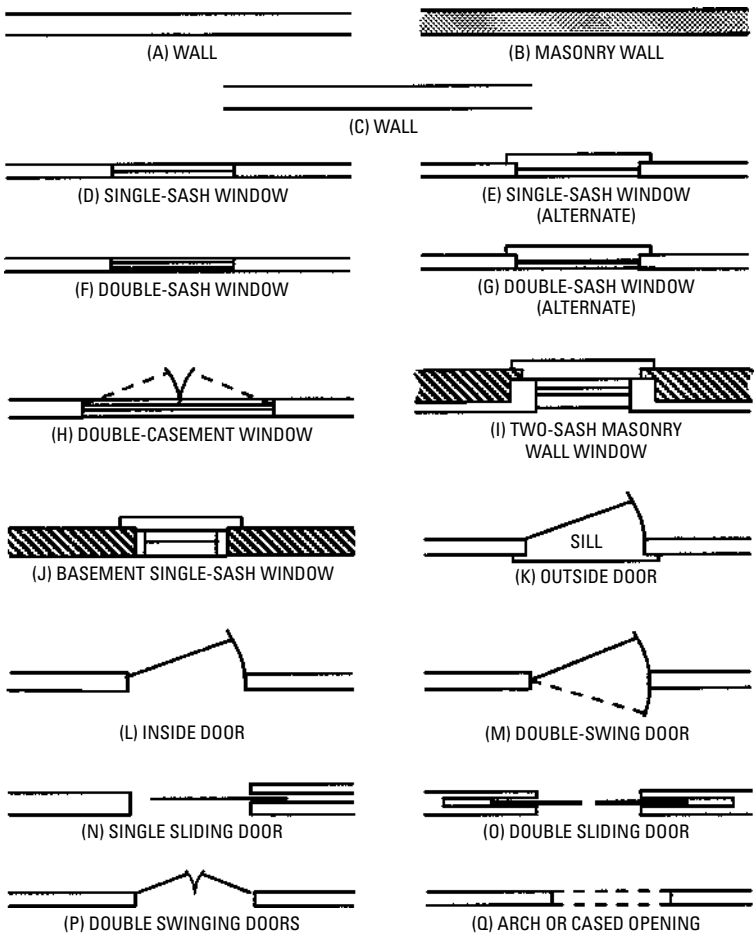

Figure 9-22 Symbols for walls, windows, and openings.

such lines as are necessary to give all of the required information. Moreover, the reader will not have to puzzle over a mass of lines that complicate the drawing. The same thing holds true for dimensions and other data. A good drawing is accurate and complete, though simple, and is therefore easily understood (read) by the carpenter.

Figure 9-21 shows the lines generally used on drawings.

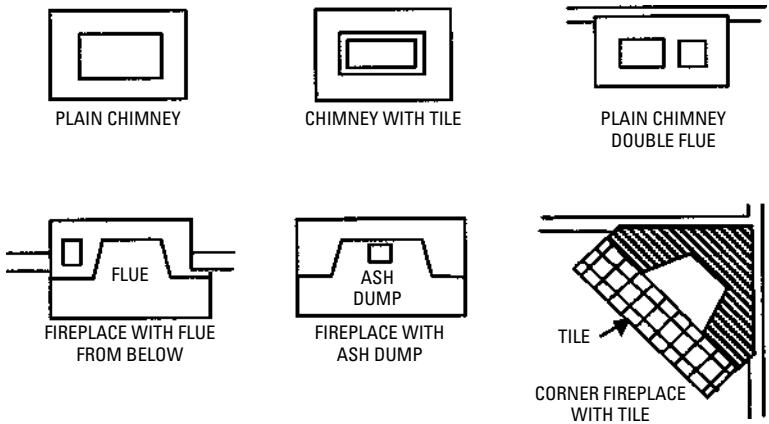

Figure 9-23 Symbols for chimneys and fireplaces.

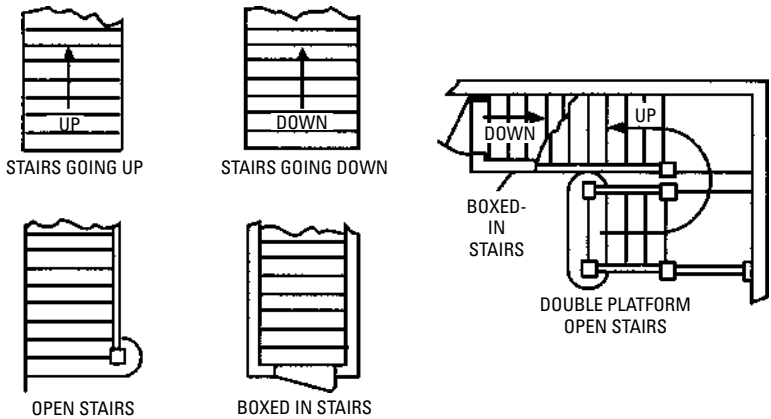

Figure 9-24 Symbols for stairs.

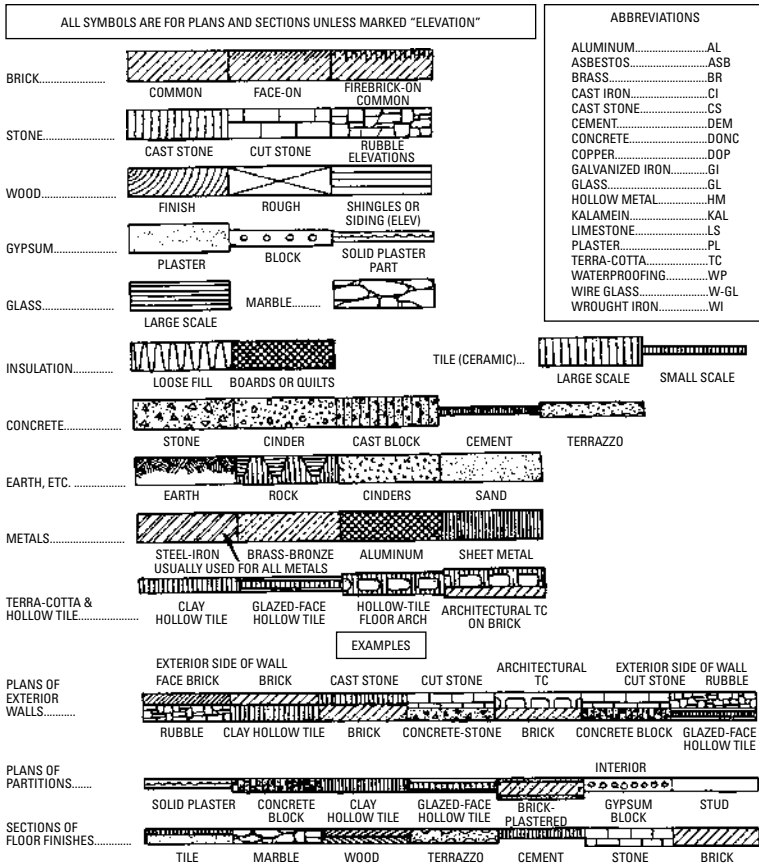

Figure 9-25 Symbols for various materials.

Walls of frame buildings are represented on floor plans by two parallel lines spaced at a distance apart equal to the wall thickness (see Figure 9-22A). Masonry walls are shown on a floor plan by cross sectional lines, as shown in Figure 9-22B. Walls of all types of construction may also be shown as in Figure 9-22C, by heavy dark lines that save time in drawing and give a better print.

There are many and varied types of window construction, the symbols for some being shown in Figure 9-22D–J. The specifications should show the materials and types of construction, thickness of glass, and type of glass to be used. These details may be listed as a supplement to the specifications, or on the drawings if there is room.

There are many types of doors that will be used, the symbols for some of these being shown in Figure 9-22K–Q. There may also be special doors used, and where this is called for, a drawing showing the detail should accompany the main drawing. Detailed sketches or drawing inserts should show all details of sills, especially where masonry construction is to be used. In drawings, the dashed line should be avoided where it is intended to indicate some part that is in view. The dashed line is ordinarily intended to represent some hidden feature or part.

Figure 9-23 shows a few conventions or symbols for chimneys and fireplaces. There may be special features that should be shown in additional drawings. In each case where details are required, a notation should be added referring to the detail drawings.

Stairs must be identified as to their direction, and whether they are boxed or open. Figure 9-24 shows some methods of identification. Arrows show the direction of the stairs.

Figure 9-25 shows symbols for the identification of materials.

If confused as to an abbreviation while reading a drawing, considering the nature of the work will be helpful in interpreting the abbreviations. It should be understood that these abbreviations relate only to one part of the subject. For every field (such as carpentry, electrical work, and so on), there are many conventions relating to each individual field. Table 9-1 shows common abbreviations used in construction drawings.

Table 9-1 Common Abbreviations Used in Construction Prints

Construction Feature	*Abbreviation*
Access door	AD
Access panel	AP
Acoustic	ACST
Aggregate	AGGR
Aluminum	AL
Anchor bolt	AB
Angle	ANG
Apartment	APT
Area	A
Area drain	AD
Asbestos	ASB

(continued)

Table 9-1 (continued)

Construction Feature	*Abbreviation*
Asbestos board	AB
Asphalt	ASPH
Asphalt tile	A Tile
Automatic washing machine	AWM
Basement	BSMT
Bathroom	B
Bath tub	BT
Beam	BM
Bearing plat	BRG PL
Bedroom	BR
Blocking	BLKG
Blueprint	BP
Boiler	BLR
Bolts	BT
Book shelves	BK SH
Boundary	BDY
Brass	BR
Broom closet	BC
Building line	BL
Cabinet	CAB
Caulking	CLKG
Casing	CSG
Catch basin	CB
Cellar	CEL
Cement floor	CEM FL
Center	CTR
Center-to-center	C to C
Center line	C/L
Ceramic	CER
Channel	CHAN
Cleanout	CO
Clear glass	CL GL
Closet	CLO
Cold air	CA
Cold water	CW
Conduit	CND

(continued)

Table 9-1 (continued)

Construction Feature	*Abbreviation*
Counter	CTR
Cubic feet	CU FT or FT^3
Detail	DET
Diagram	DIAG
Dining alcove	DA
Dining room	DR
Double-acting door	DAD
Double-strength glass	DSG
Drain	D or DR
Electric panel	EP
End to end	E to E
Excavate	EXC
Expansion joint	EXP JT
Finished floor	FIN FL
Firebrick	FRBK
Fireplace	FP
Fireproof	FPRF
Flooring	FLG
Flush	FL
Footing	FTG
Foundation	FND
Frame	FR
Garage	GAR
Gas	G
Gage	GA
Gypsum	GYP
Hall	H
Hardware	HWD
Hose bibb	HB
Hot air	HA
Hot-water tank	HWT
I-beam	I
Inside diameter	ID
Insulation	INS
Iron	I
Kitchen	K

(continued)

Table 9-1 (continued)

Construction Feature	*Abbreviation*
Knocked down	KD
Landing	LDG
Lath	LTH
Living room	LR
Main	MN
Matched and dressed	M & D
Maximum	MAX
Medicine cabinet	MC
Minimum	MIN
Miscellaneous	MISC
Mixture	MIX
Mortar	MOR
On-center	OC
Pantry	PAN
Partition	PARTN
Plaster	PLAS
Plate	PL
Porch	P
Precast	PRCST
Prefabricated	PREFAB
Pull switch	PS
Radiator	RAD
Recessed	REC
Refrigerator	REF
Register	REG
Revision	REV
Riser	R
Rivet	RIV
Room	R or RM
Rubber tile	R Tile
Screen	SCR
Section	SECT
Sewer	SEW
Shelving	SHELV
Shower	SH
Single-strength glass	SSG

(continued)

Table 9-1 (continued)

Construction Feature	*Abbreviation*
Sink	S or SK
Soil pipe	SP
Square feet	SQ FT or FT^2
Stairs	ST
Standard	STD
Switch	SW or S
Storage	STG
Telephone	TEL
Thermostat	T or THERMO
Tongue-and-groove	T&G
Unexcavated	UNEXC
Vent	V
Vinyl tile	V Tile
Washroom	WR
Water	W
Water closet	WC
Water heater	WH
Weatherstripping	WS

Summary

Drawings are a means of communication between the designers and the builders. They are used on the construction site for reference by the trade workers and as a record of how the building was actually built. The building is represented with various views (such as plans and elevations). Complex views of the building are shown with sections. Sections are like a slice of the building that opens up the interior to a view. A floor plan is a horizontal slice.

Drawings are made to scale. There is a consistent relationship between the length of lines on the drawing and the size of the building and its parts.

Drawings are developed in phases that allow the designer to set the owner's requirements down on paper and show how the design can be changed to meet various functional and legal requirements.

A drawing is made up of symbols that represent the various parts of the building and where those parts will be installed during construction. The symbols are usually standardized so everyone knows what they mean. Simple symbols like lines are used to define the basic

shape of the structure. Complex symbols indicate doors, windows, stairways, and other common parts of the building.

Review Questions

1. What is the basic purpose of using drawings in construction of a building? How is this accomplished?
2. Why are the drawings updated during construction?
3. What part of a building does an elevation show?
4. What part of a building does a section show?
5. Why are drawings reduced in size by scaling?
6. Who, besides the carpenters and builders, uses the drawings of a construction project?
7. What do graphic symbols do in a drawing?
8. Draw the symbol for a door and a window.

Chapter 10

Building Styles Explored

Today's homes may resemble those from the New England days of the 1600s, or they may look more like a manufactured-in-the-factory house. There is a large range of styles and looks available for those looking for a new home. Contractors have plans and even model homes to aid in making a selection. There are even certain suggested styles and limits to housing designed for the retired couple with no adult children living with them. Everywhere you look, you see evidence of a variety of tastes expressed in living quarters. Here, we will be exploring some of those styles and learn how to make an educated choice for your next home.

The Two-Story New England Colonial House

Many houses of the Colonial style of architecture, built in the late 1600s, are still standing in the New England states.

Many of the original houses were built to accommodate larger families, but certain architectural details were common to all of them (see Figures 10-1 and 10-2). One feature found in most of the original houses is moderate to steep roof slopes, often the one-third pitch. This was necessary to allow the use of wood shingles or shakes, about the only roof covering then generally available. Narrow eaves (with little or no projection at the gables) had the lap siding cut against wide corner boards. Plank frames were used for the windows, and there were no casings (or very narrow casings) outside. Entrances, however, were usually elaborate, sometimes with finely scrolled and carved pediments. Good, authentic replicas of many of these entrances are obtained today. In the later and more pretentious houses built in this era, sidelights were often used at the entrances, and sometimes the doors were double. The relatively wide pilasters at the sides were usually fluted or molded. A type of door that originated in England is paneled three high and two wide, with the top pair of panels smaller (nearly square). The rails in the upper part of the door form the Christian cross.

The front of the house was symmetrical about the central entrance. Although there were some exceptions to this design, symmetry was the rule. Invariably, second-story windows were placed directly over lower story openings. Small covered entrances were uncommon, porches and verandas virtually unknown. The central entrance hall was universal, with the stairway to the upper floor. In the earlier and smaller Colonial houses, the stairway

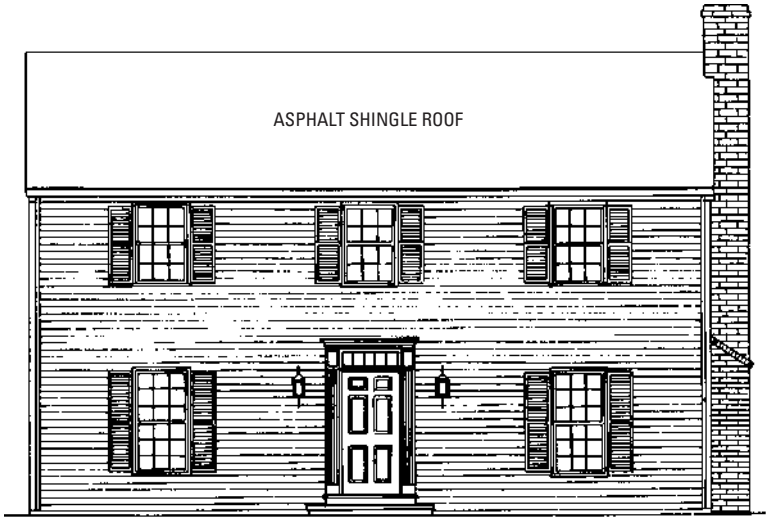

Figure 10-1 Front elevation of a Colonial-type home.

Figure 10-2 Right side elevation of a Colonial-type home.

was often steep and tortuous, often with tricky winders. However, in the later and better homes, the stairs were often beautiful, tastefully designed and elaborately carved, with superb craftsmanship. Such stairs cannot be duplicated today.

While the basic design was rectangular, many of the Colonial homes had attached ells or sheds. With a roof continuous down over a one-story shed at the rear, this house becomes the well-known *saltbox*, with a claim to fame all its own. All types of Colonial homes use windows of small glass or lights, often rectangular, sometimes diamond-shaped. In the early days, the muntins were often made of lead. Later, wood muntins were used. Large sheets of glass were almost unknown, and very expensive.

A predominant feature in the Colonial house is the *privacy* afforded, the privacy that is lacking, and so often deplored, in many modern designs. There are no unnecessarily large areas of glass to give one the eerie feeling of being spied upon at night.

Aside from the fact that this house is excellent architecturally, it need not be an expensive house (relative to other houses) to build. The downstairs bathroom is convenient, and the two complete bathrooms upstairs help prevent congestion (see Figures 10-3 and 10-4). A complete basement is suggested for convenience to heating and

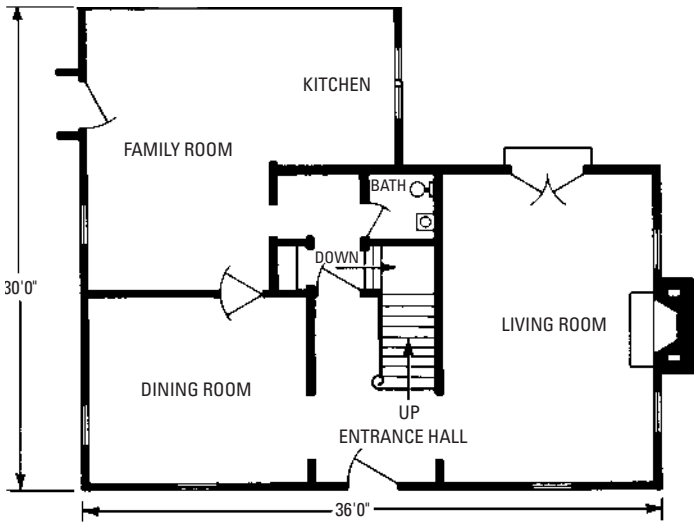

Figure 10-3 First-floor plan of a Colonial-type home.

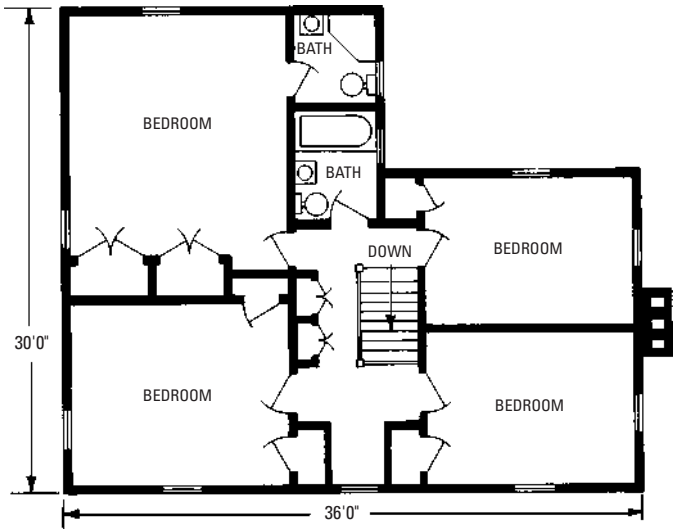

Figure 10-4 Second-floor plan of a Colonial-type home

plumbing, plus storage space. Included in the original plans is an adjacent garage. It has a covered breezeway connecting with the outside door of the family room.

A House of Modern Architecture

Architects who have pride in their originality often prefer an unusual custom design (see Figure 10-5). This does not mean that a house of unusual appearance such as this one, no matter how

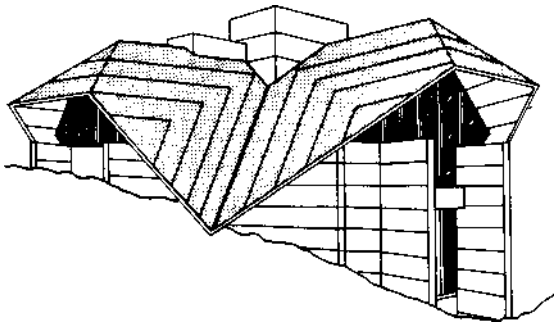

Figure 10-5 A house of modern architecture.

good the appearance, can be placed on the odd lot in a street of Colonial houses and have a pleasing effect. Houses such as this one need a proper setting, preferably a rugged and individual setting.

The design is bold, and reflects the lively imagination of its designer. Usually, modern practice makes use of large areas of glass, but these areas are effectively shaded by the wide overhangs, which make them much more acceptable than the large unprotected areas of glass in many modern designs.

The outstanding feature of this house is the roof. Although decidedly unusual, the pitches are regular (all 45°), and the roof is composed of simple intersecting planes. The roof framing is actually rather simple, for there are no *warped* planes. The architect has used diagonal lines on the roof instead of the commonly used vertical and horizontal accents. The effect is striking, especially on a sloping site, but working drawings are available that adapt it to level sites as well.

Although imaginative and unusual, the plan is in right-angled shapes for ease and economy of construction. The roof is formed of intersecting planes, with steep pitches. The roof treatment is important, because it dominates the entire view.

In the plans, designations of rooms, and dimensions, are intentionally omitted (see Figure 10-6 and Figure 10-7). Naturally, the uses of the rooms will depend upon the contour of the ground. The large room with the fireplace will undoubtedly be used for a living room, depending upon whether the entrance is at the second-story level (as it may be if the house is built on a sloping site), or at the first floor level (as it may be if the ground is level). The plan is flexible enough to allow considerable leeway in deciding upon the exact room arrangements and their subsequent uses.

In general, the walls are 12-inch lightweight concrete blocks, their cavities filled with insulation, with head joints cut smooth, and horizontal joints struck. The roof is sheathed with plywood (see Figure 10-8). With a roof so steep, many types of roof coverings may be used, but probably one of the modern types of mopped-down roof covers would be most acceptable, with granule-coated roll roofing as a cap sheet. The horizontal and diagonal rooflines are obtained by the use of battens.

The Contemporary House

Figure 10-9 shows a house called *contemporary architecture*. It embodies most of the features desired in homes of the present era, including the following:

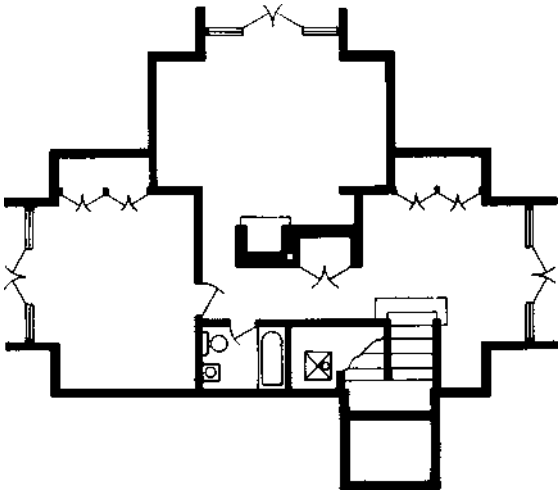

Figure 10-6 The lower-floor level of the modern house. If the entrance is at this level, the room with the fireplace will certainly be used as the living room.

- House is a split-level (not a new idea, but always popular).
- House allows adequate windows in the lower-level rooms, and the stairway down is short.
- Living room is moderately large, and, with the connected dining area, it is ample for the needs of most moderately large families.
- Bedrooms are all of practical, usable size. The spare room in the basement is a useful stand-by.
- Access door to the two-car garage is convenient.
- Family room and the party room in the basement remove some of the inevitable activities of a large family from the living room.
- Concrete patio at the rear is convenient for the usual outdoor living activities of the modern family.

Glass has been used with discretion. Although the glazed areas are moderately large, they are not unreasonably so, at the expense of privacy and to the detriment of heating and air conditioning.

Architecturally, this house is susceptible to many variations of materials. In the house illustrated in Figure 10-9, uncoursed native

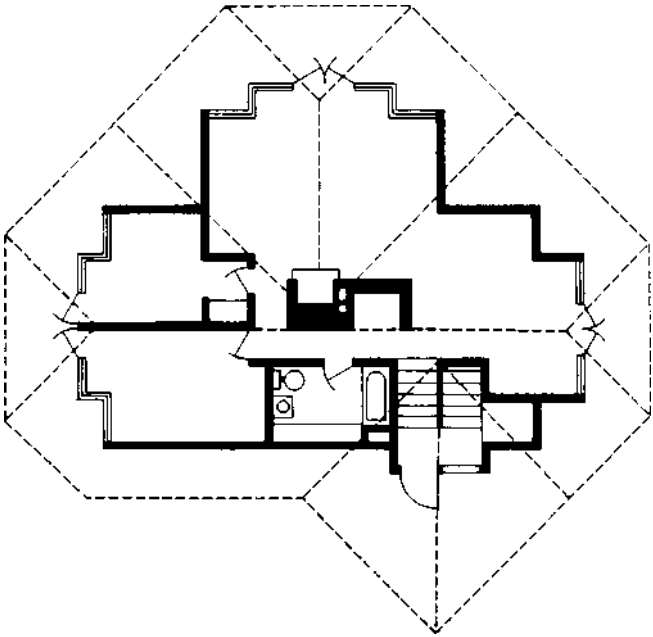

Figure 10-7 The upper-floor level. If the entrance is at this level, the room with the fireplace will certainly be used as the living room.

stone has been tastefully used, with vertical siding on the overhanging upper story.

The roof has sufficient pitch to fill its primary function that is to shed water, but it also gives a lower *spread-out* look to the house.

It is possible with the proper use of a steel beam to enlarge the party or family room. This could be done by removing the wall between party room and spare room, to make one big room in an L shape (see Figure 10-10).

Utility Pole-Type Building

The pole-type building requires no foundation (see Figure 10-11). It is mounted on round pole vertical members that are set in the ground and adequately anchored to resist uplift. For most buildings of appreciable size, the poles are set in round holes about 5 feet deep. Projecting nails or lag screws are placed in the pole 6 to 8 inches from the bottom, or holes are bored through which tight-fitting pieces of rod are driven to project a couple of inches on

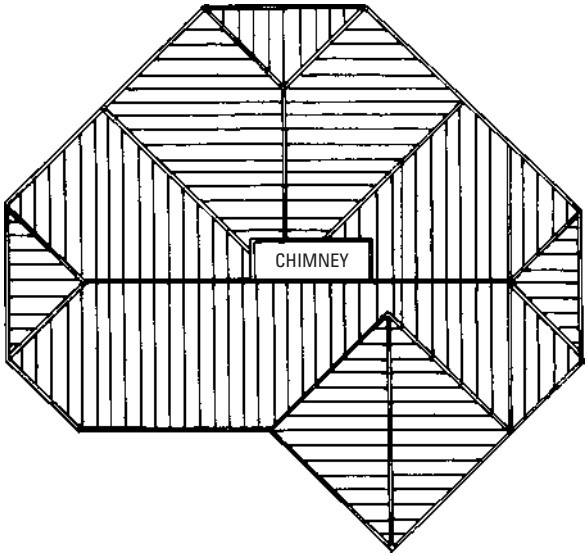

Figure 10-8 The roof framing plan of the modern house.

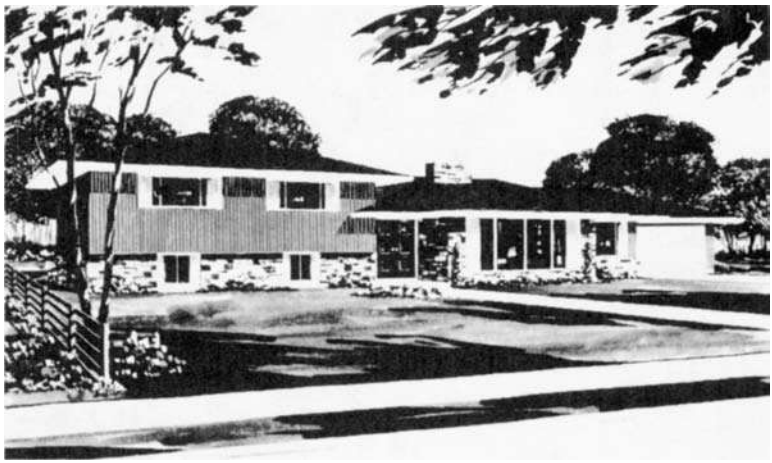

Figure 10-9 A split-level house.

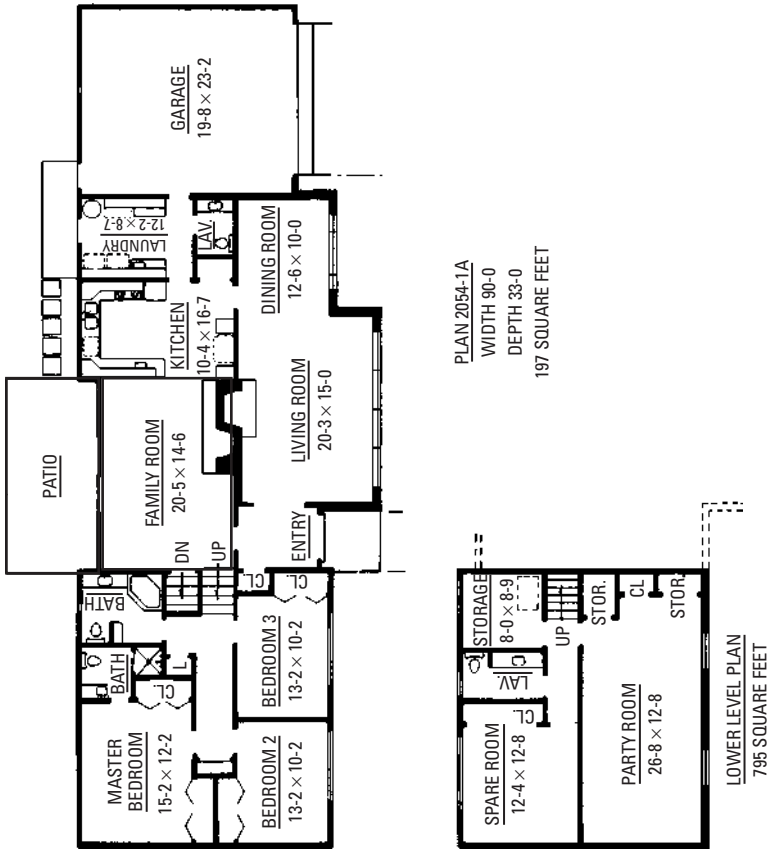

Figure 10-10 Floor plans for the contemporary house.

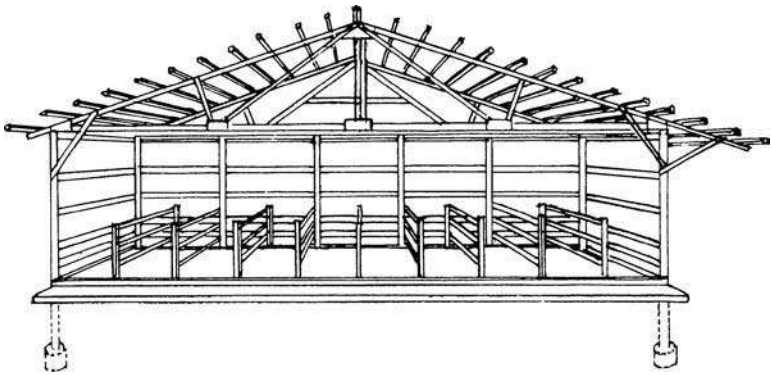

Figure 10-11 A pole-type building.

each side. After the poles are placed and positioned in the holes provided, about 12 inches of concrete is poured around the bottom ends and the holes filled with earth (see Figure 10-12).

Satisfactory joining of rafters and other framing members to the round poles may be made to properly use any standard connecting device (including plain bolts, split-ring connectors, toothed-ring connectors and plain-shank common nails).

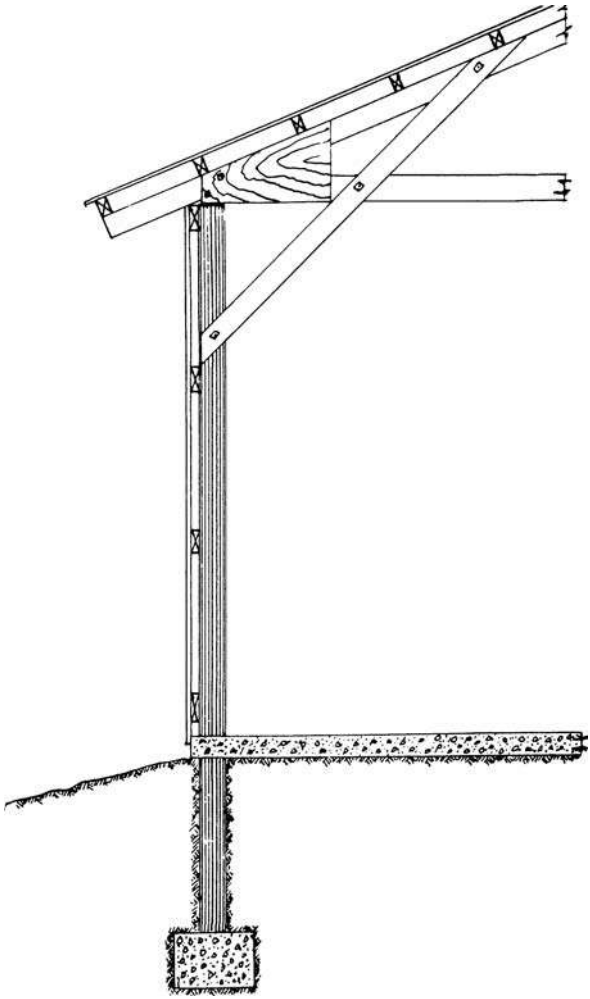

Figure 10-12 Enlarged detail of a pole-and-truss connection.

The idea of the pole-type building is not new. This type of construction was well-known at least 75 years ago, but it was generally used only on inexpensive and temporary buildings. The modern pole building can hardly be placed in this category. The poles used are now pressure-treated with a preservative and have a long life even under most adverse conditions. Modern timber fastenings are far superior to anything that our forefathers had available.

Where rafters or trusses connect to the poles, it is preferable that the poles be flattened at the contacting surfaces (see Figure 10-13). Cast-iron spike-grids, with one side curved to conform to the round side of the pole, have been satisfactorily used. However, simple bolts are also often used.

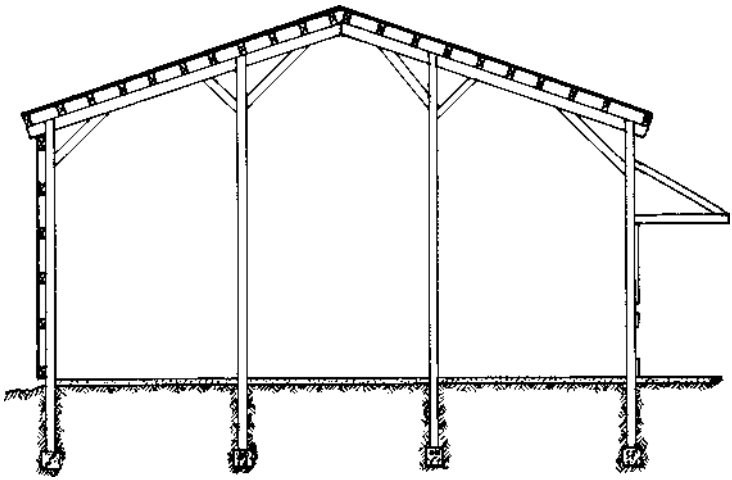

Figure 10-13 A bent of a pole-type hay storage barn, with feeding fence and canopy at the right. Roofs of these buildings may be trussed, eliminating the interior poles. Bents may be set 12 to 16 feet on centers.

Summary

Many of the two-story New England houses built in the 1600s are still standing today. Features found in most original houses were steep roof slopes, wood shingles, narrow eaves, and wide corner boards. The siding was always narrow when used in conjunction with the wide corner boards. Entrances, however, were usually elaborate, sometimes with finely scrolled and cornered pediments.

The contemporary house could be considered a modern home of the present era. Among the many styles in contemporary homes is the split-level, popularly called the tri-level. There are advantages to this type of house. The same space built in a split-level house will cost less than when it is built in a one-story.

A pole-type building (the most popular being the barn style) requires no foundation. The buildings are framed on round poles set in the ground and adequately anchored to resist uplift. The idea for pole-type buildings is not new, as this type of construction was well known at least 75 years ago.

Review Questions

1. What type of house was constructed in the 1600s?
2. What is a split-level house?
3. What are some advantages in the split-level house?
4. Give a few advantages and disadvantages of a basement.
5. What is a pole-type building?
6. What type of house is referred to as a ranch house?
7. What is the primary purpose of a roof?
8. Why would houses located in heavy snow areas have high-pitched roofs?
9. How old is the pole-type building for houses?
10. The _____ house could be considered a modern house of the present era.

Appendix

Construction Hardware

Information provided in this appendix is reproduced with permission from USP Lumber Connectors (United Steel Products Company, Inc.), a company that has been manufacturing and marketing construction hardware for more than 40 years. This information has been culled from the *Engineered Lumber Construction Hardware Product Guide and Design Manual* (Montgomery, Minnesota: United Steel Products Company, 1996).

You may contact the company's corporate offices at the following address:

Box 80
703 Rogers Drive
Montgomery, Minnesota 56069
1-800-328-5934

NDS Standards

The load resistance values presented in this appendix reflect the calculation criteria set forth in the *National Design Specification for Wood Construction* (NDS) published by the American Forest and Paper Association.

Because of this update, load values presented in previous publications may not match those presented in this catalog. The values shown in this appendix supercede those previously printed. An ongoing process is now underway to incorporate these new values in USP code reports. This will be done when each report is scheduled for reevaluation. Any questions about current code listings should be directed to the Technical Service Department at 1-800-328-5934.

Note

1) USP Lumber Connectors reserves the right to change specifications, designs and models without notice and liability for such changes. This appendix may not be reproduced in whole or in part without the prior written approval of USP Lumber Connectors.

2) This appendix reflects design changes and design load adjustments to some USP products. The information presented in this appendix supersedes all conflicting information published in previous documents

3) This appendix was designed as a general reference for the USP Product Line. Various specialized publications have been developed for design professionals, truss manufacturers, contractors, retail dealers, and building material wholesalers. Product load values may vary from one publication to another because of recent product testing or changes in regulations.

4) To achieve the allowable loads presented in this appendix, all specified fasteners must be used and proper installation procedures observed. Verify that the dimensions of supporting members are sufficient to receive specified fasteners. No product modifications are allowed without the written permission of USP Lumber Connectors.

5) Some connector models are listed more than once to indicate additional nailing options.

6) Nails specified as 8d, 10d, 16d, and 20d are common wire nails.

7) Bolts specified conform to ASTM A 307 standards or better

8) Products are sized for standard dimensional lumber or EWP sizes. For unusual supporting conditions, excessive shrinkage, or hostile environments, contact USP Lumber Connectors.

Code Evaluations

Most USP structural products listed in this catalog have been evaluated or are in the submittal stage for evaluation from one or more of the following code authorities:

- *CABO*—Council of American Building Officials (NER Report), National Evaluation Service.
- *ICBO*—International Conference of Building Officials.
- *BOCA*—Building Officials and Code Administrators International Inc.
- *SBCCI*—Southern Building Code Congress International, Inc.
- *METRO*—Dade County, FL.
- *DSA*—Division of State Architect, California.
- *LA CITY*—City of Los Angeles, California.
- *TDI*—Texas Department of Insurance.

Product Design Loads

The design loads listed are the lowest result obtained by one of the following methods:

- The ultimate tested load divided by three.
- Load producing 1/8-inch deflection.
- Nail I Bolt values based on 1991 NDS.

Durations of load adjustments for mechanical fastenings are as follows:

- Floor/Design Loads—No increase for duration of load.
- Roof Snow—115% of Design Load for two-month duration of load.
- Roof Nonsnow—125% of Design Load for seven-day duration of load.
- Uplift—133% or 160% of Design Load for wind I seismic loading.

All calculated load values shown in this publication reflect criteria established by the *1991 National Design Specification* (NDS).

Header Materials

- *I-Joist*—I-Joist with similar width to the supported I-joist
- *LVL*—Laminated Veneer Lumber
- *PSL*—Parallel Strand Lumber
- *LSL*—Laminated Strand Lumber
- *DF/SP*—Douglas Fir Larch/Southern Yellow Pine Grade 2 or better
- *SPF*—Spruce/Pine/Fir
- *2× Nailer*—DF/SYP Nailer

Note

Charts without specific header callouts ensure DF/SYP.

Testing

On all structurally rated products, USP performs full-scale testing to meet ASTM 0 1761, which is the testing standard recognized by all model agencies. All final testing is witnessed by an independent third-party testing laboratory.

Fasteners

A. *Nails*—Shall be common wire nails. Where special nails are required for the hanger, the manufacturer will furnish all 20d nails and larger, which are shorter than common wire nails of

same pennyweight. Special nails shall be furnished by manufacturer.

B. *Bolts*—Where required, shall be of nominal size shown on drawings and shall be furnished by the contractor.

C. *Expanding Anchors*—Use where shown on the drawing; furnished by contractor.

D. *Lag Bolts*—Shall be nominal size shown in drawings; furnished by contractor.

Execution

Examination

A. Verify that the wood that is to support the hanger is flat, is free of large cracks or splits, and is of the species noted on the plans. If the lumber or wood is not Douglas Fir-Larch or Southern Yellow Pine, reduction of the manufacturer's published load values, in accordance with the Building Code, is required. (A design professional shall reduce the published values in accordance with UBC Chapter 23 or the National Design specification of the American Forest and Paper Association.)

Installation

A. Use only nails of size shown in manufacturer's literature.

B. Fill all nail holes with nails as shown in manufacturer's literature.

C. Except where noted in manufacturer's literature, all nails are common wire nails and should be driven at right angles to the hanger body and seated fully.

D. *Do not* use roofing nails, which are not equivalent to common wire nails. DO NOT use nails that are shorter than $1^1/_2$ inches.

E. *Do not* use pneumatic or electric driven staples.

F. If power nail drivers are used, the nails shall be driven through the hole furnished in the hanger and shall not be driven through the hanger where no hole exists.

G. If wood I-joists are used, the maximum nail that can be driven into the top or bottom flange is a 9 gauge $\times$ $1^1/_2$ inches long (10d $\times$ $1^1/_2$ inches). The maximum nail that can be driven into the web is a 16d nail.

H. When installing hangers to laminated veneer lumber (LVL),

the maximum size nail that can be driven into the narrow face is 16d and the maximum size nail that can be driven into the wide face is 20d.

I. Hangers shall be installed tight against the supporting member.

J. The carried member shall be of sufficient length to maintain full bearing on the seat of the hanger. The carried member shall not fit tight (squeeze fit) against the carrying member. A gap of up to $^1/_4$ inch between the carried member and the carrying member is acceptable.

K. Before nailing the carried member into the hanger, take care to fully seat the member so that no space exists between the hanger and the bottom of the carried member. Where safe and practical, do not fasten the carried members to the hanger until floor sheathing has been installed. On some hangers, two diamond holes are located in the bottom of the seat that allow for wood screws to be installed. The screws pull the wood into the seat for positive bearing. Use 2 each #6 × $1^1/_4$-inch wood screws or #6 × $1^1/_4$-inch drywall screws.

L. Where bolting, use a wood template to accurately locate the holes. Drill hole $^1/_{32}$ inch minimum to $^1/_{16}$ inch maximum larger than the bolt diameter it is to receive. The bolt is to be tapped through the hole and connecting device with a hammer or mallet. Always use a washer under the head or nut if it is not in contact with a steel plate or hanger.

M. Lag bolts are not to be used in lieu of bolts unless specified.

N. Screws are not to be used in lieu of nails unless specified.

O. Use only stainless steel nails with stainless steel hangers.

P. Where wood connecting devices are attached to concrete, take care to install the device plumb, square, and true. If practical, use a plywood template to accurately locate the device.

Q. Anchor bolts, furnished and installed under Section 03300 *Cast-in-Place Concrete* shall be accurately located, plumb, and true before wood connecting devices are attached to the anchor bolts.

R. Where prefabricated structural wood is installed into the hanger, follow the written instructions of the prefabricated wood manufacturer in nailing, bearing support, bracing, and installation.

Wood I-Joist Installation to Wood

Sloped I-Joists

Use hangers with sloped seats whenever the slope exceeds the following: $^1/_2$:12 for seat bearing lengths of $2^1/_2$ inches or less; $^3/_8$:12 for bearing lengths between $2^1/_2$ inches and $3^1/_2$ inches; and $^1/_4$:12 for bearing lengths in excess of $3^1/_2$ inches.

Multiple I-Joists

Fasten together multiple plies of wood I-joists, in accordance with the manufacturer's installation guidelines, such that the joists act as a single unit.

Rotation on I-Joists

It may be necessary to install straps, blocking or sheathing to restrain torsional rotation of a supporting wood I-joist when using top mount I-joist hangers.

Fasteners

Install only the specified nails. The flanges of wood I-joists may split if larger diameter nails or longer nails are installed. Do not install nails larger than 16 penny common (0.162 inch diameter) into the web stiffeners in the wood I-joist.

Backer Blocks

Pattern the nails used to install backer blocks or web stiffeners in wood I-joists to avoid splitting the block. The nail pattern should be sufficiently spaced and avoid the same grain line, particularly with solid sawn backer blocks. Backer blocks must be installed on wood I-joists acting as the header, or supporting member. Install in accordance with the I-joist manufacturer's installation guidelines, with the backer blocks seated against the bottom flange of a supporting header when using the face mount hangers, and against the top flange if using top mount hangers. The nails installed to hangers mounted to an I-joist header must penetrate through the web and into the backer block on the opposite side.

Top Flange Hangers

The thickness of the metal and nail heads on top mount hangers must be evaluated for the effect on subsequent sheathing. Ensure the top mount hanger is installed such that the flanges of the hanger are not "over-spread" which tends to elevate the supported I-joist, causing uneven floor surfaces and squeaking. Similarly, ensure the hanger is installed plumb such that the face flanges of the hanger are mounted firmly against the wide-face surface of the header.

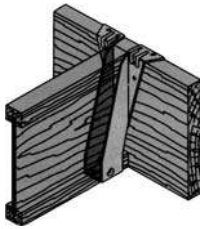

FLUSH FRAMING

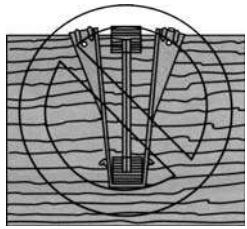

HANGER OVER-SPREAD

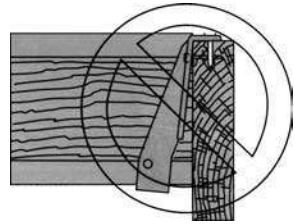

HANGER NOT PLUMB

Support Height and Lateral Stability

Hangers without web stiffeners must support I-joist top flange and provide minimum 250-pound lateral resistance with no more than 1/8″ horizontal deflection.

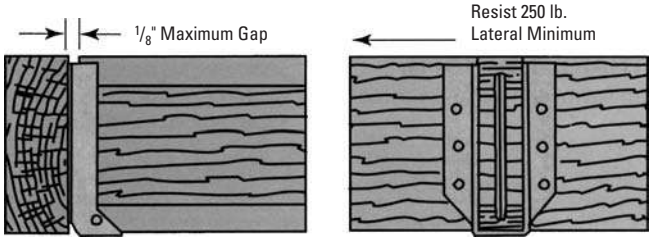

Without Web Stiffeners

Hangers for joists with web stiffeners must support a minimum of 60% of joist depth.

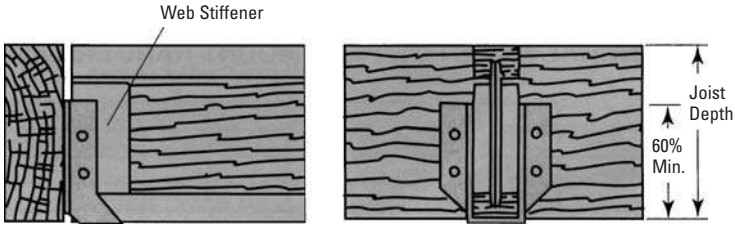

With Web Stiffeners or Solid Members

Wood I-Joist Installation to Steel I-Beam

Correct Hanger Attachment to Nailer

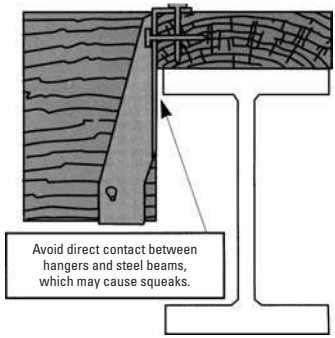

A nailer or sill plate is considered to be any wood member attached to a steel beam, concrete block wall, stem wall, or other type of support which is unsuitable for nailing, and is used as a nailing surface for top mount hangers to hold beams or joists.

Nailer Sized Correctly

Top flange of hanger is fully supported and recommended nails have full penetration into nailer, resulting in a carried member hanging safely at the proper height.

The nailer must be sized to fit the support width as shown and be of sufficient thickness to satisfy recommended top flange nailing requirements.

Note

A design professional must specify nailer attachment to steel beams.

Wrong Nailer Size Can Cause Component Failure

Too Narrow

Top flange not fully supported can cause nail breakout. Or, by fully supporting top flange, hanger is tilted back causing lifting of carrying member which results in uneven surfaces and squeaky floors.

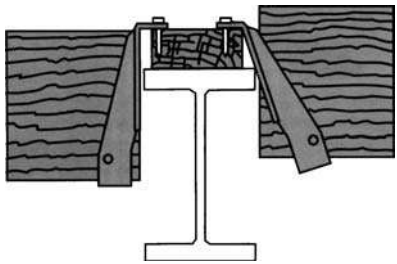

Too Wide

Loading can cause cross grain breaking of nailer. The recommended nailer overhang is $^1/_4''$ maximum per side.

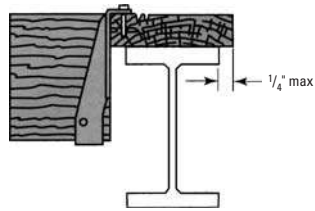

Too Thin

Top flange nailing cannot fully penetrate nailer, causing reduced allowable loads. Never use hangers which require multiple face nails since the allowable loads are dependent on all nail holes being used.

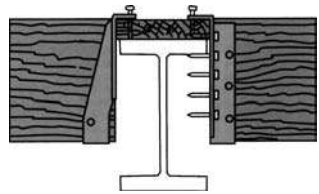

Welded Installation

Where shown in charts, hangers may be welded.
The minimum weld to the top flange is listed in the chart below.

Steel	*Weld*
10 gauge or lighter	$^1/_8'' \times 1^1/_2''$ each end
7 gauge	$^3/_{16}'' \times 1^1/_2''$ each end
3 gauge	$^1/_4'' \times 1^1/_2''$ each end

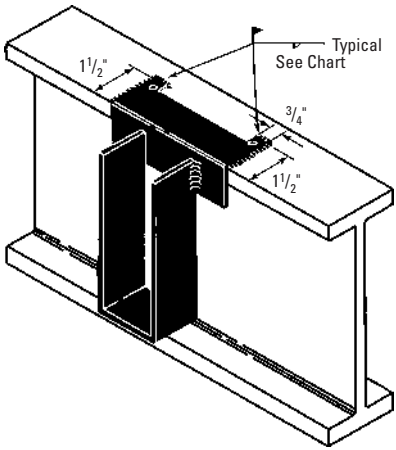

When welding galvanized steel, remove zinc coating by grinding before welding. Chip and touch up weld with primer.

Caution

Use proper respiratory protection when welding galvanized steel.

Top-Mount Hangers THO Series

This top mount hanger is engineered for I-joist to header applications. The THO offers full lateral support of the I-joist top chord which eliminates the need for web stiffeners in most applications. Raised dimple nailing guides help assure correct 45° nailing into the I-joist bottom chord. The new Seat Cleat™ prongs allow builders to"pop" I-joists into hangers for quick, positive seating, Seat Cleats™ are not yet available on all THO models.

Finish ..... G60 galvanizing
Code Listing ..... NER 478
U.S. Patent ..... #5,217,317
Seat Cleat Patent ..... applied for

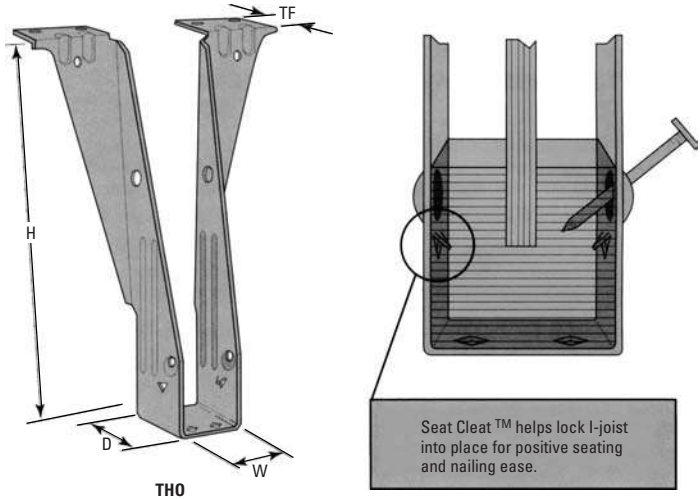

Beam and Purlin Hangers BPH and HBPH Series

These BPH hangers are used to support LVL, LSL and PSL beams and headers in light-to-medium load conditions. The slantback design incorporates strength and economy. An open back design allows the builder to install the BPH after the beam is in place.

Finish ..... G60 galvanizing
Code Listing ..... NER pending

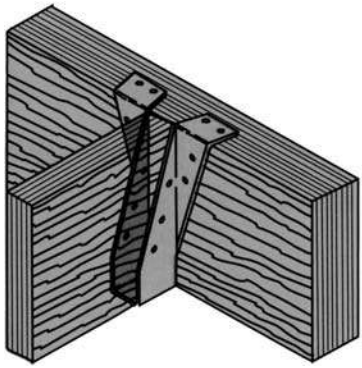

Typical **BPH** Installation

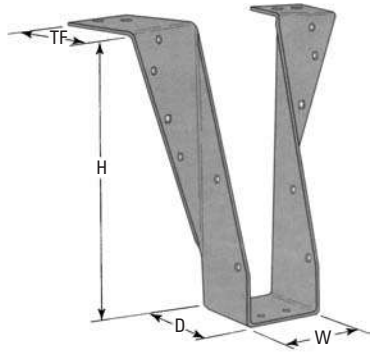

BPH / HBPH Similar

Beam Hangers LBH and HLBH Series

These heavy-duty beam hangers are specially designed for use in concentrated load conditions. The continuous top angle design offers high load values with minimum nailing. Many specialty options are available.

Finish ..... paint
Code Listing ..... NER pending
Specialty Options ..... see page 22

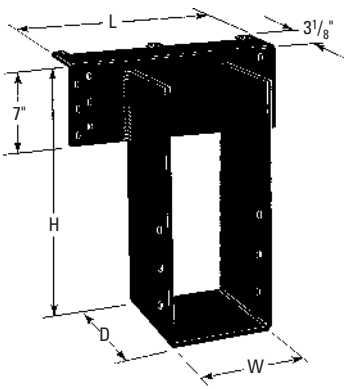

HLBH

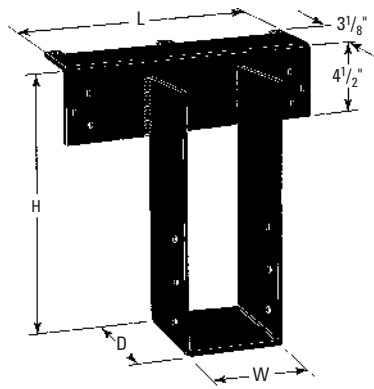

LBH / KGLT Similar

I-Joist Hangers PHI Series

The PHI series is specifically designed to support single-ply I-joists on I-joist headers. The PHI features a continuous top flange for increased strength. These hangers have also been engineered so that the top flange "depth" will match hanger seat width. This ideally places the top flange nails behind the I-joist's center web when the support joist is the same width. Joist nailing and top lateral support allows for installation without web stiffeners.

Finish ..... paint
Code Listing ..... NER pending
Specialty Options ..... see page 22

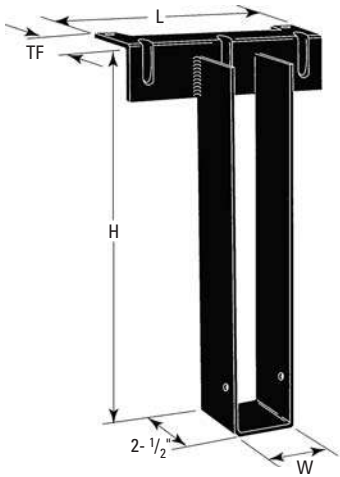

PHI

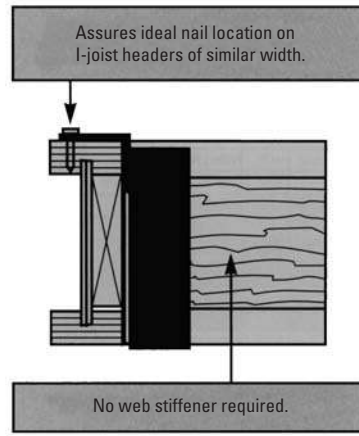

Beam and Purlin Hangers PH, PHM, PHX Series

Used to hang LVL, LSL and PSL beams and headers in medium load conditions using standard nails. The continuous top flange offers the best loading capacity with minimum nailing. A wide range of specialty options is available for diverse applications.

Finish ..... paint
Code Listing ..... NER pending
Specialty Options ..... see page 22

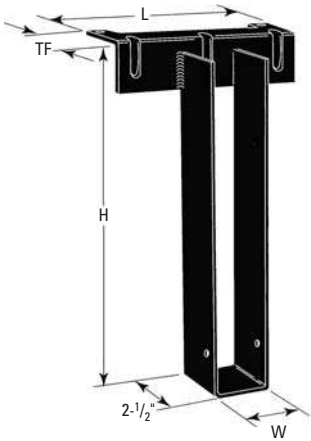

PH / PHM similar / PHX similar

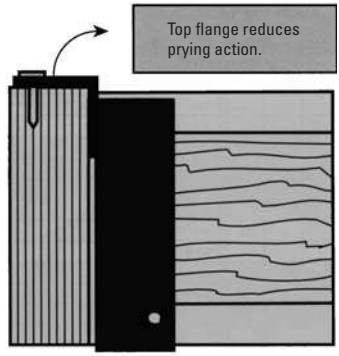

Masonry Hangers MPH Series

Specially designed to work with standard 6-inch concrete block or larger wall construction. The MPH eliminates the need for masons to fabricate special seats to support I-joists or composite wood beams. They can be easily slid into position while a mason is laying block and fastened by setting two 16d duplex nails into the concrete bed. Continuous top flange style gives strength with minimal fastening.

Finish ..... paint
Code Listings ..... NER 505
Specialty Options ..... see page 22

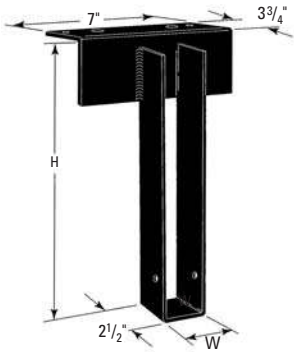

MPH

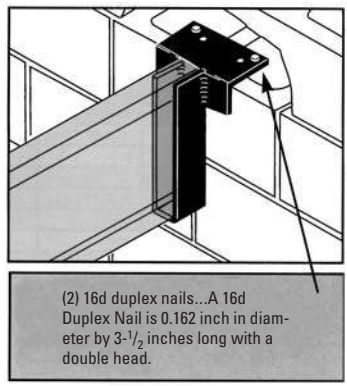

Top Mount Hangers

Joist (Inches)	USP Stock No.	Reference Number	Web STFNR Reqd	Steel Gauge	Dimensions W	H	D	L	TF	Fastener Schedule Header	Joist	100% Allowable Loads (LBS.) I-Joist	LVL	PSL	LSL	DF/SYP	SPF	†2 × Nailer	††Welded	†††Masonry
7 × 30	HBPH7130	HB1430-2	√	7	$7^1/_8$	30	3	—	$2^1/_2$	(10)NA25	(6)NA25	—	—	—	—	5345	3980	—	—	—
	PHX7130	HW1430-2	√	$^3/_{10}$	$7^1/_8$	30	$2^1/_2$	10	3	(4)16d	(2)10d	—	6015	5680	5755	5045	4230	—	—	—
	LBH7130	GLTV430-2	√	10	$7^1/_8$	30	4	12	$3^1/_8$	(9) NA16DRS	(6)16d	—	6500	6560	6010	6400	6400	—	—	—
	HLBH7130	HGLTV430-3	√	7	$7^1/_8$	30	6	12	$3^1/_8$	(15)NA16DRS	(6)16d	—	10620	10370	9600	9600	8915	—	—	—
	KGLT3530-2	—	√	$^3/_7$	$7^1/_8$	30	5	10	$2^1/_2$	(10)NA25	(6)NA25	—	—	—	—	8205	—	—	—	—
7 × 32	PHX7132	HWI432-2	√	$^3/_{10}$	$7^1/_8$	32	$2^1/_2$	10	3	(4)16d	(2)10d	—	6015	5680	5755	5045	4230	—	—	—
	LBH7132	GLTV432-2	√	10	$7^1/_8$	32	4	12	$3^1/_8$	(9)NA16DRS	(6)16d	—	6500	6560	6010	6400	6400	—	—	—
	HLBH7132	HGLTV432-2	√	7	$7^1/_8$	32	6	12	$3^1/_8$	(15)NA16DRS	(6)16d	—	10620	10370	9600	9600	8915	—	—	—

NOTE: See HEADER MATERIALS on page 3 for specific header information.
†For 2× Nailer installation, substitute 10d × $1^1/_2$″ nails for header nailing and fill top and upper face nails, If any. Refer to nailer installation on page 9.
††Refer to welded installation detail on page 8.
†††Refer to masonry installation detail on page 10.

MPH, PH, PHI, PHM, and PHX Series Options

Option	*Skewed*	*Sloped*	*Sloped/Skewed*	*Top Flange Offset*	*Saddle*	*Sloped Top Flange*
RANGE	1° to 60°	1° to 45°	skew 1° to 45° sloped 1° to 45°	—	—	0° to 35°
ALLOWABLE LOADS	100% of table load.	100% of table load.	100% OF TABLE LOAD UP TO MAX LOAD OF: PHX - 3500 lbs. PHM - 3000 lbs. PH - 2500 lbs.	HANGER WIDTH: %OF TABLE LOAD: $3^1/_2''$ or less 60% $3^9/_{16}''$ to $5^1/_2''$ 75% $5^9/_{16}''$ to $7^1/_2''$ 85%	100% of Table Load per side.	Reduce-allowable table loads using Straight-Line Interpolation.
ORDERING	Add SK, angle required and right (R) or left (L), to product number. Ex. PH1795-SK45R	Add SL, slope required, and up (U) or down (D), to product number. Ex. PH1795-SL30D	See sloped and skewed. Ex. PH1795-SK45RSL30D	Add OS offset required, and right (R) or left (L), to product number. Ex. PH1795-OSL	Add SA, and saddle width required to product number. Ex. PH1795-SA = $5^1/_2''$	Add SF, angle required, and right (R) or left (L), to product number. Ex. PH1795-SF30L

Sloped/skewed hanger with skews greater than 15° may have all joist nailing on outside flange. Sloped or sloped/skewed hangers with slopes greater than 15° may have additional joist nails added.
All sloped, skewed or combinations require bevel cut on joist in all applications and bearing stiffeners with I-joists.

LBH and HLBH Series Options

Option	*Skewed*	*Sloped*	*Sloped/Skewed*	*Top Flange Offset*	*Saddle*
RANGE	1° to 50°	1° to 45°	skew 1° to 45° sloped 1° to 45°	—	—
ALLOWABLE LOADS	LBH - 4725 lbs. max. HLBH - 8070 lbs. max. LBH / HLBH - 50% of uplift load on skew greater than 15°.	LBH - 4110 lbs. max. HLBH - 7000 lbs. max.	LBH - 3900 lbs. max. HLBH - 6650 lbs. max. LBH/HLBH- 50% of uplift load on skew greater than 15°.	60% of table load for LBH. 45% of table load for HLBH.	100% of table load per side.
ORDERING	Add SK, angle required, and right (R) or left (L), to product number. Ex. LBH3595-SK 45R	Add SL, slope required, and up (U) or down (D), to product number. Ex. LBH3595-SL30D	See sloped and skewed. Ex. LBH3595-SK45RSL30D	Add OS, offset required, and right (R) or left (L), to product number. Ex. LBH3595-OSL	Add SA, and saddle width required to product number. Ex. LBH3595-SA=5 1/2″

Sloped / skewed hanger with skews greater than 15° may have all joist nailing on outside flange. Sloped or sloped / skewed hangers with slopes greater than 15° may have additional joist nails added. All sloped, skewed or combinations require bevel cut on joist in all applications and bearing stiffeners with I-joists.

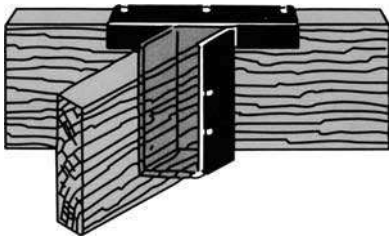

Left **Skewed** Seat

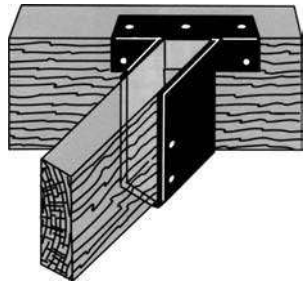

Down **Sloped**
& Left **Skewed** Seat

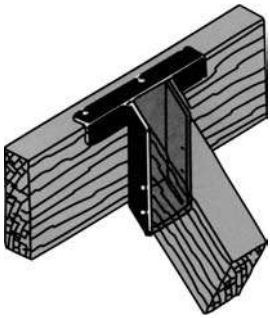

Down **Sloped** Seat

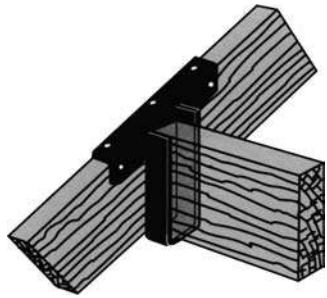

Sloped Top Flange
Down Left, Low Point

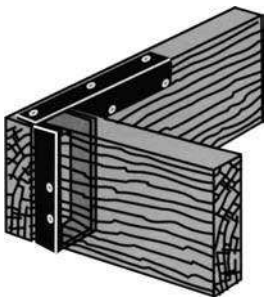

Right **Top Flange Offset**

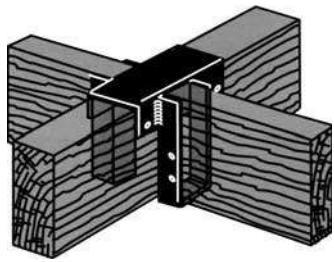

Saddle

Face Mount Hangers HD and HHD Series

These HD/HHD models are specifically designed to support LVL, LSL and PSL beams and headers in medium load conditions. The HD's slant-back design is strong but economical. Both the HD and HHD have open backs, allowing builders to install the hanger after the header or joists are in place. An extra-wide nailing flange also allows for easy field installation.

Finish ..... G60 galvanizing
Code Listing ..... NER 478
Specialty Options ..... see page 27

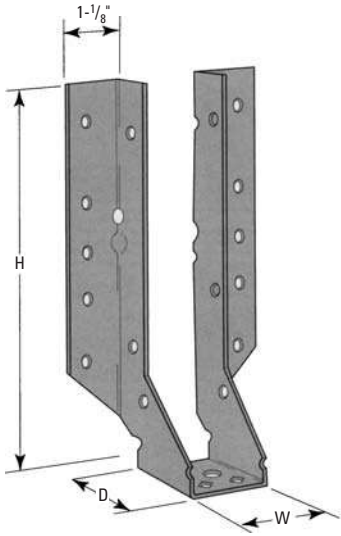

HD

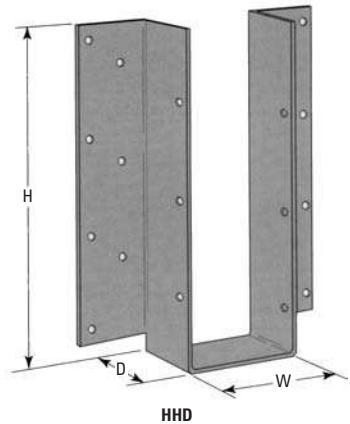

HHD

Note

NA20D nails included with all HHD series hangers.

Face Mount Hanger THF Series

The THF is specially designed to support the top chords of I-joists in depths up to 16 inches. This eliminates the need for web stiffeners in most applications. Reinforcement ribs give the THF added strength. Raised dimple nailing guides help assure correct 45° nailing into the I-joist bottom chord. The new Seat Cleat™ prongs (patent applied for) allow builders to "pop" I-joists into hangers for quick, positive seating. Seat Cleats™ are not yet available for all THF models.

Finish ..... G60 galvanizing
Code Listings ..... NER 478
U.S. Patent ..... #5,217,317
Seat Cleat Patent ..... applied for
Specialty Options ..... see page 27

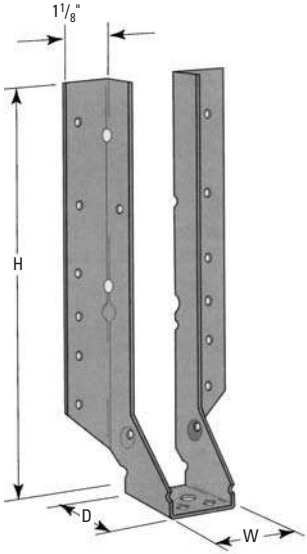

THF Single

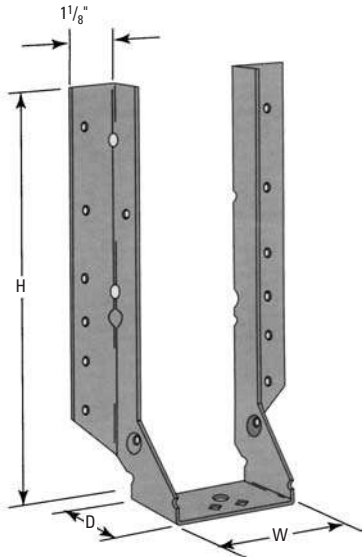

THF Double

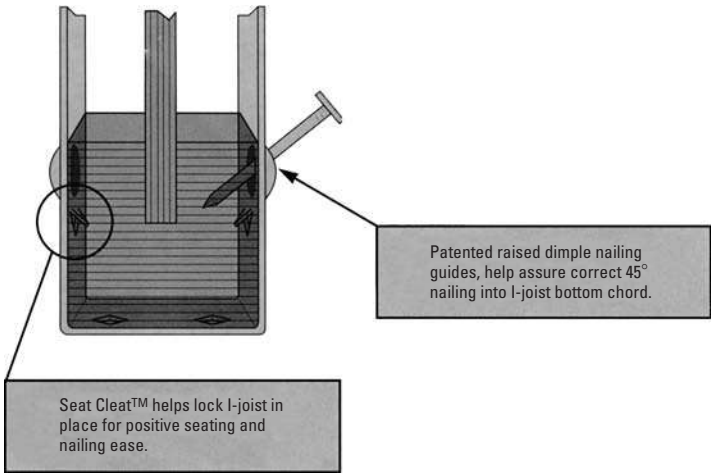

Heavy-Duty Hangers THD Series

Deep seats allow the THD to handle higher loads. The THD offers economical costs from an automated manufacturing process. This model features wide header flanges for easier fastening using standard nails.

Finish ..... G60 galvanizing
Code Listings ..... NER 478

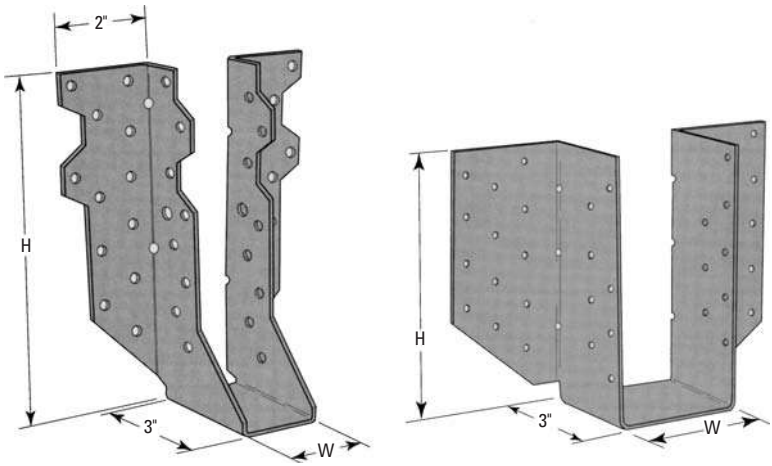

THD Configurations

Heavy-Duty Hangers THDH Series

Special slant nailing and deep seats allow the THDH to handle higher loads. This model features wide header flanges for easier fastening using standard nails.

Finish ..... G60 galvanizing
Code Listing ..... NER pending
U.S. Patent ..... #5,217,317

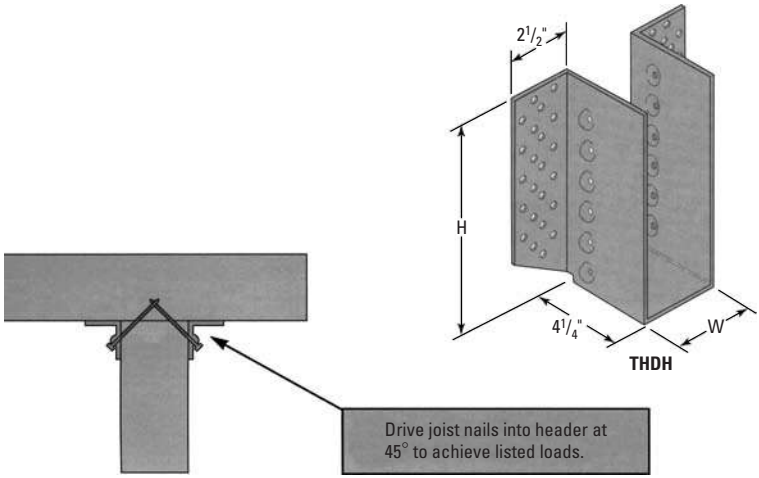

HD, HHD and THF Series Options

Option	*Skewed*	*Sloped*	*Sloped/Skewed*	*Inverted Flange*
RANGE	1° to 67½° when width is 1¾″ or less. 1° to 50° on all others.	1° to 45°	See sloped and skewed.	Not available in widths less than 2¼″.
ALLOWABLE LOADS	100% of table load. 75% of uplift load on skews greater than 15°.	100% of table load.	80% of table load. 75% of uplift load on skews greater than 15°.	100% of table load. 65% of table load when nailing into the support members end grain.
ORDERING	Add SK, angle required, and right (R) or left (L), to product number. Ex. HD17925-SK45R	Add SL, slope required, and up (U) or down (D), to product number. Ex. HD17925-SL30D	See sloped and skewed. Ex. HD17925-SK45RSL30D	Add IF, to product number. Ex. HD410-IF

Sloped/skewed hanger with skews greater than 15° may have all joist nailing on outside flange. Sloped or sloped/skewed hangers with slopes greater than 15° may have additional joist nails added. All sloped, skewed or combinations require bevel cut on joist in all applications and bearing stiffeners with I-joists.

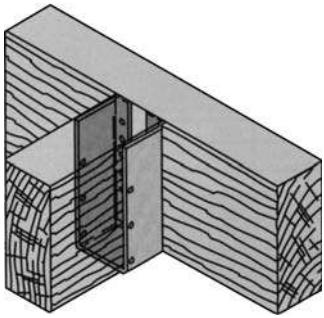

Inverted Flange

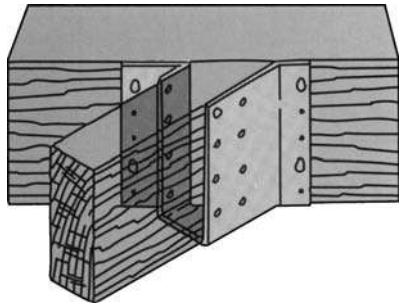

Left **Skewed** Seat

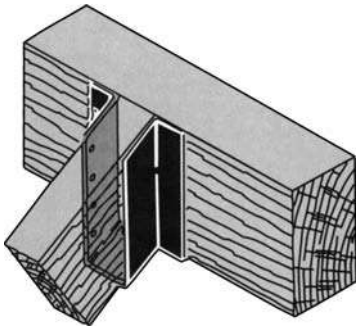

Down **Sloped** Seat

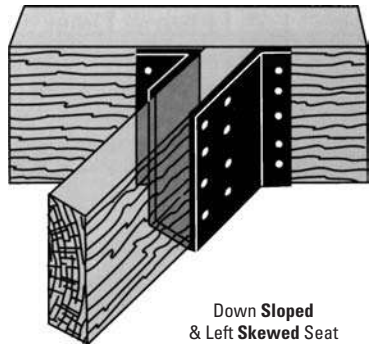

Down **Sloped**
& Left **Skewed** Seat

Knee Braces—KVB Series

USP Stock No.	Ref. No.	Dimensions D	W	L	Fastener Schedule Beam	Joist	Allowable Tension Loads (LBS) 100%	115%	125%	133%	160%
KVB5	VB5	10–15 Beam Depth	3 1/4	28 5/16	(4) NA25	(12) NA25	1050	1210	1315	1400	1680
KVB7	VB7	15–22 Beam Depth	5 1/4	39 5/16	(6) NA25	(12) NA25	1050	1210	1315	1400	1680
KVB8	VB8	22 1/2–28 1/2 Beam Depth	5 1/4	45 5/16	(6) NA25	(12) NA25	1050	1210	1315	1400	1680
KVB10	VB10	28 1/2–36 Beam Depth	6 7/8	56 5/16	(6) NA25	(12) NA25	1050	1210	1315	1400	1680
KVB12	VB12	36–42 Beam Depth	6 7/8	68 5/16	(6) NA25	(12) NA25	1050	1210	1315	1400	1680
KVBI5	—	10–15 Beam Depth	3 1/2	28 9/16	(4) 10d	(12) 10d	475	545	595	635	760
KVBI7	—	15–22 Beam Depth	5 1/4	39 5/16	(6) 10d	(12) 10d	715	820	895	950	1140
KVBI8	—	22 1/2–28 1/2 Beam Depth	5 1/4	45 5/16	(6) 10d	(12) 10d	715	820	895	950	1140
KVBI10	—	28 1/2–36 Beam Depth	7	56 7/16	(6) 10d	(12) 10d	715	820	895	950	1140
KVBI12	—	36–42 Beam Depth	7	68 7/16	(6) 10d	(12) 10d	715	820	895	950	1140

NA25 nails are included with the braces.

Knee Braces—KVB and KVBI Series

Knee Braces restrict lateral movement of beam under seismic stress. The KVB can be retrofitted into existing framing. NA25 nails are shipped with KVB Knee Brace and arrive attached to the connector in convenient poly bags.

The new KVBI Knee Brace is designed to be used with I-joist purlins and installs with 10d nails (not supplied).

Finish ..... G60 galvanizing
Material ..... 12 gauge steel
Code Listings ..... ICBO 2725
..... L.A. City RR25104

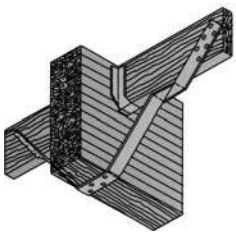

Typical **KVB** Installation

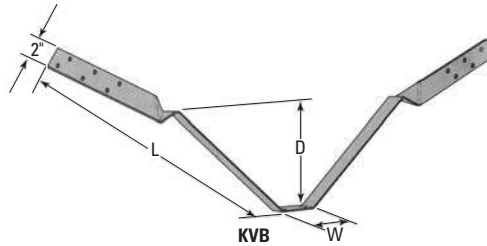

Hinge Connectors—KHC Series

USP Stock No.	Ref. No.	Beam Width	Dimensions Min W	H	PD	T	Bolt Schedule Carrying Member	Carried Member	Allowable Loads (LBS) 560 PSI	660 PSI
KHC55	HCA5-5	5 1/8	5 1/4	12	5	3/4	(2) 3/4	(2) 3/4	14415	16655
KHC56	HCA5-6	5 1/8	5 1/4	17	6	3/4	(2) 3/4	(2) 3/4	17300	19990
KHC57	HCA5-7	5 1/8	5 1/4	23 1/2	7	3/4	(2) 3/4	(2) 3/4	20180	23320
KHC59	HCA5-9	5 1/8	5 1/4	38 1/2	9	3/4	(2) 3/4	(2) 3/4	25945	29980
KHC525	—-	5 1/4	5 3/8	12	5	3/4	(2) 3/4	(2) 3/4	14415	16655
KHC526	—-	5 1/4	5 3/8	17	6	3/4	(2) 3/4	(2) 3/4	17300	19990
KHC527	—-	5 1/4	5 3/8	23 1/2	7	3/4	(2) 3/4	(2) 3/4	20180	23320
KHC529	—-	5 1/4	5 3/8	38 1/2	9	3/4	(2) 3/4	(2) 3/4	25945	29980
KHC75	HCA7-5	6 3/4	6 7/8	13	5	1	(2) 3/4	(2) 3/4	18985	21940
KHC76	HCA7-6	6 3/4	6 7/8	19	6	1	(2) 3/4	(2) 3/4	22780	26325
KHC77	HCA7-7	6 3/4	6 7/8	25 1/2	7	1	(2) 3/4	(2) 3/4	26580	30715
KHC79	HCA7-9	6 3/4	6 7/8	43 1/2	9	1	(2) 3/4	(2) 3/4	34170	39490
KHC725	—-	7	7 1/8	13	5	1	(2) 3/4	(2) 3/4	18985	21940
KHC726	—-	7	7 1/8	19	6	1	(2) 3/4	(2) 3/4	22780	26325
KHC727	—-	7	7 1/8	25 1/2	7	1	(2) 3/4	(2) 3/4	26580	30715
KHC729	—-	7	7 1/8	42 1/2	9	1	(2) 3/4	(2) 3/4	34170	39490
KHC95	HCA9-5	8 3/4	8 7/8	13	5	1 1/4	(2) 3/4	(2) 3/4	24610	28440
KHC96	HCA9-6	8 3/4	8 7/8	19	6	1 1/4	(2) 3/4	(2) 3/4	29530	34125
KHC97	HCA9-7	8 3/4	8 7/8	25 1/2	7	1 1/4	(2) 3/4	(2) 3/4	34450	39815
KHC99	HCA9-9	8 3/4	8 7/8	42 1/2	9	1 1/4	(2) 3/4	(2) 3/4	44300	51190

The minimum height is for loads shown. For heights less than the minimum shown reduce the allowable loads in direct proportion.
All bolts are 3/4 inch, and shall meet or exceed the specifications of ASTM A307.
KHC Allowable Loads are normal loads; no increase is permitted.

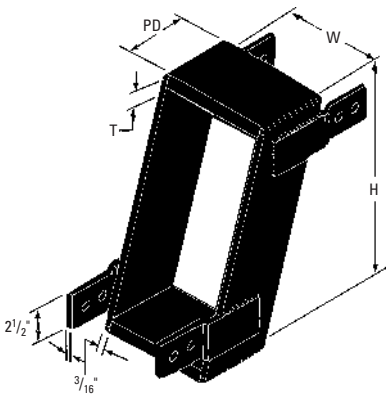

KHC

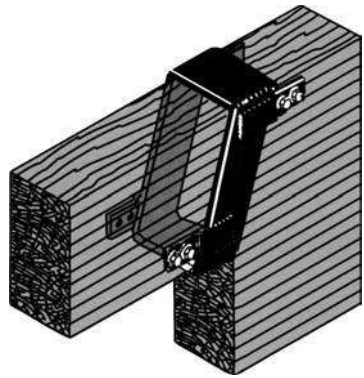

Typical **KHC** Installation

Seismic Straps—KHCST Series

USP Stock No.	Ref. No.	Dimensions L	Dimensions W	Bolt Schedule	Allowable Loads (LBS.) 133%	Allowable Loads (LBS.) 160%
KHCST2	HCST2	22	$3^1/_2$	$(4)^3/_4$	6065	6065
KHCST3	HCST3	28	$3^1/_2$	$(6)^3/_4$	8815	8815
KHCST4	HCST4	34	$3^1/_2$	$(8)^3/_4$	11225	11225

Allowable loads are for straps used in pairs, and are increased 33% or 60% for wind or seismic loads; no further increase is permitted.
Assume a minimum wood thickness of $3^1/_8$ inches.
Seismic straps shall be used with the KHC hinge connectors.
To order KHCST straps with round holes, add an "R" to the part number, as in "KHCSTR2".

Hinge Connectors KHC Series
Seismic Straps KHCST Series

KHC connectors are used to connect ends of glulam beams having equal widths. These connectors allow long continuous spans, eliminating unnecessary columns. Special dimension variations and models with seismic tabs are available upon request. KHCST straps can be installed during construction or added as a retrofit item. The strap's shape is specially designed to accommodate installation over a USP hinge connector. Order one strap per side.

Material, KHC series .....
..... 7 gauge $^3/_4$″, 1″ & $1^1/_4$″ steel
Material, KHCST series ..... 3 gauge steel
Finish..... paint
Code Listings..... ICBO 2725
.....L.A. City RR25104

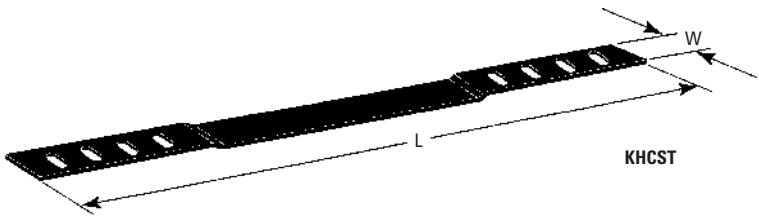

Skewed 45° Hangers SKH Series

The SKH series provides stock face-mount hangers at preset 45-degree skews, right or left. The hanger design allows for some installation flexibility—it will accommodate a 40- to 50-degree skew. All models provide uplift capacity and some eliminate the need for miter cuts. A convenient slant nail design provides time savings and ensures proper installation.

Finish ..... G60 galvanizing
Code Listing ..... NER 478

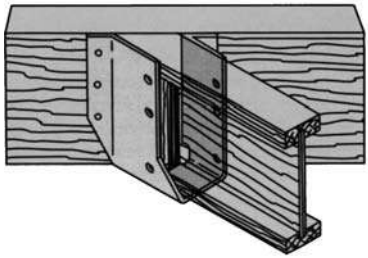

Typical **SKH** Installation

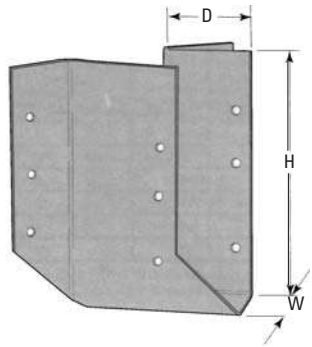

SKH

Right Skew

Left Skew

Adjustable Strap Hangers—MSH Series

The MSH is field adjustable; builders can use the flanges in overhang, face mount or combination style. The wide header strap allows for a double nailing row and higher load values. An open-back design allows installation after a member is placed in position.

Finish ..... G60 galvanizing
Code Listings ..... SBCCI 9494
..... NER pending

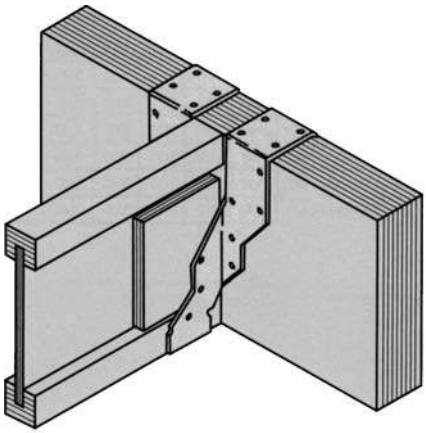

MSH Top Mount Max

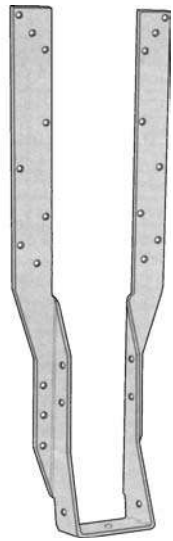

MSH422

Adjustable Rafter-to-Plate Connectors—TMP Series

								Allowable Loads (LBS.)				
								Floor	Roof		Uplift	
				Dimensions		Fastener Schedule			Snow	Non-Snow		
Refter Width (Inches)	USP Stock No.	Ref. No.	Steel Gauge	W	L	Plate	Rafter	100%	115%	125%	133%	160%
$1^1/_2$	TMP2	VPA2	18	$1^9/_{16}$	$5^9/_{16}$	(6)10d	(4)10d × $1^1/_2$	990	990	990	220	220
$1^3/_4$	TMP175	VPA25	18	$1^{13}/_{16}$	$5^9/_{16}$	(6)10d	(4)10d × $1^1/_2$	1150	1150	1150	220	220
$2^1/_3$	TMP23	VPA35	18	$2^3/_8$	$6^3/_8$	(6)10d	(4)10d × $1^1/_2$	1785	1970	1970	220	220
$2^1/_2$ or $2^5/_8$	TMP25	VPA2.68	18	$2^{11}/_{16}$	$6^3/_8$	(6)10d	(4)10d × $1^1/_2$	1825	1970	1970	220	220
3	TMP31	—	18	$3^1/_8$	$7^5/_{16}$	(6)10d	(4)10d × $1^1/_2$	1955	1970	1970	220	220
$3^1/_2$	TMP4	VPA4	18	$3^9/_{16}$	$7^5/_{16}$	(6)10d	(4)10d × $1^1/_2$	1970	1970	1970	220	220

No increase for duration of load for uplift is permitted; loads governed by test results.
10d × $1^1/_2$ nails are 9 gauge (0.148 inch diameter) by $1^1/_2$ inches long.

Adjustable Rafter-to-Plate Connectors TMP Series

With the TMP, builders can make rafter-to-plate connections at pitches from $^{1}/_{12}$ to $^{6}/_{12}$. This model automatically adjusts to any pitch within its range. There's no need to preset the pitches for every connector, and there's no need for time-consuming bird's-mouth notching or bevel plate installation. This model is available in the popular I-joist sizes.

Finish ..... G60 galvanizing
Code Listing ..... NER 478
U.S. Patent ..... # 5,217,317

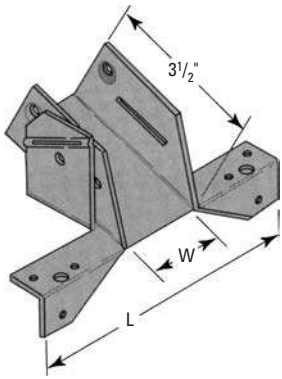

TMP

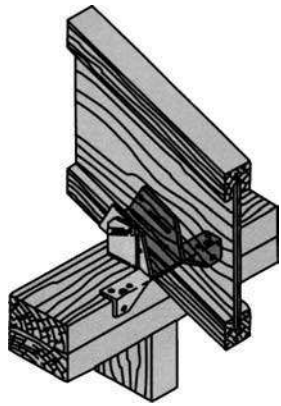

Typical **TMP** Installation

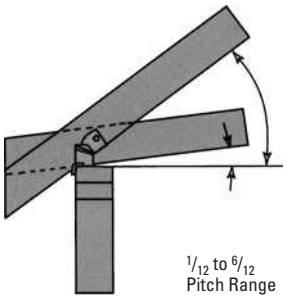

Adjustable Rafter-to-Plate Connectors—TMPH Series

Rafter Width (Inches)	USP Stock No.	Ref. No.	Dimensions W	Dimensions L	Dimensions H	Fastener Schedule Plate	Fastener Schedule Rafter	Allowable Loads (LBS.) According to Pitch 6/12	7/12	8/12	9/12	10/12	11/12	12/12	13/12	14/12	Uplift 133%	Uplift 160%
1 1/2	TMPH2	—	1 9/16	6 9/16	2 1/2	(10) 10d	(8) 10d × 1 1/2	3190	3170	3110	3000	2840	2640	2420	2180	1945	200	200
1 3/4	TMPH175	—	1 13/16	6 9/16	2 3/8	(10) 10d	(8) 10d × 1 1/2	3190	3170	3110	3000	2840	2640	2420	2180	1945	200	200
2 5/16	TMPH23	—	2 3/8	7 3/8	2 1/2	(10) 10d	(8) 10d × 1 1/2	3190	3170	3110	3000	2840	2640	2420	2180	1945	200	200
2 1/2	TMPH25	—	2 11/16	7 3/8	2 5/16	(10) 10d	(8) 10d × 1 1/2	3190	3170	3110	3000	2840	2640	2420	2180	1945	200	200
3	TMPH31	—	3 1/8	8 9/16	2 11/16	(10) 10d	(8) 10d × 1 1/2	3190	3170	3110	3000	2840	2640	2420	2180	1945	200	200
3 1/2	TMPH4	—	3 9/16	8 9/16	2 1/2	(10) 10d	(8) 10d × 1 1/2	3190	3170	3110	3000	2840	2640	2420	2180	1945	200	200

Web stiffeners are required for all wood I-joist installations.
Design Loads may not be increased.
10d × 1 1/2 inch nails are 9 gauge (0.148 inch diameter) by 1 1/2 inches long.

Adjustable Rafter-to-Ridge Connectors TMPH Series

With the TMPH, builders can make rafter-to-plate connections at pitches from $^{6}/_{12}$ to $^{14}/_{12}$. This model automatically adjust to any pitch within their ranges. There's no need to preset the pitches for every connector, and there's no need for time-consuming bird's-mouth notching or bevel plate installation. This model is available in the popular I-joist sizes.

Finish..... G60 galvanizing
Code Listing ..... NER pending
U.S. Patent ..... #5,230,198

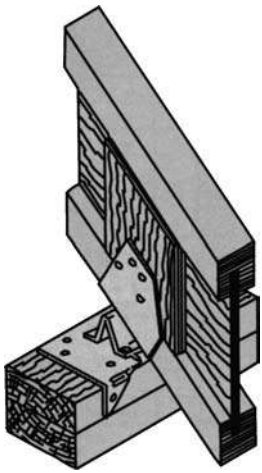

Typical **TMPH** Installation

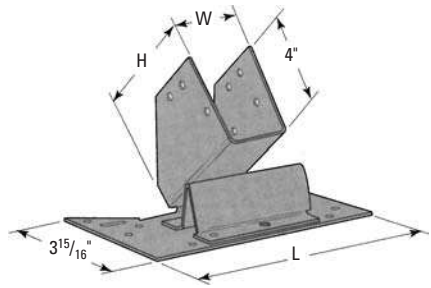

TMPH

TMP and TMPH Installation Procedure:

A) Nail clip to outside edge of wall plate.

B) Insert joist with downward pressure to seat the self-adjusting support.

C) Nail joist in place being sure to nail through the matching slots.

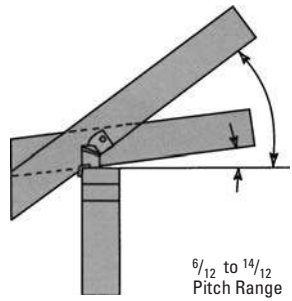

Adjustable Rafter-to-Ridge Connectors—TMU Series

Rafter Width (Inches)	USP Stock No.	Ref. No.	Steel Gauge	Dimensions W	Dimensions H	Dimensions L	Fastener Schedule Header	Fastener Schedule Joist	Allowable Loads (LBS.) Vertical Loads Floor 100%	Vertical Loads Roof Snow 115%	Vertical Loads Roof Non-Snow 125%	Uplift 133%	Uplift 160%
$1\frac{1}{2}$	TMU26	LSU26	18	$1\frac{9}{16}$	$5\frac{1}{4}$	$5\frac{1}{4}$	(10) 10d	(7) 10d × $1\frac{1}{2}$	1120	1290	1305	640	640
$1\frac{1}{2}$	TMU210	LSSU210	18	$1\frac{9}{16}$	$8\frac{5}{8}$	$5\frac{1}{4}$	(12) 10d	(11) 10d × $1\frac{1}{2}$	1340	1340	1340	770	770
$1\frac{3}{4}$	TMU175	—	18	$1\frac{13}{16}$	$5\frac{1}{4}$	$5\frac{1}{4}$	(10) 10d	(7) 10d × $1\frac{1}{2}$	1120	1290	1305	640	640
$1\frac{3}{4}$	TMU179	LSSUI25	18	$1\frac{13}{16}$	$8\frac{5}{8}$	$5\frac{1}{4}$	(12) 10d	(11) 10d × $1\frac{1}{2}$	1340	1340	1340	770	770
$2\frac{5}{16}$	TMU23	LSSUI35	18	$2\frac{3}{8}$	$8\frac{7}{8}$	$6\frac{3}{4}$	(12) 10d	(11) 10d × $1\frac{1}{2}$	1345	1545	1680	1340	1605
$2\frac{1}{2}$	TMU25	LSSUI2.68	18	$2\frac{11}{16}$	$8\frac{7}{8}$	$6\frac{3}{4}$	(12) 10d	(11) 10d × $1\frac{1}{2}$	1345	1545	1680	1340	1605
3	TMU31	LSSUI210-2	18	$3\frac{1}{8}$	7	8	(12) 10d	(11) 10d × $1\frac{1}{2}$	1345	1545	1680	1340	1605
$3\frac{1}{2}$	TMU48	LSSU410	18	$3\frac{9}{16}$	7	8	(12) 10d	(11) 10d × $1\frac{1}{2}$	1345	1545	1680	1340	1605

Uplift Loads have been increased $33\frac{1}{3}$% or 60% for wind and seismic loads; no further increases shall be permitted.
10d × $1\frac{1}{2}$ nails are 9 gauge (0.148 inch diameter) × $1\frac{1}{2}$ inches long.
Web stiffeners are required for all Wood I-joist installations.
A tie strap is required for all installations when sloped down at $\frac{7}{12}$ to $\frac{12}{12}$ pitch-use LSTI22 with (4) 10d × $1\frac{1}{2}$ nails.
per end for TMU26, TMU210, TMU175 and TMU179, use KSTI223 with (5) 10d × $1\frac{1}{2}$ nails per end for all others.
10d nails require a header thickness of 1.76 inches to achieve load values shown.
If attached to 2 × lumber, use 85% of listed values.

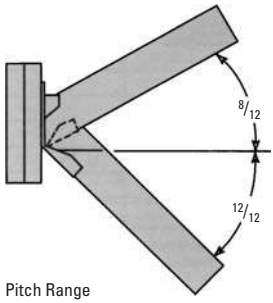

Pitch Range

Adjustable Rafter-to-Ridge Connectors TMU Series

The TMU is designed to connect rafters to ridge beams in vaulted roof structures. These connectors were originally developed to help builders avoid ordering custom sloped and skewed hangers for every situation. The TMU is easily field adjusted to meet a variety of skew and / or slope applications. Shapes to any pitch from 8/12 up through 12/12 down.

Finish ..... G60 galvanizing
Code Listing ..... NER 478
U.S. Patent ..... # 5,217,317

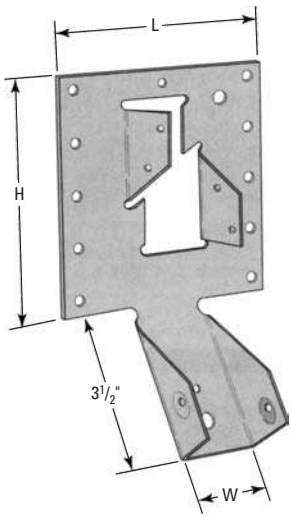

TMU

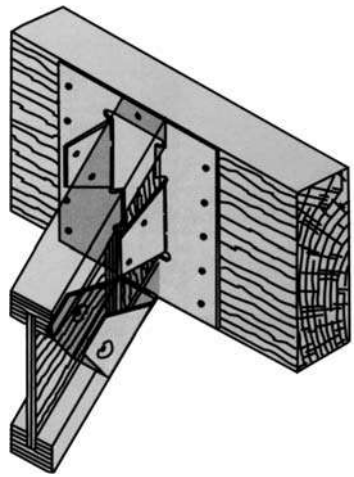

Typical **TMU** Installation

Skewed Installation

1. Install onto end of rafter using recommended nail schedule for joist. Note: When sloping TMU26/175 up, attach hanger to web before nailing to bottom flange.
2. Bend flange to desired angle, and nail in place.
3. Bend opposite flange as shown, and finish nailing.

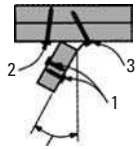

SKEW TO 45°

TMU26, 210, 175 & 179

3/4"×4"×7" Plywood Spacer with (6) 16d Nails

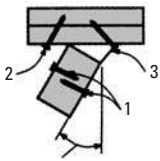

SKEW TO 30°

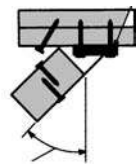

SKEW 31° TO 45°

TMU23, 25 ,31 & 48

TMU Installation Procedure

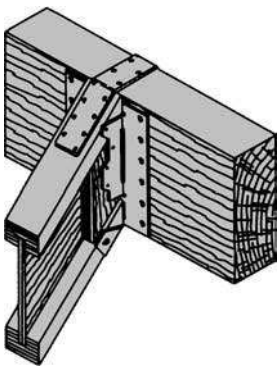

Web stiffeners required for all wood I-joist installations. Tie strap required for all wood I-joist installations when sloped 30° to 45°, or 7:12 thru 12:12. Use LSTI-22 for $1^1/_2''$ & $1^3/_4''$ width joists, use KSTI-223 for $2^5/_{16}''$ or wider joists.

Strap Ties—KSTI and LST122 Series

USP Stock No.	Ref. No.	Steel Gauge	Dimensions W	Dimensions L	Nail Schedule	Allowable Loads (LBS.) Nails 133%	Allowable Loads (LBS.) Nails 160%
KSTI223	MSTI23	18	$2^1/_{16}$	23	(23) 10d × $1^1/_2$	1645	1975
KSTI226	MSTI26	12	$2^1/_{16}$	26	(26) 10d × $1^1/_2$	1735	2080
KSTI236	MSTI36	12	$2^1/_{16}$	36	(36) 10d × $1^1/_2$	2400	2880
KSTI248	MSTI48	12	$2^1/_{16}$	48	(48) 10d × $1^1/_2$	2300	3840
KSTI260	MSTI60	12	$2^1/_{16}$	60	(60) 10d × $1^1/_2$	4000	4800
KSTI272	MSTI72	12	$2^1/_{16}$	72	(72) 10d × $1^1/_2$	4510	4950
LSTI22	—	20	$1^1/_4$	24	(16) 10d × $1^1/_2$	950	950

Allowable Loads have been increased $33^1/_3$% or 60% for wind and seismic loads; no further increase shall be permitted. 10d × $1^1/_2$ nails are 9 gauge (0.148 inch diameter) by $1^1/_2$ inches long.

Strap Ties KSTI and LSTI22 Series

These strap ties are designed specifically for installation along wood I-joist flanges.

Installation Note

The LSTI22 is for use with 1-$^1/_2$″ and 1-$^3/_4$″ I-joists. The KSTI is for use with $2^5/_{16}$″ and larger I-joists.

Finish ..... G60 galvanizing
Code Listings ..... SBCCI 9494
..... NER pending

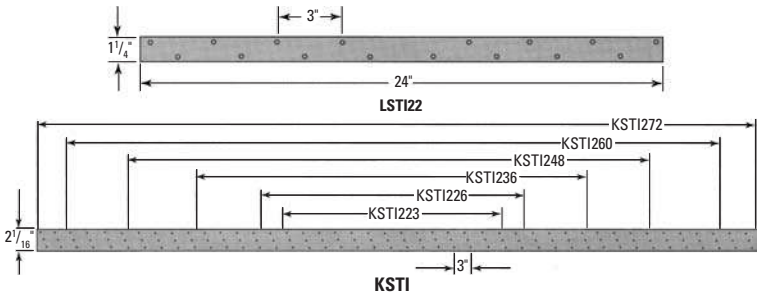

Strap Ties—LSTA and MSTA Series

USP Stock No.	Ref. No.	Steel Gauge	Dimensions W	Dimensions L	*10d Nails Per End	10d Nails Allowable Loads (LBS.) 100%	133%	160%	*10d × 1 1/2 Nails Per End	10d × 1 1/2 Nails Allowable Loads (LBS.) 100%	133%	160%
LSTA9	LSTA9	20	1 1/4	9	4	445	590	710	4	375	500	600
LSTA12	LSTA12	20	1 1/4	12	5	555	740	800	5	470	625	750
LSTA15	LSTA15	20	1 1/4	15	6	670	895	945	6	570	760	910
LSTA18	LSTA18	20	1 1/4	18	7	790	1055	1200	7	665	885	1065
LSTA21	LSTA21	20	1 1/4	21	8	900	1200	1200	8	760	1015	1200
LSTA24	LSTA24	18	1 1/4	24	9	1025	1370	1605	9	865	1150	1380
LSTA30	LSTA30	18	1 1/4	30	11	1200	1605	1605	11	1055	1410	1605
LSTA36	LSTA36	18	1 1/4	36	11	1200	1605	1605	13	1200	1605	1605
MSTA9	MSTA9	18	1 1/4	9	4	445	590	710	4	375	500	600
MSTA12	MSTA12	18	1 1/4	12	5	555	740	890	5	470	625	750
MSTA15	MSTA15	18	1 1/4	15	6	665	890	1065	6	565	750	900
MSTA18	MSTA18	18	1 1/4	18	7	775	1035	1070	7	660	875	1055
MSTA21	MSTA21	18	1 1/4	21	8	895	1195	1260	8	760	1015	1215
MSTA24	MSTA24	16	1 1/4	24	9	1035	1380	1655	9	875	1165	1395
MSTA30	MSTA30	16	1 1/4	30	11	1265	1685	2000	11	1065	1425	1705
MSTA36	MSTA36	16	1 1/4	36	13	1495	1995	2000	13	1260	1680	2000
MSTA48	MSTA48	16	1 1/4	48	14	1500	2000	2000	16	1500	2000	2000

Loads have been increased 33 1/3% or 60% for wind and seismic loads; no further increase shall be permitted.
When 10d nails are used, the minimum wood thickness shall be 1 3/4 inches.
If used on 2× lumber, use the table for 10d× 1 1/2 inch nails.
*Minimum number of fasteners needed in each end to develop strap strength.

Strap Ties LSTA and MSTA Series

With a $1^1/_4$ inch width, the 20-gauge LSTA and the heavier MSTA straps are designed especially for use with $1^1/_2$ inch members. The lighter gauge allows builders to bend these straps for many framing applications.

Finish ..... G60 galvanizing
Code Listing ..... NER 505

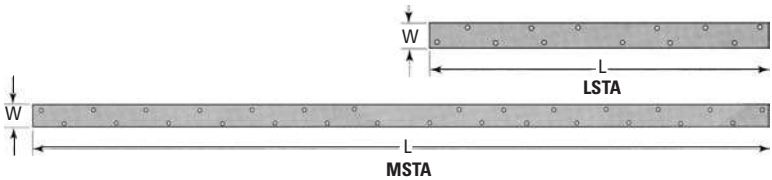

Purlin Anchors PAI Series

Purlin Anchors are engineered to meet seismic and high wind load requirements when securing I-joist purlins and beams to concrete or masonry. Loads are based on a minimum 4-inch embedment into concrete and should be reduced accordingly for masonry or other applications.

Finish ..... G60 galvanizing
Code Listing ..... NER 505

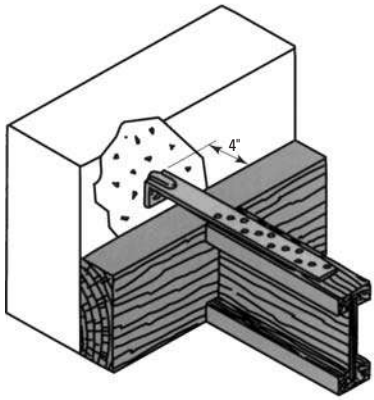

Typical **PAI** Installation

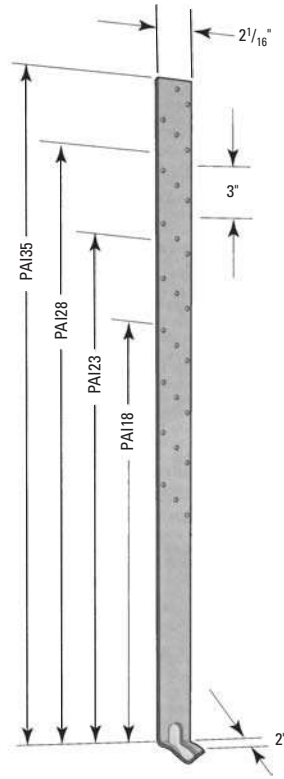

PAI

USP Stock No.	Ref. No.	Dimensions	Steel Gauge	Fastener Schedule Nails	Fastener Schedule Bolts	Minimum EMBED.	Allowable Loads (LBS.) Nails 133%	Allowable Loads (LBS.) Nails 160%
PAI18	PAI18	2 × 18 1/2	12	(12) 10d × 1 1/2	—	4	1600	1920
PAI23	PAI23	2 × 23 1/2	12	(18) 10d × 1 1/2	—	4	2400	2880
PAI28	PAI28	2 × 28 1/2	12	(24) 10d × 1 1/2	—	4	3125	3125
PAI35	PAI35	2 × 35 1/2	12	(26) 10d × 1 1/2	—	4	3125	3125

Allowable Loads have been increased 33 1/3% or 60% for wind and seismic loads; no further increase shall be permitted.
10d × 1 1/2 nails are 9 gauge (0.148 inch diameter) by 1 1/2 inches long.

Bracing—N Series

Sized for Joist to Plate Lateral Bracing				
I-Beam Height	*Joist Spacing Inches			
	12	16	19.2	24
9 1/4	N16	N16	N16	N27
9 1/2	N16	N16	N16	N27
11 1/4	N16	N16	N27	N27
11 7/8	N16	N27	N27	N27
12	N16	N27	N27	N27
14	N16	N27	N27	N30
16	N27	N27	N27	N30
18	N27	N27	N27	N30/N36
20	N27	N27	N27/N30	N36
22	N27	N30	N30	N36
24	N30	N30	N30/N36	N36
26	N30	N36	N36	N36
28	N36	N36	N36	N42
30	N36	N36	N36/N42	N42

*Requires (2) 8d common nails or NA11 nails at each end. Nails are to be driven into narrow face of each joist unless noted otherwise.

Sized for Joist to Joist X Bridging				
I-Beam Height	*Joist Spacing Inches			
	12	16	19.2	24
9 1/4	N16	N16	N27	N27
9 1/2	N16	N16	N27	N27
11 1/4	N16	N27	N27	N27/N30
11 7/8	N16	N27	N27	N30
12	N16	N27	N27	N30
14	N16	N27	N27	N30
16	N27	N27	N27	N30
18	N27	N27	N27/N30	N30/N36
20	N27	N27/N30	N30	N36
22	N27/N30	N30	N30/N36	N36
24	N30	N30	N36	N36
26	N36	N36	N36	N42
28	N36	N36	N36/N42	N42
30	N36	N36/N42	N42	N42

*Requires (2) 8d common nails or NA11 nails at each end. Nails are to be driven into top/bottom face of each joist unless noted otherwise.

Bracing and Bridging N Series

N series bridging is quickly installed and economical. This style must be installed prior to the subfloor. Use (2) 8d nails at each end and leave a slight space between the units to avoid noise generating contact.

Material ..... 20 gauge steel
Finish ..... G60 galvanizing
Code Listing ..... NER 505

N 16

Typical **N** Bracing Installation

Wet Anchors—WE and WA Series

USP Stock No.	Ref. No.	Post Size	Dimensions W	Dimensions H1	Dimensions H2	Dimensions L	Fastener Schedule Nails	Fastener Schedule Bolts	Allowable Loads (LBS.) L1 Nails 133%	L1 Nails 160%	L1 Bolts 133%	L1 Bolts 160%	L2 Nails 133%	L2 Nails 160%	L2 Bolts 133%	L2 Bolts 160%	Uplift Nails 133%	Uplift Nails 160%	Uplift Bolts 133%	Uplift Bolts 160%
WE44	PB44	4 × 4	$3^{9}/_{16}$	$4^{13}/_{16}$	$3^{3}/_{8}$	$3^{1}/_{4}$	(12) 16d	$(2)^{1}/_{2}$	1135	1135	910	910	1180	1180	1020	1020	1405	1405	1430	1430
WE46	PB46	4 × 6	$5^{1}/_{2}$	$4^{5}/_{8}$	$3^{1}/_{4}$	$3^{3}/_{8}$	(12) 16d	$(2)^{1}/_{2}$	1135	1135	910	910	1180	1180	1020	1020	1405	1405	1430	1430
WA44	—	4 × 4	$3^{9}/_{16}$	4	5	2	(12) 16d	$(2)^{1}/_{2}$	1575	1575	1525	1830	675	675	1290	1290	1375	1375	1415	1695
WA45	—	4 × 6	$3^{9}/_{16}$	4	5	2	(12) 16d	$(2)^{1}/_{2}$	1575	1575	1525	1830	675	675	1290	1290	1375	1375	1415	1695
WA55	—	6 × 6	$5^{3}/_{8}$	6	5	$2^{1}/_{2}$	(12) 16d	$(2)^{1}/_{2}$	1575	1575	1525	1830	675	675	1290	1290	1375	1375	1415	1695
WA77	—	8 × 8	$7^{1}/_{8}$	$7^{1}/_{2}$	7	3	(12) 16d	$(2)^{1}/_{2}$	1575	1575	1525	1830	675	675	1290	1290	1375	1375	1415	1695

Uplift Loads have been increased $33^{1}/_{3}$% or 60% for wind and seismic loads; no further increase shall be permitted.
Not recommended for non-top-supported installations, such as fences.
H2 is minimum embedment length of anchor into concrete.
Bolt and nail values cannot be combined.

Wet Anchors WE and WA Series

Wet Post Anchors secure posts to concrete in light and medium duty applications such as patio covers, carports or decks. The built-in stand-off plate keeps the post from making contact with the ground and damaging moisture. The WE features a formed, one-piece design which offers additional economy. **Not recommended for fence post applications.**

Material ..... 12 gauge steel
Finish ..... G60 galvanizing
Code Listing, WA series ..... NER 505
Code Listings, WE series ..... ICBO 2039
..... NER pending
Code Listing, WE44..... SBCCI 9494

Typical **WE44** Installation

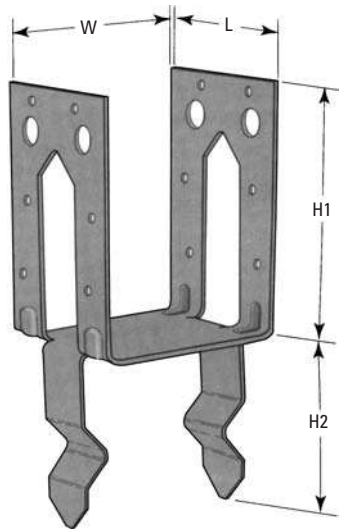

WE

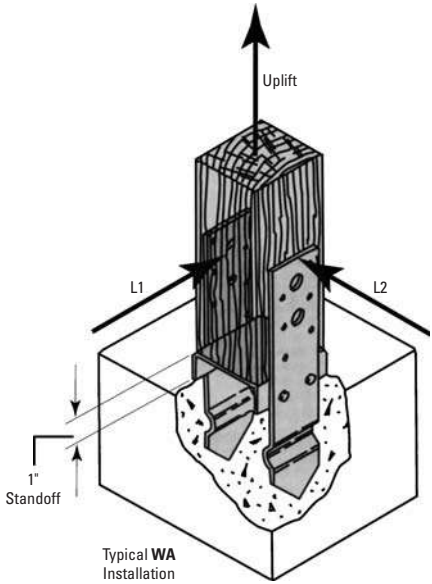

Typical **WA** Installation

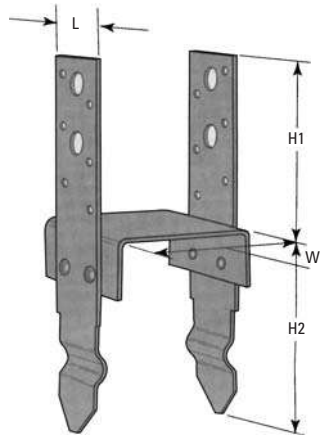

WA

Heavy Column Bases—HBM, CBE, and KCB Series

USP Stock No.	Ref. No.	Column Size	Dimensions W1	Dimensions W2	Dimensions H	Dimensions L	Fastener Schedule Bolts	Fastener Schedule Nails	Allowable Load (LBS.) Uplift Bolts 133%	Allowable Load (LBS.) Uplift Bolts 160%	Allowable Load (LBS.) Uplift Nails 133%	Allowable Load (LBS.) Uplift Nails 160%
HBM44	CBS44	$3^1/_2 \times 3^1/_2$	$3^9/_{16}$	$3^9/_{16}$	$7^5/_8$	2	$(2)^5/_8$	—	5650	4300	—	—
HBM46	CBS46	$3^1/_2 \times 5^1/_2$	$3^9/_{16}$	$5^1/_2$	$7^5/_8$	2	$(2)^5/_8$	—	5650	4300	—	—
HBM66	CBS66	$5^1/_2 \times 5^1/_2$	$5^1/_2$	$5^1/_2$	$7^5/_8$	2	$(2)^5/_8$	—	5640	4285	—	—
CBE44	LCB44	$3^1/_2 \times 3^1/_2$	$3^9/_{16}$	$3^1/_2$	$7^1/_2$	2	$(2)^1/_2$	(12)16d	3585	4300	2240	2690
CBE45	—	$3^1/_2 \times 5^1/_4$	$3^9/_{16}$	$5^3/_8$	$7^1/_2$	2	$(2)^1/_2$	(12)16d	3585	4300	2240	2690
CBE47	—	$3^1/_2 \times 7$	$3^9/_{16}$	$7^1/_8$	$7^1/_2$	2	$(2)^1/_2$	(12)16d	3585	4300	2240	2690
CBE55	—	$5^1/_4 \times 5^1/_4$	$5^1/_2$	$5^3/_8$	$7^1/_2$	2	$(2)^1/_2$	(12)16d	3570	4285	2240	2690
KCB44	CB44	$3^1/_2 \times 3^1/_2$	$3^9/_{16}$	$3^9/_{16}$	$8^7/_8$	2	$(2)^5/_8$	—	5650	6780	—	—
KCB45	—	$3^1/_2 \times 5^1/_4$	$3^9/_{16}$	$5^3/_8$	$8^7/_8$	2	$(2)^5/_8$	—	5650	6780	—	—
KCB47	—	$3^1/_2 \times 7$	$3^9/_{16}$	$7^1/_8$	$8^7/_8$	2	$(2)^5/_8$	—	5650	6780	—	—
KCB55	—	$5^1/_4 \times 5^1/_4$	$5^1/_2$	$5^3/_8$	$8^7/_8$	3	$(2)^5/_8$	—	5640	6670	—	—
KCB57	—	$5^1/_4 \times 7$	$5^1/_2$	$7^1/_8$	$8^7/_8$	3	$(2)^5/_8$	—	5640	6680	—	—
KCB77	—	7×7	$7^1/_2$	$7^1/_8$	$9^3/_4$	3	$(2)^3/_4$	—	8075	8185	—	—

Uplift Loads have been increased $33^1/_3$% or 60% for wind and seismic loads, no further increase shall be permitted.
Not recommended for fences.
Minimum edge distance is 3 inches.

Heavy Column Bases HBM, CBE, and KCB Series

These sturdy bases secure heavy columns to concrete footings or foundations. The base is set in position prior to the pour to assure proper alignment and installation. Designed for structural integrity in high wind or seismic applications. **These models are not recommended for fence post applications.**

Installation Note

For KCB installations, the base plate must be flush with the concrete surface to develop full bearing capacity.

Materials, HBM & CBE series . . . . . 12 gauge steel
Materials, KCB series . . . . . $^{3}/_{16}''$ or $^{1}/_{4}''$ steel
Finish, CBE series . . . . . G60 galvanizing
Finish, HBM & KCB series . . . . . paint
Code Listings, KCB & CBE series . . . . . SBCCI 9494
. . . . . NER pending
Code Listings, HBM . . . . . ICBO 2039
. . . . . L.A. City RR23888
Code Listings, KCB . . . . . ICBO 2725
. . . . . L.A. City RR25104

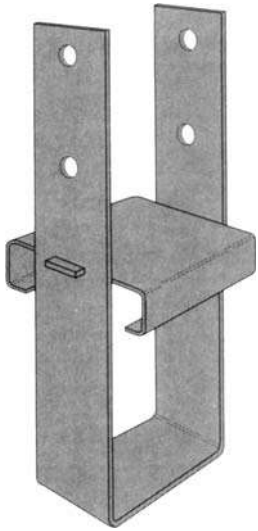

HBM

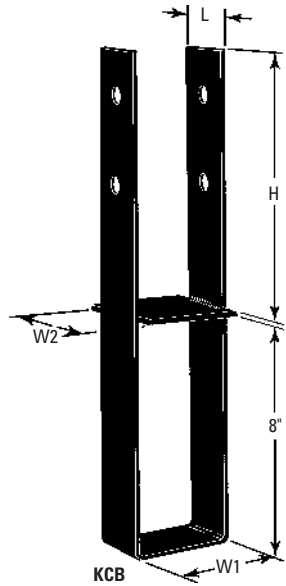

KCB

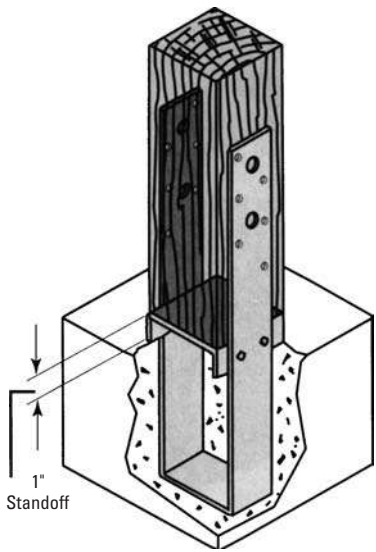

Typical **CBE** Installation

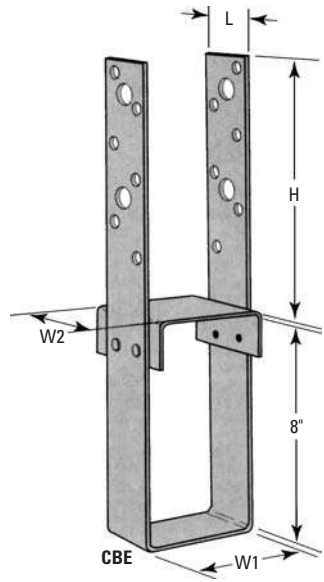

CBE

Post Caps—PCM and EPCM Series

USP Stock No.	Ref. No.	Post (Inches)	Beam (Inches)	Steel Gauge	Dimensions W1	W2	L1	L2	Fastener Schedule Post	Beam	Allowable Loads (LBS.) Lateral A1 133%	160%	Lateral A2 133%	160%	Uplift 133%	160%
PCM4416	PC44-16	3 1/2	3 1/2	16	3 9/16	3 9/16	3 1/16	11	(8)16d	(12)16d	1030	1030	835	835	1050	1050
PCM4516	—	3 1/2	5 1/4	16	3 9/16	5 3/8	3 1/16	13	(8)16d	(12)16d	1030	1030	835	835	1050	1050
PCM4716	—	3 1/2	7	16	3 9/16	7 1/8	3 1/16	15	(8)16d	(12)16d	1030	1030	835	835	1050	1050
PCM5416	—	5 1/4	3 1/2	16	5 3/8	3 9/16	3 1/2	11	(8)16d	(12)16d	1030	1030	835	835	1050	1050
PCM5516	—	5 1/4	5 1/4	16	5 3/8	5 3/8	3 1/2	13	(8)16d	(12)16d	1030	1030	835	835	1050	1050
PCM44	PC44	3 1/2	3 1/2	12	3 9/16	3 9/16	3 1/16	11	(8)16d	(12)16d	1930	1930	1495	1790	1050	1050
PCM45	—	3 1/2	5 1/4	12	3 9/16	5 3/8	3 1/16	13	(8)16d	(12)16d	1930	1930	1495	1790	1050	1050
PCM47	—	3 1/2	7	12	3 9/16	7 1/8	3 1/16	15	(8)16d	(12)16d	1930	1930	1495	1790	1395	1395
PCM54	—	5 1/4	3 1/2	12	5 3/8	3 9/16	3 1/2	11	(8)16d	(12)16d	1930	1930	1495	1790	1395	1395
PCM55	—	5 1/4	5 1/4	12	5 3/8	5 3/8	3 1/2	13	(8)16d	(12)16d	1930	1930	1495	1790	1395	1395
PCM57	—	5 1/4	7	12	5 3/8	7 1/8	3 1/2	15	(8)16d	(12)16d	1930	1930	1495	1790	1395	1395
PCM75	—	7	5 1/4	12	7 1/8	5 3/8	5 1/2	13	(8)16d	(12)16d	1930	1930	1495	1790	1395	1395
PCM77	—	7	7	12	7 1/8	7 1/8	5 1/2	15	(8)16d	(12)16d	1930	1930	1495	1790	1395	1395

(*continued*)

(continued)

USP Stock No.	Ref. No.	Post (Inches)	Beam (Inches)	Steel Gauge	Dimensions W1	W2	L1	L2	Fastener Schedule Post	Beam	Allowable Loads (LBS.) Lateral A1 133%	160%	Lateral A2 133%	160%	Uplift 133%	160%
EPCM4416	EPC44-16	3 1/2	3 1/2	16	3 9/16	3 9/16	3 1/16	7 1/4	(8)16d	(8)16d	1020	1020	1055	1055	805	805
EPCM4516	—	3 1/2	5 1/4	16	3 9/16	5 3/8	3 1/16	9 1/4	(8)16d	(8)16d	1020	1020	1055	1055	805	805
EPCM4716	—	3 1/2	7	16	3 9/16	7 1/8	3 1/16	11 1/4	(8)16d	(8)16d	1020	1020	1055	1055	805	805
EPCM5416	—	5 1/4	3 1/2	16	5 3/8	3 9/16	3 1/2	7 1/4	(8)16d	(8)16d	1020	1020	1055	1055	805	805
EPCM5516	—	5 1/4	5 1/4	16	5 3/8	5 3/8	3 1/2	9 1/4	(8)16d	(8)16d	1020	1020	1055	1055	805	805
EPCM44	EPC44	3 1/2	3 1/2	12	3 9/16	3 9/16	3 1/16	7 1/4	(8)16d	(8)16d	1180	1180	1495	1610	1480	1480
EPCM45	—	3 1/2	5 1/4	12	3 9/16	5 3/8	3 1/16	3 1/16	(8)16d	(8)16d	1180	1180	1495	1610	1480	1480
EPCM47	—	3 1/2	7	12	3 9/16	7 1/8	3 1/16	11 1/4	(8)16d	(8)16d	1180	1180	1495	1610	1480	1480
EPCM54	—	5 1/4	3 1/2	12	5 3/8	3 9/16	3 1/2	7 1/4	(8)16d	(8)16d	1180	1180	1495	1610	1480	1480
EPCM55	—	5 1/4	5 1/4	12	5 3/8	5 3/8	3 1/2	9 1/4	(8)16d	(8)16d	1180	1180	1495	1610	1480	1480
EPCM57	—	5 1/4	7	12	5 3/8	7 1/8	3 1/2	11 1/4	(8)16d	(8)16d	1180	1180	1495	1610	1480	1480
EPCM75	—	7	5 1/4	12	7 1/8	5 3/8	5 1/2	11 1/4	(8)16d	(8)16d	1180	1180	1495	1610	1480	1480
EPCM77	—	7	7	12	7 1/8	7 1/8	5 1/2	13 1/4	(8)16d	(8)16d	1180	1180	1495	1610	1480	1480

Uplift loads and lateral loads have been increased 33 1/3% or 60% for wind and seismic loads. No further increase shall be permitted.
Allowable loads are for nails only.

Post Caps PCM and EPCM Series

Post caps provide a positive connection for medium duty post-to-beam applications. The extended side plate design in these products also function as tie straps where splices occur. These models fasten with standard nails.

Finish ..... G60 galvanizing
Code Listing ..... NER 505

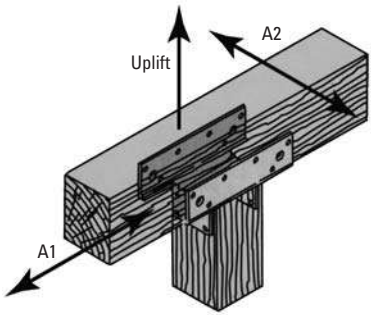

Typical **PCM** Installation

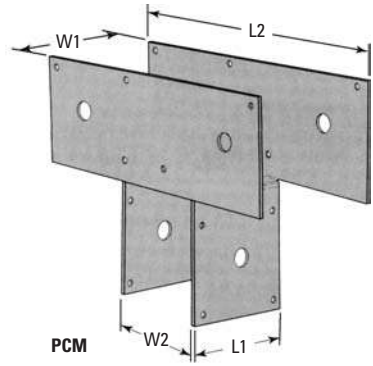

PCM

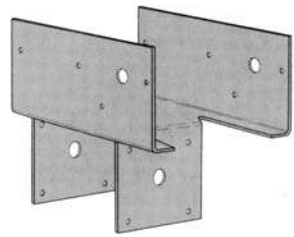

EPCM

Column Caps—KCC and KECC Series

			Dimensions					Bolt Schedule							
						L				Bearing Load		Uplift KCC		Uplift KECC	
USP Stock No.	Ref. No.	Steel Gauge	W1	W2	H	KCC	KECC	Beam	Column or Post	KCC	KECC	133%	160%	133%	160%
KCC44	CC44	7	3 5/8	3 5/8	4	7	5 1/2	(2) 5/8	(2) 5/8	15315	12030	2985	3585	1635	1960
KCC45	—	7	3 5/8	5 3/8	6 1/2	11	8 1/2	(4) 5/8	(2) 5/8	24065	18595	5270	6310	3210	3850
KCC47	—	7	3 5/8	7 1/8	6 1/2	11	9 1/2	(4) 5/8	(2) 5/8	24065	20780	5270	6310	3210	3850
KCC54	—	7	5 3/8	3 5/8	6 1/2	11	7 1/2	(4) 5/8	(2) 5/8	37810	25780	5255	6310	3385	4060
KCC55	—	7	5 3/8	5 3/8	6 1/2	11	7 1/2	(4) 5/8	(2) 5/8	37815	25780	5255	6310	3385	4060
KCC57	CC6 - 7 1/8	7	5 3/8	7 1/8	6 1/2	11	9 1/2	(4) 5/8	(2) 5/8	37815	32655	5255	6310	3385	4060
KCC75X	CC7 1/8-6	3	7 1/8	5 3/8	8	13	10 1/2	(4) 3/4	(2) 3/4	54600	44100	6230	7255	4665	5595
KCC77X	CC7 1/8 - 7 1/8	3	7 1/8	7 1/8	8	13	10 1/2	(4) 3/4	(2) 3/4	54600	44100	6230	7255	4665	5595

Bearing loads are based on 560 psi perpendicular to grain loading for glulam sizes and 8× and wider, and 625 psi for all others.
Uplift loads are increased 33% and 60% for wind or seismic loads; no further increase is permitted.
Beams shall be designed to support the required loads. Beam shear may limit loads to less than listed loads for device.
Posts are assumed adequate to support listed loads.
Bearing loads shall not be increased for duration of load.

Column Caps KCC and KECC Series

Designed for heavy duty beam-to-post connections, column caps offer strong reinforcement and support at critical framing junctions. A number of special sizes and style variations are available. Straps are normally centered on the U-bracket, but any variation may be specified. For end column caps, add an "E" after the "K" in the part number, as in "KECC325-4." Straps may be rotated 90 degrees on special order where the W2 dimension is less than or equal to the W1 dimension. Welding is performed by certified welders and routinely inspected by a code-accepted quality control agency.

Finish ..... paint
Code Listings ..... ICBO 2725
..... L.A. City RR25104

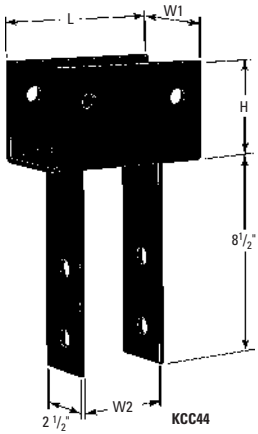

KCC44

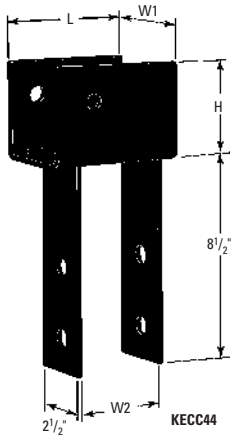

KECC44

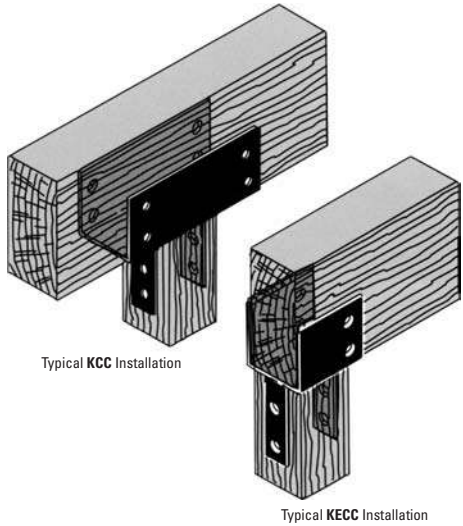

Typical **KCC** Installation

Typical **KECC** Installation

KCC and KECC Series Options

A number of size and style variations are available for KCC and KECC Column Caps on a custom order basis. Possible variations include rough or custom sizes and altered bolt pattern. A number of standard options are also available as detailed by the adjacent illustrations. These options include: KCCO, KCCT, KCCOB and 90° strap rotation.

Note

Consult retail price list or your technical assistance representative for details and ordering information.

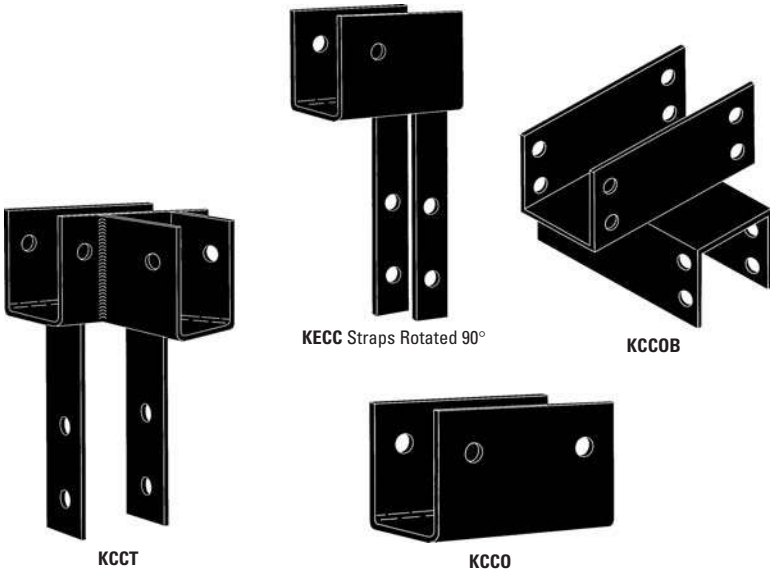

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

This Course Consists Of:

- Compass Basics
- Orienting a Map For North Reference
- Traveling to a Target
- Backstops, Handrails and Aiming Off
- Topo Map Basics
- Pace Counting
- Finding Yourself
- UTM Coordinate System
- Using GPS with a Map
- Understanding And Adjusting For Declination

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Compass Types

Baseplate

Accessory

Mirrored – Sighting (Preferred)

Lensatic

Digital

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Boxing A Compass

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

MAP ORIENTATION
With Compass Adjusted For Declination

1) Adjust dial so "N" is at index line.

2) Align edge of compass with north / south reference line.

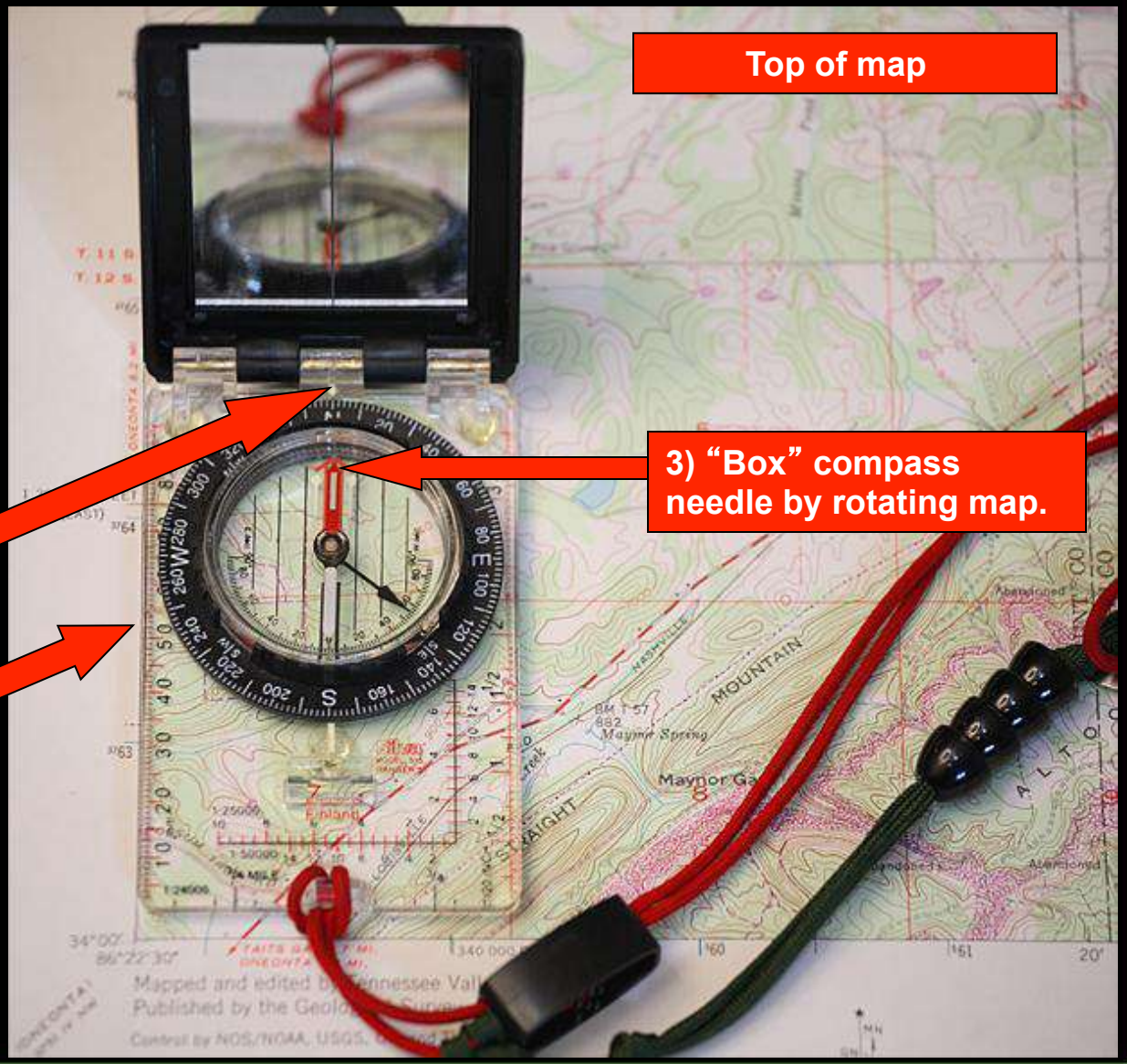

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Finding An Azimuth With Compass

(With Map Oriented To North)

Traveling To a Target On The Map – Step 1

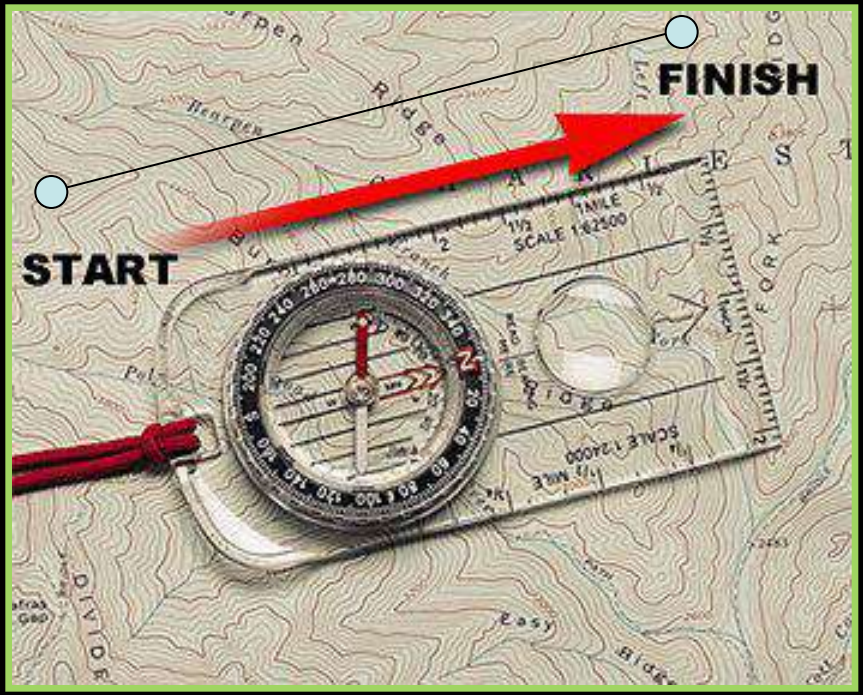

1) Draw a line on the map from your starting point to your finish point.
2) Orient the map to north.
3) Place the compass on the map with the edge of the compass on the line and the bearing arrow pointing at your destination

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Traveling To a Target On The Map – Step 2

Set Compass Heading - Turn the dial on the compass until "N" outline arrow boxes the compass needle. Your direction to your target (in degrees) is read at the Index Line on the Dial.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Traveling To a Target On The Map – Step 3

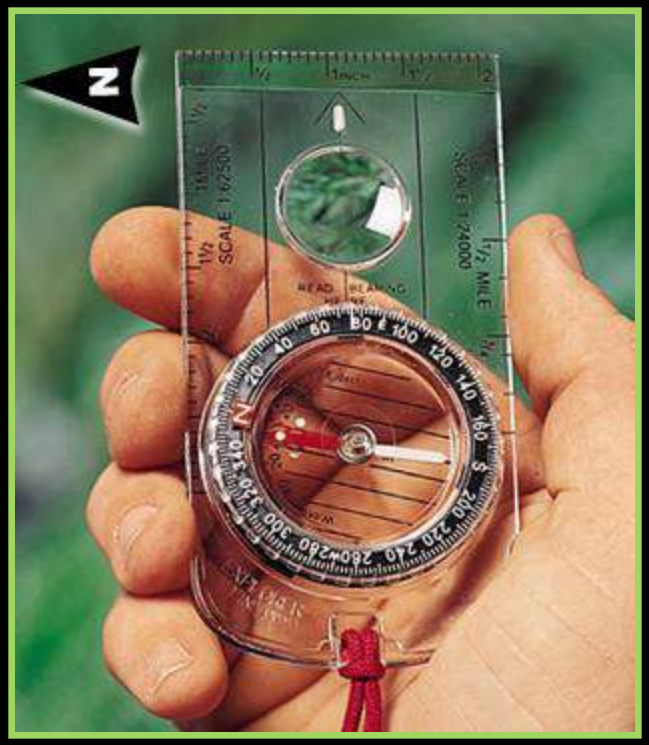

Follow Your Heading - Remove the compass from the map and hold it level, so the Magnetic Needle is free to turn. Turn your body until the red end of the Needle aligns with the Orienting Arrow (boxed) and "N" on the Dial. Using the Direction of Travel Arrow, sight a distant landmark and move to it. Repeat this process until you reach your destination.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Finding An Azimuth With Compass

(Without Map Oriented To North)

- Draw a line from start to finish.
- Lay the edge of your compass on the line in the direction you are traveling.
- Rotate the compass bezel until the orienting lines align with your North Reference Lines (Make sure that North is pointing to North on the map).
- Pick up your compass and travel to the destination.
- Note: The compass shown has already been adjusted for declination.

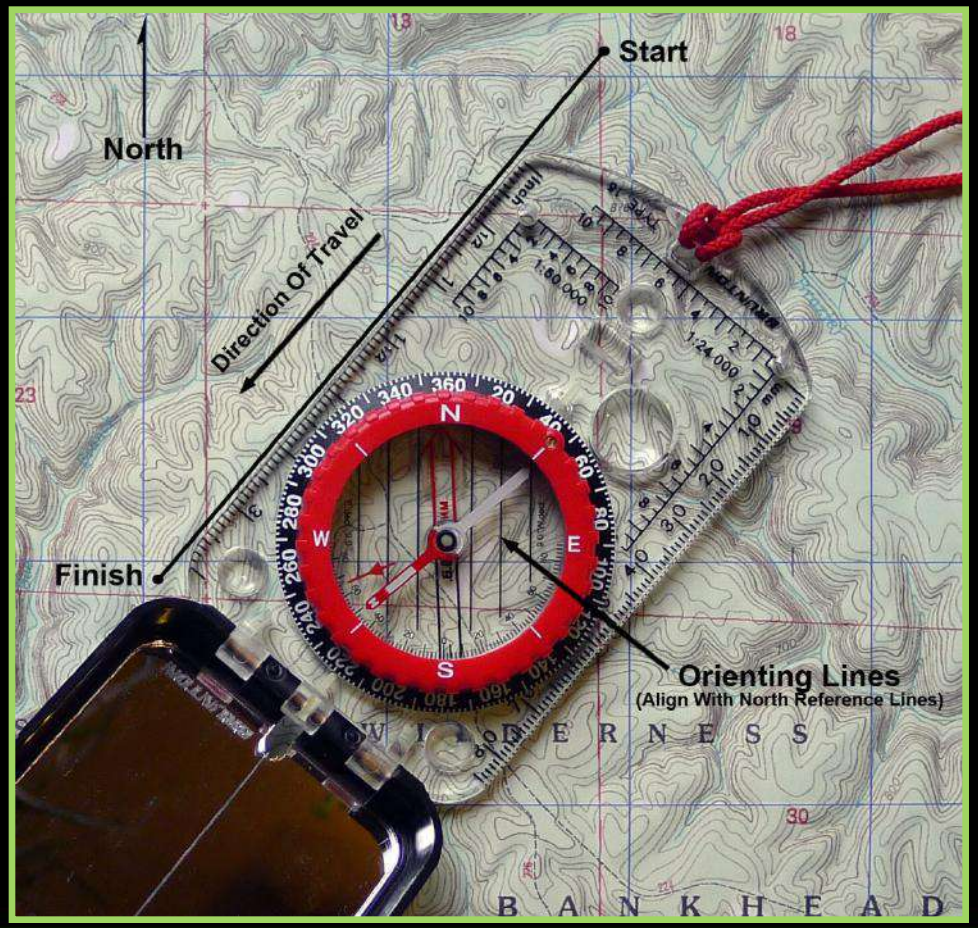

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Finding An Azimuth With Compass Card

(Without Map Oriented To North)

- Draw a line from start to finish.
- Lay the compass rose on the line so the line crosses through the center of the rose.
- Align N / S line or orienting lines on the compass rose with North Reference Line.
- Read your azimuth on the "Finish" side of the compass rose.
- Note: unless you are using a compass adjusted for declination, you will have to add or subtract the declination value to the value that you acquire from the compass rose.

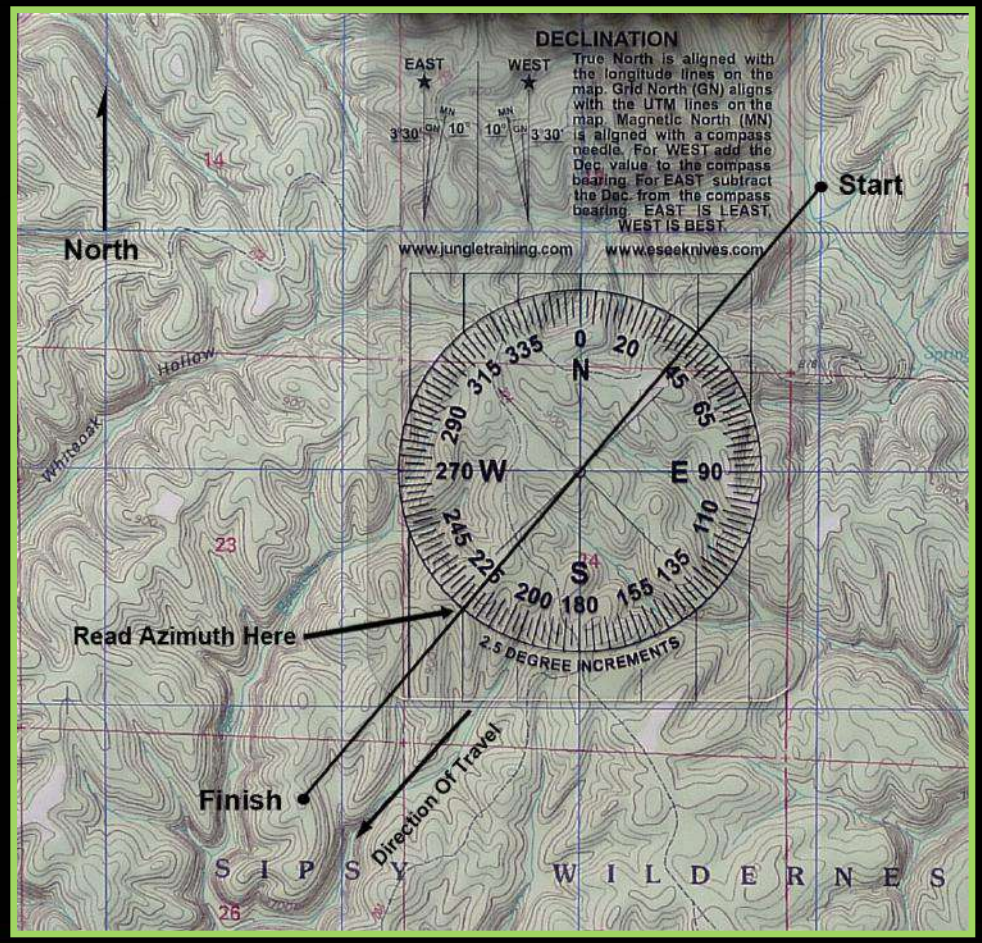

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Traveling To a Target With Sighting Compass

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Posting Up

In addition to using natural landmarks as targets for your direction of travel, you can also post up a member of your team as a target. Another option is the Back Azimuth.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Back Azimuth

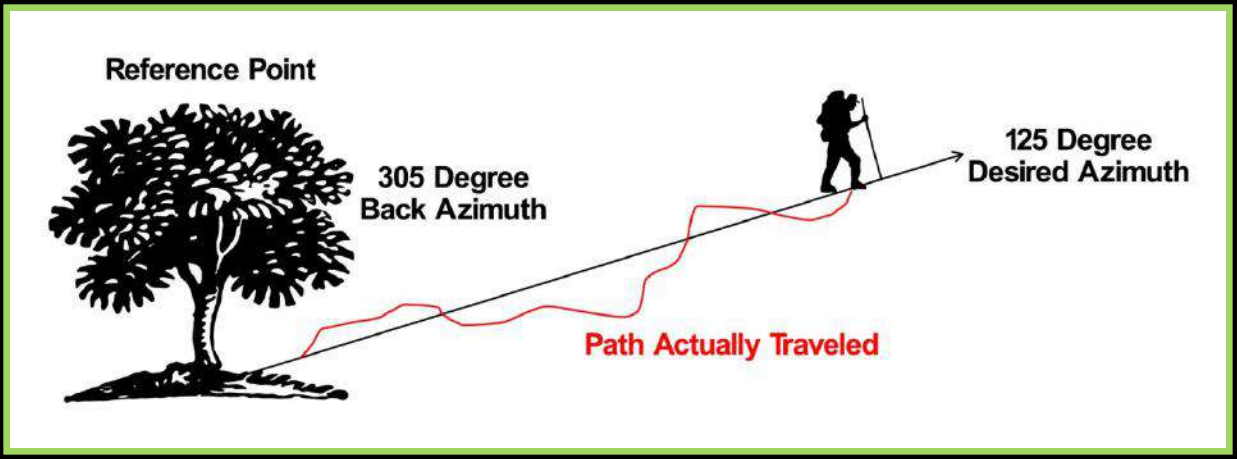

Use a Back Azimuth if you need to verify you are still on your desired direction of travel. This is especially handy when crossing rivers or boxing around objects.

Back Azimuth is 180 degrees from your target azimuth. If your target azimuth is more than 180 degrees, then subtract 180. If it is less than 180, then add 180.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Boxing Around Objects

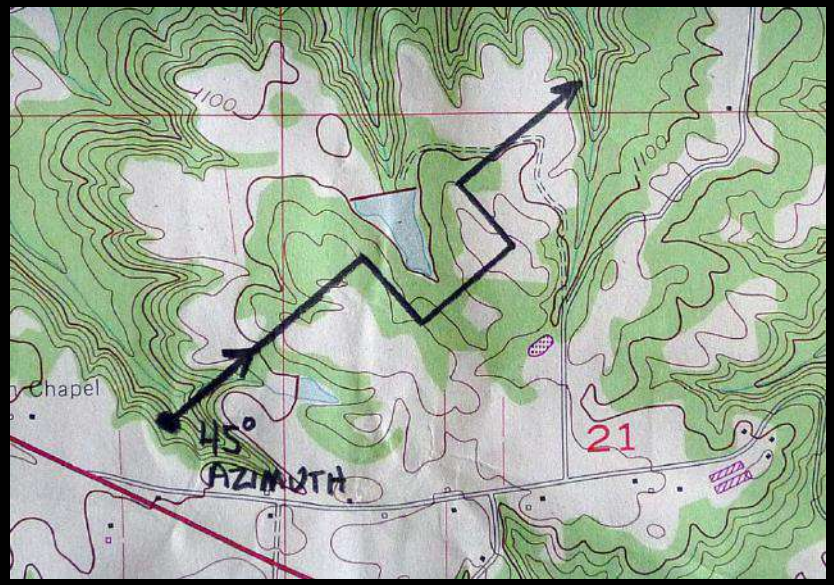

- Direction of original travel is 45 degrees
- Turn 90 degrees to a new azimuth of 135 (count your paces)
- Turn 90 degrees back to the original azimuth of 45 degrees until you are past obstacle
- Turn 90 degrees to a new azimuth of 315 and pace the same amount as the first turn
- Turn 90 degrees back to the original azimuth of 45 degrees and continue on

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

LATERAL DRIFT – MISSING YOUR TARGET

If you cannot find your target after walking your azimuth the correct distance, then mark the spot where you think it should be and began a structured grid search of the area. Make note of your azimuths, pace count and direction changes. Start your grid small then go larger if you cannot locate the target.

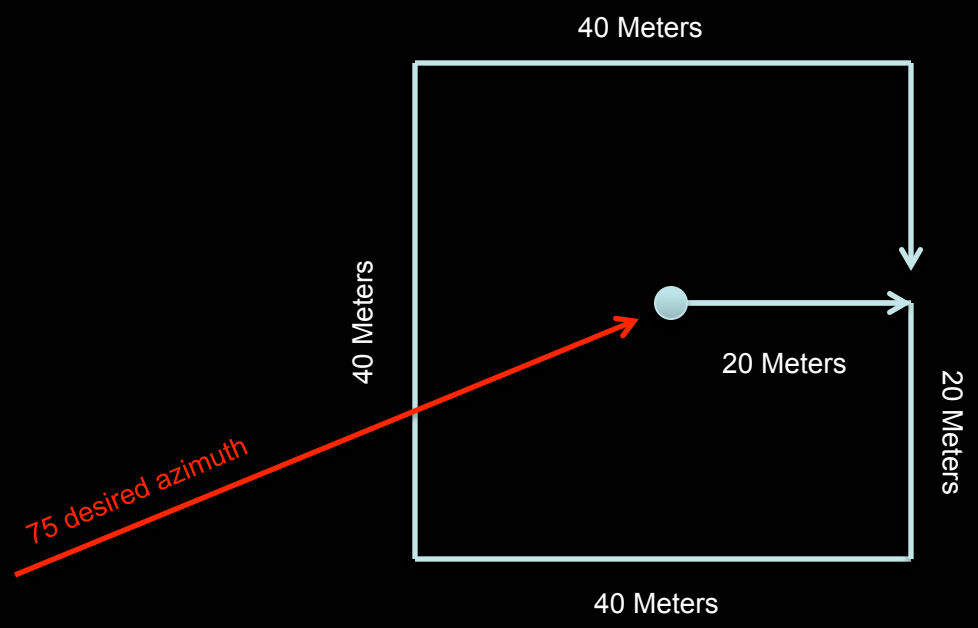

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Gross Direction Navigation

As obvious from this simple road map, the lost hiker only needs to travel a westerly direction to cross a road and reach civilization. Always know which general direction crosses a main artery!

When using a map and compass always have a "panic azimuth."

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Gross Direction Navigation Aids

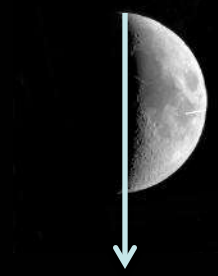

Quarter Moon

S

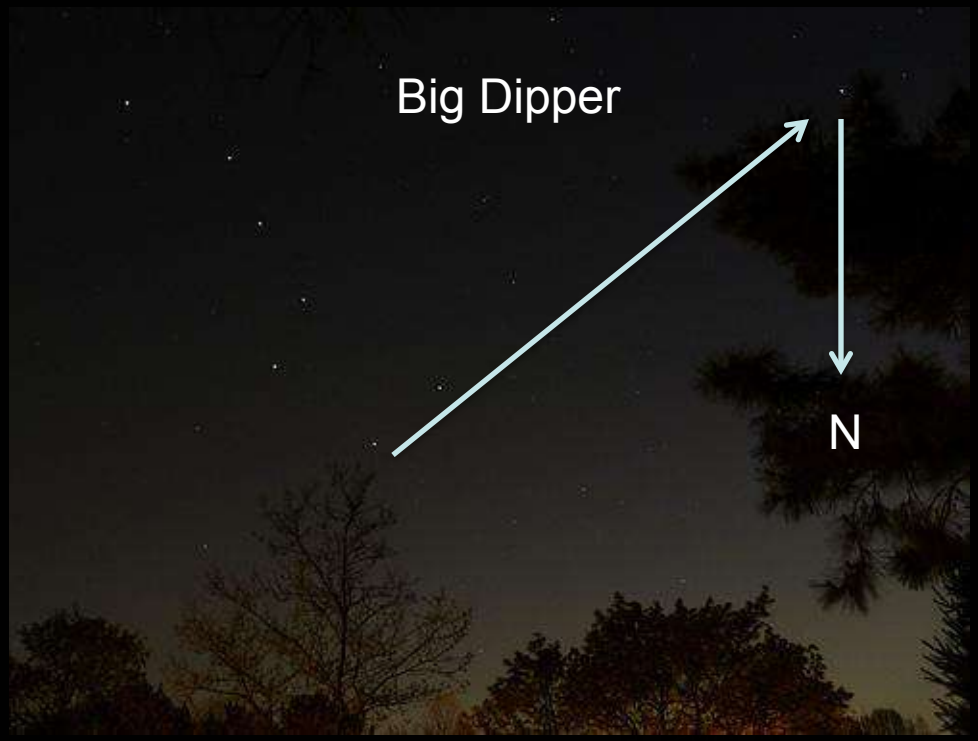

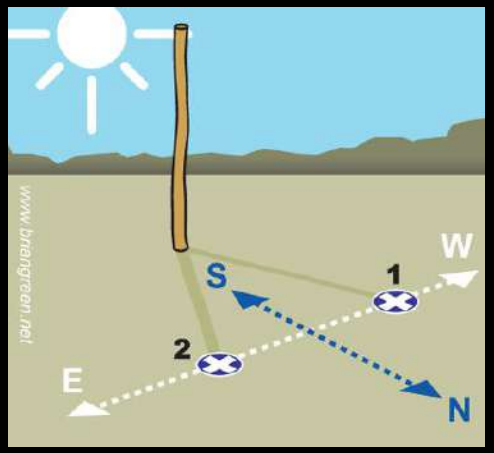

Shadow Stick

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Terrain Association Navigation

This is the most widely used method of navigation. The navigator plans his route so that he moves from terrain feature to terrain feature. An automobile driver in a city uses this technique as he moves along a street or series of streets, guiding on intersections or features such as stores, parks or houses. Like the driver, the navigator selects routes or *streets* between key points or *intersections*. These key points can be lakes, mountains, roads or any other terrain feature readily recognized on a map. Before you hike a new area, familiarize yourself with the terrain by studying maps, satellite photos or any other available information.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Backstops, Handrails & Aiming Off

BACKSTOPS are features that typically run perpendicular to your azimuth and located beyond your target point. If you run into your backstop then you know you have traveled too far and missed your target point. Backstops can be rivers, roads, mountains, railroad tracks or any other useable feature.

HANDRAILS are features that run parallel to your azimuth. They can be rivers, roads, mountains, or any other useable feature that the navigator can use to continue traveling in the desired direction.

AIMING OFF is a process of deliberately adding or subtracting to the desired azimuth so the navigator knows which way to go once they reach a feature (such as a trail or road) that leads to the target.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Backstops, Handrails & Aiming Off

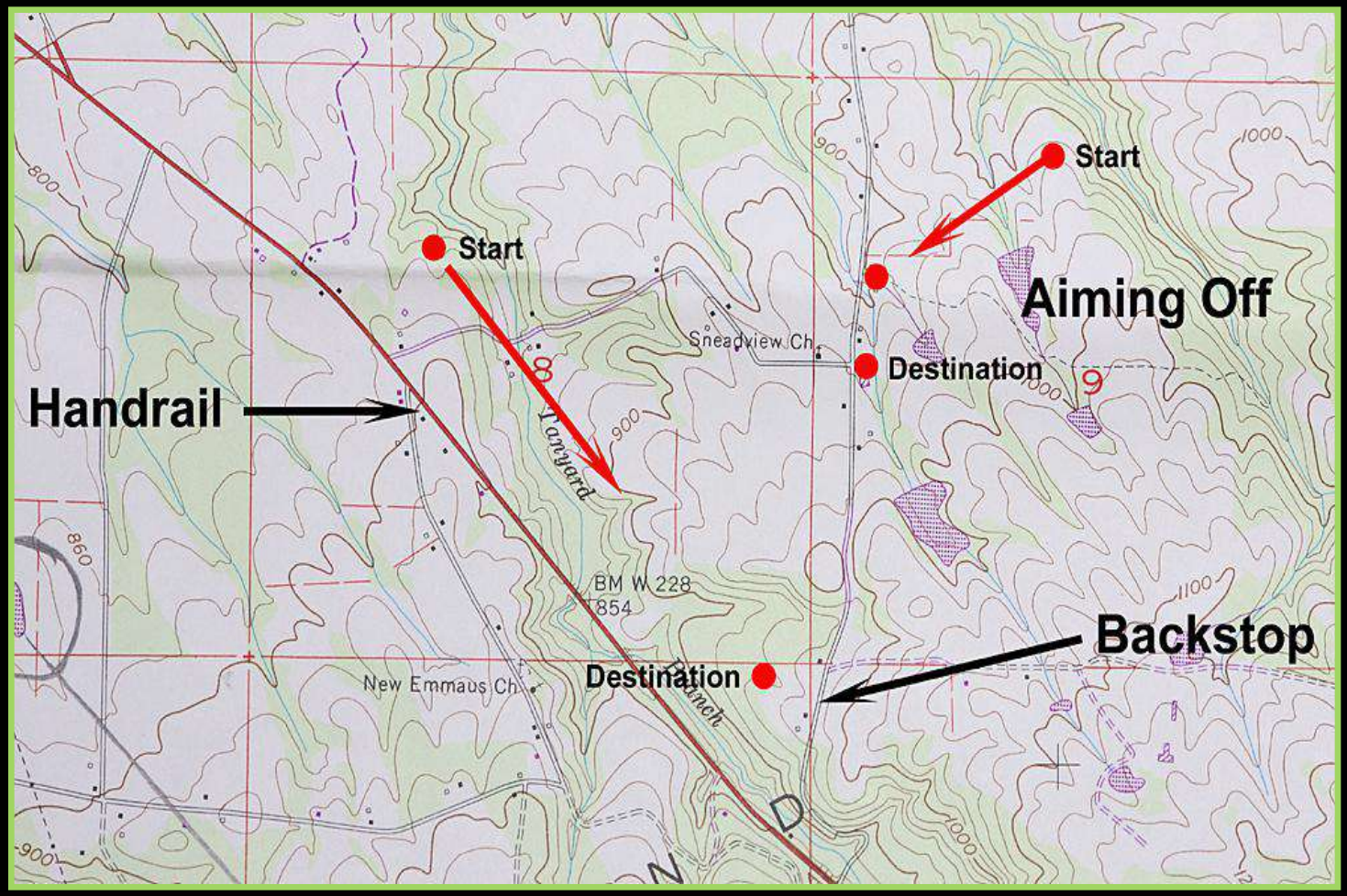

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

TOPO MAP BASICS

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

BASICS SYMBOLS

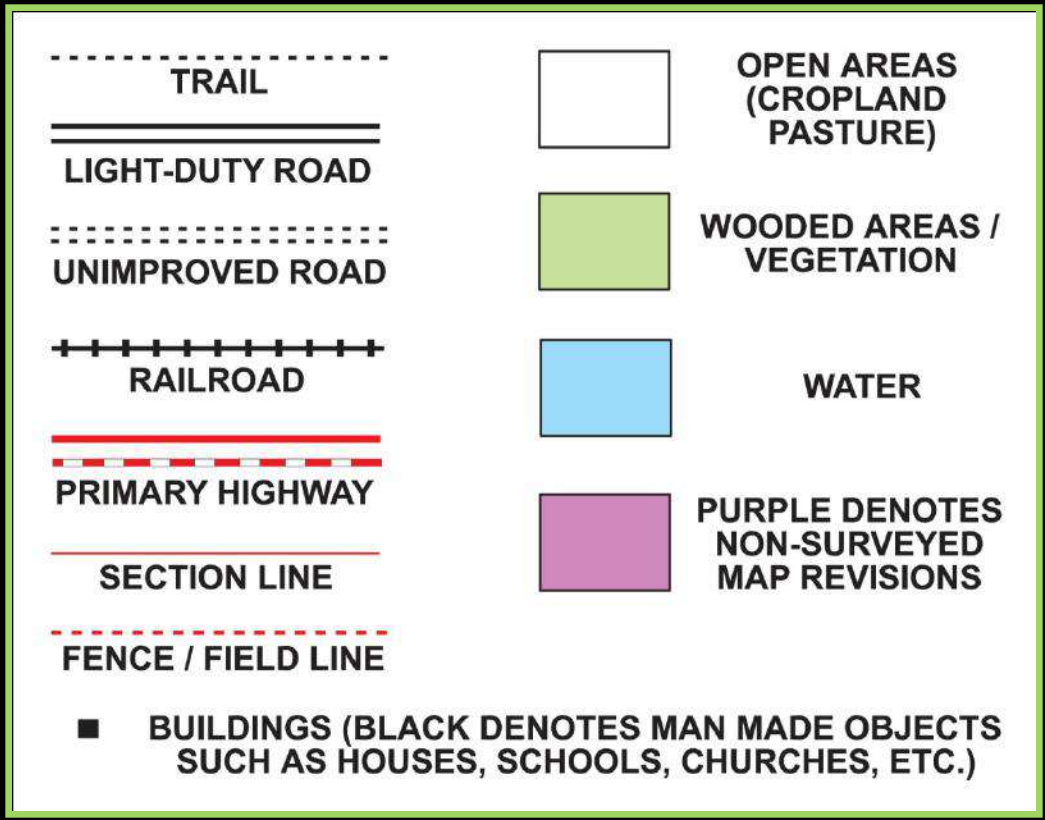

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Map Scale & Datum

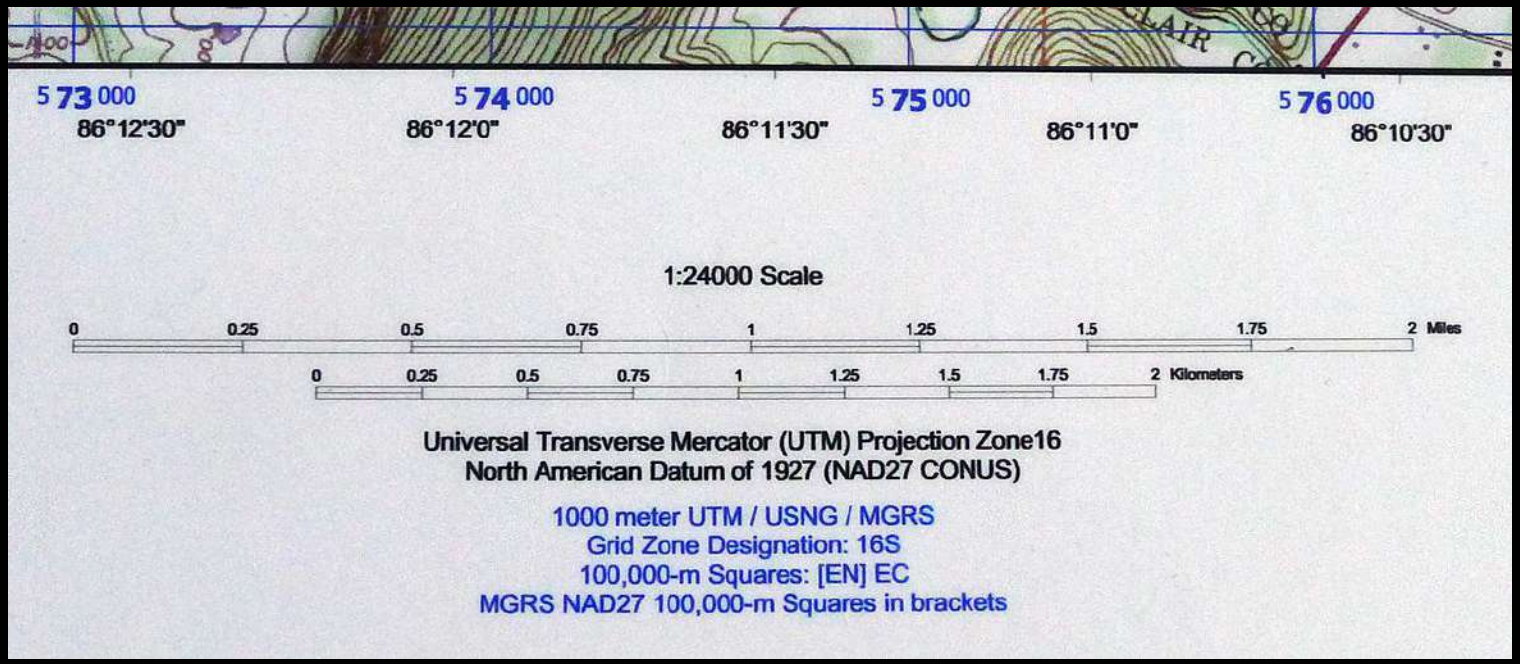

Always set your GPS to the map datum shown on your map!
Always use the proper scale card for measuring!

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

CONTOUR LINES

The closer the contour lines, the steeper the terrain. Check the map for the contour intervals. Contour lines will also have elevation markings.

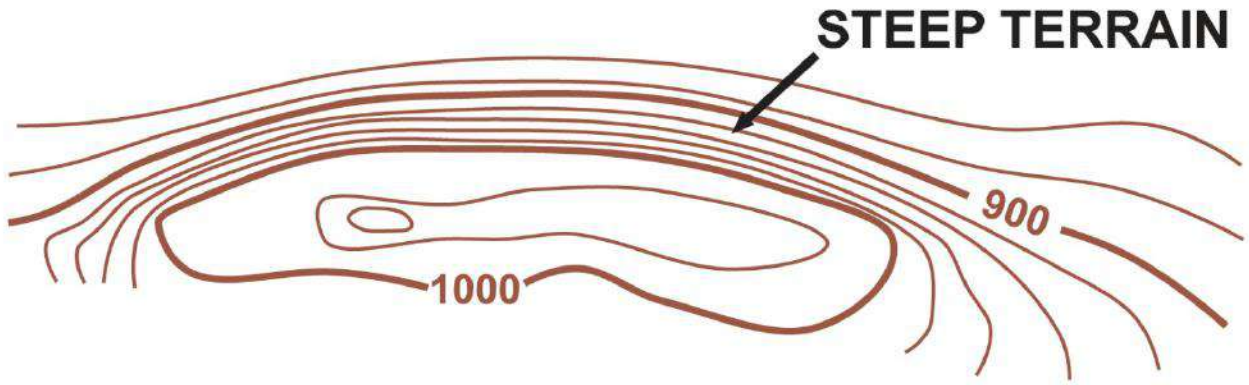

Example above shows a map with contour intervals of 20 feet.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

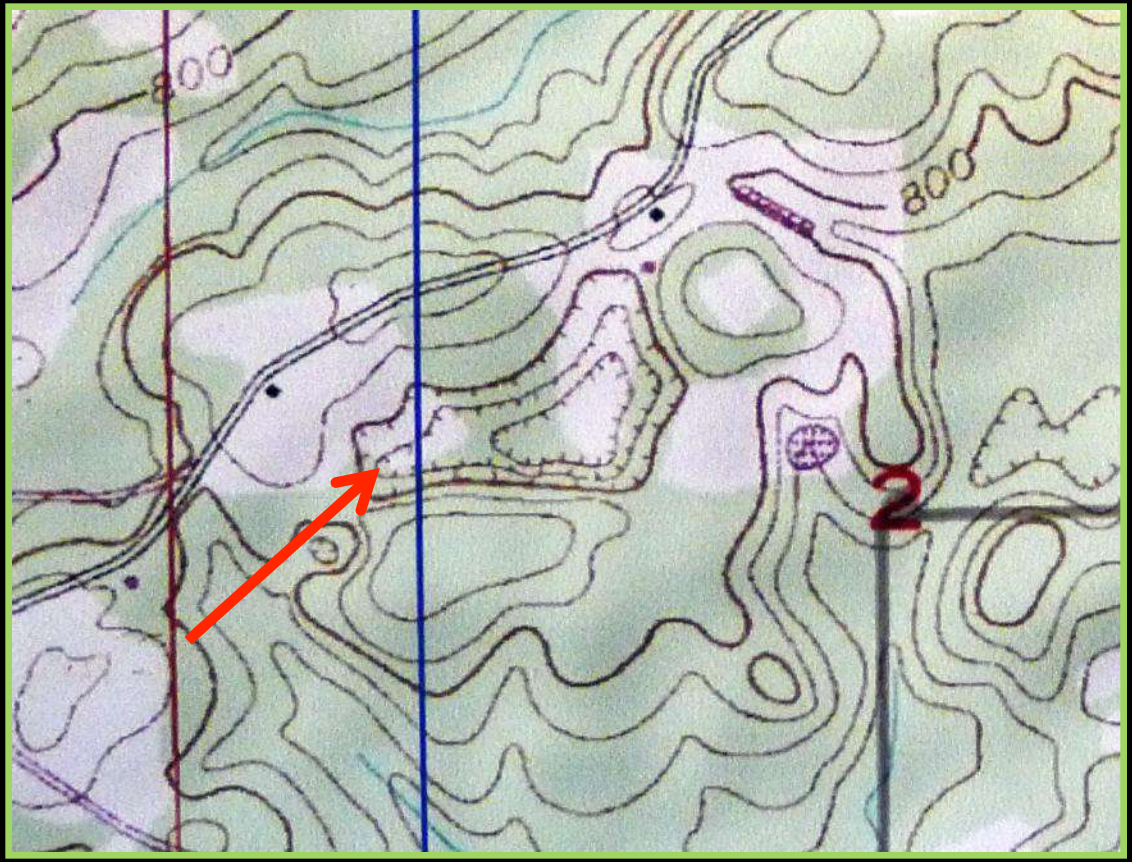

Depression

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

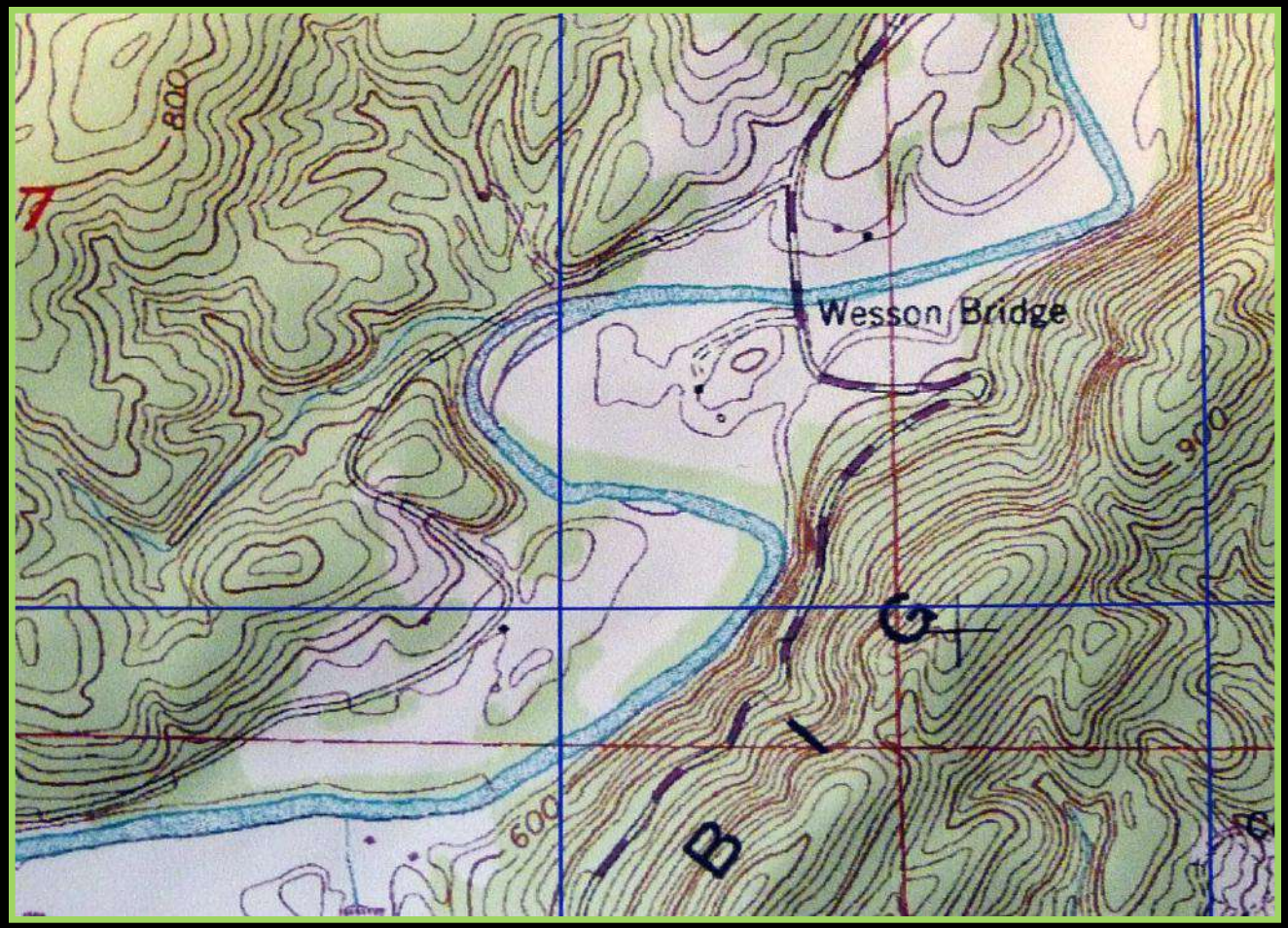

Continuous Water Flow

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

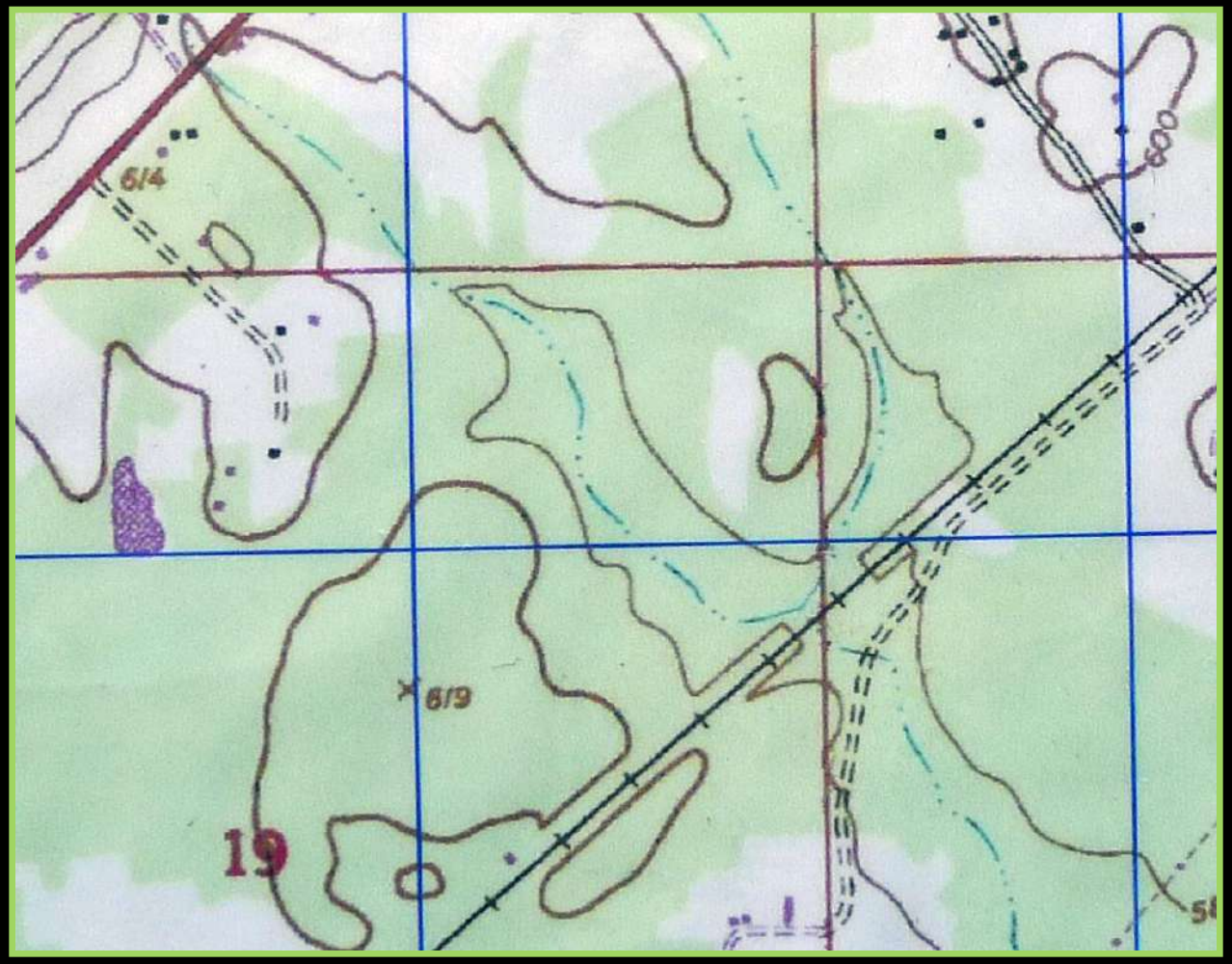

Intermittent Water Flow

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Clear Areas / Fields

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

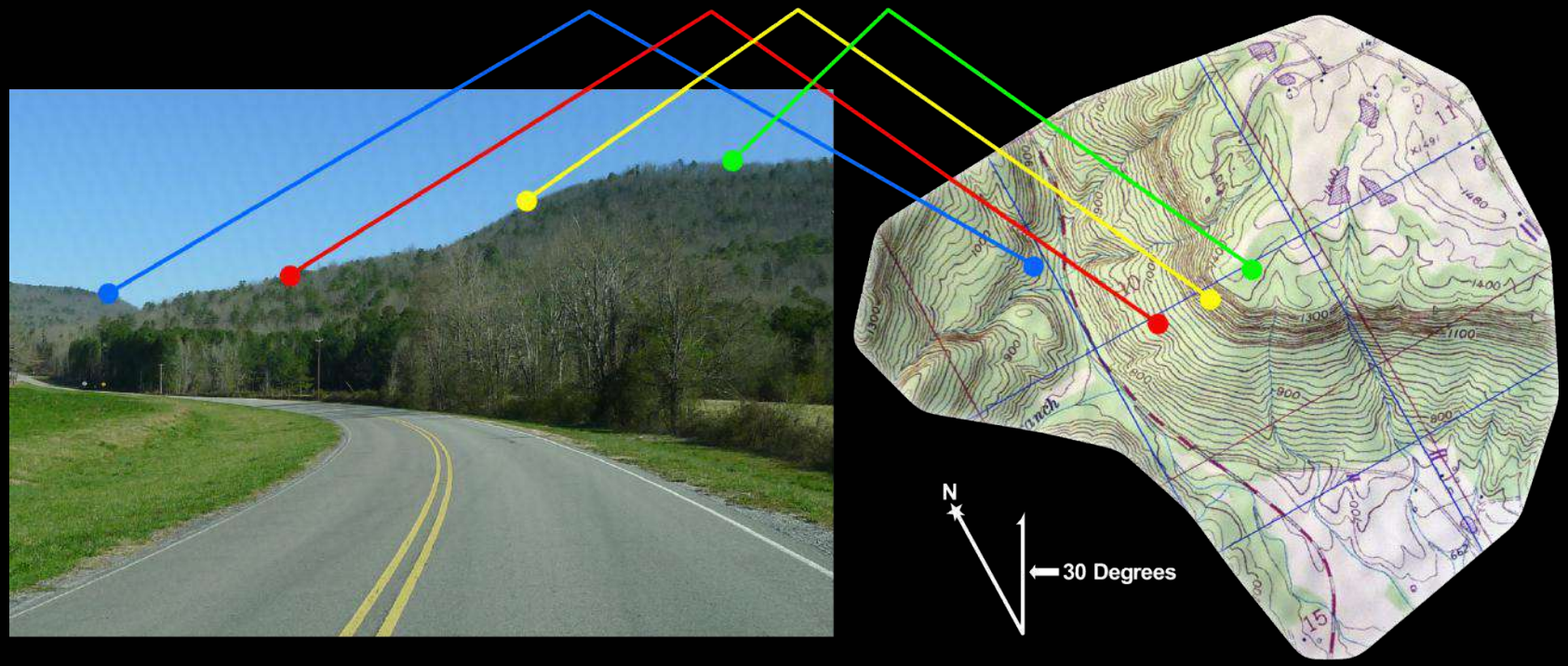

Actual View

Topo View

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

Actual View

Topo View

300 Degrees

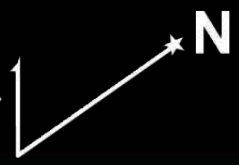

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

DETERMINING DISTANCE

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Determining Distance On Map

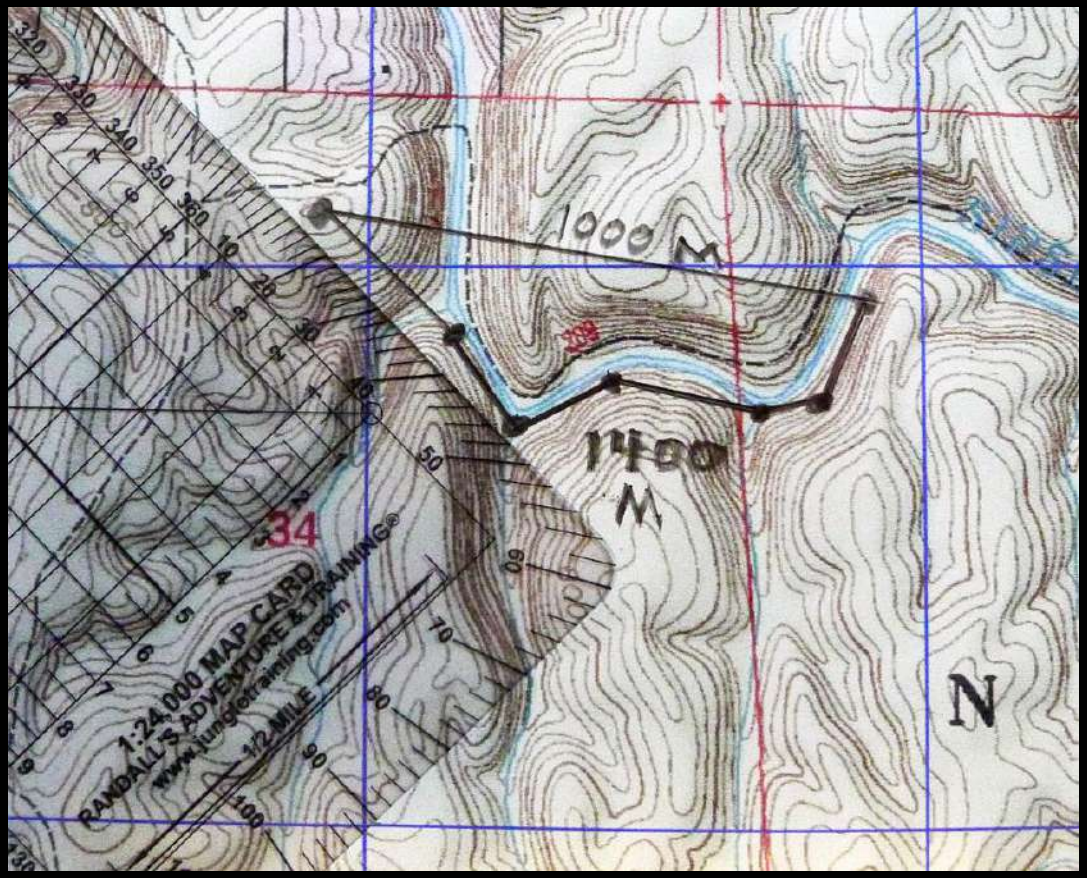

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

DETERMINE DISTANCE TRAVELED BY PACE COUNT

In thick jungle, where landmarks can not always be seen to track your position, pace counting is the best way of measuring distance. Pace counting will allow the navigator to estimate where he is at any given time. To be accurate, the navigator must practice pacing over different types of terrain. First you have to do some calculations. Measure out exactly 100 meters on three types of ground. Flat easy terrain, rougher terrain with some slope and then steep hill terrain. Then on each measured course count your paces (every time your left foot touches the ground or every 2 steps = 1 pace). You will have 3 different pace counts for different types of terrain. Once finished, memorize your pace count for all 3 types.

Averages		
Flat easy terrain	100 meters	65 paces
Rougher terrain with some slope	100 meters	75 paces
Steep hill terrain	100 meters	95 paces

General Rule: You have never traveled as far as you think you have.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

CONDITIONS THAT AFFECT PACE COUNT

- Slopes – Your pace lengthens on a downslope and shortens on an upgrade.
- Winds – A headwind shortens the pace and a tailwind increases it.
- Surfaces – Sand, gravel, mud, snow and similar surfaces tend to shorten your pace.
- Elements – Falling rain or snow causes the pace to be reduced in length.
- Clothing – Excess clothing or shoes with poor traction affect the pace length.
- Visibility – Poor visibility due to rain, snow, or fog will shorten your pace.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Ranger Pace Count Beads

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

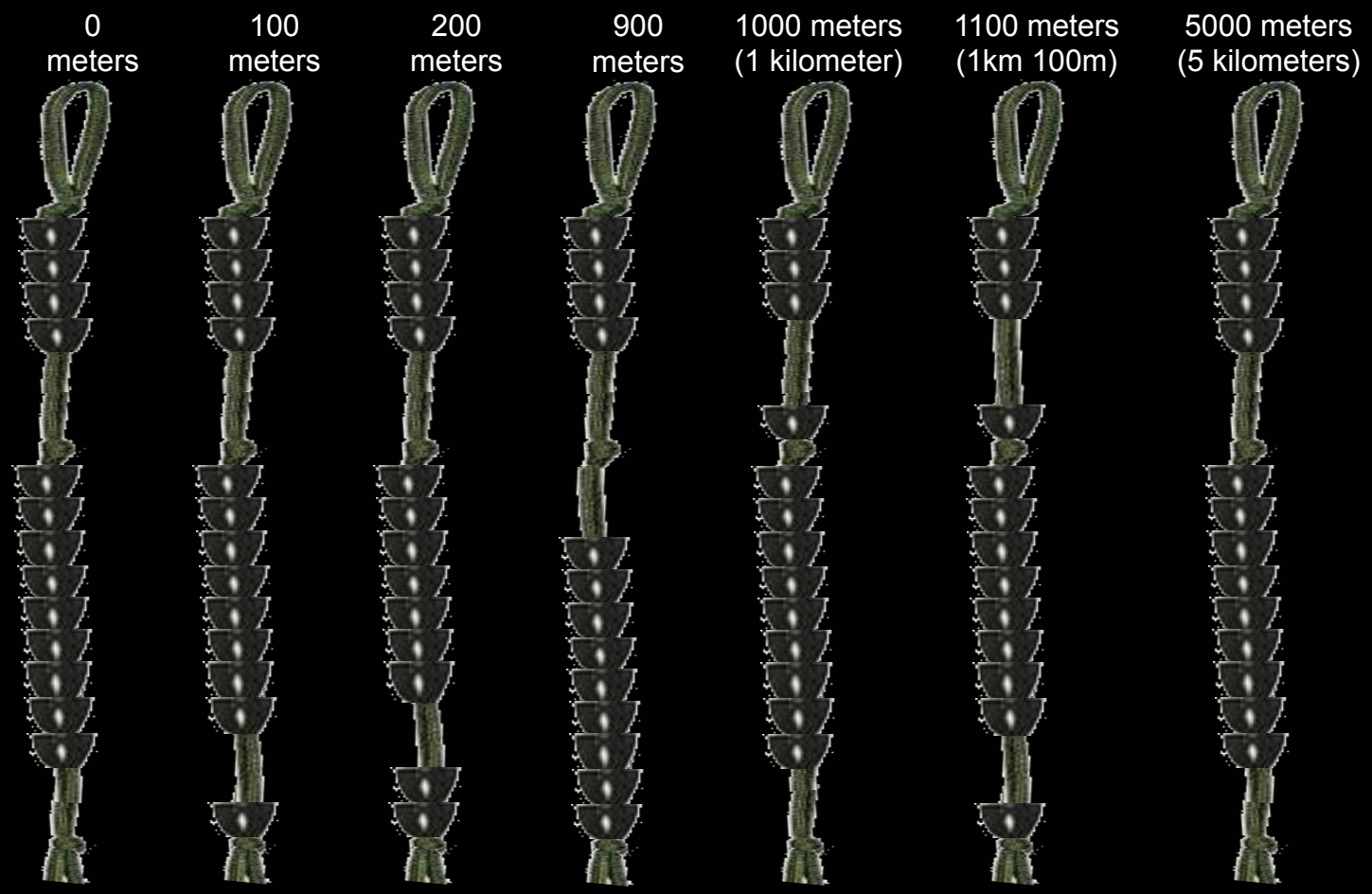

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

FINDING YOURSELF

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Baselines

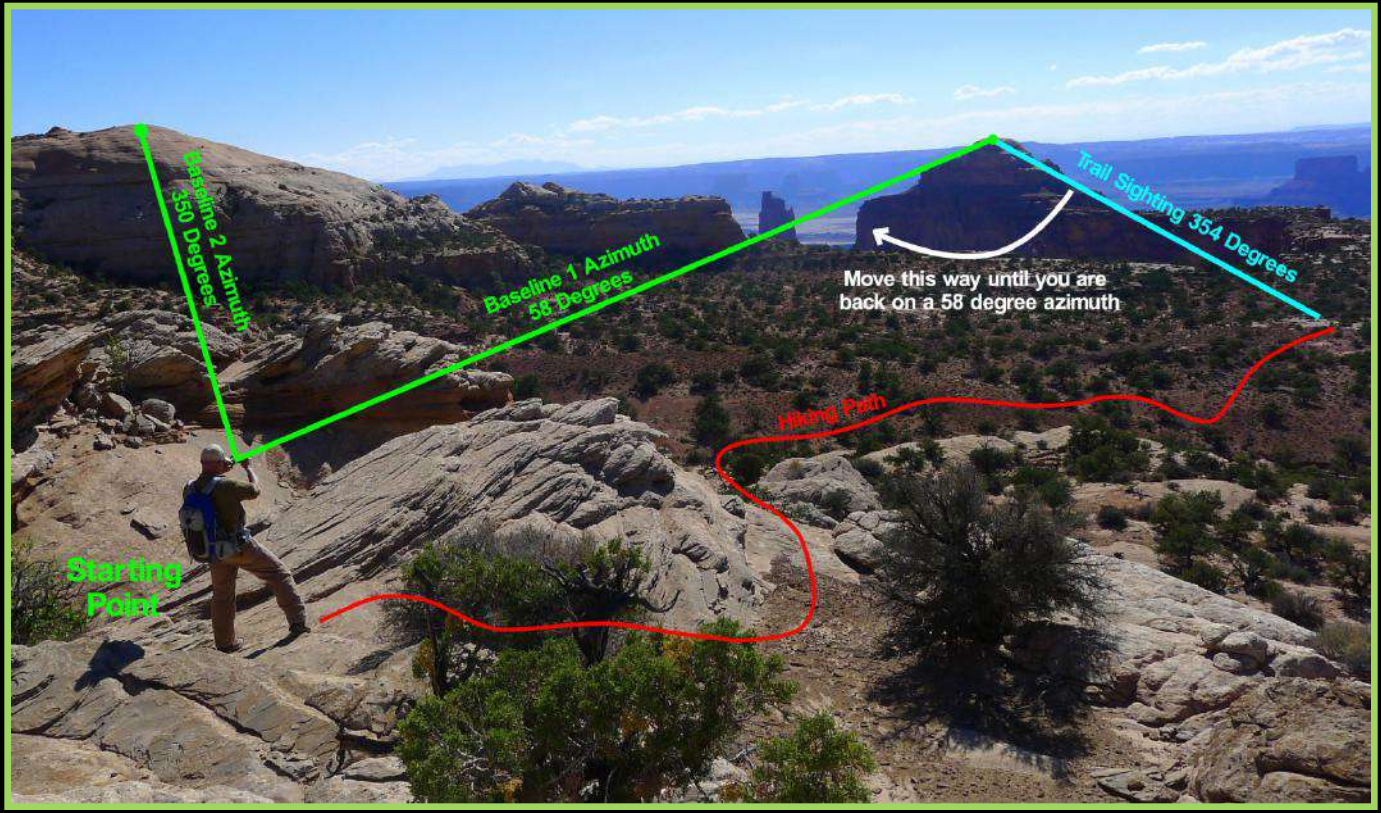

Find two prominent features. Shoot an azimuth to the first one and write it down. Shoot an azimuth to the second one and write it down. You can now travel freely. If you get lost, shoot an azimuth to the first feature and move until you get back on your first azimuth. Shoot an azimuth to your second feature and move until you get back on your second azimuth. You will be back at your starting point once both azimuths are the same as the ones you wrote down.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RESECTION

- Orient your map and secure its position. Locate a landmark on the map that you can actually see in the landscape. Take a compass bearing to that landmark by sighting and then rotating compass dial until the needle is boxed. Once you have a bearing, do not move the compass dial. Lay the corner of the compass on the map landmark and rotate the whole compass (not the dial) until the needle is boxed. If done properly the corner of the compass will still be over your landmark. Draw a line along the edge of the compass base.

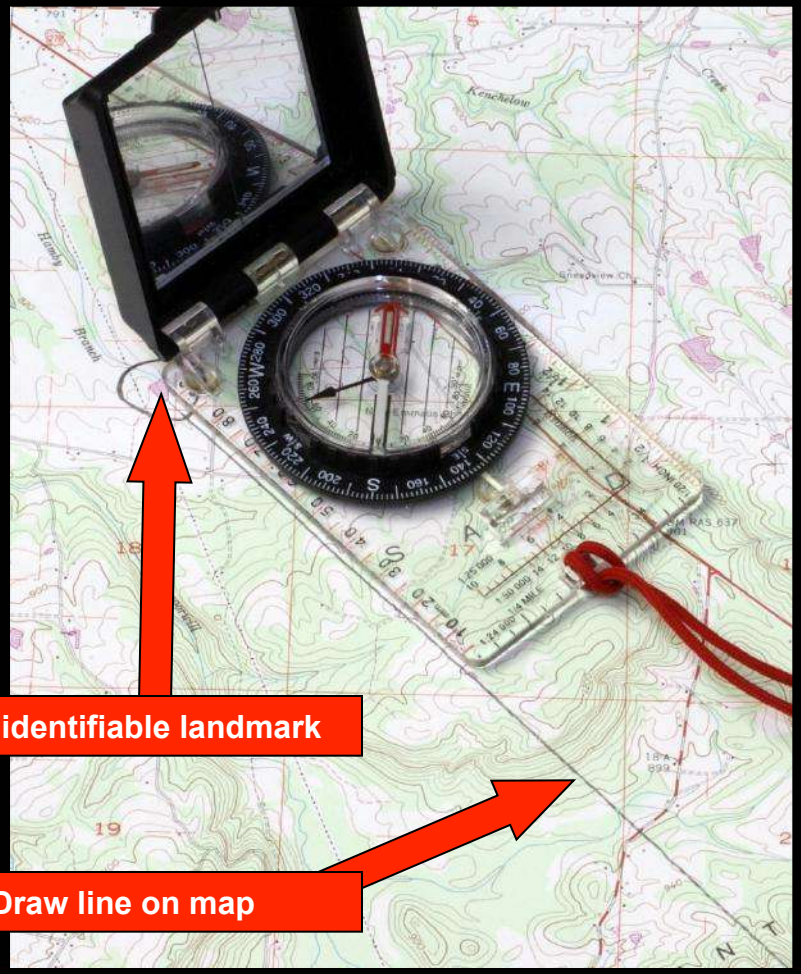

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RESECTION - Step 2

- Find a second identifiable landmark on the map and repeat the process from the previous slide. Where the two lines intersect is roughly where you are on the map.
- Finding a third point on the map and repeating the process is called "triangulation" and increases the accuracy of finding your location.

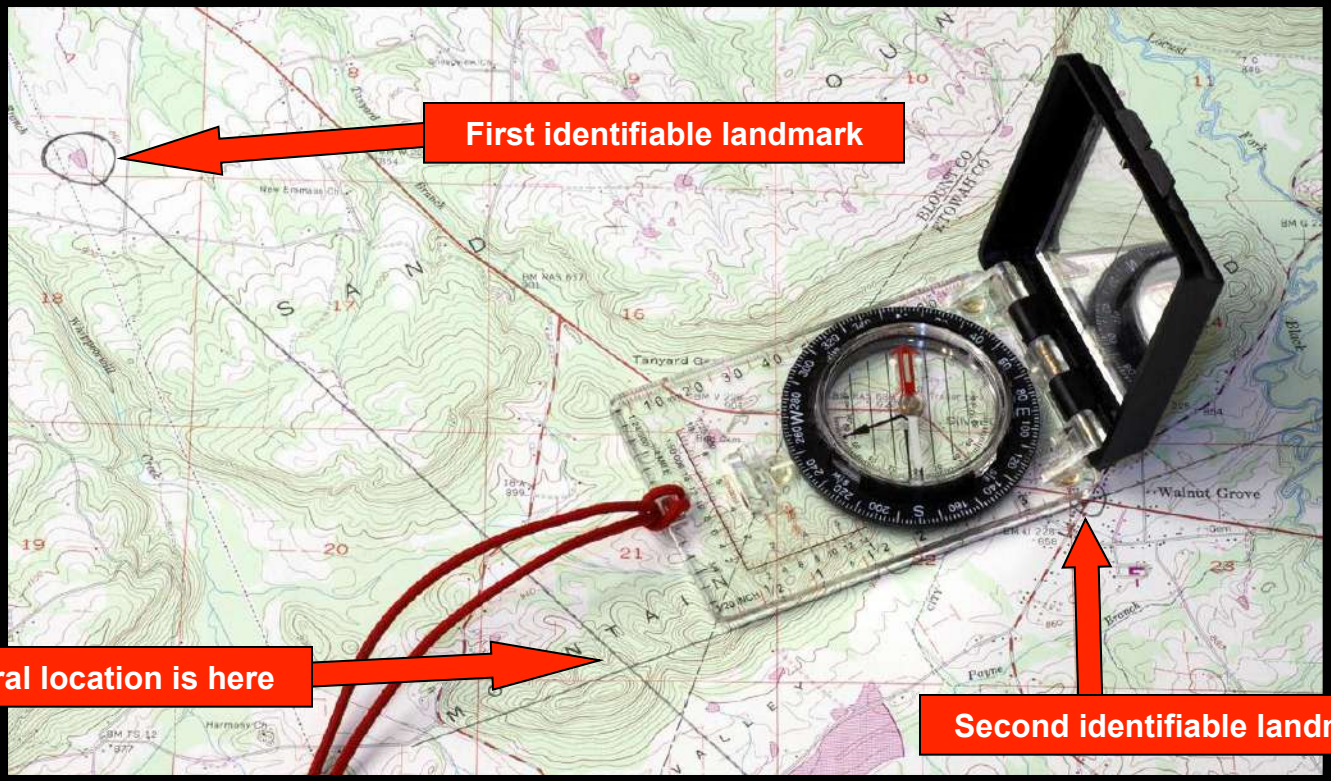

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

if you are already on a known location (such as a road, river or mountain ridge) but not sure exactly where you are on this feature, you may only need to shoot one azimuth to determine your exact position on the map

INTERSECTION is the reverse of RESECTION. An example of intersection is seeing smoke from a forest fire and determining where the fire is located on a map.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

PLOTTING COORDINATES

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Map Grids

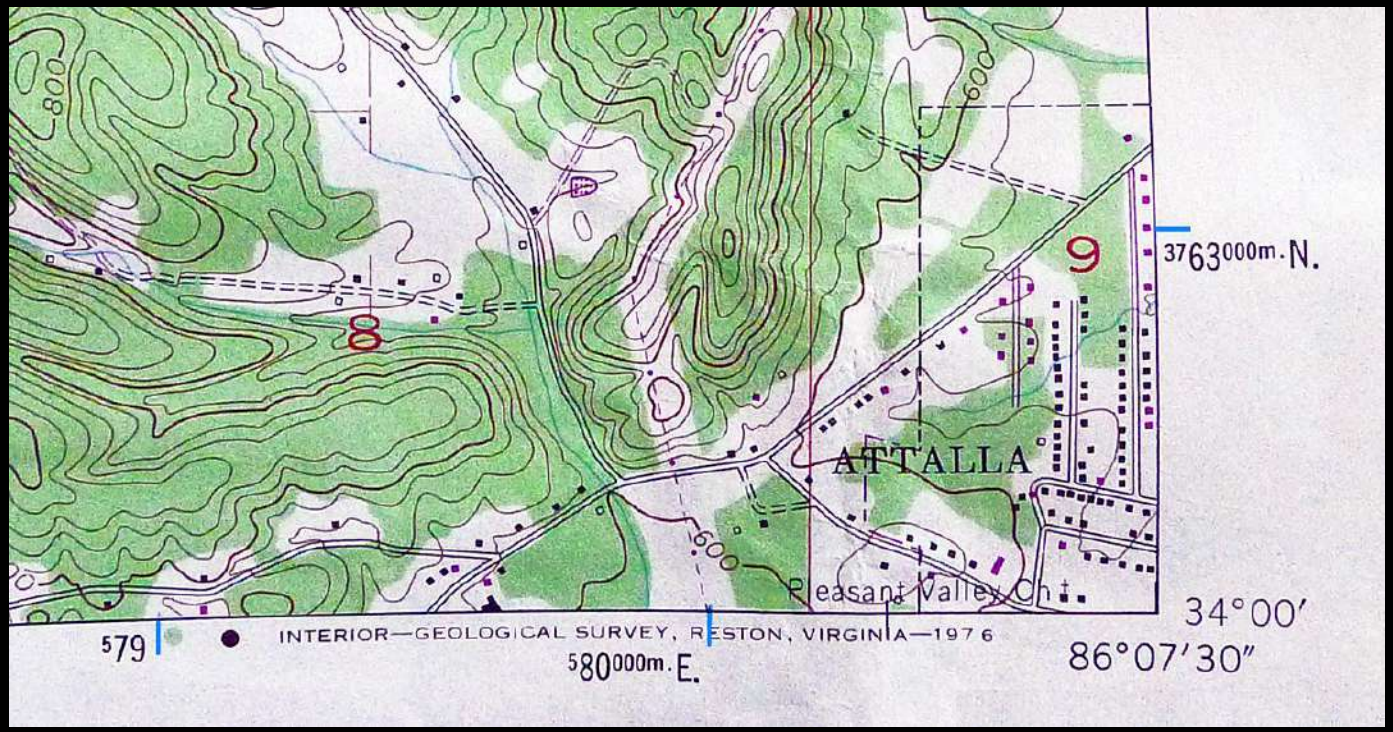

Always set your GPS to the grid system you are using on the map

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Universal Transverse Mercator (UTM)

UTM is a rectangular coordinate system based on the latitude and longitude (geographic) coordinate system. With UTM, the earth is divided into 60 zones that allows it to be projected onto maps with minimal distortion. All coordinates are expressed in meters.

The UTM grid system is the easiest method of relaying your position to others, finding yourself on a map after getting a location fix with a GPS unit, or navigating to a position on a map using a GPS unit.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

UTM: 60 Zones and 20 Latitude Bands

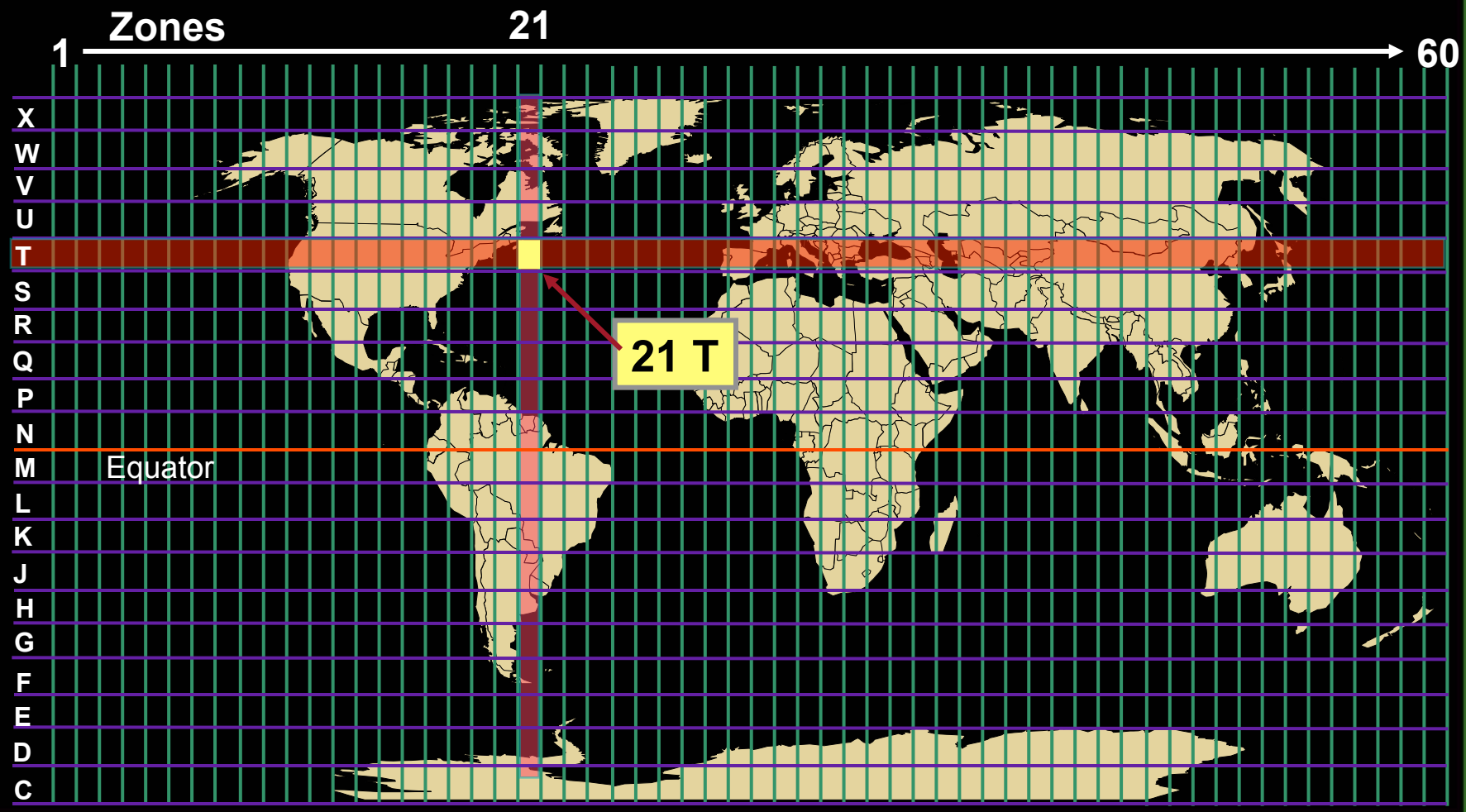

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

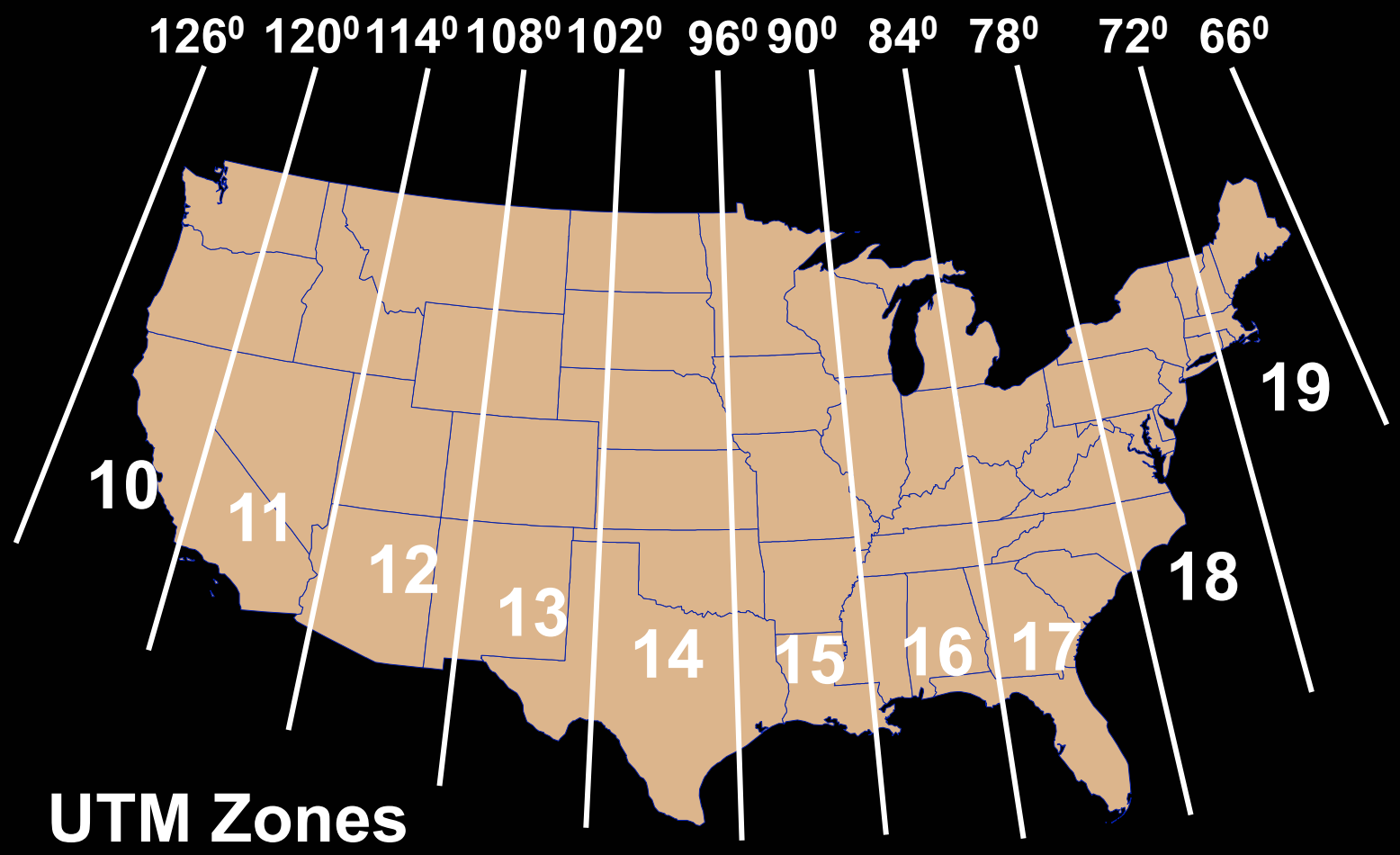

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Universal Transverse Mercator (UTM)

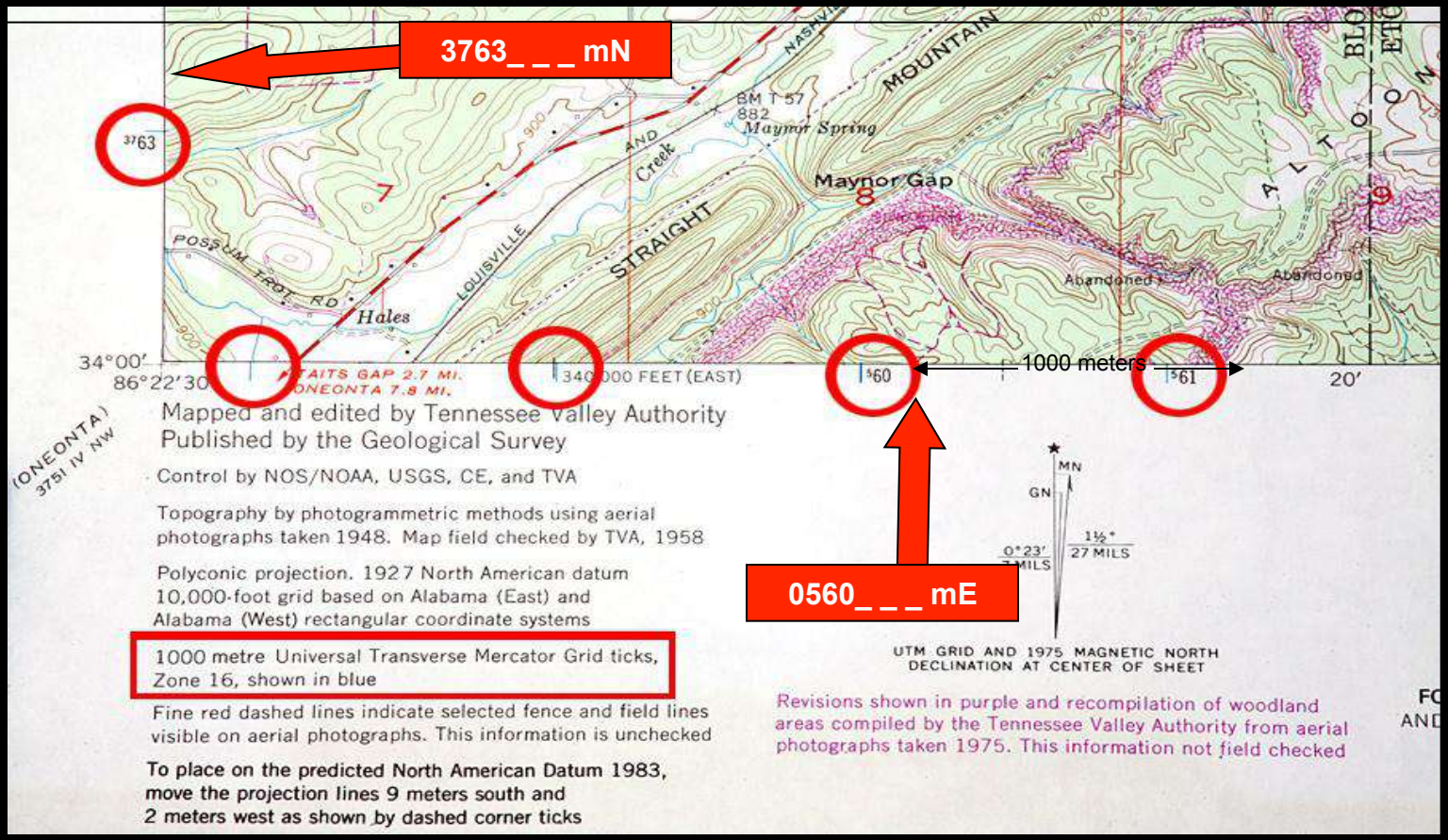

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Easting:

Is a 6-digit number to designate the EAST/WEST position in the grid. The grids are the widest at the equator. Since each grid has a different width, a system was created to measure the distance across a grid starting from the vertical center-line of each grid as opposed to starting from the left or right edge of a grid. The center line of each grid is called the central meridian. The central meridian is always assigned an easting value of 500,000 meters East. It is expressed as 500,000 mE. As you move west of the central meridian the easting will be a number less than 500,000. As you move to the east of the central meridian the easting will be a number greater than 500,000. An easting of zero will never occur, since a 6° wide zone is never more than 674,000 meters wide.

Northing:

The second number is called the Northing. It is a 7-digit number that designates how many meters you are north or south of the equator.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

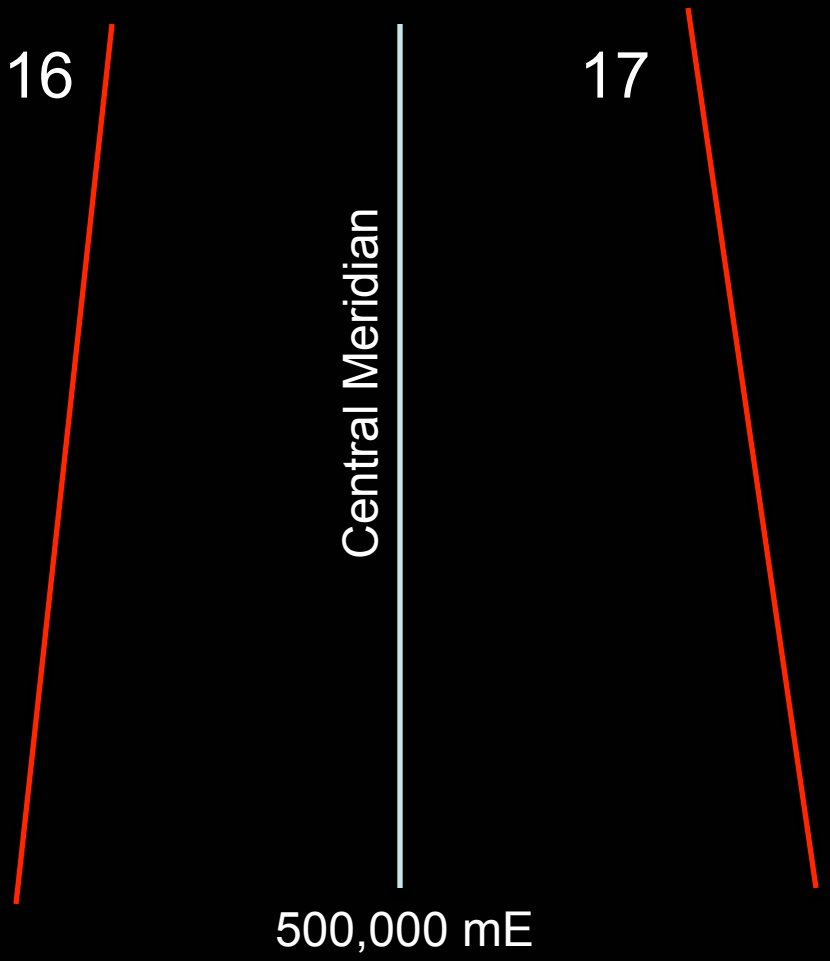

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

10,000 meter digit

100,000 meter digit(s)

1,000 meter digits

16 S 0572450 E

Map Zone

3763650 N

You only have to plot the last 3 numbers. The rest of the coordinate values are provided on the map

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

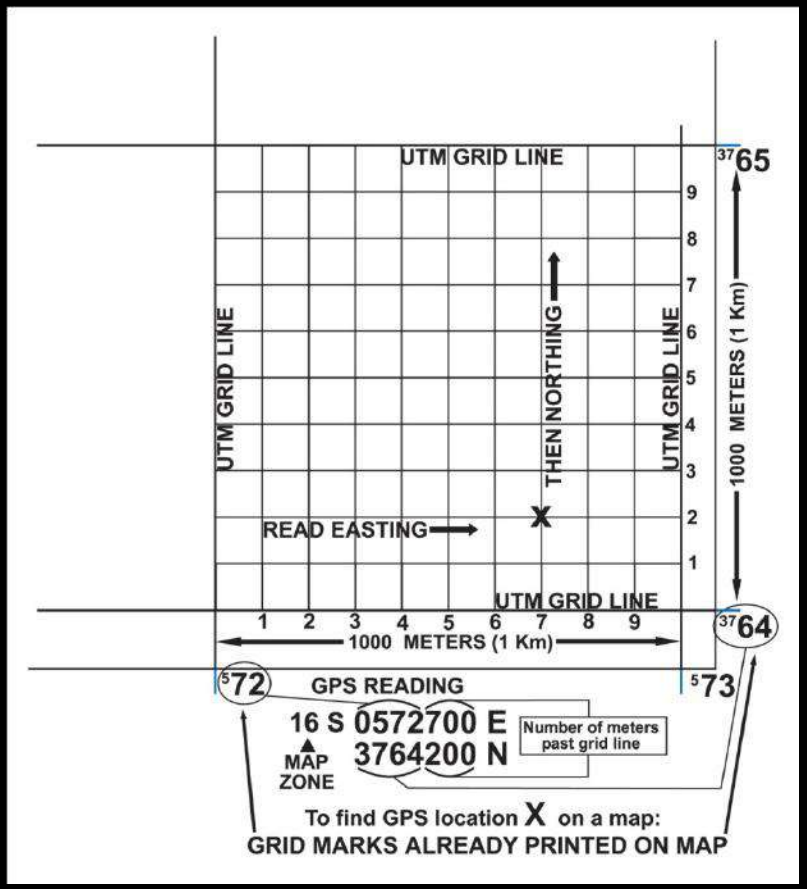

Read EASTING First

Then Read NORTHING

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

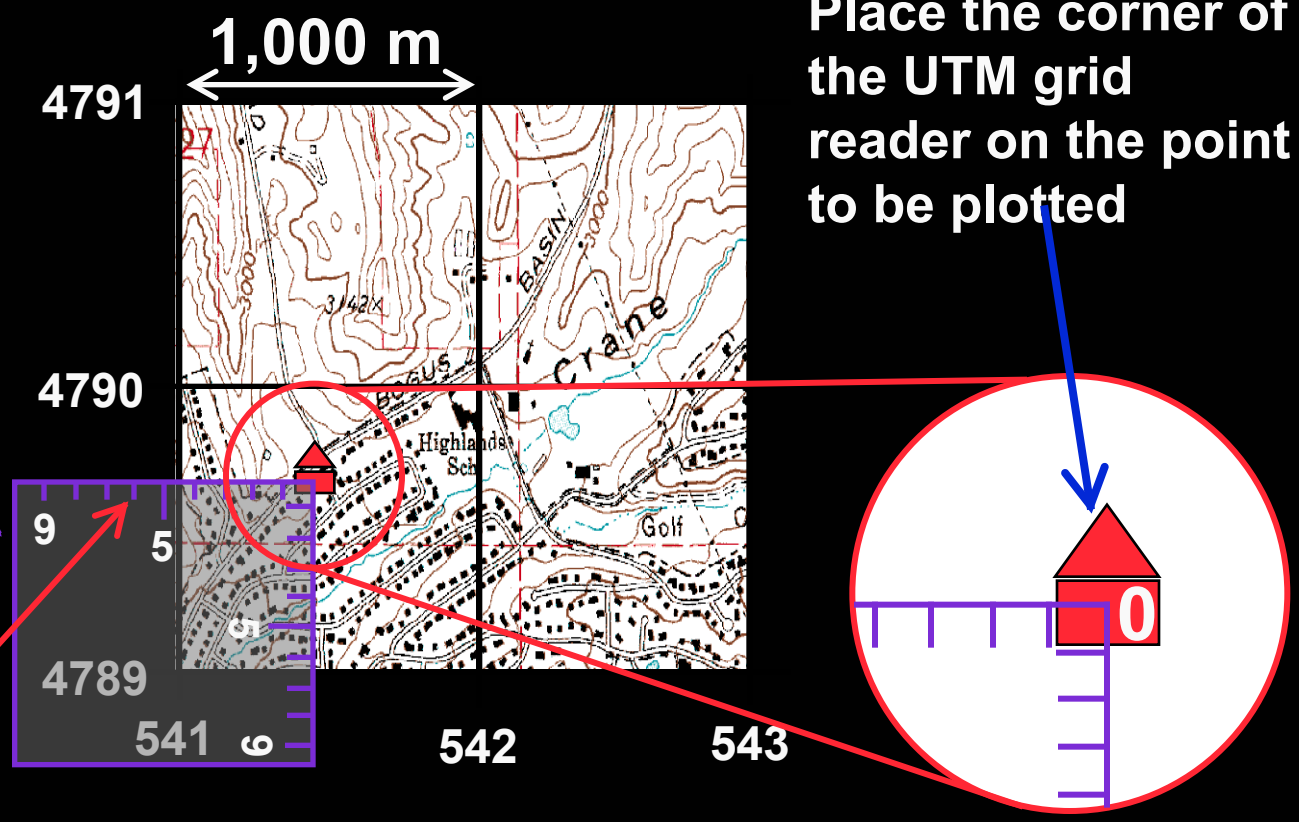

House coordinates = 0541450mE
4789650mN

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Using GPS with Map

When using a GPS to acquire an azimuth to a location, setup your GPS for Magnetic North Heading if you are not adjusting declination on your compass. Use True North Heading if your are adjusting for declination on your compass.

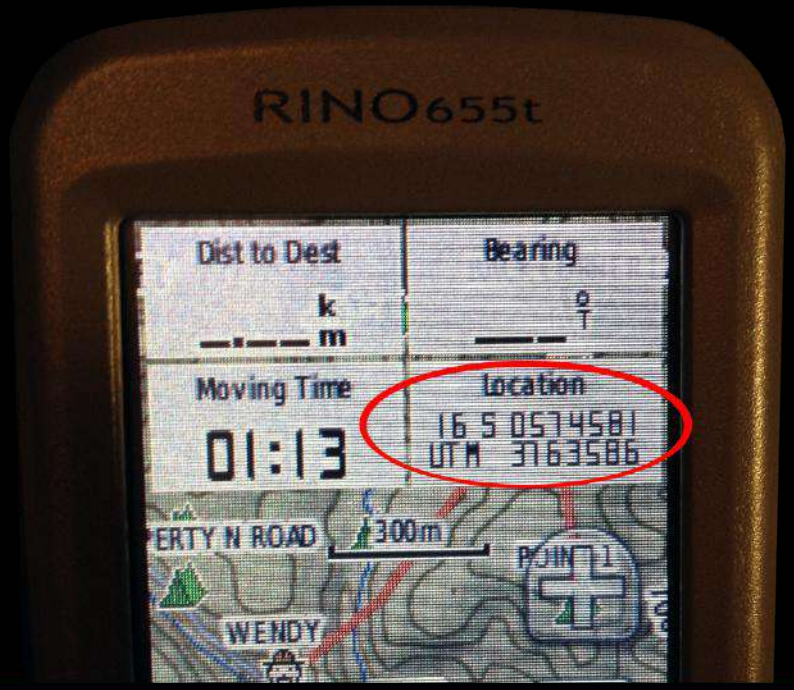

Plotting GPS to Map

1) Find current location with GPS
2) Use scale card to plot current location on the map

Plotting Map to GPS

1) Plot location on map of where you want to go with a scale card
2) Enter plotted coordinates into the GPS as a new waypoint

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

MGRS System

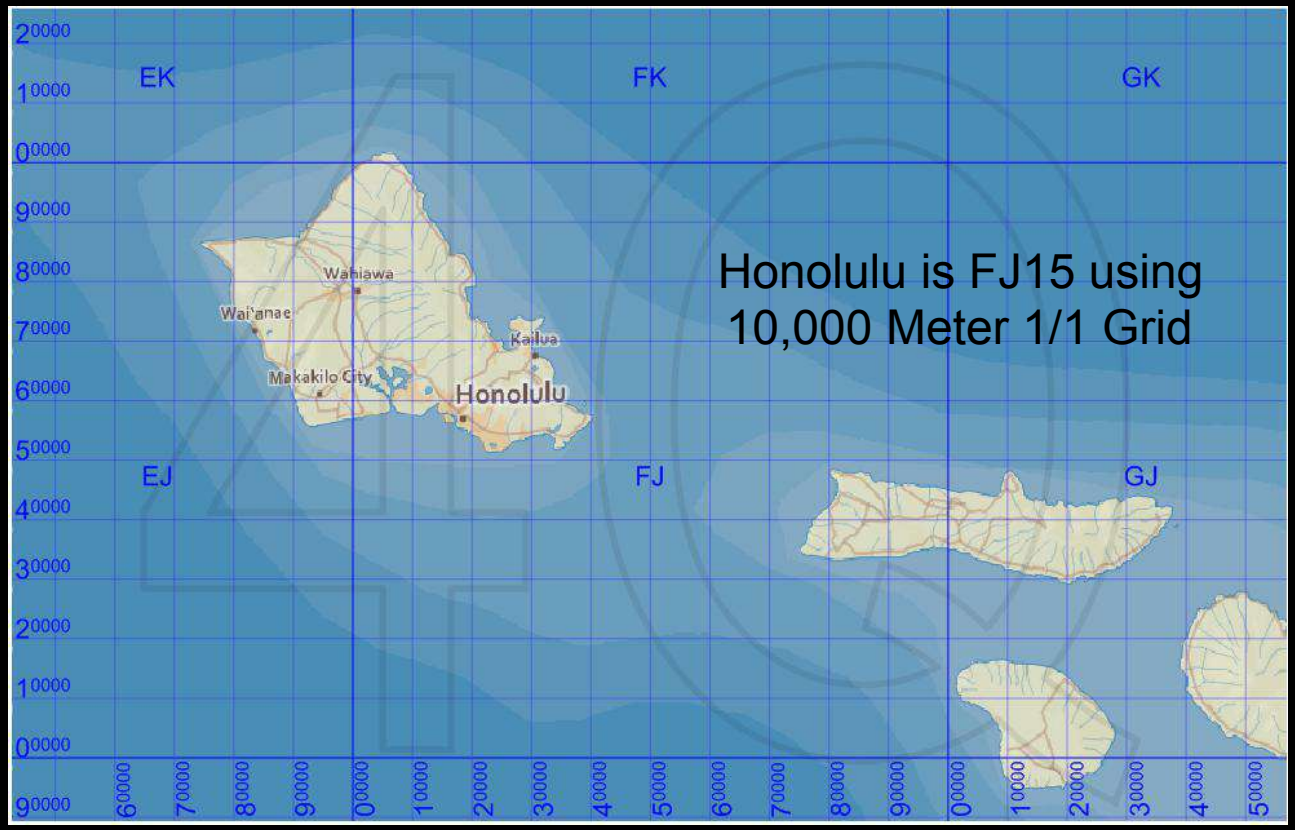

5/5 = 1 Meter, 4/4 = 10 meter, 3/3 = 100 meter,
2/2 = 1000 meter, 1/1 = 10,000 meter

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Why do we not use Latitude / Longitude for Land Nav?

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Communicating Location

General Location
Type Of Navigation Grid Used
Grid Location (Easting Then Northing)
Direction Of Travel If Moving

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Declination

Magnetic Declination is the difference between True North and Magnetic North. A compass needle ALWAYS points to Magnetic North.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

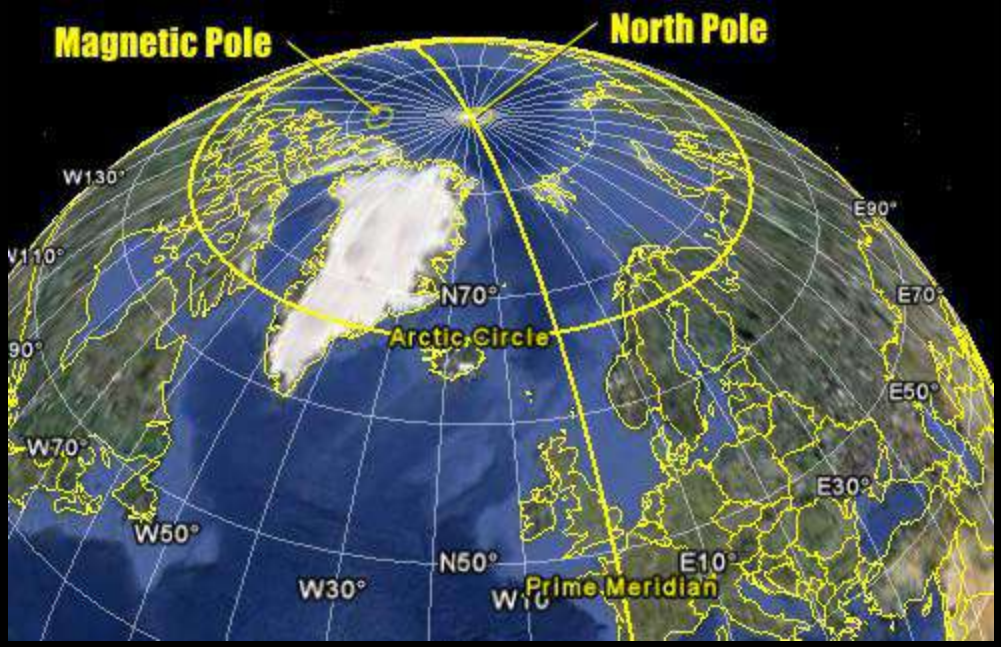

Difference in True North
And Magnetic North

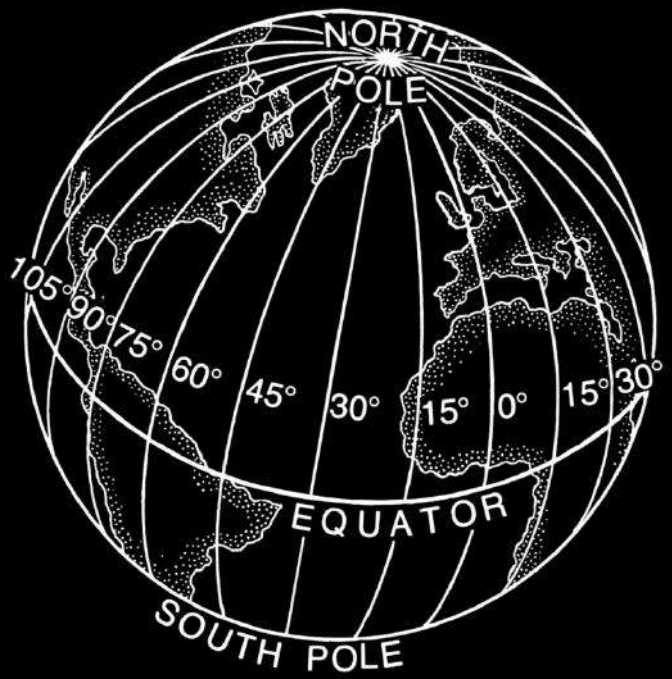

Longitude Lines
(True North)

Magnetic North is moving approximately 40 miles per year.

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Magnetic Declination

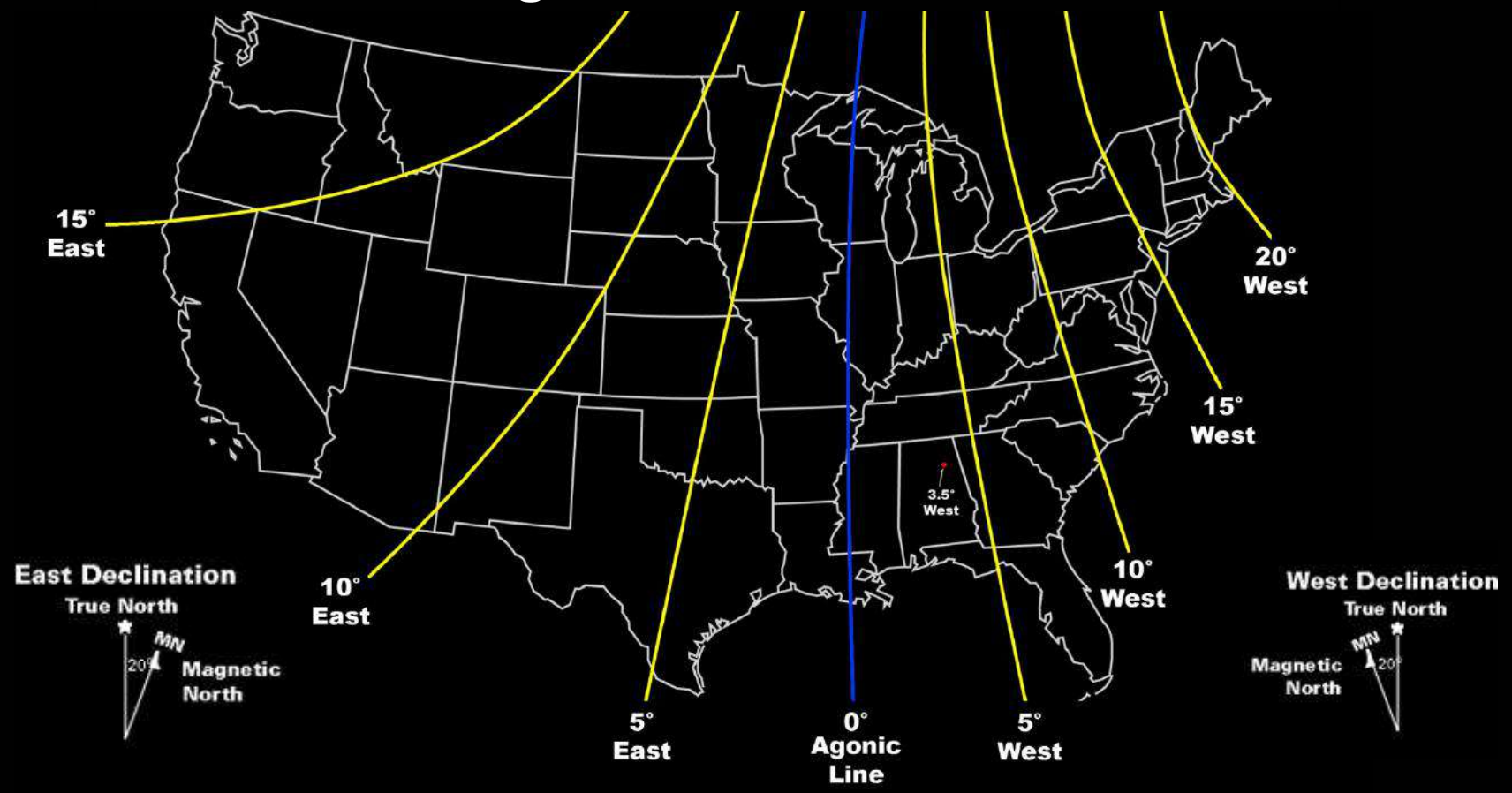

If you do not have a compass that adjusts for declination remember to add or subtract declination to your azimuth. Add for West – Subtract for East
EAST IS LEAST - WEST IS BEST

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Depending on your map, it may printed with True North Reference lines or Grid North Reference lines

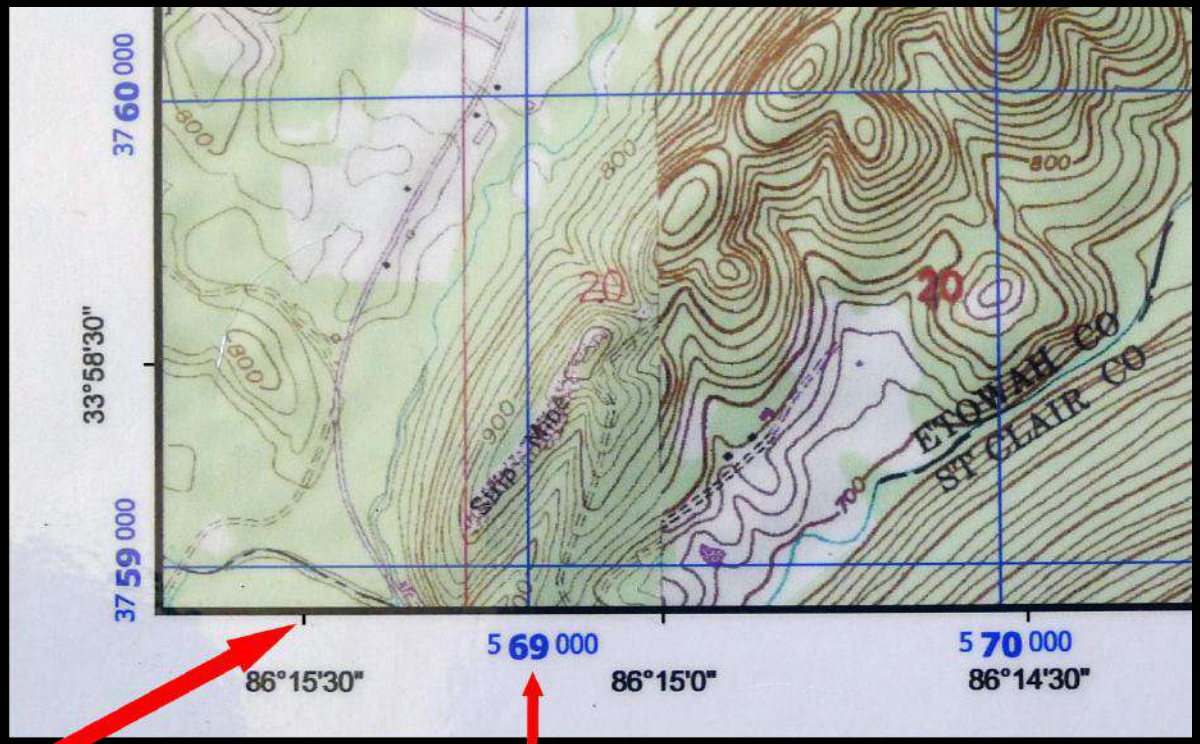

True North
(Longitude)

Grid
North
(UTM Or Other Grid)

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

Adjusting For Declination For Map Use

Declination is the difference between true north and magnetic north when using the longitude lines as north reference.

Declination is the difference or spread between grid north and magnetic north when using the UTM grid lines as north reference.

At the bottom of your map, if the MN line is left of the True North Line it will always be West Declination. If the MN line is right of the True North Line it will always be East Declination

If you are using a compass that is not adjusted for declination, add the amount for west declination, subtract the amount for east declination. If your compass is adjustable for declination, set it for the north reference declination value you are using. No further addition or subtraction is required.

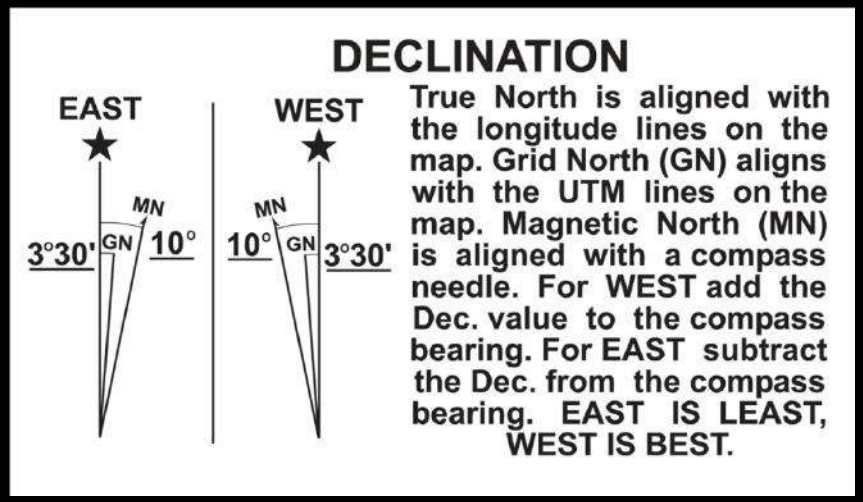

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

True North / Magnetic North / Grid North

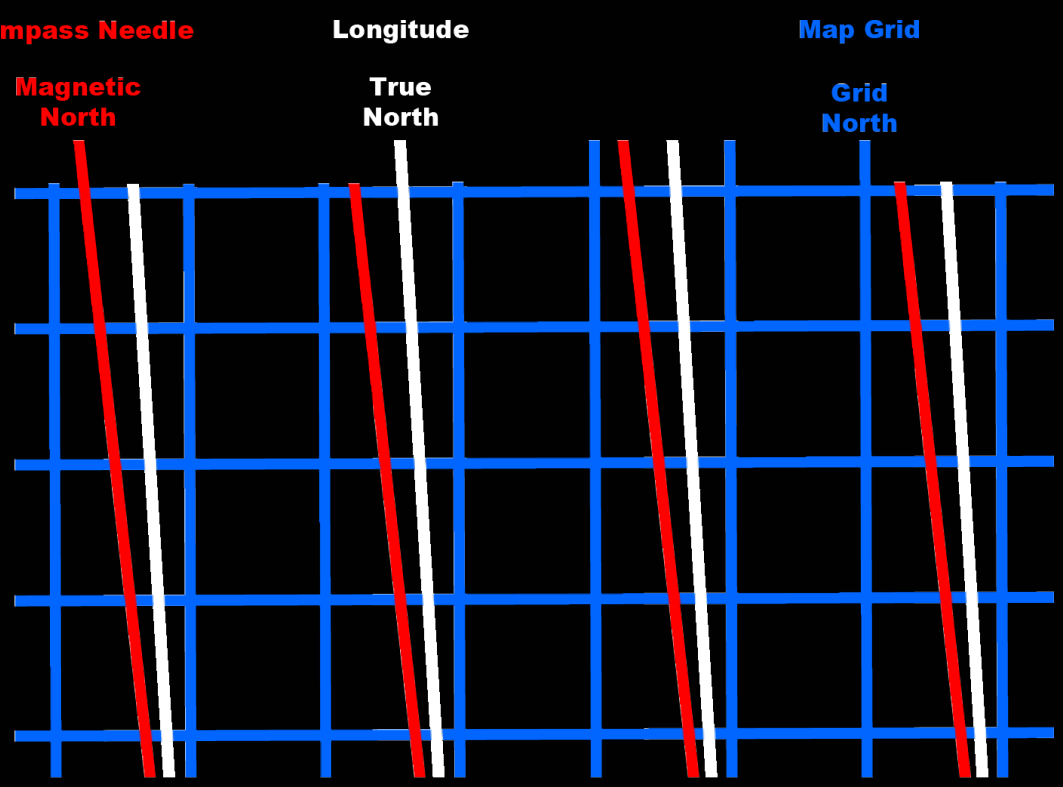

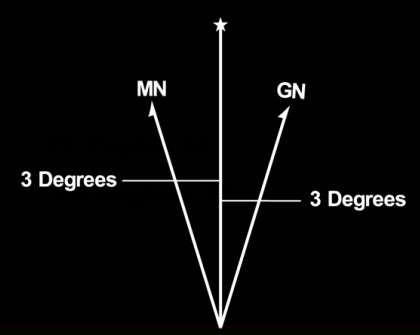

If using Grid North as Reference then declination would be set to 6 degrees West
If using True North as Reference then declination would be set to 3 degrees West

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

RANDALL'S ADVENTURE TRAINING ®
LAND NAVIGATION

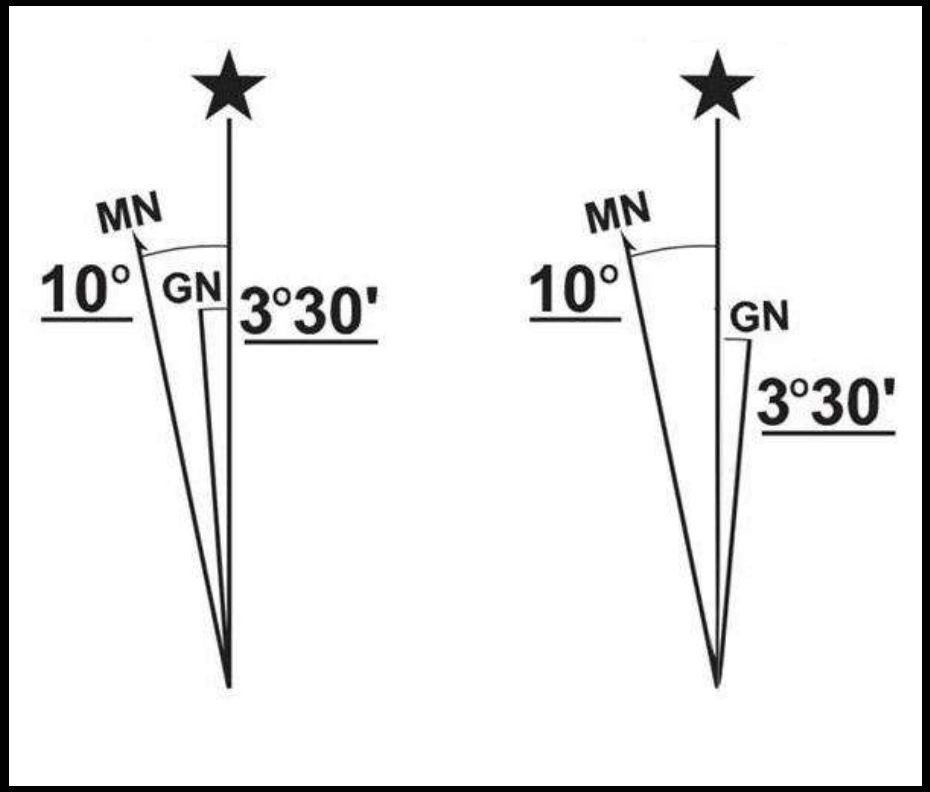

What is our declination value for each of the diagrams to use Grid North Reference?

COPYRIGHT © 2013 RANDALL'S ADVENTURE TRAINING ®

USER ADVICE
this book could save your life

THE SURVIVAL HANDBOOK

ESSENTIAL SKILLS FOR OUTDOOR ADVENTURE

LONDON, NEW YORK, MUNICH, MELBOURNE, DELHI

Senior Editors Nicky Munro, Richard Gilbert

Senior Art Editors Michael Duffy, Gillian Andrews

Editors Bob Bridle, Chris Hawkes, Philip Morgan, Gill Pitts, Clare Weber

US Editor Chuck Wills

Designers Sarah-Anne Arnold, Katie Eke, Brian Flynn, Phil Gamble, Sharon Spencer

Lead Illustrator Mike Garland
Assisting Illustrators Darren Awuah, Phil Gamble, Peter Liddiard, Mark Walker

Production Editor Tony Phipps
Production Controller Norma Weir

Managing Editor Stephanie Farrow
Managing Art Editor Lee Griffiths

Art Director Bryn Walls
Publisher Jonathan Metcalf

First American Edition, 2009

Published in the United States by
DK Publishing, 375 Hudson Street
New York, New York 10014

09 10 11 12 10 9 8 7 6 5 4 3 2 1

SD308—April 09

Copyright © 2009 Dorling Kindersley Limited
All rights reserved

Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form, or by any means (electronic, mechanical, photocopying, recording, or otherwise), without the prior written permission of both the copyright owner and the above publisher of this book.

Published in Great Britain by Dorling Kindersley Limited.

A catalog record for this book is available from the Library of Congress.

ISBN 978-07566-4279-2

DK books are available at special discounts when purchased in bulk for sales promotions, premiums, fund-raising, or educational use. For details, contact: DK Publishing Special Markets, 375 Hudson Street, New York, New York 10014 or SpecialSales@dk.com.

Printed and bound in Hong Kong by Hung Hing

Discover more at
www.dk.com

IMPORTANT NOTICE

Some of the techniques desribed in this book should be used only in dire emergencies, when the survival of individuals depends upon them. The publisher cannot be held responsible for any injuries, damage, loss or prosecutions resulting from the use or misuse of the information in this book. Do not practice these techniques on private land without the owner's permission, and obey all laws relating to the protection of land, property, plants, and animals.

CONTENTS

8 Introduction

1 BEFORE YOU GO

16 **Introduction: Prepare yourself**
18 Getting into shape
20 Mental preparation
22 Planning your journey
24 Emergency plan of action
26 **Introduction: Know your environment**
28 Temperate environments
30 Tropical environments
32 Mountain environments
34 Desert environments
36 Cold environments
38 Marine environments
40 **Introduction: Get the right gear**
42 Choosing your kit
44 Choosing a backpack
46 Dressing for the outdoors
50 Extreme survival—in the wilderness
52 Sleeping systems
56 Eating on the trail
58 Camping stoves
60 Your survival kit

2 ON THE TRAIL

64 **Introduction: Find your way**
66 Maps and map-reading
68 Get your bearings
72 Route finding
74 Navigating without a compass
76 Natural navigation
78 How weather works
80 Weather phenomena
82 Understanding local weather
84 **Introduction: Make a move**
86 Traveling on foot
88 Crossing rivers
90 Scrambling and climbing
92 Extreme survival—in the mountains
94 Moving over snow
98 Using pack animals
100 Four-wheel driving
104 Kayaking and canoeing
106 Building a raft
108 Swimming

3 CAMP CRAFT

112 **Introduction: Camp essentials**
114 Organizing your site
116 Staying clean
118 Making fire
120 The elements of fire
122 Char cloth and fire can
124 Types of fire
126 Making sparks and flames
128 Extreme survival—in the desert
130 Fire by friction
134 **Introduction: Manual skills**
136 Man-made cordage
138 Natural cordage
142 Tying knots
146 Using cutting tools

4 TAKING SHELTER

154 **Introduction: Taking shelter**
156 Hollows and caves
158 Use-anywhere shelters
162 Forest shelters
166 Tropical shelters
172 Extreme survival—in the jungle
174 Desert shelters
176 Shoreline shelters
178 Snow shelters

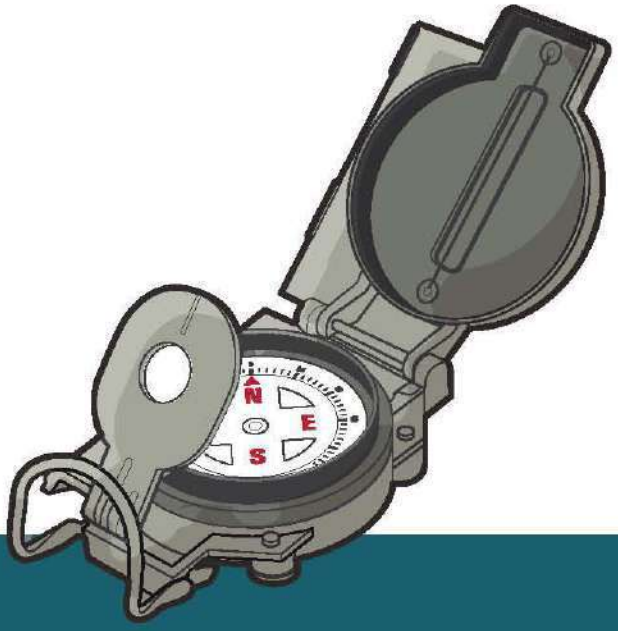

5 WATER AND FOOD

184 **Introduction: Find and treat water**
186 The importance of water
188 Finding water: Temperate climates
192 Finding water: Hot climates
194 Finding water: Cold climates
196 Finding water: At sea
198 Carrying and storing water
200 Treating water
202 **Introduction: Find and prepare food**
204 Wild cooking
206 Edible plants
208 Catching fish
212 Preparing fish
214 Extreme survival—adrift at sea
216 Trapping animals
220 Preparing small mammals
222 Preparing large mammals
224 Preparing other animals
226 Catching birds
228 Preparing birds

6 IN AN EMERGENCY

232 **Introduction: In an emergency**
234 Assessing your situation
236 Attracting rescuers
242 Wild animals
244 Extreme survival—in cold conditions
246 Environmental hazards
250 Surviving at sea

7 FIRST AID

258 **Introduction: First aid**
260 First aid essentials
262 Flesh wounds
266 Bites and stings
268 Poisonous plants and intestinal parasites
270 Bones, joints, and muscles
272 Treating exposure
274 Treating shock
275 Breathing difficulties
276 Unconsciousness
278 Moving a casualty

APPENDIX

Wild food
280 Temperate plants
282 Desert plants
284 Tropical plants
286 Plants in cold climates
288 Coastal plants
290 Mammals
292 Birds
294 Amphibians and reptiles
296 Fish
298 Invertebrates
Natural Dangers
300 Wild animals
302 Insects and arachnids
304 Poisonous plants

306 Resources
308 Glossary
312 Index
319 About the author
319 Acknowledgements

PREPARE YOURSELF

Most survival situations arise in one of two ways: either you are thrust into a situation not of your making and beyond your control, or (as is more often the case) a situation develops because of a sequence of events that could have been avoided had you recognized the danger signs and acted on them at the earliest opportunity. Unfortunately, most survival situations occur as a result of ignorance, arrogance, an overconfident belief in one's own ability, or because the forces of nature have been underestimated.

In the military forces, troops prepare for survival situations by learning the basic principles and techniques and then practicing them until they become second nature. The more you practice a particular survival skill, the more you'll understand how and why it works, and the more you'll understand your own strengths and weaknesses. Many helicopter crew members who have been involved in a survival situation, for example, report that, having ditched their aircraft in the sea, they could not remember how they subsequently

In this section YOU WILL DISCOVER...

- how keeping **in shape** can keep you **out of trouble...**
- the importance of a **positive mental attitude...**
- that where there's a **will to survive** there's a **way to survive...**
- the difference between a **coping strategy** and an **aggravating factor...**
- how **"show-stoppers"** could **scupper your trip...**
- why an **emergency plan of action** could **save your life...**

escaped from their aircraft and successfully boarded their liferaft. Thanks to their intense training regime, their actions had become instinctive and therefore subconscious.

Whether you're preparing for an overnight camping trip or a year in Africa, the more prepared you are to meet the challenges posed by a particular environment, the more likely you are to be able to cope—both physically and mentally—should you then find yourself confronted with a survival situation.

It would be foolish to think that the amount of prior preparation required for an expedition is directly related to the length of the trip or its perceived danger.

ACCIDENTS CAN HAPPEN anywhere. One of the most important things you can do to improve your chances of survival is to let people know where you're going and when you'll be back. Leave an itinerary of your trip with family or friends, and arrange a time when you will call to let them know you are safe.

In many instances, simply making sure you have a cell phone with you can prevent an accident becoming a major survival incident. In 2007, a professional athlete, out running with her dog in Moab, Utah, slipped on ice, fell approximately 60 ft (18 m), and broke her pelvis. Having left her cell phone in her vehicle, she managed to survive for two nights in the freezing desert before rescuers were finally lead to her by her dog.

In a separate incident in 2008, a farm worker in the UK caught his arm in machinery. Unable to attract attention and not having his cell phone with him, he was faced with a stark choice when the machinery caught fire: bleed to death, burn to death, or cut off his own arm with a pocket knife—he chose the latter.

> "Never assume that the amount of **prior preparation** required is directly related to the length of a trip or its **perceived danger**"

GETTING INTO SHAPE

IN ANY SURVIVAL SITUATION you're likely to be at your best, both physically and mentally, in the moments before the situation occurs. From that point on, through lack of sleep, food, and water, your situation will deteriorate until your rescue. Being in good physical condition will help you overcome the challenges you'll face in a survival situation.

KNOWING YOUR LIMITS
Having an idea of how fit you are and knowing how much you are capable of achieving before entering the wilderness means you can set yourself realistic goals—which will help to keep you out of danger.

THE BENEFITS OF EXERCISE

Over time, exercise induces changes such as weight loss, as well as improved posture, physique, strength, agility, mental alertness, and stamina. These are all vital attributes should you find yourself in a survival situation.

HOW FIT ARE YOU?

A fit heart pumps blood more slowly and efficiently than an unfit one. The hearts of women, children, and older people beat faster than those of young adult males. Take your pulse first thing in the morning—this is known as the "base rate." As the table below shows, the faster your pulse returns to its base rate after a period of exercise, the fitter you are. Note that < denotes "less than" and > denotes "more than".

AGE	20-29	30-39	40-49	50+
RATING	BASE RATE			
MEN				
Good	<69	<71	<73	<75
Average	70-85	72-87	74-89	76-91
Poor	>85	>87	>89	>91
WOMEN				
Good	<77	<79	<81	<83
Average	78-94	80-96	82-98	84-100
Poor	>94	>96	>98	>100

AGE	20-29	30-39	40-49	50+
RATING	BEATS AFTER EXERCISE			
MEN				
Good	<85	<87	<89	<91
Average	86-101	88-102	90-105	92-107
Poor	>101	>102	>105	>107
WOMEN				
Good	<93	<95	<97	<99
Average	94-110	96-112	98-114	100-116
Poor	>110	>112	>114	>116

FIT FOR THE CHALLENGE

Survival fitness is not about trying to run the 100 meters in 10 seconds; it's about stamina and endurance, and about knowing your physical limitations and being able to work with them. It is also about understanding that you may have to push yourself beyond your limits and that having a positive mental attitude is paramount to survival—the mind will often give up long before the body has reached its limit. Bear in mind that every task in a survival situation will sap your energy.

EXERCISING REGULARLY

Any effective training schedule must include at least three 45-minute sessions per week, with exercises that raise the heartbeat to more than 120 beats per minute.

STARTING A FITNESS PROGRAM

To spend time in the wilderness with a pack on your back requires a combination of strength and aerobic fitness. Training in the gym before you set off will strengthen your heart, lungs, and leg muscles, and will increase your stamina levels when you are out on the trail.

- Seek expert instruction and guidance wherever possible.
- Start gradually and then build up your routine progressively.
- Never try to work through injuries—rest and seek medical advice.
- Devise a relevant fitness program; there are many websites and organizations that provide you with detailed fitness programs for various levels of specific activities, from training for a trek through the jungle to mountain-biking along the Appalachian Trail.
- Duplicating in training what you'd like to be able to achieve on the trail will allow you to build up reference points about how your body works, and how it copes in different situations. The more you know about how you perform, the better equipped you'll be on the trail.

WARMING UP AND COOLING DOWN

Time spent warming up and cooling down after exercise will help to improve your endurance levels and will also accelerate the recovery process. Get into the habit of starting and ending your session with a five-minute jog.

GO FOR A CHECK-UP

It's a sad fact that the majority of us don't have regular medical check-ups. Many people work on the principle that you only need to see a doctor or dentist if something is wrong. However, you should always make sure that you start any adventure or wilderness trek in top condition. Therefore, before you set out, pay a visit to your doctor and dentist and make sure that your body and teeth are in good condition.

VISIT THE DOCTOR

- Let the doctor know of any ailments that have been a concern to you during the course of the previous year.
- Tell the doctor where you are going and enquire about relevant inoculations or medication that you should be taking with you.

VISIT THE DENTIST

- Have any problems with your teeth dealt with before you go. Any minor dental problem will almost certainly turn into full-blown toothache when you least want it to.

USEFUL EXERCISES

If you already get regular exercise, you'll find yourself far more comfortable in a survival situation than your more sedentary counterparts. Any pre-expedition training regime should include plenty of stretching, aerobic exercise, and weight-training.

STRETCHING YOUR UPPER BODY AND BACK

Regularly stretching your arms, neck, chest, and shoulders will help your body to maintain a strong core. This will be of great use when it comes to activities such as rock climbing or using trekking poles. Undergoing a regular back-stretching routine will loosen the back muscles, making them more flexible and less susceptible to injury, and will increase both the back's range of motion and its endurance.

STRETCHING YOUR LEGS

Because your leg muscles bear the brunt of the work, leg stiffness is a common complaint at the end of a long day on the trail. Stretching your legs will improve your flexibility, increase blood circulation, and relax your muscles. Concentrate on your calves, quads, and hamstrings.

AEROBIC EXERCISES

Also called cardiovascular fitness, aerobic fitness refers to the ability of your heart, blood vessels, and lungs to supply oxygen and nutrients to the rest of your body during sustained physical activity. Regular aerobic exercise—such as swimming, jogging, or cycling—will reduce the risk of developing coronary heart disease or high blood pressure, will help you to manage your weight and increase your stamina levels, and will make your heart stronger and more efficient, thus improving blood flow around your body.

WEIGHT-TRAINING

Increased muscle strength will enhance your ability to perform everyday tasks such as lifting, carrying, and walking. A regular weight-training program will improve your posture, build muscle density around the joints and bones, improve your sense of balance, help your stress management abilities, aid sleep, and reduce the chances of you sustaining an injury. In addition, research has shown that a regular resistance-training program can improve your metabolic rate by up to 15 percent, which will help you shed any unwanted weight at a faster rate.

MENTAL PREPARATION

REGARDLESS OF WHETHER you're backpacking through a remote area, or on a day hike in familiar territory, a situation may arise that changes your circumstances for the worse. Very quickly you move into the unknown, which causes tremendous psychological and emotional stress, known as "psychogenic shock." Understanding this will help you to deal with it better and reduce its impact.

INDIVIDUAL REACTIONS TO DISASTER

People react to survival situations in different ways, although you can expect to find some common emotional reactions in victims who are experiencing, or have survived, a disaster situation. You may experience one or more of them during or after any survival experience or trauma.

YOUR RESPONSE TO DISASTER SITUATIONS

Your psychological response to a survival situation is crucial. Statistics show that 95 percent of people who die with psychological trauma die within the first three days. Losing the will to survive—or suffering psychological disorders that prevent you from coping with the physical conditions—is your main concern. If you break down psychologically, your chances of overcoming a situation will be compromised.

PSYCHOLOGICAL PROGRESSION

It's useful to examine how people are likely to react in a survival situation. Using this knowledge, it's possible to prepare mentally for such eventualities and, in so doing, lessen the impact if the worst should happen to you when out on the trail. Normal psychological reactions to disaster tend to occur in a set pattern of four stages: the pre-impact period, the impact period, the recoil period, and the post-trauma period (see right). Contrary to popular belief, people don't normally panic, although it can be contagious if someone does.

AGGRAVATING FACTORS	HUNGER	THIRST
Reactions to disaster can result from a direct blow to the psychological system, such as extreme shock, but they can also be brought on, or aggravated by, other factors. As with all psychological problems, knowing what these aggravating factors are, and attempting to avoid them—or at least recognizing what they might lead to—will maximize your chances of preventing or overcoming the problem. The most common aggravating factors are: hunger, thirst, fatigue, seasickness, and hypothermia.	Initially, hunger is not a problem, but a long-term lack of food will cause psychological changes to occur. Symptoms include: ▪ Apathy ▪ Irritability ▪ Depression ▪ Lack of concentration	Thirst is a serious problem, especially for survivors at sea or in the desert, and its effects are more acute than hunger. Agitation is commonplace; other symptoms include: ▪ Irrational behavior (see box, right) ▪ Delusions ▪ Visual hallucinations

COPING STRATEGIES	TRAINING	MOTIVATION
There are many things you can do to prepare yourself psychologically for a survival situation—such as learning about what you should expect if the worst were to happen—and to cope better if you find yourself in one. As with all survival skills, prior knowledge is power, and will help you to deal with a survival situation far more successfully. The main areas to think about are: training, motivation, attachment, hope, acceptance, and helping others. Developing coping strategies is an important technique for survival.	People who are properly prepared, who know their environment and how to use their equipment, and have an understanding of what to expect in a survival situation, will be far more effective if they find themselves in one. Adequate training and practice using your equipment will help you to function effectively at an automatic level. Prior knowledge is key to your survival.	Often known as "the will to survive," motivation involves a refusal to accept death, and to hang on to the belief that you were not meant to die under these conditions. It involves overcoming the emotional and physical discomforts of extreme conditions. Linked to motivation is the ability to establish goals, work out the steps to those goals, and to follow those steps through.

PANIC
Panic arises from the fear of what might happen rather than what has happened. It tends to occur when people are trapped, or if there is a time limit to their escape.

HYPERACTIVITY
Hyperactive victims are easily distracted, and are full of chatter, ideas, and often unhelpful suggestions. Sufferers can reach this stage after a state of depression.

GUILT
Some sufferers feel guilty for surviving, and for not having done enough for others—and some irrationally blame themselves for bringing about the incident.

DEPRESSION
Depressed people will sit among chaos and debris vacantly gazing and not replying to questions. They're unaware of their situation and unable to help themselves, so risk further injury.

ANGER
Aggression, anger, and hostility are common reactions to trauma. They're often irrational and may even be directed at the rescuers or medical staff trying to help them.

SUICIDE
Disaster victims have been known to commit suicide immediately after being rescued, in some cases when they're already safe in the hospital. Victims should be closely monitored.

Pre-impact period
The "pre-impact" period is divided into two stages:
- Threat: danger exists but, though obvious to those who recognize it, those who will not accept it respond with denial and under-activity.
- Warning: threat of danger is now apparent to all; response is now likely to be over-activity.

Impact Period
This is the life-threatening stage. Statistically, individuals behave in one of three ways:
- 10–20 percent of people are calm and retain full awareness.
- Up to 75 percent of people are stunned, bewildered, and unable to react rationally.
- 10–25 percent exhibit extreme behavior, such as screaming.

Recoil period
This follows on directly from the impact period; for example, victims may have escaped a sinking ship and are in the liferafts. It can last for up to three days, but generally lasts for around three hours. In most cases, it is characterized by a gradual return to normal reasoning abilities, awareness, and emotional expression.

Post-trauma period
If the recoil period is not fully successful, individuals may develop psychiatric disorders. The full impact of the incident becomes apparent and a range of emotions—guilt, depression, anxiety, aimlessness, and a feeling of bereavement—may develop. These are often referred to as Post-Traumatic Stress Disorders (PTSD).

FATIGUE
In many cases, physical exhaustion is present from the outset. At other times, it may result from sleep deprivation and the physical hardship endured over time. Most survivors agree that fatigue overwhelms them, but when they want to sleep they can't—they have an inability to relax. Fatigue causes a deterioration in mental and physical performance, followed by a psychological and physical debility.

SEASICKNESS
Seasickness often brings about an overwhelming desire to curl up and die, which in survival situations can easily become a reality. It's important not to give in to this urge. Fight seasickness with the following methods:
- Keep a fixed point such as the horizon in sight.
- Take small sips of water (not salt water) if you have sufficient supplies—but ration them if you're in a life or death situation.

HYPOTHERMIA
Hypothermia (see p.273) produces both physical and psychological effects—the psychological consequences occur early in the condition, and cause:
- Loss of concentration
- Loss of memory
- Motor impairment
- Faulty decision-making
- Irrational behavior

IRRATIONAL BEHAVIOR
Irrational behavior can take many forms. Examples include the earthquake victims who were found collecting flowers instead of helping the injured, and famously, the band of the *Titanic*, who played while the ship sank rather than trying to save themselves.

ATTACHMENT
One of the strongest motivating forces for survival is the desire to be reunited with principal figures of attachment in your life. These may include:
- Husbands
- Wives
- Partners
- Children
- Grandchildren
- Close friends

HOPE
To hope means to entertain ideas that a distressing situation will improve and get better. In any survival situation it's important to cling on to hope, despite information or perceptions to the contrary. Thinking positively will help to ward off psychological trauma. It's often easier to be optimistic in a group situation than it is if you're on your own, as people can support each other.

ACCEPTANCE
An inability to accept one's situation or condition leads to frustration, anger, and irrational behavior, and it's important to avoid these feelings in a survival situation. The ability to accept the situation doesn't equate to giving in to it—far from it. Those who have this ability, and know when to be active and when to be passive often have a better chance of long-term survival.

HELPING OTHERS
First, monitor your own condition and check that you're really up to the task. Determine who is genuinely disturbed rather than showing "normal" reactions. Psychological first aid is only required for those who are failing to recover. Simple words of comfort and interest will make the majority who are numbed more responsive. Those who are disturbed should be monitored closely. Avoid giving sedatives.

PLANNING YOUR JOURNEY

WHETHER YOUR TRIP involves a day out with your backpack, or a full expedition over weeks, crossing continents in four-wheel-drive vehicles, you need to plan it very carefully. The plan for your day out won't be as detailed as it would be for the longer expedition, but it's equally important. It's a good idea to have a basic planning outline for your most regular trips, to which you can add supplementary information as your trips become more involved.

WARNING!
The social customs of a country are not the same all over the world. Many cultural differences stem from ancient cultural or religious pasts, and are taken extremely seriously. While ignoring or breaking some customs may lead only to embarrassment, breaking others can result in a penalty, punishment, or even imprisonment. Cultural differences may require females in your group to dress so that they don't show bare arms or legs, or they may dictate which hand you use to greet someone politely. Always research a country's customs thoroughly when planning your trip.

MINIMIZING THE "IF ONLY"

There's no way that you can plan for every eventuality—there are just too many variables—but what you can do is look at the type of trip you intend to take, and ensure that if a situation arises you're not left wishing you'd done something differently. The time to minimize the chances of an "if only" situation is during the planning stage. Look at the potential problems and risks, plan to avoid them, and equip yourself with the knowledge and/or equipment to deal with them if they arise.

THE SIX P'S

Remember the six p's: **P**rior **P**lanning and **P**reparation **P**revents **P**oor **P**erformance. Research has shown that the longer and more complicated the trip is, the less likely there is to be a major "survival" situation. This type of trip is likely to be well organized, and potential problems will have been considered. This means that they can either be avoided, or that there will be mechanisms in place to deal with them. In many ways, simply having a good understanding of how to deal with a situation, and being able to interpret the basic principles of survival, can prevent a minor problem from escalating into a disaster. In a survival situation, it may be your knowledge, combined with your ability to improvise, that determines whether you become a survivor or a statistic.

ORGANIZATIONAL PRIORITIES

When planning for a trip always start with the most important things—known as the "show-stoppers." These are generally the things that would actually stop the trip from happening in the first place if not organized in advance. If you then work backward from the show-stoppers to items that would simply make the trip more comfortable, most other things will fall into place. The chart on the right lists the things that you definitely cannot do without.

SHOW-STOPPERS	
MONEY	▪ You'll need enough to cover your needs plus extra for emergencies. ▪ It must be in the correct currency for the country you are visiting. ▪ Ensure your ATM card is set up for use in that particular country. ▪ Make sure you have a secure way of storing your money.
PASSPORT	▪ Your passport must be valid. If you need to renew it, do so well in advance. ▪ Some countries require your passport to be valid for a number of months after your visit. ▪ Keep your passport number in several places, such as inside your survival kit. ▪ Keep your passport safe, and in a waterproof container, such as a Ziploc bag.
VISA	▪ Research the regulations regarding visas relevant to the countries you're traveling to. ▪ Find out how to apply, how far in advance you should apply, whether or not you can apply in the country, and what other documents you need. ▪ As with your passport, always keep your visa safe, dry, and in good condition.
VACCINATION	▪ Check the regulations for the country—many have strict vaccination policies. ▪ Ensure you have the vaccinations, and any boosters, within the correct timeframe. ▪ Some inoculations last only for six months, so for extended trips you might have to have more while away. This can usually be arranged through local hospitals or clinics.
TICKETS	▪ Make sure you have the correct tickets for your journey. ▪ Read the tickets and make sure they have the correct names, dates, and locations. ▪ Should you need to prove your movements, always keep your tickets in a safe place and never throw them away—often the return journey is included on the same ticket.
INSURANCE	▪ It's advisable to get insurance against your trip being cancelled. ▪ Ensure that your insurance will cover your medical expenses should the worst happen. You don't have to be climbing Everest—even a twisted or broken ankle when hill-walking could prove to be very expensive if you don't have adequate insurance.

YOUR TEAM

If you're embarking on a trip with a group, remember that team dynamics will play an important part in the success, or otherwise, of your trip. Stressful situations—and particularly survival situations—can bring out the best, or the worst, in people. When planning a long trip, it's always a good idea to plan several shorter trips beforehand, to use as practice sessions. These will not only help you decide what equipment to take, and provide you with an opportunity to practice using it, but will also allow team members to assess how they work together as a group. This can help you to organize your team more effectively.

MIXED-GENDER GROUPS

If the team has both male and female members, it's important to factor this into the planning stage. You'll need to consider sleeping and washing arrangements, who carries what, who is responsible for what, and so on. All this should be planned in detail before you set off. It should also be noted that taking it for granted that the women will cook the meals while the men make the shelters is not the best way to start an expedition.

MIXED-AGE GROUPS

When planning your trip, remember that different age groups may have different levels of fitness, which will affect the pace of your group. However, what older members may lack in pace, they may make up for in expedition experience.

PREPARATION THROUGH TRAINING

Training yourself both mentally and physically—and practicing with the equipment you'll be using—may seem like the obvious thing to do, but in many cases the obvious can often get overlooked. You'll get the most out of your trip if you're mentally and physically prepared to a level that means you can operate within your own capabilities and comfort zone. This will allow you to enjoy and appreciate the experience, as opposed to just getting yourself through it.

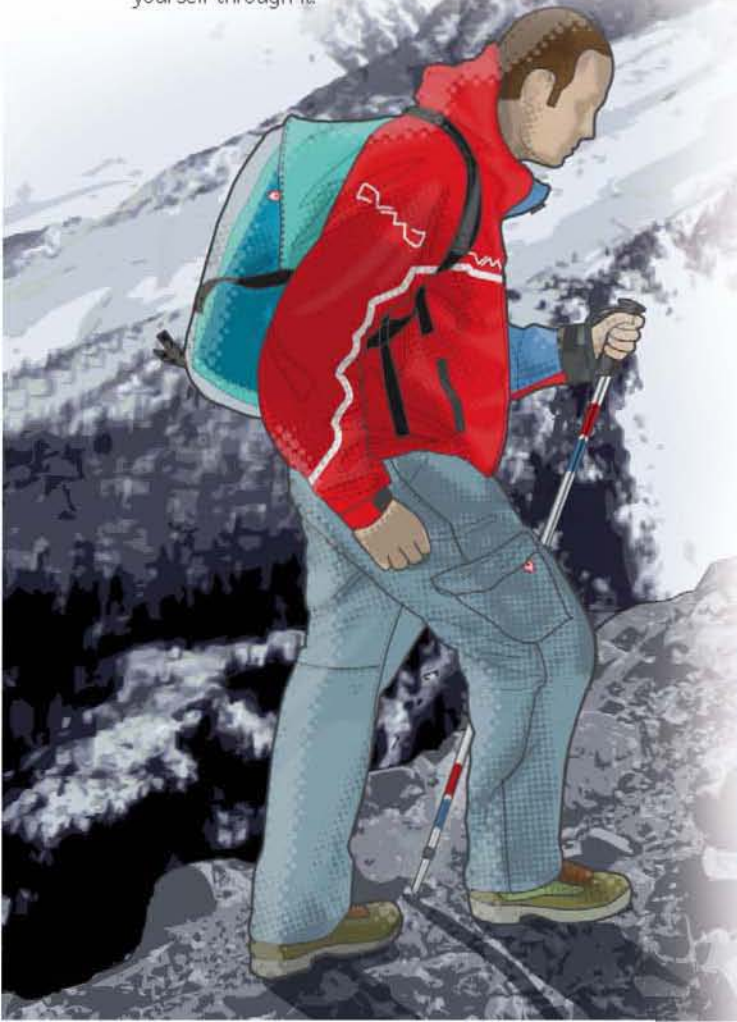

TRAINING YOURSELF

Duplicate in training what you intend to do on your trip. Build up your training gradually, over a period of weeks and months, and take into account the following:

- **The environment**: research the weather conditions you'll be facing, and look at the extremes as well as the average. For example, desert areas may be hot during the day, but can drop to below freezing at night.
- **Weight:** increase the amount of weight you carry until you're eventually carrying what you intend to take. This will not only condition you to the weight but also help you to decide just what's important to take and what's not.
- **Distance:** if your trip involves covering a certain distance a day, then train for that distance. This will give you an indication of whether it's achievable and sustainable.
- **Language**: if visiting a country in which your native language is not widely spoken, it will be beneficial if you can learn some useful phrases. Take a phrase book or an electronic translator.

TRAINING WITH EQUIPMENT

Use your equipment as much as you can, and find out the best way to operate it through practice under realistic conditions (see box, below). This will highlight its strengths and weaknesses, and allow you to determine both its capabilities and your own—the compass may be working perfectly, but practice may highlight the fact that you're not confident enough using it, in which case you'd need to do further training before you set off on your expedition. The wrong time to be trying to find the jack and spare wheel on your rented vehicle is at night, in the rain, with the mosquitoes looking at you as their next free meal. Before you set off, always think about the various skill demands required, and ensure that you're capable of addressing those demands.

OPERATING UNDER REALISTIC CONDITIONS

When training with new equipment, always practice using it under realistic conditions. For example, if you're going to be using your GPS in cold conditions, can you operate it using the gloves you have selected? If you're pitching a tent, have you got all the necessary components, and can you put it up in the dark and rain?

EMERGENCY PLAN OF ACTION

UNFORTUNATELY, EVEN THE MOST meticulously planned, thoroughly equipped, and best-executed trip can run into difficulties. Unforecasted strong winds could trap you and your kayak on an island overnight, or a sprained ankle could leave you unable to climb down the rocks you climbed up so easily. Each of these survival scenarios are difficult to predict, but could easily happen.

USING YOUR EPA

An EPA should contain up-to-date information about you as a person. This may change very little over the years, so you may only need to create a standard EPA and just update the details specific to each trip.

RUNNING INTO PROBLEMS

It's important to make sure that you have done all you can to help yourself, and help any rescue attempt that might take place if you do have any problems. Always remember that there are two sides to any survival or rescue situation: the part that you play and the part that the rescue services play. However, the effectiveness of the rescue services can be greatly increased when they have access to all relevant information—in many cases this arrives too late. It's vital to keep people informed of your intended whereabouts, so that they can raise the alarm if you deviate from your plans.

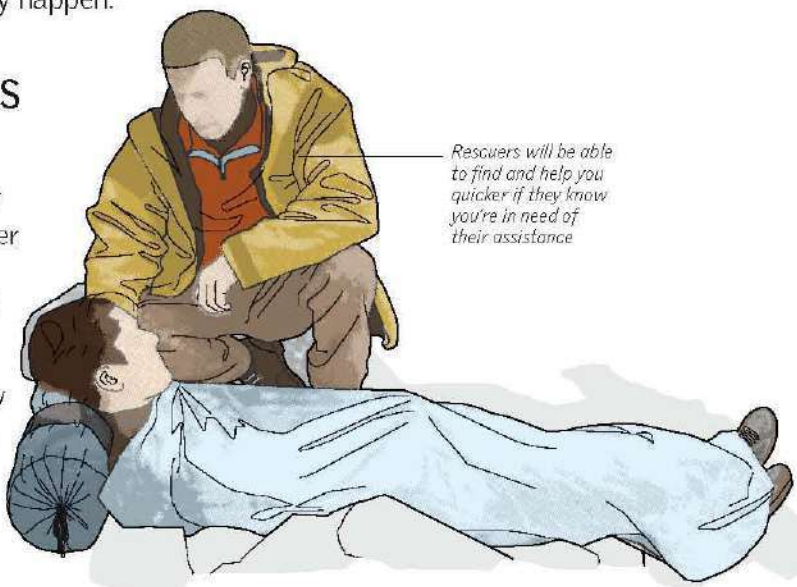

Rescuers will be able to find and help you quicker if they know you're in need of their assistance

WORST-CASE SCENARIO

In the military, every mission that's undertaken, particularly in a theater of conflict, has plans in place for the worst-case scenario. Each part of the mission is meticulously planned, and the team completes a form that states what their basic intention will be should anything happen during the various stages.

IF THE WORST HAPPENS

Having planned for the worst, should the team find themselves in difficulty, the rescue group will not have to try to second-guess what the team will do—they will have a clear indication of the team's intent and can plan effectively and decisively around this information. The team will be found quicker as a result. It's a good idea to apply this principle to your own trip.

KEEPING PEOPLE INFORMED

A good way of keeping your friends, family, and the relevant rescue services informed of your whereabouts is to write down the details of your trip, including pertinent itinerary places and dates, so that if you don't reach your destination when planned, the alarm can be raised. Similar to the "worst-case scenario" principle practised by the military (see left), an "emergency plan of action" (EPA) should be prepared by anyone venturing into the great outdoors (see right). Give a copy to your next of kin and group members, and keep a copy on your person. Where applicable, inform local services, such as park and ranger stations, of your intentions. Remember to inform these people when you arrive safely at your destination.

WRITING AN EMERGENCY PLAN OF ACTION (EPA)

The best way to determine what should be included in your plan is to look at the worst-case scenario you could find yourself in, and ask yourself what information your next of kin would need to know about you and your intentions if they felt they had to raise the alarm. In addition, if you did go missing—particularly in another country—there are many things that the relevant rescue services would need to aid their efforts, for example a recent photograph, your passport details, what equipment you have with you, what languages you can speak, and your skills. The clearer their understanding of you, your abilities, and your intentions, the easier their job will be. In addition, the more information your next of kin have, the more proactive they will feel in your rescue.

EMERGENCY PLAN OF ACTION FORM

Full name as appears on passport:
John William Smith

Date of birth:
(dd/mm/yy)
05/28/60

Height: *5'10" (178cm)*
Weight: *168 pounds (76 kilos)*
Hair color: *Brown*

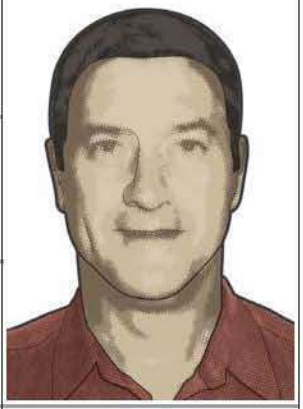

Passport number:
2008XXXXX63
Expires:
11/03/13

Driver's license number:
JHY280771smit
Expires:
12/28/22

Distinguishing marks (scars, tattoos):
Small scar—center of forehead
Large scar—right hand, middle finger
Chinese symbol tattoo—right arm

Languages spoken (fluent/basic):
English—native
French—basic
German—basic

Medication:
Medication—anti-malaria tablets

Allergies:
Amoxicillin

Swimmer: *Strong swimmer*
Outdoor skills/experience:
Attended basic military survival training.
Attended basic bushcraft course
Experienced in living outdoors

Next of kin 1: *Father*
William Smith
1018 Furlong Avenue
Brunswick, Maine,
USA 04555

Next of kin 2: *Brother*
Andrew Smith
1023 Parkglen
Ashford, Kent,
TM24 5HZ
UK

Tel: *(001) 55 555 2356*

Tel : *(0044) (0)155 555 2357*

Email: *willsmith@internet.com*

Email: *andrewsmith@internet.com*

Trip details:
Campsite 1 = Grid ST456654
Campsite 2 = Grid ST654987
Vehicles: Landrover 1 = white, reg MH55 555
Landrover 2 = blue, reg MH56 555
Group = Ben Jones, Kim Smith, and myself
Day 1: Park Landrover 2 at Campsite 2, and drive in Landrover 1 to Campsite 1
Day 2: Follow well-defined path along the Derwent Line Trail, aiming to camp overnight at Grid 4561559
Day 3: Continue along the Derwent Line Trail, aiming to be at Campsite 2 by mid-afternoon. Camp overnight at Campsite 2
Day 4: Travel in Landrover 2 to campsite 1 and retrieve Landrover 1

Foreseeable problems/intentions:
Day 1: None
Day 2: None but will use Ranger Station 18, grid 555555 (Tel. 666 6666) as an emergency rendezvous point
Day 3: None but will use Ranger Station 19, grid 666666 (Tel. 555 5555) as an emergency rendezvous point
Day 4: None

Communications plan:
Will speak to Dad on the morning of Day 1 and try to phone during the trek but am unsure of cell reception once on the trail so don't worry if you hear nothing.
Will phone Dad again when we reach campsite 2 on day 3.

My mobile: *07979 555555*
My email: *jws@internet.com*
Alt. No: *Ben 05555 555555*
Alt. No: *Kim 05555 555555*
Alt. No: *Campsite 1 555 555 55555*
Alt. No: *Campsite 2 555 555 55555*

Date: *November 23, 2009*

KNOW YOUR ENVIRONMENT

The continued survival of the human race can be attributed to our ability to adapt to our environment. While we may have lost some of our ancient ancestors' survival skills, we have, in their place, learned new skills as and when they have become necessary. The issue we face today is that the gap between the skills we once had and the skills we now have grows ever wider as we rely more heavily on modern technology. Therefore, when you head off into the wilderness, it is important to fully prepare for the environment.

Before a trip, research how the native inhabitants dress, work, and eat. How they have adapted to their way of life will help you to understand the environment and allow you to select the best gear, adopt the best techniques, and learn the correct skills. This is crucial given that most survival situations arise as a result of a sequence of events that could have been avoided—while you may have no control over the aircraft crashing, for example, you can recognize a change in the weather and choose whether to continue or turn back.

In this section YOU WILL DISCOVER...

- the difference between **tundra** and **taiga...**
- why you should **step-up** the size of your **survival footprint...**
- the **way to go** when it comes to **snowy peaks** or **swampy creeks...**
- how to avoid **getting lost** in the **permafrost...**
- the best way to **stay at the top** of the **food chain...**
- how to **feel at home** where the tropical **butterflies roam...**
- how a high-visibility **survival suit** could really **get you noticed...**

The four basic principles of survival are:

protection, location, water, and food. In most survival situations, this is also the order in which you should prioritize them.

P **PROTECTION** You must stay in a condition that allows you to be proactive in your continued survival and rescue. Physically, you should protect yourself against injury, the elements, and wildlife. Mentally, you should protect yourself against emotions that could rob you of the will to live: fear, guilt, despondency, and depression, for example. The best way to achieve this level of protection is to light and maintain a fire. Not only does it offer physical protection against the elements and wildlife, but it also provides a sense of security and familiarity that can help normalize even the most dire situation.

L **LOCATION** Your second priority is to recognize the importance of your location to your chances of survival and rescue. You will usually have two options: stay or go. Your preferred option should be to remain where you are and use anything at your disposal to mark your location to aid rescuers in their efforts to find you. If you can't stay where you are (perhaps because you are in imminent danger) you may have no option but to move to another location that provides either a better chance of survival or rescue, or both. Select a location aid that offers you the best chance of attracting attention in the environment in which you're traveling.

W **WATER** Put simply, water is the essence of life. While you may be able to survive for a few days without it, your ability to function and carry out even simple mental and physical tasks will be dramatically reduced in less than 24 hours. However, if you are injured, if the weather conditions are very hot, and if your workload is particularly heavy, for example, your survival time without water could be reduced to just a few hours. You should learn how to procure water in the particular environment in which you're traveling, and understand the ways in which a lack of water affects you.

F **FOOD** The importance of food is directly related to the length of time you are in a survival situation: the longer the situation lasts, the more important food will become in helping you stay fit and healthy. Even with a moderate workload, going without food for five to seven days will not kill you. You will, of course, feel hungry, you will grow tired, your movements will slow, and your body will lose its ability to repair itself. However, unless you're malnourished before you enter a survival situation, you are unlikely to starve to death within a week. Your body needs water to digest food, so always remember to prioritize water over food.

> "We rely so heavily on **modern technology**: water at the **turn of a tap**; heat at the **flick of a switch**; food raised, harvested and **prepared by others**"

TEMPERATE ENVIRONMENTS

THE TEMPERATE ZONES are the two regions between the Tropic of Cancer and Arctic circle, and the Tropic of Capricorn and Antarctic circle. Typified by seasonal variations in climate, conditions can include baking hot summers, freezing cold winters, and rainfall all year round. Landscape features range from forests and snow-covered peaks to grasslands and deserts. Although most temperate regions are highly populated, don't be lulled into a false sense of security—a worst-case scenario can still arise just a few miles from help.

WARNING!
Brown bears inhabit extensive wilderness areas in the northern temperate zone, and are the most dangerous temperate animal. The Grizzly, found in North America, is the most likely to attack humans.

TEMPERATE FEATURES

Although they contain a wide range of environmental features, the temperate zone's typical feature is forest, ranging from deciduous trees that shed their leaves in fall to coniferous trees that retain their leaves all year round. Grassland predominates where forests have been cleared, while highland areas contain hills and mountains. The abundance of rain means that rivers and lakes are common, and swampy wetlands form in areas with poor drainage.

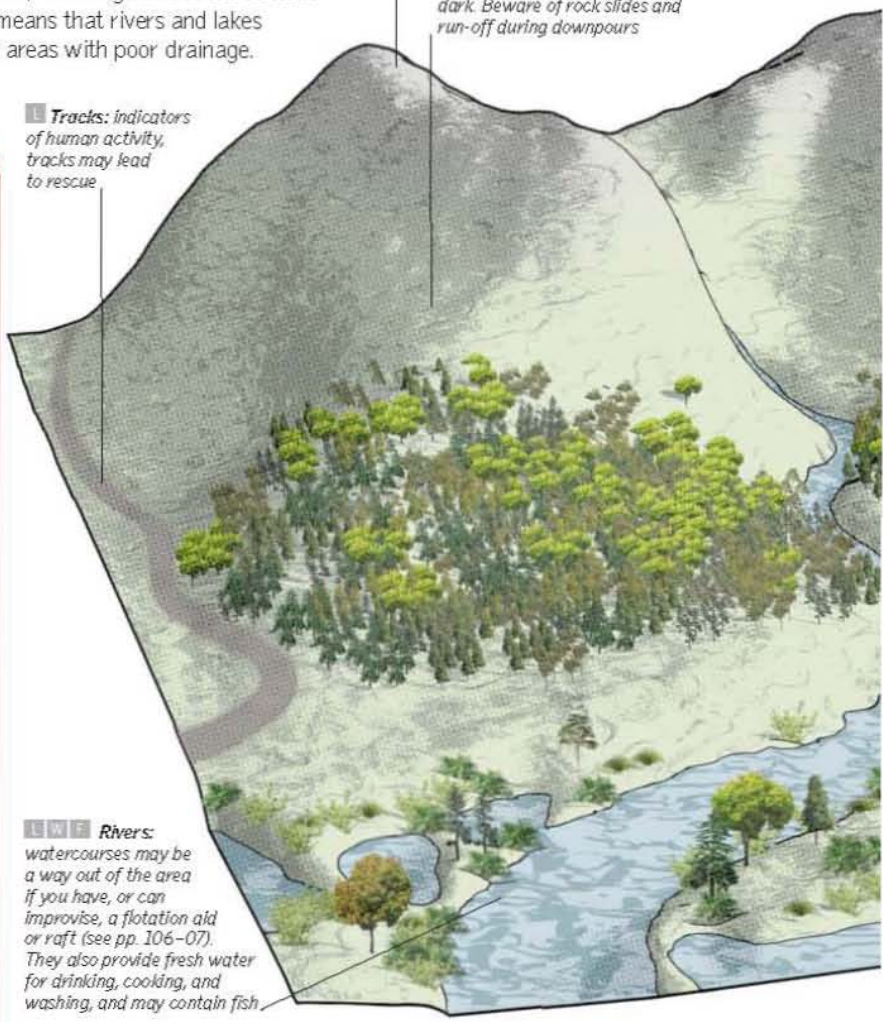

TEMPERATE ESSENTIALS

Climate and terrain can vary widely in temperate areas, so preparation for a range of eventualities is essential:

- Don't underestimate the temperate environment. Although the vast majority of the human race lives in temperate zones, the diversity of terrain and weather means that survival equipment and knowledge must be broad enough to cope with a wide range of situations and conditions.
- Temperate weather can change very quickly, so check the local forecast before you set off, and carry a small AM/FM radio so you can listen to local weather reports.
- Plan a realistic and achievable route, and prepare an EPA (see pp. 24–25). Be ready to re-assess your route during the trip.
- Take clothing for the full range of conditions you may encounter.
- Carry a survival kit (see pp. 60–61), knife, emergency equipment, cell phone, and first-aid kit (see pp. 260–61)—and know how to use them.
- Carry adequate water, and equipment to collect and purify more if necessary.
- Carry some form of basic shelter, even if only going out for the day.
- Always take a map and compass, and consider using a GPS as an aid.

WHERE TO FIND TEMPERATE AREAS

Deciduous forests are scattered throughout the temperate zones, but the largest occur in eastern North America, western Europe, and east Asia. Extensive coniferous forests are found in the higher latitudes of North America and Eurasia, while grassland is most common in continental interiors.

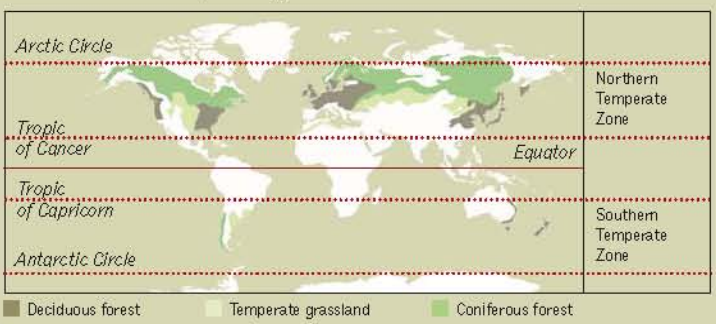

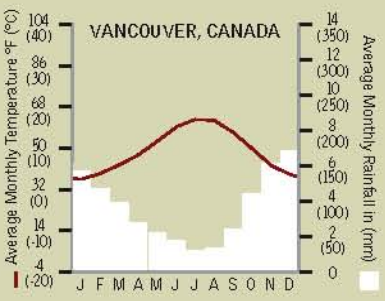

TYPICAL CLIMATE

Vancouver lies on the Pacific coast of Canada, and experiences a typical temperate climate of warm summers and cool, wet winters.

SURVIVING IN TEMPERATE AREAS

Most temperate environments have a relatively mild climate and good natural resources, making them favorable places for long-term survival. Water can be found in most areas, wood for building shelters and lighting fires is usually abundant, and different types of edible plants and animals can be found at different times of year if you know where to look. Potentially the greatest threat is hypothermia (see pp. 272–73), a particular problem in cold, wet and windy conditions, especially in winter and at night, when temperatures tend to fall even further.

WILDLIFE AND PLANT LIFE

Plant and animal life can be abundant in temperate areas and represents both a valuable food resource and a potential hazard. Shoots and stems, leaves, roots, nuts, or berries of some plants are edible, but only when positive identification—or at the least, the Universal Edibility Test (see p. 206)—is achieved. Small mammals, birds, insects, reptiles, and fish can also be found, but may be difficult to catch and kill, and must be prepared and cooked properly. Wildlife can also pose a threat—snakes, spiders, and scorpions may bite and, though rare, bears, wolves, and cougars may attack if cornered.

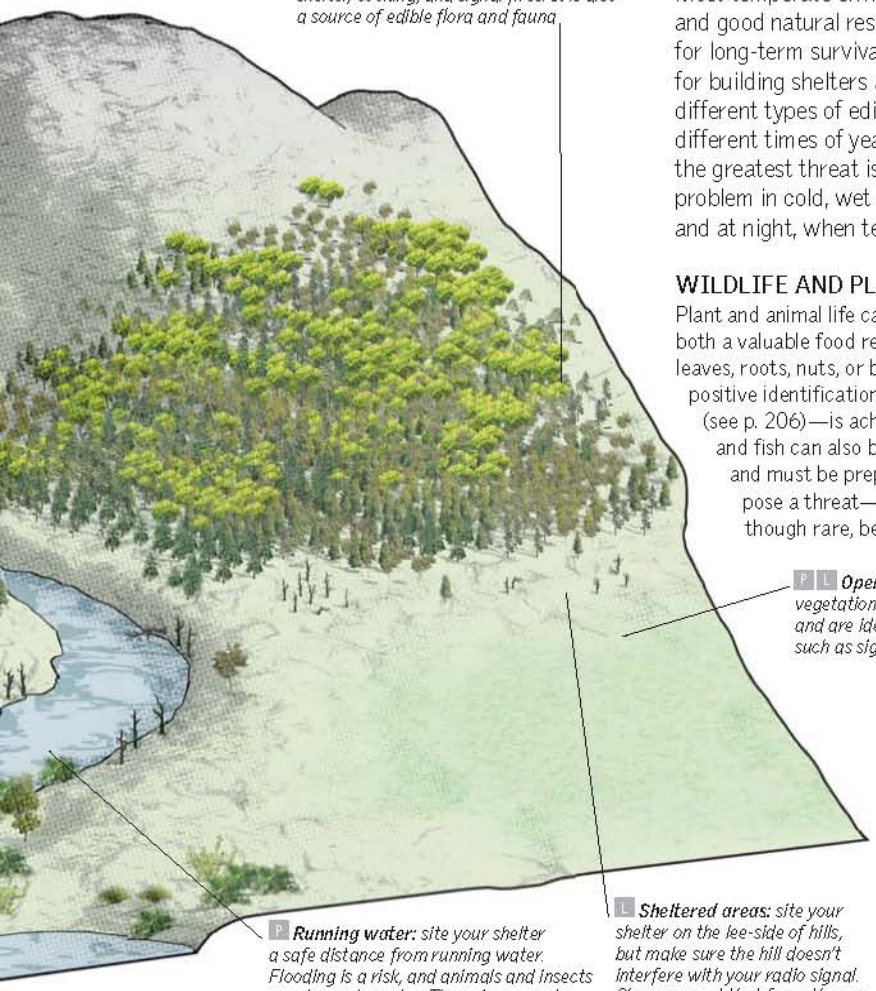

P L F ***Woodland:*** *tree cover affords protection from the elements, and wood for shelter, cooking, and signal fires. It is also a source of edible flora and fauna*

P L ***Open ground:*** *areas with no vegetation offer a clear view of dangers, and are ideal for deploying location aids, such as signal fires*

P ***Running water:*** *site your shelter a safe distance from running water. Flooding is a risk, and animals and insects are drawn to water. The noise may also hide the sound of animals or rescuers*

L ***Sheltered areas:*** *site your shelter on the lee-side of hills, but make sure the hill doesn't interfere with your radio signal. Choose a spot that faces the sun to maximize warmth and light*

FIND OUT MORE...

P **PROTECTION** **Shelters** pp. 156–65, 178–81 **Fire** pp. 118–33, 204–05 **Dangers** pp. 242–49, 300–05

L **LOCATION** **Navigation** pp. 66–77 **Movement** pp. 86–89 **Signaling** pp. 236–41

W **WATER** **Finding** pp. 188–91 **Purifying** pp. 200–01

F **FOOD** **Plants** pp. 206–07, 280–81 **Animals** pp. 208–13, 216–29, 290–99

TROPICAL ENVIRONMENTS

THE TROPICAL ZONE stretches from the Tropic of Cancer to the Tropic of Capricorn, and is centered over the Equator. A range of environmental features occurs in this area depending on the local climate, from lush, humid, biodiverse tropical rain forests to dry, more sparsely vegetated tropical scrub. With adequate preparation and caution you should be able to survive unsupported for extended periods—the jungle holds just as many resources as dangers.

WARNING!
Mosquitoes are responsible for more fatalities than any other creature in tropical areas. They carry a range of diseases including malaria and yellow fever, which kill millions of people every year.

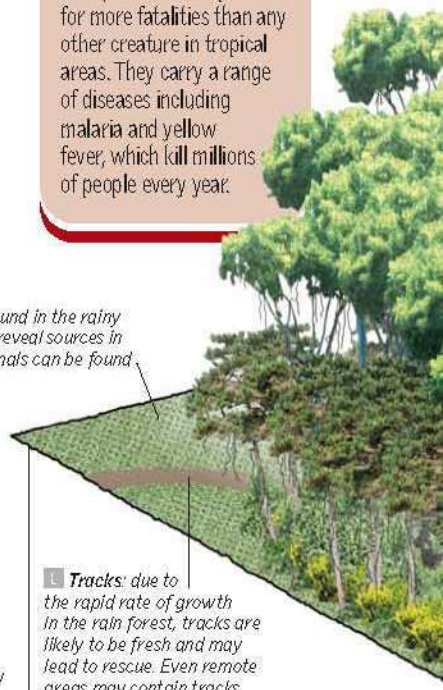

W T ***Tropical scrub**: water can be found in the rainy season, while animal movements may reveal sources in the dry season. Edible plants and animals can be found*

L ***Tracks**: due to the rapid rate of growth in the rain forest, tracks are likely to be fresh and may lead to rescue. Even remote areas may contain tracks left by loggers, prospectors, or local people*

P L ***Tropical scrub:** provides shade from the sun and materials for fire and shelter, but may be home to dangerous wild animals. Dense vegetation during the rainy season may impede visibility*

TROPICAL ESSENTIALS

Rain forests contain everything you need to stay alive, so remember the following tips when venturing out:

- While the jungle teems with predatory wildlife (such as big cats, crocodiles, and anacondas) that sees you as part of the food chain, it is the small creatures that can make life miserable. Most animals in the jungle want to avoid you as much as you want to avoid them—making a noise will scare most away.
- High humidity encourages infections, so keep everything covered (sleeves down, gloves on), and wash at every opportunity.
- Always sleep off the ground.
- Boil or treat all water. Some water found in plants can be safe to drink, but if in doubt, and if the water is yellow, milky, or very cloudy, treat before drinking.
- Many plants have defense mechanisms and can emit toxic liquids that may sting or burn—if you can't identify it, leave it alone.
- Dry tinder can be hard to find, so if you come across any, collect it and keep it safe and dry—lighting a fire in a moist, humid environment can be challenging.
- The smoke of a signal fire must penetrate through all the layers of the jungle, so light your fire where the canopy is sparse, such as the bend of a river or a clear-cut area.
- Navigation can be difficult, as you may only be able to see a few yards in front of you. Use dead reckoning—walking a short distance to a recognizable feature on your bearing—or pace-counting (see pp. 72–73).
- Rivers in the jungle usually run downhill to civilization and, eventually, the coast.
- Don't fight the jungle—tune in to its rhythm and work with it rather than against it.

TROPICAL FEATURES

Tropical rain forest, which occurs within 10 degrees of the Equator, is the predominant environmental feature in the tropics. But there are several other environment types just a few degrees of latitude away.

VERDANT RAIN FOREST

Tropical rain forests occur in areas with steady year-round temperatures and rainfall. Annual rainfall can reach around 6½–10 ft (2–3 m) and daytime temperatures may reach 90°F (30°C), dropping to 70°F (20°C) during the night. Monsoon (or "seasonal") rain forests occur in areas with a wet and dry season, while montane rain forests, also known as "cloud forests," occur in mountainous areas.

SCRUB AND SWAMPS

Also known as "thorn forest," tropical scrub consists of low, woody plants with thorns, which usually grow in clusters separated by patches of bare ground (grasses are uncommon). Leaves are shed in the dry season, forming a dense, herbaceous layer in the wet season. Swamps are another common tropical feature, and may consist of fresh or saltwater. Freshwater swamps are found in low-lying, inland areas, and consist of masses of undergrowth, reeds, grasses, occasional short palms, and islands. Saltwater swamps often contain mangroves, and occur in coastal areas that are prone to tidal flooding so are best navigated by boat. Visibility in both types of swamp is poor, and movement is difficult.

SURVIVING IN THE RAIN FOREST

While natural resources are abundant in the rain forest, the heat, humidity, number of animals, and voracious rate of vegetal growth can make it an uncomfortable place. Water, materials for shelter and fire, and edible plants may all be found, although identification of plants is crucial to avoid poisonous species. Animal life is everywhere, so sleeping platforms must be built off the ground to avoid biting insects, snakes, and spiders. The greatest danger is becoming lost—navigation is difficult due to dense undergrowth at ground level, and rescuers will struggle to find you beneath the thick jungle canopy.

P L W F ***Rain forest:*** *fuel for cooking and signaling fires and material for shelter may be found, although check for deadfalls and other dangers. Water from rain and rivers, and edible plants and animals can be found*

FIND OUT MORE...

P **PROTECTION** **Shelters** pp. 158–61, 166–71 **Fire** pp. 118–33, 204–05 **Dangers** pp. 167, 242–49

L **LOCATION** **Navigation** pp. 66–77 **Movement** pp. 106–07 **Signaling** pp. 236–41

W **WATER** **Finding** pp. 192–93 **Purifying** pp. 200–01

F **FOOD** **Plants** pp. 206–07, 284–85 **Animals** pp. 208–13, 224–29, 292–99

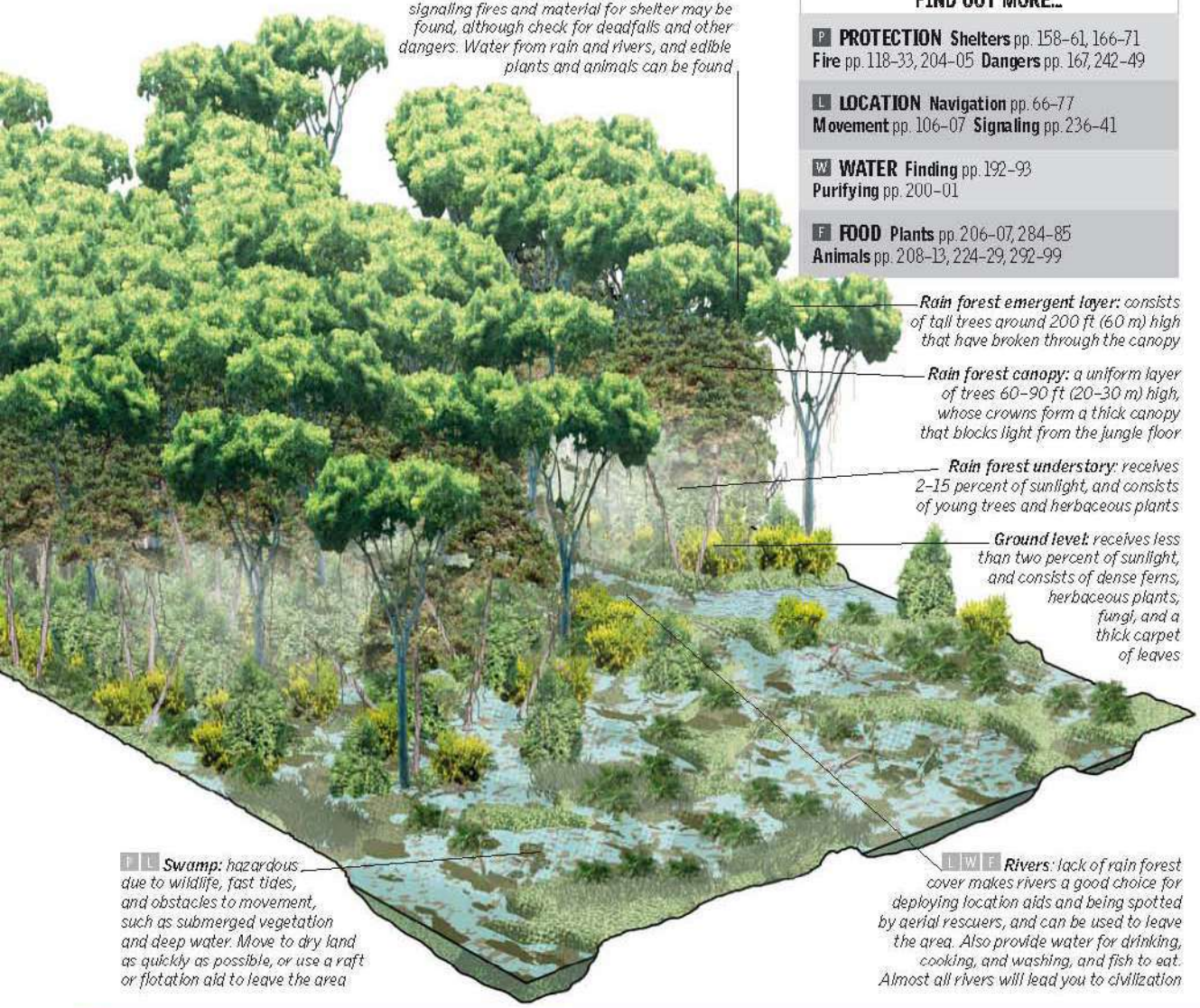

Rain forest emergent layer: *consists of tall trees around 200 ft (60 m) high that have broken through the canopy*

Rain forest canopy: *a uniform layer of trees 60–90 ft (20–30 m) high, whose crowns form a thick canopy that blocks light from the jungle floor*

Rain forest understory: *receives 2–15 percent of sunlight, and consists of young trees and herbaceous plants*

Ground level: *receives less than two percent of sunlight, and consists of dense ferns, herbaceous plants, fungi, and a thick carpet of leaves*

F L ***Swamp:*** *hazardous due to wildlife, fast tides, and obstacles to movement, such as submerged vegetation and deep water. Move to dry land as quickly as possible, or use a raft or flotation aid to leave the area*

L W F ***Rivers:*** *lack of rain forest cover makes rivers a good choice for deploying location aids and being spotted by aerial rescuers, and can be used to leave the area. Also provide water for drinking, cooking, and washing, and fish to eat. Almost all rivers will lead you to civilization*

WHERE TO FIND TROPICAL AREAS

Tropical rain forests are clustered around the Equator. The largest rain forest areas are found in Central and South America, sub-Saharan Africa, southeast Asia, northern Australia, and several Pacific islands.

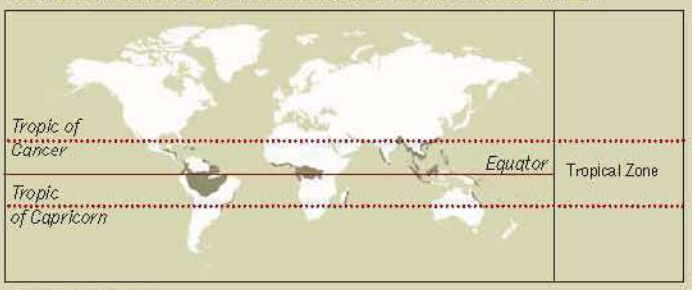

Rain forest areas

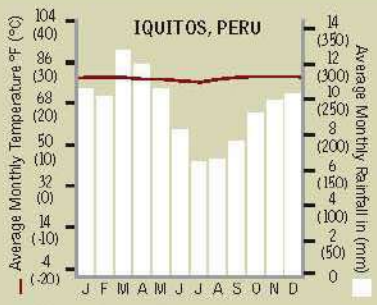

TYPICAL CLIMATE

Iquitos sits on the Amazon river in the heart of the Peruvian rain forest. It lies just south of the Equator and has a typical hot, wet tropical climate.

MOUNTAIN ENVIRONMENTS

THE MOUNTAIN ENVIRONMENT is one of the harshest places for humans to survive. Defined as landmasses with a summit above 2,000 ft (600 m), mountains can be very dangerous places due to their elevation and terrain. Lower temperatures and poor weather are more likely at higher altitudes, so there is a significant risk of hypothermia, frostbite, and altitude sickness, while snow, ice, and precipitous terrain present further hazards. Survival may depend on your ability to descend to areas with better prospects of survival and rescue.

MOUNTAIN FEATURES

Above the treeline, mountain terrain is typically barren, featuring bare rocks, gravel, boulders, and snow and ice, which are most extensive during the winter but may persist at high elevations during the summer. At lower levels, coniferous forest is the predominant feature, dissected by streams and rivers fed by snowmelt. Topographical features include scree slopes, cliffs, ravines, boulder fields, snowslopes, and glaciers.

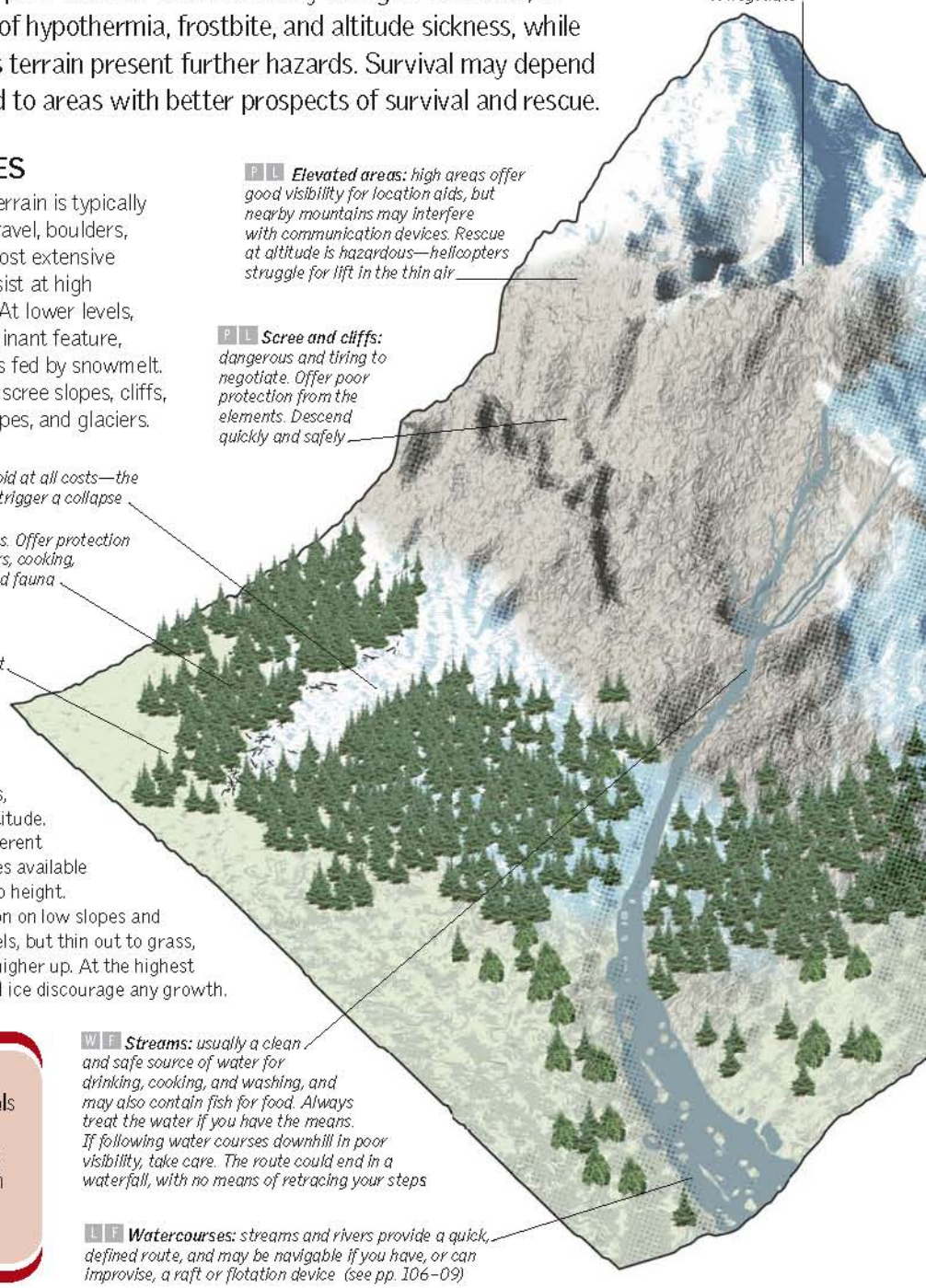

ELEVATION CHANGE

Mountains of sufficient height may have vastly different ecosystems at different elevations, due to harsh conditions at high altitude. Plant and animal life varies at different elevations, so the natural resources available for survival will differ according to height. Mixed deciduous trees are common on low slopes and conifers grow at intermediate levels, but thin out to grass, isolated shrubs, moss, and lichen higher up. At the highest levels, high winds, frost, snow, and ice discourage any growth.

WARNING!
One of the most dangerous animals in mountain areas is the cougar, also known as the puma, panther, and mountain lion. Found in North and South America, more than 20 percent of attacks are fatal.

SURVIVING IN THE MOUNTAINS

The prospects for survival are good at lower elevations, where trees provide material for shelter and fire, rivers provide water, and edible plants and animals are likely to be present. At higher elevations vegetation is scarce, so there are fewer options for shelter and food, although water from streams or snow is likely to be available. There may be a risk of avalanche on snow-covered slopes, and crevasses pose a hidden danger in glaciated areas. Cold-related injuries pose the greatest threat (see pp. 270–71).

HIGH-LEVEL HAZARDS

The primary threats to survival at high altitude are cold-related ailments, the most dangerous of which is hypothermia. This occurs when the body's core temperature falls below 95°F (35°C)—if allowed to drop as low as 86°F (30°C), it can be fatal. Frostbite may also occur in the extremities in freezing or cold and windy weather, and may lead to permanent tissue damage. Altitude sickness, which can cause pulmonary or cerebral edema—and in extreme cases, death—is also a risk above 8,000 ft (2,500 m).

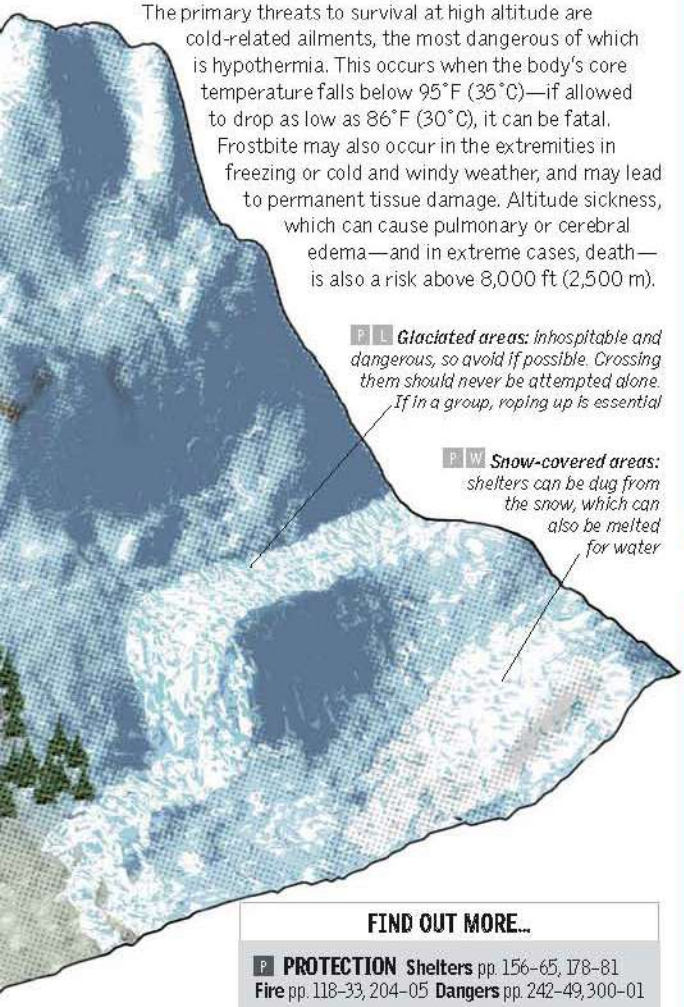

FIND OUT MORE...

P **PROTECTION** **Shelters** pp. 156–65, 178–81 **Fire** pp. 118–33, 204–05 **Dangers** pp. 242–49, 300–01

L **LOCATION** **Navigation** pp. 66–77 **Movement** pp. 90–91, 94–97 **Signaling** pp. 236–41

W **WATER** **Finding** pp. 188–91, 194–95 **Purifying** pp. 200–01

F **FOOD** **Plants** pp. 206–07, 280–81, 286–87 **Animals** pp. 208–13, 216–23, 290–97

MOUNTAIN DISTRIBUTION

Mountains are found on every continent, but the major ranges are the Himalayas and Karakoram in Asia, the Andes in South America, the Rocky Mountains in North America, and the Alps and Pyrenees in Europe.

Mountain ranges

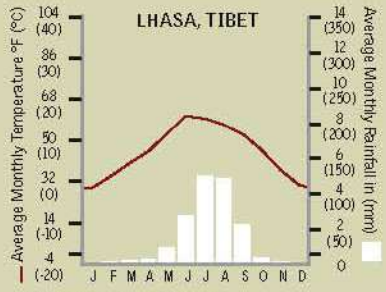

TYPICAL CLIMATE

Lhasa, the capital city of Tibet, lies at 11,972 ft (3,649 m) on the Tibetan Plateau, sometimes known as "the roof of the world." The high altitude means that winters are cool, and precipitation is relatively low because it mostly falls as snow, which has just 10 percent of the density of rain.

MOUNTAIN ESSENTIALS

High altitude and lower oxygen levels place higher-than-normal demands on even a fit body, so be prepared. Survival in the mountains is also tough on equipment.

- Respect nature and err on the side of caution—mountains are unforgiving, and rescue is unlikely to be quick or easy.
- Plan an achievable and safe route, and prepare an Emergency Plan of Action (see pp. 24–25).
- Dress in layers (see pp. 46–47). Start a walk lightly dressed (cold) and add or remove layers as necessary.
- Wear a hat (up to 70 percent of body heat is lost through the head) and gloves—cold fingers will hinder your functions.
- Tie loose items of clothing (hat, gloves, sunglasses) to your person—a lost glove could result in a lost hand, or life.
- Take a flashlight—weather changes and unforeseen problems may mean you are on the mountain in low light or darkness.
- Always carry basic equipment to give adequate protection if forced to camp overnight, such as a bivy sack and sleeping bag.
- Check avalanche warnings and carry an avalanche transceiver.
- Contact with snow makes clothing wet and less effective.
- If moving over snow, make a pair of improvised snow shoes (see pp. 94–95).

DESERT ENVIRONMENTS

THE DESERT ENVIRONMENT is a hostile place for human survival due to temperature extremes and the scarcity—or absence—of water. Most deserts are either hot and dry or cold and dry, and all typically receive less than 10 in (250 mm) of annual rainfall. They are areas of extremes, where heat exhaustion and hypothermia are threats, and flash floods may rapidly replace dry conditions. A harsh place to survive, the desert should only be entered by those who are prepared.

WARNING!
Desert areas are home to a range of venomous snakes, including the black mamba of eastern and southern Africa. It is particularly dangerous due to its aggressive nature and highly venomous bite.

DESERT FEATURES

The popular depiction of a desert is of a dry, sandy landscape with searing skies and temperatures among the highest in the world. Such areas do exist, but the desert environment has far greater diversity, from cold deserts in the polar regions to highlands, grassland, and broken, rocky terrain that contain "wadis" (seasonal stream beds).

HOT, DRY DESERTS

Temperate and tropical deserts are usually hot and dry due to high daytime temperatures, which leads to low precipitation and high evaporation. There may be months or even years between rainfall, which usually occurs in dramatic bursts but then quickly drains into seasonal stream beds (wadis), or evaporates either in the air or soon after touching the hot ground. Very little rain soaks into the ground, which results in sparse vegetation that has evolved to maximize water, usually via extensive root systems that collect moisture, waxy skin and leaves that prevent water loss, and stems or roots that hold water. At night, in winter, and at higher altitudes, freezing temperatures and frost may occur.

GRASSLANDS

Known as savanna in tropical areas, grasslands often border desert regions, and experience a similarly dry climate for most of the year. The key difference is that they also experience a wet season, which accounts for the more diverse and extensive vegetation. Temperate grassland features grasses, shrubs, and small pockets of trees, while savanna is typically more varied, with denser tree cover, bushes, and grasses that grow tall during the wet season but die back in periods of drought. Wildlife is more prevalent, including predators. Wildfires may spread during the dry season, and play an important role by destroying dead plant matter and adding fertile ash to the soil.

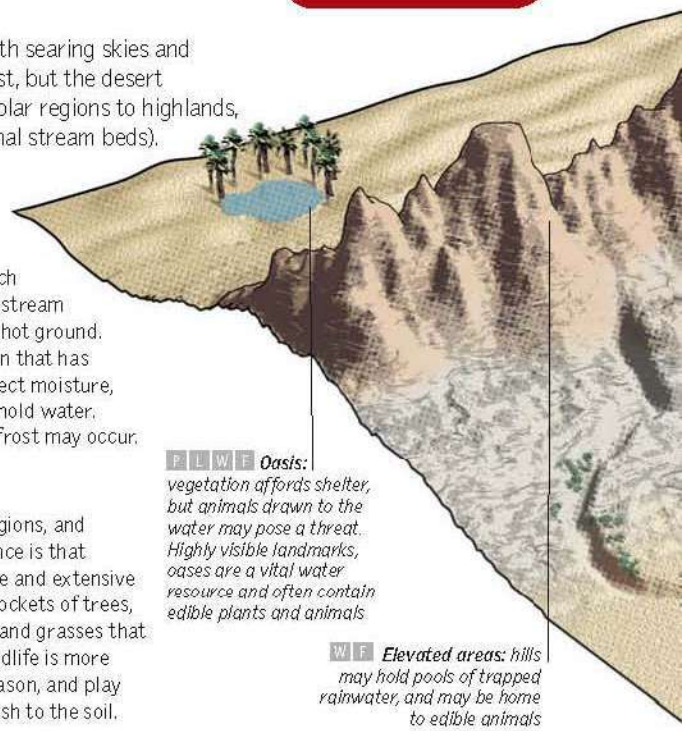

WHERE TO FIND DESERT AREAS

Although deserts occur on every continent, the largest hot desert in the world is the Sahara in North Africa, which forms part of a desert band stretching through the Middle East and into south and central Asia.

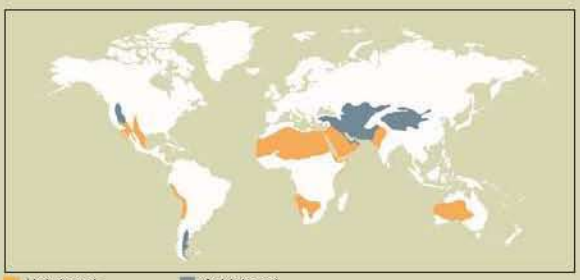

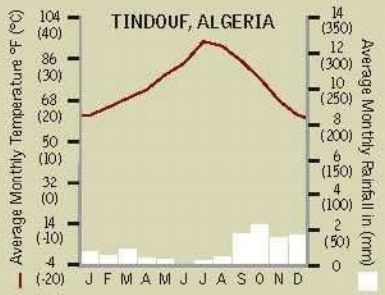

TYPICAL CLIMATE

Tindouf lies in the heart of the Sahara, and experiences a typical hot-desert climate. Precipitation is virtually zero all year round, while temperatures peak in the summer.

SURVIVING IN THE DESERT

Hostile temperatures and few natural resources limit chances of survival in the desert. Water and shelter are scarce, if not non-existent, edible vegetation is limited, and animals hide from the sun in the day. Wadis, grasslands, and higher ground all hold better prospects. The greatest dangers are dehydration and heat exhaustion (see pp. 272–73), although African savanna areas may be home to large, dangerous mammals. The main killers are big cats, hippos, rhinos, elephants, and crocodiles.

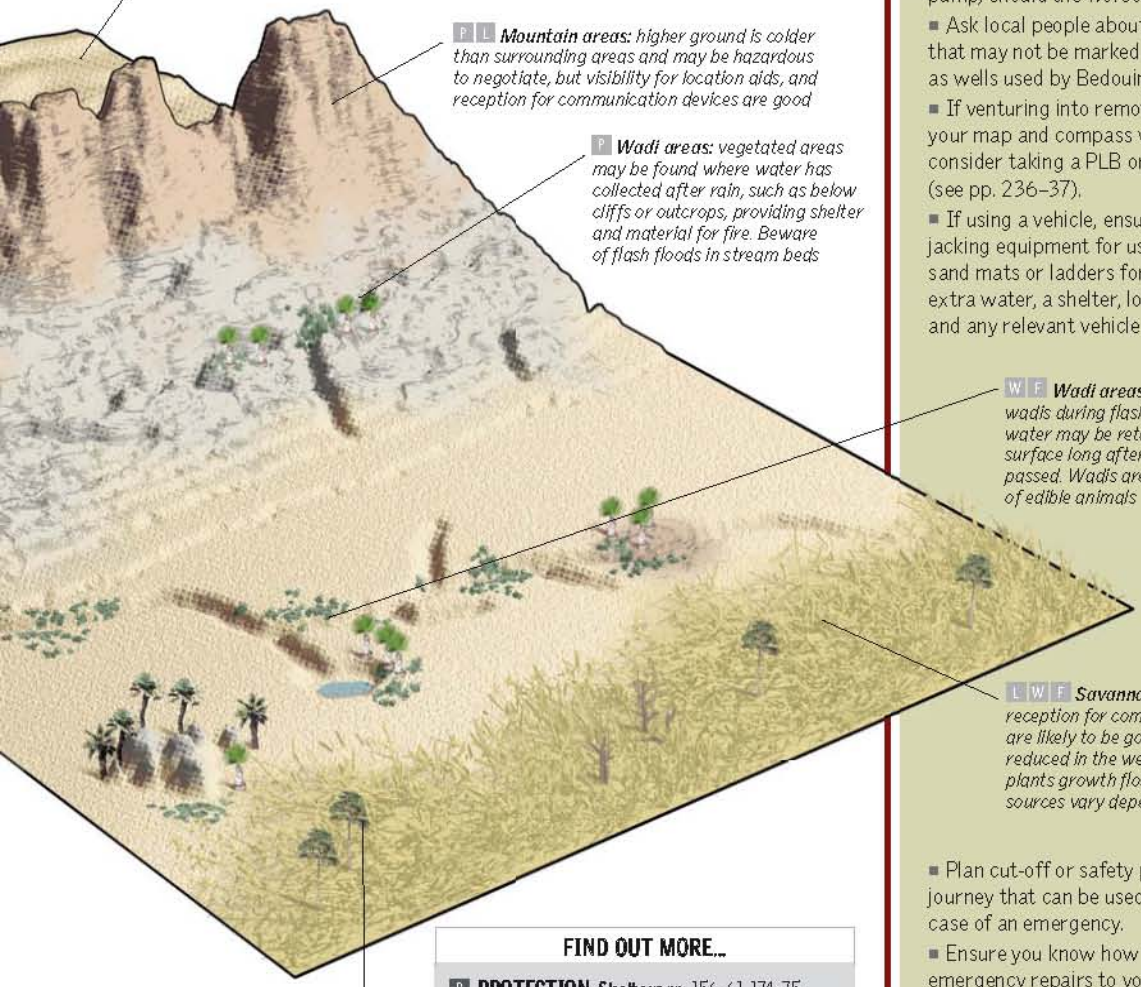

FIND OUT MORE...

P PROTECTION **Shelters** pp. 156–61, 174–75 **Fire** pp. 118–33, 204–05 **Dangers** pp. 242–49, 300–05

L LOCATION **Navigation** pp. 66–77 **Movement** pp. 100–01 **Signaling** pp. 236–41

W WATER **Finding** pp. 192–93 **Purifying** pp. 200–01

F FOOD **Plants** pp. 206–07, 282–83 **Animals** pp. 216–21, 224–29, 292–95

DESERT ESSENTIALS

Survival in the extreme conditions of the desert is impossible without full preparation. Consider the following:

- Always prepare an EPA to notify someone of your plans before entering a desert area (see pp. 24–25).
- Water is life. Do not underestimate your needs, carry extra just in case, and carry equipment to maximize your chances of finding and procuring more (binoculars, surgical tubing, filtration/purification pump) should the worst happen.
- Ask local people about water sources that may not be marked on maps, such as wells used by Bedouin tribes.
- If venturing into remote areas, augment your map and compass with a GPS, and consider taking a PLB or satellite phone (see pp. 236–37).
- If using a vehicle, ensure you carry jacking equipment for use in soft ground, sand mats or ladders for self-recovery, extra water, a shelter, location aids, and any relevant vehicle spares.
- Plan cut-off or safety points into your journey that can be used to divert to in case of an emergency.
- Ensure you know how to accomplish emergency repairs to your vehicle.
- Ask local people about known dangers, such as impassable roads, misleading tracks, and soft-sand areas.
- When using a ladder for self-recovery, tie it to the back of the vehicle so that it's dragged behind once crossed, and can be retrieved when you reach solid ground.
- If your vehicle breaks down, leave it only if staying is no longer safe or feasible—rescuers will search for it first.

COLD ENVIRONMENTS

LOW TEMPERATURES are a potentially lethal hazard, but as long as you can conserve body heat, survival in the cold is quite possible. The ice-covered Arctic and Antarctic polar regions, sub-polar tundra and taiga, as well as parts of most temperate zones during the winter, can all be classed as cold environments. Wet, windy weather can drive moderately cold temperatures down further, increasing the risk of the most lethal cold-weather condition—hypothermia.

WARNING!
The largest predatory land mammal on Earth, the polar bear inhabits the Arctic. Attacks are most likely when the bears are hungry, and are usually fatal.

COLD ENVIRONMENT FEATURES

Outside temperate areas, the typical cold environments are the polar areas, which feature glacial ice sheets and sea ice; tundra, which contains permafrost (permanently frozen soil), small shrubs, mosses, and lichens; and taiga, with extensive conifer forests. Almost half of the northern hemisphere's landmass can be classified as a cold region, due to the influence of cold air masses (see pp. 78–83), while ocean currents and altitude can also have a cooling effect.

POLAR WASTES

The northern polar area is situated over the Arctic Ocean, so largely consists of sea ice, the extent and depth of which varies throughout the year. Because of the lack of land, melting ice is a serious danger. The southern polar area is located over the continent of Antarctica, and contains the largest glacial ice sheet on Earth. Both areas contain ice shelves—glaciers that have extended over the sea—parts of which break off to form icebergs.

TUNDRA AND TAIGA

Latitudes that neighbor the polar areas are home to tundra, an environment of permafrost and vegetation that is stunted due to low temperatures. Taiga is located further away from the poles, where temperatures are high enough to support coniferous forests.

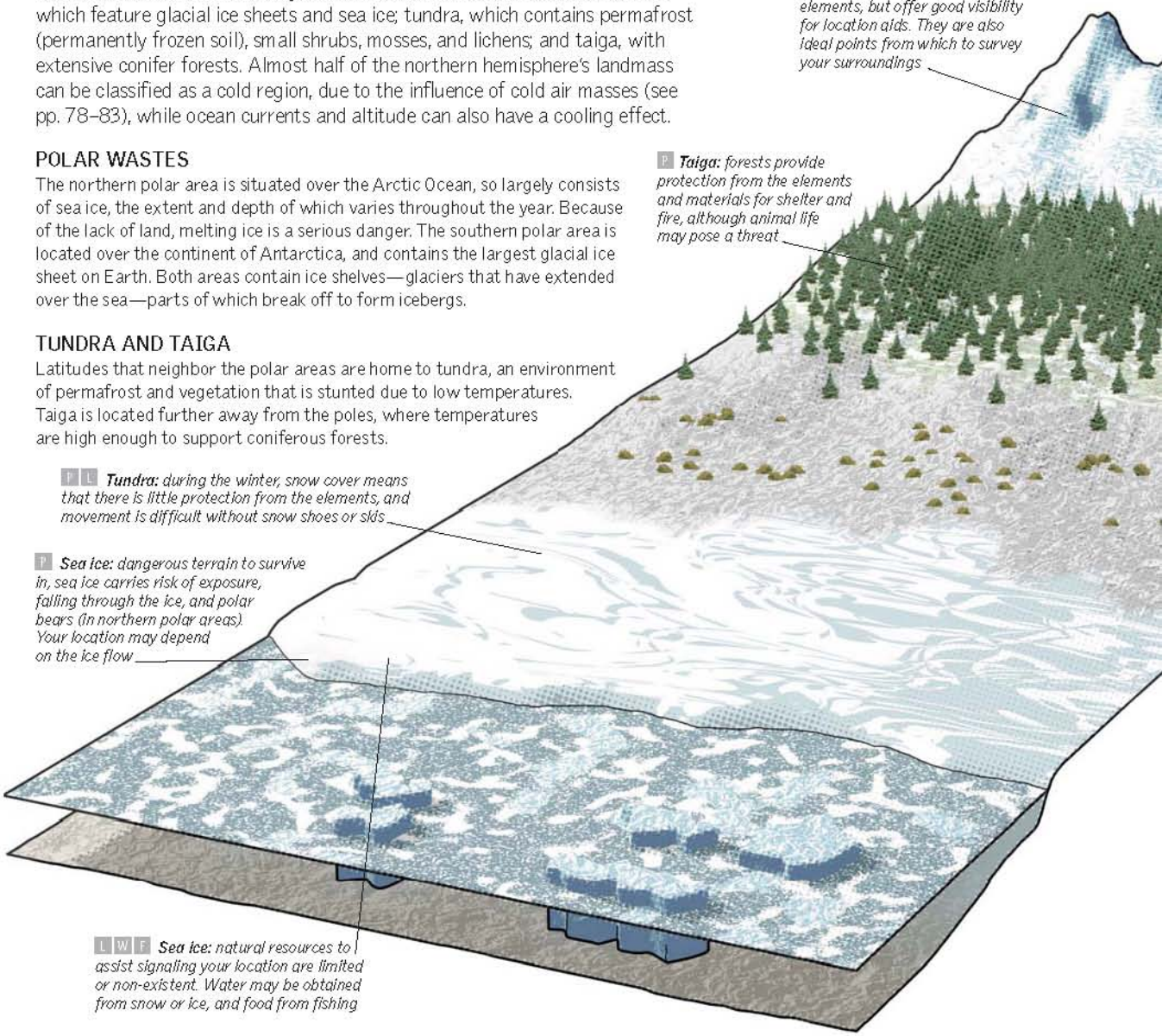

SURVIVING IN THE COLD

Natural resources may be scarce in cold environments, so your survival is likely to depend on the equipment and supplies you have with you. Shelters can be dug from the snow to help you stay warm and avoid exposure to freezing temperatures, but material for making fire is limited in polar and tundra areas. Survival is more feasible in the taiga, where wood is available, and fresh water and edible flora and fauna can also be found. The greatest dangers are hypothermia (see pp. 272–73) and—in northern polar areas—polar bears.

LOW-TEMPERATURE DANGER

Heat loss occurs in several ways, and minimizing it is essential to survival in cold conditions. It's important to avoid over-exertion, since heat is lost through sweating, and to cover your head and wear layered clothing to trap radiated body heat. Heat is also lost through conduction—direct contact between the skin and a cooler surface—particularly when wet, since water conducts heat away from the body 25 times faster than air. For this reason, it's vital to wear insulated and waterproof clothing, and to stay dry. Respiration also involves heat loss—each inhaled breath contains cold air, which is warmed as it passes through the lungs, and laden with moisture when exhaled. Reduce heat loss by covering your nose and mouth with a ski-mask or scarf, or by breathing warmer air next to a fire or in a warm shelter.

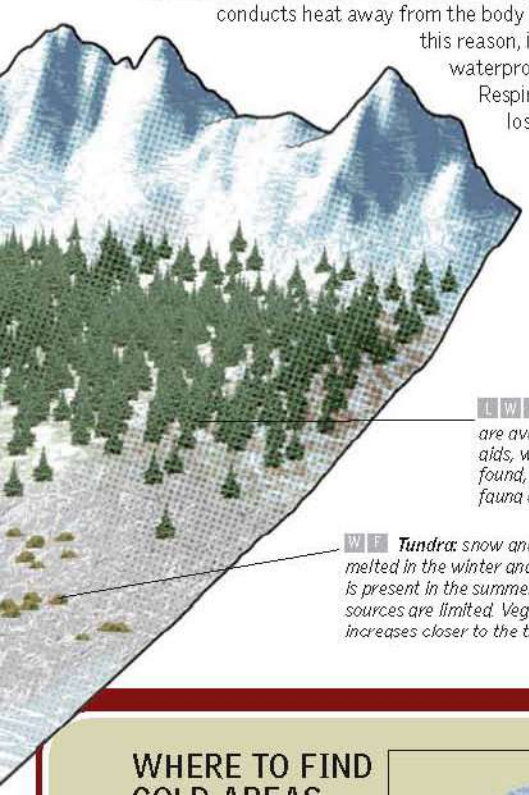

L W F *Taiga: natural resources are available for making location aids, water sources may be found, and edible flora and fauna are likely to be present*

W F *Tundra: snow and ice can be melted in the winter and ground water is present in the summer, although sources are limited. Vegetation increases closer to the treeline*

COLD ESSENTIALS

The main threats in cold environments are hypothermia and exposure, so ensure you are fully prepared:

- Dress in loose-fitting layers of clothing (see pp. 46–47), avoid overheating, and ensure your clothing stays dry and clean.
- If your hands are cold, don't warm them with your breath, which will make them wet—tuck them under your armpits.
- Regularly check your extremities (face, toes, hands, and ears) for frostnip, the first stage of frostbite (see p. 273).
- Wind-chill is dangerous, so take shelter from the wind at every opportunity, particularly if you are in a survival situation.
- Always ensure that your shelter is well ventilated—keep vent holes clear and check regularly, especially when snow falls.
- Get off the cold ground, snow, or ice—sit on your pack or make a sleeping platform using boughs to avoid losing body heat.
- If fire is your primary means of warmth, triple the amount of firewood you think you need—you will need enough to last the night.

FIND OUT MORE...

P **PROTECTION** **Shelters** pp. 156–65, 178–79 **Fire** pp. 118–33, 204–05 **Dangers** pp. 242–49

L **LOCATION** **Navigation** pp. 66–67 **Movement** pp. 94–97 **Signaling** pp. 236–41

W **WATER** **Finding** pp. 194–95 **Purifying** pp. 200–01

F **FOOD** **Plants** pp. 206–07, 286–87 **Animals** pp. 208–13, 216–23, 290–91, 296–97

WHERE TO FIND COLD AREAS

The Arctic and Antarctic polar areas are the coldest places on Earth, and are found at the furthest points from the Equator. Tundra exists next to polar regions, and is bordered by taiga, which is found in northern Eurasia and North America.

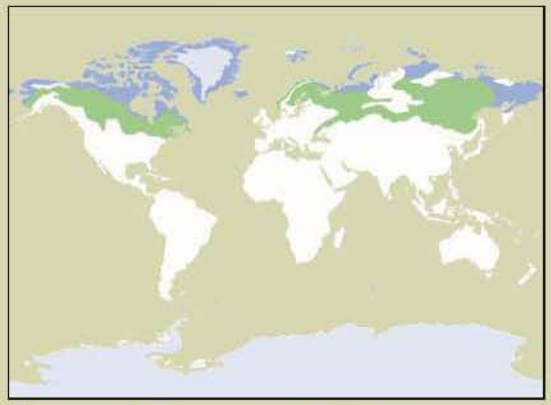

Polar areas · Tundra areas · Taiga areas

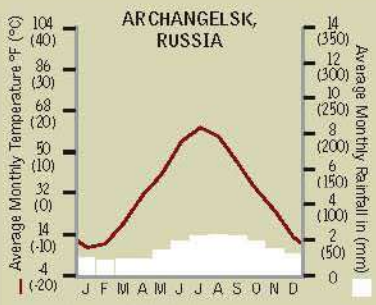

TYPICAL CLIMATE

Archangelsk is a Russian port on the Barents Sea. Located in the taiga zone, it experiences below-zero temperatures for half the year and low precipitation.

MARINE ENVIRONMENTS

POTENTIALLY THE MOST HARSH and challenging of all environments, the sea is unlike any other due to the lack of the most essential requirement for human life—fresh water. Around 70 percent of the Earth's surface is covered by saltwater, ranging from the freezing, wind-torn polar oceans to the warm tropical seas. Winds and currents govern your movement, protection is limited, and natural resources are virtually nil, so reaching land is your best hope for survival.

WARNING!
There are sharks in every ocean, but those in tropical waters are the most aggressive. Around 20 of the hundreds of species have attacked humans, and the most dangerous types are the great white, tiger, and bull sharks.

MARINE FEATURES

The marine environment ranges from the coastal margins, which are shallow enough— 650 ft (200 m) or less—to support a huge array of animal and plant life, to the open oceans, the deepest of which plunge to more than 33,000 ft (10,000 m).

SHELTERED COAST TO OPEN SEA

The coastal margins are home to the majority of ocean life, and include terrestrial areas—the shoreline and intertidal zone—and aquatic—the sea up to the edge of the continental shelf. A variety of environmental features are found, such as sand, rock, and pebble beaches, dunes, cliffs, estuaries, mud flats, mangroves, lagoons, kelp forests, and coral reefs. The open ocean is a more barren environment that supports less life, and is notable for its vast extent, comprising 92 percent of all saltwater. Survival depends on your equipment, supplies, and ingenuity.

OCEAN CONDITIONS

Your protection and location at sea depends largely on the weather—you may need to deal with anything from exposure to the hot sun, to cold, wet, windy conditions and high seas. Conditions are toughest in the South Atlantic Ocean, where winds (known as the "Roaring Forties") are strong all year round, and sea ice is extensive in the winter. Severe seasonal storms occur: hurricanes in the tropical areas of the western Atlantic, typhoons in the western Pacific. India and Southeast Asia also experience severe weather during the monsoon seasons.

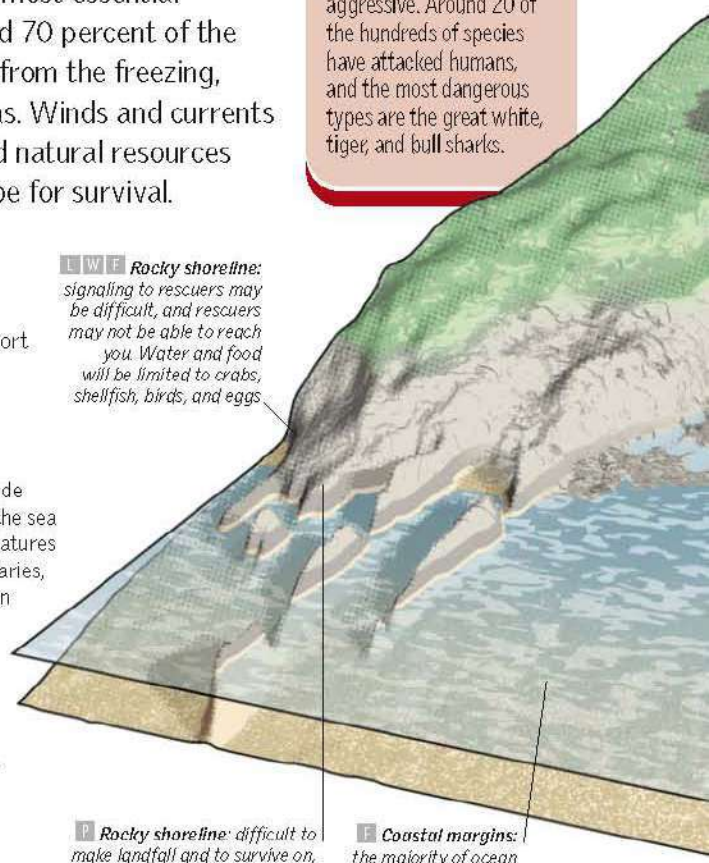

WINDS AND CURRENTS

Winds and currents follow prevailing patterns, so if you know roughly where you are, you can figure out where they'll carry you. With a drift of just 1 knot, a liferaft can move 24 miles (38 km) from its last known position in 24 hours. It may also drift up to 35° either side of the wind, further widening the search area to more than 400 sq miles (1,000 sq km) in the first 24-hour period, and increasing exponentially over time. Conversely, favorable currents can also be used to help you drift to landfall faster.

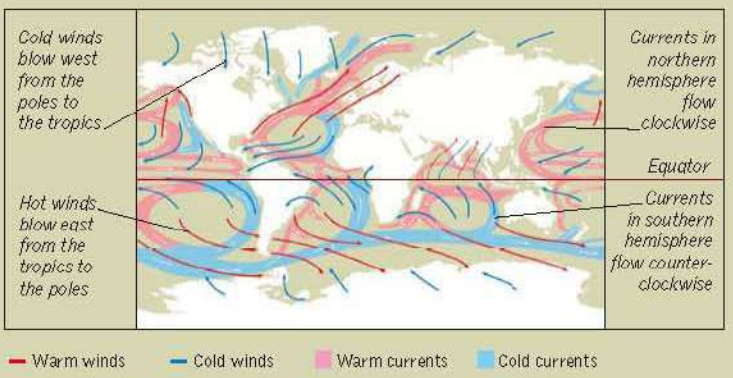

P L **Beach:** *above the high-tide mark, beaches may provide protection from wind and rain, and materials for building shelters, lighting fires, and improvising location aids*

W T **Beach stream:** *streams may drain on to beaches, providing water for drinking, cooking, and washing. Edible plants and animals may be present*

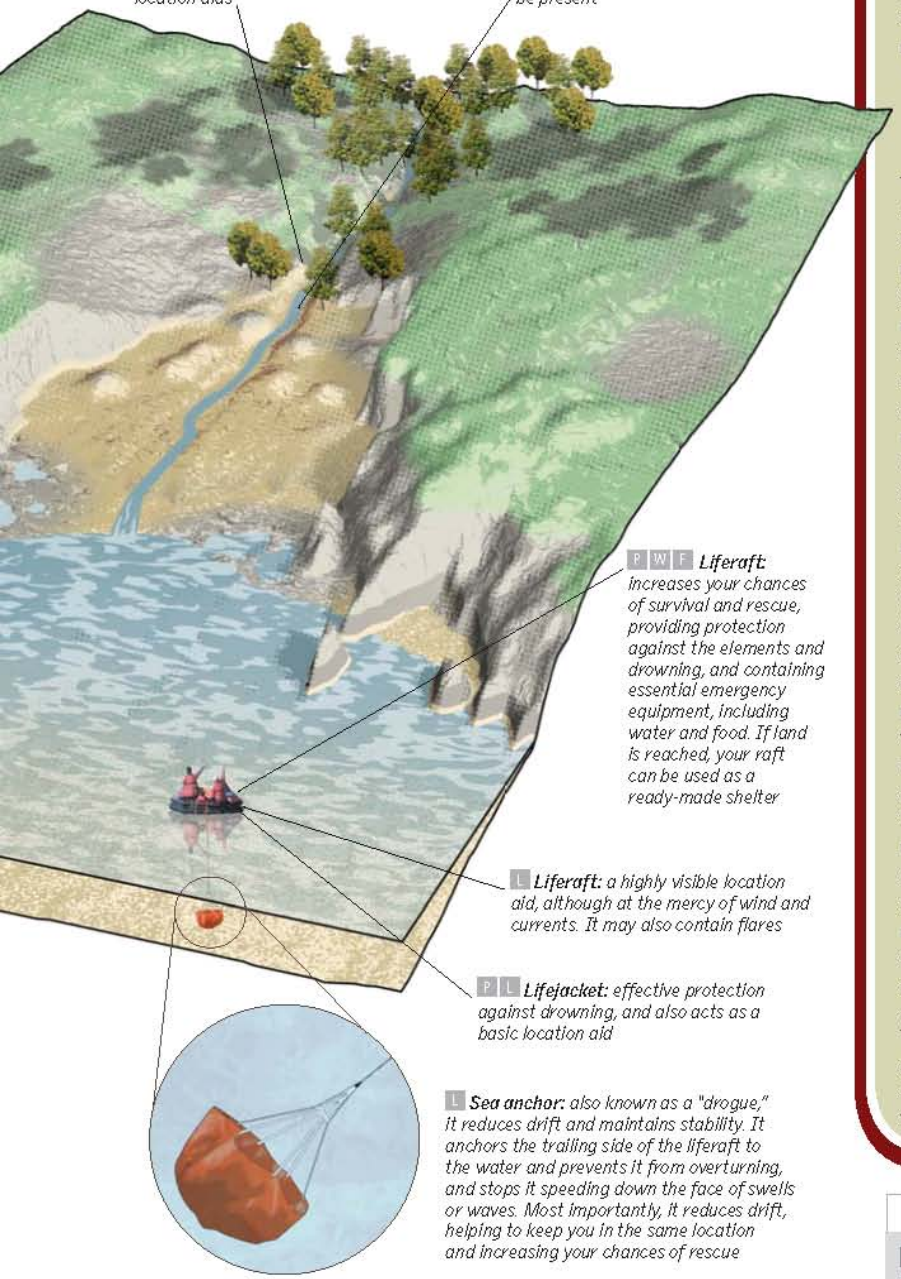

P W F **Liferaft:** *increases your chances of survival and rescue, providing protection against the elements and drowning, and containing essential emergency equipment, including water and food. If land is reached, your raft can be used as a ready-made shelter*

L **Liferaft:** *a highly visible location aid, although at the mercy of wind and currents. It may also contain flares*

P L **Lifejacket:** *effective protection against drowning, and also acts as a basic location aid*

L **Sea anchor:** *also known as a "drogue," it reduces drift and maintains stability. It anchors the trailing side of the liferaft to the water and prevents it from overturning, and stops it speeding down the face of swells or waves. Most importantly, it reduces drift, helping to keep you in the same location and increasing your chances of rescue*

SURVIVING AT SEA

Natural resources in the open ocean are virtually nil, so improvising shelter or location aids is limited to what you have with you. Seasickness, especially in a small liferaft, can seriously affect your ability to remain both hydrated and motivated. Unless you can reach land, your only drinking water will come from rain or mechanical desalination devices—consuming urine or saltwater will only increase your rate of dehydration. Reaching land or shipping lanes greatly increases your chances of rescue.

MARINE ESSENTIALS

Your chances of survival at sea will be greatly enhanced by good preparation, so consider the following:

- Pack sea-survival equipment as if your life depends on it—it might! Other environments allow you to improvise survival and location aids, but not the sea.
- Take emergency immersion-survival suits. They will protect you against the elements and aid flotation, and their color (high-visibility orange or red) is an aid to location.
- Always carry a Personal Locator Beacon (PLB), just in case.
- Take several means of obtaining water, such as a reverse-osmosis pump, which turns saltwater into drinking water, a saltwater desalination kit, a solar still, and rain catchers.
- Always keep a pre-packed emergency bag handy. It should contain; first aid kit, reverse-osmosis pump, emergency beacon, solar still, water, flares, fishing kit. (see pp. 250).
- If your vessel sinks, do not abandon it until absolutely necessary. Stay nearby unless it is unsafe to do so—anything that gives you a bigger "survival footprint" will increase your chances of being rescued, and you may be able to retrieve provisions from the vessel.
- If imminent rescue is unlikely, you will need to ration your water. Consider how much you have, how much can be procured, the chances of seasickness (which leads to dehydration), and the likelihood of being rescued later on.
- Protect yourself from the elements (sun, wind, cold, heat, and salt) as well as you can—prevention is better than cure.
- The sea, especially in rough conditions, can sap the will to live quicker than any other environment. Seasickness is literally draining, resulting in lost fluid, energy, and motivation. Take anti-seasickness tablets at the earliest opportunity—even before abandoning ship—so that you can keep them down, and keep taking them as recommended.

FIND OUT MORE...

P **PROTECTION** **Shelters** pp. 156-61, 176-77 **Fire** pp. 118-33, 204-05 **Dangers** pp. 250-55, 300-01

L **LOCATION** **Movement** pp. 106-07 **Signaling** pp. 236-41

W **WATER** **Finding** pp. 196-97 **Treating** pp. 200-01

F **FOOD** **Plants** pp. 206-07, 288-89 **Animals** pp. 208-13, 224-29, 292-99

GET THE RIGHT GEAR

Whether you're equipping yourself for a lengthy expedition or just a day trip, it's important to choose the correct equipment and clothing. The wrong time to realize that your gear isn't up to the job is when your life depends on it doing what it was designed to do.

When selecting gear, it's always a good idea to work backward from a worst-case-scenario survival situation. Think about what you would need to survive if the worst happened. This is your "first-line" equipment, which you should carry with you at all times. It should enable you to address the basic principles of survival relevant to the environment you are in, and consists of the clothing you would wear and your basic survival equipment—your survival tin and belt-order (see pp. 43, 61).

Once you have organized this basic equipment, you can then decide what pieces of essential gear would make your expedition enjoyable rather than just survivable. This will probably consist of a bivy or tent, a sleeping bag, cooking equipment and food, and something to carry it all in.

In this section YOU WILL DISCOVER...

- the **meaning** of the term "**bomb-burst gear**"...
- that it's always **good to share** when you're **in a group**...
- which **backpacks** are the **best backpacks** to pack...
- why you should **hike around the house** to **break in your boots**...
- a good reason to take a **three-season tent**...
- how to **sleep** like a **mummy**...
- what **gorp** is and **when** you should **eat it**...

It's at this stage that you should conduct a short trial to check that your chosen gear will fit in your pack and that you can comfortably carry it.

Finally, any spare capacity can be utilized for non-essential "luxury" items, such as an inflatable mattress and pillow, an MP3 player, or a book. Double-check that your gear works and that you know how to use it properly. The more you understand how and why a piece of equipment works, the better able you will be to improvise should it be damaged or lost.

With each venture into the wilderness, your confidence in your abilities and that of your equipment will increase. As it does, you may want to explore the world of ultralight camping.

ULTRALIGHT CAMPING refers to the careful selection of your three largest items of gear to dramatically reduce the load you have to carry. A standard tent, pack, and sleeping bag have a combined weight of approximately 20 lb (9kg). Ultralight camping has the potential to halve the weight of your "big three."

- Take a hooped bivy instead of a tent and choose a low season-rated sleeping bag (wear your day clothes inside the bag for extra warmth).
- You can then reduce the size of your pack because your tent and sleeping bag are smaller.

Trekking with less equipment but still staying safe allows you greater freedom to enjoy the wilderness—carrying less weight puts less strain on your body, enabling you to cover greater distances. Ultralight camping can become addictive and is safe if you use common sense—but never compromise on the quality of equipment. Try:

- Using a hexamine stove to cook food and boil water—you won't have to carry a stove and fuel.
- Taking one pot to cook food and boil water.
- Using natural materials as a mattress.

> "The **wrong time** to realize that **your equipment** isn't up to the **task in hand** is when **your life depends** on it doing what it was **designed to do**"

CHOOSING YOUR GEAR

FORWARD PLANNING IS ESSENTIAL when you're deciding on what equipment to take on your trip. You'll need to assess your own personal requirements, the likely demands of weather and terrain, and the amount of gear you'll actually be able to transport. With a little forethought, you'll be as well equipped as you can be for any situation.

PACKING FOR YOUR TRIP

You will first need to weigh up your particular kit requirements against the limitations of your chosen mode of transport (see below). It's then vital to organize and prioritize your gear (see opposite) to ensure that any items you may need in a survival scenario are always close at hand, should the worst happen.

KEY PACKING TIPS

- Pack your gear in reverse order: the things you want to get to first should be the last things you pack.
- Stash items you may need during the day in side or lid pockets, for easy access.
- Coil and tape long straps so that they aren't a snagging hazard.
- Use a dry bag inside your backpack.
- Ensure that all pockets are fully closed and zipped.
- Air out anything packed wet as soon as you stop for the day.

HOW MUCH TO PACK

The environment of the region you're traveling to will dictate the type of equipment you will need to take with you, but your proposed mode of transport is the main constraint on the quantity of kit you can take (see chart, right). The more gear you have to carry yourself, the fewer luxuries you can take: the added weight will make traveling uncomfortable, use up too much valuable energy, slow your progress, and also limit the distance you can travel each day. It's also worth considering whether you can reduce the weight and volume of the kit items you're taking (see box, below).

LIGHTENING THE LOAD

When you're going to be traveling on foot, it's important to get rid of all the excess weight and volume you can, even if it's only by a small amount. The larger and heavier your backpack, the harder you will find the journey. If you have already ruthlessly pared your gear down to the bare essentials, try to reduce the weight or volume of the kit itself. First concentrate on the big three items: tent, sleeping bag, and backpack. Consider using a hooped bivy or tarpaulin rather than a tent; look for a lighter sleeping bag—some can weigh as little as 2lb (0.9 kg); and consider buying a backpack with a lighter frame.

WEIGHT AND SPACE RESTRICTIONS

ON FOOT	Backpacks (see pp. 44–45) are available in a wide range of volumes—from 30 liters (a day pack) to 80 liters (suitable for a trip of a week or more). Bear in mind that you'll have to carry whatever gear you take with you, and that the weight of the pack will determine what you're realistically able to achieve each day in terms of distance and activity. Pack only the bare minimum: there's no room for luxuries when traveling on foot.
MOUNTAIN/ TOURING BIKE	Pannier bags for bicycles are designed to fit on front, rear, or side pannier racks, and are available in a range of shapes, sizes, and capacities (from 5 to 50 liters). Larger bags will provide you with enough space to pack essentials, such as tent, sleeping bag, cooking equipment, and spare clothes, but you'll be restricted by the weight, and also by the need to carry cycle-maintenance gear, such as spare tires, inner tubes, cables, and tools.
MOTORCYCLE	Side panniers have a capacity of up to 80 liters and tank bags up to 25 liters, and weight is not a major issue. However, you'll need to take extra protective equipment, maintenance tools (including spare tires), and, in some cases, spare fuel, all of which will impact on available space.
PACK ANIMAL	Ideal for rough terrain, such as mountains or desert, pack animals can carry serious weight. A horse can transport up to 20 percent of its body weight; a camel can carry a weight of up to 990 lb (450 kg)—although 330 lb (150 kg) is more common—but you'll need to factor in the animal's food and water.
VEHICLE	The size of the vehicle will determine the amount and the weight of gear you can take with you, but you shouldn't have to compromise; extra equipment can be packed in peli boxes—waterproof storage containers with a hard, plastic exterior—that can be strapped to the outside of the vehicle. The main issue will be the weight-versus-fuel ratio: the heavier the vehicle, the more fuel it will use.
CANOE/ KAYAK/BOAT	Some kayaks and canoes are designed to bear loads of up to 660 lb (300 kg), including the craft's occupant(s), so you should be able to take all of your essential gear, plus a few luxuries. Given your proximity to the water, keeping your gear dry is the main priority. Weight and space restrictions don't apply for larger vessels.

PRIORITIZING YOUR KIT

However much gear you decide to take, you should group individual items into one of three categories—first-line, second-line, and third-line—according to their importance to your survival in a worst-case scenario. It's very unlikely that you'll be able to carry all of your equipment all of the time, and in an emergency situation, you may not have time to gather it all together. However, by prioritizing your gear beforehand, you'll have all your essential items on your person, or close enough to grab at a second's notice.

FIRST-LINE GEAR

Known as "bomb-burst gear" in the military, first-line gear (see right) is your basic survival equipment. If something goes wrong and you have to abandon the bulk of your gear to avoid injury, or death, what you're standing in is all you'll have to help you survive. First-line gear should therefore comprise crucial items of outdoor clothing, along with essential items for navigation and safety. Your bushcraft knife, firesteel, and belt pouch make up your belt order (see pp. 60–61). You'll need to risk-assess your particular situation and adapt your kit priorities accordingly, as conditions change, meaning that your first-line gear may alter as the day progresses. Bear in mind that if a removed layer of clothing goes into your backpack, it's no longer part of your first-line kit.

SECOND-LINE KIT

Second-line gear is everything you would need to stay safe on a daily basis under normal conditions. It can either be packed in a small day pack, or contained in pouches on a belt, the idea being that you carry it at all times. For example, a climber may decide to leave the bulk of his or her equipment (third-line gear) while attempting to reach the summit of a peak. By proceeding with first- and second-line gear, the climber is sacrificing equipment for weight and speed, but is still safe, as they have the essentials on them. Examples of second-line gear can include:

- A spare set of clothes, a bivy sack, and cordage
- Emergency rations and first-aid kit (see p. 260)
- Hexamine stove (see p. 58–59) and items to make a hot drink
- Matchless fireset (see p. 118), and a metal cup

THIRD-LINE GEAR

Also known as your "sustainment load," third-line gear is essentially the equipment you need to keep yourself alive and to function for longer periods of time. How much third-line gear you have will ultimately depend on your mode of transport and the amount of equipment you can carry. Examples of third-line gear include:

- A form of shelter—a tent or a tarpaulin
- Cooking utensils, such as a stove or cooking pot
- A backpack
- Food stores
- A sleeping bag and sleeping mat
- Any large water containers or hydration packs
- Wash kit and sanitary items

FLASHLIGHTS

A flashlight is a piece of first-line gear, even if you don't intend to be out in the dark, because situations can change unexpectedly. Modern flashlights use LEDs, and are small, lightweight, powerful, and very efficient. Opt for a "hands-free" headlamp and tape a spare set of batteries to your head-band. Also include two miniature Photon (LED) flashlights in your survival kit (see pp. 60–61).

CHOOSING A BACKPACK

BACKPACKS ARE DESIGNED to enable hikers to carry large loads comfortably. They are made from a variety of materials, most of them waterproof. However, you should always use a separate waterproof bag inside your backpack to ensure your sleeping bag and spare clothes are kept dry; they are of little use in a survival situation if they get wet.

CARRY ONLY WHAT YOU NEED
Regardless of your pack size, you are the one who has to carry it. If in a group, many communal items can be shared. Try not to double up on basics (a party of four does not need four tubes of toothpaste, for example).

WHAT DO YOU NEED?

Your choice of backpack will depend on a number of factors. How long is your trip? How much do you intend to carry? What type of gear will you be taking with you? The answers to these questions will determine the size of pack—known as the "capacity"—that is right for you. Pack sizes range from 30 liters (a day pack) to 80 liters or more (suitable for trips lasting a week or more).

BACKPACK FRAMES

If you have chosen a medium- to large-sized pack, you will have to choose between an internal- or an external-frame pack.

INTERNAL-FRAME PACKS

Internal-frame packs are narrower in profile than their external-frame counterparts and tend to be longer, with either one or two internal compartments and very few pockets on the outside. The frame is integrated within the pack and usually consists of stays or flat bars, about 1 in (25 mm) wide and 1/10in (3 mm) thick, made from plastic. Straps on the outside allow you to compress the pack, which prevents the items inside the pack from shifting and throwing you off-balance.

DAY PACK

Day hikers can sometimes find themselves in trouble because they are not carrying the appropriate equipment to protect them from the elements. A day pack provides you with the means to carry essential items such as food, water, maps, compasses, and protection from cold and wet weather on a day trip.

EXTERNAL-FRAME PACKS

An external frame houses the pack on a lightweight tubing framework. They are great for heavy loads, as the pack sits more squarely on your hips. They are also cooler to carry than internal-frame packs, as the air can circulate between your back and the pack. External-frame packs usually have more pockets on the outside, which allows you to pack your items by category, rather than packing everything into one or two compartments. This makes locating items easier and allows you to unpack only the compartment you need, rather than the entire pack. Chest and waist straps ensure a comfortable fit.

FITTING A BACKPACK

Once you have chosen a backpack of the right size, design, and features for your trip, you will have to make sure that it's a comfortable fit. You will need someone's help, but here are a few tips for ensuring your new pack sits comfortably on your back.

1 First measure your back. The pack you choose should be based on the length of your torso.

2 Make sure the pack is a comfortable fit, allowing for any extra layers of clothing.

3 You can now fine-tune the pack's fit by adjusting the shoulder and hip straps.

PACKING A BACKPACK

Always pack in reverse order: the items you need first should go in last. Heavy items should be placed close to your back to prevent the pack pulling away from your shoulders.

MAKING A SIMPLE PACK FRAME

In a survival situation, having something to carry your gear will be a huge advantage. All you need to make your own improvised backpack is a knife, some wood, some cord, and a waterproof sheet.

1 Cut a light bough about 1 ft (30 cm) below the point where it branches, leaving about 3 ft (1 m) above the fork.

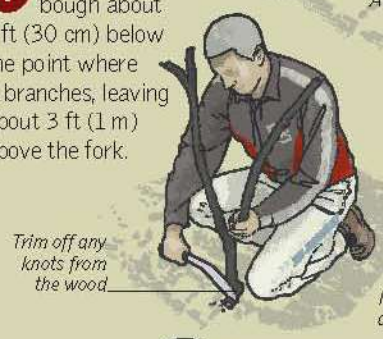

2 Cut notches on all three ends of the bough and tie lengths of cord around the notches to serve as straps to go over your shoulders.

3 Wrap your gear in a groundsheet or waterproof poncho and tie it as high as possible onto the frame on the side opposite the straps.

4 Don't be tempted to overload your improvised backpack, even though it should be able to support a good weight.

DRESSING FOR THE OUTDOORS

MODERN OUTDOOR CLOTHING is technologically advanced and highly sophisticated. New materials and designs are extremely lightweight, durable, and versatile. To make the most of your gear, choose the fabrics and combinations of clothes most suited to the environment and conditions in which you are traveling.

REGULATING BODY HEAT

Don't be tempted to overdress. Sweat can soak you as much as rain, and a sudden change of temperature can leave you vulnerable to hypothermia. Wearing layers helps you to control your body temperature.

LAYERING CLOTHES

The rule of layering is simple: several light layers are better than one heavy layer. Wearing multiple layers gives you flexibility to fine-tune your temperature by taking off or putting on layers to regulate your body heat. Choose fabrics such as wool, fleece, microfleece, and down—all of which are good insulators.

HOW LAYERING WORKS

Layering is effective because it traps air between the layers of clothes and helps you stay warm in any environment. If you wear the correct layers in the right order, your clothing will move sweat away from your body, keep moisture out, and help to keep you insulated. Wear wicking fabrics, such as polypropylene, in hot and cold weather.

THE LAYERING SYSTEM

The top (outer) layer repels rain, while dampness is wicked away from the skin by the base layer closest to the skin. The mid-layers insulate the body and help retain warmth.

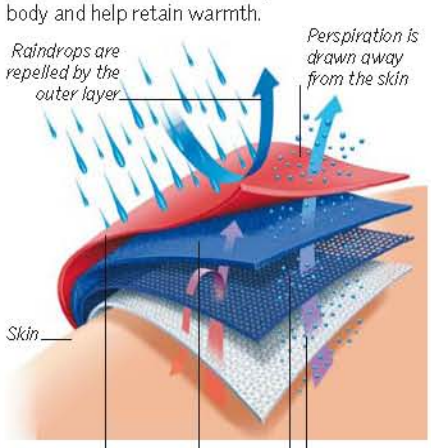

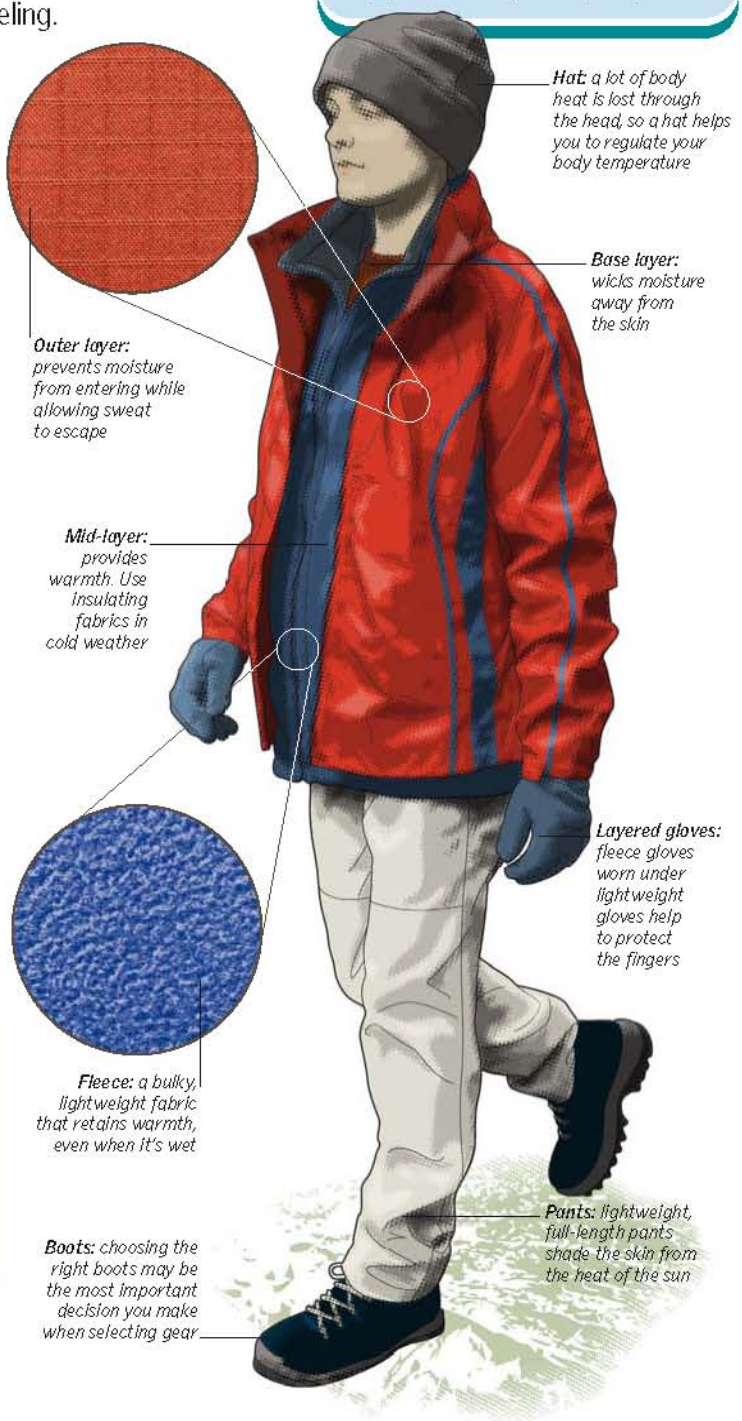

WARM-WEATHER CLOTHING

It's vital to stay as cool as possible in warm weather to avoid heat exhaustion or heat stroke (see p. 272). Too much exposure to direct sunlight can cause sunburn and dehydration, and heat rashes can be exacerbated if sweat is not able to evaporate properly. Choose breathable fabrics that will keep you cool and protect your skin from the sun. Stay hydrated and always wear a hat.

COLD-WEATHER CLOTHING

In cold weather, you need to pay particular attention to your clothing layers. Use a lightweight base layer, which will wick moisture from your skin, several layers of warm, insulating clothing, and an outer shell that's windproof and waterproof. The secret to staying comfortable is to adjust your layers as your body temperature changes. Remove layers as your body temperature rises to avoid sweating.

WET-WEATHER CLOTHING

Sudden downpours can occur almost anywhere and it is easy to be caught out by them. The best fabrics to wear in such conditions are those that are breathable, waterproof, and allow sweat to escape—the best known of these is Gore-Tex®. Always carry the gear necessary to stay warm and dry, and make sure you can access it quickly if you are caught out— it can be the difference between comfort and misery.

HOW BREATHABLE WATERPROOF FABRIC WORKS

Breathable waterproof fabric is a laminate of three layers: two layers of nylon on either side of a thin layer of microporous teflon, which contains tiny pores—50 times smaller than the width of a human hair—that keep rain droplets out but allow perspiration (water vapor) to escape.

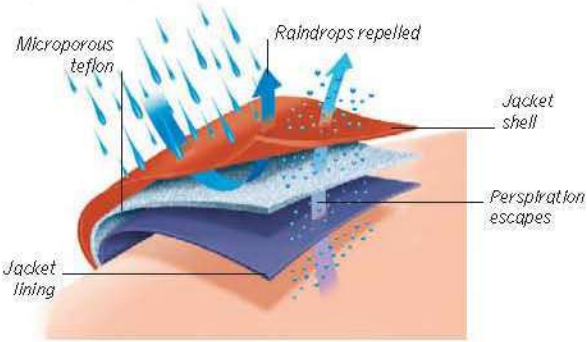

SOCKS

Socks are a vital part of your footwear system. Their function is two-fold. Firstly, they cushion the feet and prevent boots from rubbing and causing blisters. Secondly, they keep the feet warm and dry by wicking moisture away from the feet. Note that socks made from breathable waterproof fabric can be worn with regular boots but should not be combined with breathable waterproof boots.

WICKING SOCK

A lightweight inner sock that helps wick moisture away from the skin and reduces friction.

WALKING SOCK

The thick fabric of the walking sock offers maximum cushioning, comfort, and protection.

GAITERS

Gaiters are useful, both in temperate areas for keeping the feet and lower pant legs dry, and in polar regions for keeping out snow.

FOOTWEAR

A number of factors determine which type of footwear is right for you. The first concerns your personal requirements, which include the shape of your feet and the amount of support you need. You also need to consider how far you are traveling and the type of terrain you'll be covering. Then there is cost: footwear varies from the relatively cheap to the extremely costly. Bear in mind that if you intend to spend a lot of time in the wilderness, it pays to invest a little extra in comfortable, sturdy footwear.

BREAKING IN BOOTS

Your boots and feet should coexist in perfect harmony when you're on the trail. When you buy a pair of boots, wear them around the house and go on short hikes to make sure you have broken them in.

ANATOMY OF A SOLE

A boot sole has several layers to protect and support your feet. Look for a good tread on the outsole, and for cushioning beneath the heel and toe areas.

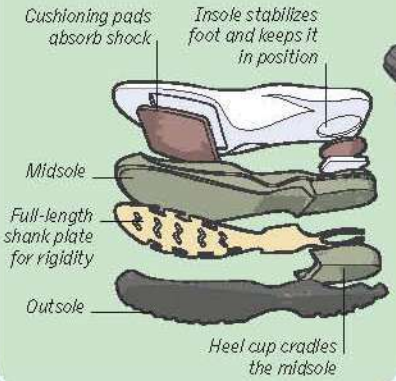

SANDALS

Modern sandal designs offer great support, and are extremely comfortable, while offering additional ventilation.

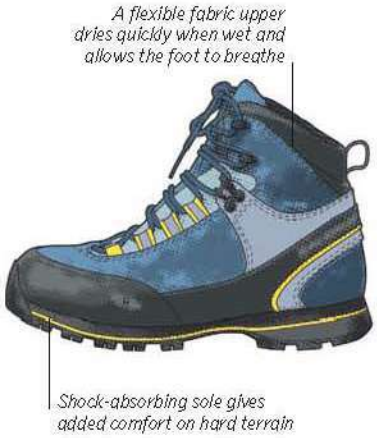

LIGHTWEIGHT BOOT

Fabric and leather hybrids are increasingly popular because they combine the support and traction of a heavier shoe with the flexibility of a sports shoe, so there is less friction on the foot.

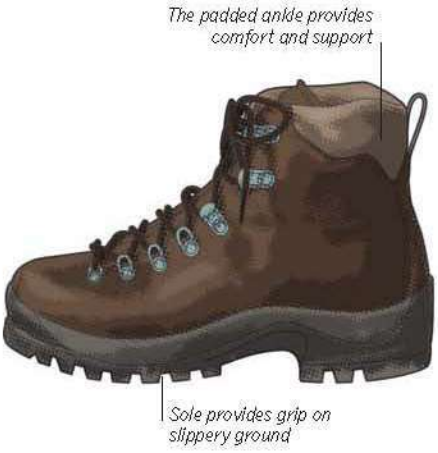

HIKING BOOTS

A compromise between weight, durability, and protection produces a good, all-round leather boot with a sole of hard-wearing rubber, such as Vibram®, water-resistant uppers, and valuable support for the ankles.

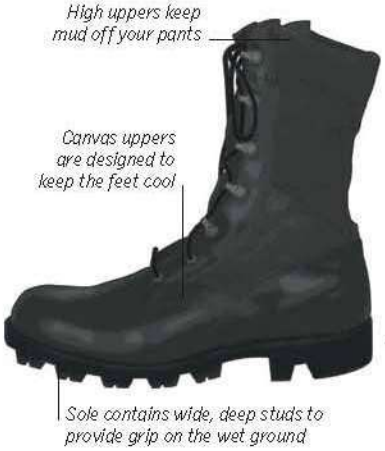

JUNGLE-TREKKING BOOT

This is a high boot made from rot-proof leather and canvas with a directly molded sole. Breather holes in the instep aid ventilation and help to drain moisture in hot or humid conditions.

CLIMBING/HIKING HYBRID BOOT

More flexible than traditional climbing boots, modern hybrid boots are designed to be worn with crampons and to keep the feet warm in very cold environments.

TRUE-LIFE ACCOUNT

EXTREME SURVIVAL— IN THE WILDERNESS

USEFUL EQUIPMENT

- Walking staff
- Pocket chainsaw
- Whistle, flashlight, signal mirror
- Animal deterrent spray
- Water purifiers
- Small AM/FM radio
- Map, compass, GPS
- Survival tin, bushcraft knife
- Cell/satellite phone
- Poncho/bivy sack

BLAKE STANFIELD, 38, AND HIS 65-YEAR-OLD FATHER, Neil, endured being trapped beneath river ice, encounters with bears, and four days and nights in the Alaskan wilderness without food. They survived thanks to good survival skills, knowledge of the local area, and an abundance of water and firewood.

Father and son were dropped by floatplane on Friday, June 6, 2003, at a remote point on the Koyukuk River, deep inside Alaska. They planned to spend seven days rafting approximately 90 miles (145 km) to the town of Bettles. The trip had hardly begun when disaster struck—the raft crashed at speed into a massive sheet of ice, throwing the pair into the freezing water. They were both caught by the strong current and pulled under the ice. Somehow Blake was washed ashore, and he used a branch to haul his father to the riverbank.

Blake quickly realized that Neil was close to hypothermia, so used his waterproof lighter to light a fire of sticks and pine needles. Far from rescue, with their supplies lost and no one expecting their return for six days, their prospects looked bleak. With night approaching they built a shelter.

"THE RAFT CRASHED AT SPEED INTO A MASSIVE SHEET OF ICE"

The next morning, Blake decided to hike the 65 miles (105 km) to Bettles to get help. On Sunday his progress was brought to a halt by the confluence of two large rivers—the Tinayguk and Koyukuk. However, Blake realized that he had reached a landmark on the flight path of supply planes that frequently flew into the interior, so he built a signal fire and waited. Two days later, a pilot spotted him and radioed his position to a US Army base 200 miles (320 km) away in Fairbanks, then dropped him a two-way radio and supplies. The pilot then continued up the river to find Neil, dropping him a sleeping bag, tent, and supplies, before radioing his position to the Army. Both men were picked up the next day—exhausted, malnourished, and close to hypothermia, but alive.

WHAT TO DO

ARE YOU IN DANGER?

NO YES

If you are in a group, try to help any others who are in danger

Get yourself out of it:
Elements—Find or improvise immediate shelter
Animals—Avoid confrontation and move away from danger
Injury—Stabilize condition and apply first aid

ASSESS YOUR SITUATION
See pages 234–35

DOES ANYONE KNOW YOU WILL BE MISSING OR WHERE YOU ARE?

NO YES

If no one knows you are missing or where you are, you will need to notify people of your plight by any means at your disposal

If you are missed, a rescue party will almost certainly be dispatched to find you

DO YOU HAVE ANY MEANS OF COMMUNICATION?

NO YES

You are faced with surviving for an indefinite period—until you are located or you find help

If you have a cell or satellite phone, let someone know your predicament. If your situation is serious enough to be worthy of emergency rescue, and you have a Personal Locator Beacon (PLB), you should consider this option

CAN YOU SURVIVE WHERE YOU ARE? *

NO YES

If you cannot survive where you are and there are no physical reasons why you should remain, you will have to move to a location that offers a better chance of survival, rescue, or both

Address the Principles of Survival: Protection, Location, Water, Food

YOU WILL HAVE TO MOVE **

YOU SHOULD STAY **

DO

- Make an informed decision on the best location to move to—find an elevated position from which to choose a suitable area for survival and rescue
- Regulate your clothing to avoid overheating when moving and hypothermia when static
- Use or improvise a walking staff to help reduce trips and falls
- Improvise shelter when not moving
- Plan your route around potential or known water sources. Filter and purify all water where possible
- Have aids to location accessible while moving and deployed when static

DON'T

- Ignore your fire—be on constant lookout for dry tinder/kindling and fuel
- Walk faster than the pace of the slowest person in the group
- Be careless when walking downhill—a twisted ankle could prove fatal
- Under or overdress. Start off a walk lightly dressed and add or remove layers as necessary

DON'T

- Leave food in your campsite as it will be at risk from predators
- Shelter too deep in the woods despite the protection it gives you from elements. Remember: your location aids need to be seen
- Eat unidentified food, as it could exacerbate your situation through illness. Food is not a priority in a short-term situation

DO

- Check your shelter site before building for natural hazards such as: insects, flooding, rock falls, wild animals, and deadfalls
- Inventory your supplies and ration them
- Keep a fire going: once it's established you can use it to purify water, benefit from the warmth, and signal rescue
- Fill plastic bags or spare clothing with dry foliage and use as a mattress or pillow to insulate you from the cold/damp ground
- If in a group, give everyone something to do, even the children. It will keep them occupied and lessen the worry

* If you cannot survive where you are, but you also cannot move owing to injury or other factors, you must do everything you can to attract rescue.
** If your situation changes (for instance, you are "moving" to find help, and you find a suitable location in which you can stay and survive) consult the alternative "Do" and "Don'ts."

SLEEPING SYSTEMS

REST IS AS ESSENTIAL for survival as food. A good night's sleep can offset much of the worry and stress of a difficult situation. Making sure your shelter is suitable for the environment you're in, and choosing the right sleeping bag, can be vital factors in survival.

SLEEPING BAGS

Although there are many different shapes and styles of sleeping bag, the one you choose should have enough padding to keep you warm at night even without a tent. A sleeping bag must never be allowed to get wet, so always keep it inside a waterproof cover, such as the bag's stuff sack or a bivy sack. Always try a sleeping bag before you buy it. If it's too snug, it will be less effective.

TIPS ON BUYING A SLEEPING BAG

There are many types of bag available. Here are a few things you should consider to ensure you buy the right sleeping bag for you.

- Determine the lowest temperature you're likely to encounter on your trip and choose a bag that will perform at this temperature.
- Synthetic bags are cheaper and easier to clean than down bags, and continue to provide insulation even when wet.
- Down bags are more expensive than synthetic bags, but provide a better warmth-to-weight ratio and last longer.
- A mummy-shaped bag provides better insulation than a rectangular-shaped bag, but has less room to move around in.

DOWN OR SYNTHETIC

Sleeping bags are made with either a down or a synthetic filling. Down is made up of highly specialized insulating feathers, normally from ducks or geese, and is unsurpassed when it comes to a warmth-against-weight ratio. However, it is ineffective when wet and can cause an allergic reaction in some people. Synthetic fillings range from simple hollow fibers all the way up to complex fibers designed to mimic the structure of down. They still retain some of their insulating capacity if they get wet.

TEMPERATURE RATINGS

Because manufacturers have no standardized way of registering how warm a sleeping bag is, temperature ratings for sleeping bags remain an inexact science. Some manufacturers give a bag both a comfort rating and a survival rating—known as a "season rating.' These give some indication as to how manufacturers expect a sleeping bag to perform.

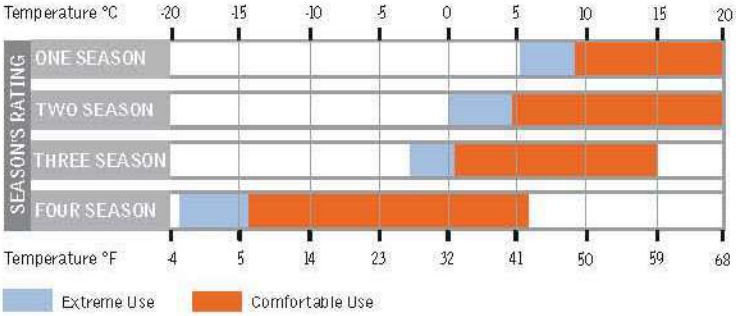

CONSTRUCTION

The principle of all sleeping bags is the same: to trap air and prevent its circulation so that the body heats the trapped air and keeps you warm while you sleep. This can be achieved in a number of ways.

BOXWALL

Used in cold-weather down bags, the filling is contained within box-like sections to minimize bunching.

SEWN-THROUGH

Filling is in separate oval channels, although heat is still lost through the stitching.

OFFSET CHANNELS

The filling is laid out in two layers to prevent bunching and heat loss through the stitching.

SHINGLE

Slanted layers of overlapping fibers that fill with warm air to aid insulation.

USING YOUR SLEEPING BAG

Before using your sleeping bag, always check inside for spiders and other dangers, and shake it to make sure the filling is distributed evenly. During use, if the bag is too warm, simply unzip it to cool down, or use the bag as a quilt. If it's too cold, either put on some clothing, or use a silk liner or waterproof bivi bag to increase its performance. Always air your sleeping bag well after using it.

SLEEPING ACCESSORIES

You'll need to get a sleeping mat as well as a sleeping bag to ensure a comfortable, dry night's sleep. Additional gear, such as bivy sacks, sleeping bag liners, and inflatable pillows, will also increase comfort.

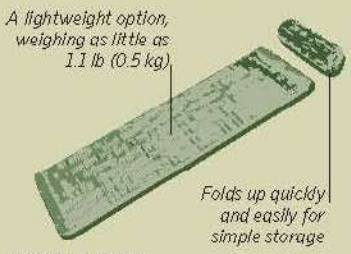

FOAM MATS

The lightest and simplest type of mat, foam mats are best used on soft surfaces, such as sand or pine needles.

AIR MATS

They may be heavier and more expensive than foam mats, but air mats provide superior comfort and thermal insulation.

SLEEPING BAG LINER

Usually made of cotton or silk, sleeping bag liners trap air between your body and the sleeping bag.

BIVY SACKS

A simple waterproof bag that is used over a sleeping bag to keep it dry or used on its own as a shelter.

INFLATABLE PILLOW

An inflatable pillow adds some luxury to expeditions. Alternatively, fill a stuff sack with spare clothing.

SPACE BLANKET

A lightweight foil blanket for use in emergencies, space blankets can also be used as a temporary canopy.

PORTABLE SHELTERS

Shelter from the elements at night is essential for long-term survival. A variety of products are available to you, depending on the conditions you're about to enter and the amount you're prepared to carry. You can choose from lightweight options, such as hammocks and variations on bivy sacks, and the more traditional alternatives, tents, which are heavier but can accommodate up to eight people.

USING A HAMMOCK

If you have a hammock you can camp even if the ground is muddy, rocky, or on an incline. The advantage of a hammock is that it's lightweight and can be erected just about anywhere. Modern hammocks provide a shelter and bed solution in one portable package.

USING A TARP TENT

A cross between a tarpaulin and a tent, a tarp tent is a highly versatile, extremely light form of shelter that's capable of withstanding even the most extreme conditions. Groundsheets should be laid out on the ground for added protection, and a suitable sleeping bag is essential. A one-person version can weigh as little as 18½ oz (0.52 kg).

USING A HOOPED BIVY

The hooped bivy is a waterproof and breathable bivy sack that has been designed to form a one-person shelter. It usually incorporates pegging points and a zippered and hooped entrance that form a small vestibule for your backpack. Many incorporate a heavy-duty groundsheet and can be set up and taken down in a matter of minutes.

USING A TENT

Tents keep the rain and wind out while retaining warmth. The outer flysheet should be strong, impermeable, taut, and able to withstand high winds. The inner tent—which may be attached to a flysheet—is loose and made of permeable material to minimize condensation but retain heat. The inner and outer parts of the tent must not touch each other; if they do, heat may be lost and condensation may form on the underside of the flysheet, then seep into the tent.

WARNING!
In some climactic conditions, such as extreme high winds, severe storms, or heavy snow cover, stakes and guylines may not be enough to anchor the tent securely to the ground. This could result in the tent being blown away—and could have severe consequences for the tent's occupants. In such conditions, use extra means—such as rocks or logs—to anchor your tent securely.

THREE-SEASON TENT

Three-season tents are designed for use in a variety of climates. The inner layer is made of a lightweight material that offers both ventilation and protection against insects. The flysheet—which should extend about 4 in (10 cm) above the ground—and the groundsheet are made of waterproof fabric.

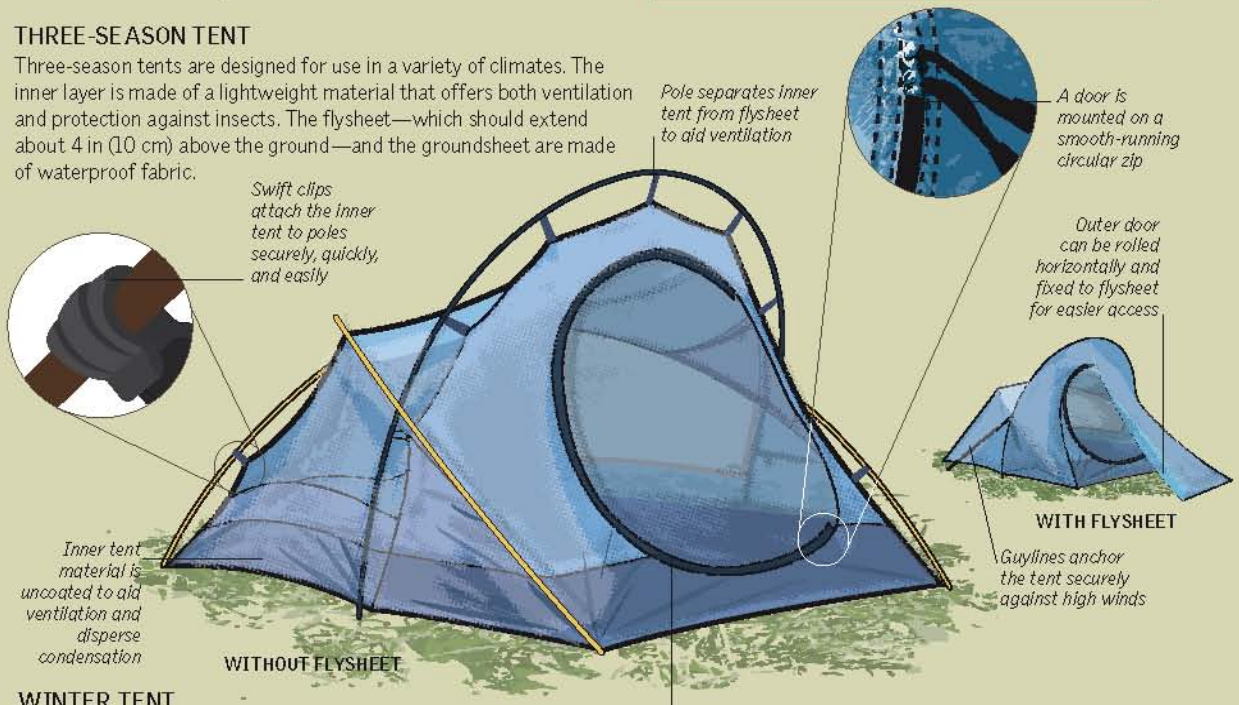

WINTER TENT

For cold-weather climates, look for a tent that offers extra-strong poles, storm windows, and ample pockets so that you can store your gear inside the tent. However, these extra features increase the tent's weight. The dome shape sheds snow well, and withstands high winds.

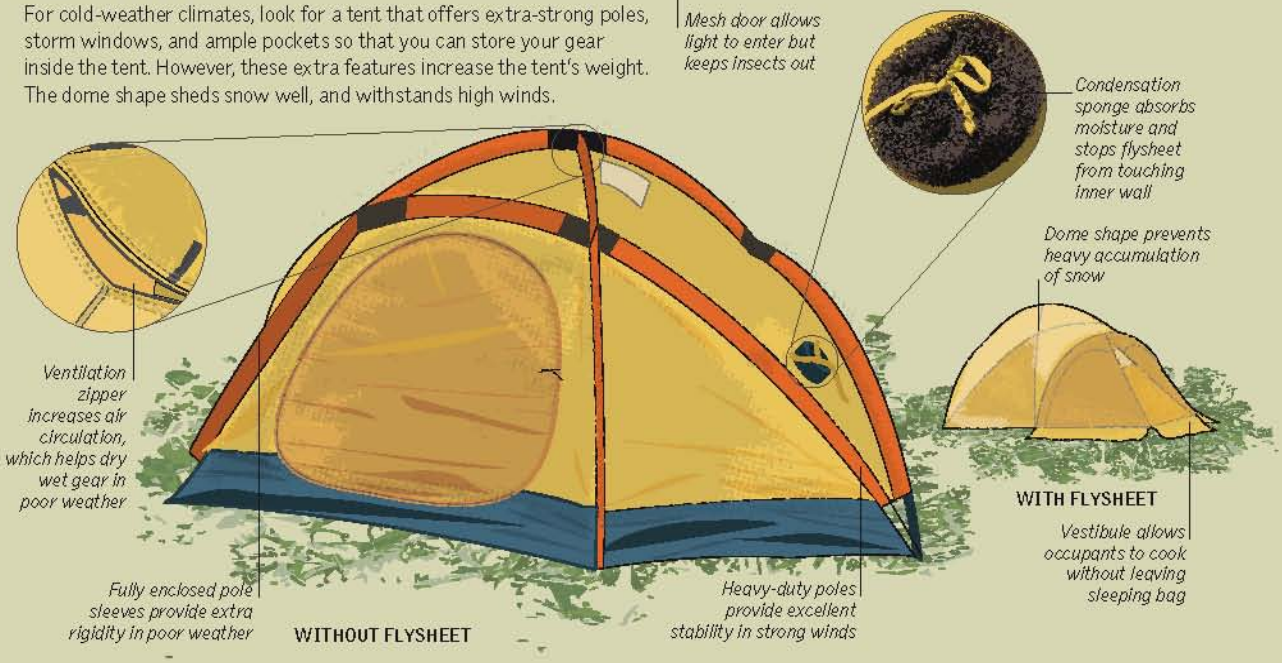

EATING ON THE TRAIL

THE FOOD YOU EAT determines how efficiently your body will function. This is never more important than when you're undertaking an expedition that will test your physical and mental endurance beyond normal levels.

FOOD FOR LIFE

Your body requires energy to fuel the chemical and biological reactions that take place within it, and to power your muscles. This, and other nutrients, comes from the food you eat (see below).

BODY FAT AND SURVIVAL

Being physically fit and active keeps you strong and less prone to injury, but being slightly overweight can actually be an advantage in a survival situation: 1 lb (½ kilo) of body fat converts to roughly 3,500 calories—the amount of energy you would usually expend during an average day.

TYPES OF NUTRITION	EXAMPLES	BENEFITS
CARBOHYDRATE	▪ Rice, pasta, bread, and cereals ▪ Fresh and dried fruit ▪ Root vegetables ▪ Chocolate and candy	▪ Major source of energy, as it converts most readily to glucose, which is your body's preferred source of fuel.
PROTEIN	▪ Meat and poultry ▪ Fish, such as tuna and salmon ▪ Eggs and dairy products ▪ Beans, legumes, and seeds	▪ Essential for growth and repair of muscle tissue. ▪ Increases your stamina and energy levels, and helps you fight fatigue. ▪ Makes you feel satisfied after eating.
FAT	▪ Nuts and seeds ▪ Dairy products, such as butter and cheese ▪ Meats, such as bacon and sausage ▪ Cooking oil	▪ Good source of energy. ▪ Makes you feel satisfied after eating.
FIBER	▪ Fresh and dried fruit, especially the skins ▪ Vegetables, beans, and seeds ▪ Whole-grain cereals ▪ Whole-meal bread, whole-meal pasta, and brown rice	▪ Provides bulk to help your digestion. ▪ Helps to make you feel full.
VITAMINS AND MINERALS	▪ Fresh and dried fruit ▪ Green vegetables ▪ Meat and fish ▪ Nuts and seeds	▪ Essential for growth and the repair of your body's tissues. ▪ Help you to maintain healthy teeth and eyes. ▪ Aids bloodflow and the production of energy.

BALANCING YOUR ENERGY LEVELS

You replace the energy that you expend on a daily basis through the food you eat. The energy obtained from food is measured in calories, and you require a certain number of calories each day to remain alive. If you consume too few calories, you will lose weight. However, eating too many calories will lead to weight gain, which can cause long-term health issues. The number of calories you require per day depends on many things, such as your age and level of activity (see right). When you don't replace the energy expended and continue to work, you start to use the energy store contained in your body's fat.

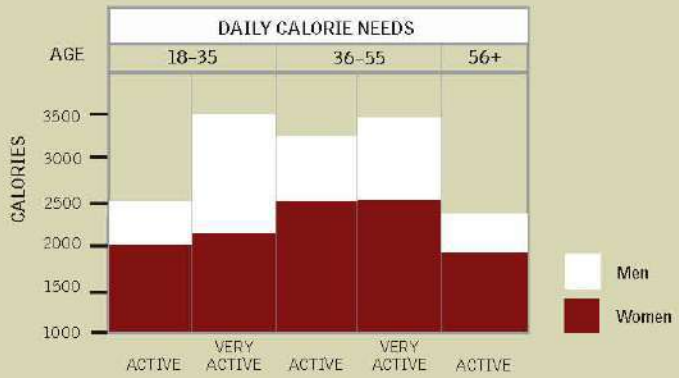

WILDERNESS DINING

Even if you're only going on a short hike, you should take food with you, such as fruit and chocolate, to replace the energy you're using. For longer expeditions, you'll need to take a number of different foodstuffs—for energy and for variety.

WARNING!
In a survival situation, don't eat for the first 24 hours unless to replace expended energy. Ration your food, as digestion uses up water.

TYPES OF FOOD	EXAMPLES	BENEFITS
BREAKFAST CEREALS AND DRIED FRUIT	▪ Trail mix—also known as "gorp" ▪ Muesli or granola ▪ Oatmeal ▪ Dried fruit	▪ Provide a vital source of carbohydrates, vitamins, and minerals to set you up at the start of the day. ▪ Contain dietary fiber to aid digestion.
TRAIL SNACKS	▪ Cookies and crackers ▪ Energy bars and cereal/granola bars ▪ Fruits, such as apples, pears, and oranges ▪ Nuts and seeds	▪ High in carbohydrates to help you maintain energy levels. ▪ Help to keep your hunger pangs at bay between main meals.
HIGH-CALORIE SNACKS	▪ Dehydrated ice cream ▪ Chocolate ▪ Hard candies ▪ Pudding mixes	▪ Provide a quick boost of energy, when required. ▪ Offer a comfort factor, although they are of limited nutritional value.
PRE-PREPARED MAIN MEALS	▪ Boil-in-the-bag meals—you can buy a wide variety of fully prepared meals, sealed in a foil bag. Cook for five to seven minutes. ▪ Freeze-dried and dehydrated meals—simply add boiling water, stir, and wait about five minutes. Extremely lightweight and take up little space. ▪ Canned meals—a great variety of them are available, but they are bulky and heavy.	▪ Provide a balanced range of food groups, vitamins, and minerals to help your body to recover and repair itself after the day's exertions. ▪ Quick and easy to prepare, and require few utensils. ▪ Many require minimal fuel to cook them, or water to rehydrate them. ▪ Can be bulked out with beans, legumes, and grains. ▪ Smaller cans may have a pull-top to open them.
MEAT, FISH, AND CHEESE	▪ Canned fish, such as sardines or tuna ▪ Cured meats, such as ham, salami, and beef jerky ▪ Hard cheese, such as cheddar and parmesan	▪ A great source of proteins and fats. ▪ Easy to store and long-lasting, although cans are heavier than other sources of food.
BEANS, LEGUMES, AND GRAINS	▪ Lentils and dried peas ▪ Dried beans and pearl barley ▪ Rice and pasta ▪ Couscous	▪ Can be used to bulk up main meals and help you to feel full. ▪ An important source of carbohydrates. ▪ Also provide fiber and protein.
MIXERS	▪ Flour ▪ Salt ▪ Suet ▪ Sugar	▪ Flour and salt can be used to make dough. ▪ Sugar and salt help to make wild food palatable. ▪ Suet is an invaluable source of fat that can be used to supplement meals.
FLAVORINGS	▪ Powdered soup ▪ Tomato puree ▪ Powdered sauces ▪ Gravy cubes	▪ Add flavor and variety to your meals. ▪ Especially useful if you're cooking with wild food.
BEVERAGES	▪ Tea ▪ Coffee ▪ Chocolate powder ▪ Dried milk	▪ Provide warmth and comfort, although they are of limited nutritional value. ▪ Chocolate powder is high in carbohydrates. ▪ Dried milk is an excellent source of calcium.

CAMPING STOVES

ALTHOUGH NOTHING CAN BEAT the satisfaction of building your own campfire, the convenience and reliability of purpose-made camping stoves make them an essential piece of cooking equipment, especially in areas where open fires are banned.

THE BENEFITS OF CAMPING STOVES

Modern camping stoves are very light and collapse down to an incredibly small, compact size. Gas-fueled stoves are by far the most common, but liquid-fueled stoves are also worth considering. Both have their own advantages and disadvantages (see below), but your choice will be dictated by space and weight constraints, the distance you're traveling, and environmental factors. A hexamine stove is a good item to have in a survival situation (see opposite).

CHOOSING A GAS- OR LIQUID-FUELED STOVE

Gas-fueled stoves are safer, cleaner, lighter, and easier to use than liquid-fueled stoves, but are less economical, and not suitable for low-temperature, high-altitude use. Liquid stoves offer greater stability, fuel versatility and economy, but you need to pump and prime (preheat) them before use, and they are less clean.

TRANSPORTING FUEL

There are many factors to consider when transporting fuel to make your expedition safer:

- When hiking, pack your fuel container upright in the middle of your backpack, and surround with clothing to protect it from knocks. Make sure you can access it easily when you reach camp. Use similar principles when traveling by other methods.
- Ensure that your fuel bottles look very different from your water bottles to prevent confusion.
- If you're going on a long expedition, check the availability of the fuel, and consider how you will dispose of cylinders safely.

	GAS-FUELED STOVES		
TYPE OF STOVE	BUTANE OR ISOBUTENE	PROPANE	BLENDED FUEL
INFORMATION	• Both fuels sold in disposable, pressurized canisters.	• Fuel sold in disposable, pressurized canisters. • Used in most portable stoves and barbecue grills.	• Fuel sold in disposable, pressurized canisters. • Is a mix of butane, propane, and/or isobutene.
PROS	• Lightweight, sealed fuel container. • Butane very efficient and burns at a high temperature. • Isobutene more efficient than butane, and performs better in cold conditions.	• Lightweight, sealed fuel container. • Fuel burns at a very hot, steady heat, and produces an almost sootless flame. • Performs very well in cold conditions.	• Lightweight, sealed fuel container. • Fuel safer than pure propane, and performs better than either pure butane or isobutene in cold conditions. • Isobutene blends give a more efficient flame.
CONS	• Higher cost and lower heat output than liquid fuels. • Fuel efficiency reduced at higher altitudes. • Butane less efficient than isobutene in temperatures below 50°F (10°C).	• Higher cost and lower heat output than liquid fuels. • Fuel efficiency reduced at higher altitudes. • Highly combustible, so less safe than other fuels.	• Higher cost and lower heat output than liquid fuels. • Fuel efficiency reduced at higher altitudes.

USING CAMPING STOVES SAFELY

Always take care when storing the fuel and using your stove. Use only the stove manufacturer's recommended fuel, and check all connections and fittings before each and every use. Also:

- Clear the cooking area of dry vegetation and leaves, in case you spill fuel or knock over your stove.
- Place your stove on a level and stable surface.
- Keep naked flames and sparks away from the fuel, especially when filling or refilling the stove. Before refilling, make sure the flame is out and the stove has cooled down.
- Use the cooker in a well-ventilated space. Liquid fuels, in particular, can give off poisonous fumes, such as carbon monoxide, and all cookers also burn vital oxygen.
- Don't use a stove inside your tent. In addition to the risk of poisonous fumes, fire is also a significant danger. A confined space with limited access, your tent can be reduced to ashes in minutes, along with everything inside, and you may not be able to escape in time to protect yourself. Losing your tent and your equipment is bad enough, without you receiving third-degree burns at the same time.
- Don't leave a burning stove unattended. If it topples over, it could easily start a fire. If the stove flame goes out, a spark or naked flame could also ignite the vapors and cause an explosion.

THE HEXAMINE STOVE

A very useful piece of emergency equipment, the hexamine stove (or hexi-stove) burns solid blocks of hexamethyl-enetetramine (also known as "methenamine"), and folds away compactly. The stove acts as a small windbreak and cooking platform, while the fuel blocks it uses provide a very reliable means of starting a fire in any conditions. The blocks produce a smokeless flame, don't liquefy while burning, and leave no ashes.

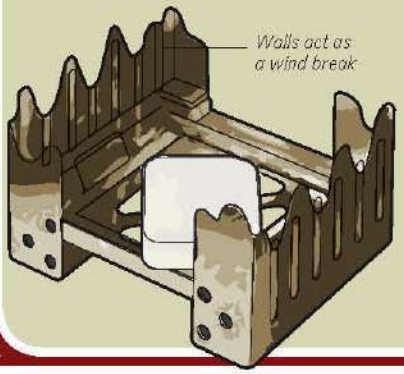

	LIQUID-FUELED STOVES		
TYPE OF STOVE	WHITE (COLEMAN) FUEL	MULTI-FUEL	DENATURED ALCOHOL
INFORMATION	• Fuel sold in disposable canisters.	• Stove can use a variety of fuels, from unleaded gas, kerosene (paraffin), aviation fuel, and white fuel, to diesel—most of which are readily available.	• Fuel sold in disposable plastic containers. • Stove basically consists of a vessel to burn the fuel in, and a windshield that also acts as a cooking platform.
PROS	• Fuel can withstand low temperatures, high altitudes, and most weather conditions. • Produces a clean flame.	• Stove uses a range of fuels. • Fuel is non-pressurized, so can be stowed safely in almost any airtight container. • Stove tends to perform reliably, efficiently, and quickly, producing a hot flame.	• Fuel is non-pressurized, so can be stowed safely in almost any airtight container. • Economical and produces a clean flame. • Stove is lightweight, stable, and safe. • Usually supplied with pots and pans.
CONS	• Stove must be pumped to generate required pressure for use. • Fuel needs to be primed.	• Potential for spillage of flammable fuel. • Stove must be pumped to generate required pressure for use. • Fuel needs to be primed. • Some fuels produce soot and noxious fumes.	• Potential for spillage of flammable fuel. • Stove slower to heat up than other liquid stoves, because it's not pumped. • Flame vulnerable to wind and other environmental conditions.

YOUR SURVIVAL KIT

A BASIC SURVIVAL KIT is an essential item to take with you on any outdoor expedition. It should be compact enough to carry on your person at all times, and its contents should address the key principles of survival: protection, location, water, and food.

LAYERS OF A SURVIVAL TIN

EXAMPLE OF A SURVIVAL KIT

Choose a tin with a waterproof seal and locking clasps. While you can buy ready-made kits, you should always adapt the contents, both to your needs and to the environment you'll be in. Ideally, items in the survival kit should be high-quality and multi-purpose—for example, the tin itself could be used as a cup, a small cooking pot, or even a signal mirror. Your kit should contain the following:

LAYER 1

- **Blister medical pads and bandages:** useful for minor cuts and blisters. Include a range of bandages (both waterproof and fabric) in a selection of sizes. Waterproof bandages can also be used to mend holes in tents and tarpaulins.

LAYER 2

- **Antiseptic wipes:** for treating wounds and cleaning bites. Can also be iginited by a spark and used to get a fire going.

LAYER 3

- **Water purification tablets:** choose from tablets or iodine (but don't use iodine if you're allergic to it, or to shellfish). Water will need to be filtered before being purified.

LAYER 6

- **Survival saw, or pocket chainsaw:** can be wrapped around the inside of the tin or cut in half, if space is limited.
- **Single-edged razor:** multi-purpose—from skinning an animal to cutting cord. Store in its packaging.
- **Needle and cotton:** use strong, waxed cotton, pre-threaded through the needle.

LAYER 7

- **Flashlights:** two small Photon (LED) lights—one white and one red—taped in "off" position.
- **Mini multi-tool:** see panel, opposite.
- **Chainsaw handles:** use with pocket chainsaw.
- **High-viz card, signal mirror:** location aids.
- **Compass:** an emergency back-up.
- **Flint, fire steel, tampons:** for starting fires.
- **Snare wire:** for animal traps and lashing.

LAYER 8

- **Waterproof matches and tinder balls:** for starting fires. Store the matches in a small, resealable plastic bag.
- **Pencil:** sharpened at both ends.
- **Potassium permanganate:** dissolve in water to sterilize water at low concentration, and to clean wounds at a high concentration. Can also be used with sugar to make fire. Store in a waterproof container.

CHOOSING A MULTI-TOOL

A useful item to carry with you on your trip, multi-tools are usually designed with a specific task or hobby in mind, and include useful features for survival, such as a small compass, pliers, saw, and various blades. Use one that has quality tools that most suit your needs, with a comfortable grip, and ensure that the blades are lockable to prevent injuries. Carry your multi-tool in a secure pocket, or in a pouch on your belt, and include a miniature version in your survival tin (see below). Bear in mind that your multi-tool should be an addition to, not a replacement for, your bushcraft knife (see p. 146).

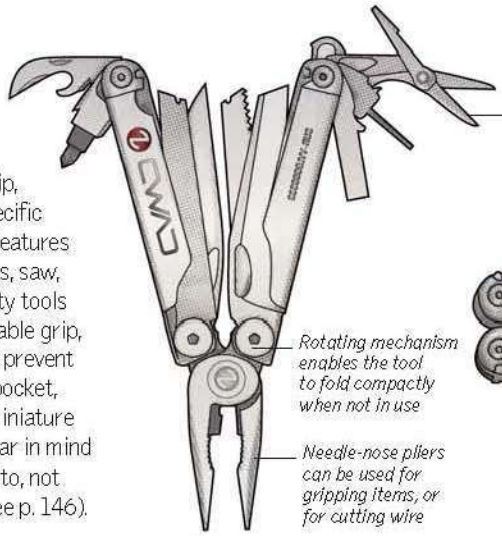

LAYER 4

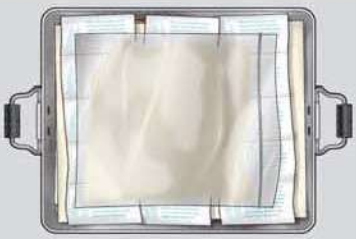

- **Petroleum jelly:** apply to chapped lips, rashes, and sores. Can also be smeared onto tampons to make them burn longer. Store in a small, resealable plastic bag.

LAYER 5

- **Waterproof notepaper:** for drawing maps or leaving messages.
- **Photograph of loved ones:** a psychological incentive in a survival scenario.
- **Credit card:** an effective means of extracting insect stings (see pp. 266–67).
- **Money:** wrapped in cellophane.

LAYER 9

- **Non-lubricated condoms:** can be used to carry water, or as a waterproof cover for smaller items, such as your mobile phone.
- **Mini fishing kit:** if you are near water, fish can be easier to catch, prepare, and cook than mammals. Fishing line can be also used for other survival tasks. Should contain a selection of hooks, flies, swivels, and split-shots.

TIN LID SEALED WITH TAPE

- **Sailmaker's needles:** multi-purpose—can be used as an arrow point, or for mending tents and tarpaulins. Wide eyes are best.
- **Safety pins:** for securing clothing, or mending your sleeping bag or tent.
- **Mini glowsticks (cyalumes):** useful for emergency lighting, and as a location aid.
- **Single-edged razor:** multi-purpose (see Layer 6). Store in its packaging.

ADDITIONAL USEFUL ITEMS TO INCLUDE

While your survival tin's size may be restrictive, you can always improve your kit with items that can fit on your belt, or in a belt pouch, known as a "belt order." This will form part of your first-line equipment (see p. 43).

- **Space blanket or aluminum foil:** can be used as a signaling device; for shelter; to carry, store, and heat water; or to cook in. Many are dual-sided: one silver, the other green for camouflage, or orange to stand out.
- **Plastic bags:** you can never have too many. Numerous uses—from a water carrier to a transpiration bag.
- **Medicines (such as painkillers and antibiotics):** should not replace your main first-aid kit, but ensures you have some basics if you're separated from your gear.
- **Small candle:** once lit, this will provide a reliable flame that you can build your fire around.
- **Nylon stockings:** can be used for warmth, or as an improvised water filter, mosquito net, or fishing net.
- **Small AM/FM radio:** battery- or solar-powered.
- **Surgical tubing:** enables you to reach water in otherwise inaccessible rock crevices.
- **Fire tin or matchless fire set:** self-contained methods of starting a fire, when no natural fuel is available.

ON THE TRAIL

FIND YOUR WAY

Before you set off on an expedition, you should have at least a basic understanding of how to read a map and use a compass. Your ability to correctly assess a map of the area you intend to visit will allow you to make informed decisions during your preparation. If you understand the area and terrain, your chances of getting lost will be reduced, and you will be able to continually evaluate your progress and therefore alter your plans as necessary. You will also be able to plan the safest and most appropriate route, and locate water, shelter, and areas that will allow you to use your location aids properly. If you're proficient with a map and compass, you'll have no cause to worry about getting lost or straying off-track and will be free to enjoy your outdoors experience.

In a survival situation, you will be faced with many tough decisions. You may have to decide whether to stay where you are and await rescue or move to an area that offers a better chance of survival and rescue. Your ability to navigate effectively—whether

In this section YOU WILL DISCOVER...

- how to **use your map** to tell the **lie of the land...**
- that the **path ahead** is as long as a **piece of string...**
- the difference between **eastings** and **northings** and an **ERV** and a **GPS...**
- how a **detour** could be the **most direct route...**
- how to **navigate** using your **hair...**
- why an **anvil of cumulonimbus** is no **fair-weather cloud...**

by using a map and compass or by using natural features—will play a major role in your decision-making process. Whilst a Global Positioning System (GPS) is an excellent aid, it relies on batteries and technology—both of which can fail.

An understanding of weather patterns will also allow you to evaluate conditions and make informed decisions as you travel. The ability to assess your situation and modify your plans means that you will be able to avoid many potential survival situations.

The ability to take a bearing (see pp. 68–71) and navigate using pace counting (see p. 72) could prove to be crucial skills in a survival situation.

TEST HOW ACCURATELY you can walk on a set bearing and pace out a set distance by following the exercise below. If you have stayed on your bearings and your pacing has been accurate, you should finish at your starting point. Pick a suitable area where you can walk at least 330 ft (100 m) in any direction. Don't cheat by heading for your marker! For this exercise to be useful, make sure you follow your bearings and count your paces to navigate.

1. Place a marker on the ground. Dial a bearing onto your compass (110° in our example).
2. Walk on this bearing, counting your paces until you think you have traveled 330ft (100m), then stop.
3. Add 120° to your original bearing. Dial this new bearing (230° in our example) onto your compass.
4. Pace out another 330 ft (100 m) on your new bearing, then stop.
5. Add another 120° to your latest bearing. Dial a last bearing (350° in our example) onto your compass.
6. Walk on your final bearing for a further 330 ft (100 m). You should be back at your starting point.

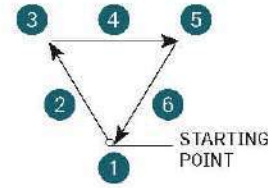

> Never underestimate the **combined power** of a map, a compass, a GPS unit, and the **skills and knowledge** to use them well—your life may **depend on it**

MAPS AND MAP-READING

A MAP IS A TWO-DIMENSIONAL representation of a three-dimensional area—from a map you can determine distance and height on the ground. If you are able to read a map, and can interpret the information it contains, you can visualize what an area looks like and use these features as landmarks to make navigation easier. Your map is very important, so keep it safe.

UNDERSTANDING MAPS

Although there are many types of map available—with varying levels of detail and scale—topographic maps are best for hiking. They show important features, such as rivers, roads, railroads, paths, buildings, and forested areas, and also depict the lie of the land through the use of contour lines (see opposite) to represent height.

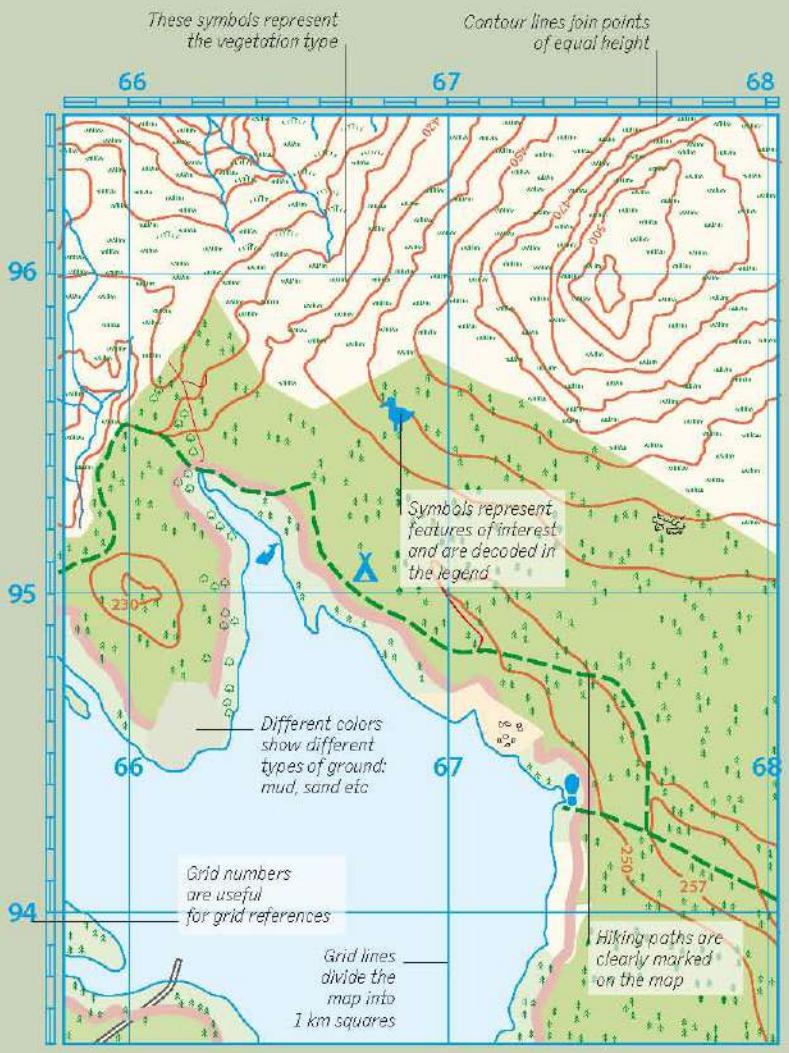

THE LEGEND

Topographic maps incorporate a legend, or key, which deciphers the information shown on the map. It's important to familiarize yourself with the symbols used, as this will help you visualize what's being represented on the map more effectively. Some examples are listed below:

HEIGHTS AND NATURAL FEATURES

Water

Mud

Sand; sand and shingle

Vertical face/cliff

Loose rock

Boulders

Outcrop

Scree

Coniferous trees

Non-coniferous trees

Managed woodland

Orchard

Scrub

Bracken, heath, or rough grassland

Marsh, reeds, or wetlands

TOURIST AND RECREATIONAL INFORMATION

Nature reserve

Campsite

Fishing area

Walks/Trails

THE SCALE

Hiking maps are always drawn to a scale provided in the legend. This is a ratio of how much you would have to enlarge the map to reach actual size. 1:25,000—whereby 4 cm on the map equates to 1 km on the ground (2½ in to 1 mile)—is the most useful level of detail for hikers. A smaller-scale of 1:50,000, for example, will provide a more basic overview of the terrain.

MEASURING DISTANCE

Maps are drawn to scale so that you can use them to accurately estimate distances on the ground. Being able to measure distance is important because it means you can calculate the most direct—and energy-efficient—route to your destination. In a survival situation, every last bit of energy counts, so the shorter the distance, the better.

USING THE GRID LINES

The simplest method of measuring distance on a map is to use the grid lines—on a scale 1:25,000 map each grid square represents 1 km (traveling diagonally across a square, it's approximately 1.5 km). You can also lay a piece of paper between the two points, mark the start and end of your route, and place it underneath the scale line to read the distance.

USING STRING OR SOLDER

Invariably you will deviate from a straight line and will have to navigate around obstacles or bends. A far better way of measuring distance is therefore to take a piece of string, curve it around your intended route, and transfer it to the scale line. Solder wire (use wire that is lead-free) is even more accurate because it holds its shape on the map and remains flexible.

CONTOUR LINES

Topographic maps feature lines called contour lines. A contour joins points of equal height above sea level, and allows the topography of the ground to be depicted in detail. The contour interval is specified in the legend—for maps with a 1:25,000 scale you would usually expect to see a 5-meter vertical interval between each line, although for maps of mountainous regions this interval may be 10 meters. The ability to look at the contour lines, and imagine how they translate to the ground, is a skill that takes a little while to acquire, but once mastered will allow you to read a map more proficiently.

USING CONTOURS

Knowing how steep the ground is will greatly improve your navigational skills, as well as your route-planning (see p. 73). Walking up and down hills uses a lot of energy, so it is far better to follow the contour lines on your map to go around hills instead.

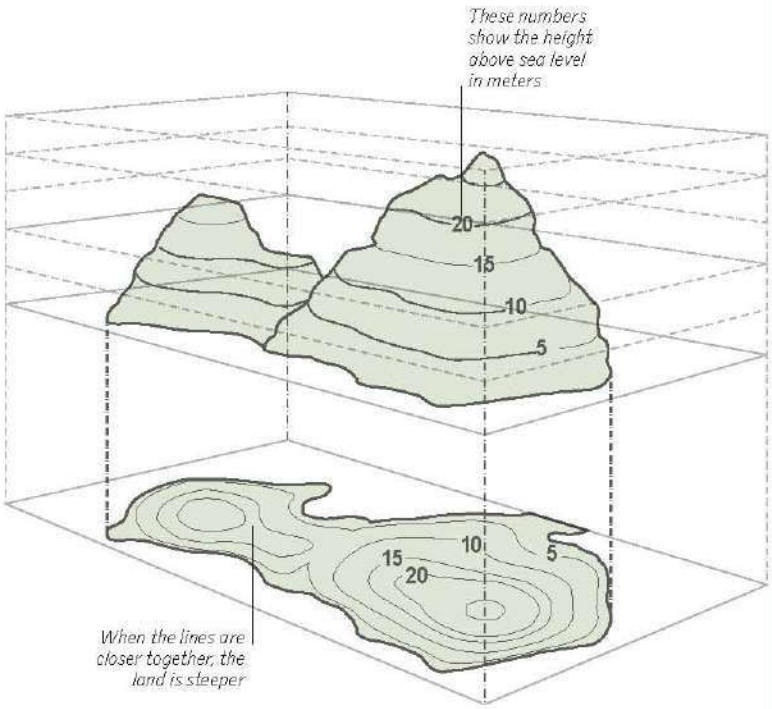

GRID REFERENCES

Maps contain grid lines, which help you locate a specific point anywhere on a map, using a unique number known as a grid reference. The vertical grid lines are called "eastings," as they increase in value as they travel east on the map. The horizontal lines are called "northings," where these perpendicular lines cross creates the grid.

WORKING OUT GRID REFERENCES

Use the numbers on the grid lines and apply the easting number first. On a scale 1:25,000 map, where the grid lines are 1 km apart, the shaded area on the grid below would have a four-figure grid reference of 2046, and indicate a grid square that is 1 km by 1 km. To be more accurate, use a six-figure number—mentally divide the square into tenths. The cross would therefore have a grid reference of 185445.

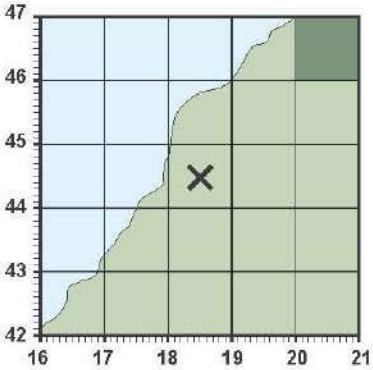

GET YOUR BEARINGS

IN ADDITION TO understanding and reading maps, it's important to know how to orientate your map to the land so you can use it to navigate. It's possible to do this by sight, although in most cases you'll need a more reliable method and for this you should use a compass. Use your compass to determine direction, orientate your map and yourself, take and plot bearings, and navigate.

CARDINAL POINTS

The four cardinal points of the compass are:

- North (N)—0°/360°
- East (E)—90°
- South (S)—180°
- West (W)—270°

HOW A COMPASS WORKS

A compass needle is a magnetized piece of metal that, when allowed to rotate freely, will orientate itself to the North and South magnetic poles. Always hold a compass level.

STANDARD COMPASS

This standard orienteering Silva compass is a good compass for hiking. It allows you to set your map and work out distances using the printed scale on the base.

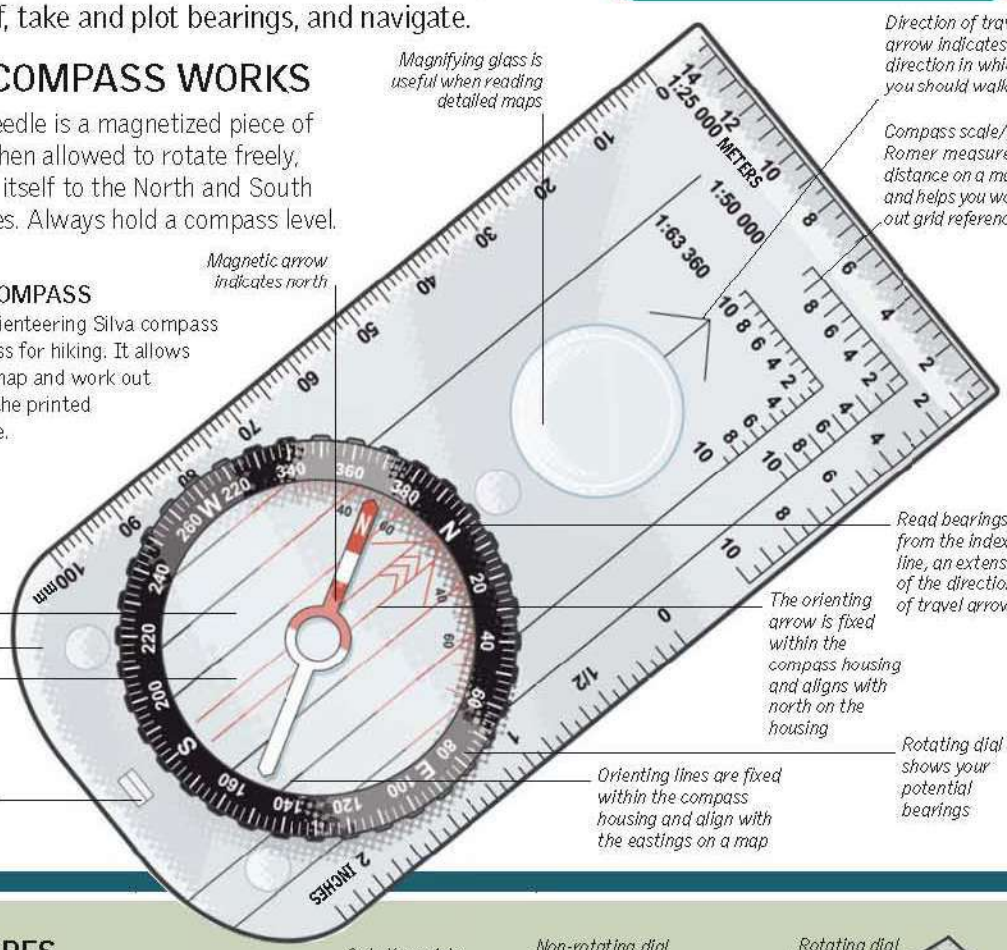

OTHER TYPES OF COMPASS

There are many different types of compass, ranging from simple button compasses to complex instruments that include features such as sighting mirrors. Use a quality compass that suits your needs, and carry a simpler backup for emergencies.

BUTTON

The simplest compass available—its small size and basic level of detail makes it ideal as a backup compass.

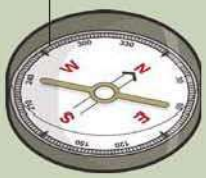

FIXED DIAL

As well as the basic points of the compass, degrees and bearings are also shown on the immovable dial.

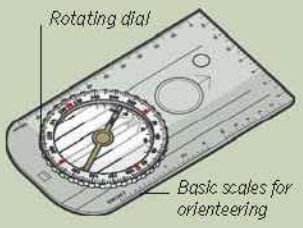

BASIC ORIENTEERING

With a rotating dial, this has simpler markings than the standard compass above, and is ideal for beginners.

SETTING YOUR MAP

Walking with your map set allows you to read the ground from the map as you pass over it, and to recognize and predict features as you progress, which means that you'll soon realize if you are heading off course. In some cases, when you have a good view of the terrain around you, and know the approximate area you're in, you can simply rotate your map until the features line up with the corresponding features on the ground. However, using your compass is a far more accurate method.

1 Rotate the dial so that "N" sits under the index line. Lay your map flat on the ground and ensure there is nothing nearby that could affect your compass (see box below).
- Lay the compass on the map so its edge runs parallel with a vertical grid line (easting).

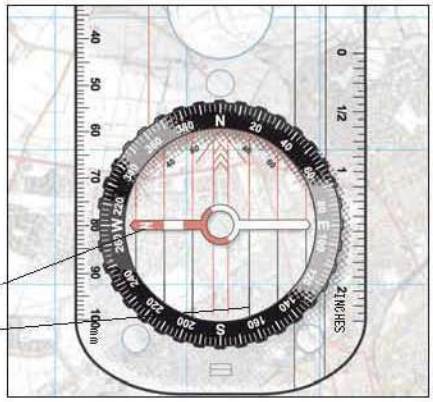

2 Keeping orienting lines aligned with the grid lines on the map, rotate the entire map until the north magnetic needle on the compass sits inside the orienting arrow. The map is now set to magnetic north and should basically line up with the features that surround you.
- If magnetic variation in your area is high (+ 5°), compensate accordingly (see right).

LENSATIC
The lensatic compass is an excellent instrument for taking very precise navigational measurements.

TAKING AN ACCURATE READING

Always hold the compass level to allow the needle to rotate easily. A compass is simply a magnetized piece of metal and, as such, is susceptible to interference. To avoid this, never use your compass near:
- Metal or other magnetized objects
- Electric currents, such as overhead high-tension power cables
- Buildings and vehicles—these often contain metal and electricity that might affect the accuracy of your compass readings

MAGNETIC VARIATION

Map legends refer to north in three ways: "true north," "grid north," and "magnetic north." The angle between magnetic north and grid north is known as "magnetic variation," and is provided in the map legend. True north is the direction of a meridian of longitude that converges on the North Pole. Grid north runs parallel to the vertical grid lines on a map and differs from true north because a map is flat. Magnetic north is the direction indicated by a magnetic compass.

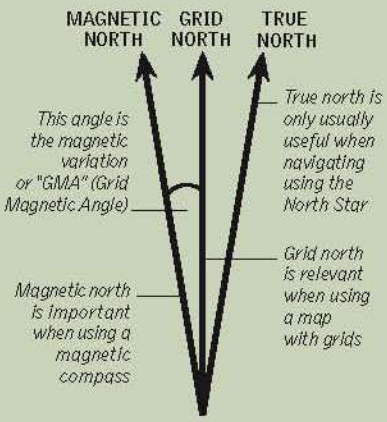

COMPENSATING FOR VARIATION

When converting a magnetic bearing to a grid bearing, or vice versa, you have to adjust for magnetic variation. When the variation is west, use the phrases "Mag to grid—get rid" or "Grid to mag—add." If the variation is east, the opposite applies.

1 Check the legend of your map to establish the magnetic variation. This depends on your location, and whether it is east or west of grid north.

2 If the variation is 0°, there's no magnetic variation affecting the compass, so make no adjustments.

3 If converting a magnetic bearing to a grid bearing with a 12° west variation, take off the 12°. With an east variation, add it on.

4 If converting a grid bearing to a magnetic bearing with a 12° west variation, add on the 12°. With an east variation, take it off.

TAKING BEARINGS

Always give any compass work your full attention. Rushing it, especially when working out bearings, can lead to navigational errors that could, at best, involve more walking and, at worst, get you lost.

USING A MAP

Using your map to work out the direction in which you need to walk is simple. Use your compass as a protractor to work out your bearing, and then to keep you on track.

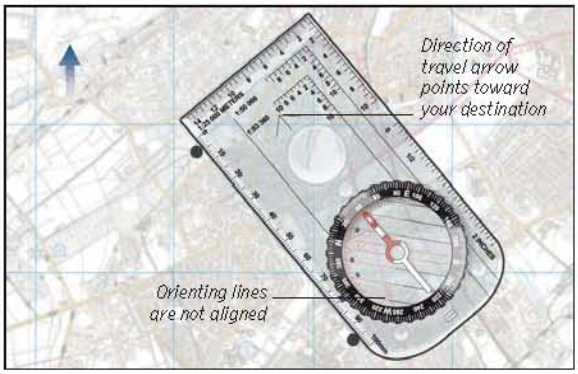

BACK BEARING

A back bearing enables you to find your position by taking a bearing from a feature back to you. To do this, take a bearing to a point in the normal way and either add or subtract 180 degrees. You can also read the bearing exactly opposite to the index line. This is useful when working out the bearing from a feature back to you and transferring it to your map (see below).

1 Lay your map on a flat surface, ensuring that there's nothing nearby to interfere with the compass reading (see box, p. 69).

- Lay the edge of your compass so that it runs between the point you want to navigate from and the point you want to navigate toward.
- Ensure that the direction of travel arrow on the compass is pointing in the direction you want to travel on the map.

USING FEATURES ON THE GROUND

Sometimes you may need to take a bearing to a specific point to navigate toward it. The point may be a feature that you can see at the time but may subsequently lose sight of during your journey because of the terrain. You can also plot that bearing and others on a map in order to work out your own position (see right, and panel, far right).

TAKING A BEARING ON A FEATURE

Pointing your compass at the feature, hold the base plate steady and rotate the compass housing until the orienting arrow sits directly under the north needle. Read the bearing—this is the magnetic bearing you would walk on to navigate to the feature.

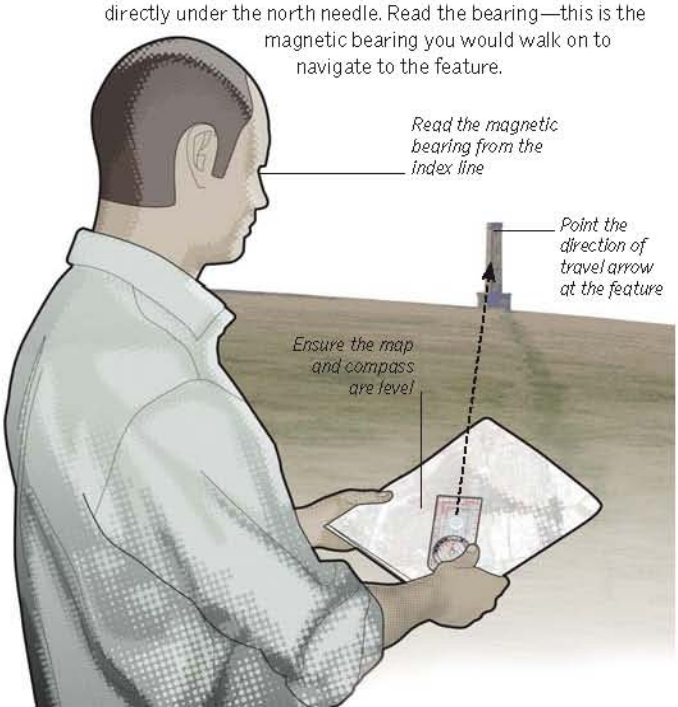

TRANSFERRING A MAGNETIC BEARING TO A GRID BEARING

It is important that you know how to transfer a compass bearing from a feature (a magnetic bearing) onto a map (a grid bearing). For this example, imagine that the magnetic variation is 12° west.

- Take a bearing on your chosen feature (see left). In this example it is 45° magnetic.
- You want to plot this bearing on your map, which has a grid, so remember the phrase "mag to grid—get rid." You would therefore subtract the magnetic variation (12°) from the magnetic bearing (45°), which equals 33°.
- Dial this revised bearing into your compass.

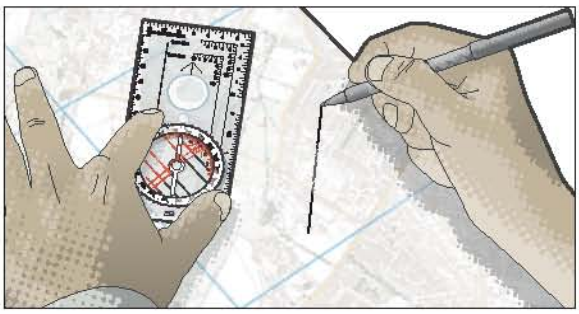

TRANSFERRING THE BEARING ONTO THE MAP

Place the top left corner of the compass base plate over the feature on your map. Keeping it there, rotate the entire compass until the orienting lines are parallel with the vertical grid lines. Draw a line from the feature down the left side of the base plate to map your bearing.

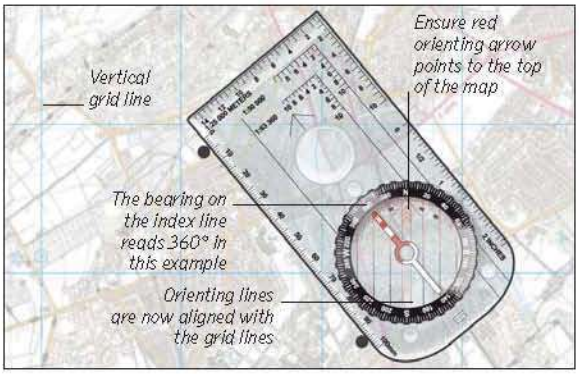

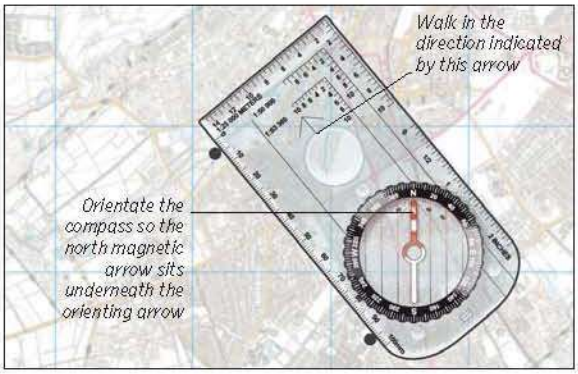

2 Turn the compass housing around until the orienting arrow and the orienting lines within the housing line up with the vertical grid lines on the map. Read the bearing between the two points from the index line on the compass.

- For this grid bearing to deliver you successfully to your destination, you'll need to convert it to a magnetic bearing by using the magnetic variation information on the map legend (see p. 69). Add or subtract your figure, and adjust your compass accordingly.

3 In order to walk on this bearing you must now orientate your compass.

- Hold the compass level and at a height that allows you to comfortably look squarely down on it (close to your chest is a good position).
- Turn your body until the north end of the compass needle sits inside the orienting arrow. The direction of travel arrow is now pointing exactly in the direction that you need to walk.

FINDING YOUR POSITION

If you are unsure of your position but can see features on the ground that you also recognize marked on your map, you can take bearings on these features with your compass and transfer them onto your map (see far left, and left) to accurately determine where you are. This process is called "triangulation" (also known as "resection" in military terminology). You'll need to transfer your compass bearings into grid bearings.

AT A KNOWN FEATURE

If you are at or on a known feature on a map—such as a river, road, or track—and you can see another recognizable feature, you can take a bearing on that feature and mark it on your map. Where that line crosses your known feature is your position. In the example below, you know you are somewhere along the banks of a river and can see a church that you can identify on your map.

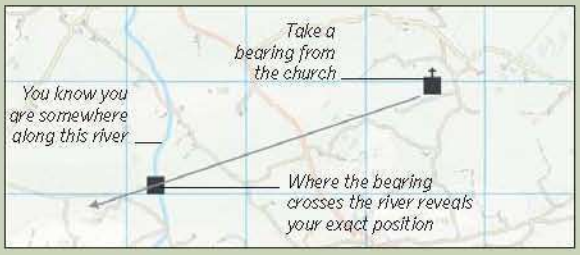

NOT AT A KNOWN FEATURE

If you are not at a known feature, but can see other recognizable features on the ground that you can identify and locate on your map, you can take bearings on these features and transfer them to your map. In order to do this you need two features that are at least 1 km (6/10 mile) away and at least 40° apart. After you have transferred both bearings onto your map, the point at which the two lines cross will reveal your exact location. For greater accuracy, plot a third feature onto your map. The lines will cross and form a triangle—your position will be inside this triangle.

1 Using your compass, take bearings to the features on the ground.

- Transfer these bearings to back bearings (see box, above left), and draw them on the map from your chosen features.

2 Extend these lines further over the map until they cross each other.

- The point at which the lines cross shows your location.
- If you want to be more accurate, repeat the process with a third feature.

ROUTE FINDING

WHETHER YOU ARE PLANNING to walk a specific route, or are in a survival situation and need to move to a safer area or to one that offers better chances of rescue, being able to study your map and calculate what you can achieve in a certain time can mean the difference between reaching your destination or spending a night in the wilderness.

ROUTE PLANNING

Break your route down into small sections. This will focus your navigation and make the overall distance seem less daunting. If you're in a group, designate an emergency rendezvous (ERV) point within each section—if someone gets separated you should all head there. If you can, include water sources and a safety point, such as a campsite, on your route.

CALCULATING DISTANCE

There are several methods of calculating the distance you are walking, and a seasoned hiker will always use at least two of them at any one time.

PACE COUNTING

Pace counting involves knowing how many paces you take to cover a set distance and then counting them as you travel. It is reasonably accurate. Distances are usually calculated in meters and kilometers, and most people take approximately 60 paces (120 steps) every 100 m. Try any of these methods:

- Cut 10 notches in your walking staff. Move an elastic band down a notch every 100 m.
- Use pace counting beads (a piece of webbing with two sets of beads divided by a central knot). One side contains nine beads to count off every 100 m, and the other has four beads to count off kilometers.
- Place 10 small pebbles in your pocket. Every time you cover 100 m, transfer a pebble to the other pocket.

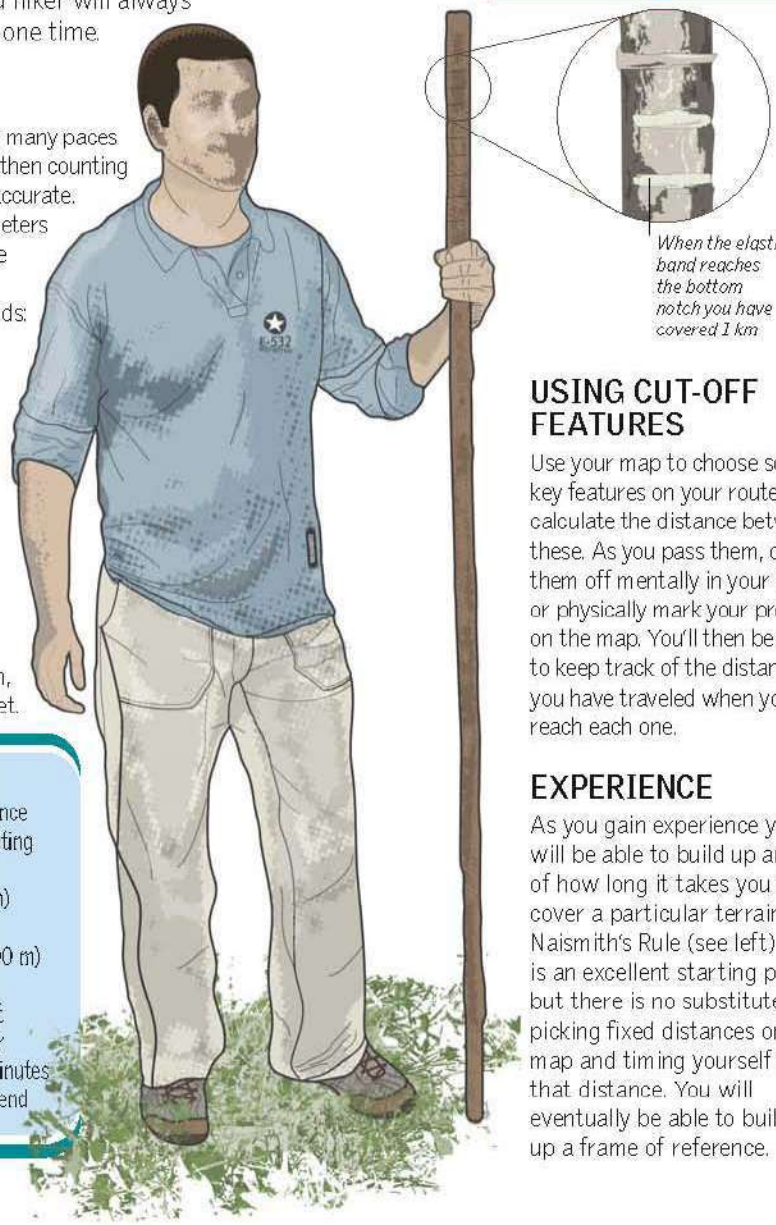

When the elastic band reaches the bottom notch you have covered 1 km

NAISMITH'S RULE

Naismaith's rule takes into account distance and topography, and is used for estimating the duration of hikes.

- Allow one hour for every 3 miles (5 km) you will travel
- Add 30 minutes for every 985 ft (300 m) you will climb
- Subtract 10 minutes for every 985 ft (300 m) you will descend. However, for very steep slopes you should add 10 minutes for every 985 ft (300 m) you will descend

USING CUT-OFF FEATURES

Use your map to choose some key features on your route, and calculate the distance between these. As you pass them, check them off mentally in your head or physically mark your progress on the map. You'll then be able to keep track of the distance you have traveled when you reach each one.

EXPERIENCE

As you gain experience you will be able to build up an idea of how long it takes you to cover a particular terrain. Naismith's Rule (see left) is an excellent starting point, but there is no substitute for picking fixed distances on your map and timing yourself over that distance. You will eventually be able to build up a frame of reference.

NAVIGATION TECHNIQUES

When navigating across land, you're less likely to get lost if you take direct bearings from one feature to another. Unfortunately this may not always be possible; obstacles such as lakes and swamps may be directly in your path and it may be easier to walk around some features rather than walk over or through them.

PRINCIPLES OF NAVIGATION

You will never get lost if you know:

- Where you started from
- What bearing or course you have been traveling on
- How far you have traveled

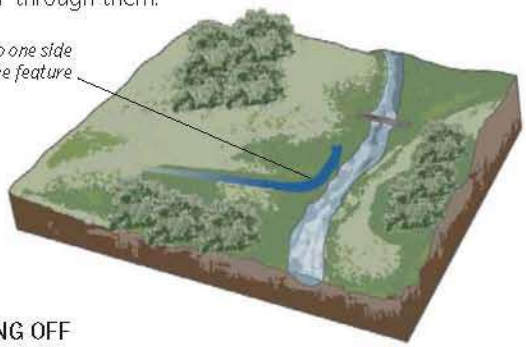

AIMING OFF

It's easy to find yourself slightly off-course after a while. If you were aiming for a small footbridge over a stream and didn't arrive exactly at the bridge, you would need to guess whether to turn left or right to reach it. By deliberately aiming off to one side (also called "deliberate off-set") you can guarantee this direction.

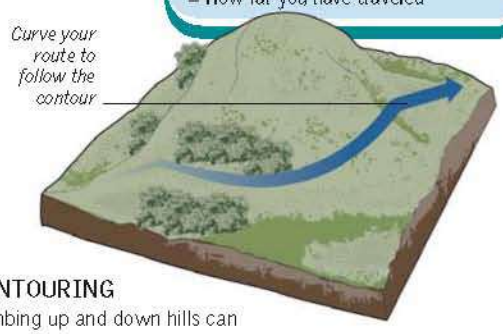

CONTOURING

Climbing up and down hills can expend a lot of energy and may not be the most effective method of navigating a particular terrain. Instead, use a technique called "contour navigation." This involves walking at the same height around a feature, which will conserve energy.

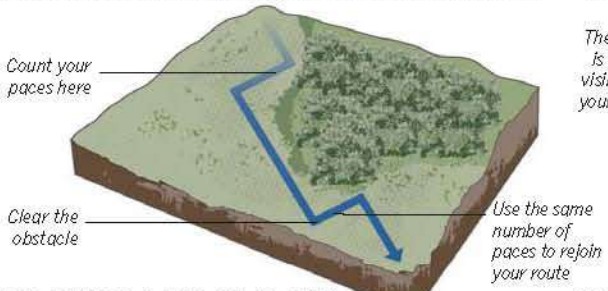

DETOURING (BOXING AN OBJECT)

If an obstacle makes a straight-line bearing impossible, use your compass to calculate four 90° turns, which you will then need to walk on to pass the obstacle. Count your paces on the first and third detours to return to your original, intended route as accurately as possible.

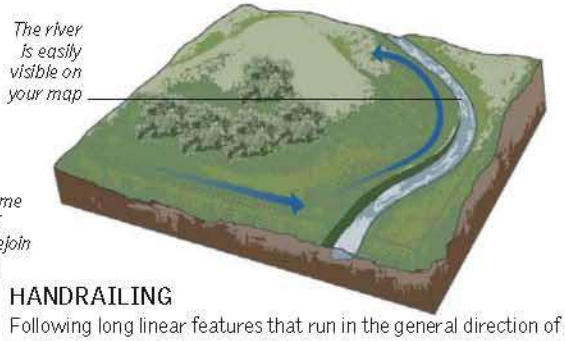

HANDRAILING

Following long linear features that run in the general direction of your travel—such as rivers, roads, or paths—can be an effective way of navigating. Because you use the features rather than your compass, navigation becomes simpler, as the features are usually easy to follow. A slight detour toward one can be worthwhile.

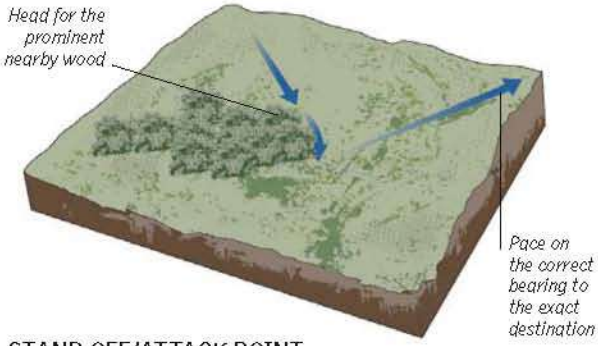

STAND OFF/ATTACK POINT

Useful when navigating to a specific point that may be difficult to locate, this technique involves aiming initially for a nearby prominent feature. Calculate a distance and bearing from it and use pacing to accurately locate the specific point. If you fail to find it, simply return to the prominent feature and start again.

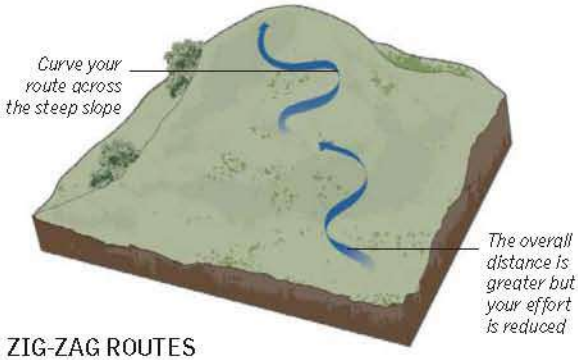

ZIG-ZAG ROUTES

You may have no alternative but to climb a steep slope and this can be exhausting. However, if you choose a zig-zag route up the slope you can reduce the effort required to achieve the climb. This will dramatically reduce the strain on your legs, ankles, lungs, and heart. It is also effective when walking down steep slopes.

NAVIGATING WITHOUT A COMPASS

IF YOU DON'T HAVE A COMPASS you can use a few simple items from your survival kit to determine direction and enable you to navigate reasonably accurately. Advances in technology have also made satellite-navigation aids more readily available.

TOOLS AND MATERIALS

You will need an improvised needle or razor blade, plus one of the following:

- Knife
- Hair
- Silk
- Magnet
- Battery and wire; paper and tape

IMPROVISING A COMPASS

It's relatively simple to make an improvised compass using a piece of magnetized ferrous metal. How accurate your compass is depends on the materials you have available to you and your own ingenuity.

SOURCING AND MAGNETIZING THE NEEDLE

If you have a compass but it's damaged, you may still be able to use the needle from it, which will already be magnetized. If this is unusable, you'll need to find a piece of ferrous metal to magnetize. Suitable items you could use include:

- A needle from your survival tin or sewing kit
- A paperclip, opened up and straightened out
- A razor blade from your survival kit
- A small nail or straightened staple taken from a fence

Once you have your improvised needle, you will need to magnetize it. The smaller and thinner the needle is, the easier this will be. Use one of the methods shown here.

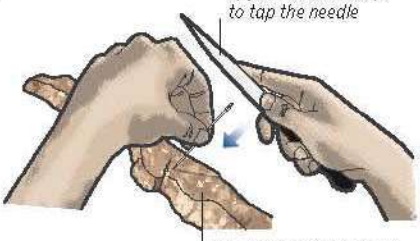

THE TAPPING METHOD

Align the needle as close to the north-south line as you can determine, hold the needle at an angle of 45 degrees, and gently tap the end of it with another piece of metal. Lightly tapping it into a hard piece of wood will increase the effect.

USING YOUR HAIR

Hold the sharp end of the needle perpendicular to your head and—taking care not to hurt yourself—stroke the needle in one direction through your hair, using careful and deliberate strokes. Repeat until the needle is magnetized.

ALLOWING THE NEEDLE TO FLOAT FREELY

Once you have made your improvised needle using one of the methods above, you need to find a way to allow it to turn freely so that it is able to indicate direction. Take care to protect your needle from elements such as the wind, which will affect its movement.

SUSPENSION METHOD

The advantage of the suspension method is that the equipment is portable and can be reused. It works best with a magnetized razor blade, which will balance well. Attach the magnetized blade to a cotton thread and suspend it inside a plastic bottle. If the bottle's neck is not wide enough to fit the blade through, remove the base of the bottle instead.

FLOATING METHOD

In a sheltered place, float the needle on the surface of some water—a puddle or a small non-magnetic container filled with water, for example. Balance the needle on a small, dry leaf (or piece of paper, piece of bark, blade of grass, or inside a shortened straw).

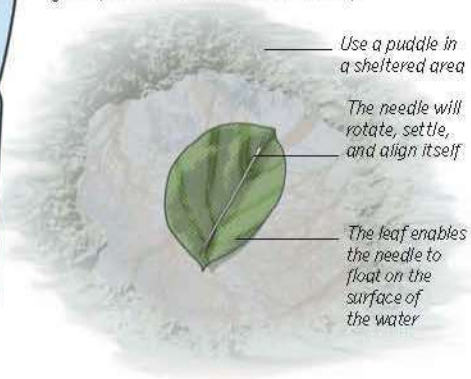

PRINCIPLES OF MAGNETIZING

In general, the longer you work on the needle the stronger the magnetization will be and the longer it will last. To tell when it is magnetized sufficiently, hold the needle up against another metal object—if it is attracted to the metal and has enough strength to hold itself against it then it will be strong enough. Once the needle is magnetized you will need to allow it to float freely (see below), and then determine which end points north by using natural aids, such as the sun (see pp. 76–77). Mark the north end of the needle with a pen or a small scratch.

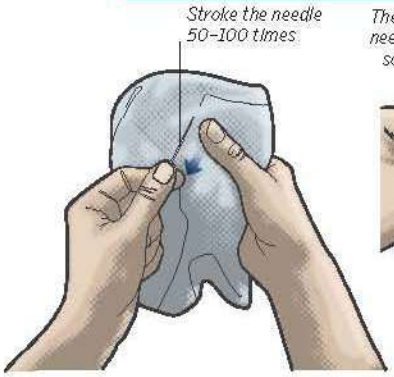

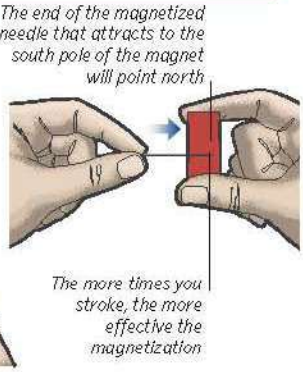

USING SILK

This works on a similar principle to the hair method but is more effective. If you have anything made from silk, such as a sleeping bag liner or thermal clothing, stroke the needle repeatedly in one direction against the silk.

USING A MAGNET

Stroke the magnet along the length of the needle repeatedly in one direction. It's a good idea to carry a magnet with you at all times, although you should never keep it near your compass as it will affect its accuracy.

USING ELECTRICITY

The most effective way to magnetize a needle is to pass a small electrical current around it. Use a battery and insulated wire; alternatively, use brass snare wire from your survival kit and insulate it using something non-conductive, such as paper.

1. Wrap the needle in a small piece of paper, which will insulate it from the electrical current.

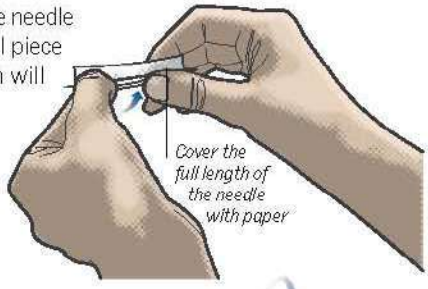

2. Wrap the wire tightly around the full length of the insulated needle.

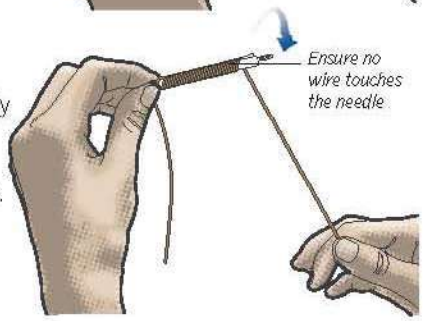

3. Attach the wire to the battery until battery starts to get warm—this indicates the process is complete.

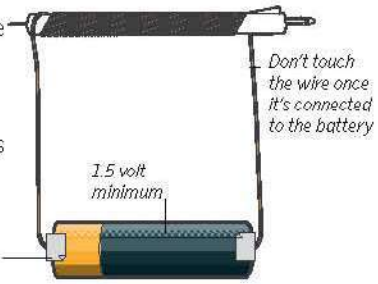

USING GPS TECHNOLOGY

A Global Positioning System (GPS) is a hand-held unit that uses 24 orbiting satellites to triangulate your position on the Earth's surface to within meters. A GPS will allow you to work out straight-line distances and bearings to and from points, but unless it incorporates mapping it will not show you the best way to get there or take into account hazards on your route. Use it in conjunction with your map and compass.

TAKING EFFECTIVE GPS READINGS

A GPS needs to have a clear view of the sky. Anything that obstructs its signal, such as tall buildings or heavy tree canopy, will reduce its ability to lock onto satellites.

NATURAL NAVIGATION

IF YOU HAVE LOST or damaged your compass and don't have the materials necessary to make an improvised version, use natural indicators to determine direction. The east-west rotation of the Earth means that you can orientate yourself according to the position of the sun, moon, and stars. All you need are some very basic materials and these simple techniques.

USING THE SUN

When visible, the sun is the clearest natural signpost to the four cardinal points (north, south, east, and west). It rises approximately in the east and sets approximately in the west and, at midday, is due south in the northern hemisphere and due north in the southern hemisphere. Use the sun's course across the sky to determine direction and approximate the time.

SHADOW STICK BASICS

Use a shadow stick to determine direction and time anywhere between the Arctic (66.5°N) and Antarctic Circles (66.5°S).

- In the northern hemisphere the shadow of the stick will be on the north side of the east-west line
- In the southern hemisphere the reverse is true: the shadow of the stick will be on the south side of the east-west line.
- When the shadow is at its shortest, it is midday.

ORIENTATION

Tracking the movement of the sun across the sky using a shadow stick will provide an indication of its direction of travel. The sun moves from east to west at 15° an hour.

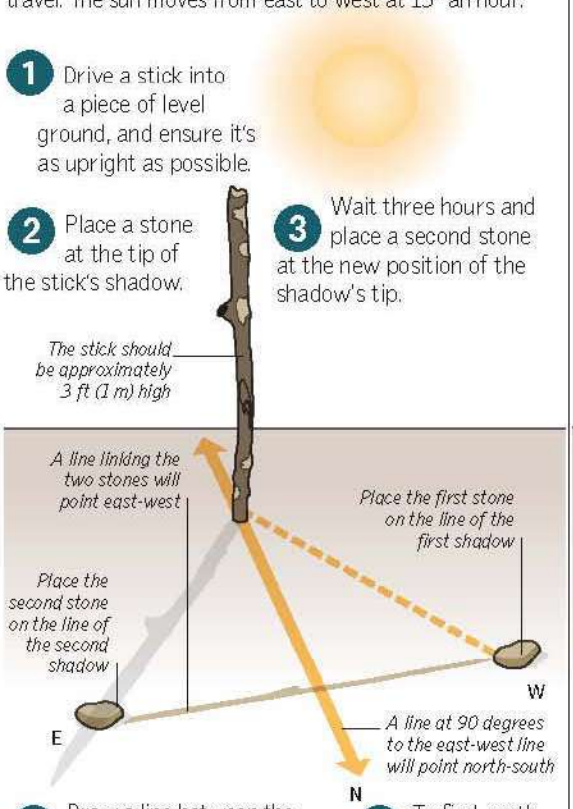

1. Drive a stick into a piece of level ground, and ensure it's as upright as possible.
2. Place a stone at the tip of the stick's shadow.
3. Wait three hours and place a second stone at the new position of the shadow's tip.
4. Draw a line between the stones to find east-west; the first stone will be at due west and the second at due east.
5. To find north-south, mark a line at 90° to the east-west line.

TELLING THE TIME

Once you have established the east-west and north-south lines, you can turn the shadow-stick apparatus into a sundial, to give an approximate idea of the time.

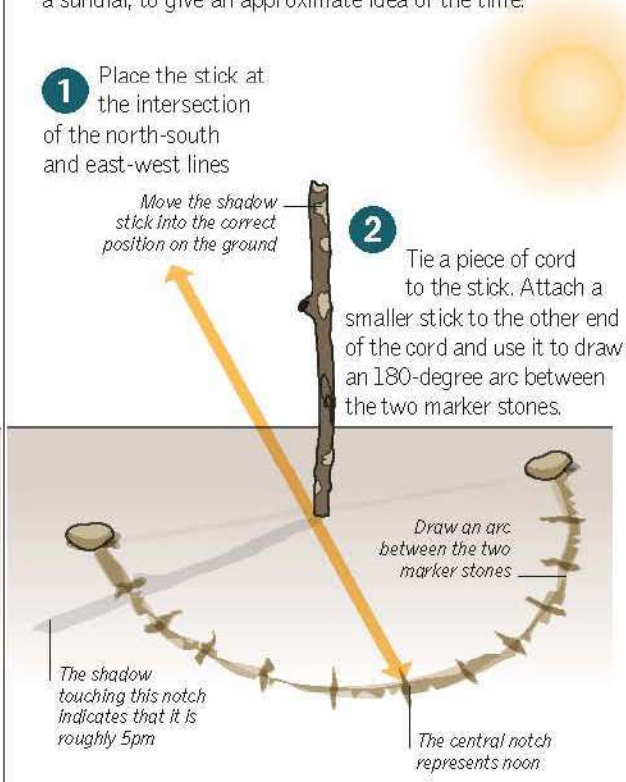

1. Place the stick at the intersection of the north-south and east-west lines
2. Tie a piece of cord to the stick. Attach a smaller stick to the other end of the cord and use it to draw an 180-degree arc between the two marker stones.
3. Divide the arc into 12 equal sections and mark each division with a notch. The notches represent one hour of time, from 6am to 6pm with noon at the middle notch.

USING THE STARS

It's possible to orientate yourself by spotting certain recognizable stars. In the northern hemisphere, Polaris is located above due north on the horizon, from which you can determine east, west, and south. Find it by locating the Plow (also known as the Big Dipper). In the southern hemisphere, you can work out the approximate position of south on the horizon by finding the Southern Cross.

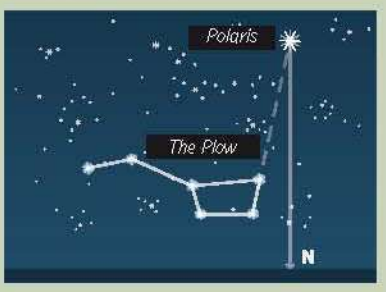

NORTHERN HEMISPHERE

Find the Plow and project a line from the top of the two stars that form its front. Follow this line until you find Polaris, which is located approximately four times the distance from the Plow as the distance between its two front stars.

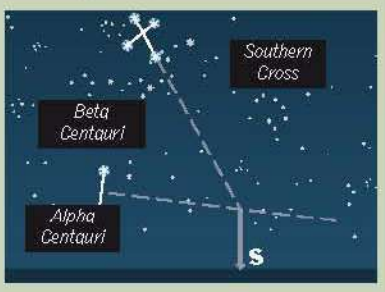

SOUTHERN HEMISPHERE

Project a line from the longer axis of the Southern Cross until you find a dark area of sky. Project a second line at 90° from the mid-point between two bright stars in the Centaurus constellation. Due south is below the point where these lines meet.

USING THE MOON

Reflecting the light of the sun, the moon rises in the east and sets in the west, so can be used for orientation. A shadow stick will work on a cloudless, full-moon night.

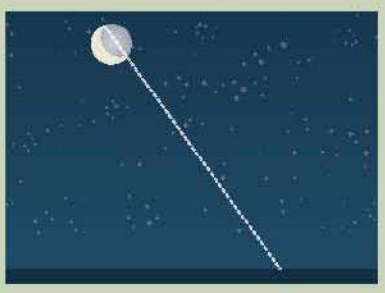

CRESCENT MOON

Although not an entirely accurate method, a line vectored between two horns of a crescent (quarter) moon will lead to a point that is approximately south on the horizon in the northern hemisphere, and roughly north in the southern hemisphere.

USING AN ANALOG WATCH

If you can see the sun, you can use an analog watch as a protractor to determine an approximate direction. Ensure it is set to the correct local time and that you have taken daylight savings (DST) into account. If you don't have a watch but know the time, simply draw a watch face on a piece of paper, marking 12 o'clock and the hour hand. This method is increasingly less effective as you near the Equator.

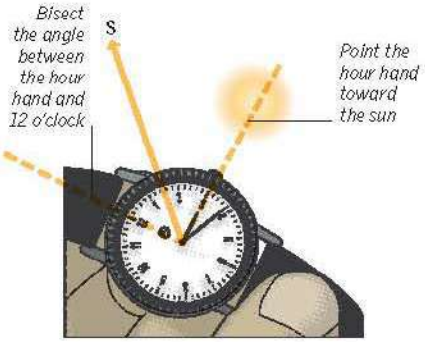

NORTHERN HEMISPHERE

In the northern hemisphere the cardinal point nearest to the sun is south. Point the hour hand of your watch at the sun and bisect the angle between the hour hand and 12 o'clock. This will be due south.

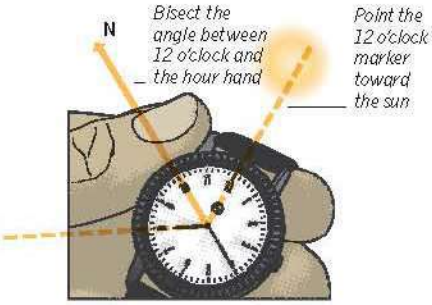

SOUTHERN HEMISPHERE

In the southern hemisphere, the cardinal point nearest to the sun is north. Point the 12 o'clock marker on your watch at the sun and bisect the angle between 12 o'clock and the hour hand on the watch. This will be due north.

NATURAL SIGNPOSTS

Nature responds to the elements in a range of ways, some of which can be studied for orientation tips. This is useful if you know the predominant wind direction of an area.

TREES AND PLANTS

- Windswept trees point away from the wind.
- Tree growth is most lush on the side that faces the sun (south in the northern hemisphere, north in the southern hemisphere).
- Some plants, such as the barrel cacti, twist toward the sun as they grow.
- Moss and lichen grow out of direct sunlight, on the shady side of rocks or trees.

ANIMALS AND INSECTS

- In very windy areas, small animals and birds tend to nest or burrow on the lee-side of hills.
- Spiders spin webs out of the wind, so an area with a lot of broken webs could indicate a recent change of wind direction, away from the dominant wind.

SNOW AND ICE

- In powder-snow conditions, snow "dunes" often form parallel to the prevailing wind.
- Frost erosion is most severe on slopes facing the sun (south in the northern hemisphere, north in the southern hemisphere).

HOW WEATHER WORKS

CHANGES IN THE WEATHER can have a significant effect on your expedition—or chances of survival if something goes wrong—so it's important to be as prepared as possible for all the conditions you might face. Check the weather report for the days ahead thoroughly, and pack and wear appropriate clothing. If the weather forecast is so bad that it could make traveling or navigation difficult, rethink your plans.

UNDERSTANDING THE WEATHER

Weather is created by the movement of air currents, the moisture content of the air, and the meeting of warm and cold fronts. Looking at weather maps before you leave for your trip will help you track these elements. You'll then need to use your understanding to forecast the effects of these movements on the ground.

WHAT IS WEATHER?

The term "weather" relates to current conditions on the ground—such as the temperature, and whether or not it is windy or raining—while "climate" refers to a region's conditions over a longer period of time. Changes in the weather are caused by alterations in air pressure and temperature. Extreme weather, such as a hurricane, occurs when these changes are more marked than usual. Weather is monitored by meteorologists, and very accurate predictions can be made. However, weather can still cause surprises, so always prepare for all possible scenarios when packing or dressing for an expedition.

READING WEATHER CHARTS

It can be helpful to compare weather charts with regular topographic land maps. Just as the gradient is steeper where contour lines appear closer together on land maps, the wind is stronger where isobar contours appear closer together on weather charts.

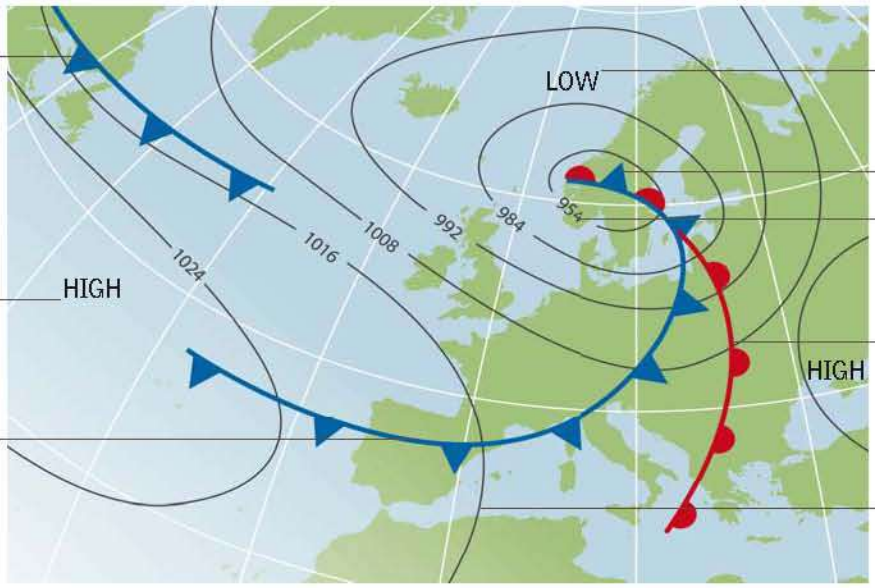

Fronts move in the direction that their symbols point toward

Areas of high pressure are generally associated with warm weather

A blue line with triangles represents a cold front, and marks the leading edge of cold air

Areas of low pressure are associated with cooler weather

Occluded fronts are fronts where two air masses merge

The lower the value of the central pressure, the more severe the wind and rain

A red line with semi-circles represents a warm front, and marks the leading edge of warm air

Isobars join areas where the air pressure is the same

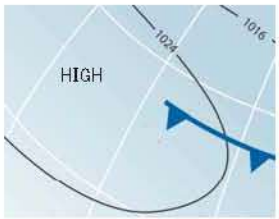

HIGH PRESSURE

Under conditions of high pressure, air descends in a spiral formation, and warms. The water vapor does not condense into clouds and we would expect to observe fairly clear skies.

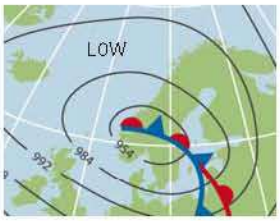

LOW PRESSURE

Under conditions of low pressure, air spirals upward—parallel to the center of the isobars—and cools. The air condenses to form clouds, and skies are usually overcast.

GLOBAL CIRCULATION

Warm air from the equator rises and moves toward the poles in each hemisphere. As it cools it descends and travels back toward the equator. Because of the Earth's rotation, three separate cells of circulating air in each hemisphere develop, which produce predictable wind and pressure patterns.

AIR MASSES

Large bodies of air with a particular temperature and humidity help weather forecasters predict the forthcoming weather and are known as air masses. They are largely defined by the area that they originated from. In general, northerly winds are colder than southerly winds, and air that has tracked over the sea ("maritime air mass") accumulates moisture and is more cloudy than air that has tracked over the land ("continental air mass"). The boundaries between tropical and polar air become warm and cold fronts.

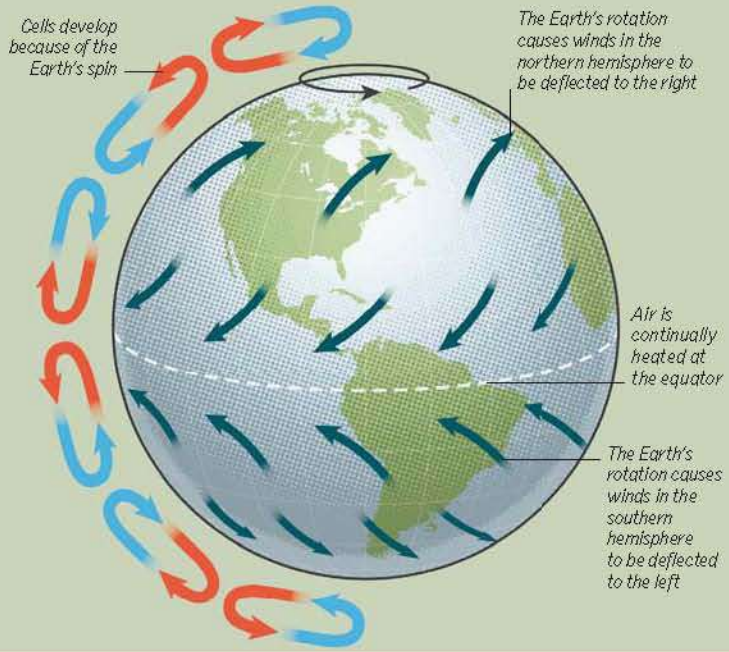

ON THE GROUND

As a weather system travels over the ground, it changes in predictable ways. Knowing how the system will progress is an important part of interpreting your weather map or forecast.

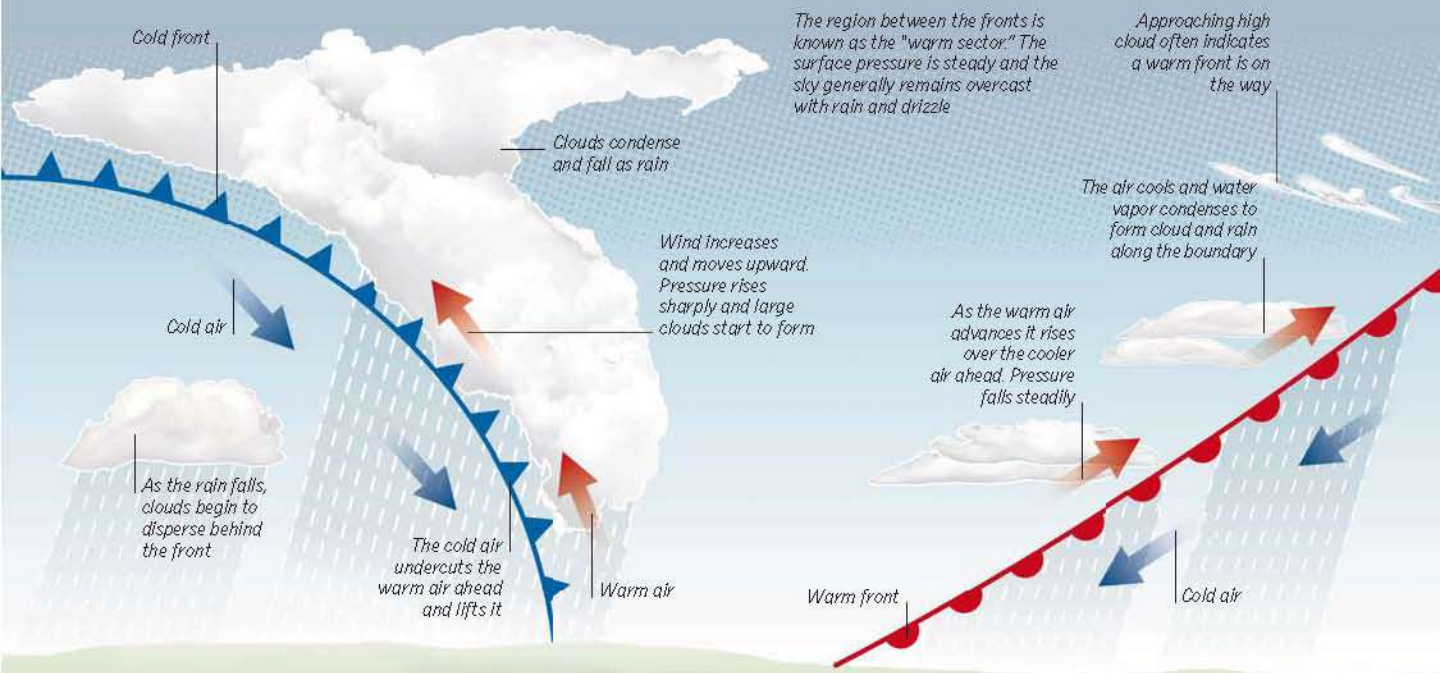

WEATHER PHENOMENA

LOOKING AT THE CLOUDS can help you to read approaching weather, which is invaluable when you're on the move with no access to weather forecasts or charts. Knowing how to recognize a storm cloud, for example, will ensure that you have enough warning to seek shelter or to change into appropriate clothing. If low cloud threatens to impair your visibility, use your compass to navigate and proceed with caution.

HOW CLOUDS FORM

Clouds form by a cooling process. This process can be compared to breathing on cold glass, when condensation appears because invisible water vapor in your breath, cooled to a temperature known as the "dew point," condenses into its visible liquid state. Clouds form in a very similar way. Air temperature falls at an average rate of two degrees celcius per 1,000 ft (300 m). When air containing moisture—perhaps because it has travelled over sea—reaches the height at which it meets its dew point temperature, the water vapor condenses to form visible cloud.

READING THE CLOUDS

Clouds are condensing masses of water vapor that provide precipitation and reflect solar radiation. They are categorized by height—into low, medium, and high clouds—and further defined by their shape. The shape of a cloud is determined by the way in which warm air rises, and is an indicator of the air stability.

HIGH CLOUDS

Fair-weather clouds are white and high in the sky. If there are no clouds, expect excellent conditions. Blankets of high cloud progressively invading the sky can signify the onset of bad weather. High clouds include:

- Anvil of cumulonimbus
- Cirrus
- Cirrustratus

MEDIUM CLOUDS

Thick layers of medium-layer cloud give heavy, persistent rain, especially if the clouds are dark and gray. The clouds you are most likely to see at this level are:

- Altocumulus
- Nimbostratus
- Altostratus

LOW CLOUDS

Low clouds have clearly defined edges and can indicate whether rain will fall in short downpours (cumulus) or persistently (stratus). Common clouds are:

- Cumulonimbus
- Stratus
- Stratocumulus
- Cumulus

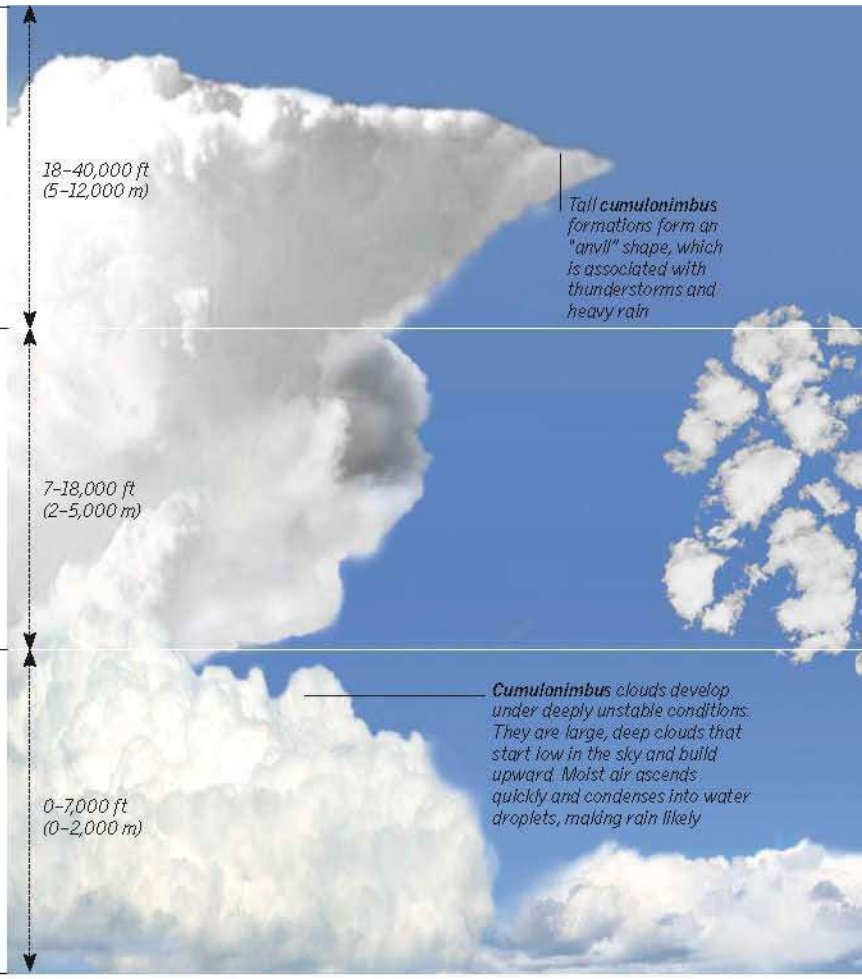

WEATHER RISKS

The weather can have a massive impact on your trip, bringing with it added dangers. It can affect your visibility levels, and the safety of the ground you're hiking over. In a survival situation, the temperature, and whether or not it's raining, can have a huge effect on your chances.

WARNING!
Never shelter under a solitary tree during a lightning storm—get in your vehicle if possible or find low ground.

HEAVY RAIN
If heavy rain falls quickly, the ground may become saturated, slippery, or flood. Seek shelter or proceed cautiously in waterproof clothing.

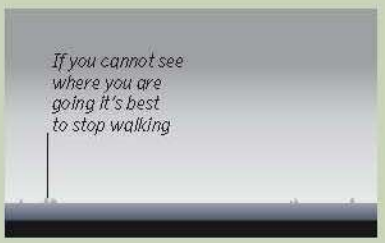

FOG
Essentially a cloud in contact with the ground, fog reduces visibility, so beware of dangerous terrain, especially on mountains.

LIGHTNING
Lightning is an atmospheric discharge of electricity—it will strike the first object it encounters on its route to earth, so avoid high, exposed places.

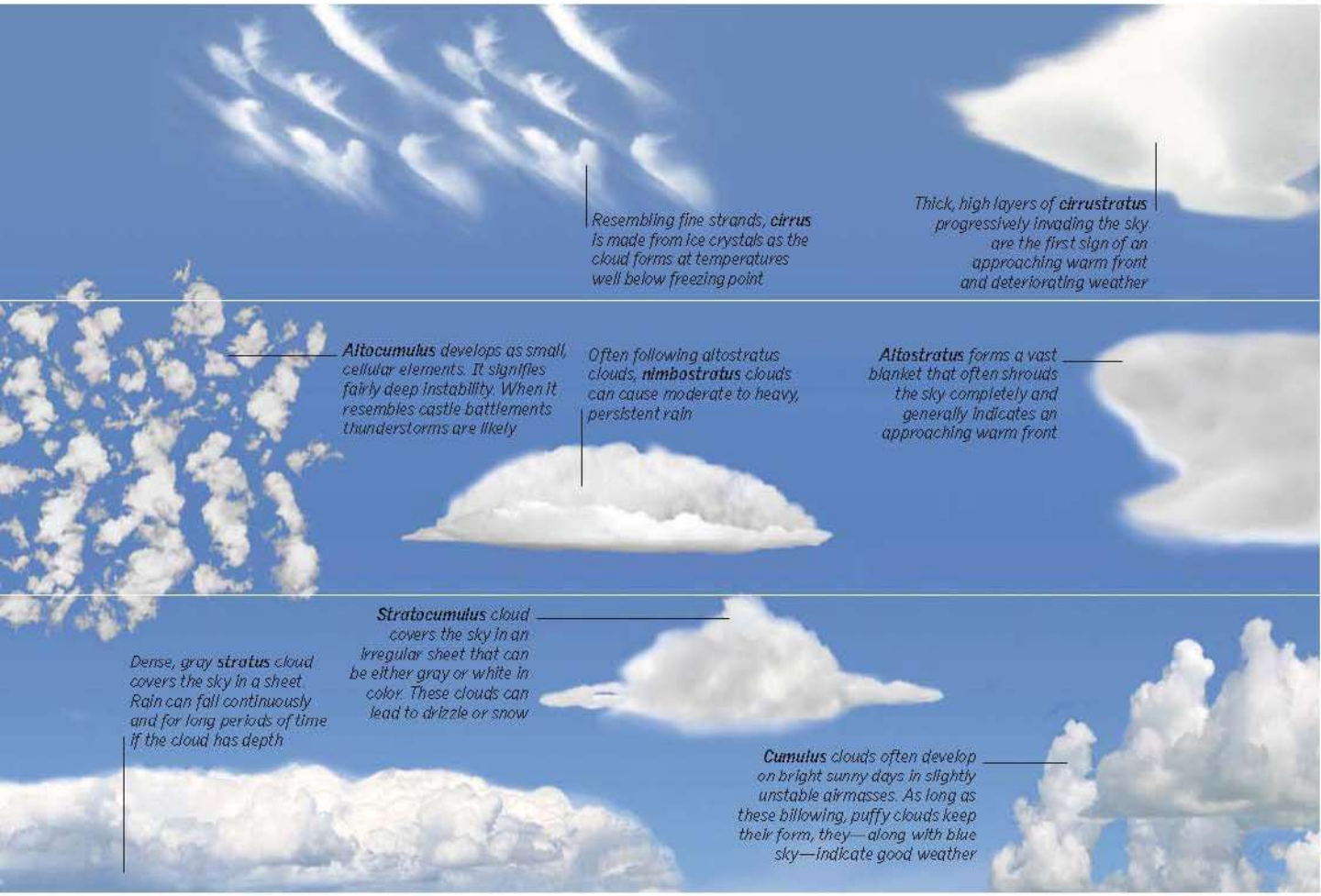

UNDERSTANDING LOCAL WEATHER

WEATHER IS AFFECTED—and can be predicted—by the local geography surrounding you. High ground forces air to rise and cool, and the relative temperatures of land and sea produce predictable effects.

REGIONAL EFFECTS

It helps to understand how these predictable weather patterns occur in various geographic areas. They may influence your decisions when on the trail.

THE FOHN EFFECT

The leeward side of high ground is warmer and more sheltered than the windward side. Air rises as it travels over obstacles and, if the air contains moisture, the water vapor condenses to form cloud once cooled to a certain level. The air loses water at the summit, then descends and warms on the leeward side.

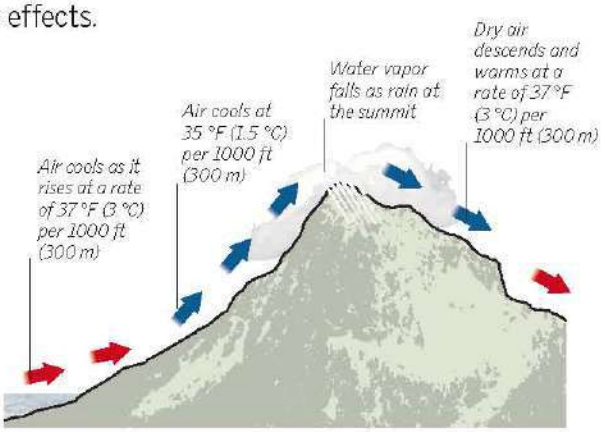

ANABATIC AND KATABATIC WINDS

Winds that flow up and down slopes during certain atmospheric conditions are known as "anabatic" and "katabatic." They typically occur in mountainous regions.

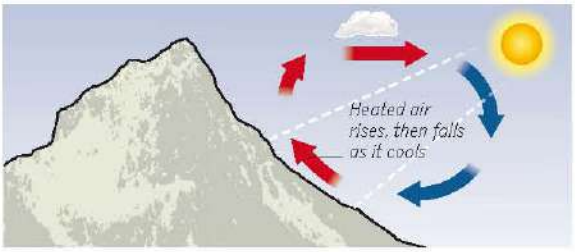

ANABATIC WINDS

During the day, the surface of sloping terrain heats up. The air rises and creates a gentle upslope breeze. Anabatic winds are lighter than katabatic winds because they act against gravity.

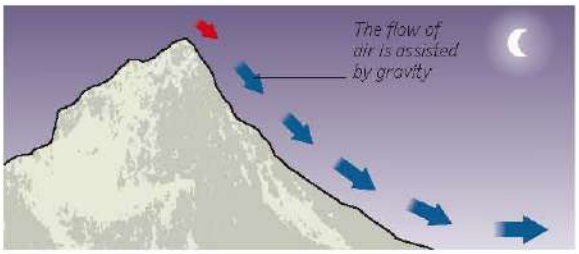

KATABATIC WINDS

Katabatic winds form on clear nights with a light breeze. Air in contact with the ground cools and its density increases, causing it to flow down the slopes of the hillside.

SEA AND LAND BREEZES

Sea breezes often develop along coastlines on clear sunny days, whereas land breezes tend to develop along coastlines on cloudless nights.

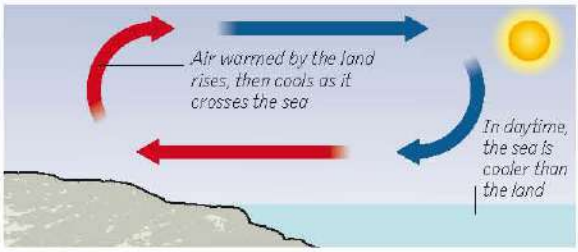

SEA BREEZE

Warm air rises over land during the afternoon, drawing in cool air to replace it along the coast. The result of this coastal air circulation by day is a wind that blows from sea to land.

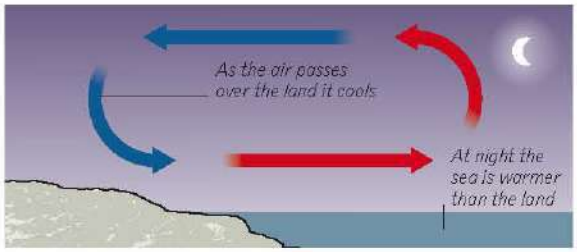

LAND BREEZE

At night, the sea becomes comparatively warmer and air begins to rise. Air from the neighboring land is drawn toward the sea. The result is a wind that blows from land to sea.

NATURAL WEATHER FORECASTERS

These natural indicators are based on observation and can be useful if you have no alternative means of predicting the weather.

LOOKING AT THE SKY

- If the sky is red at dawn there is moisture in the air and a potential storm ahead. A red sky at night often indicates good weather to come.
- A rainbow usually indicates good weather is on the way, or a light shower.

PLANTS AND FLOWERS

- The scent of plants and flowers is often stronger before rain.
- Certain flowers, such as Scarlet Pimpernel and Morning Glory, are said to close up if bad weather is on the way.
- Pine cones are one of the best natural forecasters—their scales absorb moisture in the air and close up if wet weather is approaching, and unfurl in dry air.

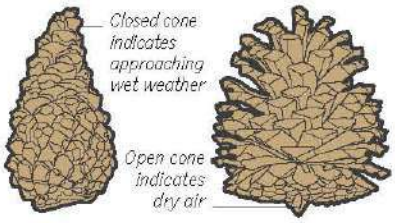

WATCHING ANIMALS

- Animals migrating from high to low areas may indicate that a storm is on the way.
- Cows often lie down before it rains.
- Wool reacts to moisture in the air by swelling and straightening. It shrivels in dry weather conditions.
- Humans can also sense atmospheric change—some people suffer headaches before a thunderstorm.

CHANGING WEATHER

Watch out for alterations in the wind direction or strength as this may lead to a change in the weather. A dry, steady wind that changes direction or decreases in strength often indicates that rain is on its way.

THE BEAUFORT SCALE

Providing visual references for the effects of wind speed on land and at sea, the Beaufort Scale was designed by Sir Francis Beaufort in 1805. It provides a simple way of estimating wind speeds without the need for equipment, and is still in common usage today. The scale ranges from calm to hurricane, and is numbered from zero to 12.

BEAUFORT NUMBER	WIND DESCRIPTION	WIND SPEED MPH (KPH)	WIND EFFECT ON LAND AND AT SEA
0	Calm	0 (0)	Smoke rises vertically. Sea like a mirror.
1	Light air	1–2 (1–3)	Smoke drifts gently. Sea ripples like scales.
2	Light breeze	3–7 (4–11)	Leaves rustle. Wind felt on skin. Small wavelets.
3	Gentle breeze	8–12 (12–19)	Leaves and twigs move. Large wavelets with scattered whitecaps.
4	Moderate breeze	13–18 (20–29)	Small branches move. Small waves with frequent whitecaps.
5	Fresh wind	19–24 (30–39)	Small trees begin to sway. Moderate waves with some foam and spray.
6	Strong wind	25–31 (40–50)	Umbrella usage is difficult. Large waves of 10 ft (3 m) with some spray.
7	Near gale	32–38 (51–61)	Whole trees sway. Sea heaps up and foam begins to streak.
8	Gale	39–46 (62–74)	Walking is difficult. Moderately high waves of more than 18 ft (5 m).
9	Severe gale	47–54 (75–87)	Damage to roofs. High waves with toppling crests.
10	Storm	55–63 (88–101)	Trees uprooted or broken off. Very high waves and sea surface appears white.
11	Severe storm	64–74 (102–119)	Houses damaged. Exceptionally high waves of more than 38 ft (11 m).
12	Hurricane	More than 75 (More than 120)	Buildings destroyed. Huge waves of more than 46 ft (14 m).

MAKE A MOVE

While some survival situations are simply unavoidable, many are entered into because basic techniques relevant to a particular mode of transport have not been followed correctly. This could be due to a lack of knowledge, a loss of concentration, an individual's over-confident belief in his or her own equipment or skills, or as a result of sheer recklessness. Therefore, before you venture into an unfamiliar environment, make sure that you carefully research the type of terrain you will be encountering and investigate the best method of traveling safely and efficiently across it. Knowing the correct techniques for scrambling up a mountain or walking down a steep hillside, for example, could mean the difference between an enjoyable day out and a life-threatening survival situation arising from a fractured ankle in a remote location. Equally, being able to regain control of your vehicle after it has entered a skid could quite simply save your life. Thoroughly researching the terrain will also allow you to select the correct

In this section YOU WILL DISCOVER...

- how to make a **paddle** (so you're never **up a creek** without one)...
- the difference between your **finger shelf** and your **hand jam...**
- how to **pull** your own **pulk** and **scramble** over **scree...**
- when to **blaze a trail** or **ski without skis...**
- how to tell your **front-wheel skid** from your **skidoo...**
- when to let the **camel train** take the **strain...**
- how **chewing gum** can stop that **sinking feeling...**

equipment and most appropriate clothing, and enable you to familiarize yourself with the survival techniques relevant to a particular environment.

Whether you're planning to travel on foot, by boat, on horseback, in a 4WD vehicle, or by boat, you must consider not only your own capabilities and those of other members of your team, but also the capabilities of your equipment. Remember, pushing anything beyond its limits—whether that be a person, an animal, or a vehicle—will invariably lead to it failing.

A walking staff is one of the simplest yet most important survival aids you will ever need. It's the first piece of equipment to improvise should you find yourself in a survival situation.

THE "SURVIVOR'S THIRD LEG"—as a walking staff is also known—increases your ability to support yourself by allowing you to have two points of contact with the ground at any one time. This will reduce the chance of you slipping—a crucial factor given that your ability to walk is your main means of rescue; reduce your mobility and you seriously reduce your ability to survive.

A versatile tool, your survivor's third leg can be employed in many different survival situations. Among its many uses, it can be employed to:

- Support you as you walk
- Protect your face when you're walking through thickets or gorse
- Check for adequate support when crossing marshy ground
- Test the ground ahead for obstructions
- Check the depth of water when you're crossing streams and rivers
- Protect you against wild animals
- Form a ridgepole for your shelter
- Help you with your pace counting
- Spear fish or catch game
- Dig up roots or plants

"Your ability to walk may be your **primary means of rescue**; reduce your mobility and you dramatically reduce your **chances of survival**"

TRAVELING ON FOOT

HIKING IS A GREAT WAY to explore wilderness terrain. It's important to have a decent level of fitness before you set out, and to wear and carry appropriate clothing and equipment. Hiking requires basic skills—using the correct techniques will help you move more efficiently, and ensure that your trip is safe.

BASIC WALKING SKILLS

Aim for a slow, even pace that can be maintained for the duration of the hike by all members of the group. A good way to maintain this pace is to develop a hiking rhythm. Take regular rest breaks and, if you're walking in a group, ensure everyone knows the route.

UPHILL TECHNIQUES

When traveling uphill, lean forward slightly, maintain your momentum but shorten your stride. When pushing upward, keep your feet flat on the ground.

- Carry a walking staff for support as you ascend.
- Move your legs forward from the hips.

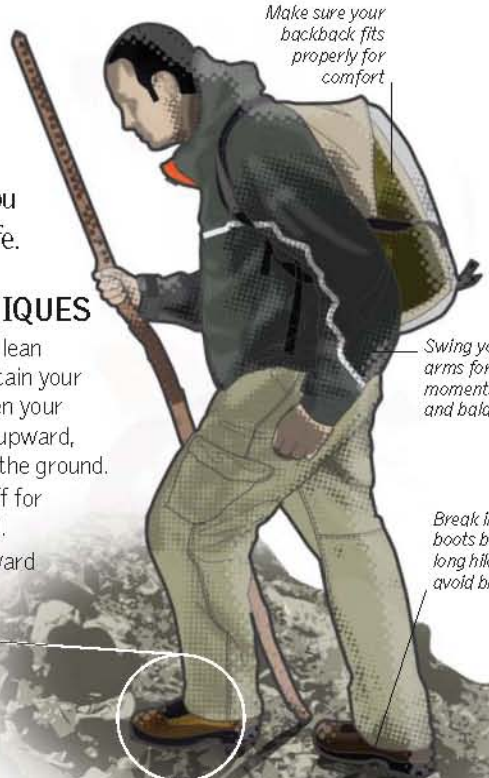

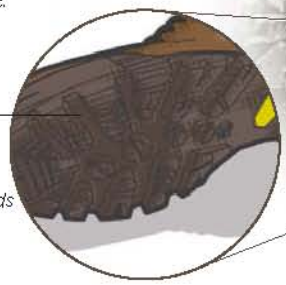

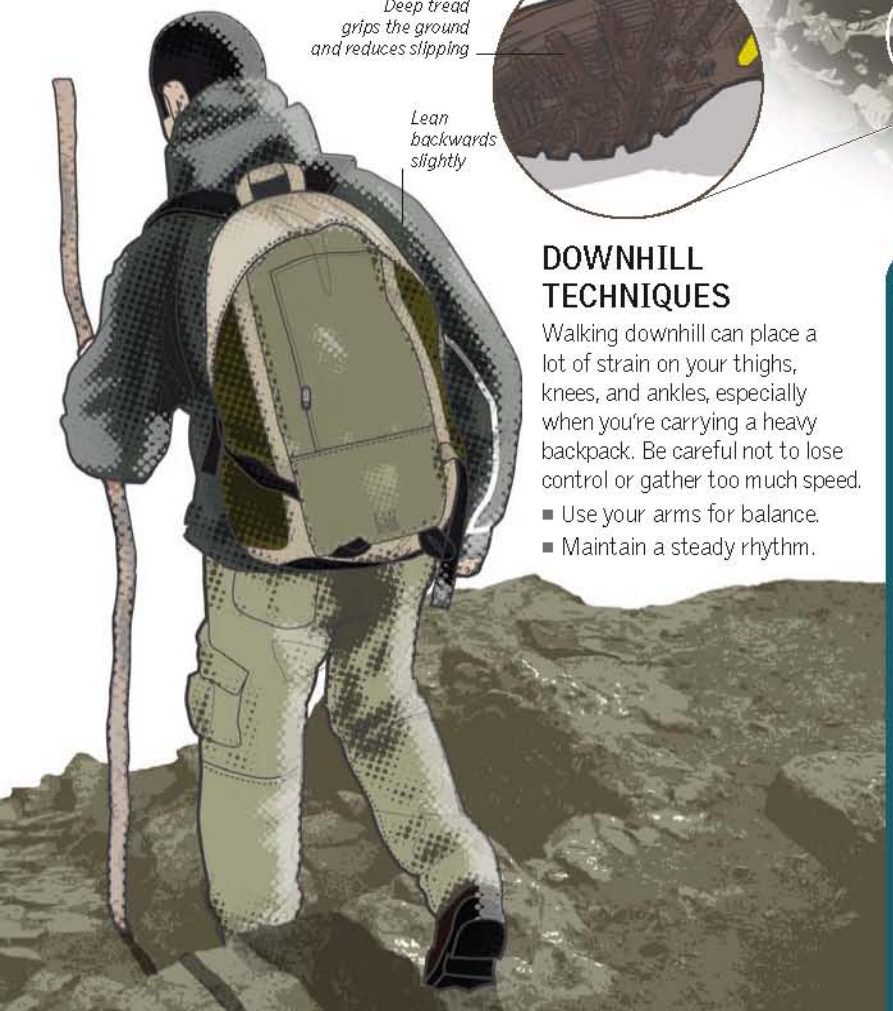

DOWNHILL TECHNIQUES

Walking downhill can place a lot of strain on your thighs, knees, and ankles, especially when you're carrying a heavy backpack. Be careful not to lose control or gather too much speed.

- Use your arms for balance.
- Maintain a steady rhythm.

WALKING AT NIGHT

Unless you're in the desert and it's cooler to move in darkness, avoid walking at night because of the added risks posed by navigational difficulties, and predators that hunt at night. If there's no alternative, try the following:

- Use your flashlight or improvise a torch by setting alight some birch bark or similar.
- If this isn't possible and you've time, close your eyes for 20 minutes to allow them to adjust to night vision.
- Use your walking stick or pole to feel in front of you for obstacles, tripping hazards, or sudden drops in the ground.
- Keep your pace slow and deliberate, and check your compass regularly.

NEGOTIATING DIFFICULT GROUND

One of the most challenging terrains to hike over is "scree," a mass of small rocks that slides underfoot making uphill and downhill travel difficult. The slippery nature of the surface can make progress slow, and care should be taken to avoid falling and injuring yourself. Hiking on scree is hard work, but using the correct techniques will help you to advance confidently, efficiently, and safely.

TRAVERSING SCREE

Choose a zig-zag route and look for a path that contains similar-sized rocks. Walk sideways across the slope, taking small steps and testing rocks for stability before placing your full weight on them.

Rocks at the edge of the slope are likely to be larger and more stable

ASCENDING SCREE

Tread carefully, kicking your toes into the slope and testing the step before transferring your weight to that leg. Alternatively, walk with your feet splayed, placing your weight on the inside of each foot.

DESCENDING SCREE

"Screeing" downhill combines sliding with slow-motion jogging. When you have the technique it can be great fun, but avoid larger rocks to prevent ankle injuries.

TRAIL MARKERS

Your map and compass should be your primary method of navigation, but keep an eye out for trail markers, or "blazes," during your hike.

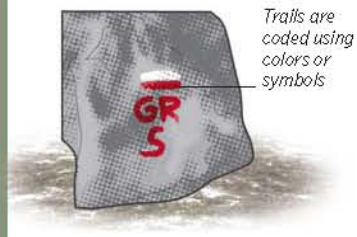

Trails are coded using colors or symbols

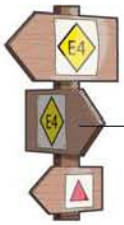

Arrows indicate direction

ROCKS

Trail blazes painted onto rocks and boulders are especially common in rocky terrain. They may be low on the ground so look carefully.

SIGNPOSTS

Made from wood, metal, or plastic, these signposts are particularly useful in areas where there are few rocks or trees on which to blaze.

Stacks should be visible from a distance

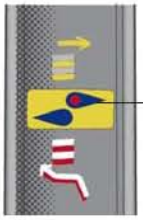

The same sign may indicate several trails

CAIRNS

Piles of rocks, known as "cairns," are designed to be visible in thick fog. Their size can vary from a few rocks to a large stack of boulders.

DIRECTIONAL SIGNS

When reading markers, look out for painted arrows or variations in the blazing, such as a bend. This indicates a change of direction in the trail.

Face the direction of travel

Knees are your center of gravity

1 Ensuring the scree is fine and deep, use gentle hops to launch yourself down the slope. Keep balanced and let gravity do the hard work.

Stay as upright as possible

Use your arms for balance

2 As you build up momentum and rhythm, dig your heels into the slope and slide a short distance with each step, avoiding leaning forward.

Use your arms to break your fall

Keep your knees bent as if you're skiing downhill

3 If you lose your balance temporarily, use your arms to steady yourself, relax your knees, and continue your descent.

CROSSING RIVERS

RIVER CROSSINGS ARE DANGEROUS and should be avoided unless absolutely necessary. In a survival situation, once you are cold and wet—which can lead to hypothermia—it is hard to get warm and dry. Always check your map for routes around the river, and choose the safest available crossing point.

WADING BASICS
Wear footwear when wading to protect your feet from rocks and other dangers. Remove pants to keep them dry and decrease resistance in the water. Use a walking staff for added support.

CROSSING SAFELY

Before you get into the water, make sure you have a change of clothes or means of getting dry on the other side. In cold conditions, collect everything you need to get a fire going (tinder, kindling, and dry wood), and keep this dry as you cross.

CHOOSING YOUR CROSSING POINT

Walking upstream generally leads to shallower water, but be aware that even shallow water can have strong currents, as can water that looks calm on the surface. Always check for bridges further up- or downstream.

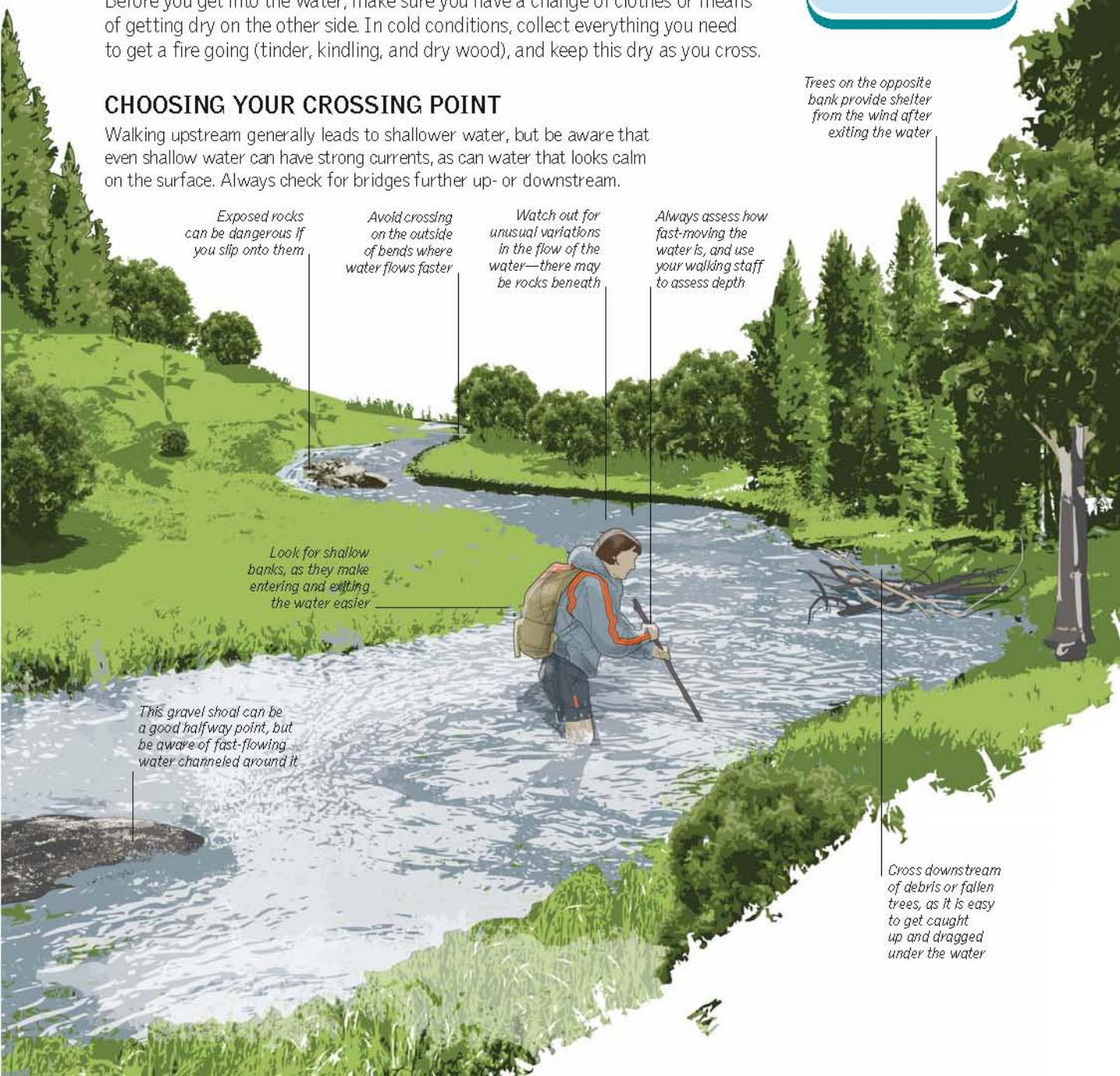

WADING WITH OTHERS

Crossing in a group is safer than crossing alone. Linking your arms together creates a stronger, more stable structure against the current and provides backup for anyone who falls. Loosen the straps of your backpack, and place only one arm through them, so you can release the pack quickly if you fall.

WARNING!
Never cross white water, or rivers that have flooded. In these situations it is safer not to cross at **all** than to risk injury.

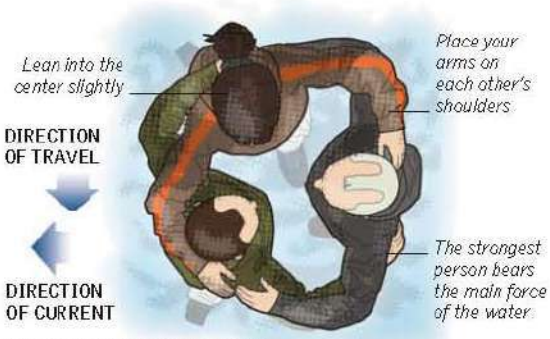

CROSSING IN A HUDDLE

Positioning the strongest person upstream, with the others providing stability and support, link your arms tightly and take short, deliberate steps across the river.

CROSSING IN A LINE

Keeping well balanced, cross the river perpendicular to the current. Move slowly and position each step carefully to avoid being swept away by the force of the water.

WADING ALONE

Wading across a river alone is not ideal, but if you have no option then your walking staff or "survivor's third leg" (see p. 72)—or simply any long stout stick—will provide additional support and balance. Use it as a probe to alert you to any sudden changes of depth in the river bed.

USING A WALKING STAFF

Facing the current, walk diagonally across the river, placing each foot down firmly and deliberately, and leaning on your walking staff for support. Your feet and your staff are your three points of contact—keep at least two of them in contact with the river bed at all times.

CROSSING WITH ROPES

Using ropes is a good option for dangerous crossings, but they can get tangled, and drag you under the water. Always aim for the easiest and safest option—using ropes is complicated and should be your last resort.

1 A "safety" rope is tied to a tree and then carried across by the strongest person, who also takes a "crossing" rope with a carabina tied centrally.

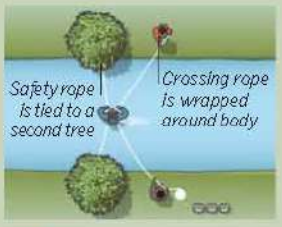

2 The safety rope is tied to a tree and the carabina clipped to the safety rope. The carabina is pulled back and fixed to the next person to cross.

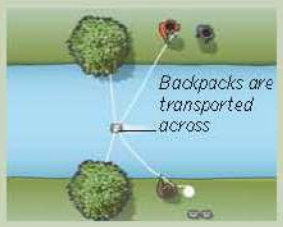

3 He or she crosses upstream of the safety rope, holding it for support. Backpacks can be clipped to the carabina and transported, too.

4 The final person unties the safety rope from the tree, wraps it around his or her body, and wades across, supported by the others.

SCRAMBLING AND CLIMBING

CLIMBING WITHOUT ROPES is known as "scrambling." As with roped techniques, the idea is to maintain three points of contact with the rock at all times—either two feet and one hand or both hands and one foot. Your legs should power the climb, with your arms used predominantly for balance. Proceed carefully, making sure you are always balanced and confident of your next move.

WARNING!
Climbing is a dangerous activity that should only be considered as a last resort. Ideally, plan ahead using your map to find a way around the obstacle.

HAND AND FOOT HOLDS

When planning your route, choose your hand and foot holds carefully. Don't reach too far, and test each hold for stability before using it to support your weight.

EDGES

For very small footholds, place the inside edge of your foot in the hold in the rock to take the pressure off your toes.

FINGER SHELVES

Curl your fingers over the rock for a secure fingerhold. The larger the fingerhold, the more secure it will be.

LARGE POCKETS

Placing your foot securely within the pocket, balance yourself so both your arms and legs take the strain.

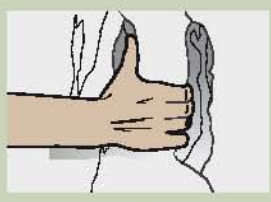

SIDE PULLS

Side pulls can be used to maintain balance or to pull yourself across the rock. Grip the rock tightly.

PROTRUSIONS

If possible, place the entire sole of your foot on the protrusion. If only the ball of your foot will fit, keep your heel low.

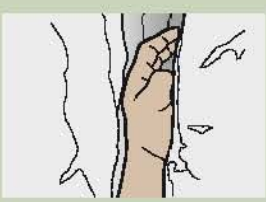

HAND JAMS

Insert your hand into the crack, push your thumb into your palm, arch your hand, and wedge it tightly inside.

BASIC CLIMBING TECHNIQUES

Always climb within your capabilities. It's important not to take any risks, as it's far safer to descend and start the climb again than to chance a fall. Before you start to climb, plan the easiest and safest route.

SCRAMBLING

When moving across the rock you will need to combine a variety of techniques to negotiate different obstacles. Always plan your moves several steps in advance, and keep three points of contact with the rock at all times.

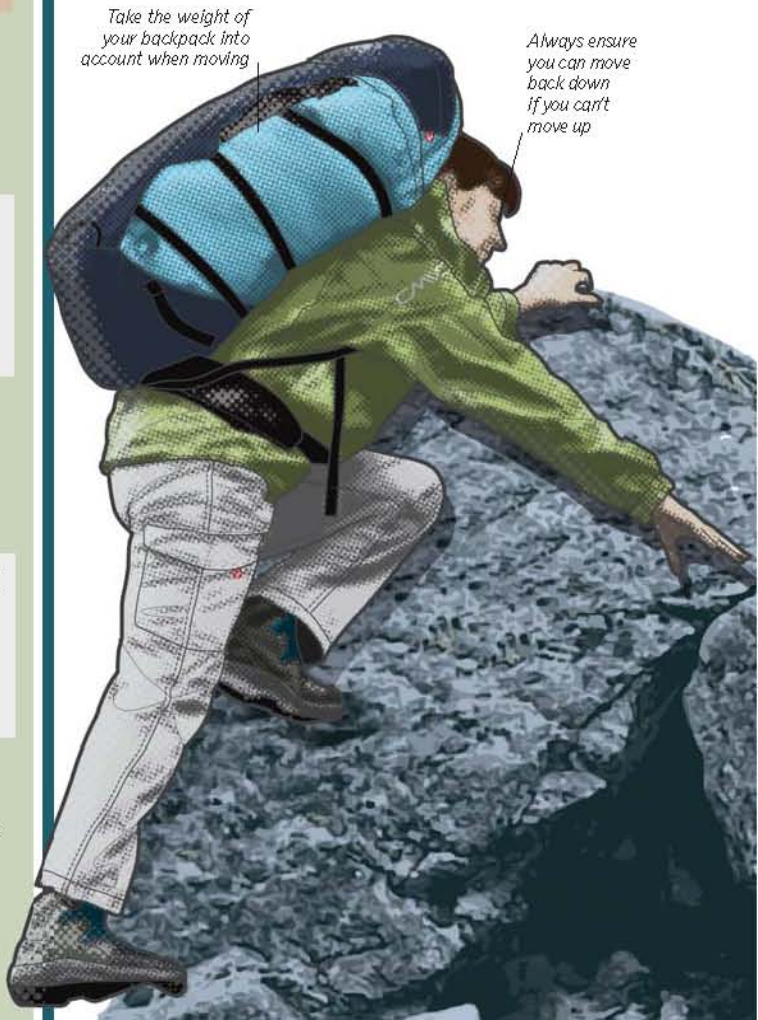

MANTELLING

The mantelling technique is used to climb overhangs in the rock. Use your lead ankle and then your knee to lever your body up over the obstacle. It's a physically challenging technique, but useful.

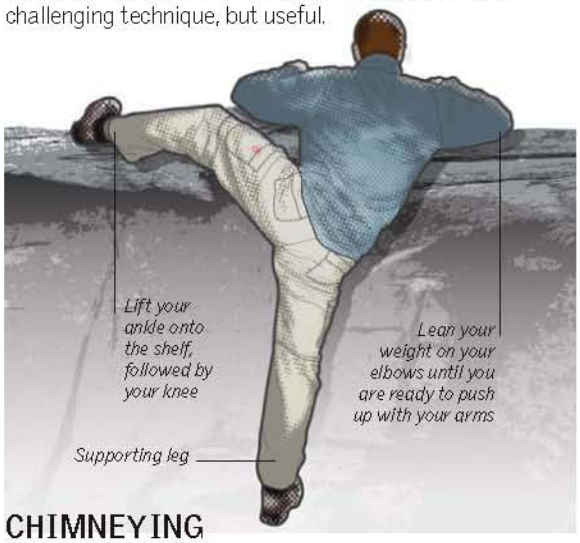

CHIMNEYING

To climb up the inside of large rock clefts, or "chimneys," use the chimneying technique. To ascend, move your back and hands up the wall, and push and follow with your legs. It's easy to get stuck at the top so plan your exit route carefully.

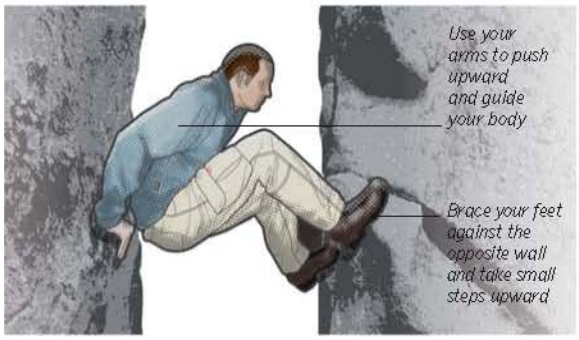

STRADDLING

If the chimney is relatively wide, you may need to alter your body position to straddle the gap. With a leg and arm on each wall, inch yourself up the rock using your legs to push your body upward.

CLIMBING WITH EQUIPMENT

The advantages of climbing with ropes cannot be underestimated as far as safety is concerned, although a lot of specialized equipment is required. As well as ropes, helmets, and harnesses, climbers can also use screws to secure themselves to the rock as they progress each stage.

BELAYING

Belaying is a technique used by climbing partners to safeguard each other during a climb. The belayer secures himself to the rock and stands on a suitable ledge while the lead climber ascends. Once at the top, the lead climber secures himself and acts as the belayer.

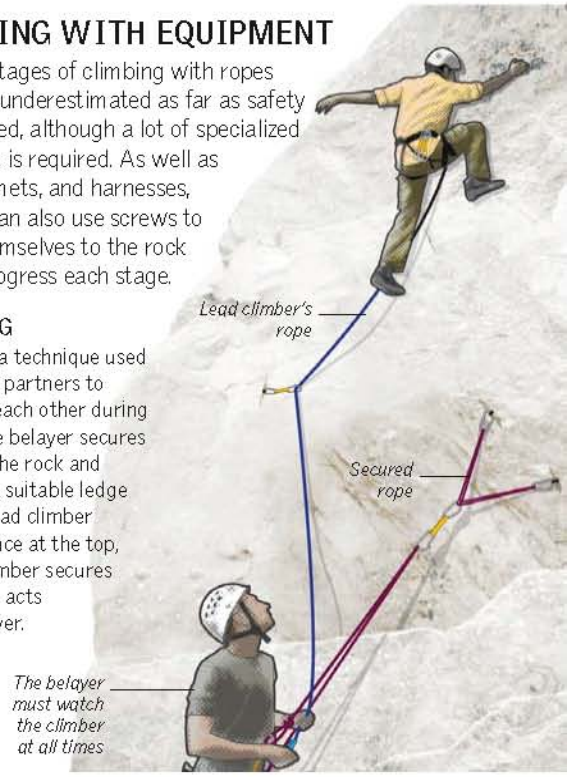

ICE CLIMBING

Ice climbers utilize similar techniques to rock climbers, although they carry ice axes and wear crampons on their feet, to help them grip the ice.

MOVING UPWARD

As with rock climbing and scrambling, ice climbers ascend using power from their legs and arms. Crampons and ice axes are used to create holds in the rock.

TRUE-LIFE ACCOUNT

EXTREME SURVIVAL— IN THE MOUNTAINS

USEFUL EQUIPMENT

- Two-way radio
- Avalanche beacon and probe
- Waterproof clothing
- Spare clothing layers
- Collapsible shovel
- Climbing equipment, if needed
- Signal flares
- Map, compass, GPS
- Survival tin, bushcraft knife
- Cell/satellite phone
- Poncho/bivy sack

STEVEN GREEN, 32, AN EXPERIENCED HILL WALKER from Dumfries, Scotland, spent four nights lying on a remote mountainside after a life-threatening fall. Suffering from a broken jaw and fractured skull, he survived via a combination of good preparation, quick thinking, and improvisation.

On Thursday, October 7, 1999, Steven set out alone for a trek in the hills of Wester Ross in the northwest Highlands of Scotland. He was well equipped with, among other things, a cell phone, bivy sack, pita bread, and water—and had taken the wise precaution (or so he thought) of informing his girlfriend of his intended route.

"HE SLIPPED ON WET GRASS AND FELL DOWN A WATERFALL"

Following an all-day trek and an overnight camp, he was descending the mountain on Friday when he slipped on wet grass and fell down a waterfall, coming to rest in the shallow, rocky riverbed. He knew he would not survive for long in the cold water so, despite being in intense pain, he crawled to the relative safety of the bank, then wisely got into his bivy sack for warmth and waited. He sustained himself by eating pita bread, which he softened in water because he couldn't chew. His phone was damaged in the fall so he was unable to call for help. When he failed to return home on time, the alarm was raised by Steven's girlfriend, but she had forgotten to write down his intended route so no one knew his exact location.

Steven experienced a further four days and nights of exposure while local and RAF rescue teams and search-and-rescue dogs combed the mountains for him. As the days passed they feared the worst, but when they located Steven on Tuesday morning—after finding his car and, inside it, his route map—they found him in good spirits. At hospital Steven was diagnosed with a fractured skull, broken jaw, missing front teeth, cuts and bruises, and a leg injury—but he was alive.

WHAT TO DO

ARE YOU IN DANGER?

NO YES

If you are in a group, try to help any others who are in danger

Get yourself out of it:
Elements—Find or improvise immediate shelter
Animals—Avoid confrontation and move away from danger
Injury—Stabilize condition and apply first aid

ASSESS YOUR SITUATION
See pages 234–35

DOES ANYONE KNOW YOU WILL BE MISSING OR WHERE YOU ARE?

NO YES

If no one knows you are missing or where you are, you will need to notify people of your plight by any means at your disposal

If you are missed, a rescue party will almost certainly be dispatched to find you

DO YOU HAVE ANY MEANS OF COMMUNICATION?

NO YES

You are faced with surviving for an indefinite period—until you are located or you find help

If you have a cell or satellite phone, let someone know your predicament. If your situation is serious enough to be worthy of emergency rescue, and you have a Personal Locator Beacon (PLB), you should consider this option

CAN YOU SURVIVE WHERE YOU ARE? *

NO YES

If you cannot survive where you are and there are no physical reasons why you should remain, you will have to move to a location that offers a better chance of survival, rescue, or both

Address the Principles of Survival: Protection, Location, Water, Food

YOU WILL HAVE TO MOVE **

YOU SHOULD STAY **

DO

- Keep all clothing dry and clean
- Collect drinking water from fast-moving streams; filter and disinfect if you have the means
- Be on constant lookout for signs of cold-related injuries such as frostnip, frostbite, and hypothermia
- Regulate your clothing to avoid overheating when moving and hypothermia when static
- Use a walking staff to aid safe movement
- Watch the weather closely and be prepared to change your plans at short notice—mountain climates are highly unpredictable
- Improvise shelter when not on the move

DON'T

- Descend hills in a careless manner; zig-zagging across hills is less tiring and puts less strain on leg muscles
- Travel at too fast a pace—high altitude equals less air and this will place greater demands on even a fit person
- Sweat too much, as the moisture will chill you further
- Ignore any opportunities to collect dry tinder/kindling

DON'T

- Allow your extremities to get too cold as those areas are most susceptible to frostbite
- Overlook the dangers of carbon monoxide poisoning in cramped shelter. Don't let candles, stoves, or fires burn all night
- Breathe on your hands to warm them up: you are exhaling warm air that you will have to replace with cold

DO

- Select a shelter site that offers protection from the elements. Build it no bigger than it needs to be
- Fill plastic bags or spare clothing with dry foliage and use as a mattress or pillow to insulate you from the cold, damp ground
- Deploy all your aids to location and prepare for immediate use
- Check upstream for the quality of your water source
- Light a fire and (if in a group) take turns tending it to keep it going all night
- Continually re-assess your situation and adapt your actions as necessary.
- Be constantly alert for signs of rescue

* If you cannot survive where you are, but you also cannot move owing to injury or other factors, you must do everything you can to attract rescue.
** If your situation changes (for instance, you are "moving" to find help, and you find a suitable location in which you can stay and survive) consult the alternative "Do" and "Don'ts."

MOVING OVER SNOW

PREPARATION IS ESSENTIAL when hiking over frozen terrain. Not only must you be physically fit—the conditions make for slow, exhausting progress—but you must also have the right equipment and know how to use it. Wearing snowshoes or skis, and breathable and layered clothing to regulate your temperature, is essential. With the correct techniques you can hike safely and enjoy the surroundings.

WARNING!
Walking over deep snow without snowshoes is called "post-holing" and should be avoided unless absolutely necessary. Sinking into the snow will leave you exhausted and very wet, which in cold conditions can quickly lead to hypothermia. The exertion can also cause sweating and a dangerous reduction in your body temperature when you stop.

SNOW AND ICE

Knowing how to deal with different types of snow and ice is invaluable when hiking over frozen terrain.

DEEP SNOW

- If in a group, walk in single file and take turns at the front: the most strenuous position, as you're creating the path.
- Avoid rocks—in spring, they absorb heat and the snow above becomes unstable.

FROZEN CRUST

- Use a walking staff to test the snow ahead. An ice crust above deep snow may take your weight, but progress carefully.
- Later in the day you may come across melted depressions, called "sun cups." Cross on the rims to avoid sinking into the snow.

SLOPES

- Kick firmly into the slope and test your weight before ascending. When descending, you can use a technique called "boot skiing" (skiing without your skis).
- Choose your route based on the conditions: travel in straight lines if you are able to, or in zig-zags if the terrain is steep.

ICE

- Always progress carefully, using your walking staff to test the ice, especially over rivers and lakes. If in a group, rope yourselves together for safety.
- Wear crampons for extra grip. On very steep slopes, use your ice ax to cut steps.
- Use your ice ax to halt a fall by turning to face the slope and digging it into the ice.

GLACIERS

- Never attempt to cross glaciers without a guide—glaciers require specialized skills.

WAYS OF TRAVELING

When traveling over snow, your main aim is to get to your destination as safely as possible, without expending too much energy or losing too much body heat. Methods range from improvised snowshoes to motor vehicles.

USING SNOWSHOES AND SKIS

Snowshoes and skis are an effective way of traveling over snow. They work by spreading your body weight over a larger surface area, which enables you to walk on the surface of the snow rather than sink into it. Always set off cold, as you'll warm up quickly, and add or remove layers as required.

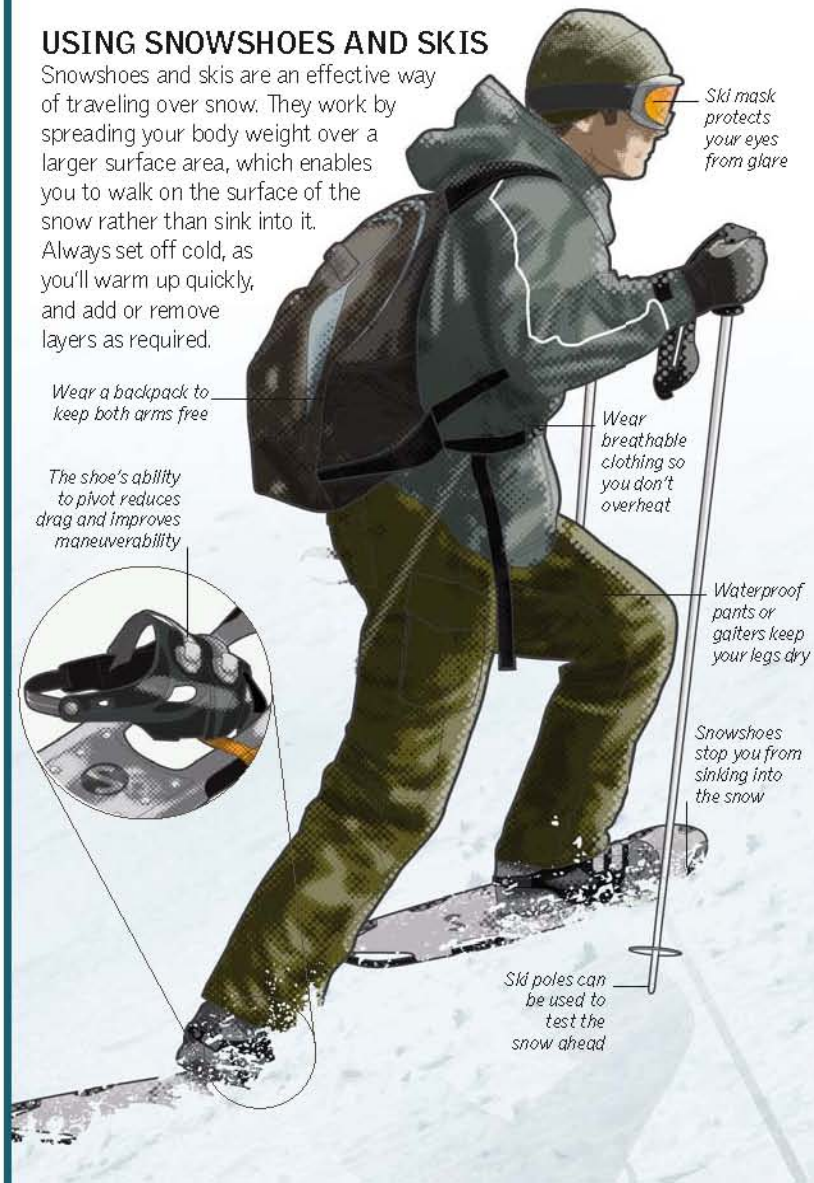

MAKING IMPROVISED SNOWSHOES

If you don't have any pre-manufactured snowshoes—for example, in a survival situation—you can make some very simple shoes to help you negotiate the snow more effectively, using your knife, some green wood, and cordage. They may take time to build, but will save time and effort in the long run.

1 Cut five lengths of green wood. They should be as thick as your thumb and the same length as the distance from your foot to your armpit. Cut three shorter lengths for the cross-pieces.

- Lash the ends of the five longer pieces together securely using cordage.
- Calculate where the ball of your foot will be positioned on the shoe and lash a cross-piece across the five lengths. Ensure the shoe will balance.

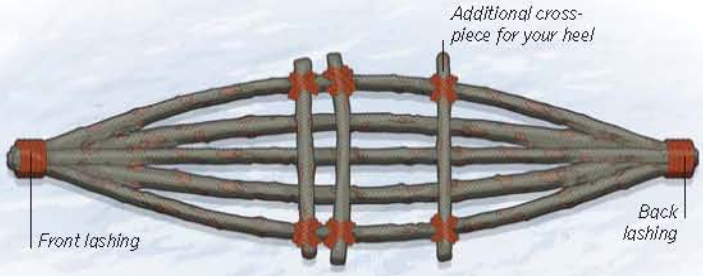

2 Lash the five loose lengths of wood together at the back of the shoe. It's important to make this lashing as secure as possible.

- Fix the second cross-piece roughly 2 in (5 cm) behind the first.
- Lash the third cross-piece where your heel will rest.
- Repeat steps 1 and 2 to make a second shoe before progressing to step 3.

3 Place your foot on top of the snow shoe, ensuring that the ball of your foot sits directly over the front cross-piece and that your heel is positioned on the back cross-piece.

- Tie your boot to the snow shoe using whatever cordage you have, but ensure the heel is allowed to remain free to pivot. Repeat for the other foot.

ALTERNATIVE METHODS

These alternative methods will also spread your weight on the snow. If you have no gaiters, tie plastic bags around the bottom of your legs to keep them dry.

USING BRANCHES

Ideal for short distances, a simple way to get you out of deep snow and to a road or track is to attach branches to your feet using cordage. Select a tree, such as pine, that has strong, close branches.

USING SAPLINGS

You will need branches that have some flexibility. Gently bend the longest branch into a teardrop shape and lash the two ends together. Heating the saplings over a fire will make them easier to bend.

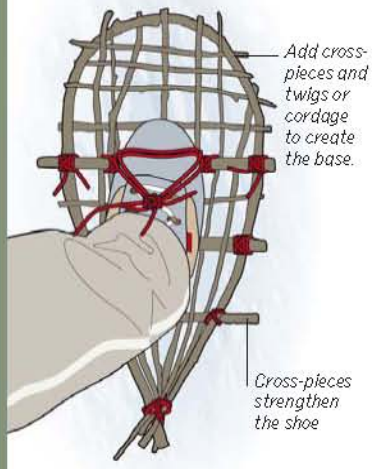

MAKING A SLED

If you find yourself in a survival situation with heavy equipment to carry over snow, making a simple sled or "pulk" (see panel, right) will help you transport it more efficiently. You could also build one to transport an injured person or small child. You can make your sled as large or as small as you need, providing you follow these basic principles.

TOOLS AND MATERIALS

- Survival saw or pocket chainsaw
- Knife
- Cordage
- Forked branch (for bulk of sled)
- Two branches (for bracing pieces)
- Sticks (for cross-pieces)

1 Using a short saw or penknife, cut a forked branch to form the base of the sled.

- Tie the two ends of the fork to the main branch with cord. This creates tension and forces the branches to curve and act as runners.

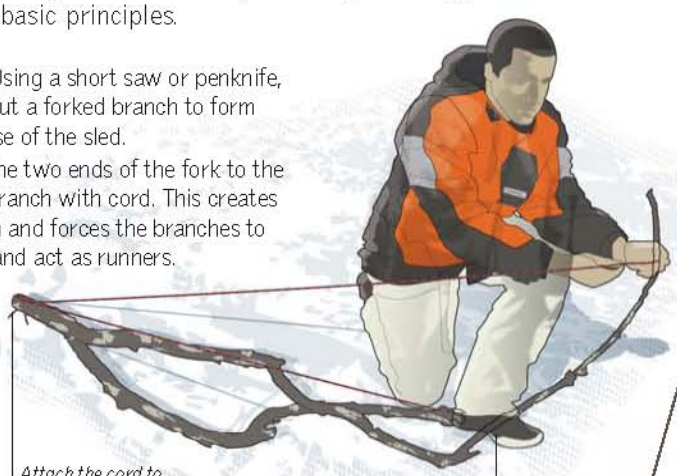

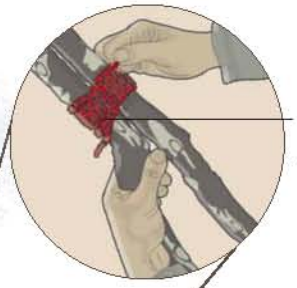

2 Lash an additional branch to one of the runners to create a bracing piece.

- Repeat on the other side.
- The bracing pieces strengthen the branches and prevent any cord from touching the ground—if it does, it will wear out quickly and disintegrate once the sled is on the move.

ANIMAL POWER

Huskies, or sled dogs, are ideal for transporting people and equipment across winter terrain. Because of their thick coat they can endure very low temperatures, and their large, furred feet allow them to move quickly over snow. Huskies work well as a team and can pull heavy loads easily and efficiently.

TRAVELING WITH DOGS

Although huskies are easy to care for and relatively simple to handle, you should not undertake an expedition with dogs unless you are traveling with an expert or have first undergone extensive training. The same goes for all expeditions in snowy terrain.

HUMAN POWER

Pulling your equipment behind you using a pulk is an extremely efficient way of moving over snow. Pulks are small, low-slung toboggans, typically made from lightweight plastic, and come in many different sizes.

TRAVELING WITH A PULK

Although pulks are the most efficient method of carrying heavy loads over snow using human power, they can be hard work, especially in softer snow. Wear breathable clothing to allow excess body heat to escape. Know how to release your harness quickly in an emergency. If you are in a group, one person can be harnessed to the rear of the pulk to act as a brakeman when going downhill.

WARNING!
Take care when traveling down slopes, as the pulk can easily pick up speed and become difficult to control if you have no brakeman.

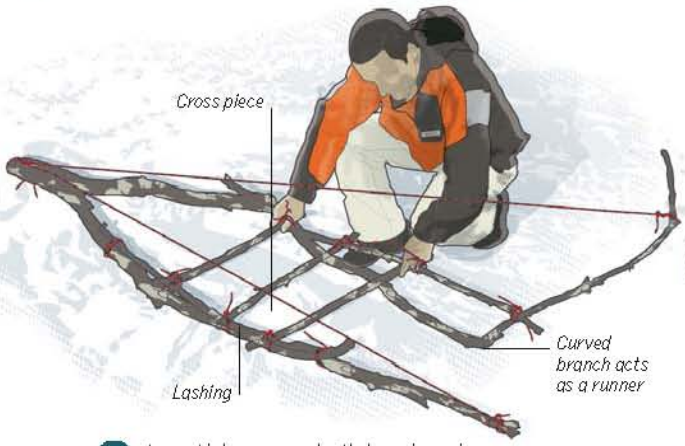

3 Lay sticks across both bracing pieces and lash them together. These cross-pieces strengthen the structure.

- You can use as many sticks as you can find, but three or four should suffice.
- These cross-pieces will also form the main carrying platform of your sled.

4 Make a handle for the sled using a small branch. The branch should be as smooth as possible to make pulling the sled more comfortable.

- Attach the handle to the front of the sled using a long piece of cord.

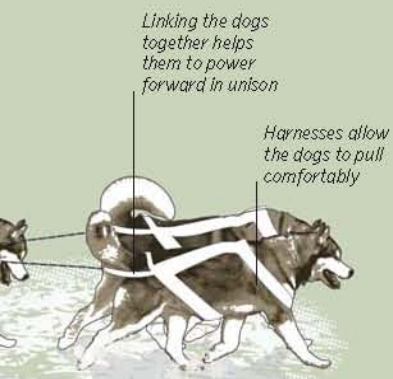

MOTOR POWER

Snowmobiles are practical and quick, and have revolutionized arctic travel. They are easier to handle than huskies, and can reach relatively high speeds. However, although you can cover great distances in short periods of time, if anything goes wrong you may be far from help. Always carry survival equipment when you travel.

BEFORE YOU GO

Prior to a snowmobile trip, regardless of distance, it is vital to check your vehicle maintenance and ensure that you have sufficient fuel for the trip. Plan your route carefully and inform someone of your plans so that they can raise the alarm if you don't reach your destination when planned. Dress warmly, wear goggles or a ski mask to reduce glare, and drive safely.

USING PACK ANIMALS

PACK ANIMALS CAN BE USED to transport heavy loads when hiking over wilderness terrain. Ideally suited to carrying substantial loads, they can cover great distances in areas where vehicles would struggle. When you are organized, and know how to care for the animals, traveling with pack animals can be very rewarding.

PACKING TIPS

- Know the capabilities of your animal, and what weight they can safely carry over the necessary distance.
- Before loading the animal, groom it well and check for sores or ticks.
- Place heavier items at the bottom of your panniers and lighter items at the top.

LOADING THE ANIMAL

Stand on level ground, making sure that the animal has all four legs placed firmly on the ground. Before placing the loaded panniers onto the animal, lift them to ensure they are well balanced. Rearrange if one is lighter. You may need to hobble or blindfold the animal to keep it still while loading.

ANIMAL WELFARE

Remember to transport food, water, and supplies for the animal as well as for yourself. Their welfare is paramount, so treat them well.

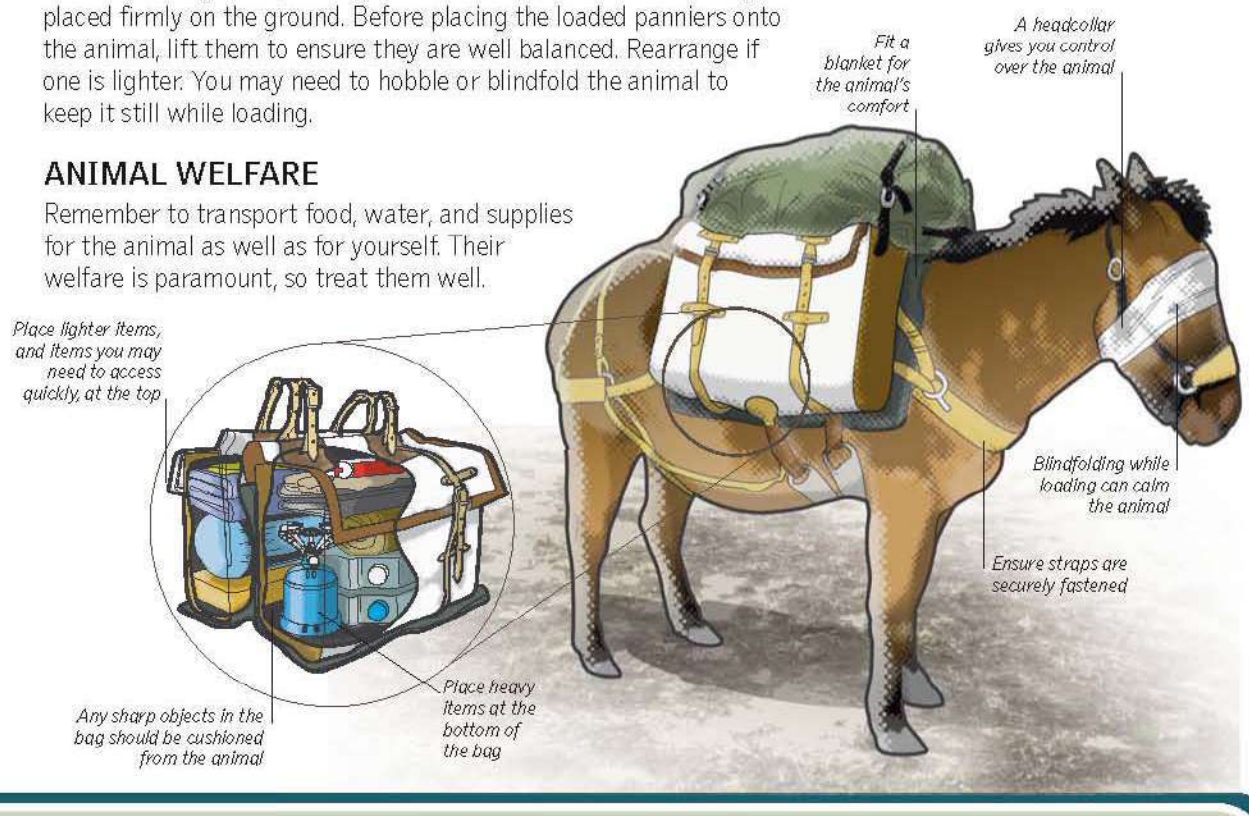

PACK TRAINS

Roping animals together on the trail is standard practice when traveling with more than one animal. When tying them together, ensure the ropes are long enough to allow the animals to walk comfortably, and short enough to prevent tripping.

THE PEOPLE

Appoint a driver to lead the group and take control of navigation. You will need sufficient people to control and care for the animals.

USING PACK ANIMALS TO SURVIVE

In an emergency situation, your priorities change significantly when using pack animals. Unpack the panniers and remove any heavy items not essential for survival. If necessary, you can climb on and ride to safety or, in a worst-case scenario, your animal can be used for food.

CARRYING INJURED PEOPLE

If a member of your group is injured, build a makeshift stretcher (see pp. 278–79) and use the animal to drag him or her to safety.

WARNING!
Do not attempt to use pack animals without expert help or extensive training. The animals require a lot of care, and you will need to organize your days meticulously.

CHOOSING YOUR ANIMAL

The availability of pack animals depends on the region you are traveling in. Within any region, your choice of animal should take into account the load you are expecting them to carry, the distance you require them to travel, and the terrain you plan to cross. However, remember that individual animals' abilities can vary depending on their age and size.

ANIMAL	REGION	AVERAGE LOAD	ADVANTAGES	DISADVANTAGES
Horse	Worldwide	▪ 175–240 lb (80–110 kg) ▪ 20 percent of body weight	▪ Easy to train, with a good temperament ▪ Strong ▪ Can tolerate heat ▪ Can manage steep terrain	▪ Require grooming ▪ Can stray from camp if not securely tethered
Mule	Eurasia, Americas	▪ 120–180 lb (55–80 kg) ▪ 20 percent of body weight	▪ Will stick by a mare closely—this can be used to keep them moving and to keep them within the camp at night ▪ Very hardy and can manage steep terrain	▪ Young mules can be easily startled, so good training is essential ▪ Stubborn
Husky	Arctic	▪ 85lb (40kg) ▪ 7 dogs can pull 600lb (270 kg)	▪ Hardy—can cope well with snow and cold conditions ▪ Can travel fast	▪ Need a lot of fresh meat, which must be carried ▪ Prone to fight among themselves
Camel	Central Asia, North Africa, Australia	▪ 198–308 lb (90–140 kg) ▪ 30 percent of body weight	▪ Versatile on different terrains ▪ Can drink up to a quarter of its body weight and then go without water for several days	▪ Strong-willed and difficult to control ▪ Violent—can spit and bite
Llama	Andes	▪ 77–123 lb (35–55 kg) ▪ 25–30 percent of body weight	▪ Minimal environmental impact ▪ Travels well at high altitudes on difficult and steep terrain	▪ Makes horses and mules nervous ▪ Can be difficult to control if poorly trained
Elephant (Indian)	South Asia	▪ 1,650–2,750 lb (750–1,250 kg) ▪ 25 percent of body weight	▪ Can carry very heavy loads ▪ Can manage steep terrain	▪ Slow-moving ▪ Requires large amounts of food and water ▪ Takes time to train
Ox	Eurasia, Americas	▪ 300–450 lb (135–205 kg) ▪ 30 percent of body weight	▪ Hardworking ▪ Very strong ▪ Surefooted and can manage steep terrain	▪ Slow-moving ▪ Stubborn

FOUR-WHEEL DRIVING

A FOUR-WHEEL-DRIVE (4WD) vehicle can cover large areas of difficult terrain that would be inaccessible in a 2WD vehicle. With an experienced driver, most 4WD vehicles can tackle deep mud, water, snow, ice, and sand.

CHOOSING YOUR VEHICLE

When choosing your vehicle it's important to consider what you'll be using it for. Large vehicles have more internal space than smaller ones, but may find it harder to negotiate very rough ground without getting stuck. Similarly, very powerful vehicles can cross almost any terrain but use a lot of fuel, so may not be a viable option for long-term expeditions.

EXPEDITION CHECKLIST

- Check your map and inform others of your route and intended timeframe.
- Ensure the vehicle is fit for the journey, with all necessary spares. Check the fuel, oil, water, brakes, and hydraulic fluids, and check that there are no leaks.
- Inspect the tire treads, wheel nuts, lights, and steering arms.
- Always carry spare water, wheel(s), and fuel, as well as survival equipment, sand ladders, a 12v heavy-duty tire inflator, winching strap, and first aid kit.

GENERAL DRIVING TECHNIQUES

Four-wheel-drive vehicles enable you to manually switch between two-wheel-drive, for driving on roads, and four-wheel-drive, for low-traction conditions, such as on soft ground. The difference is the number of wheels powered by the engine at any one time.

WARNING!
Don't hook your thumbs around the steering wheel when driving over rough terrain. Hitting a rut can jerk the wheel and break them.

CHOOSING A ROUTE

When unsure about the conditions immediately ahead, walk the route first, checking for potential problems and obstacles. Where necessary, mark the route you have walked and follow these markers when you drive. Ask yourself whether you really need to go that way and, if so, which route is best to avoid getting stuck. And, if you do get stuck, what are your options for self-recovery or escape?

WHEN TO ENGAGE 4WD

Engaging 4WD uses up a lot of fuel and should not be used on hard roads because of the risk of damaging tires and gears. Always choose the four-wheel-drive option on difficult terrain. It allows you to travel in a very low gear and gives the vehicle a far superior traction. Just before starting to cross the rough ground, stop the vehicle and engage the four-wheel-drive.

DRIVING AS A TEAM

Driving off-road over rough terrain is both mentally and physically tiring. Take regular breaks and share the driving if in a group. You should always have at least two people in the vehicle:

- A driver—who takes responsibility for powering and steering the vehicle.
- A spotter—who takes responsibility for navigation and helps the driver pick a good route across the terrain. This may involve getting out of the vehicle and guiding the driver across a difficult section of ground.

DRIVING OVER DIFFICULT TERRAIN

A 4WD vehicle handles difficult terrain well because all four wheels can be powered by the engine at any one time (gas engines are generally more powerful, but diesel engines last longer and work well at low speeds). Using a few simple techniques, you will find that you are able to drive over terrain—and to places—that would otherwise be inaccessible. Always remember that the principle behind 4WD is to reduce the chances of you getting stuck—not to allow you to go further until you do get stuck.

MAXIMUM TRACTION

Your vehicle's traction on soft surfaces can be improved by lowering your tire pressure slightly. As a guide, place a brick ½in (1 cm) away from the side of a rear tire on flat ground, deflate the tire until it touches the brick, measure this pressure, and apply to all four tires. Make sure you can reinflate them.

DRIVING ON SAND

In soft sand the tires tend to move the sand from the front of the tire to the rear of the tire. If forward movement is halted for even a few seconds the wheel can dig itself into a hole. To prevent this, continually steer from side to side, so that the tire steers out of its own ruts. Avoid rapid changes in speed.

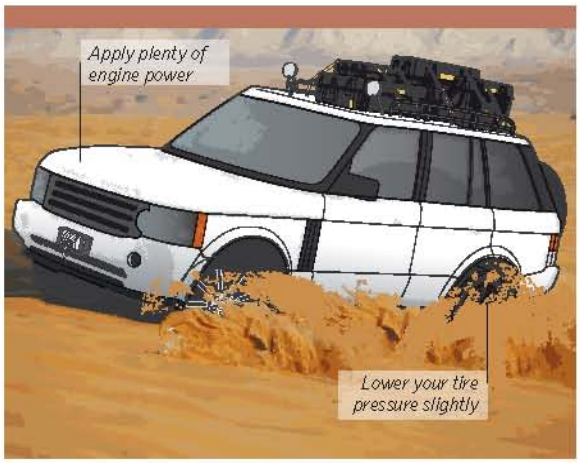

MUDDY TRACKS

Driving in mud requires concentration and the ability to adapt your driving to suit the conditions. In deep mud, use wide tires or lower the tire pressure slightly; however, if there is a hard surface below the mud this will make things worse. Steering outside existing track ruts is most likely to ensure the best traction.

DRIVING ON SNOW AND ICE

Snow and ice require very smooth driving methods. Apply gradual pressure to the accelerator and brakes to avoid wheel spin, use low gears, especially when traveling downhill, and avoid changing gears unnecessarily. Using snow chains will increase traction, and you should practice fitting them before you need them.

CROSSING WATER

Always walk your route before driving across water. If the water seems too deep or the current too fast, don't attempt the crossing. It's important to drive at the correct speed—too fast will send water everywhere, but driving too slowly may flood the engine bay.

RECOVERY FROM SOFT GROUND

Although using the correct techniques will certainly improve your chances of crossing soft ground successfully, it's important to know what to do if you do get stuck. Ideally you should never set out on a driving expedition with fewer than two vehicles. A second vehicle can be used for winching, dealing with breakdowns, or driving for help.

EQUIPMENT LIST

- Winching mechanism and strap
- Cable or tow rope
- Spare tire (inflated)
- Shovel and saw
- Sleeping bag or blanket
- Branches or stakes
- Sand ladders or mats

BASIC RECOVERY TECHNIQUES

When you get stuck, it's tempting to keep trying to aggressively free your vehicle. However, churning up what solid ground is left—and digging yourself into deeper ruts—will only make things more difficult in the long run. Stop, evaluate your options (reversing or pushing, digging, using branches, and winching), and calmly decide how best to achieve your aim. Don't act in haste—an ill-thought-out plan could leave you in more trouble than you were in originally.

REVERSING OR PUSHING

If you can't free the vehicle using four-wheel drive, try alternating between reversing and driving forward in first gear.

- If this has no effect, ask the passengers to get out and push while you drive forward.
- If you are making the situation worse, stop immediately and try an alternative solution.

DIGGING

If reversing and pushing don't work, the next step toward freeing your vehicle is to dig down in front of the wheels in an attempt to create a slope that you can then drive up.

- Dig out the sand in front of each tire to create an upward slope
- Drive very gently up the slope. Avoid revving the engine as this can cause the wheels to spin and lose their grip on the ground.

USING BRANCHES

If the vehicle still won't move, place branches, wooden planks, sand ladders, or blankets—in fact, anything that will increase your traction—in front of the wheels. The idea is to give the tires something to grip, and it should get you moving fairly quickly.

- Without revving the engine too much, gently ease the vehicle forward onto the branches or other material.
- Maintain a slow, steady speed and continue until you are back on firmer ground.
- Once clear, remember to stop and pick up your equipment, and remove any obstacles from the track.

WINCHING

If none of the basic techniques work, and your vehicle is still stuck, it's time to consider winching. Use a cable attached to an electric-powered winching mechanism to pull the vehicle out of the hole via a strong anchor point. You can winch to another vehicle if you're traveling in convoy, although you do run the risk of both vehicles getting stuck. When winching, take time to assess your options—using a natural anchor point is the easiest method and should be your first approach.

NATURAL ANCHOR POINTS

Trees, rocks, roots, or deadfalls can all be used as anchor points. When using a tree, always place the strap or cable near the ground, and use a winch strap where possible to avoid damaging the tree. If the tree looks as if it may not be able to provide the required support, tie the tree to others in the vicinity. If you intend to use rocks, ensure they are large enough and firmly embedded in the ground.

BURIED ANCHOR POINTS

If there are no natural anchor points, you can construct an improvised anchor by burying objects such as logs or a spare wheel. Dig a hole in the ground that's at least 3 ft (1 m) deep, attach your winch cable to your improvised anchor, then bury the anchor in the hole. Refill the hole to secure it. If you're using a tire, use the tire lever behind your spare wheel as an attachment point for the cable.

WINCHING TO A STAKE

A series of long stakes in the ground can also provide an improvised anchor point. You will need a long, sturdy main stake—to which the winch cable is attached—and a series of supporting stakes, lashed together for additional strength. Push the stakes securely into the ground at a slight angle. Be careful not to stand near the stakes when winching, in case they come loose.

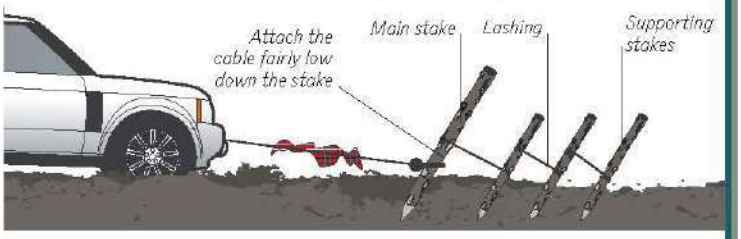

WARNING!
When winching, don't stand within range of the cable in case it snaps under the strain. A breaking cable can cut a tree—or a person—in half. Place a blanket or sleeping bag over it to absorb some of the shock and reduce whiplash.

EMERGENCY ACTION

Knowing what to do in certain emergency situations will help you stay calm if the worst happens.

FRONT SKIDS

The term "skid" covers any kind of slide in which the wheels fail to grip the road. In a front skid—also called an "understeer"—the front wheels fail to turn when you steer so the vehicle continues to travel forward.

- Steer in the direction of the skid but be careful not to oversteer. If you're on ice and skidding straight, step on the clutch or shift to neutral.
- Once you have gained control, correct your course and continue driving. Consult your driver's manual if you have anti-lock brakes.

REAR SKIDS

In a rear skid, the back wheels lose control and the vehicle rotates more than usual and can spin.

- Steer in the direction of the skid but face toward where you want to travel.
- Once facing in the right direction, bring the steering wheel back to the center and continue to steer, regardless of what the rear is doing, until you have gained total control.

BRAKE FAILURE

Before setting out on a drive, always check to see whether there is brake fluid leaking from the vehicle. If so, don't travel. If the brakes fail when on the move, do the following:

- Without turning off the engine, shift down the gears to reduce speed while maintaining control.
- Once the vehicle reaches a speed of less than 25 mph (40 kph) apply the handbrake, keeping a firm grip on the steering wheel.

ACCELERATOR MALFUNCTION

If the accelerator sticks and your engine will not slow down, you can decelerate by putting the vehicle into neutral and applying the brakes.

- If it's safe to do so, switch off the engine, but you'll lose functions like power steering and lights.
- Steer safely to a stop. If possible, avoid using the handbrake as this may cause a skid—although it may be necessary if you need to avoid a collision.

KAYAKING AND CANOEING

TRAVELING IN A KAYAK OR CANOE is a good way of negotiating river systems and open water. With your equipment safely stowed in your boat, expeditions can range from short day trips to year-long adventures.

KAYAKING

Kayaks are an extremely efficient method of water transport, especially on open sea, rivers, and lakes. Their narrow design and light frame make them easy to maneuver.

CAPSIZING IN A KAYAK

Capsizing is an almost inevitable part of kayaking in rough water. The "wet exit"—slipping out of the kayak completely while still holding on to it—is sometimes the only option, but wherever possible use an eskimo roll instead.

ESKIMO ROLL

Mastering the eskimo roll enables you to avoid getting your legs wet and having to climb back in your kayak. With practice, it should become one fluid movement.

1 From the upside-down position, twist your body to the side of the kayak.

- Gripping your paddle, firmly push your arms out of the water, then rest them on the side of the kayak.

2 With your head near the surface, sweep your body and paddle away from the side of the kayak.

- At the same time, flick your hip away to start the rotation of the kayak.

3 Continue the hip flick until the kayak returns to an upright position in the water.

- Straighten your back so you are sitting upright again, and continue to paddle your course.

WARNING!

- Always wear a lifejacket. Empty water bottles in your pockets can be used as makeshift flotation devices in an emergency.
- In a survival situation, never leave your kayak or canoe unless it is life-threatening not to do so. A kayak is highly visible, especially from the air, and is far more likely to be spotted first by rescuers.

BASIC PADDLING SKILLS

Kayak paddles have two blades that sweep alternately through the water, propelling the kayak forward. To steer, paddle on the opposite side to the direction in which you want to go. Some kayaks have a foot-operated rudder.

CATCH

Place your lead blade firmly in the water so that it enters near to your feet.

BRACE

Rotate your torso and lead blade in preparation for the propulsive stroke.

STROKE

Sweep your lead blade firmly through the water to propel the kayak forward.

TRANSITION

As the blade exits the water, repeat the catch on the other side and continue paddling.

CANOEING

As with kayaks, canoes can hold one or two people with equipment, and are powered by the body from a seated position. They are generally wider than kayaks, however, and their decks are open. Canoe paddles have one blade as opposed to two.

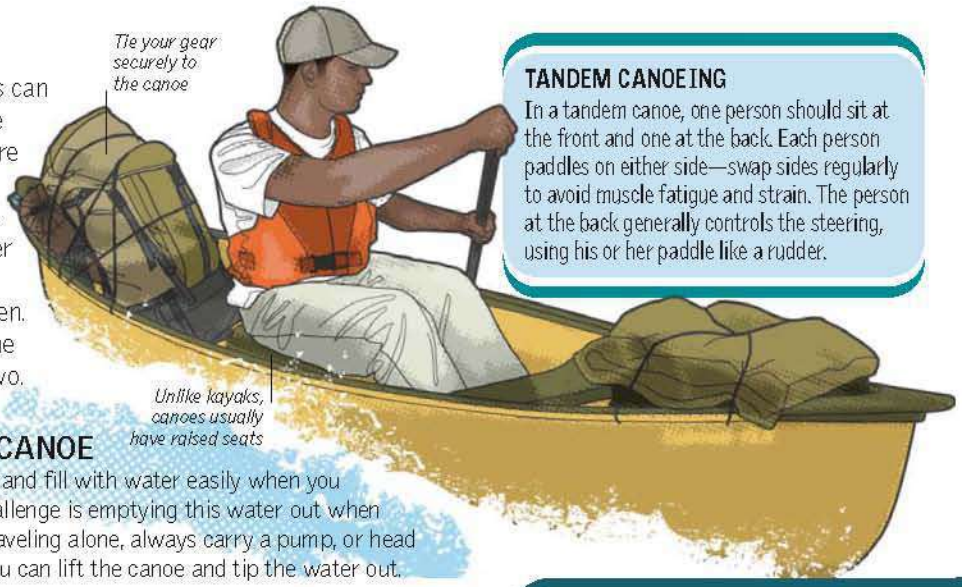

TANDEM CANOEING

In a tandem canoe, one person should sit at the front and one at the back. Each person paddles on either side—swap sides regularly to avoid muscle fatigue and strain. The person at the back generally controls the steering, using his or her paddle like a rudder.

CAPSIZING IN A CANOE

Because canoes are open and fill with water easily when you capsize, your greatest challenge is emptying this water out when the canoe is righted. If traveling alone, always carry a pump, or head for shallower water so you can lift the canoe and tip the water out.

CANOE-OVER-CANOE RESCUE

If you are traveling with another canoe and are far from shallow water, the best option is the canoe-over-canoe rescue. The capsized canoe is rotated upside-down and pushed onto the rescue vessel, which empties the water out. The people on the rescue craft then rotate the canoe back to an upright position and return it to the water.

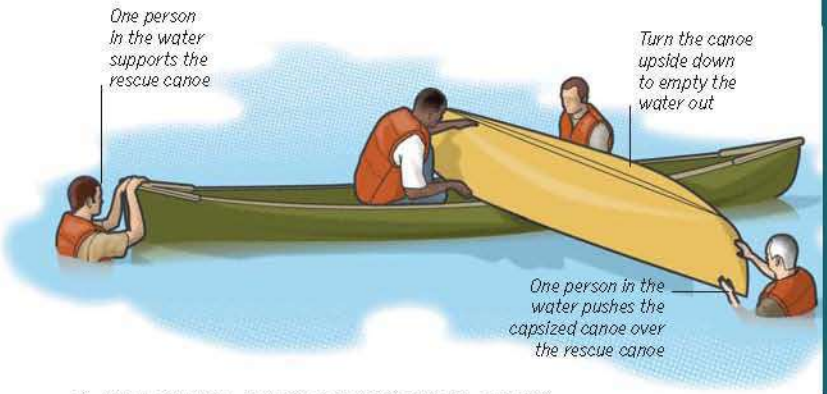

KAYAKING AND CANOEING TIPS

- Research your route and weather conditions thoroughly.
- Don't overload the boat, and ensure it's well balanced.
- Carry a high-volume pump so you can rescue yourself when alone.
- If watertight bulkheads are not fitted on your kayak, use airbags to add buoyancy and reduce the amount of water that can enter a compartment.
- Use waterproof dry bags to keep your kit dry.
- Clip valuables onto the boat or yourself. If they don't float, attach a float to them so they won't sink if they fall in the water.
- Carry waterproof material and bungee cord to repair blown hatches.
- Never enter a cave alone, or without head protection.
- Fit a leash to your paddle and attach it to the craft. If you then have to perform a rescue you can throw it in the water and keep both arms free.
- Chewing gum is great for the short-term repair of small holes in the craft, as is black masking tape and plumber's repair tape.

BASIC PADDLING SKILLS

When canoeing solo, use a "switch" style of paddling to ensure your canoe travels in a straight line—if the canoe starts to veer off course, switch sides.

CATCH

Sit in the middle of the canoe, ensuring that it's well balanced. Reach forward and thrust the blade into the water.

DOWN STROKE

Drive the paddle swiftly and firmly down into the water so that it assumes a vertical position.

PULL STROKE

Firmly pull the paddle blade backward through the water. This will propel the canoe forward.

REPEAT STROKE

Lift the blade out of the water toward the rear of the canoe and return to the starting position to repeat the stroke.

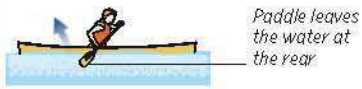

BUILDING A RAFT

IF VENTURING INTO the wilderness, it is useful to know how to build an improvised raft or flotation aid. You may find that a major water obstacle lies between you and rescue, or you may be in an area in which dense undergrowth makes cross-country travel difficult but river travel relatively easy. However, in a survival situation you are unlikely to have a lifejacket, so assess the risks carefully.

WARNING!
Apart from the brushwood raft (below), most improvised rafts will float half-submerged, so you'll be constantly sitting or kneeling in water. This could lead to hypothermia in certain conditions, so build an additional raised platform if necessary.

TOOLS AND MATERIALS
- Knife or small saw
- Poncho, shelter sheet, or tarpaulin
- Long sticks, brushwood, and thatch
- Cordage

MAKING A BRUSHWOOD RAFT

If you have a poncho, shelter sheet, or tarpaulin, you can construct a brushwood raft. This is a one-man raft, but if built correctly will keep you completely out of the water. Build it as close to the water as possible so you don't have to carry it far.

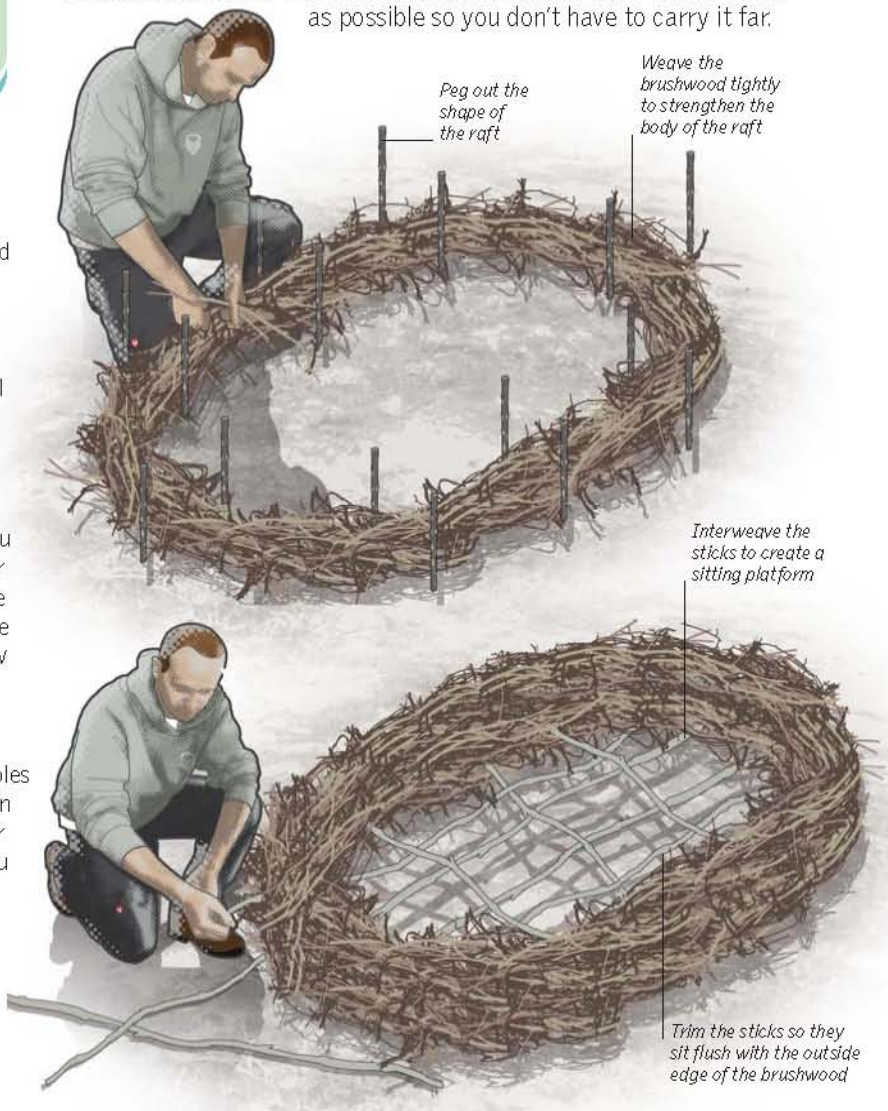

1 Construct your raft near a suitable launching place.
- Lay your poncho on the ground to determine the size of the finished raft. Make sure there's enough material to come up the sides and gather on top.

2 Peg out sticks to form an oval shape, the size and shape of your intended raft.
- Interlace brushwood and thatch tightly through the sticks to form the sides of the raft. The tighter you weave the brushwood, the stronger the body of the raft will be, and the better it will float. The height of the brushwood will also determine how high the raft sits in the water.

3 Form a sitting platform by pushing a selection of long poles or sticks through the sides of woven brushwood. Weave the sticks under and over each other as much as you can to create a stable structure.
- Using your knife or a small saw, trim the sticks to size so that they don't protrude from the sides.
- Remove the pegs from around the brushwood.

MAKING A PADDLE

In most cases, you can use the current of the river for momentum rather than having to paddle, and your walking staff to maneuver the raft. However, where there is no current you will have to provide the propulsion yourself, and will need to improvise some form of paddle.

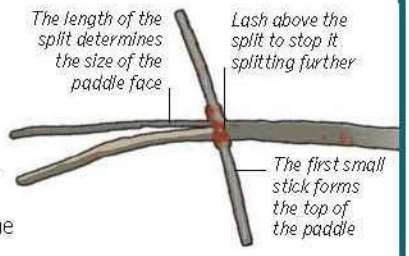

1 Find a suitable length of green wood to form the handle. This should be as wide as possible but still comfortable to grip.

- In the end of the wood, make a split long enough to form the paddle face.
- Collect smaller branches and insert one of them into the split. Lash into place.

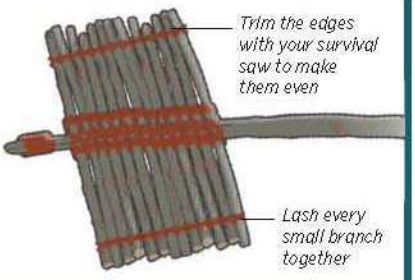

2 Continue to force smaller branches into the split, lashing each one as you progress until you have a sufficient paddle area.

- Securely lash the two open splits of the handle together.
- To make the paddle more rigid, lash the ends together.

4 Slide your poncho underneath the raft to form a waterproof seal.

- Push the poncho hood through to the inside and tie its neck securely to ensure it's completely water tight.
- Pull the poncho up the sides of the raft and lash securely across the top.

5 Fill the space underneath the sitting platform with natural materials, such as additional brushwood, grass, or moss. You could also use any item that would aid buoyancy—such as empty plastic bottles or waterproof bags filled with air and tied off.

- Drag or push the raft into the water, checking to see whether it floats in shallower water before loading your equipment and climbing aboard.

ALTERNATIVE RAFTS

Use whatever materials are available to you to make your raft. Look out for logs, bamboo, and discarded oil drums, as they are naturally buoyant materials that can be used to make an effective raft.

LOG RAFT

To make a log raft, use dry wood—ideally standing deadwood—as this will float higher in the water. Cut notches in the logs to allow the cross-pieces to sit snugly.

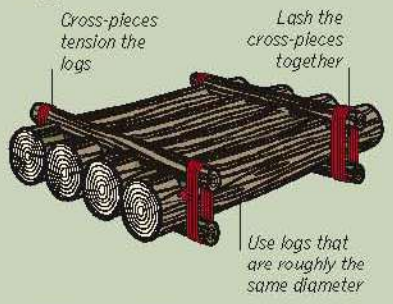

BAMBOO RAFT

Bamboo is made up of hollow compartmentalized sections, which means it is ideal for raft-making. Bamboo rafts are much lighter than log rafts.

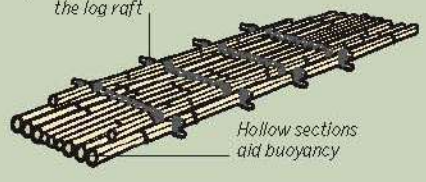

DRUM RAFT

If oil drums are available, they make ideal rafts. Always take care when handling chemical drums, as they may once have contained toxins.

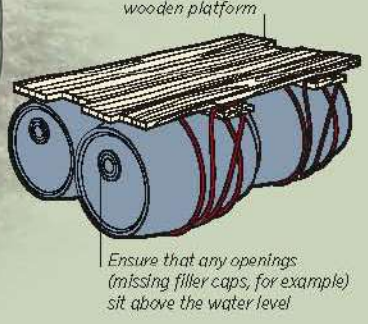

SWIMMING

WHEN THE WATER IS TOO DEEP to wade and you don't have the materials to build a raft, you may need to swim to negotiate a water obstacle. Before you get into the water, plan how you will get warm and dry when you get out. Making a float will help to keep clothing and equipment dry.

WARNING!
As with any survival activity involving water, try to avoid getting wet—find a way around the obstacle if possible, or find shallower water to wade across. The cold water increases your risk of hypothermia so ensure you have the necessary equipment for building a fire. Never enter water where there are likely to be dangers such as crocodiles or hippos.

SURVIVAL SWIMMING

Unless you need to cross rocky ground to enter the water, remove your boots and most of your clothing to reduce drag and keep them dry. Place them either in a waterproof survival bag or in an improvised poncho float (see opposite) and float them across.

SWIMMING ACROSS RIVERS

Choose a safe place to cross (see p. 88), and plan where you will exit the water.

- Take the current into account as you are likely to drift downstream slightly.
- Lower yourself in gently—never jump or dive into the water.
- If the water is cold, gently immerse yourself until your body has recovered from the initial shock.

THE CROSSING

Choosing a stroke you feel comfortable with (and that will expend the least energy), swim across your route with your float in one arm.

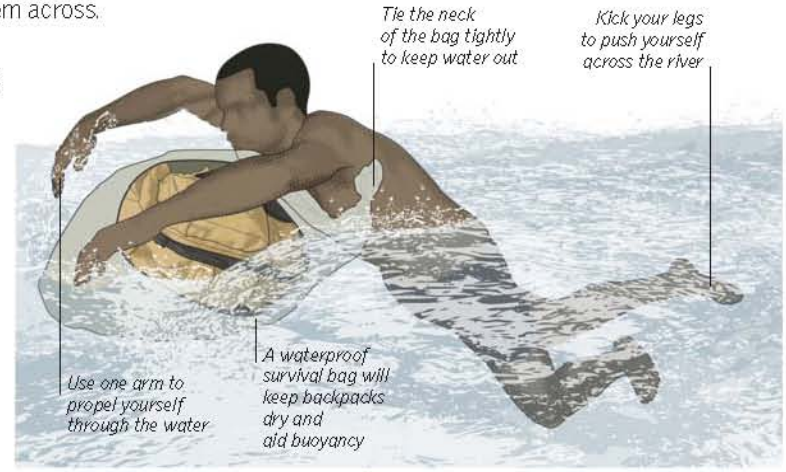

SWIMMING DOWN RIVERS

Avoid swimming up or down rivers—walking the route instead is far safer. However, if you accidentally fall in and find yourself traveling downstream in fast-moving water, there are ways of protecting yourself until you reach safety.

DEFENSIVE SWIMMING

The aim of defensive swimming is to adopt a position that keeps you as safe as possible and prevents your feet from becoming ensnared in rocks, which, in a strong current, could push you under the water. Assume the defensive swimming position until you reach shallower water, and can stand up and climb onto the bank.

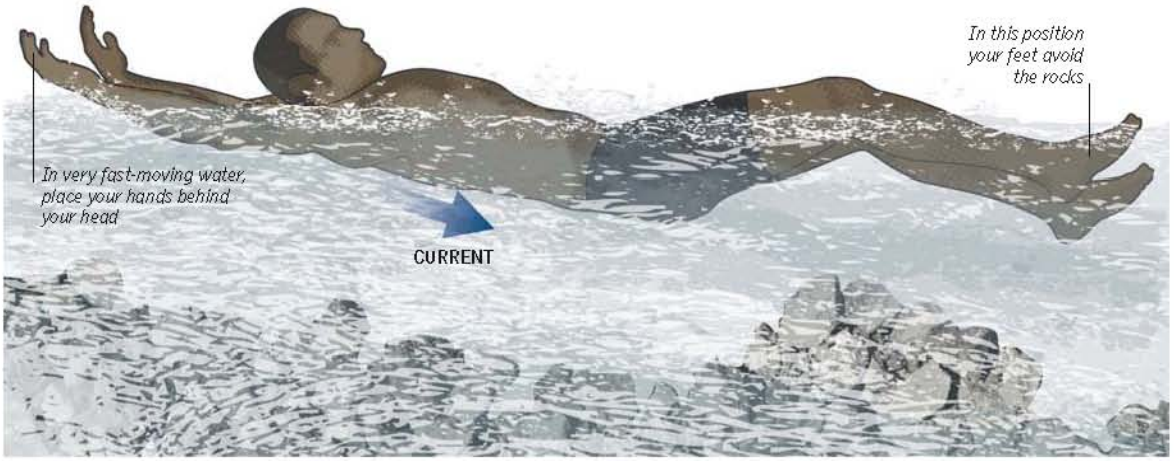

SWIMMING STROKES

It's important to know what the different swimming strokes are, although you're unlikely to stick to them strictly in a survival situation. If you're a confident swimmer, always choose the stroke that will use the least amount of energy, such as breaststroke. If you're less experienced you should avoid getting in the water unless in a life-or-death situation.

DOG PADDLE

An easy stroke suitable for less confident swimmers, this involves simply paddling forward with your arms and legs. When you paddle with your left arm, kick with your right leg and vice versa.

FRONT CRAWL

Also known as "freestyle," front crawl uses a lot of energy so is not the best stroke for a normal survival situation. However, you can build up a lot of speed, which may sometimes be required.

BREAST STROKE

This is perhaps the most common stroke for crossing rivers. Your arm and leg movements should be made simultaneously and smoothly. It is a good option if you're pulling a float.

BACK STROKE

If you know the water is safe, this stroke can be a good choice, as it uses little energy. However, not being able to see where you're going can be a disadvantage, especially in rocky areas.

MAKING A PONCHO FLOAT

Constructing a poncho float will keep your belongings dry and provide limited flotation as you cross the river. If you don't have a poncho, use any large piece of waterproof material.

1 Push your poncho hood to the inner side, and tightly secure its neck with the drawstrings or cordage.

- Lay the poncho on the ground with the inner side facing upward.
- Place your gear on top.

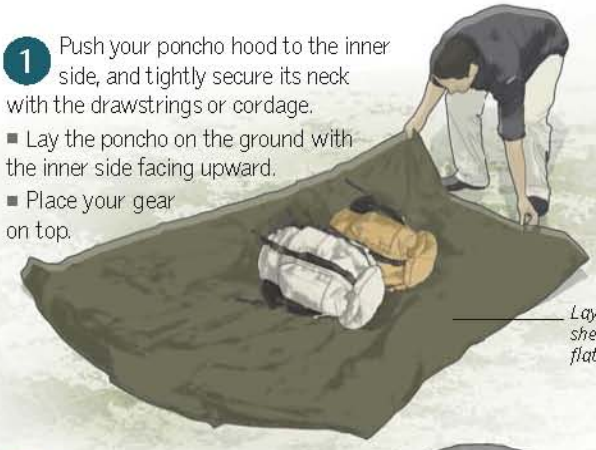

2 Bring one side of the poncho up and over the equipment that is piled together in the middle.

- Repeat with the other three sides to create a rectangular parcel.
- Fold the corners, and ensure that they won't allow water to enter the float.

3 Wrap carefully to make it watertight.

- At this point, if you have a second poncho, repeat from step one, placing the float face down on top of the second sheet and wrapping again.
- If available, place brushwood inside the float for added buoyancy.

4 Tie ropes, bootlaces, vines, or improvised cordage around the float to secure it.

- As you enter the water, gently lower the float in with you. Pull it along as you cross the river.

CAMP CRAFT

CAMP ESSENTIALS

Whether you intend to remain where you are and wait for rescue, move to a safer area and await rescue there, or walk out of the survival situation yourself, you're probably going to need to select a site on which to put up a shelter. This could be for just a single night or for a longer-term stay.

A sound understanding of what constitutes a good location will allow you to address the basic principles of survival. For example, a well-selected site will help keep you safe from danger while still allowing you to deploy your location aids effectively. It should also provide you with adequate materials for building a shelter and a fire, and offer an accessible supply of water, both for drinking and for hygiene.

A well-organized site will not only give you a sense of purpose and order, but will also provide a safe environment for yourself and your equipment. Designating an area for storing equipment and tools, for example, will help prevent vital items from being lost and will reduce the likelihood of you or members of your team being injured.

In this section YOU WILL DISCOVER...

- how to **make a shower** before you **take a shower...**
- why **keeping clean** is good for both **body and soul...**
- how to work up a **lather** with **wood ash...**
- when a **thatched screen** could **protect your modesty...**
- what turns a **hard candy** into a **firestarter...**
- the **difference** between **cattails** and **fire dogs...**

A fire is an integral part of any campsite. It can be used for warmth, for purifying water, for cooking, for signaling to potential rescuers, for protection against wild animals, and for providing light when darkness falls. It also provides a sense of security. The psychological effects of being able to start a fire in a survival situation should never be underestimated; neither should the psychological effects of *not* being able to start one. Even in a survival situation, a simple camp can give a sense of normality and "home."

You can produce a spark—and therefore light a fire—by using a device such as a firesteel or a ranger's flint and steel (see p. 127).

A FIRESTEEL COMPRISES two main parts: the material that will produce the spark (usually a rod made from ferrocerium or magnesium alloy) and a sharp striker device (usually a knife blade or short piece of hacksaw blade). When the striker is drawn over the rod, a spark is produced.

To control the spark created by these two moving parts, follow the method outlined below. It will allow you to direct the spark accurately, and reduce the chance of you knocking and scattering your tinder. This can be a problem if you hold the rod next to the tinder and strike down into it—especially if you are cold, wet, and tired and your hands are shaking.

1. Place the rod in the center of the tinder. Then position your striker onto the rod. Lock the hand holding the striker into position.

2. Pull the rod up and away from the tinder, drawing it against the striker. By drawing away from the tinder you avoid the danger of disrupting or scattering it.

3. To direct the sparks, alter the angle at which you pull the rod up and away.

> The **psychological effects** of being able to **start a fire** should not be underestimated; neither should the effects of **not being able** to start one

ORGANIZING YOUR SITE

WHERE YOU CHOOSE to set up your shelter depends on the environment, but always take into account the four principles of survival: protection, location, water, and food. First, make sure there are no obvious dangers, and that you're able to signal for rescue. If possible, select a location close to a water source.

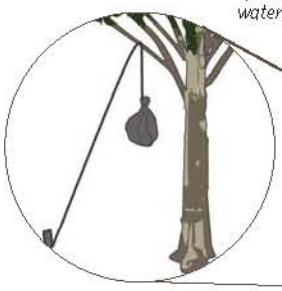

Select a place for a latrine that's downwind of the camp and downstream from the water source (see p. 117)

Pitch your camp near a source of wood for shelter and for fuel

If predators are in your area, hang unused food in a bag 10 ft (3 m) off the ground and 3 ft (1 m) away from the trunk or branches of a tree

SITE FEATURES

Once you're happy with your chosen site—and confident it doesn't hold any forseeable dangers—you can organize the area to accommodate your needs, and to make your time in camp easier and more efficient.

CAMP ADMINISTRATION

Whether you're in a group or on your own, it's important to organize your site, and quickly establish disciplines and routines to ensure camp safety and to reduce the risk of accidents. Designate specific areas—for storing equipment, firewood, cooking, and sleeping—and specific routines for when you do things (see panel, right).

When you take off your boots, prop them upside down on a stick to ensure nothing crawls inside. Never put them too close to a fire to dry

Assign an open area for location aids, such as a large "V", and as a potential landing site for a rescue helicopter (see pp. 236–41)

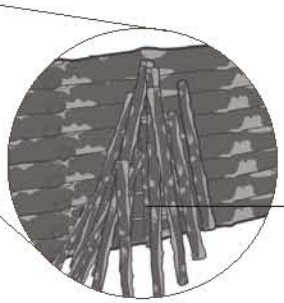

Collect and cut the fuel you need before it gets dark and keep it dry in an upright stack

Create a safe area for cutting wood. Use a tree stump as a platform for cutting and chopping

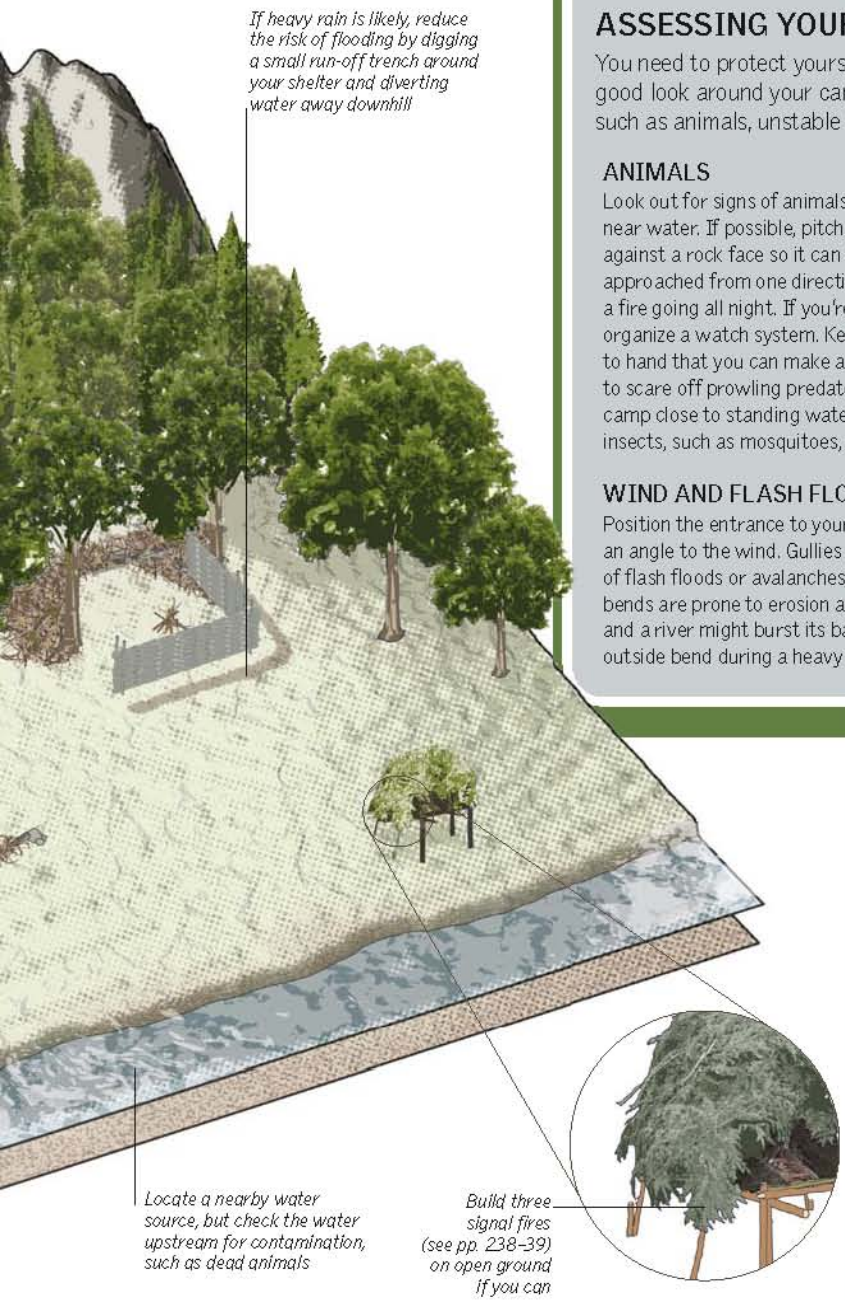

ASSESSING YOUR ENVIRONMENT

You need to protect yourself from injury, threats, and risks, so have a good look around your campsite and assess it for any potential dangers, such as animals, unstable rocks or trees, and the likelihood of flooding.

ANIMALS

Look out for signs of animals, especially near water. If possible, pitch your camp against a rock face so it can only be approached from one direction. Keep a fire going all night. If you're in a group, organize a watch system. Keep things to hand that you can make a noise with to scare off prowling predators. Don't camp close to standing water where insects, such as mosquitoes, swarm.

WIND AND FLASH FLOODS

Position the entrance to your shelter at an angle to the wind. Gullies run the risk of flash floods or avalanches; inside river bends are prone to erosion and floods; and a river might burst its banks on an outside bend during a heavy downpour.

STANDING DEADWOOD

These are dead trees that haven't yet fallen. Heavy wind, or the weight of rain or snow, can make them fall. This is the best type of wood for kindling and fire fuel (see p. 121).

DEADFALLS

These are dangerous branches that have broken off a tree, but haven't yet fallen to the ground. Some trees, such as beech, ash, and yew, drop their branches without warning.

ROCKFALLS AND ICE FALLS

If you camp next to rocks, check for cracks and fissures. Fires below them can cause rockfalls. When it's cold, ice sheets can suddenly fall from the rocks.

WHERE NOT TO CAMP

- Don't set up your shelter on sloping, poorly drained ground.
- Don't pitch your camp too close to water because of the risk of flooding and the presence of insects and animals.
- Avoid camping near noisy water sources, such as waterfalls, because they mask any other sounds that could alert you to the presence of threats, such as wild animals, or even possible rescue, such as a helicopter or emergency whistles.

CAMP TIMETABLE

In most cases, when you arrive at an area in which you intend to remain for a while (known as "going static"), start planning your activities by working back from a cut-off time—usually when it gets dark.

THREE HOURS UNTIL DARK

You arrive in the area. Drink water, stow what is no longer needed (map, compass, and so on), and change from wet clothing, but keep one set of clothes dry at all times. Scout the area for the most suitable campsite.

THEN...

Start pitching your camp. Build your shelter, including bedding. Gather tinder, kindling, and fuel and make a safe firebase and reflector. Collect water and forage for food. Prepare all aids to location (for both night and day).

ONE HOUR UNTIL DARK

During twilight, finish organizing the site and see to your personal needs—wash, use the latrine, and check your equipment. If you're in a group, make sure everyone knows where the emergency equipment is and who has been designated each task (lighting the signal fire, shining the flashlight, blowing the whistle). Set up a watch system for the fire.

STAYING CLEAN

PERSONAL HYGIENE IN THE FIELD is an important element of protection. Keeping yourself clean and healthy helps to ensure that your body is working at its most efficient and reduces the risk of illness. How you feel physically also has a direct impact on how you feel psychologically. It's a state of mind: if you let your personal hygiene slip, it's just a matter of time before everything else starts to follow.

CLEAN CLOTHING

The condition of your clothing and equipment can affect your state of mind. If your clothing is dirty and unkempt, your attitude to survival will lack discipline. Try to keep your clothing clean and in good repair.

HYGIENE CHECKLIST

Personal hygiene is about keeping clean and healthy, so develop a daily routine that ensures you correctly use personal protection aids (tablets, insect repellents, sunblock), and safely handle food, water, and cooking and eating utensils. Keep your clothes clean, and attend to your bodily functions (see opposite).

HAIR

You don't need to shampoo your hair—let it make its own oils and minerals and establish a natural balance. Wash out incidental smells, such as fire smoke, with hot or cold water.

SCALP

Check your head for insects and bites every morning and night. Local knowledge will help—for example, look for ticks if you are in a wooded area populated by deer.

EYES

Wear sunglasses or a hat to protect your eyes from bright sunlight and snow glare. Rinse your eyes with water twice a day to protect against infections like conjunctivitis.

EARS

Carefully check your ears for foreign objects with a clean, wet finger, especially if you're sleeping rough on the ground.

TEETH AND GUMS

Use clean fingers to rub your teeth and gums, or make a toothpaste from baking soda, or a mouthwash from salt and water.

BODY

Every two days wash your armpits, crotch, hands, feet, and toes with running water to keep fungal infections at bay.

MAKING A SHOWER

Standing under a shower can work wonders for your spirit—it removes accumulated dirt and sweat and makes your body feel refreshed. An improvised shower is quick and easy to make from a metal or a plastic container.

1 Turn the container upside down on a flat surface.
- Punch holes in the bottom with the bradawl on your penknife or the point of your knife.

2 Make a hole about 1 in (2.5 cm) down from the rim of the container.
- Make a matching hole on the opposite side of the container.

3 Smooth out any rough edges around the two holes.
- Thread a piece of cordage more than 2 ft (60 cm) long through the two holes.

4 Pull the cord through until you have the same length—about 1 ft (30 cm)—on either side.
- Tie the two strands together with an overhand knot (see p. 143).

BODILY FUNCTIONS

In a genuine survival situation, you don't eat as much food as usual and your toilet functions—particularly with regards to solid waste—will reduce dramatically after a day or so. Nevertheless, it's important to maintain a routine, especially if you are in a group.

IN TRANSIT

If you're moving every day or so, it's not worth making a latrine, so just attend to your bodily functions as follows:

- **Urination:** choose a tree away from the water source and downwind of the camp.
- **Defecation:** dig a hole at the base of a tree, fill the hole afterward, and mark it with stones or two crossed sticks. Clean yourself (but don't use the hand you eat with) with toilet paper, leaves, or grass, then running water. Wash your hands and fingernails. Burn or bury used toilet paper.
- **Menstruation:** if you don't have tampons or sanitary napkins with you, use something cotton and washable, or even sphagnum moss. Burn or bury whatever you have used.

GOING STATIC

If you're going static by staying in one place for more than a few days, build a latrine downwind of your camp, and downhill and away from your water source. Dig a deep trench, either next to or between two trees, and make a seat from two poles. Cover the deposits with sand or soil to reduce the smell and keep flies away. When you leave, dismantle the latrine, fill the trench, and mark the area with stones or crossed sticks.

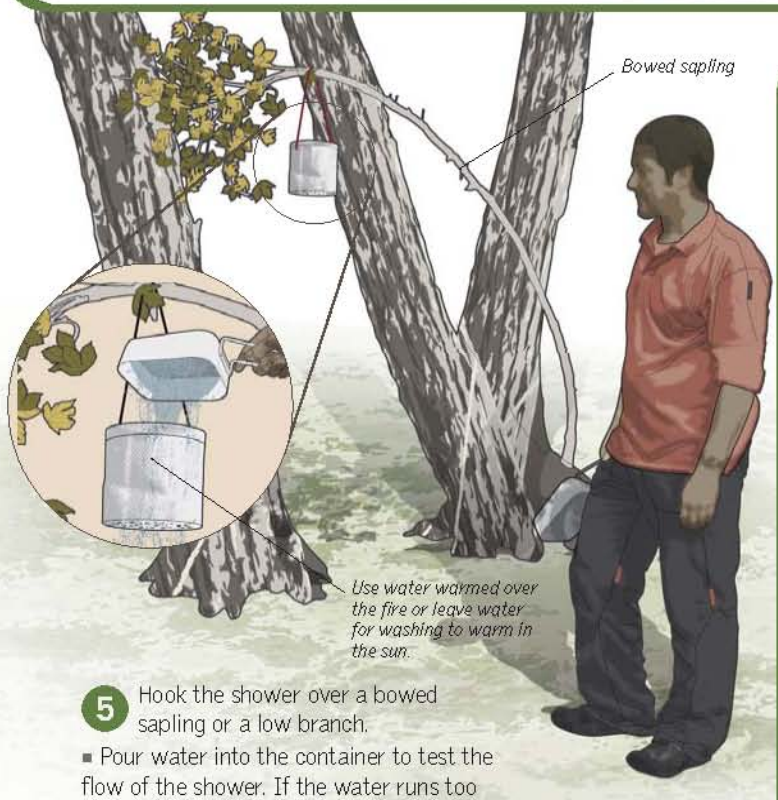

5 Hook the shower over a bowed sapling or a low branch.

- Pour water into the container to test the flow of the shower. If the water runs too quickly, put leaves in the bottom to slow it.

LATHERING UP

Camping soap is a concentrated antibacterial liquid soap that can be used without water. Decant a little into a small container, such as an old 35mm film container, and it will last for weeks. Alternatively, you can make soap from natural materials.

MAKING SOAP

You can make natural soap from various sources that contain a substance called saponin, which —when mixed with water—has a cleaning effect.

- **Birch leaves:** Select young leaves and place them in a container (even a plastic bag will do). Add some cold water, then add some hot or boiling water (whichever your container can handle without melting). Agitate the mixture—this allows the saponin in the leaves to dissolve into a natural soap.
- **Soapwort:** Agitate soapwort roots in water until they foam up. Let the light foam settle before using it to wash yourself and your clothes.
- **Horse chestnut:** Dip horse chestnut leaves in warm or hot water then remove. Squeeze the leaves in your hand to produce saponin.
- **Wood ash:** Mix wood ash with water. Don't use this method too often, as it can dry out your skin.

MAKING FIRE

THE ABILITY TO MAKE and maintain a fire can be a significant psychological factor between determining whether you do all you can to survive or just give up. Fire gives you a sense of "being," and, like a shelter, it can transform a clearing under a tree into a "home."

A PORTABLE KIT

It's important to gather all your materials before you start—not only your tinder, kindling, and fuel, but also your means of lighting a fire. This involves a fair amount of organization, so a portable kit that keeps it all in one place is much more convenient.

FIRE PRECAUTIONS

If you intend to have a fire in front of your shelter, bear in mind some basic rules:

- Don't build the fire too close to the shelter as it may get out of control or a spark may blow into the shelter and ignite it.
- Make sure you have fully extinguished your fire before you decamp and leave the site. Pour water (if you have enough) onto the fire and the immediate surrounding area, or use damp soil, sand, or earth.

COMPONENTS OF FIRE

Three crucial components—oxygen, heat, and fuel—need to be present in order for a fire to start and be sustained. Although you don't have any control over the resulting chemical reaction, known as combustion, you can develop the skill of lighting a fire. The key is to achieve the best balance between the components. If your fire isn't going well, get back to basics and ask yourself which one of the components is working against you.

OXYGEN

Although oxygen is vital for combustion to happen, it's all too easy to prevent it from getting to the flame by smothering the fire with too much wood. If your initial fire looks as though it's dying, try fanning it with your hand or a map to create a draft that feeds oxygen to the fire.

HEAT

Heat is essential for igniting the fuel. In most cases this heat can be generated either by a spark (such as flint and steel), by a chemical reaction and friction (such as matches or potassium permanganate), or by friction alone (such as a bow drill). (See pp. 126–27 and pp. 130–33.)

FUEL

Once the fire gets going, you need to have fuel to burn (see pp. 120–21). You should start with small, dry pieces of fuel that will catch the flame and generate enough heat to then burn increasingly large pieces of fuel.

MATCHLESS FIRESET TIN

Working on the principle of "Don't make things any harder than they need be," this matchless fireset, designed for the military, contains a one-stop solution to getting a fire going—regardless of the weather conditions.

INSIDE THE TIN

The small tin, which is waterproof when taped, contains everything you need to produce a fire: a sparking device, tinder, kindling, and fuel. Hexamine from the fuel blocks is scraped onto cotton wool, then a spark from the flint and steel ignites the cotton wool long enough for the fuel blocks to light.

USING THE TIN

You can use the tin itself to boil water. Raise the tin 1 in (2.5 cm) off the ground with two sticks and light a quarter of a hexamine fuel block underneath. After a while, the water boils so you can have a potentially life-saving drink.

PREPARING BEFOREHAND

More than any other survival task, starting a fire needs good prior preparation. If you don't have the correct materials and in sufficient quantities, you will probably fail in your attempts to start a fire. Preparing the ground, your materials, and your equipment will usually make firelighting much easier and more likely to succeed.

CHOOSING YOUR GROUND

You need to be careful when selecting the place in which you intend to start your fire. Clear the ground before building your fire. Never light a fire directly on the ground and watch that your fire doesn't spread or burn out of control.

KEY POINTS TO REMEMBER

- Carry some form of firelighter with you at all times.
- Practice your skills before you need them—in different conditions and using different materials. You'll soon learn what works and what doesn't.
- Collect tinder as soon as you enter the forest. If it's wet, dry it in your pockets against your warm body.
- Collect everything you need to get a fire going, then multiply it by ten. If you're in a dire situation, you may have only one or two chances of getting a fire going before you either run out of matches or tinder, or you suffer from exhaustion or hypothermia.

CAMP FIRE DO'S

- Make sure you have enough wood close by to fuel your fire. Carrying wood to your fire uses energy, which may be at a premium in a survival situation.
- Rake the area around the spot where you intend to start the fire and brush away leaves or anything else that may ignite and start a forest fire.
- Check the ground for tree roots. Your fire could set an exposed root alight, or even one just under the surface. Once a root starts to burn underground, the heat works its way along the root and could start a forest fire.
- If you're trying to get rescued, choose a spot for the fire where passing vehicles and aircraft can see it.

CAMP FIRE DON'TS

- Don't use your hands as a rake to clear the ground, because you may get bitten by an insect or a snake. Use your feet or a branch instead.
- Don't build a fire next to old logs or fallen trees, as they may catch fire. They may not look alight but they can smolder for several days and a heavy breeze could subsequently fan the embers and start a blaze.
- Don't build your fire under an overhanging branch, or leaves, as the heat from the fire can soon dry them out and then set them alight.
- Don't position your fire so that the wind affects the way the fire burns, or blows smoke or flames into your shelter.

THE ELEMENTS OF FIRE

THE THREE MATERIAL ELEMENTS you need to build a fire are tinder, kindling, and fuel. They must be dry and collected in sufficient quantities. A well-made feather stick effectively provides tinder, kindling, and fuel on one piece of wood—and it needs only a spark to set it alight.

EMERGENCY TINDER

Carry tinder, such as a tampon or cotton wool balls, in a 35mm film canister or a Ziploc bag in your emergency kit. For each fire, use some cotton wool and a small piece of tampon. Cotton wool balls smeared with petroleum jelly make a flame last ten times longer.

MAKING A FEATHER STICK

Between four and six well-made feather sticks provide enough combustible material to get a fire going. Keep a few already made and packed away to use in an emergency. The best feather sticks are made from standing dead wood, but just about any type of dry wood will do. If you use small, dead branches that have snapped off a tree, remove the bark first.

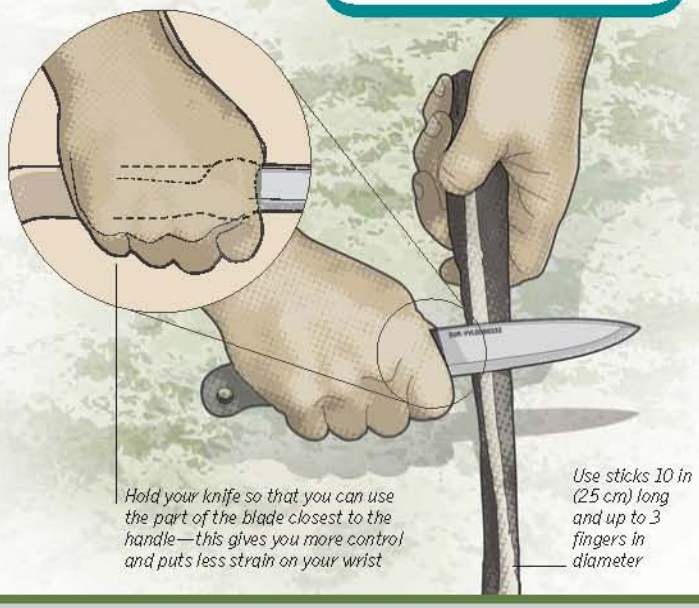

1 Choose the side of a stick with an even grain and no knots. Work on a hard, flat surface to stop the stick slipping.

- Lay the blade flat on the stick and run it all the way down without cutting into the wood—do this ten times. This gives you a feel for the wood and how your knife moves over it.

TINDER

The first element you use to make a fire is dry, combustible material called tinder. You may have tinder in your equipment (see pp. 122–23), or you may need to find natural or other man-made sources. The key to success is to experiment with what's around you before you actually need it. Make sure it's dry—leave it in the sun if it's damp. When you use some tinder, replace it at the earliest opportunity.

TYPES OF TINDER

Natural sources: Feather sticks (see above), shavings from the outside of a bamboo stem, bark shavings, fine wood dust, pine pitch, plant and animal down, powdered dung, and fire dog (a charred stick). Also, birch bark (silver birch bark can be lit even when it's wet), clematis, honeysuckle, cattail, dry grass, dead and dry moss, and some fungi. The best tinder for friction methods of firelighting (see pp. 130–33) is tree bark, dry grass, fungi, lichen, plant roots, fibers, and down.

Man-made sources: Cotton wool balls, tampons, char cloth, fire tin (see p. 123), lint, tissue paper, camera film, strips of rubber, and candle wick.

TINDER BUNDLE

A good way to prepare your tinder to take a spark or a coal is to make it into a tinder bundle. Vigorously tease, rub, and pull the fibers with your fingers until it becomes a ball the size of a grapefruit. Push the finest, most combustible material into the interior of the ball. Try mixing up your tinder to make it more effective—for example, dry grass, birch bark, and cattail.

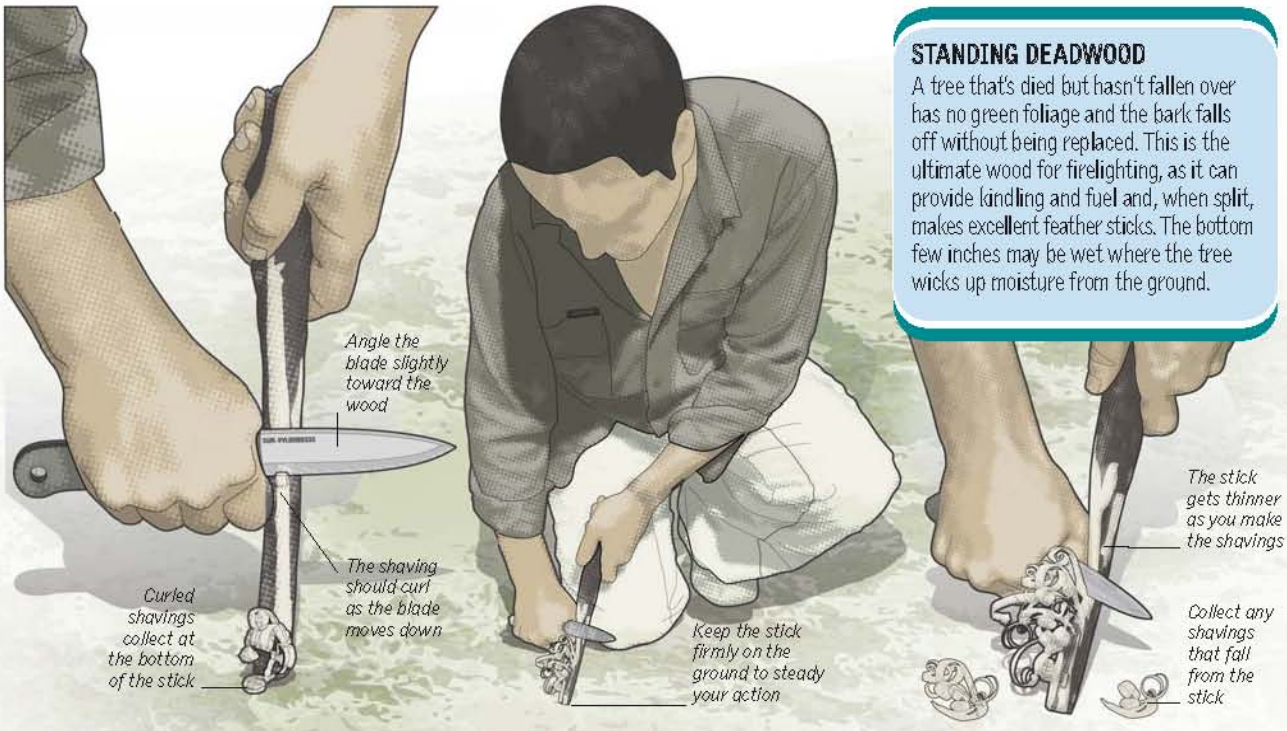

STANDING DEADWOOD

A tree that's died but hasn't fallen over has no green foliage and the bark falls off without being replaced. This is the ultimate wood for firelighting, as it can provide kindling and fuel and, when split, makes excellent feather sticks. The bottom few inches may be wet where the tree wicks up moisture from the ground.

2 Now tilt the angle of the blade slightly toward the wood.

- Keeping a part at the top of the wood to use as a handle, run the blade down the wood to cut a shaving.
- Stop just before the bottom so that the shaving stays attached to the stick.

3 Turn the stick slightly and run the knife down the edge created in step 2 to create a second shaving. Turn the stick again and repeat, working around the stick.

- Try to get a rhythm going, and put your body behind the cutting action.

4 Once you have mastered the technique, continue to feather the rest of the stick.

- When you've finished you will have a thin stick with curled shavings still attached, and ready for use as kindling.

KINDLING

Kindling is the second material element of firelighting and, when dry, is added to burning tinder. Kindling can be as thin as a match or as thick as a finger. You should be able to snap it with your hands. If it doesn't break with a crack then it's probably not completely dry. If it's damp, remove the outer bark (this retains most of the moisture) and break the kindling down into small sticks 6 in (15 cm) long.

TYPES OF KINDLING

- Soft-wood twigs are very combustible, while wood with flammable resin, such as pine, burns hot and long.
- You can use some types of tinder as kindling—for example, bark, palm leaves, pine needles, grass, ground lichens, and ferns—but you will need larger quantities for kindling than you needed for tinder.

FUEL

Initially, your fire needs constant tending, but once it can sustain itself for five minutes it's established and you can add increasingly larger fuel to create a good heart—a bed of hot coals that sustains the fire with minimum effort on your part. The fuel should be about as thick as your wrist or forearm. At first, use dry fuel split into sticks that catch alight easily. Add green (live) wood and larger logs (whole or split) later, once when the fire is established. In wet conditions, build your fire under cover to keep off the rain.

TYPES OF FUEL

- Hard woods from mainly deciduous trees (such as oak, maple, ash, beech, and birch) burn hot and long, produce good coals, and are more efficient. But they are hard to get going.
- Soft woods from conifers (such as pine, fir, and spruce) are easier to light as they can contain resin, but burn faster and produce less heat than hard woods. They also make more smoke.
- Peat is found on well-drained moorland and can be cut with a knife. However, it needs a good air supply.
- Charcoal is lightweight, smokeless, and burns hot.
- Dried animal droppings provide a good smokeless fuel.

CHAR CLOTH AND FIRE CAN

BEFORE SETTING OFF ON A TRIP, add either some char cloth or a fire can to your backpack. These excellent forms of tinder are reliable and potentially life-saving. Alternatively, you can take some mayasticks. Char cloth is very easy to make, so keep some dry in your emergency fire kit or lining the bottom of your survival kit (see pp. 60–61).

TOOLS AND MATERIALS
- A can with a tight-fitting lid, such as a small shoe-polish can
- Nail
- 100 percent cotton cloth
- Knife or scissors
- Spark or flame

MAKING CHAR CLOTH

Char cloth is cotton cloth that's been combusted in the absence of oxygen (pyrolysis). It's lightweight, takes up almost no space, and produces an ember from even a weak spark extremely well. Char cloth works only when it's completely dry, so keep it in a watertight container.

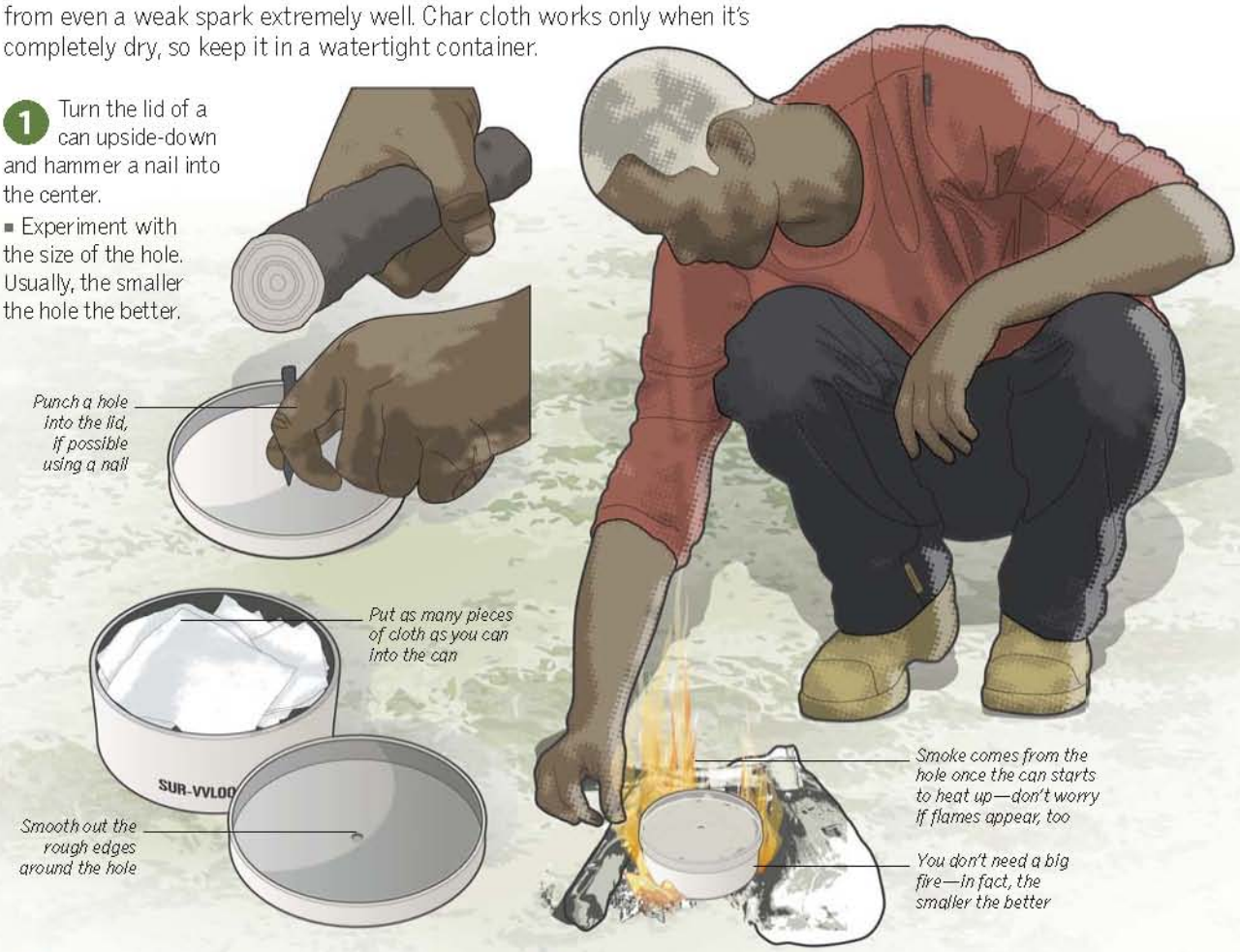

1 Turn the lid of a can upside-down and hammer a nail into the center.
- Experiment with the size of the hole. Usually, the smaller the hole the better.

2 Cut 100 percent cotton cloth into pieces that fit into the can without folding.
- Vary the sizes of the pieces so they don't just stack flatly on top of each other, but at the same time don't squash them in.
- Securely place the lid on the can.

3 Place the can on a fire to burn off all the oxygen inside the can, or place it on top of some good coals scraped to the side of your camp fire.
- When the smoke stops the process is complete.
- Safely remove the can from the fire.
- Don't open the can until it's cool.

MAKING A FIRE CAN

You can use a fire can to start a stubborn fire when conditions are less than perfect, or to boil water, do some basic cooking, or warm your hands on a cold day. Once lit, a fire can burns for hours with a concentrated, controllable flame that produces no smoke. When it starts to fail, you can either replenish the existing cardboard with more wax or replace the cardboard and start again.

TOOLS AND MATERIALS

- A can with a tight-fitting lid, such as a small shoe-polish tin
- Cardboard (ribbed or plain)
- Candle and a match or lighter
- Knife or scissors

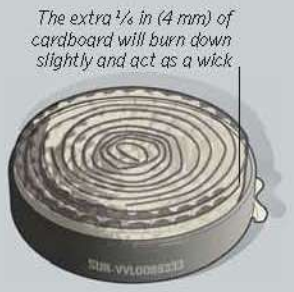

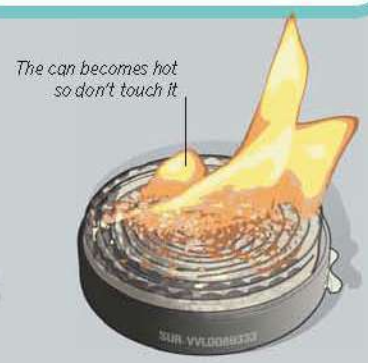

1 Cut out a long, thin piece of cardboard that's 1/6 in (4 mm) wider than the depth of your can.

- If it's ribbed, cut it across the grain.
- Roll it tightly along its length until the roll just fits inside the can.

2 Light your candle and let the melting wax drop into the can.

- Let the wax soak the cardboard and fill up the can—it's a slow process.
- Stop when the wax nears the top.
- Allow the wax to harden.

3 When the can has cooled down, hold it at an angle and light the top of the cardboard with the candle.

- The flame should be concentrated and spread across the top of the can.

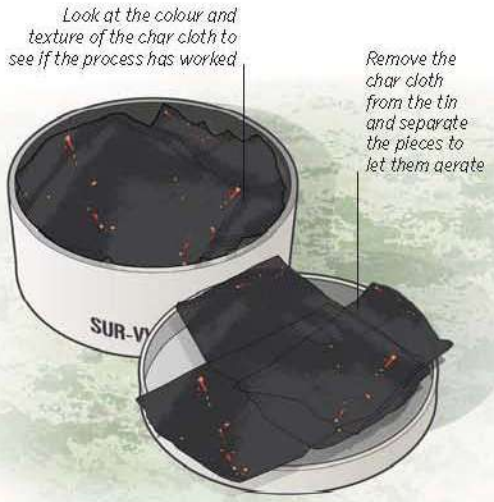

4 Once the tin has cooled down, remove the lid and examine the charred cloth.

- The cloth should be completely black and have a semi-rigid, but softish texture. If it's fawn or brown, put it back on the fire and leave for longer.
- Cloth that's brittle and crumbles is of no use.
- Check the cloth works by striking a spark on to a piece; the spark should create a small red ember.

FIRE WADS AND MAYASTICKS

Fire wads are rolled strips of newspaper soaked in a fuel. Once dry, they make handy waterproof tinder. Mayasticks are pieces of resinous fatwood from the highlands of Mexico and Guatemala. The pieces are easy to light, even when wet, and generate a hot flame.

MAKING FIRE WADS AND MAYASTICKS

- To make fire wads, tightly roll strips of newspaper into a tube 4 in (10 cm) long and tie them with cord. Soak them in melted wax and allow to dry. Light the center with a match.
- To light a mayastick, cut a little sliver of wood with your knife and lift it from the surface of the stick. When a little resin seeps out, set it alight with a flint and steel (see p. 127).

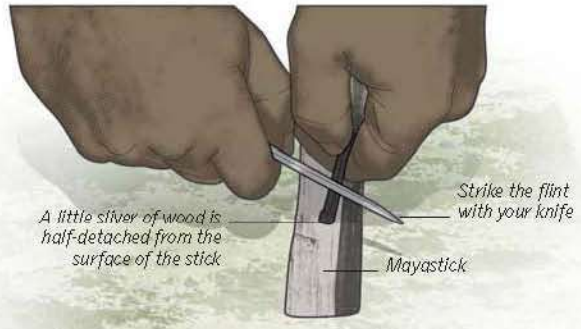

TYPES OF FIRE

ONCE YOU HAVE COLLECTED your tinder, kindling, and fuel, and made sure it's dry and ready at hand, you can set about making your fire. There are many different types of fire to choose from, depending on what you need the fire for (see below). Before you start to build your fire, make sure you have prepared the ground so that everything is safe (see pp. 118–19).

CHOOSING A FIRE

If you have a choice of fuel, as well as time to invest in building a specific type of fire, look at the choices below and determine which one is likely to suit your immediate needs.

- Your primary consideration should be the function of the fire: warmth is probably your most urgent requirement, but other uses include cooking, signaling (see pp. 238–39), drying wet clothing, and disposing of waste. You may want a particular type of fire because it lasts all night.
- Consider the availability of the components you need to build the fire—for example, the right fuel and the best ground. A rule of thumb is to estimate how much you think you'll need, then double it.

FIRE ESSENTIALS

There's plenty of advice surrounding the pros and cons of making fires, but three general tips will help:

- Never make things any harder than they need to be. Choose a fire that requires the least effort for the maximum gain.
- It is more efficient to build a small fire and sit close to it, than to build a large fire and sit far away from it.
- If all your wood appears wet, remove the bark and split the wood—the center will usually be dry.
- Once the fire is established, place damp tinder, kindling, and larger pieces of wood close to the fire to dry out. Keep watch on them so that they don't dry out to the point where they catch fire!

TYPE OF FIRE	TEEPEE	STONE-LINED	AUTOMATIC	LONG LOG
			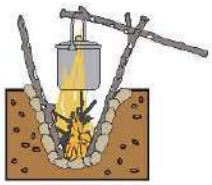	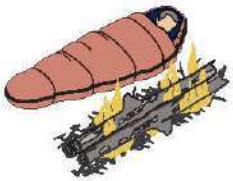
HOW	Surround the tinder ball with kindling in the shape of a teepee. Arrange small, medium, and large fuel logs in a square at the foot of the teepee.	Arrange large non-porous stones in a circle, place a tinder ball in the middle, with kindling around it. Add fuel logs when the fire is established.	Line a hole 3 ft (1 m) deep with non-porous stones. Put tinder and kindling inside and rest large logs against the sides so they drop down as they burn.	Put tinder, kindling, and fuel logs in a depression 6 ft (2 m) long. Lay two long fuel logs on top of the burning fire.
WHERE	Areas with abundant fuel, because the fire requires lots of firewood	Windy locations Popular areas with existing fire rings and where low-impact camping is vital	Rock-free earth or sand, because it's easy to dig	Forested areas, because the fire requires long logs for fuel
WHY	Quick to light Wet wood can be burned because it's dried by the heat of the inner fire	Stones shelter fire from wind Using an existing fire ring reduces the fire's impact on the environment	Self-feeding once lit, which means you don't need to keep adding fuel	Long-lasting (the fire can stay alight all night) Emits a great deal of heat
USES	Warmth Cooking Signaling (if green vegetation is burned)	Warmth Cooking Signaling (if green vegetation is burned)	Cooking Signaling (if green vegetation is burned)	Warmth (in cold weather, build a fire either side of you). Cooking (once embers have formed)

LIGHT YOUR FIRE

There are numerous ways to build a fire and get one going, and everyone has their favorite. The following example is a tried-and-trusted method that's versatile and works well in a variety of conditions.

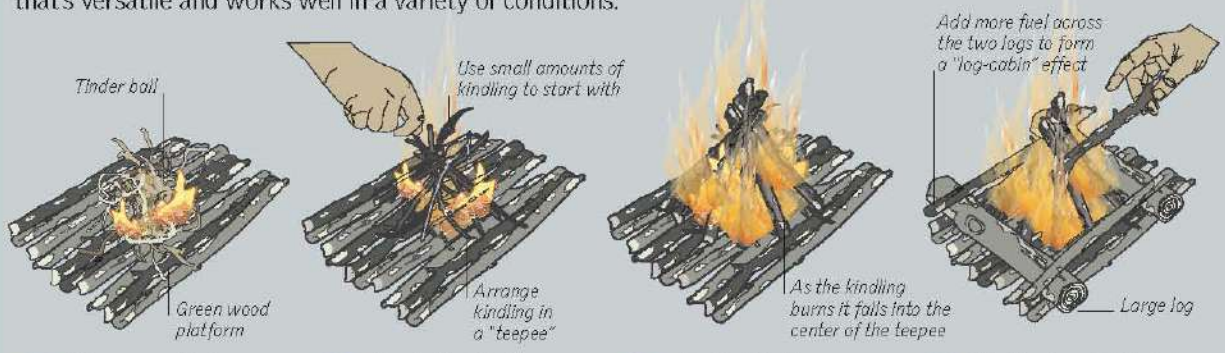

1 Place your tinder ball (see p. 120) on a platform composed of green wood.

- Light the tinder with your chosen method (see pp. 126–27) and let it catch.

2 Gently lay kindling by the tinder ball.

- Build a kindling "teepee", as this lets the fire breathe where the heat is strongest.
- Feather sticks also make ideal kindling (see p. 120).

3 As the kindling catches fire and the flames grow, add larger pieces of kindling.

- Continue to add kindling, gradually building up to split logs (see p. 148).

4 The fire is established when you can leave it alone for five minutes without it going out.

- Then lay a large log on the wind-facing side and one on the other side.

TYPE OF FIRE	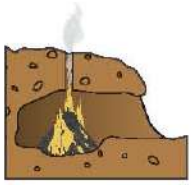			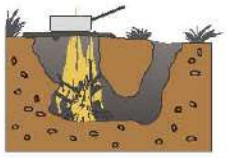
	SNAKE HOLE	STAR-SHAPED	HUNTER	DAKOTA HOLE
HOW	• Create a hole in the side of a bank and a chimney up through the earth. Light a fire inside using any fuel.	• Build a fire from tinder, kindling, and fuel. • Arrange four logs so they meet in the middle. • Push logs in as they burn.	• Make a fire out of tinder, kindling, and any type of fuel. Place two long logs either side of the fire in a V-shape.	• Dig a large hole for the fire and a slightly smaller hole for the chimney, with a tunnel linking the two. Use small logs as fuel and cook at ground level.
WHERE	• Windy locations • Rock-free earth or sand bank	• Forested areas, because of the logs the fire needs	• Cold or windy locations	• Anywhere you can dig a hole, because the fire requires very little fuel once established
WHY	• Chimney creates a draft, giving a high-temperature fire • Sheltered from the elements	• Long-lasting • Good embers for cooking	• Hardwood logs shelter the fire from the wind • Produces a great deal of heat	• Concentrated heat • Flames are below ground so fire is hidden
USES	• Warmth • Cooking • Waste disposal	• Cooking—balance a pot on the logs • Heating water	• Warmth • Cooking	• Cooking • Keeping warm • Drying wet clothing

MAKING SPARKS AND FLAMES

LIGHTING TINDER IS THE FIRST STAGE of making a fire. Matches and lighters do this instantly, but there are a number of ways to create a spark that you can then use to coax your tinder into a flame. This is relatively easy when everything is dry, but with patience and persistence it is also possible when conditions are bad.

IGNITION DEVICES

There are various ingenious methods of creating that all-important spark or flame to ignite your fire. If you don't have matches or a lighter, you'll need to use another device, such as a flint and steel. You can also improvise by using an external energy source, such as:

- Focusing the sun's heat with a magnifying glass or beverage can.
- Making sparks with a battery.
- Creating a chemical reaction using potassium permanganate.

STRIKING A MATCH

It may seem simple enough, but there's a way of striking a match—known as "commando style"—that reliably produces a flame in all kinds of conditions.

1. Hold the box in one hand and a match between the thumb and first two fingers of the other.
   - Strike the match firmly away from you.

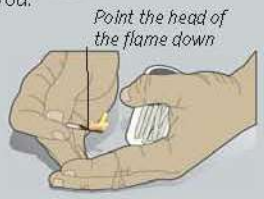

2. When the match lights, immediately cup your hands to protect the flame.
   - Let the flame burn a little way along the stem before using it to light the tinder.

TYPE OF DEVICE	MATCHES/LIGHTERS	MAGNIFYING GLASS	BEVERAGE CAN	BATTERY
			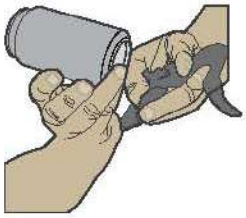	
GEAR	• Waterproof box of matches. • Cigarette lighter taped to a piece of cord around your neck. • Dry tinder	• Magnifying glass • Dry tinder	• Empty beverage can • Dry tinder	• Flashlight and its battery • Wire or wire wool • Dry tinder
WHAT TO DO	• To light a match "commando style," strike it away from you on the box and then cup it in your hands (see above).	• With the magnifying glass, focus bright sunlight onto some dry tinder and create a hot spot. Hold the magnifying glass steady until the tinder catches alight.	• Polish the bottom of the can (see Signal mirror, p. 245). • Catch bright sunlight on the shiny surface and reflect it onto tinder to create a hot spot. Hold the can steady until the tinder catches alight.	• Lay the wire across the battery terminals to create some sparks. • Remove the bulb from the flashlight and place wire wool over the terminal. Switch on the flashlight to create sparks.
COMMENTS	• Waterproof matches are usually just standard matches that have been coated with wax and varnish. • Always have a lighter around your neck on a piece of cord.	• When you choose a compass, make sure it has a magnifying glass incorporated into it (for reading the details on maps). • You can also use the lens in reading spectacles.	• You'll need to practice this technique so that you can rely on it in a survival situation.	• The thinner the wire the better this will work, especially with lower voltage batteries (1.5v). • Use this method for a short period only, otherwise you will drain the battery.

CREATING A CHEMICAL FLAME

Potassium permanganate is an extremely useful item to have in your survival kit, because you can use the chemical not only to sterilize water and wounds, but also to create a spark to light a fire. You will need some sugar to make it work, so use some from your survival rations or crush a hard candy.

WARNING!
Potassium permanganate is a strong oxidizer that can, when mixed with certain chemicals, create an explosive mixture. It can also stain your clothes and skin.

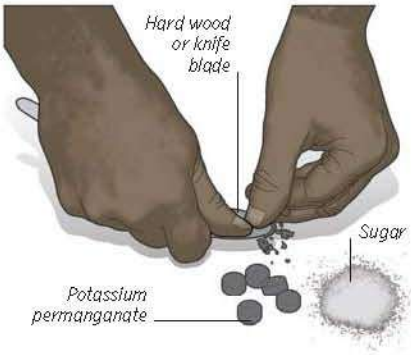

1. On a hard surface, such as a flat piece of wood or rock, crush some potassium permanganate. Have an equal amount of sugar ready.

2. Mix the ingredients. With the back of your knife, push down on the mixture and drag it along the hard surface to create friction.

3. Continue to push down and drag your blade until you get a spark.
   - The flame dies down quickly so have your tinder at the ready.

TYPE OF DEVICE	FIRESTEEL	FLINT AND STEEL	ONE-HANDED STRIKER	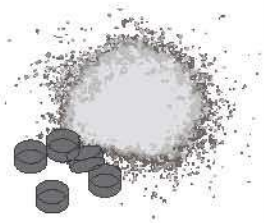 POTASSIUM PERMANGANATE
GEAR	▪ Ferrocerium firesteel ▪ Metal striker ▪ Dry tinder	▪ Magnesium-alloy rod ▪ Steel striker (usually a short piece of hacksaw blade) ▪ Dry tinder	▪ A magnesium-alloy rod and steel device that can be operated using only one hand ▪ Dry tinder	▪ Potassium permanganate ▪ Sugar ▪ Knife ▪ Dry tinder
WHAT TO DO	▪ Hold striker next to the tinder. ▪ Place the firesteel under the striker, directly on to the tinder. ▪ Draw the firesteel back across the striker to direct the sparks onto the tinder. The firesteel moves, not the striker.	▪ Follow instructions for the firesteel (left).	▪ Place the flint rod in the tinder. ▪ Press the thumb button onto the rod and push the handle down the length of the rod. ▪ Pushing harder creates greater friction, which leads to a more intense spark.	▪ Mix potassium permanganate and sugar in equal amounts on a hard surface. ▪ Press down on the mixture with your knife and drag the blade along to create a spark (see above).
COMMENTS	▪ Lasts for about 12,000 strikes. ▪ Temperature reaches 3,000°C. ▪ Works in all weathers and at any altitude.	▪ There are many variations and kits; one has a magnesium block for the holder and the striker shaves slivers of magnesium into the tinder to help with ignition.	▪ Works in all weathers. ▪ Has a safety feature that prevents accidental use. ▪ Designed for fighter pilots, who may injure an arm or hand after ejecting.	▪ You can also mix a small amount of potassium permanganate with glycol or antifreeze. Quickly wrap in paper and put on the ground. Warning: stand well clear as the combustion can be sudden and dangerous.

TRUE-LIFE ACCOUNT

EXTREME SURVIVAL—IN THE DESERT

USEFUL EQUIPMENT

- Personal Locator Beacon (PLB)
- Sunglasses and sunscreen
- Cotton scarf or hat
- Signal mirror and signal flares
- Map, compass, GPS
- Survival tin, bushcraft knife
- Cell/satellite phone
- Poncho/bivy sack

DESPITE A LACK OF SURVIVAL KNOWLEDGE, a party of five survived searing temperatures, dehydration, hunger, and the threat of animal attack for six days when their plane crashed in the Kalahari Desert in Botswana.

On Wednesday, March 1, 2000, Carl du Plessis and three associates—Mike and Lynette Nikolic, and Nebojea Graorac—set off on a business flight from the capital of Botswana, Gaborone, bound for Maun, a popular tourist town in the north of the country. However, the aircraft developed engine trouble during the flight and was forced to crash-land. Critically, radio contact was lost, so no distress call was emitted before the plane went down. The pilot and all four passengers on board suffered burns during the crash and subsequent fire, but two sustained even more serious injuries—Lynette Nikolic damaged her spine and left arm, and Graorac suffered a punctured lung.

"CRITICALLY, RADIO CONTACT WAS LOST, SO NO DISTRESS CALL WAS EMITTED BEFORE THE PLANE WENT DOWN"

The following morning, Du Plessis and pilot Costa Marcandonatos decided to walk the estimated four days to Maun to find help. Their plan was misguided from the start—Maun was actually around some 190 miles (300 km) to the west, and other settlements to the east were much closer. They attempted to navigate by the sun but, lacking a machete, couldn't penetrate the dense bush. They were forced to follow elephant tracks to find watering holes, meeting elephants on one occasion, and heard the roar of lions at night. By Saturday they had walked 56 miles (90 km), but their zig-zag route had taken them only 20 miles (30 km) from the crash site.

Despite their mistakes, they stumbled upon a manned hunting lodge the following day. After their initial efforts to radio the authorities had failed, a helicopter finally reached the crash site on Monday, March 6, airlifting the casualties to hospital for emergency treatment six days after the crash.

WHAT TO DO

ARE YOU IN DANGER?

NO YES

If you are in a group, try to help any others who are in danger

Get yourself out of it:
Sun/Heat—Find or improvise immediate shelter
Animals—Avoid confrontation and move away from danger
Injury—Stabilize condition and apply first aid

ASSESS YOUR SITUATION
See pages 234–35

DOES ANYONE KNOW YOU WILL BE MISSING OR WHERE YOU ARE?

NO YES

If no one knows you are missing or where you are, you will need to notify people of your plight by any means at your disposal

If you are missed, a rescue party will almost certainly be dispatched to find you

DO YOU HAVE ANY MEANS OF COMMUNICATION?

NO YES

You are faced with surviving for an indefinite period—until you are located or you find help

If you have a cell or satellite phone, let someone know your predicament. If your situation is serious enough to be worthy of emergency rescue, and you have a Personal Locator Beacon (PLB), you should consider this option

CAN YOU SURVIVE WHERE YOU ARE? *

NO YES

If you cannot survive where you are and there are no physical reasons why you should remain, you will have to move to a location that offers either a better chance of survival, rescue, or both

Address the Principles of Survival: Protection, Location, Water, Food

YOU WILL HAVE TO MOVE **

YOU SHOULD STAY **

DO

- Make an informed decision on the best location to move to
- Leave clear indications of your intent (written messages or signs) if abandoning a vehicle
- Have aids to location accessible while moving and deployed when static
- Protect yourself against glare from the sun and windburn
- Seek or improvise shelter when not moving and seize all opportunities to collect fuel for a fire—deserts get cold at night
- Be on constant lookout for signs of water or civilization—such as green vegetation, converging animal/human tracks, or circling birds

DON'T

- Sit directly on the hot ground when you stop
- Travel during the hottest part of the day
- Take risks—a twisted ankle caused by running down a sand dune could be fatal
- Force the pace—travel at the speed of the slowest person in the group
- Shelter in dry river beds, because of the potential risk of flash floods

DON'T

- Leave a broken down vehicle unless absolutely necessary—it's easier to see and is what rescuers will be looking for
- Waste energy unnecessarily—be idle unless there is something that requires doing
- Ignore your fire—use anything that burns (wood, vehicle tires) to generate heat and smoke

DO

- Seek or improvise shade and work during the coolest part of the day/night. A shelter dug even 6 in (15 cm) below ground level will provide a much cooler place to rest
- Ration sweat, not your water; you require around 1¾ pints (1 liter) per hour in temps above 100°F (38°C), half of this if it's cooler
- Continually re-assess your water situation and options for augmenting your supplies
- Leave everything out overnight that could collect dew so you can drink it
- Prepare all of your aids to location for immediate use. Be constantly alert for signs of rescue

* If you cannot survive where you are, but you also cannot move owing to injury or other factors, you must do everything you can to attract rescue.
** If your situation changes (for instance, you are "moving" to find help, and you find a suitable location in which you can stay and survive) consult the alternative "Do" and "Don'ts."

FIRE BY FRICTION

THIS METHOD USES PRIMITIVE EQUIPMENT, such as a bow drill, and requires knowledge, skill, practice, the correct wood, time, and effort. However, once learned it's possibly the most rewarding of all survival skills. Your first fire created from a coal produced by a fireset that you have made yourself will be a fire you will never forget. In a genuine survival situation, however, you would always start a fire with the quickest and easiest method available.

TOOLS AND MATERIALS

- Cordage such as para cord, fishing line, shoelaces, or bailing twine
- Wood
- Knife
- Bradawl (optional)

THE FOUR SKILLS

There are many ways to make fire by friction, including a hand drill, bamboo fire saw (see p. 133), and bow drill. Each has its own advantages, depending on the materials available and your practical skills. Regardless of the method, however, there are four main elements to producing fire by friction—each a skill on its own:

- Identifying and procuring the correct types of wood.
- Manufacturing the individual pieces that make the fireset.
- Using the fireset to produce a coal.
- Nurturing the coal into a flame.

MAKING A BOW DRILL

The bow drill set is one of the most efficient methods of making fire by friction. Try to find the right type of wood for each part, especially the drill and hearth board (see Testing Wood, opposite).

1 To make the bow, cut a stick to the same length as the distance between your armpit and your fingertip.

- Cut a notch (or make a hole with a bradawl) near each end.
- Tie cord with an arbor knot (see pp. 142–43) to one notch and with half hitches (see pp. 144–45) to the other. Leave only enough slack to wrap around your finger.

2 To make the drill, cut a straight piece of wood that has no knots.

- The drill needs to be about 1 in (25 mm) in diameter and about 8 in (20 cm) long.

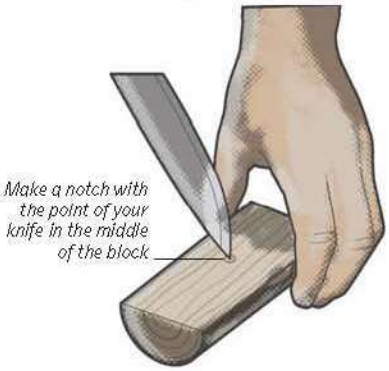

3 Make the bearing block from a piece of hard wood about 3–4 in (7.5–10 cm) across.

- Cut it to a length of 4–5 in (12–15 cm), split it in two, and make a notch in it.
- The drill's dull point fits into the notch, while you hold the block when you are bowing (see p. 132).

TESTING WOOD

Standing deadwood is ideal for the drill and hearth board as it's dry. Soft deadwood is easier to use than hard deadwood, but if you can't tell the difference, use these tests.

THUMBNAIL TEST

Cut through the bark of a branch, exposing a patch of wood beneath. If you dent this wood easily with your thumbnail, it's a soft wood. If you can only make a small impression, then it's a hard wood.

1 Hold your knife at a slight angle to the wood. Push it with a thumb on top of the blade.

2 Press your thumbnail at an angle of 90 degrees to the wood in the exposed patch.

CHAR TEST

Each type of wood produces a distinctive char and coal. Some, such as hazel and lime, make a coal easily but the fireset will wear out quickly. The char test helps you discover which woods make a good coal. Drill into a softwood hearth board until it smokes, continue for five seconds and then stop. Look at the char produced:

- If it's like a fine powder it will probably produce a coal.
- If it's coarse, or appears to disintegrate into nothing, it probably won't make a coal.

4 Make the hearth board from a flat piece of wood about 1 in (25 mm) thick. You may have to split some wood to get what you need.

- Hold your drill so that it sits about ⅔ in (10 mm) from the edge of the board.
- Mark this point with your knife and notch out a small hole to help the drill get started.

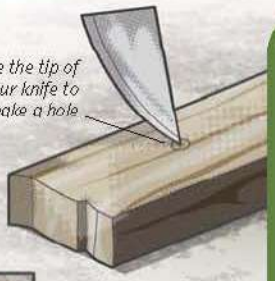

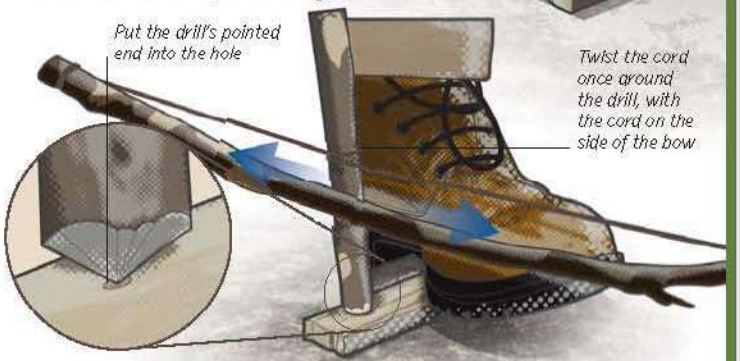

5 Set up the bow, drill, and bearing block (see also p. 132) and bow the drill into the hole in the hearth board until the hole is as wide as the drill. Don't go all the way through.

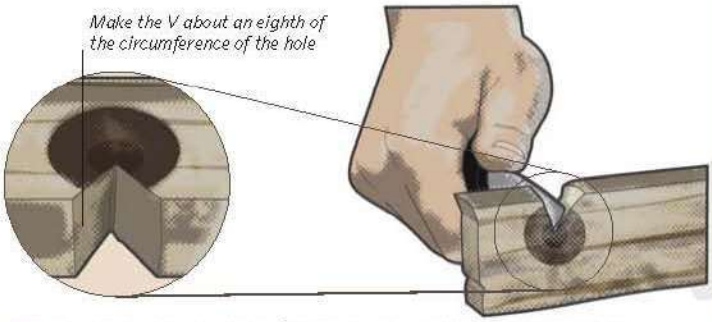

6 Cut a V to the center of the hole through the hearth board. The bottom of this V is where the char collects and the ember forms in the ember pan (see p. 132).

THE FINISHED FIRESET

The bow drill set works best when the drill and hearth board are made from the same wood, because both parts then wear down evenly. This may not always be possible, so experiment with different woods.

- Good woods to start with are sycamore, hazel, ivy, lime, willow, and sotol (not found in Europe).
- Also useful are alder, birch, cedar, pine, spruce, maple, oak, and poplar.

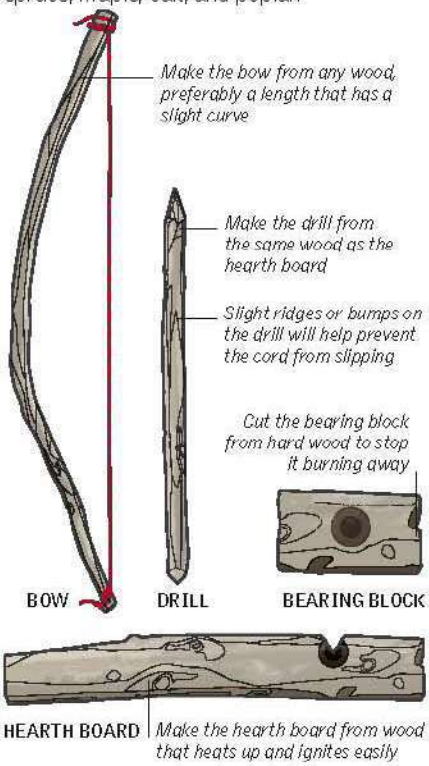

USING A BOW DRILL

Before you use a bow drill, have your dry tinder bundle, kindling, and fuel beside you. Until you become proficient at using a drill, you will use lots of energy, so beware of overheating and dehydration. Turning a coal into a fire is a skill in itself, so use every opportunity to practice making tinder bundles and lighting them with a small coal from your camp fire.

TOOLS AND MATERIALS

- Bow drill set (see pp. 130–31)
- Thin, dry piece of bark to use as an ember pan
- Buffed tinder, kindling, and fuel

GETTING INTO POSITION

Place the ember pan on a flat, dry surface and align it with the V on the hearth board. Rest one foot on the board. Loop the bow cord once around the drill, starting on the inside. Insert the drill into the hole on the hearth board, slot the block on top (with a green leaf in the notch for lubrication), and lock your wrist into your shin as support. Kneel on the other leg and bear down on the block.

1 Hold the bow horizontal to the ground and parallel to your body.

- Lean slightly forward to apply downward pressure to the drill and back to reduce it.
- Start bowing slowly, using the full length of the bow. Aim for an even, flowing rhythm.
- Breathing steadily as you bow, gradually increase speed and pressure until smoke appears. Vigorously bow with maximum effort ten more times—this usually produces a coal in the ember pan.

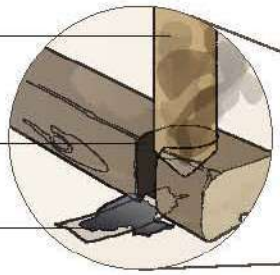

2 Gently remove the drill and lift the hearth board up and away from the smoking coal in the ember pan.

- If the coal is sticking to the board, gently tap the board with your knife to loosen it.
- Carefully lift the ember pan off the ground and hold it in the air to see if the coal is glowing.
- If the coal in the ember is just smoking, and not glowing, then gently fan it with your hand until you see it glow.

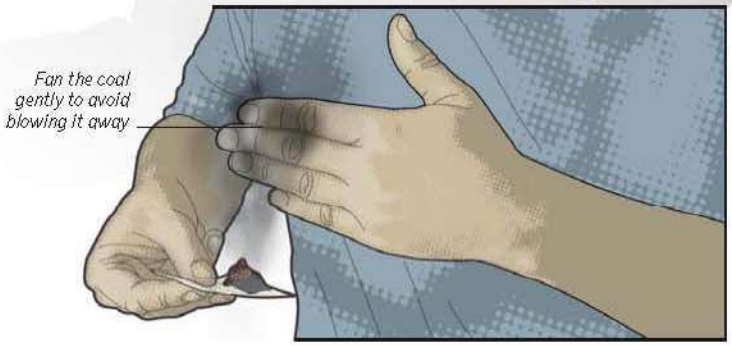

3 Gently transfer the glowing coal into your tinder bundle.

- The coal must touch the tinder to transfer its heat, so squeeze the tinder around the coal, but don't crush it or starve it of oxygen.

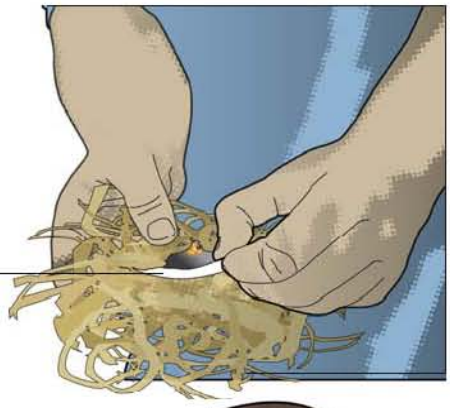

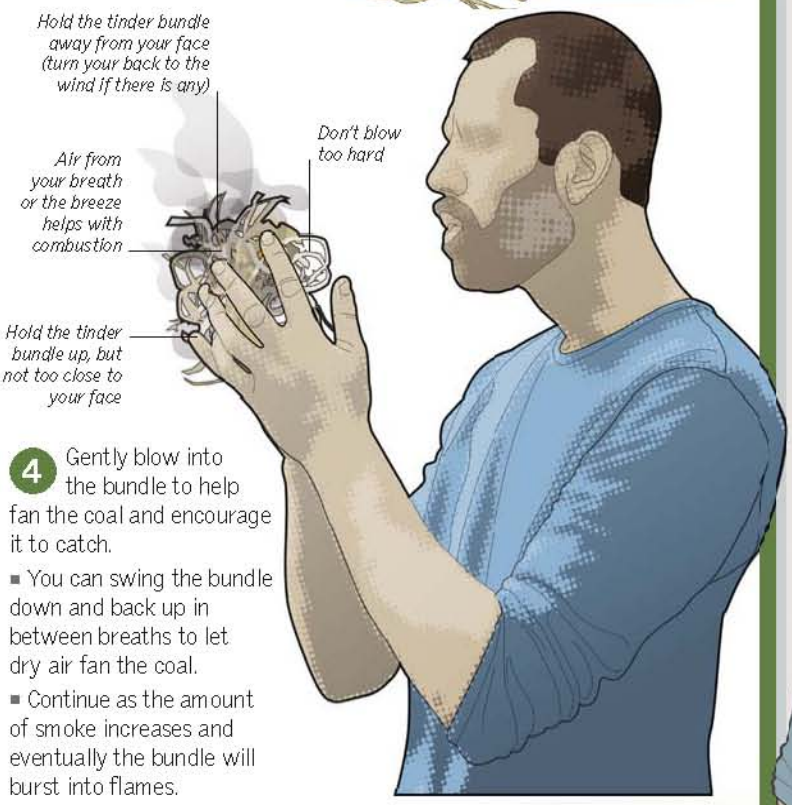

4 Gently blow into the bundle to help fan the coal and encourage it to catch.

- You can swing the bundle down and back up in between breaths to let dry air fan the coal.
- Continue as the amount of smoke increases and eventually the bundle will burst into flames.

5 Don't throw the burning bundle on the ground, just calmly place it where you intend to have your fire. Start building the fire around the bundle (see pp. 124–25).

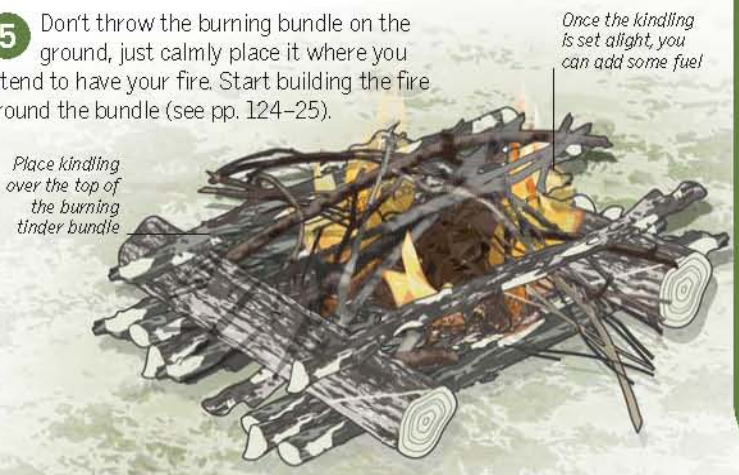

BAMBOO FIRE SAW

This friction method uses a saw technique to light tinder collected from scraping the outside of the bamboo with your knife.

MAKING THE FIRE SAW

You need a bamboo section 18 in (45 cm) long. Split it lengthways into two equal pieces.

1 Sharpen the edge of the first piece with your knife and then cut a V-shaped notch in the outside middle of the second piece.

2 Hold tinder over the notch in the second piece of bamboo with another bamboo strip that's snapped but not fully broken.

- Kneel down and brace the sharpened piece of bamboo vertically between the ground and your stomach (place a pad on your stomach for comfort).
- Place the notch and tinder over the sharp edge of the first piece, and rub the second piece up and down in a rhythmic movement.

3 When the friction produces lots of smoke, continue for 20 strokes more and you should have a coal.

MANUAL SKILLS

Learning skills that you can use in the wilderness is always a "work in progress." Having taught thousands of students over several years, I am yet to conduct a course in which I didn't learn a new skill. For example, this could be a more efficient way to erect a shelter, how to tie a new knot, or the best way to carry out a task more safely and efficiently.

In the military, we invariably teach students how to accomplish a particular task using only the most basic gear. By learning how to perform a task using the absolute basic equipment, they will be able to understand the important elements of the task and will have no choice but to get them right in order to succeed. This in turn helps students to appreciate how and why certain techniques do and don't work, and gives them an opportunity to use their improvisational skills to modify basic techniques in order to get the most out of them. Knowing the correct method of using a knife, for example, will not only improve safety on an expedition, but will also mean that only the minimum

In this section YOU WILL DISCOVER...

- the merits of a **hank** in **your pocket**...
- how to strip a **nettle stem** and not **get stung**...
- when **to twist** and when **to roll**...
- how to tell a **working end** from a **standing end**...
- the uses of a **sheet bend** or a **Siberian hitch**...
- the **difference** between a **parang** and a **kukri**...
- how to **fell a tree** with a **thumping stick**...

amount of valuable energy is being used to complete a task. Equally, knowing how to tie a few simple knots, which can be used in the majority of survival situations, will allow you to use whatever cordage you have in the most efficient way possible.

If you find yourself in a survival situation with more than the most basic gear, life will be much easier. However, the ability to improvise what you need when you don't have it could mean the difference between continued survival and despair.

If you give someone with little understanding of knots a length of cord and ask them to complete a task, they will invariably use the whole length of cord. If, however, you teach someone two simple knots and set them the same task, they will invariably use only the necessary amount of cord.

USING THE BEST KNOT, lashing, or hitch for a particular task means that you will use cord more efficiently and, therefore, ensure that your supply will go further. In a survival situation, where you may be moving every day, you need to be able to use the same pieces of cordage day after day.

- Use knots that don't require cutting and that can be untied quickly and easily (trying to undo a knot with cold, wet hands in the dark, and when you are tired and miserable, can be an emotional experience that usually ends in frustration and several pieces of cut cordage).
- Before you set off into the wilderness, learn a few basic knots, such as the Siberian hitch and the taut-line hitch, which are great for rigging a simple poncho shelter.
- Practice these knots until you can tie—and untie—them with your eyes closed.
- If you are trekking in a cold environment, practice tying them wearing gloves.

> "The **ability** to **improvise** what you **need** when you **don't have it** can mean the **difference** between **continued survival** and **despair**"

MAN-MADE CORDAGE

A VITAL PIECE OF SURVIVAL EQUIPMENT, man-made cordage is one of those small items that has many uses: building a shelter, repairing equipment and clothing, making traps and nets, and producing fire by friction using a bow drill. Always pack plenty.

USING CORDAGE

It is vital that you use your limited supply of cordage as efficiently as possible.

- Use only the minimum cordage that is absolutely necessary to accomplish each task.
- Use simple, strong knots that, if possible, can be released without cutting. This leaves the cordage intact for the next task.

PACKING CORDAGE

Before you set off on your travels, check how much cordage you will need. Make sure you pack it in two places—in your backpack and on your person, in case you get separated from your pack. Make sure, too, that the needle you pack in your survival kit has an eye big enough for the inner strands of the cordage.

PARACORD

Parachute cord, or "paracord," is a type of cordage originally developed for the rigging lines of parachutes and then adopted as the standard utility cordage by most of the world's military forces. It is a lightweight, nylon rope composed of 32 braided strands. Each strand contains a number of smaller threads that can also be used. In most situations, paracord is a very good option because it's readily available, strong, and packs down small.

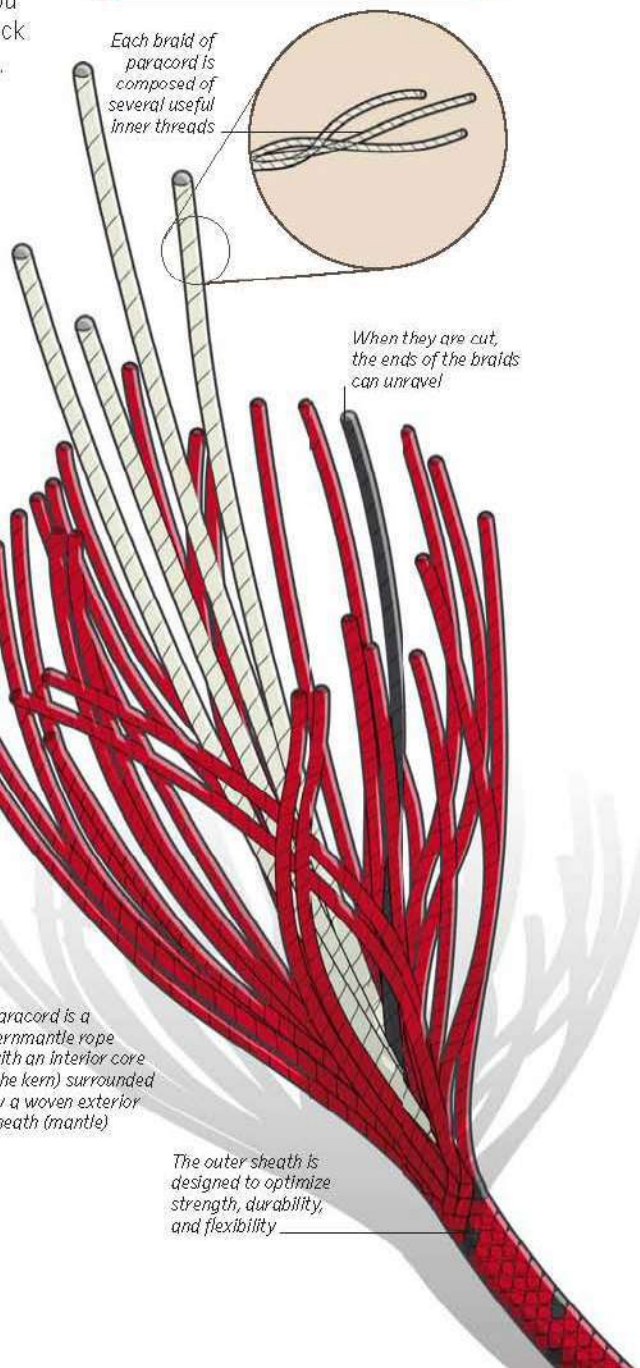

SURVIVAL TIP

Green is the ideal color for military paracord as it remains camouflaged. In a survival situation, red is best because it stands out—it's easier to find if you drop it and can be used as an aid to location. Don't pack just one very long piece of paracord (150ft/30m, for example). It's much better to cut the paracord into 30 ft (10 m) lengths and tie each of them into a coil called a hank (see opposite and below).

TEN HANKS

A hank can fit in the palm of your hand. It takes up very little room so you can easily stuff up to 10 of them into your pack and pockets, and leave them there until you need them.

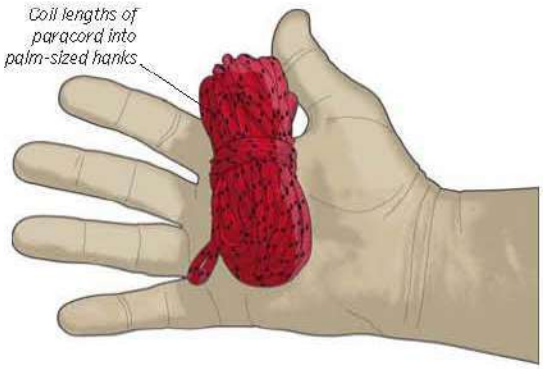

TYING A HANK

It's best to keep your cordage in a hank—you can untie it when you need a piece and retie it to store it again. Tying a hank is straightforward and soon becomes second nature, because you'll do it again and again.

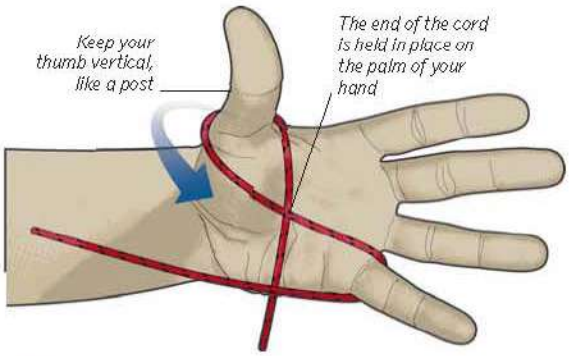

1 Hold out your hand (you can use either hand) with the thumb and the fingers spread.

- Lay the end of the cord on your palm, then loop the cord in a figure-eight around your thumb and little finger.

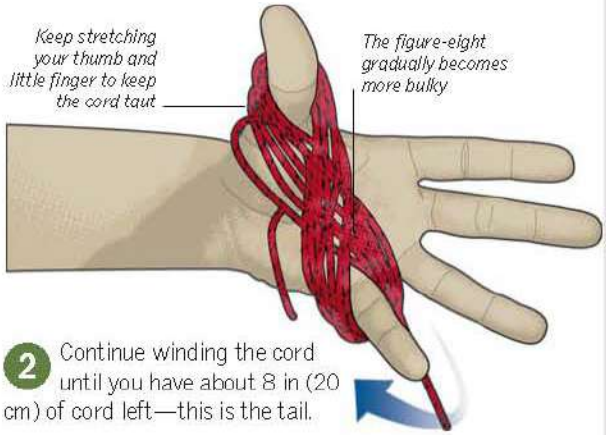

2 Continue winding the cord until you have about 8 in (20 cm) of cord left—this is the tail.

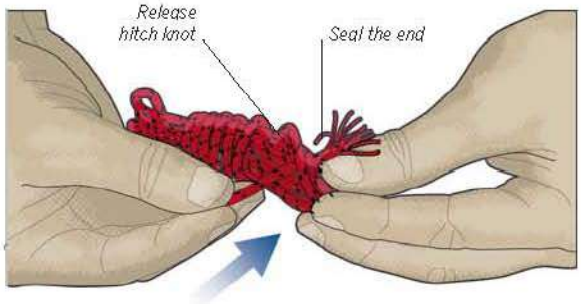

3 Take the hanked cordage off your thumb and little finger.

- Secure the hank by tying the tail around it with a release hitch knot (see p. 143).

OTHER TYPES OF CORDAGE

In a survival situation you may have to improvise and use other types of cordage, especially if you didn't bring any paracord with you or have run out. Before you start making it from natural materials (see pp. 138–41), see what else you might have that would be quicker and easier to adapt.

SHOELACES

The normal laces of shoes and boots are often strong enough to use as cord. However, the members of many military units, who are limited to the amount of equipment they can stow or carry, replace their bootlaces with paracord. They use twice as much as they need and wrap the extra length around the top of their boots before tying it. This gives them a ready supply of extra cordage that they can cut off as needed. You can easily do the same—but never sacrifice all the cordage you are using as shoelaces as you'll need to make sure you can still walk comfortably.

BELTS, CLOTHING, AND HATS

You can cut any type of belt, whatever it's made of (leather, canvas, and so on) into strips to use as cordage. You can cut clothing into strips, too. However, you don't have to destroy a shirt completely: start at the bottom hem and cut 1 in (2.5 cm) strips all the way round. You can even use your hat—a 1 in (2.5 cm) strip from the brim of a jungle hat is just over 3 ft (1 m) long. Better still, coil paracord around the crown of your hat and stitch it in place (just the front and the back).

RAIDING YOUR KIT

You may have other items in your equipment that you could use as cordage, such as guylines from tents or draw cords from clothing or sleeping bags. Dental floss, especially the waxed type, is extremely strong and can be used effectively for lashing or sewing. If you have a vehicle, you probably have a towrope; take a part of it and unravel the rope into smaller pieces of cordage.

BE A SCAVENGER

Few places are untouched by modern man. Even in remote areas you'll come across discarded items from the "civilized world"—from plastic bottles to bailing twine. In a situation where improvisation could be the difference between life and death, collect anything you could use to your advantage.

SEALING PARACORD

Paracord can be awkward to use because its ends, if left unprotected, eventually unravel. Try to seal the ends after you have cut them by holding a flame next to them until they melt. Be careful, as the melted nylon can drip and burn.

NATURAL CORDAGE

IF YOU DON'T HAVE any man-made cordage, you may be able to make a natural substitute. Use the roots and stems of plants to produce improvised rope for erecting shelters, and use sinew from the tendons of larger animals for strong whipping and sewing.

WARNING!
Take extreme care when you collect and work with plant fibers, especially removing plants such as green ferns from the ground. Bending a stem at its base and pulling it up can expose fibers that are razor-sharp. Cut the plant with a knife or wear gloves. The juice of some plants can irritate the skin, so wash your hands when you've finished.

PREPARING NETTLE FIBERS

Old stinging nettles or wood nettles with long stems have the most fibers. Wearing gloves (or with your hands covered), strip the leaves from the nettles by grasping the stem at the base and pulling your clenched fist up the full length of the stem.

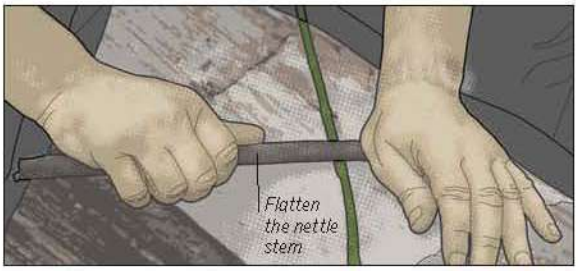

1 Sit astride a log and lay a nettle stem in front of you. Roll a smooth, rounded stick backward and forward on the stem, pressing down hard.

- Continue rolling until the whole stem is crushed.

2 Tease open the crushed tissue along the whole length of the stem of the nettle with your thumbnail. This exposes the spongy pith within.

CHOOSING YOUR RAW MATERIALS

Many kinds of raw material can be used to make natural cordage, from bark and roots to stems and sinews. Search your immediate environment first to see what can be used—you may not have to look too far.

	BARK	ROOTS
FACTS	■ Tree trunks have layers between the bark and heartwood called the "cambium" layers, which are best for cordage.	■ Many tree roots make good cordage and lashings, but evergreen trees, such as pine, fir, spruce, and cedar, work best.
FINDING	■ The bark of trees such as willow and linden make good cordage. ■ Use other types of bark, such as clematis or honeysuckle, as they break from the tree or vine.	■ Look for new roots near the surface of the soil; these are flexible. Cut slightly thicker roots than you need, as you have to remove the bark. Remove only a few roots from each tree and repair any damage you do to the topsoil.
PREPARING	■ Strip the bark from a dead tree, then beat it to loosen the inner cambium fibers. ■ Pull off the cambium fibers in long strips. Use them as they are (clematis and honeysuckle), or braid them into stronger lengths.	■ Cut a slit along the root. Peel the bark and let the bare root dry and shrink before splitting in half or into quarters. ■ Keep the bark for tinder or kindling. You could also put it on a fire to produce smoke that keeps away flies.

3 Bend the stem in the middle of its length. When you bend it, the inner pith will break away from the outer skin, making it easy for you to separate them.

MAKING CORDAGE FROM PLANTS

To begin with, you must have access to enough quantities of the plant. Don't start making cordage with the only plant in the area. There should also be some basic characteristics present in order for plant material to work well as cordage.

- The fibers need to be long enough. If you have to splice them together to make workable lengths, you will weaken the line.
- If you're braiding pieces together to make a stronger line, use rougher fibers, as they will bite together better. Shiny or smooth fibers tend to unravel easily.
- The fibers need to be strong. Pull on them to see how much strain they take before breaking. They also need to be pliable enough so they don't break when you bend and tie them.

4 Carefully peel the outer skin from the inner pith.

- With the finger and thumb of one hand, gently pull the layer outward, while pushing down on it with the thumb of the other hand.
- You'll be left with long fibrous strips of outer skin, which you can make into short bindings or natural cordage (see p. 140).

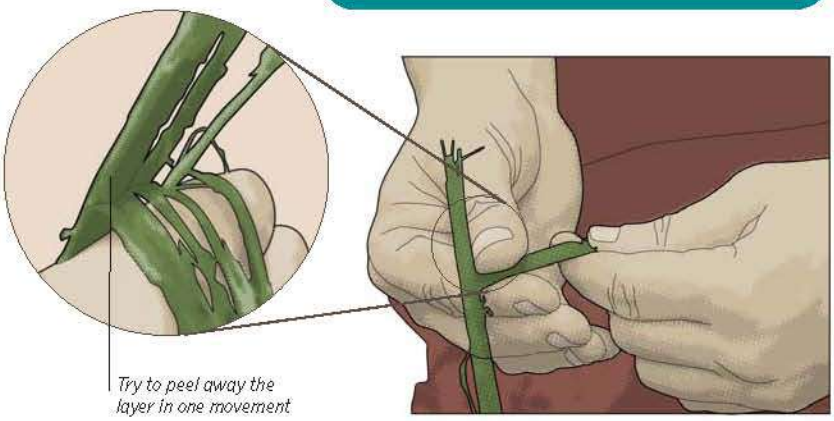

	WITHIES	LEAVES	SINEW
FACTS	■ The strong, flexible stems of willow, birch, ash, and hazel are called "withies" and are used for thatching and in gardening. ■ Withies make useful and robust lashing and are best used in spring or summer.	■ The leaves of many plants, such as the lily family, agave, and sisal hemp, contain useful fibers.	■ Animal sinew makes strong and versatile cordage—some sinews can support the weight of a man. ■ Sinew has been used to bind arrow heads to shafts and to sew together everyday clothing, such as leather and mukluks.
FINDING	■ Choose a young sapling or branch that's long and flexible. The fewer the stems or branches the better, as they will need to be removed.	■ To find out if the leaves of a plant contain useful fibers, simply tear a leaf apart to see if it breaks into stringy layers.	■ Sinews are the tendons connecting bone to muscle. The largest are on both sides of the spinal column, running parallel to it. ■ Shorter lengths can be taken from the calves of the hind legs.
PREPARING	■ Remove stems or branches. Grasp the wand at the base and twist until the fibers break, working your way to the tip. Bend the wand into an S-shape and crank the middle of the wand to loosen the fibers. Then cut the wand from the tree and use the withies.	■ Soak the leaves in water to make the inner layers swell and burst. Bacteria work on the cell tissue, separating it from the fibers. ■ Rinse the fibers in fresh water to stop further bacterial decay and then dry them.	■ Take a tendon from an animal's body, remove its outer sheath, then clean it and let it dry. Separate the dried fibers and use them as individual threads, or splice or braid them together to form stronger cords. You can soften dried sinew with water or saliva.

MAKING NATURAL CORDAGE

Once you have prepared a sufficient quantity of fibers, it's best to let them dry before using them—hang them up in the sun or near a fire. In a survival situation, you can use them before drying. Twisting or rolling the fibers are the two main methods of making natural cordage. As with most survival tasks, practice makes perfect.

TOOLS AND MATERIALS
Enough prepared natural fibers (see pp. 138–39) to give you the length of cordage that you require

TWISTED CORDAGE

Start your first length of cordage toward one end, not in the middle, to prevent all the subsequent splices from coinciding at the same point. This provides strength, as each splice is a potential weak point. If they are thin enough, the twisted cordage fibers can then be plaited to create a thicker cord that's even stronger.

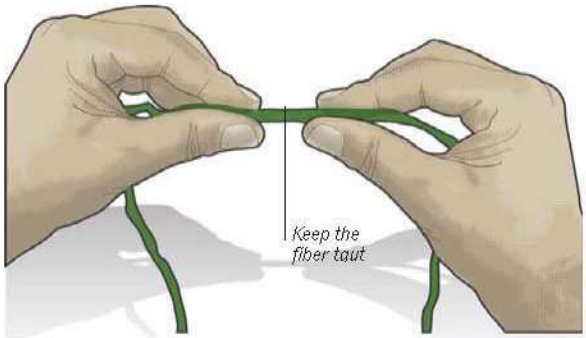

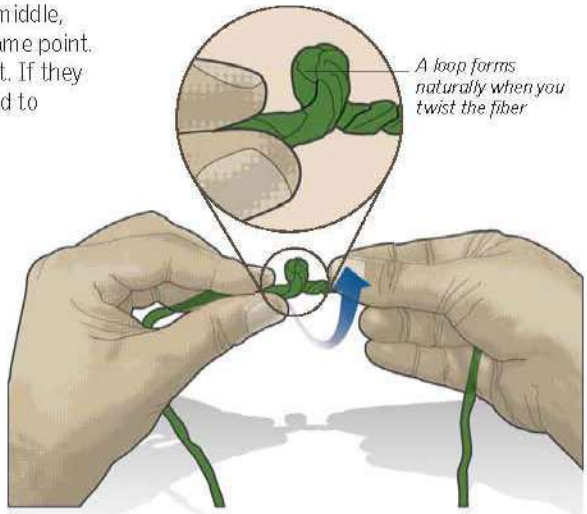

1 Grip the first length of fiber a third of the way along its length between your left thumb and forefinger.
- Grasp the fiber with the thumb and forefinger of your right hand 1 in (3 cm) from your left.

2 Twist the fiber with your right hand until a tight loop is formed.
- Keeping the tension in the cord, inch your left thumb and forefinger forward and clamp down on the loop.

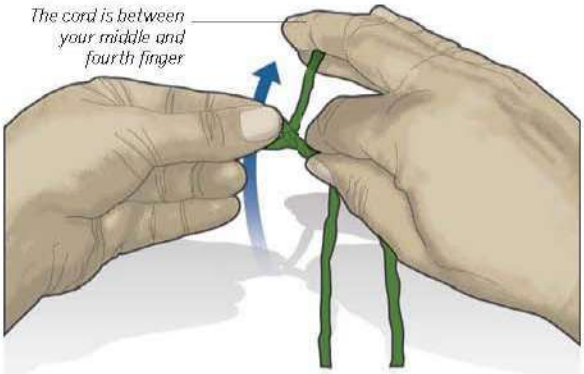

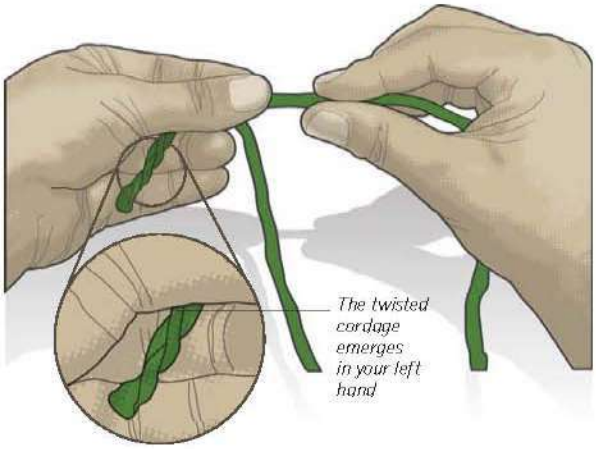

4 Sweep the middle and fourth fingers of your right hand upward, pulling the left-hand cord under the captive piece.
- Inch your left thumb and forefinger forward, and clamp down on the new twist.

5 Now that the new twist is clamped by your left hand, you can release your right thumb and forefinger, letting the lower cord fall.
- Grip the upper cord 1 in (3 cm) from your left fingers with your right thumb and forefinger.

SIMPLE ALTERNATIVE TWIST

You need to start with two fibers of different lengths so that when you splice on additional lengths the splices are staggered. Start by tying one end of one fiber with one end of the other.

- Clamp tied ends in your left thumb and forefinger and the two tails in your right thumb and forefinger about 2 in (5 cm) down.
- Roll both fibers between your fingers away from you for one complete roll. Repeat this until you have about 3 in (7.5 cm) of rolled fibers, then clamp them in your right thumb and forefinger.
- Release your left thumb and forefinger and the rolled fibers will naturally twist together to form a strand of cordage.
- Repeat the process until you have enough for your needs.

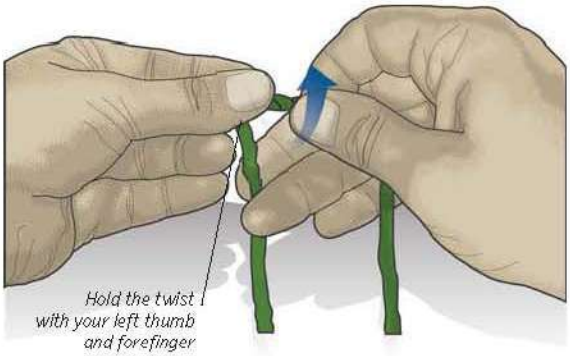

3 Move your right hand back so that 1 in (3 cm) of cord is visible and apply another twist.

- At the same time reach forward with the middle and fourth fingers of your right hand and grasp the cord hanging from your left fingers.

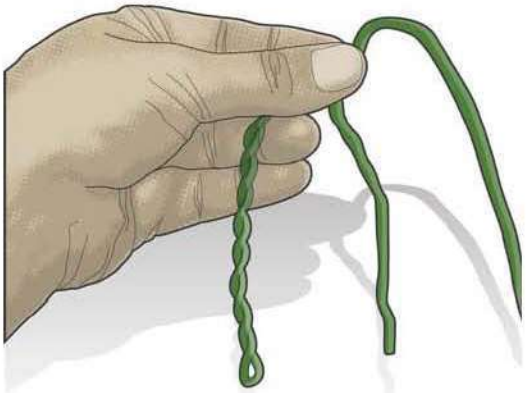

6 Repeat Steps 3, 4, and 5 as you work your way along the length of cord.

- To splice in an extra length, overlap the new piece 2 in (6 cm) along the existing fiber and twist the two together between your left thumb and forefinger. Then continue with the process as before.

ROLLED CORDAGE

This method of rolling cordage, which is also known as the "pygmy roll," is easier to do if you're sitting down. As with the simple alternative twist (see left), you need to start with two fibers of different lengths and tie one end of one fiber to one end of the other.

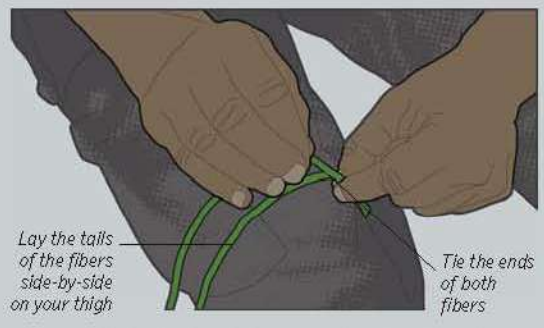

1 Hold the tied ends between the thumb and forefinger of your left hand.

- Lay the two tails of the fibers on your right thigh.

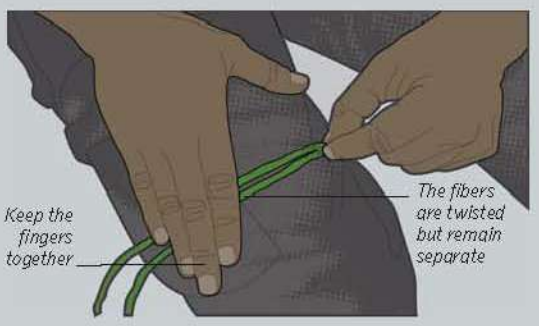

2 Use the flat palm of your right hand to roll both fibers simultaneously away from you. You may find it easier to use the flat of your fingers.

- At the end of the roll, clamp the fibers down onto your thigh with the tips of your fingers.

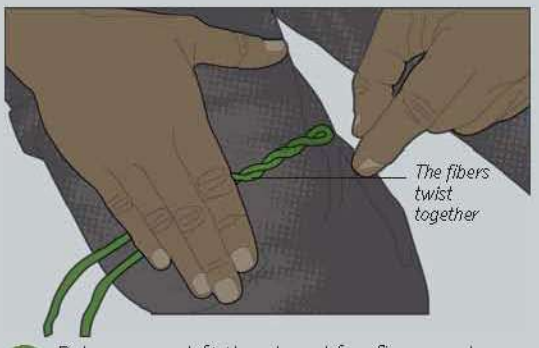

3 Release your left thumb and forefinger and the rolled fibers will naturally twist together to form a strand of cordage.

- Repeat the process, splicing in more cordage as required.

TYING KNOTS

THERE ARE HUNDREDS of different types of knot in existence, but by practicing the handful of straightforward knots shown here and overleaf, you should be able to accomplish most survival tasks. Start by tying each knot in the situation you expect to use it—for example, erecting a poncho shelter—and then practice it at least 20 times.

THE ENDS OF A LINE
The end of a rope or line that takes the most active part in knot tying is the working end. The other end is called the standing end—it's more passive, as the knot is tied around it.

SLIP KNOT

A slip knot is a good example of a simple knot that has many practical uses in survival situations.

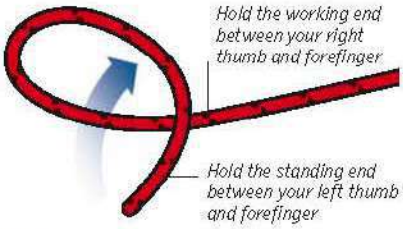

1 Twist the standing line over the working line to form a loop.
- Hold the loop between your left thumb and forefinger.

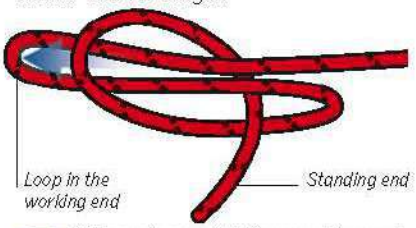

2 Make a loop with the working end and pass it into the first loop.
- Collect the second loop with your left thumb and forefinger and pull it through.

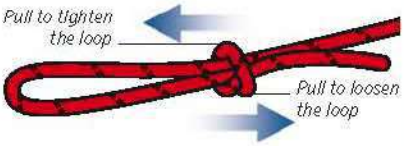

3 Gather both ends in your right thumb and forefinger and pull them away from the loop in your left hand.
- Put the loop over whatever you want to secure—for example, a button-tie on your bivy sack—and pull the knot tight.

ARBOR KNOT

This all-purpose knot has several survival uses. When erecting a poncho or a tarp (see Chapter Four), it can be used as a tie-off knot to secure one end of a line to a fixed point, such as a tree or peg, where no adjustment is required. It can also be used as a lashing to secure two poles together—for example, in an A-frame shelter (see p. 164)—with just one turn of cordage.

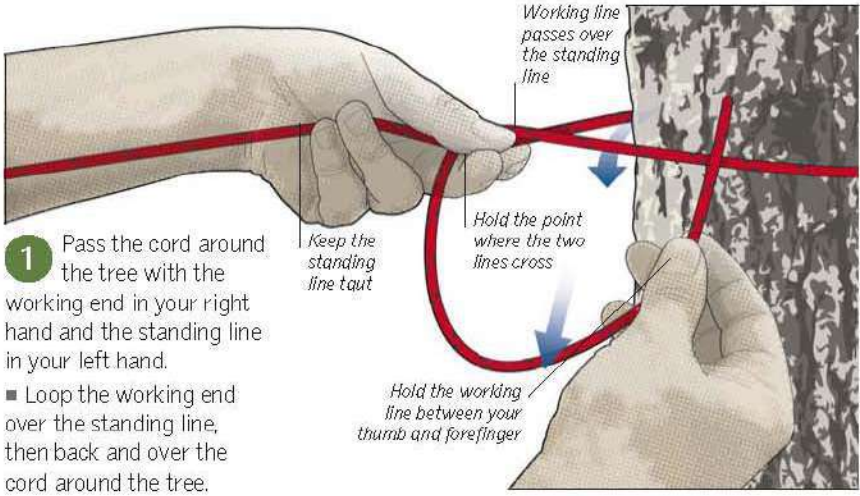

1 Pass the cord around the tree with the working end in your right hand and the standing line in your left hand.
- Loop the working end over the standing line, then back and over the cord around the tree.

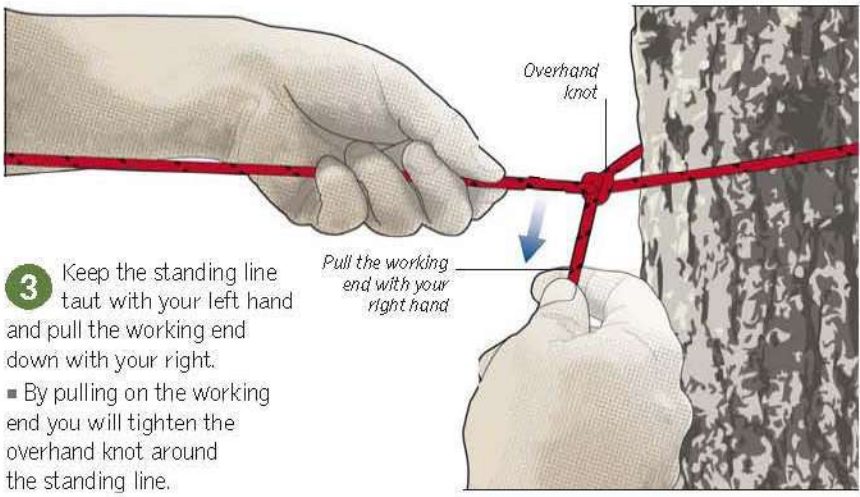

3 Keep the standing line taut with your left hand and pull the working end down with your right.
- By pulling on the working end you will tighten the overhand knot around the standing line.

OVERHAND KNOTS

An overhand knot is the first part of the knot you use to tie the laces on your shoes. One overhand knot tied around another forms a stop knot or they can be tied as a double overhand knot around a fixture.

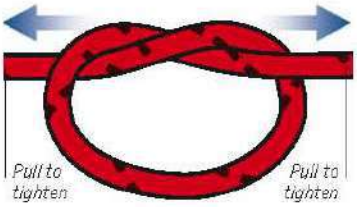

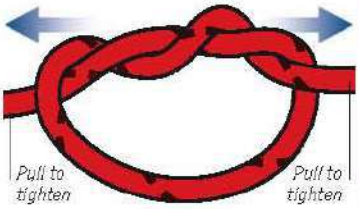

OVERHAND KNOT

Twist the left line over the right to form a loop. Hold it with your right thumb and forefinger.

- Pass the working end through the loop with your left forefinger and thumb and pull both ends to tighten.

DOUBLE OVERHAND KNOT

Follow the sequence for the overhand knot (left), but pass the working end through the loop twice before pulling the two ends.

- A stop knot used at the end of a line can stop the line from slipping.

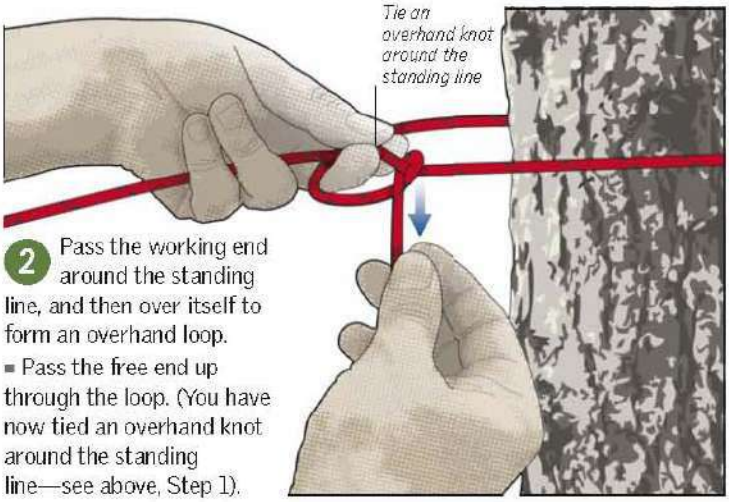

2 Pass the working end around the standing line, and then over itself to form an overhand loop.

- Pass the free end up through the loop. (You have now tied an overhand knot around the standing line—see above, Step 1).

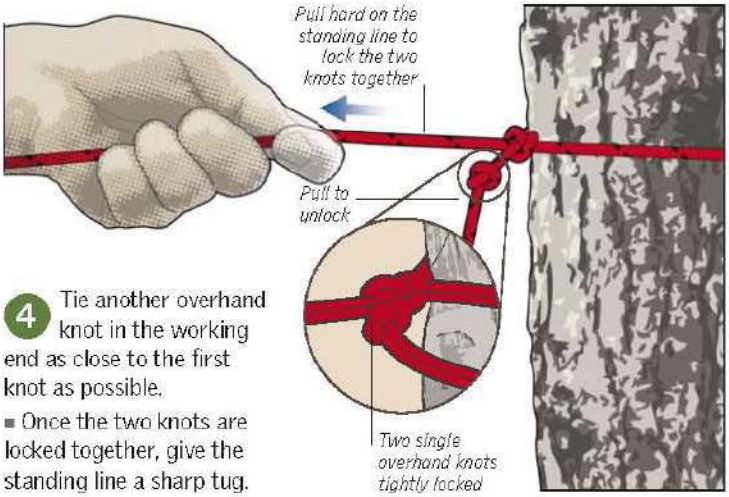

4 Tie another overhand knot in the working end as close to the first knot as possible.

- Once the two knots are locked together, give the standing line a sharp tug.

DOUBLE CHAIN FASTENING KNOT

This knot securely attaches the free end of a line to a fixture, particularly when weight is involved and you don't want the knot to slip or come loose.

1 Loop the working end around the fixture, such as a tree or pole.

- Pass the working end over and around the standing line.

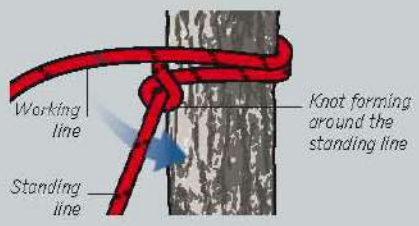

2 Pass the working end back around the fixture and then bring it back under the standing line.

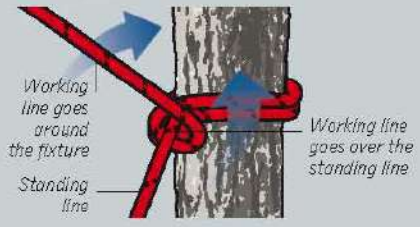

3 Pass the working end over the top and around the standing line again, and once more around the fixture.

4 As you bring round the working line, thread it underneath itself to complete the knot.

- Lock the knot neatly into position.

DOUBLE SHEET BEND

The sheet bend, also known as the "weaver's hitch," is an extremely useful survival knot as it securely ties together two ropes. Double sheet bends are even more secure and are recommended for tying ropes of different thickness. In this instance you use the thinner of the two ropes to tie the knot.

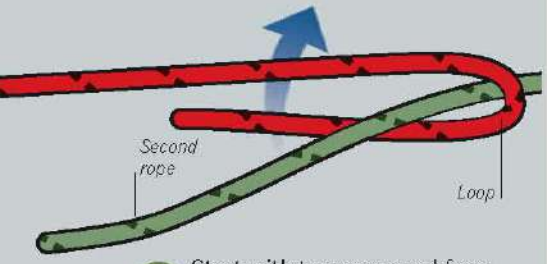

1 Start with two ropes and form a loop in the working end of one of them.

- Pass the end of the second rope through the loop.

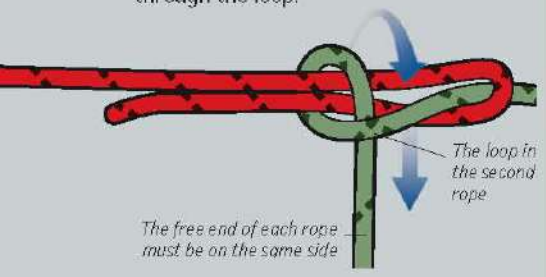

2 Pass the free end of the second rope first under, then up and over the loop. Pass the end through the new loop you have created.

- To finish a single sheet bend, pull the standing end of each rope to tighten it.

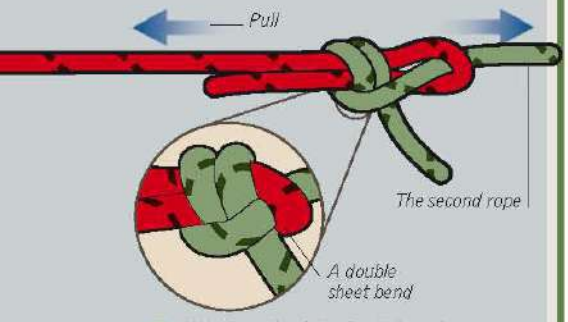

3 To tie a double sheet bend, pass the free end of the second rope through the loop (as in step 2) again.

- Tighten the knot as before by pulling the free end of each rope.

SIBERIAN HITCH

This knot, which is also known as the "Evenk knot," is good for attaching a rope to a fixture—for example, when securing a ridgeline to a tree for a poncho shelter (see p. 158).

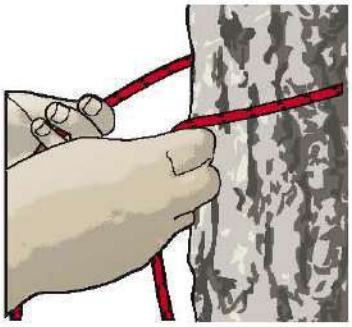

1 Pass the cord around the tree with the working line in your right hand and the standing line in your left hand.

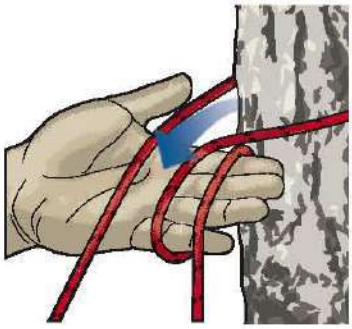

2 Lay the standing line on the palm of your left hand.

- Loop the working line one and a half times around your fingers.

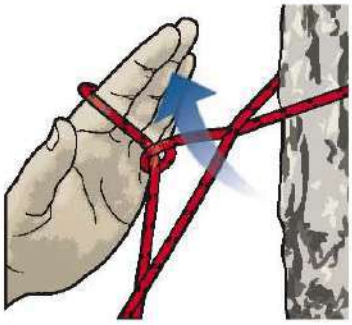

3 Bring your left hand under the standing line.

- Turn the palm of your hand so that one loop twists over another.

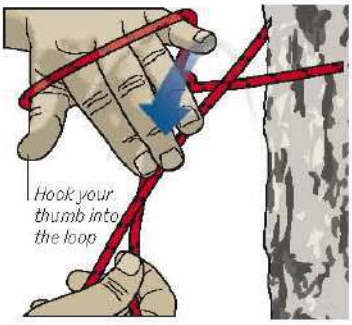

4 Hold the working end in your right thumb and forefinger.

- Bring the loop that is around your left fingers over the standing line.

5 Pinch the working line with the fingers of your left hand.

- Pull it up through the loop to tighten the knot on itself.

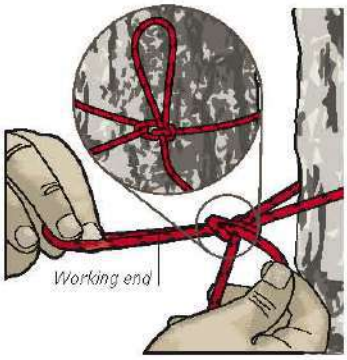

6 Pull the working end again, to tighten the knot around the tree.

- To free the knot, pull the end of the standing line.

TAUT LINE HITCH

This knot is widely used in survival and outdoor activities because it can be adjusted to increase or decrease the tension in a fixed line—for example, a guide rope on a tent or a mooring on a boat in tidal waters.

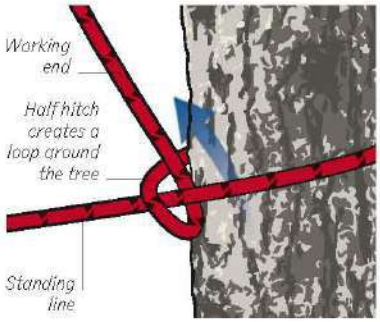

1 Attach the standing line to an anchor point, such as a tent, and pass the working line around a tree.

- Tie a half hitch (see below) with the working end around the standing line.

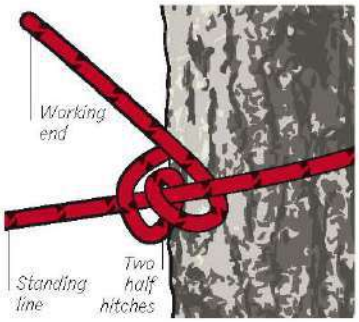

2 Tie another half hitch next to the first one, so that the two half hitches sit side by side around the standing line.

- Don't tighten the half hitches yet.

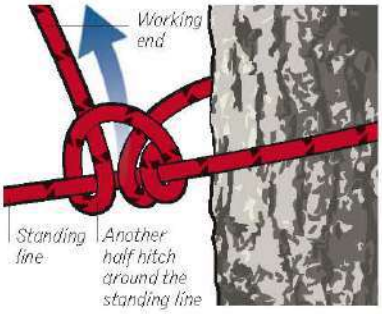

3 Pass the working end under the standing line and back through the new loop it created (but outside of the loops in Steps 1 and 2).

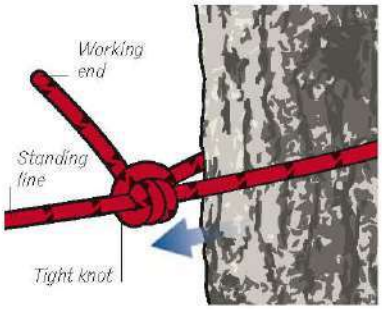

4 Tighten the knot by pulling both the lines.

- Pull the knot further from the anchor point to increase the tension.

QUICK KNOT RELEASE

You can dismantle a tent or free a boat from its mooring more quickly by finishing off the taut line hitch (see left) with a quick-release knot.

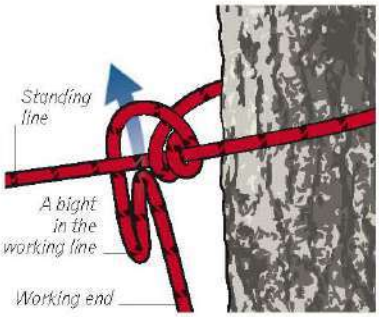

1 Repeat Step 3 of the taut line hitch (see left), but this time make a loop, known as a "bight," in the working line and pass it—rather than just the end—through the loop.

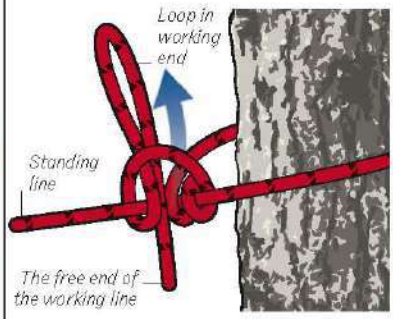

2 Tighten the knot as in Step 4 (see left).

- Leave the working end free so you can pull it quickly to release the knot.

HALF HITCH

The half hitch is a simple knot that forms the basis for other knots (see taut line hitch, above). It's not a secure knot and, when used alone, it tends to slip. You can double or treble the knot to remedy this. Two half hitches under tension can be difficult to undo without cutting the cord, especially after a few days, if the knot has dried out after being wet.

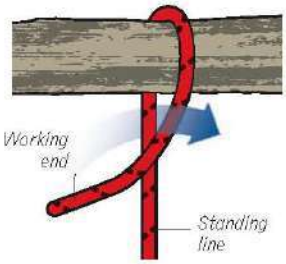

1 Pass the cord around the fixture, such as a bar or tree.

- Pass the working end around the standing line and back through the loop.

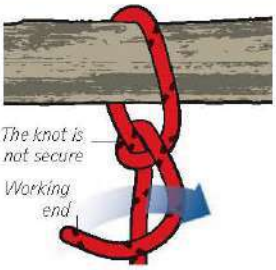

2 Hold the standing line and push the half hitch knot toward the fixture to tighten it.

- Repeat the sequence and tie another half hitch.

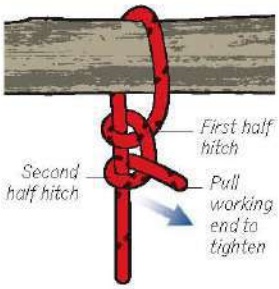

3 Push the second knot tight to the first hitch knot for more security.

- If you need to, finish off the working end with an overhand knot (see p. 143).

USING CUTTING TOOLS

ANYONE VENTURING INTO THE WILDERNESS should carry some kind of knife, preferably a bushcraft knife. It's probably the most important survival tool after knowledge. Although a good knife will help you through many survival situations, look at the environment you are traveling to and choose your cutting tools according to their potential uses—for example, you may need a saw as well as a knife.

WARNING!
If you drop a bushcraft knife, never attempt to catch it because there's a very good chance you will lose your fingers!

BUSHCRAFT KNIFE

An experienced person can accomplish most survival skills, such as making feather sticks (see pp. 120–21), with a quality bushcraft knife. This type of cutting tool has a number of crucial features that will stand you in good stead in every situation.

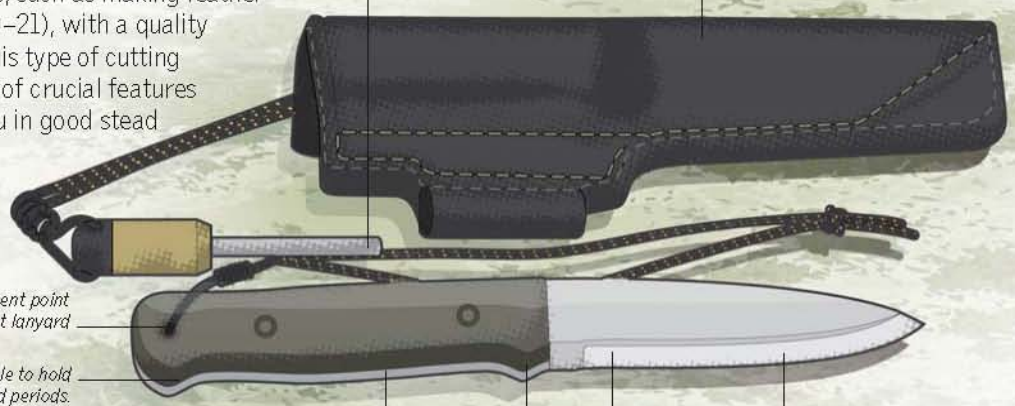

LARGE KNIVES

All over the world, people living primitive lives have two essential items: a cooking pot and a large working blade, either a parang or machete.

- Parangs and machetes are similar types of knife. The parang is associated with the Malay people and there are many varieties available. Machetes are more common—you can still see machetes being used in Africa today that were made 40 years ago from old Landrover suspension springs! The kukri is a similar knife used by the Ghurkas of Nepal.
- A parang or machete blade is about 18–20 in (45–50 cm) long and is used primarily for cutting and slashing but, like the ax, can be used for intricate work.
- Always wear a wrist lanyard when using a machete, parang, or kukri—and pay attention to what you are doing.

KUKRI

MACHETE/PARANG

KNIFE KNOW-HOW

- A bushcraft knife is used primarily for carving, cutting, and splitting small logs. If you need to chop large logs, use either a machete or a saw.
- A knife doesn't need to be longer than 8–9 in (20–22.5 cm).
- Don't use a knife with a serrated edge when working with wood, as the serrations make it hard to use the part of the blade closest to the handle (essential for making feather sticks, etc).
- Keep your knife safe when not in use. Clean and dry the blade and return it to its sheath.
- Wear your sheath on your belt or around your neck on a lanyard. Belt sheaths may be high (the handle is above the belt) or low (the handle is below the belt, which allows you to sit down without the handle poking you in the side).

SHARPENING A KNIFE

It's very important to keep the blade of your knife sharp. Blunt blades are dangerous because they are likely to bounce or slip off the wood you're working on. Also, blades that have completely lost their cutting edge are very difficult to sharpen.

SHARPENING STONES

Keep a good sharpening stone at home to ensure you always set out with a knife that is sharp.

- In the field, carry a smaller, lightweight device for quickly re-sharpening a blade that is beginning to lose its edge.
- There is a good range of small sharpening stones (medium one side, fine the other) that fit into a neat leather pouch.
- Small automatic sharpeners are available to buy that use two sets of ceramic stones to sharpen a knife pulled between them.

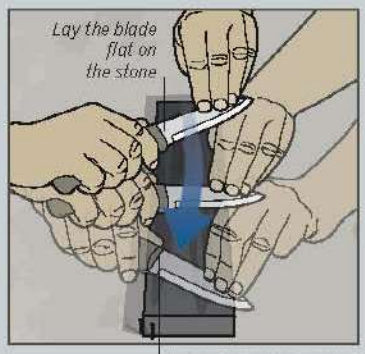

1 Hold the knife in a forehand grip, with the blade pointing away from you. Support the tip of the blade with the thumb and fingers of your other hand.

- Push the knife so the full blade arcs over the stone in one sweep.
- Lift the blade and return to the start, repeating the action 6–8 times.

2 Repeat the sequence for the other side of the blade, but reversing the crucial elements.

- Hold the knife with a backhand grip, with the blade pointing toward you at the far end of the stone.
- Pull the knife in an arc over the stone in one draw. Lift and return to the start, repeating the action 6–8 times.

WORKING WITH YOUR KNIFE

The way you hold your bushcraft knife can help you accomplish a variety of cutting tasks easily and safely. There are several essential grips, including the forehand, backhand, and chest lever. In most cases, you need to grip the handle firmly at the end nearest the blade so that you can cut with the part of the blade nearest the handle.

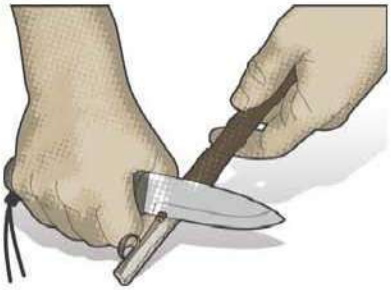

FOREHAND GRIP

This grip is the most natural and the one most often used. It allows your back and arm to provide the power and strength, while your wrist provides the control. Placing one or both thumbs on the back of the blade allows you additional control for finer, more detailed work. Make sure the blade is facing down. Otherwise you are likely to give yourself a serious cut.

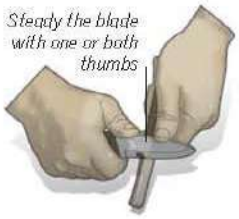

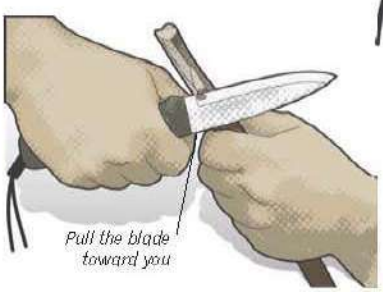

BACKHAND GRIP

This is the same as the forehand grip, except the blade faces up toward you. You can use it when you want to see and control exactly where the blade is going, or when you're cutting something and a follow-through would injure you—for example, when sharpening a stick.

CHEST LEVER

This is an extremely effective technique in which you hold the knife in a backhand grip and use the muscles of your arm and chest to provide power and strength. You can cut close to your body while your wrists control the cutting action.

SAFELY PASSING A KNIFE

A group may only have one good knife for everybody to use, so you'll need to pass it to each other frequently. Pass the knife handle first, with the blade up. This simple and safe method is easily learned and can be adopted by everyone.

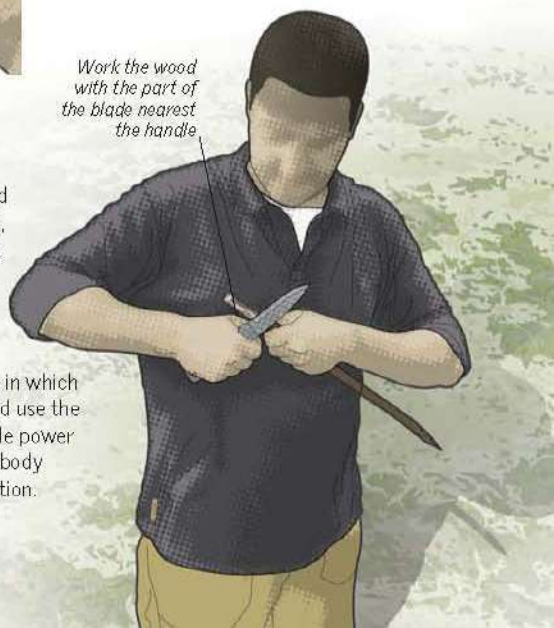

TREE-FELLING WITH A BLADE

For shelters and fuel, the largest tree you are likely to cut down will be about 4 in (10 cm) in diameter. For anything larger you'd need an ax, but for smaller ones use a saw (see pp. 150–51). If you don't have a saw, use the blade of your bushcraft knife or a machete or parang. Whichever blade you use, keep its lanyard around your wrist.

1 Figure out the lean of the tree and start your first cut on the opposite side. This stops your blade jamming in the cut as the tree leans further.

- Hit the top of the blade with a thumping stick (club).
- Move the blade about ⅝ in (15 mm) to one side of the first cut, and then angle the blade at 45 degrees toward the first cut.
- Continue to alternate this action from side to side, cutting a V into the wood.

WARNING!
Ensure your blade is sharp. Keep what you're cutting between you and the blade, with the blade pointing away.

SPLITTING FIREWOOD

When you split wood into fuel for a fire, use the largest blade you have, work on a solid surface, and avoid wood that has large knots as these prevent the blade from making a clean cut.

SAFE WORKING AREA
Chop wood in a safe area that other people won't wander into and where there are no snagging hazards or objects to impede your swing and knock your knife off-target.

LARGE LOGS

If you only have a bushcraft knife to work with, and you need to split large logs into smaller pieces, the following method works well.

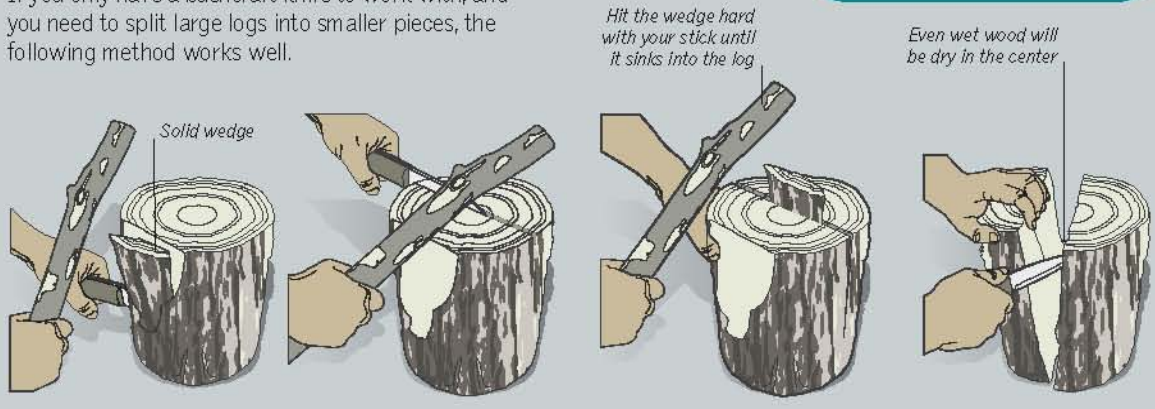

1 Hit the back of your knife with a stick and cut a solid wedge from the side of the log.

2 Use your knife and a thumping stick to start the initial cut across the top of the log.

3 Hit the wedge into the initial cut. Keep doing this until the log splits. Use more wedges if you need to.

4 If the log does not split completely, force it open with your hands or pry it with your blade.

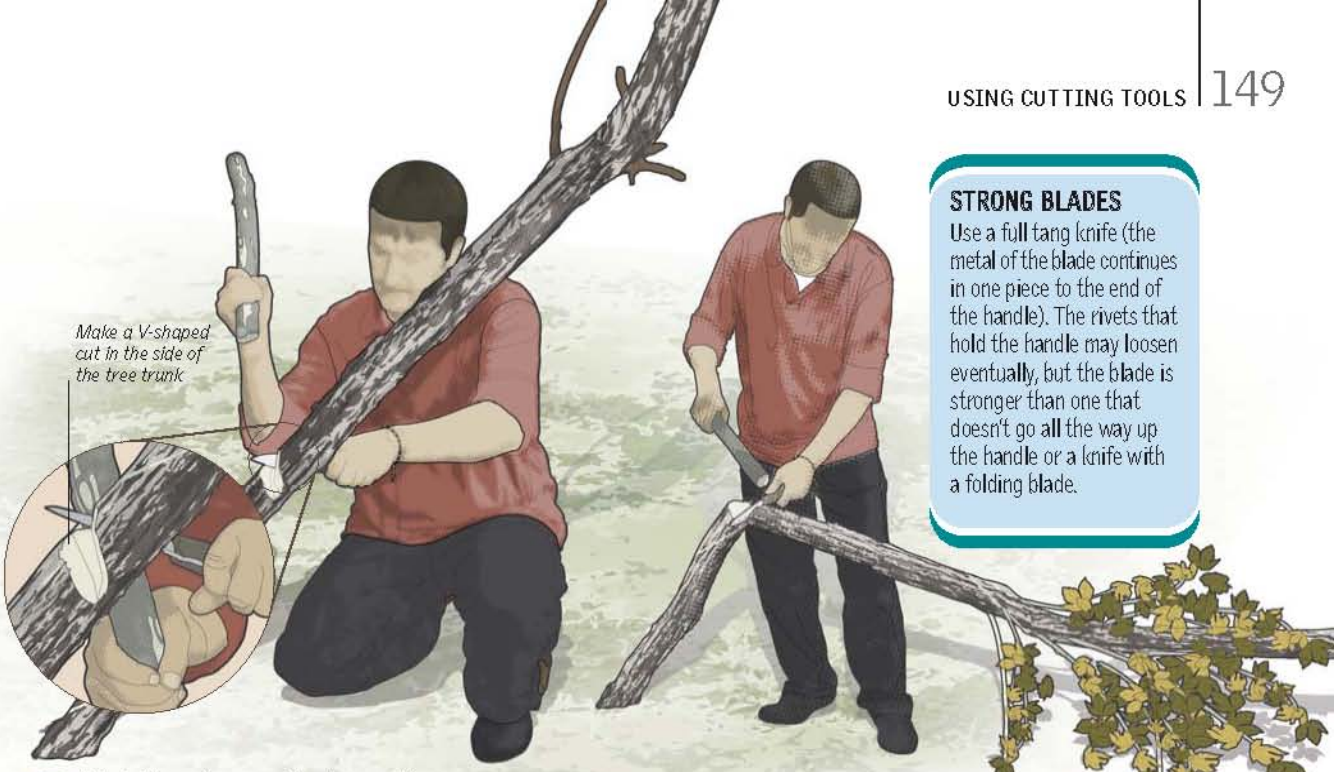

STRONG BLADES
Use a full tang knife (the metal of the blade continues in one piece to the end of the handle). The rivets that hold the handle may loosen eventually, but the blade is stronger than one that doesn't go all the way up the handle or a knife with a folding blade.

2 Work the cut around the tree until you've almost reached the middle from every direction except from below.

- If you try to cut completely through the wood from one side it will get increasingly difficult as the blade goes deeper.

3 Once the cuts are deep enough to weaken the wood, you will be able to snap the tree in two.

- Take care that the tree doesn't fall on your feet.

SMALL LOGS

Splitting a small log can be more tricky than splitting a large log, because it's difficult to balance it safely while you hit the blade with your thumping stick.

1 Place the log on a solid surface and put your knife in the middle of the end of the log you're about to chop.

- Hit the back of the blade firmly with your stick.

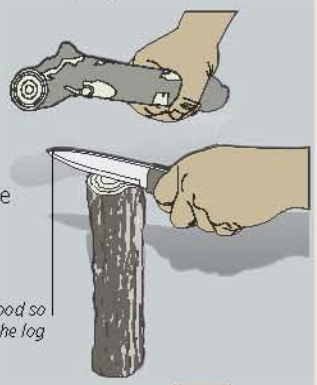

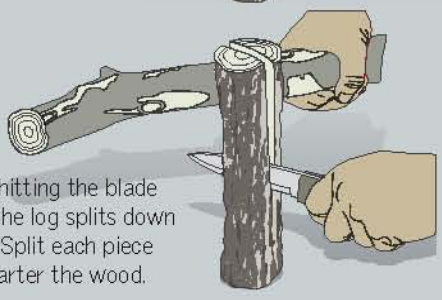

2 Keep hitting the blade until the log splits down the middle. Split each piece again to quarter the wood.

4 To trim the branches from the fallen tree, start at the bottom and work your way to the top. This allows you a clear, short swing for your thumping stick.

- Keep the tree between you and the blade, and trim each branch in the direction of its growth.
- If you're using a machete, you don't need a thumping stick, as the heavy blade has enough momentum of its own.

USING A SAW

A saw is much better than a bushcraft knife for larger jobs, such as preparing logs for a fire or cutting down the poles needed to build a shelter. There are several types of pocket saw available that are lightweight and take up very little room in your backpack or survival tin.

USING A POCKET CHAINSAW

Using a pocket chainsaw to cut down a small tree is safe and not especially hard work. You can use the same saw to trim the branches and then cut the wood into manageable sections.

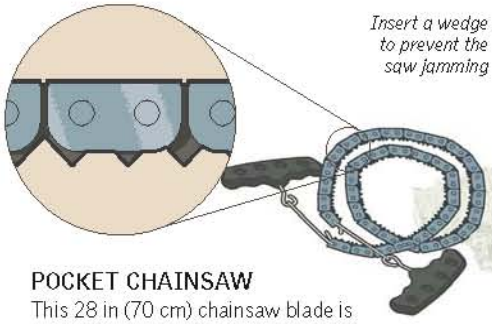

POCKET CHAINSAW
This 28 in (70 cm) chainsaw blade is compact and has attachment rings and handles. It can cut through a 3 in (7.5 cm) diameter tree in under 10 seconds.

1 Determine the natural lean of the tree and start to saw on the opposite side.
- Pull the saw back and forth.
- Knock a wedge in behind the saw when halfway through to stop the tree from jamming the blade.

2 Slow down when the saw has nearly cut through the tree.
- Allow the weight of the tree to snap the cut and break it in two.

3 Use the same back-and-forth rhythm to trim the branches.
- Clear the branches away as you trim them from the tree.

4 Saw the trimmed trunk into a measured section.
- When you're halfway through, prop up the trunk to make sawing easier.

FOLDING SAWS

Folding saws are strong, light, and convenient. Some, such as the Sandvik Laplander, have a serrated blade that folds into the handle (see right). Others, such as the backpacker's (see below), can be dismantled for stowing.

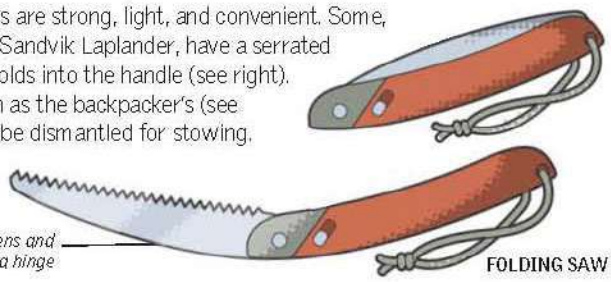

FOLDING SAW

BACKPACKER'S FOLDING SAW

These saws are generally lightweight and strong, but are not as effective as the pocket chainsaw. If space is tight, then you can pack only the blade and improvise a bow saw, using natural materials found in the woods.

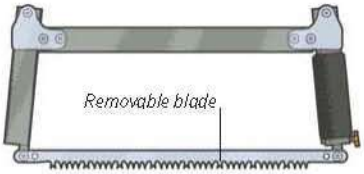

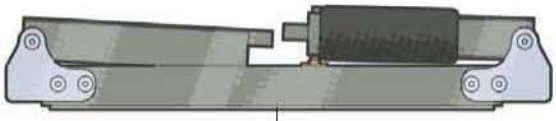

BACKPACKER'S FOLDING SAW

COMMANDO SAW

This saw takes up very little space and consists of a serrated wire blade with a swivel ring at each end. It's a useful piece of survival equipment, although it tends to snag and snap easily unless you keep it completely straight. You can easily convert it into an improvised bow saw, and, with a little more effort, make it into a very effective tension saw.

COMMANDO SAW

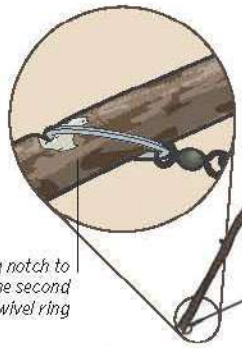

MAKING A BOW SAW

Cut a sapling approx 1 in (2.5 cm) in diameter and 3 ft (90 cm) long with a forked end. Hook the swivel ring of your commando saw over the fork and place the forked end on the ground. Ease the sapling into the other ring and make a notch for the ring to fit in.

IF YOU HAVE NO TOOLS

Even without any cutting tools, it's still possible to safely break wood down to size. The answer is snapping—or even burning—the wood. Don't be tempted to stand branches and sticks against a log or rock and then stamp on them in the middle. In a genuine survival situation your feet are your only means of transportation so don't do anything that risks injuring them.

SNAPPING THE WOOD

You need to limit the size of wood you're trying to break—don't try to snap branches that are too broad. Use wood that's dead, because green living wood just bends and refuses to break completely.

- Find two trees that are 1–2 ft (30–60 cm) apart—or a tree with two trunks close together—to use as a fulcrum for the wood. Place the wood in the fulcrum.
- Adopt a firm and steady stance and, with one foot in front of the other, pull the wood toward you until it breaks. Pulling toward you keeps your center of gravity over your spread feet, while pushing away from you causes you to stumble forward when the wood breaks.

BURNING THE WOOD

If you have a fire going, place the wood on the fire where you want it to break. Leave it on the fire until it burns all the way through. Alternatively, when it has burned halfway through, use the snapping method described above—but don't work on the wood while it's still burning or hot.

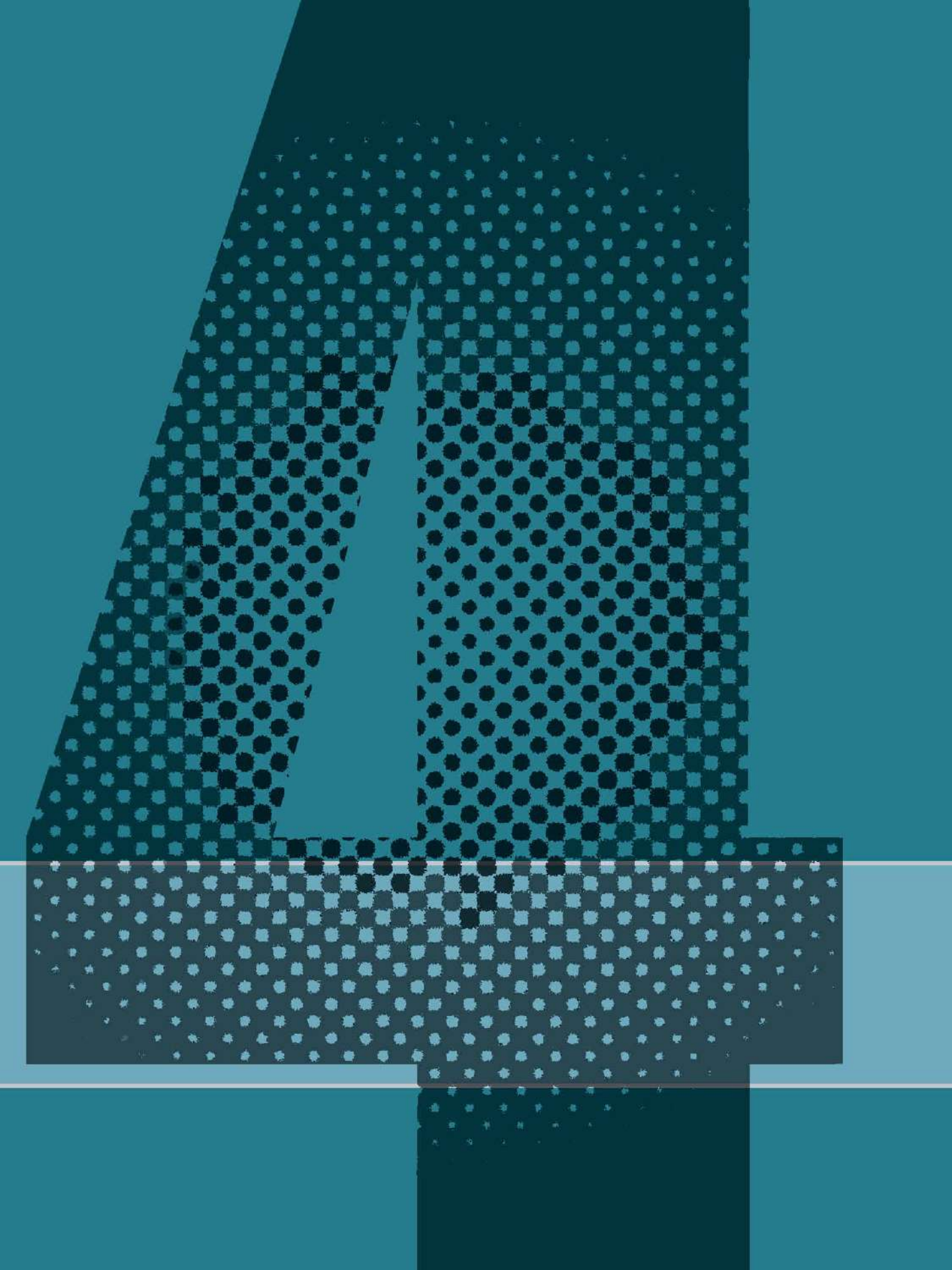

TAKING SHELTER

TAKING SHELTER

A shelter is your primary means of protection in a survival situation. It can be somewhere to simply keep dry during a sudden downpour or a place to spend several nights while out in the wilds. It's important, therefore, that you understand how to correctly construct your shelter. After all, its effectiveness could mean the difference between relative comfort and abject misery!

In a survival situation, it's important to play a proactive role in your own continued survival and rescue. You can only achieve this effectively if you're in a good condition—both physically and mentally. A day or night spent exposed to the elements will determine how well or how badly you perform the next day. A restless and uncomfortable night, for example, will lead to sleep deprivation, which can make you moody, irritable, and easily frustrated. This in turn can lead to a lack of concentration and irrational thinking—none of which are helpful in a survival situation.

Cold nights spent in the wilderness seem to last longer than they should, so time spent ensuring that you can

In this section YOU WILL DISCOVER...

- when to **lie low** in a hollow or **camp in a cave...**
- how to **modify** your **poncho** to put up a **shelter...**
- the difference between a **bothy** and a **bivy...**
- how to **bed down** with **duck down** or **wake up** in a **wickiup...**
- the **importance** of using your **noggins** to **secure** your shelter...
- how to **configure** your **fig leaves** to keep out the **rain...**

always enjoy as warm, dry, safe, and comfortable a night as possible is rarely wasted.

With shelter-building—as with other areas of survival—you should always apply the principle of expending the least amount of energy for the maximum amount of gain. This could mean simply exploiting the options nature has already provided, such as caves or hollows. Finally, plan for the worst conditions—it may be sunny when you build your shelter but pouring with rain at three in the morning!

On any expedition, you should aim to carry equipment—such as a basic bivy or a tarpaulin—that could be used to form a shelter should you need to protect yourself from the elements.

A SPACE BLANKET, however, is quite simply an essential piece of gear—if you pack nothing else, make sure you carry one and keep it with you at all times. Space blankets pack down very small and are available with one side silver (to reflect heat) and one side bright orange (to aid location). No matter what environment you're exploring, a space blanket can offer you immediate protection from the elements.

With a little ingenuity, a space blanket can also be converted into a basic shelter. You will find instructions on how to build a variety of shelters throughout this chapter, but always remember the following general tips, too:

- First look for natural sheltering places
- Look out for hazards such as signs of flooding
- Construct your shelter well before nightfall
- Clear the ground before you start to build
- Position the entrance at right angles to the wind
- Build for the worst possible weather conditions
- Make sure your shelter is sturdy and secure
- Don't over-exert yourself as you build
- Raise your sleeping area off the ground
- If on a slope, sleep with your head higher than your feet

> The **effectiveness of a shelter** can mean the difference between relative comfort and **abject misery**

HOLLOWS AND CAVES

IF YOU DON'T HAVE a shelter of your own, look for shelter opportunities that nature has already provided for you, such as a cave or a natural hollow in the ground. However, avoid natural hollows in low-lying areas that may flood and, if you're on a slope, keep well away from hollows exposed to run-off water.

TOOLS AND MATERIALS

- Sturdy digging stick
- Strong branches (or poles) and logs
- Saw, such as pocket chainsaw
- Tree bark, turf, leaves, forest mulch, moss, bracken, and pine boughs for the roof

BEDDING

Cover the floor of a shelter with some form of bedding to prevent your body from losing heat to the cold ground through conduction. After all the energy you've used to build the shelter, why waste it on an uncomfortable night? Use the driest materials you can find (see list below).

MAKE YOURSELF COMFORTABLE

Here are some ideas to help you enjoy a good night's sleep in your shelter:

- Collect twice as much bedding as you think you need: 6 in (15 cm) of holly, for example, compresses to about 2 in (5 cm) after you've been lying on it for a while.
- Place logs around the bedded area to stop the bedding from shifting and spreading.
- If the ground slopes, sleep with your head higher than your feet.

BEDDING MATERIALS

- Feathers, such as duck and goose down, are the best bedding as they retain warmth but not moisture.
- You're unlikely to find enough feathers, so also try pine and spruce boughs, dry leaves, moss, bracken, grasses, and holly (cover it to prevent it from pricking you).

NATURAL HOLLOWS

A hollow that may be too shallow to sit in can still provide shelter from the wind if you lie in it—and, with a basic roof, keep you dry too.

1 If the hollow isn't deep enough to lie in, remove some soil with a digging stick.

- Beware wet soil, as it could be prone to flooding.
- If the ground is too hard to dig, build up the sides with logs for extra depth.
- Lay bedding material on the ground before continuing.

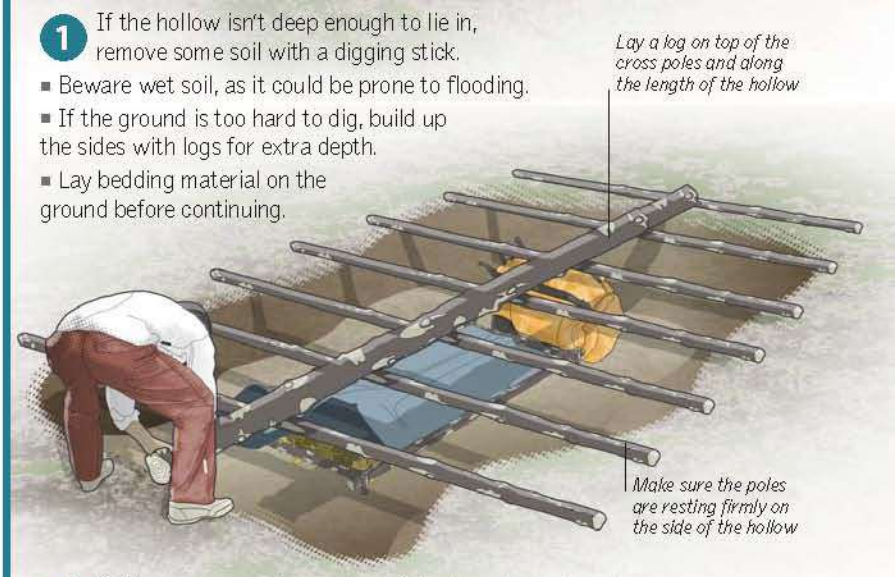

2 With your saw, cut a number of sturdy poles or branches (see pp. 150–51) that are long enough to place across the hollow to create a supporting roof.

- Cut a thicker, longer log and place it across the poles. This will create height and slope for a pitched roof. Make sure the log isn't too heavy.

TAKING SHELTER IN CAVES

Caves are ready-made shelters that are usually dry and secure. There are some dangers, including animals, poor air quality, and water. Many caves are outlets for streams or are connected to pools. Don't go any further in than you can see, as there may be hidden drops, slippery surfaces, and crevasses. Avoid old and disused mines, as they may be prone to collapse.

ANIMAL DANGERS

Many animals, such as bears, bats, insects, spiders, and snakes, shelter in caves, so check the ground around the entrance and inside for signs of tracks and nests. Vampire bats can be dangerous as they are known to carry rabies. Bat droppings (guano) form a thick layer on the ground and, in large quantities, can be highly combustible.

POOR AIR QUALITY

If a cave makes you feel light-headed or nauseous, leave at once because the air may be foul due to excess carbon dioxide. Other warning signs are an increased pulse and breathing rate. If a flame starts to dim, or turn blue, leave the cave immediately—this may not be a sign of excess carbon dioxide but a lack of oxygen, which can be worse.

LIGHTING FIRES

If you build a fire at the front of a cave the smoke can blow back inside and the fire can block your exit if it gets too big. Build a fire toward the rear of the cave, but make sure that there is a sufficient flow of air for the smoke to escape. You can erect a small fire reflector or make a pile of rocks at the entrance to reflect heat back into the cave.

COASTAL CAVES

Coastal caves share the same dangers as any other cave, but they are also subject to the tides, which can rise very quickly—in some areas they can advance hundreds of yards in a matter of minutes.

THE DANGERS OF COASTAL CAVES

When you're assessing a coastal cave, look out for the following warning signs of flooding during high tide:

- A line of seaweed, driftwood, or flotsam and jetsam—either in the cave or further back on the beach either side of the entrance.
- A damp or wet smell. The cave may even look wet if the tide has recently gone out.
- Rock pools in the cave or in front of it.

3 Place shorter branches, sticks, or poles over the top of the log and cross branches to create the pitched roof.

- Pack these materials tightly together and make sure they're firmly in position.

4 Leave an entrance at one end of the shelter with enough space to get in and out without disturbing the roof.

- Ensure the entrance doesn't face the wind—an angle of 90 degrees to the wind is best.
- Insulate the pitched roof with as much foliage as possible.
- Start the final insulation layer at the ground, working up to the top of the roof to create an overlap for rainwater to run off.
- In wet conditions secure your poncho or space blanket over the top.

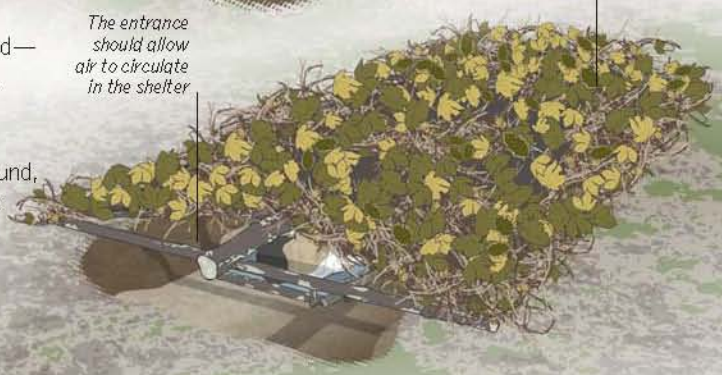

USE-ANYWHERE SHELTERS

WITH A PONCHO OR BIVY SACK (see pp. 160–61), you can make a shelter for use in any environment and in any conditions. You can also adapt any type of material—a survival space blanket, ground sheet, or tarpaulin—in the same way.

TOOLS AND MATERIALS
- Poncho, bivy sack, groundsheet, space blanket, or tarpaulin
- Cordage or elastic bungees
- Pegs; stick or pole
- Rock or log for hammering
- Survival knife or penknife

ONE-POLE PONCHO

This shelter is very quick and easy to erect and take down—useful when energy levels are low or when you just want to get out of the rain. If you need to be rescued, put the most visible color on the outside to increase the chance of being spotted.

1 Cut three 2 ft (60 cm) lengths of cordage and one 6 ft (2 m) length.
- Tie the shorter lengths into loops using an overhand knot (see p. 143).
- Tie the long length to one corner with a Siberian hitch knot (see p. 144).
- Fasten the cordage loops to the corner grommets.

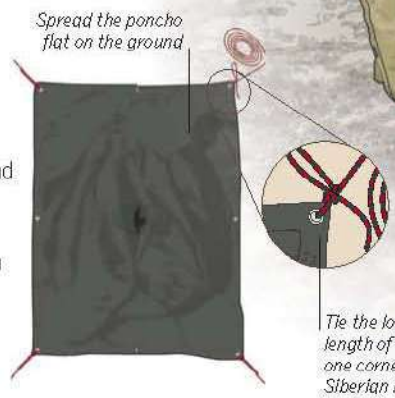

2 Tie the long cord to a tree with another Siberian hitch knot at a height of 3 ft (1 m) off the ground.
- If no tree is available, run the cordage over a stick (staked into the ground) and peg it down.

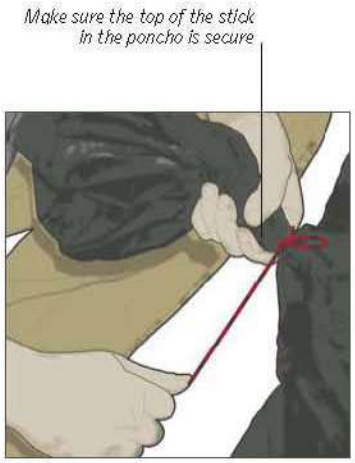

3 Place one end of a 3-ft (1-m) long stick in the hood of the poncho.
- Seal the hood by tying a simple knot.
- Wind several turns of cordage around the knot and secure it with an arbor knot (see p. 142).

4 Peg out the left and right corners, then raise the stick to form a dome.
- Pull the rear corner taut and peg it out to add tension, making the structure rigid.
- To enter the shelter, curl your body around the center pole.

CORDED A-FRAME

This poncho A-frame creates a tentlike shelter between two trees. If you have enough cordage, a single ridgeline can be tied between the trees; if not, use two cordage loops, as shown below. Alternatively, you can use a ridgepole at the same height.

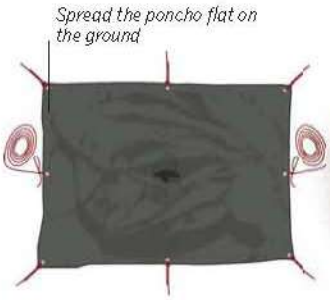

1 Find two trees 2 ft (60 cm) further apart than the length of your poncho.

- Attach cordage loops 1 ft (30 cm) long to each grommet on the longer sides of your poncho and 3-ft (1-m) loops to the end ones.
- Tie a knot around the hood to seal it.

2 Tie the end of one of the longer cords to a tree with a double chain fastening knot (see p. 143) at a height of 3 ft (1 m).

- Tie the other long cord to the second tree with a taut line hitch knot (see p. 145) and adjust the hitch to pull it tight.

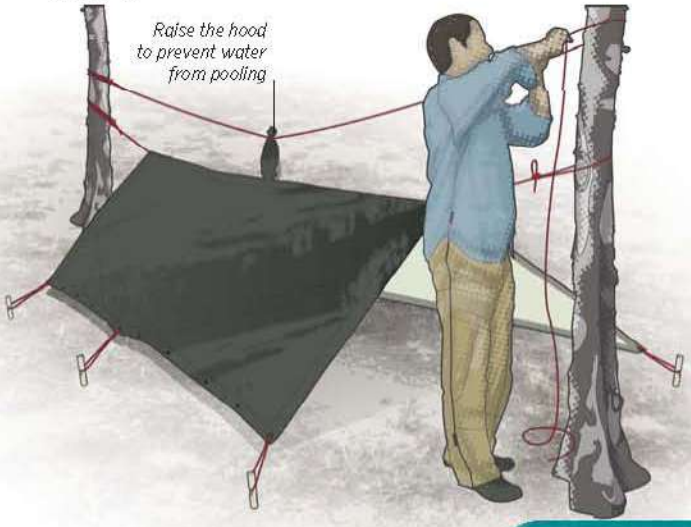

3 Peg the middle loop of each side to pull the poncho taut. Peg out the four remaining loops, pulling tight as you go.

- If you have enough cordage, raise the hood by tying it either to a second horizontal line or to a vertical line passed over a branch above.
- Place long boughs along each side of the shelter, with shorter boughs at each end to hold in bedding material.

NEED TO KNOW

- Block one end of the shelter with branches, boughs, and mulch to help retain heat.
- If your material has no grommets, use button ties instead (see p. 161).

TENT PEGS

You can make your own tent pegs and then either reuse them or replace them with new ones when necessary. Discard old pegs that naturally decompose over time.

MAKING TENT PEGS

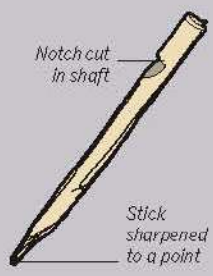

- Pegs should be about 9i n (22.5 cm) long and 1 in (2.5 cm) in diameter. Anything less will either snap under pressure or pull out in high winds.
- Cut wood for the pegs from green trees.
- Don't use wood from the ground as it will be in some stage of decay.
- Dry out new pegs slowly over the embers of a fire to harden them. Remove them when they're a light brown color. Wet wood sounds deep and dull when you tap it; as it dries the sound gets higher and crisper.
- Never drive a peg in with your foot. Use a stone or a heavy stick instead. If you misjudge it, the peg may stab you through the ankle or pierce your boot and even your foot.

ALTERNATIVE DESIGN

If you're in a group and you each have a poncho, you can clip two ponchos together and use one as a built-in floor.

USING TWO PONCHOS

First clip the poppers of both ponchos on their two long sides to form a tube. The top one is the A-frame, the bottom one the groundsheet.

- Pack all your bedding material under the bottom poncho and hold it in place by laying branches at either end.
- Assemble the shelter with the top poncho in the same way as the single poncho version.

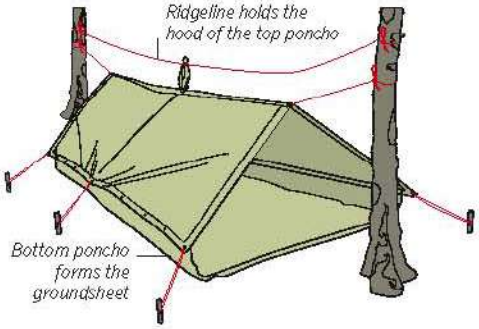

BIVY PUP TENT

A bivy sack is a lightweight, waterproof cover for a sleeping bag and is made out of a breathable material that cuts down condensation. It's an essential component of ultralight camping (see p. 52) and can very easily be turned into a single-person pup tent. The advantage of making a tent from the bivy sack is that, in wet conditions, the rain runs off the side and doesn't seep to the inside as quickly as it does when the bag is flat on the ground.

TOOLS AND MATERIALS

- Bivy sack
- Saplings
- Button ties
- Wooden pegs
- Cordage and knife
- Rock or log for hammering

1 Lay the bivy sack flat on the ground where you intend to set it up.

- If you have the flap of the hood on top it will form a door that hangs down. However, if the flap is on the bottom you can tie it up as a door.
- Attach a button tie (see panel, opposite) with a loop to each of the bottom corners of the bivy sack.
- Peg the button-tie loops into the ground.

2 Attach a button tie with a long length of cordage to the center of the bivy sack's opening and hold it up.

- Pull down the right and left sides of the opening and place a button tie and loop where they meet the ground.
- Peg out the button-tie loops into the ground.

4 Cut a pole the same height as the entrance (you could use a hooped sapling or crossed sticks instead).

- Stand the pole at the entrance, run the middle button-tie cord over the top, and peg out the cord tightly.
- Readjust all the pegs to ensure tight and even tension.

5 When you want to get into the bivy pup tent, simply remove the opening pole and climb in feet first.

- Make sure you avoid the long cord as you wriggle in.

OTHER SHELTERS

There are other temporary shelters you can erect by adapting the techniques, tools, and materials outlined for both the use-anywhere shelter and the bivy pup tent. Which shelter you choose depends on whether you have a poncho, tarpaulin, or bothy bag.

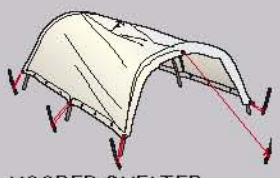

HOOPED SHELTER

If there are saplings or soft wood trees, you can form three hoops and peg out your poncho over them.

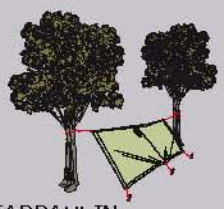

TARPAULIN

Attach one side of the tarp to a line between two trees and peg the other end down to the ground at an angle.

BOTHY BAG

A bothy bag, or storm shelter, is a large, light, nylon bag that protects one or two people from the elements.

3 Place a hooped sapling (or two crossed sticks) over the bottom of the bivy sack.

- Attach the button tie that's 18 in (45 cm) from the foot of the bivy sack to the hoop to raise it and provide more space for your feet.
- Tightly peg out the cord from the hoop.

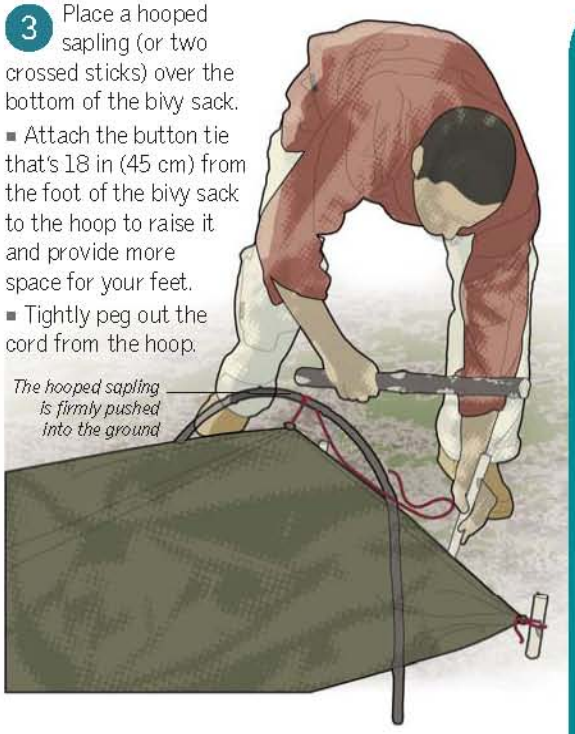

MAKING A BUTTON TIE

Button ties provide a simple, secure fastening for staking out ponchos and tarpaulins that have no loops or grommets to which you could attach a line. As you don't have to cut holes in the material, it remains waterproof and less prone to ripping.

1 Encase a small, round, smooth stone (or similar) in material to form a "button". Prepare cord, with the open loop of a slip knot (see p. 142) at one end.

2 Place the open loop of the slip knot over the neck of the button and pull it tight.

3 At the other end of the cord make a simple loop to place over your shelter pegs.

6 After you have climbed in, replace the pole under the long cord.

- Use the hood of the bivy sack as a door and open and close as necessary.

FOREST SHELTERS

IF YOU FIND YOURSELF in a survival situation in a temperate forest without a poncho, bivy sack, or tarp, you can make a shelter from natural materials. For example, you can construct a lean-to or an A-frame, or simply adapt a fallen tree. A forest lean-to is easier for a group to build because you can share the workload and don't have to use up too much precious energy.

TOOLS AND MATERIALS
- Poles and stakes
- Saw and knife
- Cordage
- Saplings
- Foliage for the roof covering

BUILDING A FOREST LEAN-TO

A lean-to has a sloping roof that leans against a horizontal ridgepole. It works best on a flat area between two trees or vertical supports secured firmly into the ground. If you're building a lean-to for a group, ask everyone to lie down next to each other and add an additional 6 in (15 cm) per person—this will establish how wide the shelter needs to be.

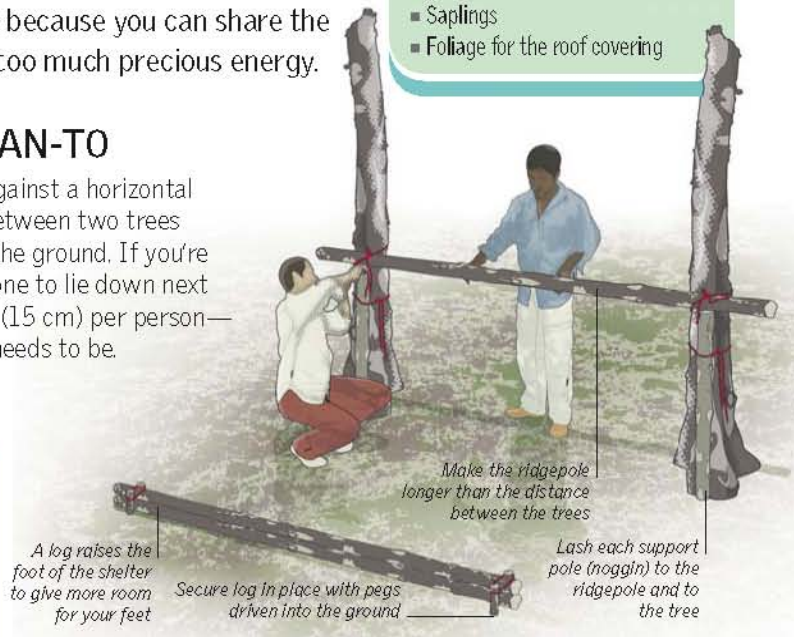

1 Place the ridgepole against the trees at the height you require and lash it with an arbor knot (see p. 142) to both trees.
- Place a support pole (noggin) under the ridgepole and lash to the trees.
- Firmly peg a log where you want the foot of the shelter to be.

4 Apply the covering to the woven skeleton of the roof and walls. How much you apply depends on the weather you're experiencing.
- Lay large materials—for example, pine or fir boughs, large leaves, clumps of moss—as a base to stop smaller materials from falling through.

BUILDING IN A FOREST

Temperate forests, whether they consist of deciduous or coniferous trees or both, provide many opportunities for finding or building a shelter. Expend the least amount of energy for the most amount of protection—first see what nature can provide before building a shelter yourself.

POINTS TO REMEMBER

If you do need to make a shelter in a forest, here are some tips:
- Select a place close to water but away from water hazards, such as flooding, animals, and insects.
- Before you start, check for hazards such as deadfalls (sudden branch falls), rockfalls, and flash floods.
- Select a place that gives you maximum protection from the elements and with all the materials you need nearby.
- Collect everything you need before you start building and before it gets dark.
- Make sure your aids to location can be seen or activated quickly (signal fires, heliograph, flares, radio signals).
- Think safety: if possible, wear gloves to clear leaves and debris from the ground to protect you from spiders and snakes.
- Layer down your clothing as you work to prevent overheating.
- Make your shelter as waterproof as you can and ensure your bedding raises you at least 4 in (10 cm) off the ground.

TREE SHELTERS

The branches, roots, or trunk of a fallen tree can make a basic shelter. Make sure the tree is safe before you use it, as a further fall could injure you. Dead trees are probably dry, so don't light a fire under them. Check the base of the tree for snakes, spiders, and nests of insects.

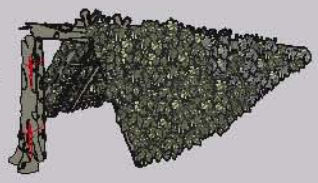

BROKEN BRANCHES

A branch that has snapped from a tree trunk but hasn't yet fallen makes a good ridgepole. Prop smaller branches and roofing material up against it.

UPTURNED ROOTS

The base of an uprooted tree can give good protection from the elements. Make sure it's safe and won't flood if it rains, then add boughs and foliage.

FALLEN TRUNK

Lay sticks or branches against a fallen trunk to form a pitched roof. Weave saplings across them and cover with bark, turf, moss, leaves, and mulch.

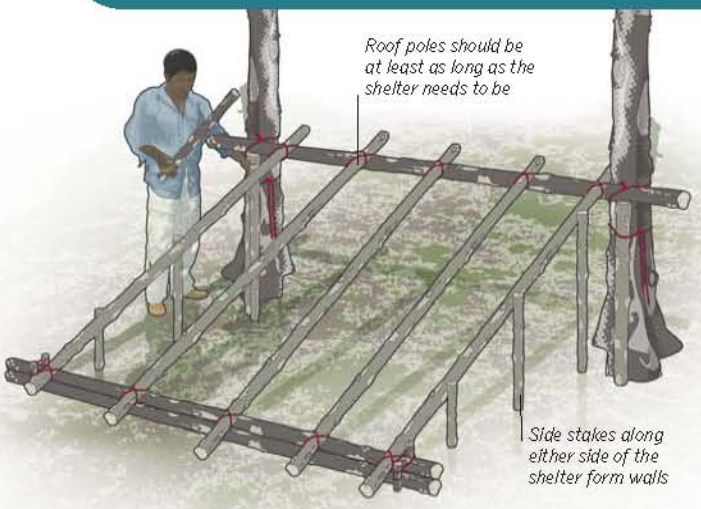

2 Place five roof poles evenly along the ridgepole so that they slope back to the foot of the lean-to.

- Keep the outside poles between the trees to stop them moving outward.
- Lash each roof pole to the ridgepole.
- Hammer in side stakes alongside the outside poles.

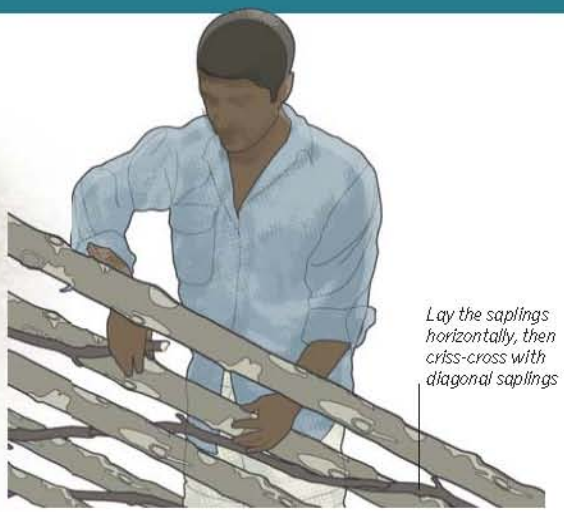

3 Weave saplings in and out through the roof poles, working across and down from the ridgepole.

- Alternate the weave of each row of saplings, first horizontally and then diagonally.
- Weave smaller saplings through side stakes to form walls.

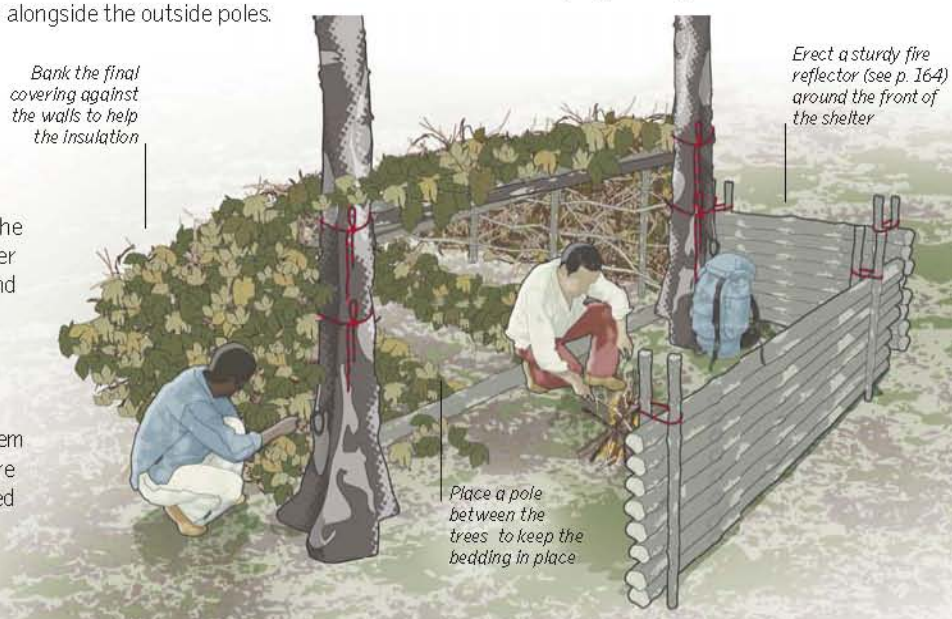

5 With everyone in the same lean-to shelter you need to light and tend only one fire.

- Layer the inside floor with dry material to form bedding.
- Establish a watch system during the night to ensure that the fire is maintained and controlled.

FOREST A-FRAME

A-frames can take several hours to build, but if you're staying in one place for a few days it may be worth expending the energy. One-person A-frames are relatively simple to construct and are easy to keep warm. They can be adapted into multi-person shelters to accommodate groups of people.

TOOLS AND MATERIALS

- Poles of various lengths
- Saplings
- Knife
- Saw (or ax)
- Cordage
- Covering for walls—such as pine boughs, leaves, and moss

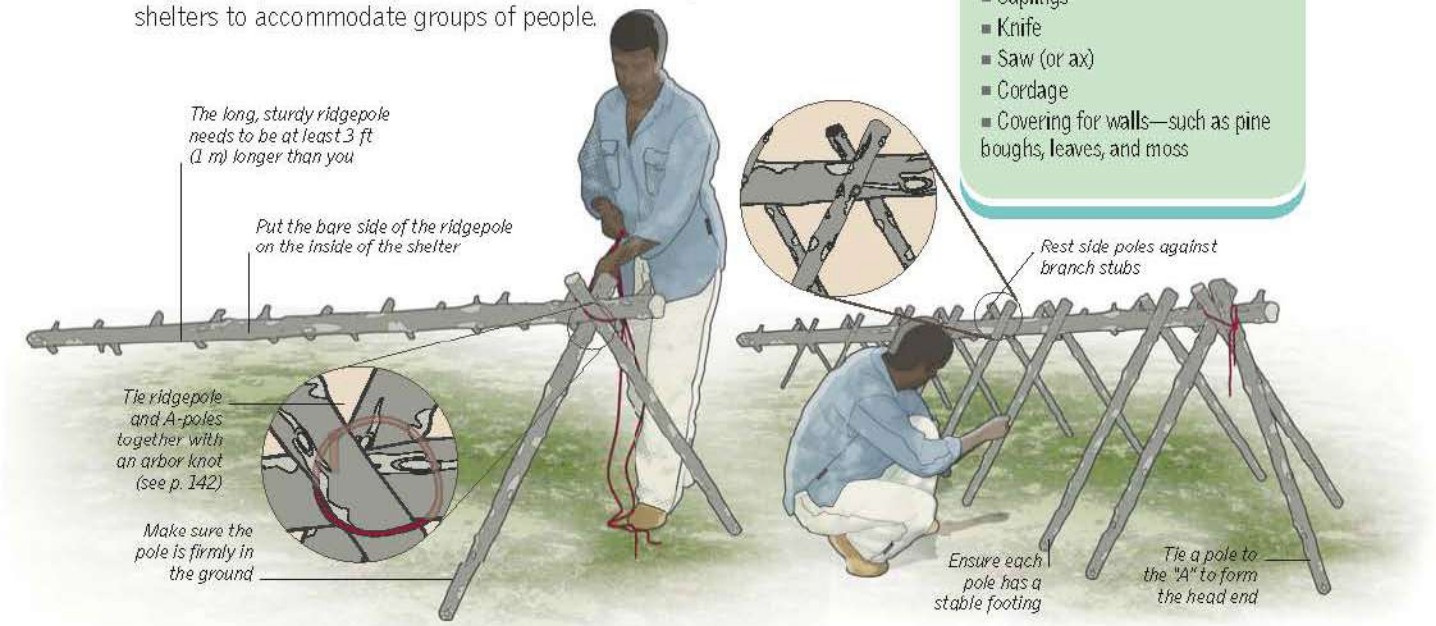

1 Smooth one side of the ridgepole with your knife, leaving one side bare and one with branch stubs.

- Hammer two poles into the ground to form an "A".
- Rest one end of the ridgepole on top of the "A" and the other end on the ground.

2 Lay a series of poles against the branch stubs on the ridgepole. These will be of decreasing height as you work your way from front (the "A") to back (the ground).

- At each junction, rest the side poles against the stubs or tie them to the ridgepole with an arbor knot (see p. 142).

FIRE REFLECTOR

A fire radiates heat in all directions—up, down, and 360 degrees around. A fire reflector makes a fire more efficient as it directs heat into your shelter.

Construct your reflector so that you can build a fire about 3 ft (1 m) from the entrance to your shelter. You can sit in the entrance and tend the fire at arms' length—any closer and you would be too hot and the shelter might catch fire; any further away and you lose too much heat. An L-shaped end to the reflector will retain more heat.

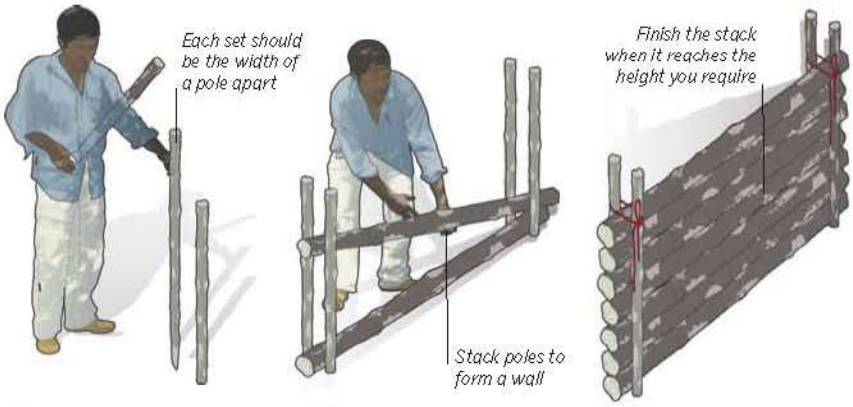

1 Check the air flow will be across the proposed reflector (to disperse smoke).

- Hammer two poles into the ground. Put a second set in at poles' length apart.

2 If the fire is going to be close to the reflector, use green wood to prevent the heat setting it on fire.

- Put poles between the sets to form a wall.

3 Tie the top of the upright poles together.

- Make sure the wall is as long as the shelter's entrance to retain heat and keep out wind and rain.

ALTERNATIVE SHELTERS

Two other options are a front-opening A-frame and a one-person lean-to. For the former, secure a ridgepole or line to a tree at the height you want the entrance to be, then adapt the forest A-frame (see opposite). For the latter, tie the ridgepole to two forked sticks hammered into the ground and apply a cover similar to the forest lean-to (see pp. 162–63).

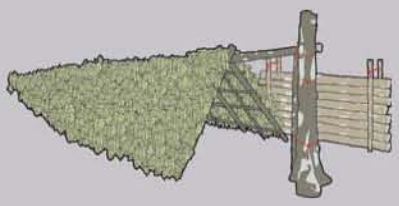

FRONT OPENING A-FRAME
Wedge a noggin under the ridgepole and lash it to the tree with an arbor knot to take some of the weight. Build a fire reflector to the side with the prevailing wind.

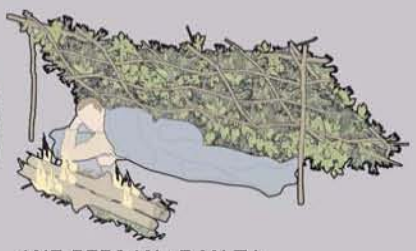

ONE-PERSON LEAN-TO
A small one-person lean-to, together with a long log fire built in front, provides you with a warm, dry, and comfortable shelter for an overnight stay.

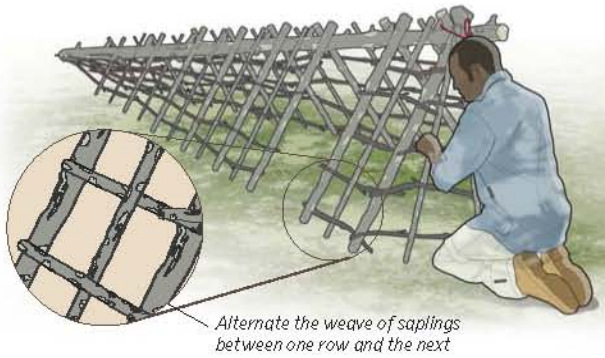

3 Weave four or five rows of saplings horizontally through the side poles to make the walls into a strong framework for the covering.

- Leave an entrance in the side near the front. Make sure it's large enough for you to sit inside comfortably.

4 Cover the framework with a thick layer of natural material, such as pine boughs, small branches, and sturdy twigs. Weave the materials into a "thatch".

- On each side, work your way gradually from the back of the shelter to the front, and from the bottom to the top.

NEED TO KNOW
Build your shelter so that the wind blows across the front of it at a slight angle, rather than into it. This maintains its effectiveness and prevents smoke entering from the fire. If the shelter backs on to the wind, there's a risk the wind will blow around it and curl into the entrance.

5 Cover the first layer with leaves, moss, and mulch, starting from the ground and working up to the ridgepole.

- Build a fire reflector (see opposite) in front of the entrance to help direct heat towards you.
- Stow your equipment and pack in the front of the shelter.

TROPICAL SHELTERS

TROPICAL RAIN FORESTS CAN BE the easiest environments for building shelters because of the wealth of available materials. Once you've found a suitable location just decide what type of shelter you need. The most important thing is to raise the shelter off the ground. Make it as comfortable as you can, because quality sleep is important.

TOOLS AND MATERIALS
- Poncho fitted with press-studs
- Cordage
- Long, heavy poles or wide bamboo stalks
- Spacer bars
- Knife, saw, machete, or parang
- Needle
- Strong, waxed, cotton thread

MODIFIED PONCHO BED

Most ponchos have press-studs to allow you to clip up the edges or to clip two ponchos together. They also have grommet holes that you can use as lashing points. Here, a poncho is modified into a tropical bed.

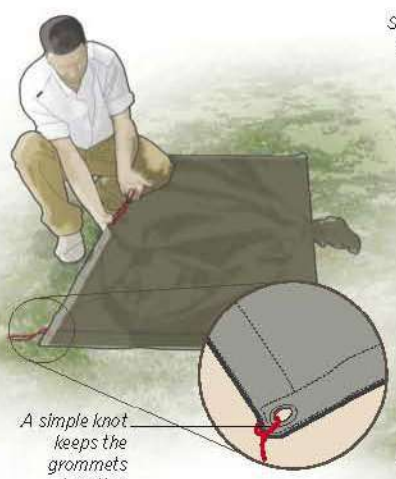

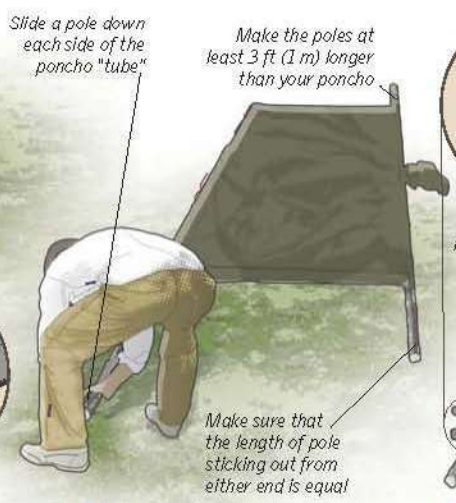

1 Clip the press-studs of the two sides of your poncho together to form a kind of "tube".
- Tie the grommets on the clipped side with pieces of cordage so that the press-studs don't come undone.

2 Cut two strong support poles that can take your weight.
- Insert the two support poles inside the "tube" of your poncho, one down each side.

3 Cut two "spacer bars" that will fit perpendicularly across the support poles or, if you can find them, cut poles with a natural fork in them.
- Tie the spacer bars into position using an arbor knot (see p. 142).

HAMMOCKS

Many types of hammock are designed for the kind of hot weather you get in the tropics, and are often fitted with mosquito nets. In fact, hammocks make the best overnight shelter, as long as you can find two trees the right distance apart. The Hennessy hammock is a good choice because it packs down small.

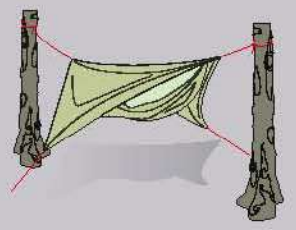

HENNESSY HAMMOCK
This ultralight, quick-drying hammock is both easy to set up and comfortable to sleep in (see p. 54). It's equipped with a self-closing door.

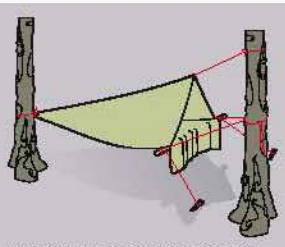

PARACHUTE SHELTER
Fold the panels of a parachute into a triangle. Lash a spacer bar to the opening and peg it out. Tie off the apex of the triangle further up the tree.

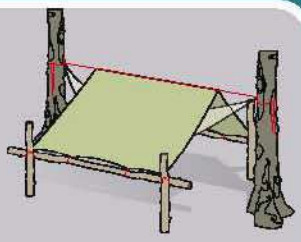

HAMMOCK AND TARP
Tie a ridgeline above the hammock. Drape the tarp over the line and anchor it on either side to a horizontal pole tied to two uprights.

ALTERNATIVE METHODS

You can easily modify other forms of sheeting, such as a tarpaulin, shelter sheet, and groundsheet. You will need a decent needle and plenty of strong, waxed, cotton thread (see Survival kits, pp. 60–61). There are two methods, which you can do either on the trail or as part of your prior preparation. Adding spacer bars keeps the poles apart and the sleeping platform tight. Both of these methods produce an excellent emergency stretcher.

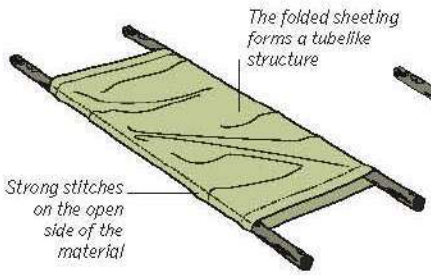

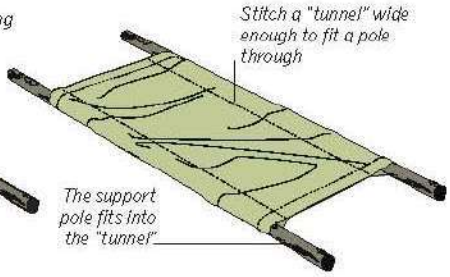

METHOD ONE

Lay the material on the ground and fold it once to create a wide "tube." Stitch the length of the open sides and then insert a pole along each side of the tube.

METHOD TWO

Fold the sheet and stitch a "tunnel" along both sides just wide enough to take a pole that will support your weight. Push the support poles through the tunnels.

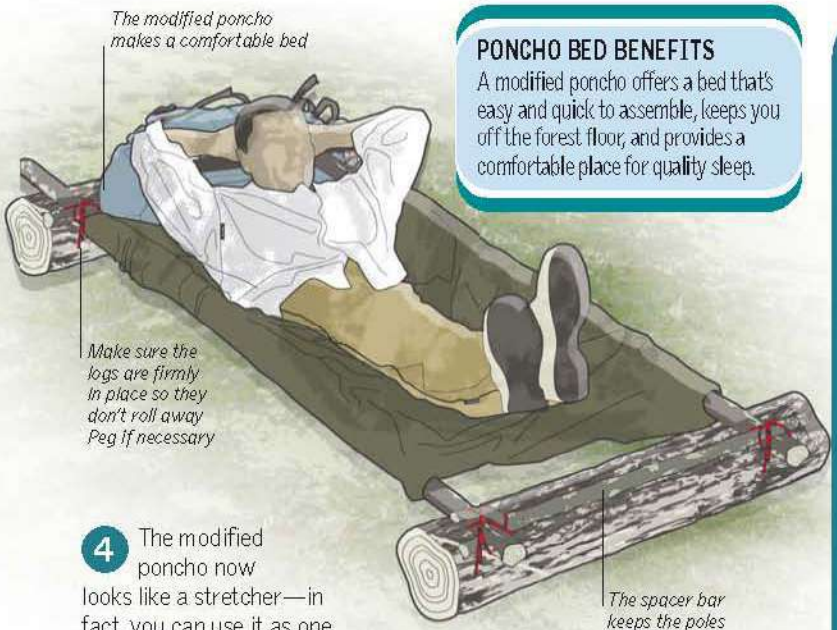

PONCHO BED BENEFITS

A modified poncho offers a bed that's easy and quick to assemble, keeps you off the forest floor, and provides a comfortable place for quality sleep.

4 The modified poncho now looks like a stretcher—in fact, you can use it as one in an emergency.

- Raise the poncho off the ground by resting it either on thick logs or sturdy poles.

ALTERNATIVE DESIGN

If you have a second poncho, shelter sheet, or tarp you can create a shelter to keep the rain off your modified poncho (see also Corded A-frame, p. 159).

- Tie a ridgeline between two trees and hang the poncho, sheet, or tarp over the top of it.
- If you're using a second poncho, tie its hood to another ridgeline.
- Peg out the four corners of the poncho, sheet, or tarpaulin.

BUILDING IN THE JUNGLE

Shelters in the rain forest need to be quick to erect, just big enough for your needs, and safe from animals.

CHECKLIST

If you're thinking about building a shelter in a tropical forest, the following tips may help you:

- Check for hazards such as deadfalls (sudden branch falls)—a major cause of injuries during military jungle training.
- You'll need a sharp cutting tool—ideally a parang or machete, but a decent bushcraft knife or small ax will suffice.
- Clear the ground around your shelter to deter animals. Use a makeshift brush, never your bare hands, to avoid being bitten by snakes or spiders.
- Build your shelter far enough off the ground to avoid being bothered by insects, snakes, or any other animals—particularly those that move around at night.
- Start building your shelter well before darkness sets in, which, in tropical zones, usually happens very quickly. Using a large knife in reduced light or by the light of a flashlight can be very dangerous.
- Work at a rate that you can manage. Humid conditions can very quickly lead to dehydration and heat-related injuries such as heat stress and heat stroke. Your body tries to keep itself cool by sweating, so don't work too fast, drink water frequently, and take regular breaks.
- Make your shelter secure—you don't want to make repairs at night.
- Inside your shelter use a full mosquito net and a mosquito head-net.
- A fire will deter insects and animals.

JUNGLE A-FRAME

An A-frame is relatively easy to make. If you have a poncho, tarpaulin, shelter sheet, groundsheet, or another type of sheet, you'll need to modify it first to make it into a bed (see pp. 166–67). If you do not have a poncho, lay branches across the poles to form a sleeping platform.

1 Cut seven long poles that will take your weight. Tie two poles together with an arbor knot (see p. 142) to form an "A" and tie the joint to a tree or branch of a tree (as shown).

2 Tie two more poles together to form an "A" for the other end.

- Position these in line with the first "A".
- The distance between the two "A"s needs to be at least 2 ft (60 cm) longer than your height.

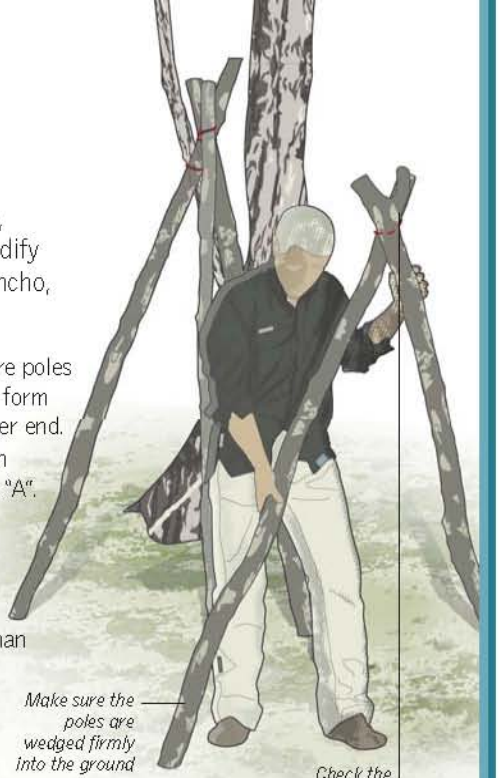

3 Put a ridgepole on top of the two "A"s.

- Tie this to both junctions for extra stability.

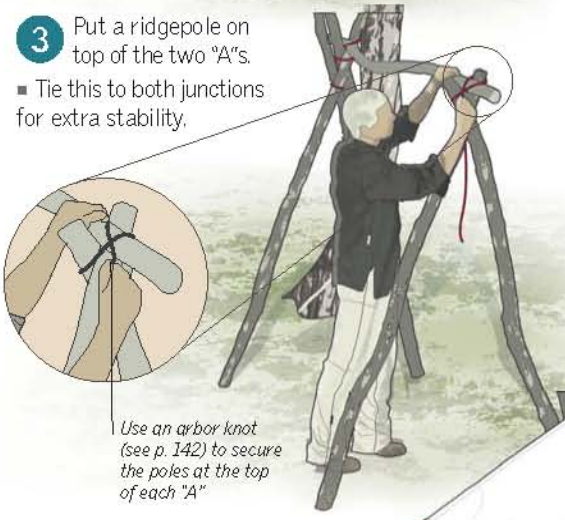

4 Position both poles of your poncho bed (see p. 166) on the outside of the A-frame.

- Move the poles down the outside of the frame legs until the sheet is taut.

5 Place a tarpaulin or shelter sheet over the ridgepole.

- Securely tie a guyline to each corner with a Siberian hitch (see p. 144).
- Run each guyline to a peg or tree and secure it with a taut line hitch (see p. 145).

TOOLS AND MATERIALS

- Long poles and ridgepole
- Cordage and pegs
- Knife, saw, machete, or parang
- Modified poncho bed
- Tarpaulin, groundsheet, poncho, or shelter sheet

BAMBOO LEAN-TO

If you can find enough bamboo you can make a type of lean-to shelter. Find a spot close to the bamboo supply to reduce the amount you have to carry. If you need a raised sleeping platform, adapt the method used for the jungle hut (see p. 174).

TOOLS AND MATERIALS

- Thick wooden poles and bamboo poles
- Knife, saw, machete, or parang
- Cordage and pegs
- Poncho and tarpaulin (if you have them)

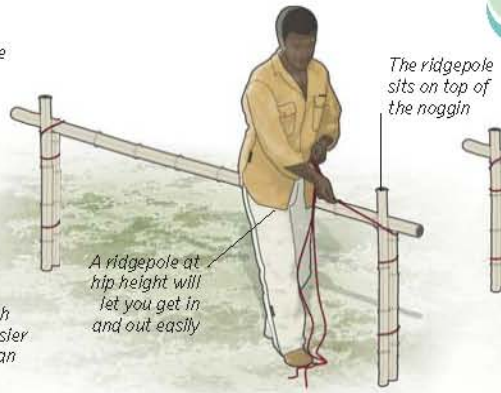

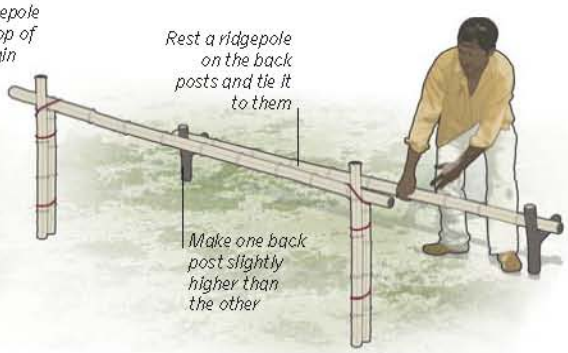

1 Cut four thick, wooden support posts.

- Hammer two posts into the ground, wide enough to let you lie lengthways between them.

2 Tie a noggin to each post for support. Make sure they are level and at the height you need for an entrance.

- Place a wooden ridgepole on top of the noggins and tie it to the posts.

3 Make two short, forked back posts to give the shelter a sloped roof.

- Align them with the front posts and far enough back so the roof is wide enough to shelter you.
- Hammer them into the ground and tie a second ridgepole to them.

4 Cut lengths of bamboo long enough to reach between the two ridgepoles.

- Split these by wedging your knife into each pole and hammering it down.
- Prepare enough split bamboo lengths to cover the roof area two times.

5 Starting from one end, lay lengths of split bamboo, hollow side up, side-by-side, between the ridgepoles.

- Work your way back along the roof, this time laying the lengths of split bamboo with the hollow side down.

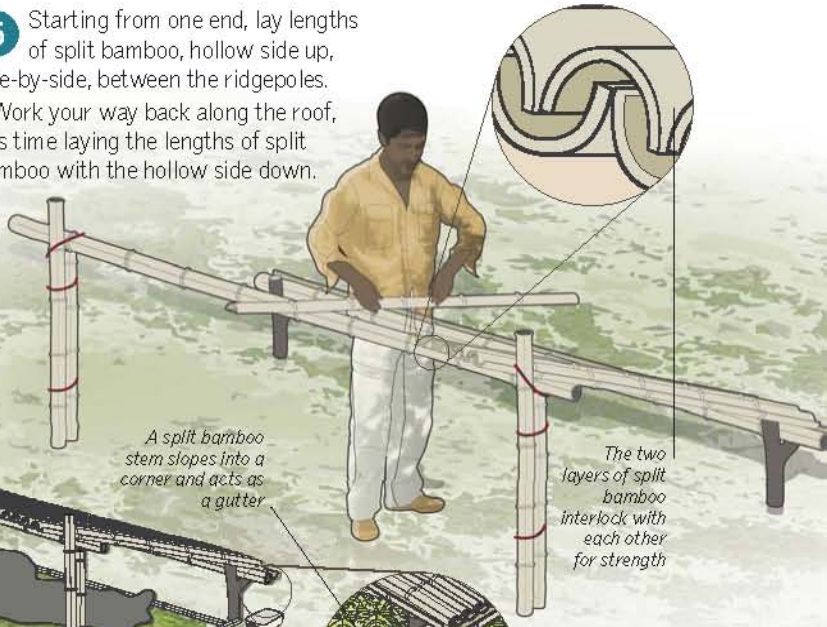

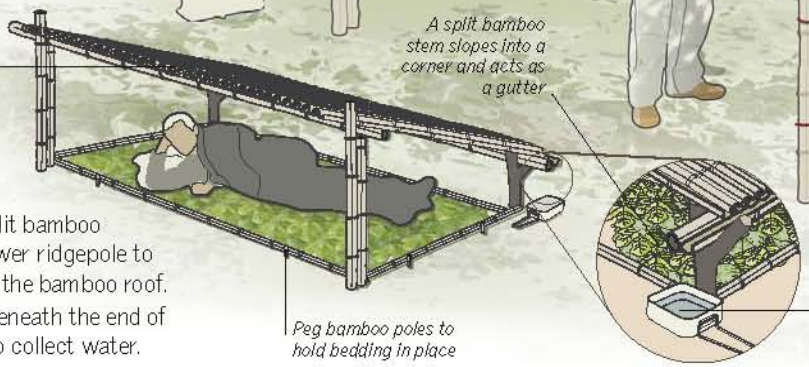

6 Lash a long split bamboo stem to the lower ridgepole to form a gutter under the bamboo roof.

- Put a pan or pot beneath the end of this sloping gutter to collect water.

JUNGLE HUT

With a little extra effort you can make a more permanent and substantial shelter than a jungle A-frame. The raised apex roof of a jungle hut allows you to sit on the platform in relative comfort. This method uses a framework of four posts, but using two trees instead of two of the poles would make it more rigid.

TOOLS AND MATERIALS
- Strong posts and poles
- Branches for platform
- Knife, saw, machete, or parang
- Digging tool
- Broad leaves for thatching the roof

1 Dig four strong posts into the ground to form a rectangle at least 1 ft (30 cm) longer and wider than yourself.

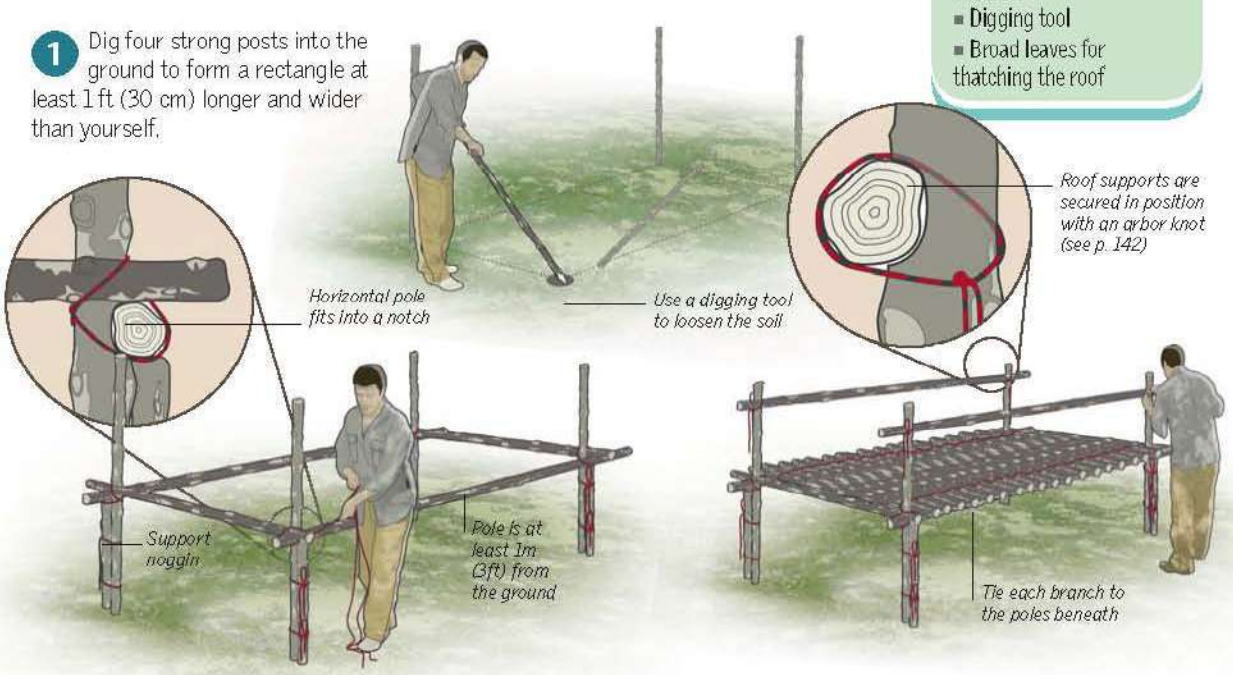

2 With your knife or saw, cut a notch into each post deep enough to fit the curve of the horizontal pole.
- Wedge the four horizontal poles into their notches and secure the junction using an arbor knot (see p. 142).
- Lash a noggin to each post to support the poles.

3 Lay a row of stripped branches side-by-side across the horizontal poles to form a raised floor.
- To make the roof supports, cut a notch on the inside of each post about 3 ft (1 m) higher than the platform.
- Tie a pole into the notches on the posts.

THATCHING LEAVES

By taking advantage of the shapes of some tropical leaves, you can make a substantial, long-lasting roof or walls. The bigger and broader leaves involve less work. If you use rattan palm leaves you can create an alternating interweave that gives a tighter thatch than the overlapping method described above.

RATTAN LEAVES

The rattan palm has leaves that are composed of rows of smaller leaflets. You can split the leaves in two and hang the halves in layers on the framework of your roof. Alternatively, you can use them whole, by folding the leaflets from one side and individually weaving them with leaflets from the other side.

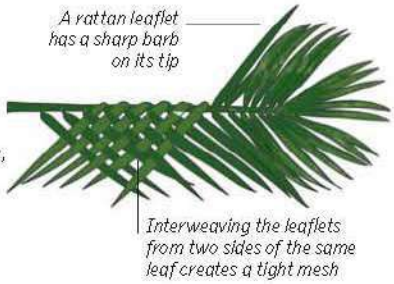

LONG, BROAD LEAVES

Many tropical plants have long, broad leaves that are good for thatching. These include some banyan figs, rubber plants, and types of banana. Arrange these leaves side-by-side on the poles of your roof frame, tuck them over the pole, and sew them into position with strips of vine. The tips of the leaves in one row point down over the leaves of the row underneath.

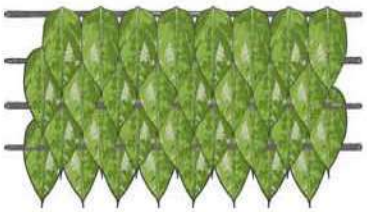

ALTERNATIVE SHELTERS

Wickiups are found in various forms around the world and are adapted as local needs and materials dictate. For example, Native Americans made them as temporary shelters in the plains, where they were also known as teepees. The pygmies of the rain forest used saplings to form a hemisphere that, once covered with natural material, created a warm, dry shelter.

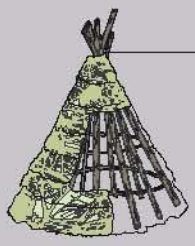

WICKIUP

The wickiup is a cluster of straight poles that are lashed together at the top, with an interwoven framework that's covered with animal hides or grass.

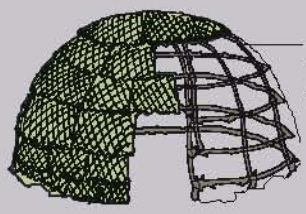

PYGMY HUT

The domed pygmy hut is made from a circle of bent saplings or limber poles secured firmly in the ground, then lashed and thatched with natural materials.

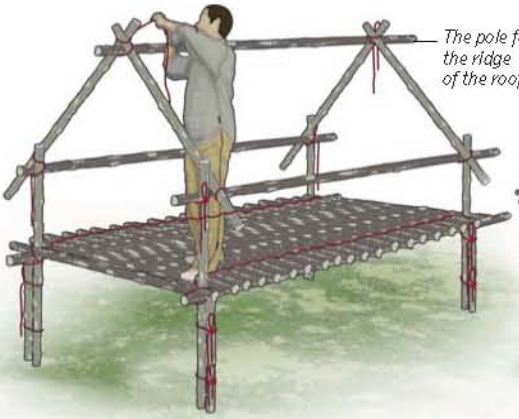

4 Tie poles together to form two A-frames. Tie these on top of the roof support poles at each end of the framework.

- Place a ridgepole across the apex of the two A-frames and tie it in place.

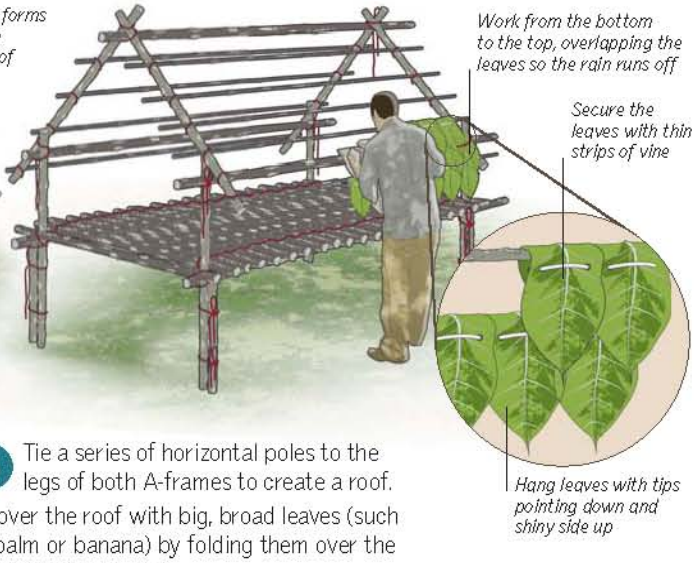

5 Tie a series of horizontal poles to the legs of both A-frames to create a roof.

- Cover the roof with big, broad leaves (such as palm or banana) by folding them over the horizontal roof poles.

THREE-LOBED LEAVES

Leaves with three lobes, such as the familiar fig leaf, can simply be hooked over the poles of your roof with the stem pointing upward. On each row of leaves you need to hang the left and right lobes either behind the pole or in front of it in an alternating pattern. The middle lobe then hangs down over two leaves on the row underneath.

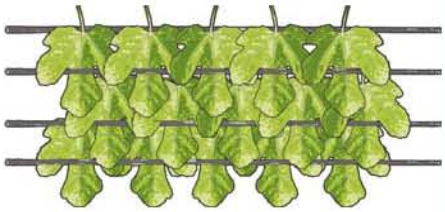

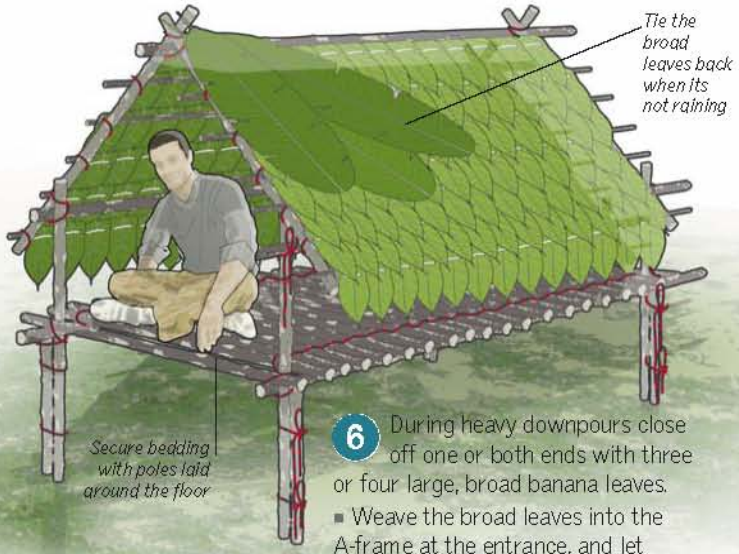

6 During heavy downpours close off one or both ends with three or four large, broad banana leaves.

- Weave the broad leaves into the A-frame at the entrance, and let them hang down.

TRUE-LIFE ACCOUNT

EXTREME SURVIVAL—IN THE JUNGLE

USEFUL EQUIPMENT

- Machete/parang
- Tarpaulin/shelter sheet
- Hammock and mosquito net
- Insect repellant
- Signal flares and whistle
- Flashlight and batteries
- First-aid kit
- Map, compass, GPS
- Survival tin, bushcraft knife
- Cell/satellite phone
- Poncho/bivy sack

FRENCHMEN GUILHEM NAYRAL AND LOÏC PILLOIS lost their way on a 78 mile (125 km) trek through virgin rain forest in French Guiana. They survived for 51 days in the jungle by drinking rainwater and eating palm seeds, snakes, and insects—although Nayral narrowly escaped death after eating a poisonous spider.

The pair set off from the Grand Kanori rapids on the Approuague River on Wednesday, February 14, 2007, bound for Saül, a former mining village at the center of the country. They planned to spend 12 days hiking, and carried sufficient food supplies, a compass and map, machete, a tarpaulin, and two hammocks. They soon found that the going was tough—some days it would take several hours of hacking through vines to hike just one mile. On the morning of February 26, when Nayral and Pillois had been due to reach Saül, the pair found themselves far from civilization and in unexpected terrain.

Knowing that a search operation would be launched, they set up camp and waited to be rescued. They built a shelter and divided the tasks—Nayral in charge of food and Pillois the fire, which was kept alight constantly to attract rescuers. They had abundant water from rainfall, but were reduced to scavenging for palm seeds, beetles, snakes, frogs, and spiders.

"THE PAIR FOUND THEMSELVES FAR FROM CIVILIZATION AND IN UNEXPECTED TERRAIN"

Occasionally, helicopters passed overhead, but failed to spot them through the dense canopy of foliage. After waiting for three weeks, the pair decided to abandon camp and trek west toward Saül. After another week of trekking for just three hours a day, unknowingly just 3 miles (5 km) from Saül, Nayral was incapacitated after eating a still-poisonous half-cooked spider. With no choice but to leave his companion behind, Pillois pushed on and reached Saül, returning with a helicopter to rescue Nayral—dehydrated, with intestinal poisoning, and plagued by parasites—on April 5, 51 days after beginning his trek.

WHAT TO DO

ARE YOU IN DANGER?

NO YES

If you are in a group, try to help any others who are in danger

Get yourself out of it:
Sun/Heat/Humidity—Slow down to the pace of the jungle. Find or improvise immediate shelter
Animals—Only six percent of snakes are poisonous but everything in the jungle will try to protect itself
Injury—Stabilize condition and apply first aid

ASSESS YOUR SITUATION
See pages 234–35

DOES ANYONE KNOW YOU WILL BE MISSING OR WHERE YOU ARE?

NO YES

If no one knows you are missing or where you are, you will need to notify people of your plight by any means at your disposal

If you are missed, a rescue party will almost certainly be dispatched to find you

DO YOU HAVE ANY MEANS OF COMMUNICATION?

NO YES

You are faced with surviving for an indefinite period—until you are located or you find help

If you have a cell or satellite phone, let someone know your predicament. If your situation is serious enough to be worthy of emergency rescue, and you have a Personal Locator Beacon (PLB), you should consider this option

CAN YOU SURVIVE WHERE YOU ARE? *

NO YES

If you cannot survive where you are and there are no physical reasons why you should remain, you will have to move to a location that offers a better chance of survival, rescue, or both

Address the Principles of Survival: Protection, Location, Water, Food

YOU WILL HAVE TO MOVE **

YOU SHOULD STAY **

DO

- Make an informed decision on the best location to move to
- Use line of sight to navigate on your bearing, as visibility may be less than 33 ft (10 m)
- Improvise shelter when not moving, and sleep off the ground, clear of the damp floor and animals
- Be on constant lookout for dry tinder and fuel
- Follow water courses downstream. Transport in the jungle relies on rivers, so settlements are most likely to be found alongside rivers
- Step onto logs so that you can see what is on the other side, rather than stepping straight over onto an unseen snake

DON'T

- Use your hands to clear undergrowth—a machete or walking staff is better suited
- Drink untreated water—boil or treat all of your water before drinking
- Leave it too late in the day to stop and make camp—three hours before sunset is recommended
- Keep too quiet. Make a noise as you progress to warn animals for your approach

DON'T

- Let the oppressive nature of the jungle overwhelm you. Slow down to its pace, don't fight against it
- Let your firewood get damp—store dry tinder and split or quarter wood to get to the dry inner core
- Eat what you cannot identify as edible—this could result in you becoming so ill that you cannot function

DO

- Select a shelter site where you can sleep off the ground and where your location aids will be most effective
- Use a mosquito net if you have one; if not, put damp foliage on your fire to repel insects, or cover exposed skin with mud
- Deploy all your aids to location and prepare them for immediate use. Be constantly alert for signs of rescue
- Keep your flesh covered despite the heat—the high humidity encourages infections. Also wash at every opportunity
- Keep a fire going—it aids location and wards off insects

* If you cannot survive where you are, but you also cannot move owing to injury or other factors, you must do everything you can to attract rescue.
** If your situation changes (for instance, you are "moving" to find help, and you find a suitable location in which you can stay and survive) consult the alternative "Do" and "Don'ts."

DESERT SHELTERS

IF YOU'RE PLANNING A TRIP into the desert, take something you can use to improvise an immediate shelter from the sun—for example, a shelter sheet, tarpaulin, poncho, space blanket, or even a couple of pieces of parachute material that will pack down into almost nothing. You can either build a shelter called a "scrape" in a natural hollow or erect a quick shelter above the ground using your poncho.

TOOLS AND MATERIALS
- Cordage and pegs
- Digging tool
- Knife
- Poncho, space blanket, tarp, or parachute material
- Rocks or ballast material
- Strong posts

BUILDING IN THE DESERT

Erecting a natural shelter in the desert is a challenge because of the heat and potential lack of materials, so try to find a site that's shaded by a tree or shrub.

POINTS TO REMEMBER

The following tips will help you deal with the extremes of temperature:

- Never try to build your shelter during the hottest part of the day.
- Ration your sweat and not your water. Avoid exerting yourself. If you start to sweat then stop what you're doing and take a break in the shade (30 minutes under an unfolded map or a space blanket and have a drink). If in a group, share the work.
- Don't leave it until the last minute to find a suitable location for your shelter. Make the decision early and plan accordingly.
- Never build a shelter in a low-lying area, dry river bed, or wadi as they are potentially at risk of flash floods.
- Avoid the top of large, isolated hills or mountains because of the danger from lightning and extreme winds.
- Try to set up your shelter on a small rise, where the temperature can be as much as 10 degrees warmer at night (cold air sinks).
- Ensure the opening faces north in the Northern Hemisphere and south in the Southern Hemisphere so the sun doesn't shine directly in during the day.
- Try to dig down to create a depression, as the ground is cooler below the surface.
- Build your shelter for the worst possible conditions and not the conditions at the time. Desert weather can change very quickly and dramatically, and the ferocious winds will rip apart any shelter that is not secure.
- If your shelter sheet has a shiny side, make sure it faces up to reflect heat and act as an aid to location from above.

DESERT SCRAPE

If you have cord, you can dig a scrape and use the cord to peg out the sheet above the scrape. If you don't have cord the sheet will have to be held in place by other means, such as soil, sand, or rocks. With all layered desert shelters try to maintain tautness and separation between the layers.

CORDLESS SCRAPE

If you don't have enough cord to peg out a sheet, you can make a scrape by either digging down or building up the sides with rocks or sand mounds placed at regular intervals. Anchor your sheet with rocks.

1 If the hollow isn't deep enough to lie in, make a digging tool and scrape out the soil from the bottom to build up the sides of the hollow.

2 Spread your poncho, shelter sheet or tarp across the top of the hollow.
- Make sure the edges of the material overlap the sides of the hollow.

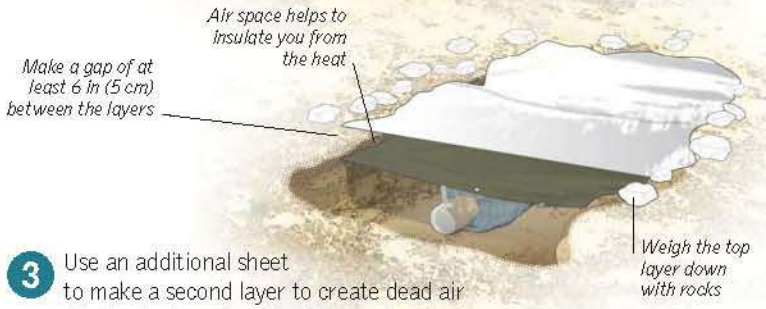

3 Use an additional sheet to make a second layer to create dead air space that helps to reduce the temperature below.
- If you only have one piece of material to use as cover, try to double it over to create the two layers.

QUICK SHELTER

If you can't find a hollow, erect a poncho shelter (see pp. 158–59) in a place that keeps you cool during the day—for example, under existing shade, such as trees or bushes, or at the top of a slight rise in order to benefit from any cool breeze.

1 Find a site beside a tree or secure a post next to where you want the shelter's opening to be.

- Rig a ridgeline to the tree or post, and peg your poncho over it to form a shelter.
- Repeat with a second poncho, shelter sheet, or space blanket to create a separate layer.

DAY-TIME CONFIGURATION

2 At night, de-rig the outer shelter and use it as a sleeping cover for warmth.

- Prevent heat escaping by securing the sides and bottom with rocks, stones, shingle, or sand.
- Sleep on some bedding if possible.

NIGHT-TIME CONFIGURATION

PARA SCRAPE

If you have cordage you can tie your sheet to four posts. If you haven't got enough cordage for four, use your backpack or a pile of stones instead.

- Prepare the depression or hollow as for a cordless scrape (see left).
- Use trees if available, or improvise posts from wood, piles of stones, or your backpack.
- Tie the sheet to the trees or the posts above the depression, leaving a gap for air to flow over.
- Create a second layer at least 6 in (15 cm) above the first.
- If foliage is available, place it between the layers to maintain separation.

SUN PROTECTION

You should always carry immediate protection against the sun. You may be able to create shade using something you've brought, such as a trekking umbrella, tent, or space blanket.

TREKKING UMBRELLA

A small, lightweight umbrella protects against sun, wind, and rain, creates cool shade, and doubles as a walking stick. Some have reflective material, or a flashlight in the handle.

SPACE BLANKET

This should be part of your basic survival equipment and will afford immediate relief from the sun. Remember that it's almost impossible to pack a space blanket to its original size once it's unpacked.

SHORELINE SHELTERS

IF YOU FIND YOURSELF in a survival situation, and can make your way to a shoreline or beach, it may offer you particular benefits. Rescue is more likely here than further inland as your aids to location work better in open areas, people are more likely to live on the coasts, and useful flotsam washes up twice daily on even the remotest of beaches.

TOOLS AND MATERIALS
- Digging tool
- Driftwood and/or rocks
- Broad-leaved plants or grass, or poncho or tarp
- Knife
- Cordage and pegs

DRIFTWOOD SHELTER

If there's enough driftwood around on the shoreline, you can use it to make a variety of shelters covered in this chapter. A hole-in-the-ground driftwood shelter offers simple protection, but may need continued attention if the sand is very fine or gets wet.

1 Make a digging tool and dig a trench in the sand.
- The trench must be above the high-tide line, long enough for you and your equipment, wide enough to be comfortable, and deep enough so you can roll over without disturbing the roof.

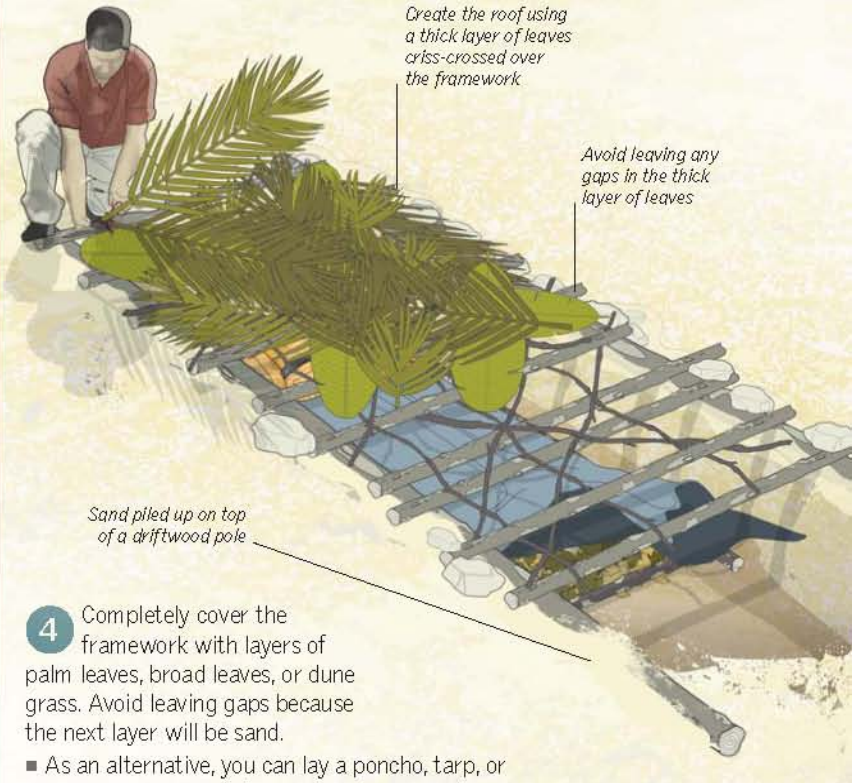

4 Completely cover the framework with layers of palm leaves, broad leaves, or dune grass. Avoid leaving gaps because the next layer will be sand.
- As an alternative, you can lay a poncho, tarp, or shelter sheet over the framework. Weigh it down with sand or stones, or peg it out if you have cordage.

BESIDE THE WATER

Shorelines vary with the environment—from tropical beaches, where sleeping out under the stars is not a problem, to rugged coasts in higher latitudes, where spending a night unprotected from the elements would be suicidal.

POINTS TO REMEMBER

If you intend to spend a night on a shoreline, exposed to the elements, the following tips may help you:
- Build the shelter above the highest high-tide mark on a seashore (see opposite) and above the highest watermark on a river or lake beach. If in doubt, move slightly further inland, where the protection may be better and materials more easily available.
- Try to establish what the weather's going to do by looking at the sky, watching the water, or detecting changes in the wind (see pp. 82–83).
- Plan for the worst, or at least have a back-up location should things not work out.
- Check the area for insects, such as midges, mosquitoes, gnats, and horseflies.
- Look for signs of wildlife, such as crabs and even turtles as sources of food.
- Finish your shelter and collect water and firewood before it gets dark.
- Beware the effects of sun and windburn, even on overcast days.
- Utilize driftwood and other building materials. Remember, any wood that's been in the sea is likely to be very heavy.

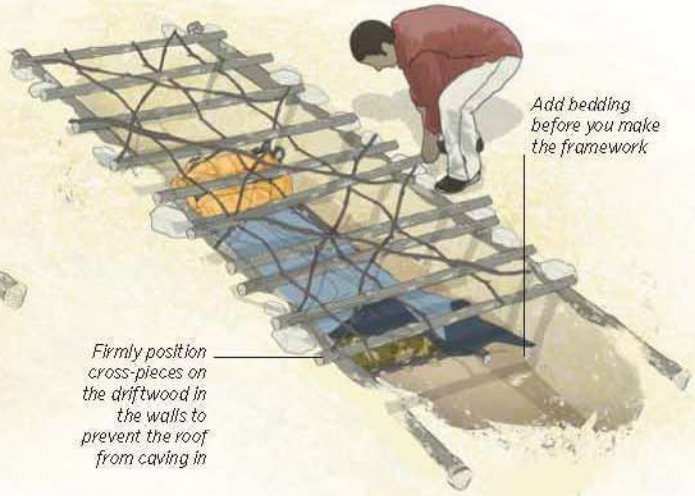

2 Find two pieces of driftwood that are as long as your trench. Several shorter pieces in a line will work equally well if you can't find long pieces.

- Place the driftwood beside the trench and pile on sand and rocks to build up the height of the walls.

3 Place pieces of driftwood across the trench to form the framework of a basic roof.

- Weave smaller pieces of driftwood into the framework, making sure they sit firmly on the driftwood in the walls.

5 Once you have covered the framework sufficiently, finish the shelter with a thick layer of sand or soil.

- The more cover you put on the roof, the more protected from the elements you will be. However, take care not to put so much weight on it that it collapses.

USING A LIFERAFT AS SHELTER

Remember, if you have abandoned ship in a liferaft, it can be used as a ready-made immediate shelter—even on land.

ON THE BEACH

The best place for a shelter is the landward side of the backshore: you can watch the sea, your daily fire can be seen from the sea, and you can see your shelter as you scavenge along the beach. In bad weather, however, the landward side of the dune crest is the best place.

PARTS OF A BEACH

A beach is shaped by the tides and the berm (a natural ridge) consists of deposits of the materials that make up the changing shoreline (sand, shingle, shale, and so on).

- At the top of the berm is a crest. A slope, or face, leads down from the crest to the water. At the very bottom of the face there may be a trough.
- The storm beach extends inland and this is where the wind and storms blow small particles of sand.
- Dunes form where the wind creates larger deposits of sand behind the beach.
- Longshore bars (sandbanks) may lie further out to sea and are formed where the waves first start to break.

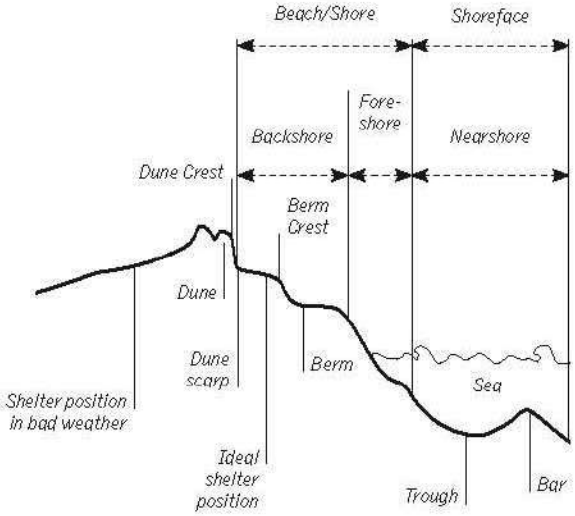

SNOW SHELTERS

THE TYPE OF SNOW SHELTER you can build depends on the kind of snow, the equipment you have, and the opportunities offered by the environment—for example, wooded areas are usually better than open areas because they provide protection and natural materials. You could build a snow trench, snow cave, tree pit, snow ledge, or a quinzhee.

TOOLS AND MATERIALS

- Boughs and branches
- Bedding
- Long pole
- Spade or pan for shovelling snow
- Long knife or saw
- Knife (or other long blade)
- Tarpaulin or shelter sheet

COMPACTED SNOW TRENCH

If the snow is compact enough, and if you have a long knife or snow saw, you can cut out blocks of snow to form a trench and then use the blocks to form an apex roof. This design requires a lot of effort and a bit of practice, but gives you a solid shelter with some additional height.

1 Use the blade of your knife to cut a rectangular trench that is the width, length, and depth you require.

- Cut blocks from the rectangle and place them at the side of the trench.

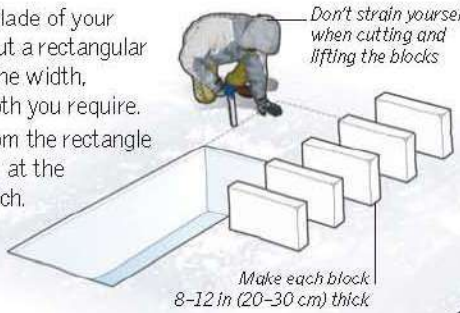

2 Trim the ends and rest two blocks against each other at the required angle for your roof.

- Use the other blocks to complete the roof and cut a triangle for the foot end.

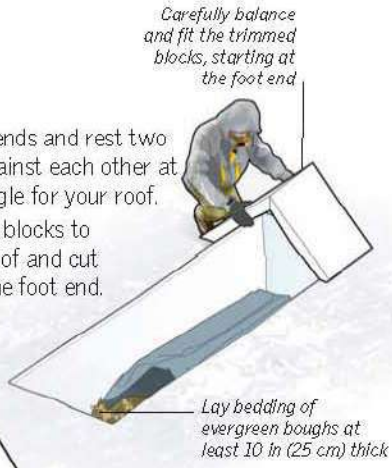

3 Pack snow into any gaps between the blocks.

- Pile extra snow on top of the blocks to add insulation.
- Create a doorway from snow blocks or your backpack.
- Don't block the doorway completely, but allow for adequate ventilation.

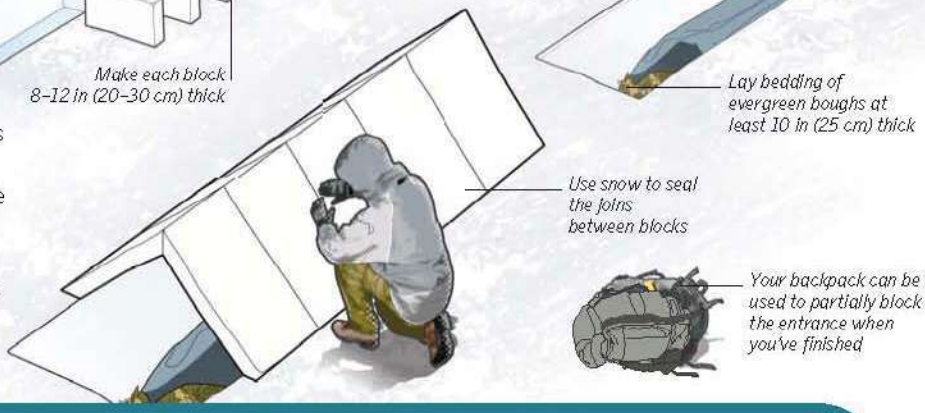

BUILDING SNOW SHELTERS

Study the snow around you to determine whether it's compacted or uncompacted before you start.

POINTS TO REMEMBER

The following tips may help if you intend to build and stay in a snow shelter:

- Make a shelter that's just big enough for you and your equipment. Don't spend hours building a shelter for only one night: do the least amount of work for the maximum amount of protection.
- If you need to peg out guylines in the snow, tie a short stick 6–12 in (15–30 cm) long to a line or cordage and bury it in the snow. Compact the snow down on top and, when the snow sinters (hardens), your peg will be held secure.
- Snow is an excellent insulator. Fresh, uncompacted snow is typically 90–95 percent trapped air. Since the air barely moves, the snow can keep you warm and dry if used correctly.
- Check your site for hazards, such as snowdrifts, freezing winds, avalanche, cornice collapse, and big animals.
- Create ventilation holes—make one near the ground to let fresh air in and one at the top to let air escape. Check the holes are clear every 1–2 hours.
- Keep tools inside the shelter in case you have to dig your way out.
- If you leave the shelter, mark the entrance. Take basic survival equipment with you: it's better to have and not need than to need and not have.
- Brush snow off your equipment and clothing before entering the shelter.
- Tie all vital equipment to yourself so that you can't accidentally drop and lose it in deep snow.

FOUND SHELTERS

If you're lucky you may find a tree that has space under it so you don't have to dig a trench. For example, under a low bough or in a tree pit—an area where a heavy and deep snowfall has built up around a tree, leaving little or no snow under the lower branches. Use your walking stick to check the depth of the snow, and also to check for pockets of air.

LOW BOUGH

Find an evergreen tree with a large branch lying on top of the snow. Dig down from the leeward side, excavate a pit, insulate the floor with boughs, and make sure you have adequate ventilation.

TREE PIT

Test the snow around a tree that has large, low, snowbound branches. Little or no resistance indicates a pit. Dig from the leeward side, place boughs on the floor, and ventilate the pit.

FIGHTER TRENCH

If the snow is soft, a fighter trench is quick and easy to build. In an emergency, you can even make a trench by kicking out the soft snow. First, find a location that protects you from the elements as much as possible, then test the depth of the snow with your walking stick or a pole.

WARMING AND COOLING TIPS

Whether you're building a shelter or resting inside it, the following tips may help you stay at the right temperature:

- Remove layers of clothing as you work, perhaps to your base layers with a waterproof layer on top. Keep clothing dry so that when you stop working you can put it on again.
- Keep shelters at a constant temperature. If snow melts and then re-freezes, it stops insulating.
- Avoid heating your shelter too much. Even a candle flame can raise the temperature inside by four degrees.

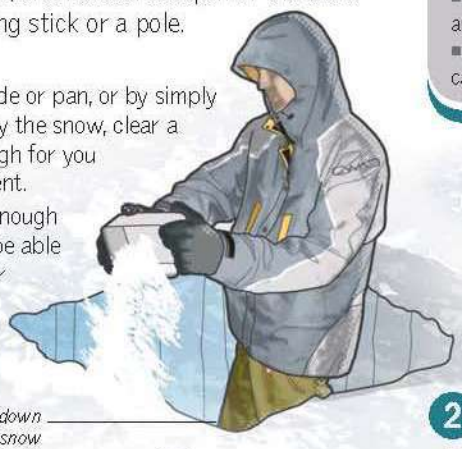

1 Using a spade or pan, or by simply kicking away the snow, clear a trench large enough for you and your equipment.

- Allow at least enough depth for you to be able to roll over in your sleep without disturbing the roof.

2 Create a framework of branches and boughs across the trench.

- Make sure you have enough roofing material on top to stop the snow from penetrating through.

3 If you have a shelter sheet or tarpaulin, place it on top of the framework to form an extra insulating layer.

- Alternatively, you can use it to cover your bedding.

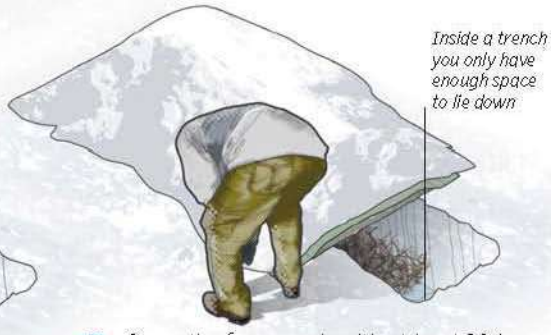

4 Cover the framework with at least 12 in (30 cm) of snow to act as insulation.

- Dig a small pit at the entrance to allow easy access.
- At the entrance, you can make a small, controlled fire on a green log base.

QUINZHEE

A quinzhee is a dome-shaped shelter made by hollowing out a pile of settled snow. It's an overnight shelter that is easier to construct than the more permanent igloo ("igloo" means "home"), which is made from cut blocks of snow and requires skill and knowledge. You can't stand up in a quinzhee, but you can sit upright or just curl up.

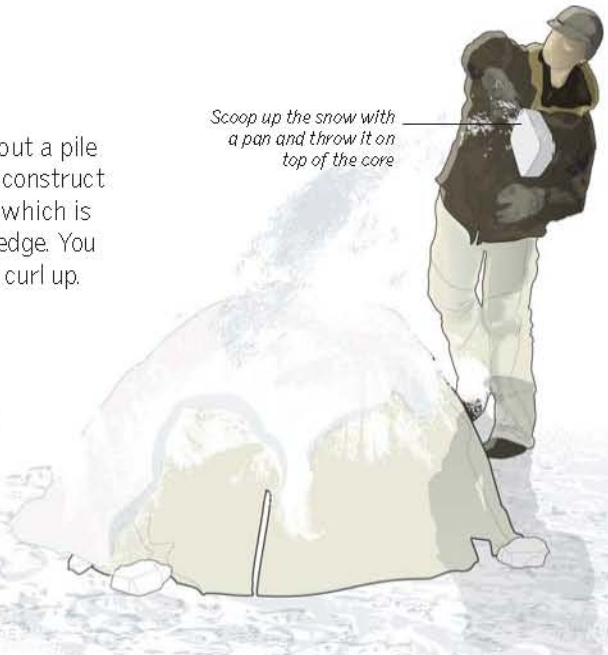

1 Find a relatively flat area that's covered with an abundance of snow.

- Mark a circle for your quinzhee, including walls that will be about 10 in (25 cm) thick, and stamp down the snow.
- Use your backpack and boughs or leaves covered with a tarpaulin to form the core of the shelter. Site the doorway at 90 degrees to the prevailing wind.

2 Use a container, pan, snowshoe, or other suitable item to gather as much soft snow as you can.

- Pile snow on the shelter's core to form a dome of the required height. Build up layers of snow until you have a covering of at least 10 in (25 cm), smoothing out each layer as you go.

3 Smooth out the snow on the dome and leave it to sinter (harden) for 1–3 hours, depending on the type of snow and the ambient temperature.

- Keep active during this waiting period, particularly if it's windy: collect wood, build a fire (away from the quinzhee), and prepare yourself for the night.

4 To get an even thickness in the roof and walls, push guide-sticks of equal length through the snow toward the center of the dome.

- Build a small compact mound in front of the dome.

SNOW CAVE

A snow cave provides good protection but it takes a lot of work. Suitable locations might be compact snow on the lee side of a hill or an established snowdrift with a firm crust. Avoid snow that's newly fallen, powdery, loose, or shallow and uneven. The cave should be high enough to sit up in and deep enough for you and your equipment. Locate the entrance 45 degrees downwind to keep it free of drift.

BUILDING THE CAVE

- Dig a tunnel 3 ft (1 m) long into the snow bank, then excavate a cave on two levels. The lowest is the tunnel, while the other is a ledge at least 2 ft (70 cm) higher, and wide enough to sleep on.
- Create an arched ceiling to give the cave more support. Make it at least 18 in (45 cm) thick so it can take the weight of heavy snow.
- Mark the area with boughs to warn people where the shelter is.
- Insulate the sleeping ledge with plenty of natural material.
- Make at least one ventilation hole in the roof, but not facing into the wind.

SNOW LEDGE

You can build a snow ledge from a well-packed drift of snow out of the wind, or at least perpendicular to it.

MAKING THE SNOW LEDGE

- Dig out blocks to form a doorway about 2 ft (70 cm) into the snow. Dig out blocks to form a horizontal rectangle above.
- Dig upward to create a sleeping platform and then form an arched ceiling.
- Place the cut blocks across the horizontal rectangle and seal any gaps with snow.
- Make a ventilation hole in the roof and one lower down to let fresh air circulate.

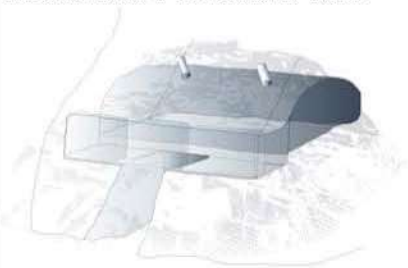

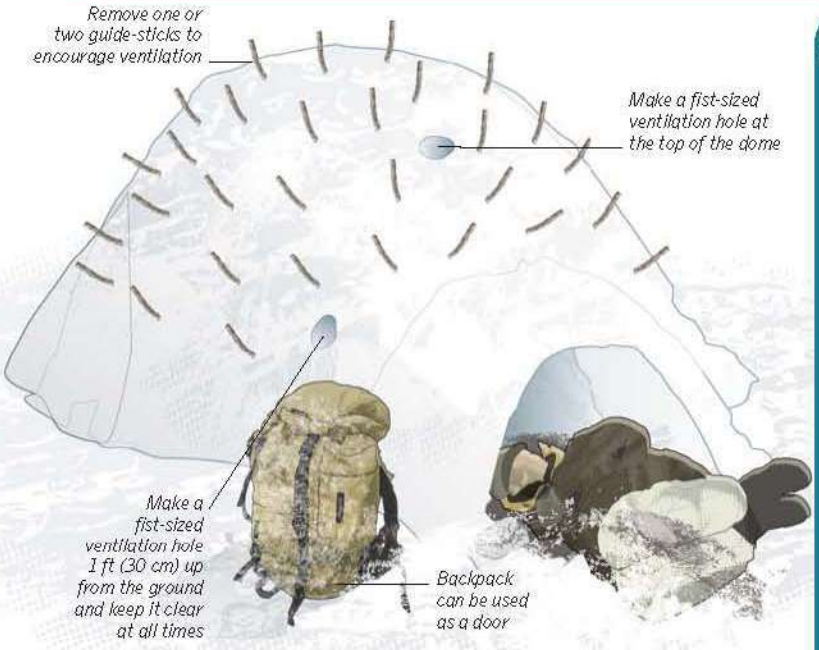

5 When the snow has hardened, burrow into the mound, remove your backpack and tarpaulin, and excavate snow from the core.

- Use the guide-sticks to keep the walls at least 10 in (25 cm) thick.
- Don't use your hands to dig as they will get cold and wet.
- Smooth out the snow on the inside to prevent drips from forming.
- Build a raised sleeping platform. This creates a cold sink (an area into which the coldest air falls), taking cold air away from your sleeping area.

CARBON MONOXIDE

Carbon monoxide is an odorless gas that's produced when there's not enough oxygen to create carbon dioxide from burning fuel. Carbon monoxide poisoning can be fatal in a well-insulated, non-ventilated environment.

PREVENTION IS BETTER THAN CURE

Create one or more holes, 3–4 in (7.5–10 cm) in diameter, at the base and top of your shelter. Make sure the air flows out unobstructed.

SIGNS OF POISONING

Carbon monoxide poisoning is cumulative and can build up over a few days. Mild effects are fatigue, faintness, and flu-like symptoms. As it progresses, the effects are severe headaches, nausea, and decreased mental coordination.

TAKE ACTION

Get into the fresh air at once. You need to breathe fresh air for at least four hours to reduce the amount of carbon monoxide in your system by half.

DETECTORS

Carbon monoxide detectors are widely available, but tend to be made for the home or RV. Some models are battery-operated, but they're too bulky for camping. You can wear a patch that changes color when carbon monoxide is in the air, giving an early visual warning.

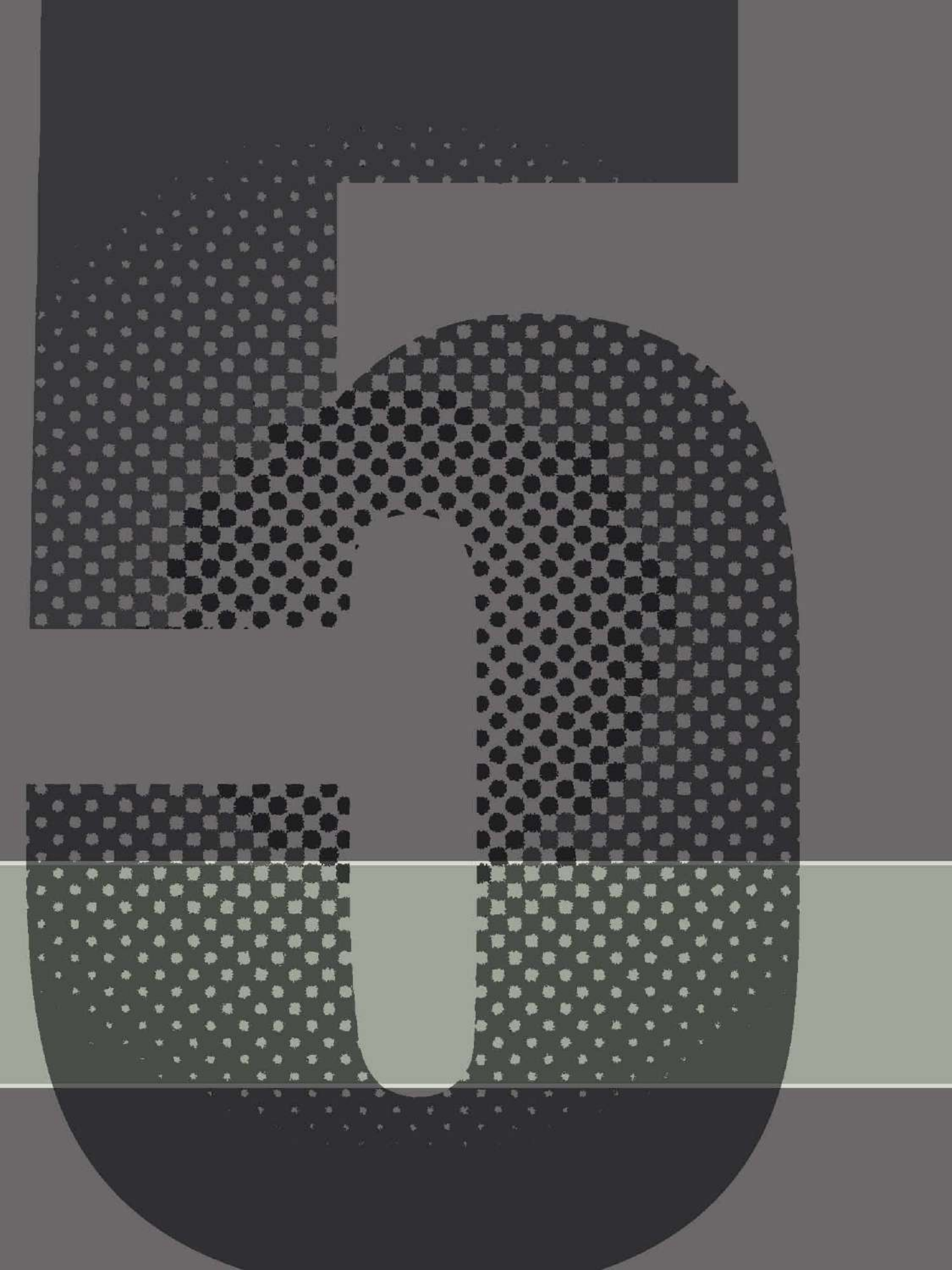

WATER AND FOOD

FIND AND TREAT WATER

The importance of water, even in a short-term survival situation, should never be underestimated. Water is essential to life, and a regular intake of 2–3 liters (4⅕–6⅓ pints) a day is needed just to maintain your water balance and prevent dehydration. The amount required can increase dramatically depending on factors such as the temperature of the environment, your age and physical condition, your workload, and whether you have been injured. It's not unusual for UK Royal Marines operating in desert or jungle environments to require 14 liters (3¾ gallons) per day. We tend to take water for granted, not appreciating just how important it is until we don't have any—at which point it becomes the most important thing in the world.

You should always plan your treks around your need for water and your ability to replace it as required. There are many hydration systems available and many small and efficient methods of filtering and purifying water on the trail. In a survival situation, you should always strive to filter and purify any

In this section YOU WILL DISCOVER...

- that **birds** can **show you** the **way** to **water**...
- when to **suck** on a **small pebble**...
- how to **make** a **gypsy well** and a **solar still**...
- why a **Finnish marshmallow** could **save your life**...
- how to **absorb water** without **drinking it**...
- how to **improvise** a **basic bladder**...
- the **importance** of **surgical tubing**...

water before drinking it. Bringing water to the boil will kill all water-borne diseases. In the short-term, stomach bugs from contaminated water may not kill you but they can seriously affect your ability to carry out other survival tasks. However, if you have no choice, it's better to drink contaminated water than not to drink at all. That way, a doctor will at least be able to treat you, whereas dehydration will kill you—and death can't be treated! Never drink urine or salt water, as these will only dehydrate you more.

If a source of water is undrinkable, such as salt or stagnant water, or even urine, you can still produce drinking water if you have the means to start and maintain a fire.

THERE ARE MANY WAYS to purify water, but if you have a fire you will always be able to distill it—and therefore make it drinkable. Build your fire close to the water source (if the water can be poured into a container or a hole dug in the ground where it will not seep away quickly, this will make the process more efficient). When the fire is established, place rocks in it (don't use slate or other layered rocks as they might shatter). Once the rocks are heated, use a forked stick or similar to transfer them into the salt or foul water. Then, suspend a piece of absorbent material, such as a T-shirt or moss, over the hot rocks to collect the steam. The steam will condense on the material and the resulting water, which you can wring out, will be fit to drink. If you have a cooking pot or survival tin, you can simply fill it with water, boil it over a fire, and collect the steam in the same way.

If you have absolutely no means of treating or boiling water, you should try to:

- Find the clearest flowing water and collect it from the surface.
- Filter debris out, even if only through a sock.
- Remember, it's better to drink foul water than not to drink at all.

> We tend to take **water for granted** until we don't have any—at which point it becomes the most **important thing in the world**

THE IMPORTANCE OF WATER

YOU REQUIRE A STEADY supply of water to sustain yourself in a survival situation and without it you will dehydrate. Left unchecked, dehydration will end in death. To survive, a balance between water intake and loss must exist.

WHY YOU NEED WATER

Water is essential to life. It is needed, directly or indirectly, for every physical and chemical process that takes place in your body. Here are just a few of the functions that water performs:

- **Delivery service:** water carries oxygen, nutrients, and other essentials around the body.
- **Waste remover:** the kidneys use water to flush out toxins via urine.
- **Coolant:** water regulates the body's temperature.
- **Breathing aid:** the lungs use water to moisten inhaled air so that it doesn't irritate the sensitive pulmonary linings.
- **Sensory aid:** water helps conduct nervous impulses around the body.
- **Shock absorber:** water protects the vital organs and provides lubrication around the joints.

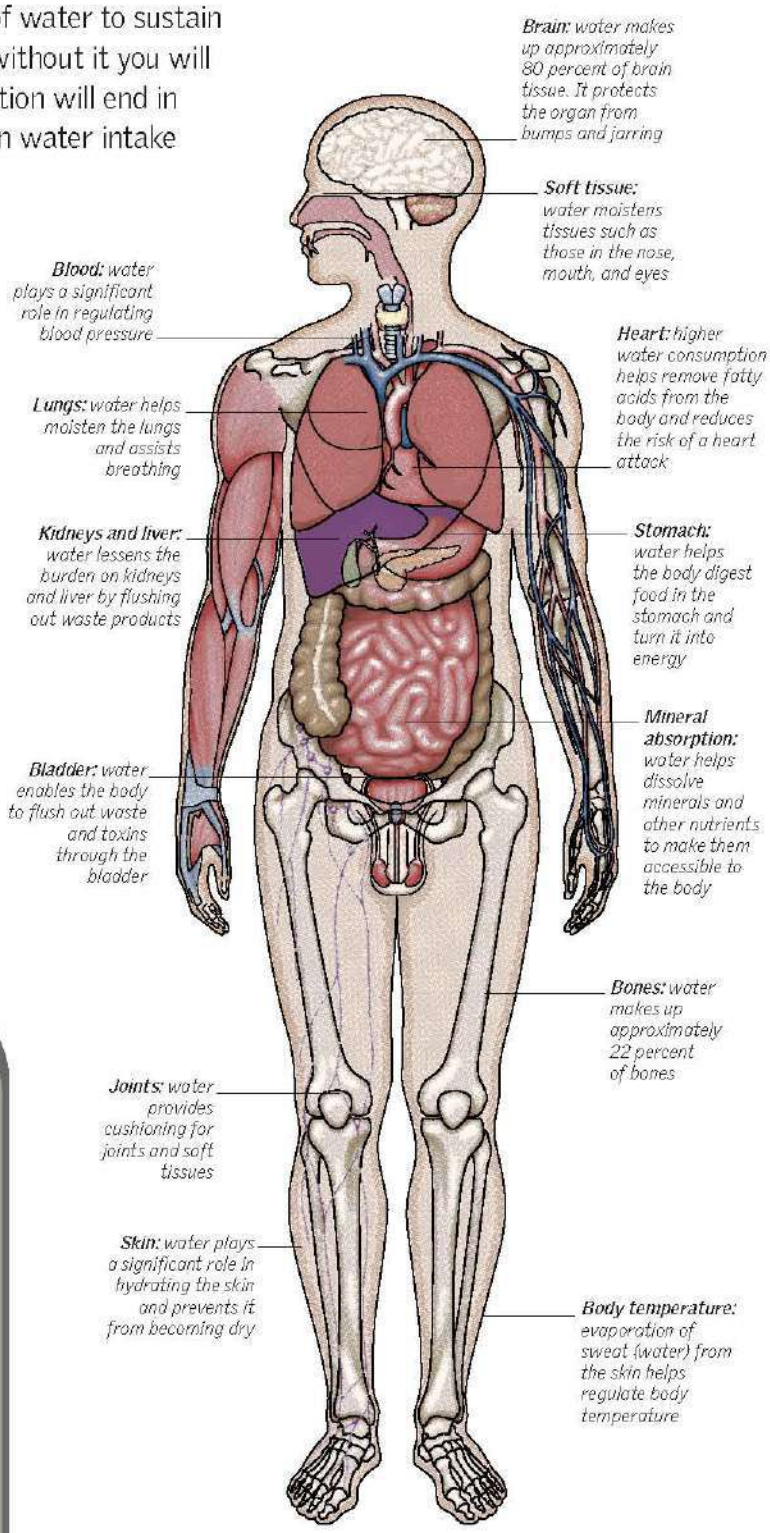

HOW MUCH WATER DO YOU NEED?

How much water you need to survive a particular situation is dictated by a number of factors, such as your physical state, the environment you are in, and your exertion levels. Even when resting in the shade, the average person will lose more than 1 liter (1¾ pints) of water each day just through breathing and urination, a figure that increases dramatically once loss of water through sweat is taken into account. A minimum of 3 liters (5¼ pints) per day is required to remain healthy in a survival situation, with this amount increasing for higher temperatures and heavier workloads.

WHAT IS DEHYDRATION?

Dehydration occurs when you fail to replace the water your body loses. It's vital to recognize the symptoms early. Factors that can lead to dehydration include high and low temperatures, humidity, work-rate, clothing, body size, fitness levels, and injury.

THE EFFECTS OF WATER LOSS		
1–5% LOST	**6–10% LOST**	**11–12% LOST**
▪ Thirst ▪ Discomfort ▪ Urine becomes darker ▪ Loss of appetite ▪ Impatience ▪ Drowsiness ▪ Lethargy ▪ Nausea ▪ Headache	▪ Dizziness ▪ Dry mouth ▪ Blueness of extremities ▪ Slurred speech ▪ Swollen tongue ▪ Blurred vision ▪ Tingling in limbs ▪ Inability to walk ▪ Difficulty in breathing	▪ Stiffness of joints ▪ Deafness ▪ Defective vision ▪ Shriveled skin ▪ Lack of feeling in skin ▪ Inability to swallow ▪ Delirium ▪ Unconsciousness ▪ Death

WATER-BORNE DISEASES

Water-borne diseases are caused by ingesting water contaminated by the faeces or urine of humans or animals that contains protozoa, viruses, bacteria, or intestinal parasites. Globally, they cause 10 million deaths a year.

	DISEASE	SYMPTOMS
PROTOZOA	Cryptosporidium	Loss of appetite, nausea, and abdominal pain, usually followed by profuse, foul-smelling, watery diarrhea, and vomiting.
	Giardiasis	Loss of appetite, lethargy, fever, vomiting, diarrhea, blood in the urine, and abdominal cramps.
VIRUS	Infectious hepatitis (Hepatitis A)	Nausea, loss of appetite, mild fever, aching muscles, dark-colored urine, jaundice, and abdominal pain.
BACTERIUM	Amoebic dysentery	Feeling of fatigue and listlessness. Feces may be solid, but will smell foul and contain blood and mucus.
	Bacillary dysentery (Shigellosis)	Fever, abdominal pain, muscle cramps, high temperature, and blood, pus, and mucus in stools.
	Cholera	Vomiting, poor circulation, cold and clammy skin, muscle cramps, rapid dehydration, and increased heart-rate.
	E. coliform	Diarrhea and vomiting. Can cause death in vulnerable groups such as the very young or the elderly.
	Leptospirosis	Jaundiced appearance, lethargy, high temperature, aching muscles, and vomiting. Can be fatal if not diagnosed early.
	Salmonella	Nausea, diarrhea, headaches, stomach cramps, fever, possible blood in the feces, and vomiting.
PARASITES	Bilharzia	Irritation to the urinary tract and blood in urine, rash or itchy skin, abdominal pain, cough, diarrhea, fever, and fatigue.
	Hookworms	Anaemia and lethargy. Larvae travel to lungs and are coughed up and swallowed into the stomach, where they grow into worms.

TOO MUCH WATER?

Hyponatraemia is a condition caused when excess water accumulates in the body at a higher rate than it can be excreted. It results in a diminished sodium concentration in the body's plasma and the swelling of the body's cells. It can lead to a swollen brain and other neurological problems and, in extreme cases, coma and death. The way to prevent hyponatraemia is to control the amount of water that you drink, and to regulate your body's salt intake. If you don't have salt or sodium tablets in your survival tin, you can filter saltwater through fabric to sift out the salt content.

RATIONING YOUR WATER

If your water supplies are limited, you'll have to use what rations you have efficiently until you're rescued. If your water rations will not last that long, you'll have to procure water yourself. There is much debate about the advantages and disadvantages of drinking no water for the first 24 hours of a survival situation, but at this early stage it's best to make sure you're adequately hydrated. Your particular circumstances will dictate what's best to do, but always consider the following:

- The incident that put you in the survival situation may have been both dramatic and stressful; this will make you thirsty.
- In the first 24 hours of a survival situation, you'll be addressing the principles of survival: protection (shelter) and location (see p. 27). This is hard, thirsty work.
- Physical factors—such as seasickness, injury, or the environment you're in, such as a desert—may dictate that water-rationing is not a viable option.
- If you only have a limited water supply, but drink nothing for the first 24 hours, you may end up being so dehydrated that what little water you have in your possession will have no positive effect on your dehydrated state.

FINDING WATER: TEMPERATE CLIMATES

THE ABILITY TO LOCATE drinking water can challenge even the most experienced survivor and the challenges can vary dramatically according to the environment and the local conditions. It is, therefore, extremely important that you're aware of all the potential sources of water for the environment you're in.

WARNING!
In a survival situation, even if you think your chances of rescue are high and that help is imminent, you should start to look for water sources as soon as you have addressed the immediate problems of protection and shelter (see pp. 154–81). Remember that you can survive for up to three weeks without food; without water, you won't be able to survive for more than a few days.

COLLECTING WATER

You can find water in a range of sources, which vary in terms of quality and accessibility. Your priority is to locate the best and most accessible source of water in your immediate environment.

CATCHING RAINWATER

Collecting rain as it falls is the safest way to procure drinking water. It will need no treatment prior to drinking, as long as the catchment device itself has not been contaminated. Any number of non-porous materials—such as tarpaulins, ponchos, flysheets, survival blankets, or even large leaves—can be used as a rain-catchment device. Bear in mind that the larger the surface area of the material, the more rainwater you'll be able to catch.

MATERIALS AND TOOLS
- Tarpaulin
- Four sticks
- String
- Heavy stone
- Container

1. Select a place as close to your camp as possible where your catchment device will be exposed to the most possible rain.

2. Firmly secure the tarpaulin to four stakes (sticks of equal length could be used) using string. Make sure that one end is higher than the other to provide a natural run-off for the water.

3. Place a heavy stone in the middle of the tarpaulin, roughly two-thirds of the way toward the lower end, to create a channel into which the rain will run down from the tarpaulin's sides.

4. Place a container, such as a pan or can, beneath the end of the channel to collect the water as it starts to flow off the tarpaulin.

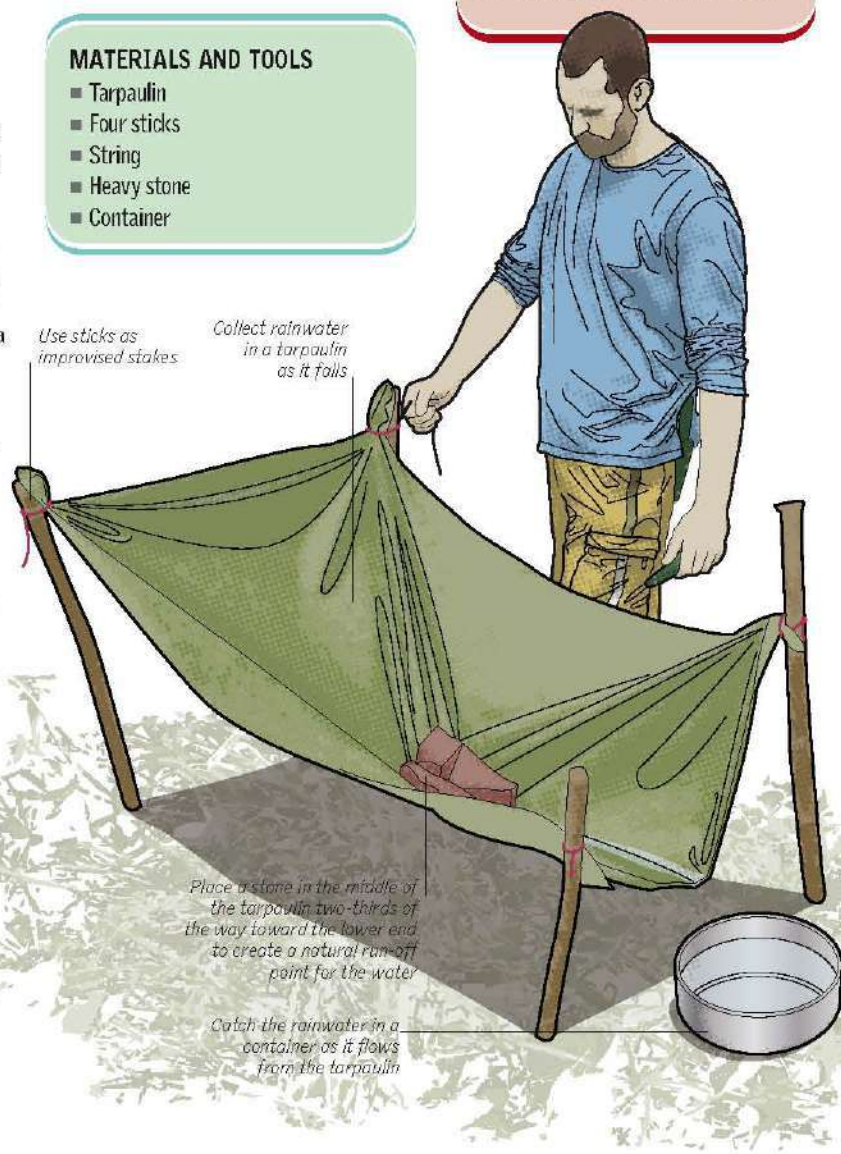

LOCATING OTHER NATURAL WATER SOURCES

If rainwater is not available, there may be other natural sources of water, from easily visible streams and rivers, to bores and holes, which can be hidden. Whatever your source, you should always treat the water before drinking.

SOURCE	CHARACTERISTICS
SPRINGS	These occur when the water is forced to the surface as a result of subterranean pressures or from gravitational flow from higher sources. Found in low-lying areas, springs normally provide a permanent water source. Contrasting green vegetation is an indicator of their presence.
STREAMS AND RIVERS	Rivers and streams are an invaluable source of water, but there are still a few things you need to bear in mind when collecting water from them. The closer to the mountaintop the river or stream is, the clearer the fast-running water will be. The further downstream the water travels, the more likely it is to pick up minerals, debris, and pollutants that might be harmful to you. If possible, either check upstream for dead animals that may have contaminated the water, or follow the water downstream. Always try to collect the fast-flowing water near to the surface. In arid areas, rivers and streams tend only to flow during floods and will contain more pollutants and debris.
ROCK HOLES	Usually found in high ground, rock holes are natural collectors of rainwater. If the water appears to be trapped deep down, you can use your surgical tubing to retrieve it. Make sure you filter and purify the water before you drink it.
WELLS AND BORES	In some areas, you may find wells or old bore holes. Wells, which may be featured on local maps, can be deep and covered, making water procurement more difficult. In remote areas, wells are covered and marked in certain ways by the locals—find out what the markers are for your area.
LAKES AND PONDS	Rivers, streams, and water run-offs all flow into lakes or ponds. If you're collecting water from these sources, always try to collect it as it runs into the body of water, as lakes and ponds are more static than flowing water and can, therefore, become increasingly stagnant. Try to find the cleanest-looking area and avoid areas where debris has collected or where algae has formed. Note that the presence of fish indicates that the water is still oxygenated.
SEEPAGE	Usually located at the base of cliffs, high ground, or rocky outcrops, seepage is caused by slow-running channels that drain off these features.
SOAKS	Soaks are found close to rivers and creeks in low-lying areas, and are normally lower than the existing water table. The presence of soaks is often indicated by vegetation and they may be subject to pollution as a result of their use by animals. Make sure you filter and purify the water before you drink it.

AVOIDING POTENTIAL DANGERS

Most water sources are likely to be used by animals for drinking, bathing, urinating, and defecating, so always filter and purify any water collected before use (see pp. 200–01); the only exception to this rule is if you have collected rainwater. When collecting water, you should also be aware of the following dangers.

- It is possible that you'll encounter dangerous animals either using the water source, or on the way to or from the water source.
- Almost all major water sources will have a ranking system to determine which animals can use it. If all of the small gazelles suddenly disappear, ask yourself why.
- If you're collecting water from rivers, be aware of the potential dangers of river wildlife, such as crocodiles and snakes.
- If using dry riverbeds during the rainy season, be aware that flash floods can move quicker than you can run.

LOCATING HIDDEN WATER SOURCES

The presence of water is usually indicated by signs of life, such as green vegetation, animal tracks, or human habitation. Even if the terrain you're in appears lifeless, there may be plenty of indicators of a possible water source.

USE THE TERRAIN

- Observe the landscape for patches of green. Be aware that vegetation may not need obvious surface water for survival and may get its water from deep roots that tap into moisture below the surface.
- Water is subject to gravity and is more likely to be found downhill or in low-lying areas, such as valleys, dry riverbeds, narrow canyons, gullies, and at the base of cliffs or rock formations. Green vegetation forms next to a river and decreases as the ground rises away from the water source.
- Water will often seep inland on coastlines, leaving behind wetlands that contain water with tolerable levels of salt or that can be distilled in a solar still (see p. 189) to procure fresh water.

WATCH THE ANIMALS

- The Bedouin listen to the twittering of birds at dawn and dusk and follow their flight path to discover where they drink.
- Flocks of birds circling over one spot are usually flying over a water source. This does not apply to meat-eating birds, such as vultures, eagles, or hawks, who get their fluid requirements from the meat they eat.
- All finches and grain-eaters need a regular supply of water. Observe their flight patterns to locate a water source.
- Bees need water, so beehives are never far from a water source.
- Animal tracks, especially those of herd animals, will often lead to a water source. Look for converging sets of tracks.
- Flies stay close to water and the presence of mosquitoes almost certainly means that water is nearby.
- Look out for herd animals, such as buffalo, hippos, elephants, impala, and wildebeest, as they depend on water.

PROCURING WATER

Even if you find yourself in an environment without any obvious water sources, it doesn't necessarily mean that water will not be available to you. There are a number of water-procurement techniques that may mean the difference between you making it through a survival situation or not.

SOURCING WATER FROM DEW

Dew is water in the form of droplets that form on exposed surfaces in the early morning or late evening and can provide an invaluable supply of fresh water. It occurs when the temperature of a surface is low enough to allow the moisture in the warmer air above it to condense. Dew can easily be collected from any non-porous surface—such as a car roof or a tarpaulin—with a piece of cloth that can then be wrung out into a container. You can also harvest or trap dew.

HARVESTING DEW

You can harvest dew by walking through a field of long grass before sunrise or late in the evening with a piece of absorbent material—such as rags or a T-shirt—tied around your ankles.

1 Tie the material tightly around each ankle and walk through the dew-covered grass. It will absorb the dew as you move.

2 Wring the rags to extract the water. Repeat the process until you have an ample supply of water or the dew has evaporated.

DELAYING DEHYDRATION

When water is scarce, the following points will help you delay the onset of dehydration:

- Conserve what water you have and use it as efficiently as possible.
- Work only in the coolest part of the day and avoid sweating.
- If the sun is shining, seek shade and keep covered.
- Suck a small, smooth button or pebble to help stimulate saliva and remove the sensation of thirst.
- Avoid eating protein-rich food as it requires more water to digest than those in other food groups.

MAKING A DEW TRAP

Dig a hole about 18 in (45 cm) deep, line with a plastic sheet, and fill with smooth, clean stones. Water will condense on the stones overnight. Harvest as early as possible the next morning to ensure it does not evaporate.

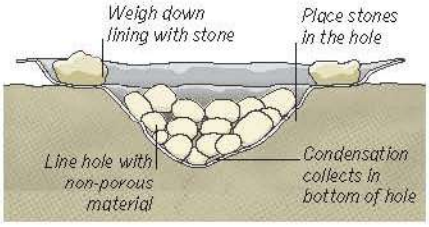

SOURCING WATER FROM PLANTS

Transpiration is the evaporation of water from a plant, primarily from its leaves. You can collect this vapor to boost your fresh water supply. All you need is a clear plastic bag.

MAKING A TRANSPIRATION BAG

Place a smooth rock in the lower corner of the plastic bag, and place the bag over the leaves of a tree branch, tying the end. As water evaporates from the leaves, it will condense on the inside of the plastic bag and collect at its lowest point.

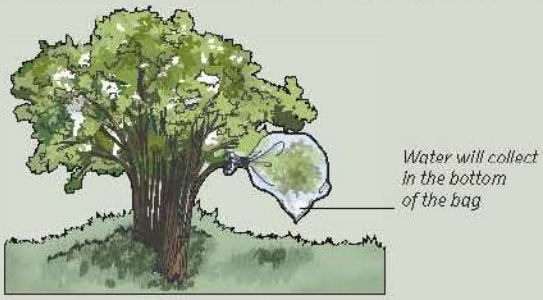

MAKING A VEGETATION BAG

Cut green vegetation and place it in a plastic bag. Place a smooth rock in the lower corner of the bag and tie off the open end. Secure the bag in direct sunlight. The sun causes water in the leaves to evaporate, condensation forms on the plastic bag and drips to the lower corner of the bag.

MAKING A SOLAR STILL

A solar still works using the same principle as a vegetation bag (see opposite). It collects potable water from the vapor that is produced by vegetation, water that is unfit to drink, or moisture from the ground.

TOOLS AND MATERIALS

- Spade or shovel
- Two containers
- Vegetation and/or undrinkable water
- Piece of fabric or clothing
- Plastic sheet, tarpaulin, or space blanket
- Stones
- Surgical tubing

1 Ideally find, or dig, a hole in the ground at least 2 ft (0.6 m) wide and 2 ft (0.6 m) deep. Place an empty container in the center of the hole. Fill the hole with vegetation, a receptacle containing undrinkable water, such as saltwater or urine, or fabric soaked in undrinkable water.

2 Cover the hole with a plastic sheet, and secure it in place with stones. Place a stone in the center of the sheet to create a run-off point for the water. The sun's heat will evaporate water from the vegetation, or distill the undrinkable water, producing pure water vapor. The water vapor, which is now free of contaminants, will condense on the underside of the plastic sheet, and drip into the container where it can be collected.

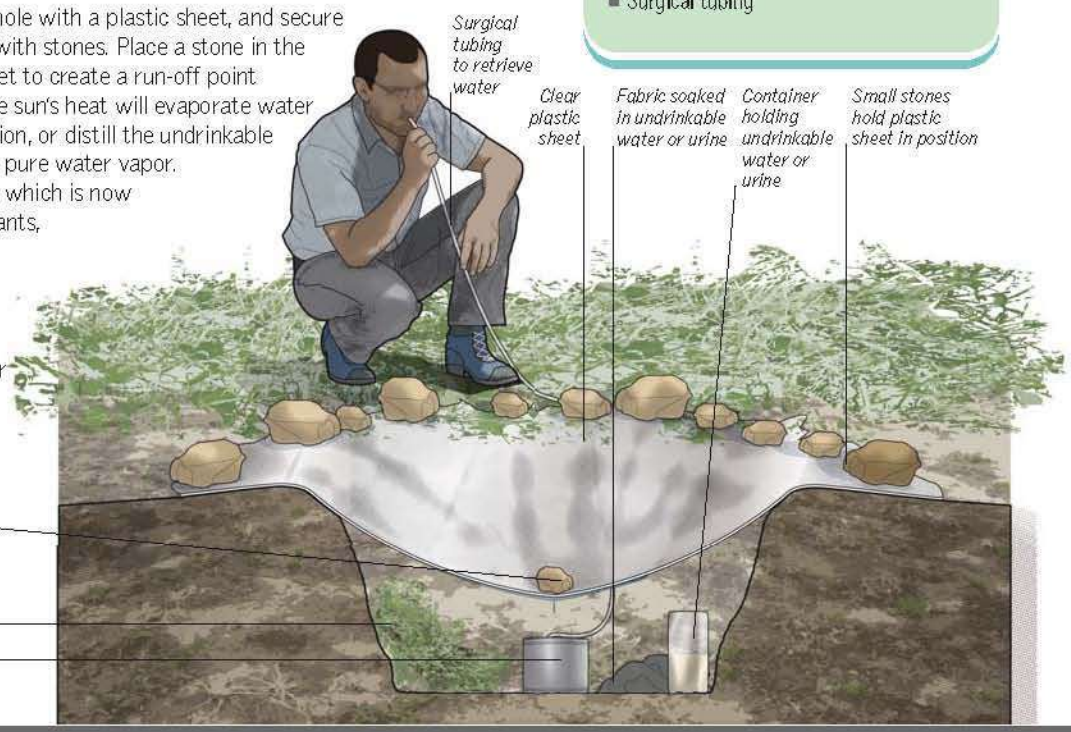

MAKING A GYPSY WELL

A good method of using the ground to create cleaner water from stagnant water sources, a gypsy well can also be used to collect water from saturated ground. The water collected through this method will still need to be treated before it is safe to drink (see p. 201).

1 Dig a hole a few feet away from a stagnant water source. It should be wide enough for you to be able to dip into it with a container and at least 12 in (30 cm) below the first layer of saturated soil. The hole will quickly fill with water.

2 Bail out the muddy water and allow the hole to refill—you may have to repeat this process several times. Once the water is clear, it can be collected, treated, and used. Cover the well when not in use to prevent debris or small animals falling in.

SURGICAL TUBING

Always pack a length of surgical tubing in your survival kit. It doesn't take up much space and is an extremely versatile piece of equipment:

- It can be used as an improvised straw to reach water trapped in rocks, hollows, and trees.
- It allows you to access water procured from a solar still without taking the still apart.
- It can be built into transpiration and vegetation bags so that you can remove water without having to open the bag.

FINDING WATER: HOT CLIMATES

YOUR NEED FOR WATER increases in hot climates as your body starts to use more of its water supply to regulate its temperature through perspiration. If you fail to drink more than you perspire, you will start to dehydrate and even approaching the first stages of dehydration can affect your chances of survival. Hot climates can be divided into two groups: hot-humid and hot-dry.

WARNING!
Water sources are plentiful in jungle and rain forest environments and, under normal circumstances, you should have no problem getting hold of enough water to satisfy your needs. However, rivers and streams may not be available during certain seasons, or if you have climbed too high, so knowledge of other water procurement methods could be crucial to your survival.

HOT-HUMID CLIMATES

The hot-humid conditions found in jungles and rain forests mean that procuring water is rarely an issue. However, the body's need for water in these conditions should not be underestimated; it is not unusual to have to drink up to 3 ¾ gallons (14 liters) a day to avoid dehydration.

SOURCING WATER FROM VEGETATION

Many plants, such as pitcher plants, have hollow parts that collect rainfall or dew. Some trees store and catch rainwater in natural receptacles, such as cracks or hollows. In an emergency, life-saving liquid can be garnered from a tree's roots or sap. You can find water trapped in the sections of green bamboo by carefully cutting into the bamboo with your machete or knife; or collect small, unripe coconuts and quench your thirst with the fluid they contain; or make a spigot and tap into the water contained inside the water tree. Life-saving fluid is everywhere in the jungle, and you don't have to look far to find it.

After making a second cut in the water vine, the fluid it contains will start to flow

WATER VINES

Found throughout the jungles and rain forests of tropical regions, water vines are easily identifiable by their size and shape and can provide an excellent sort of fresh water. However, bear in mind that not all water vines are water-bearing; not all contain drinkable water and some even contain poisonous sap.

1. Most water-bearing vines are about 2 in (5 cm) in diameter. If you think you have found one, make a small cut in the vine with a machete and check the color of the sap. If the fluid is milky, don't drink it; if it's clear, then the water in the vine will be safe to drink, so cut through the vine as high as you can with a knife or machete.

2. Cut off the vine at a point lower than the first cut. The liquid, which has a neutral, fruity taste, will start to flow. Don't let the vine touch your mouth as the bark may contain irritants.

3. The pores in the upper end of the vine may re-close, stopping the flow of water. To rectify this, simply cut the top of the vine again with your machete.

CATCHING RAINWATER

Catching rainwater is the best way to procure water: it is passive and requires no energy to collect once you place your containers. There are many forms of catchment device, but make sure you filter and purify the water (see pp. 200–01) before drinking it.

BAMBOO ROOF

Construct a sloped bamboo roof with a bamboo gutter. This could be the roof of your shelter (see p. 169), but if water procurement is a problem, you will have to construct additional bamboo roofs.

WIDE-LEAF ROOF

If you're in an area where there are wide-leafed plants, it's easy to construct a roof from them. Overlay the leaves as you would with roof tiles, working from the bottom to the top (see p. 170). Placed in this way, they will allow the water to run to the bottom. A length of bamboo cut in half lengthways can be placed as a collection gutter at the bottom.

BAMBOO DRAINPIPE

Observe rainwater's route down a tree trunk and tie a length of bamboo that has been cut in half lengthwise in its path. Place the other end of the bamboo into a suitable container.

DRIP RAGS

Wrapping any absorbent material around a leaning tree, such as a rag or a T-shirt, will result in the water running down the tree and soaking the material. Shape the rag to form a low point from which water will drip and place a suitable container underneath to catch the water.

HOT-DRY CLIMATES

Anyone venturing into this environment should have sufficient water for their needs, plus an emergency supply just in case, otherwise they have no right to enter the desert. Green vegetation usually signifies water or moisture in some form, and many techniques of procuring water in temperate climates (see pp. 188–91) may work in some desert conditions.

SOURCES OF EMERGENCY FLUID

If there is no surface water to be found, and you have no other means of procuring water, a water-yielding plant may be your only option. In some plants, the clear sap, fruit, or trapped rainwater, may quench your immediate thirst, but do not rely on these sources to keep you alive for long.

BARREL-SHAPED CACTUS

Barrel-shaped cacti contain a milky fluid that's safe to drink. Carefully remove the top of the cactus with a machete, mash up the flesh inside with a stick to make a pulp, and extract the juice from the pulp by sucking through a hollow grass stem. Alternatively, use a cloth to soak up as much of the fluid as possible and then wring the material to extract the fluid. The rewards from both techniques are minimal.

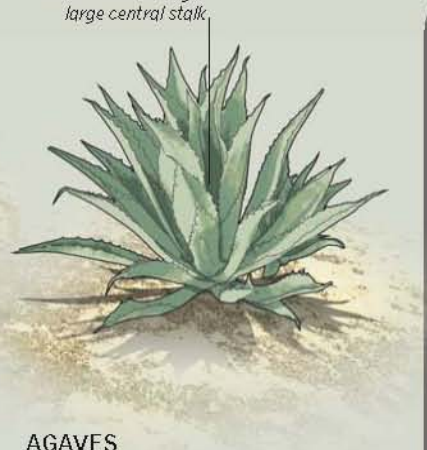

AGAVES

Native to Mexico and the southern and western United States, agaves have a rosette of thick, fleshy leaves containing fluid that is safe to drink. Cut the huge flower stalk with a knife or machete and collect the juice.

PRICKLY PEARS

Found in low-spreading clumps measuring 3 ft (0.9 m) in height, and native to dry, sandy soils throughout the world, the prickly pear cactus has an edible fruit that can provide a life-saving amount of fluid.

FINDING UNDERGROUND WATER

Water is life in the desert; wherever there is life there'll be water. Unusual clusters of green plants may indicate a minor presence of water; an abundance of greenery may indicate a more substantial water source. Water is rarely found above a depth of 6 ft (1.8 m) below the surface, so you will have to dig for it. If you do, make sure you only do so during the coolest part of the day. Bear in mind that, despite what you may have seen in movies, or read in other books, your chances of procuring water from the inside bend of a dry wadi in a real survival situation are slim to non-existent. In fact, you're more likely to waste what water you do have looking for the dry wadi and die in the process!

THINK LATERALLY

The chances are that if you're having to procure water in the desert you'll already be in a desperate situation. Bear in mind that, in addition to the methods mentioned above, if you have knowledge of the techniques outlined in the temperate environment section (see pp.190–91)—such as solar stills, dew traps, and vegetation and transpiration bags—it could be enough to keep you alive in the desert.

CHECKLIST FOR HOT-DRY CLIMATES

Even the best-laid plans can hit unforeseen problems, but many problems encountered in a desert environment can be avoided with some prior preparation.

- Always start hydrated.
- Carry enough water for your needs plus emergency water; your emergency supply should be enough to get you out of danger.
- Monitor your progress against the water you use. If you're using more than you thought, re-evaluate what you want to accomplish. It's better to turn back and learn from the mistake than to push on and create a survival situation that need not exist.
- Cache water ahead if necessary.
- Check your map for probable water sources. Confirm the reliability of these sources with locals and ask if there are any sources, such as wells, that are not shown on the map.
- Mark your map, or waypoint your GPS, with any water sources you sight as you progress. It's better to go back to a known source than to move on with nothing more than hope.
- To keep the water cool, always keep the bottles in the shade or in a windy location.

FINDING WATER: COLD CLIMATES

ADOPT THE SAME METHODS of trying to find drinking water in cold climates as you would in temperate ones (see pp. 186–91), although the ability to procure water becomes a major problem in freezing temperatures. The cold-weather survivor faces a dilemma: he or she is surrounded by water, but could die of thirst because most of that water is frozen.

WARNING!
Never try to melt ice or snow in your mouth, as it can cause freezing injuries to your mouth and lips. What's more, your body will expend heat as it melts the ice, and this could cause you to cross that very fine survival line between being cold or being hypothermic.

CHECKLIST FOR COLD CLIMATES

You should prioritize water usage in cold-weather conditions in the same manner as you would in desert conditions (see p. 193).

- One of your first thoughts should be about procuring water. You will have to find a water source that's close to everything you need to build and maintain a fire.
- Always look for an alternative water source before trying to melt snow or ice. It's easier, less time-consuming, and more fuel-efficient to fill your water containers with natural meltwater than it is to melt snow or ice by the heat of a fire.
- Your ability to procure water in freezing conditions will be directly related to your ability to start and maintain a fire.
- Be aware that it takes time and patience to melt snow and ice. You'll also need enough fuel to maintain a fire for a considerable period of time.
- Regulate your body heat to minimize overheating and sweating.
- Keep drinking water close to your body to prevent it from freezing, but avoid having water containers directly next to your skin. Instead, keep them between layers of clothing and use the warmer air trapped between the layers to help raise the water's temperature.
- Do not use recently frozen sea water, as it contains high levels of salt.

MELTING ICE AND SNOW

If you have the choice between melting ice and snow, favor ice, as it melts more quickly than snow and is up to 17 times denser. However, if you can't find ice, use dense, compact snow. Always look for the whitest, purest-looking ice or snow.

MELTING ICE

If you have some water to begin with, pour some into a container and heat it over a fire. Break the ice into small pieces—rather than adding it in one lump—and keep adding pieces to the container to melt them. Keep the water hot, but not boiling, so that you avoid losing water through evaporation.

USING A HOT PLATFORM

If you don't already have water for the method above, you can melt ice slowly on a gently sloping platform fashioned from any flat piece of stone or wood positioned above a fire.

1. Build a fire. Search for a large stone with a flat surface and two logs, or smaller stones, that are strong enough to support it. Place the small logs or stones on either side of the fire as a support for the flat-surfaced stone. Make sure that the platform is at an angle. This will create a natural run-off for the ice as it melts. Place a block of ice in the center of the platform.

2. As the fire starts to heat up the platform, the ice will begin to melt. The meltwater will flow off the platform, where it can be collected in a container, such as a mess tin.

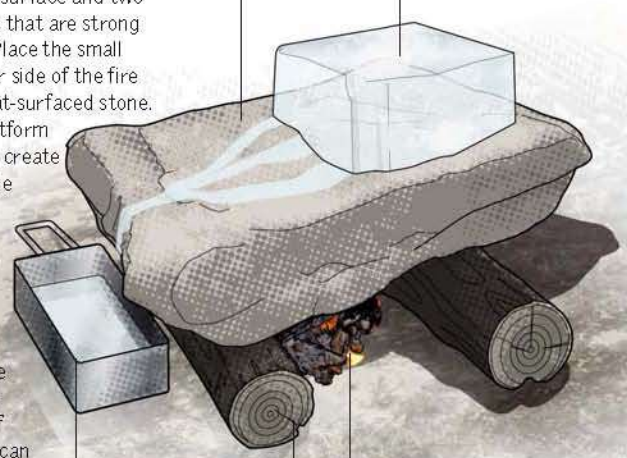

MELTING SNOW

If you already have water, follow the techniques for melting ice (see opposite), heating a little water and adding the snow little by little. Don't pack the snow too tightly; if an air pocket forms, the heat from the fire will be absorbed by the metal container rather than by the snow and could result in the fire burning a hole through the metal container before it melts the snow.

MAKING A FINNISH MARSHMALLOW

Cut a solid piece of dense snow—often referred to as a "marshmallow" or a "snowman's head"—and skewer it with a stick. Secure the stick in the ground close enough to a fire that it receives heat, and position a suitable container underneath it to capture the water as it melts.

STORING WATER

Snow is a great insulator: even if the temperature dips to -40°F (-40°C), water in a bottle will remain largely unfrozen if placed under at least 1 ft (0.3 m) of snow. Make sure you store the bottles upside down. That way, if some of the water does freeze, it will freeze at the bottom of the bottle and not at the top.

USING A MELTING SACK

Using a similar principle to the Finnish marshmallow (see left), this technique involves using an improvised sack—made from any porous material, such as a T-shirt or a sock—suspended near to a fire. The heat from the fire will start to melt the snow, which can then be collected in a well-positioned container.

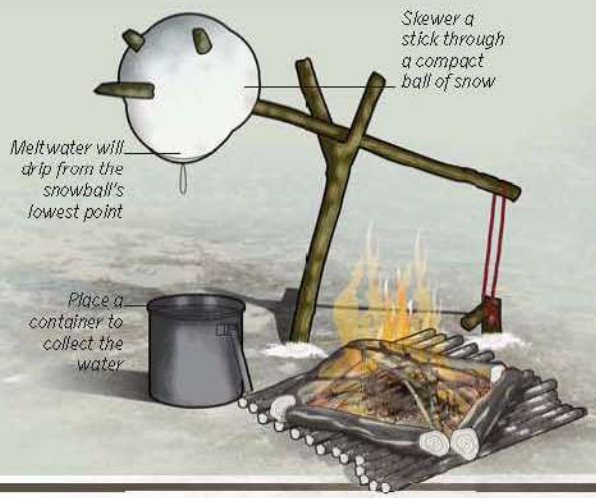

IN A WORST-CASE SCENARIO

If you're caught out on barren snow or ice, you're unlikely to have access to natural fuel to burn in order to melt snow or ice into water. As long as you have your second-line survival equipment (see p. 42), you'll have all you need to procure life-saving water.

USING YOUR FIRE TIN

Set up your fire tin on the ground and shelter it from any wind using either your body or your pack. Using your survival tin as a container, place small quantities of ice or snow into the tin. Light your fire tin and place your survival tin over the flame. Add more ice or snow as it starts to melt.

USING YOUR MATCHLESS FIRESET

Place the lid of the matchless fireset tin on the ground. Light a hexi fuel block using the cotton wool and flint/striker. Using the fireset container or your survival tin, melt small quantities of ice or snow over the flame. Once the ice or snow has melted, let the container cool so you don't burn yourself. The water is then safe to drink.

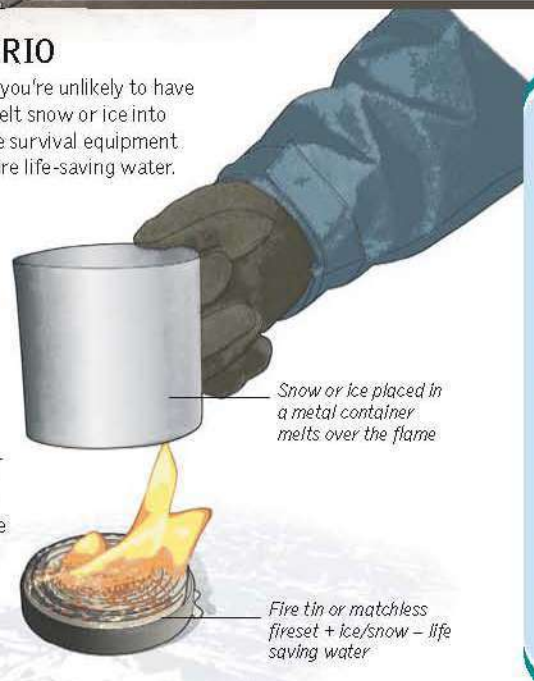

USING BODY HEAT

When Inuit hunters capture a caribou, they empty out the contents of its stomach, turn the stomach inside out, fill it with snow, and tie it shut with a length of intestine. They then put the stomach back in the cavity while they skin the caribou. By the time they have finished, the heat given off by the animal's body has melted the snow. The Inuit then open the bag carefully and suck out the water through a block of snow to filter it.

FINDING WATER: AT SEA

OF ALL ENVIRONMENTS, the sea is possibly the most difficult in a survival situation. It offers no natural resources for protection against the extremes of temperatures, wind, rain, and sea state, and provides little to aid location. Being surrounded by water that you can't drink only adds to the difficulty. Some devices are capable of making seawater safe do drink, but if you don't have one, you'll have to find a way of procuring enough fresh water to keep you alive.

WARNING!
Never drink salt water. Its salt concentration is three times higher than that of blood and ingesting it will dehydrate you. Continued use over a prolonged period will lead to kidney failure and, ultimately, death.

CONSERVING YOUR WATER RATIONS AT SEA

Rationing fresh water supplies when you're adrift at sea is a sensible precaution, as you have no idea how long it will be before you're rescued or before you reach land. Here are a few tips to help you conserve what water you have while you're at sea:

- Fix your daily water ration after taking stock of the amount of water you have, the output of solar stills and desalting kits, and the number and physical condition of your party.
- Prevent fresh water supplies from becoming contaminated by saltwater.
- Keep water supplies well shaded, both from the overhead sun and from the glare off the sea's surface.
- In hot conditions, dampening your clothes with saltwater can help to lower your body temperature—but don't overdo it. This is a trade-off between cooling yourself down and the saltwater boils and rashes that will result from continued exposure.
- Don't exert yourself. Relax and sleep whenever possible.
- Use every container you have—even a simple trash bag—to collect rainwater, and keep them well sealed and attached to the raft.
- If you don't have water, don't eat. Protein consumption will hasten the onset of dehydration.

COLLECTING FRESH WATER

If you find yourself adrift with no hope of immediate rescue, obtaining drinking water will be a major priority. If you don't have a solar still or a reverse-osmosis pump, you must find another means of procuring fresh water. Fortunately, there are several ways of doing so.

GATHERING RAINWATER

Most modern liferafts incorporate a built-in rainwater catchment system that channels rainwater and dew from the outer surface of the liferaft into collection pockets inside the liferaft. However, even if you're not in a liferaft, you can construct a similar system using a tarpaulin or any other waterproof material. Watch the clouds, be ready for the possibility of any showers (see pp. 80–81) and spread your tarpaulin in a bowl shape to catch the largest amount of rainwater possible. Always place a tarpaulin before nightfall, so that you don't miss out on any overnight rainfall.

HARVESTING DEW

At night, secure the tarpaulin like a sunshade and turn up its edges to capture dew. It's also possible to harvest any dew that may have collected on the sides of the raft using a sponge or cloth that you then wring out.

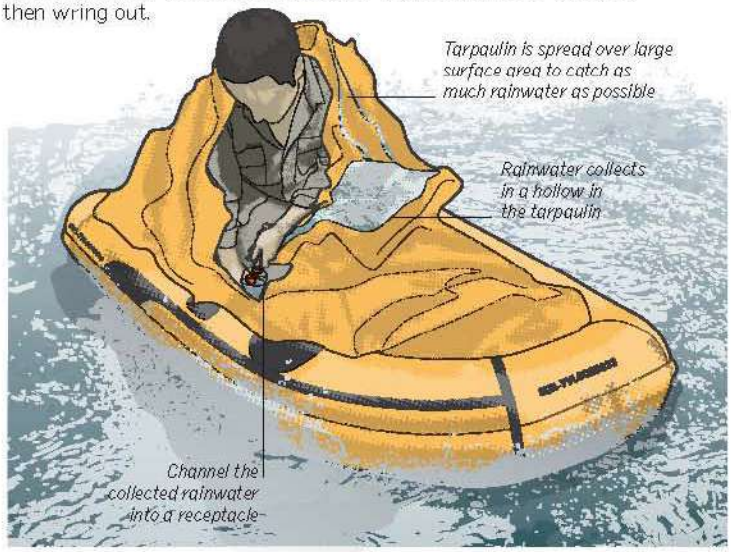

TREATING SALTWATER

If you have no means of collecting rainwater or dew, there are several products capable of turning undrinkable saltwater into fresh water. Although these products are standard issue on most liferafts, you should always try to have at least one of them with you if you are venturing into a marine environment.

SOLAR STILL

Solar stills are a simple way of distilling water using the power of the sun. Saltwater is placed at the bottom of the container, where it is evaporated by the sun through clear plastic. Pure water condenses on the top of the plastic and drips down to the side, where it can be collected, often via a tube. Most solar stills on modern liferafts are inflatable.

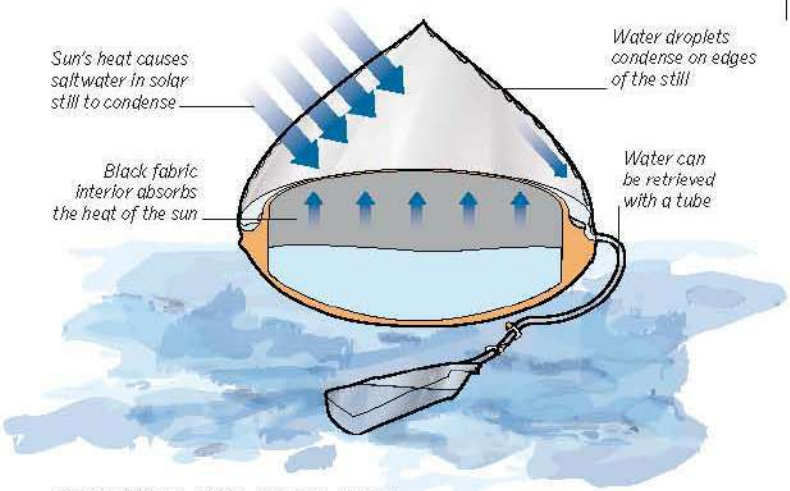

MAKE YOUR OWN SOLAR STILL

If you have a chance to gather together the right materials, it's easy to make a small solar still. All you need are two containers (one larger than the other), a sheet of plastic, some string, and a weight to form a natural run-off point for the water as it condenses under the heat of the sun. If you have some surgical tubing, you can use it to retrieve the fresh water without taking the still apart.

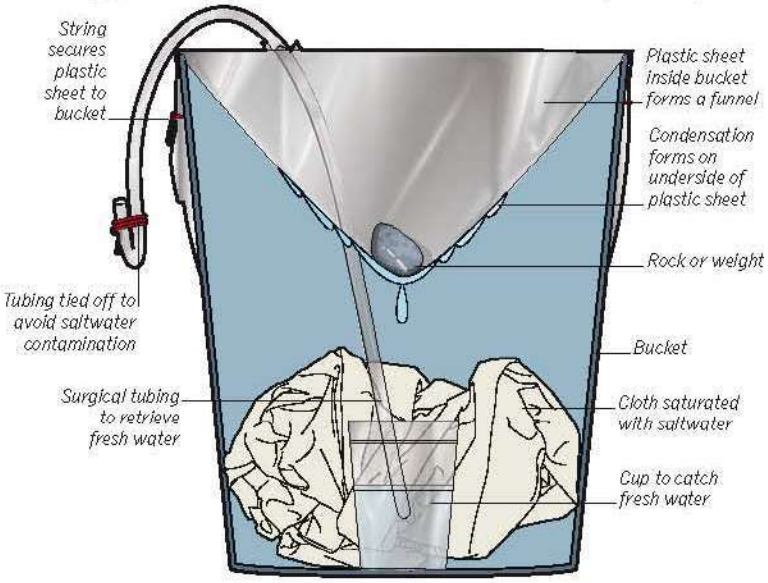

REVERSE-OSMOSIS PUMP

These hand-powered devices pump seawater at a very high pressure through a membrane that filters out the salt. Depending on the model, they can produce around 23 liters (49 pints) of fresh water per day.

DESALTING KIT

These kits turn seawater into freshwater through a process called "ion exchange." Because they only produce small amounts of fresh water over several hours, use desalting kits only during long periods of overcast weather when you can't use a solar still.

THE LAST RESORT

In absolute emergencies, potentially life-saving forms of liquid can be obtained from the ocean.

SEA ICE

In Arctic seas, you can obtain drinking water from old sea ice. This ice is bluish in color, has rounded corners, splinters easily and, more importantly, is nearly free of salt. New ice is gray, milky, hard, and salty. Water from icebergs is fresh, but icebergs are dangerous to approach, so you should only use them in an emergency.

FISH

Drink the aqueous fluid found along the spine and in the eyes of large fish. Carefully cut the fish in half to get to the fluid along the spine and suck the eye. If you are so short of water that you need to do this, do not drink any of the other fluids: they are rich in protein and fat and your body will use up more of its water digesting them than it obtains from them.

SEA TURTLE

Sea turtle blood has a salt concentration similar to that of humans. The blood can be collected by slitting the turtle's throat. Note that, although this may help prolong survival, sea turtles are an endangered species, so you should only kill one as a last resort.

REHYDRATION ENEMA

If you have some water that is not salty or poisonous, but is too foul to take orally, you can absorb as much as a pint a day—enough to keep you alive—through the large intestine using an improvised tubing device.

CARRYING AND STORING WATER

YOU WILL NEED TO make many decisions in any survival situation, some of which could mean the difference between life and death. One of these decisions will be whether to stay and wait for rescue or to attempt self-rescue. A major factor in this decision will be the availability of water and your ability to carry and store it.

WARNING!
Mark your containers clearly. Late at night in camp it's easy to confuse bottles that contain treated water with ones that contain untreated water or camping fuel.

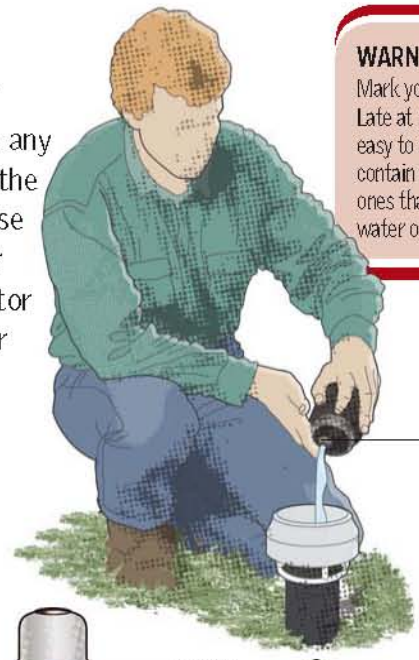

WATER CONTAINERS

There are many different kinds of container for carrying both hot and cold liquids. They range from solid plastic or steel flasks or bottles, to collapsible waterproof bags that can be folded up when they're empty.

PLASTIC WATER BOTTLE
Strong and lightweight, with screw-off tops or push-pull nozzles.

METAL WATER BOTTLE
Slightly heavier, but stronger, than plastic water bottles.

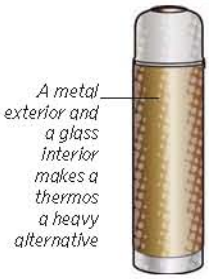

THERMOS FLASK
Although very heavy, thermos flasks allow you to store either hot or cold water.

MILITARY WATER BOTTLE
A standard issue bottle used by most of the world's military forces.

COLLAPSIBLE WATER CANTEEN
Heavy-duty plastic water canteen that can be hung round the neck.

HYDRATION SYSTEMS

Hydration systems usually consist of a water-storage container (called a "reservoir"), an on-demand, one-way drinking tube, and a harness. However, a common problem with these systems is that, because you can't see the water, you can suddenly discover you have used all of it without realizing.

COLLAPSIBLE WATER BOWL
Easy to pack, light, and can also be used to seperate items in your backpack.

WATER BAGS
Useful for carrying water from its source back to your camp. Can also be used to keep your equipment dry.

IMPROVISING WATER CONTAINERS

In an emergency, there's a strong chance that you won't have the luxury of having water bottles or storage systems at your disposal. However, with luck—and a little ingenuity—you should be able to find something that allows you to store and carry enough water to satisfy your needs.

SEA SURVIVAL SUIT

Because they are designed to keep water out, they can also be used to keep water in. Most survival suits are packed with a layer of chalk to prevent the material from sticking together, so wash it out before use.

WATERPROOF CLOTHING

Many types of waterproof clothing can be adapted to hold water. Jacket sleeves and trouser legs can be knotted to form a basic bladder; Gore-Tex® socks will hold water; and some waterproof backpacks can be used for water storage.

DISCARDED MATERIALS

Always be on the lookout for anything that can hold water, from empty plastic bags and bottles, to large industrial-type containers. Clean and sterilize found items before use.

GOURDS

The shell of a hollowed-out and dried fruit—a gourd—can hold water. Gourds made from large fruits, such as squashes, pumpkins, and melons, can hold a considerable amount of water.

EMERGENCY WATER

Sachets of emergency water can be bought and stored for emergency use. They usually come in a packet that contains five 1⅞fl oz (50 ml) sachets of water, each being little more than a mouthful. These should always only be used as a last resort. Always try to procure water by other means before turning to your emergency water supply.

BAMBOO

A bamboo cane will provide a natural cup if you cut 1 in (2½cm) below one joint and then 1 in (2½cm) below the next. Take care to smooth the edges after cutting.

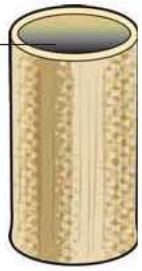

MAKING A MINI-RESERVOIR

Your ability to store sufficient water for your needs in a survival situation may be limited and it may be impractical —because of injury or distance—to travel back and forth for water every time you need it. If you have some basic materials, making a mini-reservoir to catch rainwater (see also p. 188) solves this problem.

1. To reduce the amount of effort required, choose a site that offers the least resistance (such as a natural hollow). Dig a shallow pit using whatever materials you have to hand, such as a stick.

2. Smooth the edges of the pit and line it with waterproof material, such as a survival blanket. Weight the edges with rocks, soil, or logs, making sure that dirt can't run into the pond when it rains.

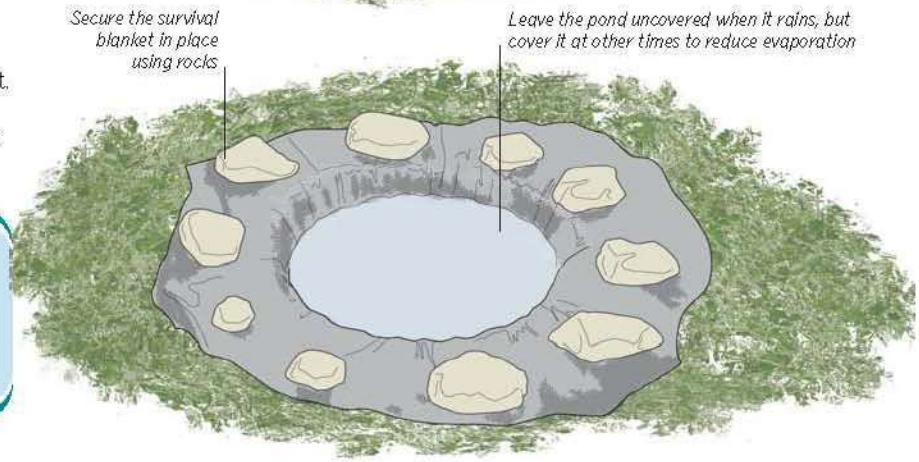

WATER HYGIENE

Water stored in a mini-reservoir for any period of time will need to be filtered and purified before it's safe to drink (see pp. 200–01).

TREATING WATER

WITH THE EXCEPTION OF rainwater, all other water procured in a survival situation should be treated before it is safe to drink to remove or destroy harmful pathogens and microorganisms that could lead to gastrointestinal illness.

FILTERING WATER

If you don't have a device that filters and disinfects the water, you will have to accomplish the same task in two stages. Before purifying the water, you will have to filter it to remove any debris. You could use the popular Millbank bag (see opposite), but it may be necessary for you to construct your own improvised filter.

MILLBANK BAG

Used extensively by military forces around the world, the Millbank bag is an effective water filter that packs down small and can be used many times to produce large quantities of filtered water. Set up the bag as soon as you arrive in camp, because the process is quite slow: the Millbank bag will filter 1 liter (2.1 pints) of water in five minutes. Note that water filtered in this way will still need to be purified.

BUILDING A TRIPOD FILTER

If you're not carrying a filtration system with you such as a Millbank bag, you could make an improvised filter. All you need are three sticks to form a tripod, and some materials to create three separate layers.

1. Use three sticks or a bent sapling to form a tripod. Using any materials you have, form layers, starting at the top with the coarsest material and working your way down using finer materials—such as parachute silk or nylon—as you go.
2. Pour water into the top layer. It will become filtered as it passes through the increasingly fine materials.

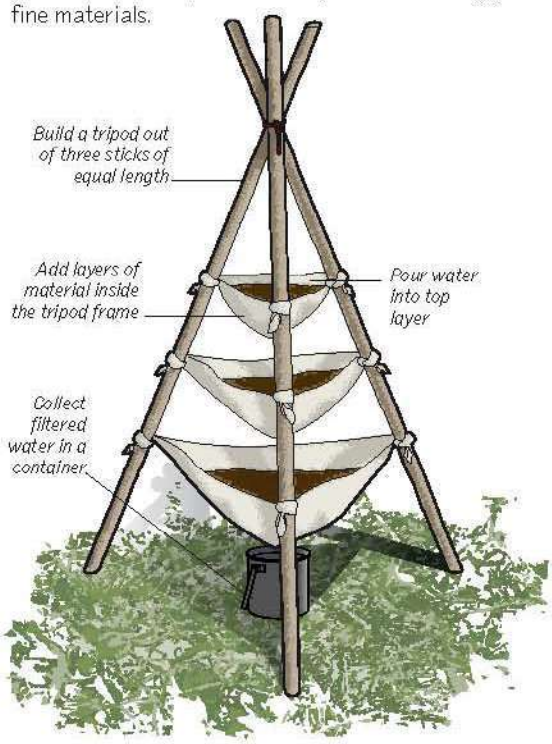

MAKING A BOTTLE FILTER

To make an improvised bottle filter, take a container, such as a plastic bottle, and cut off the bottom (or make a hole). Note that a sock used in the same manner is still effective.

1. Hang the bottle upside down from a branch. Fill the bottle with layers of different materials working from coarse to fine as you work your way down the container.
2. Pour the water into the top end of the bottle and allow it to work its way down through the layers.

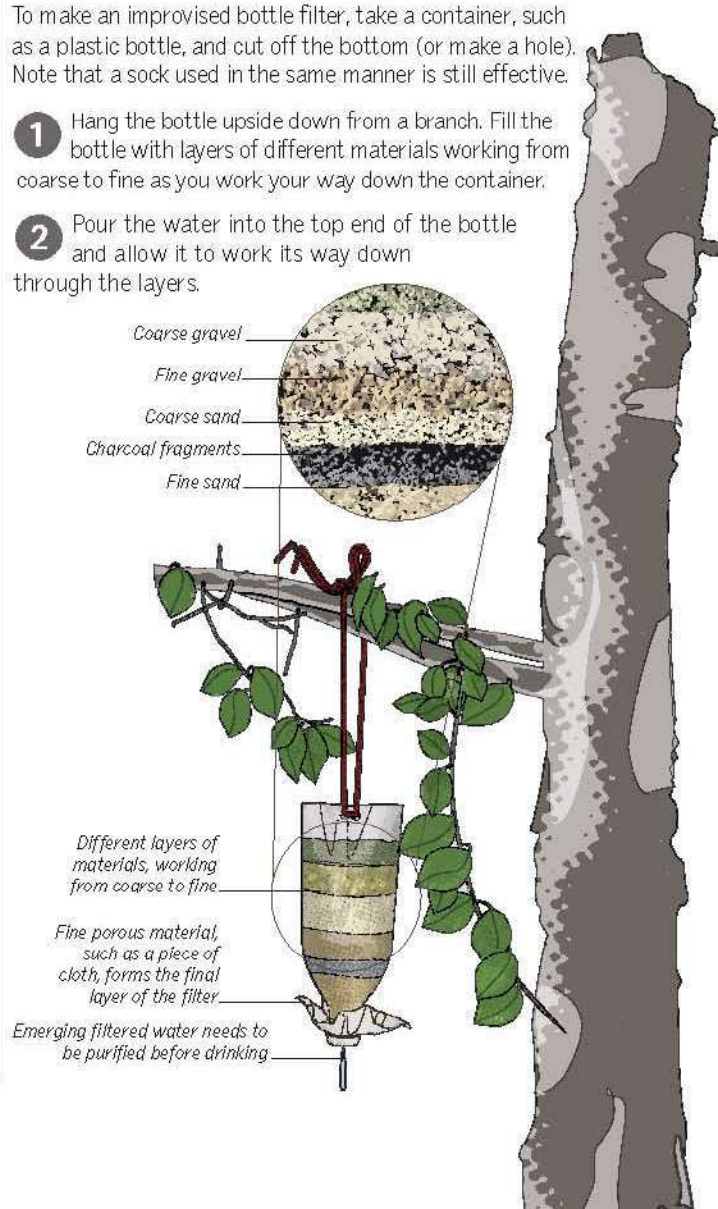

DISINFECTING WATER

If you drink untreated water, you run the risk of becoming infected with a water-borne disease (see p. 187), so it's vital that you treat any water first. If you have the ability to start a fire, the most effective way of making water safe to drink is by boiling it. However, if you are unable to make a fire, there are several devices available that are capable of filtering and purifying water to make it safe for drinking.

MINI PORTABLE WATER PURIFIERS

These are specially designed units that filter the water and then purify it, by pumping the contaminated water through either microfilters, chemicals, or a combination of both. Sizes vary from small emergency pumps capable of purifying up to 13 gallons (50 liters) of water, to larger units that can filter huge quantities of water.

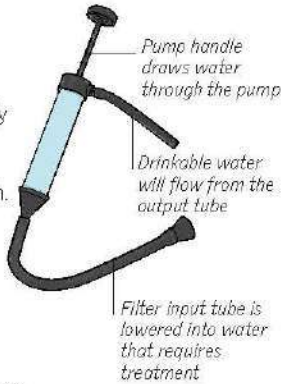

GRAVITY/PRESSURE FILTERS

These devices are incorporated within drinking bottles. The water is either allowed to flow naturally through the system via gravity, or is squeezed through by the operator. All bottles of this type usually employ: a filter to remove sediment and organic contaminants; a micron filter to remove protozoa; and a chemical that kills water-borne bacteria and viruses.

SURVIVAL STRAWS

Compact emergency water purifiers, survival straws contain a filter system and employ either carbon or iodine resin systems to eliminate water-borne diseases and harmful chemicals. You need to get the water to a point where you can reach it with your straw. If you want to draw a supply of water to carry with you, you'll have to draw the water into your mouth and decant it into a container—a laborious process.

OTHER METHODS

If you can't boil water, or if you don't have a water-purification device, you'll have to rely on non-mechanical techniques. The concentration and contact time required for some of these methods is dictated by the quality and temperature of the water being treated.

"BIG BUBBLES, NO TROUBLES"

Microorganisms and virtually all intestinal pathogens are killed at temperatures well below boiling point. The process of bringing water to the boil is sufficient to disinfect it—continuing to boil it just wastes fuel, time, and water.

METHOD	DESCRIPTION
Iodine (Liquid and Tablets)	Iodine—which destroys bacteria, viruses, and cysts—can be used to disinfect water effectively and conveniently. Its action is dependent on its concentration, the water temperature, and duration of contact—a concentration of 8 mgs per liter at 68°F (20°C) will destroy all pathogens if left for ten minutes.
Chlorine tablets	Chlorine-based tablets will destroy most bacteria, but are less effective for viruses and cysts. They are more effective when used in combination with phosphoric acid and will destroy both Giardiasis and Cryptosporidium.
Potassium Permanganate	Potassium permanganate can be bought at most pharmacies. Mix a few granules with your water until it turns light pink. Leave for at least 30 minutes before drinking.
Bleach	Adding unscented household bleach is the cheapest way of adding chlorine to water (it contains 5 percent sodium hypochlorite). Be careful to add just one drop of bleach per liter of water—two if the water is cloudy—and leave for at least 30 minutes before drinking. This method is not always effective against Giardiasis and Cryptosporidium.
Ultraviolet (UV) Light	When many harmful microorganisms are exposed to UV light, the process of light absorption disrupts the cell's DNA, rendering the organism harmless. The quality of the water will affect the amount of exposure to UV light required: the cloudier the water the harder it will be for UV to penetrate.
UV Passive	Fill plastic bottles with water, replace the lids, and place them in direct sunlight, preferably on a dark surface. The sun's rays will kill the bacteria that cause common water-borne diseases.
UV Active (Steripen)	A small UV purifier that is placed in pre-filtered water and activated for a short time. Some models can purify up to 1 liter (2.1 pints) of water in as little as 48 seconds.

FIND AND PREPARE FOOD

Your body converts food into fuel, which provides you with heat and energy, and helps you to recover from hard work, injury, or sickness. If you are healthy, your body can survive for weeks without food by using the reserves stored in its tissues—although you will use approximately 70 calories per hour just breathing and up to 5,500 calories a day if laboring hard.

In a short-term survival situation, food should not be your major priority. You would probably have eaten recently and, if you'd prepared properly, you should have some basic emergency food in your pack. While you may go through food withdrawal symptoms—when your stomach complains because the food it's expecting doesn't arrive—you're not going to die of starvation within a few days. However, the body will react to its fuel not being replaced: hunger, a lack of energy, and a deterioration in coordination can be expected after a few days. If the opportunity to procure food arises, it should always be taken. Eat little and often but always make sure you have sufficient water to digest it.

In this section YOU WILL DISCOVER...

- how to **cook** with **hot rocks**...
- how to **make** a **fishing reel** from a **beverage can**...
- why a **snare** must have a **perfect end**...
- how to **lasso a lizard** and **snare a squirrel**...
- that there's **more than one way** to **skin a rabbit**...
- which **grubs taste** like **scrambled eggs**...
- how to **catch** a **bird** in a **bush**...

In a long-term survival situation, your survival priorities will change and the need for food in order to simply survive will become more important. There is a thin line between food not being your priority and then subsequently finding that you're in no physical condition to do anything about it when it does become a priority—you should regularly reassess your situation and alter your plans accordingly. It takes effort, skill, and a certain amount of luck to obtain food in the wild, especially if you're not in your natural environment.

When gathering food in the wild, always ensure that the energy gained from the food is more than the energy you expended in procuring it, otherwise it's a wasteful exercise.

FOOD THAT'S EASY TO FIND and gather should always be your first priority:

- **Plants** are easy to collect and, as long as they're readily available in the environment you're in, should be your first choice for food. However, make sure that you are absolutely sure that they are edible—mistakenly eating the wrong leaf or berry could cause vomiting and diarrhea, making your situation worse.
- **Fishing** requires little effort once the lines or traps have been set, and they will work for you around the clock. Fish is high in protein and relatively simple to prepare and cook.
- **Insects, reptiles, and amphibians** may also be available, but be careful that you don't expend more energy in catching them than you gain in nutrients from eating them. Remember also that many insects, snakes, and amphibians are poisonous, but they can still be used as bait for fish and mammals.
- **Birds and mammals** have their own survival mechanisms and are wary of humans, especially in remote areas where contact with humans is limited. Even if caught, the bird or mammal will need to be killed, plucked or skinned, and cooked.

> "Always ensure that the **energy gained** from the **food** is **more than** the **energy you expended** in **procuring it**, otherwise it's a **wasteful exercise**"

WILD COOKING

IN A SURVIVAL SITUATION, you must cook every item of food that you're not sure about in order to kill any parasites or harmful bacteria. While cooking reduces the nutritional value of food, it does help to improve the taste of many wild foods and to make them more digestible than if eaten raw.

FOUND MEAT
If you find a dead animal, you can eat the meat if you cut it into small pieces and boil it for 30 minutes. Don't touch it if you have cuts on your hands—you don't know how it died.

COOKING OVER AN OPEN FIRE

Use the fierce flames of a fire to boil water then, when the flames have died down, use the steady heat of the embers to cook on (see p. 121). The most basic method of cooking is roasting, using a spit made from green wood to suspend the food over a fire. Make sure the meat is thoroughly cooked.

BOILING OR STEWING

When you boil food, fat and natural juices are retained in the water. It is important to drink the water to obtain the maximum nutrition from the meal, unless you have been boiling toxic substances out of the food.

DAMPER BREAD

This simple method for making yeast-free bread was developed by stockmen working out in the Australian bush.

1. Mix flour and water (and a pinch of salt, if you have it) into a pliable dough, then roll it into a long, thick sausage shape.

2. Wind the dough around a stick, then hold it over the embers of a fire, turning regularly until it browns. It will slide easily off the stick when cooked.

STEAMING FOOD

Steaming leaches fewer nutrients away than boiling and is a particularly good way of cooking fish and green vegetables—fresh leaves will be ready to eat in just a few minutes. To steam food, it must be suspended above boiling water in some way.

BAMBOO STEAMER

Bamboo stems are tough, hollow, and divided into sections. Use a sharp stick to make small holes in the walls dividing a three-section piece of cane, leaving the wall at the base intact. Pour water into the stem until it's just below the bottom section ring. Add food to the top section and cover with grass and a loose-fitting bamboo lid. Lean the stem over a fire, propping it against a forked stick.

CLAY BAKING

Baking food in clay requires no cooking utensils. Animals must be cleaned and gutted first, but can otherwise just be covered in soft clay. When the meat is cooked, skin, spines, or feathers will remain embedded in the clay. Cooking root vegetables or fish in this way removes their skins, losing valuable nutrients, so it's best to wrap them in leaves first.

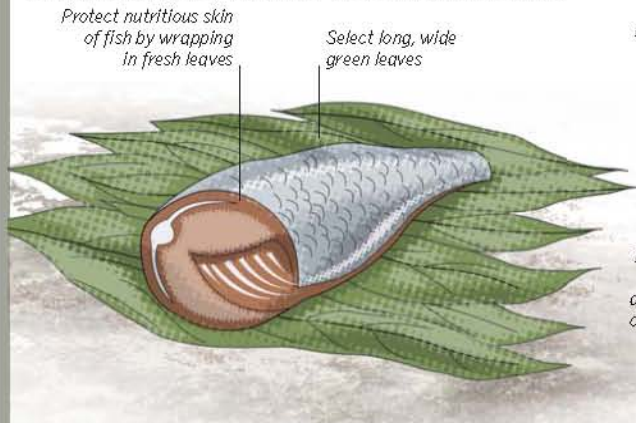

1 Wrap the food in a bundle of fresh green leaves and tie them in place with some long strands of grass to make a secure parcel. Use only leaves from plants that you have identified as non-toxic.

2 Cover the food parcel with an even layer of clay, making sure it's well sealed.

3 Place the parcel of clay in a bed of hot embers and build a fire up on top of it. Cooking time will be between 30 and 60 minutes, depending on the size of the food item. Break open the clay and remove the food.

COOKING WITH HOT ROCKS

Rocks take a while to heat up but they stay hot for a long time, allowing food to bake steadily on them. To reduce cooking time, cover the food with some birch bark or a flat piece of wood. When cooking on rocks, don't use slate or other layered rocks as they are likely to shatter when heated. Another method of cooking with hot rocks is to place them in a pit. The food is covered with leaves and placed on the rocks, then the pit is filled in to retain the heat. The food is left buried until it is ready to eat.

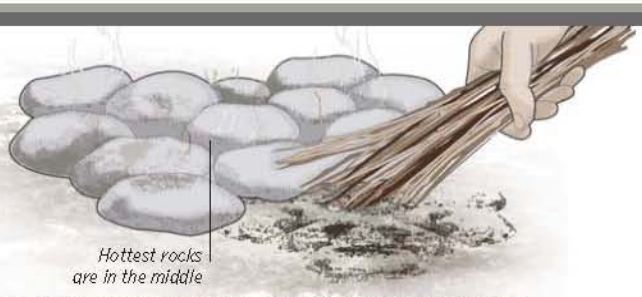

2 Brush the embers and ash off the rocks, taking care not to touch the rocks with your hands as they will now be extremely hot.

1 Place some large, reasonably flat rocks close together. Light a fire on top of the rocks using some tinder and dry sticks, then leave the fire to burn down to hot ashes while you prepare your food.

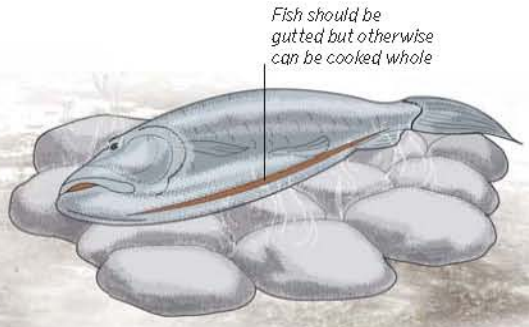

3 Place food on top of the rocks to bake. The hottest rocks will be in the center, so items that must be cooked more slowly should be put near the edges. Keep adding more food to cook until the rocks have cooled.

EDIBLE PLANTS

IN A SHORT-TERM survival situation, food is not a priority—identifying edible plants uses energy, requires knowledge and skill, and the risks of getting it wrong far outweigh the benefits. In a long-term survival situation, however, your priorities would change, so being able to identify edible plants is advantageous.

GATHERING PLANTS

When foraging for plants, take a bag or can with you and be careful not to crush what you collect. Gather only a few species to lessen the chances of mixing in something inedible or poisonous. Don't assume that because birds or mammals have eaten a plant it's safe for you to eat too. Unless you are absolutely certain, you will need to carry out the following test to find out whether a plant is edible.

WARNING!
The Universal Edibility Test (see below) does not apply to mushrooms. For this reason alone, you should never eat any mushrooms that you have gathered (especially in a survival situation) unless you are 100 percent positive that you can identify each specific mushroom as being edible. In 2008, a British woman died after eating a poisonous mushroom that had been foraged by mistake with edible mushrooms.

DIGGING FOR ROOTS

Roots and tubers are a good source of carbohydrate and their skins contain vitamins. Take care when gathering in spring as some plants will only have small shoots and will be hard to identify.

1. Cut a stick from a hard-wood tree, such as oak, sharpen one end to a chisel-shape then harden it in a fire.

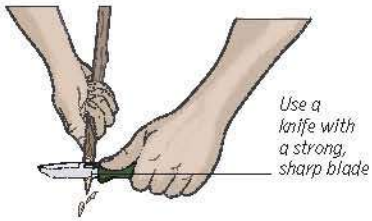

2. Choose a larger plant, then dig down at its side, loosening the earth around the root until it can be pulled out in one piece.

UNIVERSAL EDIBILITY TEST

This test enables you to check whether a plant is safe to eat. If you're in a group, only one of you should test the plant. Make sure that you have plenty of drinking water and firewood with you (see box, right), and eat nothing for eight hours beforehand. Test only one type of plant at a time, use only one part of that plant—roots, leaves, stalk, buds, fruit—at a time, and test it in the same state in which it will be eaten: either raw or cooked. Make sure that the plant you are testing is in plentiful supply (there's no point testing it otherwise) and avoid plants with milky or soapy sap, or bright colors—they are usually Nature's warning signal.

1. First, inspect the plant—it should look fresh and in good condition. Avoid anything slimy. Next, sniff the plant. Discard it if it smells bad. If it smells of peaches or bitter almonds, it may contain cyanide.

2. Take a small portion of the plant part you're testing and gently rub it on the sensitive skin inside your elbow or wrist. Wait for 15 minutes. If no irritation, stinging sensation, rash, or swelling develops, proceed to Step 3.

NEEDLE TEA

The needles of evergreen trees such as pine and spruce are rich in vitamins A and C and can be used to make a refreshing—and potentially life-saving—tea.

1. For each cup of tea, collect two teaspoons of fresh green needles and bruise them with a stone.

2. Drop the needles into boiling water, then let them infuse for 10 minutes. Keep the pot warm and stir occasionally.

3. Strain the tea through a cloth tied over a container. Sweeten with sugar or honey, if you have any.

3. Touch the corner of your mouth with the plant and wait 15 minutes. If no adverse reaction occurs, touch your lip and tongue with it. Wait another 15 minutes. If you suffer no ill effect, proceed to Step 4.

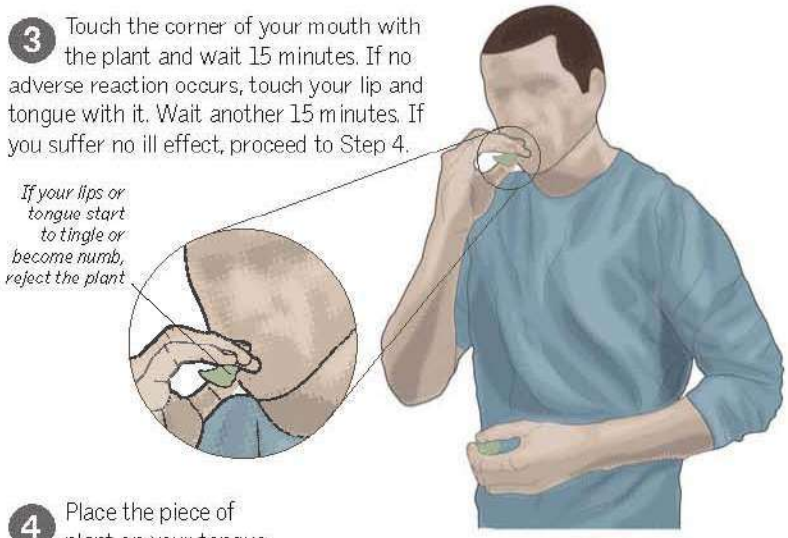

4. Place the piece of plant on your tongue, wait 15 minutes, then chew it and hold it in your mouth for 15 minutes more. If there are no ill effects, swallow the piece and wait eight hours. Eat nothing else but continue to drink water.

5. If no ill effect occurs, eat a quarter of a cup of the same part of the plant, prepared in the same way, then wait another eight hours. If there is no reaction, it is safe to eat. Eat little and often rather than gorging.

WHAT TO DO IF YOU HAVE AN ADVERSE REACTION

If at any stage you experience any type of adverse reaction to a plant during the edibility test, you should:

- Stop the test immediately.
- Avoid eating anything else until the symptoms cease.
- Wash the affected area thoroughly (if the reaction is external).
- Attempt to induce vomiting (if the reaction comes after eating the plant) by drinking salt water or by pushing a finger into the back of your throat.
- Drink lots of warm water (this is why you must ensure that you have a plentiful supply of purified water and firewood before you begin the test).
- Crush a teaspoonful of charcoal taken from a partially burned log and mix it into a paste with warm water, then swallow the mixture. It will either induce vomiting or, if you can keep it down, it may absorb the toxins.
- Try swallowing a paste made from white wood ash and water to reduce stomach pain.

If the tester is unconscious:
Do not induce vomiting if the tester is unconscious—get medical help and monitor the casualty.

CATCHING FISH

FISH CAN BE HOOKED, netted, trapped, speared, or even caught by hand if you're lucky. Observe their behavior—where and when they feed, and what they eat—to determine which method to try. It must be noted that some fishing techniques are illegal in some parts of the world and should only be used in a genuine survival situation.

BAIT AND LURES

Fish have a keen sense of smell to help them find their preferred food, so if one type of bait doesn't work, try another. Predatory fish are attracted to live bait by its movement—a worm wriggling on a hook, for example. You can make artificial bait out of shiny metal, cloth, or feathers. Try to imitate the fish's natural prey, such as an insect skimming across the water, to lure the fish into biting.

FISHING TACKLE

Fishing equipment can be made from all kinds of material. Your survival kit should have some fishing line and a few hooks, but if you don't have a kit you can improvise. You can use a stick as a makeshift rod, for example.

MAKING A FISHING REEL

You can make a reel with an empty beverage can. Tie one end of a line to the tab, then wrap it around the can until you have about 2 ft (60 cm) left. Attach a hook, float, and some sinkers. Hold the top of the can in one hand and the float in the other. Point the bottom of the can toward where you want the hook to land and throw the float. The rest of the line will unwind and follow. If a fish bites, give the line a jerk then wind it in around the can.

IMPROVISED HOOKS

Fishhooks can be fashioned from any piece of metal, such as a nail, pin, safety pin, needle, or some wire. You can also make them from natural materials, such as thorns, hard wood, coconut shells, bone, spines, or seashells. If you don't have any fishing line you can use parachute cord (see p. 136) or make cordage from plant fibers (see p. 138).

SINGLE THORN
Cut a 1 in (2.5 cm) length of bramble stem with a large, strong thorn. Tie a line to a notch at the other end.

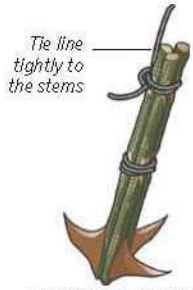

MANY THORNS
Tie three strong thorns together securely with some fishing line. Cut a notch in the stems and attach the line.

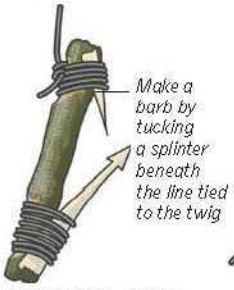

WOOD OR BONE
Carve a splinter of bone into an arrow point or take a sliver of hard wood, like oak, and tie to a twig.

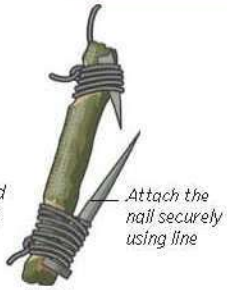

NAIL
Cut a notch at one end of a small piece of hard wood. Place the head of a nail in the notch and tie it in place.

SAFETY PIN
Remove the safety clasp. Bend the sharp point of the pin round to form a hook. Attach line to the wire loop.

REMOVING A FISHHOOK

You should never try to remove a fishhook that is deeply embedded in flesh. Cut the fishing line as close to the hook as possible, then put a pad around the exposed part of the hook and secure with a bandage. Seek medical help as soon as possible and monitor for signs of infection. However, in a survival situation, you can try to remove the hook by doing the following:

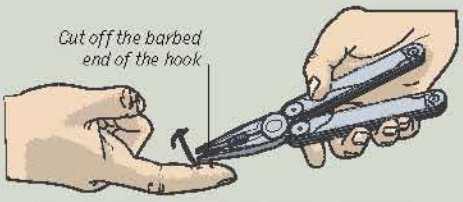

1 If the barb is visible, cut it off with a pair of pliers. If you can't see the barb, firmly and quickly push the hook in further until the barb emerges.

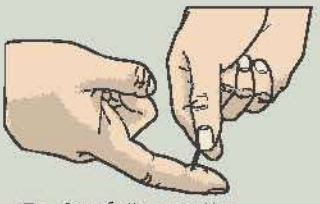

2 Carefully ease the hook out by its eye. Clean the wound then cover it with a bandage.

TYPES OF BAIT

Live bait includes creatures such as:
- Worms and slugs
- Grasshoppers, crickets, and beetles
- Maggots and caterpillars
- Frogs (including large tadpoles)
- Small fish—to catch larger fish

Examples of inanimate bait include:
- Meat, guts, and reproductive organs recovered from animals
- Nuts and small fruits
- Bread, cheese, and pasta

FEATHER LURE

Make a hook, then attach it to a line. Tie a brightly colored feather just above the hook. Slowly move the lure across the water's surface to attract the fish.

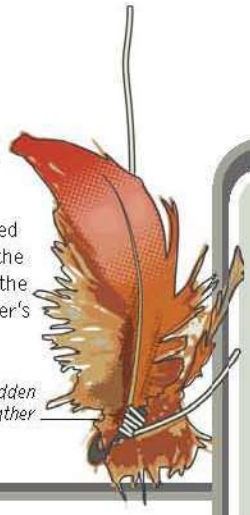

FLOATS AND SINKERS

A float keeps a baited hook at the best depth to attract the species of fish you're trying to catch. A sinker (weight) placed below the hook also helps to hold it in position. Your survival kit should include a few split-shot weights but, if it doesn't, you can tie small stones to the line instead. If the float bobs below the water's surface, you may have hooked a fish.

MAKING A FLOAT

You can use any natural material that floats, such as a piece of bark or a rose hip (see below). If you find a bird's flight feather, you can make a float by trimming the feather until you are left with just the hollow quill, then folding it in half and tying the two ends together.

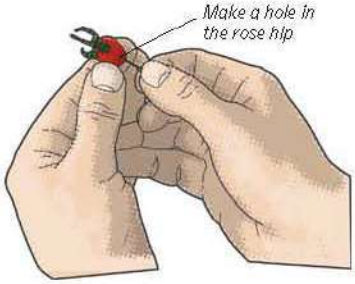

1 Using a sharp pin, a piece of wire, or a long thorn, pierce through the center of a rose hip.

2 Thread fishing line through the hole. If you don't have line, use a length of fine cordage.

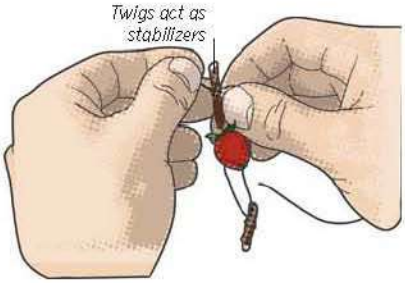

3 Tie small twigs above and below the rose hip. These will act as stabilizers and will also prevent the float from slipping up or down the line.

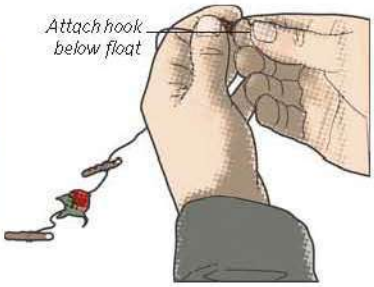

4 Attach the hook (in this example, a modified safety pin) to the line at the required depth below the float. If you have one, tie a sinker below the hook.

LINE FISHING

The more hooks you can get in the water, the greater your chances of catching a fish. All the methods shown below are passive—you simply set them and leave them to work for you.

SELF-STRIKING LINE

If a fish bites at the bait, the trigger will be jerked out of the catch and the bent rod will fly up. The line will be pulled tight, embedding the hook in the fish's mouth.

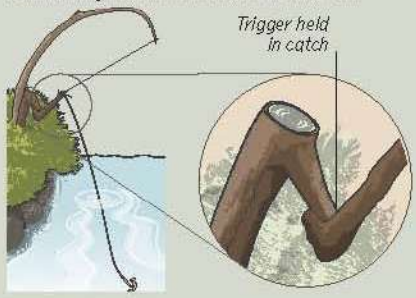

ICE FISHING

First check that the ice is at least 2 in (5 cm) thick and can bear your weight, then cut a hole about 1 ft (30 cm) in diameter over the deepest part of the lake.

1 Tie a hooked and baited line to a stick with a flag at the other end and a central crosspiece.

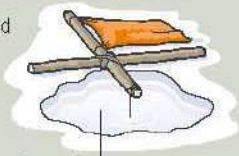

2 If a fish takes the bait, the crosspiece will be jerked over the hole and the flagpole pulled upright.

NIGHT LINES

To increase your chances of making a catch, you need to attract fish that live at all depths. Tie a rock to one end of a line and attach hooks with live bait at intervals along it. Tie the other end to a post at the edge of the bank and throw the line in the water, leaving it there overnight.

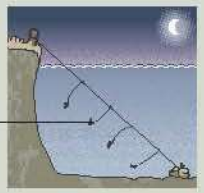

NETS, TRAPS, AND HARPOONS

You can leave a hooked line in the water, but unless you regularly check it any fish you catch may be eaten by bigger fish. In a survival situation, nets and traps are more convenient methods of fishing because once set you can leave them and they work for you all the time.

WARNING!
Gill nets and traps are illegal fishing methods in many parts of the world, so use them only in an emergency.

MAKING A DIPPING NET

Fish that are too small to hook or harpoon may still be large enough to catch with a dipping net. Such fish are usually found at the edges of streams and lakes, and around rocks in pools. You can make a simple dipping net if you can split a branch or find a forked sapling and have a mosquito net—or a spare item of clothing, such as a sleeveless vest, T-shirt, or stockings.

1 Make two small nicks in the hem of the mosquito net, vest, or stockings, then thread the forks of the sapling through them.

2 Cut another nick in the other side of the hem, where the forks meet, and pull them through. Tie the ends together with cordage, then push them back inside the hem.

3 Tie off the net (or the vest above the armholes and neck), then either cut off the excess material or invert the net. This will prevent extra drag when you use the net.

MAKING A GILL NET

Once it has been set in a river, a gill net is a highly effective way of catching fish, whether they are swimming up- or downstream. However, a gill net should only be set for short periods of time as fish of all sizes can be entangled or injured. Make the lines out of paracord or natural cordage (see pp. 136–39).

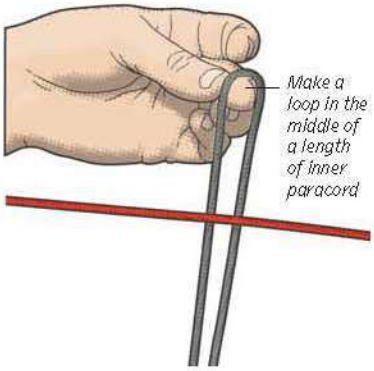

1 Take a section of the outer sheath of some paracord and suspend it between two trees. Loop the inner paracord behind this line.

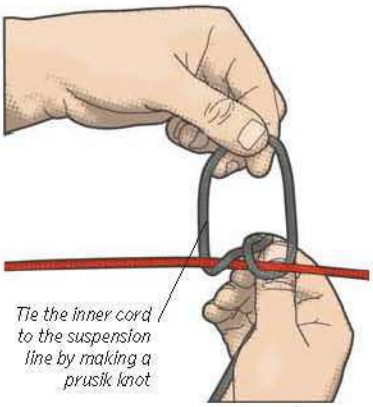

2 Holding the two ends of the inner cord in your other hand, pass them around the suspended line and through the loop.

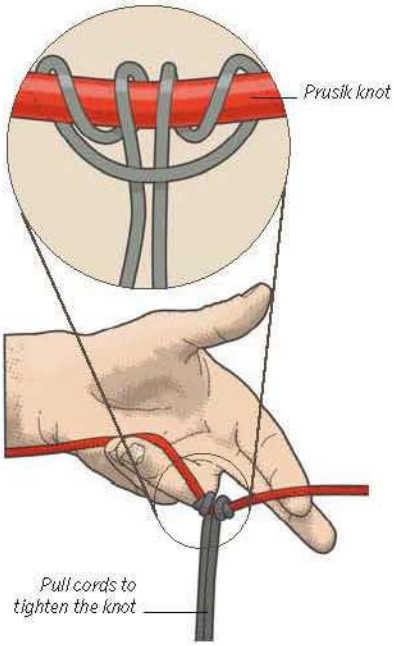

3 Pull the knot tight. Repeat the process every 1½ in (4 cm) along the suspended line for the required width of the net.

MAKING A HARPOON

Using a harpoon to spear fish takes time, patience, and a certain level of skill. Keep the point of your harpoon in the water to avoid splashing and scaring off the fish. Strike quickly when you see a potential catch, aiming just ahead of it. A spiked harpoon is the simplest to make.

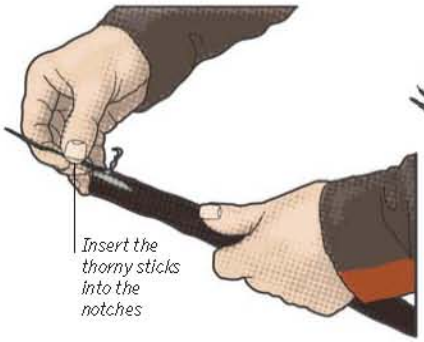

1 Gather a few short, thorny sticks to form the barbs of the harpoon. Cut a long, thick, straight branch and make notches around one end.

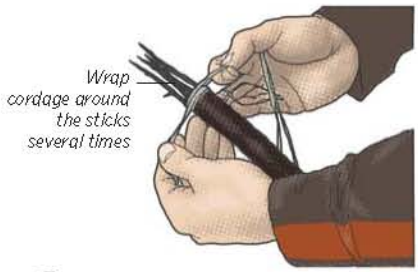

2 Bind the thorny sticks tightly to the shaft with cordage. They will need to be able to support some of the weight of the fish so wrap them securely.

3 When using the harpoon, take care not to damage the barbs by driving them against rocks or onto the riverbed. Use a dipping net, if you have one, to land the fish.

4 Make the mesh by tying a cord from one pair of cords to a cord from an adjacent pair, using a simple overhand knot (see p. 143).
- Continue alternating along and down the line. Tie stones to the ends of the net to hold it down once it is in the water.

5 To set the gill net, hang the suspension line between trees or poles on either side of the river, submerging the net in the water to a depth of about 6 in (15 cm).
- The net can also be used to catch birds (see p. 226).

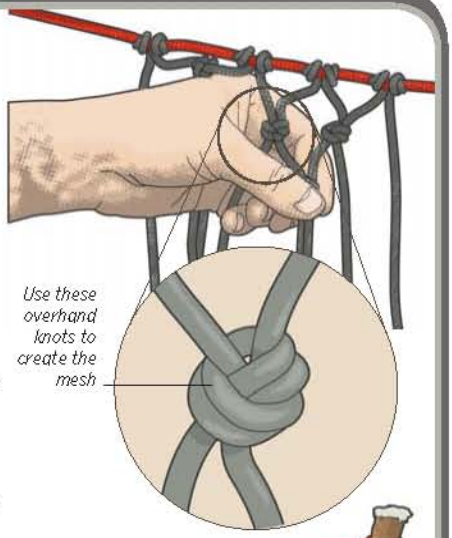

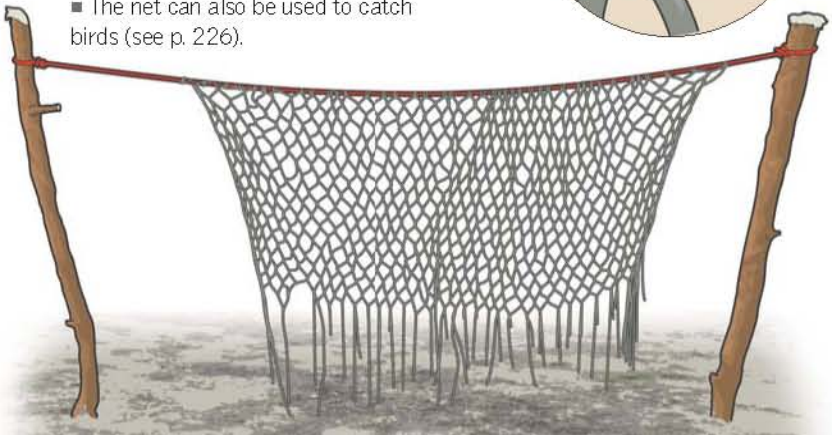

TRAPS

Fish traps can be made from a variety of materials. You can use rocks to build a wall or you can make a basket from sticks and cordage. You can even use man-made materials, such as bottles. If you only succeed in catching very small fish you can still use them as bait.

BOTTLE TRAP

Also known as a "minnow trap," this method uses a large plastic bottle to catch small fish. Cut the bottle in two just below the shoulder. Invert the neck and insert it into the cut end, then tie the two pieces together. Make holes in the plastic with a hot pin so the trap will sink. Bait the trap, then place it in a stream. Check it regularly to remove any fish caught and replace the bait.

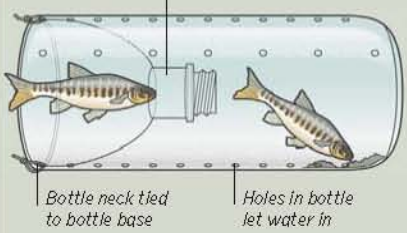

SINGLE-WALL TRAP

If your camp is near a tidal creek, you can build a curved wall of large rocks out from the bank. Pick the location at high tide and build a low wall at low tide. Fish may be trapped in the pool created between the wall and the bank as the tide recedes.

PREPARING FISH

ALL FRESHWATER FISH are edible but they must be cooked before eating as they are often infested with parasites and harbor bacteria. Saltwater fish are safer to eat raw, if necessary, but will taste better cooked. Never eat a fish that has pale gills, sunken eyes, flabby skin or flesh, or that smells bad.

WARNING!
When handling fish be careful not to accidentally touch your eyes. The slime may contain bacteria that can cause a painful inflammation of the eye covering (conjunctivitis). Make sure you wash your hands clean of all slime. If you have any cuts on your hands, cover them before touching the fish.

FISH HYGIENE

As soon as you have landed a fish, it should be killed and gutted, then cooked and eaten as quickly as possible, particularly in a warm climate. It doesn't take long for fish to go bad because their slimy skin provides a breeding site for flies and bacteria. In a cold climate, you can delay filleting for up to 12 hours, which will make the job easier.

FILLETING A FISH

Filleting removes the parts of a fish that might quickly go bad, while leaving as much flesh as possible. With some species, you may find it easier to remove the bones after the fish has been cooked. Boil the bones and the head to make a nutritious stock. This must be done immediately and the stock should be kept in a cool place and drunk within a few hours.

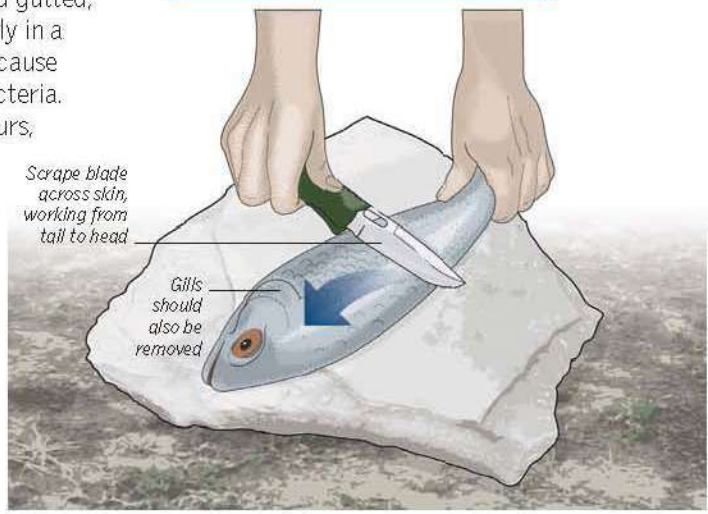

1 Once you have killed the fish, cut its throat to bleed it and cut out the gills. Wipe the slime off its skin to make it less slippery.
- Most fish don't need to be skinned—in fact, the skin is nutritious—but eels and catfish do (see panel, opposite).

2 Fish can be cooked with their scales on but, if you have time, it is best to remove them, particularly if they are large, as they can be a choking hazard.
- Hold the fish by the tail and scrape off the scales, holding the blade away from you and moving toward the head.

4 Pull out the internal organs then spread the fish open to check that you have removed everything. Keep the roe to eat (it lies down the side of the fish).
- Wash the fish thoroughly, inside and out.

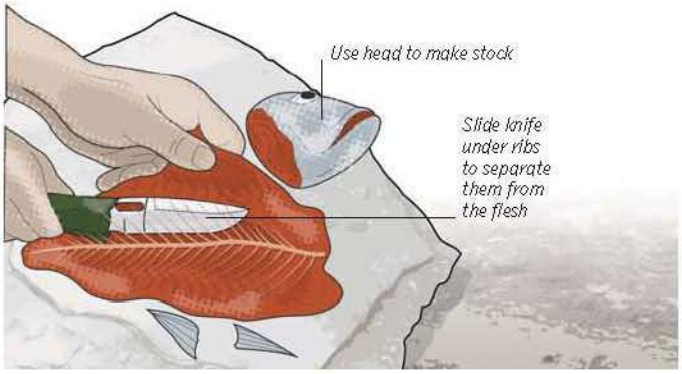

5 Cut off the head, tail, and fins. Open out the body and slide the knife under the ribs to separate them from the body, working toward the head. Fish oils can make the knife slippery, so you can use your thumbs instead.
- Repeat for the ribs on the other side of the backbone.

SMALL FISH

Any fish that is less than 6 in (15 cm) long doesn't need to be filleted. After gutting, they can be fried, grilled, or roasted whole. Keeping the heads and tails on helps to prevent the fish from falling to pieces.

SKEWERING FISH

Small fish, such as perch, can be grilled over an ember fire on a skewer to make a simple, yet tasty and nutritious meal. Impale the fishes on a green stick and hold them close to the embers. They won't take long to cook.

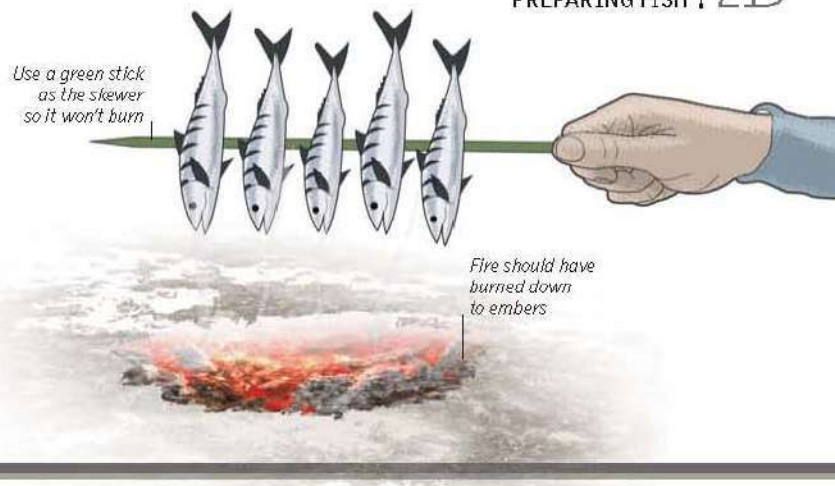

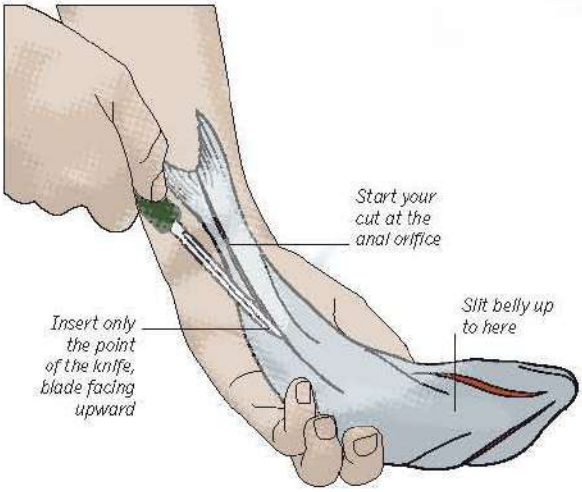

3 Holding the fish with its tail toward you, insert the point of a sharp knife, blade up, into the anal orifice and slit the fish open along its belly to its throat. This prevents you from puncturing the internal organs.

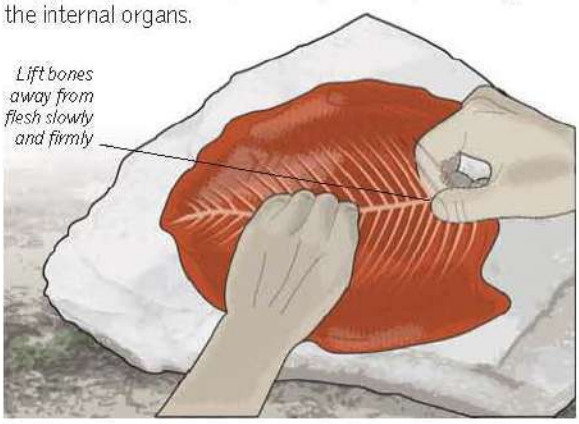

6 Separate the top and bottom of the backbone from the flesh with the tip of your knife. Carefully pull the backbone and ribs away from the flesh in one piece.
- If you have any tweezers or needle-nose pliers, use them to remove any small bones that remain.

SKINNING FISH

Eels and catfish are tasty but have to be skinned and gutted before cooking. To gut the fish, either use the filleting method shown opposite or Step 3 below. Catfish have a cartilage skeleton and can simply be cut crosswise into steaks.

1 Pass a stake through the gills of the fish and support on two strong uprights.
- Cut around the skin below the head with a sharp knife.
- Cut around the fins.

2 Separate the top of the skin from the flesh then peel the skin downward.
- You will need to use both hands to get a good grip, and a firm, steady action. If the fish is very large, you may need to slit the skin.

3 Remove the fish from the stake and break its backbone. When you pull the head off, the guts will come away with it.
- Remove tail and fins.

TRUE-LIFE ACCOUNT

EXTREME SURVIVAL—ADRIFT AT SEA

USEFUL EQUIPMENT

- Liferaft, lifejacket, survival suit
- Water catchment devices
- Solar still/reverse-osmosis pump
- Emergency Locator Beacon
- Marine VHF radio
- First-aid kit, sunscreen, sunglasses
- Flashlight and batteries
- Map, compass, GPS
- Survival tin, bushcraft knife
- Cell/satellite phone
- Poncho/bivy sack

IN 1972, A FAMILY OF FIVE SURVIVED for 38 days adrift in the Pacific Ocean. The Robertson family were 18 months into a round-the-world sailing trip when their 43 ft (13 m) wooden schooner *Lucette* sank. They survived thanks to a combination of good seamanship, improvisation, and good fortune.

The family set off from Falmouth, England, on January 27, 1971—father Dougal at the helm, wife Lynn, and children Anne (who disembarked in the Bahamas), Douglas, and twins Neil and Sandy as crew—and safely navigated the Atlantic and Caribbean. On June 15, 1972, however, about 200 miles (320 km) west of the Galapagos, disaster struck—a pod of killer whales charged the boat, splintering the hull and holing her irreparably.

"A POD OF KILLER WHALES CHARGED THE BOAT...HOLING HER IRREPARABLY"

With *Lucette* sinking rapidly and no time to radio a distress call, the crew launched their inflatable liferaft and roped it to the boat's 10 ft (3 m) solid-hull dinghy, the *Ednamair*, which they used as a towboat after improvising a sail. Their supplies amounted to 2 gallons (10 liters) of water, a bag of onions, oranges and lemons, vitamin-fortified bread, glucose, four fishing hooks, a fishing line, a first-aid kit, a kitchen knife, and eight signal flares. They sailed north toward the Doldrums to find rain, which they collected with the use of a tarpaulin, and caught fish, eating some of the meat raw and drying the rest in the sun to be stored as rations.

After 16 days, the liferaft had deteriorated to such an extent that the family was forced to transfer to the dinghy, using the remnants of the raft as a canopy to provide shelter and aid in the collection of rainwater. They used the wind and currents to sail northeast toward Central America, and built up sufficient rations over the next three weeks to provide energy for the extra exertion of rowing toward the coast. Luckily for them, there was no need—their ordeal came to an end on July 23, when a Japanese fishing boat spotted a signal flare and picked them up.

WHAT TO DO

ARE YOU IN DANGER? +

NO YES

- If you are in a group, try to help any others who are in danger
- Prepare for the possibility of abandoning ship, and try to steer closer to either land or known shipping lanes
- Delegate responsibilities
- Ensure lifejackets and liferafts are ready for deployment

Get yourself out of it:
Sinking vessel—You need water, location aids, and protection from drowning and the elements
Animals—Try not to splash in the water as this will attract sharks
Injury—Stabilize condition and apply first aid

ASSESS YOUR SITUATION
See pages 234–35

DOES ANYONE KNOW YOU WILL BE MISSING OR WHERE YOU ARE?

NO YES

If no one knows you are missing or where you are, you will need to notify people of your plight by any means at your disposal

If you are missed, a rescue party will almost certainly be dispatched to find you

DO YOU HAVE ANY MEANS OF COMMUNICATION?

NO YES

You are faced with surviving for an indefinite period—until you are located or you find help

If you have a cell or satellite phone, let someone know your predicament. If your situation is serious enough to be worthy of emergency rescue, and you have a Personal Locator Beacon (PLB), you should consider this option

CAN YOU SURVIVE WHERE YOU ARE? *

NO YES

Abandon ship in a controlled manner and deploy all the liferafts. Always try to enter the liferaft dry. Use the rescue line to reach people in distress. Follow the "immediate actions" instructions printed inside the liferaft

Address the Principles of Survival: Protection, Location, Water, Food

YOU WILL HAVE TO MOVE **

DO

- Deploy the drogue—this will lessen drifting
- Inventory all food, water, and equipment and start rationing
- Prepare water procurement devices, such as solar stills and reverse osmosis pumps
- Protect yourself from the elements—sun, wind, and salt spray
- If you have no liferaft, then huddle together in a group with children in the center. If alone, adopt the H.E.L.P. position
- Improvise flotation aids from anything that can float or hold trapped air. Good examples are plastic bags, plastic bottles, and knotted clothing

DON'T

- Cut the painter to the vessel until you know it will sink, as the vessel is what people will be looking for
- Drink seawater under any circumstances—this will only increase your rate of dehydration
- Eat unless you have sufficient water to digest the food—fish is high in protein and requires plenty of water to digest

YOU SHOULD STAY **

DON'T

- Neglect to take anti-seasickness tablets. Vomiting will dehydrate you and the effects are very demotivating
- Forget to wear your survival equipment and ensure everyone knows how to operate it
- Try to second-guess the rescuers when they arrive. Do exactly as they say, they know what they are doing

DO

- Ensure that the rescue services are kept aware of your situation and updated with relevant information as the situation dictates
- Prepare to abandon ship: pack lifejackets and suitable clothing
- Ensure you know how to operate your survival equipment and that you have basic aids to location such as a whistle, flashlight, or plastic water bottle on your person
- Clear the deck of loose objects and be prepared to drop sail should rescue by helicopter be attempted
- Make an inventory and ration supplies

+ Never abandon a vessel unless you have to—searchers will be trying to locate it. Use the vessel's equipment while you can, and before you leave collect essential items to aid your survival.
* If you cannot survive where you are, but you also cannot move owing to injury or other factors, you must do everything you can to attract rescue.
** If your situation changes (for instance, you are "moving" to find help, and you find a suitable location in which you can stay and survive) consult the alternative "Do" and "Don'ts."

TRAPPING ANIMALS

ALTHOUGH YOU SHOULD be prepared to take down sitting prey if you get the opportunity, trapping small animals is easier than hunting them—it requires less skill and energy and leaves you free to carry on with other tasks. One of the simplest traps is a snare.

MAKING A SNARE

Stainless steel snares of various strengths can be bought ready-made, with a running loop, or eye, at one end and a securing loop at the other. However, most survival kits contain a length of single-strand brass wire that can be used to make a snare.

1. Decide on the strength of snare you want (see panel, below) and double or quadruple the wire accordingly.
- Pass the strands around a stick, place it on the ground, then loop the loose ends around a second stick.
- Rotate the second stick until the strands have entwined, forming a single, thick wire. Remove the sticks.

ALTERNATIVE MATERIALS

Guitar strings make great snares as they already have a "perfect end." Natural cordage can also be used.

STRENGTHS OF SNARE

The single-strand wire found in most survival kits isn't strong enough to hold most animals you would want to catch. You'll need to double or quadruple the strands by winding them together to increase their strength.

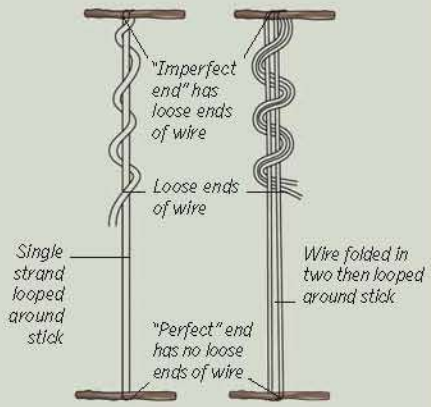

TWO STRANDS
A two-strand snare will be sufficient to bear the weight of a squirrel-sized animal.

FOUR STRANDS
A four-strand snare will be strong enough to hold onto an animal the size of a rabbit.

KILLING A RABBIT

Unless you're an expert, breaking a rabbit's neck by holding its head and pulling on its legs can be difficult and could result in you dislocating its hips, causing it more pain. The best method is to strike the rabbit on the back of the neck with a solid stick, thus avoiding having to pick it up and risk being bitten or scratched. The eyes will glass over immediately when the animal is dead.

LOOK AND LISTEN

To be successful at hunting and trapping, you need to decide which types of animal you are going to try to catch. To do that, you need to find out what animals are in the vicinity and where exactly they are to be found. Prey animals will use their keen senses to avoid being caught—so you will have to use your senses to find them. Look for any signs of animals, both on the ground and in the trees. And use your ears—you may be able to hear an animal even when you can't see it.

SIGNS TO LOOK OUT FOR:

- Runs, trails, and tracks
- Droppings
- Chewed or rubbed vegetation
- Feeding and watering areas
- Lairs, dens, and resting sites

SQUIRREL POLE

Make several two-strand snares with nooses 3 in (8 cm) in diameter and stake-out cords 8 in (30 cm) in length. Place them around a long pole, with the lowest 2 ft (60 cm) off the ground. If you snare a squirrel, leave it, as others will come to investigate.

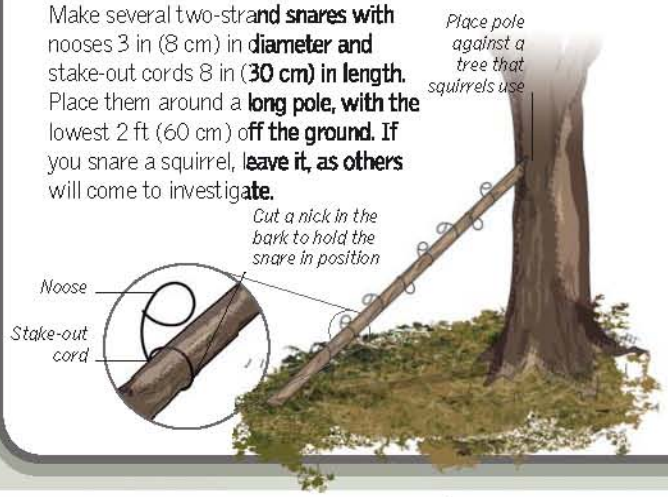

2 To make the noose, pass the imperfect end, or securing loop, through the perfect end, or running loop. The running loop has no loose ends of wire so it won't snag in an animal's fur and prevent the noose from tightening as it struggles.

- Use a single prusik knot (see p. 210) to tie a length of cord to the securing loop.

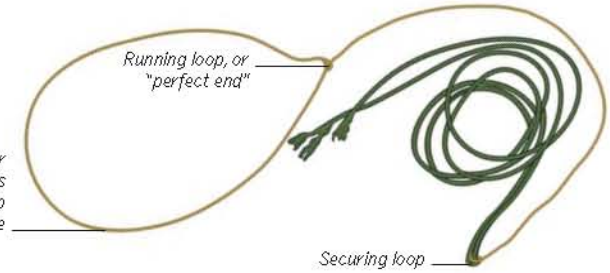

3 Before setting the snare, either bury it in the ground for a few hours or pass it over a flame for a few seconds to remove any human scent and dull its surface.

- Don't waste snares; always set them on runs you know the animals are using and set as many as you can.
- Use sticks to make a support from which to suspend the snare, or bend saplings into an arch.

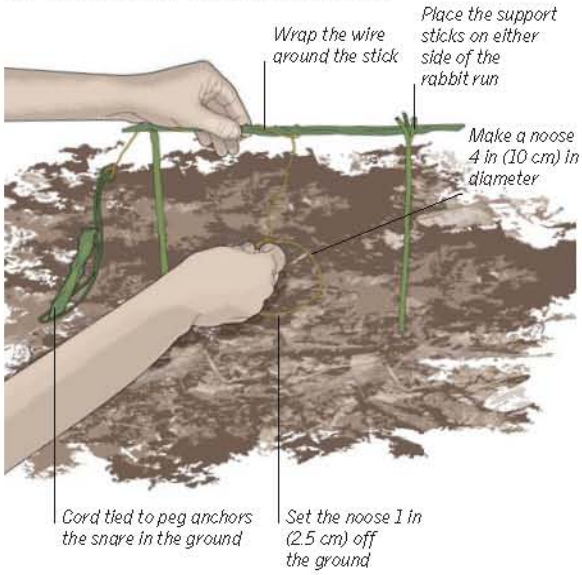

4 Avoid disturbing or treading on the run, as the disturbance or your scent could alert the animal. Camouflage the frame with vegetation, which you can also use to funnel the animal toward the snare.

- Use natural vegetation, such as holly or other prickly shrubs, to form a funnel on either side of the snare.
- The funnel helps to ensure that the animal has no option but to pass through the snare.

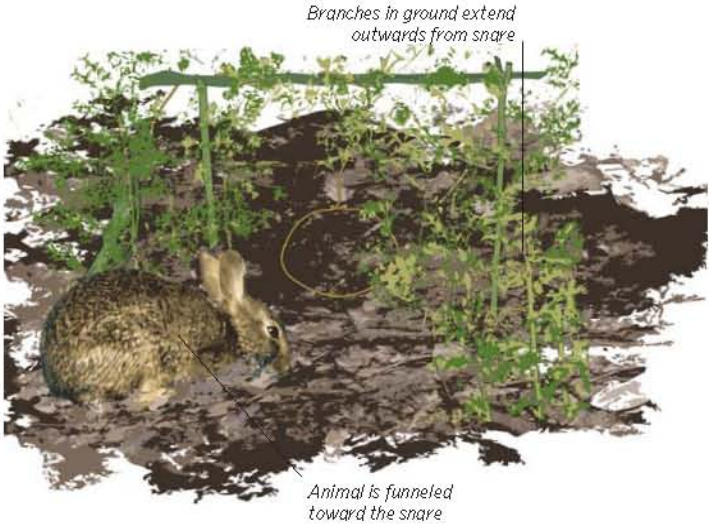

WEAPONS AND TECHNIQUES

A spear is the simplest weapon to make and use, but if you have the time and materials you can make more complicated weapons, although they require more skill to use. Learning other basic techniques, such as catching insects, is also invaluable when hunting.

THROWING STAR

With a throwing star you're four times more likely to hit prey than with a spear, and, if you don't succeed in wounding the animal with one of the sharp ends, the weight of the weapon will at least stun it. Keep it close at hand and ready to use, should the opportunity arise.

1 Find two sticks 2 in (5 cm) thick and 18 in (45 cm) long. Make a square-cut joint in the center of each, then sharpen each end.

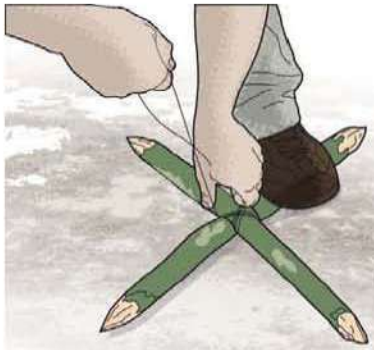

2 Join the sticks together by overlapping the two square-cut joints and lashing them to each other with paracord or natural cordage.

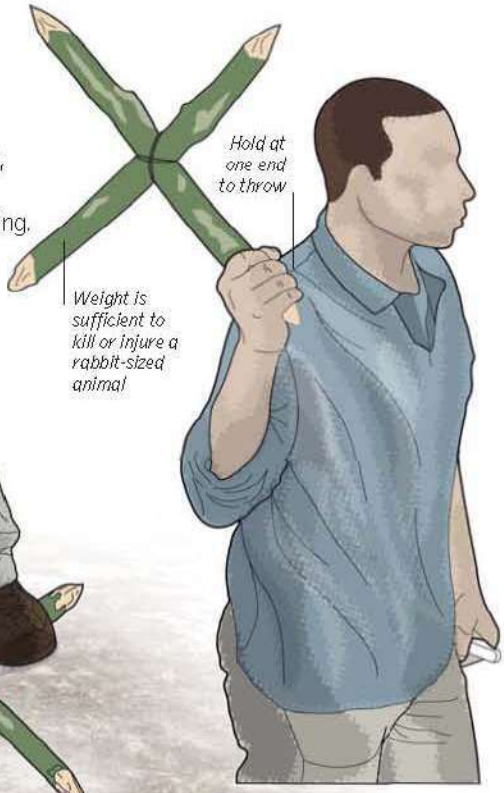

3 You can exert more force if you're throwing from a standing position. Aim for the animal's legs.

RODENT SKEWER

A forked spear can be used to catch small mammals in their burrows. Thrust the pointed end into the hole. When you can feel the animal, twist the stick until you have snagged its fur, then carefully pull it out.

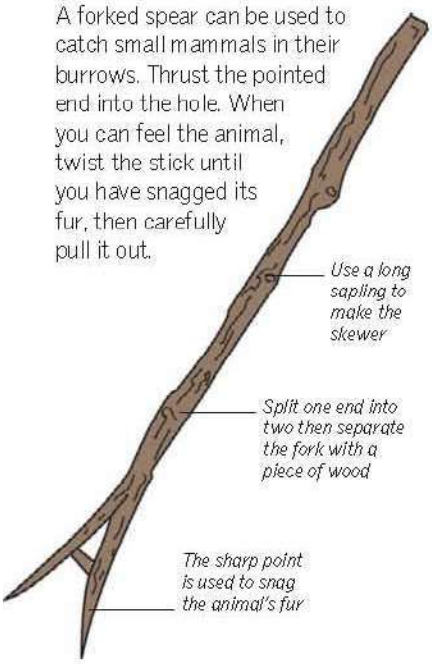

SLINGSHOT

A slingshot can be used to kill small animals. Cut a strong forked branch and make a notch in each fork. Take the rubber tubing from your survival kit (see pp. 60–61) and thread it through a piece of leather or plastic. Tie the ends of the tubing to the notches. Place a pebble in the pouch, stretch back the sling, and take aim.

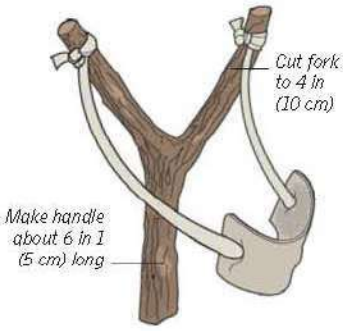

NOOSE STICK

You can use a noose stick to lasso lizards, slow-moving mammals, or roosting birds, jerking the pole to tighten the noose around the animal's neck. Use a wire or cordage snare (see pp. 216–17). When stalking prey, move slowly and very quietly.

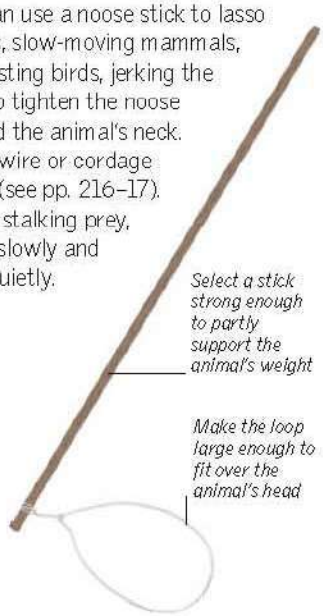

EASY PICKINGS

Ants are social insects, and most species are aggressive in defense of their nest. They have a stinging bite and some species will also squirt formic acid at their attacker. However, they make a nutritious meal if gathered with care. In summer months in northern temperate regions, you can eat the larvae of wood ants. And, if you use the technique shown below, you can get the ants to do all the work of collecting the larvae together for you. You'll have to vandalize the nest, but don't remove all of it.

ANT CUISINE

Ant larvae are best fried—they taste like shrimp. Adult ants also make a tasty snack—in Bogotá, Colombia, movie theaters sell roasted leaf-cutter ant abdomens instead of popcorn.

1 Place a tarpaulin in a sunny patch next to the nest. Scoop the nest material, ants, and larvae into the middle of the tarpaulin.

2 Lay some sticks on the tarpaulin, close to the edge, then fold over the sides to create shade. The ants will carry the larvae into the shade.

3 After a while, throw back the folded sides of the tarpaulin and scoop up the larvae, which look like fat grains of white rice.

CATCHING INSECTS

Many flying insects can be eaten, but catching them takes a bit of ingenuity. Crawling insects, such as ants (see above) and termites, can provide a good meal if you can collect them in sufficient quantities.

TERMITE FISHING

Termites are found in the tropics and subtropics. Some species live in vast numbers inside mounds, which they build out of mud and saliva. If a foreign object breaches the walls, the termites will attack it with their powerful jaws.

1 Cut a long, thin, straight stick then peel it until smooth. Push it slowly into the termite hill.

2 Remove the stick and scrape off the attached termites into a container ready for frying or roasting.

NIGHT FLIERS

Nocturnal flying insects, such as moths, are attracted to light. Stretch a white sheet between two branches and peg the bottom taut above a bowl of water. Hang a flashlight behind the sheet. Insects will fly into the sheet and fall into the water.

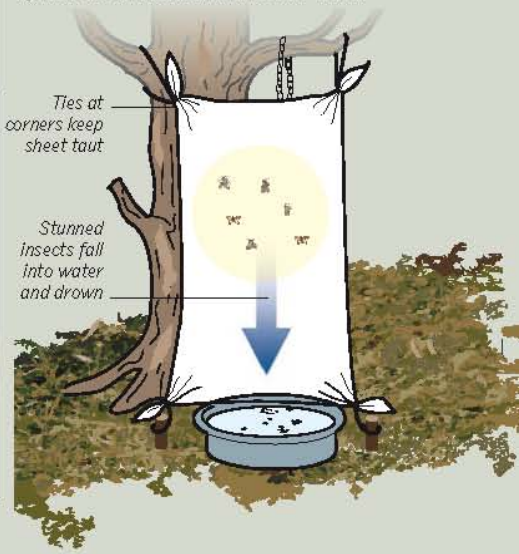

PREPARING SMALL MAMMALS

PARASITES SUCH AS FLEAS and lice will leave the body once a warm-blooded animal is dead so, if your circumstances permit, leave the carcass to cool before preparing it for cooking. In a hot climate, place the body in the shade. However, don't let it get too cold, as skin is more easily removed when a body is warm.

WARNING!
Rabbits, hares, and rodents may be infected with a bacteria called tularemia, which can be fatal to humans. Do not touch these animals with your bare hands if you have a scratch on your skin. If you have no gloves, cover your hands in soap lather before handling the animal, and wash your hands when you have finished. The germ is destroyed by heat, so cook the meat well. Myxamatosis, a viral disease that affects rabbits' mucous glands, is not harmful to humans.

RABBIT-SIZED MAMMALS

The method shown here applies to all furred mammals approximately the same size as a rabbit. Before gutting the animal, remove any urine by holding the body by the forelegs and progressively squeezing down from the chest toward the bowels, making sure that you direct the spray away from your own body. You'll need to remove any scent glands, which are usually inside the forelegs or around the anus, in case the musk taints the meat (take particular care when handling a skunk).

GUTTING

No part of a mammal carcass should be wasted. If your camp is near water, keep the guts for fish bait. The offal should be removed carefully and eaten if healthy—discard any organ that's pale or spotty. Liver is rich in essential vitamins and minerals and needs little cooking. It should be eaten as fresh as possible (removing the gall bladder from the middle first). Kidneys, which in most mammals are surrounded by fat, are also a valuable source of nutrition.

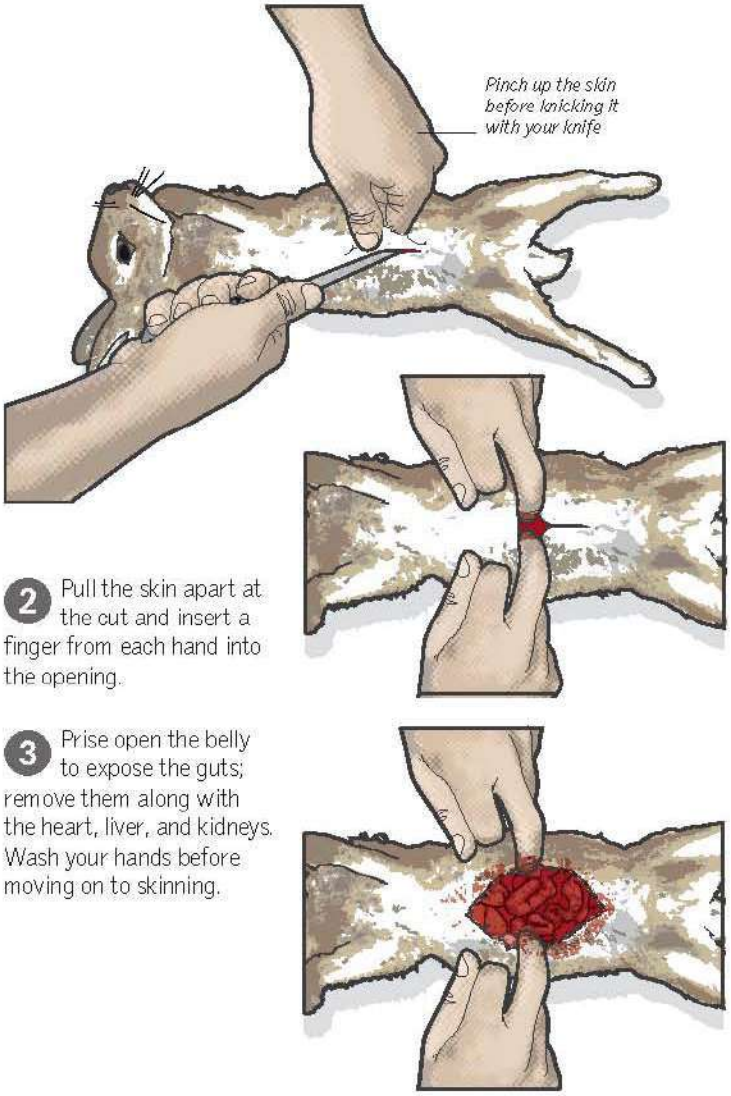

1. Place the animal on its back on clean ground, such as a bed of pine boughs, with its head pointing toward you. Cut a small hole in its belly with the point of your knife, taking care not to pierce the guts.

2. Pull the skin apart at the cut and insert a finger from each hand into the opening.

3. Prise open the belly to expose the guts; remove them along with the heart, liver, and kidneys. Wash your hands before moving on to skinning.

WARNING!
Rabbits make an easy meal, but their flesh lacks fat and vitamins so, if your diet consists entirely of rabbit, your body will use its own vitamins and minerals to digest the meat. These are then lost from your body in your feces and, if not replaced, you'll get diarrhea and so become hungry and weak. Eating more rabbit makes the condition worse and you will eventually die of starvation. You must supplement your diet with some vegetation and fatty food.

SKINNING

If you wish to keep the skin for use later (to make a pair of mittens, for example), you should remove it as carefully as possible. You'll also need to cure it; stretch the skin as tight as possible and leave it in the sun or hang it close to the heat of a fire to dry. Rubbing wood ash into the skin will help to speed up the process.

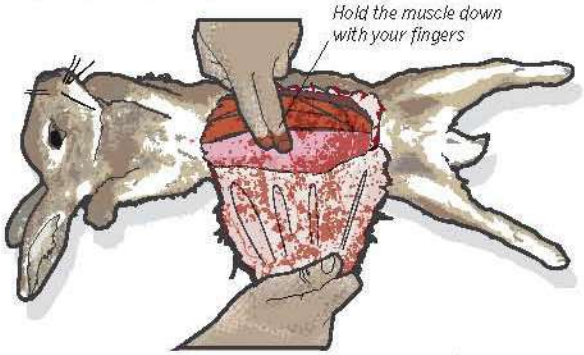

1 Beginning at the belly, separate the skin from the muscle surrounding the gut cavity. You'll find that it pulls away quite easily. When you have reached the back on one side, repeat the process on the other.

GLOVE SKINNING

The preferred survival method of skinning and gutting a rabbit-sized mammal is to suspend it by its rear legs from a strong branch as this keeps the animal off the dirty ground.

- If you have no knife to hand, snap off the lower part of a foreleg and use the sharp edge of the broken bone to cut the skin.
- Gut the animal, then cut the skin around all four paws and between the rear legs.
- Tug the skin down toward the head.
- If you wish, pull the skin over the head. Otherwise just cut the head off.

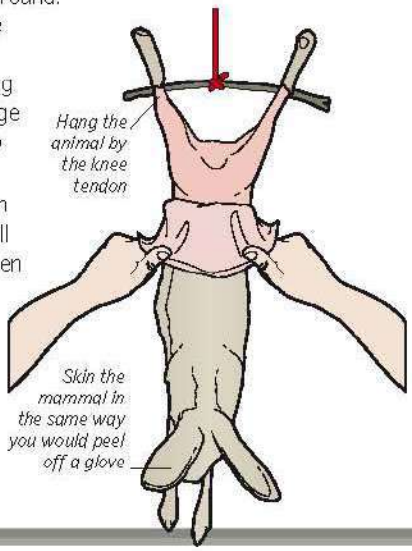

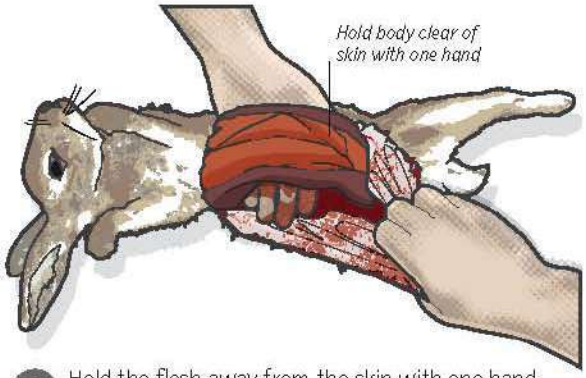

2 Hold the flesh away from the skin with one hand and pull the skin over the rear legs, one by one. The hindquarters are now free of skin and fur, except for the tail, which should be cut off and the scent glands removed.

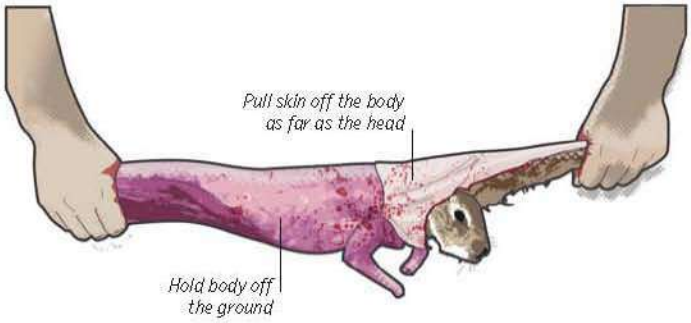

3 Hold the animal by the rear legs and pull the skin forward, easing out both forelegs. Pull the skin over the neck and then cut off the head. Wash the body to remove any traces of fur.

HOW TO SKIN A SQUIRREL

Squirrel flesh is tender and tasty. Once you have skinned the squirrel, you can simply skewer it on a spit and roast it whole over a fire. Alternatively, you can cut the meat into pieces and make a stew. A large squirrel will feed two people.

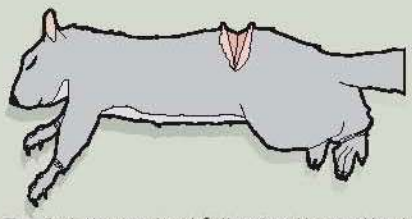

1 Gut the squirrel following the method shown opposite. Cut the skin around the paws, then cut through the skin across the middle of its back.

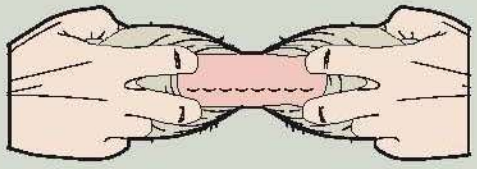

2 Insert two fingers under the skin on either side of the cut and pull the two pieces apart and off the body. Wash off any traces of fur.

PREPARING LARGE MAMMALS

ALL LARGE MAMMALS are edible, but you must not eat the livers of polar bears or bearded seals as they contain toxic levels of vitamin A. Wild game is sometimes infected with roundworm larvae and, if you eat the meat raw or undercooked, it will infect you with parasites, so always cook all wild game thoroughly.

WARNING!
Approach all game with caution, as the animal may still be alive and most wounded animals are extremely dangerous. To check, touch its eye with the tip of a long stick (or the muzzle of a gun). Even if the animal is unconscious, it will blink if it's alive. Always skin, gut, and butcher game at the site of the kill. You don't want the smell of fresh blood to attract predators or scavengers to your camp.

PREPARING THE KILL

Lifting large animals, such as deer, uses energy so, if you're alone, prepare the carcass on the ground. However, if you're in a group and circumstances permit, it's best to suspend the animal by its hocks (the tarsal joints of its hind legs) from a branch, as it will bleed better and will be easier to skin and gut. If you can't find a suitable tree from which to hang it, build a frame.

BLEEDING

Bleeding is essential to preserve the meat and stops it from tasting too gamy. It also helps to cool the carcass. Slit the animal's throat from ear to ear and let the blood drain out. Blood is rich in vitamins and minerals, including salt, so, if you're able to, collect it in a container to use later in a stew. Cover the container to protect the blood from flies and to keep it cool.

SKINNING

If you wish to preserve the hide, it's better to take the skin off before gutting. Cut the belly skin from throat to tail, cutting around the genitals. Cut along each leg from above the foot to the belly. Pull the hide off the carcass, severing connective tissue as necessary. If you're skinning on the ground and you don't want to keep the hide, you can use it to protect the meat—remove it fully only after jointing is complete.

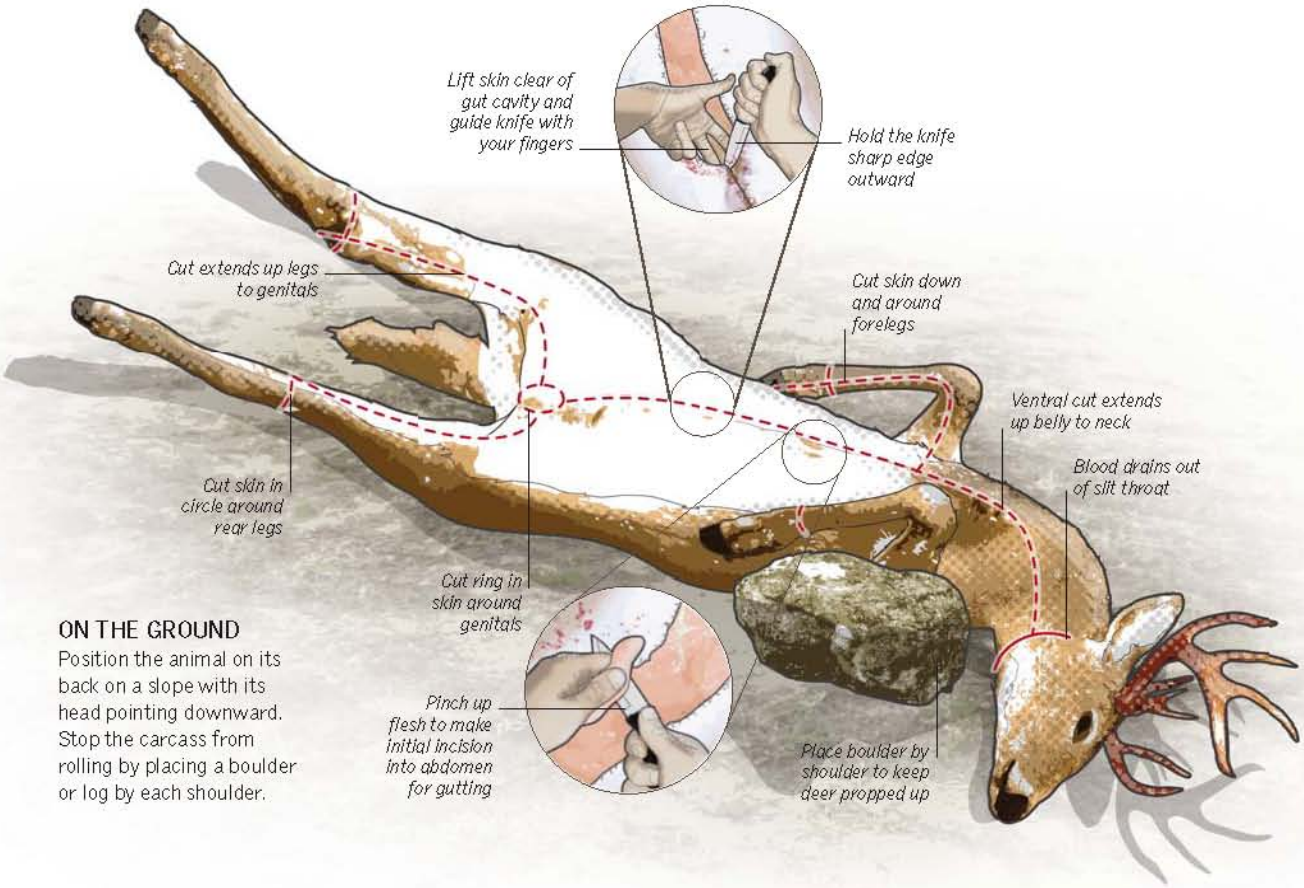

ON THE GROUND
Position the animal on its back on a slope with its head pointing downward. Stop the carcass from rolling by placing a boulder or log by each shoulder.

GUTTING

The method shown here is for gutting a carcass on its back. When finished, inspect the heart, liver, and kidneys for signs of worms or other parasites and, if the organs are healthy, keep them to eat. If the liver is spotted, a sign of disease, discard all internal organs and boil the meat.

1 To avoid piercing the internal organs, pinch up the abdomen near the breastbone and make an incision big enough to insert two fingers.

- Use your fingers to guide the knife, cutting toward, then around, the anus with the sharp blade-edge upward.

2 Using a saw, if you have one, cut through the outer muscle and the breastbone to open up the chest cavity,

3 Using the knife, slice through the diaphragm muscle, which separates the chest cavity from the gut cavity, getting as close to the spine as you can.

- Remove the liver, taking care not to cut the gall bladder in the center.

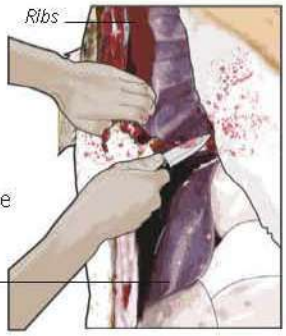

4 Reach up through the chest cavity and cut the windpipe and esophagus. Hold them with one hand and pull out all the internal organs as one unit.

- Check that the anus is clear of feces, pushing a hand through if you can't see daylight.

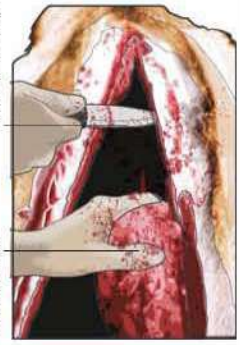

KEEP THE GUTS

If you're able to carry them, take the guts, reproductive organs, and glands back to your camp in a sealed container. You can use them as bait in traplines or for fishing. Keep the fat that surrounds the intestines for cooking.

JOINTING

Large game should be cut into manageable pieces that can be carried back to camp. Boning the meat helps to reduce the weight. Bag or wrap the meat as you remove it to keep it clean and free of flies, and keep it in the shade in hot climates. How a carcass is divided partly depends on the species, but the hindquarters contain the steaks and the best cuts.

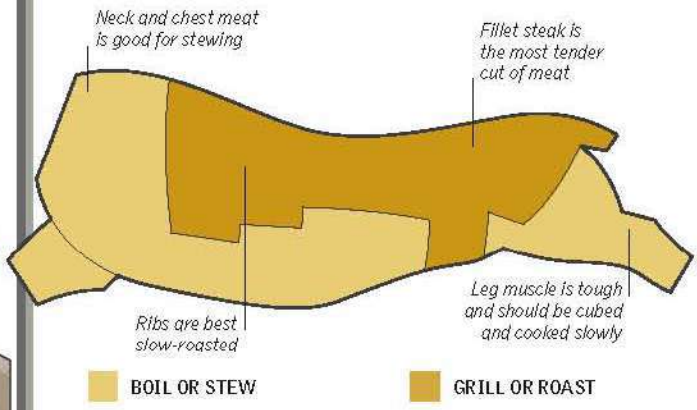

USES FOR HIDES

Keep the skin if you can. When dried, it's light and can be used as a blanket, or it can be made into an item of clothing. Hide is one of the best materials for lashes. Sinew also makes good cordage (see p. 140).

TANNING THE HIDE

The best way to clean a skin is to stretch it on a frame and then carefully scrape off the fat and any remaining scraps of flesh, using a flint or a piece of bone. Every mammal, with the exception of buffalo, has enough brain matter for use in tanning its own hide. Mash the brains well in warm water, then apply the mixture to the hairless side of the skin and leave it to dry for 24 hours.

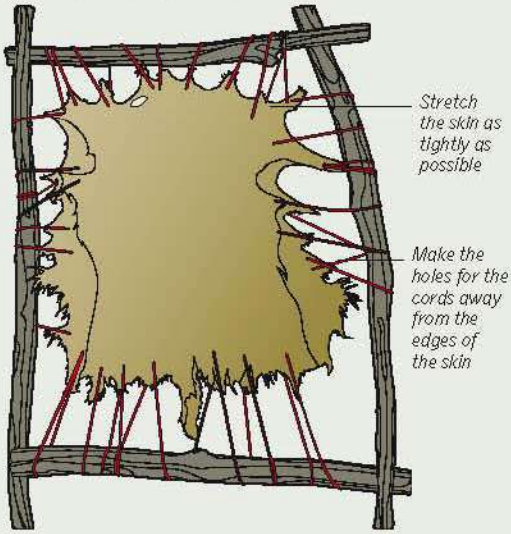

PREPARING OTHER ANIMALS

ALL TYPES OF ANIMAL should be considered when searching for your next meal. Invertebrates are far more plentiful and widespread than other animals, and take less energy to gather and prepare. Reptiles and amphibians can also provide a vital source of nutrition.

EATING FOR SURVIVAL

The thought of eating grubs or grasshoppers might not appeal, but experimentation is vital in a survival situation—don't forget that in some cultures such creatures are highly prized, if prepared properly. However, some animals are less palatable than others, in which case you should chop them finely and add to a stew.

SLUGS, SNAILS, AND WORMS

Slugs should be avoided as they often feed on poisonous fungi. To prepare snails, either starve them or feed them safe food such as wild garlic for 24 hours to purge their guts before cooking. Either boil them in water or bake them in their shells in hot embers until the juices bubble over. Avoid all marine snails and any terrestrial snail with a brightly colored shell as they may be toxic. All earthworms are edible. Place them in salty water until they are purged, then boil, though they can be eaten raw if necessary.

EDIBLE INSECTS

Insects are mostly protein and make good emergency food. Avoid any hairy or brightly colored insects or those that emit a foul smell (including larvae). As a general rule, avoid adults that sting or bite, but ants and honey bees can be eaten if collected carefully. Most insects, except those with a hard carapace, like beetles, can be eaten raw.

HOPPING INSECTS

Grasshoppers and crickets have large leg muscles, and most are quite tasty when cooked (avoid brightly colored ones). Remove the antennas, wings, and leg spurs, and roast to kill any parasites.

WITCHETTY GRUBS

Witchetty grubs are a highly nutritious Australian bush delicacy. They can be eaten raw but, if roasted quickly in hot ashes, they taste of scrambled eggs. Treat palm grubs in the same way.

Large, white grubs high in protein and calcium

AMPHIBIANS AND REPTILES

All frogs in the *Rana* genus (see pp. 292–93) are edible. Skin and then boil or roast them. The best, most meaty part is the hind legs. Don't handle or eat toads or brightly colored tropical frogs as many have highly toxic skin secretions. Lizards and snakes are a good source of protein. They must be skinned and gutted, then either roasted on a stick if small, or cut into small pieces and boiled. All snake flesh is edible—and tastes like chicken—apart from the head, which must be removed.

1 Hold the dead snake firmly behind the head and cut it off about 6 in (15 cm) down the body.

2 Slit the belly with a sharp knife and remove the entrails and other organs. Either throw them away or keep as bait.

3 Peel the skin back, then grasp the body in one hand and the skin in the other and pull it off.

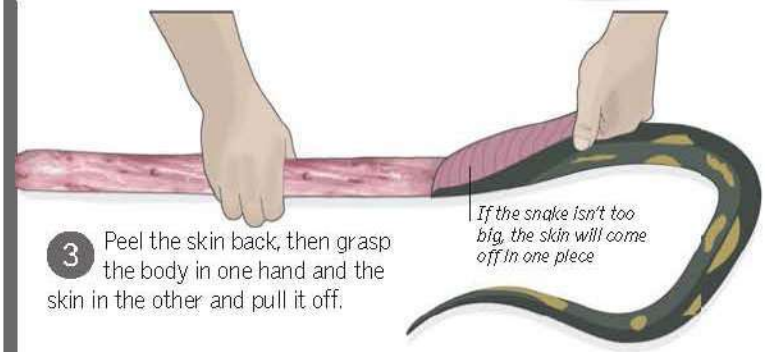

PREPARING A CRAB

Avoid their strong claws (tie them up if you can). To kill a crab, plunge it in boiling water or stab it through the eye socket or the orifice under the shell flap on the body. Cook the crab for about 15 minutes.

1 Place the cooked crab on its back, then twist off its legs and the claws. They all contain edible meat, so crack them open with a rock.

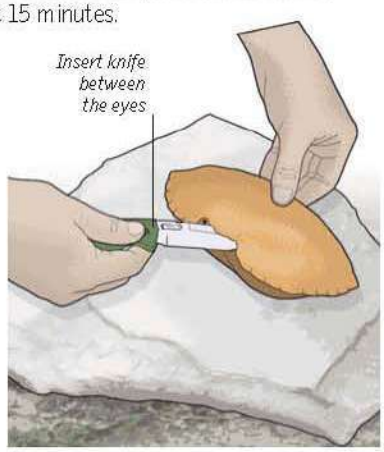

2 Open the shell by inserting the point of your knife between the two halves and twisting it. Lift back the lower shell.

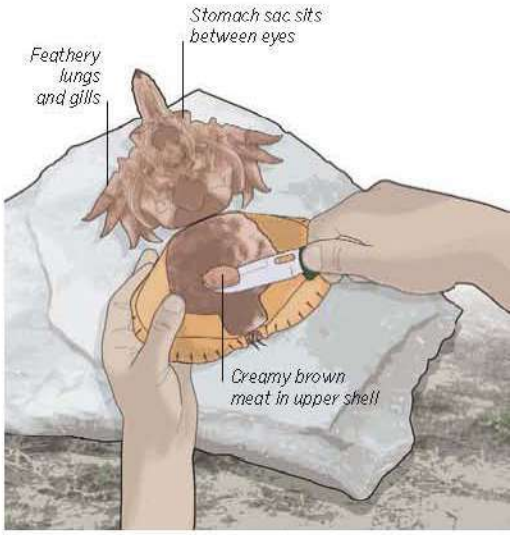

3 Remove and discard any green matter and the lungs, gills, and stomach as they're all poisonous. Scoop the meat out of the shell.

SHELLFISH

Shellfish can be found in streams and on lake and sea shores. Don't collect marine shellfish that are not covered at high tide or are near any source of pollution. Shellfish must be alive when you collect them—bivalves, such as mussels, will close their shell if you tap them, and univalves, such as limpets, will cling tightly to their rock—and you must cook and eat them immediately.

WARNING!
Tropical species of mussel should be avoided in the summer as they may be poisonous. The black mussel, which is found on Arctic and sub-Arctic coasts, may be poisonous at any time of the year. The symptoms of mussel poisoning are acute nausea, vomiting, diarrhea, abdominal pain, and fever.

HOW TO SHUCK AN OYSTER

Oysters are rich in vitamins and minerals. To eat an oyster raw, first open its shell by inserting a blunt knife into the hinge at the thicker, more pointed end of the oyster and twisting. However, in a survival situation, play safe and boil oysters in their shells for five minutes after their shells open. Do not eat any whose shells do not open during cooking.

SHRIMPS, CRABS, AND LOBSTERS

Small crustaceans, such as shrimps and crawfish, should be cooked in boiling water for five minutes. Larger crustaceans, such as some crabs and lobsters, should be boiled for up to 20 minutes. A tasty alternative to boiling shrimps or prawns is to roast them on skewers made of green wood, placed in the embers of a fire.

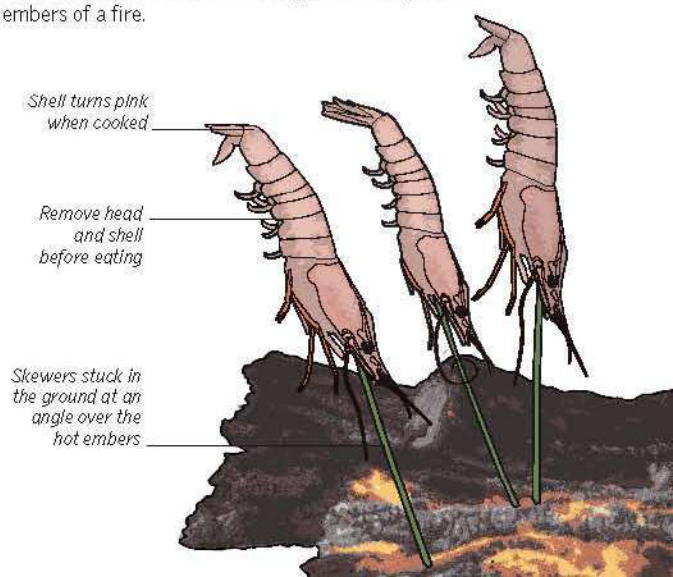

CATCHING BIRDS

BIRDS CAN BE TRAPPED in a variety of ways—in snares, nets, tunnels, or cages—and they can be hunted with noose sticks or any projectile weapon. The trick is to observe their behavior and to match your technique—and bait—to the bird. You need to discover their regular roosts, feeding spots, and flight paths.

WARNING!
Eggs are easy to collect, particularly from ground-nesters, although gulls, for example, will attack you if you try. However, taking eggs from nests is illegal in many countries and should only be done if absolutely necessary.

FIGURE-4 CAGE TRAP

This trap enables you to catch ground-feeding birds (or rabbit-sized mammals) without killing or wounding them, so you can choose when you want to kill them and eat the meat fresh. The "Figure 4" is the trigger that's used to drop the cage over the animal.

MAKING THE TRIGGER

To construct the Figure 4 trigger you need to make three notched sticks (bait stick, upright stick, and release stick) and then link them together as shown below.

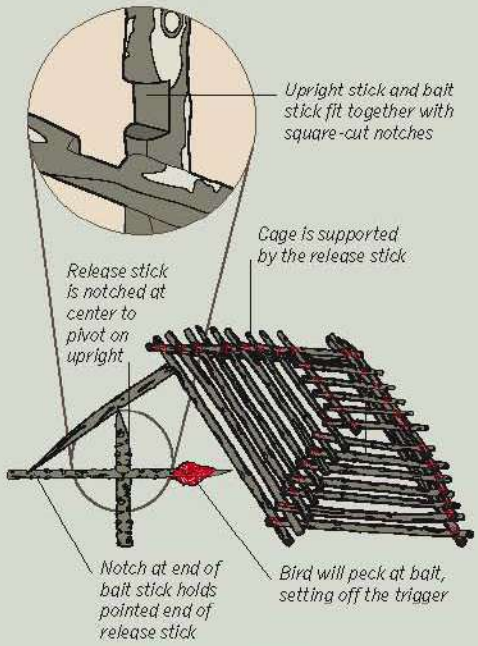

MAKING THE CAGE

Create a pyramid of sticks by placing progressively shorter sticks across longer sticks and lashing them together at the ends. Balance the cage over the bait.

ON THE GROUND

Gamebirds, such as quail, and inquisitive scavengers, such as crows, can be lured into traps if you use the right bait. Migratory waterfowl, such as ducks and geese, molt in the late summer, which means they are easier to catch as they can't fly away.

WALK-IN TUNNEL TRAP

A tunnel trap can be used to catch gamebirds, which have stiff feathers that lie in one direction and bend only with difficulty. As a bird tries to retreat from the trap, its feathers will become wedged in the tunnel walls.

- Dig a near-horizontal, funnel-shaped tunnel close to the ground.
- Lay a trail of bait (seed or berries) leading to the rear of the tunnel.
- As it eats the bait, the bird will move deeper into the tunnel and will be unable to back out.

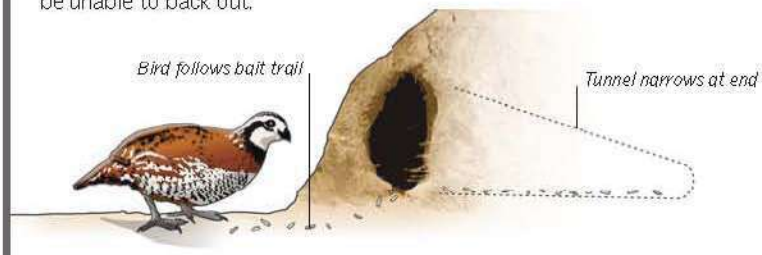

ON THE WING

A fine net stretched between two trees across a flight path is an effective method of catching birds in flight. Throwing a bola requires more skill, and works by entangling the bird in the spinning ropes.

MAKING A BOLA

To make a bola, use an overhand knot (see p. 143) to tie three 3-ft (1-m) lengths of cord together about 3 in (8 cm) from one end. Find three rocks weighing about 7 oz (200 g), wrap each one in a piece of cloth, and tie it up with the free end of a cord.

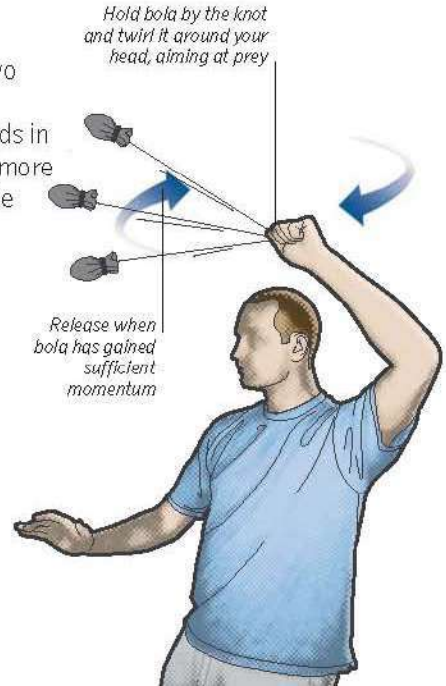

ON THE NEST

Birds are creatures of habit and will usually roost in the same place every night, such as in the branches of a tree as protection against ground predators. They rarely move once it's dark. Many birds are more accessible when breeding as, once you've located the nest, you'll always know where to find them. However, you should only take nesting birds in a genuine survival situation.

MAKING A SNARE STICK

A snare stick is used to trap birds roosting in trees. Remove a bird once it has been snared, as its squawking and fluttering will alert other birds to the danger and scare them off.

- Make several single-strand snares (see pp. 216–27) with a loop diameter of about 1–2 in (2.5–5 cm), depending on the size of the bird.
- Place them close together along a stick, making notches in the stick to hold them in position with the loops uppermost.
- Tie the stick on top of a branch where you have seen birds roosting.

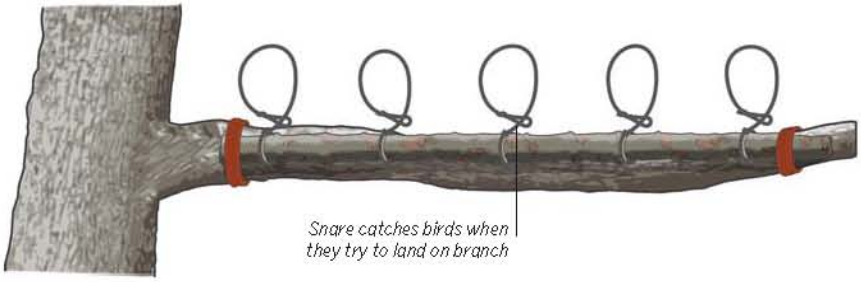

PHEASANT POLE

Pheasants roost in trees at night, usually returning to the same branch, and they won't leave once it's dark. To snare a sleeping pheasant, raise the pole up in front of it. The bird will wake and peck at the foil. As its head goes through the loop, pull the cordage and the snare will close around the bird's neck.

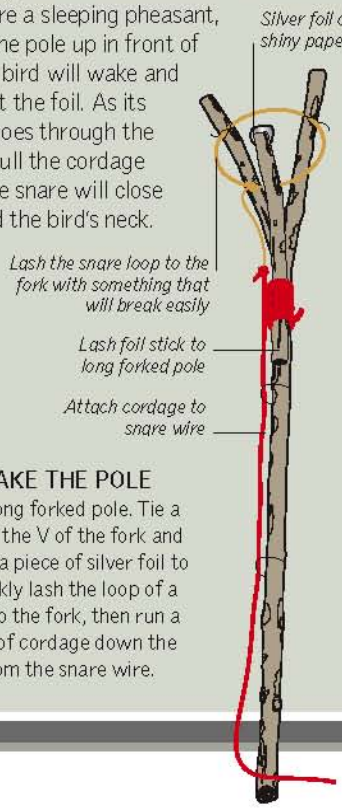

TO MAKE THE POLE

Cut a long forked pole. Tie a stick in the V of the fork and attach a piece of silver foil to it. Weakly lash the loop of a snare to the fork, then run a length of cordage down the pole from the snare wire.

BOUGH FENCE TRAP

This method is used to trap ducks, which often go ashore during the day to sun themselves, or at night to roost. They tend to favor small islands as protection against predators.

- Cut several saplings and make double-stranded snares (see p. 216).
- Place the boughs a few feet from the water's edge, and bend them into a series of overlapping hoops to form a fence around the island.
- **If you only have a few snares, place them over the ducks' inward and outward tracks. Placing logs between the snares will then force the birds to pass through the hoops in order to get onto the island.**
- **Suspend a snare from the center of each bough arch.**

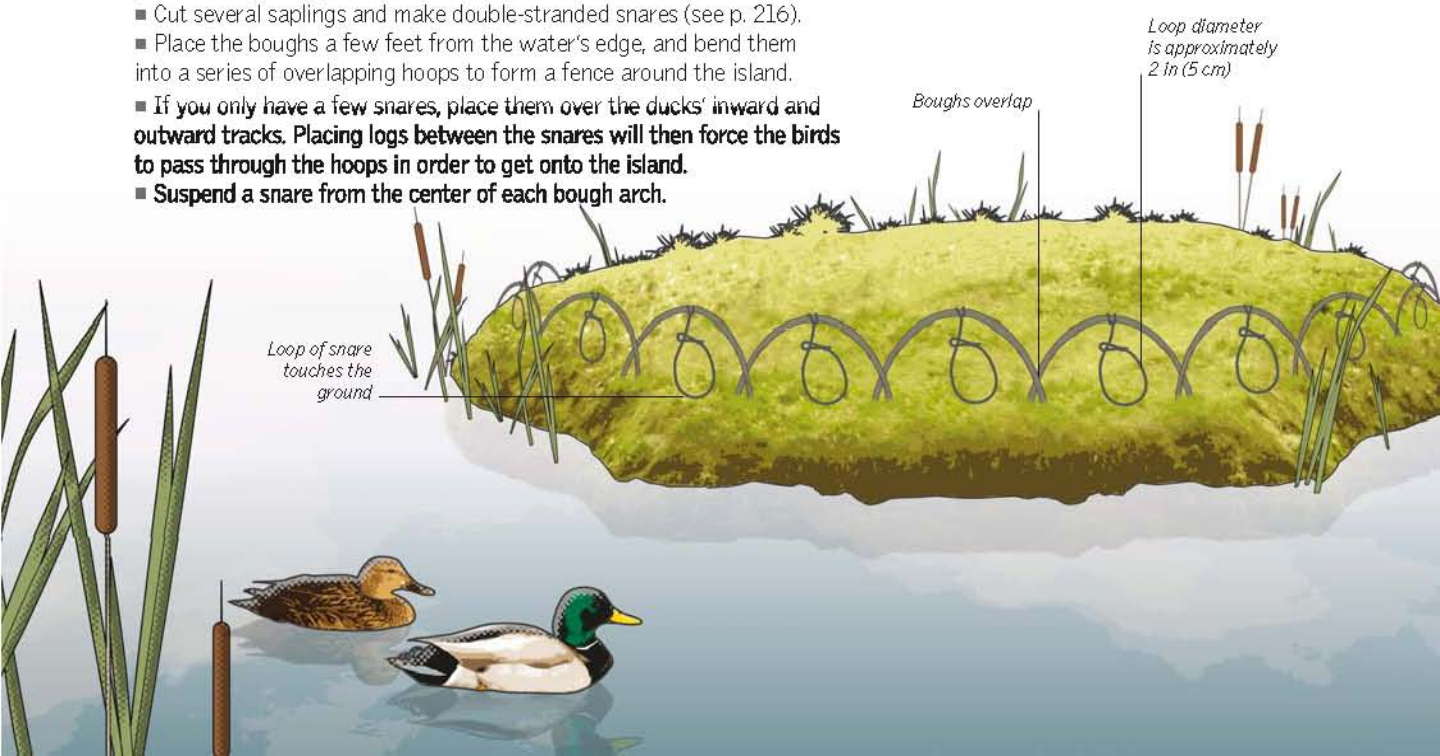

PREPARING BIRDS

GAMEBIRDS AND WATERFOWL are the most commonly eaten wild birds, but all birds are edible if prepared correctly. Skinning a bird is the quickest way to get at the meat, but this removes the nutritional value of the skin, so always pluck a bird if you have the time to do so.

GETTING STARTED

If the bird hasn't already been killed by the method of capture you'll need to dispatch it—for example, by stretching its neck and cutting its throat (for the best way to kill gamebirds, see below). Whichever method you use to kill the bird, you must bleed it before plucking, but make sure you pluck it while the body is still warm.

WARNING!
Handle scavengers and carrion eaters, such as crows, buzzards, and vultures, as little as possible, because they are more likely to be infested with lice and ticks, and are prone to infection.
- The meat of such birds must be boiled for at least 30 minutes to kill any infectious organisms. This will also tenderize the meat.
- Always wash your hands after handling any bird.

KILLING GAMEBIRDS

To quickly and humanely dispatch a gamebird, such as a pheasant, fold its wings into the body and hold it under your arm. Cover its head with your jacket to calm it down (this also reduces the chances of you being injured by its claws or beak). Still holding the bird firmly, take a 2 ft (60 cm) long, thick stick and place it on the ground. Put the bird's head underneath the stick, place your feet on the stick either side of the head, and pull the bird up sharply by the legs—the bird's head will come off.

POACHER'S METHOD
The poacher's way to quickly get the meat from a dead pheasant that hasn't been decapitated is to place it face down on the ground with its head toward you, put a foot on each wing, take hold of the legs, and pull up sharply. This motion will tear the legs and breast meat away from the rest of the bird. Give it a sharp flick to detach the guts. In a survival situation, this is a good way to get meat off the bird if you don't have a knife or razor.

PLUCKING A BIRD

It's not essential, but scalding a bird in a bucket of hot (not boiling) water for a couple of minutes usually helps to loosen the feathers; the exceptions are waterfowl and seabirds—their feathers will tighten instead. Be careful not to overscald the bird or the skin will start to cook.

1. Lay the bird on its back, take a few breast feathers between your finger and thumb, and begin to pluck.
- Tug sharply, pulling the feathers away from the direction in which they lie.

2. Don't try to pluck too many feathers at once or you'll tear the skin. Work your way steadily around the front and back until you've plucked the whole body.
- Keep the feathers (unless the bird is a scavenger, see box, above) to use for tinder, insulation, or fishing lures.

OPTIONAL EXTRAS

Keep the crop (esophagus) contents of a gamebird—you can use the seeds and berries as bait to catch other birds. If the bird is female and has eggs in the oviduct, keep them and eat them. You should also eat the heart and the liver if they are in good condition; discard them if they look old or spotty, or show signs of parasites.

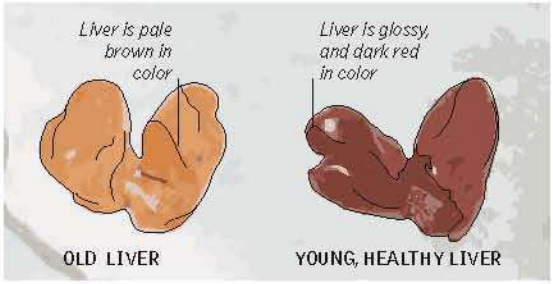

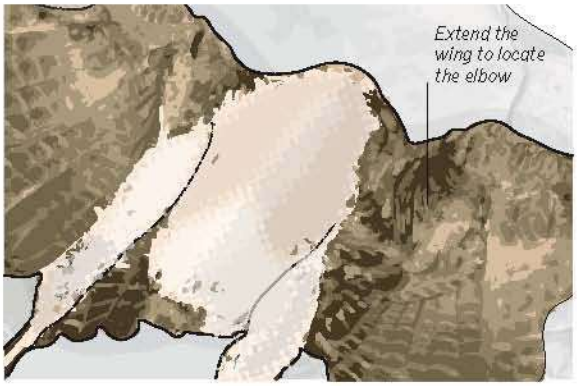

3 Pluck the legs in the same way, then turn to the wings. Locate the elbows and cut off the lower wings at this joint. Pluck the upper wings.

- If the bird's head is still attached, cut it off as close as possible to the body.

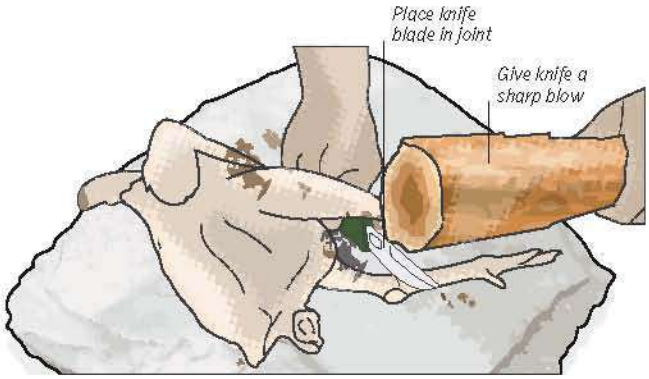

4 To remove the feet, find the ankle, insert the blade of a sharp knife into the joint, and press down hard.

- If you're dealing with a large bird, such as a turkey, hit the back of the knife with a thick stick for extra force.

DRAWING METHODS

The usual way to remove the offal from a bird (known as "drawing") is to take the crop (esophagus) out of the neck end and the rest of the offal out of the rear end. The survival method is to make a cut from the throat to the tail with a sharp knife, reach in, and pull out all the offal (see below). Should you find yourself without a knife, you can still butcher a bird by pulling the skin apart with your fingers and working your way up the breast to the neck. The offal can be removed by ripping open the skin over the belly.

1 Hold the bird in one hand, breast upward and with its rear end toward you. Insert the point of a sharp knife into the throat.

- Make a single incision down toward the tail, being careful to not pierce the guts.

2 Reach into the gut cavity and pull out the offal, taking care not to break any eggs. Keep the liver and heart if healthy (see above left).

- Wash the bird clean with cold water and wash your hands thoroughly before further handling.

IN AN EMERGENCY

IN AN EMERGENCY

You're only a survivor when you have been rescued. This means that you must be able to either get yourself out of the predicament you are in (self-rescue) or be rescued from that situation by others. Sometimes you will be able to choose whether to undertake self-rescue; at other times the decision will be out of your hands. There may be several reasons why you cannot achieve self-rescue: you could be utterly lost; local conditions such as flooding or bad weather could trap you; or injury to yourself or a group member could make movement impossible. In this situation, the onus is on you to attract attention. You must be able to make contact with rescuers, using location aids you have with you or those you can improvise. Never delay any form of rescue because of the embarrassment factor—the only important factor is the outcome, and I would always rather be embarrassed and alive than eventually found dead!

Preparation is key—informing people of your intentions and timeframes will at least have someone wondering why you are not back yet. Equally, taking the best

In this chapter YOU WILL DISCOVER...

- how to **recognize** the **enemies** of **survival**...
- why **LEOs** are so **important**...
- when to **set fire** to a **log cabin**...
- that **tinsel** isn't just for **Christmas**...
- what to **do** if a **hippo yawns**...
- that you **should go** with the **flow** in an **avalanche**...
- how to **inflate** your **pants**...

location aids for your environment and knowing how best to use them will increase the chances of you being found.

In many survival situations, a major decision will be whether you remain where you are or move to a location that offers a better chance of survival, rescue, or both. There are numerous factors that will dictate your best option but, in general, it's always best to stay where you are. It's all too easy to make a rash decision and attempt to walk out of a situation only to put yourself in even greater danger.

A location aid can make the difference between life and death. A comparatively recent invention is the multi-tasking Skystreme device.

SKYSTREME is a silver, inflatable foil kite, which weighs just 1½ oz (43 g) and packs down to a small and convenient size. It can perform four tasks in a survival situation:

Location aid The kite is orally inflated and its wedge shape means it can lift off from the ground unaided in 4 mph (6 kph) winds.

- The 165 ft (50 m) of line attached to the kite (this much cordage is invaluable in itself) allows it to rise above tree levels and reflect sunlight. It can be seen by the naked eye at a distance of 2 miles (3 km).
- At night, you can hang a small flashlight or lightstick underneath the kite to aid visibility.
- The metallic surface reflects radar and can be detected by aircraft at a distance of 10 miles (17 km). A UK Royal Marine on a polar crossing flew his Skystreme behind his sled so that support aircraft could find and track him.

Emergency first aid splint The kite can be inflated around a broken or sprained limb.

Thermal vest The kite can be inflated and put inside clothing to act as a body warmer.

Water storage Instead of being inflated with air, the kite can be used to carry and store water.

> "You are not a **survivor** until you have **been rescued**"

ASSESSING YOUR SITUATION

ONCE YOU'RE OUT of immediate danger, assess your situation and plan accordingly. At this initial stage, it's crucial to think clearly—the decisions you make now could mean the difference between life and death. In most cases, remaining where you are is the preferred option, but no two situations are ever the same: the circumstances, environment, conditions, and you—the individual—all have a major impact on what can and cannot be achieved.

THE STRATEGY FOR STAYING ALIVE

In an emergency situation, think of the four priorities of survival: protection, location, water, and food. Your situation will determine which is the most important. In most cases, as long as you're in no further danger from injury or the elements, you should focus your efforts on establishing a safe location and getting yourself rescued. The strategy for staying alive, known by the acronym SURVIVAL (see below), gives you a framework to use and helps you remember what you need to do to remain alive and get rescued.

THE RULE OF THREES

A common understanding among those concerned with survival is that various key time limits have a numerical value linked to the number three. Remembering the "rule of threes" can help you focus your decision-making, particularly if you are injured, at risk of further injury, or in immediate danger from the elements. In most cases:

- **Three seconds** is the psychological reaction time for making a decision.
- **Three minutes** is the length of time your brain can do without oxygen before it suffers irreparable damage.
- **Three hours** is the critical time you can survive unprotected in extreme climates.
- **Three days** is the approximate length of time you can live without water.
- **Three weeks** is the approximate length of time you can live without food.

SIZE UP YOUR SITUATION	USE ALL YOUR SENSES	REMEMBER WHERE YOU ARE	VANQUISH FEAR
• First, assess the particulars of your surroundings, physical condition, and equipment. • If you're in a group, you can share tasks and responsibilities, but remember that everyone will be affected by the consequences of your decisions.	• Most people react to a true survival situation through either training (they automatically do what they've been trained to do in a particular situation) or by instinct (they automatically do what their mind and body tell them to do).	• In any survival situation, it always helps to know where you are so that you can make the best decisions about what to do and where to go next.	• Fear and panic can be formidable enemies, so it's imperative that you have the knowledge and training to counteract them and prevent them from making your situation worse.
• Your surroundings: every environment has its own idiosyncrasies—hot and dry, hot and wet, cold, exposed, or enclosed. Determine what you need to do in order to adapt to that environment. • Your equipment: assess your equipment and consider how it can best be used in your particular situation. • Your physical condition: remove yourself and others from danger. Check for injuries and administer first aid as necessary. Remember that the trauma and stress of a survival situation may cause you to overlook, or subconsciously ignore, injuries that you yourself may have received.	• Whatever has happened, approach your predicament in a calm and rational manner. The situation requires careful thought and planning. • If you act in haste, you may overlook important factors, lose vital equipment, or simply make matters worse. The saying "Undue haste makes waste" is especially true in a survival situation. • Listen to your subconscious survival senses and gut feelings, and learn to act on them when they send you warning signals.	• Knowing your precise location can clarify whether rescuers are likely to find you or if you'll have to rescue yourself. You'll get a good idea of the obstacles you may face, whether you should remain where you are, and where best to position your aids to location. • If you have made an Emergency Plan of Action (see pp. 24–25), someone will know your approximate location, and when you're due back. • Channel your efforts into making sure that when rescuers are looking for you, your aids to location are in place.	• If uncontrolled, fear and panic can destroy your ability to make intelligent and rational decisions. They can cause you to react to your feelings and imagination rather than to your actual situation and your abilities. Moreover, they can incapacitate you and drain your energy, thus triggering other negative emotions. • If you're in a group, your responses can have a direct effect on others—positive responses are productive and can motivate, while negative responses can undermine confidence and morale.

THE ENEMIES OF SURVIVAL

In a survival situation, there are seven factors, known as "the enemies of survival," that can work against you. In many cases, you can deal with them by knowing what they are and understanding their effects. One way to memorize these is by using a mnemonic such as: "Be Prepared To Face These Hostile Factors".

BOREDOM AND LONELINESS

When boredom sets in, you become inactive and lose the ability to deal with your situation effectively, so you need to keep busy. Loneliness makes you overwhelmed by what you need to achieve, leading to a feeling of helplessness.

PAIN

If you're injured, don't ignore the pain. Attend to a minor injury, as it could grow into a major problem that could impair your ability to survive. A positive mental attitude coupled with keeping busy helps to distract the mind from pain.

THIRST

Thirst is not a good indicator of the body's need for water. Your body can be dehydrated before you feel thirsty. Stay ahead of dehydration, rather than have to deal with it. Prioritize your need for water in a way that's relevant to your environment.

FATIGUE

Tiredness leads to mistakes that, at best, cause frustration and, at worst, may result in injury or death. In a survival situation, it's unlikely that you'll replace the energy you use effectively, so everything becomes harder to achieve. Never underestimate the importance of quality rest to your physical and mental wellbeing.

TEMPERATURE

Temperature is a major factor in any survival situation and it will be affected by wind, rain, and humidity. You should dress to suit the environment you're in and be aware of the signs and symptoms of temperature-related injuries, such as dehydration, hypothermia, heat stress, and heat stroke (see pp. 272–73).

HUNGER

In a short-term survival situation (one to five days), procuring food is not a high priority. You can offset your reduced energy and stamina levels by drinking water and pacing yourself so you work within your limits. However, take every opportunity to procure food without expending energy.

FEAR

Fear is one of our body's greatest survival tools, as it can stimulate you, so that you're ready to act—however, it can also debilitate. Fear is good as long as you have control over it, and the key to controlling fear in a survival situation is knowledge.

IMPROVISE	VALUE LIVING AND LIFE	ACT LIKE THE LOCALS	LEARN BASIC SKILLS
▪ The true skill of a survivor is to understand what's required and improvise solutions to particular problems. Do you have the skills and knowledge to keep yourself alive and in a condition to be proactive in your own rescue?	▪ Some people without training and equipment have survived the most horrendous situations. In many cases, this was simply because they had the will to live and refused to give up!	▪ Whatever environment you're trying to survive in, you can be sure that the local or indigenous people and the local wildlife have developed ways of adapting to it in order to survive.	▪ Learning basic skills increases your chances of survival. Without training your prospects of survival are down to luck, which is never the best place to start. There is a saying: "Luck favors those who are best prepared."
▪ You may start out with all the right equipment, but it may get lost or broken, or simply wear out. Your ability to improvise may mean the difference between your continued struggle to survive in relative comfort, or absolute misery. ▪ Think laterally, like the climber who was stranded on the side of a mountain with no aids to location. He used the flash on his camera to signal his location to a rescue helicopter. Improvise and overcome!	▪ The stories of prisoners of war often reveal what kept them alive: religious beliefs, thoughts of family and friends, or a determination not to let the enemy win. While these alone may not always be enough, they're certainly a key factor in any survival situation. ▪ It helps to bear in mind that hardship means different things to different people and in different cultures. ▪ Survival is about dealing with hardships and having the will to live. If the will to live is not there, then just having the knowledge and equipment may not be enough.	▪ Look at how the local people dress and act: in hot countries they leave manual work until the coolest parts of the day and work in a slow and deliberate manner to reduce sweating and therefore conserve water. ▪ If you're in a desert, for example, watch where animals go to find shade: they're mostly nocturnal and spend the day underground. Learn their tricks to find water, like the darkling beetle in the Namib Desert that drinks water from the fog that condenses on its carapace. ▪ If you're in a jungle, pay attention when the animals go quiet or quickly leave an area—danger is usually around the corner.	▪ Prior preparation is the key to survival: discover what you need to know about the environment you're going to; familiarize yourself with all your equipment; and practice your basic skills until they become second nature. This thorough preparation will help you to combat the fear of the unknown and give you the self-confidence to meet the challenges of any survival situation you may be in.

ATTRACTING RESCUERS

TO BE A SURVIVOR you need to rescue yourself or be rescued by others. If you can't rescue yourself—perhaps you're injured, completely lost, or trapped by bad weather—you must attract attention with location aids that you've either brought with you or have improvised.

AIDS TO LOCATION

Location aids can save your life, so make sure you know how to use them effectively. A helicopter on a search pattern may make only one pass over an area before moving on, so you'll only have a few minutes to act decisively.

COMMUNICATING WITH SATELLITES

When activated, a Personal Locator Beacon (PLB) transmits a radio distress signal to two complementary satellite systems called LEOSAR and GEOSAR. Together, these form the "COSPAS-SARSAT" system. The signal is then relayed to a rescue coordination center closest to the beacon's location. PLBs mainly use 406Mhz; the military also use 243.5Mhz and 282.8Mhz.

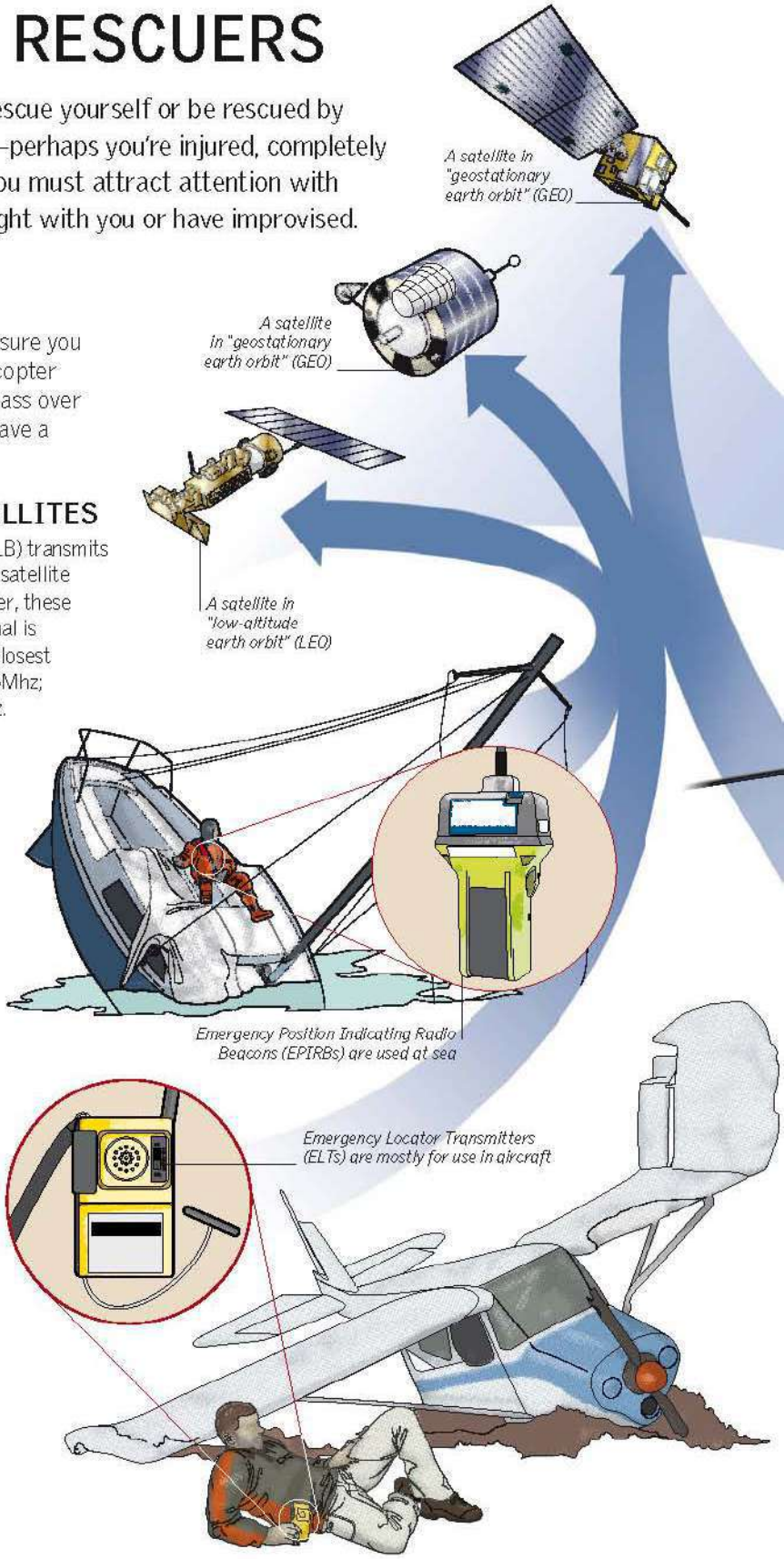

ATTRACTING ATTENTION

When you're in a survival situation, you need to employ your location aids in the best way possible. There are three main principles for attracting attention:

- **Attract:** pick a place that maximizes your chances of attracting attention using your location aids, such as open or high ground. Place your aids carefully so they can be detected from as wide an area as possible. The more obvious the signal the better.
- **Hold:** you need to hold the attention of rescuers by maintaining the signal until they indicate they've seen you. Try to send some critical information (the type of assistance you require or the number of survivors and their condition, for example) via a May Day, or Help message (see pp. 237–41).
- **Direct:** regardless of which location aids you use, once rescuers have been alerted, do all you can to direct them to your present location. If you have voice communication, direct them to your position. If you leave notes before moving on from places, make sure they contain dates and precise details of your intentions.

SHOULD I STAY OR SHOULD I GO?

You need to decide whether to stay where you are or to move to a location where you have a better chance of rescue or self-rescue. If you move on, remember the following:

- Keep your aids to location at hand. It's no good having your signal mirror at the bottom of your backpack when you may have only seconds to attract a passing vehicle or plane.
- Deploy your aids to location at the end of each day, even though it takes effort and you'll be moving in the morning.
- To show where you've been and where you're heading, leave markers, such as a note left in a visible position and visual clues on the ground or vegetation, to indicate your direction of travel.

MOBILE PHONES AND RADIOS

Always take a cell or satellite phone with you when traveling—they are essential in an emergency. On a sea trip, take a marine VHF radio to contact would-be rescuers.

CELL AND SATELLITE PHONES

Wherever you are in the world, you can rent or buy a cell phone linked to local networks. Choose one with a GPS unit for fixing and racking your position and a camera so you can send pictures of your location and any injuries you've sustained. Alternatively, rent or buy a satellite phone that's connected to the Iridium Satellite Phone System, in which 66 low-earth orbiting (LEO) satellites provide complete coverage of Earth (including oceans, airways, and polar regions).

MARINE VHF RADIOS

All large ships and most motorized small craft are equipped with marine VHF radios. These hand-held units transmit and receive on frequencies between 156 to 174MHz—usually on Channel 16, the international calling and distress channel. Channel 9 can also be used in some places. Transmission power ranges between 1 and 25 watts, giving a maximum range of up to about 60 miles (110 km) between aerials mounted on tall ships and hills, and 5 miles (9 km) between aerials mounted on small boats at sea level. Your VHF radio should be waterproof, able to float, and kept on charge. Follow the instructions whenever transmitting and receiving on VHF.

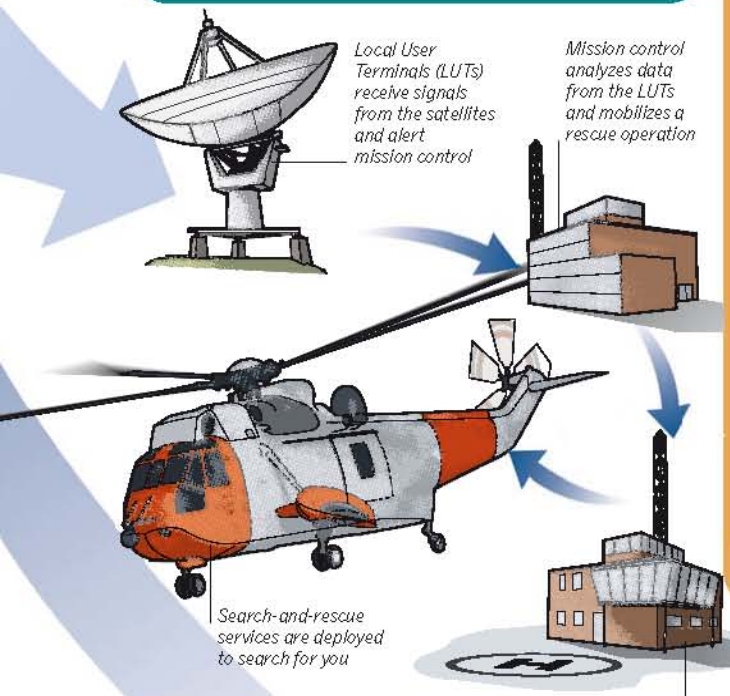

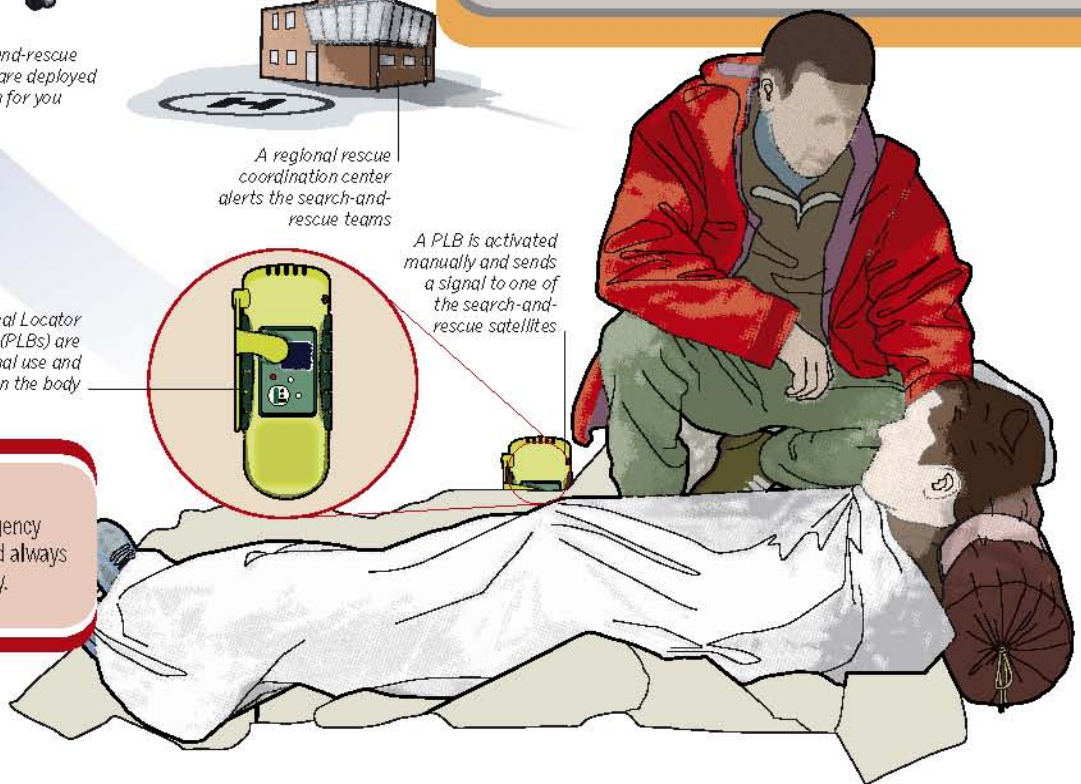

WARNING!
PLBs are for emergency use only and should always be used responsibly.

BUILDING SIGNAL FIRES

A signal fire is a very effective way of attracting the attention of rescuers, but you do need the right materials and building one requires some effort. A well-constructed signal fire located in a good position will generate a large amount of smoke that can be seen from a long way away. Two variations—the dome and the log cabin—are shown here.

TOOLS AND MATERIALS
- Knife and saw
- Poles and long saplings
- Green vegetation
- Tinder, kindling, and fuel
- Lighter or matches

SIGNAL FIRE ESSENTIALS

Whichever signal fire you make, there are several important principles to follow in order to make it as effective as possible.

FIRE FORMATION

If you can, prepare three signal fires in a recognized formation, such as a triangle or in a straight line. If the fires are random, they could be misread as a bush fire or a group of native fires. The distance between each fire should be at least 65 ft (20 m) but needs to be dictated by how quickly you can effectively light each one in turn.

FIRE COMPONENTS

Essentially, each signal fire is composed of a ready-to-light firebase made of dry tinder, kindling, and fuel. When you add the merest spark or flame to the tinder, the firebase immediately turns into a sustainable fire.
- Site the firebase off the ground to prevent damp from reaching it, and to allow enough airflow to help it ignite more effectively.
- Cover the top of the firebase with large amounts of green vegetation and anything else that produces smoke, such as tires. The covering keeps the firebase dry.
- Place a second stack of green vegetation near the firebase and add this as the original vegetation burns out.

AT THE READY

Once you have prepared the signal fire, you need to keep the following nearby, ready for use at a moment's notice:
- Dry tinder in a waterproof container underneath the firebase.
- Hexamine fuel tablets, stove fuel, gas, paper, or birch bark to guarantee the fire gets going again if it starts to dwindle.
- Something to create a spark or flame.
- A witch's broom, made of a cleft stick stuffed with kindling or bark, next to your camp fire, so you can quickly light the broom and transfer it to the signal fire.

MAKING A DOME SIGNAL FIRE

Prepare a large fire on a raised platform under a dome-shaped structure made from bent saplings. Fueled by air from below, the fire creates plumes of smoke from the green vegetation. If you have no saplings, use poles to a make a teepee shape.

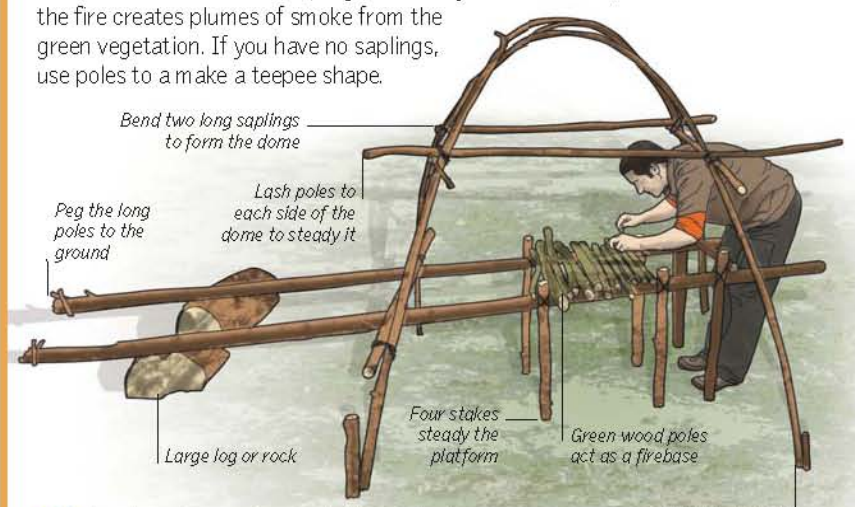

1 Lay two long poles parallel to each other. Prop them up over a log or rock at one end, and tie them to four stakes at the other. Lay lengths of green wood side by side on the long poles under the dome, to act as a firebase.
- Bend two long saplings at 90 degrees to each other to form a dome.

4 Layer green vegetation on top of the dome to form a roof over the fire platform.
- Keep additional green vegetation nearby to add to the fire when it's alight.

MAKING A LOG CABIN SIGNAL FIRE

This signal fire is known as the "log cabin" because of the way the fuel wood is stacked.

- The framework for the fire is a stack of green wood, made up of pairs of poles arranged in alternating layers set at 90 degrees to each other.
- Place an additional supply of green vegetation nearby to be used as required, but not so close that it catches fire by accident.

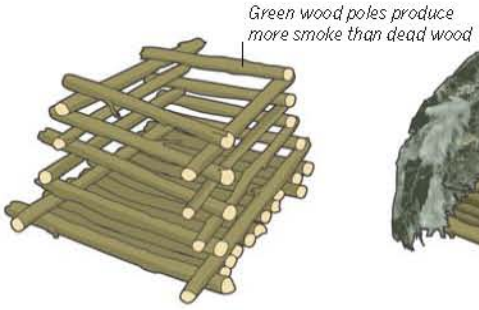

1 Lay a platform of green wood on the ground and build a firebase of dry tinder, kindling, and fuel on top.
- Build a "log cabin" over the firebase using green wood.

2 Heap green vegetation over the platform to generate thick plumes of smoke when the fire is lit.
- Leave an access point at the front to allow you to light the firebase.

2 Prepare a fire (see pp. 124–25) on the platform and load it with green vegetation.
- Lay the vegetation close enough to the fuel below to catch easily when the fire is lit—but don't smother it.

3 Prepare some kindling bundles which will help turn the initial flame into a fire.
- Make the bundles by breaking small, dead branches (as thick as a small finger) from the lower trunks of trees.
- Fold them into a small bundle held together with a thin branch or vine.

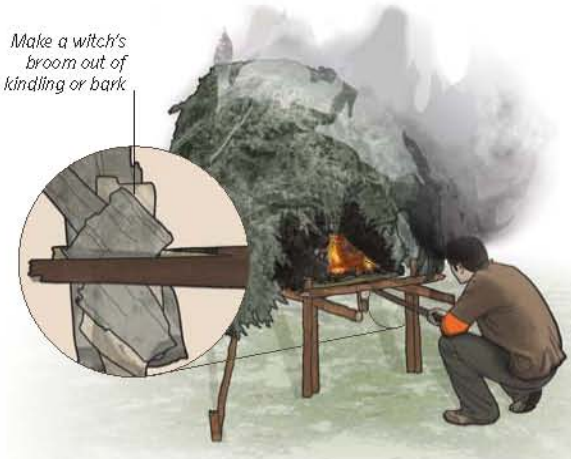

5 Light a witch's broom (see panel, left) from your camp fire, take it to the dome, and light the fire.
- Use kindling bundles (see Step 3) to fuel the fire.

6 When smoke billows upward, light a preprepared signal flare (see p. 241), if you have one, taped or tied to a long pole.
- Use the pole to position the flare as high into the smoke as you can.

OTHER RESCUE SIGNALS

As well as signal fires, there are other visual devices for attracting potential rescuers. Some are unusual and others, such as a whistle, quite obvious. Whatever device you use, you have to try to hold the rescuers' attention by persisting with your signal.

LIGHT WINDMILL

You can create a highly visible, illuminated "windmill" effect by whirling a chemical lightstick (cyalume) in front of you attached to the end of a cord 3 ft (1 m) long. Depending on local conditions it can be seen up to 2 miles (3 km) away—or further by an aircraft. Blow your whistle at the same time, using the International Distress Signal of six blasts over one minute, then one minute's silence. The reply is three short blasts.

CREATING A STROBE EFFECT

LED strobe lights are compact, robust, waterproof, and are sometimes incorporated within standard flashlights. They are very powerful and can be seen from a long way away. Many strobes can be set to produce a variety of sequenced flashes, including flashing SOS in Morse Code. Even if you don't have a dedicated strobe light, you can still switch your regular flashlight on and off repeatedly to attract attention.

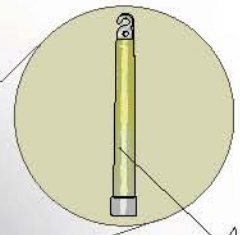

TINSEL TREE

The signaling device known as a "tinsel tree" is best used when you're remaining in one place for a while. It's effective only during daylight hours, so you'll need to find a suitable and easily visible tree or bush in a sunny position.

- Cut a silver survival blanket, a roll of aluminum foil, or any reflective material into strips.
- Attach the strips to the tree so that they move in the breeze, catching the sunlight and glinting like small mirrors.

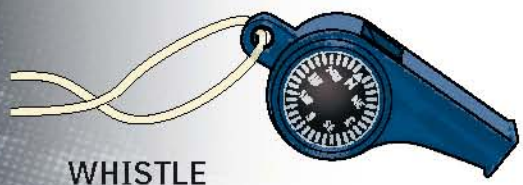

WHISTLE

Take a whistle with you when you venture out into the wilderness because sound, especially in remote areas where there's little or no noise pollution, travels incredible distances. A whistle is an item of first-line gear (see pp. 42–43), and should be kept on a cord around your neck.

SIGNALING DISTRESS

Two of the most recognized signs of distress are a little red star shooting up into the sky or a billowing cloud of orange smoke. There are many different types of signal flares and rockets, ranging from simple, handheld devices that fire flares into the sky to specialized kits designed to penetrate thick jungle canopy. Usually found in the emergency survival packs of aircraft and liferafts, flares are also available from specialized shops.

SIGNAL FLARES

One end of a signal flare has an orange-colored smoke signal for daytime use. The other end has a flare for night-time use (but it can be operated during the day, too). Remember the following:

- Follow the instructions on the outside of the device.
- Wear gloves to protect your hands.
- The heat from a flare can damage a liferaft, so keep your flare well clear when you light it.
- Don't discard flares unless you have used both ends.

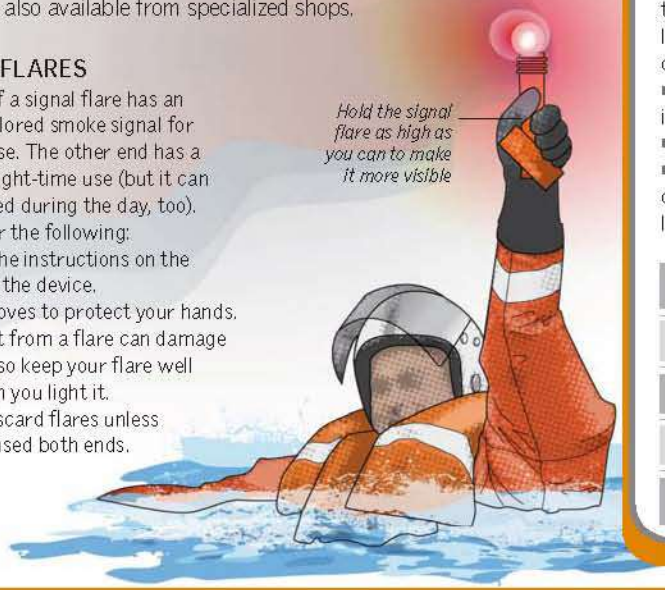

GROUND-TO-AIR MARKERS

You can improvise internationally recognized emergency signals on the ground that rescuers in the air can see.

GETTING NOTICED

To make a ground-to-air marker, use anything that contrasts with the ground, such as orange lifejackets, seaweed, clothing, rocks, branches of trees, or soil.

- Whatever you use, make sure the message is big and visible from all directions.
- Check on your markers regularly.
- When appropriate, use one of the emergency codes shown below. SOS or HELP written in big letters will attract attention too.

V	Require assistance
X	Require medical assistance
N	No/Negative
Y	Yes/Affirmative
↑	Proceed this way (arrow points)

SIGNAL MIRROR

A signal mirror, or "heliograph," has a shiny surface that reflects the sun and sends flashes over distances exceeding 30 miles (50 km), depending on the strength of the sun, the size of the mirror, and the clarity of the air. You can even use the light from the moon when it's full. By interrupting the flashes you can send messages in Morse Code.

IMPROVISING A SIGNAL MIRROR

If you don't have a heliograph or any other mirror, try using anything with a shiny, reflective surface, such as a foil food packet, a CD, or the bottom of a beverage can.

WARNING!
Don't dazzle your would-be rescuers by continuing to flash directly at them. Using your signal mirror intermittently is more effective for attracting attention.

1. Polish the base of a can to a shine with a slightly abrasive paste. You can use charcoal and water, toothpaste, or even chocolate.

2. Hold the can in front of your face with the polished base facing the sun.
   - Direct the flash of reflected sunlight on to the palm of your other hand to practice controlling the light.
   - Aim the can toward a rescuer and send a signal, either by moving your palm up and down to interrupt the flash or by directing the flash between a "V" made with your thumb and fingers.

WILD ANIMALS

WALKING IN THE WILDS of Alaska and coming across a fresh deposit of bear feces is a sobering experience. It quickly dawns on you that you're potentially a part of the natural food chain. What do you have to do to protect yourself, and will it be effective or just provoke the animal? Fortunately, the survival instincts of wild animals are such that, with only a few exceptions, most animals, regardless of their size, will avoid confronting you. However, they will defend themselves by attacking you if provoked, cornered, or surprised, particularly when they have young.

FERAL ANIMALS

Wherever you go in the world there are animals that have escaped into the wild from zoos, wildlife parks, and domestic situations. These feral animals include pigs, cats, dogs, mink, and muntjacs. In many cases, they have become particularly aggressive and seem to have lost any fear of confrontation with humans. Try to find out which feral animals are common in the region you're traveling in and avoid them if at all possible.

	BEARS	BIG CATS	HIPPOS	ELEPHANTS	CROCODILIANS
ANIMAL FACTS	• Brown bears stand taller then black bears and weigh more. Polar bears are the largest land predator. All bears are powerful, with claws on their strong paws. Brown bears can run faster than humans. • Bears are found in the Northern Hemisphere and parts of South America.	• Tigers are the largest cat, with lions a close second. All big cats have sharp teeth and claws. • Other dangerous cats include cougars (also called pumas, mountain lions, and panthers), leopards, and jaguars. • Big cats are found on all continents, except for Australasia.	• The hippopotamus is the third-largest animal in Africa after the elephant and white rhino. • Hippos can weigh more than 3 tons. They have long, razor-sharp incisors and tusk-like canines.	• African elephants have larger ears than Indian elephants and are more aggressive. • Elephants stand up to 13 ft (4 m) tall and weigh up to 6 tons. • They can reach speeds of 25–30 mph (40–48 kph).	• Crocodilians include alligators, gharials, and crocodiles. • They can stay submerged for more than an hour at a time, can swim up to 20 mph (32 kph), and run as fast as 11 mph (17 kph) over short distances. • Crocodilians live in many subtropical and tropical parts of the world.
PRECAUTIONS	• Look for signs of bears and check with locals about bear activity. • Stow and cook food away from the campsite. Keep unused food and garbage out of reach. • Avoid thickets and streams where bears rest and feed. • Carry a stick, knife, or spray (see opposite).	• Avoid coming into contact with big cats. They are most unlikely to attack unless you provoke them or threaten their cubs. • An exception is the cougar, which is responsible for an increasing number of unprovoked attacks in urban areas of North America.	• Avoid provoking a hippo. Many encounters are the result of hitting a partially submerged hippo with a canoe or boat. Almost all hippo attacks are fatal. • Stay vigilant on or beside rivers that hippos are known to frequent. • Don't get between a hippo and water.	• Keep away from places that elephants frequent, such as watering holes. • If you do come across them, don't get too close, especially if there's a baby elephant nearby. • Look for a safe place to retreat to, such as a vehicle, rocky outcrop, or a tree, before you need it.	• Stay away from waters and river banks where crocodilians live. Be vigilant at all times. • If you need to go near the water, watch the area for at least 30 minutes first. • Don't go to the same spot twice, as crocodiles may lie in wait the next time. • Keep a defensive weapon, such as a knife, close by.
WHAT TO DO IF...	• Make a noise to let a bear know you're there. • If you see a bear, stay calm. Make yourself look as big as possible by raising your arms. • Walk slowly backward. Don't run. If the bear follows, stop and hold your ground. • If the bear attacks, play dead, or fight back.	• If a big cat approaches, stare at its eyes, shout, and make a noise to confuse it. Use a spray if you need to (see opposite). • Don't turn your back or run away. • If a big cat sees you, do not crouch or bend down. An upright human makes less attractive prey than a four-legged animal.	• If you encounter a hippo, go back on your tracks and find another route. • If a hippo yawns at you, he's not feeling sleepy—this is a threat. He's showing you his teeth, tusks, and jaws that can snap a canoe in half! Make every effort to escape.	• If an elephant squares up to you with flared ears, trumpeting, and kicking the dirt in front of it, then back away. • If an elephant charges, run to your safe place (see above). • As a last resort, play dead and hope the elephant loses interest.	• If you meet a crocodilian on land, run away. • If a crocodilian does grab you, it may let go—if it does, run. • If it drags you into the water, fight back. Stab under its throat with a knife, hit or poke its eyes, strike its nostrils hard, or bang the large valve at the back of its throat.

ARMING YOURSELF

If you're traveling in areas where encounters with wild animals are possible, be ready with a weapon such as a knife, large stick, or deterrent spray. Remember, avoidance is always better than confrontation.

PEPPER SPRAY

One type of deterrent spray contains the pepper ingredients capsaicin and capsaicinoids. It shoots a cloud or stream that irritates an animal's eyes. The animal usually withdraws, but take care until the danger has passed.

	SHARKS	SNAKES	AMPHIBIANS
ANIMAL FACTS	▪ There are more than 450 species of shark, but only a few of them pose any danger to humans. These include great white, bull, and tiger sharks. ▪ Most attacks occur in coastal waters in the tropics and subtropics, especially where the water is murky or stirred up by the surf of breaking waves.	▪ Less than 15 percent of the 3000+ types of poisonous snake are regarded as being dangerous to humans. ▪ Depending on the snake, venom either affects the blood, the nervous system, or the heart. ▪ Snakes live in every part of the world except for very cold environments.	▪ Many frogs, toads, and salamanders secrete poison through their skin. The most poisonous frog known is the golden poison-dart frog from South America. ▪ Amphibians live where the climate is sufficiently warm and wet for them to breed.
PRECAUTIONS	▪ Avoid shark-infested waters or waters where sharks are known to feed or frequent. Ask the locals about recent sightings of shark activity. ▪ Carry something, such as a pole or spear, that you can use as a weapon.	▪ Wear long pants and boots, and cover your neck. ▪ As you walk, tap the ground in front of you with a stick to alert snakes to avoid you. ▪ Step onto logs rather than over them. ▪ At night, place your boots upside down on sticks. ▪ Don't put your hands into holes or cracks where snakes might be hiding.	▪ You can touch a poisonous frog or toad with no ill effect. The poison only takes effect if it enters your system via an open wound or your mouth or eyes. If this happens, get emergency help without delay.
WHAT TO DO IF...	▪ If you see a shark, stay calm and move to safety. ▪ If it makes a move toward you, swim away smoothly but keep watching it. ▪ If it rushes at you, hit it on the nose with whatever you have at hand. ▪ If it grabs you, aggressively strike its eyes or gills.	▪ If you come across a snake, try to remain completely still. Most snakes will instinctively move away and are more likely to attack a moving target. ▪ Don't panic. If you have a stick, slowly bring it to a position ready for use. ▪ If a snake attacks, hit it hard on the head.	▪ If contact with a poisonous amphibian is unavoidable, immediately wash the affected area. Keep the water you use away from open cuts or abrasions. ▪ Thoroughly wash your hands and don't put your fingers in your mouth or rub your eyes.

DO YOUR RESEARCH

Spend some time researching the area you'll be traveling in and find out which animals present the greatest dangers. Discover as much as you can about their habits and how best to avoid coming into contact with them.

SPOT THE SIGNS

Understand the habits of these dangerous creatures: when do they forage, sleep, and drink? What are their habits?

- Learn to recognize the natural signs that indicate their presence: tracks or prints; feces/droppings; territory markings, such as flattened vegetation where the animal rubs its scent on the ground; telltale signs, such as where bears have clawed the bark from trees.
- Be aware of other animals disappearing quickly from a watering hole, or a noisy jungle suddenly going quiet.
- Ask local people if there have been any reports of animals attacking people.
- Find out the best ways to ward off an attack or defend yourself should you have no option (see left). Adopting a certain stance or submissive posture, or avoiding or maintaining eye contact, could make the difference between safety or danger, life or death.

TRUE-LIFE ACCOUNT

EXTREME SURVIVAL—IN COLD CONDITIONS

USEFUL EQUIPMENT

- Waterproof, layered clothing
- Trekking pole or ice ax
- Collapsible snow shovel
- Light, whistle, avalanche beacon
- Tarpaulin
- Gore-Tex® gloves and gaitors
- Map, compass, GPS
- Survival tin, bushcraft knife
- Cell/satellite phone
- Poncho/bivy sack

34-YEAR-OLD ICE HOCKEY PLAYER ERIC LEMARQUE became lost in the snowy wilderness of the Sierra Nevada mountains in California after snowboarding at Mammoth Mountain ski resort. Despite being under-equipped and ill-prepared, his improvisation helped him to survive for seven days.

LeMarque was snowboarding alone late on Friday, February 6, 2004, setting off down an unmarked run of virgin powder snow just as the lifts were shutting for the day. Coming to a stop on a flat section, and with visibility reduced to just 10 ft (3 m), he realized he had lost the trail. Wearing uninsulated ski trousers and jacket, and with just an MP3 player, a cell phone with a dead battery, and wet matches, he was ill-prepared for survival, so he began to search for the trail.

LeMarque chose the wrong direction and walked away from the ski resort. Realizing he faced a night in the open, he used his snowboard to dig a crude trench, lining it with bark scraped from a pine tree for insulation from the snow. He tried to light a fire with pine needles and lint from his clothing, but his matches were too wet. He also ate pine needles and bark.

"WITH VISIBILITY REDUCED TO JUST 10 FT (3 M), HE REALIZED HE HAD LOST THE TRAIL"

Over the next five days, LeMarque walked further into the wilderness, leaving scraps of clothing for rescuers and attempting to signal to passing planes with the blue LCD screen of his MP3 player. Fortunately for LeMarque, rescuers spotted his snowboard tracks and followed it for 24 hours, finding him on February 13. He was barely conscious, dehydrated, hypothermic, and malnourished—having lost 35 lb (16 kg) in body weight—and was suffering from severely frostbitten feet, which later required both legs to be amputated below the knees. But he was alive.

WHAT TO DO

ARE YOU IN DANGER?

NO YES

If you are in a group, try to help any others who are in danger

Get yourself out of it:
Cold/wind/wet—These can quickly lead to hypothermia, so get yourself out of the elements quickly
Animals—Avoid confrontation and move away from danger
Injury—Stabilize condition and apply first aid

ASSESS YOUR SITUATION
See pages 234–35

DOES ANYONE KNOW YOU WILL BE MISSING OR WHERE YOU ARE?

NO YES

If no one knows you are missing or where you are, you will need to notify people of your plight by any means at your disposal

If you are missed, a rescue party will almost certainly be dispatched to find you

DO YOU HAVE ANY MEANS OF COMMUNICATION?

NO YES

You are faced with surviving for an indefinite period—until you are located or you find help

If you have a cell or satellite phone, let someone know your predicament. If your situation is serious enough to be worthy of emergency rescue, and you have a Personal Locator Beacon (PLB), you should consider this option

CAN YOU SURVIVE WHERE YOU ARE? *

NO YES

If you cannot survive where you are and there are no physical reasons why you should remain, you will have to move to a location that offers either a better chance of survival, rescue, or both

Address the Principles of Survival: Protection, Location, Water, Food

YOU WILL HAVE TO MOVE **

YOU SHOULD STAY **

DO

- Make an informed decision on the best location to move to
- Keep hydrated by wrapping snow in a wet item of clothing and suck it as it melts—but only if you are walking or working and generating heat
- Improvise a walking staff that can be used to check depth and quality of snow, and unseen drop-offs
- Check regularly for signs of frostnip or frostbite and hypothermia
- Have aids to location accessible while moving and deployed while static
- Protect all extremities from the elements—tie gloves to cord threaded through your jacket so they don't get lost

DON'T

- Underestimate the need for water just because it is cold—you are just as likely to get dehydrated in a cold environment as in a hot one
- Post-hole through virgin snow—it is exhausting. Instead, improvise a pair of snow shoes
- Sleep directly on the ground. Ventilate your shelter and first check for natural dangers (avalanche)

DON'T

- Use your body heat to melt snow as it lowers your body temperature and can induce hypothermia
- Breathe air onto cold hands—breath contains moisture which will then cool and conduct heat away from your hands
- Sit or lay directly on the cold ground. Use whatever is available to improvise a sitting or sleeping platform

DO

- Select a suitable shelter site away from dangers. Build it big enough for you and your equipment. Incorporate a cold sink and a sleeping platform higher than the sink
- Keep cutting and digging tools in the shelter in case a very heavy snow fall or an avalanche requires you to dig your way out
- Mark the entrance to the shelter so you can find it easily
- Keep a fire going: once established, you can use it to melt snow and benefit from warmth
- Prepare all of your aids to location for immediate use

* If you cannot survive where you are, but you also cannot move owing to injury or other factors, you must do everything you can to attract rescue.
** If your situation changes (for instance, you are "moving" to find help, and you find a suitable location in which you can stay and survive) consult the alternative "Do" and "Don'ts."

ENVIRONMENTAL HAZARDS

SOME PARTS OF THE WORLD have a reputation for sudden or extreme weather conditions, and unpredictable environmental hazards that cause havoc for everyone, not just travelers. Before you set off, make sure that you're aware of any such potential hazards in the areas you're visiting, and be ready for them, just in case.

BEING PREPARED

Many environmental hazards—from avalanches and volcanoes to forest fires, tornadoes, and hurricanes—can suddenly propel you into an emergency situation. While you can never be sure of escaping the worst excesses that Nature can throw at you, you can prepare yourself. A combination of the right equipment, local knowledge, and an awareness of evasive techniques will mean that you shouldn't have to find yourself in a survival situation thinking "if only".

COMMONSENSE ADVICE

No fail-safe guidelines will protect you from every environmental danger and hazard, but there is some commonsense advice that you will do well to heed before you venture into the wilderness:

- Check the prevailing weather conditions of your destination at the time of your journey. Log onto the internet or tune into local radio stations.
- Find out if you're heading into a particularly risky season of weather—for example, monsoons in India or hurricanes in the Caribbean.
- Take an emergency survival kit (see pp. 58-59).
- Learn the relevant survival techniques either before you go or as soon as you arrive at your destination—you may not get a second chance.
- Don't challenge Nature—it rarely loses.

EQUIPPING YOURSELF FOR AN AVALANCHE REGION

Before heading for the mountains, find out whether the prevailing conditions make avalanches more likely (see right).

- Take a collapsible shovel for moving snow and a probe for checking the depth of the snow.
- Keep some survival aids in your pockets in case you lose your backpack.
- Carry an avalanche transceiver—when activated, its signal can be detected by rescue services.

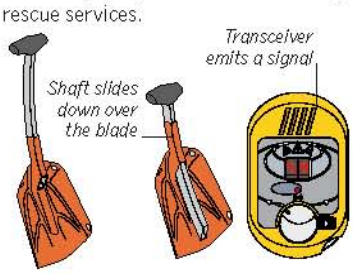

COLLAPSIBLE SHOVEL

AVALANCHE TRANSCEIVER

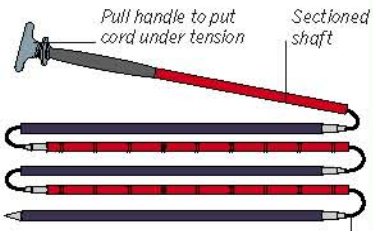

SNOW PROBE

SURVIVING AN AVALANCHE

There's always the danger of an avalanche on slopes that face away from the sun in the middle of winter, when a fresh layer of heavy snow sits on top of a weak layer of snow. An avalanche may be triggered when snow is disturbed by loud noises, an earth tremor, or the movement of skiers or snowboarders. Learn the warning signs, take the right equipment, and practice the emergency steps in case you get caught.

SPOTTING THE WARNING SIGNS

As you cross the snowbound mountains, hills, and valleys, look out for signs that warn of a possible avalanche:

- Convex slopes at an angle between 30 and 45 degrees.
- Slopes without trees or rocks.
- Loose, dry snow that doesn't settle.
- Soft, newly fallen snow that's more than 1 ft (30 cm) deep.
- Snow that sounds hollow.
- Snow that falls as crystals or pellets.
- Snow that falls at more than 1 in (2.5 cm) an hour.

TAKING EVASIVE ACTION

If you see or hear an avalanche, and you think it might be coming your way, take evasive action at once:

- Activate your avalanche transceiver in case you get caught in the snow.
- Try to take cover—for example, under a solid rock overhang if you can see one nearby (see below).
- If you can't find or reach suitable cover, try to sidestep the avalanche by skiing out of the way—at right angles to its potential path.

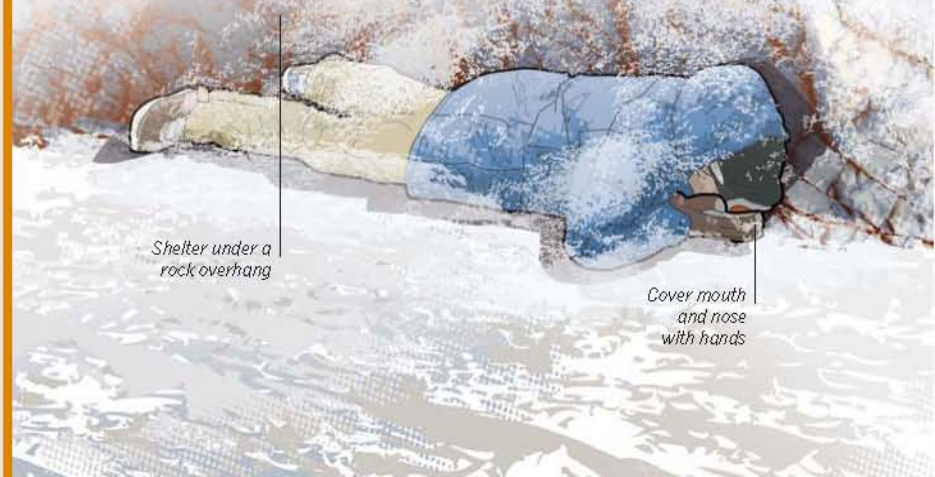

HAZARDOUS GROUND

All kinds of local conditions can prove treacherous if you don't keep your wits about you. Walking through wetlands can be risky, as the ground can suddenly give way under foot and you can find yourself in a swamp or, worse still, a patch of quicksand.

SURVIVING A SWAMP

Freshwater swamps are found in low-lying inland areas and contain masses of thorny undergrowth, reeds, and grasses, and there may be dangerous animals, such as crocodilians and snakes. The water may be foul and mosquitoes, which can cause malaria, are often present. Moving through a swamp is difficult: you can be on solid ground one moment, and chest-deep in water the next. Where possible, try to build a raft or some kind of flotation aid to help you escape. Use the waterways to navigate your way to open water, where you have the best chance of either being seen by rescuers or finding your way back to civilization.

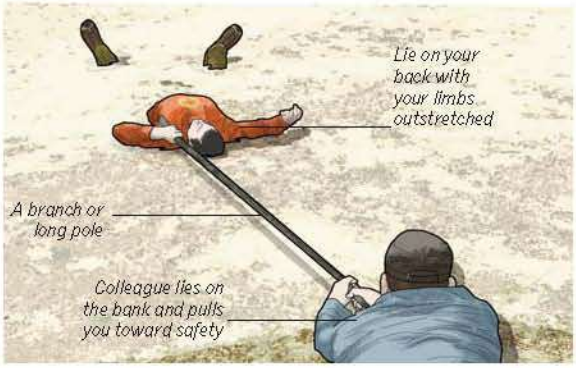

ESCAPING FROM QUICKSAND

If you step into quicksand, try to fall onto your back with your limbs outstretched to spread your weight. Try to get to the bank by paddling with your hands. Don't struggle as you will sink faster. If you're with someone, he or she needs to lie on firm ground and pull you to safety, reaching you with the help of a long pole, rope, or branch.

ESCAPING FROM AN AVALANCHE

If you can't take evasive action and find yourself below an advancing avalanche with no prospect of escape, try to keep calm and remember the following advice:

- Remove the bindings from your skis and the loops from your ski poles. Put your rucksack on one shoulder, but ditch it rather than risk dislocating your shoulder.
- If you're overtaken by snow, "swim" with the flow (see above) to make a big space around you before the snow compacts and "sets". Use your hands or a collapsible shovel to clear the snow around your head and to create a space for you to breathe, as often people suffocate before they freeze.
- If you don't know which way is up, dribble saliva to find out which way is down—then dig in the opposite direction.
- Get out quickly, as time is of the essence, and shout when you hear potential rescuers to attract their attention.

ESCAPING A VOLCANIC ERUPTION

Find out if there's an active volcano in the region you're visiting. If there is, and its eruption is imminent, leave the area as soon as you can. It will pump enormous quantities of lava from the mantle below the crust and produce huge amounts of ash, toxic gases, debris, and mud. However, if you do get caught in an eruption:

- Be careful when driving to safety, as the ash and mud make roads slippery.
- Take cover to avoid the flying fiery debris, hot gases, and suffocating clouds of ash. Sulfur dioxide in the ash chokes your lungs and, when mixed with rain to form sulfuric acid, burns your skin.
- If you're caught in an ash cloud, cover your face with a wet cloth (or use a mask if you have one). Wash your skin afterward.

SURVIVING AN EARTHQUAKE

Find out if an earthquake is likely to happen in the region you're visiting. If you feel an earthquake coming:

- Get into the open, away from structures or trees that can fall on you. Then lie down.
- Stay in open ground until aftershocks and tremors have stopped altogether.
- If you're in a building, go to the lowest floor, and stay beside a wall or under a sturdy table. Turn off the gas if you have time.
- If you're in a vehicle, stop, but stay inside.

SURVIVING A FOREST FIRE

When vegetation on the ground is tinder dry, the slightest spark can start a fire that gets out of control and sweeps through a forest. A forest fire may be caused by lightning, a piece of glass focusing the sun's rays, a discarded cigarette, or a spark from a camp fire. Forest fires produce intense heat, thick smoke, and toxic gases—and they use up all the available oxygen in the air, making it impossible to breathe.

PREPARATION AND SPOTTING THE WARNING SIGNS

If you're going to a place where forest fires are a hazard, find out from local radio reports whether conditions are making them more likely. If you venture into a forest, make sure you carry a cell phone or some electronic means of signaling, and let others know your route and destination. The following warning signs may help you gain time before a fire is upon you:

- You'll probably smell a fire first, and you'll hear it crackling as it burns before you see it. If you smell a fire and you notice animals growing agitated, a forest fire may be close by.
- The smoke will help you establish how close the fire is. The direction of the smoke tells you which way the wind is blowing.

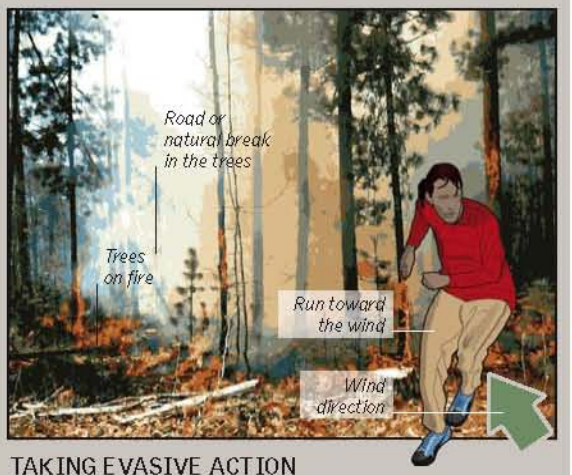

TAKING EVASIVE ACTION

If the wind is blowing toward the fire then move into it quickly. However, if the wind is behind the fire you could be in serious danger because the fire will be moving very fast.

- Try to find a river, lake, road, or natural break in the forest. Stay there until you're rescued or the fire has passed you by.
- Don't go up to high ground as fire is drawn faster uphill.
- Many forest fires spread on a wide front, so avoiding the fire by going around it may be impossible.
- If the fire is upon you, and the wall of flame is fragmented, the best course of action may be to run through the flames: take off man-made clothes (as they will melt onto you), cover as much of your skin as you can, and douse yourself in water if you have any. Take a deep breath, press a damp cloth to your mouth and nose, pick a spot where the wall of flames is thinnest, and run without stopping until you're through.
- If you're in a vehicle, stay inside. Park it as far away from the trees as possible, turn off the engine, close the windows, lie down on the floor, and cover yourself if you can.
- If escape is impossible, try to dig and bury yourself under soil. Dampen all clothing and get as low as possible.

EXTREME WEATHER

High winds and torrential rain are common seasonal occurrences across the world. From tornadoes and whirlwinds to tropical storms such as hurricanes and typhoons, these extreme weather patterns cause enormous damage and threaten many lives.

AVOIDING HURRICANES AND TORNADOES

Weather forecasts can predict approximate directions and paths of impending hurricanes and local emergency services will give advice on the predicted severity. If you're planning a trek in an area known for hurricanes and tornados, check long-range weather forecasts and be prepared to cancel or alter your plans if necessary—if you're caught in the open your chances of survival are extremely limited. If you're going to be in the wilderness during a "storm season" then take a battery- or solar-powered radio, check for weather updates, and be prepared to head back to civilization. If your home is in a tornado or hurricane area, follow the advice of local services, but most importantly be prepared with a plan that will ensure you don't get caught out.

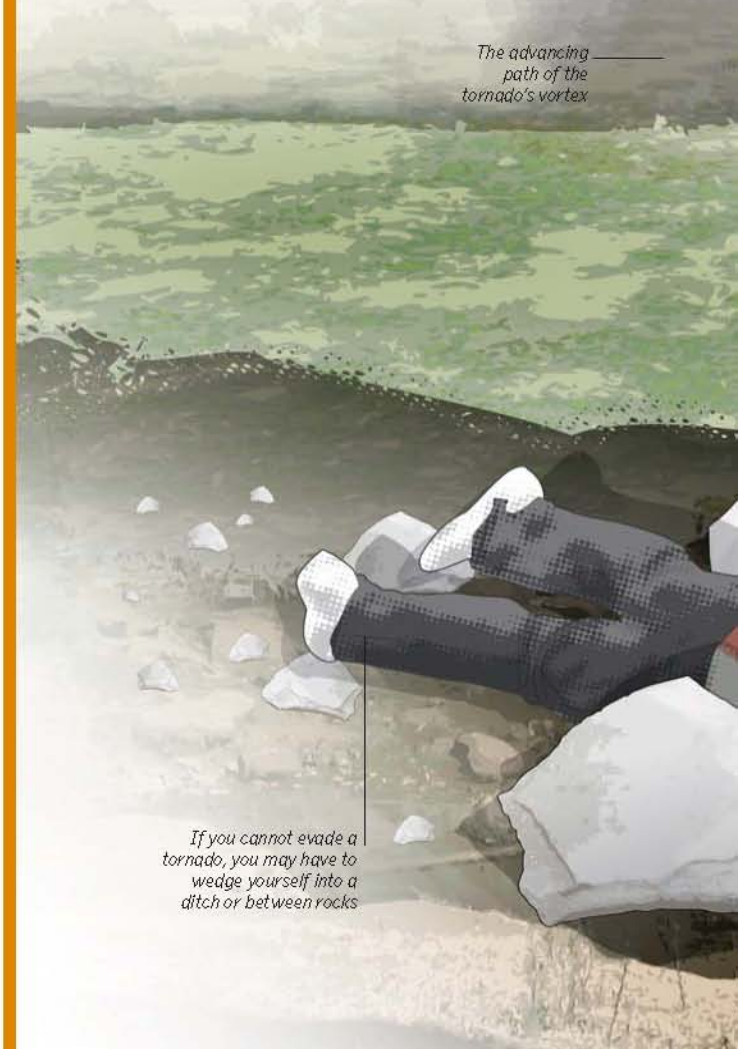

PREPARING TO EVACUATE

The arrival of a hurricane or tornado is a time when you can find yourself in a survival situation on your own home turf. The following advice can help you prepare to evacuate.

- Secure anything outside that could be picked up by winds and cause damage, such as rubbish bins and garden furniture.
- Prepare an evacuation plan (including pets if you have them). Find out where you will evacuate to and make several route plans to the location in case roads are unusable. Fuel your vehicle as soon as a hurricane warning is forecast.
- Put together a hurricane survival pack that contains what you need for a 72-hour period. Include drinking water, a change of clothing, non-perishable food, sleeping bags, radio, flashlights, and contact numbers for family and emergency services.
- Protect your house by securely boarding up all windows and doors. Turn off the water, gas, and electricity.

STAYING PUT

If you decide to stay put, you should still follow the guidelines above. As the storm approaches, move everyone and your supplies to an underground shelter or a room without windows.

- Monitor the radio reports and comply with the advice of the emergency services.
- Don't go outside until the "all-clear" has been given.

TAKING COVER OUTDOORS

If it appears that you are, or could be, in the path of a tornado or hurricane, do everything you can to move out of its way—choose a direction at a right angle to its path.

- If you're driving but can't get out of the way, try to park in an underpass. Leave the vehicle and wedge yourself into a secure position as high as you can under the underpass. If you're in the open, shelter upwind of your abandoned vehicle, so that it does not become a hazard to you.
- If you're out in the open, head for the lowest ground and find a ditch or depression (see below), in which you can shelter, or wedge yourself between rocks.

SURVIVING AN ELECTRICAL STORM

An electrical storm may develop when warm air rises and meets colder air. Electricity sparks between water droplets in the clouds, forming lightning that takes the shortest route to the ground. The following advice will help you avoid being struck by lightning, or minimize its effects if you are:

- Avoid getting caught in open ground—seek shelter but not under a lone tree: when lightning strikes a single tree a tremendous voltage fans out from its base.
- Remain in your vehicle if there is no other cover: it acts as a Faraday Cage (a metallic enclosure that prevents the entry or escape of an electromagnetic field).
- If caught in the open: make yourself as small as possible and limit the amount of area you cover. Do not lay down or stand up, crouch down low with feet together and hands off the ground.

SURVIVING A FLASH FLOOD

A sudden deluge of rain doesn't always drain away quickly, instead flowing rampantly over the surface of the land in torrents, and causing flash floods. Soil, animals, vegetation, and even buildings can be quickly washed away. There may be landslides, too, and rivers may break their banks. The following advice can help in the event of a flash flood:

- If you're inside a building, move to the upper floors, taking with you essentials such as bedding, food, and matches. Unless the building is threatened, stay there until the waters have receded or you have been rescued.
- If you're outside, head for higher ground.
- Never walk or drive through a flood.
- Filter and boil water for drinking, as the sources of water around you may have been contaminated. Alternatively, collect rainwater to drink.

ESCAPING A SANDSTORM

If you can see a sandstorm coming, mark your direction of travel before it strikes and find somewhere safe to shelter—for example, behind some rocks. Face away from the direction of the wind and cover yourself as completely as you can, particularly your head, face, and neck.

SURVIVING AT SEA

THE SEA IS ARGUABLY THE HARSHEST environment of all and you could be thrown into a sea-survival situation for any number of reasons: bad weather, fire, mechanical failure, or collision. However, there's plenty of good advice to follow and equipment designed to protect you.

BE PREPARED
If you're venturing onto the open seas, you should plan for the worst and be prepared with knowledge, skills, and equipment.

KNOW YOUR VESSEL

Regardless of the size of your vessel, find out everything you can about what to do in the event of a major problem that requires you to abandon ship and take to the water. The points below are generic and may not apply to all situations or types of vessel.

LARGE PASSENGER VESSELS

If you're a passenger on a large vessel, such as a ferry, liner, or cruise ship, make sure you learn the safety procedures.

- Attend the "Abandon Ship" drills.
- Find out which emergency alarms indicate fire, collision, and abandon ship.
- Find out where the lifejackets are stowed, and how to put them on and operate them.
- Learn the escape routes.
- Locate the Emergency Lifeboat Stations and find out what your responsibilities are.
- If you have children or people with special needs with you, make sure you have a system for getting them on deck and providing them with suitable survival equipment.

SMALL VESSELS

These craft include yachts, small boats, canoes, and kayaks, so crew members need to agree on an emergency plan of action and delegate responsibilities relevant to the situation and the skills of each person.

- Know where the emergency equipment is stowed and how it's operated.
- If your vessel has an EPIRB (see p. 236), make sure everyone knows where it is and how it's operated. Those activated automatically are released hydrostatically from a bracket at a water depth of 3–10 ft (1–3 m). The buoyant EPIRB then floats to the surface and begins transmitting.
- Make sure your EPIRB has been registered so that, once the signal is detected, the rescue services know who it belongs to.
- Keep a grab bag to hand (see box, right).

SURVIVAL SUIT AND LIFEJACKET

A survival suit is designed to keep you warm and dry in extreme conditions and rough seas, and a lifejacket will keep you afloat with your head out of the water, even if you're unconscious.

PROTECTION

Survival suits and lifejackets are equipped with various protection features to help you survive in the sea.

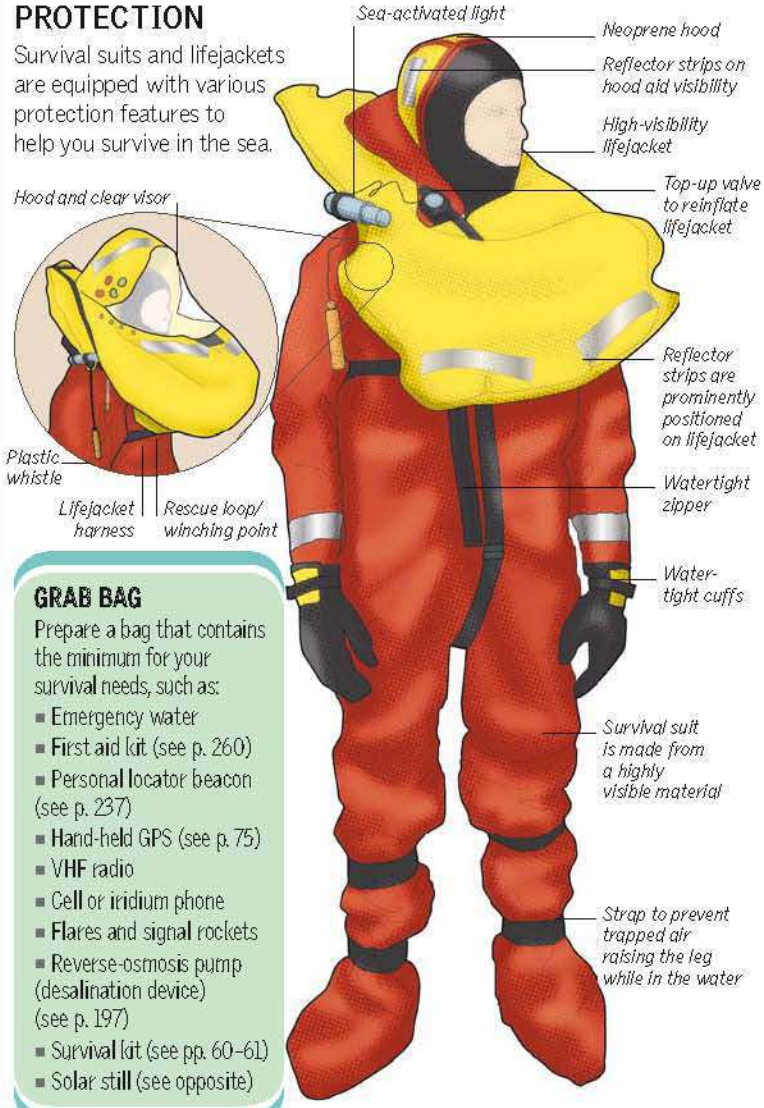

GRAB BAG
Prepare a bag that contains the minimum for your survival needs, such as:

- Emergency water
- First aid kit (see p. 260)
- Personal locator beacon (see p. 237)
- Hand-held GPS (see p. 75)
- VHF radio
- Cell or iridium phone
- Flares and signal rockets
- Reverse-osmosis pump (desalination device) (see p. 197)
- Survival kit (see pp. 60–61)
- Solar still (see opposite)

TYPES OF LIFERAFT

Single-seater and multi-seater liferafts contain similar design features and survival aids. As well as being able to accommodate more people, multi-seaters carry larger quantities of fresh water and more anti-sea sickness tablets, for example.

MULTI-SEATER LIFERAFT

Many vessels carry multi-seater liferafts in a valise or a hard container. Multi-seaters can accommodate between four and 25 people and may be open-topped or covered. Many larger multi-seaters may also include locator beacons (see p. 236), paddles, and a solar still (see below).

SINGLE-SEAT LIFERAFT

Only one person can fit into a single-seat liferaft. You can inflate it with carbon dioxide, by pumping air in manually, or by oral inflation.

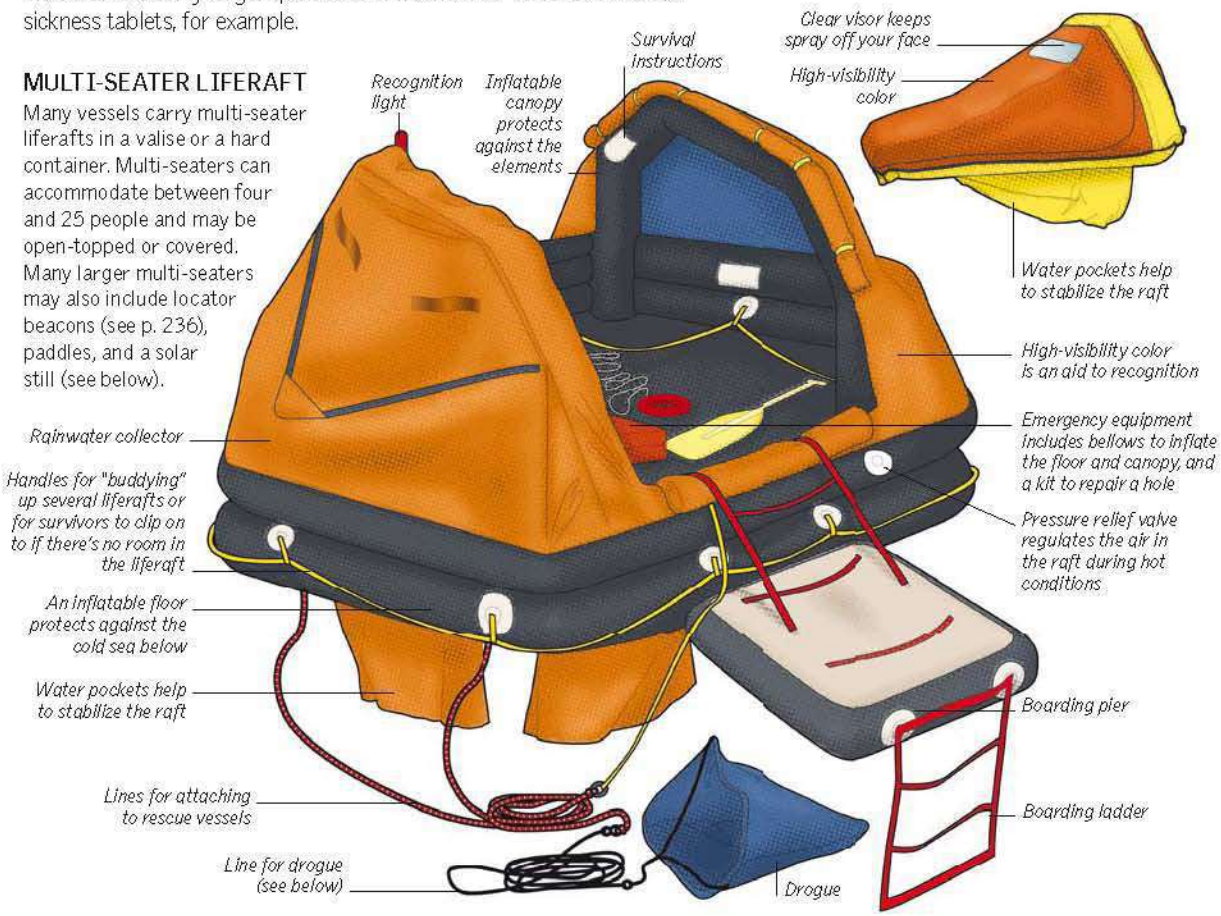

SOLAR STILL

A solar still is a light, compact, and easy-to-use device for producing drinking water from sea water. Heat from the sun evaporates the salty water inside the still. Condensation on the walls collects in a channel around the rim and is directed to a store. Depending on the prevailing conditions and the availablilty of sunshine, a solar still can produce as much as 2 liters (3½ pints) of fresh drinking water a day.

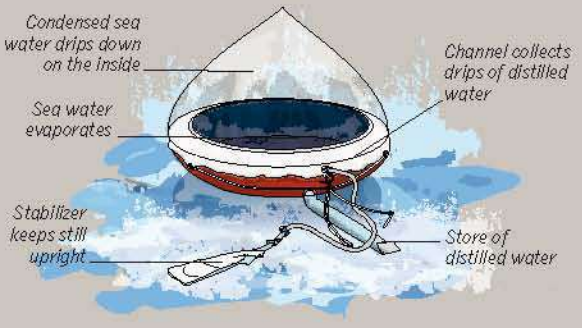

DROGUE

A sea anchor, known as a "drogue," is a critical piece of equipment because it helps to reduce drift and keeps the liferaft stable and seaworthy, particularly in heavy seas.

STEADYING A RAFT

A drogue stops a liferaft from overturning by creating drag that "anchors" the trailing side of the raft to the water.

- A drogue can position a liferaft either downwind or 90 degrees to the wind.
- It helps to keep the raft near the ditching location, thereby improving the chance of rescue. Even in a 2-knot current a liferaft can drift 50 miles (80 km) a day.

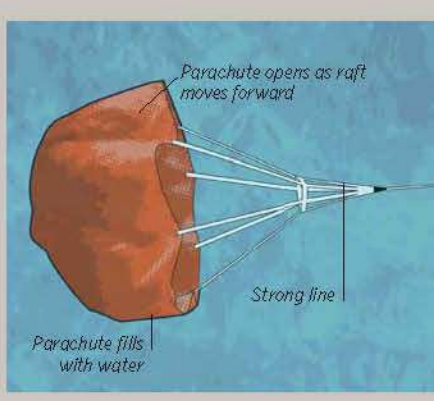

ABANDONING SHIP

Unless your vessel is an immediate danger to you, don't abandon it until you really need to. Even a badly damaged vessel can provide you with protection from the elements, equipment such as radios and flares, and provisions such as water and food. In addition, it is a big target for rescuers to spot. Make every effort to keep the vessel afloat.

THE DANGERS THAT AWAIT

When you do decide to abandon ship, you will be faced with the following dangers. If you have prepared yourself properly before heading out to sea, you will—in many cases, at least—be able to deal with them.

- Hypothermia brought on by inadequate clothing and exposure to wet, windy, and rainy conditions.
- Drowning because you have no lifejacket.
- Dehydration caused by a lack of water or an injury.
- Malnutrition due to a lack of food.
- Cold shock (see p. 254) due to sudden immersion in cold water.

BEFORE ABANDONING SHIP

Where time and circumstances permit, don't abandon your ship until absolutely necessary. Send a May Day signal with your name, position, group size, physical condition, and circumstances. Then carry out the following measures:

- Activate the 406 EPIRB (see p. 236).
- Ensure everyone on the vessel wears layers of clothing, and has a survival suit and lifejacket (see p. 250).
- Check the liferafts are ready for launch.
- Get the grab bag (see p. 250).
- Prepare to take the jerry-cans of water on board the liferafts.
- Fill spare containers with water.
- Gather up as much food as you can.

TAKING TO THE LIFERAFT

If you have to abandon ship, make every effort to launch all available liferafts. Even the craft that are not used will make it easier for search and rescue teams to detect the "footprint" of the survivors. They also contain additional supplies of water, flares, and other useful items.

- Many liferafts have a painter. This is a line that attaches the liferaft to the vessel to ensure it does not blow away when thrown overboard and inflated.
- Should the vessel sink, the painter has a weak link that breaks under pressure, or you can cut it.

PREVENTING HYPOTHERMIA

Getting wet greatly reduces your survival chances. At 41°F (5°C), a normally dressed person has only a 50 percent chance of surviving for one hour. You are six times more likely to survive in protective clothing.

- Climb into the liferaft carefully so that you don't get wet.
- Wear several layers of clothing to trap air. Even wet layers retain some heat around the body. Don't forget your head, hands, and feet.
- Wear a survival suit (see p. 250) to increase your survival time.

1 Move slowly and deliberately into the liferaft, keeping out of the water, and trying to stay as dry as possible. If you have to enter the water, climb down or lower yourself into it, rather than jumping.

- Look for the Day-Glo instruction patch inside the liferaft. Follow the recommended advice under "Immediate Actions" (see panel, opposite).

IN THE LIFERAFT

Once you have successfully negotiated the tricky procedure of abandoning ship and climbing into the liferaft, there are a number of actions you need to take. These are prioritized into immediate, secondary, and subsequent.

IMMEDIATE ACTIONS

- Inflate the floor of the liferaft with the bellows, while carrying out a roll call of your group to check for missing members.
- If the vessel is still afloat, keep the raft attached via the painter. Someone should be ready to cut the painter close to the vessel in case it starts to sink.
- Once clear of the vessel, set up the drogue (see p. 251).
- In difficult weather conditions, close the entrances of the raft in order to keep in heat, and keep out wind, rain, seawater, or spray.
- Bale out water, check for leaks, use the sponge to dry the liferaft, and use the leak stopper and clamps if necessary.

SECONDARY ACTIONS

- Treat the injured.
- Take anti-seasickness tablets.
- Post lookouts.
- Bring liferafts together.
- Warm up as best you can.

WARNING!
Never drink seawater, because the salt it contains will increase your rate of dehydration. If you're in hot conditions without any water, dehydration can set in within one hour.

SUBSEQUENT ACTIONS

- Delegate a leader based on experience.
- Find out who has useful skills, such as first aid and sea survival training.
- Ask everyone to look for sharp objects that could puncture the liferaft. At the same time, ask them if they have anything that could have a survival use.
- Find out what survival aids you have. Prepare aids to location (see p. 236) and show everyone how to operate them.
- Establish a routine and detail a watch system for inside and outside the liferaft.
- Delegate one member of the group to make repairs, another to keep the rations, and another to administer first aid.
- Establish how much water and food you have and start to ration in accordance with your particular situation.
- Procure water as soon as possible—don't wait until you need it. Deploy solar stills (see p. 251) and use the reverse-osmosis pump (see p. 197).

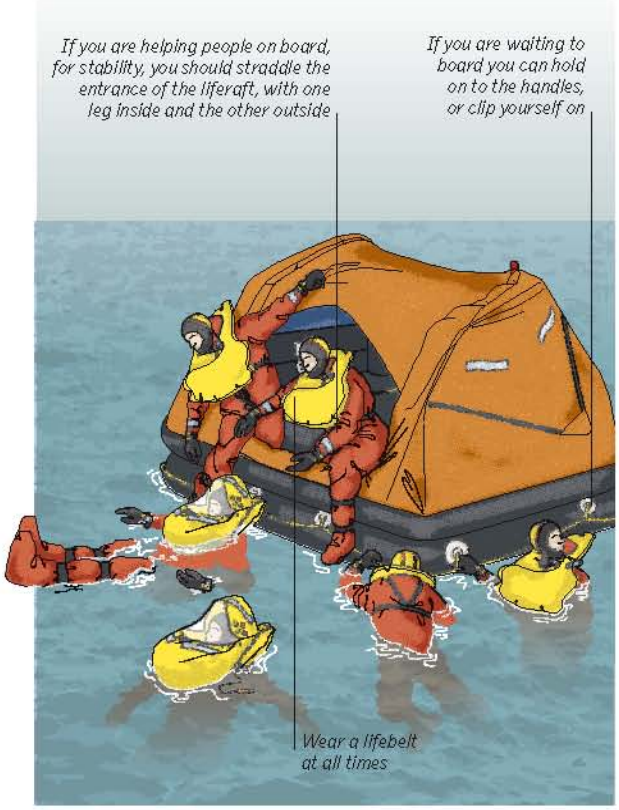

2 If your group does end up in the water, the first two people on board the liferaft should help others on, one at a time, lifting them under the arms.

- Don't overload the raft. Those least vulnerable can hang on to the raft's handles, or tie their lifejacket lifeline to one.

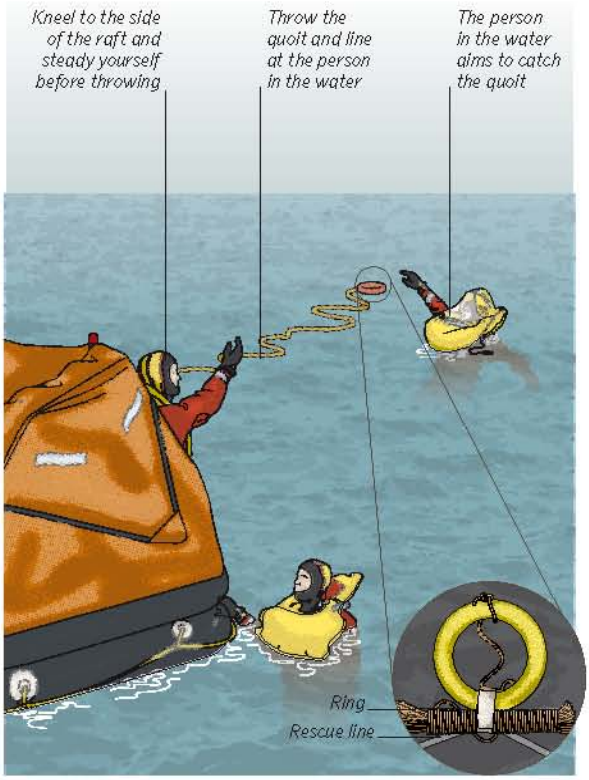

3 Look for anyone who is struggling to make their way toward you. Use the rescue line and ring to pull them to safety in the liferaft.

- Avoid entering the sea unless you have to rescue an unconscious person from the water.

TAKING TO THE WATER

About two-thirds of people who drown in open water were within 10 ft (3 m) of a safe refuge and 60 percent of them were "good" swimmers. Here are some useful tips to prevent drowning.

JUMPING INTO THE WATER

Abandoning ship can be a dangerous procedure. If you have to enter the water, climb down using ropes or nets. If you have no option but to jump into the sea, these straightforward steps should improve your chances of survival. Jumping into the water is a last resort, as you will be prone to cold shock (see box, right) and at immediate risk of hypothermia and drowning. If you have to do it, first ensure your lifejacket is fitted correctly and check the surface of the water to find a safe entry point. Watch out for people, debris, and burning fuel. If you can't avoid the fuel, swim under it and, before coming up for air, poke your hand through the surface to ensure you're clear of it. As you come up, keep your face down to protect your nose, eyes, and mouth.

1 Once you've decided on your entry point, stand on the lowest part of the vessel.

- Close mouth and pinch nose to stop water from entering.

2 Keep yourself upright with your back straight, then jump clear of the vessel.

- Cross your ankles and lock them together before entering the water.

3 Once you're in the water, inflate your lifejacket.

- If you're in a survival suit, raise your arms and gently pull one seal away from your wrist to let the excess air escape.

4 Lie on your back and swim as calmly as you can away from hazards, such as burning fuel.

- Push your head back onto the neck of the lifejacket to help raise your hips.
- Keep your feet and knees together, and use a backward butterfly stroke to swim toward the liferaft.

COLD SHOCK

Cold shock response is the body's reaction to immersion in very cold water—for example, if abandoning a ship in winter in the North Atlantic Ocean. It is a common cause of death in such circumstances. Symptoms include gasping for breath and hyperventilation, which can lead to the inhalation of water, disorientation, panic, and the possible onset of hypothermia. A sudden increase in blood pressure and heart rate can cause cardiac problems in some people. Prolonged immersion in water will make it hard to perform physical movements, making swimming, climbing aboard a liferaft, or firing a flare extremely difficult.

- Body type or mental conditioning can help some people survive swimming in icy water.
- Dressing in layers and/or wearing a survival suit improves your chances of survival.
- Avoid entering the water if you possibly can.

IF YOU HAVE NO LIFERAFT

Even without a liferaft, you stand a better chance of surviving at sea if you're in a group. More survivors create a larger target for rescuers to see, and being with others can be good for morale.

- Collect any floating debris before it disappears with the current, as this can increase your "footprint" so that rescuers can see you more easily.
- Determine what location aids and other equipment are available and prepare these for use.
- If there are children or injured among your group, place them in the center of the group and huddle close together.
- If you are alone, get into a position known as H.E.L.P. (see below), which stands for "Heat Escape Lessening Position." This will help to keep heat within the core of your body.

FLOATING FACE DOWN

Don't panic if you have to enter the water without a lifejacket. Your body's natural buoyancy will keep at least the top of your head above water. To keep your face above water, too, make small windmill motions with your arms extended. However, if the water's rough, your only option may be to float face down.

STAY POSITIVE

It's very important to remain positive and to focus on your situation right now rather than on what might happen in a few hours or tomorrow.

1 It's important to relax, even though this seems hard when your life is in danger.

- Let your face lie in the water, and put your arms out in front of you.

2 Begin to exhale into the water as you raise your head up.

- Lift your head as it breaks through the surface and empty your lungs completely before inhaling more air.

3 With fresh air in your lungs, duck your face back into the water, keeping your mouth closed.

- Let your body float again, before you repeat the sequence.

IMPROVISING A BUOYANCY AID

If you're in the water and wearing pants, you can improvise a buoyancy aid that will help you to keep your head above water. Initially, it may be awkward to organize, but the benefit outweighs the effort involved.

1 Take off your pants and tie the legs together near the bottom. Tighten the knot as much as you can with your teeth.

- Flick the pants over your head from behind until they fill with air. Tread water as you do this.

2 Quickly catch the waist with your hands and grip it tight to hold air in.

- Put your head between the trouser legs and float.
- You will need to repeat the process regularly.

SECONDARY DROWNING

A person who nearly drowns inhales water that can cause potentially fatal chemical and biological changes in the lungs. This is known as "secondary drowning." It can be caused by inhaling fresh or salty water—even as little as 1fl oz (30 ml). Secondary drowning may occur 24-72 hours after immersion and, while uncommon, it's not rare. Take the following precautions to guard against secondary drowning:

- Monitor anyone who has nearly drowned or has spent some time in water.
- Ask the person to take a couple of deep breaths and check for any pain or discomfort that can't be accounted for.
- Look for symptoms such as coughing, breathing difficulties, chest pain, and saliva that looks like foam.
- Make sure the person is sitting in an upright position.
- If possible, provide oxygen—the blood levels of oxygen can fall quickly—and lots of rest and reassurance.

FIRST AID

FIRST AID

Regardless of the events that put you in a survival situation, a major factor that will dictate your options and actions will be whether anyone has been injured. Someone's survival may depend on the treatment they receive at the time of the incident, and during the ongoing survival ordeal, so it's essential that everyone should be proficient and confident in basic first aid techniques. In a true survival situation, the term "seek medical help" really means "deal with it yourself." While it would be impossible to be able to treat all injuries—a task even a paramedic would find daunting without the right equipment—the vast majority can be either dealt with or stabilized using a combination of basic first-aid techniques and common sense.

The "Protection" element of the survival principles (see p. 27) applies to your situation at all times—you should continually look at the consequences of every action in relation to protection against injury. Prevention is always a better option than cure. In hot climates, being able to recognize the first stages

In this chapter YOU WILL DISCOVER...

- that **maggots** have their **uses**...
- how to **improvise goggles** to prevent **snow blindness**...
- when to **stop**, **drop**, and **roll**...
- how to treat **snake bites** and **jellyfish stings**...
- when to **eat charcoal** or drink **bark tea**...
- how to prevent **frostnip** from becoming **frostbite**...
- when to **perform** a **firefighter's lift**...

of heat stress will allow you to take action with them before they develop into life-threatening dehydration and heatstroke. In cold climates, being able to recognize the first stages of frostnip will allow you to stop it from becoming frostbite. In many cases, further injury can be avoided by adopting the basic principles detailed throughout this book, such as being able to assess the best course of action in a given situation, plan a route and move safely over terrain, and protect yourself against the elements.

The will to survive is often the only factor that determines whether you live or die—regardless of your equipment, training, knowledge, and skills.

WHEN FACED WITH everything that man and nature can throw at you, when there appears to be no hope, you'll be faced with two choices: will you accept your situation and wait and see what hand fate deals you, or will you endure the pain and discomfort and fight for you own survival?

This determination was clearly shown by Aron Ralston during a hike in Blue John Canyon, Utah. After accidentally dislodging a 800 lb (363 kg) boulder, which pinned his right arm, he was faced with a bleak outlook. After five days, aware that no one knew he was missing and having run out of water, Ralston decided to amputate his own arm with his knife, applied a tourniquet, and hiked off to safety.

In another famous example, Simon Yates took the agonizing decision to cut the rope that held his injured climbing partner, Joe Simpson (who Yates thought had died), over a crevasse in the Peruvian Andes. Incredibly, Simpson survived the fall and chose to endure. He spent three days without food and only splashes of water from melting ice as he crawled and hopped the 5 miles (8 km) back over frozen mountainous terrain to reach their camp.

> "The **vast majority** of **injuries** can be either **dealt with** or **stabilized** using a **combination** of **basic first aid techniques** and **common sense**"

FIRST AID ESSENTIALS

SAFETY IS KEY on any expedition. Before you set off, make sure that you and everyone in your party has the necessary medical equipment—especially medications. If someone is injured, he or she should be treated immediately. If you can't contact the emergency services (and if there are enough of you), one person should stay with the casualty while two others go for help.

WARNING!
Protect yourself from danger at all times. You can't help anyone if you become a casualty as well. If the area is unsafe, don't approach the casualty—get emergency help and monitor the casualty's condition from a safe distance.

BASIC FIRST AID KIT

Keep your first aid kit (see panel, opposite) dry and readily accessible. Check the seals on sterile dressings; if they're not intact they're not sterile. Replace anything you use as soon as you can.

PRIORITIES AFTER AN INCIDENT

Assess a situation quickly and methodically. Find out what happened. Check casualties for life-threatening conditions such as unconsciousness or severe bleeding (see p. 264) and treat those first.

- **Response** Is the casualty conscious or unconscious? If the casualty is responding to you, he or she is conscious. Shake the shoulders gently if you're not sure.
- **Airway** Is it open and clear? If the casualty can talk, it is. If he or she is unconscious, open and clear it (see p. 276).
- **Breathing** Is it normal? Treat any difficulty such as asthma (see p. 275). If he or she is unconscious and not breathing, call for emergency help and begin CPR (see p. 277).
- **Circulation** Are there any signs of severe bleeding? If so, treat immediately. Once life-threatening conditions are under control, you can make a more detailed assessment. Examine the casualty methodically from head to toe. Ask how the incident occurred, as it can indicate likely injuries.

Tale fabric, waterproof, and hypoallergenic plasters

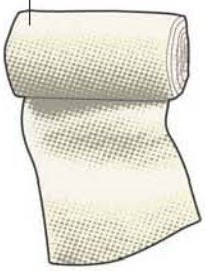
Gauze roller bandage to secure dressings—can be washed and reused

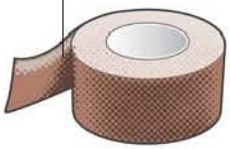
Zinc-oxide tape can be used to secure dressings

IMPROVISED SLINGS

Hand, arm, or shoulder injuries need to be immobilized and supported in a raised position. If you don't have a triangular bandage with you, use a piece of strong cloth about 3 ft (1 m) square, folded in half to form a triangle (see p. 270). You can also use your jacket or even rucksack straps. The jacket corner "sling" is the only one that provides enough support for a hand, wrist, or forearm injury. Ask the casualty to support the arm with his or her other hand while you secure the sling.

JACKET CORNER
To support an injured forearm or hand, fold the jacket up over the arm and pin it.

BUTTON-UP JACKET
Undo one of the buttons and slide the injured arm into the opening for support.

BELT SUPPORT
Support an upper arm injury in a raised position with a belt looped into a figure-eight.

PINNED SLEEVE
Pin a sleeve to a jacket or the strap of a backpack for support.

SHOULDER STRAP
Rest a sprain by tucking your hand in your backpack strap.

PROTECTING AGAINST INFECTION

Disposable gloves prevent cross-infection between you and the casualty; they must be latex-free, since contact with latex can cause an allergic reaction. Antiseptic wipes are also invaluable when cleaning wounds.

LATEX-FREE DISPOSABLE GLOVES

ANTISEPTIC WIPES

COMBINED STERILE DRESSING

This is a sealed dressing consisting of a pad attached to a bandage. It's easy to apply, and can be used as a sling. It should be taped on to your backpack strap for easy access in an emergency.

CHECK LIST

Make sure that your first aid kit and medicines are suited to the environment you will be visiting.

BASIC KIT

- Alcohol-free antiseptic wipes
- Latex-free disposable gloves
- Alcohol gel for handwashing
- Antiseptic cream
- Antibiotic eye ointment
- Adhesive dressings—fabric, waterproof, and hypoallergenic
- Gel blister bandages
- Combined sterile dressings, or sterile pads and bandages in assorted sizes
- Roller bandages—take self-adhesive for supporting joints and gauze for securing dressings
- Two triangular bandages
- Micropore or zinc oxide tape
- Scissors and tweezers
- Safety pins
- Disposable syringes

PERSONAL MEDICATION

- Painkillers
- Anti-inflammatories
- Medical alert bracelet/pendant
- Prescription medicines such as asthma inhaler and/or adrenaline (epinephrine) autoinjector
- Antihistamines
- Anti-diarrhea medicine
- Packets of oral rehydration salts
- Hydrocortisone cream

ENVIRONMENT-SPECIFIC EXTRAS

- Malaria tablets
- Mosquito repellent
- Anti-poison-ivy cream
- Sunblock
- Tick remover
- DEET powder for removing leeches

FLESH WOUNDS

ANY INJURY THAT BREAKS THE SKIN carries a risk of infection as germs can enter the body. These can come from the cause of injury, air, dirt, or clothing embedded in the wound. In the wilderness, keeping the injured area clean can be a real challenge, but it is essential. Tetanus is a potentially lethal infection caused by bacteria that live in soil. It can be prevented by immunization, so ensure that your vaccinations are up to date.

MAGGOT THERAPY

If a wound is seriously infected and you have no antibiotics, expose it to flies for one day, then cover. Maggots will develop and eat any dead tissue. Check daily, and flush the maggots out of the wound with sterile water before they start to eat healthy tissue.

BLISTERS AND BRUISING

A blister is a fluid-filled "bubble" of skin that occurs when skin is rubbed repeatedly against a surface (a friction burn). A bruise is bleeding into the skin and surrounding tissue from a blow that does not break the surface of the skin.

HOW TO TREAT BLISTERS

The ideal treatment for a blister is to rest and wait until it has healed, but this may not be possible in a survival situation. Cover it with a gel blister dressing if you have one or, if the blister is large, you may have to pierce it to enable you to continue walking. However, never pierce a blister caused by a burn—you risk infection.

1 If you have to burst a blister, first sterilize a needle by holding it in a flame until it's red hot, then letting it cool.
- Clean the area with water or wipes, pat dry, then pierce the edge of the blister.

2 Carefully apply pressure to the side of the blister opposite the hole made by the needle. Continue to apply pressure until all of the fluid has been squeezed out.

3 Gently clean and dry the blister, then protect the wound from infection by applying a dressing.
- Use a gel blister plaster or padded moleskin, if possible.
- Alternatively, use a pad held in place with zinc-oxide tape.

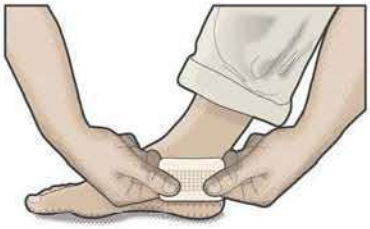

BLISTER PREVENTION

Following these simple rules when out hiking should prevent blisters from developing in the first place:
- Ensure boots or shoes fit properly and are well "broken in" before you set out.
- Always wear clean, dry, comfortable socks next to your skin. Avoid wearing two pairs as they may bunch up, increasing friction.
- Keep toenails cut short and straight.
- Remove boots and air your feet during rest periods.
- If you feel a "hot spot," treat it before it becomes a problem: stop immediately and apply moleskin, a gel pad, or zinc-oxide tape.

HOW TO TREAT BRUISING

To reduce the swelling and pain of a bruise, raise the injured area, and apply a cold compress. If you have access to cold running water, ice, or snow, soak a cloth and hold it against the bruised area for at least ten minutes. Severe bruising may indicate a more serious injury, such as a broken bone (see p. 271) or internal injury (see Shock, p. 274), which will require immediate treatment.

WOUNDS AND BLEEDING

Severe bleeding can be distressing for both you and the casualty, but it can usually be controlled by a combination of direct pressure and elevation of the injury. Remain calm and reassure the casualty while treating the wound. Treat for shock (see p. 274).

FOREIGN OBJECT IN A WOUND

Any loose foreign objects, such as pieces of dirt or gravel, should be removed from a wound, otherwise they may cause infection or delay healing. Either rinse them off with cold running water or carefully pick loose pieces off the wound with tweezers. However, if an object is embedded in a wound don't remove it, but treat as shown below.

1 Don't try to remove the object as it may be plugging bleeding. Control bleeding by pressing firmly on either side of the wound.

- Push the edges of the cut together but take care not to press directly on the object.
- Raise the wound above the level of the heart.

2 Place a piece of gauze over the object to protect it, then build up padding on either side (rolled bandages are ideal). Bandage over the pads and the object.

- Check the circulation beyond the bandage every ten minutes (see p. 264).

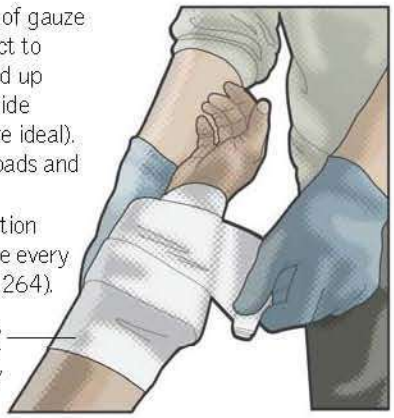

MINOR CUTS OR GRAZES

Any break in the skin, however minor, needs to be cleaned and protected from infection. Rinse the wound with clean cold water and pat it dry. Then cover the injury; the dressing pad must be larger than the wound. For small grazes, use a plaster; for larger ones, use a sterile pad and a bandage.

NATURAL WOUND DRESSING

The birch polypore, or razorstrop fungus, which grows on the sides of old birch trees, can be used as a natural wound dressing. Cut a thin slice from the top of the fungus and secure it over the wound. If you are in any doubt about the identity of the fungus, do not use it.

TYPES OF WOUND

Different types of object and force produce different kinds of wound. It's useful to identify the type of wound you or the casualty has incurred so that the correct method of treatment can be applied.

BRUISE (CONTUSION)

A blunt blow will break blood vessels under the skin, causing blood to leak into the tissues. This results in a bruise: the skin is tender, swollen, and blue-black in color.

GRAZE (ABRASION)

A friction burn from a rope, or a sliding fall, will scrape off the top layers of skin, leaving a raw, tender area. Grazes often contain embedded foreign matter.

TEAR (LACERATION)

If the skin is torn open, the wound may not bleed as badly as an incised wound, but a larger area of tissue may be damaged and vulnerable to infection.

PUNCTURE WOUND

Sharp objects such as nails or sea urchin spines can puncture the skin. The entry hole will be small but the wound will be deep, with a high risk of infection.

INCISED WOUND

If a sharp-edged object cuts across the skin, blood vessels will be sliced open and bleeding will be severe. Nerves or tendons may also be damaged.

STAB WOUND

A penetrating wound from a long, bladed instrument is a very serious injury. A stab wound to the torso can damage vital organs and cause internal bleeding.

GUNSHOT WOUND

Check if the casualty has an exit wound. Treat entry and exit wounds separately.

ENTRY WOUND

Bullets drive deep into, or through, the body. They leave a small, neat entrance wound and cause serious internal damage and contamination.

EXIT WOUND

If a bullet passes through the body, the exit wound will be large and ragged. Don't remove the bullet if there is no exit wound.

SEVERE EXTERNAL BLEEDING

Control the bleeding with direct pressure over a wound and, if possible, call for emergency help. Never use a tourniquet as this can cause severe tissue damage. Life-threatening shock is likely to develop if blood loss is severe (see p. 274).

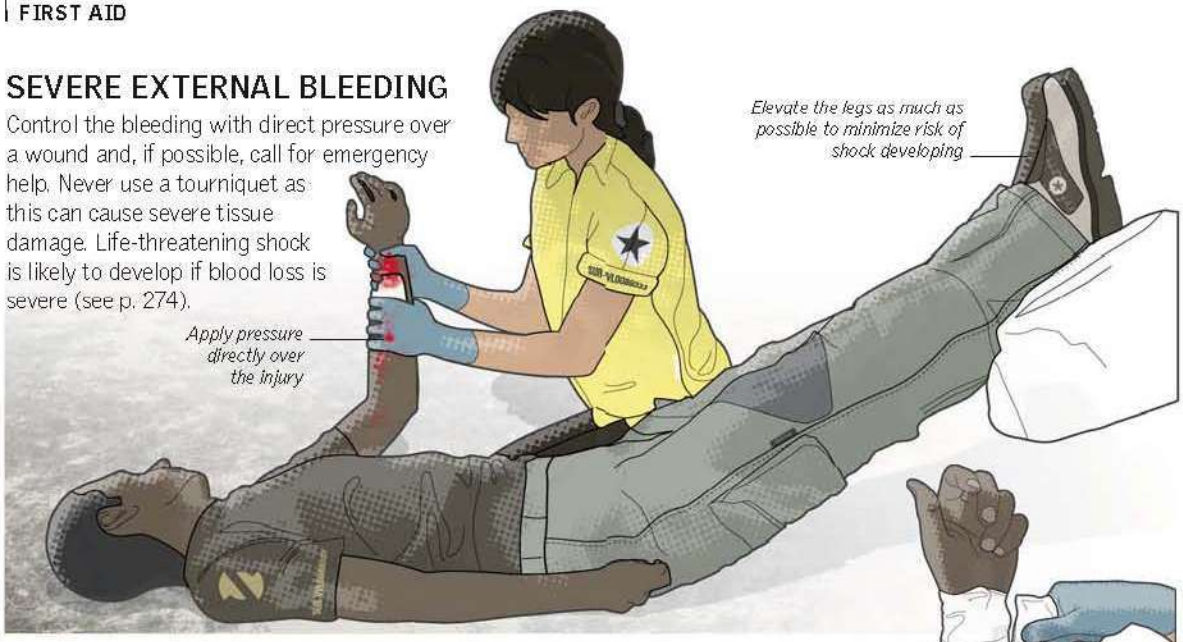

1 Remove or cut away clothing to expose the wound. Apply direct pressure to the wound, over a sterile dressing or pad if you have one. Raise the injured limb above the level of the casualty's heart to reduce blood flow to the area. Help him to lie down and raise his legs.

2 Secure the dressing with a bandage. If blood soaks through, apply a second dressing on top of the first.

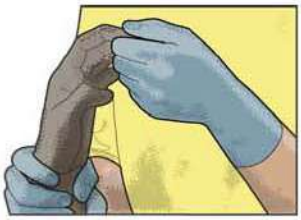

3 Every ten minutes, check the bandage is not too tight. Gently press a fingernail beyond the dressing. If the skin color does not return quickly, rebandage more loosely.

VARICOSE VEINS

When the one-way valves in veins fail, blood pools behind them, causing raised knobbly skin. The taut, thin-walled veins can easily be burst by a knock. Bleeding will be profuse.

1 Help the casualty to lie down and raise and support the injured area as high as you can. This reduces bleeding straight away. Expose the wound and apply direct pressure over a sterile dressing or pad.

2 Bandage the pad firmly to maintain pressure on the wound. Keep the area raised. Check the bandage is not too tight (see step 3, above). If necessary, loosen it but maintain pressure.

SCALP WOUNDS

A scalp wound can bleed profusely, making it appear worse than it actually is. However, it may mask a more serious head injury. If a casualty becomes drowsy, has a headache, or double vision, get emergency help if possible.

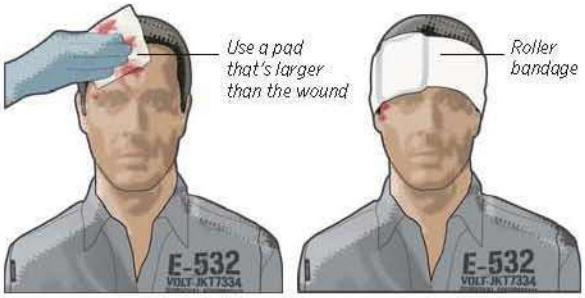

1 Sit the casualty on the floor. Carefully replace loose flaps of skin, cover the wound with a sterile pad, and apply pressure.

2 Secure the dressing with a roller bandage. If the casualty doesn't quickly recover or if he deteriorates, get help.

EYE INJURIES

The eye can be seriously injured by a blow or contact with sharp objects, such as a tree branch, risking scarring, infection, or even loss of vision. Always wear eye protection when hiking through dense vegetation.

1 Lay the casualty down with his head on your knee. Cover the injured eye. Ask him to keep both eyes still, as moving one eye causes the other to move. If an object is sticking out of the eye, pad securely around it.

2 Secure the dressing with a bandage. If you're on your own, fix the dressing in place with plasters or tape and try not to move your eyes.

FOREIGN OBJECT IN THE EYE

If you can see an object on the surface of the eye, try lifting it off with the corner of a handkerchief, or wash it out with sterile water, washing away from the good eye. Don't remove anything that is sticking to the eye.

SNOW BLINDESS

Also known as "flash burn," this occurs if the surface of the eye is damaged through exposure to ultraviolet light, such as the glare from sun reflected off snow or water. Wearing sunglasses can prevent this. In an emergency, make your own goggles from cardboard or birch bark. If a person is affected by snow blindness, give him gauze pads to hold against his eyes. Bandage them in place if help is delayed.

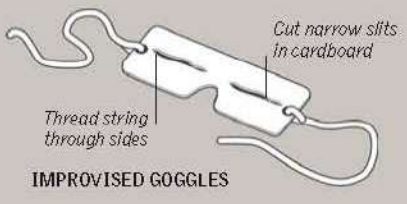

IMPROVISED GOGGLES

BURNS AND SCALDS

There is a serious risk of infection with all burns. Burns may affect only the outermost layer of skin, the upper layers, or the full thickness. Severe burns will affect all three layers. If the burn is larger than the casualty's hand it needs hospital treatment.

SMALL SUPERFICIAL BURNS

Cool the injury for ten minutes with cold running water or any cold, harmless liquid. Remove jewelry or watches from the affected area before it swells. Cover the burn to prevent infection.

LARGE OR DEEP BURNS

If a burn is extensive or deep, then fluid will be lost from the body and life-threatening shock is likely to develop (see p. 274). If the injury occured in a fire, the casualty could also have burns to his air passages so may have breathing difficulties. Don't burst any blisters as you will increase the risk of infection.

STOP, DROP, AND ROLL

If clothing is on fire:
- Stop moving
- Drop to the ground
- Roll over on the ground until flames are extinguished

1 Help the casualty lie down and protect him from the ground if possible. Cool the injury by dousing it with cold water for ten minutes; this also reduces swelling and relieves pain.

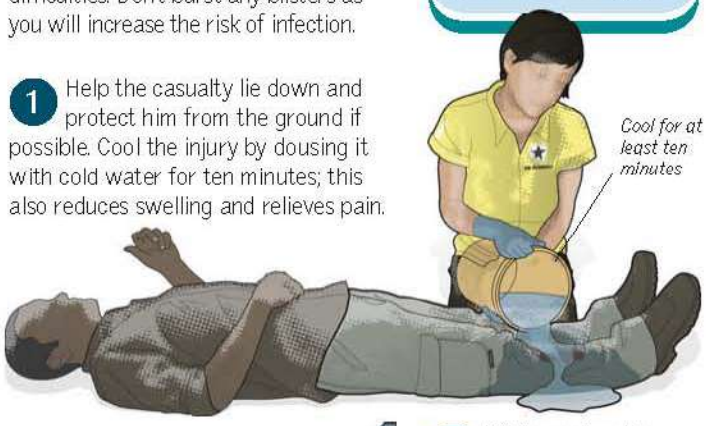

2 While cooling the burn, remove or cut away clothing from around the burn. Don't touch or remove anything that's sticking to the burn.

3 Cover the entire area with plastic kitchen wrap, a clean plastic bag, or a clean, lint-free dressing. Get emergency help as soon as possible.

BITES AND STINGS

MANY BITES OR INSECT STINGS are painful, but most can be treated with simple first aid. However, there is a risk of an allergic reaction called anaphylactic shock. If a casualty develops a red blotchy rash, watery eyes or puffiness around the eyes, and/or breathing difficulties, get urgent medical help (see p. 274).

TREATING BITES

An injury that breaks the skin carries a risk of infection. This is greatest with animal bites, since the mouth harbors so many germs. For a snake bite, you'll need to get emergency help, as the casualty needs to be transported on a stretcher in the treatment position to prevent the venom spreading through the body.

PORCUPINE QUILLS

If you are pierced by a porcupine's barbed quills, you must remove them or they will continue to work their way into the flesh and may pierce a vital organ.

- Cut off the ends of the quills, which are hollow, to allow them to deflate slightly.
- Pull the quills out, using pliers if you have them.
- Clean the puncture wounds and apply an antiseptic ointment.

SNAKE BITES

Relatively few snakes are venomous, but it's generally safer to assume that all of them are. Most snakes cause a painful bite that leaves small puncture marks in the skin; however, the bite of a poisonous snake may in fact be painless. Symptoms also include nausea and vomiting, disturbed vision, and breathing difficulties.

1 Keep the casualty calm. Help her to lie down with her head, chest, and shoulders supported.

- Advise her to keep still.
- Make a note, if you can, of the time the bite occurred.

2 Apply a bandage around the site of the injury. Do not wash the area, remove any footwear or clothing, or attempt to suck the venom.

3 Tie another bandage around the affected limb that extends from the bite as far up the limb as possible.

- Check the circulation (see p. 264).
- Immobilize the affected limb.
- Move the casualty only with the use of a stretcher.

MAMMAL BITES

Bites from mammals—animals or humans—carry a serious risk of infection because the sharp teeth cause puncture wounds that transport bacteria deep into the tissue (see p. 263). As always, wear gloves to protect yourself. The bite can also crush surrounding tissue, and bleeding may be severe (see p. 264).

CLEAN WOUND

Raise the injury. Wash the wound with clean gauze and water. Pat dry, and cover it with a wound dressing.

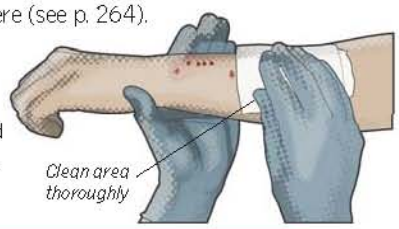

SNAKE IDENTIFICATION

Identify the snake if you can, since this will help medical teams find the correct antivenom. If you're not certain of the identity, make a note of its color or distinguishing features. If it's safe, put the snake in a secure container, but bear in mind that venom is active even if a snake is dead. Don't wash the venom off the bite; it can be used to identify the antivenom.

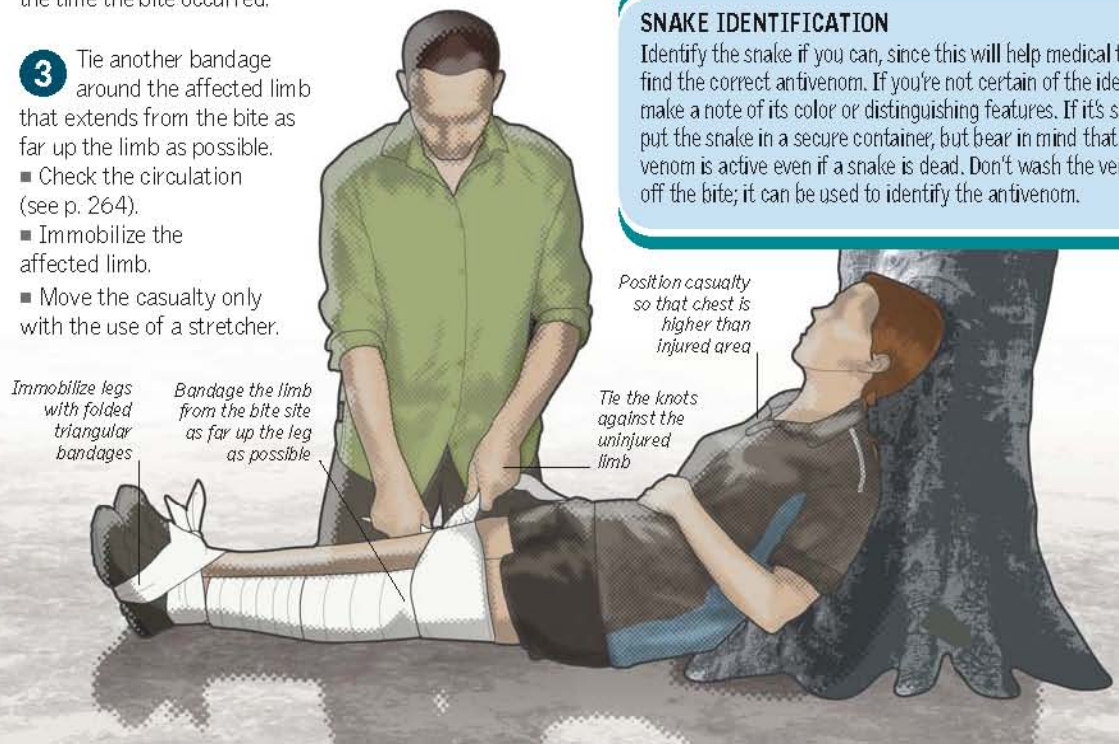

TICK BITES

Ticks are tiny, spider-like parasites that live in grass or woodland. They feed on blood, attaching themselves to the skin with spiked mouthparts, and swell to the size of a pea. They carry Lyme Disease, so must be removed.

TRADITIONAL METHOD

Using tweezers, grasp the tick's head as close to the skin as you can. Pull the head upward using steady pressure (don't twist). Keep the tick in a container so it can be checked for Lyme Disease.

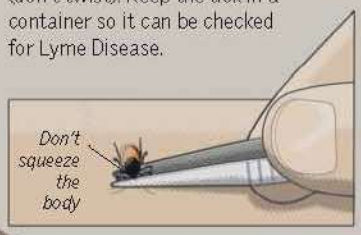

SPECIALIZED REMOVAL HOOK

Tick extractor tools that "unscrew" the mouthparts of the tick from the skin are available from pet shops. Slide the hook along the skin to grab the tick. Raise the hook very slightly and rotate it to lift the tick clear.

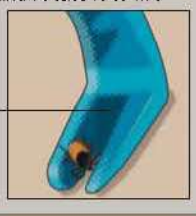

HANDLING LEECHES

When in leech-infested territory, inspect your clothing and limbs every few minutes. Never pull a leech off if it's attached, as the jaws will remain in the skin and cause infection. To remove:

- Apply the juice of a raw lime, DEET, alcohol, or a dash of salt. If you're a smoker, put your cigarette butts in a piece of cloth, moisten the cloth, and squeeze nicotine onto the leech.
- Once the leech has dropped off, wash the area to remove the anticoagulant in and around the wound. Treat any bleeding (see p. 263) and cover the wound.

TREATING STINGS

Many stings are painful, but they are rarely life-threatening. Scorpion stings can be very painful and cause severe illness, and treatment should be as for snake bite (see opposite). Multiple insect stings of any type can produce a more serious reaction (see anaphylactic shock, p. 274).

WARNING!
A sting in the mouth or throat can cause swelling that blocks the airway. To minimize this, give the casualty cold water to sip.

INSECT STING

A bee, wasp, or hornet sting is often painful and followed by swelling and redness at the site. Some people have an allergy to stings, so monitor the casualty for signs of anaphylactic shock.

1 If the sting is visible, scrape it off sideways with the edge of a credit card, knife, or a fingernail.

- Don't squeeze the sac as you may squeeze more venom into the area.

2 Raise the affected part and place a cold pad against it for at least ten minutes to minimize swelling.

- Monitor the casualty for signs of allergy such as wheezing, or swelling around the face.

SEA CREATURES

When touched, sea anemones, corals, jellyfish, and Portuguese men-of-war release venomous cells that stick to the skin. Treat jellyfish stings as below; apply a cold compress to other stings to minimize swelling and relieve pain. Creatures such as weever fish have sharp spines that, if trodden on, become embedded, and may become infected.

JELLYFISH STING

Pour vinegar or seawater over the area to neutralize the sting. Help the casualty to sit down; treat as for snake bite (see opposite).

SEA URCHIN SPINES

Immerse the injured part in water as hot as the casualty can tolerate for about 30 minutes. Get medical help, as the spines must be removed.

POISONOUS PLANTS AND INTESTINAL PARASITES

IT'S ESSENTIAL TO MAINTAIN personal hygiene in camp, to purify all drinking water from natural sources (see pp. 198–201), and to observe the rules of food safety if you wish to remain healthy in the wilderness. You should also learn how to recognize and avoid the poisonous plants in the region you're traveling in.

CONTACT POISONING

Poisonous plants don't always have to be ingested to cause damage. If they come into contact with the skin, the result can be pain, swelling, redness, a rash, and itching. Get medical help fast and, in the meantime, rinse the skin with cold water for 20 minutes. If the water splashes an eye, rinse for ten minutes. Although plants are the most common cause of contact poisoning, chemicals such as camp fuel can also be harmful.

WARNING!
If you think you've handled a poisonous plant, don't touch especially sensitive parts of your body such as the eyes, mouth, or genitals until you have washed your hands thoroughly.
- If you develop blisters on your skin, don't scratch them—however much they itch. If you break the skin, you'll run the risk of getting an infection.
- As well as painful blistering, some people may have an extreme reaction to the toxin. They should be monitored and treated for shock (see p. 274). Get medical help as soon as possible.

POISONOUS PLANTS

Poison ivy, poison oak, and poison sumac are the most well known of the plants that contain urushiol, an irritant oil. If you damage the plant and get the oil on your skin, you must wash it off immediately with soap and cold water. You should also remove and wash any contaminated clothing to prevent the oil spreading. Apply an anti-poison-ivy cream as soon as possible. Many people will develop an itchy rash and often painful blistering within 4 to 24 hours of contact with the plant (see warning panel, above).

POISON SUMAC

Found in wet acid swamps in eastern North America, poison sumac can grow as tall as 18 ft (6 m).

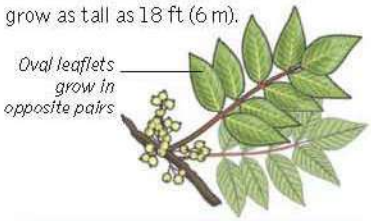

POISON IVY

Native to wooded areas of North America, poison ivy is now found worldwide.

STINGING NETTLES

Found in many countries, stinging nettles usually cause only a temporary stinging sensation where the plant's hairs touch the skin. Apply a soothing cold compress or rub the affected area with a dock leaf if you can find one. Watch for an itchy red rash that indicates an allergic reaction.

POISON OAK

Like poison ivy, the leaflets of poison oak grow in threes. This plant is found in wooded parts of North America.

NATURAL REMEDIES

Where nature causes a problem, she sometimes provides the solution. The following counter the effects of urushiol:

- Jewelweed has a sticky juice that will dry out blisters in a few days. Cut a piece of stem, split it down the middle, and rub the pieces on the affected skin.

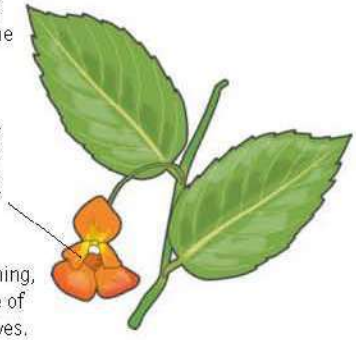

- To soothe itching, make a poultice of witch hazel leaves. Mash up the leaves, adding water if too dry, then apply to the affected area.
- Try washing with a tannic acid solution. Tannic acid is found in tea, or you can make it from oak bark (see opposite).
- Tea tree oil, from the leaves of the melaleuca tree, is also said to counter the effects of urushiol. Apply directly to the affected area.

SWALLOWED POISON

Try to find out what the casualty ingested, how much, and when. If you suspect the casualty has swallowed a caustic chemical, such as camp fuel, do not induce vomiting as it will burn again on the way back up. Get medical help and monitor the casualty. If the lips are burnt, give frequent sips of cool milk or water. If you know the casualty has swallowed a poisonous plant or fungus and he or she is conscious, induce vomiting by tickling the back of the throat. Dilute the poison by getting him or her to drink large quantities of water or milk, or tea mixed with charcoal.

DIARRHEA AND VOMITING

In the wilderness, diarrhea and vomiting can kill, because they lead to dehydration (see p. 272) or even shock (see p. 274). The most likely causes are food poisoning or drinking contaminated water, although infectious diseases are a possibility, so get medical help if symptoms persist. You should rest, keep warm, and replace lost fluids. If you're hungry, you can eat small amounts of bland foods such as pasta for 24 hours.

NATURAL REMEDIES

In a survival situation, there are various natural remedies you can try to stop diarrhea or alleviate symptoms such as stomach pain, although some are more palatable than others:

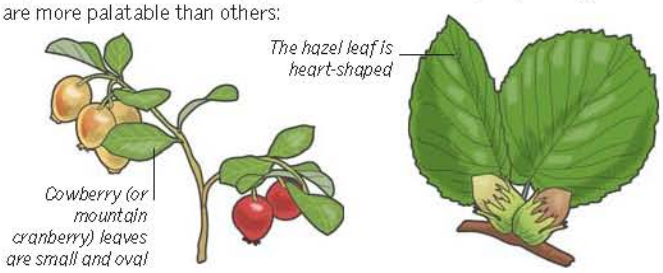

- **Tea:** Drink tea made from hazel, cowberry, or cranberry leaves.
- **Bark:** Pull some bark off a tree (preferably oak), remove the inner bark and boil it for at least 12 hours, adding more water as necessary. The resulting black brew smells and tastes vile, but it contains tannic acid and will cure diarrhea. Drink one cup every two hours.
- **Charcoal**: Take a partially burned piece of wood, scrape off the char, then swallow about a handful with water.
- **Bones:** Burn to ashes then grind, or grind bones into a powder between two rocks. Make a paste with water, then swallow about a tablespoonful.
- **Chalk:** Grind into a powder, mix with water to a paste, then swallow about a tablespoonful.
- **Ash:** Make a paste of wood ash and water, then swallow. This will alleviate stomach pain.

REST AND REHYDRATE

Drink water to maintain your fluid levels. To help replace lost salts, dissolve a packet of rehydration salts in the water, or one teaspoon of salt in a liter of water, before drinking.

INTESTINAL PARASITES

There are two main types of intestinal parasite: helminths (tapeworms, pinworms, and roundworms) and protozoa (giardia, for example). The usual causes of infection are ingesting contaminated water or food, or poor personal hygiene. Symptoms include nausea or vomiting, diarrhea (see left), dysentry, bloating, stomach pain, weight loss, or a rash or itching around the rectum. Seek medical advice as soon as possible.

WILDERNESS CURE FOR WORMS

If you have passed a worm in your stool, swallow a couple of tablespoons of kerosene. While this method may make you sick, it will make the worms sicker. Gasoline will also work but is not as effective.

HOW TO AVOID BEING INFECTED

Prevention is far better than cure and to avoid being infected by parasites present in water or spread in feces you should always do the following in the wilderness:

- Boil all water, or use other reliable methods of purification (see pp. 198–201).
- Don't clean your teeth or rinse your mouth with water that isn't pure.
- Don't swim or stand in rivers or lakes in places where you could be at risk unless absolutely necessary.
- Cover any cuts or wounds on your skin.
- Maintain strict personal hygiene in camp (see pp. 116–17) and when handling food.
- Boil any meat you think might be infected for at least 20 minutes, or ideally until it falls off the bone, before eating it.

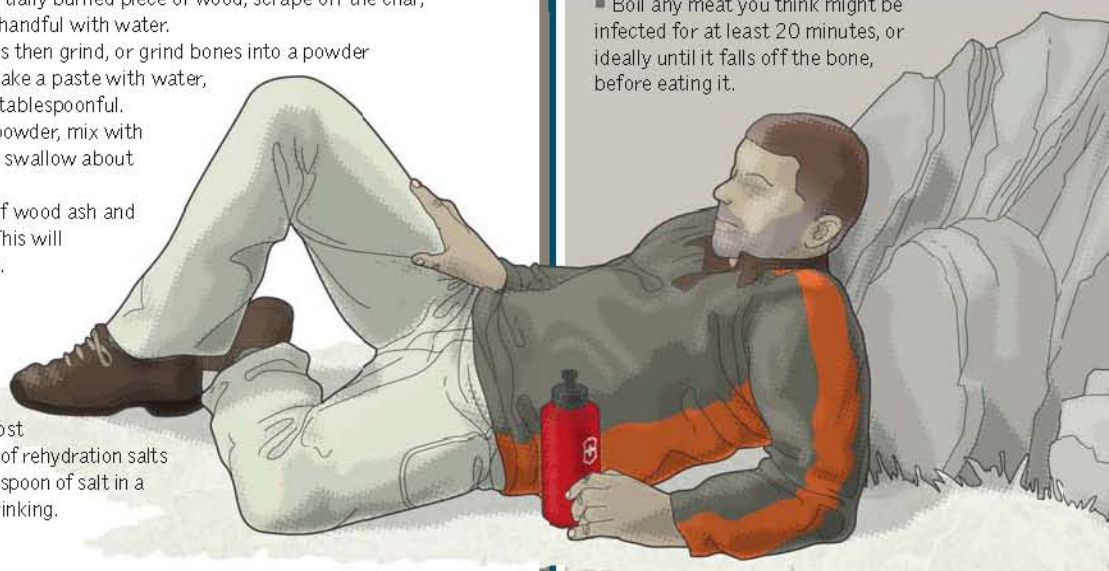

BONES, JOINTS, AND MUSCLES

IT CAN BE DIFFICULT TO TELL whether an injury is a sprain, a broken bone, or a dislocation. The ends of broken bones can move, damaging blood vessels or nerves nearby, so treat the casualty in the position found and immobilize the injured area before letting him or her move. Anyone with a spine injury or broken leg must be carried by stretcher.

CRAMP

This painful muscle spasm can be caused by dehydration and a reduction of body salts through perspiration—so make sure you have enough to drink when exercising. Sit down, rest, and stretch the affected muscles.

FOOT

Help the person stand on his good foot and stretch the muscles to reverse the spasm. Once the cramp has eased, massage the affected area of the foot.

CALF

Sit the person down and support the affected leg. Help him to straighten his leg, and flex his toes to reverse the spasm. Then massage the painful muscle.

THIGH

If the cramp is in the back of the thigh, straighten the leg to stretch the muscle; if it's in the front of the thigh, bend the leg. Once the pain eases, massage the affected area.

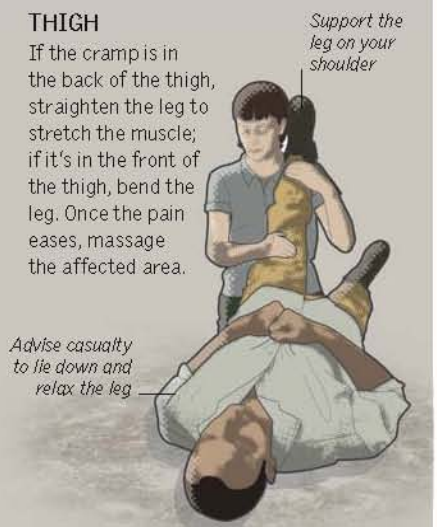

SPRAINS AND STRAINS

A strain is a pulled muscle. A sprain occurs when ligaments that hold a joint together are damaged. The ideal treatment is to raise the injury, cool, and rest it. If you provide comfortable support, gentle movement can help the injury. If in any doubt, treat the injury as a broken bone (see opposite).

1 Rest and support the injury. Wrap a cold compress around it for at least 10 minutes to reduce swelling and bruising.

2 Leave the compress in place, or wrap padding around the injury. Apply a crêpe bandage from below the injury to the next joint.

ARM INJURY

Falling onto an outstretched hand can cause a broken wrist, forearm, upper arm, or collar bone. Support the affected arm in a sling. If a casualty can't bend his arm, he may have injured his elbow, in which case don't use a sling. Instead wrap padding around the joint and secure the arm to the body with triangular bandages. To make sure that the bandage isn't too tight, check the wrist pulse.

USING A SLING

Slide a triangular bandage between the arm and the chest. Bring the front up over the arm and tie a reef knot on the uninjured side.

LEG INJURY

Injuries to the legs can be serious, as any fracture is likely to be unstable, which means that the ends of the bones can move easily, and could pierce one of the large blood vessels in the leg, resulting in severe bleeding. Don't move the casualty unless you have to, and, even then, only when the legs have been immobilized. If you see any signs of shock (see p. 274) ensure the head is low, but do not raise the legs.

HAND INJURY

Injuries to hands are often complicated by bruising or bleeding. Raise the injured hand and treat bleeding with direct pressure (see p. 264). Remove jewelry before the area starts to swell. Wrap the hand in padding and support it in a raised position with a sling.

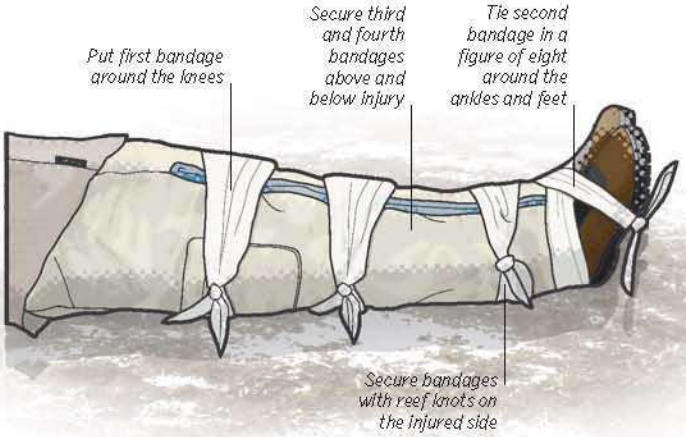

1 Lay the casualty down and support the injury to minimize further damage. Call for emergency help. If this is nearby just maintain this support. You can put rolled-up coats or blankets on either side for extra support.

2 If help is delayed or you need to transport the casualty, put bandages around the knees and ankles (and pelvis, if the thigh is injured), and above and below the injury. Place padding between the legs, then tie the bandages.

SPINAL INJURY

If a person falls and lands on his back, or falls from a height, it is best to assume that he has a spinal and, probably, a head injury. Don't move him—support his head and neck in line with the rest of his back. Moving him could damage the spinal cord, which may result in permanent loss of movement below the injured area. Call for emergency help, or send someone to get help while you stay with the casualty. If you have to move him because he is in danger, use the log-roll technique (see p. 279).

SUPPORT THE HEAD AND NECK

Kneel or lie behind the casualty's head. Keep your arms steady by resting your elbows on your thighs and place your hands on either side of his head to keep it in line with his body. Wait for help.

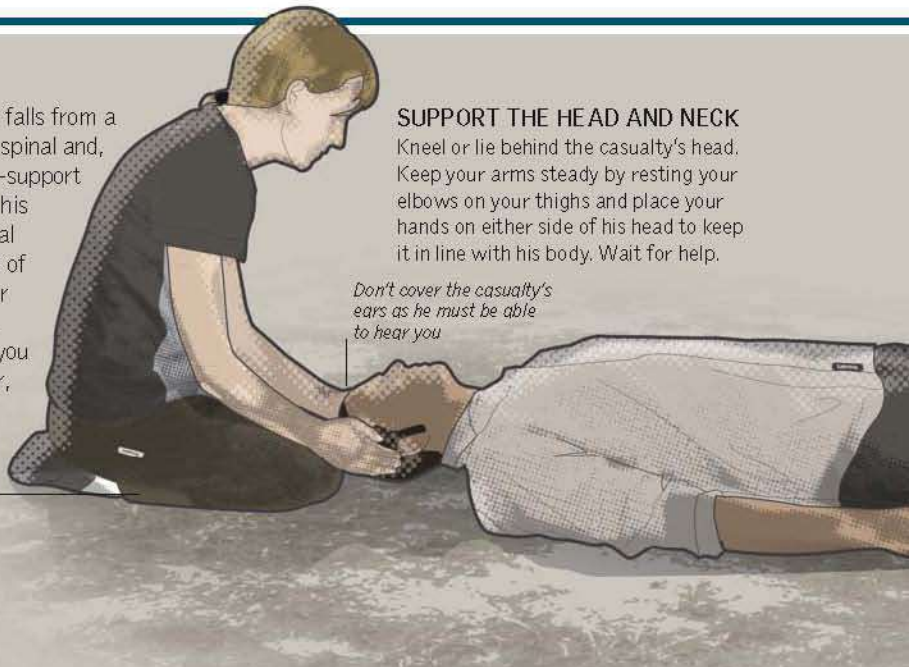

TREATING EXPOSURE

EXTREMES OF HEAT AND COLD can prevent the body's temperature-regulation mechanisms from functioning properly. Both extremes can cause life-threatening injures, so it's vital to act quickly. Don't leave a casualty—call for help or, if possible, send someone to seek help while you treat him or her.

WARNING!
If a casualty becomes unconscious, be prepared to give cardiopulmonary resuscitation (see p. 277).

HEAT INJURIES

In hot weather, wear a hat, reapply sunscreen frequently, and stay in the shade when you can to avoid heat injuries such as sunburn. You'll quickly become dehydrated if you don't drink enough water to replace fluids lost through sweating.

DEHYDRATION

Help the casualty to sit down and give him fluids to drink; water is usually sufficient, but rehydration salts mixed with water is best. If the casualty complains of cramp, help him stretch the affected muscles, then massage them firmly (see p. 270).

HEAT EXHAUSTION

If a casualty feels dizzy and starts to sweat profusely, but has cold, clammy skin, get him into the shade and give him fluids to drink. Help him to lie down then raise his legs—support his feet on a backpack—to help improve blood flow to the brain. Monitor him while he recovers.

HEATSTROKE

This life-threatening condition may follow heat exhaustion or develop with no warning. Heatstroke causes the body's temperature-control mechanism to fail. If a person complains of headache, feels dizzy, has hot, dry skin, and begins losing consciousness, he or she may have heatstroke and will need urgent medical help.

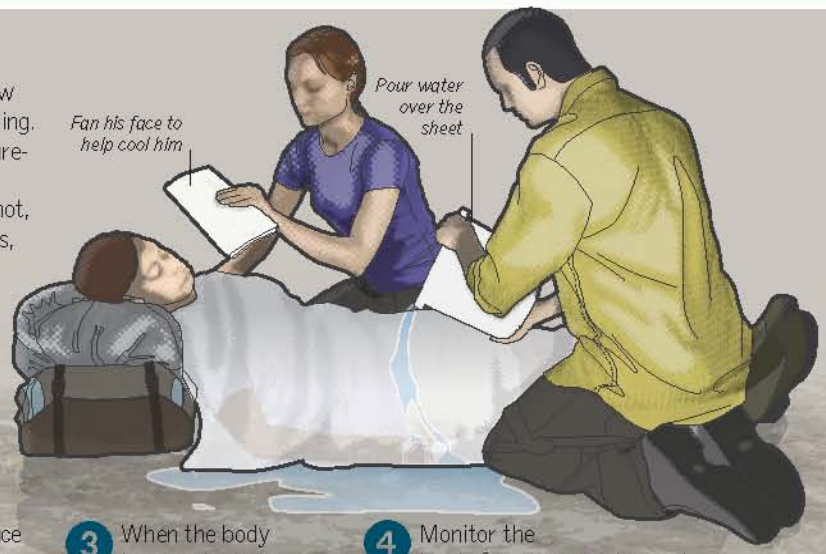

1. Move the casualty to as cool a place as possible—out of the sun. Help him to sit or lie down with his head raised and remove all of his outer clothing.

2. Do anything possible to quickly reduce the body temperature. Ideally, wrap him in a cold, wet sheet. Keep the sheet cool by continually pouring water over it.

3. When the body temperature drops below 104°F (40°C), replace the wet sheet with a dry one.

4. Monitor the level of response, pulse, and breathing. If his temperature starts to rise again, repeat the treatment.

COLD INJURIES

Exposure to cold can result in parts of the body freezing (frostnip and frostbite) or the body's core temperature becoming dangerously low (hypothermia).

FROSTNIP AND FROSTBITE

Frostnip is the freezing of the top layer of skin, usually on the face and extremities. The skin turns numb, white, and hard. Untreated, it can lead to frostbite, which is much more serious—the deeper tissues, and even the bone, freeze. The skin turns white or blue and feels solidly frozen.

WARM GRADUALLY

Frostnip and frostbite can both be treated by warming the affected area, though frostbite requires more intensive treatment. Warm the affected area with body heat—place the casualty's hands in his own armpits, or place his feet in your armpits. Remove any rings and raise the injured part to reduce swelling. Ideally, place the affected area in warm water. Dress the injury in sterile dressings.

HYPOTHERMIA

This life-threatening condition develops if the body's core temperature falls below 95°F (35°C). Treatment aims to prevent further heat loss. It's vital that a casualty is warmed up gradually. If he or she is warmed up too quickly, blood is diverted away from vital organs, such as the heart and brain, to the skin, which can actually speed up cooling of the body. A casualty with hypothermia must be moved on a stretcher.

DROWNING

If a person has been immersed in cold water, there is a high risk of hypothermia. In addition, the cold can cause the heart to stop, or throat spasms can block the airway. Water can also enter the lungs and cause secondary drowning (see p. 255) hours after the person appeared to have recovered.

RESCUING A CASUALTY

If you have rescued a person from water, help him or her to lie down with the head low. Replace wet clothes with dry ones. Treat for hypothermia (see left). If the person is unconscious and not breathing, give five rescue breaths before beginning chest compressions (see p. 277).

ALTITUDE SICKNESS

Symptoms of altitude sickness include nausea, loss of appetite, shortness of breath, and a headache that's not relieved by medication. The casualty may also have difficulty sleeping and will feel unwell. The only solution is to start the descent immediately and remain at a lower altitude for a few days. Severe cases will need to be carried.

1 Help the casualty to a sheltered place where he should rest to prevent his body temperature falling further. Send someone else to get help.

2 If possible, and if there is no risk of further cold, remove any wet clothing and replace with warm, dry clothes—but don't give up your own clothes.

3 Put a thick layer of dry leaves underneath him to insulate him from the ground. Help him to lie down in a sleeping bag and, if you have one, cover him with a survival blanket.

TREATING SHOCK

SHOCK IS A LIFE-THREATENING CONDITION that occurs if the circulatory system fails. The most common cause is severe bleeding, but it can also be a consequence of burns. Initially, there will be a rapid pulse and pale clammy skin. As the condition progresses, breathing becomes rapid and shallow, pulse weakens, and skin becomes pale gray-blue. If untreated, unconsciousness results.

ANAPHYLACTIC SHOCK

Anaphylactic shock is a rare but severe allergic reaction affecting the whole body. People who know they are susceptible to this carry a special adrenaline (epinephrine) autoinjector for use in an emergency. If the casualty has one, but is too weak to use it, take off the safety cap and, holding it in your fist, press it against the casualty's thigh (through clothing if necessary).

DEALING WITH SHOCK

Do not give the casualty anything to eat or drink as an anaesthetic may be needed; moisten his lips with water if he's thirsty. Call for help; the casualty must be transported in the treatment position.

1 Treat the cause of shock, for example bleeding or burns (see p. 265). Suspect shock, too, if you notice any symptoms, yet can't see any obvious injury—it could be caused by internal bleeding.

2 Help the casualty to lie down; insulate him from the ground with blankets or bracken. Raise and support his legs as high as you can above the level of his heart.

3 Loosen tight clothing, for example at the neck, chest, and waist. Keep his head low; this may prevent him from losing consciousness. Keep the casualty warm; cover him with a blanket or sleeping bag if you have one.

4 Shelter the casualty as much as you can, but don't move him unnecessarily. Monitor his level of response, breathing, and pulse while you wait for help. Begin resuscitation if he loses consciousness (see p. 277).

BREATHING DIFFICULTIES

RESPIRATORY PROBLEMS need prompt treatment because they can prevent sufficient oxygen reaching the body tissues. The cause may be temporary, for example choking, suffocation, or smoke inhalation, or it can be a long-term condition such as asthma that requires medication. You may need to get emergency help.

WARNING!
If at any stage the casualty loses consciousness, open the airway (see p. 276) and check breathing; the throat muscles may relax enough to allow breathing. If he or she is not breathing, begin resuscitation as this may dislodge the blockage (see p. 277).

CHOKING

When an object becomes stuck in the throat it can cause a muscular spasm that blocks the airway. Always ask the casualty if she is choking, to make sure. If she can speak, cough, or breathe, the obstruction is mild and she will probably be able to clear it herself.

1. If the casualty is breathing, tell her to continue coughing. If she can't speak or cough, help her to bend forward. Support her upper body and give up to five back blows in between her shoulder blades with the heel of your hand.

2. If back blows fail, stand behind her. Put your arms around her abdomen, clench one fist and grasp it with the other hand. Pull sharply inwards and upward up to five times. Check the mouth and remove any obvious blockage.

3. If the obstruction has still not cleared, repeat the five back slaps and five abdominal thrusts three times, then call for help if it's not already on the way. Continue until help arrives or the casualty becomes unconscious (see pp. 276–77).

ASTHMA

This is a condition in which breathing becomes difficult because the muscles in the air passages go into spasm. Most people who suffer from asthma carry inhalers with them. Many have two inhalers—a brown or white "preventer" inhaler and a blue "reliever" inhaler to use in an attack. If the person does not have any medication, sit him or her down and call for emergency help.

1. Sit the casualty down and advise her to take a puff from her reliever inhaler. Tell her to take slow, deep breaths and to sit as upright as she can. The attack should start to ease in a few minutes.

2. If the attack does not ease, tell her to take another dose from her inhaler and rest while she recovers. If the attack still does not ease, or is becoming worse, emergency help is needed as she could lose consciousness.

WHAT TO DO IN A SURVIVAL SITUATION

The best advice is prevention—know what triggers the asthma and avoid those things if at all possible. However, should someone have an asthma attack and not have an inhaler, try the following:

- Tell her to exhale as completely as she can. This expels the "stale" air, with little or no oxygen in it. It may be difficult—and will seem strange—to blow air out when the basic instinct is to gulp air in, but this does work.
- Then, tell her to inhale, slowly and steadily, and to close her eyes to help calm herself while doing so.

UNCONSCIOUSNESS

IF SOMEONE FALLS UNCONSCIOUS, your priority is to make sure their airway is open so that they can breathe. Call for emergency help immediately (ideally ask someone to do this while you treat the casualty). Don't move the casualty and don't leave him or her alone unless you have to go and get help.

CHECK FOR RESPONSE

Gently shake the casualty's shoulders (if it's a child, tap the shoulders). Talk to him and watch for a response. If he's alert, he's conscious. If, for example, he reacts weakly, he may not be fully conscious—monitor him for any change (deterioration or improvement). If there's no response, he's unconscious.

OPEN THE AIRWAY

If an unconscious casualty is on his back, he's at risk of swallowing his tongue—thereby blocking his air passages. Tilting the head and lifting the chin will "lift" the tongue, clearing the air passage.

CHECK THE BREATHING

Tilt a casualty's head back with one hand and lift the chin with two fingers of your other hand; don't press on the soft tissues under the chin. Keep the airway open, and look, listen, and feel for normal breathing. If the casualty is breathing normally, place him in the recovery position (see below). If he's not breathing, begin chest compressions right away (see opposite).

1 Place a hand on the forehead to tilt the head; lift the chin.

2 Look along the casualty's chest; listen, and feel, for breath against your cheek for no more than ten seconds.

RECOVERY POSITION

If an unconscious casualty is breathing, place him in the recovery position to keep his airway open and clear. Remove anything bulky from his pockets. Kneel beside him. Bend the arm nearest you at a right angle to his body then bring the other arm across his chest until his hand rests against his near cheek, and hold it there. Bend the far leg at the knee, and, still holding the knee, pull the casualty toward you until he is on his side.

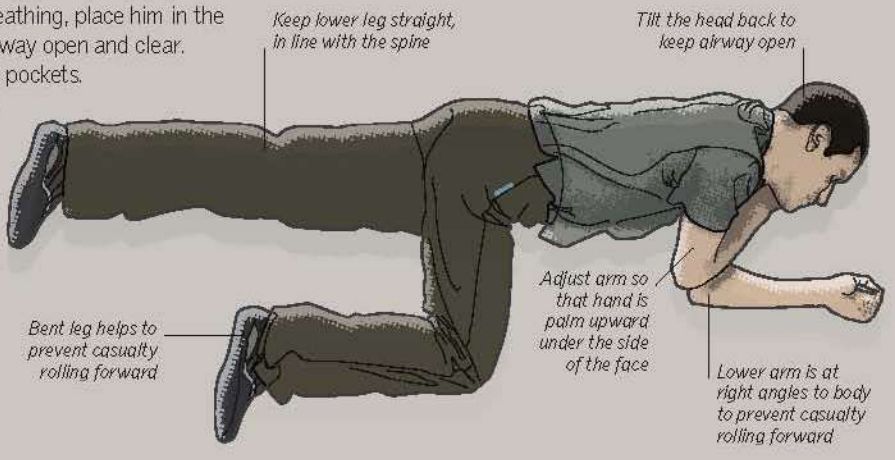

CARDIOPULMONARY RESUSCITATION (CPR)

If a casualty is not breathing, you must try to keep the body supplied with oxygen by using chest compressions and rescue breaths until emergency help arrives. This is known as cardiopulmonary resuscitation, or CPR. If an adult collapses, the cause is most likely to be a heart problem so treat as below. If you have rescued an unconscious casualty from water, start with rescue breaths as for a child (see right). If you are unable to achieve rescue breaths, you can give chest compressions alone.

HOW TO GIVE CPR

Kneel beside the casualty, level with his chest so that you don't have to change position. If you have someone else with you, take it in turns to give CPR so you don't become too exhausted. Change over at the end of each two-minute cycle.

1 Put one hand on the center of the casualty's chest—make sure you don't press on the lower abdomen, the tip of the breastbone, or the ribs.

2 Place the heel of your other hand on top of the first and link your fingers together. Keep your fingers off the casualty's chest.

3 Begin chest compressions. Lean over and, keeping your arms straight, press straight down on the casualty's chest, depressing it by 1½–2 in (4–5 cm). Release the pressure and let the chest come back up, but don't move your hands. Repeat 30 times.

4 Tilt the casualty's head to open the airway and pinch his nose to close the nostrils. Let his mouth fall open slightly. Lift his chin with the fingers of your other hand.

5 To begin rescue breaths, take a normal breath and seal your lips over those of the casualty. Blow into his mouth until you see his chest rise, then lift your mouth away and watch his chest fall. If his chest doesn't rise, adjust his head and try again. Repeat to give a second breath, but don't make more than two attempts at giving rescue breaths before compressing again.

6 Continue the cycle of 30 compressions followed by two rescue breaths until the casualty recovers, help arrives, or you are too exhausted to keep going.

- If at any stage the casualty starts breathing normally, place him in the recovery position (see opposite) and monitor his condition until help arrives.

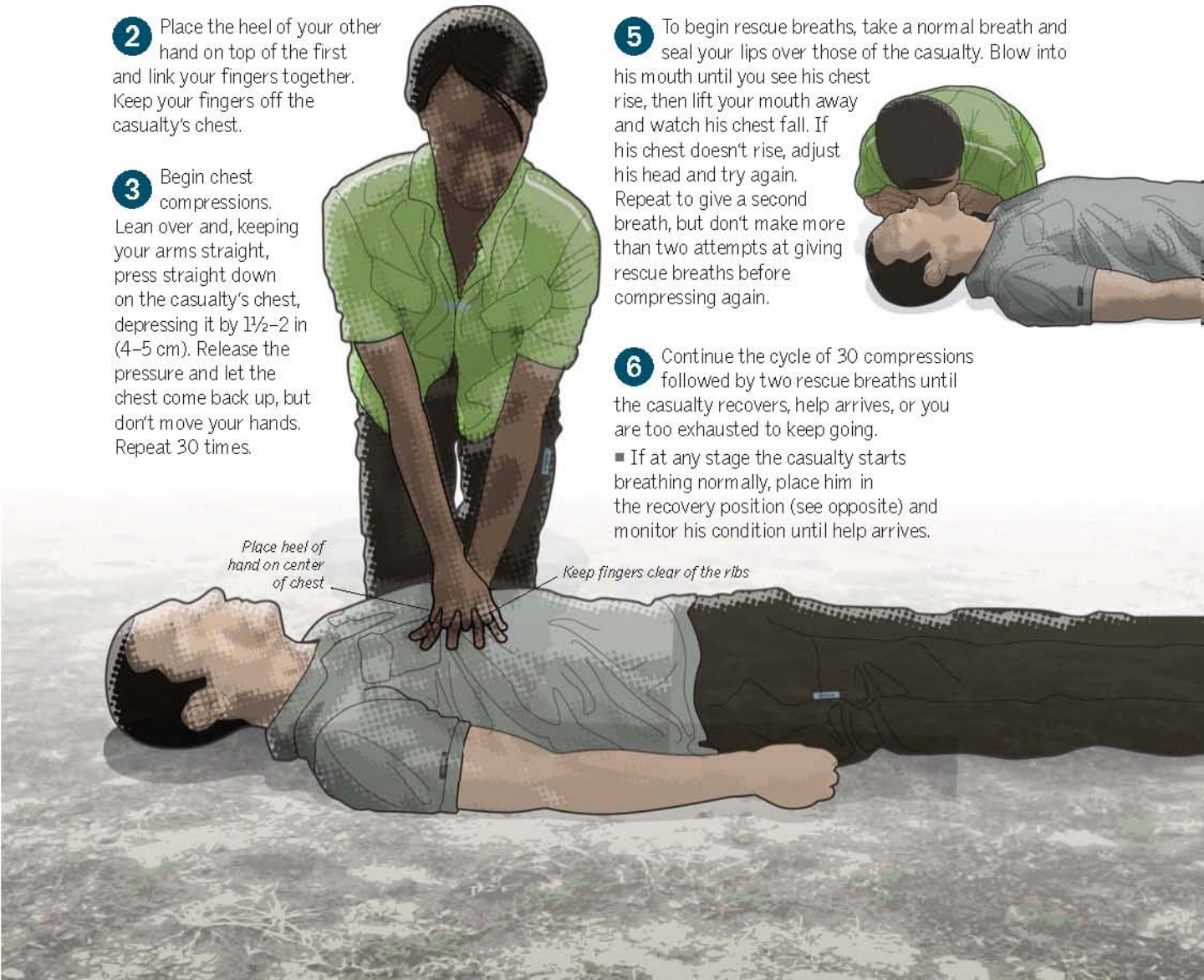

CPR ON A CHILD

- Start by giving the child five rescue breaths.
- Give 30 compressions using the heel of one hand only and slightly less pressure than you would for an adult. Depress the chest by approximately one third of its depth.
- Continue with 30 compressions followed by two rescue breaths until the child recovers, help arrives, or you become exhausted.
- If you are on your own, give CPR for one minute before you try to call for help.

MOVING A CASUALTY

IDEALLY, AN INJURED PERSON should be treated in the position in which he or she was found. Make the casualty as comfortable as possible and wait for help. In a survival situation, if you have to move someone, it's essential to immobilize an injury first to avoid aggravating it.

WARNING!
Move a casualty only if his life is in immediate danger if he is left where you found him. Even then, only move him if by doing so you will not injure yourself. Don't otherwise attempt to move a casualty, as it will make his injuries worse—stay with the casualty and send others for help.

PREPARING FOR A MOVE

Plan your move before you start. Choose a method appropriate to the injury. Don't attempt to move anyone by yourself if you have help. Encourage the casualty to do as much as possible himself.

FIREFIGHTER'S LIFT

This is a technique that can be used to move a conscious casualty if you're on your own and need to carry an injured person a short distance. It gets its name from the firefighters who originally used it. Don't use it if a casualty has head or facial injuries or a broken arm or leg. To prevent injuring your back, use your legs to power the move.

1 Support the casualty's injury with padding and bandages. Help him to stand up.
- Stand at right angles to the casualty, then squat down in front of him, staying as close to him as possible.
- Your shoulder should be level with the top of the casualty's legs.

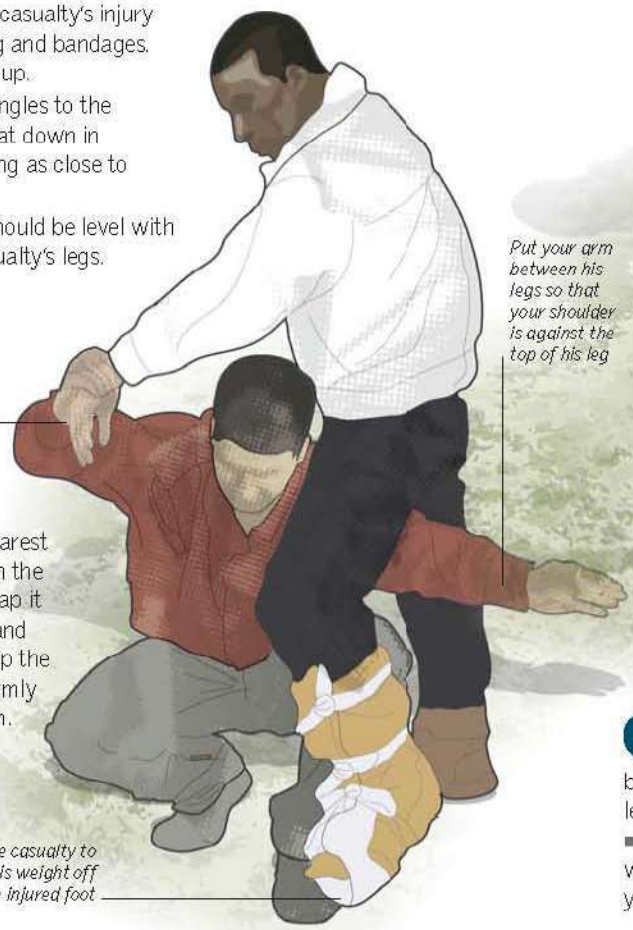

2 Pass your nearest arm between the casualty's legs, wrap it around his thigh, and grasp his leg. Grasp the casualty's wrist firmly with the other arm.

3 With your feet shoulder-width apart to ensure you're well balanced, encourage the casualty to lean across your shoulders.
- Keeping your back straight, hold his wrist firmly, and use the strength of your legs to stand up.

IMPROVISING A STRETCHER

If you need to move a casualty who is unconscious, has a broken leg, or a spine injury, he must be carried by stretcher. Ideally, call the emergency services, and wait for them to arrive. If you can't contact them and you need to get the casualty to medical assistance, you may need to make a stretcher. Use a hurdle or gate, or make the stretcher from poles and coats.

WARNING!
Rescuers should not use their own clothing to make a stretcher if this might put them at risk of exposure.

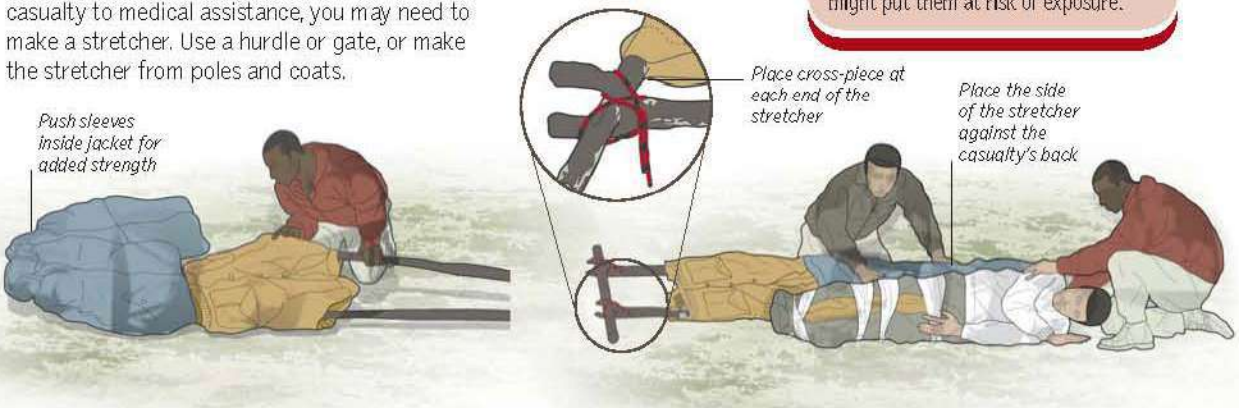

1 Zip or button up two or three jackets. Cut two support poles, strong enough to bear the casualty's weight and 3 ft (1 m) longer than the casualty. Lash a short forked branch across each end (see pp. 166–67) to keep the support poles apart.

2 Immobilize any injury (see p. 271), and roll the casualty onto his uninjured side. One person should support his head while others help to keep his body straight, ideally one for the upper body and one for the legs. Slide the stretcher in place then roll him gently back on to it.

TWO-PERSON SEAT

If there are two of you, this method can be used to carry a conscious casualty who can't walk, but who can support himself with his arms. Nominate one person to be in charge of the move and give the instructions. Keep your backs straight at all times.

1 Stand facing each other behind the casualty. Grasp your left wrist with your right hand, then grasp each other's free wrist to make a seat.

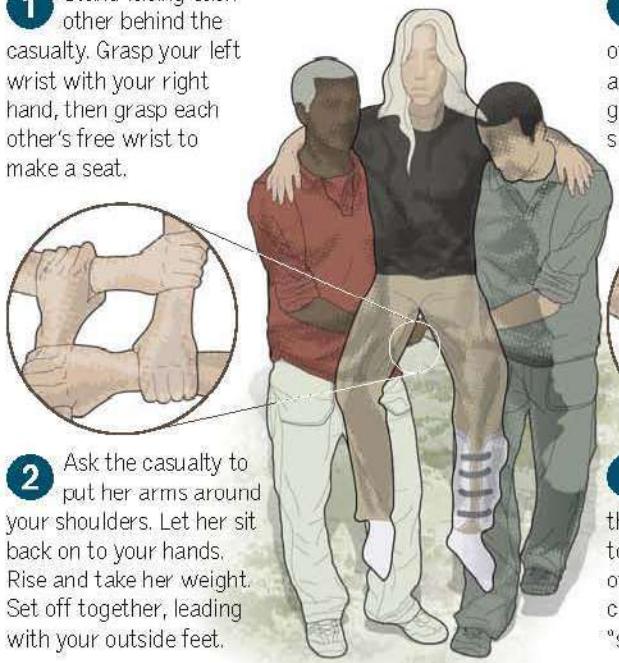

2 Ask the casualty to put her arms around your shoulders. Let her sit back on to your hands. Rise and take her weight. Set off together, leading with your outside feet.

SUPPORTING THE UPPER BODY

If a casualty has an injured arm, the seat carry (see left) can be adapted to support the upper body. As before, one of you should direct the move and always move together. Keep your backs straight as you walk, and stop if it causes you or the casualty any discomfort.

1 Stand facing each other, on either side of the casualty. Put your arms around her back and grasp her clothing on the side farthest from you.

2 Pass your free hands behind the casualty's thighs. Link your fingers together or grasp each other's wrists. Help the casualty back on to the "seat," and lift her.

WILD FOOD

MANY POISONOUS PLANTS in temperate regions resemble edible ones, so you should only eat plants that you can positively identify. If you're in any doubt at all, carry out the Universal Edibility Test (UET) first (see pp. 206–07). Don't forget that some plants are edible only at certain stages of their growth. Never eat wild plants if you have any known

NAME	WHERE	IDENTIFICATION	EDIBLE PART
BISTORT/KNOTWEED *Polygonum* spp.	▪ Damp, grassy habitats, such as meadows, in North America and Eurasia.	▪ Perennial plant, 16–40 in (40–100 cm) tall. Green, triangular leaves. Small pink or white flowers form dense, cylindrical flowerheads at end of long, unbranched stem.	▪ Young shoots and leaves ▪ Roots
BRAMBLE/BLACKBERRY *Rubus fruticosus*	▪ In most temperate habitats worldwide, but particularly scrubland, hedges, and woodland.	▪ Scrambling deciduous shrub with thorny stems forming thickets. Green, toothed leaves. White or deep pink five-petaled flowers. Fruit ripens to blue-black in late summer.	▪ Fruit ▪ Young shoots ▪ Leaves
BUCKWHEAT *Fagopyrum esculentum*	▪ Open grassland worldwide.	▪ Red-stemmed plant up to 2 ft (60 cm) tall. Green, triangular leaves. Small pink, five-petaled flowers grow in clusters. Triangular seeds.	▪ Seeds
DANDELION *Taraxacum* spp.	▪ Widely distributed in temperate areas of Eurasia; introduced to the Americas, Australia, and New Zealand.	▪ Rosette of green, jagged-edged leaves grows close to the ground from a central taproot. Large, bright yellow flower head matures into a spherical "clock" of several seeds.	▪ Leaves ▪ Roots
DOG ROSE *Rosa canina*	▪ In hedgerows, woodland margins, and scrubland in Europe, northwest Africa, and western Asia; introduced to other regions.	▪ Scrambling deciduous shrub with spiny stems. Dark green, toothed leaves. Scentless, white, pink or deep pink five-petaled flowers. Fruit, or "hip," is orange-red and oval.	▪ Hips ▪ Buds and flowers ▪ Young leaves
STINGING/ COMMON NETTLE *Urtica dioica*	▪ By hedgerows and woodland margins in North America, Eurasia, and North Africa.	▪ Perennial herbaceous plant, 20–60 in (50–150 cm) tall. Green, oval, toothed leaves covered in stiff stinging hairs. Small green flowers sometimes have a red tinge.	▪ Young shoots and leaves
SUGAR MAPLE *Acer saccharum*	▪ Woods and forests in northeastern North America.	▪ Deciduous tree, usually 82–115 ft (25–35 m) tall. Gray-brown bark smooth on young trees, furrowed and flaky on old trees. Green, five-lobed leaves turn bright red in fall.	▪ Sap ▪ Inner bark
WALNUT *Juglans* spp.	▪ Temperate areas in Eurasia and North America. Common in the Himalayas.	▪ Large deciduous tree up to 82 ft (25 m) tall with distinctive gnarled bark. Green leaves have narrow leaflets. Nut has a thick, green husk, which rots away once it falls to the ground.	▪ Nuts
WATER CHESTNUT AND WATER CALTROP *Trapa natans* and *T. bicornis*	▪ In slow-moving water in warm temperate areas of Eurasia and Africa; introduced to North America and Australia.	▪ Green, triangular, saw-toothed leaves float on surface. White four-petaled flowers. Fruit is borne underwater and has four sharp spines and one hard gray seed.	▪ Seeds
WILD GRAPE *Vitis* spp.	▪ In most habitats in North America and Eurasia.	▪ Sprawling, high-climbing vine with large, lobed leaves. Fruit grows as hanging bunches of berries, which are amber or dark purple when ripe.	▪ Fruit ▪ Young leaves

TEMPERATE PLANTS

allergies, a pre-existing medical condition, or are pregnant. The UET doesn't apply to mushrooms, so never eat any mushrooms unless you can positively identify every single one as being edible. As part of your preparation for your trip, you should familiarize yourself with the edible plants in the region you're visiting, and their seasonal availability.

PREPARATION	SIMILAR PLANTS	BEWARE OF	EXTRA INFORMATION
▪ Young leaves and shoots can be eaten raw or boiled. ▪ Soak roots, then boil or roast.	▪ Alpine bistort (*P. viviparum*) is common on rocky soils in northern alpine regions and the Arctic.		▪ Roots are twice twisted and are said to resemble two snakes. They contain starch and can be used to make flour.
▪ Fruit and peeled young shoots are edible raw. ▪ Infuse the leaves to make tea.	▪ Wild raspberry (*R. idaeus*) fruit ripens to a bright red. Tea made from the leaves can be used to treat diarrhea.	▪ Catching yourself on thorns.	▪ Blackberries and raspberries are high in vitamin C and sugars. Blackberry leaf tea can be used to treat colds and coughs.
▪ Hull and grind the seeds to make buckwheat flour. ▪ Roast seeds and cook with water to make porridge.		▪ Buckwheat leaves are edible but can cause photosensitivity of the skin if eaten in large quantities.	
▪ Young leaves can be eaten raw; boil old leaves to remove the bitter taste (change the water once). ▪ Boil the roots.			▪ Roots can be roasted and ground to make a substitute for coffee. The leaves are high in vitamins A and C, iron, and calcium.
▪ Buds and flowers are edible raw. ▪ Chew the pulp of raw hips, or dry and eat later. ▪ Boil leaves in water to make tea.		▪ The hips of all wild roses (*Rosa* spp.) are edible, but do not eat the seeds if they are prickly as they will irritate the throat and stomach.	▪ Rose hips remain on the shrub all year. They are rich in vitamin C. Boil crushed hips in water then strain to make a nutritious syrup.
▪ Boil the plants for 10 minutes to destroy the formic acid and histamine in the stinging hairs.	▪ Dead nettles (*Lamium* spp.) are also edible if cooked. They have heart-shaped leaves and white or purple flowers. No stinging hairs.	▪ Gather and handle the plants carefully, wearing gloves if you have them, to avoid being stung by the hairs.	▪ Fresh nettles are rich in protein and vitamin K. The fibers in mature stems can be woven into cordage (see pp. 138–41).
▪ Boil sap into a high-energy syrup. ▪ Eat inner bark raw or boil into a gelatinous mass, then roast and grind into a flour.	▪ Red maple (*A. rubrum*) also produces sweet sap. ▪ White birch (*Betula* spp.) has edible inner bark.		▪ To collect the sap, cut a V-shape into the trunk, make a hole below the cut, and insert a leaf to guide the drips into a container below.
▪ Crack the shell and eat the ripe nut raw.			▪ Rich in fat, protein, and vitamins. ▪ Crushed green husks of black walnut (*J. nigra*) are toxic to fish. Use to poison and catch them.
▪ Seeds can be eaten raw or roasted.			▪ Water chestnuts are a source of carbohydrate.
▪ Ripe fruit is best eaten raw. ▪ Boil young leaves.		▪ Moonseed vines (*Menispermum* spp.) have grapelike fruits, which are poisonous. Unlike edible grapes, they only have one seed.	▪ Ripe grapes are rich in vitamin C and sugar. Water can be obtained from the vine stem (see p. 192).

WILD FOOD

SOME OF THE PLANTS listed below are good sources of water/liquid. In a survival situation in the desert, don't eat food if you don't have any water, as your body will use up its own water to digest it. Only eat plants that you can positively identify and, if you are in any doubt at all, carry out the Universal Edibility Test (UET) first (see pp. 206–07).

NAME	WHERE	IDENTIFICATION
ACACIA/WATTLE *Acacia* spp.	▪ Africa, southern Asia, Australia, and the Americas.	▪ Thorny, medium-sized trees with gray-white bark. Green leaves divided into many small, oval leaflets. Small, ball-shaped yellow, white or pink flowers.
AGAVE *Agave* spp.	▪ Southern USA and Mexico, Central America, the Caribbean, and northeastern South America.	▪ Rosette of fleshy leaves with sharp tip and spiny margin. Central, tall flower stalk.
AMARANTH *Amaranthus* spp.	▪ The Americas, Africa, and Asia.	▪ Tall herb with alternating leaves and erect fleshy stems. Tiny brown/black seeds (one plant produces 40,000–60,000 seeds).
BAOBAB *Adansonia* spp.	▪ Africa and northern Australia.	▪ Large trees with swollen trunks, up to 30 ft (9 m) in diameter. Green leaves have 5–7 digit-like leaflets. Oblong fruit up to 8 in (20 cm) long.
CAROB TREE *Ceratonia siliqua*	▪ Mediterranean, northern Africa, Middle East, and India.	▪ Evergreen tree up to 50 ft (15 m) tall. Shiny green leaves up to 8 in (20 cm) long. Small red flowers. Flat leathery seed pods; dark-green/black when ripe. Hard brown seeds.
CARRION FLOWER *Hoodia* spp.	▪ Southwest Africa.	▪ Numerous thorny, succulent stems up to 6 ft (2 m) tall. Star-shaped or shallowly bell-shaped flowers give off the smell of rotting meat.
DATE PALM *Phoenix dactylifera*	▪ West and North Africa, Middle East, and India; introduced to Mexico and USA.	▪ Tall slender palms with a large crown of green leaves, divided into many narrow leaflets. Fruit reddish-brown when ripe.
PRICKLY PEAR CACTUS *Opuntia* spp.	▪ The Americas; introduced to the Caribbean, the Mediterranean, Africa, and Australia.	▪ Thick, jointed, flat, pad-like, green stems covered in spines. Red or yellow flowers. Fruit red when ripe.
WILD GOURDS *Cucurbitaceae*	▪ Kalahari Desert (Namibia), the Sahara, islands in the Mediterranean, the Middle East, and southeastern India.	▪ Ground-trailing green vine. Bright yellow flowers. Orange-sized green or yellow fruit.

DESERT PLANTS

Don't forget that some plants are edible only at certain stages of their growth. Don't eat wild plants if you have any known allergies, a preexisting medical condition, or are pregnant. When planning your trip, it's advisable to familiarize yourself with the edible plants in the region you are visiting and their seasonal availability.

EDIBLE PART	PREPARATION	BEWARE OF	EXTRA INFORMATION
▪ Seeds (from dark brown pods) ▪ Young leaves and shoots	▪ Roast seeds. ▪ Boil leaves and shoots.	▪ Catching yourself on thorns.	▪ Roots may be tapped for water.
▪ Stalk, before in flower ▪ Buds and flowers	▪ Roast the stalk. ▪ Boil buds and flowers.	▪ Juice from many species can cause acute contact dermatitis, with reddening and blistering lasting 1–2 weeks. Leaves have needle-like ends.	▪ Roasted stalk tastes sweet, like molasses.
▪ Young shoots and leaves ▪ Seeds	▪ Eat shoots and leaves raw or boiled or stir-fried. ▪ Remove chaff from seeds, then cook like popcorn or grind into flour to make bread.	▪ Amaranth leaves are high in oxalic acid, which can irritate the gut and cause kidney stones, so eat in moderation.	▪ Amaranth leaves are a common leaf vegetable in the tropics and warm temperate regions.
▪ Shoots and young leaves ▪ Fruit ▪ Seeds	▪ Eat the fruit pulp raw. ▪ Boil shoots and leaves. ▪ Roast seeds.		▪ Tap the trunk for water. Kalahari bushmen suck water through holes in the trunk using the hollow stems of grasses.
▪ Pods ▪ Seeds	▪ Eat the sweet, nutritious pulp of the pods raw. ▪ Grind the seeds into flour and make into porridge.		
▪ Stems	▪ Tap the stems for their water.		▪ The stems contain an appetite suppressant. Kalahari bushmen chew the stems of several *Hoodia* species before and during their long, arduous hunts.
▪ Fruit ▪ Growing tip of the palm (heart) ▪ Young leaves ▪ Sap	▪ Eat the fruit (dates) raw or sun-dried. ▪ Boil the leaves and heart. ▪ Boil the sap down to a syrup.		▪ Date palms always grow near water. The leaves can be used to thatch a shelter.
▪ Fruit ▪ Seeds ▪ Pads	▪ Peel fruit and eat raw. ▪ Roast and grind seeds to a flour. ▪ Boil or grill young pads, first removing thorns by peeling or scorching over a fire.	▪ Spines. Avoid any similar-looking plant that has milky sap as it may be poisonous.	▪ The non-milky sap from the pads is a good source of water.
▪ Fruit (Tsamma melon only) ▪ Flowers ▪ Seeds ▪ Young leaves and shoots	▪ Pound fruit to a pulp and eat raw. ▪ Eat the flowers raw. ▪ Roast or boil seeds. ▪ Chew leaves and shoots for water.	▪ The fruit of the wild desert gourd or colocynth (*Citrullus colocynthis*) is very bitter and is a strong laxative.	▪ The Tsamma melon (*C. lanatus*) is also known as the wild watermelon. The fruit pulp is a good source of water.

WILD FOOD

A GREAT NUMBER OF PLANTS flourish in the tropics, growing all year round in the warm and humid conditions—those listed below are just a few of the most common varieties. When you plan your trip, you should familiarize yourself with the edible plants in the region you are visiting. In tropical forests, most fruits are borne high in the canopy—out of reach,

NAME	WHERE	IDENTIFICATION	EDIBLE PART
BAMBOO *Bambuseae*	▪ Tropical and subtropical forests.	▪ Tree-like grasses with woody, segmented, hollow stems ranging from black to green to gold in color. Green blade-like leaves. Species range in height from 1⅔–11½ ft (0.5–3.5 m).	▪ Young shoots ▪ Seeds
BANANA AND PLANTAIN *Musa* spp.	▪ Native to Australia and Southeast Asia; introduced to other tropical regions.	▪ Tree-like herbaceous plants up to 33 ft (10 m) tall. Large, green, strap-like, split leaves. Flowers and fruit grow in dense hanging clusters.	▪ Fruit ▪ Buds and flowers ▪ Shoots and young stems ▪ Roots
BRAZIL NUT *Bertholletia excelsa*	▪ Along river banks in rain forests in South America.	▪ Large dry-season deciduous tree up to 150 ft (45 m) tall. Green, oblong, crinkly leaves. Yellow flowers. Fruit is the size of a coconut with a hard, woody shell and contains 8–24 seeds.	▪ Seeds or "nuts"
CEYLON/VINE SPINACH *Basella alba*	▪ Widespread in tropical forests.	▪ Trailing, vine-like plant reaching 100 ft (30 m) in length. Beet-red stems. Fleshy, oval or heart-shaped, green or red-purple leaves.	▪ Young leaves and stems
FIG *Ficus* spp.	▪ Variety of habitats in tropical and subtropical regions.	▪ Evergreen tree with long aerial roots growing from the trunk and branches. Leathery green leaves. Pear-shaped fruit grow directly from the trunk or branches.	▪ Ripe fruit
PAPAYA/PAWPAW *Carica papaya*	▪ Native to tropical rain forests in the Americas; introduced to other tropical and some temperate regions.	▪ Small tree up to 20 ft (6 m) high, with a soft hollow trunk. Large green leaves have seven lobes. Large melon-like fruit turn yellow or orange when ripe.	▪ Fruit ▪ Young flowers, leaves, and stems
PEANUT *Arachis hypogaea*	▪ Native to tropical rain forests in the Americas; introduced to other tropical and some temperate regions.	▪ Small, bushy plant up to 1⅔ ft (50 cm) tall. Paired, oval green leaves grow four to a stem. Yellow flowers. Underground legume (pod) contains 1–4 nuts.	▪ Seeds or "nuts"
SAGO PALM *Metroxylon sagu*	▪ Damp lowlands of tropical rain forests in Southeast Asia; introduced elsewhere.	▪ Spiny-trunked palm up to 33 ft (10 m) tall. Crown of pinnate (feather-like) green leaves.	▪ Pith ▪ Young shoots ▪ Young nuts
WATER LILY *Nymphaea*	▪ Lakes, ponds, and rivers in tropical and subtropical regions worldwide. Also found in temperate regions.	▪ Flat, green, heart-shaped leaves float on surface of water. Large, white, yellow, pink, or blue fragrant flowers.	▪ Tubers ▪ Stems ▪ Seeds

TROPICAL PLANTS

unless you can climb the tree. Only eat plants that you can positively identify. If you are in any doubt at all, carry out the Universal Edibility Test (UET) first (see pp. 206–07). Don't forget that some plants are edible only at certain stages of their growth. Don't eat wild plants if you have a known allergy, a preexisting medical condition, or are pregnant.

PREPARATION	SIMILAR PLANTS	BEWARE OF	EXTRA INFORMATION
▪ Split the tough outer sheath of the shoots, and boil or steam. ▪ Boil the seeds or grind, mix with water, make into cakes, and bake.		▪ Shoots of the giant bamboo or volohosy (*Catharlostachys madagascariensis*) contain prussic acid (cyanide).	▪ Bamboo stems often hold water (see p. 190). The stems can be used for cooking (see p. 204) and building a shelter (see p. 169).
▪ Fruit and flowers are edible raw or cooked. ▪ Cook shoots, stems, and roots.			▪ Banana and plantain fruit is rich in potassium and vitamins A, B6, and C. Hard, unripe plantain fruit is only edible when cooked.
▪ Break fruit capsule open, crack the nut's shell, and eat raw.	▪ The sapucaia tree (*Lecythis* spp.) produces edible seeds in a similar fruit capsule. They are called paradise nuts.		▪ Ripe fruit capsules fall to the ground. Rodents and monkeys are attracted to the nuts, which are rich in fats and selenium.
▪ Steam, boil, stew, or stir-fry the leaves and stems.		▪ Don't eat too much, as the plant can have a mild laxative effect.	▪ Young leaves and stems are rich in vitamins.
▪ Eat the fruit raw or cooked.		▪ Don't eat any hard, woody fruit, or those covered in hairs.	▪ Edible figs are soft when ripe and green, red, or black in color.
▪ Flesh of ripe fruit is edible raw. Unripe fruit is edible cooked. ▪ Flowers, leaves, and stems must be boiled, and the water must be changed at least once.		▪ Don't get the milky juice of green, unripe fruit in your eyes as it will cause temporary blindness.	▪ Ripe fruit is rich in vitamin C. Unripe fruit will ripen quickly if placed in the sun.
▪ Remove the shell and eat the nuts raw.		▪ Don't handle or eat if you are allergic to peanuts.	▪ Peanuts are a good source of protein, B vitamins, and minerals.
▪ Crush the pith, knead it in water, strain into a pot, and let the fine sago settle. Squeeze and let it dry. Mix to a paste with boiling water.	▪ Many palms have edible parts. The sugar palm (*Arenga pinnata*) has an edible sap which can be boiled down to a thick syrup.	▪ Don't eat a palm's fruit unless you have positively identified it as an edible variety as the fruit of some species contain harmful crystals.	▪ A mature palm will yield 330–660 lb (150–300 kg) of sago starch—almost pure carbohydrate.
▪ Peel and slice the tubers, which can be eaten raw. ▪ Stems are best cooked. ▪ Dry and grind the seeds into flour.	▪ All parts of the lotus (*Nelumbo* spp.) are edible raw or cooked.		▪ The tubers are rich in starch. The seeds have a bitter taste but they are edible.

APPENDIX

WILD FOOD

WHEN PREPARING FOR YOUR TRIP, you should familiarize yourself with the edible plants in the region you are visiting and their seasonal availability. Only eat plants that you can positively identify and, if you are in any doubt at all, carry out the Universal Edibility Test (UET) first (see pp. 206–07). Don't forget that some plants are edible only

	NAME	WHERE	IDENTIFICATION
LICHENS	**ICELAND MOSS** *Cetraria islandica*	▪ Mountainous areas in Arctic, sub-Arctic, and cold temperate regions in North America and Europe; lava slopes and plains of Iceland.	▪ Mat-forming lichen, up to 4 in (10 cm) high. Gray-green to pale chestnut branches, rolled into tubes terminating in flattened lobes with fringed edges.
LICHENS	**REINDEER MOSS** *Cladonia rangiferina*	▪ Tundra, bogs, and open woodlands in Arctic, sub-Arctic, and northern temperate regions.	▪ Mat-forming lichen, 2–4 in (5–10 cm) high. Gray, rounded branches resemble antlers.
LICHENS	**ROCK TRIPE** *Umbilicaria* spp.	▪ On rocks in Arctic, sub-Arctic, and northern temperate regions.	▪ Rounded lichens with curling edges, usually gray or brown.
PLANTS	**ARCTIC WILLOW** *Salix arctica*	▪ Tundra in North America, Europe, and Asia; mountainous areas in some northern temperate regions.	▪ Mat-forming shrub, 1–2 ft (30–60 cm) high. Rounded, shiny green leaves. Flowers grow as yellow catkins.
PLANTS	**BEARBERRY** *Arctostaphylos uva-ursi*	▪ Mountainous areas in Arctic and sub-Arctic regions.	▪ Dwarf evergreen shrub. Thick, leathery, club-shaped green leaves. Pink or white flowers. Bright red berries grow in clusters.
PLANTS	**BOG LABRADOR TEA** *Rhododendron groenlandicum*	▪ Bogs and alpine areas in Arctic, sub-Arctic, and temperate regions in North America and Europe.	▪ Evergreen shrub, 1–3 ft (30–90 cm) high. Narrow, leathery, green leaves are hairy underneath and fragrant. Small, fragrant, white flowers grow in clusters and are sticky.
PLANTS	**CLOUDBERRY** *Rubus chamaemorus*	▪ Bogs, marshes, and wet meadows in alpine and sub-Arctic regions of North America, Europe, and Asia.	▪ Perennial shrub, 4–10 in (10–25 cm) high. Soft, green leaves have 5–7 lobes. White flowers have five petals. Raspberry-sized, amber berries grow at the top of the plant.
PLANTS	**CROWBERRY** *Empetrum nigrum*	▪ Tundra, moorlands, bogs, and spruce forests in Arctic, sub-Arctic, and northern temperate regions. Also found in the Andes.	▪ Dwarf, evergreen, mat-forming shrub. Short, pale green, needle-like leaves. Small, purple-red flowers. Black berries.
PLANTS	**KERGUELEN CABBAGE** *Pringlea antiscorbutica*	▪ Rocky areas of sub-Antarctic islands in the Indian and Southern Oceans.	▪ Cabbage-like plant.

PLANTS IN COLD CLIMATES

at certain stages of their growth. Don't eat wild plants if you have any known allergies, a preexisting medical condition, or are pregnant. There are no poisonous types of lichen (althought see rock tripe entry, below), but they must all be soaked in water overnight and boiled well before they can be eaten.

EDIBLE PART	PREPARATION	BEWARE OF	EXTRA INFORMATION
▪ All	▪ Soak for several hours then boil well.	▪ All lichens contain an acid that will cause stomach irritation unless soaked and boiled before eating.	▪ Lichens are low in protein but high in carbohydrates (lichen-starch).
▪ All	▪ Soak for several hours then boil well.	▪ All lichens contain an acid that will cause stomach irritation unless soaked and boiled before eating.	▪ High in vitamins A and B. Important source of food for reindeer (caribou). Partially digested lichens from the reindeer's stomach are considered a delicacy by herders.
▪ All	▪ Soak for several hours then boil well.	▪ There have been some reports of rock tripe poisoning so apply the Universal Edibility Test first.	
▪ Young shoots ▪ Leaves ▪ Young roots	▪ Peel off outer bark and eat inner shoots raw. ▪ Eat the leaves raw. ▪ Peel roots and eat raw.		▪ Arctic willow leaves have 7–10 times more vitamin C than an orange.
▪ Berries	▪ Cook before eating.	▪ Bears feeding on the berries.	▪ Forms mats in Arctic regions.
▪ Leaves	▪ Infuse to make tea.		▪ Pick individual leaves rather than whole branches and harvest from several shrubs. Northern Labrador tea (*R. tomentosum*) leaves also make a fragrant tea. The plant grows in peaty soils and on tundra.
▪ Berries	▪ Edible raw.	▪ Bears feeding on the berries.	▪ Berries initially pink, ripening to amber in fall. High in vitamin C.
▪ Berries	▪ Eat raw or cooked.	▪ Bears feeding on the berries.	▪ Previous year's berries will last on plant until spring if not picked. Fresh berries can be dried for storage. Low in vitamins.
▪ Leaves	▪ Boil well.		▪ Leaves have a bitter taste so must be well-boiled. High in potassium and vitamin C. Antarctic explorers and whalers used to eat Kerguelen cabbage to ward off scurvy, hence its scientific name.

APPENDIX

WILD FOOD

THERE ARE NO POISONOUS SEAWEEDS, but some may cause gastrointestinal upset. The seaweeds listed below are common and safe to eat if gathered when still growing, but should only be eaten in small quantities at first. Don't eat seaweed if you are short of fresh water to drink. Only eat plants that you can positively identify and, if you are in

	NAME	WHERE	IDENTIFICATION
SEAWEEDS	**GUTWEED** *Enteromorpha* spp.	▪ Saltmarshes and rock pools in cold-temperate waters worldwide.	▪ Pale or bright green, tubular, unbranched stems, 8–16 in (20–40 cm) long.
SEAWEEDS	**KELP** *Alaria* spp., *Laminaria* spp., and *Macrocystis* spp.	▪ Rocky coasts of the Atlantic and Pacific oceans.	▪ Very long, strap-like, olive-green to brown fronds. Giant kelp (*M. pyrifera*) is the largest seaweed on Earth, reaching lengths of 150 ft (45 m).
SEAWEEDS	**LAVER** *Porphyra* spp.	▪ Rocky coasts worldwide.	▪ Very thin, irregularly shaped membranous frond, up to 20 in (50 cm) long. Color varies from olive-green to purple-brown or blackish.
SEASHORE PLANTS	**COCONUT PALM** *Cocos nucifera*	▪ Sandy, rocky, and coral coasts in tropics and subtropics worldwide.	▪ Woody perennial tree, up to 72 ft (22 m) tall, with a crown of pinnate (feather-like) leaves. Gray trunk is ringed with growth scars.
SEASHORE PLANTS	**GLASSWORT/ MARSH SAMPHIRE** *Salicornia* spp.	▪ Salt marshes and coastal mudflats of western and eastern North America, western Europe, and the Mediterranean.	▪ Bright green, jointed, fleshy stems, 4–12 in (10–30 cm) high. Scale-like leaves and tiny flowers are sunk into the stems.
SEASHORE PLANTS	**ORACHE/SALTBUSH** *Atriplex* spp.	▪ Sandy and shingle beaches worldwide.	▪ Sprawling plant with spikes of small, greenish flowers. Pale- or silvery-green leaves are either triangular or spear-shaped, sometimes lobed.
SEASHORE PLANTS	**SCREW PINE** *Pandanus* spp.	▪ Tropical coasts from Madagascar to southern Asia and islands in the southwestern Pacific.	▪ Tree up to 30 ft (9 m) tall, supported by stilt-like aerial roots. Strap-like, saw-toothed leaves are grouped in spirals. Large, globular knobbly fruit ripens from green to orange or red.
SEASHORE PLANTS	**SCURVY-GRASS** *Cochlearia* spp.	▪ Salt marshes and rocky coasts of northern North America and northern Europe and Asia.	▪ Creeping plant, 4–16 in (10–40 cm) high, with thick, fleshy, dark green, heart-shaped leaves. Small, white, four-petaled flowers.
SEASHORE PLANTS	**SEA BEET** *Beta vulgaris maritima*	▪ Salt marshes, shingle beaches, and cliffs in Europe.	▪ Sprawling plant up to 40 in (1 m) high, with glossy green, stems and leaves often tinged red. Tiny green flowers grow in clusters on the stems.
SEASHORE PLANTS	**SEAROCKET** *Cakile* spp.	▪ Sandy shores in North America, Europe, Asia, and Australia.	▪ Up to 16 in (40 cm) high with deeply lobed, fleshy, green leaves. Pale lavender, four-petaled flowers.

COASTAL PLANTS

any doubt at all, carry out the Universal Edibility Test (UET) first (see pp. 206–07). Don't eat wild plants if you have any known allergies, a preexisting medical condition, or are pregnant. When planning your trip, it is advisable to familiarize yourself with the edible plants in the region you are visiting and their seasonal availability.

	EDIBLE PART	PREPARATION	BEWARE OF	EXTRA INFORMATION
	▪ All	▪ Eat raw or dry in the sun and grind to a powder, for ease of storage.		▪ Gutweed is a good indicator of where fresh water runs across a seashore as it is particularly abundant in such sites.
	▪ Fronds	▪ Best boiled, though young fronds can be eaten raw.	▪ Large doses of iodine can be harmful.	▪ Sugarwrack (*L. saccharina*) tastes sweet, as its name implies, and can be eaten fried or boiled. Kelp is high in iodine, small amounts of which are needed for good health.
	▪ All	▪ Boil, then mash.		▪ The Welsh make a cake of out of laver and oatmeal, called laverbread. The Japanese produce thousands of tons of paper-thin sheets of dried, processed laver, called nori.
	▪ Seeds (coconuts)	▪ Drink the milk fresh. ▪ Eat the flesh raw or sun-dried.	▪ Mature coconut milk needs extra water to be digested. It's also a laxative, so don't drink too much.	▪ Green, unripe coconuts are a good source of water (see p. 190). Coconut milk is rich in sugar and vitamins and contains protein. The oily flesh is also nutritious.
	▪ Stems	▪ Steam or boil.		▪ Also known as Poor Man's Asparagus. Rock samphire (*Crithmum maritimum*), which is found on cliffs and rocky shores, is unrelated but has edible succulent leaves.
	▪ Leaves	▪ Young leaves can be eaten raw, older leaves should be boiled.		▪ *Atriplex* species are restricted to saline soils. As well as coasts, they are also found inland on the shores of alkaline lakes and in deserts. Gray saltbush (*A. cinerea*) is a type of bushfood in Australia.
	▪ Fruit	▪ Inner flesh of ripe fruit can be eaten raw. Partly ripe fruit can be eaten if baked for two hours.	▪ The leaves have small, thorn-like spines at the edges, which can scrape and irritate the skin.	▪ Unripe, green fruit is inedible.
	▪ Leaves	▪ Eat the leaves raw or pulp and drink.		▪ Scurvy-grass leaves are rich in vitamin C— sailors used to eat (or drink) them to prevent scurvy. They are very bitter, so are best leached in water, but you should eat them in a survival situation.
	▪ Leaves	▪ Eat raw or boiled.		▪ Sea beet is the wild ancestor of beets, sugar beets, and Swiss chard.
	▪ Leaves ▪ Young pods	▪ Leaves and young pods can be eaten raw.		▪ Searocket leaves have a peppery taste. One species of *Cakile* grows in the deserts of the Arabian Peninsula.

WILD FOOD

ALL MAMMALS ARE EDIBLE, but some species or groups of mammals are in danger of extinction and are protected by law; the mammal groups listed below contain many species that are common in their areas of distribution, and are not usually dangerous if approached and handled correctly. However, when preparing for your trip, you must

	NAME	WHERE	IDENTIFICATION
AERIAL	**BATS** *Chiroptera*	▪ Temperate and tropical regions worldwide.	▪ The only mammals with true wings and the ability to fly. Wingspans vary from 6 in (15 cm) to over 5 ft (1.5 m) in the large fruit bats, or flying foxes (*Megachiroptera*).
TREES / GROUND	**TREE SQUIRRELS** *Sciurus*	▪ North and South America, Europe, and temperate parts of Asia.	▪ Small or medium-sized rodents, with large, bushy tails.
TREES / GROUND	**PORCUPINES** *Hystricomorpha*	▪ North and South America, Africa, and tropical Asia.	▪ Large, rounded rodents, with a coat of sharp quills.
GROUND / BURROWS	**ANTELOPE AND DEER** *Bovidae* and *Cervidae*	▪ North and South America, Africa, and Eurasia.	▪ Hoofed mammals, with long legs and barrel-shaped bodies. Male antelope have permanent horns; male deer have antlers, which they shed and regrow each year.
GROUND / BURROWS	**GUINEA PIGS** *Cavia* spp.	▪ Northwestern, central, and southeastern South America.	▪ Small rodent, with dark, coarse fur, short legs, and no tail.
GROUND / BURROWS	**HEDGEHOGS** *Erinaceinae*	▪ Europe, Africa, and Asia; introduced to New Zealand.	▪ Small, short-legged insectivorous mammal, with spines on back and sides of body.
GROUND / BURROWS	**KANGAROOS** *Macropus* spp.	▪ Australia.	▪ The largest of the marsupials (pouched mammals), with long, strong tails, large back legs, and short forelegs.
GROUND / BURROWS	**RABBITS AND HARES** *Leporidae*	▪ Most habitats worldwide, from arctic tundra to semi-desert.	▪ Small or medium-sized herbivorous mammals, with large ears and small, round, furry tails. Fur is usually brown or dark gray; the Arctic hare (*Lepus arcticus*) has a winter coat: white with black ear tips.
GROUND / AQUATIC	**BEAVER** *Castor* spp.	▪ North America and northwestern Eurasia.	▪ Large semi-aquatic rodents, with coarse, brown fur, webbed feet, and a flat, scaly tail.
GROUND / AQUATIC	**CANE RATS** *Thryonomys* spp.	▪ Sub-Saharan Africa.	▪ Large rodents, with coarse, brown fur and a long, naked tail. The greater cane rat (*T. swinderianus*) is semi-aquatic.

MAMMALS

familiarize yourself with those mammals that can be hunted in the region you are visiting and their seasonal movements. Some countries and/or states will require you to have a hunting license or permit. For trapping and preparation methods, see pp. 216–23, and for cooking suggestions, see pp. 204–05.

HOW TO FIND	BEWARE OF	EXTRA INFORMATION
▪ Active at night. Roost in colonies in caves and trees during the day. Temperate species hibernate in the winter.	▪ Sharp teeth. Some bats carry rabies or other diseases (see pp. 300–01), so cook the meat thoroughly. Bats are protected in Europe and some parts of North America.	▪ Large, plump, fruit-eating bats, such as the Indian flying-fox (*Pteropus giganteus*), are considered a delicacy in many tropical areas.
▪ Mostly active by day, feeding in branches and on the ground on shoots, nuts, and birds' eggs. Nest in trees in a range of woodland habitats.	▪ Sharp teeth and claws.	▪ Flying squirrels (*Pteromyini*) are nocturnal, tree-living members of the squirrel (*Sciuridae*) family. They glide from tree to tree rather than fly.
▪ Some New World porcupines climb trees to feed, but all Old World porcupines spend all their time on the ground across a range of habitats.	▪ Barbed quills (see p. 266 for how to remove).	▪ Slow-moving, so can easily be run down and speared. Unlike most small game, the meat is fatty rather than lean.
▪ Antelope are mainly found in savanna, or marshland habitats, and deer mainly in woodland, forests, or tundra.	▪ Horns or antlers.	▪ Most active at dawn and dusk. Usually live in groups.
▪ Feed mostly at dawn and dusk, in shrubby grassland in mountainous regions.	▪ Sharp teeth.	▪ Have communal feeding runways. Traps can be baited with leafy vegetables.
▪ Habitats range from woodland, hedgerows, and grassland to desert. Feed at night on small animals, such as worms and insects.	▪ Sharp spines. Usually infested with parasites, so handle carefully and cook well.	▪ Traditionally prepared for eating by covering with clay and baking in embers of a fire (see p. 205).
▪ Live in open savanna woodland, feeding mainly at night on vegetation. Gather at waterholes in times of drought.	▪ Sharp claws and a powerful kick.	▪ Kangaroos are protected in some Australian states, but can be hunted (if you have a permit) in others.
▪ Rabbits live in burrows, often in large numbers, coming above ground to feed on vegetation, using well-worn runs. Hares live above ground.	▪ Rabbits and hares may be infected with germs, so handle them carefully and cook well. Their meat is very lean, and must be eaten with green vegetables (see p. 220).	▪ In most areas, these are the first animals to try to trap. The European rabbit (*Oryctolagus cuniculus*) is an invasive pest species in Australia.
▪ Look for the distinctive mud-and-stick lodges in the middle of a pond or lake. Beavers leave the lodge at night to feed on water plants and trees along the riverbank.	▪ Strong, chisel-like teeth. The Eurasian beaver (*Castor fiber*) is protected in many of the countries in which it is found.	▪ Use regular runs along streams. The tail can be eaten as well as the flesh.
▪ Feed at night on reeds and grasses in marshland and on riverbanks, or on grasses in moist savanna, and on rocky hillsides.	▪ Sharp teeth.	▪ Cane rats are a valuable source of bush meat in West and Central Africa. "True" rats (*Rattus* spp.) are edible but often carry diseases, so handle with care.

WILD FOOD

ALL BIRDS ARE EDIBLE, although a few taste horrible (birds-of-paradise, for example), and pitohuis have poisonous feathers and skin. Some species or groups of birds are in danger of extinction and are protected by law; the bird groups listed below contain many species that are common in their areas of distribution. When planning your trip,

NAMES	WHERE	IDENTIFICATION
GAMEBIRDS *Galliformes*	▪ Worldwide, except Antarctica.	▪ Typically have plump bodies, small heads, and short, rounded wings. Range in size from quail, which can weigh as little as 2 oz (50 g), to the wild turkey (*Meleagris gallopavo*), which can be as heavy as 22 lb (10 kg).
OSTRICH *Struthio camelus*	▪ From West to East Africa south of the Sahara, and southern Africa.	▪ The world's tallest and heaviest bird, reaching 220 lb (100 kg), the ostrich is flightless. Long neck and long, strong legs, with two-toed feet. Males have black and white plumage, females brown.
OWLS *Strigiformes*	▪ Every continent, except Antarctica.	▪ Upright posture and a flattened face with large, forward-facing eyes. Strong, hooked bill and sharp talons (claws).
PIGEONS AND DOVES *Columbiformes*	▪ Worldwide, except the polar regions.	▪ Plump, full-breasted bodies, with a small head and bill. Thick, soft plumage, which can be brown or gray, or brightly colored in some tropical species.
STORKS AND HERONS *Ciconiiformes*	▪ Worldwide, except the polar regions.	▪ All have large bodies, with long legs, long necks, and bills. Range height from 10 in (25 cm) for the smallest bitterns to 5 ft (1.5 m) for the largest storks.
VULTURES *Accipitridae* and *Cathartidae*	▪ All continents, except Antarctica and Australasia.	▪ Large carrion-eating birds, with large wings, powerful feet with sharp talons, and a strongly hooked bill. Head and neck is often bald or sparsely feathered.
WADERS AND GULLS *Charadriiformes*	▪ Worldwide.	▪ Vary greatly in size and shape, but most have subdued brown, gray, black, or white plumage. Waders (or shorebirds) usually have lightweight bodies and long legs. Gulls have compact bodies.
WATERFOWL *Anseriformes*	▪ Worldwide, except Antarctica.	▪ Typically have plump bodies, powerful wings, and short legs with webbed feet. Most species have a broad, flattened bill. Ducks usually have short necks, while geese and swans have long necks.

BIRDS

you must familiarize yourself with those birds that can be hunted in the region you are visiting and their seasonal availability. Some countries and/or states will require you to have a hunting license or permit. For trapping and preparation methods, see pp. 226–29; for cooking suggestions, see pp. 204–05.

	HOW TO FIND	BEWARE OF	EXTRA INFORMATION
	▪ Wide range of terrestrial habitats, from high mountains, to tropical forests, to Arctic tundra.	▪ Short but sharp beak and claws. In some species (including pheasants), the males have sharp spurs on their ankles.	▪ Nearly all gamebirds feed and nest on the ground. At night, many gamebirds roost in trees.
	▪ Open semi-arid plains (from desert to savanna) and open woodland.	▪ Can run fast and deliver a powerful kick. Very protective of eggs in the nest and the young birds.	▪ The eggs are the largest of any bird, weighing 3 lb (1.4 kg). The nests are communal and contain up to 40 eggs.
	▪ In most habitats, from tundra to dense forest.	▪ Sharp bill and talons.	▪ Most species hunt at night and roost in trees during the day. Usually nest in holes in trees, under rocky overhangs, or, sometimes, in burrows.
	▪ Variety of forest habitats, from open woodland to dense, tropical rain forest, and grassland and semi-arid areas.		▪ Often feed in flocks on the ground or in trees, and usually roost communally in trees. Can be taken by hand from the roost if you approach slowly and quietly.
	▪ Freshwater habitats, ranging from wetlands, rivers, lakes, marshes, swamps, mangroves, and lagoons to tidal mudflats.	▪ Sharp bill.	▪ Usually feed alone at the water's edge, but often nest communally in trees.
	▪ Open areas, including mountains, plains, deserts, and savannas.	▪ Sharp bill and talons. Handle as little as possible, as prone to parasites and infection. Boil meat for at least 30 minutes.	▪ Carrion eaters and scavengers, such as crows (*Corvus* spp.) and gulls (see below), can be lured into traps baited with meat.
	▪ Tundra and a variety of wetland and coastal habitats, including tidal mudflats, beaches, and cliffs.	▪ Sharp bill. Gulls and terns will aggressively defend their nest sites.	▪ Most birds in this group nest on the ground or on rock ledges, laying about four well-camouflaged eggs. Often nest in large colonies.
	▪ Arctic tundra, wetlands, rivers, and lakes. Ducks and swans usually feed on the water, while geese graze on the land.	▪ Geese and swans can be very aggressive, particularly in the breeding season. Mute swans (*Cygnus olor*) can weigh up to 26 lb (12 kg).	▪ Many species migrate between their breeding grounds in the Arctic and their wintering grounds further south. They molt their feathers in the late summer, rendering them flightless for a couple of weeks and thus easier to catch.

WILD FOOD

SOME AMPHIBIAN AND REPTILE SPECIES are at risk of extinction and are protected by law; the animals listed below are common in their areas of distribution, and are non-venomous. As part of preparation for your trip, it is advisable to familiarize yourself with the reptiles and amphibians in the region you are visiting. Many countries and/or states

	NAME	WHERE	IDENTIFICATION
FROGS	**AFRICAN CLAWED FROG** *Xenopus* spp.	▪ Sub-Saharan Africa; introduced to North America, South America, and Europe.	▪ Flattened, brown body, 2¼–5 in (6–13 cm) long, with a line of white "stitch marks" along the sides. Clawed toes on the muscular legs.
FROGS	**AMERICAN BULLFROG** *Rana catesbeiana*	▪ North America; introduced to Europe and Asia.	▪ Body is green with brown markings above and white below, 3½–8 in (9–20 cm) long. Large legs and large eardrums.
FROGS	**SOUTH AMERICAN BULLFROG** *Leptodactylus pentadactylus*	▪ Central and northern South America.	▪ Smooth yellow or pale brown body with some dark markings, 3–9 in (8–22 cm) long.
TURTLES	**AFRICAN HELMETED TURTLE** *Pelomedusa subrufa*	▪ Sub-Saharan Africa.	▪ Flattened, brown shell, 8–12½ in (20–32 cm) long.
TURTLES	**ASIAN LEAF TURTLE** *Cyclemus dentata*	▪ Southeast Asia.	▪ Oval, light to dark brown shell, 6–9½ in (15–24 cm) long, with serrated edge near tail. Reddish brown head and legs.
TURTLES	**PAINTED TURTLE** *Chrysemys picta*	▪ Southern Canada, USA, and northern Mexico.	▪ Flattened, smooth shell, 6–10 in (15–25 cm) long; brown above and yellow below, sometimes patterned. Yellow or red stripes on the neck.
SNAKES	**CARPET PYTHON** *Morelia spilota*	▪ Indonesia, New Guinea, and Australia.	▪ Several subspecies, all of which have a bold pattern of irregular markings, which can be reddish brown, brown, black or gray. Average length is 6½ ft (2 m), but can reach 13 ft (4 m).
SNAKES	**COMMON BOA** *Boa constrictor*	▪ Central America, South America, and some Caribbean islands.	▪ Several subspecies but all have characteristic dark saddle markings along the back and a dark stripe behind each eye. Narrow head and pointed snout.
SNAKES	**COMMON EGG-EATING SNAKE** *Dasypeltis scabra*	▪ Sub-saharan Africa	▪ Reddish-brown or gray body with dark, angular markings. A slender snake, 28–39 in (70–100 cm) long. Rounded snout.
SNAKES	**COMMON RATSNAKE** *Elaphe obsoleta*	▪ Southern Canada, and central and eastern USA	▪ Subspecies coloration varies from bright yellow-orange to pale gray with darker blotches. Length range is 4–6 ft (1.2–1.8 m). Long head with rounded snout.
LIZARDS	**COMMON/VIVIPAROUS LIZARD** *Zootoca vivipara*	▪ Europe, extending to the Arctic Circle, and central to eastern Asia, including Japan.	▪ Body is usually brown olive, sometimes black, with males having bright yellow or orange bellies and females creamy white ones. Length range is 4–7 in (10–18 cm).
LIZARDS	**GREEN IGUANA** *Iguana iguana*	▪ Central America and northern South America.	▪ Green or grayish body, 3¼–6½ ft (1–2 m) long, including the long striped tail. Long, stout legs for climbing trees. Adults have a fleshy dewlap beneath the throat, which is large in males.
LIZARDS	**RAINBOW LIZARD** *Agama agama*	▪ West, Central, and East Africa.	▪ Males turn brightly colored in the sun, with an orange-red head and a blue or turquoise body; females and juvenile males remain gray. Length range is 12–16 in (30–40 cm).
	WESTERN FENCE LIZARD *Sceloporus occidentalis*	▪ Southwestern USA and northwestern Mexico.	▪ Brown body with raised, pointed scales. Blue patches on belly are most pronounced in males. Length range is 6–9 in (15–23 cm).

AMPHIBIANS AND REPTILES

will require you to have a hunting license or permit. See p. 218 for some trapping methods and p. 224 for how to prepare the animal for cooking. Avoid brightly colored tropical frogs as they are often highly toxic. In addition, don't eat box turtles, as they sometimes eat poisonous fungi and their flesh may be toxic.

	HOW TO FIND	SIMILAR SPECIES	EXTRA INFORMATION
	▪ Wetlands and waterholes in savannas and tropical and subtropical forests.		▪ Also known as the African clawed toad.
	▪ Wetlands.	▪ All frogs in the *Rana* genus are edible. The genus includes New World true frogs and the European common frog (*R. temporaria*).	▪ American bullfrog legs can be the size of chicken drumsticks. All temperate species of frog hibernate in the winter.
	▪ Tropical forests and rain forests.		▪ Active at night. If picked up, the South American bullfrog emits a loud scream to startle the predator into dropping it.
	▪ Watering holes and rain pools in open country.		▪ In the rainy season, this turtle wanders from pool to pool to forage. In the dry season, it buries itself in mud.
	▪ Shallow streams in mountains or lowlands.		▪ Very active on land and in water.
	▪ Lakes, ponds, and slow-moving streams and rivers.		▪ Active by day. You will often find several painted turtles piled up together on a log in the water, basking in the sun.
	▪ Wide range of habitats.		▪ The carpet python is non-venomous, but it can inflict a nasty bite. It is one of Australia's most widespread snakes.
	▪ Wide range of habitats, from tropical forest to dry savanna.	▪ There are several species in the boa family, which includes the world's largest snake, the green anaconda (*Eunectes murinus*).	▪ Kills its prey by coiling around it and suffocating it. Don't tackle large common boas—they can be up to 13 ft (4 m) long.
	▪ Desert and open habitats.		▪ This nocturnal African snake feeds on eggs during the bird-breeding season then fasts for the rest of the year. It has no teeth.
	▪ Rocky hillsides with open woodland.	▪ Ratsnakes (*Elaphe* spp.), which include the brightly colored corn snake (*E. guttata*), are found in both the New and Old Worlds.	▪ Like the common boa (see above), all ratsnakes kill their prey by constriction.
	▪ Wide range of terrestrial habitats.		▪ There's no point trying to catch a common lizard by its tail as, like many lizards, it can shed the tail and regrow it.
	▪ Tropical forest and rain forest.		▪ Defends itself by lashing out with its tail and claws.
	▪ Open habitats.		▪ Rainbow lizard can shed and regrow its tail.
	▪ On rocks and other prominent places in temperate and coniferous woodlands.	▪ The eastern fence lizard (*S. undulatus*) is widely distributed across southeastern USA and northeastern Mexico.	▪ Fence lizard can shed and regrow its tail.

WILD FOOD

WHEN PLANNING YOUR TRIP, it is advisable to familiarize yourself with the species of fish in the region you are visiting, their seasonal availability, and the best method of catching them (see pp. 208–09 for how to make tackle and techniques to use in a survival situation). Many countries and/or states will require you to have a fishing

NAME	WHERE	IDENTIFICATION
BARRAMUNDI *Lates calcarifer*	▪ Western and eastern Indian Ocean, and northwestern and western Pacific Ocean.	▪ Up to 6½ft (2 m) long, with a rounded tail fin. Dark greenish-gray upper body fading to silver below.
COMMON/BRONZE BREAM *Abramis brama*	▪ Europe to Central Asia.	▪ Up to 2¾ft (82 cm) long, with a deep, narrow, bronze-colored body and a deeply forked tail.
COMMON/KING CARP *Cyprinus carpio*	▪ Europe and Asia; introduced in North America and Australasia.	▪ Up to 5 ft (1.5 m) long, with a dark gold upper body paling to silver below.
EUROPEAN TURBOT *Psetta maxima*	▪ Northeast Atlantic Ocean up to the Arctic Circle and the Mediterranean.	▪ Almost circular flatfish, up to 3¼ft (1 m) in diameter. Sandy brown with brown or black speckles on top.
NILE PERCH *Lates niloticus*	▪ Northern, central, and eastern Africa.	▪ Up to 6¼ft (1.9 m) long, with a rounded tail fin. Dark gray-blue upper body fading to silver below.
RUDD *Scardinius erythrophthalmus*	▪ Europe and Asia.	▪ Up to 17 in (45 cm) long, with a deeply forked tail. Dark greenish upper body paling to silver below. Red pelvic, pectoral, and anal fins.
ATLANTIC SALMON *Salmo salar*	▪ North Atlantic and Arctic oceans, and Baltic Sea and adjacent rivers; introduced to Argentina and Australasia.	▪ Up to 5 ft (1.5 m) long, with a powerful, streamlined, silver blue-green body.
TARPON *Megalops atlanticus*	▪ Eastern and western Atlantic Ocean, the Gulf of Mexico, and the Caribbean.	▪ Up to 8 ft (2.5 m) long, with a deeply forked tail. Bright silver body, with large, hard scales.
TENCH *Tinca tinca*	▪ Europe and Asia; introduced in North America.	▪ Olive green upper body, golden below, with a square tail fin. Small barbel at corners of mouth. Average length is 27 in (70 cm).
BROWN TROUT AND SEA TROUT *Salmo trutta*	▪ Temperate waters worldwide.	▪ Streamlined body with an average length of 3¼ft (1 m). The brown trout has a brownish body flecked with distinctive black and red spots. The sea-going form, the sea trout, is silver-blue with black spots.

FISH

license or permit, and some species of fish can only be caught at certain times of the year. Before fishing, you should also check with the locals that fish in the area are safe to eat—in some regions the waters may be contaminated. See pp. 212–13 for how to prepare a fish for cooking.

	HOW TO FIND	SIMILAR FISH	EXTRA INFORMATION
	• Tropical, slow-moving creeks and estuaries, hiding in mangrove roots and rocky outcrops.		• Beware of the sharp dorsal fin.
	• Bottom of still and slow-moving waters, typically in lakes, rivers and ponds. Swim in large shoals.	• Silver bream (*Blicca bjoerkna*) and white-eye bream (*Abramis sapa*).	
	• Still and slow-moving waters, often among vegetation near banks of lakes or ponds.	• Wild carp, mirrored carp, and leather carp are all varieties of the common carp.	
	• Sandy, rocky, or mixed sea beds in shallow coastal and brackish water.	• Other flatfish (*Pleuronectiformes*) found in shallow water, such as sole, plaice, and flounder.	
	• Lakes and large rivers.	• Other perch (*Perciformes*).	• Beware of the sharp dorsal fin and teeth (the Nile perch is a voracious predator). This species weighs up to 440 lb (200 kg).
	• Still and slow-moving waters, often near banks of ponds and marshlands.	• Roach (*Rutilus rutilus*)	
	• Cold, fast-flowing rivers and coastal waters.	• Pacific salmon (*Oncorhynchus* spp.)	• The salmon run (when the fish migrate from the sea to their birth rivers) is in the autumn, with spawning between November and January.
	• Estuaries, lagoons, tidal flats and mangrove swamps. Feed in shoals.	• Indo-Pacific tarpon (*M. cyprinoides*), which is smaller and rarer.	• This species is hard to hook because of its bony mouth. It weighs up to 160kg (350lb).
	• Still and slow-moving waters, in ponds, lakes and rivers. The best time of day to catch tench is at dawn, at the edges of dense vegetation.		
	• Streams, rivers, lakes, and coastal waters.	• Rainbow trout and steelhead trout (*Oncorhynchus mykiss*) and the cutthroat trout (*O. clarki*).	• Sea-going forms migrate from the sea to rivers to spawn in the autumn. Beware of small bones when eating.

WILD FOOD

SOME SPECIES OF INVERTEBRATE are in danger of extinction and are protected by law; the animals listed below are common in their areas of distribution. When preparing for your trip, it is advisable to familiarize yourself with the edible invertebrates in the region you are visiting and the best way to catch them (see p. 219 for some techniques to use in a survival

	NAME	WHERE	IDENTIFICATION
MOLLUSCS	**COMMON/BLUE MUSSEL** *Mytilus edulis*	▪ Intertidal zones of estuaries and coasts of North and southeastern Atlantic and northeastern and southwestern Pacific.	▪ Dark-shelled bivalve, reaching 4–6 in (10–15 cm) in length.
MOLLUSCS	**GIANT AFRICAN LAND SNAILS** *Achatina* spp.	▪ Sub-tropical and tropical parts of Africa; introduced to the Asia-Pacific region.	▪ Large terrestrial snails, reaching up to 12 in (30 cm) in length when body is extended, and 4 in (10 cm) in diameter. Whorled, conical, brownish shell with darker bands.
ARTHROPODS	**GRASSHOPPERS** *Acrididae*	▪ Worldwide, on vegetation and the ground.	▪ Winged insects with powerful hindlegs, ⅜–3¼ in (1–8 cm) long, usually with camouflage coloring and patterning.
ARTHROPODS	**HONEYPOT/HONEY ANTS** *Myrmecocystus* spp.	▪ Semi-arid and arid regions of southwest USA, Mexico, Africa, and Australia.	▪ Segmented bodies with six legs and a pair of antennae, and a constricted waist. Range in colour and size, from 1/50–½in (2–12mm). Special members of the colony, called repletes, have abdomens that can swell to the size of a grape.
ARTHROPODS	**WITCHETTY GRUB** *Endoxyla leucomochla*	▪ Underground, in the roots of the witchetty bush (*Acacia kempeana*), central Australia.	▪ The larva of this species of cossid moth reaches about 2¾ in (7 cm) in length. It is white with a brown head.
CRUSTACEANS	**CRAYFISH/CRAWFISH/ YABBIES/KOURAS** *Astacidea*	▪ Worldwide, in freshwater streams.	▪ Segmented bodies, which can be sandy yellow, green, dark brown, or blue-gray in color. Ten legs, the front two of which are large claws. Average length is 4 in (8 cm), but some species grow much larger.
CRUSTACEANS	**WOODLICE** *Oniscidea* and *Armadillidiidae*	▪ Worldwide, in damp terrestrial microhabitats, such as in rotting wood and leaf litter.	▪ Flattened, segmented bodies up to ¾ in (2 cm) in length. Gray or light brown or black in color. Pill millipedes will roll up into a ball when threatened.
WORMS	**COMMON EARTHWORM** *Lumbricus terrestris*	▪ Temperate regions of Europe; introduced to most parts of the world.	▪ Reddish worm, reaching 14 in (35 cm) in length when extended.

INVERTEBRATES

situation). Many countries and/or states will require you to have a hunting license or permit, and some species of shellfish can only be caught at certain times of the year. Also check with the locals that the shellfish are safe to eat—in some regions the waters may be contaminated. See pp. 224–25 for some preparation and cooking methods.

SIMILAR INVERTEBRATES	BEWARE OF	EXTRA INFORMATION
▪ Most species of marine and freshwater bivalves and univalves are edible if alive and healthy when gathered.	▪ Don't collect marine shellfish that aren't covered at high tide. The black mussel (*Musculus niger*), found in Arctic waters, can be poisonous year-round. Mussels in tropical zones are poisonous during the summer.	▪ Healthy bivalves should close their shells when tapped and the shells should open when cooked. Healthy univalves, such as limpets, should be difficult to pry off the rocks on which they live.
▪ Many terrestrial and freshwater snails are edible. The winkle or common periwinkle (*Littorina littorea*) is an edible sea snail.	▪ Avoid any terrestrial snails with brightly colored shells and all sea snails unless you can positively identify them as an edible species, as they may be poisonous.	▪ Snails should be starved for 24 hours or fed a diet of edible green leaves to purge their guts before cooking. Boil for at least 10 minutes.
▪ Most species of cricket (*Gryllidae*) and katydid (*Tettigoniidae*) are edible.	▪ Avoid any brightly colored grasshoppers as they may be toxic.	▪ Remove the antennae, wings, and leg spurs, and roast to kill any parasites.
▪ Most species of ant are edible if gathered carefully. The larvae of wood ants (*Formica* spp.) in northern temperate regions make a nutritious meal in summer months. The abdomen of the green or weaver ant (*Oecophylla smaragdina*) found in southeast Asia and Australia tastes like citrus sherbert. Termites (*Isoptera*) are also edible. See p. 219 for catching methods.	▪ Most species of ant are aggressive in defense of their nest and have a stinging bite. Some will then squirt formic acid at the site of the bite.	▪ The swollen abdomens of the repletes contain a nutritious fluid. ▪ The repletes live deep underground in the ants' nest so you would have to dig them out.
▪ The term "witchetty grub" is also applied to the edible larvae of other cossid moths, ghost moths (*Hepialidae*), and longhorn beetles (*Cerambycidae*). Palm grubs are the edible larvae of the palm weevil (*Rhynchophorus* Spp.). They live in the trunks of sago palms (see pp. 284–285) in Southeast Asia.	▪ Do not eat any insect larvae that are already dead when you find them, or that look sick, smell bad, or that irritate the skin if handled.	▪ Should be eaten raw or roasted quickly in hot embers.
▪ Marine decapods such as crabs, lobsters, and prawns are also edible.	▪ May contain harmful parasites so cook well.	▪ Keep alive until ready to eat. See p. 225 for preparation and cooking methods.
		▪ Boil or fry gently. Will turn slightly pink, like shrimp, when cooked.
▪ All earthworms are edible.		▪ The common earthworm is unusual in that it feeds at the surface, making it easier to find and gather than other worms.

NATURAL DANGERS

FOR INFORMATION ON LARGER dangerous animals, such as bears, big cats, and sharks, and venomous animals, such as snakes, see pp. 242–43. When planning your trip, it is advisable to familiarize yourself with the potentially harmful wildlife in the region you are visiting, the nature of the threat, and how to avoid it if possible. In the case of suspected

	NAME	WHERE	IDENTIFICATION
MAMMALS	**VAMPIRE BATS** *Desmodontinae*	▪ In rain forests, deserts, and grasslands, from Mexico to northern and central South America.	▪ Small bats, with a body length of 2¾–3½ in (7–9 cm) and a wingspan of 14–16 in (35–40 cm). Dark brownish gray fur, paler on the belly. Razor-like upper incisors.
MAMMALS	**RATS** *Rattus* spp.	▪ Worldwide, except for the polar regions.	▪ Medium-sized rodents, with a minimum body length of 5 in (12 cm), and long, naked tails. Coarse brown, gray-brown, or black fur, paler on belly. Pink feet.
SNAKES	**SPITTING COBRAS AND RINKHALS** *Naja* spp. and *Hemachatus haemachatus*	▪ Warm temperate, subtropical, and tropical regions of Asia and Africa.	▪ Pale brown to black snakes up to 6 ft (2 m) long. When threatened, will rear up and spread their hoods, revealing striped patterns on their necks.
FISH	**ELECTRIC EEL** *Electrophorus electricus*	▪ Amazon and Orinoco river systems, South America, often in shallow water.	▪ Not a true eel, but an eel-like fish up to 8¼ ft (2.5 m) long and can be as thick as a human thigh. Continuous fin along lower body.
FISH	**STONEFISH** *Synanceia verrucosa*	▪ In tropical coastal waters of northern Indian Ocean and southwestern Pacific Ocean.	▪ Up to 16 in (40 cm) long, this lumpy fish can match its color to the rock or sediment on which it lies while waiting for passing prey. Venomous spines in the dorsal fin.
OTHER ANIMALS	**BOX JELLYFISH/ SEA WASPS** *Cubomedusae*	▪ At the surface of tropical waters of southwest Pacific and eastern Indian Ocean.	▪ Box-shaped, transparent jellyfish, up to 10 in (25 cm) in diameter. Up to 15 long tentacles, which bear stinging cells, at each corner.
OTHER ANIMALS	**CENTIPEDES** *Chilopoda*	▪ In soil, leaf litter, cracks, and crevices in temperate, subtropical, and tropical regions worldwide.	▪ Elongated, flattened bodies divided into at least 16 segments, most of which have one pair of legs. Tropical species are typically brightly colored—yellow, red, orange, or green—with dark stripes. The world's largest centipede, *Scolpendra gigantea*, reaches 12 in (30 cm) in length.
OTHER ANIMALS	**CONE SHELLS/ CONE SNAILS** *Conidae*	▪ In inter-tidal zones and on coral reefs in warm temperate, subtropical, and tropical seas and oceans worldwide.	▪ Brightly colored and patterned, cone-shaped shells, up to 9 in (23 cm) long. These predatory snails have venomous, harpoon-like mouthparts.

WILD ANIMALS

poisoning, call the emergency services immediately. For further information regarding treatment for bites and stings, see pp. 266–67. If a casualty becomes unconscious, open the airway and check breathing (see p. 276). Be ready to begin CPR—chest compressions and rescue breaths (see p. 277), use a plastic face mask or face shield, if you have one.

DANGER	TREATMENT	HOW TO AVOID
▪ Many mammal species are infected with rabies, but because vampire bats feed by biting their victims, then lapping blood from the wound, the virus can be transmitted via their saliva. Symptoms include fever, headache, fear (especially of water), and seizures.	▪ If bitten by a vampire bat, or other potentially rabid mammal, clean the wound thoroughly and seek immediate medical help. Symptoms usually appear 2–8 weeks after infection, by which time the disease is almost always fatal.	▪ Vampire bats roost in hollow trees and caves, emerging at night to feed. Don't shelter in bat caves (bat dung can cause a respiratory disease). Keep covered at night, and sleep under a mosquito net. Get vaccinated before traveling to a high-risk area.
▪ Can be infested with parasites and can also carry infectious diseases. Leptospirosis is a water-borne bacterial disease spread by infected rats' urine; the acute form in humans is called Weil's disease.	▪ If you've been in contact with potentially contaminated water and develop flu-like symptoms (fever, headache, muscle pain) you must get your blood tested as soon as possible. Antibiotics are effective in the early stages of the disease.	▪ If handling a rat, beware of its teeth and claws, and wash thoroughly afterward. Cook the meat thoroughly. In rat-infested areas avoid getting water into any cuts; if you do, wash with soap and boiled water.
▪ Will eject a fine spray of venom from their fangs upward the eyes of an aggressor. Causes intense pain, tearing, and discharge from the eyes, spasm and swelling of the eyelids. Ulceration of the cornea can lead to infection resulting in blindness.	▪ Call the emergency services. Rinse the eyes with cool water for 10 minutes. Apply an antibiotic eye ointment.	▪ If you come across a spitting cobra, try to remain completely still and allow the snake to move away, if possible (see p. 243).
▪ Can generate shocks of up to 600 volts, used to stun or kill prey when hunting, or in self-defense. Such a shock can be discharged up to 8 hours after its death, and is potentially lethal to humans.	▪ Call the emergency services. If the casualty becomes unconscious, open the airway and check breathing (see p. 276). Be ready to begin CPR (see p. 277).	▪ Keep your boots on if crossing a tributary and take care when washing. These waters are also home to piranhas (*Serrasalminae*), which have razor-sharp teeth.
▪ The world's most venomous fish. If trodden on, the spines inject poison into the puncture wound. This is extremely painful and can be fatal. The injured body part will swell and muscular paralysis may set in.	▪ Call the emergency services—anti-venin is available. Immerse the injured part in water as hot as the casualty can stand for at least 30 minutes (see p. 267). Don't immobilize the injured limb. Be ready to begin CPR if the casualty loses consciousness (see pp. 276–77).	▪ Wear something on your feet when walking in shallow water. Weaverfish (*Trachinidae*) and scorpionfish (*Scorpaenidae*) also have venomous stings; the treatment is the same.
▪ Sting is extremely painful—it will damage the skin and can cause permanent scarring. In severe cases, box jellyfish (*Chironex fleckeri*) stings can lead to cardiac arrest and death within minutes.	▪ Call the emergency services—anti-venin is available. Flood the stung area with vinegar or sea water for at least 30 seconds to neutralize the stinging cells (see p. 267). Be ready to begin CPR if the casualty loses consciousness (see pp. 276–77).	▪ Never touch jellyfish, even when dead. All jellyfish, sea anemones, corals, and the Portuguese man-of-war (*Physalia physalis*) release venomous cells when touched. Some are more toxic than others.
▪ Have a pair of large, venomous claws on their heads, which in some species can be dangerous to humans. The sting is painful and leads to swelling. May result in anaphylactic shock.	▪ Call the emergency services. Apply a cold compress for at least 10 minutes to reduce pain and swelling. Monitor the casualty for anaphylactic shock (see p. 274) and be ready to begin CPR (see p. 277).	▪ As a general rule, avoid all multiple-legged arthropods. If you discover one on your body, gently brush it off in the direction in which it is walking.
▪ Will fire its harpoon in self-defense if touched. Sting of most species is painful; that of the larger tropical species can be fatal. Symptoms include swelling and numbness, and may be followed by severe breathing difficulties.	▪ Call the emergency services and treat as for snake bite (see p. 266). Monitor the casualty for shock (see p. 274) and be ready to begin CPR (see p. 277).	▪ Don't pick up any cone shells in rock pools or when snorkeling or diving. (There is no anti-venin; treatment entails providing life support until the venom is metabolized by the casualty's body.)

APPENDIX

NATURAL DANGERS

IN THE CASE OF SUSPECTED POISONING from a spider bite or a scorpion sting, you must call the emergency services immediately. For further information regarding treatment for bites and stings, see pp. 266–67. If a casualty becomes unsconscious, open the airway and check breathing (see p. 276). Be ready to begin CPR—chest compressions and rescue

	NAME	WHERE	IDENTIFICATION
INSECTS	**BEES, WASPS, AND HORNETS** *Hymenoptera*	▪ In a variety of terrestrial habitats worldwide.	▪ Narrow-waisted bodies, often striped black and yellow. Bees are hairy, while wasps and hornets have hairless bodies. Up to 1½ in (3.5 cm) long. Bees have barbed stingers, wasps and hornet stingers are smooth.
INSECTS	**MOSQUITOES** *Culicidae*	▪ In a variety of terrestrial habitats near water worldwide, especially in warmer regions.	▪ Narrow bodies and long, slender legs. Up to ¾ in (2 cm) long. Females have syringe-like mouthparts for piercing skin.
ARACHNIDS	**BROWN RECLUSE/ FIDDLEBACK SPIDER** *Loxosceles reclusa*	▪ Southern midwestern states of USA, south to the Gulf of Mexico.	▪ Small brown spider, about ½ in (1.25 cm) long) with a distinctive dark, violin-shaped mark on the back of its cephalothorax (head).
ARACHNIDS	**FUNNEL-WEB SPIDERS** *Atrax robustus* and *Hadronyche* spp.	▪ Moist, cool, sheltered habitats in eastern and southern Australia.	▪ Large spiders, up to 1½ in (4.5 cm) long, with glossy dark brown or black bodies and short legs.
ARACHNIDS	**TARANTULAS** *Theraphosidae*	▪ Deserts and forests in subtropical and tropical regions worldwide.	▪ Very large spiders with bodies up to 5 in (12 cm) long and a legspan of up to 11 in (28 cm). Bodies and legs are covered in bristly hairs. Pale brown to black in color, with pink, red, brown, or black markings.
ARACHNIDS	**WIDOW SPIDERS** *Latrodectus* spp.	▪ Warm temperate, subtropical, and tropical regions worldwide.	▪ Small spiders, with dark, glossy bodies and distinctive red, yellow, or white markings on the abdomen, which are shaped like an hourglass in some species.
ARACHNIDS	**SCORPIONS** *Scorpiones*	▪ Deserts, grasslands, woodland, and forests in warm temperate, subtropical, and tropical regions worldwide.	▪ Segmented bodies, with large, claw-like pedipalps and sting-bearing tails. Yellow, brown, or black bodies, 3–8 in (8–20 cm) long, depending on the species.
ARACHNIDS	**TICKS** *Ixodidae*	▪ Grassland and woodland habitats worldwide.	▪ Rounded bodies with no distinct divisions. Yellow to red- or black-brown in color. Up to ⅔ in (1 cm) long, but larger after feeding. They have spiked mouthparts, which they use to attach themselves to the skin of warm-blooded animals to feed on their blood.

INSECTS AND ARACHNIDS

breaths (see p. 277). If you need to give rescue breaths and there is poison in the casualty's mouth, or on his or her face, use a plastic face mask or face shield, if you have one. Before your trip, it is advisable to familiarize yourself with the harmful insects and arachnids (spiders, scorpions, and ticks) in the region you are visiting, and the nature of the threat.

DANGER	TREATMENT	HOW TO AVOID
▪ When a bee stings you, the stinger is ripped out of its body with the venom sac. Wasps and hornets can sting repeatedly. The stings are painful, followed by redness and swelling. A single sting may result in anaphylactic shock (see p. 274)	▪ Scrape the sting off sideways at the base of the sting if it is visible, using a fingernail or a credit card. Raise the affected part and place a cold compress against it for at least 10 minutes (see p. 267). Monitor for signs of allergic reaction.	▪ Watch out for flowers where bees may be feeding. Be careful when gathering fruit, or cleaning fish or game, as wasps or hornets will be attracted to the smell. Don't disturb their nests.
▪ Female mosquitoes feed on blood and their bites can irritate the skin. They are carriers of infectious diseases such as malaria, yellow fever, and dengue fever.	▪ Apply an antihistamine cream to the bites if they itch, and do not scratch them, as there is a risk of infection. If casualty develops a headache and fever, get medical help. Get casualty to rest, drink plenty of water, and take anti-malarials, if directed.	▪ Cover your arms and legs after sunset, use mosquito repellent, and sleep under a mosquito net. Take anti-malarial medication as directed. Get a yellow fever vaccination before traveling, where advised.
▪ Bite can cause fever, chills, vomiting, and, pain in the joints. In most cases the bite is minor, but in some cases the tissue around the wound will die, leaving a deep scar. Rarely fatal.	▪ Sit casualty down, treat as for snakebite (see p. 266), and call the emergency services. Apply a cold compress to the site of the sting. Monitor for signs of allergic reaction, such as wheezing, and anaphylactic shock (see p. 274).	▪ This spider likes to hide in dark places, so be careful where you put your hands and bare feet, and always check your bedding.
▪ Bite is very painful and cause profuse sweating, nausea, vomiting, tingling around the mouth and tongue, salivation, and weakness. Rarely fatal.	▪ Sit casualty down, treat as for snakebite (see p. 266), and call the emergency services—anti-venin is available. Monitor for signs of allergic reaction and anaphylactic shock (see p. 274).	▪ Take care when moving rocks and logs.
▪ Bite can be painful but is only mildly venomous. If provoked, tarantulas will flick irritating hairs from their abdomen which will cause itching and swelling of the skin and nasal passages, and watering of the eyes.	▪ Give pain relief and apply a cold compress to the site of the bite. Monitor for signs of allergic reaction, such as wheezing, and anaphylactic shock (see p. 274). Wash any hairs away with cool water.	▪ Tarantulas are nocturnal and hunt prey in trees and on the ground. Move away if you disturb one; don't touch it.
▪ Bite is painful and results in localized redness and swelling. More severe reactions include profuse sweating, abdominal and chest pains, nausea, and vomiting. Rarely fatal.	▪ Apply a cold compress (ice if available) and give pain relief. If symptoms worsen, call the emergency services – anti-venin is available. Monitor for signs of allergic reaction, such as wheezing, and anaphylactic shock (see p. 274).	▪ Widow spiders, which include Australia's redback spider (*L. hasselti*), build their webs in dry, sheltered spots, such as in shrubs or among rocks or logs, so take care when moving among them.
▪ Scorpion stings can be very painful and may cause severe illness. A few species have neurotoxins in their venom, resulting in temporary paralysis for 1–2 days. Rarely fatal and only in young children, the elderly, or the ill or infirm.	▪ If swelling occurs, sit casualty down, treat as for snake bite (see p. 266). Apply an ice pack or cold compress to the site of the sting. Call the emergency services, monitor for anaphylactic shock (see p. 274) and allergic reaction. Try to identify the scorpion.	▪ Scorpions are mostly nocturnal and take shelter during the day. Check your boots and clothing before putting them on and check your bedding at night. Take care when gathering firewood or turning over rocks.
▪ Ticks are carriers of infectious diseases such as Lyme disease and Rocky Mountain spotted fever. Sufferers of Lyme disease may develop a red, ringed "bull's-eye" skin lesion at the bite site.	▪ Carefully remove an attached tick with tweezers or a specialized tool (see p. 267). Clean and monitor the wound for symptoms. Get medical help if casualty develops a severe headache, stiff neck, or a fever.	▪ Wear light-shaded, tightly woven clothing and apply tick repellent. Check your skin and clothing for ticks at regular intervals.

NATURAL DANGERS

IN THE CASE OF SUSPECTED PLANT POISONING, you must call the emergency services immediately. If a casualty becomes unconscious, open the airway and check their breathing (see p. 276). Be ready to begin CPR—chest compressions and rescue breaths (see p. 277). If you need to give rescue breaths, use a plastic face mask or face shield, if you have one, to

	NAME	WHERE	IDENTIFICATION	POISONOUS PART
CONTACT POISONING	**COWHAGE/COWAGE/ COWITCH** *Mucuna pruriens*	▪ Scrub and light woodland in the the tropics and USA.	▪ Vine-like plant with green, oval leaflets growing in threes. Spikes of hairy, dull purplish flowers. Hairy brown seedpods.	▪ Flowers ▪ Pods
CONTACT POISONING	**POISON IVY** *Toxicodendron radicans*	▪ Wooded areas in North America; widely introduced elsewhere.	▪ See p. 268 for illustration.	▪ All
CONTACT POISONING	**WHITE/GRAY MANGROVE** *Avicennia marina*	▪ Mangrove swamps from tropical Africa to Indonesia and Australasia.	▪ Slender tree with pale bark and long aerial roots. Green, oblong, leathery leaves. Yellow flowers and fleshy, pale orange seed capsules.	▪ Sap
INGESTION POISONING	**BANEBERRY** *Actaea* spp.	▪ Temperate woodland in the northern hemisphere.	▪ Low shrub with, toothed green leaves. Small white flowers. White, black, or red berries grow in clusters.	▪ All, but the berries are the most poisonous.
INGESTION POISONING	**CASTOR OIL PLANT/ CASTOR BEAN** *Ricinus communis*	▪ Throughout the tropics in scrub and wasteland; introduced to temperate regions.	▪ Tall shrub with glossy dark green or dark purple, star-like leaves. Spikes of yellow flowers. Spiky red pods.	▪ All, but the seeds contain particularly high levels of the poison ricin.
INGESTION POISONING	**DEATH CAMAS/ DEATH LILY** *Zigadenus* spp.	▪ Grassy, wooded, and rocky areas in temperate regions of North America.	▪ Long grass-like leaves arise from a bulb. Flowers have six white petals with a green heart at base.	▪ All
INGESTION POISONING	**PANGI** *Pangium edule*	▪ Tropical forests in Southeast Asia.	▪ Tall tree with green, heart-shaped leaves. Green flower spikes. Large brownish pear-shaped fruits grow in clusters.	▪ All, but especially the seeds, which contain prussic acid (cyanide).
INGESTION POISONING	**PHYSIC NUT** *Jatropha curcas*	▪ Woodland throughout the tropics and southern USA.	▪ Large shrub with large, green, ivy-like leaves. Small greenish-yellow flowers. Apple-sized yellow fruit has large seeds.	▪ All, but particularly the seeds.
INGESTION POISONING	**POISON HEMLOCK** *Conium maculatum*	▪ Wet or moist ground in temperate parts of Eurasia; introduced to North America and Australia.	▪ Tall herb with branching purple-spotted stem; toothed leaves; small clusters of white flowers grown in flat umbrellas.	▪ All, but the root has the highest level of toxins.
INGESTION POISONING	**STRYCHNINE TREE** *Strychnos nux-vomica*	▪ Tropical and subtropical forests in India, Southeast Asia, and Australia.	▪ Evergreen tree with paired oval leaves. Small clusters of greenish flowers grow at end of branches. Large orange-red berries.	▪ All, but the seeds contain lethal levels of strychnine.

POISONOUS PLANTS

protect yourself from any poison on the casualty's mouth. When planning your trip, it is advisable to familiarize yourself with the poisonous plants in the region you are visiting. For further information regarding plant contact poisoning and ingestion poisoning, and their treatment in a survival situation, see pp. 268–69.

	EFFECT	TREATMENT	ADDITIONAL INFO
	▪ Skin irritation. Contact with the eyes can cause blindness.	▪ Call the emergency services. Rinse the skin with cold water for 20 minutes, the eye(s) for 10 minutes. Don't let the water collect under the casualty or you.	
	▪ Reddening, itching, swelling, and blistering of skin within 4–24 hours of contact with the plant. Some people may suffer an allergic reaction called anaphylactic shock (see p. 274).	▪ Wash the affected area with soap and cold water. Apply anti-poison ivy cream. Monitor and treat for shock (see p. 274) if necessary. For natural remedies, see p. 268.	▪ Western poison oak (*T. diversilobum*), Atlantic poison oak (*T. pubescens*), and poison sumac (*T. vernix*) also cause contact poisoning (see p. 268).
	▪ The sap blisters the skin and will cause temporary blindness if it gets in the eyes.	▪ Call the emergency services. Rinse the skin with cold water for 20 minutes, the eye(s) for 10 minutes. Monitor casualty and treat for shock (see p. 274).	▪ Blinding or milky mangrove (*Exoecaria agallocha*) also has poisonous sap.
	▪ Can be fatal—ingestion of berries can lead to cardiac arrest and death. Other symptoms include dizziness, vomiting, and severe internal irritation.	▪ Call the emergency services. Induce vomiting if casualty is still conscious. Dilute the poison by getting him or her to drink large quantities of water or milk, or tea mixed with charcoal.	
	▪ Can be fatal—ingestion of raw seeds causes severe diarrhea.	▪ Call the emergency services. Induce vomiting if casualty is still conscious. Dilute the poison by getting him or her to drink large quantities of water or milk, or tea mixed with charcoal.	▪ The seedpods explode when ripe, scattering the large, oval, glossy, brown-mottled seeds, which can be mistaken for beans.
	▪ Deadly poisonous. Symptoms include profuse salivation, vomiting, diarrhea, confusion, slow, irregular heartbeat, low temperature, difficulty breathing, and unconsciousness, leading to death.	▪ Call the emergency services. Induce vomiting if casualty is still conscious. Dilute the poison by getting him or her to drink large quantities of water or milk, or tea mixed with charcoal.	▪ This plant looks similar to wild onions, which are edible, but the death camas does not have the onion smell.
	▪ Deadly poisonous. Causes unconsciousness and severe breathing difficulties, which, if untreated, may lead to death.	▪ Call the emergency services (antidotes are available). Induce vomiting if casualty is still conscious. Dilute the poison (see Death Camas, above). Keep casualty warm and at rest.	▪ If the casualty is unconscious and stops breathing, give rescue breaths, using a face shield. Alternatively, use an oxygen bag and a mask, if possible.
	▪ Severe diarrhea and vomiting.	▪ Call the emergency services. Induce vomiting if casualty is still conscious. Dilute the poison by getting him or her to drink large quantities of water or milk, or tea mixed with charcoal.	▪ The seeds look like betel nuts and taste sweet. Several species in the *Jatropha* genus are poisonous, for example, the aptly named bellyache bush (*J. gossypiifolia*).
	▪ Deadly poisonous. A small amount can result in death by respiratory failure from muscular paralysis. Early symptoms include nausea, vomiting, and rapid heartbeat.	▪ Call the emergency services. Induce vomiting if casualty is still conscious. Dilute the poison (see Physic Nut, above). If the casualty is unconscious and stops breathing, begin CPR, using a face shield.	▪ Water hemlock or spotted cowbane (*Cicuta* spp.), which looks similar, is also deadly poisonous. Don't mistake these plants for wild carrots or wild parsnips.
	▪ Deadly poisonous. Muscle spasms occur 10–20 minutes after exposure, leading to seizures followed by death from severe breathing difficulties or exhaustion within 2–3 hours.	▪ Call the emergency services. Induce vomiting if casualty is still conscious. Dilute the poison by getting him or her to drink large quantities of water or milk, or tea mixed with charcoal.	▪ Warning: strychnine can also enter the system through contact with the eyes.

RESOURCES

SCOUTS AND SCOUTING

Boy Scouts of America

1325 West Walnut Hill Lane
P.O. Box 152079, Irving,
Texas 75015-2079

Phone: 972-580-2000

Fax: 972-580-2413

www.scouting.org

Boy Scouts of America prepares young people to make ethical and moral choices over their lifetimes by instilling in them the values of the Scout Oath and Law.

Girl Scouts of the USA

420 Fifth Avenue
New York, New York 10018-2798

Phone: 800-GSUSA-4-U [800-478-7248] or 212-852-8000

www.girlscouts.org

Girl Scouts is one of the largest organizations dedicated specifically to girls. It encourages them to develop skills and values such as leadership, social conscience, and self-worth that will stand them in good stead in the wider world.

National Scouting Museum

Boy Scouts of America, S505,
1329 West Walnut Hill Lane, Irving,
Texas 75038

Phone: 800-303-3047 and 972-580-2100

e-mail: nsmuseum@netbsa.org

www.bsamuseum.org

This museum is a tribute to the rich history of the Boy Scouts of America and features a Norman Rockwell art gallery, virtual reality adventures, hands-on learning experiences, and a collection that traces the Scouting movement from its beginnings.

WEATHER AND ENVIRONMENTAL HAZARDS

National Oceanic and Atmospheric Administration (NOAA)

1401 Constitution Avenue, NW,
Room 6217, Washington, DC 20230

Phone: 202-482-6090

Fax: 202-482-3154

www.noaawatch.gov

The NOAA has several departments, such as the National Severe Storms Laboratory (NSSL), the Storm Prediction Center (SPC), which is part of the National Weather Service (NWS), and the National Centers for Environmental Prediction (NCEP). The SPC provides regular, timely, and accurate storm forecasts, and watches out for severe thunderstorms, tornadoes, heavy rain, heavy snow, and fire weather events over the contiguous United States.

National Weather Service

1325 East West Highway, Silver Spring,
Maryland 20910

www.weather.gov/om/severeweather/index

Provides weather, hydrologic, and climate forecasts and observations for the whole country and the waters and oceans around it. The service offers a national information database for the use of public, private, and global communities.

NOAA Weather Radio All Hazards (NWR)

www.weather.gov/nwr

A nationwide network of radio stations broadcasts continuous weather information directly from the nearest National Weather Service office. NWR broadcasts official Weather Service warnings, watches, forecasts, and other hazard information 24 hours a day, seven days a week.

National Interagency Coordination Center (NICC)

www.nifc.gov/nicc/index.htm

The NICC serves as a focal point for coordinating the national mobilization of resources, such as heavy airtankers, helicopters, smokejumpers, area command teams, and Remote Automated Weather Stations (RAWS), for wildland fire and other incidents throughout the United States.

The Forest Fire Lookout Association (FFLA)

2590 W. Versailles Drive, Coeur d'Alene,
Idaho 83815

Phone: 208-765-1714 and 800-GRN-TREE

Email: weblookout@imbris.net

www.firelookout.org

The FFLA investigates former forest-fire lookout sites, ground cabins, and early forest fire detection methods to encourage public groups and others in restoring forest-fire lookouts.

U.S. Geological Service

Western Distribution Branch,
P.O. Box 25286, Denver, Colorado 80225

Phone: 303-236-7477

www.usgs.gov

This multi-disciplinary science organization focuses on biology, geography, geology, geospatial information, and water, and is dedicated to the timely, relevant, and impartial study of the US landscape, and natural resources and hazards.

DesertUSA

www.desertusa.com

This website is a comprehensive resource for North American deserts and Southwest destinations. It includes information about desert biomes, and the ways plants and animals adapt to them, and about national and state parks.

HEALTH AND MEDICAL

Centers for Disease Control and Prevention

1600 Clifton Rd, Atlanta, Georgia 30333

Phone: 800-232-4636 and 888-232-6348

Email: cdcinfo@cdc.gov

www.cdc.gov

Provides the expertise, information, advice, and tools that individuals and communities need to protect their health, including the promotion of healthy living and the prevention of disease, injury, and disability.

U.S. Department of Health and Human Services

200 Independence Avenue, S.W.,
Washington, D.C. 20201

Phone: 202-619-0257

Toll Free: 1-877-696-6775

www.hhs.gov/disasters/index.shtml

For information and advice on coping with disasters and related medical emergencies.

American Medical Association

515 N. State Street, Chicago, Illinois 60610

Phone: 800-621-8335

www.ama-assn.org/ama/pub/category/6206.html

The AMA's Center for Public Health Preparedness and Disaster Response is an important national educational resource for enhancing the disaster preparedness and response capabilities of both civilian and military providers.

American Red Cross

National Headquarters, 2025 E Street NW
Washington, DC 20006

Phone: 800-733-2767

www.redcross.org/services

The nation's premier response organization that provides relief for communities during natural and manmade emergencies. It belongs to a worldwide movement that offers neutral humanitarian care and assistance to victims of war and devastating natural disasters such as earthquakes.

National Capital Poison Center

3201 New Mexico Ave, Suite 310,
Washington D.C. 20016

Emergency Phone: 800 222 1222

www.poison.org

Affiliated with The George Washington University Medical Center, the Poison Center is committed to prevent poisonings, save lives, and limit injury from poisoning.

SEARCH AND RESCUE

National Association for Search & Rescue (NASAR)

PO Box 232020, Centerville,
Virginia 20120-2020

Phone: 703-222-6277

Toll Free: 877-893-0702

Email: info@nasar.org

www.nasar.org/nasar

A not-for-profit association dedicated to the advancement of professional, literary, and scientific knowledge in fields related to search and rescue throughout the United States and around the world.

National Institute for Urban Search & Rescue

PO Box 91648, Santa Barbara,
California 93190

Phone: 800-767-0093 and 805-569-5066

Cell: 805-798-0169

Email: niusr@cox.net

www.niusr.org

A non-profit organization dedicated to finding ways of saving lives by improving disaster readiness and response through public awareness, collaboration, research, and engineering.

Mountain Rescue Association

www.mra.org

Represents highly skilled and active mountain rescue teams throughout the country, who work for local government authorities and who are dedicated to saving lives through rescue and mountain safety education.

International Association of Dive Rescue Specialists

201 North Link Lane, Fort Collins,
Colorado 80524-2712

Phone (toll free): 800-IADRS-911

Phone (international): 1-970-482-1562

Fax (international): 970-482-0893

www.iadrs.org

An association of public safety divers and water rescue personnel who, over the years, have educated and assisted professionals in all 50 states and 15 foreign countries.

Association of Air Medical Services

526 King Street, Suite 415,
Alexandria, Virgina 22314

Phone: 703-836-8732

Fax: 703-836-8920

www.aams.org

An international not-for-profit organization that represents providers of air and surface medical transport systems, and aims to ensure that everyone has access to quality air medical and critical care support.

FEDERAL

Environmental Protection Agency (EPA)

EPA Headquarters, Ariel Rios Building
1200 Pennsylvania Avenue, N.W.,
Washington, DC 20460

EPA Hotlines: Various phone numbers and websites for assistance.

www.epa.gov/ebtpages/
emeremergencypreparedness.html

The EPA develops, implements, and coordinates preparations for chemical and other emergencies in order to be able to respond quickly and effectively to environmental crises and to keep the public informed about hazards in their community.

Federal Emergency Management Agency (FEMA)

500 C Street S.W., Washington, D.C. 20472

Phone: 800-621-FEMA-3362

TDD: TTY users can dial 1-800-462-7585 to use the Federal Relay Service.

www.fema.gov

FEMA aims to reduce the loss of life and property, and protect the country from all hazards, including natural disasters, acts of terrorism, and other manmade disasters.

U.S. Dept of the Interior National Parks Pass

1849 C Street, NW, Washington, DC 20240

Phone: 202-208-3100

Email: webteam@ios.doi.gov

www.doi.gov/parkpass.html

You can buy an annual pass for entry into national parks from the US Geological Survey store and through the government's federal lands recreation web portal at www.recreation.gov.

CANADA

National Search and Rescue Secretariat (NSS)

400-275 Slater Street, Ottawa,
Ontario K1A 0K2

Phone: (toll-free) 1-800-727-9414

Email: inquiry@nss.gc.ca

www.nss.gc.ca

An independent government agency that coordinates the search-and-rescue requirements of the military, coastguard, police, transport, parks, and meteorological services.

Search and Rescue Volunteer Association of Canada

8 Paradise Road, Paradise,
Newfoundland and Labrador A1L 3B4

Phone (toll free): 1-866-9SARVAC
(1-866-972-7822)

Phone (office): 709-368-5533

Fax: 709-368-1298

Email: info@sarvac.ca

www.sarvac.ca

A not-for-profit organization that provides a national search-and-rescue service, and is committed to fostering the exchange of information between other rescue services.

Meteorological Service of Canada

www.msc-smc.ec.gc.ca

Provides information and conducts research on water, climate, weather, atmospheric science, and other environmental issues.

Environment Canada

Inquiry Center, 351 St. Joseph Boulevard
Place Vincent Massey, 8th Floor
Gatineau, Quebec K1A 0H3

Phone: 1-800-668-6767
or 819-997-2800

Fax: 819-994-1412

TTY: 819-994-0736

Email: enviroinfo@ec.gc.ca

www.ec.gc.ca

Government department committed to preserving and enhancing the country's environment, and to providing weather forecasts.

GLOSSARY

A-frame A type of shelter design in which the roof is suspended by a single support running from the front to the back, resting on an "A"-shaped framework at either end.
Air mass A large body of air in a weather system that has its own temperature and humidity.
Altitude sickness An illness brought on by low air pressure at high altitudes.
Anabatic winds Daytime, upslope winds.
Anaphylactic shock A dangerous allergic reaction brought on by a bite or sting.
Attack point See *Stand-off*.
Avalanche transceiver A personal beacon that can be activated to alert rescuers by someone caught in an avalanche.

Base layer The bottom layer of clothing that sits directly against the skin and is designed to "wick" sweat away from the body.
Base rate The heart rate, as measured when the body is at rest.
Beaufort scale A way of describing wind speed on land and at sea, using a scale from 0 (calm) to 12 (hurricane).
Belaying A rope technique used by climbing partners to safeguard each other.
Bivi pup tent A basic one-man tent that uses a bivi bag as a covering.
Bivi bag A large waterproof sack suitable for use as a basic shelter; smaller and lighter than a tent.
Blazes A general term for markers on a trail.
Bola A hunting device made of a rope with weights attached, designed to be spun and thrown.
Boot skiing A technique for descending snow-covered slopes without skis.
Bothy bag A large waterproof bag that provides shelter from the wind, rain or sun.
Bow-and-drill set One of the oldest methods of creating a glowing ember (coal) for use in lighting a fire.
Boxing an object See *Detouring*.
Breathable clothing Clothing that allows sweat in the form of water vapour to be carried away from the body but is waterproof from the outside so keeps liquids (rain or snow) out.
Bungee A type of elasticated cord, often with hooks at either end.
Bushcraft knife A versatile survival knife used primarily for carving, cutting, and splitting small logs.
Button compass A small, simple compass, ideal as a back-up compass.
Button tie A method of attaching cordage to fabric that has no hoops or grommets; useful in that it avoids ripping a hole in the fabric, maintaining any waterproof qualities of the fabric, and allows both fabric and cordage to be reused subsequently for other purposes.

Cairn Trail marker consisting of a pile of rocks of varying sizes, designed to be visible in fog.
Canoe An open-deck paddle boat, for one or more people, and equipment.
Carabina A metal clip used for joining ropes or attachment.
Carbon monoxide A colourless, odourless, tasteless, but highly toxic gas.
Cardiopulmonary resuscitation (CPR) An emergency system of chest compressions and rescue breaths designed to restart a casualty's heart and lungs.
Cerebral oedema Accumulation of excess water in the brain that causes reduced brain function and potentially fatal swelling.
Char cloth Cotton cloth that has been combusted in the absence of oxygen. Used for lighting fires.
Chimneying A climbing technique used for climbing up the inside of large rock clefts.
Cirrus A wispy cloud that forms high in the sky and is made from ice crystals.
Cold shock An involuntary gasp reflex followed by hyperventilation, caused by sudden immersion in very cold water. In some circumstances cold shock response can lead to the inhalation of water, disorientation, panic, hypothermia, cardiac problems, and death.
Commando saw A small saw consisting of a serrated wire blade with a ring at each end.
Compacted snow trench A shelter dug into compacted snow; the excavated snow can be cut into blocks and used to form a roof.
Compass baseplate A plate to which some compasses are fixed, containing additional markings used for orientation and navigation.
Compass scale Scaled markings on a compass baseplate that measure distances on a map and help in working out grid references.
Compass An instrument used for orientation and navigation, using a freely rotating needle that indicates the direction of magnetic north.
Coniferous tree A needle-leaved, cone-bearing, mostly evergreen tree, such as pine, spruce, or fir.
Continental air mass Air that has tracked over land and carries comparatively less moisture than maritime air mass.
Contour lines Lines on a map that show points of equal height above sea level, thus detailing the changing height of natural features.
Contour navigation Walking at a constant height around a high natural obstacle.
Contouring See *Contour navigation*.
Contusion A bruise.
Cordage A type of light rope – an essential piece of survival equipment.
COSPAS-SARSAT system A satellite system, made up of low-altitude earth-orbiting (LEOSAR) and geostationary earth-orbiting (GEOSAR) satellites, that picks up signals from distress beacons.
Cramp Painful muscle spasm, often a result of dehydration.
Crampons Spiked metal plates that attach to boots to provide grip on icy surfaces.
Cumulonimbus A type of cloud that starts low in the sky and builds upwards, producing short, heavy downpours or thunderstorms.
Cumulus A billowing, puffy cloud that is generally small and develops on bright, sunny days, indicating fine weather.

Damper bread Yeast-free bread, suitable for making on a campfire.
Deadfall A trap in which a weight falls onto the prey. Also a mass of fallen dead timber.
Defensive swimming A type of swimming technique designed to keep the swimmer protected from obstacles in the water.
Degree A unit of latitude or longitude, equal to 1/360 of a circle.
Dehydration A low level of water in the body. A very dangerous condition if not reversed.
Deliberate off-set (aiming off) Deliberately aiming left or right of a known feature.
Desalting kit An emergency desalination kit for turning saltwater into freshwater.
Detouring Walking around an obstacle as a method of clearing it, using a point in the distance as a reference.
Dew point The temperature at which water vapour held in the air cools to become liquid.
Dew trap A method of collecting dew to form usable water.
Digging stick A sturdy piece of wood with a pointed end.
Dipping net A fishing net used for scooping up fish that are too small to catch on a line.
Drill and flywheel A friction-based method of starting a fire.
Drip rag A method of collecting rainwater dripping down a tree using cloth.
Drogue A sea anchor used to stabilize a vessel.
Dysentery A water-borne disease that causes severe diarrhoea.

Eastings Vertical grid lines on a map; they increase in value the further east they are.
Emergency Locator Transmitter (ELT) A beacon that can send a distress signal to an orbiting satellite to initiate a rescue; mostly for use in aircraft.
Emergency plan of action (EPA) A document left with relevant authorities that contains vital details about you and your intended route, should you need to be rescued.
Emergency Position-Indicating Radio Beacon (EPIRB) A beacon that can send a distress signal to an orbiting satellite to initiate a rescue; mostly used at sea.
Emergency rendezvous (ERV) A pre-determined point at which all members of a group should meet in the event of separation.
Equator An imaginary circle running around the Earth's diameter that is equidistant from the North and South Poles at all points.
Eskimo roll A technique for righting a capsized kayak.

Faggot A bundle of kindling.
Feather stick A stick that has been feathered to use as kindling and fuel.
Fighter trench A type of snow trench dug into soft snow. Ideally uses a tarpaulin for the roof.
Finnish marshmallow A method of melting chunks of ice over a fire then collecting the water in a container.
Fire tin A tin filled with waxed cardboard, used for lighting a campfire in difficult conditions, or as a basic cooking device.
Firebase A stable, non-flammable base suitable for building a campfire on.
Firefighter's lift A method of moving a casualty by carrying them over your shoulder.
Fire set A set of components necessary to generate a glowing ember that can start a fire.
Firesteel A metal bar which, when hit with a striker, produces a spark for lighting campfires.
Flash burn See *Snow blindness*.
Föhn effect A dry downslope wind which occurs on the lee-side of a mountain or hill.
Four-wheel drive (4WD) A system on a vehicle in which drive is provided by all four wheels, rather than just two; results in superior traction.
Frostbite A serious condition caused by excessive exposure to extreme cold, in which body tissues actually freeze, sometimes to the bone. Often starts with the extremities.
Frostnip The freezing of the top layer of skin, usually on the face and extremities, caused by exposure to extreme cold. Can develop into frostbite if left untreated.

Gaiters Protective fabric that wraps around the lower leg and ankle, to keep out water and protect against sharp objects.
Gill net A fishing net in the form of a mesh stretched across a stream.
Global positioning system (GPS) A hand-held unit that uses orbiting satellites to determine the user's position accurately.
Gourd The shell of hollowed-out, dried fruit often used to carry water.
Grab bag A pre-prepared bag of essential survival equipment suitable for quickly picking up in an emergency at sea.
Gravity filter An effective method of filtering and purifying water, incorporated into a bottle.
Grid bearing A horizontal direction expressed in degrees east or west of north or south.
Grid magnetic angle (GMA) The difference between magnetic north and grid north, expressed as an angle.
Grid north A northerly direction that runs parallel to vertical gridlines on a map. Different from true north because a map is flat.
Grid reference A method of pinpointing the location of a place or object anywhere on a map, using coordinates provided by a numbered grid system printed on the map.
Grommet A metal-ringed eyelet in clothing.
Groundsheet A waterproof sheet.
Gypsy well A method of using a hole in the ground to filter non-potable water.

Hand jam A crack in a rockface suitable for wedging your whole hand into when climbing.
Handrailing A method of navigating using long linear features on the landscape that run in the general direction of your travel, such as rivers, roads, or paths.
Hank A small coil of cordage.
Heat exhaustion (heat stress) Dehydration that results in dizziness and profuse sweating; often a precursor of heatstroke.
Heatstroke A life-threatening condition brought on by severe overheating.
HELP (Heat Escape Lessening Posture) A posture that reduces heat-loss from the body when floating in water.
Hexamine stove A stove that uses hexamine solid fuel instead of gas or liquid fuel.
Hot platform A method of melting ice over a fire and collecting the liquid in a container.

Hyperthermia An abnormally high body temperature; potentially fatal.
Hyponatraemia An accumulation of excess water in the body, leading to a dangerously low concentration of sodium.
Hypothermia A life-threatening condition caused by exposure to cold. Symptoms usually appear in the following order: uncontrollable shivering (which may stop as the body temperature lowers); irrational or out-of-character behaviour; confusion, mood swings, and withdrawal; pale, cold skin; and slurred speech and/or stumbling. Watching out for and counteracting the early symptoms may help prevent them developing into a full-blown potentially fatal condition.
Hypoxia Inadequate levels of oxygen in the blood or body tissues. Potentially fatal.

Igloo A shelter constructed from cut blocks of snow.
Iodine A chemical used to purify water.
Isobar contours Lines on a weather chart that connect points at which the barometric pressure is the same.

Katabatic winds Cool, light downslope winds that form on clear nights.
Kayak A closed-deck paddle boat for one or more people.
Kindling Small pieces of fuel for a fire, added to burning tinder.
Kukri A large-bladed knife from Nepal, used mainly for chopping.

Land breeze A night-time wind that blows from land to sea.
Lanyard A small cord or rope used for securing or suspending an item.
Latitude The angular distance north or south of the Equator, in degrees along a meridian.
Layering Wearing several light layers of clothes rather than just one or two thick ones. Layers can be added or taken off to maintain a consistent body temperature.
Lean-to A shelter with a sloping roof that leans against a horizontal ridgepole.
Lee-side The side that is sheltered from the wind.
Lensatic compass A type of compass used for precise navigational work.
Longitude The angular distance east or west across the Earth between one meridian and the prime meridian at Greenwich, England.

Machete A large-bladed knife used mainly for chopping, approximately 45–50cm (18–20in) long.
Magnetic north The direction indicated by a magnetic compass.
Magnetic variation The difference between magnetic north and grid north.
Mantelling A technique used for climbing overhanging rock.
Maritime air mass Air that has tracked over the sea, typically containing moisture.
Mayday The internationally recognized distress signal for a grave emergency.
Matchless fire set Small kit that includes everything you need to start a fire – tinder, fuel, and spark device.
Melting sack An improvised sack that can hold snow or ice and is positioned close to a fire to provide water.
Meridian An imaginary arc on the Earth's surface from the North to the South Pole, connecting all locations running along it.
Mid-layer An insulating layer of clothing that sits between the base layer and the outer layer.
Motor impairment A limitation or loss of muscle control or movement.

Naismith's Rule A method of calculating the time it will take to arrive at a destination, taking into account distance and topography.
Natural hollow A natural dip in the ground that can be used as an emergency shelter.
New World Non-Eurasian and non-African regions of the world – specifically the Americas and Australia.
Noggin A piece of wood used to help support a structure, reducing the amount of cordage required.
Non-potable water Water that is not suitable for drinking.
Northings The horizontal grid lines on a map; increase in value the further north they are.

Occluded front A point at which two air masses meet in a weather system.
Old World Europe, Asia, and Africa.
Outer layer Outer clothing; ideally it should keep out rain but let sweat (as water vapour) escape.

Pace counting A method of determining distance; involves knowing how many paces you take to cover a set distance.
Pack animal An animal such as a mule or horse used for carrying heavy loads.
Paracord A useful type of cordage originally developed as rigging lines for parachutes. Usually comprises an outer sheath over an inner section of yarn.
Parang A large-bladed knife from Malaysia, used mainly for chopping.
Pathogen A micro-organism, especially bacteria or fungi, that causes disease.
Permafrost Permanently frozen soil.
Personal Locator Beacon (PLB) A beacon that can transmit a distress signal to an orbiting satellite to initiate a rescue.
Poncho A multi-purpose waterproof outer garment made from one sheet of fabric; also useful for building a basic shelter or bed.
Post-holing A technique for walking in deep snow without snow shoes.
Potable water Water suitable for drinking.
Potassium permanganate A chemical used for a variety of tasks, from lighting fires to purifying water.
Precipitation Moisture in a cloud that may fall as rain, snow, or hail.
Prismatic compass See *Lensatic compass*.
Protractor A device for measuring angles.
Psychogenic shock A very high level of psychological and emotional stress brought on by a sudden disaster situation.
Pulk A simple plastic sled used for carrying equipment over snow.
Pulmonary oedema Swelling and fluid accumulation in the lungs that causes potentially fatal breathing problems.
Pygmy hut A domed hut made from a circle of bent saplings or limber poles and thatched with natural materials.
Pygmy roll A method of rolling fibres together to make cordage.

Quicksand A bed of loose, wet sand that yields easily to pressure and can engulf anything on its surface.
Quinzhee A basic dome-shaped, snow-covered overnight shelter.

Ranger flint and steel A rod of flint which, when hit with a striker, produces a spark for lighting campfires.
Recovery position The position an unconscious person should be placed in to minimize further injury and help recovery.
Reflector A device for directing heat from a campfire towards shelter.
Reverse-osmosis pump A survival pump system that turns saltwater into freshwater.
Ridgepole A long horizontal pole that forms the apex of a roof.
Romer measure See *Compass scale*.

Salting A salt marsh.
Scrambling Climbing without ropes.
Scrape A type of basic shelter built in a natural hollow or depression in the ground.
Sea breeze A daytime wind that blows from sea to land.
Secondary drowning Potentially fatal biological changes that can take place in the lungs after a person nearly drowns, caused by the body's response to inhaling water.
Shock A life-threatening condition that occurs if the circulatory system fails, often triggered by severe bleeding, burns, or sudden cold.
Show-stopper A failure in planning that could cause delays to, or stop a trip or expedition.
Sighting mirror An advanced component on a high-end compass used for helping in orientation and navigation.
Signal fire A fire specifically designed to produce lots of smoke to attract attention.
Signal flare A hand-held distress flare that gives off orange smoke in daylight and a bright light at night.
Signal mirror A simple signalling device that works by reflecting sunlight.
Silva compass A type of basic compass useful for hiking.
Sip well A method of extracting water trapped under rocks by sandy ground.
Skidoo A motorized vehicle used for travelling over snow and ice.
Slingshot A hunting weapon that fires small stones and is made from a Y-shaped stick with an elastic strip stretched between the prongs.
Snare A wire noose used for trapping animals.
Snow blindness Eye damage caused by ultraviolet light reflected off snow or water.
Snow cave An effective cave-like shelter dug into compacted snow on the lee-side of a hill.
Snowshoes Wide-soled strap-on shoes designed to stop you sinking into snow.
Soak A water source found close to rivers and creeks in areas that are usually lower-lying than the existing water table.
Solar still A method of extracting drinking water from any source of moisture by a process of evaporation and condensation.
Space blanket A plastic blanket covered in aluminum foil that can be used as a shelter or a reflective signalling device; to carry, store, and heat water; or to cook in.
Stand-off A technique for navigating to a point that may be hard to locate, by orientating from a nearby prominent feature.
Standing deadwood Wood from a tree that has died but is still standing. Hazardous, as branches or the whole tree may fall, but makes excellent firewood.
Straddling A climbing technique used for ascending wide chimneys.
Stratus Dense, grey cloud that forms a sheet. Rain can fall for long periods of time if the cloud has depth.
Strobe light An effective signalling device in the form of a rapidly flashing LED light.
Survival blanket See *Space blanket*.
Survival kit Essential items for survival carried on your person.
Survival straw A compact emergency water purifier.
Survival suit A waterproof, buoyant suit designed for survival in open water.
Survival tin An essential, compact container carried on the person that addresses basic survival needs: protection, location, water, and food.

Taiga A subpolar region characterized by coniferous forest.
Tarp Shortened form of "tarpaulin", a type of sturdy waterproof sheet.
Temperate climate A climate without extremes of temperature or rainfall.
Tetanus A potentially lethal infection caused by bacteria that live in soil.
Throwing star An improvised four-pointed hunting weapon used for throwing at prey.
Tinder A dry, light, combustible material that is the first fuel used when lighting a fire.
Topographic map A map that shows the main features on a landscape, and elevation via use of contour lines.
Transpiration bag A method of evaporating, condensing, and siphoning off water from foliage using a plastic bag and sunlight.
Triangulation A method of determining your location by taking bearings with a compass from recognizable landscape features.
True north The direction of a meridian of longitude that converges on the North Pole.
Tundra A polar region characterized by permafrost and stunted vegetation.

Universal Edibility Test (UET) A test that checks whether a plant is safe to eat via a methodical process of testing small amounts on the skin, in the mouth, and in the stomach.
UV active purification Purifying water via a battery-operated ultraviolet (UV) purifier.
UV passive purification Purifying water by leaving filled bottles in strong direct sunlight. The UV rays eventually kill most pathogens.

Vegetation bag See *Transpiration bag*.
VHF radio Very high frequency radio transmitter; usually used at sea.

Water balance The difference between water lost from the body through sweat and the water taken in through drinking.
Water purifier A device that filters and purifies water by pumping it through micro filters, chemicals, or a combination of both.
Weather chart A map of major weather systems and their predicted direction of travel.
Wicking material A material that moves moisture away from your body to evaporate.
Wickiup A small hut made from straight poles lashed together at the top, with an interwoven framework covered with animal hides or grass.
Witch's broom A cleft stick stuffed with thin, dry bark suitable for lighting a fire.
Withies The strong, flexible stems of plants, such as willow, birch, ash, and hazel.

Zig-zag route Method of climbing a steep slope using a zig-zag path to reduce the effort required to achieve the climb. Also effective when walking down steep slopes.

INDEX

A

A-frames
  forest 164–65
  front opening 165
  jungle 168
  poncho shelters 159, 167, 175
abrasions (grazes) 263
abandoning ship 250, 252–55
ABC (airway/breathing/circulation), first aid 260, 276–77
aerobic exercise 19
agaves 193, 282–83
air quality, caves 157
aiming off 73
air masses 79
allergies and anaphylactic shock 266, 267, 274
altitude sickness 33, 273
altocumulus 80, 81
altostratus 80, 81
amoebic dysentery 187
amphibians and reptiles
  catching and eating 203, 218, 224, 294–95
  dangers from 243, 266, 300–01
  see also snakes
anabatic winds 82
analogue watches, navigating using 77
anaphylactic shock 266, 267, 274
anchor points, winching 103
anchors, sea 39, 251
animals
  dangers from 115, 157, 189, 242–43, 300–3
  edible 290–99
  locating water 189
  navigation aids 77
  traveling with 42, 96–97, 98–99
  weather forecasting 83
  see also specific animals (e.g. bears); specific types (e.g. mammals)
ants 219, 298–99
arbor knot 142
arm injuries 270
ash
  for soap 117
  for stomach pain 269
asthma 275
automatic fires 124
avalanches 32, 246–47

B

bacillary dysentery 187
backpacks 41, 42, 44–45, 108
bacterial infections 185, 187, 212, 220, 262, 267, 301, 303
baits and lures 208–09, 223, 229
baking food 204, 205
bamboo 284–85
  fire saw 133
  lean-tos 169
  rafts 107
  steamer 204
  water sources and collection 192, 199
bark
  cordage 138
  tea 269
  tinder and kindling 120, 121, 238
bats 157, 290–91, 300–01
beaches 39, 177
beacons, phones, and radios 39, 43, 61, 232–33, 236–37, 246, 250, 252
bearings, compass 64–65, 68–71
bears 28, 36, 157, 242
Beaufort scale 83
beds and bedding 41, 52–53, 156, 166–67
bees 224, 267, 302–3
belaying 91
berms 177
big cats 32, 242
bilharzia 187
birds
  catching and eating 203, 204, 205, 218, 226–29, 292–93
  locating water 189
bites and stings 116, 266–67, 268, 300–03
bivy sacks 53
  bivy pup tent 160–61
  hooped bivy 41, 54
blankets, space 53, 61, 155, 175
blazes (trail markers) 87
bleeding
  birds, for food 228
  first aid 263–64
  mammals, for food 222
blended fuel stoves 58
blisters 262, 268
boats *see* ships and boats
body fat 56
boiling
  food 204, 212
  water 118, 185, 200
bolas 226
bones, joints, and muscles injuries 233, 270–1
boot skiing 94
boots 46, 49, 114
bothy bags 161
bottle filters 200
bottle traps, fishing 211
bough fence traps 227
bow drills 130–33
boxing an object (detouring) 73
branches
  as snowshoes 95
  vehicle recovery 102
bread, damper 204
breathable waterproof fabric 48
breathing difficulties
  first aid 275
  in shelters 157, 181
breezes and winds 38, 79, 82–83
brown bears 28, 242
bruising 262, 263
brushwood rafts 106–07
buddy-buddy sleeping system 163
buoyancy aids (floats) *see* lifejackets
burns and scalds 265
bushcraft knives 146–49
butane stoves 58
button compasses 68
button ties 161

C

cacti 193, 282–83
cage traps, figure 4 226
cairns 87
calorie needs 56, 202
camels 42, 99
camping soap 117
camping stoves 58–59
camps
  organization and administration 112–13, 114–15
  see also specific areas (e.g. latrines)
canoeing and kayaking 42, 104–05
capacity, packs and bags 41, 42, 44
capsizing 104, 105
carbohydrates 56
carbon dioxide 157
carbon monoxide poisoning 181
cardinal points 68, 76, 77
cardiopulmonary resuscitation (CPR) 277
cats, big 32, 242
caves 157, 181
cell phones 43, 237
chainsaws 60, 150
char cloths 122–23
char test, softwoods and hardwoods 131
charts
  maps and map reading 43, 64–65, 66–71
  weather 78
chemical flames 127
chemical lightsticks 61, 240
chimneying 91
chlorine tablets 201
choking 275
cholera 187
cirrostratus 80, 81
cirrus 80, 81
clay baking, food 205
climate 78
  see also specific environments (e.g. temperate environment)
climbing and scrambling 90–91
clothing 46–49, 94
  cleanliness 116
  cultural differences 22
  on fire 265
  hats 43, 46, 47, 48
clothing, improvising with
  buoyancy aids 255
  cordage 137
  dipping nets 210
  slings 261
  stretchers 279
  water collection and treatment 185, 190, 192, 195, 199, 200

clouds 78, 79, 80–81
coastal caves 157
coasts *see* marine environments
cold environments 36–37
  clothing 47, 94
  cold shock 254
  dos and don'ts 51, 245
  edible plants 286–87
  hypothermia and frostbite 21, 29, 33, 37, 252, 273
  tents 55
  water sources 194–95
  see also ice; snow
cold fronts, weather 78, 79
cold shock 254
cold-weather clothing 47, 94
commando saws 151
commando-style match striking 126
communications equipment 43, 61, 232–33, 236–37
compacted snow trench 178
compasses and bearings 43, 64–65, 68–71, 74–75
components of fire 118, 238
condoms 61
conjunctivitis 212
constellations, navigating by 77
continental air masses 79
contour lines 66, 67
contouring 73
contusions (bruising) 262, 263
cooking
  fires, rocks, and pits 119, 204–05
  fish preparation 212–13
  invertebrates 219, 224–25
  mammals preparation 220–23
  stoves 58–59
coping strategies 20–21, 113, 234–35, 259
cordage 136–41
cordless scrape shelters 174
cougars 32, 242
CPR (cardiopulmonary resuscitation) 277
cardiovascular fitness 18, 19
crabs 225
cramp 270
crickets and grasshoppers 224, 298–99
crocodilians 189, 242
crossing rivers 88–89, 101, 108–09
crustaceans 225
cryptosporidium 187
cumulonimbus 80
cumulus 80, 81
cuts and grazes 263
cutting tools 60, 146–51
cyalumes 61, 240

D

Dakota hole fires 125
damper bread 204
day packs 42, 44
deadfalls 115
deadwood, standing 115, 121, 131
defensive swimming 108
dehydration and rehydration 187, 190, 193, 196, 253, 269, 272
deliberate off-set 73
denatured alcohol stoves 59
desalting kits 197
desert environments 34–35
  clothing 47
  dos and don'ts 129
  edible plants 282–83
  four-wheel driving 35, 101
  night walking 86
  shelters 174–75
  water sources 34, 35, 193
detouring (boxing an object) 73
dew harvesting and traps 190, 196
dew point 80
diarrhea and vomiting 187, 203, 269
digging
  ponds 199
  roots and tubers 206
  underground water 193
  vehicle recovery 102
dipping nets 210
disaster situations and emergencies
  causes 16–17, 26, 84
  dos and don'ts 27, 51, 93, 129, 173, 215, 232–33, 245
  emergency plans of action (EPAs) 24–25, 250
  emergency rendezvous points (ERVs) 72
  environmental hazards 246–49
  four-wheel driving (4WDs) 35, 103
  gear priorities 43
  mental preparation and responses 20–21, 234–35, 259
  PLWF (protection, location, water, and food) 11, 27, 258
  rescue services 24
  rescue signaling 39, 43, 232–33, 236–41, 250, 252
  survival at sea 16–17, 39, 196–97, 214–15, 250–55
  survival kits 60–61
  SURVIVAL strategies 234–35
  water sources and rationing 39, 187, 188–97, 199, 251
  see also first aid
diseases *see* infections
distance measuring 67, 72
distilling water 185
distress signals 39, 232–33, 236–41, 250, 252
dogs 17, 96–97, 99
dome signal fires 238–39
dos and don'ts
  disaster situations and emergencies 27, 51, 93, 129, 173, 215, 232–33, 245
  fires 118, 119, 124
double overhand knot 143
double sheet bend 144
down sleeping bags 52–53
downhill walking 86, 87
drawing birds (food preparation) 229
driftwood shelters 176–77
drip rags 192
driving techniques, 4WDs 35, 100–03
drogues 39, 251
drowning 254, 255, 273
drum rafts 107
duck traps 227
dunes, sand 35, 177, 189
dysentery 187

E

E. coliform infection 187
earthquakes 247
earthworms 224, 298–99
eastings 67
edible plants 203, 206–07, 280–89
eggs 226
electric current, magnetizing compass needles 75
electrical storms 81, 249
elements of fire *see* fuel; kindling; tinder
elephants 99, 189, 242
ember pans 132
emergencies *see* disaster situations and emergencies
emergency signaling codes 240, 241, 252
emergency transmitters 39, 232–33, 236–37, 246, 250, 252
emergency water packs 199
emotional responses
  disaster situations 20–21, 113, 234–35, 259
  team dynamics 23
environmental hazards 246–49
environments
  understanding and preparing for 23, 26
  see also specific environments (e.g. cold environments)
EPAs (emergency plans of action) 24–25, 250
EPIRBs (emergency position indicating radio beacons) 236, 250, 252
equipment *see* gear and equipment
ERVs (emergency rendezvous points) 72
eskimo rolls 104
Evenk knot 144
exercise and fitness 18–19
exposure, first aid 272–73
external-frame packs 44
extreme weather 246, 248–49
eyes
  care 116
  problems and injuries 212, 265

F

fallen trees, shelters 163
fatigue 21, 235
fats 56
fear 234, 235
feather lures 209
feather sticks 120–21
features, use in navigation 70–71, 72, 73
feelings *see* emotional responses
felling trees 148–51
feral animals 242
fibre (nutrient) 56
fighter trench 179
figs 171, 284–85
filleting fish 212–13
filtering water 184, 200, 201
Finnish marshmallow 195
fire saws, bamboo 133
fire tins 123
fire wads 123
firefighter's lift 278
fires
  benefits 27, 113
  burns, first aid 265
  in caves 157
  clothing on fire 265

AF REGULATION 64-4
VOLUME I

Search and Rescue

SURVIVAL TRAINING

SURVIVE • RETURN • HONOR

15 JULY 1985

DEPARTMENT OF THE AIR FORCE

DEPARTMENT OF THE AIR FORCE
Headquarters US Air Force
Washington DC 20330-5000

AF REGULATION 64-4
VOLUME I
15 July 1985

Search and Rescue

SURVIVAL TRAINING

This regulation describes the various environmental conditions affecting human survival, and describes individual activities necessary to enable that survival. This regulation is for instructor and student use in formal and USAF survival and survival continuation training. This regulation also applies to US Air Force Reserve and Air National Guard units and members. Sources used to compile this regulation are listed in the bibliography, attachment 1.

	Paragraph	Page
Part One — The Elements of Surviving		
Chapter 1 — Mission		
Introduction	1–1	22
Aircrew Mission	1–2	22
Goals	1–3	23
Survival	1–4	23
Decisions	1–5	23
Elements	1–6	23
Chapter 2 — Conditions Affecting Survival		
Introduction	2–1	24
Environmental Conditions	2–2	24
The Survivors Condition	2–3	24
Duration – The Time Condition	2–4	27
Sociopolitical Condition	2–5	27
Induced Conditions	2–6	27
Chapter 3 – The Survivor's Needs		
Introduction	3–1	30
Maintaining Life	3–2	30
Maintaining Honor	3–3	31
Returning	3–4	32
Part Two — Psychological Aspects of Survival		
Chapter 4 — Contributing Factors		
Introduction	4–1	36
Survival Stresses	4–2	36
Pain	4–3	37
Thirst and Dehydration	4–4	37
Cold and Heat	4–5	38
Hunger	4–6	39
Frustration	4–7	40
Fatigue	4–8	40

Supersedes AFM 64-3, 15 August 1969. (See signature page for summary of changes.)
No. of Printed Pages: 579
OPR: ATC/DONZ (MSgt R.J. Paetz)
Approved by: ATC/DON (Col W.W. Miller)
Editor: M.M. Rastellini
Distribution: F

This publication includes copyrighted material.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402-5000

	Paragraph	Page
Sleep Deprivation	4–9	42
Isolation	4–10	42
Insecurity	4–11	42
Loss of Self–Esteem	4–12	43
Loss of Self–Determination	4–13	44
Depression	4–14	44
Chapter 5 — Emotional Reactions		
Introduction	5–1	45
Fear	5–2	45
Anxiety	5–3	46
Panic	5–4	46
Hate	5–5	47
Resentment	5–6	47
Anger	5–7	47
Impatience	5–8	48
Dependence	5–9	48
Loneliness	5–10	49
Boredom	5–11	49
Hopelessness	5–12	49
Summary	5–13	50
Chapter 6 — The Will to Survive		
Introduction	6–1	51
Overcoming Stress	6–2	51
Crisis Period	6–3	51
The Coping Period	6–4	52
Attitude	6–5	52
Optimism	6–6	52
Summary	6–7	52
Part Three — Basic Survival Medicine		
Chapter 7 — Survival Medicine		
Introduction	7–1	53
Procedures and Expedients	7–2	53
Hygiene	7–3	54
Care of the Mouth and Teeth	7–4	54
Care of the Feet	7–5	55
Clothing and Bedding	7–6	55
Rest	7–7	55
Rules for Avoiding Illness	7–8	56
General Management of Injuries	7–9	56
Illnesses	7–10	60
Chapter 8 – PW Medicine		
Introduction	8–1	61
History	8–2	61
Trauma	8–3	62
Gastrointestinal Problems	8–4	62
Hepatitis	8–5	63
Nutritional Deficiencies	8–6	63
Communicable Diseases	8–7	64
Skin Diseases	8–8	64
Fungal Infection	8–9	64
Dental Problems	8–10	65
Burns	8–11	65
Lacerations and Infection	8–12	65
Fraction/Sprains	8–13	66

	Paragraph	Page
Summary	8–14	67
Conclusion	8–15	68
Part Four — Facts and Conditions Affecting a Survivor		
Chapter 9 — Weather		
Introduction	9–1	69
Knowledge of Weather	9–2	69
Elements Affecting Weather	9–3	69
Circulation of the Atmosphere	9–4	71
Storms	9–5	76
Weather Forecasting	9–6	77
Summary	9–7	78
Chapter 10 — Geographic Principles		
Introduction	10–1	80
Components of the Environment	10–2	80
Effects of Climate on Terrain	10–3	83
Effects of Terrain on Climate	10–4	84
Effects of Climate and Terrain on Life Forms	10–5	85
Chapter 11 — Environmental Characteristics		
Introduction	11–1	88
The Köppen–Geiger System	11–2	88
Tropical Climates	11–3	88
Dry Climates	11–4	100
Warm Temperature Climate	11–5	105
Snow Climates	11–6	116
Ice Climates	11–7	120
Open Seas	11–8	124
Ocean Currents	11–9	126
Climates Conditions	11–10	126
Life Forms	11–11	127
Chapter 12 — Local People		
Introduction	12–1	138
Contact with People	12–2	138
Survivor's Behavior	12–3	138
Changing Political Allegiance	12–4	139
Part Five — Personal Protection		
Chapter 13 — Proper Body Temperature		
Introduction	13–1	140
Body Temperatures	13–2	141
Heat Transfer	13–3	141
Chapter 14 — Clothing		
Introduction	14–1	143
Protection	14–2	143
Clothing Materials	14–3	143
Natural Materials	14–4	143
Synthetic Materials	14–5	143
Types of Insulation	14–6	143
Insulation Measurement	14–7	144
Clothing Wear in Snow and Ice Areas	14–8	147
Care of the Feet	14–9	149
Clothing in the Summer Arctic	14–10	149
Clothing at Sea	14–11	150
Antiexposure Garments	14–12	151
Warm Oceans	14–13	153

	Paragraph	Page
Tropical Climates	14–14	153
Dry Climates	14–15	153
Chapter 15 — Shelter		
Introduction	15–1	155
Shelter Considerations	15–2	155
Location	15–3	156
Immediate Action Shelters	15–4	157
Improvised Shelters	15–5	157
Shelters for Warm Temperature Areas	15–6	157
Maintenance and Improvements	15–7	159
Construction of Specific Shelters	15–8	159
Shelter for Tropical Areas	15–9	161
Specific Shelter for Tropical Environments	15–10	165
Shelter for Dry Climates	15–11	165
Principles of Desert Shelters	15–12	168
Shelter for Snow and Ice Areas	15–13	168
General Construction Techniques	15–14	171
Shelter Living	15–15	171
Shelter Considerations for Arctic and Arctic-like Areas	15–16	175
Shelter for Open Seas	15–17	175
Chapter 16 — Firecraft		
Introduction	16–1	177
Elements of Fire	16–2	177
Fire Location	16–3	178
Fire Site Preparation	16–4	178
Firemaking with Matches (or Lighter)	16–5	179
Heat Sources	16–6	179
Firemaking with Special Equipment	16–7	183
Burning Aircraft Fuel	16–8	184
Useful Firecraft Hints	16–9	184
Fire Lays	16–10	185
Chapter 17 — Equipment		
Introduction	17–1	188
Types of Kits	17–2	188
Issued Equipment	17–3	188
Improvised Equipment	17–4	192
Miscellaneous Improvised Equipment	17–5	195
Ropes and Knots	17–6	200
Personal Survival Kits	17–7	217
Part Six — Sustenance		
Chapter 18 — Food		
Introduction	18–1	219
Nutrition	18–2	219
Food	18–3	220
Animal Food	18–4	221
Plant Food	18–5	234
Food in Tropical Climates	18–6	240
Food in Dry Climates	18–7	244
Food in Snow and Ice Climates	18–8	246
Food in Open Seas	18–9	251
Preparing Animal Food	18–10	255
Cooking	18–11	264
Preparing Food in Enemy Areas	18–12	267
Preserving Food	18–13	269

	Paragraph	Page
Preparing Plant Food	18–14	272
Chapter 19 — Water		
Introduction	19–1	275
Water Requirements	19–2	275
Water Sources	19–3	277
Water in Snow and Ice Areas	19–4	279
Water on the Open Seas	19–5	280
Water in Tropical Areas	19–6	281
Water in Dry Areas	19–7	283
Preparation of Water for Consumption	19–8	287
Part Seven — Travel		
Chapter 20 — Land Navigation		
Introduction	20–1	290
Maps	20–2	290
Aeronautical Charts	20–3	291
Information Contained in Margin	20–4	298
Topographical Map Symbols and Colors	20–5	307
Coordinate Systems	20–6	307
Elevation and Relief	20–7	318
Representative Fraction (RF)	20–8	320
Graphic (Bar) Scales	20–9	321
Using a Map and Compass, and Expressing Direction	20–10	323
Chapter 21 – Land Travel		
Introduction	21–1	352
Decision to Stay or Travel	21–2	352
Travel	21–3	353
Land Travel Techniques	21–4	353
Mountain Walking Techniques	21–5	362
Burden Carrying	21–6	363
Chapter 22 — Rough Land Travel and Evacuation Techniques		
Introduction	22–1	369
Safety Coordination	22–2	369
Climbing Ropes	22–3	369
Specialized Knots for Climbing and Evacuation	22–4	370
Special Equipment	22–5	373
The Successful Climb	22–6	379
Belaying	22–7	383
Communication	22–8	385
Anchors	22–9	386
Climbing	22–10	388
Rappelling	22–11	392
Overland Snow Travel	22–2	395
Snow and Ice Climbing Procedures and Techniques	22–13	398
Glaciers and Glacial Travel	22–14	405
Evacuation Principles and Techniques	22–15	410
Chapter 23 — Water Travel		
Introduction	23–1	427
River Travel	23–2	427
Using Safe Judgment and Rules for River Travel	23–3	427
River Hydraulics	23–4	433
Emergency Situations	23–5	442
Improvised Rafts	23–6	445
Fording Streams	23–7	445

	Paragraph	Page
Traveling on Open Seas	23–8	446
Physical Considerations	23–9	450
Life Preserver Use	23–10	451
Raft Procedures	23–11	451
Making a Landfall	23–12	456
Methods of Getting Ashore	23–13	456
Part Eight — Signaling and Recovery		
Chapter 24 — Signaling		
Introduction	24–1	460
Manufactured Signals	24–2	461
Improvised Signals	24–3	465
Chapter 25 — Recovery Principles		
Introduction	25–1	474
National Search and Rescue (SAR) Plan	25–2	474
Survivor Responsibilities	25–3	474
Recovery Site	25–4	475
Recovery Procedures	25–5	475
Pickup Devices	25–6	477
Preparation for Open Sea Recovery	25–7	480
Part Nine — Evasion		
Chapter 26 — Legal and Moral Obligations		
Introduction	26–1	482
Definitions	26–2	482
Military Drive	26–3	482
Chapter 27 — Factors of Successful Evasion		
Basic Principles	27–1	485
Preparation	27–2	485
Opportunity	27–3	486
Motivation	27–4	487
Evasion Principles	27–5	487
Chapter 28 — Camouflage		
Introduction	28–1	491
Types of Observation	28–2	491
Comparison of Direct and Indirect Observation	28–3	492
Preventing Recognition	28–4	492
Factors of Recognition	28–5	493
Principles and Methods of Camouflage	28–6	495
Individual Camouflage	28–7	499
Concealment in Various Geographic Areas	28–8	503
Concealment Factor for Areas other than Temperate	28–9	506
Camouflage Techniques for Shelter Areas	28–10	507
Firecraft Under Evasion Conditions	28–11	509
Sustenance for Evasion	28–12	512
Encampment Security Systems	28–13	513
Evasion Movement	28–14	518
Searching Terrain	28–15	519
Movement Techniques Which Limit the Potential for Detection of an Evader (Single)	28–16	520
Barriers to Evasion Movement	28–17	528
Evasion Aids	28–18	539
Combat Recovery	28–19	543

	Paragraph	Page
Part Ten — Induced Conditions — Nuclear, Biological, and Chemical (NBC)		
Chapter 29 — Nuclear Conditions		
Introduction	29–1	547
Effects of Nuclear Weapons	29–2	547
Types of Nuclear Bursts	29–3	549
Injuries	29–4	550
Types of Residual Radiation	29–5	552
Effects of Fallout on the Individual	29–6	552
Radiation Detection	29–7	553
Body Reactions to Radiation	29–8	554
Countermeasures Against Penetrating External Radiation	29–9	556
Shelter	29–10	558
Water	29–11	561
Food	29–12	562
Chapter 30 — Biological Conditions		
Introduction	30–1	568
Detection	30–2	568
Climate	30–3	568
Terrain	30–4	569
Protection	30–5	569
Tips for a Survivor	30–6	569
Chapter 31 — Chemical Conditions		
Introduction	31–1	570
Chemical Groups	31–2	570
Detection	31–3	572
Protection	31–4	573
Avoiding Chemical Agents	31–5	574
Decontamination of Chemical Agents	31–6	575
Tips for a Survivor	31–7	575

Attachment

1.	Bibliography	577

Figures

1-1.	Elements of Surviving	22
1-2.	Survival Triangle	23
2-1.	Five Basic Conditions	24
2-2.	Cold Front Occlusion	25
2-3.	Terrain	26
2-4.	Life Forms	26
2-5.	Survivor's Condition	27
2-6.	Code of Conduct	28
2-7.	Induced Conditions	29
3-1.	Survivor's Needs	30
3-2.	Personal Protection	31
3-3.	Shelters	32
3-4.	Fire	32
3-5.	Water	33
3 6.	Food	33
3-7.	Self-Aid	33
3-8.	Health and Morale	34
3-9.	Recovery	34
3-10.	Travel	35
3-11.	Burden Carrying	35
4-1.	Psychological Aspects	36

Part One

THE ELEMENTS OF SURVIVING

Chapter 1

MISSION

1-1. Introduction. An ejection sequence, a bailout, or crash landing ends one mission for the crew but starts another—to successfully return from a survival situation. Are they prepared? Can they handle the new mission, not knowing what it entails? Unfortunately, many aircrew members are not fully aware of their new mission or are not fully prepared to carry it out. All instructors teaching aircrew survival must prepare the aircrew member to face and successfully complete this new mission. (Figure 1-1 shows situations a member might encounter.)

1-2. Aircrew Mission. The moment an aircrew member leaves the aircraft and encounters a survival situation, the assigned mission is to: "return to friendly control without giving aid or comfort to the enemy, to return early and in good physical and mental condition."

a. On first impressions, "friendly control" seems to relate to a combat situation. Even in peacetime, however, the environment may be quite hostile. Imagine parachuting into the arctic when it's minus 40°F. Would an aircrew member consider this "friendly?" No. If the aircraft is forced to crash-land in the desert where temperatures may soar above 120°F, would this be agreeable? Hardly. The possibilities for encountering hostile conditions affecting human survival are endless. Crewmembers who egress an aircraft may confront situations difficult to endure.

b. The second segment of the mission, "without giving aid or comfort to the enemy," is directly related to a combat environment. This part of the mission may be most effectively fulfilled by following the moral guide—the Code of Conduct. Remember, however, that the Code of Conduct is useful to a survivor at all times and in all situations. Moral obligations apply to the peacetime situation as well as to the wartime situation.

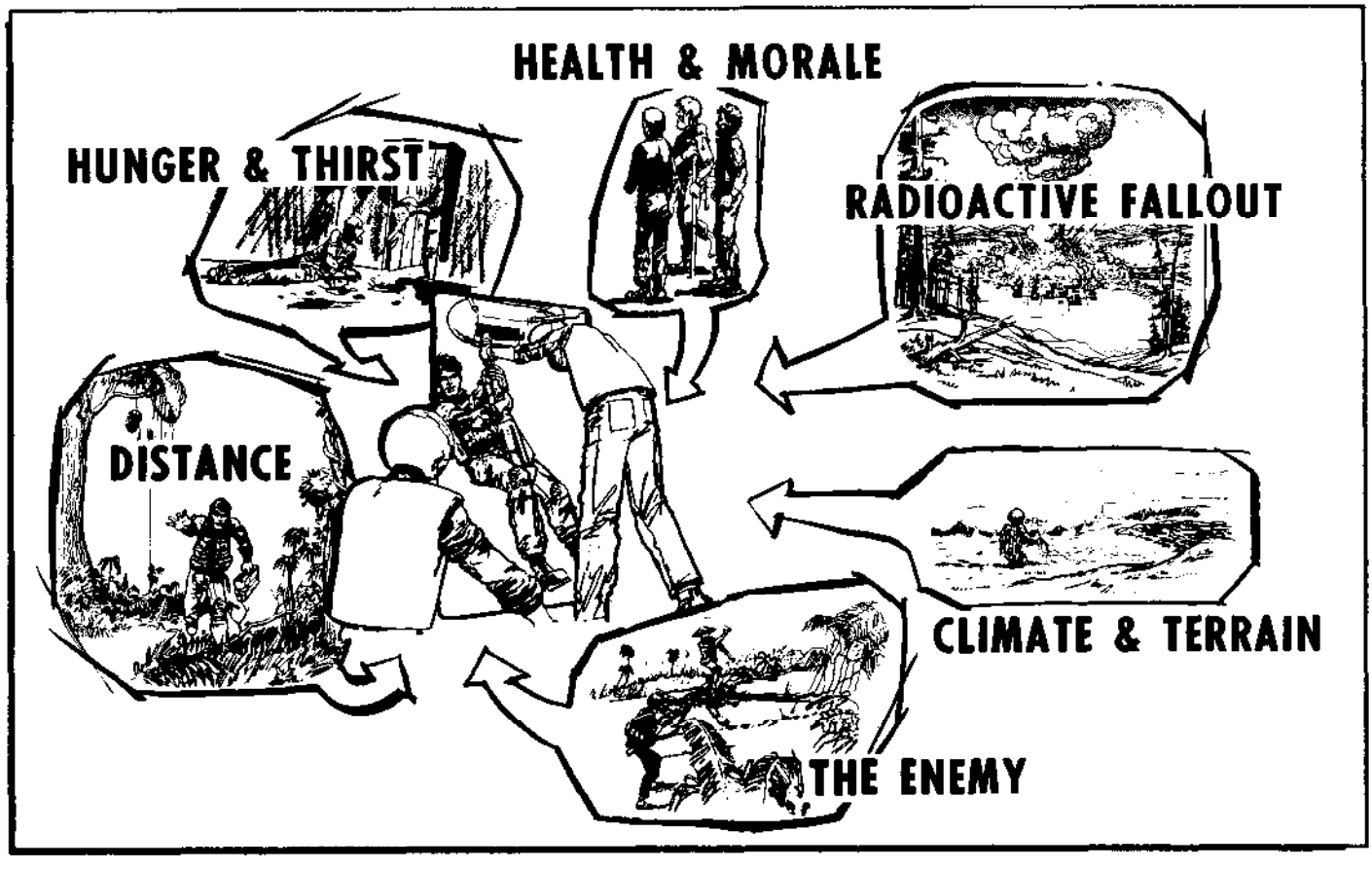

Figure 1-1. Elements of Surviving.

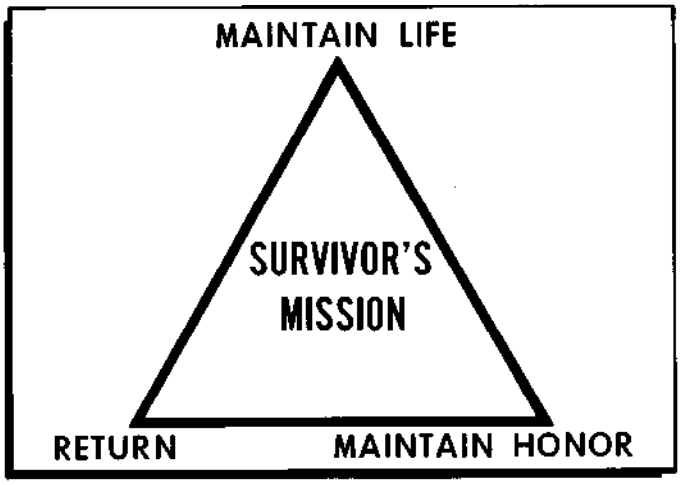

Figure 1-2. Survival Triangle.

c. The final phase of the mission is "to return early and in good physical and mental condition." A key factor in successful completion of this part of the mission may be the *will to survive.* This will is present, in varying degrees, in all human beings. Although successful survival is based on many factors, those who maintain this important attribute will increase their chance of success.

1-3. Goals. Categorizing this mission into organizational components, the three goals or duties of a survivor are to maintain life, maintain honor, and return. Survival training instructors and formal survival training courses provide training in the skills, knowledge, and attitudes necessary for an aircrew member to successfully perform the fundamental survival duties shown in figure 1-2.

1-4. Survival. Surviving is extremely stressful and difficult. The survivor may be constantly faced with hazardous and difficult situations. The stresses, hardships, and hazards (typical of a survival episode) are caused by the cumulative effects of existing conditions. (See chapter 2 pertaining to conditions affecting survival.) Maintaining life and honor and returning, regardless of the conditions, may make surviving difficult or unpleasant. The survivor's mission forms the basis for identifying and organizing the major needs of a survivor. (See survivor's needs in chapter 3.)

1-5. Decisions. The decisions survivors make and the actions taken in order to survive determine their prognosis for surviving.

1-6. Elements. The three primary elements of the survivor's mission are: the conditions affecting survival, the survivor's needs, and the means for surviving.

Chapter 2

CONDITIONS AFFECTING SURVIVAL

2-1. Introduction. Five basic conditions affect every survival situation (figure 2-1). These conditions may vary in importance or degree of influence from one situation to another and from individual to individual. At the onset, these conditions can be considered to be neutral—being neither for nor against the survivor, and should be looked upon as neither an advantage nor a disadvantage. The aircrew member may succumb to their effects—or use them to best advantage. These conditions exist in each survival episode, and they will have great bearing on the survivor's every need, decision, and action.

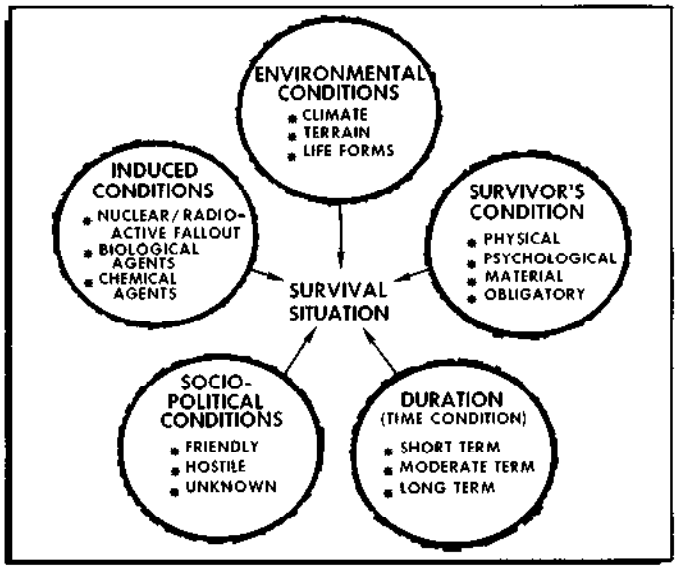

Figure 2-1. Five Basic Conditions.

2-2. Environmental Conditions. Climate, terrain, and life forms are the basic components of all environments. These components can present special problems for the survivor. Each component can be used to the survivor's advantage. Knowledge of these conditions may very well contribute to the success of the survival mission.

a. Climate. Temperature, moisture, and wind are the basic climatic elements. Extreme cold or hot temperatures, complicated by moisture (rain, humidity, dew, snow, etc.) or lack of moisture, and the possibility of wind, may have a life threatening impact on the survivor's needs, decisions, and actions. The primary concern, resulting from the effects of climate, is the need for personal protection. Climatic conditions also have a significant impact on other aspects of survival (for example, the availability of water and food, the need and ability to travel, recovery capabilities, physical and psychological problems, etc.) (figure 2-2).

b. Terrain. Mountains, prairies, hills, and lowlands, are only a few examples of the infinite variety of land forms which describe "terrain." Each of the land forms have a different effect on a survivor's needs, decisions, and actions. A survivor may find a combination of several terrain forms in a given situation. The existing terrain will affect the survivor's needs and activities in such areas as travel, recovery, sustenance, and, to a lesser extent, personal protection. Depending on its form, terrain may afford security and concealment for an evader; cause travel to be easy or difficult; provide protection from cold, heat, moisture, wind, or nuclear, biological, chemical (NBC) conditions; or make surviving a *seemingly* impossible task (figure 2-3).

c. Life Forms. For survival and survival training purposes, there are two basic life forms—plant life and animal life (other than human). NOTE: The special relationship and effects of people on the survival episode are covered separately. Geographic areas are often identified in terms of the abundance of life (or lack thereof). For example, the barren arctic or desert, primary (or secondary) forests, the tropical rain forest, the polar ice cap, etc., all produce images regarding the quantities of life forms. These examples can have special meaning not only in terms of the hazards or needs they create, but also in how a survivor can use available life forms (figure 2-4).

(1) Plant Life. There are hundreds of thousands of different types and species of plant life. In some instances, geographic areas are identified by the dominant types of plant life within that area. Examples of this are savannas, tundra, deciduous forests, etc. Some species of plant life can be used advantageously by a survivor—if not for the food or the water, then for improvising camouflage, shelter, or providing for other needs.

(2) Animal Life. Reptiles, amphibians, birds, fish, insects, and mammals are life forms which directly affect a survivor. These creatures affect the survivor by posing hazards (which must be taken into consideration), or by satisfying needs.

2-3. The Survivor's Condition. The survivor's condition and the influence it has in each survival episode is often overlooked. The primary factors which constitute the survivor's condition can best be described by the four categories shown in figure 2-5. Aircrew members must prepare themselves in each of these areas before each mission, and be in a state of "constant readiness" for the possibility of a "survival mission." Crewmembers must be aware of the role a survivor's condition plays both before and during the survival episode.

a. Physical. The physical condition and the fitness level of the survivor are major factors affecting

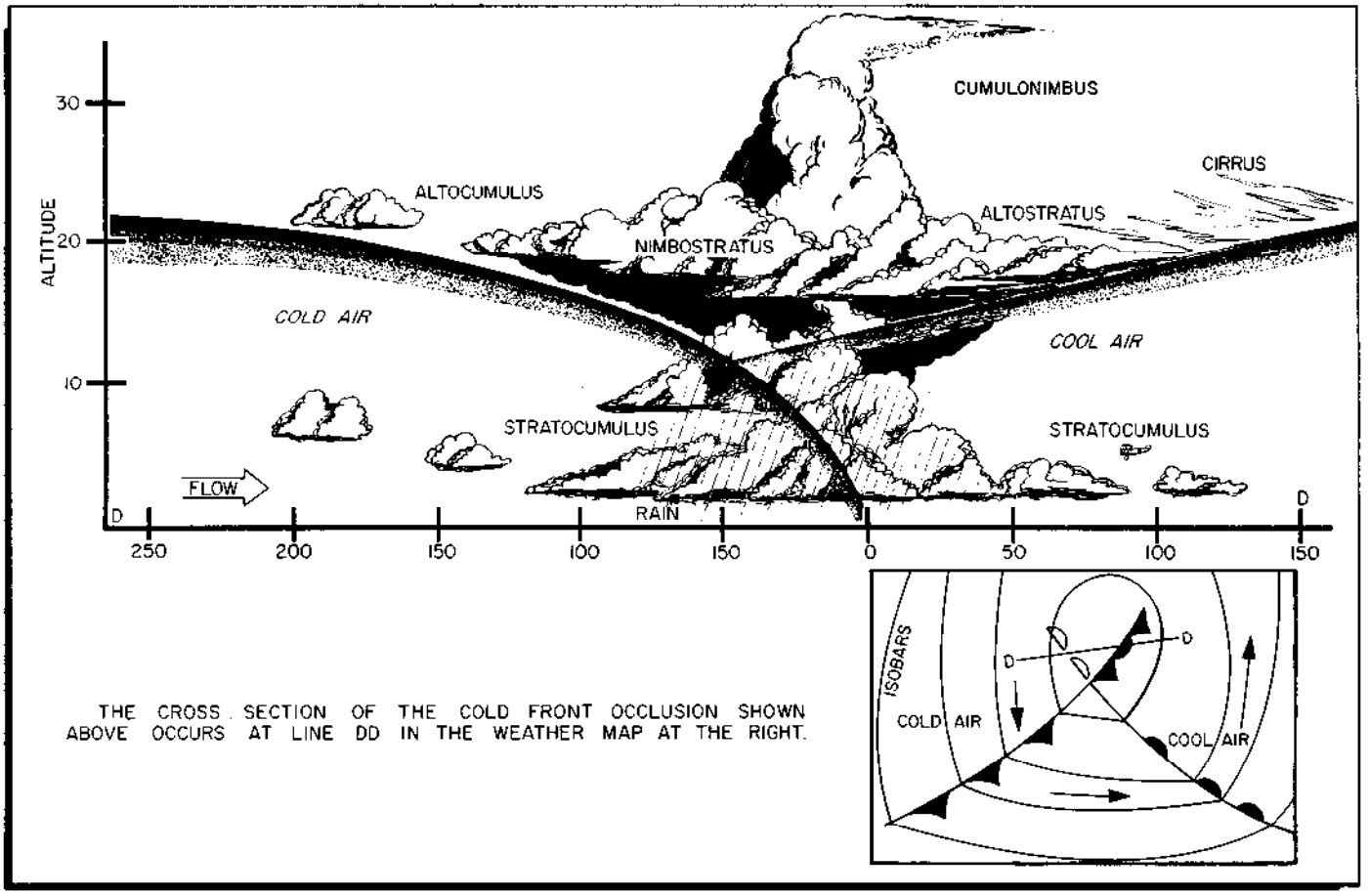

Figure 2-2. Cold Front Occlusion.

survivability. Aircrew members who are physically fit will be better prepared to face survival episodes than those who are not. Further, a survivor's physical condition (injured or uninjured) during the initial phase of a survival episode will be a direct result of circumstances surrounding the ejection, bailout, parachute landing, or crash landing. In short, high levels of physical fitness and good post-egress physical condition will enhance a survivor's ability to cope with such diverse variables as: (1) temperature extremes, (2) rest or lack of it, (3) water availability, (4) food availability, and (5) extended survival episodes. In the last instance, physical weakness may increase as a result of nutritional deficiencies, disease, etc.

b. Psychological. Survivors' psychological state greatly influences their ability to successfully return from a survival situation.

(1) Psychological effectiveness in a survival episode (including captivity) results from effectively coping with the following factors:

(a) Initial shock - Finding oneself in a survival situation following the stress of ejection, bailout, or crash landing.

(b) Pain - Naturally occurring or induced by coercive manipulation.

(c) Hunger - Naturally occurring or induced by coercive manipulation.

(d) Thirst - Naturally occurring or induced by coercive manipulation.

(e) Cold or Heat - Naturally occurring or induced by coercive manipulation.

(f) Frustration - Naturally occurring or induced by coercive manipulation.

(g) Fatigue (including Sleep Deprivation). - Naturally occurring or induced by coercive manipulation.

(h) Isolation - Includes forced (captivity) and the extended duration of any episode.

(i) Insecurity - Induced by anxiety and self-doubts.

(j) Loss of self-esteem - Most often induced by coercive manipulation.

(k) Loss of self-determination - Most often induced by coercive manipulation.

(l) Depression - Mental "lows."

(2) A survivor may experience emotional reactions during a survival episode due to the previously stated

Figure 2-3. Terrain.

factors, previous (life) experiences (including training) and the survivor's psychological tendencies. Emotional reactions commonly occurring in survival (including captivity) situations are:

(a) Boredom - sometimes combined with loneliness.

(b) Loneliness.

(c) Impatience.

(d) Dependency.

(e) Humiliation.

(f) Resentment.

(g) Anger - sometimes included as a subelement of hate.

(h) Hate.

(i) Anxiety.

(j) Fear - often included as a part of panic or anxiety.

(k) Panic.

Figure 2-4. Life Forms.

(3) Psychologically survival episodes may be divided into "crisis" phases and "coping" phases. The initial crisis period will occur at the onset of the survival situation. During this initial period, "thinking" as well as "emotional control" may be disorganized. Judgment is impaired, and behavior may be irrational (possibly to the point of panic). Once the initial crisis is under control, the coping phase begins and the survivor is able to respond positively to the situation. Crisis periods may well recur, especially during extended situations (captivity). A survivor must strive to control if avoidance is impossible.

(4) The most important psychological tool that will affect the outcome of a survival situation is the *will to survive.* Without it, the survivor is surely doomed to failure—a strong will is the best assurance of survival.

c. Material. At the beginning of a survival episode, the clothing and equipment in the aircrew member's possession, the contents of available survival kits, and salvageable resources from the parachute or aircraft are the sum total of the survivor's material assets. Adequate premission preparations are required (must be stressed during training). Once the survival episode has started, special attention must be given to the care, use, and storage of all materials to ensure they continue to be

SURVIVOR'S CONDITION

- ○ PHYSICAL
- ○ PSYCHOLOGICAL
- ○ MATERIAL
- ○ OBLIGATORY

Figure 2-5. Survivor's Condition.

serviceable and available. Items of clothing and equipment should be selectively augmented with improvised items.

(1) Clothing appropriate to anticipated environmental conditions (on the ground) should be worn or carried as aircraft space and mission permit.

(2) The equipment available to a survivor affects all decisions, needs, and actions. The survivor's ability to improvise may provide ways to meet some needs.

d. Legal and Moral Obligations. A survivor has both legal and moral obligations or responsibilities. Whether in peacetime or combat, the survivor's responsibilities as a member of the military service continues. Legal obligations are expressly identified in the Geneva Conventions, Uniform Code of Military Justice (UCMJ), and Air Force directives and policies. Moral obligations are expressed in the Code of Conduct. (See figure 2-6.)

(1) Other responsibilities influence behavior during survival episodes and influence the *will to survive.* Examples include feelings of obligation or responsibilities to family, self, and(or) spiritual beliefs.

(2) A survivor's individual perception of responsibilities influence survival needs, and affect the psychological state of the individual both during and after the survival episode. These perceptions will be reconciled either consciously through rational thought or subconsciously through attitude changes. Training specifically structured to foster and maintain positive attitudes provides a key asset to survival.

2-4. Duration—The Time Condition. The duration of the survival episode has a major effect upon the aircrew member's needs. Every decision and action will be driven in part by an assessment of when recovery or return is probable. Air superiority, rescue capabilities, the distances involved, climatic conditions, the ability to locate the survivor, or captivity are major factors which directly influence the duration (time condition) of the survival episode. A survivor can never be certain that rescue is imminent.

2-5. Sociopolitical Condition. The people a survivor contacts, their social customs, cultural heritage, and political attitudes will affect the survivor's status. Warfare is one type of sociopolitical condition, and people of different cultures are another. Due to these sociopolitical differences, the interpersonal relationship between the survivor and any people with whom contact is established is crucial to surviving. To a survivor, the attitude of the people contacted will be friendly, hostile, or unknown.

a. Friendly People. The survivor who comes into contact with friendly people, or at least those willing (to some degree) to provide aid, is indeed fortunate. Immediate return to home, family, or home station, however, may be delayed. When in direct association with even the friendliest of people, it is essential to maintain their friendship. These people may be of a completely different culture in which a commonplace American habit may be a gross and serious insult. In other instances, the friendly people may be active insurgents in their country and constantly in fear of discovery. Every survivor action, in these instances, must be appropriate and acceptable to ensure continued assistance.

b. Hostile People. A state of war need not exist for a survivor to encounter hostility in people. With few exceptions, any contact with hostile people must be avoided. If captured, regardless of the political or social reasons, the survivor must make all efforts to adhere to the Code of Conduct and the legal obligations of the UCMJ, the Geneva Conventions, and USAF policy.

c. Unknown People. The survivor should consider all factors before contacting unknown people. Some primitive cultures and closed societies still exist in which outsiders are considered a threat. In other areas of the world, differing political and social attitudes can place a survivor "at risk" in contacting unknown people.

2-6. Induced Conditions. Any form of warlike activity results in "induced conditions." Three comparatively new induced conditions may occur during combat operations. Nuclear warfare and the resultant residual radiation, biological warfare, and chemical warfare (NBC) create life-threatening conditions from which a survivor needs immediate protection. The longevity of NBC conditions further complicates a survivor's other needs, decisions, and actions (figure 2-7).

U.S. FIGHTING MAN'S

CODE OF CONDUCT

I

I am an American fighting man. I serve in the forces which guard my country and our way of life. I am prepared to give my life in their defense.

II

I will never surrender of my own free will. If in command, I will never surrender my men while they still have the means to resist.

III

If I am captured, I will continue to resist by all means available. I will make every effort to escape and aid others to escape. I will accept neither parole nor special favors from the enemy.

IV

If I become a prisoner of war, I will keep faith with my fellow prisoners. I will give no information or take part in any action which might be harmful to my comrades. If I am senior, I will take command. If not, I will obey the lawful orders of those appointed over me and will back them up in every way.

V

When questioned, should I become a prisoner of war, I am required to give name, rank, service number, and date of birth. I will evade answering further questions to the utmost of my ability. I will make no oral or written statements disloyal to my country and its allies or harmful to their cause.

VI

I will never forget that I am an American fighting man, responsible for my actions, and dedicated to the principles which made my country free. I will trust in my God and in the United States of America.

Figure 2-6. Code of Conduct.

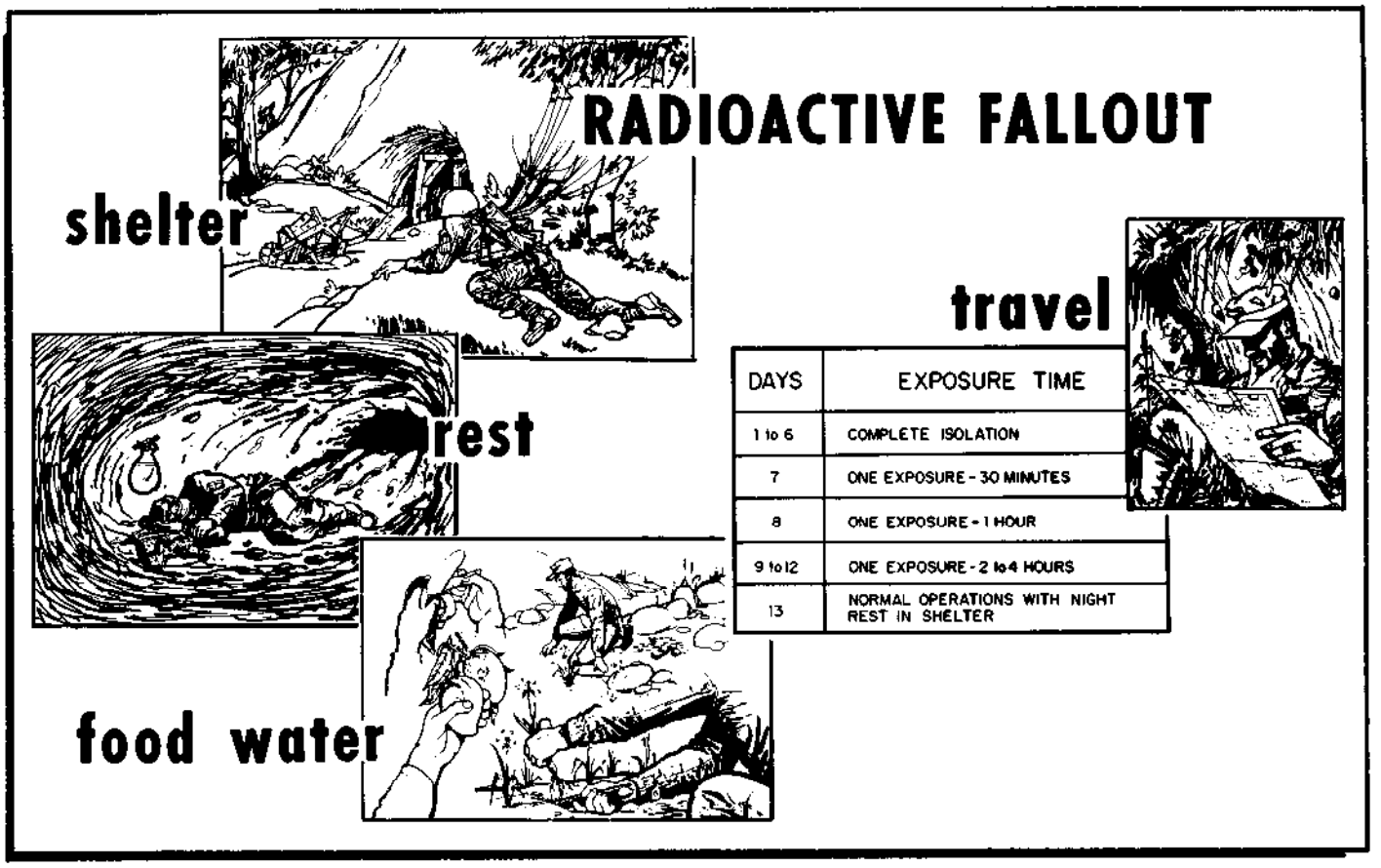

DAYS	EXPOSURE TIME
1 to 6	COMPLETE ISOLATION
7	ONE EXPOSURE - 30 MINUTES
8	ONE EXPOSURE - 1 HOUR
9 to 12	ONE EXPOSURE - 2 to 4 HOURS
13	NORMAL OPERATIONS WITH NIGHT REST IN SHELTER

Figure 2-7. Induced Conditions.

Chapter 3

THE SURVIVOR'S NEEDS

3-1. Introduction. The three fundamental goals of a survivor—to maintain life, maintain honor, and return—may be further divided into eight basic needs. In a noncombatant situation, these needs include: personal protection, sustenance, health, travel, and communications (signaling for recovery). During combat, additional needs must be fulfilled. They are: evasion, resistance if captured, and escape if captured. Meeting the individual's needs during the survival episode is essential to achieving the survivor's fundamental goals (figure 3-1).

Figure 3-1. Survivor's Needs.

3-2. Maintaining Life. Three elementary needs of a survivor in any situation which are categorized as the integral components of maintaining life are: personal protection, sustenance, and health.

a. Personal Protection. The human body is comparatively fragile. Without protection, the effects of environmental conditions (climate, terrain, and life forms) and of induced conditions (radiological, biological agents, and chemical agents) may be fatal. The survivor's primary defenses against the effects of the environment are clothing, equipment, shelter, and fire. Additionally, clothing, equipment, and shelter are the primary defenses against some of the effects of induced conditions (figure 3-2).

(1) The need for adequate clothing and its proper care and use cannot be overemphasized. The human body's tolerance for temperature extremes is very limited. However, its ability to regulate heating and cooling is extraordinary. The availability of clothing and its proper use is extremely important to a survivor in using these abilities of the body. Clothing also provides excellent protection against the external effects of alpha and beta radiation, and may serve as a shield against the external effects of some chemical or biological agents.

(2) Survival equipment is designed to aid survivors throughout their episode. It must be cared for to maintain its effectiveness. Items found in a survival kit or aircraft can be used to help satisfy the eight basic needs. Quite often, however, a survivor must improvise to overcome an equipment shortage or deficiency.

(3) The survivor's need for shelter is twofold—as a place to rest and for protection from the effects of the environmental and(or) induced conditions (NBC). The duration of the survival episode will have some effect on shelter choice. In areas that are warm and dry, the survivor's need is easily satisfied using natural resting places. In cold climates, the criticality of shelter can be measured in minutes, and rest is of little immediate concern. Similarly, in areas of residual radiation, the criticality of shelter may also be measured in minutes (figure 3-3).

(4) Fire serves many survivor needs: purifying water, cooking and preserving food, signaling, and providing a source of heat to warm the body and dry clothing (figure 3-4).

b. Sustenance. Survivors need food and water to maintain normal body functions and to provide strength, energy, and endurance to overcome the physical stresses of survival.

(1) Water. The survivor must be constantly aware of the body's continuing need for water (figure 3-5).

(2) Food. During the first hours of a survival situation, the need for food receives little attention. During the first 2 or 3 days, hunger becomes a nagging aggravation which a survivor can overcome. The first major food crisis occurs when the loss of energy, stamina, and strength begin to affect the survivor's physical capabilities. The second major food crisis is more insidious. A marked increase in irritability and other attitudes may occur as the starvation process continues. Early and continuous attention must be given to obtaining and using any and all available food. Most people have food preferences. The natural tendency to avoid certain types of food is a major problem which must be overcome early in the survival situation. The starvation process ultimately overcomes all food aversions. The successful survivor overcomes these aversions before physical or psychological deterioration sets in (figure 3-6).

c. Health (Physical and Psychological). The survivor must be the doctor, nurse, corpsman, psychologist, and cheerleader. Self-aid is the survivor's sole recourse.

(1) Prevention. The need for preventive medicine and safety cannot be overemphasized. Attention to sani-

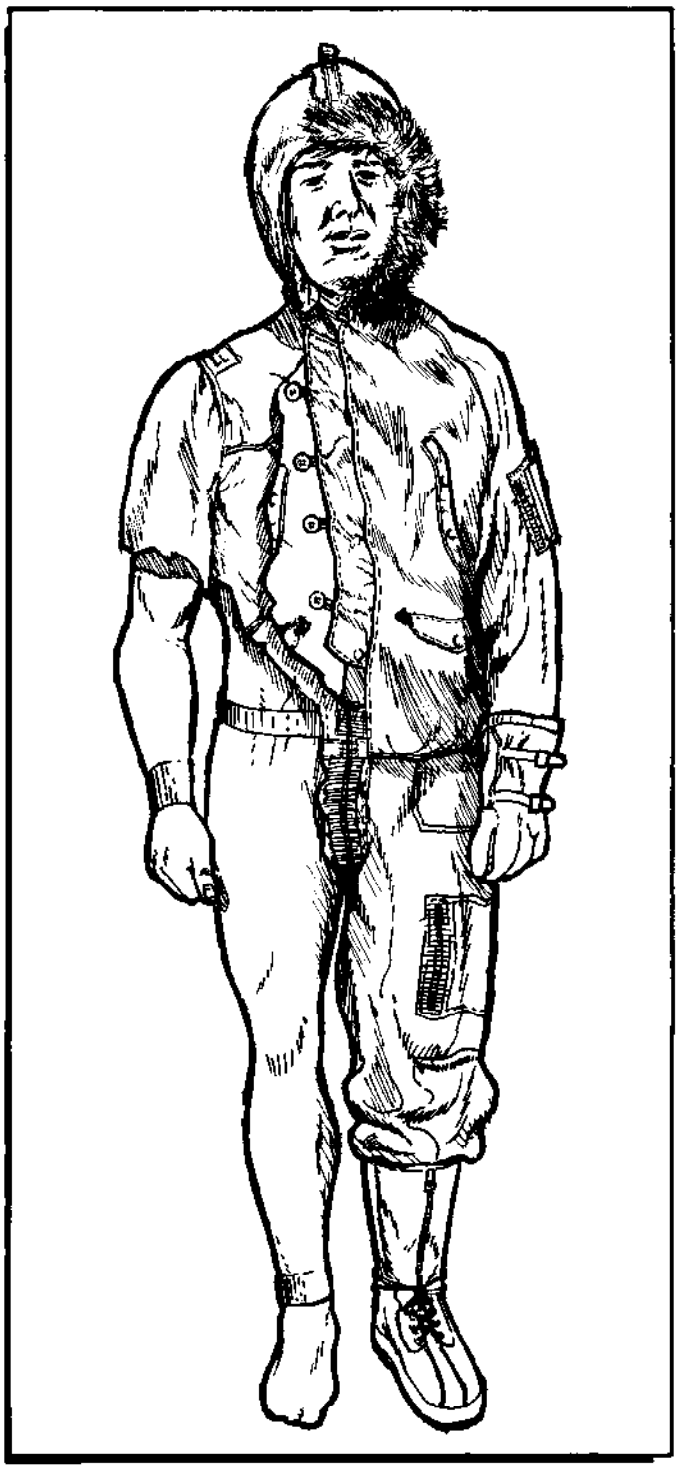

Figure 3-2. Personal Protection.

tation and personal hygiene is a major factor in preventing physical, morale, and attitudinal problems.

(a) The need for cleanliness in the treatment of injuries and illness is self-evident. The prisoner of war (PW) who used maggots to eat away rotting flesh caused by infection is a dramatic example. Prevention is much more preferred than such drastic procedures.

(b) Safety must be foremost in the mind of the survivor; carelessness is caused by ignorance and(or) poor judgment or bad luck. One miscalculation with a knife or ax can result in self-inflicted injury or death.

(2) Self-Aid:

(a) Injuries frequently occur during ejection, bailout, parachute landing, or ditching. Other post-egress factors may also cause injury. In the event of injury, the survivor's existence may depend on the ability to perform self-aid. In many instances, common first aid procedures will suffice; in others, more primitive techniques will be required (figure 3-7).

(b) Illness and the need to treat it is more commonly associated with long-term situations such as an extended evasion episode or captivity. When preventive techniques have failed, the survivor must treat symptoms of disease in the absence of professional medical care.

(3) Psychological Health. Perhaps the survivor's greatest need is the need for emotional stability and a positive, optimistic attitude. An individual's ability to cope with psychological stresses will enhance successful survival. Optimism, determination, dedication, and humor, as well as many other psychological attributes, are all helpful for a survivor to overcome psychological stresses (figure 3-8).

3-3. Maintaining Honor. Three elementary needs which a survivor may experience during combat survival situations are categorized as integral components of maintaining honor. These three elementary needs are: (a) avoiding capture or evading, (b) resisting (if captured), and (c) escaping (if captured).

a. Avoiding Capture. Evasion will be one of the most difficult and hazardous situations a survivor will face. However difficult and hazardous evasion may be, captivity is always worse. During an evasion episode, the survivor has two fundamental tasks. The first is to use concealment techniques. The second is to use evasion movement techniques. The effective use of camouflage is common to both of these activities.

(1) Hiding oneself and all signs of presence are the evader's greatest needs. Experience indicates that the survivor who uses effective concealment techniques has a better chance of evading capture. Capture results most frequently when the evader is moving.

(2) The evader's need to move depends on a variety of needs such as recovery, food, water, better shelter, etc. Evasion movement is more successful when proven techniques are used.

b. Resisting. The PW's need to resist is self-evident. This need is both a moral and a legal obligation. Resistance is much more than refusing to divulge some bit of classified information. Fundamentally, resisting is two distinctly separate behaviors expected of the prisoner:

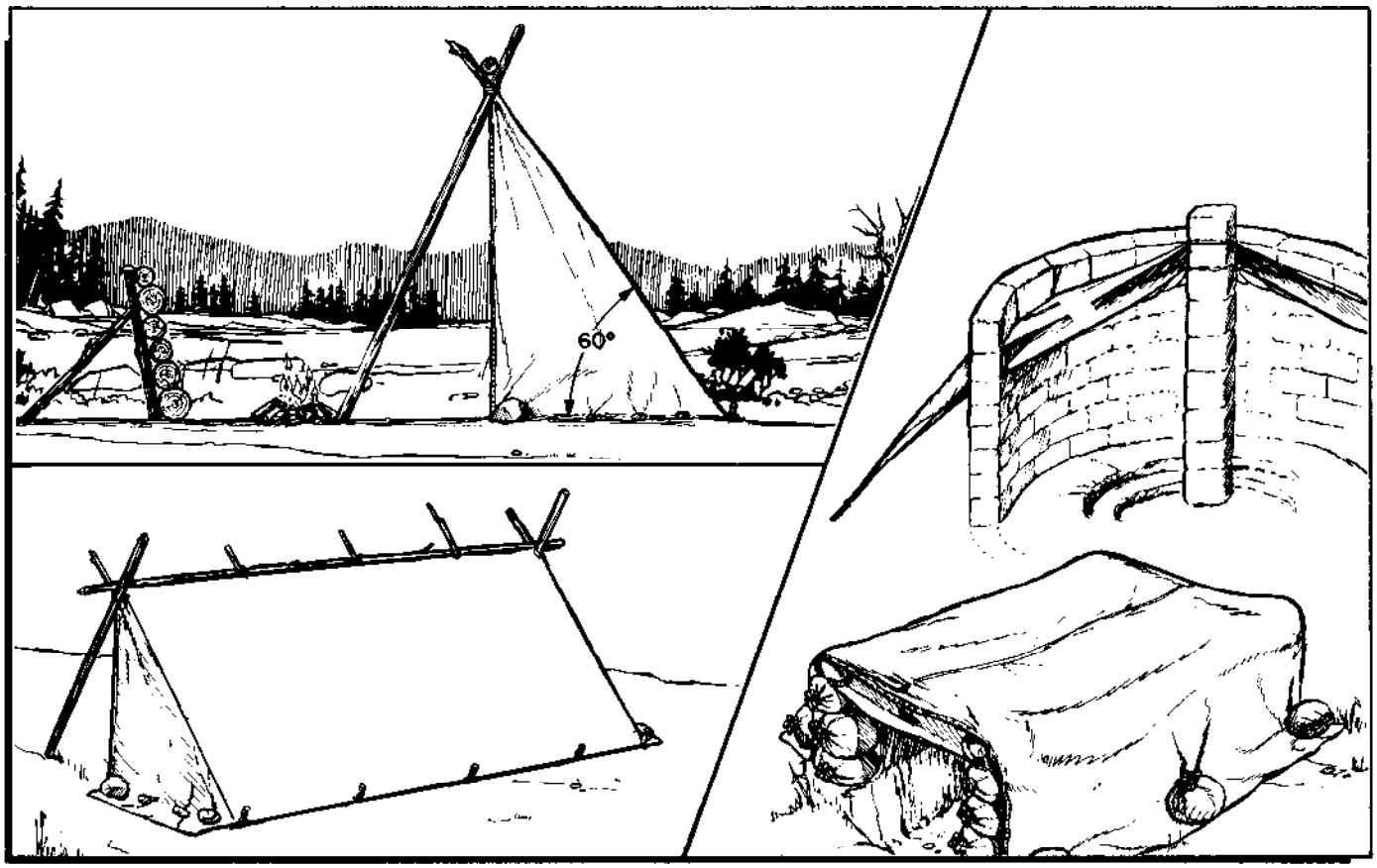

Figure 3-3. Shelters.

(1) Complying with legal and authorized requirements only.

(2) Disrupting enemy activity through resisting, subtle harrassment, and tying up enemy guards who could be used on the front lines.

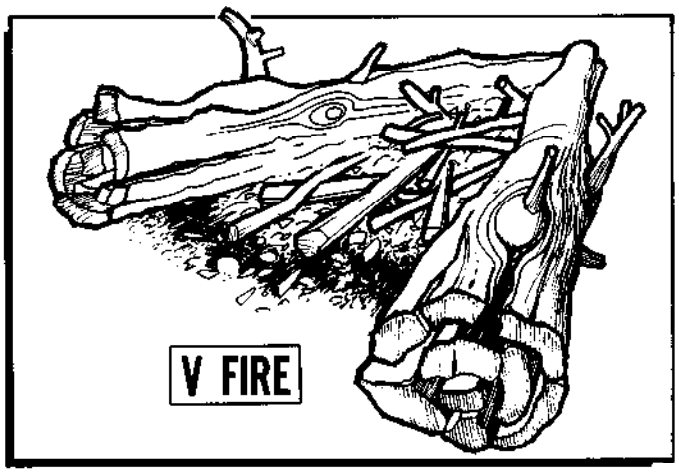

Figure 3-4. Fire.

c. Escaping (When Possible and Authorized). Escape is neither easy nor without danger. The Code of Conduct states a survivor should make every effort to escape and aid others to escape.

3-4. Returning. The need to return is satisfied by successful completion of one or both of the basic tasks confronting the survivor: aiding with recovery and traveling (on land or water).

a. Aiding With Recovery. For survivors or evaders to effectively aid in recovery, they must be able to make their position and the situation on the ground known. This is done either electronically, visually, or both (figure 3-9).

(1) Electronic signaling covers a wide spectrum of techniques. As problems such as security and safety during combat become significant factors, procedures for using electronic signaling to facilitate recovery become increasingly complex.

(2) Visual signaling is primarily the technique for attracting attention and pinpointing an exact location for rescuers. Simple messages or information may also be transmitted with visual signals.

b. Travel On Land. A survivor may need to move on land for a variety of reasons, ranging from going for

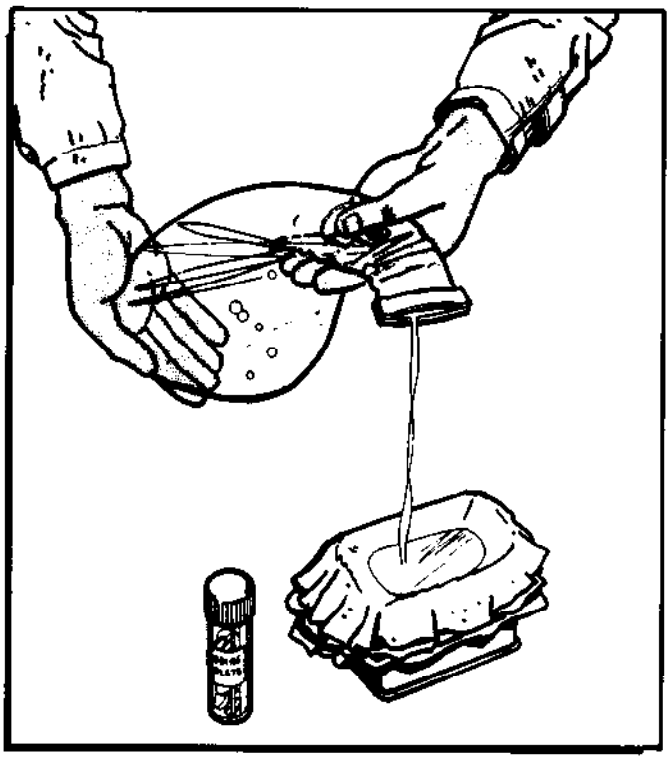

Figure 3-5. Water.

water to attempting to walk out of the situation. In any survival episode, the survivor must weigh the need to travel against capabilities and(or) safety (figure 3-10). Factors to consider may include:

Figure 3-6. Food.

(1) The ability to walk or traverse existing terrain. In a nonsurvival situation, a twisted or sprained ankle is an inconvenience accompanied by some temporary pain and restricted activity. A survivor who loses the mobility, due to injury, to obtain food, water, and shelter, can face death. There is a safe and effective way to travel across almost any type of terrain.

Figure 3-7. Self-Aid.

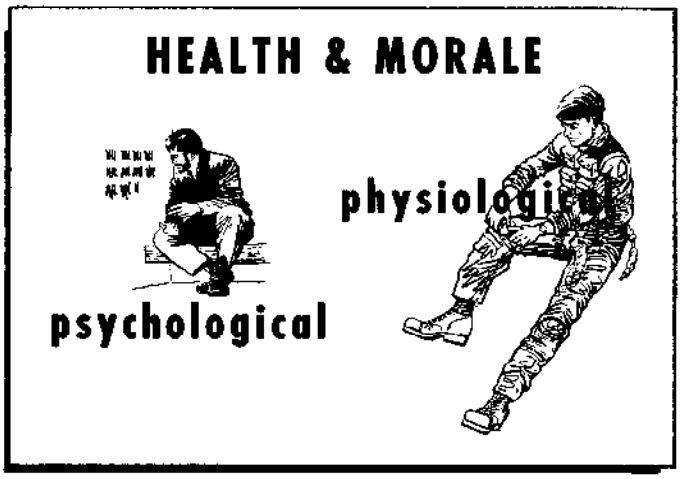

Figure 3-8. Health and Morale.

(2) The need to transport personal possessions (burden carrying). There are numerous documented instances of survivors abandoning equipment and clothing simply because carrying was a bother. Later, the abandoned materials were not available when needed to save life, limb, or aid in rescue. Burden carrying need not be difficult or physically stressful. There are many simple ways for a survivor to carry the necessities of life (figure 3-11).

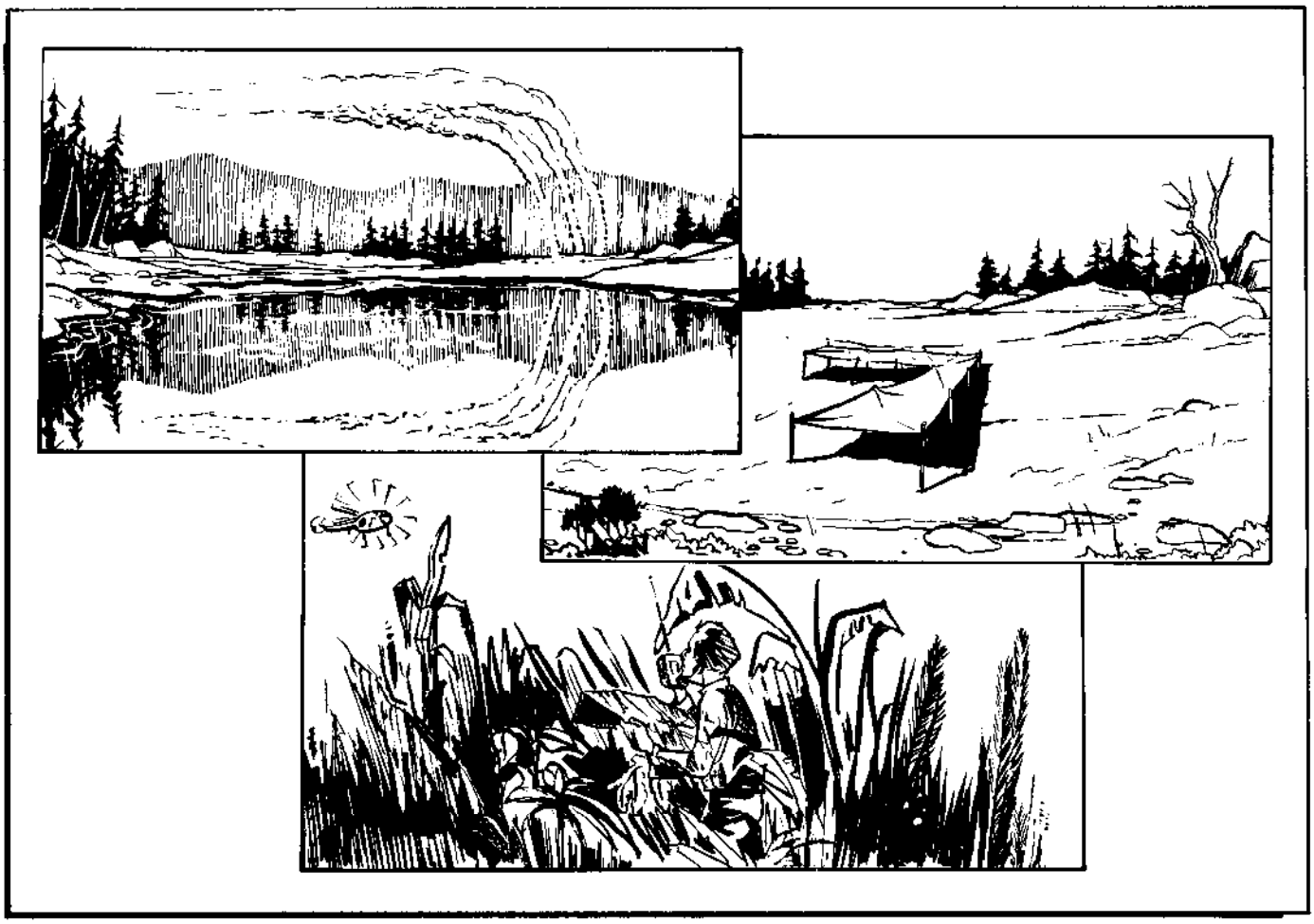

Figure 3-9. Recovery.

(3) The ability to determine present position. Maps, compasses, star charts, Weems plotters, etc., permit accurate determination of position during extended travel. Yet, the knowledgeable, skillful, and alert survivor can do well without a full complement of these aids. Constant awareness, logic, and training in nature's clues to navigation may allow a survivor to determine general location even in the absence of detailed navigation aids.

(4) Restrictions or limitations to select and maintain a course of travel. The tools used in determining position are the tools used to maintain a course of travel. A straight line course to a destination is usually the simplest, but may not always be the best course for travel. Travel courses may need to be varied for diverse reasons, such as to get food or water, to enhance covert travels or to avoid hazardous or impassable obstacles or terrain. Careful planning and route selection before and during travel is essential.

c. Travel On Water. Two differing circumstances may require survivors to travel on water. First, those who crash-land or parachute into the open sea are confronted with one type of situation. Second, survivors who find a river or stream which leads in a desirable direction are faced with a different situation. In each instance, however, a common element is to stay afloat.

(1) The survivor's initial problems on the open sea are often directly related to the winds and size of the waves. Simply getting into a liferaft and staying there are often very difficult tasks. On the open sea, the winds and ocean currents have a significant effect on the direction of travel. As the survivor comes closer to shore, the direction in which the tide is flowing also becomes a factor. There are some techniques a survivor can use to aid with stabilizing the raft, controlling the direction and rate of travel, and increasing safety.

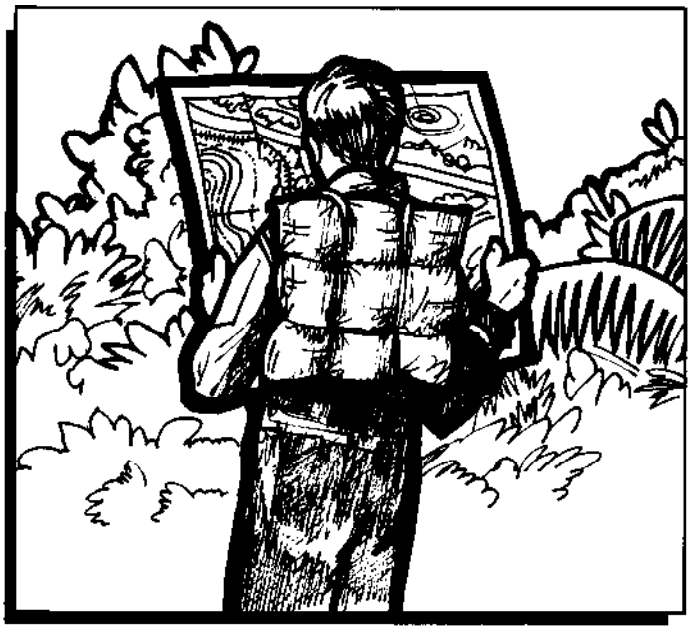

Figure 3-10. Travel.

(2) Survivors using rivers or streams for travel face both hazards and advantages as compared to overland travel. First, floating with the current is far less difficult than traveling overland. An abundance of food and water are usually readily available. Even in densely forested areas, effective signaling sites are generally available along streambeds. A survivor must use care and caution to avoid drowning, the most serious hazard associated with river travel.

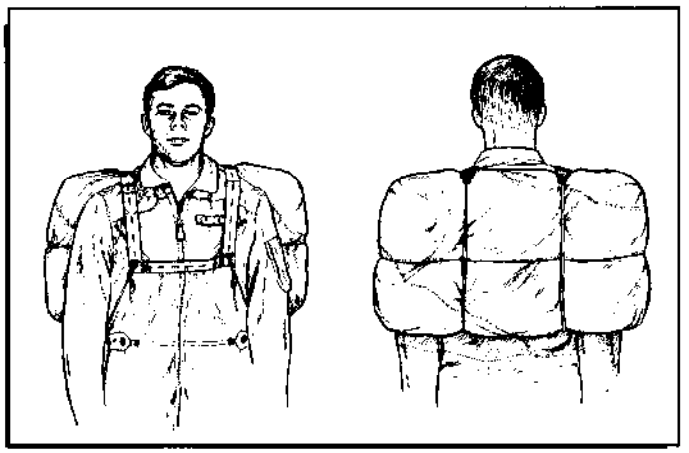

Figure 3-11. Burden Carrying.

Part Two

PSYCHOLOGICAL ASPECTS OF SURVIVAL

Chapter 4

CONTRIBUTING FACTORS

4-1. Introduction. Aircrew members in a survival situation must recognize that coping with the psychological aspects of survival are at least as important as handling the environmental factors. In virtually any survival episode, the aircrew will be in an environment that can support human life. The survivors' problems will be compounded because they never really expected to bail out or crash-land in the jungle, over the ocean, or anywhere else. No matter how well prepared, aircrews probably will never completely convince themselves that "it can happen to them." However, the records show it can happen. Before aircrew members learn about the physical aspects of survival, they must first understand that psychological problems may occur and that solutions to those problems must be found if the survival episode is to reach a successful conclusion (figure 4-1).

4-2. Survival Stresses:

a. The emotional aspects associated with survival must be completely understood just as survival conditions and equipment are understood. An important factor bearing on success or failure in a survival episode is the individual's psychological state. Maintaining an even, positive psychological state or outlook depends on the individual's ability to cope with many factors. Some include:

(1) Understanding how various physiological and emotional signs, feelings, and expressions affect one's bodily needs and mental attitude.

(2) Managing physical and emotional reactions to stressful situations.

(3) Knowing individual tolerance limits, both psychological and physical.

(4) Exerting a positive influence on companions.

b. Nature has endowed everyone with biological mechanisms which aid in adapting to stress. The bodily changes resulting from fear and anger, for example, tend to increase alertness and provide extra energy to either run away or fight. These and other mechanisms can hinder a person under survival conditions. For instance, a survivor in a raft could cast aside reason and drink sea water to quench a thrist; or, evaders in enemy territory, driven by hunger pangs, could expose themselves to capture when searching for food. These examples illustrate how "normal" reactions to stress could create problems for a survivor.

c. Two of the gravest threats to successful survival are concessions to comfort and apathy. Both threats represent attitudes which must be avoided. To survive, a person must focus planning and effort on fundamental needs.

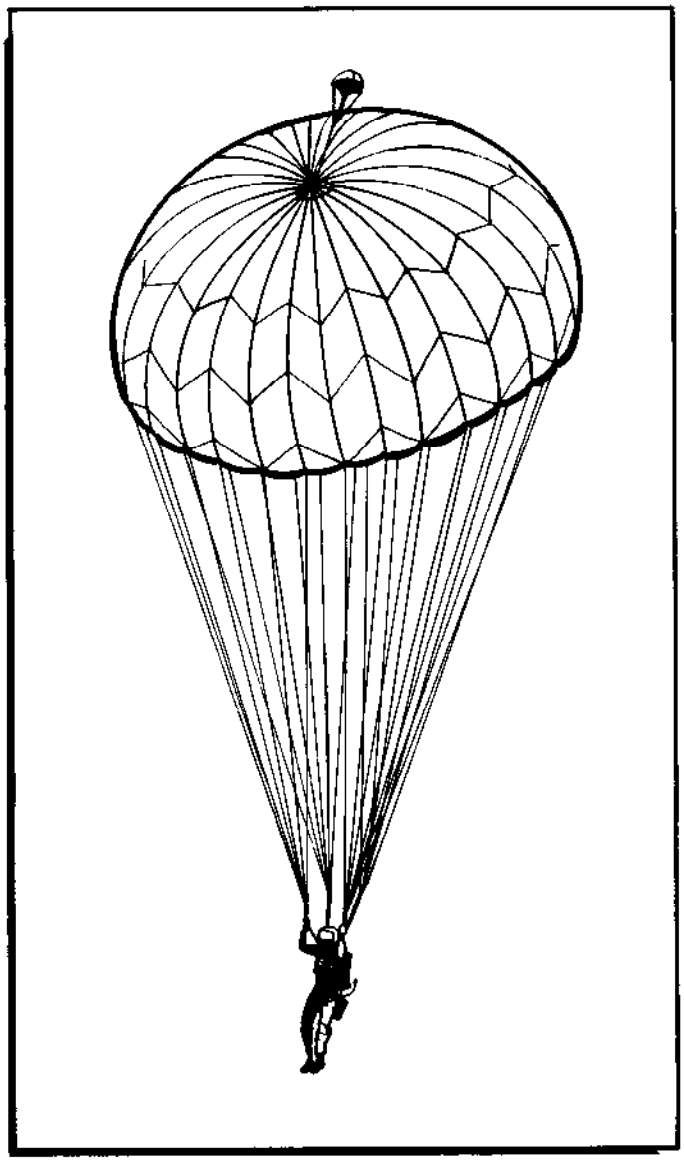

Figure 4-1. Psychological Aspects.

(1) Many people consider comfort their greatest need. Yet, comfort is not essential to human survival. Survivors must value life more than comfort, and be willing to tolerate heat, hunger, dirt, itching, pain, and any other discomfort. Recognizing discomfort as tem-

Figure 4-2. Pain.

porary will help survivors concentrate on effective action.

(2) As the will to keep trying lessens, drowsiness, mental numbness, and indifference will result in apathy. This apathy usually builds on slowly, but ultimately takes over and leaves a survivor helpless. Physical factors can contribute to apathy. Exhaustion due to prolonged exposure to the elements, loss of body fluids (dehydration), fatigue, weakness, or injury are all conditions which can contribute to apathy. Proper planning and sound decisions can help a survivor avoid these conditions. Finally, survivors must watch for signs of apathy in companions and help prevent it. The first signs are resignation, quietness, lack of communication, loss of appetite, and withdrawal from the group. Preventive measures could include maintaining group morale by planning, activity, and getting the organized participation of all members.

d. Many common stresses cause reactions which can be recognized and dealt with appropriately in survival situations. A survivor must understand that stresses and reactions often occur at the same time. Although survivors will face many stresses, the following common stresses will occur in virtually all survival episodes: pain, thirst, cold and heat, hunger, frustration, fatigue, sleep deprivation, isolation, insecurity, loss of self-esteem, loss of self-determination, and depression.

4-3. Pain:

a. Pain, like fever, is a warning signal calling attention to an injury or damage to some part of the body. Pain is discomforting but is not, in itself, harmful or dangerous. Pain can be controlled, and in an extremely grave situation, survival must take priority over giving in to pain (figure 4-2).

b. The biological function of pain is to protect an injured part by warning the individual to rest it or avoid using it. In a survival situation, the normal pain warnings may have to be ignored in order to meet more critical needs. People have been known to complete a fight with a fractured hand, to run on a fractured or sprained ankle, to land an aircraft despite severely burned hands, and to ignore pain during periods of intense concentration and determined effort. Concentration and intense effort can actually stop or reduce feelings of pain. Sometimes this concentration may be all that is needed to survive.

c. A survivor must understand the following facts about pain:

(1) Despite pain, a survivor can move in order to live.

(2) Pain can be reduced by:

(a) Understanding its source and nature.

(b) Recognizing pain as a discomfort to be tolerated.

(c) Concentrating on necessities like thinking, planning, and keeping busy.

(d) Developing confidence and self-respect. When personal goals are maintaining life, honor, and returning, and these goals are valued highly enough, a survivor can tolerate almost anything.

4-4. Thirst and Dehydration:

a. The lack of water and its accompanying problems of thirst and dehydration are among the most critical problems facing survivors. Thirst, like fear and pain, can be tolerated if the will to carry on, supported by calm, purposeful activity, is strong. Although thirst indicates the body's need for water, it does not indicate how much water is needed. If a person drinks only enough to satisfy thirst, it is still possible to slowly dehydrate. Prevention of thirst and the more debilitating dehydration is possible if survivors drink plenty of water any time it is available, and especially when eating (figure 4-3).

b. When the body's water balance is not maintained, thirst and discomfort result. Ultimately, a water imbalance will result in dehydration. The need for water may be increased if the person:

(1) Has a fever.

(2) Is fearful.

(3) Perspires unnecessarily.

(4) Rations water rather than sweat.

c. Dehydration decreases the body's efficiency or ability to function. Minor degrees of dehydration may not noticeably affect a survivor's performance, but as it becomes more severe, body functioning will become increasingly impaired. Slight dehydration and thirst can

Figure 4-3. Thirst.

also cause irrational behavior. One survivor described it:

> "The next thing I remember was being awakened by an unforgettable sensation of thirst. I began to move about aimlessly and finally found a pool of water."
>
> "We finally found water. In the water were two dead deer with horns locked. We went down to the water and drank away. It was the best damned drink of water I ever had in my life. I didn't taste the stench of the deer at all."

While prevention is the best way to avoid dehydration, virtually any degree of dehydration is reversible simply by drinking water.

4-5. Cold and Heat. The average normal body temperature for a person is 98.6°F. Victims have survived a body temperature as low as 20°F below normal, but consciousness is clouded and thinking numbed at a much smaller drop. An increase of 6 to 8 degrees above normal for any prolonged period may prove fatal. Any deviation from normal temperature, even as little as 1 or 2 degrees, reduces efficiency.

a. Cold is a serious stress since even in mild degrees it lowers efficiency. Extreme cold numbs the mind and dulls the will to do anything except get warm again. Cold numbs the body by lowering the flow of blood to the extremities, and results in sleepiness. Survivors have endured prolonged cold and dampness through exercise, proper hygiene procedures, shelter, and food. Wearing proper clothing and having the proper climatic survival equipment when flying in cold weather areas are essential to enhance survivability (figure 4-4).

(1) One survivor described cold and its effect:

> "Because of the cold water, my energy was going rapidly and all I could do was to hook my left arm over one side of the raft, hang on, and watch the low flying planes as they buzzed me...As time progressed, the numbing increased...and even seemed to impair my thinking."

(2) Another survivor remembered survival training and acted accordingly:

> "About this time, my feet began getting cold. I remembered part of the briefing I had received about feet freezing so I immediately took action. I thought about my shoes, and with my jack knife, cut off the bottom of my Mark II immersion suit and put them over my shoes. My feet immediately felt warmer and the rubber feet of the immersion suit kept the soles of my shoes dry."

b. Just as "numbness" is the principal symptom of cold, "weakness" is the principle symptom of heat. Most people can adjust to high temperatures, whether in the hold of a ship or in a harvest field on the Kansas prairie. It may take from 2 days to a week before circulation, breathing, heart action, and sweat glands are all adjusted to a hot climate. Heat stress also accentuates dehydration, which was discussed earlier. In addition to the problem of water, there are many other sources of discomfort and impaired efficiency which are directly attributable to heat or to the environmental conditions in hot climates. Extreme temperature changes, from extremely hot days to very cold nights, are experienced in desert and plains areas. Proper use of clothing and shelters can decrease the adverse effects of such extremes (figure 4-5).

c. Bright sun has a tremendous effect on eyes and any exposed skin. Direct sunlight or rays reflecting off the terrain require dark glasses or improvised eye protec-

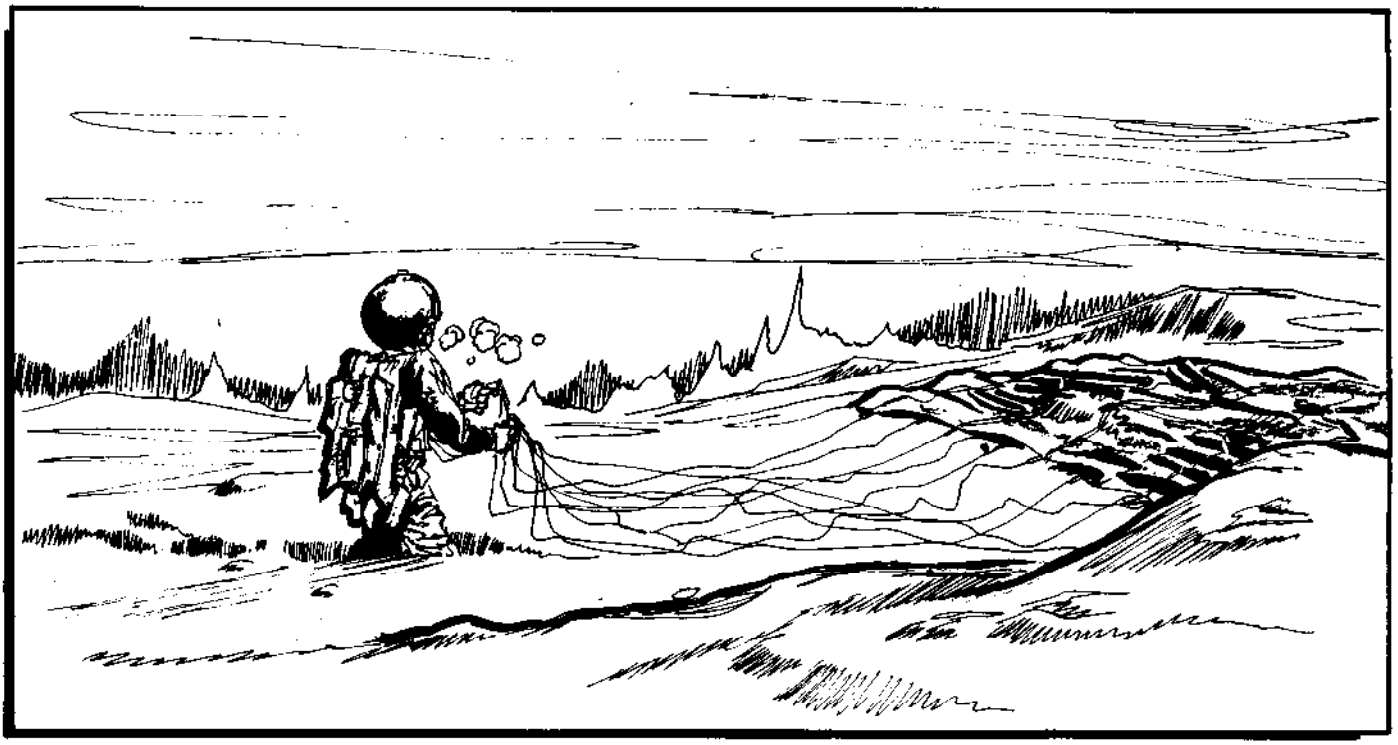

Figure 4-4. Cold.

tors. Previous suntanning provides little protection; protective clothing is important.

d. Blowing wind, in hot summer, has been reported to get on some survivors' nerves. Wind can constitute an

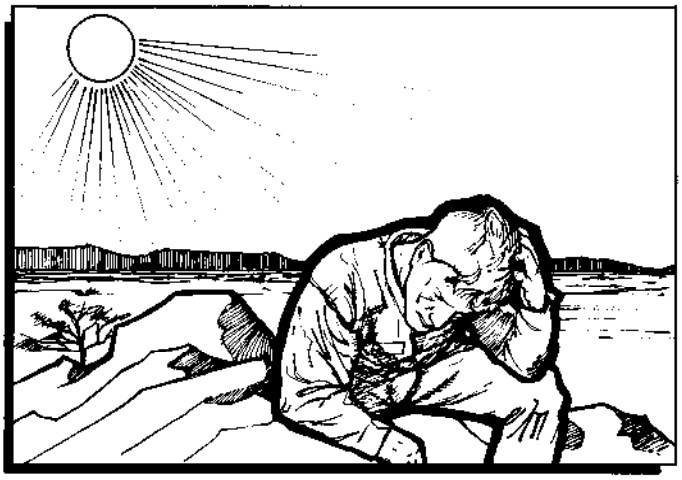

Figure 4-5. Heat.

additional source of discomfort and difficulty in desert areas when it carries particles of sand and dirt. Protection against sand and dirt can be provided by tying a cloth around the head after cutting slits for vision.

e. Acute fear has been experienced among survivors in sandstorms and snowstorms. This fear results from both the terrific impact of the storm itself and its obliteration of landmarks showing direction of travel. Finding or improving shelter for protection from the storm itself is important.

f. Loss of moisture, drying of the mouth and mucous membranes, and accelerate dehydration can be caused by breathing through the mouth and talking. Survivors must learn to keep their mouths shut in desert winds as well as in cold weather.

g. Mirages and illusions of many kinds are common in desert areas. These illusions not only distort visual perception but sometimes account for serious incidents. In the desert, distances are usually greater than they appear and, under certain conditions, mirages obstruct accurate vision. Inverted reflections are a common occurrence.

4-6. Hunger. A considerable amount of edible material (which survivors may not initially regard as food) may be available under survival conditions. Hunger and semistarvation are more commonly experienced among survivors than thirst and dehydration. Research has revealed no evidence of permanent damage nor any decrease in mental efficiency from short periods of total fasting (figure 4-6).

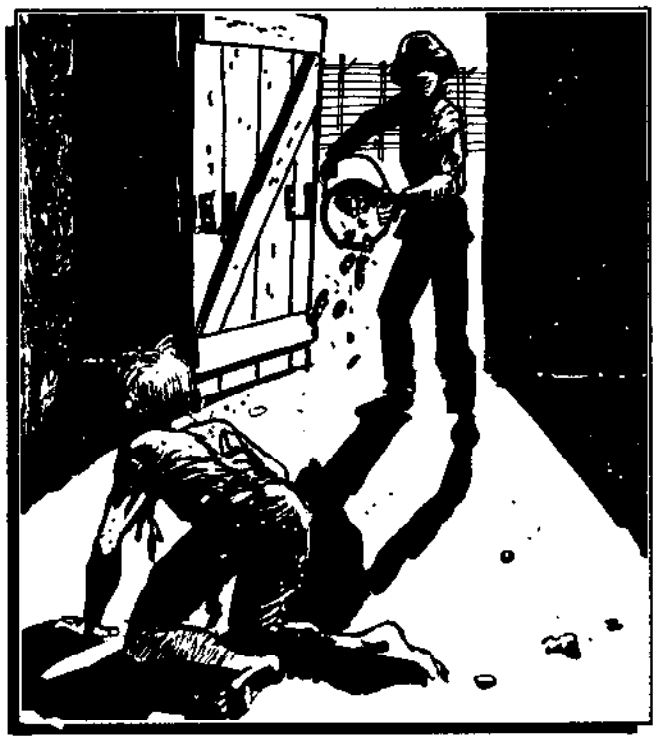

Figure 4-6. Hunger.

a. The prolonged and rigorous Minnesota semistarvation studies during World War II revealed the following behavioral changes:

(1) Dominance of the hunger drive over other drives.

(2) Lack of spontaneous activity.

(3) Tired and weak feeling.

(4) Inability to do physical tasks.

(5) Dislike of being touched or caressed in any way.

(6) Quick susceptibility to cold.

(7) Dullness of all emotional responses (fear, shame, love, etc.)

(8) Lack of interest in others—apathy.

(9) Dullness and boredom.

(10) Limited patience and self-control.

(11) Lack of a sense of humor.

(12) Moodiness—reaction of resignation.

b. Frequently, in the excitement of some survival, evasion, and escape episodes, hunger is forgotten. Survivors have gone for considerable lengths of time without food or awareness of hunger pains. An early effort should be made to procure and consume food to reduce the stresses brought on by food deprivation. Both the physical and psychological effects described are reversed when food and a protective environment are restored. Return to normal is slow and the time necessary for the return increases with the severity of starvation. If food deprivation is complete and only water is ingested, the pangs of hunger disappear in a few days, but even then the mood changes of depression and irritability occur. The individual tendency is still to search for food to prevent starvation and such efforts might continue as long as strength and self-control permit. When the food supply is limited, even strong friendships are threatened.

c. Food aversion may result in hunger. Adverse group opinion may discourage those who might try foods unfamiliar to them. In some groups, the barrier would be broken by someone eating the particular food rather than starving. The solitary individual has only personal prejudices to overcome and will often try strange foods.

d. Controlling hunger during survival episodes is relatively easy if the survivor can adjust to discomfort and adapt to primitive conditions. This man would rather survive than be fussy:

> "Some men would almost starve before eating the food. There was a soup made of lamb's head with the lamb's eyes floating around in it....When there was a new prisoner, I would try to find a seat next to him so I could eat the food he refused."

4-7. Frustration. Frustration occurs when one's efforts are stopped, either by obstacles blocking progress toward a goal or by not having a realistic goal. It can also occur if the feeling of self-worth or self-respect is lost (figure 4-7).

a. A wide range of obstacles, both environmental and internal, can lead to frustration. Frustrating conditions often create anger, accompanied by a tendency to attack and remove the obstacles to goals.

b. Frustration must be controlled by channeling energies into a positive and worthwhile obtainable goal. The survivor should complete the easier tasks before attempting more challenging ones. This will not only instill self-confidence, but also relieve frustration.

4-8. Fatigue. In a survival episode, a survivor must continually cope with fatigue and avoid the accompanying strain and loss of efficiency. A survivor must be aware of the dangers of over-exertion. In many cases, a survivor may already be experiencing strain and reduced efficiency as a result of other stresses such as heat or cold, dehydration, hunger, or fear. A survivor must judge capacity to walk, carry, lift, or do necessary work, and plan and act accordingly. During an emergency, considerable exertion may be necessary to cope with the situation. If an individual understands fatigue and the attitudes and feelings generated by various kinds of effort, that individual should be able to call on available reserves of energy when they are needed (figure 4-8).

a. A survivor must avoid complete exhaustion which may lead to physical and psychological changes. A survivor should be able to distinguish between exhaustion and being uncomfortably tired. Although a person

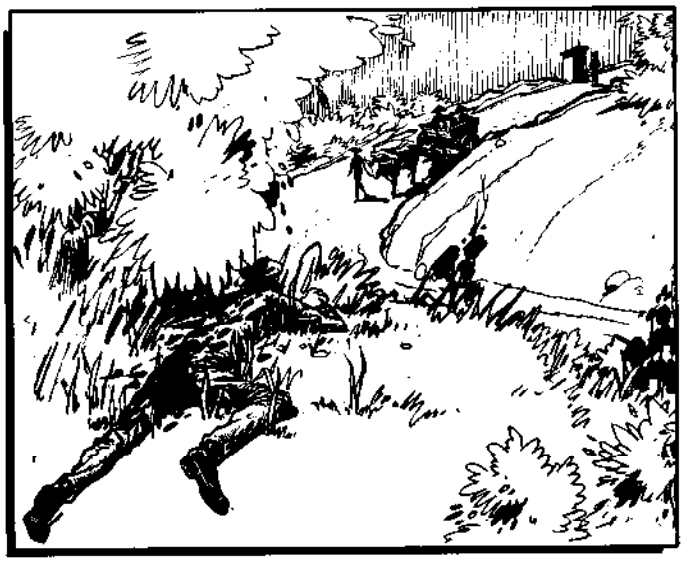

Figure 4-7. Frustration.

should avoid working to complete exhaustion, in emergencies certain tasks must be done in spite of fatigue.

(1) Rest is a basic factor for recovery from fatigue and is also important in resisting further fatigue. It is

Figure 4-8. Fatigue.

essential that the rest (following fatiguing effort) be sufficient to permit complete recovery; otherwise, the residual fatigue will accumulate and require longer periods of rest to recover from subsequent effort. During the early stages of fatigue proper rest provides a rapid recovery. This is true of muscular fatigue as well as mental fatigue. Sleep is the most complete form of rest available and is basic to recovery from fatigue.

(2) Short rest breaks during extended stress periods can improve total output. There are five ways in which rest breaks are beneficial:

(a) They provide opportunities for partial recovery from fatigue.

(b) They help reduce energy expenditure.

(c) They increase efficiency by enabling a person to take maximum advantage of planned rest.

(d) They relieve boredom by breaking up the uniformity and monotony of the task.

(e) They increase morale and motivation.

(3) Survivors should rest before output shows a definite decline. If rest breaks are longer, fewer may be required. When efforts are highly strenuous or monotonous, rest breaks should be more frequent. Rest breaks providing relaxation are the most effective. In mental work, mild exercise may be more relaxing. When work is monotonous, changes of activity, conversation, and humor are effective relaxants. In deciding on the amount and frequency of rest periods, the loss of efficiency resulting from longer hours of effort must be weighed against the absolute requirements of the survival situation.

(4) Fatigue can be reduced by working "smarter." A survivor can do this in two practical ways:

(a) Adjust the pace of the effort. Balance the load, the rate, and the time period. For example, walking at a normal rate is a more economical effort than fast walking.

(b) Adjust the technique of work. The way in which work is done has a great bearing on reducing fatigue. Economy of effort is most important. Rhythmic movements suited to the task are best.

(5) Mutual group support, cooperation, and competent leadership are important factors in maintaining group morale and efficiency, thereby reducing stress and fatigue. A survivor usually feels tired and weary before the physiological limit is reached. In addition, other stresses experienced at the same time; such as cold, hunger, fear, or despair, can intensify fatigue. The feeling of fatigue involves not only the physical reaction to effort, but also subtle changes in attitudes and motivation. Remember, a person has reserves of energy to cope with an important emergency even when feeling very tired.

b. As in the case of other stresses, even a moderate amount of fatigue reduces efficiency. To control fatigue, it is wise to observe a program of periodic rest. Because the main objective—to establish contact with friendly forces—survivors may overestimate their strength and risk exhaustion. On the other hand, neither an isolated individual nor a group leader should underestimate the capacity of the individual or the group on the basis of fatigue. The only sound basis for judgment must be gained from training and past experience. In training, a person should form an opinion of individual capacity based on actual experience. Likewise, a group leader must form an opinion of the capacities of fellow aircrew members. This group didn't think:

"By nightfall, we were completely bushed ... We decided to wrap ourselves in the 'chute instead of making a shelter. We were too tired even to build a fire. We just cut some pine boughs, rolled ourselves in the nylon and went to sleep...and so, of course, it rained, and not lightly. We stood it until we were soaked, and then we struggled out and made a shelter. Since it was pitch dark, we didn't get the sags out of the canopy, so the water didn't all run off. Just a hell of a lot of it came through. Our hip and leg joints ached as though we had acute rheumatism. Being wet and cold accentuated the pain. We changed positions every 10 minutes, after gritting our teeth to stay put that long."

Figure 4-9. Sleep Deprivation.

4-9. Sleep Deprivation. The effects of sleep loss are closely related to those of fatigue. Sleeping at unaccustomed times, sleeping under strange circumstances (in a strange place, in noise, in light, or in other distractions), or missing part or all of the accustomed amount of sleep will cause a person to react with feelings of weariness, irritability, emotion tension, and some loss of efficiency. The extent of an individual's reaction depends on the amount of disturbance and on other stress factors which may be present at the same time (figure 4-9).

a. Strong motivation is one of the principal factors in helping to compensate for the impairing effects of sleep loss. Superior physic and mental conditioning, opportunities to rest, food and water, and companions help in enduring sleep deprivation. If a person is in reasonably good physical and mental condition, sleep deprivation can be endured 5 days or more without damage, although efficiency during the latter stages may be poor. A person must learn to get as much sleep and rest as possible. Restorative effects of sleep are felt even after "catnaps." In some instances, survivors may need to stay awake. Activity, movement, conversation, eating, and drinking are some of the ways a person can stimulate the body to stay awake.

b. When one is deprived of sleep, sleepiness usually comes in waves. A person may suddenly be sleepy immediately after a period of feeling wide awake. If this can be controlled, the feeling will soon pass and the person will be wide awake again until the next wave appears. As the duration of sleep deprivation increases, these periods between waves of sleepiness become shorter. The need to sleep may be so strong in some people after a long period of deprivation that they become desperate and do careless or dangerous things in order to escape this stress.

4-10. Isolation. Loneliness, helplessness, and despair which are experienced by survivors when they are isolated are among the most severe survival stresses. People often take their associations with family, friends, military collegues, and others for granted. But survivors soon begin to miss the daily interaction with other people. However, these, like the other stresses already discussed, can be conquered. Isolation can be controlled and overcome by knowledge, understanding, deliberate countermeasures, and a determined will to resist it (figure 4-10).

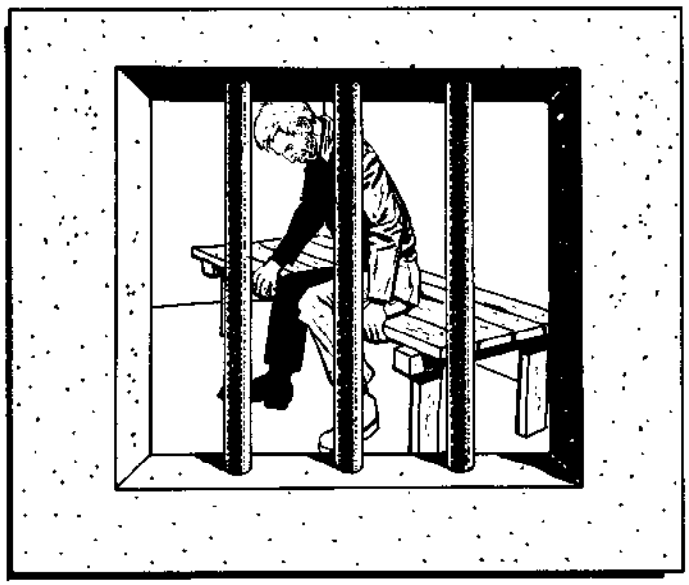

Figure 4-10. Isolation.

4-11. Insecurity. Insecurity is the survivor's feeling of helplessness or inadequacy resulting from varied stresses and anxieties. These anxieties may be caused by uncertainty regarding individual goals, abilities, and the future in a survival situation. Feelings of insecurity may

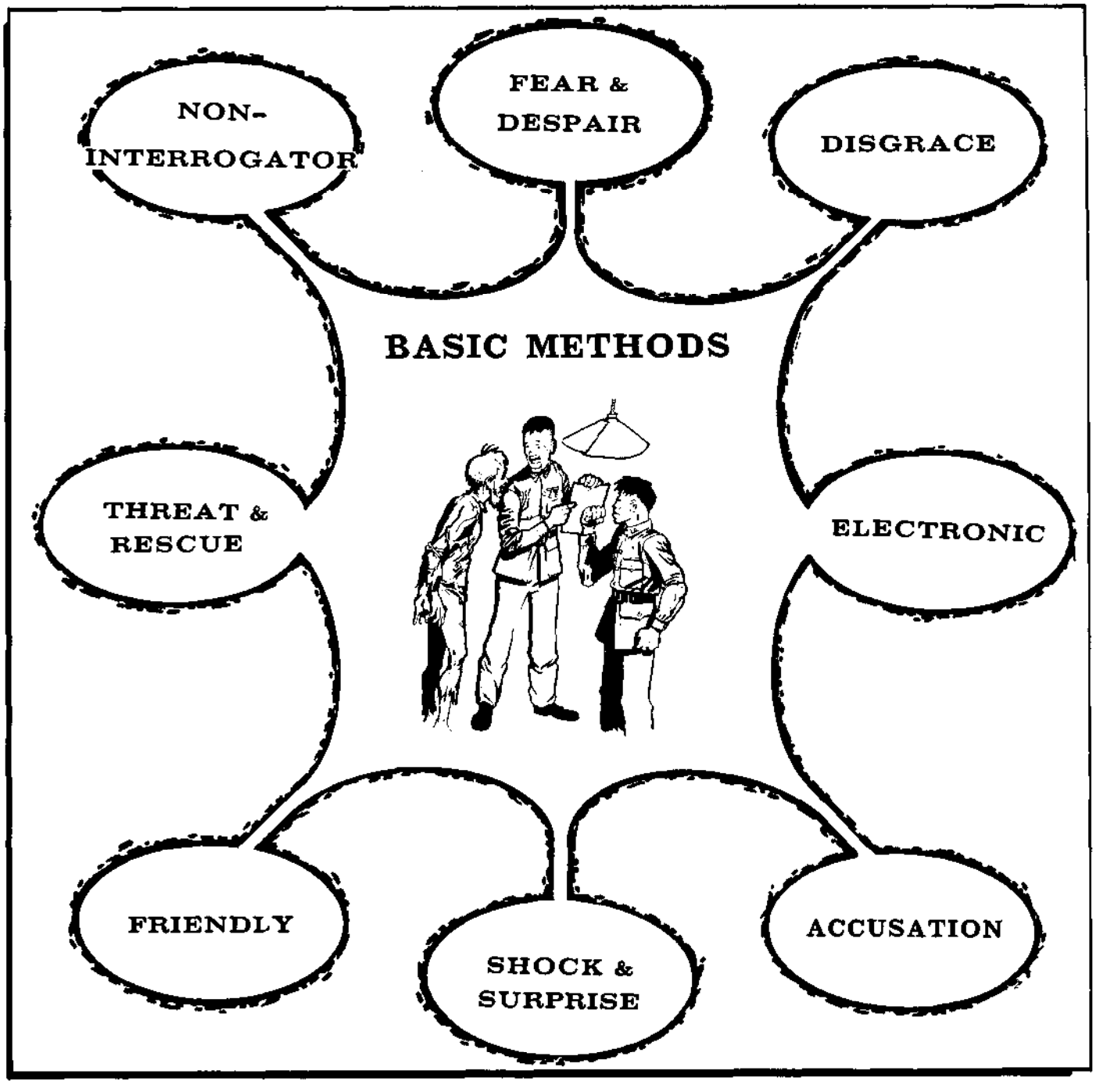

Figure 4-11. Loss of Self-Esteem.

have widely different effects on the survivor's behavior. A survivor should establish challenging but attainable goals. The better a survivor feels about individual abilities to achieve goals and adequately meet personal needs, the less insecure the survivor will feel.

4-12. Loss of Self-Esteem. Self-esteem is the state or quality of having personal self-respect and pride. Lack of (or loss of) self-esteem in a survivor may bring on depression and a change in perspective and goals. A loss of self-esteem may occur in individuals in captivity. Humiliation and other factors brought on by the captor may cause them to doubt their own worth. Humiliation comes from the feeling of losing pride or self-respect by being disgraced or dishonored, and is associated with the loss of self-esteem. Prisoners must maintain their pride and not become ashamed either because they are PWs or because of the things that happen to them as a

result of being a PW. The survivor who "loses face" (both personally and with the enemy) becomes more vulnerable to captor exploitation attempts. To solve this problem, survivors should try to maintain proper perspective about both the situation and themselves. Their feelings of self-worth may be bolstered if they recall the implied commitment in the Code of Conduct—PWs will not be forgotten (figure 4-11).

4-13. Loss of Self-Determination. A self-determined person is relatively free from external controls or influences over his or her actions. In everyday society, these "controls and influences" are the laws and customs of our society and of the self-imposed elements of our personalities. In a survival situation, the "controls and influences" can be very different. Survivors may feel as if events, circumstances, and (in some cases) other people, are in control of the situation. Some factors which may cause individuals to feel they have lost the power of self-determination are a harsh captor, captivity, bad weather, or rescue forces that make time or movement demands. This lack of self-determination is more perceived than actual. Survivors must decide how unpleasant factors will be allowed to affect their mental state. They must have the self-confidence, fostered by experience and training, to live with their feelings and decisions, and to accept responsibility for both the way they feel and how they let those feelings affect them.

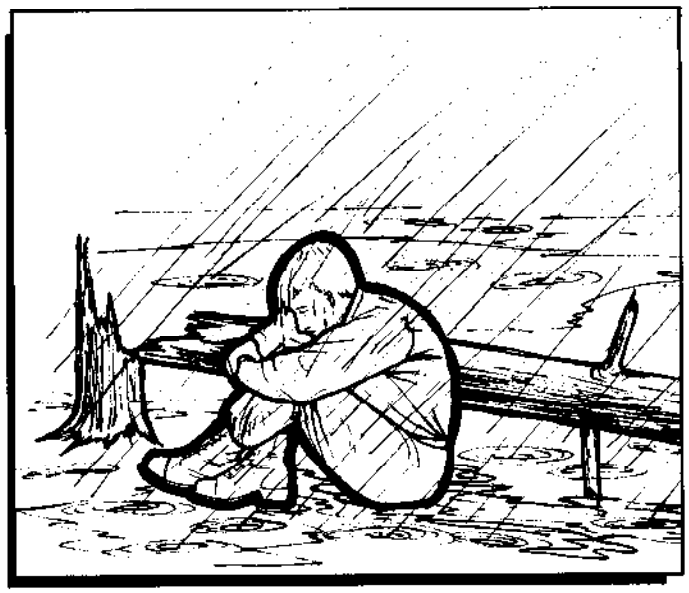

Figure 4-12. Depression.

4-14. Depression. As a survivor, depression is the biggest psychological problem that has to be conquered. It should be acknowledged that everyone has mental "highs" as well as mental "lows." People experiencing long periods of sadness or other negative feelings are suffering from depression. A normal mood associated with the sadness, grief, disappointment, or loneliness that everyone experiences at times is also described as depression. Most of the emotional changes in mood are temporary and do not become chronic. Depressed survivors may feel fearful, guilty, or helpless. They may lose interest in the basic needs of life. Many cases of depression also involve pain, fatigue, loss of appetite, or other physical ailments. Some depressed survivors try to injure or kill themselves (figure 4-12).

a. Psychiatrists have several theories as to the cause of depression. Some feel a person who, in everyday life and under normal conditions, experiences many periods of depression would probably have a difficult time in a survival situation. The main reason depression is a most difficult problem is that it can affect a wide range of psychological responses. The factors can become mutually reinforcing. For example, fatigue may lead to a feeling of depression. Depression may increase the feeling of fatigue, and this, in turn, leads to deeper depression and so on.

b. Depression usually begins after a survivor has met the basic needs for sustaining life, such as water, shelter, and food. Once the survivor's basic needs are met, there is often too much time for that person to dwell on the past, the present predicament, and on future problems. The survivor must be aware of the necessity to keep the mind and body active to eliminate the feeling of depression. One way to keep busy (daily) is by checking and improving shelters, signals, and food supply.

Chapter 5

EMOTIONAL REACTIONS

5-1. Introduction. Survivors may depend more upon their emotional reactions to a situation than upon calm, careful analysis of potential danger—the enemy, the weather, the terrain, the nature of the in-flight emergency, etc. Whether they will panic from fear, or use it as a stimulant for greater sharpness, is more dependent on the survivor's reactions to the situation than on the situation itself. Although there are many reactions to stress, the following are the most common and will be discussed in detail: fear, anxiety, panic, hate, resentment, anger, impatience, dependency, loneliness, boredom, and hopelessness.

5-2. Fear. Fear can SAVE A LIFE—or it can COST ONE. Some people are at their best when they are scared. Many downed fliers faced with survival emergencies have been surprised at how well they remembered their training, how quickly they could think and react, and what strength they had. The experience gave them a new confidence in themselves. On the other hand, some people become paralyzed when faced with the simplest survival situation. Some of them have been able "to snap themselves out of it" before it was too late. In other cases, a fellow aircrew member was on hand to assist them. However, others have not been so fortunate. They are not listed among the survivors (figure 5-1).

Figure 5-1. Fear.

a. How a person will react to fear depends more upon the individual than it does upon the situation. This has been demonstrated both in actual survival situations and in laboratory experiments. It isn't always the physically strong or the happy-go-lucky people who handle fear most effectively. Timid and anxious people have met emergencies with remarkable coolness and strength.

b. Anyone who faces life-threatening emergencies experiences fear. Fear is conscious when it results from a recognized situation (such as an immediate prospect of bailout) or when experienced as apprehension of impending disaster. Fear also occurs at a subconscious level and creates feelings of uneasiness, general discomfort, worry, or depression. Fear may vary widely in intensity, duration, and frequency of occurrence, and affect behavior across the spectrum from mild uneasiness to complete disorganization and panic. People have many fears; some are learned through personal experiences, and others are deliberately taught to them. Fear in children is directed through negative learning, as they are taught to be afraid of the dark, of animals, of noise, or of teachers. These fears may control behavior, and a survivor may react to feelings and imagination rather than to the problem causing fear.

c. When fantasy distorts a moderate danger into a major catastrophe, or vice versa, behavior can become abnormal. There is a general tendency to underestimate and this leads to reckless, foolhardy behavior. The principal means of fighting fear (in this case) is to pretend that it does not exist. There are no sharp lines between recklessness and bravery. It is necessary to check behavior constantly to maintain proper control.

d. One or more of the following signs or symptoms may occur in those who are afraid. However, they may also appear in circumstances other than fear.

(1) Quickening of pulse; trembling.

(2) Dilation of pupils.

(3) Increased muscular tension and fatigue.

(4) Perspiration of palms of hands, soles of feet, and armpits.

(5) Dryness of mouth and throat; higher pitch of voice; stammering.

(6) Feeling of "butterflies in the stomach," emptiness of the stomach, faintness, and nausea.

e. Accompanying these physical symptoms are the following common psychological symptoms:

(1) Irritability; increased hostility.

(2) Talkativeness in early stages, leading finally to speechlessness.

(3) Confusion, forgetfulness, and inability to concentrate.

(4) Feelings of unreality, flight, panic, or stupor.

f. Throughout military history, many people have coped successfully with the most strenuous odds. In adapting to fear, they have found support in previous training and experience. There is no limit to human

control of fear. Survivors must take action to control fear. They cannot run away from fear. Appropriate actions should be to:

(1) Understand fear.

(2) Admit that it exists.

(3) Accept fear as reality.

g. Training can help survivors recognize what individual reactions may be. Using prior training, survivors should learn to think, plan, and act logically, even when afraid.

h. To effectively cope with fear, a survivor must:

(1) Develop confidence. Use training opportunities; increase capabilities by keeping physically and mentally fit; know what equipment is available and how to use it; learn as much as possible about all aspects of survival.

(2) Be prepared. Accept the possibility that "it can happen to me." Be properly equipped and clothed at all times; have a plan ready. Hope for the best, but be prepared to cope with the worst.

(3) Keep informed. Listen carefully and pay attention to all briefings. Know when danger threatens and be prepared if it comes; increase knowledge of survival environments to reduce the "unknown."

(4) Keep busy at all times. Prevent hunger, thirst, fatigue, idleness, and ignorance about the situation, since these increase fear.

(5) Know how fellow crewmembers react to fear. Learn to work together in emergencies—to live, work, plan, and help each other as a team.

(6) Practice religion. Don't be ashamed of having spiritual faith.

(7) Cultivate "good" survival attitudes. Keep the mind on a main goal and keep everything else in perspective. Learn to tolerate discomfort. Don't exert energy to satisfy minor desires which may conflict with the overall goal—to survive.

(8) Cultivate mutual support. The greatest support under severe stress may come from a tightly knit group. Teamwork reduces fear while making the efforts of every person more effective.

(9) Exercise leadership. The most important test of leadership and perhaps its greatest value lies in the stress situation.

(10) Practice discipline. Attitudes and habits of discipline developed in training carry over into other situations. A disciplined group has a better chance of survival than an undisciplined group.

(11) Lead by example. Calm behavior and demonstration of control are contagious. Both reduce fear and inspire courage.

i. Every person has goals and desires. The greatest values exercise the greatest influence. Because of strong religious, moral, or patriotic values, people have been known to face torture and death calmly rather than reveal information or compromise a principle. Fear can kill or it can save lives. It is a normal reaction to danger. By understanding and controlling fear through training, knowledge, and effective group action, fear can be overcome.

5-3. Anxiety:

a. Anxiety is a universal human reaction. Its presence can be felt when changes occur which affect an individual's safety, plans, or methods of living. It is generally felt when individuals perceive something bad is about to happen. A common description of anxiety is "butterflies in the stomach." Anxiety creates feelings of uneasiness, general discomfort, worry, or depression. Anxiety and fear differ mainly in intensity. Anxiety is a milder reaction and the specific cause(s) may not be readily apparent, whereas fear is a strong reaction to a specific, known cause. Common characteristics of anxiety are: fear of the future, indecision, feelings of helplessness, resentment (figure 5-2).

Figure 5-2. Anxiety.

b. To overcome anxiety, the individual must take positive action by adopting a simple plan. It is essential to keep your mind off of your injuries and do something constructive. For example, one PW began to try and teach English to the Chinese and to learn Chinese from them.

5-4. Panic. In the face of danger, a person may panic or "freeze" and cease to function in an organized manner. A person experiencing panic may have no conscious control over individual actions. Uncontrollable, irrational behavior is common in emergency situations. Anybody can panic, but some people go to pieces more easily than others. Panic is brought on by a sudden overwhelming fear, and can often spread quickly through a group of people. Every effort must be made

to bolster morale and calm the panic with leadership and discipline. Panic has the same signs as fear and should be controlled in the same manner as fear. This survivor allowed pain to panic him.

> "His parachute caught in the tree, and he found himself suspended about five feet above the ground...one leg strap was released while he balanced in this aerial position and he immediately slipped toward the ground.

Figure 5-3. Panic.

> In doing so, his left leg caught in the webbing and he was suspended by one leg with his head down. Unfortunately, the pilot's head touched an ant hill and biting ants immediately swarmed over him. Apparently, in desperation, the flier pulled his gun and fired five rounds into the webbing holding his foot. When he did not succeed in breaking the harness by shooting at it, he placed the last shot in his head and thus took his own life. It was obvious from the discoverer's report that if the pilot had even tried to turn around or to swing himself from his inverted position, he could have reached either the aerial roots or the latticed trunk of the tree. With these branches, he should have been able to pull himself from the harness...The fact that his head was in a nest of stinging ants only added to his panic, which led to the action that took his life." (Figure 5-3)

5-5. Hate. Hate—feelings of intense dislike, extreme aversion, or hostility—is a powerful emotion which can have both positive and negative effects on a survivor. An understanding of the emotion and its causes is the key to learning to control it. Hate is an acquired emotion rooted in a person's knowledge or perceptions. The accuracy or inaccuracy of the information is irrelevant to learning to hate.

a. Any person, any object, or anything that may be understood intellectually, such as political concepts or religious dogma, can promote feelings of hate. Feelings of hate (usually accompanied with a desire for vengence, revenge, or retribution) have sustained former prisoners of war through their harsh ordeals. If an individual loses perspective while under the influence of hate and reacts emotionally, rational solutions to problems may be overlooked, and the survivor may be endangered.

b. To effectively deal with this emotional reaction, the survivor must first examine the reasons why the feeling of hate is present. Once that has been determined, survivors should then decide what to do about those feelings. Whatever approach is selected, it should be as constructive as possible. Survivors must not allow hate to control them.

5-6. Resentment. Resentment is the experiencing of an emotional state of displeasure or indignation toward some act, remark, or person that has been regarded as causing personal insult or injury. Luck and fate may play a role in any survival situation. A hapless survivor may feel jealous resentment toward a fellow PW, travel partner, etc., if that other person is perceived to be enjoying a success or advantage not presently experienced by the observer. The survivor must understand that events cannot always go as expected. It is detrimental to morale and could affect survival chances if feelings of resentment over another's attainments become too strong. Imagined slights or insults are common. The survivor should try to maintain a sense of humor and perspective about ongoing events and realize that stress and lack of self-confidence play roles in bringing on feelings of resentment.

5-7. Anger. Anger is a strong feeling of displeasure and belligerence aroused by a real or supposed wrong. People become angry when they cannot fulfill a basic need

or desire which seems important to them. When anger is not relieved, it may turn into a more enduring attitude of hostility, characterized by a desire to hurt or destroy the person or thing causing the frustration. When anger is intense, the survivor loses control over the situation, resulting in impulsive behavior which may be destructive in nature. Anger is a normal response which can serve a useful purpose when carefully controlled. If the situation warrants and there is no threat to survival, one could yell or scream, take a walk, do some vigorous exercise, or just get away from the source of the anger, even if only for a few minutes. Here is a man who couldn't hold it.

> "I tried patiently to operate it (radio) in every way I had been shown. Growing more angry and disappointed at its failure, I tore the aerial off, threw the cord away, beat the battery on the rocks, then threw the pieces all over the hillside. I was sure disappointed." (See figure 5-4.)

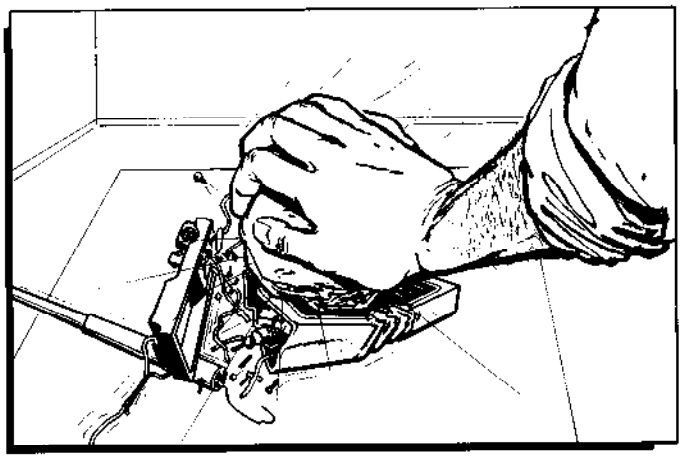

Figure 5-4. Anger.

5-8. Impatience:

a. The psychological stresses brought about by feelings of impatience can quickly manifest themselves in physical ways. Internally, the effects of impatience can cause changes in physical and mental well-being. Survivors who allow impatience to control their behavior may find that their efforts prove to be counterproductive and possibly dangerous. For example, evaders who don't have the ability or willingness to suppress annoyance when confronted with delay may expose themselves to capture or injury.

b. Potential survivors must understand they have to bear pain, misfortune, and annoyance without complaint. In the past, many survivors have displayed tremendous endurance, both mental and physical, in times of distress or misfortune. While not every survivor will be able to display such strength of character in all situations, each person should learn to recognize the things which may make them impatient to avoid acting unwisely. This survivor couldn't wait:

> "I became very impatient. I had planned to wait until night to travel but I just couldn't wait. I left the ditch about noon and walked for about two hours until I was caught."

5-9. Dependence. The captivity environment is the prime area where a survivor may experience feelings of dependency. The captor will try to develop in prisoners feelings of need, support, and trust for the captor. By regulating the availability of basic needs like food, water, clothing, social contact, and medical care, captors show their power and control over the prisoners' fate. Through emphasis on the prisoners' inability to meet their own basic needs, captors seek to establish strong feelings of prisoner dependency. This dependency can make prisoners extremely vulnerable to captor exploitation—a major captor objective. PW recognition of this captor tactic is key to countering it. Survivors must understand that, despite captor controls, they do control their own lives. Meeting even one physical or mental need can provide a PW with a "victory" and provide the foundation for continued resistance against exploitation (figure 5-5).

Figure 5-5. Dependence.

5-10. Loneliness. Loneliness can be very debilitating during a survival episode. Some people learn to control and manipulate their environment and become more self-sufficient while adapting to changes. Others rely on protective persons, routines, and familiarity of surroundings to function and obtain satisfaction (figure 5-6).

a. The ability to combat feelings of loneliness during a survival episode must be developed long before the episode occurs. Self-confidence and self-sufficiency are key factors in coping with loneliness. People develop these attributes by developing and demonstrating competence in performing tasks. As the degree of competence increases, so does self-confidence and self-sufficiency. Military training, more specifically survival training, is designed to provide individuals with the competence and self-sufficiency to cope with and adapt to survival living.

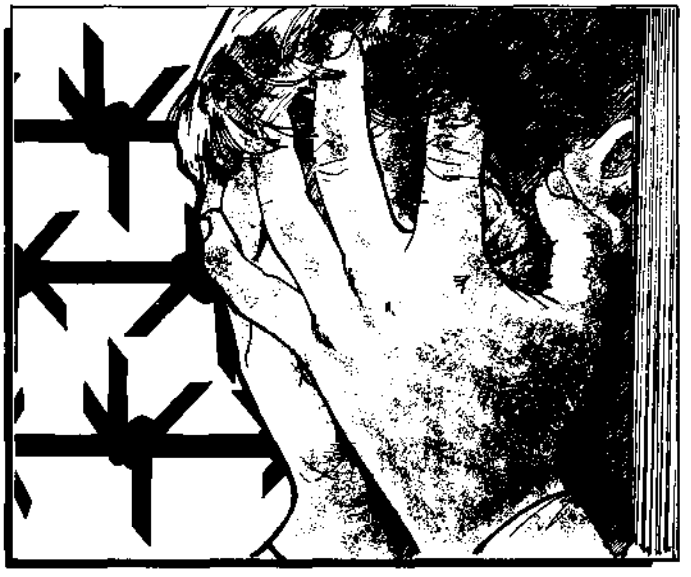

Figure 5-6. Loneliness.

b. In a survival situation, the countermeasure to conquer loneliness is to be active, to plan and think purposely. Development of self-sufficiency is the primary protection since all countermeasures in survival require the survivor to have the ability to practice self-control.

5-11. Boredom. Boredom and fatigue are related and frequently confused. Boredom is accompanied by a lack of interest and may include feelings of strain, anxiety, or depression, particularly when no relief is in sight and the person is frustrated. Relief from boredom must be based on correction of the two basic sources, repetitiveness and uniformity. Boredom can be relieved by a variation of methods—rotation of duties, broadening the scope of a particular task or job, taking rest breaks, or other techniques of diversification which may actually interfere with efficient performance of the job. The ungratifying nature of a task can be counteracted by clearing up its meaning, objectives, and, in some cases, its relation to the total plan.

a. This survivor couldn't think of anything to do: "The underground representative took me to a house to wait for another member of the underground to pick me up. This was the worst part of the whole experience—this waiting. I just sat in the house and waited for two weeks. I thought I would go mad." (See figure 5-7.)

b. This survivor invented something to do: "Not knowing what to do, I decided to kill all the bugs. There were a lot of spiders, the big ones that do not hurt humans, so I killed the flies and gave them to the spiders to eat."

Figure 5-7. Boredom.

5-12. Hopelessness. Hopelessness stems from negative feelings—regardless of actions taken, success is impossible, or the certainty that future events will turn out for the worst no matter what a person tries to do. Feelings of hopelessness can occur at virtually any time during a survival episode. Survivors have experienced loss of hope in trying to maintain health due to an inability to care for sickness, broken bones, or injuries; considering their chances of returning home alive; seeing their loved ones again; or believing in their physical or mental ability to deal with the situation; for example, evade long distances or not give information to an interrogator (figure 5-8).

Figure 5-8. Hopelessness.

a. During situations where physical exhaustion or exposure to the elements affects the mind, a person may begin to lose hope. The term "give-up-itis" was coined in Korea to describe the feeling of "hopelessness." During captivity, deaths occurred for no apparent cause. These individuals actually willed themselves to die or at least did not will themselves to live. The original premise (in the minds of such people) is that they are going to die. To them, the situation seemed totally futile and they had passively abandoned themselves to fate. It was possible to follow the process step by step. The people who died withdrew themselves from the group, became despondent, then lay down and gave up. In some cases, death followed rapidly.

b. One way to treat hopelessness is to eliminate the cause of the stress. Rest, comfort, and morale building activities can help eliminate this psychological problem. Another method used in Korea was to make the person so angry that the person wanted to get up and attack the tormentors. A positive attitude has a powerful influence on morale and combating the feeling of hopelessness.

c. Since many stress situations cannot be dealt with successfully by either withdrawal or direct attack, it may be necessary to work out a compromise solution. The action may entail changing a survivor's method of operation or accepting substitute goals.

d. Evaders faced with starvation may compromise with their conscience and steal "just this one time." They may ignore their food aversion and eat worms, bugs, or even human flesh. A related form of compromise is acceptance of substitute means to achieve the same goals.

5-13. Summary. All the psychological factors may be overcome by survivors if they can recognize the problem, work out alternative solutions, decide on an appropriate course of action, take action, and evaluate the results. Perhaps the most difficult step in this sequence is deciding on an appropriate course of action. Survivors may face either one or several psychological problems. These problems are quite dangerous and must be effectively controlled or countered for survival to continue.

Chapter 6

THE WILL TO SURVIVE

6-1. Introduction. The will to survive is defined as the desire to live despite seemingly insurmountable mental and(or) physical obstacles. The tools for survival are furnished by the military, the individual, and the environment. The training for survival comes from survival training publications, instruction, and the individual's own efforts. But tools and training are not enough without a *will to survive.* In fact, the records prove that "will" alone has been the deciding factor in many survival cases. While these accounts are not classic examples of "how to survive," they illustrate that a single-minded survivor with a powerful *will to survive* can overcome most hardships. There are cases where people have eaten their belts for nourishment, boiled water in their boots to drink as broth, or have eaten human flesh—though this certainly wasn't their cultural instinct.

a. One incident where the *will to survive* was the deciding factor between life and death involved a man stranded in the Arizona desert for 8 days without food and water. He traveled more than 150 miles during searing daytime temperatures, losing 25 percent of his body weight due to the lack of water (usually 10 percent loss causes death). His blood became so thick that the lacerations he received could not bleed until he had been rescued and received large quantities of water. When he started on that journey, something must have clicked in his mind telling him to live, regardless of any obstacles which might confront him. And live he did—on guts and will alone! (See figure 6-1.)

b. Let's flip a coin and check the other side of "will." Our location is the Canadian wilderness. A pilot ran into engine trouble and chose to deadstick his plane onto a frozen lake rather than punch out. He did a beautiful job and slid to a stop in the middle of the lake. He left the aircraft and examined it for damage. After surveying the area, he noticed a wooded shoreline only 200 yards away where food and shelter could be provided—he decided to go there. Approximately halfway there, he changed his mind and returned to the cockpit of his aircraft where he smoked a cigar, took out his pistol, and blew his brains out. Less than 24 hours later, a rescue team found him. Why did he give up? Why was he unable to survive? Why did he take his own life? On the other hand, why do people eat their belts or drink broth from their boots? No one really knows, but it's all related to the *will to survive.*

6-2. Overcoming Stress. The ability of the mind to overcome stress and hardship becomes most apparent when there appears to be little chance of a person surviving. When there appears to be no escape from the situation, the "will" enables a person to begin to win "the battle of the mind." This mental attitude can bridge the gap between the crisis period and the coping period.

Figure 6-1. Will to Survive.

6-3. Crisis Period:

a. The crisis period is the point at which the person realizes the gravity of the situation and understands that the problem will not go away. At this stage, action is needed. Most people will experience shock in this stage as a result of not being ready to face this new challenge. Most will recover control of their faculties, especially if they have been prepared through knowledge and training.

b. Shock during a crisis is normally a response to being overcome with anxiety. Thinking will be disorganized. At this stage, direction will be required because the individual is being controlled by the environment. The person's center of control is external. In a group survival episode, a natural leader may appear who will direct and reassure the others. But if the situation continues to control the individual or the group, the response may be panic, behavior may be irrational, and judgment is impaired. In a lone-survivor episode, the individual must gain control of the situation and respond constructively. In either case, survivors must evaluate the situation and develop a plan of action.

During the evaluation, the survivor must determine the most critical needs to improve the chance of living and being rescued.

6-4. The Coping Period. The coping period begins after the survivor recognizes the gravity of the situation and resolves to endure it rather than succumb. The survivor must tolerate the effects of physical and emotional stresses. These stresses can cause anxiety which becomes the greatest obstacle to self-control and solving problems. Coping with the situation requires considerable internal control. For example, the survivor must often subdue urgent desires to travel when that would be counterproductive and dangerous. A person must have patience to sit in an emergency action shelter while confronted with an empty stomach, aching muscles, numb toes, and suppressed feelings of depression and hopelessness. Those who fail to think constructively may panic. This could begin a series of mistakes which result in further exhaustion, injury, and sometimes death. Death comes not from hunger pains but from the inability to manage or control emotions and thought processes.

6-5. Attitude. The survivor's attitude is the most important element of the *will to survive.* With the proper attitude, almost anything is possible. The desire to live is sometimes based on the feelings toward another person and(or) thing. Love and hatred are two emotional extremes which have moved people to do exceptional things physically and mentally. The lack of a *will to survive* can sometimes be identified by the individual's motivation to meet essential survival needs, emotional control resulting in reckless, paniclike behavior, and self-esteem.

a. It is essential to strengthen the *will to survive* during an emergency. The first step is to avoid a tendency to panic or "fly off the handle." Sit down, relax, and analyze the situation rationally. Once thoughts are collected and thinking is clear, the next step is to make decisions. In normal living, people can avoid decisions and let others do their planning. But in a survival situation, this will seldom work. Failure to decide on a course of action is actually a decision for inaction. This lack of decisionmaking may even result in death. However, decisiveness must be tempered with flexibility and planning for unforeseen circumstances. As an example, an aircrew member down in an arctic nontactical situation decides to construct a shelter for protection from the elements. The planning and actions must allow sufficient flexibility so the aircrew can monitor the area for indications of rescuers and be prepared to make contact—visually, electronically, etc.—with potential rescuers.

b. Tolerance is the next topic of concern. A survivor or evader will have to deal with many physical and psychological discomforts, such as unfamiliar animals, insects, loneliness, and depression. Aircrew members are trained to tolerate uncomfortable situations. That training must be applied to deal with the stress of environments.

c. Survivors in both tactical and nontactical situations must face and overcome fears to strengthen the *will to survive.* These fears may be founded or unfounded, be generated by the survivor's uncertainty or lack of confidence, or be based on the proximity of enemy forces. Indeed, fear may be caused by a wide variety of real and imagined dangers. Despite the source of the fear, survivors must recognize fear and make a conscious effort to overcome it.

6-6. Optimism. One of a survivor's key assets is optimism—hope and faith. Survivors must maintain a positive, optimistic outlook on their circumstance and how well they are doing. Prayer or meditation can be helpful. How a survivor maintains optimism is not so important as its use.

6-7. Summary. Survivors do not choose or welcome their fate and would escape it if they could. They are trapped in a world of seemingly total domination—a world hostile to life and any sign of dignity or resistance. The survival mission is not an easy one, but it is one in which success can be achieved. This has been an introduction to the concepts and ideas that can help an aircrew member return. Having the *will to survive* is what it's all about!

Part Three

BASIC SURVIVAL MEDICINE

Chapter 7

SURVIVAL MEDICINE

7-1. Introduction:

a. Foremost, among the many things which can compromise a survivor's ability to return are medical problems encountered during ejection, parachute descent, and(or) parachute landing. In the Southeast Asian conflict, some 30 percent of approximately 1,000 US Air Force survivors, including 322 returned PWs, were injured by the time they disentangled themselves from their parachutes. The most frequently reported injuries were fractures, strains, sprains, and dislocations, as well as burns and other types of wounds (figure 7-1).

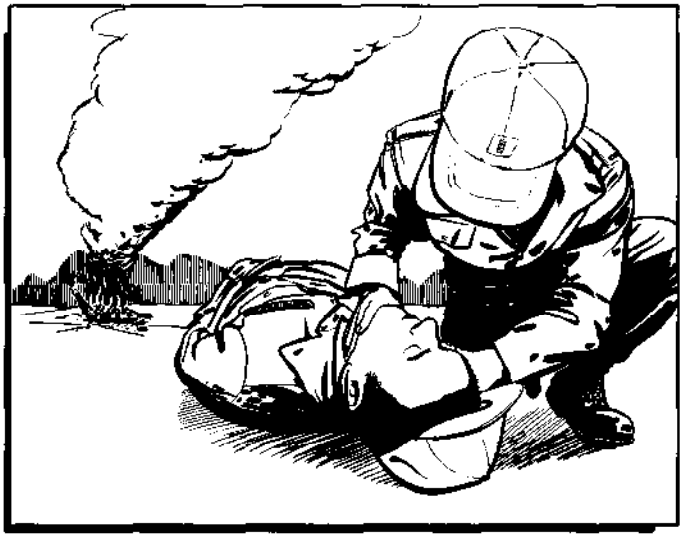

Figure 7-1. Survival Medicine.

b. Injuries and illnesses peculiar to certain environments can reduce survival expectancy. In cold climates, and often in an open sea survival situation, exposure to extreme cold can produce serious tissue trauma, such as frostbite, or death from hypothermia. Exposure to heat in warm climates, and in certain areas on the open seas, can produce heat cramps, heat exhaustion, or life-threatening heatstroke.

c. Illnesses contracted during evasion or in a captivity environment can interfere with successful survival. Among these are gastrointestinal disorders, respiratory diseases, skin infections and infestations, malaria, typhus, cholera, etc.

d. A review of the survival experiences from World War II, Korea, and Southeast Asia indicates that, while US military personnel generally knew how to administer first aid to others, there was a marked inability to administer self-aid. Further, only the most basic medical care had been taught to most military people. Lastly, it was repeatedly emphasized that even minor injuries or ailments, when ignored, became major problems in a survival situation. Thus, prompt attention to the most minor medical problem is essential in a survival episode. Applying principles of survival medicine should enable military members to maintain health and well-being in a hostile or nonhostile environment until rescued and returned to friendly control.

e. Information in this chapter and chapter 8 is a basic reference to self-aid techniques used by PWs in captivity and techniques found in folk medicine. The information describes procedures which can maintain health in medically austere situations. It includes items used to prevent and treat injuries and illnesses. Because there is no "typical" survival situation, the approach to self-aid must be flexible, placing emphasis on using what is available to treat the injury or illness. Further, survivors recognize that medical treatment offered by people of other cultures may be far different from our own. For example, in the rural areas of Vietnam, a poultice of python meat was and is used to treat internal lower back pain. Such treatment may be repugnant to some US military personnel; however, medical aid offered to survivors in non-US cultures may be the best available in the given circumstance.

f. The procedures in this chapter and chapter 8 must be viewed in the reality of a true survival situation. The results of treatment may be substandard compared with present medical standards. However, these procedures will not compromise professional medical care which becomes available following rescue. Moreover, in the context of a survival situation, they may represent the best available treatment to extend the individual's survival expectancy.

7-2. Procedures and Expedients. Survival medicine encompasses procedures and expedients that are:

a. Required and available for the preservation of health and the prevention, improvement, or treatment of injuries and illnesses encountered during survival.

b. Suitable for application by nonmedical personnel to themselves or comrades in the circumstances of the survival situation.

(1) Survival medicine is more than first aid in the conventional sense. It approaches final definitive treatment in that it is not dependent upon the availability of technical medical assistance within a reasonable period of time.

(2) To avoid duplication of information generally available, the basic principles of first aid will not be repeated, nor will the psychological factors affecting survival which were covered in part two.

7-3. Hygiene. In a survival situation, cleanliness is essential to prevent infection. Adequate personal cleanliness will not only protect against disease germs that are present in the individual's surroundings but will also protect the group by reducing the spread of these germs (figure 7-2).

Figure 7-2. Hygiene.

a. Washing, particularly the face, hands, and feet, reduces the chances of infection from small scratches and abrasions. A daily bath or shower with hot water and soap is ideal. If no tub or shower is available, the body should be cleaned with a cloth and soapy water, paying particular attention to the body creases (armpits, groin, etc.), face, ears, hands, and feet. After this type of "bath," the body should be rinsed thoroughly with clear water to remove all traces of soap which could cause irritation.

b. Soap, although an aid, is not essential to keeping clean. Ashes, sand, loamy soil, and other expedients may be used to clean the body and cooking utensils.

c. When water is in short supply, the survivor should take an "air bath." All clothing should be removed and the body simply exposed to the air. Exposure to sunshine is ideal, but even on an overcast day or indoors, a 2-hour exposure of the naked body to the air will refresh the body. Care should be taken to avoid sunburn when bathing in this manner. Exposure in the shade, shelter, sleeping bag, etc., will help if the weather conditions do not permit direct exposure.

d. Hair should be kept trimmed, preferably 2 inches or less in length, and the face should be clean-shaven. Hair provides a surface for the attachment of parasites and the growth of bacteria. Keeping the hair short and the face clean-shaven will provide less habitat for these organisms. At least once a week, the hair should be washed with soap and water. When water is in short supply, the hair should be combed or brushed thoroughly and covered to keep it clean. It should be inspected weekly for fleas, lice, and other parasites. When parasites are discovered, they should be removed.

e. The principal means of infecting food and open wounds is contact with unclean hands. Hands should be washed with soap and water, if available, after handling any material which is likely to carry germs. This is especially important after each visit to the latrine, when caring for the sick and injured, and before handling food, food utensils, or drinking water. The fingers should be kept out of the mouth and the fingernails kept closely trimmed and clean. A scratch from a long fingernail could develop into a serious infection.

7-4. Care of the Mouth and Teeth. Application of the following fundamentals of oral hygiene will prevent tooth decay and gum disease:

a. The mouth and teeth should be cleansed thoroughly with a toothbrush and dentifrice at least once each day. When a toothbrush is not available, a "chewing stick" can be fashioned from a twig. The twig is washed, then chewed on one end until it is frayed and brushlike. The teeth can then be brushed very thoroughly with the stick, taking care to clean all tooth surfaces. If necessary, a clean strip of cloth can be wrapped around the finger and rubbed on the teeth to wipe away food particles which have collected on them. When neither toothpaste nor toothpowder are available, salt, soap, or baking soda can be used as substitute dentifrices. Parachute inner core can be used by separating the filaments of the inner core and using this as a dental floss. Gargling with willow bark tea will help protect the teeth.

b. Food debris which has accumulated between the teeth should be removed by using dental floss or toothpicks. The latter can be fashioned from small twigs.

c. Gum tissues should be stimulated by rubbing them vigorously with a clean finger each day.

d. Use as much care cleaning dentures and other dental appliances, removable or fixed, as when cleaning natural teeth. Dentures and removable bridges should be removed and cleaned with a denture brush or "chew stick" at least once each day. The tissue under the dentures should be brushed or rubbed regularly for proper stimulation. Removable dental applicances should be

removed at night or for a 2- to 3-hour period during the day.

7-5. Care of the Feet. Proper care of the feet is of utmost importance in a survival situation, especially if the survivor has to travel. Serious foot trouble can be prevented by observing the following simple rules:

a. The feet should be washed, dried thoroughly, and massaged each day. If water is in short supply, the feet should be "air cleaned" along with the rest of the body (figure 7-3).

b. Toenails should be trimmed straight across to prevent the development of ingrown toenails.

c. Boots should be broken in before wearing them on *any* mission. They should fit properly, neither so tight that they bind and cause pressure spots nor so loose that they permit the foot to slide forward and backward when walking. Insoles should be improvised to reduce any friction spots inside the shoes.

Figure 7-3. Care of Feet.

d. Socks should be large enough to allow the toes to move freely but not so loose that they wrinkle. Wool socks should be at least one size larger than cotton socks to allow for shrinkage. Socks with holes should be properly darned before they are worn. Wearing socks with holes or socks that are poorly repaired may cause blisters. Clots of wool on the inside and outside should be removed from wool socks because they may cause blisters. Socks should be changed and washed thoroughly with soap and water each day. Woolen socks should be washed in cool water to lessen shrinkage. In camp, freshly laundered socks should be stretched to facilitate drying by hanging in the sun or in an air current. While traveling, a damp pair of socks can be dried by placing them inside layers of clothing or hanging them on the outside of the pack. If socks become damp, they should be exchanged for dry ones at the first opportunity.

e. When traveling, the feet should be examined regularly to see if there are any red spots or blisters. If detected in the early stages of development, tender areas should be covered with adhesive tape to prevent blister formation.

7-6. Clothing and Bedding. Clothing and bedding become contaminated with any disease germs which may be present on the skin, in the stool, in the urine, or in secretions of the nose and throat. Therefore, keeping clothing and bedding as clean as possible will decrease the chances of skin infection and decrease the possibility of parasite infestation. Outer clothing should be washed with soap and water when it becomes soiled. Under clothing and socks should be changed daily. If water is in short supply, clothing should be "air cleaned." For air cleaning, the clothing is shaken out of doors, then aired and sunned for 2 hours. Clothing cleaned in this manner should be worn in rotation. Sleeping bags should be turned inside out, fluffed, and aired after each use. Bed linen should be changed at least once a week, and the blankets, pillows, and mattresses should be aired and sunned (figure 7-4).

7-7. Rest. Rest is necessary for the survivor because it not only restores physical and mental vigor, but also promotes healing during an illness or after an injury.

a. In the initial stage of the survival episode, rest is particularly important. After those tasks requiring immediate attention are done, the survivor should inventory available resources, decide upon a plan of action, and even have a meal. This "planning session" will provide a rest period without the survivor having a feeling of "doing nothing."

b. If possible, regular rest periods should be planned in each day's activities. The amount of time allotted for rest will depend on a number of factors, including the survivor's physical condition, the presence of hostile forces, etc., but usually, 10 minutes each hour is sufficient. During these rest periods, the survivor should change either from physical activity to complete rest or from mental activity to physical activity as the case may be. The survivor must learn to become comfortable and to rest under less than ideal conditions.

Figure 7-4. Bedding.

7-8. Rules for Avoiding Illness. In a survival situation, whether short-term or long-term, the dangers of disease are multiplied. Application of the following simple guidelines regarding personal hygiene will enable the survivor to safeguard personal health and the health of others:

a. ALL water obtained from natural sources should be purified before consumption.

b. The ground in the camp area should not be soiled with urine or feces. Latrines should be used, if available. When no latrines are available, individuals should dig "cat holes" and cover their waste.

c. Fingers and other contaminated objects should never be put into the mouth. Hands should be washed before handling any food or drinking water, before using the fingers in the care of the mouth and teeth, before and after caring for the sick and injured, and after handling any material likely to carry disease germs.

d. After each meal, all eating utensils should be cleaned and disinfected in boiling water.

e. The mouth and teeth should be cleansed thoroughly at least once each day. Most dental problems associated with long-term survival episodes can be prevented by using a toothbrush and toothpaste to remove accumulated food debris. If necessary, devices for cleaning the teeth should be improvised.

f. Bites and insects can be avoided by keeping the body clean, by wearing proper protective clothing, and by using head net, improvised bed nets, and insect repellants.

g. Wet clothing should be exchanged for dry clothing as soon as possible to avoid unnecessary body heat loss.

h. Personal items such as canteens, pipes, towels, toothbrushes, handkerchiefs, and shaving items should not be shared with others.

i. All food scraps, cans, and refuse should be removed from the camp area and buried.

j. If possible, a survivor should get 7 or 8 hours of sleep each night.

k. Aircrew members should keep all immunization "shots" current.

7-9. General Management of Injuries:

a. Bleeding. Control of bleeding is most important in survival situations where replacement transfusions are not possible. Immediate steps should be taken to stop the flow of blood, regardless of its source. The method used should be commensurate with the type and degree of bleeding. The tourniquet, when required and properly used, will save life. If improperly used, it may cost the life of the survivor. The basic characteristics of a tourniquet and the methods of its use are well covered in standard first aid texts; however, certain points merit emphasis in the survival situation. A tourniquet should be used only after every alternate method has been attempted. If unable to get to medical aid within 2 hours, after 20 minutes, gradually loosen the tourniquet. If bleeding has stopped, remove the tourniquet; if bleeding continues, reapply and leave in place. The tourniquet should be applied as near the site of the bleeding as possible, between the wound and the heart, to reduce the amount of tissue lost.

b. Pain:

(1) Control of Pain. The control of pain accompanying disease or injury under survival situations is both difficult and essential. In addition to its morale-breaking discomfort, pain contributes to shock and makes the survivor more vulnerable to enemy influences. Ideally, pain should be eliminated by the removal of the cause. However, this is not always immediately possible, hence measures for the control of pain are beneficial.

(2) Position, Heat, and Cold. The part of the body that is hurting should be put at rest, or at least its activity restricted as much as possible. The position selected should be the one giving the most comfort, and be the easiest to maintain. Splints and bandages may be necessary to maintain the immobilization. Elevation of the injured part, with immobilization, is particularly beneficial in the throbbing type pain such as is typical of the "mashed" finger. Open wounds should be cleansed, foreign bodies removed, and a clean dressing applied to protect the wound from the air and chance contacts with environmental objects. Generally, the application of warmth reduces pain—toothache, bursitis, etc. However, in some conditions, application of cold has the same effect—strains and sprains. Warmth or cold is best applied by using water due to its high specific heat, and the survivor can try both to determine which is most beneficial.

(3) Pain Killers. Drugs are very effective in reducing pain; however, they are not likely to be available in the survival situation. Hence, the importance of the above "natural" procedures. Aspirin, APCs, and such tablets are primarily intended to combat the discomforts of colds and upper respiratory diseases, and, at best, will just take the edge off severe pain. They should be taken, however, if available. If no aspirin is available, there are some parts of vegetation which can be used. For example, most of the willows have been used for their pain-relieving and fever-lowering properties for hundreds of years. The fresh bark contains salicin, which probably decomposes into salicylic acid in the human body. Wintergreen, also known as checkerberry, was used by some Indians for body aches and pains. The leaves are made into a tea. The boiled bark of the magnolia tree helps relieve internal pains and fever, and has been known to stop dysentery. To be really effective in control of pain, stronger narcotic drugs such as codeine and morphine are required. During active hostilities, morphine may be available in aircraft and individual first aid kits.

c. Shock:

(1) Circulatory Reaction. Shock in some degree accompanies all injuries to the body, and frequently it is the most serious consequence of the injury. In essence, shock is a circulatory reaction of the body (as a whole) to an injury (mechanical or emotional). While the changes to the circulatory system initially favor body resistance to the injury (by ensuring adequate blood supply to vital structures), they may progress to the point of circulatory failure and death. All aircrew members should be familiar with the signs and symptoms of shock so that the condition may be anticipated, recognized, and dealt with effectively. However, the best survival approach is to treat ALL moderate and severe injuries for shock. No harm will be done, and such treatment will speed recovery.

(2) Fluids. Normally, fluids administered by mouth are generally prohibited in the treatment of shock following severe injury. Such fluids are poorly absorbed when given by mouth, and they may interfere with later administration of anesthesia for surgery. In survival medicine, however, the situation is different in that the treatment being given is the final treatment. Survivors cannot be deprived of water for long periods just because they have been injured; in fact, their recovery depends upon adequate hydration. Small amounts of warm water, warm tea, or warm coffee given frequently early in shock are beneficial if the patient is conscious, can swallow, and has no internal injuries. In later shock, fluids by mouth are less effective as they are not absorbed from the intestines. Burns, particularly, require large amounts of water to replace fluid lost from injured areas. Alcohol should never be given to a person in shock or who may go into shock.

(3) Psychogenic Shock. Psychogenic shock is frequently noted during the period immediately following an emergency; for example, bailout. Psychogenic shock, which occurs even without injury, requires attention to limit it, both in degree and duration. The degree of this post-impact shock varies widely among individuals but its occurrence is almost universal. In reality, the survivor has passed through two major emergencies almost simultaneously; the aircraft incident leading to the survival situation, and the situation itself. Should the survivor be injured (and the majority of them are), a third emergency is added. It is not uncommon, then, that some psychogenic reaction with circulatory implications occurs. Resistance to this type of shock depends upon the individual's personality and the amount of training previously received. Treatment consists of stopping all activities (when possible), relaxing, evaluating the situation, and formulating a plan of action before the survival situation begins.

d. Fractures:

(1) Proper immobilization of fractures, dislocations, and sprains is even more important in survival medicine than in conventional first aid. Rather than merely making the patient comfortable during transport to eventual treatment, in survival medicine, the initial immobilization is part of the ultimate treatment. Immobilizing body parts to help control pain was discussed earlier. In addition, immobilization in proper position hastens healing of fractures and improves the ultimate functional result. In the survival situation, the immobilization must suffice for a relatively long period of time and permit the patient to maintain a fairly high degree of mobility. Materials for splinting and bandaging are available in most survival situations, and proper techniques are detailed in most first aid manuals.

(2) The reduction of fractures is normally beyond the scope of first aid; however, in the prolonged survival situation, the correction of bone deformities is necessary to hasten healing and obtain the greatest functional result. The best time for manipulation of a fracture is in the period immediately following the injury, before painful muscle spasms ensue. Traction is applied until overriding fragments of bone are brought into line, (check by the other limb) and the extremity is firmly immobilized. Frequently, it is advantageous to continue traction after reduction to ensure the proper alignment of the bones.

(3) As plaster casts are not available in the survival situation, improvising an immobilization device is necessary. This may be done by using several parallel, pliable willow branches, woven together with vines or parachute lines. Use care so that the extremity is not constricted when swelling follows the injury. In an escape and evasion situation, it may be necessary to preserve the mobility of the survivor after reduction of the fracture. This is difficult in fractures of the lower extremities, although tree limbs may be improvised as

crutches. With companions, the use of improvised litters may be possible.

(4) Reduction of dislocated joints is done similar to that of fractures. Gentle, but firm, traction is applied and the extremity is manipulated until it "snaps" back into place. If the survivor is alone, the problem is complicated but not impossible. Traction can still be applied by using gravity. The distal portion of the extremity is tied to (or wedged) into the fork of a tree or similar point of fixation. The weight of the body is then allowed to exert the necessary traction, with the joint being manipulated until the dislocation is reduced.

e. Infection:

(1) Infection is a serious threat to the survivor. The inevitable delay in definite medical treatment and the reality of the survival situation increases the chances of wound infection. Antibiotics may not be available in sufficient amounts in the survival situation. In survival medicine, one must place more emphasis on the prevention and control of infection by applying techniques used before the advent of antibiotics.

(2) Unfortunately, survivors have little control over the amount and type of infection introduced at the time of injury. However, they can help control the infection by wearing clean clothes. Use care to prevent additional infection into wounds. Wounds, regardless of the type or severity, should not be touched with dirty hands or objects. One exception to this rule is the essential control of arterial bleeding. Clothing should be removed from wounds to avoid contamination surrounding skin areas.

(3) All wounds should be promptly cleansed. Water is the most universally available cleaning agent, and should be (preferably) sterile. At sea level, sterilize water, by placing it in a covered container and boiling it for 10 minutes. Above 3,000 feet, water should be boiled for 1 hour (in a covered container) to ensure adequate sterilization. The water will remain sterile and can be stored indefinitely as long as it is covered.

(a) Irrigate wounds rather than scrubbing to minimize additional damage to the tissue. Foreign material should be washed from the wound to remove sources of continued infection. The skin adjacent to wounds should be washed thoroughly before bandaging. When water is not available for cleaning wounds, the survivor should consider the use of urine. Urine may well be the most nearly sterile of all fluids available and, in some cultures, is preferred for cleaning wounds. Survivors should use urine from the midstream of the urine flow.

(b) While soap is not essential to clean wounds, a bar of medicated soap placed in a personal survival kit and used routinely would do much to prevent the infection of seemingly inconsequential injuries. External antiseptics are best used for cleaning abrasions, scratches, and the skin areas adjacent to lacerations. Used in deep, larger wounds, antiseptics produce further tissue damage.

(c) Nature also provides antiseptics which can be used for wound care. The American mountain ash is found from Newfoundland south to North Carolina and its inner bark has antiseptic properties. The red berries contain ascorbic acid and have been eaten to cure scurvy. The Sweet Gum bark is still officially recognized as being an antiseptic agent. Water from boiled Sweet Gum leaves can also be used as antiseptic for wounds.

f. The "Open Treatment" Method. This is the only safe way to manage survival wounds. No effort should be made to close open wounds by suturing or by other procedure. In fact, it may be necessary to open the wound even more to avoid entrapment or infection and to promote drainage. The term "open" does not mean that dressings should not be used. Good surgery requires that although wounds are not "closed," nerves, bone, and blood vessels should be covered with tissue. Such judgment may be beyond the capability of the aircrew member, but protection of vital structures will aid in the recovery and ultimate function. A notable exception to "open treatment" is the early closure of facial wounds which interfere with breathing, eating, or drinking. Wounds, left open, heal by formation of infection resistant granulation tissue (proud flesh). This tissue is easily recognized by its moist red granular appearance, a good sign in any wound.

g. Dressings and Bandages. After cleansing, all wounds should be covered with a clean dressing. The dressing should be sterile; however, in the survival situation, any clean cloth will help to protect the wound from further infection. A proper bandage will anchor the dressing to the wound and afford further protection. Bandages should be snug enough to prevent slippage, yet not constrictive. Slight pressure will reduce discomfort in most wounds and help stop bleeding. Once in place, dressings should not be changed too frequently unless required. External soiling does not reduce the effectiveness of a dressing, and pain and some tissue damage will accompany any removal. In addition, changing dressings increases the danger of infection.

h. Physiological "Logistics." Despite all precautions, some degree of infection is almost universal in survival wounds. This is the primary reason for the "open" treatment advocated above. The human body has a tremendous capacity for combating infections if it is permitted to do so. The importance of proper rest and nutrition to wound healing and control of infection has been mentioned. In addition, the "logistics" of the injured part should be improved. The injury should be immobilized in a position to favor adequate circulation, both to and from the wound. Avoid constrictive clothing or bandages. Applying heat to an infected wound further aids in mobilizing local body defense measures. Lukewarm saltwater soaks will help draw out infection and promote oozing of fluids from the wound, thereby removing toxic products. Poultices, made of clean clay, shred-

ded bark of most trees, ground grass seed, etc., do the same thing.

i. Drainage. Adequate natural drainage of infected areas promotes healing. Generally, wicks or drains are unnecessary. On occasion, however, it may be better to remove an accumulation of pus (abscess) and insert light, loose packing to ensure continuous drainage. The knife or other instrument used in making the incision for drainage must be sterilized to avoid introducing other types of organisms. The best way to sterilize in the field is with heat, dry or moist.

j. Antibiotics. Antibiotics, when available, should be taken for the control of infection. Consensus is that the drug should be of the so-called "broad spectrum type;" that is, be effective against any micro-organisms rather than specific for just one or two types. The exact amount to be included in survival kits will vary with the drug and basic assumptions as to the number and types of infections to be expected. Remember that antibiotics are potency-dated items (shelf-life about 4 years), and including them in survival kits requires kit inspection and drug replacement with active medical stocks.

k. Debridement. (The surgical removal of lacerated, devitalized, or contaminated tissue.) The debridement of severe wounds may be necessary to minimize infection (particularly of the gas gangrene type) and to reduce septic (toxic) shock. In essence, debridement is the removal of foreign material and dead or dying tissue. The procedure requires skill and should only be done by nonmedical personnel in case of dire emergency. If required, follow these general rules. Dead skin must be cut away. Muscle may be trimmed back to a point where bleeding starts and gross discoloration ceases. Fat which is damaged tends to die and should be cut away. Bone and nerves should be conserved where possible and protected from further damage. Provide ample natural drainage for the potentially infected wound and delay final closure of the wound.

l. Burns:

(1) Burns, frequently encountered in aircraft accidents and subsequent survival episodes, pose serious problems. Burns cause severe pain, increase the probability of shock and infection, and offer an avenue for the loss of considerable body fluids and salts. Direct initial treatment toward relieving pain and preventing infection. Covering the wound with a clean dressing of any type reduces the pain and chance for infection. Further, such protection enhances the mobility of the patient and the capability for performing other vital survival functions. In burns about the face and neck, ensure the victim has an open airway. If necessary, cricothyroidotemy should be done before the patient develops extreme difficulties. Burns of the face and hands are particularly serious in a survival situation as they interfere with the capability of survivors to meet their own needs. Soaking certain barks (willow, oak, maple) in water soothes and protects burns by astringent action. This is a function of the acid content of the bark used.

(2) Maintenance of body fluids and salts is essential to recover from burns. The only way to administer fluids in a survival situation is by mouth; hence the casualty should ingest sufficient water early before the nausea and vomiting of toxicity intervenes. Consuming the eyes and blood (both cooked) of animals can help restore electrolyte levels if salt tablets are not available. NOTE: The survivor may also pack salt in personal survival kits to replace electrolytes (¼ teaspoon per quart of water).

m. Lacerations: Lacerations (cuts) are best left open due to the probability of infection. Clean thoroughly, remove foreign material, and apply a protective dressing. Frequently, immobilization will hasten the healing of major lacerations. On occasion (tactical), it may be necessary to close (cover) the wound, despite the danger of infection, in order to control bleeding or increase the mobility of the patient. If a needle is available, thread may be procured from parachute lines, fabric, or clothing, and the wound closed by "suturing." If suturing is required, place the stitches individually, and far enough apart to permit drainage of underlying parts. Do not worry about the cosmetic effect; just approximate the tissue. For scalp wounds, hair may be used to close after the wound is cleansed. Infection is less a danger in this area due to the rich blood supply.

n. Head Injuries. Injuries to the head pose additional problems related to brain damage as well as interfering with breathing and eating. Bleeding is more profuse in the face and head area, but infections have more difficulty in taking hold. This makes it somewhat safer to close such wounds earlier to maintain function. Cricothyroidotemy may be necessary if breathing becomes difficult due to obstruction of the upper airways. In the event of unconsciousness, watch the patient closely and keep him or her still. Even in the face of mild or impending shock, keep the head level or even slightly elevated if there is reason to expect brain damage. Do not give fluids or morphine to unconscious persons.

o. Abdominal Wounds. Wounds of the abdomen are particularly serious in the survival situation. Such wounds, without immediate and adequate surgery, have an extremely high mortality rate and render patients totally unable to care for themselves. If intestines are not extruded through the wound, a secure bandage should be applied to keep this from occurring. If intestine is extruded, do not replace it due to the almost certain threat of fatal peritonitis. Cover the extruded bowel with a large dressing and keep the dressing wet with any fluid that is fit to drink, or urine. The patient should lie on the back and avoid any motions that increase intra-abdominal pressure which might extrude more bowel. Keep the survivor in an immobile state or move on a litter. "Nature" will eventually take care of

the problem; either through death, or walling-off of the damaged area.

p. Chest Injuries. Injuries of the chest are common, painful and disabling. Severe bruises of the chest or fractures of the ribs require that the chest be immobilized to prevent large painful movements of the chest wall. The bandage is applied while the patient deeply exhales. In the survival situation, it may be necessary for survivors to wrap their own chest. This is more difficult but can be done by attaching one end of the long bandage (parachute material) to a tree or other fixed object, holding the other end in the hand, and slowly rolling body toward the tree, keeping enough counterpressure on the bandage to ensure a tight fit.

q. Sucking Chest Wounds. These wounds are easily recognized by the sucking noise and appearance of foam or bubbles in the wound. These wounds must be closed immediately before serious respiratory and circulatory complications occur. Ideally, the patient should attempt to exhale while holding the mouth and nose closed (Valsalva) as the wound is closed. This inflates the lungs and reduces the air trapped in the pleural cavity. Frequently, a taped, airtight dressing is all that is needed, but sometimes it is necessary to put in a stitch or two to make sure the wound is closed.

r. Eye Injuries. Eye injuries are quite serious in a survival situation due to pain and interference with other survival functions. The techniques for removing foreign bodies and for treating snow blindness are covered in standard first aid manuals. More serious eye injuries involving disruption of the contents of the orbit may require that the lids of the affected eye be taped closed or covered to prevent infection.

s. Thorns and Splinters. Thorns and splinters are frequently encountered in survival situations. Reduce their danger by wearing gloves and proper footgear. Their prompt removal is quite important to prevent infection. Wounds made by these agents are quite deep compared to their width which increases chances of infection by those organisms (such as tetanus) which grow best in the absence of oxygen. Removal of splinters is aided by the availability of a sharp instrument (needle or knife), needle nose pliers, or tweezers. Take care to get all of the foreign body out; sometimes it is best to open the wound sufficiently to properly cleanse it and to allow air to enter the wound. When cleaned, treat as any other wound.

t. Blisters and Abrasions. Care for blisters and abrasions promptly. Foot care is extremely important in the survival situation. If redness or pain is noted, the survivor should stop (if at all possible) to find and correct the cause. Frequently, a protective dressing or bandage and(or) adhesive will be sufficient to prevent a blister. If a blister occurs, do not remove the top. Apply a sterile (or clean) dressing. Small abrasions should receive attention to prevent infection. Using soap with a mild antiseptic will minimize the infection of small abrasions which may not come to the attention of the survivor.

u. Insect Bites. Bites of insects, leeches, ticks, chiggers, etc., pose several hazards. Many of these organisms transmit diseases, and the bite itself is likely to become infected, especially if it itches and the survivor scratches it. The body should be inspected frequently for ticks, leeches, etc., and these should be removed immediately. If appropriate and possible, the survivor should avoid infested areas. These parasites can best be removed by applying heat or other irritant to them to encourage a relaxation of their hold on the host. Then the entire organism may be gently detached from the skin, without leaving parts of the head imbedded. Treat such wounds as any other wound. Applying cold wet dressings will reduce itching, scratching, and swelling.

7-10. Illnesses. Many illnesses which are minor in a normal medical environment become major in a survival situation when the individual is alone without medications or medical care. Survivors should use standard methods (treat symptoms) to prevent expected diseases since treatment in a survival situation is so difficult. Key preventive methods are to maintain a current immunization record, maintain a proper diet, and exercise.

a. Food Poisoning. Food poisoning is a significant threat to survivors. Due to sporadic food availability, excess foods must be preserved and saved for future consumption. Methods for food preservation vary with the global area and situation. Bacterial contamination of food sources has historically caused much more difficulty in survival situations than the ingestion of so-called poisonous plants and animals. Similarly, dysentery or water-borne diseases can be controlled by proper sanitation and personal hygiene.

b. Treatment of Food Poisoning. If the food poisoning is due to preformed toxin; staphylococcus, botulism, etc. (acute symptoms of nausea, vomiting, and diarrhea soon after ingestion of the contaminated food), supportive treatment is best. Keep the patient quiet and lying down, and ensure the patient drinks substantial quantities of water. If the poisoning is due to ingestion of bacteria which grow within the body (delayed gradual onset of same symptoms), take antibiotics (if available). In both cases, symptoms may be alleviated by frequently eating small amounts of fine, clean charcoal. In PW situations, if chalk is available, reduce it to powder, and eat to coat and soothe the intestines. Proper sanitation and personal hygiene will help prevent spreading infection to others in the party or continuing reinfection of the patient.

Chapter 8

PW MEDICINE

8-1. Introduction:

a. Imprisoned PWs are, in the physical sense at least, under the control of their captors. Thus, the application of survival medicine principles will depend on the amount of medical service and supplies the captors can, and will, give to their prisoners. An enemy may both withhold supplies and confiscate survivor's supplies. Some potential enemies (even if they wanted to provide PW medical support) have such low standards of medical practice that their best efforts could jeopardize the recovery of the patient (figure 8-1).

b. An interesting and important sociological problem arises in getting medical care for PWs. How far should prisoners go in their efforts to get adequate rations and medicines for themselves or those for whom they are responsible? The Code of Conduct is quite specific concerning consorting with the enemy. Individuals must use considerable judgment in deciding whether to forget the welfare of fellow prisoners in order to follow the letter of the Code. Even more questionable is the individual who will offer such a justification for personal actions. Again, these questions involve more than purely medical consideration. In combat, there are apt to be frequent situations in which medical considerations are outweighed by more important ones.

8-2. History:

a. As in past wars, there were professional medical personnel among the captives in North Vietnam; however, these personnel were not allowed to care for the sick and injured as in the past. Medical care and assistance from the captors were limited and generally below comparable standards of the United States. Yet 566 men returned, most in good physical and psychological condition, having relied to a large extent on their own ingenuity, knowledge, and common sense in treating wounds and diseases. They were able to recall childhood first aid, to learn by trial and error, and to use available resources. Despite their measures of success in this respect, many released personnel felt that with some prior training, considerable improvement in self-help techniques was possible even in the most primitive conditions.

Figure 8-1. PW Medicine.

b. To determine how the services could help and to assist future PWs to care for themselves if the situation required it, the Medical Section of the Air Force Intelligence Service, with the Surgeon General of the Air Force, sponsored a 5-day seminar to examine the pertinent medical experiences of captivity and to recommend appropriate additions and changes in training techniques. As a basis for seminar discussion, Air Training Command provided data on the major diseases, wounds, and ailments, and the treatment methods used by the captives in Southeast Asia. Transcripts (325) of debriefing material were screened for medical data. Significant disease categories were established for analysis simplification based on the freqency of the problems encountered.

PROBLEMS	MAJOR CATEGORIES ESTABLISHED
Dysentery	Trauma (lacerations, burns, fractures)
Fungus	Gastrointestinal problems
Dental problems	Communicable diseases
Intestinal problems	Nutritional diseases
Fractures	Dermatological ailments
Lacerations	Dental problems
Respiratory ailments	
Burns	

In examining these major categories, attention was focused on those medical problems considered significant by the prisoners themselves in evaluating their primitive practices (self-help).

8-3. Trauma:

a. Most of the prisoners began their captivity experience with precapture injuries—burns, wounds, fractures, and lacerations. Other injuries were the result of physical abuse while a prisoner. Most of these individuals, upon their return, expressed a need to know more about managing their injuries in captivity and also what to expect about the long-term effects of injuries. It was not evident to them that the practice of a few simple rules will generally lead to acceptable results in wound treatment, and that much can be done after repatriation to correct cosmetic and functional defects.

b. The groundwork for management of injuries should begin well before an individual enters the captivity or survival environment. The treatment of injuries in survival or captivity depends primarily on providing the body the best possible circumstances to "repair itself." It is vital, therefore, to have the body in the best possible physical condition before exposure to survival or captivity. This means good cardiovascular conditioning, good muscle strength and tone, and good nutritional status. Physiological and nutritional status will markedly influence the rate and degree of healing in response to injury. The opportunity for maintaining the best possible physical conditioning and nutritional status in captivity will be greatly reduced. (Once in a captivity or survival setting, it is important to do everything possible to maintain a good physiological and nutritional status.)

8-4. Gastrointestinal Problems:

a. Diarrhea. This was a common ailment in the prison environment, not only in Vietnam, but also in WW II and Korea. It plagued the forces of North Vietnam and the allied forces. This was the second most frequent malady afflicting the Viet Cong forces. The causative factors of this almost epidemic state were varied. A variety of infectious agents gaining access to the body by use of contaminated food and water certainly contributed to the problem. Equally important as causative agents were the low level of sanitation and hygiene practices within the camps. Psychogenic responses to unappetizing diet, nutritional disturbances, and viral manifestations also contributed.

(1) Captor Therapy. This consisted primarily of local or imported antidiarrheal agents, antibiotics, and vitamins. Appropriate diet therapy was instituted.

(2) Captive Self-Therapy. After instituting diet restrictions (solid food denial and increased liquid intake), afflicted personnel were administered "concoctions" of banana skins, charcoal, chalk, or tree bark tea.

(3) Treatment Evaluation. The accepted therapy for diarrhea focuses on the causative factors which in the captivity experience were largely neglected. From a symptomatic perspective, the principles of self-treatment are simple to master: restrict intake to nonirritating foods (avoid vegetables and fruits), establish hygienic standards, increase fluid intake, and, when available, use antidiarrheal agents. The prisoners often resorted to a more exotic therapeutic regimen consisting of banana skins, charcoal, chalk, salt restriction, rice, or coffee. Charcoal, chalk, and the juice of tree barks have a scientific basis for their therapeutic success. Inasmuch as diarrhea was a source of concern and a disability for the North Vietnamese as well as the captives, therapy was often offered on request and was appropriate and successful.

(4) Conclusion. Diarrhea was frequent among PWs during captivity. Seldom fatal, it was disabling and a source of concern to those afflicted. Most captives were treated on demand and improved. This condition lends itself to some form of self-therapy through an understanding of its physiological derangements. The PW re-

sponded with intelligence, common sense, and a reasonably effective self-help regimen.

b. Dysentery. From a symptomatic perspective, dysentery is a severe form of diarrhea with passage of mucous and blood. Treatment and conclusions are similar to those for diarrhea.

c. Worms and Intestinal Parasites. Worms were extremely common among the captives. Twenty-eight percent of the released prisoners indicated worms as a significant medical problem during captivity. Worms often caused gastrointestinal problems similar to those resulting from a variety of other causes. The pin worm appears to have been the primary cause. This is not surprising, as its distribution is worldwide and the most common cause of helminthic infection of people in the United States. It requires no intermediate host; hence, infection is more rapidly acquired under poor hygienic conditions so commonly found in warm climates and conditions similar to the captivity environment. Seldom fatal, worms are significant, as they can lower the general resistance of the patient and may have an adverse affect on any intercurrent illness.

(1) Captor Therapy. This consisted of antihelminthic agents (worm medicine) dispensed without regularity, but with satisfactory results.

(2) Captive Therapy. The nuisance and irritating aspects of worms led to severe rectal itching, insomnia, and restlessness. This motivated the prisoner to find some form of successful self-therapy. Prevention was a simple and readily obtainable goal. Shoes were worn when possible; hands were washed after defecation; and fingernails were trimmed close and frequently. Peppers, popular throughout the centuries in medicine, contain certain substances chemically similar to morphine. They are effective as a counter-irritant for decreasing bowel activity. Other "house remedies" popular among the captives included drinking saltwater (a glass of water with 4 tablespoons of salt added), eating tobacco from cigarettes (chewing up to two or three cigarettes and swallowing them), and infrequently drinking various amounts of kerosene. All of these remedies have some degree of therapeutic effectiveness, but are not without danger and therefore deserve further comment. Saltwater alters the environment in the gastrointestinal tract and can cause diarrhea and vomiting. Too large an amount can have harmful effects on body fluid mechanisms and can lead to respiratory complications and death. Tobacco contains nicotine and historically was popular in the 19th century as an emetic expectorant and was used for the treatment of intestinal parasites. Nicotine is, however, one of the most toxic of all drugs and can cause death when more than 60 mg is ingested. A single cigarette contains about 30 mg of nicotine, so the captives who ate two or more cigarettes had been using a cure more dangerous than the disease. Kerosene is also toxic with 3 to 4 ounces capable of causing death. It is particularly destructive to the lungs and if through vomiting it were to make its way into the trachea and eventually the lungs, the complications would then again be far worse than the presence of worms.

(3) Treatment Evaluation. The antihelminthics therapy used by the captors was extremely effective. The problem during confinement was the nonavailability of such medication on demand. In addition, the inability to practice proper hygienic standards assured the continuation of, and reinfection with, worms.

(4) Conclusion. Worm infection in confinement is common and expected. It is seldom fatal, but contributes to general disability and mental depression due to its nuisance symptoms. Under certain circumstances, worms can assist in the spread of other diseases. The principle to follow in self-care is simple—use as high a hygienic standard as possible, and use medication causing bowel paristalsis and worm expulsion. Substances which interfere with the environment of the worms will aid in their expulsion. The toxic "house remedies" must be weighed against their possible complications.

8-5. Hepatitis. Infection of the liver was fairly common in some camps and present among the prison population throughout the captivity experience. Diagnosis was usually made on the basis of change of skin color to yellow (jaundice).

a. Captor Treatment. The Vietnamese seemed to have followed the standard therapy of rest, dietary management, and vitamin supplementation. They also displayed a heightened fear of the disease and avoided direct contact, when possible, with those afflicted.

b. Captive Therapy. For the most part, it parallels the therapy of the captors. This disease allows for little ingenuity or inventiveness of therapy.

c. Comments. Hepatitis is worldwide. Presumably most cases of hepatitis in captivity were viral in origin and easily disseminated to fellow prisoners. Conditions of poor sanitation and hygiene with close communal living foster its spread. Prevention through proper hygienic practices is the most effective tool. Equally important is an understanding of the disease characteristics. The majority of the cases recovered completely and less than 1 percent succumbed to this disease.

8-6. Nutritional Deficiencies. Symptoms attributed to malnutrition were frequent in the early years of confinement and continued up through 1969. The use of polished rice and the lack of fresh fruits and vegetables contributed to vitamin and protein deficiencies. From 1969 through 1973, food supplements were provided, and by release time, few obvious manifestations of diseases were present among those returning. The primary problems during the early years were vitamin deficiencies.

a. Vitamin B Deficiency (Beri-Beri). Presumably present among several PWs (especially those confined to the Briarpatch (Xom Ap Lo) about 15 miles west of

Sontay), it was rarely diagnosed on return. Its primary manifestation was pain in the feet described by the captives as "like a minor frostbite that turned to shooting pains."

(1) Captor Therapy. Prisoners were treated with vitamin injections and increased caloric content.

(2) Captive Therapy. Increasing caloric intake by eating anything of value. No specific self-care program existed for this malady.

(3) Comments. Beri-beri is a nutritional disease resulting from a deficiency of vitamin B (Thiamine). It is widespread in the Orient and in tropical areas where polished rice is a basic dietary staple. Of the various forms of the disease, dry beri-beri would seem most important to the confinement condition. Early signs and symptoms of the disease include muscle weakness and atrophy, loss of vibratory sensation over parts of the extremities, numbness, and tingling in the feet. From the comments of the PWs, it is difficult to formulate a diagnosis. Modern therapy consists of vitamin B or sources of the vitamin in food (such as green peas, cereal grains, and unpolished rice).

(4) Conclusion. In Vietnam, the possible early onset of the dry form of beri-beri was encountered. This is supported by the symptoms described and by the existence of dietary shortages of vitamin B and other nutritional deficiency.

b. Vitamin A Deficiency. There were several reported cases of decreased vision (primarily at night) attributed to vitamin A deficiency. This problem usually occurred during periods of punishment or politically provoked action when food was withheld as part of the discipline. The condition responded well to increased caloric intake and deserves little special mention. An understanding of the transient nature of this problem and its remedial response to therapy is important.

8-7. Communicable Diseases. Some communicable diseases were endemic in North Vietnam and certainly responsible for large scale disability among the personnel of the enemy forces. Plague, cholera, and malaria are frequent and a serious public health menace. Thanks to the immunization practices of the American forces, these diseases were of little concern to Americans during their captivity.

8-8. Skin Diseases:

a. Lesions. Dermatological lesions were common to the various prison experiences. Their importance lies not in their lethality (as they apparently did not cause any deaths), but for their irritant quality and the debilitating and grating effect on morale and mental health. Boils, fungi, heat rash, and insect bites appeared frequently and remained a problem throughout the captivity experience.

b. Boils and Blisters. A deep-seated infection usually involves the hair follicles and adjacent subcutaneous tissue, especially parts exposed to constant irritation.

(1) Captor Treatment. Prisoner complaints about the presence of boils usually brought about some action by the captors. Treatment varied considerably and obviously depended on the knowledge of medics, doctors treating their prisoners, the availability of medical supplies, and the current camp policy. For the most part, systemic antibiotics, sulfa, and tetracycline were administered. In other instances, the boils were lanced or excised and treated with topical astringents.

(2) Captive Treatment. As the medics normally responded to pleas about boils, self-treatment was practiced primarily when there was distrust of captor techniques. Prisoners would attempt to lance the boil with any sharp instrument such as needle, wire, splinters, etc., and exude their contents by applying pressure. The area was then covered with toothpaste and, when available, iodine.

(3) Comments. As noted above, the boil is an infection of hair follicles. It is more frequent in warm weather and aggravated by sweat which provides ideal conditions for the bacteria. Boils seldom appear singularly. Once present, they are disseminated by fingers, clothing, and discharges from the nose, throat, and groin. Modern therapy consists of hot compresses to hasten localization, and then conservative incision and drainage. Topical antibiotics and systemic antibiotics are then used. Boils increase in frequency with a decrease in resistance as seen in malnutrition and exhaustion states in a tropical environment. This almost mimics the prison conditions.

(4) Conclusions. Self-help treatment is limited. Of importance here is sterility when handling the boils, cleanliness, exposure to sunlight, keeping the skin dry, and getting adequate nutrition. The disease is self-limiting and not fatal. The application of any material or medication with a detergent effect may be used (soaks in saline, soap, iodine, and topical antibiotics).

8-9. Fungal Infections. Fungal infections were also a common skin problem for those in Southeast Asia captivity. As with other skin lesions, they are significant for their noxious characteristics and weakening effect on morale and mental health. Superficial fungal infections of the skin are widespread throughout the world. Their frequency among PWs reflects the favorable circumstances of captivity for cultivating fungal infections.

a. Captor Therapy. Treatment consisted of medication described by many PWs as iodine and the occasional use of sulfa powder.

b. Captive Therapy. Treatment (often the result of memory of childhood experiences and trial and error observations) consisted of the removal of body hair (to prevent or improve symptoms in the case of heat rash), exposure to sunlight to dry out fungal lesions, and de-

velopment of effective techniques to foster body cooling and to decrease heat generation. Considerable effort was directed at keeping the body clean.

c. Comments. Superficial Dermatoses (skin lesions) due to fungi were common. Their invasive powers are at best uniformly weak, and because of this, infections are limited to the superficial portions of the skin and seldom by themselves fatal. Modern therapy since 1958 has relied heavily on an oral antifungal agent effective against many superficial fungi. This drug is expensive and not available in many parts of the world. Several lotions and emulsions can be used with some success. Elemental iodine is widely used as a germicide and fungicide. It is an effective antiseptic and obviously found favor in North Vietnam because of its availability. Without professional therapy, self-help, although limited in scope, can be effective. The principle of wet soaks for dry lesions and dry soaks for wet lesions is a fairly reliable guide. The use of the Sun as a drying agent can also be very effective.

d. Conclusions. Skin problems are common to the captivity environment. More importantly, extreme personal discomfort, accompanied by infection, was detrimental to the physical and mental well-being of the prisoner.

8-10. Dental Problems. These were common among all captives, not only during confinement, but also before capture. They were secondary to facial injury during egress, or caused by physical abuse during interrogation. Periodontitis (inflammation of tissue surrounding the tooth), pyorrhea (discharge of pus), and damage to teeth consistent with poor hygiene and "wear and tear" were also present.

a. Specific Complaints. Pain associated with the common toothache represented one of the most distressing problems faced by the PW. It affected the PW's nutrition and robbed the PW of the physical pleasure of eating (a highlight of isolated captivity). The inability of the PW to adequately deal with this problem caused persistent anxiety and decreased the ability to practice successful resistance techniques. In a few isolated instances, PWs actually considered collaboration with the captors in exchange for treatment and relief from tooth pain.

b. Captor Treatment. Treatment varied considerably and was no doubt influenced by political considerations. "Dentists" were infrequently available in camps before 1969. Cavities were filled, although usually inadequately, with subsequent loss of the filling. Use of local anesthesia also varied depending on the dentist providing care.

c. Captive Therapy. The PWs often chose to treat themselves rather than seek or accept prison dentistry when it was available. Abscesses were lanced with sharp instruments made locally out of wood, bamboo, or whatever was available. Brushing was excessive, again using whatever was available; chew sticks common to Asia were widely used. Aspirin (ASA), when available, was applied directly to the tooth or cavity.

d. Commentary. The self-help practices noted above had many positive aspects. The basic principle of maintaining a well-planned cleaning program using fiber, brushes, or branches certainly contributed to the relatively low incidence of cavities and infection among the prisoners. The lancing of abscesses using bamboo sticks, although not a professional maneuver, has merit insofar as the pressure is relieved and the tendency to develop into cellulitis (widespread infection) decreased. The application of aspirin directly into the cavity should be discouraged as might the application of any other substance not directly produced for this purpose.

e. Conclusions. The most effective tool against dental complications in captivity is proper preventive dentistry. The present program of the three services, if adhered to, is adequate to ensure a high state of dental hygiene while captive.

8-11. Burns. Burns were an extremely frequent injury among PWs. Severity ranged from first through third degree and occurred frequently on hands and arms.

a. Captor Treatment. For the most part, burns were treated by captors by cleaning the burns and applying antiseptics and bandages. The results obtained were, by and large, inadequate, with frequent infections and long-term debilitation.

b. Captive Treatment. No specific treatment was developed among the PWs for burns. Reliance for some form of therapy was almost completely left to the captor.

c. Commentary. Burns are extremely painful and can severely interfere with the ability to escape or to survive in captivity. The basic principle here is prevention.

d. Prevention. Adequate protection of exposed surfaces while flying (flame-retardant suits, gloves, boots, and helmet with visor down) is the best preventive action.

8-12. Lacerations and Infections:

a. Treatment. Captor treatment for lacerations and infections reflected the medical standards in North Vietnam and their domestic priorities. Wound and infection treatment varied considerably from being adequate to substandard and malpracticed. Obviously, the availability of trained physicians, a changing political climate, and difficulty in obtaining sophisticated medical supplies and equipment dictated and influenced the quality of the care delivered. The prisoners could do little professionally with this type of injury. As with diseases, the maintenance of good nutritional standards, cleanliness, and "buddy self-care" were the basic treatments.

b. Comments. When soft tissue is split, torn, or cut, there are three primary concerns—bleeding, infection, and healing of the wound.

(1) Bleeding is the first concern and must be controlled as soon as possible. Most bleeding can be controlled by direct pressure on the wound and that should be the first treatment used. If that fails, the next line of defense would be the use of classic pressure points to stop hemorrhaging. And the last method for controlling hemorrhage would be the tourniquet. The tourniquet should be used only as a last resort. Even in more favorable circumstances where the tourniquet can be applied as a first aid measure and left in place until trained medical personnel remove it, the tourniquet may result in the loss of the limb. The tourniquet should be used only when all other measures have failed, and it is a life and death matter. To control bleeding by direct pressure on the wound, sufficient pressure must be exerted to stop the bleeding, and that pressure must be maintained long enough to "seal off" the bleeding surfaces. Alternate pressing and then releasing to see if the wound is still bleeding is not desirable. It is best to apply the pressure and keep it in place for up to 20 minutes. Oozing blood from a wound of an extremity can be slowed or stopped by elevating the wound above the level of the heart.

(2) The next concern is infection. In survival or captivity, consider all breaks in the skin due to mechanical trauma contaminated, and treat appropriately. Even superficial scratches should be cleaned with soap and water and treated with antiseptics, if available. Antiseptics should generally not be used in wounds which go beneath the skin's surface since they may produce tissue damage which will delay healing. Open wounds must be thoroughly cleansed with boiled water. Bits of debris such as clothing, plant materials, etc., should be rinsed out of wounds by pouring large amounts of water into the wounds and ensuring that even the deepest parts are clean. In a fresh wound where bleeding has been a problem, care must be taken not to irrigate so vigorously that clots are washed away and the bleeding resumes. Allow a period of an hour or so after the bleeding has been stopped before beginning irrigation with the boiled water. Begin gently at first, removing unhealthy tissue, increasing the vigor of the irrigation over a period of time. If the wound must be cleaned, use great care to avoid doing additional damage to the wound. The wound should be left open to promote cleansing and drainage of infection. In captivity, frequently deep open wounds will become infested with maggots. The natural tendency is to remove these maggots, but actually, they do a good job of cleansing a wound by removing dead tissue. Maggots may, however, damage healthy tissue when the dead tissue is removed. So the maggots should be removed if they start to affect healthy tissue. Remember that it is imperative that the wound be left open and allowed to drain.

(3) An open wound will heal by a process known as secondary intention or granulation. During the healing phase, the wound should be kept as clean and dry as possible. For protection, the wound may be covered with clean dressings to absorb the drainage and to prevent additional trauma to the wound. These dressings may be loosely held in place with bandages (clean parachute material may be used for dressings and bandages). The bandages should not be tight enough to close the wound or to impair circulation. At the time of dressing change, boiled water may be used to gently rinse the wound. The wound may then be air dried and a clean dressing applied. (The old dressing may be boiled, dried, and reused.) Nutritional status is interrelated with the healing process, and it is important to consume all foods available to provide the best possible opportunity for healing.

c. Conclusions. Obviously the PW is at a distinct disadvantage in treating wounds, lacerations, and infections without modern medicine. Yet, knowledge of the basic principles mentioned above, locally available equipment and resources, and optimism and common sense can help a survivor to maintain life.

8-13. Fractures and Sprains. Fractures and sprains often occurred during shootdown and(or) egress from the aircraft. They also occurred during evasion attempts.

a. Captor Treatment. As with other treatment, treatment of sprains and fractures varied considerably depending on the severity of the injury and the resources available for treatment. Even after immediate treatment or surgical procedures, there was little followup therapy. Prisoners were usually returned to camp to care for themselves or to rely on the help of fellow prisoners.

b. Captive Treatment. Captive therapy was primarily that of helping each other to exercise or immobilize the injured area, and in severe cases, to provide nursing care.

c. Comments:

(1) An acute nonpenetrating injury to a muscle or joint can best be managed by applying cold as soon as possible after the injury. Icepacks or cold compresses should be used intermittently for up to 48 hours following the injury. This will minimize hemorrhage and disability. Be careful not to use snow or ice to the point where frostbite or cold injury occurs. As the injured part begins to become numb, the ice should be removed to permit rewarming of the tissues. Then the ice can be reapplied. Following a period of 48 to 72 hours, the cold treatment can be replaced by warm packs to the affected part. A "sprain or strain" may involve a wide variety of damage ranging from a simple bruise to deep hemorrhage or actual tearing of muscle fibers, ligaments, or tendons. While it is difficult to establish specific guidelines for treatment in the absence of a specific diagnosis, in general, injuries of this type require some period of rest (immobilization) to allow healing. The period of rest is followed by a period of rehabilitation (massage and exercise) to restore function. For what appears to be a simple superficial muscle problem, a period of 5 to 10

days rest followed by a gradual progressive increase in exercise is desirable. Pain should be a limiting factor. If exercise produces significant pain, the exercise program should be reduced or discontinued. In captivity, it is probably safest to treat severe injuries to a major joint like a fracture with immobilization (splint, cast) for a period of 4 to 6 weeks before beginning movement of the joint.

(2) Bone fractures are of two general types, open and closed. The open fracture is associated with a break in the skin over the fracture site which may range all the way from a broken bone protruding through the skin to a simple puncture from a bone splinter. The general goals of fracture management are: restore the fracture to a functional alignment; immobilize the fracture to permit healing of the bone; and rehabilitation. Restoring or reducing the fracture simply means realigning the pieces of bones, putting the broken ends together as close to the original position as possible. The natural ability of the body to heal a broken bone is remarkable and it is not necessary that an extremity fracture be completely straight for satisfactory healing to occur. In general, however, it is better if the broken bone ends are approximated so that they do not override. Fractures are almost always associated with muscle spasms which become stronger with time. The force of these muscle spasms tends to cause the ends of the broken bones to override one another, so the fracture should be reduced as soon as possible. To overcome the muscle spasm, force must be exerted to reestablish the length of the extremity. Once the ends of the bone are realigned, the force of the muscle spasm tends to hold the bones together. At this point, closed fractures are ready to be immobilized, but open fractures require treatment of the soft tissue injury in the manner outlined earlier. In other words, the wound must be cleansed and dressed, then the extremity should be immobilized. The immobilization preserves the alignment of the fracture and prevents movement of the fractured parts which would delay healing. For fractures of long bones of the body, it becomes important to immobilize the joints above and below the fracture site to prevent movement of the bone ends. In a fracture of the mid forearm, for example, both the wrist and the elbow should be immobilized. In immobilizing a joint, it should be fixed in a "neutral" or functional position. That is, neither completely straight nor completely flexed or bent, but in a position about midway between. In splinting a finger, for example, the finger should be curved to about the same position the finger would naturally assume at rest.

(3) A splint of any rigid material such as boards, branches, bamboo, metal boot insoles, or even tightly rolled newspaper may be almost as effective as plaster or mud casts. In conditions such as continuous exposure to wetness, the splint can be cared for more effectively than the plaster or mud cast. In cases where there is a soft tissue wound in close proximity to the fracture, the splint method of immobilization is more desirable than a closed cast because it permits change of dressing, cleaning, and monitoring of the soft tissue injury. The fracture site should be loosely wrapped with parachute cloth or soft plant fibers; then the splints can be tied in place extending at least the entire length of the broken bone and preferably fashioned in such a way as to immobilize the joint above and below the fracture site. The splints should not be fastened so tightly to the extremity that circulation is impaired. Since swelling is likely to occur, the bindings of the splint will have to be loosened periodically to prevent the shutting off of the blood supply.

(4) The time required for immobilization to ensure complete healing is very difficult to estimate. In captivity, it must be assumed that healing time will, in general, be prolonged. This means that for a fracture of the upper extremity of a "nonweight-bearing bone," immobilization might have to be maintained for 8 weeks or more to ensure complete healing. For a fracture of the lower extremity or a "weight-bearing bone," it might require 10 or more weeks of immobilization.

(5) Following the period of immobilization and fracture healing, a program of rehabilitation is required to restore normal functioning. Muscle tone must be reestablished and the range of motion of immobilized joints must be restored. In cases where joints have been immobilized, the rehabilitation program should be started with "passive range of motion exercises." This means moving the joint through a range of motion without using the muscles which are normally used to move that joint. For example, if the left wrist has been immobilized, a person would begin the rehabilitation program by using the right hand to passively move the left wrist through a range of motion which can be tolerated without pain. When some freedom of motion of the joint has been achieved, the individual should begin actively increasing that range of motion using the muscles of the joint involved. Do not be overly forceful in the exercise program—use pain as a guideline—the exercise should not produce more than minimal discomfort. Over a period of time, the joint movement should get progressively greater until the full range of motion is restored. Also, exercises should be started to restore the tone and strength of muscles which have been immobilized. Again, pain should be the limiting point of the program and progression should not be so rapid as to produce more than a minimal amount of discomfort.

8-14. Summary. Common sense and basic understanding of the type of injuries are most helpful in avoiding complication and debilitation. Adequate nourishment and maintenance of physical condition will materially assist healing of burns, fractures, lacerations, and other injuries—the body will repair itself.

8-15. Conclusions. In the management of trauma and burns in captivity or survival, remember that the body will do the healing or repair, and the purpose of the "treater" is to provide the body with the best possible atmosphere to conduct that self-repair. Some general principles are:

a. Be in the best possible physical, emotional, and nutritional status before being exposed to the potential survival or captivity setting.

b. Minimize the risk of injury at the time of survival or captivity by following appropriate safety procedures and properly using protective equipment.

c. Maintain the best possible nutritional status while in captivity or the survival setting.

d. Don't overtreat!!! Overly vigorous treatment can do more harm than good.

e. Use cold applications for relief of pain and to minimize disability from burns and soft tissue strains or sprains.

f. Clean all wounds by gentle irrigation with large amounts of the cleanest water available.

g. Leave wounds open.

h. Splint fractures in a functional position.

i. After the bone has healed, begin an exercise program to restore function.

j. Remember that even improperly healed wounds or fractures may be improved by cosmetic or rehabilitative surgery and treatment upon rescue or repatriation.

Part Four

FACTS AND CONDITIONS AFFECTING A SURVIVOR

Chapter 9

WEATHER

9-1. Introduction. History records many attempts by people of ancient cultures to understand the heavens. In the primitive past, the ability to predict weather was of primary importance. It was by observing the stars and other celestial bodies that these early societies could predict the coming of the seasons and therefore the weather patterns that played such a large role in their survivability. When people today are forced to live under the primitive conditions of a survival situation, they are no different from those who have struggled before them against those same conditions.

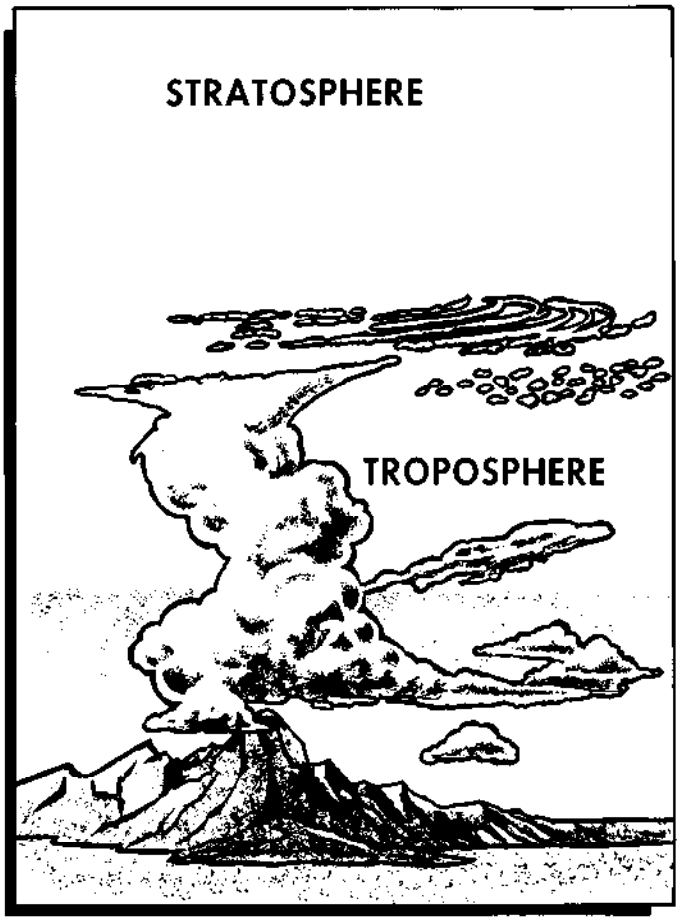

Figure 9-1. Structure of the Atmosphere.

9-2. Knowledge of Weather. However, today's participants in the age old struggle against nature must (out of necessity) still be concerned with the effect of weather. It is still true that weather cannot be controlled, but the person who is prepared (through knowledge) will be more successful.

a. Weather is not the same as climate. Weather is the state of the atmosphere, with respect to wind, temperature, cloudiness, moisture, pressure, humidity, etc. Climate, on the other hand, is the type of weather condition generally prevailing over a region throughout the year, averaged over a series of years.

b. The atmosphere extends upward from the surface of the Earth for a great many miles, gradually thinning as it approaches its upper limit. Near the Earth's surface, the air is relatively warm due to contact with the Earth. As altitude increases, the temperature decreases by about 3.5°F for every 1,000 feet until air temperature reaches about 67°F below zero at 7 miles above the Earth.

c. To understand where the weather patterns originate, a brief familiarization of the "layers" or structure of the atmosphere is needed (figure 9-1). The atmosphere is divided into two layers. The upper layer is the "stratosphere" where the temperature remains constant. The lower layer is the "troposphere" where the temperature changes. Nearly all weather occurs in this lower layer which begins at the Earth's surface and extends upwards for 6 to 10 miles.

9-3. Elements Affecting Weather. Weather conditions in the troposphere and on the Earth are affected by four elements: temperature, air pressure, wind, and moisture.

a. Temperature is the measure of the warmth or coldness of an object or substance and, for this discussion, the various parts of the atmosphere. The sunlight entering the atmosphere reaches the Earth's surface and warms both the ground and the seas. Heat from the ground and the seas then warms the atmosphere. The atmosphere absorbs the heat and prevents it from escaping into space. This process is called the greenhouse effect because it resembles the way a greenhouse works. Once the Sun sets, the ground cools more slowly than the air because it is a better conductor of heat. At night the ground is warmer than the air, especially under a clear dry sky. The ground cools more slowly than the humid nights. Temperature changes near the ground for other reasons. Dark surfaces are warmer than light-colored surfaces. Evening air settles in low areas and valleys creating spots colder than higher elevations.

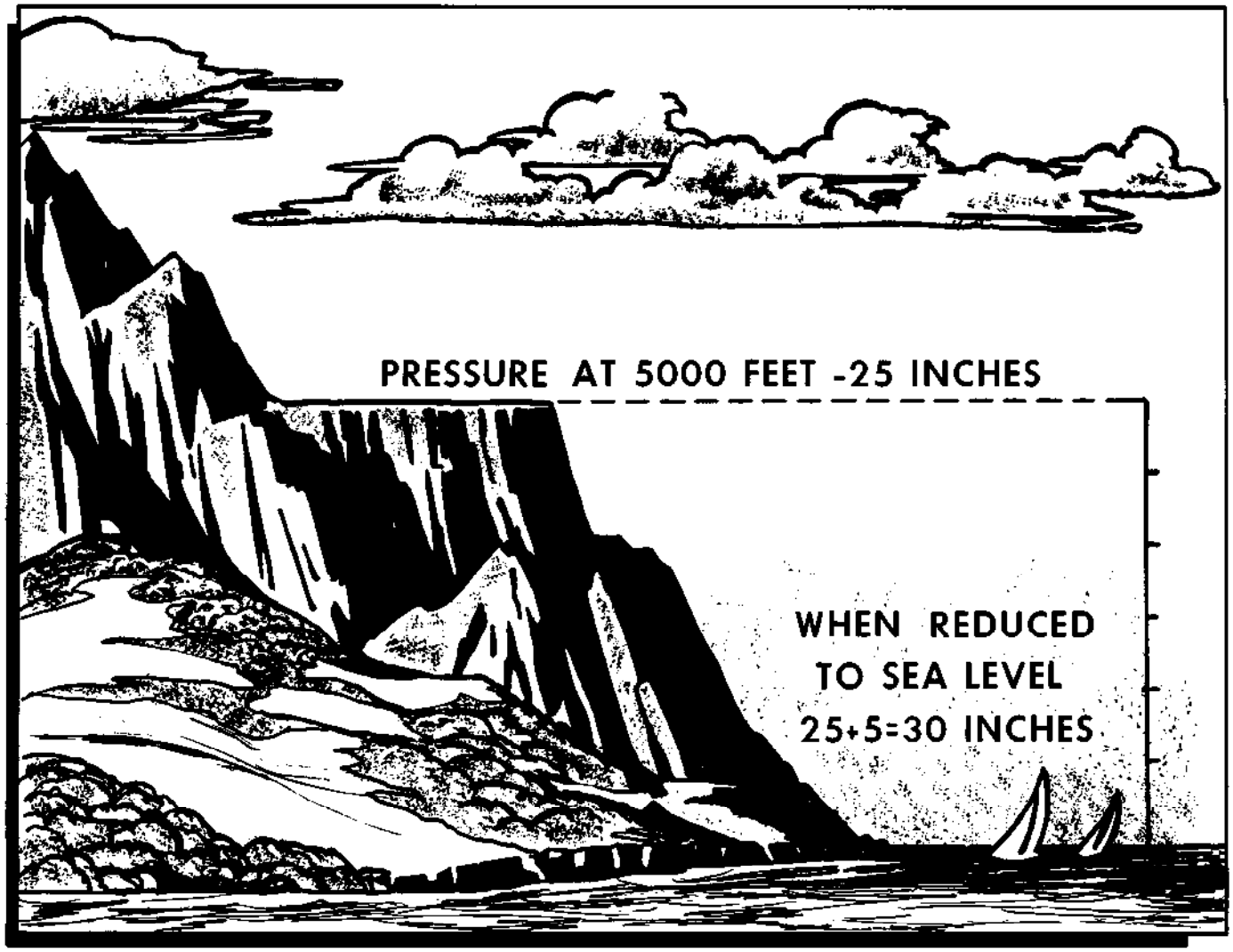

Figure 9-2. Air Pressure.

Seas, lakes, and ponds retain heat and create warmer temperatures at night near shore. The opposite is true during the day, especially in the spring when lakes are cold. On the beach, daytime temperatures will be cooler than the temperatures on land further from shore. Knowing this will help determine where a survivor should build a shelter.

b. Air pressure is the force of the atmosphere pushing on the Earth. The air pressure is greatly affected by temperature. Cool air weighs more than warm air. As a result, warm air puts less pressure on the Earth than does cool air. A low-pressure area is formed by warm air whereas cool air forms a high-pressure area (figure 9-2).

c. Wind is the movement of air from a high-pressure area to a low-pressure area. The larger the difference in pressure, the stronger the wind. On a global scale, the air around the Equator is replaced by the colder air around the poles (figure 9-3). This same convection of air on a smaller scale causes valley winds to blow upslope during the day and down the mountainside at night. Cool air blows in from the ocean during the day due to the heating and rising of the air above the land and reverses at night (figures 9-4 and 9-5). This movement of air creates winds throughout the world. When cool air moves into a low-pressure area, it forces the air that was already there to move upward. The rising air expands and cools.

d. Moisture enters the atmosphere in the form of water vapor. Great quantities of water evaporate each day from the land and oceans causing vapor in the air called humidity. The higher the humidity, the higher the moisture content in the air. Air holding as much moisture as possible is saturated. The temperature at which the air becomes saturated is called the dew point. When the temperature falls below the dew point, moisture in the air condenses into drops of water. Low clouds called fog may develop when warm, moist air near the ground is cooled to its dew point. A cooling of the air may also

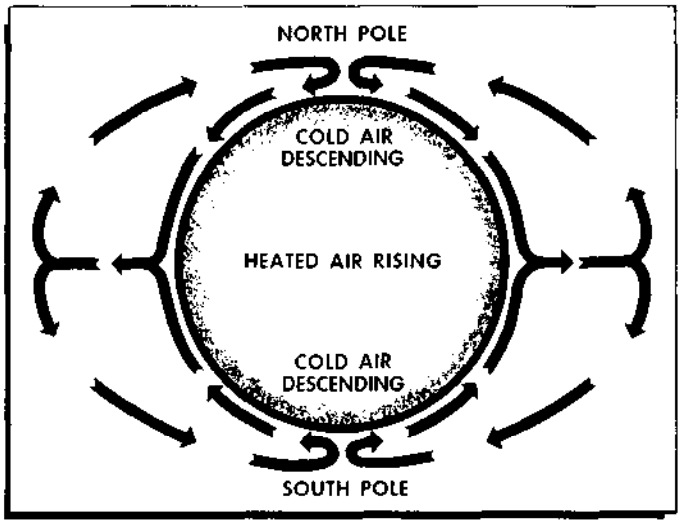

Figure 9-3. Wind Movement.

cause moisture to fall to the Earth as precipitation (rain, snow, sleet, or hail).

9-4. Circulation of the Atmosphere. If the Earth did not rotate, wind would move directly from the high-pressure areas of the poles to the low-pressure areas of the Equator. The movement of air between the poles and the Equator would go on constantly.

a. The rotation of the Earth prevents winds from the poles and the Equator from moving directly north or south. The Earth rotates from west to east, and as a result, winds moving toward the Equator seem to curve toward the west. Winds moving away from the Equator seem to curve toward the east. This is known as the Coriolis effect (figure 9-6). This effect results in winds circling the Earth in wide bands. These prevailing winds are divided into six belts which are known as the trade winds, the prevailing westerlies, and the polar easterlies; all three are found in both the Northern and Southern Hemispheres (figure 9-7).

(1) The winds blowing toward the Equator are known as the trade winds. The air above the Equator is so hot it is always rising. The north and south trade winds move in to take the place of the rising air. The Coriolis effect makes the trade winds appear to move from the east. The weather in the region of the trade winds moves from east to west because of the Earth's rotation. The doldrums is the region where the trade winds from the north and south meet near the Equator. The doldrums is usually calm, but it is quite rainy and may have periods of gusty winds.

(2) The prevailing westerlies blow away from the Equator. They occur north of the trade winds in the Northern Hemisphere and south of the trade winds in the Southern Hemisphere. The prevailing westerlies seem to move from the west because of the Coriolis effect. The weather in the region of these winds blows

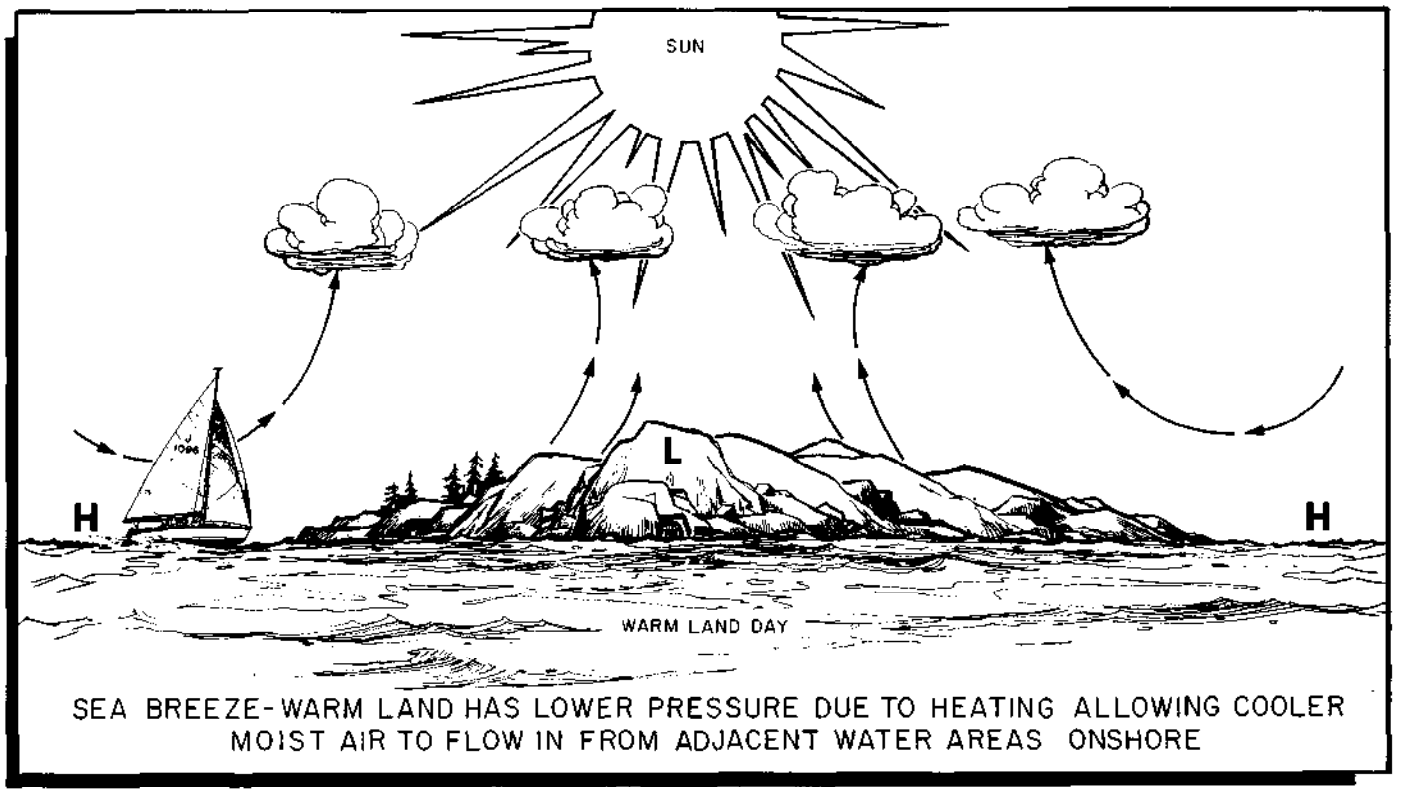

Figure 9-4. Air Transfer (Daytime).

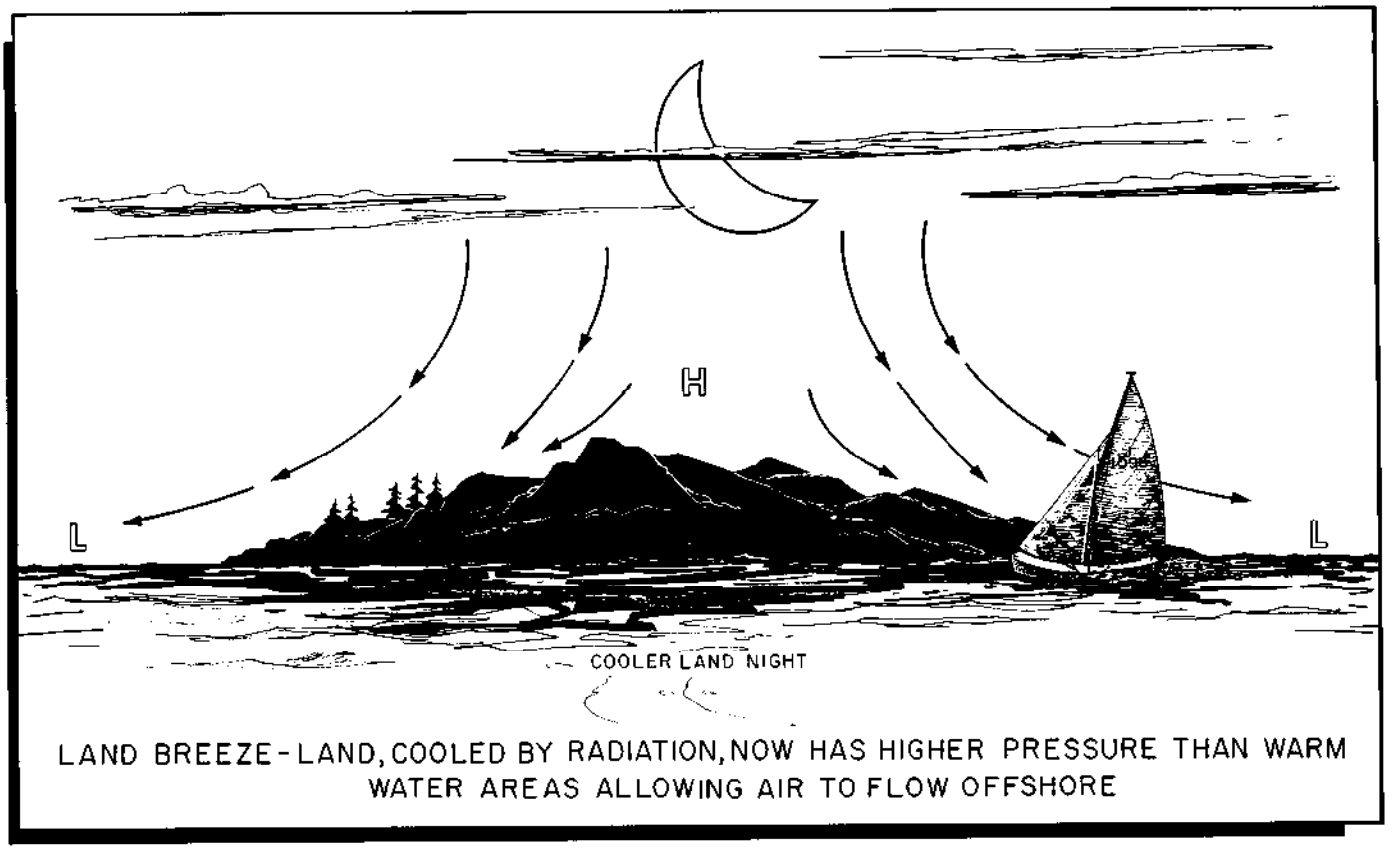

Figure 9-5. Air Transfer (Nighttime).

from west to east. The prevailing westerlies move across most of the United States and Canada, and are divided from the trade winds in a region called the Horse Latitudes. The air in the Horse Latitudes blows downward to fill the space which was left between the prevailing westerlies and the trade winds. The winds are very light in the Horse Latitudes.

(3) The winds from the North and South Poles are known as the polar easterlies. Because the air is so cold, making it heavy, the air above the poles sinks downward. The air spreads out when it reaches the ground and moves toward the Equator. The weather in the region of the polar easterlies moves from east to west with the Coriolis effect making the winds seem to blow from the east. The polar front is the meeting place of the polar easterlies and the prevailing westerlies and is a cloudy, rainy region. Above the polar front is a band of west winds called the jet stream. The jet stream occurs about 5 to 7 miles above the ground. Its winds may exceed 200 miles per hour.

b. Pressure systems are highs or lows covering areas as big as 1 million square miles. Most pressure systems found in the United States and Canada develop along the polar front. There, the cold winds of the polar easterlies and the warmer winds of the prevailing westerlies move past one another and create swirling winds called eddies. These eddies are carried eastward across the United States and Canada by the prevailing westerlies. There are two kinds of eddies: cyclones and anticyclones (figure 9-8).

(1) Cyclones formed by eddies are not the same as the storms known as cyclones. The winds of the eddies that create cyclones swirl inward toward a center of low pressure. A low-pressure system is formed by the cyclone and its low-pressure region. Because of the rotation of the Earth, cyclones that build north of the Equator blow in a counterclockwise direction. Cyclones that form south of the Equator move in a clockwise direction. Cyclones in North America generally approach on brisk winds, bringing cloudy skies and usually rain or snow.

(2) Anticyclones swirl outward around a center of high pressure, forming a high-pressure system. Anticyclones move in a clockwise direction north of the Equator and counterclockwise south of the Equator. Anticyclones come after cyclones, bringing dry, clearing weather and light winds.

c. Airmasses depend largely on the temperature and moisture of the areas in which they originate. Airmasses may cover 5 million square miles. As they move away from their source regions and pass over land and sea, the airmasses are constantly being modified through

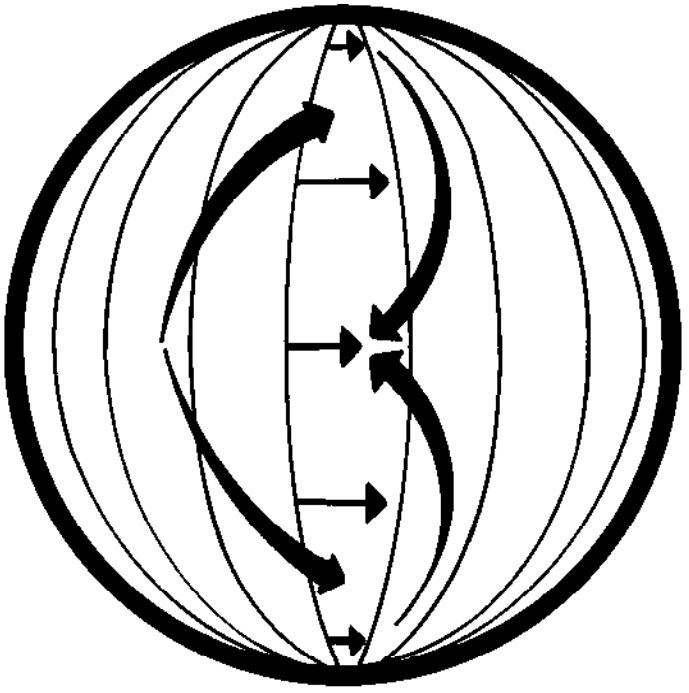

Figure 9-6. Coriolis Effect.

heating or cooling from below, lifting or subsiding, absorbing or losing moisture. In general, however, they retain some of their original characteristics and can be recognized and identified.

(1) There are four major types of airmasses:

(a) Continental polar - cold and dry.

(b) Continental tropical - hot and dry.

(c) Maritime polar - cool and moist.

(d) Maritime tropical - warm and moist.

(2) In North America, the continental polar airmass over northern Canada blows cold, dry air into southern Canada and the United States. Maritime polar airmasses off the northeast and northwest coasts of North America bring cool, damp weather to the continent. Maritime tropical airmasses from the southeast and southwest coasts bring warm, muggy weather. The polar airmasses are strongest in the winter, and the tropical airmasses are strongest in the summer. During the winter, a cold arctic airmass from the North Pole also influences the weather of North America. A continental tropical airmass forms over the southwest United States during the warm months but disappears in the winter.

d. When two different airmasses meet, they do not ordinarily mix (unless their temperatures, pressures, and relative humidities happen to be very similar). Instead, they set up boundaries called frontal zones, or "fronts." The colder airmass moves under the warmer airmass in the form of a wedge. If the boundary is not moving, it is termed a stationary front. Usually, however, the boundary moves along the Earth's surface, and as one airmass withdraws from a given area, it is replaced by another airmass. This action creates a moving front. If warmer air is replacing colder air, the front is called "warm;" if colder air is replacing warmer air, the front is called "cold." Most changes in the weather occur along fronts. The movement of fronts depends on the formation of pressure systems. Cyclones push fronts along at speeds of 20 to 30 miles per hour. Anticyclones blow into an area after a front has passed.

(1) When a warm front moves forward, the warm air slides up over the wedge of colder air lying ahead of it (figure 9-9). This warm air usually has high humidity. As this warm air is lifted, its temperature is lowered. As the lifting process continues, condensation occurs, low nimbostratus and stratus clouds form from which rain develops. The rain falls through the cooler air below, increasing its moisture content. Any reduction of temperature in the colder air, which might be caused by upslope motion or cooling of the ground after sunset, may result in extensive fog. As the warm air progresses up the slope, with constantly falling temperature, clouds appear at increasing heights in the form of altostratus and cirrostratus, if the warm air is stable. If the warm air is unstable, cumulonimbus clouds and altocumulus clouds will form and frequently produce thunderstorms. Finally, the air is forced up near the stratosphere and in the freezing temperatures at that level, the condensation appears as thin wisps of cirrus clouds. The upslope movement is very gradual, rising about 1,000 feet every 20 miles. Thus, the cirrus clouds, forming at perhaps 25,000 feet altitude, may appear as far as 500 miles in advance of the point on the ground which marks the position of the front. Warm fronts produce more gradual changes in the weather than do cold fronts. The changes depend chiefly on the humidity of the advancing warm airmass. If the air is dry, cirrus clouds may form and there will be little or no precipitation. If the air is humid, light, steady rain or snow may fall for several days. Warm fronts usually have light winds. The passing of a warm front brings a sharp rise in temperature, clearing skies, and an increase in humidity.

(2) When the cold front moves forward, it acts like a snowplow, sliding under the warmer air and tossing it aloft. This causes sudden changes in the weather. In fast-moving cold fronts, friction retards the front near the ground, which brings about a steeper frontal surface. This steep frontal surface results in a narrower band of weather concentrated along the forward edge of the front. If the warm air is stable, an overcast sky may occur for some distance ahead of the front, accompanied by general rain. If the warm air is conditionally unstable, scattered thunderstorms and showers may form in the warm air. In some cases, an almost continuous line of thunderstorms is formed and called a "squall line." Behind the fast moving cold front there is usually rapid clearing, with gusty and turbulent surface winds and colder temperatures. The slope of a cold front is much steeper than that of a warm front and the progress

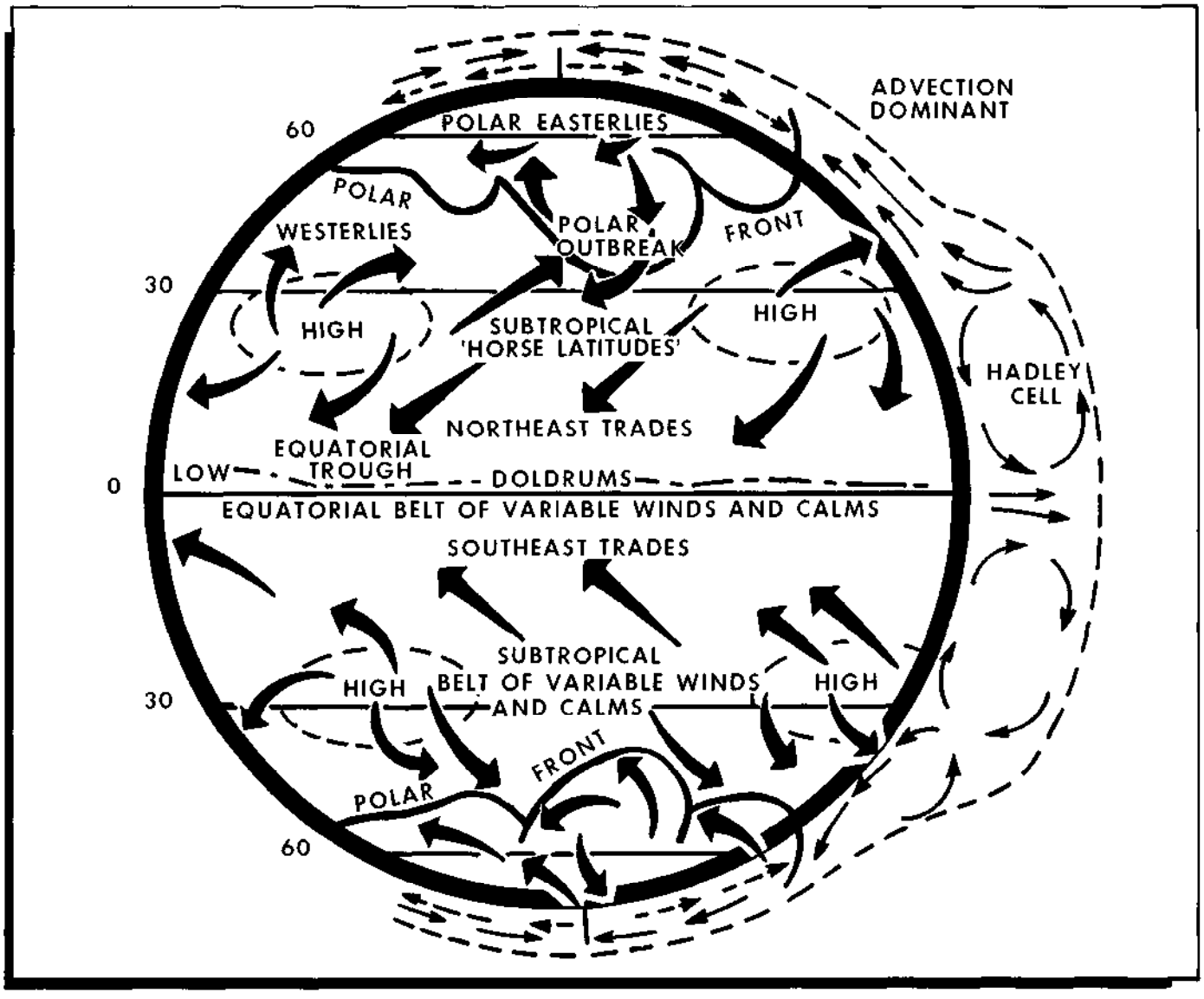

Figure 9-7. Atmosphere Circulation.

is generally more rapid—usually from 20 to 35 miles an hour, although in extreme cases, cold fronts have been known to move at 60 miles per hour (figure 9-10). Weather activity is more violent and usually takes place directly at the front instead of in advance of the front. However, especially in late afternoon during the warm season, a squall line will frequently develop as much as 50 to 200 miles in advance of the actual cold front. Whereas warm front dangers lie in low ceilings and visibilities, cold front dangers lie chiefly in sudden storms with high and gusty winds. Unlike the warm front, the cold front arrives almost unannounced, makes a complete change in the weather within the space of a few hours, and passes on. The squall line is ordinarily quite narrow—50 to 100 miles in width—but is likely to extend for hundreds of miles in length, frequently lying across the entire United States in a line running from northeast to southwest. Altostratus clouds sometimes form slightly ahead of the front, but these are seldom more than 100 miles in advance. After the front has passed, the weather clears rapidly with cooler, drier air.

(3) One other form of front with which the survivor should become familiar is the "occluded front" (figure 9-11). Cold fronts travel about twice as fast as warm fronts. As a result, cold fronts often catch up to warm fronts. When a cold front reaches a warm front, an occluded front develops. Meteorologists subdivide occlusions into two types: cold-front occlusions and warm-front occlusions. In a cold-front occlusion, the air behind the cold front is colder than the air ahead of the warm front. The weather of a cold-front occlusion re-

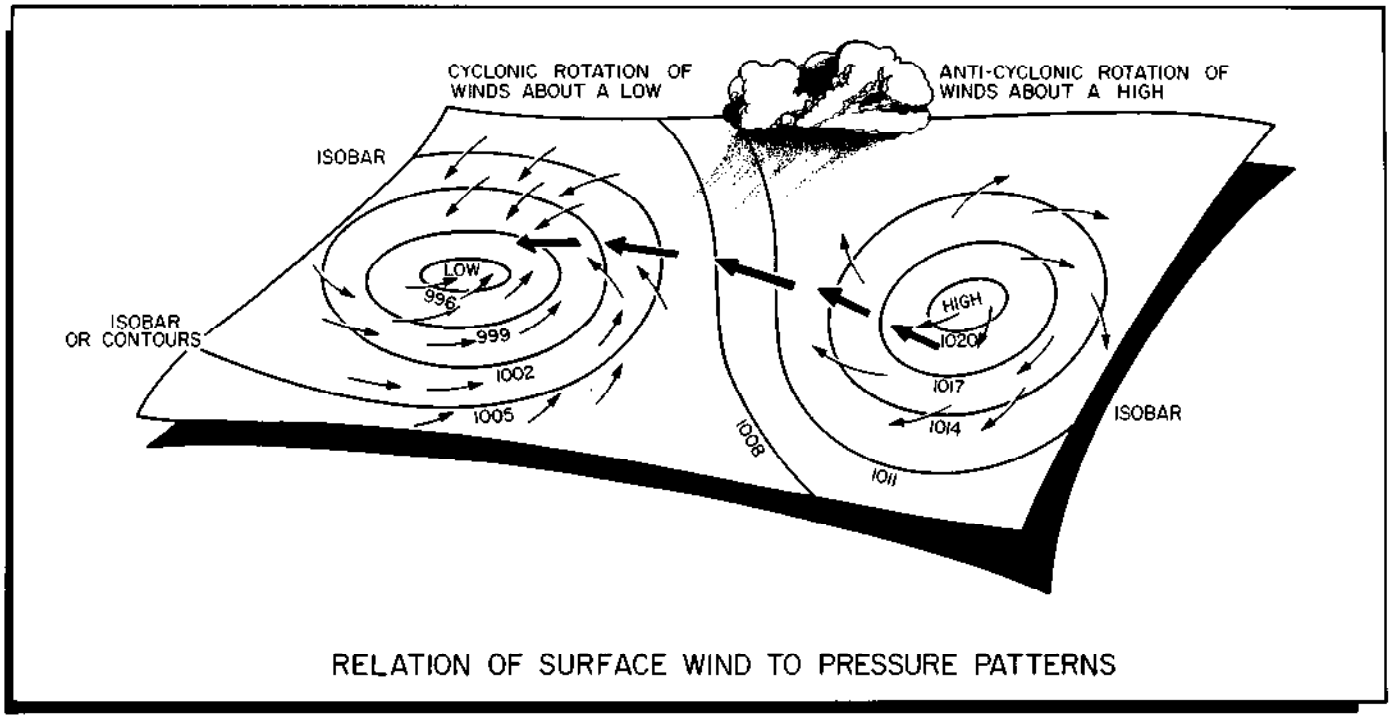

Figure 9-8. Cyclonic and Anticyclonic Rotation.

sembles that of a cold front. When the air behind the cold front is warmer than the air ahead of the warm front, it is known as a warm-front occlusion. Warm-front occlusion weather is similar to a warm front. These fronts produce milder weather than do cold or warm fronts.

(4) Stationary fronts are another type of front which occurs when airmasses meet but move very slow-

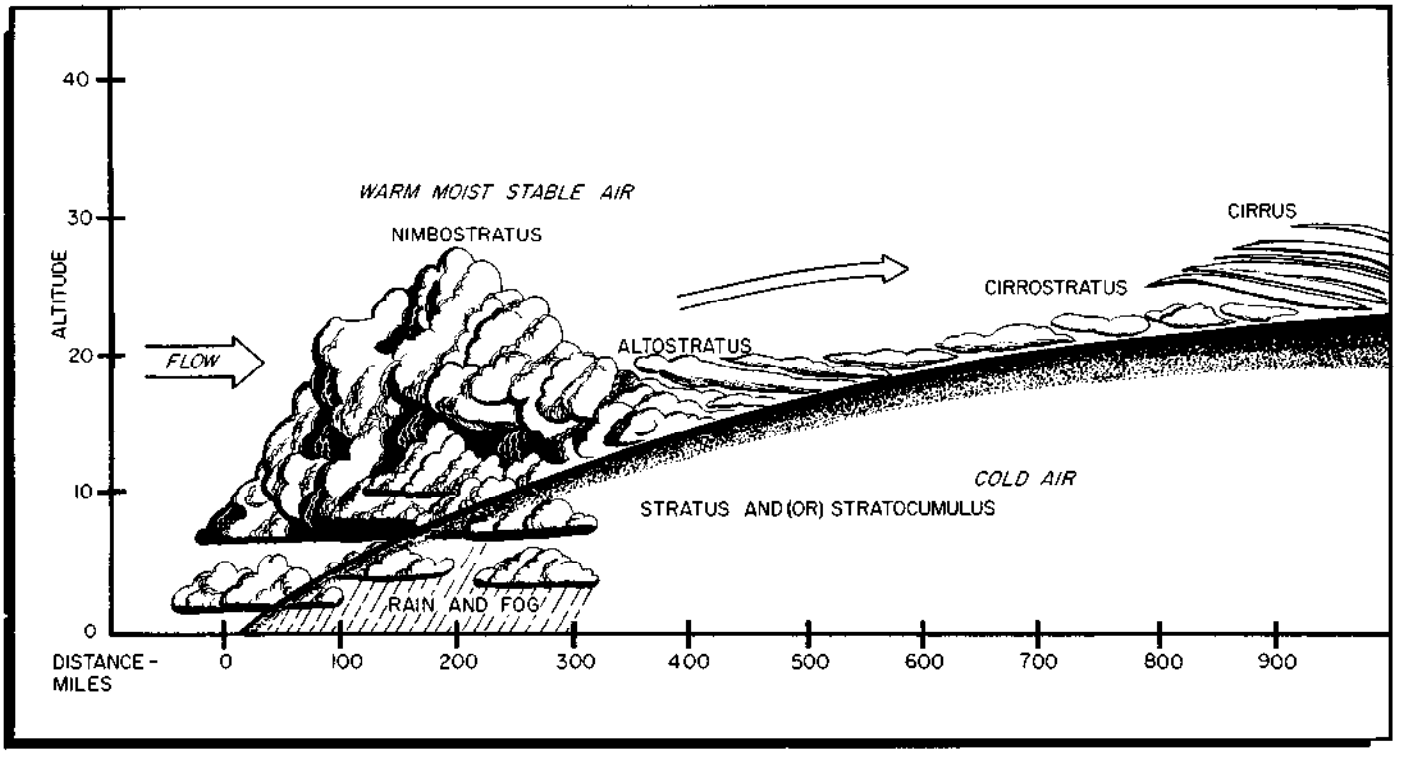

Figure 9-9. Stable Air Warm Front.

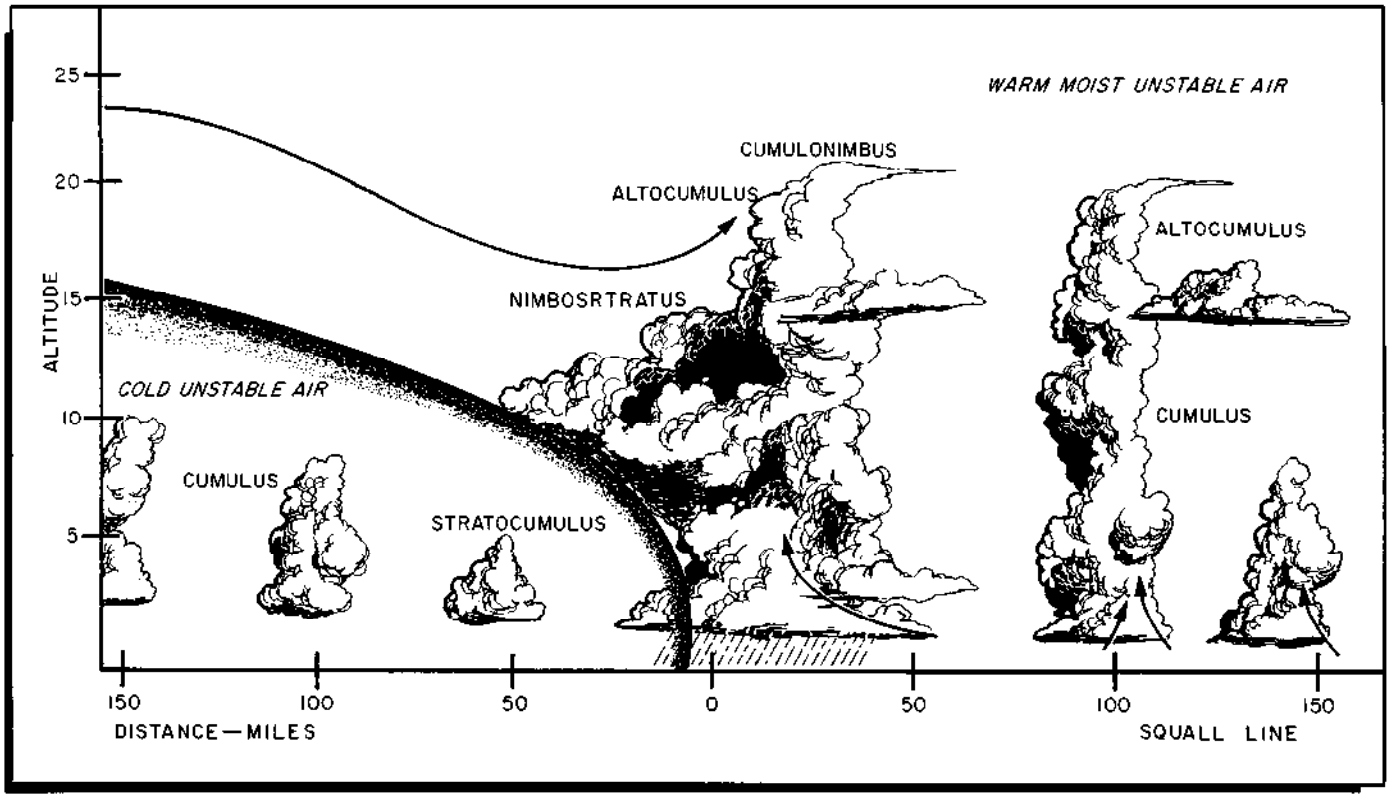

Figure 9-10. Fast-Moving Cold Front.

ly. It may remain over an area for several days bringing moderate weather.

9-5. Storms. The four main types of violent weather a person should be familiar with are thunderstorms, winter storms, tornadoes, and hurricanes.

a. Thunderstorms are the most frequent kinds of storms. As many as 50,000 thunderstorms occur throughout the world each day. Under some conditions, the rapid lifting of moist, warm air results in thunderstorms and dramatic cloud formations (figure 9-12). They develop from tall, puffy cumulonimbus clouds. Clouds may tower 5 to 10 miles high during hot, humid days. The temperatures inside the clouds are well below freezing. The air currents inside the clouds move up and down as fast as 5,000 feet per minute. Heavy rain is common because water vapor condenses rapidly in the air. Lightning and thunder occur during the life of a thunderstorm. When the sound of thunder is heard, a survivor should seek shelter immediately. Lightning causes more fatalities than any other type of weather phenomenon. In the United States alone more than 200 lightning deaths occur each year. Another reason for seeking shelter immediately is to escape the hail which sometimes accompanies the thunderstorm. Hail, which can grow as large as baseballs, is most noted for damaging crops, but a powerful storm can bring injuries, even fatalities, to survivors if shelter is not available.

b. Tornadoes are the most violent form of thunderstorms. Under certain conditions, violent thunderstorms will generate winds swirling in a funnel shape with rotational speeds of up to 400 miles per hour which extends out of the bottom of the thunderstorm. When this funnel-shaped cloud touches the surface, it can cause major destruction. The path of a tornado is narrow, usually not more than a couple of hundred yards wide. Tornadoes form in advance of a cold front and are usually accompanied by heavy rain and thun- southern areas of the United States.

c. Winter storms include ice storms and blizzards. An ice storm may occur when the temperature is just below freezing. During this storm, precipitation falls as rain but freezes on contact with the ground. A coating of ice forms on the ground and makes it very hazardous to the traveler. Snowstorms with high winds and low temperatures are called blizzards. The wind blows at 35 miles per hour or more during a blizzard, and the temperture may be 10°F or less. Blowing snow makes it impossible to travel because of low visibility and drifting.

d. A hurricane or typhoon, the most feared of storms, has a far more widespread pattern than a tornado. The storm forms near the Equator over the oceans and is a

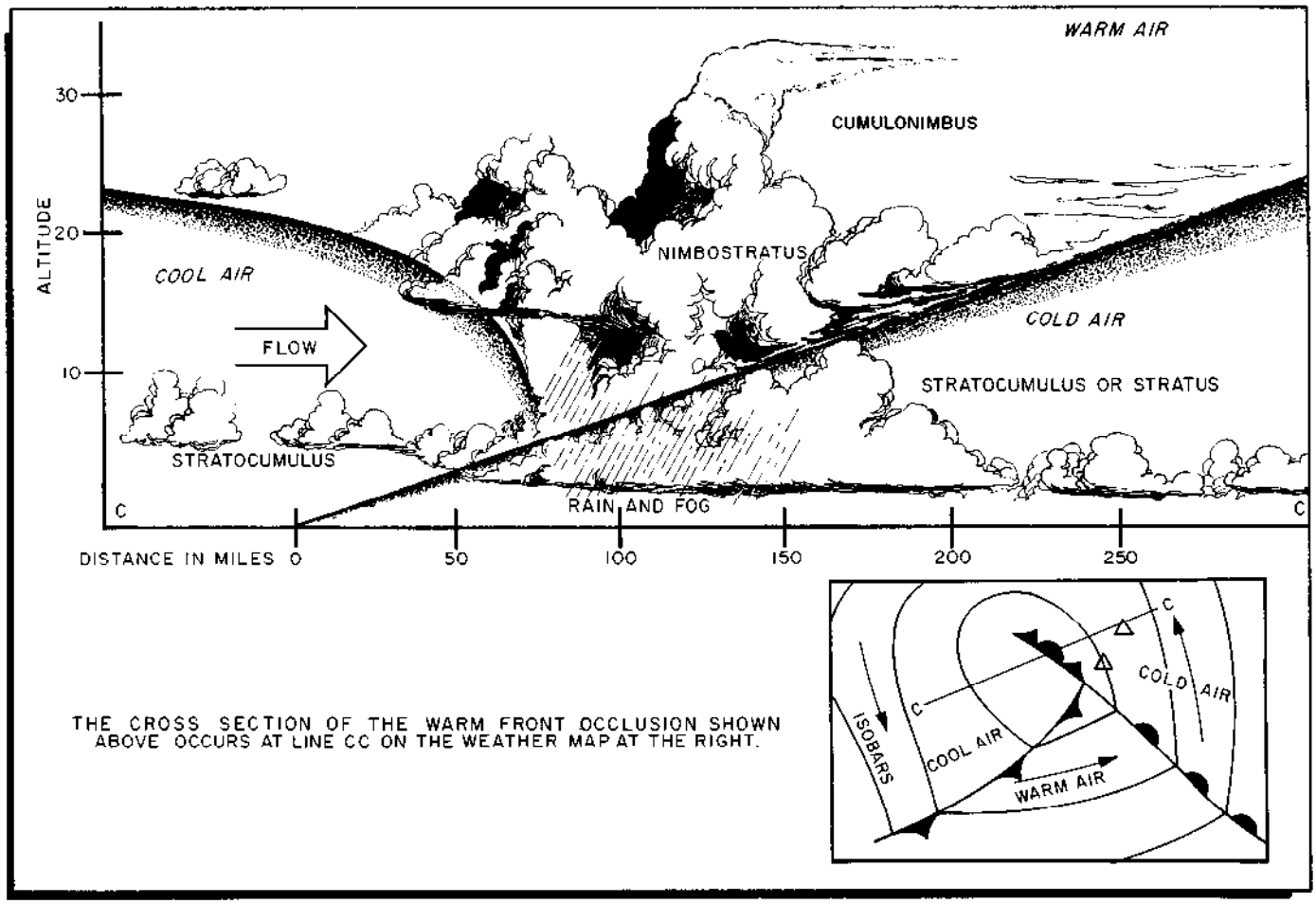

Figure 9-11. A Warm-Front Occlusion.

large low-pressure area, about 500 miles in diameter. Winds swirl around the center (eye) of the storm at speeds over 75 miles per hour and can reach 190 miles per hour. Hurricanes break up over land and often bring destructive winds and floods. Thunderstorms often form within hurricanes and can produce tornadoes. Most hurricanes occurring in the United States sweep over the West Indies and strike the southeastern coast of the country. An early indication of a hurricane is a wind from an unusual direction, like the replacement of the normal flow of the trade winds from an easterly direction. The arrival of high waves and swells at sea coming from an unusual direction may also give some warning. The high waves and swells are moving faster than the storm and may give several days warning.

9-6. Weather Forecasting. Weather forecasting enables survivors to make plans based on probable changes in the weather. Forecasts help survivors decide what clothes to wear and type of shelter to build. During an evasion situation, it may help survivors determine when to travel. While accurate weather prediction or forecasting normally requires special instruments, an awareness of changing weather patterns and attention to existing conditions can help a survivor or evader prepare for and, when appropriate, use changing weather conditions to enhance their survivability. The following are some elementary weather indicators which could help predict the weather and help save lives.

a. Clouds which move higher are good signs of fair weather. Lower clouds indicate an increase in humidity, which in all probability means precipitation (figure 9-13).

b. The Moon, Sun, and stars are all weather indicators. A ring around the Moon or Sun means rain (figure 9-14). The ring is created when tiny ice particles in fine cirrus clouds scatter the light of the Moon and the Sun in different directions. When stars appear to twinkle, it indicates that strong winds are not far off, and will become strong surface winds within a few hours. Also, a large number of stars in the heavens show clear visibility with a good chance of frost or dew.

c. "Low-hanging" clouds over mountains mean a weather change (figure 9-13). If they get larger during

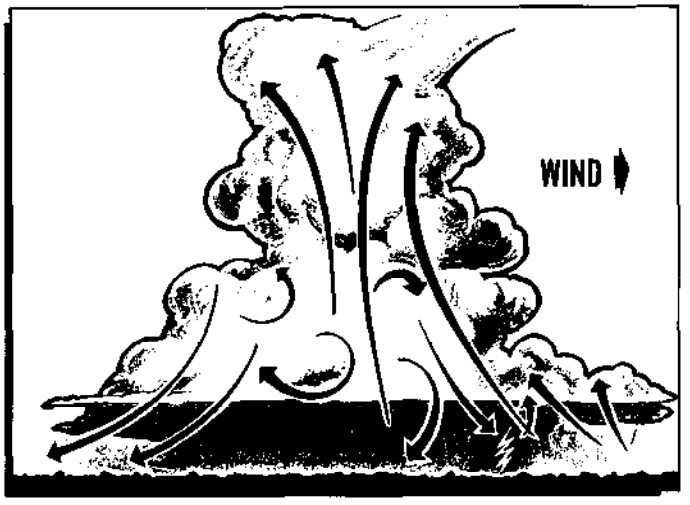

Figure 9-12. Thunderstorms.

the daytime, bad weather will arrive shortly. Diminishing clouds mean dry weather is on its way. Storms are often preceded by high thin cirrus clouds arriving from the west. When these thicken and are obscured by lower clouds, the chances increase for the arrival of rain or snow.

d. The old saying "red skies at night, sailor's delight; red skies at morning, sailors take warning," has validity. The morning Sun turning the eastern sky crimson often signals the arrival of stormy weather. As the storm moves east, clouds may turn red as a clearing western sky opens for the setting Sun.

e. "The farther the sight, the nearer the rain," is a seaman's chant. When bad weather is near, the air pressure goes down and the atmosphere becomes clearer. High atmospheric pressure with stable and dusty air means fair weather.

f. A cold front arriving in the mountains during the summer usually means several hours of rain and thunderstorms. However, the passing of a cold front means several days of clear, dry weather.

g. A morning rainbow is often followed by a squall. An afternoon rainbow means unsettled weather, while an evening rainbow marks a passing storm. A faint rainbow around the Sun precedes colder weather.

h. Stormy weather will probably follow within hours when flowers seem to be much more fragrant.

i. People say "when sounds are clear, rain is near," because sound travels farther before storms.

j. Even birds can help predict the weather. Water birds fly low across the water when a storm is approaching. Birds will huddle close together before a storm.

k. The flowers of many plants, like the dandelion, will close as humidity increases and rain is on the horizon.

l. As humidity increases, the rocks in high mountain areas will "sweat" and provide an indication of forthcoming rain.

m. Lightning can tell survivors something by noting the color and compass direction. If the lightning looks white when seen through clear air and is located in the west or northwest, survivors would know the storm is headed toward them. Storms to the south or east will normally pass to the east. Red or colored lightning is seen at a distance in storms that will pass to the north or south.

n. Smoke, rising from a fire then sinking low to the ground, can indicate that a storm is approaching.

9-7. Summary. Even with the modern equipment available, forecasting tomorrow's weather is often difficult. This chapter provided background information and tips to use to teach survivors to predict weather. By understanding the basic characteristics and actions of weather, the survivor can better prepare for its effects.

Figure 9-13. Low Clouds.

Figure 9-14. Rings Around Moon or Sun.

Chapter 10

GEOGRAPHIC PRINCIPLES

10-1. Introduction. Geographic principles bring together the three major components of the environment (terrain, life forms, and climate) for the purpose of understanding the relationship between survivors and the physical environment around them. The more survivors know about these environmental components, the better they can help themselves in a survival situation. This chapter provides a brief introduction to these complex topics.

10-2. Components of the Environment:

a. Terrain is defined as a geographic area consisting of land and its features. The landmass of the Earth is covered with a variety of topography, including mountains, valleys, plateaus, and plains.

(1) The mountains vary greatly in size, structure, and steepness of slopes. For example, there is as much contrast between the large volcanic Cascade Mountains and those of the Rocky Mountains as there is between the Rocky Mountains and the Appalachian Mountains. Most major mountain systems will have corresponding foothills (figure 10-1).

(2) With two exceptions, valleys are formed as mountains are pushed up. The exceptions are massive gorges formed by glacial action and valleys carved out by wind and water erosion (figure 10-2).

(3) Plateaus are elevated and comparatively large, level expanses of land. Throughout the southwest, examples of the typical plateau can be seen. These plateaus were formed when a volcano deposited either lava or ash over a softer sedimentary area. Through years of erosion, the volcanic "cap" broke loose in places and allowed the softer ground to be carried away. This type of plateau is the least common; however, it is the largest. The Columbia Plateau of Washington State is one example which covers 200,000 square miles (figure 10-3).

(4) The water forms of the Earth include oceans, seas, lakes, rivers, streams, ponds, and ice.

(a) Oceans comprise approximately 70 percent of the Earth's surface. The major oceans include the Pacific, Atlantic, Indian, and Arctic. Oceans have an enormous effect on land, not only in their physical contact but in their effect on weather. In most cases, lakes today are descended from much larger lakes or seas.

(b) Ice covers 10 percent of the Earth's surface. This permanent ice is found in two forms—pack ice and glaciers. Pack ice (normally 7 to 15 feet thick) is frozen sea water and may be as much as 150-feet thick. Those pieces which break off form ice islands. The two permanent icepacks on Earth are found near the North and South Poles—Arctic and Antarctic. The polar regions, which are thousands of feet thick, partially, but never completely, thaw. An icecap is a combination of pack ice and ice sheets. The term is usually applied to an ice plate limited to high mountain and plateau areas. During glacial periods, an icecap will spread over the surrounding lowlands (figure 10-4).

b. Life forms can best be described in terms of vegetation and animal life, with special emphasis on humans (which are covered later).

(1) There are hundreds of plant species on Earth. An in-depth study is obviously impossible. To understand the plant kingdom better, it is important to understand basic plant functions and adaptations they have made to exist in diverse environments. Vegetation will be categorized into either trees or plants.

(a) Of all the variety in species and types, trees can be divided simply into two types: coniferous or deciduous. Conifers are generally considered to be cone-bearing, evergreen trees. Some examples of conifers are pine, fir, and spruce. Deciduous trees are those which lose their leaves in winter and are generally considered as "hardwood." Some examples are maple, aspen, oak, and alder.

(b) For discussion, we will divide plants into two categories: annuals and perennials. Annuals complete their life cycle in 1 year. They produce many seeds and regenerate from seed. Climatic conditions may not be conducive for growth the following year, so seeds may remain dormant for many years. A classic example is the 1977 desert bloom in Death Valley. Plants bloomed for the first time in 80 years. Perennials are plants which last year after year without regeneration from seed.

(2) As with plants, the discussion of animal life has to be limited. Animals will be classified as either warmblooded or coldblooded. Using this division as a basic, it will be easier to describe animal adaptations to extreme climatic conditions. Warmblooded animals are generally recognized as cold-adapted animals and include all birds and mammals. Obviously, humans are a part of this classification because they are cold-adapted. Coldblooded animals gain heat from the environment. These are animals adapted for life in warm or moderate climates (lizards, snakes, etc.)

c. Climate can be described as an average condition of the weather at any given place. However, this description must be expanded to include the seasonal variations and extremes as well as the averages in terms of the climatic elements. In some areas, the climate is so domineering that the corresponding biome is named either in part or as a whole by the climate. Examples are the cloud forests and rain forests. The climate can only be described in terms of its various elements—temperature, moisture, and wind.

©National Geographic Society, Photo by Luis Marden

©National Geographic Society, Photo by George F. Mobley

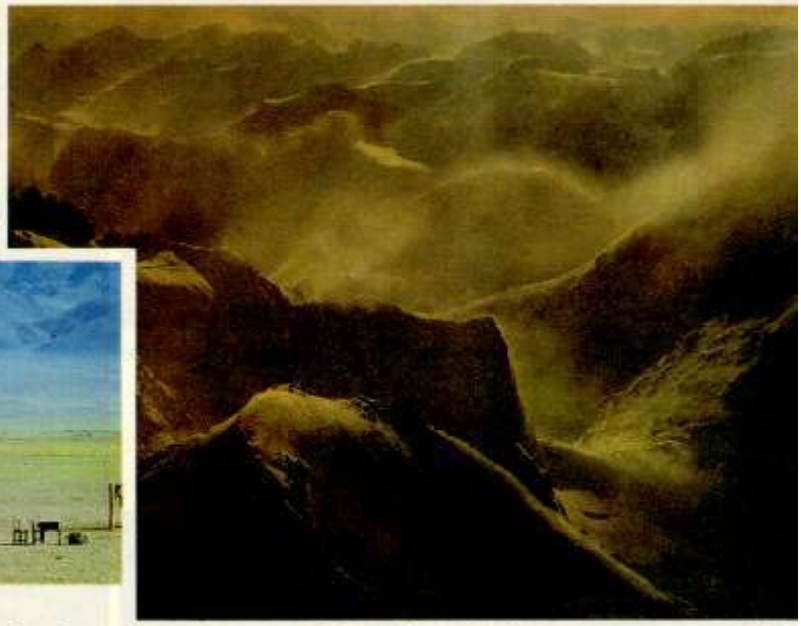

©National Geographic Society, Photo by George F. Mobley

©National Geographic Society, Photo by Dean Conger

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

Figure 10-1. Composite of Mountains.

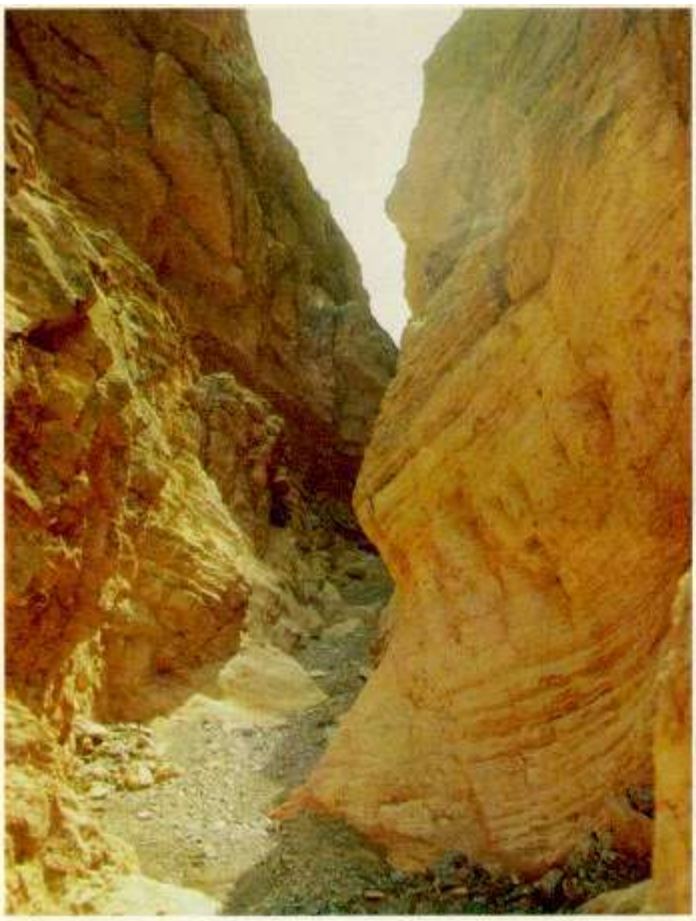

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

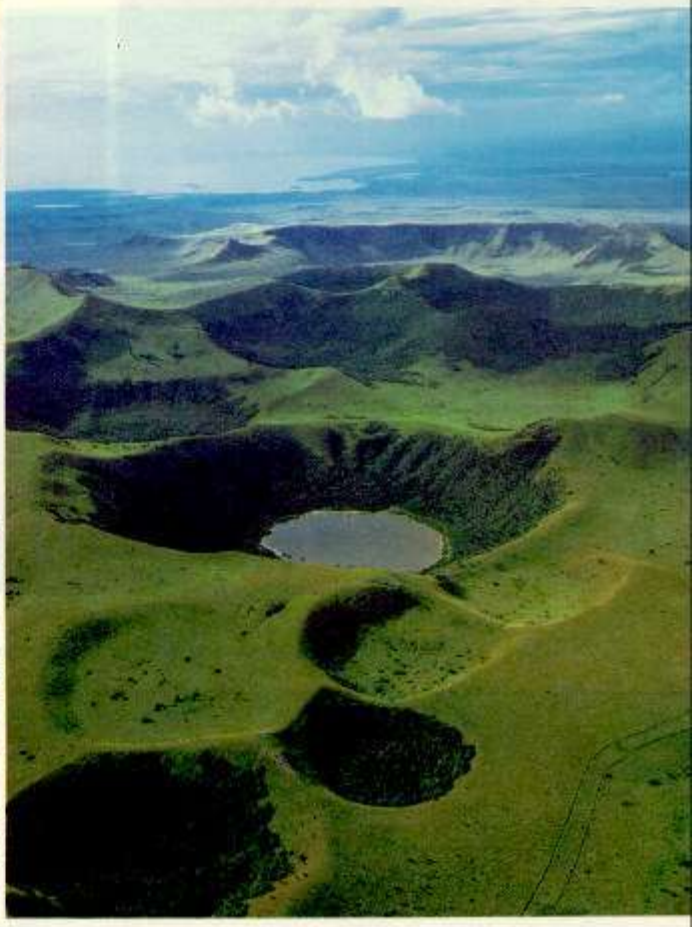

©National Geographic Society, Photo by George F. Mobley

Figure 10-2. Gorges and Valleys.

(1) The atmosphere gains only about 20 percent of its temperature from the direct rays of the Sun. Most of the atmospheric heat gain comes from the Earth radiating that heat (energy) back into the atmosphere and being trapped. This is known as the greenhouse effect.

© Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

Figure 10-3. Plateau.

(2) Thinking of the atmosphere as a greenhouse, it is easier to understand the relationship water has in this "closed system." As water evaporates, the amount of water vapor the atmosphere can absorb depends solely upon temperature. The dew point is achieved when the amount of water vapor in the air equals the maximum volume the air will hold at a given air temperature. Lowering of air temperature creates condensation. Condensation appears in the form of clouds, fog, and dew. Any additional temperature reduction results in precipitation, such as rain. If the temperature of the dew point is below freezing, precipitation may appear in the form of hail or snow.

(3) Variation in air pressure is the primary cause of wind. When air is heated, it creates an area of lower pressure. As air cools, the pressure increases. Air movement occurs as the pressure tries to equalize, thus creating wind. Because wind is also a control of climate, people need to know why and how it affects climate. Let's look at wind in two aspects: localized wind (low altitude) and upper-air wind (high altitude). Localized wind is formed at low altitude, occurring due to dynamic topographical features and fluctuating air temperature and pressure. High-altitude winds surrounding the Earth are bands of stable high- and low-pressure areas (cells). Predictable winds move off these cells which are referred to as jetstreams. These high-altitude winds control weather.

10-3. Effects of Climate on Terrain. The major effect climate has on terrain is erosion. Erosion can occur directly from heavy precipitation or indirectly by the accumulation of snow on snowpacks and glaciers. Wind and temperature both have erosion potential.

a. Heavy precipitation or melting water from icepacks and glaciers can create deep ravines by cutting into mountainous areas. Broad flood basins along major rivers can also aid in the development of river deltas in lakes, oceans, and deep fjords. The action of glaciers throughout the years has carved out deep, broad valleys with steep valley walls (figure 10-5).

b. The effects of wind erosion are greatest in barren, dry areas. The Great Arches National Park has some of the most dramatic examples of the effect of wind erosion. This type of erosion is caused by the wind driving sand and dust particles against an exposed rock or soil

Figure 10-4. Pack Ice.

surface, causing it to be worn away by the impact of the particles in an abrasive action (figure 10-6). Another form of wind erosion involves the movement of loose particles lying upon the ground surface which may be lifted into the air or rolled along the ground. Dry river beds, beaches, areas of recently formed glacial deposits, and dry areas of sandy or rocky ground are highly susceptible to this type of erosion. Sand dunes are attributed to this phenomenon (figure 10-7).

c. Frost action will have a weathering or eroding effect on rock land formations and ground surfaces. The frost action is the repeated growth and melting of ice crystals in the pore space or fractures of soil and rock. The tremendous force of growing ice crystals can exert a pressure great enough to pry apart rock. Many scree and talus slopes are caused by this action. Where soil water freezes, it tends to form ice layers parallel with the ground surface, heaving the soil upward unevenly. The peat moss mounds of the tundra are an example of this action. The net effect of frost action will be dependent on the amount of surface moisture.

10-4. Effects of Terrain on Climate. The effect of terrain on climate is not nearly as subtle as the effect of climate on terrain. Three major factors exist which must be considered when studying the effects of terrain on weather.

a. Moisture for most major weather systems comes from the evaporation of the oceans of the world. The temperature, location, and flow of ocean currents, combined with the prevailing winds will affect how much water will evaporate into the atmosphere. The warmer the ocean and corresponding current, the greater the rate of evaporation. Since the currents are deflected by landmasses, many warm currents flow parallel to major continents. When this moisture is blown inland by the prevailing winds, the net effect is the creation of a wet maritime climate, such as that found along the west coasts of Canada, Washington, Oregon, and Central Europe. If the temperature of the ocean and currents is cold, very little moisture will be yielded to the atmosphere. Examples of this occur along the Pacific coastline of Peru and Chile and along the Atlantic coastline of Angola and Southwestern Africa.

b. The interior of large continent masses are dry because of the distance which isolates them from the effects of maritime climates. The large continents of the Northern Hemisphere create dry, high-pressure cells which isolate the interiors from the lower pressure moist air cells and keep them from having much effect. The climate is referred to as the Continental Climate. The concept will be explained in the next chapter—Environmental Characteristics.

c. Mountains serve as moisture barriers, separating the maritime influenced climates from the continental influenced climates. The barrier effect of mountains on weather will be dependent on the height, length, and width of the range and the severity of the weather

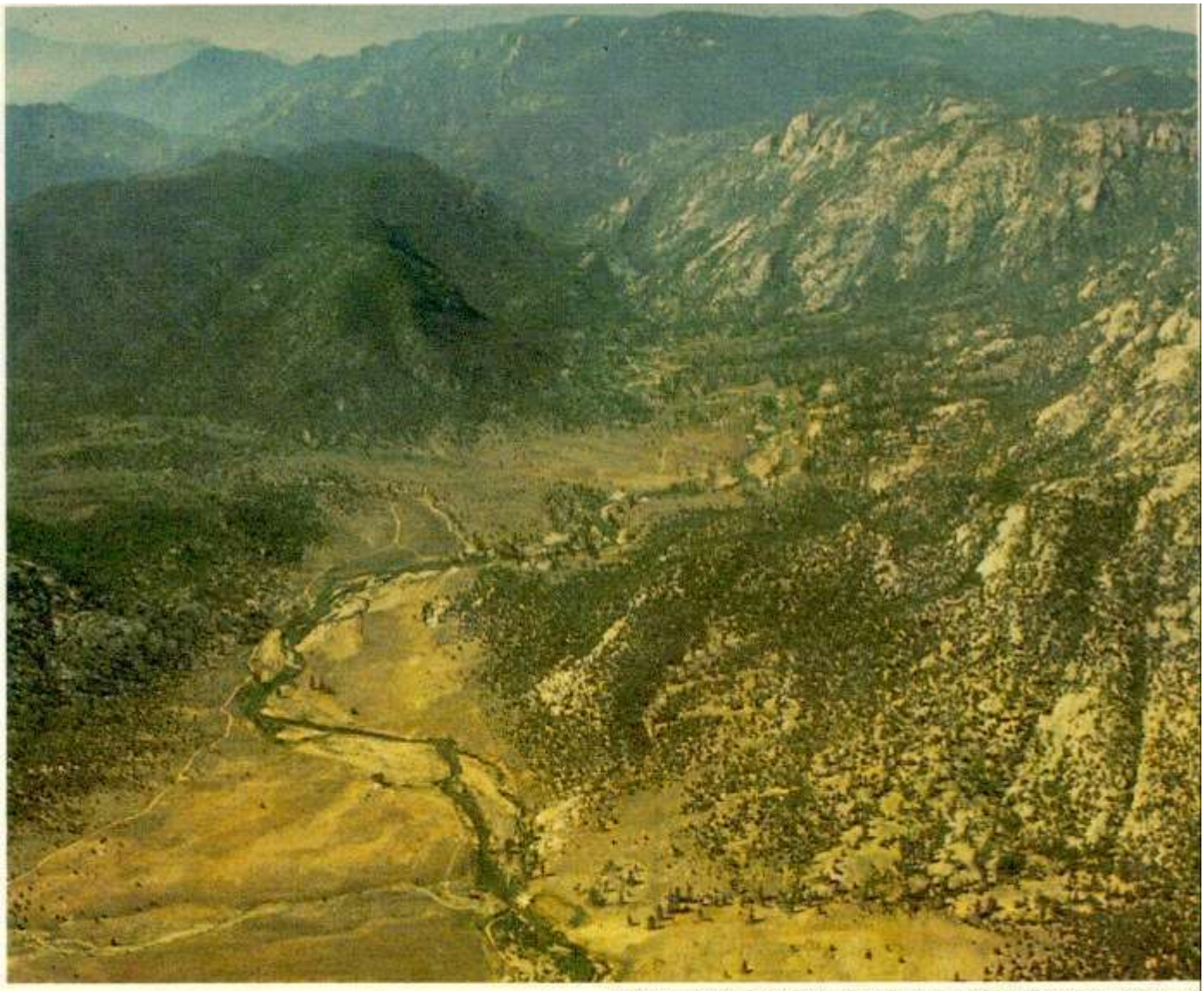

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

Figure 10-5. Valleys.

fronts. In many cases, a lack of precipitation will extend for several hundred miles beyond the mountains. An example of this phenomenon occurs in the western states. The Cascade and Sierra Mountains block a great deal of Pacific Ocean moisture from the the inland deserts of Washington, Oregon, and Nevada. The Rocky Mountains further block most of the moisture which is left in the atmosphere. Only the high cirrus clouds escape the barrier effect of these mountains. Another example can be seen in Asia. The Himalayan Mountains serve as a very effective barrier, blocking the Asiatic monsoon from central interior Asia, which helps create the Gobi Desert.

10-5. Effects of Climate and Terrain on Life Forms:

a. Since plants require water and light, climate will greatly affect the type and number of plants in an area.

(1) In areas with a great deal of rainfall, plants will be plentiful. In these areas, plants must compete for available sunlight. In areas where the primary vegetation has been knocked down (by clear cutting, landslides, or along flood basins of rivers), a thick secondary growth will occur. In time, the secondary growth, if undisturbed, will become a climax forest. Some of the trees in these areas may grow to 300 feet. Because of the shade, vines and shade-tolerant perennials may sparsely cover the ground.

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

Figure 10-6. Erosion.

(2) In contrast, in areas where the amount of rainfall is limited, the plants must compete for the available water. The number of plants will also be sparse. Due to the harsh climatic and soil conditions, plant life is typically hardy. Through millions of years of evolution, plants have developed the following survival characteristics:

(a) Production of many seeds which germinate when water does come.

(b) Shallow root systems gather water quickly when they can.

(c) Ability to store water (cacti and other succulents).

(d) Rough, textured leaves (transpiration).

(e) Production of toxins (kill off competing plants).

b. Vegetation is also affected by the terrain. In mountainous regions, the clouds begin to lose moisture as they pass over the tops of the mountains. The result is more water is available for the growth of vegetation. However, with any increase in elevation, the temperature becomes colder. This exposure to colder temperature has a drastic effect on plant life.

c. Generally, animal life is mostly dependent on two factors: water and vegetation. The greater the rainfall, the greater the number of animals. Conversely, the drier an area, the less vegetation there will be to support animals. The location of small animals is determined by the secondary growth and gound cover, used for protection.

d. Temperature also affects the habits of animals. For example, animals may burrow to protect themselves from extreme heat or extreme cold or will be more active at night in hot, dry regions. Animals also respond physiologically to temperature extremes. During extreme cold, some species of mammals enter into a state of winter dormancy (hibernation). It is a special case of temperature regulation in which animals lower the setting of their "thermostats" to maintain lower than normal body temperatures in order to save energy while maintaining minimum body functions essential for survival. This is important since their normal food supply is not always available during the winter. During periods of excessive heat, some species of fish, reptiles, mammals, and amphibians will enter into a "summer sleep" called estivation. Estivation is a state in which the animal's body functions and activities are greatly reduced. Estivation and hibernation are not merely a result of temperature regulation but rather are methods by which the organisms survive unfavorable periods.

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

Figure 10-7. Sand Dunes.

Chapter 11

ENVIRONMENTAL CHARACTERISTICS

11-1. Introduction. Most survivors will not have a choice as to where they have to survive. The ease or difficulty in maintaining life and honor and returning are dependent on the types and extremes of the climate, terrain, and life forms in the immediate area. The Köppen-Geiger System of Climate Classification will be used in this chapter as the basis for organizing the discussion of environmental characteristics. This system has become the most widely used of climatic classifications for geographical purposes.

11-2. The Köppen-Geiger System. This system defines each climate according to fixed values of temperature and precipitation, computed according to the averages of the year or of individual months. A climate system based on these data has a great advantage in that the area covered by each subtype of climate can be outlined for large parts of the world. The five major climate groups are designated as: (See figure 11-1.)

a. Tropical Climates. Average temperature of each month is above 64.4°F. These climates have no winter season. Annual rainfall is large and exceeds annual evaporation.

b. Dry Climates. Potential evaporation exceeds precipitation on the average throughout the year. No water surplus; hence, no permanent streams originate in dry climate zones.

c. Warm Temperate Climates. Coldest month has an average temperature under 64.4°F, but above 26.6°F. The warm temperate climates thus have both a summer and winter season.

d. Snow Climates. Coldest month average temperature under 26.6°F. Average temperature of warmest month above 50°F.

e. Ice Climates. Average temperature of warmest month below 50°F. These climates have no true summer.

11-3. Tropical Climates:

a. Tropics. Some people think of the tropics as an enormous and forbidding tropical rain forest through which every step taken must be hacked out and where every inch of the way is crawling with danger. Actually, much of the tropics is not rain forest. What rain forest there is must be traveled with some labor and difficulty. The tropical area may be rain forest, mangrove or other swamps, open grassy plains, or semi-dry brushland. The tropical area may also have deserts or cold mountainous districts. There is in fact, a variety of tropical climates. Each region, while subject to the general climatic condition of its own zone, may show special modifications locally. Each general climate is a whole range of basic minor climates. In all their diversity, the climates of the tropics have the following in common:

(1) An almost constant length of day and night, a length that varies by no more than half an hour at the Equator to 1 hour at the limits of the tropics. The plant life thus has an evenly distributed period of daylight throughout the year.

(2) Temperature variation throughout the tropics is minimal—9°F to 18°F.

(3) There is no systematic pattern of major tropical landforms. There are high rugged mountains; such as the Andes of South America, karst formations as in Southeast Asia, plateaus like the Deccan of India, hilly lands like those which back the Republic of Guinea in Africa, and both large and small plains like the extensive one of the upper Amazon River or the restricted plain of the Irrawaddy River in Burma. The arrangement of all these landforms is part of the pattern of the larger landmasses, not of the tropics alone.

b. Vegetation:

(1) The jungles in South America, Asia, and Africa are more correctly called tropical rain forests. These forests form a belt around the entire globe, bisected somewhat equally by the Equator. However, the tropical rain forest belt is not a continuous one, even in any of the various regions in which it occurs. Usually it is broken by mountain ranges, plateaus, and even by small semi-desert areas, according to the irregular pattern of climate which regulates the actual distribution of rain forest.

(2) Some of the leading characteristics of the tropical rain forest common to those areas in South America, in Asia, and in Africa, are:

(a) Temperatures average close to 80°F for every month.

(b) Vegetation consists of three stories.

(c) High rainfall (80 inches or more) distributed fairly evenly throughout the year.

(d) Areas of occurrence lie between 23.5 North and 23.5 South Latitudes.

(e) Evergreen trees predominate; many of large girth up to 10 feet in diameter, with thick leathery leaves.

(f) Vines (lianas) and air plants (epiphytes) are abundant.

(g) Herbs, grasses, and bushes are rare in the understory.

(h) Uniformity.

(i) Tree bark thin, green, smooth, and usually lacking fissures.

(3) The majority of plants that grow in the forest of the rainy tropics are woody and of the dimensions of trees. Trees form the principal elements of the vegeta-

Figure 11-1. Köppen-Geiger System.

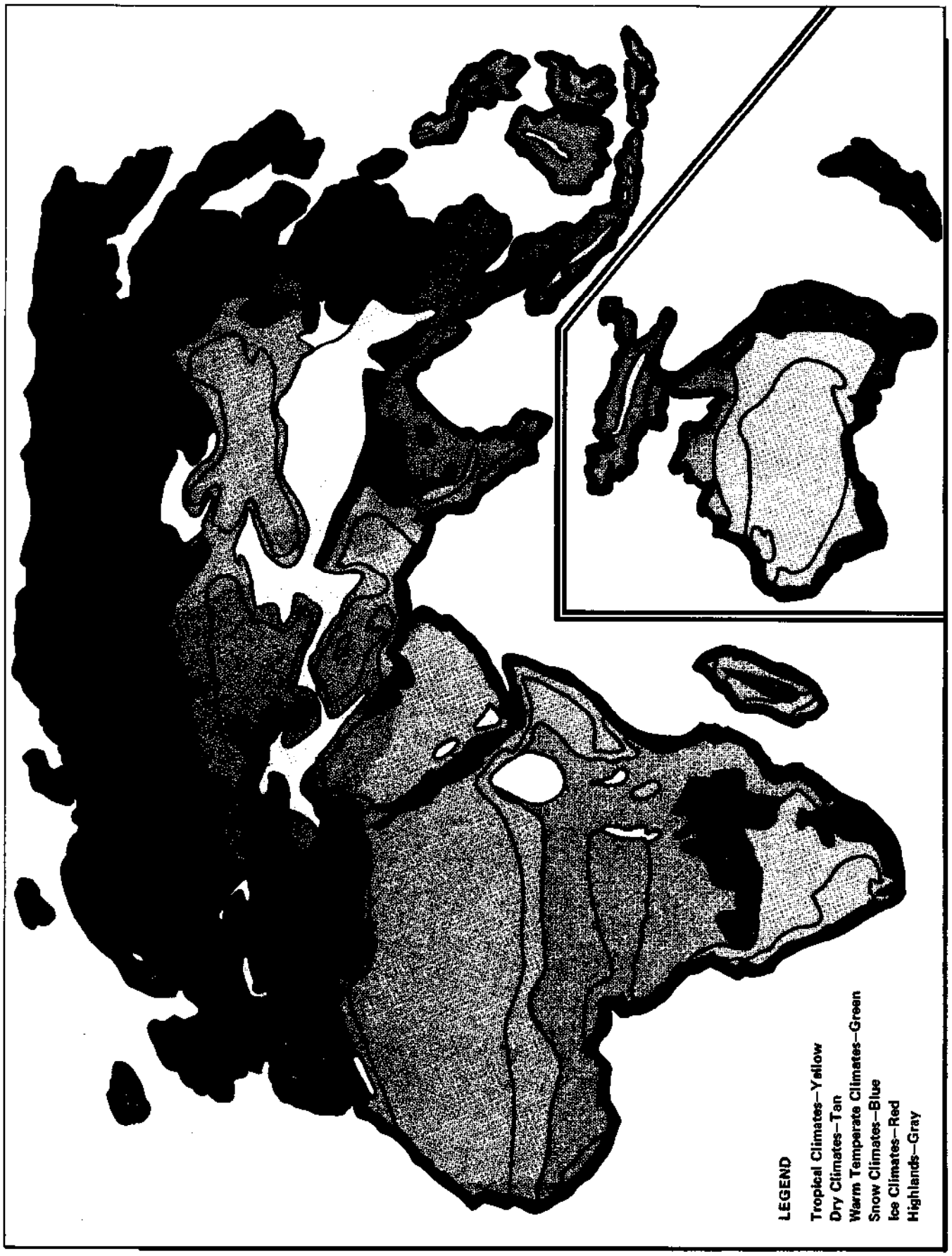

Figure 11-1. Köppen-Geiger System. (cont)

tion. The vines and air plants that grow on the trunks and branches of trees are woody. Grasses and herbs, which are common in the temperate woods of the United States, are rare in the tropical rain forest. The undergrowth consists of woody plants—seedling and sapling trees, shrubs, and young woody climbers. The bamboos, which are really grasses, grow to giant proportions, 20 to 80 feet high in some cases. Bamboo thickets in parts of some rain forests are very difficult to penetrate. The plants that produce edible parts in the jungle are often scattered, and require searching to find several of the same kind. A tropical rain forest (figure 11-2) has a wider variety of trees than any other area in the world. Scientists have counted 179 species in one 8.5-acre area in South America. An area this size in a forest in the United States would have fewer than seven species of trees.

(4) The average height of the taller trees in the rain forest is rarely more than 150 to 180 feet. Old giants of the tropical rain forest attain 300 feet in height, but this is extremely rare. Trees more than 10 feet in diameter are also rare in the jungle. The trunks are, as a rule, straight and slender and do not branch until near the top. The base of many trees is provided with plank buttresses, flag-like outgrowths which are common in all tropical forests. The majority of mature tropical trees have large, leathery, dark-green leaves which resemble laurel leaves in size, shape, and texture. The general appearance is monotonous, and large and strikingly colored flowers are uncommon. Most of the trees and shrubs have inconspicuous flowers, often greenish or whitish.

(5) Travel books often give a misleading impression of the density of tropical forests. On riverbanks or in clearings, where much light reaches the ground, there is a dense growth which is often quite impenetrable. But in the interior of an old undisturbed forest, it is not difficult to walk in any direction. Photographs give an exaggerated notion of the density of the undergrowth. It is usually possible to see another person at least 60 feet away.

(6) The abundance of climbing plants is one of the characteristic features of rain forest vegetation. The great majority of these climbers are woody and many have stems of great length and thickness. Stems as thick as a man's thigh are not uncommon. Some lianas cling closely to the trees that support them, but most ascend to the forest canopy like cables or hang down in loops or

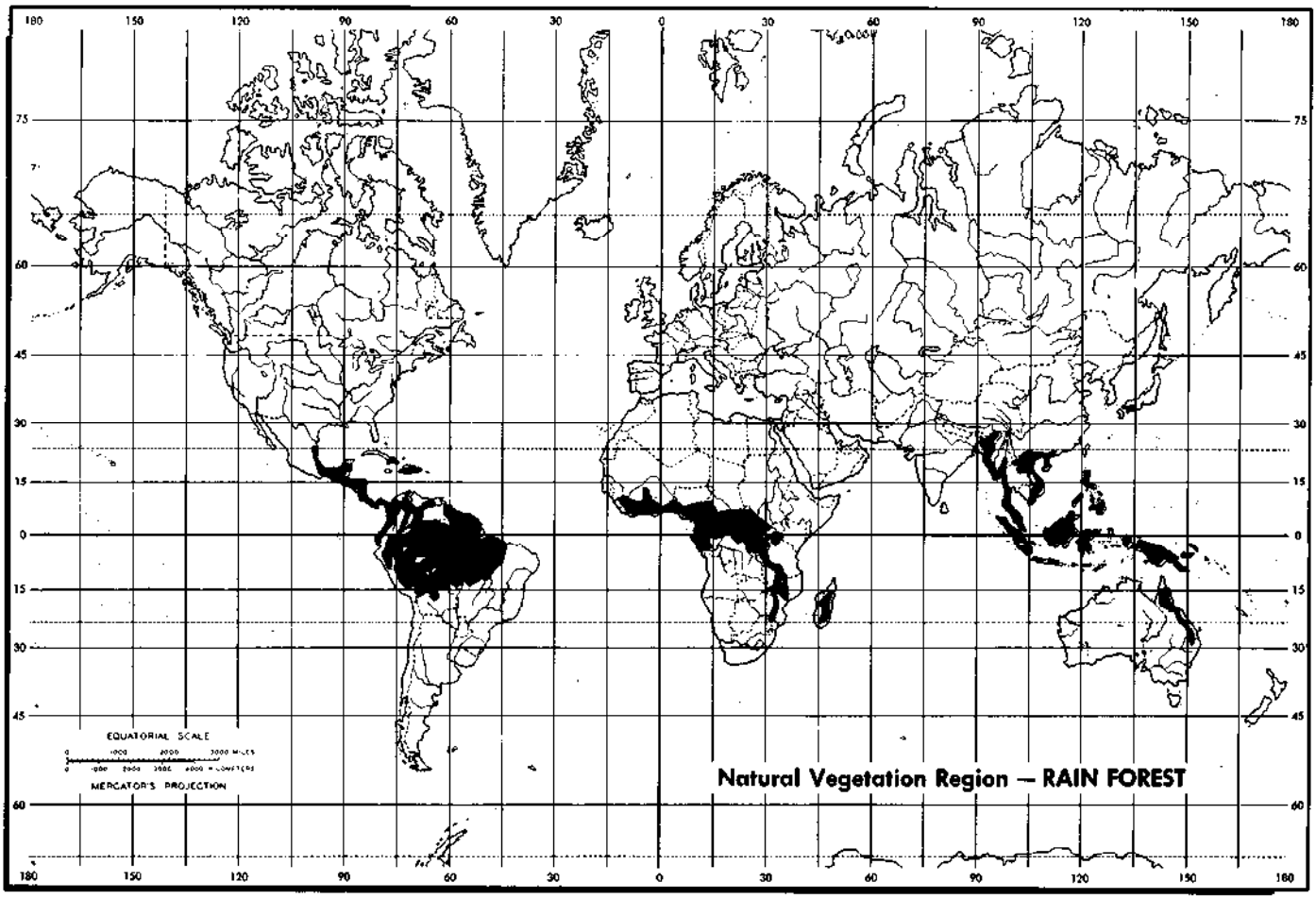

Figure 11-2. Rain Forest.

festoons. In the rain forest, there is no winter or spring, only perpetual midsummer. The appearance of the vegetation is much the same at any time of year. There are seasons of maximum flowering during which more species bloom than at any other time, and also seasons of maximum production of young leaves, plant growth, and reproduction is continuous and some flowers can be found at any time. The margins of a tropical rain forest clearing and areas around abandoned dwellings abound in edible plants. However, in the center of the virgin rain forest, trying to find food is more difficult due to accessibility. The lofty trees are so tall that fruits and nuts are generally out of reach.

c. Distribution of Tropical Rain Forests:

(1) In the Americas, the largest continuous mass of rain forest is found in the basin of the Amazon River. This extends west to the lower slopes of the Andes and east to the Atlantic coast of the continent; it is broken only by relatively small areas of savanna and deciduous forest. This great South American rain forest extends south into the region of the Gran Chaco (south-central South America) and north along the eastern side of Central America into southern Mexico and into the Antilles chain of the West Indies. In the extreme northwest of South America (Ecuador, Colombia), there is a narrow belt of rain forest, separated from the Amazonian forest by a wide expanse of deciduous forest, extending from about latitude 6 degrees South to a little beyond the Tropic of Capricorn. The distribution of rain forest in Central America is perhaps less well known than any other major tropical region. The main areas are below the 500-foot elevation (figure 11-3).

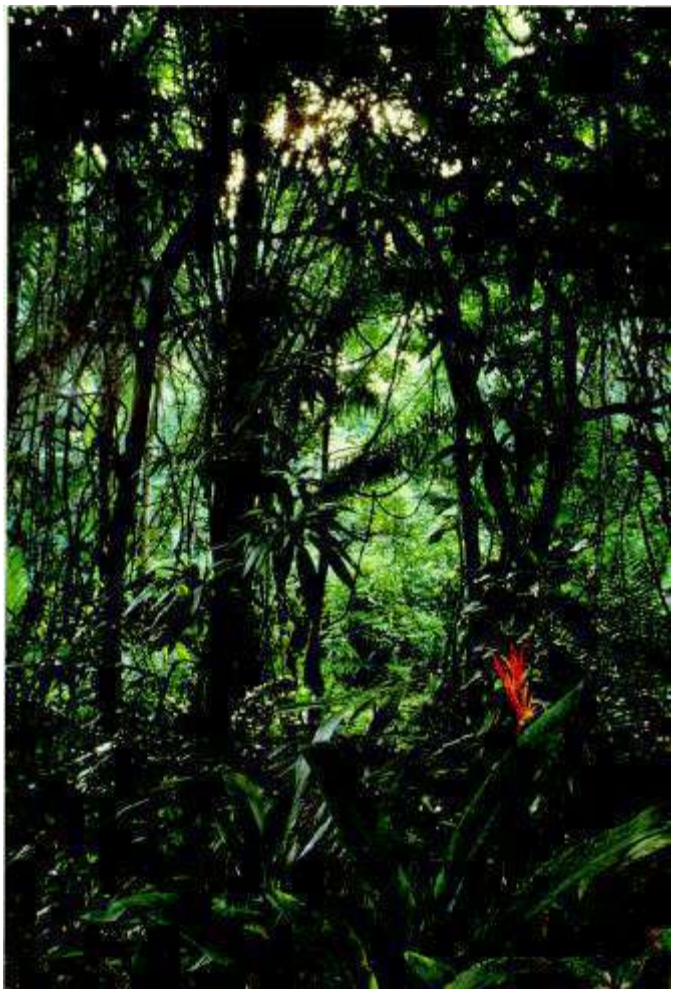

©National Geographic Society, Photo by James Stanfield

Figure 11-3. Rain Forest (El Salvador).

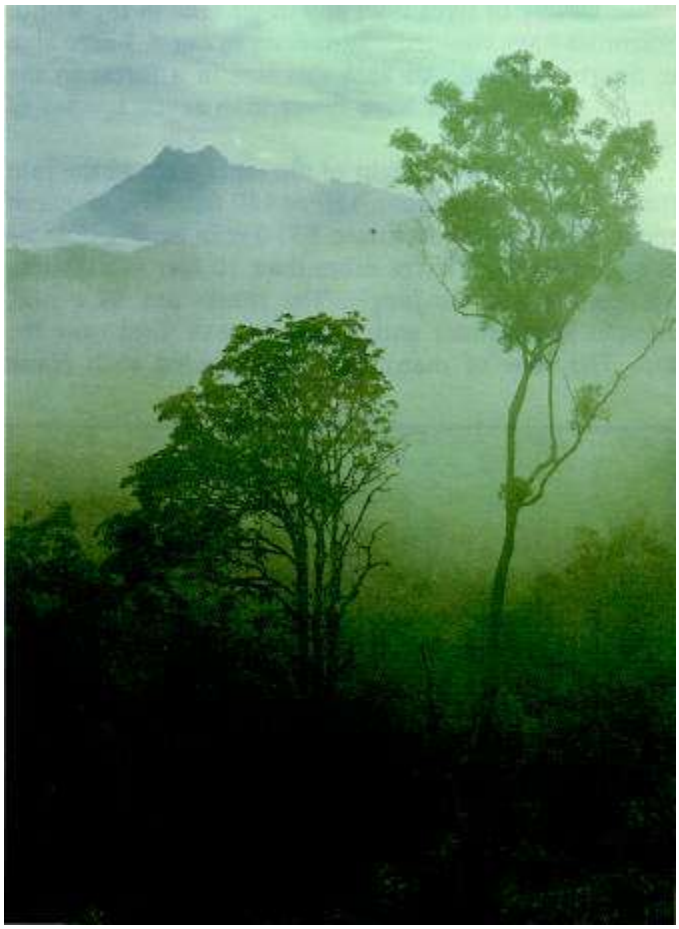

©National Geographic Society, Photo by Winfield Parks

Figure 11-4. Rain Forest (Borneo).

(2) In Africa, the largest area of rain forest lies in the Congo basin and extends westward into the Republic of Cameroon. As a narrow strip, the forest continues still farther west, parallel to the Gulf of Guinea, through Nigeria and the Gold Coast of Liberia and Guinea. Southward from the Congo basin, the forest extends towards Rhodesia.

(3) In the eastern tropics, the rain forest extends from Ceylon and Western India to Southeast Asia and the Philippines, as well as through the Malay Archipela-

go to New Guinea. The largest continuous areas are in New Guinea, the Malay Peninsula, and the adjoining islands of Sumatra and Borneo, where the Indo-Malayan rain forest reaches its greatest luxuriance and floral wealth (figure 11-4).

(a) In India, the area of rain forest is not large, but it is found locally in the western and eastern Ghats (coastal ranges) and, more extensively, in the lower part of the eastern Himalayas, the Khasia hills, and Assam.

(b) In Burma and Southeast Asia, the rain forest is developed only locally; the principal vegetation being the monsoon forest. The monsoon forest is a tropical type which is partly leafless at certain seasons.

(c) In the eastern Sunda Islands from western Java to New Guinea, the seasonal drought (due to the dry east monsoon from Australia) is too severe for the development of a rain forest, except in locally favorable situations (figure 11-5).

(d) In Australia, the tropical rain forest of Indo-Malaya is continued south as a narrow strip along the eastern coast of Queensland. Rain forest also extends into the islands of the western Pacific (Solomons, New Hebrides, Fiji, Samoa, etc.). (See figure 11-6.)

d. Food Plants. Some of the available food plants in the rain forest include:

(1) Indian or Tropical Almond.
(2) Rose Apple.
(3) East Indian Arrowroot.
(4) Bullock's Heart.
(5) Sugar Cane.
(6) Cattail.
(7) Bael Fruit.
(8) Water Chestnut (Trapa Nut).
(9) Bamboo.
(10) Chufa (Nut Grass).
(11) Goa Bean.
(12) Luffa Sponge (Wild Gourd).
(13) Yam Bean.
(14) Wild Fig.
(15) Bignay.
(16) Wild Grape.
(17) Lotus Lilly.
(18) Water Lettuce.
(19) Breadfruit.
(20) Canna Lily.
(21) Bracken (Fern).
(22) Sego Palm.
(23) Tree Fern.
(24) Palm Sugar.
(25) Mango.
(26) Papaya.
(27) Italian Millet.
(28) Screw Pine.
(29) Pearl Millet.
(30) Plantain.
(31) Mulberry.
(32) Batolo Plum.

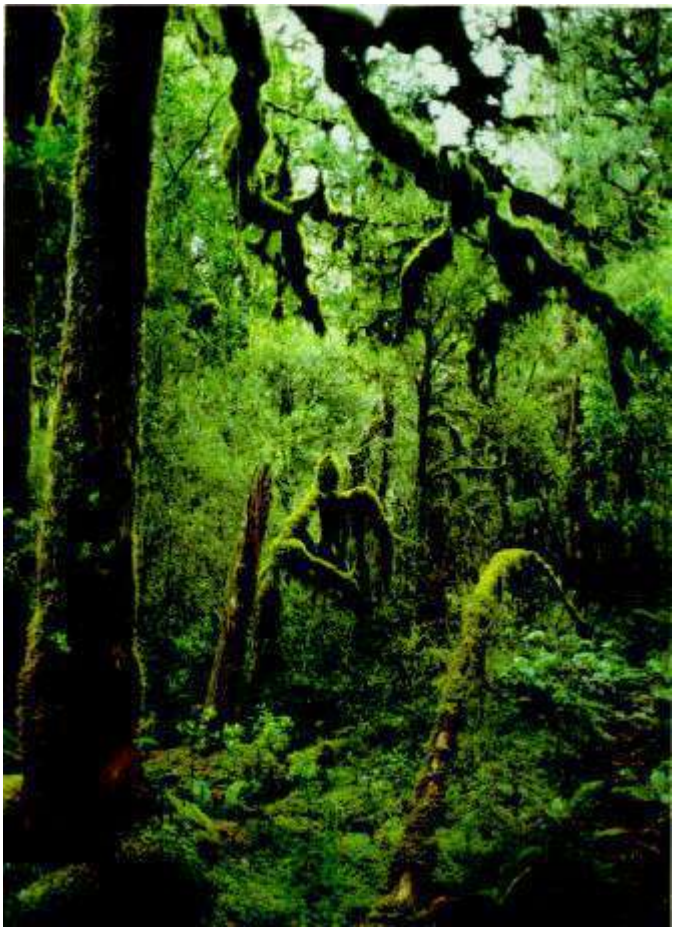

©National Geographic Society, Photo by Robert E. Gilka

Figure 11-5. Rain Forest (New Zealand).

(33) Cashew Nut.
(34) Pokeweed.
(35) Buri Palm.
(36) Polypody.
(37) Fishtail Palm.
(38) Air Potato (ubi tuber).
(39) Coconut Palm.
(40) Purslane.
(41) Nipa Palm.
(42) Rice.
(43) Rattan Palm.
(44) Soursop.
(45) Ceylon Spinach.
(46) Water Lilly.
(47) Sterculia.
(48) Sweetsop.
(49) Tamarind.
(50) Taro.
(51) Ti Plant.
(52) Horseradish Tree.
(53) Tropical Yam.

©National Geographic Society, Photo by Luis Marden

Figure 11-6. Rain Forest (Tahiti).

e. Semi-Evergreen Seasonal Forest:

(1) In character, the semi-evergreen seasonal forest in Central and South America and Africa corresponds essentially to the monsoon forest of Asia. Characteristics of the semi-evergreen forest are:

(a) Two stories of tree strata—upper story 60 to 80 feet high; lower story 20 to 45 feet high.

(b) Large trees are rare; average diameter about 2 feet.

(c) Seasonal drought causes leaf fall; more in dry years.

(2) The peculiar distribution of the rainy season and the dry season which occurs in the countries bordering the Bay of Bengal in southeastern Asia brings on the monsoon climate. The monsoons of India, Burma, and Southeast Asia are of two types. The dry monsoon occurs from November to April, when the dry northern winds from central Asia bring long periods of clear weather with only intermittent rain. The wet monsoon occurs from May to October, when the southern winds from the Bay of Bengal bring rain, usually in torrents, that lasts for days and often weeks at a time. During the dry season, most leaves drop completely off, giving the landscape a wintry appearance, but as soon as the monsoon rains begin, the foliage reappears immediately.

f. Plant Foods of the Semi-Evergreen Seasonal Forest. These foods include:

(1) Agave (Century Plant).
(2) Amaranth.
(3) Bael Fruit.
(4) Banana.
(5) Tropical or Indian Almond.
(6) Rose Apple.
(7) Bamboo.
(8) Goa Bean.
(9) Yam Bean.
(10) Mango.
(11) Purslane.
(12) Mulberry.
(13) Bignay.
(14) Italian Millet.
(15) Soursop.
(16) Cashew Nut.
(17) Breadfruit.
(18) Pearl Millet.
(19) Ceylon Spinach.
(20) Sterculia.
(21) Sugar Cane.
(22) English Acorns (Oak).
(23) Luffa Sponge (Wild Gourd).
(24) Cattail.
(25) Buri Palm.
(26) Sweetsop.
(27) Chestnut.
(28) Water Chestnut (Trapa Nut).
(29) Rattan Palm.
(30) Tamarind.
(31) Chufa (Nut Grass).
(32) Papaya.
(33) Taro.
(34) Ti Plant.
(35) Wild Fig.
(36) Screw Pine.
(37) Horseradish Tree.
(38) Tree Fern.
(39) Plaintain.
(40) Tropical Yam.
(41) Wild Grape.
(42) Pokeweed.
(43) Water Lettuce.
(44) Polypody.
(45) Canna Lily.
(46) Air Potato (ubi tuber).
(47) Lilly Lotus.
(48) Water Lily.

g. Tropical Scrub and Thorn Forest (figure 11-7):

(1) Chief Characteristics of the Tropical Scrub and Thorn Forest:

(a) Definite dry season, with wet season varying in length from year to year. Rains appear mainly as downpours from thunderstorms.

(b) Trees are leafless during dry season; average height is 20 to 30 feet with tangled undergrowth in places (figure 11-8).

(c) Ground is bare except for a few tufted plants in bunches; grasses are not common.

(d) Plants with thorns are predominate.

(e) Fires occur at intervals.

(2) Food Plants:

(a) Within the tropical scrub and thorn forest areas, the survivor will find it difficult to get food plants in the dry season. During the height of the drought period, the primary kinds of foods come from the following plant parts:

-1. Tubers.
-2. Bulbs.
-3. Pitch.
-4. Nuts.
-5. Rootstalks.
-6. Corms.
-7. Gums and Resins.
-8. Seeds and Grains.

(b) During the rainy season in the tropical scrub and thorn forest, plant food is considerably more abundant. At this time, the survivor should look for the following edible plants:

-1.Sweet Acacia.
-2.Wild Chicory.
-3.St. John's Bread.
-4.Wild Caper.
-5.Agave (Century Plant).
-6.Wild Fig.
-7.Almond.
-8.Cashew Nut.
-9.Baobab.
-10.Juniper.
-11.Tamarind.
-12.Tropical Yam.
-13.Sea Orach.
-14.Prickly Pear.
-15.Wild Pistachio.
-16.Air Potato (ubi tuber).

h. Tropical Savanna. (See figure 11-9.)

(1) General Characteristics of the Savanna:

(a) Savannas lie wholly within the tropical zone in South America and Africa.

(b) The savanna looks like a broad, grassy meadow with trees spaced at wide intervals.

(c) The grasses of the tropical savanna often exceed the height of a man. However, none of the savanna grasses are sod-forming in the manner of lawn grasses, but are bunch grasses with a definite space between each grass plant.

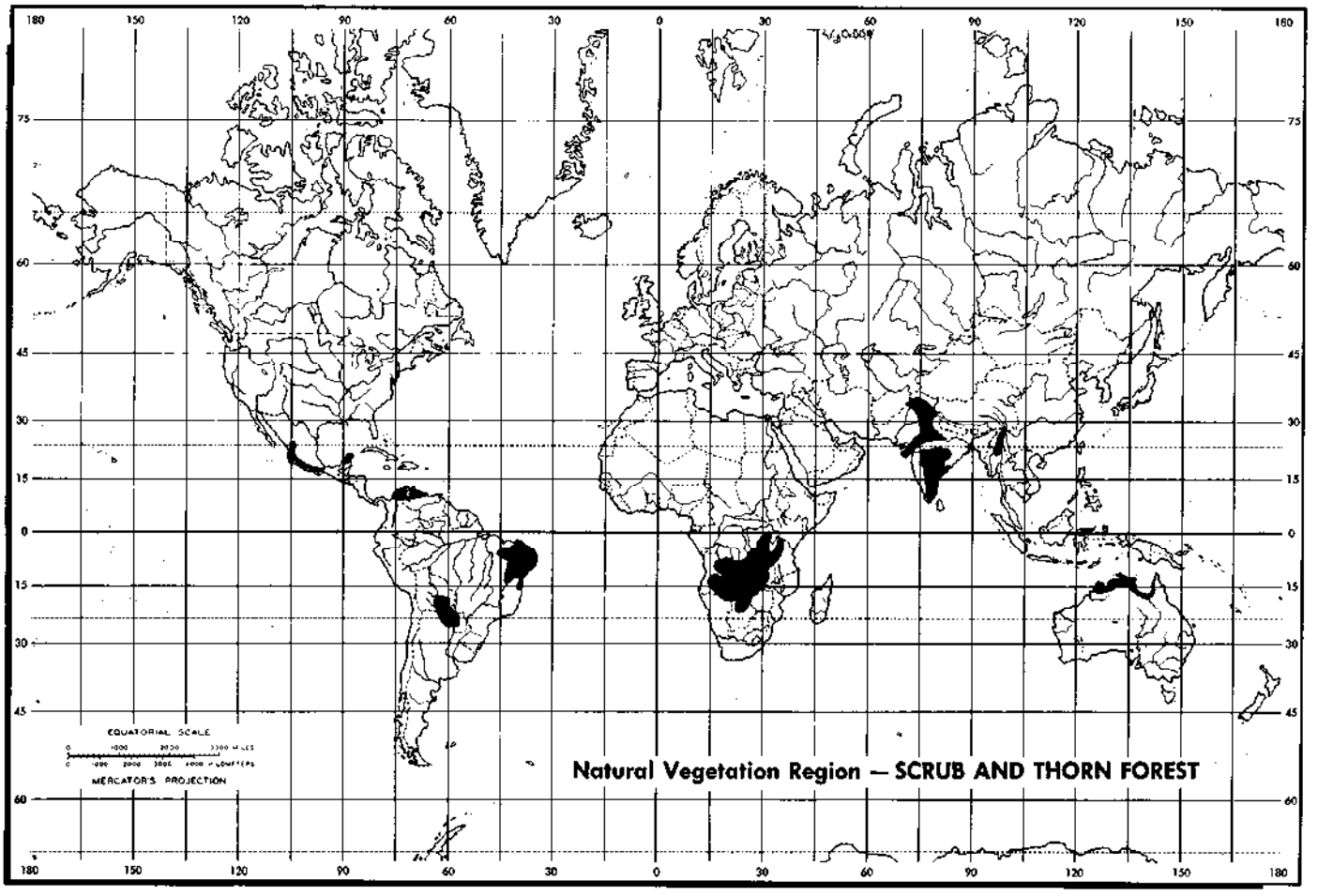

Figure 11-7. Shrub and Thorn Forest.

©National Geographic Society, Photo by Gilbert M. Grosvenor

Figure 11-8. Tropical Shrub or Thorn Forest (Kenya).

(d) The soil in the savanna is frequently red.

(e) The scattered trees usually appear stunted and gnarled like old apple

(f) Palms may be found on savannas.

(2) Savanna of South America. For the most part, the vegetation is of the bunch-grass type. A long, dry season alternates with a rainy season. In these areas, both high and low grasses are present. Bright colored flowers appear between the grass bunches during the rainy season. The grains from the numerous grasses are useful as survival food, as well as the underground parts of the many seasonal plants that appear with and following the rains.

(3) Savanna in Africa. The high grass tropical savanna of Africa is dominated by very tall, coarse grasses which grow from 5 to 15 feet high. Unless the natives burn the grass during the dry season, the savanna becomes almost impenetrable. This type of savanna occurs in a broad belt surrounding the tropical rain forest and extends from western Africa eastward beyond the Nile River. From the Nile, it extends southward and westward. The tropical bunch grass savanna comprises the greatest part of the African savanna consisting of grasses about 3 feet tall. The African savanna has both dwarf and large trees. The most renowned of these large trees is the monkeybread or baobab (figure 11-10).

(4) Food Plants. The food plants found on the savanna are also found in other vegetation areas.

(a) Amaranth.	(g) Wild Chicory.
(b) Wild Crabapple.	(h) Wild Fig.
(c) Purslane.	(i) Tamarind.
(d) Wild Apple.	(j) Chufa (Nut Grass).
(e) Wild Dock.	(k) Water Plaintain.
(f) Wild Sorrel.	(l) Water Lily.

i. Animal Life. The tropics abound in animal life. The tremendous varieties of animal species found in tropical areas throughout the world preclude discussions of each animal. It is essential that the survivor realize that just as people have an inherent fear of some animals, most animals also fear people. With some exceptions, animals of the tropics will withdraw from any encounter with humans. Being primarily nocturnal animals, most

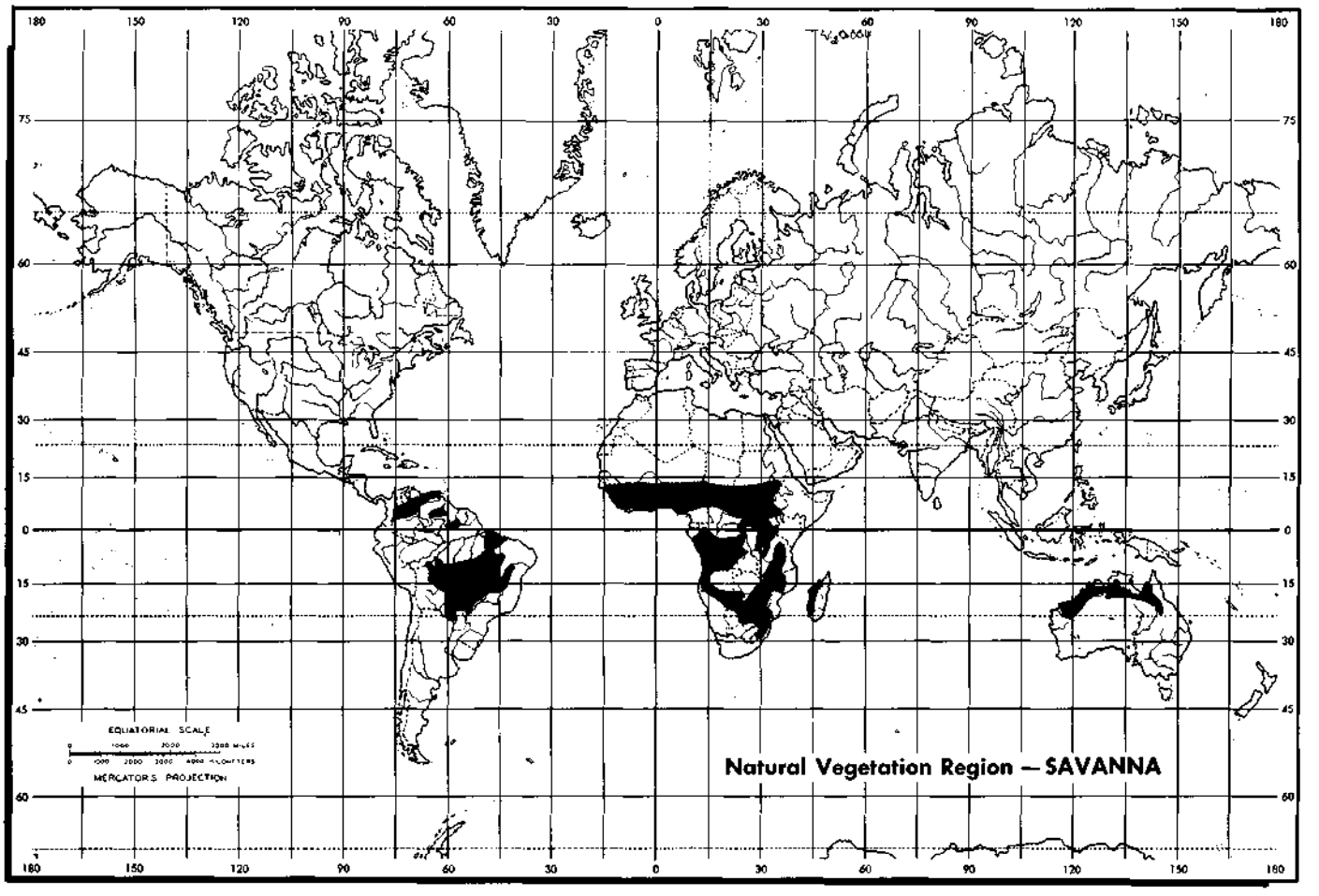

Figure 11-9. Savanna.

will never be seen by the survivor. By becoming familiar with the wild inhabitants of the tropics, the survivor will better understand this type of environment and will respect, not fear, the surroundings in which survival takes place.

(1) All tropic areas have members of the pig family. By habit, pigs are gregarious and are omnivorous in diet. They will eat any small animals they can kill, although they feed mainly on roots, tubers, and other vegetable substances. The most common species found in the Old World tropics are the peccary, the Indian wild boar, the Babirussa of Celebes, and the Central African Giant Forest Hog. In Central and South American tropics, peccaries are common. These pigs are represented by two species, the "white-lipped" peccary and the "collared" peccary. Both are grizzled black color, distinguishable by markings from which they derive their names. The white-lipped peccary, the larger of the two (height of approximately 18 inches), is black with white under the snout, and has the reputation of being the more ferocious. The collared peccary, reaching a height of 14 inches, is identified by the white or gray band around the body where the neck joins the shoulder. The collared peccary often travels in groups of 5 to 15. While alone, they are not particularly dangerous, but a pack can effectively repel any enemy and can make short work of a jaguar, cougar, or human. Both types of peccaries have musk glands which are located 4 inches up from the tail on the spine. This gland must be removed soon after the animal is killed, otherwise the flesh will become tainted and unfit for consumption.

(2) Tropic areas harbor many species of reptiles and amphibians. Most of them are edible when skinned and cooked. Hazards from these animals are mostly imagined; however, some are venomous or dangerous if encountered. Individual species of crocodylidae family (alligators, crocodiles, caiman, and gavials) are usually only abundant in remote areas away from humans. Most dangerous are the saltwater crocodiles of the Far East and the Nile crocodiles in Africa. Poisonous snakes, while numerous in the tropics, are rarely seen and pose little danger to the wary survivor. There are no known poisonous lizards in the tropics. Several species of frogs and toads contain poisonous skin secretions.

PLAINS OF TANGANYIKA

©National Geographic Society, Photo by W. Robert Moore

KENYA

©National Geographic Society, Photo by Melvin M. Payne

NAIROBI, KENYA

©National Geographic Society, Photo by W. Robert Moore

IHOSY, MADAGASCAR

©National Geographic Society, Photo by Al Moldvay

Figure 11-10. Savanna (Ihosy, Madagascar; Plains of Tanganyika; Kenya; and Nairobi, Kenya).

The large, pan-tropic toad, Bufo Marimum, exudes a particularly irritating secretion if handled roughly. Aside from skin irritations, these amphibians pose little danger to humans, unless the secretion gets into the eyes, where it may cause blurred vision, intense burning, and possible blindness.

(3) All tropical areas of the world abound in monkeys. They are very curious animals and this fact may be used to the survivor's advantage in trying to procure one for food. Only the very large species of monkey, such as the mandrill or baboon, could constitute any danger to humans.

(4) Tropic areas also have a large number of the various species of mice, rats, squirrels, and rabbits.

(5) Members of the cat family are found in jungles throughout the world. The ocelot abounds in the jungles of Central and South America. It is a small, lean, savage cat whose coloring closely resembles that of the jaguar and will attain a weight of approximately 40 pounds and a length of 3 feet when fully grown. Cats such as the leopard are found in the tropics of the old world. The leopard is one of the most wary of beasts, becoming a powerful fighter when wounded or cornered. Unlike lions and tigers, the leopard can climb trees with ease; therefore, caution should be used when hunting this animal.

(6) Deer are found in most jungle areas; however, their population is normally small. In the Asian jungles,

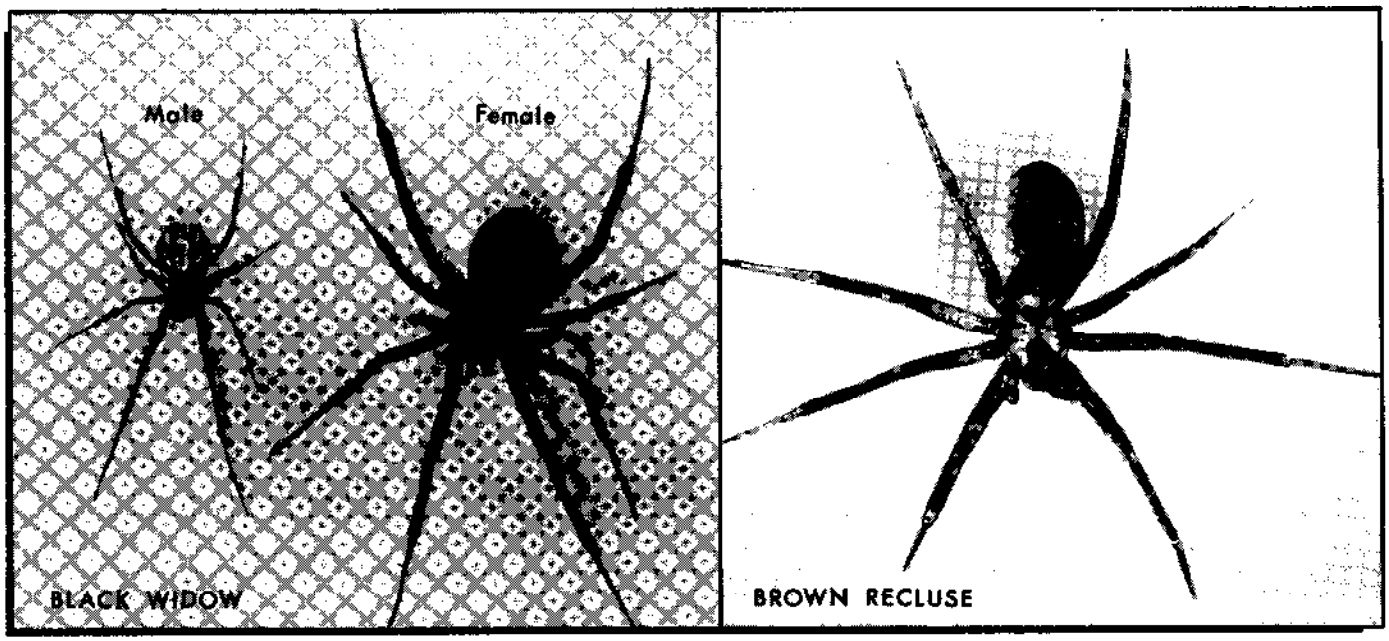

Figure 11-11. Black Widow and Brown Recluse Spiders.

several species of deer frequent the low, marshy areas adjacent to rivers. In the Central and South American jungles, two species of deer are most common. The jungle species is found in thick upland forests. It is much smaller than the North American species and seldom attains a weight of more than 80 pounds. Another deer found in the Central and South American jungles is the "brocket" or "jungle deer." This small reddish-brown deer, which attains a height of about 23 inches, is extremely shy and is found mostly in dense cover since it has no defense against other animals.

(7) The real dangers lie in the insects located in the jungle, which can pass on diseases or parasites.

(a) Malaria may be the worst enemy. It is transmitted by mosquitoes, which are normally encountered from late afternoon until early morning. Guard against bites by camping away from swamps on high land and sleep under mosquito netting, if available; otherwise, use mud on the face as a protection against insects. Wear full clothing, especially at night, and tuck pants into the tops of socks or shoes. Wear mosquito head net and gloves. Take antimalaria tablets (if available) according to directions.

(b) The greatest number of ant species is found in the jungle regions of the world. Nesting sites may be in the ground or in the trees. Ants can be a considerable nuisance especially if near a campsite. They inflict pain by biting, stinging, or squirting a spray of formic acid. Before selecting a campsite, a close check of the area should be made for any nests or trails of ants.

(c) Ticks may be numerous, especially in grassy places. Use a protected area and undress often, inspecting all parts of the body for ticks, leeches, bed bugs, and other pests. If there are several people in the group, examine each other.

(d) Fleas are common in dry, dusty buildings. The females will burrow under the toenails or into the skin to lay their eggs. Remove them as soon as possible. In India and southern China, bubonic plague is a constant threat. Rat fleas carry this disease and discovery of dead rats usually means a plague epidemic in the rat population. Fleas may also transmit typhus fever and in many parts of the tropics, rats also carry parasites which cause jaundice and other fevers. Keep food in rat-proof containers or in rodent-proof caches. Do not sleep with any food in the shelter!

(e) In many parts of the Far East, a type of typhus fever is carried by tiny red mites. These mites resemble the chiggers of southern and southwestern United States. They live in the soil and are common in tall grass, cut-over jungle, or stream banks. When a person lies or sits on the ground, the mites emerge from the soil, crawl through clothes, and bite. Usually people don't know they have been bitten, as the bite is painless and does not itch. Mite typhus is a serious disease and the survivor should take preventive measures to avoid this pest. The survivor should clear the camping ground and burn it off, sleep above the ground, and treat clothing with insect repellant.

(f) Leeches are primarily aquatic and their dependence on moisture largely determines their distribution. The aquatic leeches are normally found in still, freshwater lakes, ponds, and waterholes. They are attracted by disturbances in the water and by a chemical sense.

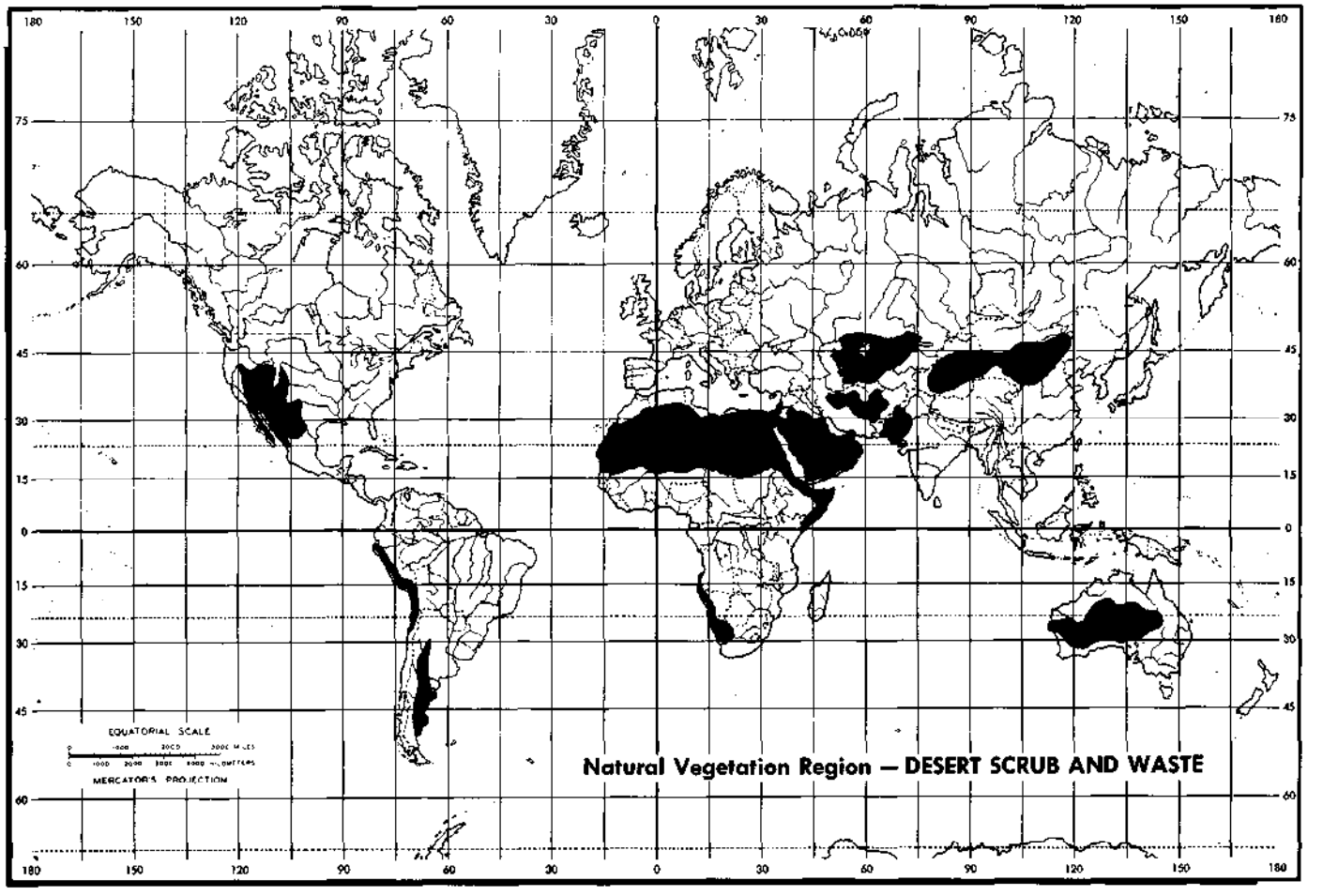

Figure 11-12. Desert Scrub and Waste Areas.

Land leeches are quite bloodthirsty and easily aroused by a combination of odor, light, temperature, and mechanical sense. These leeches are the most feared of all since they may enter air passages from which they cannot escape once they have fed and become distended. Normally, there is little pain when leeches attach themselves, and after they fill with blood, they drop off unnoticed. Some leeches, living in springs and wells, may enter the mouth or nostrils when drinking and may cause bleeding and obstruction.

(g) Spiders, scorpions, hairy caterpillers, and centipedes are often abundant. The survivor should shake out shoes, socks, and clothing and inspect bedding morning and evening. A few spiders have poisonous bites which may cause severe pain. The black widow and the brown recluse spiders are venomous and should be considered very dangerous (figure 11-11). The large spiders called tarantulas rarely bite, but if touched, the short, hard hairs which cover them may come off and irritate the skin. Centipedes bite if touched and their bite is like that of a wasp's sting. Avoid all types of many-legged insects. Scorpions are real pests as they like to hide in clothing, bedding, or shoes and strike without being touched. Their sting can cause illness or death.

j. Population. Density of human population varies with the climate. Cultivation is difficult in areas of tropical rain forests along the Equator. The torrential rains leach out the soil and weeds grow rapidly. Consequently, cultivated food sources must be supplemented by game and other products of the forest. Villages are usually scattered along rivers since movement is easier by water than through the dense forest. Numerous people are also located along coastal areas where farming takes place and people can obtain food from the sea.

11-4. Dry Climates. Dry climates are generally thought of as hot, barren areas that receive scanty rainfall. Rainfall is limited but dry climates are not barren wastelands and many kinds of plants and animals thrive (figure 11-12).

a. Deserts:

(1) Most deserts are located between the latitudes of 15 and 35 degrees on each side of the Equator and are

©National Geographic Society, Photo by Thomas J. Abercrombie

Figure 11-13. The Sahara.

dry regions where the annual evaporation rate exceeds the annual precipitation rate (generally less than 10 inches of rain annually). Extremes of temperature are as characteristic of deserts as is lack of rain and great distances. Hot days and cool nights are usual. A daily low-high of 45°F in the Sahara Desert and a 25°F to 35°F difference in the Gobi Desert is the rule. The difference between summer and winter temperatures is also extreme.

(2) Deserts occupy nearly 20 percent of the Earth's land surface, but only about 4 percent of the world's population lives there. The term "desert" is applied to a variety of areas. There are alkali deserts, rock deserts, and sand deserts. Some are barren gravel plains without a spear of grass, a bush, or cactus for a hundred miles. In other deserts, there are grasses and thorny bushes where camels, goats, or even sheep find a subsistent diet. Anywhere they are found, deserts are places of extremes. They can be extremely dry, hot, cold, and often devoid of plants, trees, lakes, or rivers. Most important to the survivor is the extremely long time (distance) between water sources. The vast deserts of North Africa, Arabia, Iran, West Pakistan, the Mojave Desert of the southwestern United States and northern Mexico, the interior Kalahari Desert of South Africa, and the Australian Desert are major examples of this type.

(3) Desert areas and climatic characteristics and seasonal variations of world deserts are as follows:

(a) Sahara Desert. The Sahara is the largest desert in the world. It stretches across North Africa from the Atlantic Ocean to the Red Sea and from the Mediterranean and the Sahara Atlas Mountains in the north to the Niger River in tropical Africa. It consists of 3 million square miles of level plains and jagged mountains, rocky plateaus, and graceful sand dunes. There are thousands of square miles where there is not a spear of grass, not a bush or tree, nor a sign of any vegetation. But Sahara oases—low spots in the desert where water can be reached for irrigation—are among the most densely populated areas in the world. Date groves and garden patches supporting 1,000 people per square mile are surrounded by barren plains devoid of life. Only 10

percent of the Sahara is sandy. The greater part of the desert is flat gravel plain from which the sand has been blown away and accumulated in limited areas forming dunes. There are rocky mountains rising 11,000 feet above sea level and there are a few depressions 50 to 100 feet below sea level. The change from plain to mountain is abrupt in the Sahara. Mountains generally go straight up from the plain like jagged skyscrapers from a city street. Sharp rising mountains on a level plain are especially noticeable in many desert landscapes because there is no vegetation to modify that abruptness. The lack of trees or bushes makes even occasional foothills appear more abrupt than in temperate climates (figure 11-13).

(b) Arabian Desert. Some geographers consider the Arabian Desert as a continuation of the Sahara. Half a million square miles in area, the Arabian Desert covers most of the Arabian Peninsula except for fertile fringes along the Mediterranean Sea, Red Sea, Arabian Sea, and the valleys of the Tigris-Euphrates Rivers. Along much of the Arabian coastline, the desert meets the sea. There is more sand in the Arabian Desert than in the Sahara and there are fewer date grove oases. These are on the east side of the desert at Gatif, Hofuf, and Medina. Also, there is some rain in Arabia each year, in contrast to the decades in the Sahara that pass without a drop. Arabia has more widespread vegetation, but nomads find scanty pasture for their flocks of sheep and goats and must depend on wells for water. Oil is carried across the desert in pipelines which are regularly patrolled by aircraft. Pumping stations are located at intervals. All these evidences of modern civilization have increased the well-being of the desert people and, as a result, chances for a safe journey afoot. However, the desert of Arabia is rugged, and native Arabs still get lost and die from dehydration.

(c) Gobi Desert - "Waterless Place." As used here, "Gobi" means only the 125,000 square mile basin or saucer-like plateau north of China which includes Inner and Outer Mongolia. On all sides of the Gobi, mountains form the rim of the basin. The basin itself slopes so gently that much of it appears to be a level plain. The Gobi has rocks, buttes, and numerous badlands, or deeply gullied areas (figure 11-14). For a hundred miles or so around the rim of the Gobi, there is a band of grassland. In average years, the Chinese find this to be a productive farmland. In drought years, agriculture retreats. Moving toward the center of the Gobi, there is less and less rainfall; soil becomes thinner, and grass grows in scattered bunches. This is the home of the Mongol herdsman. His wealth is chiefly horses, but he also raises sheep, goats, camels, and a few cattle. Beyond the rich grassland, the Gobi floor is a mosiac of tiny pebbles which often glisten in sunlight. These pebbles were once mixed with the sand and soil of the area, but in the course of centuries, the soil has been washed or blown away and the pebbles have been left behind as loose pavement. What rain there is in the Gobi drains

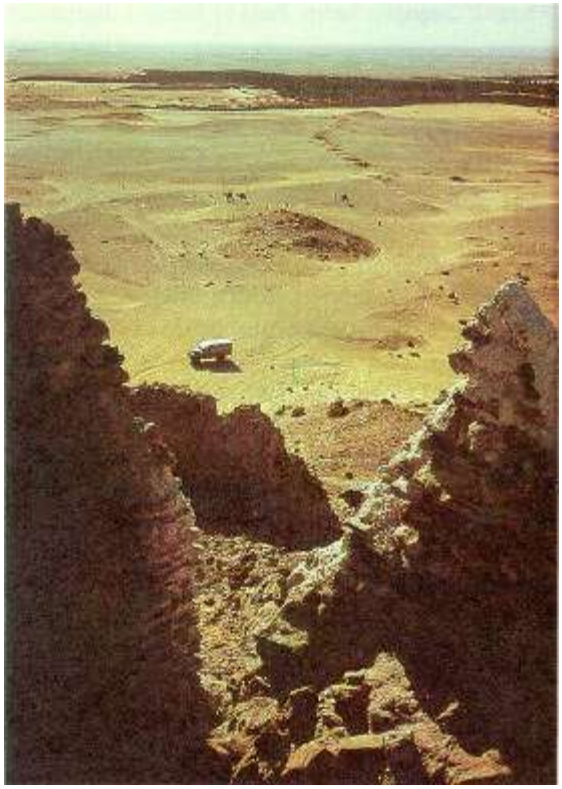

©National Geographic Society, Photo by Dean Conger

←——— ©National Geographic Society, Photo by George Rodger

Figure 11-14. The Gobi.

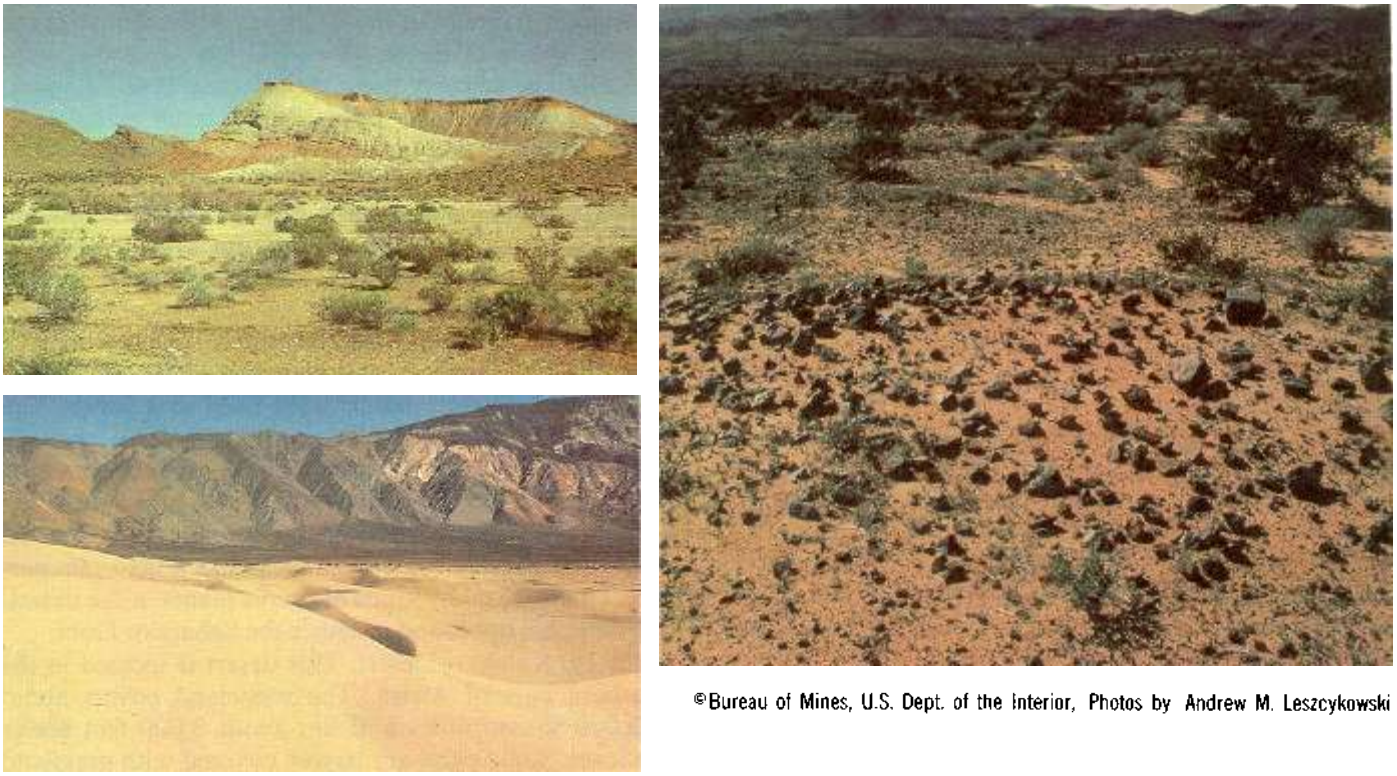

©Bureau of Mines, U.S. Dept. of the Interior, Photos by Andrew M. Leszcykowski

Figure 11-15. Composite of American Deserts.

toward the basin; almost none of it cuts through the mountain rim to the ocean. There are some distinct and well-channeled watercourses, but these are usually dry. Many are remnants of prehistoric drainage systems. In the east, numerous shallow salt lakes are scattered over the plain. They vary in size and number with the changes of rainfall in the area. Sand dunes are found in the eastern and western Gobi, but these features are not as pronounced as they are in certain sections of the Sahara Desert. The Gobi is not a starkly barren wasteland like the great African Desert. Grass grows everywhere, although it is often scanty. Mongols live in collective type farm systems and habitations instead of being concentrated in oases (figure 11-14).

(d) Australian Desert. More than one-third of Australia's total area is desert. Rainfall in the area is unpredictable with an average of less than 10 inches per year. There are three connecting deserts which occupy western Australia and one desert located in the center of the continent. They are the Great Sandy Desert, Great Victorian Desert, Gibson Desert, and Simpon or Arunta Desert. The three largest deserts, Sandy, Victoria, and Arunta are of the sandy type, held in place by vegetation. The Gibson in the western portion, is a stone-type desert. Most of the deserts of Australia have elevations of 1,000 to 2,000 feet.

(e) Atacama-Peruvian Desert of South America. Generally, there are two regions of desert in South America. The first, and by far the largest, is along the west coast, beginning in the southern part of Ecuador, extending the entire coastline of Peru, and reaching nearly as far south as Valparaiso in Chile. This region, of about 2,000 miles in length and approximately 100 miles in width, is classified as true desert. Even so, along the shoreline of Peru as far south as Africa and inland a few miles, there is often a low-cloud or misty-fog layer. The layer is approximately 1,000 feet thick and produces a fine drizzle. Because of this frequent cloud cover and other phenomena, the temperature along the coastal desert is remarkably cool, averaging about 72°F in the summer daytime and about 55°F in the winter daytime. From about 30 degrees south, the cloud cover does not exist and this region may truly be called rainless. The rare and uncertain showers are valueless for cultivated vegetation. Behind the coastal ranges in the higher elevations, the dryness is at a maximum. In the nitrate fields of the Great Atcama Desert, the air is very dry and the slightest shower is very rare. Here the summer daytime temperatures are from 85°F to 90°F. The second desert region is entirely in Argentina (east of the Andes) extending in a finger-like strip from about 30 degrees south, southwest to about 50 degrees south. This region is approximately 1,200 miles long and 100

miles wide. In this highly dissected plateau region, the temperature ranges from a yearly average of 63°F in the north to 47°F in the south. The average annual rainfall pattern is from about 4 inches in the north to about 6 inches in the south.

(f) Southwest Deserts of the United States and Mexico:

-1. These desert areas have four major subdivisions:

-a. Great Basin—the basin between the Rocky Mountains and the Sierra Nevada-Cascade Ranges of southern Nevada and western Utah.

-b. Mojave Desert— Southwestern California.

-c. Sonoran Desert—Southeastern California across southern Arizona into the southwest corner of New Mexico and from Sonora and Baja, California, into Mexico.

-d. Chihuahuan Desert—Lies to the east of the Great Sierra Madre Occidental, spreading north into southwest Texas, southern New Mexico, and southeast corner of Arizona (figure 11-15).

-2. The flat plains with scanty vegetation and abruptly rising buttes of our Southwest are reminders of both the Gobi and Sahara. But the spectacular rock-walled canyons found in the Southwest have few counterparts in the deserts of Africa and Asia. The gullied badlands of the Gobi resemble similar formations in both the Southwest and the Dakotas, but our desert rivers—the lower Colorado, lower Rio Grande, and tributaries such as the Gila and the Pecos—have a more regular supply of water than is found in Old World deserts. The Nile and Niger are, in part, desert rivers but get their water from tropical Africa. They are desert immigrant rivers (like the Colorado, which collects the melting snows of the southern Rockies) and gain sufficient volume to carry them through the desert country. In general, the southwest deserts have more varied vegetation, greater variety of scenery, and more rugged landscape than either the Gobi or the Sahara. In all three areas, it is often a long time (distance) between water sources. Death Valley, a part of the Mojave lying in southern California, probably has more waterholes and more vegetation than exist in vast stretches of the Sahara. The evil reputation of the Valley appears to have been started by unwise travelers who were too terrified to make intelligent searches for food and water. The dryness of the Death Valley atmosphere is unquestioned, but it lacks the vast barren plains stretching from horizon to horizon in the Sahara. Compared to the Sahara, the desert country of southwestern United States and Mexico sometimes looks like a luxuriant garden. There are many kinds of cactus plants in the desert, but these are not found in either the Sahara or Gobi.

(g) Kalahari Desert. This desert is located in the southern part of Africa. The wasteland covers about 200,000 square miles and lies about 3,000 feet above sea level. Some parts are largely covered with grassland and scrubby trees. The climate is similar to that of the Atacama-Peruvian Desert of South America.

b. Vegetation. The following are the more common xerophytic plants (those plants that can live with a limited water supply) which are found in the major deserts of the world.

(1) Cactus Family (Cactacae). Most cactus fruits and leaves have spines to protect them from birds and

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

Figure 11-16. Prickly Pear.

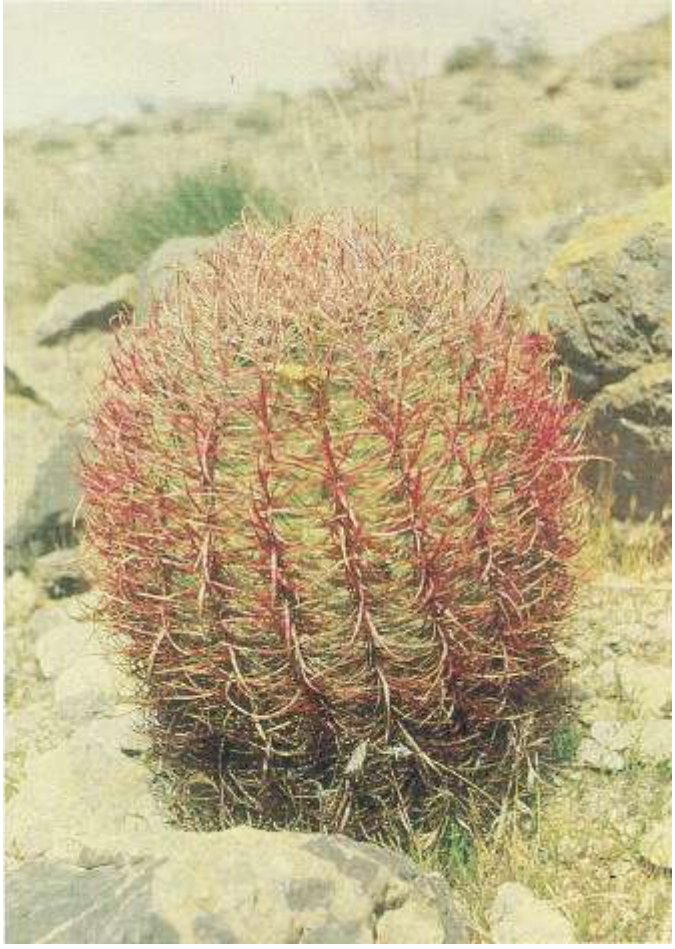

©Bureau of Mines, U.S. Dept. of the Interior, Photo by Andrew M. Leszcykowski

Figure 11-17. Barrel Cactus.

animals seeking water stored in the stems and leaves. Flat cactus leaves can be boiled and eaten and the flowering fruit is edible.

(a) Prickly Pear. Is native and most abundant in the American deserts, but has been introduced into the Gobi, Sahara, and Australian Deserts, and other parts of the world (figure 11-16).

(b) Barrel Cactus. Found in many places, but is only native to the North American deserts. It grows to 5 or 6 feet high (figure 11-17).

(c) Suguaro (Giant) Cactus. Abundant in southern Arizona and in Sonora, Mexico. Can grow up to 50 feet tall.

(2) Wild Onion. Found in the Great Basin of the Southwest and in the Gobi Desert. The bulbs are edible if they look, smell, and taste like an onion or garlic.

(3) Wild Tulip. Found in the Gobi and Sahara Deserts. The bulbs can be eaten.

(4) Shrubs:

(a) Abal. Grows to about 4 feet tall in sandy deserts. The fresh flowers can be eaten. The dry twigs can be crushed and used as a tea substitute. It is found in the Sahara and Arabian Deserts.

(b) Acacia. Most common in the Sahara, Gobi, and Australian desert regions, and in the warmer and drier parts of America. The beans can be crushed and cooked as porridge. It is spiny with many branches and grows to 10 feet tall; roots yield water 4 to 5 feet from the tree trunk (figure 11-18).

(c) Saxaul. Found on the salt deserts of the Gobi Desert. The bark acts as water storage and is a good water source.

(d) St John's Bread. Found along the border of the Mediterranean coast of the Sahara and across the Arabian Desert. Grows 40 to 50 feet tall and seeds can be pulverized and cooked as a porridge (most nutritious plant food in the Middle East).

(e) Juniper. Found in the mountainous areas of the American deserts.

(f) Vines. Wild desert gourds are found in the Sahara and Arabian Deserts. They have a vine which grows from 8 to 30 feet long, and produce a melon-like poisonous fruit. The seed can be eaten when roasted or boiled. Flowers are also edible (figure 11-19).

(g) Succulent Plants. They are filled with juices and store moisture to survive. The surface is covered with a layer of wax or a blanket of fine hairs for protection against the heat. The moisture is contained in a tough cellulos that is not digestible and must be manually broken down to release the water.

(h) Creosote Bush. This is the most widespread and successful plant of the American deserts (height from 2 to 10 feet).

(5) Dates. Occur in groves around desert oases of the lower areas of the Sahara Desert.

c. Animal Life. There are over 5,000 species of birds, reptiles, mammals, and insects found in desert areas. The raven, dove, woodpecker, owl, and hawk are common bird species. Reptiles such as lizards and snakes are numerous due to their adaptation to desert areas and ability to conserve body fluids. Many types of mammals live in desert areas and are primarily found near water sources.

d. Human Population. Humans are greatly influenced by the presence of water and they live close to rivers, wells, cisterns, or oases. For example, in the Sahara Desert, the 2 million people inhabiting this area are located near about 50 desert oases and small coastal cities. In the Gobi Desert, the Mongols live in scattered camps and move from one well to the next as they travel. In the southwest deserts of America, the population is greater along the Colorado and Rio Grande rivers.

11-5. Warm Temperate Climates:

a. The temperate zone is the area or region between the Tropic of Cancer and the Artic Circle and between the Tropic of Capricorn and the Antarctic Circle. The

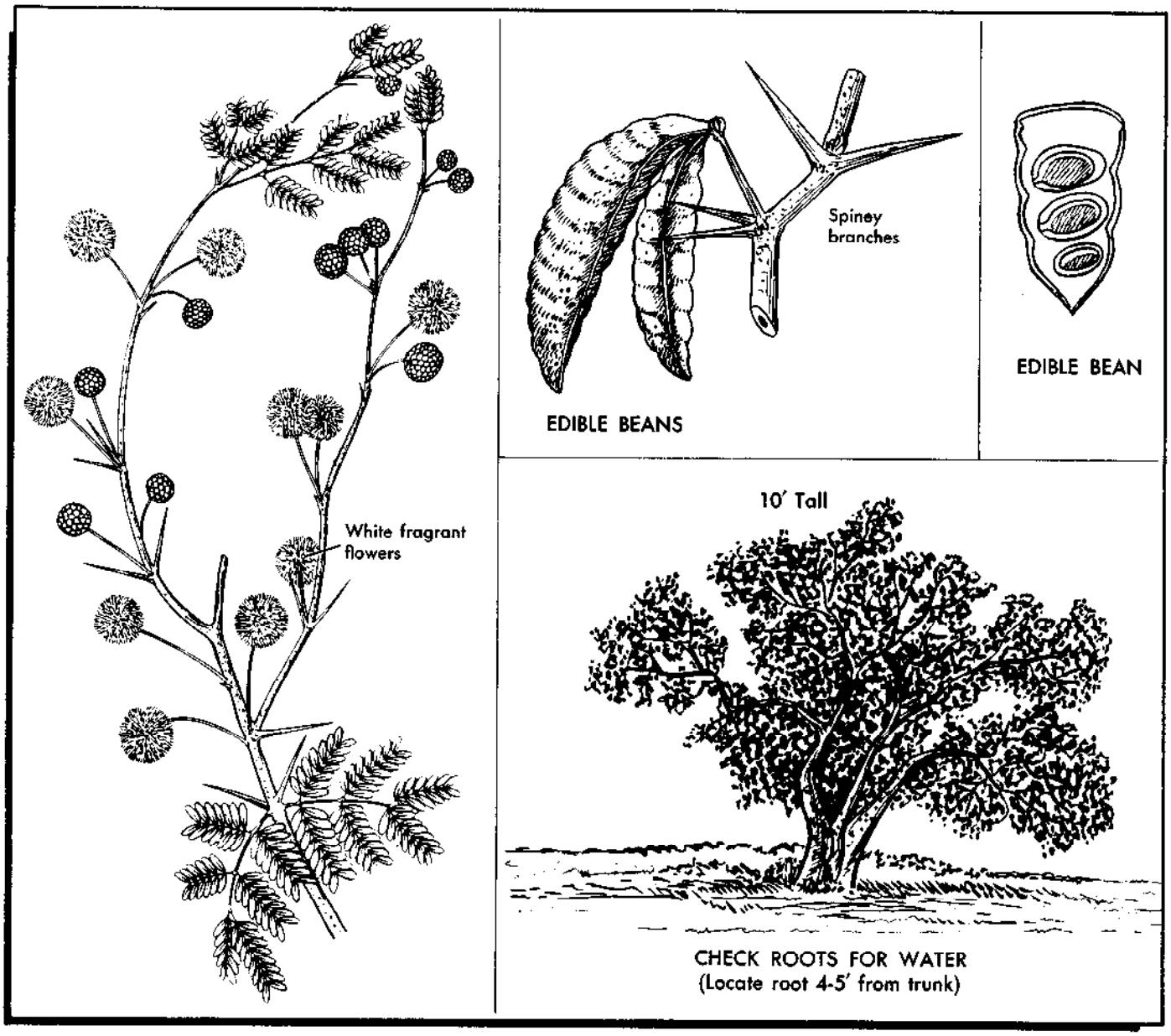

Figure 11-18. Acacia.

latitudes which comprise the temperate zone are 23½° north latitude to 66½° north latitude and 23½° south latitude to 66½° south latitude.

b. There are two main types of climate which comprise the temperate group—mild type, dominated by oceanic or marine climate; and a more severe one called continental climate.

(1) The temperate oceanic climate is the result of warm ocean currents where the westerly winds carrying moisture have a warming effect on the landmass. This oceanic type climate cannot develop over an extensive area on the eastern or leeward side of large continents in the middle latitudes. The extended effect of the ocean climate can be limited by mountain ranges. Such is the case with the Olympic, Cascade, and Rocky Mountains. As the oceanic weather system moves across the Olympic Mountains, it drops nearly 300 inches of precipitation annually. On the windward side of the Cascades, the annual precipitation ranges from 80 to 120 inches annually. In contrast, the region from the leeward side of the Cascades is a relatively dry area, receiving between 10 to 20 inches precipitation annually. As the system moves across the Rocky Mountains, most of the remaining moisture is lost.

(2) The temperate continental climate is a land-controlled climate which is a product of broad middle latitude continents. Because of this, the continental climate is not found in the Southern Hemisphere. This type of climate is very characteristic of the leeward side of mountain barriers and eastern North America and

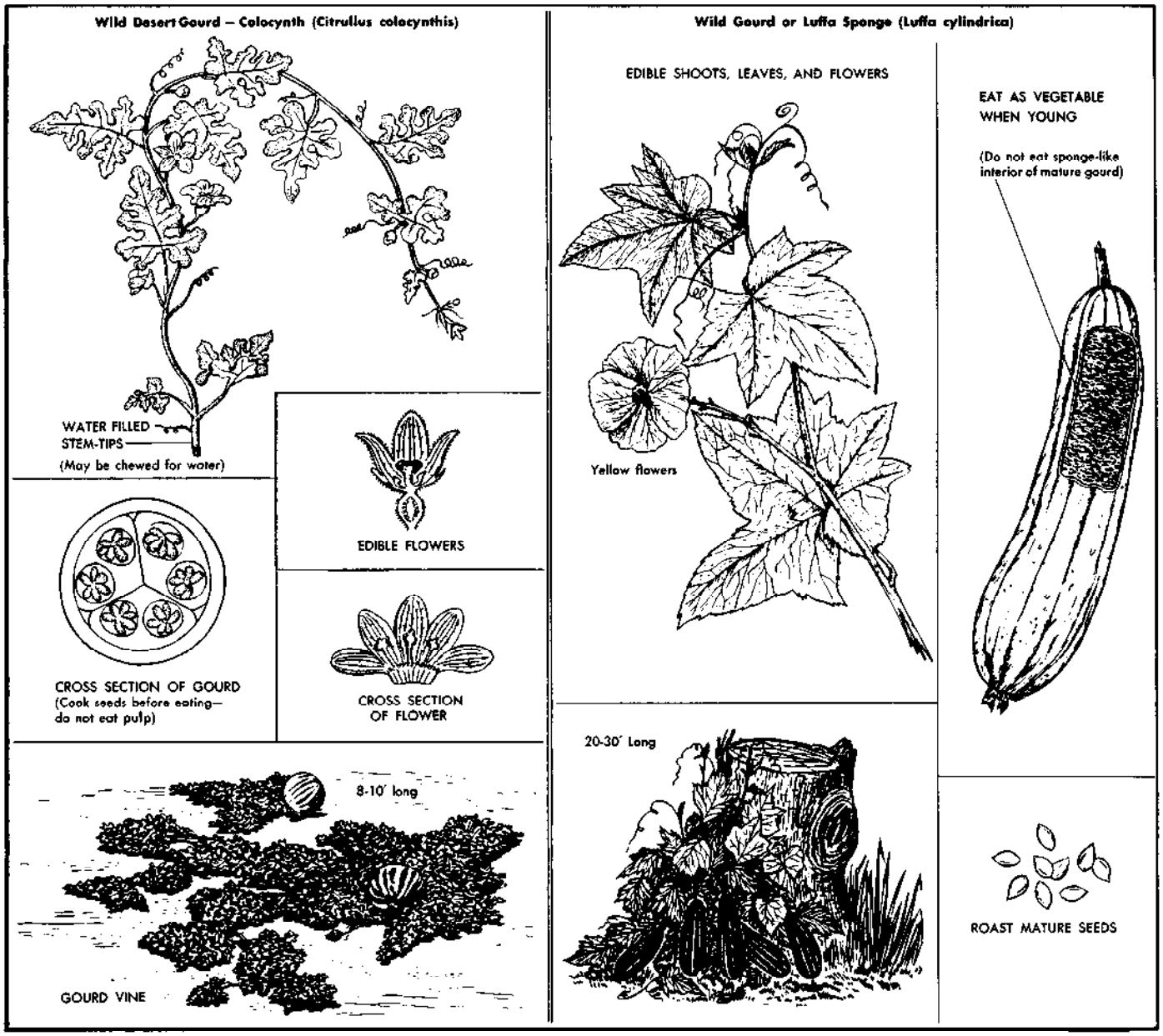

Figure 11-19. Wild Desert Gourd.

Asia. These areas are associated with dry interiors since there are few major warm-water sources available for formation of water systems. The average temperature in the winter and summer are not only extreme but also variable from one year to the next. The severe winter temperature is caused by the polar airflow toward the Equator, and neither winter nor summer temperatures are moderated by the effects of large water masses (oceans).

c. The climate within the temperate zone varies greatly in temperature, precipitation, and wind. The temperate (midlatitude) zone is divided into four major climate zones which are controlled by both tropic and polar airmasses.

(1) The humid subtropical zone is located generally between 20 and 30 degrees north and south latitude. This climate also tends to occur on the east coast of the continents which are at these latitudes. An example of this zone in the United States is the area between Missouri to lower New York and east Texas to Florida. The temperature ranges from 75°F to 80°F in the summer months to 27°F to 50°F in the winter months. The total average precipitation is 30 to 60 inches or more. During the summer months, convectional rainfall is common and thunderstorms frequent. In the winter, the rain is

©National Park Service, U.S. Dept. of the Interior, Photo by M. Woodbridge Williams

Figure 11-20. Marine West Coast Climate (Olympic Peninsula).

more widespread and is usually associated with passing midlatitude cyclones. The wind has a great influence in this area. The area is affected by both the prevailing westerlies and eastern trade winds. During the summer, the winds are influenced by eastern moist maritime airmass flows. Winters are influenced by westerly continental polar airflows. The weather is also influenced by low latitudes. The equatorial current which turns poleward forms warm currents (Gulf Stream and Japanese and Brazilian) that parallel the coasts.

(2) The marine west coast climate (figure 11-20) is sometimes referred to as the temperate oceanic climate. This climate is generally between 40 and 60 degrees north and south latitudes, on the west side of the continent. Examples are the west coasts of Washington to Alaska, Chile, nearly all of Europe, and New Zealand. The summer months are cool with average temperatures of 60°F to 70°F. The winter months are mild with temperatures averaging 27°F to 50°F. The total average rainfall ranges from 20 to 200 inches. Since the maritime climates are under the influence of the westerly winds all year, rainfall is nearly uniform from season to season. These climates are probably more cloudy than any other. They are characterized by widespread stratus and nimbostratus clouds and frequent fog. One of the main reasons for the tremendous rainfall in these climatic areas is the warm ocean currents. These currents yield moisture to the air which is blown inland by the westerly winds (figures 11-21 and 11-22).

(3) Middle latitude desert and steppe climates of complex origins are found generally between latitudes 35 to 50 degrees and in the interior of Asia and North America. Mountain ranges serve as barriers to the moist maritime airmasses, thus resulting in low levels of precipitation. In summer, these interiors generate tropical airmasses, while in winter they are overrun by polar air masses originating in Canada and Siberia. Deserts are also characterized by considerable differences between the average summer and winter temperatures. Of greater importance are the vast semi-dry steppes. Their annual precipitation of 10 to 20 inches supports short-grass vegetation. They comprise the great sheep and cattle ranges of the world; for example, the veldt of South Africa and the American Great Plains support vast numbers of animals.

(4) The Mediterranean climate is sometimes referred to as subtropical dry summer climate. It is generally located from 30 to 45 degrees north and south lati-

©National Geographic Society, Photo by John Scofield

Figure 11-21. Maritime Climate (Southern Dahomey, Africa).

tudes. Examples of this climate occur in the Mediterranean region, most notably Spain, Italy, and Greece. Summer temperatures usually average 75°F to 80°F; but in coastal locations near cool currents, the average is 5°F to 10°F lower. Typical temperature averages for the coldest months are 45°F to 55°F. Coastal locations are usually somewhat warmer in the winter than inland locations. Total annual rainfall is normally 15 to 30 inches along the equatorial margins and increases poleward. This climate is a transitional zone between the dry west coast desert and the wet west coast climate. The westerly winds and cold ocean currents are the controlling influences of the Mediterranean climate. An example of a cold current which affects climates are the Humbolt Current (Peru Current) along the coast of Chile, Peru, and California.

d. Major topographical characteristics found in temperate regions are:

(1) Mountains. Areas of steep slopes with local relief of more than 2,000 feet. Examples of this land form are the Rocky Mountains of North America, the Andes Mountains of South America, and the Himalayan Mountains of Asia.

(2) High Tablelands. Upland surfaces over 5,000 feet in elevation and having local relief of less than 1,000 feet, except where cut by widely separated canyons such as the High Tableland of the Wyoming Basin.

(3) Hills and Low Tablelands. Hill areas having local relief of more than 325 feet but less than 2,000 feet. At the ocean shoreland, however, local relief may be as low as 200 feet. A low tableland is an area less than 5,000 feet in elevation with local relief less than 325 feet, but which (unlike plains) either does not reach the sea or where it does, terminates in a bluff overlooking a low coastal plain. Examples of this terrain can be found in the Appalachian Mountains, Quebec, Southern Argentina.

(4) Plains. Surfaces with local relief of less than 325 feet. On the marine side, the surface slopes gently to the sea. Plains rising continuously inland may attain elevations of high plains—over 2,000 feet. The greatest expanses of plains occur in the center of the North American Continent, Eastern Europe, and Western Asia (figure 11-23).

(5) Depressions. Basins surrounded by mountains, hills, or tablelands which abruptly outline the basins. Examples of depressions can be found in the southwestern United States.

e. There are several biomes of plants and animal life within the temperature zone, and the characteristic life forms are dependent upon climatic characteristics within a specific area. The biomes are named for the plants most plentiful in the area.

(1) Coniferous Forests (figure 11-24). These occur in a broad band across the northern portions of the continents of North America, Europe, and Asia. The northern boundary is the tundra and the southern limits are generally around 50 degrees north latitude. However, this zone extends down to 35 degrees north latitude in the mountainous regions of the western North American continent. This biome corresponds with the humid continental climate, except in the mountainous portions of North America below 50 degrees north latitude. The main life forms in this zone are the conifers or needle leaf, cone-producing trees, such as pines, firs, spruces, and hemlock. In these areas, the trees may grow closer together, not being severely limited by a need for sunlight, and are subject to frequent fires caused by lightning. When this occurs, the ecological succession is reversed, allowing low shrubs to spring up in the burned-over areas. Although the conifer is the predominate tree, there is more subclimax or secondary growth in these biomes than in climax forests (mature or primary forests). In these areas, the pines, alders, aspen, and poplars are the dominant trees. The dominant shrubs are heather, small maples, and yews. If forced to survive in these areas for long periods, especially in winter, the survivor will find that edible food plants are scarce (figure 11-25).

Photo by Daniel Yacko

©National Geographic Society, Photo by Bianca Lavies

Figure 11-22. Red Mangrove and Swamps.

(a) In winter, the primary edible food plants available are:

-1.Rootstalks.
-2.Bulbs.
-3.Roots.
-4.Seeds.
-5.Resins (from pines).
-6.Infusion (teas) from evergreen needles.
-7.Bark (inner part).

(b) During the summer months, many more plants are available for food, including:

-1.Nuts.
-2.Sweet Acacia.
-3.Water Plantain.
-4.Shoots (potherbs).
-5.Wild Apple.
-6.Polypody.
-7.Leaves (potherbs).
-8.Baobab.
-9.Wild Rhubarb.
-10.Pollen (cattail).
-11.Beechnut.
-12.Flowering Bush.
-13.Flowers.
-14.Braken (Fern).
-15.Wild Sorrel.
-16.Fruits (dessert).
-17.Wild Calla
-18.Cattail. (water Arum).
-19.Fiddleheads (Ferns).
-20.Chufa (Nut Grass).
-21.Chestnut.
-22.Wild Crabapple.
-23.Wild Dock.
-24.Chicory.
-25.Wild Filbert (Hazelnut).
-26.Wild Grape.
-27.Juniper.
-28.Common Jujube.
-29.Pine Nuts.
-30.Spreading Wood Fern.
-31.Wild Lily.
-32.English Oak (Acorn).
-33.Tree Fern.
-34.Water Lily (Temperate Zone).

(c) In many places, the ground is thickly covered with mosses and there may be a few varieties of early flowering plants and many berry-bearing shrubs which invite birds and mammals into the open areas. Some of the largest herbivores (plant eaters) live in these evergreen forests—caribou, reindeer, moose, and deer. The small herbivores may include porcupines, several species of squirrels, mice, and rabbits. The carnivores (flesh eaters) which feed upon the smaller animals include black bear, gray wolf, lynx, wolverine, red fox, and weasel. Multitudes of insects provide food for the birds. (These insects may also present a menace to the survivor). A large variety of birds feed not only on insects but also on plants.

(2) Deciduous Forests. Decodipis (broad leaved) forests are found extensively in the eastern portion of the United States; in Europe, between 40 to 50 degrees north latitude; and also in eastern portions of the USSR, China, Korea, and Japan from 35 to 50 degrees north latitude. This biome corresponds with the subtropic and humid continental climatic zones; the area in which any deciduous forest group determines the predominant trees or climax vegetation found there. Here are a few examples: In north central United States, Beech and Maple trees assume the dominant role; in Wisconsin and Minnesota, it is Basswood and Maple; in the eastern and southern regions, the dominant trees are

©National Geographic Society, Photo by George B. Schaller

Figure 11-23. Plains.

Oak and Hickory (figure 11-26). There are also spots in this biome where pines and broadleaf evergreens grow.

(3) Deciduous and Mixed Deciduous-Coniferous Forest:

(a) Deciduous and mixed deciduous-coniferous forests manifest the following characteristics:

-1. Warm summer with rain; winters cold and drier; short drought periods.

-2. Only three stories of vegetation (trees, scrubs, herbs).

-3. Broadleaf trees without leaves in winter.

-4. Mature trees, uniform in height.

-5. Unimpeded view into interior of forest.

-6. Few herbs, ferns, mosses in summer, and abundance of edible fungi in spring and autumn.

-7. Trunks of trees covered with thick-fissured, dark-colored bark.

-8. Resting buds enclosed in hard scaly protecting leaves frequently covered with gum or resin.

-9. For the most part, leaves are thin and delicate, rarely thick and leathery like those of tropical rain forest trees.

(b) The deciduous and mixed deciduous-coniferous forests that predominate over much of eastern United States are typical of this vegetation type. The deciduous forest is wholly temperate in character. By contrast with the tropical evergreen forest with its richly shaded but chiefly dark glossy green canopy, the broadleaved temperate forest extends in a uniformly bright green expanse. The temperate deciduous and mixed deciduous-coniferous forest vegetation type occupies extensive areas in several parts of the world (figure 11-27).

-1. North America. Eastern United States.

-2. South America. Southern Chile, southeastern Brazil.

-3. Europe. Western and northern Europe, southern Scandanavia, southeastern Europe (Balkans).

-4. Asia. South central Siberia, southeastern Siberia and part of Manchuria, Korea (throughout), Japan (throughout), China (throughout except the extreme south and extreme north).

-5. Oceania. New Zealand.

(c) A general characteristic of a climax forest is the stratification of layers of plant growth similar to the canopy systems in the tropical rain forest. In a climax forest, there are usually a limited number of flowering plants, ferns, and shrubs for ground cover. The edible food plants in the vegetation zone are numerous, and a large array of edible species are available and include:

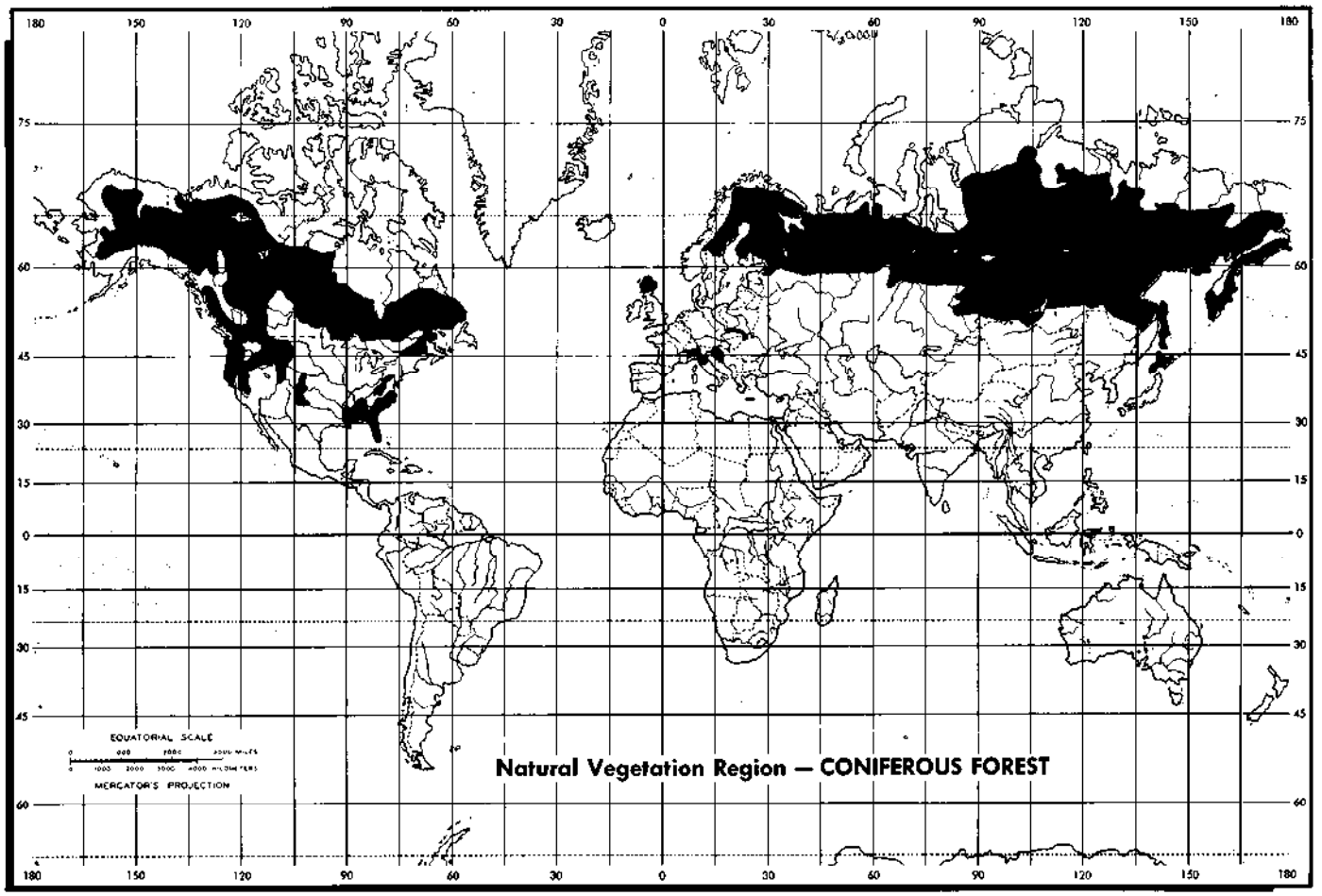

Figure 11-24. Coniferous Forest.

-1.Amarath.
-2.Chestnut, Water (Trapa Nut).
-3.Wild Lily.
-4.Wild Apple.
-5.Chicory.
-6.Lotus Lily.
-7.Beechnut.
-8.Chufa (Nut Grass).
-9.Mulberry.
-10.Braken (Fern).
-11.Wild Filbert (Hazelnut).
-12.English Oak (Acorn).
-13.Wild Calla (Water Arum).
-14.Wild Grape.
-15.Wild Onion.
-16.Cattail.
-17.Common Jujube.
-18.Pine Nuts.
-19.Chestnut.
-20.Juniper.
-21.Polypody.
-22.Air Potato (ubi Tuber).
-23.Tree Fern.
-24.Wild Dock.
-25.Purslane.
-26.Wild Tulip.
-27.Water Plantain.
-28.Wild Rhubarb.
-29.Walnut.
-30.Pokeweed.
-31.Flowering Rush.
-32.Water Lily.
-33.Tropical Yam.
-34.Wild Sorrel.
-35.Wild Crabapple.

(d) Animal life associated with deciduous forests is more varied and plentiful than in evergreen forests, though some animals such as certain species of deer, squirrels, martins, lynx, and wildcats are common in both areas. Wolves, foxes, and other small carnivores (flesh-eating animals) feed mainly on small rodents. Some forest dwellers, such as rodents, dig their dens below the ground while other dens are dug near streams where food and shelter are found. In the aquatic environment, the beaver builds dams for food and shelter. Muskrat, otter, and mink also seek the water's edge, while snakes, turtles, and frogs are found in the streams or lakes.

(4) Steppes and Prairies: (See figure 11-28.)

(a) The part of Russia extending from the Volga River through central Asia to the Gobi Desert has been referred to as the steppes. However, as a vegetation type, the steppe grasslands occur in many other parts of the world. The rainfall in steppe areas averages 15 to 30 inches per year, as compared to prairie areas which average 30 to 40 inches per year. The general aspect of a steppe area, like the prairie, is a broad treeless expanse of open countryside which may be quite rolling in places. The principal steppe areas are:

Photo by Daniel Yacko

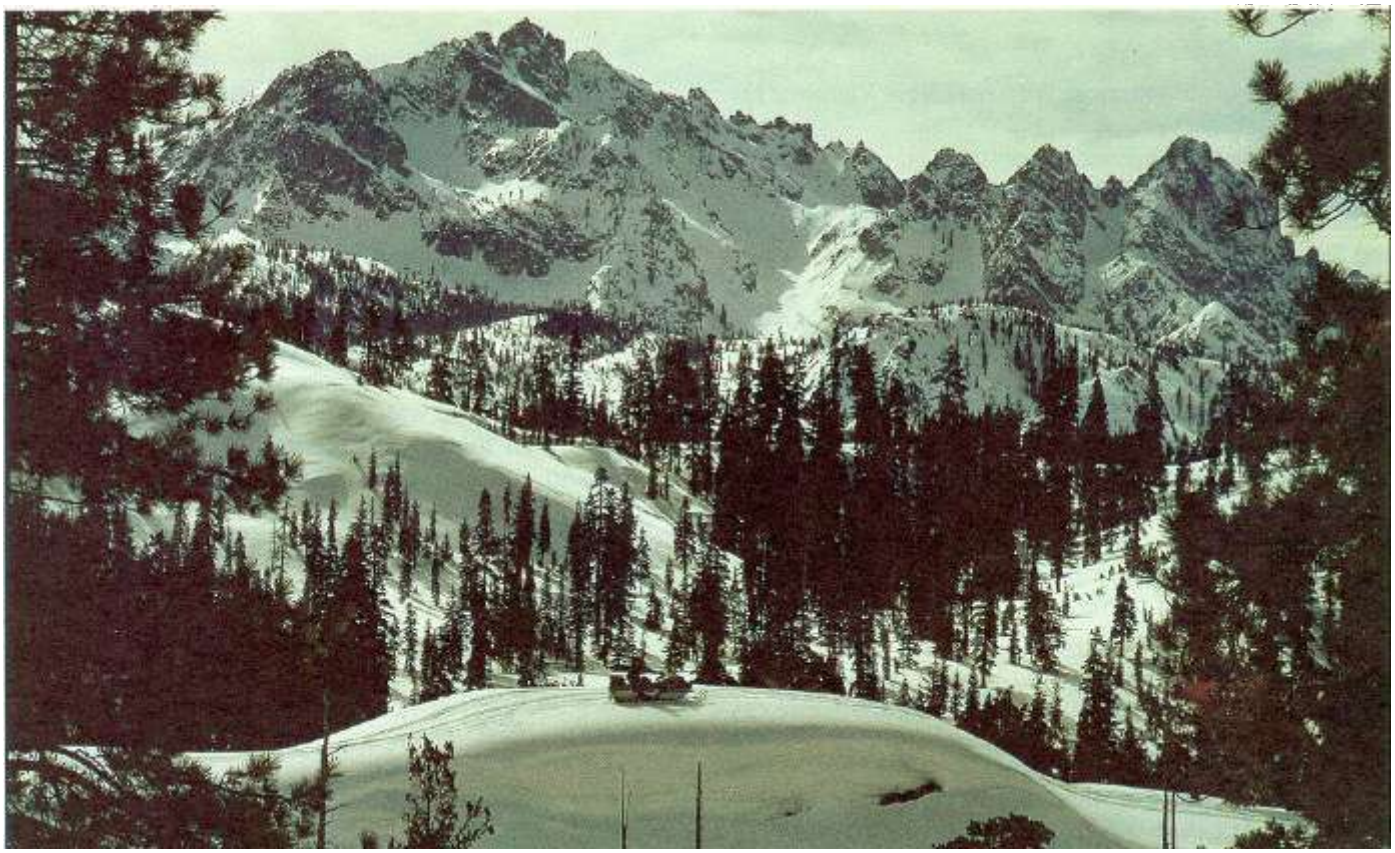

©National Geographic Society, Photo by David Hiser

Figure 11-25. Coniferous Forest.

Photo by Daniel Yacko

Figure 11-26. Deciduous Forest.

-1. North America. Western Great Plains of the United States.

-2. South America. Argentina.

-3. Africa. Narrow belt extending across Africa on the southern rim of the Sahara, and parts of Ethiopia and Kenya.

-4. Europe. Southeastern Russia.

-5. Asia, Turkey, Iran, Baluchistan, Pakistan, Turkestan, and a broad belt through central Asia.

-6. Australia. Fringes of the great central desert, especially in eastern Australia.

(b) The prairie and steppe areas are very closely related. However, the true prairie supports a somewhat different flora than the steppe areas, and for this reason, it is important that they be discussed separately. The chief distinction between prairie and steppe is the seasonal distribution of rainfall.

	PRAIRIE	STEPPE
Rainfall per year	30-40 inches	15-30 inches
Subsoil	Permanently moist	Permanently dry

In both, the precipitation comes during the short growing season (spring). Summers are hot with intermittent showers. The primary prairie regions of the world are:

-1. North America. South central Canada, and east central United States.

-2. South America. Northeastern Argentina, Uruguay, Paraguay, and Brazil.

-3. Africa. Union of South Africa.

-4. Europe. Parts of Hungary, Rumania, and Russia (Ukraine and in a belt extending through central Russia to the Urals).

-5. Asia. Manchuria.

(c) The main plants in these biomes are grasses. Due to different conditions various characteristic grasses grow in specific areas on the prairies. The tall grasses are found near the edges of deciduous forests where larger amounts of water are available. The mid-grasses grow farther west, close to the Great Basin within the United States with short grasses growing in the rain shadows of the mountains. Wild flowers and other annuals are found throughout these regions. On the fringes of the desert, desert plants may have moved into the grasslands. The following are food plants of the steppes:

-1.Sweet Acacia.
-2.Wild Chicory.
-3.Wild Rose.
-4.Amaranth.
-5.Chufa (Nut Grass).
-6.Wild Sorrel.
-7.Baobab.
-8.Wild Dock.
-9.Wild Tulip.
-10.Wild Calla (Water Arum).
-11.Wild Onion.
-12.Water Lily.
-13.Cattail.
-14.Sea Orach.

(d) Common herbivores of the prairies are ground squirrels, prairie dogs, rabbits, gophers, and a great many species of mice. These are preyed upon by

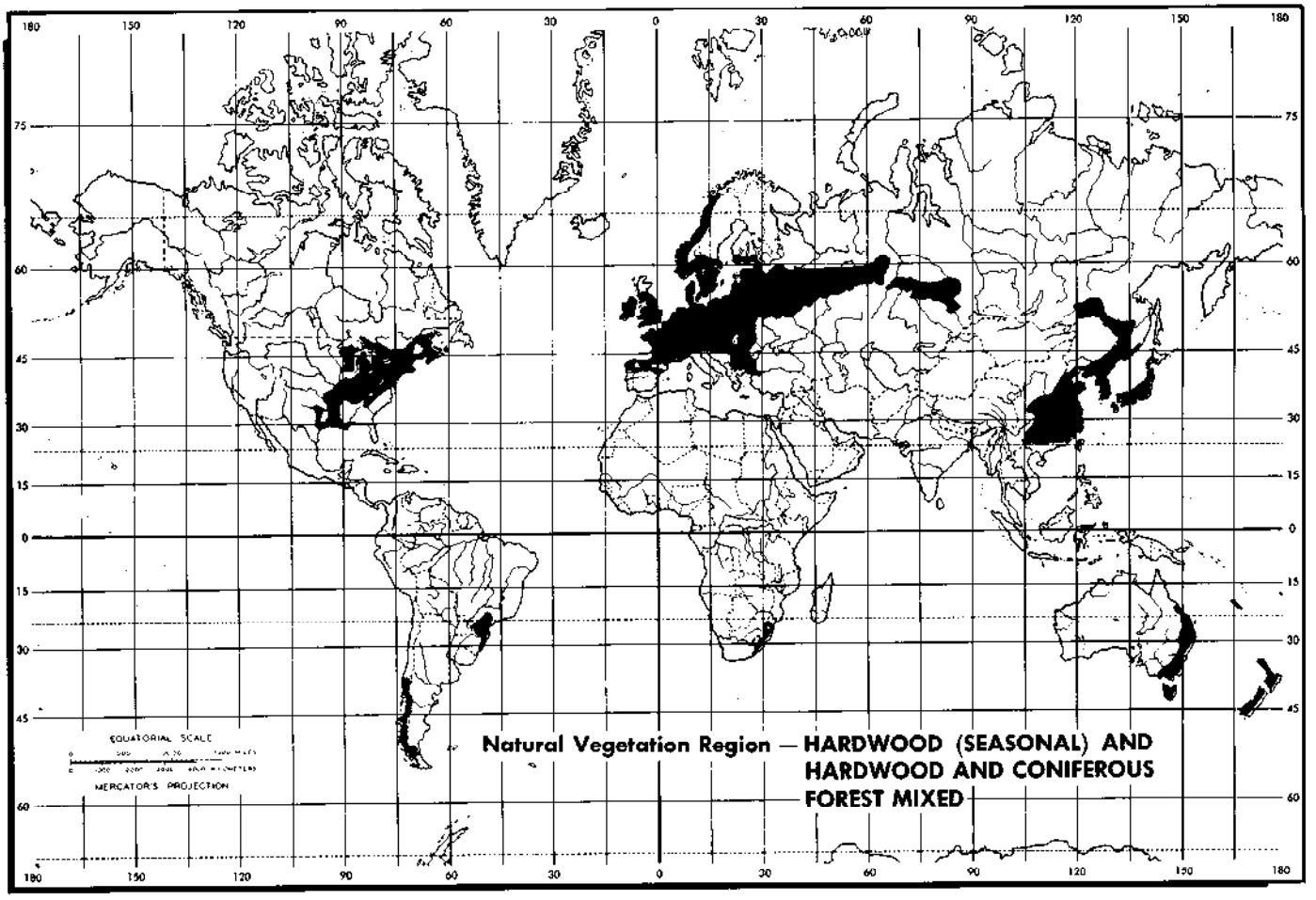

Figure 11-27. Hardwood (Seasonal).

badgers, coyotes, foxes, skunks, and hawks. Prairie animals travel in packs or herds which serve both to protect their individual members and assist in hunting prey. They typically have excellent vision and sense of smell, but their hearing, though keen, may be impaired by the noise of the pack or herd.

(e) A number of birds nest among the grasses. These include the meadowlark, prairie chicken, and grouse. During the dry season, some of these birds migrate to places better suited to raising their young.

(f) Insects like grasshoppers are well adapted to a grassland environment. The natural enemies of such insects are birds and reptiles which in turn become the prey of owls and hawks.

(5) Evergreen Scrub Forests. These biomes occur in southern California, in countries around the Mediterranean Sea, and in southern portions of Australia and correspond with the Mediterranean climate (figure 11-29).

(a) The major life form in this area is vegetation composed of broad-leaved evergreen shrubs, bushes, and trees usually less than 8 feet tall. This vegetation generally forms thickets. Sage and evergreen oaks are the dominant plants in North America in areas with rainfall between 20 and 30 inches. Areas with less rainfall or poorer soil have fewer, more drought-resistant shrubs such as manzanita. Scrub forest vegetation becomes extremely dry by late summer. The hot, quick fires that commonly occur during this period are necessary for germination of many shrub seeds and also serve to clear away dense ground cover. This ground cover is difficult for the survivor to penetrate. The branches are tough, wiry, and difficult to bend. Trees are usually widely scattered, except where they occur in groves near a stream. Usually, both trees and shrubs have undivided leaves. Grasses and brightly colored spring-flowering bulbs and other flowers may also be found. The survivor will find relatively few kinds of edible plant food within the scrub forest. During the growing season—usually only the spring months—the following kinds of plant foods are available:

-1.Agave (Century Plant). (See figure 11-30.)
-2.Wild Dock.
-3.Wild Pistachio.

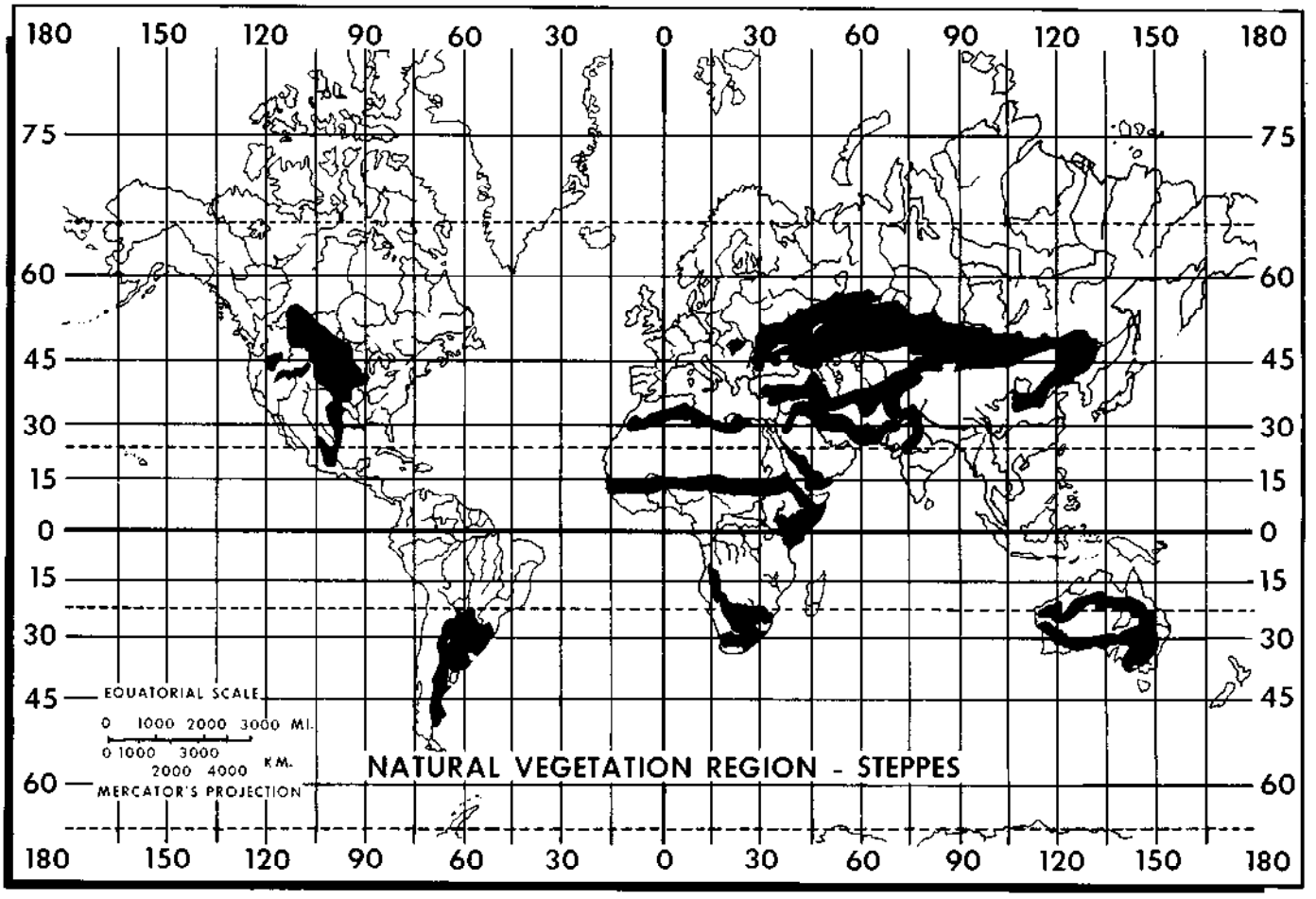

Figure 11-28. Steppes.

-4.Almond.
-5.Wild Grape.
-6.Wild Sorrel.
-7.Wild Apple.
-8.Juniper.
-9.Wild Tulip.
-10.Chicory.
-11.English Oak (Acorn).
-12.Walnut.
-13.Wild Crabapple.
-14.Wild Onion.

(b) Deer and birds usually inhabit these forests only during the wet season, which is the growth period for most scrub forest plants. Small dull colored animals such as lizards, rabbits, chipmunks, and quail are year-round residents.

11-6. Snow Climates. "Snow climates" are defined as the interior continental areas of the two great landmasses of North America and Eurasia that lie between 35 and 70 degrees north latitude. The tree line provides the best natural boundary for a topographical description of the snow climate areas. There are definite differences between the forest area to the south and the tundra to the north in snow-cover characteristics, wind conditions, animal types, and vegetation. Snow climates are comprised of two separate climate types: continental subarctic and humid continental.

a. The continental subarctic climate is one of vast extremes. The temperature may range from -108°F to 110°F. Temperatures may also fluctuate 40 to 50 degrees within a few hours. This area includes several climate subtypes. The largest areas run from Alaska to Labrador and Scandinavia to Siberia. They are cold, snowy forest climates, moist all year, with cool, short summers. A colder climate is found in northern Siberia which has very cold winters with an average cold temperature of -36°F. Another area is found in northeastern Asia where the climate is a cold, snowy forest climate with dry winters. Winter is the dominant season of the continental subarctic climate. Because freezing temperatures occur for 6 to 7 months, all moisture in the ground is frozen to a depth of many feet.

b. The humid continental climates are generally located between 35 and 60 degrees north latitude. For the most part, these climates are located in central and eastern parts of continents of the middle latitudes. These climates are a battle zone of polar and tropical

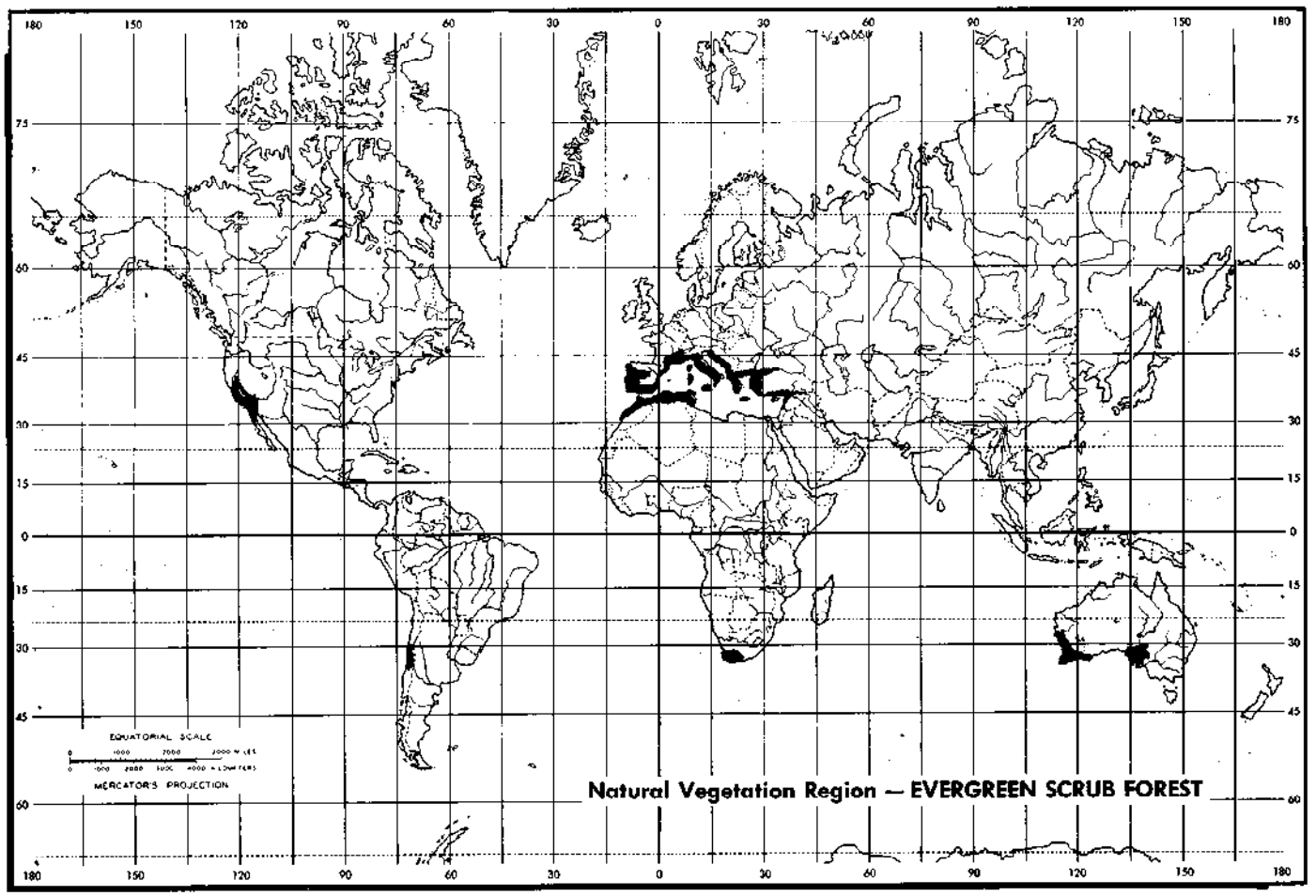

Figure 11-29. Evergreen Scrub Forest.

airmasses. Seasonal contrasts are strong and the weather is highly variable. In North America, this climate extends from New England westward beyond the Great Lakes region into the Great Plains and into the prairie provinces of Canada. This climate can also be found in central Asia. The summers are cooler and shorter than in any other climate in the temperate zone with the exception of the highland (Alpine) subarctic climate. The summer temperatures range from 60°F to 70°F. The winter temperatures range from -15°F to 26°F. The precipitation for the year varies from 10 to 40 inches. A higher percentage of the precipitation is snow, with less snow occurring in areas along the coasts. The weather is influenced by the polar easterly winds and the subtropical westerly winds. The effect of ocean currents on this continental climate is minimal. This climate is dominated by the high- or low-pressure cells centered in interiors of the continent.

c. Both climate regions have seasonal extremes of daylight and darkness resulting from the tilt of the Earth's axis (figure 11-31). Snow climate nights are long, even continuous in winter; conversely, north of the Arctic Circle, the Sun is visible at midnight at least once a year. Darkness presents a number of problems to the survivor. No heat is received directly from the Sun in midwinter, thus the cold reaches extremes. Outside activities are limited to necessity, although the light from the Moon, stars, and auroras, shining on a light ground surface, is of some help. Confinement to cramped quarters adds boredom to discomfort, and depression becomes a dominant mood as time drags on. Fortunately, the period of complete darkness does not last long.

d. The terrain of the snow climate areas coincides with a great belt of needle-leaf forests. This region is found in the higher middle latitudes. Its poleward side usually borders on tundra and its southern margin usually adjoins continental temperate climates. This area is like the tundra because it has poor drainage. As a result, there are an abundance of lakes and swamps. The coastlines vary from gentle plains sweeping down to the ocean to steep, rugged cliffs. Glaciers are a predominate feature of the high altitudes (6,000-feet elevation or above). These glaciers flow down to lower elevations or terminate at the ocean.

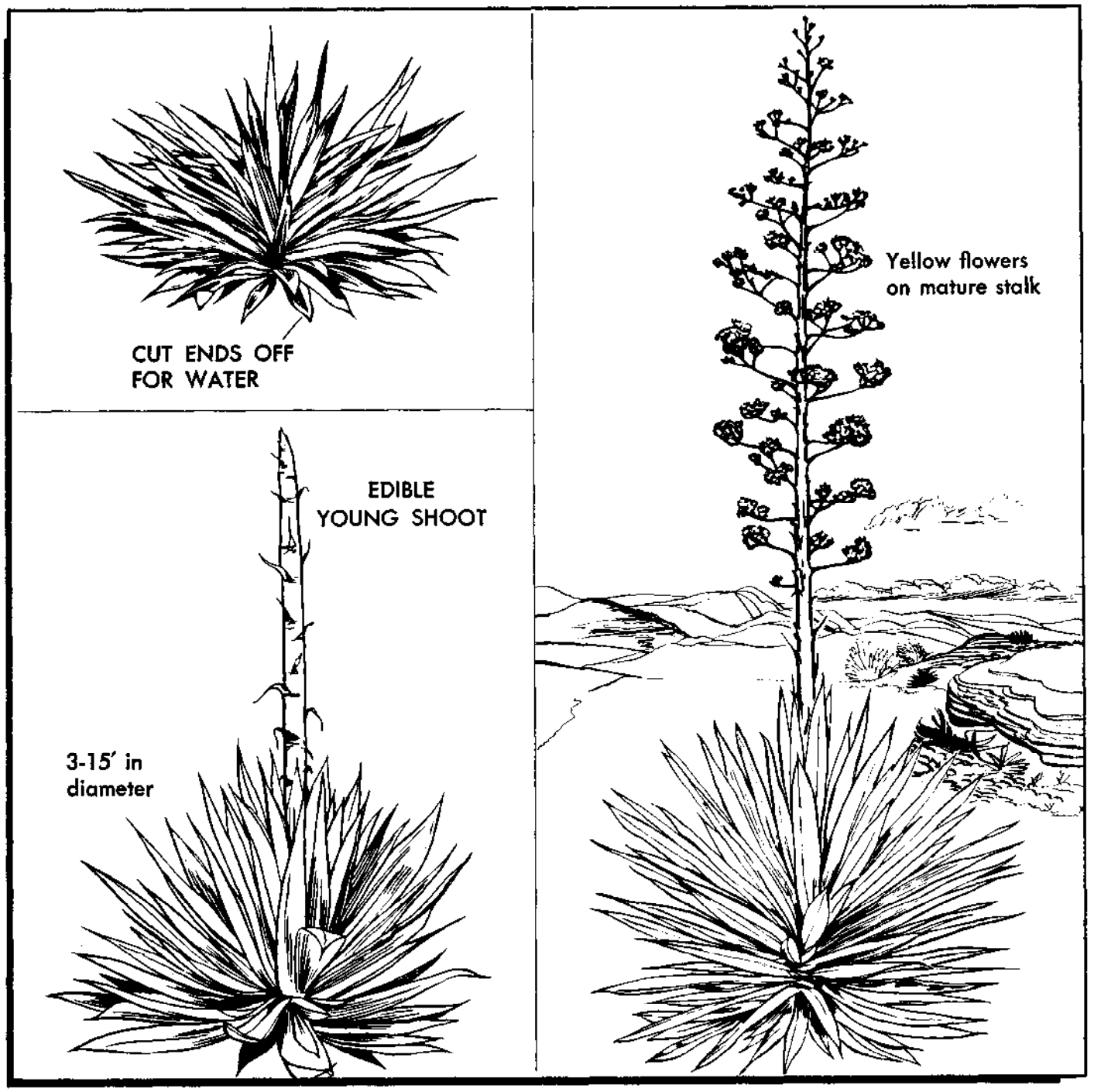

Figure 11-30. Agave (Century Plant).

e. The vegetation is similar to that found in more temperate zones; however, the cold temperatures have caused variations in the physical appearance of the plants. Dark evergreen forests thrive south of the tree line. They consist mainly of cedar, spruce, fir, and pine, mingled with birch. These subarctic forests are called taigas. A transitional zone lies between the taiga and the tundra. In this zone, the trees are sparse and seldom grow over 40 feet tall. Dwarf willow, birch, and alder mix with evergreens, and reindeer moss sometimes forms a thick carpet (figure 11-32).

f. Depending on the time of year and the place, chances for obtaining animal food vary considerably. Shorelines are normally scraped clean of all animals and plants by winter ice. Inland animals are migratory.

(1) Large Arctic Game. Caribou and reindeer migrate throughout northern Canada and Alaska (figure 11-33). In northern Siberia, they migrate inland to al-

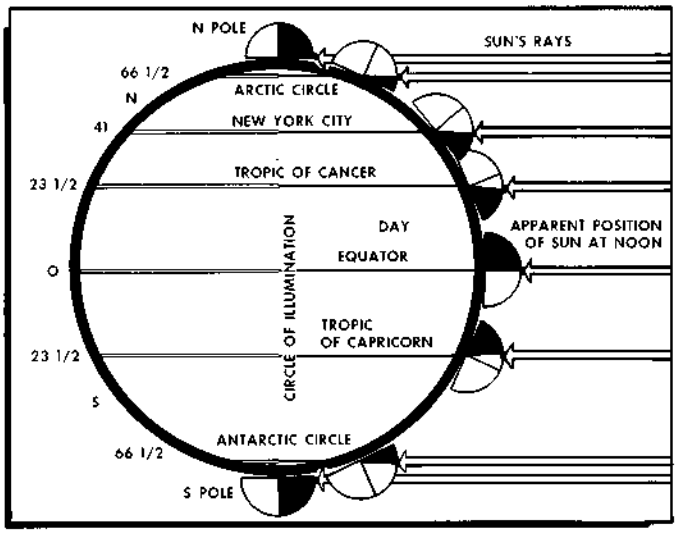

Figure 11-31. Circle of Illumination.

most 50 degrees north latitude. Some are found in west Greenland. All move close to the sea or into the high mountains in summer. In winter, they feed on the tundra. Musk oxen may be found in northern Greenland and on the islands of the Canadian archipelago. Sheep descend to lower elevations and to valley-feeding grounds in the winter. Wolves usually run in pairs or groups. Foxes are solitary and are seen most frequently when mice and lemming are abundant. Bears are dangerous, especially when wounded, startled, or with their young. They generally shun areas of human habitation.

(2) Small Land Game. Tundra animals include snowshoe and arctic hare, lemming, mice, and ground squirrels. They may be trapped or shot in the winter or summer. Most prefer some cover and can be found in shallow ravines or in groves of short willows. Ground squirrels and marmots hibernate in the winter. In summer, ground squirrels are abundant along sandy banks of large streams. Marmots live in the mountains, among the rocks, usually near the edge of a meadow or in deep soil—much like woodchucks. To find the burrow in rocky areas, look for a large patch of orange-colored lichen on rocks. This plant grows best on animal or bird dung, and the marmot always seeks relief in the same spot, not far from a well-hidden entrance.

g. The arctic is the breeding ground for many birds. In summer, ducks, geese, loons, and swans build their nests near ponds on the coastal plains or bordering lakes

Photo by Daniel Yacko

Figure 11-32. Subarctic Forest.

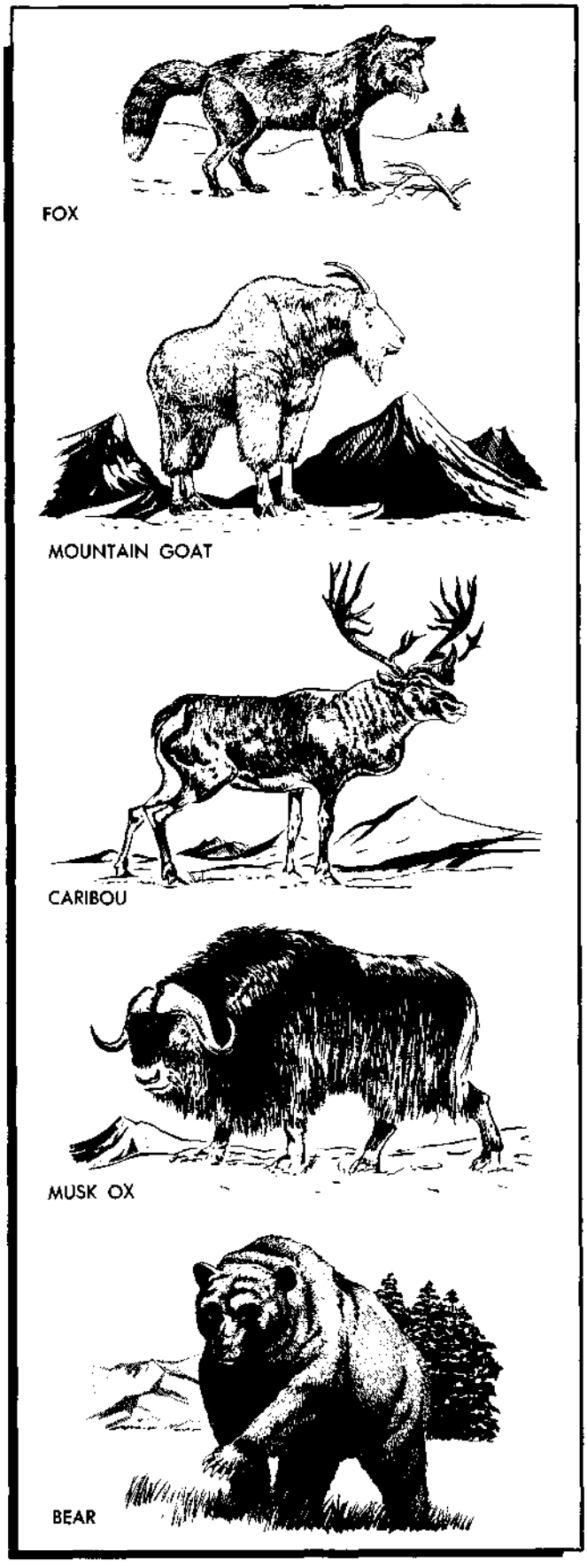

Figure 11-33. Arctic Game.

or rivers of the low tundra. A few ducks on a small pond usually indicates that setting birds may be found and flushed from the surrounding shores. Swans and loons normally nest on small, grassy islands in the lakes. Geese crowd together near large rivers or lakes. Smaller wading birds customarily fly from pond to pond. Grouse and ptarmigan, are common in the swampy forest regions of Siberia. Sea birds may be found on cliffs or small islands off the coast. Their nesting areas can often be located by their flights to and from their feeding grounds. Jaeger gulls are common over the tundra, and frequently rest on higher hillocks. In the winter, fewer birds are available because of migratory patterns. Ravens, grouse, ptarmigan, and owls are the primary birds available. Ptarmigan are seen in pairs or flocks, feeding along grassy or willow-covered slopes.

h. Arctic and tom cod, sculpin, eelpout, and other fish may be caught in the ocean. The inland lakes and rivers of the surrounding coastal tundra generally have plenty of fish which are easily caught during the warmer season. In the North Pacific and in the North Atlantic extending slightly northward into the Arctic Sea, the coastal waters are rich in all seafoods. Varieties include fish, crawfish, snails, clams, and oysters, and one of the world's largest and meatiest crabs—the king crab of the Aleutian Islands and Bering Sea areas. In the spring (breeding season), this crab comes close to shore and may be caught on fish lines set in deep water or by lowering baited lines through holes cut in the ice. Do not eat shellfish that are not covered at high tide. Never eat any type of shellfish that is dead when found, or any that do not close tightly when touched. Poisonous fish are rarer in the arctic than in the tropics. Some fish, such as sculpins, lay poisonous eggs; but eggs of the salmon, herring, or freshwater sturgeon are safe to eat. In arctic or subarctic areas, the black mussel may be very poisonous. If mussels are the only available food, select only those in deep inlets far from the coast. Remove the black meat (liver) and eat the white meat. Arctic shark meat is also poisonous (high concentration of vitamin A).

11-7. Ice Climates. There are three separate climates in the category of ice climates: marine subarctic climate, tundra climate, and icecap climate (figure 11-34).

a. Marine Subarctic Climate. Key characteristics of this climate are the persistence of cloudy skies and strong winds (sometimes in excess of 100 miles per hour) and a high percentage of days with precipitation. The region lies between 50 and 60 degrees north latitude and 45 to 60 degrees south latitude. The marine subarctic climate is found on the windward coasts, on islands, and over wide expanses of ocean in the Bering Sea and the North Atlantic, touching points of Greenland, Iceland, and Norway. In the Southern Hemisphere, this climate is found on small landmasses.

b. Tundra Climate. The tundra region lies north of 55 degrees north latitude and south of 50 degrees south latitude. The average temperature of the warmest month is below 50°F. Proximity to the ocean and persistent cloud cover keep summer air temperatures down despite abundant solar energy at this latitude near the summer solstice (figures 11-35 and 11-36).

c. Icecap Climate. There are three vast regions of ice on the Earth. They are Greenland and Antarctic continental icecaps and the larger area of floating sea ice in the Arctic Ocean. The continental icecaps differ in various ways, both physically and climatically, from the polar sea ice and can be treated separately (figure 11-37).

(1) Greenland. The largest island in the world is Greenland. Most of the island lies north of the Arctic Circle and ice covers about 85 percent of it. The warmest region of the island is in the southwestern coast. The average summer temperature is 50°F. The coldest region is the center of the icecap. The temperature there averages -53°F in the winter.

(2) The Antarctic. The Antarctic lies in a unique triangle formed by South America, Africa, and Australia. Surrounding the continent are portions of the Atlantic, Pacific, and Indian Oceans. The area is almost entirely enclosed by the Antarctic Circle. The climate is considered as one of the harshest in the world. The average temperature remains below 0°F all year. In the winter months, the mean temperature is from -40°F to -80°F. Winter temperatures inland often drop below -100°F. Great storms and blizzards (with accompanying high winds) range over the entire area due to both the continent's great elevation and by being completely surrounded by warm ocean water.

(3) Sea Ice on the Arctic Ocean. Ice on the Arctic Ocean includes frozen sea water and icebergs that have broken off glaciers. This ice remains frozen near the North Pole year around. Near the coast, the sea ice melts during the summer. Currents, tides, and winds may cause it to fold and form high ridges called pressure ridges. One piece of ice may slide over another causing a formation called rafted ice. When the ice breaks into sections separated by water, these sections are called leads. Great explosions and rolling thunder are caused by the breaking and folding of the ice.

d. Terrain. The terrain of the true ice climates encompasses nearly every variation known. Much of the landmass is composed of tundra. In its true form, the tundra

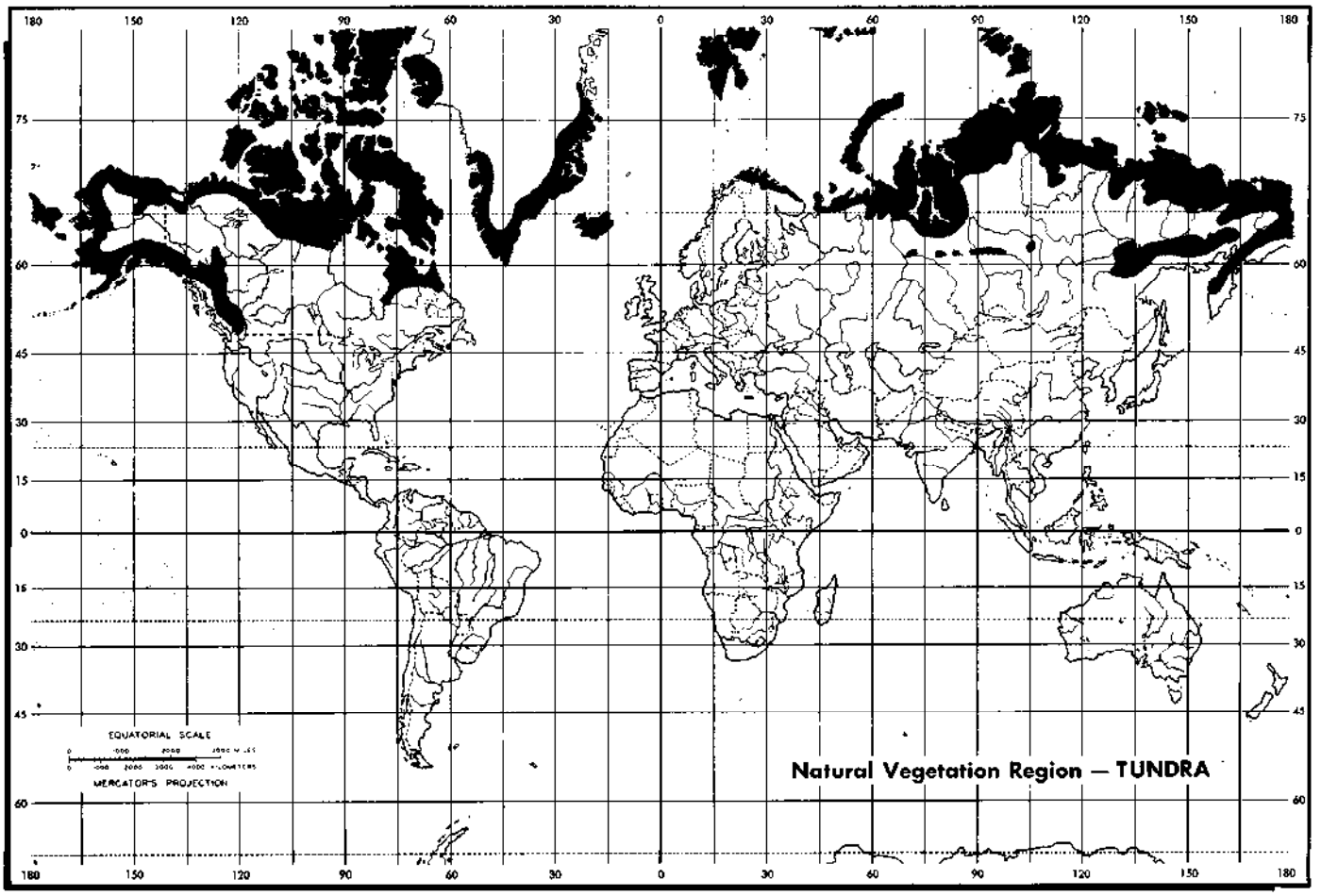

Figure 11-34. Tundra.

Photo by James Tourtillotte

Figure 11-35. Winter Tundra.

is treeless. Vast rugged mountain ranges are found in the area and rise several thousand feet above the surrounding areas. Steep terrain, snow and ice fields, glaciers, and very high wind conditions make this area a very desolate place. Continental glaciers such as the icecaps covering Greenland and the Antarctic continent are large expanses of wind-swept ice moving slowly toward the sea. Ice thickness in continental glacier areas can exceed 10,000 feet.

e. Vegetation:

(1) Shrub Tundra. In Russia, the area surrounding the Lena River is known as a typical shrub tundra environment. Shrubs, herbs, and mosses occur in this zone. Arctic birch predominates but other shrubs occur and several may be useful as supplementary food such as the crystal tea ledum (Labrador tea), willows, and the bog bilberry. In this same shrub zone a lower herbaceous layer occurs which is composed of black crowberry, several grasses, and the cowberry. On the ground, mosses and lichens are present in abundance. The shrubs on the open tundra reach a height of only 3 to 4 feet but in valleys and along the rivers, the same shrubs may reach the height of a person.

(2) Wooded Tundra. The region immediately adjoining the treeless tundra is an extension of the coniferous areas of the south. These subarctic wooded areas include a variety of tree species of which the genus Picea (Spruce) predominates. On the Kola Peninsula of northeastern Scandinavia, these northernmost forests are birch. Siberian spruce occurs between the White Sea and the Urals. Siberian larch occurs between the Urals and the Pyasina River. Dahurian larch occurs between the Pyasina River and the upper reaches of the Anadyr River. In extreme northeastern Asia, Mongolian poplar, Korean willow, and birch are found along the rivers. The trees extending into the tundra are distinguished by their stunted growth (except in river valleys where they reach 18 to 24 feet) and sparseness. Permanent ground frost, or permafrost, penetrates most parts of the true tundra and the northern limits of the forest belt closely coincide with the southern limits of permafrost. A few different plants will cover very large areas so that extensive stands of a single variety of plant are common in the arctic tundra. All tundra plants are small in stature compared to the plants in the warmer climates of more southerly latitudes. The arctic willow and birch, for instance, spread along the ground in the tundra to form

Figure 11-36. Summer Tundra.

large mats. Stunted growth in all the woody plants is the rule, although there are many evergreen plants and hardy bulbous or tuberous plants. Lichens, especially reindeer moss, are widespread in the tundra. As mentioned before, the plant life of the tundra is remarkably uniform in its distribution. Some species are common to all three areas, but other species are more restricted in their distribution. The tundra also contains many species of vegetation found in the forest regions to the south (figure 11-38).

(3) Bogs. The tundra has often been classified as a continuous bog, but this is far from the truth. Many bogs do exist. There are also many hilly and even mountainous areas with considerably drier soil. The moss, or sphagnum bog is less common than the sedge bog. A characteristic of more southern tundras is the development of large peat mounds 9 to 15 feet high and 15 to 75 feet in diameter. These mounds have been formed by ground upheavals caused by freezing water. Many edible plants grow on these bog mounds, such as the cloudberry, dwarf arctic birch, bog bilberry, black crowberry, crystal tea ledum, sheathed cotton sedge, cowberry, and others.

f. Animal Life. Compared to other parts of the world, animal life is poor in species but rich in numbers. Large animals such as caribou, reindeer, and musk oxen migrate through the tundra areas. Carnivores—wolves, foxes, lynx, wolverines, and bears—range through the landmass area and polar bears, seals, walruses, and foxes are found far out on the sea ice. Small animals are the most abundant animal life found and include hares, lemmings, marmots, mink, fishers, and porcupines.

(1) Bird life is very limited during the winter months, mainly owls and ptarmigan, but during the summer months millions of migratory waterfowl nest in the arctic tundra. Species include ducks, geese, cranes, loons, and swans, nesting in and around the swamps, bogs, and lakes of the tundra. The coastal areas are home for many species of sea birds during the summer months. The coastal waters and iceflows are rich in a variety of marine life such as seals, walruses, whales, crustaceans, and fishes.

(2) The freshwater rivers, lakes, and streams are teeming with many varieties of fish—salmon, trout, and grayling. Due to the amount of surface water in the tundra area, there are a large variety of insects. Some 40 to 60 species of mosquitoes, flies, and gnats inhabit the area.

(3) In the Antarctic, animals are virtually nonexistent. Only the lowest forms of animal life can live mainly on mosses and lichens. Marine animals, particularly whales, seals, and penguins, are found along the coastal regions. Sea birds are abundant in the summer and nest on the coastal regions and the islands. There are a few species of wingless insects, lice, ticks, mites, etc., which live off the bird population.

©National Geographic Society, Photo by Peter Schledermann

Figure 11-37. Ice Climate.

11-8. Open Seas. The Köppen-Geiger system of climate classification has been used to describe the environmental characteristics of the landmasses. However, this system is not used to categorize the largest area of the world—the oceans. They are simply divided by their names and locations. All limits of oceans, seas, etc., are arbitrary, as there is only one global sea. The terms "sea" and "ocean" are often used interchangeably in reference to saltwater. However, from a geographic point of view, a sea is a body of water that is substantially smaller than an ocean or is part of an ocean.

a. The seas cover 70.8 percent of the Earth's surface. The waters are not evenly distributed, covering 61 percent of the surface in the Northern Hemisphere and 81 percent in the Southern Hemisphere. Traditionally, the seas are divided into four oceans: Atlantic, Pacific, Indian, and Arctic. These, with their fringing gulfs and smaller seas, make up the world's seas. If the land features of the Earth were smoothed out, the seas would cover the entire globe to a depth of 12,000 feet. Mount Everest, the tallest mountain peak (29,028 feet), would disappear without a trace in the 37,800-foot deep Marianas Trench in the western Pacific Ocean. The sea floor is made up of mountains, valleys, great plains, and deep trenches. The deepest trenches and tallest mountains are found in the north Pacific. The sea floor features do, to some extent, influence the surface properties of the seas; that is, currents, waves, and tides.

b. The average salinity of the seas is usually taken as 3.5 percent. Higher values occur at or near the surface in areas where high temperatures and strong, dry winds favor evaporation. The highest salinities occur in semi-landlocked seas at mid-latitudes such as the Red Sea, the Persian Gulf, and the Mediterranean Sea. The Pacific Ocean is the largest and is over twice the size of the Atlantic or Indian Oceans. Its size allows for greater climatic variations and more widespread influence. Due to similar latitudinal references the Atlantic and Pacific Oceans have many similar characteristics. The Indian Ocean is slightly smaller than the Atlantic, but is more significantly influenced by a continental landmass than

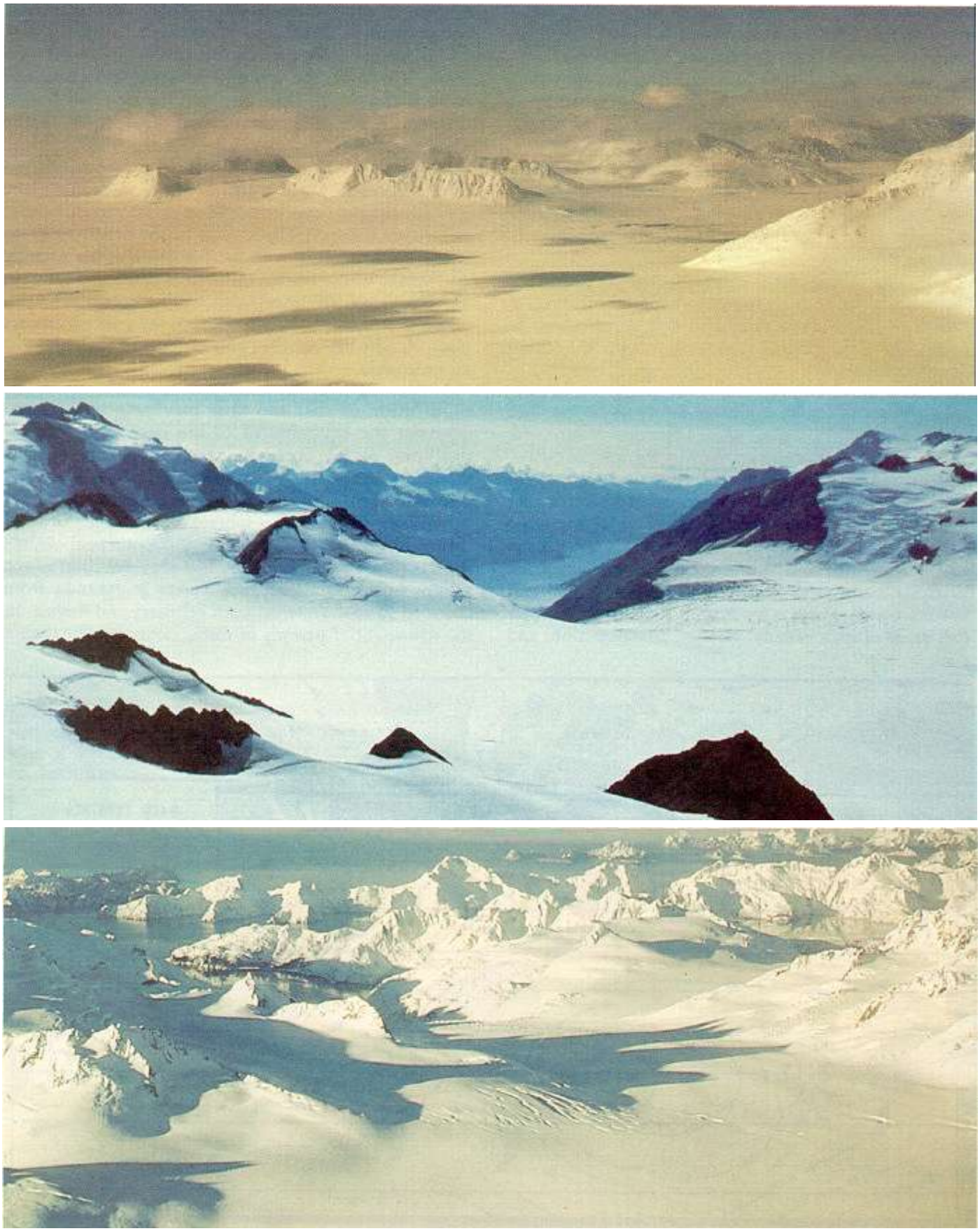

Figure 11-38. Arctic Zone Composite.

any other ocean. The Arctic Ocean is generally recognized as that body of water which lies north of 75 degrees latitude and is nearly enclosed by landmasses.

c. Within each of these four major oceans, numerous subdivisions known as seas, may be geographically aligned along indistinct boundaries (island chains; geography of ocean floor). Examples are:

(1) The Coral Sea is an arm of the South Pacific Ocean lying east of Queensland, Australia, and west of New Hebrides and New Caledonia. It extends from the Solomom Islands on the north to the Chesterfield Islands on the south.

(2) The Bering Sea is located between Alaska and Eastern Siberia, with its southern boundary formed by the arc of the Alaskan Peninsula and the Aleutian Islands. The Bering Strait connects it with the Arctic Ocean to the North.

d. Many water bodies are partially enclosed by land and are known as gulfs. An example would be the Gulf of Mexico.

11-9. Ocean Currents. The ocean has a complex circulation system made up of a variety of currents and countercurrents. These currents move at a rate from barely measureable to about 5.75 miles per hour. They may be relatively cold or warm currents and influence the climate and environment that exists on land and over the ocean. There is a constant movement of water from areas of high density, salinity, concentration, and pressure to areas of low density, salinity, concentration, and pressure in an attempt to establish an equilibrium. These factors influence the movement of ocean currents. However, the primary influence on ocean currents is the wind. They also may be diverted by the Coriolis Force and Continental Deflection (figure 11-39).

11-10. Climatic Conditions. To fully understand the general climatic conditions and seasonal variations that exist over the global sea, each major ocean must be examined separately, with the exception of the Atlantic and Pacific whose similar latitudinal references result in like characteristics (exceptions will be noted). The two physical phenomenons which have the greatest impact upon climate are currents and systems of high and low air pressures.

a. Currents with their basic characteristics of being either warm or cold and their inevitable convergence influence the environment of the open seas. Equally significant as the ocean influence on typical weather sequences (for example, temperature, wind, precipitation, and storms) are semi- and quasi-permanent centers of high and low atmospheric pressures. To observe their effect on climate, imagine a hypothetical voyage from the Pole to the Equator. The southern limit of the solidly frozen arctic icepack varies in latitude from about 65 to 75 degrees between February and August. In the winter, brief periods of calm, clear weather with a

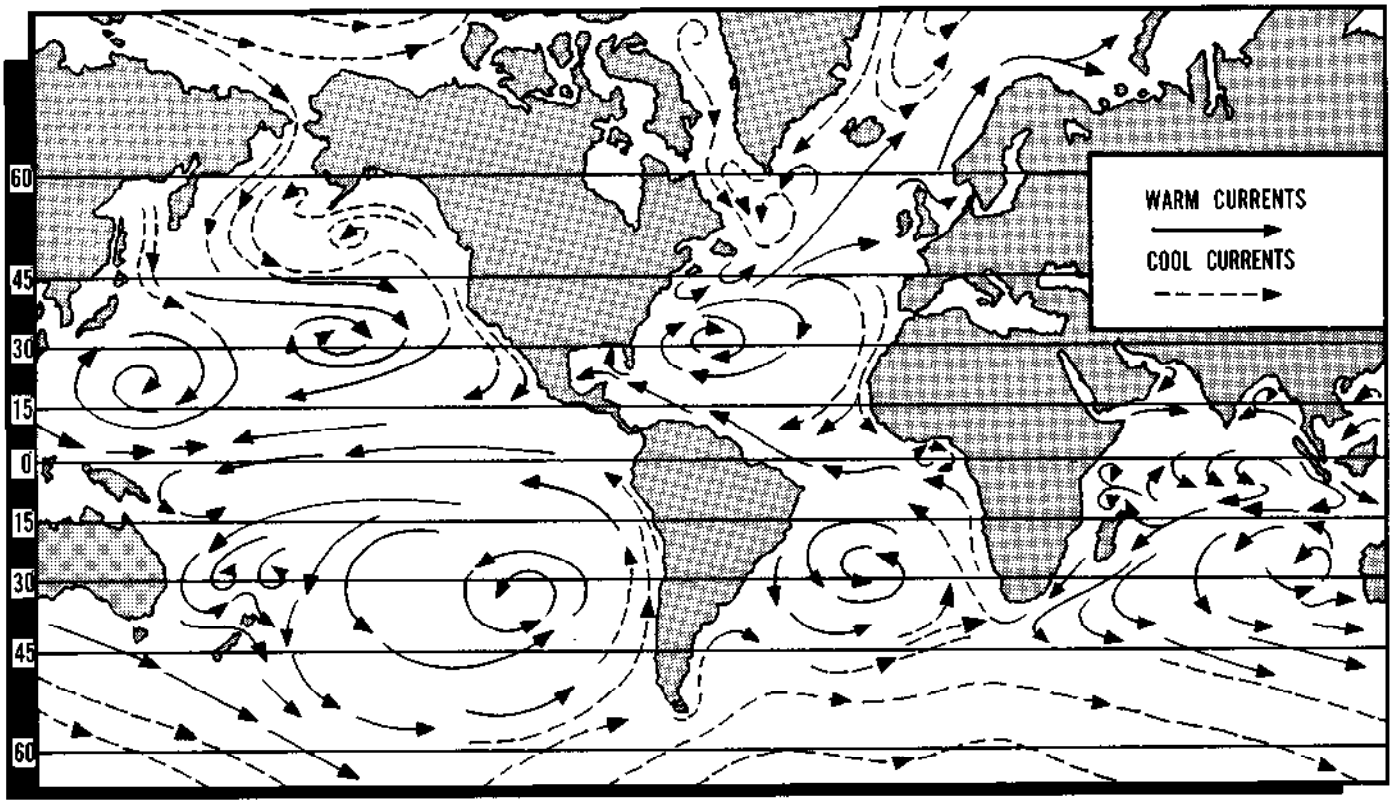

Figure 11-39. Ocean Currents.

mean temperature of -5°F are interspersed between passages of cyclonic storms characterized by snow, winds 30 to 40 mph, temperatures from -20°F to -30°F, and gale-force winds 30 percent of the time. In the summer, frequent periods of several days of calm or light variable winds with temperatures in the mid-40's may be experienced. Skies are uniformly overcast with layers of stratus or nimbostratus clouds. Dense fog banks are prevalent during calms. Rain or drizzle may continue for weeks at a time. One of the most stormy regions in the hemisphere is in the middle of the prevailing westerlies at 50 degrees north latitude. In the winter, calms are rare with winds of 15 to 20 knots and temperature near freezing. Every 2 to 3 days, a pale sun and scattered clouds give way to cumulostratus clouds and rain squalls. Wind intensity may reach 50 to 60 knots with temperatures dropping to -10°F to -15°F as the rain turns to sleet, soft snow, or hail. In the summer, protracted periods of fog, low stratus clouds, and drizzle exist with moderate breezes. The weather improves in the fall with a week or so of calm, clear weather in late September. As we move south to 40 degrees north latitude and the horse latitudes, the semipermanent high-pressure centers result in generally fair, clear weather with a tendency toward dryness. In the winter, temperatures hover near 50°F and summer brings temperatures into the 70's with calms existing one-fourth of the time. Below 25 degrees north latitude, in the heart of the trade wind belt, winds of 5 to 15 mph are normal. Endless bands of cumulus clouds and clear sky exist with little difference between summer and winter. Daytime temperatures range from 70°F to 80°F.

b. In the Atlantic, Pacific, and Indian Oceans between 5 degrees north latitude and 5 degrees south latitude, an equatorial trough of low pressure forms a belt where no prevailing surface winds exist and is known as the doldrums. Instead, the lack of extreme pressure gradients result in shifting winds and calms which exist as much as one-third of the time. Intense solar heating results in violent thunderstorms associated with strong squall winds. The convergence of these equatorial winds and trade winds from the intertropical front can be seen at a great distance because of towering cumulus clouds rising to 30,000 feet.

c. In the vicinity of the intertropical front, heavy convective showers are quite common. Across the Atlantic and Eastern Pacific, the front is usually north of the Equator. Over the western Pacific, west of 180-degrees longitude, the doldrum belt oscillates considerably. Areas north of the Equator receive their heavy rainfall from June to September. Areas south of the Equator receive their heaviest precipitation between December and March. The meteorological sequence described above may be interrupted by periods of extreme weather centered around low pressure.

d. Waterspouts are the marine equivalent of tornadoes attached to the base of a cumulus or cumulonimbus cloud. They are common off the Atlantic and Gulf coasts of the United States and along the coasts of Japan and China during any season. They are usually seen around noon when solor heating is the greatest. They are small in diameter (10 to 100 yards) and short in duration (10 to 15 minutes). Waterspouts generally exhibit less intensity than overland tornadoes.

e. Hurricanes and typhoons are synonyms for tropical cyclones whose maximum winds exceed 75 mph. They occur in the warm western sectors of all oceans during summer and fall. Winds may reach 170 to 230 mph. The lifespan of tropical cyclones ranges from 1 to 2 weeks. In the middle and high latitudes, extratropical cyclones contrast with tropical cyclones in several ways. There is no warm, clear eye, but rather, a cold region of heavy precipitation. Sustained winds are more moderate (70 to 80 mph). Extratropical cyclones may persist for 2 to 3 days at a fixed location.

f. All ocean currents have a profound influence on climate since the properties of the surface largely determine the properties of the various airmasses. The following are a few examples.

(1) The cold water of the Peru or Humbolt currents has a tremendous affect on the climate of Peru and Chile. The cold air that lies over the current is warmed as it reaches land, increasing its capacity to hold moisture. The warm air does not give up the moisture until it passes over the high Andes Mountains. This accounts for the dry climate of the coast of Chili and Peru and a more temperate climate toward the Equator than is usually found in the lower latitudes.

(2) Where the Labrador current contacts the warm gulf stream, fog prevails and steep temperature gradients are present. The northeast coast of North America has much colder climates than the west coast of Europe at the same latitude.

(3) The warm gulf stream current accounts for the continually warm and pleasant weather in the Caribbean Sea and the Gulf of Mexico.

(4) The winds blowing off the warm water of the Norwegian and east and west Greenland currents account for the unusually mild climates in northern Europe. At the same latitude elsewhere, the temperatures are usually much colder.

11-11. Life Forms. Life forms in the seas range from one-celled animals (protozoan) to complex aquatic mammals. The fish and aquatic mammals rule the sea and are of the most concern to anyone in a survival situation on the open seas. The majority of fishes and mammals can be used as food sources, but some must be considered as a hazard to life; such as, sharks, whales, barracudas, eels, sea snakes, rays, and jellyfish.

a. Sharks. (See figure 11-40.)

(1) Most sharks are scavengers, continuously on the move for food. If none is available, they lose interest and swim on. Even in warm oceans where most attacks

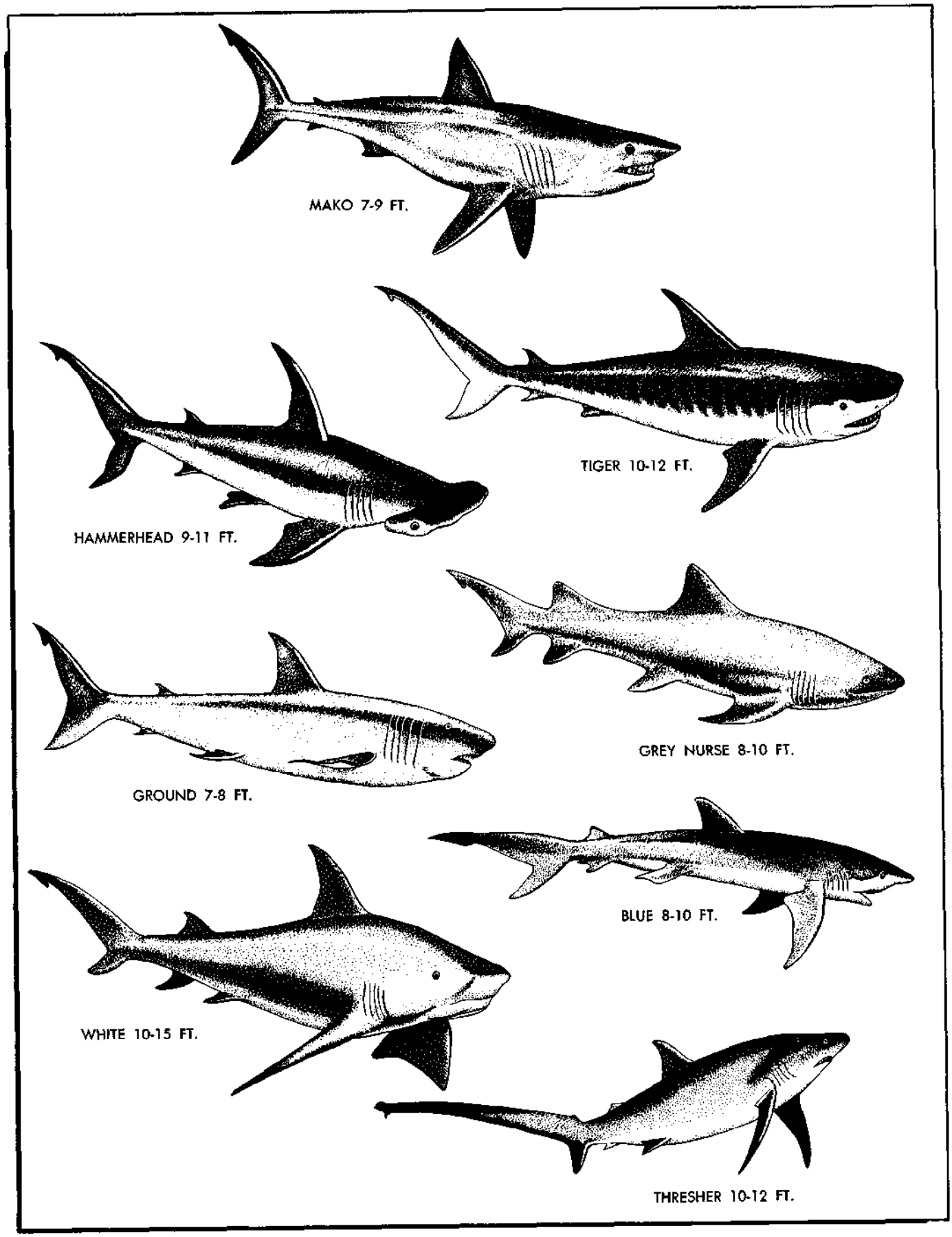

Figure 11-40. Sharks.

occur, the risk can be reduced by knowing what to do and how to do it. Sharks live in almost all oceans, seas, and in river mouths. Normally, there isn't a shark problem in areas of colder water due to the temperature of the water decreasing swim activities. Sharks vary greatly in size, but there is no close relationship between the size of a shark and the risk of attack.

(2) Hungry sharks sometimes follow fish up to the surface and into shallow waters along the shore. When sharks explore such waters, they are more likely to come in contact with people. Sharks seem to feed most actively during the night and particularly at dusk and dawn. After dark, they show an increased tendency to move toward the surface and into shore waters. Evidence indicates that a shark first locates food by smell or sound. Such things as garbage, body wastes, and blood probably stimulate the desire for food. A shark is also attracted by weak fluttery movements similar to those of a wounded fish. While a shark will investigate any large floating object as a possible food source, it probably will not attack a human unless it is hungry. Often the shark will swim away after investigating. At other times, it may approach and circle the object once or twice, or it may swim close and nudge the object with its snout. When swimming, a shark cannot stop suddenly or turn quickly in a tight circle. A shark rarely jumps out of the water to take food; however, it may grasp its prey near the surface. For this reason, people on rafts are relatively safe unless they dangle their hands, arms, feet or legs in the water.

(3) Individuals on or in the water must keep a sharp lookout for sharks. Clothing and shoes should be worn. If sharks have been noticed, survivors must be especially careful of the methods in which body wastes are eliminated and must avoid dumping blood and garbage. Vomiting, when it cannot be prevented should be done into a container or hand and thrown as far away as possible.

(a) If a group in the water is threatened or attacked by a shark, they should bunch together, form a tight circle, and face outward so an approaching shark can be seen. Ward off attack by kicking or stiff-arming the shark. Striking with the bare hand should be used only as a last resort; instead, survivors should use a hard and heavy object.

(b) Individuals should stay as quiet as possible and float to save energy. If it is necessary to swim, they should use strong, regular strokes, not frantic irregular movements.

(c) When alone, swimmers should stay away from schools of fish. If a single shark threatens at close range, the swimmer should use strong, regular swimming movements. Feinting toward the shark may scare it away.

(d) The survivor should not swim directly away from the shark, but face the shark and swim to one side, with strong rhythmic movements.

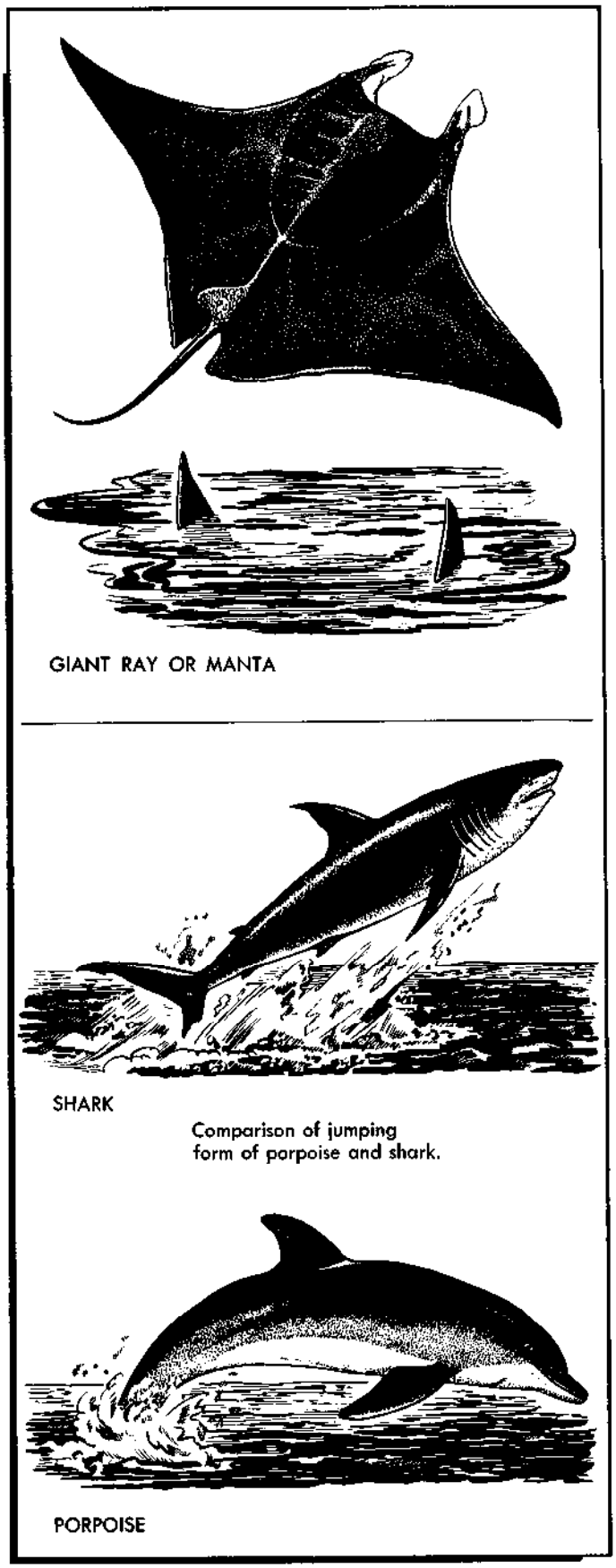

Figure 11-41. Animals Sometimes Mistaken for Sharks.

(c) If a shark threatens to attack or damage a raft, jabbing the snout or gills with an oar may discourage it. Check for sharks around and under the raft before going into the water.

(4) Other animals are sometimes mistaken for sharks.

(a) A school of porpoises or dolphins gracefully breaking the surface, blowing and grunting, may look alarming. Actually, it should be a reassuring sight, because porpoises and dolphins are enemies of sharks. Porpoises and dolphins are harmless to humans (figure 11-41).

(b) Giant rays or mantas, which also appear in tropical waters, may be mistaken for sharks. A swimming ray curls up the tips of its fins, and when seen from water level, the fins somewhat resemble the fins on the backs of two sharks swimming side by side. In deep water, all rays are harmless to swimmers; however, some are dangerous if stepped on in shallow waters (figure 11-41).

b. Grouper or Sea Bass. These fish do not constitute the same degree of hazard as sharks; however, these carnivorous fish are curious, bold, and have a never-ending appetite. Sea bass are most commonly found around rocks, caverns, old wrecks, and caves. Stay away from these areas.

c. Killer Whales. The killer whale has the reputation of being a fearless, ruthless, and ferocious creature. These fast swimmers are found in all oceans and seas, from the tropics to both polar regions. If encountered, a survivor can be assured there are others nearby since they hunt in packs of up to 40 creatures. They have been known to attack anything that swims or floats. If an initial attack is survived, get out of the water. The raft may afford some protection, but they have been known to come up under iceflows and knock other animals into the water. Stay out of the water. On thin ice, do not stand near seals, etc., as the whale may mistake the human form for a seal. However, the probability of being attacked by a killer whale is slim. If an aircrew member is attacked, it will probably be due to the fact that this intelligent whale simply mistook the person for its regular diet.

d. Barracuda. There are 20-odd species of barracuda; some are more feared in certain parts of the world than are sharks. If survivors come down in any tropical or subtropical sea, they may encounter this fish. Barracuda are attracted by anything which enters the water and they seem to be particularly curious about bright objects. Accordingly, survivors should avoid dangling dogtags or other shiny pieces of equipment in the water. Dark colored clothing is also best to wear in the water if no raft is available.

e. Moray Eels. If attacked by some species of moray eel, the survivor may have to cut off its head since some eels will retain their sharp crushing grip until dead. The knife used to do this should be very sharp since their skin is tough and difficult to cut. Their bodies are very slippery and hard to hold. A survivor is most likely to come into contact with a moray eel when poking into holes and crevices around or under coral reefs. Use caution in these areas.

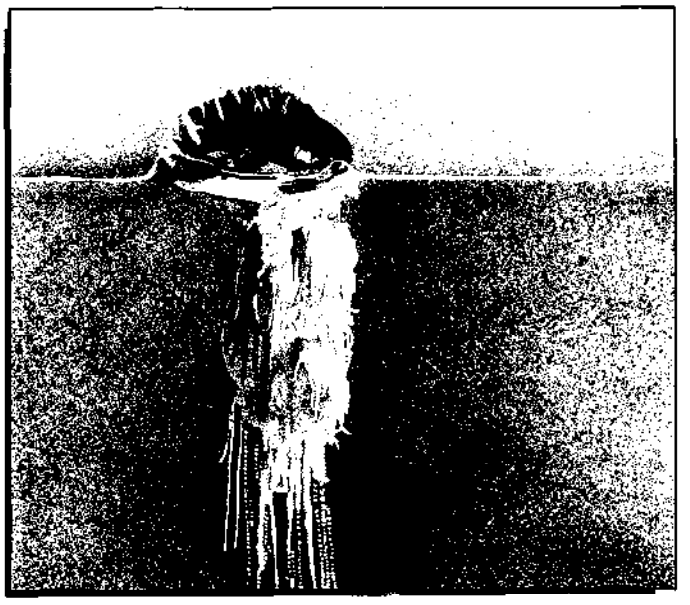

Figure 11-42. Portuguese Man-of-War.

f. Poisonous and Venomous Marine Animals (Invertebrates). There are many marine animals that have no backbone and can inflict injuries by stinging. Three major categories of invertebrates are important to the survivor.

(1) Coelenterates. This group includes jellyfish, hydroids, sea anemones, and corals. Coelenterates are all simple, many-celled organisms. They all possess tentacles equipped with stinging cells or nematocysts in addition to other technical characteristics. The family of coelenterates is divided into three major classes.

(a) Hydrozoan Class. Two of the more common members of this class are:

-1. Stinging or Fire Coral. This false coral can be found in areas of true coral reefs in warm waters.

-2. Portuguese Man-of-War or Blue Bottle. This hydroid is frequently mistaken for a true jellyfish. It is almost always found floating at the surface of the water (figure 11-42). Its stinging tentacles may extend several yards below the surface. Their float is 5 to 10 inches in length. Each tentacle may contain thousands of stinging cells. When one considers the large number of such tentacles, it is apparent that the fishing filaments of the Blue Bottle are quite a formidable venom apparatus.

(b) Anthozoa Class:

-1. Corals. Elkhorn coral and stony coral are very adaptable and have a real immunity to predators.

This helps explain why they tend to dominate reef communities. Corals are carnivores and with the use of small tentacles, capture and consume living zooplankton. Survivors should treat coral cuts by thoroughly cleaning the wound and removing any coral particles. Some coral cuts have been helped by painting them with an antiseptic solution of tincture of iodine.

-2. Sea Anemones. The sea anemone is one of the most plentiful marine creatures, with well over 1,000 species. They can be found from tide level to depths of more than 7,900 fathoms in all seas. Their size ranges from very small, (less than an inch) to over 2 feet in diameter. They eat fish, mollusks, crustaceans, and other invertebrates. Most of the stinging cells of the sea anemones are located on the outer ring of the tentacles.

(c) Scyphozoa Class—Jellyfish. There are many and varied species of jellyfish distributed throughout all seas. Their size ranges from extremely small to a diameter of 6 feet with tentacles hanging below to a depth of 100 feet. All are carnivorous. Some are transparent and glassy while others are brilliantly colored. Regardless of their size and color, they are very fragile creatures which, for the most part, depend on wind and tidal currents to help them move. Most adults can swim but this ability is weak. Whether they stay on the surface or under the surface, and to what depth, varies with each species. The stinging cells of jellyfish are located in the tentacles.

(d) Venom Apparatus of Coelenterates:

-1. All of the coelenterates have stinging cells or nemotocysts located on the tentacles. Each of these cells is like a capsule. If the survivor comes into contact with the capsule, part of it springs open and a very sharp, extremely small "thread"-type tube appears. The sharp tip of the tube penetrates the skin and the venom is injected. When coming in contact with the tentacles of any coelenterate, the survivor brushes up against literally thousands of these small stinging organs.

-2. The symptoms produced by coelenterate stings will vary according to species, where the sting is located, and the physical condition of the survivor. In general, though, the sting caused by hydroids and hydroid corals is primarily skin irritations of a local nature. Stings of the Portuguese Man-of-War may be very painful. True corals and sea anemones produce a similar reaction. Some of the sting of these organisms may be hardly noticeable, while others may cause death in 3-8 minutes. Symptoms common to all of these may vary from an immediate mild prickly or stinging sensation, like that of touching a nettle, to a burning, throbbing, shooting-type pain which may cause the survivor to become unconscious. In some cases, the pain may be localized, while in others, it may spread to the groin, armpits, or abdomen. The area in which contact was made will usually become red, followed by severe inflammation, rash, swelling, blistering, skin hemorrhages, and sometimes ulceration. In severe cases of reaction, in addition to shock, the person may experience one or more of the following: muscular cramps, lack of touch and temperature sensations, nausea, vomiting, backache, loss of speech, constriction of the throat, frothing at the mouth, delirium, paralysis, convulsions, and death. Since some of these traits appear quickly, the victim should try to get out of the water if at all possible to avoid drowning.

-3. One of the most deadly jellyfish is the sea wasp (uncommon creature which is found in tropical southern Pacific waters). This animal can cause death anywhere from 30 seconds to 3 hours after contact. Most deaths take place within 15 minutes. The pain is said to be excruciating. The sea wasp can be recognized by the long tentacles that hang down from the four corners of its squarish body.

-a. Relieve pain. Tentacles or other matter on the skin should be removed immediately. This is important because as long as this matter is on the skin, additional stinging cells may be discharged. Use clothing, seaweed, or any other available material to remove the matter. Morphine is effective in relieving pain. *DO NOT* rub the wound with anything, especially sand, as this may cause the stinging cells to be activated. *DO NOT* suck the wound.

-b. Alleviate poison effects. Suntan lotion, oil, and alcohol should be applied to the area to stop further stinging. The following local remedies have been used in various parts of the world with varying degrees of success: papain (protein destroying enzyme), sodium bicarbonate, olive oil, sugar, soap, vinegar, lemon juice, diluted ammonia solution, papaya latex, plant juices, boric acid solution, flour, baking powder, etc. (Urine—with its ammonia content—may be the only source of relief available to a survivor).

-c. Artificial respiration and cardiopulmonary resuscitation may be required. There are no known specific antidotes for most coelenterate stings; however, there is one antivenin for the sea wasp which is papain, a proteolytic enzyme in the juice of the green fruit of the papaya. Even if the survivor is in an area where the antivenin is available, it may be too late to obtain and use it. The venom acts so quickly that medical help is often too late.

-4. Jellyfish should be given a wide berth since in some species the tentacles may trail 50 feet or more from the body. After a storm in tropical areas where large numbers of jellyfish are present, the survivor may be injured by pieces of floating tentacles that have been removed from the animals during the storm. Jellyfish washed up on the beach may appear dead, but can still, in some cases, inflict painful injuries. The best prevention is to stay out of the water by getting into a raft or onto shore. If in a raft, do not let arms and legs trail over the side. The clothing (antiexposure suit) that the survivor wears should cover as much of the body as

possible. Flight clothing items currently available should provide adequate protection.

(2) Mollusks. Octopus, squid, and univalve shellfish are in this category. Mollusks make up the largest single group of biotoxic marine invertebrates of direct importance to the survivor. The phylum of mollusks is generally divided into five classes. Stinging or venomous mollusks which concern the survivor fall mainly into two categories:

(a) Gastropoda (Stomache Footers):

-1. Mollusks. These in general are unsegmented invertebrates. Sometimes their soft bodies will secrete a calcareous shell. They have a muscular foot which serves a variety of functions. Some breathe by means of a type of siphon while others use gills. Some types have jaws. In those that don't have jaws, food is obtained by a rasp-like device called a radula. In the cone shells, the radula is a barb or tooth more like a hollow, needle-like structure.

-2. Gastropods. These univalves include marine snails, slugs, as well as land and freshwater snails. It is estimated that there are over 33,000 living species of gastropods; however, only members of the genus conus are of concern to the survivor. Of these cone shells, there are over 400 species, but they will only be discussed in general terms with the emphasis placed on the more dangerous species. With few exceptions, these attractive shellfish are located in tropical or subtropical areas. All of these shells have a very highly developed venom apparatus designed for vertebrate or invertebrate creatures and are found from shallow tidal areas to depths of many hundreds of feet. The area in which the survivor may come into contact with these shellfish is in coral reefs and sandy or rubble habitat. All cone-shaped shells in these areas should be avoided. Cone shells are usually nocturnal. During the daytime, they burrow and hide in the sand, rocks, or coral; they feed at night on worms, octopus, other gastropods, and small fish. Several of these shells have caused death in humans. The venom apparatus lies within a body cavity of the animal and the animal is capable of thrusting and injecting the poison via the barb into the flesh of the victim. The cone shell is able to inflict its wound only when the head of the animal is out of the shell.

-a. Complications. The sting made by a cone shell is a puncture-type wound. The area around the wound may exhibit one or more of the following: turn blue, swelling, numbness, stinging, or burning sensation. The amount of pain will vary from person to person. Some say the pain is like a bee sting, while others find it unbearable. The numbness and tingling sensations around the site of the wound may spread rapidly, involving the whole body, especially around the lips and mouth. Complete general muscle paralysis may occur. Coma may ensue and death is usually the result of cardiac failure.

-b. Treatment. The pain comes from the injection of venom, slime, and other irritating foreign matter into the wound site. The treatment is primarily symptomatic because there is no specific treatment. Applying hot towels or soaking the affected area in hot water may relieve some of the pain. Artificial respiration may be needed.

(b) Cephalpods. This group includes the nautilis, squid, cuttlefish, and octopus. Since the octopus is the marine animal most likely to be encountered by a survivor, it is the only one that will be discussed. The head of this animal is large and contains well-developed eyes. The mouth is surrounded by eight legs equipped with many suckers. It can move rapidly by expelling water from its body cavity, though it usually glides or creeps over the bottom. Most octopuses live in water ranging from very shallow to depths of over 100 fathoms. All are carnivorous and feed on crabs, and other mollusks. Octopuses like to hide in holes or underwater caves—avoid these areas.

-1. Complications. The sharp parrot-like beak of the octopus makes two small puncture wounds into which a toxic solution or venom is injected. Pain is usually felt immediately in the form of a burning, itching, or stinging sensation. Bleeding from the wound is usually very profuse which may indicate the venom contains an anticoagulant. The area around the wound, and in some cases the entire appendage, may swell, turn red, and feel hot. There has been one report of a fatal octopus bite. This death was attributed to the blue ringed octopus (Octopus Maculosus) (figure 11-43). This small octopus is usually only 3 or 4 inches across although some may be slightly larger. Found throughout the Indo-Pacific area, this octopus is not aggressive toward humans. Because its bite is so dangerous, it should not be handled at any time. When this animal is disturbed the intensity of its blue rings varies rapidly on a light yellow or cream to brown background.

Figure 11-43. Blue Ringed Octopus.

-2. Treatment. Treat for shock, stop bleeding, clean the wound area since more venomous saliva could be in the area, and treat symptoms as they arise. There is no known cure for the venom of the Blue Ringed Octopus.

(3) Echinoderms. Sea cucumbers, starfish, and sea urchins are members of this group. Sea urchins comprise the most dangerous type of echinoderms. Sea urchins have rounded, egg-shaped, or flattened bodies. They have hard shells that carry spines. In some species, the spines are venomous and present a hazard if stepped on or handled. Some urchins are nocturnal. They all tend to be omnivorous, eating algae, mollusks, and other small organisms. They can be found in tidal pools or in areas of great depth in many parts of the world. Sea urchins are not good food sources. At certain times of the year, certain species can be poisonous.

(a) Complications. The needle-sharp points of sea urchin spines are able to penetrate the flesh easily. These spines are also very brittle and tend to break off while still attached to the wound and are very difficult to withdraw. Stepping on one of these spines produces an immediate and very intense burning sensation. The area of pain will also swell, turn red, and ache. Numbness and muscular paralysis, swelling of the face, and a change in the pulse have also been reported. Secondary infection usually sets in. While some deaths have been reported, other victims have experienced loss of speech, respiratory distress, and paralysis. The paralysis will last from 15 minutes to 6 hours.

(b) Treatment. Spines (pedicellaria) that are detached from the animal will continue to secrete venom into the wound. The spines of some species will be easily dislodged whereas others must be surgically removed. There will also be some discoloration due to a dye the animal secretes—do not be disturbed by this. Some experts say to apply grease to allow the spines to be scraped off. Others advise leaving them alone since some of the spines will dissolve in the wound within 24 to 48 hours. Still other experts say to apply citrus juice, if available, or soak the area in vinegar several times a day to dissolve them.

(c) Prevention. No sea urchin should be handled. The spines can penetrate leather and canvas with ease.

g. Venomous and Poisonous Marine Animals (Vertebrates). These fish can be divided into two general groups—fish that sting and fish that are poisonous to eat (figure 11-44).

(1) Venomous Spine Fish (Fish That Sting):

(a) Types of fish in this group are:

-1. Spiny dog fish.

-2. Stingrays. Includes whiprays, batrays, butterfly rays, cow-nosed rays, and round stingrays.

-3. Rat fish.

-4. Weever fish.

-5. Catfish.

-6. Toad fish.

-7. Scorpion fish.

-8. Surgeon fish.

-9. Rabbit fish.

-10. Star gazers.

NOTE: For all wounds from these types of fish, aid should be directed to three areas: alleviating the pain of the sting, trying to halt the effects of the venom, and preventing infection.

(b) Certain types of these fish have up to 18 spines. The pain caused by the sting of one of these spines is so great in some species that the victim may scream and thrash about wildly. In one case, a man stung in the face by a weever fish begged for bystanders to shoot him, even after two shots of morphine sulfate. Many of these fish are bottom dwellers who will not move out of the way when being approached by humans. Instead, they will lie quietly camouflaged, put up their spines, and simply wait for the unlucky individual to step on them. Other people have been injured by them while trying to remove them from fishing nets and fishing lines. In cases where humans are stung by stingrays, the barbs on the sharp spines may cause severe lacerations as well as introduce poison. These wounds should be irrigated without delay. Puncture wounds from the fish are small and make removal of the poison a difficult process. It may be necessary to remove the barb. A procedure which is fairly successful is to make a small cut across the wound (debride) and then apply suction. Even if no incision is made, suction should be tried since it is important to remove as much of the venom as possible. The more poison removed, the better. Morphine does not relieve the pain of some of these venoms. Most doctors agree that the injured part should be soaked in hot water from 30 minutes to 1 hour. The

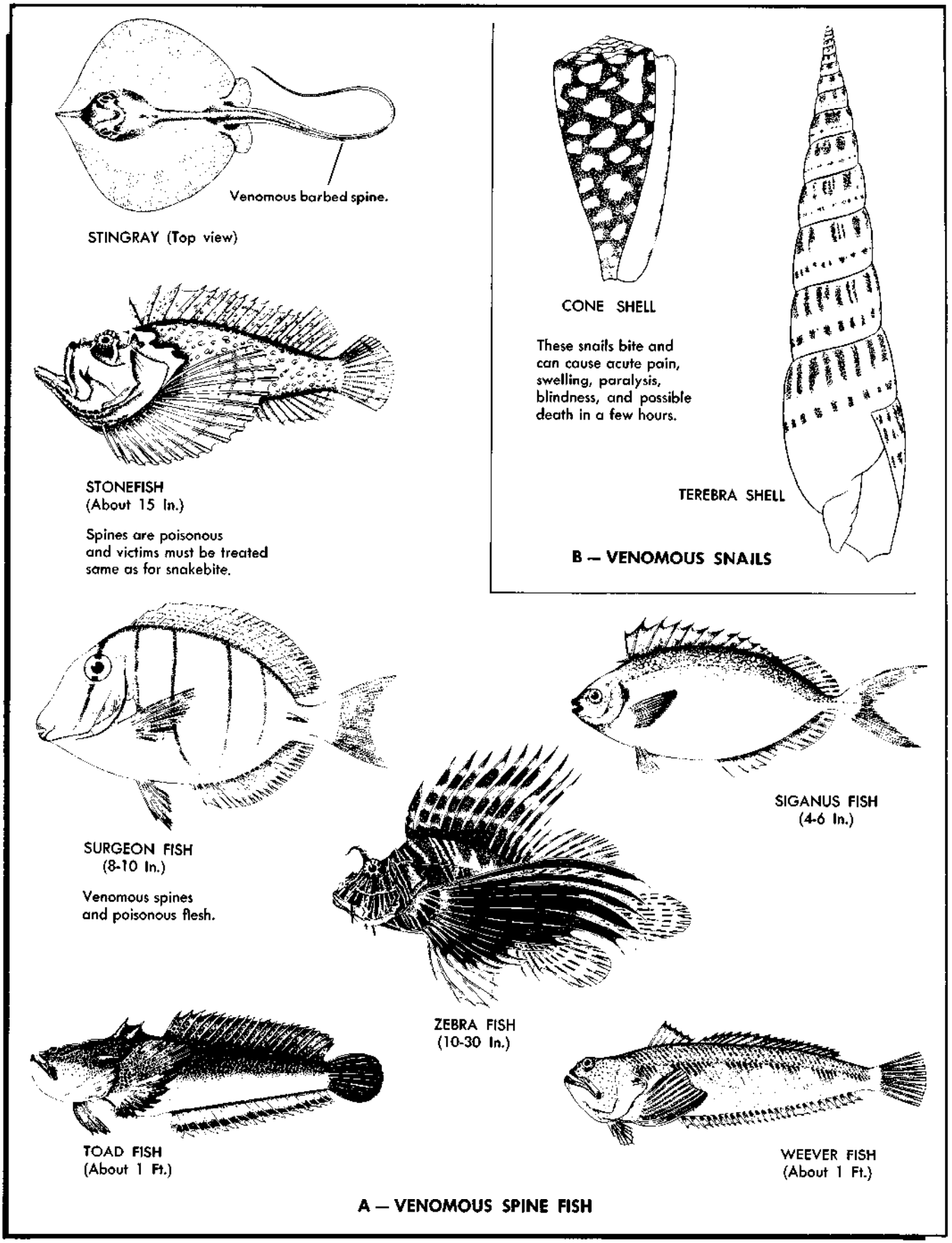

Figure 11-44. Venomous Spine Fish and Snails.

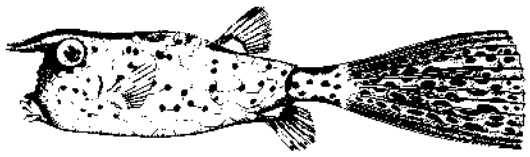

COWFISH
(6-12 In.)

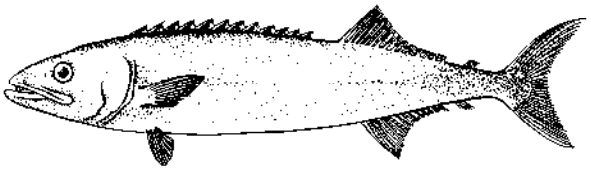

OIL FISH
(3-5 Ft.)

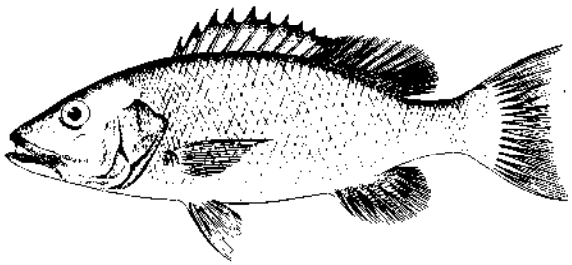

REDSNAPPER FISH
(2-3 Ft.)

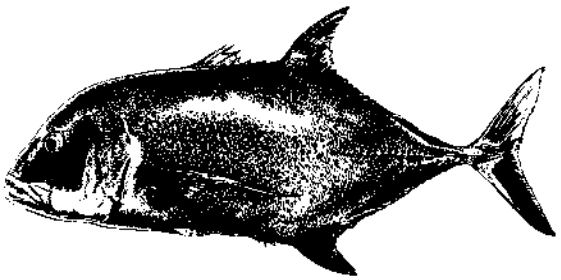

JACKFISH
(About 2 Ft.)

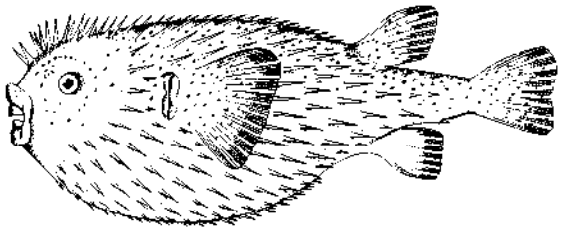

PORCUPINE FISH
(About 1 Ft.)

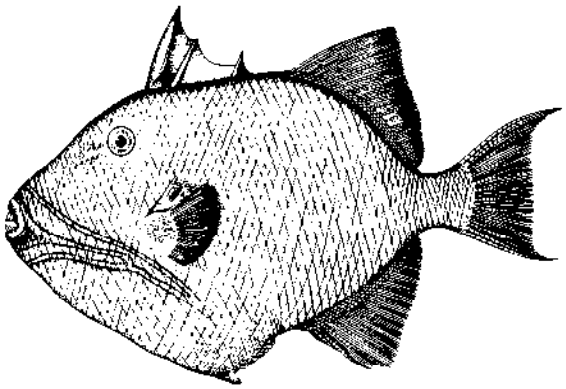

TRIGGER FISH
(1-2 Ft.)

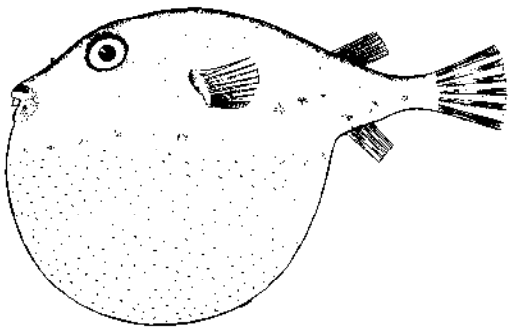

PUFFER FISH
(10-15 In.)

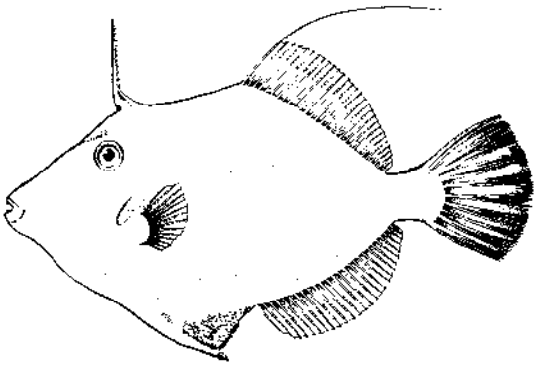

THORNFISH
(About 1 Ft.)

Figure 11-45. Fish With Poisonous Flesh.

temperature of the water should be as hot as the patient can stand without injury. If the wound is on the face or body, hot moist cloth compresses can be used. The use of heat in this manner may weaken the effect of the poison in some cases. After soaking the wound, clean it again, if necessary. Cover the area of the wound with antiseptic and a clean sterile dressing. If antibiotics are available, it may be advisable to use them to help prevent infection. Treatment for shock is wise. Artificial respiration may be needed since some venoms may cause cardiac failure, convulsions, or respiratory distress.

(c) For fish that are poisonous to eat, see figure 11-45.

-1. There is no known way to detect a poisonous fish merely by its appearance. Fish that are poisonous in one area may be safe to eat in another. In general, bottom dwellers and feeders, especially those associated with coral reefs, should be suspect. Also, unusually large predator-type fish should be eaten with caution. The internal organs and roe of all tropical marine fish should never be eaten, as those parts contain a higher concentration of poison.

-2. Under certain conditions, where the survivor may be required to eat questionable fish, rules should be followed. A fish will be safer if it can be caught away from reefs or entrances to lagoons. Once the fish has been secured, the "marine animal edibility test" should be used. The fish should be cut into thin strips and boiled in successive changes of water for an hour or more. This may help since some, but not all, of the toxins are water soluble. Further, it should be noted, that normal cooking techniques and temperatures will not weaken or destroy poisons.

-3. If boiling is not possible, cut the meat into thin strips and soak in changes of sea water for an hour or so, squeezing the meat juices out as thoroughly as possible. A survivor should eat only a small portion of the flesh and wait 12 hours to see if any symptoms arise (if the fish will not spoil). Remember that the degree of poisoning is directly related to how much fish is eaten. If in doubt, do not eat it. The advice of native people on eating tropical marine fish may not be valid. In many instances they check edibility by first feeding fish portions to their dogs and cats.

-4. Treatment. As soon as any symptoms arise, vomiting should be induced by administering warm saltwater or the whites of eggs. If these procedures don't work, try sticking a finger down the person's throat. A laxative should also be given to the victim if one is available. The victim may have to be protected from injury during convulsions. If the victim starts to foam at the mouth and exhibits signs of respiratory distress, a cricothyroidotemy may have to be performed. Morphine may help relieve pain in some cases. If the victim complains of severe itching, cool showers may give some relief. Treat any other symptoms as they arise.

(2) Poisonous Marine Turtles:

(a) Species. There are over 265 species of marine turtles. Of these, only five have been reported as poisonous and dangerous to the survivor. Many of these species are commonly eaten, but for some unknown reason, these same turtles become extremely toxic under certain conditions. Basically, the main species to be concerned with are the green, the hawksbill, and the leatherback turtles. These turtles are found mainly in tropical and subtropical seas but can also be found in temperate waters.

(b) Origin. The origin of turtle poison is unknown but some investigators suggest it comes from the poisonous marine algae eaten by the turtles. It should be noted that a species of turtle may be safe to eat in one area but deadly in another. There is absolutely no way a survivor can distinguish between a poisonous and nonpoisonous sea turtle just by looking at it or by examining any part of it. Toxicity may occur at any time of the year; however, the most dangerous months appear to be the warmer months. The degree of freshness also has nothing to do with how poisonous the turtle is.

(c) Complications. The symptoms will vary with the amount of turtle ingested. Symptoms will develop within a few hours to a few days after eating the food. These symptoms include nausea, vomiting, diarrhea, pain, sweating, coldness in the extremities, vertigo, dry and burning lips and tongue, tightness of the chest, drooling, and difficulty in swallowing. Other victims reported a heavy feeling of the head, a white coating on the tongue, diminished reflexes, coma, and sleepiness. About 44 percent of the victims poisoned by marine turtles die.

(d) Treatment. There is no known antidote for this kind of poisoning. There is no specific treatment—treat symptomatically.

(e) Prevention. If there is the slightest suspicion about the edibility of a marine turtle, it should not be eaten, or at least the marine animal edibility test should be used. Turtle liver is especially dangerous to eat because of its high vitamin A content.

h. Birds. There are roughly 260 species of sea birds. Most of the birds travel only a few miles out to sea but the albatross ranges across the seas far from any landmasses.

i. Red Tide. Red tide is a name used to describe the reddish or brownish coloration in saltwater, resulting from tiny plants and organisms called plankton, which suddenly increase tremendously in numbers. Red tides appear in waters worldwide. In the United States, they are most common off the coasts of Florida, Texas, and southern California. Although most red tides are harmless, some may kill fish and other water creatures. Still other types of red tides do not kill sea life, but cause the shellfish feeding on them to be poisonous. Some of these creatures secrete poisons which can paralyze and kill fish, or can kill fish by using nearly all of the oxygen in

the water. Although the exact reason for the sudden increase of the plankton is unknown, there is evidence that shows favorable food, temperature, sunlight, water currents, and salt in the water will increase the population. It is not unusual for it to remain from a few hours to several months. A survivor should not eat any fish that are found dead.

Chapter 12

LOCAL PEOPLE

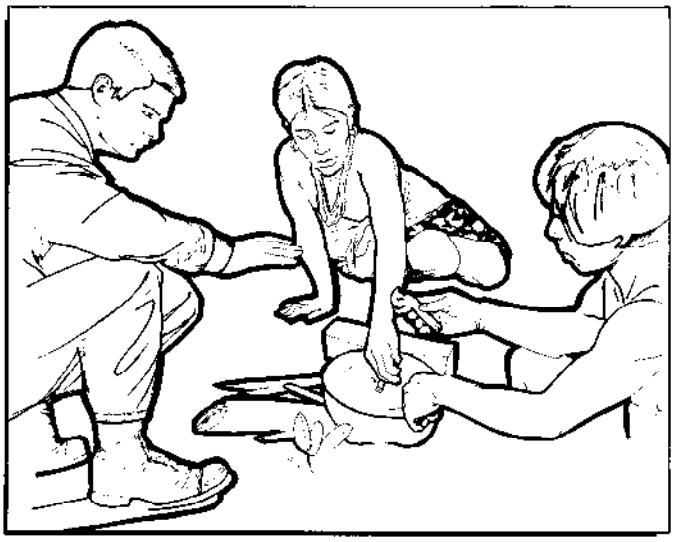

Figure 12-1. Local People.

12-1. Introduction. One evader concluded with the following advice: "My advice is, 'When in Rome, do as the Romans do!' Show interest in their country, and they will go overboard to help you!" One of the most frequently given bits of advice is to accept, respect, and adapt to the ways of the people among whom survivors find themselves. This is good advice, but there are a number of important problems involved in putting this advice into practice (figure 12-1).

12-2. Contact With People. The survivor must give serious consideration to people. Are they people with a primitive culture? Are they farmers, fishermen, friendly people, or enemies? To the survivor, "cross-cultural contact" can vary quite radically in scope. It could mean interpersonal relationships with people of an extremely different (primative) culture, or contacts with people who are culturally modern by our standards. A culture is identified by standards of behavior that are considered proper and acceptable for the members and may or may not conform to our idea of propriety. Regardless of who these people are, the survivor can expect they will have different laws, social and economic values, and political and religious beliefs.

a. People will be friendly, unfriendly, or choose to ignore the survivor. Their attitude may be unknown. If the people are known to be friendly, the survivor must make every attempt to keep them that way by being courteous and respecting the religion, politics, social customs, habits, and all other aspects of their culture. If the people are known to be enemies or are unknowns, the survivor should make every effort to avoid any contact and leave no sign of presence. Therefore, a basic knowledge of the daily habits of the local people can be extremely important in this attempt. An exception might be, if after careful and covert observation, it is determined an unknown people are friendly, contact might be made if assistance is absolutely necessary.

b. Generally, there is little to fear and everything to gain from thoughtful contact with the local peoples of friendly or neutral countries. Familiarity with local customs, displaying common decency, and most importantly, showing respect for their customs should help a survivor avoid trouble and possibly gain needed assistance. To make contact, a survivor should wait until only one person is near and, if possible, let that person make the initial approach. Most people will be willing to help a survivor who appears to be in need; however, political attitudes and training or propaganda efforts can change the attitudes of otherwise friendly people. Conversely, in nominally unfriendly countries, many people, particularly in remote areas, may feel abused or ignored by their politicians, and may be more friendly toward outsiders.

c. The key to successful contact with local peoples is to be friendly, courteous, and patient. Displaying fear, displaying weapons, and making sudden or threatening movements can cause a local person to fear a survivor which can, in turn, prompt a hostile response. When attempting contact, smile frequently. Many local peoples may be shy and seem unapproachable or they may ignore the survivor. Approach them slowly and don't rush matters.

12-3. Survivor's Behavior:

a. Salt, tobacco, silver money, and similar items should be used discreetly in trade. Paper money is well known worldwide. Don't overpay; it may lead later to embarrassment and even danger. Treat people with respect and do not laugh at or bully them.

b. Sign language or acting out needs or questions can be very effective. Many people are accustomed to it and communicate using nonverbal sign language. Aircrew members should learn a few words and phrases of the local language in and around their area of operations. Attempting to speak someone's language is an excellent way to show respect for their culture. Since English is widely used, some of the local people may understand a few words of English.

c. Certain areas may be taboo. They range from religious or sacred places to diseased or danger areas. In some areas, certain animals must not be killed. A survi-

vor must learn what the rules are and follow them. The survivor must be observant and learn as much as possible. This will not only help in strengthening relations, but new knowledge and skills may be very important later. The downed aircrew member should seek advice on local hazards and find out from friendly people where there are hostile people. Keep in mind though, that frequently, people, as in our culture, insist others are hostile because they also do not understand different cultures and distant peoples. The people that generally can be trusted, in their opinion, are their immediate neighbors—much the same as in our own neighborhood. Local people, like ourselves, suffer from diseases which are contagious. The survivor should build a separate dwelling, if possible, and avoid physical contact without seeming to do so. Personal preparation of food and drink is desirable if it can be done without giving offense. Frequently, the use of "personal or religious custom" as an explanation for isolationist behavior will be accepted by the local people.

d. Trading or barter is common in more primitive societies. Hard coin is usually good, whether for its exchange value or as jewelry or trinkets. In isolated places, matches, tobacco, salt, razor blades, empty containers, or cloth may be worth more than any form of money.

e. The survivor must be very cautious when touching people. Many people consider "touching" taboo and such actions may be dangerous. Sexual contact should be avoided.

f. Hospitality among some people is such a strong cultural trait they may seriously reduce their own supplies to make certain a stranger or visitor is fed. What is offered should be accepted and shared equally with all present. The survivor should eat in the same way they eat and, most importantly, attempt to eat all that is offered. If any promises are made, they must be kept. Personal property and local customs and manners, even if they seem odd, must be respected. Some kind of payment for food, supplies, etc., should be made.

g. Privacy must be respected and a survivor should not enter a house unless invited.

12-4. Changing Political Allegiance. In today's world of fast-paced international politics and "shuttle diplomacy," political attitudes and commitments within nations are subject to rapid change. The population of many countries, especially politically hostile countries, must not be considered friendly just because they do not demonstrate open hostility. Unless briefed to the contrary, avoid all contact with such people.

Part Five

PERSONAL PROTECTION

Chapter 13

PROPER BODY TEMPERATURE

13-1. Introduction. In a survival situation the two key requirements for personal protection are maintenance of proper body temperature and prevention of injury. The means for providing personal protection are many and varied. They include the following general categories: clothing, shelter, equipment, and fire. These individual items are not necessary for survival in every situation; however, all four will be essential in some

WIND SPEED		COOLING POWER OF WIND EXPRESSED AS "EQUIVALENT CHILL TEMPERATURE"																				
KNOTS	MPH	TEMPERATURE (°F)																				
Calm	Calm	40	35	30	25	20	15	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-60
		EQUIVALENT CHILL TEMPERATURE																				
3 - 6	5	35	30	25	20	15	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55	-65	-70
7 - 10	10	30	20	15	10	5	0	-10	-15	-20	-25	-35	-40	-45	-50	-60	-65	-70	-75	-80	-90	-95
11 - 15	15	25	15	10	0	-5	-10	-20	-25	-30	-40	-45	-50	-60	-65	-70	-80	-85	-90	-100	-105	-110
16 - 19	20	20	10	5	0	-10	-15	-25	-30	-35	-45	-50	-60	-65	-75	-80	-85	-95	-100	-110	-115	-120
20 - 23	25	15	10	0	-5	-15	-20	-30	-35	-45	-50	-60	-65	-75	-80	-90	-95	-105	-110	-120	-125	-135
24 - 28	30	10	5	0	-10	-20	-25	-30	-40	-50	-55	-65	-70	-80	-85	-95	-100	-110	-115	-125	-130	-140
29 - 32	35	10	5	-5	-10	-20	-30	-35	-40	-50	-60	-65	-75	-80	-90	-100	-105	-115	-120	-130	-135	-145
33 - 36	40	10	0	-5	-15	-20	-30	-35	-45	-55	-60	-70	-75	-85	-95	-100	-110	-115	-125	-130	-140	-150

WINDS ABOVE 40 HAVE LITTLE ADDITIONAL EFFECT.	LITTLE DANGER	INCREASING DANGER (Flesh may freeze within 1 min)	GREAT DANGER (Flesh may freeze within 30 seconds)

DANGER OF FREEZING EXPOSED FLESH FOR PROPERLY CLOTHED PERSONS

INSTRUCTIONS

MEASURE LOCAL TEMPERATURE AND WIND SPEED IF POSSIBLE; IF NOT, ESTIMATE. ENTER TABLE AT CLOSEST 5°F INTERVAL ALONG THE TOP AND WITH APPROPRIATE WIND SPEED ALONG LEFT SIDE. INTERSECTION GIVES APPROXIMATE EQUIVALENT CHILL TEMPERATURE. THAT IS, THE TEMPERATURE THAT WOULD CAUSE THE SAME RATE OF COOLING UNDER CALM CONDITIONS.

NOTES

WIND
1. THIS TABLE WAS CONSTRUCTED USING MILES PER HOUR (MPH), HOWEVER, A SCALE GIVING THE EQUIVALENT RANGE IN KNOTS HAS BEEN INCLUDED ON THE CHART TO FACILITATE ITS USE WITH EITHER UNIT.
2. WIND MAY BE CALM BUT FREEZING DANGER GREAT IF PERSON IS EXPOSED IN MOVING VEHICLE, UNDER HELICOPTER ROTORS, IN PROPELLOR BLAST, ETC. IT IS THE RATE OF RELATIVE AIR MOVEMENTS THAT COUNTS AND THE COOLING EFFECT IS THE SAME WHETHER YOU ARE MOVING THROUGH THE AIR OR IT IS BLOWING PAST YOU
3. EFFECT OF WIND WILL BE LESS IF PERSON HAS EVEN SLIGHT PROTECTION FOR EXPOSED PARTS. LIGHT GLOVES ON HANDS, PARKA HOOD SHIELDING FACE, ETC.

ACTIVITY — DANGER IS LESS IF SUBJECT IS ACTIVE. A PERSON PRODUCES ABOUT 100 WATTS (341 BTUs) OF HEAT STANDING STILL BUT UP TO 1000 WATTS (3413 BTUs) IN VIGOROUS ACTIVITY LIKE CROSS-COUNTRY SKIING.

PROPER USE OF CLOTHING AND ADEQUATE DIET ARE BOTH IMPORTANT.

COMMON SENSE — THERE IS NO SUBSTITUTE FOR IT. THE TABLE SERVES ONLY AS A GUIDE TO THE COOLING EFFECT OF THE WIND ON BARE FLESH WHEN THE PERSON IS FIRST EXPOSED. GENERAL BODY COOLING AND MANY OTHER FACTORS AFFECT THE RISK OF FREEZING INJURY.

Figure 13-1. Windchill Chart.

environments. In this part of the regulation, the conditions which affect the body temperature, the physical principles of heat transfer, and the methods of coping with these conditions will be covered.

13-2. Body Temperature. The body functions best when core temperatures range from 96°F to 102°F. Preventing too much heat loss or gain should be a primary concern for survivors. Factors causing changes in body core temperature (excluding illness) are the climatic conditions of temperature, wind, and moisture.

a. Temperature. As a general rule, exposure to extreme temperatures can result in substantial decreases in physical efficiency. In the worst case, incapacitation and death can result.

b. Wind. Wind increases the chill effect (figure 13-1), causes dissipation of heat, and accelerates loss of body moisture.

c. Moisture — Precipitation, Ground Moisture, or Immersion. Water provides an extremely effective way to transfer heat to and from the body. When a person is hot, the whole body may be immersed in a stream or other body of water to be cooled. On the other hand, in the winter, a hot bath can be used to warm the body. When water is around the body, it tends to bring the "body" to the temperature of the liquid. An example is when a hand is burned and then placed in cold water to dissipate the heat. One way to lower body temperature is by applying water to clothing and exposing the clothed body to the wind. This action causes the heat to leave the body 25 times faster than when wearing dry clothing. This rapid heat transfer is the reason survivors must always guard against getting wet in cold environments. Consider the result of a body totally submerged in water at a temperature of 50°F and determine how long a person could survive (figures 13-2 and 13-3).

13-3. Heat Transfer. There are five ways body heat can be transferred. They are radiation, conduction, convection, evaporation, and respiration.

a. Radiation. Radiation is the primary cause of heat loss. It is defined as the transfer of heat waves from the body to the environment and (or) from the environment back to the body. For example, at a temperature of 50°F, 50 percent of the body's total heat loss can occur through an exposed head and neck. As the temperature drops, the situation gets worse. At 5°F, the loss can be 75 percent under the same circumstances. Not only can heat be lost from the head, but also from the other extremities of the body. The hands and feet radiate heat at a phenomenal rate due to the large number of capillaries present at the surface of the skin. These three areas

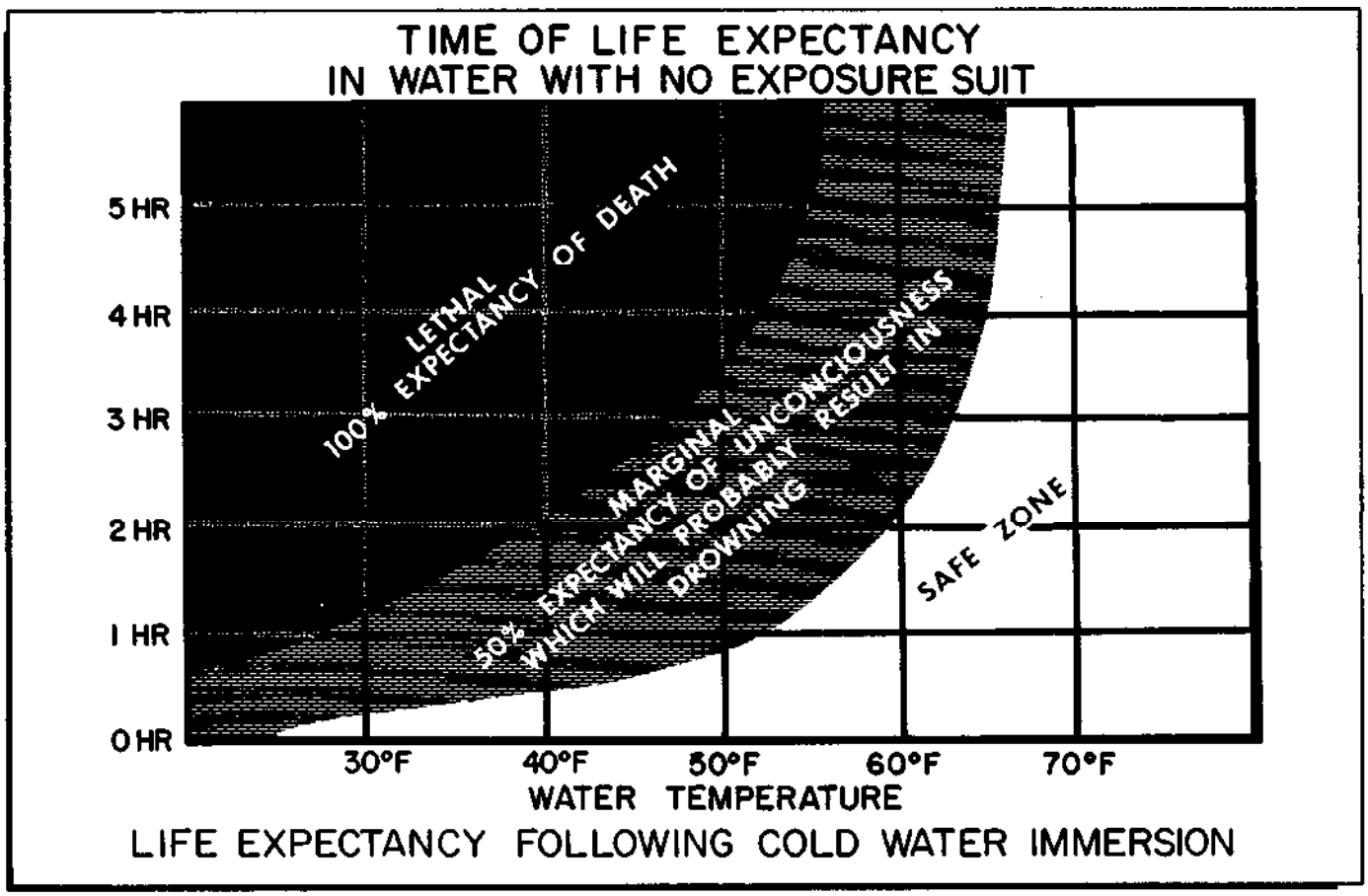

Figure 13-2. Life Expectancy Following Cold-Water Immersion.

of the body must be given particular attention during all periods of exposure to temperature extremes.

b. Conduction:

(1) Conduction is defined as the movement of heat from one molecule to another molecule within a solid object. Extreme examples of how heat is lost and gained quickly are deep frostbite and third-degree burns, both gained from touching the same piece of metal at opposite extremes of cold and heat. Heat is also lost from the body in this manner by touching objects in the cold with bare hands, by sitting on a cold log, or by kneeling on snow to build a shelter. These are practices which survivors should avoid since they can lead to overchilling the body.

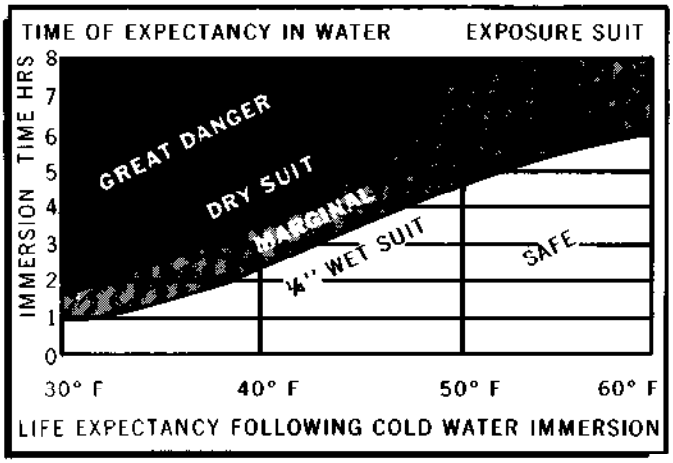

Figure 13-3. Life Expectancy Following Cold-Water Immersion (Exposure Suit.)

(2) Especially dangerous is the handling of liquid fuel at low temperatures. Unlike water which freezes at 32°F, fuel exposed to the outside temperatures will reach the same temperature as the air. The temperature of the fuel may be 10°F to 30°F below zero or colder. Spilling the fluid on exposed skin will cause instant frostbite, not only from the conduction of heat by the cold fluid, but by the further cooling effects of rapid evaporation of the liquid as it hits the skin.

c. Convection. Heat movement by means of air or wind to or from an object or body is known as convection. The human body is always warming a thin layer of air next to the skin by radiation and conduction. The temperature of this layer of air is nearly equal to that of the skin. The body stays warm when this layer of warm air remains close to the body. However, when this warm layer of air is removed by convection, the body cools down. A major function of clothing is to keep the warm layer of air close to the body; however, by removing or disturbing this warm air layer, wind can reduce body temperature. Therefore, wind can provide beneficial cooling in dry, hot conditions, or be a hazard in cold, wet conditions.

d. Evaporation. Evaporation is a process by which liquid changes into vapor, and during this process, heat within the liquid escapes to the environment. An example of this process is how a "desert water bag" works on the front of a jeep while driving in the hot desert. The wind created by the jeep helps to accelerate evaporation and causes the water in the bag to be cooled. The body also uses this method to regulate core temperature when it perspires and air circulates around the body. The evaporation method works any time the body perspires regardless of the climate. For this reason, it is essential that people wear fabrics that "breathe" in cold climates. If water vapor cannot evaporate through the clothing, it will condense, freeze, and reduce the insulation value of the clothing and cause the body temperature to go down.

e. Respiration. The respiration of air in the lungs is also a way of transferring heat. It works on the combined processes of convection, evaporation, and radiation. When breathing, the air inhaled is rarely the same temperature as the lungs. Consequently, heat is either inhaled or expelled with each breath. A person's breath can be seen in the cold as heat is lost to the outside. Because this method is so efficient at transferring heat, warm, moist oxygen is used to treat hypothermia patients in a clinical environment. Understanding how heat is transferred and the methods by which that transfer can be controlled can help survivors keep the body's core temperature in the 96°F to 102°F range. (See figure 13-4.)

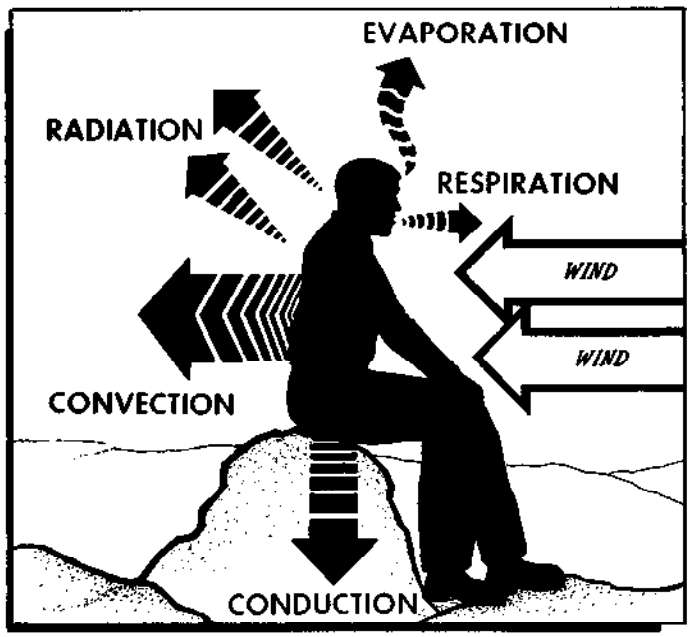

Figure 13-4. Heat Transfer.

Chapter 14

CLOTHING

14-1. Introduction. Every time people go outside they probably neglect to think about one of the most important survival-oriented assets—clothing. Clothing is often taken for granted; people tend to neglect those things which should be the most familiar to them. Clothing is an important asset to survivors and is the most immediate form of shelter. Clothing is important in staying alive, especially if food, water, shelter, and fire are limited or unobtainable. This is especially true in the first stages of an emergency situation because survivors must work to satisfy other needs. If survivors are not properly clothed, they may not survive long enough to build a fire or shelter, to find food, or to be rescued.

14-2. Protection:

a. People have worn clothing for protection since they first put on animal skins, feathers, or other coverings. In most parts of the world, people need clothing for protection from harsh climates. In snow or ice climates, people wear clothing made of fur, wool, or closely woven fabrics. They also wear warm footwear.

b. In dry climates, people wear clothing made of lightweight materials, such as cotton or linen, which have an open weave. These materials absorb perspiration and allow air to circulate around the body. People in dry climates sometimes wear white or light-colored clothes to reflect the Sun's rays. They may also wear sandals, which are cooler and more comfortable than shoes. To protect the head and neck, people wear hats as sunshades.

c. Clothing also provides protection from physical injuries caused by vegetation, terrain features, and animal life which may cause bites, stings, and cuts.

14-3. Clothing Materials:

a. Clothing is made from a variety of materials such as nylon, wool, cotton, etc. The type of material used has a significant effect on protection. Potential survivors must be aware of both the environmental conditions and the effectiveness of these different materials in order to select the best type of clothing for a particular region.

b. Clothing materials include many natural and synthetic fibers. As material is woven together, a "dead air" space is created between the material fibers. When two or three layers of material are worn, a layer of air is trapped between each layer of material creating another layer of "dead air" or insulation. The ability of these different fibers to hold "dead air" is responsible for differing insulation values.

14-4. Natural Materials. They include fur, leather, and cloth made from plant and animal fibers.

a. Fur and leather are made into some of the warmest and most durable clothing. Fur is used mainly for coats and coat linings. Leather has to be treated to make it soft and flexible and to prevent it from rotting.

b. Wool is somewhat different because it contains natural lanolin oils. Although wool is somewhat absorbent, it retains most of its insulating qualities when wet.

c. Cotton is a common plant fiber widely used to manufacture clothing. It absorbs moisture quickly and, with heat radiated from the body, will allow the moisture to pass away from the body. It does not offer much insulation when wet. It's used as an inner layer against the skin and as an outer layer with insulation (for example, wool, Dacron pile, synthetic batting) sandwiched between. The cotton protects the insulation and, therefore, provides warmth.

14-5. Synthetic Materials. Clothing manufacturers are using more and more of these materials. Many synthetic materials are stronger, more shrink-resistant, and less expensive than natural materials. Most synthetic fibers are derived from petroleum in the form of long fibers which consist of different lengths, diameters, and strengths, and sometimes have hollow cores. These fibers, woven into materials such as nylon, Dacron, and polyester, make very strong long-lasting clothing, tarps, tents, etc. Some fibers are spun into a batting type material with air space between the fibers, providing excellent insulation used inside clothing.

a. Many fabrics are blends of natural and synthetic fibers. For example, fabrics could be a mixture of cotton and polyester or wool and nylon. Nylon covered with rubber is durable and waterproof but is also heavy. There are other coverings on nylon which are waterproof but somewhat lighter and less durable. However, most coated nylon has one drawback — it will not allow for the evaporation of perspiration. Therefore, individuals may have to change the design of the garment to permit adequate ventilation (for example, wearing the garment partially unzipped).

b. Synthetic fibers are generally lighter in weight than most natural materials and have much the same insulating qualities. They work well when partially wet and dry out easily; however, they generally do not compress as well as down.

14-6. Types of Insulation:

a. Natural:

(1) Down is the soft plumage found between the skin and the contour feathers of birds. Ducks and geese are good sources for down. If used as insulation in clothing, remember that down will absorb moisture (either precipitation or perspiration) quite readily. Because of

the light weight and compressibility of down, it has wide application in cold-weather clothing and equipment. It is one of the warmest natural materials available when kept clean and dry. It provides excellent protection in cold environments; however, if the down gets wet it tends to get lumpy and loses its insulating value.

(2) Cattail plants have a worldwide distribution, with the exception of the forested regions of the far north. The cattail is a marshland plant found along lakes, ponds, and the backwaters of rivers. The fuzz on the tops of the stalks forms dead-air spaces and makes a good down-like insulation when placed between two pieces of material.

(3) Leaves from deciduous trees (those that lose their leaves each autumn) also make good insulation. To create dead-air space, leaves should be placed between two layers of material.

(4) Grasses, mosses, and other natural materials can also be used as insulation when placed between two pieces of material.

b. Synthetic:

(1) Synthetic filaments such as polyesters and acrylics absorb very little water and dry quickly. Spun synthetic filament is lighter then an equal thickness of wool and unlike down does not collapse when wet, it is also an excellent replacement for down in clothing.

(2) The nylon material in a parachute insulates well if used in the layer system because of the dead-air space. Survivors must use caution when using the parachute in cold climates. Nylon may become "cold soaked;" that is, the nylon will take on the temperature of the surrounding air. People have been known to receive frostbite when placing cold nylon against bare skin.

14-7. Insulation Measurement:

a. The next area to be considered is how well these fibers insulate from the heat or cold. The most scientific way to consider the insulating value of these fibers is to use an established criterion. The commonly accepted measurement used is a comfort level of clothing, called a "CLo" factor.

b. The CLo factor is defined as the amount of insulation which maintains normal skin temperature when the outside ambient air temperature is 70°F with a light breeze. However, the CLo factor alone is not sufficient to determine the amount of clothing required. Such variables as metabolic rate, wind conditions, and the physical makeup of the individual must be considered.

c. The body's rate of burning or metabolizing food and to produce heat varies among individuals. Therefore, some may need more insulation than others even though food intake is equal, and consequently the required CLo value must be increased. Physical activity also causes an increase in the metabolic rate and the rate of blood circulation through the body. When a person is physically active, less clothing or insulation is needed than when standing still or sitting. The effect of the wind, as shown on the windchill chart, must be considered (figure 13-1). When the combination of temperature and wind drops the chill factor to minus 100°F or lower, the prescribed CLo for protecting the body may be inapplicable (over a long period of time) without relief from the wind. For example, when the temperature is minus 60°F, the wind is blowing 60 to 70 miles per hour, and the resultant chill factor exceeds minus 150°F, clothing alone is inadequate to sustain life. Shelter is essential.

d. The physical build of a person also affects the amount of heat and cold that can be endured. For example, a very thin person will not be able to endure as low a temperature as one who has a layer of fat below the skin. Conversely, heavy people will not be able to endure extreme heat as effectively as thinner people.

e. In the Air Force clothing inventory, there are many items which fulfill the need for insulating the body. They are made of the different fibers previously mentioned, and when worn in layers, provide varying degrees of insulative CLo value. The following average zone temperature chart is a guide in determining the best combination of clothing to wear.

TEMPERATURE RANGE	CLo REQUIRED
86 to 68°F	1 - Lightweight
68 to 50°F	2 - Intermediate Weight
50 to 32°F	3 - Intermediate Weight
32 to 14°F	3.5 - Heavyweight
14 to -4°F	4.0 - Heavyweight
-4 to -40°F	4.0 - Heavyweight

The amount of CLo value per layer of fabric is determined by the loft (distance between the inner and outer surfaces) and the amount of dead air held within the fabric. Some examples of the CLo factors and some items of clothing are:

LAYERS:	1 - Aramid underwear (1 layer)	0.6 CLo
	2 - Aramid underwear (2 layers)	1.5 CLo
	3 - Quilted liners	1.9 CLo
	4 - Nomex coveralls	.6 CLo
	5 - Winter coveralls	1.2 CLo
	6 - Nomex jacket	1.9 CLo

This total amount of insulation should keep the average person warm at a low temperature. When comparing items one and two in the above example, it shows when doubling the layer of underwear, the CLo value more than doubles. This is true not only on the number one item but between all layers of any clothing system. Therefore, one gains added protection by using several very thin layers of insulation rather than two thick lay-

ers. The air held between these thin layers increases the insulation value.

f. The use of many thin layers also provides (through removal of desired number of layers) the ability to closely regulate the amount of heat retained inside the clothing. The ability to regulate body temperature helps to alleviate the problem of overheating and sweating, and preserves the effectiveness of the insulation.

g. The principle of using many thin layers of clothing can also be applied to the "sleeping system" (sleeping bag, liner, and bed). This system uses many layers of synthetic material, one inside the other, to form the amount of dead air needed to keep warm. To improve this system, a survivor should wear clean and dry clothing in layers (the layer system) in cold climates. While discussing the layer system, it is important to define the "COLDER" principle. This acronym is used to aid in remembering how to use and take care of clothing.

C - Keep clothing *C*lean.
O - Avoid *O*verheating.
L - Wear clothing *L*oose and in *L*ayers.
D - Keep clothing *D*ry.
E - *E*xamine clothing for defects or wear.
R - Keep clothing *R*epaired.

(1) Clean. Dirt and other materials inside fabrics will cause the insulation to be ineffective, abrade and cut the fibers which make up the fabric, and cause holes. Washing clothing in the field may be impractical; therefore, survivors should concentrate on using proper techniques to prevent soiling clothing.

(2) Overheating. Clothing best serves the purpose of preserving body heat when worn in layers as follows: absorbent material next to the body, insulating layers, and outer garments to protect against wind and rain. Because of the rapid change in temperature, wind, and physical exertion, garments should allow donning and removal quickly and easily. Ventilation is essential when working because enclosing the body in an airtight layer system results in perspiration which wets clothing, thus reducing its insulating qualities.

(3) Loose. Garments should be loose fitting to avoid reducing blood circulation and restricting body movement. Additionally, the garment should overhang the waist, wrists, ankles, and neck to reduce body heat loss.

(4) Dry. Keep clothing dry since a small amount of moisture in the insulation fibers will cause heat losses up to 25 times faster than dry clothing. Internally produced moisture is as damaging as is externally dampened clothing. The outer layer should protect the inner layers from moisture as well as from abrasion of fibers; for example, wool rubbing on logs or rocks, etc. The outer shell keeps dirt and other contaminants out of the clothing. Clothing can be dried in many ways. Fires are often used; however, take care to avoid burning the items. The "bare hand" test is very effective. Place one hand near the fire in the approximate place the wet items will be and count to three slowly. If this can be done without feeling excessive heat, it should be safe to dry items there. Never leave any item unattended while it is drying. Leather boots, gloves, and mitten shells require extreme care to prevent shrinkage, stiffening, and cracking. The best way to dry boots is upright beside the fire (not upside down on sticks because the moisture does not escape the boot) or simply walk them dry in the milder climates. The Sun and wind can be used to dry clothing with little supervision except for checking occasionally on the incoming weather and to make sure the article is secure. Freeze-drying is used in subzero temperatures with great success. Survivors let water freeze on or inside the item and then shake, bend, or beat it to cause the ice particles to fall free from the material. Tightly woven materials work better with this method than do open fibers.

(5) Examine. All clothing items should be inspected regularly for signs of damage or soil.

(6) Repair. Eskimos set an excellent example in the meticulous care they provide for their clothing. When damage is detected, immediately repair it.

h. The neck, head, hands, armpits, groin, and feet lose more heat than other parts of the body and require greater protection. Work with infrared film shows tremendous heat loss in those areas when not properly clothed. Survivors in a cold environment are in a real emergency situation without proper clothing. Figure 14-1 shows some examples of how military clothing works to hold body heat.

i. Models wearing samples of aircrew attire appear as spectral figures in a thermogram, an image revealing differences in infrared heat radiated from their clothing and exposed skin. White is warmest; red, yellow, green, blue, and magenta form a declining temperature scale spanning about 15 degrees; while black represents all lower temperatures. Almost the entire scale is seen on the model in boxer shorts. Warm, white spots appear on the underarm and neck. Only the shorts block radiation from the groin. Temperatures cool along the arm to dark blue fingertips far from the heat-producing torso. The addition of the next layer of clothing (Aramid long underwear) prevents heat loss except where it is tight against the body. As more layers are added, it is easy to see the areas of greatest concern are the head, hands, and feet. These areas are difficult for crewmembers to properly insulate while flying an aircraft. Mittens are ineffective due to the degraded manual dexterity. Likewise, it is difficult to feel the rudder pedal action while wearing bulky warm boots. These problems require inclusion of warm hats, mittens, and footgear (mukluk type) in survival kits during cold weather operation. Research has shown when a CLo value of 10 is used to insulate the head, hands, and feet and the rest of the body is only protected by one CLo, the average individual can be exposed to low temperatures (-10°F) comfortably for a reasonable period of time (30 to 40 minutes). When the amount of CLo value placed on the

GROUP ONE - REMOVING CLOTHES

1-1: Fully clothed

1-2: Flight jacket, wool cap and mittens, and leather shell

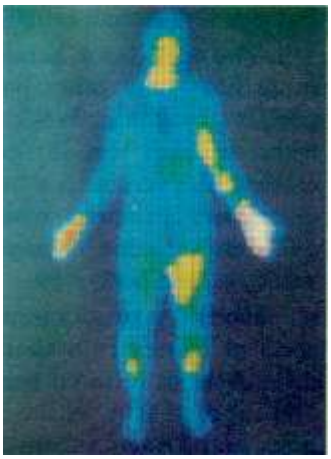

1-3: Flight suit

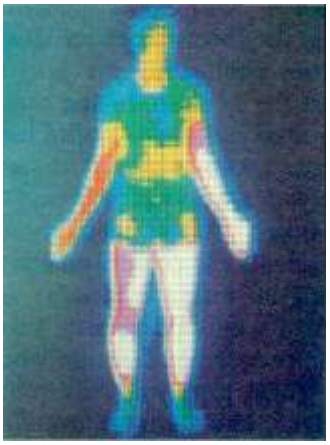

1-4: Thermals

1-5: "T" shirt and two pair cotton socks

GROUP TWO - DONNING CLOTHES

2-1: Unclothed

2-2: "T" shirt and two pair cotton socks

2-3: Thermals

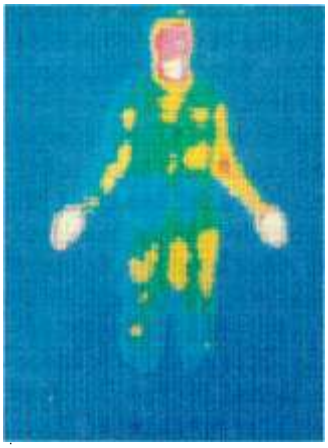

2-4: Flight suit

2-5: Flight jacket, wool cap and mittens, and leather shell

NOTE: Dark blue indicates no heat loss; the lighter the color, the greater the heat loss.

Figure 14-1. Thermogram of Body Heat Loss.

individual is reversed, the amount of time a survivor can spend in cold weather is greatly reduced due to the heat loss from their extremities. This same principle works in reverse in hot parts of the world if one submerges the head, hands, or feet in cold water, it lets the most vascular parts of the body lose heat quickly.

14-8. Clothing Wear in Snow and Ice Areas:

a. The survivor should:

(1) Avoid restricting the circulation. Clothing should not be worn so tight that it restricts the flow of blood which distributes the body heat and helps prevent frostbite. When wearing more than one pair of socks or gloves, ensure that each succeeding pair is large enough to fit comfortably over the other. Don't wear three or four pairs of socks in a shoe fitted for only one or two pairs. Release any restriction caused by twisted clothing or a tight parachute harness.

(2) Keep the head and ears covered. Survivors will lose as much as 50 percent of their total body heat from an unprotected head at 50°F.

(3) When exerting the body, prevent perspiration by opening clothing at the neck and wrists and loosening it at the waist. If the body is still warm, comfort can be obtained by taking off outer layers of clothing, one layer at a time. When work stops, the individual should put the clothing on again to prevent chilling.

(4) If boots are big enough, use dry grass, moss, or other material for added insulation around the feet. Footgear can be improvised by wrapping parachute cloth or other fabric lined with dry grass or moss for insulation.

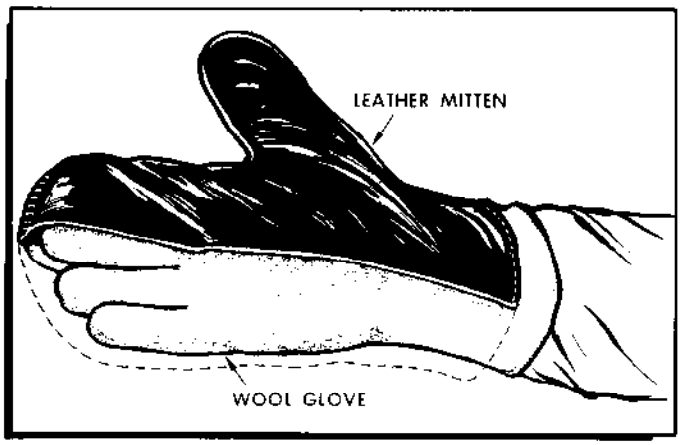

Figure 14-2. Layer System for Hands.

b. Felt booties and mukluks with the proper socks and insoles are best for dry, cold weather. Rubber-bottomed boot shoepacs with leather tops are best for wet weather. Mukluks should not be worn in wet weather. The vapor-barrier rubber boots can be worn under both conditions and are best at extremely low temperatures. The air release valve should be closed at ground level. These valves are designed to release pressure when airborne. Air should not be blown into the valves as the moisture could decrease insulation.

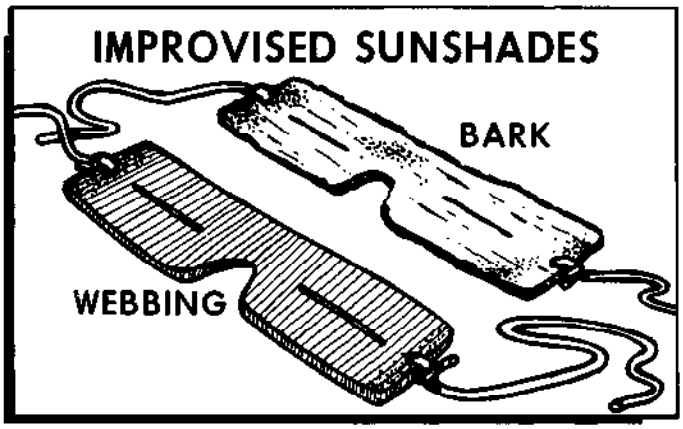

Figure 14-3. Improvised Goggles.

c. Clothing should be kept as dry as possible. Snow must be brushed from clothing before entering a shelter or going near a fire. The survivors should beat the frost out of garments before warming them, and dry them on a rack near a fire. Socks should be dried thoroughly.

d. One or two pairs of wool gloves and (or) mittens should be worn inside a waterproof shell (figure 14-2). If survivors have to expose their hands, they should warm them inside their clothing.

e. To help prevent sun or snow blindness, a survivor should wear sun or snow goggles or improvise a shield with a small horizontal slit opening (figure 14-3).

f. In strong wind or extreme cold, as a last resort, a survivor should wrap up in parachute material, if available, and get into some type of shelter or behind a windbreak. Extreme care should be taken with hard materials, such as synthetics, as they may become cold soaked and require more time to warm.

g. At night, survivors should arrange dry spare clothing loosely around and under the shoulders and hips to help keep the body warm. Wet clothes should never be worn into the sleeping bag. The moisture destroys the insulation value of the bag.

h. If survivors fall into water, they should roll in dry snow to blot up moisture, brush off the snow, and roll again until most of the water is absorbed. They should not remove footwear until they are in a shelter or beside a fire.

i. All clothing made of wool offers good protection when used as an inner layer. When wool is used next to the face and neck, survivors should be cautioned that moisture from the breath will condense on the surface and cause the insulating value to decrease. The use of a wool scarf wrapped around the mouth and nose is an excellent way to prevent cold injury, but it needs to be

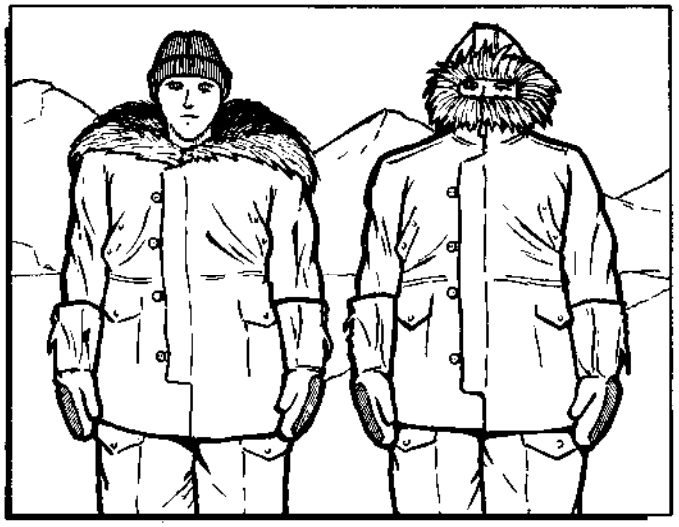

Figure 14-4. Proper Wear of Parka.

de-iced on a regular basis to prevent freezing flesh adjacent to it. An extra shell is generally worn over the warming layers to protect them and to act as a windbreak.

j. Other headgear includes the pile cap and hood. These items are most effective when used with a covering for the face in extreme cold. The pile cap is extremely warm where it is insulated, but it offers little protection for the face and back of the neck.

k. The hood is designed to funnel the radiant heat rising from the rest of the body and to recycle it to keep the neck, head, and face warm (figure 14-4). The individual's ability to tolerate cold should dictate the size of the front opening of the hood. The "tunnel" of a parka hood is usually lined with fur of some kind to act as a protecting device for the face. This same fur also helps to protect the hood from the moisture expelled during breathing. The closed tunnel holds heat close to the face longer; the open one allows the heat to escape more freely. As the frost settles on the hair of the fur, it should be shaken from time to time to keep it free of ice buildup.

l. Sleeping systems (sleeping bag, liner, and bed) are the transition "clothing" used between normal daytime activities and sleep (figure 14-7).

m. The insulating material in the sleeping bag may be synthetic or it may be down and feathers. (Feathers and down lining require extra protection from moisture). However, the covering is nylon. Survivors must realize that sleeping bags are compressed when packed and must be fluffed before use to restore insulation value. Clean and dry socks, mittens, and other clothing can be used to provide additional insulation.

n. Footgear is critical in a survival situation because walking is the only means of mobility. Therefore, care of footgear is essential both before and during a survival situation. Recommendations for care are:

(1) Ensure footgear is properly "broken-in" before flying.

(2) "Treat" footgear to ensure water-repellency (follow manufacturer's recommendations).

(3) Keep leather boots as dry as possible.

o. Mukluks have been around for thousands of years and have proven their worth in extremely cold weather. The Air Force mukluks are made of cotton duck with rubber-cleated soles and heels. (See figure 14-6.) They have slide fasteners from instep to collar, laces at instep and collar, and are 18 inches high. They are used by flying and ground personnel operating under *dry*, cold conditions in temperatures below +15°F. Survivors should change liners daily when possible.

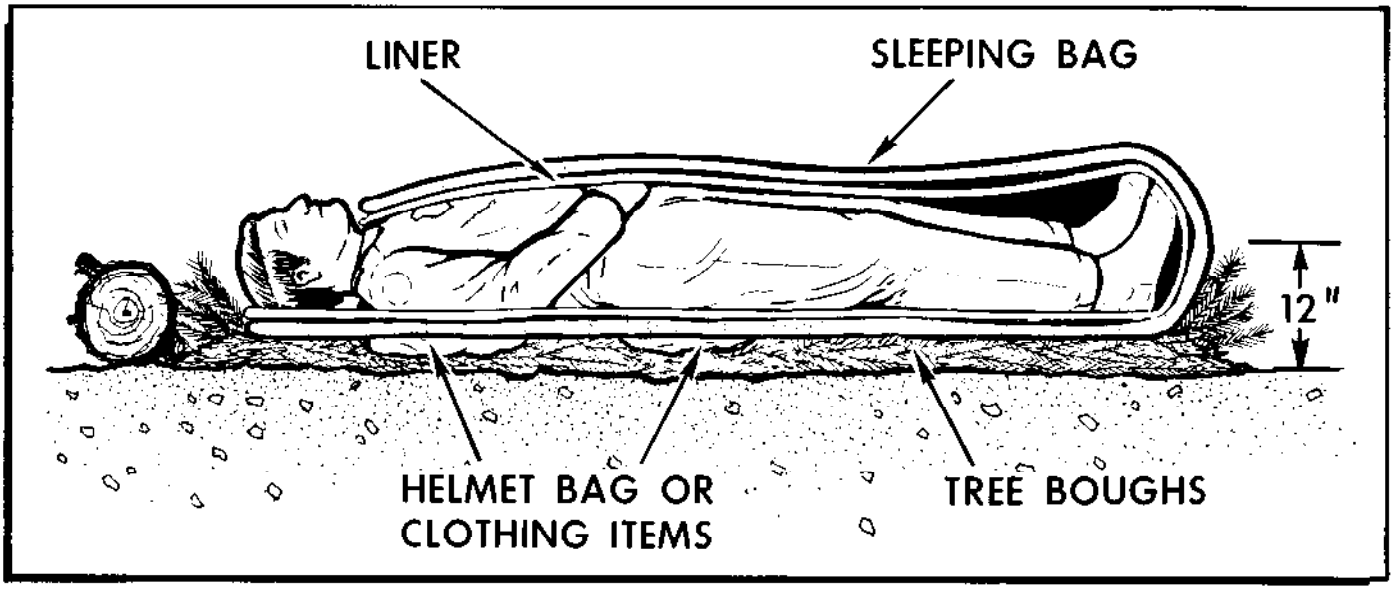

Figure 14-5. Sleeping System.

Figure 14-6. Issued Mukluks.

14-9. Care of the Feet. Foot care is critical in a survival situation. Improvising foot gear may be essential to caring for feet.

a. Moose Hock Shoe. The hock skin of a moose or caribou will provide a suitable pair of shoes (figure 14-7). Cut skin around leg at A and B. Separate from the leg and pull it over the hoof. Shape and sew up small end C. Slit skin from A to B; bore holes on each side of cut for lacing; turn inside out, and lace with rawhide, suspension line, or other suitable material.

b. Grass Insoles. Used extensively by northern natives to construct inner soles. Grass is a good insulator and will collect moisture from the feet. The survivor should use the following procedure to prepare grass for use as inner soles: Grasp a sheaf of tall grass, about one-half inch in diameter, with both hands. Rotate the hands in opposite directions. The grass will break up or "fluff" into a soft mass. Form this fluff into oblong shapes and spread it evenly throughout the shoes. The inner soles should be about an inch thick. Remove these inner soles at night and make new ones the following day.

c. Hudson Bay Duffel. A triangular piece of material used as a foot covering. To improvise this foot covering, a survivor can use the following procedures:

(1) Cut two to four layers of parachute cloth into a 30-inch square.

(2) Fold this square to form a triangle.

(3) Place the foot on this triangle with the toes pointing at one corner.

(4) Fold the front cover up over the toes.

(5) Fold the side corners, one at a time, over the instep. This completes the foot wrap. (See figure 14-8.)

d. Gaiters. Made from parachute cloth, webbing, or canvas. Gaiters help keep sand and snow out of shoes and protect the legs from bites and scratches (figure 14-9).

e. Double Socks. Cushion padding, feathers, dry grass, or fur stuffed between layers of socks. Wrap parachute or aircraft fabric around the feet and tie above the ankles. A combination of two or more types of improvised footwear may be more desirable and more efficient than any single type (figure 14-10).

14-10. Clothing in the Summer Arctic:

a. In the summer arctic, there are clouds of mosquitoes and black flies so thick a person can scarcely see through them. Survivors can protect themselves by wearing proper clothing to ensure no bare skin is exposed. A good head net and gloves should be worn.

b. Head nets must stand out from the face so they won't touch the skin. Issued head nets are either black or green. If one needs to be improvised they can be sewn to the brim of the hat or can be attached with an elastic band that fits around the crown. Black is the best color, as it can be seen through more easily than green or white. A heavy tape encasing a drawstring should be attached to the bottom of the head net for tying snugly at the collar. Hoops of wire fastened on the inside will make the net stand out from the face and at the same time allow it to be packed flat. The larger they are, the better the ventilation. But very large nets will not be as effective in wooded country where they may become snagged on brush.

c. Gloves are hot, but are a necessity where flies are found in swamps. Kid gloves with a 6-inch gauntlet closing the gap at the wrist and ending with an elastic band halfway to the elbow are best. For fine work, kid gloves with the fingers cut off are good. Cotton/Nomex work gloves are better than no protection at all, but mosquitoes will bite through them. Treating the gloves with insect repellent will help. Smoky clothing may also help to keep insects away. (See figure 14-11.)

d. A survivor should remember that mosquitoes do not often bite through two layers of cloth; therefore, a lightweight undershirt and long underwear will help. To

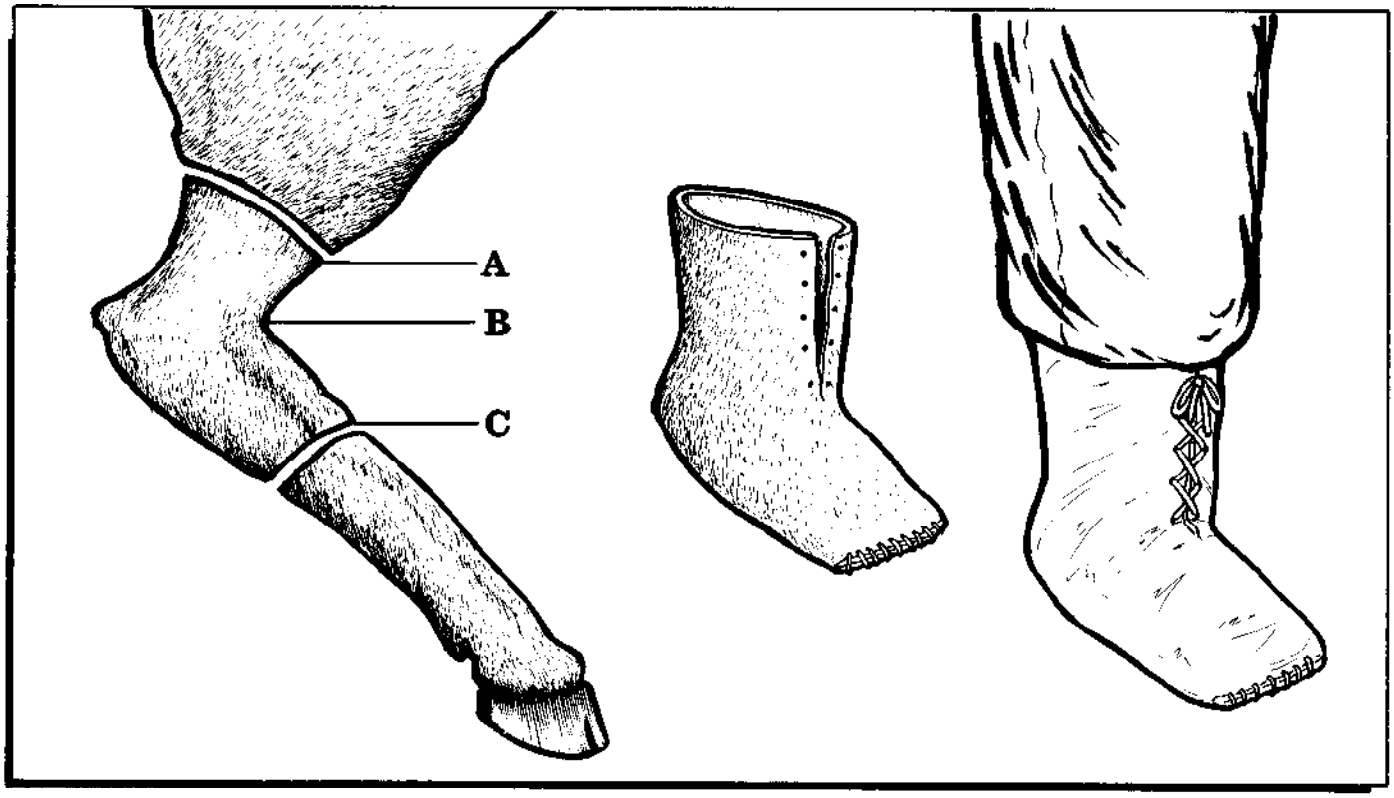

Figure 14-7. Moose Hock Shoes.

protect ankles, blouse the bottoms of trousers around boots, or wear some type of leggings (gaiters).

e. If the head net is lost or none is available, make the best of a bad situation by wearing sunglasses with improvised screened sides, plugging ears lightly with cotton, and tying a handkerchief around the neck. Treat clothing with insect repellent at night.

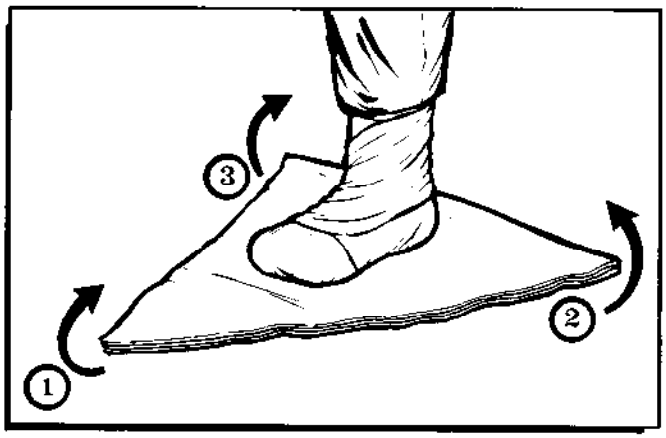

Figure 14-8. Hudson Bay Duffel.

14-11. Clothing at Sea. In cold oceans, survivors must try to stay dry and keep warm. If wet, they should use a wind screen to decrease the cooling effects of the wind. They should also remove, wring out, and replace outer

Figure 14-9. Gaiters.

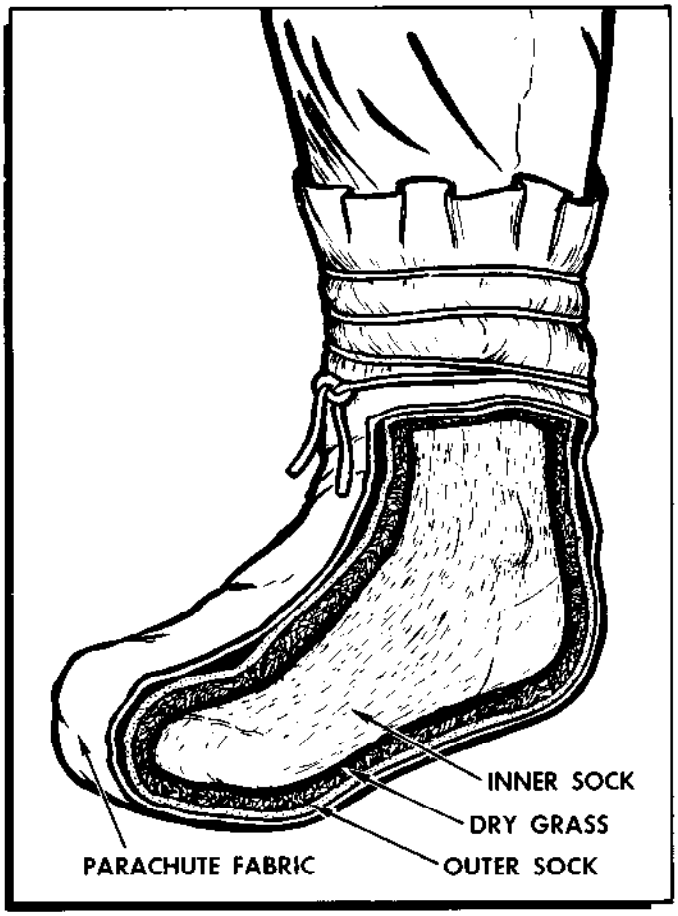

Figure 14-10. Double Socks.

Figure 14-11. Insect Protection.

garments or change into dry clothing. Hats, socks, and gloves should also be dried. If any survivors are dry, they should share extra clothes with those who are wet. Wet personnel should be given the most sheltered positions in the raft. Let them warm their hands and feet against those who are dry. Survivors should put on any extra clothing available. If no anti-exposure suits are provided, they can drape extra clothing around their shoulders and over their heads. Clothes should be loose and comfortable. Also, survivors should attempt to keep the floor of the raft dry. For insulation, covering the floor with any available material will help. Survivors should huddle together on the floor of the raft and spread extra tarpaulin, sail, or parachute material over the group. If in a 20- or 25-man raft, canopy sides can be lowered. Performing mild exercises to restore circulation may be helpful. Survivors should exercise fingers, toes, shoulders, and buttock muscles. Mild exercise will help keep the body warm, stave off muscle spasms, and possibly prevent medical problems. Survivors should warm hands under armpits and periodically raise feet slightly and hold them up for a minute or two. They should also move face muscles frequently to prevent frostbite. Shivering is the body's way of quickly generating heat and is considered normal. However, persistent shivering may lead to uncontrollable muscle spasms. They can be avoided by exercising muscles. If water is available, additional rations should be given to those suffering from exposure to cold. Survivors should eat small amounts frequently rather than one large meal.

14-12. Antiexposure Garments:

a. Assemblies. The antiexposure assemblies, both quick donning and constant wear, are designed for personnel participating in over-water flights where unprotected or prolonged exposure to the climatic conditions

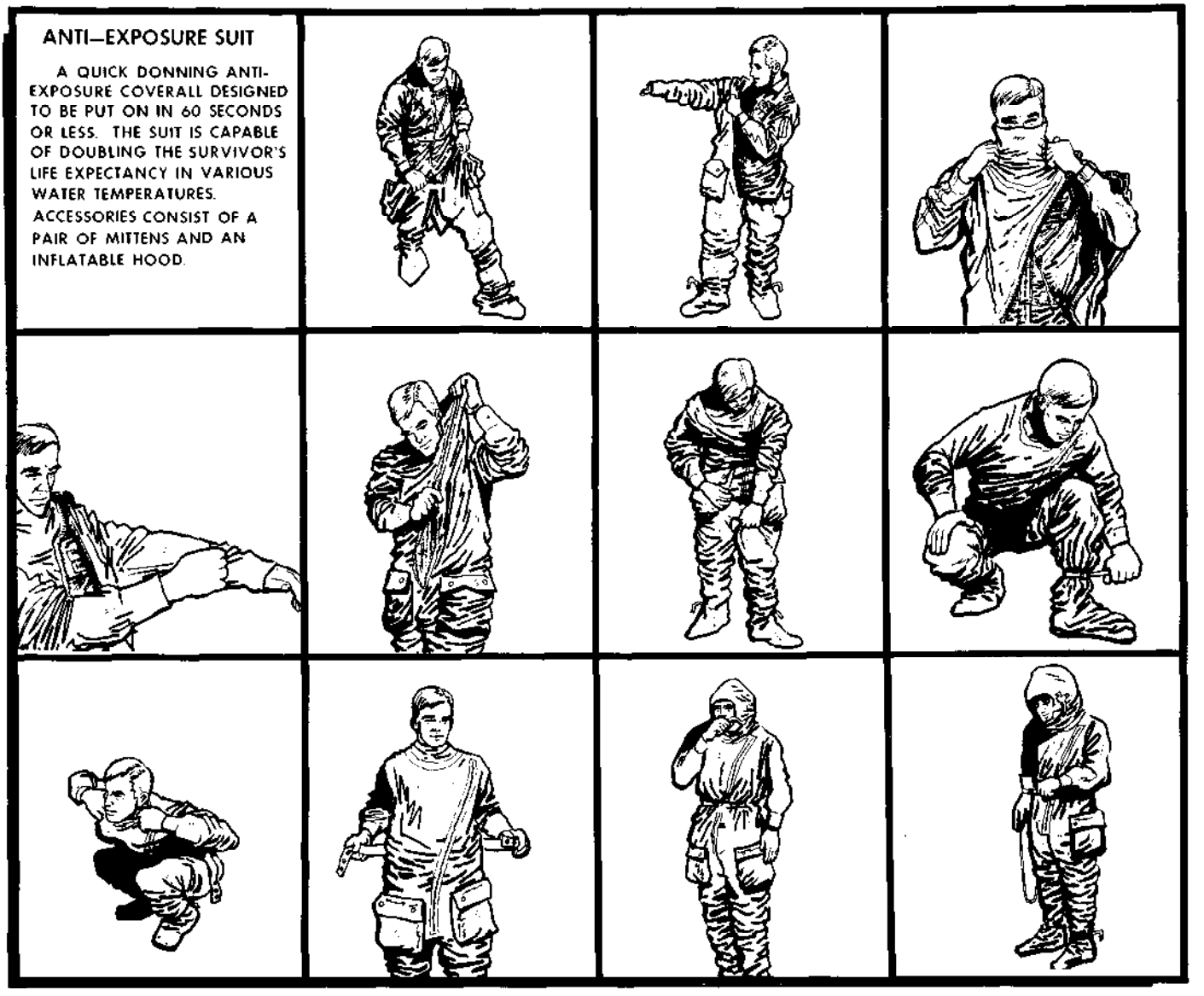

Figure 14-12. Donning Antiexposure Suit.

of cold air and (or) cold water (as a result of ditching or abandoning an aircraft) would be dangerous or could prove fatal. The suit provides protection from the wind and insulation against the chill of the ocean. The result of exposure in the water is illustrated in figures 13-2 and 13-3. Exposure time varies depending on the particular antiexposure assembly worn, the cold sensitiveness of the person, and survival procedures used.

b. Quick-Donning Antiexposure Flying Coverall. Some antiexposure coveralls are designed for quick donning (approximately 1 minute) before emergency ditching. After ditching the aircraft, the coverall protects the wearer from exposure while swimming in cold water, and from exposure to wind, spray, and rain when adrift in a liferaft.

(1) The coverall is a one-size garment made from chloroprene-coated nylon cloth. It has two expandable-type patch pockets, an adjustable waist belt, and attached boots with adjustable ankle straps. One pair of insulated, adjustable wrist strap mittens, each with a strap attached to a pocket, is provided. A hood, also attached with a strap, is in the left pocket. A carrying case with instructions and a snap fastener closure is furnished for stowing in the aircraft.

(2) To use the coverall, personnel should wear it over regular flight clothing. It is large enough to wear over the usual flight gear. The gloves and hood are stowed in the pockets of the coverall and are normally worn after boarding the liferaft.

(3) The survivor should be extremely careful when donning the coverall to prevent damage by snagging,

tearing, or puncturing it on projecting objects. After donning the coverall, the waist band and boot ankle straps should be adjusted to take up fullness. If possible, crewmembers should stoop while pulling the neck seal to expel air trapped in the suit. When jumping into the water, they should leap feet first with hands and arms close to sides or brought together above the head (figure 14-12). Note there is a constant wear exposure suit designed to be worn continuously during overwater flights where the water temperature is 60 degrees or below. The Command may waiver it to 51 degrees.

14-13. Warm Oceans. Protection against the Sun and securing drinking water are the most important problems. A survivor should keep the body covered as much as possible to avoid sunburn. A sunshade can be improvised out of any materials available or the canopy provided with the raft may be used. If the heat becomes too intense, survivors may dampen clothing with sea water to promote evaporation and cooling. The use of sunburn preventive cream or a Chapstick is advisable. Remember, the body must be kept covered completely. Exposure to the Sun increases thirst, wastes precious water, reduces the body's water content, and causes serious burns. Survivors should roll down their sleeves, pull up their socks, close their collars, wear a hat or improvised headgear, use a piece of cloth as a shield for the back of the neck, and wear sunglasses or improvise eye covers.

14-14. Tropical Climates:

a. In tropical areas, the body should be kept covered for prevention of insect bites, scratches, and sunburn.

b. When moving through vegetation, survivors should roll down their sleeves, wear gloves, and blouse the legs of their pants or tie them over their boot tops. Improvised puttees (gaiters) can be made from parachute material or any available fabric. This will protect legs from ticks and leeches.

c. Loosely worn clothing will keep survivors cooler, especially when subjected to the direct rays of the Sun.

d. Survivors should wear a head net or tie material around the head for protection against insects. The most active time for insects is at dawn and dusk. An insect repellent should be used at these times.

e. In open country or in high grass, survivors should wear a neck cloth or improvised head covering for protection from sunburn and (or) dust. They should also move carefully through tall grass, as some sharp-edged grasses can cut clothing to shreds. Survivors should dry clothing before nightfall. If an extra change of clothing is available, effort should be made to keep it clean and dry.

14-15. Dry Climates:

a. In the dry climates of the world, clothing will be needed for protection against sunburn, heat, sand, and insects. Survivors should not discard any clothing. They should keep their head and body covered and blouse the legs of pants over the tops of footwear during the day. Survivors should not roll up sleeves, but keep them rolled down and loose at the cuff to stay cool.

Figure 14-13. Protective Desert Clothing.

b. Survivors should keep in mind that the people who live in the hot dry areas of the world usually wear heavy white flowing robes which protect almost every inch of their bodies. The only areas open to the Sun are the face and the eyes. This produces an area of higher humidity between the body and the clothing, which helps keep them cooler and conserves their perspiration (figure 14-13). The white clothing also reflects the sunlight.

c. Survivors should wear a cloth neckpiece to cover the back of the neck and protect it from the Sun. A T-

shirt makes an excellent neck drape, with the extra material used as padding under the cap. If hats are not available, survivors can make headpieces like those worn by the Arabs, as shown in figure 14-13. During dust storms, they should wear a covering for the mouth and nose; parachute cloth will work.

d. If shoes are lost or if they wear out, survivors can improvise footgear. One example of this is the "Russian Sock." Parachute material can be used to improvise these socks. The parachute material is cut into strips approximately 2 feet long and 4 inches wide. These strips are wrapped bandage fashion around the feet and ankles. Socks made in this fashion will provide comfort and protection for the feet.

Chapter 15

SHELTER

15-1. Introduction. Shelter is anything that protects a survivor from the environmental hazards. The information in this chapter describes how the environment influences shelter site selection and factors which survivors must consider before constructing an adequate shelter. The techniques and procedures for constructing shelters for various types of protection are also presented.

15-2. Shelter Considerations. The location and type of shelter built by survivors vary with each survival situation. There are many things to consider when picking a site. Survivors should consider the time and energy required to establish an adequate camp, weather conditions, life forms (human, plant, and animal), terrain, and time of day. Every effort should be made to use as little energy as possible and yet attain maximum protection from the environment.

a. Time. Late afternoon is not the best time to look for a site which will meet that day's shelter requirements. If survivors wait until the last minute, they may be forced to use poor materials in unfavorable conditions. They must constantly be thinking of ways to satisfy their needs for protection from environmental hazards.

b. Weather. Weather conditions are a key consideration when selecting a shelter site. Failure to consider the weather could have disastrous results. Some major weather factors which can influence the survivor's choice of shelter type and site selection are temperature, wind, and precipitation.

(1) Temperature. Temperatures can vary considerably within a given area. Situating a campsite in low areas such as a valley in cold regions can expose survivors to low night temperatures and windchill factors. Colder temperatures are found along valley floors which are sometimes referred to as "cold air sumps." It may be advantageous to situate campsites to take advantage of the Sun. Survivors could place their shelters in open areas during the colder months for added warmth, and in shaded areas for protection from the Sun during periods of hotter weather. In some areas a compromise may have to be made. For example, in many deserts the daytime temperatures can be very high while low temperatures at night can turn water to ice. Protection from both heat and cold are needed in these areas. Shelter type and location should be chosen to provide protection from the existing temperature conditions.

(2) Wind. Wind can be either an advantage or a disadvantage depending upon the temperature of the area and the velocity of the wind. During the summer or on warm days, survivors can take advantage of the cool breezes and protection the wind provides from insects by locating their camps on knolls or spits of land. Conversely, wind can become an annoyance or even a hazard as blowing sand, dust, or snow can cause skin and eye irritation and damage to clothing and equipment. On cold days or during winter months, survivors should seek shelter sites which are protected from the effects of windchill and drifting snow.

(3) Precipitation. The many forms of precipitation (rain, sleet, hail, or snow) can also present problems for survivors. Shelter sites should be out of major drainages and other low areas to provide protection from flash floods or mud slides resulting from heavy rains. Snow can also be a great danger if shelters are placed in potential avalanche areas.

c. Life Forms. All life forms (plant, human, and animal) must be considered when selecting the campsite and the type of shelter that will be used. The "human" factor may mean the enemy or other groups from whom survivors wish to remain undetected. Information regarding this aspect of shelters and shelter site selection is in part nine of this regulation (Evasion). For a shelter to be adequate, certain factors must be considered, especially if extended survival is expected.

(1) Insect life can cause personal discomfort, disease, and injury. By locating shelters on knolls, ridges, or any other area that has a breeze or steady wind, survivors can reduce the number of flying insects in their area. Staying away from standing water sources will help to avoid mosquitoes, bees, wasps, and hornets. Ants can be a major problem; some species will vigorously defend their territories with painful stings or bites or particularly distressing pungent odors.

(2) Large and small animals can also be a problem, especially if the camp is situated near their trails or waterholes.

(3) Dead trees that are standing, and trees with dead branches should be avoided. Wind may cause them to fall, causing injuries or death. Poisonous plants, such as poison oak or poison ivy, must also be avoided when locating a shelter.

d. Terrain. Terrain hazards may not be as apparent as weather and animal life hazards, but they can be many times more dangerous. Avalanche, rock, dry streambeds, or mud-slide areas should be avoided. These areas can be recognized by either a clear path or a path of secondary vegetation, such as 1- to 15-foot tall vegetation or other new growth which extends from the top to the bottom of a hill or mountain. Survivors should not choose shelter sites at the bottom of steep slopes which may be prone to slides. Likewise, there is a danger in camping at the bottom of steep scree or talus slopes. Additionally, rock overhang must be checked for safety before using it as a shelter.

15-3. Location:

a. Four prerequisites must be satisfied when selecting a shelter location.

(1) The first is being near water, food, fuel, and a signal or recovery site.

(2) The second is that the area be safe, providing natural protection from environmental hazards.

(3) The third is that sufficient materials be available to construct the shelter. In some cases, the "shelter" may already be present. Survivors seriously limit themselves if they assume shelters *must* be a fabricated framework having predetermined dimensions and a cover of parachute material or a signal paulin. More appropriately, survivors should consider using sheltered *places* already in existence in the immediate area. This does not rule out shelters with a fabricated framework and parachute or other manufactured material covering; it simply enlarges the scope of what can be used as a survival shelter.

(4) Finally, the area chosen must be both large enough and level enough for the survivor to lie down. Personal comfort is an important fundamental for survivors to consider. An adequate shelter provides physical and mental well-being for sound rest. Adequate rest is extremely vital if survivors are to make sound decisions. Their need for rest becomes more critical as time passes and rescue or return is delayed. Before actually constructing a shelter, survivors must determine the specific purpose of the shelter. The following factors influence the type of shelter to be fabricated.

(a) Rain or other precipitation.

(b) Cold.

(c) Heat.

(d) Insects.

(e) Available materials nearby (manufactured or natural).

(f) Length of expected stay.

(g) Enemy presence in the area—evasion "shelters" are covered in part nine of the regulation (Evasion).

(h) Number and physical condition of survivors.

b. If possible, survivors should try to find a shelter which needs little work to be adequate. Using what is already there, so that complete construction of a shelter is not necessary, saves time and energy. For example, rock overhangs, caves, large crevices, fallen logs, root buttresses, or snow banks can all be modified to provide adequate shelter. Modifications may include adding snow blocks to finish off an existing tree well shelter, increasing the insulation of the shelter by using vegetation or parachute material, etc., or building a reflector fire in front of a rock overhang or cave. Survivors must consider the amount of energy required to build the shelter. It is not really wise to spend a great deal of time and energy in constructing a shelter if nature has provided a natural shelter nearby which will satisfy the survivor's needs. See figure 15-1 for examples of naturally occurring shelters.

c. The size limitations of a shelter are important only if there is either a lack of material on hand or if it is cold. Otherwise, the shelter should be large enough to be comfortable yet not so large as to cause an excessive amount of work. Any shelter, naturally occurring or otherwise, in which a fire is to be built must have a ventilation system which will provide fresh air and allow smoke and carbon monoxide to escape. Even if a fire does not produce visible smoke (such as heat tabs), the shelter must still be vented. See figure 15-27 for placement of ventilation holes in a snow cave. If a fire is to be placed outside the shelter, the opening of the shelter should be placed 90 degrees to the prevailing wind. This will reduce the chances of sparks and smoke being blown into the shelter if the wind should reverse direction in the morning and evening. This frequently occurs

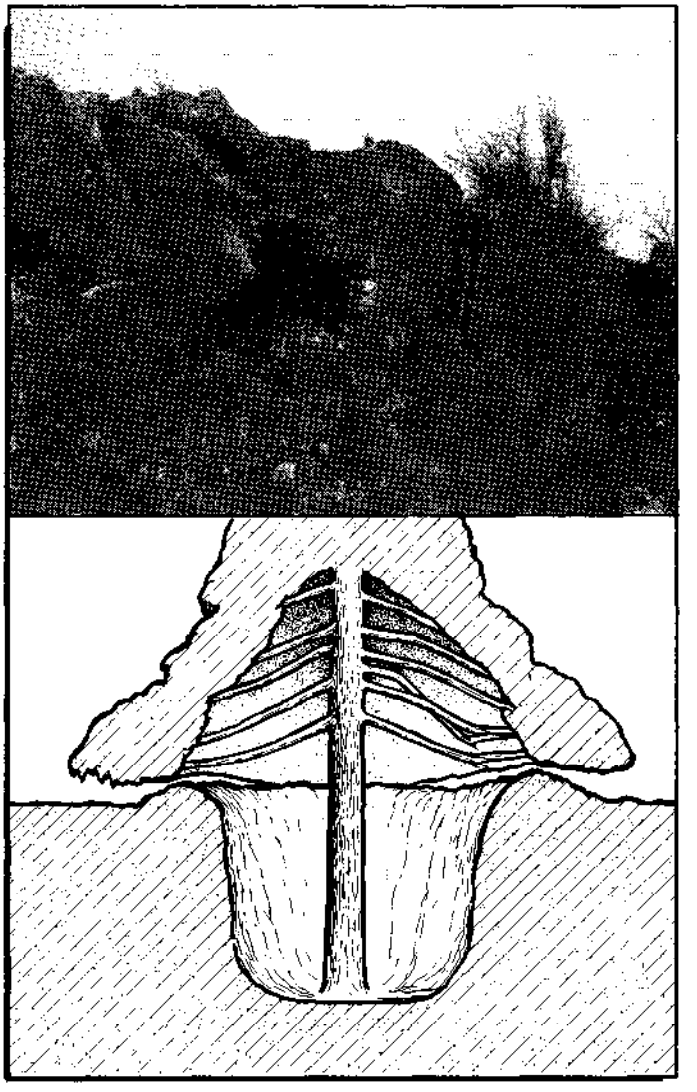

Figure 15-1. Natural Shelter.

in mountainous areas. The best fire to shelter distance is approximately 3 feet. One place where it would not be wise to build a fire is near the aircraft wreckage, especially if it is being used as a shelter. The possibility of igniting spilled lubricants or fuels is great. Survivors may decide instead to use materials from the aircraft to add to a shelter located a safe distance from the crash site.

15-4. Immediate Action Shelters. The first type of shelter that survivors may consider using, or the first type they may be forced to use, is an immediate action shelter. An immediate action shelter is one which can be erected quickly with minimum effort; for example, raft, aircraft parts, parachutes, paulin, and plastic bag. Natural formations can also shield survivors from the elements immediately, to include overhanging ledges, fallen logs, caves, and tree wells (figure 15-2). It isn't necessary to be concerned with exact shelter dimensions. Survivors should remember that if shelter is needed, use an existing shelter if at all possible. They should improvise on natural shelters or construct new shelters only if necessary. Regardless of type, the shelter must provide whatever protection is needed and, with a little ingenuity, it should be possible for survivors to protect themselves and do so quickly. In many instances, the immediate action shelters may have to serve as permanent shelters for aircrew members. For example, many aircrew members fly without parachutes, large cutting implements (axes), and entrenching tools; therefore, multiperson liferafts may be the only immediate or long-term shelter available. In this situation, multiperson liferafts must be deployed in the quickest manner possible to ensure maximum advantages are attained from the following shelter principles:

a. Set up in areas which afford maximum protection from precipitation and wind and use the basic shelter principle in paragraphs 15-2 and 15-3.

b. Anchor the raft for retention during high winds.

c. Use additional boughs, grasses, etc., for ground insulation.

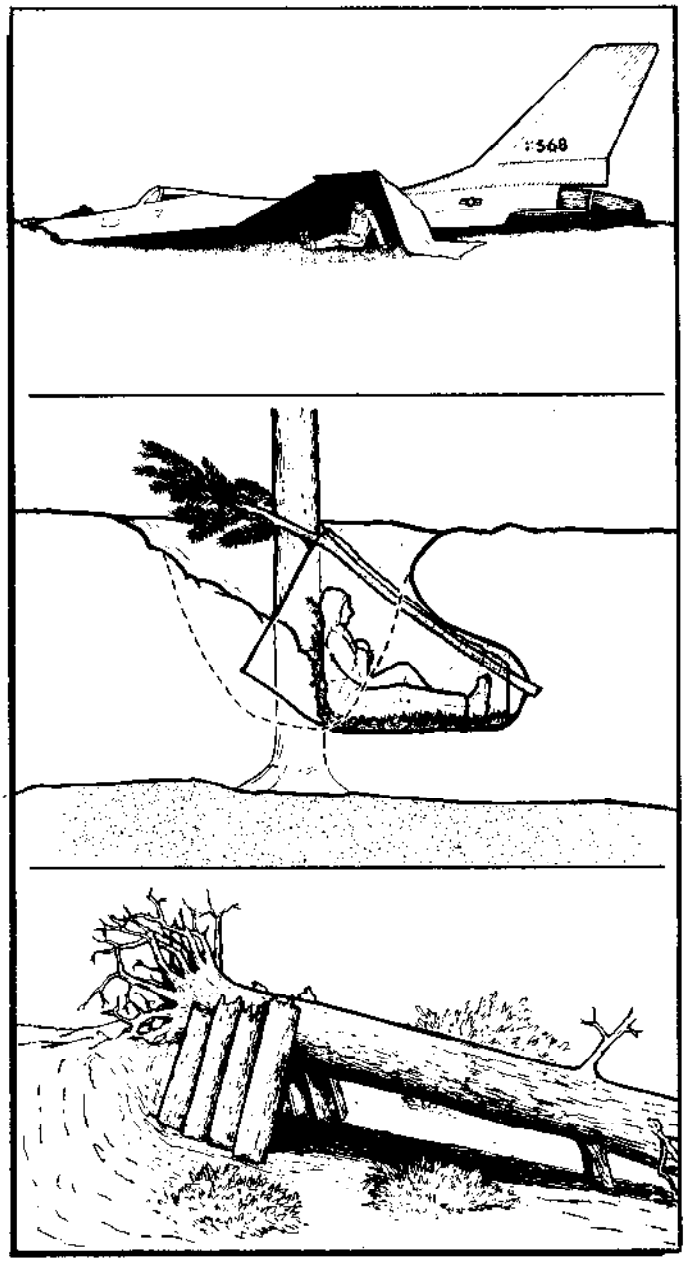

Figure 15-2. Immediate Action Shelters.

15-5. Improvised Shelters. Shelters of this type should be easy to construct and (or) dismantle in a short period of time. However, these shelters usually require more time to construct then an immediate action shelter. For this reason, survivors should only consider this type of shelter when they aren't immediately concerned with getting out of the elements. Shelters of this type include the following:

a. The "A frame" design is adaptable to all environments as it can be easily modified; for example, tropical para-hammock, temperate area "A frame," arctic thermal "A frame," and fighter trench.

b. Simple shade shelter; these are useful in dry areas.

c. Various paratepees.

d. Snow shelters; includes tree-pit shelters.

e. All other variations of the above shelter types; sod shelters, etc.

15-6. Shelters for Warm Temperature Areas:

a. If survivors are to use parachute material, they should remember that "pitch and tightness" apply to shelters designed to shed rain or snow. Parachute material is porous and will not shed moisture unless it is stretched tightly at an angle of sufficient pitch which will encourage run-off instead of penetration. An angle of 40 to 60 degrees is recommended for the "pitch" of the shelter. The material stretched over the framework

should be wrinkle-free and tight. Survivors should not touch the material when water is running over it as this will break the surface tension at that point and allow water to drip into the shelter. Two layers of parachute material, 4 to 6 inches apart, will create a more effective water repellent covering. Even during hard rain, the outer layer only lets a mist penetrate if it is pulled tight. The inner layer will then channel off any moisture which may penetrate. This layering of parachute material also creates a dead-air space that covers the shelter. This is especially beneficial in cold areas when the shelter is enclosed. Adequate insulation can also be provided by boughs, aircraft parts, snow, etc. These will be discussed in more depth in the area of cold climate shelters. A double layering of parachute material helps to trap body heat, radiating heat from the Earth's surface, and other heating sources.

b. The first step is deciding the type of shelter required. No matter which shelter is selected, the building or improvising process should be planned and orderly, following proven procedures and techniques. The second step is to select, collect, and prepare all materials needed before the actual construction; this includes framework, covering, bedding, or insulation, and implements used to secure the shelter ("dead-men," lines, stakes, etc.).

(1) For shelters that use a wooden framework, the poles or wood selected should have all the rough edges and stubs removed. Not only will this reduce the chances of the parachute fabric being ripped, but it will eliminate the chances of injury to survivors.

(2) On the outer side of a tree selected as natural shelter, some or all of the branches may be left in place as they will make a good support structure for the rest of the shelter parts.

(3) In addition to the parachute, there are many other materials which can be used as framework cover-

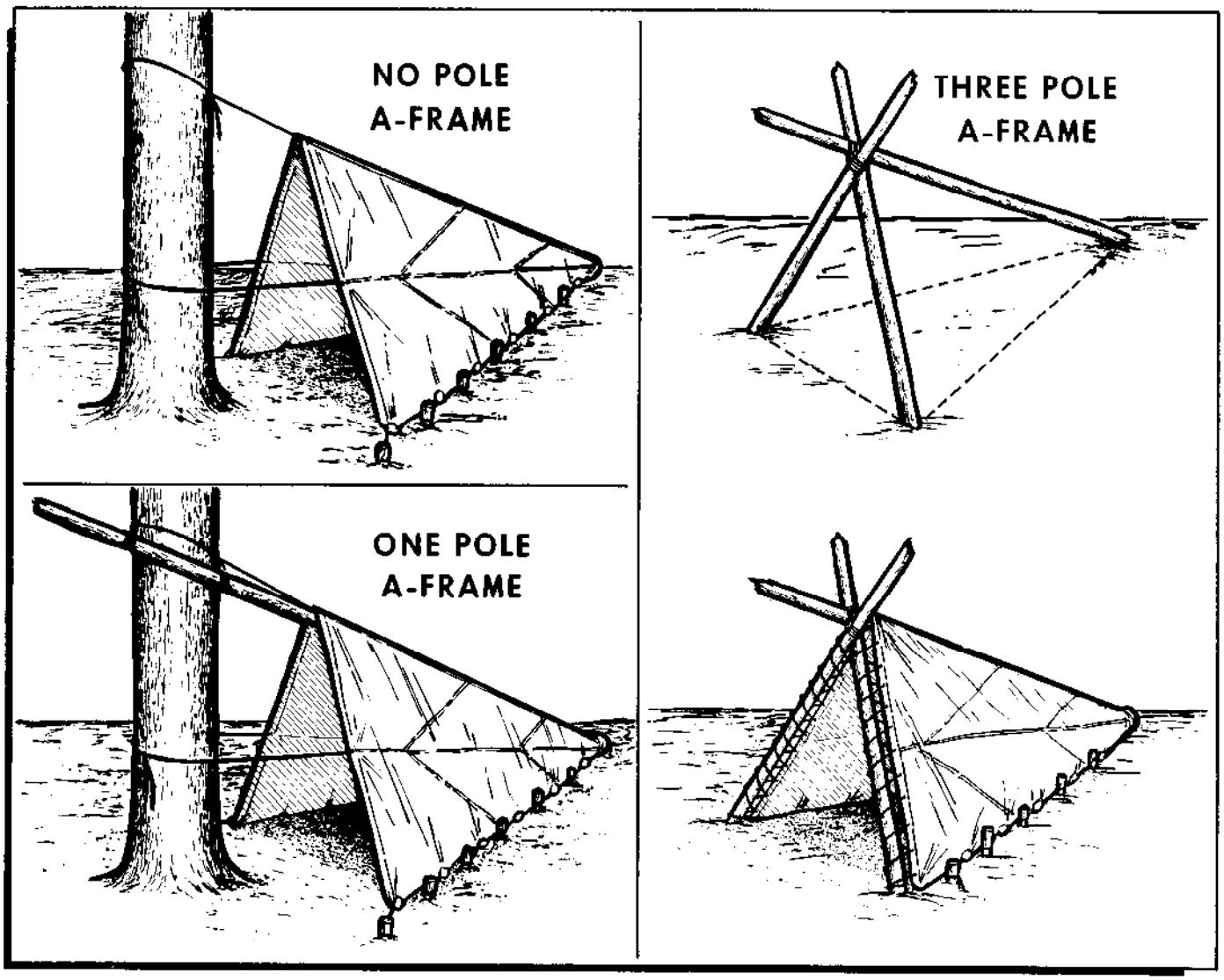

Figure 15-3. A-Frame Shelters.

ings. Some of the following are both framework and covering all in one:

(a) Bark peeled off dead trees.

(b) Boughs cut off trees.

(c) Bamboo, palm, grasses, and other vegetation cut or woven into desired patterns.

(4) If parachute material is to be used alone or in combination with natural materials, it must be changed slightly. Survivors should remove all of the lines from the parachute and then cut it to size. This will eliminate bunching and wrinkling and reduce leakage.

c. The third step in the process of shelter construction is site preparation. This includes brushing away rocks and twigs from the sleeping area and cutting back overhanging vegetation.

d. The fourth step is to actually construct the shelter, beginning with the framework. The framework is very important. It must be strong enough to support the weight of the covering and precipitation buildup of snow. It must also be sturdy enough to resist strong wind gusts.

(1) Construct the framework in one of two ways. For natural shelters, branches may be securely placed against trees or other natural objects. For parachute shelters, poles may be lashed to trees or to other poles. The support poles or branches can then be layed and (or) attached depending on their function.

(2) The pitch of the shelter is determined by the framework. A 60-degree pitch is optimum for shedding precipitation and providing shelter room.

(3) The size of the shelter is controlled by the framework. The shelter should be large enough for survivors to sit up, with adequate room to lie down and to store all personal equipment.

(4) After the basic framework has been completed, survivors can apply and secure the framework covering. The care and techniques used to apply the covering will determine the effectiveness of the shelter in shedding precipitation.

(5) When using parachute material on shelters, survivors should remove all suspension line from the material. (Excess line can be used for lashing, sewing, etc.) Next, stretch the center seam tight; then work from the back of the shelter to the front, alternating sides and securing the material to stakes or framework by using buttons and lines. When stretching the material tight, survivors should pull the material 90 degrees to the wrinkles. If material is not stretched tight, any moisture will pool in the wrinkles and leak into the shelter.

(6) If natural materials are to be used for the covering, the shingle method should be used. Starting at the bottom and working toward the top of the shelter, the bottom of each piece should overlap the top of the preceding piece. This will allow water to drain off. The material should be placed on the shelter in sufficient quantity so that survivors in the shelter cannot see through it.

15-7. Maintenance and Improvements. Once a shelter is constructed, it must be maintained. Additional modifications may make the shelter more effective and comfortable. Indian lacing (lacing the front of the shelter to the bipod) will tighten the shelter. A door may help block the wind and keep insects out. Other modifications may include a fire reflector, porch or work area, or another whole addition such as an opposing lean-to.

15-8. Construction of Specific Shelters:

a. A-Frame. The following is one way to build an A-frame shelter in a warm temperate environment using parachute material for the covering. There are as many variations of this shelter as there are builders. The procedures here will, if followed carefully, result in the completion of a safe shelter that will meet survivors' needs. For an example of this and other A-frame shelters, see figure 15-3.

(1) Materials Needed:

(a) One 12 to 18 foot long sturdy ridge pole with all projections cleaned off.

(b) Two bipod poles, approximately 7 feet long.

(c) Parachute material, normally 5 or 6 gores.

(d) Suspension lines.

(e) "Buttons," small objects placed behind gathers of material to provide a secure way of affixing suspension line to the parachute material.

(f) Approximately 14 stakes, approximately 10 inches long.

(2) Assembling the Framework:

(a) Lash (See chapter 17 — Equipment.) the two bipod poles together at eye-level height.

(b) Place the ridge pole, with the large end on the ground, into the bipod formed by the poles and secure with a square lash.

(c) The bipod structure should be 90 degrees to the ridge pole and the bipod poles should be spread out to an approximate equilateral triangle of a 60-degree pitch. A piece of line can be used to measure this.

(3) Application of Fabric:

(a) Tie off about 2 feet of the apex in a knot and tuck this under the butt end of the ridge pole. Use half hitches and clove hitches to secure the material to the base of the pole.

(b) Place the center radial seam of the parachute piece (or the center of the fabric) on the ridge pole. After pulling the material taut, use half hitches and clove hitches to secure the fabric to the front of the ridge pole.

(c) Scribe or draw a line on the ground from the butt of the ridge pole to each of bipod poles. Stake the fabric down, starting at the rear of the shelter and alternately staking from side to side to the shelter front. Use a sufficient number of stakes to ensure the parachute material is wrinkle-free.

(d) Stakes should be slanted or inclined away from the direction of pull. When tying off with a clove

hitch, the line should pass in front of the stake first and then pass under itself to allow the button and line to be pulled 90 degrees to the wrinkle.

(e) Indian lacing is the sewing or lacing of the lower lateral band with inner core or line which is secured to the bipod poles. This will remove the remaining wrinkles and further tighten the material.

(f) A rain fly, bed, and other refinements can now be added.

b. Lean-To:

(1) Materials Needed:

(a) A sturdy, smooth ridge pole (longer than the builder's body) long enough to span the distance between two sturdy trees.

(b) Support poles, 10 feet long.

(c) Stakes, suspension lines, and buttons.

(d) Parachute material (minimum of four gores).

(2) Assembling the Framework:

(a) Lash the ridge pole (between two suitable trees) about chest or shoulder high.

(b) Lay the roof support poles on the ridge pole so the roof support poles and the ground are at approximately a 60-degree angle. Lash the roof support poles to the ridge pole.

(3) Application of Fabric:

(a) Place the middle seam of the fabric on the middle support pole with lower lateral band along the ridge pole.

(b) Tie-off the middle and both sides of the lower lateral band approximately 8 to 10 inches from the ridge pole.

(c) Stake the middle of the rear of the shelter first, then alternate from side to side.

(d) The stakes that go up the sides to the front should point to the front of the shelter.

(e) Pull the lower lateral band closer to the ridge pole by indian lacing.

(f) Add bed and other refinements (reflector fire, bed logs, rain fly, etc.). See figure 15-4 for lean-to examples.

c. Paratepee, 9-Pole. The paratepee is an excellent shelter for protection from wind, rain, cold, and insects. Cooking, eating, sleeping, resting, signaling, and washing can all be done without going outdoors. The paratepee, whether 9-pole, 1-pole, or no-pole, is the only improvised shelter that provides adequate ventilation to build an inside fire. With a small fire inside, the shelter also serves as a signal at night.

(1) Materials Needed:

(a) Suspension line.

(b) Parachute material, normally 14 gores are suitable.

-1. Spread out the 14-gore section of parachute and cut off all lines at the lower lateral band, leaving about 18 inches of line attached. All other suspension lines should be stripped from the parachute.

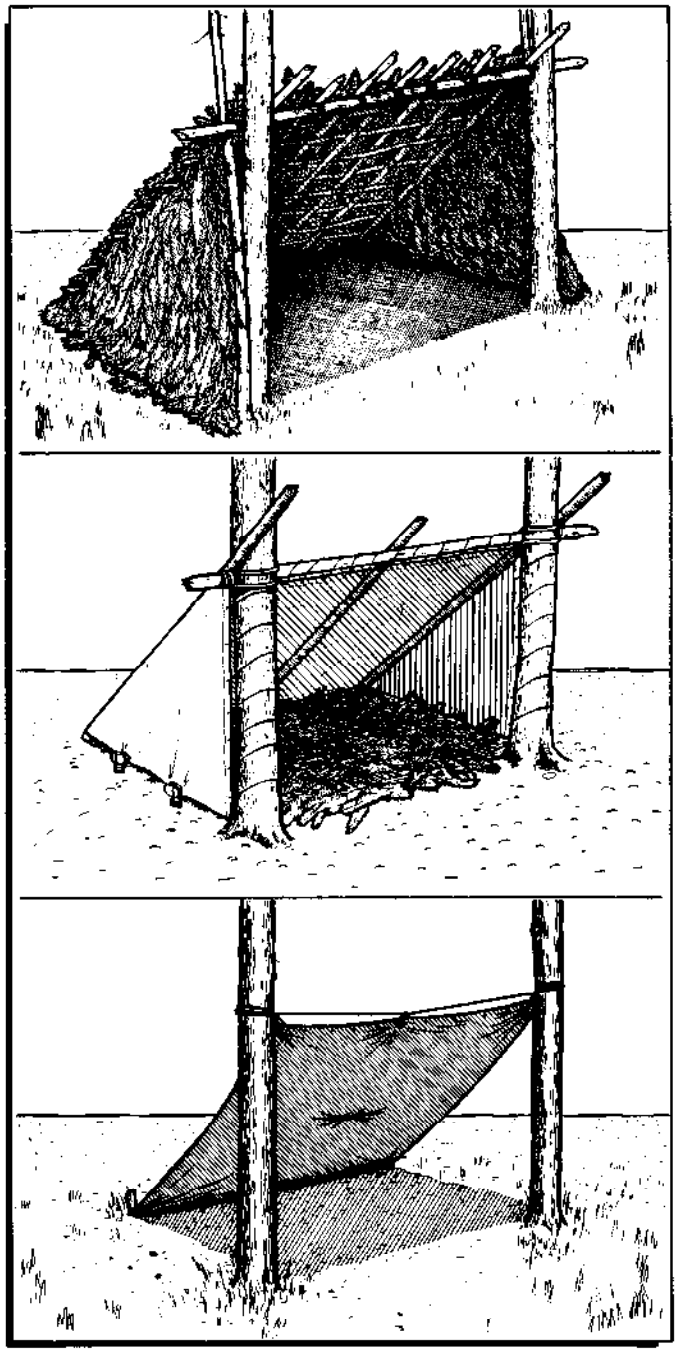

Figure 15-4. Lean-To Shelters.

-2. Sew two smoke flaps, made from two large panels of parachute material, at the apex of the 14-gore section on the outside seams. Attach suspension line with a bowline in the end to each smoke flap. The ends of the smoke flap poles will be inserted in these (see figure 15-5).

(c) Stakes.

(d) Although any number of poles may be used, 11 poles, smoothed off, each about 20 feet long, will normally provide adequate support.

(2) Assembling the Framework. (Assume 11 poles are used. Adjust instructions if different numbers are used.)

(a) Lay three poles on the ground with the butts even. Stretch the canopy along the poles. The lower lateral band should be 4 to 6 inches from the bottoms of the poles before the stretching takes place. Mark one of the poles at the apex point.

(b) Lash the three poles together, 5 to 10 inches above the marked area. (A shear lash is effective for this purpose.) These poles will form the tripod (figure 15-5).

(c) Scribe a circle approximately 12 feet in diameter in the shelter area and set the tripod so the butts of the poles are evenly spaced on the circle. Five of the remaining eight poles should be placed so the butts are evenly spaced around the 12-foot circle and the tops are laid in the apex of the tripod to form the smallest apex possible (figure 15-5).

(3) Application of Fabric:

(a) Stretch the parachute material along the tie pole. Using the suspension line attached to the middle radial seam, tie the lower lateral band to the tie pole 6 inches from the butt end. Stretch the parachute material along the middle radial seam and tie it to the tie pole using the suspension line at the apex. Lay the tie pole onto the shelter frame with the butt along the 12-foot circle and the top in the apex formed by the other poles. The tie pole should be placed directly opposite the proposed door.

(b) Move the canopy material (both sides of it) from the tie pole around the framework and tie the lower lateral band together and stake it at the door. The front can now be sewn or pegged closed, leaving 3 to 4 feet for a door. (A sewing "ladder" can be made by lashing steps up the front of the tepee (figure 15-5).

(c) Enter the shelter and move the butts of the poles outward to form a more perfect circle and until the fabric is relatively tight and smooth.

(d) Tighten the fabric and remove remaining wrinkles. Start staking directly opposite the door, and alternate from side to side, pulling the material down and to the front of the shelter. Use clove hitches or similar knots to secure material to the stakes.

(e) Insert the final two poles into the loops on the smoke flaps. The paratepee is now finished (figure 15-5).

(f) One improvement which could be made to the paratepee is the installation of a liner. This will allow a draft for a fire without making the occupants cold, since there may be a slight gap between the lower lateral band and the ground. A liner can be affixed to the inside of the paratepee by taking the remaining 14-gore piece of material and firmly staking the lower lateral band directly to the ground all the way around, leaving room for the door. The area where the liner and door meet may be sewn up. The rest of the material is brought up the inside walls and affixed to the poles with buttons (figure 15-5).

d. Paratepee, 1-Pole:

(1) Materials Needed:

(a) Normally use a 14-gore section of canopy, strip the shroud lines leaving 16- to 18-inch lengths at the lower lateral band.

(b) Stakes.

(c) Inner core and needle.

(2) Construction of the 1-Pole Paratepee:

(a) Select a shelter site and scribe a circle about 14 feet in diameter on the ground.

(b) The parachute material is staked to the ground using the lines attached at the lower lateral band. After deciding where the shelter door will be located, stake the first line (from the lower band) down securely. Proceed around the scribed line and stake down all the lines from the lateral band, making sure the parachute material is stretched taut before the line is staked down.

(c) Once all the lines are staked down, loosely attach the center pole, and, through trial and error, determine the point at which the parachute material will be pulled tight once the center pole is placed upright—securely attach the material at this point.

(d) Using a suspension line (or innercore), sew the end gores together leaving 3 or 4 feet for a door (figure 15-6).

e. Paratepee, No-Pole. For this shelter, the 14 gores of material are prepared the same way. A line is attached to the apex and thrown over a tree limb, etc., and tied off. The lower lateral band is then staked down starting opposite the door around a 12- to 14-foot circle. (See figure 15-7 for paratepee example.)

f. Sod Shelter. A framework covered with sod provides a shelter which is warm in cold weather and one that is easily made waterproof and insect-proof in the summer. The framework for a sod shelter must be strong, and it can be made of driftwood, poles, willow, etc. (Some natives use whale bones.) Sod, with a heavy growth of grass or weeds, should be used since the roots tend to hold the soil together. Cutting about 2 inches of soil along with the grass is sufficient. The size of the blocks are determined by the strength of the individual. A sod house is strong and fireproof.

15-9. Shelter for Tropical Areas. Basic considerations for shelter in tropical areas are as follows:

a. In tropical areas, especially moist tropical areas, the major environmental factors influencing both site selection and shelter types are:

(1) Moisture and dampness.

(2) Rain.

(3) Wet ground.

(4) Heat.

(5) Mud-slide areas.

(6) Dead standing trees and limbs.

(7) Insects.

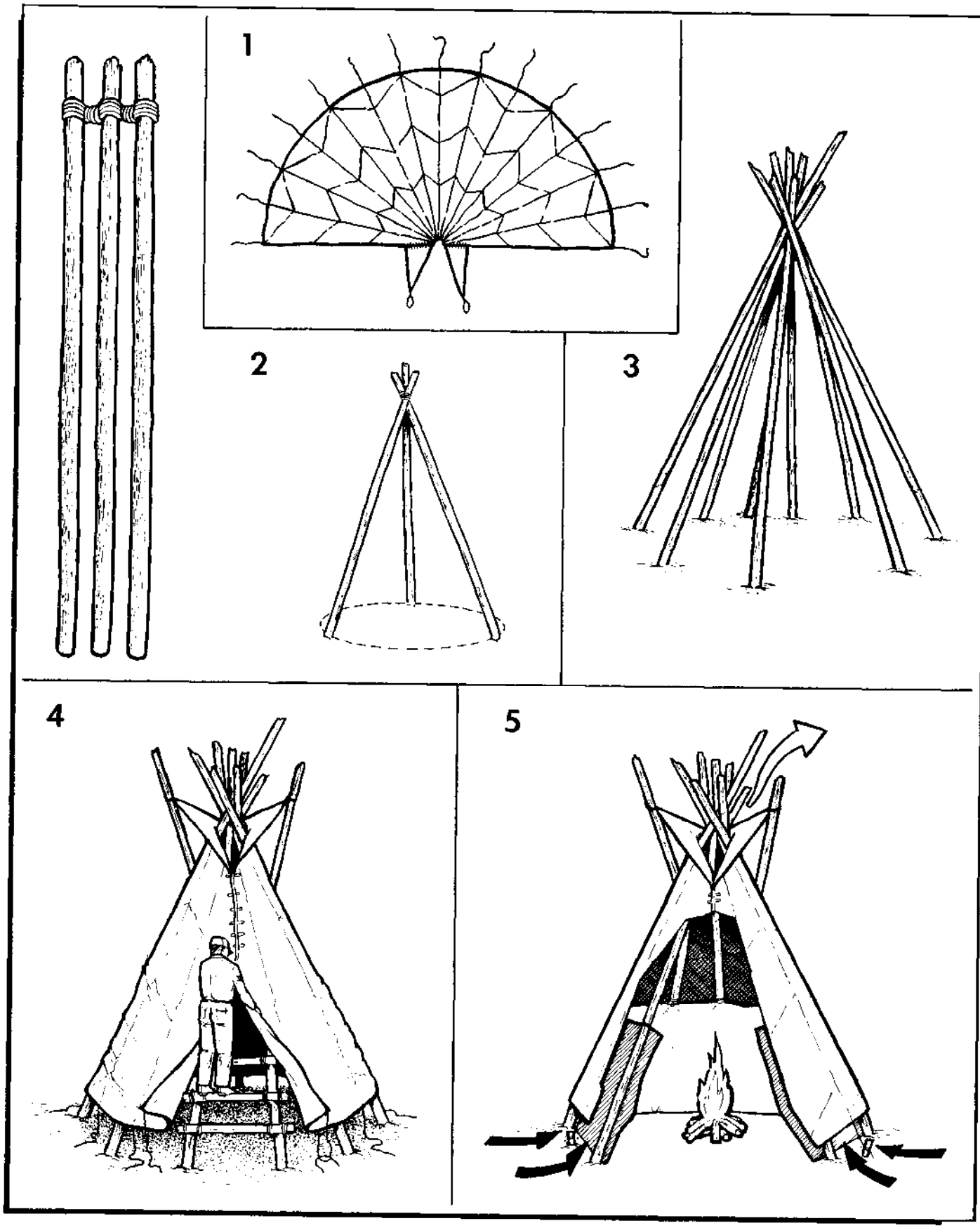

Figure 15-5. 9-Pole Tepee.

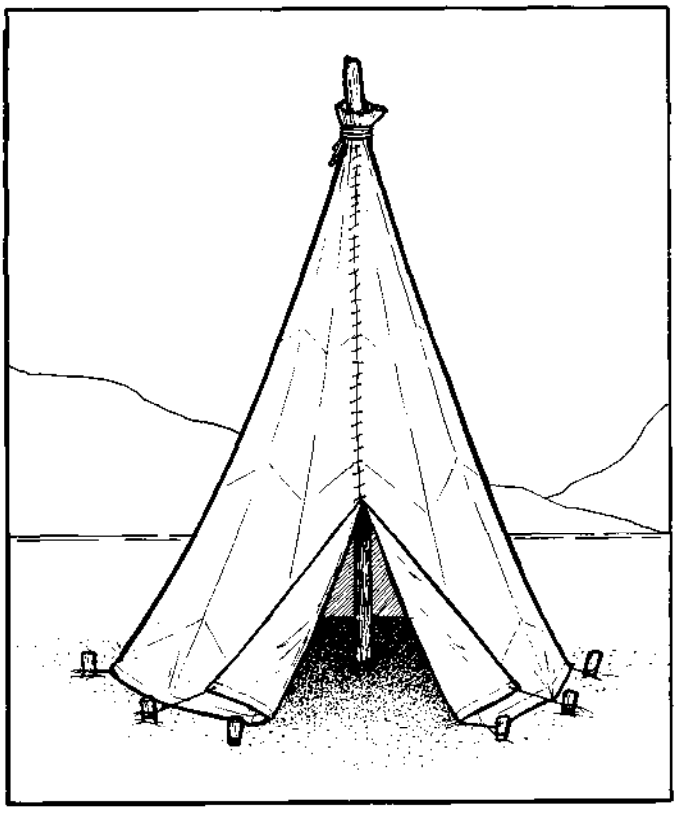

Figure 15-6. 1-Pole Tepee.

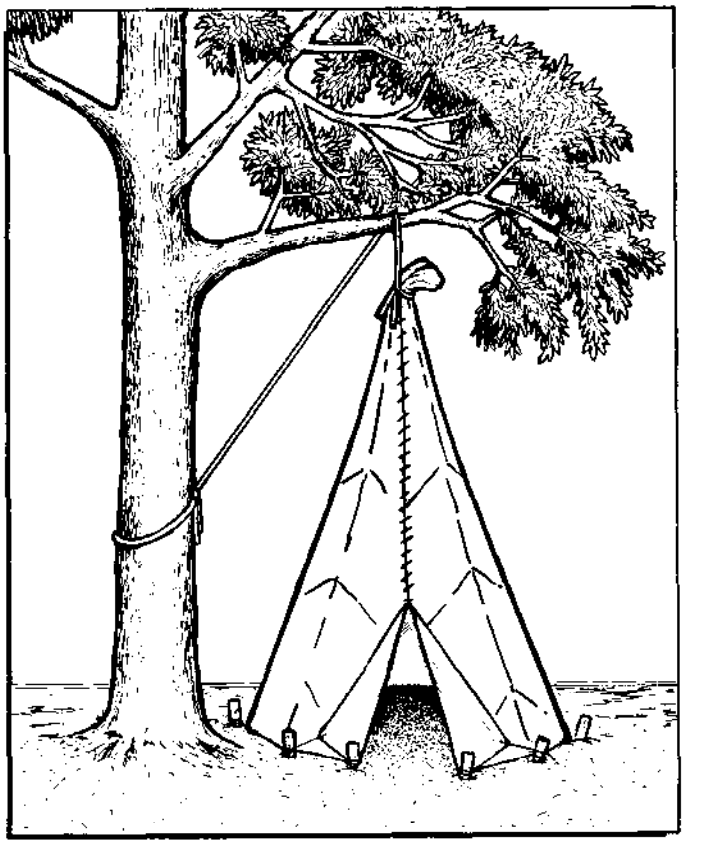

Figure 15-7. No-Pole Tepee.

b. Survivors should establish a campsite on a knoll or high spot in an open area well back from any swamps or marshy areas. The ground in these areas is drier, and there may be a breeze which will result in fewer insects.

c. Underbrush and dead vegetation should be cleared from the shelter site. Crawling insects will not be able to approach survivors as easily due to lack of cover.

d. A thick bamboo clump or matted canopy of vines for cover reflects the smoke from the campfire and discourages insects. This cover will also keep the extremely heavy early morning dew off the bedding.

e. The easiest improvised shelter is made by draping a parachute, tarpaulin, or poncho over a rope or vine stretched between two trees. One end of the canopy should be kept higher than the other; insects are discouraged by few openings in shelters and smudge fires. A hammock made from parachute material will keep the survivor off the ground and discourage ants, spiders, leeches, scorpions, and other pests.

f. In the wet jungle, survivors need shelter from dampness. If they stay with the aircraft, it should be used for shelter. They should try to make it mosquito-proof by covering openings with netting or parachute cloth.

g. A good rain shelter can be made by constructing an A-type framework and shingling it with a good thickness of palm or other broad leaf plants, pieces of bark, and mats of grass (figure 15-8).

h. Nights are cold in some mountainous tropical areas. Survivors should try to stay out of the wind and build a fire. Reflecting the heat off a rock pile or other barrier is a good idea. Some natural materials which can

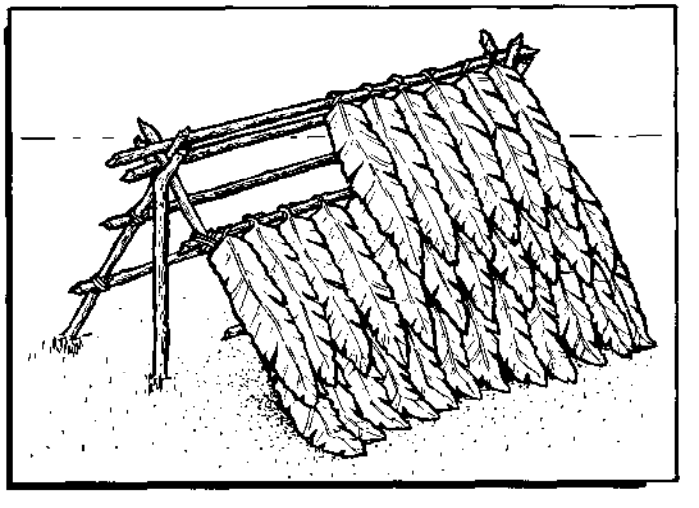

Figure 15-8. Banana Leaf A-Frame.

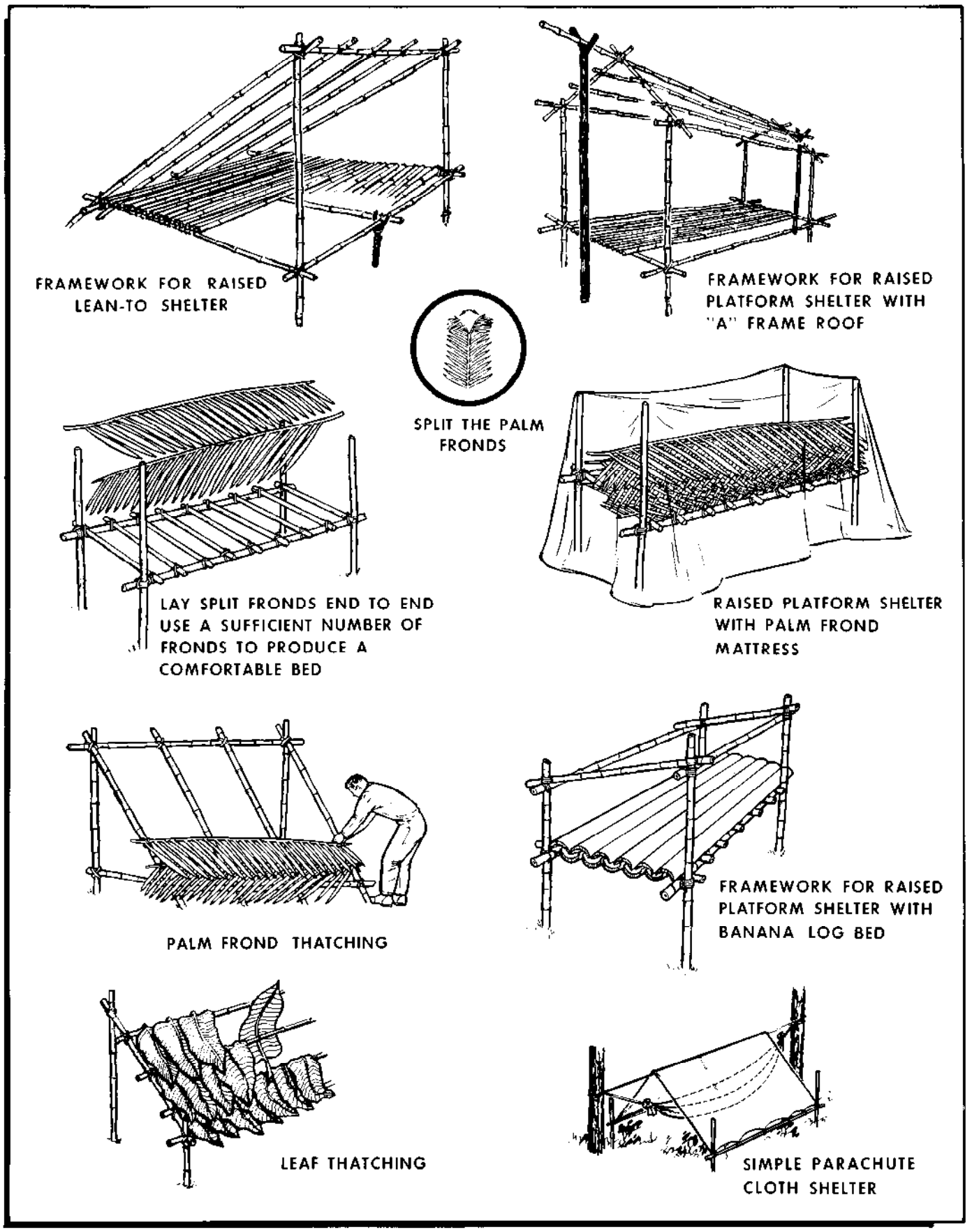

Figure 15-9. Raised Platform shelter.

be used in the shelters are green wood (dead wood may be too rotten), bamboo, and palm leaves. Vines can be used in place of suspension line for thatching roofs or floors, etc. Banana plant sections can be separated from the banana plant and fashioned to provide a mattress effect.

15-10. Specific Shelters for Tropical Environments:

a. Raised Platform Shelter (figure 15-9). This shelter has many variations. One example is four trees or vertical poles in a rectangular pattern which is a little longer and a little wider than the survivor, keeping in mind the survivor will also need protection for equipment. Two long, sturdy poles are then square lashed between the trees or vertical poles, one on each side of the intended shelter. Cross pieces can then be secured across the two horizontal poles at 6- to 12-inch intervals. This forms the platform on which a natural mattress may be constructed. Parachute material can be used as an insect net and a roof can be built over the structure using A-frame building techniques. The roof should be waterproofed with thatching laid bottom to top in a thick shingle fashion. See figure 15-9 for examples of this and other platform shelters. These shelters can also be built using three trees in a triangular pattern. At the foot of the shelter, two poles are joined to one tree.

b. Variation of Platform Shelter. A variation of the platform-type shelter is the paraplatform. A quick and comfortable bed is made by simply wrapping material around the two "frame" poles. Another method is to roll poles in the material in the same manner as for an improvised stretcher (figure 15-10).

c. Hammocks. Various parahammocks can also be made. They are more involved than a simple parachute wrapped framework and not quite as comfortable (figure 15-11).

d. Hobo Shelter. On tropical coasts and other coastal environments, if a more permanent shelter is desired as opposed to a simple shade shelter, survivors should build a "hobo" shelter. To build this shelter:

(1) Dig into the lee side of a sand dune to protect the shelter from the wind. Clear a level area large enough to lie down in and store equipment.

(2) After the area has been cleared, build a heavy driftwood framework which will support the sand.

(3) Wall sides and top with strong material (boards, driftwood, etc.) that will support the sand; leave a door opening.

(4) Slope the roof to equal the slope of the sand dune. Cover the entire shelter with parachute material to keep sand from sifting through small holes in the walls and roof.

(5) Cover with 6 to 12 inches of sand to provide protection from wind and moisture.

(6) Construct a door for the shelter (figure 15-12).

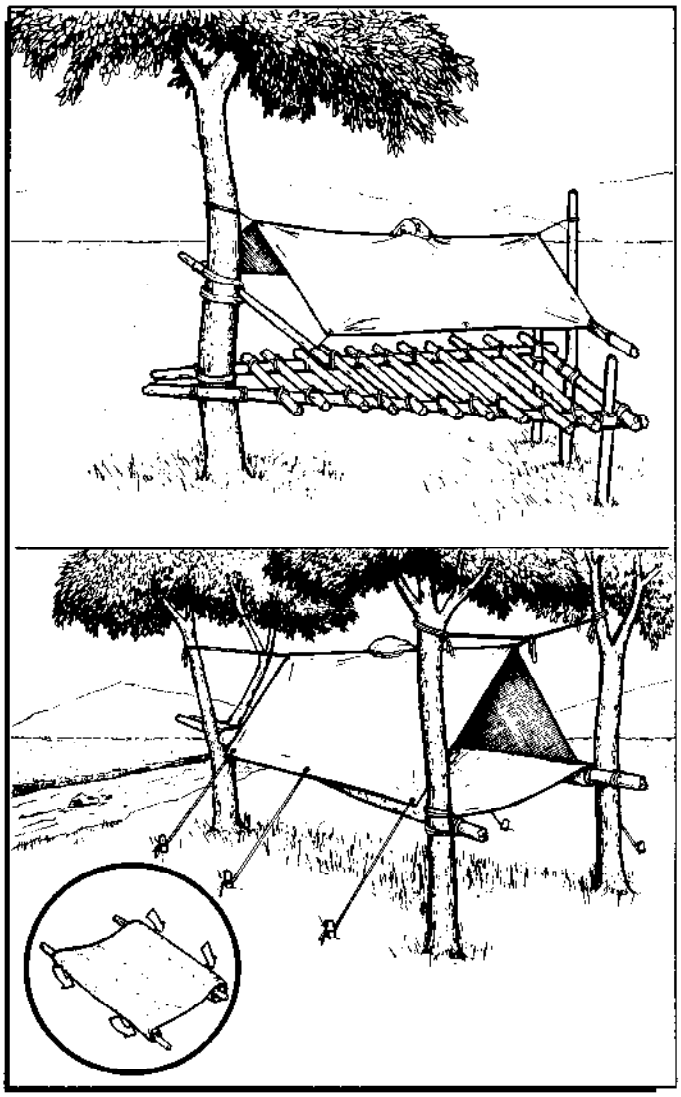

Figure 15-10. Raised Paraplatform Shelter.

15-11. Shelters for Dry Climates:

a. Natives of hot, dry areas make use of light-proof shelters with sides rolled up to take advantage of any breeze. Survivors should emulate these shade-type shelters if forced to survive in these areas. The extremes of heat *and* cold must be considered in hot areas, as most can become very cold during the night. The major problem for survivors will be escaping the heat and Sun rays.

b. Natural shelters in these areas are often limited to the shade of cliffs and the lee sides of hills, dunes, or rock formations. In some desert mountains, it is possible to find good rock shelters or cave-like protection under tumbled blocks of rocks which have fallen from cliffs. Use care to ensure that these blocks are in areas void of future rock falling activity and free from animal hazards.

c. Vegetation, if any exists, is usually stunted and armed with thorns. It may be possible to stay in the

shade by moving around the vegetation as the Sun moves. The hottest part of the day may offer few shadows because the Sun is directly overhead. Parachute material draped over bushes or rocks will provide some shade.

d. Materials which can be used in the construction of desert shelters include:

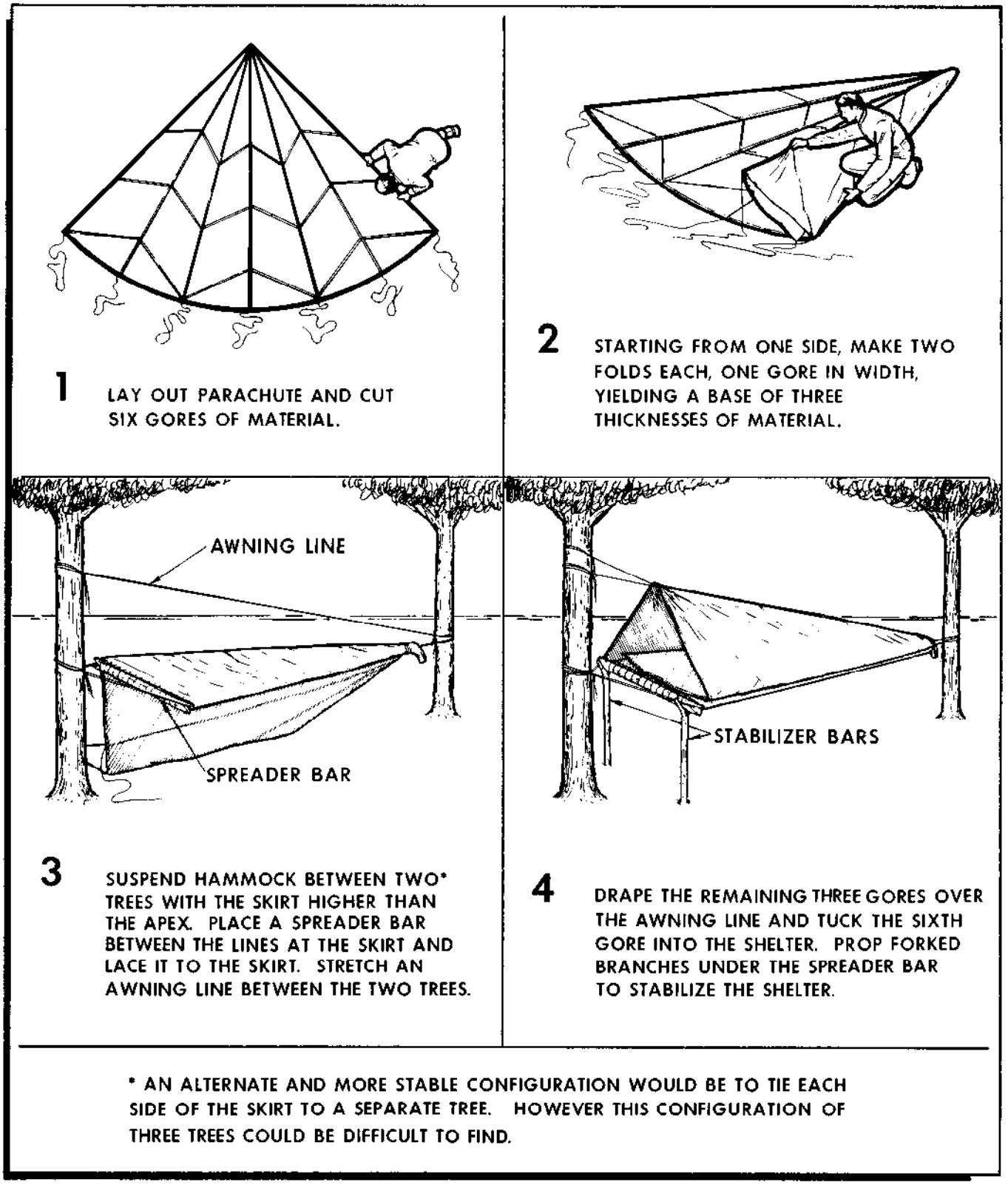

Figure 15-11. Parahammock.

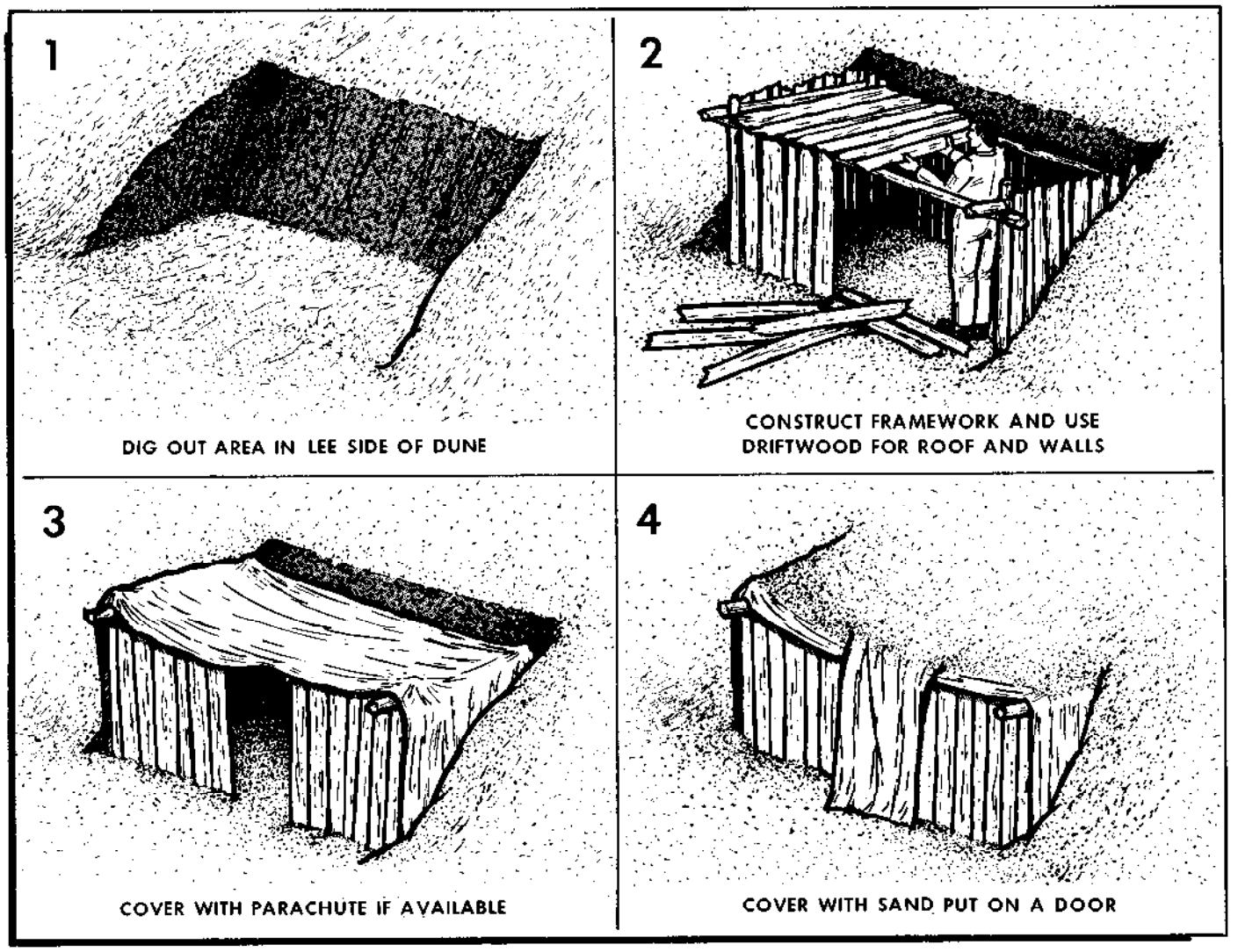

Figure 15-12. Hobo Shelter.

(1) Sand, though difficult to work with when loose, may be made into pillars by using sandbags made from parachute or any available cloth.

(2) Rock can be used in shelter construction.

(3) Vegetation such as sage brush, creosote bushes, juniper trees, and desert gourd vines are valuable building materials.

(4) Parachute canopy and suspension lines. These are perhaps the most versatile building materials available for use by survivors. When used in layers, parachute material protects survivors from the Sun's rays.

(a) The shelter should be made of dense material or have numerous layers to reduce or stop dangerous ultraviolet rays. The colors of the parachute materials used make a difference as to how much protection is provided from ultraviolet radiation. As a general rule, the order of preference should be to use as many layers as practical in the order of orange, green, tan, and white.

ULTRAVIOLET TESTS ON PARACHUTE CANOPY MATERIAL

% Ultraviolet (Short Wave 2537 A° Sunburn Rays) Blocked as compared to Direct Exposure

	1 Layer	2 Layers	3 Layers
Orange	78.2%	96.2%	99.36%
Sage Green	79.5%	96.2%	98.7%
Tan	64.1%	84.6%	93.6%
White	47.5%	61.6%	70.5%

% Ultraviolet (Long Wave 3660 A°) Blocked as Compared to Direct Exposure

	1 Layer	2 Layers	3 Layers
Orange	63.4%	92.3%	97.8%
Sage Green	60.0%	88.9%	97.8%
Tan	38.9%	66.7%	82.3%
White	28.9%	47.8%	58.9%

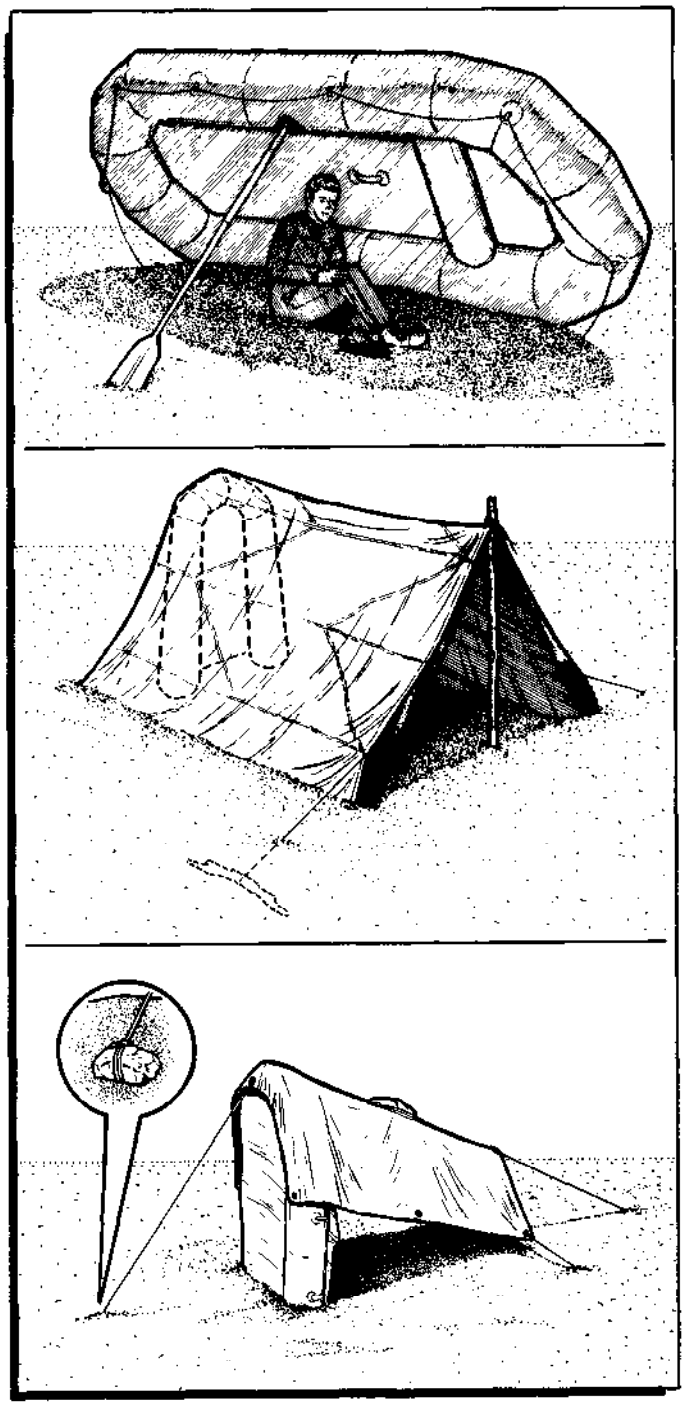

Figure 15-13. Improvised Natural Shade Shelters.

(b) The material should be kept approximately 12 to 18 inches above the individual. This allows the air to cool the underside of the material.

(c) Aircraft parts and liferafts can also be used for shade shelters. Survivors may use sections of the wing, tail, or fuselage to provide shade. However, the interior of the aircraft will quickly become superheated and should be avoided as a shelter. An inflatable raft can be tilted against a raft paddle or natural object such as a bush or rock to provide relief from the Sun (figure 15-13).

15-12. Principles of Desert Shelters:

a. The roof of a desert shelter should be multilayered so the resulting airspace reduces the inside temperature of the shelter. The layers should be separated 12 to 18 inches apart (figure 15-14).

b. Survivors should place the floor of the shelter about 18 inches above or below the desert surface to increase the cooling effect.

c. In warmer deserts, white parachute material should be used as an outer layer. Orange or sage green material should be used as an inner layer for protection from ultraviolet rays.

d. In cooler areas, multiple layers of parachute material should be used with sage green or orange material as the outer layer to absorb heat.

e. The sides of shelters should be movable in order to protect survivors during cold and (or) windy periods and to allow for ventilation during hot periods.

f. In a hot desert, shelters should be built away from large rocks which store heat during the day. Survivors may need to move to the rocky areas during the evening to take advantage of the warmth heated rocks radiate.

g. Survivors should:

(1) Build shelters on the windward sides of dunes for cooling breezes.

(2) Build shelters during early morning, late evening, or at night. However, potential survivors should recall that survivors who come down in a desert area during daylight hours must be immediately concerned with protection from the Sun and loss of water. In this case, parachute canopy material can be draped over liferaft, vegetation, or a natural terrain feature for quick shelter.

15-13. Shelters for Snow and Ice Areas:

a. The differences in arctic and arctic-like environments create the need for different shelters. Basically, there are two types of environments which may require special shelter characteristics or building principles before survivors will have adequate shelter. They are:

(1) Barren lands which include some seacoasts, icecaps, sea ice areas, and areas above the tree line.

(2) Tree-line areas.

b. Barren lands offer a limited variety of materials for shelter construction. These are snow, small shrubs, and grasses. Ridges formed by drifting or wind-packed snow may be used for wind protection (survivors should build on the lee side). In some areas, such as sea ice, windy conditions usually exist and cause the ice to shift forming pressure ridges. These areas of unstable ice and snow should be avoided at all times. Shelters which are suitable for barren-type areas include:

(1) Molded dome (figure 15-15).

(2) Snow cave (figure 15-16).

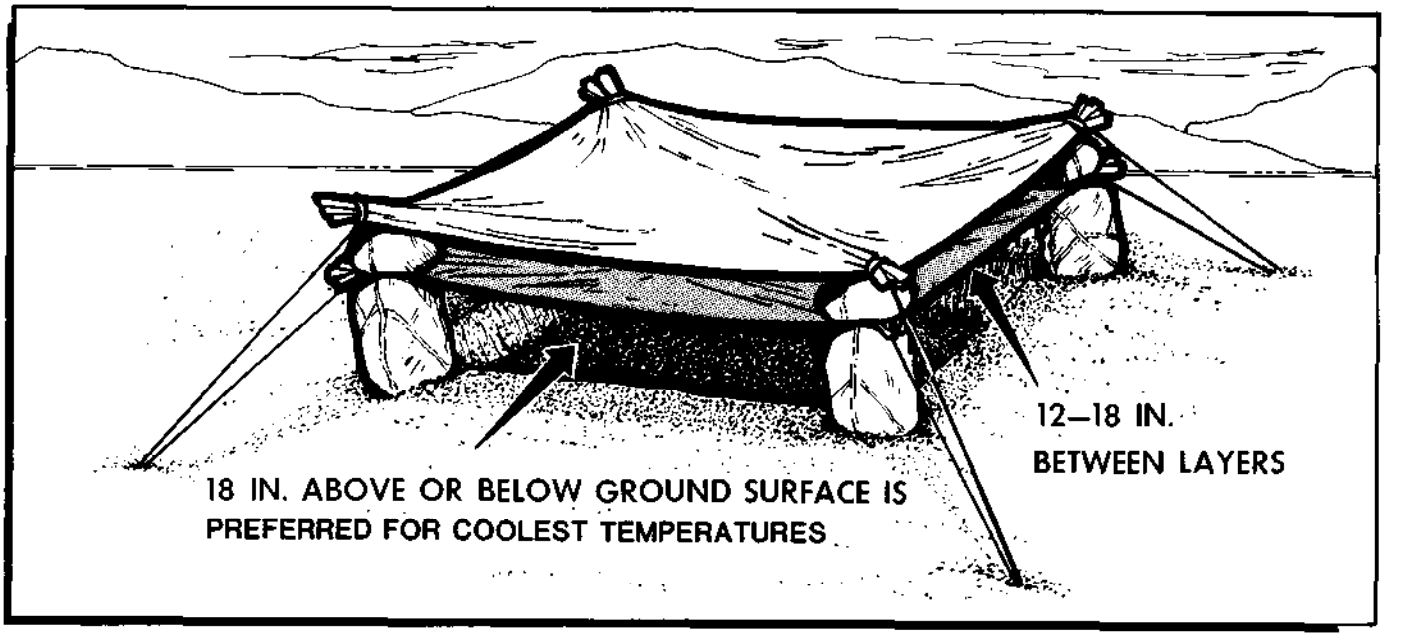

Figure 15-14. Parachute Shade Shelter.

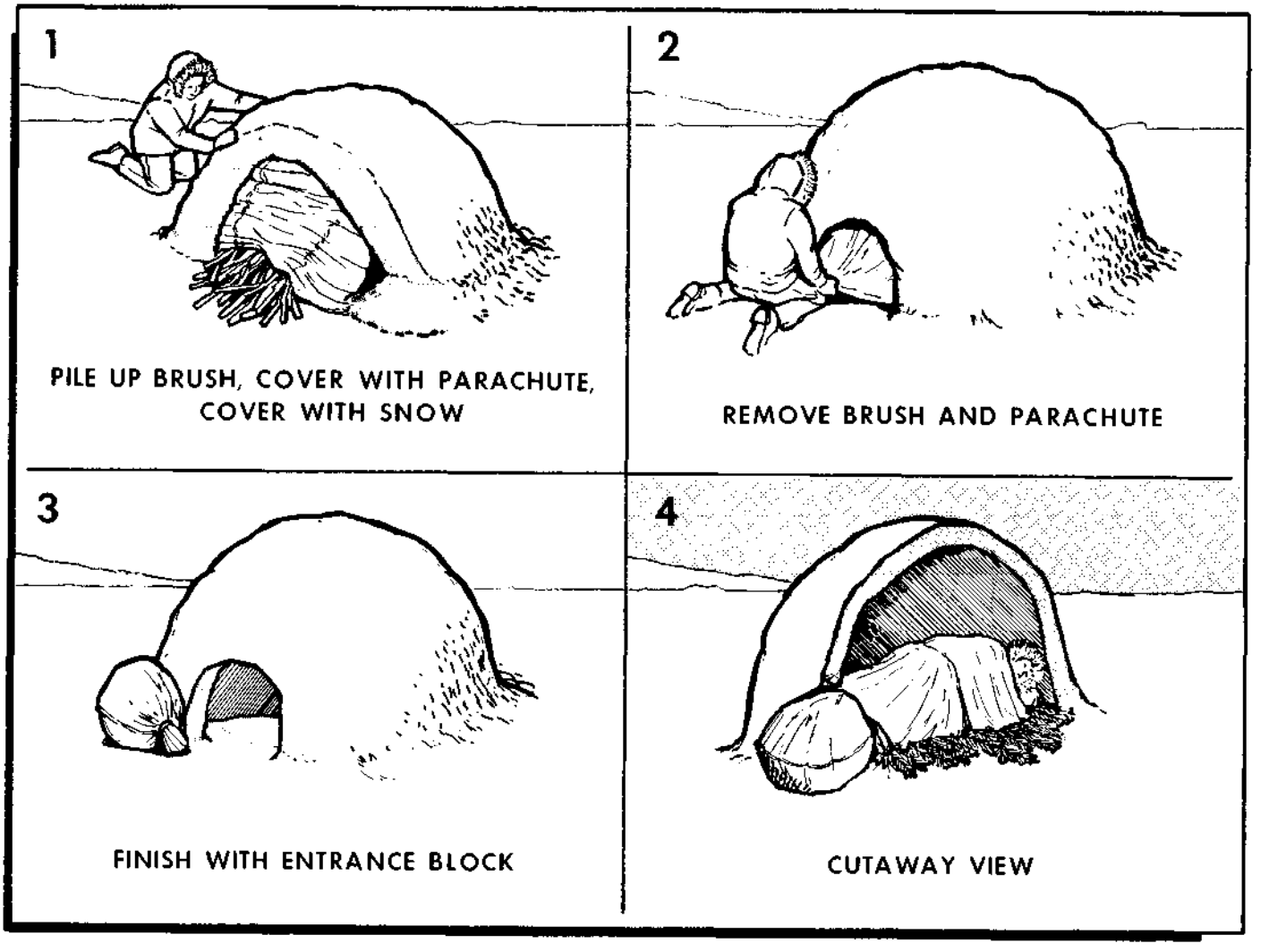

Figure 15-15. Molded Dome Shelter.

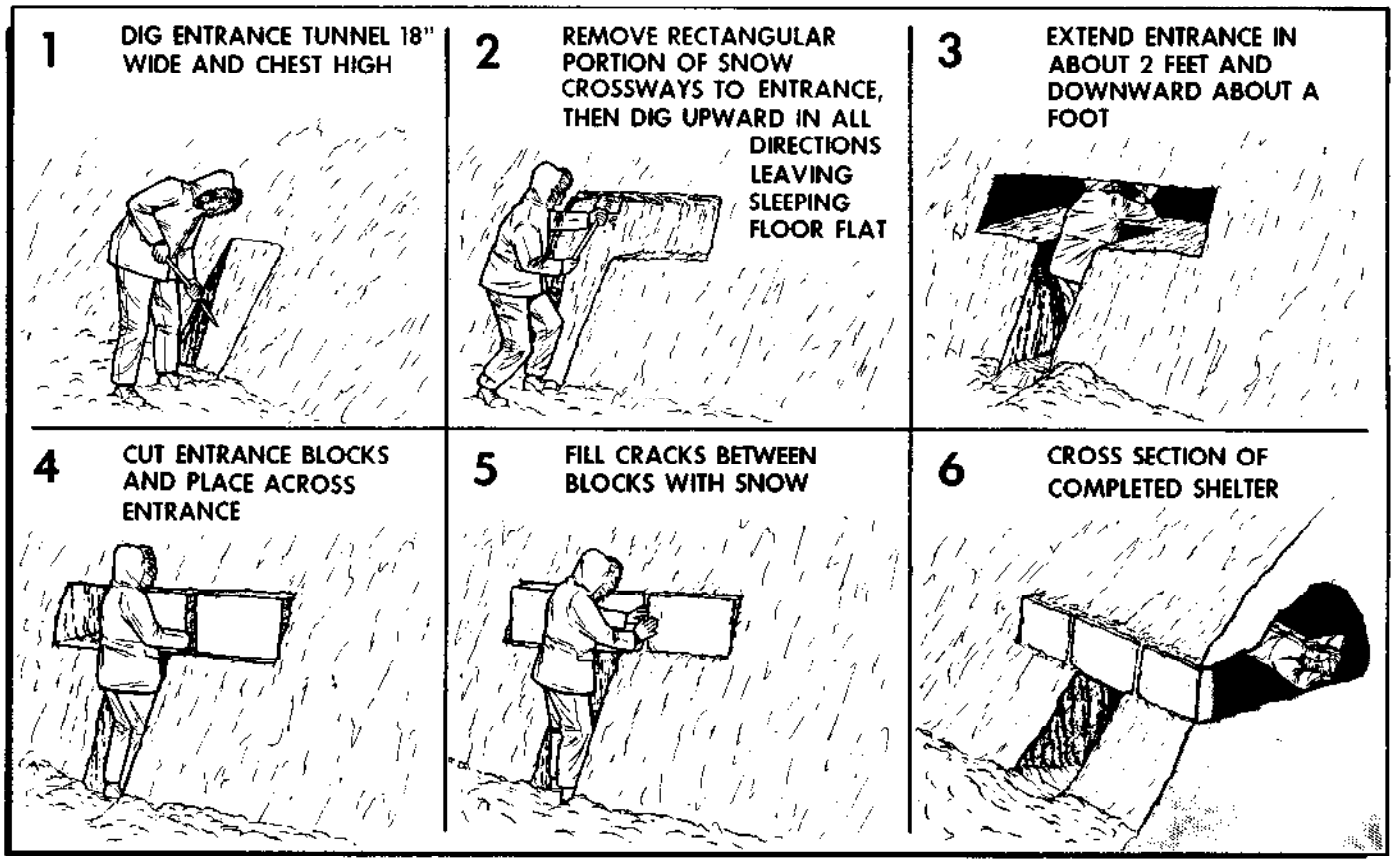

Figure 15-16. Snow Cave.

(3) Fighter trench (figure 15-17).

(4) Igloo (figure 15-18).

(5) Para-snow house (Figure 15-19).

NOTE: Of these, the ones that are quick to construct and require minimum effort and energy are the molded dome, snow cave, and fighter trench. It is important to know which of these shelters is the easiest to build since reducing or eliminating the effect of the windchill factor is essential to remaining alive.

c. In tree-covered areas, sufficient natural shelter building materials are normally available. Caution is required. Shelters built near rivers and streams may get caught in the overflow.

d. Tree-line area shelter types include:

(1) Thermal A-Frame construction (figure 15-20).

(2) Lean-to or wedge (figure 15-21).

(3) Double lean-to (figure 15-22).

(4) Fan (figure 15-23).

(5) Willow frame (figure 15-24).

(6) Tree well (figure 15-25).

e. Regardless of the type of shelter used, the use of thermal principles and insulation in arctic shelters is required. Heat radiates from bare ground and from ice masses over water. This means that shelter areas on land should be dug down to bare earth if possible (figure 15-26). A minimum of 8 inches of insulation above survivors is needed to retain heat. All openings except ventilation holes should be sealed to avoid heat loss. Leaving vent holes open is especially important if heat producing devices are used. Candles, sterno, or small oil lamps produce carbon monoxide. In addition to the ventilation hole through the roof, another may be required at the door to ensure adequate circulation of the air. (As a general rule, unless persons can see their breath, the snow shelter is too warm and should be cooled down to proclude melting and dripping.)

f. Regardless of how cold it may get outside, the temperature inside a small well-constructed snow cave will probably not be lower than -10°F. Body heat alone can raise the temperature of a snow cave 45 degrees above the outside air. A burning candle will raise the temperature 4 degrees. Burning Sterno (small size, 2⅝ oz) will raise the cave temperature about 28 degrees. However, since they cannot be heated many degrees above freezing, snow shelters provide a rather rugged life. Once the inside of the shelter "glazes" over with ice, this layer of ice should be removed by chipping it off or a new shelter built since ice reduces the insulating quality of a shelter. Maintain the old shelter until the new one is constructed. It will provide protection from the wind.

g. The aircraft should not be used as a shelter when temperatures are below freezing except in high wind conditions. Even then a thermal shelter should be constructed as soon as the conditions improve. The aircraft

will not provide adequate insulation, and the floor will usually become icy and hazardous.

15-14. General Construction Techniques:

a. All thermal shelters use a layering system consisting of the frame, parachute (if available), boughs or shrubs, and snow. The framework must be sturdy enough to support the cover and insulation. A door block should be used to minimize heat loss. Insulation should be added on sleeping areas.

b. If a barren land-type shelter is being built with snow as the only material, a long knife or digging tool is a necessity. It normally takes 2 to 3 hours of hard work to dig a snow cave, and much longer for the novice to build an igloo.

c. Survivors should dress lightly while digging and working; they can easily become overheated and dampen their clothing with perspiration which will rapidly turn to ice.

d. If possible, all shelter types should have their openings 90 degrees to the prevailing wind. The entrance to the shelter should also be screened with snowblocks stacked in a L-shape.

e. Snow on the sea ice, suitable for cutting into blocks, will usually be found in the lee of pressure ridges or ice hummocks. The packed snow is often so shallow that the snowblocks have to be cut out horizontally.

f. No matter which shelter is used, survivors should take a digging tool into the shelter at night to cope with the great amount of snow which may block the door during the night.

15-15. Shelter Living:

a. Survivors should limit the number of shelter entrances to conserve heat. Fuel is generally scarce in the

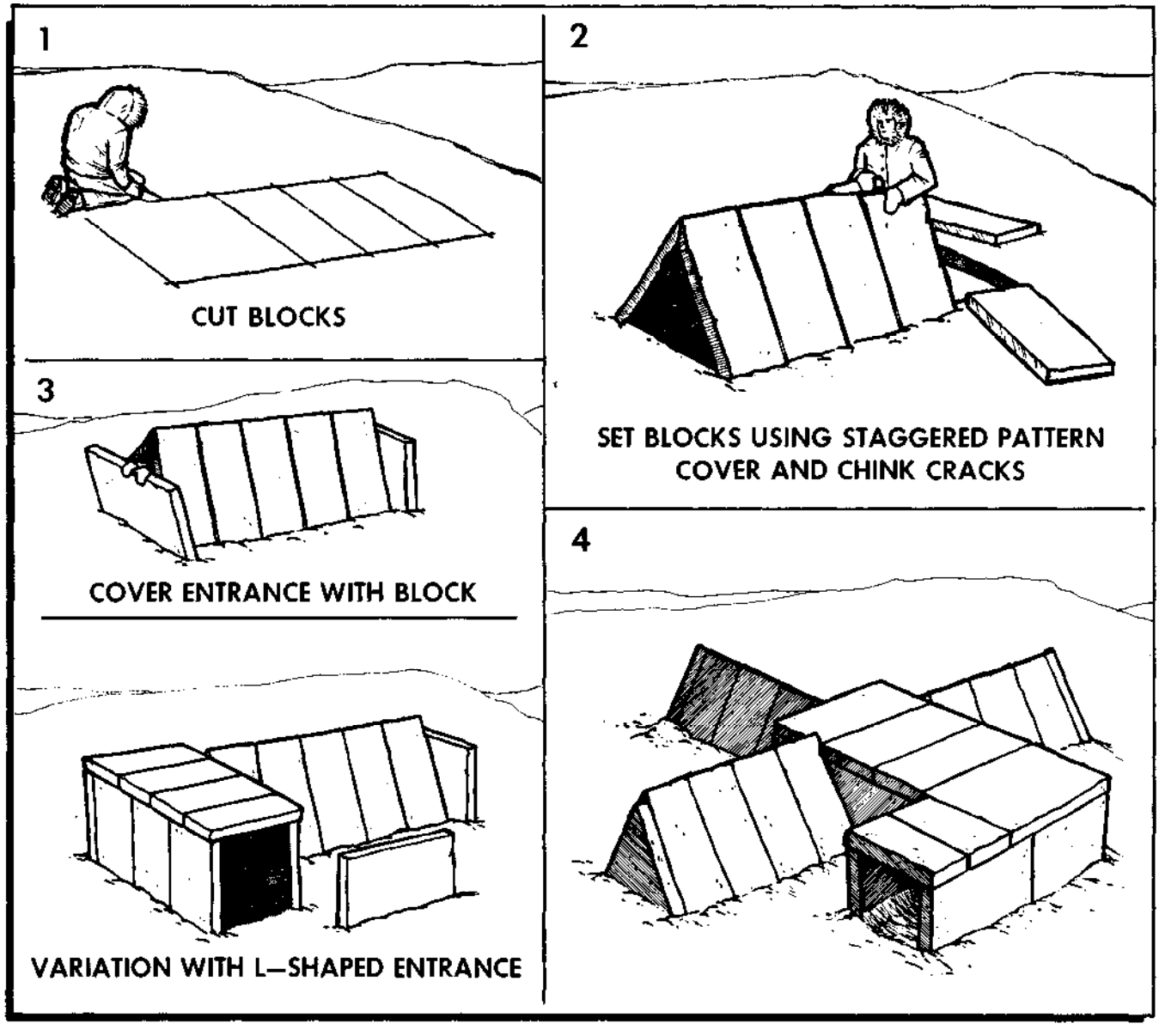

Figure 15-17. Fighter Trench.

arctic. To conserve fuel, it is important to keep the shelter entrance sealed as much as possible (figure 15-27). When it is necessary to go outside the shelter, activities such as gathering fuel, snow or ice for melting, etc., should be done. To expedite matters, a trash receptacle may be kept inside the door, and equipment may be stored in the entry way. Necessities which cannot be stored inside may be kept just outside the door. Any firearms (guns) the survivor may have must be stored outside the shelter to prevent condensation building which could cause them to malfunction.

b. A standard practice in snow shelter living is for people to relieve themselves indoors when possible. This practice conserves body heat. If the snowdrift is large enough to dig connecting snow caves, one may be used as a toilet room. If not, tin cans may be used for urinals, and snowblocks for solid waste (fecal) matter.

c. Survivors should use thick insulation under themselves when sleeping or resting even if they have a sleeping bag. They can use a thick bough bed in shingle-fashion, seat cushions, parachute, or an inverted and inflated rubber raft.

d. Outer clothing makes good mattress material. A parka makes a good footbag. The shirt and inner trousers may be rolled up for a pillow. Socks and insoles can be separated and aired in the shelter. Drying may be completed in the sleeping bag by stowing around the hips. This drying method should only be used as a last resort.

e. Keeping the sleeping bag clean, *dry*, and fluffed will give maximum warmth. To dry the bag, it should be turned inside out, frost beaten out, and warmed before the fire—taking care that it doesn't burn.

f. To keep moisture (from breath) from wetting the sleeping bag, a moisture cloth should be improvised from a piece of clothing, a towel, or parachute fabric. It can then be lightly wrapped around the head in such a way that the breath is trapped inside the cloth. A piece of fabric dries easier than a sleeping bag. If cold is experienced during the night, survivors should exercise by fluttering their feet up and down or by beating the inside of the bag with their hands. Food or hot liquids can be helpful.

g. Snow remaining in clothing will melt in a warm shelter. When the clothing is again taken outside, the water formed will turn to ice and reduce the CLo value. Brush clothes before entering the shelter. Under living conditions where drying clothing is difficult, it is easier to keep clothing from getting wet than having to dry it out later.

h. If all the snow cannot be eliminated from outer clothing, survivors should remove the clothing and store it in the entry way or on the floor away from the source of heat so it remains cold. If ice should form in clothing, it may be beaten out with a stick.

i. In the cramped quarters of any small emergency shelter, pots of food or drink can be accidentally kicked over. The cooking area, even if it is only a Sterno stove, should be located out of the way, possibly in a snow alcove.

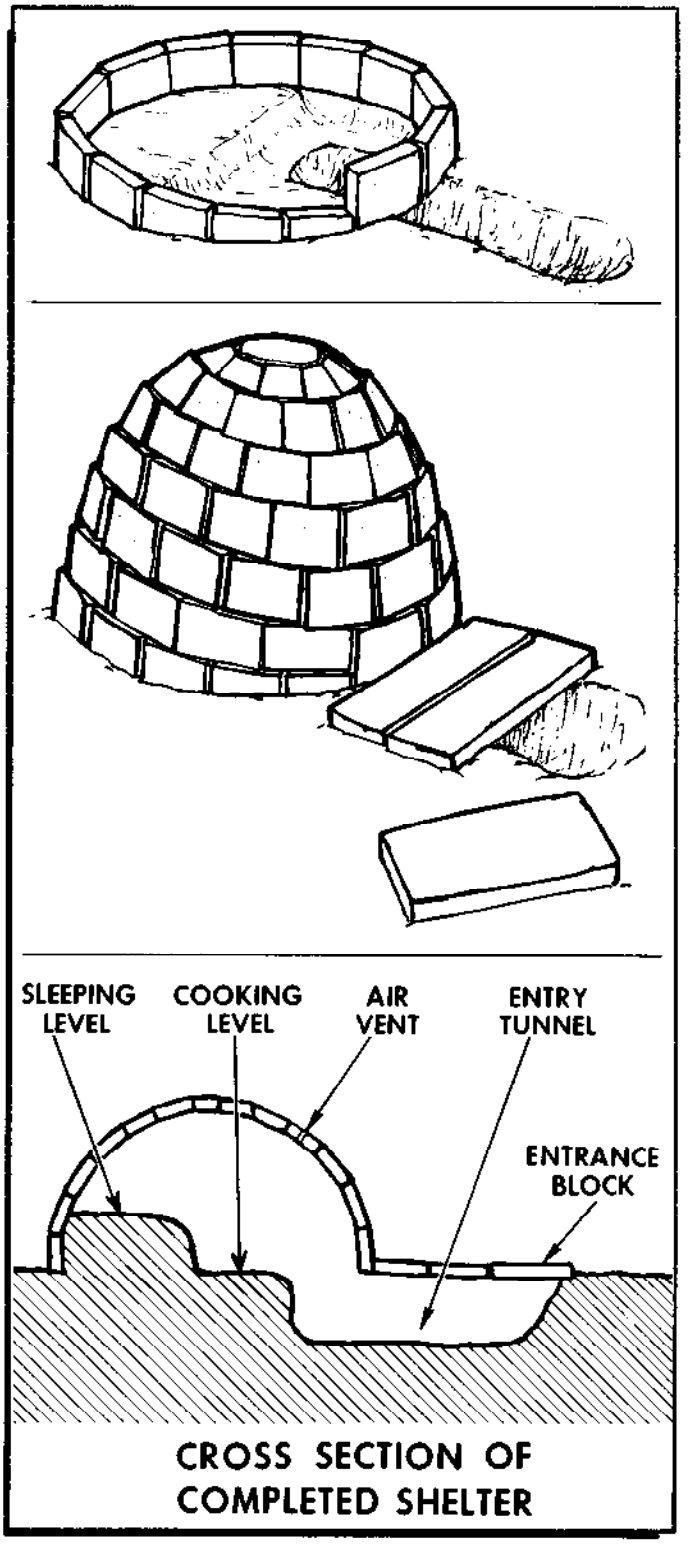

Figure 15-18. Igloo.

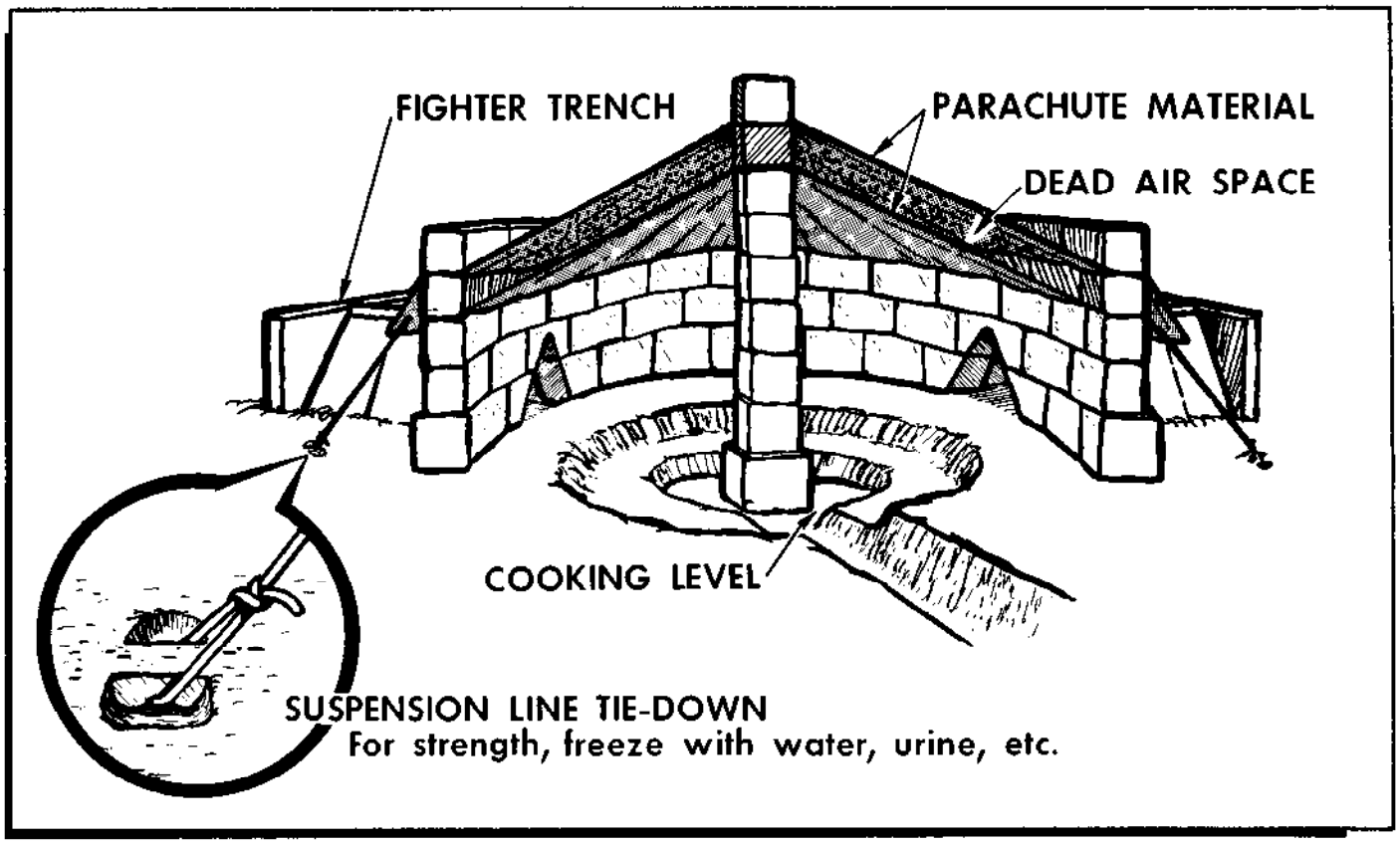

Figure 15-19. Para-Snow House.

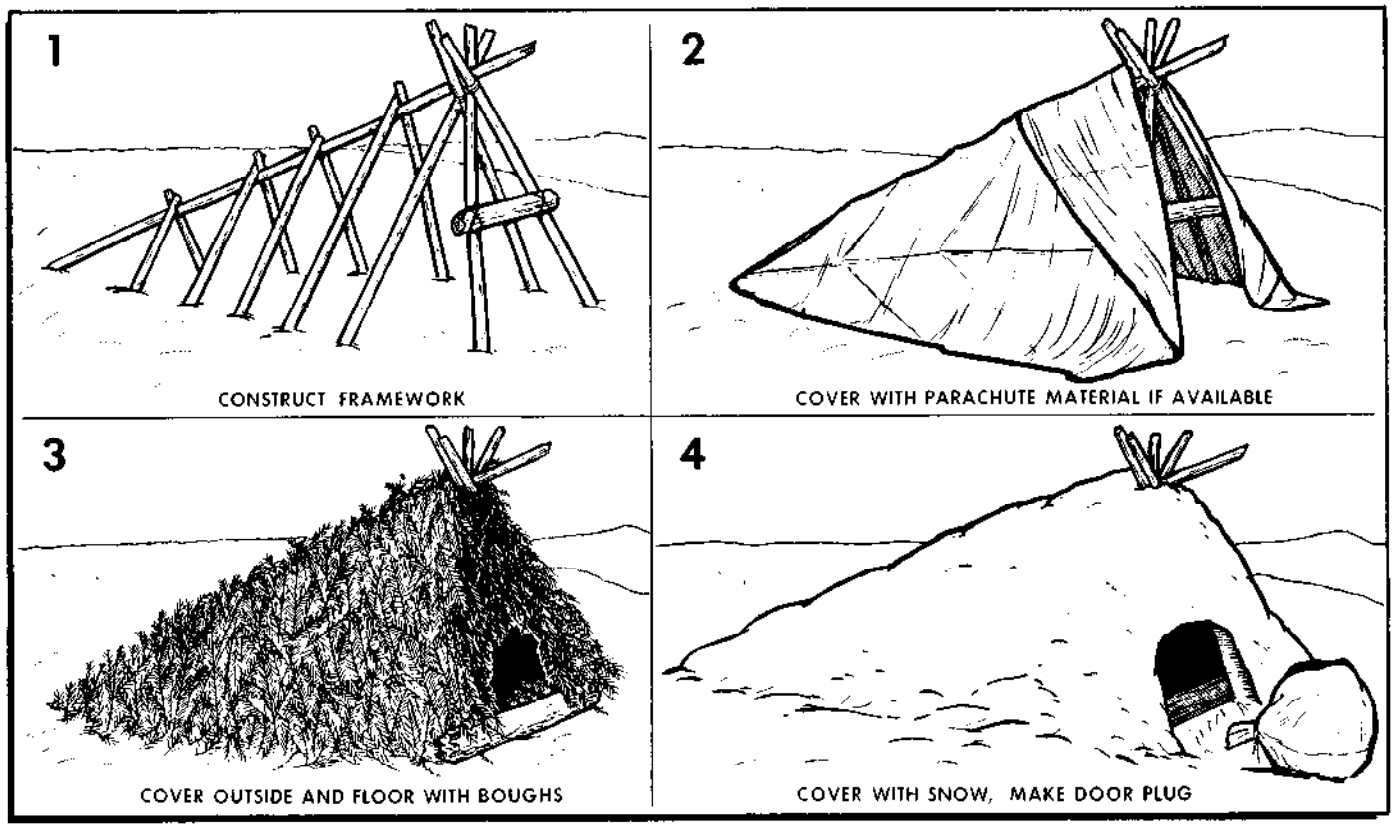

Figure 15-20. Thermal A-Frame.

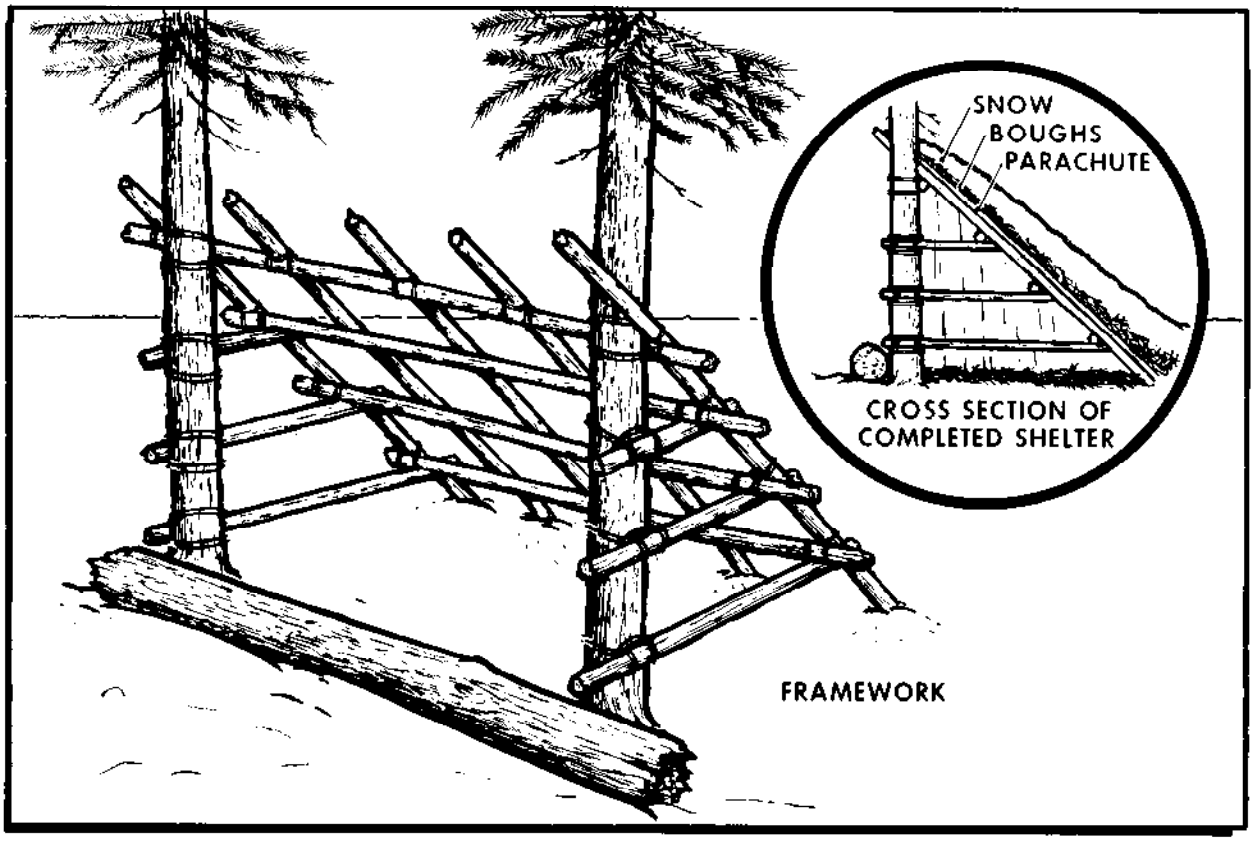

Figure 15-21. Lean-To or Wedge.

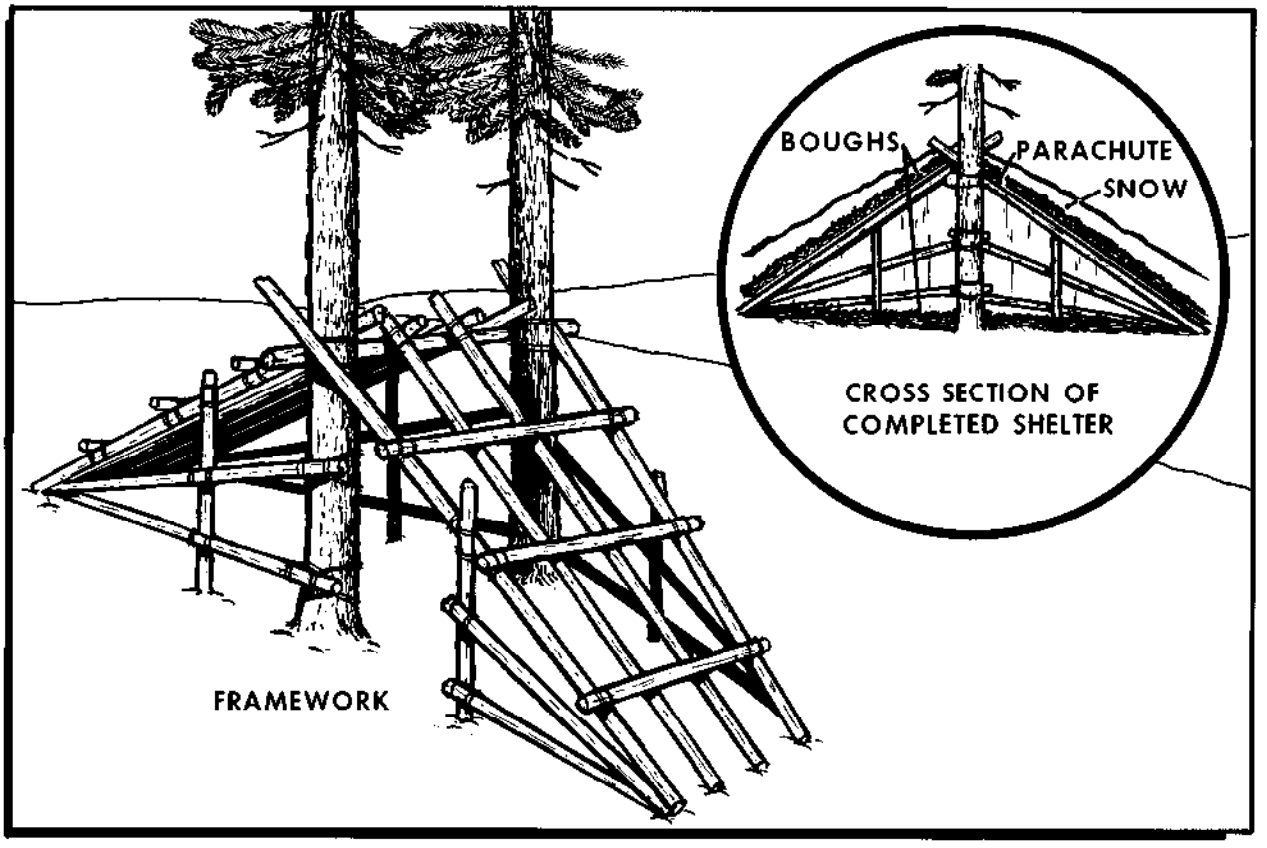

Figure 15-22. Double Lean-To.

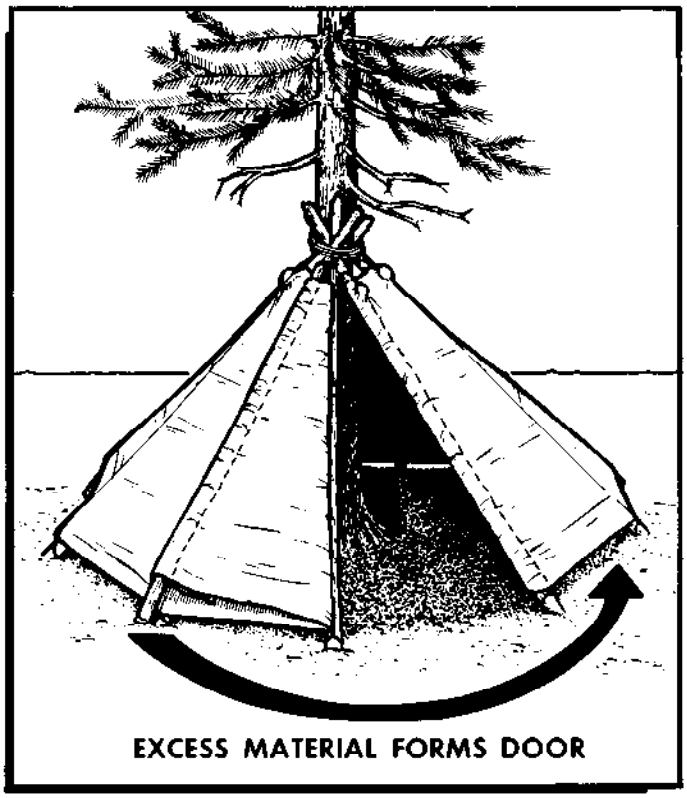

Figure 15-23. Fan Shelter.

15-16. Summer Considerations for Arctic and Arctic-Like Areas:

a. Survivors need shelter against rain and insects. They should choose a campsite near water but on high, dry ground if possible. Survivors should also stay away from thick vegetation, as mosquitoes and flies will make life miserable. A good campsite is a ridge top, cold lake shore, or a spot which gets a breeze.

b. If survivors stay with the aircraft, it can be used for shelter during the summer. They should cover openings with netting or parachute cloth to keep insects out and cook outside to avoid carbon monoxide poisoning. Fires must be built a safe distance from the aircraft.

c. Many temperate area shelters are suitable for summer arctic conditions. The paratepee (of the 1- or no-pole variety) is especially good. It will protect from precipitation and keep insects out.

15-17. Shelter for Open Seas. Personal protection from the elements is just as important on the seas as it is anywhere else. Some rafts come equipped with insulated floors, spray shields, and canopies to protect survivors from heat, cold, and water. If rafts are not so equipped or the equipment has been lost, survivors should try to improvise these items using parachute material, clothing, or other equipment.

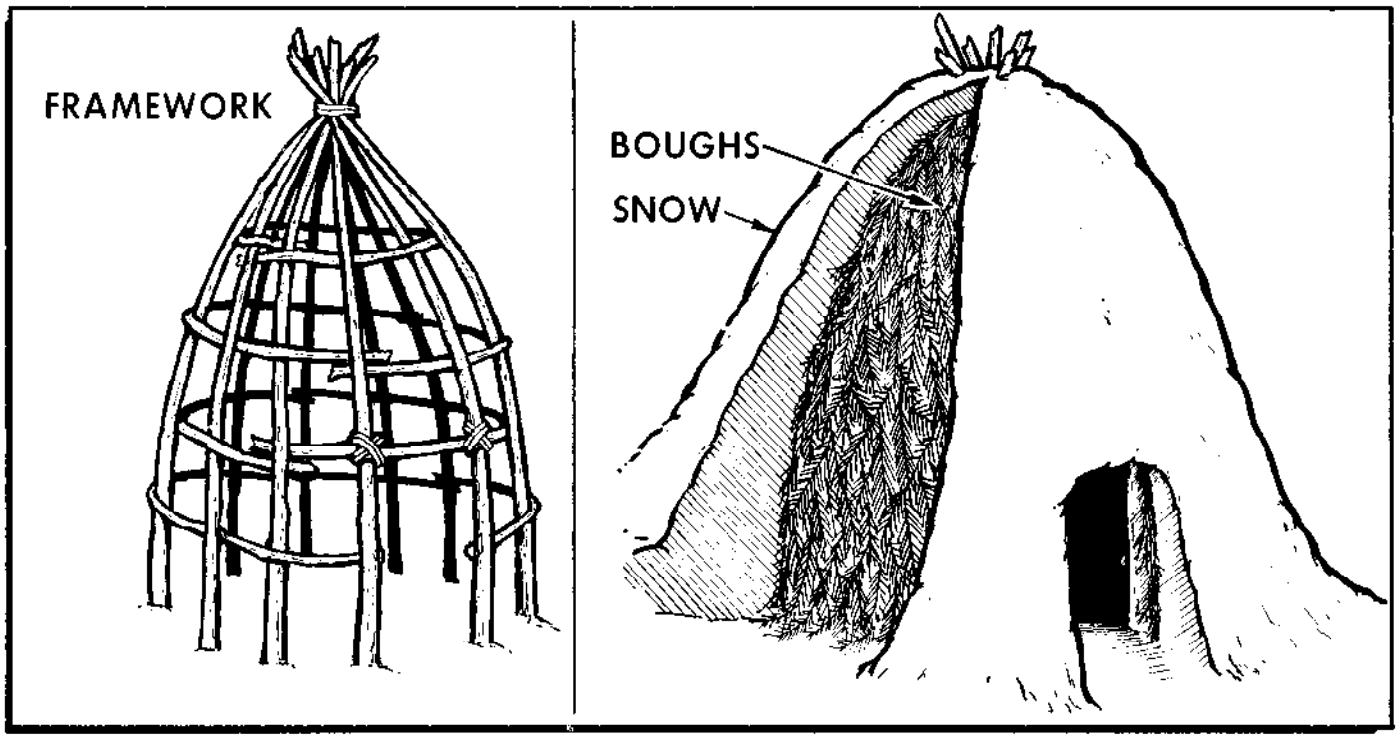

Figure 15-24. Willow Frame Shelter.

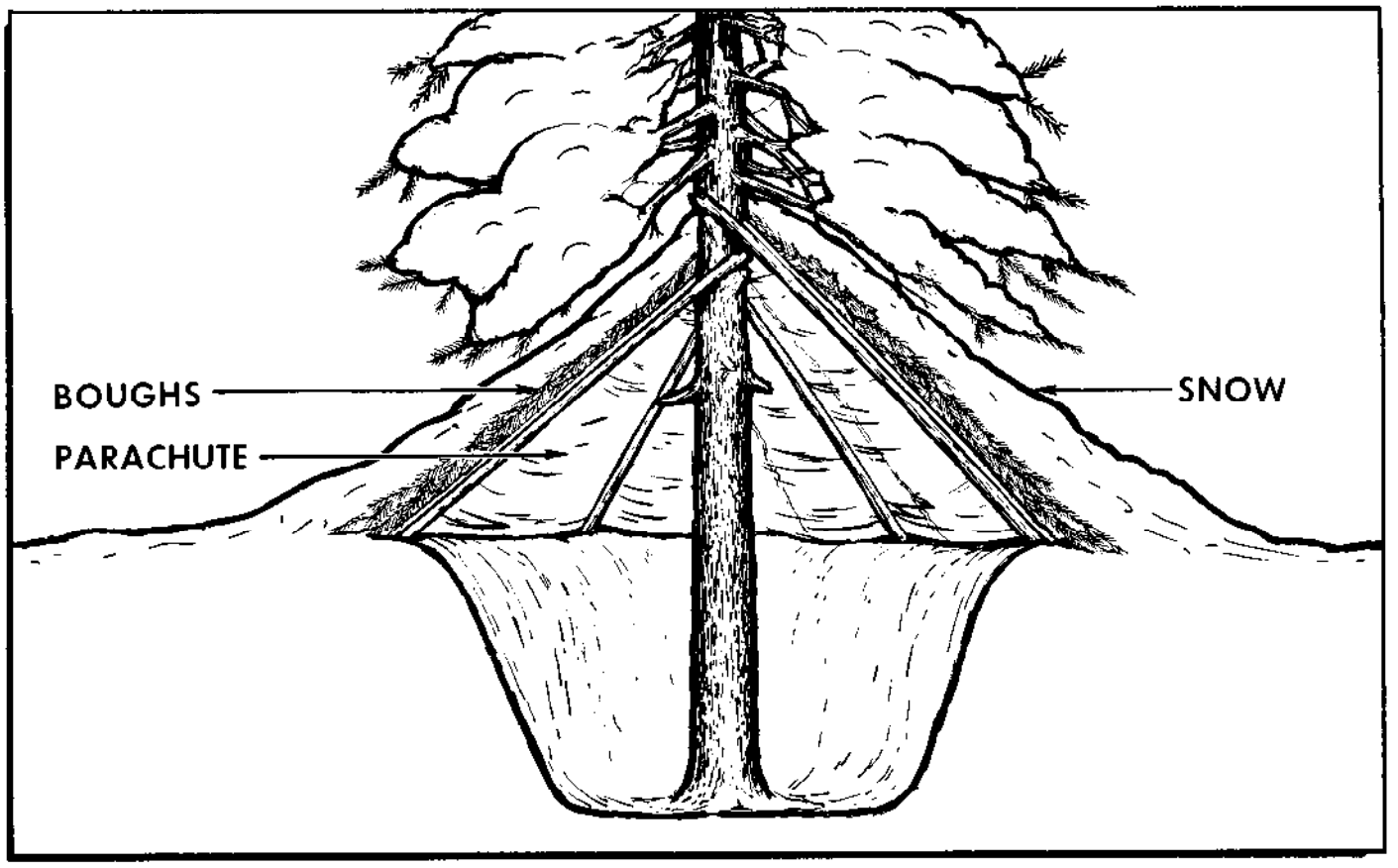

Figure 15-25. Tree Well Shelter.

Figure 15-26. Scraping Snow to Bare Earth.

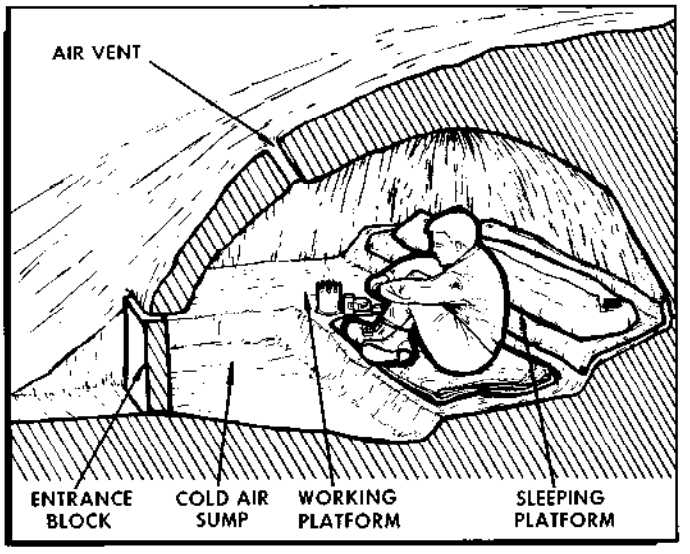

Figure 15-27. Snow Cave Shelter Living.

Chapter 16

FIRECRAFT

16-1. Introduction:

a. The need for a fire should be placed high on the list of priorities. Fire is used for warmth, light, drying clothes, signaling, making tools, cooking, and water purification. When using fire for warmth, the body uses less calories for heat and consequently requires less food. Just having a fire to sit by is a morale booster. Smoke from a fire can be used to discourage insects.

b. Avoid building a very large fire. Small fires require less fuel, are easier to control, and their heat can be concentrated. Never leave a fire unattended unless it is banked or contained. Banking a fire is done by scraping cold ashes and dry earth onto the fire, leaving enough air coming through the dirt at the top to keep the fuel smoldering. This will keep the fire safe and allow it to be rekindled from the saved coals.

16-2. Elements of Fire:

a. The three essential elements for successful fire building are fuel, heat, and oxygen. These combined elements are referred to as the "fire triangle." By limiting fuel, only a small fire is produced. If the fire is not fed properly, there is too much or too little fire. Green fuel is difficult to ignite, and the fire must be burning well before it is used for fuel. Oxygen and heat must be accessible to ignite any fuel.

b. The survivor must take time and prepare well! Preparing all of the stages of fuel and all of the parts of the fire starting apparatus is the key. To be successful at firecraft, one needs to practice and be patient.

c. The fuels used in building a fire normally fall into three categories (figure 16-1), relating to their size and flash point: tinder, kindling, and fuel.

(1) Tinder is any type of small material having a low flash point. It is easily ignited with a minimum of heat, even a spark. Tinder must be arranged to allow air (oxygen) between the hair-like, bone-dry fibers. The preparation of tinder for fire is one of the most important parts of firecraft. Dry tinder is so critical that pioneers used extreme care to have some in a waterproof "tinder box" at all times. It may be necessary to have two or three stages of tinder to get the flame to a useful size. Tinders include:

(a) The shredded bark from some trees and bushes.

(b) Cedar, birch bark, or palm fiber.

(c) Crushed fibers from dead plants.

(d) Fine, dry woodshavings, and straw/grasses.

(e) Resinous sawdust.

(f) Very fine pitch woodshavings (resinous wood from pine or sappy conifers).

(g) Bird or rodent nest linings.

(h) Seed down (milkweed, cattail, thistle).

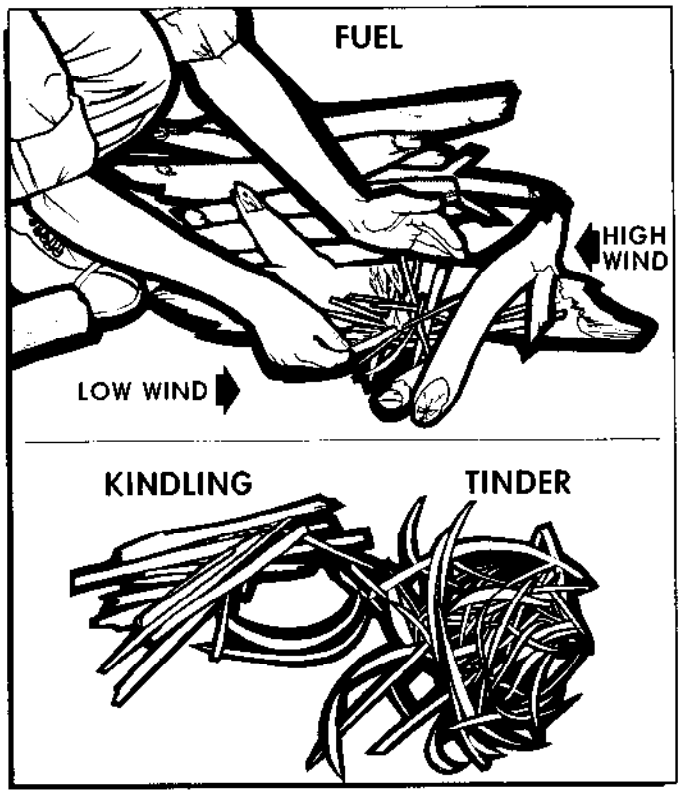

Figure 16-1. Stages of a Fire.

(i) Charred cloth.

(j) Cotton balls or lint.

(k) Steel wool.

(l) Dry powdered sap from the pine tree family (also known as pitch).

(m) Paper.

(n) Foam rubber.

(2) Kindling is the next larger stage of fuel material. It should also have a high combustible point. It is added to, or arranged over, the tinder in such a way that it ignites when the flame from the tinder reaches it. Kindling is used to bring the burning temperature up to the point where larger and less combustible fuel material can be used. Kindling includes:

(a) Dead dry small twigs or plant fibers.

(b) Dead dry thinly shaved pieces of wood, bamboo, or cane (always split bamboo as sections can explode).

(c) Coniferous seed cones and needles.

(d) "Squaw wood" from the underside of coniferous trees; dead, small branches next to the ground sheltered by the upper live part of the tree.

(e) Pieces of wood removed from the insides of larger pieces.

(f) Some plastics such as the spoon from an in-flight ration.

(g) Wood which has been soaked or doused with flammable materials; that is, wax, insect repellent, petroleum fuels, and oil.

(h) Strips of petrolatum gauze from a first aid kit.

(i) Dry split wood burns readily because it is drier inside. Also the angular portions of the wood burn easier than the bark-covered round pieces because it exposes more surface to the flame. The splitting of all fuels will cause them to burn more readily.

(3) Fuel, unlike tinder and kindling, does not have to be kept completely dry as long as there is enough kindling to raise the fuel to a combustible temperature. It is recommended that all fine materials be protected from moisture to prevent excessive smoke production. (Highly flammable liquids should not be poured on an existing fire. Even a smoldering fire can cause the liquids to explode and cause serious burns.) The type of fuel used will determine the amount of heat and light the fire will produce. Dry split hardwood trees (oak, hickory, monkey pod, ash) are less likely to produce excessive smoke and will usually provide more heat than soft woods. They may also be more difficult to break into usable sizes. Pine and other conifers are fast-burning and produce smoke unless a large flame is maintained. Rotten wood is of little value since it smolders and smokes. The weather plays an important role when selecting fuel. Standing or leaning wood is usually dry inside even if it is raining. In tropical areas, avoid selecting wood from trees that grow in swampy areas or those covered with mosses. Tropical soft woods are not usually a good fuel source. Trial and error is sometimes the best method to determine which fuel is best. After identifying the burning properties of available fuel, a selection can be made of the type needed. Recommended fuel sources are:

(a) Dry standing dead wood and dry dead branches (those that snap when broken). Dead wood is easy to split and break. It can be pounded on a rock or wedged between other objects and bent until it breaks.

(b) The insides of fallen trees and large branches may be dry even if the outside is wet. The heart wood is usually the last to rot.

(c) Green wood which can be made to burn is found almost anywhere, especially if finely split and mixed evenly with dry dead wood.

(d) In treeless areas, other natural fuels can be found. Dry grasses can be twisted into bunches. Dead cactus and other plants are available in deserts. Dry peat moss can be found along the surface of undercut streambanks. Dried animal dung, animal fats, and sometimes even coal can be found on the surface. Oil impregnated sand can also be used when available.

16-3. Fire Location. The location of a fire should be carefully selected. An old story is told of a mountain man who used his last match to light a fire built under a snow-covered tree. The heat from the fire melted the snow and it slid off the tree and put out the fire. For a survivor, this type of accident can be very demoralizing or even deadly. Locate and prepare the fire carefully.

16-4. Fire Site Preparation:

a. After a site is located, twigs, moss, grass, or duff should be cleaned away. Scrape at least a 3-foot diameter area down to bare soil for even a small fire. Larger fires require a larger area. If the fire must be built on snow, ice, or wet ground, survivors should build a platform of green logs or rocks. (Beware of wet or porous rocks, they may explode when heated.)

b. There is no need to dig a hole or make a circle of rocks in preparation for fire building. Rocks may be placed in a circle and filled with dirt, sand, or gravel to raise the fire above the moisture from wet ground. The purpose of these rocks is to hold the platform only.

c. To get the most warmth from the fire, it should be built against a rock or log reflector (figure 16-2). This will direct the heat into the shelter. Cooking fires can be walled-in by logs or stones. This will provide a platform for cooking utensils and serve as a windbreak to help keep the heat confined.

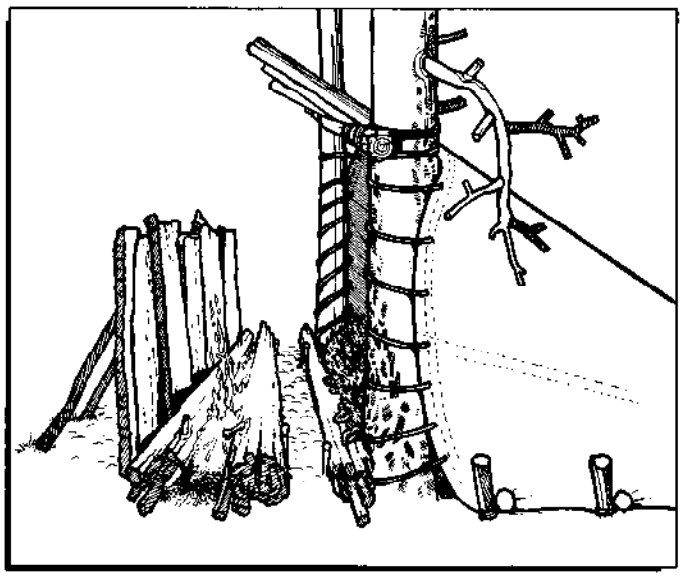

Figure 16-2. Fire Reflector.

d. After preparing the fire, all materials should be placed together and arranged by size (tinder, kindling, and fuel). As a rule of thumb, survivors should have three times the amount of tinder and kindling than is necessary for one fire. It is to their advantage to have too much than not enough. Having plenty of material on hand will prevent the possibility of the fire going out while additional material is gathered.

16-5. Firemaking With Matches (or Lighter):

a. Survivors should arrange a small amount of kindling in a low pyramid, close enough together so flames can jump from one piece to another. A small opening should be left for lighting and air circulation.

b. Matches can be conserved by using a "shave stick," or by using a loosely tied fagot of thin, dry twigs. The match must be shielded from wind while igniting the shave stick. The stick can then be applied to the lower windward side of the kindling.

c. Small pieces of wood or other fuel can be laid gently on the kindling before lighting or can be added as the kindling begins to burn. The survivors can then place smaller pieces first, adding larger pieces of fuel as the fire begins to burn. They should avoid smothering the fire by crushing the kindling with heavy wood.

d. Survivors only have a limited number of matches or other instant fire-starting devices. In a long-term situation, they should use these devices sparingly or carry fire with them when possible. Many primitive cultures carry fire (fire bundles) by using dry punk or fiberous barks (cedar) encased in a bark. Others use torches. Natural fire bundles also work well for holding the fire (figure 16-3).

e. The amount of oxygen must be just enough to keep the coals inside the dry punk burning slowly. This requires constant vigilance to control the rate of the burning process. The natural fire bundle is constructed in a cross section as shown in figure 16-3.

16-6. Heat Sources. A supply of matches, lighters, and other such devices will only last a limited time. Once the supply is depleted, they cannot be used again. If possible, before the need arises, survivors should become skilled at starting fires with more primitive means, such as friction, heat, or a sparking device. It is essential that they continually practice these procedures. The need to start a fire may arise at the most inopportune times. One of the greatest aids a survivor can have for rapid fire starting is the "tinder box" previously mentioned. Using friction, heat, and sparks are very reliable methods for those who use them on a regular basis. Therefore, survivors must practice these methods. Survivors must be aware of the problems associated with the use of primitive heat sources. If the humidity is high in the immediate area, a fire may be difficult to ignite even if all other conditions are favorable. For primitive methods to be successful, the materials must be BONE DRY. The primitive people who use these ignition methods take great care to keep their tinder, kindling, and other fuels dry, even to the point of wrapping many layers of waterproof materials around it. *PREPARATION, PRACTICE,* and *PATIENCE* in the use of primitive fire-building techniques cannot be over emphasized. A key point in all primitive methods is to ensure that the tinder is not disturbed.

a. Flint and Steel:

(1) Flint and steel is one way to produce fire without matches.

(a) To use this method, survivors must hold a piece of flint in one hand above the tinder.

(b) Grasp the steel in the other hand and strike the flint with the edge of the steel in a downward glancing blow (figure 16-4).

(2) True flint is not necessary to produce sparks. Iron pyrite and quartz will also give off sparks even if they are struck against each other. Check the area and select the best spark-producing stone as a backup for the

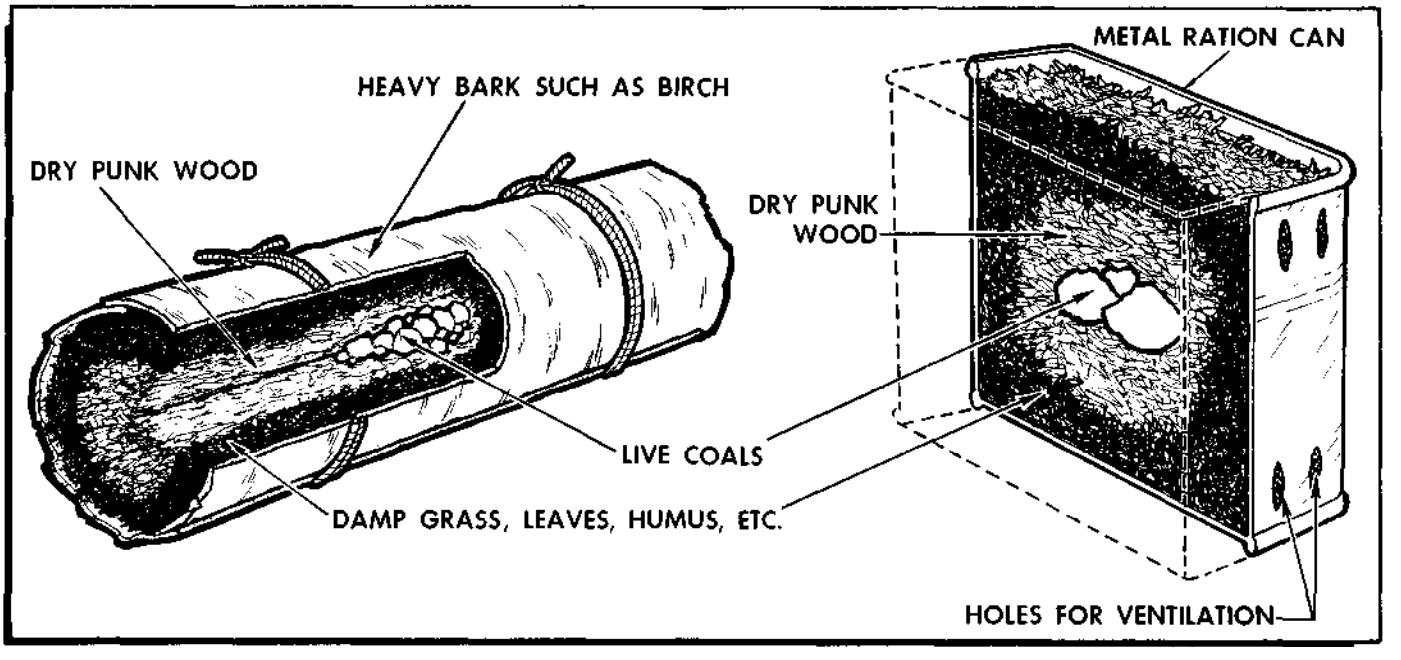

Figure 16-3. Fire Bundles.

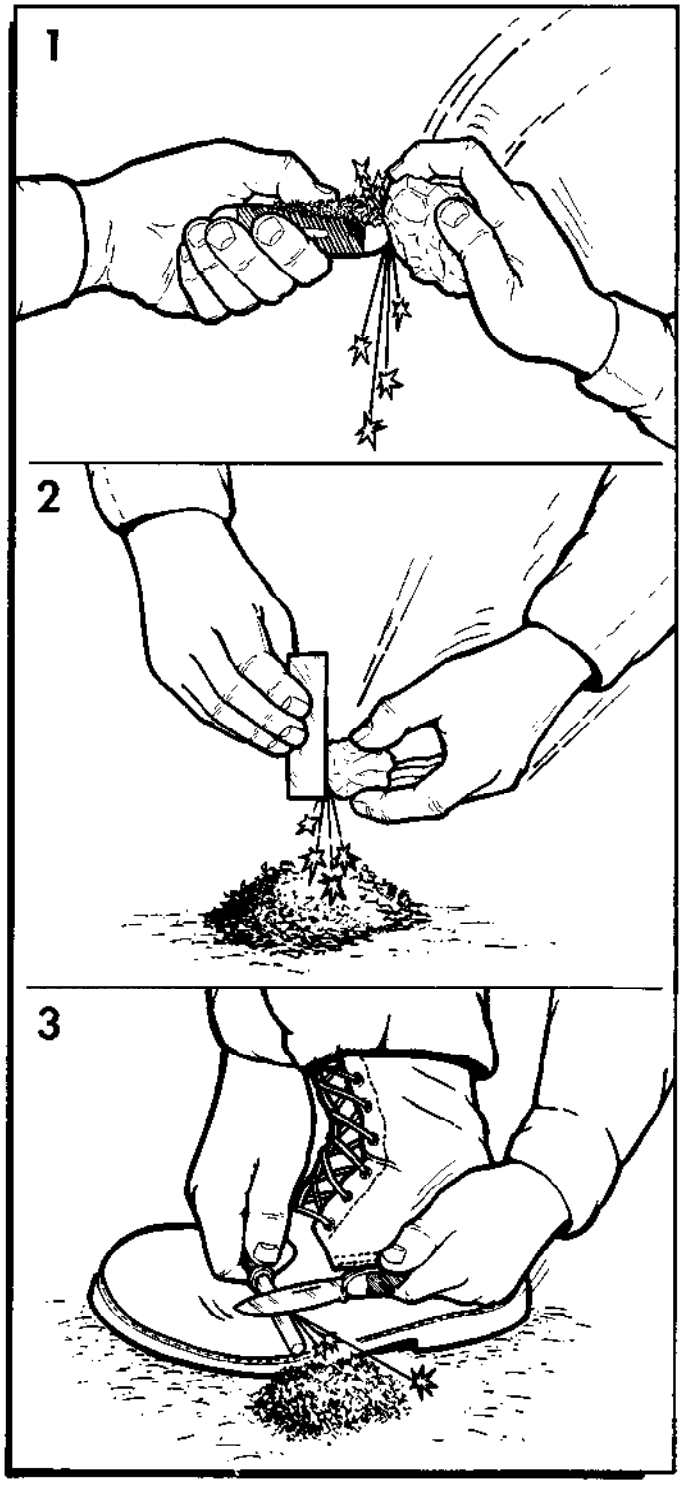

Figure 16-4. Fire Starting With Flint and Steel.

available matches. The sparks must fall on the tinder and then be blown or fanned to produce a coal and subsequent flame.

(3) Synthetic flint, such as the so-called metal match, consists of the same type material used for flints in commercial cigarette lighters. Some contain magnesium which can be scraped into tinder and into which the spark is struck. The residue from the "match" burns hot and fast and will compensate for some moisture in tinder. If issued survival kits do not contain this item and the survivors choose to make one rather than buy it, lighter flints can be glued into a groove in a small piece of wood or plastic. The survivors can then practice striking a spark by scratching the flint with a knife blade. A 90-degree angle between the blade and flint works best. The device must be held close enough for the sparks to hit the tinder, but enough distance must be allowed to avoid accidentally extinguishing the fire. Cotton balls dipped in petroleum jelly make excellent tinder with flint and steel. When the tinder ignites, additional tinder, kindling, and fuel can be added.

b. Batteries:

(1) Another method of producing fire is to use the battery of the aircraft, vehicle, storage batteries, etc. Using two insulated wires, connect one end of a wire to the positive post of the battery and the end of the other wire to the negative post. Touch the two remaining ends to the ends of a piece of noninsulated wire. This will cause a short in the electrical circuit and the noninsulated wire will begin to glow and get hot. Material coming into contact with this hot wire will ignite. Survivors should use caution when attempting to start a fire with a battery. They should ensure that sparks or flames are not produced near the battery because explosive hydrogen gas is produced and can result in serious injury (figure 16-5).

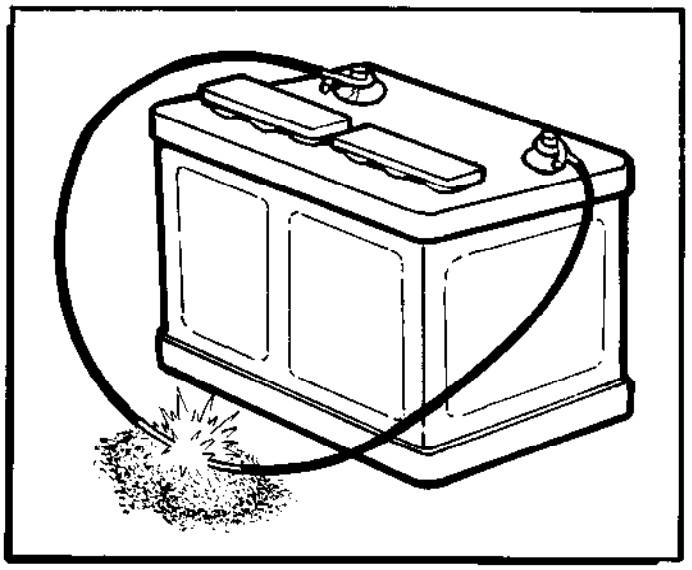

Figure 16-5. Fire Starting With Batteries.

(2) If fine grade steel wool is available, a fire may be started by stretching it between the positive and negative posts until the wire itself makes a red coal.

c. Burning Glass. If survivors have sunlight and a burning glass, a fire can be started with very little physical effort (figure 16-6). Concentrate the rays of the Sun on tinder by using the lens of a lensatic compass, a camera lens, or the lens of a flashlight which magnifies; even a convex piece of bottle glass may work. Hold the lens so that the brightest and smallest spot of concentrated light falls on the tinder. Once a whisp of smoke is produced, the tinder should be fanned or blown upon until the smoking coal becomes a flame. Powdered charcoal in the tinder will decrease the ignition time. Add kindling carefully as in any other type of fire. Practice will reduce the time it takes to light the tinder.

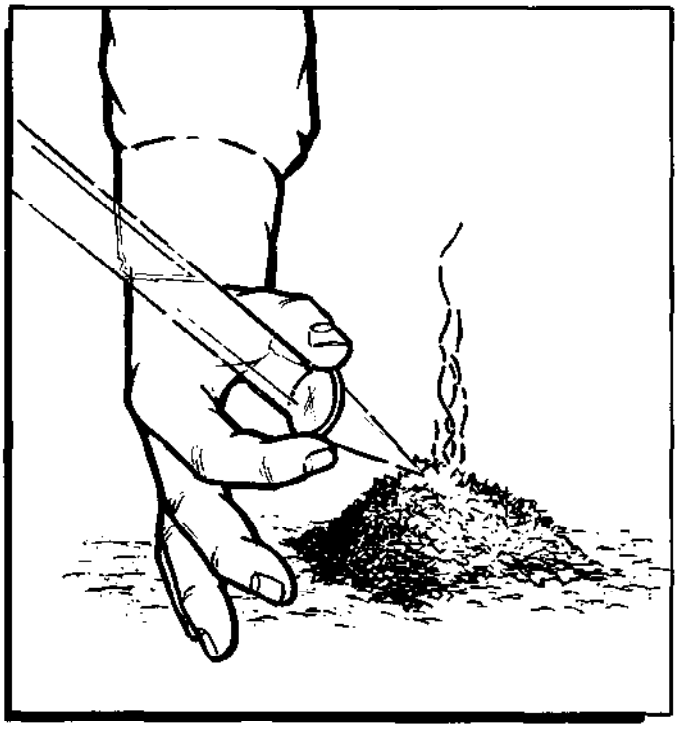

Figure 16-6. Fire Starting With Burning Glass.

d. Flashlight Reflector. A flashlight reflector can also be used to start a fire (figure 16-7). Place the tinder in the center of the reflector where the bulb is usually located. Push it up from the back of the hole until the hottest light is concentrated on the end and smoke results. If a cigarette is available, use it as a tinder for this method.

e. Bamboo Fire Saw:

(1) The bamboo fire saw is constructed from a section of dry bamboo with both end joints cut off. The section of bamboo, about 12 inches in length, is split in half lengthwise. The inner wall of one of the halves (called the "running board") is scraped or shaved thin. This is done in the middle of the running board. A notch to serve as a guide is cut in the outer sheath opposite the scraped area of the inner wall. This notch runs across the running board at a 90-degree angle (figure 16-8).

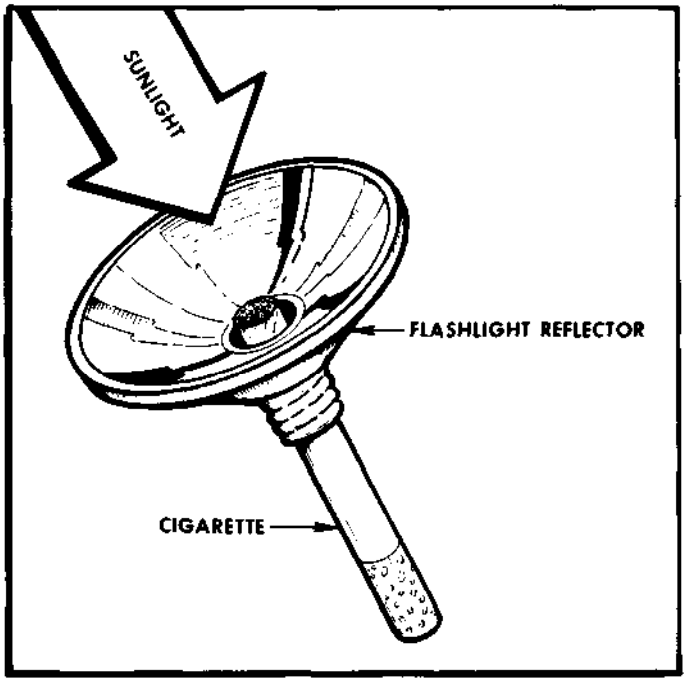

Figure 16-7. Fire Starting With Flashlight Reflector.

(2) The other half of the bamboo joint is further split in half lengthwise, and one of the resultant quarters is used as a "baseboard." One edge of the baseboard is shaved down to make a tapered cutting edge. The baseboard is then firmly secured with the cutting edge up. This may be done by staking it to the ground in any manner which does not allow it to move (figure 16-8).

(3) Tinder is made by scraping the outer sheath of the remaining quarter piece of the bamboo section. The scrapings (approximately a large handful) are then rubbed between the palms of the hands until all of the wood fibers are broken down and dust-like material no longer falls from the tinder. The ball of scrapings is then fluffed to allow maximum circulation of oxygen through the mass (figure 16-8).

(4) The finely shredded and fluffed tinder is placed in the running board directly over the shaved area, opposite the outside notch. Thin strips of bamboo should be placed lengthwise in the running board to hold the tinder in place. These strips are held stationary by the hands when grasping the ends of the running board (figure 16-8).

(5) A long, very thin sliver of bamboo (called the "pick") should be prepared for future use. One end of the running board is grasped in each hand, making sure the thin strips of bamboo are held securely in place. The running board is placed over the baseboard at a right angle, so that the cutting edge of the baseboard fits into

Figure 16-8. Bamboo Fire Saw.

the notch in the outer sheath of the running board. The running board is then slid back and forth as rapidly as possible over the cutting edge of the baseboard, with sufficient downward pressure to ensure enough friction to produce heat.

(6) As soon as "billows" of smoke rise from the tinder, the running board is picked up. The pick is used to push the glowing embers from the bottom of the running board into the mass of tinder. While the embers are being pushed into the tinder, they are gently blown upon until the tinder bursts into flame.

(7) As soon as the tinder bursts into flame, slowly add kindling in small pieces to avoid smothering the fire. Fuel is gradually added to produce the desired size fire. If the tinder is removed from the running board as soon as it flames, the running board can be reused by cutting a notch in the outer sheath next to the original notch and directly under the scraped area of the inner wall.

f. Bow and Drill:

(1) This is a friction method which has been used successfully for thousands of years. A spindle of yucca, elm, basswood, or any other straight grainwood (not softwood) should be made. The survivors should make sure that the wood is not too hard or it will create a glazed surface when friction is applied. The spindle should be 12 to 18 inches long and three-fourths inch in diameter. The sides should be octagonal, rather than round, to help create friction when spinning. Round one end and work the other end into a blunt point. The round end goes to the top upon which the socket is placed. The socket is made from a piece of hardwood large enough to hold comfortably in the palm of the hand with the curved part up and the flat side down to hold the top of the spindle. Carve or drill a hole in this side and make it smooth so it will not cause undue friction and heat production. Grease or soap can be placed in this hole to prevent friction (figure 16-9).

(2) The bow is made from a stiff branch about 3 feet long and about 1 inch in diameter. This piece should have sufficient flexibility to bend. It is similar to a bow used to shoot arrows. Tie a piece of suspension line or leather thong to both ends so that it has the same tension as that of a bow. There should be enough tension for the spindle to twist comfortably.

(3) The fireboard is made of the softwood and is about 12 inches long, three-fourths inch thick, and 3 to 6 inches wide. A small hollow should be carved in the fireboard. A V-shaped cut can then be made in from the edge of the board. This V-shape should extend into the center of the hollow where the spindle will make the hollow deeper. The object of this "V" cut is to create an angle which cuts off the edge of the spindle as it gets hot and turns to charcoal dust. This is the critical part of the fireboard and must be held steady during the operation of spinning the spindle.

(4) While kneeling on one knee, the other foot can be placed on the fireboard as shown in figure 16-9 and the tinder placed under the fireboard just beneath the V-cut. Care should be taken to avoid crushing the tinder under the fireboard. Space can be obtained by using a small, three-fourths inch diameter stick to hold up the fireboard. This allows air into the tinder where the hot powder (spindle charcoal dust) is collected.

(5) The bow string should be twisted once around the spindle. The spindle can then be placed upright into the spindle hollow (socket). The survivor may press the socket down on the spindle and fireboard. The entire apparatus must be held steady with the hand on the socket braced against the leg or knee. The spindle should begin spinning with long even slow strokes of the bow until heavy smoke is produced. The spinning should become faster until the smoke is very thick. At this point, hot powder, that can be blown into a glowing ember, has been successfully produced. The bow and spindle can then be removed from the fireboard and the

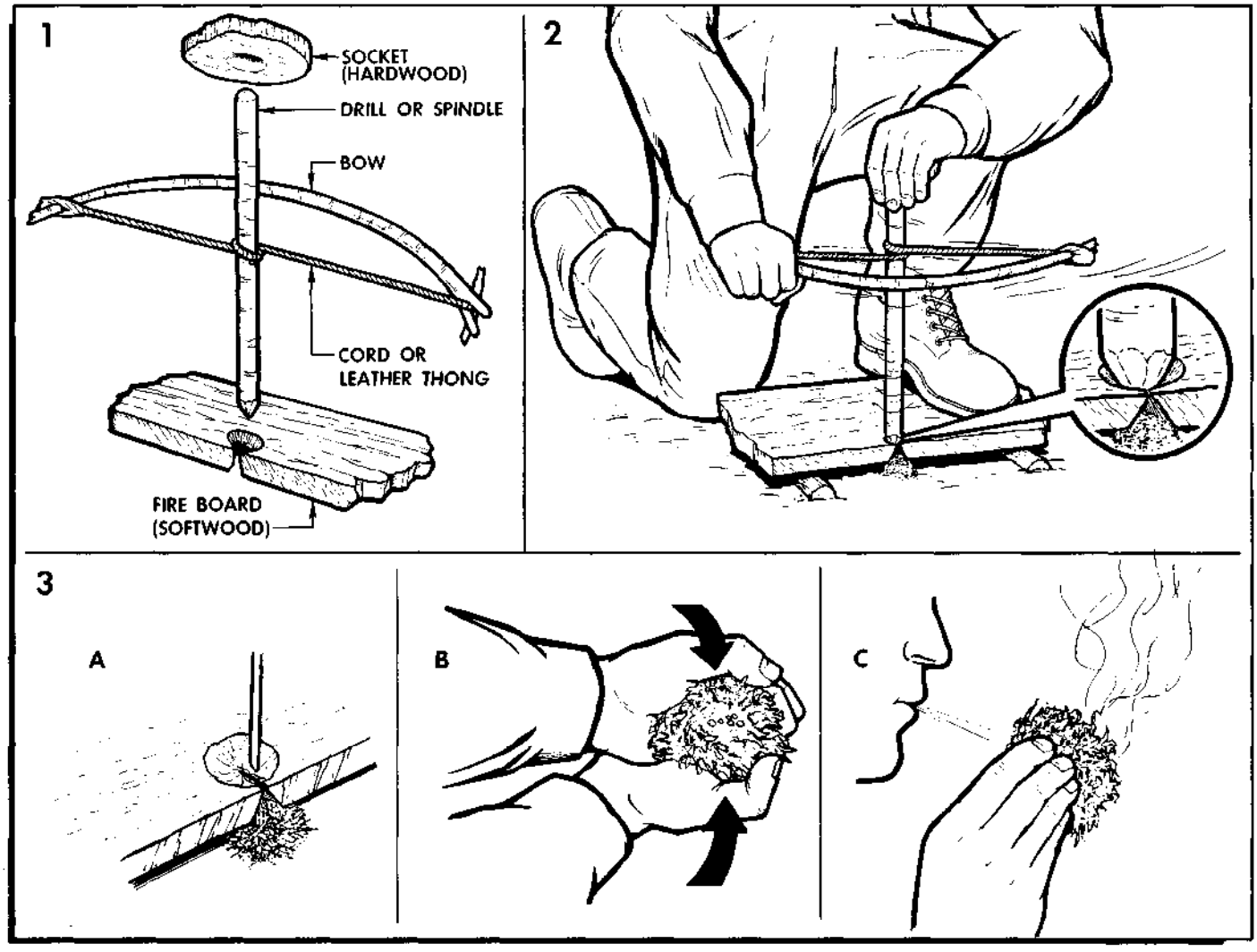

Figure 16-9. Bow and Drill.

tinder can be placed next to the glowing ember making sure not to extinguish it. The tinder must then be rolled gently around the burning ember, and blow into the embers, starting the tinder to burn. This part of the fire is most critical and should be done with care and planning.

(6) The burning tinder is then placed into the waiting fire "lay" containing more tinder and small kindling. At no time in this process should the survivor break concentration or change sequence. The successful use of these primitive methods of fire starting will require a great deal of patience. Success demands dedication and practice.

g. The Fire Thong. The fire thong, another friction method, is used in only those tropical regions where rattan is found. The system is simple and consists of a twisted rattan thong or other strong plant fiber, 4 to 6 feet long, less than 1 inch in diameter, and a 4-foot length of dry wood which is softer than rattan (dedicuous wood) (figure 16-10). Rub with a steady but increasing rhythm.

h. The Plow. The plow is a method used by some primitives and basically follows the principles of other friction methods. The wood used must not glaze with heat applied and must be able to produce powder with friction.

i. Ground Stake. Another variation can be constructed by driving a stake into the ground as shown in figure 16-11.

16-7. Firemaking With Special Equipment:

a. The night end of the day-night flare can be used as a fire starter. This means, however, that survivors must weigh the importance of a fire against the loss of a night flare.

b. Some emergency kits contain small fire starters, cans of special fuels, windproof matches, and other aids.

Figure 16-10. Fire Thong.

Survivors should save the fire starters for use in extreme cold and damp (moist) weather conditions.

c. The white plastic spoon (packed in various in-flight rations) may be the type that burns readily. The handle should be pushed deep enough into the ground to support the spoon in an upright position. Light the tip of the spoon. It will burn for about 10 minutes (long enough to dry out and ignite small tinder and kindling).

d. If a candle is available, it should be ignited to start a fire and thus prevent using more than one match. As soon as the fire is burning, the candle can be extinguished and saved for future use.

e. Tinder can be made more combustible by adding a few drops of flammable fuel/material. An example of this would be mixing the powder from an ammunition cartridge with the tinder. After preparing tinder in this manner, it should be stored in a waterproof container for future use. Care must be used in handling this mixture because the flash at ignition could burn the skin and clothing.

f. For thousands of years, the Eskimos and other northern peoples have relied heavily upon oils from animals to heat their homes. A fat stove or "Koodlik" is used by the Eskimos to burn this fuel.

g. Survivors can improvise a stove from a ration can and burn any flammable oil-type liquid or animal fats available. Here again, survivors should keep in mind that if there is only a *limited* amount of animal fat, it should be eaten to produce heat inside the body.

16-8. Burning Aircraft Fuel. On barren lands in the arctic, aircraft fuel may be the only material survivors have available for fire.

a. A stove can be improvised to burn fuel, lubricating oil, or a mixture of both (figure 16-12). The survivor should place 1 or 2 inches of sand or fine gravel in the bottom of a can or other container and add fuel. *Care should be used when lighting the fuel because it may explode.* Slots should be cut into the top of the can to let flame and smoke out, and holes punched just above the level of the sand to provide a draft. A mixture of fuel and oil will make the fire burn longer. If no can is available, a hole can be dug and filled with sand. Fuel is then poured on the sand and ignited. The survivor should not allow the fuel to collect in puddles.

b. Lubricating oil can be burned as fuel by using a wick arrangement. The wick can be made of string, rope, rag, sphagnum moss, or even a cigarette and should be placed on the edge of a receptacle filled with oil. Rags, paper, wood, or other fuel can be soaked in oil and thrown on the fire.

c. A stove can be made of any empty waxed carton by cutting off one end and punching a hole in each side near the unopened end. Survivors can stand the carton on the closed end and loosely place the fuel inside the carton. The stove can then be lit using fuel material left hanging over the end. The stove will burn from the top down.

d. Seal blubber makes a satisfactory fire without a container if gasoline or heat tablets are available to provide an initial hot flame (figure 16-13). The heat source should be ignited on the raw side of the blubber while the fur side is on the ice. A square foot of blubber burns for several hours. Once the blubber catches fire, the heat tablets can be recovered. Eskimos light a small piece of blubber and use it to kindle increasingly larger pieces. The smoke from a blubber fire is dirty, black, and heavy. The flame is very bright and can be seen for several miles. The smoke will penetrate clothing and blacken the skin.

16-9. Useful Firecraft Hints:

a. Conserve matches by only using them on properly prepared fires. They should never be used to light cigarettes or for starting unnecessary fires.

b. Carry some dry tinder in a waterproof container. It should be exposed to the Sun on dry days. Adding a little powdered charcoal will improve it. Cotton cloth is good tinder, especially if scorched or charred. It works well with a burning glass or flint and steel.

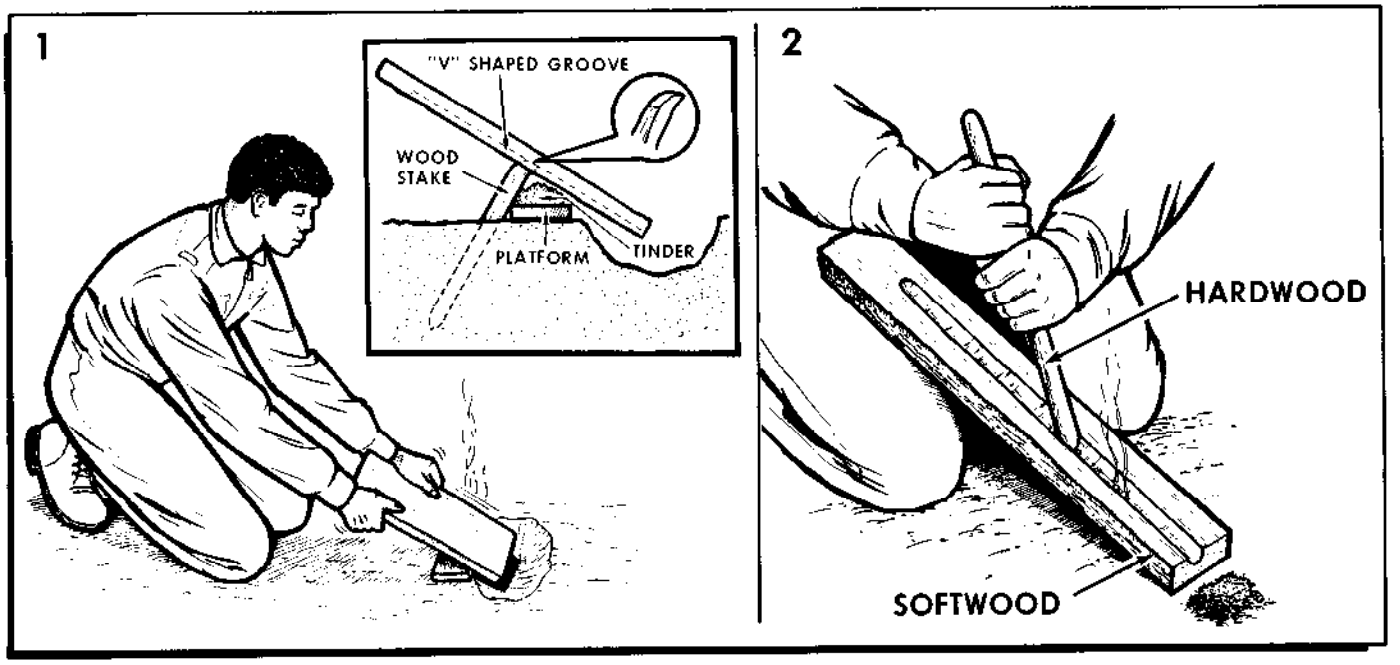

Figure 16-11. Fire Plow.

c. Remember that firemaking can be a difficult job in an arctic environment. The main problem is the availability of firemaking materials. Making a fire starts *WELL* before the match is lit. The fire must be protected from the wind. In wooded areas, standing timber and brush usually make a good windbreak but in open areas, some type of windbreak may have to be constructed. A row of snowblocks, the shelter of a ridge, or a pile of brush will work as a windbreak. It must be high enough to shield the fire from the wind. It may also act as a heat reflector if it is of solid material.

d. Remember, a platform will be required to prevent the fire from melting down through the deep snow and extinguishing it. A platform is also needed if the ground is moist or swampy. The platform can be made of green logs, metal, or any material that will not burn through very readily. Care must be taken when selecting an area for fire building. If the area has a large accumulation of humus material and (or) peat, a platform is needed to avoid igniting the material as it will tend to smolder long after the flames of the fire are extinguished. A smoldering peat fire is almost impossible to put out and may burn for years.

e. In forested areas, the debris on the ground and the lichen mat should be cleared away to mineral soil, if possible, to prevent the fire from spreading.

f. The ignition source used to ignite the fire must be quick and easily operated with hand protection such as mittens. Any number of devices will work well—matches, candle, lighter, fire starter, metal matches, etc.

16-10. Fire Lays. Most fires are built to meet specific needs or uses, either heat, light, or preparing food and water. The following configurations are the most commonly used for fires and serve one or more needs (figure 16-14).

a. Tepee:

(1) The tepee fire can be used as a light source and has a concentrated heat point directly above the apex of the tepee which is ideal for boiling water. To build:

(a) Place a large handful of tinder on the ground in the middle of the fire site.

(b) Push a stick into the ground, slanting over the tinder.

(c) Then lean a circle of kindling sticks against the slanting stick, like a tepee, with an opening toward the windward side for draft.

(2) To light the fire:

(a) Crouch in front of the fire lay with the back to the wind.

(b) Feed the fire from the downwind side, first with thin pieces of fuel, then gradually with thicker pieces.

(c) Continue feeding until the fire has reached the desired size. The tepee fire has one big drawback. It tends to fall over easily. However, it serves as an excellent starter fire.

b. Log Cabin. As the name implies, this lay looks similar to a log cabin. Log cabin fires give off a great amount of light and heat primarily because of the amount of oxygen which enters the fire. The log cabin

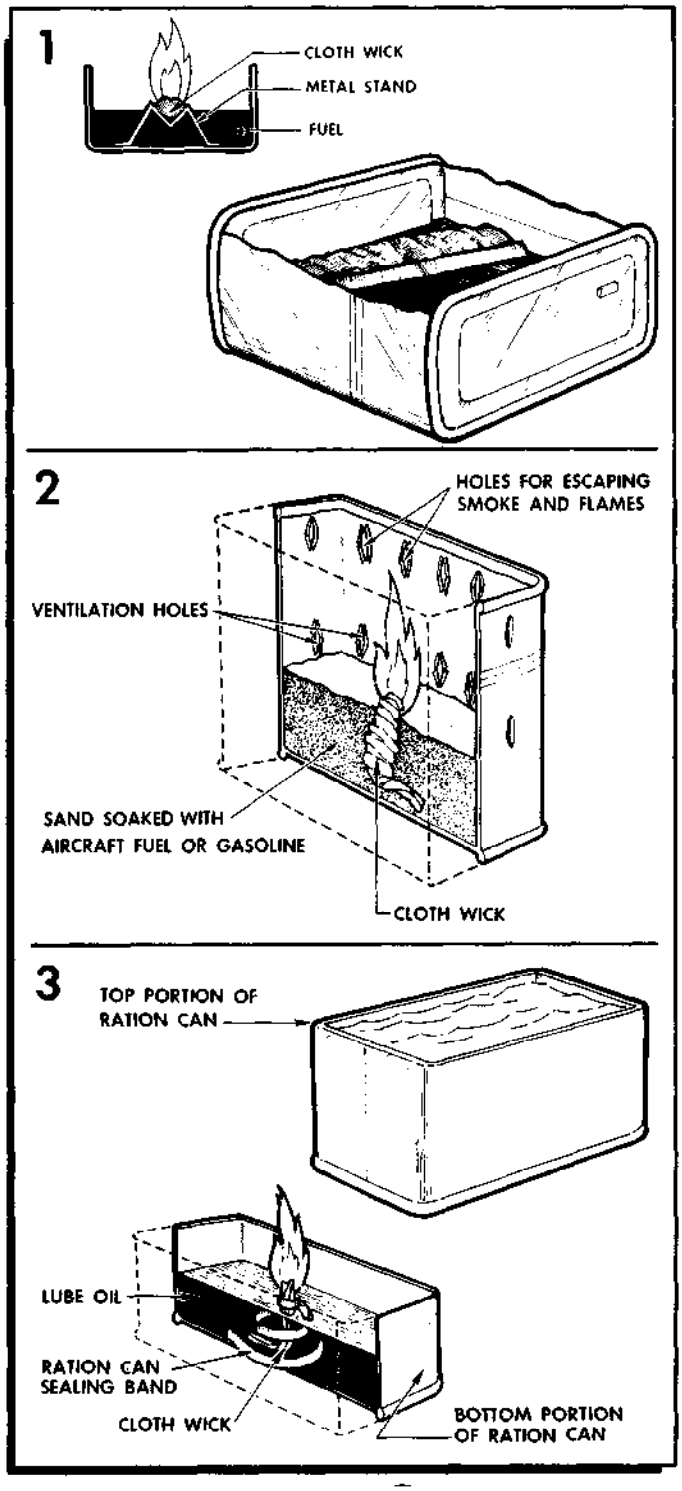

Figure 16-12. Fat and Oil Stoves.

fire creates a quick and large bed of coals and can be used for cooking or as the basis for a signal fire. If one person or a group of people are going to use the coals for cooking, the log cabin can be modified into a long fire or a keyhole fire.

c. Long Fire. The long fire begins as a trench, the length of which is layed to take advantage of existing

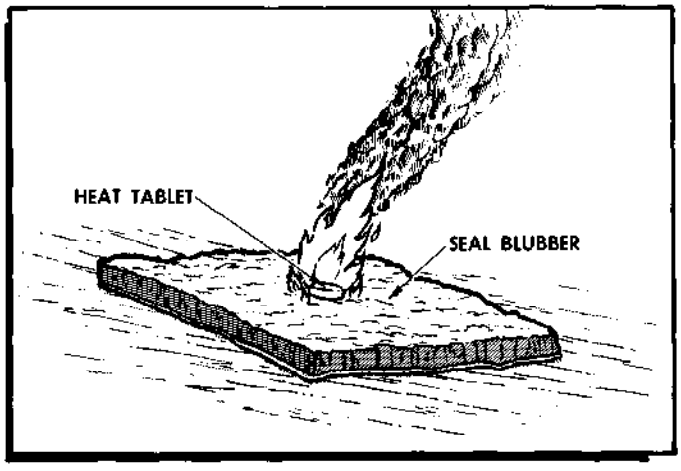

Figure 16-13. Heat Tablet/Seal Blubber

wind. The long fire can also be built above ground by using two parallel green logs to hold the coals together. These logs should be at least 6 inches in diameter and situated so the cooking utensils will rest upon the logs. Two 1-inch thick sticks can be placed under both logs, one at each end of the long fire. This is done to allow the coals to receive more air.

d. Keyhole Fire. To construct a keyhole fire, a hole is dug in the shape of an old style keyhole and does the same thing as the long fire.

e. Pyramid Fire. The pyramid fire looks similar to a log cabin fire except there are layers of fuel in place of a hollow framework. The advantage of a pyramid fire is that it burns for a long time resulting in a large bed of coals. This fire could possibly be used as an overnight fire when placed in front of a shelter opening.

f. Star Fire. This fire is used when conservation of fuel is necessary or a small fire is desired. It burns at the center of the "wheel" and must be constantly tended. Hardwood fuels work best with this type of fire.

g. "T" Fire. Used for large group cooking. The size of this lay may be adjusted to meet the group's cooking needs. In the top part of the "T," the fire is constructed and maintained as long as needed to provide hot coals for cooking in the bottom part of the "T" fire lay. The number of hot coals may be adjusted in the lower part of the "T" fire lay to regulate the cooking temperature.

h. "V" Fire. This fire lay is a modification of the long fire. The configuration allows a survivor to either block strong winds, or take advantage of light breezes. During high wind conditions, the vertex of the "V"—formed by the two outside logs—is placed in the direction from which the winds are coming, thereby sheltering the tinder (kindling) for ignition. Reversing the "lay" will funnel light breezes into the tinder (kindling) thereby facilitating ease of ignition (figure 16-1).

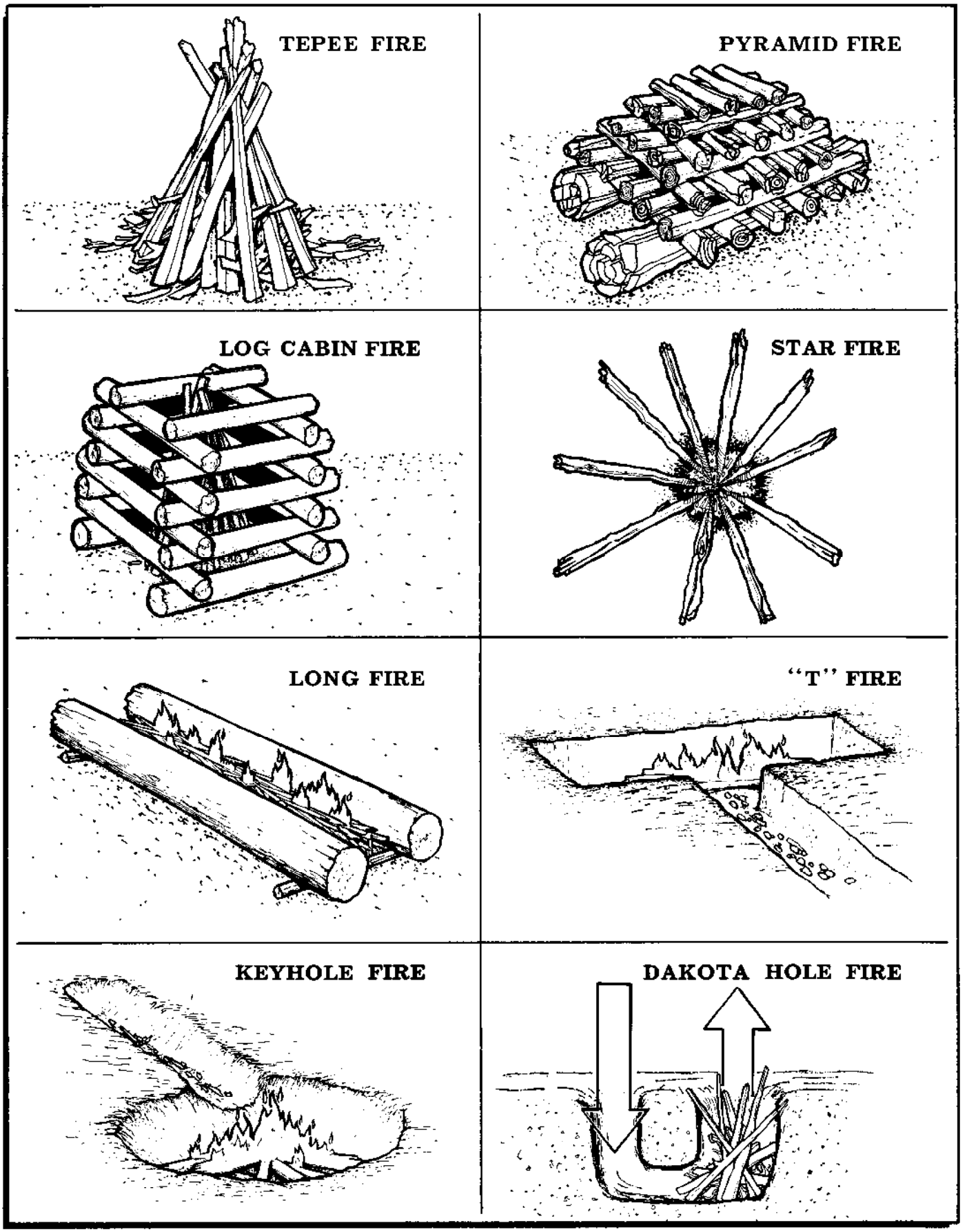

Figure 16-14. Fire Lays.

Chapter 17

EQUIPMENT

17-1. Introduction. Survivors in a survival situation have needs which must be met—food, water, clothing, shelters, etc. The survival kit contains equipment which can be used to satisfy these needs. Quite often, however, this equipment may not be available due to damage or loss. This chapter will address the care and use of issued equipment and improvising the needed equipment when not available. The uses of some issued items are covered in appropriate places throughout this regulation. The care and use of equipment (not covered elsewhere) will be addressed here.

17-2. Types of Kits:

a. All survival kits contain two types of equipment—mandatory and optional. The mandatory equipment for survival kits are:

(1) One-man liferaft (1 each).
(2) Compass (1 each).
(3) Smoke and illumination flares (2 each).
(4) Signal mirror (1 each).
(5) Hand-held launched flare (1 each).
(6) First aid kit (1 each).
(7) Survival radio (1 each).

NOTE: These items of equipment may not be mandatory for raft kit.

b. Optional items are authorized by the major air commands; this authority is delegated to subordinate commanders. These optional items are directly related to climatic conditions and the type of terrain which is being flown over. There are over 40 optional items. Here are a few examples:

(1) Sleeping bag.
(2) Strobelight with lenses.
(3) Wire saw.
(4) Water container.
(5) Survival shovel.
(6) Matchbox container.

17-3. Issued Equipment. Survival equipment is designed to aid survivors throughout their survival episode. To maintain its effectiveness, the equipment must be well cared for.

a. Electronic Equipment:

(1) Electronic signaling devices are by far the survivors' most important signaling devices. Therefore, it is important for survivors to properly care for them to ensure their continued effectiveness. In cold temperatures, the electronic signaling devices must be kept warm to prevent the batteries from becoming cold soaked.

(2) In a cold environment, if survivors speak directly into the microphone, the moisture from their breath may condense and freeze on the microphone, creating communication problems.

(3) Caution must be used when using the survival radios in a cold environment. If the radio is placed against the side of the face to communicate, frostbite could result.

(4) In a wet environment, survivors should make every effort to keep their electronic signaling devices dry. In an open-sea environment, the only recourse may be to shake the water out of the microphone before transmitting.

b. Firearms:

(1) A firearm is a precision tool. It will continue functioning only as long as it is cared for. Saltwater, perspiration, dew, and humidity can all corrode or rust a firearm until it is inoperable. If immersed in saltwater, the survivor should wash the parts in freshwater and then dry and oil them. As an expedient, one way to dry the firearm is to place it in boiling water and after removal, wipe off the excess moisture. The residual heat will evaporate most of the remaining moisture. Survivors should not use uncontrolled heat to dry the firearm as heat over 250°F can remove the temper from the springs in a short time and weaken the action.

(2) Any petroleum-based lubricants used in cold environments will stiffen or freeze causing the firearm to become inoperative. It would be better to thoroughly clean the firearm and remove all lubricant. Metal becomes brittle from cold and is, therefore, prone to breakage.

(3) A firearm was not intended for use as a club, hammer, or pry bar. To use it for any purpose other than for which it was designed, would only result in damage to the firearm.

c. Cutting Tools:

(1) A file and sharpening stone are often packed in a survival kit. The file is normally used for axes, and the stone is normally used for knives.

(2) An old axiom states that a sharp cutting tool is a safe cutting tool. Control of a cutting tool is easier to maintain if it is sharp, and the possibility of accidental injury is reduced.

(3) One of the most valuable items in any survival situation is a knife, since it has a large number of uses. Unless the knife is kept sharp, however, it falls short of its potential.

(4) A knife should be sharpened only with a stone as repeated use of a file rapidly removes steel from the blade. In some cases, it may be necessary to use a file to remove plating from the blade before using the stone.

(5) One of two methods should be used to sharpen a knife. One method is to push the blade down the stone

in a slicing motion. Then turn the blade over and draw the blade toward the body (figure 17-1).

(6) The other method is to use a circular motion the entire length of the blade; turn the blade over and repeat

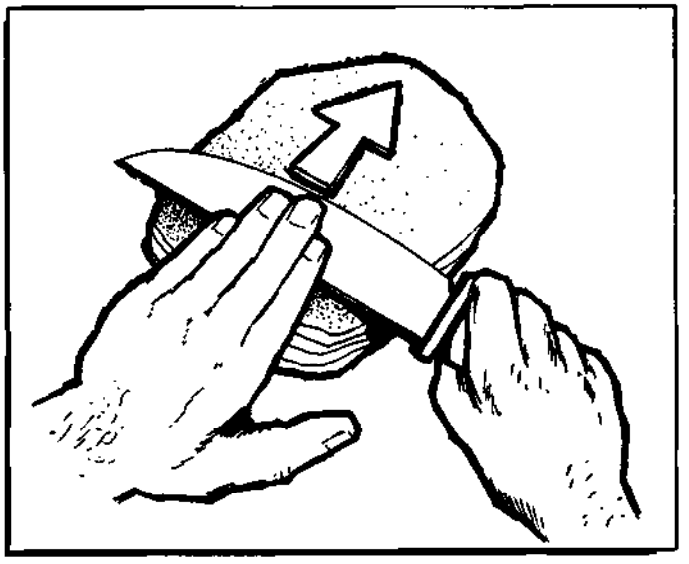

Figure 17-1. Knife Sharpening (Draw).

the process. What is done to one side of the cutting edge should also be done to the other to maintain an even cutting edge (figure 17-2).

Figure 17-2. Knife Sharpening (Circular).

(7) Most sharpening stones available to survivors will be whetstones. Water should be applied to these stones. The water will help to float away the metal removed by sharpening and make cleaning of the stone easier.

(8) If a commercial whetstone is not available, a natural whetstone can be used. Any standstone will sharpen tools, but a gray, clay-like sandstone gives better results. Quartzite should be avoided. Survivors can recognize quartzite instantly by scratching the knife blade with it—the quartz crystals will bite into steel. If no sandstone is available, granite or crystalline rock can be used. If granite is used, two pieces of the stone should be rubbed together to smooth the surface before use.

(9) As with a knife, a sharp axe will save time and energy and be much safer.

(10) A file should be used on an axe or hatchet. Survivors should file away from the cutting edge to prevent injury if the file should slip. The file should be worked from one end of the cutting edge to the other. The opposite side should be worked to the same degree. This will ensure that the cutting edge is even. After using a file, the stone may be used to hone the axe blade (figure 17-3).

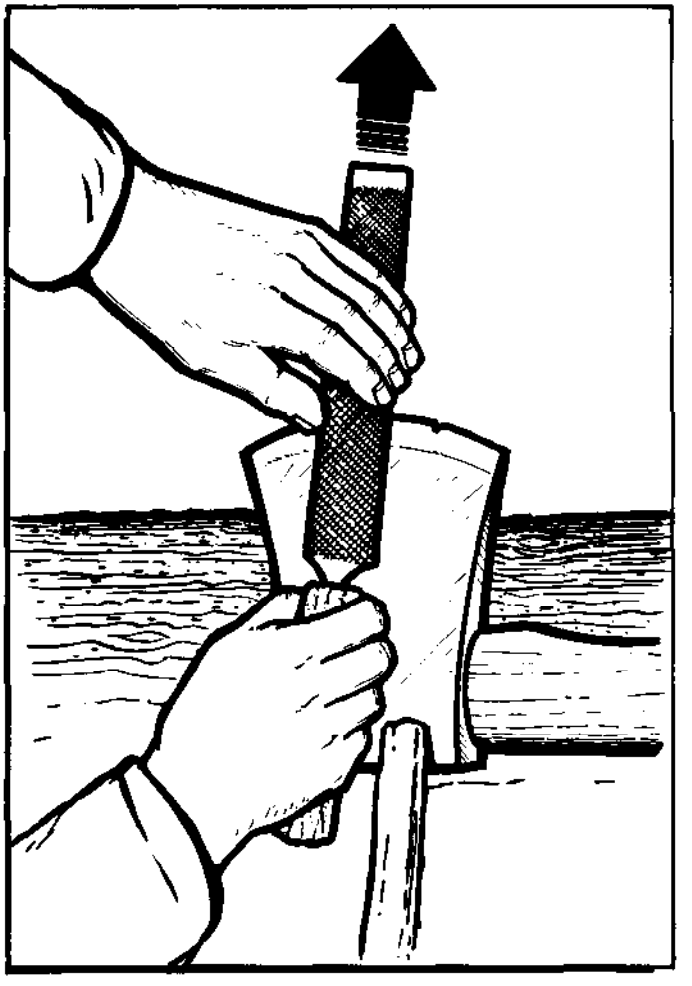

Figure 17-3. Sharpening Axe.

(11) When using an axe, don't try to cut through a tree with one blow. Rhythm and aim are more important than force. Too much power behind a swing interferes with aim. When the axe is swung properly, its weight provides all the power needed.

(12) Carving a new axe handle and mounting the axe head takes a great deal of time and effort. For this reason, a survivor should avoid actions which would require the handle to be changed. Using aim and paying attention to where the axe falls will prevent misses which could result in a cracked or broken handle. Survivors should not use an axe as a pry bar and should avoid leaving the axe out in cold weather where the handle may become brittle.

(13) A broken handle is difficult to remove from the head of the axe. Usually the most convenient way is to burn it out (figure 17-4). For a single-bit axe, bury the bit in the ground up to the handle, and build a fire over it. For a double-bit, a survivor should dig a small trench, lay the middle of the axe head over it, cover both "bits" with earth, and build the fire. The covering of earth keeps the flame from the cutting edge of the axe and saves its temper. A little water added to the earth will further ensure this protection.

(14) When improvising a new handle, a survivor can save time and trouble by making a straight handle instead of a curved one like the original. Survivors should use a young, straight piece of hardwood without knots. The wood should be whittled roughly into shape and finished by shaving. A slot should be cut into the axe-head end of the handle. After it is fitted, a thin, dry wooden wedge can then be pounded into the slot. Survivors should use the axe awhile, pound the wedge in again, then trim it off flush with the axe. The handle must be smoothed to remove splinters. The new handle can be seasoned to prevent shrinkage by "scorching" it in the fire.

d. Whittling:

(1) Whittle means to cut, trim, or shape (a stick or piece of wood) by taking off bits with a knife. Survivors should be able to use the techniques of whittling to help save time, energy, and materials as well as to prevent injuries. They will find that whittling is a necessity in constructing triggers for traps and snares, shuttles and spacers, and other improvised equipment.

(2) When whittling, survivors must hold the knife firmly and cut away from the body (figure 17-5). Wood should be cut with the grain. Branches should be trimmed as shown in figure 17-6.

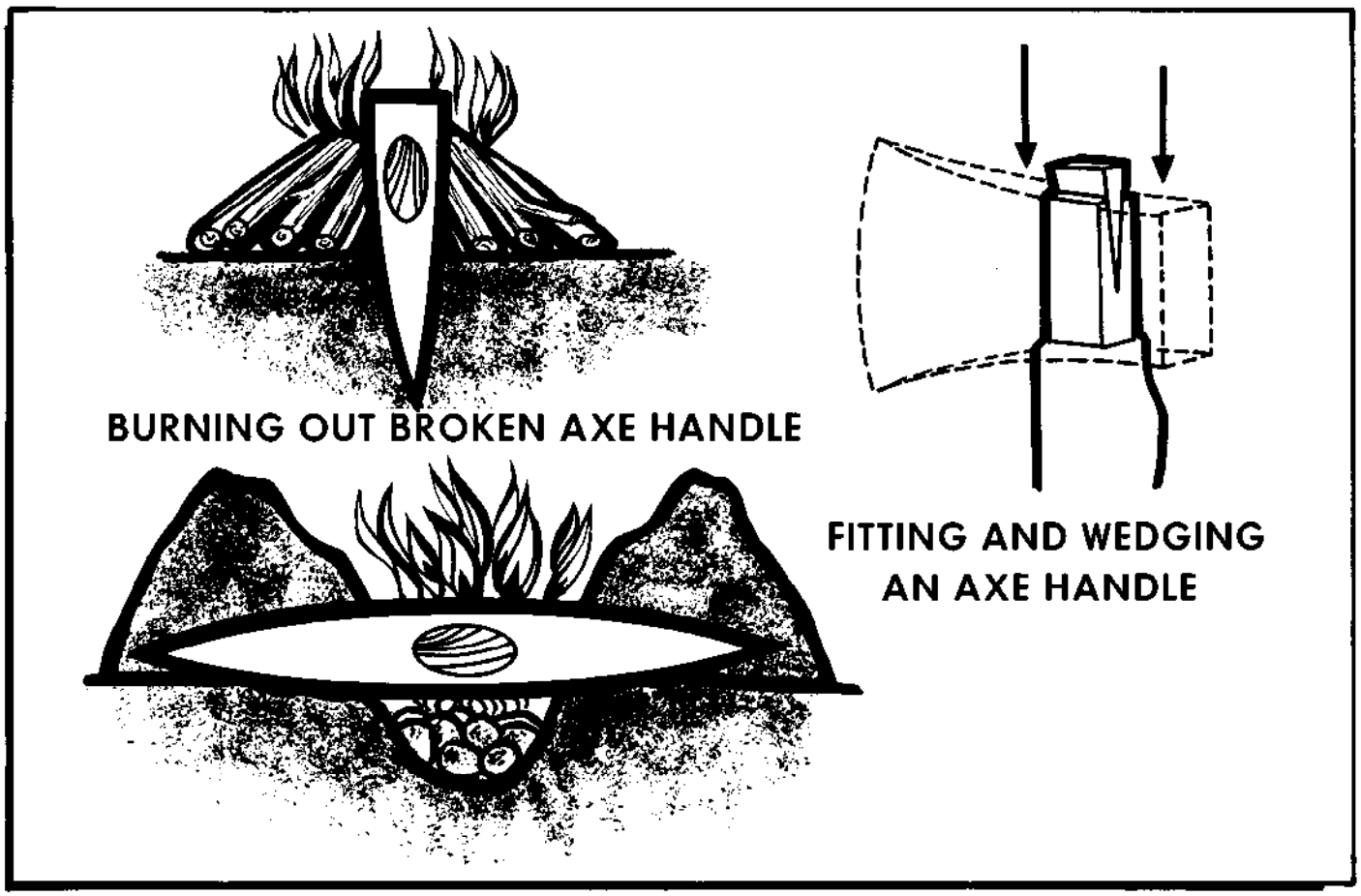

Figure 17-4. Removing Broken Axe Handle.

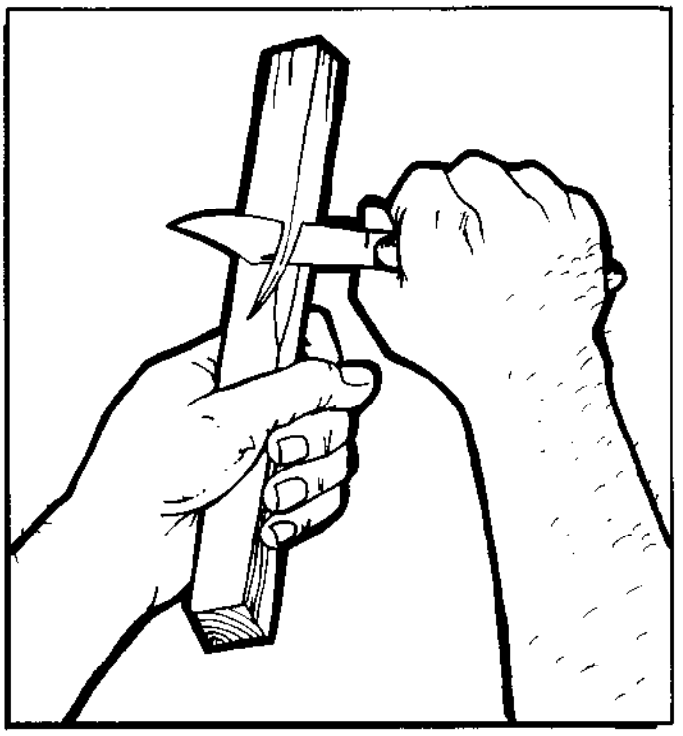

Figure 17-5. Whittling.

(3) To cut completely through a piece of wood, a series of V-cuts should be made all the way around as in figure 17-7. Once the piece of wood has been severed, the pointed end can then be trimmed.

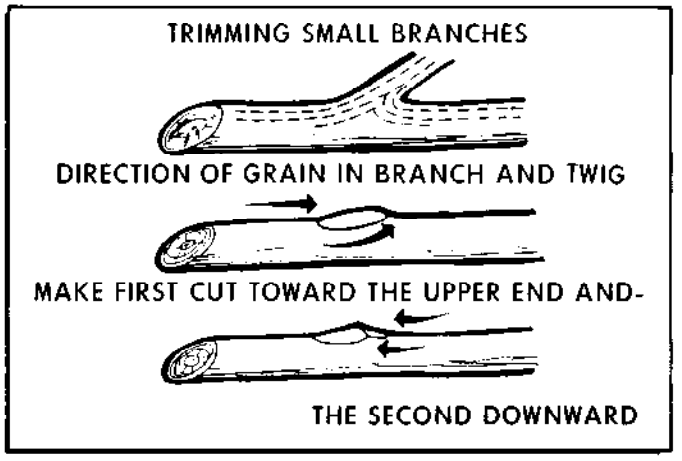

Figure 17-6. Trimming Branches.

(4) The thumb can be used to help steady the hand. Be sure and keep the thumb clear of the blade. To maintain good control of the knife, the right hand is steadied with the right thumb while the left thumb pushes the blade forward (figure 17-8). This method is very good for trimming.

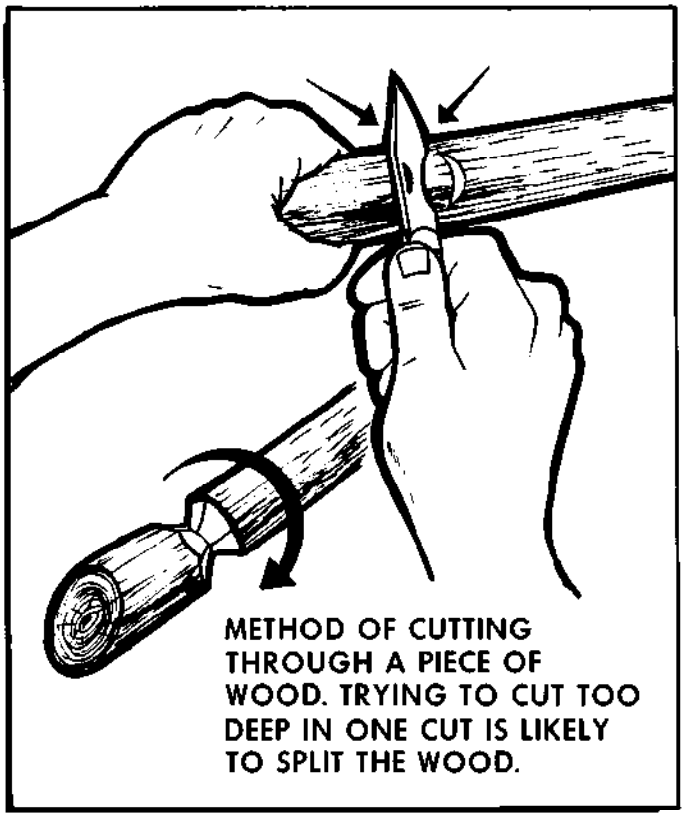

Figure 17-7. Cutting Through a Piece of Wood.

e. Felling Trees:

(1) To fell a tree, the survivor must first determine the direction in which the tree is to fall. It is best to fell the tree in the direction in which it is leaning. The lean of the tree can be found by using the axe as a plumb line

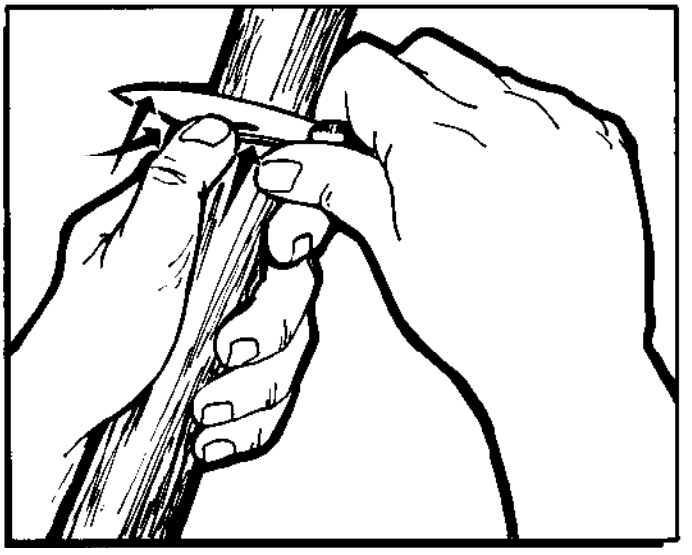

Figure 17-8. Fine Trimming.

(figure 17-9). The survivor should then clear the area around the tree from underbrush and overhanging branches to prevent injury (figure 17-10).

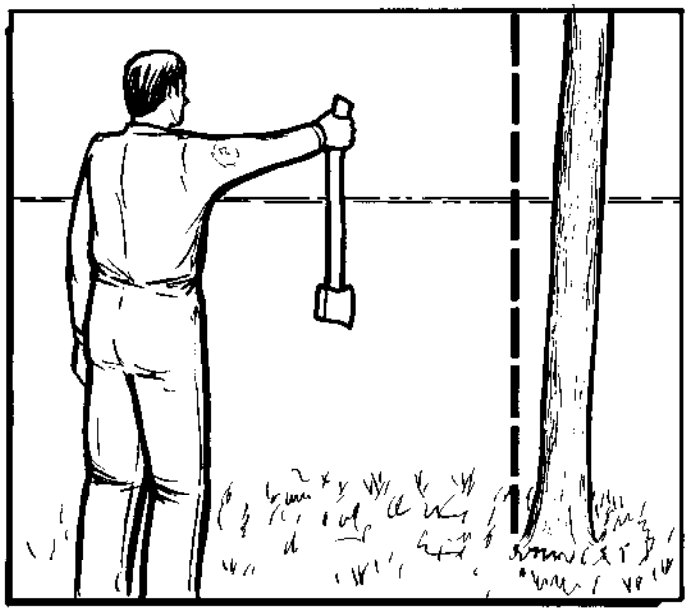

Figure 17-9. Using the Axe as a Plumb Line.

(2) The survivor should make two cuts. The first cut should be on the leaning side of the tree and close to the ground and the second cut on the opposite side and a little higher than the first cut (figure 17-11).

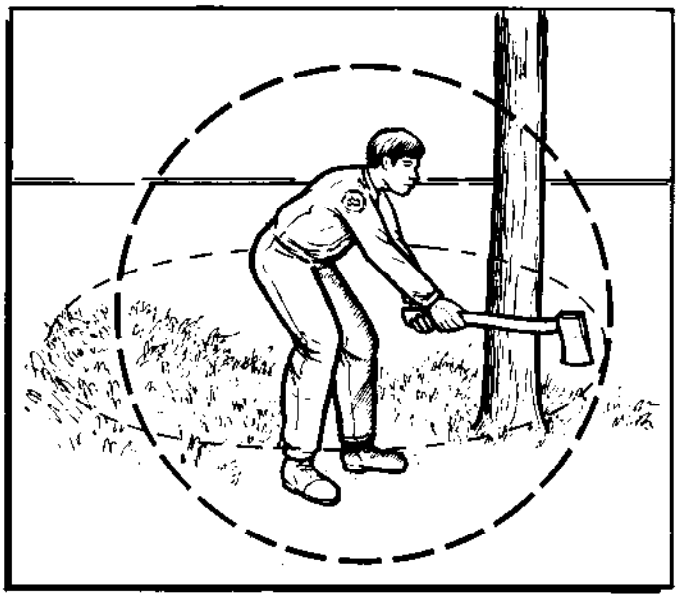

Figure 17-10. Clearing Brush from Cutting Area.

(3) Falling trees often kick back and can cause serious injury (figure 17-12), so survivors must ensure they have a clear escape route. When limbing a tree, start at

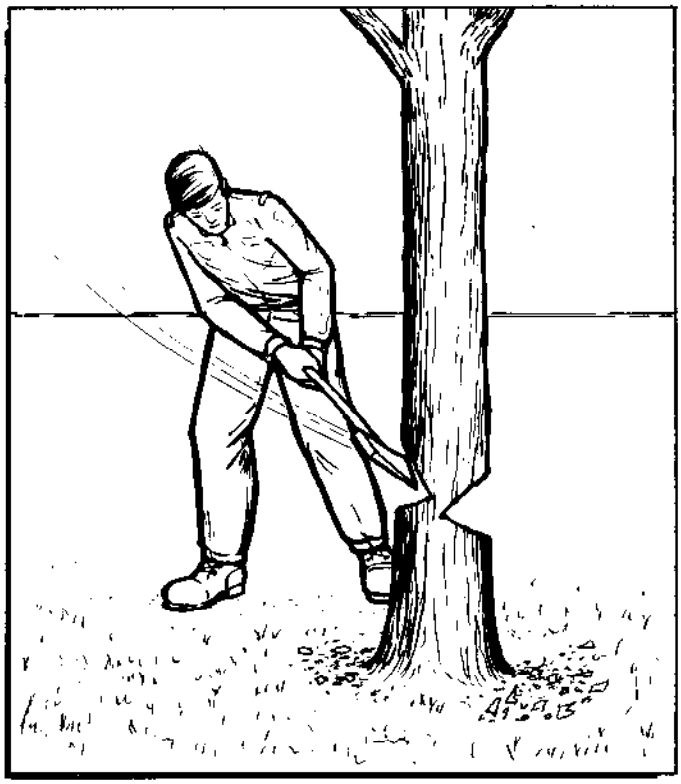

Figure 17-11. Felling Cuts.

the base of the tree and cut toward the top. This procedure will allow for easier limb removal and results in a smoother cut. For safety, the survivor should stand on one side of the trunk with the limb on the other.

(4) To prevent damage to the axe head and possible physical injury, any splitting of wood should be done on a log as in figure 17-13. The log can also be used for cutting sticks and poles (figure 17-14).

(5) To make cutting of a sapling easier, bend it over with one hand, straining grain. A slanting blow close to the ground will cut the sapling (figure 17-15).

17-4. Improvised Equipment:

a. If issued equipment is inoperative, insufficient, or nonexistent, survivors will have to rely upon their ingenuity to manufacture the needed equipment. Survivors must determine whether the need for the item outweighs the work involved to manufacture it. They will also have to evaluate their capabilities. If they have injuries, will the injuries prevent them from manufacturing the item(s)?

b. Undue haste may not only waste materials, but also waste the survivors' time and energy. Before manufacturing equipment, they should have a plan in mind.

c. The survivors' equipment needs may be met in two different ways. They may alter an existing piece of equipment to serve more than one function, or they may also construct a new piece of equipment from

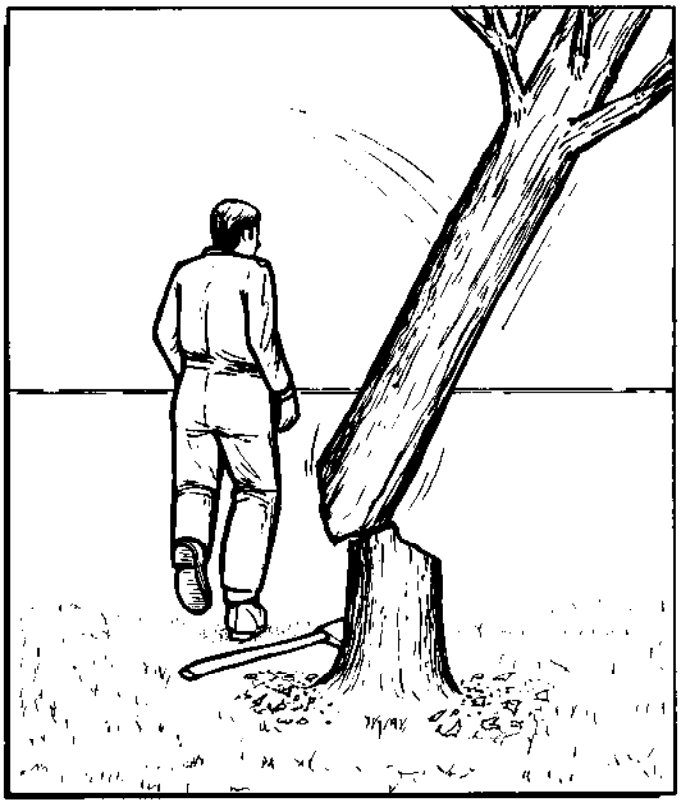

Figure 17-12. Tree Kickbacks.

available materials. Since the items survivors can improvise are limited only by their ingenuity, all improvised items cannot be covered in this regulation.

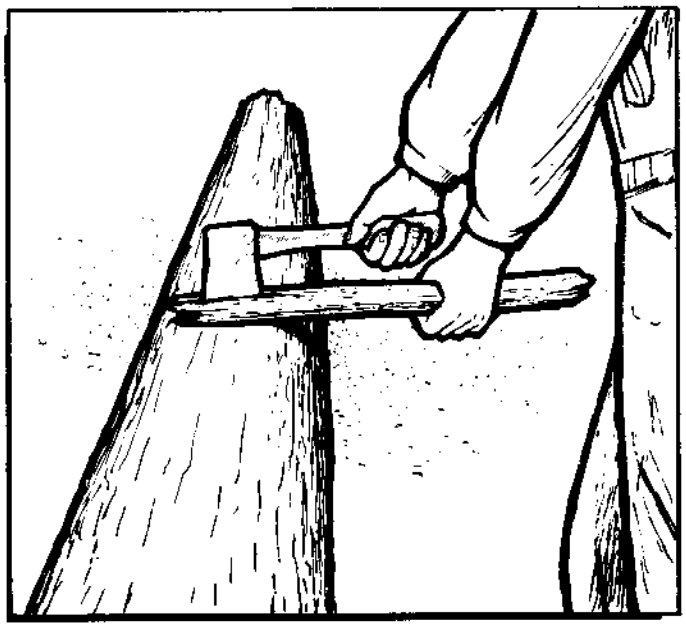

Figure 17-13. Splitting Wood.

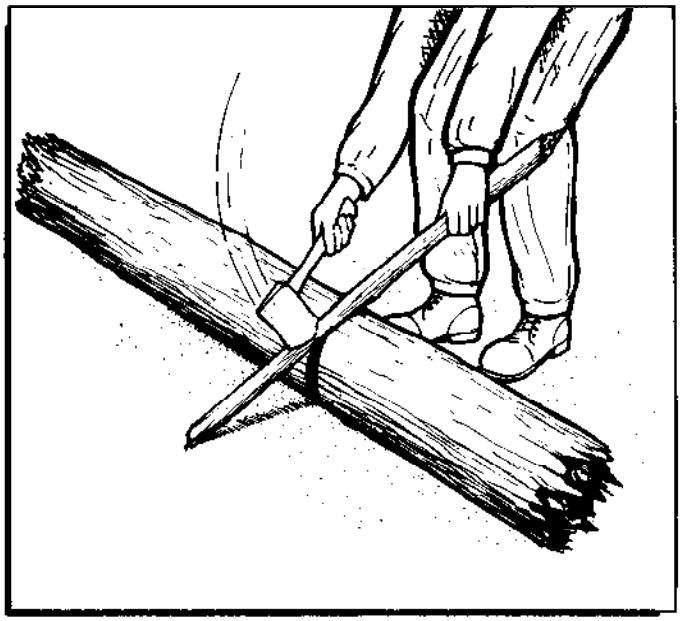

Figure 17-14. Cutting Poles.

d. The methods of manufacturing the equipment referred to in this regulation are only ideas and do not have to be strictly adhered to. Many Air Force survivors have a parachute. This device can be used to improvise a variety of needed equipment items.

e. The parachute consists of (figure 17-16):

(1) The pilot chute which deploys first and pulls the rest of the parachute out.

(2) The parachute canopy which consists of the apex (top) and the skirt or lower lateral band. The canopy material is divided by radial seams into 28 sections called gores. Each gore measures about 3 feet at the skirt and tapers to the apex. Each gore is further subdivided into four sections called panels. The canopy is normally divided into four colors. These colored areas are intended to aid the survivor in shelter construction, signaling, and camouflage.

(3) Fourteen suspension lines connect the canopy material to the harness assembly. Each piece of suspension line is 72 feet long from riser to riser and 22 feet long from riser to skirt and 14 feet from skirt to apex. The tensile strength of each piece of suspension line is 550 pounds. Each piece of suspension line contains seven to nine pieces of innercore with a tensile strength of 35 pounds. The harness assembly contains risers and webbing, buckles, snaps, "D" rings, and other hardware which can be used in improvisation.

f. The whole parachute assembly should be considered as a resource. Every piece of material and hardware can be used.

(1) To obtain the suspension lines, a survivor should cut them at the risers or, if time and conditions permit, consider disassembling the connector links. Cut

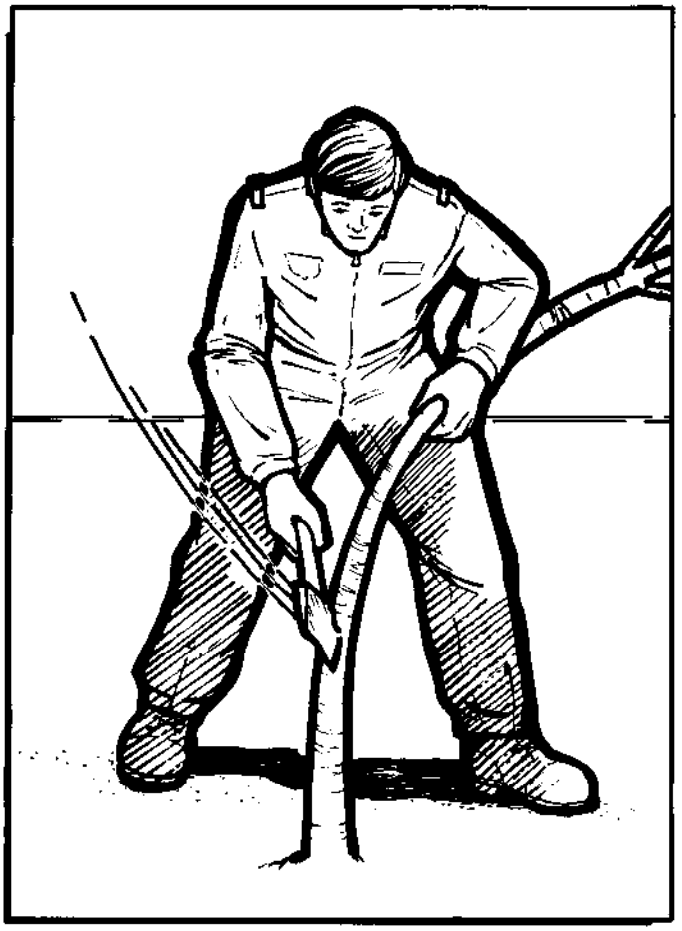

Figure 17-15. Cutting Saplings.

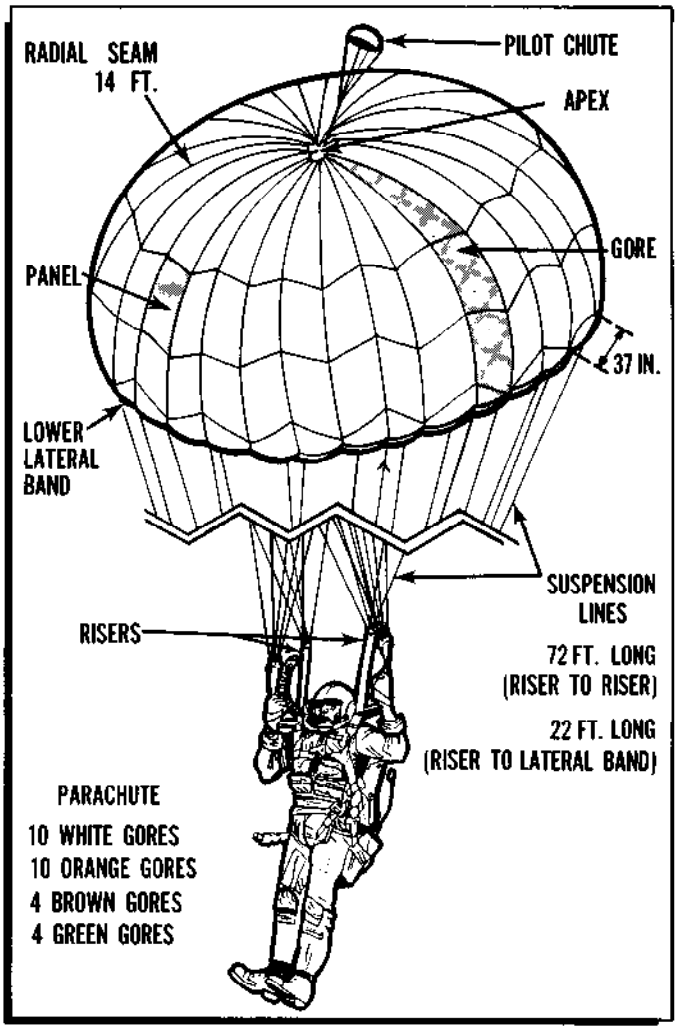

Figure 17-16. Parachute Diagram.

the suspension lines about 2 feet from the skirt of the canopy. When cutting suspension lines or dismantling the canopy/pack assembly, it will be necessary to maintain a sharp knife for safety and ease of cutting.

(2) Survivors should obtain all available suspension line due to its many uses. Even the line within the radial seams of the canopy should be stripped for possible use. The suspension line should be cut above the radial seam stitches next to the skirt end of the canopy (two places). The cut should not go all of the way through the radial seam (figure 17-17). At the apex of the canopy, and just below the radial seam stitching, a horizontal cut can be made and the suspension line extracted. The line can then be cut.

(3) For maximum use of the canopy, survivors must plan its disassembly. The quantity requirements for shelter, signaling, etc., should be thought out and planned for. Once these needs have been determined, the canopy may be cut up. The radial seam must be stretched tightly for ease of cutting. The radial seam can then be cut by holding the knife at an angle and following the center of the seam. With proper tension and the gentle pushing (or pulling) of a sharp blade there will be a controlled splitting of the canopy at the seam (figure 17-17). It helps to secure the apex either to another individual or to an immobile object such as a tree.

(4) When stripping the harness assembly, the seams of the webbing should be split so the maximum usable webbing is obtained. The harness material and webbing should not be randomly cut as it will waste much needed materials.

g. One requirement in improvising is having available material. Parachute fabric, harness, suspension lines, etc., can be used for clothing. Needles are helpful for making any type of emergency clothing. Wise survivors should always have extra sewing needles hidden somewhere on their person. A good needle or sewing awl can be made from the can-opening key from the ration tin (figure 17-18) or, as the Eskimos do, from a sliver of bone. Thread is usually available in the form of innercore. It will be to the survivor's benefit to collect small objects which may "come in handy." Wire, nails, buttons, a piece of canvas, or animal skin should not be

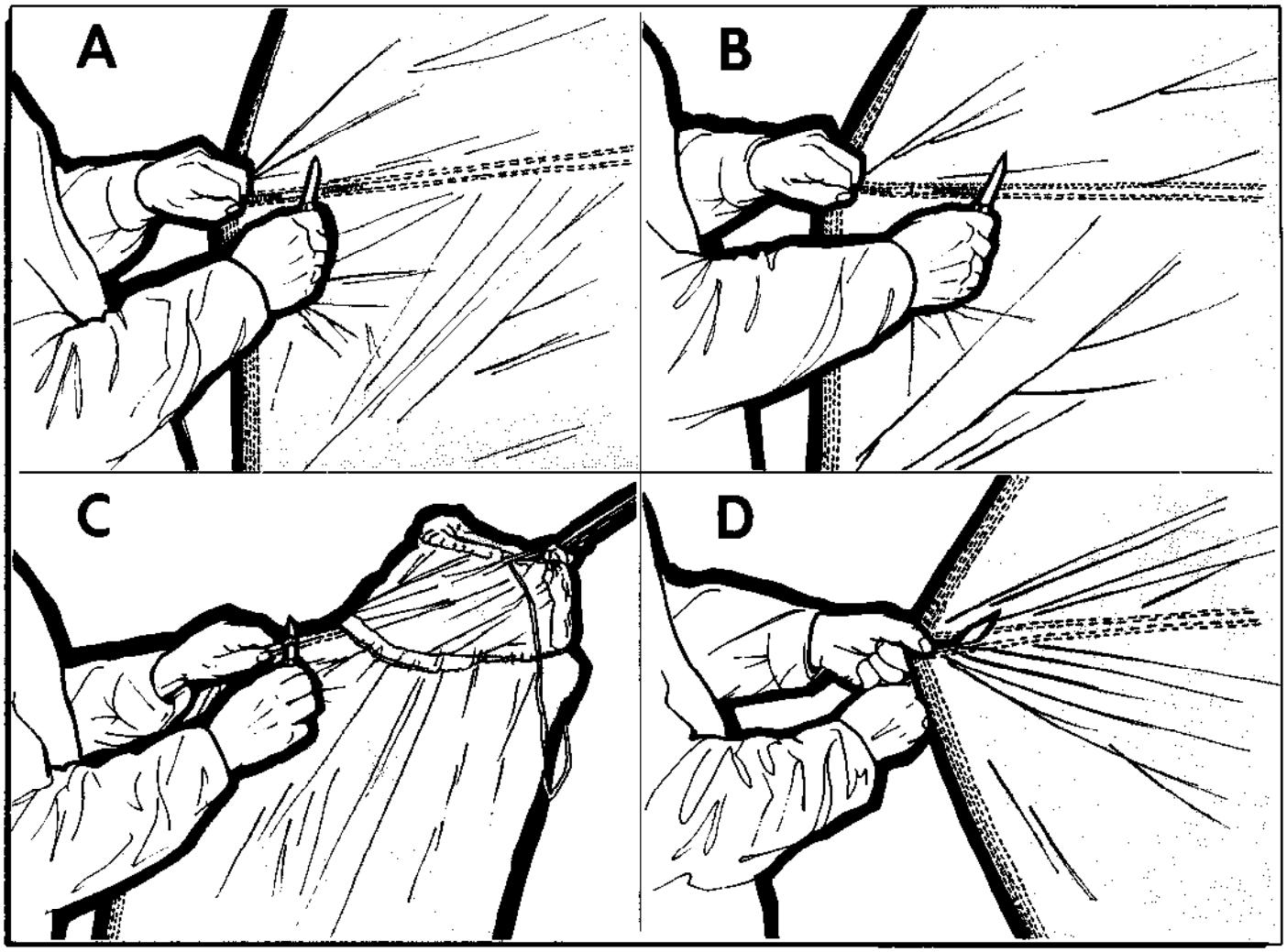

Figure 17-17. Cutting the Parachute.

discarded. Any such object may be worth its weight in gold when placed in a hip pocket or a sewing kit. Any kind of animal skin can be used for making clothing such as gloves or mittens or making a ground cover to keep the sleeping bag dry and clean. Small skins can be used for mending and for boot insoles. Mending and cleaning clothes when possible will pay dividends in health, comfort, and safety.

h. The improvised equipment survivors may need to make will probably involve sewing. The material to be sewn may be quite thick and hard to sew, and to keep from stabbing fingers and hands, a palm-type thimble can be improvised (figure 17-19). A piece of webbing, leather, or other heavy material, with a hole for the thumb, is used. A flat rock, metal, or wood is used as the thimble and this is held in place by a doughnut-shaped piece of material sewn onto the palm piece. To use, the end of the needle with the eye is placed on the thimble and the thimble is then used to push the needle through the material to be sewn.

17-5. Miscellaneous Improvised Equipment:

a. Improvised Trail-Type Snowshoes. The snowshoe frame can be made from a sapling 1 inch in diameter and 5 feet long. The sapling should be bent and spread to 12 inches at the widest point. The survivor can then include the webbing of suspension lines (figure 17-20). The foot harness, for attaching the snowshoe to the boot, is also fashioned from suspension line.

b. Improvised Bear Paw-Type Snowshoes. A sapling can be held over a heat source and bent to the shape shown in figure 17-21. Wire from the aircraft or parachute suspension line can be used for lashing and for making webbing. Snowshoes can also be quickly improvised by cutting a few pine boughs and lashing them together at the cut ends. The lashed boughs positioned with the cut ends forward can then be tied to the feet (figure 17-22).

(1) Survivors should guard against frostbite and blistering while snowshoeing. Due to the design of the

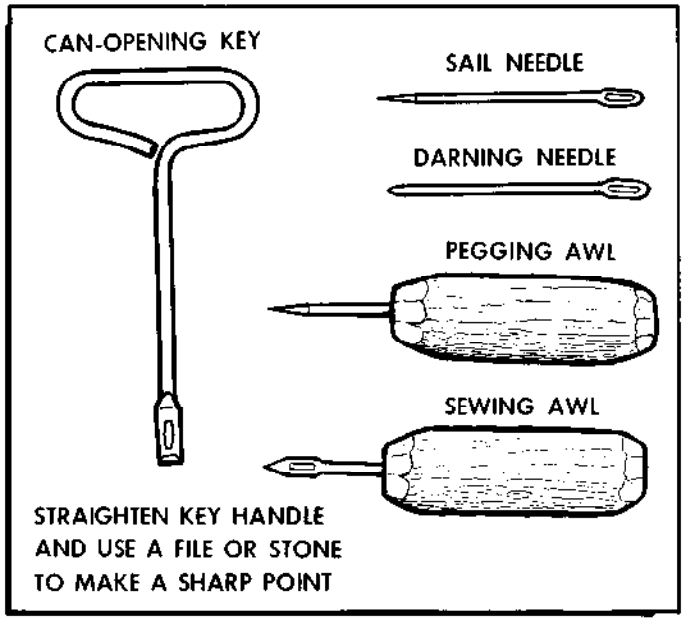

Figure 17-18. Needle and Sewing Awl.

harness, the circulation of the toes is usually restricted, and the hazard of frostbite is greater. They should check the feet carefully, stop often, take off the harness, and massage the feet when they seem to be getting cold.

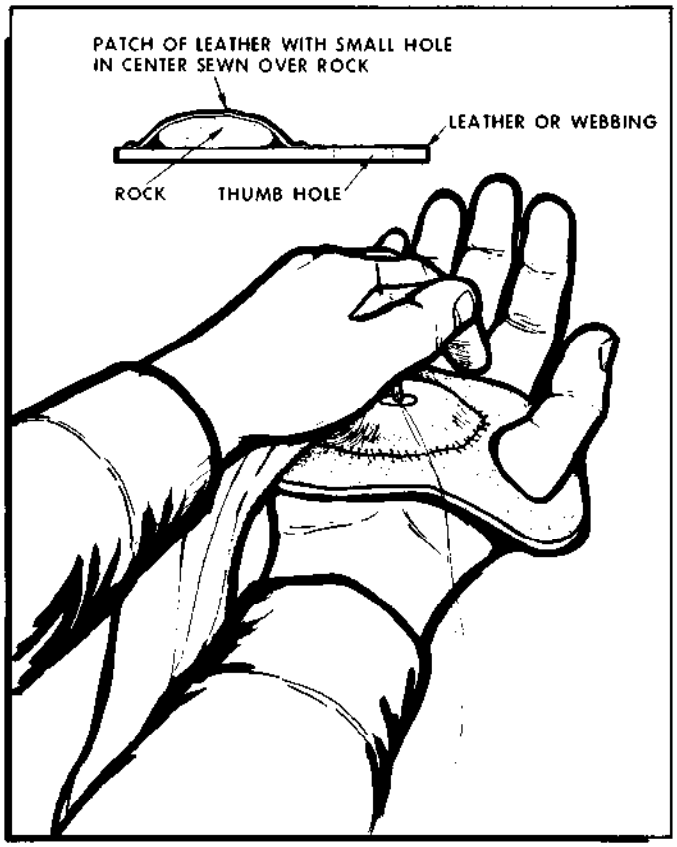

Figure 17-19. Palm Thimble.

(2) Blistering between the toes or on the ball of the foot is sometimes unavoidable in a "tenderfoot" if much snowshoeing is done. To make blisters less likely, the survivor should keep socks and insoles dry and change them regularly.

c. Sleeping Bag. Immediate action should be to use the whole parachute until conditions allow for improvising. A sleeping bag can be improvised by using four gores of parachute material or an equivalent amount of other materials (figure 17-23). The material should be folded in half lengthwise and sewn at the foot. To measure the length, the survivor should allow an extra 6 to 10 inches in addition to the individual's height. The two raw edges can then be sewn together. The two sections of the bag can be filled with cattail down, goat's beard lichen, dry grass, insulation from aircraft walls, etc. The

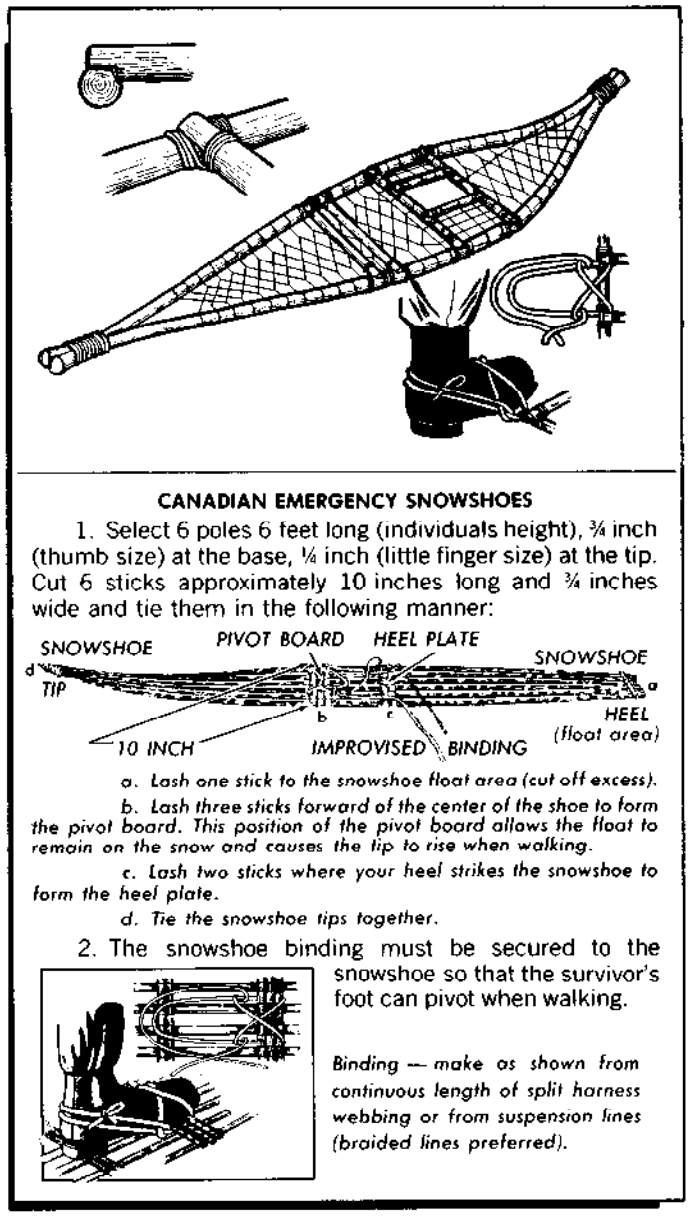

Figure 17-20. Improvised Trail Snowshoes.

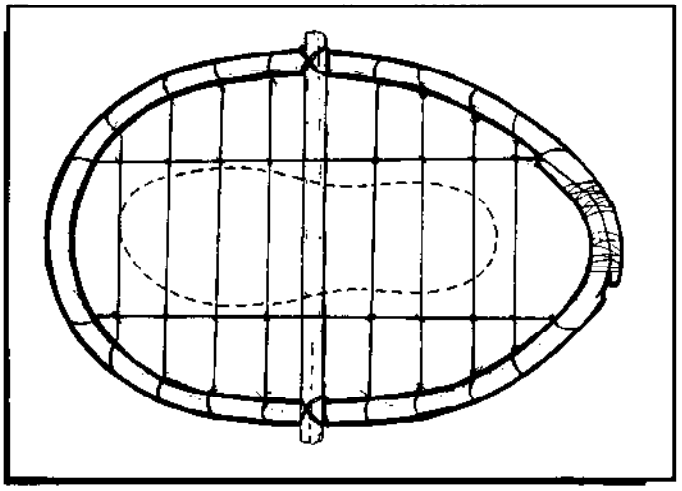

Figure 17-21. Improvised Bear Paws.

stuffed sleeping bag should then be quilted to keep the insulation from shifting. The bag can be folded in half lengthwise and the foot and open edges sewn. The length and width can be adjusted for the individual.

d. Insulating Bed:

(1) In addition to the sleeping bag, some form of ground insulation is advisable. An insulation mat will help insulate the survivor from ground moisture and the cold. Any nonpoisonous plants such as ferns and grasses will suffice. Leaves from a deciduous tree make a comfortable bed. If available, extra clothing, seat cushions, aircraft insulation, rafts, and parachute material may be used. In a coniferous forest, boughs from the trees would do well if the bed is constructed properly.

(2) The survivor should start at the foot of the proposed bed and stick the cut ends in the ground at about a 45-degree angle and very close together. The completed bed should be slightly wider and longer than the body. If the ground is frozen, a layer of dead branches can be used on the ground with the green boughs placed in the dead branches, similar to sticking them in the ground.

(3) A bough bed should be a minimum of 12 inches thick before use. This will allow sufficient insulation between the survivor and the ground once the bed is compressed. The bough bed should be fluffed up and boughs added daily to maintain its comfort and insulation capabilities.

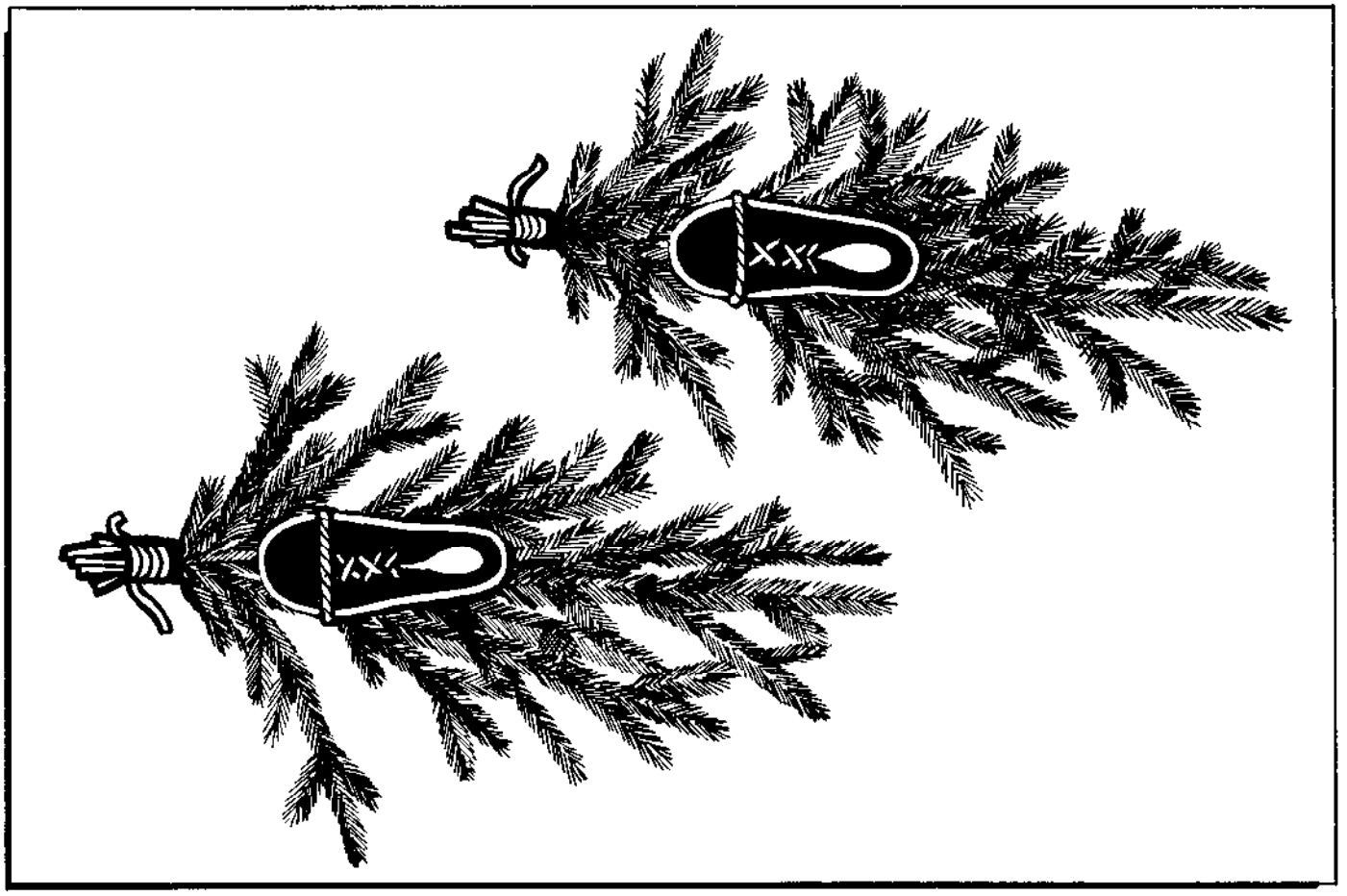

Figure 17-22. Bough Snowshoes.

(4) Spruce boughs have many sharp needles and can cause some discomfort. Also the needles on various types of pines are generally located on the ends of the boughs, and it would take an abundance of pine boughs to provide comfort and insulation. Fir boughs on the other hand, have an abundance of needles all along the boughs and the needles are rounded. These boughs are excellent for beds, providing comfort and insulation (figure 17-24).

e. Rawhide. Rawhide is a very useful material which can be made from any animal hide. Processing it is time consuming but the material obtained is strong and very durable. It can be used for making sheaths for cutting tools, lashing materials, ropes, etc.

(1) The first step in making rawhide is to remove all of the fat and muscle tissue from the hide. The large pieces can be cut off and the remainder scraped off with a dull knife or similar instrument.

(2) The next step is to remove the hair. This can be done by applying a thick layer of wood ashes to the hair side. Ashes from a hardwood fire work best. Thoroughly sprinkle water all over the ashes. This causes lye to leach out of the ash. The lye will remove the hair. The hide should be rolled with the hair side in and stored in a cool place for several days. When the hair begins to slip (check by pulling on the hair), the hide should be unrolled and placed over a log. Remove the hair by scraping it off with a dull knife. Once the hair is removed, the

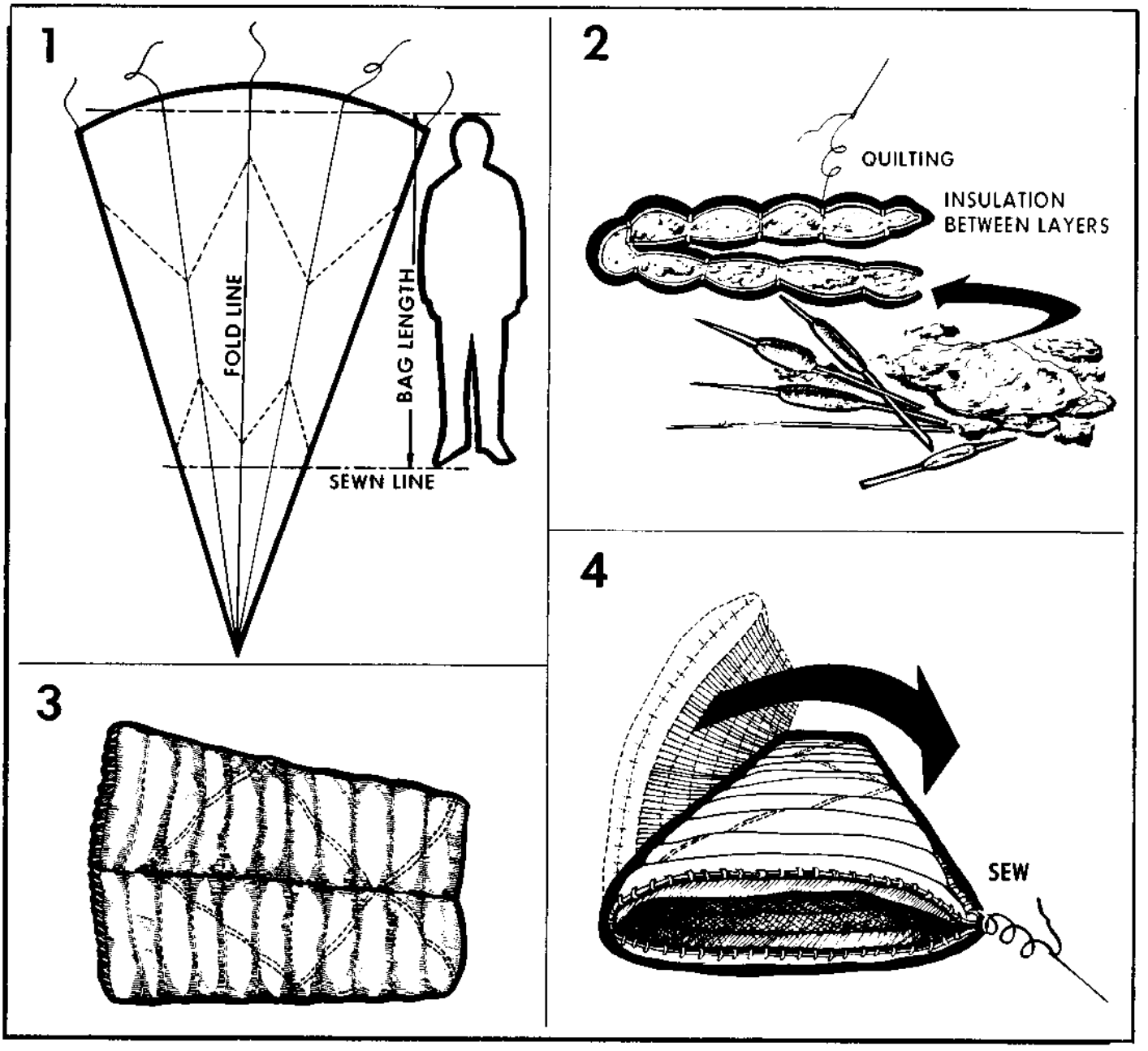

Figure 17-23. Improvised Sleeping Bag.

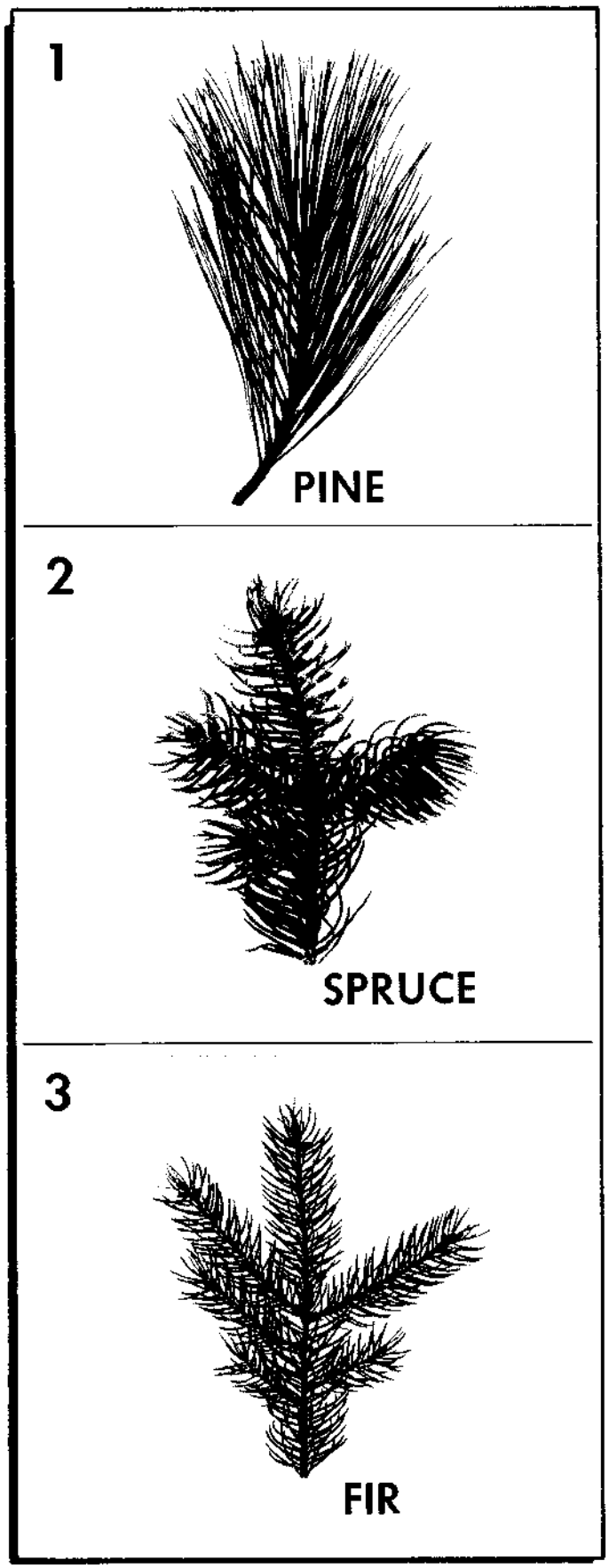

Figure 17-24. Boughs.

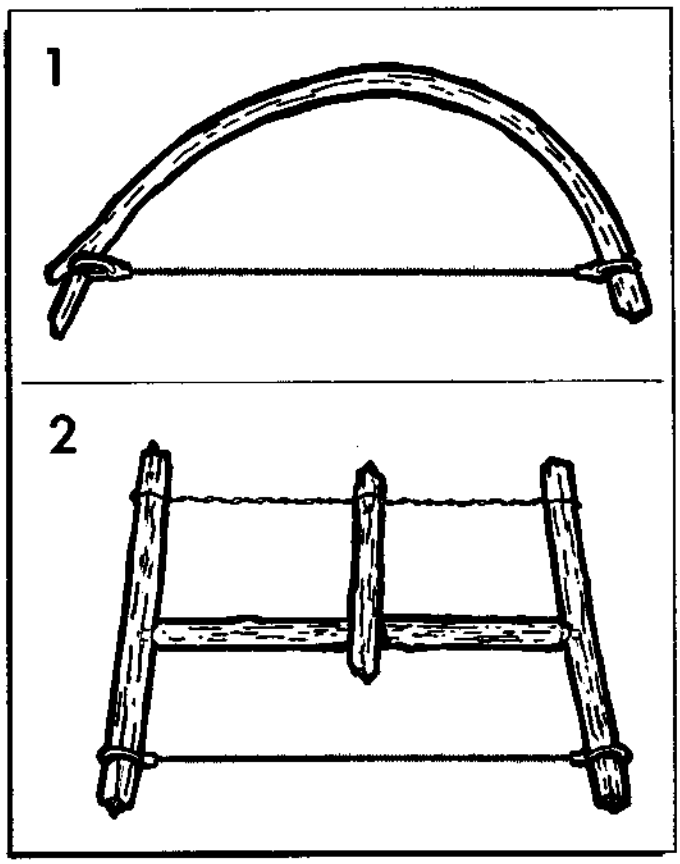

Figure 17-25. Bow Saw and Buck Saw.

hide should be thoroughly washed, stretched inside a frame, and allowed to dry slowly in the shade. When dry, rawhide is extremely hard. It can be softened by soaking in water.

f. Wire Saws. Wire or pieces of metal can be used to replace broken issued saws. With minor modifications, the survivor can construct a usable saw. A bow-saw

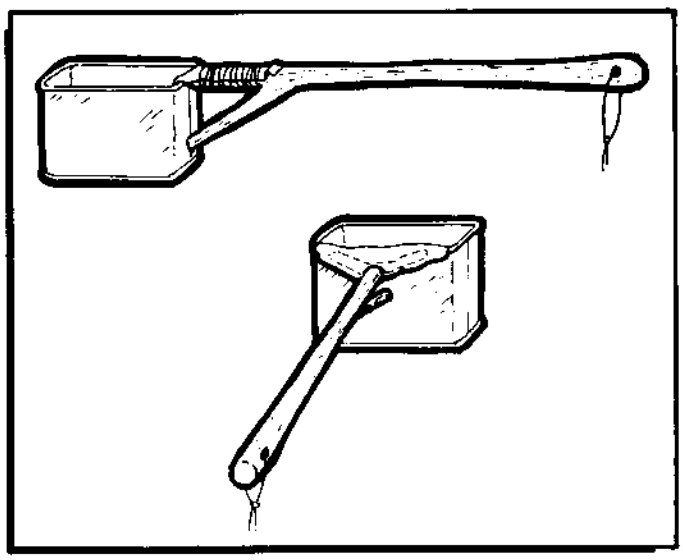

Figure 17-26. Cooking Utensils.

arrangement will help to prevent the blade from flexing. A green sapling may be used for the bow as shown in figure 17-25. If a more durable saw is required and time permits, a bucksaw may be improvised (figure 17-25). Blade tension can be maintained by use of a tightening device known as a "windlass" (figure 17-25).

g. Cooking Utensils. Ration tins can serve as adequate cooking utensils. If the end has been left intact as in figure 17-26, use a green stick long enough to prevent burning the hand while cooking. If the side has been left intact, a forked stick may be used to add support to the container (figure 17-26).

17-6. Ropes and Knots:

a. Basic Knowledge of Tying a Knot. A basic knowledge of correct rope and knot procedures will aid the survivor to do many necessary actions. Such actions as improvising equipment, building shelters, assembling packs, and providing safety devices require the use of proven techniques. Tying a knot incorrectly could result in ineffective improvised equipment, injury, or death.

b. Rope Terminology: (See figure 17-27.)

(1) Bend. A bend (called a knot in this regulation) is used to fasten two ropes together or to fasten a rope to a ring or loop.

(2) Bight. A bight is a bend or U-shaped curve in a rope.

(3) Hitch. A hitch is used to tie a rope around a timber, pipe, or post so that it will hold temporarily but can be readily untied.

(4) Knot. A knot is an interlacement of the parts of bodies, as cordage, forming a lump or knot or any tie or fastening formed with a cord, rope, or line, including bends, hitches, and splices. It is often used as a stopper to prevent a rope from passing through an opening.

(5) Line. A line (sometimes called a rope) is a single thread, string, or cord.

(6) Loop. A loop is a fold or doubling of the rope through which another rope can be passed. A temporary loop is made by a knot or a hitch. A permanent loop is made by a splice or some other permanent means.

(7) Overhand Turn or Loop. An overhand loop is made when the running end passes over the standing part.

(8) Rope. A rope (often called a line) is made of strands of fiber twisted or braided together.

(9) Round Turn. A round turn is the same as a turn, with running end leaving the circle in the same general direction as the standing part.

(10) Running End. The running end is the free or working end of a rope.

(11) Standing End. The standing end is the balance of the rope, excluding the running end.

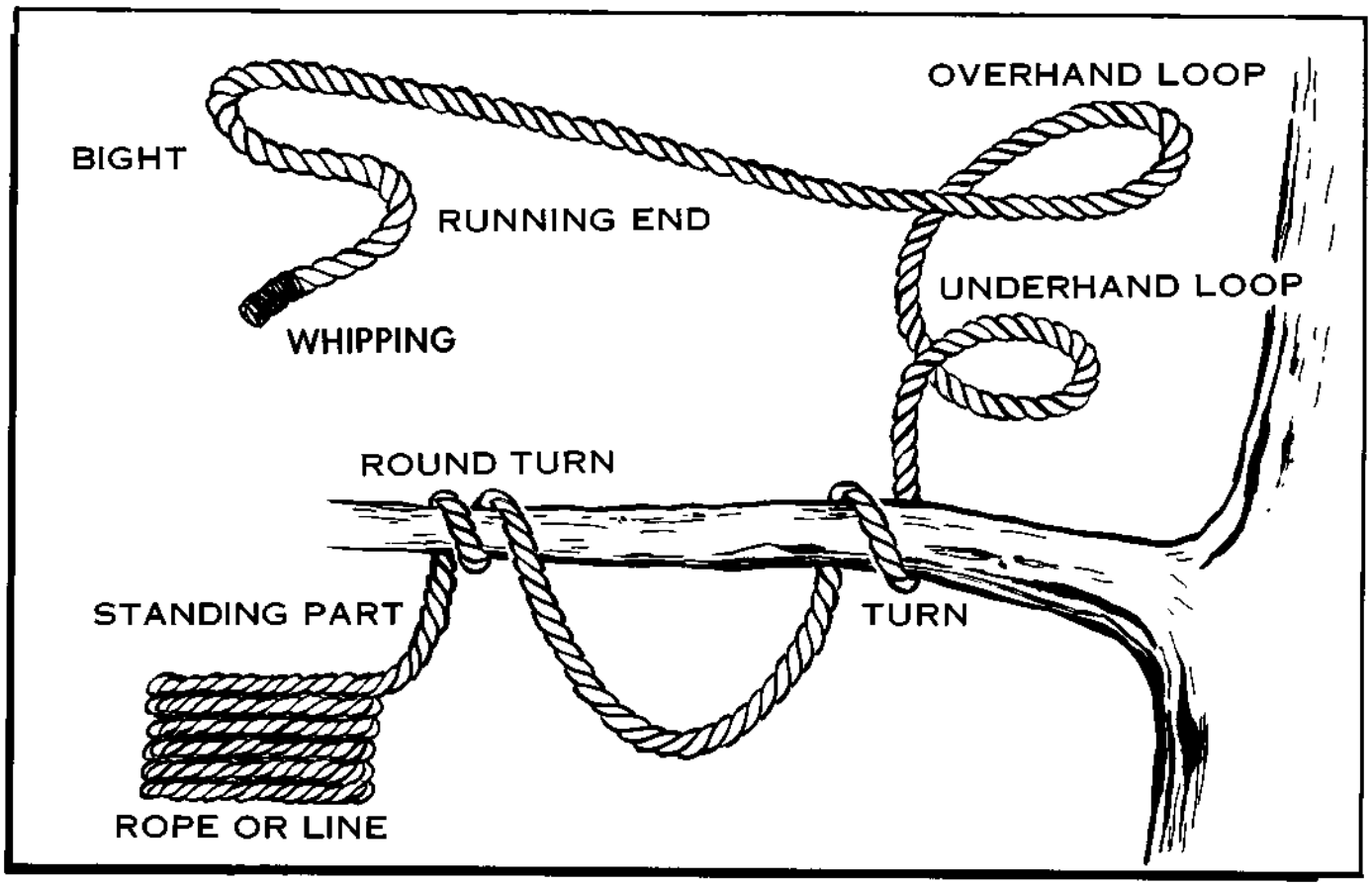

Figure 17-27. Elements of Ropes and Knots.

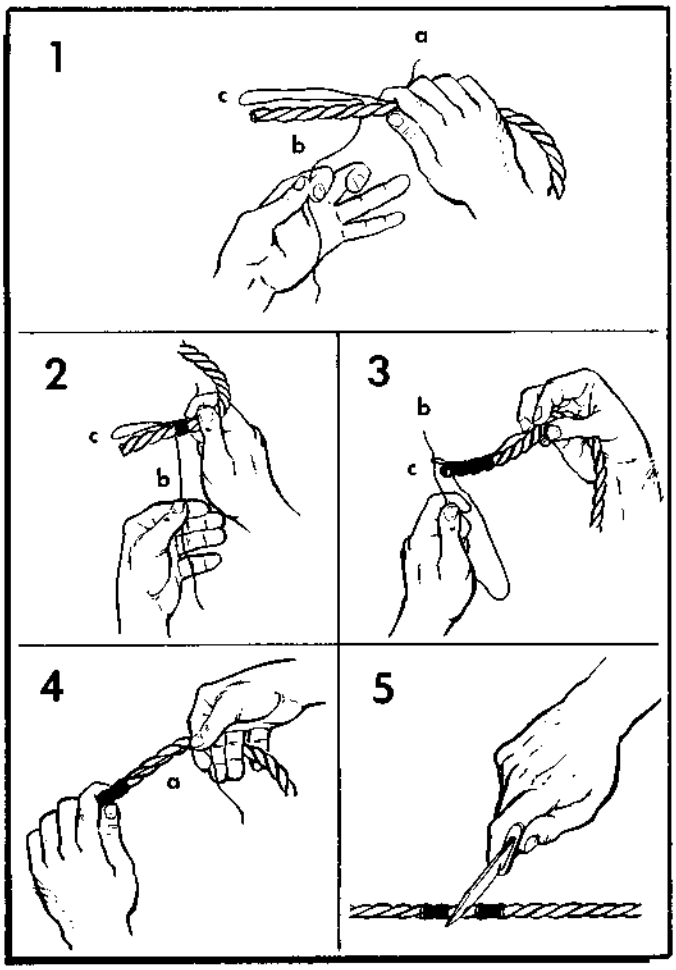

Figure 17-28. Whipping the End of a Rope.

(12) Turn. A turn describes the placing of a rope around a specific object such as a post, rail, or ring with the running end continuing in the opposite direction from the standing end.

(13) Underhand Turn or Loop. An underhand turn or loop is made when the running end passes under the standing part.

c. Whipping the Ends of a Rope. The raw, cut end of a rope has a tendency to untwist and should always be knotted or fastened in some manner. Whipping is one method of fastening the end of the rope. This method is particularly satisfactory because it does not increase the size of the rope. The whipped end of a rope will still thread through blocks or other openings. Before cutting a rope, place two whippings on the rope 1 or 2 inches apart and make the cut between the whippings (figure 17-28-5). This will prevent the cut ends from untwisting immediately after they are cut. A rope is whipped by wrapping the end tightly with a small cord. Make a bight near one end of the cord and lay both ends of the small cord along one side of the rope (figure 17-28-1). The bight should project beyond the end of the rope about one-half inch. The running end (b) of the cord should be wrapped tightly around the rope and cord (figure 17-28-2) starting at the end of the whipping which will be farthest from the end of the rope. The wrap should be in the same direction as the twist of the rope strands. Continue wrapping the cord around the rope, keeping it tight, to within about one-half inch of the end. At this point, slip the running end (b) through the bight of the cord (figure 17-28-3). The standing part of the cord (a) can then be pulled until the bight of the cord is pulled under the whipping and cord (b) is tightened (figure 17-28-4). The ends of cord (a and b) should be cut at the edge of the whipping, leaving the rope end whipped.

d. Knots at End of the Rope:

(1) Overhand Knot. The overhand knot (figure 17-29) is the most commonly used and the simplest of all knots. An overhand knot may be used to prevent the end of a rope from untwisting, to form a knot at the end of a rope, or as a part of another knot. To tie an overhand knot, make a loop near the end of the rope and pass the running end through the loop, pulling it tight.

(2) Figure-Eight Knot. The figure-eight knot (figure 17-30) is used to form a larger knot than would be formed by an overhand knot at the end of a rope. A figure-eight knot is used in the end of a rope to prevent the ends from slipping through a fastening or loop in another rope. To make the figure-eight knot, make a loop in the standing part, pass the running end around

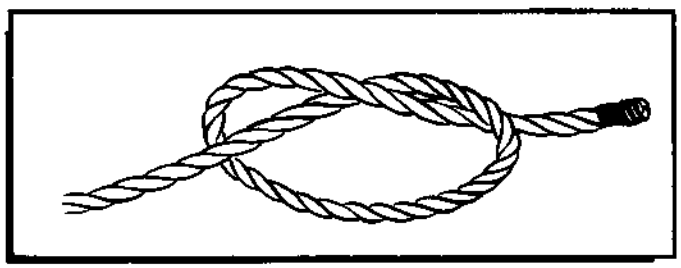

Figure 17-29. Overhand Knot.

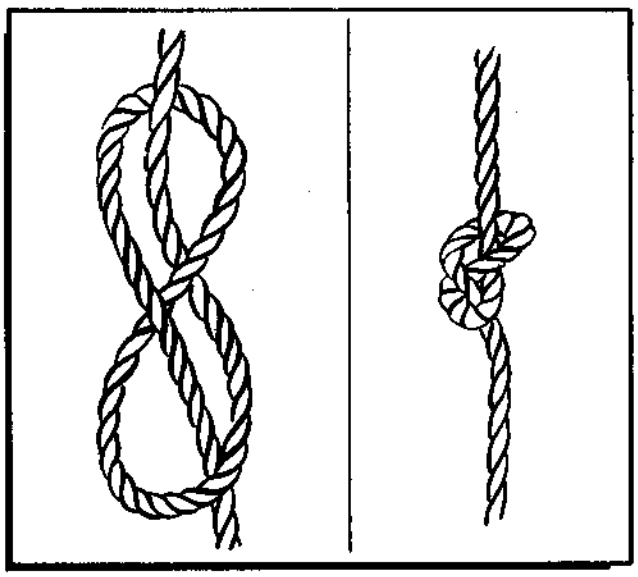

Figure 17-30. Figure-Eight Knot.

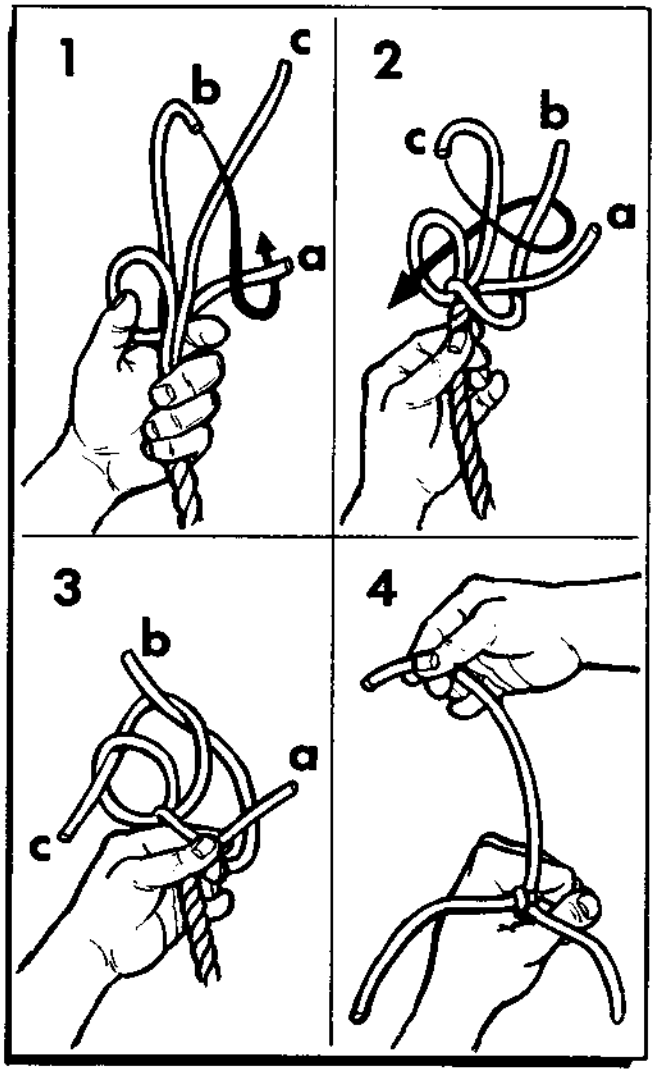

Figure 17-31. Wall Knot.

the standing part back over one side of the loop, and down through the loop. The running end can then be pulled tight.

(3) Wall Knot. The wall knot (figure 17-31) with a crown is used to prevent the end of a rope from untwisting when an enlargement is not objectionable. It also makes a desirable knot to prevent the end of the rope from slipping through small openings, as when rope handles are used on boxes. The crown or the wall knots may be used separately. To make the wall knot, untwist the strands for about five turns of the rope. A loop in strand "a" (figure 17-32-1) should be used and strand "b" brought down (figure 17-31-2) and around strand "a." Strand "c" (figure 17-31-3) can then be brought around strand "b" and through the loop in strand "a." The knot can then be tightened (figure 17-31-4) by grasping the rope in one hand and pulling each strand tight. The strands point up or away from the rope. To make a neat, round knot, the wall knot should be crowned.

(4) Crown on Wall Knot. To crown a wall knot, the end of strand "a" (figure 17-32-1) should be moved between strands "b" and "c." Next strand "c" is passed (figure 17-32-2) between strand "b" and the loop in strand "a." Line "b" is then passed over line "a" and through the bight formed by line "c" (figure 17-32-3). The knots can then be drawn tight and the loose strands cut. When the crown is finished, strands should point down or back along the rope.

e. Knots for Joining Two Ropes:

(1) Square Knot. The square knot (figure 17-33) is used for tying two ropes of equal diameter together to prevent slippage. To tie the square knot, lay the running end of each rope together but pointing in opposite directions. The running end of one rope can be passed under

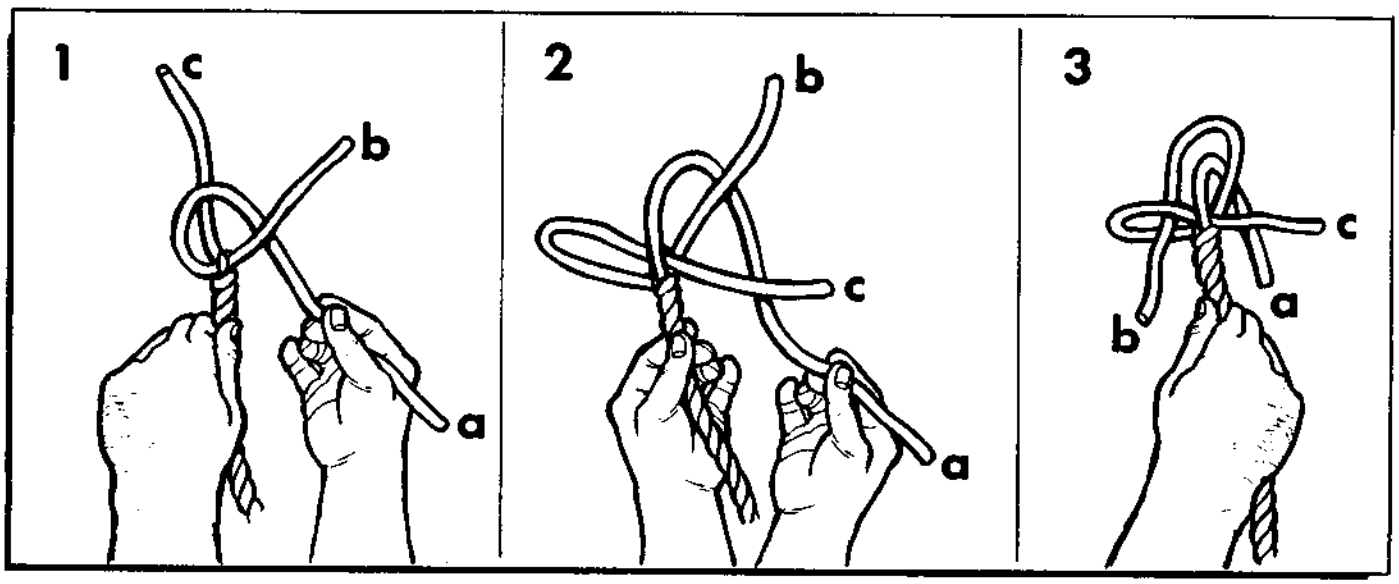

Figure 17-32. Crown on Wall Knot.

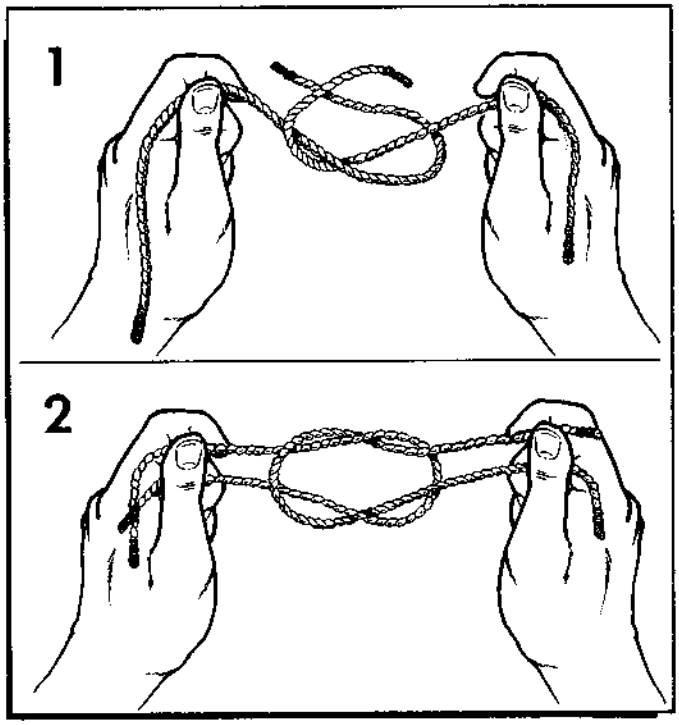

Figure 17-33. Square Knot.

the standing part of the other rope. Bring the two running ends up away from the point where they cross and crossed again (figure 17-33-1). Once each running end is parallel to its own standing part (figure 17-33-2), the two ends can be pulled tight. If each running end does not come parallel to the standing part of its own rope, the knot is called a "granny knot" (figure 17-34-1). Because it will slip under strain, the granny knot should not be used. A square knot can also be tied by making a bight in the end of one rope and feeding the running end of the other rope through and around this bight. The

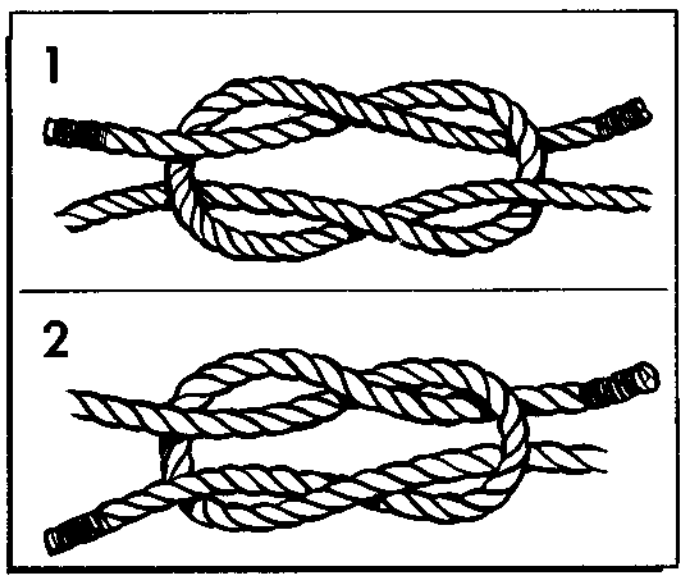

Figure 17-34. Granny and Thief Knots.

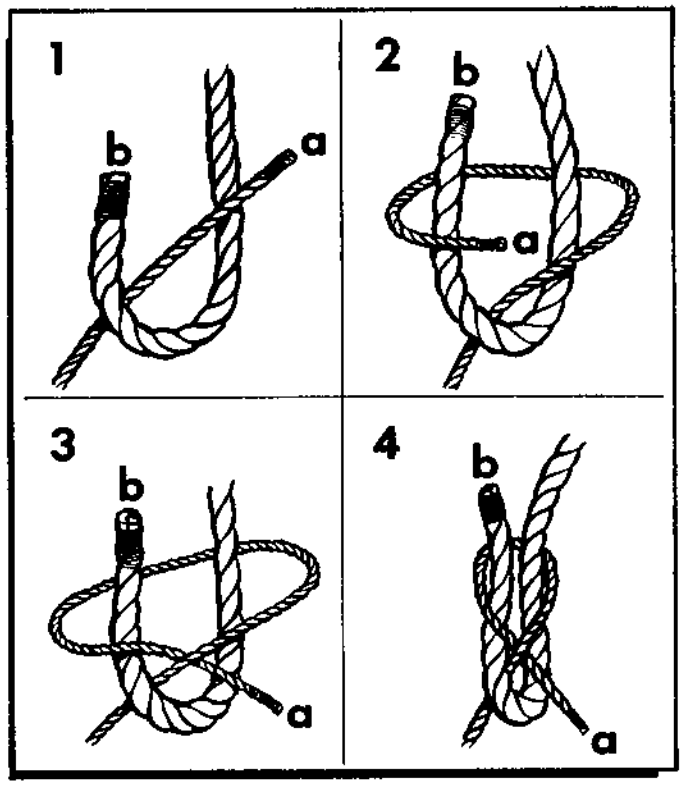

Figure 17-35. Single Sheet Bend.

running end of the second rope is routed from the standing side of the bight. If the procedure is reversed, the resulting knot will have a running end parallel to each standing part but the two running ends will not be opposite each other. This knot is called a "thief" knot (figure 17-34-2). It will slip under strain and is difficult to untie. A true square knot will draw tighter under

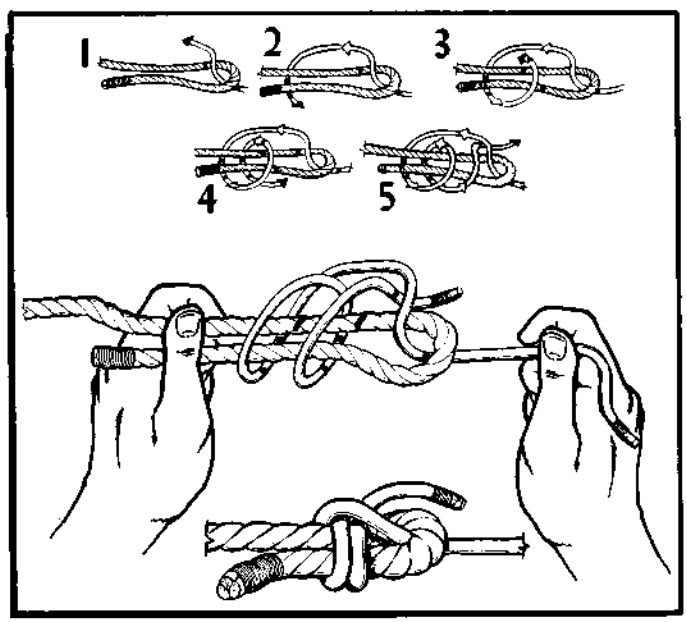

Figure 17-36. Double Sheet Bend.

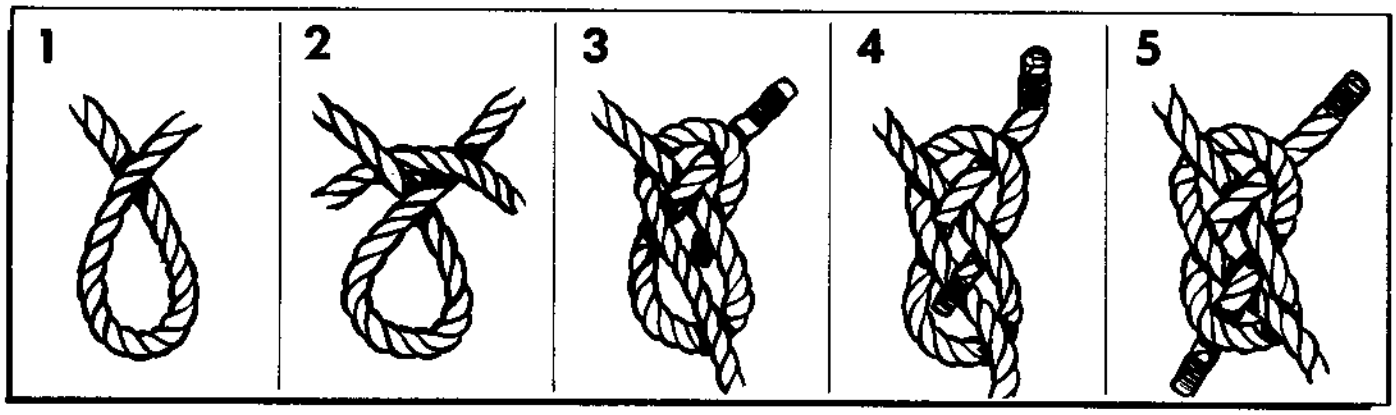

Figure 17-37. Carrick Bend.

strain. A square knot can be untied easily by grasping the bends of the two bights and pulling the knot apart.

(2) Single Sheet Bend. The use of a single sheet bend (figure 17-35), sometimes called a weaver's knot, is limited to tying together two dry ropes of unequal size. To tie the single sheet bend, the running end (a) (figure 17-35-1) of the smaller rope should pass through a bight (b) in the larger rope. The running end should continue around both parts of the larger rope (figure 17-35-2), and back under the smaller rope (figure 17-35-3). The running end can then be pulled tight (figure 17-35-4). This knot will draw tight under light loads but may loosen or slip when the tension is released.

(3) Double Sheet Bend. The double sheet bend (figure 17-36) works better than the single sheet bend for joining ropes of equal or unequal diameter, joining wet ropes, or for tying a rope to an eye. It will not slip or draw tight under heavy loads. To tie a double sheet bend, a single sheet bend is tied first. However, the running end is not pulled tight. One extra turn is taken around both sides of the bight in the larger rope with the running end for the smaller rope. Then tighten the knot.

(4) Carrick Bend. The carrick bend (figure 17-37) is used for heavy loads and for joining thin cable or heavy rope. It will not draw tight under a heavy load. To tie a carrick bend, a loop is formed (figure 17-37-1) in one rope. The running end of the other rope is passed behind the standing part (figure 17-37-2) and in front of the running part of the rope in which the loop has been formed. The running end should then be woven under one side of the loop (figure 17-37-3), through the loop, over the standing part of its own rope (figure 17-37-4), down through the loop, and under the remaining side of the loop (figure 17-37-5).

f. Knots for Making Loops:

(1) Bowline. The bowline (figure 17-38) is a useful knot for forming a loop in the end of a rope. It is also easy to untie. To tie the bowline, the running end (a) of the rope passes through the object to be affixed to the bowline and forms a loop (b) (figure 17-38-1) in the standing part of the rope. The running end (a) is then passed through the loop (figure 17-38-2) from underneath and around the standing part (figure 17-38-3) of the rope, and back through the loop from the top (figure 17-38-4). The running end passes down through the loop parallel to the rope coming up through the loop. The knot is then pulled tight.

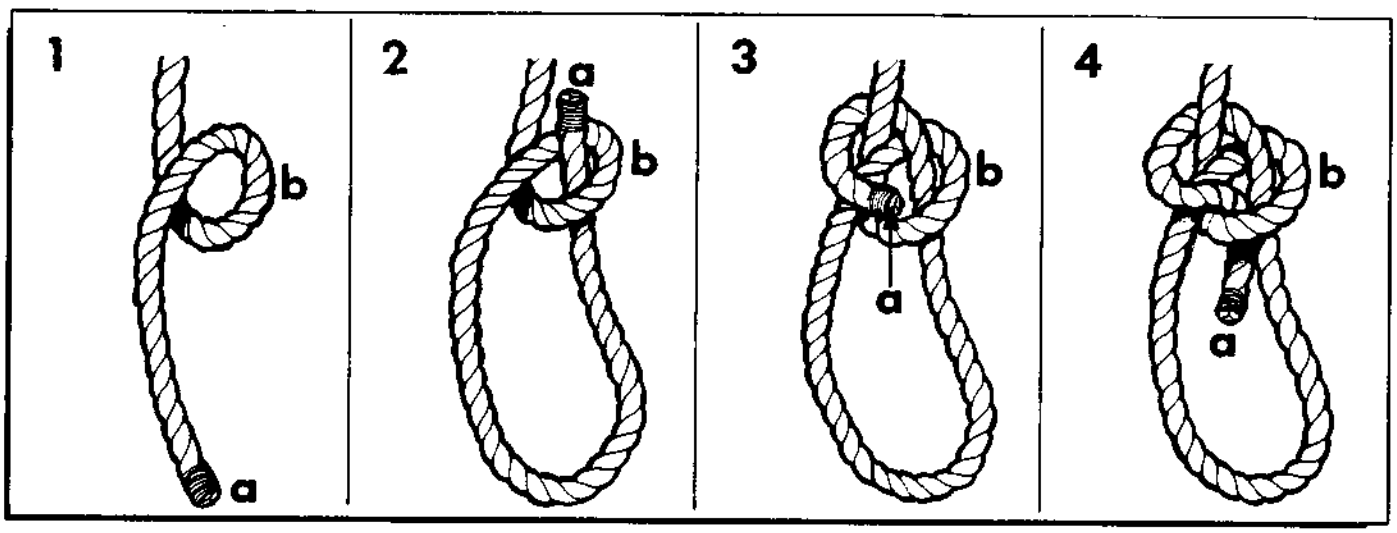

Figure 17-38. Bowline.

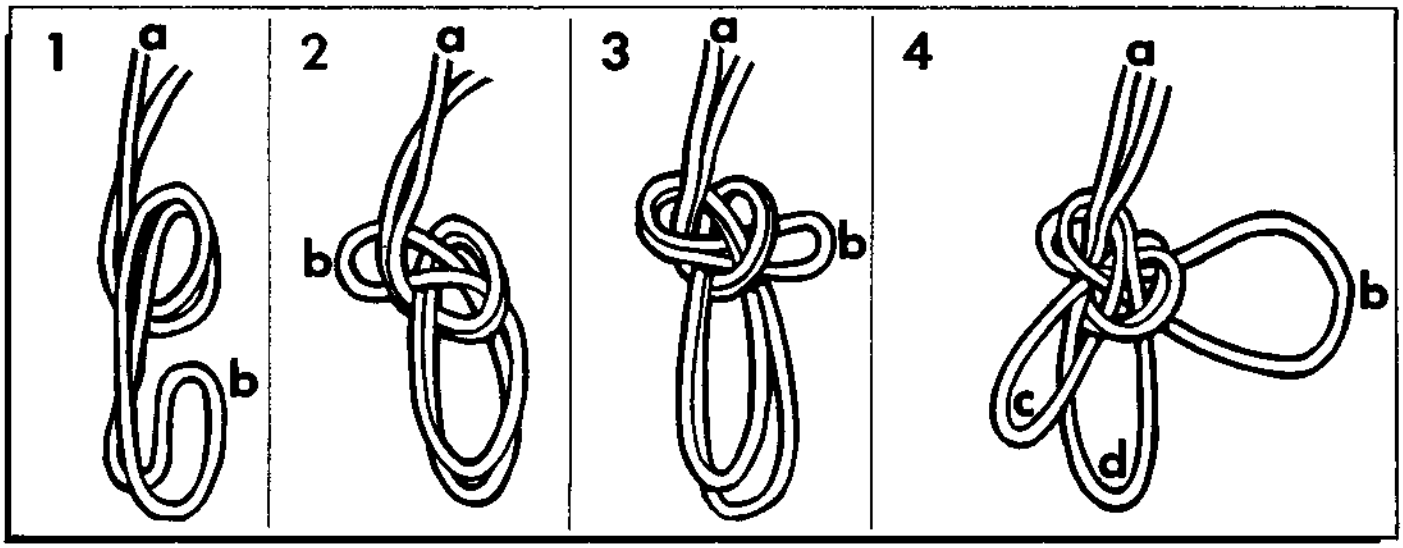

Figure 17-39. Double Bowline.

(2) Double Bowline. The double bowline (figure 17-39) with a slip knot is a rigging used by tree surgeons who work alone in trees for extended periods. It can be made and operated by one person and is comfortable as a sling or boatswain's chair (figure 17-40). A small board with notches as a seat adds to the personal comfort of the user. To tie a double bowline, the running end (a) (figure 17-39) of a line should be bent back about 10 feet along the standing part. The bight (b) is formed as the new running end and a bowline tied as described and illustrated in figure 17-38. The new running end (b) (figure 17-39) or loop is used to support the back and the remaining two loops (c) and (d) support the legs.

(3) Rolling or Magnus Hitch (figure 17-41). A rolling or Magnus hitch is a safety knot designed to make a running end fast to a suspension line with a nonslip grip yet it can be released by hand pressure bending the knot downward. The running end (a) (figure 17-41-1) is passed around the suspension line (b) twice, making two full turns downward (figure 17-41-2). The running end (a) (figure 17-41-3) is then turned upward over the two turns, again around the suspension line, and under itself (figure 17-41-3). This knot is excellent for fastening a rope to itself, a larger rope, a cable, a timber, or a post.

(4) Running Bowline. The running bowline (figure 17-42) is the basic air transport rigging knot. It provides a sling of the choker type at the end of a single line and is generally used in rigging. To tie a running bowline, make a bight (b) (figure 17-42-1) with an overhand loop (c) made in the running end (a). The running end (a) is passed around the standing part, through the loop (c) (figure 17-42-2), under, then back over the side of the bight, and back through the loop (c) (figure 17-42-3).

(5) Bowline on a Bight. It is sometimes desirable to form a loop at some point in a rope other than at the end. The bowline on a bight (figure 17-43) can be used

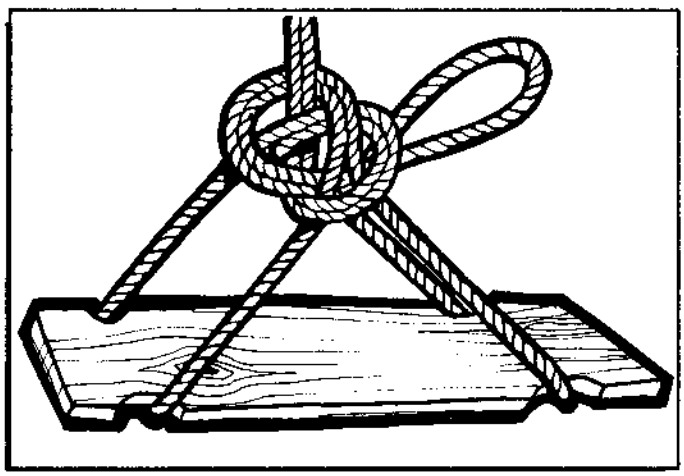

Figure 17-40. Boatswain's Chair.

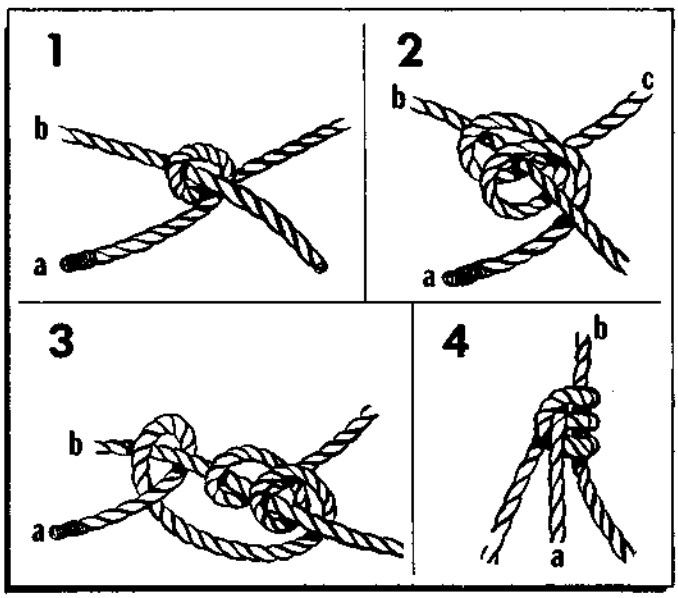

Figure 17-41. Rolling or Magnus Hitch.

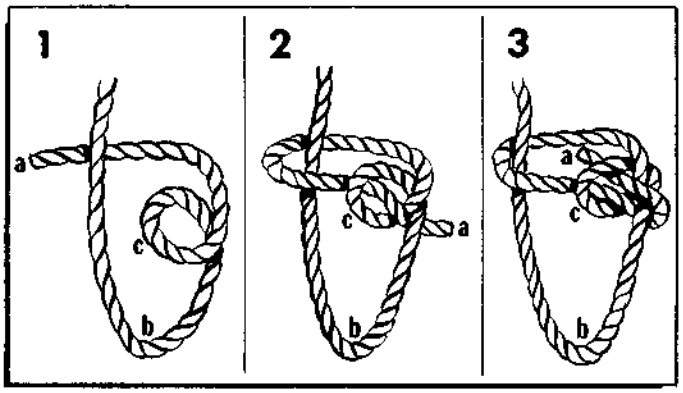

Figure 17-42. Running Bowline.

for this purpose. It is easily untied and will not slip. The same knot can be tied at the end of the rope by doubling the rope for a short section. A doubled portion of the rope is used to form a loop (b) (figure 17-43-1) as in the case of the bowline. The bight end (a) of the doubled portion is passed up through the loop (b), back down (figure 17-43-2), up around the entire knot (figure 17-43-3), and tightened (figure 17-43-4).

(6) Spanish Bowline. A Spanish bowline (figure 17-44) can be tied at any point in a rope, either at a place where the line is doubled or at an end which has been doubled back. The Spanish bowline is used in rescue work or to give a two-fold grip for lifting a pipe or other round object in a sling. To tie the Spanish bowline, a doubled portion of the rope is held in the left hand with the loop up and the center of the loop is turned back against the standing parts to form two loops (figure 17-44-1) or "rabbit ears." The two rabbit ears (c) and (d) (figure 17-44-2) are moved until they partly overlap each other. The top of the loop nearest the person is brought down toward the thumb of the left hand, being sure it is rolled over as it is brought down. The thumb is placed over this loop (figure 17-44-5) to hold it in position. The top of the remaining loop is grasped and brought down, rolling it over and placing it under the thumb. There are now four small loops, (c, d, e, and f) in the rope. The lower left-hand loop (c) is turned one-half turn and inserted from front to back of the upper left-hand loop (e). The lower right-hand loop (d) is turned (figure 17-44-4) and inserted through the upper right-hand loop (f). The two loops (c and d) which have been passed through are grasped and the rope pulled tight (figure 17-44-5).

(7) French Bowline. The French bowline (figure 17-45) is sometimes used as a sling for lifting injured

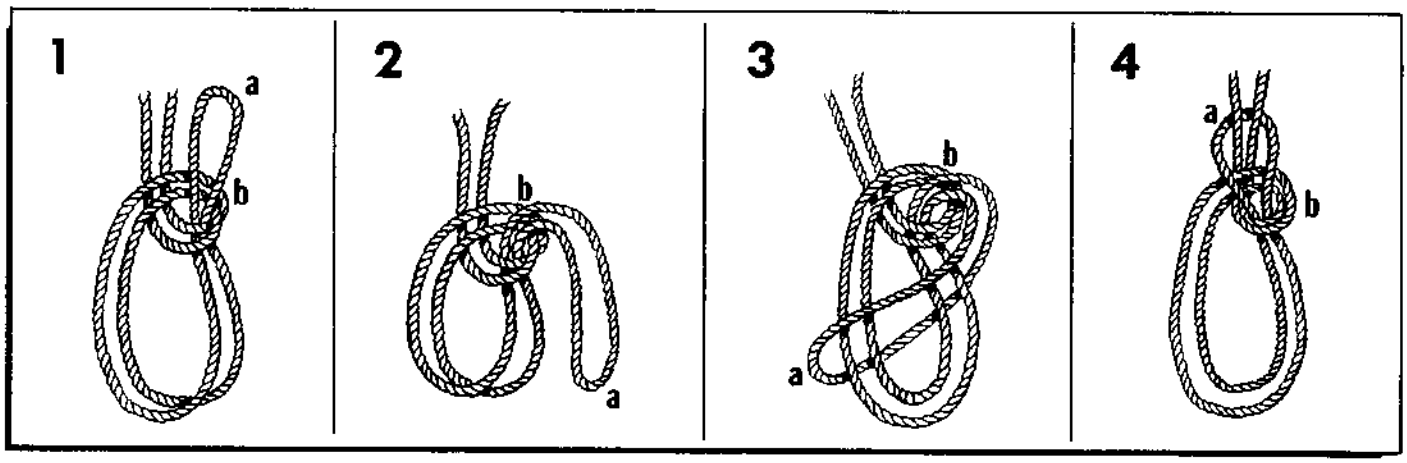

Figure 17-43. Bowline on a Bight.

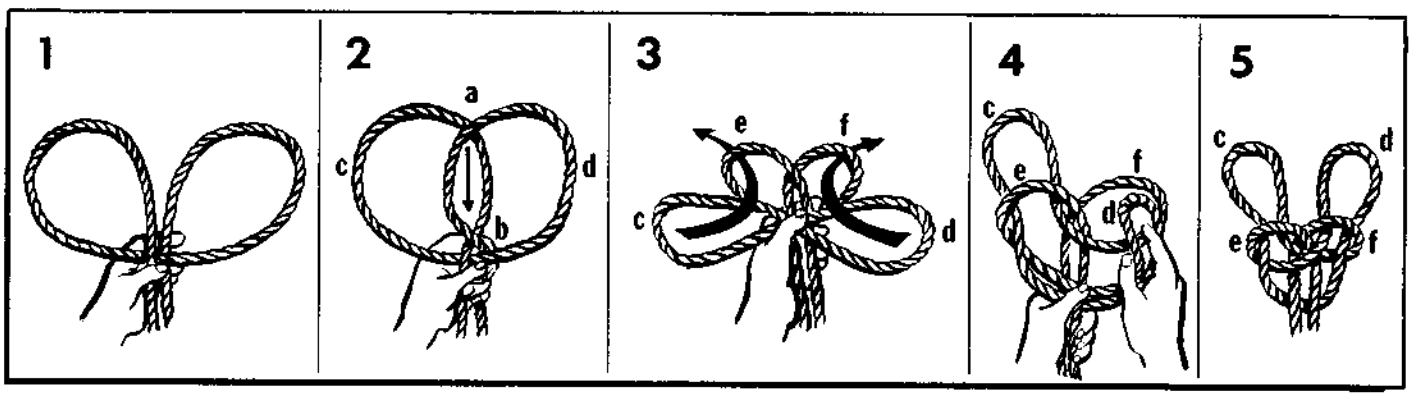

Figure 17-44. Spanish Bowline.

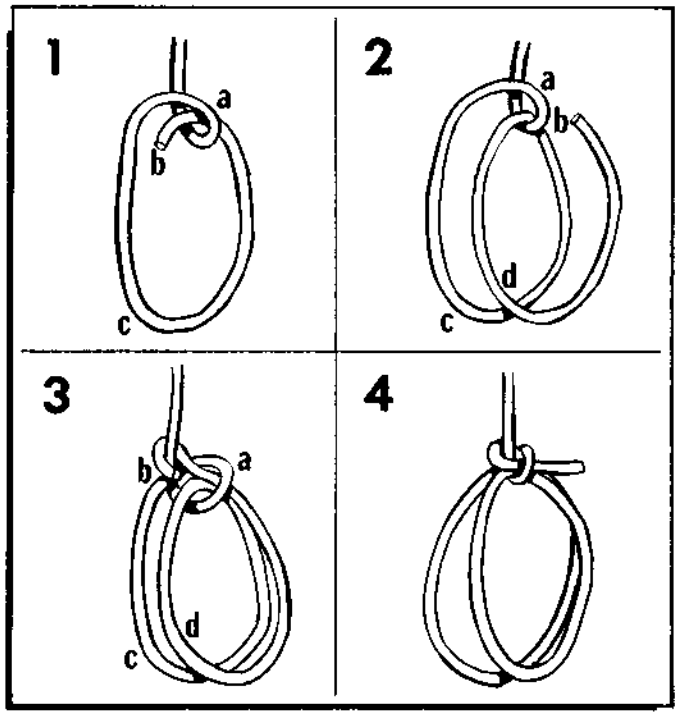

Figure 17-45. French Bowline.

people. When used in this manner, one loop is used as a seat and the other loop is used around the body under the arms. The weight of the injured person keeps the two loops tight so that the victim cannot fall out and for this reason, it is particularly useful as a sling for someone who is unconscious. The French bowline is started in the same way as the simple bowline. Make a loop (a) (figure 17-45-1) in the standing part of the rope. The running end (b) is passed through the loop from underneath and a separate loop (c) is made. The running end (b) is passed through the loop (a), again from underneath (figure 17-45-3), around the back of the standing part and back through the loop (a) so that it comes out parallel to the looped portion. The standing part of the rope is pulled to tighten the knot (figure 17-45-4), leaving two loops (c and d).

(8) Harness Hitch. The harness hitch (figure 17-46) is used to form a nonslipping loop in a rope. To make the harness hitch, form a bight (a) (figure 17-46-1) in the running end of the rope. Hold this bight in the left hand and form a second bight (b) in the standing part of the rope. The right hand is used to pass bight (b) over bight (a) (figure 17-46-2). Holding all loops in place with the left hand, the right hand is inserted through bight (a) behind the upper part of bight (b) (figure 17-46-3). The bottom (c) of the first loop is grasped and pulled up through the entire knot (figure 17-46-4), pulling it tight.

g. Hitches:

(1) Half Hitch. The half hitch (figure 17-47-1) is used to tie a rope to a timber or to another larger rope. It is not a very secure knot or hitch and is used for temporarily securing the free end of a rope. To tie a half hitch, the rope is passed around the timber, bringing the running end around the standing part, and back under itself.

(2) Timber Hitch. The timber hitch (figure 17-47-2) is used for moving heavy timbers or poles. To make the timber hitch, a half hitch is made and similarly the running end is turned about itself at least another time. These turns must be taken around the running end itself or the knot will not tighten against the pull.

(3) Timber Hitch and Half Hitch. To get a tighter hold on heavy poles for lifting or dragging a timber hitch and half hitch are combined (figure 17-47-3). The running end is passed around the timber and back under the standing part to form a half hitch. Further along the timber, a timber hitch is tied with the running end. The strain will come on the half hitch and the timber hitch will prevent the half hitch from slipping.

(4) Clove Hitch. A clove hitch (figure 17-47-4) is used to fasten a rope to a timber, pipe, or post. It can be tied at any point in a rope. To tie a clove hitch in the center of the rope, two turns are made in the rope close together. They are twisted so that the two loops lay back-to-back. These two loops are slipped over the timber or pipe to form the knot. To tie the clove hitch at the end of a rope, the rope is passed around the timber in two turns so that the first turn crosses the standing part and the running end comes up under itself on the second turn.

(5) Two Half Hitches. A quick method for tying a rope to a timber or pole is the use of two half hitches.

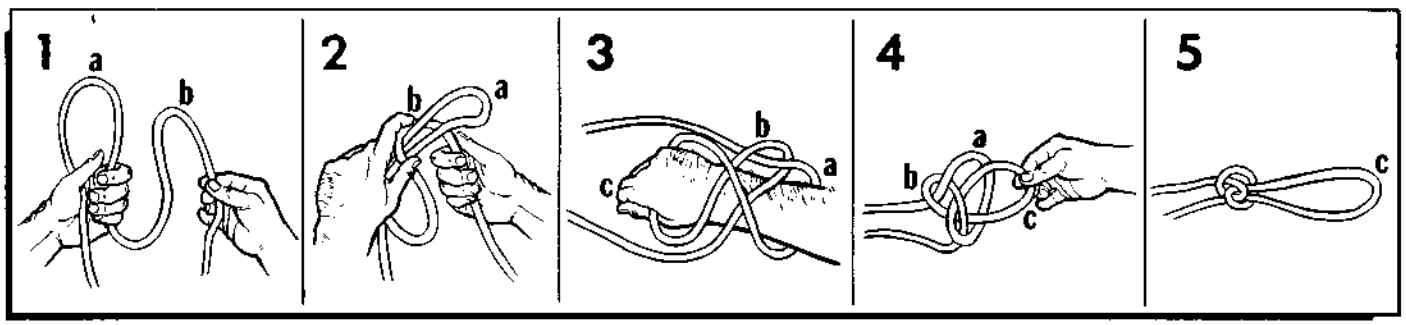

Figure 17-46. Harness Hitch.

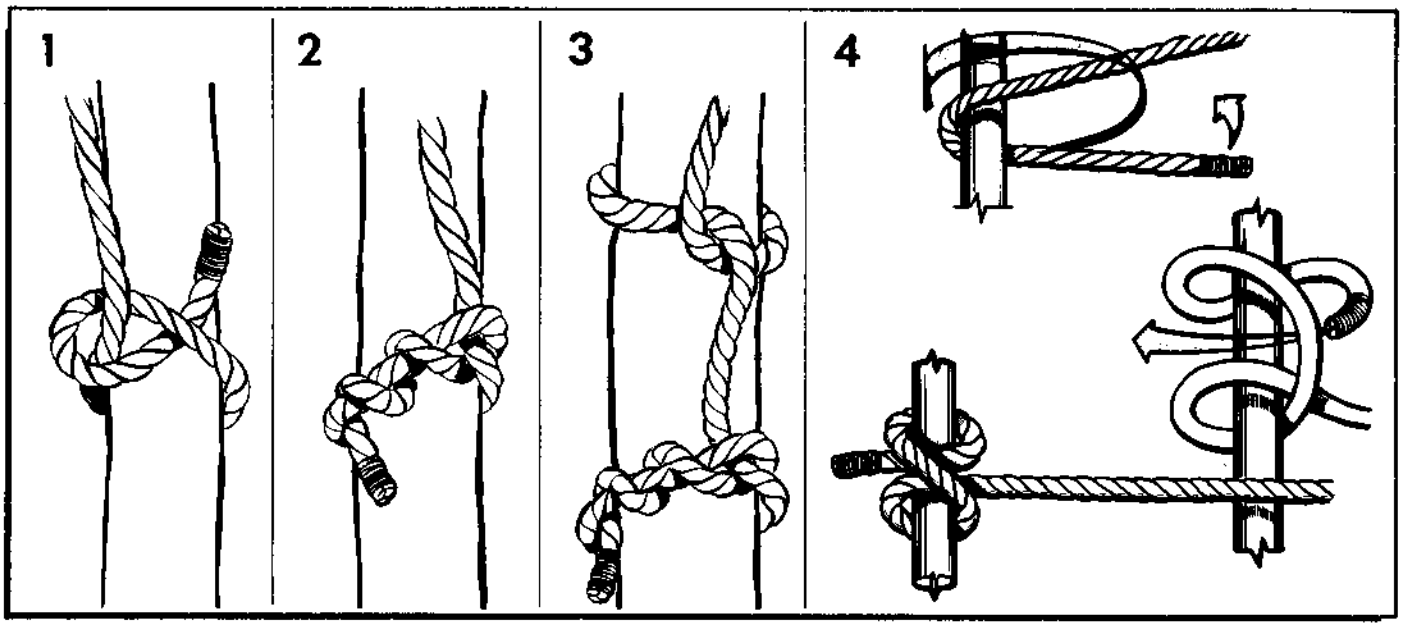

Figure 17-47. Half Hitch, Timber Hitch, and Clove Hitch.

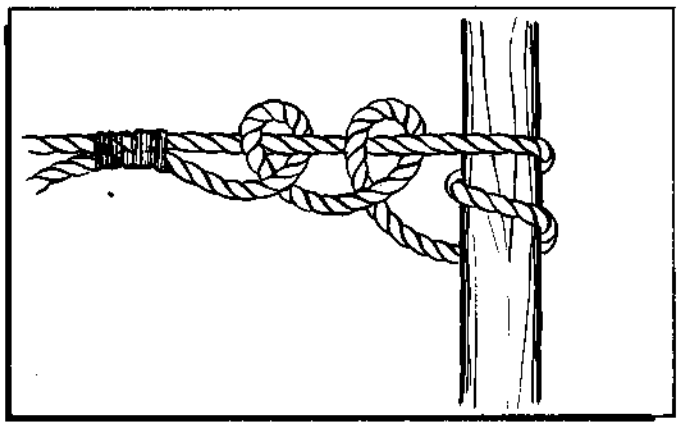

Figure 17-48. Round Turn and Two Half Hitches.

The running end of the rope is passed around the pole or timber, and a turn is taken around the standing part and under the running end. This is one half hitch. The running end is passed around the standing part of the rope and back under itself again.

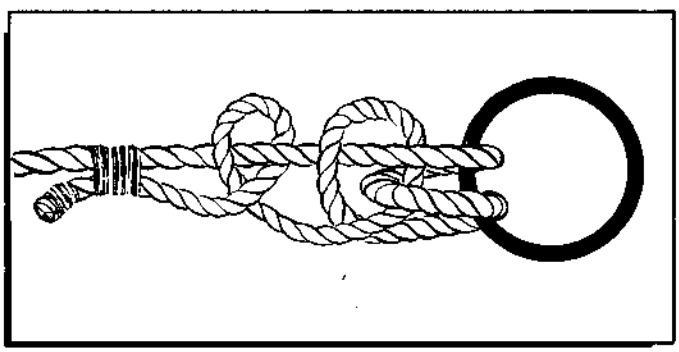

Figure 17-49. Fisherman's Bend.

(6) Round Turn and Two Half Hitches. Another hitch used for fastening a rope to a pole, timber, or spar is the round turn and two half hitches (figure 17-48). The running end of the rope is passed around the pole or spar in two complete turns, and the running end is brought around the standing part and back under itself to make a half hitch. A second half hitch is made. For greater security, the running end of the rope should be secured to the standing part.

(7) Fisherman's Bend. The fisherman's bend (figure 17-49) is used to fasten a cable or rope to an anchor, or for use where there will be a slackening and tightening motion in the rope. To make this bend, the running end of the rope is passed in two complete turns through the

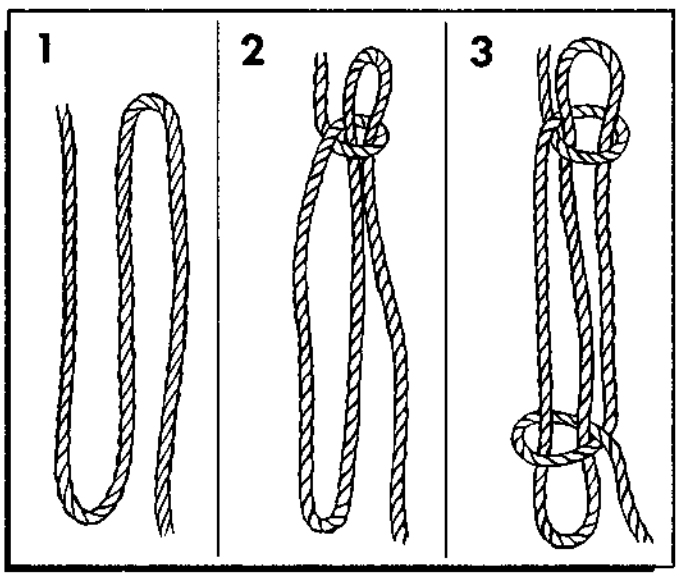

Figure 17-50. Sheep Shank.

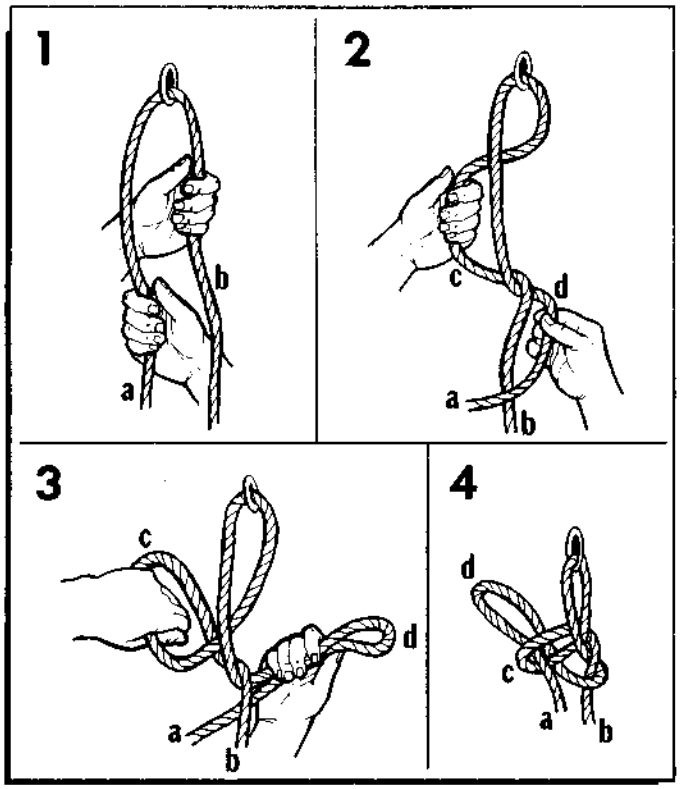

Figure 17-51. Speir Knot.

ring or object to which it is to be secured. The running end is passed around the standing part of the rope and through the loop which has just been formed around the ring. The running end is then passed around the standing part in a half hitch. The running end should be secured to the standing part.

(8) Sheepshank. A sheepshank (figure 17-50) is a method of shortening a rope, but it may also be used to take the load off a weak spot in the rope. To make the sheepshank (which is never made at the end of a rope), two bights are made in the rope so that three parts of the rope are parallel. A half hitch is made in the standing part over the end of the bight at each end.

(9) Speir Knot. A Speir knot (figure 17-51) is used when a fixed loop, a nonslip knot, and a quick release are required. It can be tied quickly and released by a pull on the running end. To tie the Speir knot, the running end (a) is passed through a ring (figure 17-51-1) or around a pipe or post and brought back on the left side of the standing part (b). Both hands are placed, palms up, under both parts of the rope with the left hand higher than the right hand; grasping the standing part (b) with the left hand and the running end (a) with the right hand. The left hand is moved to the left and the right hand to the right (figure 17-51-3) to form two bights (c and d). The left hand is twisted a half turn toward the body so that bight (c) is twisted into a loop (figure 17-51-3). Pass bight (d) over the rope and down through the loop (c). The Speir knot is tightened by pulling on the bight (d) and the standing part (b) (figure 17-51-4).

(10) Rolling Hitch (Pipe or Pole). The rolling hitch (pipe or pole) (figure 17-52) is used to secure a rope to a pipe or pole so that the rope will not slip. The standing part (a) of the rope is placed along the pipe or pole (figure 17-52-1) extending in the direction opposite to the direction the pipe or pole will be moved. Two turns (b) are taken with the running end around the standing part (a) and the pole (figure 17-52-3). The standing part (a) of the rope is reversed so that it is leading off in the direction in which the pole will be moved (figure 17-52-3) and two turns taken (c) (figure 17-52-4) with the running end (d). On the second turn around, the running end (d) is passed under the first turn (c) to secure it. To make this knot secure, a half hitch (e) (figure 17-52-6) is tied with the standing part of the rope 1 or 2 feet above the rolling hitch.

(11) Blackwall Hitch. The blackwall hitch (figure 17-53) is used for fastening a rope to a hook. To make the blackwall hitch, a bight of the rope is placed behind the hook. The running end (a) and standing part (b) are crossed through the hook so that the running end comes out at the opposite side of the hook and under the standing part.

(12) Catspaw. A catspaw can be made at the end of a rope (figure 17-54) for fastening the rope to a hook. Grasp the running end (a) of the rope in the left hand and make two bights (c and d) in the standing part (b). Hold these two bights in place with the left hand and take two turns about the junction of the two bights with the standing part of the rope. Slip the two loops (c and d) so formed over the hook.

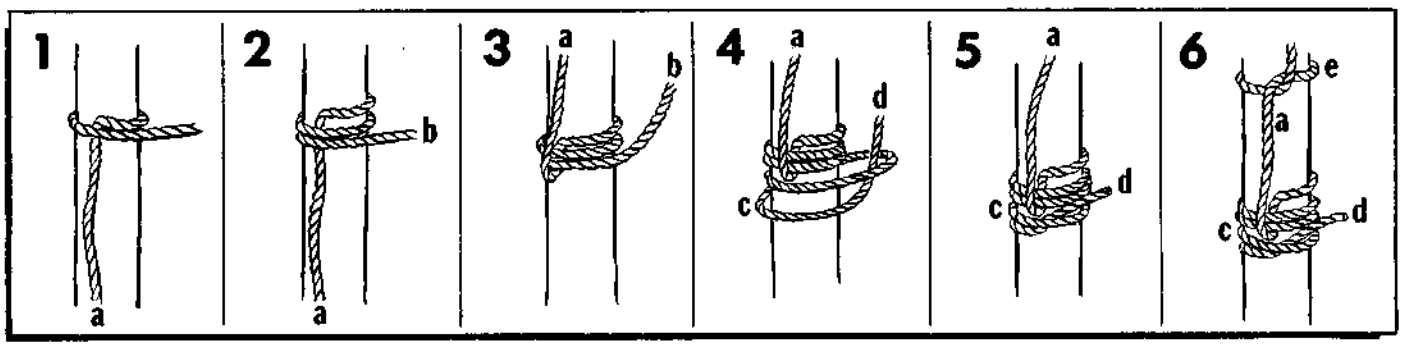

Figure 17-52. Rolling Hitch.

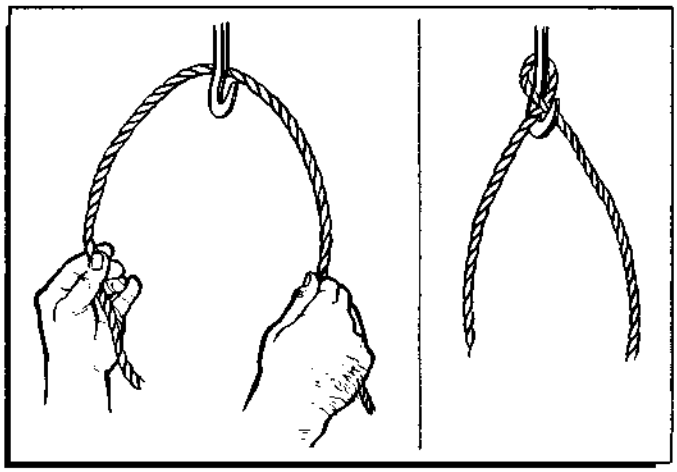

Figure 17-53. Blackwall Hitch.

(13) Scaffold Hitch. The scaffold hitch (figure 17-55) is used to support the end of a scaffold plank with a single rope. To make the scaffold hitch, the running end of the rope is layed across the top and around the plank, then up and over the standing part (figure 17-55-1). A doubled portion of the running end is brought back under the plank (figure 17-55-2) to form a bight (b) at the opposite side of the plank. The running end is taken back across the top of the plank (figure 17-55-3) until it can be passed through the bight (b). A loop is made (c) in the standing part (figure 17-55-4) above the plank. The running end is passed through the loop (c) around the standing part, and back through the loop (c).

(14) Barrel Slings. Barrel slings can be made to hold barrels horizontally or vertically. To sling a barrel horizontally (figure 17-56), a bowline is made with a long bight. The rope at the bottom of the bight is brought up over the sides of the bight. The two "ears" are thus moved foward over the end of the barrel. To sling a barrel vertically (figure 17-57) the rope is passed under the barrel and up to the top. An overhand knot is made (a) on top (figure 17-57-1). With a slight tension on the rope, the two parts (figure 17-57-2) of the overhand knot are grasped, separated and pulled down to the center of the barrel (b and c). The rope is pulled snug and a bowline tied (d) over the top of the barrel (figure 17-57-3).

h. Lashing. There are numerous items which require lashings for construction; for example, shelters, equipment racks, and smoke generators. Three types of lash-

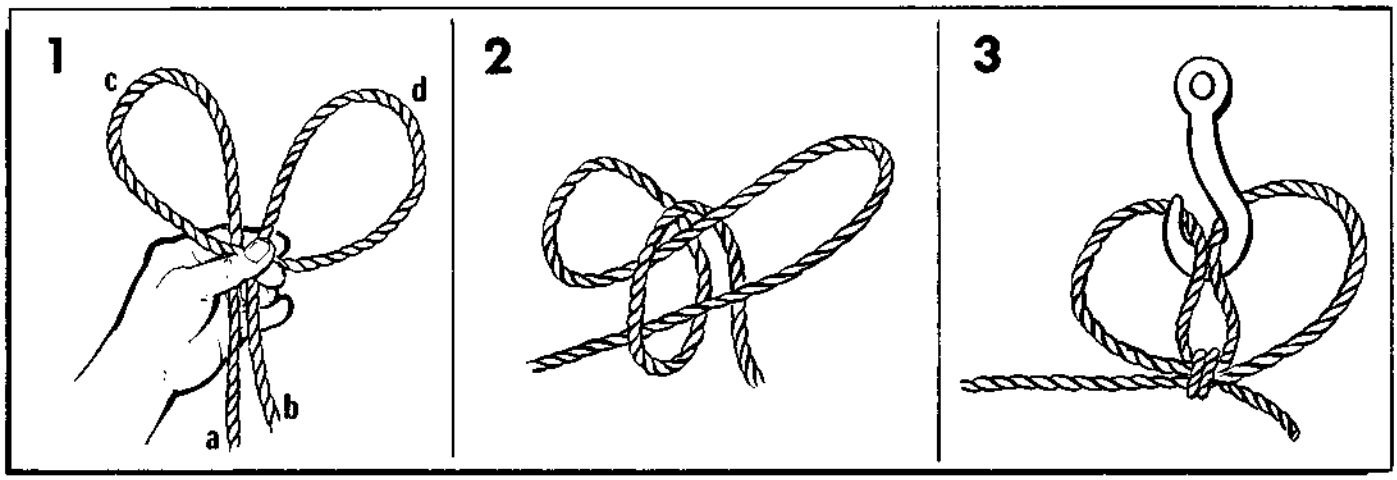

Figure 17-54. Catspaw.

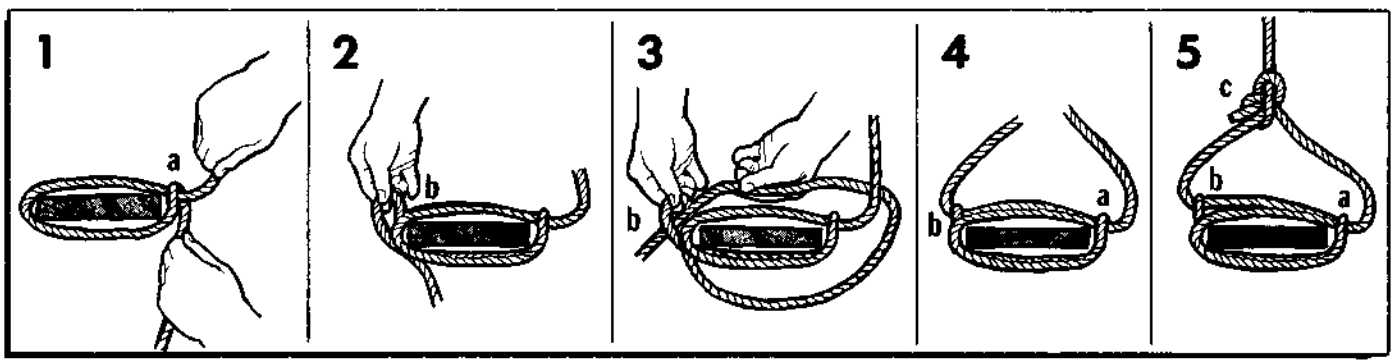

Figure 17-55. Scaffold Hitch.

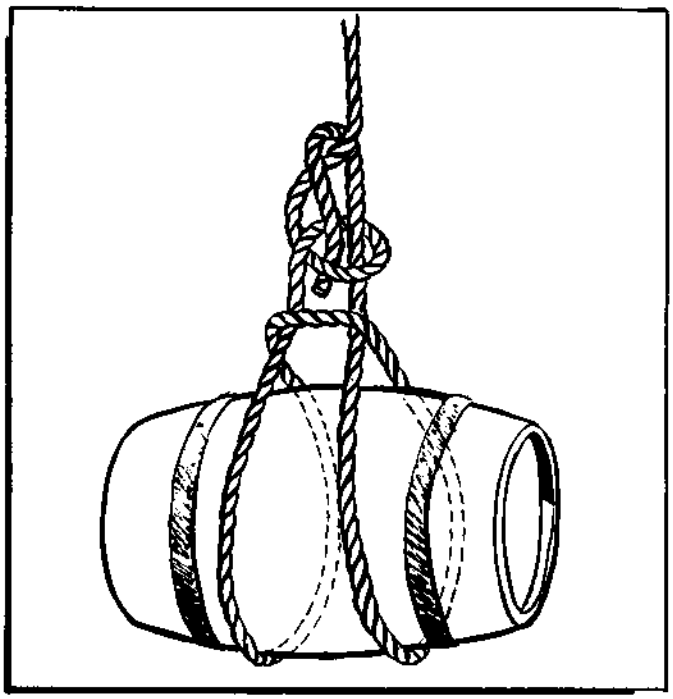

Figure 17-56. Barrel Slung Horizontally.

ings will be discussed here—the square lash, the diagonal lash, and the shear lash.

(1) Square Lash. Square lashing is started with a clove hitch around the log, immediately under the place where the crosspiece is to be located (figure 17-58-1). In laying the turns, the rope goes on the outside of the previous turn around the crosspiece, and on the inside of the previous turn around the log. The rope should be kept tight (figure 17-58-2). Three or four turns are necessary. Two or three "frapping" turns are made between the crosspieces (figure 17-58-3). The rope is pulled tight; this will bind the crosspiece tightly together. It is finished with a clove hitch around the same piece that the lashing was started on (figure 17-58-4). The square lash is used to secure one pole at right angles to another pole. Another lash that can be used for the same purpose is the diagonal lash.

(2) Diagonal Lash. The diagonal lash is started with a clove hitch around the two poles at the point of crossing. Three turns are taken around the two poles (figure 17-59-1). The turns lie beside each other, not on top of each other. Three more turns are made around the two poles, this time crosswise over the previous turns. The turns are pulled tight. A couple of frapping turns are made between the two poles, around the lashing turns, making sure they are tight (figure 17-59-2). The lashing is finished with a clove hitch around the same pole the lash was started on (figure 17-59-3).

(3) Shear Lash. The shear lash is used for lashing two or more poles in a series. The desired number of poles are placed parallel to each other and the lash is started with a clove hitch on an outer pole (figure 17-60-1). The poles are then lashed together, using seven or eight turns of the rope laid loosely beside each other (figure 17-60-2). Make frapping turns between each pole (figure 17-60-3). The lashing is finished with a clove hitch on the pole opposite that on which the lash was started (figure 17-60-4).

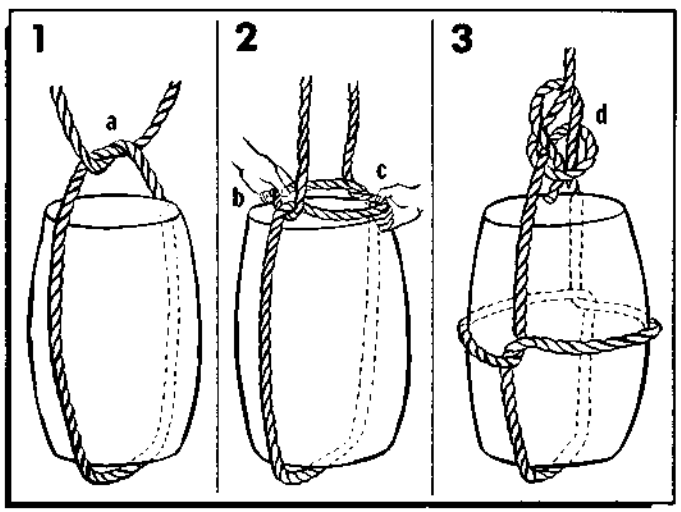

Figure 17-57. Barrel Slung Vertically.

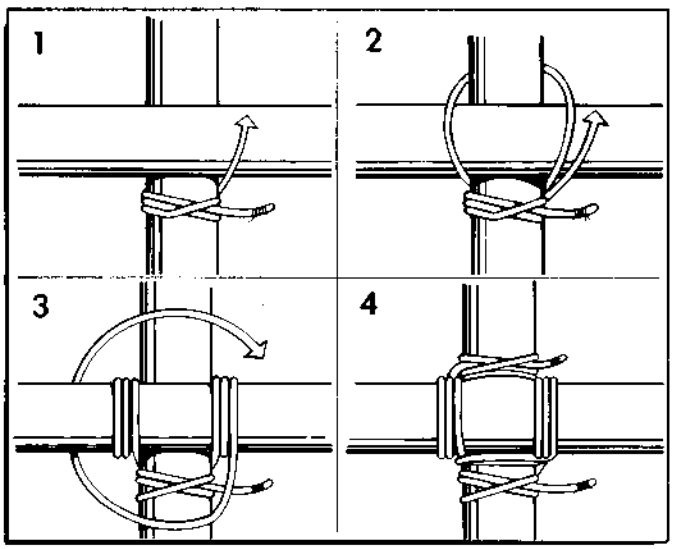

Figure 17-58. Square Lash.

i. Making Ropes and Cords. Almost any natural fibrous material can be spun into good serviceable rope or cord, and many materials which have a length of 12 to 24 inches or more can be braided. Ropes up to 3 and 4 inches in diameter can be "laid" by four people, and tensile strength for bush-made rope of 1-inch diameter range from 100 pounds to as high as 3,000 pounds.

(1) Tensile Strength. Using a three-lay rope of 1-inch diameter as standard, the following table of tensile strengths may serve to illustrate general strengths of various materials. For safety's sake, the lowest figure should always be regarded as the tensile strength.

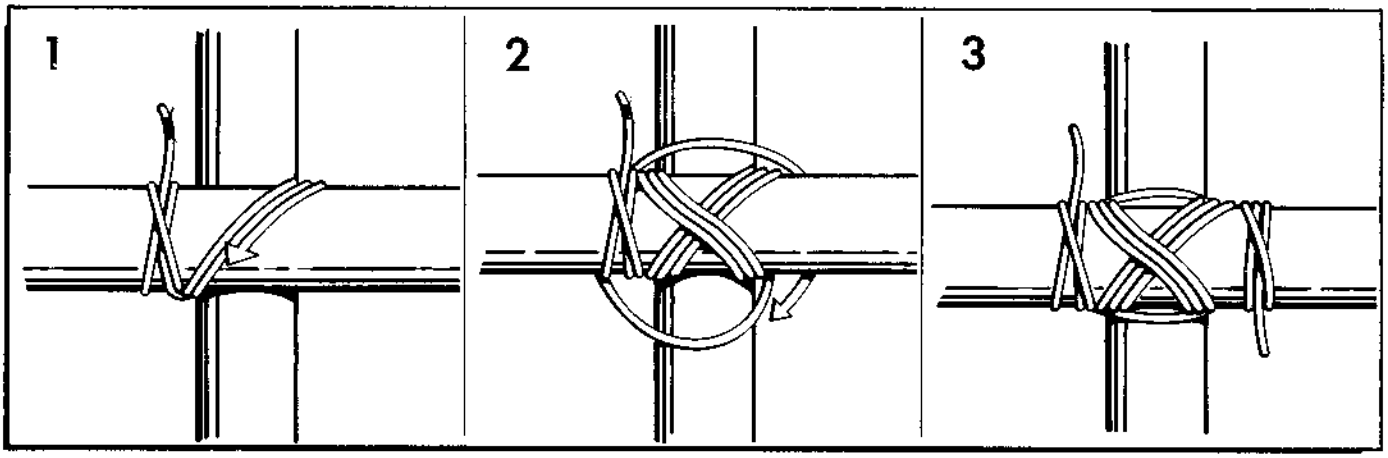

Figure 17-59. Diagonal Lash.

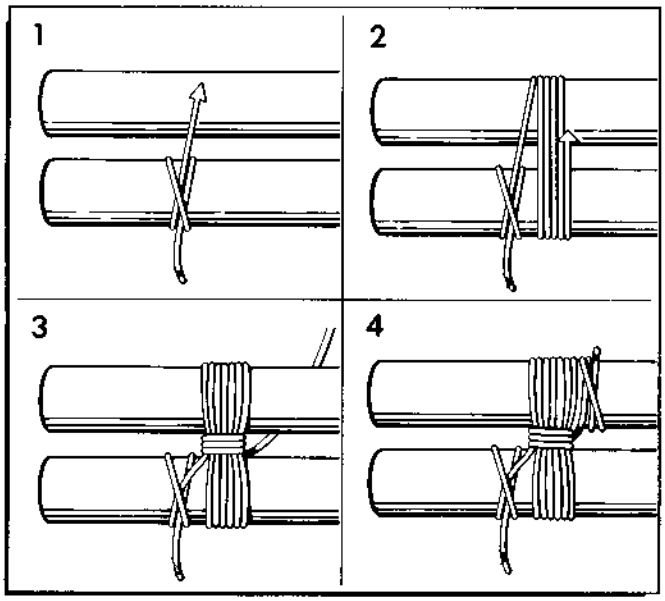

Figure 17-60. Shear Lash.

Green Grass	100 lbs to 250 lbs
Bark Fiber	500 lbs to 1,500 lbs
Palm Fiber	650 lbs to 2,000 lbs
Sedges	2,000 lbs to 2,500 lbs
Monkey Rope (Lianas)	560 lbs to 700 lbs
Lawyer Vine (Calamus)	½-inch diam, 1,200 lbs

NOTE: Doubling the diameter quadruples the tensile strength half the diameter reduces the tensile strength to one-fourth.

(2) Principles of Ropemaking Materials. To discover whether a material is suitable for rope making, it must have four qualities:

(a) It must be reasonably long in the fiber.

(b) It must have "strength."

(c) It must be pliable.

(d) It must have "grip" so the fibers will "bite" onto one another.

(3) Determining Suitability of Material. There are simple tests to determine if a material is suitable:

(a) First, pull on a length of the material to test for strength.

(b) Second, twist it between the fingers and "roll" the fibers together; if it will withstand this and not "snap" apart, an overhand knot is tied and gently tightened. If the material does not cut upon itself, but allows the knot to be pulled taut, it is suitable for ropemaking if the material will "bite" together and is not smooth or slippery.

(4) Where to Find Suitable Material. These qualities can be found in various types of plants, in ground vines, in most of the longer grasses, in some of the water reeds and rushes, in the inner barks of many trees and shrubs, and in the long hair or wool of many animals.

(5) Obtaining Fibers for Making Ropes. Some green freshly gathered materials may be "stiff" or unyielding. When this is the case, it should be passed through hot flames for a few moments. The heat treatment should cause the sap to burst through some of the cell structure, and the material thus becomes pliable. Fibers for rope making may be obtained from many sources such as:

(a) Surface roots of many shrubs and trees have strong fibrous bark.

(b) Dead inner bark of fallen branches of some species of trees and in the new growth of many trees such as willows.

(c) The fibrous material of many water and swamp growing plants and rushes.

(d) Many species of grass and weeds.

(e) Some seaweeds.

(f) Fibrous material from leaves, stalks, and trunks of many palms.

(g) Many fibrous-leaved plants such as the aloes.

(6) Gathering and Preparing Materials. There may be a high content of vegetable gum in some plants. This can often be removed by soaking the plants in water, by

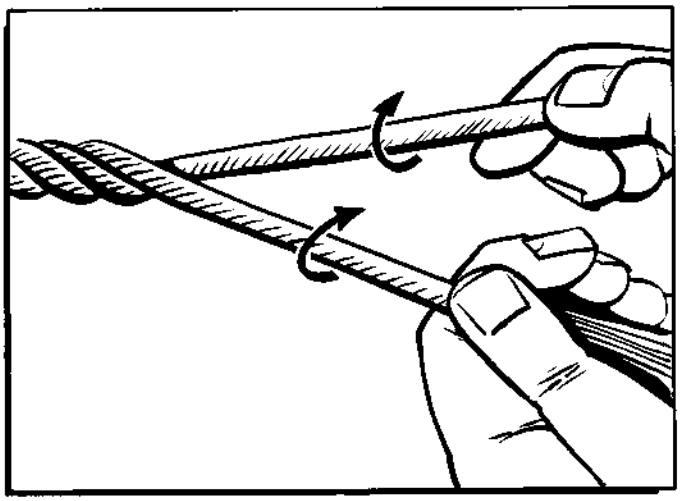

Figure 17-61. Twisting Fibers.

boiling, or by drying the material and "teasing" it into thin strips.

(a) Some of the materials have to be used green if any strength is required. The materials that should be green include the sedges, water rushes, grasses, and lianas.

(b) Palm fiber is harvested in tropical or subtropical regions. It is found at the junction of the leaf and the palm trunk, or it will be found lying on the ground beneath many palms. Palm fiber is a "natural" for making ropes and cords.

(c) Fibrous matter from the inner bark of trees and shrubs is generally more easily used if the plant is dead or half dead. Much of the natural gum will have dried out and when the material is being teased, prior to spinning, the gum or resin will fall out in fine powder.

(7) Making a Cord by Spinning with the Fingers:

(a) Use any material with long strong threads or fibers which have been previously tested for strength and pliability. The fibers are gathered into loosely held strands of even thickness. Each of these strands is twisted clockwise. The twist will hold the fibers together. The strands should be formed one-eighth inch diameter. As a general rule, there should be about 15 to 20 fibers to a strand. Two, three, or four of these strands are later twisted together, and this twisting together or "laying" is done with a counterclockwise twist, while at the same time, the separate strands which have not yet been laid up are twisted clockwise. Each strand must be of equal twist and thickness.

(b) Figure 17-61 shows the general direction of twist and the method whereby the fibers are bonded into strands. In a similar manner, the twisted strands are put together into lays, and the lays into ropes.

(c) The person who twists the strands together is called the "layer" and must see that the twisting is even, the strands are uniform, and the tension on each strand is equal. In "laying," care must be taken to ensure each of the strands is evenly "laid up;" that is, one strand does not twist around the other one.

(d) When spinning fine cords for fishing lines, snares, etc., considerable care must be taken to keep the strands uniform and the lay even. Fine thin cords of no more than 1/32-inch thickness can be spun with the fingers and are capable of taking a breaking strain of 20 to 30 pounds or more.

(e) Normally two or more people are required to spin and lay up the strands for cord. However, many native people spin cord unaided. They twist the material by running the flat of the hand along the thigh, with the fibrous material between hand and thigh; and with the free hand, they feed in fiber for the next "spin." Using this technique, one person can make long lengths of single strands. This method of making cord or rope with the fingers is slow if any considerable length of cord is required.

(f) An easier and simpler way to rapidly make lengths of rope from 50 to 100 yards or more in length is to make a rope machine and set up multiple spinners in the form of cranks. Figure 17-62 shows the details of rope spinning.

(g) To use a rope machine, each feeder holds the material under one arm and with one free hand feeds it into the strand which is being spun by the crank. The other hand lightly holds the fibers together till they are spun. As the lightly spun strands are increased in length, they must be supported on crossbars. They should not be allowed to lie on the ground. Spin strands from 20 to 100 yards before laying up. The material should not be spun in too thickly. Thick strands do not help strength in any way, rather, they tend to make a weaker rope.

(8) Setting Up a Rope Machine:

(a) When spinning ropes of 10 yards or longer, it is necessary to set crossbars every 2 or 3 yards to carry the strands as they are spun. If crossbars are not set up, the strands or rope will sag to the ground, and some of the fibers will tangle up with grass, twigs, or dirt on the ground. Also, the twisting of the free end may either be stopped or interrupted and the strand will be unevenly twisted.

(b) The easiest way to set up crossbars for the rope machine is to drive pairs of stakes into the ground about 6 feet apart and at intervals of about 6 to 10 feet. The crossbars must be smooth and free from twigs and loose portions of bark that might twist in with the spinning strands.

(c) The crossbar (a) is supported by two uprights and pierced to take the cranks (b). These cranks can be made out of natural sticks, morticed slab, and pegs, or if available, bent wire. The connecting rod (c) enables one person to turn all cranks clockwise simultaneously. Crossbars supporting the strands as they are spun are shown (d). A similar crank handle to the previous ones (b) is supported on a forked stick at the end of the rope

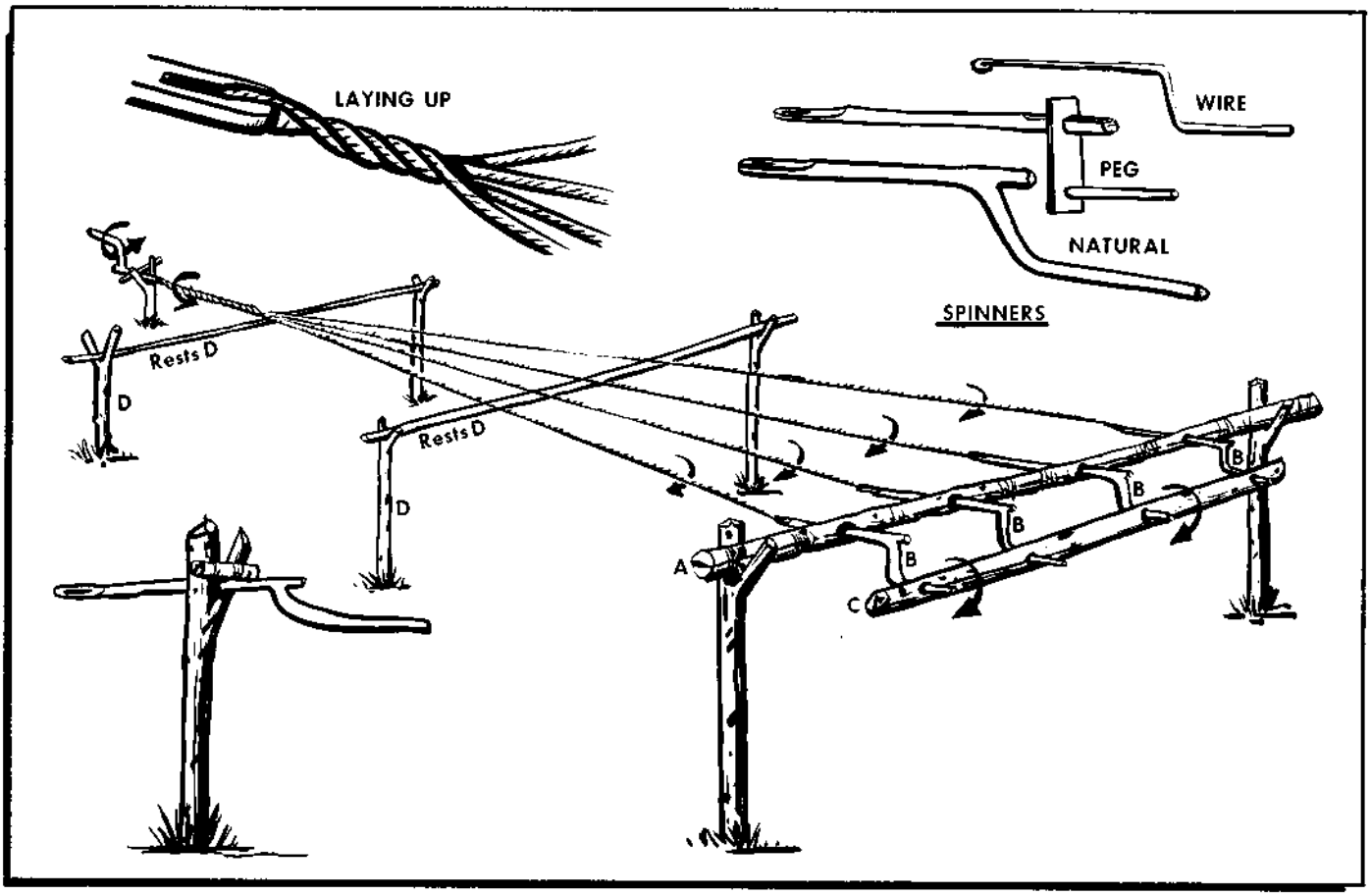

Figure 17-62. Rope Machine.

machine. This handle is turned in reverse (counterclockwise) to the cranks (c) to twist the connected strands together. These are "laid up" by one or more of the feeders.

(d) The first strand should be turned clockwise, then the laying up of the strands will be done counterclockwise and the next laying will again be clockwise. Proof that the rope is well made is that the individual fibers lay lengthways along the rope.

(e) In the process of laying up the strands, the actual twisting together or laying will take some of the original "twist" out of the strand which has not yet been laid. Therefore, it is necessary to keep twisting the strands while laying together.

(f) When making a rope too long to be spun and laid in one piece, a section is laid up and coiled on the ground at the end of the rope walk farthest from the cranks. Strands for a second length are spun, and these strands are married or spliced into the strands of the first section and then the laying up of the second section continues the rope.

(g) The actual "marrying" of the strands is done only in the last lay, which makes the rope when completed. The ends where the strands are married should be staggered in different places. By this means, rope can be made and extended in sections to a great length.

(h) After a complete length of rope is laid up, it should be passed through the fire to burn off the loose ends and fibers. This will make the rope smooth and more professional looking.

(9) Laying the Strands:

(a) The strands lie on the crossbar as they are spun. When the strands have been spun to the required length, which should not be more than about a hundred feet, they are joined together by being held at the far end. They are then ready for laying together. The turner, who is facing the cranks, twists the ends together counterclockwise, at the same time keeping full weight on the rope which is being layed up. The layer advances placing the strands side by side as they turn.

(b) It is important to learn to feed the material evenly, and lay up slowly, thereby getting a smooth even rope (figure 17-63). Do not try to rush the ropemaking. Speed in ropemaking only comes with practice. At first it will take a team of three or four up to 2 hours or more to make a 50-yard length of rope of three lays, each of three strands; that is, nine strands for a rope with a finished diameter of about 1 inch. With practice, the same three or four people will make the same rope in 15 to 20 minutes. These times do not include time for gathering material.

(c) In feeding the free ends of the strands, twist in the loose material fed in by the feeder. As the feeders move backward, they must keep a slight tension on the strands.

(10) Making Rope with a Single Spinner:

(a) Using a Single Crank. Two people can make a rope, using a single crank. A portion of the material is fastened to the eye of the crank, as with the multiple crank. Supporting crossbars, as used in a ropewalk, are required when a length of more than 20 or 30 feet is being spun.

(b) Feeding:

-1. If the feeder is holding material under the left arm, the right hand is engaged in continuously pulling material forward to the left hand which feeds it into the turning strand. These actions, done together as the feeder walks backward, govern the thickness of the strands. The left hand, lightly closed over the loose turning material, must "feel" the fibers "biting" or twisting together.

-2. When the free end of the turning strand, which is against the loose material under the arm, takes in too thick a tuft of material, the left hand is closed, and so arrests the twist of the material between the left hand and the bundle. This allows teasing out the overall "bite," with the right hand, thus maintaining a uniform thickness of the spinning strand.

(c) Thickness of Strands. Equal thickness and twist for each of the strands throughout their length are important. The thickness should not be greater than is necessary with the material being used. For a grass rope, the strand should not be more than one-fourth inch diameter; for coarse bark or palm, not more than one-eighth or three-sixteenth inch; and for fine bark, hair, or sisal fiber, not more than one-eighth inch.

(d) Common Errors in Ropemaking:

-1. There is a tendency with beginners to feed unevenly. Thin wispy sections of strand are followed by thick portions. Such feeding degrades the quality of rope. Rope made from such strands will break with less than one-fourth of the tensile strain on the material.

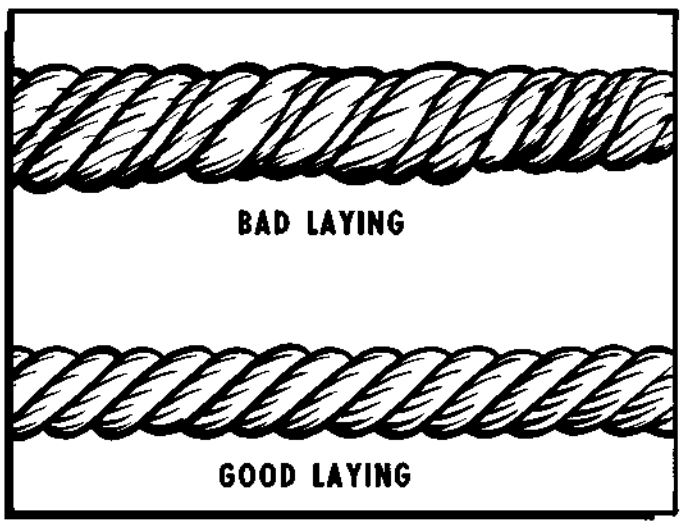

Figure 17-63. Rope Laying.

-2. Beginners are wise to twist and feed slowly. Speed, with uniformity of twist and thickness, comes with practice.

-3. Thick strands do not help. It is useless to try and spin a rope from strands an inch or more in thickness. Such a rope will break with less than half the tensile strain on the material. Spinning "thick" strands does not save time in ropemaking.

(e) Lianas, Vines, and Canes. Lianas and ground vines are natural ropes, and grow in subtropical and tropical scrub and jungle. Many are of great strength and useful for braiding, tree climbing, and other purposes. The smaller ground vines, when "braided", give great strength and flexibility. Canes and stalks of palms provide excellent material if used properly. Only the outer skin is tough and strong, and this skin will split off easily if the main stalk is bent away from the skin. This principle also applies to the splitting of lawyer cane (calamus), palm leaf stalks, and all green material. If the split starts to run off, bend the material away from the thin side, and it will gradually gain in size and come back to an even thickness with the other split side.

(f) Bark Fibers:

-1. The fibers in many barks which are suitable for "ropemaking" are located near the innermost layers. This is the bark next to the sap wood. When seeking suitable barks of green timber, cut a small section about 3 inches long and 1 inch wide. Cut this portion from the wood to the outer skin of the bark.

-2. The specimen should be peeled and the different layers tested. Green bark fibers are generally difficult to spin because of "gum" and it is better to search around for windfall dead branches and try the inner bark of these. The gum probably has leached out, and the fibers should separate easily.

-3. Many shrubs have excellent bark fiber, and here it is advisable to cut the end of a branch and peel off a strip of bark for testing. Thin bark from green shrubs is sometimes difficult to spin into fine cord and is easier to use as braid for small cords.

-4. Where it is necessary to use green bark fiber for rope spinning the gum will generally wash out when the bark is teased and soaked in water for a day or so. After removing from the water, the bark strips should be allowed to dry before shredding and teasing into fiber.

(11) Braiding. One person may require a length of rope. If there is no help available to spin materials, it is necessary to find reasonably long material. With this material, one person can braid and make suitable rope. The usual three-strand braid makes a flat rope, and while quite good, it does not have finish or shape, nor is it as "tight" as the four-strand braid. On other occa-

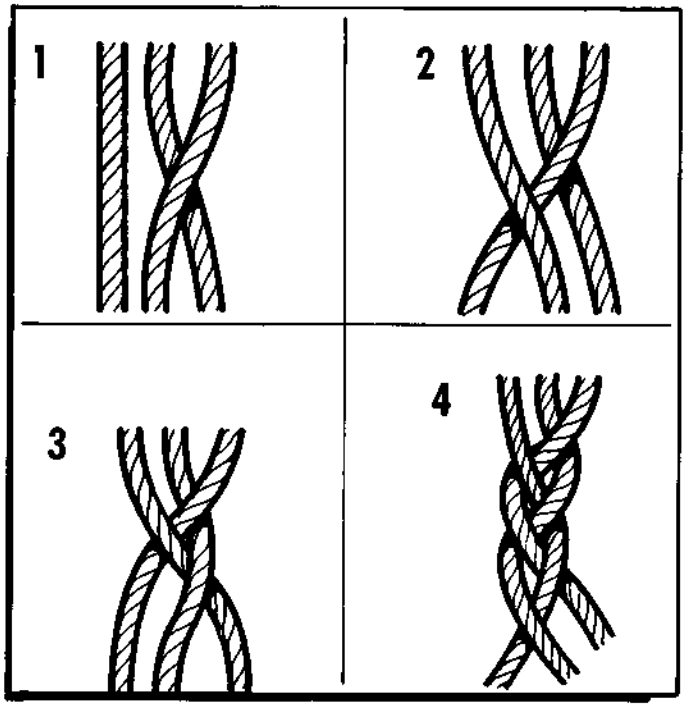

Figure 17-64. Three-Strand Braid.

sions, it may be necessary to braid broad bands for belts or for shoulder straps. There are many fancy braids which can be developed from these, but these three are basic, and essential for practical woodcraft work. A general rule for all braids is to work from the outside into the center.

(a) Three Plait:

-1. The right-hand strand is passed over the strand to the left.

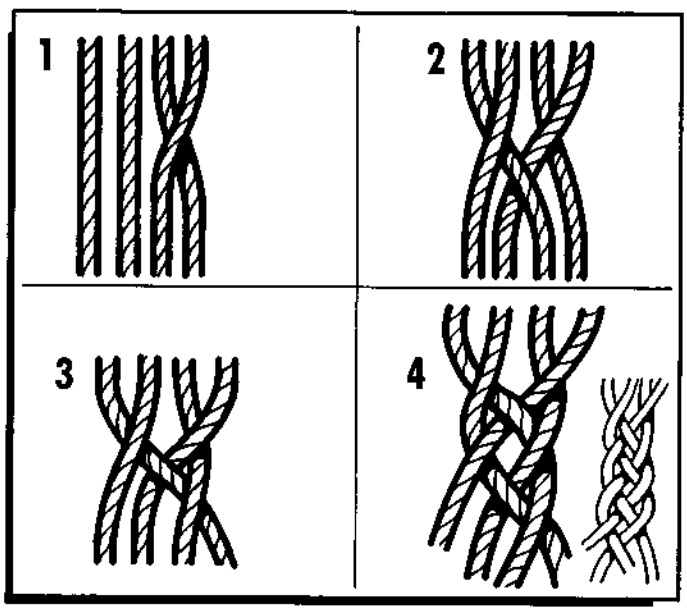

Figure 17-65. Four-Strand Braid.

-2. The left-hand strand is passed over the strand to the right.

-3. This is repeated alternately from left to right (figure 17-64).

(b) Flat Four-Strand Braid:

-1. The four strands are placed side by side. The right-hand strand is taken (figure 17-65-1) and placed over the strand to the left.

-2. The outside left-hand strand (figure 17-65-2) is laid under the next strand to itself and over what was the first strand.

-3. The outside right-hand strand is laid over the first strand to its left (figure 17-65-3).

-4. The outside left strand is placed under and over the next two strands, respectively, moving toward the right.

-5. Thereafter, the right-hand strand goes over one strand to the left, and the left-hand strand under and over to the right (figure 17-65-4).

(c) Broad Braid. Six or more strands are held flat and together.

-1. A strand in the center is passed over the next strand to the left, as in figure 17-66-1.

-2. The second strand to the left of center is passed toward the right and over the first strand so that it points toward the right (figure 17-66-2).

-3. The strand next to the first one is taken and woven under and over (figure 17-66-3).

-4. The next strands are woven from left and right alternately towards the center (figure 17-66-4 through 6). The finished braid should be tight and close (figure 17-66-7).

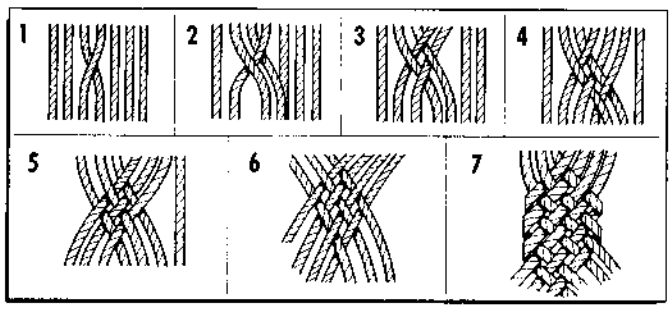

Figure 17-66. Broad Braid.

-5. To finish the broad braid:

-a. One of the center strands is laid back upon itself (figure 17-67-1).

-b. Now take the first strand which it enclosed in being folded back, and weave this back upon itself (figure 17-67-2).

-c. Strand from the opposite side is laid back and woven between the strands already braided (figure 17-67-3).

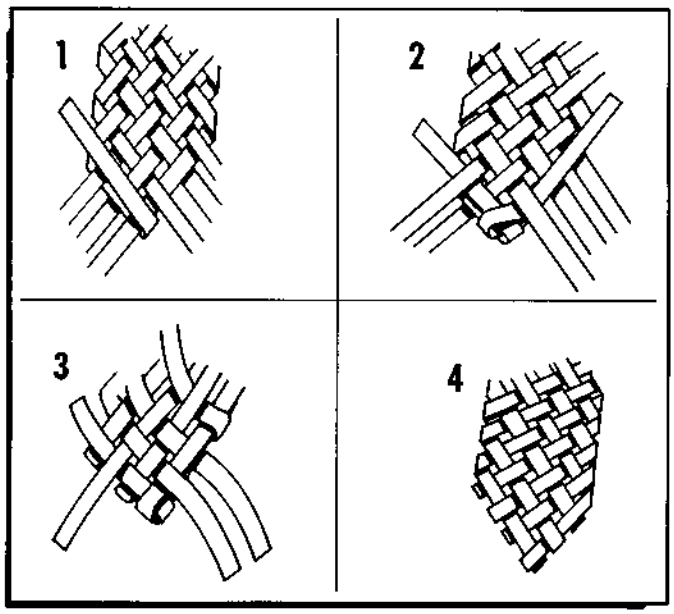

Figure 17-67. Finishing the Broad Braid.

-d. All the strands should be so woven back that no strands show an uneven pattern, and there should be a regular under-over-under of the alternating weaves (figure 17-67-4).

-e. If the braid is tight, there may be a difficulty in working the loose ends between the plaited strands.

-f. This can be done easily by sharpening a thin piece of wood to a chisel edge to open the strands sufficiently to allow the ends being finished to pass between the woven strands.

-g. It should be rolled under a bottle or other round object and made smooth for final finishing.

17-7. Personal Survival Kit:

a. Even though a survival kit may be available, aircrew members should consider assembling and carrying personal survival kits. Survival experiences have occurred where survivors hit the ground running, and because of shock and fear left their survival kits behind. If survivors have a personal survival kit in a pocket, it may improve their survival chances considerably.

b. A great deal of thought should go into preparing personal survival kits. The potential needs of the survivors must be a consideration, such as the impact of the environmental elements, type of mission to be flown (tactical or nontactical), availability of rescue, and how far to friendly forces (figure 17-68).

c. There are two basic ways to carry a personal survival kit. One way is to pack all items into one or two waterproof containers. The other way is to scatter the items throughout personal clothing. Any type of small container can be used to encase the contents of the personal survival kit. Plastic cigarette cases, soap dishes, and Band-aid boxes are excellent containers.

d. Examples of items which can be packed into a small container are:

(1) Matches.
(2) Safety pins (varied sizes).
(3) Fishhooks.
(4) Knife (small, multibladed).
(5) Button compass.
(6) Prophylactic (for water container).
(7) Bouillon cubes.
(8) Salt.
(9) Snare wire.
(10) Water purification tablets.
(11) Signal mirror.
(12) Needles.
(13) Band-aids.
(14) Aluminum Foil.
(15) Insect repellent stick.
(16) Chapstick.
(17) Soap (Antiseptic).

NOTE: All kits carried aboard the aircraft should be approved by the unit life support officer.

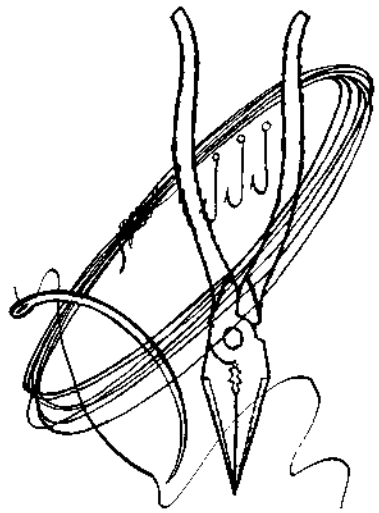

MINIMUM ESSENTIAL ITEMS

High quality pocket knife with at least two cutting blades.

Pocket compass.

Match safe with matches.

- Plastic or metallic container.
- Waterproof kitchen-type matches (cushion heads against friction), or
- Waterproof matches rolled in paraffin-soaked muslin in an easily opened container such as small soap box, toothbrush case, etc.

Needles — sailmakers, surgeons, and darning — at least one of each.

Assorted fishhooks in heavy foil, tin, or plastic holders.

Snare wire — small hank.

Needle-nosed pliers with side cutters; high quality.

Bar surgical soap or hand soap containing physohex.

Small fire starter of pyrophoric metal (some plastic match cases have a strip of the metal anchored on the bottom outside of the case).

Personal medicines.

Water purification tablets.

"Bandaids."

Insect repellent stick.

Chapstick.

GOOD TO HAVE ITEMS

*Pen-gun and flares.

*Colored cloth or scarf for signaling.

Stick-type skin dye (for camouflage).

Plastic water bottle.

*Flexible saw (wire saw).

*Sharpening stone.

Safety pins (several sizes).

Travel razor.

Small steel mirror.

6" flat bastard file.

Aluminum foil.

ADDITIONAL SUGGESTIONS

Toothbrush — small type.

Surgical tape.

Prophylactics (make good waterproof containers or canteens).

*Penlight with batteries.

Fishline.

*Fishline monofilament.

*Clear plastic bags.

Emergency ration can opener (can be taped shut and strung on dog tag chain).

Split shot — for fishing sinkers.

Gill net.

Small, high quality candles.

INDIVIDUAL MEDICAL KIT

Sterile gauze compress bandage.

Antibiotic ointment (Neomycin polymycin bacitracin opthalmic ointment is good).

Tincture of zephrine — skin antiseptic.

Aspirin tablets.

Salt tablets.

Additional medications may be desirable, depending upon nature of the mission and an individual's particular personal needs.

This should be discussed with and procured from your local flight surgeon.

*Especially valuable.

Figure 17-68. Personal Survival Kit Items.

Part Six

SUSTENANCE

Chapter 18

FOOD

18-1. Introduction. Except for the water they drink and the oxygen they breathe, survivors must meet their body needs through the intake of food. This chapter will explore the relationship of proper nutrition to physical and mental efficiency. It is extremely important that survivors maintain a proper diet at all times. A nutritionally sound body stands a much better chance of surviving. Improper diet over a long period of time may lead to a lack of stamina, slower reactions, less resistance to illness, and reduced mental alertness, all of which can cost survivors their lives in a survival situation. A knowledge of the body's nutritional requirements will help survivors select foods to supplement their rations.

18-2. Nutrition. Survivors and evaders expend much more energy in survival situations than they would in the course of their normal everyday jobs and life. Basal metabolism is the amount of energy expended by the body when it is in a resting state. The rate of basal metabolism will vary slightly with regard to the sex, age, weight, height, and race of a person. The basic energy expended, or number of calories consumed by the hour will change as a person's activity level changes. A person who is simply sitting in a warm shelter, for example, may consume anywhere from 20 to 100 calories an hour, while that same person evading through thick undergrowth with a heavy pack, would expand a greater amount of energy. In a survival situation, proper food can make the difference between success and failure.

a. The three major constituents of food are carbohydrates, fats, and proteins. Vitamins and minerals are also important as they keep certain essential body processes in good working order. It is also necessary for survivors to maintain proper water and salt levels in their bodies, as they aid in preventing certain heat disorders.

(1) Carbohydrates. Carbohydrates are composed of very simple molecules which are easily digested. Carbohydrates lose little of their energy to the process of digestion and are therefore efficient energy suppliers. Because carbohydrates supply easily used energy, many nutritionists recommend that, if possible, survivors should try to use them for up to half of their calorie intake. Examples of carbohydrates are: starches, sugars, and cellulose. These can be found in fruits, vegetables, candy, milk, cereals, legumes, and baked goods. Cellulose cannot be digested by humans, but it does provide needed roughage for the diet.

(2) Fats. Fats are more complex than carbohydrates. The energy contained in fats is more slowly released than the energy in carbohydrates. Because of this, it is a longer lasting form of energy. Fats supply certain fat-soluble vitamins. Sources of these fats and vitamins are butter, cheese, oils, nuts, egg yolks, margarine, and animal fats. If survivors eat fats before sleeping, they will sleep warmer. If fats aren't included in the diet of survivors, they can become run down and irritable. This can lead to both physical and psychological breakdown.

(3) Protein. The digestive process breaks protein down into various amino acids. These amino acids are formed into new body tissue protein, such as muscles. Some protein gives the body the exact amino acids required to rebuild itself. These proteins are referred to as "complete." Protein that lacks one or more of these essential amino acids is referred to as "incomplete." Incomplete protein examples are cheese, milk, cereal grains, and legumes. Incomplete protein, when eaten in combination with milk and beans for example, can supply an assortment of amino acids needed by the body. Some complete protein is found in fish, meat, poultry, and blood. No matter which type of protein is consumed, it will contain the most complex molecules of any food type listed.

(a) If possible, the recommended daily allowance of 2½ to 3 ounces complete protein should be consumed by each survivor each day. If only the incomplete protein is available, two, three, or even four types of foods may need to be eaten in combination so that enough amino acids are combined to form complete protein.

(b) If amino acids are introduced into the body in great numbers and some of them are not used for the rebuilding of muscle, they are changed into fuel or stored in the body as fat. Because protein contains the more complex molecules, over fats or carbohydrates, they supply energy after those forms of energy have been used up. A lack of protein causes malnutrition, skin and hair disorders, and muscle atrophy.

b. Vitamins occur in small quantities in many foods, and are essential for normal growth and health. Their chief function is to regulate the body processes. Vitamins can generally be placed into two groups: fat-soluble and water-soluble. The body only stores slight amounts of the water-soluble type. In a long survival episode where a routinely balanced diet is not available,

survivors must overcome food aversions and eat as much of a variety of vitamin-rich foods as possible. Often one or more of the four basic food groups (meat, fish, poultry; vegetables and fruits; grain and cereal; milk and milk products) are not available in the form of familiar foods, and vitamin deficiencies such as beri-beri or scurvy result. If the survivor can overcome aversions to local foods high in vitamins, these diseases as well as signs and symptoms such as depression and irritability can be warded off.

c. Adequate minerals can also be provided by a balanced diet. Minerals build and (or) repair the skeletal system and regulate normal body functions. Minerals needed by the body include iodine, calcium, iron, and salt, to name but a few. A lack of minerals can cause problems with muscle coordination, nerves, water retention, and the ability to form or maintain healthy red blood cells.

d. For survivors to maintain their efficiency, the following number of calories per day is recommended. These figures will change because of individual differences in basal metabolism, weight, etc. During warm weather survivors should consume anywhere from 3,000 to 5,000 calories per day. In cold weather the calorie intake should rise from 4,000 to 6,000 calories per day. A familiarity with the calorie and fat amounts in foods is important for survivors to meet their nutritional needs. For example, it would take quite a few mussels and dandelion greens to meet those requirements. Survivors should attempt to be familiar enough with foods that they can select or find foods that provide a high calorie intake (figure 18-1).

(1) Survivors should also be familiar with the number of calories supplied by the food in issued rations. In most situations, rations will have to be supplemented with other foods procured by survivors. If possible, survivors should limit their activities to save energy. Rationing food is a good idea since survivors never know when their ordeal will end. They should eat when they can, keeping in mind that they should maintain at least a minimum calorie intake to satisfy their basic activity needs.

(2) Caloric and fat values of selected foods are shown in the chart, and unless otherwise specified, the foods listed are raw. Depending on how survivors cook the food, the usable food value can be increased or decreased.

FOOD	CALORIES	FAT
WHOLE LARGE DUCK EGG	177	12.0
SMALL OR LARGE MOUTH BASS – 3 TO 4 OZ.	109	3.6
CLAMS – 4 TO 5 LARGE	88	.2
FRESHWATER CRAYFISH – 3 TO 4 OZ.	75	.6
EEL – 3 TO 5 OZ.	240	20.0
OCTOPUS – 3 TO 4 OZ.	76	.9
ATLANTIC SALMON – 4 OZ.	220	14.0
RAINBOW TROUT – 4 OZ.	200	11.8
BANANA – ONE SMALL	87	.3
BREADFRUIT – 3 TO 4 OZ.	105	.5
GUAVA – ONE MEDIUM	64	.7
MANGO – ONE SMALL	68	.5
WILD DUCK – 4 OZ.	230	16.0
BAKED OPOSSUM – 4 OZ.	235	10.6
WILD RABBIT – 4 OZ.	124	4.0
VENISON – 4 OZ.	128	3.1
DANDELION GREENS – ONE CUP COOKED	70	1.4
POTATO – MEDIUM	78	.2
PRICKLY PEAR – 4 OZ.	43	.2

Figure 18-1. Food and Calorie Diagram.

18-3. Food. Survivors should be able to find something to eat wherever they are. One of the best places to find food is along the seacoast, between the high and low watermark. Other likely spots are the areas between the beach and a coral reef; the marshes, mud flats, or mangrove swamps where a river flows into the ocean or into a larger river; riverbanks, inland waterholes, shores of ponds and lakes, margins of forests, natural meadows, protected mountain slopes, and abandoned cultivated fields.

a. Rations placed in survival kits have been developed especially to provide some of the proper sustenance needed during survival emergencies. When eaten as directed on the package, it will keep the survivor relatively efficient. If enough other food can be found, rations should be conserved for emergency use.

b. Consideration must be given to available food and water and how long the survival episode may last. Environmental conditions must also be considered. If a survivor is in a cold environment, more of the *proper* food will be required to provide necessary body heat. Rescue may vary from a few hours to several months, depending on the environment, operational commitments, and availability of rescue resources in that area. Available food must be rationed based on the estimated time which will elapse before being able to supplement issued rations with natural foods. If it is decided that some of the survivors should go for help, each traveler should be given twice as much food as those remaining behind. In this way, the survivors resting at the encampment and

those walking out will stay in about the same physical condition for about the same length of time.

c. If available water is less than a quart a day, avoid dry, starchy, and highly seasoned foods and meat. Keep in mind that eating increases thirst. For water conservation, the best foods to eat are those with high carbohydrate content, such as hard candy and fruit. All work requires additional food and water. When work is being performed, the survivor must increase food and water consumption to maintain physical efficiency. If food is available, it is alright to nibble throughout the day. It is preferable though to have at least two meals a day, with one being hot. Cooking usually makes food safer, more digestible, and palatable. The time spent cooking will provide a good rest period. On the other hand, some food such as sapodilla, star apple, and soursop, are not palatable unless eaten raw.

d. Native foods may be more appetizing if they are eaten by themselves. Rations and native foods usually do not mix well. In many countries, vegetables are often contaminated by human feces which the natives use as fertilizer. Dysentary is transmitted in this way. If possible, survivors should try to select and prepare their own meals. If necessary to avoid offending the natives, indicate that religious beliefs or taboos require self-preparation of food.

e. Learn to overcome food prejudices. Foods that may not look good to the survivor are often a part of the natives regular diet. Wild foods are high in mineral and vitamin content. With a few exceptions, all animals are edible when freshly killed. Avoid strange looking fish and fish with flesh that remains indented when depressed as it is probably becoming spoiled and should not be eaten. With knowledge and the ability to overcome food prejudices, a survivor can eat and sustain life in strange or hostile environment.

18-4. Animal Food. Animal food gives the most food value per pound. Anything that creeps, crawls, swims, or flies is a possible source of food. People eat grasshoppers, hairless caterpillars, wood-boring bettle larvae and pupae, ant eggs, spider bodies, and termites. Such insects are high in fat and should be cooked until dried. Everyone has probably eaten insects contained in flour, cornmeal, rice, beans, fruits, and greens in their daily foods.

a. Man as a Predator. To become successful in hunting, the hunter must go through a behavioral change and reorganize personal priorities. This means the one and only goal for the present is to kill an animal to eat. To kill this animal, the hunter must mentally become a predator. The hunter must be prepared to undergo stress in order to hunt down and kill an animal. Because of the type of weapons survivors are likely to have, it will be necessary to get very close to the animal to immobilize or kill it. This is going to require all the stealth and cunning survivors can muster. In addition to stealth and cunning, knowledge of the animal being hunted is very important. If in an unfamiliar area, survivors may learn much about the animal life of the area by studying signs such as trails, droppings, and bedding areas.

b. Animal Sign. The survivor should establish the general characteristics of the animals. The size of the tracks will give a good idea of the size of the animal. The depth of the tracks will indicate the weight of the animal. The animal dung can tell the hunter much. For example, if it is still warm or slimy, it was made very recently; if there is a large amount scattered around the area, it could well be a feeding or bedding area. The droppings may indicate what the animal feeds upon. Carnivores often have hair and bone in the dung; herbivores have coarse portions of the plants they have eaten. Many animals mark their territory by urinating or scraping areas on the ground or trees. These signs could indicate good trap or ambush sites. Following the signs (tracks, droppings, etc.) may reveal the feeding, watering, and resting areas. Well worn trails will often lead to the animal's watering place. Having made a careful study of all the signs of the animal, the hunter is in a much better position to procure it, whether electing to stalk, trap, or snare it, or lie in wait to shoot it.

c. Hunting. If survivors elect to hunt, there are some basic techniques which will be helpful and improve chances of success. Wild animals rely entirely upon their senses for their preservation. These senses are smell, vision, and hearing. Humans have lost the keenness of some of their senses like smelling, hearing, etc. To overcome this disadvantage, they have the ability to reason. As an example, some animals have a fantastic sense of smell, but this can be overcome by approaching the quarry from a downwind direction. The best times to hunt are at dawn and dusk as animals are either leaving or returning to their bedding areas. Both diurnal and nocturnal animals are active at this time. There are five basic methods of hunting:

(1) Still or Stand. This is the best method for inexperienced hunters as it involves less skill. The main principle of this method is to wait in ambush along a well-used game trail, until the quarry approaches within killing range. Morning and evening are usually the best times to still hunt. Care should be taken not to disturb the area; always wait downwind. Patience and self-control are necessary to remain motionless for long periods of time.

(2) Stalking. "Stalking" refers to the stealthily approach toward game. This method is normally used when an animal has been sighted and the hunter then proceeds to close the distance using all available cover. Stalking must be done slowly so that minimum noise is made; quick movement is easily detected by the animal. Always approach from the downwind side and move when the animal's head is down eating, drinking, or looking in another direction. The same techniques are

used in blind stalking as in the regular stalk, the main difference being that the hunter is stalking a position where the animal is expected to be while the animal is not in sight.

(3) Tracking. Tracking is very difficult unless conditions are ideal. This method involves reading all of the signs left behind by the animal, interpreting what the animal is doing, and how it can best be killed. The most common signs are trails, beds, urine, droppings, blood, tracks, and feeding signs.

(4) Driving. Some wild animals can be scared or driven in a direction where other hunters or traps have been set. This method is normally used where the game can be funneled; a valley or canyon is a good place to make a drive. More than one person is usually necessary to make a drive.

(5) Calling. Small predators may be called in by imitating an injured animal. Ducks and geese can be attracted by imitating their feeding calls. These noises can be made by sucking on the hand, blowing on a blade of grass or paper, sucking the lip, or using specially designed devices. Survivors should not call animals unless they know what they are doing as strange noises may "spook" the animal.

d. Killing Implements. It is difficult to kill animals of any size without using some type of tool or weapon. As our technology has increased in complexity, so have our killing tools. If a firearm is available, a basic knowledge of shooting and hunting techniques is necessary.

(1) Learning to become proficient with primitive weapons is important. Many primitive tribes of the world are still effectively using spears, clubs, bows and arrows, sling shots, etc., to provide food for their families. One of the limiting factors in the use of firearms is the amount of ammunition on hand. Therefore, a survivor cannot afford to waste ammunition on moving game or game which is beyond the effective range of the firearm being used. Wait for a pause in the animal's motions. The shot must be placed in a vital area with any firearm. Aim for the brain, spine, lungs, or heart (figure 18-2). A hit in these areas is usually fatal.

(2) A full-jacketed bullet often won't immediately down a larger animal hit in a vital area such as the lungs or heart. The alternative to losing the animal is tracking it to where it falls. Often it's better to wait awhile before pursuing the animal. If not pursued, it may lay down and stiffen or perhaps bleed to death. Follow the blood trail to where the game has gone down and kill it if it is

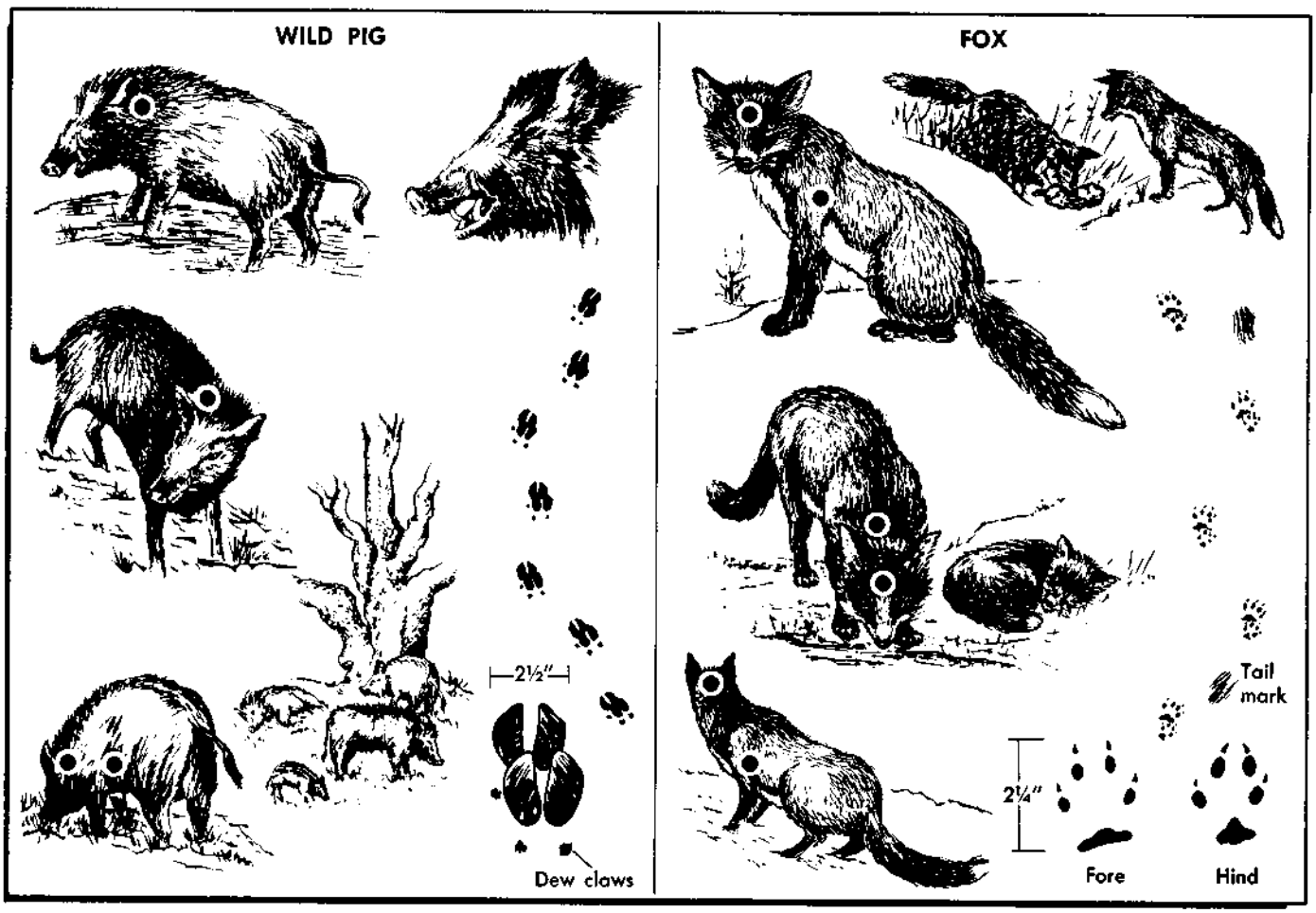

Figure 18-2. Shooting Game.

still alive. Even though ammunition might be limited, small game may be more productive than large game. Although they present smaller targets and have less meat, they are less wary, more numerous, and travel less distance to escape if wounded. A large amount of edible meat on small game can be destroyed from a bullet wound. On rodents, most of the meat is on the hindquarters and frontquarters; birds, it is the breast and legs. The survivor should try to hit a vital spot that spoils the least meat.

(3) Night hunting is usually best, since most animals move at night. A flashlight or torch may be used to shine in the animal's eyes. It will be partly blinded by the light and a survivor can get much closer than in the daytime. If no gun is available, the animal can be killed with a club or a sharpened stick used as a spear.

(4) Remember that large animals, when wounded, cornered, or with their young, can be dangerous. Be sure the animal is dead, not just wounded, unconscious, or playing "possum." Animals usually die with their eyes open and glazed-over. Poke all "dead" animals in the eye with a long sharp stick before approaching them.

(5) Small freshwater turtles can often be found sunning themselves along rivers and lakeshores. If they dash into shallow water, they can still be procured with nets, clubs etc; watch out for mouth and claws. Frogs and snakes also sun and feed along streams. Use both hands to catch a frog—one to attract it and keep it busy while grabbing it with the other. Bright cloth on a fishhook also works. All snakes are good eating and can be killed with a long stick. Both marine and dry-land lizards are edible. A noose, small fishhook baited with a bright cloth lure, slingshot, or club can be used. A slingshot can be made with a forked stick and the elastic from the parachute pack or surgical tubing found in some survival kits (figure 18-3). With practice, the slingshot can be very effective for killing small animals.

e. Snaring and Trapping. Snaring and trapping animals are ways survivors can procure animal food to supplement issued rations. Since small animals are usually more abundant than large animals, they will probably be the survivor's main source of food. Snares should be set out on a 15:1 ratio; 15 snares should be set out for every one animal expected to be caught.

(1) Using traps and snares are more advantageous then going out on foot and physically hunting the animal. The most important advantage being that traps work 24 hours a day with no assistance from the hunter. A large area can be effectively strapped with the possibility of catching many animals within the same period of time. Survivors (generally) use much less energy maintaining a trapline than is used by hunting. This means less food is required because less energy is expended.

(2) The traps or snares should be set in areas where the game is known to live or travel. Look for signs such as tracks, droppings, feeding signs, or actual sightings of the animal. If snares are used, they should be set up to catch the animal around the neck. Therefore, the loop must allow the head to pass through but not the body. Loops will vary in size from one animal to another. When placing snares, try to find a narrow area of the game trail where the animal has no choice but to enter the loop. If a narrow area cannot be found, brush or other obstacles can be arranged to funnel the animal into the snare (figure 18-4). Do not overdo the funneling; use as little as possible. Avoid disturbing the natural surroundings if possible. Do not walk on game trails, but approach 90 degrees to the trail, set the snare, and back away. Snares may also be set over holes or burrows. All snares and traps should be set during the midday because most animals are nocturnal in nature. Check snares and traps twice daily. If possible, check after sunup and before sunset. The checks should be made from a distance so any animals moving at the time of checking will not be disturbed or frightened away.

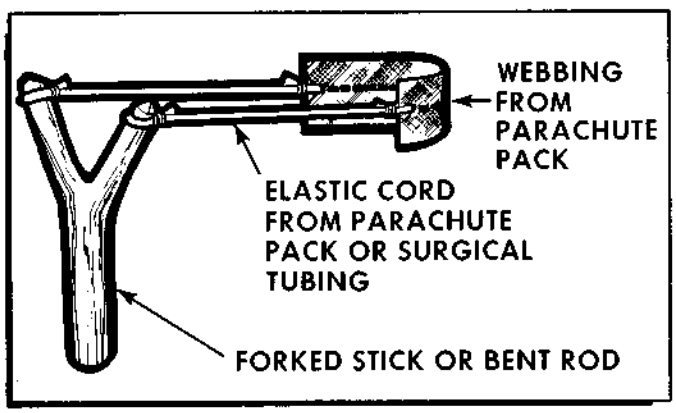

Figure 18-3. Slingshot.

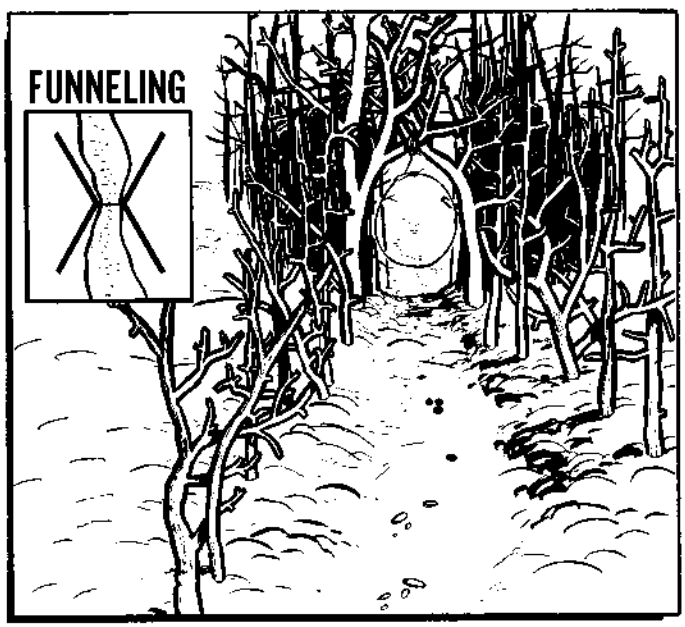

Figure 18-4. Funneling.

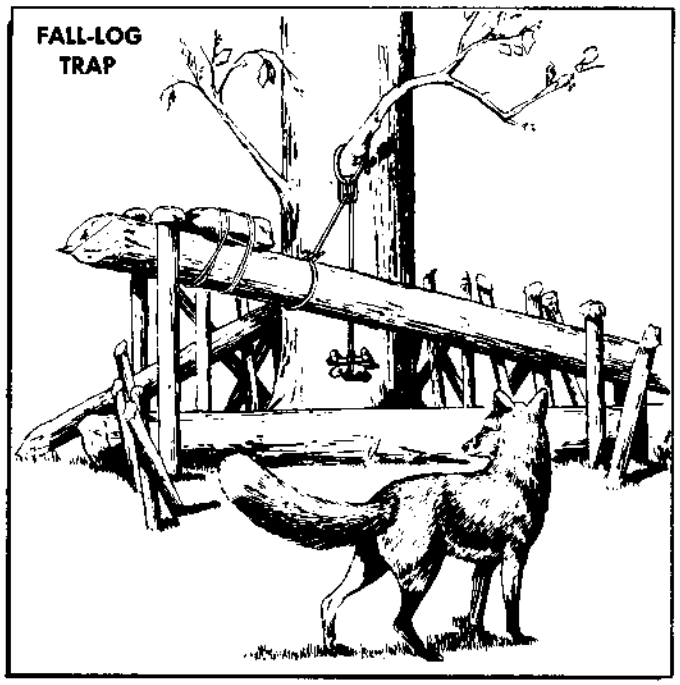

Figure 18-5. Mangle.

(3) There are three ways to immobilize or trap animals.

(a) Strangle. This is done by simply using a free-sliding noose which, when tightened around the neck, will restrict circulation of air and blood. The materials should be strong enough to hold the animal; for example, suspension line, string, wire, cable, or rawhide.

(b) Mangle. Mangle traps use a weight which is suspended over the animal's trail or over bait. When the animal trips the trigger, the weight (log) will descend and mangle the animal (figure 18-5).

(c) Hold. Any means of impeding the animal and detaining its progress would be considered a hold-type trap.

(4) The apache foot snare is an example of a hold-type trap. It is used for large browsers and grazers like deer (figure 18-6). It should be located along game trails where an obstruction, such as a log, blocks the trail. When animals jump over this obstruction, a very shallow depression is formed where their hooves land. The apache foot snare should be placed at this depression. The box trap for birds is another example of hold-type traps (figure 18-7).

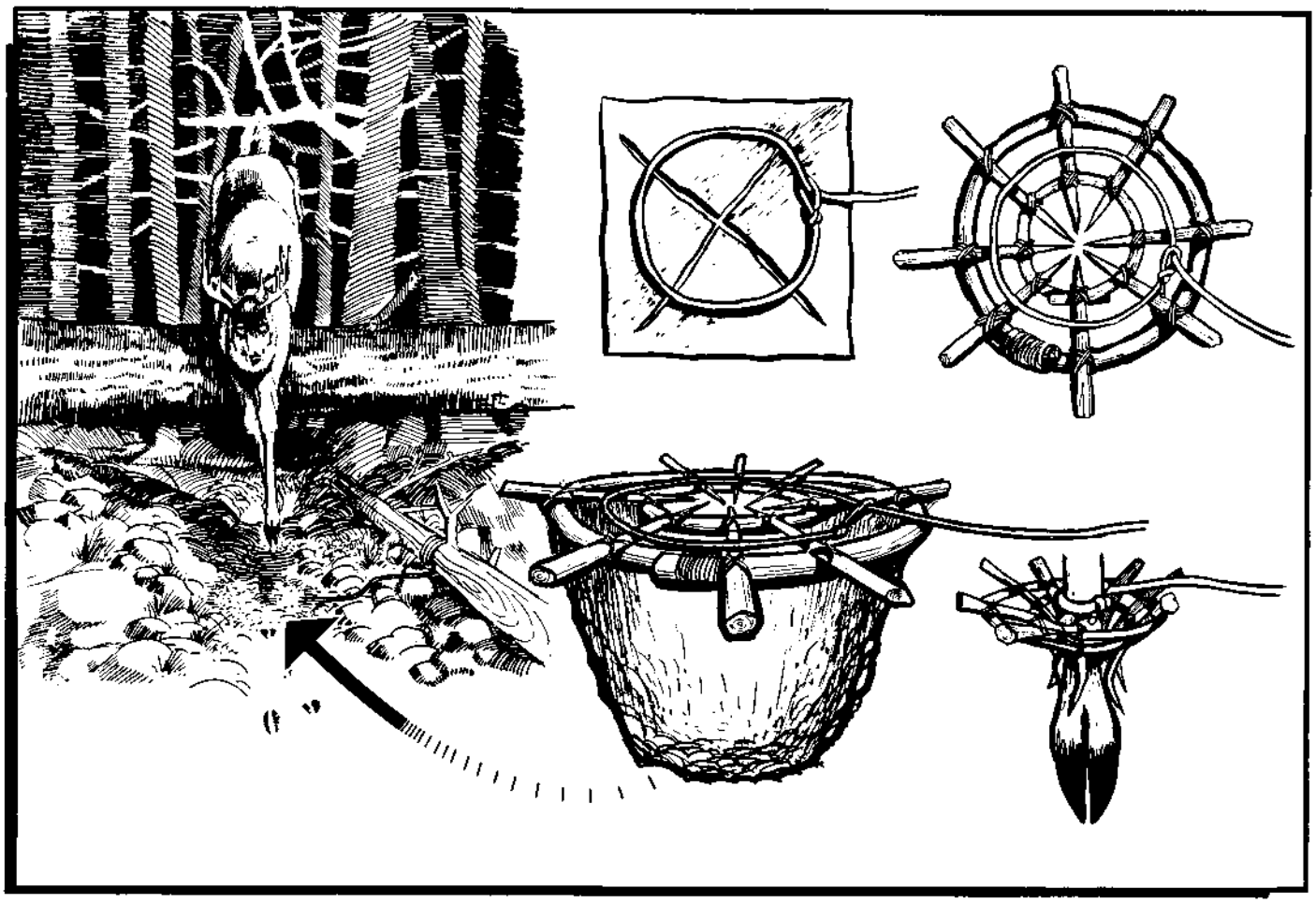

Figure 18-6. Apache Foot Snare.

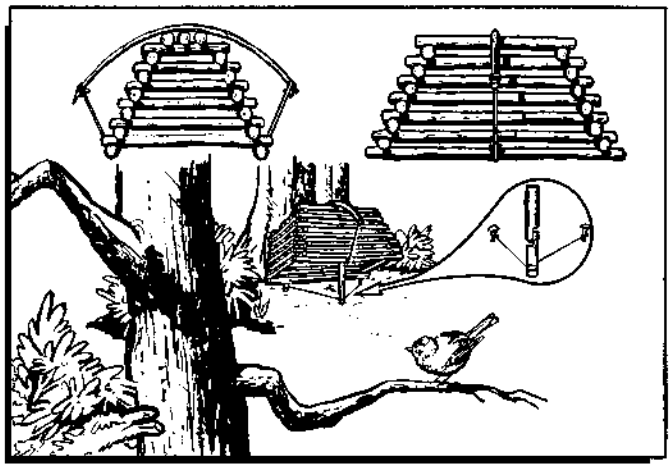

Figure 18-7. Box Trap.

(5) The simple loop is the quickest snare to construct. All snares and traps should be simple in construction with as few moving parts as possible. This loop can be constructed from any type of bare wire, suspension line, inner core, vines, long strips of green bark, clothing strips or belt, and any other material that will not break under the strain of holding the animal. If wire is being used for snares, a figure "8" or locking loop should be used (figure 18-8). Once tightened around the animal, the wire is locked into place by the figure "8" which prevents the loop from opening again. A simple loop snare is generally placed in the opening of a den, with the end of the snare anchored to a stake or similar object (figure 18-8). The simple loop snare can also be used when making a squirrel pole (figure 18-9) or with some types of trigger devices.

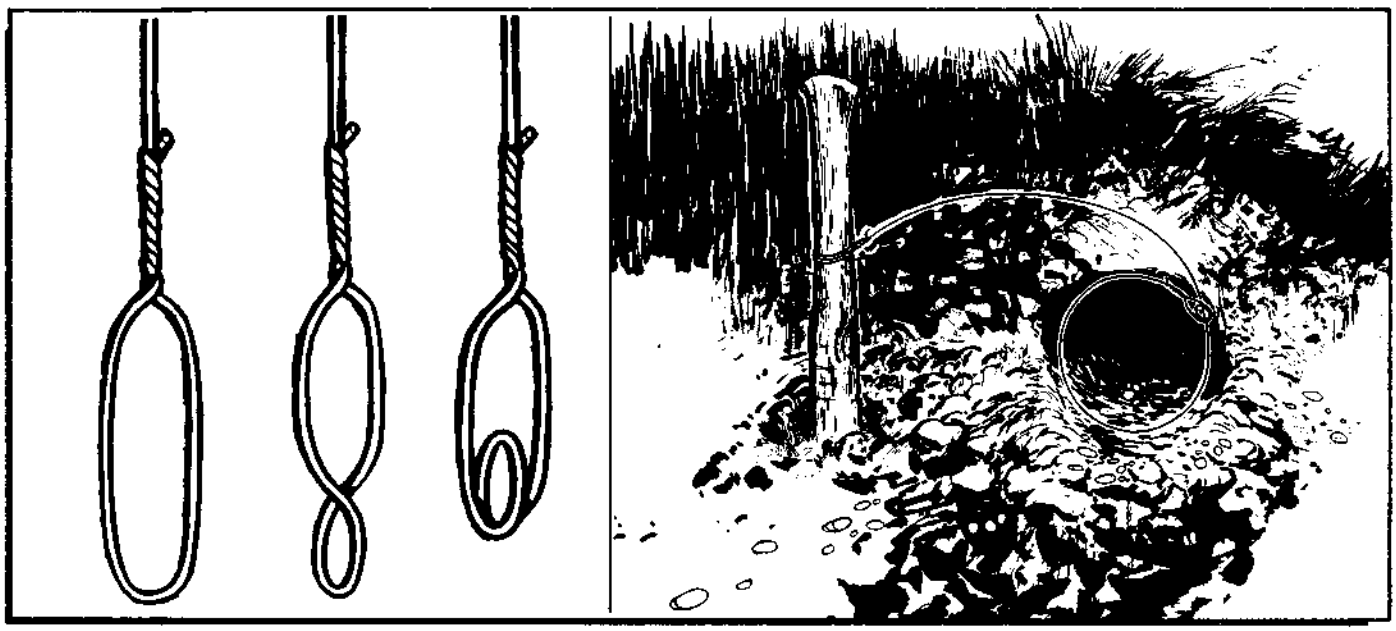

Figure 18-8. Locking Loop and Setting Noose.

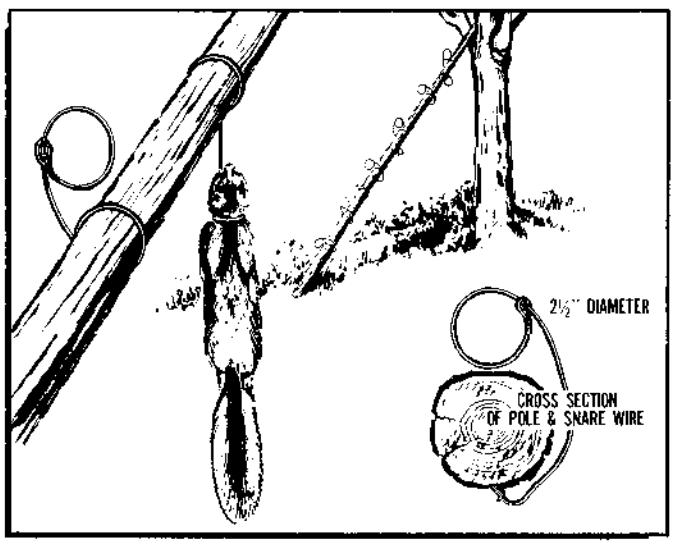

Figure 18-9. Squirrel Pole.

f. Triggers. Triggers may be used with traps. The purpose of the trigger is to set the device in motion, which will eventually strangle, mangle, or hold the animal. There are many triggers. Some of the more common ones are:

(1) Two-pin toggle with a counterweight for small to medium animals which are lifted out of the reach of predators (figure 18-10).

(2) Figure "H" with wire snare for small mammals and rodents (figure 18-11).

(3) Canadian ace for predators such as bobcat, coyote, etc., (figure 18-12).

(4) Three-pin toggle with deadfall for medium to large animals (figure 18-13). Medium and large animals can be captured using deadfalls, but this type of trap is recommended only when big game exists in large quantities to justify the great expense of time and effort spent in constructing the trap.

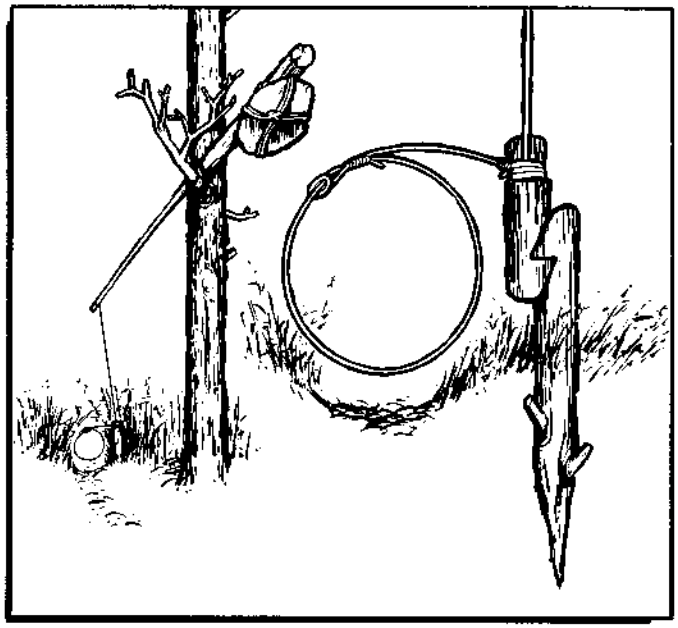

Figure 18-10. Two-Pin Toggle.

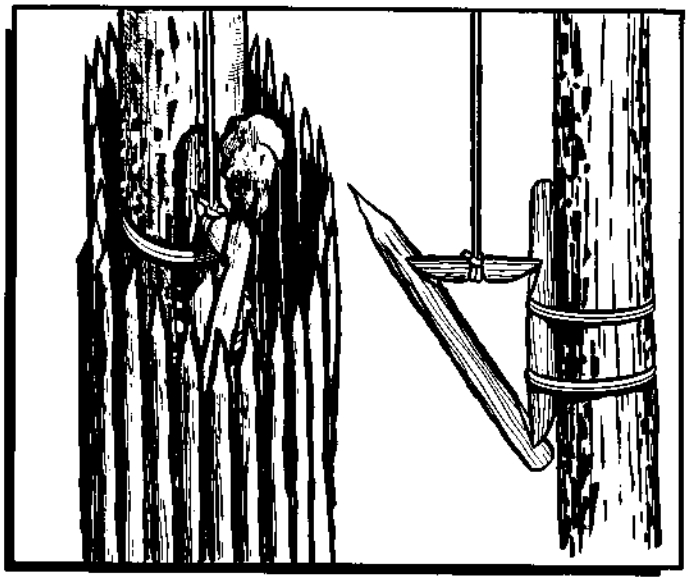

Figure 18-12. Canadian Ace.

(5) The twitch-up snare which incorporates the simple loop, can be used to catch small animals (figure 18-14). When the animal is caught, the sapling jerks it up into the air and keeps the carcass out of the reach of predators. This type of snare will not work well in cold climates, since the bent sapling will freeze in position and not spring up when released.

(6) A long forked stick can be used as a twist stick to procure ground squirrels, rabbits, etc. A den that has signs of activity must be located. Using the long forked stick, the survivor probes the hole with the forked end until something soft is felt then twisting the stick will entangle the animal's hide in the stick and the animal can be extracted (figure 18-18).

g. Birds. Birds can be caught with a gill net. The net should be set up at night vertically to the ground in some natural flyway, such as an opening in dense foliage. A small gill net on a wooden frame with a disjointed stick for a trigger can also be used. A gill net can be made by using inner core from parachute suspension line (figure 18-20).

(1) Birds can be caught on baited fishhooks (figure 18-15) or simple slipping loop snares. Bird's nest can be a source of food. All bird eggs are edible when fresh. Large wading birds such as cranes and herons often nest in mangrove swamps or in high trees near water.

(2) During molting season, birds cannot fly because of the loss of their "flight" feathers; they can be procured by clubbing or netting.

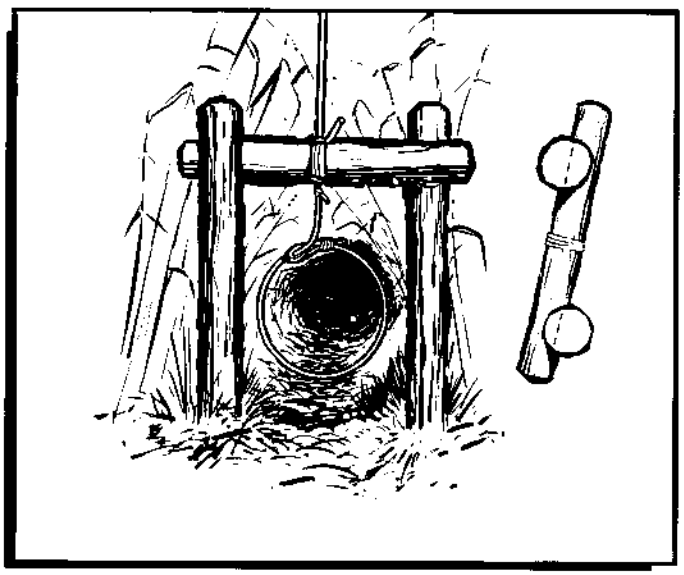

Figure 18-11. Figure H.

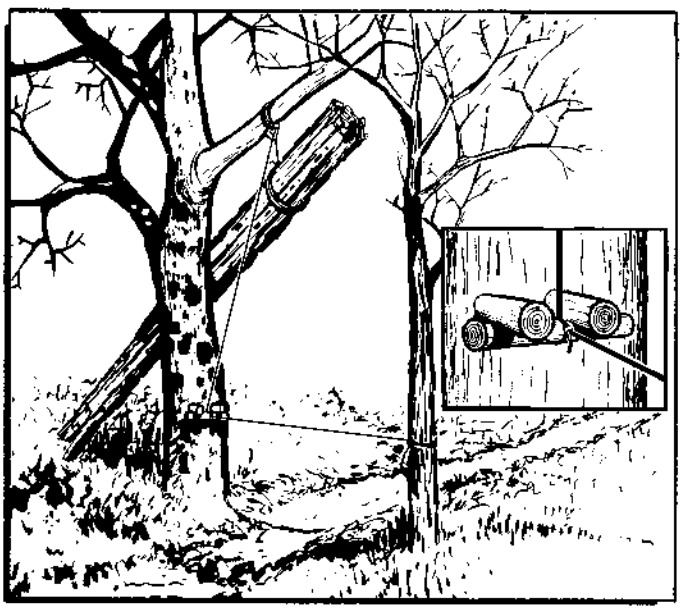

Figure 18-13. Three-Pin Toggle.

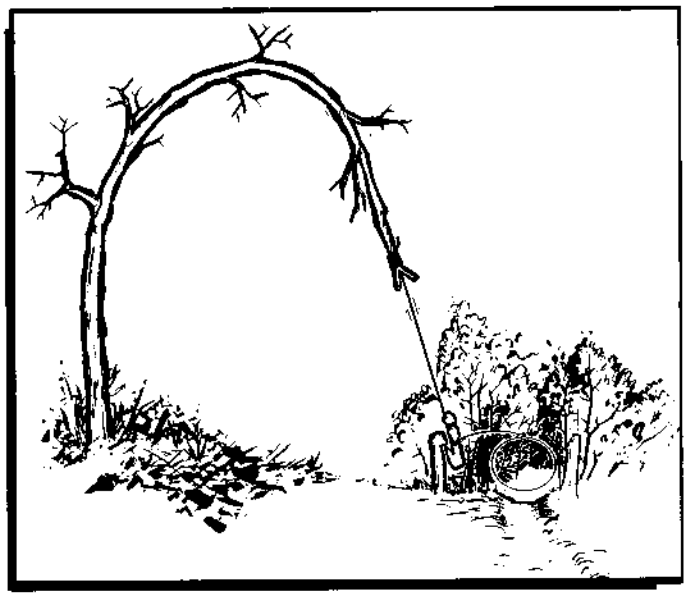

Figure 18-14. Twitch Up.

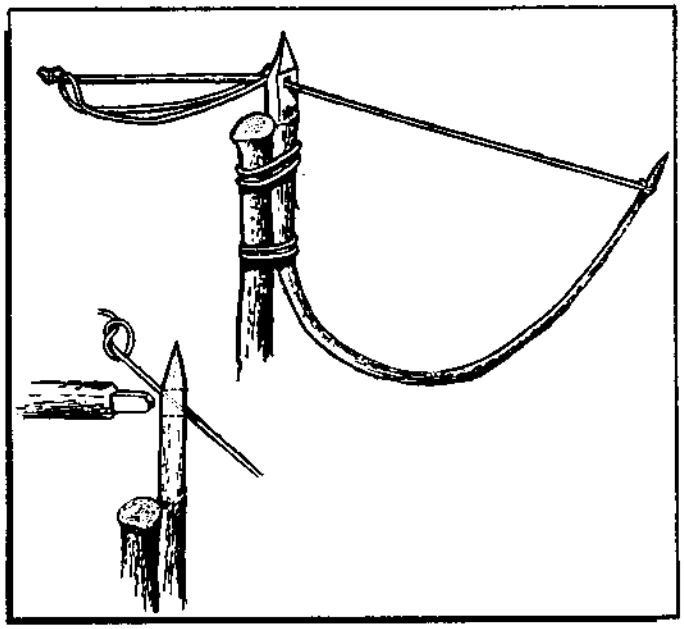

Figure 18-16. Ojibwa Bird Snare.

(3) Birds can be also caught in an Ojibway snare. This snare is made by cutting a 1- or 2-inch thick sapling at a height of 4½ to 5 feet above the ground (figure 18-16). A springy branch is then whittled flat at the butt end and a rectangular hole is cut through the flattened end. One end of a ½-inch thick stick, 15 inches long, is then whittled to fit slightly loose in the hole and the top corner of the whittled end is rounded off so the stick will easily drop away from the hole. The branch is then tied by its butt end to the top of the sapling. A length of inner core from suspension line is tied to the bottom end of the branch and the branch is bent into a bow with the line passing through the hole in the butt end. A knot is tied in the line and the 15-inch stick is then placed in the hole to lock the line in place (just behind the knot). An 8-inch loop is made at the end of the line and laid out on the 15-inch stick (spread out as well as possible). A piece of bait is placed on top of the sapling, and when a bird comes to settle on the 15-inch stick, the stick drops from the hole causing the loop to tighten around the bird's legs.

(4) When many birds frequent a particular type of bush, some simple loop snares may be set up throughout the bush. Make the snares as large as necessary for the particular type of birds that come to perch, feed, or roost there (figure 18-17).

(5) In wild, wooded areas, many larger species of birds such as spruce grouse and ptarmigan may be approached. The spruce grouse, which has merited the name of "fools hen," can be approached and killed with

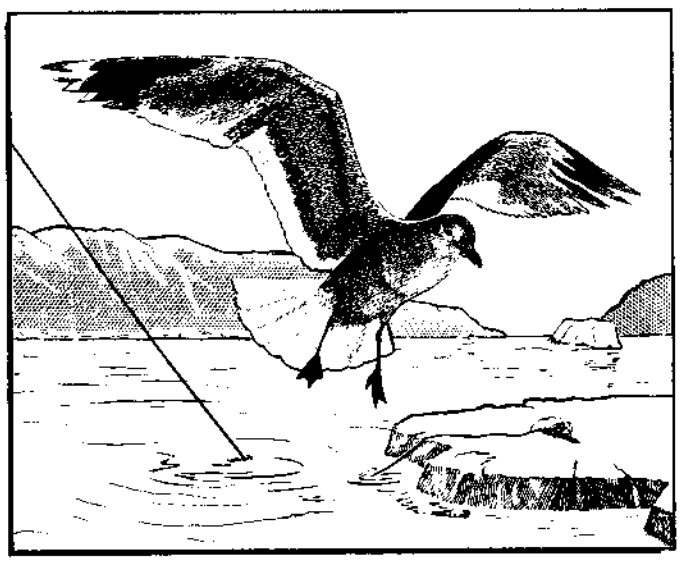

Figure 18-15. Baited Fishhook.

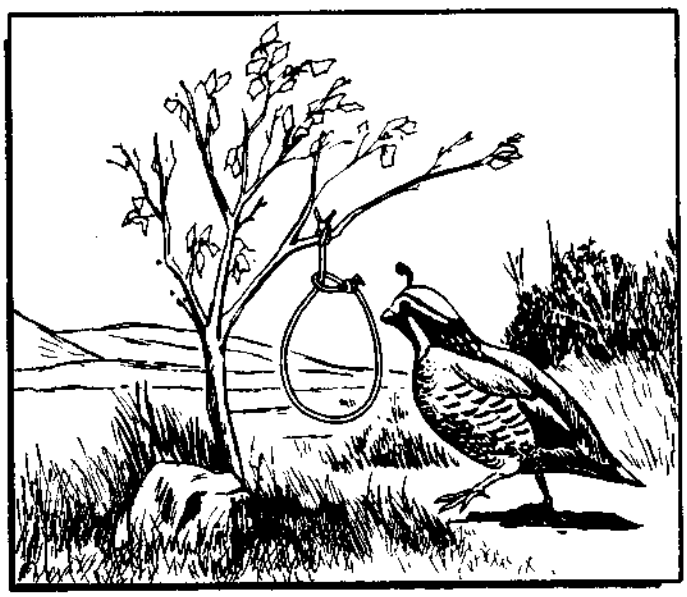

Figure 18-17. Ptarmigan or Small Game Snare.

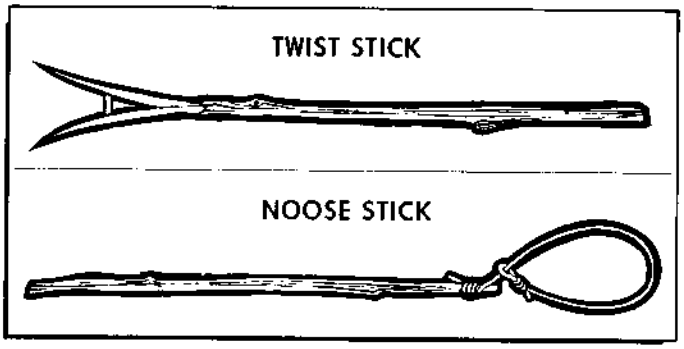

Figure 18-18. Twist Stick and Noose Stick.

a stick with little trouble. It often sits on the lower branches of trees and can be easily caught with a long stick with a loop at the end (figure 18-18).

(6) Ground feeding birds (Quail, Hungarian Partridge, Chukar) can be trapped in a trench dug into the ground. The trench should be just wide enough for the bird to walk into, so survivors must first observe the type of ground feeding birds in the area. The trench should be 2 to 3 feet long and about 10 to 12 inches deep at the deep end. The other end of the trench should be ramped down from the surface level. Bait is scattered along the surface into the pit, and after having pecked the last piece of bait the bird will not be able to get out of the pit because it can't fly out or climb out, its feathers keep it from backing out, and it can't turn around to walk out.

(7) Perching birds may be captured by using bird lime. Bird lime is a term applied to any sticky or gluey substance which is rubbed on a branch to prevent the flight of a bird which has landed on it or has flapped a wing against it. Bird lime is usually made from the sap of plants in the Euphorbia family. The common names of some of these plants are spotted spurge, cypress spurge, snow-on-the-mountain, and poinsettias. The Euphorbias have a wide range in North and Central America. The milky sap is poisonous and may cause blisters on the skin and should be handled with care. Bird lime is most effective in the desert and jungle, but

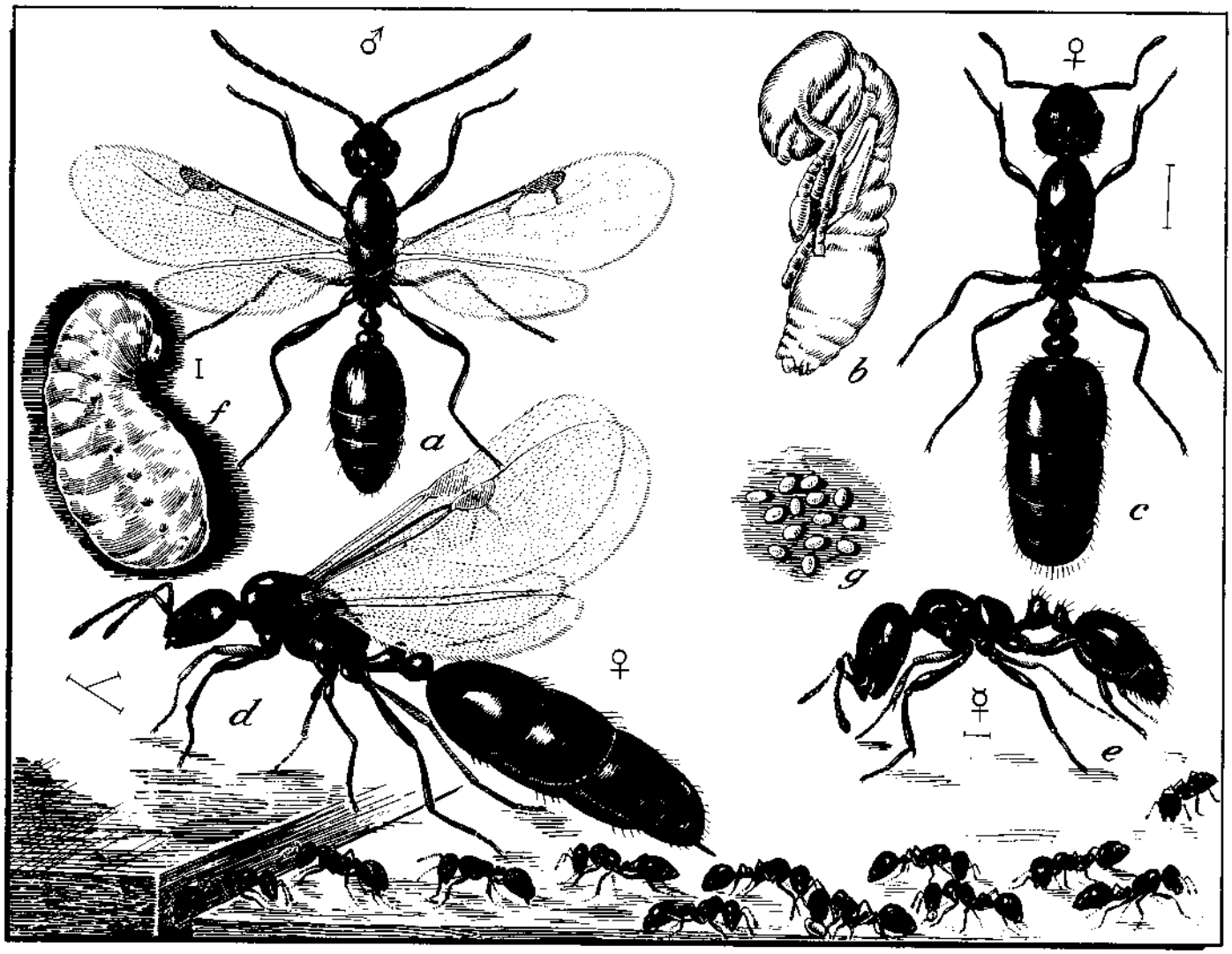

Figure 18-19. Ants.

it will not work in cold weather. Dust will make bird lime ineffective, so it should be used in spots where dust is not prevalent. The sap of the breadfruit tree makes excellent bird lime as it swells and become glutinous upon contact with air.

h. Insects. If there *ever* is a time when food aversions must be overcome, it is when survivors turn to insects as a food source.

(1) Primitive peoples eat insects and consider them great delicacies. When food is limited and insects are available, they can become a valuable food source. In some places, locusts and grasshoppers, cicadas, and crickets are eaten regularly; occasionally termites, ants, and a few species of stonefly larvae are consumed. Big beetles such as the Goliath Beetle of Africa, the Giant Water Beetles, and the big Long Horns are relished the world over. Clusters, like those of the Snipefly Atherix (that overhang the water), and the windrows of Brinefly puparia are eaten. Aquatic water bugs of Mexico are grown especially for food. All stages of growth can be eaten, including the eggs but, the large insects must be cooked to kill internal parasites.

(2) Termites and white ants are also an important food source. Strangely enough, these are closely related to cockroaches. The reason they are eaten so extensively in Africa is the fact that they occur in enormous numbers and are easily collected both from their nests and during flight. They are sometimes attracted to light in unbelievable numbers and the natives become greatly excited when the large species appear.

(3) Many American Indian tribes made a habit of eating the large carpenter ants that are sometimes pests in houses. These were eaten both raw and cooked. Even today the practice of eating them has not entirely disappeared, although they do not form an essential part of the diet of any of the inhabitants of this country (figure 18-19).

(4) It is not at all unnatural that the American Indians should have relished the honey ants in all parts of the continent where they occur. These ants are peculiar in that some of the workers become veritable storehouses for honey, their abdomens becoming more or less spherical and so greatly enlarged that they are scarcely able to move. They cluster on the ceilings and walls of their nests and disgorge part of their stored food to other inhabitants. The Indians discovered the sweetness stored in these insects and made full use of it. At first they ate the ants alive, later gathering them in quantity and crushing them so that they formed an enticing dish—one which was considered a delicacy and served to guests of distinction as a special favor. The next step in the use of the honey ant was the extraction of the pure honey by crushing the insects and straining the juices. After the honey was extracted, it was allowed to ferment, forming what is said to be a highly flavored wine.

(5) Indians of the American tropics, with a much larger ant fauna from which to choose, select the queens of the famous leafcutting or so-called umbrella ants upon which to feed, eating only the abdomens, either raw or cooked.

(6) It is natural that caterpillers, the larvae of moths and butterflies, should form a very substantial part of the food of primitive peoples because these are often of large size or occur in great abundance. In Africa, many tribes consider caterpillers choice morsels of food, and much time is spent in collecting them. Some of the native tribes recognize 20 or more different kinds of caterpillars that are edible, and are sufficiently well acquainted with the life history of the insects to know the plants upon which they feed and the time of year when they have reached the proper stage of development for collecting. Caterpillars with hairs should be avoided. If eaten, the hairs may become lodged in the throat causing irritation or infection. Today it is known that insects have nutritional or medicinal value. The praying mantis, for example, contains 58 percent protein, 12 percent fat, 3 percent ash, vitamin B complex, and vitamin A. The insect's outer skeleton is an interesting compound of sugar and amino acids.

(7) Bee larvae were eaten by the ancient Chinese. Some Chinese today eat locusts, dragonflies, and bumblebees. Cockroaches and locusts are a favorite dish in Szechuan. In Kwangtun, grasshoppers, golden June beetles, crickets, wasp larvae, and silkworm larvae are used for food.

(8) Stinging insects should have their stinging apparatus removed before they are eaten.

(9) As can be seen, insects have been used as a food source for thousands of years and will undoubtedly continue to be used. If survivors cannot overcome their aversion to insects as a food source, they will miss out on a valuable and plentiful supply of food.

i. Fishing. Fishing is one way to get food throughout the year wherever water is found. There are many ways to catch fish which include hook and line, gill nets, poisons, traps, and spearing.

(1) If an emergency fishing kit is available, there will be a hook and line in it, but if a kit is not available a hook and line will have to be procured elsewhere or improvised. Hooks can be made from wire or carved from bone or wood. The line can be made by unraveling parachute suspension line or by twisting threads from clothing or plant fibers. A piece of wire between the fishing line and the hook will help prevent the fish from biting through the line. Insects, smaller fish, shellfish, worms, or meat can be used as bait. Bait can be selected by observing what the fish are eating. Artificial lures can be made from pieces of brightly colored cloth, feathers, or bits of bright metal or foil tied to a hook. If the fish will not take the bait, try to snag or hook them in any part of the body as they swim by. In freshwater, the deepest water is usually the best place to fish. In shallow streams, the best places are pools below falls, at the foot of rapids, or behind rocks. The best time to fish is usual-

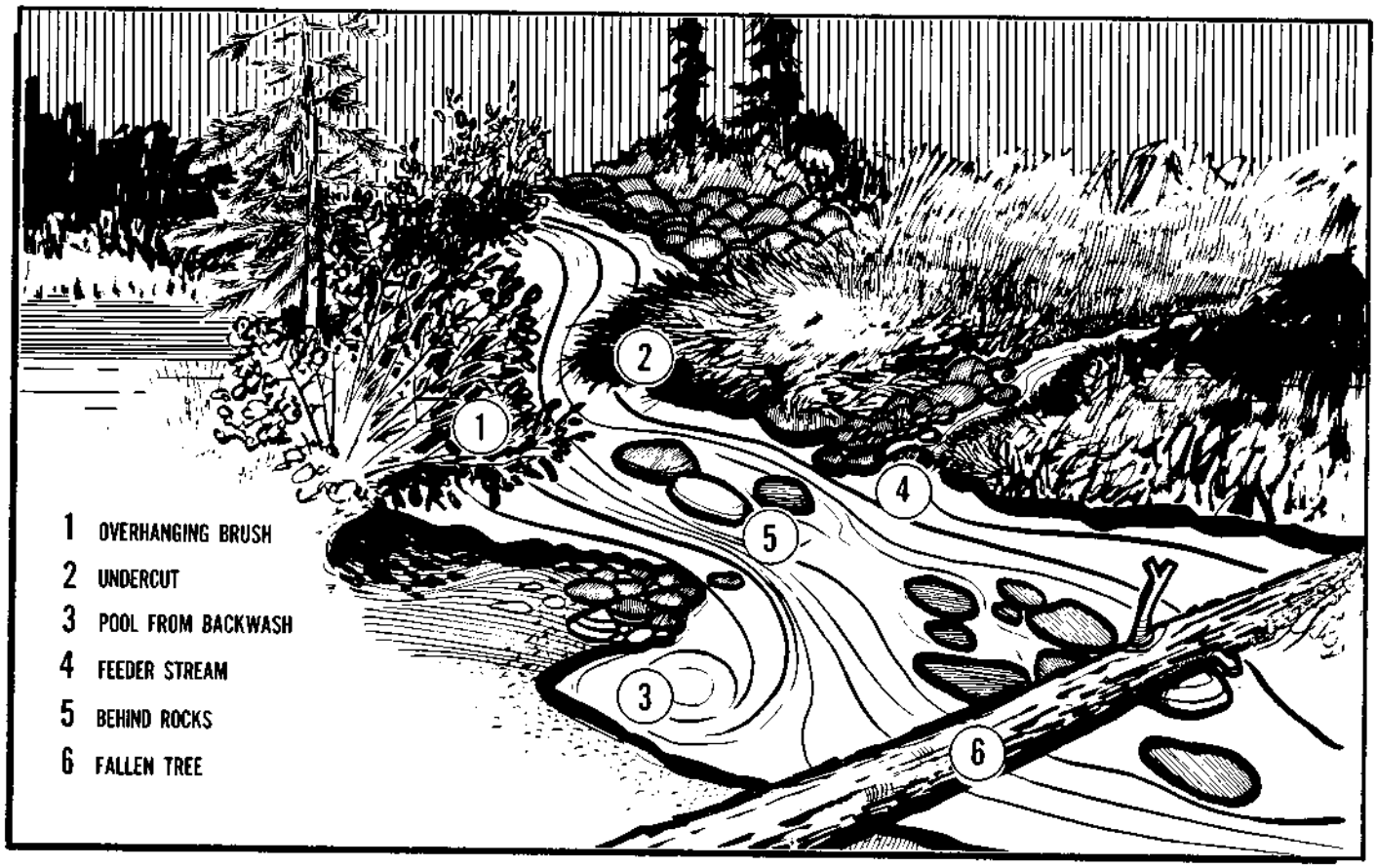

Figure 18-20. Fishing Places.

ly early morning or late evening (figure 18-20). Sometimes fishing is best at night, especially in moonlight or if a light is available to attract the fish. The survivor should be patient and fish at different depths in all kinds of water. Fishing at different times of the day and changing bait often is rewarding.

(2) The most effective fishing method is a net because it will catch fish without having to be attended (figures 18-21 and 18-22). If a gill net is used, stones can be used for anchors and wood for floats. The net should be set at a slight angle to the current to clear itself of any floating refuse that comes down the stream. The net should be checked at least twice daily (figure 18-23). A net with poles attached to each end works effectively if moved up or down a stream as rapidly as possible while moving stones and threshing the bottom or edges of the streambanks. The net should be checked every few moments so the fish cannot escape.

(3) Shrimp (prawns) live on or near the sea bottom and may be scraped up. They may be lured to the surface with light at night. A hand net made from parachute cloth or other material is excellent for catching shrimp. Lobsters are creeping crustaceans found on or near the sea bottom. A lobster trap, jig, baited hook, or dip net can be used to catch lobster. Crabs will creep, climb, and burrow and are easily caught in shallow water with a dip net or in traps baited with fish heads or animal viscera.

(4) Fishtraps (figure 18-24) are very useful for catching both freshwater and saltwater fish, especially those that move in schools. In lakes or large streams, fish tend to approach the banks and shallows in the morning and evening. Sea fish, traveling in large schools, regularly approach the shore with the incoming tide, often moving parallel to the shore guided by obstruction in the water.

(a) A fishtrap is basically an enclosure with a blind opening where two fence-like walls extend out, like a funnel, from the entrance. The time and effort put into building a fishtrap should depend on the need for food and the length of time survivors plan to stay in one spot.

(b) The trap location should be selected at high tide and the trap built at low tide. One to 2 hours of work should do the job. Consider the location, and try to adapt natural features to reduce the labors. Natural rock pools should be used on rock shores. Natural pools on the surface of reefs should be used on coral islands by blocking the opening as the tide recedes. Sandbars, and the ditches they enclose, can be used on sandy shores. The best fishing off sandy beaches is the lee side of offshore sandbars. By watching the swimming habits

of fish, a simple dam can be built which extends out into the water forming an angle with the shore. This will trap fish as they swim in their natural path. When planning a more complex brush dam, select protected bays or inlets using the narrowest area and extending one arm almost to the shore.

(c) In small, shallow streams, the fishtraps can be made with stakes or brush set into the stream bottom or weighted down with stones so that the stream is blocked except for a small narrow opening into a stone or brush pen or shallow water. Wade into the stream, herding the fish into the trap, and catch or club them when they get in shallow water. Mud-bottom streams can be trampled until cloudy and then netted. The fish are blinded and cannot avoid the nets. Freshwater crawfish and snails can be found under rocks, logs, overhanging bushes, or in mud bottoms.

(5) Fish may be confined in properly built enclosures and kept for days. In many cases, it may be advantageous to keep them alive until needed and thus ensure there is a fresh supply without danger of spoilage. Mangrove swamps are often good fishing grounds. At low tide, clusters of oysters and mussles are exposed on the mangrove "knees" or lower branches. Clams can be found in the mud at the base of trees. Crabs are very active among branches or roots and in the mud. Fish can be caught at high tide. Snails are found on mud and clinging to roots. Shellfish which are not covered at high tide or those from a colony containing diseased members should not be eaten. Some indications of diseased shellfish are shells gaping open at low tide, foul odor, and (or) milky juice.

(6) Throughout the warm regions of the world, there are various plants which the natives use for poisoning fish. The active poison in these plants is harmful only to cold-blooded animals. Survivors can eat fish killed by this poison without ill effects.

(a) In Southeast Asia, the derris plant is widely

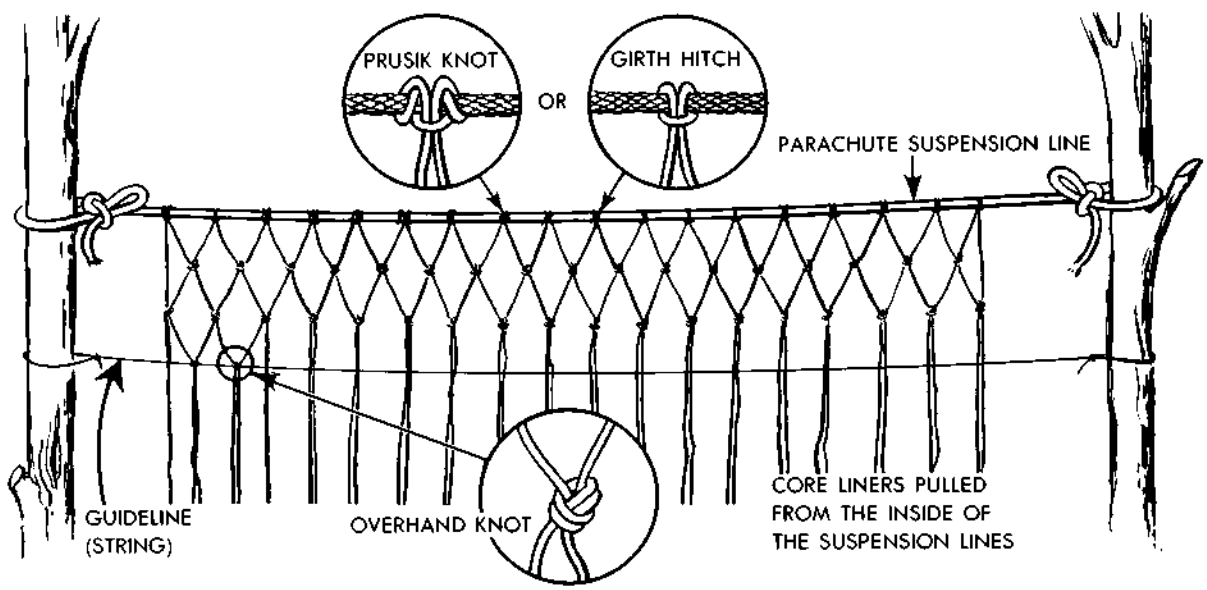

1. Suspend a suspension line casing (from which the core liners have been pulled) between two uprights, approximately at eye level.

2. Hang core liners (an even number) from the line suspended as in 1, above. These lines should be attached with a Prusik knot or girth hitch and spaced in accordance with the mesh you desire. One-inch spacing will result in a 1-inch mesh, etc. The number of lines used will be in accord with the width of the net desired. If more than one man is going to work on the net, the length of the net should be stretched between the uprights, thus providing room for more than one man to work. If only one man is to make up the net, the depth of the net should be stretched between the uprights and step 8, below, followed.

3. Start at left or right. Skip the first line and tie the second and third lines together with an overhand knot. Space according to mesh desired. Then tie fourth and fifth, sixth and seventh, etc. One line will remain at the end.

4. On the second row, tie the first and second, third and fourth, fifth and sixth, etc., to the end.

5. Third row, skip the first line and repeat step 3 above.

6. Repeat step 4, and so on.

7. You may want to use a guide line which can be moved down for each row of knots to ensure equal mesh. Guide line should run across the net on the side opposite the one you are working from so that it will be out of your way.

8. When you have stretched the depth between the uprights and get close to ground level, move the net up by rolling it on a stick and continue until the net is the desired length.

9. String suspension line casing along the sides when net is completed to strengthen it and make the net easier to set.

Figure 18-21. Making a Gill Net.

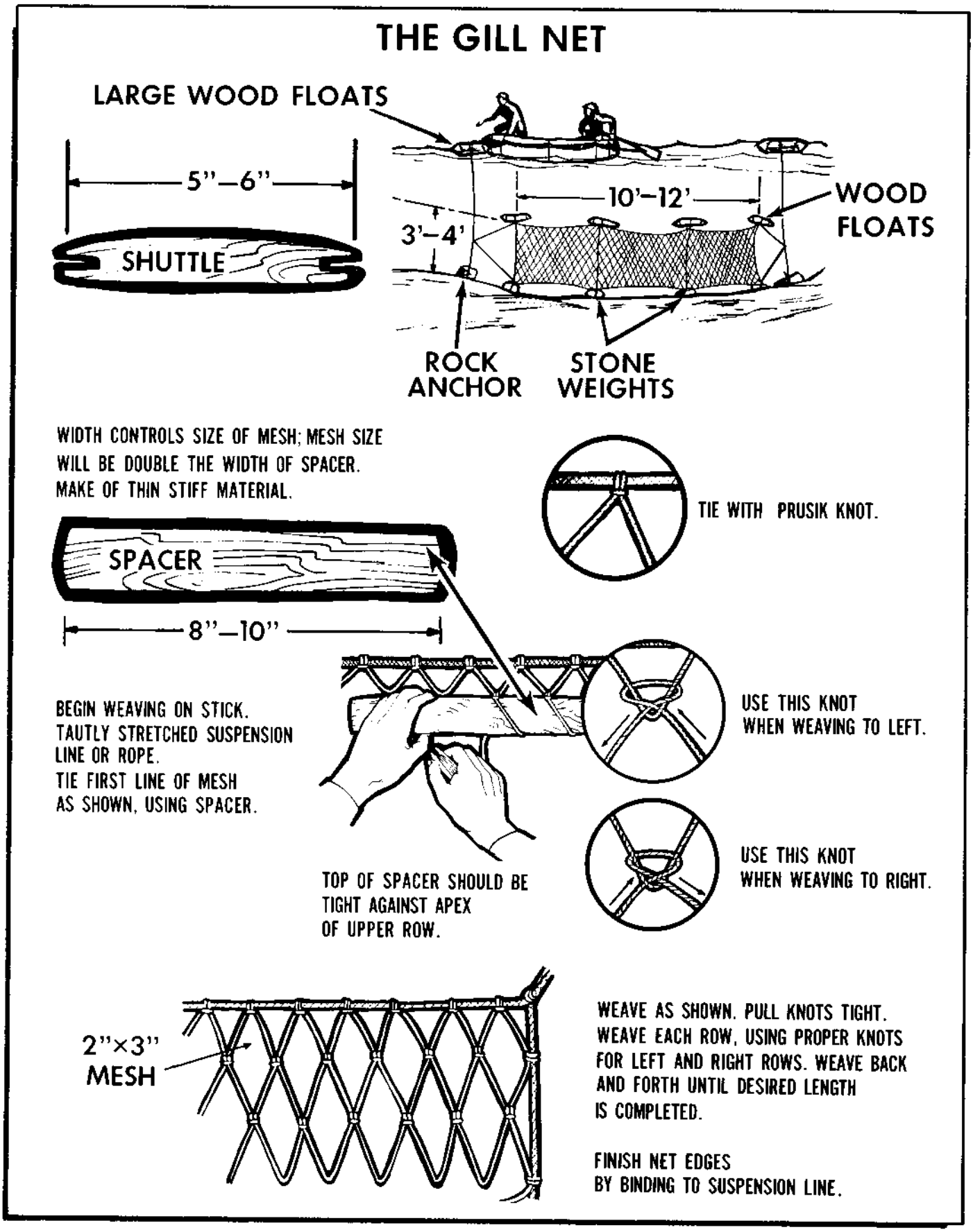

Figure 18-22. Making a Gill Net With Shuttle and Spacer.

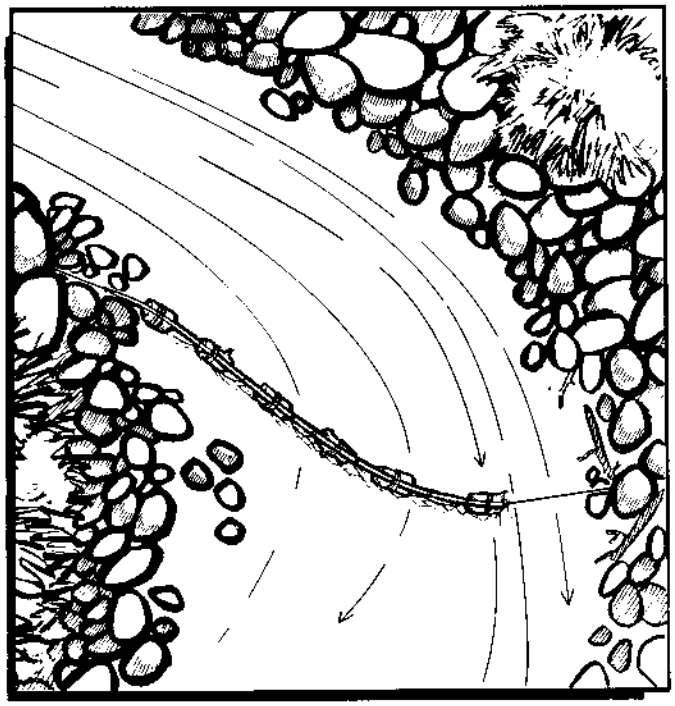

Figure 18-23. Setting the Gill Net.

used as a source of fish poison. The derris plant, a large woody vine, is also used to produce a commercial fish poison called rotenone. Commercial rotenone can be used in the same manner as crushed derris roots; it causes respiratory failure in fish, but has no ill effects on humans. However, rotenone has no effect if dusted over the surface of a pond. It should be mixed to a malted-milk consistency with a little water, and then distributed in the water. If the concentration is strong, it takes effect within 2 minutes in warm water, or it may take an hour in colder water. Fish sick enough to turn over on their backs will eventually die. An ounce of 12 percent rotenone can kill every fish for a half mile down a slow-moving stream that is about 25 feet wide. A few facts to remember about the use of rotenone are:

-1. It is very swift acting in warm water at 70°F and above.

-2. It works more slowly in cold water and is not practical in water below 55°F.

-3. It is best applied in small ponds, streams, or tidal pools.

-4. Excess usage will be wasted. However, too little will not be effective.

(b) A small container of 12 percent rotenone (one-half ounce) is a valuable addition to any emergen-

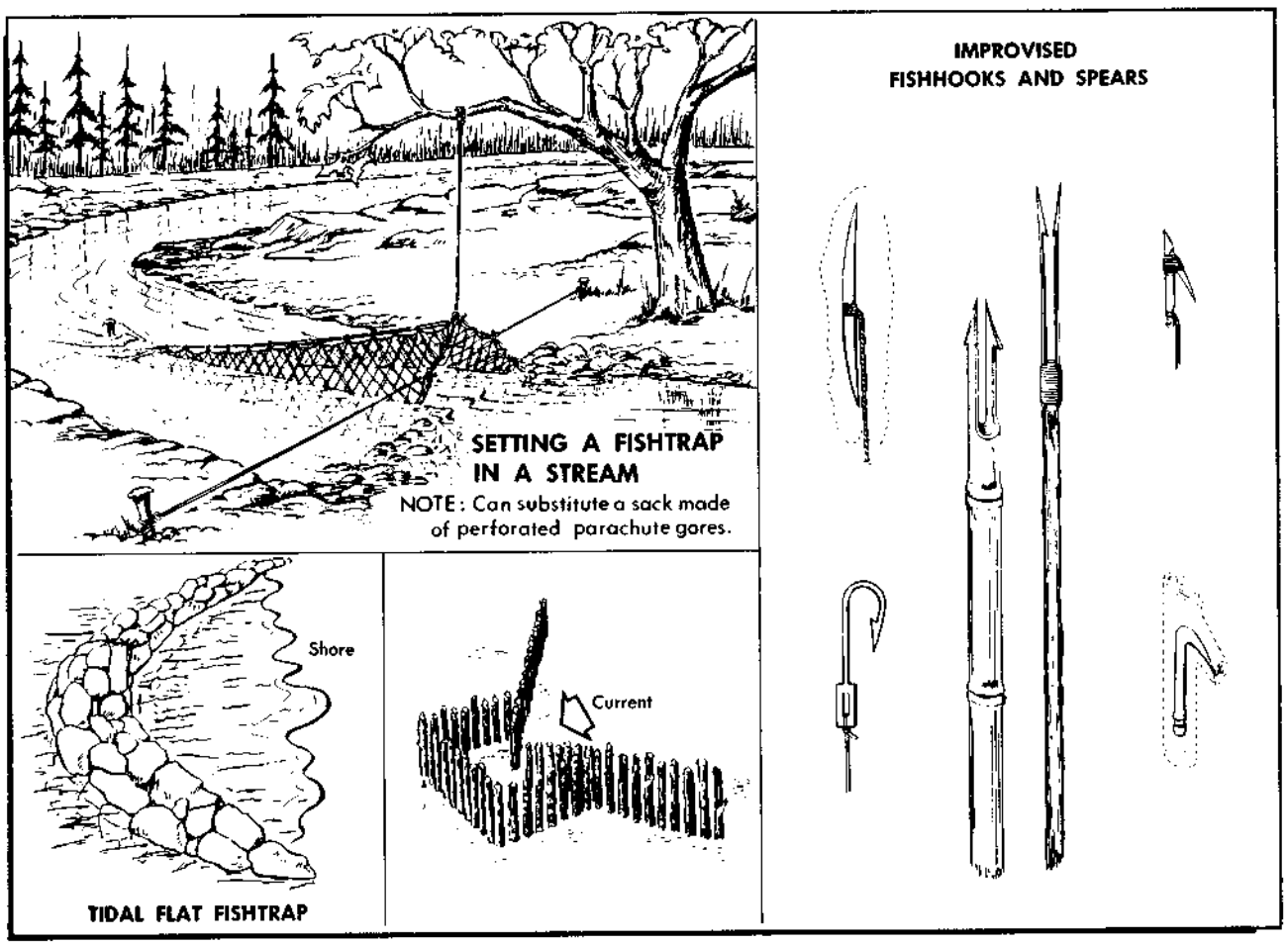

Figure 18-24. Maze-type Fishtraps.

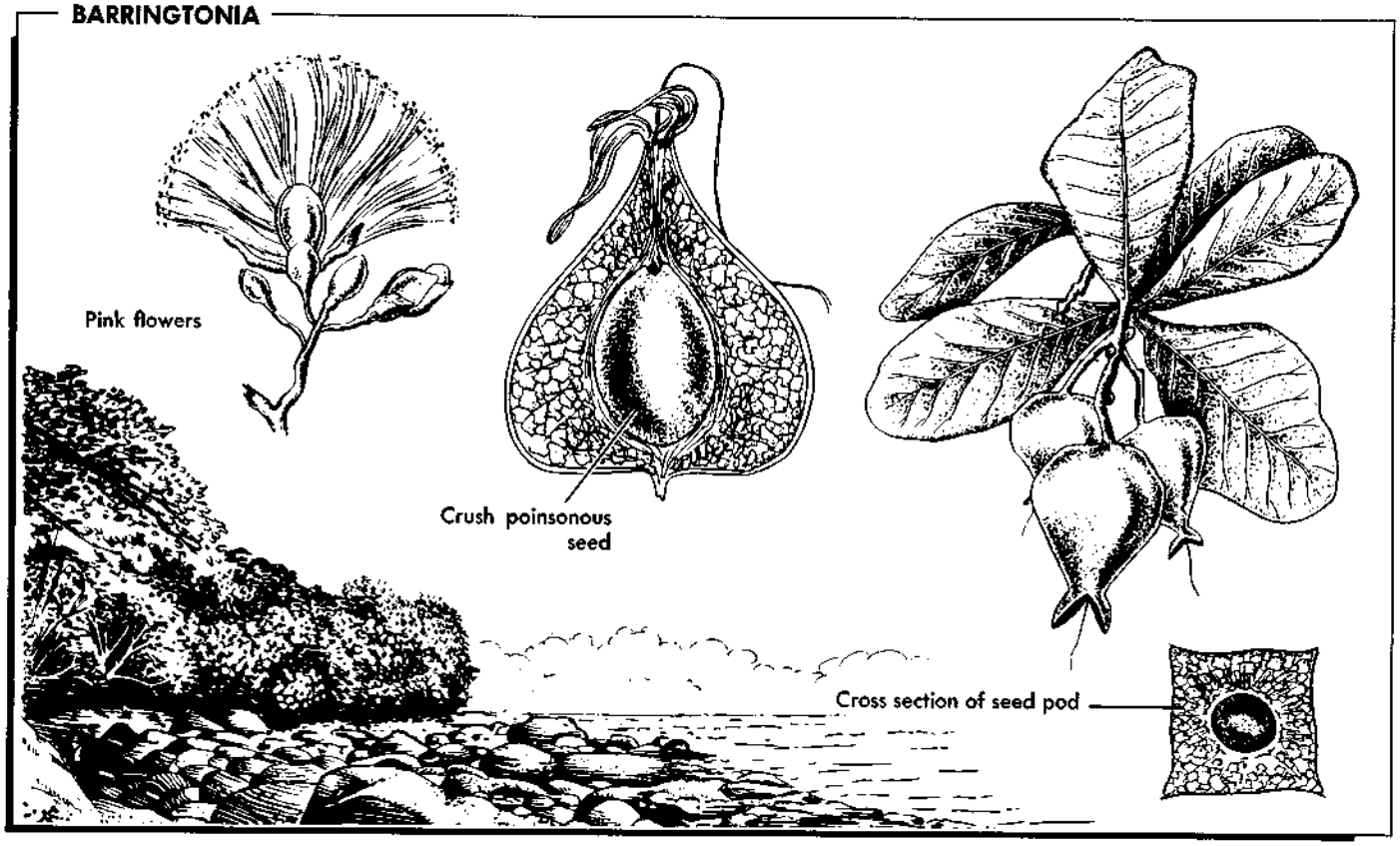

Figure 18-25. Barringtonia Plant for Poisoning Fish.

cy kit. Do not expose it unnecessarily to air or light; it retains its toxicity best if kept in a dark-colored vial. Lime thrown in a small pond or tidal pool will kill fish in the pool. Lime can be obtained by burning coral and seashells.

(c) The most common method of using fish-poison plants is to crush the plant parts (most often the roots) and mix them with water. Drop large quantities of the crushed plant into pools or the headwaters of small streams containing fish. Within a short time, the fish will rise in a helpless state to the surface. After putting in the poison, follow slowly down stream and pick up the fish as they come to the surface, sink to the bottom, or swim crazily to the bank. A stick dam or obstruction will aid in collecting fish as they float downstream. The husk of "green" black walnuts can be crushed and sprinkled into small sluggish streams and pools to act as a fish stupefying agent. In the southwest Pacific, the seeds and bark from the barringtonia tree (figure 18-25) are commonly used as a source of fish poison. The barringtonia tree usually grows along the seashore.

(7) Tickling can be effective in small streams with undercut banks or in shallow ponds left by receding flood waters. Place hands in the water and reach under the bank slowly, keeping the hands close to the bottom if possible. Move the fingers slightly until they make contact with a fish. Then work hands gently along its belly until reaching its gills. Grasp the fish firmly just behind the gills and scoop it onto land. In the tropics, this type of fishing can be dangerous due to hazardous marine life in the water such as piranhas, eels, and snakes.

18-5. Plant Food. The thought of having a diet consisting only of plant food is often distressing to stranded aircrew members. This is not the case if the survival episode is entered into with the confidence and intelligence based on knowledge or experience. If the survivors know what to look for, can identify it, and know how to prepare it properly for eating, there is no reason why they can't find sustenance. In many isolated regions, survivors who have had some previous training in plant identification can enjoy wild plant food.

a. Plants provide carbohydrates, which provide body energy and calories. Carbohydrates keep weight and energy up, and include important starches and sugars.

b. A documented and authoritative example of the value of a strictly plant diet in survival can be cited in the case of a Chinese botanist who had been drafted into the Japanese Army during World War II. Isolated with his company in a remote section of the Philippines, the Chinese botanist kept 60 of his fellow soldiers alive for 16 months by finding wild plants and preparing them properly. He selected six men to assist him, and

then found 25 examples of edible plants in the vicinity of their camp. He acquainted the men with these samples, showing them what parts of the plants could be used for food. He then sent the men out to look for similar plants and had them separate the new plants according to the original examples to avoid any poisonous plant mingling with the edible ones. The result of this effort was impressive. Though all the men had a natural desire for ordinary food, none suffered physically from the plant food diet. The report was especially valuable because the botanist kept a careful record of all the food used, the results, and the comments of the men. This case history reflects the same opinions as those found in questionnaires directed to American survivors during World War II.

c. Another advantage of a plant diet is availability. In many instances, a situation may present itself in which procuring animal food is out of the question because of injury, being unarmed, being in enemy territory, exhaustion, or being in an area which lacks wildlife. If convinced that vegetation can be depended upon for daily food needs, the next question is "where to get what and how."

(1) Experts estimate there are about 300,000 classified plants growing on the surface of the Earth, including many which thrive on mountain tops and on the floors of the oceans. There are two considerations that survivors must keep in mind when procuring plant foods. The first consideration, of course, is the plant be edible, and preferably, palatable. Next, it must be fairly abundant in the areas in which it is found. If it includes an inedible or poisonous variety in its family, the edible plant must be distinguishable to the average eye from the poisonous one. Usually a plant is selected because one special part is edible, such as the stalk, the fruit, or the nut.

(2) To aid in determining plant edibility, there are general rules which should be observed and an edibility test that should be performed. In selecting plant foods, the following should be considered. Select plants resembling those cultivated by people. It is risky to rely upon a plant (or parts thereof) being edible for human consumption simply because animals have been seen eating it (for example, horses eat leaves from poison ivy; some rodents eat poisonous mushroom). Monkeys will put poisonous plants and fruits in pouches of their mouths and spit them out later. When selecting an unknown plant as a possible food source, apply the following general rules:

(a) Mushrooms and fungi should not be selected. Fungi have toxic peptides, a protein-base poison which has no taste. There is no field test other than eating to determine whether an unknown mushroom is edible. Anyone gathering wild mushrooms for eating must be absolutely certain of the identity of every specimen picked. Some species of wild mushrooms are difficult for an expert to identify. Because of the potential for poisoning, relying on mushrooms as a viable food source is not worth the risk.

(b) Plants with umbrella-shaped flowers are to be completely avoided, although carrots, celery, dill, and parsley are members of this family. One of the most poisonous plants, poison water hemlock, is also a member of this family (figure 18-26).

(c) All of the legume family should be avoided (beans and peas). They absorb minerals from the soil and cause problems. The most common mineral absorbed is selenium. Selenium is what has given locoweed its fame. (Locoweed is a vetch.)

(d) As a general rule, all bulbs should be avoided. Examples of poisonous bulbs are tulips and death camas.

(e) White and yellow berries are to be avoided as they are almost always poisonous. Approximately one-half of all red berries are poisonous. Blue or black berries are generally safe for consumption.

(f) Aggregated fruits and berries are always edible (for example, thimbleberry, raspberry, salmonberry, and blackberry).

(g) Single fruits on a stem are generally considered safe to eat.

(h) Plants with shiny leaves are considered to be poisonous and caution should be used.

(i) A milky sap indicates a poisonous plant.

(j) Plants that are irritants to the skin should not be eaten, such as poison ivy.

(k) A plant that grows in sufficient quantity within the local area should be selected to justify the edibility test and provide a lasting source of food if the plant proves edible.

(l) Plants growing in the water or moist soil are often the most palatable.

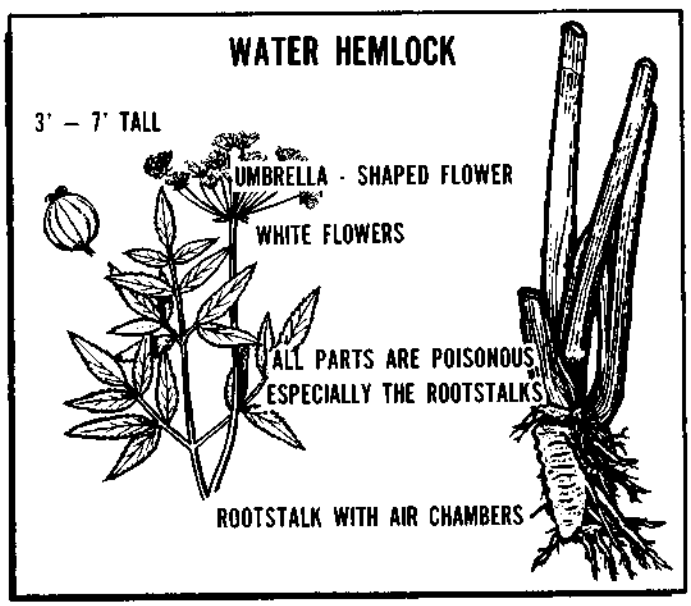

Figure 18-26. Water Hemlock.

(m) Plants are less bitter when growing in shaded areas.

(3) The previously mentioned information concerning plants is general. There are exceptions to every rule, but when selecting unknown plants for consumption, plants with these characteristics should be avoided. Plants that do not have these characteristics should be considered as possible food sources. Apply the edibility test to only one plant at a time so if some abnormality does occur, it will be obvious which plant caused the problem. Once a plant has been selected for the edibility test, proceed as follows:

(a) Crush or break part of the plant to determine the color of its sap. If the sap is clear, proceed to the next step.

(b) Touch the plant's sap or juice to the inner forearm or tip of the tongue. (A small taste of a poisonous plant will not do serious harm.) If there are no ill effects, such as a rash or burning sensation to the skin, bitterness to the taste, or numbing sensation of the tongue or lips, then proceed with the rest of the steps. (NOTE: Sometimes heavy smokers are unable to taste various poisons, such as alkaloids).

(c) Prepare the plant or plant part for consumption by boiling in two changes of water. The toxic properties of many plants are water soluble or destroyed by heat; cooking and discarding in two changes of water lessens the amount of poisonous material or removes it completely. Parboiling is a process of boiling the individual plant parts in repeated changes of water to remove bitter elements. This boiling period should last about 5 minutes.

(d) Place about 1 teaspoonful of the prepared plant food in the mouth for 5 minutes and chew but do not swallow it. A burning, nauseating, or bitter taste is a warning of possible danger. If any of these ill effects occur, remove the material from the mouth at once and discard that plant as a food source. However, if no burning sensation or other unpleasant effect occurs, swallow the plant material and wait 8 hours.

(e) If after this 8 hours there are no ill effects, such as nausea, cramps, or diarrhea, eat about 2 tablespoonfuls and wait an additional 8 hours.

(f) If no ill effects occur at the end of this 8-hour period, the plant may be considered edible.

(g) Keep in mind that any new or strange food should be eaten with restraint until the body system has become accustomed to it. The plant may be slightly toxic and harmful when large quantities are eaten.

(4) If cooking facilities are not available, survivors will not be able to boil the plant before consumption. In this case, plant food may be prepared as follows:

(a) Leach the plant by crushing the plant material and placing it in a container. Pour large quantities of cold water over it (rinse the plant parts). Leaching removes some of the bitter elements of nontoxic plants.

EDIBLE PARTS OF PLANTS	
Underground Parts	Tubers Roots and Rootstalks Bulbs
Stems and Leaves (potherbs)	Shoots and Stems Leaves Pith Bark
Flower Parts	Flowers Pollen
Fruits	Fleshy Fruits (dessert and vegetable) Seeds and Grains Nuts Seed Pods Pulps
Gums and Resins	
Saps	

Figure 18-27. Edible Parts of Plants.

(b) If leaching is not possible, survivors should follow the steps they can in the edibility test.

d. The survivor will find some plants which are completely edible, but many plants which they may find will have only one or more identifiable parts having food and thirst-quenching value. The variety of plant component parts which might contain substance of food value is shown in figure 18-27.

(1) Underground Parts:

(a) Tubers. The potato is an example of an edible tuber. Many other kinds of plants produce tubers such as the tropical yam, the Eskimo potato, and tropical water lilies. Tubers are usually found below the ground. Tubers are rich in starch and should be cooked by roasting in an earth oven or by boiling to break down the starch for ease in digestion. The following are some of the plants with edible tubers.

-1.Arrowroot, East Indian.
-2.Taro.
-3.Cassava (Tapioca).
-4.Bean, Yam.
-5.Chufa (Nut Grass).
-6.Water Lily (Tropical).
-7.Sweet Potato (Kamote).
-8.Yam Tropical.

(b) Roots and Rootstalks. Many plants produce roots which may be eaten. Edible roots are often several feet in length. In comparison, edible rootstalks are underground portions of the plant which have become thickened, and are relatively short and jointed. Both true roots and rootstalks are storage organs rich in

stored starch. The following are some of the plants with edible roots or rootstalks (rhizomes):

-1.Baobab.
-2.Pine, Screw.
-3.Bean, Goa.
-4.Plantain, Water.
-5.Bracken.
-6.Reindeer Moss.
-7.Calla, Wild (Water Arum).
-8.Rock Tripe.
-9.Pollypody.
-10.Canna Lily.
-11.Rush, Flowering.
-12.Cattail.
-13.Spinach, Ceylon.
-14.Chicory.
-15.Ti Plant.
-16.Horseradish.
-17.Tree Fern.
-18.Lotus Lily.
-19.Water Lily (Temperate Zone).
-20.Manioc.

(c) Bulbs. The most common edible bulb is the wild onion, which can easily be detected by its characteristic odor. Wild onions may be eaten uncooked, but other kinds of bulbs are more palatable if cooked. In Turkey and Central Asia, the bulb of the wild tulip may be eaten. All bulbs contain a high percentage of starch. (Some bulbs are poisonous, such as the death camas which has white or yellow flowers.) The following are some of the plants with edible bulbs:

-1.Lily, Wild.
-2.Tulip, Wild.
-3.Onion, Wild.
-4.Blue Camas.
-5.Tiger Lily.

(2) Shoots and Leaves:

(a) Shoots (Stems). All edible shoots grow in much the same fashion as asparagus. The young shoots of ferns (fiddleheads) and especially those of bamboo and numerous kinds of palms are desirable for food. Some kinds of shoots may be eaten raw, but most are better if first boiled for 5 to 10 minutes, the water drained off, and the shoots reboiled until they are sufficiently cooked for eating (parboiled). (See figure 18-28).

-1.Agave (Century Plant).
-2.Palm, Coconut.
-3.Purslane.
-4.Reindeer Moss.
-5.Bamboo.
-6.Palm, Fishtail.
-7.Bean, Goa.
-8.Palm, Nipa.
-9.Bracken.
-10.Palm, Rattan.
-11.Rhubarb, Wild.
-12.Cattail.
-13.Palm, Sago.
-14.Spinach, Ceylon.
-15.Rock Tripe.
-16.Colocynth.
-17.Palm, Sugar.
-18.Papaya.
-19.Sugar Cane.
-20.Lotus Lily.
-21.Pokeweed (poisonous roots).
-22.Sweet Potato-Kamote.
-23.Luffa Sponge.
-24.Water Lily (Tropical).
-25.Polypody.
-26.Palm, Buri.
-27.Willow, Arctic.

(b) Leaves. The leaves of spinach-type plants (potherbs), such as wild mustard, wild lettuce, and lamb quarters, may be eaten either raw or cooked. Prolonged cooking, however, destroys most of the vitamins. Plants which produce edible leaves are perhaps the most numerous of all edible plants. The young tender leaves of nearly all nonpoisonous plants are edible. The follow-

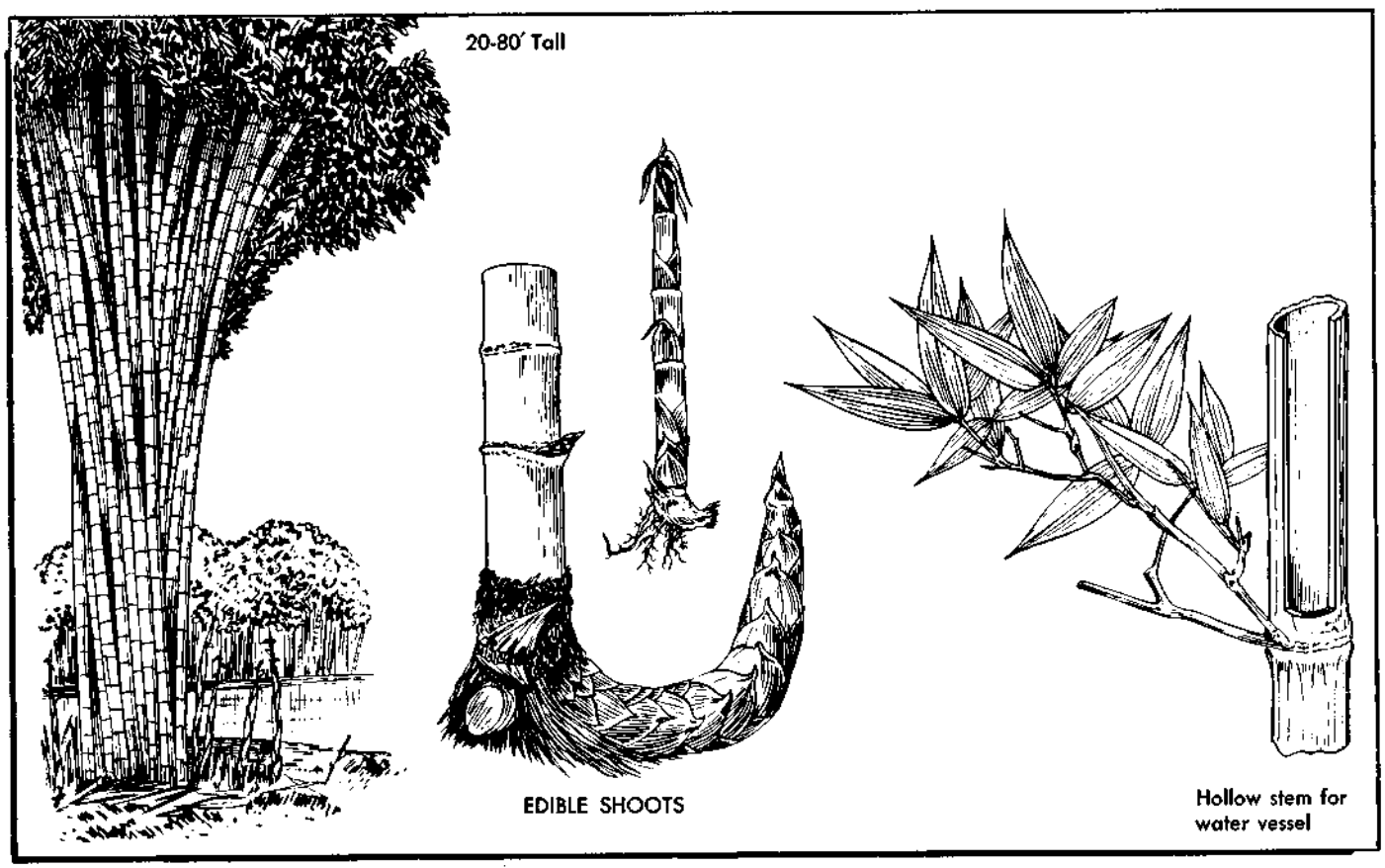

Figure 18-28. Bamboo.

ing are only some of the plants with edible leaves:

-1.Amarath.
-2.Luffa Sponge.
-3.Rock Tripe.
-4.Avocado.
-5.Mango.
-6.Sorrel, Wild.
-7.Baobab.
-8.Orach, Sea.
-9.Bean, Goa.
-10.Papaya.
-11.Spinach, Ceylon.
-12.Cassava.
-13.Chickory.
-14.Pine, Screw.
-15.Spreading Wood Fern.
-16.Dock.
-17.Plantain.
-18.Pokeweed (poisonous roots).
-19.Sweet Potato-Kamote.
-20.Tamarind.
-21.Horseradish.
-22.Prickly Pear.
-23.Taro (only after cooking).
-24.Lettuce, Water.
-25.Purslane.
-26.Ti Plant.
-27.Willow, Arctic.
-28.Lotus Lily.
-29.Reindeer Moss.

(c) Pith. Some plants have an edible pith in the center of the stem. The pith of some kinds of tropical plants is quite large. Pith of the sago palm is particularly valuable because of its high food value. The following are some of the palms with edible pith (starch):

-1.Buri.
-2.Fishtail.
-3.Sago.
-4.Coconut.
-5.Rattan.
-6.Sugar.

(d) Bark. The inner bark of a tree—the layer next to the wood—may be eaten raw or cooked. It is possible in northern areas to make flour from the inner bark of such trees as the cottonwood, aspen, birch, willow, and pine. The outer bark should be avoided in all cases because this part contains large amounts of bitter tannin. Pine bark is high in vitamin C. The outer bark of pines can be cut away and the inner bark stripped from the trunk and eaten fresh, dried, or cooked, or it may be pulverized into flour. Bark is most palatable when newly formed in spring. As food, bark is most useful in the arctic regions, where plant food is often scarce.

(3) Flower Parts:

(a) Flowers and Buds. Fresh flowers may be eaten as part of a salad or to supplement a stew. The hibiscus flower is commonly eaten throughout the southwest Pacific area. In South America, the people of the Andes eat nasturtium flowers. In India, it is common to eat the flowers of many kinds of plants as part of a vegetable curry. Flowers of desert plants may also be eaten. The

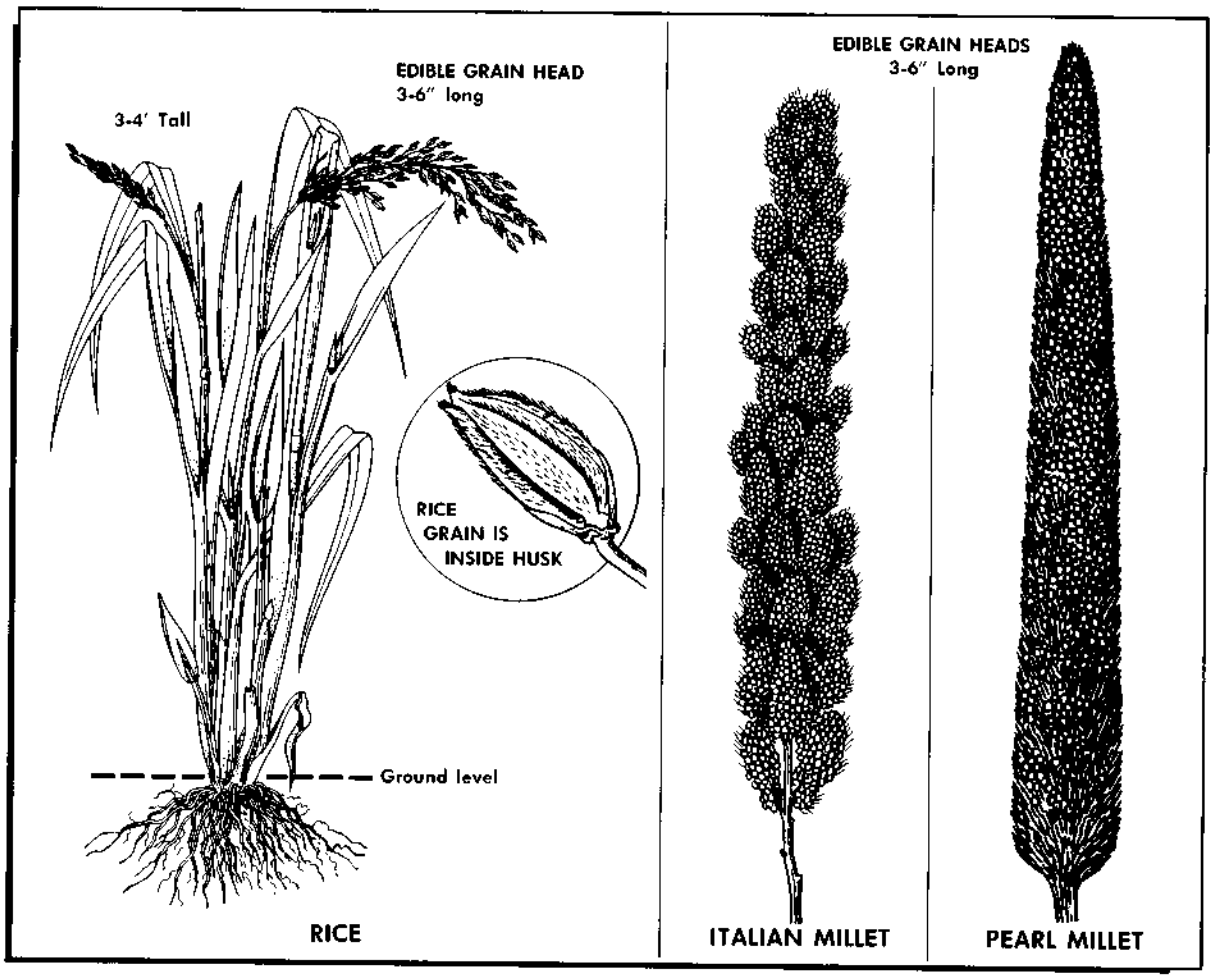

Figure 18-29. Grains.

following are plants with edible flowers:

-1.Abal.
-2.Colocynth.
-3.Papaya.
-4.Banana.
-5.Horseradish.
-6.Caper, Wild.
-7.Luffa Sponge.

(b) Pollen. Pollen looks like yellow dust. All pollen is high in food value and in some plants, especially the cattail. Quantities of pollen may easily be collected and eaten as a kind of gruel.

(4) Fruits. Edible fruits can be divided into sweet and nonsweet (vegetable) types. Both are the seed bearing parts of the plant. Sweet fruits are often plentiful in all areas of the world where plants grow. For instance, in the far north, there are blueberries and crowberries; in the temperate zones, cherries, plums, and apples; and in the American deserts, fleshy cactus fruits. Tropical areas have more kinds of edible fruit than other areas, and a list would be endless. Sweet fruits may be cooked, or for maximum vitamin content, left uncooked. Common vegetable fruits include the tomato, cucumber, and pepper.

(a) Fleshy Fruits (Sweet). The following are plants with edible fruits:

-1.Apple, Wild.
-2.Bael Fruit.
-3.Banana.
-4.Bignay.
-5.Blueberry, Wild.
-6.Bullocks Heart.
-7.Cloudberry.
-8.Crabapple.
-9.Cranberry.
-10.Fig, Wild.
-11.Grape, Wild.
-12.Huckleberry.
-13.Jackfruit.
-14.Jujube, Common.
-15.Mango.
-16.Mulberry.
-17.Papaya.
-18.Plum, Batako.
-19.Pokeberry.
-20.Prickly Pear.
-21.Rose Apple.
-22.Soursop.
-23.Sweetsop.

(b) Fleshy Fruits (Vegetables). The following are plants with edible fruits (vegetables):

-1.Breadfruit.
-2.Horseradish.
-3.Plantain.
-4.Caper, Wild.
-5.Luffa Sponge.

(c) Seeds and Grains. Seeds of many plants, such as buckwheat, ragweed, amaranth, and goosefoot, contain oils and are rich in protein. The grains of all cereals and many other grasses, including millet, are also extremely valuable sources of plant protein. They may either be ground between stones, mixed with water and cooked to make porridge, parched or roasted over hot stones. In this state, they are still wholesome and may be kept for long periods without further preparation (figure 18-29). The following are some of the plants with edible seeds and grains:

-1.Amaranth.
-2.Millet, Italian.
-3.Rice.
-4.Bamboo.
-5.Millet, Pearl.
-6.Palm, Nipa.
-7.Tamarind.
-8.Pine, Screw.
-9.Coloynth.
-10.Water Lily (Tropical).
-11.Sterculia.
-12.Baobab.
-13.Orach, Sea.
-14.St. John's Bread.
-15.Bean, Goa.
-16.Lotus Lily.
-17.Purslane.
-18.Water Lily (Temperate).
-19.Luffa Sponge.

(d) Nuts. Nuts are among the most nutritious of all raw plant foods and contain an abundance of valuable protein. Plants bearing edible nuts occur in all the climatic zones of the world and in all continents except in the arctic regions. Inhabitants of the temperate zones are familiar with walnuts, filberts, almonds, hickory nuts, acorns, hazelnuts, beechnuts, and pine nuts, to mention just a few. Tropical zones produce coconuts and other palm nuts, brazil nuts, cashew nuts, and macadamia nuts (figure 18-30). Most nuts can be eaten raw but some such as acorns, are better when cooked. The following are some of the plants with edible nuts:

-1.Almond.
-2.Chestnut, Water (Trapa Nut).
-3.Palm, Buri
-4.Almond, Indian or Tropical.
-5.Chestnut, Mountain.
-6.Palm, Coconut.
-7.Beechnut.
-8.Filbert (Hazelnut).
-9.Palm, Fishtail.
-10.Jackfruit Seeds.
-11.Oak, English (Acorn).
-12.Palm, Sago.
-13.Palm, Sugar.
-14.Pine.
-15.Pistachio, Wild.
-16.Walnut.

(e) Pulps. The pulp around the seeds of many fruits is the only part that can be eaten. Some fruits produce sweet pulp; others have a tasteless or even bitter pulp. Plants that produce edible pulp include the custard apple, inga pod, breadfruit, and tamarind. The pulp of breadfruit must be cooked, whereas in other plants, the pulp may be eaten uncooked. Use the edibility rules in all cases of doubt.

(5) Gums and Resins. Gum and resin are sap that collects and hardens on the outside surface of the plant. It is called gum if it is soft and soluble, and resin if it is hard and not soluble. Most people are familiar with the gum which exudes from cherry trees and the resin which seeps from the pine trees. These plant byproducts are edible and are a good source of nutritious food which should not be overlooked.

(6) Saps. Vines or other plant parts may be tapped as potential sources of usable liquid. The liquid is obtained by cutting the flower stalk and letting the fluid drain into some sort of container such as a bamboo section. Palm sap with its high-sugar content is highly nutritious. The following are some plants with edible sap and drinking water:

(a) Acacia, Sweet (water).
(b) Colocynth (water).
(c) Palm, Coconut (sap).
(d) Palm, Fishtail (sap).
(e) Agave (water).
(f) Cuipo Tree (water).
(g) Saxual (water).
(h) Palm, Nipa (sap).
(i) Palm, Rattan (water).
(j) Cactus (water).
(k) Grape (water).
(l) Banana (water).
(m) Palm, Sago (sap).
(n) Palm, Sugar (sap).
(o) Palm, Buri (sap).

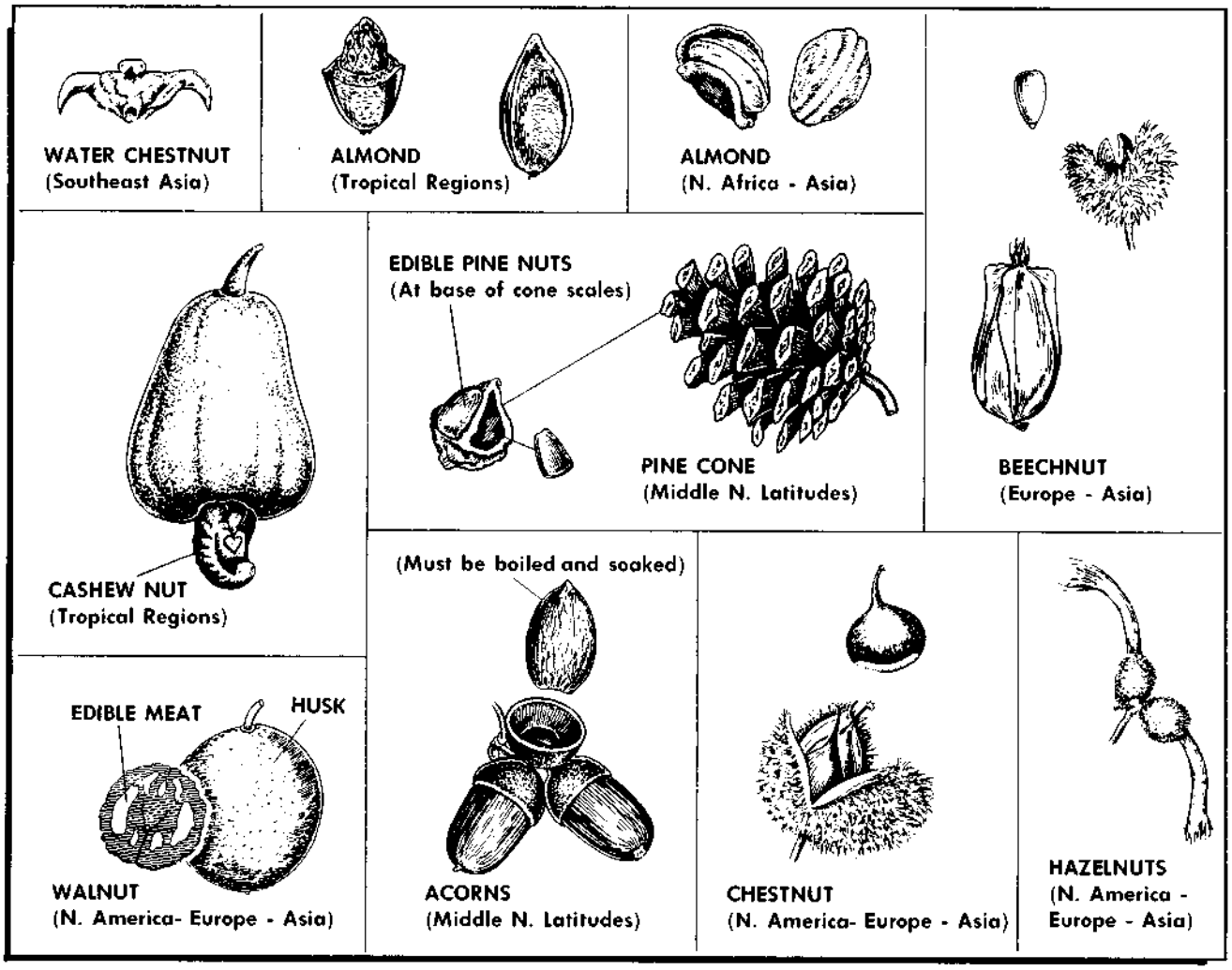

Figure 18-30. Edible Nuts.

18-6. Food in Tropical Climate. There are more types of animals in the jungles of the world than in any other region. A jungle visitor who is unaware of the life style and eating habits of these animals would not observe the presence of a large number of the animals.

a. Game trails are the normal routes along which animals travel through a jungle. Some of the animals used as food are hedgehogs, porcupines, anteaters, mice, wild pigs, deer, wild cattle, bats, squirrels, rats, monkeys, snakes, and lizards.

(1) Reptiles are located in all jungles and should not be overlooked as a food source. All snakes should be considered poisonous and extreme caution used when killing the animal for a food source. All cobras should be avoided since the spitting cobra aims for the eyes; the venom can blind if not washed out immediately. Lizards are good food, but may be difficult to capture since they can be extremely fast. A good blow to the head of a reptile will usually kill it. Crocodiles and caimans are extremely dangerous on land as well as in the water.

(2) Frogs can be poisonous; all brilliantly colored frogs should be totally avoided. Some frogs and toads in the tropics secrete substances through the skin which has a pungent odor. These frogs are often poisonous.

(3) The larger, more dangerous animals such as tigers, rhinocerous, water buffalo, and elephants are rarely seen and should be left alone. These larger animals are usually located in the open grasslands.

b. Seafood such as fish, crabs, lobsters, crayfish, and small octopi can be poked out of holes, crevices, or rock pools (figure 18-31). Survivors should be ready to spear them before they move off into deep water. If they are in deeper water, they can be teased shoreward with a baited hook, or a stick.

(1) A small heap of empty oysters shells near a hole may indicate the presence of an octopus. A baited hook placed in the hole will often catch the octopus. The survivor should allow the octopus to surround the hook and line before lifting. Octopi are not scavengers like sharks, but they are hunters, fond of spiny lobster and other crab-like fish. At night, they come into shallow water and can be easily seen and speared.

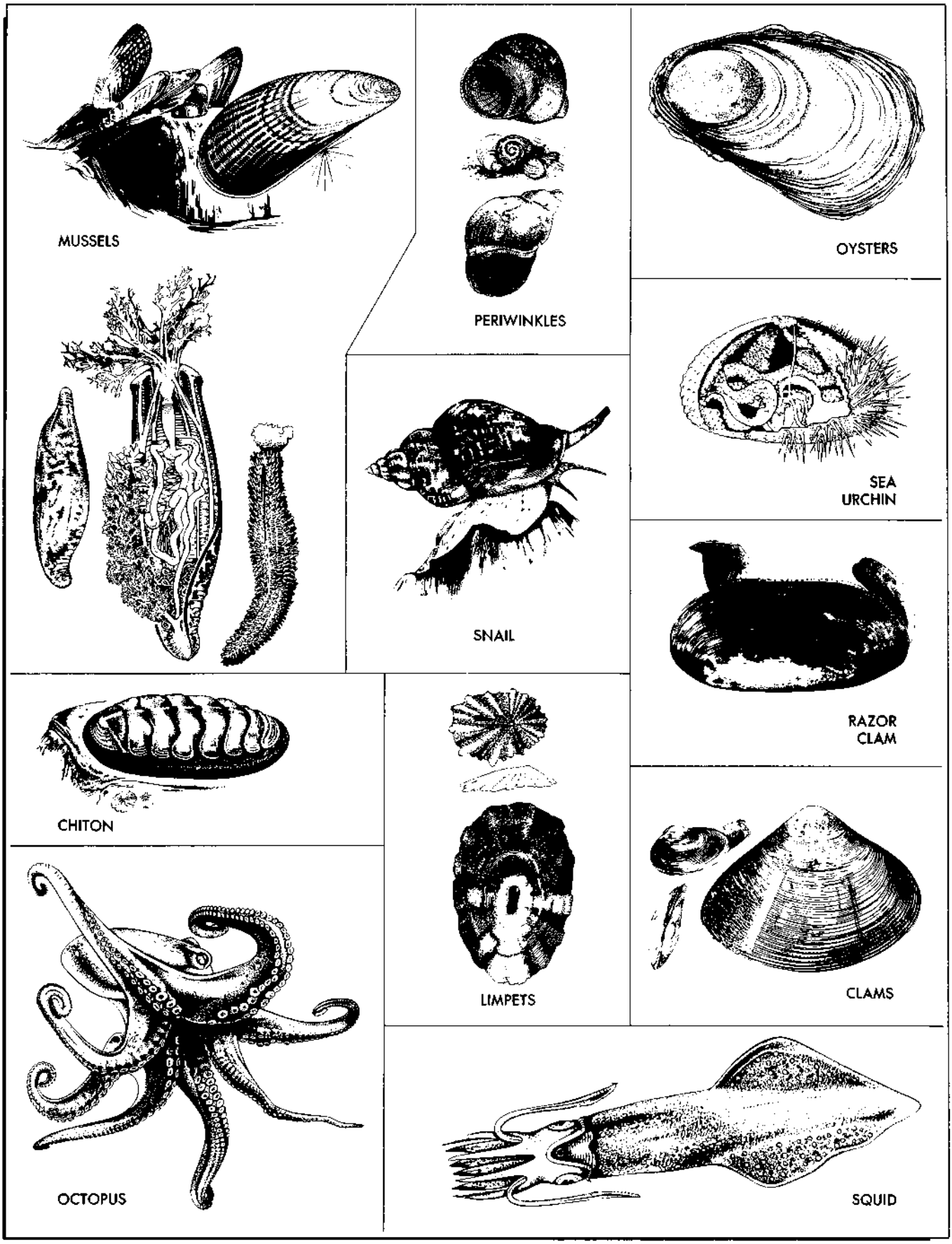

Figure 18-31. Edible Invertebrates.

(2) Snails and limpets cling to rocks and seaweed from the low-water mark up. Large snails called chitons adhere tightly to rocks just above the surf line.

(3) Mussels usually form dense colonies in rock pools, on logs, or at the bases of boulders. *Mussels are poisonous in tropical zones during the summer,* especially when seas are highly phosphorescent or reddish.

(4) Sluggish sea cucumbers and conchs (large snails) live in deep water. The sea cucumber will shoot out its stomach when excited. The stomach is not edible. The skin and the five strips of muscle can be eaten after boiling. Conches can be boiled out of their shells and have very firm flesh. Use care when picking conches up. The bottom of their "foot" has a boney covering which can severely cut the survivor who procures it.

(5) The safest fish to eat are those from the open sea or deep water beyond the reef. Silvery fishes, river eels, butterfly fishes, and flounders from bays and rivers are good to eat.

(6) Land crabs are common on tropical islands and are often found in coconut groves. An open coconut can be used for bait.

(7) A number of methods can be used for procuring fish.

(a) Hook-and-Line Fishing. This type of fishing on a rocky coast requires a lot of care to keep the line from becoming entangled or cut on sharp edges. Most shallow-water fish are nibblers. Unless the bait is well placed and hooked and the barb of the hook offset by bending, the bait may be lost without catching a fish. Use hermit crabs, snails, or the tough muscle of a shellfish as bait. Take the cracked shells and any other animal remains and drop them into the area to be fished. This brings the fish to the area and provides a better procurement opportunity. Examine stomach contents of the first fish caught to determine what the fish are feeding on.

(b) Jigging. A baited or spooned hook dipped repeatedly beneath the surface of the water is sometimes effective. This method may be used at night.

(c) Spearing. This method is difficult except when the stream is small and the fish are large and numerous during the spawning season, or when the fish congregate in pools. Make a spear by sharpening a long piece of wood, lashing two long thorns on a stick, or fashioning a bone spear point, and take a position on a rock over a fish run. Wait patiently and quietly for a fish to swim by.

(d) Chop Fishing. Chop fishing is effective at night during low tide. This method requires a torch and a machete. The fish are attracted by the light of the torch, and then they may be stunned by slashing at them with the *back* of the machete blade. Care should be taken when swinging the machete (figure 18-32).

c. The jungle environment has a uniquely favorable condition for plant and animal life. The variety and richness of plant growth in these areas are paralleled nowhere else on the Earth. Because the rainfall is distributed throughout the year and there is a lack of cold seasons, plants in the humid regions can grow, produce leaves, and flower the year round. Some plants grow very rapidly. For example, the stem of the giant bamboo may grow more than 22 inches in a single day.

(1) A survivor in search of plant food should apply some basic principles to the search. A survivor is lucky to find a plant that can readily be identified as edible. If a plant resembles a known plant, it is very likely to be of the same family and can be used. If a plant cannot be identified, the edibility test should be applied. A survivor will find many edible plants in the tropical forest, but chances of finding them in abundance are better in an area that has been cultivated in the past (secondary growth).

(2) Some plants a survivor might find:

(a) Citrus fruit trees may be found in uncultivated areas, but are primarily limited to areas of secondary growth. The many varieties of citrus fruit trees and shrubs have leaves 2 to 4 inches long alternately arranged. The leaves are leathery, shiny, and evergreen. The leaf stem is often winged. Small (usually green) spines are often present by the side of the bud. The flowers are small and white to purple in color. The fruit has a leathery rind with numerous glands and is round and fleshy with several cells (fruit sections or slices) and many seeds. The great number of wild and cultivated fruits (oranges, limes, lemons, etc.) native to the tropics are eaten raw or used in beverages.

(b) Taro can be found in both secondary growth and in virgin areas. It is usually found in the damp, swampy areas in the wild, but certain varieties can be found in the forest. It can be identified by its large heart-shaped or arrowhead-shaped leaves growing at the top of a vertical stem. The stem and leaves are usually green and rise a foot or more from a tuber at the base of the stem. Taro leaf tips point down; poisonous elephant ear points up. All varieties of taro must be cooked to break down the irritating crystals in the plant.

(c) Wild pineapple can be found in the wild, and common pineapples may be found in secondary growth areas. The wild pineapple is a coarse plant with long clustered, sword-shaped leaves with sawtoothed edges. The leaves are spirally arranged in a rosette. Flowers are violet or reddish. The wild pineapple fruit will not be as fully developed in the wild state as when cultivated. The seeds from the flower of the plant are edible as well as the fruit. The ripe fruit may be eaten raw, but the green fruit must be cooked to avoid irritation. (The leaf fibers make excellent lashing material and ropes can be manufactured from it.)

(d) Yams may be found cultivated or wild. There are many varieties of yam, but the most common has a vine with square-shaped cross section and two rows of heart-shaped leaves growing on opposite sides of the vine. The vine can be followed to the ground to locate

Figure 18-32. Chop Fishing.

the tuber. The tubers should be cooked to destroy the poisonous properties of the plant (figure 18-33).

(e) Ginger grows in the tropical forest and is a good source of flavoring for food. It is found in shaded areas of the primary forest. The ginger plant grows 5 to 6 feet high. It has seasonal white snapdragon-type flowers, some variations have red flowers. The leaves when crushed produce a very sweet odor and are used for seasoning or tea. The tea is used by primitive people to treat colds and fever.

(f) The coconut palm is found wild on the seacoast and in farmed areas inland. It is a tree 50 to 100 feet high, either straight or curved, marked with ring-like leaf scars. The base of the tree is swollen and surrounded by a mass of rootlets. The leaves are leathery and reach a length of 15 to 20 feet. (The leaves make excellent sheathing for shelter.) The fruit grows in clusters at the top of the tree. Each nut is covered with a fibered hard shell. The "heart" of the coconut palm is edible and is found at the top. (The new leaves grow out of the heart.) Cut the tree down and remove the leaves to gain access to the heart. The flower of the coconut tree is also edible and is best used as a cooked vegetable. The germinating nut is filled with a meat that can be eaten raw or cooked. There are many other varieties of palm found in the tropics which have edible hearts and fruits (figure 18-34).

(g) The papaya is an excellent source of food and can be found in secondary growth areas. The tree grows to a height of 6 to 20 feet. The large, dark green, many fingered, rough-edged leaves are clustered at the top of the plant. The fruit grows on the stem clustered under the leaves. The fruit is small in the wild state, but cultivated varieties may grow to 15 pounds. The peeled fruit can be eaten raw or cooked. The peeling should never be eaten. The green fruit is usually cooked. The milky sap of the green fruit is used as a meat tenderizer; care should be taken not to get it in the eyes. Always wash the hands after handling fresh green papayas. If some of the sap does get in the eyes, they should be washed immediately (figure 18-35).

(h) Cassava (tapioca) can be found in secondary growth areas. It can be identified by its stalk-like leaves which are deeply divided into numerous pointed sections or fingers. The woody (red) stem of the plant is slender and at points appears to be sectioned. When found growing wild in secondary growth areas, pull the trunks to find where a root grows. When one is found, a tuber can be dug. Tubers have been found growing around a portion of the stem that was covered with vegetation. The brown tuber of the plant is white inside and must be boiled or roasted. The tuber must also be peeled before boiling. (The green-stemmed species of cassava is poisonous and must be cooked in several changes of water before eating it.)

(i) Ferns can be found in the virgin tropical forest or in secondary growth areas. The new leaves (fiddle heads) at the top are the edible parts. They are covered with fuzzy hair which is easily removed by rubbing or washing. Some can be eaten raw, but as a rule, should be cooked as a vegetable (figure 18-36).

(j) Sweet sops can be found in the tropical forest. It is a small tree with simple, oblong leaves. The fruit is shaped like a blunt pine cone with thick grey-green or yellow, brittle spines. The fruit is easily split or broken when ripe, exposing numerous dark brown seeds imbedded in the cream colored, very sweet pulp.

(k) The star apple is common in the tropical forests. The tree grows up to a height of 60 feet and can be identified by the leaves which have shiny, silky, brown hairs on the bottom. The fruit looks like a small apple or plum with a smooth greenish or purple skin. The meat is greenish in color and milky in texture. When cut through the center, the brown, elongated seeds make a figure like a 6- or 10-pointed star. The fruit is sweet and eaten only when fresh. When cut, the rind will, like other parts of the tree, emit a white sticky juice or latex which is not poisonous (an exception to the milky sap rule).

(3) Of the 300,000 different kinds of wild plants in the world, a large number of them are found in the

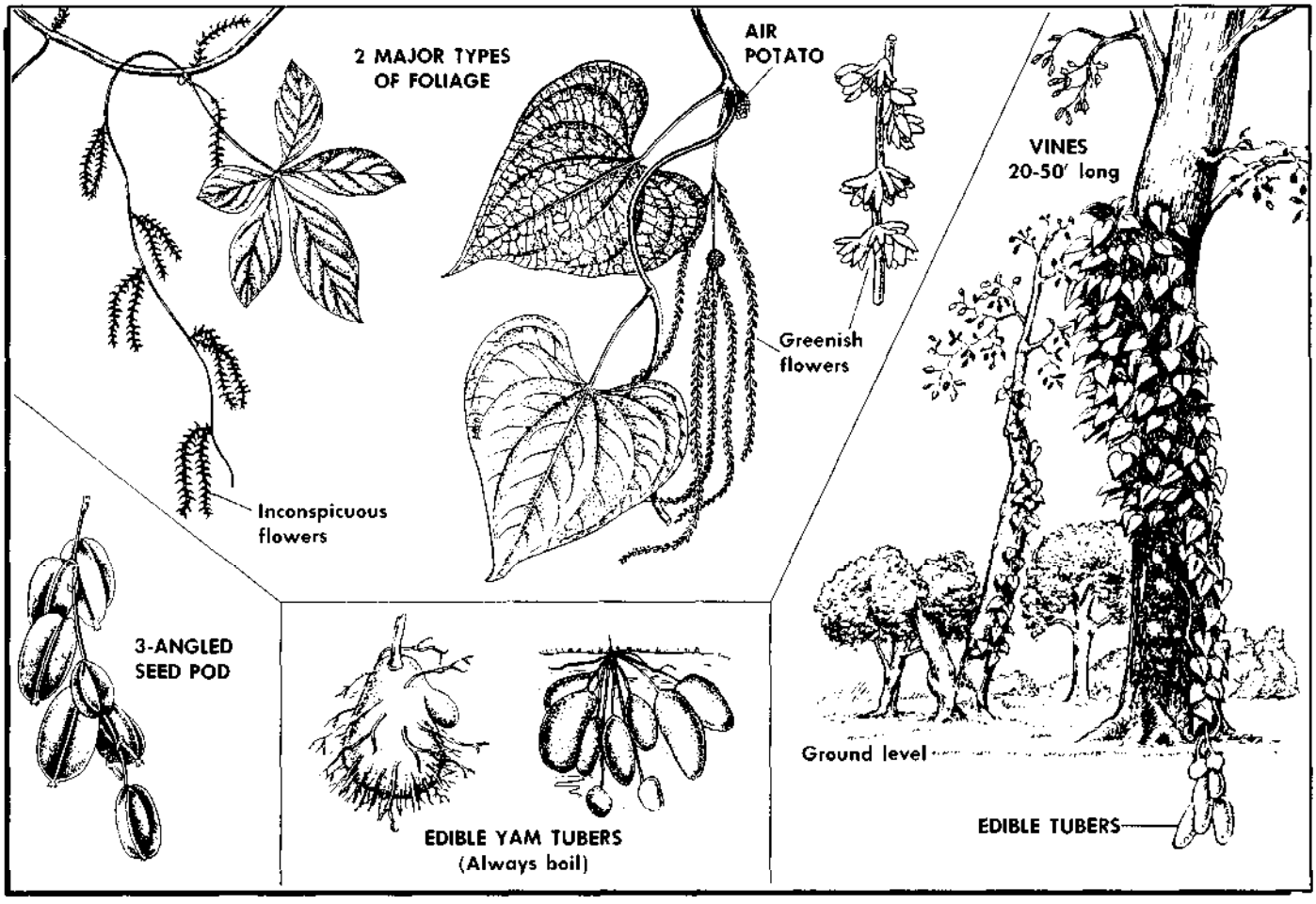

Figure 18-33. Yams.

tropics and many of them are potentially edible. Very few are deadly when eaten in small quantities. Those which are poisonous may be detected by using the edibility rules. Only a small number of jungle plants have been discussed. It would be of great benefit to anyone flying over or passing through a tropical environment to study the plant foods available in this type of environment.

18-7. Food in Dry Climates. Although not as readily available as in the tropical climate, food is available and obtainable.

a. Plant life in the desert is varied due to the different geographical areas. It must be remembered, therefore, that available plants will depend on the actual desert, the time of year, and if there has been any recent rainfall. The aircrew member should be familiar with plants in the area to be flown over.

(1) Date palms are located in most deserts and are cultivated by the native people around oases and irrigation ditches. They bear a nutritious, oblong, black fruit (when ripe).

(2) Fig trees are normally located in tropical and subtropical zones; however, a few species can be found in the deserts of Syria and Europe. Many kinds are cultivated. The fruit can be eaten when ripe. Most figs resemble a top or a small pear somewhat squashed in shape. Ripe figs vary greatly as to palatability. Many are hard, woody, covered with irritating hairs, and worthless as survival food. The edible varieties are soft, delectable, and almost hairless. They are green, red, or black when ripe.

(3) Millet, a grain bearing plant, is grown by natives around oases and other water sources in the Middle East deserts.

(4) The fruit of all cacti are edible. Some fruits are red, some yellow, but all are soft when ripe. Any of the flat leaf variety, such as the prickly pear, can be boiled and eaten as greens (like spinach) if the spines are first removed. During severe droughts, cattlemen burn off the spines and use the thick leaves for fodder. Although the cactus originates in the American deserts, the prickly pear has been introduced to the desert edges in Asia,

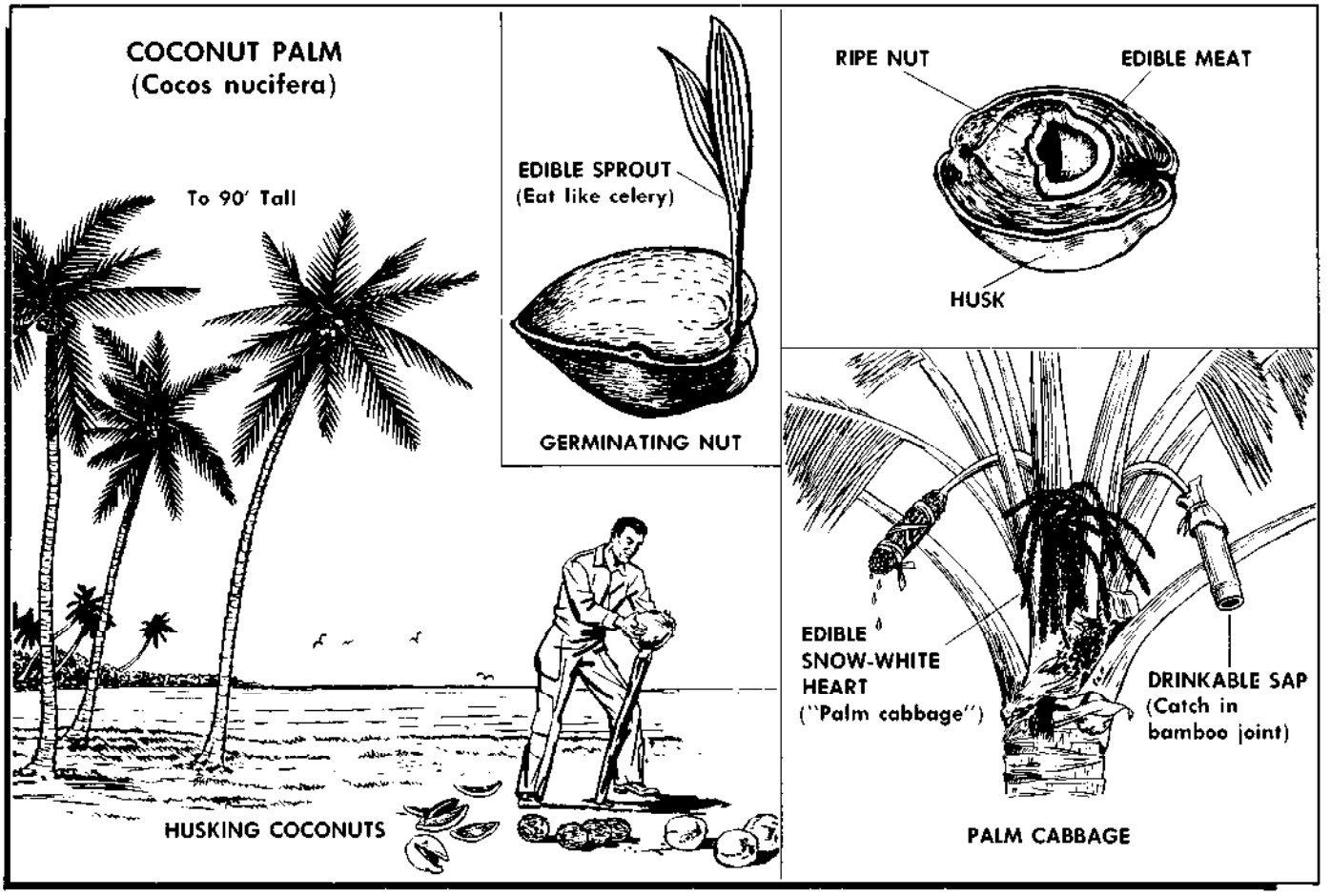

Figure 18-34. Coconut Palm.

Africa, the Near East, and Australia, where it grows profusely. Natives eat the fruit as fast as it ripens.

(5) There are two types of onions in the Gobi desert. A hot, strong, scallion-type grows in the late summer. It will improve the taste of food, but should not be used as a primary food. The highland onions grow 2 to 2.5 inches in diameter. These can be eaten like apples and the greens can also be eaten raw or cooked.

(6) All desert flowers can be eaten except those with milky or colored sap.

(7) All grasses are edible. Usually the best part is the whitish tender end that shows when the grass stalk is pulled from the ground. All grass seeds are edible.

b. Animal food sources may be used to supplement diets and provide needed protein and fats. When looking at a desert area, it is sometimes difficult to visualize an abundance of animal life existing in it. There is, however, a great quantity of animal life present. Most are edible, but some may be hazardous to a survivor during the procurement stage. Some of the abundant animal life includes:

(1) At the peak of seasonal plant growth, the desert crawls and buzzes with an enormous number and variety of beetles, ants, wasps, moths, and bugs. They appear with the first good rains and generally feed during nighttime. The Ute Indians of North America have harvested crickets, and peoples of the Middle East have roasted locusts. The human diet in Mexico and the American Indians of the Southwest frequently includes grasshoppers and caterpillars.

(2) On the Playas of the Sonora and Chichuahua deserts, several species of freshwater shrimp appear every summer in warm temporary ponds. In the Mohave Desert, where summer rains are rare, they may appear only a few times in a century.

(3) Snakes, lizards, tortoises, etc., have adapted well to the desert environment. Care must be observed when procuring them as some are hazardous, such as the Gila monster and rattlesnake. The desert tortoise, about a foot long when full grown, lives in some of the harshest regions of the Mohave and Sonora deserts. It is club footed, herbivorous, and can crawl about 20 feet per minute. The tortoise converts some of its food into water which is stored for the hot months in two sacs under the upper shell. A pint of water lasts the dry season. In spring and fall, the tortoise browses in broad

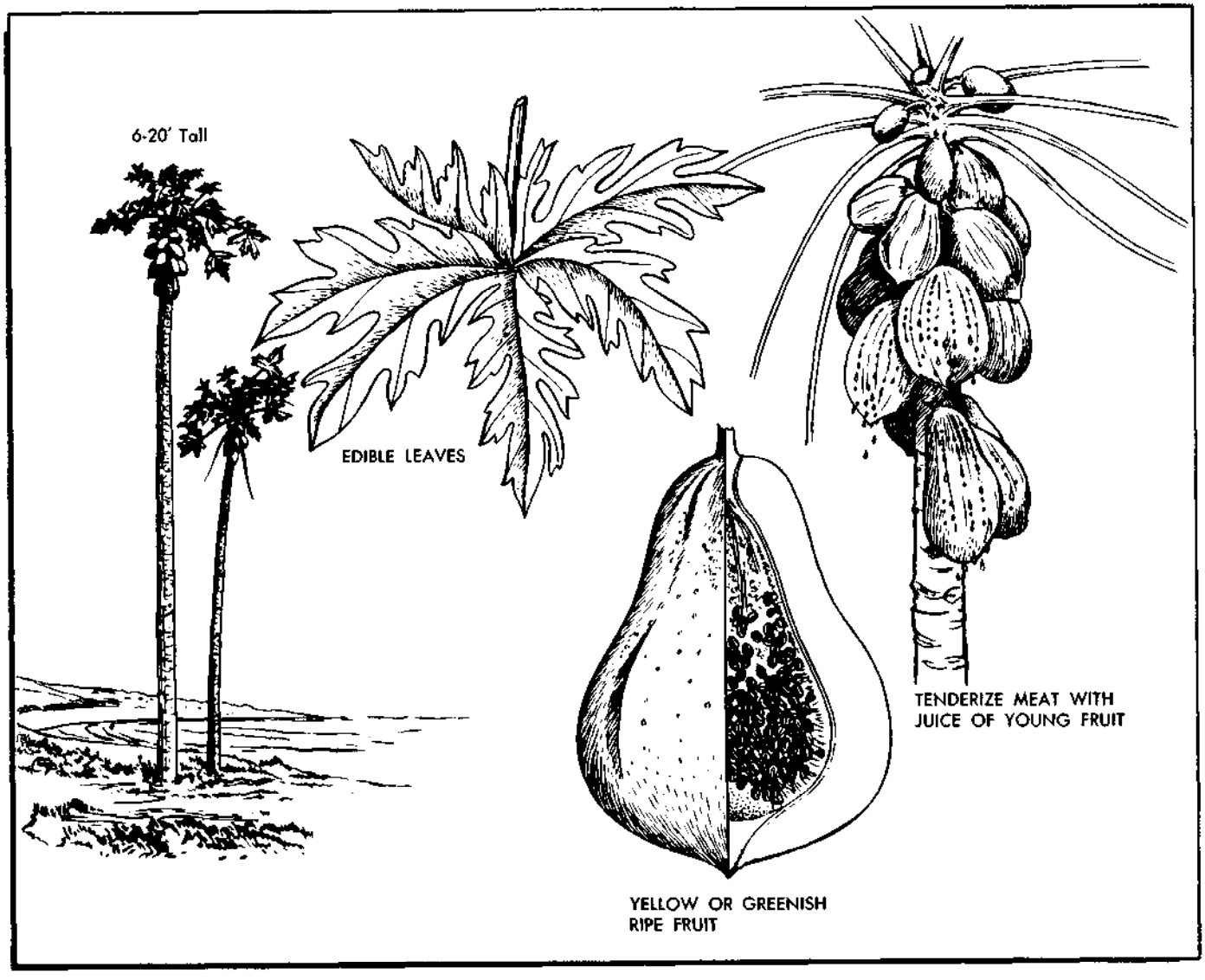

Figure 18-35. Papaya.

daylight, becoming livelier as the day warms up. In the heat of the summer, it comes out of its shallow burrow in the early morning, the late evening, or not at all.

(4) In general, desert birds stay in areas of heavier vegetation and many need water daily; therefore, most will be found within short flights of some type of water source. Many birds will migrate during the drought season. If an abundance of birds is seen, insects, vegetation, and a water source will normally be nearby.

(5) Rabbits, prairie dogs, and rats have learned to live in deserts. They remain in the shade or burrow into the ground protecting themselves from the direct sun and heated air as well as from the hot desert surface.

(6) Larger mammals are also found in the desert. This group consists of gazelles, antelope, deer, foxes, small cats, badgers, dingos, hyenas, etc., and are amazingly abundant. Most are nocturnal and generally avoid humans. They roam at night eating smaller game and insects; a few eat plants; and a few can be hazardous to a survivor. Any of these mammals should be approached with caution.

(7) Only a few of the available animals and plants have been discussed. If the possibility of having to survive in a desert area exists, the aircrew member should try to become familiar with the food source available in that area.

18-8. Food in Snow and Ice Climates. In the snow and ice climates, food is more difficult to find than water. Animal life is normally more abundant during the warm months, but it can still be found in the cold months. Fish are available in most waters during the warmer months but they congregate in deep waters, large rivers, and lakes during the cold months. Some edible plant life can be found throughout the year in most areas of the arctic.

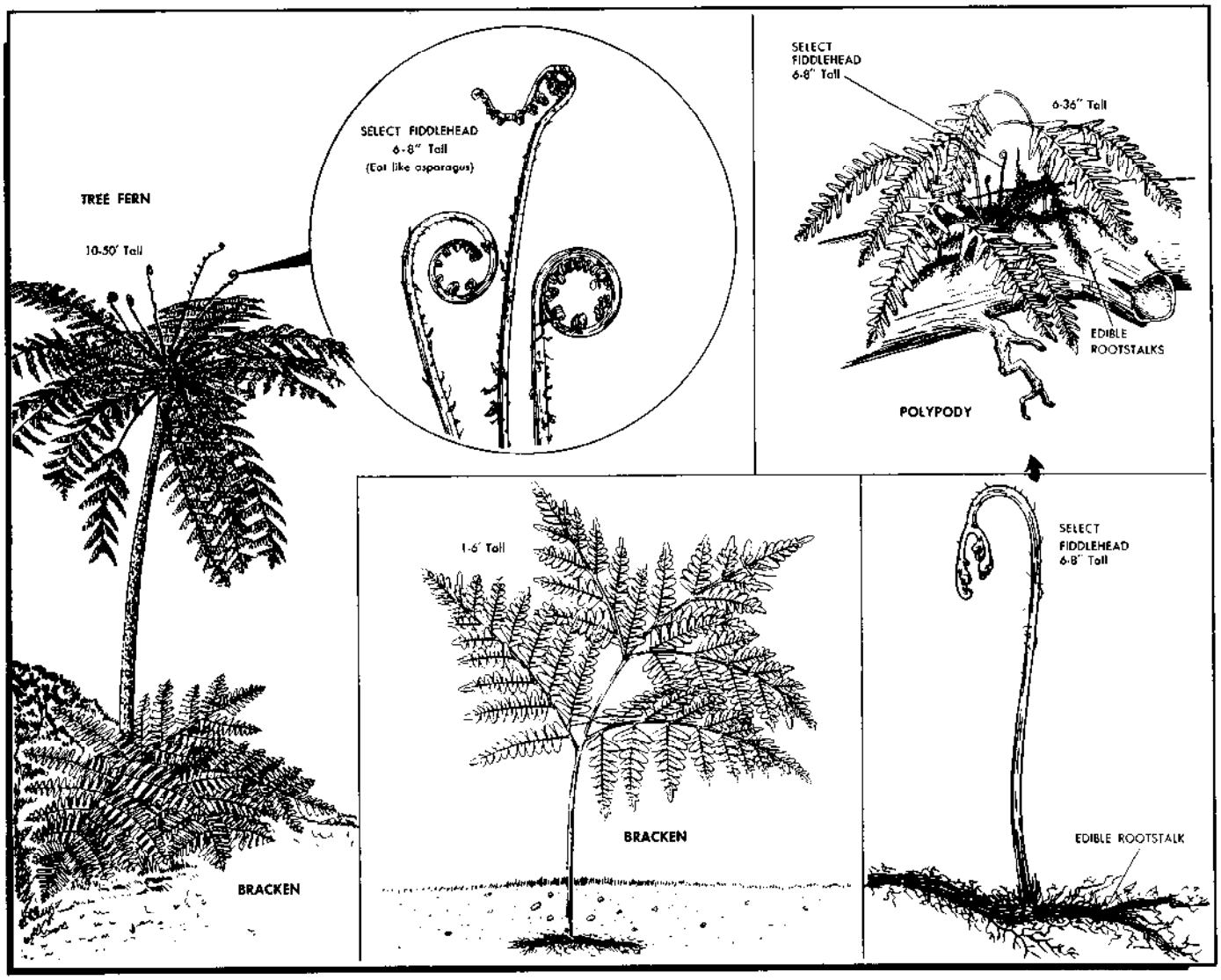

Figure 18-36. Edible Ferns.

a. All animals in the arctic regions are edible, but the livers of seals and polar bears must not be eaten because of the high concentration of vitamin A. Death could result from ingesting large quantities of the liver. On the open sea ice, game animals such as seal, walrus, polar bear, and fox are available. Many types of birds can be found during the warmer months. Fish can be caught throughout the year.

(1) Seal will probably be the main source of food when stranded on the open sea ice. They can be found in open leads, areas of thin ice, or where snow has drifted over a pressure ridge forming a cave which could have open water or very thin ice. These areas may also house polar bears which feed primarily on seals. Polar bears should be avoided.

(a) Newborn seals have trouble staying afloat or swimming and will be found on the ice in the early summer. The seal cubs can be easily killed with a club, spear, knife, or firearm and make an excellent source of food. The meat, blubber (fat), and coagulated milk in their stomachs are edible. When killing a cub, it is best to keep a lookout for the mother. She tends to protect offspring in any way possible.

(b) Seals must surface periodically to breathe. When the icepack is thin, the seals poke their noses through the ice and take a breath of air in a lead or in open water. In thick ice, the seal will chew and (or) claw a breathing hole through the ice. Normally most seals will have more than one breathing hole. In hunting seals, it is best to take a position beside a breathing hole and wait until a seal comes up to breathe, then spear or strike it on the head with a club. Seals are very sensitive to blows on or about the nose. They will often lose consciousness but not die. A hook can be suspended

through the breathing hole so it hangs down at least 6 inches below the ice. When a seal comes to breathe, it can become hooked when it tries to depart the breathing hole. Seals can be recovered by gaffing or grabbing by hand, but in some cases, the breathing hole might have to be enlarged to pull the body through. If the seal is killed in open water, a "manak" or "grapple hook" can be used to retrieve it. All seals killed in open water or those that fall into open water should be recovered immediately. During the cold months, they will float for quite awhile, but during the warm months or when a female is nursing young, they sink rapidly. This is due to the loss of body fat (figure 18-37).

(2) Birds are plentiful during the summer months and can be procured by spearing, clubbing, catching with a baited fishhook, or use of a weapon.

(3) On tundra areas, there are large game, small game, and birds available as a food source.

(a) The large game consists of caribou, musk oxen, sheep, wolf, and bears (figure 18-38). Even though the large game animals can be a food source, they will be difficult to procure if a firearm is not available. Therefore, they should be considered a hazard to a survivor without a firearm. In the spring, bears tend to congregate along rivers and streams due to the amount of food available—normally salmon. During the fall, bears will be found feeding at berry patches. During certain seasons of the year, these areas should be avoided.

(b) Small game animals of the tundra include hares, lemmings, mice, ground squirrels, marmots, and foxes. They may be trapped or killed the entire year. When snaring, it is best to use a simple loop made of strong line or wire. The wire must be a two-strand twisted wire since metal becomes brittle in the cold and breaks very easily. Other snares and triggers will be less effective in the cold climate. A gill net can be used as a snare by spreading it across a trail so that the animal will entangle itself.

(c) Surface water is generally plentiful due to the number of lakes, ponds, bogs, and marshes. Water fowl and birds are very abundant during the warm months and include ducks, terns, geese, gulls, owls, and ptarmigan. The eggs and young birds are an excellent food source and can be easily procured (figure 18-39).

(4) As in the tundra areas, the forested areas in the arctic and arctic-like areas abound in wildlife.

(a) The large game species include moose, deer, caribou, and bear.

Figure 18-37. Walruses.

Figure 18-38. Big Horn Sheep.

(b) Small game of the forests includes hares, squirrels, porcupine, muskrat, and beaver. They can be snared or trapped easily in winter or summer. Small animal trails can be found in the winter with great ease. Most animals do not like to travel in deep snow so they tend to travel the same trail most of the time and this trail will look like a small superhighway — the snow packed down well below the normal snow level. Most trails will also be located in heavy cover or undergrowth or parallel to roads and open areas. The same trails will normally be used during the summer.

(5) During the summer months, the open water provides an excellent opportunity to procure all types of fish, both freshwater and saltwater, and freshwater mussels. The mussels can be handpicked off the bottom, while the fish can be netted, speared, clubbed, or caught with a hook and line. After freezeup, fishing is still possible through the ice. Shallow lakes, rivers, or ponds can freeze completely killing off all fish life. Fish tend to congregate in the deepest water possible. A hole should be cut through the ice at the estimated deepest point. Other good locations are at outlets or where tributaries flow into lakes or ponds. The ice is normally thinner over rapid moving water and at the edges of deep streams or rivers with snowdrifts extending out from the banks. Open water is often marked by a mist or fog formed over the area by vaporizing water. All methods of procuring fish in the summer will work in the winter (figure 18-40).

(6) The ocean shores are rich hunting grounds for edible sea life such as clams, mussels, scallops, snails, limpets, sea urchins, chitons, and sea cucumbers. They can be procured most of the year wherever there is open water. Tidal pools usually contain a great number of both fish and mollusks. The fish can be netted, speared, or hand caught. All sea life can be eaten raw, but cooking usually makes it more palatable (figure 18-41).

b. The plant life of the arctic regions is generally small and stunted due to the effects of permafrost, low mean temperatures, and a short growing season.

Figure 18-39. Swans.

(1) On the barren tundra areas, a wide variety of small edible plants and shrubs exist. During the short summer months on the tundra, Labrador tea, fireweed, coltsfoot, dwarf arctic birch, willow, and numerous other plants and berries can be found. During the winter, roots, rootstalks, and frozen berries can be found beneath the snow. Lichens and mosses are abundant but should be selected carefully as some species are poisonous.

(2) In bog or swamp areas, many types of water sedge, cattail, dwarf birch, and berries are available. During spring and summer, many young shoots from these plants are easily collected.

(3) The wooded areas of the arctic contain a variety of trees (birch, spruce, poplar, aspen, and others). Many berry-producing plants can be found, such as blueberries, cranberries, raspberries, cloud berries, and crow berries. Wild rose hips, Labrador tea, alder, and other shrubs are very abundant. Many wild edible plants are highly nutritious. Greens are particularly rich in carotene (vitamin A). Leafy greens, many berries, and rose hips are all rich in ascorbic acid (vitamin C). Many roots and rootstalks contain starch and can be used as a potato substitute in stews and soups.

(4) Although there are several types of edible mushrooms, fungi, and puff–balls in the arctic, a person should avoid ingesting them because it is difficult to identify the poisonous and nonpoisonous species. During the growing season, the physical characteristics can change considerably making positive identification even more difficult.

(5) There are many poisonous plants and a few poisonous berries in the arctic. Very few cause death; many will cause extreme nausea, dizziness, abdominal pain, and diarrhea. Contact poisonous plants, such as poison ivy, are not found in the arctic. The more common poisonous plants are shown in figures 18-42 through 18-49.

(6) When selecting edible plants, select young shoots when possible as these will be the most tender. Plants should be eaten raw to obtain the most nutritive value. Some of the more common edible plants are:

(a) Dandelions generally grow with grasses but may be scattered over rather barren areas. Both leaves

Figure 18-40. Winter Fishing.

and roots are edible raw or cooked. The young leaves make good greens; the roots (when roasted) are used as a substitute for coffee.

(b) Black and white spruce are generally the northern most evergreens. These trees have short, stiff needles that grow singularly rather than in clusters like pine needles. The cones are small and have thin scales. Although the buds, needles, and stems have a strong resinous flavor, they provide essential vitamin C by chewing them raw. In spring and early summer, the inner bark can be used for food.

(c) The dwarf arctic birch is a shrub with thin tooth-edged leaves and bark which peels off in sheets. The fresh green leaves and buds are rich in vitamin C. The inner bark may also be eaten.

(d) There are many different species of willow in the arctic. Young tender shoots may be eaten as greens and the bark of the roots is also edible. They have a decidedly sour taste but contain a large amount of vitamin C (figure 18-50).

(7) Lichens are abundant and widespread in the far North and can be used as a source of emergency food. Many species are edible and rich in starch-like substances, including Iceland moss, peat moss and reindeer lichen. Beard lichen growing on trees has been used as food by Indians. However, some of it contains a bitter acid which causes irritation of the digestive tract. If lichens are boiled, dried, and powdered, this acid is removed and the powder can then be used as flour or made into a thick soup.

18-9. Food on Open Seas. Almost all sea life is not only edible, but is also an excellent source of nutrients essential to humans. The protein is complete because it contains all the essential amino acids, and the fats are similar to those of vegetables. Seafoods are high in minerals and vitamins. The majority of life in the sea (fish, birds, plants, and aquatic animals) is edible.

a. Most seaweeds are edible and are a good source of food, especially for vitamins and minerals. Some seaweeds contain as much as 25 percent protein, while others are composed of over 50 percent carbohydrates. At least 75 different species are used for food by seacoast residents around the world. For many people, es-

Figure 18-41. Shell Fish.

pecially the Japanese, seaweeds are an essential part of the diet, and the most popular varieties have been successfully farmed for hundreds of years. The high cellulose content may require gradual adaptation because of their laxative quality if they comprise a large part of the diet. As with vegetables, some species are more flavorful than others. Generally, leafy green, brown, or red seaweeds can be washed and eaten raw or dried. The following list of edible seaweeds gives a description of the plant, tells where it may be found, and in many cases, suggests a method of preparation:

(1) Common green seaweeds (figure 18-51), often called sea lettuce (Ulva lactuca), are in abundance on both sides of the Pacific and North Atlantic oceans. After washing it in clean water, it can be used as a garden lettuce.

(2) The most common edible brown seaweeds are the sugar wrack, kelp, and Irish moss (figure 18-52).

(a) The young stalks of the sugar wrack are sweet to taste. This seaweed is found on both sides of the Atlantic and on the coasts of China and Japan.

(b) Edible kelp has a short cylindrical stem and thin, wavy olive-green or brown fronds one to several feet in length. It is found in the Atlantic and Pacific oceans, usually below the high-tide line on submerged ledges and rocky bottoms. Kelp should be boiled before eating. It can be mixed with vegetables or soup.

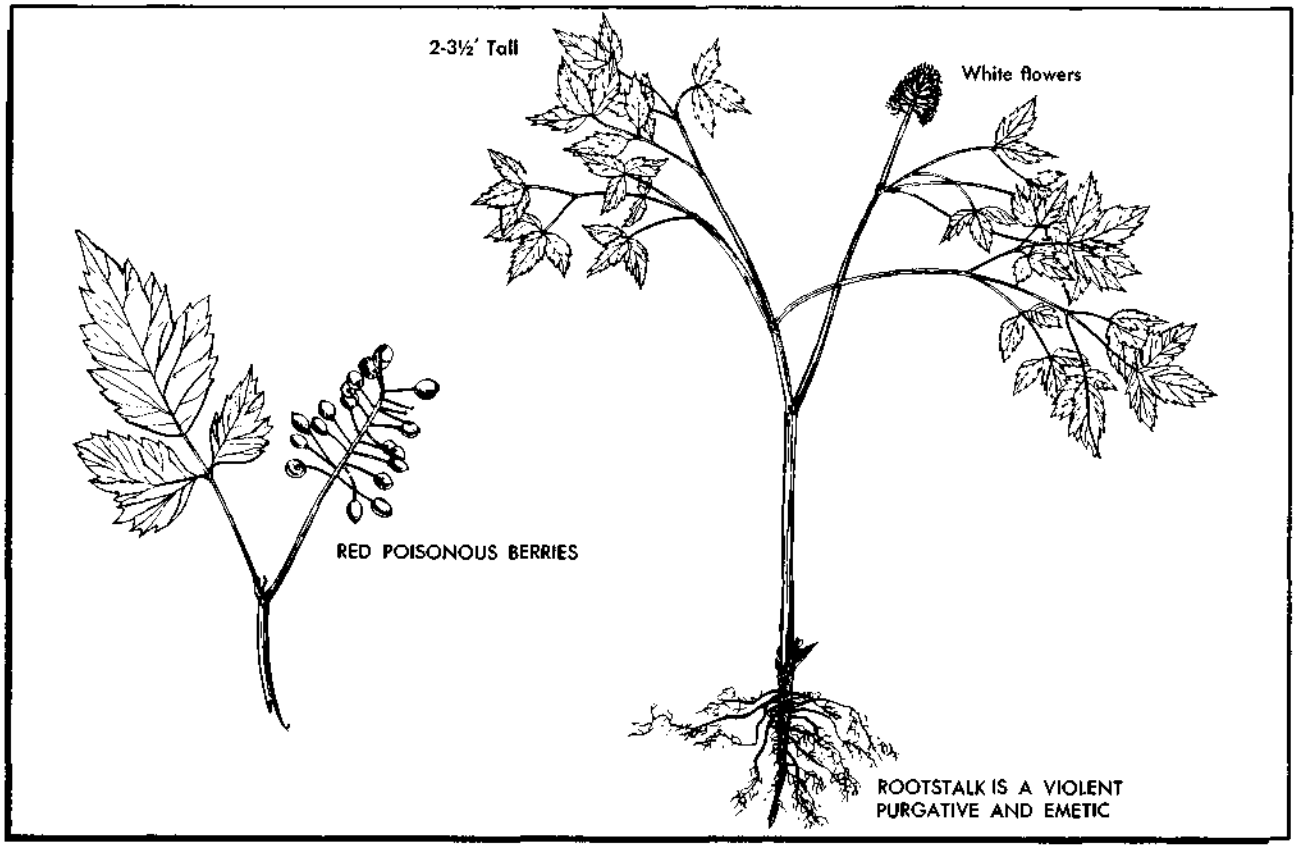

Figure 18-42. Baneberry.

(c) Irish moss, a variety of brown seaweed, is quite edible, and is often sold in market places. It is found on both sides of the Atlantic Ocean and can be identified by its tough, elastic, and leathery texture; however, when dried, it becomes crisp and shrunken. It should be boiled before eating. It can be found at or just below the high-tide line. It is sometimes found cast upon the shore.

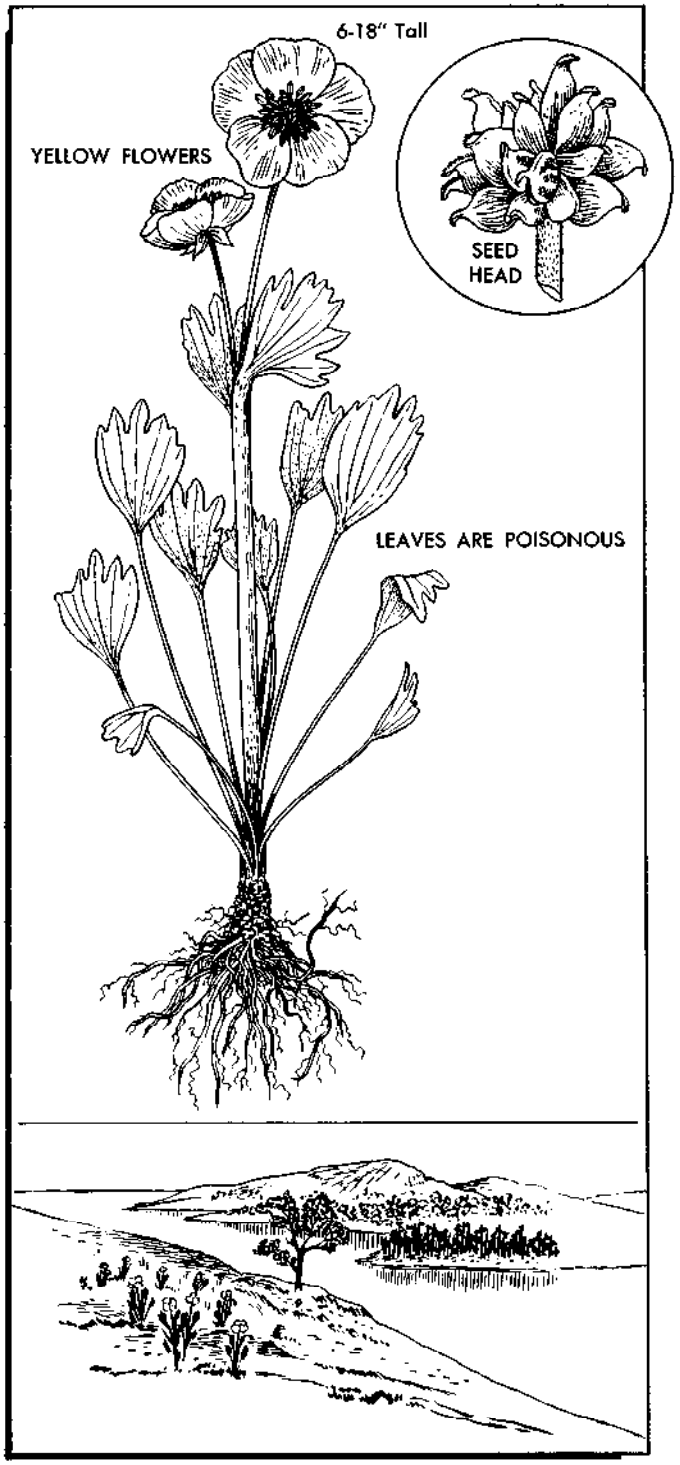

Figure 18-43. Buttercup.

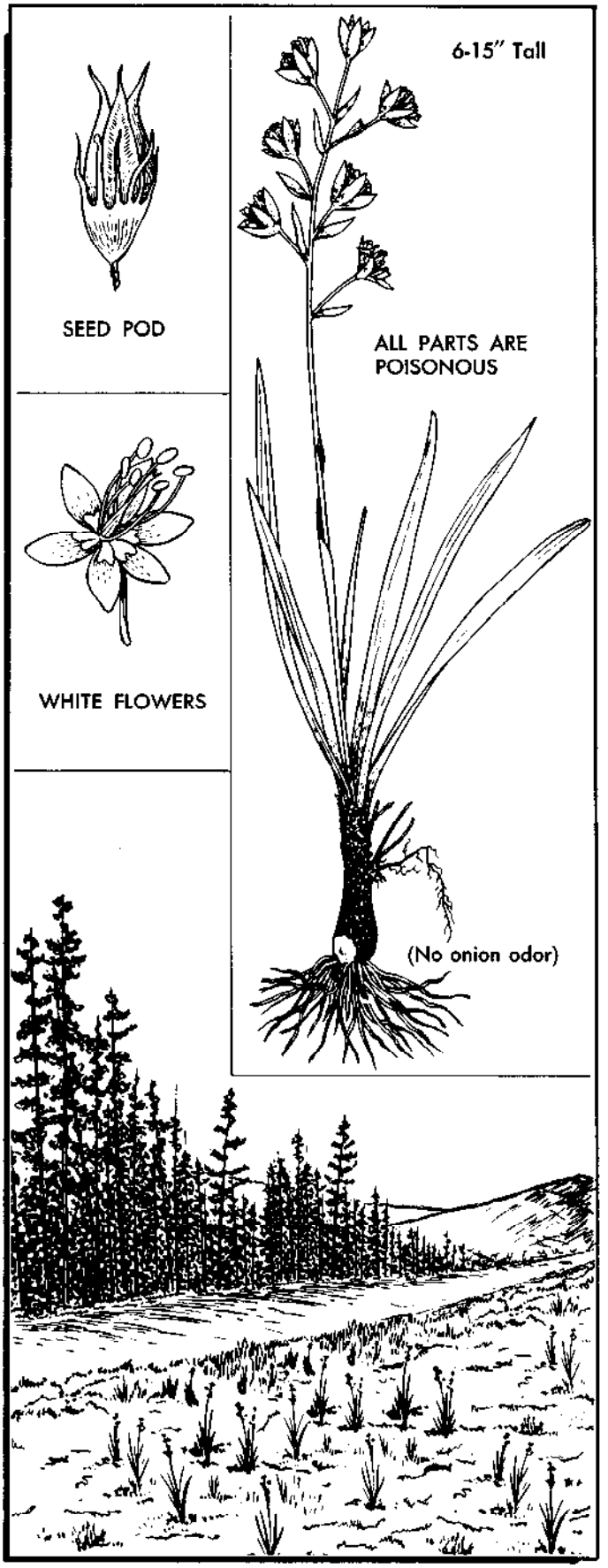

Figure 18-44. Death Camas.

(3) Red seaweeds can usually be identified by their characteristic reddish tint, especially the edible varieties. The most common and edible red seaweeds include the dulse, laver, and other warm-water varieties (figure 18-53).

(a) Dulse has a very short stem which quickly broadens into a thin, broad, fan-shaped expanse which is dark red and divided by several clefts into short, round-tipped lobes. The entire plant is from a few inches to a foot in length. It is found attached to rocks or coarser seaweeds, usually at the low-tide level, on both sides of the Atlantic Ocean and in the Mediterranean. Dulse is leathery in consistency and is sweet to the taste. If dried and rolled, it can be used as a substitute for tobacco.

(b) Laver is usually red, dark purple, or purplish-brown, and has a satiny sheen or filmy luster. Common to both the Atlantic and Pacific oceans, it has been used as food for centuries. This seaweed is used as a relish, or is cleaned and then boiled gently until tender. It can also be pulverized and added to crushed grains and fried in the form of flatcakes. During World War II, laver was chewed for its thirst-quenching value by New Zealand troops. Laver is usually found on the beach at the low-tide level.

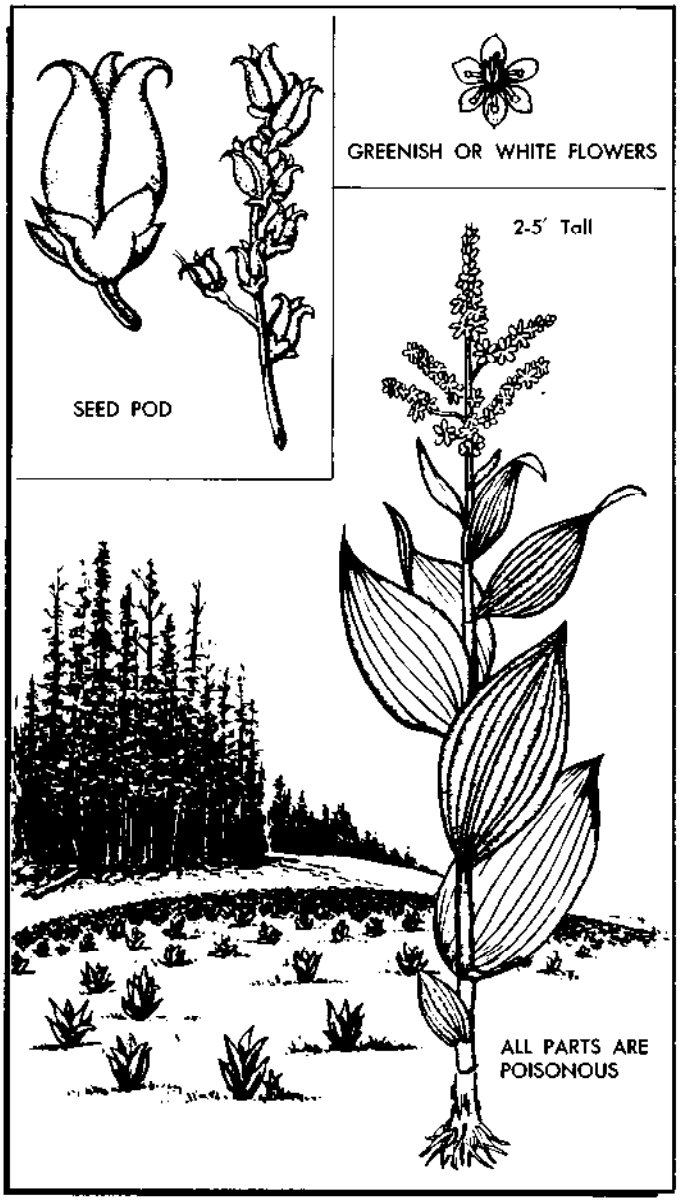

Figure 18-45. False Hellebore.

(c) A great variety of red, warm-water seaweed is found in the South Pacific area. This seaweed accounts for a large portion of the native diet. When found on the open sea, bits of floating seaweed may not only be edible but often contains tiny animals that can be used for food. The small fish and crabs can be dislodged by shaking the clump of seaweed over a container.

b. Plankton includes both minute plants and animals that drift about or swim weakly in the ocean. These basic organisms in the marine food chain are generally more common near land since their occurrence depends upon the nutrients dissolved in the water. Plankton can be caught by dragging a net through the water. The taste of the plankton will depend upon the types of organisms predominant in the area. If the population is mostly fish larvae, the plankton will taste like fish. If the population is mostly crab or shellfish larvae, the plankton will taste like crab or shellfish. Plankton contains valuable protein, carbohydrates and fats. Because of its high chiton and cellulose content, however, plankton cannot be immediately digested in large quantities. Therefore, anyone subsisting primarily on a plankton diet must gradually increase the quantities consumed. Most of the planktonic algae (phytoplankton) are smaller than the planktonic animals (zooplankton) and, although edible, are less palatable. Some plankton algae, for example, those dinoflagellates that cause "red tides" and paralytic shellfish poisoning, are toxic to humans.

(1) If a survivor is going to use plankton as a food source, there must be a sufficient supply of freshwater for drinking. Each plankton catch should be examined to remove all stinging tentacles broken from jellyfish or Portuguese man-of-war. The primarily gelatinous species may also be selectively discarded since their tissues are predominately composed of saltwater. When the plankton is found in subtropical waters during the summer months, and the presence of poisonous dinoflagellates is suspected (due to discoloration or high luminescence of the ocean), the edibility test should be applied before eating.

(2) The final precaution which a survivor may wish to take before ingesting plankton is to feel or touch the plankton to check for species that are especially spiney. The catch should be sorted (visually) or dried and crushed before eating if it contains large numbers of these spiney species.

c. If a fishing kit is available, the task of fishing will be made much easier. Small fish will usually gather under the shadow of the raft or in clumps of floating seaweed.

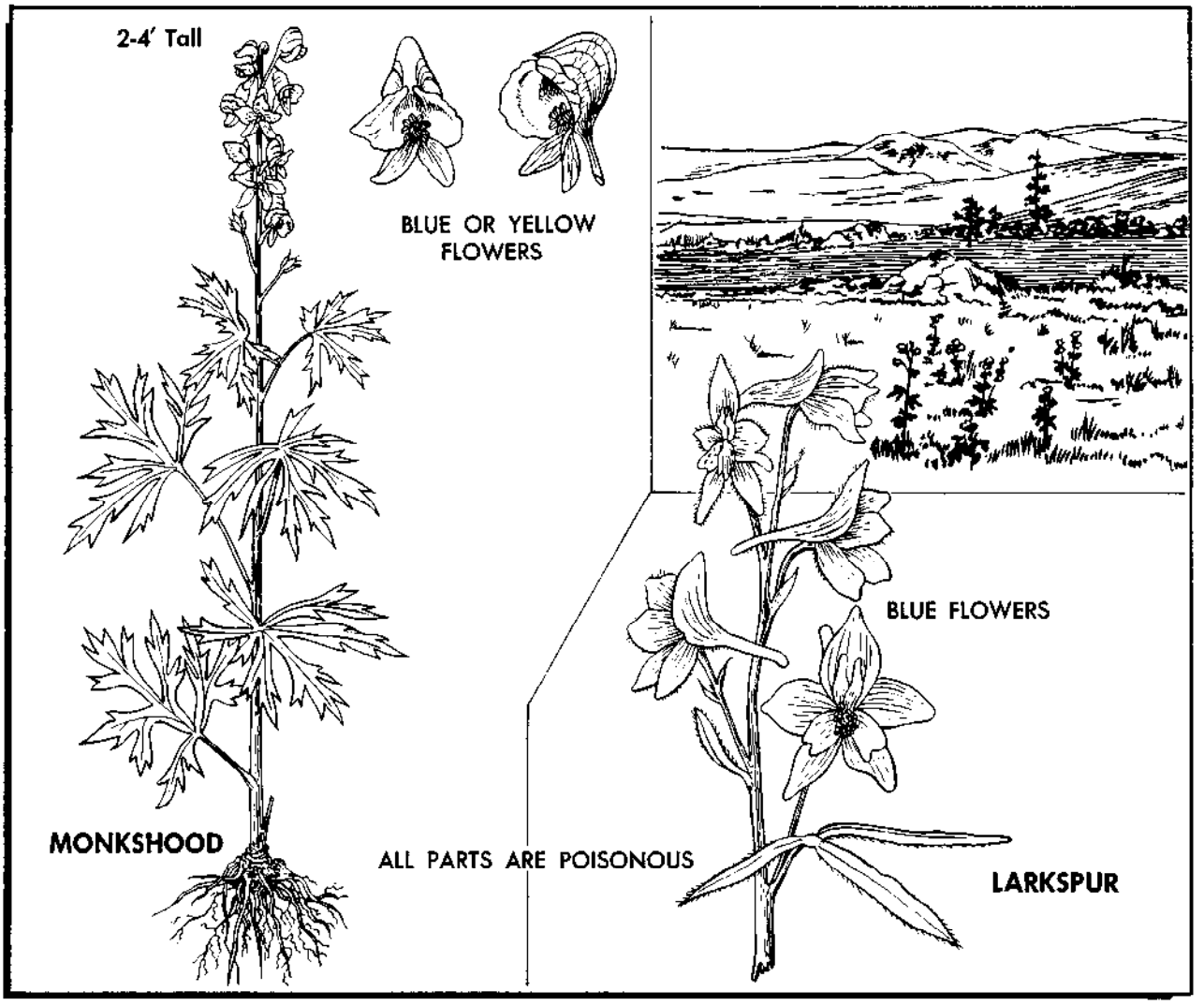

Figure 18-46. Monkshood and Larkspur.

These fish can be eaten or used as bait for larger fish. A net can also be used to procure most all sea life. Light attracts some types of fish. A flashlight or reflected moonlight can be used. It is not advisable to secure fishing lines to the body or the raft because a large fish may pull a person out of the raft or damage the raft. Fish, bait, or bright objects dangling in the water can attract large dangerous fish. All large fish should be killed outside the raft by a blow to the head or by cutting off the head.

d. Sea birds have proven to be a useful food source which may be more easily caught than fish. Survivors have reported capturing birds by using baited hooks, by grabbing, and by shooting. Freshly killed birds should be skinned, rather than plucked, to remove the oil glands. They can be eaten raw or cooked. The gullet contents can be a good food source. The flesh should be eaten or preserved immediately after cleaning. The viscera, along with any other unused parts, make good fish bait.

e. Marine mammals are rarely encountered by a person in the water, although they may be seen from a distance. Any large mammal is capable of inflicting injuries, but unless such mammals are pursued, they will generally avoid people. The killer whale (Orca) is rarely seen and, although large enough to feed on humans, has never been known to do so. Almost all sea mammals are a good source of food but difficult to obtain. The liver, especially that of any arctic or cold-water mammal, should not be eaten because of toxic concentrations of vitamin A.

f. All sea life must be cleaned, cut up, and eaten as soon as possible to avoid spoilage. Any meat left over can be preserved by sun-drying or smoking. The internal parts can be used as bait. If any doubt exists as to the edibility of a seafood, apply the "marine animal edibility test" found in chapter 11.

18-10. Preparing Animal Food. Survivors must know how to use the meat of game and fish to their advantage

and how to do this with the least effort and physical exertion. Many people have died from starvation because they had failed to take full advantage of a game carcass. They abandoned the carcass on the mistaken theory that they could get more game when needed.

a. If the animal is large, the first impulse is usually to pack the meat to camp. In some cases, it might be easier to move the camp to the meat. A procedure often advocated for transporting the kill is to use the skin as a sled for dragging the meat. When the entire animal is dragged, this method may prove satisfactory only on frozen lakes or rivers or over very smooth snow-covered terrain. In rough or brush-covered country, however, it is generally more difficult to use this method, although it will work. Large mountain animals can sometimes be dragged down a snow-filled gully to the base of the mountain. If meat is the only consideration, and the survivors do not care about the condition of the skin, mountain game can sometimes be rolled for long distances. Before transporting a whole animal, it should be gutted and the incision closed. Once the bottom of the hill is reached, almost invariably the method is either to backpack the meat to camp, making several trips if no other survivors are present, or to pack the camp to the animal. Under survival conditions, home is on the back. When the weight of the meat proves excessive and moving the camp is not practical, some of the meat could be eaten at the scene. The heart, liver, and kidneys should be eaten as soon as possible to avoid spoilage.

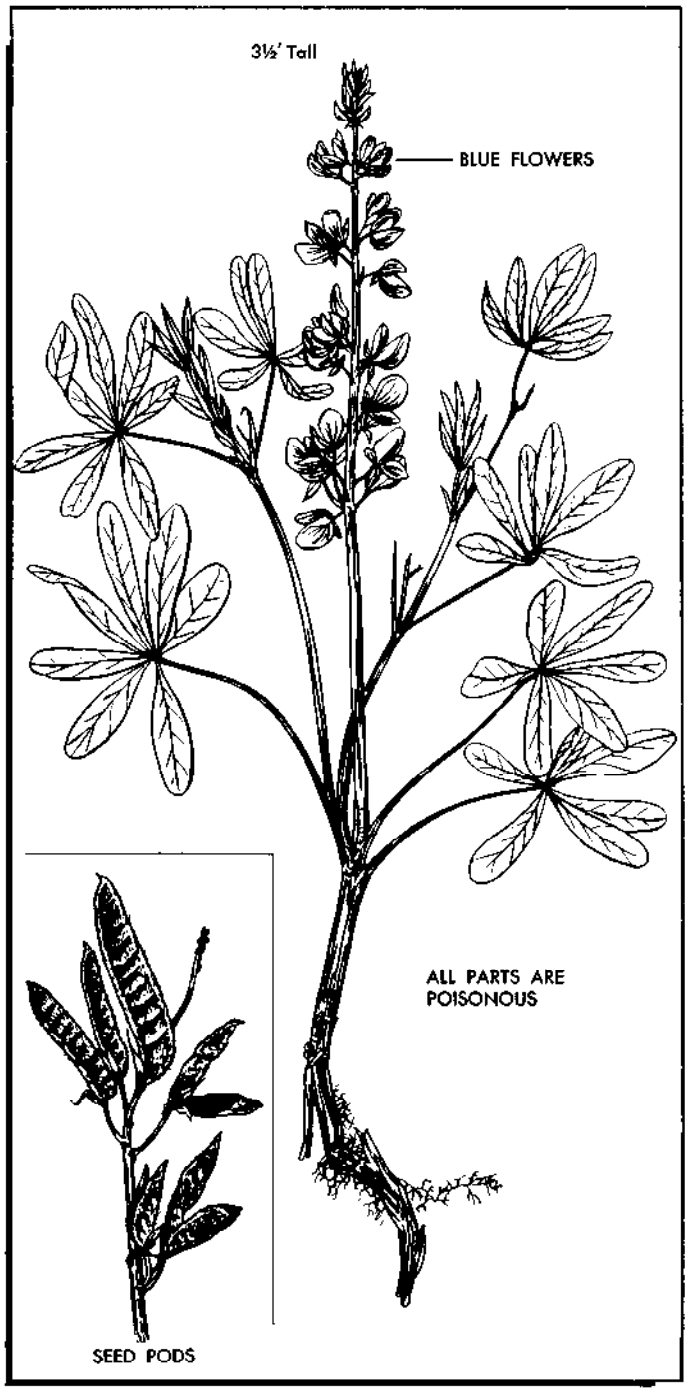

Figure 18-47. Lupine.

(1) Under survival conditions, skinning and butchering must be done carefully so that all edible meat can be saved. When the decision is made to discard the skin, a rough job can be done. However, considerations should be given to possible uses of the skin. A square of fresh skin, long enough to reach from the head to the knees, will not weigh much less when it is dried, and is an excellent ground cloth for use under a sleeping bag on frozen ground or snow. The best time to skin and butcher an animal is immediately after the kill. However, if an animal is killed late in the day, it can be gutted immediately and the other work done the next morning. An effort to keep the carcass secure from predators should be made.

(2) When preparing meat under survival conditions, all edible fat should be saved. This is especially important in cold climates, as the diet may consist almost entirely of lean meat. Fat must be eaten in order to provide a complete diet. Rabbits lack fat, and the fact that a person will die after an extended diet consisting only of rabbit meat indicates the importance of fat in a primitive diet. The same is true of birds, such as the ptarmigan.

(3) Birds should be handled in the same manner as other animals. They should be cleaned after killing and protected from flies. Birds, with the exception of sea birds, should be plucked and cooked with the skin on. Carrion-eating birds, such as vultures, must be boiled for at least 20 minutes to kill parasites before further cooking and eating. Fish-eating birds have a strong, fish-oil flavor. This may be lessened by baking them in mud or by skinning them before cooking.

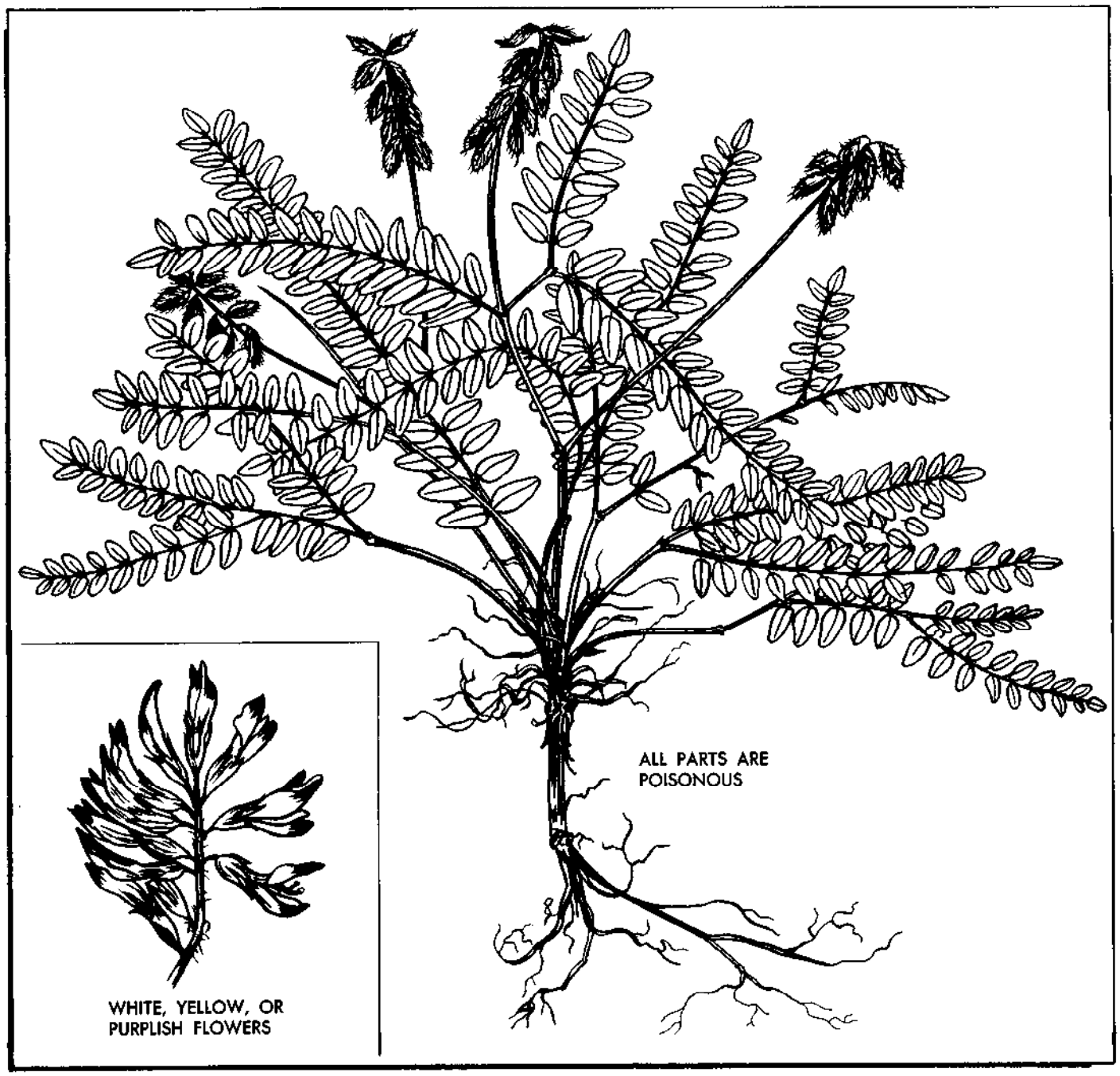

Figure 18-48. Vetch and Locoweed.

b. There are two general ways to skin animals depending upon the size: the big game method, or the glove skinning method.

(1) Survivors should use the big game method when skinning and butchering large game.

(a) The first step in skinning is to turn the animal on its back and with a sharp knife, cut through the skin on a straight line from the tail bone to a point under its neck as illustrated in figure 18-54. In making this cut, pass around the anus and, with great care, press the skin open until the first two fingers can be inserted between the skin and the thin membrane enclosing the entrails. When the fingers can be forced forward, place the blade of the knife between the fingers, blade up, with knife held firmly. While forcing the fingers forward, palm upward, follow with the knife blade, cutting the skin but not cutting the membrane.

(b) If the animal is a male, cut the skin parallel to, but not touching the penis. If the tube leading from the bladder is accidentally cut, a messy job and unclean meat will result. If the gall or urine bladders are broken, washing will help clean the meat. Otherwise, it is best

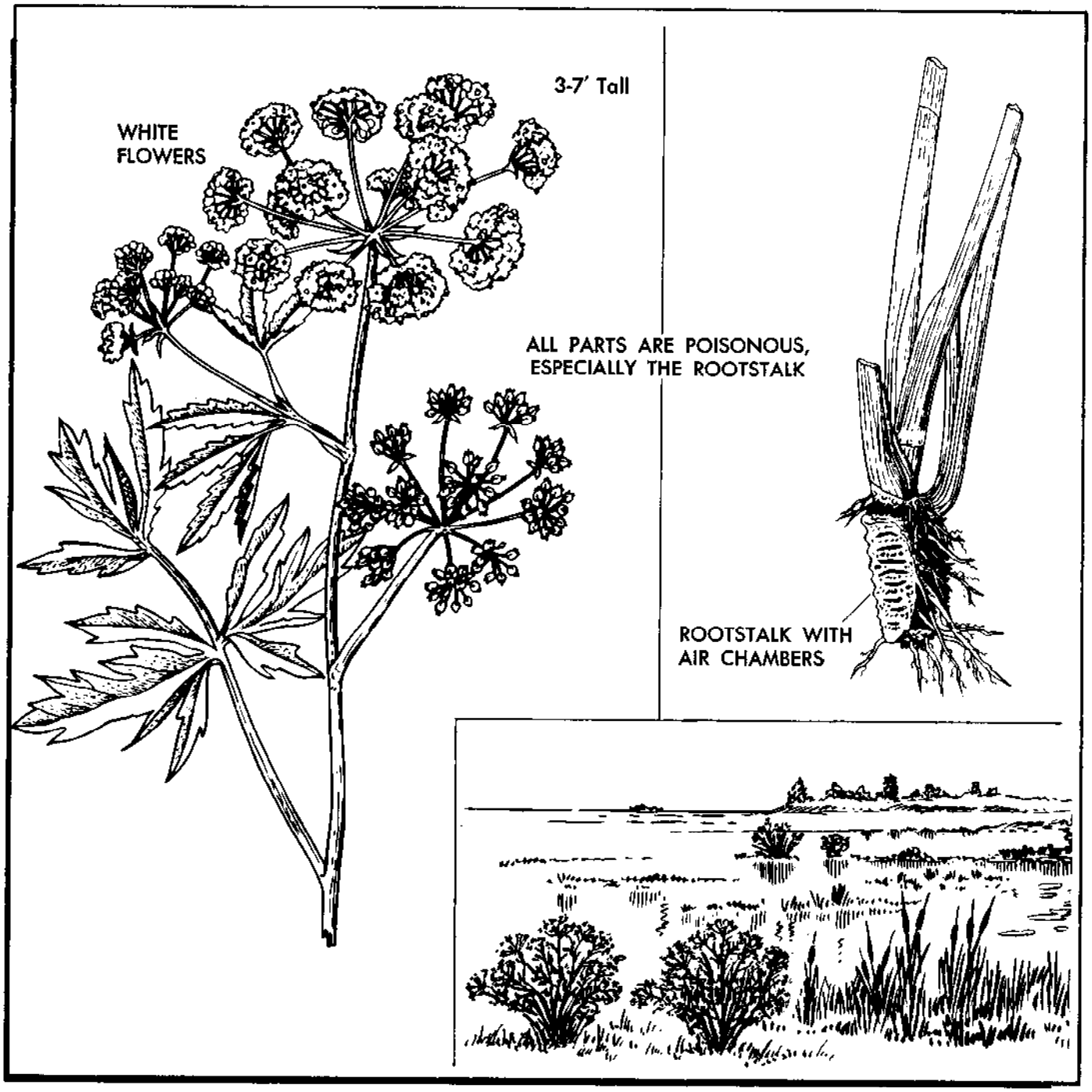

Figure 18-49. Water Hemlock.

not to wash the meat but to allow it to form a protective glaze.

(c) On reaching the ribs, it is no longer possible to force the fingers forward, because the skin adheres more strongly to flesh and bone. Furthermore, care is no longer necessary. The cut to point C can be quickly completed by alternately forcing the knife under the skin and lifting it. With the central cut completed, make side cuts consisting of incisions through the skin, running from the central cut (A-C) up the inside of each leg to the knee and hock joints. Then make cuts around the front legs just above the knees and around the hind legs above the hocks. Make the final cross cut at point C, and then cut completely around the neck and in back of the ears. Now is the time to begin skinning.

(d) On a small or medium-sized animal, one person can skin on each side. The easiest method is to begin at the corners where the cuts meet. When the animal is large, three people can skin at the same time. However, one should remember that when it is getting

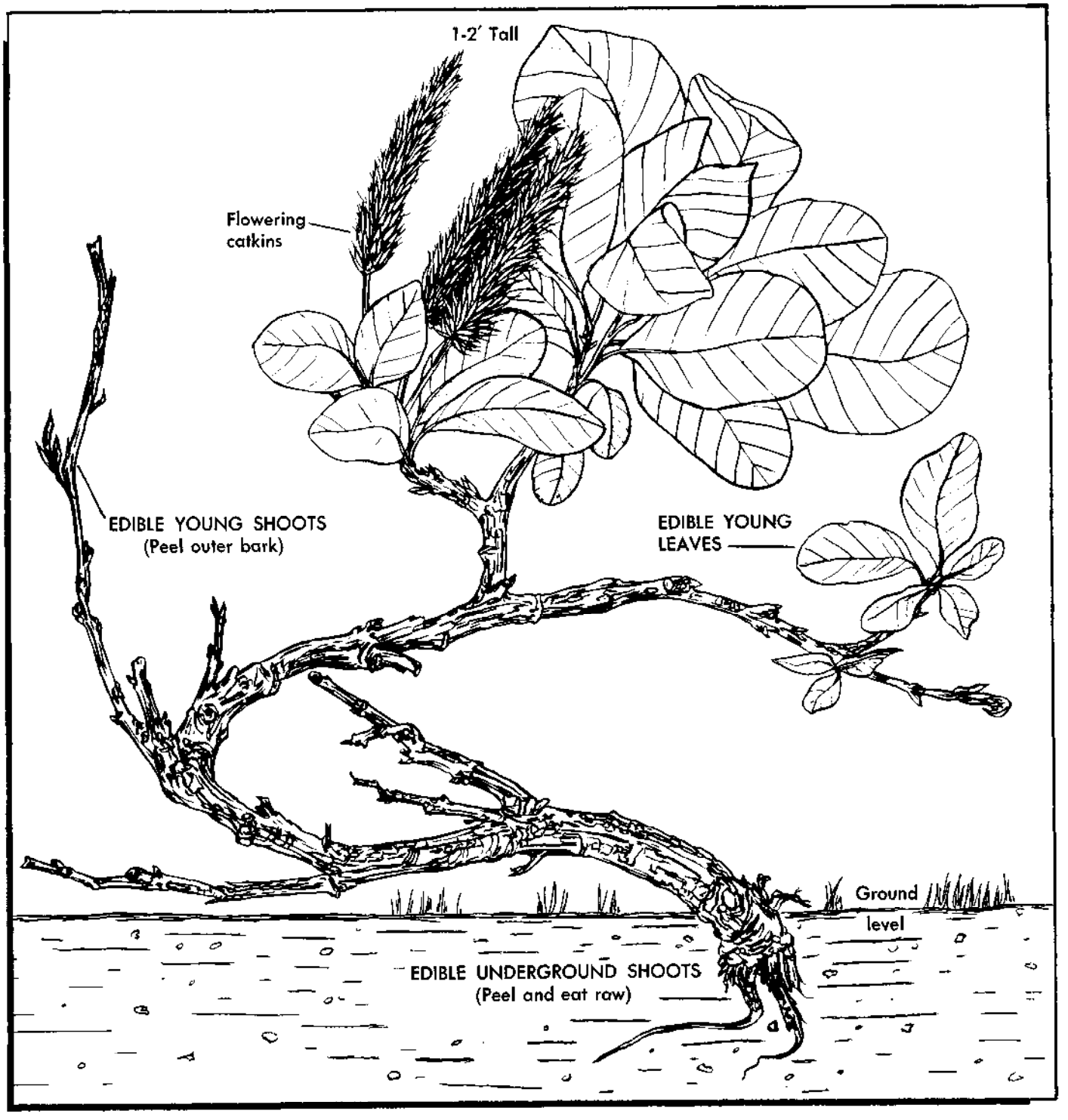

Figure 18-50. Arctic Willow.

dark and hands are clumsy because of the cold, a sharp skinning knife can make a deep wound. After skinning down the animal's side as far as possible, roll the carcass on its side to skin the back. Then spread out the loose skin to prevent the meat from touching the ground and turn the animal on the skinned side. Follow the same procedure on the opposite side until the skin is free.

(e) In opening the membrane which encloses the entrails, follow the same procedure used in cutting the skin by using the fingers of one hand as a guard for the knife and separating the intestines from membrane. This thin membrane along the ribs and sides can be cut away in order to see better. Be careful to avoid cutting the intestines or bladder. The large intestine passes through an aperture in the pelvis. This tube must be separated from the bone surrounding it with a knife. Tie

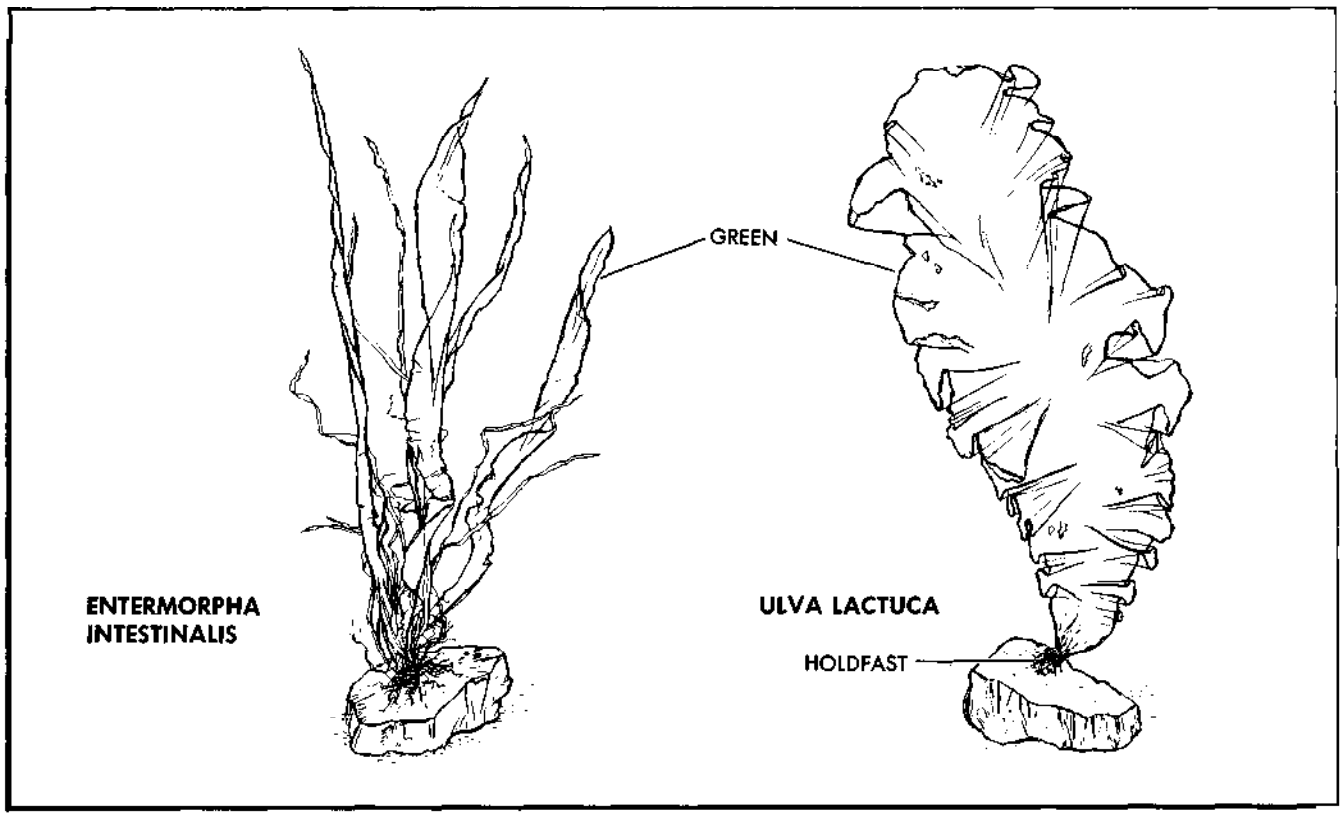

Figure 18-51. Edible Green Seaweeds.

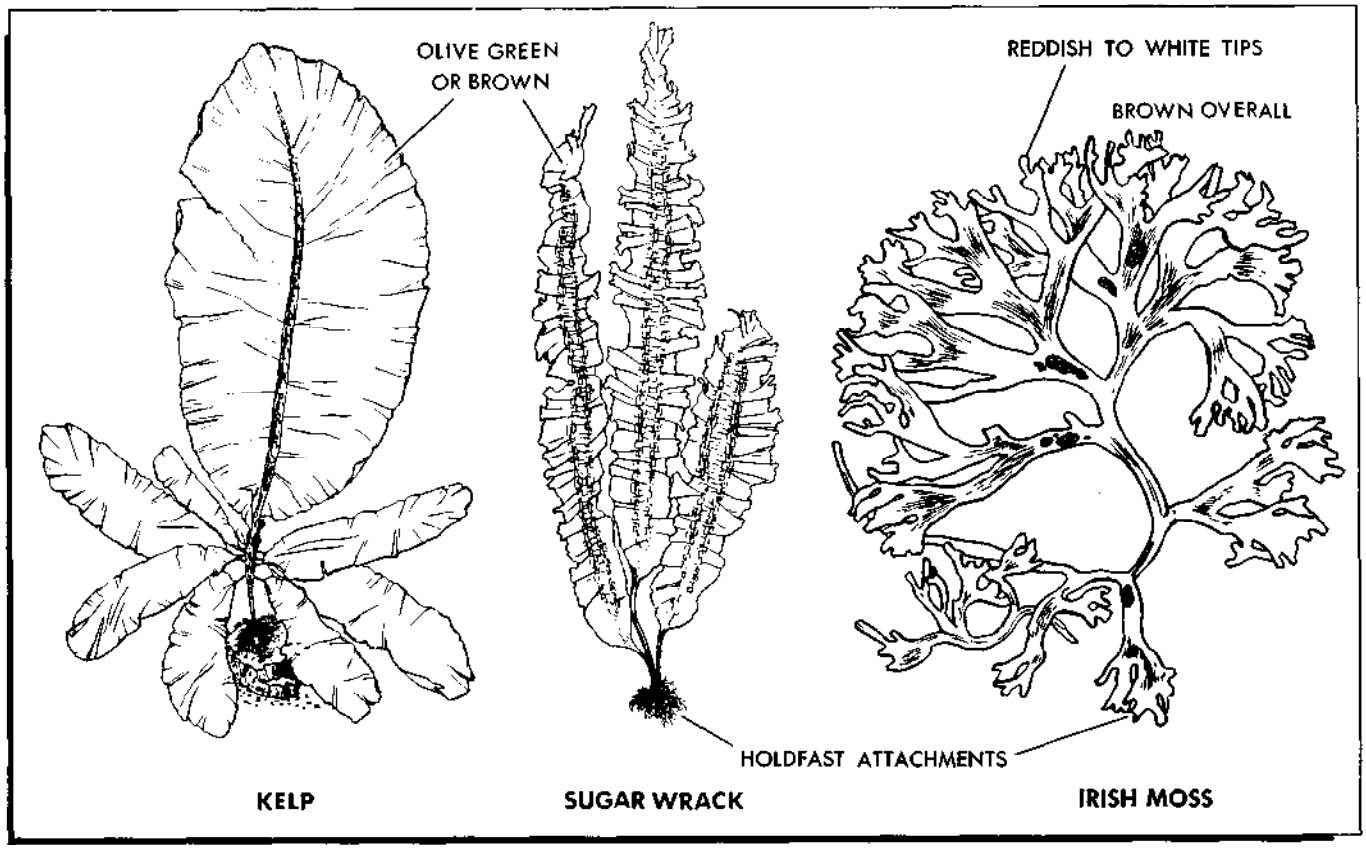

Figure 18-52. Edible Brown Seaweeds.

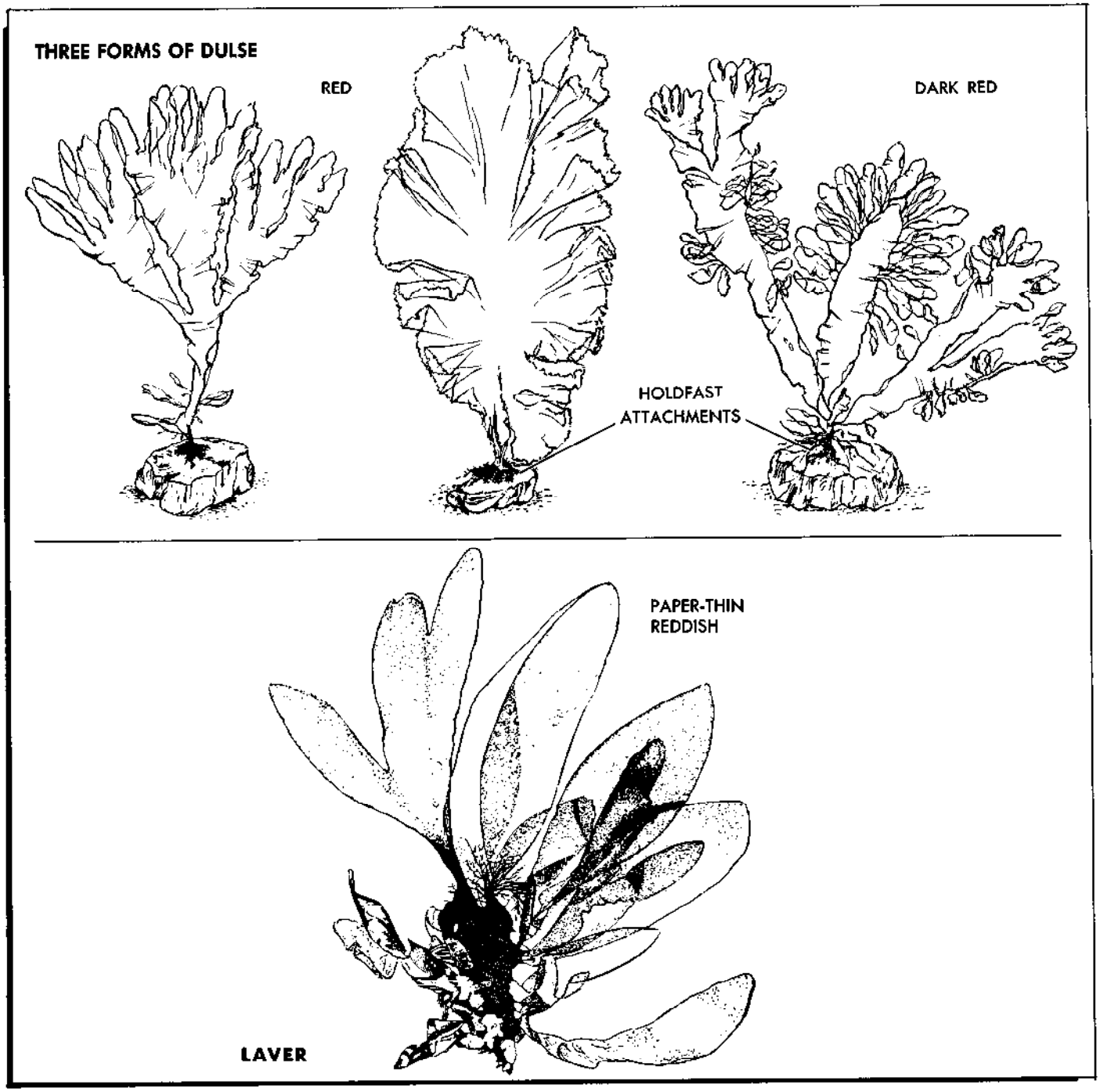

Figure 18-53. Edible Red Seaweeds.

a knot in the bladder tube to prevent the escape of urine. With these steps completed, the entrails can be easily disengaged from the back and removed from the carcass. Another method of gutting or field dressing is shown in figure 18-55. After gutting is completed, it may be advisable to hang the animal. Figure 18-56 shows two methods. (NOTE: If it is hot, gut the animal before skinning it.)

(f) The intestines of a well-conditioned animal are covered with a lace-like layer of fat which can be lifted off and placed on nearby bushes to dry for later use. The gall bladder which is attached to the liver of some animals should be carefully removed. If it should happen to rupture, the bile will taint anything it touches. Be sure to clean the knife if necessary. The kidneys are imbedded in the back, forward of the pelvis, and are covered with fat. Running forward from the kidneys on each side of the backbone are two long strips of chop-meat or muscle called tenderloin or backstrap. Eat this after the liver, heart, and kidneys as it is usually very tender. Edible meat can also be removed from the head, brisket, ribs, backbone, and pelvis.

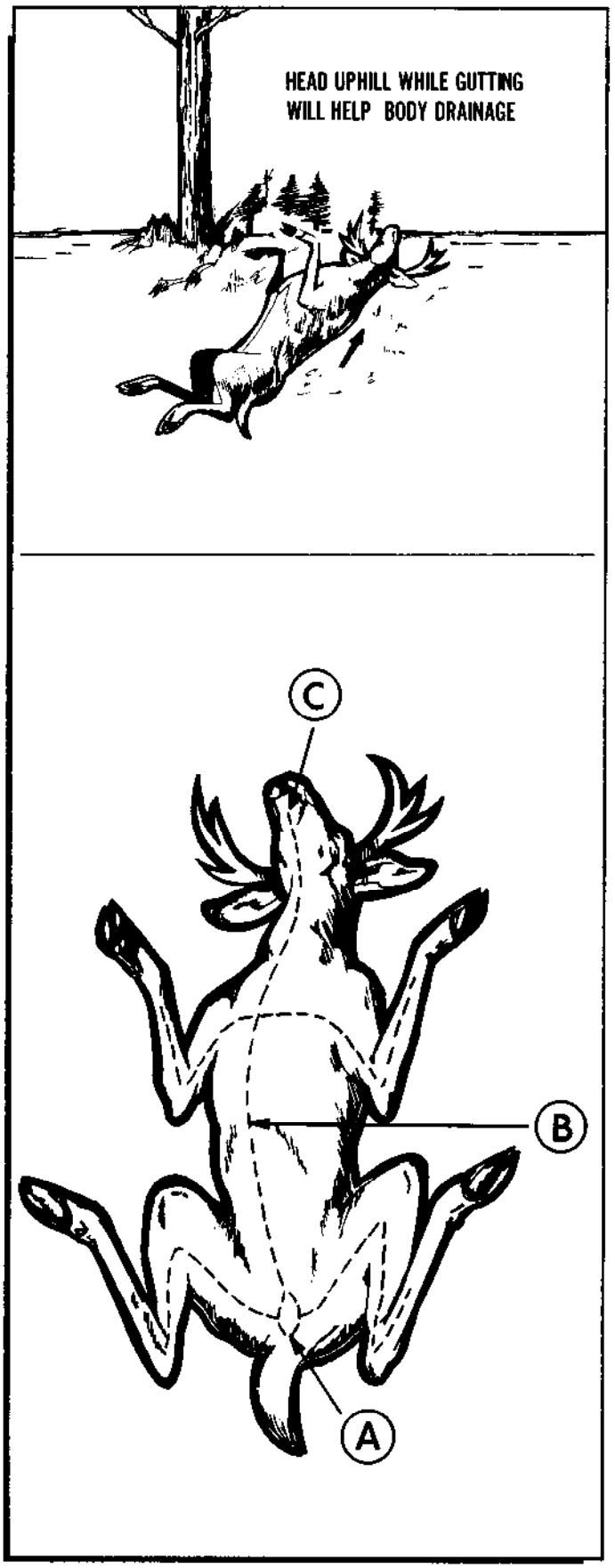

Figure 18-54. Big Game Skinning.

(g) Large animals should be quartered. To do this, cut down between the first and second rib and then sever the backbone with an axe or machete. Cut through the brisket of the front half and then chop lengthwise through the backbone to produce the front quarters. On the rear half, cut through the pelvic bone and lengthwise through the backbone. To make the load lighter and easier to transport, a knife could be used to bone the animal, thereby eliminating the weight of the bones. Butchering is the final step and is simplified for survival purposes. The main purpose is to cut the meat in manageable size portions (figure 18-57).

(2) Glove skinning is usually performed on small game (figure 18-58).

(a) The initial cuts are made down the insides of the back legs. The skin is then peeled back so that the hindquarters are bare and the tail is severed. To remove the remaining skin, pull it down over the body in much the same way a pullover sweater is removed. The head and front feet are severed to remove the skin from the body. For one-cut skinning of small game, cut across the lower back and insert two fingers under each side of the slit. By pulling quickly in opposite directions, the hide will be easily removed (figure 18-59).

(b) To remove the internal organs, a cut should be made into the abdominal cavity without puncturing the organs. This cut must run from the anus to the neck. There are muscles which connect the internal organs to the trunk and they must be severed to allow the viscera to be removed. A rabbit may be gutted by using a knifeless method with no mess and little time lost. Squeeze the entrails toward the rear resulting in a tight bulging abdomen. Raise the rabbit over the head and sling it down hard striking the forearms against the thighs. The momentum will expel the entrails through a tear in the vent (figure 18-60). Save the internal organs such as heart, liver, and kidneys, as they are nutritious. The liver should be checked for any white blotches and discarded if affected as these indicate tularemia (also known as rabbit fever). The disease is transmitted by rodents but also infects humans.

c. Cold-blooded animals are generally easy to clean and prepare.

(1) Snakes and lizards are very similar in taste and they have similar skin. Like the mammals, the skin and viscera should be removed. The easiest way to do this is to sever the head and (or) legs. In the case of a lizard, peel back enough skin so that it may be grasped securely and simply pull it down the length of the body turning the skin inside out as it goes. If the skin does not come away easily, a cut down the length of the animal can be made. This will allow the skin to part from the body more easily. The entrails are then removed and the animal is ready to cook.

(2) Except for the larger amphibians such as the bullfrog, the hind legs are the largest portion of the animal worth saving. To remove the hindquarters, sim-

ply cut through the backbone with a knife, leaving the abdomen and upper body. Pull the skin from the legs and they are ready to cook. With the bullfrogs and larger amphibians, the whole body can be eaten. The head, the skin, and viscera should be removed and discarded (use as bait to catch something else).

d. Most fish need little preparation before they are eaten. Scaling the fish before cooking is not necessary. A cut from the anus to the gills will expose the internal organs which should be removed. The gills should also be removed before cooking. The black line along the inside of the backbone is the kidney and should be removed by running a thumbnail from the tail to the head. There is some meat on the head and should not be discarded. See figure 18-61 for one method of filleting a fish.

e. All birds have feathers which can be removed in two ways: by plucking or pulling out the feathers, and by skinning. The gizzard, heart, and liver should be retained. The gizzard should be split open as it contains partially digested food and stones which must be discarded before being eaten.

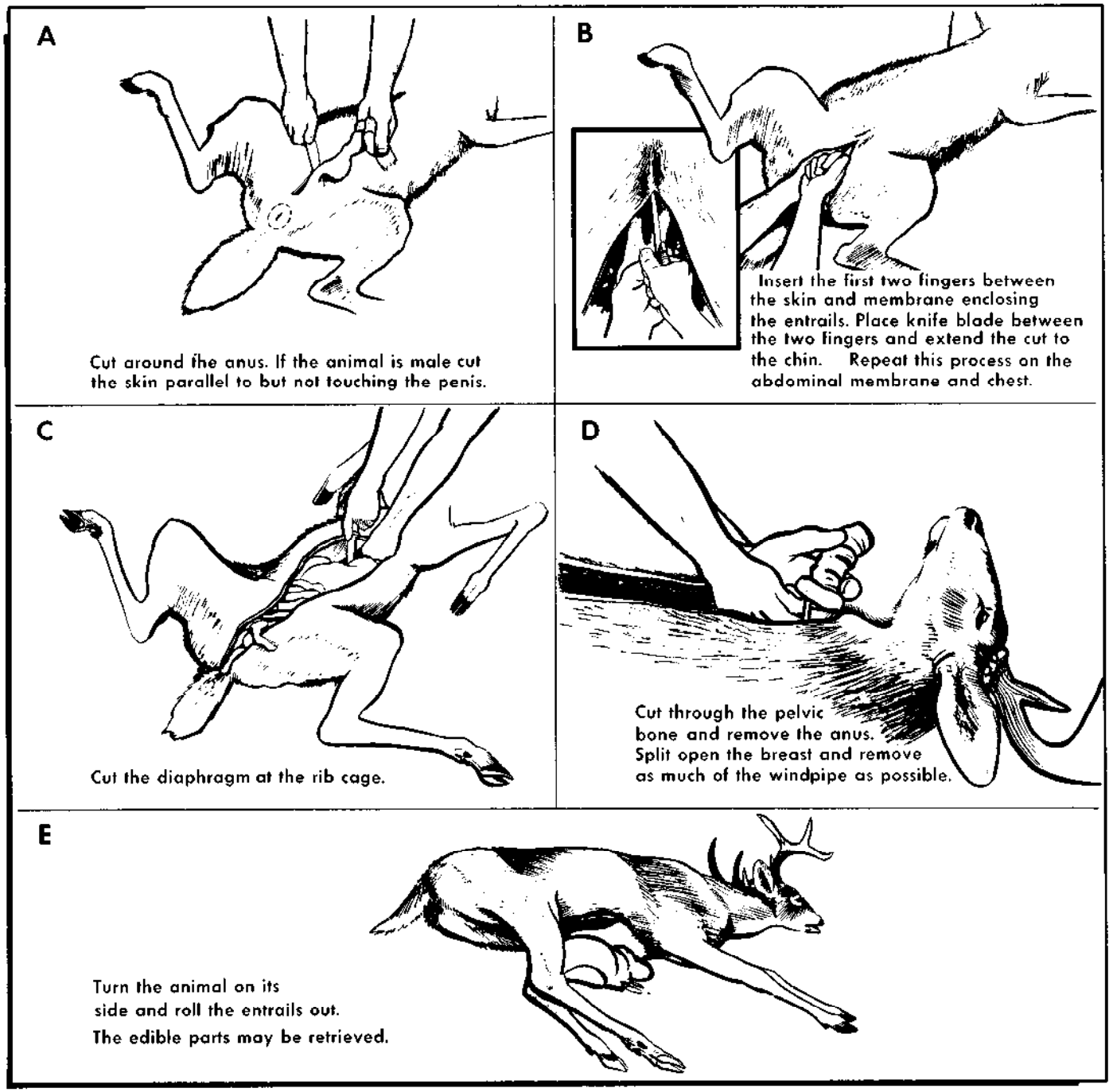

Figure 18-55. Field Dressing.

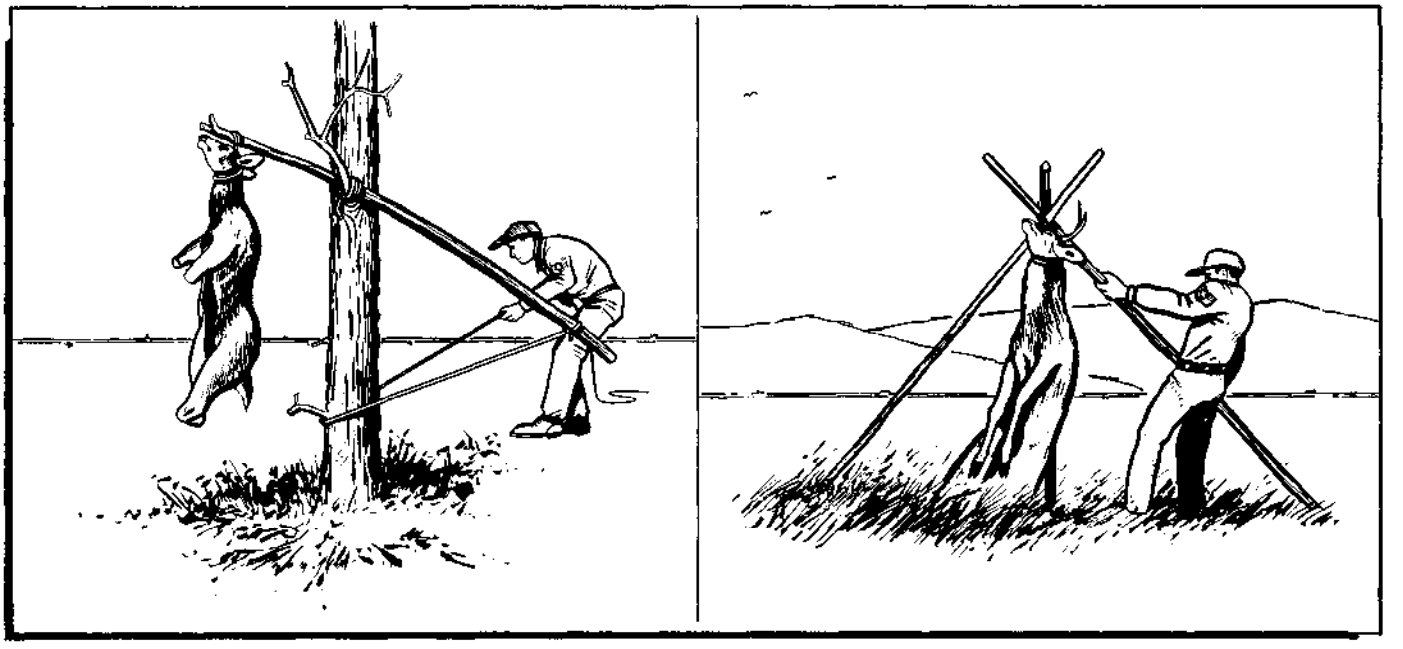

Figure 18-56. Hanging Game.

f. Insects are an excellent food source and they require little or no preparation. The main point to remember is to remove all hard portions such as the hind legs of a grasshopper and the hard wing covers of beetles. The rest is edible.

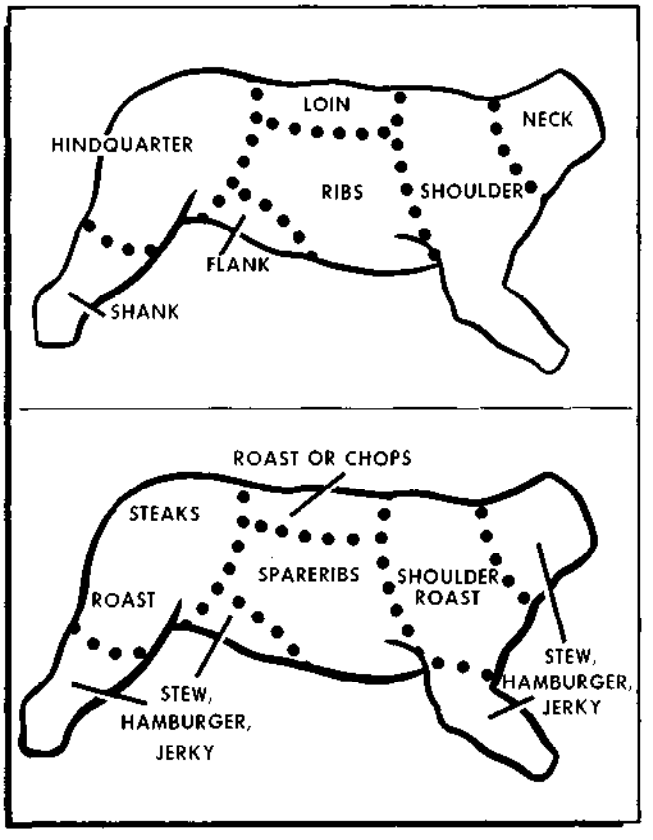

Figure 18-57. Butchering.

18-11. Cooking. All wild game, large insects (grasshoppers), freshwater fish, clams, mussels, snails, and crawfish must be thoroughly cooked to kill internal parasites. Mussels and large snails may have to be minced to make them tender.

a. Boiling is the most nutritious, simplest, and safest method of cooking (figure 18-62). Numerous containers can be used for boiling; for example, a metal container suspended above, or set beside, a heat source to boil foods. Green bamboo makes an excellent cooking container. Stone boiling is a method of boiling using super-heated rocks and a container that holds water but cannot be suspended over an open flame. Example of containers are survival kit containers, flying helmet, a hole in the ground lined with waterproof material, or a hollow log. The container is filled with food and water and then heated with super-hot stones until the water boils. Stones from a stream or damp area should not be used. The moisture in the stones may turn to steam and cause the stone to explode while the stones are being heated in the fire. The container should be covered and new stones added as the water stops boiling. The rocks can be removed with the aid of a wire secured to the rock before being put into the container or two sticks used in a chopstick fashion.

b. Baking is a good method of cooking as it is slow and is usually done by putting food into a container and cooking it slowly. Baking is often used with various types of ovens. Foods may be wrapped in wet leaves (figure 18-63) (avoid using a type of plant that will give an unpleasant flavor to what is being cooked), placed inside a metal container, or they may be packed with mud or clay and placed directly on the coals. Fish and

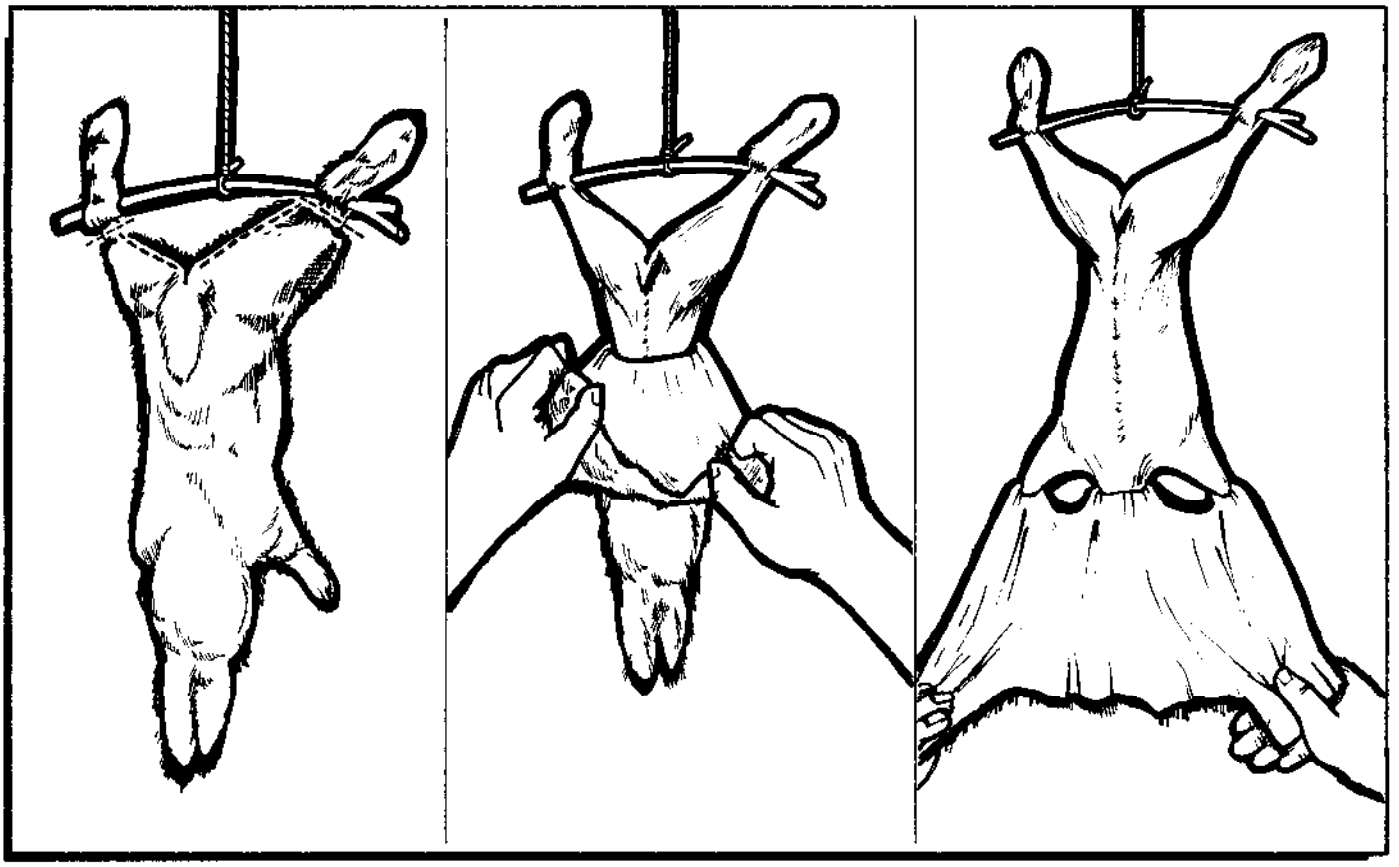

Figure 18-58. Glove Skinning.

birds packed in mud and baked must not be skinned because the scales, skin, or feathers will come off the animal when the mud or clay is removed. Clambake-style baking is done by heating a number of stones in a fire and allowing the fire to burn down to coals. A layer of wet seaweed or leaves is then placed over the hot rocks. Food such as mussels and clams in their shells are placed on the wet seaweed and (or) leaves (figure 18-64). More wet seaweed and (or) leaves and soil is used as a cover. When thoroughly steamed in their own juices, clam, oyster, and mussel shells will open and may be eaten without further preparation.

c. Any type of food can be cooked in the ground in a rock oven (figure 18-65). First, a hole is dug about 2 feet deep and 2 or 3 feet square, depending on the amount of food to be cooked. The sides and bottom are then lined with rock. Next, procure several green trees about 6 inches in diameter and long enough to bridge the hole. Firewood and grass or leaves for insulation should also be gathered. A fire is started in the hole. Two or three green trees are placed over the hole and several rocks are placed on the trees. The fire must be maintained until the green trees burn through. This indicates the fire has burned long enough to thoroughly heat the rocks and the oven is ready. The fallen rocks, fire, and ash are removed from the hole and a thin layer of dirt is spread over the bottom. The insulating material (grass, leaves, moss, etc.) is placed over the soil, then the food more insulating material on top and around the food, another thin layer of soil, and the extra hot rocks are placed on top. The hole is then filled with soil up to ground level.

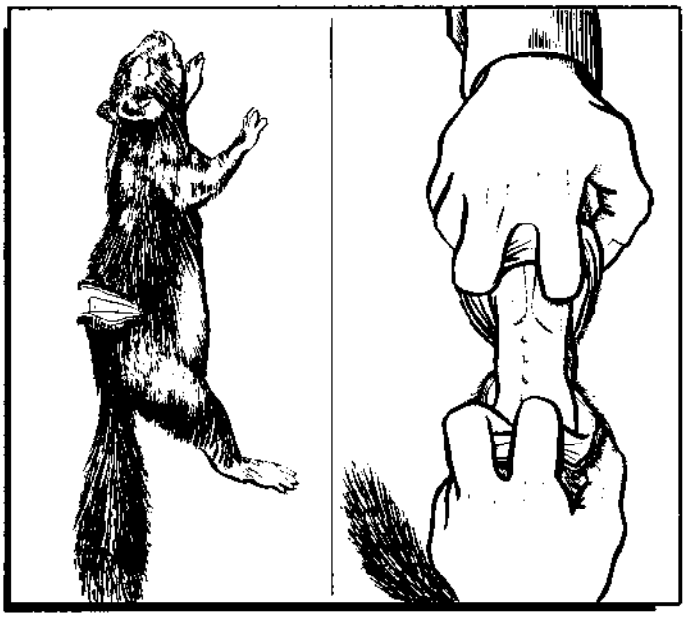

Figure 18-59. Small Animal Skinning.

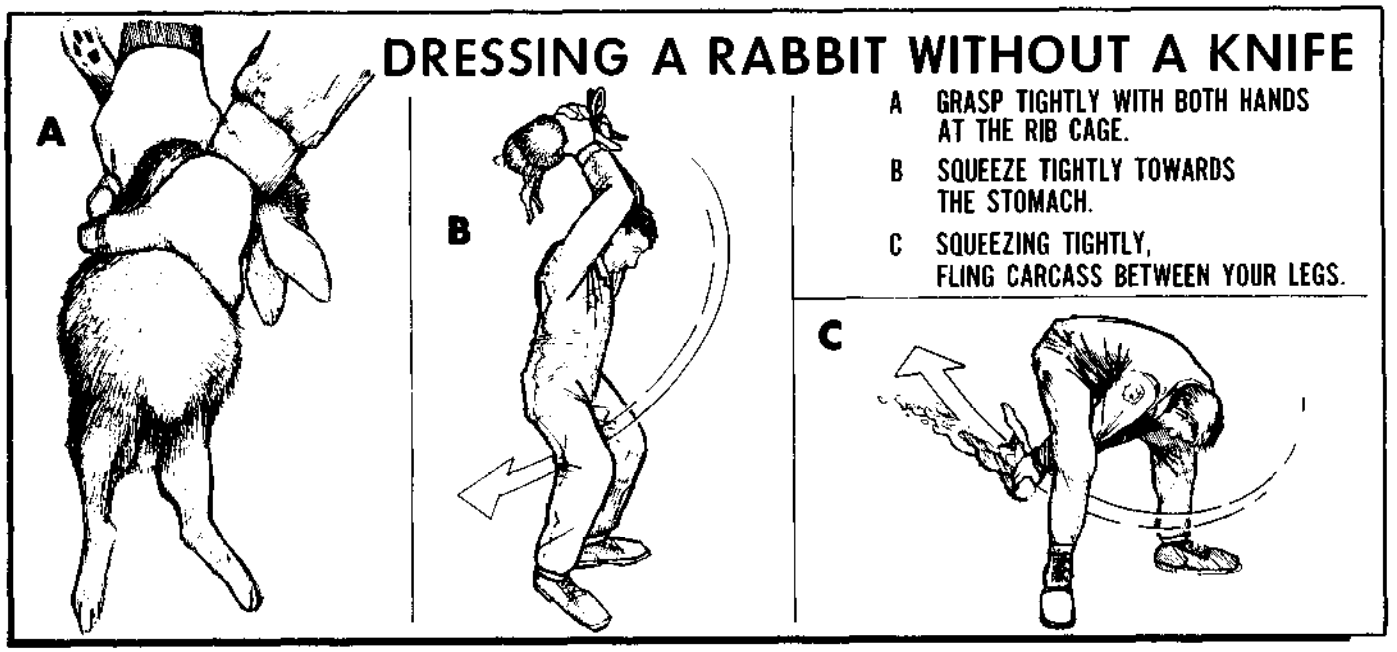

Figure 18-60. Dressing a Rabbit Without a Knife.

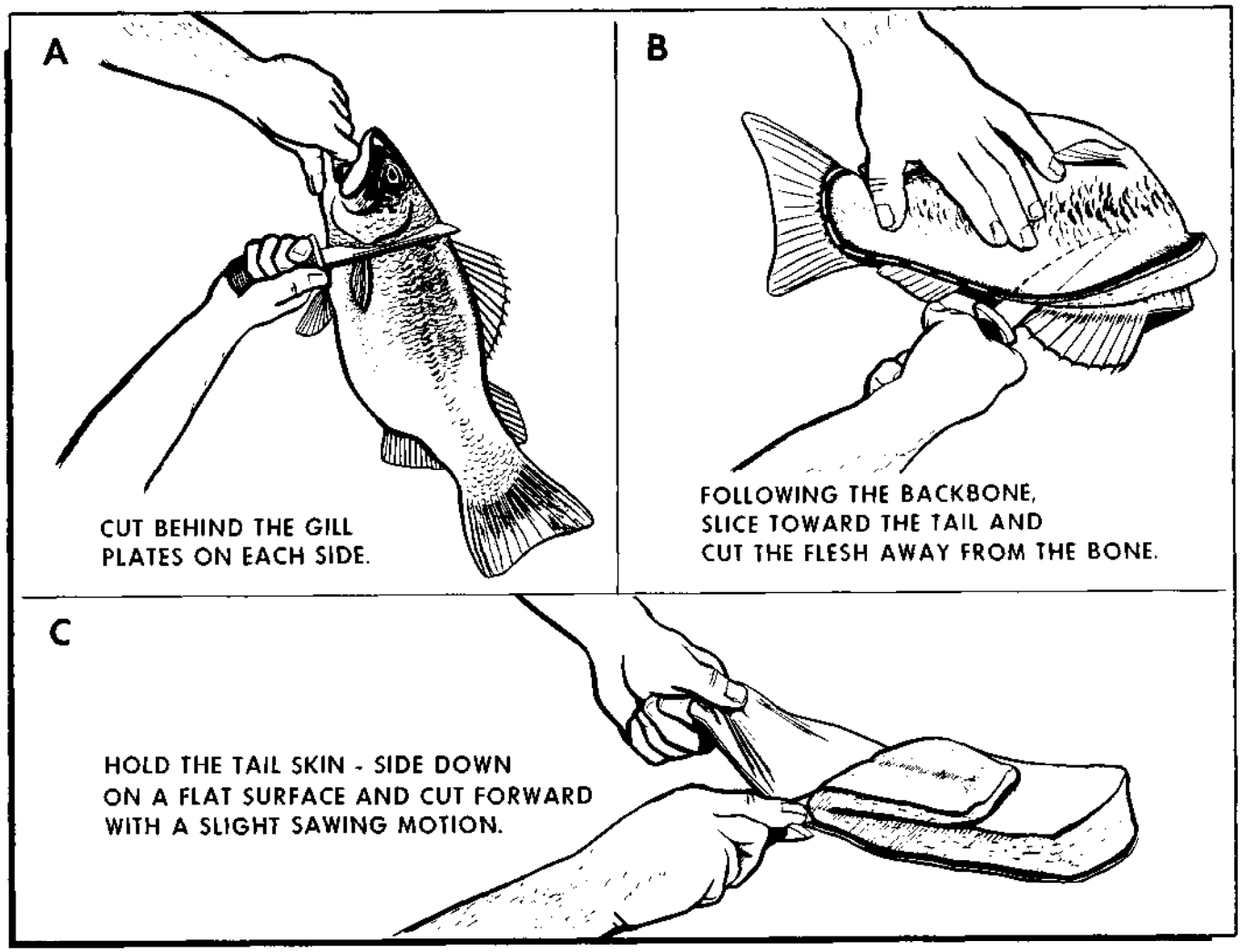

Figure 18-61. Filleting a Fish.

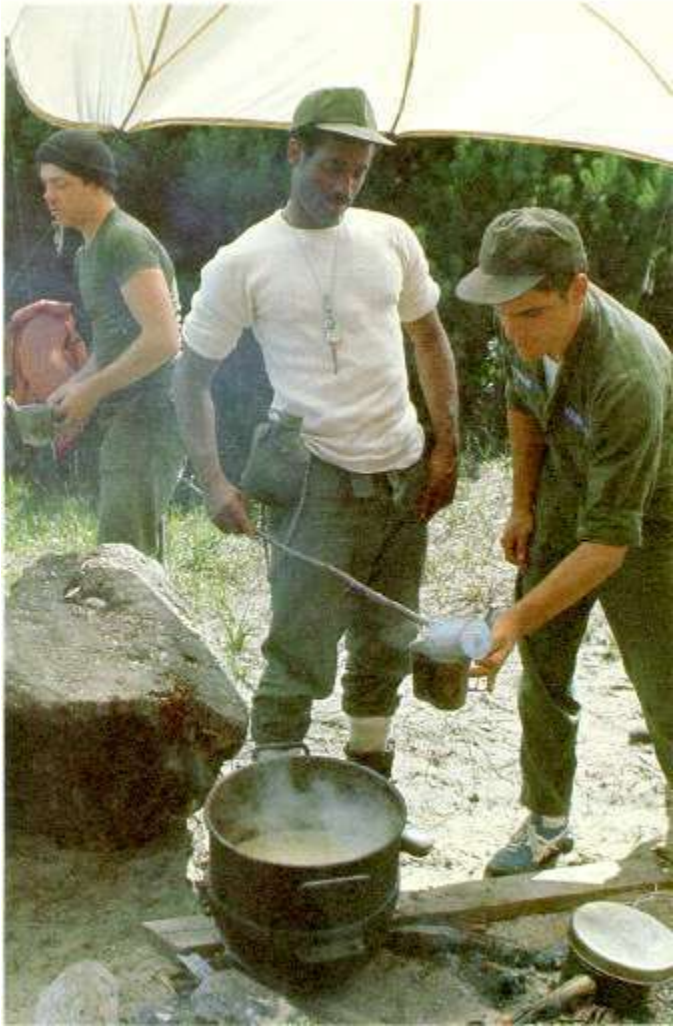

Figure 18-62. Boiling.

Small pieces of meat (steaks, chops, etc.) cook in 1½ to 2 hours and large pieces take 5 to 6 hours.

d. Roasting is less desirable as it involves exposing the food to direct heat which quickly destroys the nutritional properties (figure 18-66). Putting a piece of meat on a stick and holding it over the fire is considered roasting.

e. Broiling is the quickest way to prepare fish. A rock broiler may be made by placing a layer of small stones on top of hot coals, and laying the fish on the top. Scaling the fish before cooking is not necessary, and small fish need not be cleaned. Cooked in this manner, fish have a moist and delicious flavor. Crabs and lobsters may also be placed on the stones and broiled.

f. Meat may be cooked by laying it on a flat board or stone (planking) which is propped up close to the fire (figure 18-67). The meat will have to be turned over at least once to allow thorough cooking. The cooking time depends on how close the meat is to the fire.

g. Frying is by far the least favorable method of preparing food. It tends to make the meat tough because most all of the natural juices are cooked out of the meat. Some of the nutritional value of the meat will also be destroyed. Frying can be done on any nonporous surface which can be heated. Examples are unpainted aircraft parts, turtle shells, large seashells, flat rocks, and some survivial kit parts.

18-12. Preparing Food in Enemy Areas. The problem of preparing food in a hostile area becomes acute when a fire, even a small cooking fire, can bring about capture. After finding food in a hostile area, the problem of preparing the food in a manner which will not compromise the survivors presence must be resolved. Of course, it

Figure 18-63. Baking.

Figure 18-64. Clam Baking.

Figure 18-65. Rock Oven.

Figure 18-66. Broiling and Roasting.

would be simple to state that the best solution would be to eat the food without cooking it.

a. In some respects, this would be a more reasonable solution than it might initially seem to be. From the standpoint of palatability, it is mostly a matter of adjusting the "frame of mind." Animal foods are recognized as being palatable when cooked to a very minor degree. The need for food cannot be ignored and the situation may demand that it be eaten partially cooked or even uncooked.

b. With regard to the health considerations involved, many of the reasons for cooking are recognized as a means of destroying organisms that may be present in the food and can cause sickness or ill effects if they enter into the body. Under survival conditions in a hostile area, one may be forced to forego thorough cooking and accept the risk involved until their return to friendly forces where professional treatment is available.

c. Assuming that there will have to be a way to prepare food under hostile conditions, a survivor should be aware of some of the ways in order to achieve some degree of safety, and at the same time, improve palatability. Parasites and other organisms living in the flesh of the animals depend upon the body temperature of the animals, the moisture within the flesh of the animals, and other factors to support their life. Any action that modifies these conditions (for example, freezing or thorough drying of the meat) and kills some parasites may improve the palatability.

d. If cooking is considered necessary, use extreme care in selecting the site for a fire and ensure that security considerations are favorable. The food should be prepared in very small quantities in order to keep the size of the fire as small as possible. The use of the "Dakota Hole" configuration is more appropriate for cooking food during a tactical situation (figure 16-14).

18-13. Preserving Food. Finding natural foods is an uncertain aspect of survival. The survivor must make the best use of the available food. Food, especially meat, has a tendency to spoil within a short period of time unless it is preserved. There are many ways to preserve food; some of the most common are cooking, refrigeration, freezing, and dehydration.

Figure 18-67. Planking.

a. Cooking will slow down the decomposition of food but will not eliminate it. This is because many bacteria are present which work to break it down. Cooking methods which are the best for immediate consumption, such as boiling, are the least effective for preserving food. Food should be recooked every day until all is consumed.

b. Cooling is an effective method of storing food for short periods of time. Heat tends to accelerate the decomposition process where cooling retards decomposition. The colder food becomes, the less the likelihood of deterioration until freezing eliminates decomposition. Cooling devices available to a survivor are:

(1) Food items buried in snow will maintain a temperature of approximately 32°F.

(2) Food wrapped in waterproof material and placed in streams will remain cool in summer months. Care should be taken to ensure food is secured.

(3) Earth, below the surface, particularly in shady areas or along streams, remains cooler than the surface. A hole may be dug, lined with grass, and covered to form an effective cool storage area much the same as a root celler.

(4) When water evaporates, it tends to cool down the surrounding area. Using this fact, articles of food may be wrapped in an absorbent material such as cotton or burlap and rewetted as the water evaporates.

c. Once food is frozen, it will not decompose. Food should be frozen in meal-sized portions so refreezing is avoided.

d. Drying removes all moisture and preserves the food. Drying is done by sunning, smoking, or burying it in hot sand.

(1) For sun-drying, the food should be sliced very thin and placed in direct sunlight. Meat should be cut across the grain to improve tenderness and decrease drying time. If salt is available, it should be added to improve flavor and accelerate the drying process (figure 18-68).

(2) Smoking is a process done through the use of nonresinous wood such as willow or aspen and is used to produce smoke which adds flavor and dries the meat. A smoke rack is also necessary to contain the smoke (figure 18-69). The following are the procedures for drying meat using smoke:

(a) Cut meat very thin and across the grain. If the meat is warm and difficult to slice thin, cut the meat in 1 or 2-inch cubes and beat it thin with a clean wooden mallet (improvised).

(b) Remove fat.

(c) Hang the meat on a rack so each piece is separate.

(d) Elevate meat no less than 2 feet above coals.

Figure 18-68. Sun-Drying.

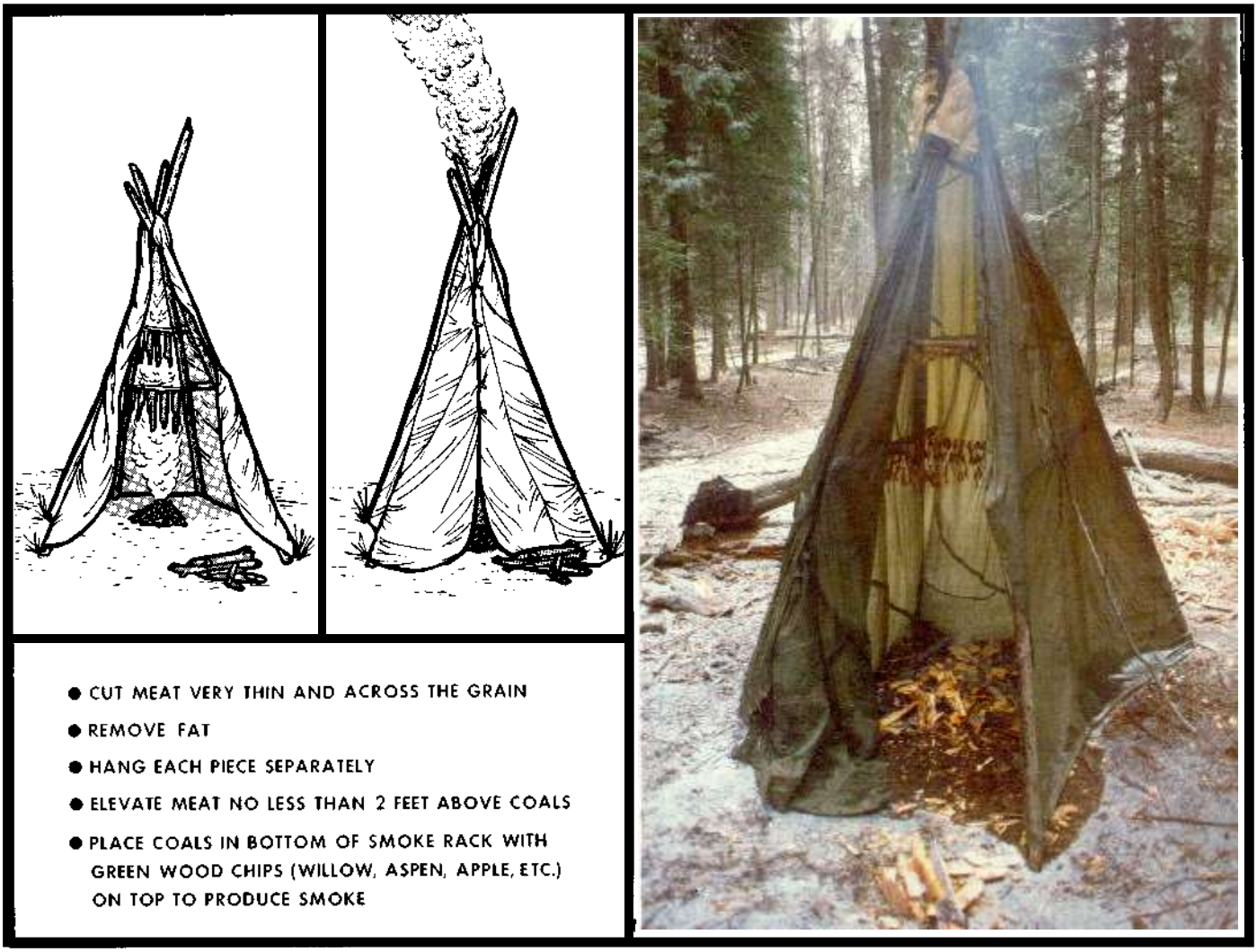

Figure 18-69. Smoke-Drying.

(e) Coals are placed in the bottom of a smoke rack with green woodchips on top to produce smoke.

e. The method used to preserve fish through warm weather is similar to that used in preserving meat (figure 18-70). When there is no danger of predatory animals disturbing the fish, the fish should be placed on available fabric and allowed to cool during the night. Early the next morning, before the air gets warm, the fish should be rolled in moist fabric (and leaves). This bundle can be placed inside the survivor's pack. During the rest periods, or when the pack is removed, it should be placed in a cool location out of the Sun's rays.

(1) Fish may be dried in the same manner described for smoking meat. To prepare fish for smoking, the heads and backbone are removed and the fish are spread flat on a grill. Thin willow branches with bark removed make skewers.

(2) Fish may also be dried in the Sun. They can be suspended from branches or spread on hot rocks. When the meat has dried, sea water or salt should be used on the outside, if available.

f. In survival environments, there are many animals and insects that will devour a survivor's food if it is not correctly stored. Protecting food from insects and birds is done by wrapping it in parachute material, wrapping and tying brush around the bundle, and finally, wrapping it with another layer of material. This creates "dead air" space making it more difficult for insects and birds to get to the food. If the outer layer is wetted, evaporation will also cool the food to some degree. In most cases, if the food is stored several feet off the ground, it will be out of reach of most animals. This can be done by hanging the food or putting it into a "cache". If the food is dehydrated, the container must be completely waterproof to prevent reabsorption. Frozen food will remain frozen only if the outside temperature remains below freezing. Burying food is a good way to store as long as scavengers are not in the area to uncover it. Insects and small animals should also be

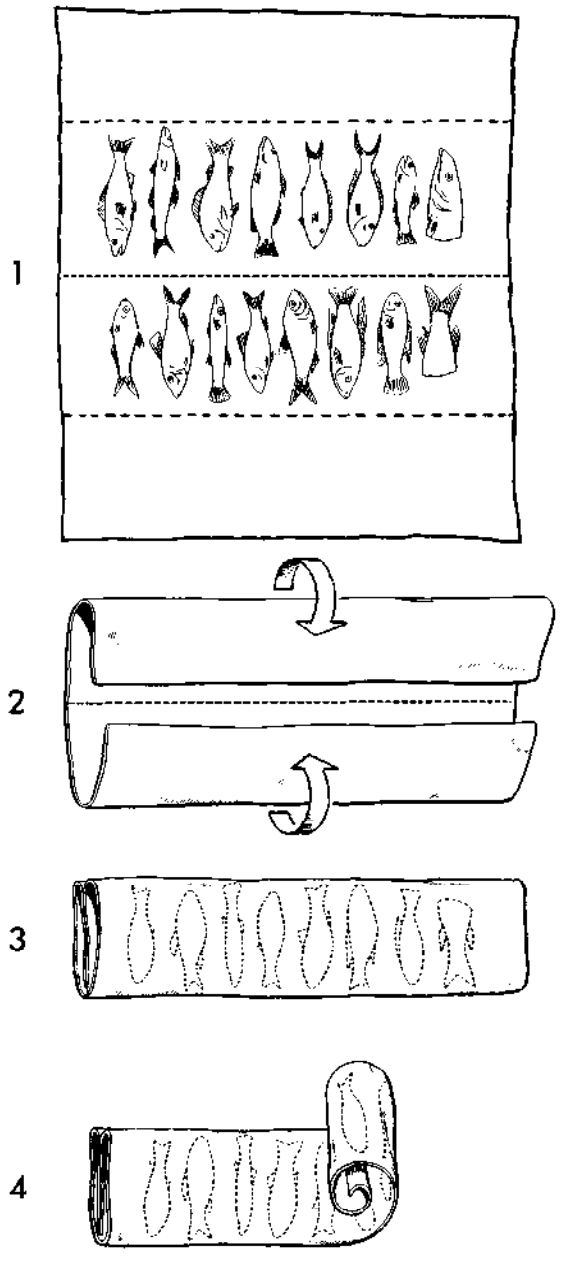

1. Arrange fish on available fabric.

2. Turn down the upper edge of wrap over the top line of fish and turn up the lower edge over the lower line.

3. Fold in the center as shown.

4. Then begin on the edge and roll the wrap. You will have a rounded roll of protected fish. This roll should be securely, but not tightly, tied and wrapped in a sleeping bag, parachute fabric or clothing, as you would do with meat.

Figure 18-70. Preserving Fish.

remembered when burying the food. Food should never be stored in the shelter as this may attract wild animals and could be hazardous to the survivors.

18-14. Preparing Plant Food. Preparing plant foods can be more involved than preparing animal life.

a. Some plant foods, such as acorns and tree bark may be bitter because of tannin. These plants will require leaching by chopping up the plant parts, and pouring several changes of fresh water over them. This will help wash out the tannin, making the plant more palatable. Other plants such as cassava and green papaya must be cooked before eating to break down the harmful enzymes and chemical crystals within them and make them safe to eat. Plants such as skunk cabbage must undergo this cooking process several times before it is safe to eat.

b. All starchy foods must be cooked since raw starch is difficult to digest. They are boiled, steamed, roasted, or fried and are eaten plain, or mixed with other wild foods. The manioc (cassava) is best cooked, because the bitter form (green stem) is poisonous when eaten raw. Starch is removed from sago palm, cycads, and other starch-producing trunks by splitting the trunk and pounding the soft, whitish inner parts with a pointed club. This pulp is washed with water and the white sago (pure starch) is drained into a container. It is washed a second time, and then it may be used directly as a flour. One trunk of the sago palm will supply a survivor's starch needs for many weeks.

c. The fiddleheads of all ferns are the curled, young succulent fronds which have the same food value as cabbage or asparagus. Practically all types of fiddleheads are covered with hair which makes them bitter. The hair can be removed by washing the fiddleheads in water. If fiddleheads are especially bitter, they should be boiled for 10 minutes and then reboiled in fresh water for 30 to 40 minutes. Wild bird eggs or meat may be cooked with the fiddleheads to form a stew.

d. Wild grasses have an abundance of seeds, which may be eaten boiled or roasted after separating the chaff from the seeds by rubbing. No known grass is poisonous. If the kernels are still soft and do not have large stiff barbs attached, they may be used for porridge. If brown or black rust is present, the seeds should not be eaten (Ergot Poisoning). To gather grass seeds, a cloth is placed on the ground and the grass heads beaten with sticks.

e. Plants that grow in wet places along margins of rivers, lakes, and ponds, and those growing directly in water are of potential value as survival food. The succulent underground parts and stems are most frequently eaten. Poisonous water plants are rare. In temperate climates, the water hemlock is the most poisonous plant found around marshes and ponds. In the tropics, the various members of the calla lily family often grow in very wet places. The leaves of the calla lily look like

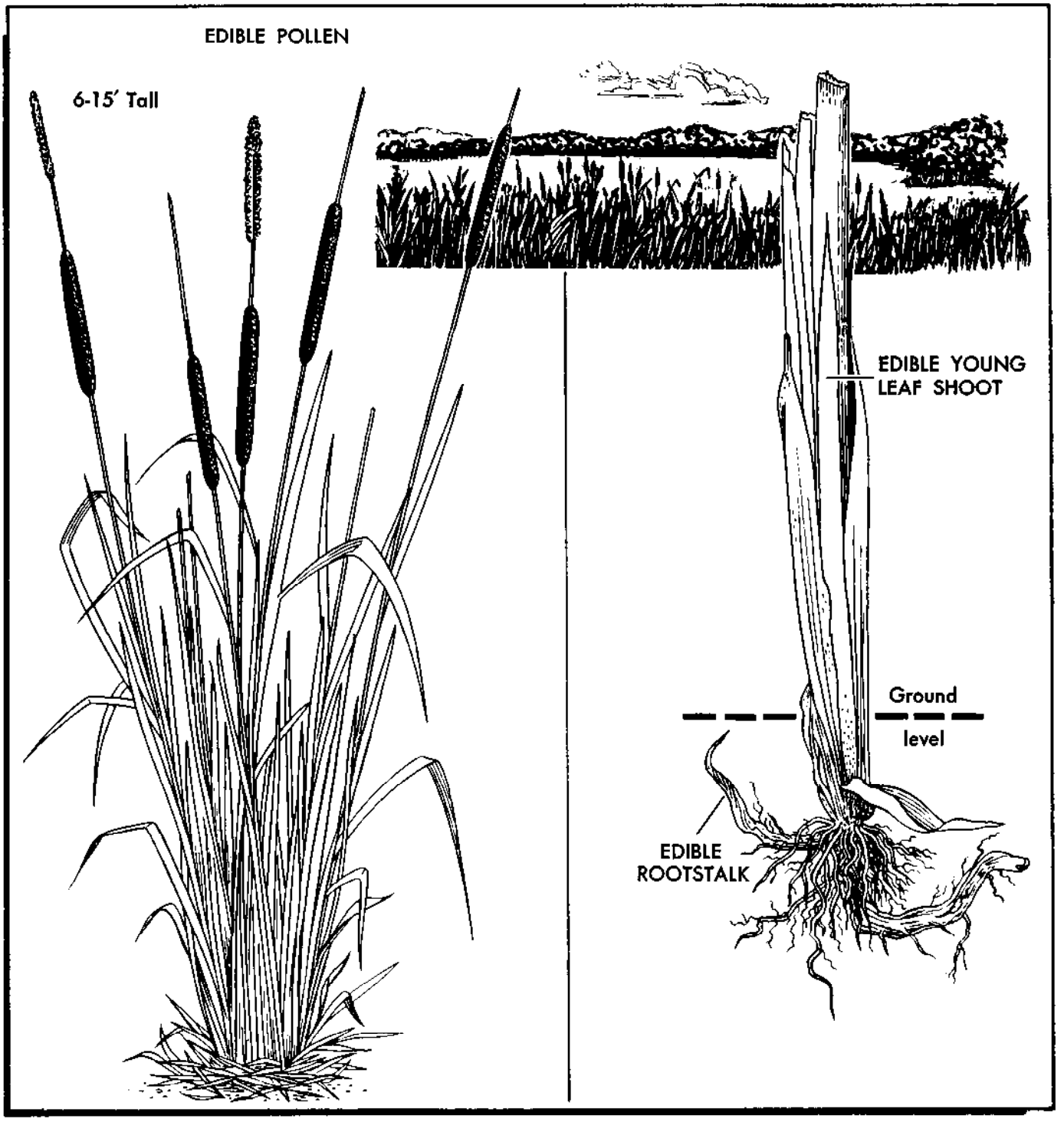

Figure 18-71. Cattails.

arrowheads. Jack-in-the-pulpit, calla lily, and sweet flag are members of the Arum family. To be eaten, the members of this plant family must be cooked in frequent changes of water to destroy the irritant crystals in the stems. Two kinds of marsh and water plants are the cattail and the water lily.

(1) The cattail (Typha) is found worldwide except in tundra regions of the far north (figure 18-71). Cattails can be found in the more moist places in desert areas of all continents as well as in the moist tropic and temperate zones of both hemispheres. The young shoots taste like asparagus. The spikes can be boiled or steamed when green and then eaten. The rootstalks, without the outer covering, are eaten boiled or raw. Cattail roots can be cut into thin strips, dried, and then ground into flour. They are 46 percent starch, 11 percent sugar, and the rest is fiber. While the plant is in flower, the yellow pollen is very abundant; this may be mixed with water

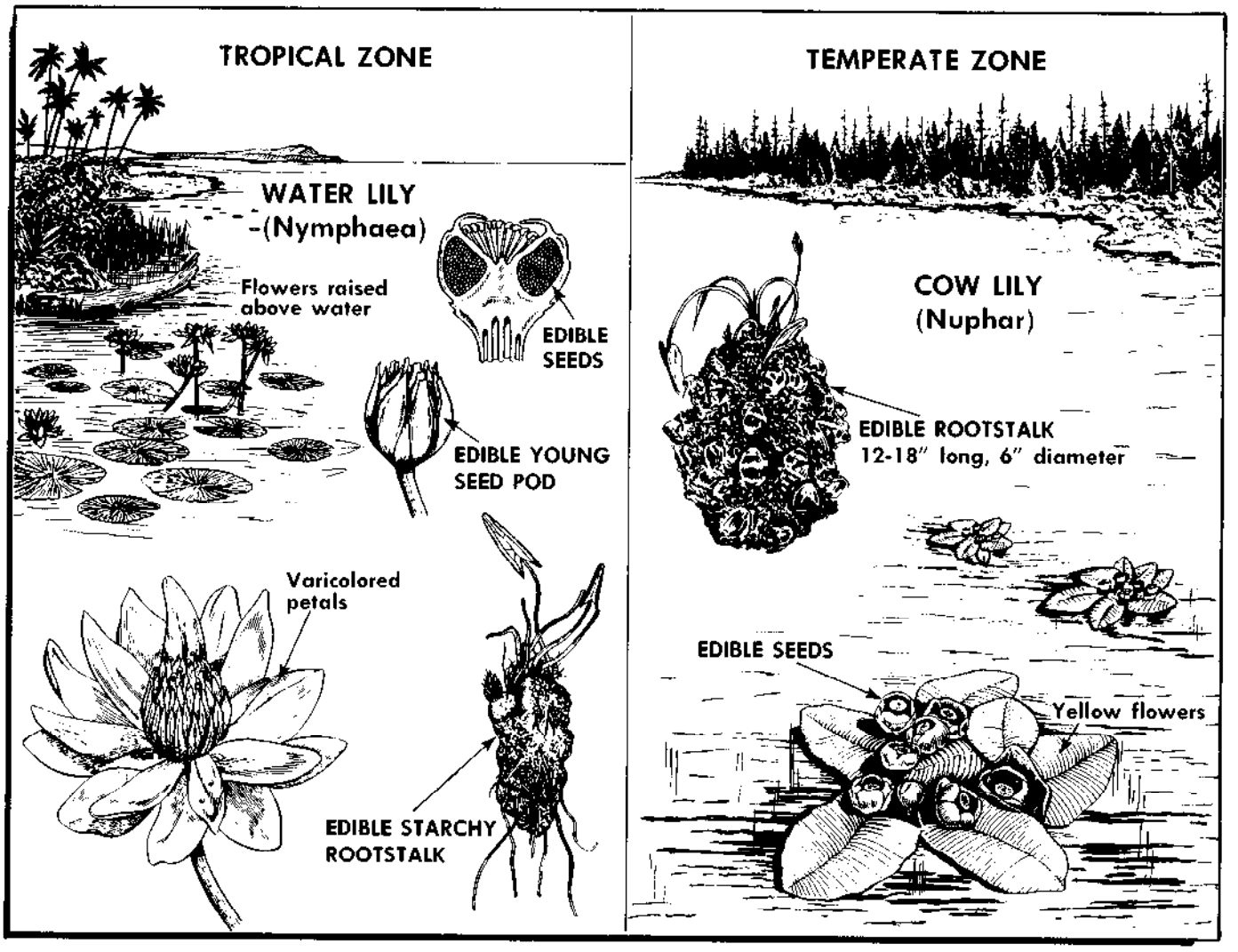

Figure 18-72. Water Lilies.

and made into small cakes and steamed as a substitute for bread.

(2) Water lilies (Nymphaea and Nuphar) occur on all the continents, but principally in southern Asia, Africa, North America, and South America (figure 18-72). Two main types are:

(a) Temperate water lilies produce enormous rootstalks and yellow or white flowers which float on the water.

(b) Tropical water lilies produce large edible tubers and flowers which are elevated above the water surface.

(3) Rootstalks or tubers may be difficult to obtain because of deep water. They are starchy and high in food value. They can be eaten either raw or boiled. Stems may be cooked in a stew. Young seed pods may be sliced and eaten as a vegetable. Seeds may be bitter, but are very nourishing. They may be parched and rubbed between stones as flour. The water lily is considered an important food source by native peoples in many parts of the world.

f. Nuts are very high in nutritional value and usually can be eaten raw. Nuts may be roasted in the fire or roasted by shaking them in a container with hot coals from the fire. They may then be ground to make a flour. If a survivor does not wish to eat a plant or plant part raw, it can be cooked using the same methods used in cooking meat—by boiling, roasting, baking, broiling, or frying.

g. If survivors have been able to procure more plant foods than can be eaten, the excess can be preserved in the same manner as animal foods. Plant foods can be dried by wind, air, sun, or fire, with or without smoke. A combination of these methods can be used. The main object is to remove the moisture. Most wild fruits can be dried. If the plant part is large, such as some tubers, it should be sliced, and then dried. Some type of protection may be necessary to prevent consumption and (or) contamination by insects. Extra fruits or berries can be carried with the survivor by wrapping them in leaves or moss.

Chapter 19

WATER

19-1. Introduction. Nearly every survival account details the need survivors had for water. Many ingenious methods of locating, procuring, purifying, and storing water are included in the recorded experiences of downed aircrew members. If survivors are located in temperate, tropic, or dry climates, water may be their first and most important need. The priority of finding water over that of obtaining food must be emphasized to potential survivors. An individual may be able to live for weeks without food, depending on the temperature and amount of energy being exerted. A person who has no water can be expected to die within days. Even in cold climate areas or places where water is abundant, survivors should attempt to keep their body fluids at a level that will maintain them in the best possible state of health. Even in relatively cold climates, the body needs 2 quarts of water per day to remain efficient (figure 19-1).

19-2. Water Requirements. Normally, with atmospheric temperature of about 68°F, the average adult requires 2 to 3 quarts of water daily.

a. This water is necessary to replace that lost daily in the following ways:

(1) Urine. Approximately 1.4 quart of water is lost in the urine.

(2) Sweat. About 0.1 quart of water is lost in the sweat.

(3) Feces. Approximately 0.2 quart of water is lost in the feces.

(4) Insensible Water Loss. When the individual is unaware water loss is actually occurring, it is referred to as insensible water loss. Insensible water loss occurs by the following mechanisms:

(a) Diffusion through the skin. Water loss through the skin occurs as a result of the actual diffusion of water molecules through the cells of the skin. The average loss of water in this manner is approximately 0.3 to 0.4 quart. Fortunately, loss of greater quantities of water by diffusion is prevented by the outermost layer of the skin, the epidermis, which acts as a barrier to this type of water loss.

(b) Evaporation through the lungs. Inhaled air initially contains very little water vapor. However, as soon as it enters the respiratory passages, the air is exposed to the fluids covering the respiratory surfaces. By the time this air enters the lungs, it has become totally saturated with moisture from these surfaces. When the air is exhaled, it is still saturated with moisture and water is lost from the body.

b. Larger quantities of water are required when water loss is increased in any one of the following circumstances:

(1) Heat Exposure. When an individual is exposed to very high temperatures, water lost in the sweat can be increased to as much as 3.5 quarts an hour. Water loss at this increased rate can deplete the body fluids in short time.

(2) Exercise. Physical activity increases the loss of water in two ways as follows:

(a) The increased respiration rate causes increased water loss by evaporation through the lungs.

(b) The increased body heat causes excessive sweating.

(3) Cold Exposure. As the temperature decreases, the amount of water vapor in the air also decreases. Therefore, breathing cold air results in increased water loss by evaporation from the lungs.

(4) High Altitude. At high altitudes, increased water loss by evaporation through the lungs occurs not only as a result of breathing cooler air but also as a result of the increased respiratory efforts required.

(5) Burns. After extensive burns, the outermost layer of the skin is destroyed. When this layer is gone, there is no longer a barrier to water loss by diffusion, and the rate of water loss in this manner can increase up to 5 quarts each day.

(6) Illness. Severe vomiting or prolonged diarrhea can lead to serious water depletion.

c. Dehydration (body fluid depletion) can occur when required body fluids are not replaced.

(1) Dehydration is accompanied by the following symptoms:

(a) Thirst.
(b) Weakness.
(c) Fatigue.
(d) Dizziness.
(e) Headache.
(f) Fever.
(g) Inelastic abdominal skin.
(h) Dry mucous membranes, that is, dry mouth and nasal passages.
(i) Infrequent urination and reduced volume. The urine is concentrated so that it is very dark in color. In severe cases, urination may be quite painful.

(2) Companions will observe the following behavioral changes in individuals suffering from dehydration:

(a) Loss of appetite.
(b) Lagging pace.
(c) Impatience.
(d) Sleepiness.
(e) Apathy.
(f) Emotional instability.
(g) Indistinct speech.
(h) Mental confusion.

(3) Dehydration is a complication which causes decreased efficiency in the performance of even the simplest task. It also predisposes survivors to the development of severe shock following minor injuries. Constriction of blood vessels in the skin as a result of dehydration increases the danger of cold injury during

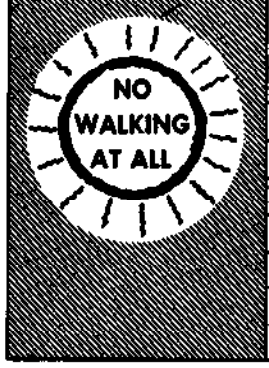

MAXIMUM DAILY TEMPERATURE (°F) IN SHADE	AVAILABLE WATER PER MAN, US QUARTS					
	0	1 Qt	2 Qts	4 Qts	10 Qts	20 Qts
	DAYS OF EXPECTED SURVIVAL					
120°	2	2	2	2.5	3	4.5
110	3	3	3.5	4	5	7
100	5	5.5	6	7	9.5	13.5
90	7	8	9	10.5	15	23
80	9	10	11	13	19	29
70	10	11	12	14	20.5	32
60	10	11	12	14	21	32
50	10	11	12	14.5	21	32

MAXIMUM DAILY TEMPERATURE (°F) IN SHADE	AVAILABLE WATER PER MAN, US QUARTS					
	0	1 Qt	2 Qts	4 Qts	10 Qts	20 Qts
	DAYS OF EXPECTED SURVIVAL					
120°	1	2	2	2.5	3	
110	2	2	2.5	3	3.5	
100	3	3.5	3.5	4.5	5.5	
90	5	5.5	5.5	6.5	8	
80	7	7.5	8	9.5	11.5	
70	7.5	8	9	10.5	13.5	
60	8	8.5	9	11	14	
50	8	8.5	9	11	14	

Figure 19-1. Water Requirements.

cold exposure. Failure to replace body fluids ultimately results in death.

(a) Proper treatment for dehydration is to replace lost body fluids. The oral intake of water is the most readily available means of correcting this deficiency. A severely dehydrated person will have little appetite. This person must be encouraged to drink small quantities of water at frequent intervals to replenish the body's fluid volume. Cold water should be warmed so the system will accept it easier.

(b) To prevent dehydration, water loss must be replaced by periodic intake of small quantities of water throughout the day. As activities or conditions intensify, the water intake should be increased accordingly. Water intake should be sufficient to maintain a minimum urinary output of 1 pint every 24 hours. Thirst is not an adequate stimulus for water intake, and a person often dehydrates when water is available. Therefore, water intake should be encouraged when the person is not thirsty. Humans cannot adjust to decreased water intake for prolonged periods of time. When water is in short supply, any available water should be consumed sensibly. If sugar is available, it should be mixed with the water, and efforts should be made to find a local water source. Until a suitable water source is located, individual water losses should be limited in the following ways:

-1. Physical activity should be limited to the absolute minimum required for survival activities. All tasks should be performed slowly and deliberately with minimal expenditure of energy. Frequent rest periods should be included in the daily schedule.

-2. In hot climates, essential activity should be conducted at night or during the cooler part of the day.

-3. In hot climates, clothing should be worn at all times because it reduces the quantity of water loss by sweating. Sweat is absorbed into the clothing evaporated from its surface in the same manner it evaporates from the body. This evaporation cools the air trapped between the clothing and the skin, causing a decrease in

the activity of the sweat glands and a subsequent reduction in water loss.

-4. In hot weather, light-colored clothing should be worn rather than dark-colored clothing. Dark-colored clothing absorbs the Sun's light rays and converts them into heat. This heat causes an increase in body temperatures which activates the sweat glands and increases water loss through sweating. Light-colored clothing, however, reflects the Sun's light rays, minimizing the increase in body temperature and subsequent water loss.

19-3. Water Sources. Survivors should be aware of both the water sources available to them and the resources at their disposal for producing water.

a. Survivors may obtain water from solar stills, desalter kits, or canned water packed in various survival kits. It would be wise for personnel, who may one day have to use these methods of procuring water, to be knowledgeable of their operating instructions and the amount of water they produce.

(1) Canned water provides 10 ounces per can.

(2) Desalter kits are limited to 1 pint per chemical bar—kits contain eight chemical bars.

(3) A "sea solar still" can produce as much as 2½ pints per day.

(4) "Land solar stills" produce varied amounts of water. This amount is directly proportionate to the amount of water available in the soil or placed into the still (vegetation, entrails, contaminated water, etc.), and the ambient temperature.

b. Aircrew members would be wise to carry water during their missions. Besides the fact that the initial shock of the survival experience sometimes produces feelings of thirst, having an additional water container can benefit survivors. The issued items (canned water, desalter kits, and solar stills) should be kept by survivors for times when no natural sources of freshwater are available.

c. Naturally occurring indicators of water are:

(1) Surface water, including streams, lakes, springs, ice, and snow.

(2) Precipitation, such as rain, snow, dew, sleet, etc.

(3) Subsurface water, which may not be as readily accessible as wells, cisterns, and underground springs and streams, can be difficult for survivors to locate and use.

d. Several indicators of possible water are:

(1) Presence of abundant vegetation of a different variety, such as deciduous growth in a coniferous area.

(2) Drainages and low-lying areas.

(3) Large clumps of plush grass.

(4) Animal trails which may lead to water. The "V" formed by intersecting trails often point toward water sources.

e. Survivors may locate and procure water as follows:

(1) Precipitation may be procured by laying a piece of nonporous material such as a poncho, piece of canvas, plastic, or metal material on the ground. If rain or snow is being collected, it may be more efficient to create a bag or funnel shape with the material so the water can be easily gathered. Dew can be collected by wiping it up with a sponge or cloth first, and then wringing it into a container (figure 19-2). Consideration should be given to the possibility of contaminating the water with dyes, preservatives, or oils on the surfaces of the objects used to collect the precipitation. Ice will yield more water per given volume than snow and requires less heat to do so. If the Sun is shining, snow or ice may be placed on a dark surface to melt (dark surfaces absorb heat, whereas light surfaces reflect heat). Ice can be found in the form of icicles on plants and trees, sheet ice on rivers, ponds, and lakes, or sea ice. If snow must be used, survivors should use snow closest to the ground. This snow is packed and will provide more water for the amount of snow than will the upper layers. When snow is to be melted for water, place a small amount of snow in the bottom of the container being used and place it over or near a fire. Snow can be added a little at a time. Survivors should allow water in the container bottom to become warm so that when more snow is added, the mixture remains slushy. This will prevent burning the bottom out of the container. Snow absorbs water, and if packed, forms an insulating airspace at the bottom of the container. When this happens, the bottom may burn out.

(2) Several things may help survivors locate ground water, such as rivers, lakes, and streams.

(a) The presence of swarming insects indicates water is near. In some places, survivors should look for signs of animal presence. For example, in damp places, animals may have scratched depressions into the ground to obtain water; insects may also hover over these areas.

(b) In the Libyan Sahara, donut-shaped mounds of camel dung often surround wells or other water sources. Bird flights can indicate direction to or from water. Pigeons and doves make their way to water regularly. They fly from water in the morning and to it in the evening. Large flocks of birds may also congregate around or at areas of water.

(c) The presence of people will indicate water. The location of this water can take many forms—stored water in containers that are carried with people who are traveling, wells, irrigation systems, pools, etc. Survivors who are evaders should be extremely cautious when approaching any water source, especially if they are in dry areas; these places may be guarded or inhabited.

(3) When no surface water is available, survivors may have to tap the Earth's supply of ground water. Access to this depends upon the type of ground—rock or loose material, clay, gravel, or sand.

(a) In rocky ground, survivors should look for springs and seepages. Limestone and lava rocks will have more and larger springs than any other rocks. Most

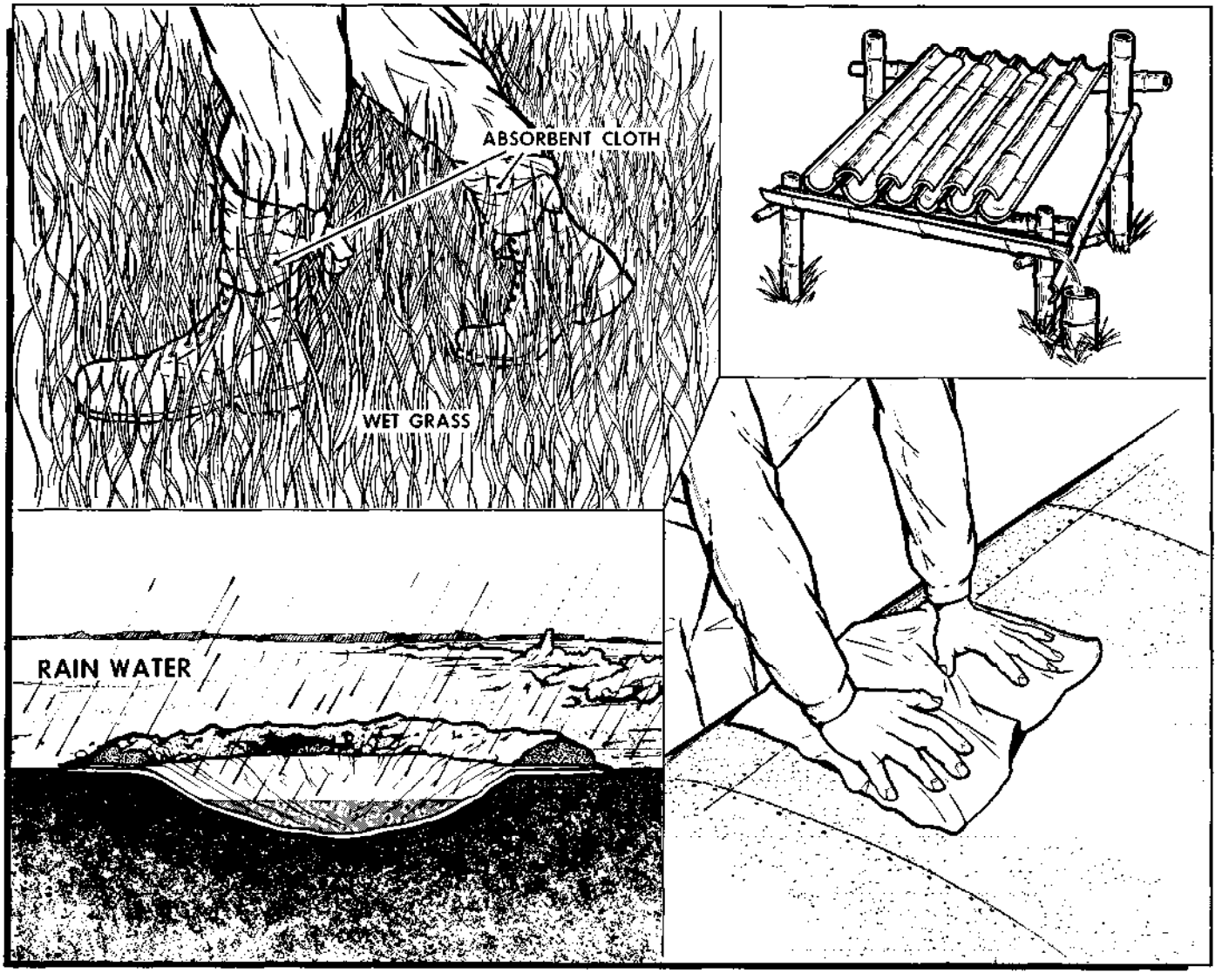

Figure 19-2. Methods of Procuring Water.

lava rocks contain millions of bubble holes; ground water may seep through them. Survivors can also look for springs along the walls of valleys that cross a lava flow. Some flows will have no bubbles but do have "organ pipe" joints—vertical cracks that part the rocks into columns a foot or more thick and 20 feet or more high. At the foot of these joints, survivors may find water creeping out as seepage, or pouring out in springs.

(b) Most common rocks, like granite, contain water only in irregular cracks. A crack in a rock with bird dung around the outside may indicate a water source that can be reached by a piece of surgical hose used as a straw or siphon.

(c) Water is more abundant and easier to find in loose sediments than in rocks. Springs are sometimes found along valley floors or down along their sloping sides. The flat benches or terraces of land above river valleys usually yield springs or seepages along their bases, even when the stream is dry. Survivors shouldn't waste time digging for water unless there are signs that water is available. Digging in the floor of a valley under a steep slope, especially if the bluff is cut in a terrace, can produce a water source. A lush green spot where a spring has been during the wet season is a good place to dig for water. Water moves slowly through clay, but many clays contain strips of sand which may yield springs. Survivors should look for a wet place on the surface of clay bluffs and try digging it out.

(d) Along coasts, water may be found by digging beach wells (figure 19-3). Locate the wells behind the first or second pressure ridge. Wells can be dug 3 to 5 feet deep and should be lined with driftwood to prevent sand from refilling the hole. Rocks should be used to line the bottom of the well to prevent stirring up sand

when procuring the water. The average well may take as long as 2 hours to produce 4 to 5 gallons of water. (Do not be discouraged if the first try is unsuccessful—dig another.)

19-4. Water in Snow and Ice Areas. Due to the extreme cold of arctic areas, water requirements are greatly increased. Increased body metabolism, respiration of cold air, and extremely low humidity play important roles in reducing the body's water content. The processes of heat production and digestion in the body also increase the need for water in colder climatic zones. The constructing of shelters and signals and the obtaining of firewood are extremely demanding tasks for survivors. Physical exertion and heat production in extreme cold place the water requirements of a survivor close to 5 or 6 quarts per day to maintain proper hydration levels. The diet of survivors will often be dehydrated rations and high protein food sources. For the body to digest and use these food sources effectively, increased water intake is essential.

a. Obtaining water need not be a serious problem in the arctic because an abundant supply of water is available from streams, lakes, ponds, snow, and ice. All surface water should be purified by some means. In the summer, surface water may be discolored but is drinkable when purified. Water obtained from glacier-fed rivers and streams may contain high concentrations of dirt or silt. By letting the water stand for a period of time, most silt will settle to the bottom; the remaining water can be strained through porous material for further filtration.

b. A "water machine" can be constructed which will produce water while the survivors are doing other tasks. It can be made by placing snow on any porous material (such as parachute or cotton), gathering up the edges, and suspending the "bag" of snow from any support near the fire. Radiant heat will melt the snow and the water will drip from the lowest point on the bag. A container should be placed below this point to catch the water (figure 19-4).

c. In some arctic areas, there may be little or no fuel supply with which to melt ice and snow for water. In this case, body heat can be used to do the job. The ice or snow can be placed in a waterproof container like a waterbag and placed between clothing layers next to the body. This cold substance should not be placed directly

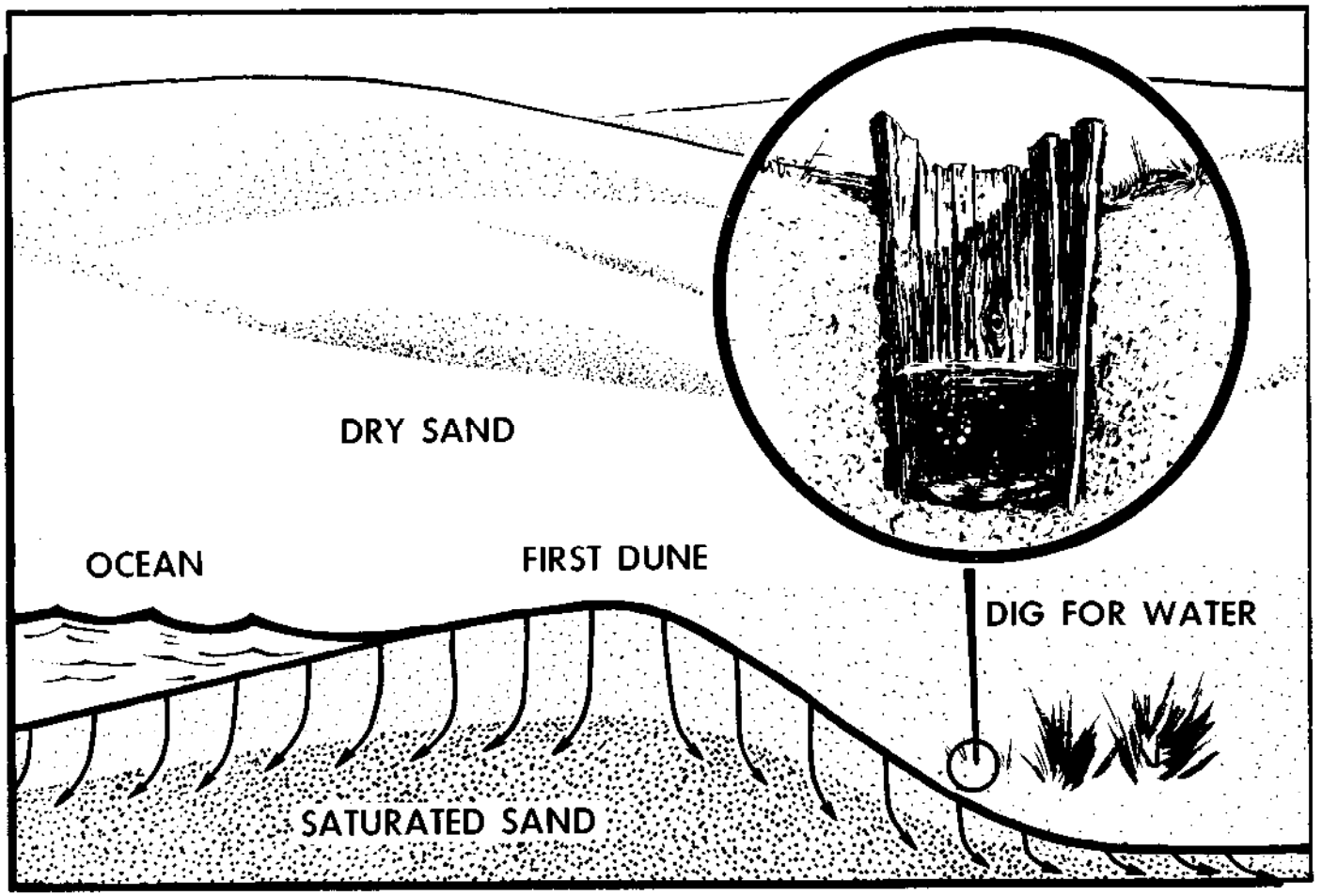

Figure 19-3. Beach Well.

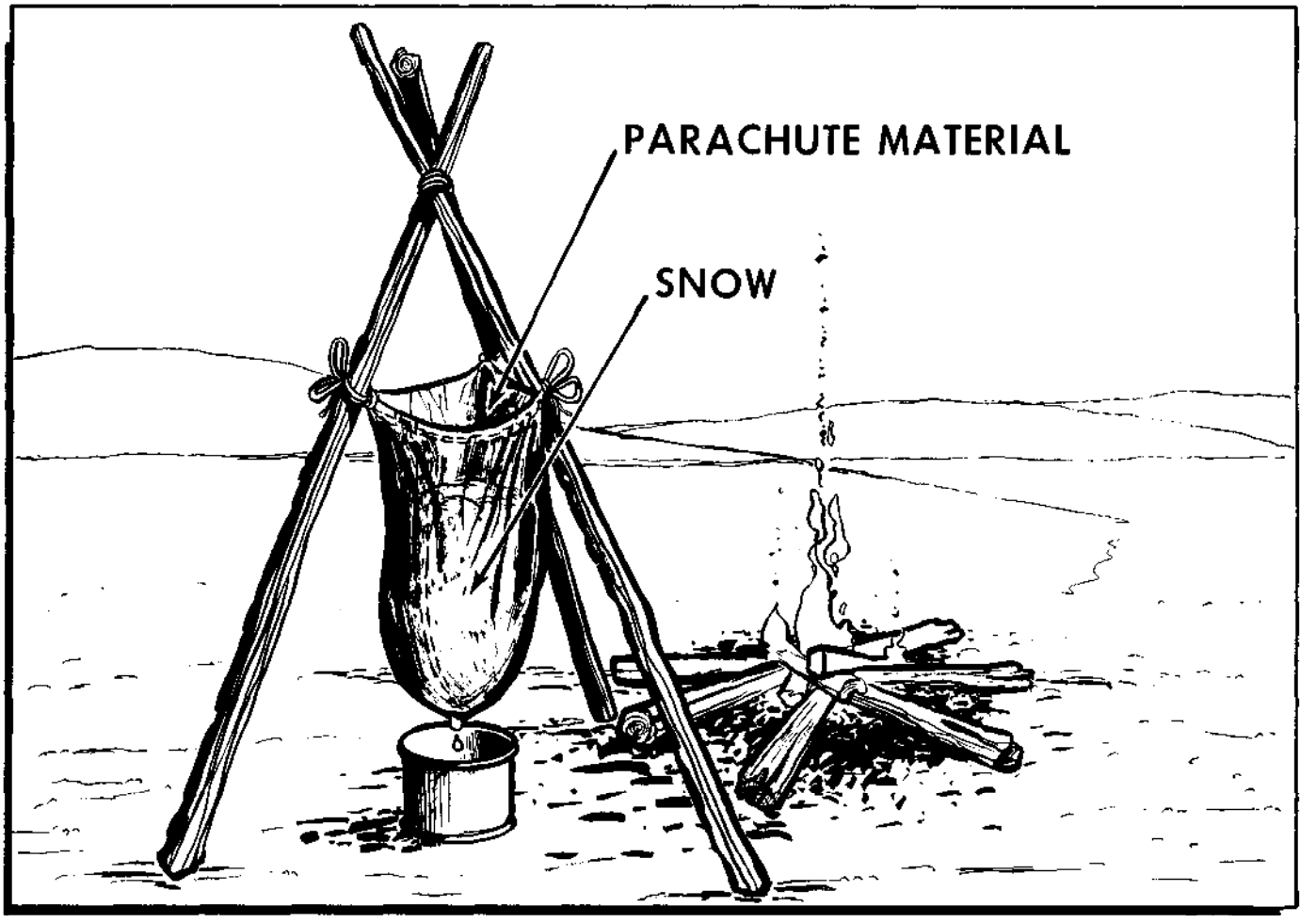

Figure 19-4. Water Machine.

next to the skin; it causes chilling and lowering of the body temperature.

d. Since icebergs are composed of freshwater, they can be a readily available source of drinking water. Survivors should use extreme caution when trying to obtain water from this source. Even large icebergs can suddenly roll over and dump survivors into the frigid sea water. If sea ice is the primary source of water, survivors should recall that like seawater itself, saltwater ice should never be ingested. To obtain water in polar regions or sea ice areas, survivors should select old sea ice, a bluish or blackish ice which shatters easily and generally has rounded corners. This ice will be almost salt-free. New sea ice is milky or gray colored with sharp edges and angles. This type of ice will not shatter or break easily. Snow and ice may be saturated with salt from blowing spray; if it tastes salty, survivors should select different snow or ice sources.

e. The ingesting of unmelted snow or ice is not recommended. Eating snow or ice lowers the body's temperature, induces dehydration, and causes minor cold injury to lips and mouth membranes. Water consumed in cold areas should be in the form of warm or hot fluids. The ingestion of cold fluids or foods increases the body's need for water and requires more body heat to warm the substance.

19-5. Water on the Open Seas. The lack of drinkable water could be a major problem on the open seas. Seawater should never be ingested in its natural state. It will cause an individual to become violently ill in a very short period of time. When water is limited and cannot be replaced by chemical or mechanical means, it must be used efficiently. As in the desert, conserving sweat not water, is the rule. Survivors should keep in the shade as much as possible and dampen clothing with seawater to keep cool. They should not over exert but relax and sleep as much as possible.

a. If it rains, survivors can collect rainwater in available containers and store it for later use. Storage containers could be cans, plastic bags, or the bladder of a life preserver. Drinking as much rainwater as possible while it is raining is advisable. If the freshwater should become contaminated with small amounts of seawater

or salt spray, it will remain safe for drinking (figure 19-5). At night and on foggy days, survivors should try to collect dew for drinking water by using a sponge, chamois, handkerchief, etc.

b. Solar stills will provide a drinkable source of water. Survivors should read the instructions immediately and set them up, using as many stills as available. (Be sure to attach them to the raft.) Desalter kits, if available, should probably be saved for the time when no other means of procuring drinking water is available. Instructions on how to use the desalter kit are on the container.

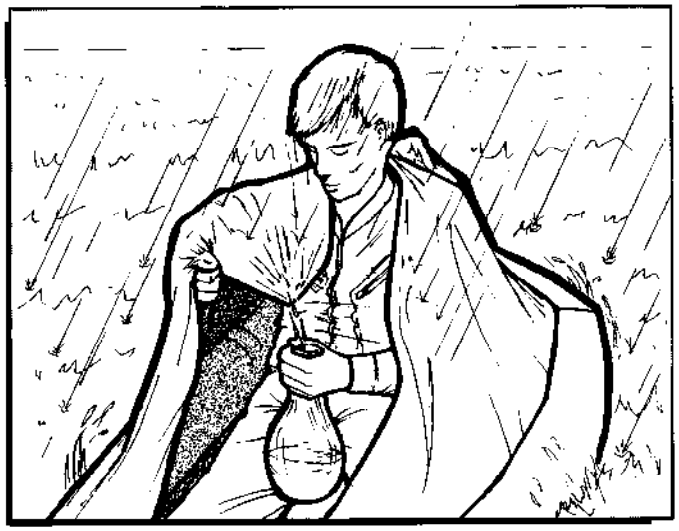

Figure 19-5. Collecting Water from Spray Shield.

c. Only water in its conventional sense should be consumed. The so-called "water substitutes" do little for the survivor, and may do much more harm than not consuming any water at all. There is no substitute for water. Fish juices and other animal fluids are of doubtful value in preventing dehyration. Fish juices contain protein which requires large amounts of water to be digested and the waste products must be excreted in the urine which increases water loss. Survivors should never drink urine—urine is body waste material and only serves to concentrate waste materials in the body and require more water to eliminate the additional waste.

19-6. Water in Tropical Area. Depending on the time of the year and type of jungle, water in the tropical climates can be plentiful; however, it is necessary to know where to look and procure it. Surface water is normally available in the form of streams, ponds, rivers, and swamps. In the savannas during the dry season, it may be necessary for the survivor to resort to digging for water in the places previously mentioned. Water obtained from these sources may need filtration and should be purified. Jungle plants can also provide survivors with water.

a. Many plants have hollow portions which can collect rainfall, dew, etc. (figure 19-6). Since there is no absolute way to tell whether this water is pure, it should be purified. The stems or the leaves of some plants have a hollow section where the stem meets the trunk. Look for water collected here. This includes any Y-shaped plants (palms or air plants). The branches of large trees often support air plants (relatives of the pineapple) whose overlapping, thickly growing leaves may hold a considerable amount of rainwater. Trees may also catch and store rainwater in natural receptacles such as cracks or hollows.

Figure 19-6. Water Collectors.

b. Pure freshwater needing no purification can be obtained from numerous plant sources. There are many varieties of vines which are potential water sources. The vines are from 50 feet to several hundred feet in length and 1 to 6 inches in diameter. They also grow like a hose along the ground and up into the trees. The leaf structure of the vine is generally high in the trees. Water

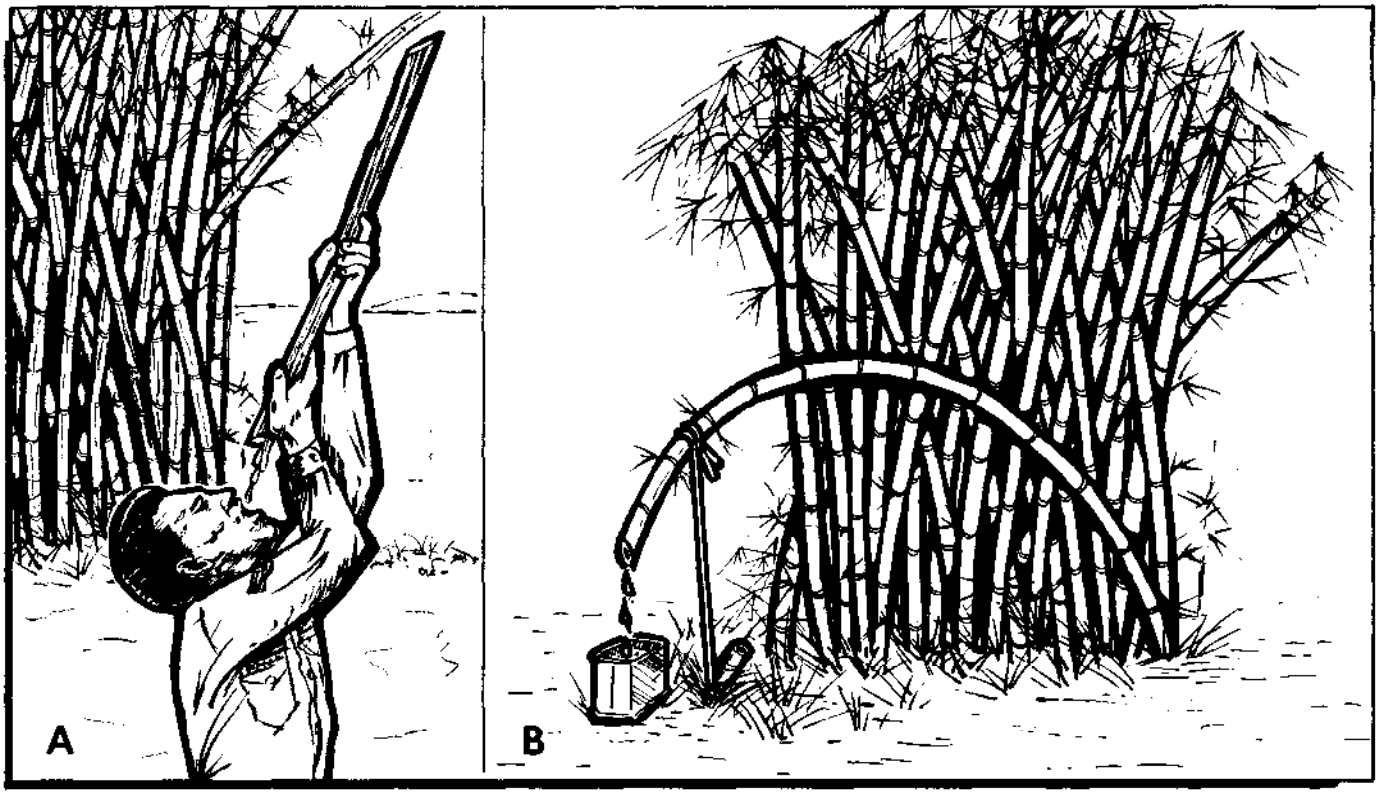

Figure 19-7. Water Vines and Bamboo.

vines are usually soft and easily cut. The smaller species may be twisted or bent easily and are usually heavy because of the water content. The water from these vines should be tested for potability. The first step in testing the water from vines is for survivors to nick the vine and watch for sap running from the cut. If milky sap is seen, the vine should be discarded; if no milky sap is observed, the vine may be a safe water vine. Survivors should cut out a section of the vine, hold that piece vertically, and observe the liquid as it flows out. If it is clear and colorless, it may be a drinkable source. If it is cloudy or milky-colored, they should discard the vine. They should let some of the liquid flow into the palm of the hand and observe it. If the liquid does not change color, they can now taste it. If it tastes like water or has a woody or sweet taste, it should be safe for drinking. Liquid with a sour or bitter taste should be avoided. Water trapped within a vine is easily obtained by cutting out a section of the vine. The vine should first be cut high above the ground and then near the ground. This will provide a long length of vine and, in addition, will tend to hide evidence of the cuts if the survivors are in an evasion situation. When drinking from the vine, it should not touch the mouth as the bark may contain irritants which could affect the lips and mouth (figure 19-7). The pores in the upper end of the section of vine may reclose, stopping the flow of water. If this occurs, survivors should cut off the end of the vine opposite the drinking end. This will reopen the pores allowing the water to flow.

c. Water from the rattan palm and spiny bamboo may be obtained in the same manner as from vines. It is not necessary to test the water if positive identification of the plant can be made. The slender stem (runner) of the rattan palm is an excellent water source. The joints are overlapping in appearance, as if one section is fitted inside the next.

d. Water may be trapped within sections of green bamboo. To determine if water is trapped within a section of bamboo, it should be shaken. If it contains water, a sloshing sound can be heard. An opening may be made in the section by making two 45-degree angle cuts, both on the same side of the section, and prying loose a piece of the section wall. The end of the section may be cut off and the water drunk or poured from the open end. The inside of the bamboo should be examined before consuming the water. If the inside walls are clean and white, the water will be safe to drink. If there are brown or black spots, fungus growth, or any discoloration, the water should be purified before consumption. Sometimes water can also be obtained by cutting the top off certain types of green bamboo, bending it over, and staking it to the ground (figure 19-7). A water container should be placed under it to catch the

dripping water. This method has also proven effective on some vines and the rattan palm.

e. Water can also be obtained from banana plants in a couple of different ways, neither of which is satisfactory in a tactical situation. First, survivors should cut a banana plant down, then a long section should be cut off which can be easily handled. The section is taken apart by slitting from one end to the other and pulling off the layers one at a time. A strip 3 inches wide, the length of the section, and just deep enough to expose the cells should be removed from the convex side. This section is folded toward the convex side to force the water from the cells of the plant. The layer must be squeezed gently to avoid forcing out any tannin into the water. Another technique for obtaining water from the banana plant is by making a "banana-well." This is done by making a bowl out of the plant stump, fairly close to the ground, by cutting out and removing the inner section of the stump (figure 19-8). Water which first enters the bowl may contain a concentration of tannin (an astringent which has the same effect as alum). A leaf from the banana plant or other plant should be placed over the bowl while it is filling to prevent contamination by insects, etc.

f. Water trees can also be a valuable source of water in some jungles. They can be identified by their blotched bark which is fairly thin and smooth. The leaves are large, leathery, fuzzy, and evergreen, and may grow as large as 8 or 9 inches. The trunks may have short outgrowths with fig-like fruit on them or long tendrils with round fruit comprised of corn kernel-shaped nuggets. In a nontactical situation, the tree can be tapped in the same manner as a rubber tree, with either a diagonal slash or a "V." When the bark is cut into, it will exude a white sap which if ingested causes temporary irritation of the urinary tract. This sap dries up quite rapidly and can easily be removed. The cut should be continued into the tree with a spigot (bamboo, knife, etc.) at the bottom of the tap to direct the water into a container. The water flows from the leaves back into the roots after sundown, so water can be procured from this source only after sundown or on overcast (cloudy) days. If survivors are in a tactical situation, they can obtain water from the tree and still conceal the procurement location. If the long tendrils are growing thickly, they can be separated and a hole bored into the tree. The white sap should be scraped off and a spigot placed below the tap with a water container to catch the water. Moving the tendrils back into place will conceal the container. Instead of boring into the tree, a couple of tendrils can be cut off or snapped off if no knife is available. The white sap should be allowed to dry and then be removed. The ends of the tendrils should be placed in a water container and the container concealed.

g. Coconuts contain a refreshing fluid. Where coconuts are available, they may be used as a water source. The fluid from a mature coconut contains oil, which when consumed in excess can cause diarrhea. There is little problem if used in moderation or with a meal and not on an empty stomach. Green unripe coconuts about the size of a grapefruit are the best for use because the fluid can be taken in large quantities without harmful effects. There is more fluid and less oils so there is less possibility of diarrhea.

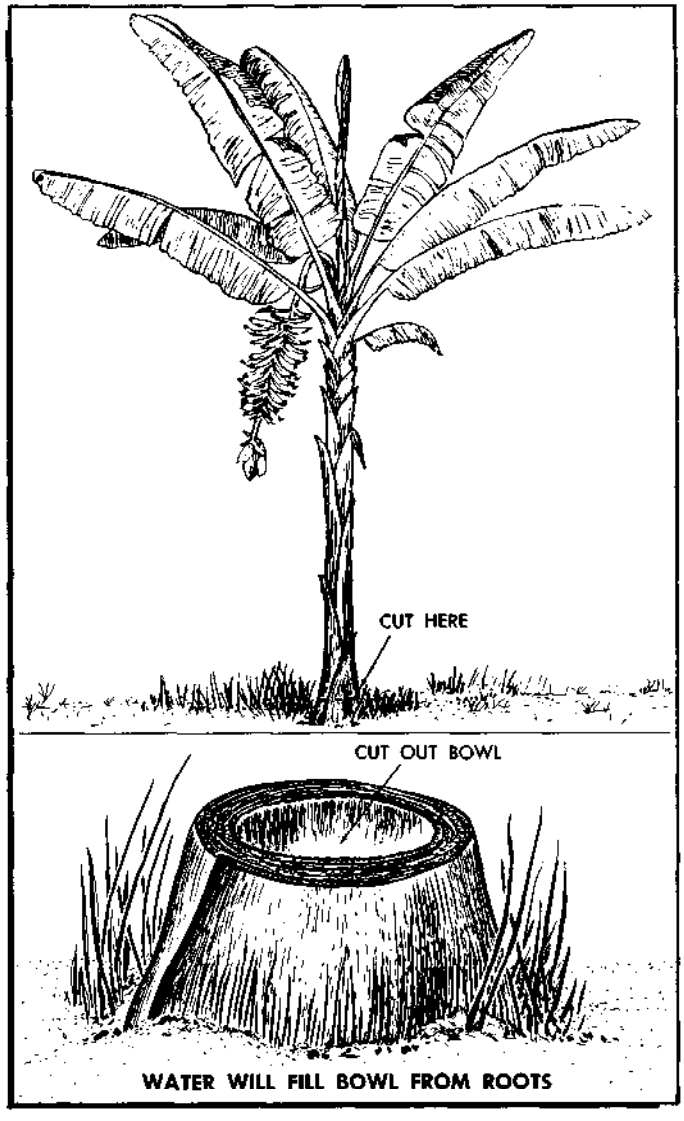

Figure 19-8. Water from Banana Plant.

h. Water can also be obtained from liquid mud. Mud can be filtered through a piece of cloth. Water taken by this method must be purified. Rainwater can be collected from a tree by wrapping a cloth around a slanted tree and arranging the bottom end of the cloth to drip into a container (figure 19-9).

19-7. Water in Dry Areas. Locating and procuring water in a dry environment can be a formidable task. Some of the ways to find water in this environment have been

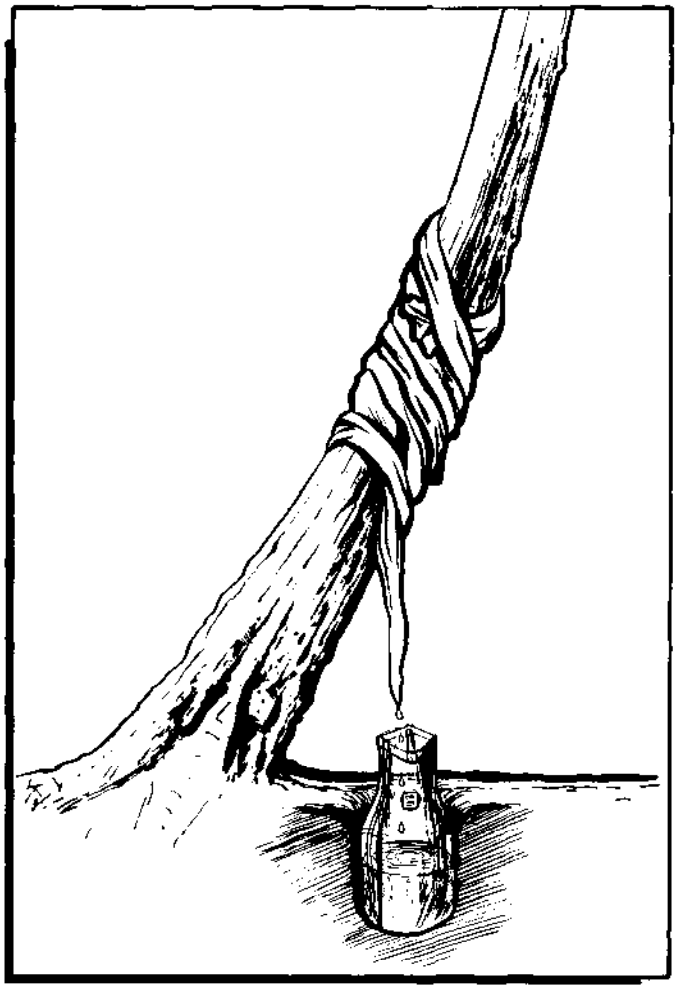

Figure 19-9. Collecting Water from Slanted Tree.

explored, such as locating a concave bend in a dry riverbed and digging for water (figure 19-10). If there is any water within a few feet of the surface, the sand will become slightly damp. Dig until water is obtained.

a. Some deserts become humid at night. The humidity may be collected in the form of dew. This dew can be collected by digging a shallow basin in the ground about 3 feet in diameter and lining it with a piece of canvas, plastic, or other suitable material. A pyramid of stones taken from a minimum of 1 foot below the surface should then be built in this basin. Dew will collect on and between the stones and trickle down onto the lining material where it can be collected and placed in a container.

b. Plants and trees having roots near the surface may be a source of water in dry areas. Water trees of dry Australia are a source of water, as their roots run out 40 to 80 feet at a depth of 2 to 9 inches under the surface. Survivors may obtain water from these roots by locating a root 4 to 5 feet from the trunk and cutting the root into 2- or 3-foot lengths. The bark can then be peeled off and the liquid from each section of root drained into a container. The liquid can also be sucked out. The trees growing in hollows or depressions will have the most water in their roots. Roots that are 1 to 2 inches thick are an ideal size. Water can be carried in these roots by plugging one end with clay.

c. Cactus-like or succulent plants may be sources of water for survivors, but they should recall that no plants should be used for water procurement which have a milky sap. The barrel cactus of the United States provides a water source. To obtain it, survivors should first cut off the top of the plant. The pulpy inside portions of the plant should then be mashed to form a watery pulp. Water may ooze out and collect in the bowl; if not, the pulp may be squeezed through a cloth directly into the mouth.

d. The solar still is a method of obtaining water that uses both vegetation and ground moisture to produce water (figure 19-11). A solar still can be made from a sheet of clear plastic stretched over a hole in the ground. The moisture in the soil and from plant parts (fleshy stems and leaves) will be extracted and collected by this emergency device. Obviously, where the soil is extremely dry and no fleshy plants are available, little, if any, water can be obtained from the still. The still may also be used to purify polluted water.

(1) The parts for the still are a piece of plastic about 6 feet square, a water collector-container or any waterproof material from which a collector-container can be fashioned, and a piece of plastic tubing about one-fourth inch in diameter and 4 to 6 feet long. The tubing is not absolutely essential but makes the still easier to use. A container can be made from such materials as plastic, aluminum foil, poncho, emergency ration tins, or a flight helmet. The tubing, when available, is fastened to the bottom of the inside of the container and used to remove drinking water from the container without disturbing the plastic. Some plastics work better than others, although any clear plastic should work if it is strong.

(2) If plants are available or if polluted water is to be purified, the still can be constructed in any convenient spot where it will receive direct sunlight throughout the day. Ease of digging will be the main consideration. If soil moisture is to be the only source of water, some sites will be better than others. Although sand generally does not retain as much moisture as clay, a wet sand will work very well. Along the seacoast or in any inland areas where brackish or polluted water is available, any wet soil, even sand, produces usable amounts of water. On cloudy days, the yield will be reduced because direct sunlight is necessary if the still is to operate at full efficiency.

(3) Certain precautions must be kept in mind. If polluted water is used, survivors should make sure that none is spilled near the rim of the hole where the plastic touches the soil and that none comes in contact with the

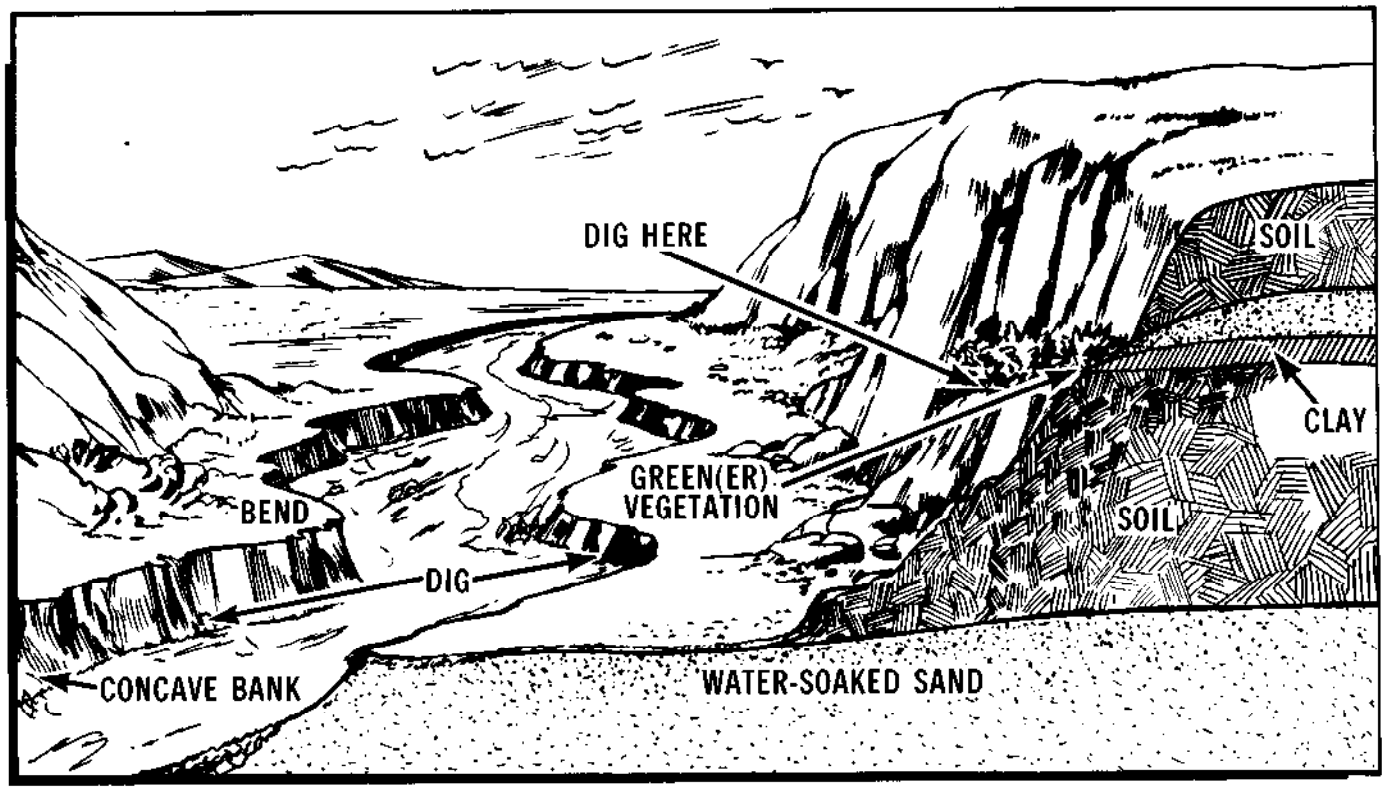

Figure 19-10. Dry Stream Bed.

container to prevent the freshly distilled water becoming contaminated. Survivors should not disturb the plastic sheet during daylight "working hours" unless it is absolutely necessary. If a plastic drinking tube is not available, raise the plastic sheet and remove the container as few times as possible during daylight hours. It takes one-half hour for the air in the still to become resaturated and the collection of water to begin after the plastic has been disturbed. Even when placed on fairly damp soil and in an area where 8 hours of light per day is directed on the solar still, the average yield is only about 1 cup per day per still. Due to the low yields obtained from this device, survivors must give consideration to the possible danger of excessive dehydration brought about by constructing the solar still. In certain circumstances, solar still returns, even over 2- or 3-day periods, will not equal the amount of body fluid lost in construction and will actually hasten dehydration.

(4) Steps survivors should follow when constructing a solar still are: Dig a bowl-shaped hole in the soil about 40 inches in diameter and 20 inches deep. Add a smaller, deeper sump in the center bottom of the hole to accommodate the container. If polluted waters are to be purified, a small trough can be dug around the side of the hole about halfway down from the top. The trough ensures that the soil wetted by the polluted water will be exposed to the sunlight and at the same time that the polluted water is prevented from running into the container. If plant material is used, line the sides of the hole with pieces of plant or its fleshy stems and leaves. Place the plastic over the hole and put soil on the edges to hold it in place. Place a rock no larger than a plum in the center of the plastic until it is about 15 inches below ground level. The plastic will now have the shape of a cone. Put more soil on the plastic around the rim of the hole to hold the cone securely in place and to prevent water-vapor loss. Straighten the plastic to form a neat cone with an angle of about 30 degrees so the water drops will run down and fall into the container. It takes about 1 hour for the air to become saturated and start condensing on the underside of the plastic cone.

e. The vegetation bag is a simpler method of water procurement. This method involves cutting foliage from trees or herbaceous plants, sealing it in a large clear plastic bag, and allowing the heat of the Sun to extract the fluids contained within. A large, heavy-duty clear plastic bag should be used. The bag should be filled with about 1 cubic yard of foliage, sealed, and exposed to the Sun. The average yield for one bag tested was 320 ml/bag 5-hour day. This method is simple to set up. The vegetation bag method of water procurement does have one primary drawback. The water produced is normally bitter to taste, caused by biological breakdown of the leaves as they lay in the water produced and super-

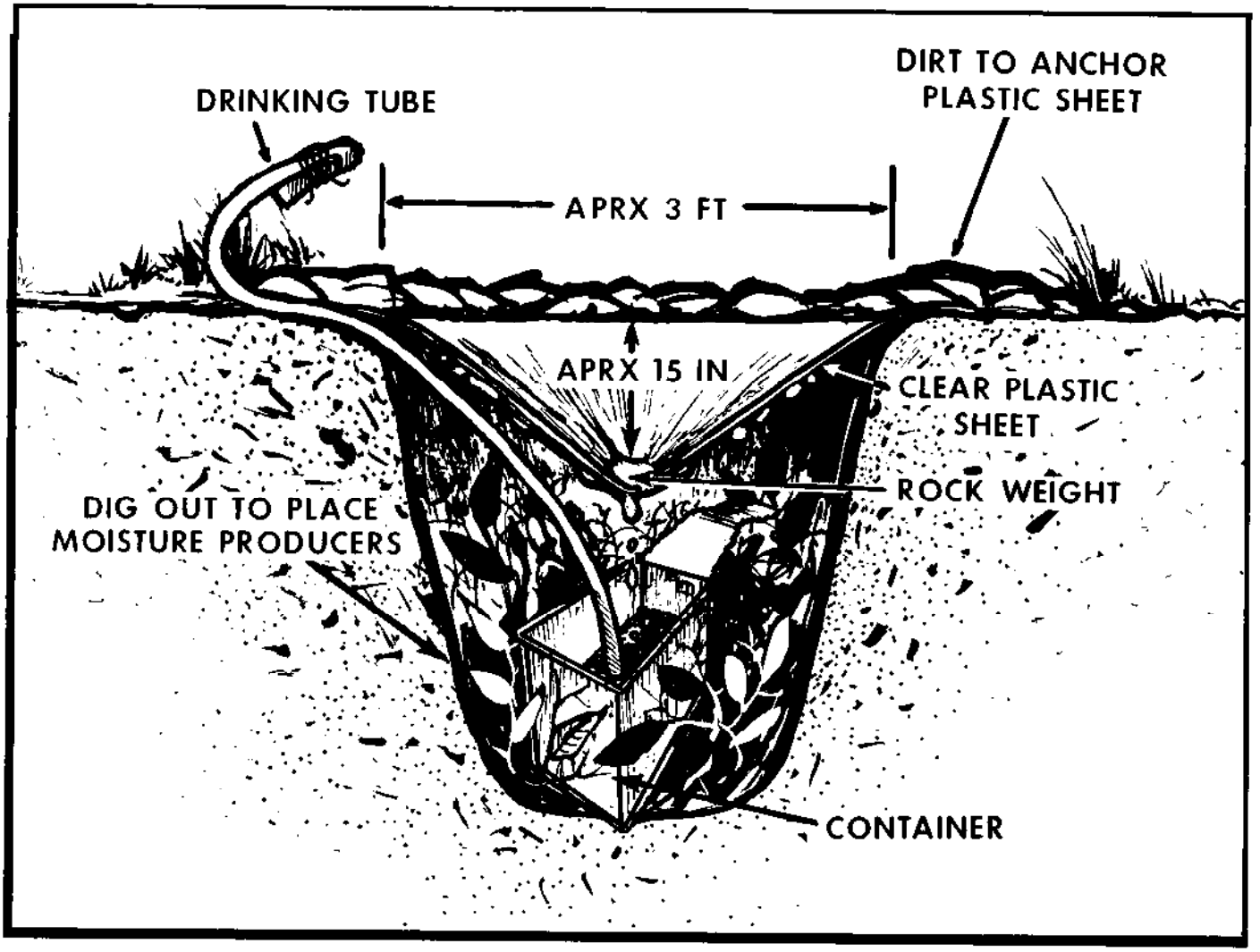

Figure 19-11. Solar Still.

heated in the moist "hothouse" environment. This method can be readily used in a survival situation, but before the water produced by certain vegetation is consumed, it should undergo the taste test. This is to guard against ingestion of cyanide-producing substances and other harmful toxins, such as plant alkaloids. (See figure 19-12.)

f. One more method of water procurement is the water transpiration bag, a method that is simple to use and has great potential for enhancing survival. This method is the vegetation bag process taken one step further. A large plastic bag is placed over a living limb of a medium-size tree or large shrub. The bag opening is sealed at the branch, and the limb is then tied down to allow collected water to flow to the corner of the bag. For a diagram of the water transpiration method, see figure 19-13.

(1) The amount of water yielded by this method will depend on the species of trees and shrubs available. During one test of this method, a transpiration bag produced approximately a gallon per day for 3 days with a plastic bag on the same limb, and with no major deterioration of the branch. Other branches yielded the same amount. Transpired water has a variety of tastes depending on whether or not the vegetation species is allowed to contact the water.

(2) The effort expended in setting up water transpiration collectors is minimal. It takes about 5 minutes' work and requires no special skills once the method has been described or demonstrated. Collecting the water in a survival situation would necessitate survivors dismantling the plastic bag at the end of the day, draining the contents and setting it up again the following day. The same branch may be reused (in some cases with almost similar yields); however, as a general rule, when vegetation abounds, a new branch should be used each day.

(3) Without a doubt, the water transpiration bag method surpasses other methods (solar stills, vegetation bag, cutting roots, barrel cactus) in yield, ease of assembly, and in most cases, taste. The benefits of having a simple plastic bag can't be over-emphasized. As a water procurer, in dry, semi-dry, or desert environments

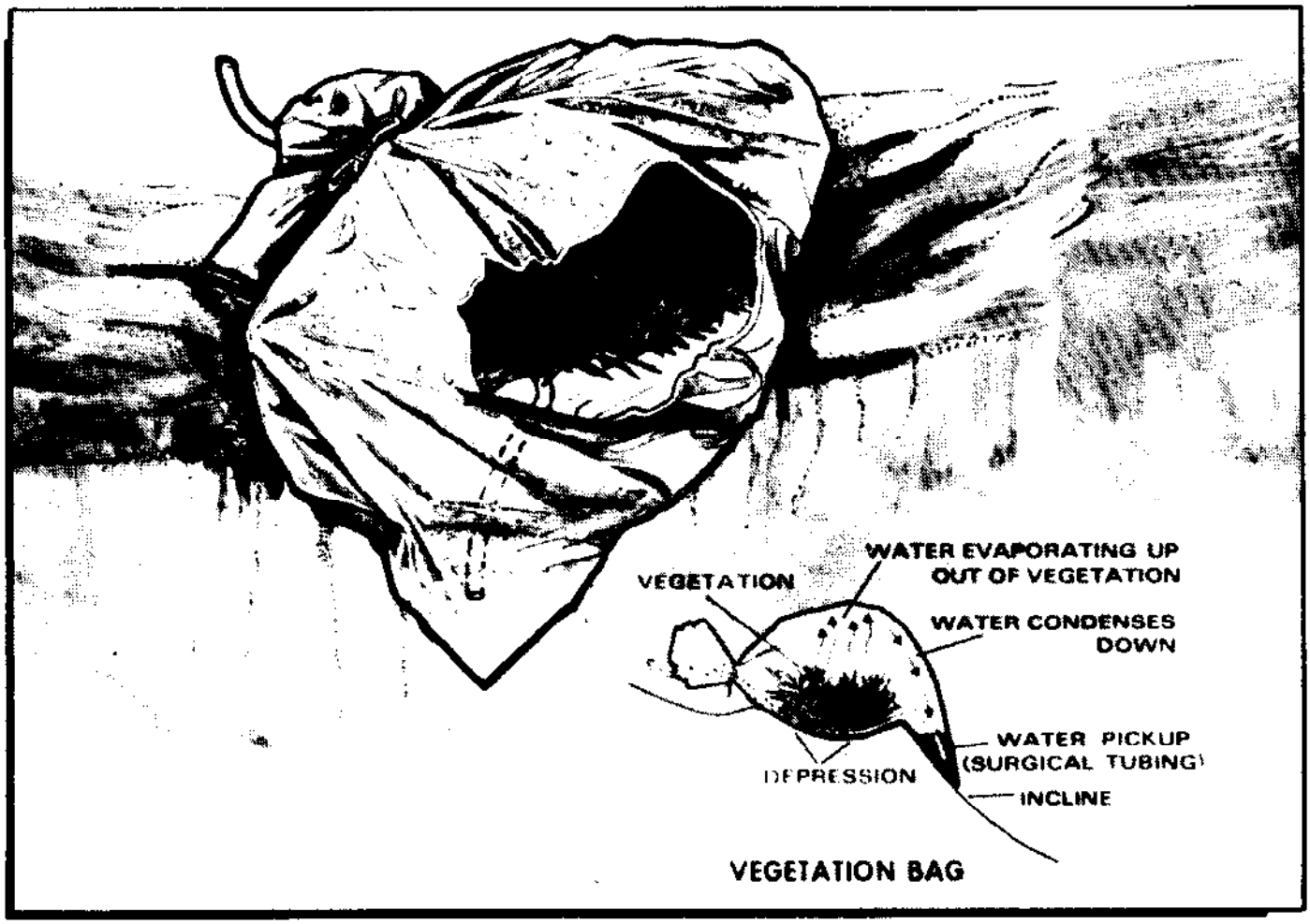

Figure 19-12. Vegetation Bag.

where low woodlands predominate, it can be used as a water transpirator; in scrubland, steppes, or treeless plains, as a vegetation bag; in sandy areas without vegetation, it can be cut up and improvised into solar stills. Up to three large, heavy-duty bags may be needed to sustain one survivor in certain situations.

19-8. Preparation of Water for Consumption:

a. The following are ways survivors can possibly determine the presence of harmful agents in the water:

(1) Strong odors, foam, or bubbles in the water.

(2) Discoloration or turbid (muddy with sediment).

(3) Water from lakes found in desert areas are sometimes salty because they have been without an outlet for extended periods of time. Magnesium or alkali salts may produce a laxative effect; if not too strong, it is drinkable.

(4) If the water gags survivors or causes gastric disturbances, drinking should be discontinued.

(5) The lack of healthy green plants growing around any water source.

b. Because of survivors' potential aversion to water from natural sources, it should be rendered as potable as possible through filtration. Filtration only removes the solid particles from water—it does not purify it. One simple and quick way of filtering is to dig a sediment hole or seepage basin along a water source and allow the soil to filter the water (figure 19-14). The seepage hole should be covered while not in use. Another way is to construct a filter—layers of parachute material stretched across a tripod (figure 19-15). Charcoal is used to eliminate bad odors and foreign materials from the water. Activated charcoal (obtained from freshly burned wood is used to filter the water). If a solid container is available for making a filter, use layers of fine-to-coarse sand and gravel along with charcoal and grass.

c. Purification of water may be done a variety of ways. The method used will be dictated by the situation (such as tactical or nontactical).

(1) Boil the water for at least 10 minutes.

(2) To use purification tablets survivors should follow instructions on the bottle. One tablet per quart of clear water; two tablets if water is cloudy. Let water

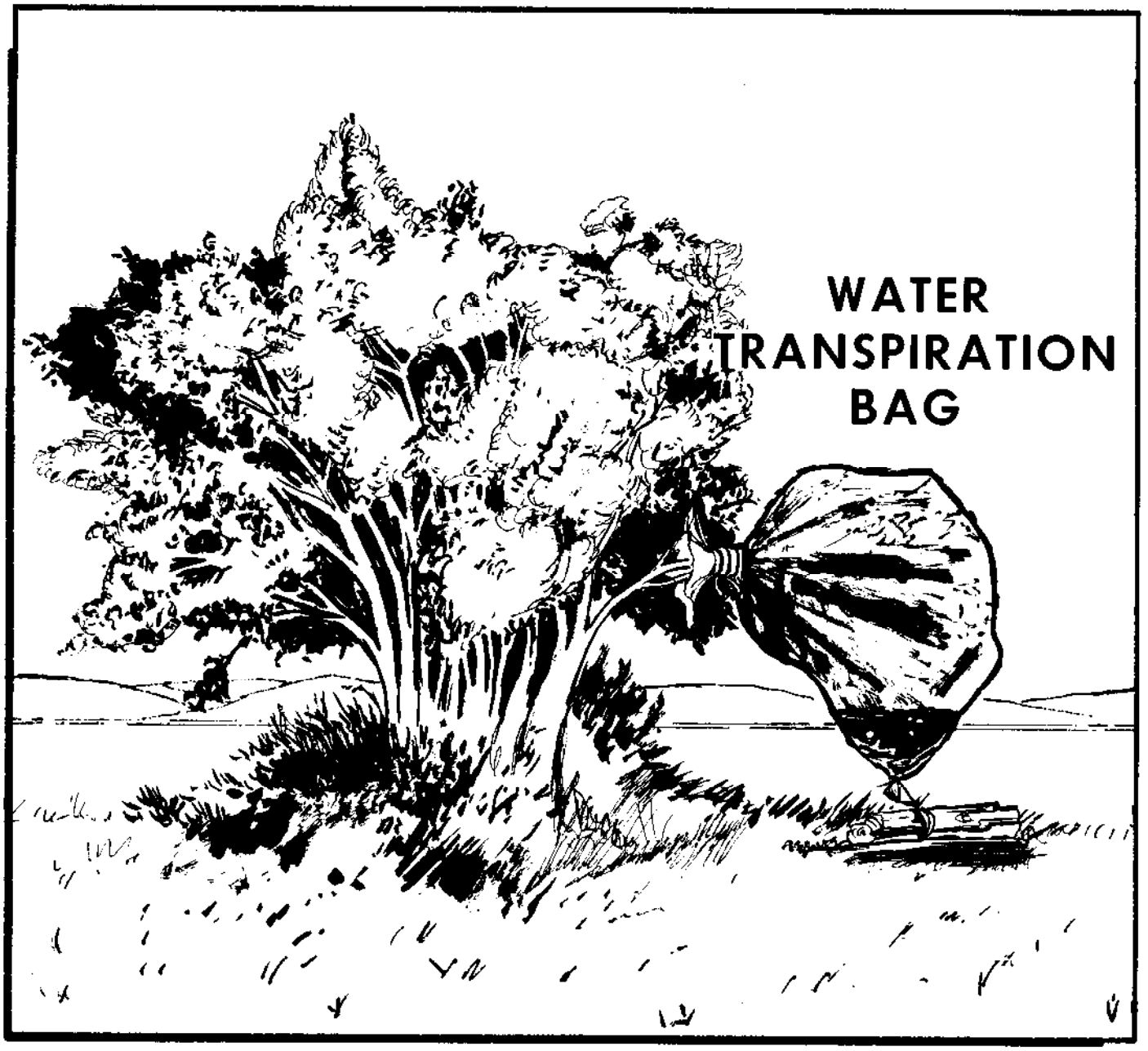

Figure 19-13. Transpiration Bag.

stand for 5 minutes (allowing the tablet time to dissolve), then shake and allow to stand for 15 minutes. Survivors should remember to turn the canteen over and allow a small amount of water to seep out and cover the neck part of the canteen. In an evasion situation, water purification tablets should be used for purifying water. If these are not available, plant sources or nonstagnant, running water obtained from a location upstream from habitatio should be consumed.

(3) Eight drops of 2½-percent iodine per quart—stir or shake and let stand for at least 10 minutes.

d. After water is found and purified, survivors may wish to store it for later consumption. The following make good containers:

(1) Waterbag.

(2) Canteen.

(3) Prophylactic inside a sock for protection of bladder.

(4) Segment of bamboo.

(5) Birch bark and pitch canteen.

(6) LPU bladder.

(7) Hood from antiexposure suit.

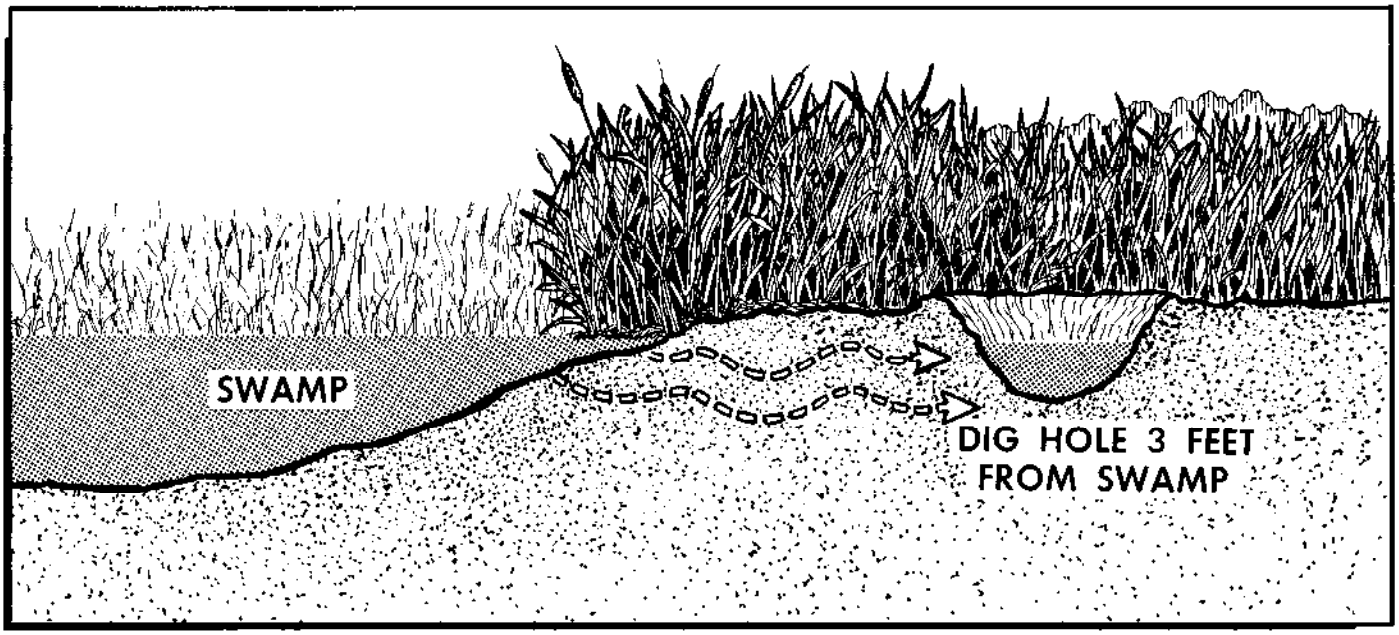

Figure 19-14. Sediment Hole.

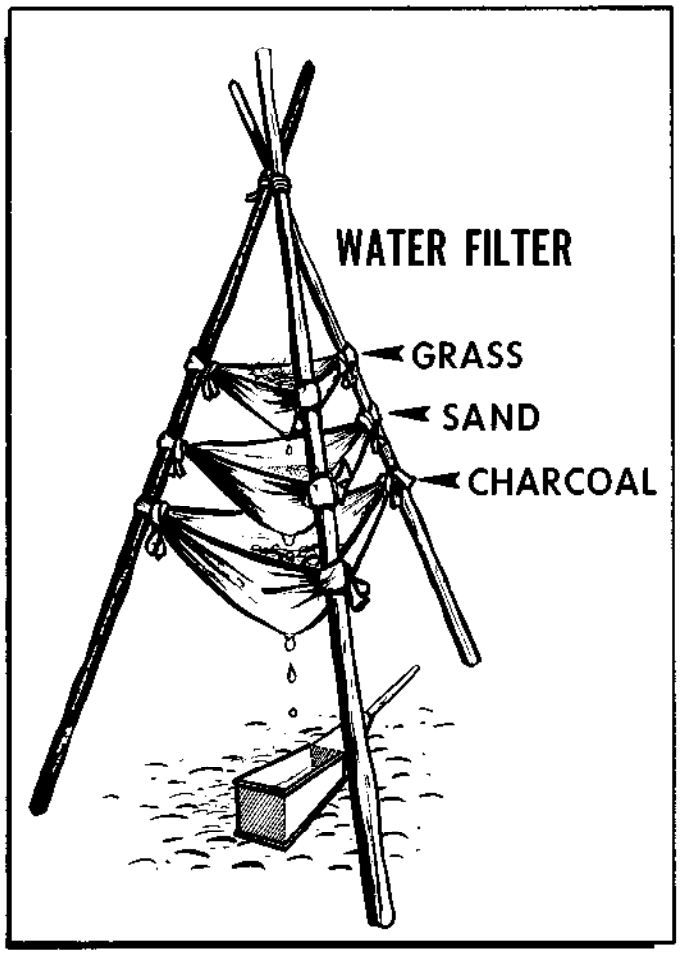

Figure 19-15. Water Filter.

Part Seven

TRAVEL

Chapter 20

LAND NAVIGATION

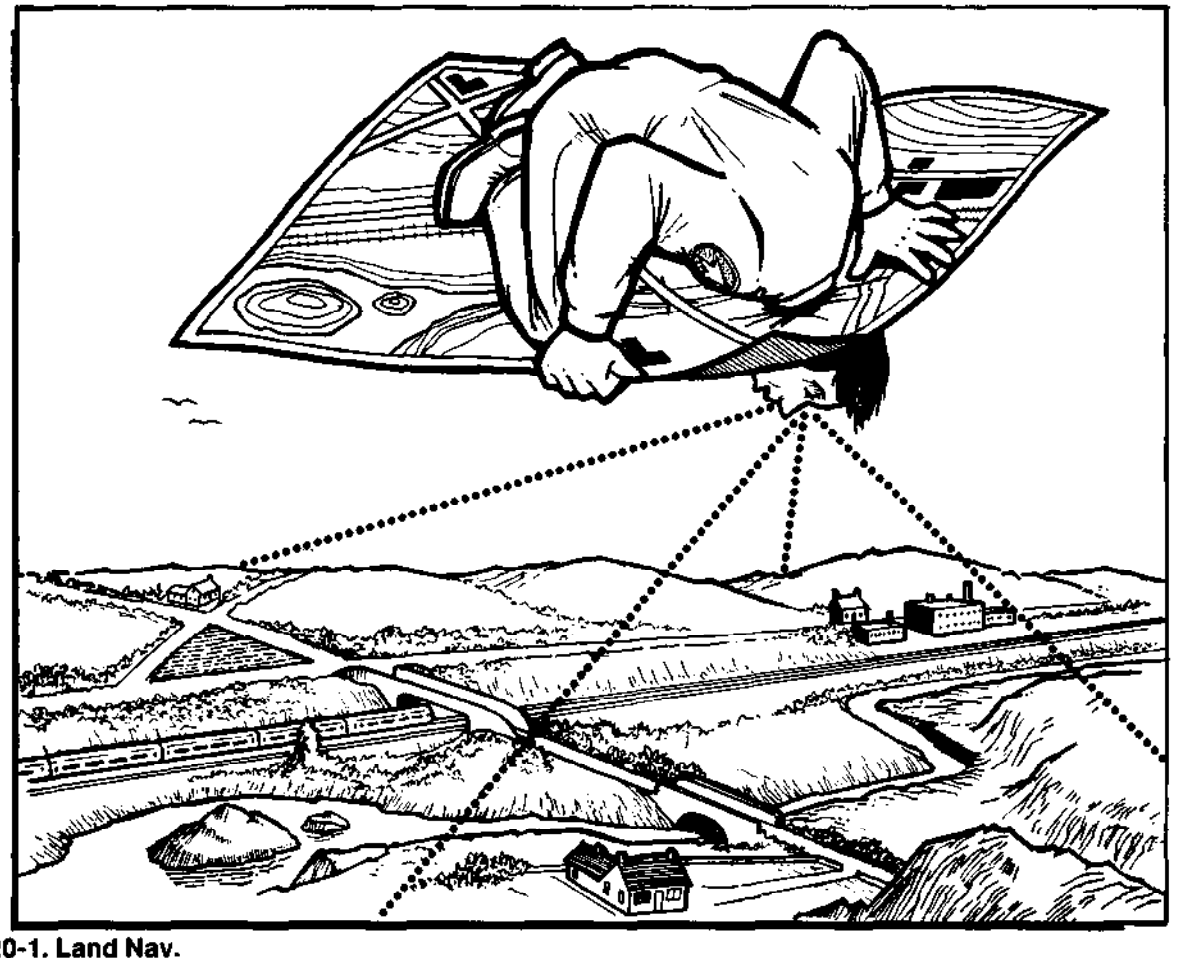

Figure 20-1. Land Nav.

20-1. Introduction:

a. Survivors must know their location in order to intelligently decide if they should wait for rescue or if they should determine a destination and (or) route to travel. If the decision is to stay, the survivors need to know their location in order to radio the information to rescue personnel. If the decision is to travel, survivors must be able to use a map to determine the best routes of travel, location of possible food and water, and hazardous areas which they should avoid.

b. This chapter provides background information in the use of the map and compass (figure 20-1).

20-2. Maps:

a. A map is a pictorial representation of the Earth's surface drawn to scale and reproduced in two dimensions. Every map should have a title, legend, scale, north arrow, grid system, and contour lines. With these components, survivors can determine the portion of the Earth's surface the map covers. Survivors should be able to understand all of the markings on the map and use them to advantage. They should also be able to determine the distance between any two points on the map and be able to align the map with true north so it conforms to the actual features on the ground.

b. A map is a conceptionalized picture of the Earth's surface as seen from above, simplified to bring out important details and lettered for added identification. A map represents what is known about the Earth rather than what can be seen by an observer. However, a map is selective in that only the information which is necessary for its intended use is included on any one map. Maps also include features which are not visible on Earth, such as parallels, meridians, and political boundaries.

c. Since it is impossible to accurately portray a round object, such as the Earth, on a flat surface, all maps have some elements of distortion. Depending on the intended use, some maps sacrifice constant scale for accuracy in measurement of angles, while others sacrifice accurate measurement of angles for a constant scale. However, most maps used for ground navigation use a compromise projection in which a slight amount of distortion is introduced into the elements which a map portrays, but in which a fairly true picture is given.

d. A planimetric map presents only the horizontal positions for the features represented. It is distinguished from a topographic map by the omission of relief in a measurable form.

e. A topographic map (figure 20-2) portrays terrain and landforms in a measurable form and the horizontal positions of the features represented. The vertical positions, or relief, are normally represented by contours. On maps showing relief, the elevations and contours are measured from a specified vertical datum plane and usually mean sea level.

f. A plastic relief map is a reproduction of an aerial photograph or a photomosaic made from a series of aerial photographs upon which grid lines, marginal data, place names, route numbers, important elevations, boundaries, approximate scale, and approximate direction have been added.

g. A PICTOMAP (figure 20-3) is the acronym for photographic image conversion by tonal masking procedures. It is a map on which the photographic imagery of a standard photomap has been converted into interpretable colors and symbols.

h. A photomosaic is an assembly of aerial photographs and is commonly called a mosiac in topographic usage. Mosiacs are useful when time does not permit the compilation of a more accurate map. The accuracy of a mosiac depends on the method used in its preparation and may vary from simply a good pictorial effect of the ground to that of a planimetric map.

i. Military city map is a topographic map, usually 1:12,500 scale, of a city, outlining streets and showing street names, important buildings, and other urban elements of military importance which are compatible with the scale of the map. The scales of military city maps can vary from 1:25,000 to 1:5,000, depending on the importance and size of the city, density of detail, and available intelligence information.

j. Special maps are for special purposes such as trafficability, communications, and assault. These are usually overprinted maps of scales smaller than 1:100,000 but larger than 1:1,000,000. Other types of special maps are those made from organosol or materials other than paper to meet the requirements of special climatic conditions.

k. A terrain model is a scale model of the terrain showing landforms, and in large scale models, industrial and cultural shapes. It is designed to provide a means for visualizing the terrain for planning or indoctrination purposes and for briefing on assault landings.

l. A special purpose map is one that has been designed or modified to give information not covered on a standard map or to elaborate on standard map data. Special purpose maps are usually in the form of an overprint. Overprints may be in the form of individual sheets or combined and bound into a study of an area. A few of the subjects covered are:

(1) Landform.
(2) Drainage characteristics.
(3) Vegetation.
(4) Climate.
(5) Coast and landing beaches.
(6) Railroads.
(7) Airfields.
(8) Urban areas.
(9) Electric power.
(10) Fuels.
(11) Surface water resources.
(12) Ground water resources.
(13) Natural construction materials.
(14) Cross-country movement.
(15) Suitability for airfield construction.
(16) Airborne operations.

20-3. Aeronautical Charts. Air navigation and planning charts are used for flight planning. Each different series of charts is constructed at a different scale and format to meet the needs of a particular type of air navigation. The air navigation and planning charts are smaller in scale and less detailed than Army maps or air target materials. The control of positional error is less critical. The following list includes the charts most commonly used in intelligence operations. They are available through the Defense Mapping Agency (DMA) Officer of Distribution Services, Washington DC. A description of each chart follows the listing:

CHART	SCALE	CODE
USAF Global Navigation and Planning Chart	1:5,000,000	GNC
USAF Jet Navigation Chart	1:3,000,000	JNC-A
USAF Operational Navigation Chart	1:1,000,000	ONC
USAF Tactical Pilotage Chart	1:500,000	TPC
USAF Jet Navigation Chart	1:2,000,000	JN
Joint Operations Graphic	1:250,000	JOG

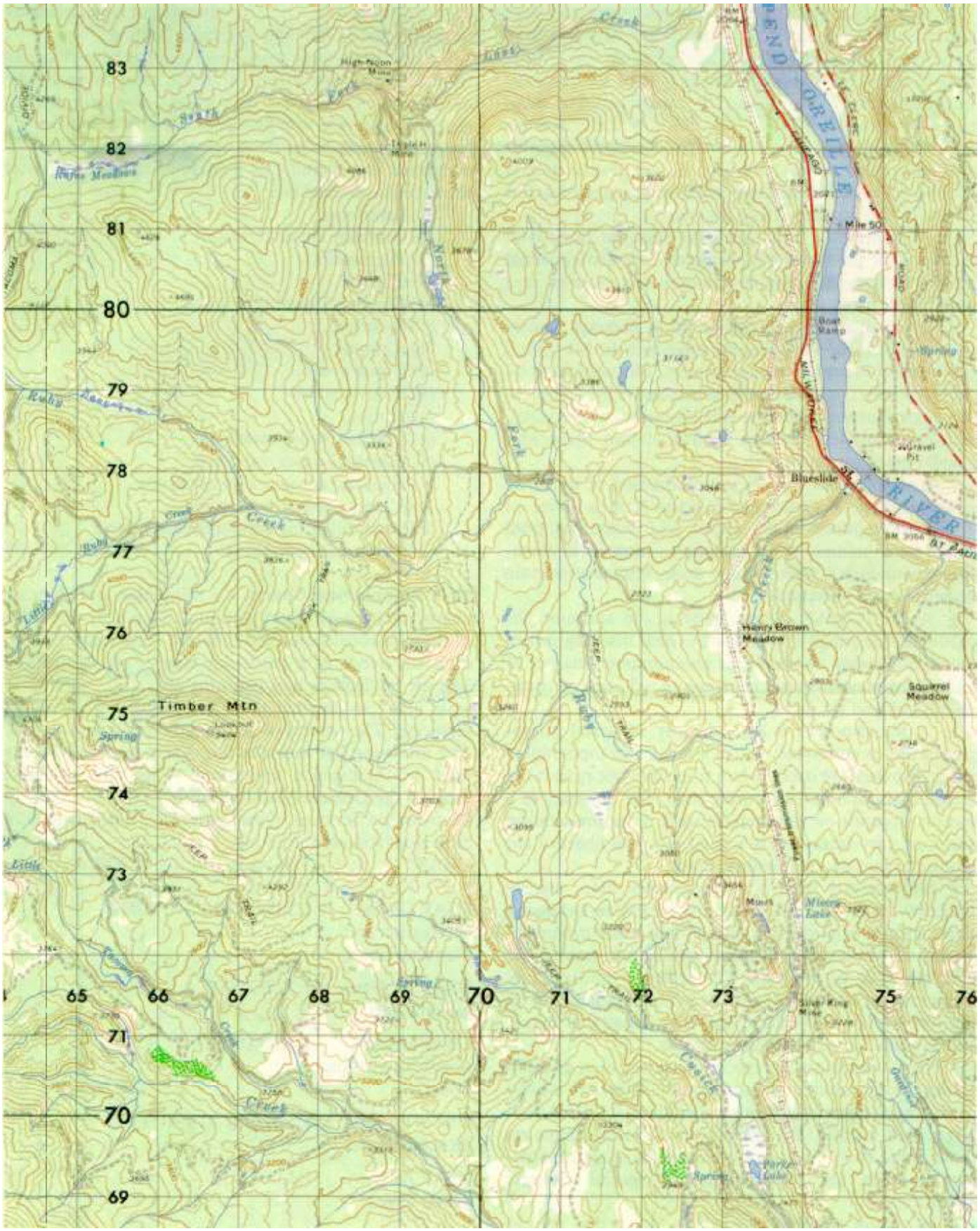

Figure 20-2. Topographic Map.

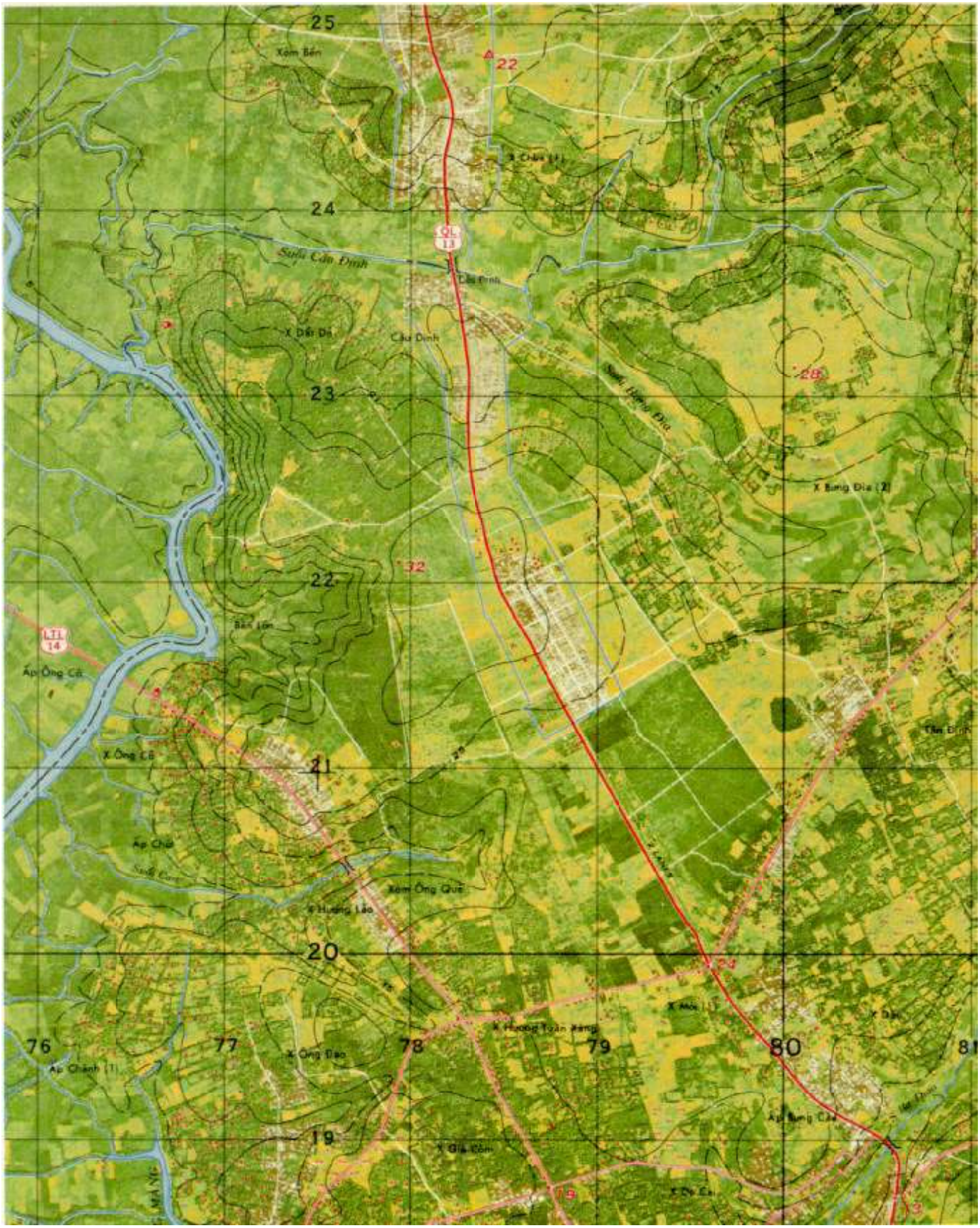

Figure 20-3. Pictomap.

a. Global Navigation and Planning Chart (GNC) (figure 20-4). This chart is designed for general planning purposes where large areas of interest and long-distance operations are involved. It serves as a navigation chart for long-range, high-altitude, and high-speed aircraft since sheet lines have been selected on the basis of primary areas of strategic interest. Several other general planning charts are available through the DMA. Some of these charts are produced on selected areas of strategic interest; others provide wide coverage. All general planning charts are produced at a small or very small scale which provides extensive area coverage on a single sheet.

b. USAF Jet Navigation Chart (JN/JNC-A) (figure 20-5). The basic JNC is produced at a scale of 1:2,000,000. The JNC-A is produced on the north polar area and in the United States at a scale of 1:3,000,000. Both jet navigation charts are printed on 41½- by 57⅜-inch sheets.

(1) The JN chart is used for preflight planning and en route navigation by long-range jet aircraft with dead reckoning, radar, celestial, and grid navigational capabilities. The charts are designed so they can be joined to produce a strip chart which provides the necessary navigational information for any intended course. Relief is indicated through the use of contours, spot elevations, and gradient tints. Large, level terrain areas are indicated by a symbol that consists of narrow, parallel lines with the elevation annotated within the symbol.

(2) Principal cities and towns and principal roads and rail networks are shown on the JN chart. The transportation network is shown in the immediate area of populated places. Lakes and principal drainage patterns are also pictured. The elevations of major lakes are indicated so that the altitude may be determined by using the aircraft radar altimeter.

c. USAF Operational Navigation Chart (ONC) (figure 20-6):

(1) The ONC was developed to meet military requirements for a chart adaptable to low-altitude navigation. The ONC is used for preflight planning and en route navigation. It is also used for operational planning, intelligence briefing and plotting, and flight planning displays.

(2) This chart covers an area of 8° of latitude and 12° of longitude. ONC sheets are identified by combining a letter and a number (figure 20-7). Letters identify 8° bands of latitude, starting at the North Pole and progressing southward. Numbers identify 12° sections of longitude from the prime meridian eastward. The successful execution of low-altitude mission depends upon visual and radar identification of ground features used as checkpoints and a rapid visual association of these features with their chart counterparts. The ONC portrays, by conventional signs and symbols, cultural features which have low-altitude checkpoint significance. Powerlines are shown (except on cities) and are indicated by the usual line and pole symbol.

(3) For certain circumstances, operational requirements may be more effectively satisfied by pictorial illustrations than by the conventional symbolization of such structures as prominent buildings, bridges, dams, towers, holding or storage tanks, stadiums, and related features. For these reasons, significant landmarks are depicted on ONCs by pictorial symbols.

(4) The ONC portrays relief in perspective so that the user gets instantaneous appreciation of relative heights, slope gradients, and the forms of ground patterns. Topographic expression, illustrated basically with contours and spot elevations, is emphasized by the use of shaded relief and terrain characteristic tints defining the overall elevation levels. ONC contour intervals and terrain characteristic tints are selected regionally. This captures the relative significance of ground forms as a complete picture, and this feature aids preflight planning and in-flight identification.

d. USAF Tactical Pilotage Chart (TPC) (figure 20-8):

(1) The TPC is produced in a coordinated series at a scale of 1:500,000. Sheet sizes are the same dimensions as the ONC sheets; however, a TPC covers only one-fourth as much area as an ONC sheet. The TPC breakdown on the ONC is illustrated in figure 20-9. A TPC is identified by the ONC identification and the letter "A," "B," "C," or "D."

(2) The TPC is used for detailed preflight planning and mission analysis. In designing the TPC, emphasis was placed on ground features which are significant for low-level, high-speed navigation, using visual and radar means. The selected ground features also permit immediate ground-chart orientation at predetermined checkpoints.

(3) Relief on the TPC is displayed by contours (intervals may vary between 100 feet and 1,000 feet), spot elevations, relief shading, and terrain characteristic tints. Cultural features such as towns and cities, principal roads, railroads, power transmission lines, boundaries, and other features of value for low-altitude visual missions are included on the TPC. Pictorial symbols are used for features which provide the best checkpoints. Other features of the TPC which enhance its tactical air navigation qualities are as follows:

(a) UTM grid overprint.

(b) Vegetation color and symbol code.

(c) Enlarged vertical obstruction symbols.

(d) Enlarged road and railroad symbols.

(e) Emphasized radio aid to navigation symbols.

(f) Foreign place name glossary.

(g) Airdrome runway patterns to scale when information is available.

(h) Spot elevation, gradient tints, and shaded relief depicted for all elevations.

(i) The highest elevation for each 15-minute quadrangle is shown in thousands and hundreds of feet.

Figure 20-4. GNC Map.

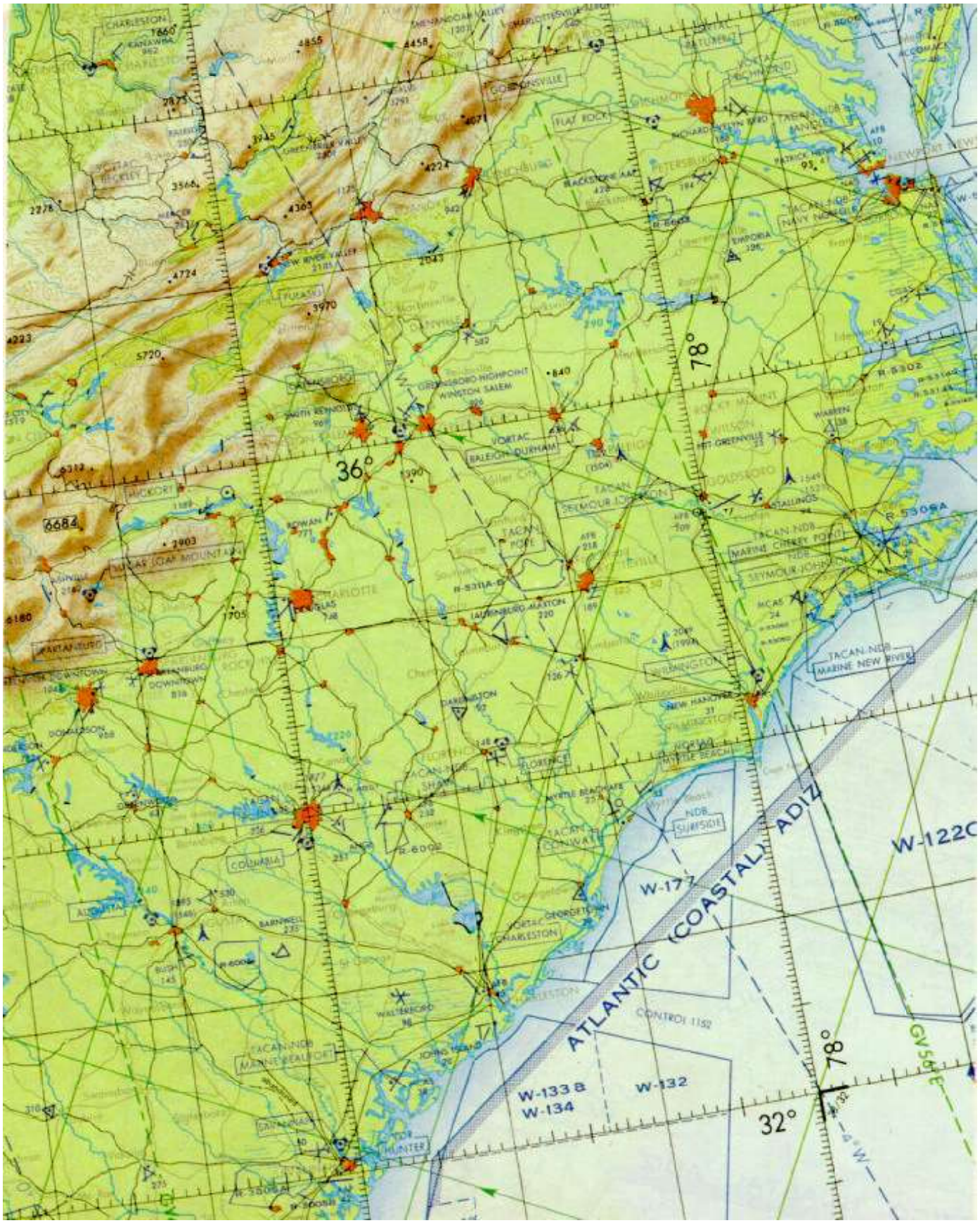

Figure 20-5. JNC Map.

Figure 20-6. ONC Map.

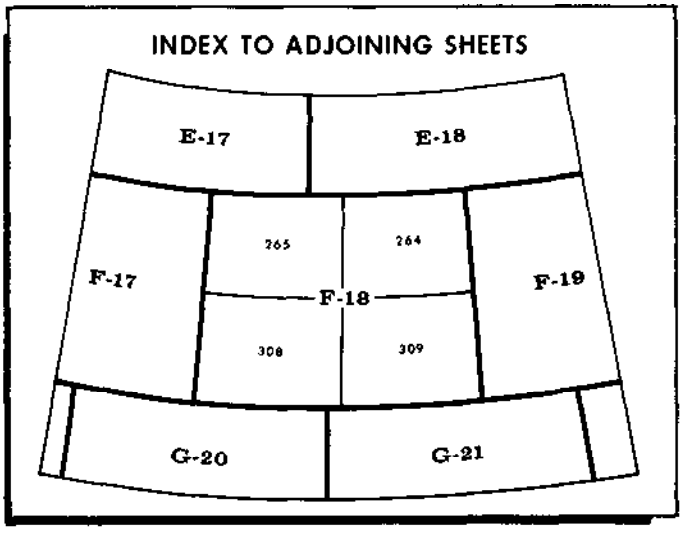

Figure 20-7. Operational Navigation Chart Index.

e. Joint Operations Graphic (JOG) (Series 1501 AIR):

(1) JOGs (figure 20-10) are series of 1:250,000 scale military maps designed for joint ground and air operations. The maps are published in ground and air editions. Both series emphasize the air-landing facilities but the air series has additional symbols to identify aids and obstructions to air navigation.

(2) JOG was designed to provide a common-scale graphic for Army, Navy, and Air Force use. Air Forces make use of it for tactical air operations, close air support, and interdiction by medium- and high-speed aircraft at low altitudes. The chart may also be used for dead reckoning and visual pilotage for short-range en route navigation. Due to its large scale, it is unsuitable for local area command planning for strategic and tactical operations.

(a) Relief on the JOG is indicated by contour lines (in feet). In some areas, the intervals may be in meters, with the approximate value in feet indicated in the margin of the chart. Spot elevations are used through all terrain levels. The ground series show elevations and contours in meters while the air series show the same elevations and contours in feet.

(b) Relief is also shown through gradient tints, supplemented by shaded relief. The highest elevations in each 15-minute quadrangle are indicated in thousands and hundreds of feet.

(c) Cultural features, such as cities, towns, roads, trails, and railroads are illustrated in detail. The locations of boundaries and power transmission lines are also shown. Vegetation is shown by symbol. Detailed drainage patterns and water tint are used to illustrate water features, such as coastlines, oceans, lakes, rivers and streams, canals, swamps, and reefs. The JOG includes aeronautical information such as airfields, fixed radio navigation and communication facilities, and all known obstructions over 200 feet above ground. If the information is available, the airfield runway patterns are shown to scale by diagram.

(d) The basic numbering system of the JOG consists of two letters and a number which identifies an area 6° in longitude by 4° in latitude. If the chart covers an area north of the Equator, the first letter is "N;" a chart covering an area south of the Equator is identified with an initial "S." The second letter identifies the 4° bands of latitude lettered north and south from the Equator. The number identifies the 6° sections of longitude which are numbered from the 180° meridian eastward. The 6° x 4° areas identified by two letters and a number from 1 to 60 are further broken down to either 12 or 16 sheets. Figure 20-11 illustrates how the sheets are numbered in each breakdown. The figure also indicates the respective latitudes at which the 12- and 16-sheet breakdown is used. Charts produced in Canada use a slightly different sheet identification system. The DOD Aeronautical Chart Catalog contains an explanation of the system.

f. DOD Evasion Charts (Figure 20-12). The Defense Mapping Agency and Aeronautical Chart and Information Center prepare DOD evasion charts. The Korea and Southeast Asia charts have been completed. The scale for these charts is 1:250,000. The charts have both longitude and latitude and the UTM grid coordinate systems. The relief is duplicated by both contour lines and shading. The magnetic variation is shown by a compass rose superimposed on the chart. The charts also indicate the direction of seasonal ocean currents. These charts may include geographic environmental data consisting of a description of the people, climate, water, food, hazards, and vegetation. A conversion of elevation bar scale may aid in communicating with other forces. The star chart is provided to aid in night navigation.

20-4. Information Contained in Margin:

a. Before using any piece of equipment, a wise operator always reads the manufacturer's book of instructions. This is also true with maps. The instructions are placed around the outer edges of the map and are known as marginal information. All maps are not the same, so it becomes necessary each time a different map is used to carefully examine the marginal information.

b. Figure 20-13 is a large-scale (1:50,000) topographic map. The circled numbers indicate the marginal information with which the map user must be familiar. The location of the marginal information will vary with each different type of map. However, the following items are on most maps. The circled numbers correspond to the item numbers listed and described below.

(1) Sheet Name (1). The sheet name is in two places; the center of the upper margin and the right side of the lower margin. Generally, a map is named after its

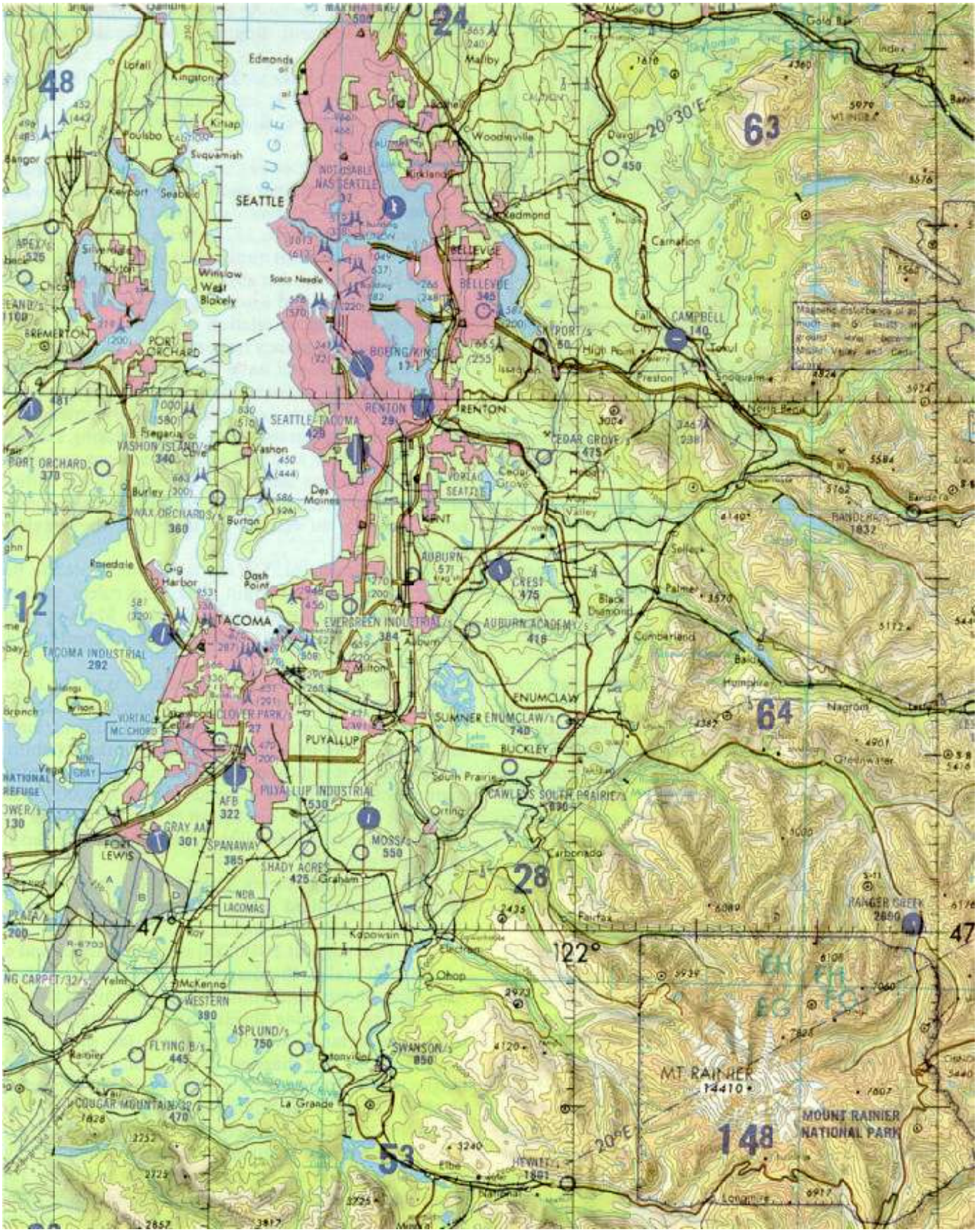

Figure 20-8. TPC Map.

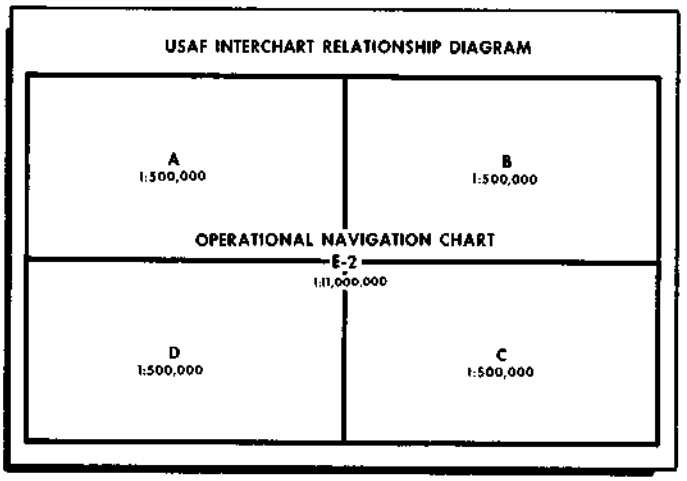

Figure 20-9. Relationship of TPC and ONC.

outstanding cultural or geographic feature. When possible, the name of the largest city on the map is used (not shown).

(2) Sheet Number (2). The sheet number is in the upper right margin and is used as a reference number for that map sheet. For maps at 1:100,000 scale and larger, sheet numbers are based on an arbitrary system which makes possible the ready orientation of maps at scales of 1:100,000, 1:50,000, and 1:25,000 (figures 20-14 and 20-15).

(3) Series Name and Scale (3):

(a) The map series name is in the upper left margin. A map series usually comprises a group of similar maps at the same scale and on the same sheet lines or format designed to cover a particular geographic area. It may also be a group of maps which serve a common purpose, such as military city maps. The name given a series is of the most prominent area. The scale note is a representative fraction which gives the ratio of map distance to the corresponding distance on the Earth's surface. For example, the scale note 1:50,000 indicates that one unit of measure on the map equals 50,000 units of the same measure on the ground.

(b) Scale. The scale is expressed as a fraction and gives the ratio of map distance to ground distance. The terms "small scale," "medium scale," and "large scale" may be confusing when read with the numbers. However, if the number is viewed as a fraction, it quickly becomes apparent the 1:600,000 of something is smaller than 1:75,000 of the same thing. Hence, the larger the number after 1:, the smaller the scale of the map.

-1. Small Scale. Maps at scales of 1:600,000 and smaller are used for general planning and strategical studies at the high echelons. The standard small scale is 1:1,000,000.

-2. Medium Scale. Maps at scales larger than 1:600,000 but smaller than 1:75,000 are used for planning operations, including the movement and concentration of troops and supplies. The standard medium scale is 1:250,000.

-3. Large Scale. Maps at scales of 1:75,000 and larger are used to meet the tactical, technical, and administrative needs of field units. The standard large scale is 1:50,000.

(4) Series Number (4). The series number appears in the upper right margin and the lower left margin. It is a comprehensive reference expressed either as a four-digit numeral (example, 1125), or as a letter, followed by a three- or four-digit numeral (example, V7915).

(5) Edition Number (5). The edition number is in the upper margin and lower left margin. It represents the age of the map in relation to other editions of the same map and the agency responsible for its production. The latest edition will have the highest number. EDITION 1 DMATC indicates the first edition prepared by the Defense Mapping Agency Topological Center. Edition numbers run consecutively; a map bearing a higher edition number is assumed to contain more recent information than the same map bearing a lower edition number. Advancement of the edition number constitutes authority to rescind or supersede the previous edition.

(6) Bar Scales (6). The bar scales are located in the center of the lower margin. They are rulers used to convert map distance to ground distance. Maps normally have three or more bar scales, each a different unit of measure.

(7) Credit Note (7). The credit note is in the lower left margin. It lists the producer, dates, and general methods of preparation or revision. This information is important to the map user in evaluating the reliability of the map as it indicates when and how the map information was obtained. On some recent 1:50,000 scale maps, the map credits are shown in tabular form in the lower margin, with reliability information presented in a coverage diagram.

(8) Adjoining Sheets Diagram (8) (not shown). Maps at all standard scales contain a diagram which illustrates the adjoining sheets.

(a) On maps at 1:100,000 and larger scales and at 1:1,000,000 scales, the diagram is called the Index to Adjoining Sheets, and consists of as many rectangles, representing adjoining sheets, as are necessary to surround the rectangle which represents the sheet under consideration. The diagram usually contains nine rectangles, but the number or names may vary depending on the location of the adjoining sheets. All represented sheets are identified by their sheet numbers. Sheets of an adjoining series, whether published or planned, that are the same scale are represented by dashed lines. The series number of the adjoining series is indicated along the appropriate side of the division line between the series (figure 20-16).

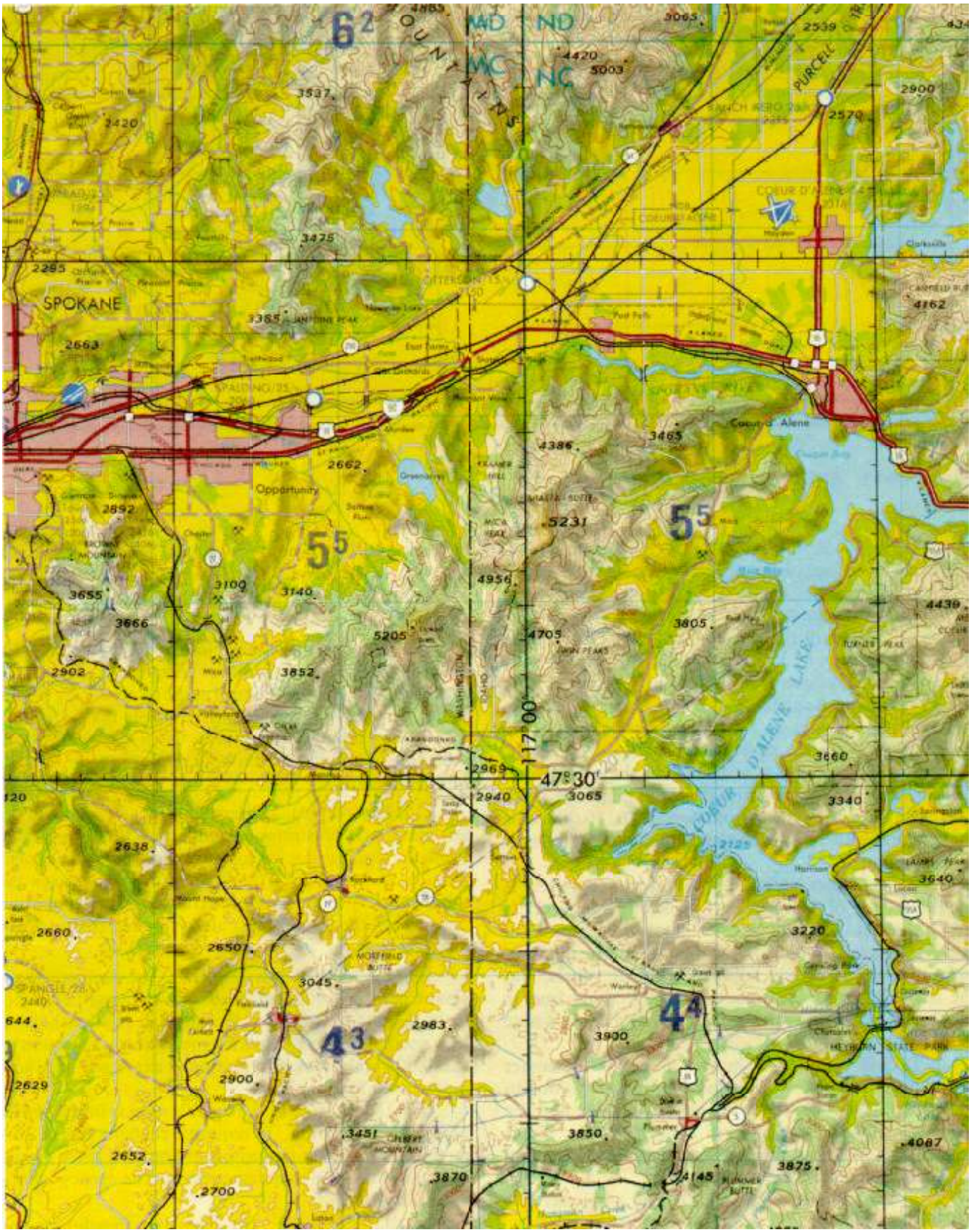

Figure 20-10. Joint Operations Graphics Map.

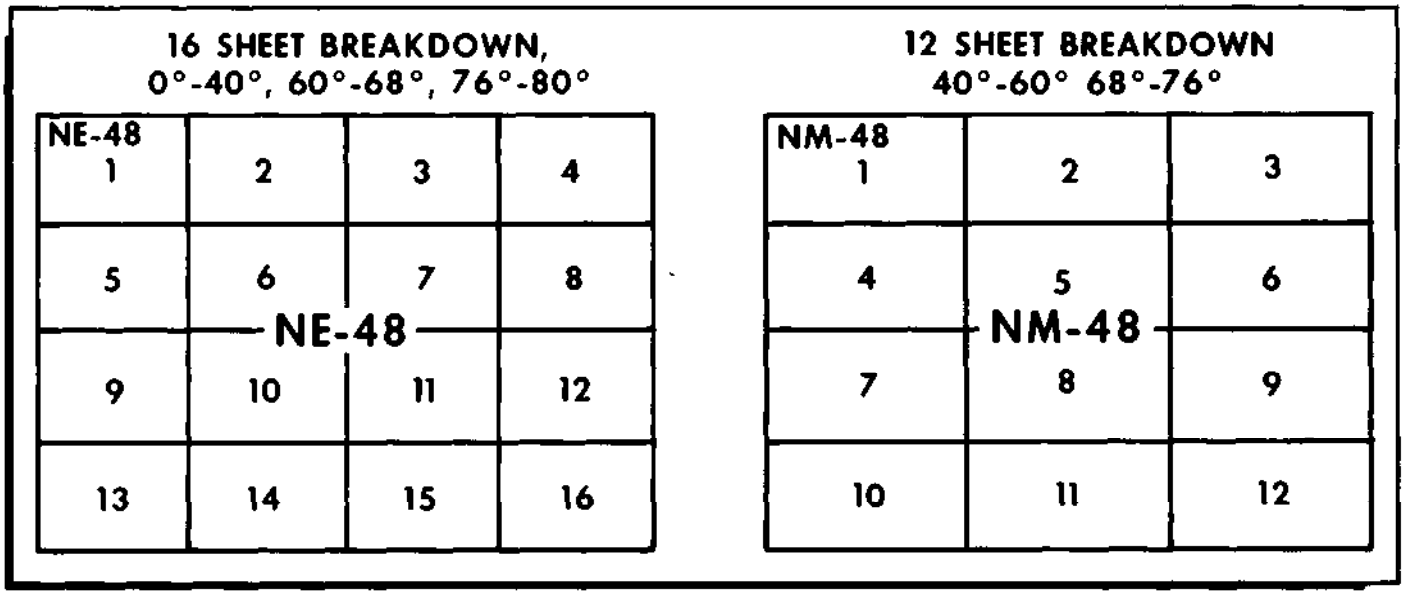

Figure 20-11. JOG Sheet Numbering System.

(b) On 1:50,000 scale maps, the sheet number and series number of the 1:250,000 scale map of the area are shown below the Index to Adjoining Sheets.

(c) On maps at 1:250,000 scale, the adjoining sheets are shown in the location diagram. Usually, the diagram consists of 25 rectangles, but the number may vary with the locations of the adjoining sheets.

(9) Index to Boundaries (9). The index to boundaries diagram appears in the lower right margin of all sheets 1:100,000 scale or larger, and 1:1,000,000 scale. This diagram, which is a miniature of the map, shows the boundaries which occur within the map area, such as county lines and state boundaries. On 1:250,000 scale maps, the boundary information is included in the location diagram.

(10) Projection Note (10). The projection system is the framework of the map. For maps, this framework is the conformal type; that is, small areas of the surface of the Earth retain their true shapes on the projection, measured angles closely approximate true values, and the scale factor is the same in all directions from a point. The projection is identified on the map by a note in the lower margin.

(11) Grid Note (11). The grid note is in the center of the lower margin. It gives information pertaining to the grid system used, the interval of grid lines, and the number of digits omitted from the grid values. Notes pertaining to overlapping or secondary grids are also included when appropriate.

(12) Grid Reference Box (12). The grid reference box has instructions for composing a grid reference.

(13) Vertical Datum Note (13). This note is in the center of the lower margin. It designates the basis for all vertical control stations, contours, and elevations appearing on the map. On JOGs at 1:250,000 scale, the vertical datum note may appear in the reliability diagram.

(14) Horizontal Datum Note (14). This note is located in the center of the lower margin. It indicates the basis for all horizontal control stations appearing on the map. This network of stations controls the horizontal positions of all mapped features. On JOGs at 1:250,000 scale, the horizontal datum note may appear in the reliability diagram.

(15) Legend (15). The legend is located in the lower left margin. It illustrates and identifies the topographic symbols used to depict some of the more prominent features on the map. The symbols are not always the same on every map. To avoid error in the interpretation of symbols, the legend must always be referred to when a map is read.

(16) Declination Diagram (16). The declination diagram is usually located in the lower margin of large-scale maps and indicates the angular relationships of true north, grid north, and magnetic north. On maps at 1:250,000 scale, this information is expressed as a note in the lower margin.

(17) User's Note (17). A user's note is in the center of the lower margin. It requests cooperation in correcting errors or omissions on the map. Errors should be marked and the map forwarded to the agency identified in the note.

(18) Unit Imprint (18). The unit imprint, in the lower left margin, identifies the agency which printed the map and the printing date. The printing date should

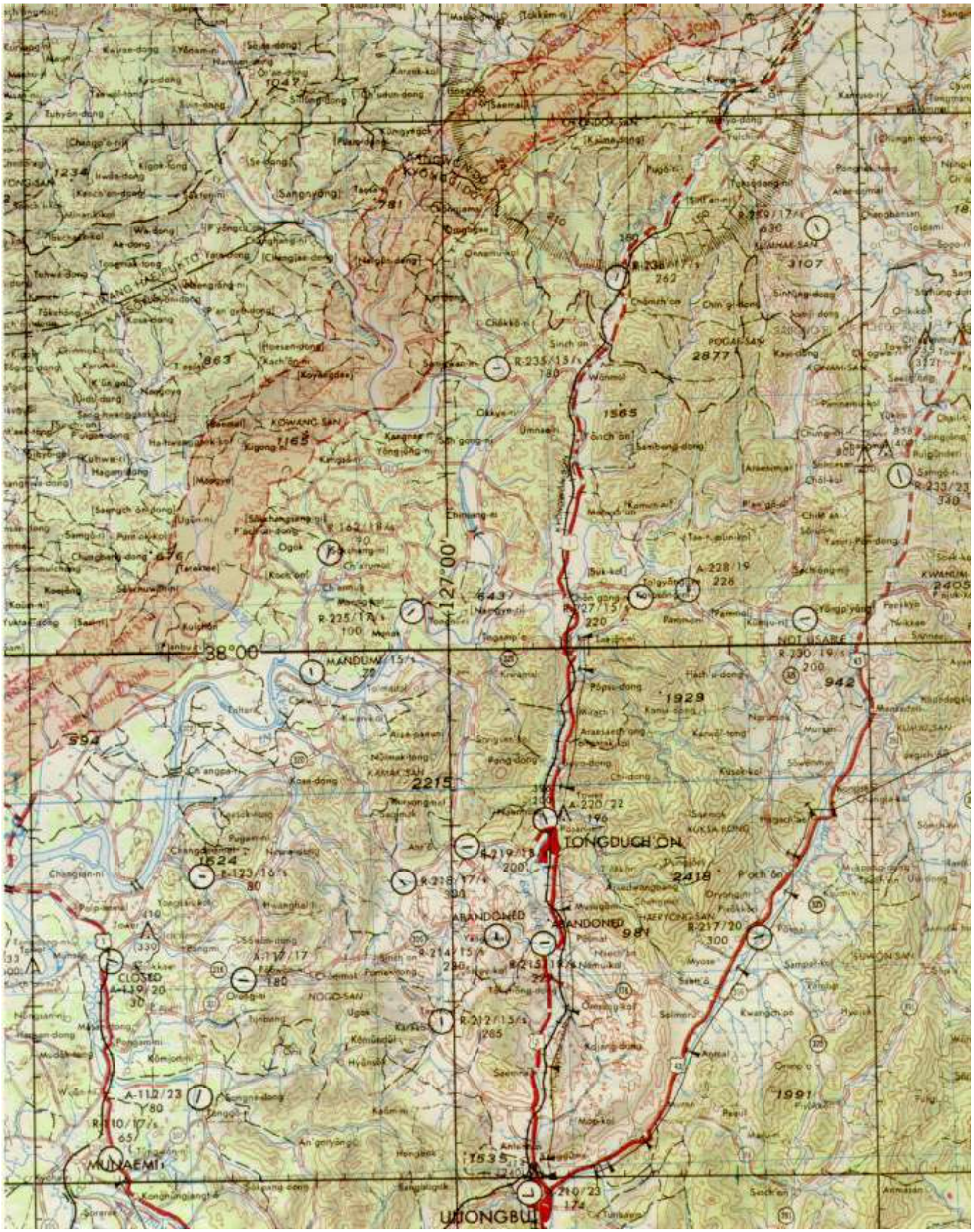

Figure 20-12. DOD Evasion Chart.

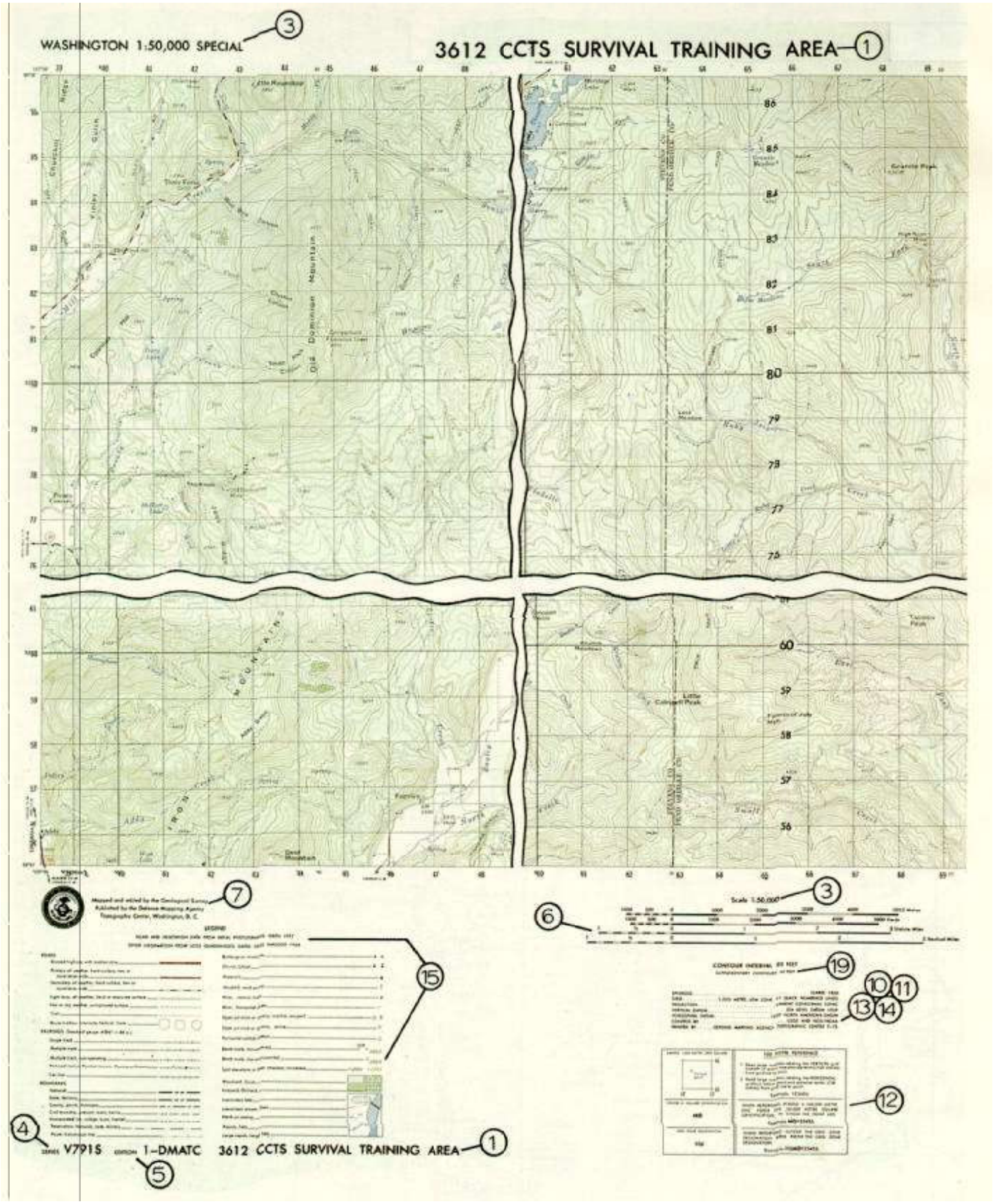

Figure 20-13. 1:50,000 Topographic Map.

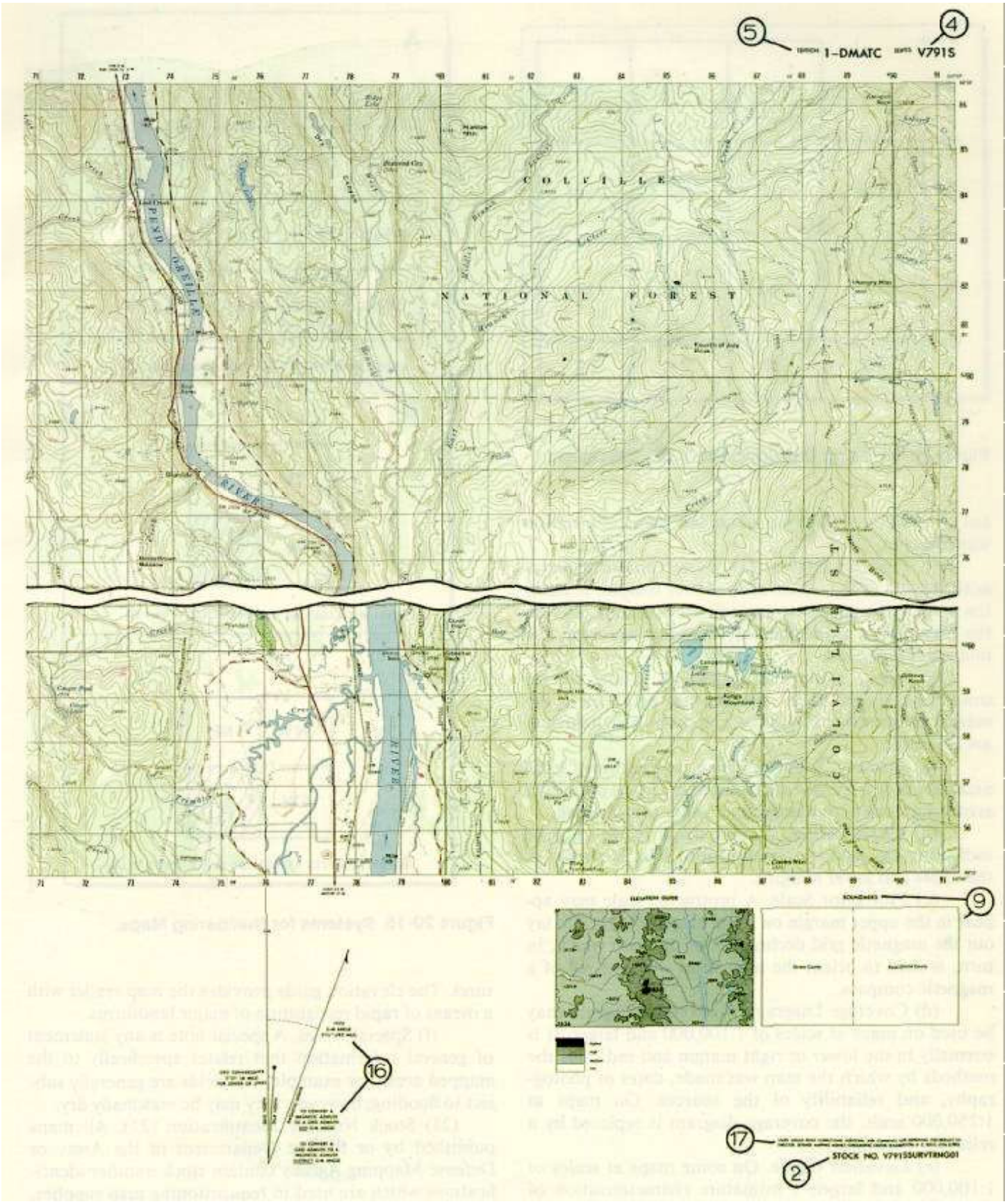

Figure 20-13. 1:50,000 Topographic Map. (continued)

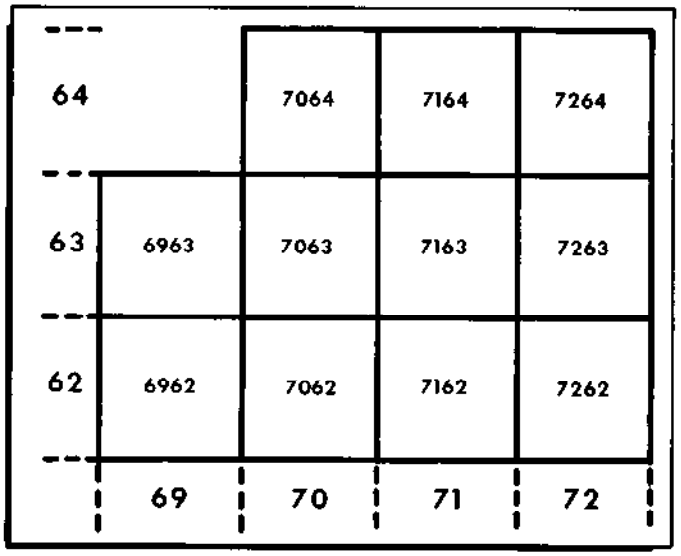

Figure 20-14. Basic Development, 1:100,000 Scale.

not be used to determine when the map information was obtained.

(19) Contour Interval (19). The contour interval note appears in the center of the lower margin. It states the vertical distance between adjacent contour lines on the map. When supplementary contours are used, the interval is indicated.

(20) Special Notes and Scales (20). Under certain conditions, special notes or scales may be added to the margin information to aid the map user. The following are examples:

(a) Glossary. A glossary is an explanation of technical terms or a translation of terms on maps of foreign areas where the native language is other than English.

(b) Classification. Certain maps require a note indicating the security classification. This is shown in the upper and lower margins.

(c) Protractor Scale. A protractor scale may appear in the upper margin on some maps. It is used to lay out the magnetic grid declination of the map which, in turn, is used to orient the map sheet with the aid of a magnetic compass.

(d) Coverage Diagram. A coverage diagram may be used on maps at scales of 1:100,000 and larger. It is normally in the lower or right margin and indicates the methods by which the map was made, dates of photography, and reliability of the sources. On maps at 1:250,000 scale, the coverage diagram is replaced by a reliabilty diagram.

(e) Elevation Guide. On some maps at scales of 1:100,000 and larger, a miniature characterization of the terrain is shown by a diagram in the lower right margin of the map. The terrain is represented by bands of elevation, spot elevations, and major drainage features. The elevation guide provides the map reader with a means of rapid recognition of major landforms.

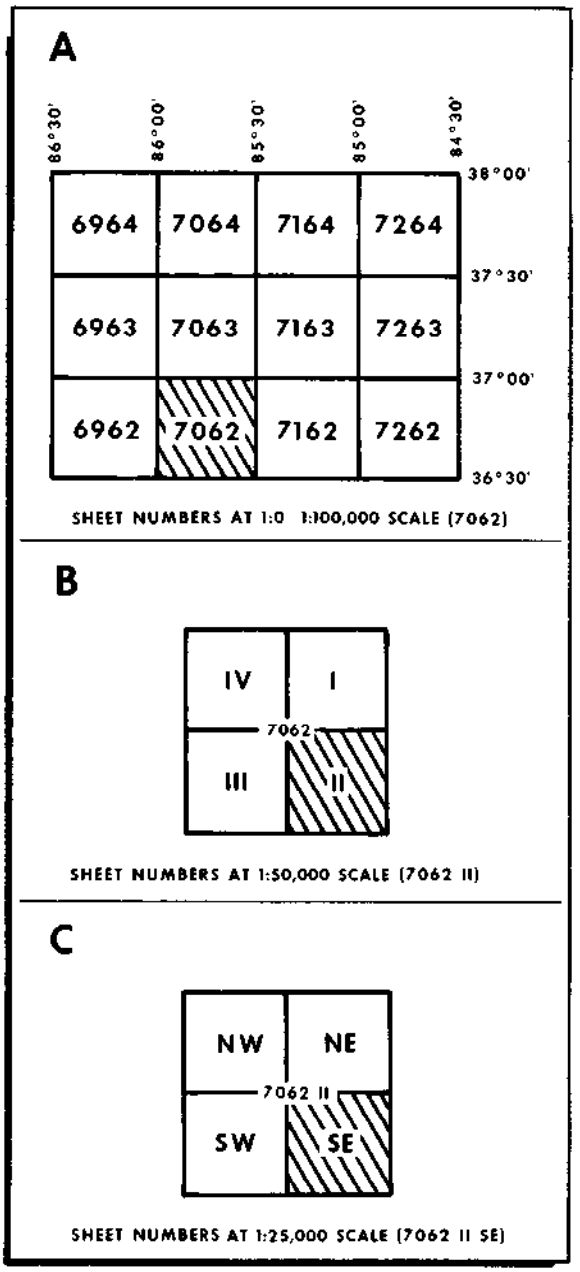

Figure 20-15. Systems for Numbering Maps.

(f) Special Notes. A special note is any statement of general information that relates specifically to the mapped area. For example, rice fields are generally subject to flooding; however, they may be seasonally dry.

(21) Stock Number Identification (21). All maps published by or for the Department of the Army or Defense Mapping Agency contain stock number identifications which are used in requisitioning map supplies. The identification consists of the words "STOCK NO." followed by a unique designation which is composed of the series number, the sheet number of the individual

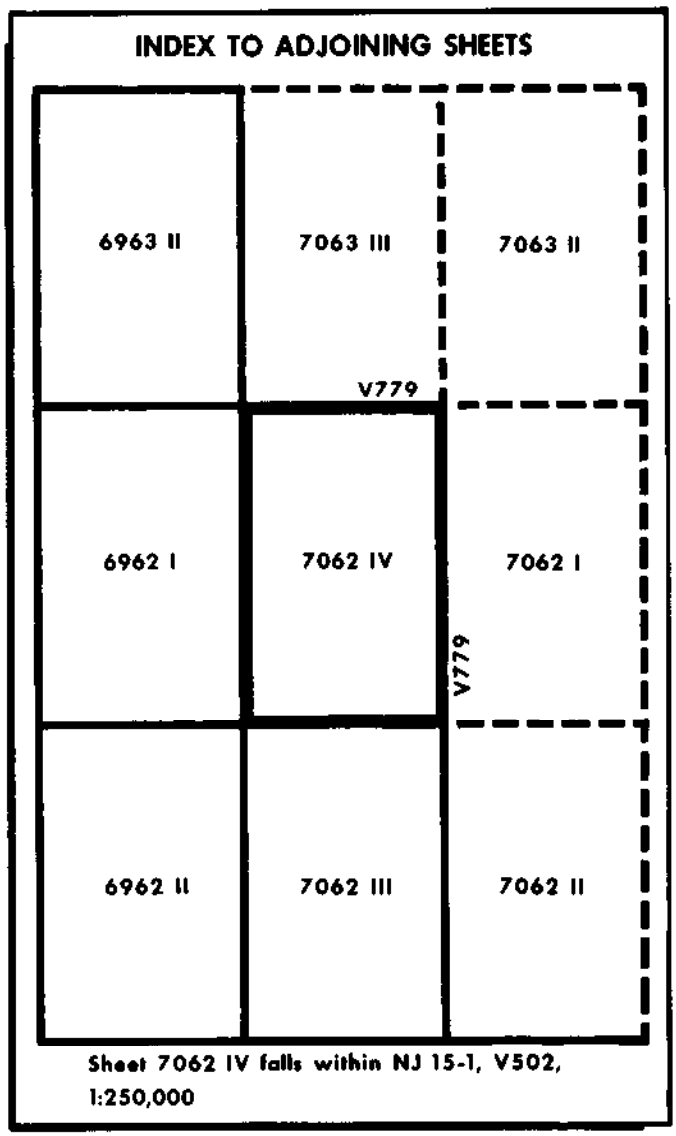

Figure 20-16. Index to Adjoining Sheets.

map, and on recently printed sheets, the edition number.

20-5. Topographic Map Symbols and Colors:

a. The purpose of a map is to permit one to visualize an area of the Earth's surface with pertinent features properly positioned. Ideally, all the features within an area would appear on the map in their true proportion, position, and shape. This, however, is not practical because many of the features would be unimportant and others would be unrecognizable because of their reduction in size. The mapmaker has been forced to use symbols to represent the natural and manmade features of the Earth's surface. These symbols resemble, as closely as possible, the actual features as viewed from above (figures 20-17 and 20-18).

b. To facilitate identification of features on the map by providing more natural appearance and contrast, the topographic symbols are usually printed in different colors, with each color identifying a class of features. The colors vary with different types of maps, but on a standard large-scale topographic map, the colors used and the features represented are:

(1) Black—the majority of cultural or manmade features.

(2) Blue—water features such as lakes, rivers, and swamps.

(3) Green—vegetation such as woods, orchards, and vineyards.

(4) Brown—all relief features such as contours.

(5) Red—main roads, built-up areas, and special features.

(6) Occasionally, other colors may be used to show special information. (These, as a rule, are indicated in the marginal information. For example, aeronautical symbols and related information for air-ground operations are shown in purple on JOGs.)

c. In the process of making a map, everything must be reduced from its size on the ground to the size which appears on the map. For purposes of clarity, this requires some of the symbols to be exaggerated. They are positioned so that the center of the symbol remains in its true location. An exception to this would be the position of a feature adjacent to a major road. If the width of the road has been exaggerated, then the feature is moved from its true position to preserve its relation to the road.

d. Army Field Manual 21-31 gives a description of topographic symbols and abbreviations authorized for use on US military maps. Figure 20-19 illustrates several of the symbols used on maps.

20-6. Coordinate Systems. The intersections of reference lines help to locate specific points on the Earth's surface. Three of the primary reference line systems are the geographic coordinate system, the reference (GEOREF) system, and the universal transverse mercator grid system (UTM). Knowing how to use these plotting systems should help a survivor to determine point locations.

a. Coordinates. Quantities that give position with respect to two reference lines are called coordinates. Thus, the intersection of F Street and 4th Avenue (figure 20-20) is the coordinate location of the Gridville Public Library. The coordinates of the local theater are D Street and 6th Avenue. One can see from this simplified example that coordinates are read at intersections of vertical and horizontal lines. The basic coordinate system used on maps and charts is the geographic military grid. The structure and use of the geographic coordinate system, the world geographic reference system, and the military grid reference system will be discussed and illustrated.

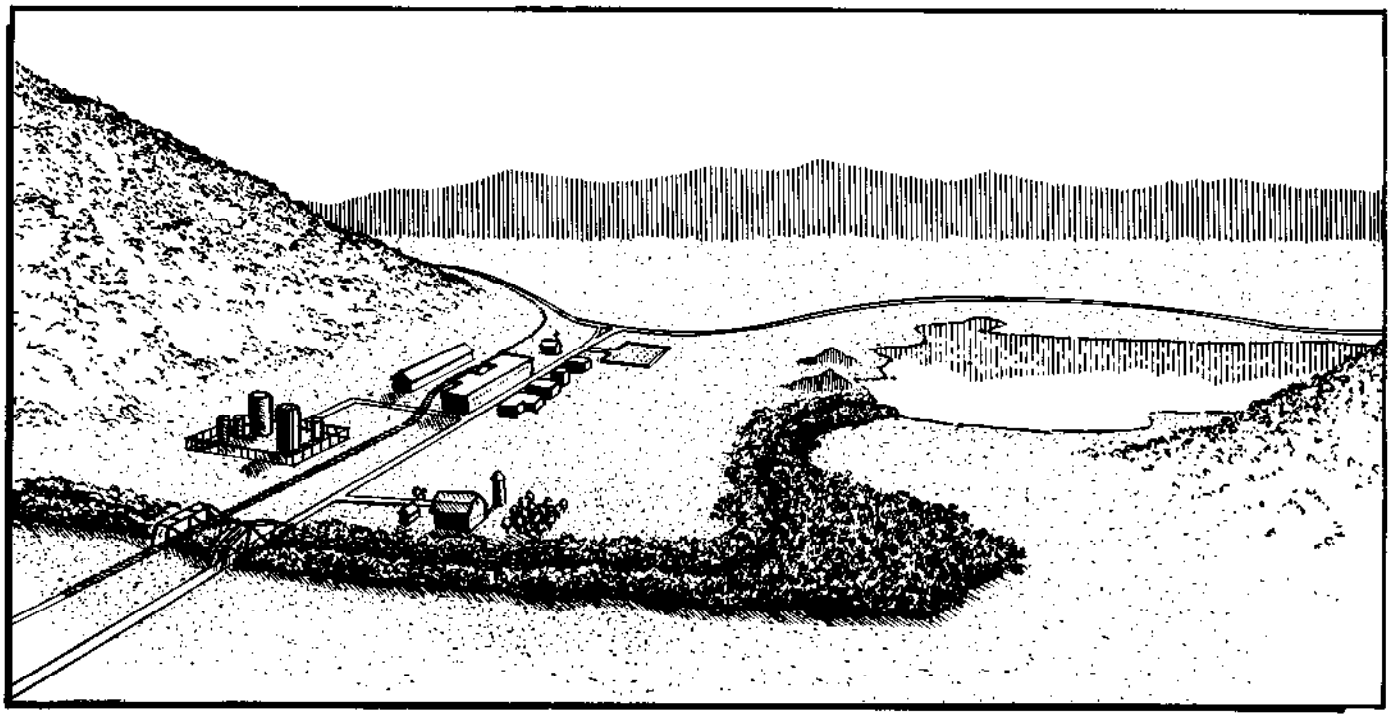

Figure 20-17. Area Viewed from Ground Position.

(1) Geographic Coordinates. The geographic coordinate system is a network of imaginary lines that circle the Earth. They are used to express Earth position or location. There are north-south lines called meridians of longitude and east-west lines named parallels of latitude. The location of any point on the Earth can be expressed in terms of the intersection of the line of latitude and the line of longitude passing through the point.

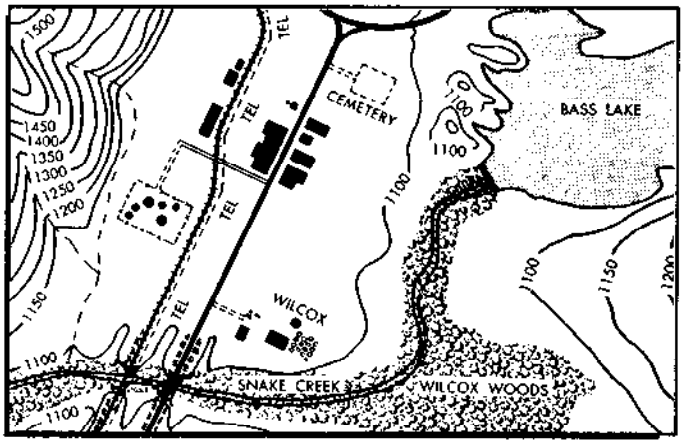

Figure 20-18. Area Viewed from Ground Position—Map.

(2) Meridians of Longitude. The lines of latitude and longitude are actually great and small circles formed by imaginary planes cutting the Earth. A great circle divides the Earth into two equal parts (halves); whereas, a small circle divides the Earth into two unequal parts. Study figure 20-21 and note that: (1) each north-south line is a great circle, and (2) each great circle passes through both the North and South Poles. Each half of each of these great circles from one pole, in either direction, to the other pole is called a meridian of longitude. The other half of the same great circle is a second meridian of longitude.

(a). Meridian is derived from the Latin word "meridianum," which means "lines that pass through the highest point on their course" (in this case, both the North and South Poles). Any angular distance measured east or west of the meridian is called longitudinal distance; hence, the term "meridian of longitude." It is necessary, of course, to assign values to the meridians to make them meaningful. The most appropriate values to use for circles are degrees (°), minutes (′), and seconds (″). Circles are customarily divided into 360° per circle, 60′ per degree, and 60″ per minute.

(b). All meridians are equal in value; hence, one of them must be assigned the value of 0° (the starting point). The meridian passing through Greenwich, England, is zero degrees (0°). This meridian is also called the prime meridian (figure 20-22). The other half of the great circle on which the prime meridian is located is designated the 180th meridian. Portions of this meridian are also called the international dateline.

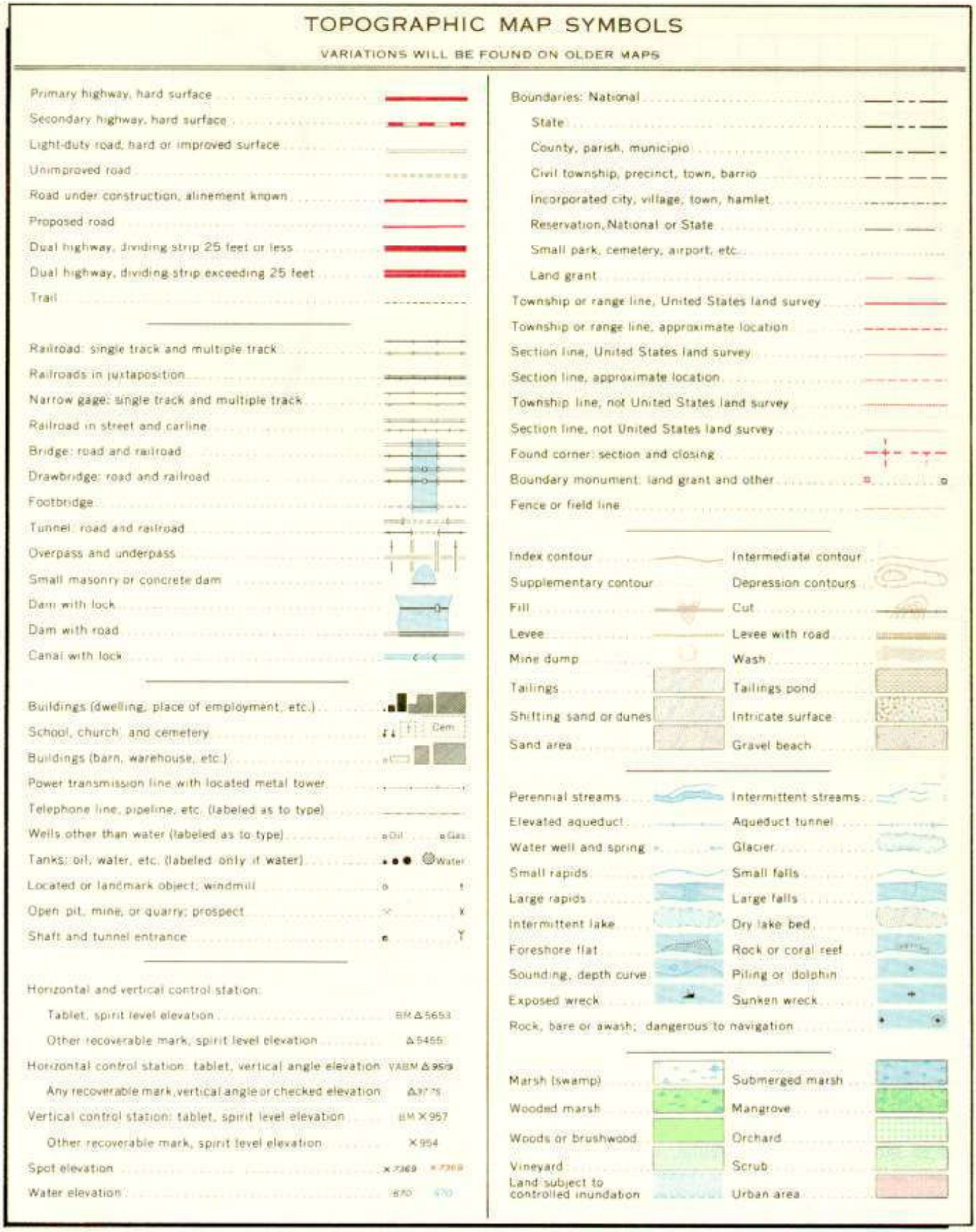

Figure 20-19. Topographic Map Symbols.

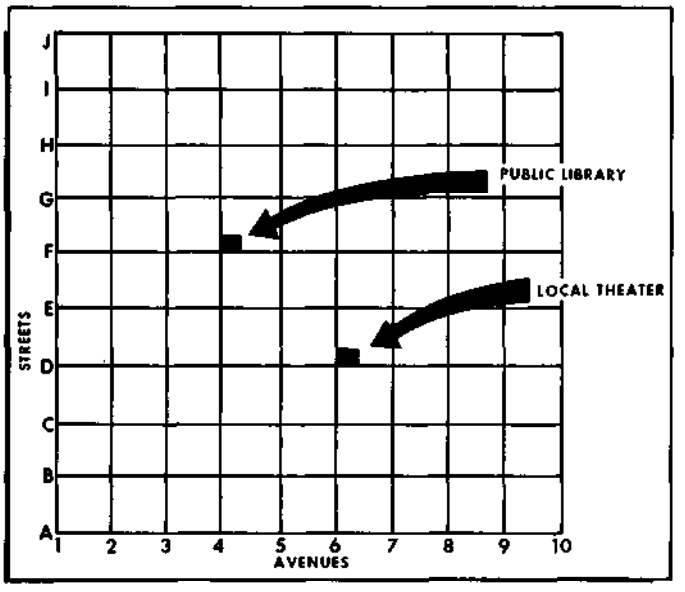

Figure 20-20. Gridville City.

(c). From the prime meridian east of the international dateline, meridians are assigned values of 0° through 180° east. Similarly, from the prime meridian

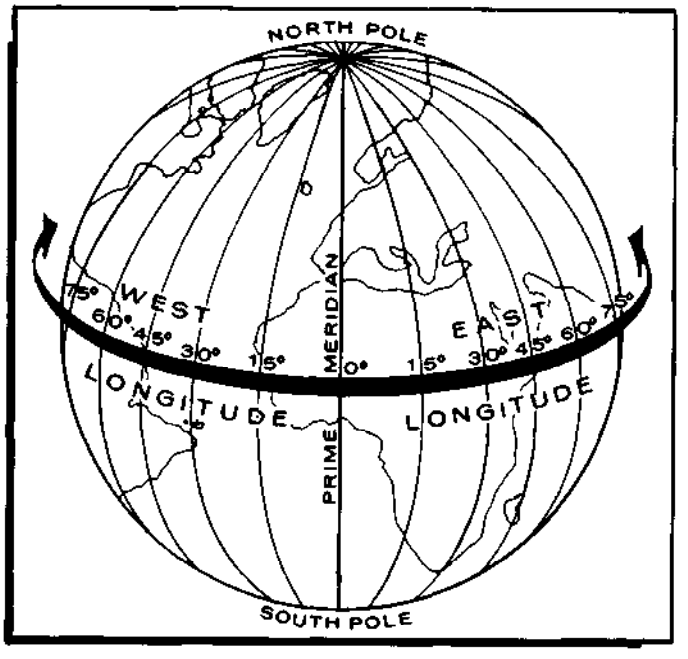

Figure 20-21. Meridian of Longitude.

west to the international dateline, meridians are assigned values of 0°through 180° west. The 0° meridian together with the 180° meridian forms a great circle which divides the Earth into east and west longitude (or hemispheres). There are 180° of east longitude plus 180° of west longitude for 360° of longitude.

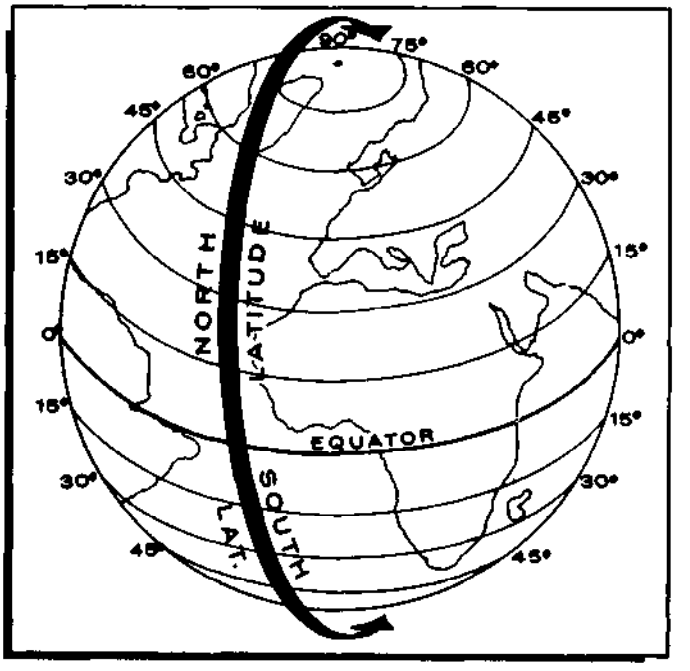

Figure 20-22. Parallels of Latitude.

b. Parallels of Latitude. Notice in figure 20-22 that the circles running in an east-west direction are of varying diameters (sizes). Only the circle designated "Equator" is a great circle. All others are small circles. Note that all circles are parallel to the Equator and run laterally around the Earth. Hence, each circle is called a parallel of latitude. Unlike meridians, which extend only halfway around the Earth, a parallel of latitude extends all the way around the Earth; for the record, the Equator is also a parallel of latitude. Since the Equator is the only great circle of latitude, it is a natural starting point for the 0° value of latitude. The North and South Poles are designated 90° north latitude and 90° south latitude, respectively. Parallels between the Equator and North Pole carry values between 0° and 90°north; parallels between the Equator and the South Pole are assigned values between 0° and 90° south.

(1) Figure 20-23 combines the lines of latitude and longitude. Lines 0°through 90° north or south latitude and 0° through 180° east or west longitude form the grid of the geographic coordinate system. Study the positions of Points A and B in figure 20-23. Determine the geographic coordinates of each in degrees. Note that point A is positioned 32° north of the Equator and 35° east of the prime meridian. The geographic position of point A, therefore, is 32°north 35° east. Point B is located 25° south of the Equator and 40° west of the prime meridian. Hence, the geographic position of point B is 25° south 40°west.

(2) Just as any point within the city of Gridville (figure 20-20) can be referenced by the intersection of

two imaginary lines, any point on the Earth's surface can be referenced by the intersection of the imaginary lines of latitude and longitude.

c. Writing Geographic Coordinates. To illustrate the proper way to write geographic coordinates, let's assume that a person needs to write the coordinates of a target. The target is located 30°20′ north of the Equator and 135°06′ east of the prime meridian. Thus, the position is located at 30°20′ north latitude and 135°06′ east longitude. By combining latitude and longitude, the position of the geographic location can be expressed as 30°20′N 135°06′E. To write these coordinates in the correct military form, eliminate the degree (°) and minute (′) symbols. Thus, the coordinates would be written 302000N1350600E.

(1) Writing geographic coordinates in the military form is necessary for wire and radio transmission of geographic coordinates. Why? The transmission equipment does not include the degree (°), minute (′), and second (″) characters in its keyboards. Coordinates are also stored in automated data processing computers which are programmed to handle coordinates in military characters or spaces. If the sequence of numbers and letters fed into a computer is less than 15 spaces, or in error, the resulting printout will be meaningless.

(2) When a position is located that is less than 10° latitude, a zero is added to the left of the degree number. For example, 7° of latitude is written as 07. Likewise, two digits always designate minutes and two digits for seconds. Thus, 7°N becomes 07N; 7°6′N becomes 0706N; and 7°6′5″N becomes 070605N. In expressing longitude, three digits are required to indicate degrees, two digits for minutes, and two digits for seconds. Thus

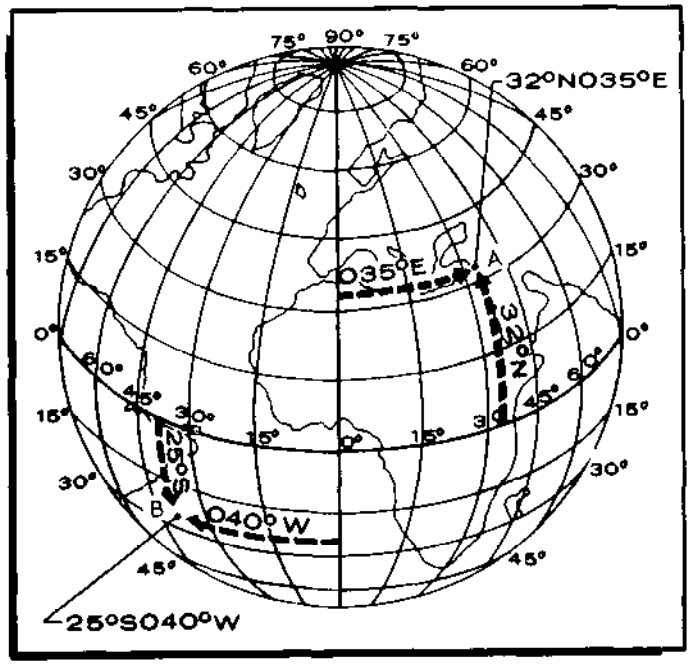

Figure 20-23. Latitudes and Longitudes.

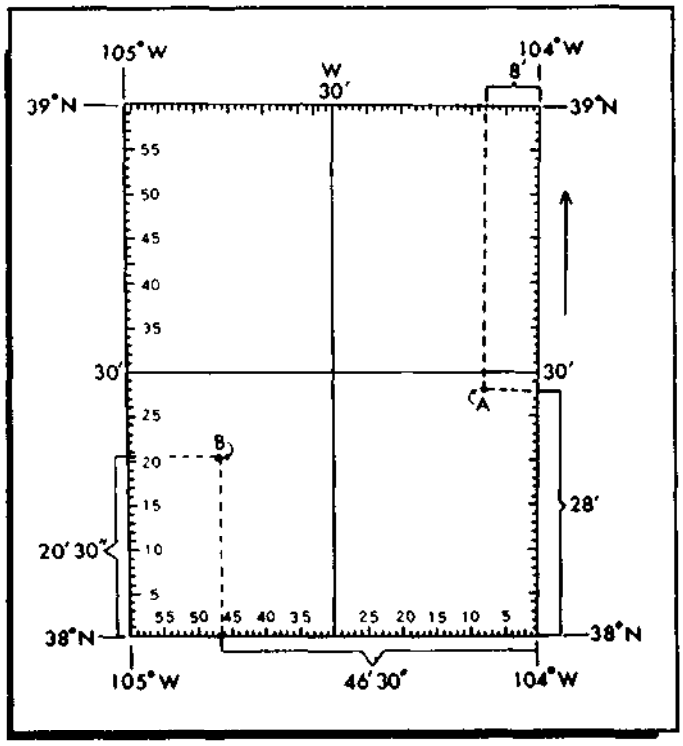

Figure 20-24. Plotting Geographic Coordinates.

8°E becomes 008E; 8°5′E becomes 00805E; and 8°5′4″E becomes 0080504E.

(3) In general, there are five rules to follow in correctly writing geographic coordinates:

(a) Write latitude first, followed by longitude.

(b) Use an even number of digits for latitude and an odd number of digits for longitude.

(c) Do not use a dash or leave a space between latitude and longitude.

(d) Use single upper case letter to indicate direction from the Equator and prime meridians.

(e) Omit the symbols for degrees, minutes, and seconds.

d. Plotting Geographic Coordinates:

(1) One can probably read the coordinates of point A and B in figure 20-24 rather easily; however, plotting points on maps from given coordinates must also be done. To do this, first get acquainted with the map being used. Assume that figure 20-24 is the map being used. Note that it covers an area from 38N to 39N and from 104W to 105W, an area of 1° by 1°. Also note that latitude and longitude are subdivided by 30′ division lines and then with tick marks into 5- and 1-minute subdivisions.

(2) Assume that the coordinates of the point which must be plotted are 382800N1040800W. Next, follow the general procedure listed below to plot the point on the map:

(a) Locate the parallel of latitude for degrees (38°N).

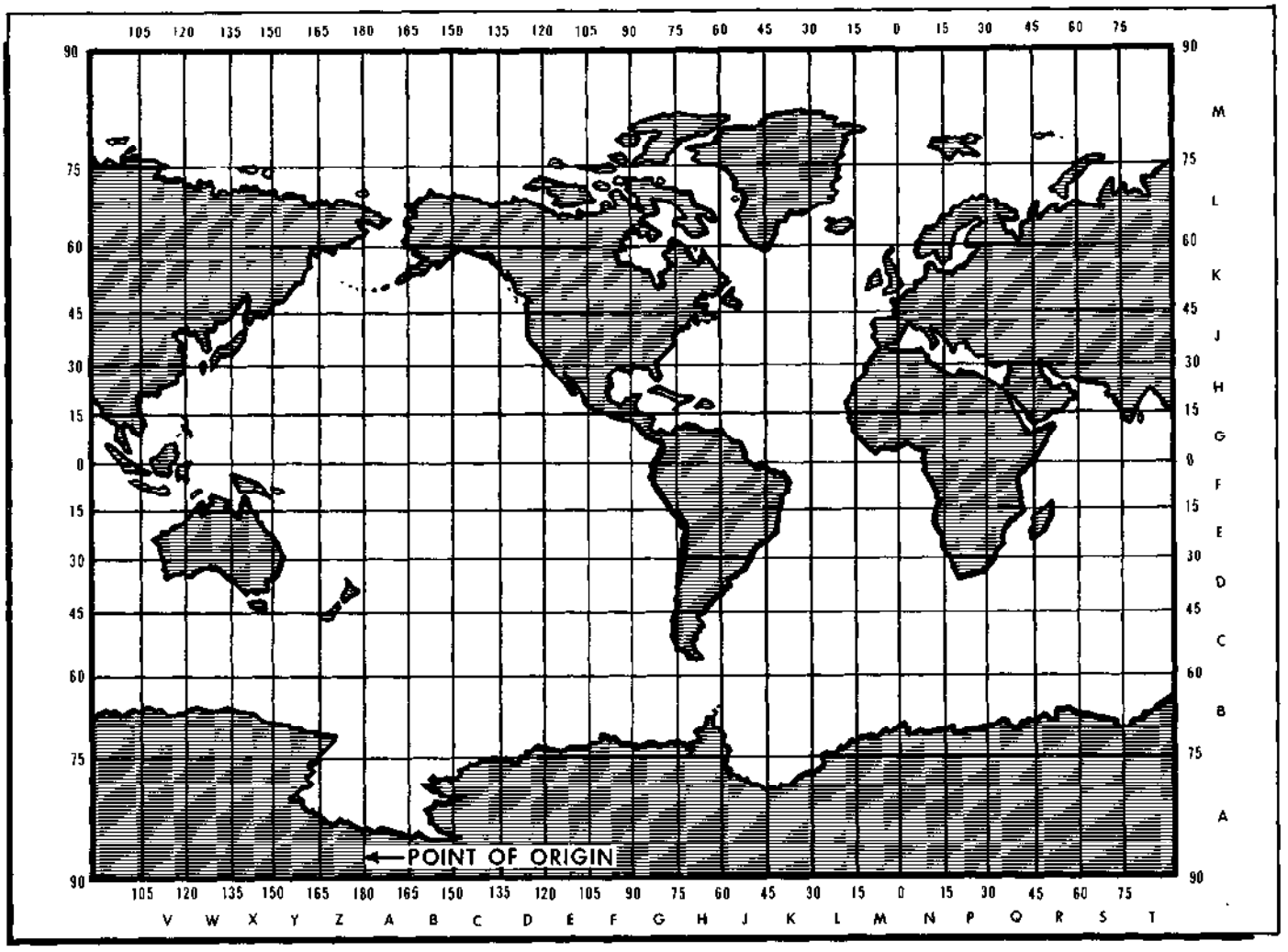

Figure 20-25. GEOREF 15-Degree Quadrants.

(b) Find the meridian of longitude for degrees (104°W).

(c) Move to the meridian (usually a tick mark) for minutes (08°W).

(d) Move to the parallel (usually a tick mark) for minutes (28′N).

(e) Plot the point on the map (point A in figure 20-24; plot at 382800N1040800W.

(3) Recovery points, rally points, and destination positions may be plotted or identified on a map or chart to enable rescue personnel, the survivors, and evaders to locate these positions. Seconds are not shown between the 1-minute tick marks on maps and charts; they must be estimated. It is easy to estimate halfway tick marks (30 seconds); one-fourth (15 seconds) and three-fourths (45 seconds) are also reasonably easy to estimate. Then, as experience is gained, people will find that on large-scale maps they can estimate the sixths (10 seconds) and eights (about 8 seconds). They cannot, however, accurately estimate to sixths or eights at the scale shown in figure 20-24.

(4) To write geographic coordinates more precisely than minutes, merely carry the coordinates out to include seconds. In the previous example, the coordinates of a target located 30°20′ north of the Equator and 135°06′ east of the prime meridian were written as 302000N1350600E. A more exact position of the target might be 30°20′05″N latitude and 135°06′16″E longitude. This more precise position is correctly written as 302005N1350616E.

e. World Geographic Reference System (GEOREF). The geographic coordinate system has several shortcomings when it is used in military operations. One objectionable feature is the large number of characters necessary to identify a location. To specify a location within 300 yards, a coordinate reading such as 241412NO141512W is necessary, with a total of 15 characters. Another objectionable feature is the diversity of directions used in applying the grid numbering system. Any particular point on a geographic grid can be north and east, north and west, south and east, or south and west. This means there are four different ways to proceed when reading various geographic coordinates.

Such a system obviously promotes errors. To overcome the disadvantages and promote speed in position reporting, other grid systems are used. We shall now examine one of these systems—that which is commonly called GEOREF. Air Force uses the GEOREF system as a reference in the control and direction of forces engaged in large area operations and operations of a global nature.

(1) GEOREF System Structure. The geographic coordinate grid serves as the base for the GEOREF system. The grid originates at the 180° meridian and the South Pole. Starting at the 180° meridian, it proceeds right, or eastward, around the world, and back to the 180° starting point. From the South Pole, it proceeds northward to the North Pole (figure 20-25).

(a) Notice in figure 20-25 that the basic layout is subdivided into 24 east-west zones and 12 north-south zones. This forms 288 quadrangles that measure 15°X15°. The 24 east-west zones are lettered "A" through "Z" omitting "I" and "O." The 12 north-south zones are lettered "A" through "M" (omitting "I"). Each quadrangle is identified by two letters and is located by reading right and up. For example, the southern tip of Florida is located in GEOREF quadrangle G-H (figure 20-25).

(b) Each of the 15° quadrangles is divided into 1° quadrangles (figure 20-26). First, they are divided to the right into 15 zones lettered "A" through "Q" (omitting "I" and "O"), then up into 15 zones lettered "A" through "Q" ("I" and "O" omitted).

(c) This system makes it possible to identify any quadrangle by four letters; for example, WGAN. The two letters designate the 15° grid zone, and the other

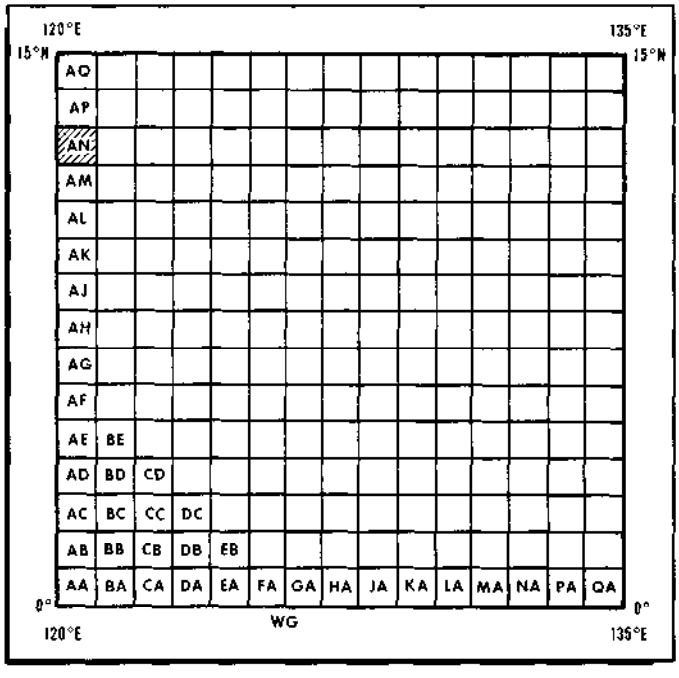

Figure 20-26. GEOREF 1-Degree Quadrants.

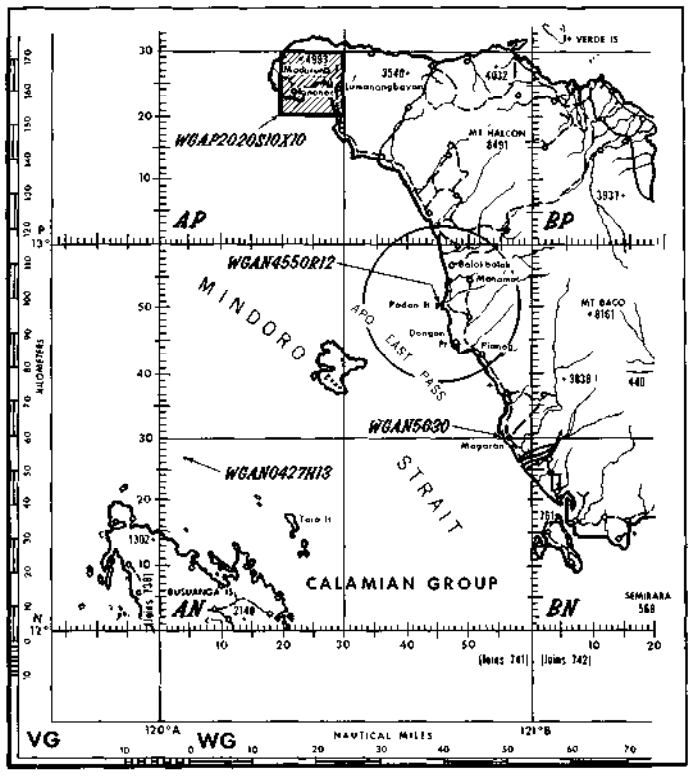

Figure 20-27. GEOREF 1-Degree Quadrants WGAN.

two letters identify a 1° quadrangle within the 15° grid zone. In figure 20-27, WGAN refers to the quadrangle situated between 120° east longitude and 12° and 13°north latitude. Notice that the 1° quadrangle WGAN is further divided by 30-minute division lines, and then with tick marks into 5- and 1-minute subdivisions.

(2) GEOREF Coordinates:

(a) Any feature within a 1° quadrangle can be located by reading the number of minutes to the right and the number of minutes up. For example, the city of Magaran (figure 20-27) can be located by proceeding as follows:

-1. 15° X 15° quadrangle identification	WG
-2. 1° X 1° quadrangle identification	WGAN
-3. Minutes to the right	WGAN 56
-4. Minutes up	WGAN 5630
-5. Full GEOREF coordinate	WGAN 5630

(b) If a reference of greater accuracy than 1 minute is required, the 1-minute tick marks may be divided into decimal values (tens or hundreds). By doing this, it is possible to locate a point within one-tenth of a minute with four letters and six numbers and within one-hundredth of a minute by four letters and eight numbers.

(3) GEOREF Special References. Another real advantage of the GEOREF system is the simplicity with

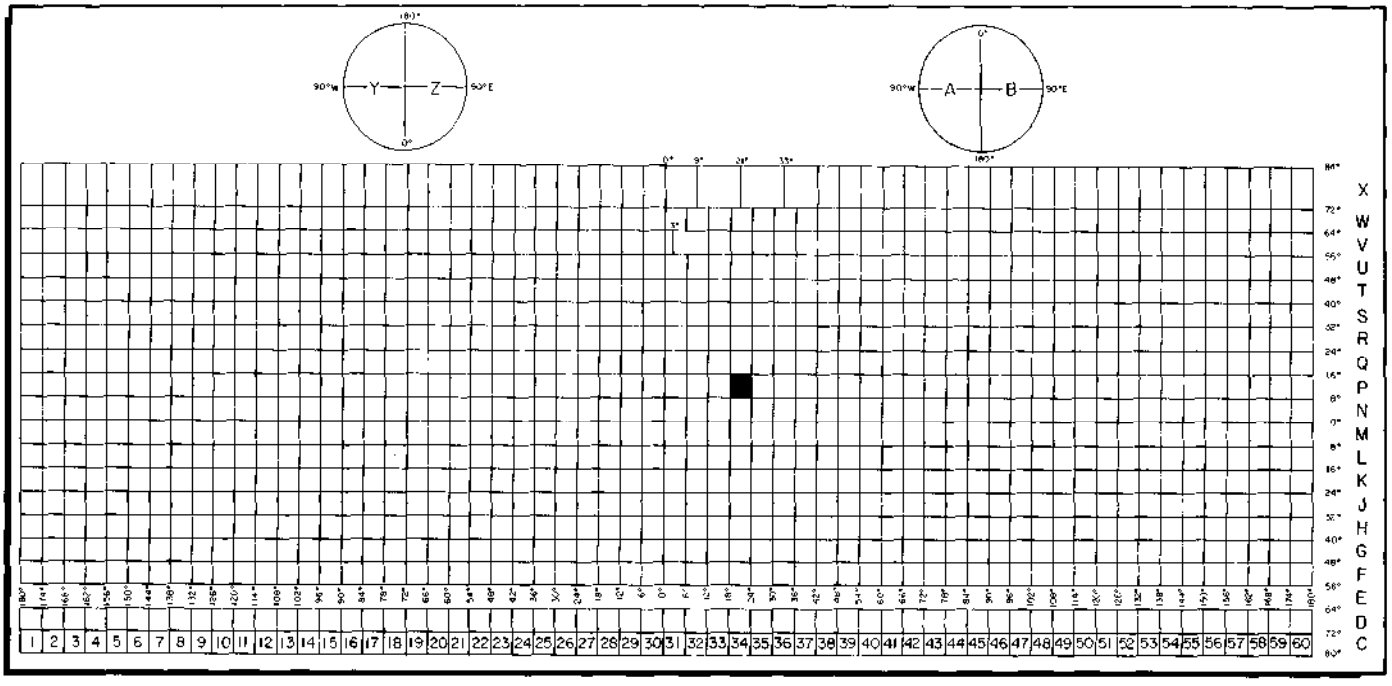

Figure 20-28. UPS Grid Zones.

which it allows a person to designate an area or indicate the elevation of a point. To designate the location and dimensions of a rectangular area, first read the GEOREF coordinates of the southwest corner of the area. Then add and "S," which denotes "side," and digits denoting the number of nautical miles that the area extends to the east. Then add an "X," denoting "by," and digits denoting the number of nautical miles the area extends to the north. An example of such a reference is WGAP2020S10X10 (figure 20-27). Circular areas are designated in much the same manner. First, read the GEOREF coordinates of the center of the area. Then add an "R," denoting radius, and digits defining the radius in nautical miles. This is also illustrated in figure 20-27 as WGAN4550R12.

(4) Military Grid Reference System. A grid is a rectangular coordinate system superimposed on a map. It consists of two sets of equally spaced parallel lines that are mutually perpendicular and form a pattern of squares. Some maps carry more than one grid. In such cases, each grid is shown in a different color or is otherwise distinguished. The military grid reference system is comprised of two grid systems. The US Army adopted the universal transverse mercator (UTM) grid for areas between 80° south latitude and 84°north latitude. For the polar caps, areas below 80° south latitude and above 84° north latitude, the universal polar stereographic (UPS) grid was adopted. The unit of measurement for the UTM and UPS grids are the meter, but the interval at which the grid lines are shown on the maps depends upon the scale.

(5) The UTM Grid System. In the UTM system, the surface of the Earth is divided into large quadrilateral grid zones (figure 20-28). Beginning at the 180th meridian, 6° columns are numbered 1 through 60 eastward with each column broken down into rows. From 80° south through 72° north, each row is 8° south-north.

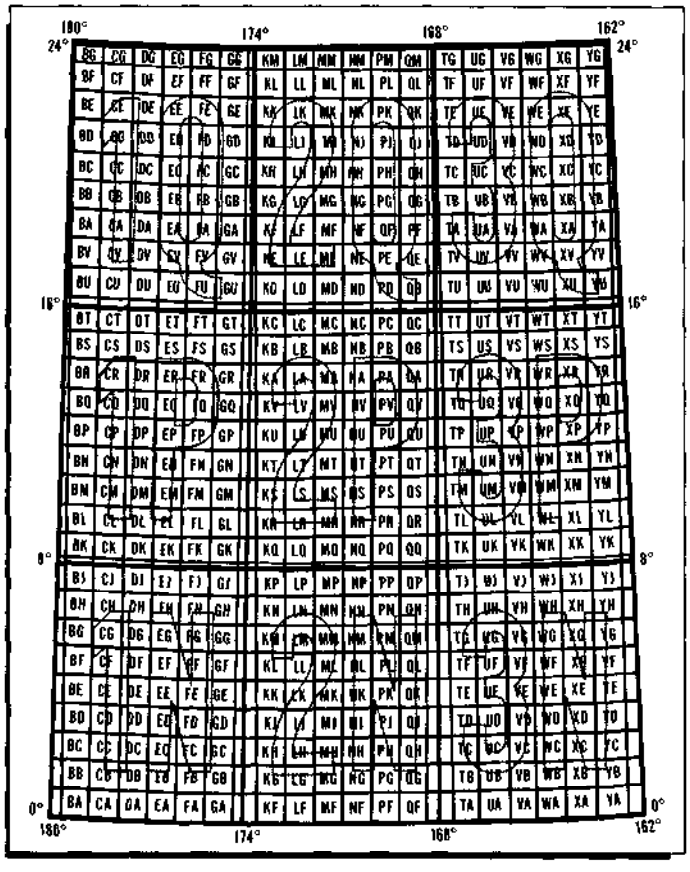

Figure 20-29. UTM Grid Zones.

The top row, 72° to 84° north is 12° south-north. The south-north rows are lettered "C" through "X" (omitting "I" and "O") as shown in figure 20-28. The grid zones are located by reading right and up. For instance, right to column 34 and up to "P" locates grid zone 34P, which is the shaded grid zone of figure 20-28. The UPS grid zones covering the polar areas are designated by a single "A," "B," "Y," or "Z" (figure 20-28).

(a) Each UTM grid zone is divided into columns and rows to form small squares measuring 100,000 meters on each side and are called 100,000-meter squares. Each square is identified with two letters. The columns are lettered "A" through "Z" (omitting "I" and "O"), starting at the 180th meridian and progressing eastward around. The 24 letters are repeated every 18° (figure 20-29). Starting at the Equator, the horizontal row 100,000-meter squares are lettered "A" through "V" (omitting "I" and "O") northward. From the Equator southward, the designation "V" through "A" is used. The letters are repeated periodically (figure 20-29).

(b) The Earth's meridians converge toward the poles. Therefore, the grid zones are not square or rectangular. The actual width of each grid zone decreases toward the poles. This condition causes partial squares to occur along the grid zones. In the far north and south latitudes, the grid zones become so narrow that 100,000-meter square designations may disappear completely. However, each full or partial 100,000-meter square within a grid zone is referenced with two letters. The first letter refers to the vertical column (left to right), and the second letter identifies the horizontal row (bottom to top). Thus, a grid zone designation plus two letters identifies or designates an area 100,000 meters on each side. Furthermore, as the UTM system is set up, no two squares with the same designation are included in a grid zone or on the same map sheet.

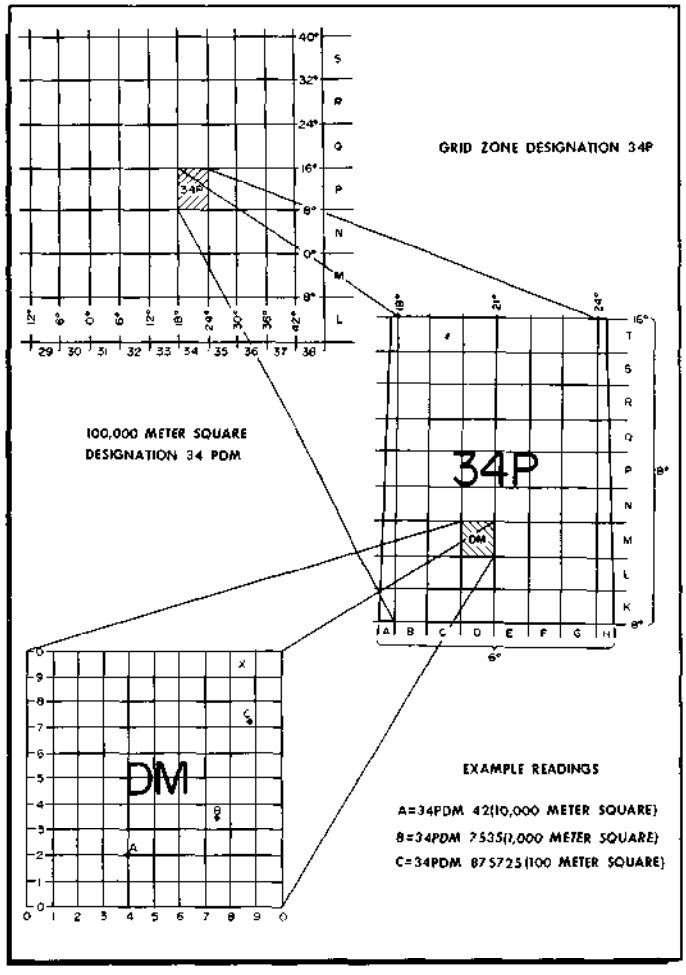

Figure 20-30. Plotting UTM Grid Coordinates.

(c) Observe grid zone 34P which is expanded in figure 20-30 to show the 100,000-meter squares. For grid zone 34P, the columns are designated "A" through "H," and the rows are designated "K" through "T" (omitting "O"). The left column begins with "A" because, as stated earlier, columns repeat the alphabet each 18°. The bottom row begins with "K" because "A" through "J" (omitting "I") was used up in the previous 8° of north latitude.

(d) Next, note the partial squares along the left and right sides of grid zone 34P (figure 20-30). Partial squares occur because the distance east and west of the central meridian of each grid zone does not contain an even number of 100,000-meter squares. The last squares, therefore, must terminate at the meridian junctions. In the north-south direction, partial 100,000-meter squares seldom occur.

(e) Figure 20-31 shows the grid reference box for a map or chart. Note the statement in the upper left corner of the grid reference box. It identifies the grid zones that are represented on the map sheet—52S and 53S. Thus, the full UTM coordinates of any point within the map area begins with either 52S or 53S.

(f) Still, it is not clear which area is 52S and which is 53S. Therefore, study the 100,000-meter square block identification located directly below the grid zone designation. From the diagram a person can see that everything to the left of the center meridian is grid zone 52S and everything to the right of the same meridian is 53S.

(g) Other important information given in the grid reference block includes: (1) sample reference and, (2) step-by-step procedures for locating or writing coordinates. Each time a new map is used, identify the sample point and write its UTM coordinates to ensure the grid breakdown for that map is understood.

(h) A troublesome and sometimes confusing situation exists where 100,000-meter squares fuse together along meridians separating grid zones. Remember, this happens every 6° around the world. Notice in figure 20-32 that the 100,000-meter squares GP and KJ are only partial squares, fusing along the center meridian (so are GQ, GN, KH, and KK). There are then full and

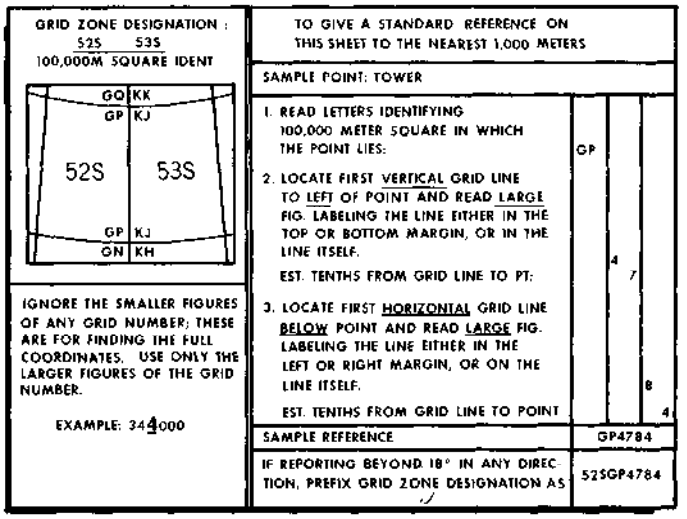

Figure 20-31. UTM Grid Reference Box.

partial 10,000 meter-squares within GP and KJ. Column 7 of the GP is comprised of partial 10,000-meter squares; columns 8 and 9 are missing because of the forced fusing along the meridian; similarly, column 2 of KJ is partial; columns O and I are missing. There is no problem in reading coordinates with full 10,000-meter squares. The tower in GP (sample point in figure 20-32) is 47 right and 84 up (omitting the grid zone and 100,000 meter square designation. All partial 100,000-meter squares are full sized in a north-south dimension). Therefore, distances up are referenced as full squares. However, partial squares, which occur in an east-west dimension, are something less than 10,000 meters long. Points within such partial squares are referenced as if the omitted part were present. That is, each partial square is imaginarily expanded into a full-sized square for reference purposes.

(i) The city of Bergen is 40 up; Celle is 35 up. Celle is three-tenths of the horizontal distance between grid line 7 and grid line 8—if there were a grid line 8. Thus, Celle is 73 right, and its full coordinates are 52SGP7335. Bergen is eight-tenths of the distance (reading left to right) between grid line 2—if there were a grid line 2—and grid line 3. Thus Bergen is 28 right, its full coordinates are 53SKJ2840.

(j) Figure 20-32 depicts a UTM grid breakdown as it normally appears on 1:250,000 scale charts. The smallest physical square is 10,000 meters on each side. However, larger scale maps with grid squares of 1,000 or even 100 meters on each side may be used. If so, add the values for the smaller grid squares. As additional digits are added, more precise points on the Earth's surface can be located.

(6) The UPS Grid System. The UPS reference system is the companion system to the UTM system. It covers the area of the world above 84° north latitude and below 80° south latitude. The UPS has a similar divisional breakdown and is read or written like the UTM system.

(a) Figure 20-33 shows the arbitrarily assigned designations for the UPS system in the North and South Pole regions. Note that from the small circles of the figure, the polar area is divided into two grid zone divisions by the 180° and 0° meridians. The west longitude half is designated "grid zone A or Y." Also notice that no numbers are used with the letters to identify the grid zones.

(b) The two grid zones "A" and "B" of the South Pole are divided into 100,000-meter squares, as shown in the large circle of figure 20-33. Each square is identified by a two-letter designation, which is assigned so no duplication exists between the two grid zones. The letters "I" and "O" are omitted, and to avoid confusion with 100,000-meter squares in adjoining UTM zones, the letters D, E, M, N, V, and W are also omitted.

(c) The UPS system is also read right and then up. Thus, the shaded 100,000-meter square at 10 o'clock in figure 20-33 is identified as AQR. (Remember, no numbers are used in identifying the grid zone.) The shaded 100,000-meter square near the South Pole of the same drawing is identified as BBM.

(d) The UPS breakdown of the North Pole region is similar to the South Pole region. Conversion of figure 20-33 to fit the North Pole would require the following changes: Substitute grid zone letters "Y" for "A" and "Z" for "B," and interchange the 0° and 180° positions on the common parallel at 80° South latitude. Designation of the 100,000-meter squares for the North Pole region is shown in figure 20-33.

(e) If the map scale is sufficiently large, the 100,000-meter squares can be subdivided into smaller squares of 10,000 meters on each side. Then, the 10,000-meter squares can be divided into 1,000-or even 100-meter squares. However, there is rarely a requirement in the polar regions for such a large-scale chart.

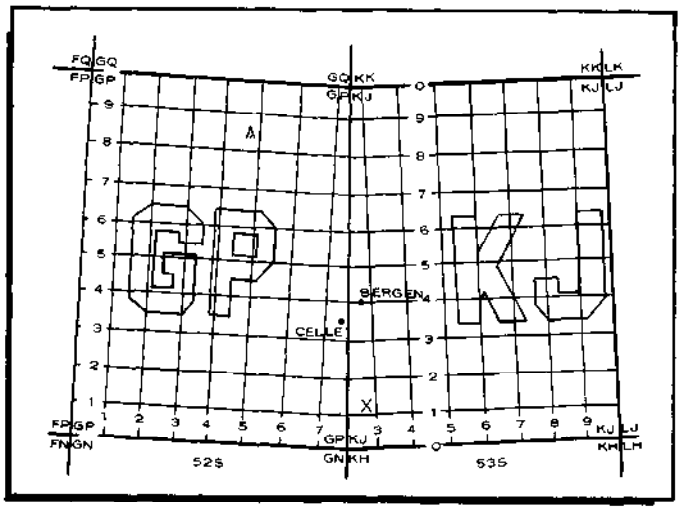

Figure 20-32. Fusion of Grid Zones 52S and 53S.

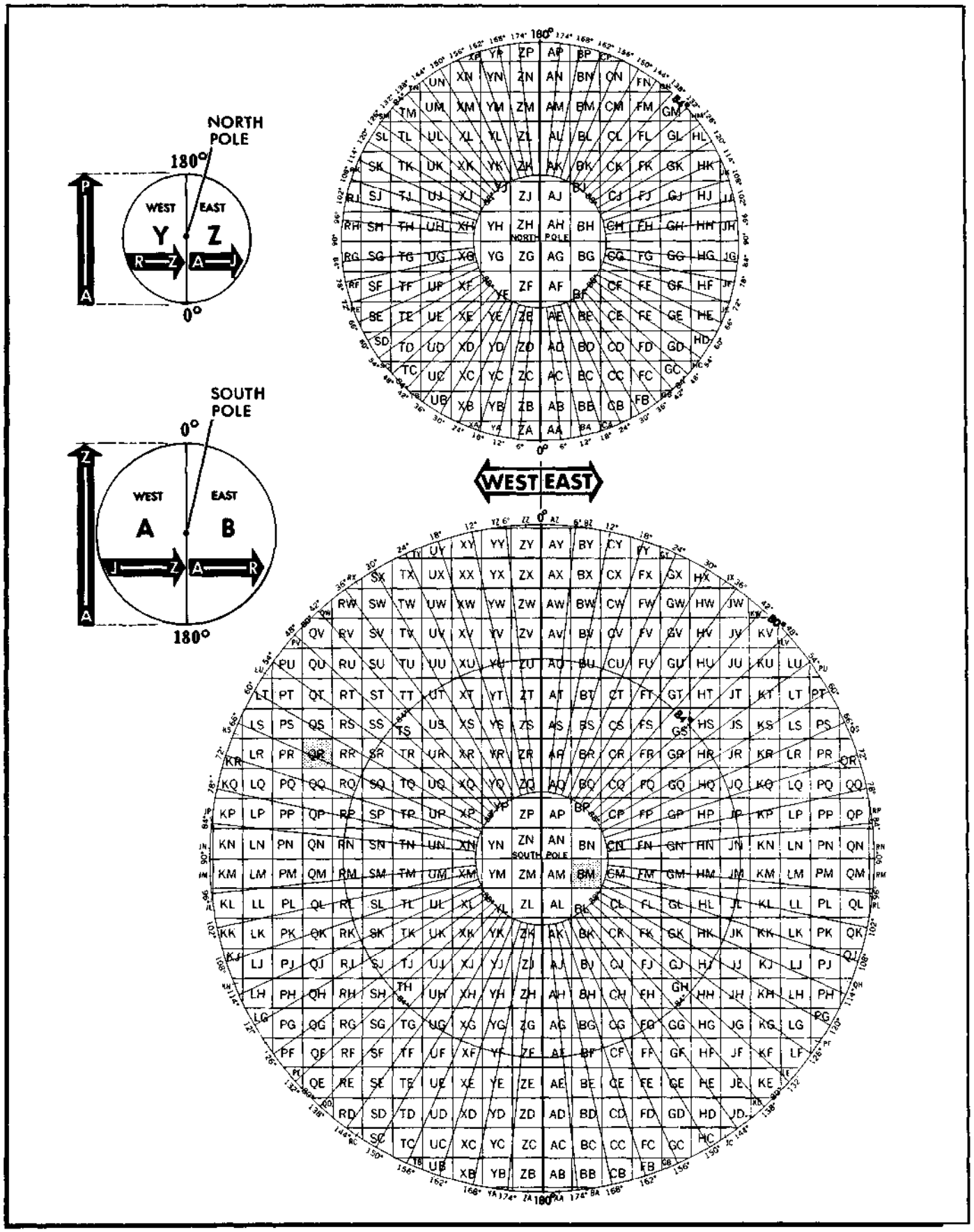

Figure 20-33. UPS Grid Zones.

Generally, a person can expect to work with small-scale charts with the grid broken down no further than 100,000-meter squares.

(7) Public Land Survey. In the western part of the United States or in areas which were not settled before the Federal Government was formed, all land is laid out in rectangular survey as established by the Government. This public land survey is based on all land being divided in relation to true north. Public land surveys all originate from six or seven initial points which are exact locations of even latitude and longitude lines which have been established astronomically.

(a) From any one of the initial points a true north-south line, referred to as the principal meridian, is established. From the same point a true east-west line, referred to as the baseline, is established. Along this principal meridian and baseline are laid out 6-mile squares or townships. Each of these townships are numbered in relation to the initial point of survey. To the east and west of the initial point, the townships are designated by range numbers; to the north and south of the initial point, the townships are designated by township numbers. Therefore, township 2 north, range 3 east, would lie between 6 and 12 miles north of the initial point and between 12 and 18 miles east of the initial point.

(b) Each township contains 36 square miles and is divided into 36 sections. A section is 1 square mile or 640 acres. The section layout on townships is the same throughout the Public Land Survey. The sections are numbered in rows back and forth beginning in the upper right-hand corner of the township and ending in the lower right-hand corner (figure 20-34).

(c) Each section is divided into quarters or quarter sections of 160 acres each. These quarter sections are named by the compass location in relation to the section. The upper right-hand quarter section is referred to as the northeast ¼, the lower right-hand quarter is the southeast ¼, the lower left-hand quarter is the southwest ¼, and the upper left-hand quarter is the northwest ¼.

(d) Each quarter section is further subdivided into quarters or four blocks of 40 acres each known as forties. The forties are also located by the compass directions. In locating a particular piece of property, the 40-compass quadrant is given first, followed by the quarter section quadrant. Thus the SW-SE means the southwest 40 of the southeast quarter section. This is the basic unit of land management and, therefore, one should become familiar with the Public Land Survey and the means of locating specific pieces of property.

20-7. Elevation and Relief. A knowledge of map symbols, grids, scale, and distance gives enough information to identify two points, locate them, measure between them, and determine how long it would take to travel between them. But what happens if there is an obstacle between the two points? The map user must become proficient in recognizing various landforms and irregularities of the Earth's surface and be able to determine the elevation and differences in height of all terrain features.

a. Datum Plane. This is the reference used for vertical measurements. The datum plane for most maps is mean or average sea level.

b. Elevation. This is defined as the height (vertical distance) of an object above or below a datum plane.

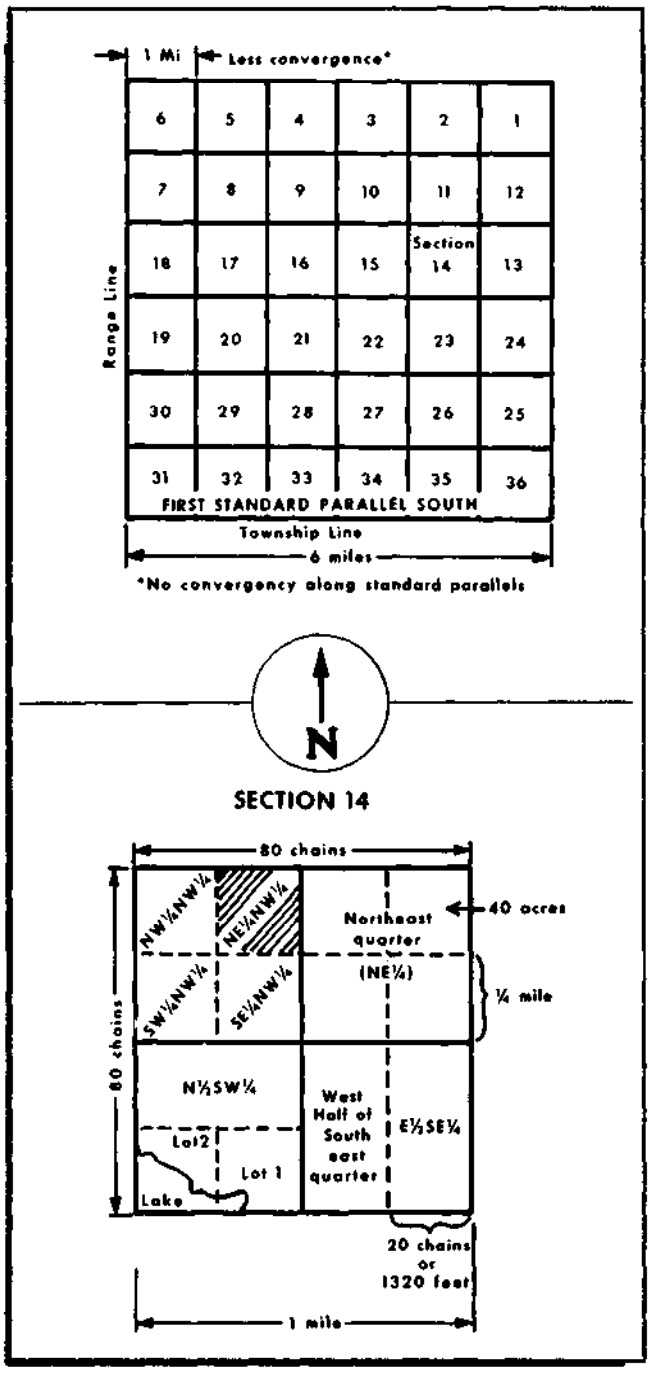

Figure 20-34. Section 14.

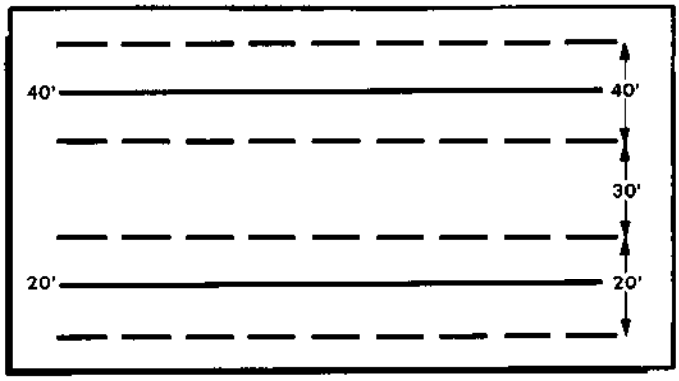

Figure 20-35. Estimating Elevation and Contour Lines

c. Relief. Relief is the representation of the shape and height of landforms and characteristic of the Earth's surface.

d. Contour Lines:

(1) There are several ways of indicating elevation and relief on maps. The most common way is by contour lines. A contour line is an imaginary line connecting points of equal elevation. Contour lines indicate a vertical distance above or below a datum plane. Starting at sea level, each contour line represents an elevation above sea level. The vertical distance between adjacent contour lines is known as the contour interval. The amount of contour interval is given in the marginal information. On most maps, the contour lines are printed in brown. Starting at zero elevation, every fifth contour line is drawn with a heavier line. These are known as index contours and somewhere along each index contour, the line is broken and its elevation is given. The contour lines falling between index contours are called intermediate contours. They are drawn with a finer line than the index contours and usually do not have their elevations given.

(2) Using the contour lines on a map, the elevation of any point may be determined by:

(a) Finding the contour interval of the map from the marginal information, and noting the amount and unit of measure.

(b) Finding the numbered contour line (or other given elevation) nearest the point for which elevation is being sought.

(c) Determining the direction of slope from the numbered contour line to the desired point.

(d) Counting the number of contour lines that must be crossed to go from the numbered line to the desired point and noting the direction—up or down. The number of lines crossed multiplied by the contour interval is the distance above or below the starting value. If the desired point is on a contour line, its elevation is that of the contour; for a point between contours,

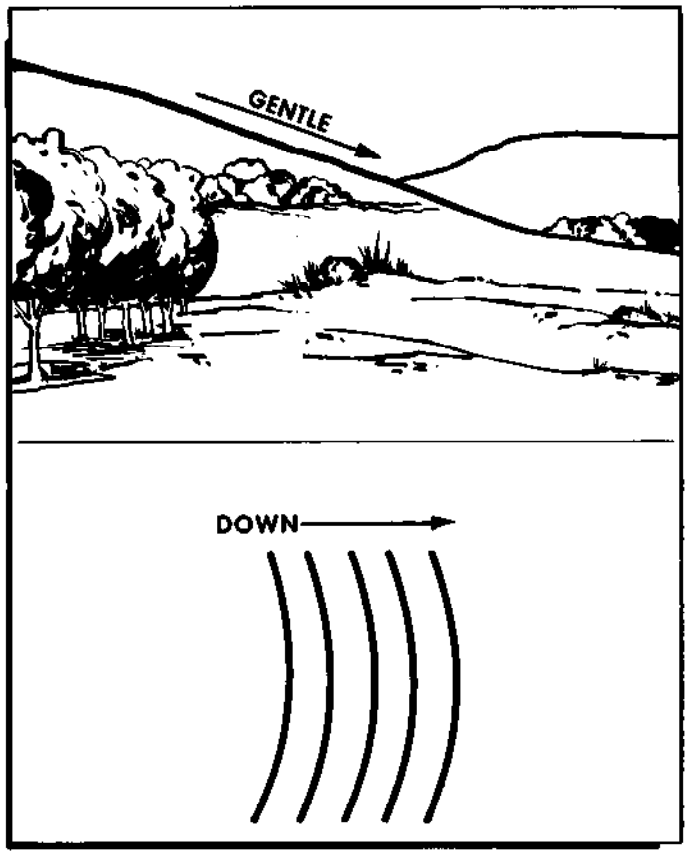

Figure 20-36. Uniform Gentle Slope.

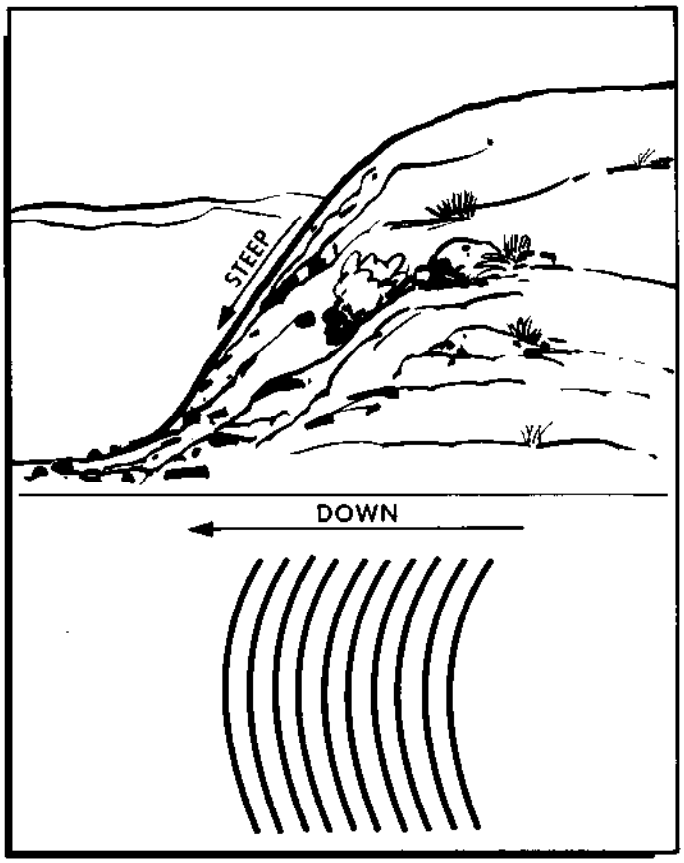

Figure 20-37. Uniform Steep Slope.

most military needs are satisfied by estimating the elevation to an accuracy of one-half the contour interval. All points less than one-fourth the distance between the lines are considered to be at the same elevation as the line. All points one-fourth to three-fourths the distance from the lower line are considered to be at an elevation one-half the contour interval above the lower line (figure 20-35).

(e) To estimate the elevation of the top of an unmarked hill, add half the contour interval to the elevation of the highest contour line around the hill. To estimate the elevation of the bottom of a depression, subtract half the contour interval from the value of the lowest contour around the depression.

(f) On maps where the index and intermediate contour lines do not show the elevation and relief in as much detail as may be needed, supplementary contour may be used. These contour lines are dashed brown lines, usually at one-half the contour interval for the map. A note in the marginal information indicates the interval used. They are used exactly as are the solid contour lines.

(g) On some maps contour lines may not meet the standards of accuracy but are sufficiently accurate in both value and interval to be shown as contour rather than as form lines. In such cases, the contours are considered as approximate and are shown with a dashed symbol; elevation values are given at intervals along the heavier (index contour) dashed lines. The contour note in the map margin identifies them as approximate contours.

(h) In addition to the contour lines, bench marks and spot elevations are used to indicate points of known elevation on the map. Bench marks, the more accurate of the two, are symbolized by a black X, as X BM 124. The elevation value shown in black refers to the center of the X. Spot elevations shown in brown generally are located at road junctions, on hilltops, and other promiment landforms. The symbol designates an accurate horizontal control point. When a bench mark and a horizontal control point are located at the same point, the symbol BM is used.

(i) The spacing of the contour lines indicates the nature of the slope. Contour lines evenly spaced and wide apart indicate a uniform, gentle slope (figure 20-36). Contour lines evenly spaced and close together indicate a uniform, steep slope. The closer the contour lines to each other, the steeper the slope (figure 20-37). Contour lines closely spaced at the top and widely spaced at the bottom indicate a concave slope (figure 20-38). Contour lines widely spaced at the top and closely spaced at the bottom indicate a convex slope (figure 20-39).

(j) To show the relationship of land formations to each other and how they are symbolized on a contour map, stylized panoramic sketches of the major relief

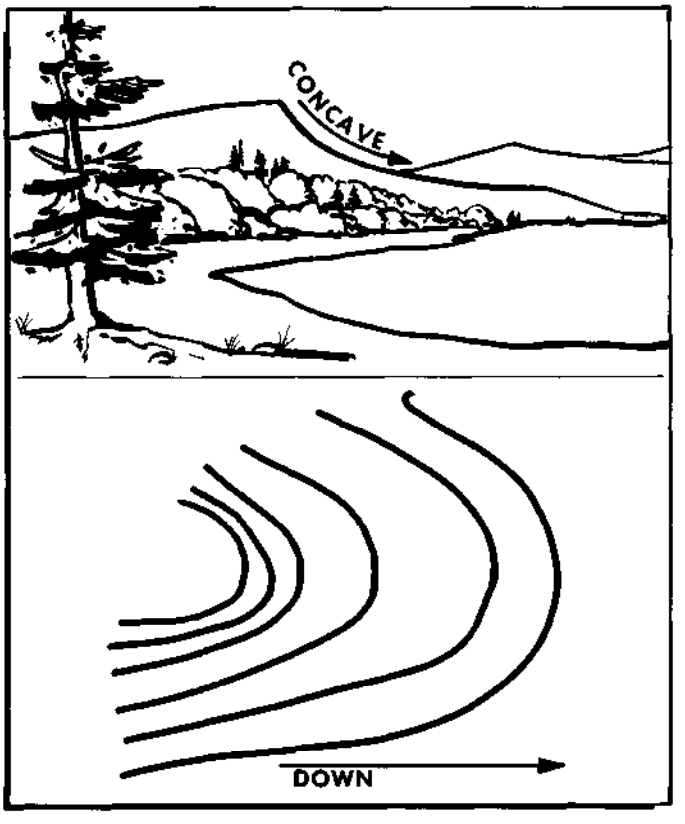

Figure 20-38. Concave Slope.

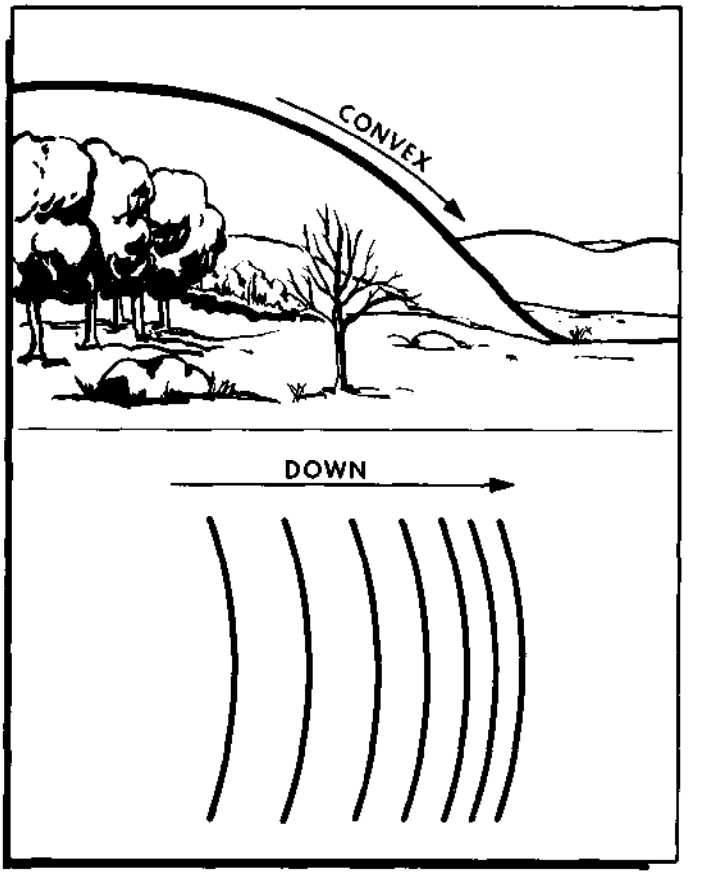

Figure 20-39. Convex Slope.

formations were drawn and a contour map of each sketch developed. Each figure (figure 20-40 through 20-46) shows a sketch and a map with a different relief feature and its characteristic contour pattern.

(3) Hill. A point or small area of high ground (figure 20-40). When one is located on a hilltop, the ground slopes down in all directions.

(4) Valley. Usually a stream course which has at least a limited extent of reasonably level ground bordered on the sides by higher ground (figure 20-41A). The valley generally has maneuvering room within its confines. Contours indicating a valley are U-shaped and tend to parallel a major stream before crossing it. The more gradual the fall of a stream, the farther each contour inner part. The curve of the contour crossing always points upstream.

(5) Drainage. A less-developed stream course in which there is essentially no level ground and, therefore, little or no maneuvering room within its confines (figure 20-41B). The ground slopes upward on each side and toward the head of the drainage. Drainages occur frequently along the sides of ridges, at right angles to the valleys between the ridges. Contours indicating a drainage are V-shaped, with the point of the "V" toward the head of the drainage.

(6) Ridge. A range of hills or mountains with normally minor variations along its crest (figure 20-42A). The ridge is not simply a line of hills; all points of the ridge crest are appreciably higher than the ground on both sides of the ridge.

(7) Finger Ridge. A ridge or line of elevation projecting from or subordinate to the main body of a mountain or mountain range (figure 20-42B). A finger ridge is often formed by two roughly parallel streams cutting drainages down the side of a ridge.

(8) Saddle. A dip or low point along the crest of a ridge. A saddle is not necessarily the lower ground between two hilltops; it may simply be a dip or break along an otherwise level ridge crest (figure 20-43).

(9) Depression. A low point or sinkhole surrounded on all sides by higher ground (figure 20-44).

(10) Cuts and Fills. Manmade features by which the bed of a road or railroad is graded or leveled off by

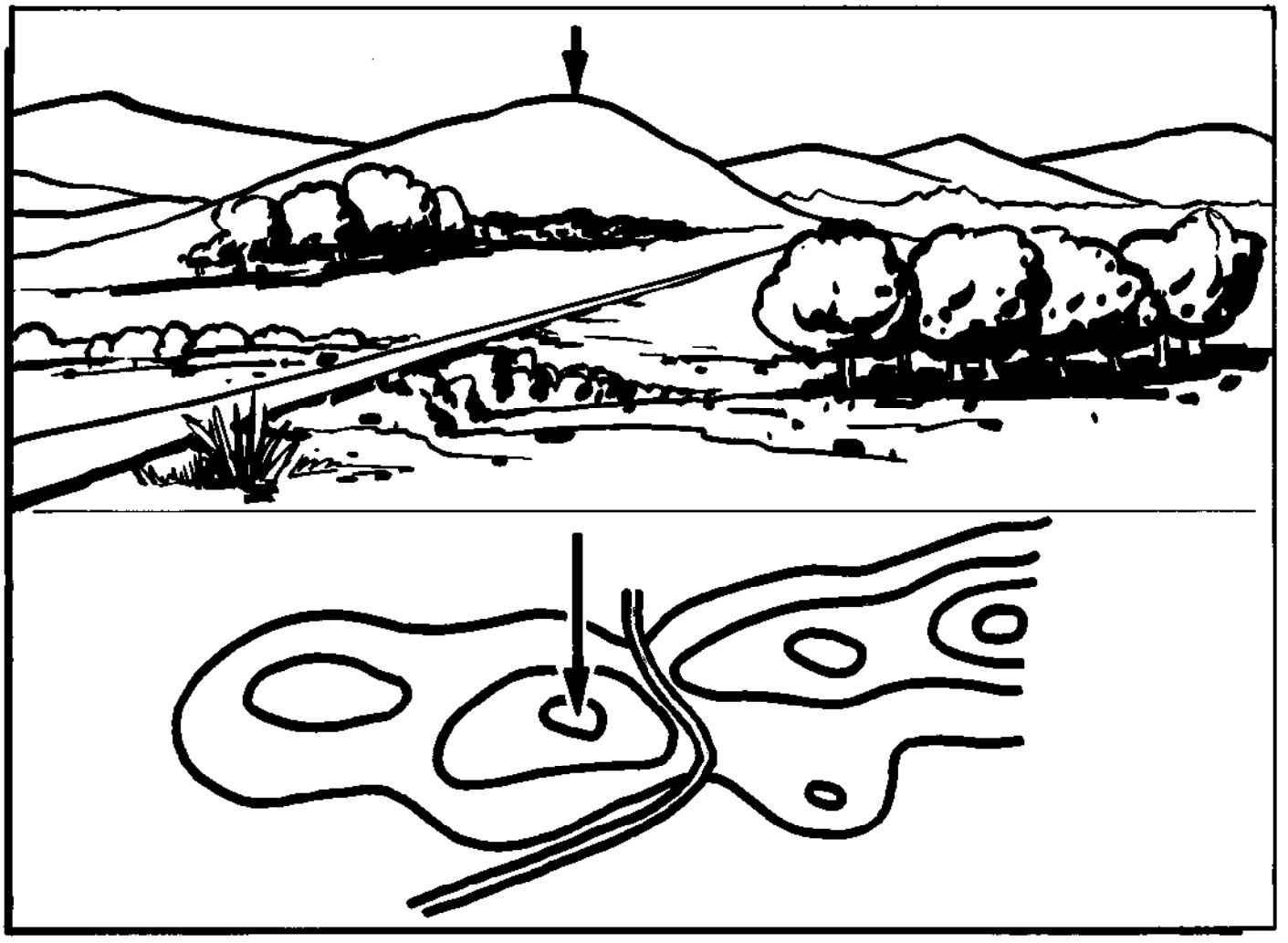

Figure 20-40. Hill.

cutting through high areas (figure 20-45A) and filling in low areas (figure 20-45B) along the right-of-way.

(11) Cliff. A vertical or near vertical slope (figure 20-46). When a slope is so steep that it cannot be shown at the contour interval without the contours fusing, it is shown by a ticked "carrying" contour(s). The ticks always point toward lower ground.

20-8. Representative Fraction (RF):

a. The numerical scale of a map expresses the ratio of horizontal distance on the map to the corresponding horizontal distance on the ground. It usually is written as a fraction, called the representative fraction (RF). The representative fraction is always written with the map distance as one (1). It is independent of any unit of measure. An RF of 1/50,000 or 1:50,000 means that one (1) unit of measure on the map is equal to 50,000 of the same units of measure on the ground.

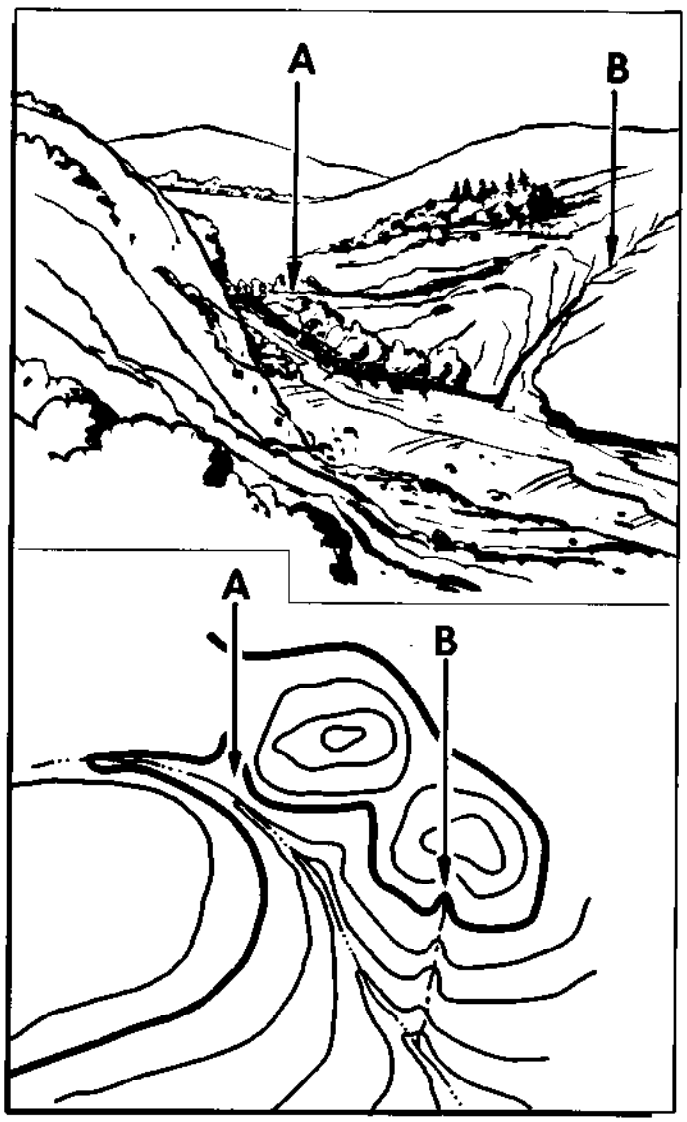

Figure 20-41. (A) Valley (B) Drainage.

b. The ground distance between two points is determined by measuring between the points on the map and multiplying the map measurement by the denominator of the RF.

Example: RF = 1:50,000 or $\frac{1}{50,000}$

Map distance = 5 units (CM)

5 X 50,000 - 250,000 units (CM) of ground distance (figure 20-47).

c. When determining ground distance from a map, the scale of the map affects the accuracy. As the scale becomes smaller, the accuracy of measurement decreases because some of the features on the map must be exaggerated so that they may be readily identified.

20-9. Graphic (Bar) Scales:

a. On most military maps, there is another method of determining ground distance. It is by means of the graphic (bar) scales. A graphic scale is a ruler printed on the map on which distances on the map may be measured as actual ground distances. To the right of the zero (0), the scale is marked in full units of measure and is called the primary scale. The part to the left of zero (0) is divided into tenths of a unit and is called the extension scale. Most maps have three or more graphic scales, each of which measures distance in a different unit of measure (figure 20-48).

b. To determine a straight-line ground distance between two points on a map, lay a straight-edged piece of paper on the map so that the edge of the paper touches both points. Mark the edge of the paper at each point. Move the paper down to the graphic scale and read the ground distance between the points. Be sure to use the scale that measures in the unit of measure desired (figure 20-49).

c. To measure distance along a winding road, stream, or any other curved line, the straightedge of a piece of paper is used again. Mark one end of the paper and place it at the point from which the curved line is to be measured. Align the edge of the paper along a straight portion and mark both the map and the paper at the end of the aligned portion. Keeping both marks together, place the point of the pencil on the mark on the paper to hold it in place. Pivot the paper until another approximately straight portion is aligned and again mark on the map and the paper. Continue in this manner until measurement is complete. Then place the paper on the graphic scale and read the ground distance (figure 20-50).

d. Often, marginal notes give the road distance from the edge of the map to a town, highway, or junction of the map. If the road distance is desired from a point on the map to such a point off the map, measure the distance to the edge of the map and add the distance speci-

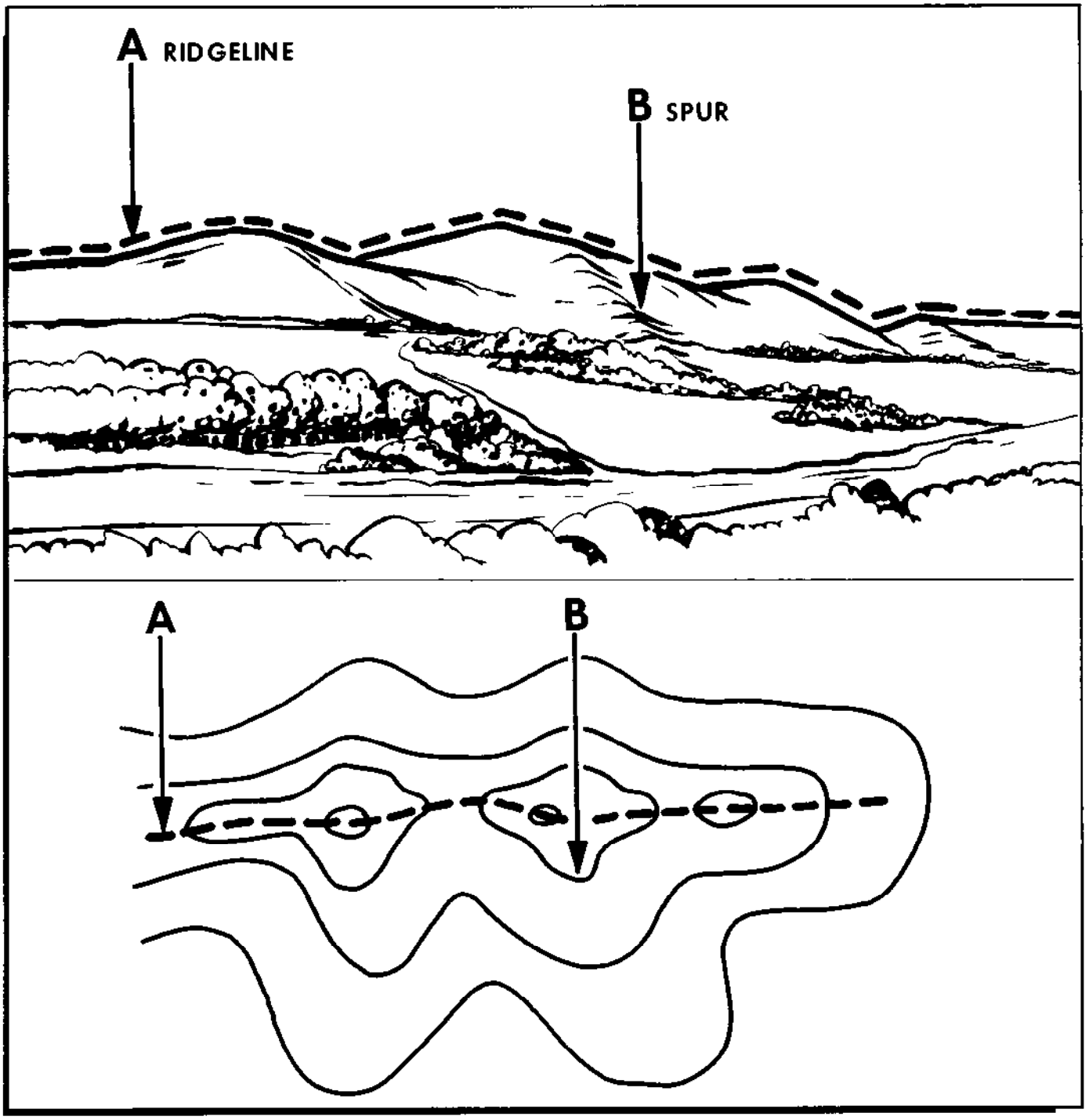

Figure 20-42. (a) Ridge Line (b) Finger Ridge.

fied in the marginal note to that measurement. Be sure the unit of measure is the same (figure 20-51).

20-10. Using a Map and Compass, and Expressing Direction:

a. To use a map, the map must correspond to the lay of the land, and the user must have a knowledge of direction and how the map relates to the cardinal directions. In essence, to use a map for land navigation, the map must be "oriented" to the lay of the land. This is usually done with a compass. On most maps, either a declination diagram, compass rose, and lines of map magnetic variations are provided to inform the user of the difference between magnetic north and true north.

b. Directions are expressed in everyday life as right, left, straight ahead, etc.; but the question arises, "to the right of what?" Military personnel require a method of expressing direction which is accurate, adaptable for use

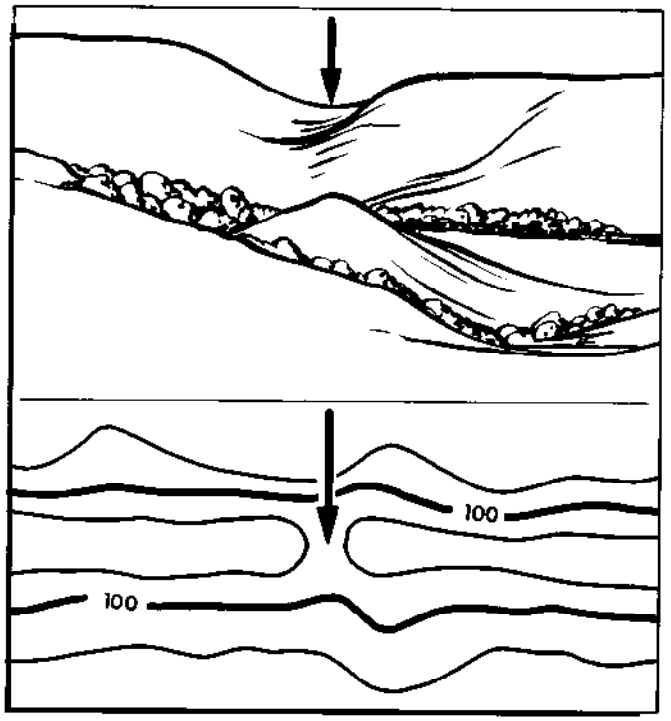

Figure 20-43. Saddle.

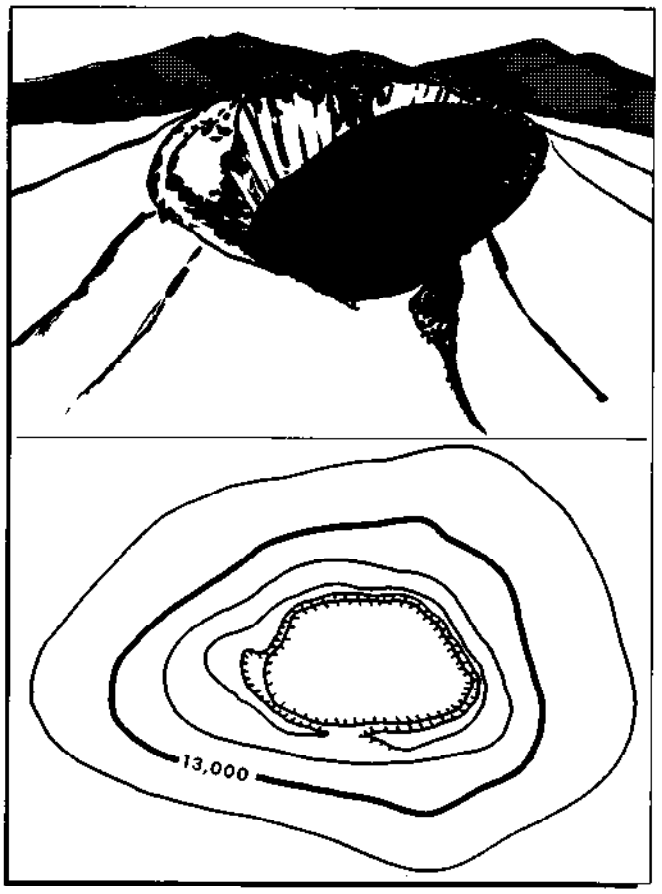

Figure 20-44. Depression.

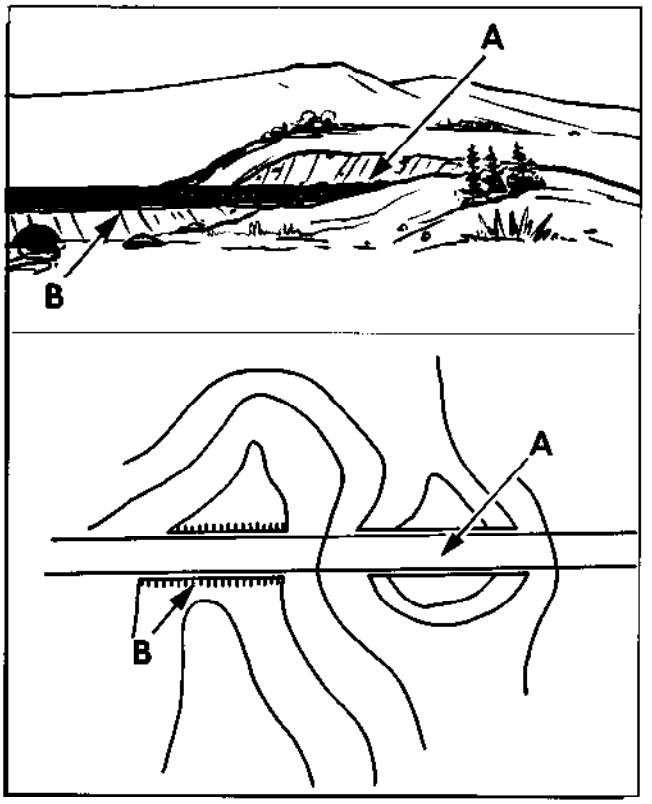

Figure 20-45. (A) Cut (B) Fill.

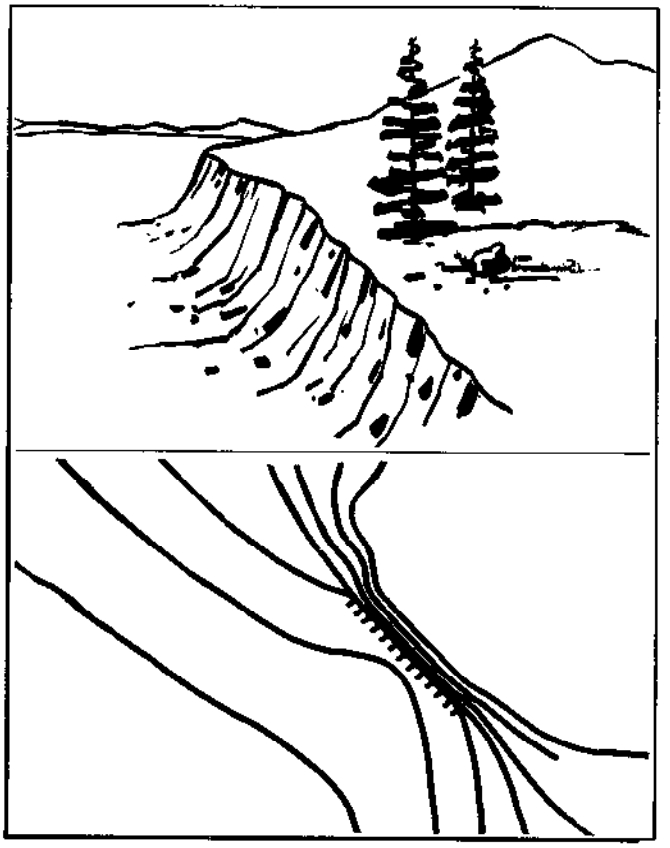

Figure 20-46. Cliff.

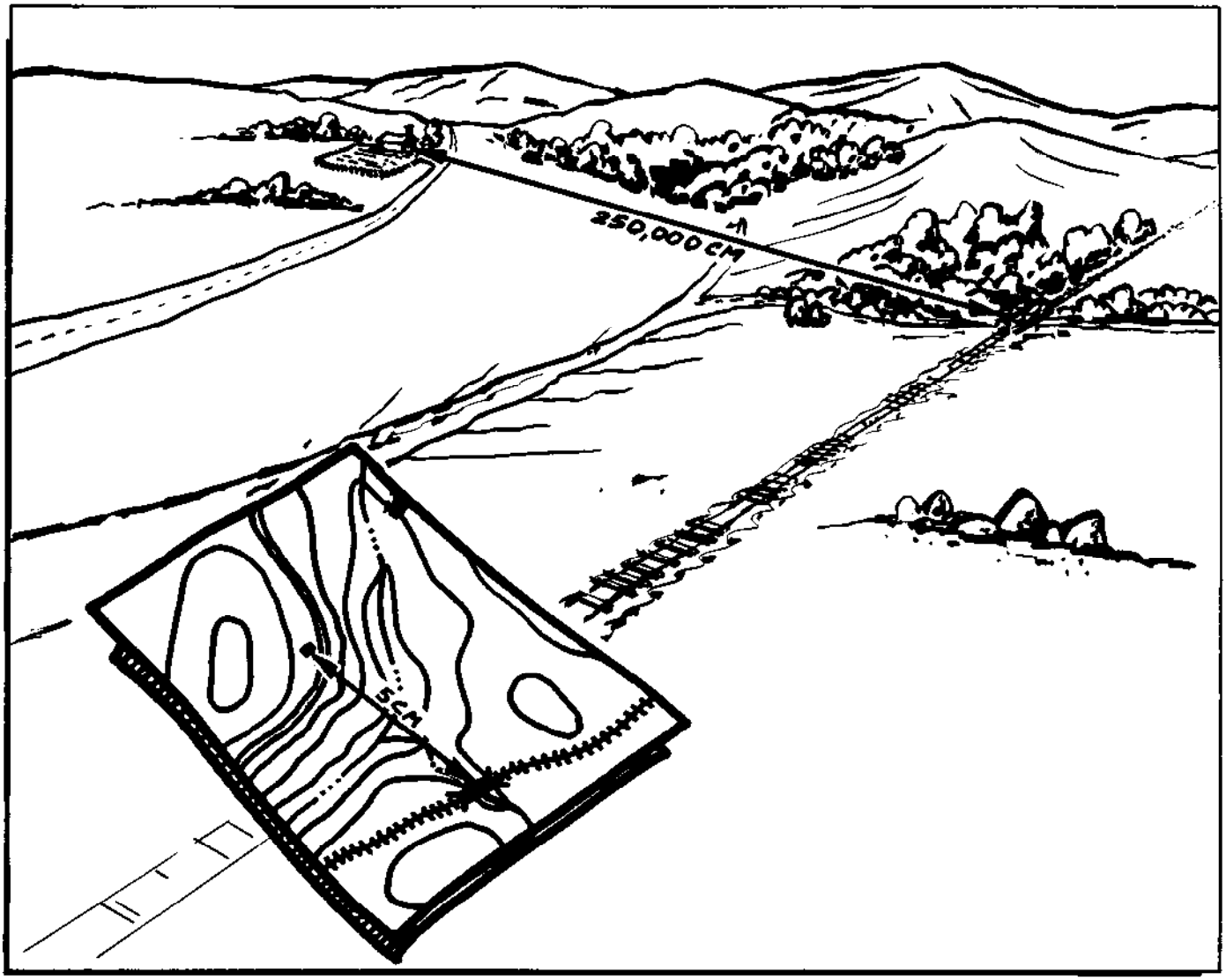

Figure 20-47. Ground Distance.

in any area of the world, and has a common unit of measure. Directions are expressed as units of angular measure. The most commonly used unit of angular measure is the degree with its subdivisions of minutes and seconds.

(1) Baselines. To measure anything, there must always be a starting point or zero measurement. To express a direction as a unit of angular measure, there must be a starting point or zero measure and a point of reference. These two points designate the base or refer-

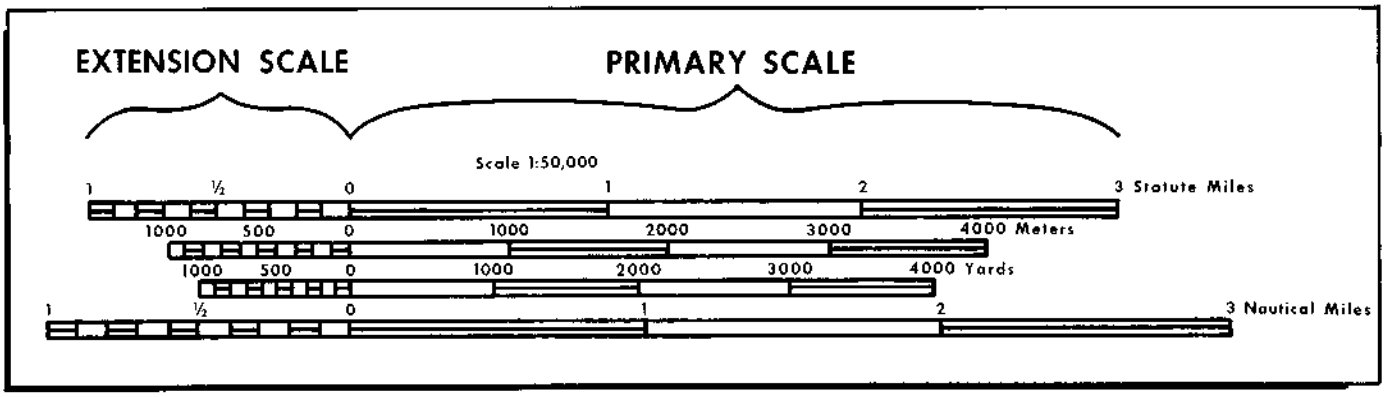

Figure 20-48. Graphic Bar Scale.

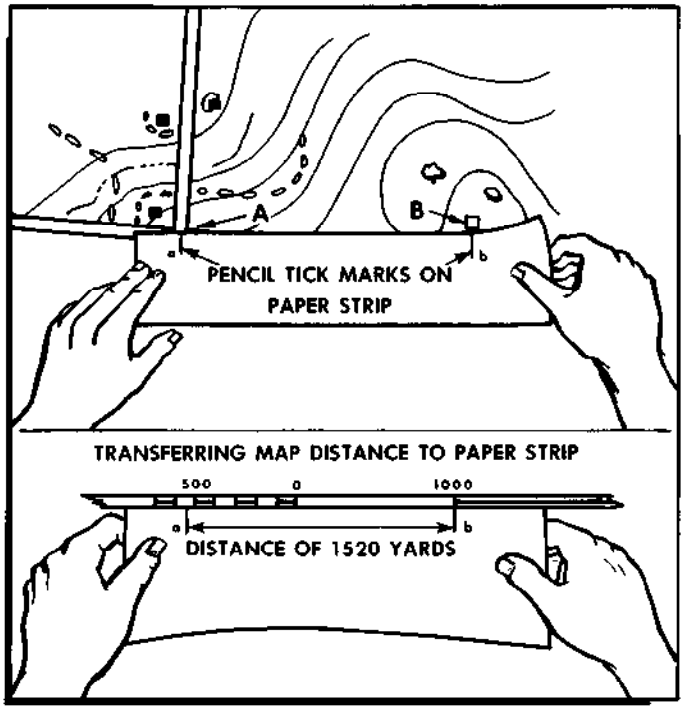

Figure 20-49. Measuring Straight Line Map Distances.

ence line. There are three baselines—true north, magnetic north, and grid north. Those most commonly used are magnetic and grid—the magnetic when working with a compass, and the grid when working with a military map.

(a) True north—a line from any position on the Earth's surface to the North Pole. All lines of longitude are true north lines. True north is usually symbolized by a star (figure 20-52).

(b) Magnetic north—the direction to the north magnetic pole, as indicated by the north-seeking needle of a magnetic instrument. Magnetic north is usually symbolized by a half arrowhead (figure 20-52).

(c) Grid north—the north established by the vertical grid lines on the map. Grid north may be symbolized by the letters GN or the letter Y.

(2) Azimuth and Back Azimuth:

(a) The most common method used by the military for expressing a direction is azimuths. An azimuth is defined as a horizontal angle, measured in a clockwise manner from a north baseline. When the azimuth between two points on a map is desired, the points are joined by a straight line and a protractor is used to measure the angle between north and the drawn line. This measured angle is the azimuth of the drawn line (figure 20-53). When using an azimuth, the point from which the azimuth originates is imagined to be the center of the azimuth circle (figure 20-54). Azimuths take their name from the baseline from which they are measured; true azimuths from true north, magnetic azimuths from magnetic north, and grid azimuths from grid north (figure 20-52). Therefore, any given direction can be expressed in three different ways: a grid azimuth if measured on a military map, a magnetic azimuth if measured by a compass, or a true azimuth if measured from a meridian of longitude.

(b) A back azimuth is the reverse direction of an azimuth. It is comparable to doing an "about face." To obtain a back azimuth from an azimuth, add 180° if the azimuth is 180° or less, or subtract 180° if the azimuth is 180° or more (figure 20-55). The back azimuth of 180° may be stated as either 000° or 360°.

(3) Declination Diagram. A declination diagram is placed on most large-scale maps to enable the user to properly orient the map. The diagram shows the interrelationship of magnetic north, grid north, and true north (figure 20-56). On medium-scale maps, declination information is shown by a note in the map margin.

(a) Declination is the angular difference between true north and magnetic or grid north. There are two declinations, a magnetic declination (figure 20-57) and a grid declination.

(b) Grid-Magnetic (G-M) Angle is an arc indicated by a dashed line, which connects the grid north and the magnetic north prongs. The value of this arc (G-M ANGLE) states the size of the angle between grid north and magnetic north and the year it was prepared. This value is expressed to the nearest ½°, with mil equivalents shown to the nearest 10 mils.

(c) Grid Convergence is an arc, indicated by a dashed line, which connects the prongs for true north and grid north. The value of the angle for the center of the sheet is given to the nearest full minute with its equivalent to the nearest mil. These data are shown in the form of a grid convergence note.

(d) Conversion notes may also appear with the diagram explaining the use of the G-M angle. One note

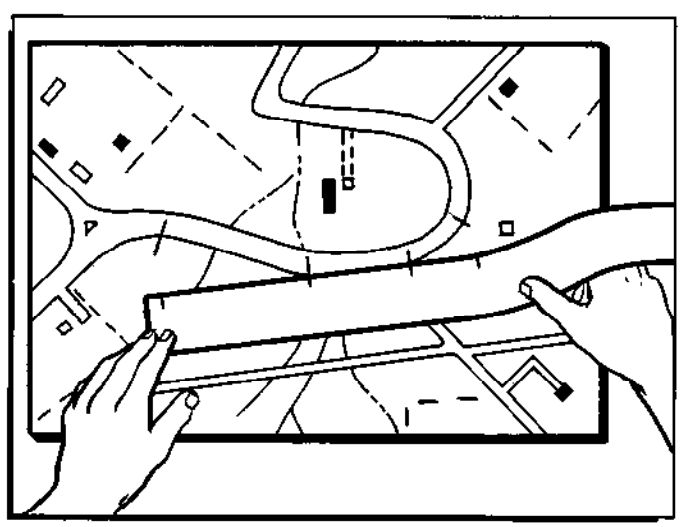

Figure 20-50. Measuring Curved Line Distances.

B–5. Conversion Factors

One	*Inches*	*Feet*	*Yards*	*Statute miles*	*Nautical miles*	*Millimeters*
Inch	1	0.0833	0.0277	.....	.....	25.40
Foot	12	1	0.333	.....	.....	304.8
Yard	36	3	1	0.00056	.....	914.4
Statute Mile	63,360	5,280	1,760	1	0.8684	.....
Nautical Mile	72,963	6,080	2,026	1.1516	1	.....
Millimeter	0.0394	0.0033	0.0011	.....	.....	1
Centimeter	0.3937	0.0328	0.0109	.....	.....	10
Decimeter	3.937	0.328	0.1093	.....	.....	100
Meter	39.37	3.2808	1.0936	0.0006	0.0005	1,000
Decameter	393.7	32.81	10.94	0.0062	0.0054	10,000
Hectometer	3,937	328.1	109.4	0.0621	0.0539	100,000
Kilometer	39,370	3,281	1,094	0.6214	0.5396	1,000,000
Myriameter	393,700	32,808	10,936	6.2137	5.3959	10,000,000

One	*cm*	*dm*	*m*	*dkm*	*hm*	*km*	*mym*
Inch	2.540	0.2540	0.0254	0.0025	0.0003	.....	.....
Foot	30.48	3.048	0.3048	0.0305	0.0030	0.0003	.....
Yard	91.44	9.144	0.9144	0.0914	0.0091	0.0009	.....
Statute Mile	160,930	16,093	1,609	160.9	16.09	1.6093	0.1609
Nautical Mile	185,325	18,532	1,853	185.3	18.53	1.8532	0.1853
Millimeter	0.1	0.01	0.001	0.0001	.....	.....	.....
Centimeter	1	0.1	0.01	0.001	0.0001	.....	.....
Decimeter	10	1	0.1	0.01	0.001	0.0001	.....
Meter	100	10	1	0.1	0.01	0.001	0.0001
Decameter	1,000	100	10	1	0.1	0.01	0.001
Hectometer	10,000	1,000	100	10	1	0.1	0.01
Kilometer	100,000	10,000	1,000	100	10	1	0.1
Myriameter	1,000,000	100,000	10,000	1,000	100	10	1

Example I

Problem: Reduce 76 centimeters to (?) inches

$$76 \text{ cm} \times 0.3937 = 29 \text{ inches.}$$

Answer: There are 29 inches in 76 centimeters.

Example II

Problem: How many feet are there in 2.74 meters?

$$\frac{2.74}{.3048} = 9 \text{ feet}$$

Answer: There are approximately 9 feet in 2.74 meters.

Figure 20-51. Conversion Factors.

provides instructions for converting magnetic azimuth to grid azimuth; the other note provides for converting grid azimuth to magnetic azimuth. The conversion (add or subtract) is governed by the direction of the magnetic north prong relative to the grid north prong.

(e) The grid north prong is aligned with the easting grid lines on the map, and on most maps is formed by an extension of an easting grid line into the margin. The angles between the prongs are seldom plotted exactly. The relative position of the directions is obtained from the diagram, but the numerical value should not be measured from it. For example, if the amount of declination from grid north to magnetic north is 1°, the arc shown in the diagram may be exaggerated if mea-

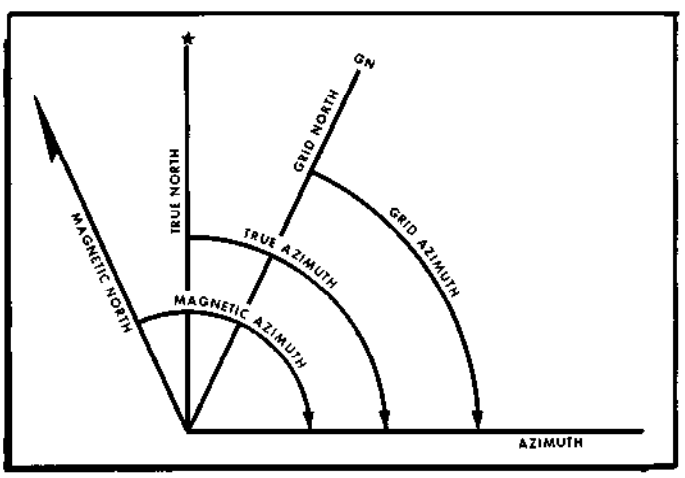

Figure 20-52. True, Grid and Magnetic Azimuths.

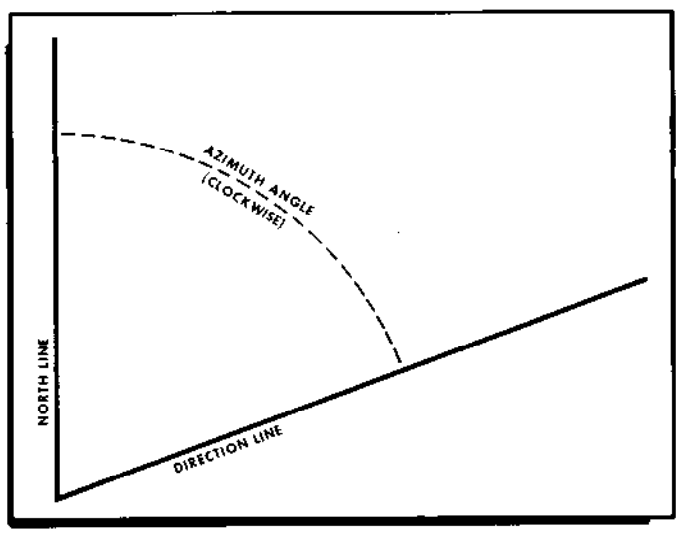

Figure 20-53. Azimuth Angle.

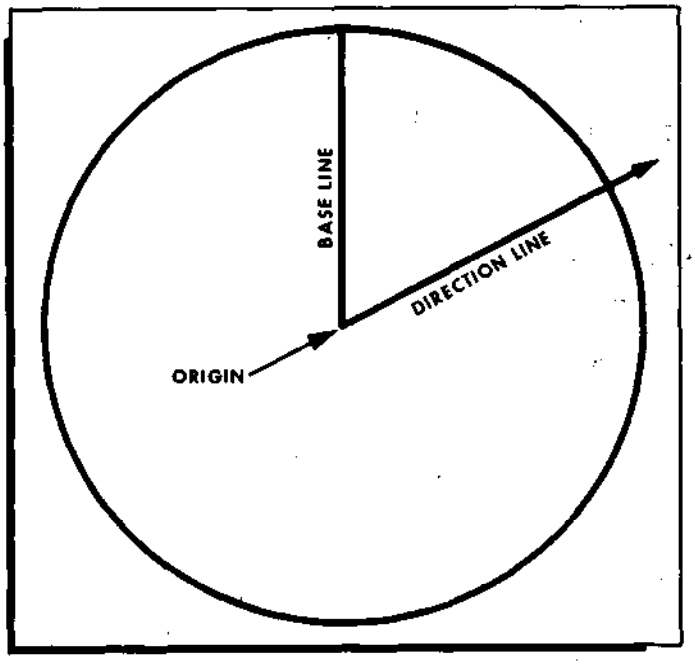

Figure 20-54. Origin of Azimuth Circle.

sured, having an actual value of 5°. The position of the three prongs in relation to each other varies according to the declination data for each map.

(f) Some older maps have a note under the declination diagram which gives the magnetic declination for a certain year and the amount of annual change. The annual change is so small when compared to the ½° value of the G-M angle that it is no longer shown on standard large scale maps.

(4) Protractors. Protractors come in several forms—full circle, half circle, square, and rectangular (figure 20-58). All of them divide a circle into units of angular measure, and regardless of their shape, consist of a scale around the outer edge and an index mark. The index mark is the center of the protractor circle from which all the direction lines radiate.

(a) To determine the grid azimuth of a line from one point to another on the map from (A to B or C to D) (figure 20-59), draw a line connecting the two points.

-1. Place the index of the protractor at that point where the line crosses a vertical (north-south) grid line.

-2. Keeping the index at this point, align the 0° - 180° line of the protractor on the vertical grid line.

-3. Read the value of the angle from the scale; this is the grid azimuth to the point.

(b) To plot a direction line from a known point on a map (figure 20-60):

-1. Construct a north-south grid line through the known point:

-a. Generally, align the 0° - 180° line of the protractor in a north-south direction through the known point.

-b. Holding the 0° - 180° line of the protractor on the known point, slide the protractor in the north-south direction until the horizontal line of the protractor (connecting the protractor index and the 90° tick mark) is aligned on an east-west grid line.

-c. Then draw a line connecting 0°, the known point, and 180°.

-2. Holding the 0° - 180° line on the north-south line, slide the protractor index to the known point.

-3. Make a mark on the map at the required angle. (In an evasion situation, do not mark on the map.)

-4. Draw a line from the known point through the mark made on the map. This is the grid direction line.

(5) The Compass and Its Uses:

(a) The magnetic compass is the most commonly used and simplest instrument for measuring directions and angles in the field. The lensatic compass (figure 20-61) is the standard magnetic compass for military use today.

(b) The lensatic compass must always be held level and firm when sighting on an object and reading an azimuth (figure 20-62). There are several techniques for holding the compass and sighting. One way is to align the sighting slot with the hairline on the front sight

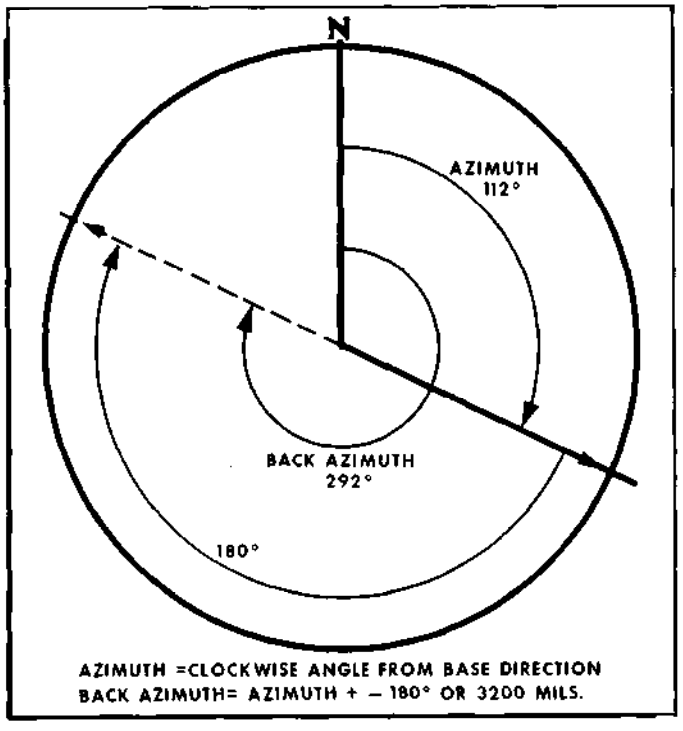

Figure 20-55. Azimuth and Back Azimuth.

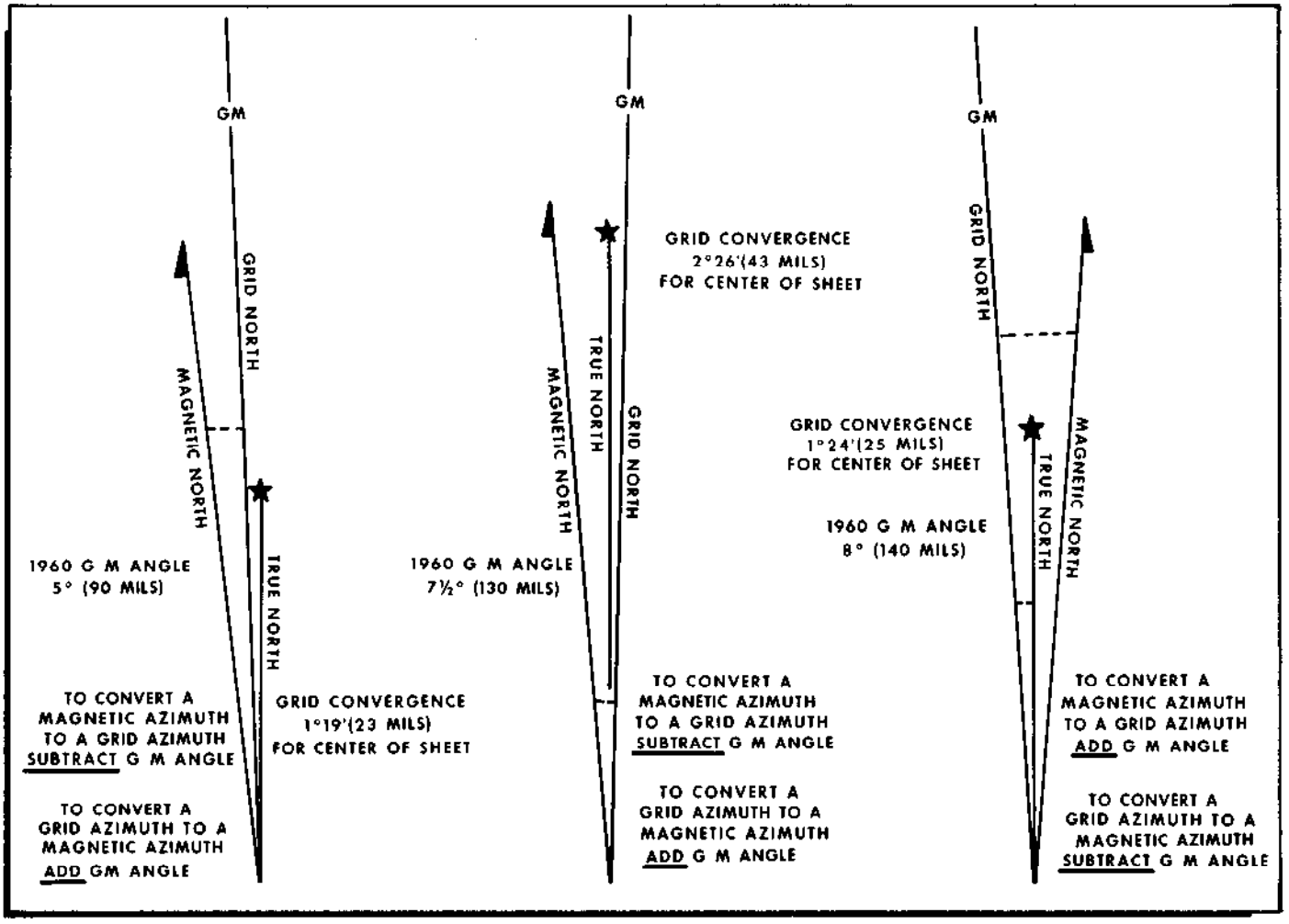

Figure 20-56. Declination Diagram (East and West).

in the cover and the target. The azimuth can then be read by glancing down at the dial through the lens. This technique provides a reading precise enough to use.

(6) Night Use of the Compass. For night use, special features of the compass include the luminous markings, the bezel ring, and two luminous sighting dots. Turning the bezel ring counterclockwise causes an increase in azimuth, while turning it clockwise causes a decrease. The bezel ring has a stop and spring which allows turns at 3° intervals per click and holds it at any desired position. One accepted method for determining compass directions at night is:

(a) Rotate the bezel ring until the luminous line is over the black index line.

(b) Hold the compass with one hand and rotate the bezel ring in a counterclockwise direction with the other hand to the number of clicks required. The number of clicks is determined by dividing the value of the required azimuth by 3. For example, for an azimuth of 51°, the bezel ring would be rotated 17 clicks counterclockwise (figure 20-63).

(c) Turn the compass until the north arrow is directly under the luminous line on the bezel.

(d) Hold the compass open and level in the palm of the left hand with the thumb along the side of the compass. In this manner, the compass can be held consistently in the same position. Position the compass approximately halfway between the chin and the belt, pointing to the direct front. (Practice in daylight will make a person proficient in pointing the compass the same way every time.) Looking directly down into the compass, turn the body until the north arrow is under the luminous line. Then proceed forward in the direction of the luminous sighting dots (figure 20-61). When the compass is to be used in darkness, an initial azimuth should be set while light is still available. With this initial azimuth as a base, any other azimuth which is a multiple of 3° can be established through use of the clicking feature of the bezel ring. The magnetic compass is a delicate instrument, especially the dial balance. The survivor should take care in its use. Compass readings should never be taken near visible masses of iron or electrical circuits.

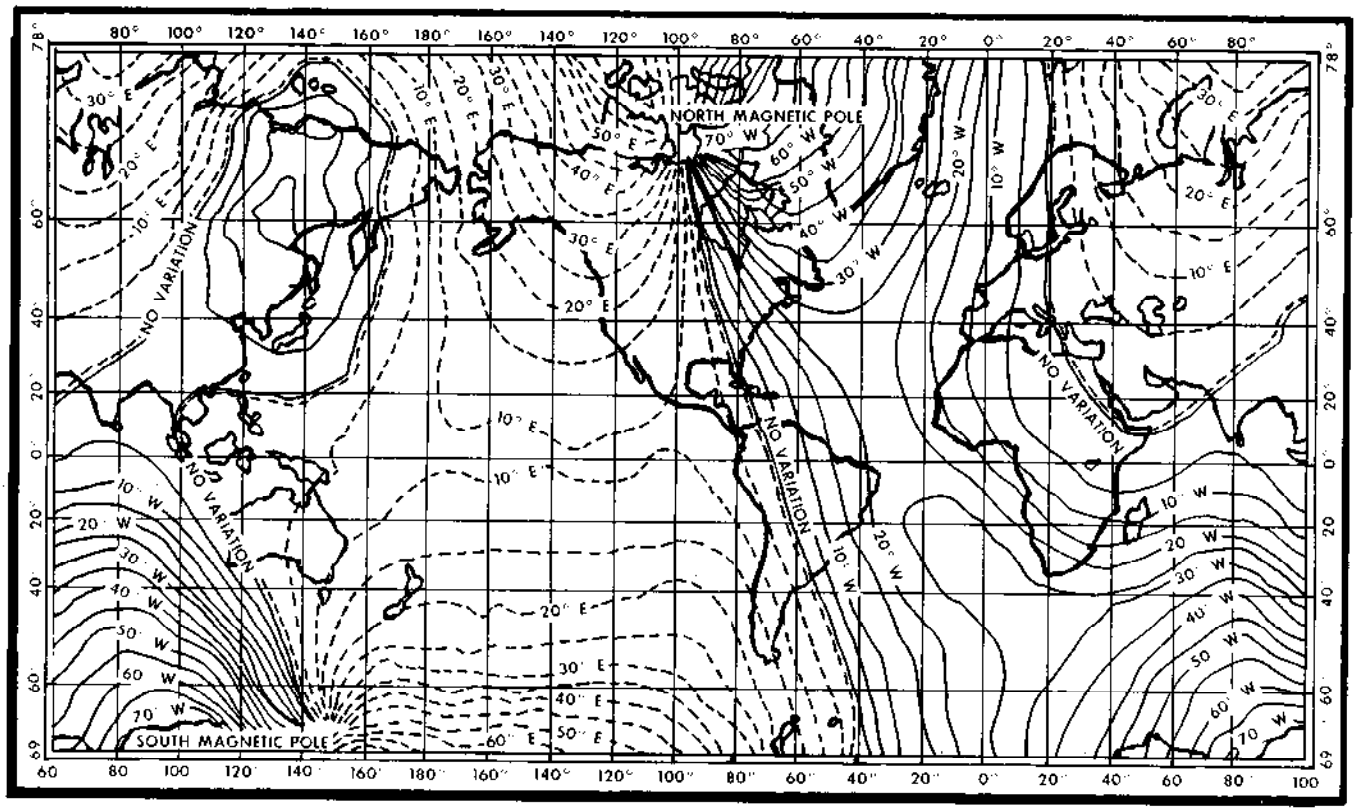

Figure 20-57. Lines of Magnetic Variation.

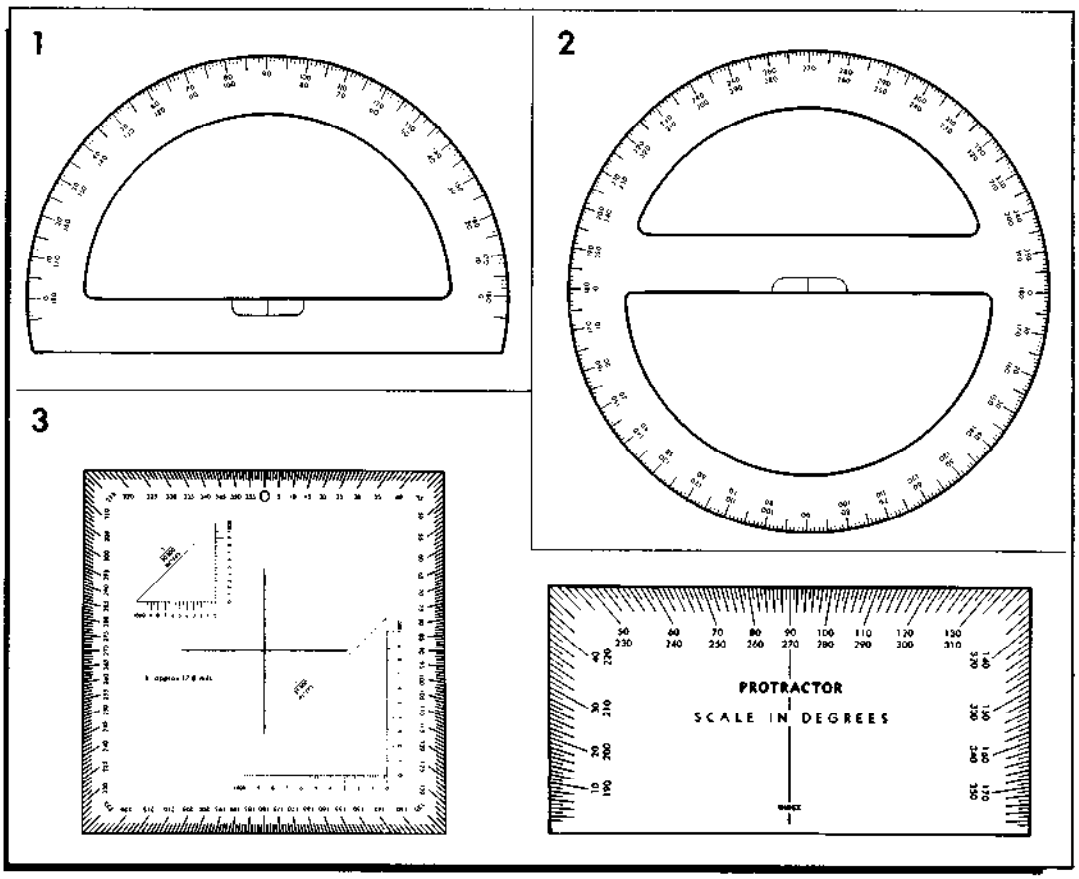

Figure 20-58. Types of Protractors.

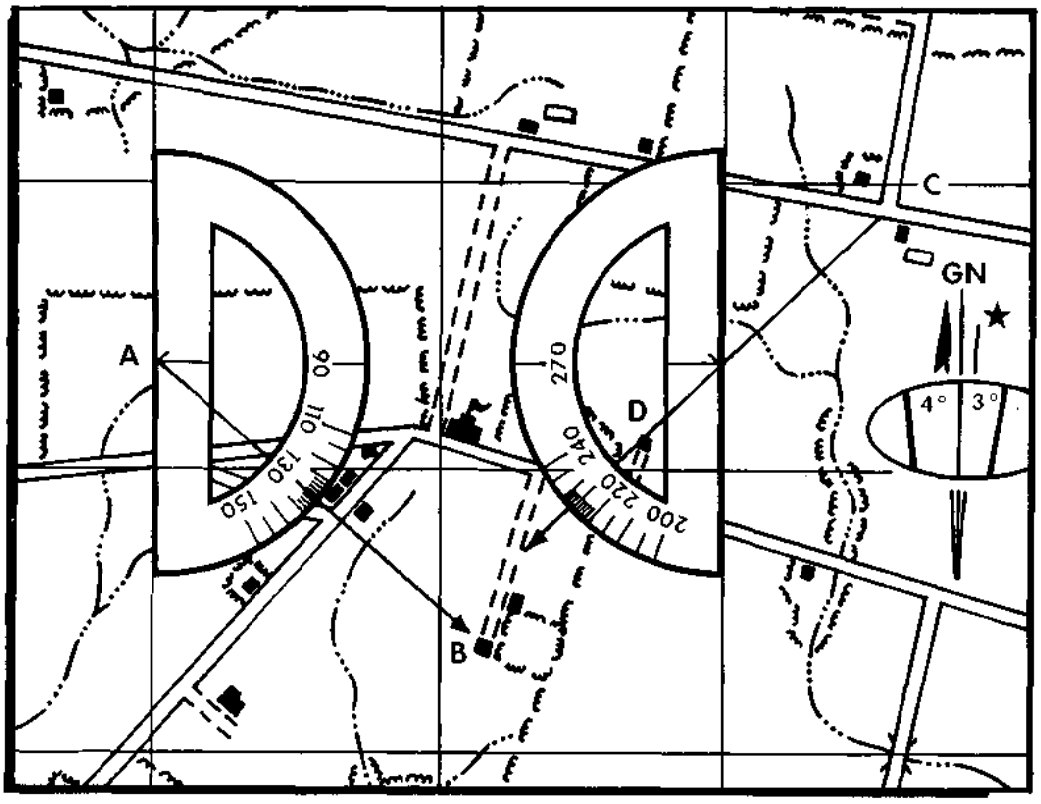

Figure 20-59. Measuring an Azimuth on a Map.

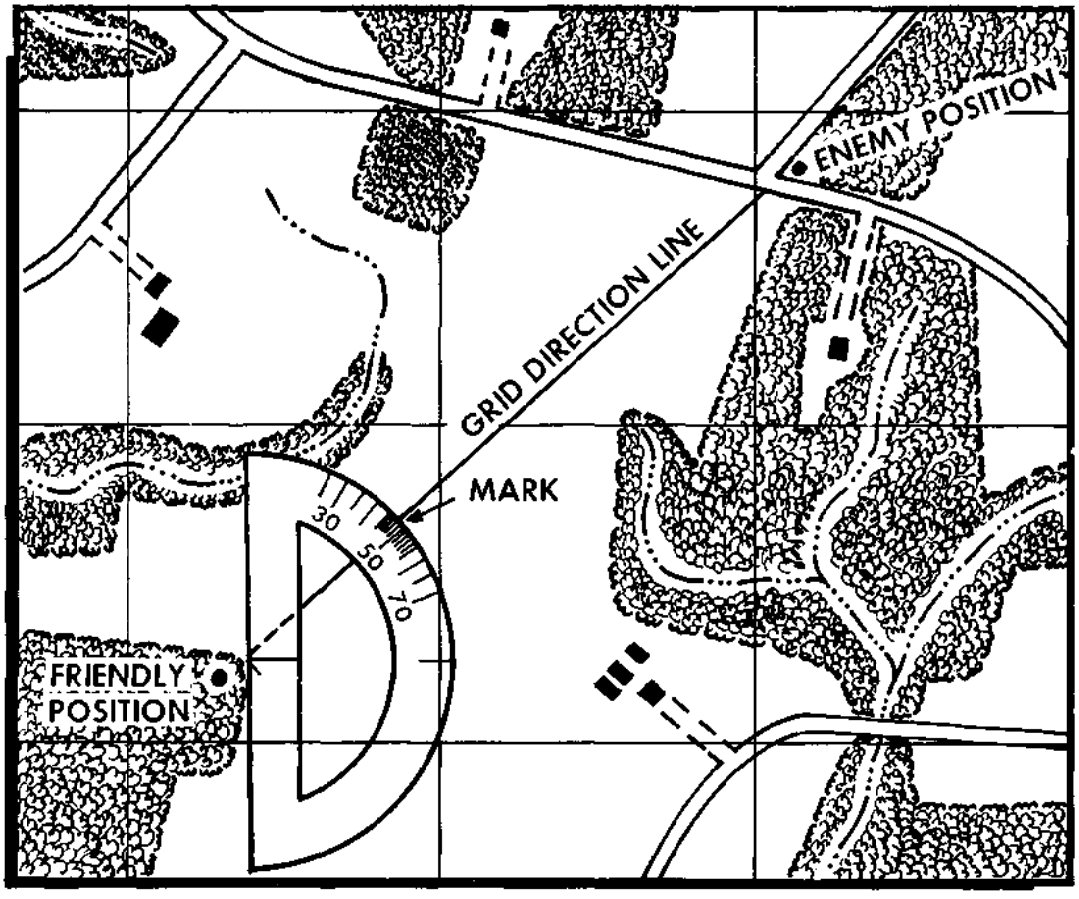

Figure 20-60. Plotting an Azimuth on a Map.

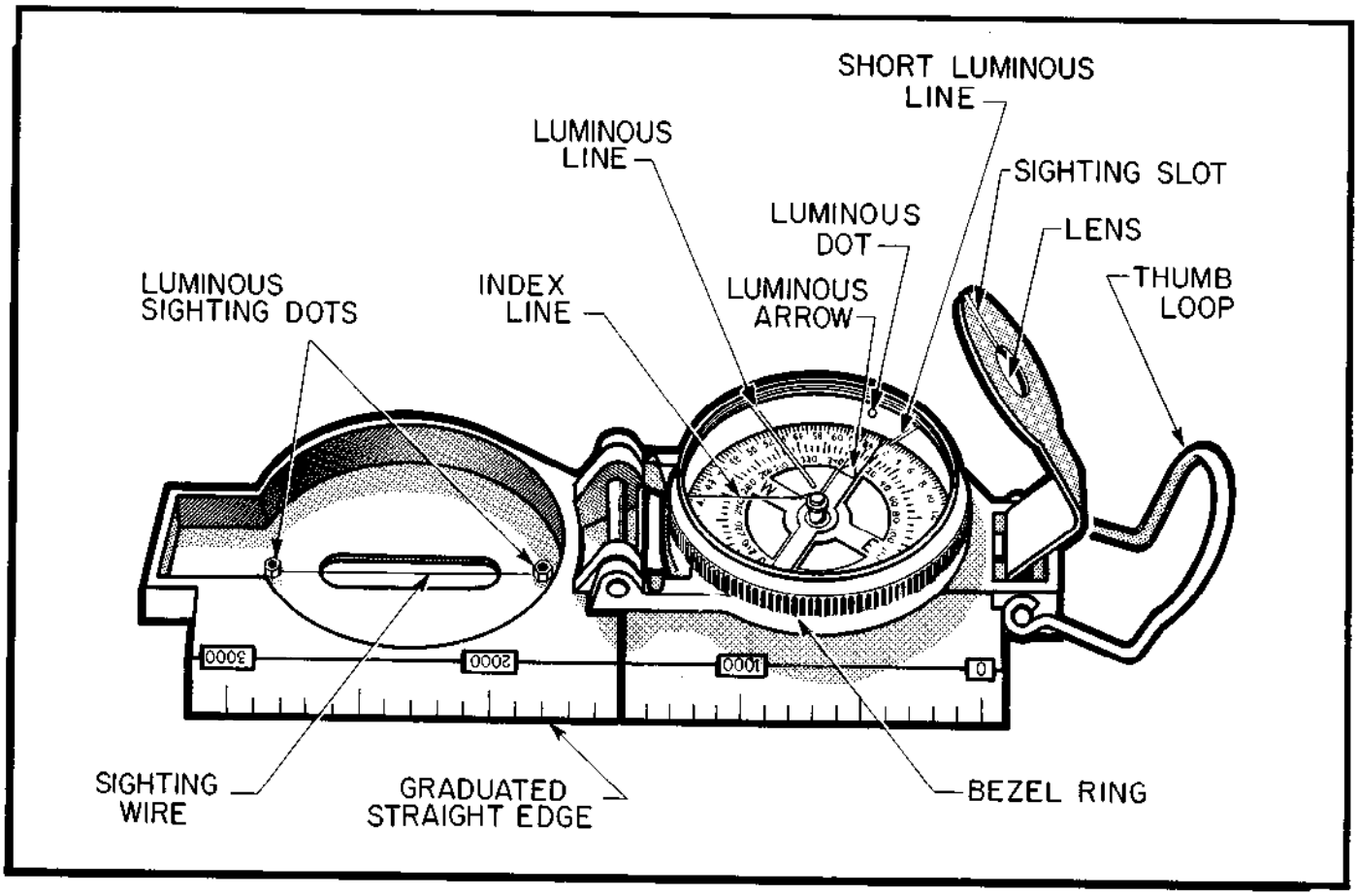

Figure 20-61. Lensatic Compass.

Figure 20-62. Holding the Compass.

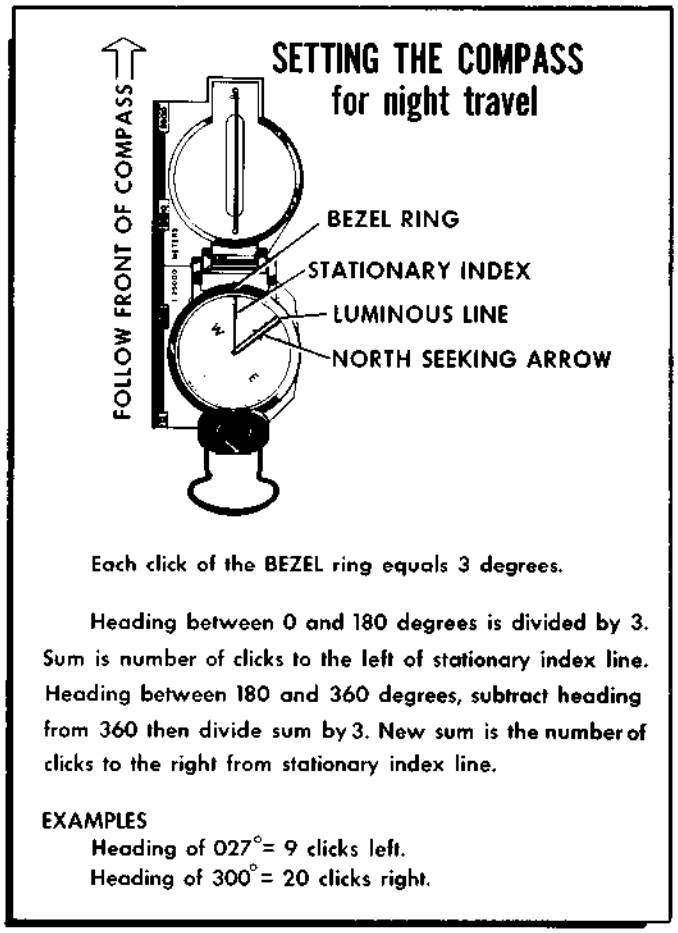

Figure 20-63. Night Travel.

(7) Map Orientation:

(a) A map is oriented when it is in a horizontal position with its north and south corresponding to north and south on the ground. The best way to orient a map is with a compass. (NOTE: Caution should be used to ensure nothing (metal, mine ore, etc.) in the area will alter the compass reading.)

(b) With the map in a horizontal position, the lensatic compass is placed parallel to a north-south grid line with the cover side of the compass pointing toward the top of the map. This will place the black index line on the dial of the compass parallel to grid north. Since the needle on the compass points to magnetic north, a declination diagram is (on the face of the compass) formed by the index line and the compass needle.

(c) Rotate the map and compass until the directions of the declination diagram formed by the black index line and compass needle match the directions shown on the declination diagram printed in the margin of the map. The map is then oriented (grid north).

(d) If the magnetic north arrow on the map is to the left of grid north, the compass reading will equal the G-M angle (given in the declination diagram). If the magnetic north is to the right of grid north, the compass reading will equal 360° minus the G-M angle. In figure 20-64, the declination diagram illustrates a magnetic north to the right of grid north and the compass reading will be 360° minus 21½° or 338½°.

(e) Remember to point the compass north arrow in the same direction as the magnetic north arrow, and the compass reading (equal to the G-M angle or the 360° minus G-M angle) will be quite apparent.

(f) In summary, if the variation is to the east of true north or the magnetic north arrow of the declination diagram is to the east (right) of the grid north line, subtract the degrees of variation from 360°. If it is to the left (west), add to 000°. East is least and west is best.

(g) If a grid line is not used, a true north-south line can be used. True north-south lines are longitudinal lines or lines formed by the vertical lines on a tick map (assuming the top of the map is north). The same procedure is used if magnetic variation is figured from true north—not grid north.

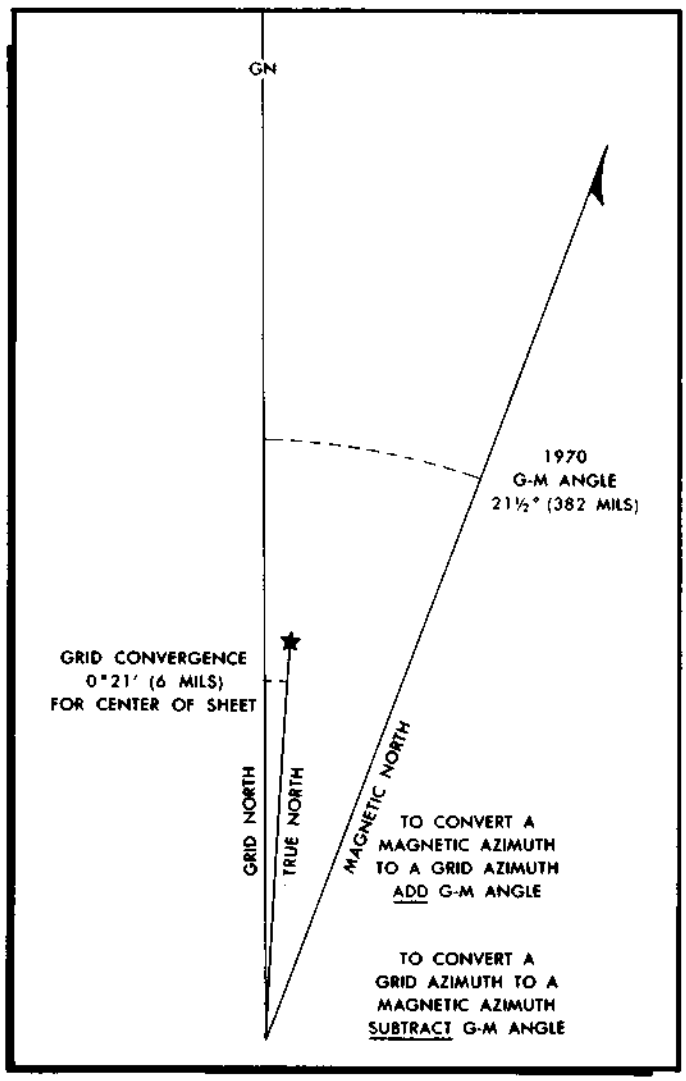

Figure 20-64. Declination Diagram.

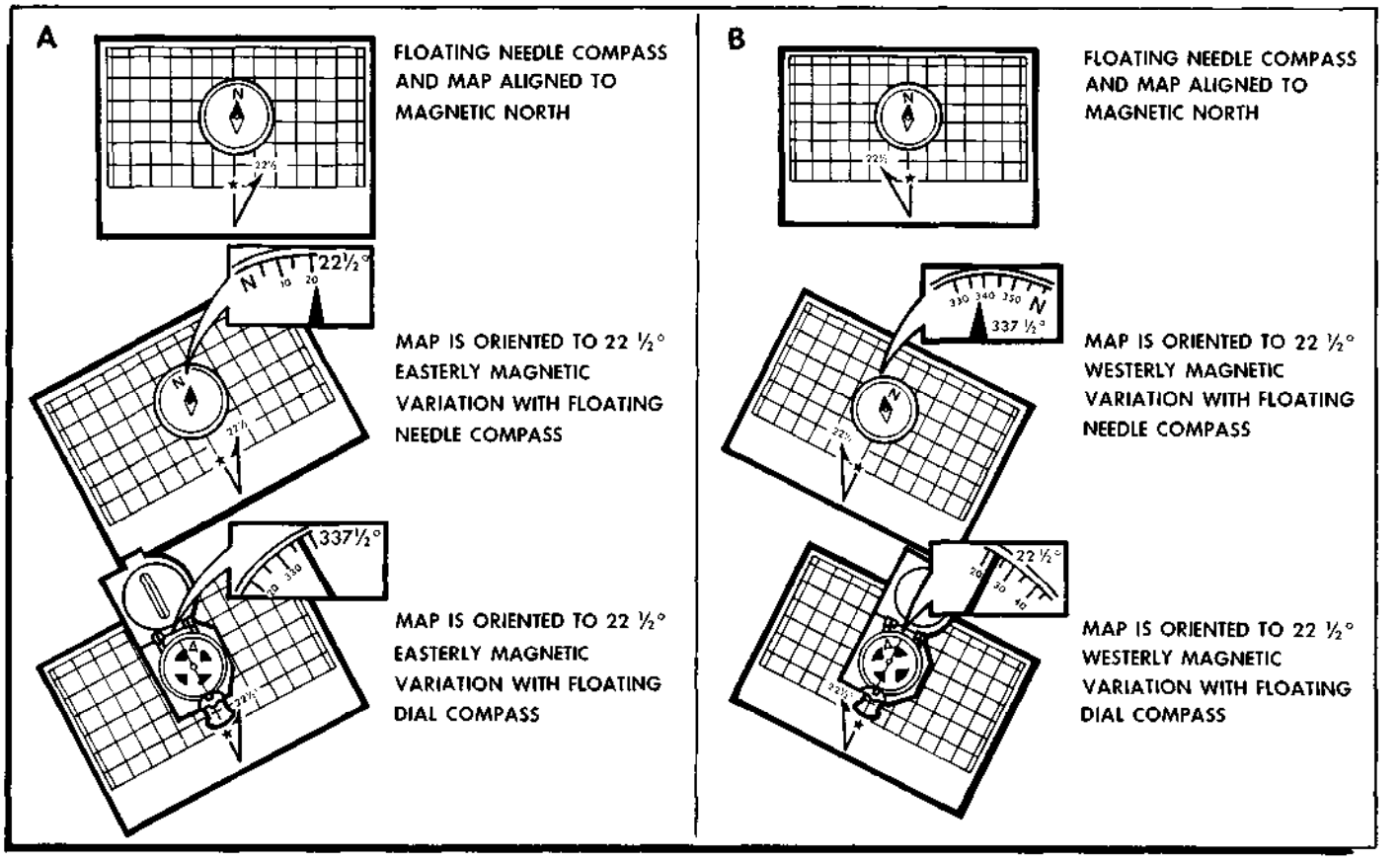

Figure 20-65. Floating Needle Compass.

(h) A floating needle compass (figures 20-65a and 20-65b) has a needle with a north direction marked on it. The degree and direction marks are stationary on the bottom inside of the compass. The button and wrist compasses may be floating dial or floating needle. To determine the heading, line up the north-seeking arrow over 360° by rotating the compass. Then read the desired heading. Orienting a map with a floating needle compass is similar to the method used with the floating dial. The only exception is with the adjustment for magnetic variation. If magnetic variation is to the east, turn the map and the compass to the left (the north axis of the compass should be aligned with the map north) so that the magnetic north-seeking arrow is pointing at the number of degrees on the compass which corresponds with the angle of declination.

(i) When a compass is not available, map orientation requires a careful examination of the map and the ground to find linear features common to both, such as roads, railroads, fence lines, power lines, etc. By aligning the feature on the map with the same feature on the ground (figure 20-66), the map is oriented. Orientation by this method must be checked to prevent the reversal of directions which may occur if only one linear feature is used. This reversal may be prevented by aligning two or more map features (terrain or manmade). If no second linear feature is visible but the map user's position is known, a prominent object may be used. With the

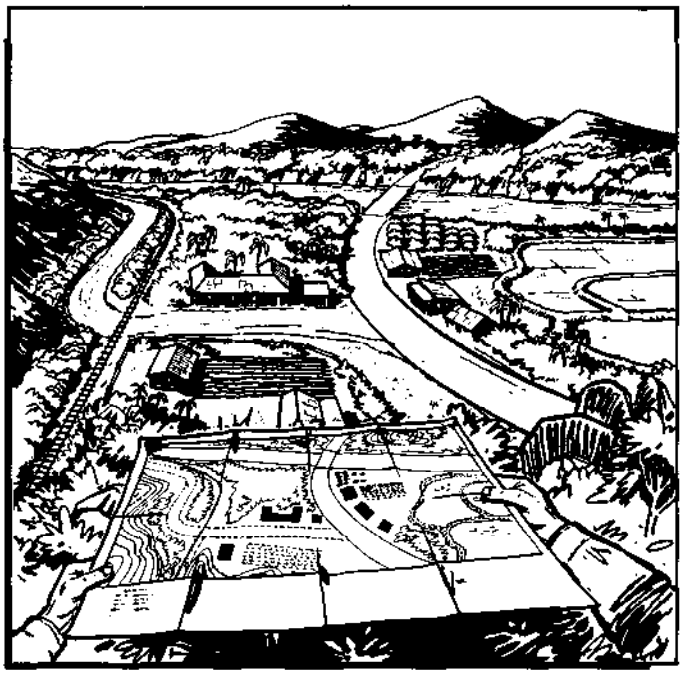

Figure 20-66. Map Orientation by Inspection.

prominent object and the user's position connected with a straight line on the map, the map is rotated until the line points toward the feature.

(j) If two prominent objects are visible and plotted on the map and the position is not known, move to one of the plotted and known positions, place the straightedge or protractor on the line between the plotted positions, turn the protractor and the map until the other plotted and visible point is seen along the edge. The map is then oriented.

(k) When a compass is not available and there are no recognizable prominent landforms or other features, a map may be oriented by any of the field expedient methods we will now discuss.

(8) Determining Cardinal Directions Using Field Expedients:

*(a) Shadow tip method of determining direction and time. This simple method of finding direction by the Sun consists of only three basic steps (figure 20-67).

-1. Step 1. Place a stick or branch into the ground at a fairly level spot where a distinct shadow will be cast. Mark the shadow tip with a stone, twig, or other means.

-2. Step 2. Wait until the shadow tip moves a few inches. If a 4-foot stick is being used, about 10 minutes should be sufficient. Mark the new position of the shadow tip in the same way as the first.

-3. Step 3. Draw a straight line through the two marks to obtain an approximate east-west line. If uncertain which direction is east and which is west, observe this simple rule: The Sun "rises in the east and sets in the west" (but rarely DUE east and DUE west). The shadow tip moves in just the opposite direction. Therefore, the first shadow-tip mark is always in the west direction, and the second mark in the east direction, anyplace on Earth.

(b) A line drawn at right angles to the east-west line at any point is the approximate north-south line, which will help orient a person to any desired direction of travel.

(c) Inclining the stick to obtain a more convenient shadow does not impair the accuracy of the shadow-tip method. Therefore, a traveler on sloping ground or in highly vegetated terrain need not waste valuable time looking for a large level area. A flat spot, the size of the hand, is all that is necessary for shadow-tip markings and the base of the stick can be either above, below, or to one side of it. Also, any stationary object (the end of a tree limb or the notch where branches are jointed) serves just as well as an implanted stick because only

*From *Better Ways of Pathfinding*, by Robert S. Owendoff, 153 Cundry Drive, Falls Church VA 22046. 1964§ by Stackpole Company, Harrisburg PA. All rights reserved by copyright owner (author).

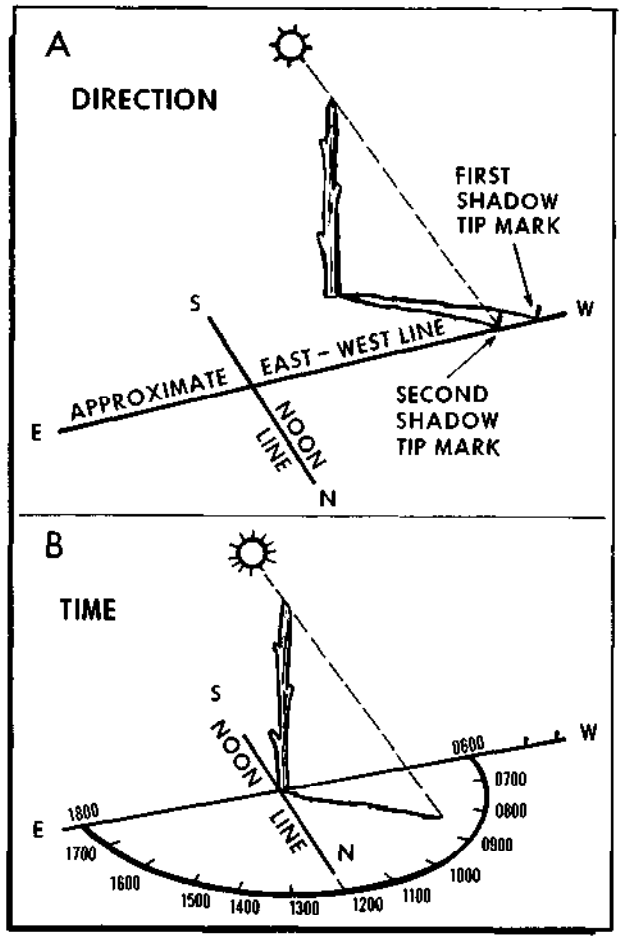

Figure 20-67. Determining Time and Direction by Shadow.

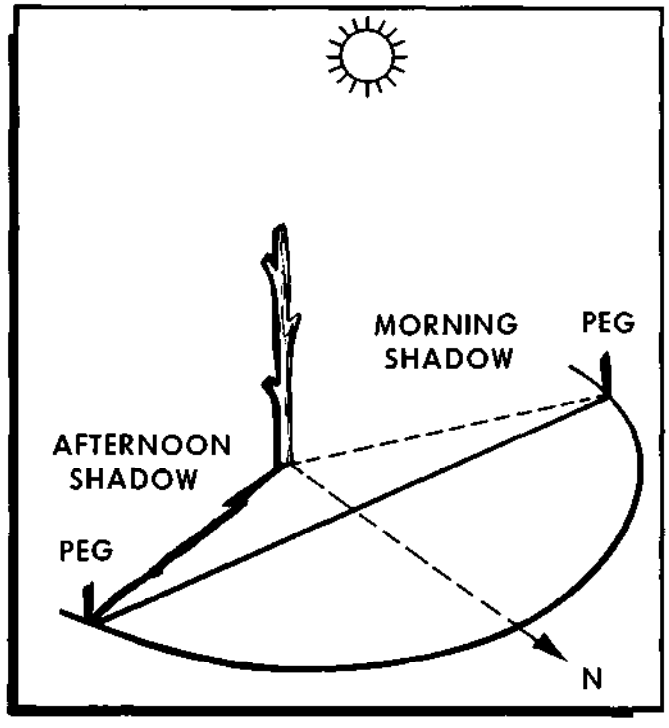

Figure 20-68. Equal Shadow Method of Determining Direction.

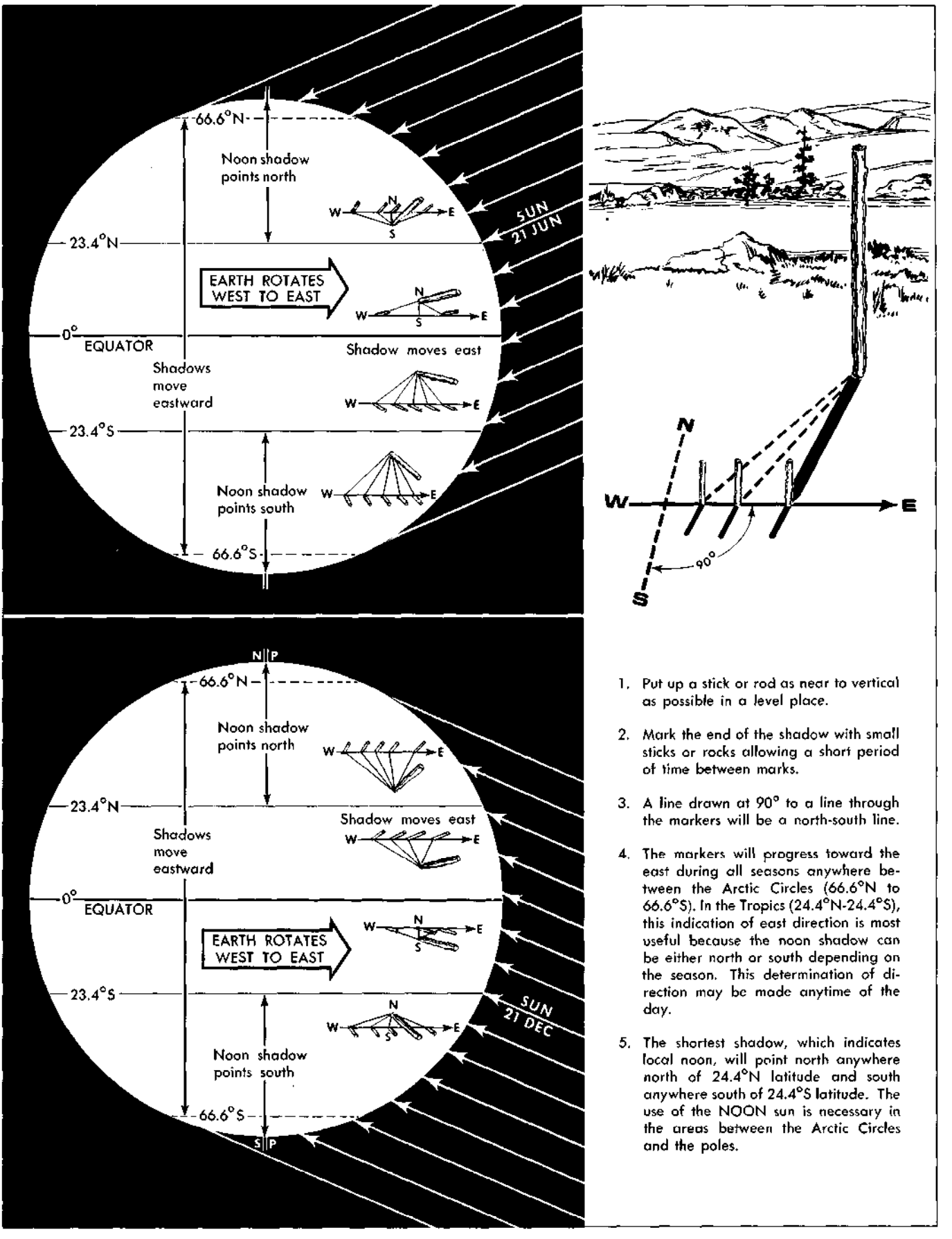

1. Put up a stick or rod as near to vertical as possible in a level place.
2. Mark the end of the shadow with small sticks or rocks allowing a short period of time between marks.
3. A line drawn at 90° to a line through the markers will be a north-south line.
4. The markers will progress toward the east during all seasons anywhere between the Arctic Circles (66.6°N to 66.6°S). In the Tropics (24.4°N-24.4°S), this indication of east direction is most useful because the noon shadow can be either north or south depending on the season. This determination of direction may be made anytime of the day.
5. The shortest shadow, which indicates local noon, will point north anywhere north of 24.4°N latitude and south anywhere south of 24.4°S latitude. The use of the NOON sun is necessary in the areas between the Arctic Circles and the poles.

Figure 20-69. Stick and Shadow Method of Determining Direction.

the shadow tip is marked.

(d) The shadow-tip method can also be used to find the approximate time of day (figure 20-67B).

-1. To find the time of day, move the stick to the intersection of the east-west line and the north-south line, and set it vertically in the ground. The west part of the east-west line indicates the time is 0600 and the east part is 1800, ANYWHERE on Earth, because the basic rule always applies.

-2. The north-south line now becomes the noon line. The shadow of the stick is an hour hand in the shadow clock and with it the time can be estimated using the noon line and 6 o'clock line as the guides. Depending on the location and the season, the shadow may move either clockwise or counterclockwise, but this does not alter the manner of reading the shadow clock.

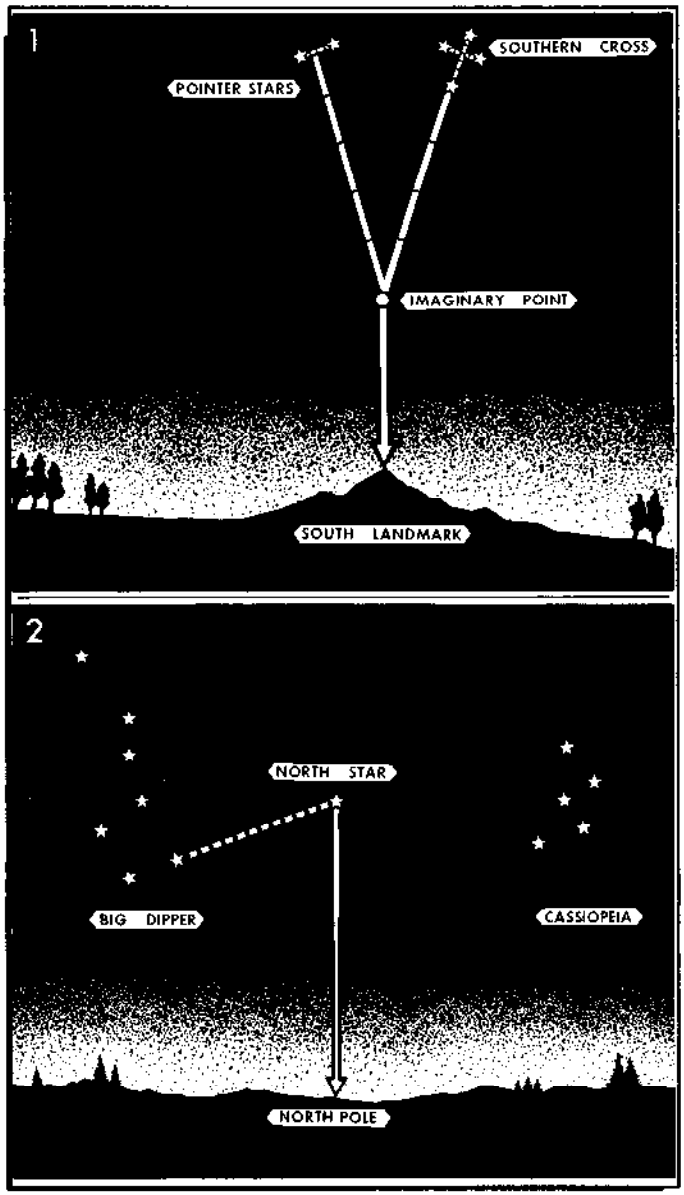

Figure 20-70. Determination of Direction by Using the Stars.

-3. The shadow clock is not a timepiece in the ordinary sense. It always reads 0600 at sunrise and 1800 at sunset. However, it does provide a satisfactory means of telling time in the absence of properly set watches. Being able to establish the time of day is important for such purposes as keeping a rendezvous, prearranged concerted action by separated persons or groups, and estimating the remaining duration of daylight. Shadow-clock time is closest to conventional clock time at mid-day, but the spacing of the other hours, compared to conventional time, varies somewhat with the locality and date.

(e) The shadow-tip system is ineffective for use beyond 66½° latitude in either hemisphere due to the position of the Sun above the horizon. Whether the Sun is north or south of a survivor at mid-day depends on the latitude. North of 23.4°N, the Sun is always due south at local noon and the shadow points north. South of 23.4°S, the Sun is always due north at local noon and the shadow points south. In the tropics, the Sun can be either north or south at noon, depending on the date and location but the shadow progresses to the east regardless of the date.

(f) Equal-shadow method of determining direction (Figures 20-68 and 20-69). This variation of the shadow-tip method is more accurate and may be used at all latitudes less than 66° at all times of the year.

-1. Step 1. Place a stick or branch into the ground vertically at a level spot where a shadow at least 12 inches long will be cast. Mark the shadow tip with a stone, twig, or other means. This must be done 5 to 10 minutes before noon (when the Sun is at its highest point (zenith)).

-2. Step 2. Trace an arc using the shadow as the radius and the base of the stick as the center. A piece of string, shoelace, or a second stick may be used to do this.

-3. Step 3. As noon is approached, the shadow becomes shorter. After noon, the shadow lengthens until it crosses the arc. Mark the spot as soon as the shadow tip touches the arc a second time.

-4. Step 4. Draw a straight line through the two marks to obtain an east-west line.

(g) Although this is the most accurate version of the shadow-tip method, it must be performed around noon. It requires the observer to watch the shadow and complete step 3 at the exact time the shadow tip touches the arc.

(h) At night, the stars may be used to determine the north line in the northern hemisphere or the south line in the southern hemisphere. Figure 20-70 shows how this is done.

(i) A watch can be used to determine the approximate true north or south (figure 20-71). In the northern hemisphere, the hour hand is pointed toward the Sun. A south line can be found midway between the hour hand

and 1200 standard time. During daylight savings time, the north-south line is midway between the hour hand and 1300. If there is any doubt as to which end of the line is north, remember that the Sun is in the east before noon and in the west in the afternoon.

(j) The watch may also be used to determine direction in the Southern Hemisphere; however, the method is different. The 1200-hour dial is pointed toward the Sun, and halfway between 1200 and the hour hand will be a north line. During daylight savings time, the north line lies midway between the hour hand and 1300.

(k) The watch method can be in error, especially in the extreme latitudes, and may cause "circling." To avoid this, make a shadow clock and set the watch to the time indicated. After traveling for an hour, take another shadow-clock reading.

(9) Determining Specific Position. When using a map and compass, the map must be oriented using the method described earlier in this chapter. Next, locate two or three known positions on the ground and the map. Using the compass, shoot an azimuth to one of the known positions (figure 20-72). Once the azimuth is determined, recheck the orientation of the map and plot the azimuth on the map. To plot the azimuth, place the front corner of the straightedge of the compass on the corresponding point on the map. Rotate the compass until the determined azimuth is directly beneath the stationary index line. Then draw a line along the straightedge of the compass and extend the line past the estimated position on the map. Repeat this procedure for the second point (figure 20-72). If only two azimuths are used, the technique is referred to as biangulation (figure 20-72). If a third azimuth is plotted to check the accuracy of the first two, the technique is called triangulation (figure 20-72). When using three lines, a triangle of error may be formed. If the triangle is large, the work should be checked. However, if a small triangle is formed, the user should evaluate the terrain to determine the actual position. One azimuth may be used with a linear land feature such as a river, road, railroad, etc., to determine specific position (figure 20-72).

(10) Determining Specific Location Without a Compass. A true north-south line determined by the stick and shadow, Sun and watch, or celestial constella-

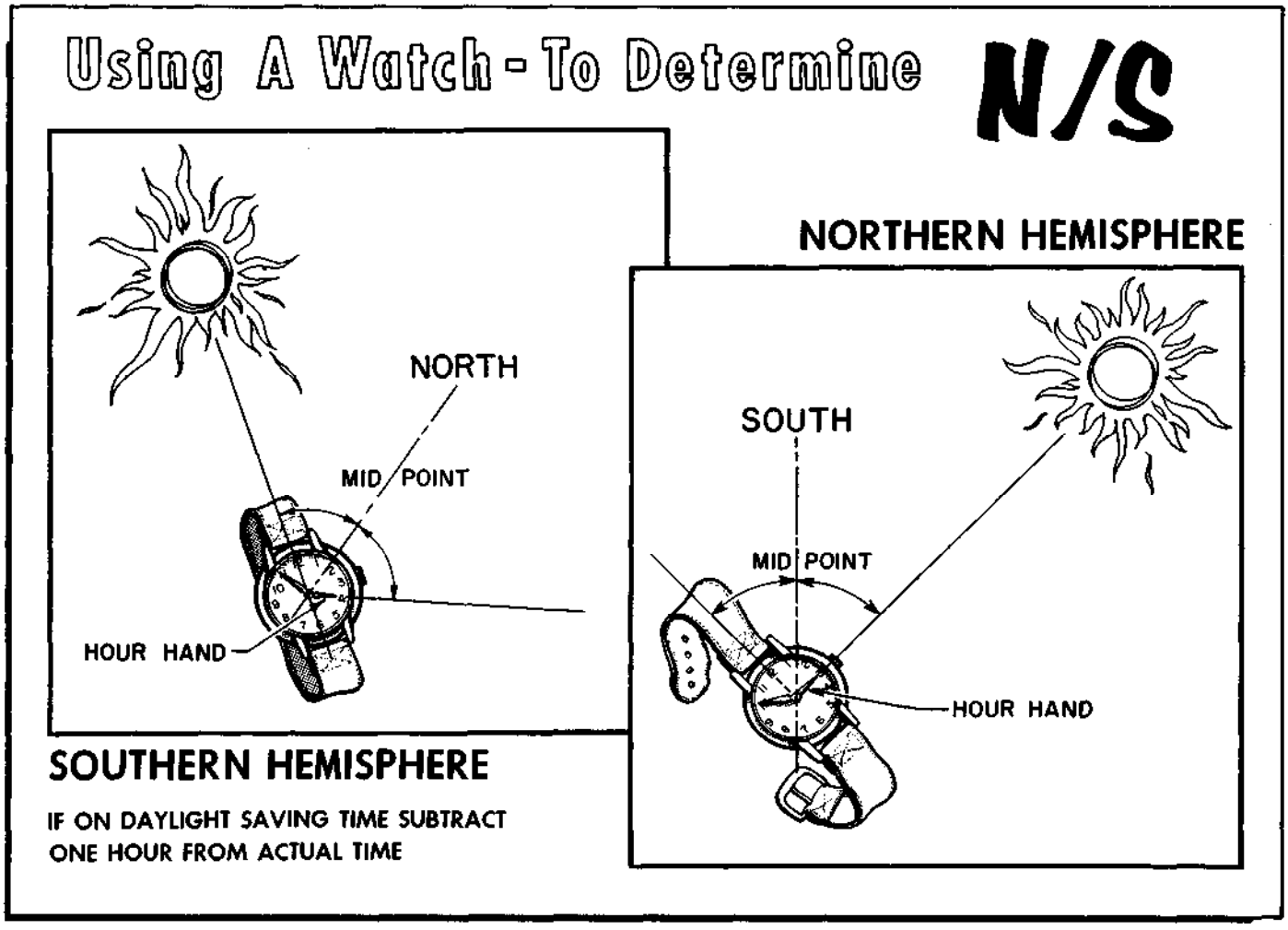

Figure 20-71. Directions Using a Watch.

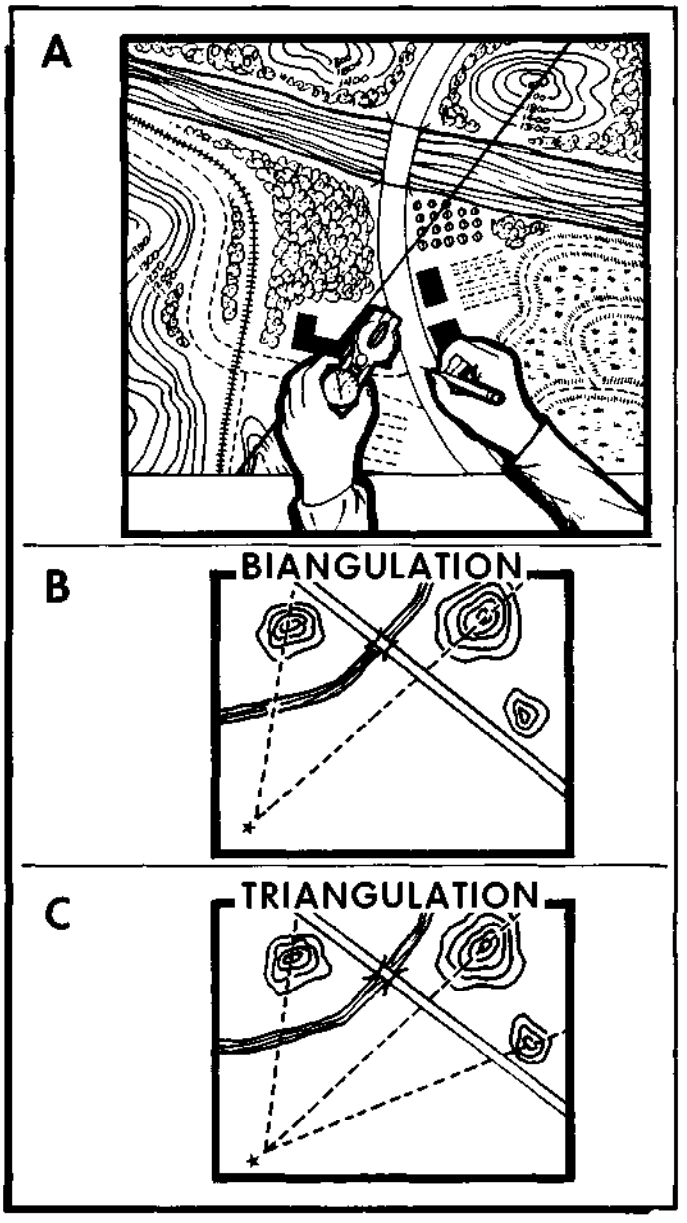

Figure 20-72. Azimuth, Biangulation, and Triangulation.

tion method may be used to orient the map without a compass. However, visible major land features can be used to orient the map to the lay of the land. Once the map is oriented, identify two or three landmarks and mark them on the map. Lay a straightedge on the map with the center of the straightedge at a known position as a pivot point and rotate the straightedge until the known position of the map is aligned with present position, and draw a line. Repeat this for the second and third position. Each time a line of position is plotted, the map must still be aligned with true north and south. If three lines of position are plotted and form a small triangle, use terrain evaluation to determine present position. If they form a large triangle, recheck calculations for errors.

(11) Dead Reckoning:

(a) Dead reckoning is the process of locating one's position by plotting the course and distance from the last known location. In areas where maps exist, even poor ones, travel is guided by them. It is a matter of knowing one's position at all times by associating the map features with the ground features. A great portion of the globe is unmapped or only small scale maps are available. The survivor may be required to travel in these areas without a usable map. Although these areas could be anywhere, they are more likely to be found in frozen wastelands and deserts.

(b) For many centuries, mariners used dead reckoning to navigate their ships when they were out of sight of land or during bad weather, and it is just as applicable to navigation on land. Movement on land must be carefully planned. In military movement, the starting location and destination are known, and if a map is available, they are carefully plotted along with any known intermediate features along the route. These intermediate features, if clearly recognizable on the ground, serve as checkpoints. If a map is not available, the plotting is done on a blank sheet of paper. A scale is selected so the entire route will fit on one sheet. A north direction is clearly established. The starting point and destination are then plotted in relationship to each other. If the terrain and enemy situations permit, the ideal course is a straight line from starting point to destination. This is seldom possible or practicable. The route of

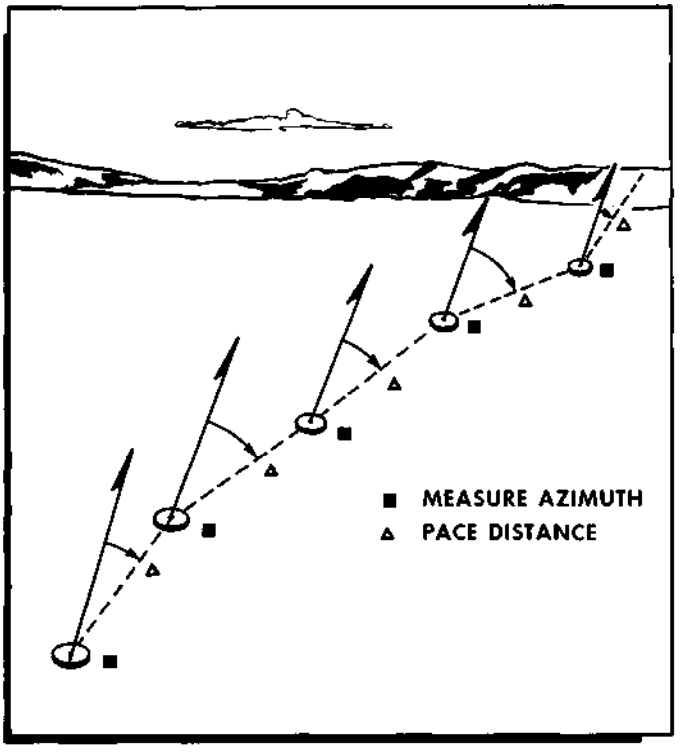

Figure 20-73. Compass Navigation on Foot.

1	2	3	4	5	6	7
ODOMETER READING AT START OF EACH COURSE	DISTANCE IN MILES	FORWARD AZIMUTH (MAGNETIC)	DECLINATION CORRECTION	DEVIATION CORRECTION	TRUE AZIMUTH	NOTES
A 4750						
	6	17°	13°	+3°	33°	
B 4756						
	9	358°	"	+2°	13°	
C 4765						
	8	341°	"	+1°	355°	
D 4773						
	1	314°	"	0°	327°	
E 4774						
	1.5	341°	"	+1°	355°	
F 4775	5					
	1.5	322°	"	0°	335°	
G 4777						
	1	312°	"	0°	325°	
H 4778						
	12	300°	12°	−1°	311°	
I 4790						
	6	341°	"	+1°	354°	
J 4796						
	6	302°	"	−1°	313°	
K 4802						
	20	319°	"	0°	331°	
4810						Wahoo River
						Crossing
4814						Cut 2 miles
L 4824			Base Camp (data)			

Figure 20-74. Sample Log.

travel usually consists of several courses, with an azimuth established at the starting point for the first course to be followed. Distance measurement begins with the departure and continues through the first course until a change in direction is made. A new azimuth is established for the second course and the distance is measured until a second change of direction is made, and so on. Records of all data are kept and all positions are plotted.

(c) A pace (for our purposes) is equal to the distance covered every time the same foot touches the ground (surveyor's paces). To measure distance, count the number of paces in a given course and convert to the map unit. Usually, paces are counted in hundreds, and hundreds can be kept track of in many ways: making notes in a record book; counting individual fingers; placing small objects such as pebbles into an empty pocket; tying knots in a string; or using a mechanical hand counter. Distances measured this way are only approximate, but with practice can become very accurate. It is important that each person who uses dead reckoning navigation establish the length of an average pace. This is done by pacing a measured course many times and computing the mean (figure 20-73). In the field, an average pace must often be adjusted because of the following conditions:

-1. Slopes. The pace lengthens on a downgrade and shortens on an upgrade.

-2. Winds. A headwind shortens the pace while a tailwind increases it.

-3. Surfaces. Sand, gravel, mud and similar surface materials tend to shorten the pace.

-4. Elements. Snow, rain, or ice reduces the length of the pace.

-5. Clothing. Excess weight of clothing shortens the pace while the type of shoes affects traction and therefore the pace length.

(d) A log (figure 20-74) should be used for navigation, by dead reckoning, to record all of the distances and azimuths while traveling. Often, relatively short stretches of travel cannot be traversed in a straight course because of some natural features such as a river, or a steep, rugged slope. This break in normal navigation is shown on the log to ensure proper plotting.

(e) The course of travel may be plotted directly on the face of the map or on a separate piece of paper at the same scale as the map. If the latter method is chosen, the complete plot can be transferred to the map sheet, if at least one point of the plot is also shown on the map. The actual plotting can be done by protractor and scale. The degree of accuracy obtained depends upon the quality of draftmanship, the environmental conditions, and the care taken in obtaining data while en route. Figure 20-75 illustrates a paper plot of the data obtained for the log sample in figure 20-74. It should be noted that four of the courses from A to H are short and have been plotted as a single course, equal to the sum of the four distances and using a mean azimuth of the four. This is recommended because it saves time without a loss of accuracy. If possible, a plot should be tied into at least one known intermediate point along the route. This is done by directing the route to pass near or over a point. If the plotted position of the intermediate point differs from its known location, discard the previous plot and start a new plot from the true location. The previous plot may be inspected to see if there is a detectable constant error applicable to future plots; otherwise, it is of no further use.

(f) An offset is a planned magnetic deviation to the right or left of an azimuth to an objective. It is used when approaching a linear feature from the side, and a point along the linear feature (such as a road junction) is the objective. Because of errors in the compass, or in map reading, one may reach the linear feature and not know whether the objective lies to the right or left. A deliberate offset by a known number of degrees in a known direction compensates for possible errors and ensures that, upon reaching the linear feature, the user knows whether to go right or left to reach the objective.

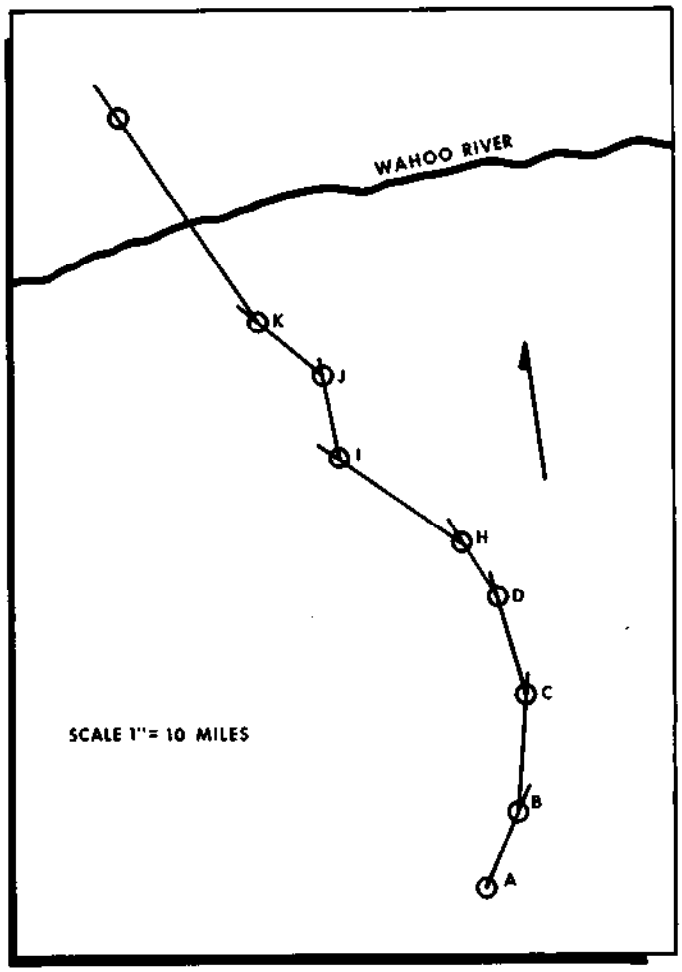

Figure 20-75. Separate Paper Plot.

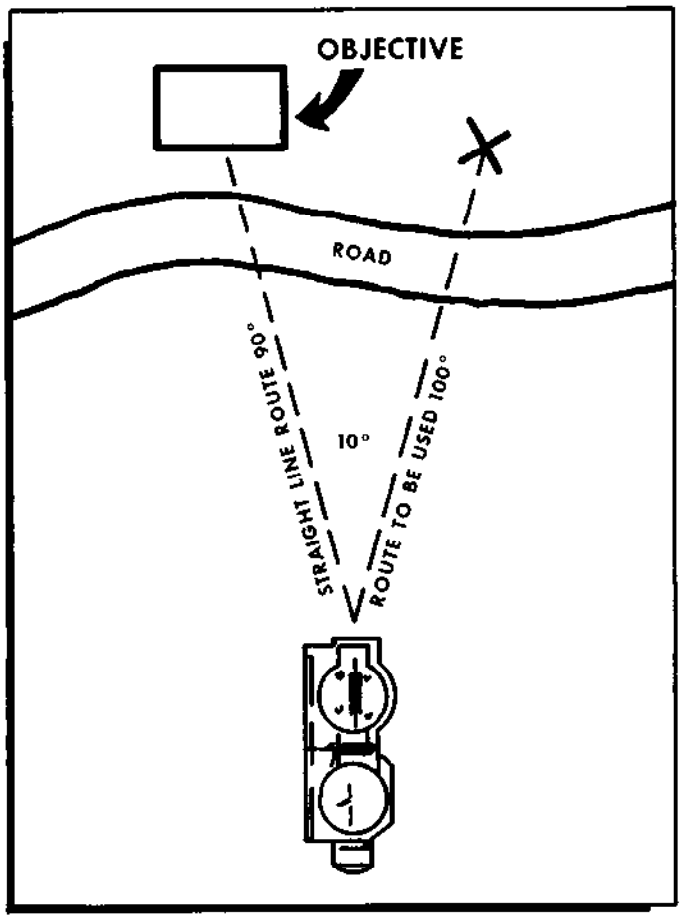

Figure 20-76. Deliberate Offset.

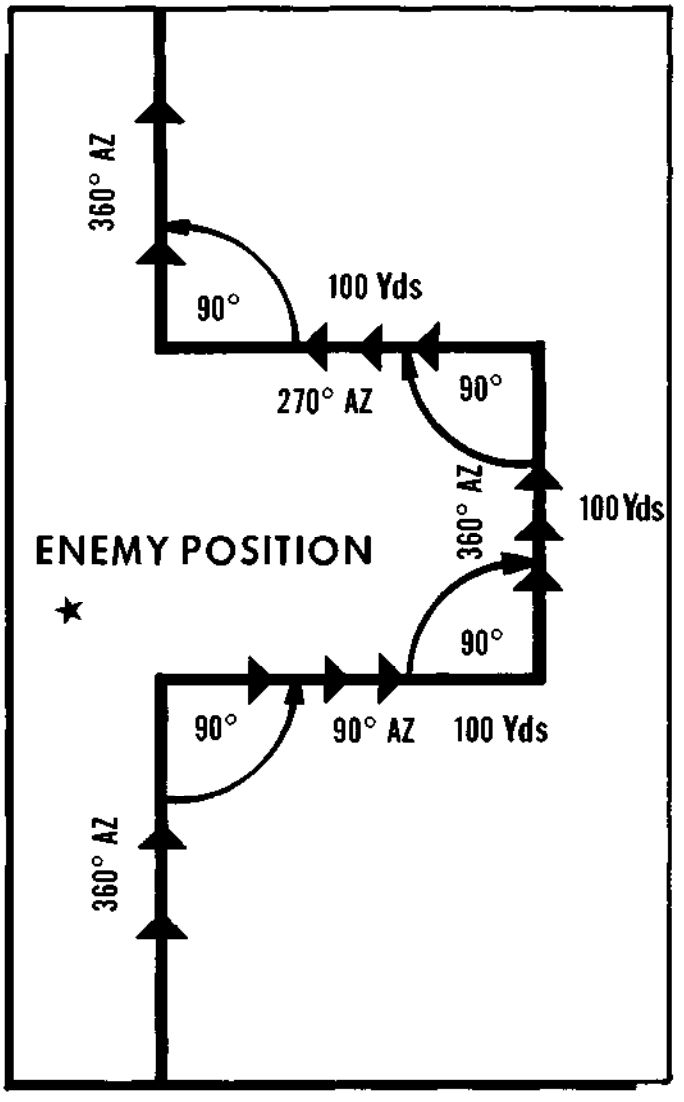

Figure 20-77. Detour Around Energy Position.

Figure 20-76 shows an example of the use of offset to approach an objective. It should be remembered that the distance from "X" to the objective will vary directly with the distance to be traveled and the number of degrees offset. Each degree offset will move the course about 20 feet to the right or left for each 1,000 feet traveled. For example: In figure 20-76, the number of degrees offset is 10 to the right. If the distance traveled to "X" is 1,000 feet, then "X" is located about 200 feet to the right of the objective.

(g) Figure 20-77 shows an example of how to bypass enemy positions or obstacles by detouring around them and maintaining orientation by moving at right angles for specified distances; for example, moving on an azimuth 360°and wish to bypass an obstacle or position. Change direction to 90° and travel for 100 yards, change direction back to 360° and travel for 100 yards, change direction to 270° and travel for 100 yards, then change direction to 360°, and back on the original azimuth. Bypassing an unexpected obstacle at night is done in the same way.

(12) Polar Coordinates:

(a) A point on the map may be determined or plotted from a known point by giving a direction and a distance along the direction line. This method of point location uses polar coordinates (figure 20-78). The reference direction is normally expressed as an azimuth, and the distance is determined by any convenient unit of measurement such as meters or yards.

(b) Polar coordinates are especially useful in the field because magnetic azimuth is determined from the compass and the distance can be estimated.

(13) Position Determination:

(a) Determining Latitude. (From the Sun at sunrise and sunset), Figure 20-79 shows the true azimuth of the rising Sun and the relative bearing of the setting Sun for all of the months in the year in the Northern and Southern Hemispheres (the table assumes a level horizon and is inaccurate in mountainous terrain).

-1. Latitude can be determined by using a compass to find the angle of the Sun at sunrise or sunset (subtracting or adding magnetic variation) and the date. According to the chart in figure 20-79, on January 26th, the azimuth of the rising Sun will be 120° to the left when facing the Sun in the Northern Hemisphere (it would be 120° to the right for setting Sun); therefore, the latitude would be 50°. If in the Southern Hemisphere, the direction of the Sun would be the opposite.

-2. The table does not list every day of the year, nor does it list every degree of latitude. If accuracy is desired within 1° of azimuth, interpolation may be nec-

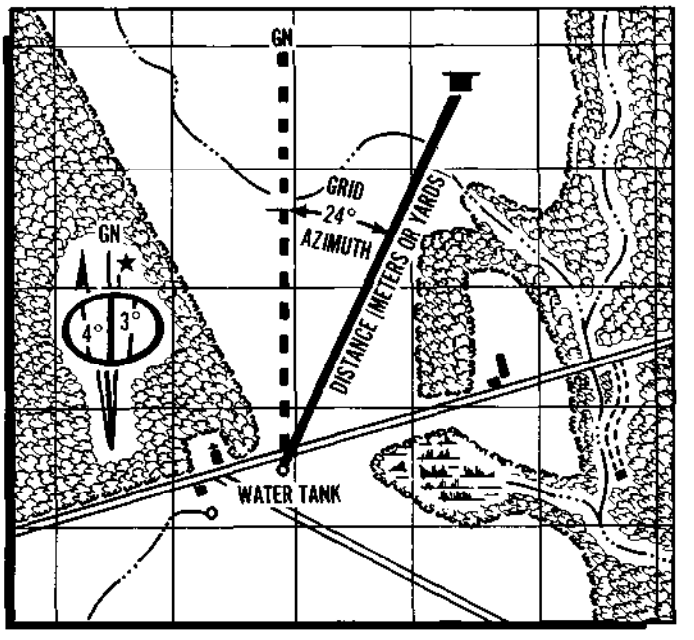

Figure 20-78. Polar Coordinates Used to Designate Position on Map.

DATE		Angle to North from the rising or setting Sun (level terrain) LATITUDE												
		0°	5°	10°	15°	20°	25°	30°	35°	40°	45°	50°	55°	60°
JANUARY	1	113	113	113	114	115	116	117	118	121	124	127	133	141
	6	112	113	113	113	114	115	116	118	120	123	127	132	140
	11	112	112	112	113	113	114	115	117	119	122	125	130	138
	16	111	111	111	112	112	113	114	116	118	120	124	129	136
	21	110	110	110	111	111	112	113	115	117	119	122	127	133
	26	109	109	109	109	110	111	112	113	115	117	120	124	130
FEBRUARY	1	107	107	108	108	108	109	110	111	113	115	117	121	126
	6	106	106	106	106	107	107	108	109	111	113	115	118	123
	11	104	104	105	105	105	106	107	108	109	110	112	116	120
	16	103	103	103	103	103	104	105	106	107	108	110	112	116
	21	101	101	101	101	101	102	102	103	104	105	107	109	112
	26	99	99	99	99	100	100	100	101	102	103	104	106	108
MARCH	1	98	98	98	98	99	99	99	100	100	101	102	104	106
	6	96	96	96	96	96	97	97	97	98	98	99	100	102
	11	94	94	94	94	94	94	95	95	95	96	96	97	98
	16	92	92	92	92	92	92	92	92	93	93	93	93	94
	21	90	90	90	90	90	90	90	90	90	90	90	90	90
	26	88	88	88	88	88	88	88	88	87	87	87	87	86
APRIL	1	86	86	86	86	85	85	85	85	84	84	83	82	81
	6	84	84	84	83	83	83	83	82	82	81	80	79	77
	11	82	82	82	82	81	81	81	80	80	79	77	76	74
	16	80	80	80	80	79	79	78	78	77	76	74	72	70
	21	78	78	78	78	78	77	76	76	75	73	72	69	66
	26	77	77	76	76	76	75	75	74	72	71	69	66	63
MAY	1	75	75	75	74	74	73	73	72	70	69	66	63	59
	6	74	74	73	73	73	72	71	70	68	67	64	61	56
	11	72	72	72	72	71	70	69	68	67	64	62	58	52
	16	71	71	71	70	70	69	68	67	65	63	60	55	49
	21	70	70	70	69	69	68	67	65	63	61	58	53	47
	26	69	69	69	68	68	67	66	64	62	60	56	51	44
JUNE	1	68	68	68	67	66	66	64	63	61	58	54	49	41
	6	67	67	67	67	66	65	64	62	60	57	53	48	40
	11	67	67	67	66	65	64	63	62	59	56	53	47	39
	16	67	67	67	66	65	64	63	62	59	56	53	47	39
	21	67	67	67	66	65	64	63	62	59	56	53	47	39
	26	67	67	67	66	65	64	63	62	59	56	53	47	39
JULY	1	67	67	67	66	65	64	63	62	59	56	53	47	39
	6	67	67	67	66	66	65	64	62	60	57	53	48	40
	11	68	68	68	67	66	65	64	63	61	58	54	49	41
	16	69	68	68	68	67	66	65	64	62	59	55	50	43
	21	69	69	69	69	68	67	66	65	63	60	57	52	45
	26	70	70	70	70	69	68	67	66	64	62	59	54	48
AUGUST	1	72	72	72	71	71	70	69	68	66	64	61	57	51
	6	73	73	73	73	72	71	71	69	68	66	63	60	55
	11	75	75	74	74	74	73	72	71	70	68	66	63	58
	16	76	76	76	76	75	75	74	73	72	70	68	65	61
	21	78	78	77	77	77	76	76	75	74	72	71	68	65
	26	79	79	79	79	79	78	78	77	76	75	73	71	68
SEPTEMBER	1	82	82	82	81	81	81	80	80	79	78	77	75	73
	6	83	83	83	83	83	83	82	82	81	81	80	78	77
	11	85	85	85	85	85	85	85	84	84	83	83	82	81
	16	87	87	87	87	87	87	87	86	86	86	85	85	84
	21	89	89	89	89	89	89	89	89	89	89	88	88	88
	26	91	91	91	91	91	91	91	91	91	91	92	92	92
OCTOBER	1	93	93	93	93	93	93	93	94	94	94	95	95	96
	6	95	95	95	95	95	96	96	96	97	97	98	99	100
	11	97	97	97	97	97	98	98	99	99	100	101	102	104
	16	99	99	99	99	99	100	100	101	101	102	104	105	108
	21	101	101	101	101	101	102	102	103	104	105	107	109	112
	26	102	102	103	103	103	104	104	105	106	108	109	112	115
NOVEMBER	1	104	104	105	105	105	106	107	108	109	110	113	116	120
	6	106	106	106	107	107	108	109	110	111	113	115	119	123
	11	107	107	108	108	108	109	110	111	113	115	117	121	126
	16	109	109	109	109	110	111	112	113	115	117	120	124	130
	21	110	110	110	111	111	112	113	114	116	119	122	126	133
	26	111	111	111	112	112	113	114	116	118	120	124	128	135
DECEMBER	1	112	112	112	113	113	114	115	117	119	122	125	130	138
	6	112	112	113	113	114	115	116	118	120	123	126	132	140
	11	113	113	113	114	115	116	117	118	121	124	127	133	141
	16	113	113	113	114	115	116	117	118	121	124	127	133	141
	21	113	113	113	114	115	116	117	118	121	124	127	133	141
	26	113	113	113	114	115	116	117	118	121	124	127	133	141

NOTE: When the Sun is rising, the angle is reckoned from East to North.
When the Sun is setting, the angle is reckoned from West to North.

Figure 20-79. Finding Direction from the Rising or Setting Sun.

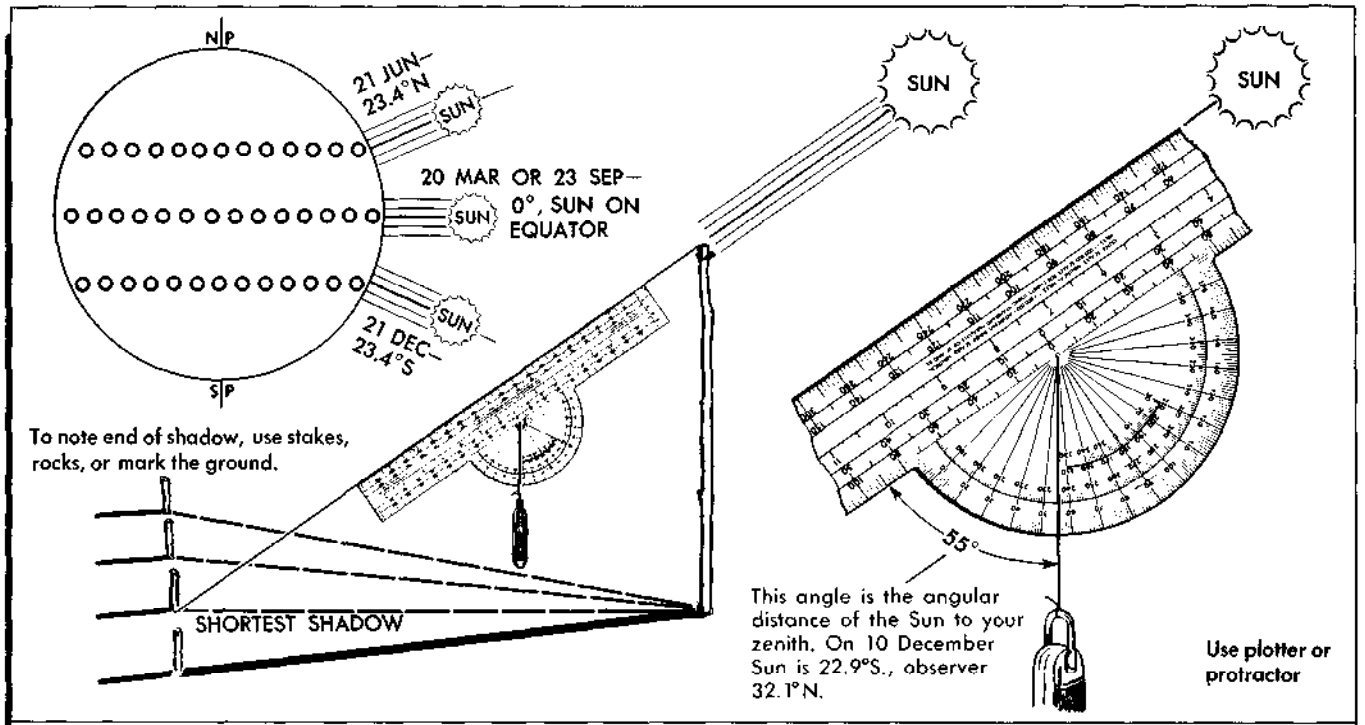

DECLINATION OF SUN

(IN DEGREES AND TENTHS OF A DEGREE)

Declination is tabulated to the nearest tenth of a degree rather than to the nearest minute of arc. To convert 1/10° (0.1°) to minutes, multiply by 6. (ie. 27.9°=27° 54')

DAY	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1	S 23.1	S 17.5	S 7.7	N 4.4	N 15.0	N 22.0	N 23.1	N 18.1	N 8.4	S 3.1	S 14.3	S 21.8
2	23.0	17.2	7.3	4.8	15.3	22.1	23.1	17.9	8.1	3.4	14.6	21.9
3	22.9	16.9	6.9	5.2	15.6	22.3	23.0	17.6	7.7	3.8	15.0	22.1
4	22.9	16.6	6.6	5.6	15.9	22.4	22.9	17.3	7.3	4.2	15.3	22.2
5	22.8	16.3	6.2	5.9	16.2	22.5	22.8	17.1	7.0	4.6	15.6	22.3
6	S 22.7	S 16.0	S 5.8	N 6.3	N 16.4	N 22.6	N 22.7	N 16.8	N 6.6	S 5.0	S 15.9	S 22.5
7	22.5	15.7	5.4	6.7	16.7	22.7	22.6	16.5	6.2	5.4	16.2	22.6
8	22.4	15.4	5.0	7.1	17.0	22.8	22.5	16.3	5.8	5.7	16.5	22.7
9	22.3	15.1	4.6	7.4	17.3	22.9	22.4	16.0	5.5	6.1	16.8	22.8
10	22.2	14.8	4.2	7.8	17.5	23.0	22.3	15.7	5.1	6.5	17.1	22.9
11	S 22.0	S 14.5	S 3.8	N 8.2	N 17.8	N 23.1	N 22.2	N 15.4	N 4.7	S 6.9	S 17.3	S 23.0
12	21.9	14.1	3.5	8.6	18.0	23.1	22.0	15.1	4.3	7.3	17.6	23.1
13	21.7	13.8	3.1	8.9	18.3	23.2	21.9	14.8	3.9	7.6	17.9	23.1
14	21.5	13.5	2.7	9.3	18.5	23.2	21.7	14.5	3.6	8.0	18.1	23.2
15	21.4	13.1	2.3	9.6	18.8	23.3	21.6	14.2	3.2	8.4	18.4	23.3
16	S 21.2	S 12.8	S 1.9	N 10.0	N 19.0	N 23.3	N 21.4	N 13.9	N 2.8	S 8.8	S 18.7	S 23.3
17	21.0	12.4	1.5	10.4	19.2	23.4	21.3	13.5	2.4	9.1	18.9	23.3
18	20.8	12.1	1.1	10.7	19.5	23.4	21.1	13.2	2.0	9.5	19.1	23.4
19	20.6	11.7	0.7	11.1	19.7	23.4	20.9	12.9	1.6	9.9	19.4	23.4
20	20.4	11.4	0.3	11.4	19.9	23.4	20.7	12.6	1.2	10.2	19.6	23.4
21	S 20.2	S 11.0	N 0.1	N 11.7	N 20.1	N 23.4	N 20.5	N 12.2	N 0.8	S 10.6	S 19.8	S 23.4
22	20.0	10.7	0.5	12.1	20.3	23.4	20.4	11.9	0.5	10.9	20.1	23.4
23	19.8	10.3	0.9	12.4	20.5	23.4	20.2	11.6	N 0.1	11.3	20.3	23.4
24	19.5	9.9	1.3	12.7	20.7	23.4	20.0	11.2	S 0.3	11.6	20.5	23.4
25	19.3	9.6	1.7	13.1	20.9	23.4	19.7	10.9	0.7	12.0	20.7	23.4
26	S 19.0	S 9.2	N 2.1	N 13.4	N 21.1	N 23.4	N 19.5	N 10.5	S 1.1	S 12.3	S 20.9	S 23.4
27	18.8	8.8	2.5	13.7	21.2	23.3	19.3	10.2	1.5	12.7	21.1	23.3
28	18.5	8.5	2.9	14.0	21.4	23.3	19.1	9.8	1.9	13.0	21.3	23.3
29	18.3	8.1	3.2	14.4	21.6	23.3	18.8	9.5	2.3	13.3	21.4	23.3
30	18.0	...	3.6	14.7	21.7	23.2	18.6	9.1	2.7	13.7	21.6	23.2
31	S 17.7	...	N 4.0	...	N 21.9	...	N 18.4	N 8.8	...	S 14.0	...	S 23.1

EXAMPLE: On 10 December the declination of the Sun is 22.9°S., so observers who measure the zenith distance as 0° would know that they are at latitude 22.9°S. If they measure a zenith distance of 5° with the Sun south of this zenith, they are 5° north of 22.9°S, or at a latitude 17.9°S; and if the Sun is north, they are 5° south of 22.9°S, or latitude 27.9°S.

Figure 20-80. Determining Latitude by Noon Sun.

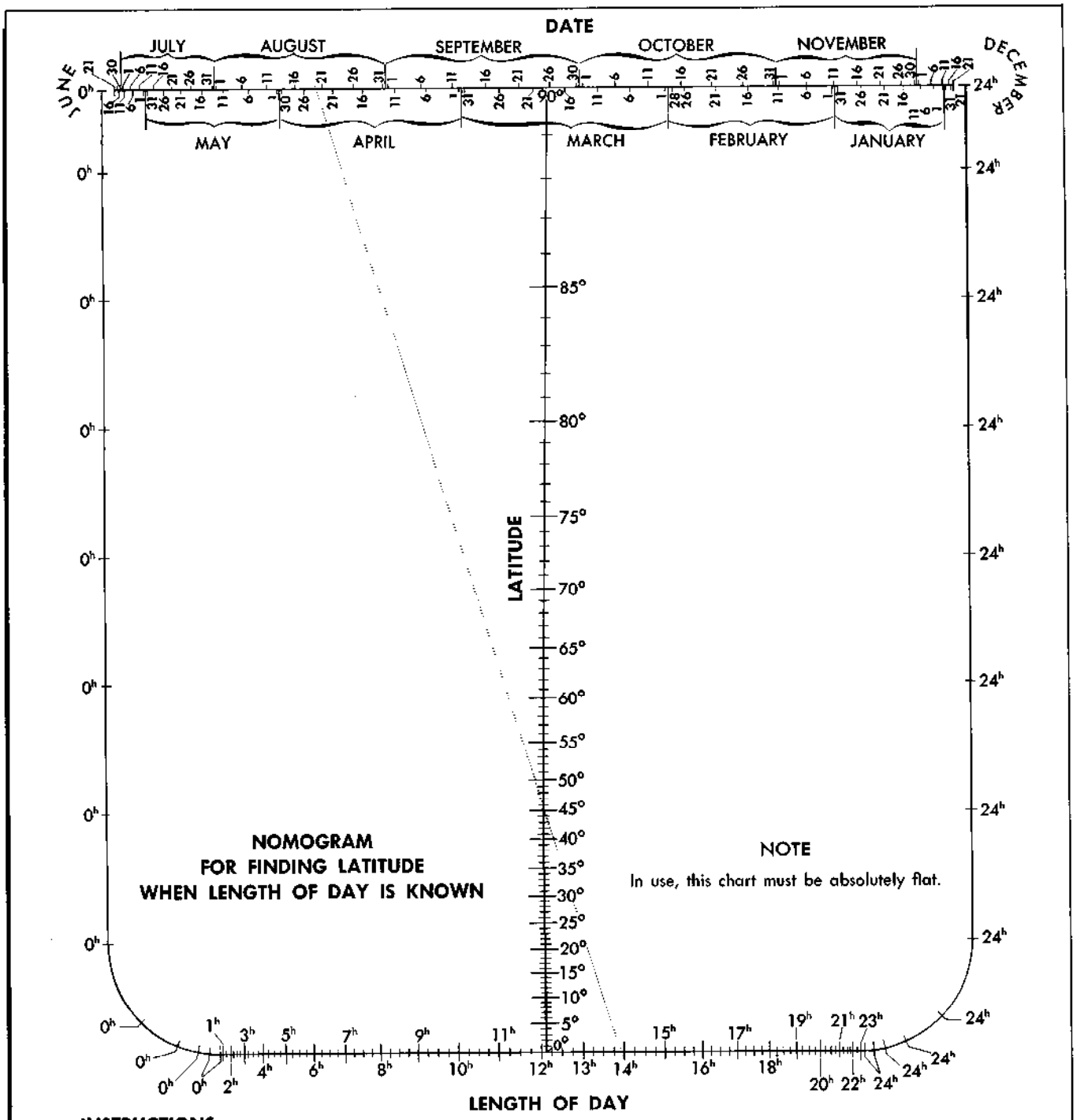

INSTRUCTIONS

To find your latitude:

In Northern Latitudes:

1. Find length of the day from the instant the top of the Sun appears above the ocean horizon to the instant it disappears below the horizon. This instant is often marked by a green flash.
2. Lay a straight edge or stretch across the Nomogram, connecting the observed length of the day on the Length of Day Scale with the date on the Date Scale.
3. Read your latitude on the Latitude Scale.

EXAMPLE: On August 20, observed length of the day is 13 hours and 54 minutes. Latitude is 45°30′N.

In Southern Latitudes:

Add 6 months to the date and proceed as in northern latitudes.

EXAMPLE: On 11 May observed length of day is 10 hours and 4 minutes. Adding 6 months gives 11 November. Latitude is 41°30′S.

Figure 20-81. Nomogram.

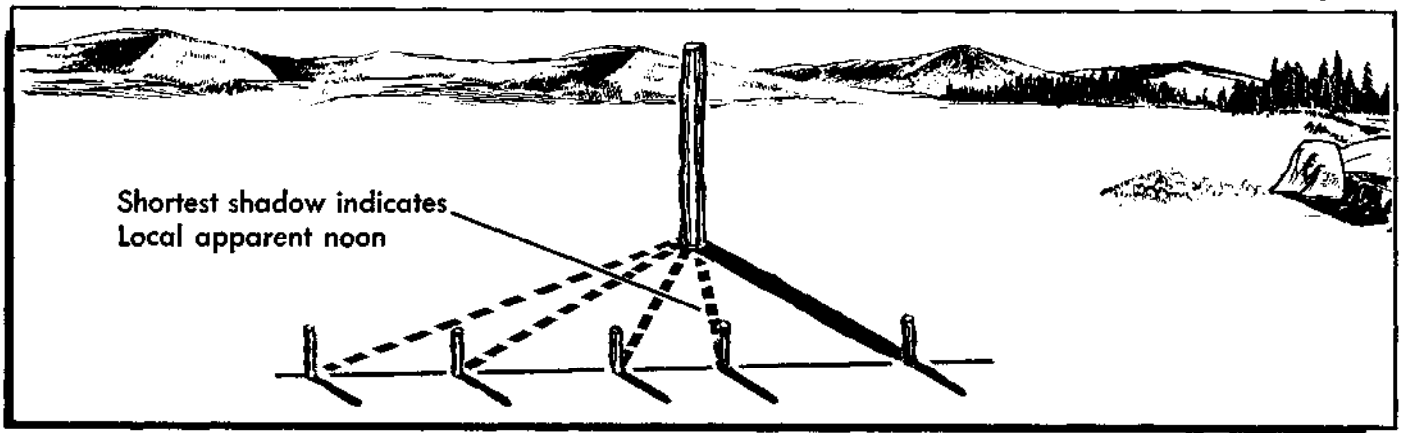

Figure 20-82. Stick and Shadow.

essary (split the difference) between the values given in the table. For example, between 45° latitude and 50° latitude is 5°. The difference in latitudes (5°) and the difference in azimuths (3°) split (5/3) is 1°⅔′(1°40′), so the more accurate degree of latitude would be 46°40′ latitude.

(b) Latitude by Noon Altitude of the Sun. On any given day, there is only one latitude on Earth where the Sun passes directly overhead or through the zenith at noon. In all latitudes north of this, the Sun passes to the south of the zenith; and in those south of it, the Sun passes to the north. For each 1° change of latitude, the zenith distance also changes by 1 degree. Figure 20-80 gives the latitude for each day of the year where the Sun is in the zenith at noon. If a Weems plotter or other protractor is available, maximum altitude of the Sun should be used to find latitude by measuring the angular distance of the Sun from the zenith at noon. Local noon can be found using the methods described earlier. Stretch a string from the top of a stick to the point where the end of the noon shadow rested, place the plotter along the string and drop a plumb line from the center of the plotter. The intersection of the plumb line with the outer scale of the plotter shows the angular distance of the Sun from the zenith.

(c) Latitudes by Length of Day. This method is used most effectively while on open seas. When in any latitude between 60°N and 60°S, the exact latitude within 30 nautical miles (½°) can be determined if the length of the day within 1 minute is known. This is true throughout the year except for about 10 days before and 10 after the equinoxes—approximately 11-31 March and 12 September-2 October. During these two periods, the day is about the same length at all latitudes. A level horizon is required to time sunrise and sunset accurately. Find the length of day from the instant the top of the Sun first appears above the ocean horizon to the instant it disappears below the horizon. This instant is often marked by a green flash. Write down the times of sunrise and sunset. Don't count on remembering them.

Date	Eq. of Time*	Date	Eq. of Time*	Date	Eq. of Time*	Date	Eq. of Time*	Date	Eq. of Time*	Date	Eq. of Time*
Jan. 1	−3.5 min.	Mar. 4	−12.0	May 2	+3.0 min.	Aug. 4	−6.0	Oct. 1	+10.0 min.	Dec. 1	+11.0
2	−4.0	8	−11.0	14	+3.8	12	−5.0	4	+11.0	4	+10.0
4	−5.0	12	−10.0	May 28	+3.0	17	−4.0	7	+12.0	6	+9.0
7	−6.0	16	−9.0			22	−3.0	11	+13.0	9	+8.0
9	−7.0	19	−8.0	June 4	+2.0	26	−2.0	15	+14.0	11	+7.0
12	−8.0	22	−7.0	9	+1.0	Aug. 29	−1.0	20	+15.0	13	+6.0
14	−9.0	26	−6.0	14	0.0	Sept. 1	0.0	Oct. 27	+16.0	15	+5.0
17	−10.0	Mar. 29	−5.0	19	−1.0	5	+1.0			17	+4.0
20	−11.0			23	−2.0	8	+2.0			19	+3.0
24	−12.0	Apr. 1	−4.0	June 28	−3.0	10	+3.0	Nov. 4	+16.4	21	+2.0
Jan. 28	−13.0	5	−3.0			13	+4.0	11	+16.0	23	+1.0
		8	−2.0	July 3	−4.0	16	+5.0	17	+15.0	25	0.0
Feb. 4	−14.0	12	−1.0	9	−5.0	19	+6.0	22	+14.0	27	−1.0
13	−14.3	16	0.0	18	−6.0	22	+7.0	25	+13.0	29	−2.0
19	−14.0	20	+1.0	July 27	−6.6	25	+8.0	Nov. 28	+12.0	Dec. 31	−3.0
Feb. 28	−13.0	Apr. 25	+2.0			Sep. 28	+9.0				

* Add plus values to mean time and subtract minus values from mean time to get apparent time.

Figure 20-83. Equation of Time.

Note that only the length of day counts in the determination of latitude; a watch may have an unknown error and yet serve to determine this factor. If only one water horizon is available, as on a seacoast, find local noon by the stick and shadow method. The length of day is twice the interval from sunrise to noon or from noon to sunset. Knowing the length of day, latitude can be found by using the nomogram shown in figure 20-81.

(d) Longitude from Local Apparent Noon. To find longitude, a survivor must know the correct time and the rate at which a watch gains or loses time. If this rate and the time the watch was last set is known, the correct time can be computed by adding or subtracting the gain or loss. Correct the zone time on the watch to Greenwich time; for example, if the watch is on eastern standard time, add 5 hours to get Greenwich time. Longitude can be determined by timing the moment a celestial body passes the meridian. The easiest body to use is the Sun. Use one of the following methods:

-1. Stick and Shadow. Put up a stick or rod (figure 20-82) as nearly vertical as possible in a level place. Check the alignment of the stick by sighting along the line of a makeshift plumb bob. (To make a plumb bob, tie any heavy object to a string and let it hang free. The line of the string indicates the vertical.) Sometime before midday, begin marking the position of the end of the stick's shadow. Note the time for each mark. Continue marking until the shadow definitely lengthens. The time of the shortest shadow is the time when the Sun passed the local meridian or local apparent noon. A survivor will probably have to estimate the position of the shortest shadow by finding a line midway between two shadows of equal length, one before noon and one after. If the times of sunrise and sunset are accurately determined on a water horizon, local noon will be midway between these times.

-2. Double Plumb Bob. Erect two plumb bobs about 1 foot apart so that both strings line up on Polaris, much the same as a gun sight. Plumb bobs should be set up when Polaris is on the meridian and has no east-west correction. The next day, when the shadows of the two plumb bobs coincide, they will indicate local apparent noon.

-3. Mark Down the Greenwich Time of Local Apparent Noon. The next step is to correct this observed time of meridian passage for the equation of time; that is, the number of minutes the real Sun is ahead of or behind the mean Sun. (The mean Sun was invented by astronomers to simplify the problems of measuring time. Mean Sun rolls along the Equator at a constant rate of 15° per hour. The real Sun is not so considerate; it changes its angular rate of travel around the Earth with the seasons.) Figure 20-83 gives the value in minutes of time to be added to or subtracted from mean (watch) time to get apparent (Sun) time.

-4. After computing the Greenwich time of local noon, the difference of longitude between the survivor's position and Greenwich can be found by converting the interval between 1200 Greenwich and the local noon from time to arc. Remember that 1 hour equals 15° of longitude, 4 minutes equal 1° of longitude, and 4 seconds equal 1′ of longitude. Example: The

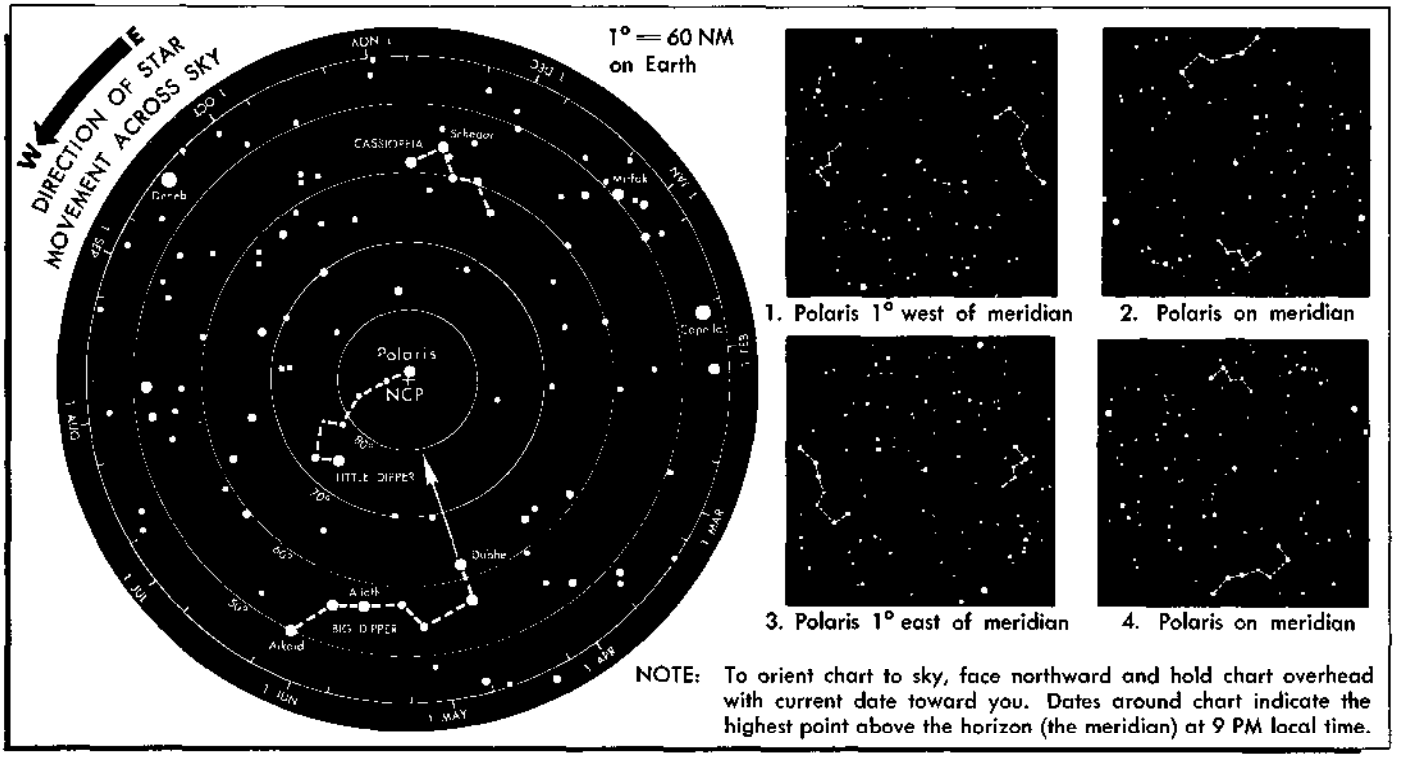

Figure 20-84. Finding Direction from Polaris.

survivor's watch is on eastern standard time, and it normally loses 30 seconds a day. It hasn't been set for 4 days. The local noon is timed at 15:08 on the watch on 4 February. Watch correction is 4 X 30 seconds, or plus 2 minutes. Zone time correction is plus 5 hours. Greenwich time is 15:08 plus 2 minutes plus 5 hours or 20:10. The equation of time for 4 February is minus 14 minutes. Local noon is 20:10 minus 14 minutes or 19:56 Greenwich. The difference in time between Greenwich and present position is 19:56 minus 12:00 or 7:56. A time of 7:56 equals 119°of longitude. Since local noon is later than Greenwich noon, the survivor is west of Greenwich, longitude is 119°W.

(e) Direction and Position Finding at Night:

-1. Direction from Polaris. In the Northern Hemisphere, one star, Polaris (the Pole Star), is never more than approximately 1° from the North Celestial Pole. In other words, the line from any observer in the Northern Hemisphere to the Pole Star is never more than 1° away from true north. Find the Pole Star by locating the Big Dipper or Cassiopeia, two groups of stars which are very close to the North Celestial Pole. The two stars on the outer edge of the Big Dipper are called pointers because they point almost directly to Polaris. If the pointers are obscured by clouds, Polaris can be identified by its relationship to the constellation Cassiopeia. Figure 20-84 indicates the relation between the Big Dipper, Polaris, and Cassiopeia.

-2. Direction from the Southern Cross. In the Southern Hemisphere, Polaris is not visible. There the Southern Cross is the most distinctive constellation. When flying south, the Southern Cross appears shortly before Polaris drops from sight astern. An imaginary line through the long axis of the Southern Cross, or True Cross, points toward the South Pole. The True Cross should not be confused with a larger cross nearby known as the False Cross, which is less bright and more widely spaced. Two of the four stars in the True Cross are among the brightest stars in the heavens; they are the stars on the southern and eastern arms. Those of the northern and western arms are not as conspicuous but are bright.

-3. There is no conspicuous star above the South Pole to correspond to Polaris above the North Pole. In fact, the point where such a star would be, if one existed, lies in a region devoid of stars. This point is so dark in comparison with the rest of the sky that it is known as the Coalsack. Figure 20-85 shows the True Cross and—to the west of it—the False Cross.

(f) Finding Due East and West by Equatorial Stars. Due to the altitude of Polaris above the horizon, it may sometimes be difficult to use as a bearing. To use a point directly on the horizon may be more convenient.

-1. The celestial equator, which is a projection of the Earth's equator onto the imaginary celestial sphere, always intersects the horizon line at the due east and west points of the compass. Therefore, any star on the celestial equator rises due east and sets due west (disallowing a small error because of atmospheric refraction). This holds true for all latitudes except those of the North and South Poles, where the celestial equator

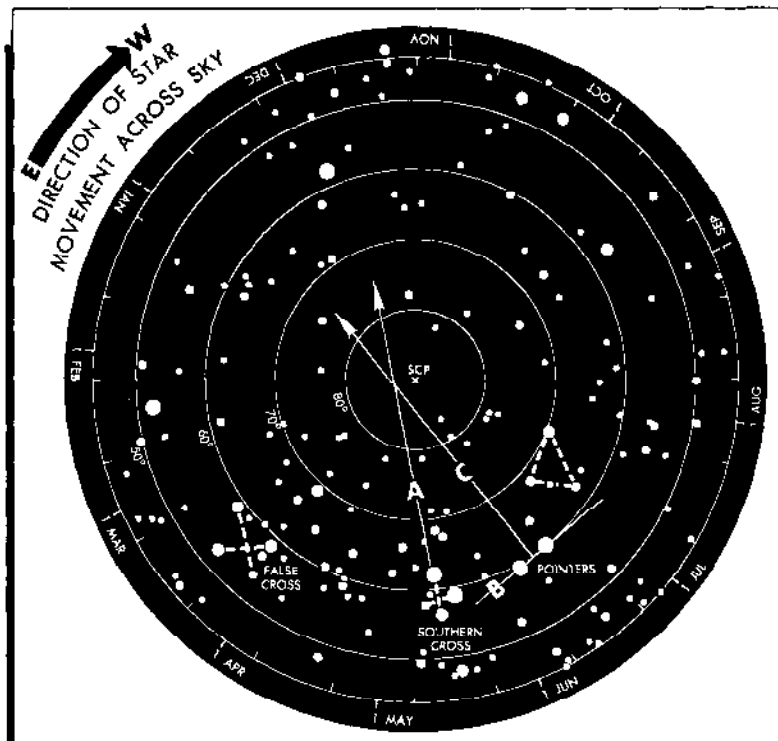

TO LOCATE THE SOUTH CELESTIAL POLE:

1. Extend an imaginary line (A) along the long axis of the True Cross to the south.
2. Join the two bright stars to the east of the Cross with an imaginary line (B). Bisect this line with one at right angles (C) and let it extend southward to intersect line (A).
3. The intersection of line (C) with the line through the Cross (A) is a few degrees from the South Celestial Pole (approximately 5 or 6 full moon widths).

NOTE: To orient chart to sky, face southward and hold chart overhead with current date toward you. Dates around chart indicate the highest point above the horizon (the meridian) at 9 PM local time.

Figure 20-85. Finding Direction from Southern Cross.

and the horizon have a common plane. However, if a survivor is at the North or South Pole, it will probably be known, so this technique can be assumed to be of universal use.

-2. Certain difficulties arise in the practical use of this technique. Unless a survivor is quite familiar with the constellations, it may be difficult to spot a specific rising star as it first appears above the eastern horizon. It will probably be simpler to depend upon the identification of an equatorial star before it sets in the west.

-3. Another problem is caused by atmospheric extinction. As stars near the horizon, they grow fainter in brightness because the line of sight between the observer's eyes and the star passes through a constantly thickening atmosphere. Therefore, faint stars disappear from view before they actually set. However, a fairly accurate estimate of the setting point of a star can be made some time before it actally sets. The atmospheric conditions of the area have a great effect on obstructing a star's light as it sets. Atmospheric haze, for example, is much less a problem on deserts than along temperate zone coastal strips.

-4. Figure 20-86 shows the brighter stars and some prominent star groups which lie along the celestial equator. There are few bright stars actually on the celestial equator. However, there are a number of stars which lie quite near it, so an approximation within a degree or so can be made. Also, a rough knowledge of the more conspicuous equatorial constellations will give the survivor a continuing checkpoint for maintaining orientation.

(g) Finding Latitude from Polaris. A survivor can find the latitude in the Northern Hemisphere north of 10°N by measuring the angular altitude of Polaris above the horizon, as shown in figure 20-87.

(h) Finding Direction (North) from Overhead Stars that are not in the General Location of the Celestial Poles:

-1. At times, survivors may not be able to locate Polaris (the North Star) due to partial cloud cover, or its position below the observer's horizon. In this situation, it would seem that survivors would be unable to locate direction. Fortunately, survivors who wish to initially find direction or who wish to check a course of

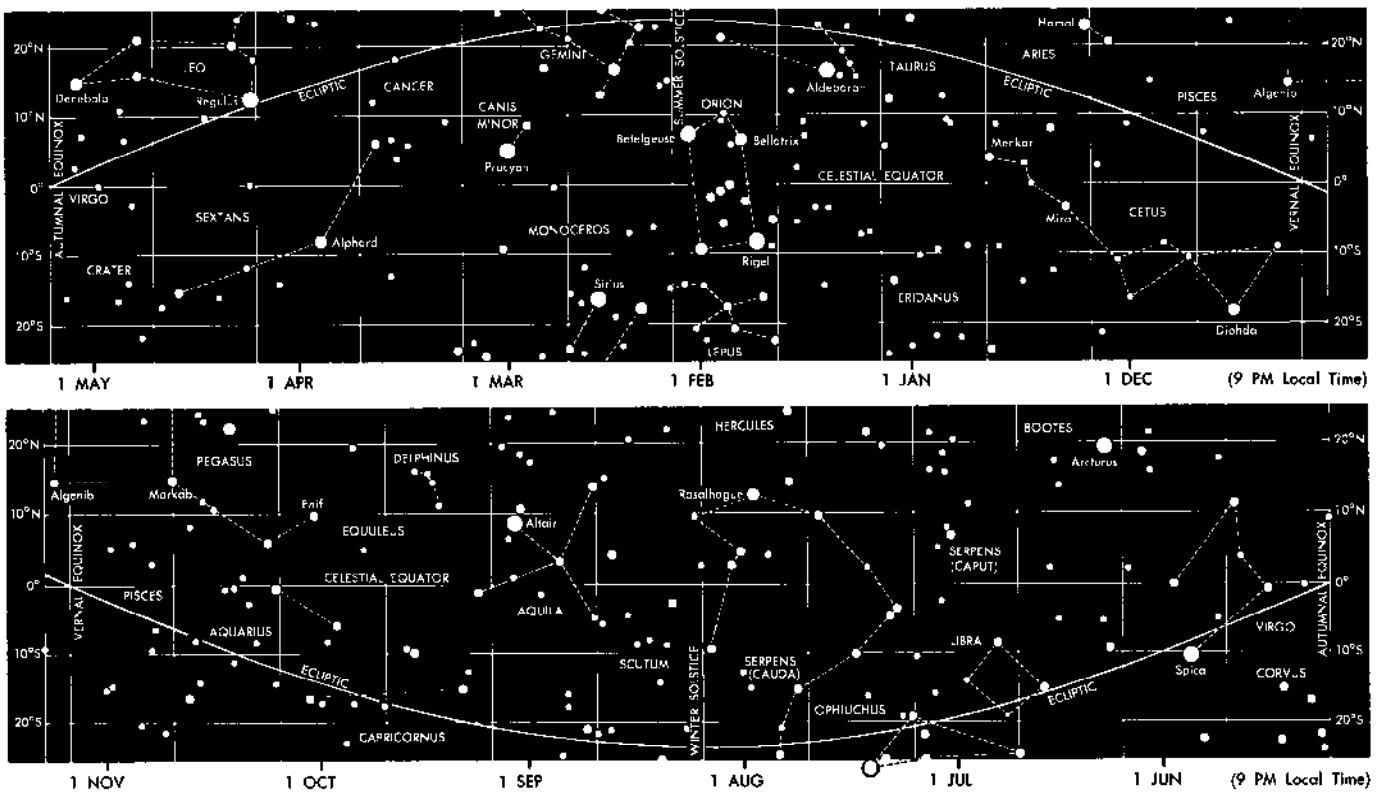

The star charts on this page show only those star-groups which lie near the celestial equator. Each map covers a strip approximately 50° wide (north and south) and 180° wide (east and west). Most of the more conspicuous naked eye stars are shown, the brighter stars being indicated by larger dots. Some star-groups have been joined by dashed lines to aid in identification.

Below each chart will be found the months of the year. On the first of each month the stars shown above the month name will be on the celestial meridian (highest point of their path above the horizon) at 9 p.m. local time.

The celestial equator makes a 45° angle with the southern horizon at latitude 45°N. At 60°N latitude it makes a 30° angle. In other words, degrees of latitude subtracted from 90° gives the angle formed by the celestial equator and the southern horizon.

The preceding information should enable the survivor to identify the equatorial stars to be used in determining due east or west. When observing stars rising in the east, the charts should be slanted down and to the left; when observing stars setting in the west, they should be slanted down and to the right.

Figure 20-86. Charts of Equatorial Stars.

HOW TO USE WEEMS PLOTTER TO MEASURE ALTITUDE OF POLARIS

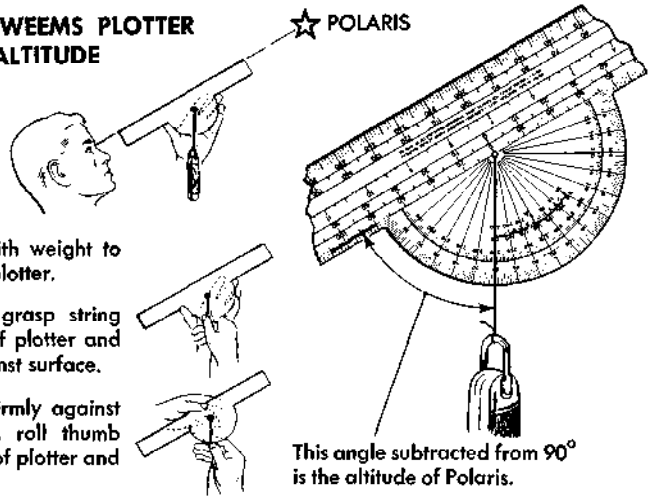

1. Attach string with weight to center hole of plotter.
2. After sighting, grasp string at outer edge of plotter and hold firmly against surface.
3. Holding string firmly against face of plotter, roll thumb under the edge of plotter and take reading.

HOW TO USE ANY PROTRACTOR TO MEASURE ALTITUDE OF POLARIS

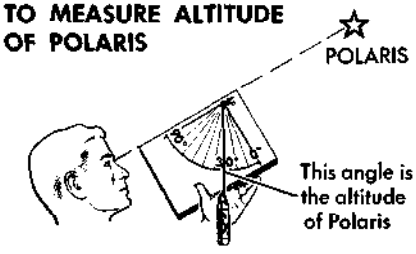

Any protractor (such as the one printed in AFM 64-5) can be used to find latitude.

1. Attach string with weight to Point "P".
2. After sighting, grasp string at edge of protractor.
3. Hold string in this position and take reading. This gives the altitude of Polaris.

CORRECTION FOR OBSERVED ALTITUDE OF POLARIS

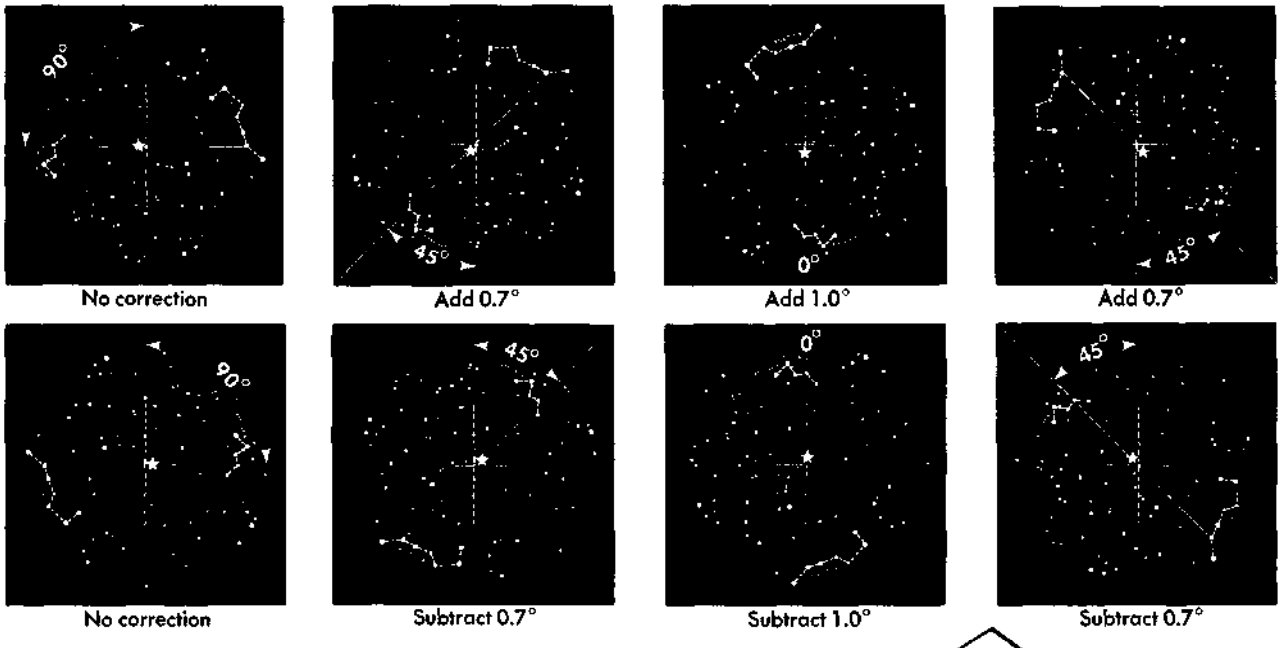

The star diagrams above are drawn for angles of 0°, 45°, and 90° between the vertical dotted line thru the pole and the line thru Cassiopeia and the Big Dipper (Ursa Major). For intermediate positions the angle may be estimated and the correction taken from the Correction Table. Subtract corrections given in the table when Polaris is above the horizontal line thru the pole and add corrections when Polaris is below this line.

Note that the correction changes very slowly near the time when the correction is greatest and hence an error in estimation of the position has little effect at this time.

CORRECTION TABLE

ANGLE	CORRECTION	ANGLE	CORRECTION
0°	1.0°	50°	0.6°
10°	1.0°	60°	0.5°
20°	0.9°	70°	0.3°
30°	0.9°	80°	0.2°
40°	0.8°	90°	No correction

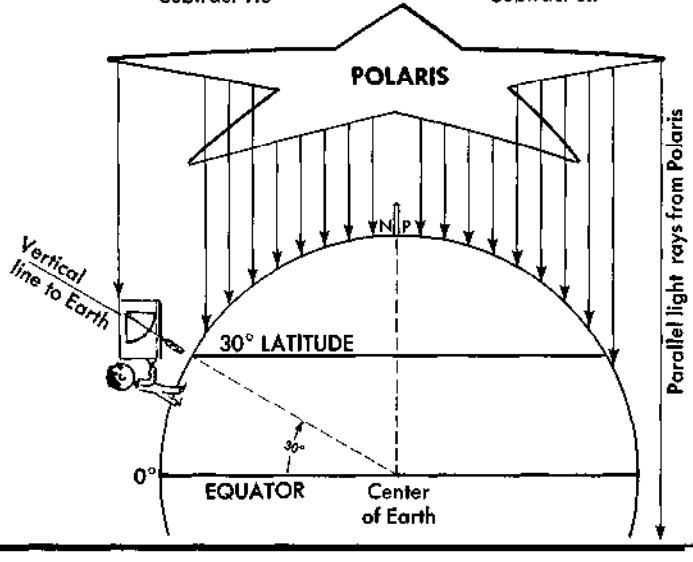

Figure 20-87. Finding Latitude by Polaris.

travel during the night need not worry about being lost or unable to travel if Polaris cannot be identified.

-2. The following is an adaptation of the stick and shadow method of direction finding. This method is based on the principles that all the heavenly bodies (Sun, Moon, planets, and stars) rise (generally) in the east and set (generally) in the west. This technique can be used anywhere on Earth with any stars except those which are circumpolar. Circumpolar stars are those which appear to travel around Polaris instead of apparently "moving" from east to west.

-3. To use this technique, survivors should keep in mind that they may use *any star other than a circumpolar one.*

-4. Survivors who wish to know general direction must prepare a device to aid them. This can be done by placing a stick (about 5 feet in length) at a slight angle in the ground in an open area (figure 20-88). Thin material (suspension line, string, vine, braided cloth, etc.) is then attached to the tip of the stick. This material should be longer than what is required to reach the ground (figure 20-88).

-5. The survivor should lie on the back with the head next to this hanging line and pull the cord up to the temple area and hold it tautly.

-6. Next, the survivor moves around on the ground until the taut line is pointing directly at the selected, bright, noncircumpolar star (or planet).

-7. The taut line is now in position to simulate the star's (or planet's) shadow. Survivors should remember that this method of finding direction is an adaptation of the Sun, stick, and shadow approach. Here the more distant stars and (or) planets take the place of our Sun. Since these objects are too distant from the Earth to create shadow, the string represents the shadow.

-8. With the taut line simulating the star's shadow, survivors should mark the point on the ground where the line touches with a stick, stone, etc. The survivor should repeat this sighting on the same star (or planet) after about 15 to 20 minutes (marking the spot at which the line "shadow" touches the ground). A line scribed on the ground which connects these two points will run west-east (as the stars and planets move from east to west, the "shadow" will move in the opposite direction). The first mark will be in the west. Drawing a line perpendicular to the west-east line, a survivor will have a north-south line and be able to travel.

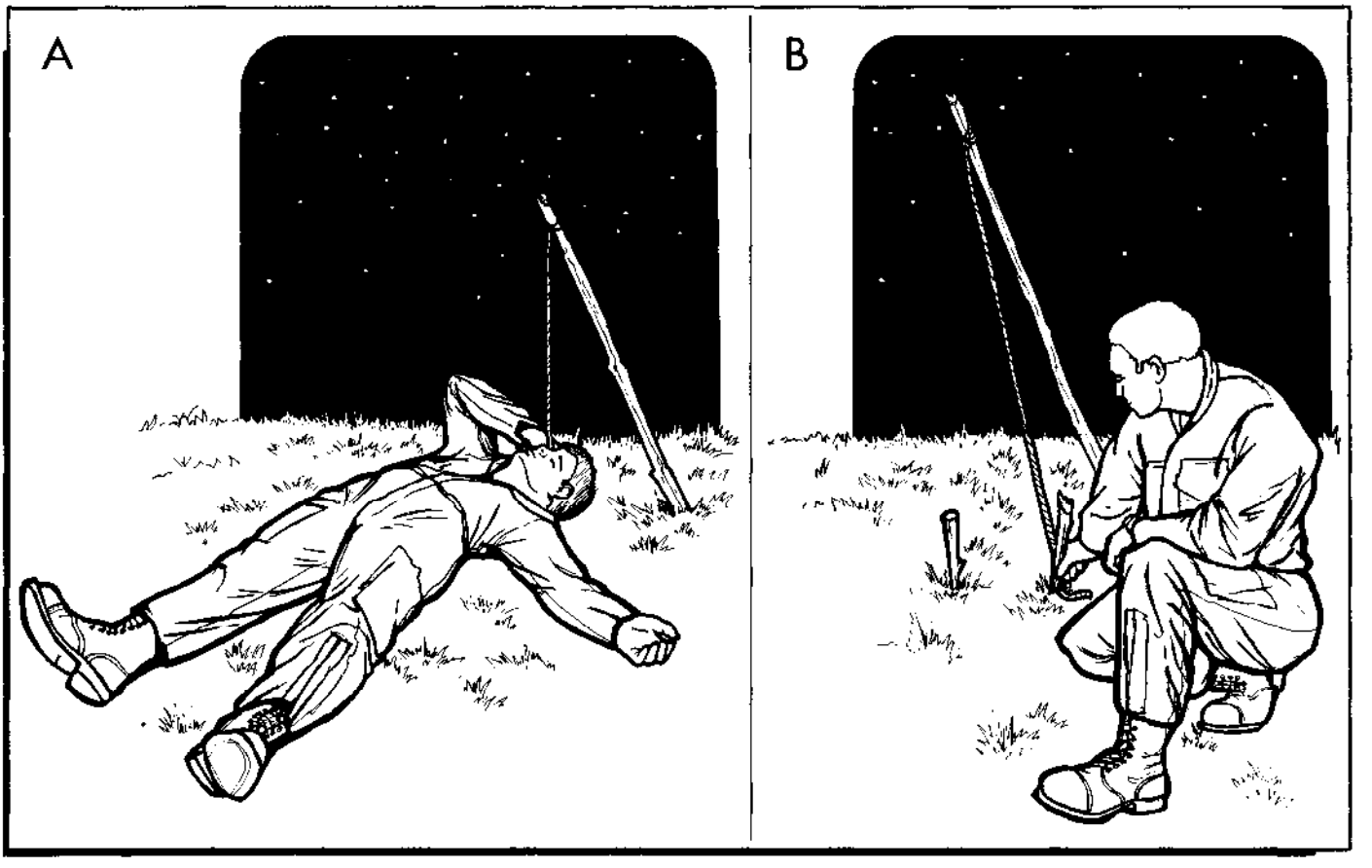

Figure 20-88. Stick and String Direction Finding.

Chapter 21

LAND TRAVEL

21-1. Introduction:

a. In any survival situation following an aircraft emergency, a decision must be made to either move or remain as close as possible to the parachute or crash site. In this chapter, land travel will be discussed and the various considerations that survivors should address before determining if travel is or is not a necessity.

b. Survivors may need to carry supplies and equipment while traveling to sustain life. For this reason, the techniques of backpacking and improvised packing are discussed to help a person do this task.

c. As a survivor, the ability to walk effectively is important in conserving energy and safety. Additionally, in rough terrain, travel may need to be done with the aid of a rope. The techniques of ascending and descending steep terrain are fundamental to understanding and performing rescue from rough terrain. These techniques, as well as techniques for snow travel, are covered. Travel may not be easy, but a knowledgeable traveler can travel safely and effectively while saving time and energy.

21-2. Decision to Stay or Travel. In hostile areas, the decision to travel is normally automatic. To stay in the vicinity of the crash or parachute landing can lead to capture. In friendly areas, a choice exists. The best advice is to stay with the aircraft. Most rescues have been made when downed aircrews remained with the aircraft.

a. Survivors should only leave the area when they are certain of their location and know that water, shelter, food, and help can be reached, or after having waited several days, they are convinced that rescue is not coming and they are equipped to travel.

b. Before making any decision, survivors should consider their personal physical condition and the condition of others in the party when estimating their ability to sustain travel. If people are injured, they should try to get help. If travel for help is required, they should send the people who are in the best physical and mental condition. Send two people if possible. To travel alone is dangerous. Before any decision is made, survivors should consider all of the facts.

(1) If the decision is to stay, these problems should be considered:

(a) Environmental conditions.

(b) Health and body care; camp sanitation.

(c) Rest and shelter.

(d) Water supplies.

(e) Food.

(2) If the decision is to travel: In addition to the primary survival problems of providing food, water, and shelter, the following must be considered:

(a) Direction of travel and why.

(b) Travel plan.

(c) Equipment required.

c. Before departing the site, survivors should leave information at their aircraft (nontactical situation only) stating departure time, destination, route of travel, personal condition, and available supplies.

d. From the air, it is easier to spot the aircraft than it is to spot people traveling on the ground. Someone may have seen the aircraft crash and investigate. The aircraft or parts from it can provide shelter, signaling aids, and other equipment (cowling for reflecting signals, tubing for shelter framework, gasoline and oil for fires, etc.). Avoiding the hazards and difficulties of travel is another reason to stay with the aircraft. Rescue chances are good if survivors made radio contact, landing was made on course or near a traveled air route, and weather and air observation conditions are good.

e. Present location must be known to decide intelligently whether to wait for rescue or to determine a destination and route of travel. The survivors should try to locate their position by studying maps, landmarks, and flight data, or by taking celestial observations. Downed personnel should try to determine the nearest rescue point, the distance to it, the possible difficulties and hazards of travel, and the probable facilities and supplies en route and at the destination.

f. There are a number of other factors that should be considered when deciding to travel.

(1) The equipment and materials required for cross-country travel should be analyzed. Travel is extremely risky unless the necessities of survival are available to provide support during travel. Survivors should have sufficient water to reach the next probable water source indicated on a map or chart and enough food to last until they can procure additional food. To leave shelter to travel in adverse weather conditions is foolhardy unless in an escape and evasion situation.

(2) In addition to the basic requirements, the physical condition of the survivor must be considered in any decision to travel. If in good condition, the survivor should be able to move an appreciable distance, but if the survivor is not in good condition or is injured, the ability to travel extended distances may be reduced. Analyze all injuries received during the emergency. For example, if a leg or ankle injury occurred during landing, this must be considered before traveling.

(3) If possible, survivors should avoid making any decision immediately after the emergency. They should wait a period of time to allow for recovery from the mental—if not the physical—shock resulting from the emergency. When shock has subsided survivors can then evaluate the situation, analyze the factors involved, and make valid decisions.

21-3. Travel. Once the survivors decide to travel, there are several considerations that apply regardless of the circumstances.

a. The ranking person must assume leadership, and the party must work as a team to ensure that all tasks are done in an equitable manner. Full use should be made of any survival experience or knowledge possessed by members of the group, and the leader is responsible for ensuring that the talents of all survivors are used.

b. Survivors should keep the body's energy output at a steady rate to reduce the effects of unaccustomed physical demands.

(1) A realistic pace should be maintained to save energy. It increases durability and keeps body temperature stable because it reduces the practice of quick starts and lengthy rests. More importantly, a moderate, realistic pace is essential in high altitudes in avoiding the risks of lapse of judgment and hallucinations due to lack of oxygen (hypoxia). Travel speed should provide for each survivor's physical condition and daily needs, and the group pace should be governed by the pace of the slowest group member. Additionally, rhythmic breathing should be practiced to prevent headache, nausea, lack of appetite, and irritability.

(2) Rest stops should be short since it requires added energy to begin again after cooling off. Survivors should wear their clothing in layers (layer system) and make adjustments to provide for climate, temperature, and precipitation. It is better to start with extra clothing and stop and shed a layer when beginning to warm up.

(3) Wearing loose clothing provides for air circulation, allows body moisture to evaporate, and retains body heat. Loose clothing also allows freedom of movement.

(4) Travelers should keep in mind when planning travel time and distance that the larger the group, the slower the progress will be. Time must be added for those survivors who must acclimate themselves to the climate, altitudes, and the task of backpacking. Survivors should also allow time for unexpected obstacles and problems which could occur.

(5) Proper nutrition and water are essential to building and preserving energy and strength. Several small meals a day are preferred to a couple of large ones so that calories and fluids are constantly available to keep the body and mind in the best possible condition. Survivors should try to have water and a snack available while trekking, and they should eat and drink often to restore energy and prevent chills in cold temperatures. This also applies at night.

21-4. Land Travel Techniques. Land travel techniques are based largely on experience, which is acquired through performance. However, experience can be partially replaced by the intelligent application of specialized practices that can be learned through instruction and observation. For example, travel routes may be established by observing the direction of a bird's flight, the actions of wild animals, the way a tree grows, or even the shape of a snowdrift. Bearings read from a compass, the Sun, or stars will improve on these observations and confirm original headings. All observations are influenced by the location and physical characteristics of the area where they are made and by the season of the year.

a. Route Finding. The novice should follow a compass line, whereas the experienced person follows lines of least resistance by realizing that a curved route may be faster and easier under certain circumstances. Use game trails when they follow a projected course only. For example, trails made by migrating caribou are frequently extensive and useful. On scree or rockslides, mountain sheep trails may be helpful. Game trails offer varying prospects, such as the chance of securing game or locating waterholes. Successful land travel requires knowledge beyond mere travel techniques. Survivors should have at least a general idea of the location of their starting place and their ultimate destination. They should also have knowledge of the people and terrain through which they will travel. If the population is hostile, they must adapt their entire method of travel and mode of living to this condition.

b. Wilderness. Wilderness travel requires constant awareness. A novice views a landscape from the top of a hill with care and interest, and says, "let's go." The experienced person carefully surveys the surrounding countryside. A distant blur may be mist or smoke; a faint, winding line on a far-off hill may be manmade or an animal trail; a blur in the lowlands may be a herd of caribou or cattle. People should plan travel only after carefully surveying the terrain. Study distant landmarks for characteristics that can be recognized from other locations or angles. Careful and intelligent observation will help survivors to correctly interpret the things they see, distant landmarks, or a broken twig at their feet. Before leaving a place, travelers should study their backtrail carefully. Survivors should know the route forward and backward. An error in route planning may make it necessary to backtrack in order to take a new course. For this reason, all trails should be marked (figure 21-1).

c. Mountain Ranges. Mountain ranges frequently affect the climate of a region and the climate in turn influences the vegetation, wildlife, and the character and number of people living in the region. For example, the oceanside of mountains has more fog, rain, and snow than the inland side of a range. Forests may grow on the oceanside, while inland, it may be semi-dry. Therefore, a complete change of survival techniques may be necessary when crossing a mountain range.

(1) Travel in mountainous country is simplified by conspicuous drainage landmarks, but it is complicated by the roughness of the terrain. A mountain traveler can

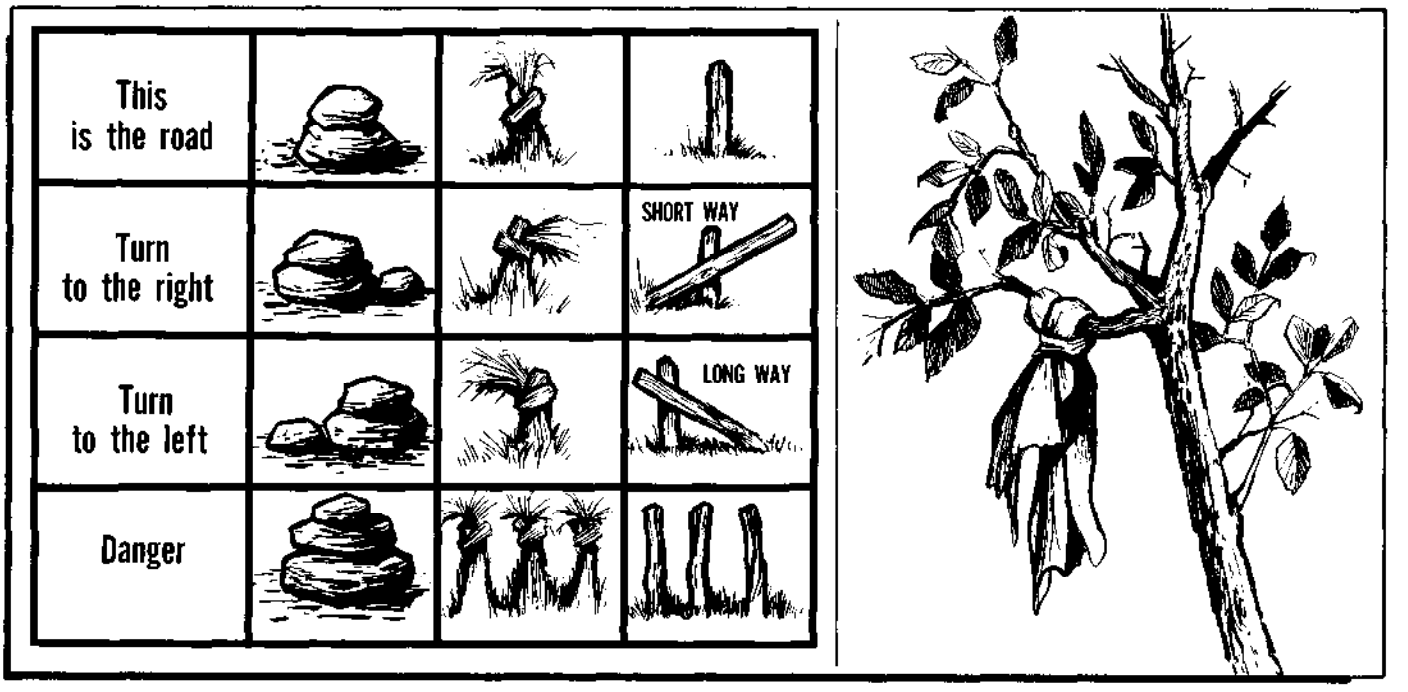

Figure 21-1. Marking a Trail.

readily determine the direction in which rivers or streams flow; however, surveying is necessary to determine if a river is safe for rafting, or if a snowfield or mountainside can be traversed safely. Mountain travel differs from travel through rolling or level country, and certain cardinal rules govern climbing methods. A group descending into a valley, where descent becomes increasingly steep and walls progressively more perpendicular, may be obliged to climb up again in order to follow a ridge until an easier descent is possible. In such a situation, rappeling with a parachute line rope may save many weary miles of travel. In mountains, travelers must avoid possible avalanches of earth, rock, and snow, as well as crevasses (deep cracks in the ice) in ice fields.

(2) In mountainous country, it may be better to travel on ridges—the snow surface is probably firmer and there is a better view of the route from above. Survivors should watch for snow and ice overhanging steep slopes. Avalanches are a hazard on steep snow-covered slopes, especially on warm days and after heavy snowfalls.

(3) Snow avalanches occur most commonly and freqently in mountainous country during wintertime, but they also occur with the warm temperatures and rainfalls of springtime. Both small and large avalanches are a serious threat to survivors traveling during winter as they have tremendous force. The natural phenomena of snow avalances is complex. It is difficult to definitely predict impending avalanches, but knowing general behaviors of avalanches and how to identify them can help people avoid avalanche hazard areas.

(a) Snow or Sluff Avalanche. The loose snow or sluff avalanche is one kind of avalanche that starts over a small area or in one specific spot. It begins small and builds up in size as it descends. As the quantity of snow increases, the avalanche moves downward as a shapeless mass with little cohesion.

(b) Terrain Factors Affecting Avalanches:

-1. Steepness. Most commonly, avalanches occur on slopes ranging from 30 to 45 degrees (60 to 100 percent grade), but large avalanches do occur on slopes ranging from 25 to 60 degrees (40 to 173 percent grade). (See figure 21-2.)

-2. Profile. The dangerous slab avalanches have more chance of occurring on convex slopes because of the angle and the gravitational pull. Concave slopes cause a danger from slides that originate at the upper, steep part of the slope (figure 21-3).

-3. Slopes. Midwinter snowslides usually occur on north-facing slopes. This is because the north slopes do not receive the required sunlight which would heat and stabilize the snowpack. South-facing slope slides occur on sunny, spring days when sufficient warmth melts the snow crystals causing them to change into wet, watery, slippery nowslidessnow. Leeward slopes are dangerous because the wind blows the snow into well packed drifts just below the crest. If the drifts have not adhered to the snow underneath, a slab avalanche can occur. Windward slopes generally have less snow and are compact. It is usually strong enough to resist move-

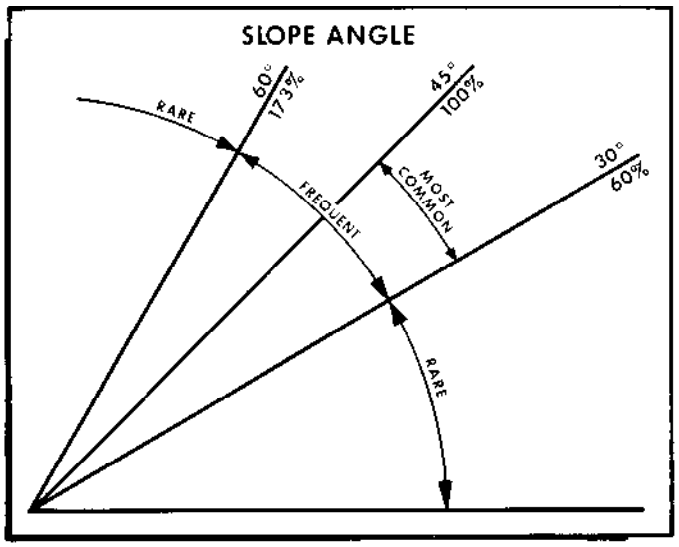

Figure 21-2. Slope Angle.

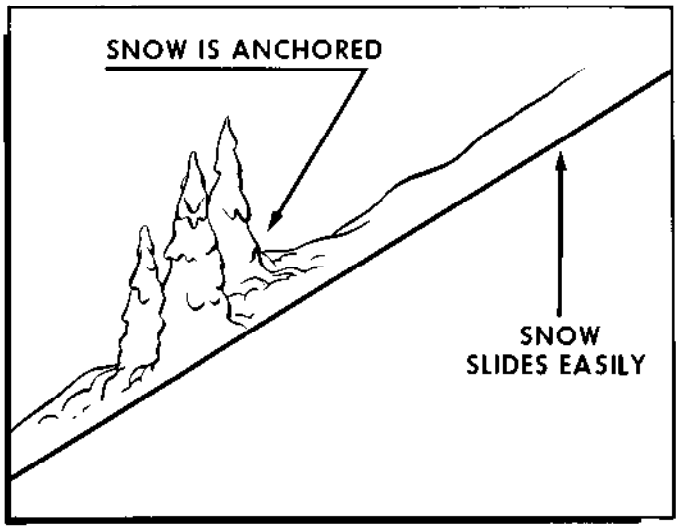

Figure 21-4. Snowslides.

ment, but avalanches may still occur with warm temperature and moisture.

-4. Surface Features. Most avalanches are common on smooth, grassy slopes that offer no resistance.

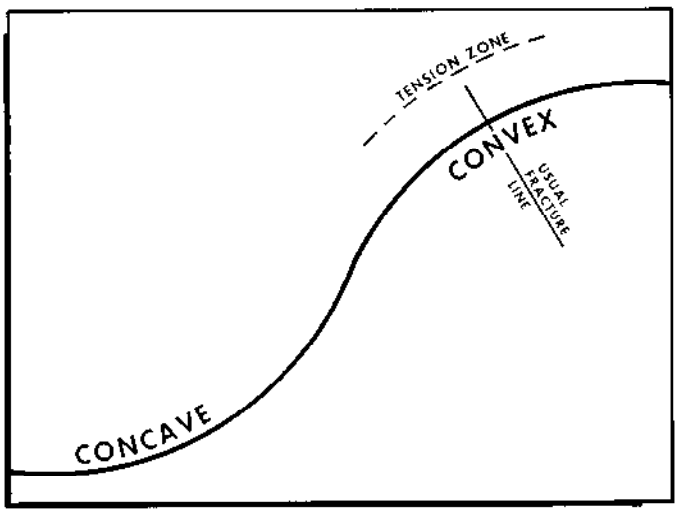

Figure 21-3. Profile of Slope.

Brush, trees, and large rocks bind and anchor the snow, but avalanches can still occur in tree areas (figure 21-4).

(c) Weather Factors:

-1. Old snow depth covers up natural anchors (rocks, brush, and fallen trees) so that the new snowfall slides easier. The type of old snow surface is important. Sun crests or smooth surface snows are unstable; whereas a rough, jagged surface would offer stability and an anchorage. A loose snow layer underneath is far more hazardous than a compacted one as the upper layer of snow will slide more easily with no rough texture to restrain it. Travelers should check the underlying snow by using a rod or stick.

-2. Winds, 15 miles per hour or more, cause the danger of avalanches to develop rapidly. Leeward slopes will collect snow that has been blown from the windward sides, forming slabs or sluffs, depending upon the temperature and moisture. Snow plumes or cornices indicate this condition (figure 21-5).

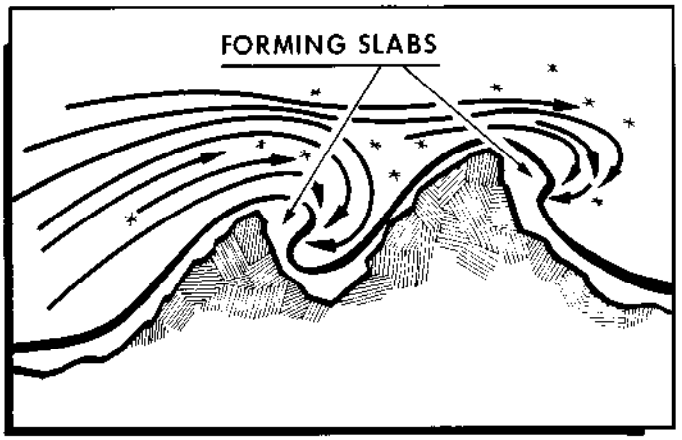

Figure 21-5. Forming Slides.

-3. A high percentage of all avalanches occur during or shortly after storms. Layers of different types of snow from different storms will cause unstable snow because the bond between layers will vary in strength. The rate of snowfall also has a significant effect on stability. A heavy snowfall spread out over several days is not as dangerous as a heavy snowfall in a few hours because slow accumulation allows time for the snow to settle and stabilize. A large amount of snow over a short

period of time results in the snow constantly changing and building up, giving it little time to settle and stabilize. If the snow is light and dry, little settling or cohesion occurs, resulting in instability.

-4. Under extremely cold temperatures, snow is unstable. In temperatures around freezing or just above, the snow tends to settle and stabilize quickly. Storms, starting at low temperature with light, dry snow which are followed by rising temperature, cause the top layer of snow to be moist and heavy, providing opportune conditions for avalanching. The light, dry snow underneath lacks the strength and elastic bondage necessary to hold the heavier, wetter snow deposited on top; therefore, the upper layer slides off. Also, extreme temperature differences between night and day cause the same problems. Rapid changes in weather conditions cause adjustments and movement within the snowpack. Survivors should be alert to rapid changes in winds, temperatures, and snowfall which may affect snow stability.

-5. Avalanches of wet snow are more likely to occur on south slopes. Sun, rainstorms, or warmer temperatures brought by spring weather are absorbed by the snow causing it to become less stable.

(d) Warning Signs. Avalanches generally occur in the same area. After a path has been smoothed, it's easier for another avalanche to occur. Steep, open gullies and slopes, pushed over small trees, limbs broken off, and tumbled rocks, are signs of slide paths. Snowballs tumbling downhill or sliding snow is an indication of an avalanche area on leeward slopes. If the snow echoes or sounds hollow, conditions are dangerous. If the snow cracks and the cracks persist or run, the danger of a slab avalanche is imminent. The deeper the snow, the more the terrain features will be obscured. Knowledge of common terrain features can help survivors visualize what they may be up against, what to avoid, and the safest areas to travel. Knowing the general weather pattern for the area is helpful. Survivors should try to determine what kind of weather will be coming by observing and knowing the signs that indicate certain weather conditions.

(e) Route Selection. If avoiding the mountains and avalanche danger areas is impossible, there are precautions survivors should take when confronting dangerous slopes. They should decide which slopes will be the safest by analyzing the factors that determine what makes one slope safe and another deadly. Study the slope terrain and keep in mind why avalanches occur.

-1. When survivors decide to cross a slope, one person at a time should cross. If all go together, they should not tie together since there is no way one person can hold another against an avalanche. Instead, they should tie a contrasting color line about 100 feet long (using suspension line or PLD tape) to each person. If they should get caught in an avalanche, the line will help identify their position if it is exposed. Survivors should select escape routes before and throughout the climb and keep these routes in mind at all times. They should also stay to the fall line (natural path an object will follow when falling or sliding down the slope) when climbing and not zig-zag or climb a different route because it seems easier. Staying to the fall line will prevent making fractures and at the same time, compact the snow, making it more stable for others who follow. If traversing, they should travel above the danger area. Survivors should travel quickly and quietly to avoid extended exposure to the probable danger of avalanches.

-2. If caught in an avalanche, any equipment that weighs survivors down should be dropped. The pack, snowshoes, and any other articles should be jettisoned.

-3. The standard rule is to use swimming motions to try to move towards the snow surface. Further, survivors should go for the sides and not try to out swim the avalanche. If near the surface, they should try to keep one arm or hand above the surface to mark their position. If buried, a person should inhale deeply (nose down) before the snow stops moving to make room for their chest. Trapped persons should try to make breathing space around their faces. They shouldn't struggle, but should relax and conserve their energy and oxygen. Only when fellow survivors or rescuers are nearby should the trapped individual shout. Rescue should be done as quickly as possible. Avalanche victims generally have a 50 percent chance of surviving after 30 minutes have passed because the snow will set up (harden).

-4. Glaciers may offer emergency travel routes across mountain ranges. Glacier crossing demands special knowledge, techniques, and equipment such as the use of a lifeline and poles for locating crevasses. There are many places where mountain ranges can be negotiated on foot in a single day by following glaciers. Survivors must be especially careful on glaciers and watch for crevasses covered by snow. If traveling in groups of three persons or more, they should be roped together at intervals of 30 to 40 feet. Every step should be probed with a pole. Snow-bridged crevasses should be crossed at right angles to their course. The strongest part of the bridge can be found by probing. When crossing a bridged crevasse, weight should be distributed over a large area by crawling or by wearing snowshoes.

-5. When forested areas are dense, river travel and ridges usually afford the easiest travel routes. In open forests, land travel is easier and offers a better selection of travel routes; however, such forests may not offer sufficient cover or concealment for escape and evasion travel.

-6. After a fire, windstorm, or logging operation, second-growth timber usually grows thick. It is worse after it grows about 20 feet high since any space between the trees is filled in by branches and the overhead timber isn't (yet) thick enough to cause the lower

branches to die from lack of sunlight.

-a. Deciduous brush also contributes to the overgrowth. Blowdowns, avalanche fans, and logging slush are difficult to negotiate. Such obstructions, even a few hundred feet may require major changes to the original travel route.

-b. Scrub cedar (subalpine fir) is hard to penetrate. There are tactics that can be used to make travel easier. Survivors can use fallen trees as walkways to provide a short route of travel through the scrub cedar to a clear area. Gloves should be worn when penetrating thorny vegetation. Overlaying bushes can be separated to allow passage. When land is steep, brush can be used to provide handholds if it is strong and anchored well.

-c. Brush can be dangerous. Survivors should be aware of the possibility of slipping while going downhill. Therefore, they should ensure each step is firmly placed. Survivors should be aware of travel difficulties presented by cliffs, boulders, and ravines which are covered by brush.

-7. Do not travel through dense brush if it can be avoided.

-a. Travel on trails rather than taking shortcuts through the brush. Brush is frequently easier to travel through (over) during the winter season when it is covered by snow and when snowshoes are available (improvised).

-b. During the summer, avoid avalanche tracks because the debris may be difficult to penetrate. Traveling on the "timber cones" between avalanche paths is best when climbing a valley.

-c. The heaviest timber is the best area to travel because little or no brush will be growing on the forest floor.

-d. Try to avoid areas near creeks and valley floors because they have more brush and trees than the valley walls and ridges. However, traveling in the stream channels may be preferable when the area is covered with dense brush and vegetation. Survivors may have to wade, but the stream may offer the best route through the brush.

-e. Traveling high above the brush at the timberline may be worthwhile if the bottom and sides of the valley look futile.

d. Snow and Ice Areas. Travel in snow and ice areas is not recommended except to move from an unsafe to a safe area, or from an area that has few natural resources to an area of greater resources (shelter material, food, and signaling area).

(1) Before traveling to a possible rescue site, town, village, or cabin, travelers should know their approximate position and the location of the desired site. The greatest hazard in snow and ice areas is the intense cold and high winds. Judging distance is difficult due to the lack of landmarks and the clear arctic air. Image distortion is a common phenomena. "White-out" conditions exist and the survivor should not travel during this time. A white-out condition occurs when there is complete snow cover and the clouds are so thick and uniform that light reflected by the snow is about the same intensity as that from the sky. If traveling during bad weather, great care must be taken to avoid becoming disoriented or falling into crevasses, over cliffs or high snow ridges, or walking into open leads. A walking stick is very useful to probe the area in the line of travel.

(2) Strong winds often sweep unchecked across tundra areas (due to the lack of vegetation) causing white-out conditions. Because of blowing snow, fog, and lack of landmarks, a compass is a must for travel, yet it is still difficult to navigate a true course since the magnetic variation in the high latitudes (polar areas) is often extreme.

(3) During the summer months, the area is a mass of bogs, swamps, and standing water. Crossing these areas will be difficult at best. Rain and fog are common. Insects such as mosquitoes, midges, and black flies can and will cause the survivor physical discomfort and may cause travel problems. If the body is not completely covered with clothing, or if survivors do not use a head net or insect repellant, insect bites may be severe and infection can set in.

(4) In mountainous country, it is often best to travel along ridge lines because it provides a firmer walking surface and there is usually less vegetation to contend with. High winds make travel impractical if not impossible at times. Glaciers have many hidden dangers. Glacial streams may run just under the surface of the snow or ice, creating weak spots, or they may run on the surface and cause slick ice. Crevasses which run across the glacier can be a few feet to several hundred feet deep. Quite often crevasses are covered over with a thin layer of snow, making them practically invisible. Survivors could fall into crevasses and sustain severe injuries or death. If glacier travel is required, it is best to use a probe pole to test the footing ahead.

(5) Summer travel in timbered areas should not present any major problems; however, travel on ridges is preferred since the terrain is drier and there are usually fewer insects. During the cold months, snow may be deep and travel will be difficult without some type of snowshoes or skis. Travel is generally easier on frozen rivers, streams, and lakes since there is less snow or wind-packed snow and they are easier to walk on.

(6) River and stream travel can be hazardous. Rivers comparatively straight are that way because of the volume of water flow and extremely fast currents. These rivers tend to have very thin ice in the winter (cold climates), especially where snowbanks extend out over the water. If an object protrudes through the ice, the immediate area will be weak and should be avoided if possible. Where two rivers and streams come together, the current is swift and the ice will be weaker than the ice on the rest of the river. Very often after freezeup, the

source of the river or stream dries up so rapidly that air pockets are formed under the ice and can be dangerous if fallen into. During the runoff months (spring and summer), rivers and streams usually have a large volume of water which is very cold and can cause cold injuries. Wading across or down rivers and streams should be done with proper footwear and exposure protection due to the depth, swiftness, unsure footing, and coldness of the water. Generally, streams are too small and shallow for rafting. Streams are often bordered by high cliffs or banks at the headwaters. As a stream progresses, its banks are often choked with alder, devil's club, and other thick vegetation making traveling very slow and difficult. Many smaller streams will simply lead the traveler to a bog or swamp where they end, causing more problems for the survivor.

(7) Sea ice conditions vary greatly from place to place and season to season. During the winter months, there is generally little open water except between the edges of floes. Crossing from one floe to another can be done by jumping across the open-water area, but footing may be dangerous. When large floes are touching each other, the ice between is usually ground into brash ice by the action of the floes against each other and this ground-up ice will not support a survivor's weight. Pressure ridges are long ridges in sea ice caused by the horizontal pressure of two ice floes coming together. Pressure ridges may be 100 feet high and several miles long; they may occur in a gulf or bay, or on polar seas. They must be crossed with great care because of the ruggedness of ice formations, weak ice in the area, and possibility of open water covered with a thin layer of snow or ground-up ice. During summer months, the ice surface becomes very rough and covered with water. The ice also becomes soft and honeycombed (candlestick ice) even though the air temperature may be below freezing. Traveling over sea ice in the summer months is very dangerous.

(8) Icebergs are great masses of ice and are driven by currents and winds; about two-thirds of the iceberg is below the surface of the sea. Icebergs in open seas are always dangerous because the ice under the water will melt faster than the surface exposed to the air, upsetting the equilibrium and toppling them over. The resulting waves can throw small pieces of ice in all directions. Avoid pinnacle-shaped icebergs—low, flat-topped icebergs are safer.

e. Dry Climates. Before traveling in the desert, the decision to travel must be weighed against the environmental factors of terrain and climate, condition of survivors, possibility of rescue, and the amount of water and food required.

(1) The time of day for traveling is greatly dependent on two significant factors; the first and most apparent is temperature, and the second is type of terrain. For example, in rocky or mountainous deserts, the eroded drainages and canyons may not be seen at night and could result in a serious fall. Additionally, manmade features such as mining shafts or pits and irrigation channels could cause similar problems. If the temperature is not conducive to day travel, survivors should travel during the cooler parts of the day (in early morning or late evening). Traveling on moonlit nights is another possibility; however, survivors must be aware that moonlight can cast deceiving shadows. This problem can be decreased by scanning the ground to allow the night sensitive portions of the eye time to pick up the slight differences in lighting. In hot desert areas where these hazards do not occur, traveling at night is a very practical solution. During the winter in the midlatitude deserts, the cold temperatures make day travel most sensible.

(2) There are three types of deserts: mountain, rocky plateau, and sandy or dune deserts. These deserts can present difficult travel problems.

(a) Mountain deserts are characterized by scattered ranges or areas of barren hills or mountains, separated by dry, flat basins. High ground may rise gradually or abruptly from flat areas to a height of several thousand feet above sea level. Most desert rainfall occurs at high elevations and the rapid runoff causes flash floods, eroding deep gullies or ravines and depositing sand and gravel around the edges of the basins. These floods are a problem on high and low grounds. The flood waters rapidly evaporate, leaving the land as barren as before, except for plush vegetation which rapidly becomes dormant. Basins without shallow lakes will have alkaline flats which can cause problems with chemical burns and can destroy clothing and equipment.

(b) Rocky plateau deserts have relief interspersed by extensive flat areas solid or broken rock at or near the surface. They may be cut by dry, steep-walled, eroded valleys, known as wadis in the Middle East and arroyos or canyons in the United States and Mexico. The narrower of these valleys can be extremely dangerous to humans and material due to flash flooding. Travel in these valleys may present another problem: a survivor can lose site of reference points and travel farther than intended. The Golan Heights in Israel is an example of a rocky plateau desert.

(c) Sandy (dune) deserts are extensive flat areas covered with sand or gravel. They are the product of ancient and modern wind erosion. "Flat" is relative in this case, as some areas contain sand dunes that are over 1,000 feet high and 10 to 15 miles long. Travel in such terrain depends on windward or leeward gradients of the dunes and texture of sand. These dunes help a survivor determine general direction. Longitudinal dunes are continuous banks of sand at even heights that lie parallel with the dominant wind. Other areas, however, may be totally flat for distances in excess of 2 miles. Plant life varies from none to scrub reaching over 6 feet. Examples of this desert include the ergs of the Sahara Desert, Empty Quarter of the Arabian Desert,

areas of California and New Mexico, and the Kalahari Desert in South Africa. A seif dune has forms similar to a drift behind a rock. Its length lies in the direction of the prevailing winds. Additionally, the horseshoe-shaped crescent dune has a hollow portion that faces downward. Ripples caused by wind in the sand may indicate the direction of the prevailing winds. These ripples generally lie perpendicular to the prevailing winds. In deserts, it is easier to travel on the windward side of the tops of dunes. Even though these ridges may not lay in a straight line and may wander, they offer a better route of travel than traveling in straight lines. A great deal of energy and time can be expended walking up and down dunes, especially in the loose sand on the leeward side of dunes.

(3) People who travel the desert at night orient themselves by the stars and Moon. Those who traveled by day use compasses, when available, and the Sun. Survivors should use all directional aids during emergency travel and each aid should be frequently cross-checked against each other. For desert travel, a compass is a valuable piece of equipment.

(a) Without a compass, landmarks must be used for local navigation. This can lead to difficulties. Mirages cause considerable trouble. Ground haze throughout the day may obscure vision. Distances are deceptive in the deserts and survivors have reported difficulty in estimating distances and the size of objects. In southern Egypt, one survivor reported large boulders always appeared smaller than they were and in other cases small obstacles appeared insurmountable. Survivors in Saudi Arabia and in Tunisia warned that it is difficult to maintain a single landmark in navigation. Several groups reported they found it necessary to take turns keeping an eye on a specific mountain, peak, or object which was their goal. Objects have a way of vanishing in some cases when the eye is moved for an instant, and in other cases, many peaks or hills looked alike and caused difficulties in determining the original object. In Tunisia, twin peaks are not reliable landmarks because of their frequent occurrence. (Survivors have found after a short time of traveling they may have up to a dozen twin peaks for reference in the same vicinity.) The Great Sand Sea (Egypt and Libya) was the emergency landing site of several groups of survivors and caused navigational difficulties. In these rolling sand hills, it is impossible to keep one object in view, and even footmarks fail to provide a reliable backtrail for determining travel directions. The extreme flatness of other stretches of desert terrain in North Africa also makes navigation difficult. With no landmarks to follow, no objectives to sight, survivors may walk in circles or large arcs before realizing their difficulties.

(b) A Marine pilot who made an emergency landing in the Arizona desert took the precaution of immediately spreading his parachute on the ground and putting rocks on the edges to ensure maximum visibility from the air. Then he decided to walk to his crashed plane, a distance he estimated to be 500 yards from his landing spot. He reached the plane and found it gutted by fire, and spent 5 days wandering the flat desert trying to find his parachute.

(c) Navigational difficulties of a different type may be experienced in Ethiopia, Kenya, and Somalia. Here the density of the thorn brush, even though it was primarily acacia with small leaves, makes it extremely hard to navigate from one point to another. In this area, survivors should follow animal trails and hope they lead to rivers or waterholes. Elephant trails seem to offer the best and clearest route.

(d) In the Sinai Desert area and in portions of Egypt, travel routes may be used; survivors can "stay put" on the trails. One survivor, who made a trail, encountered a camel caravan almost immediately, athough he reported it bothered him that he had not seen them approach, as they suddenly appeared out of a mirage. Another commented it was awfully hard to be alone in his section of the desert, for in every direction, he saw wandering tribes, camel herds, or people watching him. Two survivors independently suggested that survivors pay attention to the wind as an aid in navigation. One survivor, on the Arabian Peninsula, noted the wind blew consistently from the same direction. The other, in the Libyan Desert, made the same comment and said he was able to judge his direction of travel by the angle at which the wind blew his clothes or struck his body. Survivors in certain areas may orient themselves to the prevailing winds once it is established that these are consistent.

(4) People who walked across the North African deserts had much to say about the local environment and little of it was complimentary. The extreme temperatures bothered them most. It was extremely hot during the day and often bitter cold at night, especially during January and February.

(a) The bright sunlight was hard on their eyes, extremities, and exposed skin. The blinding effects of the Sun reflecting off the terrain caused many persons to express concern regarding sunglasses. Several built fires and smoked their goggles to obtain protection against the glare. Lenses alone do not adequately protect the eyes from glare. Sunglasses may be improvised to reduce this problem. Light-skinned individuals tended to sunburn faster and more severely than darker-skinned individuals. Some reported that no amount of previous suntanning seemed to make any difference. The heat affected their feet and hands. Survivors reported that the surface became so hot their feet were blistered through their shoes. Exposure of bare hands to the sunlight resulted in painful burns. Placing sunburned hands in bare armpits gave considerable relief since the armpits were one of the few places on the body where a person could find continuous perspiration to aid in cooling.

(b) The persistent winds of desert areas seem to provide no cooling effect, and several survivors found the constant blowing of the wind "got on their nerves." More significant is the fact that the constant winds usually carried an amount of sand or dirt particles. These particles got in eyes, ears, nostrils, and mouth and caused irritations which were often severe. Additionally, this persistent wind also caused earaches. One survivor reported that the abrasions of the eyes by the particles of dust reached a point where first the man's eyes watered so much that he could not see, and eventually the watering stopped and "emery cloth eyelids remained," making life miserable for him.

(c) Extreme winds blew sandstorms which lasted from a few minutes to months. Generally, survivors reported they could see the approach of such storms and were able to take proper precautions; however, sandstorms completely surprised a few groups and they had difficulty navigating. None of the survivors who experienced sandstorms in the northern desert underestimated the power and danger of such storms. Protection from the storms was uppermost in their minds. Most survivors used rock cairns, natural ledges, boulders, depressions, or wells for shelter. Survivors had time to dig depressions and rig a shelter from blankets, parachutes, or tarpaulins. A few wrapped themselves in their parachutes and endured the storm in a prone position.

(d) Nearly all of these people had some comment to make on orientation before, during, and after a sandstorm. They warned specifically that it is necessary to adequately mark the direction of travel before the storm. A few survivors said when the storm was over they had no idea which way they had been traveling and all their landmarks were forgotten, obliterated, or indistinguishable. The general plan for marking travel routes before a sandstorm is to place a stick to indicate direction. One survivor oriented himself with one rock a few feet in front of his position. He commented after the storm, that one point was not adequate and recommended using a row of stones, sticks, or heavy gear about 10 yards in length to give adequate direction following such an emergency.

(e) Several survivors reported they learned the hard way to keep their mouths shut in the desert. This meant breathing through the mouth caused drying and talking not only got on the other's nerves but also caused excessive drying of the mucous membranes.

(5) Mirages are common in desert areas. They are optical phenomena due to refraction of light by uneven density distribution in the lower layer of the air. The most common desert mirage occurs during the heat of the day when the air close to the ground is much warmer than the air aloft. Under this condition, atmospheric refraction is less than normal and the image of the distant low sky appears on the ground looking like a sheet of water. Distant objects may appear to be reflected in the "water." When the air close to the ground is much colder than the air aloft, as in the early morning under a clear sky, atmospheric refraction is greater than normal. When this condition occurs, distant objects appear larger and closer than they are and objects below the normal horizon are visible. Unless the density distribution in the lower layers is such that the light rays from an object reach the observer along two or more paths, they will see a distorted image or multiple images of the object.

(a) Reports of mirages were very common in the survival episodes examined. In most cases, they were recognized as mirages and only caused minor difficulties. No survivor reported these mirages actually represented bodies of water. While traveling, the survivor experienced problems as a result of mirages. Distances could not be judged because intermediate terrain was obscured by the mirage. Mirages hampered vision and navigation since it concealed objects. Additionally, mirages "magnified some objects and concealed others." One man hunting in the heat of the day reported that when he sighted an animal, it ran into or hid in a mirage. The lower layer of hot air which causes the mirage, commonly called desert haze, hampers vision and distorts objects. Signaling difficulties resulted from this since sighting a reflection on an object was apparently very difficult due to the low haze on the desert.

(b) Several survivors reported cases of imaginary illusions which were due to the haze or mirages. One group looked for a hill, for a viewpoint, so long that the entire party began to see hills in all directions. They finally held a conference to iron out their difficulties and all settled on one hill which the group should approach. Everyone in the party saw the hill, and the group walked an estimated 9 miles looking for the hill which never existed and which eventually disappeared into the desert flatness. Dawn and dusk illusions also occurred and were reported in the survivors' stories. One group was severely troubled with the false-dawn spectral light on the western horizon. The fact that the Sun at first appeared to be rising in the west caused anxious moments. Another party had one person who claimed he saw a flashing beacon on the evening horizon. The pilot explained the illusion as the occasional refraction of bright starlight near the horizon, through the residual heat waves of the cliff before them. But one crewmember was so convinced it was a beacon that he started walking to investigate this object and was never seen again.

(6) The following are manmade characteristics of the desert:

(a) Roads and trails are scarce in the desert, as complex road systems are not needed. Most existing road systems have been used for centuries to connect centers of commerce or important religious shrines, such as Mecca and Medina in Saudi Arabia. These roads are now supplemented by routes for transporting oil or other mineral deposits to collection points. Rudimentary trails exist in many deserts for caravans and nomadic tribesmen. These trails have wells or oases ap-

proximately every 20 to 40 miles, although there are waterless stretches of over 100 miles. These trails vary in width from a few yards to over 800 yards. Vehicle travel in mountainous desert terrain may be severely restricted. Available routes may be easily blocked by hostile people or climatic conditions. Passes may be blocked by snow in the winter. The travel distance on foot or by animal between two points in the mountains may be less than a tenth of the distance required if vehicles are used to make the trip.

(b) Apart from nomadic tribesmen who live in tents, desert inhabitants live in thick-walled structures with small windows, usually built of masonry or a mud and straw (adobe) mixture. The ruins of earlier civilizations are scattered across the deserts. Ancient posts and forts invariably command important avenues of approach and frequently dominate the only available passes in difficult terrain. Control of these positions may be imperative for forces intent on dominating the immediate area.

(c) Exploration for and exploitation of natural resources, of which oil is the best known, occurs in many desert areas, especially the Middle East. Wells, pipelines, refineries, quarries, and crushing plants may lead a survivor to rescue or captivity. Additionally, pipelines are often elevated, making them visible from a distance.

(d) Many desert areas are irrigated for agricultural and habitation purposes. Agriculture and irrigation canals are signs which can lead a survivor to people.

f. Tropical Climates. The inexperienced person's view of jungle travel may range from difficult to nearly impossible. However, with patience and good planning, the best and least difficult route can be selected. In some cases, the easiest routes are rivers, trails, and ridge lines. However, there may be hazards associated with these routes.

(1) Rivers and streams may be overgrown making them difficult to reach and impossible to raft. These waterways may also be infested with leeches. Trails may have traps or animal pits set on them. Trails can also lead to a dead end or into thick brush or swamps. Ridges may end abruptly at a cliff. The vegetation along a ridge may also conceal crevices or extend out past cliffs, making the cliff unnoticed until it's too late.

(2) The machete is the best aid to survival in the jungle. However, survivors should not use it unless there is no other way. They should part the brush rather than cut it if possible. If the machete must be used, cut at a down-and-out angle, instead of flat and level, as this method requires less effort.

(3) Survivors should take their time and not hurry. This allows them to observe their surroundings and gives better insight as to the best route of travel. Watch the ground for the best footing as some areas may be slippery or give way easily. Avoid grabbing bushes or plants when traveling. Falling may be a painful experience as many plants have sharp edges, thorns, or hooks. (NOTE: Wear gloves and fully button clothes for personal protection.)

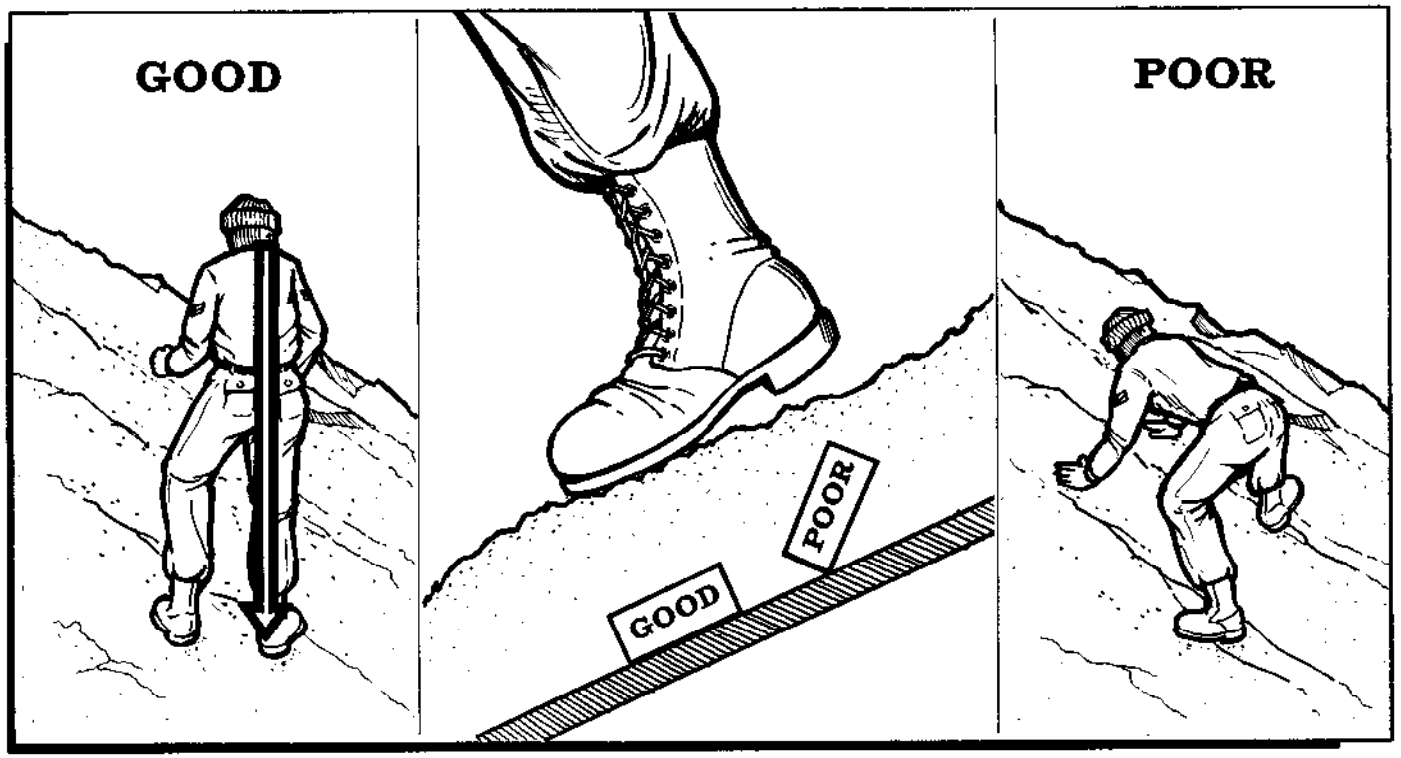

Figure 21-6. Weight of Body Over Feet.

(4) Quicksand can be a problem. In appearance, quicksand looks just like the surrounding area with an absence of vegetation. It is usually located near the mouths of large rivers and on flat shores. The simplest description of quicksand is a natural water tank filled with sand and supplied with water. The bottom consists of clay or other substances capable of holding water. The sand grains are rounded, as opposed to normal sharper-edged sand. This is caused by water movement which also prevents it from settling and stabilizing. The density of this sand-water solution will support a person's body weight. The danger if a survivor panics may be drowning. In quicksand, the survivor should use the spread-eagle position to help disperse the body weight to keep from sinking and a swimming technique to return to solid ground. (NOTE: Remember to avoid panicking and struggling, and spread out and swim or pull along the surface.)

21-5. Mountain Walking Techniques. Depending upon the terrain formation, mountain walking can be divided into four different techniques—walking on hard ground, walking on grassy slopes, walking on scree slopes, and walking on talus slopes. Included in all of these techniques are two fundamental rules which must be mastered in order to expend a minimum of energy and time. These fundamentals are: The weight of the body must be kept directly over the feet (figure 21-6), and the sole of the boot must be placed flat on the ground (figure 21-6). These fundamentals are most easily accomplished by taking small steps at a slow steady pace. An angle of descent which is too steep should be avoided and any indentations or protrusions of the ground, however small, should be used to advantage.

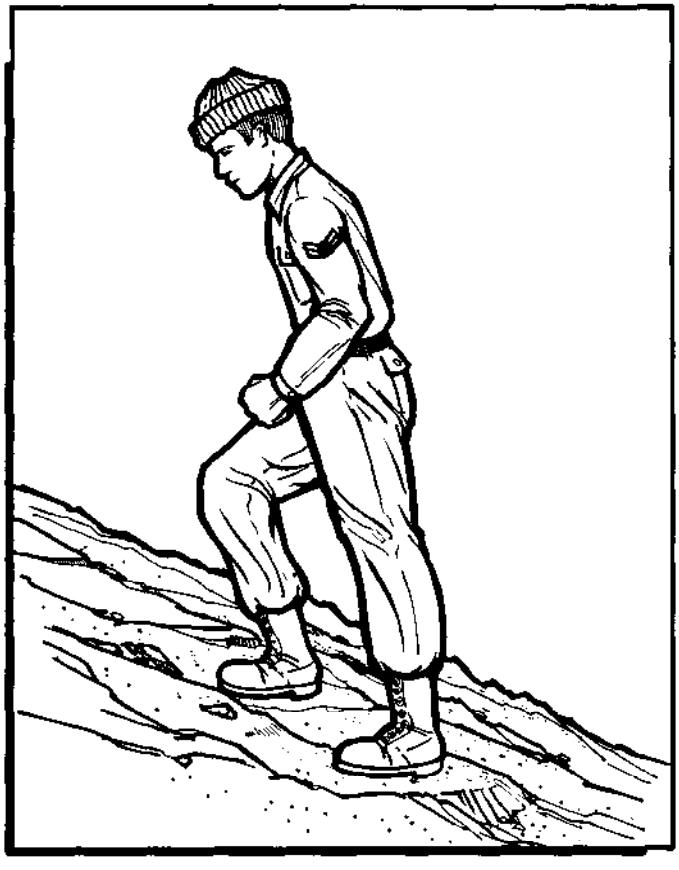

Figure 21-7. Locking Knees.

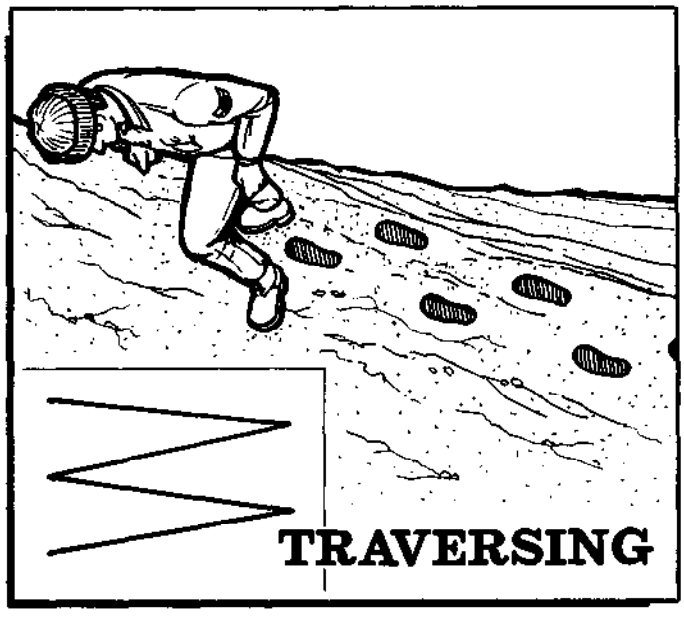

Figure 21-8. Traversing.

a. Hard ground is usually considered to be firmly packed dirt which will not give way under the weight of a person's step. When ascending, the above fundamentals should be applied with the addition of locking the knees on every step in order to rest the leg muscles (figure 21-7). When steep slopes are encountered, they can be traversed easier than climbed straight up. Turning at the end of each traverse should be done by stepping off in the new direction with the uphill foot (figure 21-8). This prevents crossing the feet and possible loss of balance. In traversing, the full sole principle is done by rolling the ankle away from the hill on each step. For narrow stretches, the herringbone step may be used; that is, ascending straight up a slope with toes pointed out (figure 21-9) and using the principles stated above. Descending is usually easiest by coming straight down a slope without traversing. The back must be kept straight and the knees bent (figure 21-10) in such a manner that they take up the slack of each step. Again, remember the weight must be directly over the feet, and the full sole must be placed on the ground with every step.

b. Grassy slopes are usually made up of small tussocks of growth rather than one continuous field. In ascending, the techniques previously mentioned are applicable; however, it is better to step on the upper side of each tussock (figure 21-11) where the ground is more level than on the lower side. Descending is best done by traversing.

Figure 21-9. Walking Uphill.

c. Scree slopes consist of small rocks and gravel which have collected below rock ridges and cliffs. The size of the scree varies from small particles to the size of a fist. Occasionally, it is a mixture of all size rocks, but normally scree slopes will be made up of rocks the same size. Ascending scree slopes is difficult, tiring, and should be avoided when possible. All principles of ascending hard ground apply, but each step must be picked carefully so the foot will not slide down when weight is placed on it. This is best done by kicking in with the toe of the upper foot so a step is formed in the scree (figure 21-12). After determining the step is stable, carefully transfer weight from the lower foot to upper and repeat the process. The best method for descending scree is to come straight down the slope with feet in a slight pigeon-toed position using a short shuffling step with the knees bent and back straight (figure 21-13). When several climbers descend a scree slope together, they should be as close together as possible, one behind the other, to prevent injury from dislodged rocks. Since there is a tendency to run down scree slopes, care must be taken to ensure that this is avoided and control is not lost. By leaning forward, one can obtain greater control. When a scree slope must be traversed with no gain or loss of altitude, use the hop-skip method. This is a hopping motion in which the lower foot takes all the weight and the upper foot is used for balance.

d. Talus slopes are similar in makeup to the scree slopes, except the rock pieces are larger. The technique of walking on talus is to step on top of, and on the uphill side of, the rocks (figure 21-14). This prevents them from tilting and rolling downhill. All other previously mentioned fundamentals are applicable. Usually, talus is easier to ascend and traverse, while scree is a more desirable avenue of descent.

21-6. Burden Carrying. Backpacking is essential when heavy loads must be carried for distances. Using a suitable harness and following certain approved packing and carrying techniques can eliminate unnecessary hardships and help in transporting the load with greater safety and comfort. Carrying a burden initially creates mental irritation and fatigue, either of which can lower morale. Survivors should keep their minds occupied with other thoughts when packing a heavy load. Adjustments should be made during each rest stop to improve the fit and comfort of the pack. Additionally, the rate of travel should be adjusted to the weight of the pack and the environmental characteristics of the terrain being crossed.

a. Burden carrying is a task that must be done in most survival environments. Often survivors must quickly

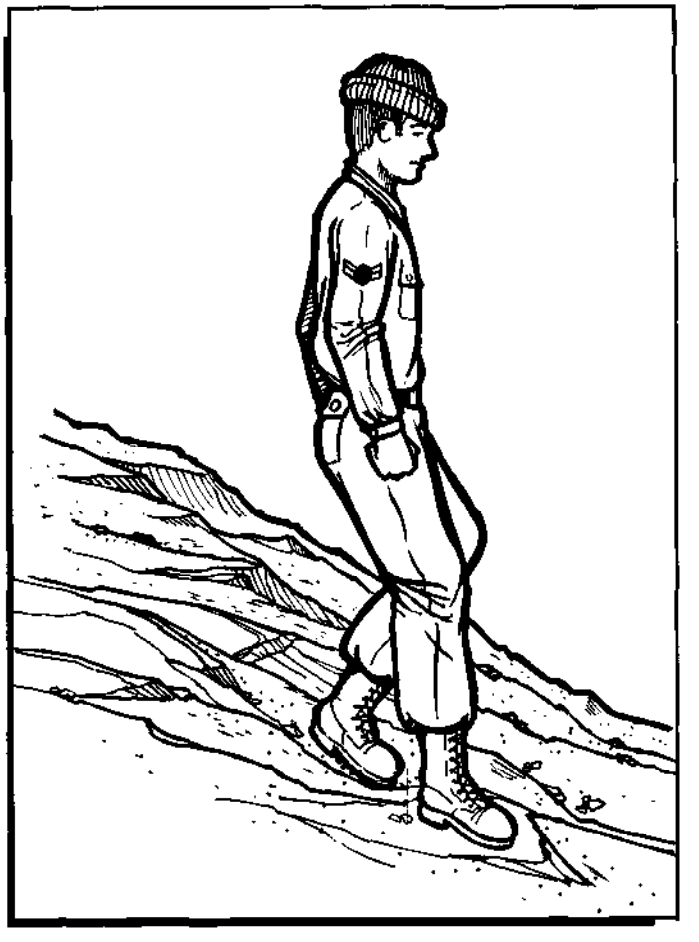

Figure 21-10. Walking Downhill.

Figure 21-11. Walking on Grassy Slopes.

gather their equipment and move out without the assistance of a good pack. The gear may have to be carried in the arms while rapidly leaving the area. In such an instance, it would be better to fashion a roll of the gear and wear it over the shoulder, time permitting. When time is not a factor, it may be desirable to make a semirigid pack such as as square pack. The convenience of being able to keep track of equipment, particularly small items, can be critical in survival situations.

(1) Packsack. A packsack can be fashioned from available survival kit containers or several layers of the fabric from the parachute canopy. The sack is sewed with inner core and a needle.

(2) Square Pack:

(a) The following instructions explain how to improvise a square pack (figure 21-15). Lay a rectangle of

Figure 21-12. Walking on Scree Slopes.

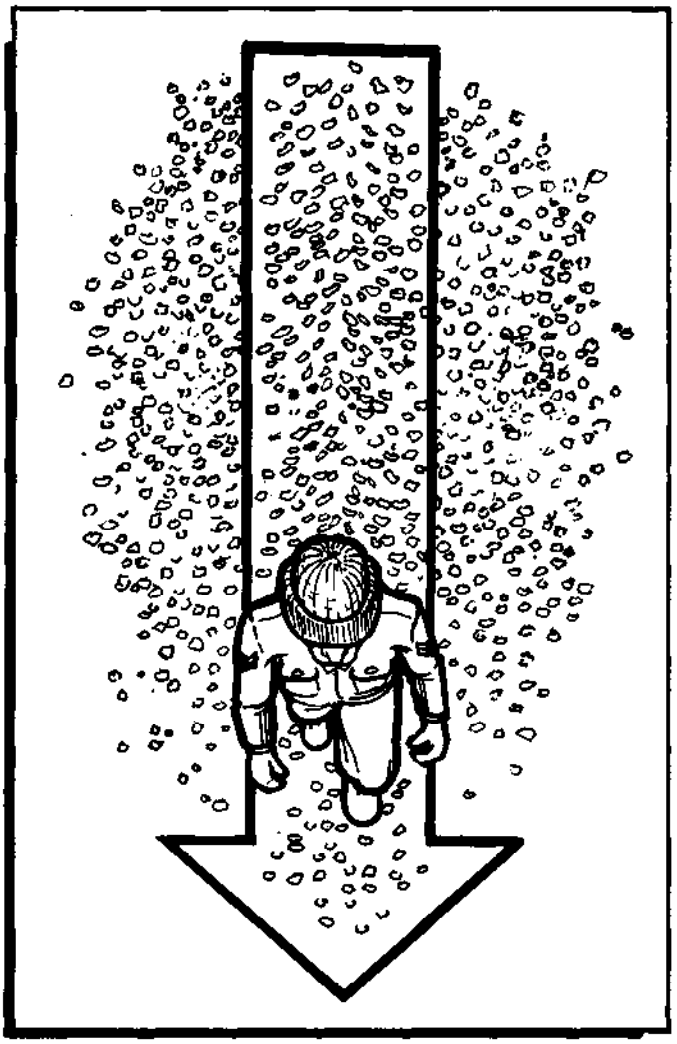

Figure 21-13. Walking Down Scree Slopes.

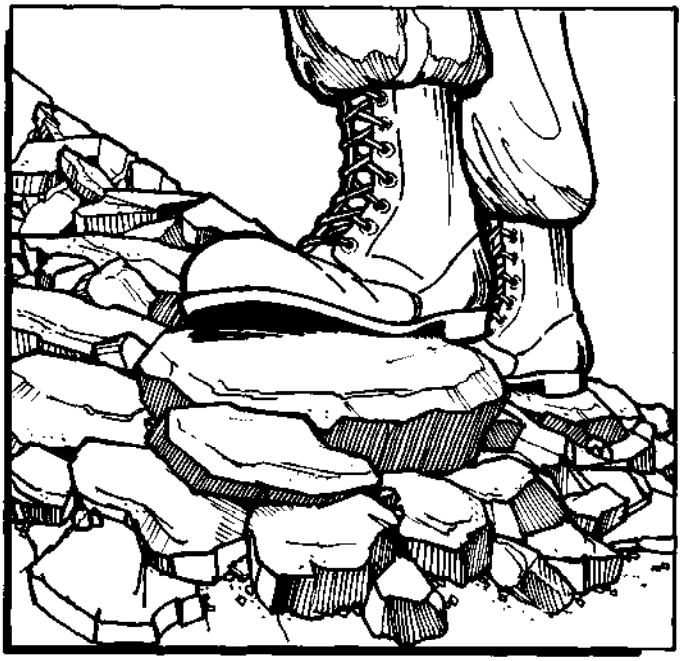

Figure 21-14. Walking on Talus Slopes.

material, waterproof if available, (5 feet by 5 feet minimum size) flat on the ground (examples are plastic, tarp, poncho, paulin 77, etc.) (illustration A). Visualize the material divided into squares like a tic-tac-toe board. The largest piece of soft, bulky equipment (sleeping bag, parachute canopy, etc.) is placed in the center square in an "S" fold. This places the softest item in the pack against the tarp which rests on the back while traveling. If using a poncho, place the sleeping bag just below the hood opening (illustration B). Place hard, heavy objects between the top layer and the middle layer of the "S" fold near the top of the pack. Soft items can be placed between the middle and bottom layer (illustration C).

(b) After all desired items are inside the folds, tie the inner pack in the fashion shown in illustration D. Start with a 1-inch diameter loop in the end of a long piece of parachute suspension line or other suitable line and loop it around the "S" fold laterally. Standing at the bottom of the pack, divide it into thirds and secure the running end of the line to the loop with a trucker's hitch. Both of these hitches should be at the intersection of the thirds so as to divide the pack vertically into thirds. Wrap the running ends around the pack at 90 degrees (working toward the center) to the line and when it crosses another line, use a jam (postal) hitch to secure it and pull both ways to ensure tightness in all directions. When returning to the original Figure 22-42 starting position, use the loop of the tied trucker's hitch to secure another trucker's hitch and the inner part is complete (illustration D). The waterproof materials are then folded around the inner pack as shown in illustrations E and F. Tie the "outer" pack in the same manner ensuring that it is waterproof with all edges folded in securely. If a poncho is used, the head portion may be used to get into the pack if necessary. However, it must be properly secured to ensure that the inner items are protected. With a square pack constructed in this manner, there is no reason why the equipment should get wet. (NOTE: With practice, an excellent pack can be

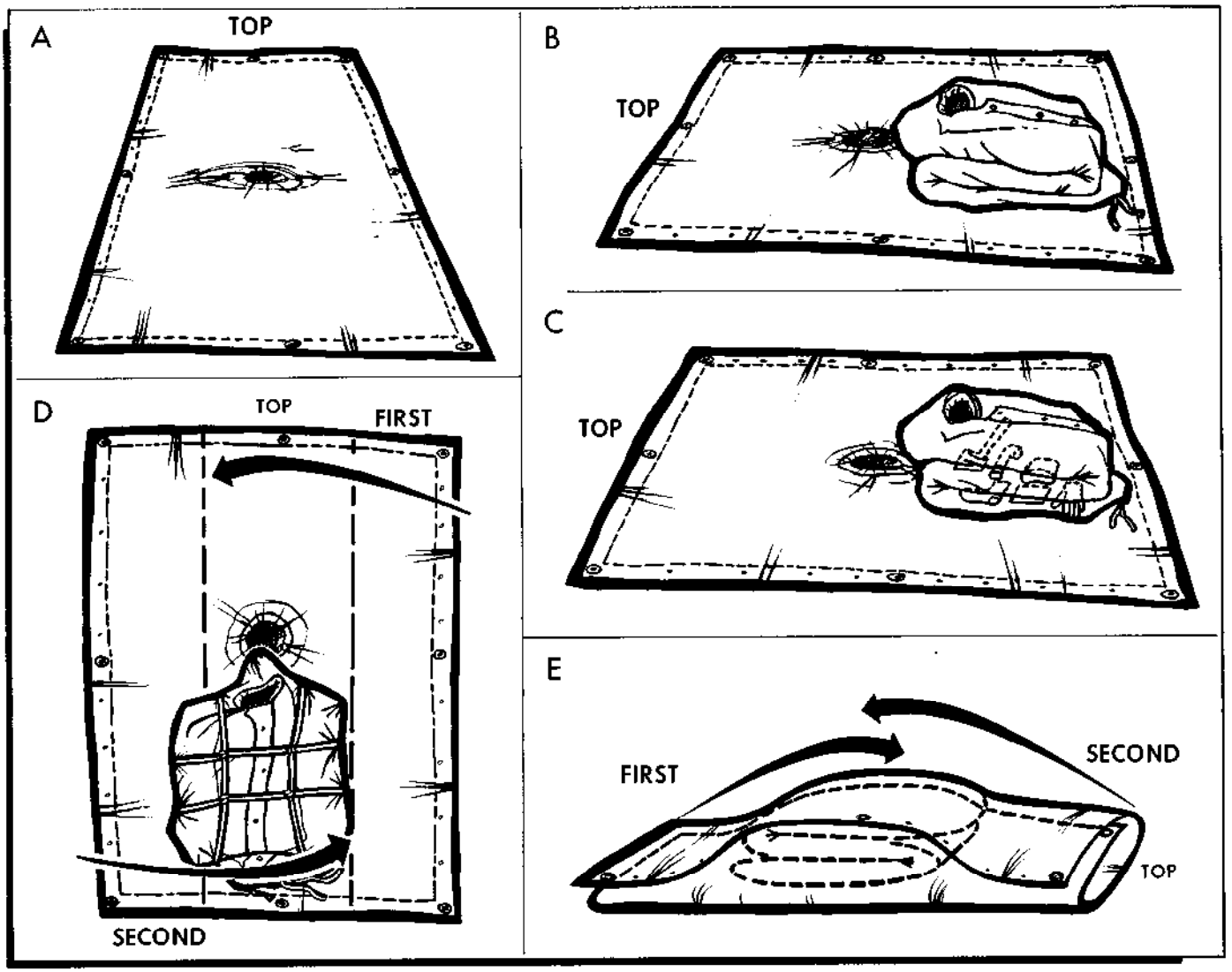

Figure 21-15. Square Pack.

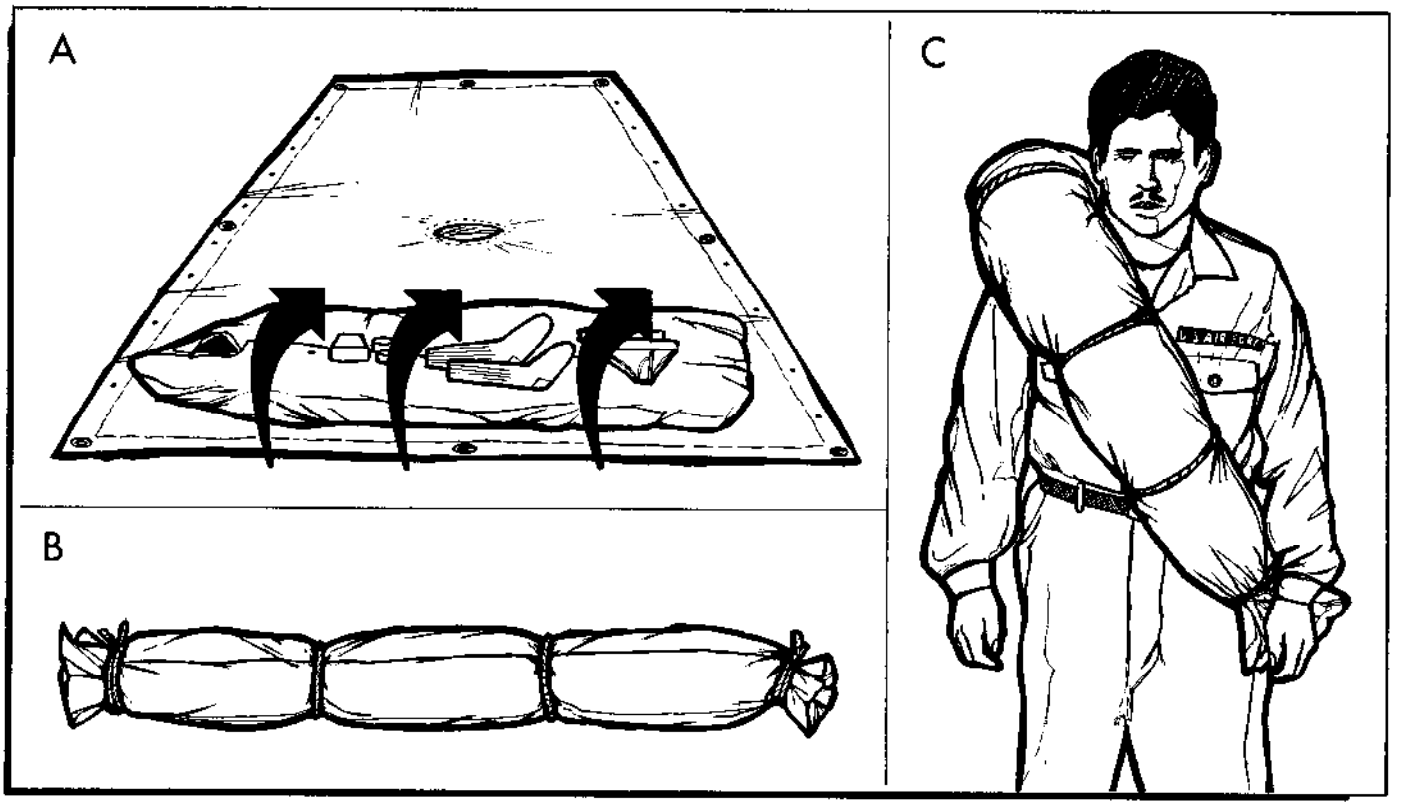

Figure 21-16. Horseshoe Pack.

constructed by tying the inner pack and outer cover simultaneously.)

(3) Horseshoe Pack. The horseshoe pack has been referred to by many names in history including the "Johnny Reb Pack," "cigarette pack," "bed roll," and others. It is simple to make and use and relatively comfortable to carry over one shoulder. It is constructed as follows (figure 21-16): Lay available square-shaped (preferably 5 feet by 5 feet) material (waterproof if available) flat on the ground (illustration A) and place all gear on the long edge of the material, leaving about 6 inches at each end. All hard items should be padded. Roll the material with the gear to the opposite edge of the square (illustration B). Tie each end securely. Place at least two or three evenly spaced ties around the roll. Bring both tied ends together and secure. This pack is compact and comfortable if all hard, heavy items are packed well inside the padding of the soft gear. If one's shoulder is injured, the pack can be carried on the other shoulder. It is easy to put on and remove (illustration C).

b. The most widely used improvised packstrap is called an Alaskan packstrap (figure 21-17). This type of packstrap can be fashioned out of any pliable and strong material. Some suitable materials for constructing the packstrap are animal skins, canvas, and parachute harness webbing. The pack should be worn so it can be released from the strap with a single pull of the cord in the event of an emergency, such as falling into water. The knot securing the pack should be made with an end readily available which can be pulled to drop the pack quickly; for example, a trucker's hitch with safety for normal terrain travel and with the safety removed when in areas of danger, such as water or rough terrain.

(1) Some advantages of the Alaskan packstrap are:

(a) Small in bulk and light in weight.

(b) Easily carried in a pocket while traveling.

(c) Quickly released in an emergency.

(d) Can be adjusted to efficiently pack items of a variety of shapes and sizes.

(e) Can be used with a tumpline to help distribute the weight of the pack over the shoulders, neck, and chest, thereby eliminating sore muscles and chafed areas.

(2) Some disadvantages of the Alaskan packstrap are:

(a) Difficult to put on (without practice).

(b) Experience and ingenuity are necessary to use it with maximum efficiency.

c. The following principles should be considered when packing and carrying a pack:

(1) The pack or burden-carrying device should be adequate for the intended job.

(2) The pack or burden may be adaptable to a pack frame. The pack frame could have a belly band to distribute the weight between the shoulders and hips and prevent undue swaying of the pack. Pack frames are also used to carry other burdens such as meat, brush, and firewood.

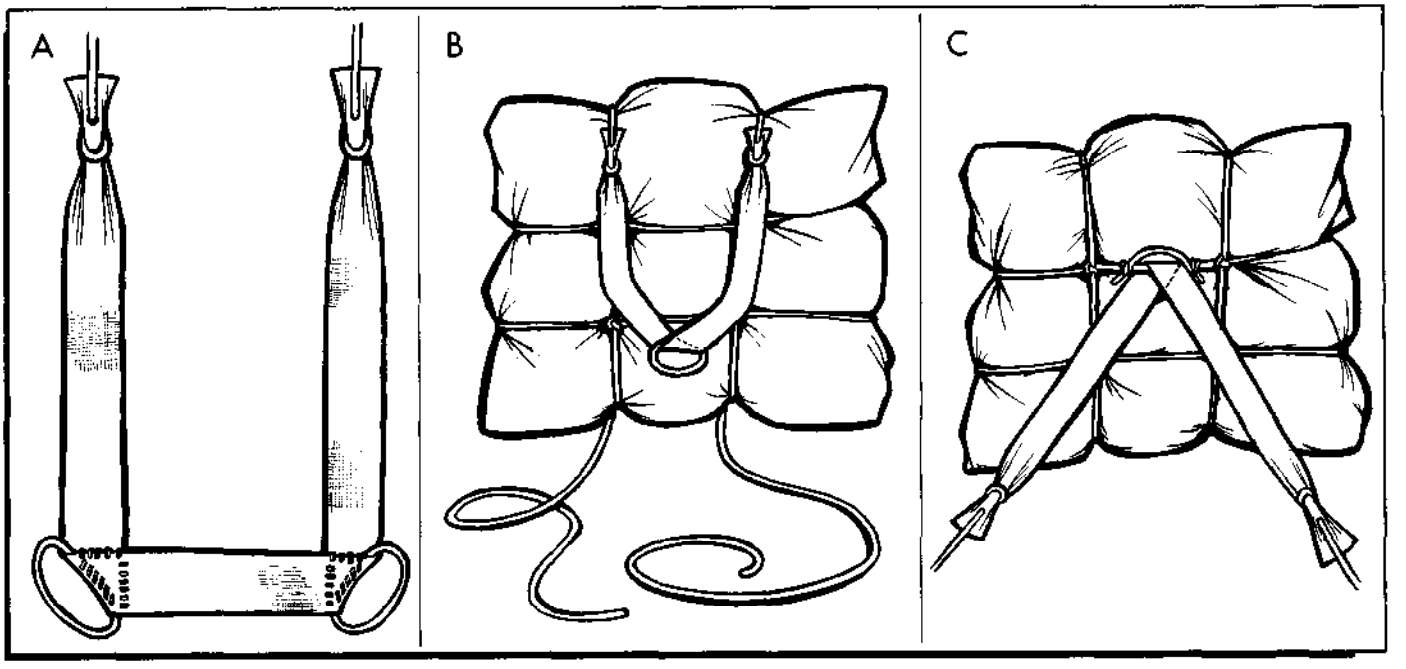

Figure 21-17. Alaskan Packstrap.

(3) Proper weight distribution is achieved by ensuring that the weight is equally apportioned on each side of the pack and as close as possible to the body's center of gravity. This enhances balance and the ability to walk in an upright position. If heavy objects are attached to the outside of the pack, the body will be forced to lean forward. A pack bundle without a frame or packboard should be carried high on the back or shoulders. For travel on level terrain, weight can be carried high. When traveling on rough terrain, weight should be carried low or midway on the back to help maintain balance and footing.

(4) Emergency and other essential items (extra and (or) protective clothing, first-aid kit, radio, flashlight, etc.) should be readily available by being placed in the top of the pack.

(5) Fragile items are protected by padding them with extra clothing or some soft material and placing them in the pack where they won't shift or bounce around. Hard and (or) sharp objects cannot damage the pack or other items if cutting edges are properly sheathed, padded, and not pointed toward the bearer. Items outside the pack should be firmly secured but not protruding where they could snag on branches and rocks.

(6) Adjust and carry the pack so that overloading or straining of muscles or muscle groups is avoided. When using a pack, the straps should be adjusted so they ride comfortably on the trapezius muscles and avoid movement when walking. Back support should be tight and placed to ensure good ventilation and support. During breaks on the trail, rest using the proper position to ease the weight of the pack and take the strain off muscles. (See figure 21-18 for methods of resting.) A comfortable pack is adjustable to the physique of the person. A waistband will support 80 to 90 percent of the weight and is fitted relatively tight. The waistband should be cinched down around the pelvic girdle/crest area to avoid constricting circulation or restricting muscle movement.

d. A tumpline is an excellent aid to burden carrying since it transfers part of the weight of the pack to the skeletal system (figure 21-19).

(1) Tumpline Construction:

(a) Attach a soft band, which rests on the upper forehead, to the pack by using light line. Make the band of any strong, soft material, such as animal skin with hair, tanned skin, an old sock, or parachute cloth. Make the band long enough to reach over the forehead and down to a point opposite each ear. A tumpline does not require any sewing.

(b) Adjust the tumpline to fit the head by making loops at the ends. It is difficult to reach down and behind to make necessary adjustments of the pack, but a person can easily reach up and adjust the pack by using the loops on either side of the forehead.

(c) Make mainstrings from rawhide or parachute suspension lines. Tie them to the lower corners of the pack, bring them up to the loops at the ends of the tumpline, and tie them with a slipknot. Experience is needed to estimate proper adjustments before putting on the pack; however, adjustments can always be made after the packstraps are adjusted.

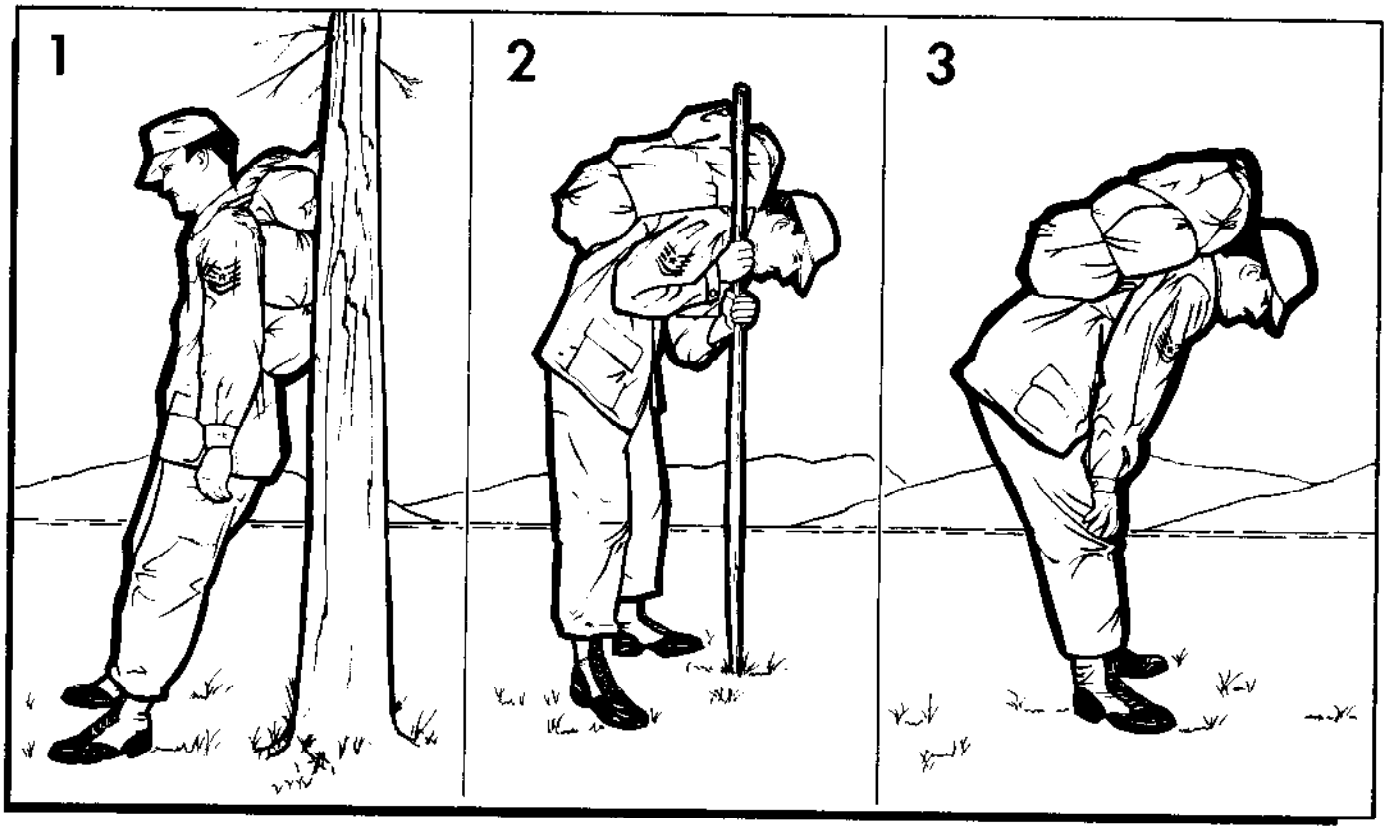

Figure 21-18. Methods of Resting.

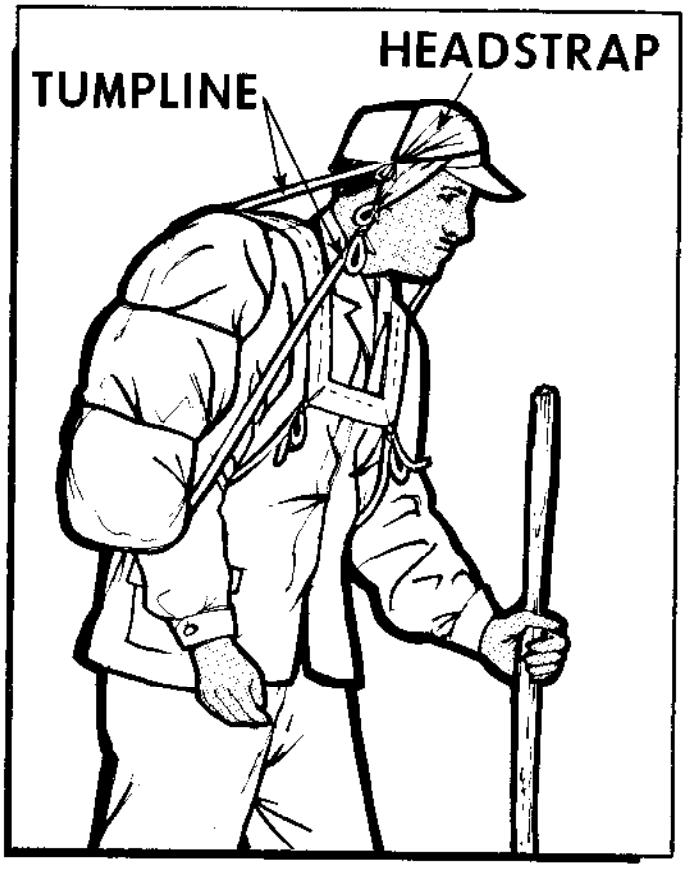

Figure 21-19. Tumpline.

(2) Tumpline Use. The tumpline should be tight enough to transfer about half of the weight of the pack/burden from the shoulders to the head. Occasionally, a heavy pack may cut off the blood circulation to the shoulders and arms. In such cases, a tumpline is of great value. By slight adjustments, most of the weight can be transferred to the head and neck, thereby loosening the shoulder straps and permitting circulation to return to numb arms. A tumpline may cause the neck muscles to be sore for the first few days due to the unusual strain placed on them; however, this discomfort soon disappears. With practice, heavy weights can be supported with only the neck and head. A tumpline can also be used to pack animal carcasses, firewood, or equipment. Since it can be rolled up and carried in a pocket, it can be a real aid to survival.

Chapter 22

ROUGH LAND TRAVEL AND EVACUATION TECHNIQUES

22-1. Introduction. There are survival situations where traveling over rough terrain is required. However, if it is necessary to travel in such areas, specialized skills, knowledge, and equipment are required. These skills may include rope management, specialized knots, belaying and climbing techniques, prusik climbing, rope bridges, rappelling, litter evacuation, etc. The amount and type of equipment needed will depend on the type of operation (climbing, evacuation, etc.).

22-2. Safety Considerations. A safety rope must be used when there is danger of the rescue team or climber falling. The rescue team leader will identify the members of the team who will climb together. Any member of the team may request to be roped during any mountain operation. The rope will be maintained as long as requested by any member. The rescue team leader will decide whether or not the entire team will rope up. Environmental factors may require a rapid retreat. During these circumstances, when speed is critical, it may be desirable to unrope. The rescue team leader's decision to unrope may be overridden by any team member who desires to remain roped. This option pertains to areas such as ridge climbs and low-grade climbs when roping is not required. Solo climbing in severe terrain is not recommended.

22-3. Climbing Ropes. Climbing ropes are made of synthetic fibers, such as nylon. The two essential qualities of a climbing rope are tensile strength and the ability to absorb the shock of a fall. This combination of elongation plus strength varies with the type of rope construction.

a. There are two basic types. The tensile strengths average 5,500 pounds, and the construction designs are called the Kernmantle and the Hauser Lay.

(1) Kernmantle rope consists of a core of braided or twisted strands protected by a braided sheath. They have a 6 to 8 percent stretch factor. Depending upon use, Kernmantle ropes are 9 to 11 millimeter in diameter; standard lengths are 150 feet and 165 feet. Ropes for wet conditions are also available.

(2) Mountain-Lay (three strand hard lay) ropes consist of a right-hand lay of three main strands twisted around each other. These ropes have a 9 to 13 percent stretch (twice as much as Kernmantle ropes) factor. They are constructed in 120-, 150-, and 165-foot lengths and are three-eights to seven-sixteenths inches in diameter.

b. Climbing ropes and anchor slings require special care. Stepping on the ropes will grind abrasive dirt particles into them which will cut the fibers of the rope. Contact between the rope and sharp corners or edges of rocks should be avoided. This may cut the rope. If the rope must run over a sharp edge, it should be padded (figure 22-1). The ropes should be kept dry because wet ropes are not as strong and collect more dirt. Do not hang ropes on sharp edges. Do not let one rope rub against another during use. Nylon rubbing on nylon will create friction which can burn through the rope(s). This is called weld abrasion. Smoking around ropes should be prohibited and they should not contact any source of excessive heat. Petroleum products will accelerate deterioration of nylon. When not in use, protect the rope from the ultraviolet rays of the Sun which also have a deteriorating effect on nylon.

c. Before constructing any rope system, the ropes should be backcoiled or layed out. This is done by simply removing the rope from the coil and, starting with one end, arranging the rope into neat overlapping loops on the ground. This inspection ensures the rope is free of knots or kinks (figure 22-1).

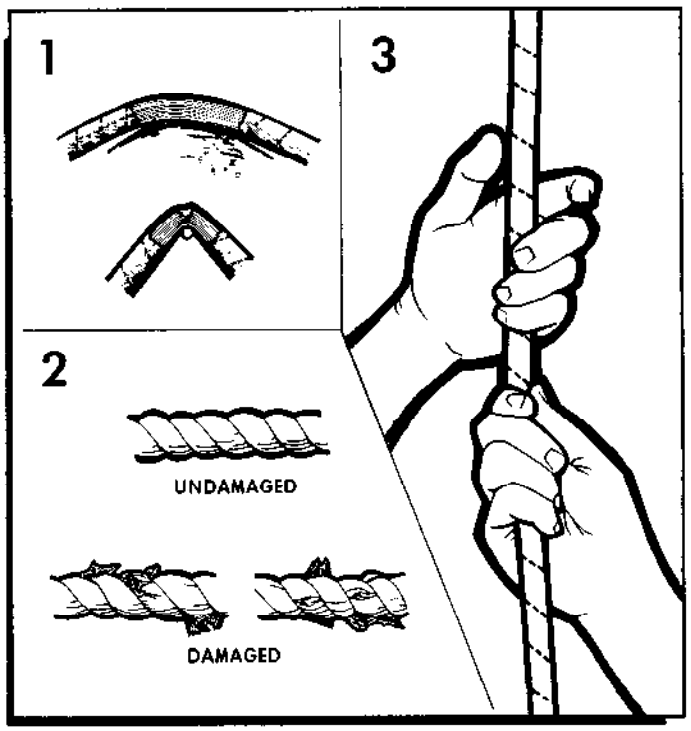

Figure 22-1. Damaged Rope.

(1) Before throwing the rope, one end should be secured. This end should be the standing end on the backcoil. Next, several small loops are taken from the working end and placed in one hand, several more loops are placed in the other hand. The throw is done by raising the hand closest to the rope end (running end)

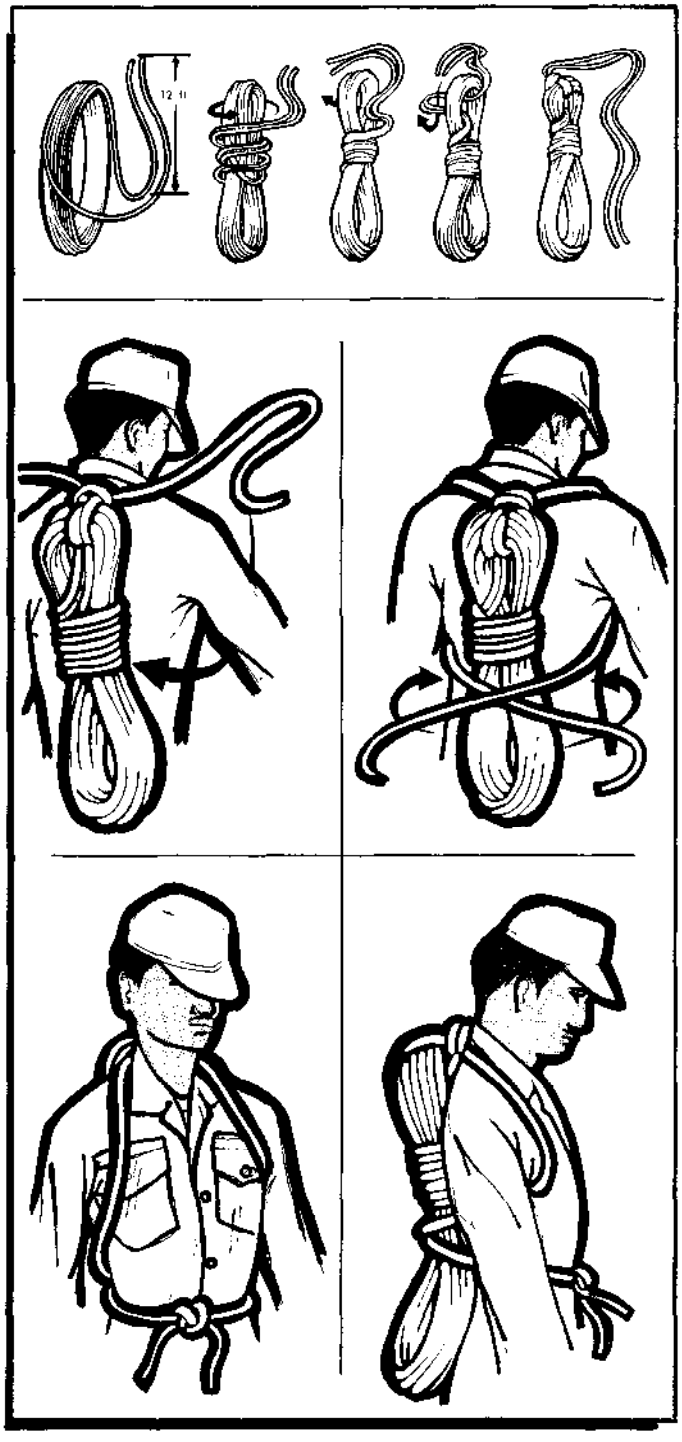

Figure 22-2. The European Coil.

over the head and throwing the loops out and down in the intended direction. The loops in the other hand are allowed to drop immediately after the running end is thrown. Properly done, the remaining rope will feed out. Before throwing the rope, the term "rope" should be sounded. This will alert anyone below to the danger of falling rope. Wait for the response "clear" if there is someone below. Under normal conditions, there should be no knots in the rope or any equipment attached to the rope due to the possibility of the rope becoming entangled if it should snag on a rock, brush, etc.

(2) Two common methods are used to coil rope:

(a) The European coil is done by starting in the center of the rope, coiling the rope into 3-foot diameter loops until there is about 12 feet of rope tail left on the two ends. The ends are then wrapped three to four times around the coil, a bight is passed through the center of the coil, and the tails passed through the bight. This is then tightened down with two remaining tails of about 6 feet. The coil can now be carried by placing the coil on the back, passing one end over each shoulder, between the arm and the chest, and passing around the back to return to the front of the body. The two ends are now secured at waist level (figure 22-2).

(b) To form a mountain coil, one end of the rope is taken in one hand while the other hand is run along the rope until both hands are outstretched. The hands are then brought together forming a loop which is transferred to one hand. This is repeated, forming uniform coils, until the entire rope is coiled. To secure the coil, a 12-inch bight is made in the standing end of the rope and laid on top of the coil. Drop the last loop on the coil and take this length of rope and wrap it around the coil and the bight. The first wrap is made at the open end of the bight to lock. Continue to wrap toward the closed end of the bight until just enough rope remains to be run through the inside of the bight. Pull the running end of the bight to secure the coil (figure 22-3). The coil may now be carried draped over a pack or over one shoulder and beneath the other.

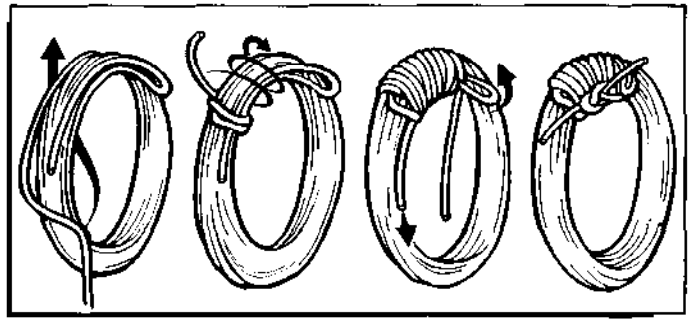

Figure 22-3. The Mountain Coil.

22-4. Specialized Knots for Climbing and Evacuation. Each of the following knots have a specific purpose. These knots have survived the test of time and are used in maintaining operations. They are designed to have the least effect on the fiber of a rope lock without slipping, and they are easy to untie when wet and icy. All knots reduce the strength of ropes; however, these knots reduce the strength of the rope as little as possible. Most

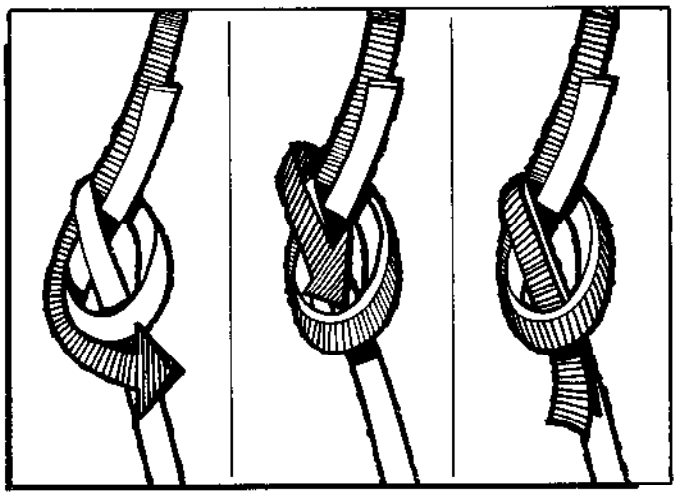

Figure 22-4. The Water Knot.

knots should be safetied with an overhand knot or two half hitches. A knot does not have to be safetied if the knot is designed for the middle of a line or if it is 15 feet or more from the end of the rope.

a. The water knot or right bend is used for joining nylon webbing (figure 22-4).

b. The double fisherman's bend is used to securely join Kernmantle or hard lay lines (figure 22-5).

c. The double figure-eight knot is used for temporarily joining Kernmantle or hard lay ropes (rappels, Tyrolean traverses) (figure 22-6A). Carabiners should be placed between lines within the knot to prevent the knot from clinching down when loaded. The figure-eight loop may be tied at the end or in the middle of a line establishing a fixed loop (figure 22-6B). If the loop is tied at the end of the line, then an overhand or single fisherman's safety knot must be used. The figure eight on a bight is used to provide two loops from a single knot to attach to two different anchors (figure 22-7).

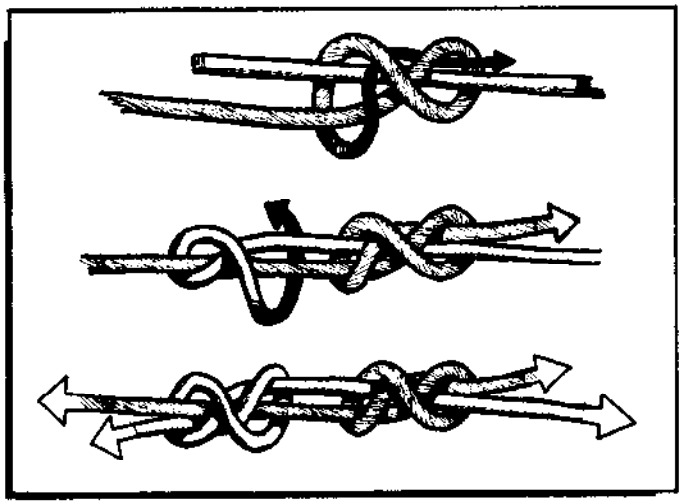

Figure 22-5. Double Fisherman's Bend.

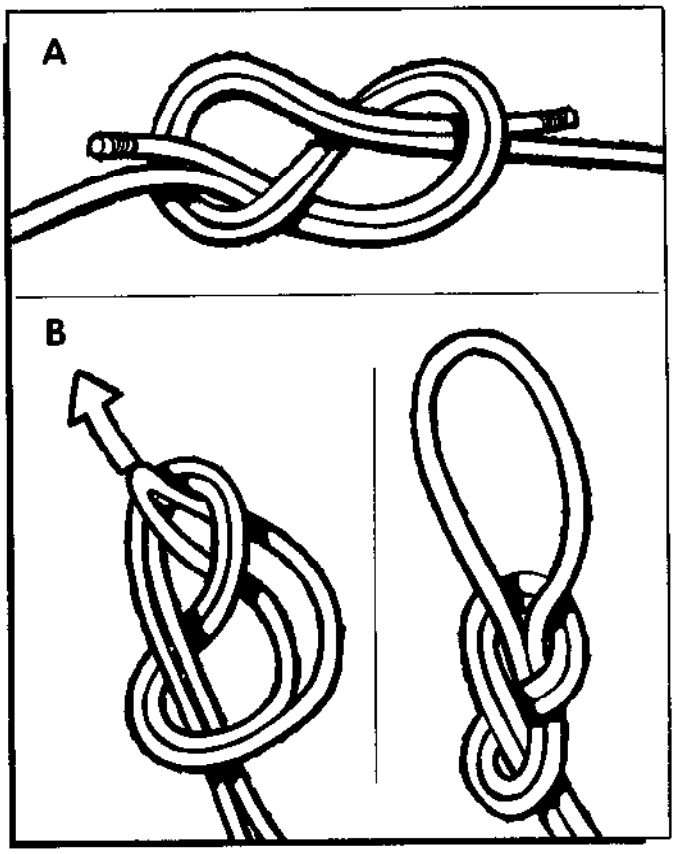

Figure 22-6. Figure-Eight Loop.

d. The butterfly is used to make a fixed loop in the middle of a line where the direction is between 120- to 180-degree angles (figure 22-8). For angles less than 120 degrees, a figure eight in a loop will suffice. For an angle of pull greater than 120 degrees, the figure eight will become weakened and begin to split.

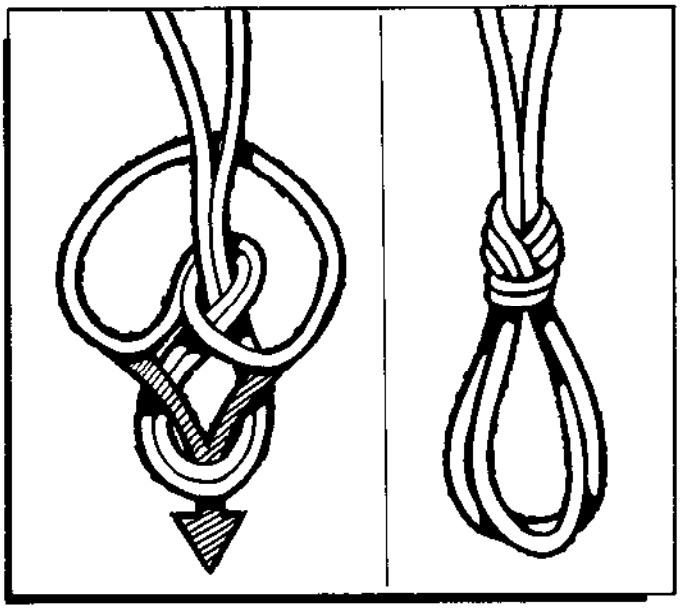

Figure 22-7. Figure-Eight with Two Loops.

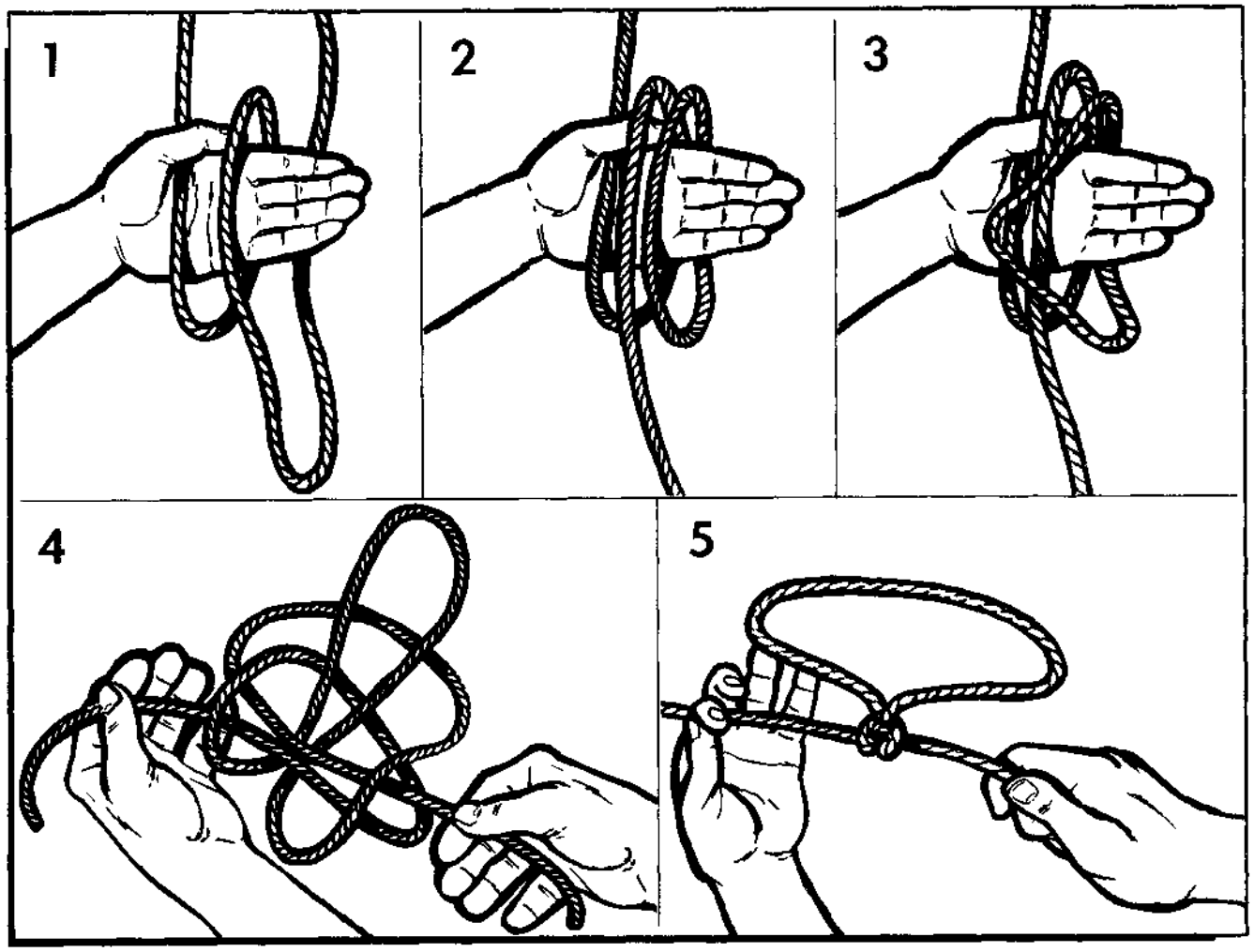

Figure 22-8. Butterfly Knot.

e. The prusik knot may be used to ascend a fixed line or safety a rappel (figure 22-9).

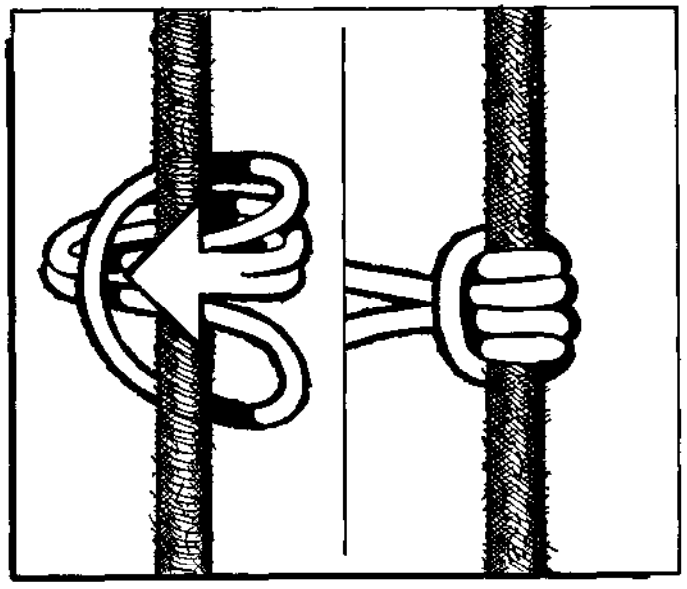

Figure 22-9. Prusik Knot.

f. A three-loop bowline is a variation of the bowline on a bight. It is used for three anchor points or as an improvised harness (figure 22-10).

g. The mariner's knot is a combination of two knots tied with a sling and a carabiner (figure 22-11) and used

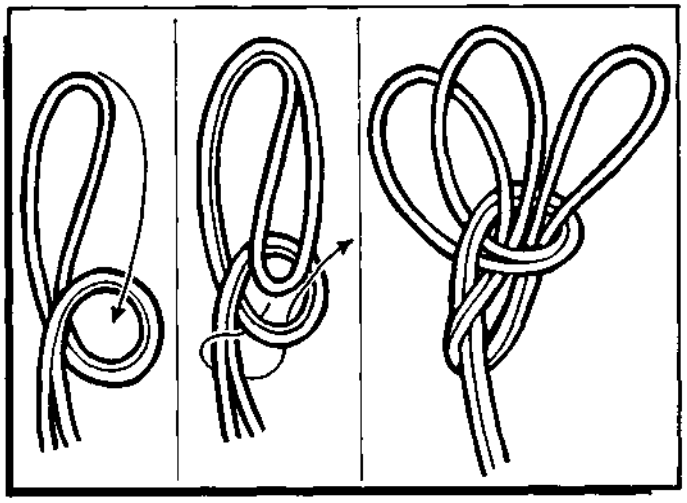

Figure 22-10. Three-Loop Bowline.

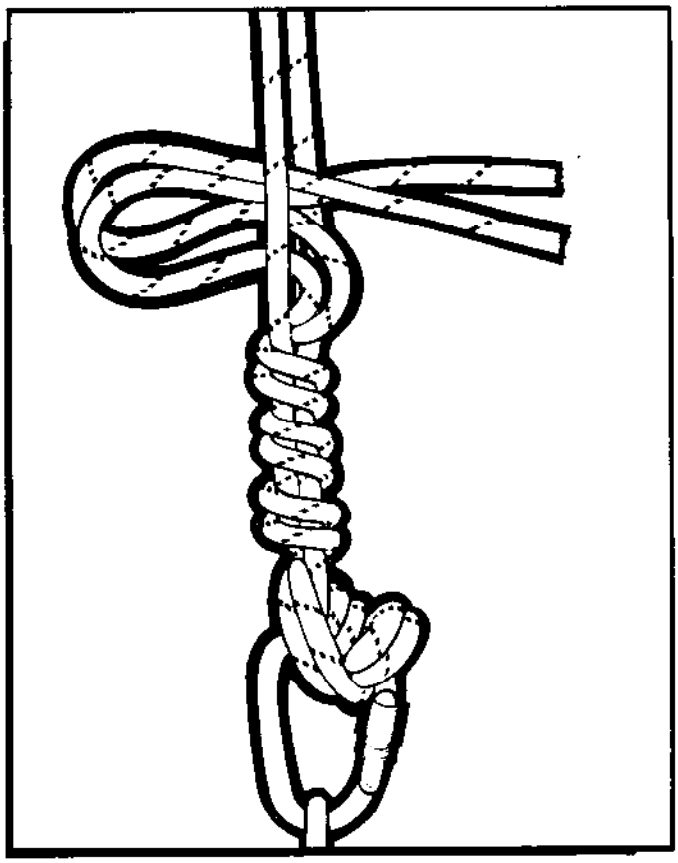

Figure 22-11. Mariner's Knot.

to transfer loads from one system to another. With a sling, a prusik is placed on the main rope (load rope). The remaining portion of the sling is then wrapped three times through the anchor carabiner and then three times around itself next to the carabiner. The remaining tail is tucked between the two sides of the sling. Tension may be gradually released by removing the wraps.

22-5. Special Equipment. Without proper equipment, safety is jeopardized and travel is impossible in severe terrain.

a. Seat Harness. The seat harness is a safety sling which is used to attach the rope to the climber or rappeller. It must be tied correctly for safety and comfort reasons. An improvised seat harness can be made of 1-inch tubular nylon tape. The tape is placed across the back so the midpoint (center) is on the hip opposite the hand that will be used for braking during belaying or rappelling. Keep the midpoint on the appropriate hip, cross the ends of the tape in front of the body, and tie half of a surgeon's knot (three or four overhand wraps) where the tapes cross (figure 22-12). The ends of the tape are brought between the legs (front to rear), around the legs, and then secured with a jam hitch to tape around the waist on both sides. The tapes are tightened by pulling down on the running ends of the tape. This must be done to prevent the tape from crossing between the legs (figure 22-12). Bring both ends around to the front and across the tape again. Then bring the tape to the opposite side of the intended brake hand and tie a square knot with an overhand knot or two half-hitch safety knots on either side of the square knot. The safety knots should be passed around as much of the tape as possible (figure 22-12). Once the seat harness has been properly tied, attach a single locking carabiner to the harness by clipping all of the web around the waist and the web of the half surgeon knot together. The gate of the carabiner should open on top and away from the climber.

b. Commercial Seat Harness. A commercial seat harness is a climbing harness which is worn around the pelvic girdle and is used with climbing ropes for descending and ascending devices and for rescue evacuations. The harness is easily donned, comfortable, allows freedom of movement, and evenly distributes the body

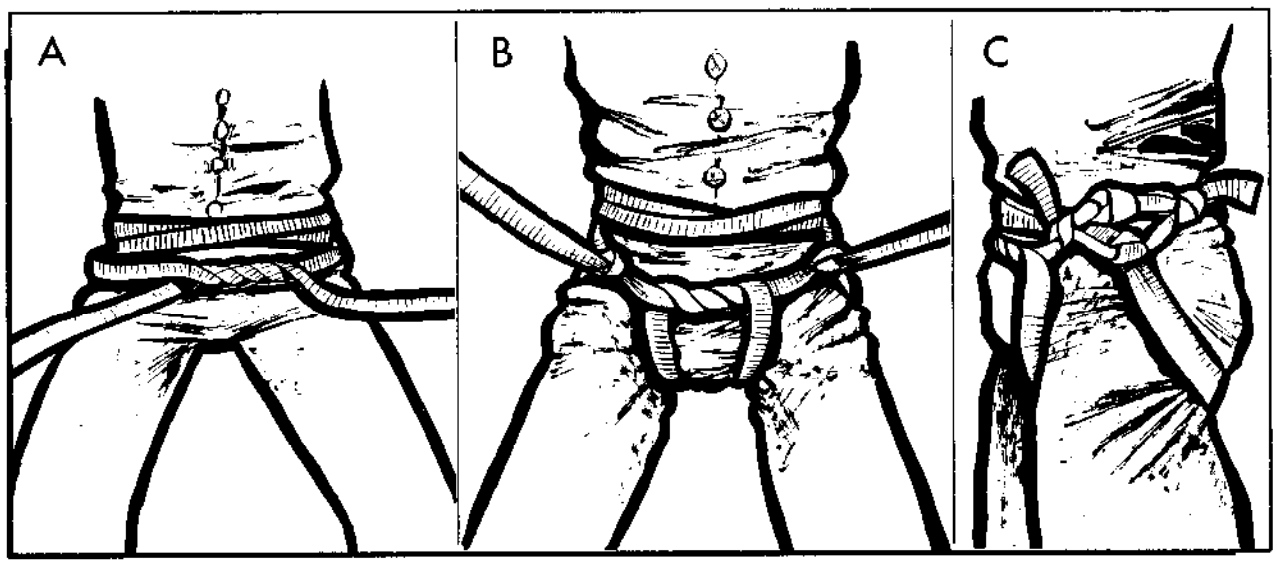

Figure 22-12. Improvised Seat Harness.

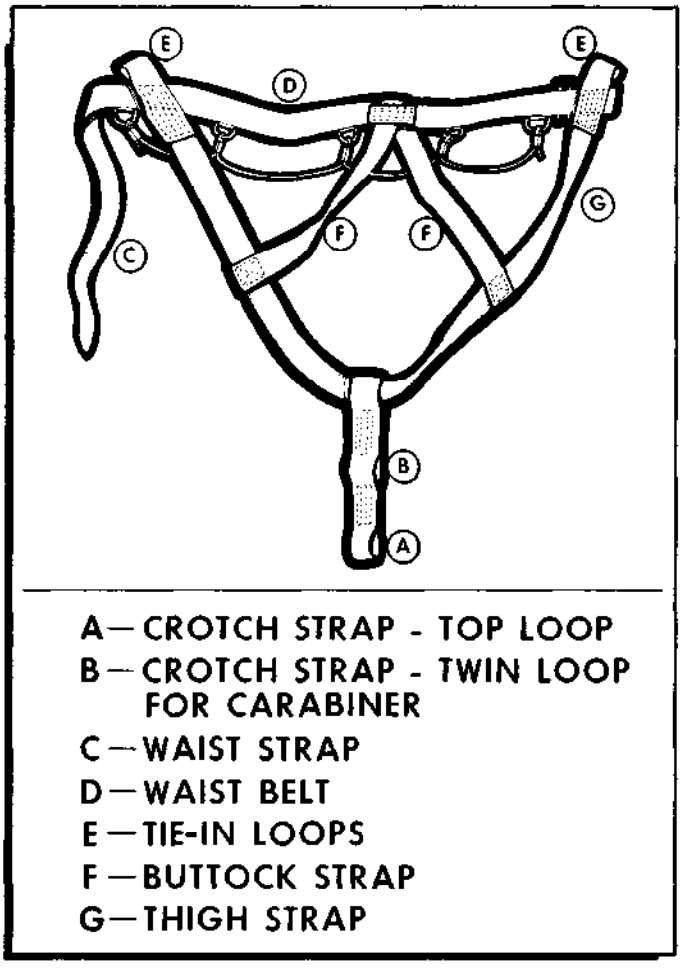

Figure 22-13. Sit Harness.

weight to the pelvic region. Figure 22-13 identifies the parts of the harness. The seat harness webbing should be cared for in the same manner as all webbing. The plastic buckle should not be subject to direct jars since it may fracture or establish a weak point on the buckle which can fail when used. It is primarily constructed of nylon webbing with box stitching at juncture points. Small metal rings are sewn on the lower half of the waist belt and connected with 300-pound line for equipment storage. The harness should have tie-in loops. Webbing should not have tears, rips, burns, or have been subjected to chemical exposure. The plastic buckle should not have fracture marks. All seams and thread should be intact. Harnesses showing signs of excessive wear, tear, rips, burns, or exposure to chemicals should be removed from service. Fractured buckles should be replaced.

c. Improvised Chest Harness. The Parisian bandolier or chest harness is a secondary safety sling used to attach to a safety sling at the top of climbs or for top belaying on high-angle rappels. A Parisian bandolier is made from a continuous loop of webbing. The loop should be 3 to 4 feet long. The knot connecting the two ends of tape together is a water knot with an overhand knot or half hitch as safeties. To don the bandolier, place one arm through the loop and bring the running end of the loop behind the back and under the opposite arm. A sheet bend is tied in the center of the chest by using the center of the tape which was passed under the climber's arm and inserted through the portion of bight formed by the tape that goes over the shoulder (figure 22-14).

d. Climbing Helmet. Climbers should wear this hard shell helmet which is designed to minimize injury dur-

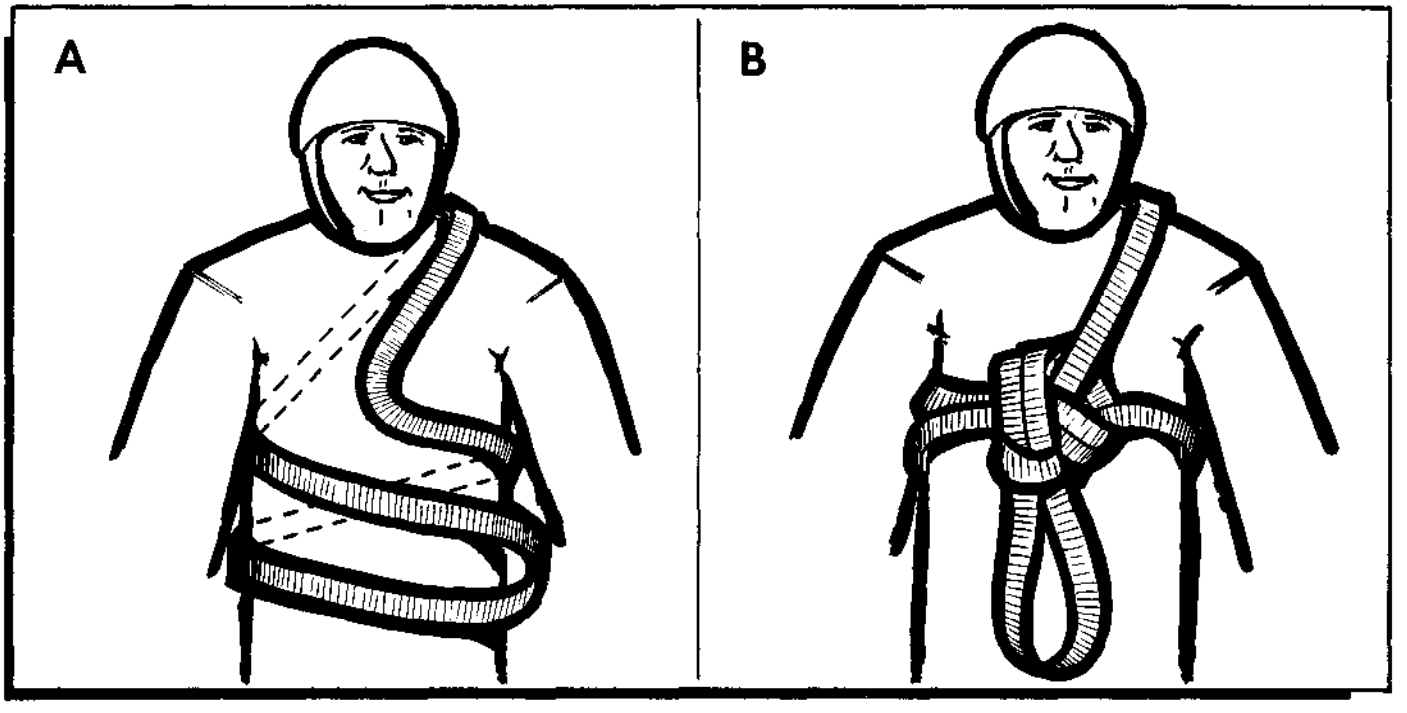

Figure 22-14. Parisian Bandolier.

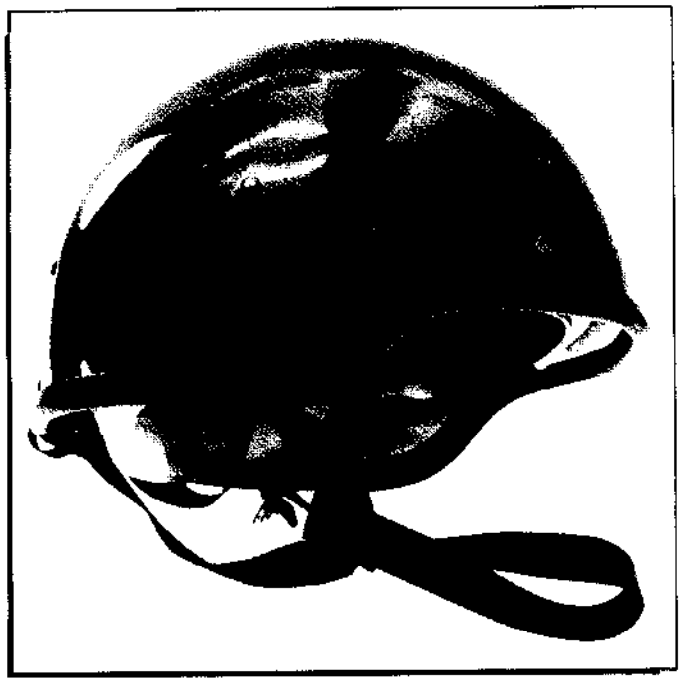

Figure 22-15. Climbing Helmet.

ing falls or when struck by a small object falling from above their position. The helmet consists of a hard shell which is held away from the head by a suspension system. The suspension system absorbs a portion of the blow to the top of the helmet. Helmet fit should be such that the side of the head is protected. A "Y" style strap system of lightweight webbing retains the helmet on the head better than a single strap. The headband is adjustable to allow donning over a watch cap in cold weather (figure 22-15). The hard shell should not have cracks or breaks. The suspension system must be securely riveted in place. The straps must be free of cuts, frays, and securely fastened to the helmet.

SIZE	MOUNTAIN-LAY	KERNMANTLE	NYLON WEB
5/16" OR 7 MM WIDE	2700	2300	
3/8" OR 9 MM WIDE	3700	3200	
7/16" OR 11 MM	5500	4800	
1/2" WIDE			1100
9/16" WIDE			1700
1" WIDE			4000

Figure 22-16. Breaking Strength of Rope.

CARABINER BREAKING STRENGTH		
TYPE	MINIMUM	MAXIMUM
STANDARD OVAL, ALLOY	3890	4210
STANDARD D, ALLOY	4065	5515
LOCKING D, ALLOY	4960	6310
LOCKING D, STEEL	GREATER THAN 11,000 POUNDS	

Figure 22-17. Breaking Strength of Carabiners.

e. Gloves. Gloves should be worn when belaying or rappelling. Since these techniques are performed frequently, the gloves should be attached to the climber (rescuer) at all times.

f. Slings. Slings have many uses in adverse terrain work—construction of anchor systems, stirrups to aid movement, attachment to chocks, improvised harness-

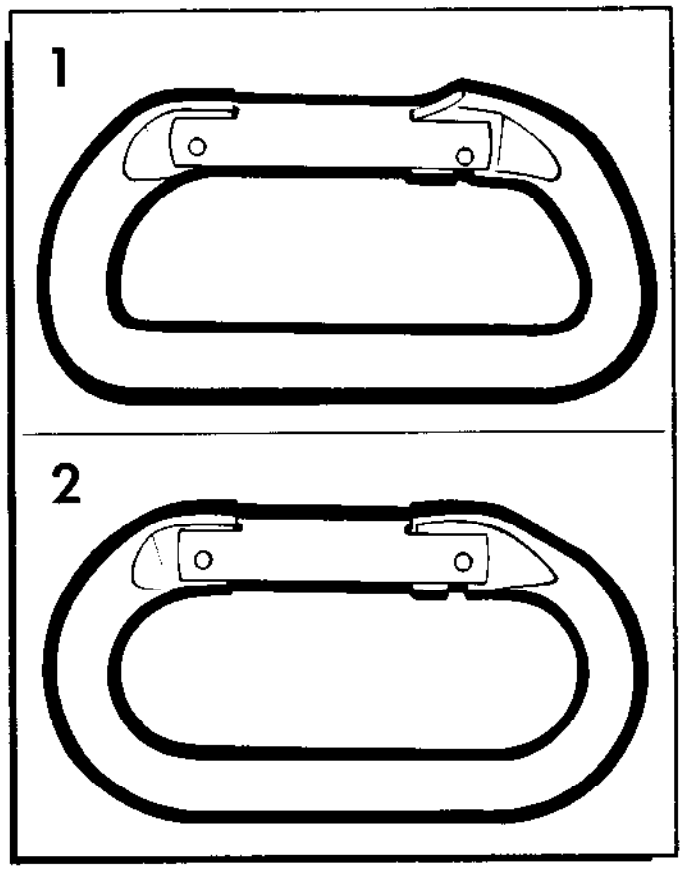

Figure 22-18. Types of Carabiners.

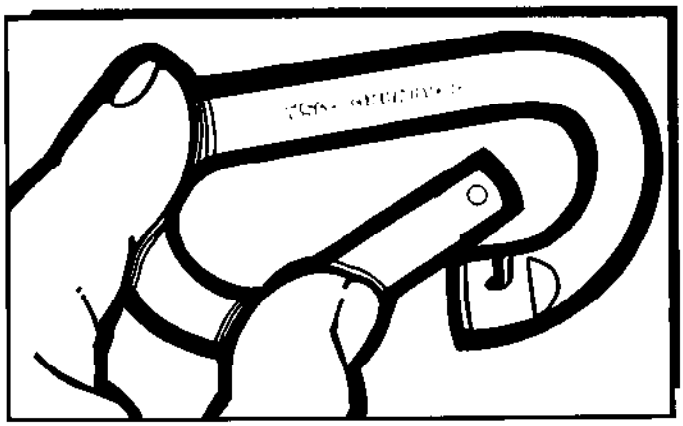

Figure 22-19. Nonlocking Carabiners.

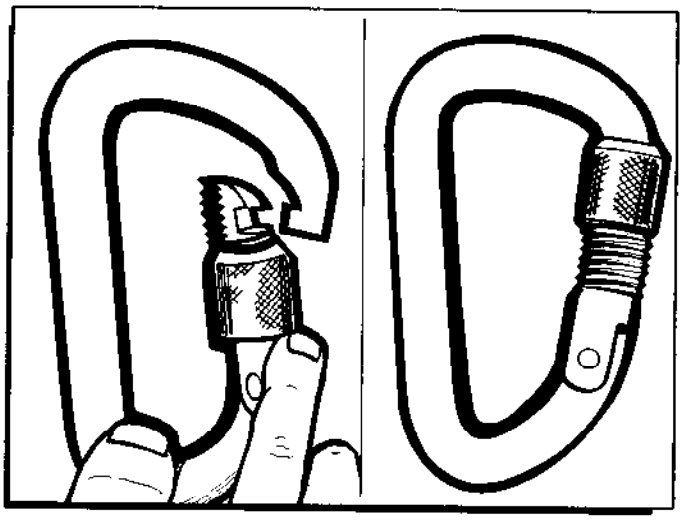

Figure 22-20. Large Locking Carabiner.

es, and to attach equipment to rope systems. Slings may be constructed from Mountain-Lay, Kernmantle, and flat or tubular nylon. The standard length for long slings is 13 feet with smaller slings cut to the individual task. For climbing, slings about 4 feet long, giving a circumference of 2 feet, are acceptable. Sling rope diameters for Mountain-Lay or Kernmantle are 5 mm, 7 mm, and 9 mm while tape slings are usually one-half inch or 1 inch wide. During construction of any system or slinging of chocks, the largest possible rope diameter or tape width should be used. Mountain-Lay and Kernmantle ropes used for slings are the same as climbing ropes. Nylon webbing is constructed of multiple strands of nylon sewn in a weave pattern throughout the length. It should be constructed in flat or tubular shapes. Specific strengths are noted in figure 22-16. Nylon webbing which is burned, torn, cut, or frayed should be removed from service. If nylon webbing is suspected to be inferior due to shock, loading, or abuse, it should not be used.

g. Carabiners. The carabiners used by climbers (rescuers) fall into two material designs—aluminum alloy and chrome vanadium steel. Both material designs may be either locking or nonlocking. Alloy metals provide the maximum strength for adverse terrain work. As with ropes, carabiners are rated differently depending on the manufacturer. Figure 22-17 depicts the breaking strengths for the various types of carabiners. Aluminum alloy and chrome vanadium steel carabiners will be oval-shaped or D-shaped (figure 22-18).

(1) Alloy standard oval (nonlocking) is shaped in an oval with a gate to allow access into the center of the oval. The gate contains a locking pin which fits into the locking notch when the gate is closed (figure 22-19). The alloy D-shape (nonlocking) is shaped in the fashion of a D which allows greater distribution of weight applied to the longer side. Features and operation are the same as the standard oval.

(2) Alloy locking D or oval are both constructed the same as previously discussed but are machined with threads on the gate and a sleeve which, when screwed clockwise, will cover the locking notch and pin for a positive lock. The steel locking D is the same construction as the alloy locking D, but is made of steel and is normally used for mountain rescue work (figure 22-20).

(3) Carabiners should not be dropped or used for other than the designed purpose since small fracture lines may develop and weaken the structure. Carabiners should not be used as a hammer nor loaded (stressed)

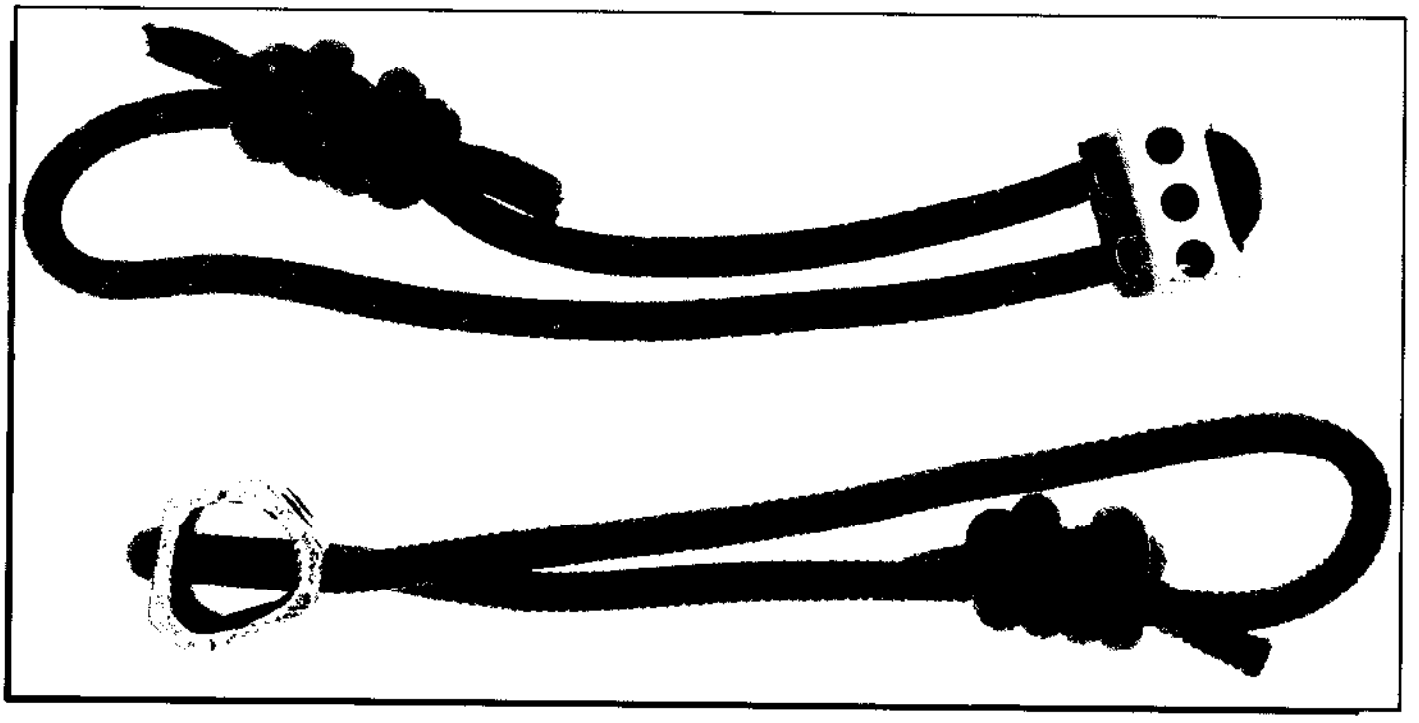

Figure 22-21. Hexcentric Chock with Kermantle Sling.

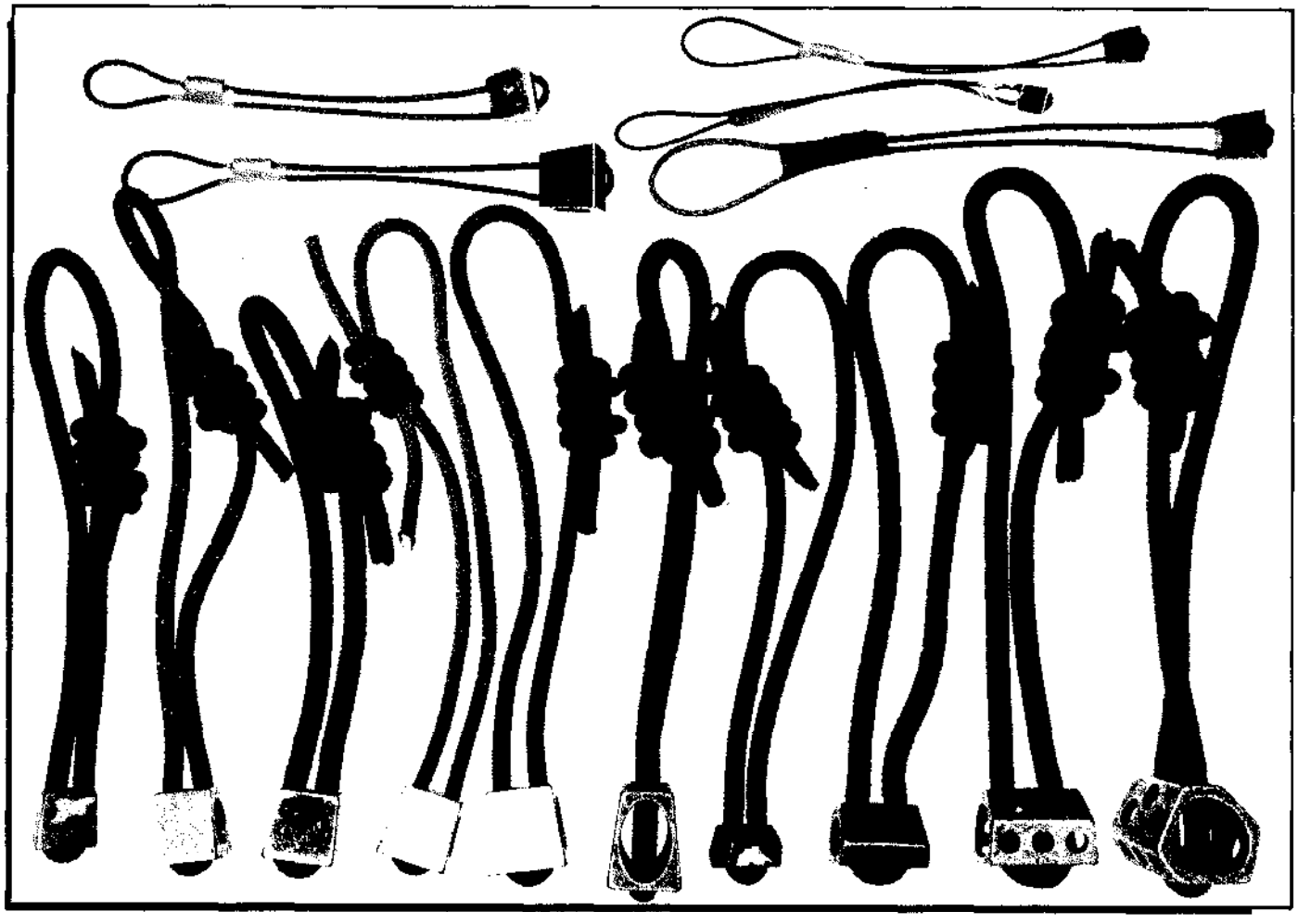

Figure 22-22. Chocks.

beyond their maximum breaking strength. The moving parts, hinge and sleeve, of locking carabiners should be kept clean for free movement. If a carabiner "binds," do not oil it — discard it! Carabiners should not be filed, stamped, or marked with an engraving tool. Colored tape or Teflon paint may be used to identify carabiners. All moving parts (gate, locking sleeve) should operate freely and the locking pin must properly align with the locking notch. Obvious fractures, regardless of size, are cause for condemning a carabiner.

h. Chocks:

(1) Chocks are metal alloy or copper shapes with unequal sides which are placed within cracks in rocks to serve as anchor points or parts of a protection system. Figures 22-21 and 22-22 show the common chocks used. Each is designed to fit within a variety of cracks in rocks. Sizes vary from one-sixteeth inch thick by one-fourth inch wide to several inches thick and wide. The various shapes with uneven sides allow one size to fit many rock openings. Carrying different sizes of chocks allows a person to choose the most suitable chock for the job. Depending on the style, chocks are constructed differently and require a sling for use.

(a) Copperhead. A cylindrical copper chock ranging from three-sixteeths inch in diameter by one-half inch long to one-half inch in diameter by 1 inch long. The relatively soft head bites well into rock and has

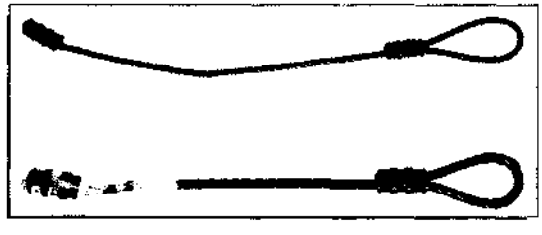

Figure 22-23. Copperheads.

little rotation after placement. It is manufactured with a cable sling attached (figure 22-23).

(b) Hexentric. A metal alloy chock constructed with six sides. Each side is a different length than the opposing side (figure 22-21). The two ends are tapered, gradually getting smaller from the back to the front of the chock. The front of the hexentric chock is the narrowest of the faces. Small hexentric chocks are manufactured with wire slings and larger ones have two holes bored through the front to back for threading of a Kernmantle sling. Kernmantle slings vary with the size of the chock, from 5 mm to 9 mm. The largest sling size possible will be used with each chock.

(c) Wedge. A solid alloy metal chock shaped in the form of a wedge. All four sides decrease in size from back to front. As with small hexentrics, small wedges are manufactured with wire slings while larger wedges require Kernmantle slings.

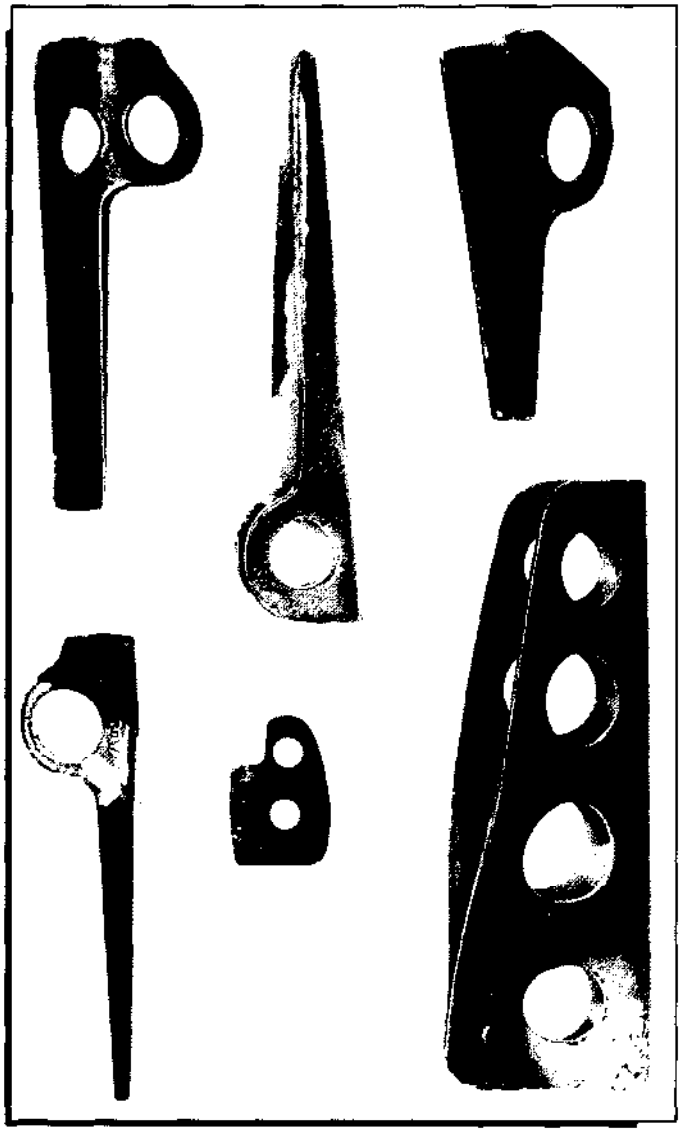

Figure 22-24. Pitons.

(2) All chocks should be treated the same as carabiners. However, chocks will take considerable abuse since they are wedged in rock, and at times, the leverage of a hammer may be required to remove them. Chocks manufactured with wire slings and wedges should be closely monitored for security. If there is doubt about the condition of a wedge, it must not be used. Rope or tape slings on chocks must meet inspection requirements. Cracks in any surface of the chock is cause to remove it from service. Small scratches or grooves on the outer surfaces are not cause for removal from service.

i. Pitons. Pitons are metal alloy spikes of various shapes, widths, and lengths which are hammered into cracks in rock surfaces for anchor points or as part of a protection system. The different shapes are identified by name; each name has the same shape but comes in various sizes in length and width. Figure 22-24 depicts the types of pitons. Pitons are divided into two different categories—blades, whose holding power results from being wedged into cracks; and angles, which hold by wedging and blade compression.

(1) Blades comprise those pitons which have flat surfaces and range in size from knife-blade thickness to one-half inch.

(2) Angles are those which have rounded or V-shaped blades and allow the sharp edges to cut into the edge of a crack. Sizes vary from a short "shallow angle" of one-half inch thick and 2½ inches long to "bongs" (large angles of 4 inches in width). The leeper, a special Z-shaped angle, is designed to give greater holding power due to its cutting ability and blade compression.

(3) Pitons are made of chrome-molybdenum alloy metal which has a high strength versus weight ratio. Multiple holes further reduce the weight but do not compromise their strength. Pitons are virtually indestructible. However, they should only be used for the designed purpose. Damage may occur during the placement of a piton by ineffective or off-center hammering. Pitons must not have cracks or be bent from the shape of manufacture. Any piton which is suspected of cracks or bends should be removed from service.

j. Figure-Eight Clog. A mechanical friction device used for belaying, rappelling, and breaking while lowering personnel and equipment. The device is formed in a figure-eight (figure 22-25) with two different size openings comprising the inner holes. The rope is passed through the larger hole of the "8" and over the connecting portion of the device. The smaller hole is attached to a carabiner. The figure-eight clog is manufactured in various sizes.

k. Protection System (Belay System). This system is commonly referred to as a belay, securing of a climber tied to the end of a rope by a stationary second climber (figure 22-26A and 22-26B).

(1) The major components of the belay system are (figure 22-27):

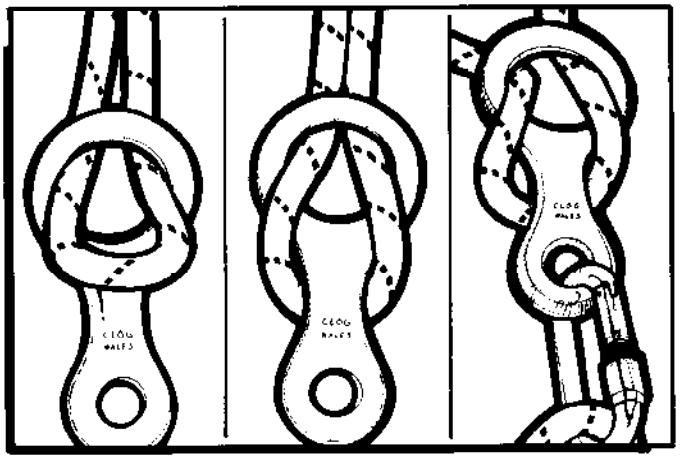

Figure 22-25. Figure-Eight Clog Direct Belay Installation.

(a) An anchor—secure point to which the belayer is attached.

(b) Belayer—individual responsible for the security of the climber.

(c) Intermediate protection point(s)—placement of slings, chocks, or pitons along the route of climb by the lead climber. The climbing rope is threaded through carabiners attached to the devices and the climbing rope clipped into the carabiner.

(d) Climbing rope size—11 mm Kernmantle or seven-sixteenths inch Goldline.

(2) The minimum equipment for a belay system is divided between rescue team and personal equipment. Rescue team equipment is comprised of the items necessary for the climbers to reach the objective. These are the climbing rope and climbing hardware, including chocks, pitons, slings, and carabiners. Individual equipment is comprised of items which allow each climber to perform belaying and climb. These equipment items are climbing helmet, seat harness, climbing boots, and gloves (worn while belaying or rappelling only).

22-6. The Successful Climb. The successful accomplishment of a climb is based on the strict application of basic principles and techniques.

a. Route Selection. Route selection can be the deciding factor in planning a climb. A direct line is seldom the proper route from a given point to the area of the survivor. Time spent at the beginning of the climbing operation in proper route selection may save a large amount of time once the operation has started. The entire route must be planned before it is carried out, with the safest route selected. Natural hazards present,

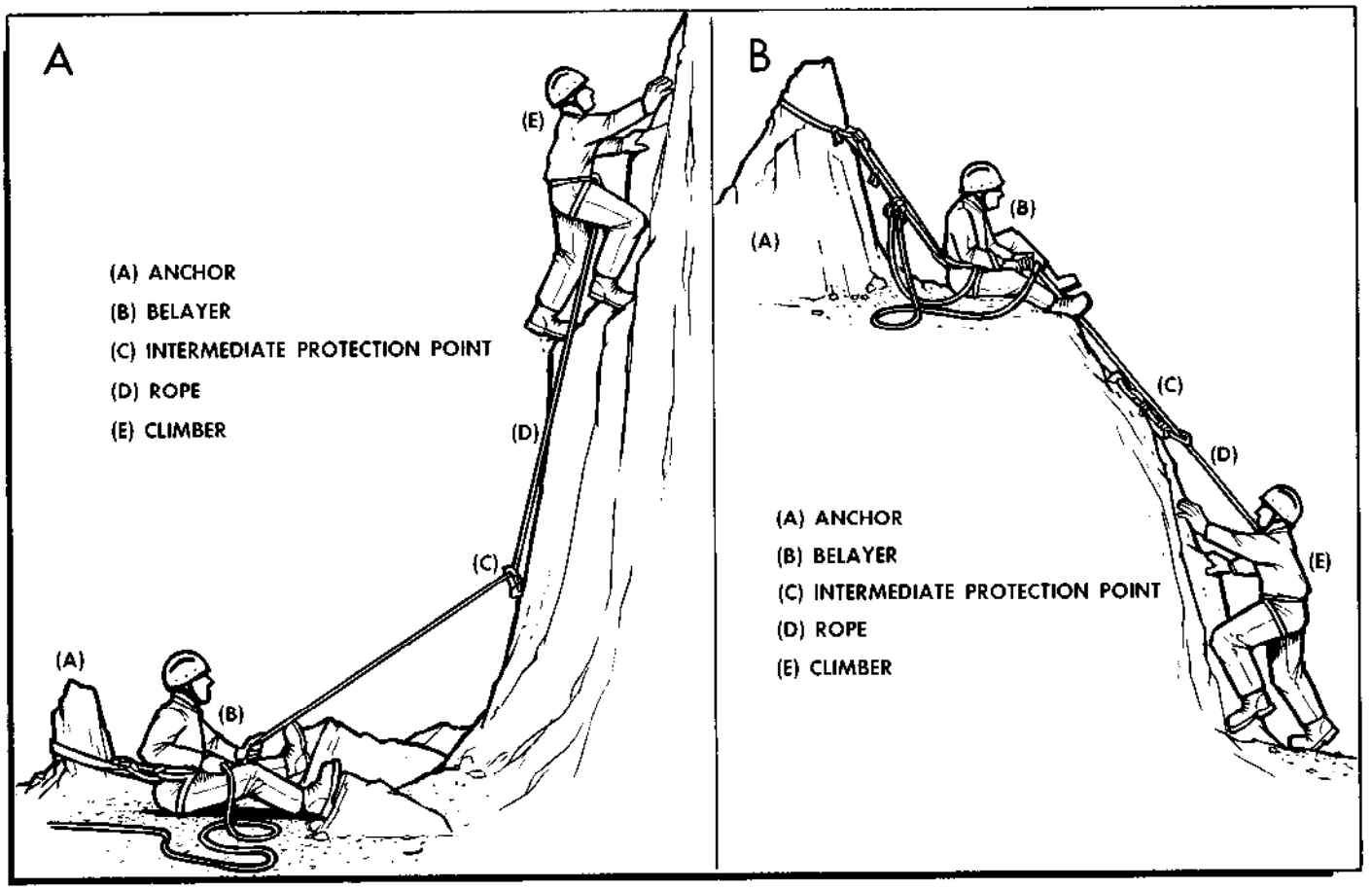

Figure 22-26. Belay System (A) Bottom (B) Top.

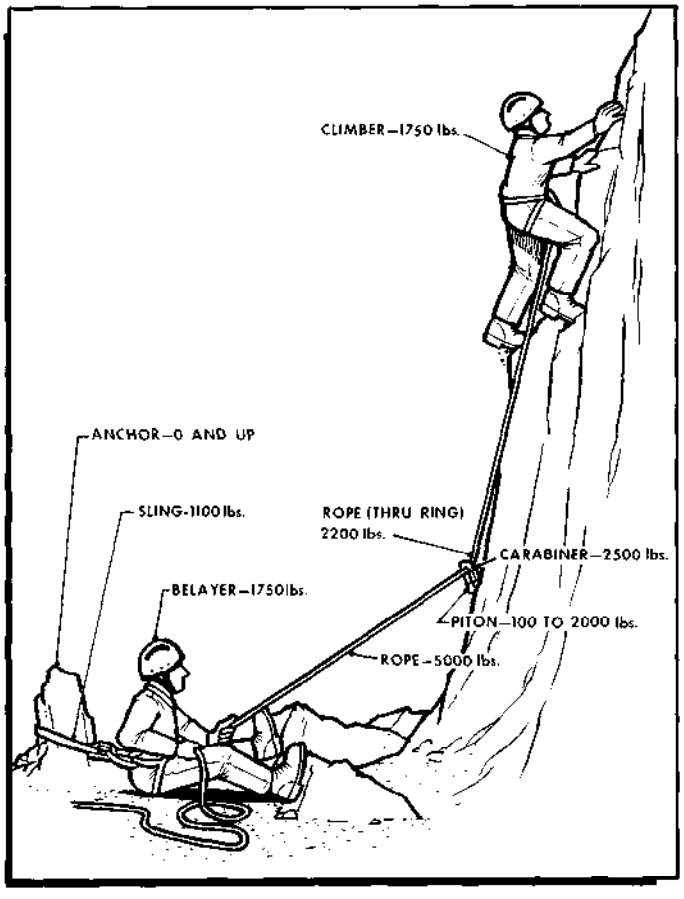

Figure 22-27. Belay Chain.

retreat routes available, time involved to perform the climb, and logistics will be major influencing factors in selecting the route.

b. Lead Climber Techniques. The lead climber is responsible for following the preplanned route of climb and altering the route as necessary. If the route of climb must be altered, consideration must be given to the experience of the climber(s). Climbing moves which are considered easy for the leader may be difficult for the second climber. Additionally, lead climbers must provide for their own personal protection during climbs to minimize the distance of a possible fall. Each intermediate protection point placed must be secure and follow as direct a line as possible.

c. Protection Placement. The basic principle to placement of protection is to decrease the distance of a fall. If the protection is not placed at frequent intervals, the climber's descent during a fall will equal twice the distance from the climber to the belayer. Slack in the belay system and rope stretch will slightly increase the distance. If a leader climbed 90 feet above the belayer without adding protection, the resultant fall will be 180 feet plus slack and stretch (figure 22-28). The belayer is helpless to stop the leader's fall in this situation. However, providing protection at intervals (figure 22-28) of 15, 20, and 25 feet, for example, will decrease the distance of a leader's fall to a little over 30 feet, making it easier for the belayer to stop the fall. Preferably, protection should be placed at about 10-foot intervals during all climbing operations to minimize the distance of a fall.

(1) The leader can reduce the drag on the rope by climbing in as straight a line as possible through protection points. The zigzagging of the climbing rope through protection points widely spaced or at abrupt angles (figure 22-29) will increase rope drag. Ideally, when points of protection are separated, a sling should be added to the anchor to keep the climbing rope in a straight line (figure 22-29). The attachment of a sling to an intermediate protection device is called a runner.

(2) Chocks on wire should always have a runner placed on the wire as shown in figure 22-30. The runner reduces the direct rope movement of the wired chock, thereby reducing possible dislodgment. For all wired and roped chocks, runners should be clipped into a carabiner at the chock, and the climbing rope clipped into an additional carabiner.

(3) Pitons may also be extended in a similar manner to that used for chocks. If sufficient runners are not available on long climbs, a chock sling may be used. The climbing leader must correctly thread the climbing rope through the carabiners attached to the piton. The carabiner should open either down and out or toward the belayer. Further, the rope running through the carabiner should run from the inside to the outside to prevent binding or the carabiner gate from opening.

d. Leader Belay. The length of the climbing rope or the length of the route climbed will determine when a belay will be established by the leader. If the route is longer than a rope length, a belay must be established by the leader to protect the first belayer's ascent to the leader's position. The belay chain is established as shown in figure 22-31. The only alteration necessary to the belay chain would be if the second climber continued the climb as the leader once the belayer's (first leader's) position is reached. This leap frogging of the leader is referred to as "climbing through". The sequence for the climbing through method begins when the first leader reaches the end of the climbing rope length. The leader selects a suitable belay position when the belayer (second climber) states that approximately 20 feet of rope remain. With a suitable site selected, one that has anchor and an area for a standing or sitting belay, the lead climber belays. The second climber then climbs up to the lead climber. Once reaching the lead climber, the second climber assumes the lead climber's role and continues the climb.

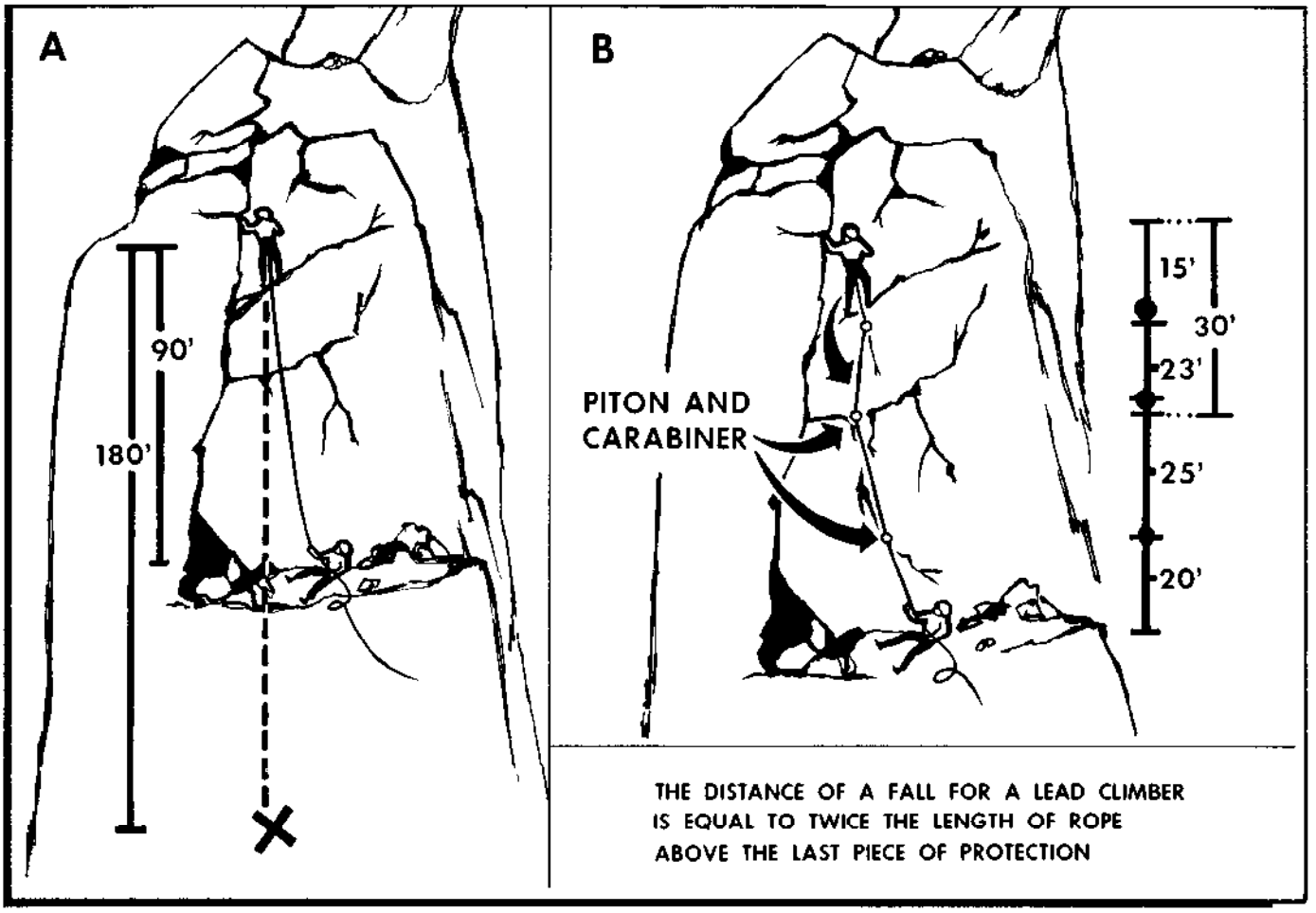

Figure 22-28. Protection Placement.

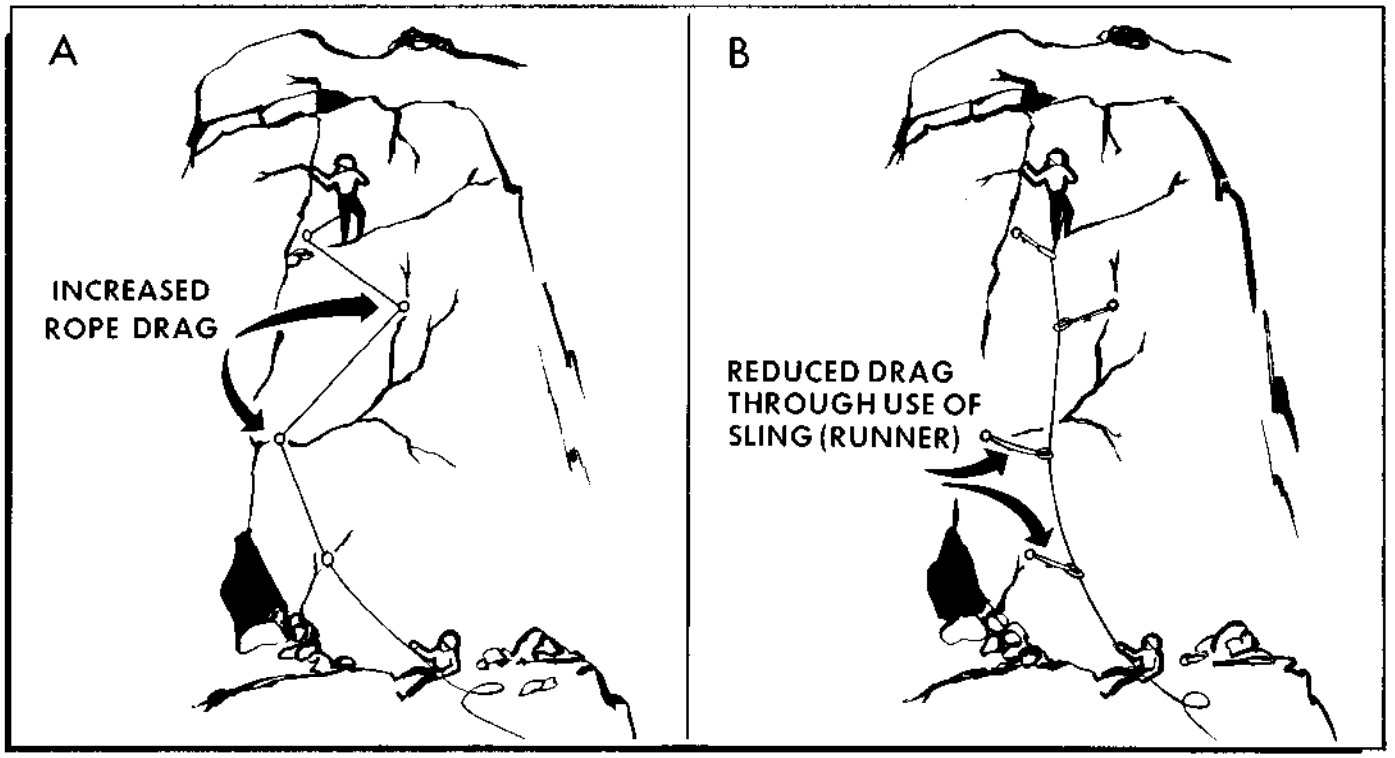

Figure 22-29. Direct Placement.

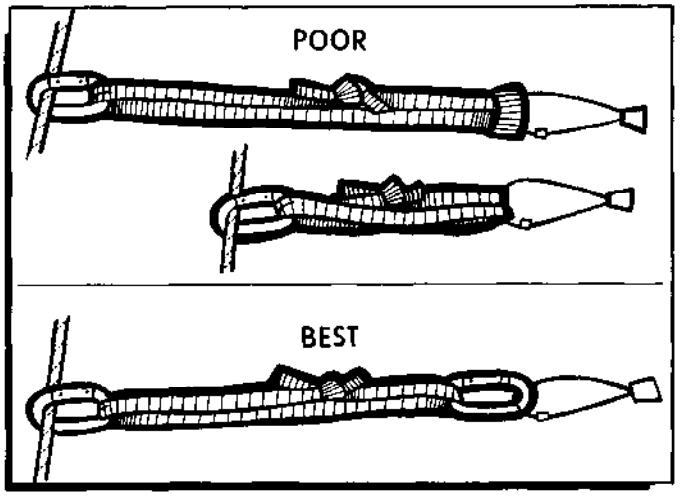

Figure 22-30. Runner Placement on Chocks.

(1) Terrain must be analyzed to find an efficient route of travel. The rescuer (climber) must make a detailed reconnaissance, noting each rock obstacle, the best approach, height, angle, type of rock, difficulty, distance between belay positions, amount of equipment, and number of trained rescuers needed to accomplish the mission on or beyond the rocks. If the strata dips toward the rescuer, holds will be difficult as the slope will be the wrong way. However, strata sloping away from the rescuer and toward the mountain mass provides natural stairs with good holds and ledges.

(2) At least two vantage points should be used so a three-dimensional understanding of the climb can be attained. Use of early morning or late afternoon light, with its longer shadows, is helpful in this respect. Actual ground reconnaissance should be made, if possible.

e. Dangers to Avoid:

(1) On long routes, changing weather will be an important consideration. Wet or icy rock can make an otherwise easy route almost impassable; cold may reduce climbing efficiency; snow may cover holds. A weather forecast should be obtained if possible. Smooth rock slabs are treacherous, especially when wet or iced after freezing rain. Ledges should then be sought. Rocks overgrown with moss, lichens, or grass become treacherous when wet. Under these conditions, cleated boots are by far better than composition soles.

(2) Tufts of grass and small bushes that appear firm may be growing from loosely packed and unanchored soil, all of which may give way if the grass or bush is pulled upon. Grass and bushes should be used only for balance by touch or as push holds—never as pull holds. Gently inclined but smooth slopes of rock may be covered with pebbles that may roll treacherously underfoot.

(3) Ridges can be free of loose rock, but topped with unstable blocks. A route along the side of a ridge just below the top is usually best. Gullies provide the best protection and often the easiest routes, but are more subject to rockfalls. The side of the gully is relatively free from this danger. Climbing slopes of talus, moraines, or other loose rock are not only tiring to the individual but dangerous because of the hazards of rolling rocks to others in the party. Rescuers should close up intervals when climbing simultaneously. In electrical

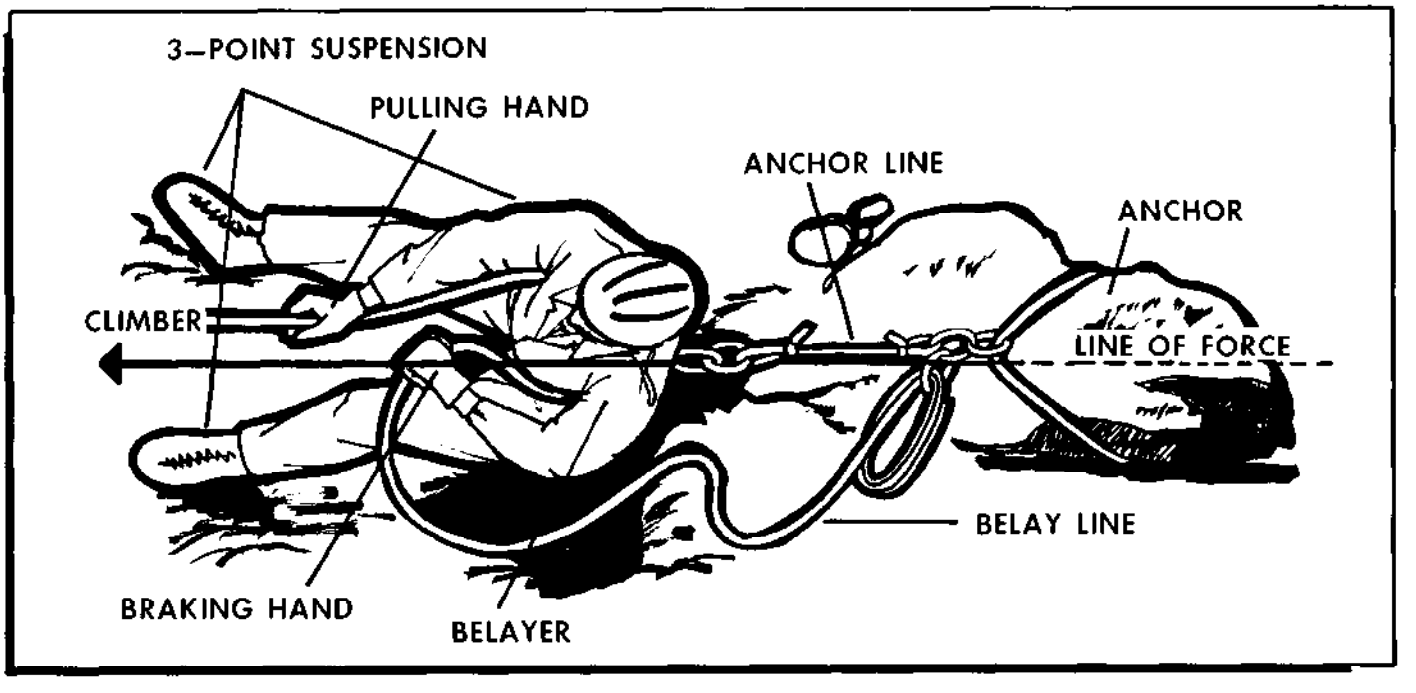

Figure 22-31. Belay Chain.

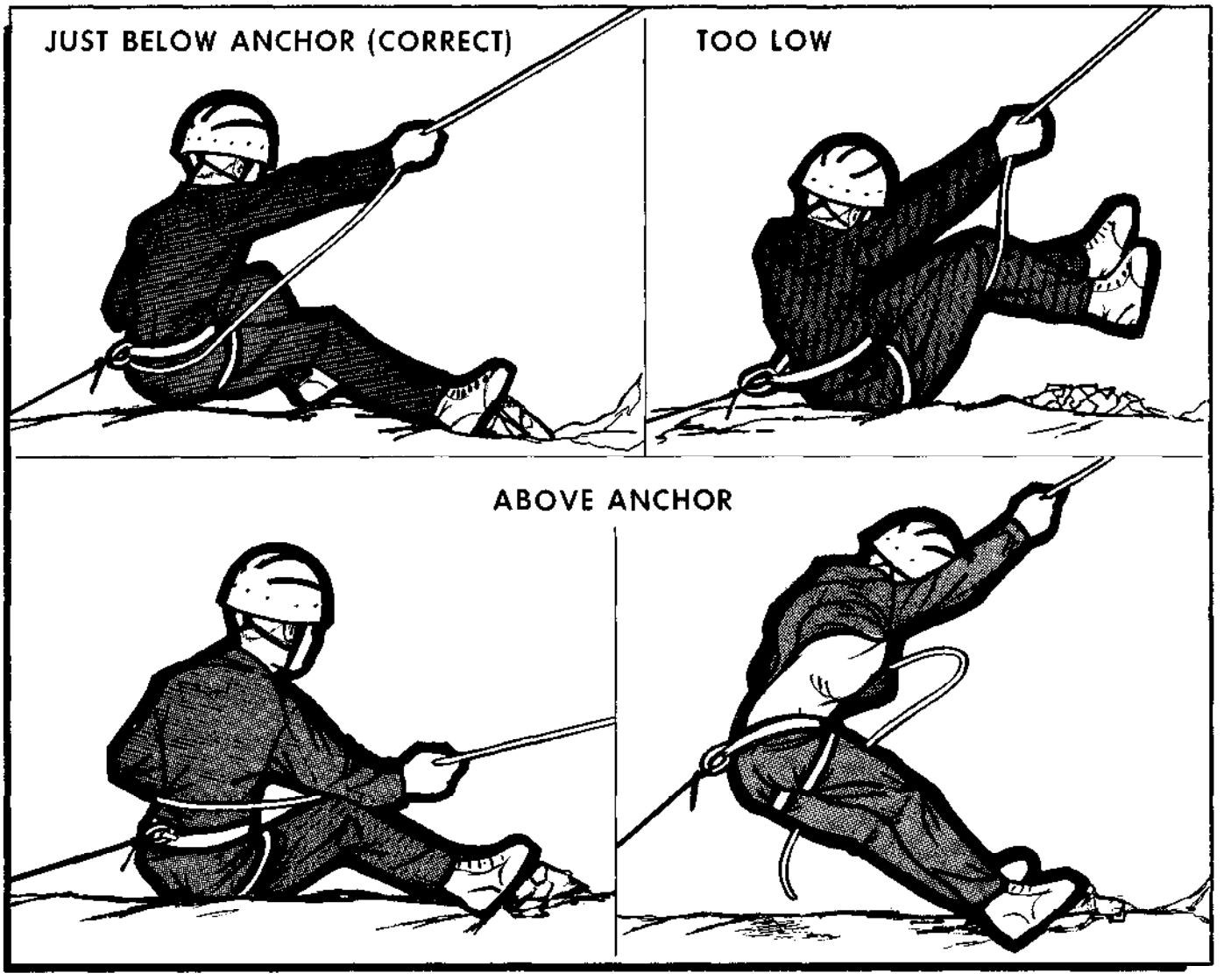

Figure 22-32. Sitting Belay.

storms, lightning can endanger the climber. Peaks, ridges, pinnacles, and lone trees should be avoided.

(4) Rockfalls are the most common mountaineering danger. The most frequent causes of rockfalls are other climbers, heavy rain and extreme temperature changes in high mountains, and resultant splitting action caused by intermittent freezing and thawing. Warning of a rockfall may be the cry "ROCK," a whistling sound, a grating sound, a thunderous crashing, or sparks where the rocks strike at night. A rockfall can be a single rock or a rockslide covering a relatively large area. Rockfalls occur on all steep slopes, particularly in gullies and chutes. Areas of frequent rockfalls may be indicated by fresh scars on the rock walls, fine dust on the talus piles, or lines, grooves, and rock-strewn areas on snow beneath cliffs. Immediate action is to seek cover, if possible. If there is not enough time to avoid the rockfall, the climber should lean into the slope to minimize exposure. Danger from falling rock can be minimized by careful climbing and route selection. The route selected must be commensurate with the ability of the least experienced team member. (NOTE: Yell "rock" when equipment is dropped.)

22-7. Belaying:

a. Belaying provides the safety factor or tension, which enables the party to climb with greater security. Without belaying skill, the use of rope in party climbing is a hazard. When climbing, a climber is belayed from above or below by another rescue team member.

(1) The belayer must run the rope through the guiding hand, which is the hand on the rope running to the climber or rescuer, and around their body to the brake hand, making certain that it will slide readily. The belayer must ensure that the remainder of the rope is laid out so it will run freely through the braking hand.

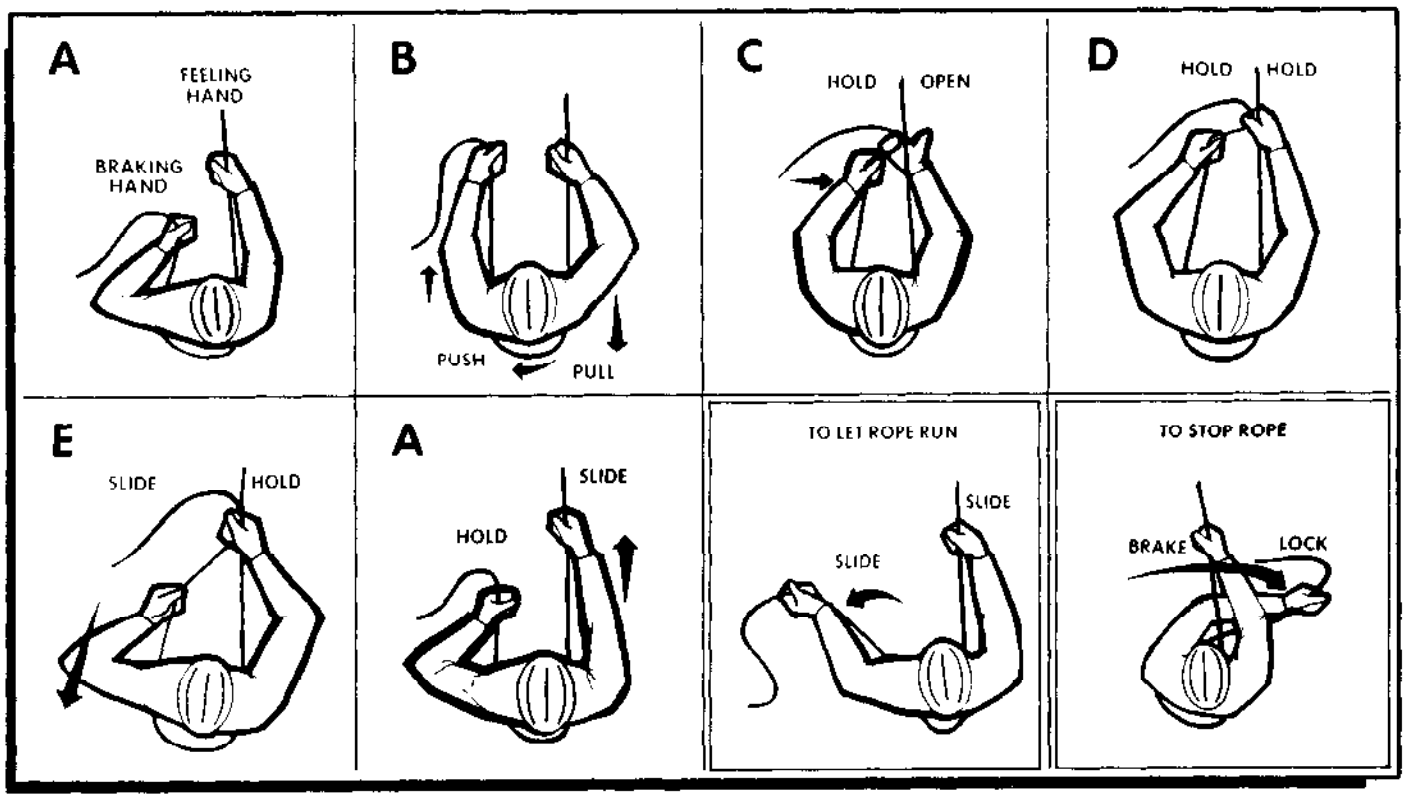

Figure 22-33. The Belay Sequence.

(2) The belayer must constantly be alert to the climber's movements in order to anticipate any needs. Avoid letting too much slack develop in the rope through constant use of the guiding hand. Keep all slack out of the rope leading to the rescuer, thus sensing any movement. If belaying a lead climber, the climber will need a constant flow of slack while climbing. If the rope is fed too slow or fast, the climber must communicate with the belayer to adjust the rate. Avoid taking up slack too suddenly to prevent throwing the climber off balance. When taking up slack, the braking hand is not brought in front of the guiding hand, but just behind the guiding hand. This allows the braking hand to slide back and to remain constantly on the rope. The braking hand is never removed from the rope during a belay.

(3) The belayer should brace well for the expected direction of pull in a fall so the force of the pull will, when possible, pull the belayer more firmly into position. A climber should neither trust nor assume a belay position which has not been personally tested.

b. The sitting belay is normally the most secure and preferred position (figure 22-32).

(1) The belayer sits and attempts to get good triangular bracing position with the legs and buttocks. Legs should be straight when possible, and the guiding hand must be on the side of the better braced leg. The rope should run around the hips. If the belay spot is back from the edge of a cliff, friction of the rope will be greater and will simplify the holding of a fall, but the direction of pull on the belayer will be directly outward. The rope must not pass over sharp edges.

(2) Even with a good belay stance, if the rope is too high or too low on the back of a belayer, the belayer may be unable to hold a falling climber (figure 22-32). With the rope too high on the back of the belayer, on a bottom belay, the rope will ride up in the belayer's armpits. (NOTE: Rope should run under the anchor.) This will pull the belayer forward and off balance if a climber were to fall. If the rope is too low on the back of a belayer, the rope will be pulled under the buttocks of the belayer. (NOTE: Rope should run on top of the anchor.) This will force the belayer to attempt to stop the fall by trying to hold onto the rope.

(3) When necessary, seek a belay position that offers cover from a rockfall.

(4) If the climber falls, the belayer should be able to perform the following movements automatically.

(a) Relax the guiding hand.

(b) Apply immediate braking action. This is done by bringing the braking hand across the chest or in front of the body (figure 22-33).

c. When a falling climber has been brought to a stop, the belayer must hold until the situation is relieved.

There are different ways to do this. If the climber is alright and can safely climb onto the rock or can be lowered to a secure ledge, all is well. However, the climber may be injured, or there may be no place to which the climber can be lowered. The belayer cannot continue holding the climber. The belayer must be relieved to assist the climber in the next step. This can be done by using a prusik sling. If the belayer is alone, the braking hand will hold the static climber while the other hand is free to anchor the belay line. This particular method is only good when using a belay anchor. Follow the procedure shown in figure 22-34. Climbers who are in good condition can either prusik up or pendulum across to a ledge.

22-8. Communications. Because constant communication between the belayer and climber is essential for safety, a standard group of climbing commands must be mastered.

a. Communications while climbing must be as simple as possible and single words are preferred. Communications in the form of commands are necessary when climbing. Commands which sound alike should be avoided. The command system follows a set pattern which leads to safe climbing and ensures that belays are used. Commands should be clear, specific, and given in a loud voice. The sequence of the system should not be broken. A review of the command system should be made before any climb. The European command has all the necessary commands to ensure safety. Only one command has a reply; the other commands are answered by the next command in the system. Sounding off the next command is not done until the requirements of the prior command are complete.

b. Commands are generally climber initiated. The belayer, in each instance, acknowledges when commands are heard and understood. If the command is not understood, the belayer should not say anything. The silence indicates to the climber that the belayer doesn't know what is happening. The command must be repeated.

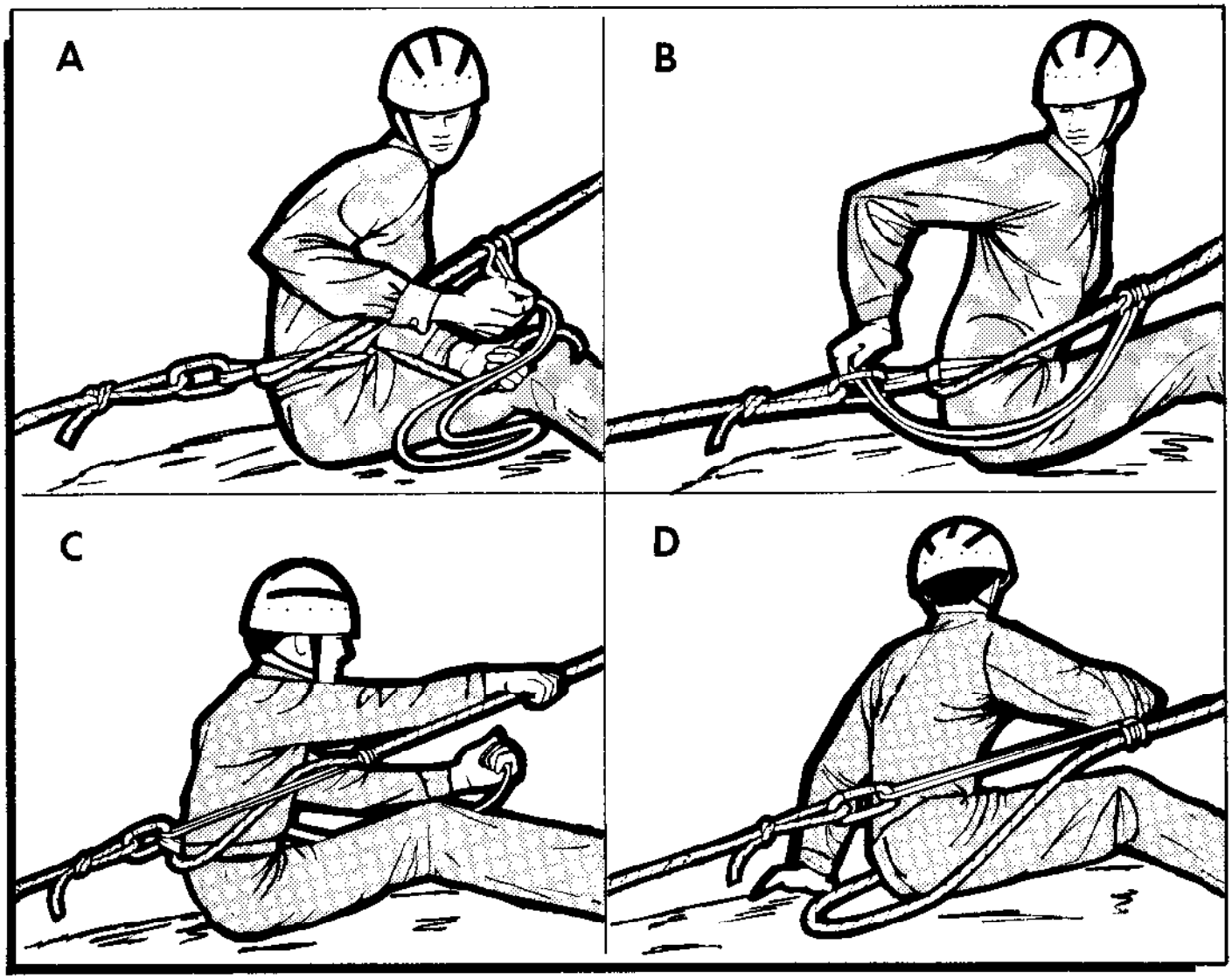

Figure 22-34. Safety Off the Belay Line.

(1) The following are the standard commands used:

(a) Initiation of a Climbing Sequence:

-1. "ON BELAY"—a signal indicating the climber is secured on the end of the rope and asks if belayer is ready.

-2. "BELAY ON"—(the belayer acknowledges the climber) means the belayer is in position and ready.

-3. "BELAY TEST"—the climber is asking to test the belay.

-4. "TEST"—signifies the belayer is ready for the climber to gradually apply weight on the rope. Releasing of the tension will indicate the test is finished.

-5. "UP ROPE"—the climber is directing the belayer to take up the slack.

-6. Belayer response: "ROPE UP."

-7. "THAT'S ME"—the climber signifies to the belayer that the tension felt on the rope is the climber.

-8. Belayer response: "THANK YOU."

-9. "CLIMBING"—the climber signifies the climber has chosen a route and is ready to climb.

-10. "CLIMB ON"—the belayer is acknowledging being ready for the climber to begin climbing. The climber will not start climbing before the command is heard.

(b) General Commands used while Climbing:

-1. "TENSION"—the climber is telling the belayer to take up all of the slack in the rope. This command is given if the climber needs assistance or there is an impending fall.

-2. Belayer responds with: "TENSION ON."

-3. "RESTING"—the climber is in a position for rest; belayer will lock brake across waist.

-4. Belayer responds with: "REST ON."

-5. When the climber begins again the command of "CLIMBING" will be given and "CLIMB ON" is given when the belayer is ready.

-6. "SLACK ___ FEET"—the climber will use this command when slack is needed in the rope to get over a difficult section or traverse. The climber should say "SLACK" and then the estimated number of feet the climber needs. Belayer will respond by repeating the number of feet of slack given. The climber will indicate that enough slack has been given by saying "THANK YOU."

-7. The command "FALLING" should be given when the climber is going to fall. This provides the belayer adequate time to apply the brake. Before continuing the climb, the climber will give the command "CLIMBING."

(c) Commands to Terminate the Climbing Sequence:

-1. "OFF BELAY"—the climber uses this command to indicate the climber is secure and through climbing.

-2. "BELAY OFF"—the belayer uses this command to indicate the brake hand is off the rope and the climber is no longer on belay.

22-9. Anchors. Anchors are secure points in the belay chain providing protection for the belayer and the climber. Anchor systems must be able to withstand high loads. The basis for any type anchor is strong, secure points for attachment.

a. Anchor Points. Anchor systems may be simple and consist of a single anchor point, or complex and consist of multiple anchor points. Anchor points are divided into two classes—natural and artificial. Anchor systems may be constructed entirely of one class, or a combination of the two.

(1) Natural Anchor Points (figure 22-35):

(a) Spike. A spike is a vertical projection of rock. For use as an anchor point, a sling is placed around the spike.

(b) Rock Bollard. A rock bollard is a large rock or portion of such a rock which has an angular surface enabling a sling or rope to be placed around it in such a manner that will not allow it to slip off. Care must be taken to ensure the bollard will not be pulled loose when subjected to a sudden load.

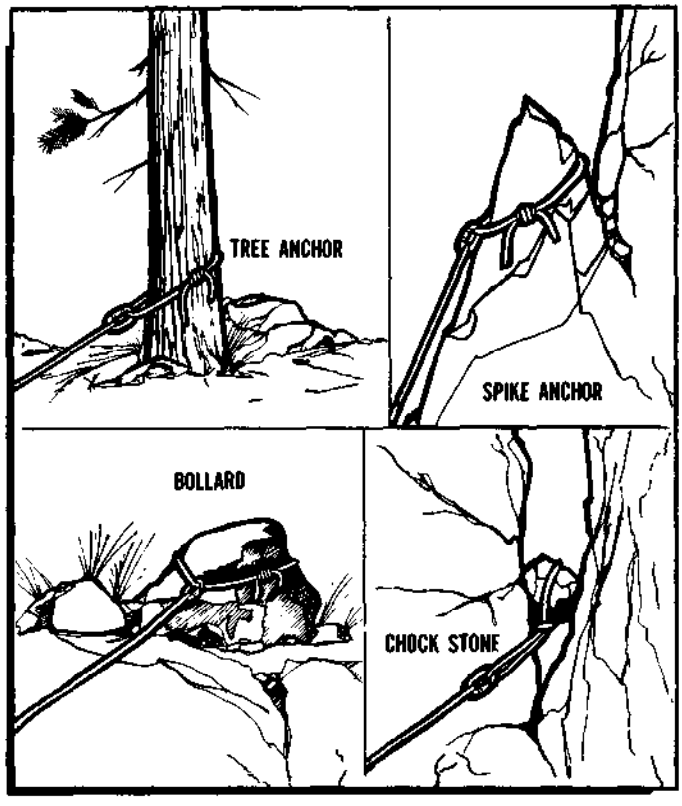

Figure 22-35. Natural Anchor Points.

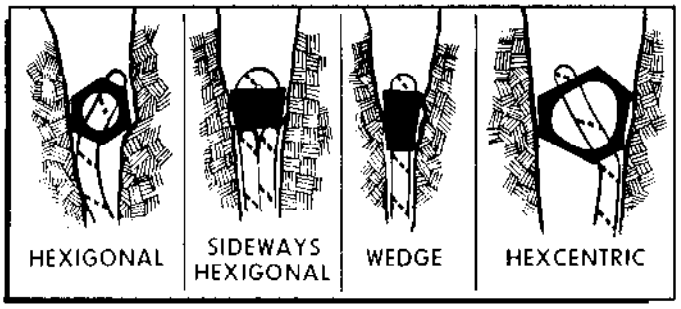

Figure 22-36. Placement of Chocks.

(c) Chockstone. A natural chockstone is a securely wedged rock providing an anchor point for a sling. In most cases, the rock is wedged within a crack.

(d) Tree. Trees often make very secure anchor points. A rope can be tied directly to the tree (or a sling doubled around the tree and connected by a carabiner). In loose or rocky soil, trees should be carefully watched and avoided if other anchor points are available.

(2) Artificial Anchor Points. Artificial anchor points are those constructed from equipment carried by the team. These are usually the chocks or pitons placed in cracks or bolts drilled in the rock.

(a) Chock Placement. The basic principle is to wedge the selected chock into a crack so that a pull in the direction of fall will not pull the chock out. The proper method is to select a crack suitable for an anchor point and select a chock to fit that crack. The chock chosen should closely fit the widest portion of the crack. Work the chock into the crack until it is securely seated. It should be seated so that the load will come on the entire chock without rotating it out of position. When the chock is in place, jerk hard on the attached sling in the direction of fall to ensure it is well seated. (Ensure the chock cannot continue to work downward to a larger area of the crack and become dislodged). A chock may also be placed so that its pull is in one direction while the second chock has a pull in the opposing direction. One sling is passed through the loop of the other chock. A force exerted downward will pull the two chocks toward each other along the axis of the crack. In vertical cracks, the lower sling should be passed through the upper for best results. The two chocks should be placed far enough apart so neither sling will reach the other chock. Figure 22-36 shows the various types of chocks

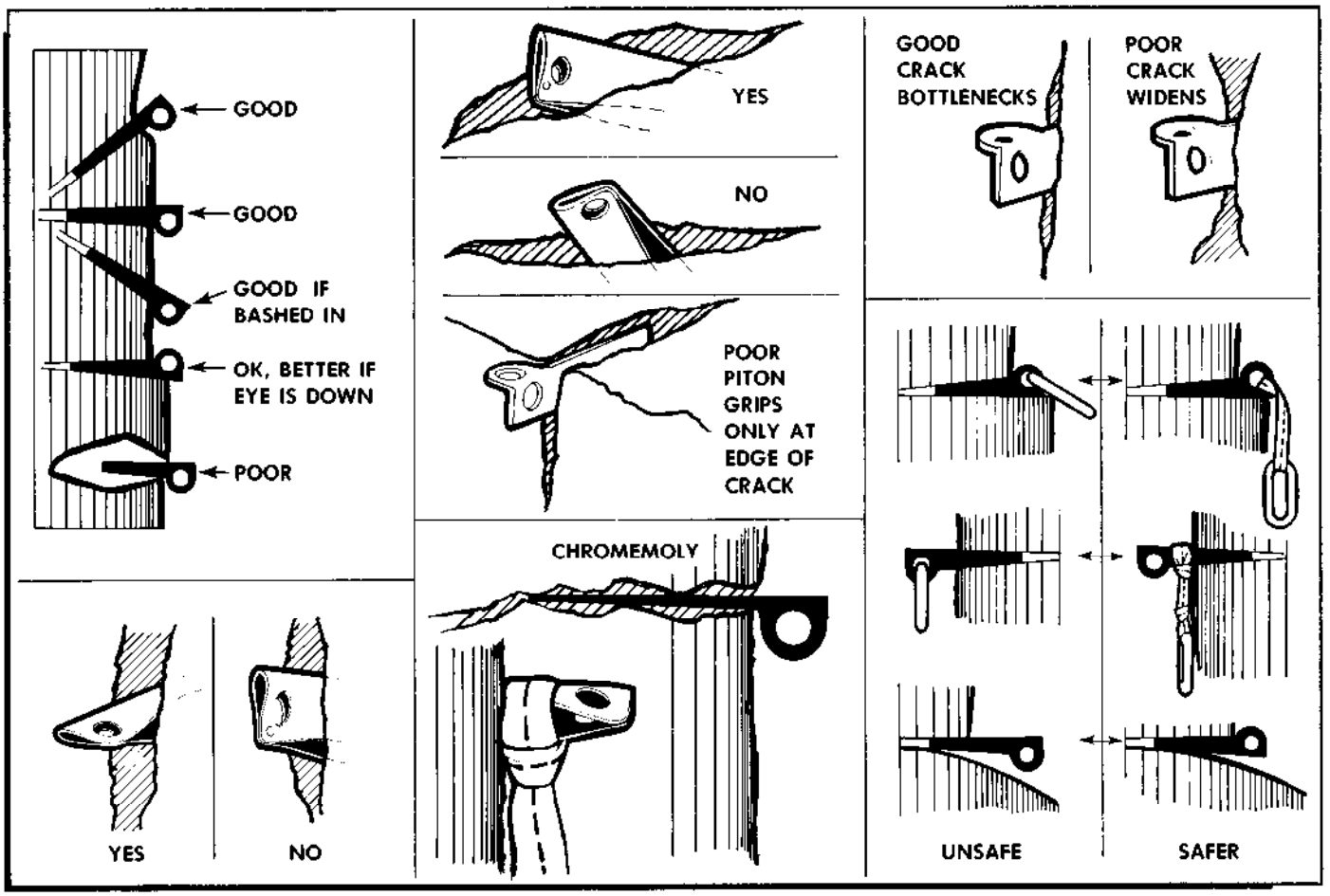

Figure 22-37. Placement of Pitons.

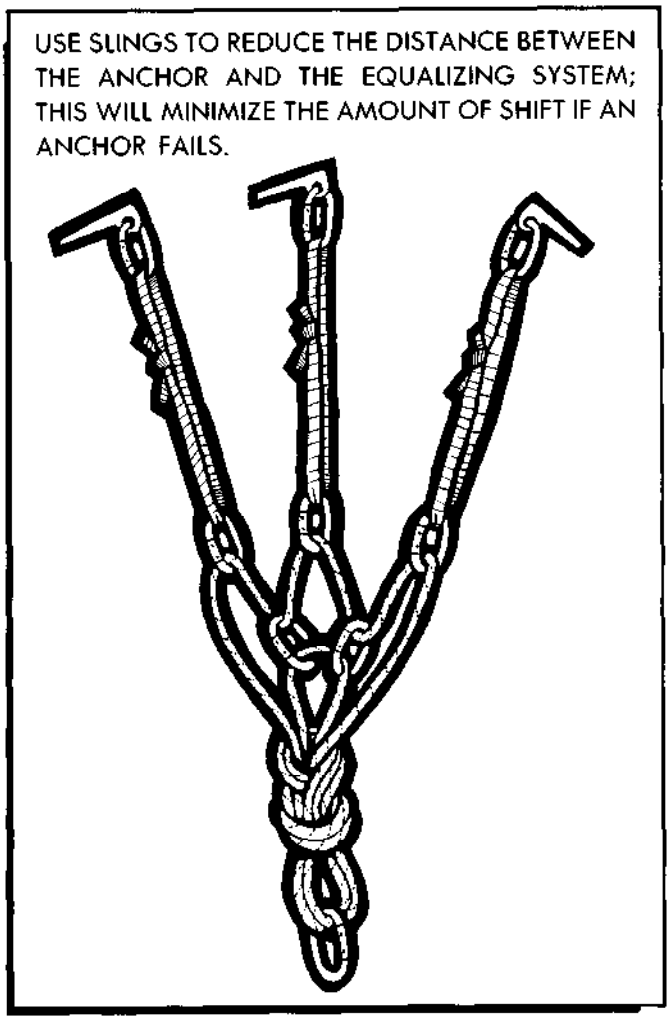

Figure 22-38. Equalizing Anchor.

and their placement. (NOTE: Chocks are preferred over pitons and bolts since they do not deface the rock.)

(b) Piton Placement. Basically, the use of pitons is a matter of locating a crack, selecting a piton which fits, driving the piton in, clipping a carabiner to the eye, and attaching the rope system to the carabiner.

-1. First look at the crack to decide the best position for driving the piton. The piton should be driven into the wider portion of the crack to reduce the likelihood of shifting or rotating under pressure. The crack must not widen or flare internally and not have a change in direction of more than 15 to 20 degrees.

-2. Before driving, the piton should fit one-half to two-thirds of the blade length into the crack. Drive until only the eye protrudes or the piton meets resistance. Do not attempt to overdrive the piton because it may fracture. Lightly tap the piton to test for movement or improper seating. If movement is noted, remove the piton and replace it with the next larger size or locate another anchor position. The proper placement of pitons in horizontal cracks is with the eye down. If the piton cannot be driven into the rock until only the eye protrudes, it may be tied off by placing a short sling around the piton with a girth or clove hitch as close to the rock as possible. In this case, the sling should be used as the attachment point and not the eye of the piton. Figure 22-37 shows the proper placement of pitons.

b. Anchor System. The purpose of an anchor system is to unite weak anchor points into a strong anchor system. Systems are divided into two classes—equalizing and nonequalizing.

(1) Equalizing systems are constructed so that if there is a change in direction of the load, the stress will be equally distributed to all anchor points. One major problem with this system is that if one point fails, the remaining points will be shock loaded (figure 22-38).

(2) Nonequalizing systems are constructed when a change of direction is not expected. The major advantage is the entire load is shared by the anchor points equally and would require the entire system to fail before coming loose. Any multipoint anchor system tied together so the attaching rope does not slip would be considered nonequalizing.

22-10. Climbing:

a. Balance Climbing. Balance climbing is the type of movement used to climb rock faces. It is a combination of the balance movement of a tightrope walker and the unbalanced climbing of a person ascending a tree or ladder. During the process of route selection, the climber should mentally climb the route to know what is expected. Climbers should not wear gloves when balance climbing.

(1) Body Position. The climber must keep good balance when climbing (the weight placed over the feet during movement). (See figure 22-39.) The feet, not the hands, should carry the weight (except on the steepest cliffs). The hands are for balance. The feet do not provide proper traction when the climber leans in toward the rock. With the body in balance, the climber moves with a slow, rhythmic motion. Three points of support, such as two feet and one hand, are used when possible. The preferred handholds are waist to shoulder high. Resting is necessary when climbing because tense muscles tire quickly. When resting, the arms should be kept low where circulation is not impaired. Use of small intermediate holds is preferable to stretching and clinging to widely separated big holds. A spread-eagle position, where a climber stretches too far (and cannot let go), should be avoided.

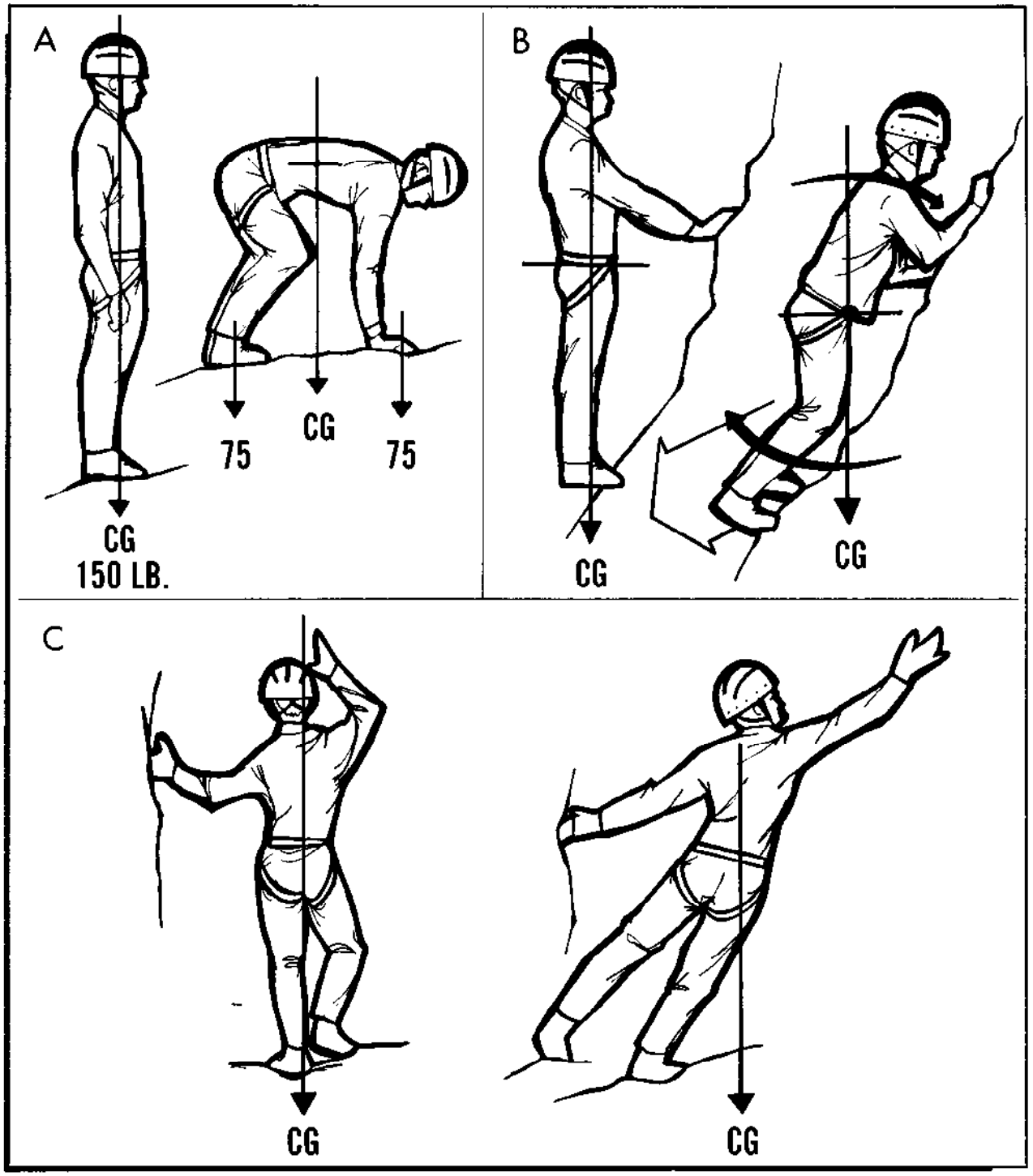

Figure 22-39. Body Position.

(2) Types of Holds:

(a) Push Holds. Push holds are desirable because they (figure 22-40) help the climber keep the arms low; however, they are more difficult to hold onto in case of a slip. A push hold is often used to advantage in combination with a pull hold.

(b) Pull Holds. Pull holds (figure 22-41) are those that are pulled down upon and are the easiest holds to use. They are also the most likely to break out.

(c) Jam Holds. Jam holds (figure 22-42) involve jamming any part of the body or extremity into a crack. This is done by putting the hand into the crack and clenching it into a fist or by placing the arm into the

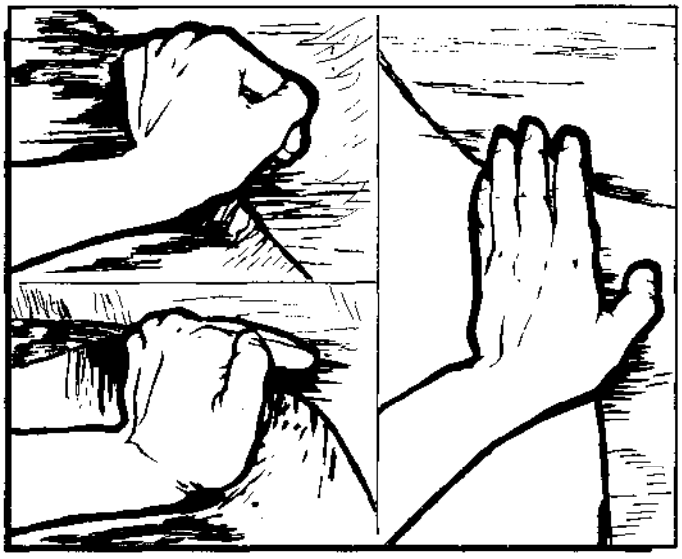

Figure 22-40. Push Holds.

crack and twisting the elbow against one side and the hand against the other side. When using the foot in a jam hold, care should be taken to ensure the boot is placed so it can be removed easily when climbing is continued.

(d) Combination Holds. The holds previously mentioned are considered basic and from these any number of combinations and variations can be used. The number of these variations depends only on the limit of the individual's imagination. Following are a few of the more common ones:

-1. The counterforce (figure 22-43) is attained by pinching a protruding part between the thumb and fingers and pulling outward or pressing inward with the arms.

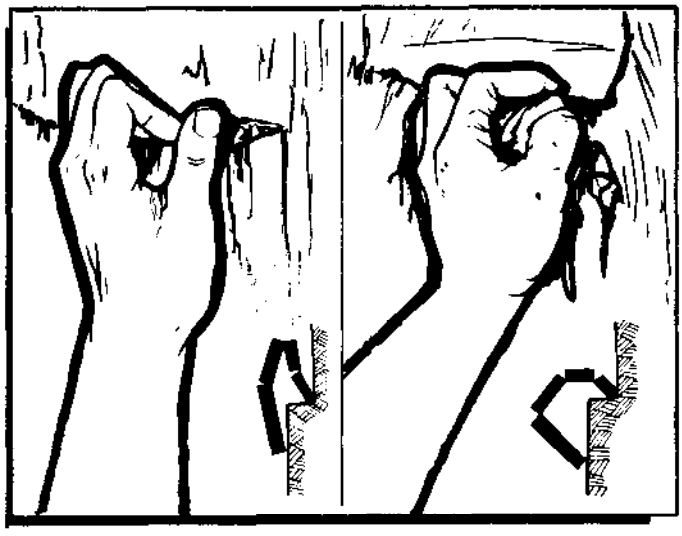

Figure 22-41. Pull Holds.

-2. The lay-back (figure 22-44) is done by leaning to one side of an offset crack with the hands pulling and the feet pushing against the offset side. Lay-backing is a classic form of force or counterforce where the hands and feet pull and push in opposite directions enabling the climber to move up in a series of shifting moves. It is very strenuous.

-3. Underclings (figure 22-45) permit cross pressure between hands and feet.

-4. Mantleshelving, or mantling, takes advantage of down pressure exerted by one or both hands on a slab or shelf. By straightening and locking the arm, the body is raised, allowing a leg to be placed on a higher hold (figure 22-46).

(e) Chimney Climb. This is a body-jam hold used in very wide cracks (figure 22-47). The arms and legs are used to apply pressure against the opposite faces of the

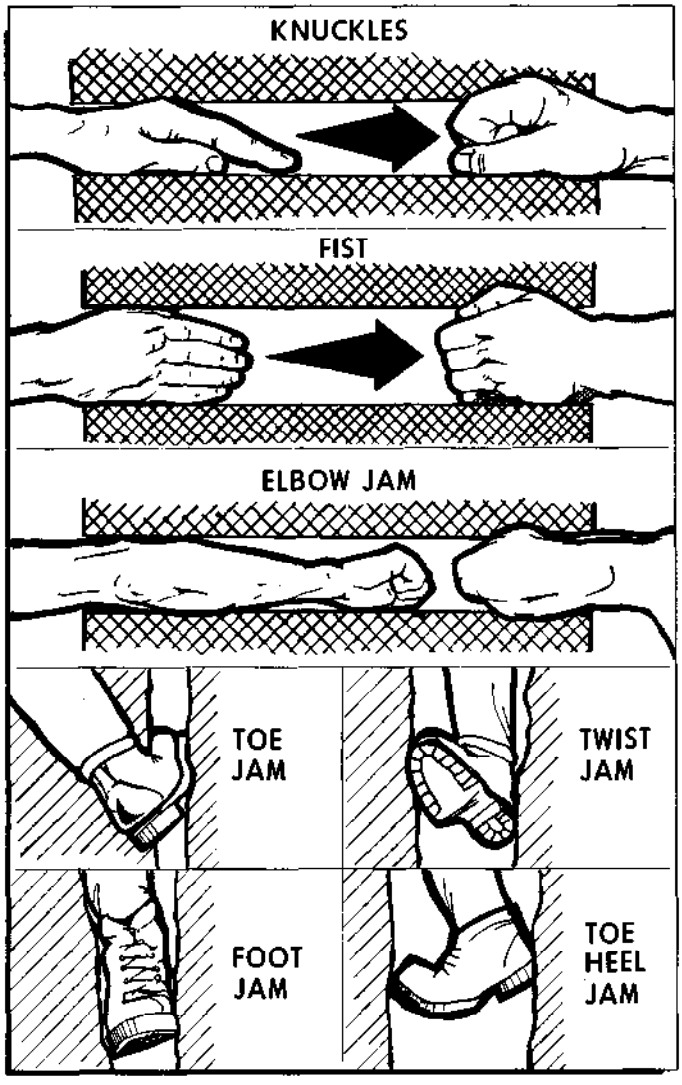

Figure 22-42. Jam Holds.

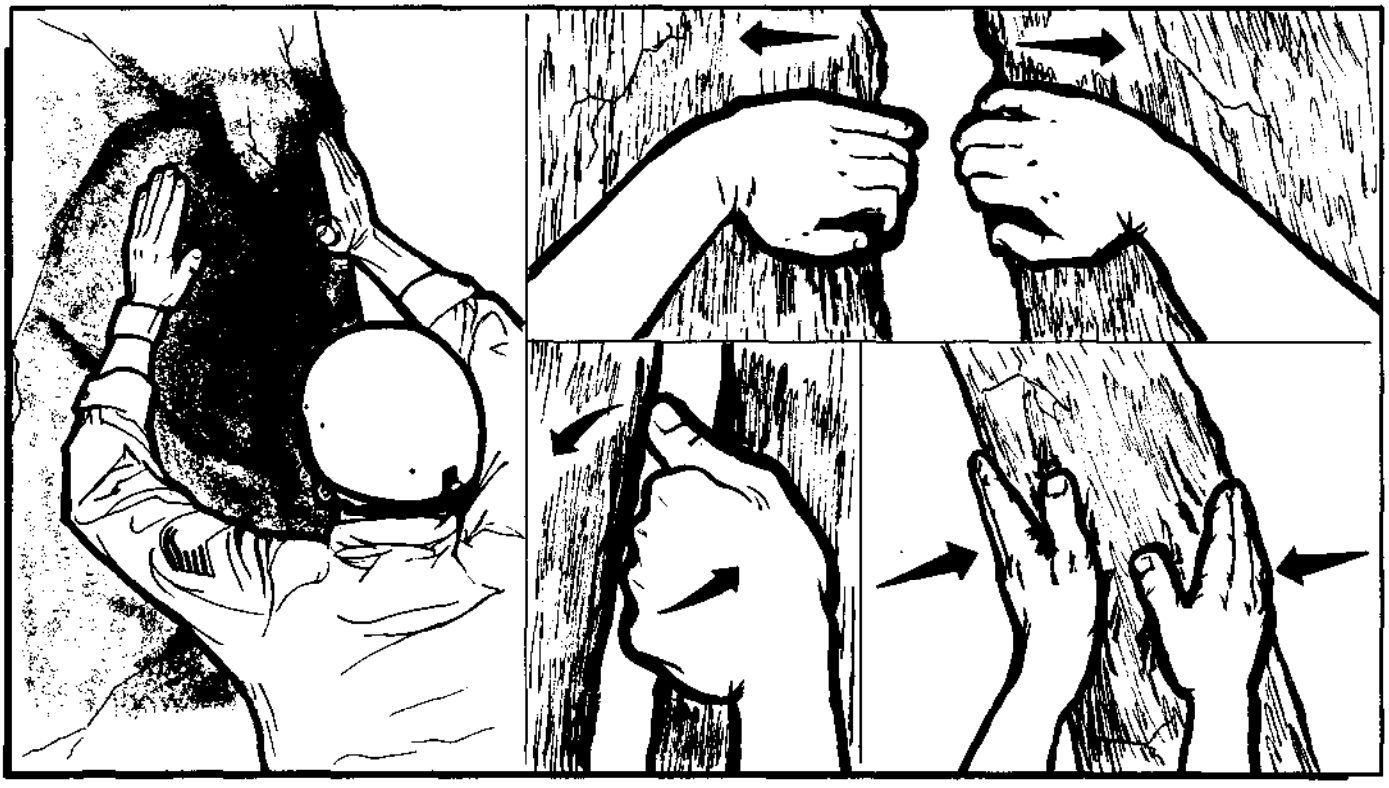

Figure 22-43. Combination Holds.

rock in a counterforce move. The outstretched hands hold the body while the legs are drawn as high as possible. The legs are flexed forcing the body up. This procedure is continued as necessary. Another method is to place the back against one wall and the legs and arms against the other and "worm" upward (figure 22-47).

b. Friction Climbing:

(1) A slab is a relatively smooth portion of rock lying at an angle. When traversing, the lower foot is pointed slightly downhill to increase balance and friction of the foot. All irregularities in the slope should be used for additional friction. On steep slabs, it may be necessary to squat with the body weight well over the feet with hands used alongside for added friction. This position may be used for ascending, traversing, or descending. A slip may result if the climber leans back or lets the buttocks down. Wet, icy, mossy, or a scree-covered slab is the most dangerous.

(2) Friction holds (figure 22-48) depend solely on the friction of hands or feet against a relatively smooth surface with a shallow hold. They are difficult to use because they give a feeling of insecurity which the inexperienced climber tries to correct by leaning close to the rock, thereby increasing the insecurity. They often serve well as intermediate holds, giving needed support while the climber moves over them; however, they would not hold if the climber decided to stop.

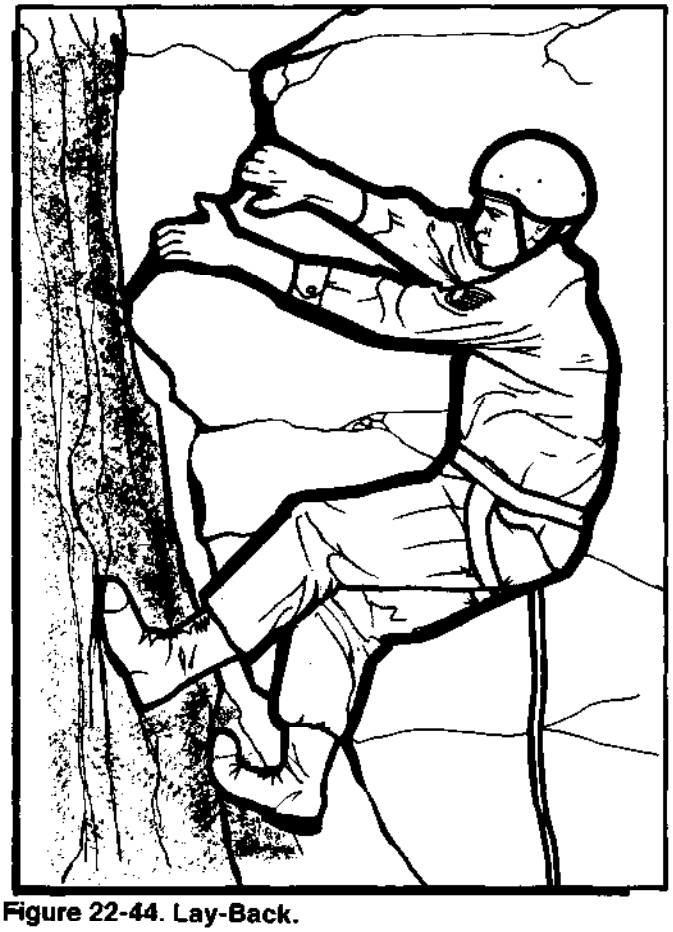

Figure 22-44. Lay-Back.

Figure 22-45. Underclings.

c. Prusik Climbing. Prusiking is a method of ascension using a fixed rope (figure 22-49). This method does not require a belay. With two prusik slings, a chest prusik sling, and a carabiner, a person can climb the length of a fixed rope. The slings are attached with prusik knots that grip tightly when loaded yet slide when the load is removed. Prusiking is strenuous and requires the use of both hands and both feet, hence it is no system for a badly injured person. Prusik slings are usually made from lengths of 7 mm Kernmantle or Mountain-Lay rope. Smaller diameter rope is less bulky and grips the climbing rope better, but in practice the knots are very difficult to work with gloved hands and is not recommended. To allow the climber to make maximum steps up with each foot in turn, the longer foot sling should extend to about nose level and the other to several inches below the waist loop. Both prusicks should have loops just big enough for the foot. The chest sling is just long enough to keep the climber from toppling backwards when rigged. Standing upright with all the weight bearing on the foot slings, the climber lifts one foot and raises its unweighted knot. Now, stepping up and shifting weight onto this sling, it is repeated with the other foot. The climb is less tiring when the foot slings are of different lengths, allowing the climber to take equal steps with both feet. There is some difficulty in sliding the knots when the rope is wet or the slings are made from laid rope. Spinning is also a potential problem under overhangs where the wall cannot be touched to maintain stability, particularly when climbing a laid rope. While a rescuer (climber) is ascending via the prusik method, companions have little to do except guard the anchors and prepare to lift the climber over the edge.

22-11. Rappelling. The climber with a rope can descend quickly by sliding down a rope which has been doubled around such anchor points as a tree, a projecting rock, or several artificial anchors secured to each other with sling rope.

a. Establishing a Rappel. In selecting the route, the climber should be sure the rope reaches the bottom or a place from which further rappels or climbing can be done. The rappel point should be carefully tested, and inspected to ensure the rope will run around it when one end is pulled from below and the area is clear of loose rocks. If a sling rope is used for a rappel point, it should be tied twice to form two separate loops. The first person down chooses a smooth route for the rope which is free of sharp rocks. Place loose rocks, which the rope might later dislodge, far enough back on ledges to be out of the way. The rappeler should ensure the rope runs freely around the rappel point when pulled from below. Each person down will give the signal "OFF RAPPEL," straighten the rope, and ensure the rope runs freely around its anchor. When silence is needed, a prearranged signal of pulling on the rope is substituted for the vocal signal. Recover the rope when the last person is down. The rope should be pulled smoothly to prevent the rising end from whipping around the rope. Climbers should stand clear of falling rope and rocks which may be dislodged. The rope should be inspected frequently if a large number of people are rappelling. Rappellers should wear gloves during rappels to protect the palms from rope burns.

b. Types of Rappel. The type of rappel is determined by the steepness of the terrain. The hasty rappel is used only on moderate pitch slopes. The body rappel may be used on moderate to severely pitched slopes, but never

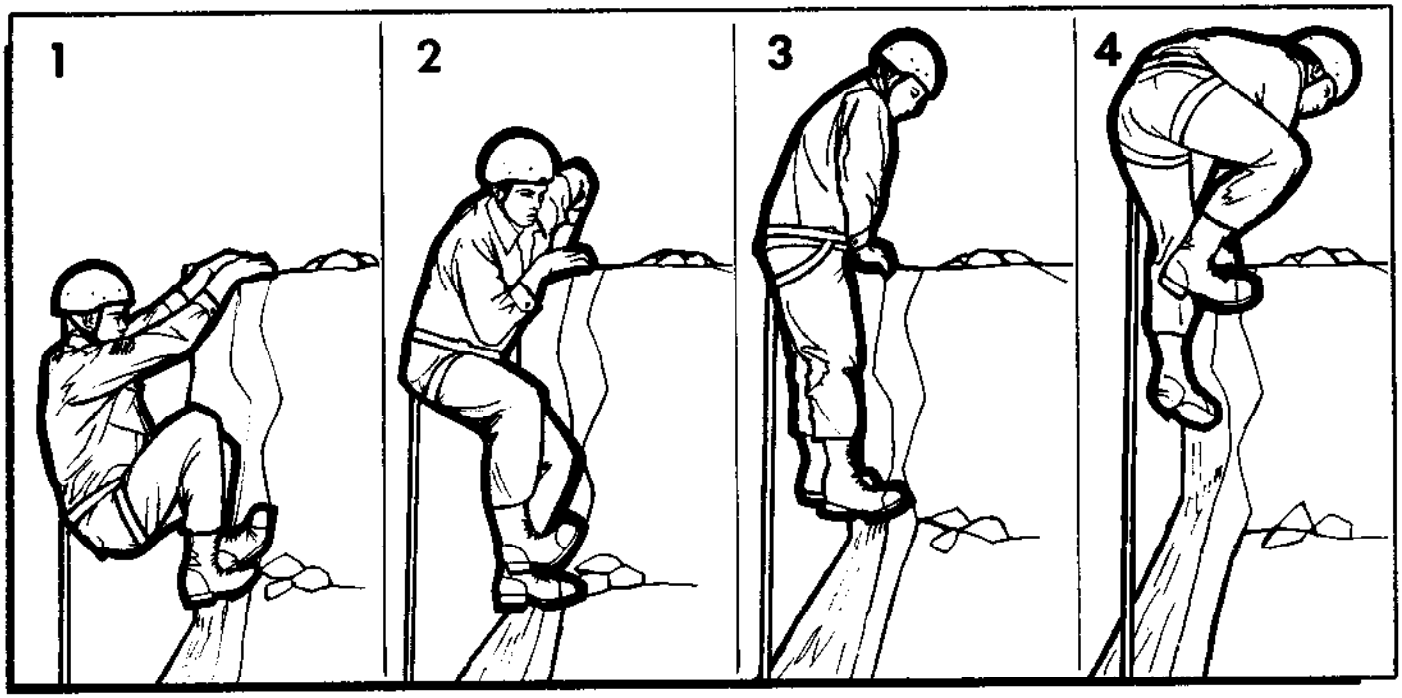

Figure 22-46. Mantleshelving.

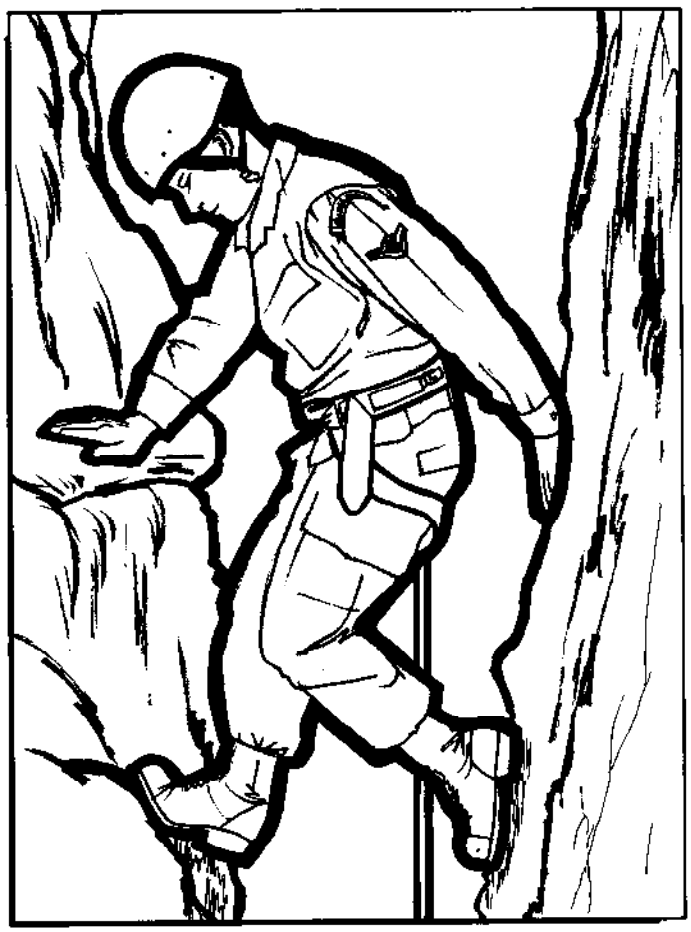

Figure 22-47. Chimney Climbing.

Figure 22-48. Friction Holds.

Figure 22-49. Prusiking.

used on overhangs. The seat rappel is used on very steep pitches, including overhangs.

(1) The Hasty Rappel (figure 22-50). Facing slightly sideways to the anchor, the climber places the ropes across the back. The hand nearest the anchor is the guiding hand. To stop, the rappeller brings the braking hand across in front of the body, and at the same time turns to face the anchor point.

(2) The Body Rappel (figure 22-51). The climber faces the anchor point and straddles the rope; then pulls the rope from behind, runs it around either hip, diagonally across the chest, and back over the opposite shoulder. From there the rope runs to the braking hand which is on the same side of the hip that the rope crosses (for example, the right hip to the left shoulder to the right hand). The climber should lead with the braking hand down and should face slightly sideways. The foot corresponding to the braking hand should precede the other at all times. The guiding hand should be used only to guide and not to brake. To rappel, lean out at a 45-degree angle to the rock. Keep the legs well spread and relatively straight for lateral stability, and the back straight since this reduces unnecessary friction. The collar should be turned up to prevent rope burns on the neck. Gloves should be worn and any other articles of clothing may be used as padding for the shoulders and buttocks. To brake, lean back and face directly into the rock so the feet are flat on the rock.

(3) Four-Carabiner Seat Rappel. Seat rappels differ from the body and hasty rappels in that the friction is primarily absorbed by a carabiner in the sling rope seat worn by the rappeller. The rappeller stands to one side of the rope (when braking with the right hand on the left and when braking with the left hand on the right). Some slack between the carabiner and the anchor point is taken up and brought through two carabiners which are attached to the harness and are horizontal and reversed and opposed (figure 22-52). Two additional carabiners are clipped from opposite sides, gates down, at 90 degrees to the original carabiners to form a friction brake bar (figure 22-53).

c. Body Position. The body must be perpendicular to the face of the rock, and the feet must be about shoulder width apart and flat on the rock (figure 22-54). To de-

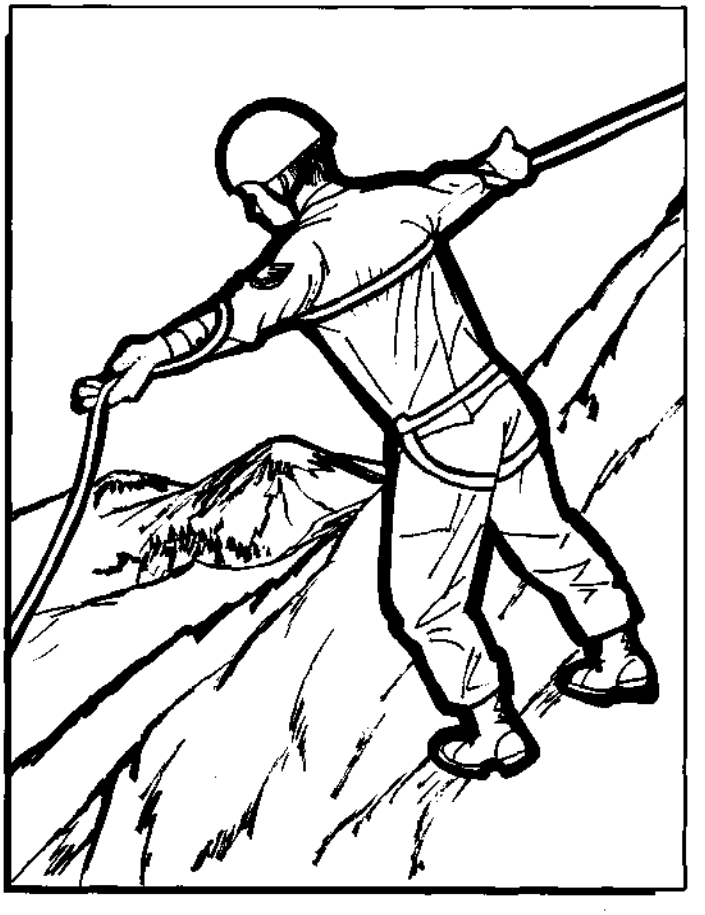

Figure 22-50. Hasty Rappel.

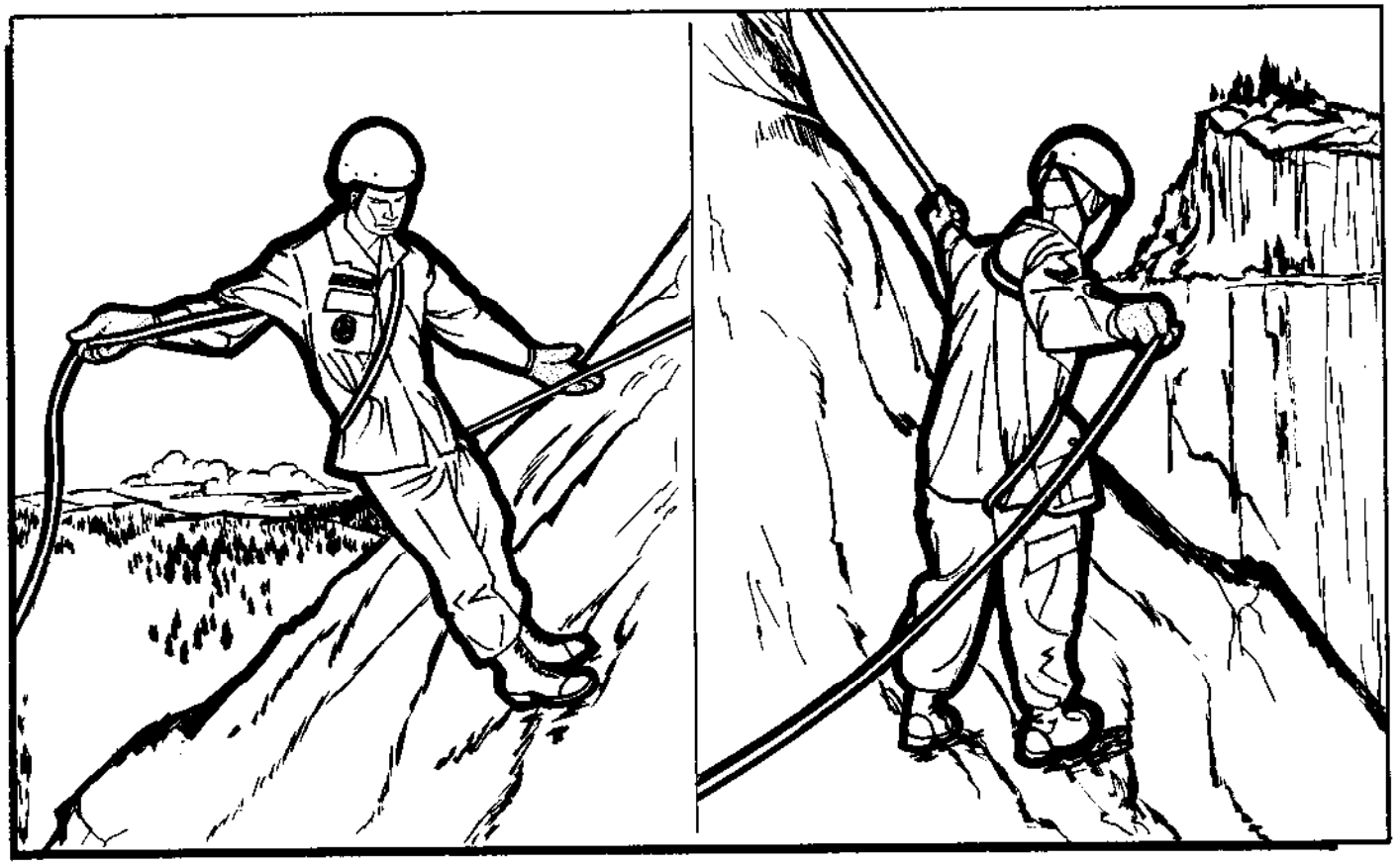

Figure 22-51. Body Rappel.

scend, hold the brake out to the side to reduce friction. To brake, cinch down on the rope and move the brake hand to the small of the back.

22-12. Overland Snow Travel. Cold-weather operations conducted in snow-covered regions magnify the difficulties of reaching a survivor and effecting an extraction. Routes of travel surveyed from the air may not be possible from the ground. A straight line from the insertion point to the objective is the most desired route; however, inherent dangers (avalanches, collapsing cornices, etc.) may necessitate an alternate route entailing a longer trek. During all operations, safety, and not ease of travel, will be the primary concern.

a. Snow Conditions. Travel time varies from hour to hour. Certain indicators may assist in the direction of travel; that is, the best snow condition is one which supports a person on or near the surface when wearing boots and the second best is calf-deep snow conditions. If possible avoid traveling in thigh- or waist-deep snow. Snowshoes should be worn when conditions dictate.

(1) South and west slopes offer hard surfaces late in the day after exposure to the Sun and the surface is refrozen. East and north slopes tend to remain soft and unstable. Walking on one side of a ridge, gully, clump of

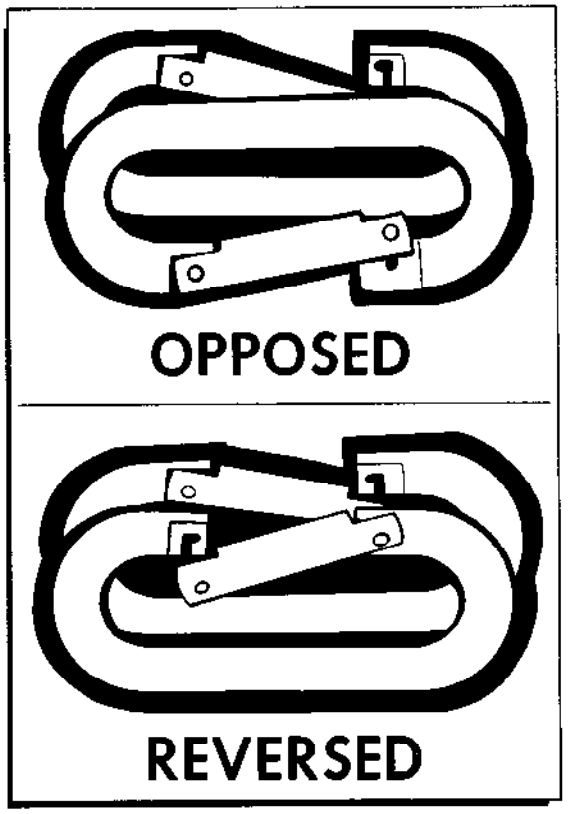

Figure 22-52. Seat Rappel (Carabiners).

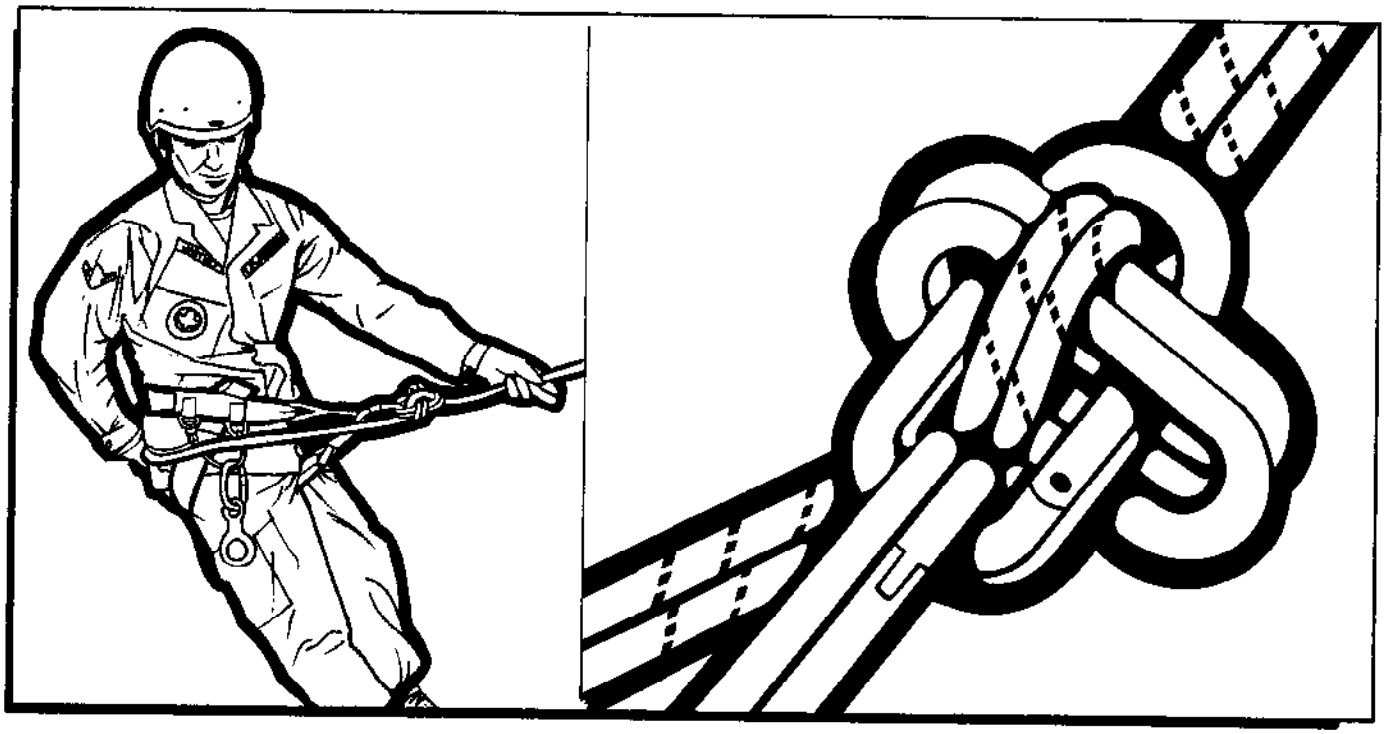

Figure 22-53. Four-Carabiner Brake System.

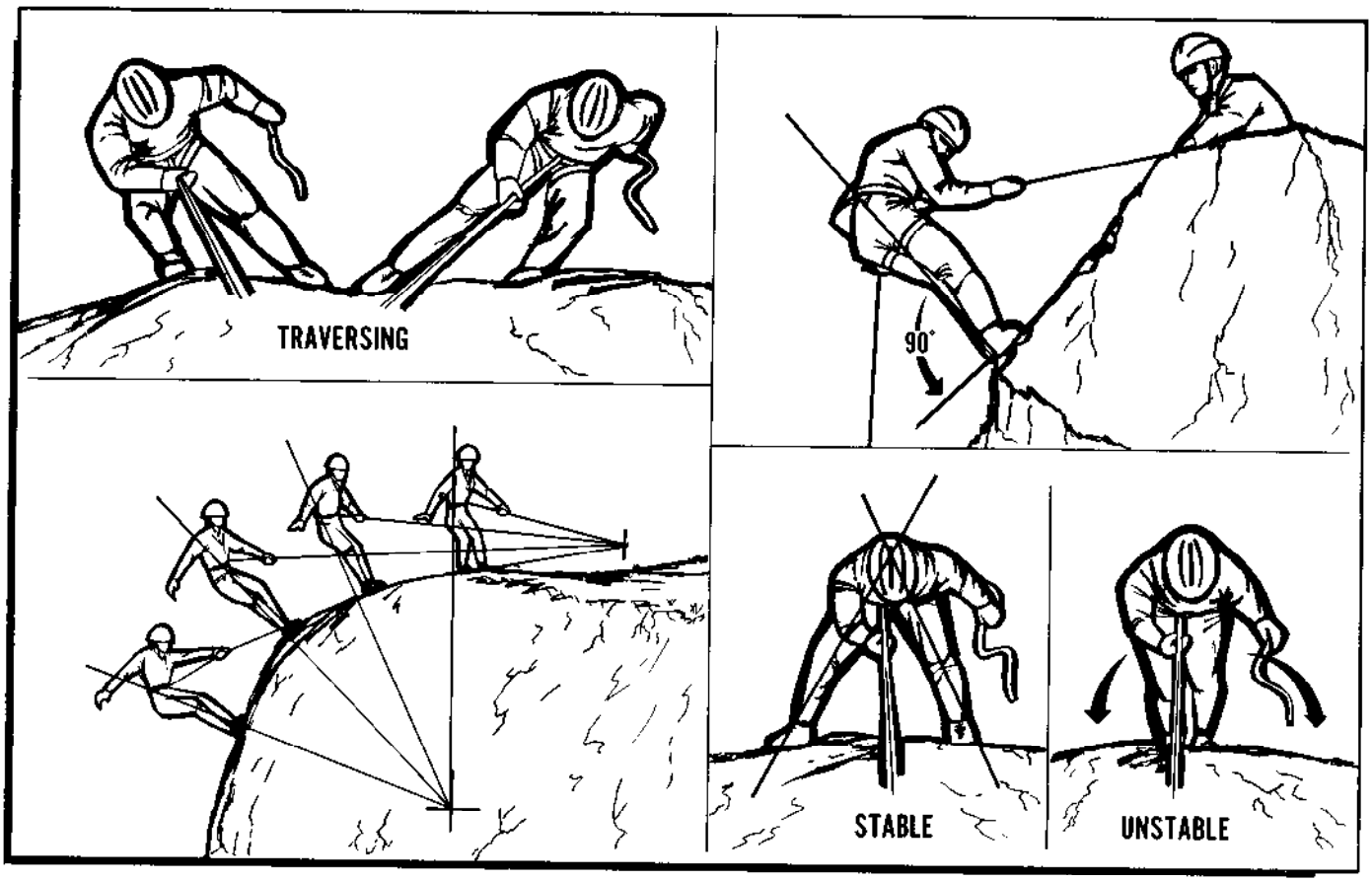

Figure 22-54. Body Position for Rappel.

trees, or large boulders is often more solid than the other side. Dirty snow absorbs more heat than clean snow; slopes darkened by rocks, dust, or uprooted vegetation usually provide more solid footing. Travel should be done in the early morning after a cold night to take advantage of stable snow conditions. Since sunlight affects the stability of snow, travel should be concentrated in shaded areas where footing should remain stable.

(2) In areas covered by early seasonal snowfall, travel between deep snow, and clear ground must be done cautiously. Snow on slopes tends to slip away from rocks on the downhill side, forming openings. These openings, called moats, are filled by subsequent snowfalls. During the snow season, moats below large rocks or cliffs may become extremely wide and deep, presenting a hazard to the rescue team.

b. Travel Speed. An over-zealous drive to reach an objective may be too fast for the endurance of the team. Fast starts at the point of insertion usually result in frequent stops for recuperation. The best way to reach an objective is to start with a steady pace and continue that pace throughout. Movement at reasonable speeds, with rest stops as required, will help prevent team "burnout." The following considerations further ensure steady advancement with minimal degradation to the team. A steady pace helps maintain an even rate of breathing. After the initial period of travel; that is, one-half hour, a shakedown rest should be initiated to adjust boots, snowshoes, crampons, packs, etc., or to remove or add layers of clothing.

c. Snowshoe Technique. A striding technique is used for movement with snowshoes. In taking a stride, the toe of the snowshoe is lifted upward to clear the snow and thrust forward. Energy is conserved by lifting the snowshoe no higher than is necessary to clear the snow. If the front of the snowshoe catches, the foot is pulled back to free it and then lifted before proceeding with the stride. The best and least exertive method of travel is a loose-kneed rocking gait in a normal rhythmic stride. Care should be taken not to step on or catch the other snowshoe.

(1) On gentle slopes, ascent is made by climbing straight upward. (Traction is generally very poor on hard-packed or crusty snow.) Steeper terrain is ascended by traversing and packing a trail similar to a shelf. When climbing, the snowshoe is placed horizontally in the snow. On hard snow, the snowshoe is placed flat on the surface with the toe of the upper one diagonally uphill to get more traction. If the snow will support the weight of a person, it is better to remove the snowshoes and temporarily proceed on foot. In turning, the best method is to swing the leg up and turn the snowshoe in the new direction of travel.

(2) Obstacles such as logs, tree stumps, ditches, and small streams should be stepped over. Care must be taken not to place too much strain on the snowshoe ends by bridging a gap, since the frame may break. In shallow snow, there is danger of catching and tearing the webbing on tree stumps or snags. Wet snow will frequently ball up under the feet, making walking uncomfortable. This snow should be knocked off with a stick or pole.

(3) Generally, ski poles are not used in snowshoeing; however, one or two poles are desirable when carrying heavy loads, especially in mountainous terrain. The bindings must not be fastened too tightly or circulation will be impaired and frostbite can occur. During stops, bindings should be checked for fit and possible readjustment.

d. Uphill Travel. Maximum altitude may be obtained with less effort by traversing a slope. A zigzag or switchback route used to traverse steep slopes places body weight over the entire foot as opposed to the balls of the feet as in a straight line uphill climb. An additional advantage to zigzagging or switchbacking is alternating the stress and strain placed on the feet, ankles, legs, and arms when a change in direction is made.

(1) When a change in direction is made, the body is temporarily out of balance. The proper method for turning on the steep slope is to pivot on the outside foot (the one away from the slope). With the upper slope on the right side, the left foot (pivot foot) is kicked directly into the slope. The body weight is transferred onto the left foot while pivoting toward the slope. The slope is then positioned on the left side and the right foot is on the outside.

(2) In soft snow on steep slopes, pit steps must be stamped in for solid footing. On hard snow, the surface is solid but slippery, and level pit steps must be made. In both cases, the steps are made by swinging the entire leg in toward the slope, not by merely pushing the boot into the snow. In hard snow, when one or two blows do not suffice, crampons should be used. Space steps evenly and close together to facilitate ease of travel and balance. Additionally, the lead climber must consider the other team members, especially those who have a shorter stride.

(3) The team should travel in single file when ascending, permitting the leader to establish the route. The physical exertion of the climbing leader is greater than that of any other team member. The climbing leader must remain alert to safeguard other team members while choosing the best route of travel. The lead function should be changed frequently to prevent exhaustion of any one individual. Team members following the leader should use the same leg swing technique to establish foot positions, improving each step as they climb. Each foot must be firmly kicked into place, securely positioning the boot in the step. In compact

snow, the kick should be somewhat low, shaving off snow during each step, thus enlarging the hole by deepening. In very soft snow, it is usually easier to bring the boot down from above, dragging a layer of snow into the step to strengthen and decrease the depth of it.

(4) When it is necessary to traverse a slope without an increase in elevation, the heels rather than the toes form the step. During the stride, the climber twists the leading leg so that the boot heel strikes the slope first, carrying most of the weight into the step. The toe is pointed up and out. Similar to the plunge step, the heel makes the platform secure by compacting the snow more effectively than the toe.

e. Descending. The route down a slope may be different from the route up a slope. Route variations may be required for descending different sides of a mountain or moving just a few feet from icy shadows onto sun-softened slopes. A good surface snow condition is ideal for descending rapidly since it yields comfortably underfoot. The primary techniques for descending snow-covered slopes are plunge stepping and descending step by step.

(1) The plunge step makes extensive use of the heels of the feet (figure 22-55) and is applicable on scree as well as snow. Ideally, the plunging route should be at an angle, one that is within the capabilities of the team and affords a safe descent. The angle at which the heel should enter the surface varies with the surface hardness. On soft snow slopes, almost any angle suffices; however, if the person leans too far forward, there is a risk of lodging the foot in a rut and inflicting injuries. On hard snow, the heel will not penetrate the surface unless it has sufficient force behind it. Failure to firmly drive the heel into the snow can cause a slip and subsequent slide. The quickest way to check a slip is to shift the weight onto the other heel, making several short, stiff-legged stomps. This technique is not intended to replace "the ice arrest" technique which is usually more effective. When roped, plunging requires coordination and awareness of all team members' progress. Speed of the team must be limited to the slowest member. Plunging is unsatisfactory when wearing crampons due to the snow compacting and sticking to them.

(2) The technique of step-by-step descending is used when the terrain is extremely steep, snow significantly deep, or circumstances dictate a slower pace. On near-vertical walls, it is necessary to face the slope and cautiously lower oneself step by step, thrusting the toe of the boot into the snow while maintaining an anchor or handhold with the axe. Once the new foothold withstands the body's full weight, the technique is repeated. On moderately angled terrain, the team can face away from the slope and descend by step-kicking with the heels.

Figure 22-55. Plunge Step.

22-13. Snow and Ice Climbing Procedures and Techniques. Snow and ice climbing differs from rock climbing, yet many of the procedures and techniques are the same. Belay tie-in commands, principles of runner placement, straight-line climbing, and placement of protection are common to snow and ice as well as rock. As expected, however, there are major differences from rock climbing.

a. Ice Axe Techniques (figure 22-56). The axe is the most important tool a climber carries. It can be used for braking assistance when a climber begins sliding down a steep snow-covered incline.

(1) Each rescuer (climber) must practice the self-arrest technique before venturing onto steep grades. Since the ice axe arrest requires the use of the ice axe, the climber must hang onto it at all times. The ice axe, whether sharp or not, is a lethal weapon when flying about on the attached cord. Physically, the climber rigs

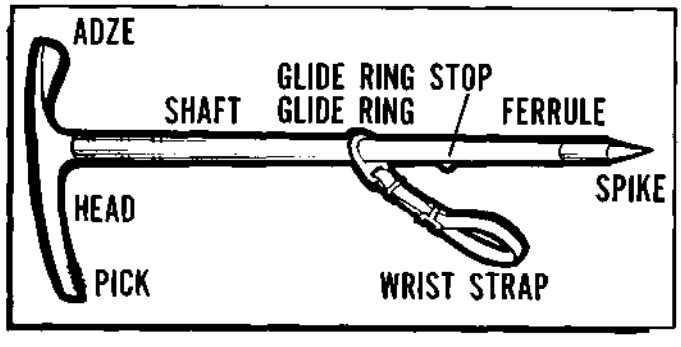

Figure 22-56. Ice Axe.

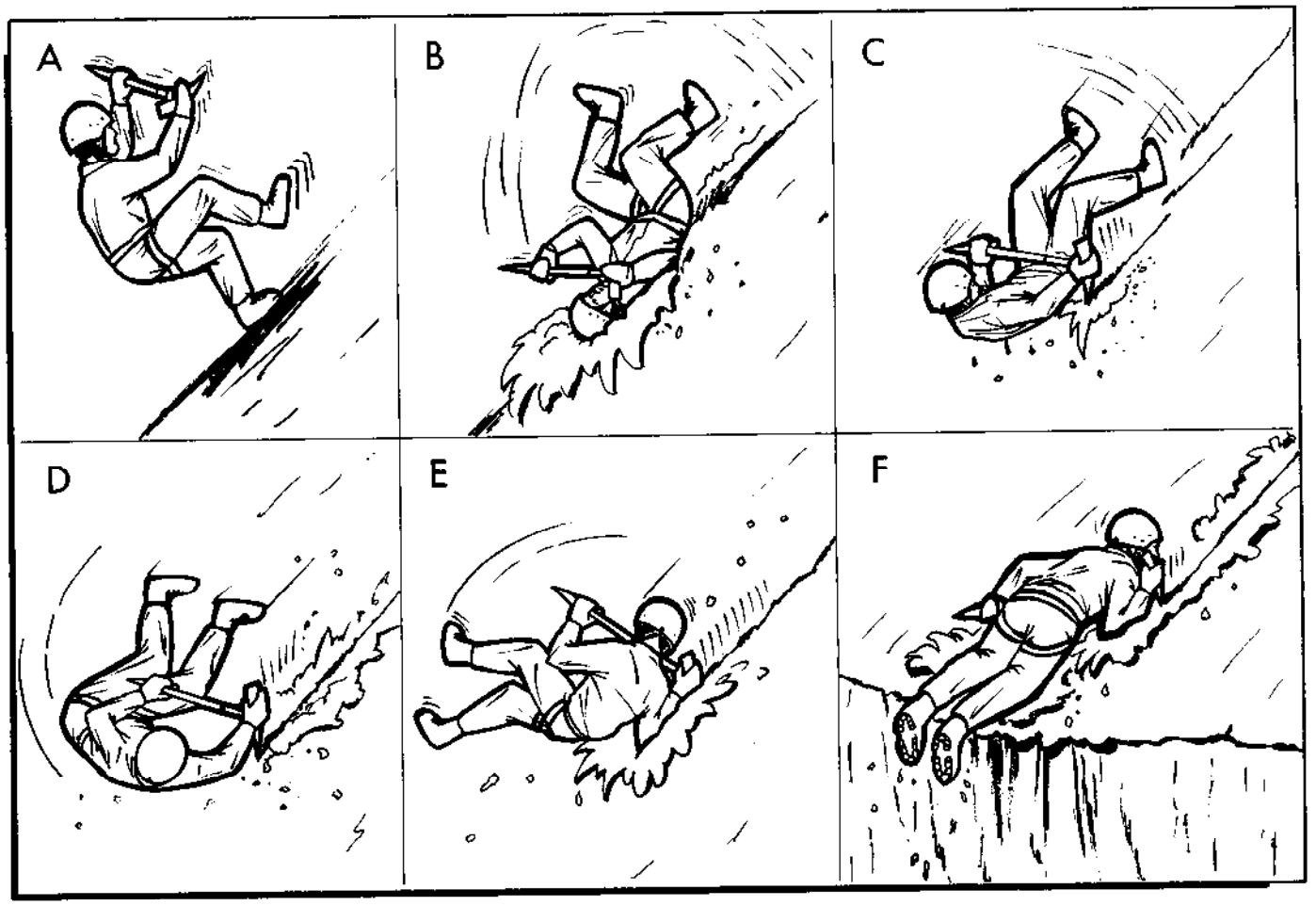

Figure 22-57. Ice Axe Self-Arrest.

for arrest by rolling down shirt sleeves, putting on mittens, securing loose gear, and most important of all, making certain the axe is held correctly. Mentally prepare by recognizing the importance of instantaneous application. A quick arrest, before the fall picks up speed, has a better chance of success than a slow arrest. Preparation for an ice axe arrest should be taken when traveling on terrain which could result in a fall.

(2) The proper method of holding an ice axe for self-arrest (see figure 22-57) is to place one hand on the head of the axe with the thumb under the adze and fingers over the pick. The other hand is placed on the shaft next to the spike. The pick is pressed into the slope just above the shoulder so the adze is near the angle formed by the neck and shoulders. The shaft should cross the chest diagonally with the spike held firmly close to the opposite hip. A short axe is held in the same position, although the spike will not reach the opposite hip. Chest and shoulders should press strongly on the shaft and the spine should be arched slightly to distribute weight primarily at the shoulders and toes. The legs should be stiff and spread apart, toes digging in (if wearing crampons, keep the toes off the surface until almost stopped), and hang on to the axe!

b. Team Arrest. The team arrest is intermediate between self-arrest and belays. When there is doubt that a person could arrest a fall, such as on crevassed glaciers and steep snowfields, and conditions are not so extreme as to make belaying necessary, the party ropes up and travels in unison. If any member falls, arrest is made by two or three axes. The rope between the climbers must be fully extended except for minimum slack carried by the second and subsequent persons to allow them to flip the rope out of the track (steps). This also allows easy compensation for pace variations. However, slack is minimized to bring the second and subsequent axes into action at the moment of need. A roped climber who falls should immediately yell "FALLING." It is not advisable to delay the alarm to see how "self-arrest" will develop because team members may hear the signal after they have been pulled into their own falls, decreasing their ability to help. When roped climber(s) hear the cry "falling," they immediately drop into self-arrest position.

c. Boot Axe Belay (figure 22-58). The boot axe belay can be set up rapidly and used when a team is moving together and belaying is only required at a few spots. The boot axe belay should be practiced until a sweep

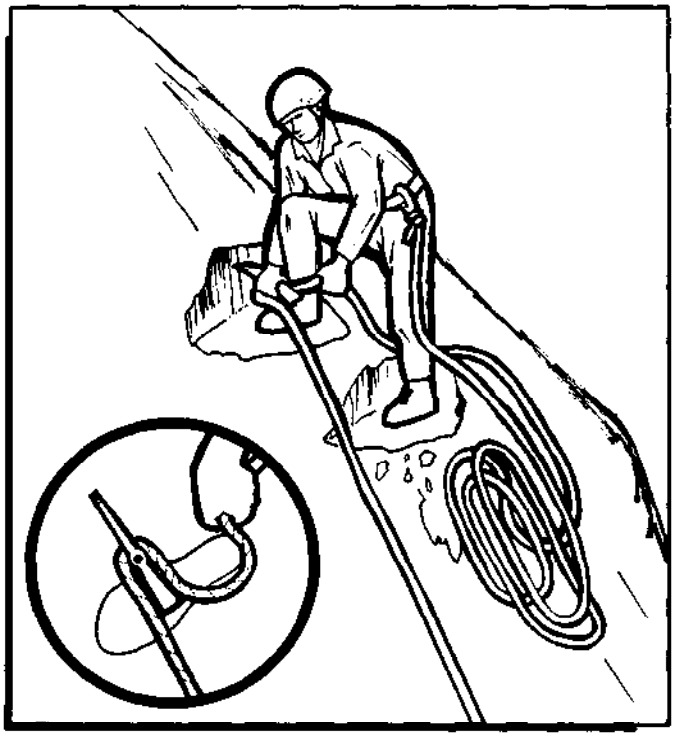

Figure 22-58. Boot Axe Belay.

and jab of the ice axe can set up the stance within a couple of seconds. The axe provides an anchor to the slope and the slope and the boot braces the axe. Both give a friction surface over which the run of rope is controlled.

(1) To prepare a boot axe belay, a firm platform, large enough for the axe and uphill boot, is stamped out in the snow. The ice axe shaft is jammed as deeply as possible, at a slight uphill angle (against the anticipated fall) into the snow at the rear of the platform. The pick is parallel to the fall line, pointing uphill, thus applying the strongest dimension of the shaft against the force of a fall. The length of the pick prevents the rope from escaping over the top of the shaft.

(2) The belayer stands below the axe, facing at a right angle to the fall line. The uphill boot is stamped into the slope against the downhill side of the shaft at a right angle to the fall line, bracing the shaft against downhill pull. The downhill boot is in a firmly compacted step below the uphill boot so that the leg is straight, stiffly bracing the belayer. The uphill hand is on the axe head in arrest grasp, bracing the shaft against downhill and lateral stress. From below, the rope crosses the toe of the boot, preventing the rope from trenching into the snow. The rope bends around the uphill side of the shaft, then down across the instep of the bracing boot, and is controlled by the downhill hand. To apply braking through greater friction, the downhill or braking hand brings the rope uphill around the heel, forming an "S" bend.

d. Crampon Techniques:

(1) Donning. When attaching the harness, the buckles should be positioned to cinch on the outward sides of the boots. Special care must be taken to strap the crampons tightly to the boots, running the strap through each attachment prong or ring. If crampons do not have heel loops, ankle straps should be long enough to be crossed behind the boot before being secured to prevent boots from sliding backward out of the crampons. Many crampons have been lost because this precaution was not taken. When trimming new straps, allowance must be made for gaiters which sometimes cover the instep of the boot. Donning is best done by laying each crampon on the snow or ice with all rings and straps outward; then place the boot on the crampon and tighten the straps. Even modern neoprene-coated nylon straps should be checked from time to time to make sure they are tight, have not been cut, and are not trailing loop strap ends which could cause the wearer to trip.

(a) If it is believed crampons may be needed, they must be carried. Conditions change rapidly; an east-facing slope may be mushy enough for step-kicking during the morning, but can become a sheet of smooth white ice in the afternoon shade. Furthermore, cramponing may contribute directly to the team's safety by enabling it to negotiate stretches of ice faster and with less fatigue than having to chop steps. The decision of whether or not to wear crampons is determined by the situation. Wearing crampons should not be considered mandatory because of venturing onto a glacier; neither should a team attempt to save time by never wearing crampons on steep, exposed icy patches just because they are fairly short. Another important guideline is to don crampons before they are needed to avoid donning them while teetering in ice steps. On mixed rock and ice climbs, constant donning and removing of crampons takes so much time that the objective may be lost.

(b) Crampons should be worn throughout the entire climb if the terrain is 50 percent or more suitable for crampons (crampons may skid or be broken on rock surfaces). Crampons are not required if the snow or ice patches are fairly short, good belays are available, and rock predominates. These alternatives are suggestions and the team leader's decision must be based on the conditions at hand.

(c) Crampons should be taken off when the snow begins to ball up badly in them and no improvement in snow conditions is anticipated. On the ascent, it may be possible to clear away the soft surface snow and climb on the ice below, but this is usually impractical and futile on the descent. Occasionally the climber should kick the crampons free of accumulated snow. The time-worn practice of striking the ice axe shaft against the crampons to knock out the snow is effective, but hard on the axe and perhaps the ankle. In situations where the crampons must be worn even though the snow balls

up in them, shuffling the feet through the snow instead of stepping over the surface tends to force the accumulated snow through the back points. The normal kicking motion of the foot generally keeps the crampons snow-free on the ascent and while traversing.

(d) On the descent, drive the toe of the boot under the surface of the snow ahead of the heel, walking on the ball of the foot. Keep the weight well forward and use short skating steps allowing the foot to slide forward and penetrate the harder sublayers.

(2) Flatfooting. Flatfooting involves a logical and natural progression of coordinated body and ice axe positions to allow the climber to move steadily and in balance while keeping all vertical points of the crampons biting into the ice. The weight is carried directly over the feet, the crampon points stamped firmly into the ice with each step, with the ankles and knees flexed to allow boot soles to remain parallel to the slope.

(a) On gentle slopes, the climber walks straight up the hill. Normally, the feet are naturally flat to the slope and the axe is used as a cane. If pointing the toes uphill becomes awkward, they may be turned outward in duck-fashion. As the slope steepens, the body is turned to face across the slope rather than up it. The feet may also point across the slope, but additional flexibility and greater security are gained by pointing the lower foot downhill. The axe is used only to maintain balance and may be carried in the cane position or the arrest grasp with either the pick or point touching the slope. (Movement is diagonal rather than straight upward and the climber takes advantage of terrain irregularities and graded slopes. Changes in direction are done as in step-kicking on snow by planting the downhill foot, turning the body toward the slope to face the opposite direction, and stepping off with the new downhill foot.)

(b) On gentler slopes, the flat-footed approach is used throughout, but it is more secure and easier on steeper slopes to initiate the turn by kicking the front points and briefly front-pointing through the turn. At some point in the turn, the grip on the axe must be reversed. The exact moment for this depends on the climber and the specific situation. However, the climber's stance must be secure when the third point of support—the axe is temporarily relinquished.

(c) On steep slopes, which approach the limit of practical use of this style of ascent, the climber relies on the axe for security of a hold as well as for balance. The axe is held in the arrest grasp with one hand on the head and the other on the shaft, above the point. The well-sharpened pick is planted firmly in the ice at about shoulder height to provide one point of suspension while a foot moves forward and the crampons are stamped in (figure 22-59).

(d) Descent follows the same general progression of foot and axe positions; descend the fall line, gradually turning the toes out as the slope gets steeper. As the slope steepens, widen the stance, flex the knees, and lean forward to keep weight over the feet, and finally, face sideways and descend with the support of the axe in the arrest position. On very steep or hard ice, it may be necessary to face the slope and front-point downwards. When flat-footing downhill, all crampon points should be stamped firmly into the ice. It may be necessary to strive to take small steps which allow the climber to maintain balance during moves; long steps require major weight shifts to adjust balance.

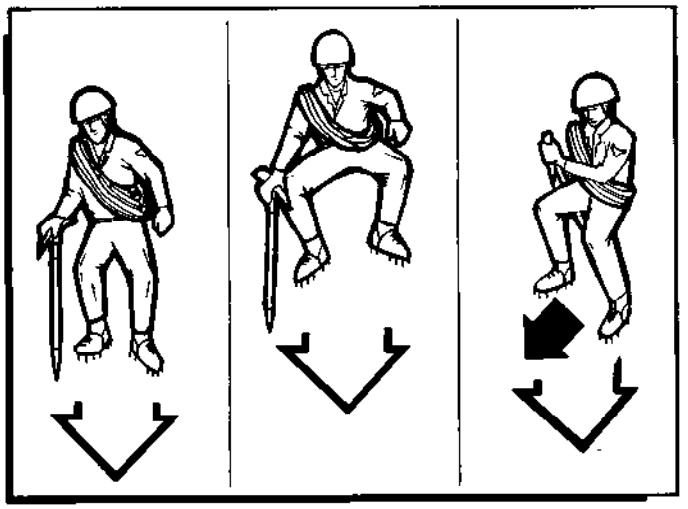

Figure 22-59. Descent with Crampons.

e. Anchors. Snow and ice conditions require the use of special devices for establishing belay anchors or placement of intermediate protection during a climb.

(1) Snow Pickets. Three- to four-foot lengths of aluminum "T" or tubular sections perform as long pitons and are suited for belaying. They must always be used in pairs or greater numbers, one anchoring the other (figure 22-60).

(2) Snow Fluke. A 12-inch piece of metal is buried with a runner coming to the surface; this can pull out if snow conditions are not just right. Better resistance to pull-out is gained with a large flat piece of metal driven into the snow surface at an angle, acting in the same manner as the fluke of an old-fashioned anchor. There is no danger of the runner being cut or weakened from wet conditions with the attachment of a wire cable. The softer the snow, the larger the size of the plate. When using flukes, it is very important that the proper angle with the surface be maintained; otherwise, wire, instead of becoming stronger (going deeper) when pulled, will become weaker (surface). Additionally, the cable may

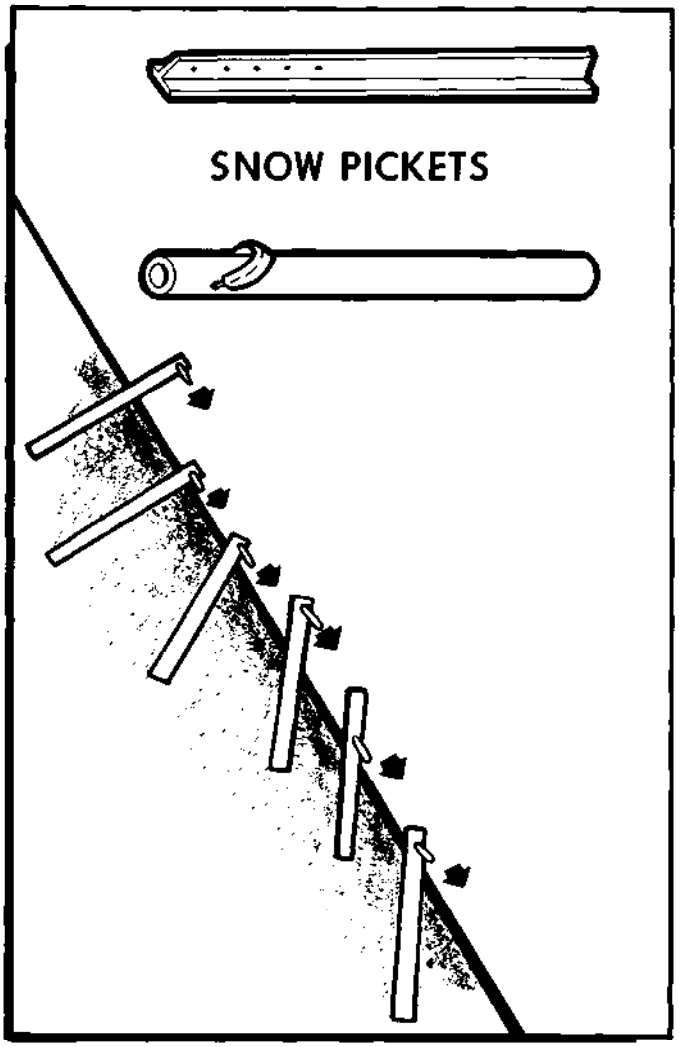

Figure 22-60. Snow Pickets.

act as a lever arm on hard snow, causing the fluke to pop out. This can be prevented by carefully cutting a channel in the snow for cables so that the pull comes directly at the plate. If attention is paid to placement, snow flukes will provide great security as belay and rappel anchors. A properly placed fluke is secure for a sitting or standing hop belay on snow (figure 22-61).

(3) Ice Screws:

(a) Tubular screws are very strong and are the most reliable (figure 22-62). They are difficult to place in hard or water ice since they tend to clog and have a large cross section. Their main advantage is that they minimize "spalling" (a crater-like splintering of the ice around the shaft of the screw) by allowing the displaced ice to work itself out through the core of the screw. If the core of the ice remaining in the screw is frozen in place, it jams the screw in subsequent placements. The ice may be removed by pushing with a length of wire or by heating with a cigarette lighter. This type of screw

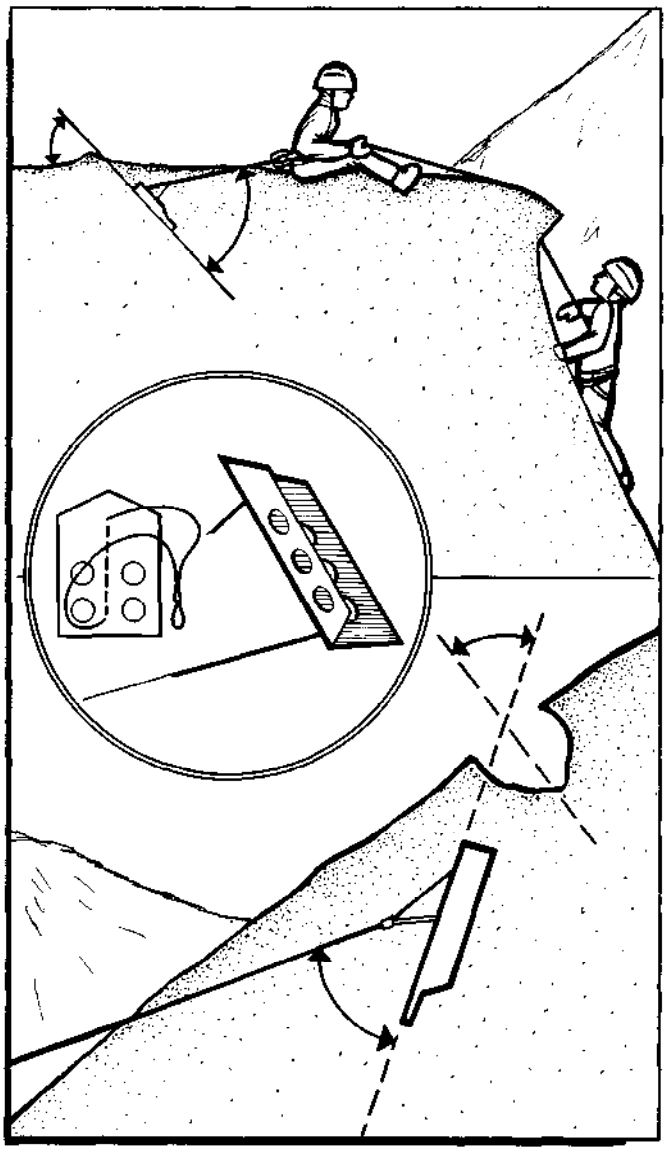

Figure 22-61. Snow Flukes.

requires both hands for placement; however, once it is started, the pick of an ice hammer or axe inserted in the eye allows the climber to gain the advantage of leverage. Removal is easy and melt-out is slow due to the large cross section.

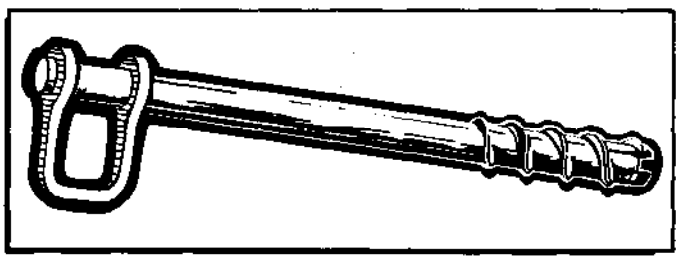

Figure 22-62. Tubular Ice Screw.

(b) Heavier "coathanger" type screws can be relied upon to stop a fall (figure 22-63). They are easier to start in hard ice than tubular screws and can often be placed with one hand, although it may be necessary to tap them while twisting as they are started. Their holding power is less than tubular screws as they tend to fracture hard ice and, under heavy loads, tend to shear through the ice because of their small cross section.

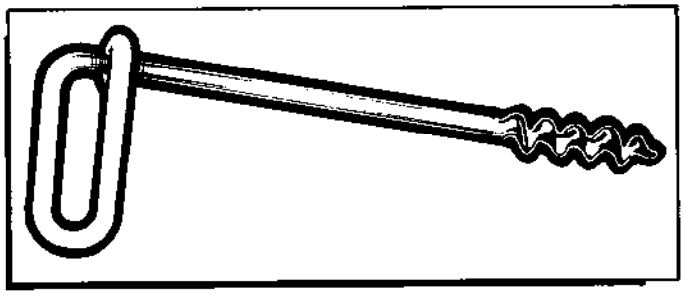

Figure 22-63. Coathanger Ice Screw.

(c) Developed as an attempt to make an easy-to-place and easy-to-remove screw, the solid screws are driven in like a piton and screwed out (figure 22-64). They offer excellent protection in water ice but are less effective in other ice forms. Melt-out is sometimes rapid because of limited thread displacement and, under load, they tend to shear through the ice as do coathanger screws.

(d) Before placement of ice screws or pitons, any soft snow or loose ice should be scraped or chopped away until a hard and trustworthy surface is reached. A small starting hole punched out with the pick or spike of the axe or hammer facilitates a good grip for the starting threads or teeth. The screw is pressed firmly into the ice and twisted in at the same time, angled slightly uphill against the anticipated direction of pull. Ice pitons are, of course, driven straight in, but must also be angled against the pull that would result from a fall (figure 22-65). If any spalling or splintering of the ice occurs, the screw should be removed and another placement tried 1 or 2 feet away. Some glacier ice will spall near the surface but by continuing to place the screw and gently chopping out the shattered ice, a deep, safe placement may be obtained. As a general rule, short screws or pitons should be used in hard ice and long ones in softer ice. They should always be placed in the ice until the eye is flush with the surface. When removing ice hardware, take care not to bend it since this diminishes its effectiveness in future use.

(4) Ice/Snow Bollard. Although not a natural anchor in itself, an ice or snow bollard is easily made from natural materials. A semicircular trench is dug in the snow or ice. The trench should be 3 to 4 feet across and 6 to 12 inches deep. Allow a larger size for poor

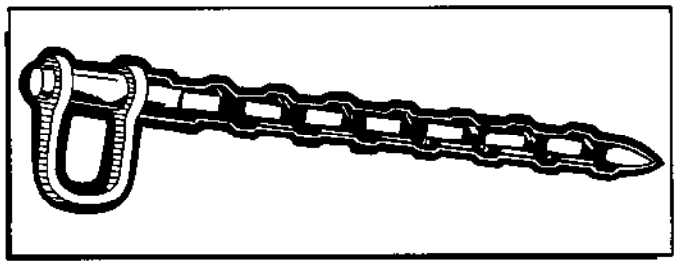

Figure 22-64. Solid Ice Screw.

snow or ice conditions. The rope can be positioned in the trench to provide a downward belay (figure 22-66).

f. Glissading. Glissading is a means of rapidly descending a slope. Consisting of two basic positions, glissading offers a speedy means of travel with less energy exerted than using the descending step-by-step or plunging techniques.

(1) When snow conditions permit, the sitting glissade position is the easiest way to descend. The climber simply sits in the snow and slides down the slope while holding the axe in an arrest position (figure 22-67). Any tendency of the body to pivot head downwards may be checked by running the spike of the axe rudder-like along the surface of the snow. Speed is increased by lying on the back to spread the body weight over a greater area and by lifting the feet in the air. Sitting back up and returning the feet to the snow surface reduces speed. On crusted or firmly consolidated snow, sit fairly erect with the heels drawn up against the buttocks and the boot soles skimming along the surface. Turns are nearly impossible in a sitting glissade; however, the spike, dragged as a rudder and assisted by body contortions, can effect a change in direction of several degrees. Obstructions on the slope are best avoided by rising into a standing glissade (figure 22-67) for the turn, and then returning to the sitting position. Speed is decreased by dragging the spike and increasing pressure on it. After the momentum has been checked by the spike, the heels are dug in for the final halt but not while sliding at a fast rate as the result is likely to be a somersault. Emergency stops at high speeds are made by arresting.

(2) The standing glissade is similar to skiing. Positioned in a semicrouch stance with the knees bent as if sitting in a chair (figure 22-67), the legs are spread laterally for stability, and one foot is advanced slightly to anticipate bumps and ruts. For additional stability, the spike of the axe can be skimmed along the surface, the shaft held alongside the knee in the arrest grasp, with the pick pointing down or to the outside away from the body. Stability is increased by widening the spread of

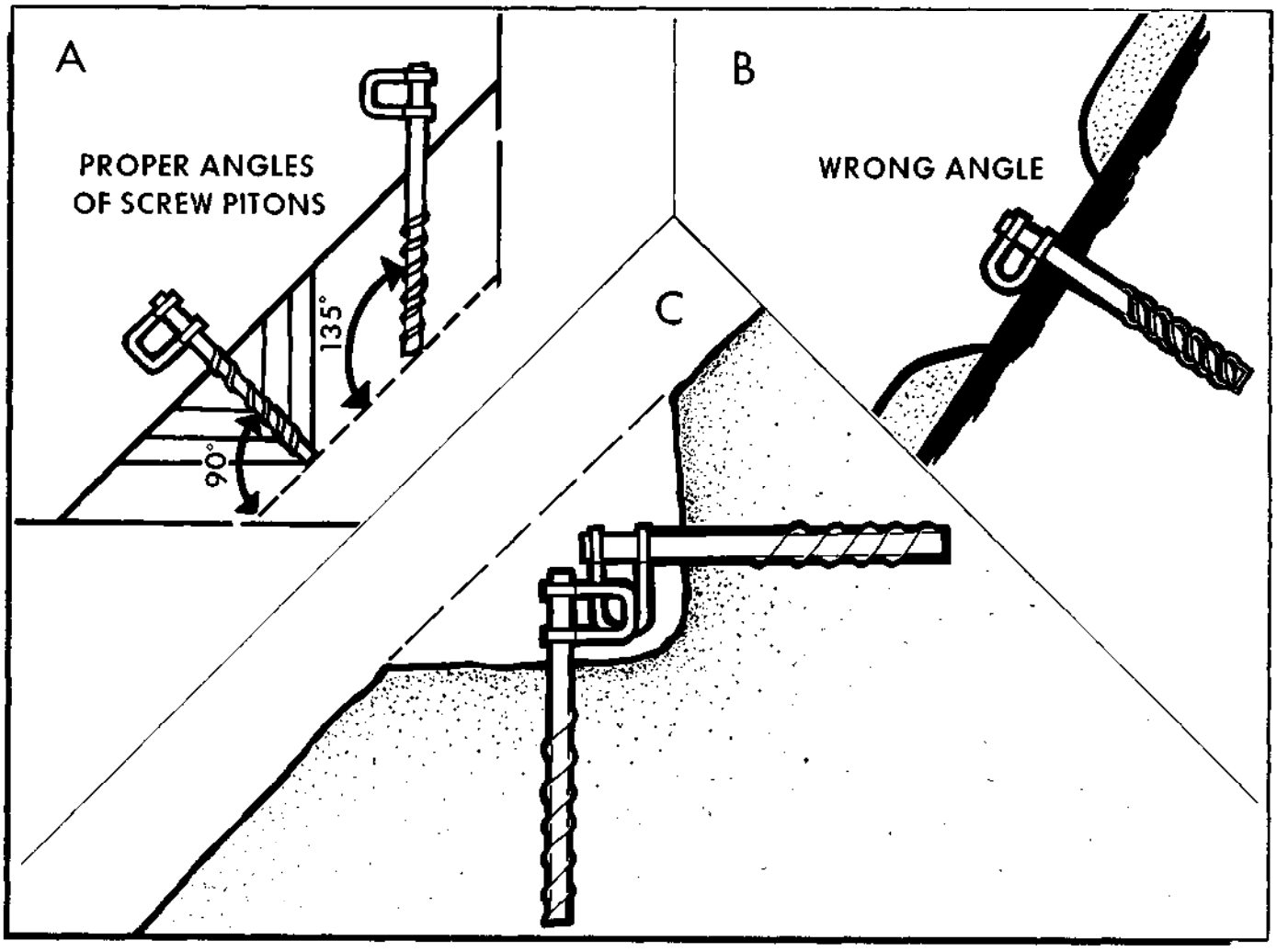

Figure 22-65. Placement of Ice Screw and Piton.

the legs, deepening the crouch, and putting more weight on the spike. A decrease in speed increases muscular strain and the technique becomes awkward and trying, although safe. Speed is increased by bringing the feet close together, reducing weight on the spike, and leaning forward until the boot soles are running flat along the surface like short skis. If the slide is too shallow, a long skating stride helps.

(3) A glissade should be made only when there is a safe runout. Unless a view of the entire run can be obtained beforehand, the first person down the run must use extreme caution, stopping frequently to study the terrain ahead. Equipment must be adjusted before beginning the descent. Crampons and other hardware must be properly stowed. Never attempt to glissade while wearing crampons as it is extremely easy to snag a crampon and be thrown down the slope. Mittens or gloves are worn to protect the hands and to maintain control of the axe. Heavy waterproof pants provide protection to the buttocks. Gaiters are also helpful for all glissading. Glissades should never be attempted in terrain where the axe safety cord is required. The hazards

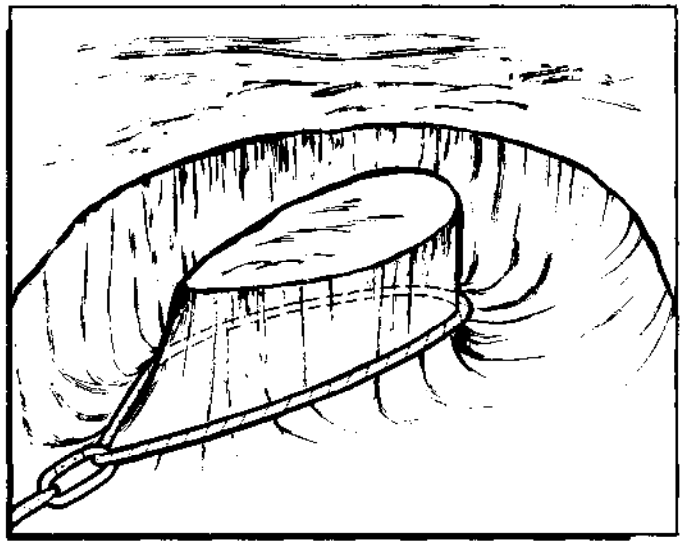

Figure 22-66. Ice/Snow Bollard.

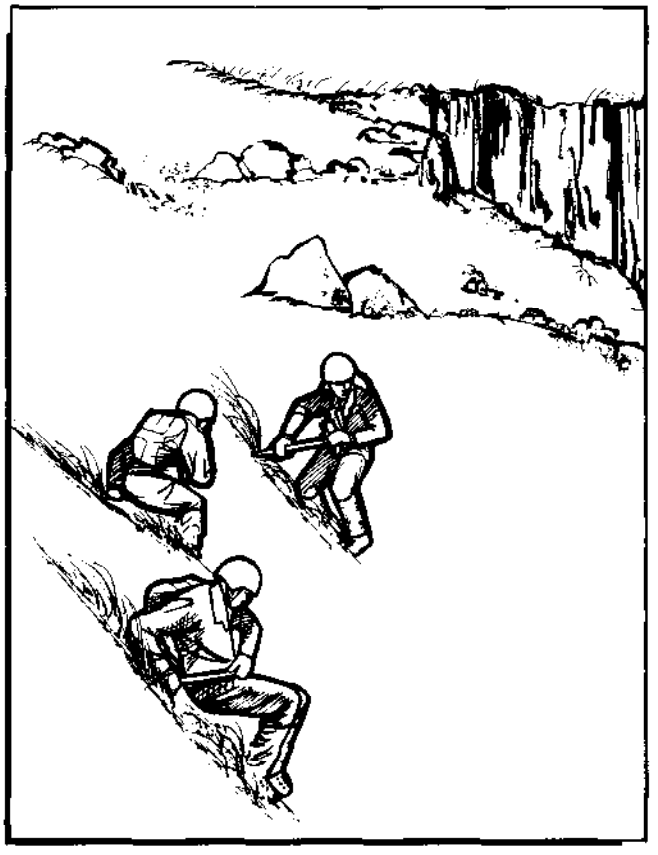

Figure 22-67. Glissading.

of a flailing ice axe should never be risked during a glissade.

22-14. Glaciers and Glacial Travel:

a. Features. To cope with the problems which can arise in using glaciers as avenues of travel, it is important to understand something of the nature and composition of glaciers.

(1) A valley glacier is essentially a river of ice and it flows at a rate of speed that depends largely on its mass and the slope of its bed. A glacier consists of two parts:

(a) The lower glacier, which has an ice surface void of snow during the summer.

(b) The upper glacier, where the ice is covered, even in summer with layers of accumulated snow that changes into glacier ice.

(2) To these two integral parts of a glacier may be added two others which, although not a part of the glacier proper, are generally adjacent to it and are of similar composition. These adjacent features, the ice and snow slopes, are immobile since they are anchored to underlying rock slopes. A large crevasse separates such slopes from the glacier proper and defines the boundary between moving and anchored ice.

(3) Ice is plastic-like near the surface, but not smooth enough to prevent cracking as the ice moves forward over irregularities in its bed. Fractures in a glacier surface, called crevasses, vary in width and depth from only a few inches to many feet. Crevasses form at right angles to the direction of greatest tension and due to a limited area, tension is usually in the same direction. Crevasses in any given area tend to be roughly parallel to each other. Generally, crevasses develop across a slope. Therefore, when traveling up the middle of a glacier, people usually encounter only transverse crevasses (crossing at right angles to the main direction of the glacier). Near the margins or edges of a glacier, the ice moves more slowly than it does in midstream. This speed differential causes the formation of crevasses diagonally upstream away from the margins or sides. While crevasses are almost certain to be encountered along the margins of a glacier and in areas where a steepening in gradient occurs, the gentlest slopes may also contain crevasses.

(4) An icefall forms where an abrupt steeping of slope occurs in the course of a glacier. These stresses are set up in many directions. As a result, the icefall consists of a varied mass of iceblocks and troughs with no well-defined trend to the many crevasses.

(5) As a glacier moves forward, debris from the valley slopes on either side is deposited on its surface. Shrinkage of the glacier from subsequent melting causes this debris to be deposited along the receding margins of the glacier. Such ridges are called lateral (side) moraines. Where two glaciers join and flow as a single river of ice, the debri on the adjoining lateral margins of the glaciers also unites and flows with the major ice stream, forming a medial (middle) moraine. (By examining the lower part of a glacier, it is often possible to tell how many tributaries have joined to form the lower trunk of the glacier.) Terminal (end) moraine is usually found where the frontage of the glacier has pushed forward as far as it can go; that is, to the point at which the rate of melting equals the speed of advance of the ice mass. This moraine may be formed of debri pushed forward by the advancing edge or it may be formed by a combination of this and other processes.

(a) Lateral and medial moraines may provide excellent avenues of travel. When the glacier is heavily crevassed, moraines may be the only practical routes. Ease of progress along moraines depends upon the stability of the debris composition. If the material consists of small rocks, pebbles, and earth, the moraine is usually loose and unstable and the crest may break away at each footstep. If large blocks compose the moraine, they have probably settled into a compact mass and progress may be easy.

(b) On moraine travel, it is best either to proceed along the crest or, in the case of lateral moraines, to follow the trough which separates it from the mountainside. Since the slopes of moraines are usually unstable, there is a great risk of spraining an ankle on them. Medial moraines are usually less pronounced than lateral moraines because a large part of their material is transported within the ice. Travel on them is usually easy but should not be relied upon as routes for long distances since they may disappear beneath the glacier surface. Only rarely is it necessary for a party traveling along or across moraines to be roped together (figure 22-68).

(6) Glacial rivers are varied in type and present numerous problems to those who must cross or navigate them. Wherever mountains and highlands exist in the arctic regions, melting snows produce concentrations of water pouring downward in a series of falls and swift chutes. Rivers flowing from icecaps, hanging piedmonts (lake-like), or serpentine (winding or valley) glaciers are all notoriously treacherous. Northern glaciers may be vast in size and the heat of the summer sun can release vast quantities of water from them. Glacier ice is extremely unpredictable. An ice field may look innocent from above, but countless subglacial streams and water reservoirs may be under its smooth surface. These reservoirs are either draining or temporarily blocked. Mile-long lakes may lie under the upper snowfield, waiting only for a slight movement in the glacier to liberate them sending their waters into the valleys below. Because of variations in the amounts of water released by the Sun's heat, all glacial rivers fluctuate in water level. The peak of the flood water usually occurs in the afternoon as a result of the noonday heat of the Sun on the ice. For some time after the peak has passed, rivers which drain glaciers may not be fordable or even navigable. However, by midnight or the following morning, the water may recede so fording is both safe and easy. When following a glacial river broken up into many shifting channels, choose routes next to the bank rather than taking a chance on getting caught between two dangerous channels.

(7) Glaciers from which torrents of water descend are called flooding glaciers. Two basic causes of such glaciers are the violent release of water which the glacier carried on its surface as lakes, or the violent release of large lakes which have been dammed up in tributary glaciers because of the blocking of the tributary valley by the main glacier. This release is caused by a crevasse or a break in the moving glacial dam; the water then roars down in an all-enveloping flood. Flooding glaciers can be recognized from above by the flood-swept character of the lower valleys. The influence of such glaciers is sometimes felt for many miles below. Prospectors have lost their lives while rafting otherwise safe rivers because a sudden flood entered by a side tributary and descended as a wall of white, rushing water.

(8) On those portions of a glacier where melting occurs, runoff water cuts deep channels in the ice surface and forms surface streams. Many such channels exceed 20 feet in depth and width. They usually have smooth sides and undercut banks. Many of these streams terminate at the margins of the glacier where in summer they contribute to the torrent that constantly flows between the ice and the lateral moraine. Size increases greatly as the heat of the day moves to an end. The greatest caution must be taken in crossing a glacial surface stream since the bed and undercut banks are usually hard, smooth ice which offers no secure footing.

(9) Some streams disappear into crevasses or into round holes known as glacial mills, and then flow as subglacial streams. Glacial mills are cut into the ice by the churning action of water. They vary in diameter. Glacial mills differ from crevasses, not only in shape but also in origin, since they do not develop as a result of moving ice. In places, the depth of a glacial mill may equal the thickness of the glacier.

b. Glacier Operations. The principal dangers and obstacles to operations in glacier areas are crevasses and icefalls. Hidden crevasses present unique problems and situations since their presence is often difficult to detect. When one is detected, often it is due to a team member having fallen through the unstable surface cover. The following techniques and procedures should be followed when performing glacier operations.

(1) Equipment Preparation. The prevention of hypothermia should be of primary importance when performing glacier operations. Sufficient protective clothing must be worn or carried to cover all climatic temperature variations. Climbers trapped in crevasses have died of hypothermia while their team members, helpless to assist from their position on the glacier surface above, were sweltering in sunshine. Backpacks should be equipped with a lanyard consisting of a 6-foot piece of line with a figure-eight and nonlocking carabiner at one end. The free end of the lanyard is attached to the pack and the unlocking carabiner is snapped into the buttock strap of the seat harness. If the climber falls into a crevasse and is suspended upside down by the weight of the pack, the pack can be released with the lanyard and the person can return to an upright position.

(2) Team Composition. The first law of glacial travel is to rope-in during travel. The principal consideration is to avoid crevasses. When stepping onto a known glacier or onto a snowfield of unknown stability,

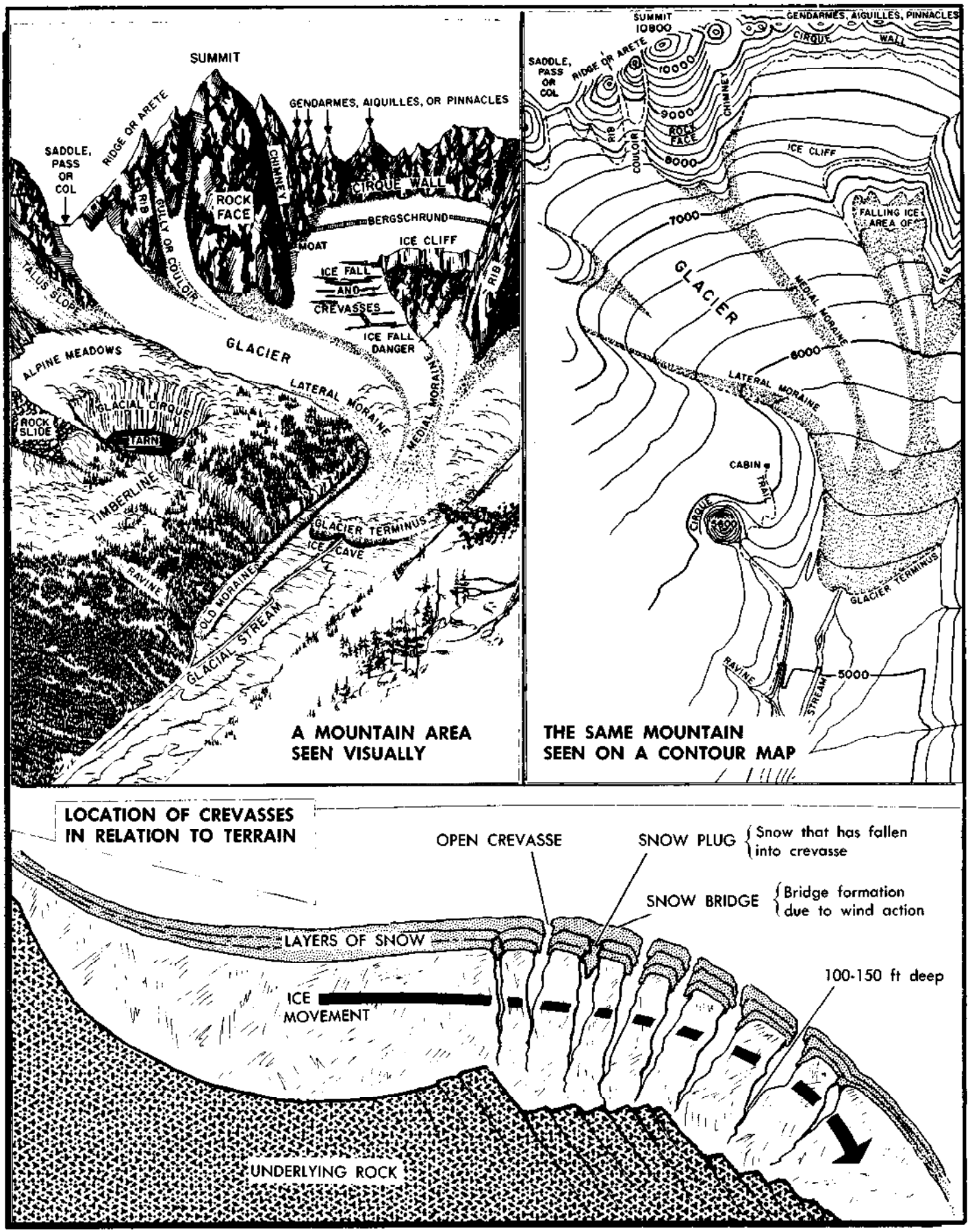

Figure 22-68. Typical Glacier Construction.

whether crevasses are visible or not, the law of roping-in remains. The only variable to this law is when avalanches present a greater hazard than the threat of crevasses. The most experienced climber in glacial travel should be the lead climber; however, if crevasses are completely masked, the lightest climber may lead. During moderate climbs, three climbers tied into a 165-foot rope is ideal. During severe climbs requiring belay, a 120-foot rope with only two climbers is recommended. If a two-person climbing team falls, the team must be arrested by a single axe. If a three-person climbing team is roped in, the rope is usually so shortened that if one climber falls, the others are often dragged in before they have time to react.

(3) Roping-In. Climbers are roped together by constructing figure-eight knots at the ends and middle of the rope. The rope is attached by passing a locking carabiner through the figure-eight and the crotch strap of the seat harness. Associated climbing equipment such as ice axes, slings, and packs are donned. When completely roped in and prepared for travel, there should not be less than 50 feet of rope between each of the climbers. The more rope between the climbers, the better the chance for a successful arrest.

c. Glacier Travel. Due to the difficulty of crevasse rescue, two or more rope teams are recommended for glacier travel since a single team is sometimes pinned down in the arrest position and members are unable to free themselves to begin rescue. Rope teams must travel close together to lend assistance to each other; however, not so close as to fall into the same crevasse. During extended periods on a glacier, skis and snowshoes are often of great value. This footgear will distribute the weight more widely than boots alone and place less strain on snow bridges. Neither skis nor snowshoes are substitutes for the rope, but may be used for easy travel.

(1) Operations in the mountains have certain limitations imposed by nature in glacial movement. Access to the end portion of a glacier may be difficult due to abruptness of the ice and possible presence of crevasses. Additional obstacles of mounting a glacier may be swift glacial streams or abrupt mountain terrain bordering the glacier ice. The same obstacles may also have to be negotiated when dismounting or mounting a valley glacier at any place along its course. Further considerations to movement on a glacier are steep sections, heavily crevassed regions, and icefalls. The use of up-to-date aerial photographs, when available, with aerial reconnaissance is a valuable means of gathering advance information about a particular glacier. The photos, however, only supplement and do not negate the advantages of surface reconnaissance conducted from available vantage points.

(2) Trail wands are used to mark the route and crevasses. The wands, especially essential to safety during periods of adverse weather, are placed every 150 feet along the route and can be used during day or night. A climbing team should not cluster close together during rest stops. If areas of safety cannot be found, the rope must be kept extended during rests just as during travel. A party establishing camp on a snow-covered glacier similarly remains roped-in for as long a period as required to safely inspect the area by stomping and probing the surface thoroughly before placing trust in the site. Hidden crevasses should always be assumed to exist in the area.

(3) Normally a team will travel in single file, stepping in the leader's footsteps or in echelon formation (figure 22-69). If a crevasse pinches out, an end run must be made (figure 22-70) even if it involves traveling half a mile to gain a few dozen feet of forward progress. The time taken to walk around is generally much less than in forcing a direct crossing. Important to remember in an end run is the possible hidden extension of a visible crevasse. A frequent error is aiming at the visible end. Unless the true or subsurface end is clearly visible during the approach, it is best to make a wide swing around the end.

(a) In late summer, the visible end is often the true end due to surface snow and ice having melted. When end runs are impractical because of the distance involved or because the end of one crevasse is adjacent to another, snow bridges may provide a crossing point. One kind consists of remnant snow cover sagging over an inner open space. Another kind, with a foundation which extends downward into the body of the glacier, is less a bridge than a solid area between two crevasses.

(b) Any bridge should be closely and completely examined before use. If overhanging snow obscures the bridge, the lead climber must explore at closer range by probing the depth and smashing at the sides while walking delicately, ready for an arrest or sudden drop. The second climber establishes a belay (figure 22-71) anchored by the third climber who is also prepared to initiate rescue if the leader falls. An excessively narrow or weak bridge may be crossed by straddling or even slithering on the stomach, thereby lowering the center of gravity and distributing the weight over a broader area. When there is doubt about the integrity of a bridge, but it is the only possible route, the lightest climber in the team should be the first across, with the following climbers walking with light steps and taking care to step exactly in the same tracks.

(c) Bridges vary in strength with changes in temperatures. In the cold of winter or early morning, the thinnest and most fragile of bridges may have incredible structural strength. However, when the ice crystals melt

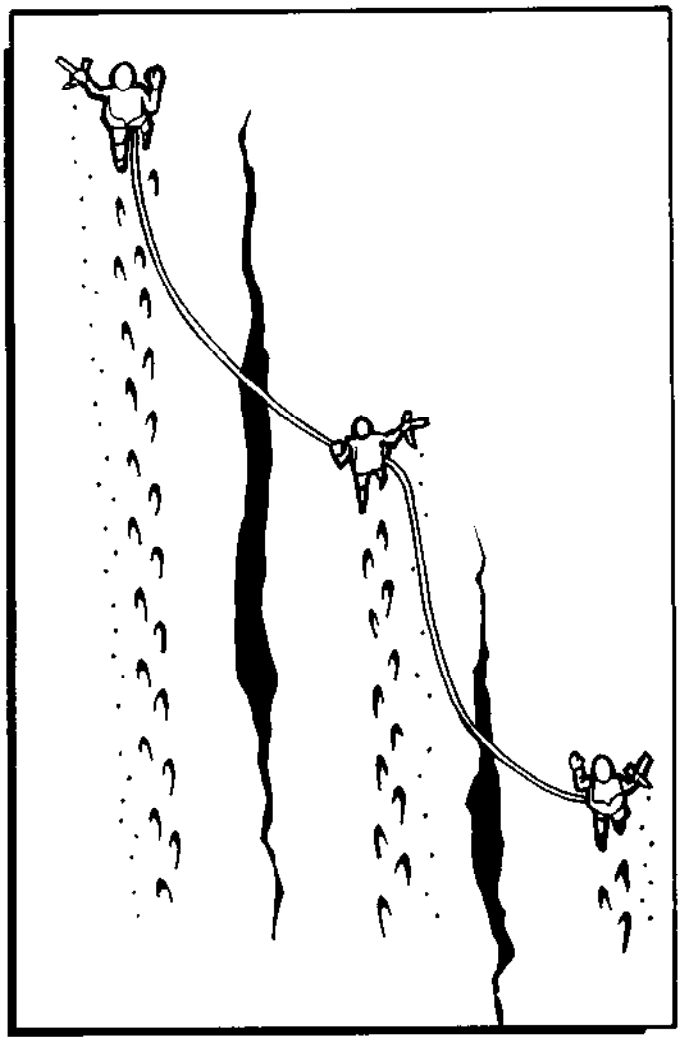

Figure 22-69. Echelon Formation.

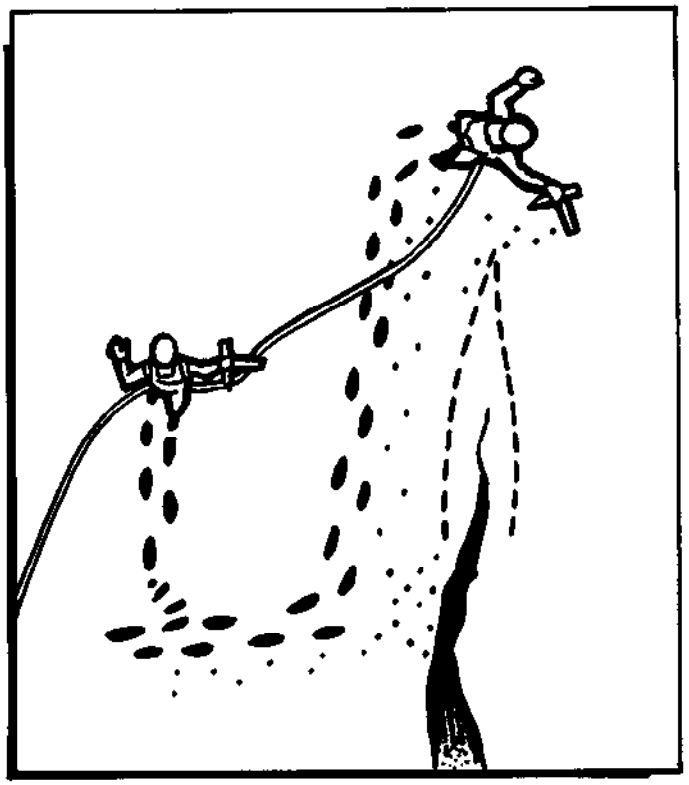

Figure 22-70. End Run.

in the afternoon temperature, even the largest bridge may suddenly collapse. Each bridge must be tested with care, being neither abandoned nor trusted until its worth is determined (figure 22-71).

(d) Narrow cracks in a bridge can be stepped across, but wider crevasses require jumping. If the jump is so long that a run is required, the approach should be carefully packed. A running jump (figure 22-72) can carry the climber further than a standing jump, although running jumps are not often practical. Most jumps are made with only two or three lead-up steps. In any case, care must be taken to locate the precise edge of the crevasse before any attempt is made to jump. Encumbering clothing and equipment must be removed before the jump, although the jumper must bear in mind the low temperature which often exists within crevasses.

d. Crevasse Rescue. Each climber must be able to effect a crevasse rescue if a team member falls into a crevasse. When a climber falls, the remaining team members must drop into a self-arrest position and stabilize their positions. All climbers should never be dragged into the crevasse. If a climber falls, the remaining team members must support the weight until one of them can establish a reliable anchor point or until the second team arrives to help. If the fallen climber is able to assist the recovery, self-extraction from the crevasse may be performed by using prusiks.

(1) A problem inherent to crevasse rescue is the imbedding of the rope (caused by the fallen climber's weight) in the ice and snow. Unless the rope is buffered with an ice axe during the climb out, it will tend to entrench itself deeper into the ice, eventually creating a deep groove in which it will be extremely difficult to use and retrieve, or it will freeze in place, rendering it useless. Corrective actions are to travel down along the rope, taking care not to drop debris on the climber, and free it from the ice. An additional method is to drop a spare rope down to the climber who shifts weight off the imbedded rope until it can be freed.

(2) If the climber is using prusiks and the action of the climb seesaws the rope into the tip of the crevasse, it will be extremely difficult to ascend the few remaining feet. A procedure to overcome this situation is for the climber to tie into the rope near the prusik. The climber then strikes the figure-eight knot from the harness and sends the end to the team above via a retrieving line.

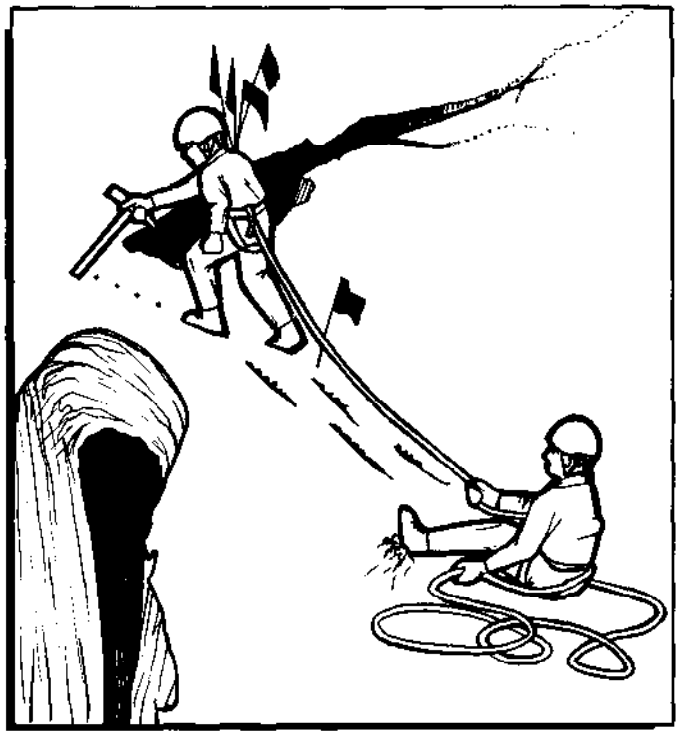

Figure 22-71. Crossing Bridged Crevasse.

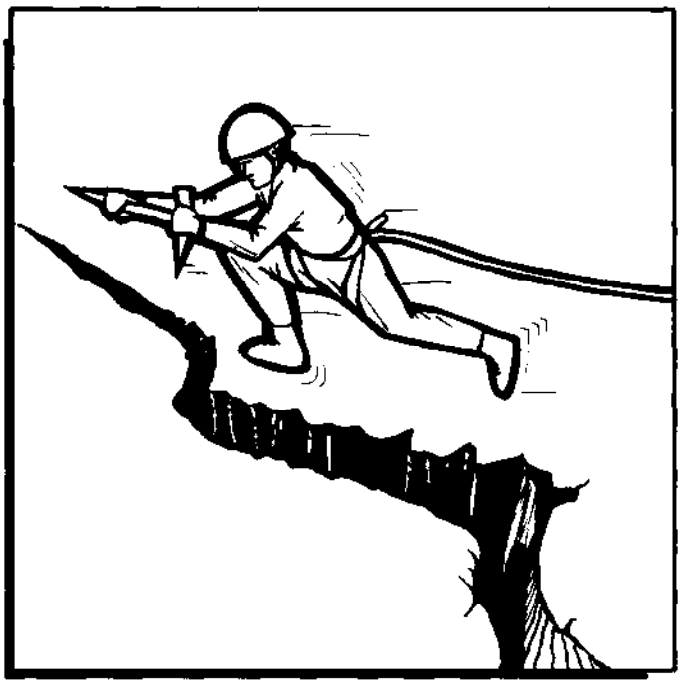

Figure 22-72. Jumping a Crevasse.

Once firmly anchored in place, the rope affords a viable route of ascent. Negotiating the final few feet of a crevasse is usually difficult due to the pressure of the rope against the lip or side of the wall. Prusiks tend to compress against or gouge into the wall rendering them nonfunctional. In most cases, the final few feet are overcome by brute strength. If a second rope is available, an alternate method can be used (figure 22-73).

(3) If a fallen climber is unable to help in the recovery, another climber may be required to enter the crevasse. Before the team member is lowered, all assurances must be determined that the assistance will enhance the outcome of the operation and not compound it. The rescuer should administer medical treatment as required, paying special attention to preventing or treating cold weather injuries as the interior of the crevasse can become extremely cold. Warm protective clothing must be used if the medical situation does not permit immediate extraction.

22-15. Evacuation Principles and Techniques. The performance of mountain rescue is not only physically demanding, but also mentally challenging. Hard and fast procedures to fit all circumstances for mountain rescue cannot be established. The team's ability to innovate will, in most instances, allow the adaptation of the basic principles, techniques, and procedures into a system suitable to effect the rescue. A normal rescue system will use anchors, belays, and various specialized systems. There are basically two methods of rescue: bringing the victim (patient) up to the rescuer's position, or evacuating the victim down from the position.

a. Safety. The establishment of rescue systems must be thoroughly tested prior to use. One missed step in setting up a rescue system may result in further injury of the victim and (or) injury to the rescuers.

b. Evacuations. Evacuation of a victim from the position in a downhill direction is an easier task than establishing a mechanical leverage for pulling a victim to the top of a hill. The victim's medical condition will dictate the method of evacuation and equipment used. The primary litter used is the Stokes litter. This is a tubular frame litter with a wire basket. The tubular main bar provides a very strong framework for mountain operations. The patient may be secured in the litter by means of several cross-body straps, or by the interlacing of slings. When an evacuation team arrives on the scene, there are two activities which should take place at the same time: (1) The patient should be treated and prepared for transporting in the Stokes litter, and (2) the anchor and mechanical brake system must be established. Within obvious restraints of time and distance, low-angle evacuations are always preferable to high-angle evacuations. Low-angle work eliminates many hazards and requires a lesser degree of knowledge and skill. Using the Stokes litter eliminates excessive knot tying and lashing essential on other litters.

c. Preparing the Braking System. The following describes the preparation of the brake system for a low-

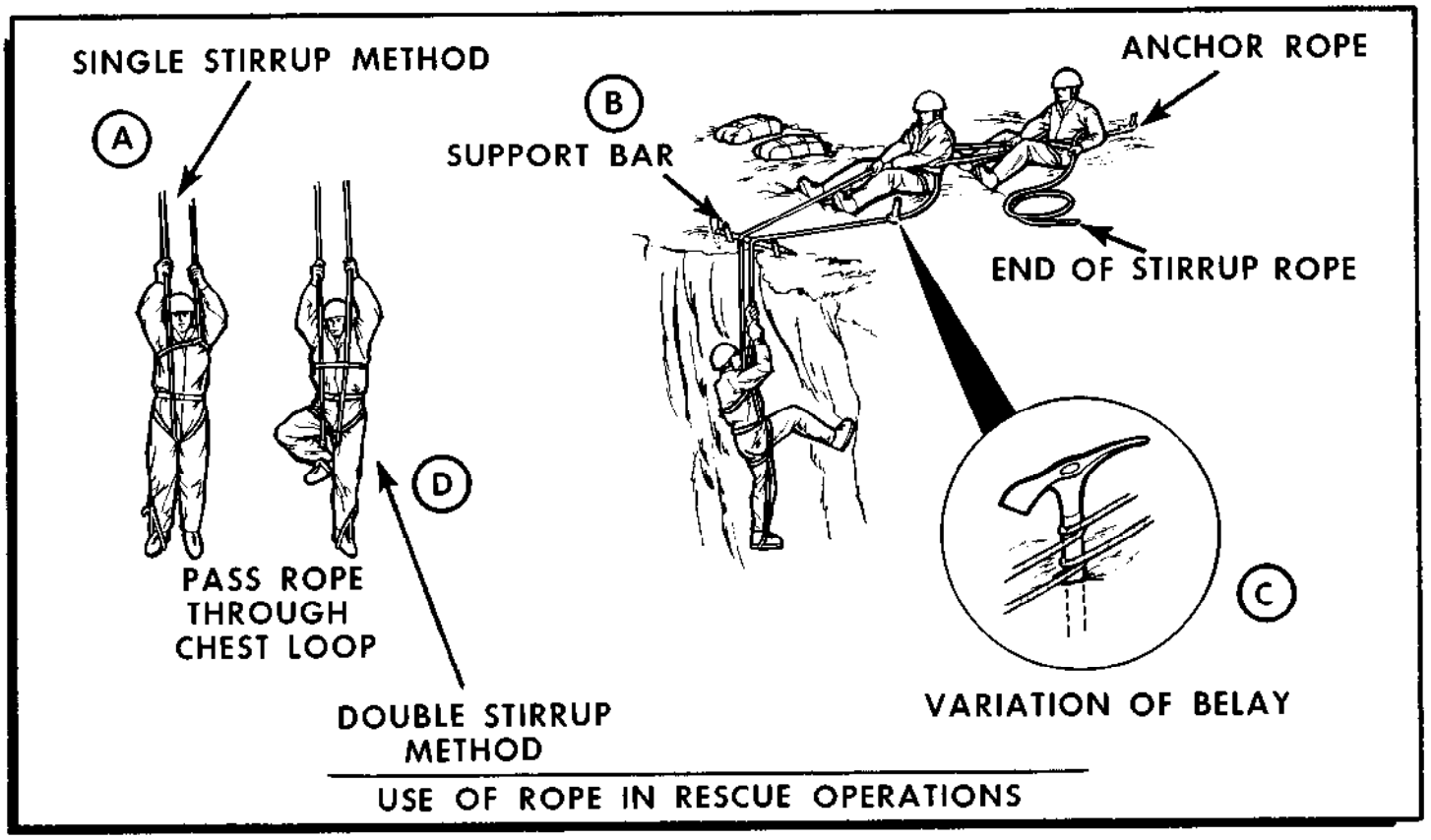

Figure 22-73. Two-Rope Crevasse Rescue.

angle litter evacuation. Establish a very sound anchor, which will be in the direction of pull for the first pitch. If a sound anchor is not available, establish an anchor system. If the terrain does not allow the brake operator ample room to safely and effectively perform the required tasks, than an 11 mm (seven-sixteenth inch) rope sling or double tape sling can be used to adjust for the distance from the anchor to the area where the brake operator will work. The mechanical braking device is securely attached to the anchor or sling. This braking device should either be a figure-eight rappel ring, or a four-carabiner brake system. The rope to be used for the litter descension should be backcoiled. If the rope is to be used for ascending, the rope should lay out along the route of ascension. The head of the litter will be attached to the figure-eight at the end of the rope with a steel-locking carabiner. If locking carabiners are not available, the rope should be tied directly to the litter using several round turns on the outer rail at the head of the Stokes litter and tied off with a bowline and a safety knot. The rope is then properly locked into the mechanical braking system.

d. Preparing the Patient for Transport. While the rope and brake system are being prepared, the litter and patient should also be prepared. The litter must be secured to prevent its loss or further injury to the patient. Additionally, the litter may be padded or insulated (blankets or foam pads) for protection. The ties for securing the feet and pelvis should be attached to the litter. Before evacuating, all emergency medical treatment appropriate to the situation should be performed (splinting fractures, maintaining an open airway, etc.). The patient should be insulated from environmental conditions such as cold, wind, or rain. The person in charge of the patient's medical condition should ensure that the patient's condition is stable enough for transporting. In mountainous terrain, the patient should be protected from further injury due to rockfall by wearing a helmet at all times. A litter team generally consists of four to six people. Fewer than six cannot withstand the fatigue of frequent or long trips while carrying an injured person.

e. Three- or Four-Man Lift. Three bearers take up positions on one side of the victim, one at the shoulder, one at the hip, and one at the knees. If one side is injured, the three bearers should be on the uninjured side. A fourth bearer, if available, takes a position on the opposite side, at the victim's hip.

(1) The bearers should kneel next to the victim. Then, simultaneously, the bearer at the victim's shoulder puts one arm under the victim's head, neck, and

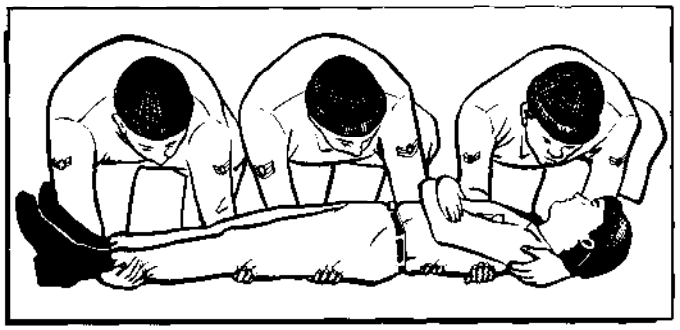

Figure 22-74. Lifting the Patient.

shoulder, and the other under the upper part of the victim's back. Each bearer at the victim's hips places one arm under the victim's back and the other under the victim's thighs. The bearer at the victim's knees places one arm under the victim's knees and the other under the ankles (figure 22-74).

(2) The person at the victim's head gives all the commands. The command "prepare to lift!" is followed by the command "lift." Immediately, all the bearers lift simultaneously and place the victim in line on their knees. If the victim needs to be moved any distance to the litter, move as shown in figure 22-75.

(3) The fourth bearer, if available, places a stretcher under the victim and against the toes of the three kneeling bearers. The command "Prepare to lower!" is followed by the command "Lower!" and the victim is gently lowered to the litter. Once properly positioned in the litter, the victim must be secured in a manner to prevent further injury. The victim may be secured to the litter in a variety of ways depending upon the evacuation route and the victim's condition.

f. Securing Patient in the Litter:

(1) The tape sling used to secure the feet is tied to the framework of the Stokes litter which separates the legs near the groin area. The tape sling should be tied with a clove hitch in the middle of the tape in a manner to prevent the tape from sliding down to the feet when pulled tight. The feet are secured by running the tape across the legs to the window on the outside of the litter then across the patient's legs to the feet. An overhand knot is made in each tape which can be passed over the corresponding foot. When the feet are secured, there should be ample room to apply tension to the head if needed. The tape is then tied at the foot of the litter to a major support bar on the inside of the litter frame. The ties on a Stokes litter should never be made on any outside rail as they are subject to abrasion. The tape slings should be tied off with a two-round turn and two half hitches. If the two-round turn does not hold tension, then a clove hitch can be used in its place.

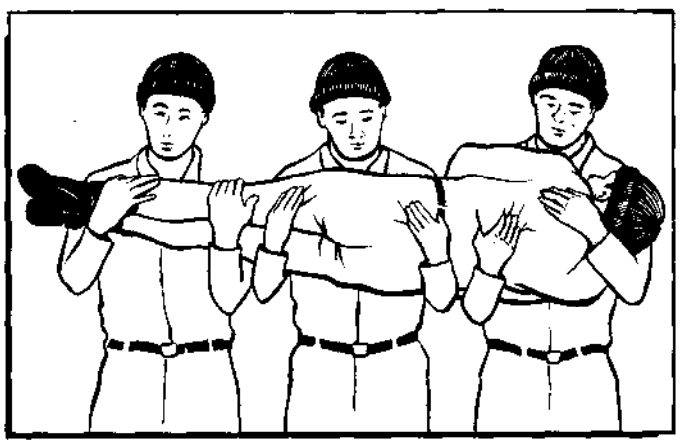

Figure 22-75. Moving the Patient.

(2) The tape used to secure the pelvis should be tied just above the tape used to secure the feet, and secured in the same fashion. Each end of the rope is passed over the leg to the larger upright cross-member of the Stokes litter between the outer rail and inner basket rail. This cross-member corresponds with the side of the hip. The tape is secured with a two-round turn and two half hitches or a clove hitch and two half hitches. The ends of the tapes are then tied together at the middle of the patient's waistline with a square knot and two half hitches on either side of the knot.

(3) The upper torso is secured by placing the middle of the tape in the center of the patient's chest and the two ends of the tape are secured to the large upright cross-member. The running ends of the tape are then passed diagonally across the patient to the cross-member which is next to the abdomen. The tape is secured again and the ends are tied at the midline of the body. The head is secured by running a tape sling over the helmet and securing the tape at the corresponding cross-members. The helmet can be used with a tape sling to provide traction; however, it is not a substitute for the neck collar (figure 22-76).

(4) Once the patient and the system are ready for the low-angle evacuation, the entire system must be doublechecked. Once the litter and patient are prepared as described, ascent or descent is made through a team of litter bearers and a belay point. The minimum essential members for a low-angle evacuation are five—a belayer and four litter bearers (figure 22-77).

g. Low-Angle Evacuation Descent. On the descent, the belay rope attached to the head of the litter runs through a mechanical belay brake system. One team member acts as the belayer. This secures the litter and aids in lowering. Another person may assist the belayer with the rope. Litter bearers take their positions on the litter, ideally three on each side. The chief medic and crew boss should be among these to ensure the patient is

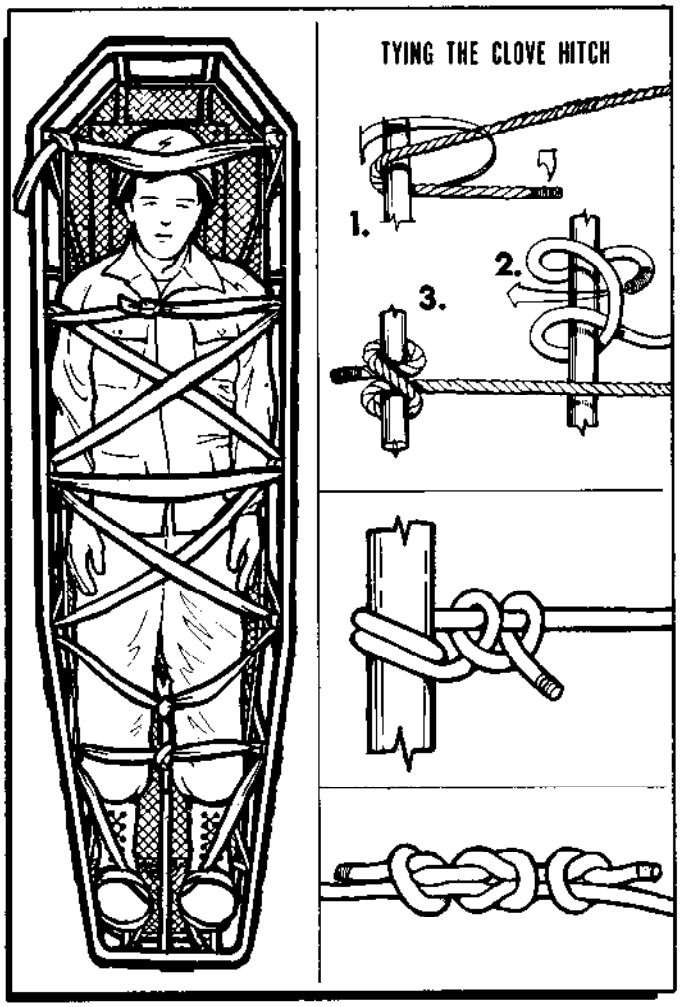

Figure 22-76. Tying the Patient into a Stokes Litter.

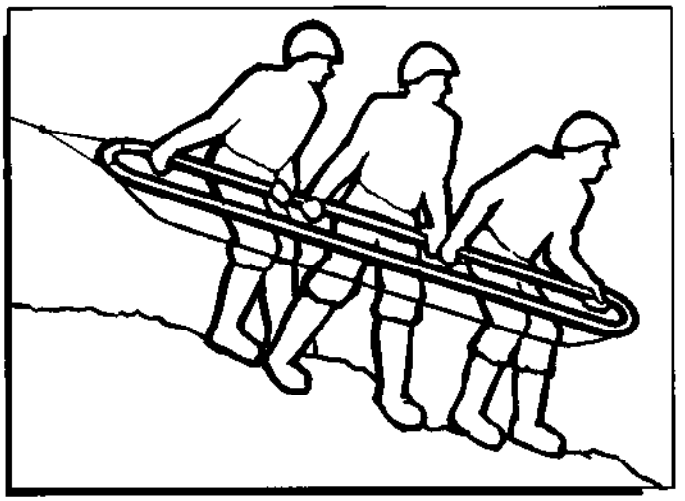

Figure 22-77. Carrying the Litter.

monitored 100 percent of the time and that effective communication exists between the litter bearers and the belayer.

(1) In descending, the most direct, practical passage which takes advantage of available trees and rocks for belay points should be used. Communication is made through a series of commands. As litter bearing is rapidly exhausting, team members should alternate roles. Additionally, a sling attached with a girth hitch to the litter may be used to transfer some of the weight from the arm to the skeletal system via the shoulder. It is also advantageous to use the belay system brake by leaning forward, thus reducing the amount of lifting required.

(2) The scout may precede the team to pick a trail, make the passage more negotiable, or make a reconnaissance so the team need not retrace its course if an impasse is encountered. The scout can also select the site for and secure the next anchor. The scout must remember that the anchors and belay stations must be less than 140 feet apart (with 150-foot rope). Any time the route of descent or ascent changes course more than 90 degrees, a new anchor and brake system must be established or a runner used to change the direction of pull from the belay to the litter. In addition, if the rope is binding against vegetation or rock formations, a sling with a carabiner should be used to pull the rope away from the obstacle.

(3) Once all rescuers have been assigned positions and understand their responsibilities, the litter can be moved. The crew boss will count to three and give the command to lift. The crew boss should coordinate the litter bearers' activities while moving over rough terrain. Standardized commands between the crew boss and belayer are used to control the rate of descent. The following are the commands which should be used:

(a) "ON BELAY?" - Crew boss.

(b) "BELAY ON" - Belayer.

(c) "ROPE" - The crew boss is telling the belayer to feed the rope out in feet.

(d) "SLOW ROPE" - The crew boss is telling the belayer to feed the rope out in inches.

(e) "BRAKE" - Can be said by anyone to avoid a fall or obstacles.

(f) "LITTER SECURE" - The crew boss is telling the belayer that the litter will not be moving and it is attached to the anchor.

(g) "OFF BELAY" - The crew boss is telling the belayer to break the belay system and begin moving equipment to the next station.

(4) When ascending a steep slope, the procedures described in the descent are generally reversed. Additional manpower is required to pull the belay rope through a four-carabiner brake system or figure-eight clog while the litter bearers lift and slowly climb. One person is required to operate the brake system. As this procedure is considerably more fatiguing than descending, the litter bearers should not try to do all the work.

However, as they are ascending the slope, they should assist and hold the litter off the ground.

h. The Buddy System. The buddy system is an evacuation for a slightly injured patient or a patient who is incapable of getting off a precipice. Equipment required is two climbing ropes, four slings, chest, and seat harness for the rescuer, optional seat harness for the patient, and sufficient equipment to construct an anchor system. The following steps are necessary:

(1) An anchor system is set up with a belay device. The first rope is backcoiled as a belay line. A figure-eight knot is tied in the end of the rope with a fisherman's safety.

(2) The second rope is coiled with large enough coils to fit around the shoulders of the patient and the rescuer (figure 22-78). Once coiled, the loops are divided into two groups so that a figure eight is formed. The patient steps into the divided coil so that each leg is through one-half of the figure eight. The knot securing the coil should be in the small of the patient's back, and the coils should be beneath the patient's arms.

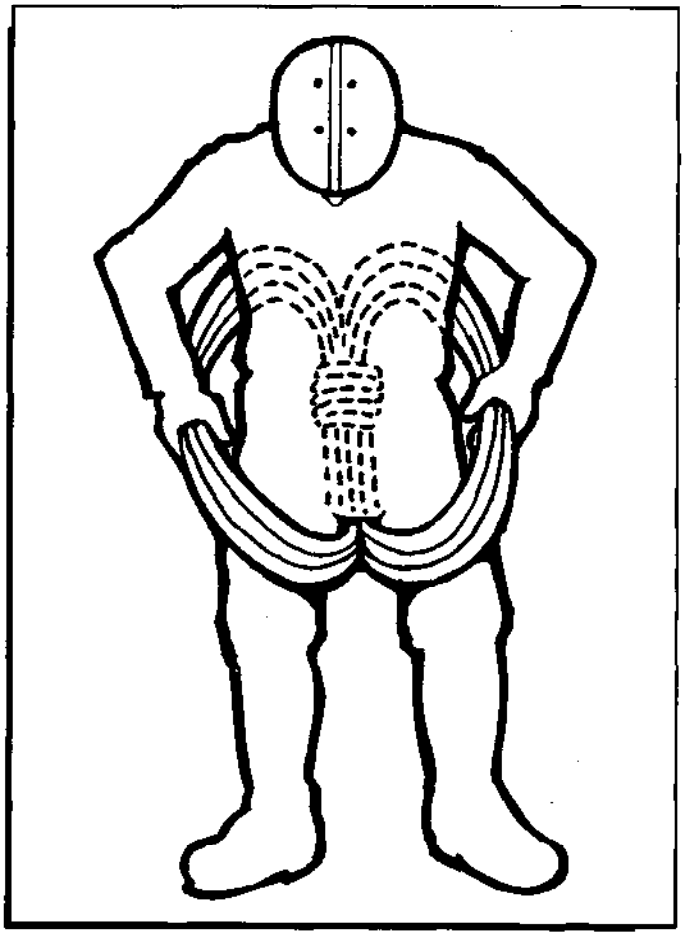

Figure 22-78. Preparing Patient for Buddy Evacuation System.

(3) The rescuer then stands in front of the patient and places each arm through the loops of the coil, half over the right shoulder and the other half over the left shoulder (figure 22-79).

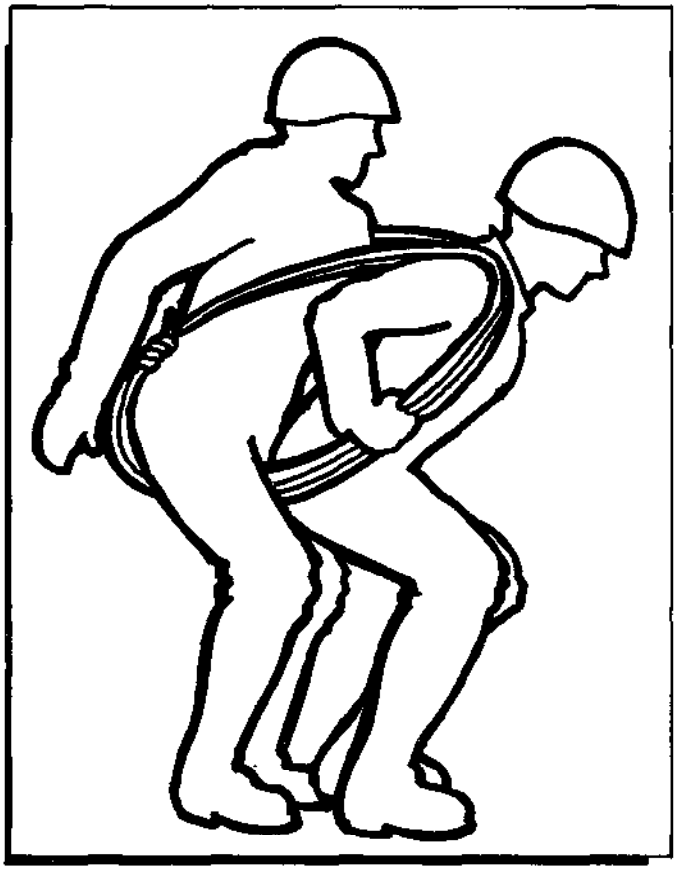

Figure 22-79. Rescuer Donning Rope and Patient.

(4) A sling is passed around the back of the patient, passing under the arms and over the shoulders of the rescuer. The sling is then wrapped around the coil as it passes over the front of the rescuer's shoulders (figure 22-80). The working ends of the sling should pass over the top of the loop formed in the wrap. Ensure the wrap is made low enough not to cross the rescuer's neck or interfere with breathing. The two ends are tied together in the center of the rescuer's chest using a square knot. The tails are taken down to the rescuer's seat harness, one tail is passed through each side, tied in a loop of the harness, and back up to the square knot. Here they are secured to the rope between the coil and square knot by a clove hitch knot on each tail (figure 22-81).

(5) To attach the party to the belay rope, the figure-eight knot on the end of the belay rope is attached to the carabiner on the crotch strap of the seat harness of the rescuer. The two tie end loops are secured with a carabiner sling; another sling is connected to the rescuer's chest harness and the belay rope with a prusik knot (figure 22-82). This sling is used to support the addition-

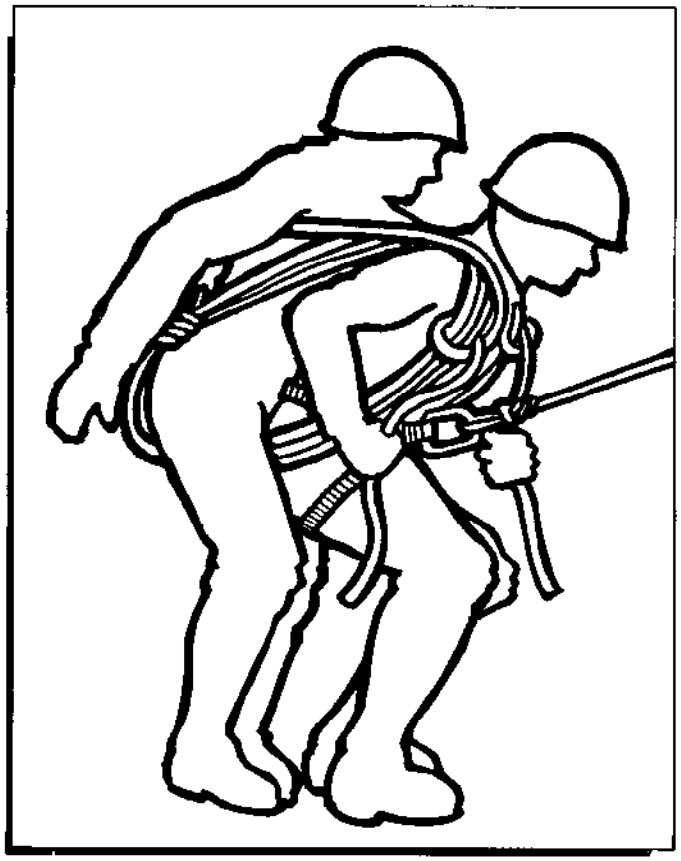

Figure 22-80. Securing the Patient for Buddy Evacuation.

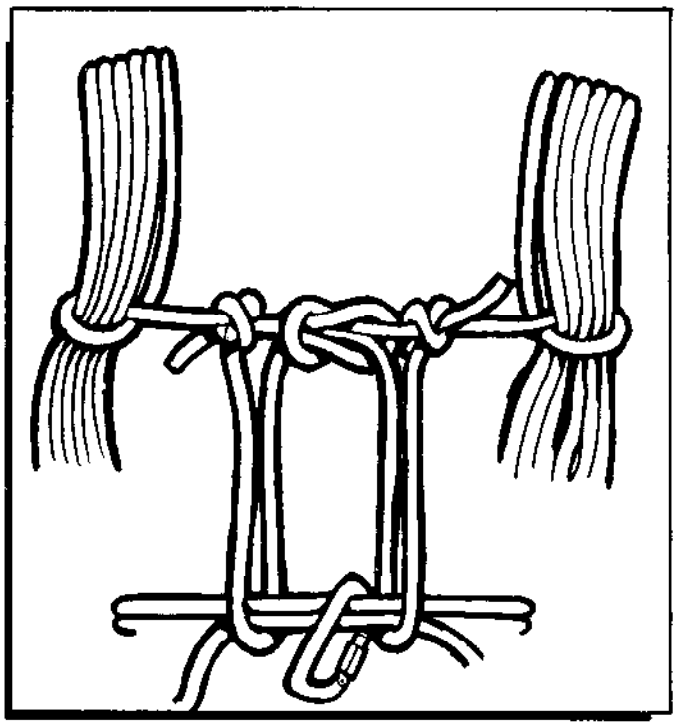

Figure 22-81. Tie-In Procedures.

al weight of the patient. This will enable the rescuer to remain perpendicular to the surface of the rock. Without this sling, the rescuer would fall over backward. The sling should be short enough so the rescuer can adjust the sling during the evacuation. The two people are then belayed down using a mechanical braking device (four-carabiner brake). (See figure 22-82.)

(6) The commands are:

(a) "ON BELAY?" - Given by rappeller.

(b) "BELAY ON" - Given by belayer.

(c) "BRAKE" - The belayer will stop the descent.

(d) "ROPE" - Belayer repeats "Rope." The belayer will feed the rope out in inches.

(e) "PATIENT SECURE" - Given by rappeller.

(f) "THANK YOU" - Given by belayer. The patient and the rescuer are in a safe location from rockfall and will not fall.

(g) "OFF BELAY" - Given by rappeller.

(h) "BELAY OFF" - Given by belayer. The rappeller is no longer secured by the belayer and the belayer can disconnect from the system.

i. Vertical Litter Evacuation. The vertical litter evacuation is used in areas where a horizontal descent is not possible. The patient is secured to the litter in the same manner as for a horizontal evacuation. Two ropes are used for the vertical evacuation. The belay rope to the litter is configured with a figure-eight knot and a locking steel carabiner. A double-tape sling is passed between the two steel frame bars forming the main body at the head of the litter (figure 22-83). The sling should pass outside of the four joining bars between the frame bars. Both ends of the sling are connected into the belay rope carabiner. Two carabiners are connected to the windows on one side of the litter. The second rope runs down the side of the litter through these carabiners to finish at the foot of the litter. A figure-eight knot is tied in the end of the rope. The seat harness of the barrelman attaches to this knot. A figure-eight knot is tied in the end of the sling and a carabiner clipped into the knot and connected to the foot of the litter (figure 22-84). A knot is added to the other end of the sling. A carabiner is added to the mariner's loop and connected to the barrelman's chest harness. As in the horizontal evacuation, the commands are given by the barrelman. The commands are the same as in the high angle horizontal litter evacuation (figure 22-85).

j. Horizontal High Angle Litter Evacuation. The horizontal litter evacuation is the preferred position for an injured patient. This position allows for easier medical treatment and is required for shock prevention. The danger of this position is the patient's exposure to rockfall. Because of its complexity, it is also the most

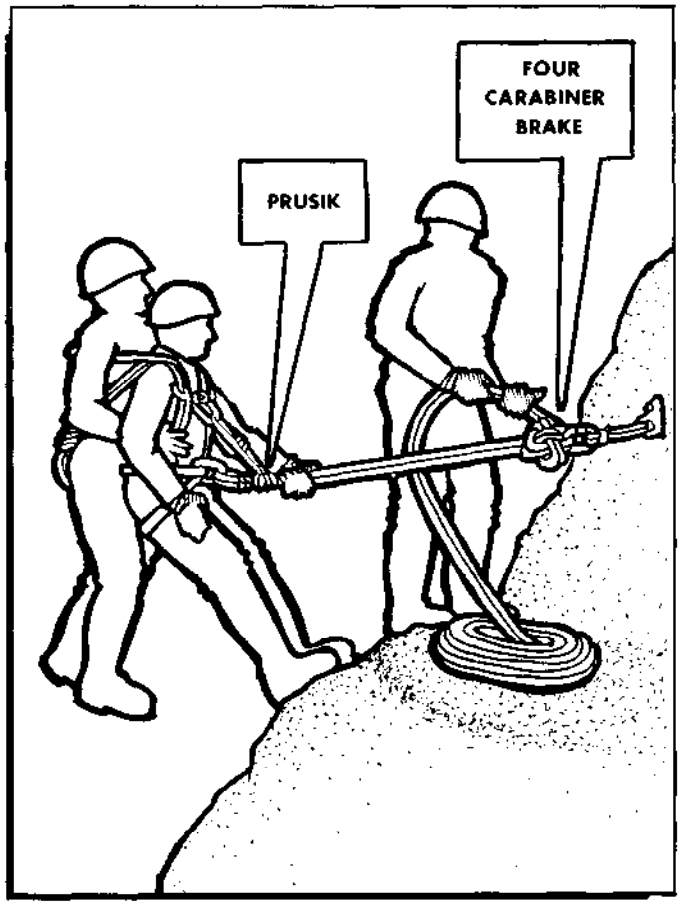

Figure 22-82. Hookup for Buddy Evacuation.

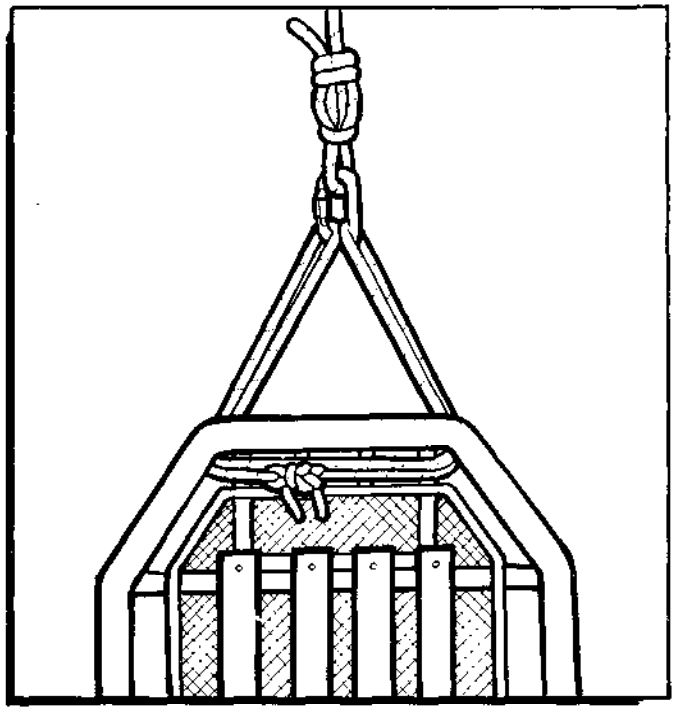

Figure 22-83. Attaching Rope to Stokes Litter.

potentially dangerous evacuation and should not be done unless no alternative exists.

(1) Team Composition. This method requires a minimum of three highly qualified team members. It also requires a great deal of equipment. It should not be attempted without the required equipment and qualified team members. Because the relative nature of the high-angle litter evacuation is potentially dangerous, the margin of safety must far exceed the stress on equipment. The safety factor is greatly increased by using equalizing anchors, two ropes for lowering a belayer, a rope handler, and a mechanical brake system. The high-angle evacuation on vertical walls is performed by only one litter bearer with the litter fully supported by the lowering ropes. However, on walls which are not vertical, it requires a great deal of strength to pick up the litter and walk it down. When an injured victim is on a steep wall, a second person will administer first aid and then assist in loading the victim into the litter. In fact, this "second person" may perform the technically most difficult part of the evacuation in loading the injured person. In some cases, a "third person" can help load a severely injured patient. The job of the barrelman is to get the litter to the victim, to see that the victim is properly loaded and secured, to give the victim a smooth evacuation by preventing the loaded litter from knocking against the rock, and to administer first aid to the victim if necessary. For first aid purposes, the litter should normally be horizontal. Proper operation of the brake controlling the litter is essential.

(2) Rigging the Litter. Two 7/16-inch Mountain Lay ropes are used for lowering, each terminating in a figure-eight knot with a 4- to 5-foot tail, attached to a large locking steel carabiner. Each tail will terminate with a figure-eight knot and safety knot. One end will be attached to the seat harness of the barrelman and the other will be attached to the patient. These ropes should be the same length for ease in rope changes, and ideally the ropes will also be matched in elasticity and of similar wear. The spiders are the nylon slings, between 30 to 36 inches long, which attach the litter to the rope. These spiders are attached to the litter with large locking carabiners. These large carabiners are clipped over the outer rail of the Stokes litter and should have their gates facing in and locked (figure 22-86). If these carabiners are not available, the spiders can be attached to the major upright supports in the window of the litter with a double round turn and two or more half hitches. There are four spiders. One group of two spiders is attached to the windows at the sides of the litter at the head. The head group of two spiders is attached to the window at the sides of the foot of the litter. The spider groups should be connected with the key locking carabiner. This carabiner is attached to both of the ropes (figure 22-87). The

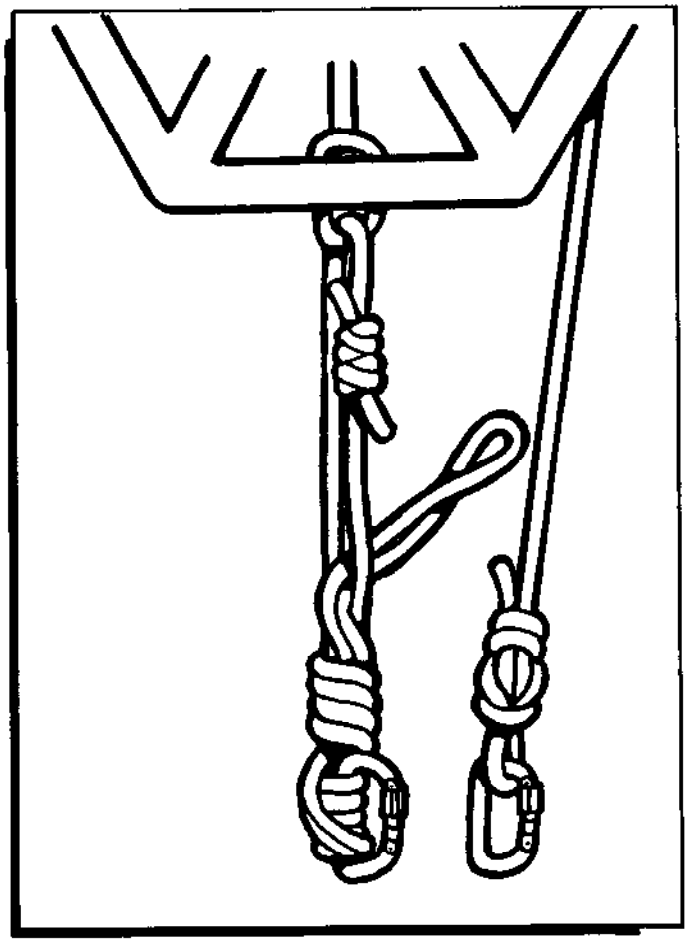

Figure 22-84. Tie-Ins for Foot of Litter.

spiders should be adjusted so the litter is horizontal or with the head only slightly elevated. The patient is tied into the litter in the same way as for the low-angle litter evacuation. However, the patient should have a chest or seat harness on. The figure-eight knot from one of the lowering ropes must be attached to the chest harness with a standard locking carabiner.

(3) The Barrelman. The barrelman is attached to the other lowering ropes by the figure-eight knot at the end of this rope. This rope is attached to the barrelman's seat harness with a standard locking carabiner. This is a safety line only. The majority of the barrelman's weight is supported with a 7 mm line which attaches the barrelman's seat harness with a standard locking carabiner to the key character. This line is secured with a prusick or Bachman's knot. The mariner's knot can be adjusted so the rescuer's feet are flat on the rock below the Stokes litter. The rescuer should be able to shield the patient with the upper body.

(4) Anchors. The anchor location should be directly above the victim if possible. For an evacuation on a large face when visibility is good, a spotter on the ground with a plumb bob can direct an accurate alignment of the anchor over the victim by communicating to the people on top. If the cliff is not vertical, the observer must observe the cliff "hang-on" and must be in the vertical plane containing the fall line through the victim. If the cliff is irregular, the plumb bob technique may be unreliable, but the observer may still be the best source of information in locating the anchor. If the anchor is placed in such a way that the rope will have to pass over a sharp corner of rock at the lip of the face,

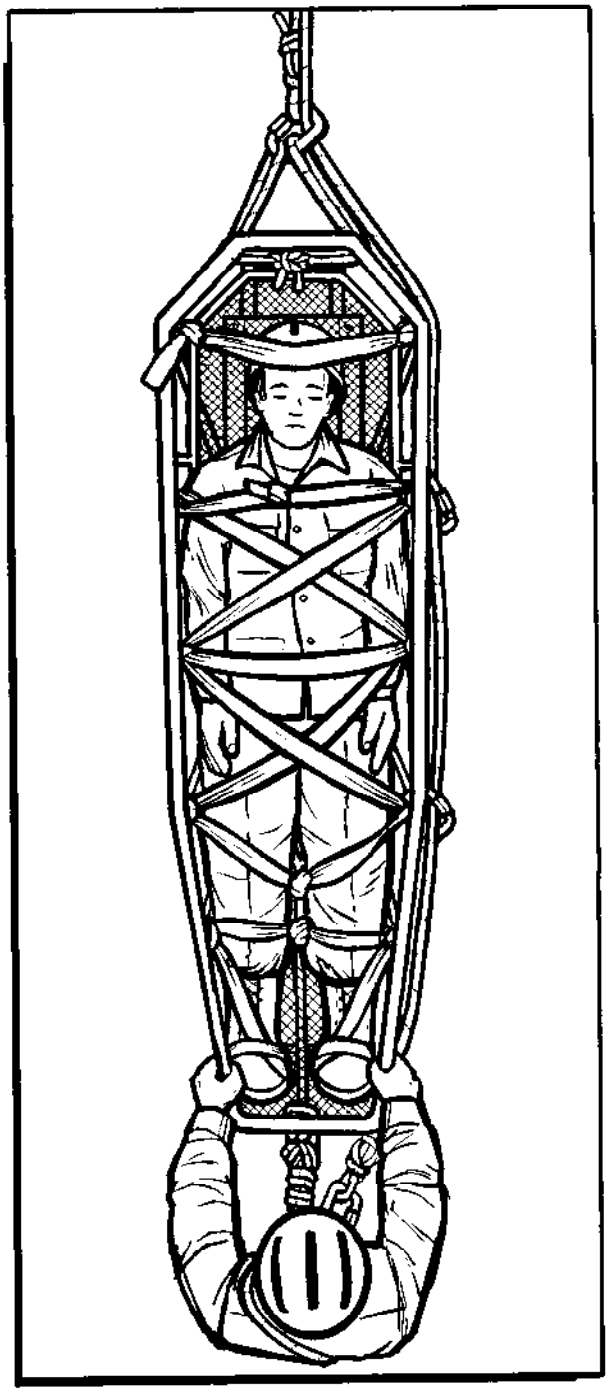

Figure 22-85. Horizontal Descent of Litter.

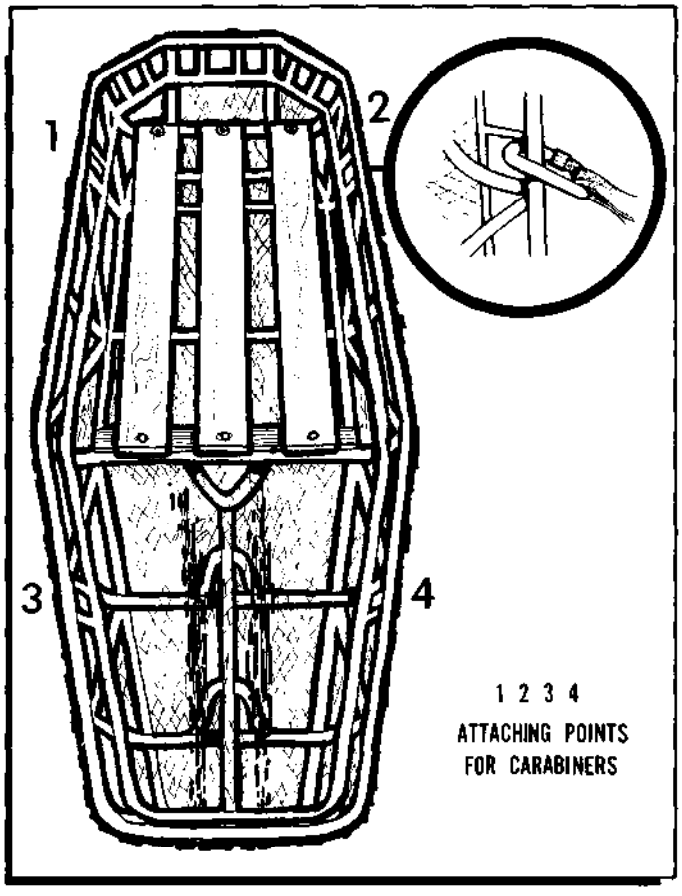

Figure 22-86. Attaching Points for Carabiners.

the corner should be "softened" by breaking the edge with a hammer and by securing padding. If possible, an equalizing anchor system (figure 22-88) should be used for anchoring the top brake systems. This is the safest anchor system to use because it will adjust for a shift in position and the shock load will be equally distributed if one of the individual anchors fail, the litter bearer falls, or the braking is of a fast descent.

(5) Braking System. The four-carabiner or figure-eight clog brake system can be used; however, the four-carabiner system is superior. This brake system will have a brake operator assigned. The brake operator obviously plays an important part in a successful evacuation, and should use gloves and avoid touching a hot braking device. To maintain full control while lowering, the brake operator should never let go of the ropes until tied off or "off belay." The operator should keep the descent smooth and steady, feeding the ropes equally into the brakes and never allowing a kink to form. A kink may jam in the brake and be difficult to release. Smoothness is easily produced if the loading by the litter is steady and has sufficient weight to require the brake operator to maintain a moderate grip on the rope and let the litter do the pulling. The litter should normally remain horizontal; thus the two lines lowering the ropes should be fed equally into the brake. The ropes may be grasped together by one hand and fed equally, even with unbalanced loads. It is often helpful to the brake operator to have a rope handler pull the rope which is not feeding rapidly enough through the brakes. This is particularly useful when the litter is lightly loaded.

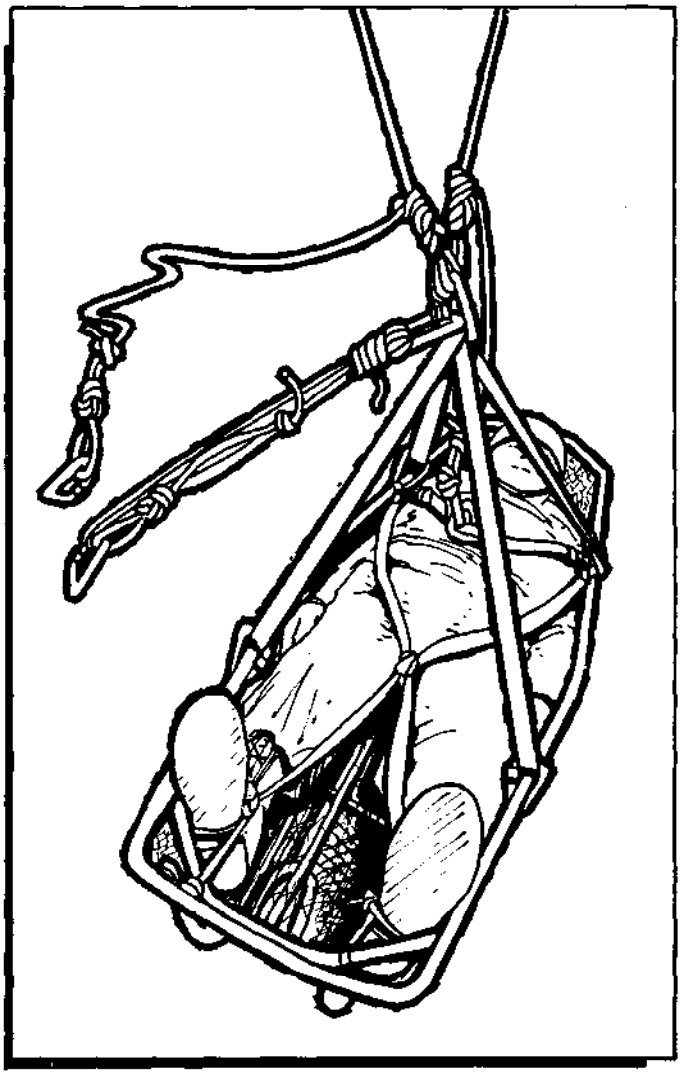

Figure 22-87. Rigging the Stokes Litter for Horizontal Evacuation.

(6) Rope Management. The proper management of the rope by the team at the top is essential for efficient evacuations. In fact, the overall job on top usually involves more thought and skill than is required by the litter bearer. The operations on top are best done by at least two people. The brake operator directly controls the speed of descent. A rope handler provides slack rope with no kinks to the operator and assists on the brake when required. The rope handler's tasks are crucial since the evacuation comes to a rapid halt if the ropes

become snarled. The ideal situation on top is to be able to lay the lowering ropes out in a straight line in their entire length. The rope handler will then have an easy time assisting the operator. If space is limited, the ropes can be stacked separately and neatly. There may be a tendency for the rope to kink at the brake, in which case the rope handler will stay busy twisting and spinning the slack rope.

(7) Handling the Litter. The barrelman will always wear a hardhat, pack, and gloves. The reason for this is that in case small rocks fall toward the helpless victim, the barrelman can lean over the victim providing a shield; and the packs and hardhats protect the barrelman. The victim will also wear a hardhat (when possible). A redundant barrelman tie-in is used for safety, and the locking of doubled carabiners should be used to prevent the barrelman tie-in from being accidentally released. Smooth handling of the litter by a skilled barrelman is the best way to ensure a safe ride down for the victim; banging the litter against the rock is very hard on the victim, and poor handling of the litter may result in injury to the barrelman. In starting the evacuation, the barrelman may have to carry the litter a short distance at the top of the face taking care that ropes are positioned over a smooth rounded edge of the cliff and the spider carabiners have gates facing in. Once their weight is on the lowering ropes, the barrelman's main job is to hold the litter away from the rock. The feet are used against the rock as they are in rappelling to maintain footing. The barrelman's position on a nearly vertical wall is to be hanging from the tie-ins which may be a comfortable seat sling. The barrelman (figure 22-88) is normally on the outside of the litter (away from the rock), gripping the outer or an underneath rail to hold the litter still farther off the rock to clear projections, etc., or to level the litter if loading is uneven. Occasionally, the rail at the cliff face is grasped, but care must be taken to avoid getting crushed hands.

(a) The descent is easy on flat, smooth, vertical walls. On a wall with obstructions, footwork techniques are the same as in rappelling. The legs should be perpendicular to the wall and spread comfortably apart. The Prusick or Bachman's knot should be adjusted so the barrelman moves the litter to the edge of a vertical face and may require readjusting so the space between the litter and the barrelman can be maintained. The barrelman will be bent at a 90-degree angle at the waist which will enable the barrelman to bend over to protect the patient or administer first aid.

(b) The litter must be secured to an anchor to keep it from sliding and falling. The lowering ropes should be removed from the spiders and the knots untied to allow the ropes to be hauled back to the summit without a knot jamming. The spiders themselves should not be dragged up the cliff because of the danger of jamming. Normally, the ropes will be pulled up and carried down rather than dropped because of the possibility of snagging or damage. After the evacuation is complete, the litter should be moved to an area sheltered from falling rock.

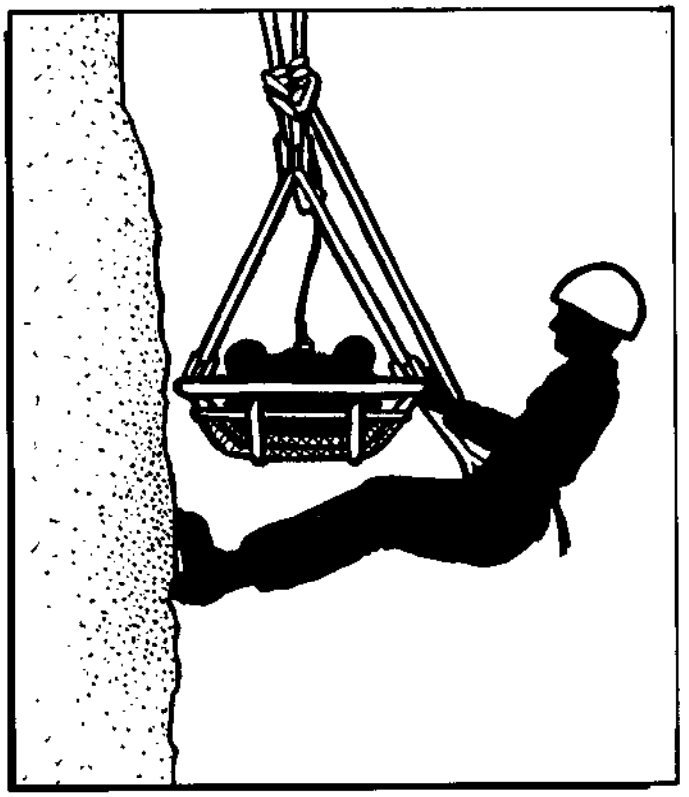

Figure 22-88. Body Position for Rescuer.

(8) Communication. The general goal of the brake operator is to lower the litter smoothy and safely as directed by the barrelman. The commands to be used are as follows:

(a) "ON BELAY" - Has the usual meaning and is used before the evacuation begins.

(b) "ROPE" - Means for the brake operator to feed the lowering ropes equally through the brake at a moderate speed.

(c) "SLOW ROPE" - Means for the brake operator to feed the lowering ropes equally through the brakes at a slow speed.

(d) "BRAKE" - Means for the brake operator to brake both lowering ropes simultaneously.

(e) "SECURE" - Means to maintain the brake until the litter is anchored or in a position where it will not fall.

(f) "UP ROPE" - Means the litter is detached and all knots and slings have been removed. The rope can be pulled up without becoming caught on rock outcroppings or in cracks.

k. Horizontal Hauling Lines and Tyrolean Traverse. A suspension system used to transport a load off the ground along a taut (static) line may be useful when

crossing streams, gorges, or other difficult terrain. These systems can be used for team movement or for transporting victims. The two-rope bridge is time consuming to set up, but in spite of the complexity the techniques can be useful in rescue. However, time should not be wasted waiting for the system to be set up. Sometimes the rigging may be done while the victim is being transported to the crossing site. Since system stresses are often great, the margin of safety is smaller than usual or desirable in rescue operations, and the rigging should be supervised by someone experienced in the techniques.

(1) Anchors. A major requirement for any suspension system is the availability of secure anchors at each end. Anchors have to be as strong for this use as in any mountain rescue. To withstand the high-line stress, the anchors may have to be a combination of several anchor points rigged to be self-equalizing. For reasonable safety, each anchor system should be able to hold at least 10 times the load to be transported.

(a) The anchors must be suitably placed in addition to being strong. To prevent the load from "bottoming out" or to minimize line tension by increasing sag, the anchors must be placed in a high location. If the rope contacts any sharp rock, the rock edges must be padded with suitable material such as leather gloves, pack, tree branches, etc., since a taut rope can be easily cut. The anchors should be 6 to 10 feet back from the edge of the stream or cliff to provide for loading and unloading. The rope should be 3 feet or more off the ground at these positions, and if good anchors do not allow this, A-frames may be improvised from sturdy tree trunks.

(b) If possible, two ropes or a double rope should be used to increase the safety margin. However, one rope will work although the strength of the system is decreased. It is very important to understand the stresses involved with this system. For example, if the rope span is 40 feet and the sag in the middle of the rope is 1 foot with a 200-pound load, the rope tension is approximately 2,500 pounds. However, if the sag is 4 feet in a 40-foot span, then the tension is 650 pounds.

(2) Establishing the Two-Rope Bridge. Connect the center of the rope through the anchor. Do not tie, but simply pass through a carabiner. If the span requires two ropes to be tied together, a double figure eight should be used with single fisherman safeties. A carabiner must be placed in the double figure eight; if not, the double figure eight will be difficult to untie. One team member will cross the obstacle carrying the two ends of the rope plus the hauling rope. Once across, an anchor is established and one end of the rope connected to it (figure 22-89). Remove as much slack as possible so the knot will be taut against a 4-inch pulley or carabiner on the opposite side when tied off to the anchor system. The other end of the rope is then put through a 3:1 mechanical advantage system.

(a) To use the mechanical advantage, a safety prusik is attached to the rope. This safety prusik must be attached to a stout anchor and be as long as practical

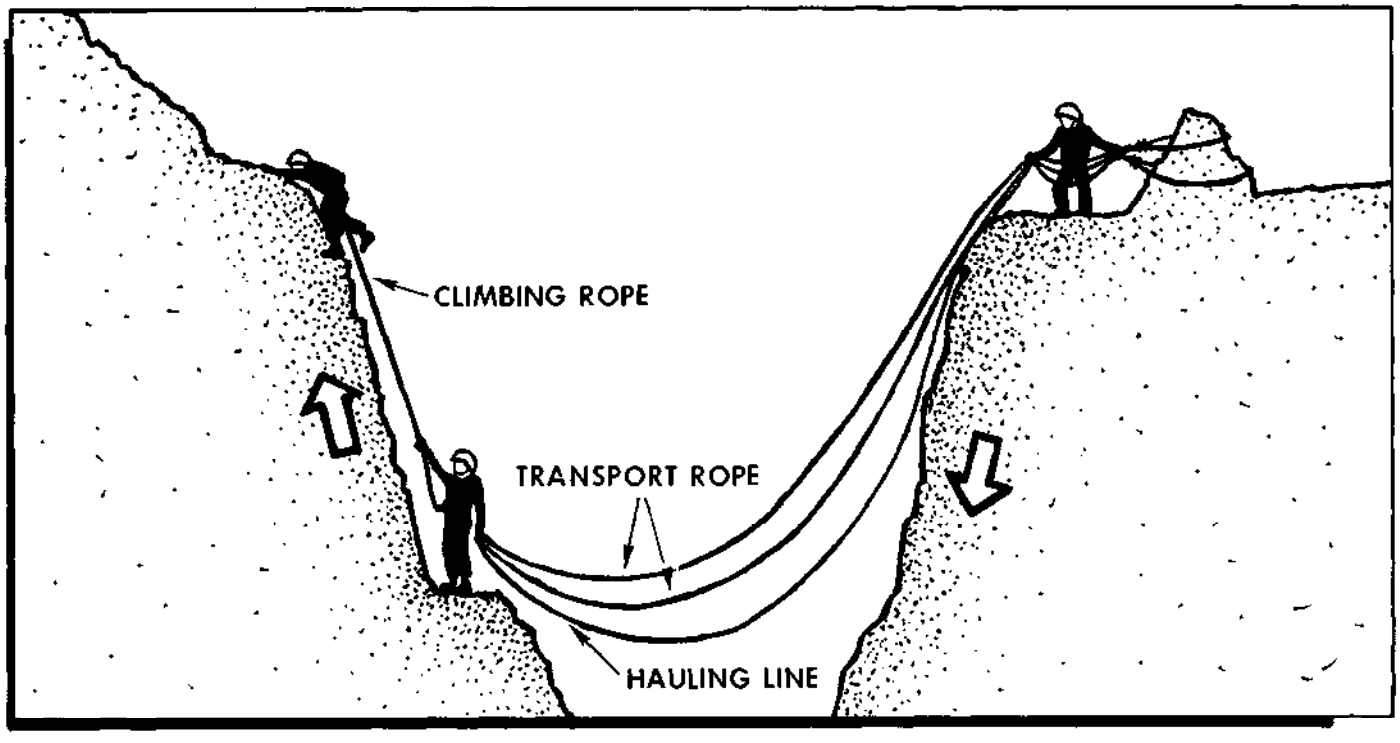

Figure 22-89. Horizontal Hand Evacuation System.

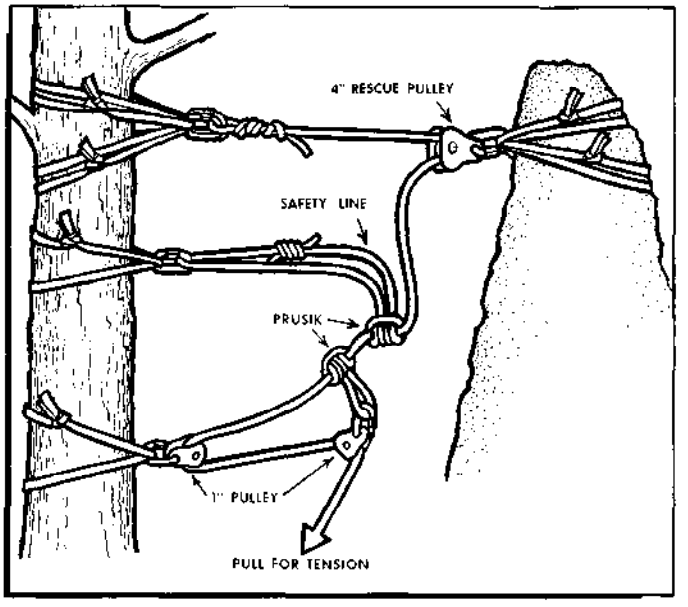

Figure 22-90. Rigging Hand Evacuation System.

to hold the tension on the rope and to slide forward as the slack is pulled out of the rope. The rope is then passed through a carabiner which is attached to an anchor. A smaller prusik sling is tied on the rope behind the first prusik. A carabiner is attached to the prusik sling. The rope is then passed through this carabiner. The rope should resemble a "Z" when the running end is pulled toward the anchor. This is called a "three to one mechanical advantage" because there are three moving lines. The force of pull on the running end of the rope will be increased (ideally) three times. This does not account for the friction of the rope passing over the carabiners or pullies.

(b) Using the mechanical advantage, the following steps should be used to tighten the rope. Pull on the hauling line or the running end of the rope until the snap links come together. Have another person keep this safety prusik slide out as far forward as possible. Then let the safety prusik take up the load, while slowly letting slack out of the running end of the rope. Then slide the second prusik out as far as possible, and repeat the entire maneuver. When the rope is tight enough, tie the rope off (figure 22-90).

(3) Crossing the Two-Rope Bridge. Half the team will now cross the rope by means of the Tyrolean traverse. This is a means of crossing a suspended rope while being secured to the rope. A pulley or carabiner is attached to the rope and clipped into the seat harness. The person then hangs under the rope and pulls hand over hand to the other side. The load is placed on the rope by pulleys or carabiners. The end of the second rope is tied to the load and used to pull the load across the rope to the other side (figure 22-91). Once the patient is secured in the litter, the slings to hold the litter horizontally can be attached. There are six slings in all, one set of three at the foot of the litter. In each of the

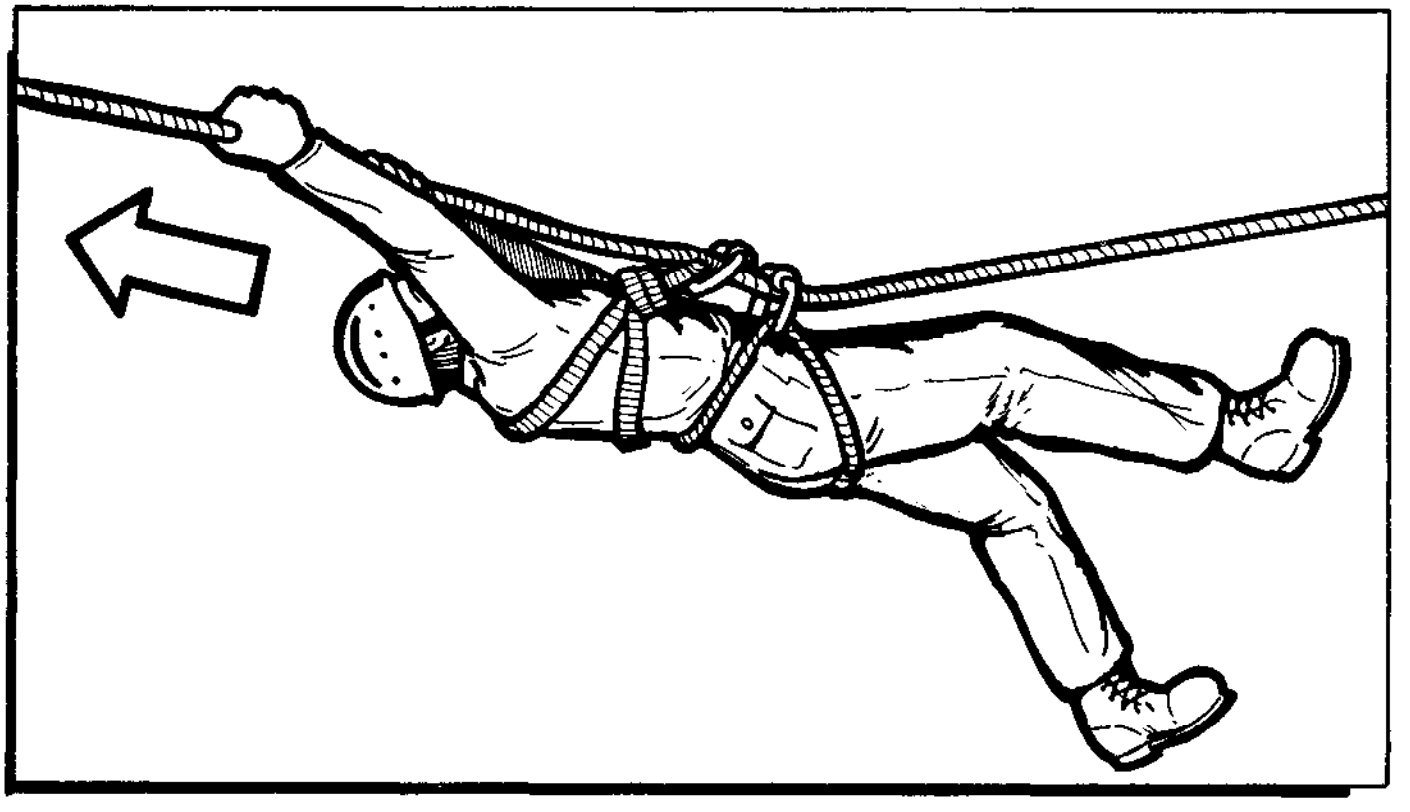

Figure 22-91. Rescuer Traversing.

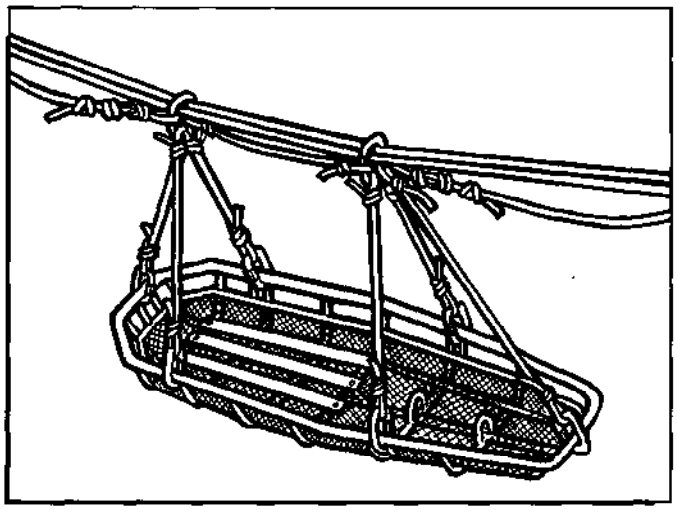

Figure 22-92. Rigging the Horizontal Hauling System.

sets, one is attached to the end of the litter and the other two to the sides of the litter. The slings are about 20 inches long and attached to the inner rail of the Stokes litter with a round turn and two half hitches (figure 22-92). The remaining team then crosses the rope. Each set of slings is attached to a steel "D" locking carabiner. A short length of rope will join the two carabiners for equal pull during the traverse. An in-haul line should be attached to the litter to control the descent or to haul across the open expanse (figure 22-93). To retrieve the rope, the anchors are removed and the rope simply pulled across the obstacle. If an artificial anchor was used for the midrope anchor, it is not retrievable (figure 22-94).

(4) Crevasse or Ravine Recoveries. The ability of a team to recover a victim below them, as in a crevasse, requires the use of a mechanical advantage. A mechanical advantage is the arrangement of ropes and use of pulleys to arrange a system which will allow a force directed on the system to produce a lifting power in multiples of the force. The primary use of a mechanical advantage is to perform vertical lifts as in crevasse or cliff rescue and to apply tension for tightening the system used in construction of hauling lines (figure 22-95). The mechanical advantage can be in different ratios depending on the number of pulleys and changes of direction. Pulleys should be used when a rope passes through a carabiner. If pulleys are not available, a carabiner can be used; however, a larger amount of friction will be present causing a loss of some of the advantage in the system. Rope placement is important in the construction of a mechanical advantage. The ropes should run side by side with no twist around each other. Two basic types are illustrated. The Z-pulley method uses a portion of the load-bearing rope to make the mechanical advantage and, therefore, is a one-rope system (figure 22-96). The second type uses a second rope to establish the mechanical advantage on the load-bearing rope (figure 22-97).

(5) Z-Pulley System. This system of mechanical advantage is used to lift a fallen climber or stranded individual out of a crevasse, up a cliff face, etc. If the fallen climber is attached to a rope, this rope may be used for the system. If the climber is not attached, it will be necessary for a member of the rescue team to rappel or be belayed down to the climber. The rope should be connected to a seat harness and a chest harness. An improvised harness may be required (figure 22-97). Pull all rope slack up from the crevasse. Near the edge, anchor a prusik sling and attach it to the rope. This keeps the rope from slipping back down over the edge.

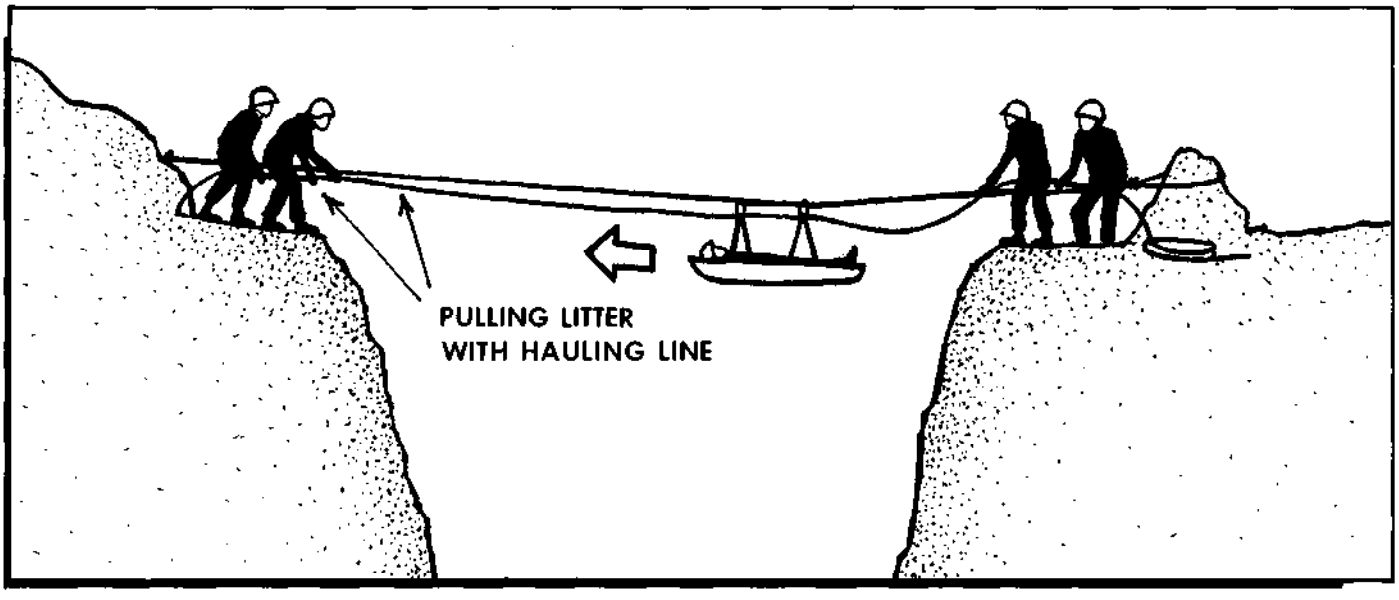

Figure 22-93. Pulling the Litter.

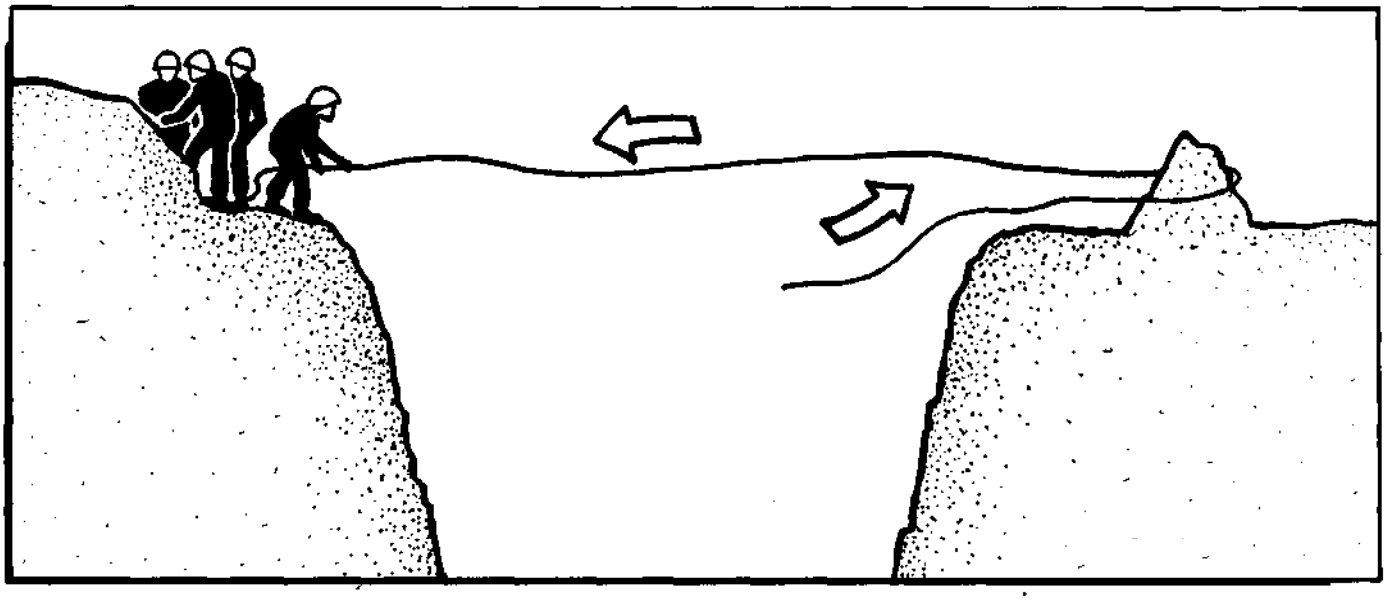

Figure 22-94. Retrieving the Rope.

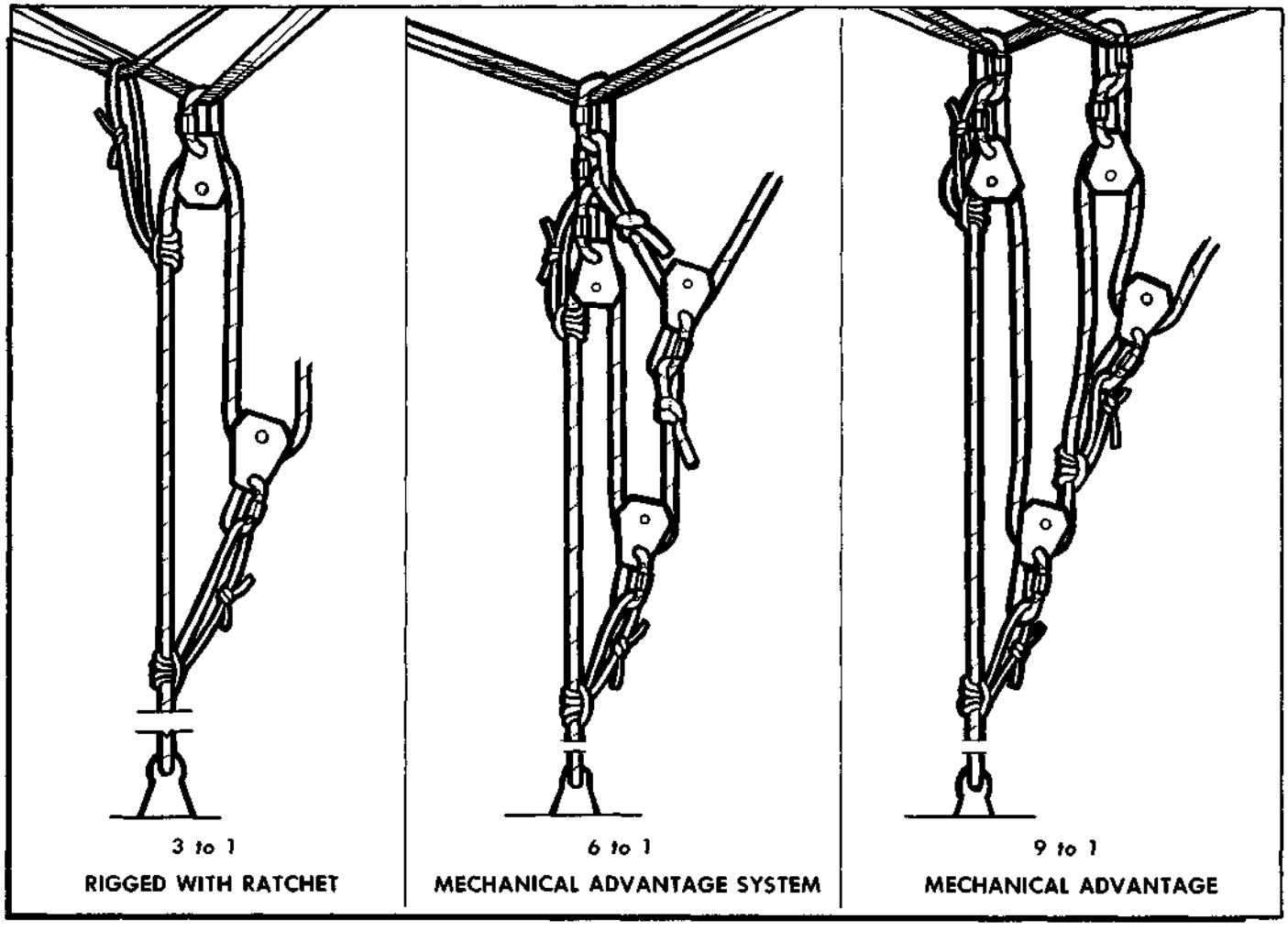

Figure 22-95. Mechanical Advantage Systems.

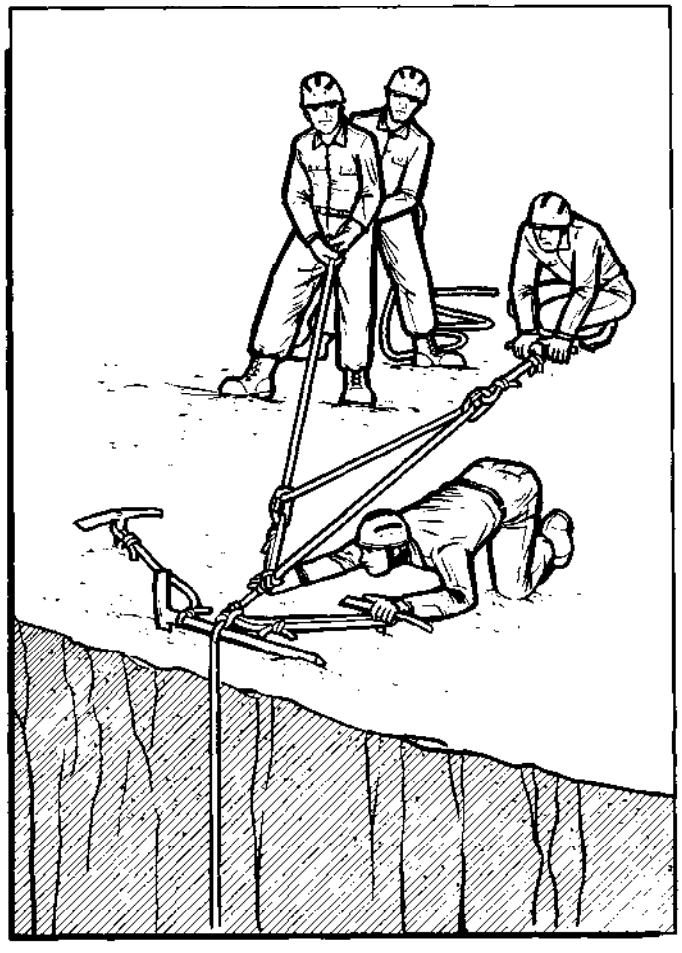

Figure 22-96. Two-Pulley System.

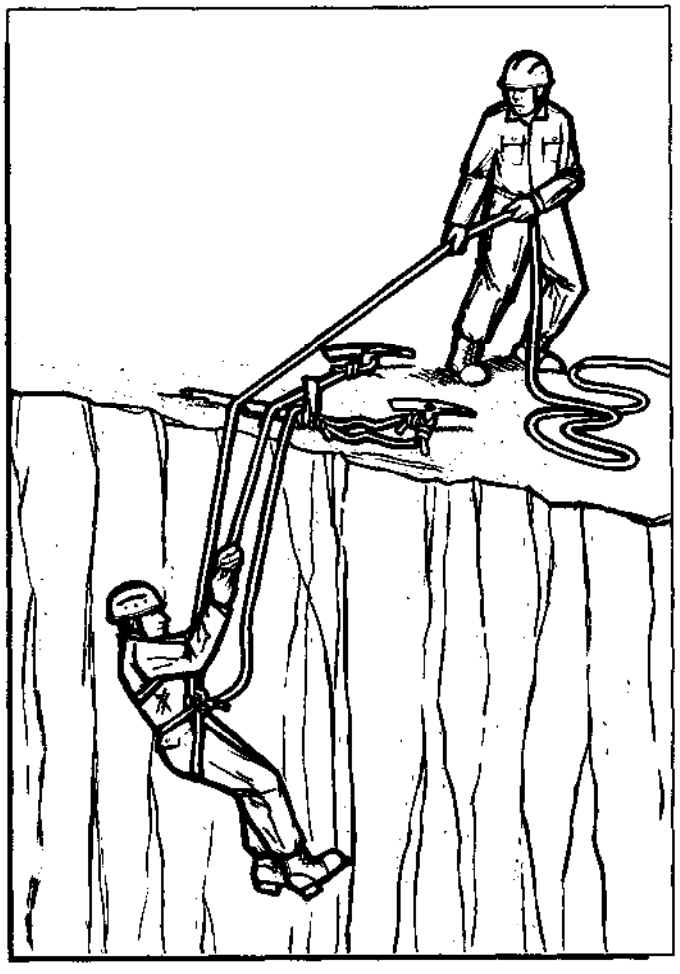

Figure 22-97. Second Rope Added.

An anchor system should be established back from the crevasse and a pulley attached to the anchor. The rope is routed from the crevasse through this pulley. The rope comes out of the pulley and runs back to the crevasse edge. Here another prusik sling is added to the rope as it comes out of the crevasse and a pulley attached to it. The rope from the anchor pulley is passed through this pulley. The system is now complete. The operation is conducted by pulling on the standing end of the rope until the second prusik sling is pulled to the anchor pulley. The prusik at the edge of the crevasse takes the load while the other prusik is reset in the original position. Continue with the lift until the climber is free of the crevasse.

(6) Added Rope Mechanical Advantage. This system requires an additional rope to form the system. As in the Z-pulley system, a rope from the fallen climber is required. Pull all rope slack up from the crevasse. Establish an anchor and connect a prusik sling to the anchor. Secure the belay rope with this prusik sling. Establish two separate anchors and anchor one end of a long sling to the anchor nearest the belay rope. Attach another prusik sling to the belay rope between the anchor prusik and crevasse edge. On the standing end of this sling, attach a carabiner using a figure-eight knot. Attach another carabiner through the body of the figure-eight knot. Place a pulley on each of these carabiners and on the remaining anchor point. Run the long sling through the pulley at the end of the sling, back through the pulley on the second anchor, and through the remaining pulley in the body of the figure-eight knot. The system is ready for operation (figure 22-96). Pull on the long sling end until the prusik with the two carabiners is pulled into the anchor system for the long sling. Allow the anchor prusik to hold the weight and reset the long sling in the initial way. Continue with this procedure until the climber is raised from the crevasse.

(7) Gorge Lift Rescue System. One rescuer rappels to the climber while the other goes around the crevasse or gorge to the other side. An anchor is established on both sides. The end of one rope is thrown across to the far side where it is attached to the anchor. The center of the rope is allowed to drop in the crevasse where the climber is attached to it by a pulley. The end of a second

Figure 22-98. George Lift.

rope is lowered to the climber and anchored to the person. The free end of the first rope is put into the mechanical advantage system. As tension is applied to the rope, the climber is lifted. When the rope is at an approximate 40-degree angle, the climber is pulled toward the edge by the ropes. As the climber nears the edge, the weight of the climber will sharpen the angle of the rope. It will be necessary to apply more tension to the rope. Continue until the climber is rescued (figure 22-98).

(8) Rope/Knot Bypass While on Belay. When using multiple ropes or adding extra ropes to a long descent, it is necessary to pass the joining knots through the belay system. The knots will not pass through the system so a method is used which will secure the rope while the belay device is removed and the knot passed. As the knot approaches the belay system, apply a brake and secure the system. Place a prusik sling on the rope below the belay device. Run this sling to the anchor and tie a mariner's knot (figure 22-99). Take the brake off the belay and allow enough rope through the belay until the

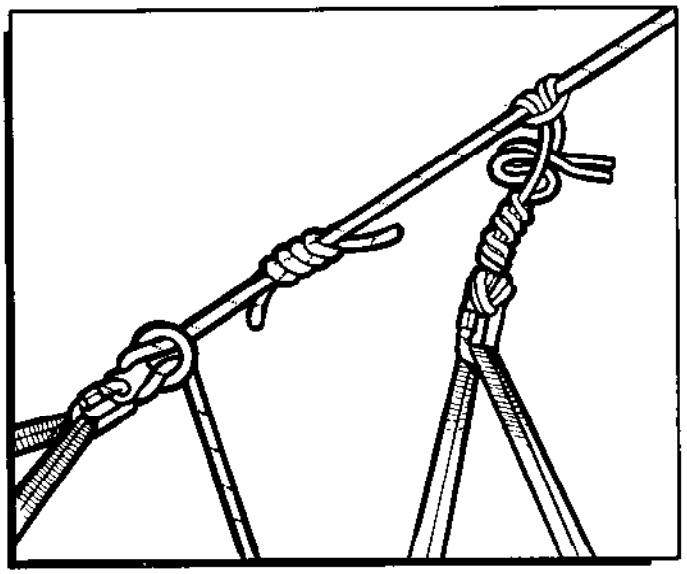

Figure 22-99. Multiple Ropes.

tension is taken by the mariner's knot. Replace the rope through the belay immediately behind the knot and secure the brake. Slowly remove the mariner's knot until the tension is again on the belay system (figure 22-100). Continue with the descent and repeat procedures as necessary.

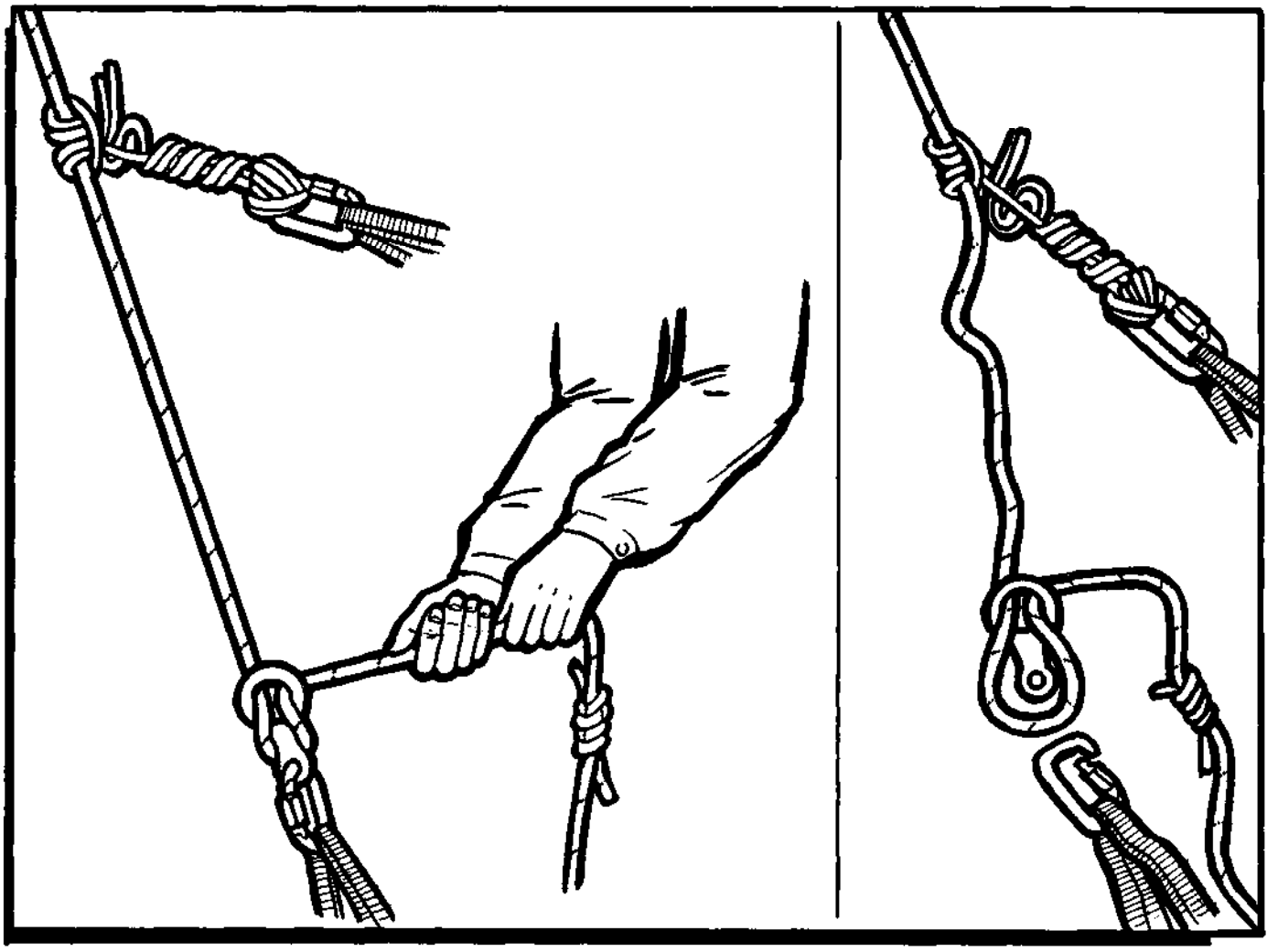

Figure 22-100. Rope Bypass.

Chapter 23

WATER TRAVEL

23-1. Introduction. In this chapter, the travel techniques which can be used on 71 percent of the Earth's surface, the water environment, will be addressed. The techniques of river and open sea travel may be adapted to other water features such as swamps and lakes. In the paragraphs concerning open sea travel, the environmental factors of oceans of the world will be considered as they relate to travel. The techniques for individual water rescue and swimming are important for recovering injured survivors and equipment. Additionally, the problem of submersion must be considered as well as how the antiexposure suit, life preservers, and liferafts can extend a survivor's life expectancy. This is why a survivor must be familiar with the individual rafts and raft procedures. The ultimate goal of a survivor on the open seas is to be rescued; however, a second goal, if rescue is not made, would be to make it to land. A thorough knowledge of water travel techniques will greatly increase the survivor's chances of reaching landfall. If done correctly, landfall can be reached with minimum loss of equipment or injury.

23-2. River Travel:

a. Rivers have been used for centuries in all environments as a safe means of travel and is the reason why most of the cities of the world are located on rivers. It is not uncommon for a river to flow at 4 or 5 knots per hour. A survivor could travel 20 to 25 miles in 5 hours of travel. This may contrast greatly with the rate of travel on land. The amount of energy required to carry survivors' equipment and other supplies and to travel 20 to 25 miles on land is much greater.

b. Each major continent has thousands of miles of navigable rivers. Some rivers such as the Nile, Amazon, Mississippi, Lena, and Mackenzie have hundreds of miles of navigable water with seldom a ripple. These navigable sections are generally found flowing through the flatlands, plains, tundras, and basins of the world. In these areas, only the temperatures of the water and the plant and animal life may present hazards. In contrast, the headwaters of rivers, like the Mackenzie, Yangtze, and Ganges, are so rough that they would best be categorized as a threat to life. This would also be true of the Snake, Salmon, and Rogue Rivers of the Northwestern United States. These rivers, although traveled by white water rafters, pose an unreasonable hazard to survivors. Survivors must take into account individual or group skills, injuries, type and severity of rapids, the temperature of the water, and direction the river flows in making the decision to travel. Even if a portage of several miles is required, the energy saved by floating on a river might warrant river travel. However, once the energy expended for portage exceeds the energy conserved by floating a section of a river, the river as a mode of travel should be abandoned.

c. In a nontactical situation, rivers will most likely carry survivors to indigenous people who could aid the survivor in meeting the basic needs for sustaining life and effecting rescue. Even if some form of civilization is not encountered, the survivors would most likely reach a lakeshore or even the seacoast. These environments, particularly the seacoast, provide transition zones from land to water which are rich in food and other survival resources. In these areas, the resources could improve the chances of survival. It is much easier to spot signs of survivors along a shore versus the interior of a landmass.

23-3. Using Safe Judgment and Rules for River Travel:

a. There are certain safety rules and guidelines that must be followed to reduce the dangers associated with river travel. Respect for these rules and guidelines is necessary to reduce the potential dangers.

b. The most important safety rule is personal preparation. Preparation should begin by thoroughly scouting the river. The conditions of the river will determine the intermittent stops. High ridges along river edges provide needed visibility to plan each leg of travel. If there are numerous bends and poor lookout points to view the river, stops are frequent. Sound judgment must be used when planning routes. Patience in planning each leg of travel helps prevent disaster. All survivors must prevents know the plans and be able to handle the route safely, considering their skills and strength. Survivors should be aware of and avoid river hazards and have alternate routes and communication signals in case flow conditions suddenly change, making the run more difficult. All rapids which cannot be seen clearly from the river should be scouted. The route should be discussed with the crew. The skills, knowledge, and abilities of the aircrew members must be considered, including swimming abilities and physical condition. Areas of high risk should not be attempted. Before reaching an area of suspected great difficulty, rafts should be beached and carried to the next point of travel; this is called portaging.

c. Before entering the raft, survivors should don life preservers and suitable clothing for adequate protection. The equipment should be tested to ensure it is serviceable. Bulkiness is not advisable due to the possibility of the raft capsizing and being weighted down with water. The antiexposure suit (if available) should be worn. Items that might absorb water should be packed in a waterproof container. The survivor should ensure that:

(1) All medical supplies, repair kit, and survival kit are in the raft.

(2) Survival kit is checked and inventoried.

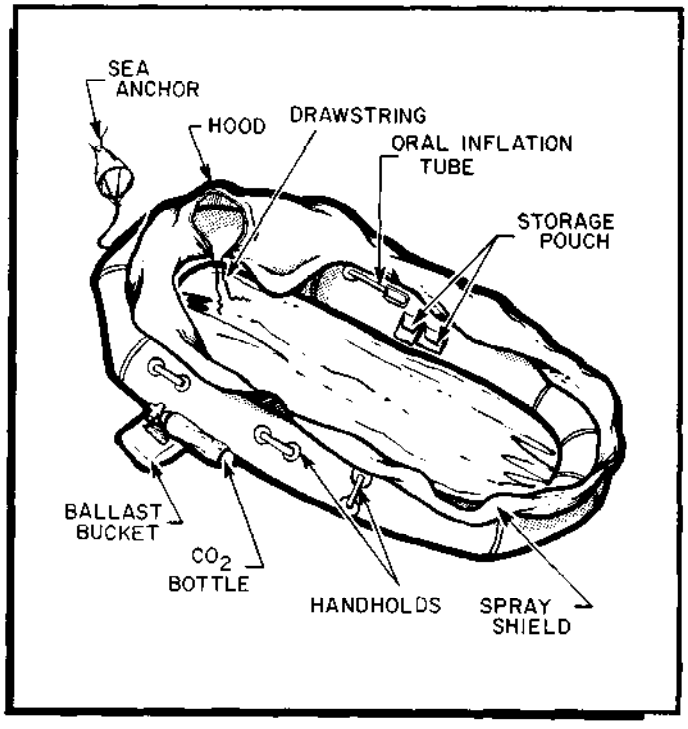

Figure 23-1. One-Man Raft.

(3) Extra efforts should be made to keep supplies and equipment in good condition.

(4) All items are secured to the raft to prevent loss and (or) injuries.

(5) Before use, the raft is checked for leaks and necessary repairs are made.

d. When using a one-man raft for river travel (figure 23-1), it may be advisable to tie or cut off the ballast bucket, fasten the spray shield in the opened position, and remove the sea anchor to prevent problems with swamping or entanglement with subsurface obstacles. Without the ballast bucket, the raft can be easily maneuvered by paddling with the backstroke or for slight adjustment, with a front stroke. When using either the backstroke or the front stroke, the survivor will find it easier if the two underarm cells of the life preserver underarms (LPU) are disconnected in front and the cells placed behind the back (figure 23-2A). This gives the survivor a full range of motion. When rough water is encountered, the survivors should fasten the LPU and face downstream.

e. One of the primary methods of avoiding hazards on the river is to slow the speed of the raft and move across the river, avoiding a collision with the obstacle. This ferry position should be initiated early to avoid large rocks and reversals in the river. If the collision obstacles are to be avoided, point the bow of the raft toward them and backstroke against the current to slow the speed of the raft's downstream progress and move it across the river. Usually, the best angle is about 45 degrees to the current. The greater the angle, the quicker the movement across the river, but this also increases the downstream speed of the raft. Decreasing the angle will slow downstream speed, but movement across the width of the river will also be decreased (figure 23-2B). The raft will be more maneuverable if it is well inflated. If the raft should pass over a rock, arch the back up to prevent injury to the buttocks or back.

f. When using multiplace rafts, the boarding ladder and sea anchor should be removed to prevent entanglement. If available, about 50 feet of line should be tied on the bow and stern of the raft to be used for tie-offs. An additional 200 feet of line should be coiled and tucked away for emergency and rescue work; one end is secured to the raft while the other end has a fixed loop. An improvised suspension line rope may be used for this; for example, three-strand braid can support about 1,000 to 1,500 pounds of pressure, or a two-strand twist strengthens the line to support about 700 to 900 pounds of pressure (figure 23-3).

g. Proper placement of equipment and personnel should equalize weight distribution to ensure stable control; overloading should be avoided. Assign personnel crew positions and responsibilities in the raft; captain (person in charge), stern paddler (maneuvers raft), and side paddlers. Twilight and night rafting should be avoided (nontactical) as poor visibility increases danger.

h. Two ways to steer a multiplace raft are:

(1) To steer a raft by using sweeps (long oar) and poles. A pole is more efficient in fairly shallow water, but a sweep is preferable in deep water. Poles and sweeps from both ends of the raft are used. The person in the bow (front) can see any obstructions ahead, and the one in the stern (rear) can follow directions for steering. Poles are also useful for pushing a raft in quiet water.

(2) Paddle techniques are used to maneuver the raft. When paddling, there are three possible body positions on a raft. The best way is to sit on the upper buoyancy tube with both legs angled to the inside of the raft. The body should be perpendicular to the sides of the raft, enabling the rafter to paddle. Another way is to sit cowboy style, straddling the upper bouyancy tube of the raft with one leg on either side, and folding at the knee with each leg back. However, the outside knee may collide with obstacles and cause injury. The third way is normally used in calmer waters because it consists of partially straddling the upper tube, with legs comfortably extended. In a smaller raft, the survivor may be able to sit down inside the raft and reach over the buoyancy tube. The following strokes can be done from these positions. Knowing the parts of paddles will help in explaining the different paddling strokes (figure 23-4).

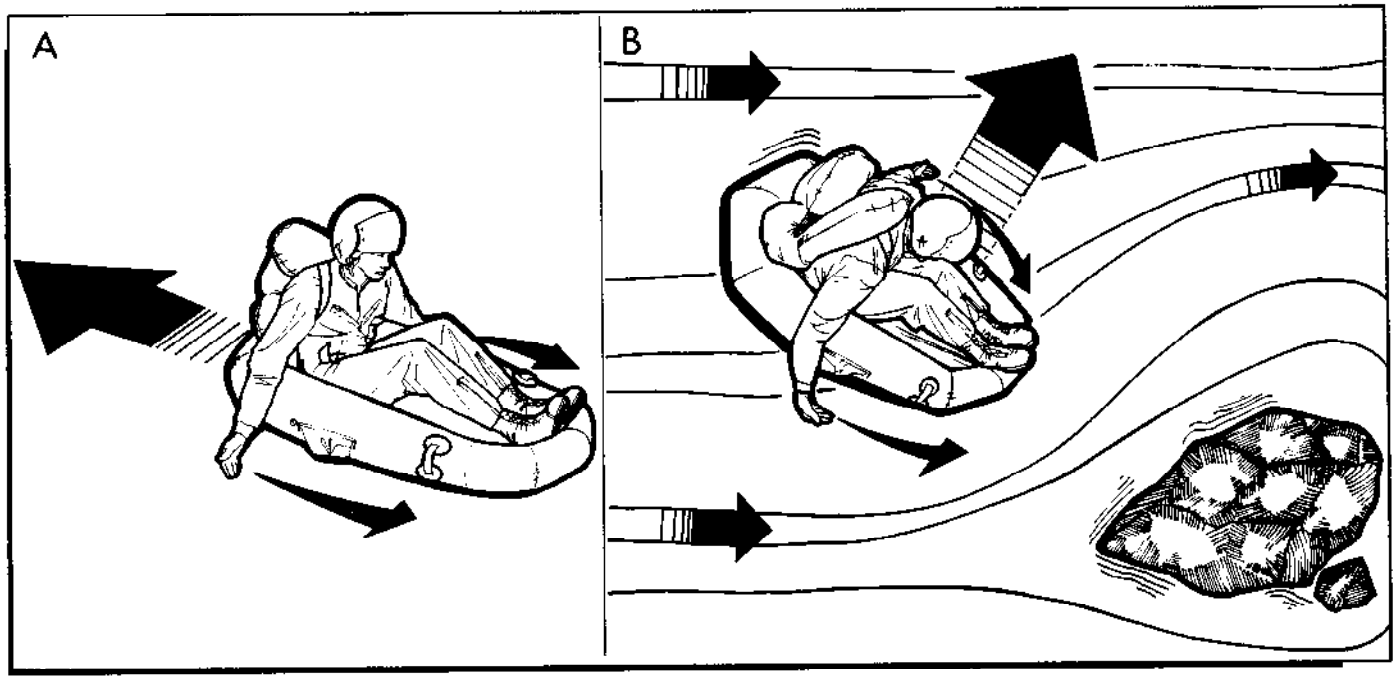

Figure 23-2. One-Man Raft (Paddling).

(a) One of the easiest is the forward stroke which is done in smooth continuous movements using these techniques:

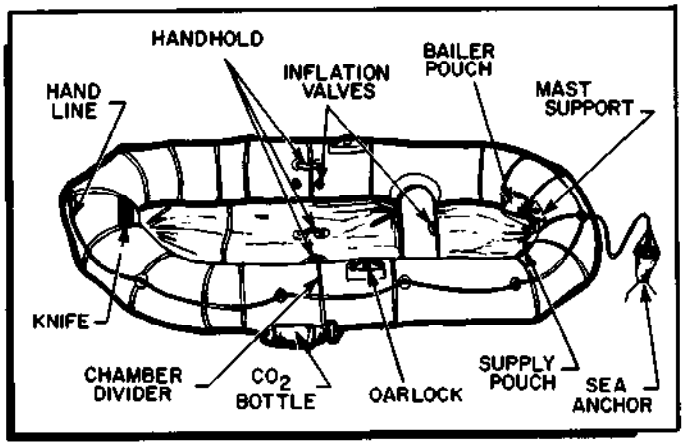

Figure 23-3. Seven-Man Raft.

-1. Thrust the blade of the paddle forward using the outboard arm, then momentarily keeping the outboard arm stiff and away from the raft, push the grip. The inboard hand is then moved forward to cut the blade deeply into the water. Continue the stroke by pushing on the grip and pulling on the shaft keeping the blade at a 90-degree angle to the raft. Stop the motion as the blade comes slightly past the hip, because a full followthrough provides little forward power and wastes valuable energy. Slide the blade out of the water by pushing down on the grip and swinging it toward the inboard hip and turning the blade at a parallel angle to the water once it has cleared the water. By paralleling

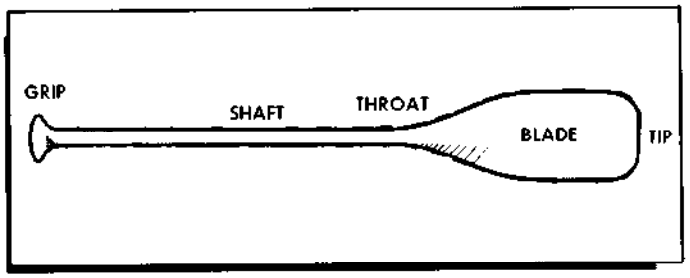

Figure 23-4. Paddle.

the blade, it cuts wind and wave resistance and saves time and energy. This cycle is repeated until the strokes are changed.

-2. In mild water, there is no need to over reach or excessively twist the upper trunk of the body. When extra speed is needed, lean deeply into the strokes which brings the entire body into play. Position the inboard hand across the tip of the paddle grip and the outboard hand halfway to three-fourths of the way down the shaft.

(b) The opposite of the forward stroke is the backward stroke. The blade is thrust into the water just behind the hip, and pressure applied by simultaneously pushing forward on the shaft and pulling back on the grip. End the stroke where the forward stroke would

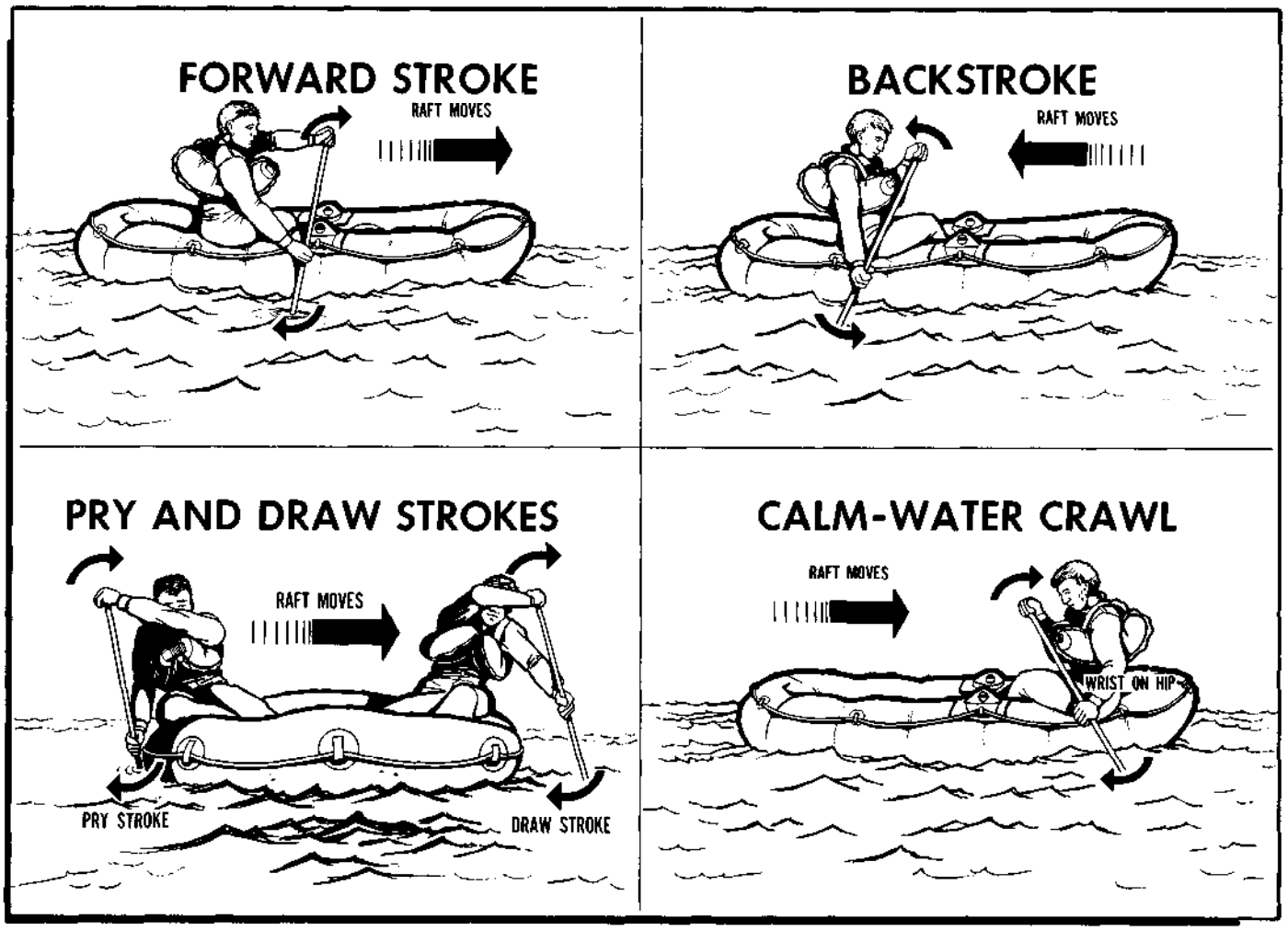

Figure 23-5. The Paddle Strokes.

begin, and again angle the blade out of the water back to the beginning of the backward stroke.

(c) The draw and pry strokes are opposite sideways strokes. These strokes are good for small sideways maneuvers and for turning the raft when used from the front or rear of the raft.

-1. Drawstroke. Reach out from the raft, dip the blade in parallel to the raft, and pull on the shaft while pushing on the grip. Pull the blade flat to the side of the raft. Pull the paddle out and repeat.

-2. Pry stroke. Dip the blade in close to the raft, and push out on the shaft while pulling in on the grip.

(d) The fifth stroke, called the calm water crawl, is used alternately with the forward stroke when paddling through long calms. Sit cowboy fashion while facing the stern and hold the paddle diagonally in front with the shaft which is held by the outboard hand against the outboard hip and the grip held by the inboard hand in front of the inboard shoulder. Extend the inboard arm to swing the blade behind, dip the blade in the water, and pull back on the grip, prying it forcefully, using the hip as a fulcrum (the point of support on which a lever works). Using the shoulder, hip, and hand as assisters, the crawl is easy yet powerful (figure 23-5).

(e) The ferry is a basic paddle maneuvering technique used to navigate bends and to sidestep obstacles in swift currents. The ferry is essentially paddling upstream at an angle to move the raft sideways in the current. Paddle rafts can ferry either with the bow (front) angled upstream or downstream. The bow-upstream ferry is stronger because it uses the more powerful and easier forward stroke. It is carried out by placing the raft at a 45-degree angle to the current with the bow angled upstream and the side toward the desired direction. The bow-downstream ferry is weaker because it uses the less powerful backstroke, but it does offer certain advantages. It enables paddlers to look ahead without straining their necks, and makes it easy to put the bow into waves (figures 23-6 and 23-7). It is carried out

by backstroking with the stern (back) angled upstream at a 45-degree angle and the side facing the desired direction.

(f) There may be times when the only way for a heavy raft to enter a small or violent eddy is with a reverse ferry (figure 23-8). The following steps may be used for an oar or paddle raft, except the paddle raft approaches the eddy bow first and finishes in a bow-upstream position:

-1. Raft approaches sideways.

-2. Raft turns around to angle its bow downstream.

-3. With careful timing, the captain should have the crew begin to pull powerfully on the paddle. The angle of the raft to the current can be close to 90 degrees, but is best at about 45 degrees.

-4. While aiming for the eddy, the crew should continue with the front stroke and gain momentum.

-5. With the crew still using the front stroke, the raft breaks through the eddy fence.

-6. With the bow in the upstream eddy current and the stern still in the downstream current, the raft is spun into a normal ferry angle. The crew continues with the front stroke while making the necessary turn to bring the boat entirely into the eddy.

-7. The raft rides easily in the eddy. (NOTE: The reverse ferry and eddy turns are not only used to enter eddies, but can also be used to dodge through tight places. The reverse ferry (or sometimes an extreme ferry) scoots the raft sideways, the eddy turn snaps the bow into a bow-downstream position, and the raft, rather than entering the eddy, rides the eddy fence past a major obstruction or hole.)

(g) The straight forward paddle is used in calm and moderate waters where there is ample maneuvering time. Simply point the bow in the desired direction and follow the forward stroke method of paddling. The back paddle is performed the exact opposite of the forward paddle. Point the stern (back) in the desired direction and follow the backstroke method of paddling (figure 23-5).

(h) To make a left turn, the left side of the raft will backpaddle, while the right side paddles forward. It's just the opposite to make the raft turn right. The right side on the raft backpaddles while the left side paddles forward—both performing the paddling maneuvers at the same time (figure 23-9).

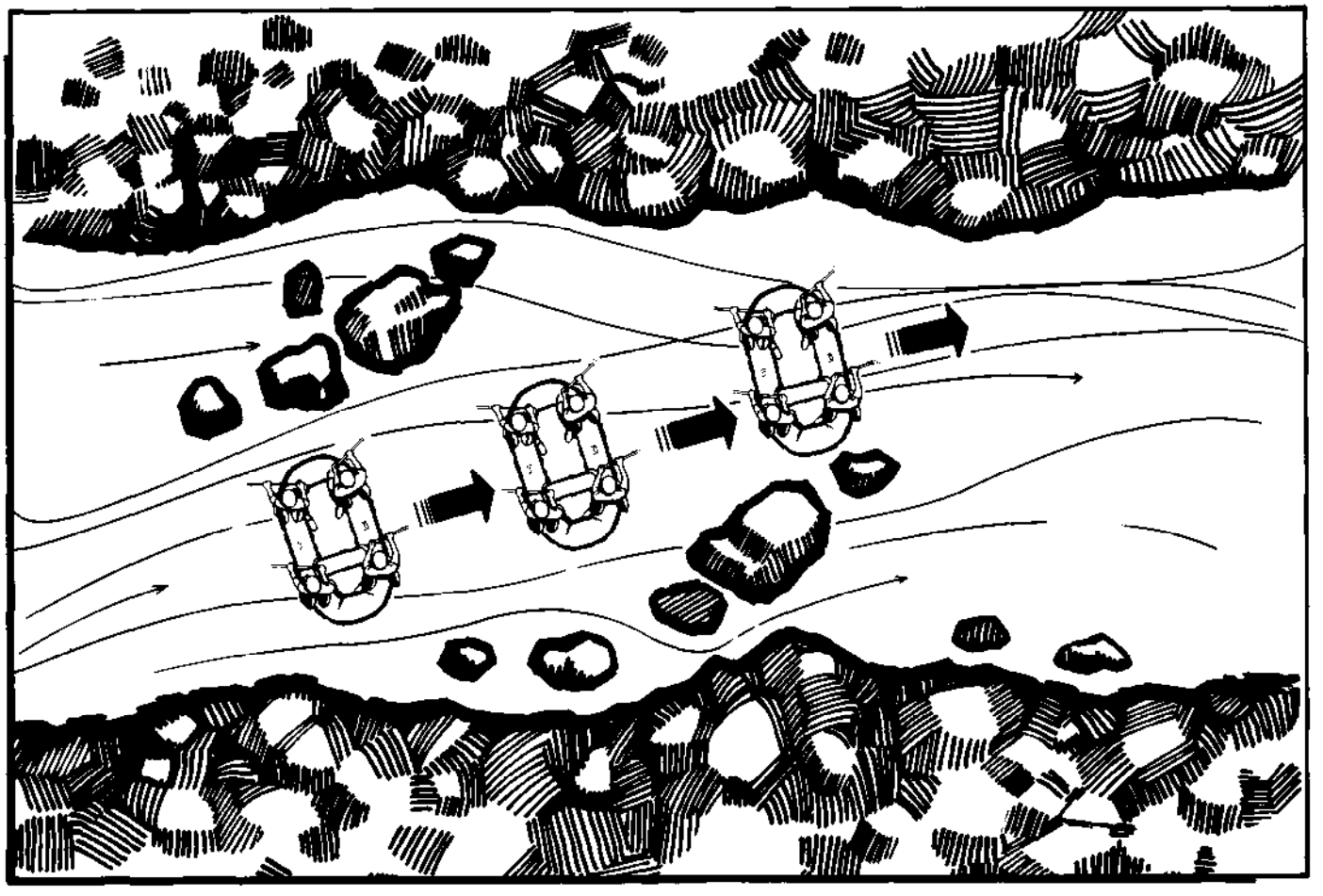

Figure 23-6. The Paddle Ferry.

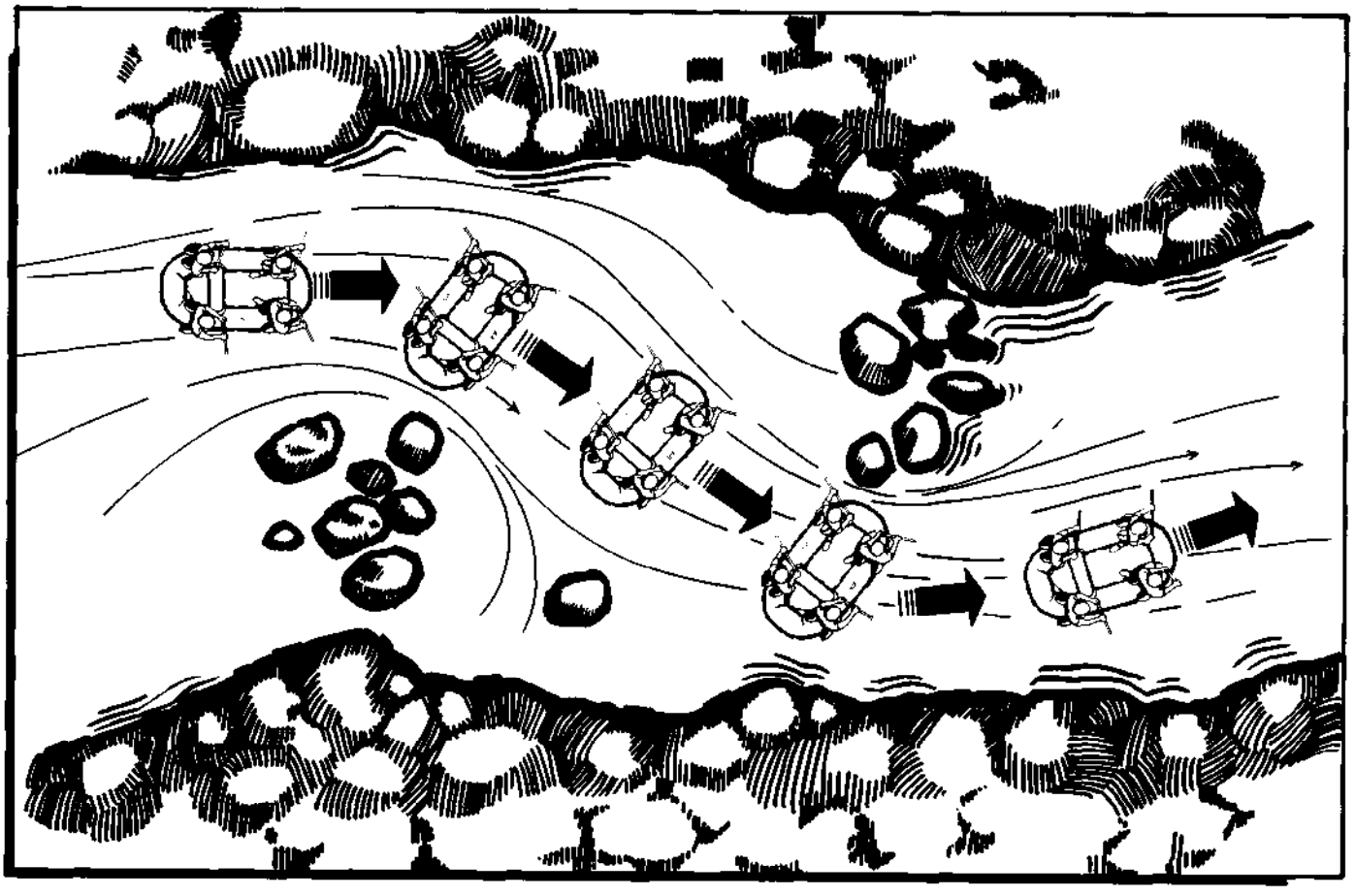

Figure 23-7. Bow Downstream Ferry.

(i) The pry and draw strokes are used to move the raft sideways (figure 23-10).

(j) Stern maneuvers are used to increase the maneuverability of the raft. The paddle used at the stern of the boat is basically a rudder which controls direction. To turn right, the paddle blade is held to the right, square against the direction of the current. To turn left, the paddle is held to the left.

(k) Other strokes, such as a forward stroke or draw stroke, used at the stern of the raft, will cause it to turn or move faster. If stroking is done slightly to the side (either right or left) of the raft, it will help move the raft in the opposite direction (figure 23-11).

(l) River travel requires fast, decisive action. Therefore, a paddle raft needs a captain to coordinate the crew's actions by the use of commands or signals. Communications between captain and crew are crucial; all members must agree on a set of short, clear commands. The following are suggested commands:

CAPTAIN'S COMMANDS:	CREW RESPONSE:
Forward	Crew paddles forward.
Backpaddle	Crew does backstroke.
Turn right	Left side paddles forward, right side does the backstroke.
Turn left	Right side paddles forward, left side does the backstroke.
Draw right	Right side uses draw stroke, left side uses pry stroke.
Draw left	Left side uses draw stroke, right side uses pry stroke.
Stop	Paddlers relax.

-1. Commands must be carried out immediately, so the crews should practice until they can snap through all the commands without hesitation. The captain controls both the direction and speed of the raft with a specific tone of voice and commands. This control, and the captain's ability to anticipate how the water ahead will affect the raft, will help avoid undercompensation and overcompensation of maneuvers through obstacles. Good captains think well ahead and

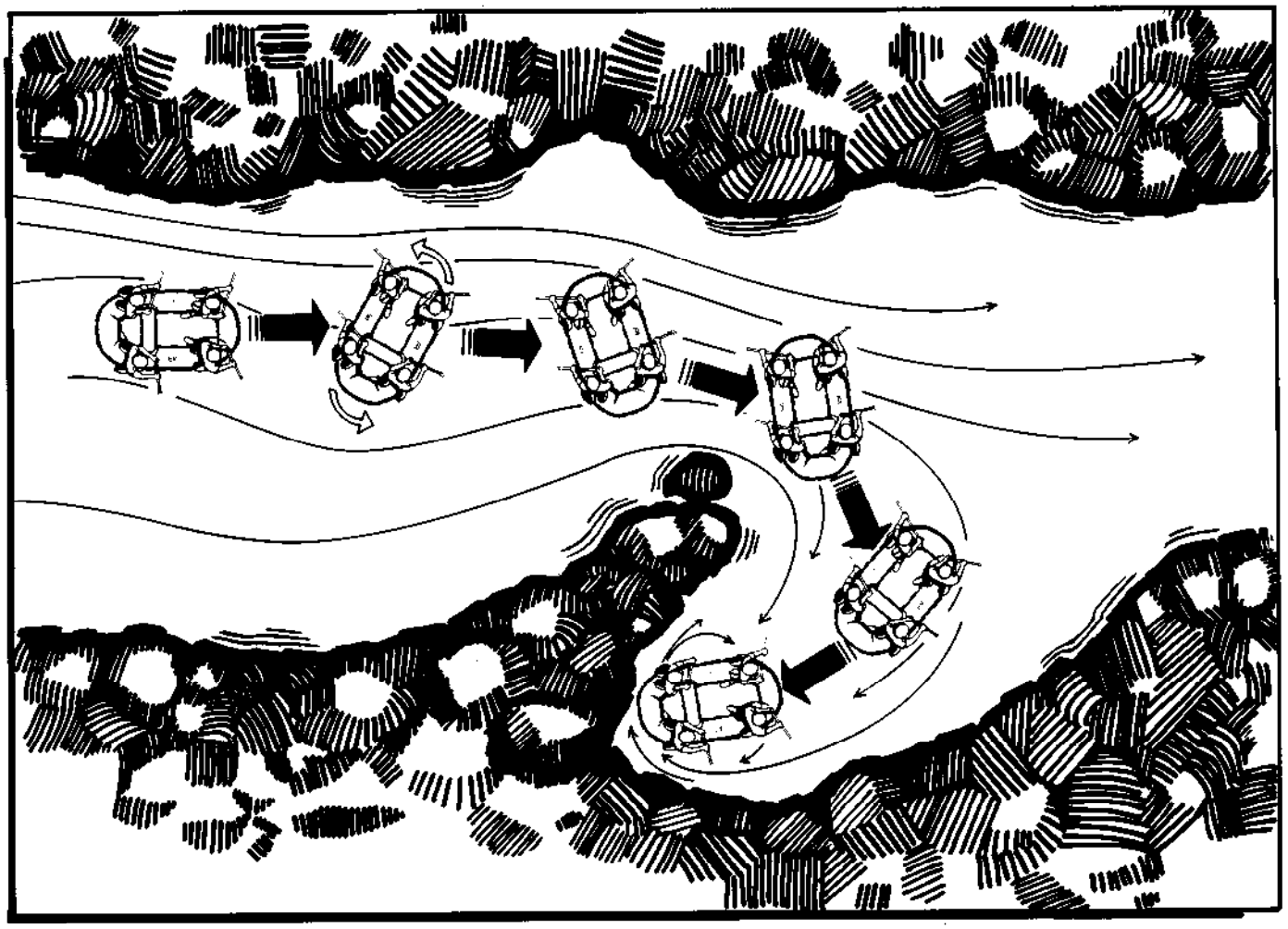

Figure 23-8. Entering an Eddy with a Reverse Ferry.

move with the river, issuing commands precisely and sparingly, working their crews as little as possible.

-2. When using commands and maneuvering the boat in harmony with the river's currents, paddling can be easy and effective, even fun. When instant action is necessary, the captain may say, "paddle at will." When time permits, the captain should introduce commands with a preparatory statement such as, "we're going to ferry to the right of that big rock. OK—(gives command)." This gives the crew time to prepare for the next response. If a command is not heard or understood, it should be repeated with zest until it is understood. If a raft member spots a better way through a rapid or channel, a fully extended arm is used to point it out. This signal, like the others agreed upon, should be repeated until it is understood.

23-4. River Hydraulics. An understanding of river hydraulics is important to the survivor. A knowledge of the types of obstacles and why they should be avoided or overcome is necessary for a safe river journey.

a. Laminar Flow. The drag produced when moving water flows over or past various types of objects and surfaces is called a laminar flow (figure 23-12). The laminar flow principle is that various layers or channels of water move at different speeds. The lower layer of the river moves more slowly than the top layer. This is due to the friction on the bottom and sides of the river which is caused by soil, vegetation, or contours of the riverbank. The layers next to the bottom and sides are the slowest; each subsequent layer will increase in speed. The top layer of the river is only affected by the air. The fastest part of the flow on smooth straight stretches of water will be between 5 and 15 percent of the river depth below the surface. Even straight running riverbeds are not smooth; they have jutting and receding banks on the sides which affect the laminar flow. The friction caused by the banks causes the sides of the flow to be slower than the midstream. The areas near the banks are more shallow and have fewer layers. When the river travels at 4 to 5 knots, turbulence begins to develop which interferes with the regular flow of the current. When this rate of riverflow is achieved, the

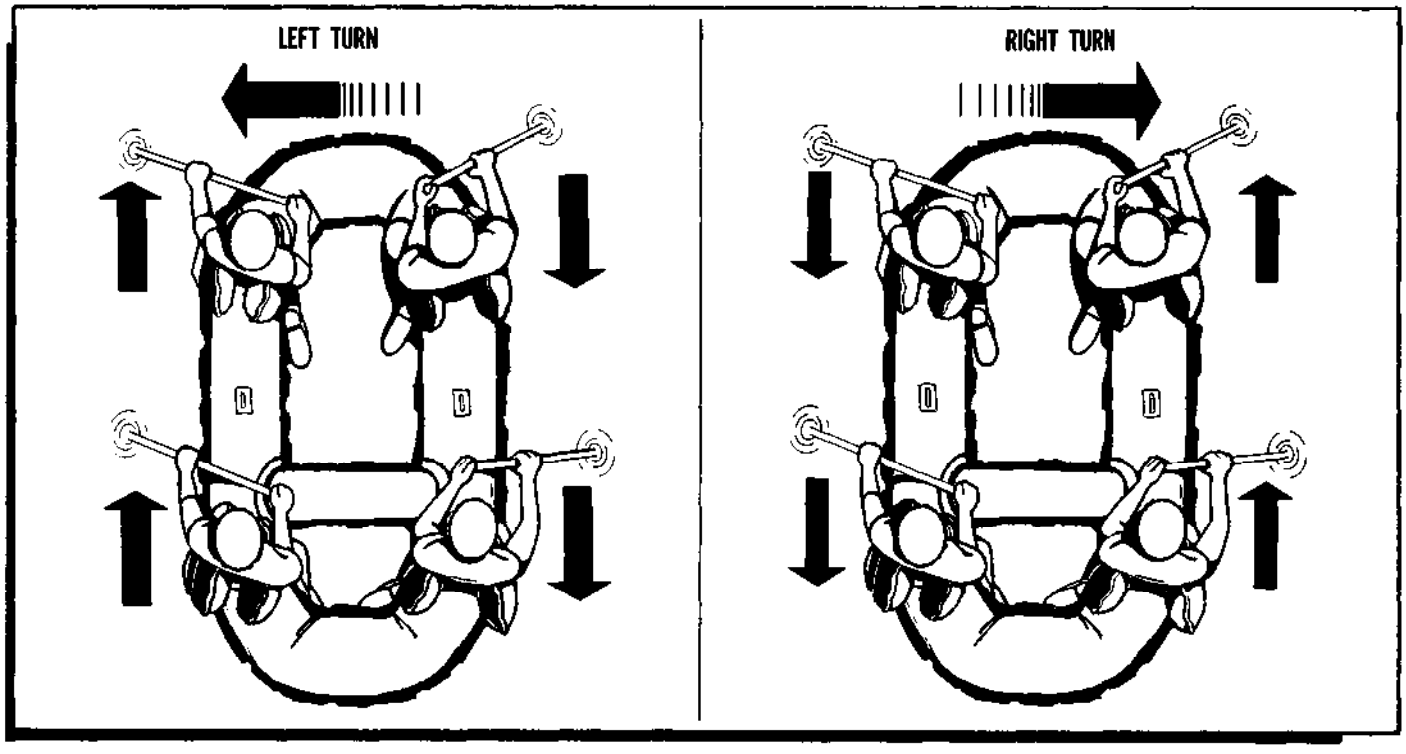

Figure 23-9. Turning the Raft to the Right or Left.

friction between the layers of water will cause whirling and spinning actions which agitates the smooth flow of water, creating more resistance.

b. River Currents:

(1) When a current of a river is deflected by obstructions, the overall downward flow of a river will

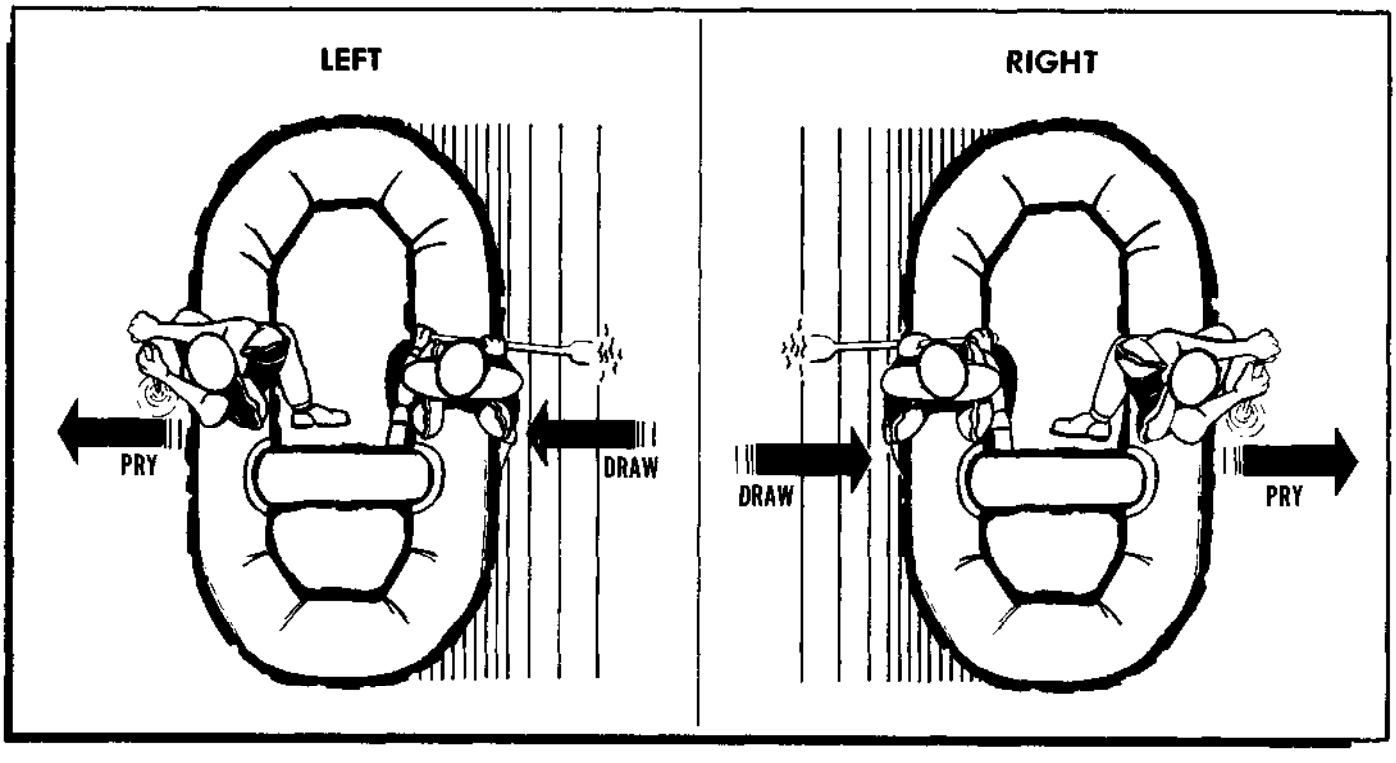

Figure 23-10. Making the Raft Move Left or Right with Pry and Draw Stroke.

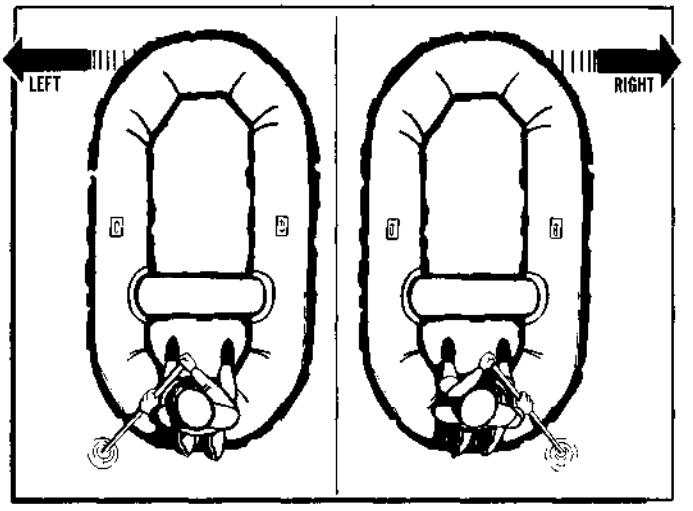

Figure 23-11. Stern Maneuvers.

respond. These responses vary from mild to radical deflection, creating direction and speed changes of waterflow. These changes are called reflex current. The reflex current responds to an obstruction such as bends or submerged rocks.

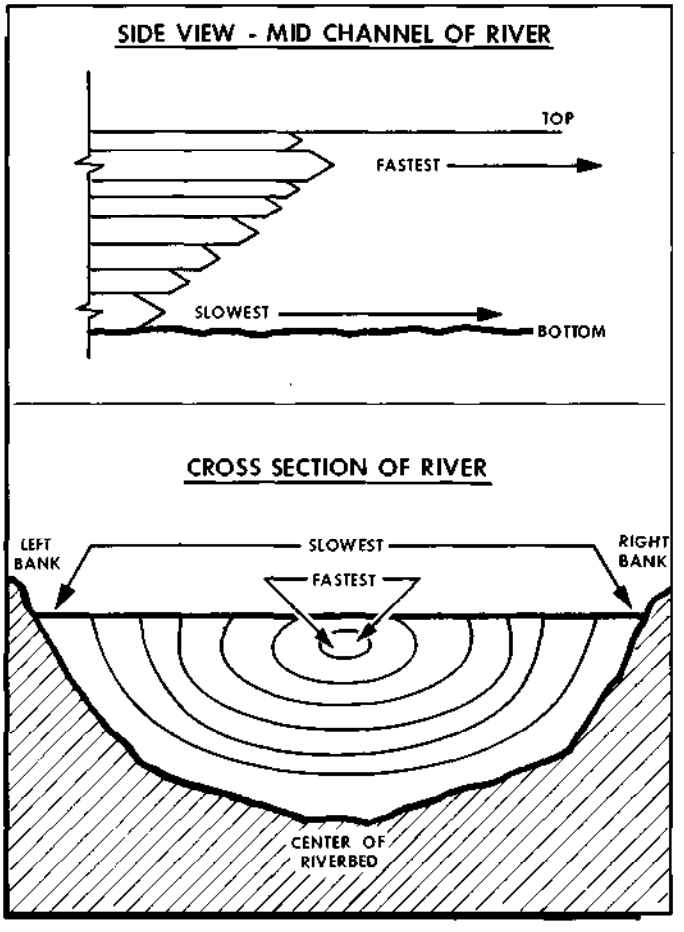

Figure 23-12. Laminar Flow.

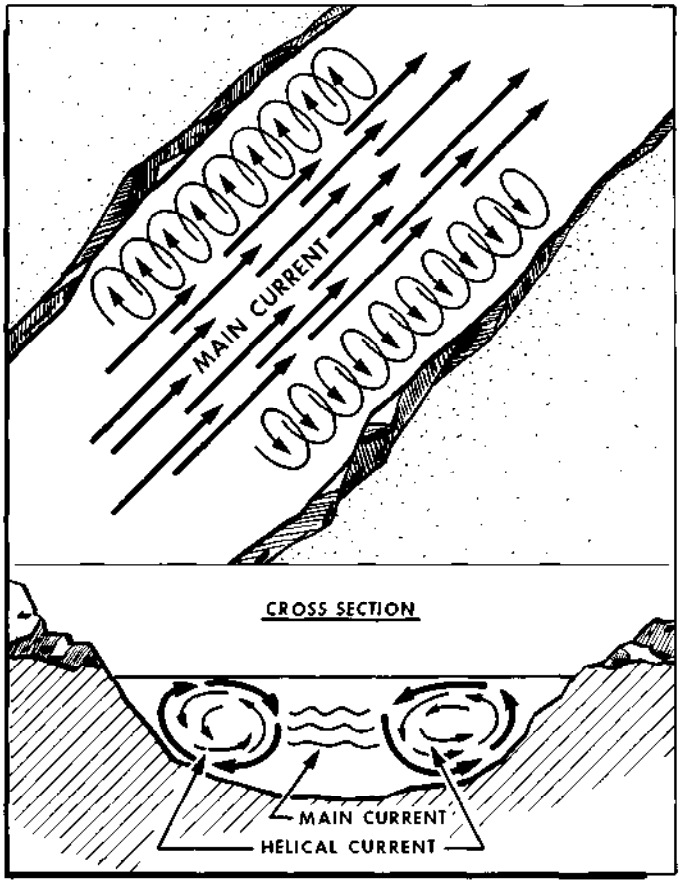

Figure 23-13. Helical Flow.

(2) One response to the laminar flow is a spiraling, coil-springing flow called a helical current (figure 23-13) which corkscrews as a result of the friction with the riverbank. Going downriver, on the left side of the supposed straight-line river, the helical flow turns clockwise to the main current, and on the right side the helical flow is counterclockwise. This results from friction and drag caused by shallowing banks combined with the strong force of the main current flowing down. The helical flow and the mainstream create a circular, whirling secondary current which travels down along a line near the point of maximum flow. Helical current flow starts along the bottom of the river going out toward the riverbank, surfacing, and then spiraling back into the mainstream at a downward angle. This flow causes floating objects around the edges to be pulled into the mainstream and held there. By understanding where the fast water is and how to observe the characteristics which show the current, a survivor can maneuver the raft to take advantage of the faster water to increase the rate of travel. Even at the quietest edge of a flow, particles are still drawn into the strongest part of the current. Laminar and helical flows are always present in fast-flowing rivers.

(3) The main channel is the deepest part of the river and can wander from bank to bank. The turbulence

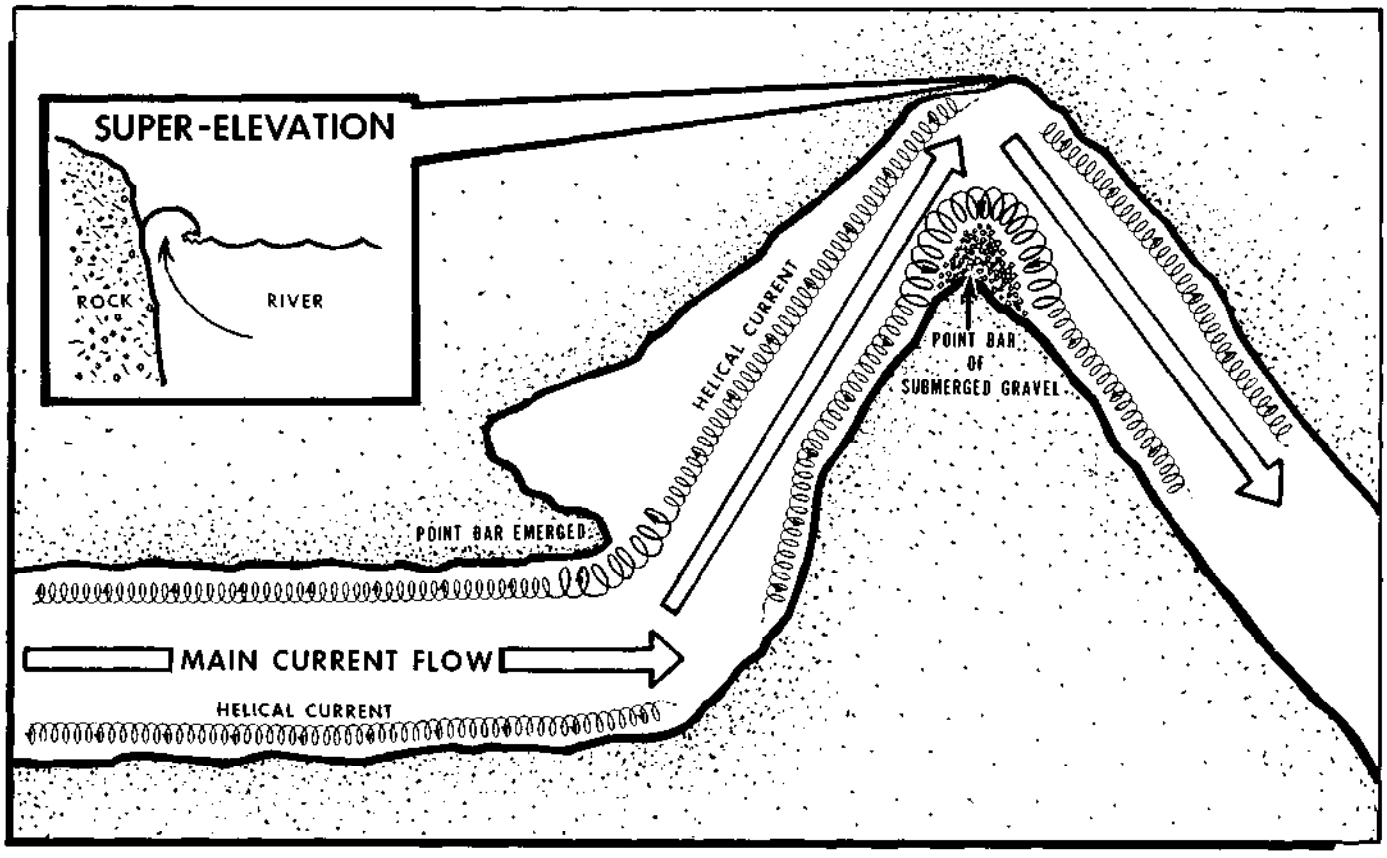

Figure 23-14. Macroturbulence.

caused by the wandering main current erodes wide curves into sharper, more defined bends, creating indirect courses.

(4) When the river makes sharp turns, the current is affected by centrifugal force swinging it wide into the outside bank. The helical current diminishes, being smothered by the laminal flow, thereby increasing the corkscrew effect on the inside of the curve. The surface water is being whirled hard in the direction of the outside curve of the bank; the faster the waterflow, the stronger the push. Floating objects are forced, with the surface water, to the outside of the curve and into the banks, usually getting lodged against and onto the shore.

(a) A powerful helical flow not only pushes the surface outward, but as it swirls up from the bottom it carries sediment up with it. The sediment and other debris is deposited at the highest point of the inside bank of the bend. The sediment is then dropped during high water, and when the water recedes, a point bar (made of sand and gravel) is revealed. The point bar generally sticks out far enough to funnel floating objects into the swiftest part of the river during high waters, avoiding the sandbar.

(b) Super-elevation is a feature where the water is being increased in volume, intensity, and height. When both the stream volume and movement are high, centrifugal force exerts another type of influence on flow characteristics. The river surface water tends to curve in a dish shape towards the outer bend, like a banked turn on a racetrack. The dished inside curve is the easiest and safest route to travel through. If maneuvered correctly, the slight rise of the water and the force of the current around the curve will cause the raft to slip gently off the wave and into the quiet pools of water below. But if the raft was maneuvered across the line of the currents, the raft may either be sucked under by the dominant helical flow, or the power and force of the river on the outside of the dish on the curve could smash and pin any floating device against the outside bank.

(5) Macroturbulence is any extreme, unpredictable turbulence (figure 23-14). It is an especially dangerous phenomenon caused by a drop or decline in the river bottom. The phenomenon also occurs when the water comes in contact with the river bend or rocks. The extreme amounts of froth created from the turbulence and the gravitational pull cause rafts to spin. Raft control is extremely limited because the lack of water viscosity causes resistance against paddles and a lack of buoyancy, making it difficult to float or maneuver. This type of

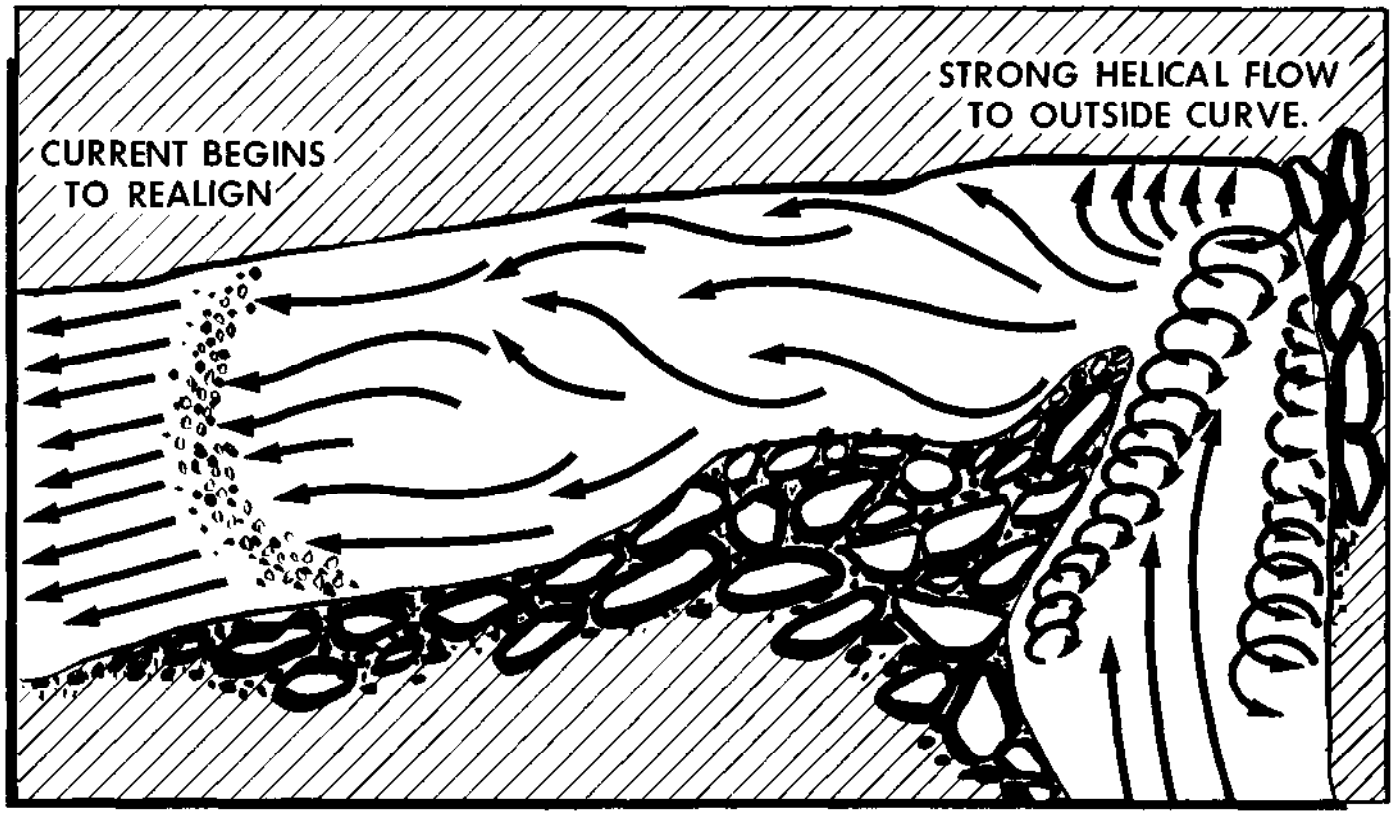

Figure 23-15. Anatomy of a River Bend.

white water can be impassable, depending upon how extreme the dip and amount of waterflow.

(6) Coming out of a sharp bend, the river currents are mixed, but the dangerous movement is still pulling. The result of the laminar flow shooting into a bank creates a helical flow effect immediately below the turn, where the river is still trying to assume a natural "straight" flow. Being a liquid, water cannot resist stress and it responds to a variety of obstacles (most common are submerged boulders) (figure 23-15).

(7) When water flows over obstructions, such as submerged boulders, the character of the laminar flow is changed. As the water flows over the top of the rock, the layers of the laminar flow increase in speed. This is known as a venturi effect. The hydraulic area is a type of "vacuum" formed as water flows around the rock. Created directly below the obstruction are confused and disordered currents which accelerate the layers of the laminar flow (figure 23-16).

(8) One type of hydraulic is the surge, which usually occurs when the current is slow and the water is deep. This hydraulic is formed downstream from an obstruction with a surge in the water volume. When obstacles no longer have the ability to hold the water back, the pressure is released. Surges present few problems if the boulder or obstruction is covered with enough waterflow to prevent contact when floating over the top. Survivors should be aware of obstructions (known as "sleepers") if the water does not sufficiently cover them (figure 23-17). Failing to recognize a sleeper can result in raft destruction and severe bodily injuries. With large sleepers, the water flows over the top creating a powerful current. This powerful, secondary current is trying to fill the vacuum created by the hydraulic downstream action.

(9) Another form of large sleeper is referred to as a dribbling fall. These are caused by minimal water flowing over submerged obstacles with considerable drop below. This type of sleeper causes a bumpy ride, reducing speed, and can capsize the raft.

(10) Breaking holes occur where a large quantity of water flows over a sleeper and the drop is not steep enough to create a suction hole (figure 23-18). A wave of standing water, much like an ocean breaker, is found downstream. This wave is stationary and can vary from 1 to 10 feet high, and even though it lacks the strong upriver flow of a suction hole, it can be a trap for rafts too small to climb up and over the crest. The size of the breaking hole and the survivor's seamanship must be considered before tackling this obstacle.

c. Suction Hole or Three-Dimensional Eddy:

(1) The vacuum created by a suction hole is strong enough to pull a survivor wearing a life preserver be-

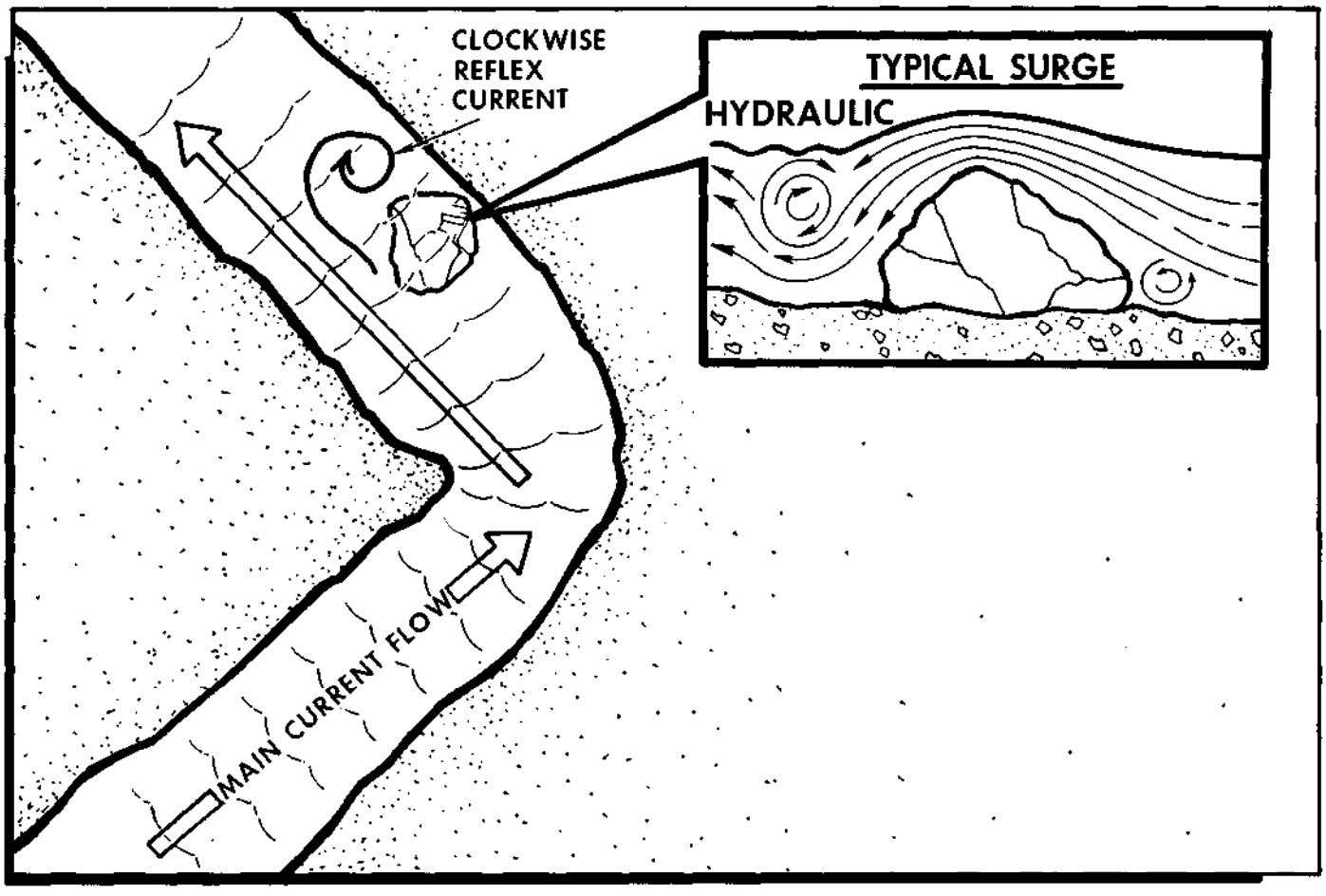

Figure 23-16. Response to a Submerged Rock.

neath the surface. If pulled down into a suction hole, the survivor will normally be whirled to the surface downriver and returned to the suction hole by the upriver flow, to be pushed under once again. Objects too buoyant to sink normally remain trapped. It's usually difficult to identify a suction hole because there is no frothing, no obvious curling water, and little noise. Extreme caution should be used when a large bulge appears in the water (figure 23-19). There are three possible ways of surviving and avoiding serious injury in suction holes. One way is to find the layer of water below the surface which is moving in the same desired direction. The second way is to reach down with a paddle or hand and feel for a current which is moving out of the hole. However, in a large suction hole the downstream flow will be too deep to reach. The survivor should attempt to cut across through the side of the eddy into the water rushing by. The final and best solution is to scout ahead and try to identify the location of suction holes and avoid them.

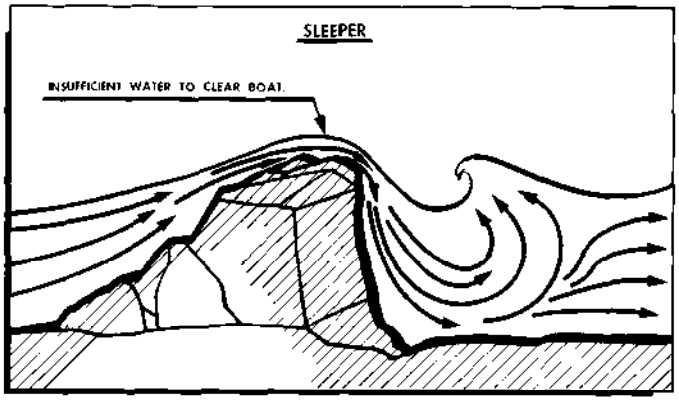

Figure 23-17. Sleeper.

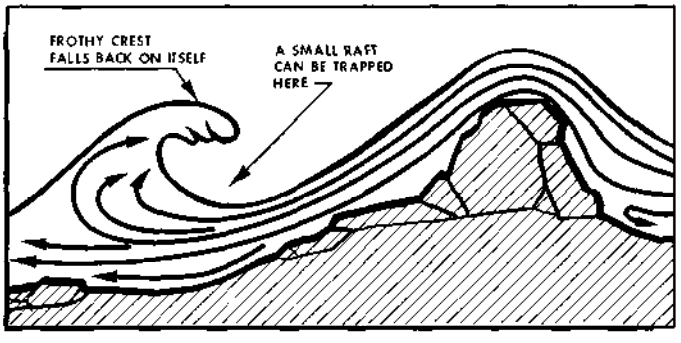

Figure 23-18. Breaking Hole.

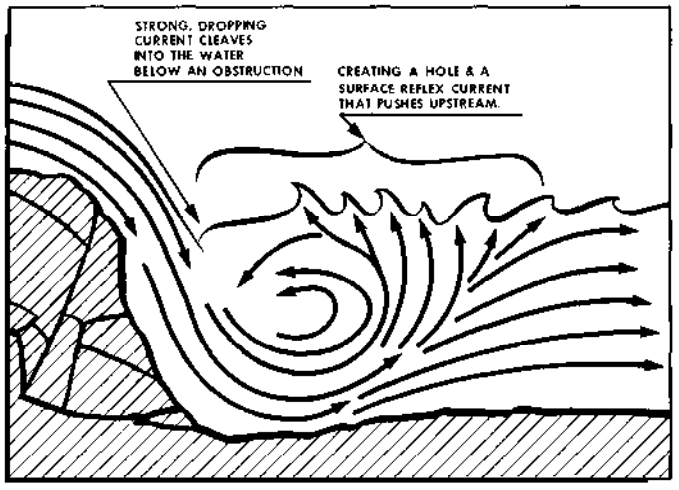

Figure 23-19. Suction Hole.

(2) An eddy (figure 23-20) is a reaction to an obstruction. The type of eddy which occurs next to the bank is caused by portions of the main current being deflected and forced to flow back upriver where it again joins the mainstream. These areas are usually associated with quiet and slow-flowing water. They are also associated with areas where the river widens or just above or below a bend in the river. An eddy has two distinct currents: the upstream current, and the downstream current. The dividing line between the two is called an eddy fence. It is a line of small whirlpools spun off the upriver current by the power of the downstream current.

(3) A two-dimensional eddy is when the tip of an obstruction is slightly above the water and causes a two-dimensional flow around the obstacle. Because of the speed and power of the current, the water is super-elevated and is significantly higher than the level of water directly behind the obstacle. This creates a hole which is filled by the flow of water around the obstacle (figure 23-21). Two-dimensional eddies which occur in midstream will create two eddy fences, one on each side

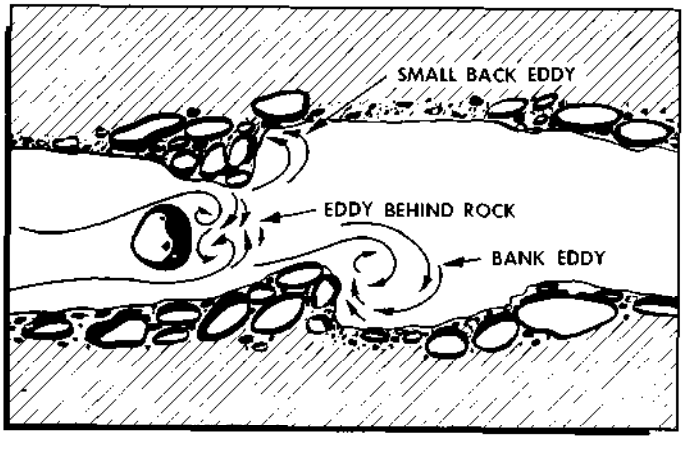

Figure 23-20. Eddies.

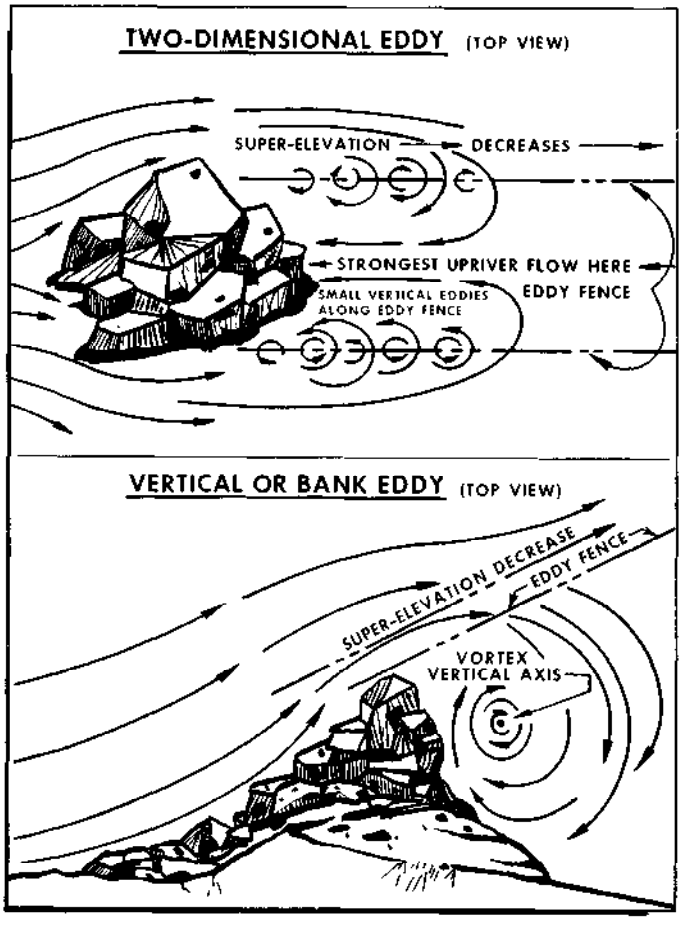

Figure 23-21. Two-Dimensional Eddies.

of the obstacle. The water will enter the depression from both sides and will travel in a circular motion, clockwise from the right bank, counterclockwise from the left bank, and back upstream directly behind the obstacle. If the projection is large enough and a strong circular motion is created, it becomes a whirlpool. The outer reaches of the swirling water are super-elevated by centrifugal force and a suction is created, similar to a drain in a bathtub. These vortexes are very rare and usually occur on huge rivers. Survivors may stop and rest where the eddies occur since there should be no strong swift currents in the eddy. If the obstacle is huge, it may be impossible to paddle fast enough to cross the eddy fence without being spun around in a pinwheeling manner.

d. Falls. In most falls, there are two reflex currents or suction holes forming, both whirling on crosscurrent axes into the falls (figure 23-22). One current falls behind the crashing main stream of water while the other falls in front. It's not too dangerous on a 2- or 3-foot drop, but when heights of 6 or more feet are present, it could be a death trap. The suction holes formed below these falls are inherently inescapable because of their power. The foam and froth formed at the bottom of falls

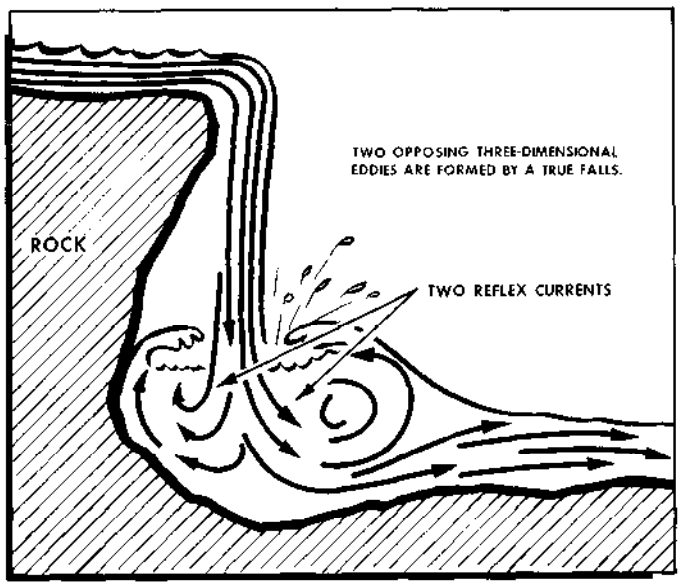

Figure 23-22. Falls.

can be dangerous. Jagged boulders and other hazards may be hidden.

e. Boils. A boil may occur below a fall or sleeper, downstream of the curling or reverse-current suction. This appears as a dome or mound-shaped water formation. Boils are the result of layers of flows hitting bottom, aimed upward, and reaching the surface parting into a flowerlike flow. The water billowing out in boils is super-oxygenated, taking away the resistance needed to push with the paddles or to suspend a survivor in a life preserver.

f. Rollers. Another difficulty found when traveling on fast rivers are rollers. Rollers are large, cresting waves caused by a variety of situations. A wave seen below a breaking hole is one type of roller. Velocity waves are another type which occur on straight stretches of fast dropping waters and caused by the drag of sandy banks and submerged sandbars. They may be large enough to overturn a raft, but are easily recognized, and are usually regular and easy to navigate by keeping the raft direction of travel in line with the crest of the wave (figure 23-23).

g. Tail Waves. Tail waves are quiet waves which are a reflex from the current hitting small rocks along the bed of the river and deflecting the current toward the surface. They are usually so calm that going over them is not noticeable.

h. Bank Rollers. Bank rollers are similar in appearance to crested tail waves and occur when there is a sharp bend in the riverbed which turns so sharply the water can't readily turn in the bend. The water slams into the outside bank and super-elevates, falling back upon itself. Small bank rollers cause few problems, but large ones cresting 5 feet or higher can capsize a raft (figure 23-24). There are three terms used to specify the severity or height of rollers; one being washboard. Washboard rollers are a series of swells which gently ripple and are safe and easy to ride. The next stage of rollers is called standing water (cresting rollers), where the speed of contours on the bottom are such that the tops of the swells fall back onto themselves. The most dangerous and insurmountable rollers are referred to as haystacks or roosters (figure 23-25). They resemble haystacks of water coming down from every direction—giant frothing bulges. They may be so high a raft cannot go over them. It would be easy to become trapped in a deep trough (depression) and be buried under tons of frothing water. These troughs also hide dangers such as sharp rocks which could tear a raft or break through the bottom of a boat.

i. Chutes. River chutes are good, logical travel routes, but they may harbor dangers. A chute (or tongue) is a

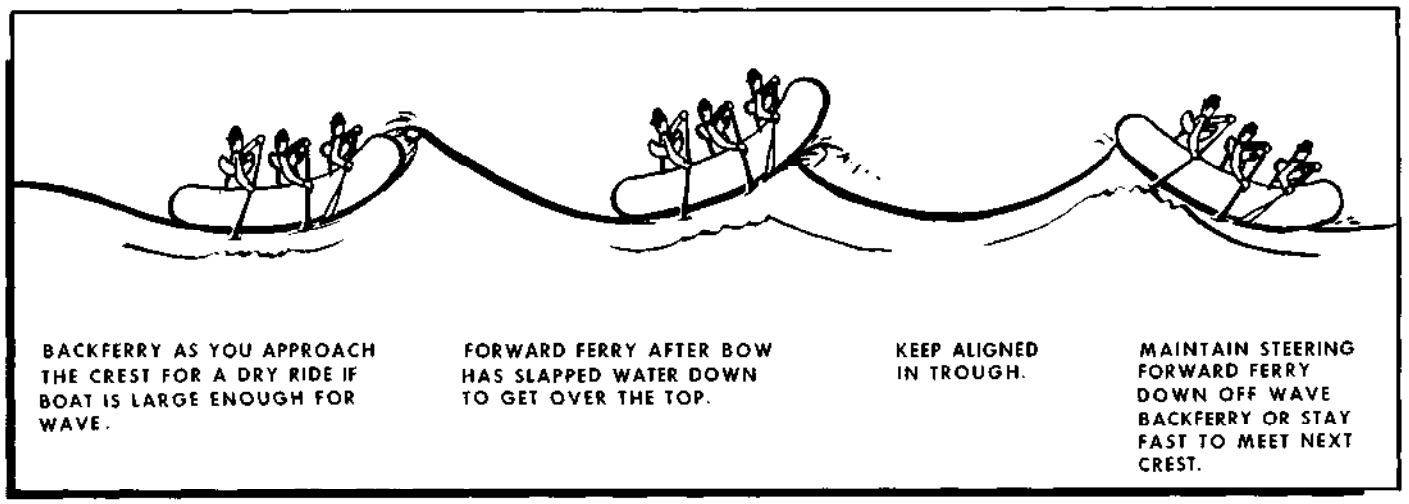

Figure 23-23. Tail Waves and Rollers.

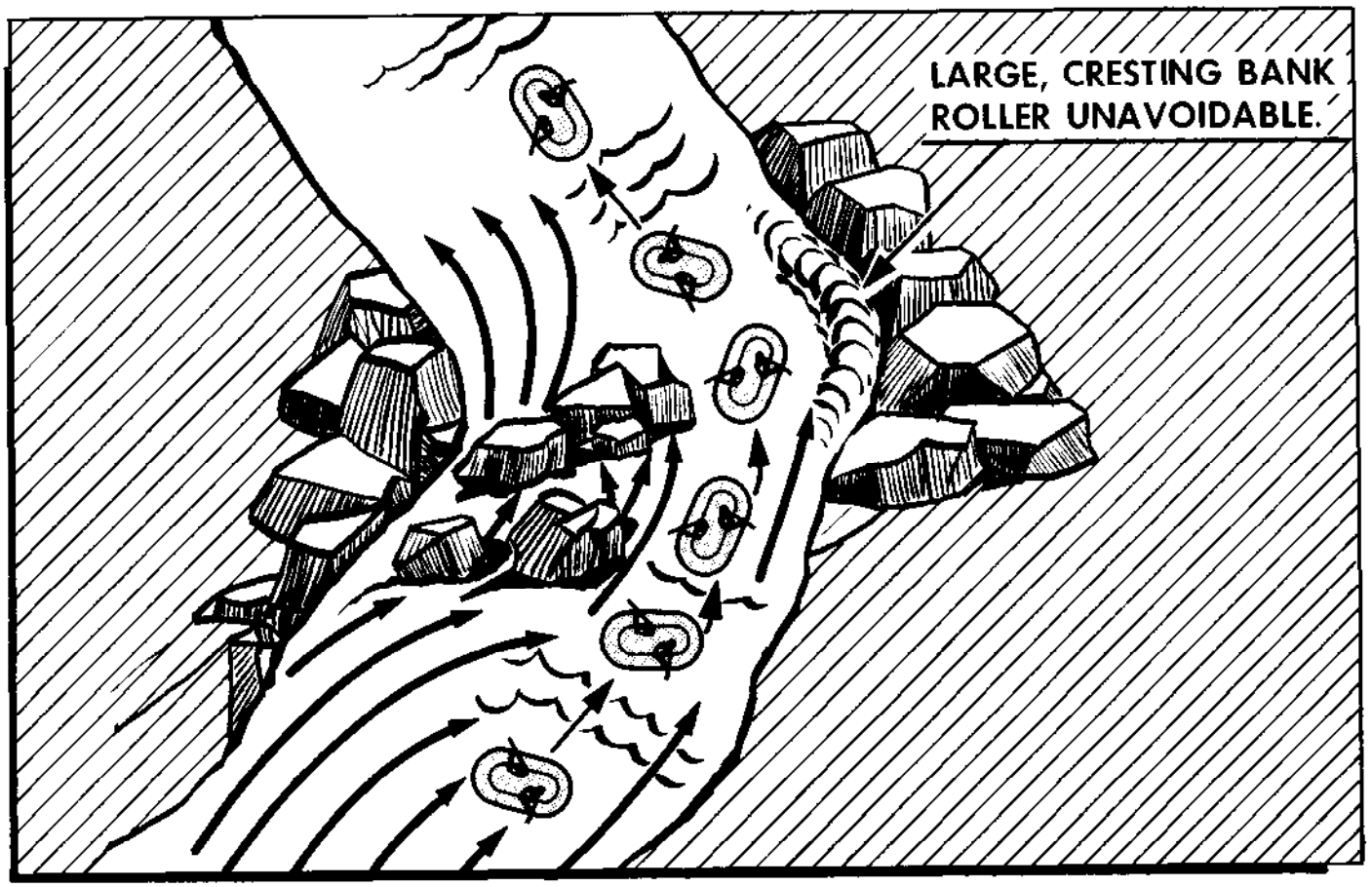

Figure 23-24. Bank Roller.

swift-running narrow passage between river obstructions caused by a damming effect. An example would be water forced between two large boulders. Because the waterflow is restricted, it accelerates and a powerful current rushes through. Because of the water velocity, there may be a suction hole on either side of the chute (figure 23-26).

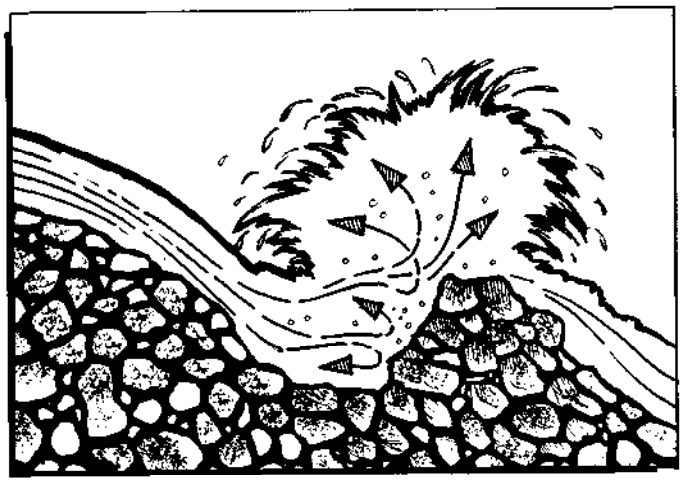

Figure 23-25. Haystack.

j. Log Jams (Tongue of the Rapids). Log jams are extremely dangerous. They consist of logs, brush, and debris collected from high waters that become lodged across the current. They remain stationary while the river flows through them. If a craft should be swept up against the stationary logs, it will be pinned in place by the current. Should the craft be tipped and swamped, it could be swept under the log jam. If the current is strong

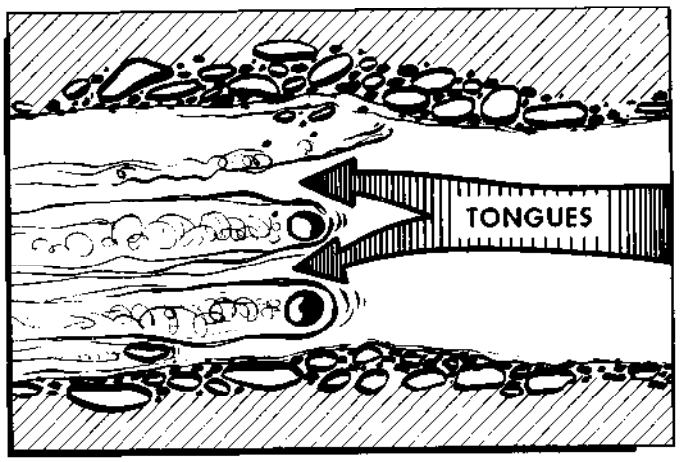

Figure 23-26. Tongue of the Rapids.

enough to do this, the occupants may also be pinned underneath the jam.

k. Sweepers. Sweepers can be the most dangerous obstructions in rivers. A sweeper is a large tree growing on a riverbank which has fallen over and is resting at or near the surface of the water. It may bounce up and down with the current. Survivors may be suddenly confronted with a sweeper which blocks the channel while rounding a bend in a river. The survivors are relatively helpless when it encounters sweepers in swift water. The only precautionary measure is to land above a bend in order to study the river ahead. (NOTE: Many people have met disaster by hitting sweepers.)

23-5. Emergency Situations:

a. Rock Collisions. Collisions with rocks above the surface of the water are common occurrences on a river. If the collision is unavoidable, survivors should spin the raft powerfully just before contact, or hit the rock bow-on. If the survivor is able to spin the raft, it will usually turn the raft off and around the rock. If they hit the rock bow-on, it will stop the raft momentarily, giving time to manipulate a spin-off with a few turn strokes. When the bow-on method is used, occupants in the stern (aft) should move to the center of the raft before impact. This allows the stern to raise and the rushing current to slide under the raft. If the stern is low, the water will pile against it causing the raft to be swamped. If survivors are colliding broadside with a rock, the entire crew must immediately jump to the side of the raft nearest the rock—always being the raft's downstream side (figure 23-27). This should be done before contact. If not, the river will flow over and suck down the raft's upstream tube. The raft will be flooded and the powerful force of the current will wrap it around the rock, possibly trapping some or all of the crew between the rock and the raft. Once it is unswamped, the broached (broadsided) raft on a rock is usually freed easily (figure 23-28). If two people push with both feet on the rock in the direction of the current, the raft will swing or slide into the pull of the current. The rest of the crew should shift to the end which is swinging into the current. These methods rarely fail; however, if the raft refuses to budge, it can be freed using the enormous power of the current. Large gear bags or sea anchors are securely tied to a long rope and secured to the end of the raft expected to swing downstream. The sea anchors (gear bags) are then tossed downstream into the current. (NOTE: A safety line should be used.)

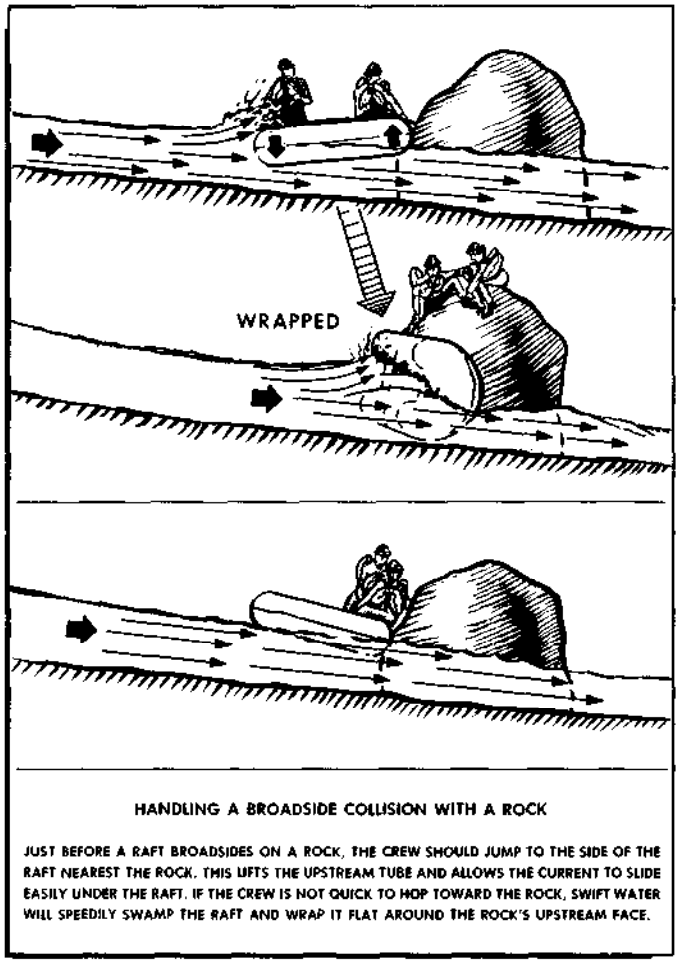

Figure 23-27. Handling Broadside Collision with Rock.

b. Freeing a Wrapped Raft. Sometimes the powerful force of swift water may pin and wrap a raft around a rock. It is unusual for a raft to be equally balanced around the rock, so it will move more easily one way than it will the other. The part of the raft with more weight and bulk should be moved toward the flow of the current. Lines are attached to at least two points on the raft so, when pulled, the force is equally distributed. One of these points is on the far end of the raft, around the tube and is called the hauling line. (NOTE and CAUTION: A small hole may need to be cut in the floor of the raft to pass the line through and around the tube

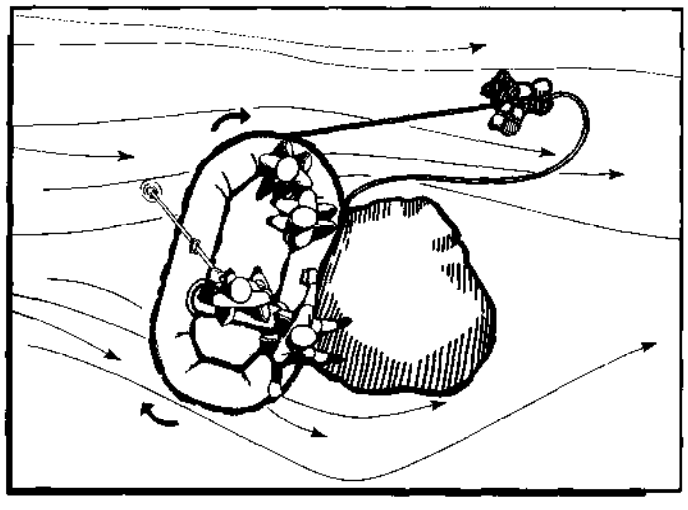

Figure 23-28. Freeing Unswamped Raft.

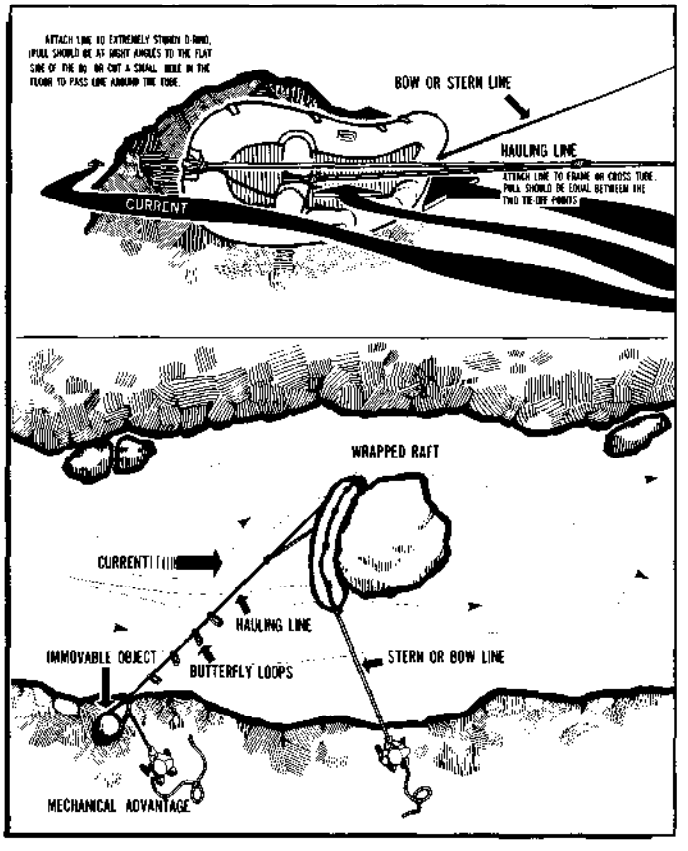

Figure 23-29. Freeing a Wrapped Raft.

if there is no ring to pass it through.) The second tie-off point can be a cross tube. One person should hold the line attached to the stern of the raft. The raft should be moved over the rock into the pull of the current. Once it is freed, the person holding the stern line can move the raft to a safe position.

c. Raft Flips. When a raft is about to flip, there is little time to react to the situation. If the raft is diving into a big hole, the primary danger is being violently thrown forward into a solid object in the raft. Survivors should protect themselves by dropping low and flattening themselves against the backside of the baggage or cross tube. If the raft is being upset by a rock, fallen tree, or other obstacles, members should jump clear of the raft to prevent being crushed against the obstruction or struck by the falling raft. If the raft is pinned flat against an obstruction, the members should stay with the raft and try to safely climb up the obstruction (figure 23-29).

d. Lining Unrunnable Rafts. Lining a raft through rapids is basically letting it run through rapids with a crew on shore controlling it by attached lines (figure 23-30). The raft should be moved slowly by maintaining tight control of the lines attached to the bow and stern. If no strong eddies or steep narrow chutes are present, one member should walk along the shore and control the raft with a strong, long line. The lines running to shore should be long enough to allow the raft full travel through the rapids.

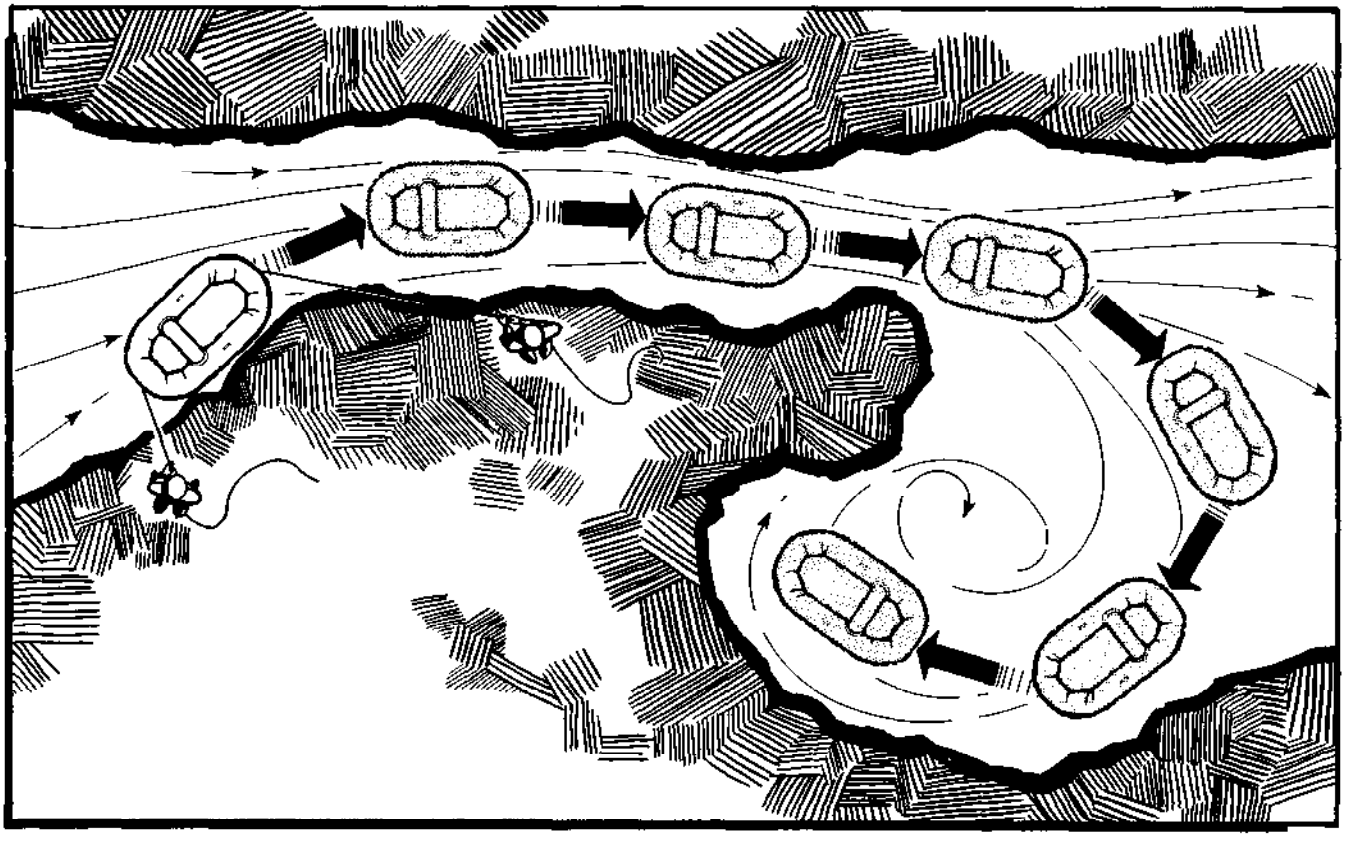

Figure 23-30. Lining a Chute.

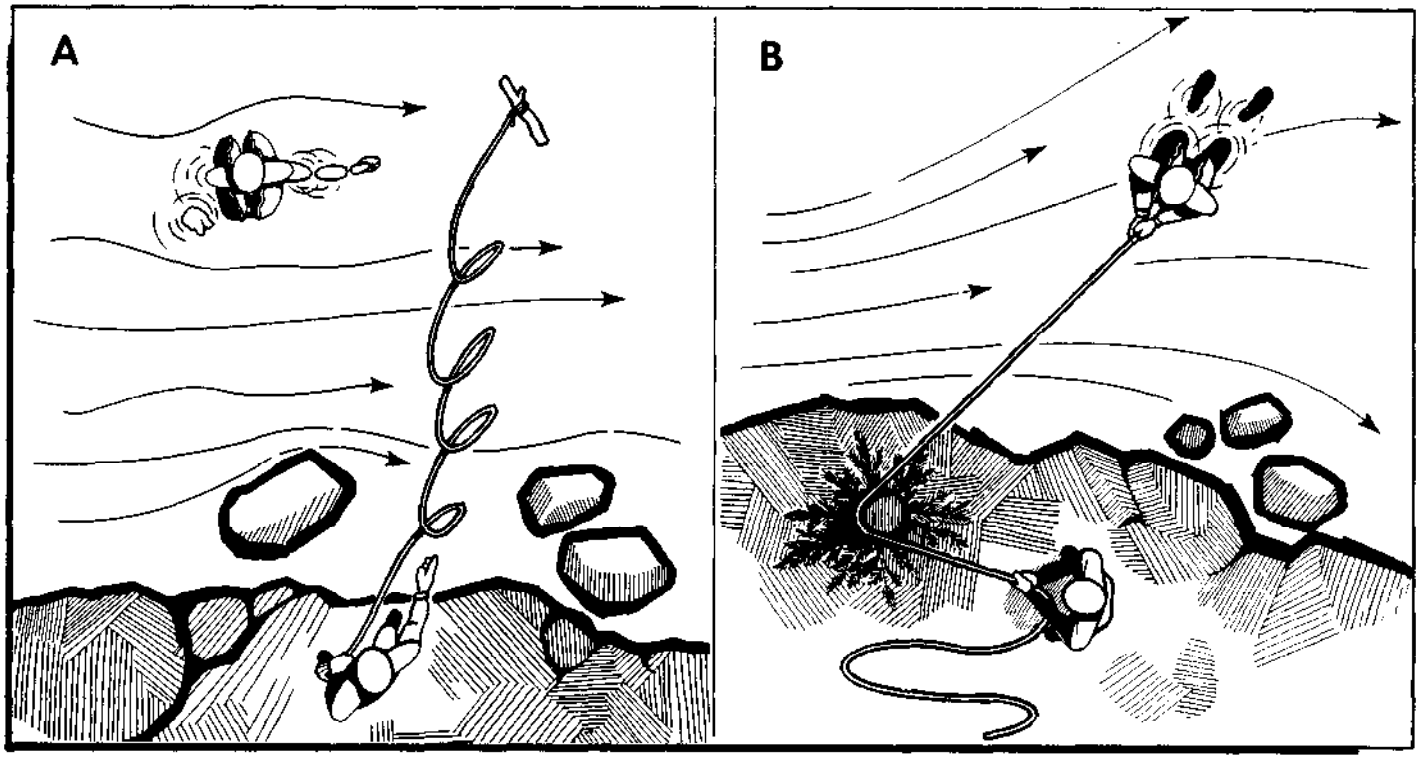

Figure 23-31. Making a Rescue.

e. Rescuing a Swimmer from Shore. The rescuer should carefully choose the right spot where the coil of rope thrown to the swimmer will not cross hazardous areas or obstacles, and yet be near a rock or tree which can be used for belaying (securing without being tied) the end of the line. The person throwing the line should make sure the line has a flotation device (life preserver) at the end before it is thrown to the swimmer. The rope should be coiled in a manner which will allow it to flow smoothly, without entanglement, to full extension. One hand holds one-half to one-third of the coil while the other throws the remainder of the coil out. The weighted coil should be thrown to a spot where the swimmer will drift, which is usually downstream, in front of the swimmer. As the rope travels out, all of the line except the last 10 or 12 feet should be uncoiled. The member

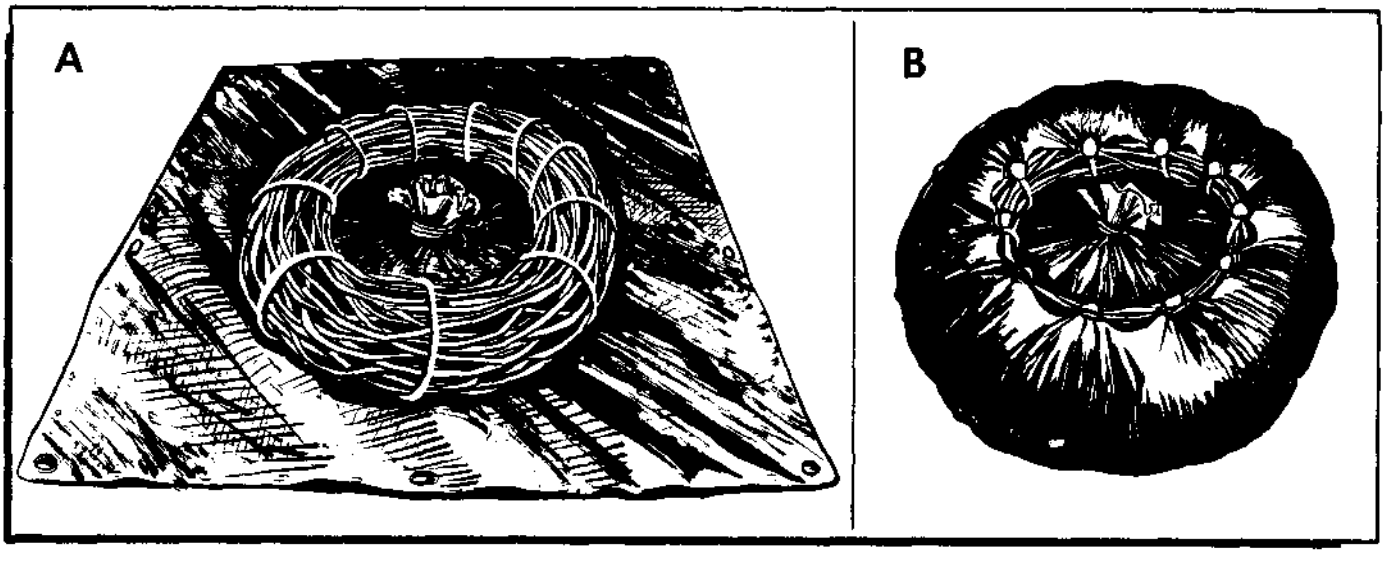

Figure 23-32. Poncho Ring Raft.

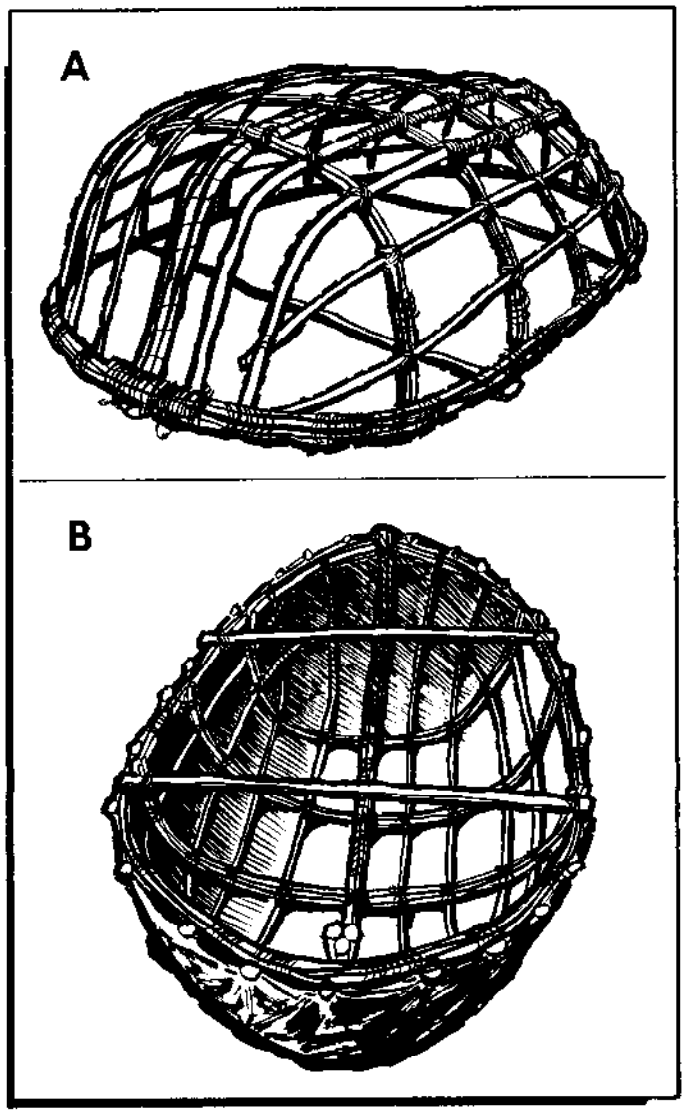

Figure 23-33. Bull Boat.

pulling the survivor should be braced or have the line around a rock or tree to hold the swimmer once the line has been reached by the survivor.

f. The Swimmer's Responsibility. The swimmer should be aware of the rescuer's location and face downstream when waiting for the hauling line. When the line is thrown, the swimmer will normally be required to swim to reach it. Once holding the line, the swimmer should be prepared for a very strong pull from the current and line. The line should be held tightly, but not wrapped around a wrist or hand! Entanglement must be avoided. The swimmer should pull, hand-over-hand, until reaching shallow water, and then use the rope for steadiness while walking to shore (figure 23-31).

23-6. Improvised Rafts:

a. Types of Flotation Devices. There are various types of flotation devices which may be improvised and used as rafts for equipment and personnel.

(1) The Poncho (donut) raft (figure 23-32) can be used for transporting equipment, but is not a good vehicle for people. The raft is constructed by using saplings or pliable willows and a waterproof cover. A hoop-shaped framework of saplings or pliable willow is constructed within a circle of stakes. The hoop is tied with cordage or suspension line and removed from the circle of stakes and placed on the waterproof cover to which it will be attached. Clothing and (or) equipment is then placed in the raft and the survivor swims, pushing the raft.

(2) The bull boat (figure 23-33) is a shallow-draft skin boat shaped like a tub and formerly used by Indians in the Great Plains area. The survivors should construct an oval frame, similar to a canoe, of willow or other pliable materials and cover the framework with waterproof material such as a signal paulin or skins. This makes a craft which is suitable for transporting equipment across a river with the survivor propelling it from behind.

(3) An emergency boat can be made by stretching a tarpaulin or light canvas cover over a skillfully shaped framework of willows and adding a well-framed keel of green wood, such as slender pieces of spruce. Gunwales (sides) of slender saplings are attached at both ends and the spreaders or thwarts are attached as in a canoe. Ribs of strong willows are tied to the keel. The ends of the ribs are bent upward and tied to the gunwales. The inside of the frame is closely covered with willows to form a deck upon which to stand. Such a boat is easy to handle and is buoyant, but lacks the strength necessary for long journeys. This boat is entirely satisfactory for ferrying a group across a broad, quiet stretch of river. When such a boat has served its purpose, the cover should be removed for later use.

(4) The vegetation raft is built of small vegetation which will float and is placed within clothing or parachute to form a raft for a survivor and (or) equipment. Plants such as water hyacinth or cattail may be used (figure 23-34).

(5) A good floating device for the single survivor can be fabricated by using two balsa logs or other lightweight wood. The logs should be placed about 2 feet apart and tied together. The survivor sits on the lines and travels with the current (figure 23-35).

(6) A dugout canoe is good transportation, but difficult to construct. One method is to build a long fire on the side to be dug out and chop away the burned material when the fire is out. Repeat this procedure as often as necessary.

b. Building a Raft. The greatest problem in raft construction (figure 23-36) is being able to construct a craft strong enough to withstand the buffeting it may have to take from rocks and swift water. Even if 6- to 8-inch

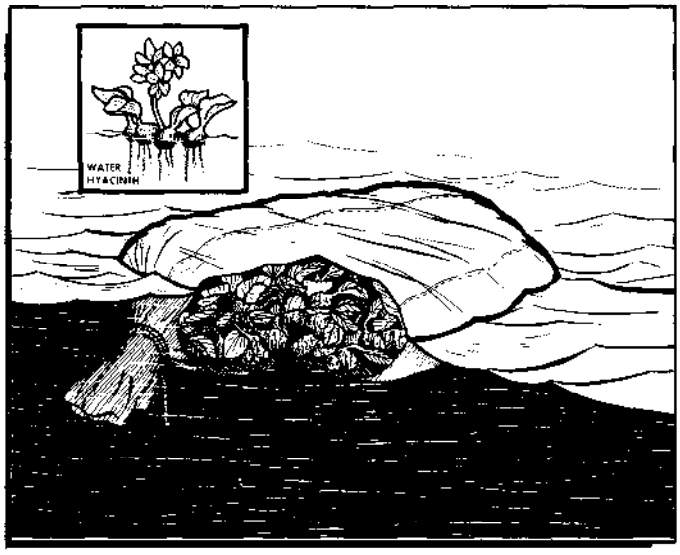

Figure 23-34. Vegetation Bag.

spikes are available, they are not satisfactory since they pull or twist out easily. Rope quickly wears out from frequent, rough contact with rocks and gravel. Northern woodsmen have evolved a construction method (figure 23-37) which requires neither spikes nor rope, yet produces a raft superior in strength. The only material required are logs, although rope is sometimes useful; the only tools needed are an axe and a sheath knife.

23-7. Fording Streams:

a. Survivors traveling on foot through wilderness areas may have to ford some streams. These can range from small, ankle-deep brooks to large rivers. Rivers are often so swift a survivor can hear boulders on the bottom being crashed together by the current. If these streams are of glacial origin, the survivor should wait for them to decrease in strength during the night hours before attempting to ford.

b. Careful study is required to find a place to safely ford a stream. If there is a high vantage point beside the river, the survivor should climb the rise and look over the river. Finding a safe crossing area may be easy if the river breaks into a number of small channels. The area on the opposite bank should be surveyed to make sure travel will be easier after crossing. When selecting a fording site, the survivor should:

(1) When possible, select a travel course which leads across the current at about a 45-degree angle downstream.

(2) Never attempt to ford a stream directly above, or close to, a deep or rapid waterfall or a deep channel. The stream should be crossed where the opposite side is comprised of shallow banks or sandbars.

(3) Avoid rocky places, since a fall may cause serious injury. However, an occasional rock which breaks the current may be of some assistance. The depth of the water is not necessarily a deterrent. Deep water may run more slowly and be safer than shallow water.

(4) Before entering the water, the survivors should have a plan of action for making the crossing. Use all possible precautions, and if the stream appears treacherous, take the steps shown in figure 23-38.

23-8. Traveling on Open Seas. Four-fifths of the Earth's surface is covered by open water. Although accounts of sea survival incidents are often gloomy, successful survival is possible; however, the raft is at the mercy of the currents and winds. There are many currents, both warm and cold, throughout the seas.

a. Currents. Sea currents flow in a clockwise direction in the Northern Hemisphere and counterclockwise in the Southern Hemisphere. This is caused by three factors: the Sun's heat, the winds, and the Earth's rotation (Coriolis Effect). Most sea currents travel at speeds of less than 5 miles per hour. Using currents as a mode of travel can be done by putting out the sea anchor and letting the current pull the raft along. Survivors should

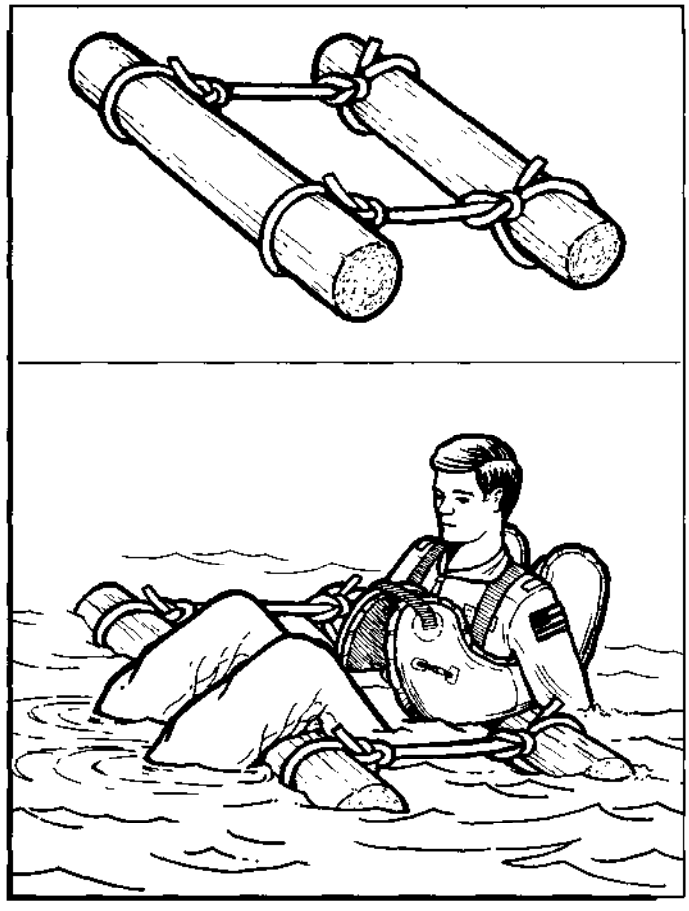

Figure 23-35. Log Flotation.

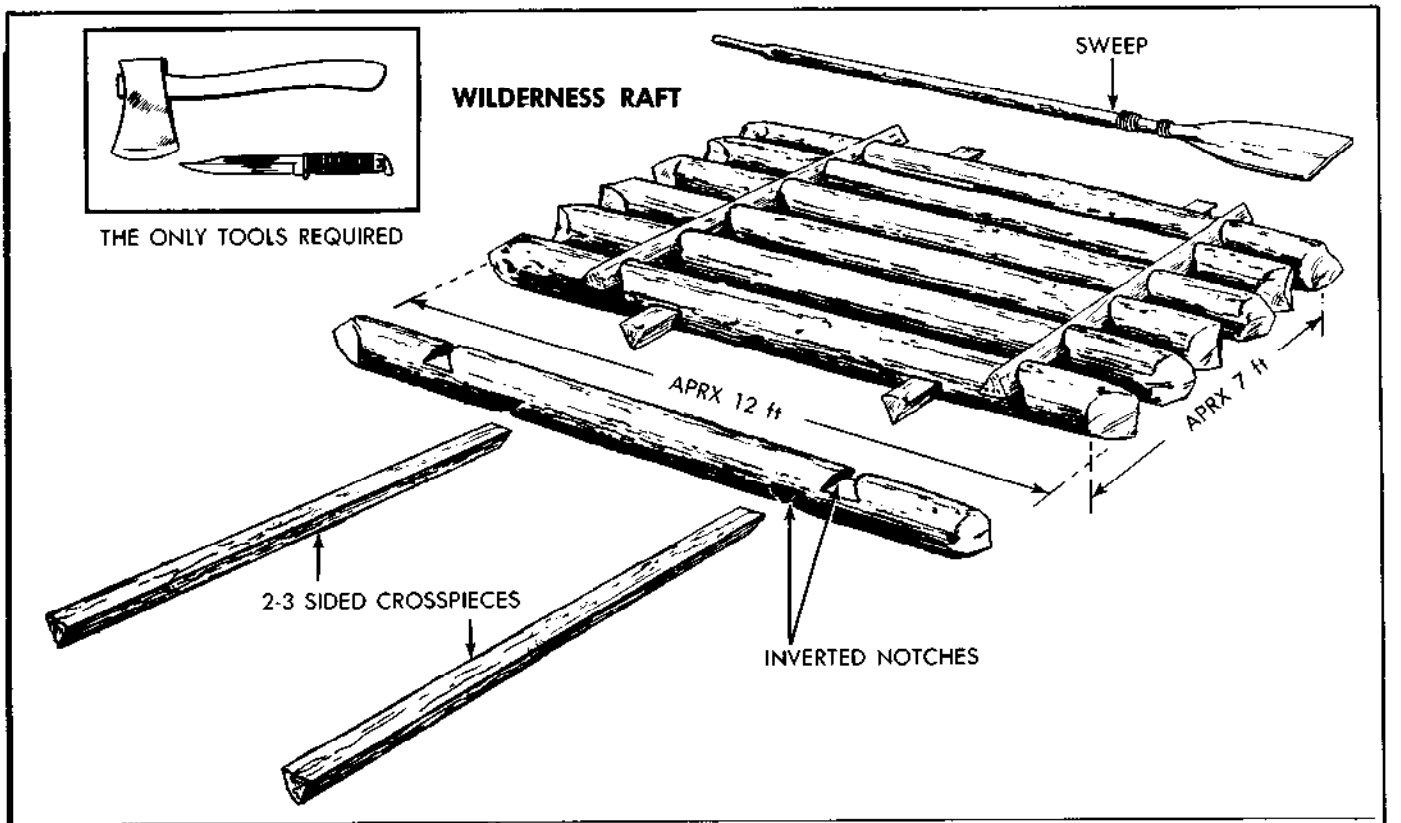

RAFT CONSTRUCTION

A raft for three persons should be about 12 feet long and 6 feet wide, depending on the size of the logs used. The logs should be 12 to 14 inches in diameter and so well matched in size that notches you make in them are level when crosspieces are driven into place.

Build the raft on two skid logs placed so that they slope downward to the bank. Smooth the logs with an ax so that the raft logs lie evenly on them. Cut two sets of slightly offset inverted notches, one in the top and bottom of both ends of each log. Make the notches broader at the base than at the outer edge of the log, as shown in the illustration. Use small poles with straight edges or a string pulled taut to make the notches. A three-sided wooden crosspiece about a foot longer than the total width of the raft is to be driven through each end of the four sets of notches.

Complete the notches on all logs at the top of the logs. Turn the logs over and drive a 3-sided crosspiece through both sets of notches on the underside of the raft. Then complete the top set of notches and drive through the two additional sets of crosspieces.

You can lash together the overhanging ends of the two crosspieces at each end of the raft to give it added strength; however, when the crosspieces are immersed in water they swell and tightly bind the raft logs together.

If the crosspieces fit too loosely, wedge them with thin, boardlike pieces of wood split from a dead log. When the raft is in water, the wood swells, and the crosspieces become very tight and strong.

Make a deck of light poles on top of the raft to keep packs and other gear dry.

Figure 23-36. Raft Construction.

use caution when traveling through areas where warm and cold currents meet. It can be a storm forming area with dense fog and high winds and waves.

b. Winds. Winds also aid raft travel. In tropical areas, the winds are easterly blowing (trade winds). In higher latitudes, they blow from the west (westerlies). To use the winds as a mode of travel, the sea anchor should be pulled into the raft and, if available, a sail should be erected.

c. Waves. Waves can be both an asset and a hazard to raft travel. Waves are normally formed by the wind. The severity of the wind determines the size of the waves. On open seas, waves range from a few inches to over 100 feet in height. Under normal conditions, waves alone will move a liferaft only a few inches at a time; therefore, using waves as a mode of propulsion is not practical. Waves are a great help in finding land or shallow areas in the sea. Ocean waves always break when they enter shallow water or when they encounter an obstruction. The force of energy of the wave depends on how abruptly the water depth decreases and the size of the waves. Breaking waves can be used as an aid to

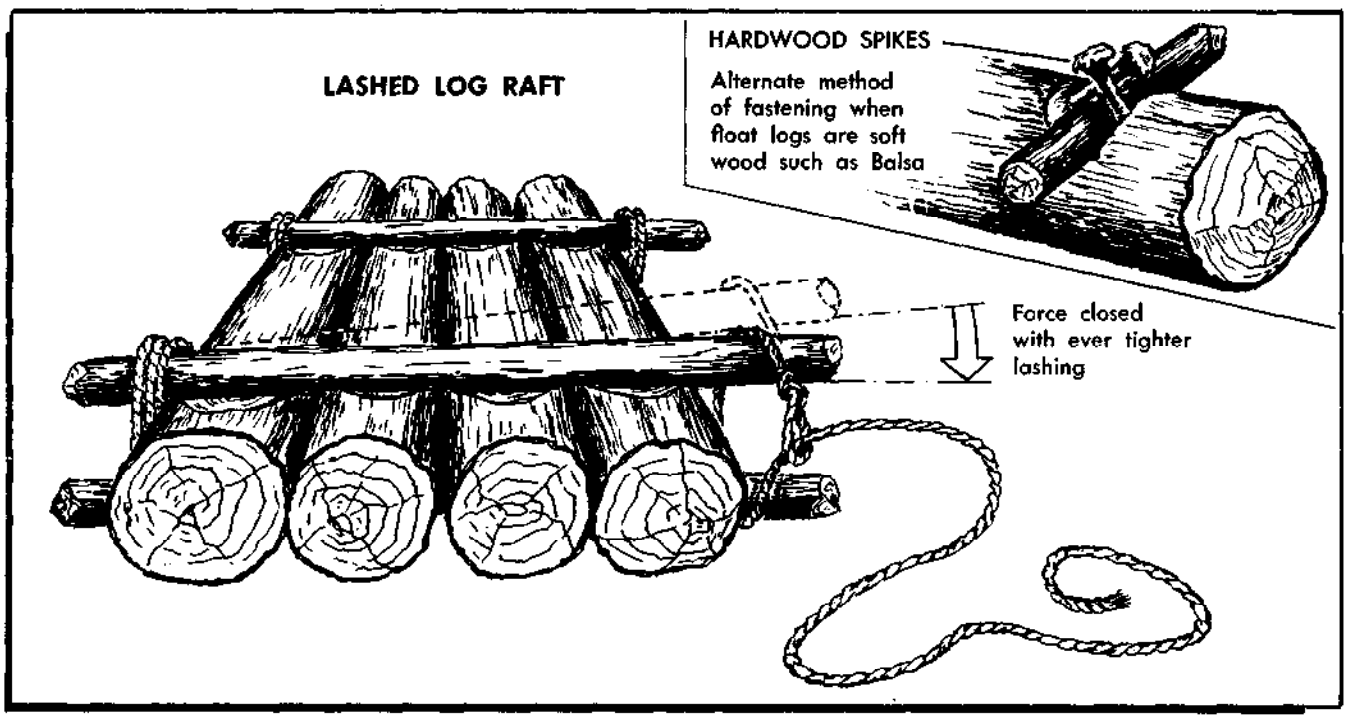

Figure 23-37. Lashed Log Raft.

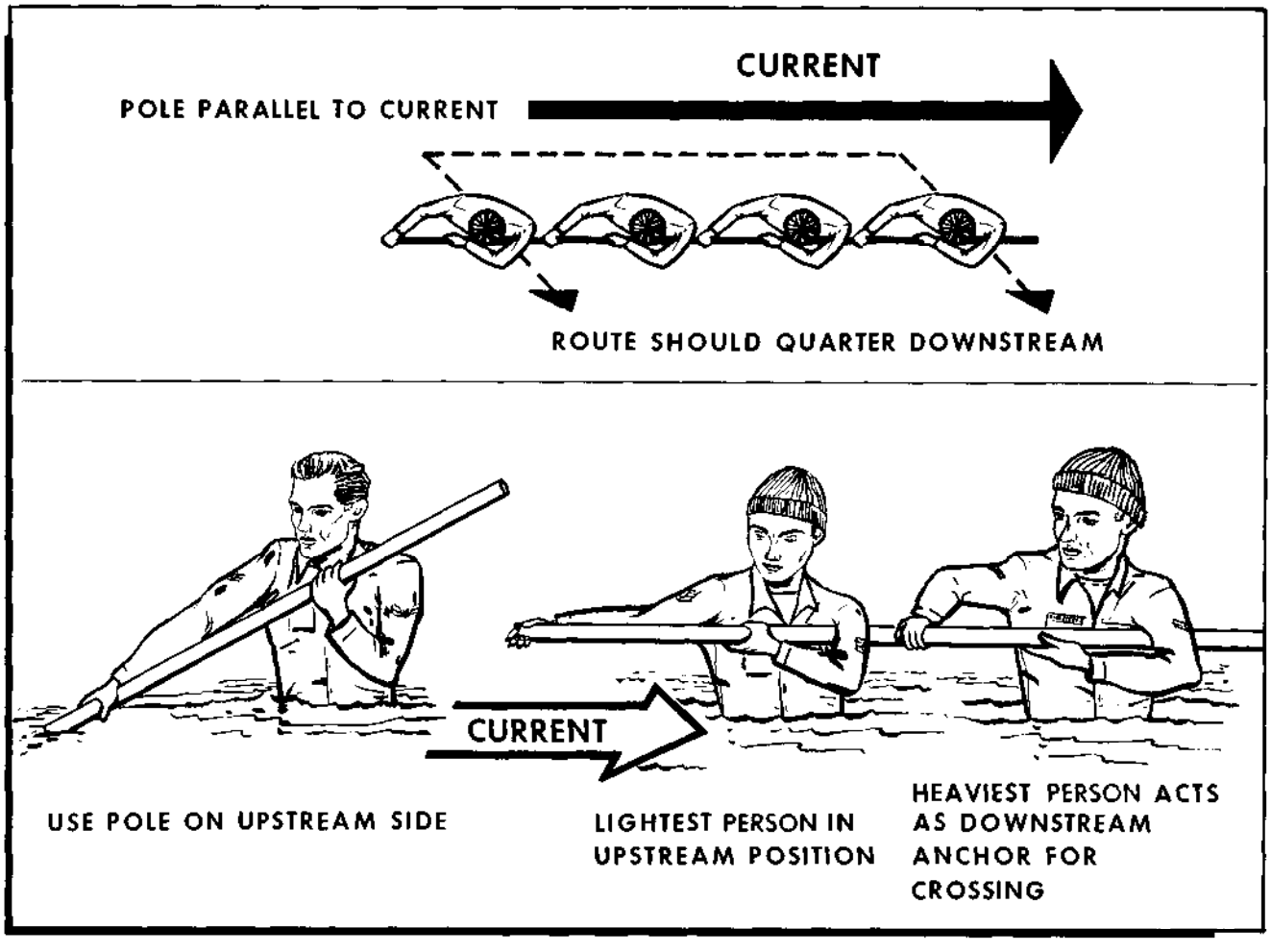

Figure 23-38. Fording a Treacherous Stream.

make a landfall. Storms at sea are probably the greatest hazard to survivors in rafts. Aside from the waves created by a storm, the wind and rain can make life in a raft very difficult. The waves and wind can capsize a raft, or throw a person out of the raft, and then will constantly fill the raft with water. Seasickness can result from gentle to severe wave action. Additionally, rescue efforts may be severely hampered by large waves.

d. Tides. Tides are another form of wave but they are very predictable. They occur twice daily and usually cause no problems for anyone in a raft. Tides may range from one to 40 feet in height depending on the area of the world. They should be considered when planning a landing. When the tide is going in, it will help propel the raft to shore. The action of the water going away from shore when the tide is going out makes landing difficult.

e. Hazards. Certain marine life must be considered a hazard to survival. Sharks, jellyfish, eels, and most reef fishes can cause serious injuries if encountered. Waste materials should be disposed of when these creatures are not present.

(1) The survivor should be aware of the saltwater (estuarine) crocodile. It is found throughout the Southeast Asian shoreline. It is a well known maneater, and is almost always found in salt or brackish water. It is more commonly found near river mouths and along the coasts; however, it has been known to swim as much as 40 miles out into the sea. Females with nests are likely to be vicious and aggressive. They will grow to a length of 30 feet but most specimens are less than 15 feet in length. Survivors should watch for this reptile while landing their raft or fishing.

(2) The survivor will normally encounter reef fishes during the landing process by stepping on them. They may also be caught while fishing. Clothing and footgear must be worn at all times whether landing raft or fishing.

(3) Coral is normally found in warm waters, along the shores of islands and mainlands. There are many different types of coral. They should be avoided since all can destroy a raft or severely injure a survivor. It is best to stay in the raft when coral is encountered. If the survivor must wade to shore, footgear and pants should be worn for protection. Moving slowly and watching every step may prevent serious injuries. Coral does not exist where freshwater enters the sea.

(4) Ships can be a welcome sight to the survivor, but they can also be a hazard. Since the raft is a small object in a very large sea, the survivor must constantly be aware that at night or during inclement weather, the raft will be difficult to see and could be struck by a large ship.

f. Early Considerations. Survivors should stay upwind and clear of the aircraft (out of fuel covered waters) but in the vicinity of the crash until the aircraft sinks. A search for survivors is usually activated around the entire area of and near the crash site. Missing personnel may be unconscious and floating low in the water. Rescue procedures are illustrated in figure 23-39.

(1) The best technique for rescuing aircrew members from the water is to throw them a line with a life preserver attached. The second is to send a swimmer (rescuer) from the raft with a line using a flotation device which will support the weight of a rescuer. This will help to conserve energy while recovering the survivor. The least acceptable technique is to send an attached swimmer without floatable devices to retreive a survivor. In all cases, the rescuer should wear a life preserver. The strength of a person in a state of panic in the water should not be underestimated. A careful approach can prevent injury to the rescuer.

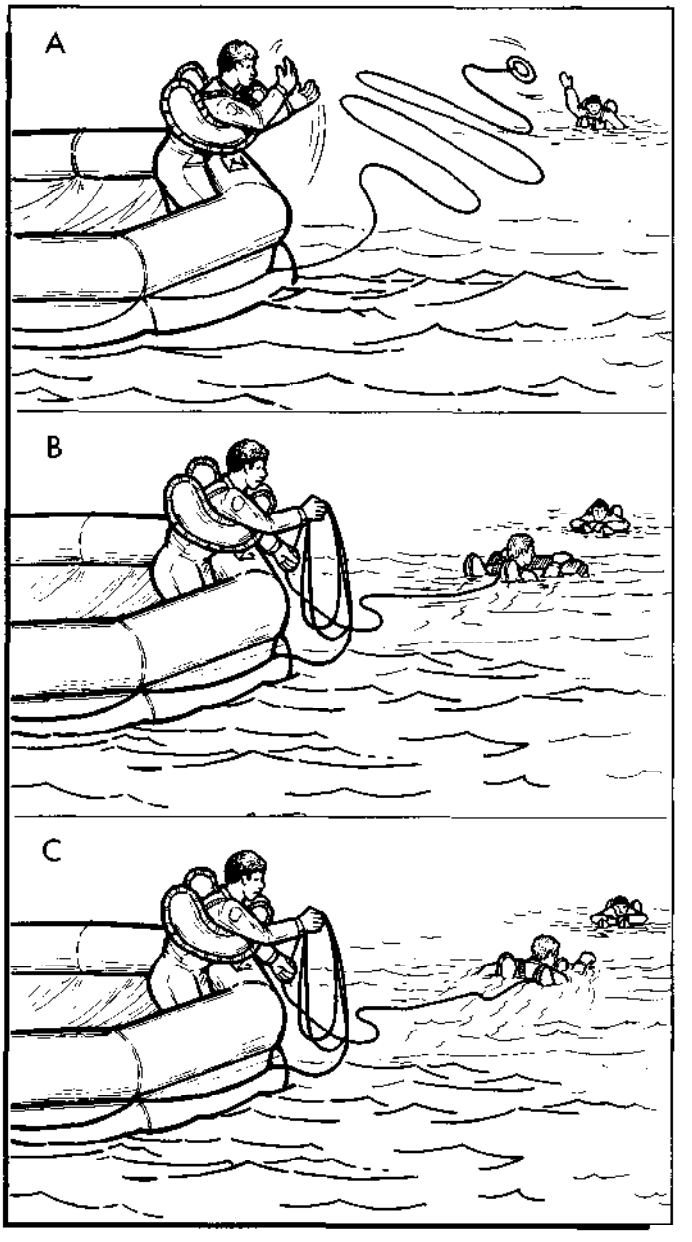

Figure 23-39. Rescue from Water.

(2) When the rescuer is approaching a survivor in trouble from behind, there is little danger of being kicked, scratched, or grabbed. The rescuer should swim to a point directly behind the survivor and grasp the backstrap of the life preserver. A sidestroke may then be used to drag the survivor to the raft.

(3) All debris from the aircraft should be inspected and salvaged (rations, canteens, thermos and other containers, parachutes, seat cushions, extra clothing, maps, etc). Secure equipment to the raft to prevent loss. Special precaution should be taken with flashlights and signaling equipment to keep them dry so they will function when needed.

(4) Rafts should be checked for inflation. Leaks, and points of possible chafing should be repaired, as required. All water should be removed from inside the raft. Care should be taken to avoid snagging the raft with shoes or sharp objects. Placing the sea anchor out will slow the rate of drift. If there is more than one raft, they should be connected with at least 25 feet of line. The lifeline attached to the outer periphery of the raft is to be used for the connection. Donning the antiexposure suit is essential in cold climates. Erecting wind breaks, spray shields, and canopies will protect survivors from the elements. Survivors should huddle together and exercise regularly to maintain body heat.

(5) Monitoring the physical condition of survivors and administering first aid to survivors is essential. If available, seasickness pills will help prevent vomiting and resulting dehydration.

(6) Survivors should prepare all available signaling equipment for immediate use.

(7) Compasses, watches, matches, and lighters will become worthless unless they are kept dry.

(8) The raft repair plugs should be attached to the raft for easy access as soon as possible.

(9) All areas of the body should be protected from the Sun. Precautions should be taken to prevent sunburn on the eyelids, under the chin, and on the backs of the ears. Sunburn cream and Chapstick will protect these areas.

(10) The leader should calmly analyze the situation and plan a course of action to include duty assignments (watch duty, procuring and rationing food and water, etc.). All survivors, except those who are badly injured or completely exhausted, are expected to perform watch duty, which should not exceed 2 hours. The survivor on watch should be looking for signs of land, passing vessels or aircraft, wreckage, seaweed, schools of fish, birds, and signs of chafing or leaking of the raft.

(11) Food and water can be conserved by saving energy. Survivors should remain calm.

(12) Maintaining a sense of humor will help keep morale high.

(13) The survivor(s) should remember that rescue at sea is a cooperative effort. Search aircraft contacts are limited by the visibility of survivors based on the availability of visual or electronic signaling devices. Visual and electronic communications can be increased by using all available signaling devices (signal mirrors, radios, signal panels, dye marker, and other available devices) when an aircraft is in the area.

(14) A log should be maintained with a record of the navigator's last fix, time of ditching, names and physical condition of survivors, ration schedule, winds, weather, direction of swells, times of sunrise and sunset, and other navigation data.

23-9. Physical Considerations:

a. The greatest problem a survivor is faced with when submerged in cold water is death due to hypothermia. When a survivor is immersed in cold water, hypothermia occurs rapidly due to the decreased insulating quality of wet clothing and the fact that water displaces the layer of still air which normally surrounds the body. Water causes a rate of heat exchange approximately 25 times greater than air at the same temperature. The following lists life expectancy times:

Temperature of Water	*Time*
70° - 60°	12 hours
60° - 50°	6 hours
50° - 40°	1 hour
40° - below	-1 hour

NOTE: These times may be increased with the wearing of an antiexposure suit.

b. The best protection for a survivor against the effects of cold water is to get into the liferaft, stay dry, and insulate the body from the cold surface of the bottom of the liferaft. If this is not possible, wear of the antiexposure suits will extend a survivor's life expectancy considerably. It's important to keep the head and neck out of the water and well insulated from the cold water effects when the temperature is below 66°F. The wearing of life preservers increases the predicted survival time just as the body position in the water increases the probability of survival. The following table shows predicted survival times for an average person in 50°F water:

Situation	*Predicted Survival Time (Hours)*
No Flotation	
Drownproofing	1.5
Treading Water	2.0
With Flotation	
Swimming	2.0
Holding Still	2.7
Help	4.0
Huddle	4.0

Figure 23-40. HELP Position.

(1) Help Body Position. Remaining still and assuming the fetal position, or heat escape lessening posture (HELP) (figure 23-40), will increase the downed crewmember's survival time. About 50 percent of the heat is lost from the head. It is therefore important to keep the head out of the water. Other areas of high heat loss are the neck, the sides, and the groin.

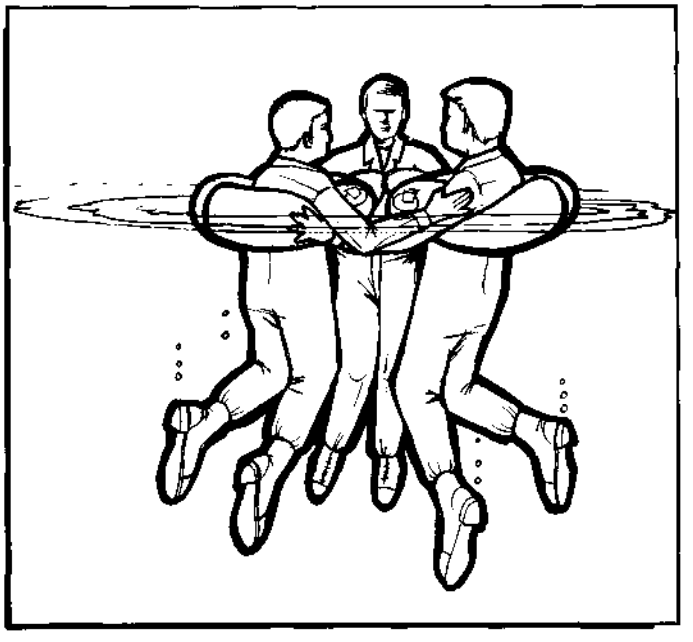

Figure 23-41. Huddling for Temperature Conservation.

(2) Huddling. If there are several survivors in the water, huddling close, side to side in a circle, body heat will be preserved (figure 23-41).

23-10. Life Preserver Use:

a. Survival Swimming Without a Life Preserver. A survivor who knows how to relax in the water is in little danger of drowning, especially in saltwater where the body is of lower density than the water. Trapped air in clothing will help buoy the survivor in the water. If in the water for long periods, the survivor will have to rest from treading water. The survivor may best do this by floating on the back. If this is not possible, the following technique should be used: Rest erect in the water and inhale; put the head face-down in the water and stroke with the arms; rest in this facedown position until there is a need to breathe again; raise the head and exhale; support the body by kicking arms and legs and inhaling; then repeat the cycle.

b. Swimming With a Life Preserver. The bulkiness of clothing, equipment, and (or) any personal injuries will necessitate the immediate need for flotation. Normally, a life preserver will be available for donning before entering the water.

(1) Proper inflation of the life preserver must be done after clearing the aircraft but, preferably, before entering the water. Upon entering the water, the two cells of the life preserver should be fastened together. Limited swimming may be done with the life preserver inflated by cupping the hands and taking strong strokes deep into the water. The life preserver may be slightly deflated to permit better arm movement.

(2) The backstroke should be used to conserve energy when traveling long distances. If aiding an injured or unconscious person, the sidestroke may have to be used. When approaching an object, it is best to use the breaststroke. If a group must swim, they should try to have the strongest swimmer in the lead with any injured persons intermingled within the group. It is best to swim in a single file.

23-11. Raft Procedures. There are three needs which can be satisfied by most of the rafts; they are personal protection, mode of travel, and evasion and camouflage.

a. One-Man Raft:

(1) The one-man raft has a main cell inflation. If the CO_2 bottle should malfunction and not inflate the raft or if the raft develops a leak, it can be inflated orally. The spray shield acts as a shelter from the cold, wind, and water. In some cases, this shield serves as insulation. The insulated bottom plays a significant role in the survivor's protection from hypothermia by limiting the conduction of the cold through the bottom of the raft (figure 23-42).

(2) Travel is more effectively made by inflating or deflating the raft to take advantage of the wind or cur-

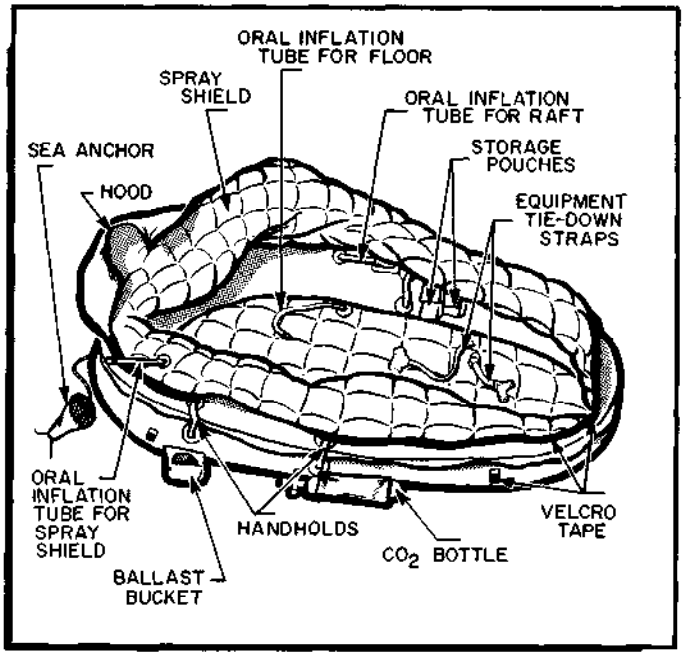

Figure 23-42. One-Man Raft with Spray Shield.

rent. The spray shield can be used as a sail while the ballast buckets serve to increase raft drag in the water. The last device which may be used to control the speed and direction of the raft is the sea anchor. (NOTE: The primary purpose of the sea anchor is to stabilize the raft.)

(3) Black rafts have been developed for use in tactical areas. These rafts blend with the background of the sea. The raft can be further modified for evasion by partially deflating which provides a low profile.

(4) The one-man raft is connected to the aircrew member by a lanyard parachuting to the water. Survivors should not swim to the raft, but pull it to their position via the lanyard. The parachute J-1 releases should be closed and the life preserver separated before boarding the raft (figure 23-43). The raft may hit the water upside down, but may be righted by approaching the bottle side and flipping it over. The spray shield must be in the raft to expose the boarding handles.

(5) If the survivor has an arm injury, boarding is best done by turning the back to the small end of the raft, pushing the raft under the buttocks and lying back (figure 23-44). Another method of boarding is to push down on the small end until one knee is inside and lie forward (figure 23-44).

(6) In rough seas, it may be easier for the survivor to grasp the small end of the raft, and, in a prone position, kick and pull into the raft. Once in the raft, lying face down, the sea anchor should be deployed and adjusted. To sit upright in the raft, one side of the seat kit might have to be disconnected and the survivor should roll to that side. The spray shield is then adjusted. There are two variations of the one-man raft, with the improved model incorporating an inflatable spray shield and floor for additional insulation. The spray shield is designed to help keep the survivor dry and warm in cold oceans, and protect them from the Sun in the hot climates (figure 23-45).

(7) The sea anchor can be adjusted to either act as a drag by slowing down the rate of travel with the current or as a means of traveling with the current. This is done by opening or closing the apex of the sea anchor. When opened, the sea anchor (figure 23-46) will act as a drag and the survivor will stay in the general area. When the sea anchor is closed (figure 23-46), it will form a pocket for the current to strike and propel the raft in the direction of the current. Additionally, the sea anchor should be adjusted so that when the raft is on the crest of a wave, the sea anchor is in the trough of the wave (figure 23-47).

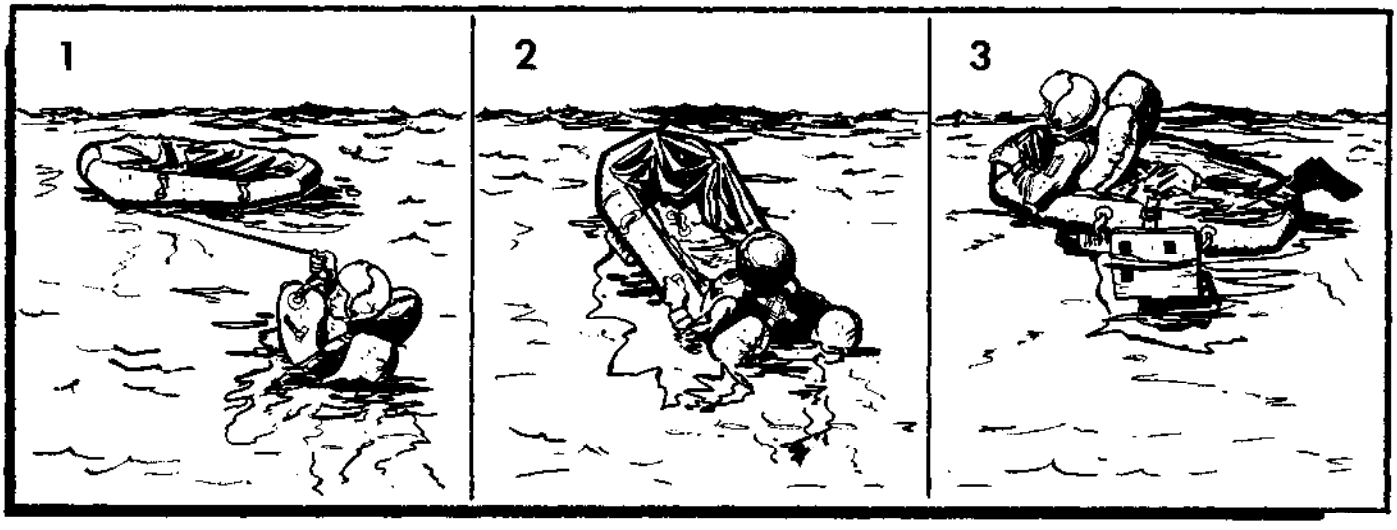

Figure 23-43. Boarding One-Man Raft.

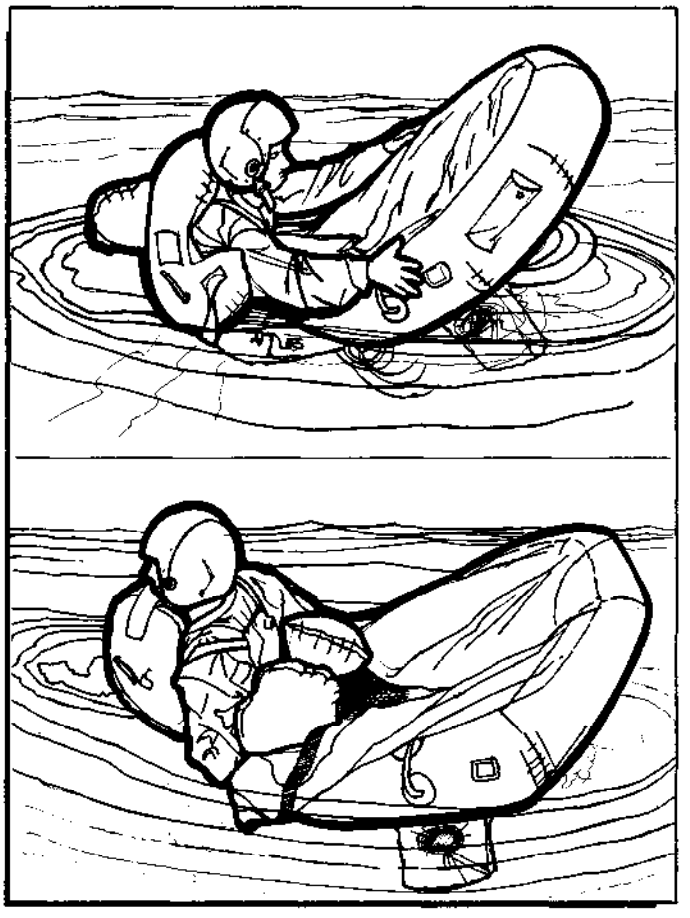

Figure 23-44. Boarding One-Man Raft (Other Methods).

b. Seven-Man Raft:

(1) The seven-man raft is found on some multiplace aircraft. It can also be found in the survival drop kit (MA-1 Kit, figure 23-48). This type of raft may inflate upside down and may, therefore, require the survivor to right the raft before boarding. The J-1 releases should be closed before boarding. The survivor should always work from the bottle side to prevent injury if the raft turns over. Facing into the wind provides additional assistance in righting the raft. The handles on the inside bottom of the raft are used for boarding (figure 23-49).

(2) The boarding ladder is used to board if someone assists in holding down the opposite side. If no assistance is available, the survivor should again work from the bottle side with the wind at the back to help hold down the raft. The survivor should separate the life preserver, grasp an oarlock and boarding handle, kick the legs to get the body prone on the water, and then kick and pull into the raft. If the survivor is weak or injured, the raft may be partially deflated to make boarding easier (figure 23-50).

Figure 23-45. One-Man Raft with Spray Shield Inflated.

(3) Manual inflation can be done by using the pump to keep buoyancy chambers and cross-seat firm, but the raft should not be over inflated. The buoyancy chambers and cross-seat should be rounded but not drum tight. Hot air expands, so on hot days, some air may be released while air may be added on cold days (figure 23-51).

c. Sailing a Raft into the Wind. Rafts are not equipped with keels, so they cannot be sailed into the wind. However, anyone can sail a raft downwind, and multiplace (except 20/25-man) rafts can be successfully sailed 10 degrees off from the direction of the wind. An attempt to sail the raft should not be made unless land is near. If the decision to sail is made and the wind is blowing toward a desired destination, survivors should fully inflate the raft, sit high, take in the sea anchor, rig a sail, and use an oar as a rudder as shown in figure 23-52.

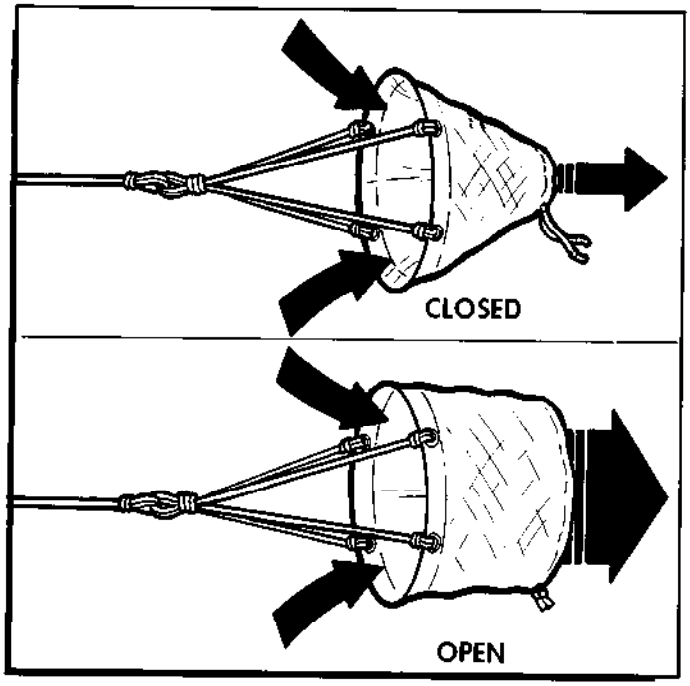

Figure 23-46. Sea Anchor.

d. Multiplace Raft. In a multiplace (except 20/25-man) raft, a square sail should be erected in the bow using oars with their extensions as the mast and crossbar (figure 23-52). A waterproof tarpaulin or parachute material may be used for the sail. If the raft has no regular mast socket and step, the mast may be erected by tying it securely to the front cross-seat using braces. The bottom of the mast must be padded to prevent it from chafing or punching a hole through the floor whether or not a socket is provided. The heel of a shoe, with the toe wedged under the seat, makes a good improvised mast step. The corners of the lower edge of the sail should not be secured. The lines attached to the corners are held with the hands so that a gust of wind will not rip the sail, break the mast, or capsize the raft. Every precaution must be taken to prevent the raft from

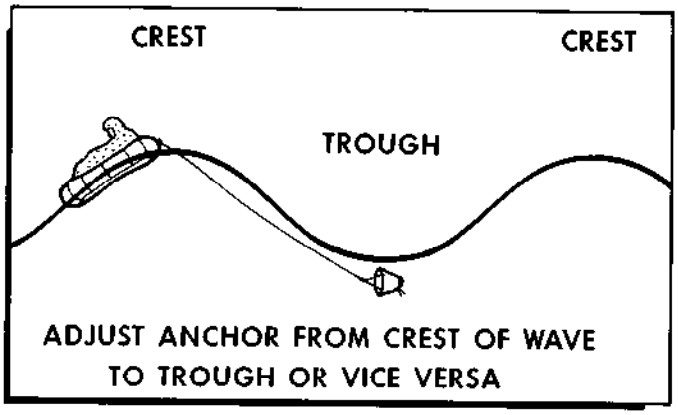

Figure 23-47. Deployment of the Sea Anchor.

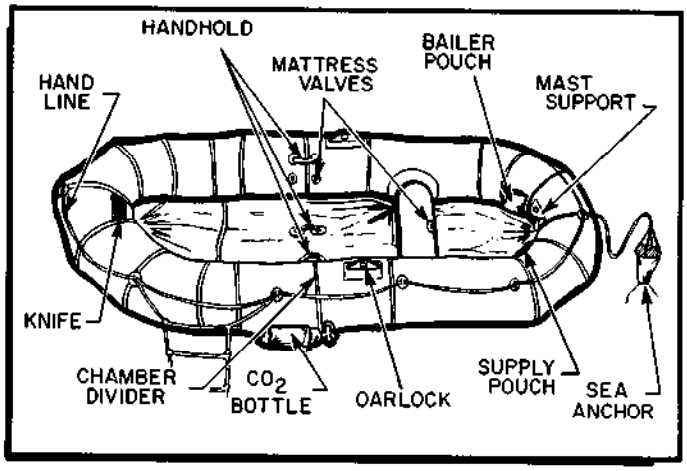

Figure 23-48. Seven-Man Raft.

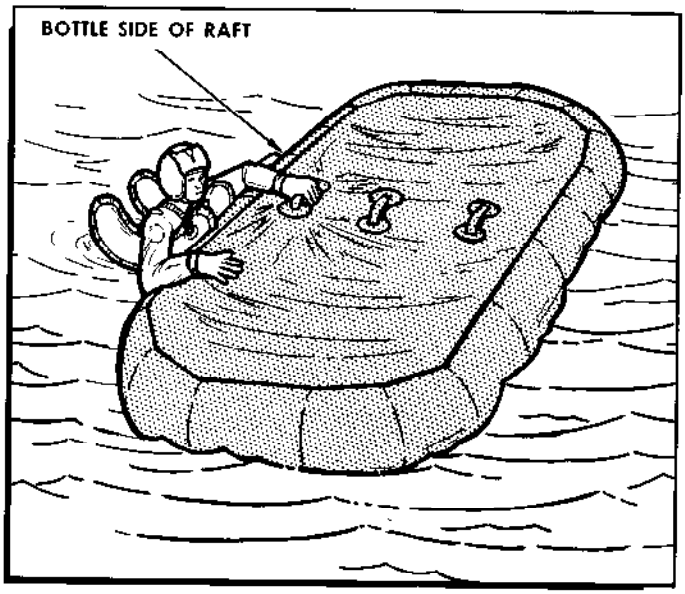

Figure 23-49. Method of Righting Raft.

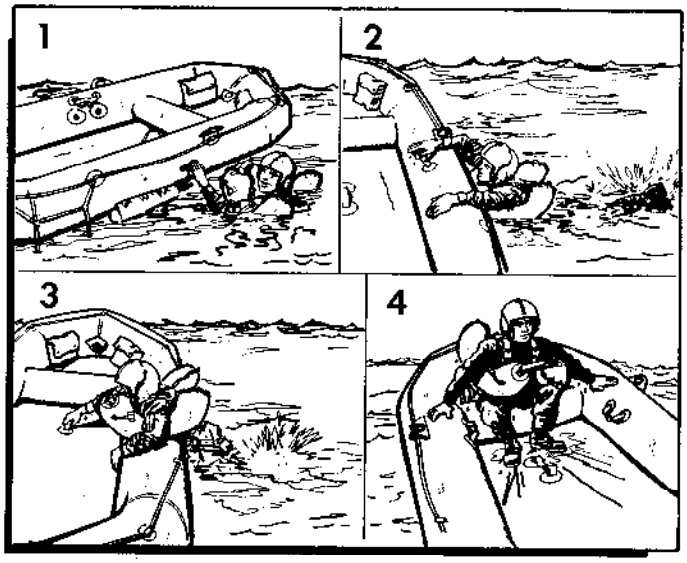

Figure 23-50. Method of Boarding Seven-Man Raft.

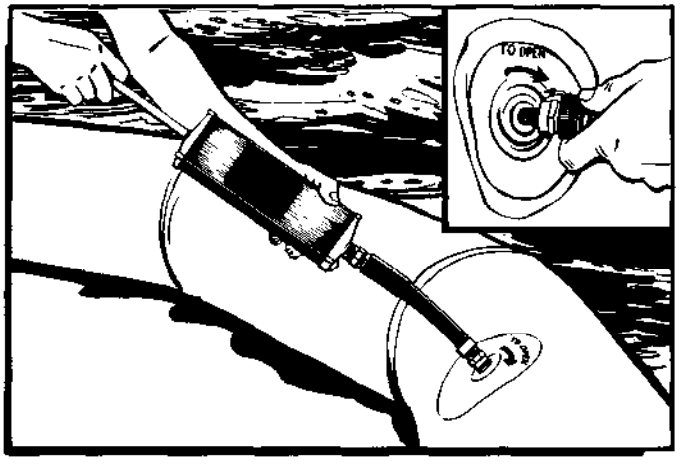

Figure 23-51. Inflating the Raft.

turning over. In rough weather, the sea anchor is kept out away from the bow. The passengers should sit low in the raft with their weight distributed to hold the upwind side down. They should also avoid sitting on the sides of the raft or standing up to prevent falling out. Sudden movements (without warning the other passengers) should be avoided. When the sea anchor is not in use, it should be tied to the raft and stowed in such a manner that it will hold immediately if the raft capsizes.

e. Twenty to Twenty Five-Man Rafts. The 20/25-man rafts may be found in multiplace aircraft (figures 23-53 and 23-54). They will be found in accessible areas of the fuselage or in raft compartments. Some may be automatically deployed from the cockpit, while others may need manual deployment. No matter how the raft lands in the water, it's ready for boarding. The accessory kit is attached by a lanyard and is retrieved by hand. The center chamber must be inflated manually with the hand pump. The 20/25-man raft should be boarded from the aircraft if possible; if not, the following steps should be taken:

(1) Approach lower boarding ramp.

(2) Separate the life preserver.

(3) Grasp the boarding handles and kick the legs to get the body into a prone position on the water's surface; then kick and pull until inside the raft.

(4) If for any reason the raft is not completely inflated, boarding will be made easier by approaching the intersection of the raft and ramp, grasping the upper boarding handle, and swinging one leg onto the center of the ramp as in mounting a horse (figure 23-55).

(5) The equalizer tube should be clamped immediately upon entering the raft to prevent deflating the entire raft, in case of puncture (figure 23-56).

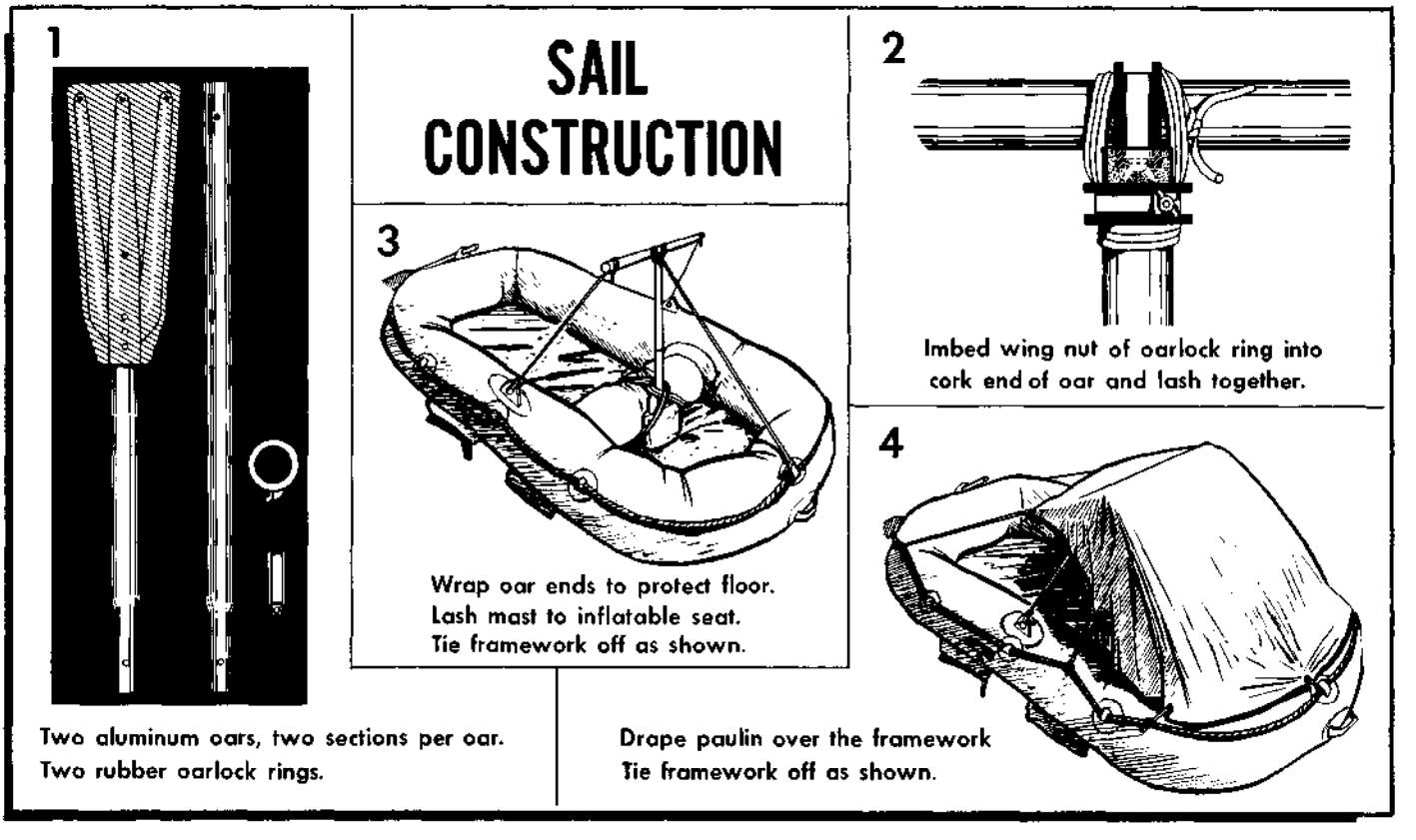

Figure 23-52. Sail Construction.

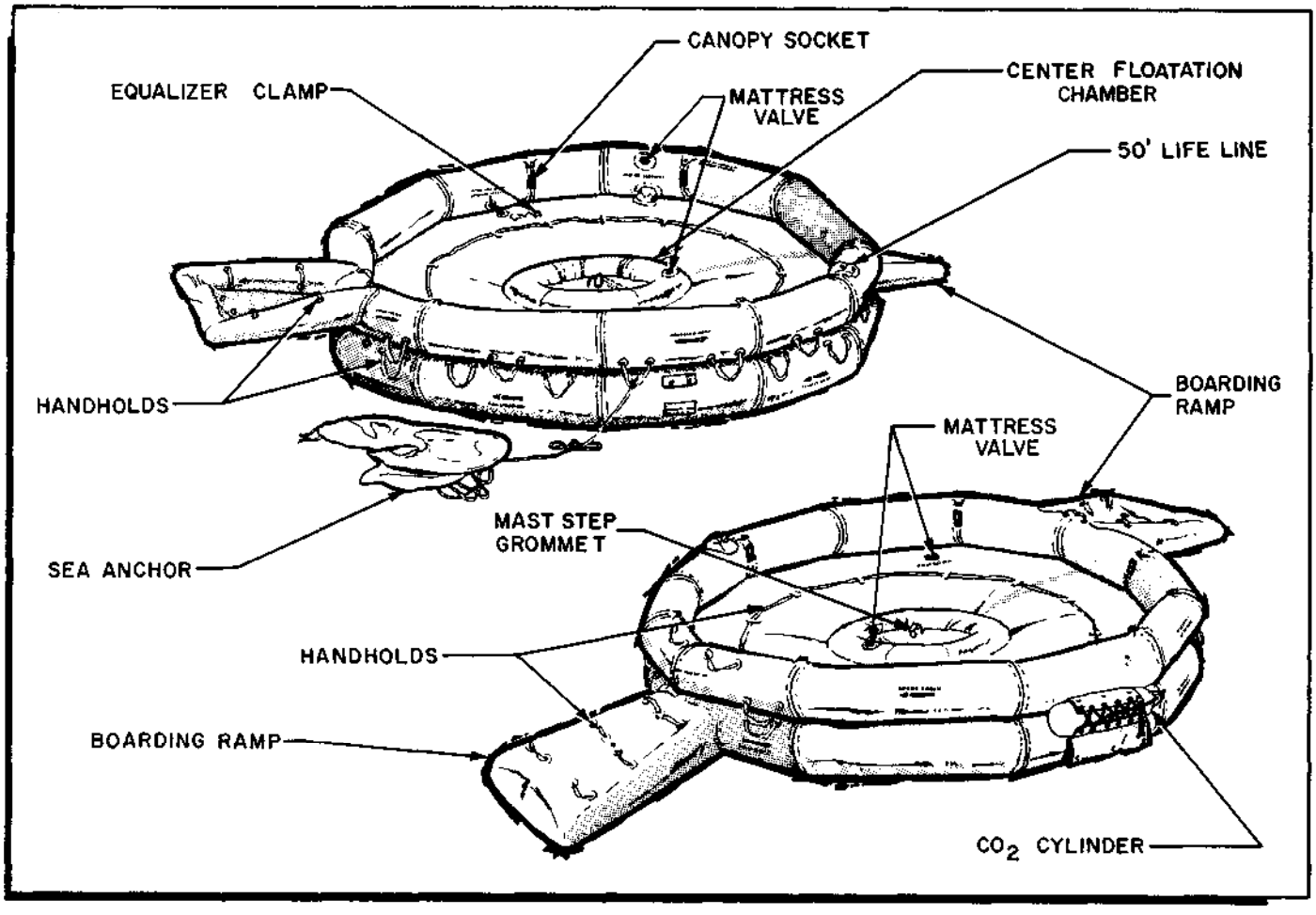

Figure 23-53. 20-Man Raft.

(6) The 20/25-man raft can be inflated by using the pump to keep the chambers and center ring firm. They should be well rounded but not drum tight (figure 23-57).

23-12. Making a Landfall. The lookout should watch carefully for signs of land. Some indications of land are:

a. A fixed cumulus cloud in a clear sky, or in a sky where all other clouds are moving, often hovers over or slightly downwind from an island.

b. In the tropics, a greenish tint in the sky is often caused by the reflection of sunlight from the shallow lagoons or shelves of coral reefs.

c. In the arctic, ice fields or snow-covered land are often indicated by light-colored reflections on clouds, quite different from the darkish gray reflection caused by open water.

d. Deep water is dark green or dark blue. Lighter color indicates shallow water, which may mean land is near.

e. In fog, mist, rain, or at night, when drifting past a nearby shore, land may be detected by characteristic odors and sounds. The musty odor of mangrove swamps and mudflats and the smell of burning wood carries a long way. The roar of surf is heard long before the surf is seen. Continued cries of sea birds from one direction indicate their roosting place on nearby land.

f. Birds are usually more abundant near land than over the open sea. The direction from which flocks fly at dawn and to which they fly at dusk may indicate the direction of land. During the day, birds are searching for food and the direction of flight has no significance unless there is a storm approaching.

g. Land may be detected by the pattern of the waves, which are refracted as they approach land. Figure 23-58 shows the form the waves assume. Land should be located by observing this pattern and turning parallel to the slightly turbulent area marked "X" on the illustration and following its direction.

23-13. Methods of Getting Ashore:

a. Swimming Ashore. This is a most difficult decision. It depends on many things. Some good swimmers

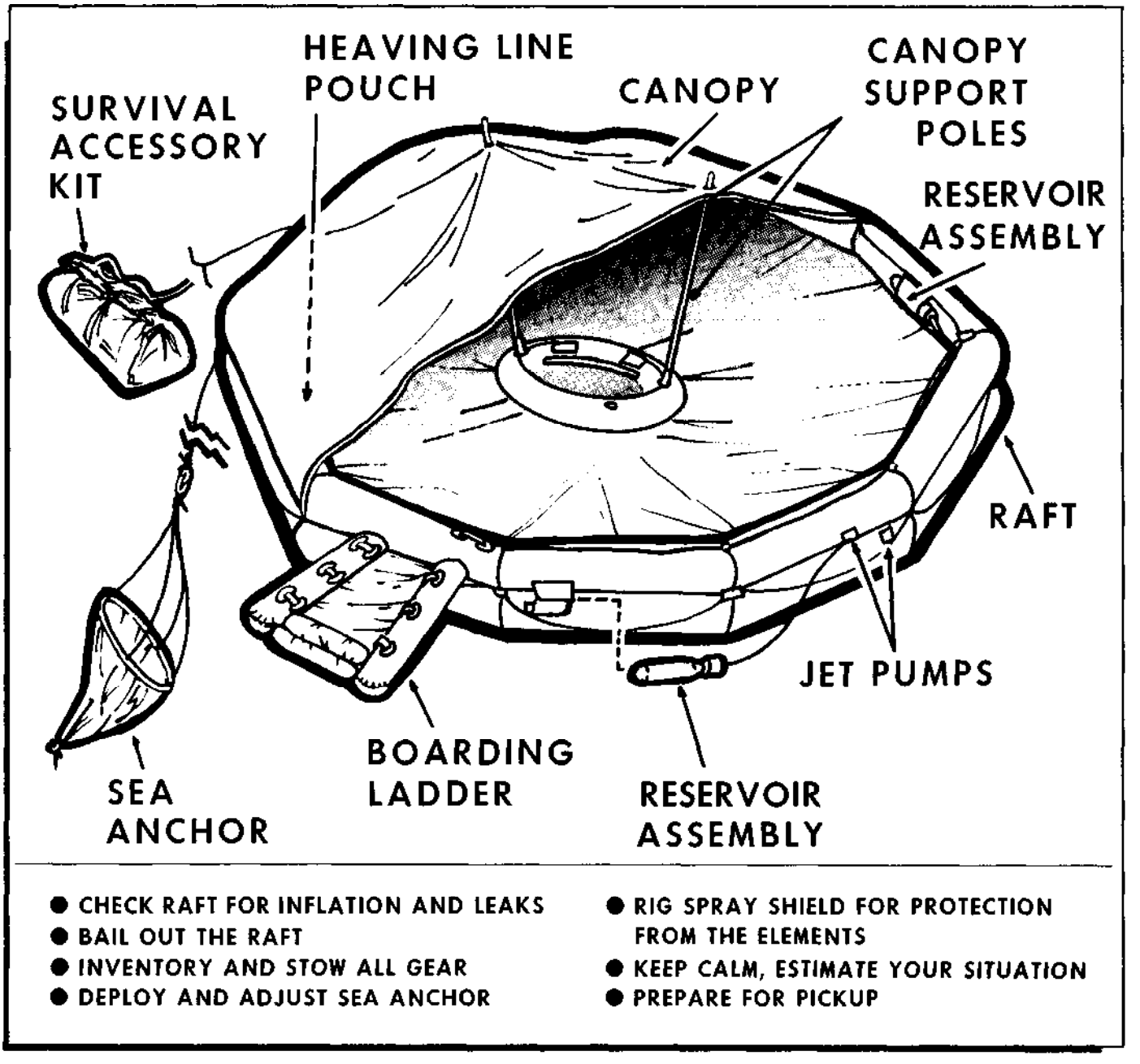

Figure 23-54. 25-Man Raft.

have been able to swim eight-tenths of a mile in 50°F water before being overcome by hypothermia. Others have not been able to swim 100 yards. Furthermore, distances on the water are very deceptive. In most instances, staying with the raft is the best course of action. If the decision is made to swim, a life preserver or other flotation aid should be used. Shoes and at least one thickness of clothing should be worn. The side or breast stroke will help conserve strength.

(1) If surf is moderate the survivor can ride in on the back of a small wave by swimming forward with it and making a shallow dive to end the ride just before the wave breaks. The swimmer should stay in the trough between waves in high surf, facing the seaward wave and submerging when the wave approaches. After the wave passes, the swimmer should work shoreward in the next trough.

(2) If the swimmer is caught in the undertow of a large wave, push off the bottom and swim to the surface and proceed shoreward. A place where the waves rush up onto the rocks should be selected if it is necessary to land on rocky shores and avoid places where the waves explode with a high white spray. After selecting the landing point, the swimmer should advance behind a

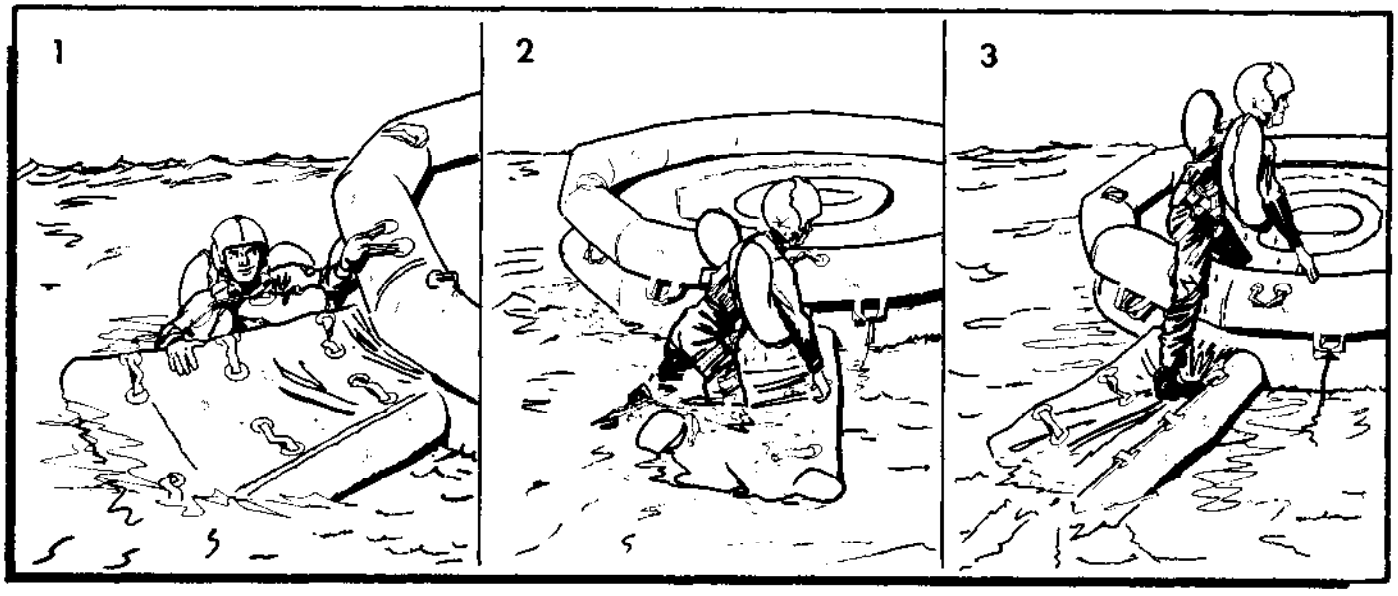

Figure 23-55. Boarding 20-Man Raft.

large wave into the breakers. The swimmer should face shoreward and take a sitting position with the feet in front, 2 or 3 feet lower than the head, so the knees are bent and the feet will absorb shocks when landing or striking submerged boulders or reefs. If the shore is not reached the first time, the survivor should swim with hands and arms only. As the next wave approaches, the sitting position with the feet forward should be repeated until a landing is made.

(3) Water is quieter in the lee of a heavy growth of seaweed. This growth can be very helpful. The swimmer should crawl over the top by grasping the vegetation with overhand movements.

(4) A rocky reef should be crossed in the same way as landing on a rocky shore. The feet should be close together with knees slightly bent in a relaxed sitting posture to cushion blows against coral.

b. Rafting Ashore. In most cases, the one-man raft can be used to make a shore landing with no danger. Going ashore in a strong surf is dangerous. The time should be taken to sail around and look for a sloping beach where the surf is gentle. The landing point should be carefully selected. Landing when the Sun is low and straight in front is not recommended. The survivor

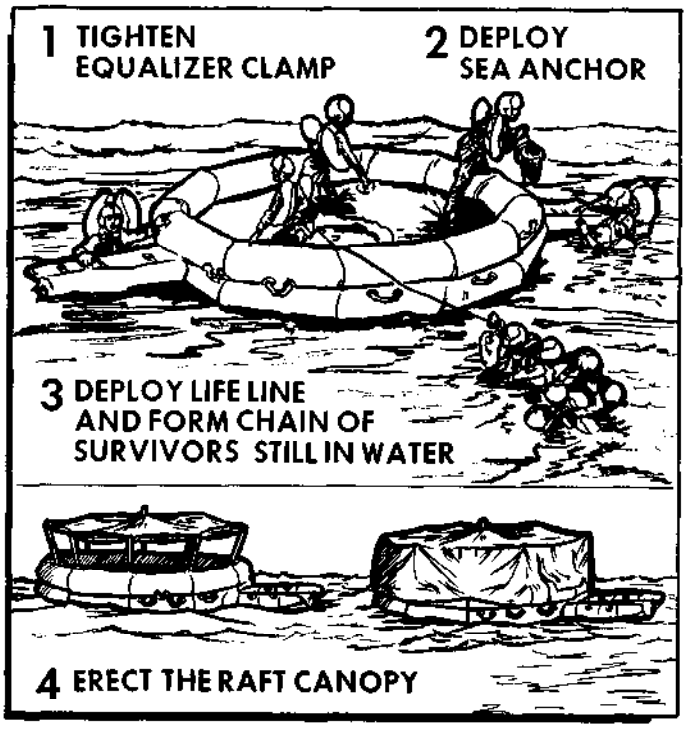

Figure 23-56. Immediate Action—Multiplace Raft.

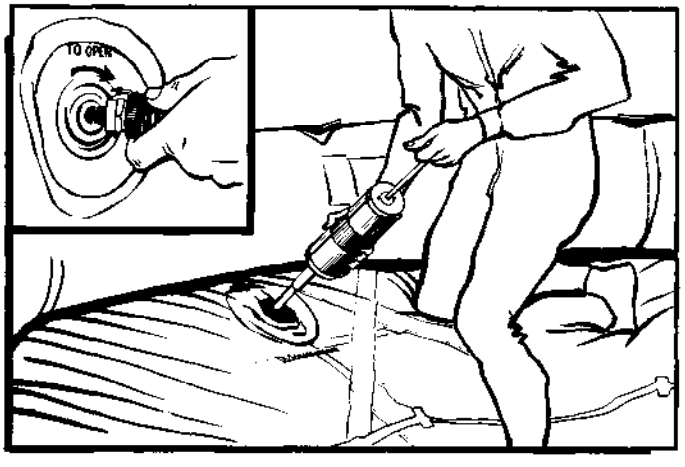

Figure 23-57. Inflating the 20-Man Raft.

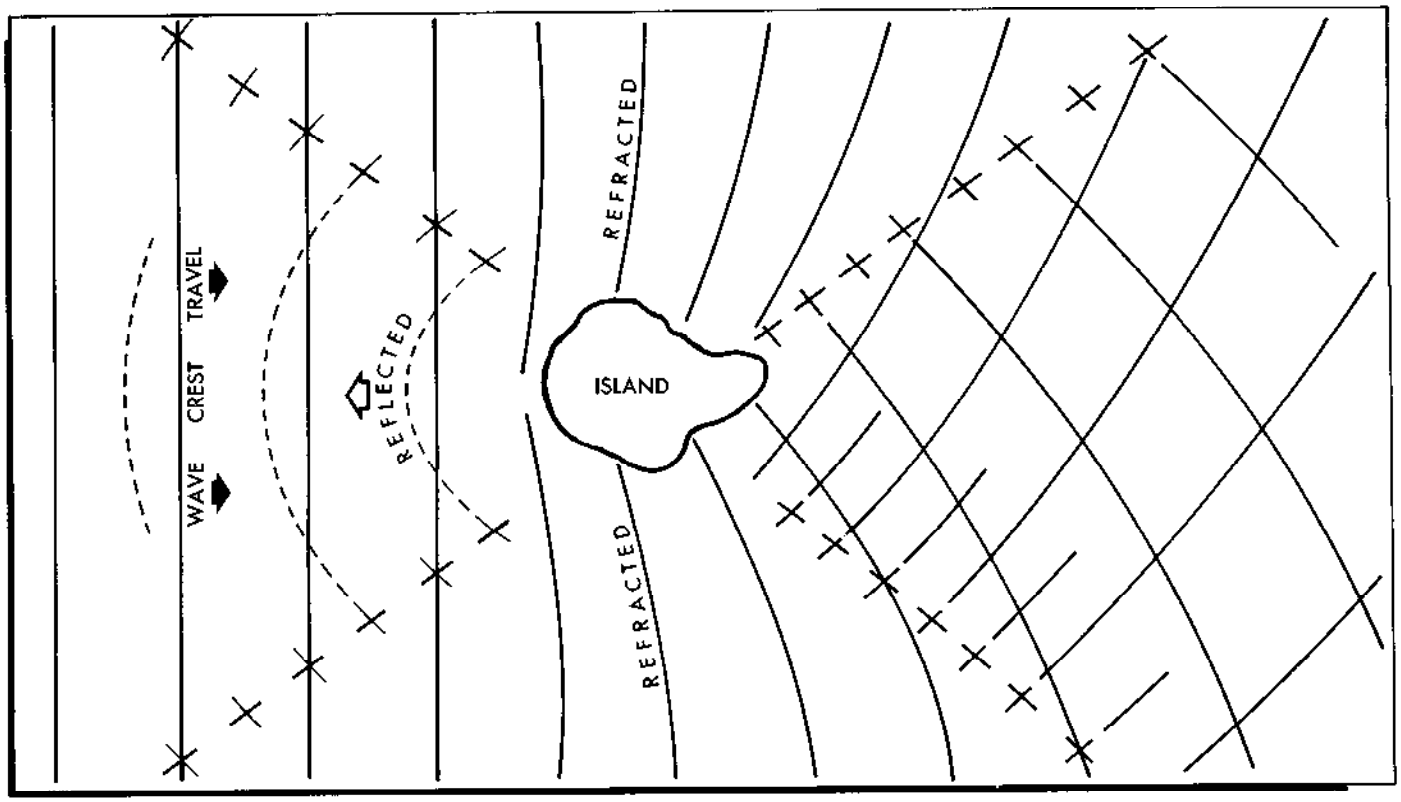

Figure 23-58. Diagram of Wave Patterns About an Island.

should look for gaps in the surf line and head for them while avoiding coral reefs and rocky cliffs. These reefs don't occur near the mouths of freshwater streams. Avoid rip currents or strong tidal currents which may carry the raft far out to sea.

(1) When going through surf, the survivor should:

(a) Take down the mast.
(b) Don clothing and shoes to avoid injuries.
(c) Adjust and fasten life preserver.
(d) Stow equipment.
(e) Use paddles to maintain control.

(f) Ensure the sea anchor is deployed to help prevent the sea from throwing the stern of the raft around and capsizing it. CAUTION: The sea anchor should not be deployed when traveling through coral.

(2) In medium surf with no wind, survivors should keep the raft from passing over a wave so rapidly that it drops suddenly after topping the crest. If the raft turns over in the surf, every effort should be made to grab hold.

(3) The survivor should ride the crest of a large wave as the raft nears the beach, staying inside until it has grounded. If there is a choice, a night landing should not be attempted. If signs of people are noted, it might be advantageous to wait for assistance.

(4) Sea-ice landings should be made on large stable floes only. Icebergs, small floes, and disintegrating floes could cause serious problems. The edge of the ice can cut, and the raft deflated. Use paddles and hands to keep the raft away from the sharp edges of the iceberg. The raft should be stored a considerable distance from the ice edge. It should be fully inflated and ready for use in case the floe breaks up.

Part Eight

SIGNALING AND RECOVERY

Chapter 24

SIGNALING

24-1. Introduction:

a. Most successful recoveries have resulted primarily because survivors were able to *assist* in their own recovery. Many rescue efforts failed because survivors lacked the knowledge and ability necessary to assist. When needed, this knowledge and ability could have made the difference between life or death—freedom or captivity (figure 24-1).

Figure 24-1. Signaling and Recovery.

b. What can survivors do to assist in their own recovery? First, they need to know what is being done to find them. Next, they need to know how to operate the communications equipment in the survival kit and when to put each item into use. Survivors should also be able to improvise signals to improve their chances of being sighted and to supplement the issued equipment.

c. It is not easy to spot one survivor, a group of survivors, or even an aircraft from the air, especially when visibility is limited. Emergency signaling equipment is designed to make a person easier to find. Emergency equipment may be used to provide rescue personnel with information about survivors' conditions, plans, position, or the availability of a rescue site where recovery vehicles might reach them (figure 24-2).

d. A part of a survivor's plan of action should be to visualize how emergencies will develop, recognize them, and, at the appropriate time, let friendly forces know about the problem. The length of time before survivors are rescued often depends on the effectiveness of emergency signals and the speed with which they can be used. Signal sites should be carefully selected. These sites should enhance the signal and have natural or manufactured materials readily available for immediate use. Survivors should avoid using pyrotechnic signals wastefully as they may be needed to enhance rescue efforts. Signals used correctly can hasten recovery and eliminate the possibility of a long, hard survival episode. Survivors should:

(1) Know how to use their emergency signals.

(2) Know when to use their signal.

Figure 24-2. Signaling.

(3) Be able to use their signals on short notice.

(4) Use signals in a manner which will not jeopardize individual safety.

e. The situation on the ground governs the type of information which survivors can furnish the rescue team, and will govern the type of signaling they should use. In nontactical survival situations, there are no limitations on the ways and means survivors may use to furnish information.

f. In hostile areas, limitations on the use of signals should be expected. The use of some signaling devices will pinpoint survivors to the enemy as well as to friendly personnel. Remember the signal enhances the visibility of the survivors.

24-2. Manufactured Signals:

a. Electronic Signals:

(1) Current line-of-sight electronic signaling devices fall into two categories. One is the transceiver type; the other is the personal locator beacon type. The transceiver type is equipped for transmitting tone or voice and receiving tone or voice. The personal locator beacon is equipped to transmit tone only. The ranges of the different radios vary depending on the altitude of the receiving aircraft, terrain factors, forest density, weather, battery strength, type(s) of radios and interference. Interference is a very important aspect of the use of these radios. If a personal locator beacon is transmitting, it will interfere with incoming and outgoing signals of the transceivers.

(2) Before using survival radios, a few basic precautions should be observed. These will help in obtaining maximum performance from the radios in survival situations.

(a) The survival radios are line-of-sight communication devices; therefore, the best transmission range will be obtained when operating in clear, unobstructed terrain.

(b) Extending from the top and bottom of the radio antenna is an area referred to as the "cone of silence." To avoid the "cone of silence" problem, keep the radio/beacon antenna orthogonal to (at a right angle to) the path of the rescue aircraft.

(c) Since the radios have the capability of transmitting a tone (beacon) without being hand-held, they can be placed upright on a flat elevated surface allowing the operator to perform other tasks.

(d) Never allow the radio antenna to ground itself on clothing, body, foliage or the ground. This will severely decrease the effective range of the signals.

(e) Conserve battery power by turning the radio off when not in use. Do not transmit or receive constantly. Use the locator beacon to supplement the radio when transmitting is done. In tactical environments, the radio should be used as stated in the premission briefing.

(f) Survival radios are designed to operate in extreme heat or cold. The life expectancy of a battery decreases as the temperature drops below freezing and exposure to extreme heat or shorting out of the battery can cause an explosion. During cold weather, the battery should be kept warm by placing it between the layers of clothing to absorb body heat, or wrapped in some type of protective material when it is not being used.

(g) Survival radios are designed to be waterproof. However, precautions should be taken to keep them out of water.

(3) Presently, a satellite monitoring system has been developed to assist in locating survivors. To activate this system (SARSAT), the transmitter is "keyed" for a minimum of 30 seconds. In a nontactical situation, leave the beacon on until rescue is heard or sighted.

b. Pyrotechnics. Care should be used when operating around flammable materials.

(1) A device containing chemicals for producing smoke or light is known as a pyrotechnic. Hand-held flares are in this category. Survivors may be required to use a variety of flares. They must know the types of flares stored in their survival kits and (or) aircraft. Aircrew members should learn how to use each type of flare before they face an emergency. Flares are designed to be used during the day or night. Day flares produce a unique bright-colored smoke which stands out very clearly against most backgrounds. Night flares are extremely bright and may be seen for miles by air, ground, or naval recovery forces.

(2) The hand-held launched flares also fall in the pyrotechnic category. They were designed to overcome the problems of terrain masking and climatic conditions. For example, a person may be faced with multilayer vegetation or atmospheric conditions known as an inversion which keeps the smoke next to the ground.

(3) Flares must be fired at the right time to be of maximum use. Smoke flares, for example, take a second or two after activation before they produce a full volume of smoke. Therefore, the flare should be ignited just before the time it can be seen by rescue personnel. These signals should not be used in tactical environments unless directed to do so.

(4) Tracer ammunition is another pyrotechnic which may be issued to aircrew members. When fired, the projectile appears as an orange-red flash the size of a golf ball. According to specifications, these tracers have a range of 1,300 feet. Tracer bullets have been detected from a distance of 6 miles, but there is usually difficulty in pinpointing the survivor. A survivor should only use

Figure 24-3. Sea Marker Dye.

this signaling device when rescue forces can be seen or heard. Do not direct this device *at* the aircraft.

(5) Because of the rapid changing technology in pyrotechnic signaling devices, an aircrew member should check regularly for new and improved models, making special note of the firing procedures and safety precautions necessary for their operations.

c. Sea Marker:

(1) Of the many dyes and metallic powders tested at various times for marking the sea, the most successful is the fluorescent, water-soluble, orange powder. When released in the sea, a highly visible, light green, fluorescent cast is produced. Sea marker dye has rapid dispersion power; a packet spreads into a slick about 150 feet in diameter and lasts an hour or more in calm weather. Rough seas will stream it into a long streak, which may disperse in 20 minutes (figure 24-3).

(2) Under ideal weather conditions, the dye can be sighted at 5 miles with the aircraft operating at 1,000 feet. The dye has also been spotted at 7 miles away from an aircraft operating at 2,000 feet.

(3) Sea marker dye should be used in friendly areas during daytime and only when there is a chance of being sighted (aircraft seen or heard in the immediate area). It is not effective in heavy fog, solid overcast and storms with high winds and waves. The release tab on the packet of dye is pulled to open for use. In calm water, the dye can be dispersed more rapidly stirring the water with paddles or hands.

(4) If left open in the raft, the escaping powder penetrates clothing, stains hands, face, and hair, and eventually may contaminate food and water. To avoid the inevitable messiness, some survivors have tied the sea marker dye to the sea anchor. Others have dipped the packet over the side, letting it drain off the side into the sea. After using the dye, it should be rewrapped to conserve the remainder of the packet.

d. Paulin Signals. The paulin is a conventional signaling device used to send specific messages to aircraft. It may be packed with some sustenance kits and multiplace liferaft accessory kits. The paulin is constructed of rubberized nylon material and is blue on one side and yellow on the other. These colors contrast against each

other so when one side is folded over the other, the designs are easily distinguished (figure 24-4). The size is 7 feet by 11 feet which is a disadvantage when folded because it makes a small signal. The paulin has numerous uses. It can be used as a camouflage cloth, sunshade, tent, or sail, or it can be used to catch drinking water. The space blanket, used as a substitute for the sleeping bag in some survival kits, can be used in the same manner as the signal paulin because it is highly reflective (silver on one side and various colors on the other side).

e. Audio Signals. Sounds carry far over water under ideal conditions; however, they are easily distorted and deadened by the wind, rain, or snow. On land, heavy foliage cuts down on the distance sound will travel. Shouting and whistling signals have been effective at short ranges for summoning rescue forces. Most contacts using these means were made at less than 200 yards, although a few reports claim success at ranges of up to a mile. A weapon can be used to attract attention by firing shots in a series of three. The number of available rounds determine whether this is practical. Survivors have used a multitude of devices to produce sound. Some examples are: striking two poles together, striking one pole against a hollow tree or log, and improvising whistles out of wood, metal, and grass.

f. Light Signals. When tested away from other manufactured lights, aircraft lights have been seen up to 85 miles. At night, a survivor should use any type of light to attract attention. A signal with a flashlight, or a light or fire in a parachute shelter, can be seen from a long distance. A flashing light (strobelight) is in most survival kits.

g. Signal Mirror.

(1) The signal mirror is probably the most underrated signaling device found in the survival kit. It is the most valuable daytime means of visual signaling. A mirror flash has been visible up to 100 miles under ideal conditions, but its value is significantly decreased unless it is used correctly. It also works on overcast days. Practice is the key to effective use of the signal mirror. Whether the mirror is factory manufactured or improvised, aim it so the beam of light reflected from its surface hits the overflying aircraft.

(2) The signal mirror's effectiveness is its greatest weakness if the survivor is in enemy territory. It is just as bright to the enemy as to the rescuer; use it wisely! Survivors should understand that even if the mirror flash is directly on the aircraft (especially if the aircraft is using terrain masking techniques), that same flash may be visible to others (possibly the enemy) who are located at the proper angle in regard to the survivors' position.

(3) In a hostile environment, the exact location of the flash is extremely important. The signal mirror should be covered when not in use. One of the easiest

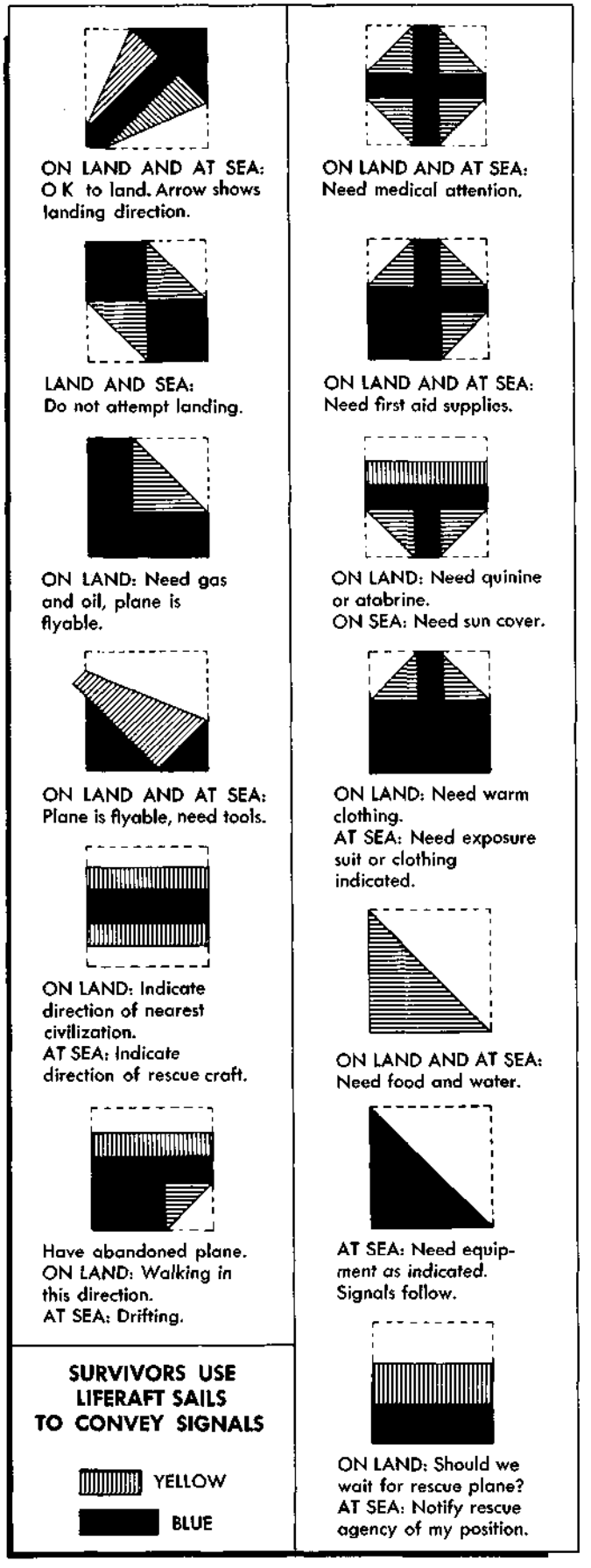

Figure 24-4. Paulin Signals.

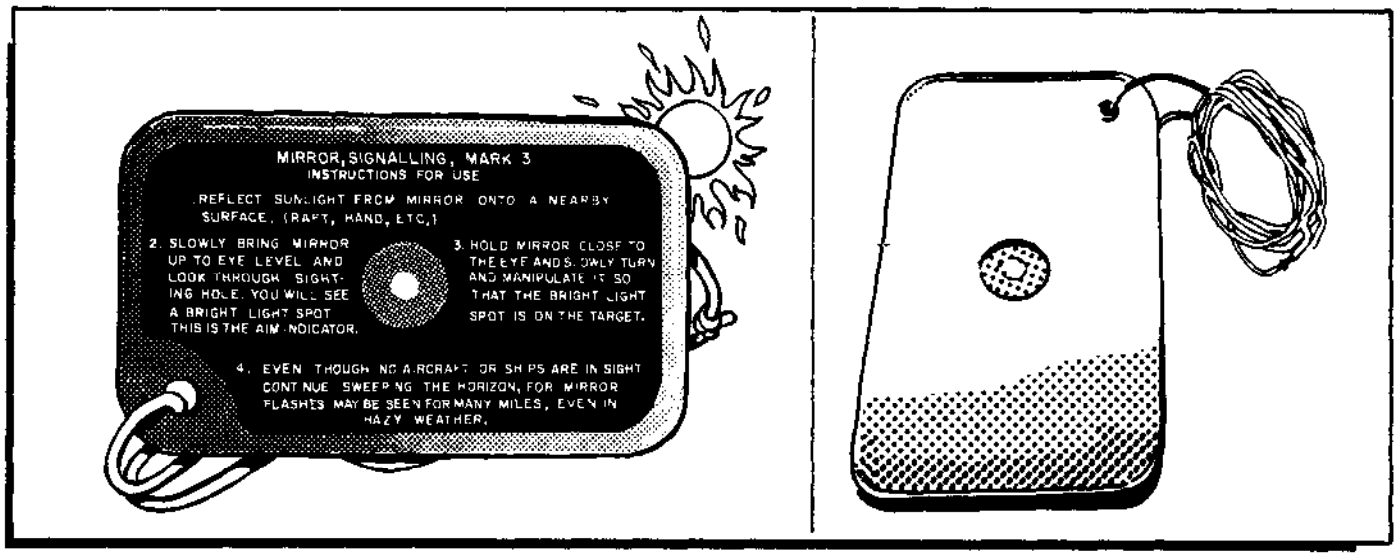

Figure 24-5. Signal Mirrors.

methods is to tie the string from the mirror around the neck and tuck the mirror in the shirt or flight suit. When the mirror is removed from inside the clothing, the hand should be placed over the mirror surface to prevent accidental flashing. The covered mirror may then be raised toward the sky and the hand withdrawn. The flash can then be directed onto the free hand and the aiming indicator (sunspot) located. This minimizes the indiscriminate flashing of surrounding terrain. When putting the mirror away, the survivor should remember to cover the mirror to prevent a flash.

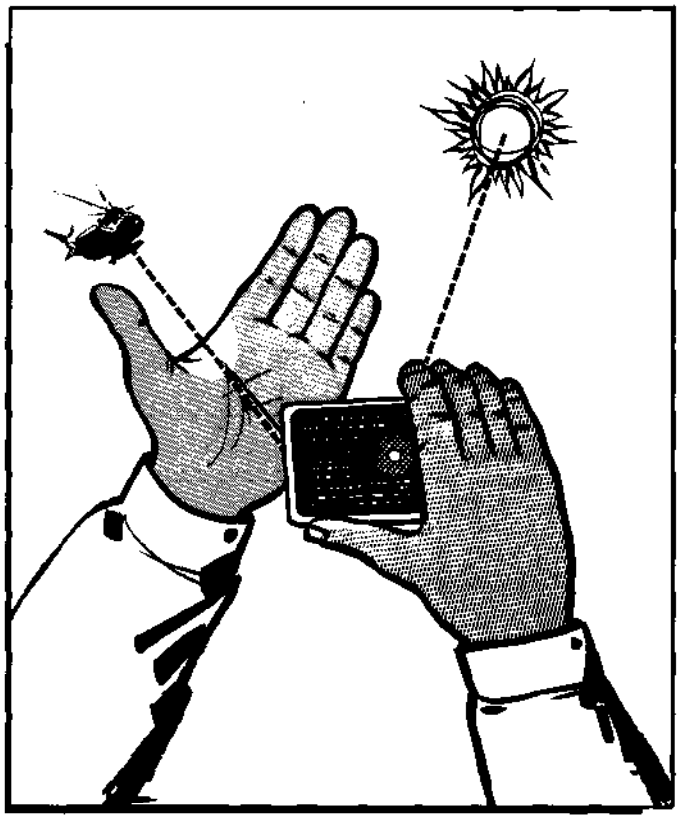

Figure 24-6. Aiming Signal Mirror.

h. Aiming Manufactured Mirrors. Instructions are printed on the back of the mirror. Survivors should:

(1) Reflect sunlight from the mirror onto a nearby surface—raft, hand, etc.

(2) Slowly bring the mirror up to eye-level and look through the sighting hole where a bright spot of light will be seen. This is the aim indicator.

(3) Hold mirror near the eye and slowly turn and manipulate it so the bright spot of light is on the target.

(4) In friendly areas, where rescue by friendly forces is anticipated, free use of the mirror is recommended. Survivors should continue to sweep the horizon even though no aircraft or ships are in sight (figure 24-5).

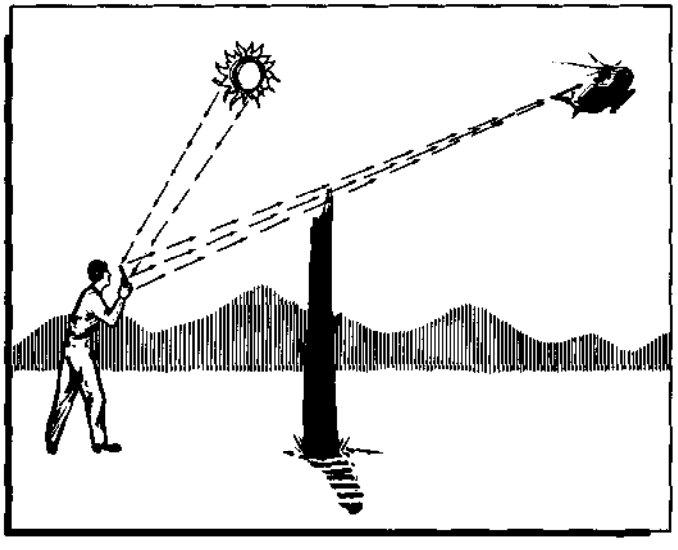

Figure 24-7. Aiming Signal Mirror-Stationary Object.

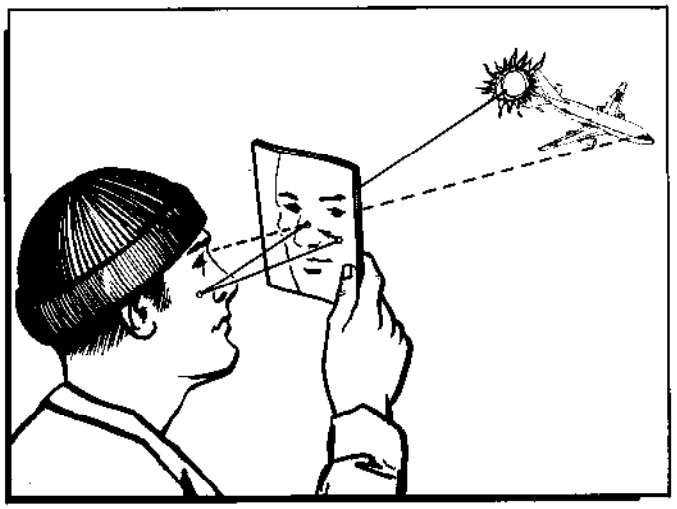

Figure 24-8. Aiming Signal Mirror—Double-Faced Mirror.

24-3. Improvised Signals:

a. Signal Mirrors. Improvised signal mirrors can be made from ration tins, parts from an aircraft, polished aluminum, glass, or the foil from rations or cigarette packs. However, the mirror must be accurately aimed if the reflection of the Sun in the mirror is to be seen by the pilot of a passing aircraft or the crew of a ship.

b. Aiming Improvised Mirrors:

(1) The simple way to aim an improvised mirror is to place one hand out in front of the mirror at arm's length and form a "V" with two fingers. With the target in the "V" the mirror can be manipulated so that the majority of light reflected passes through the "V" (figure 24-6). This method can be used with all mirrors. Another method is to use an aiming stake as shown in figure 24-7. Any object 4 to 5 feet high can serve as the point of reference.

(2) Survivors should hold the mirror so they can sight along its upper edge. Changing their position until the top of the stick and target line up, they should adjust the angle of the mirror until the beam of reflected light hits the top of the stick. If stick and target are then kept in the sighting line, the reflection will be visible to the rescue vehicle.

(3) Another method is to improvise a double-faced mirror (shiny on both sides). A sighting hole can be made in the center of the mirror.

(a) When trying to attract the attention of a friendly rescue vehicle that is no more than 90 degrees from the Sun, proceed as shown in figure 24-8.

(b) The survivor's first step will be to hold the double-faced mirror about 3 to 6 inches away from the face and sight at the rescue target through the hole in the center of the mirror. The light from the Sun shining through the hole will form a spot of light on the survivor's face. This spot will be reflected in the rear surface of the mirror. Then, aiming at the rescue vehicle through the hole, the survivor can adjust the angle of the mirror until the reflection of the spot on the face in the rear surface of the mirror lines up with, and disappears, into the sighting hole.

(c) When the reflected spot disappears and the rescue vehicle is still visible through the hole, the survivor can be sure the reflected light from the Sun is accurately aimed. The survivor may also "shimmer" the mirror by moving it rapidly over the target. This ensures that the part of the bright flash the rescuers see coincides with the position of the survivor. This (shimmering) is especially useful on a moving target.

(d) When the angle between the target and the Sun is more than 90 degrees (when the survivor is between the rescue vehicle and the Sun) a different method may be used for aiming. The survivor should adjust the angle of the mirror until the spot made by the

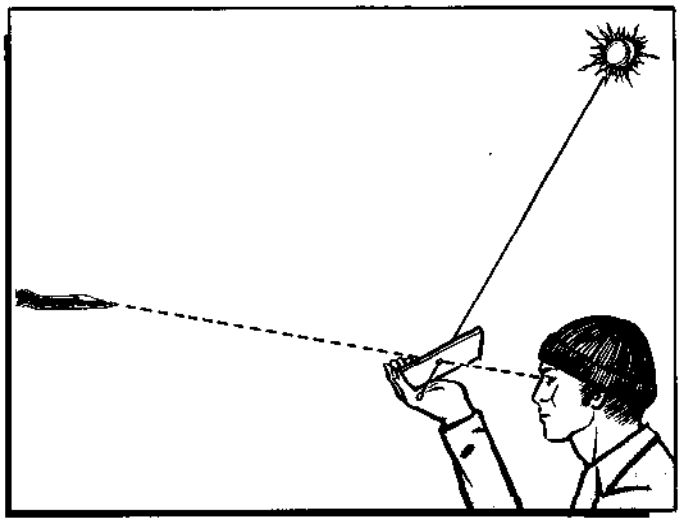

Figure 24-9. Aiming Signal Mirror—Angle Greater than 180 Degrees.

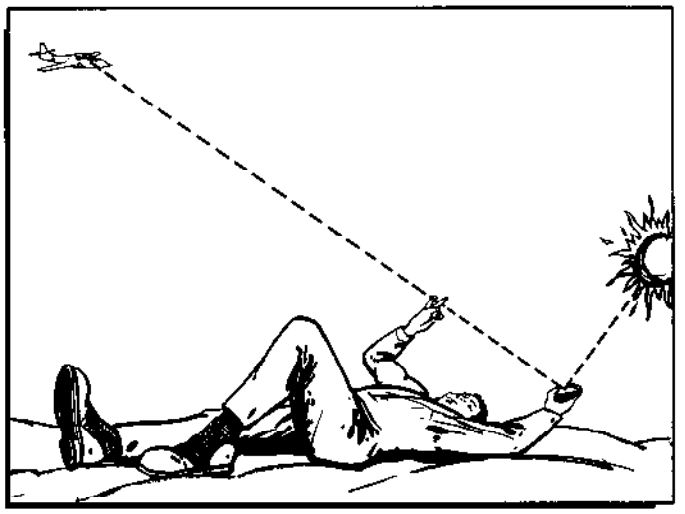

Figure 24-10. Aiming Signal Mirror—Angle Greater than 180 Degrees (Another Method).

Figure 24-11. Shelter as a Signal.

Sun's rays passing through the hole in the mirror lands on the hand instead of on the face. The reflection in the back of the mirror that comes off the hand may then be manipulated in the same way (figure 24-9).

(e) Another method used when the angle is greater than 90 degrees is to lie on the ground in a large clearing and aim the mirror using one of the methods previously discussed (figure 24-10).

c. Fire and Smoke Signals:

(1) Fire and smoke can be used to attract the attention of recovery forces. Three evenly spaced fires, 100 feet apart, arranged in a triangle or in a straight line, serve as an international distress signal. One signal fire will usually work for a survivor. During the night, the flames should be as bright as possible, and during the day, as much smoke as possible should be produced.

(2) Smoke signals are most effective on clear and calm days. They have been sighted from up to 50 miles away. High winds, rain, or snow tend to disperse the smoke and lessen the chances of it being seen. Smoke signals are not dependable when used in heavily wooded areas.

(3) The smoke produced should contrast with its background. Against snow, dark smoke is most effective. Likewise, against a dark background, white smoke is best. Smoke can be darkened by rags soaked in oil, pieces of rubber, matting, or electrical insulation, or plastic being added to the fire. Green leaves, moss, ferns, or water produce white smoke.

(4) To increase its effectiveness, the signal fire must be prepared before the recovery vehicle enters the area.

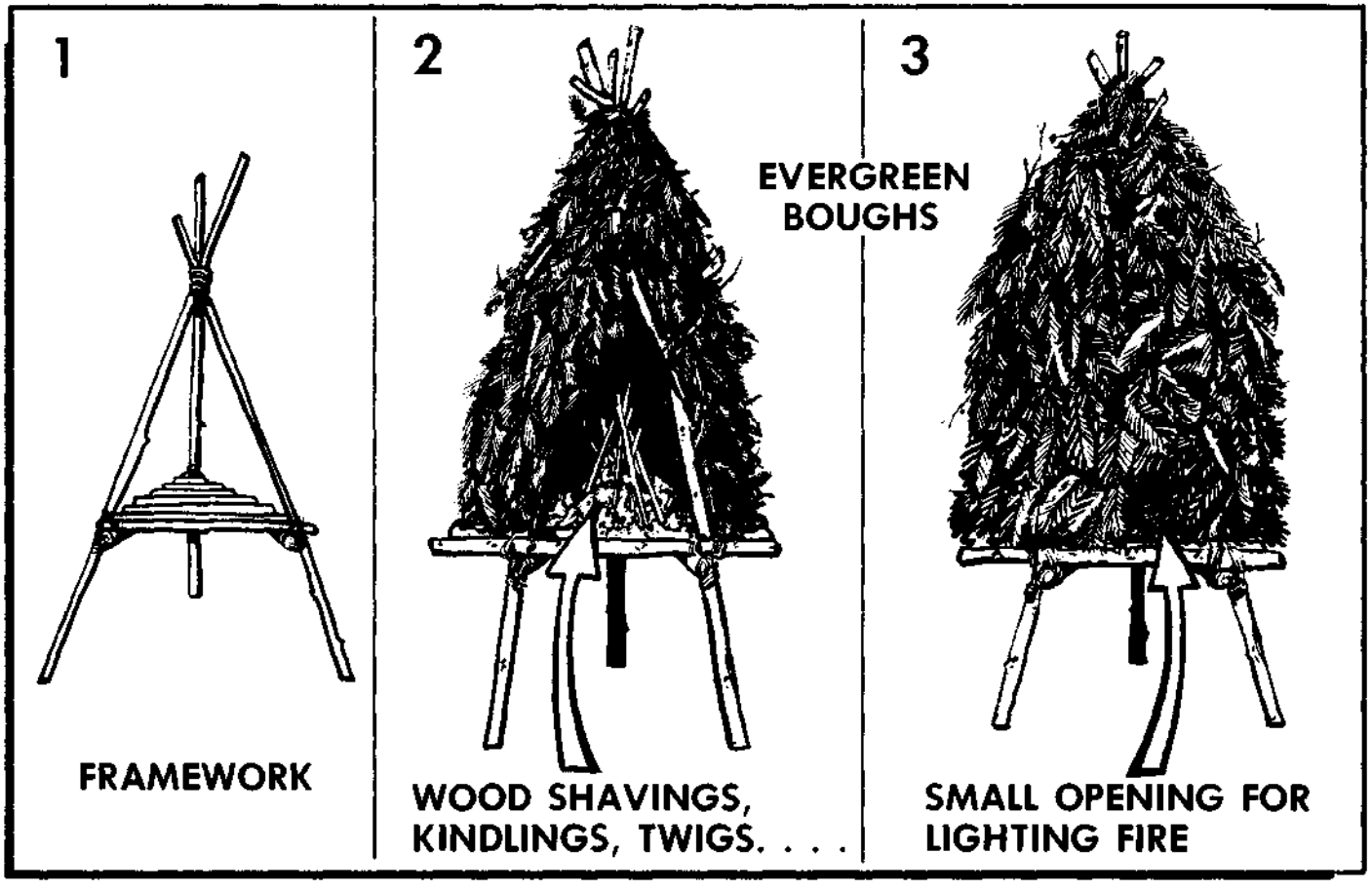

Figure 24-12. Smoke Generator—Platform.

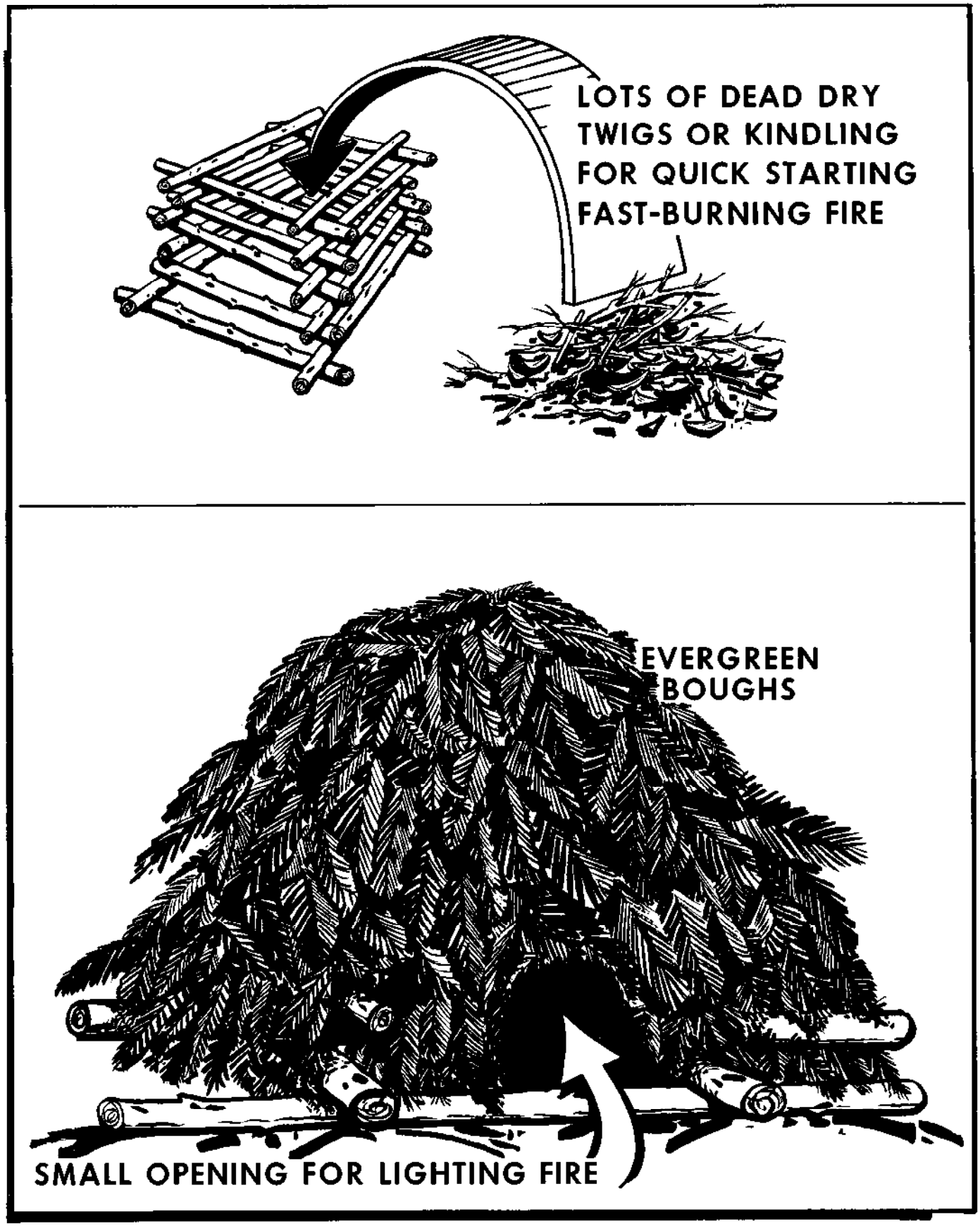

Figure 24-13. Smoke Generator—Ground.

The fires used by survivors for heat and cooking may be used as signal fires as long as the necessary materials are available in the immediate vicinity. Survivors should supplement the fire to provide the desired signal (figure 24-11).

(5) Smoke Generators:

(a) Raised Platform Generator (figure 24-12). The survivor should:

-1. Build a raised platform above wet ground or snow.

-2. Place highly combustible materials on the platform.

-3. Then place smoke-producing materials over the platform and light when search aircraft is in the immediate vicinity.

(b) Ground Smoke Generator (figure 24-13). The survivor should:

-1. Build a large log cabin fire configuration on the ground. This provides good ventilation and supports the green boughs used for producing smoke.

-2. Place smoke-producing materials over the fire lay; ignite when a search aircraft is in the immediate vicinity.

(c) Tree Torch Smoke Generator (figure 24-14). To build this device, the survivor should:

-1. Locate a tree in a clearing to prevent a forest fire hazard.

-2. Add additional smoke-producing materials.

-3. Add igniter.

-4. Light when a search aircraft is in the immediate vicinity.

(d) Fuel Smoke Generator. If survivors are with the aircraft, they can improvise a generator by burning aircraft fuels, lubricating oil, or a mixture of both. One to 2 inches of sand or fine gravel should be placed in the bottom of a container and saturated with fuel. Care should be used when lighting the fuel as an explosion may occur initially. If there is no container available, a hole can be dug in the ground, filled with sand or gravel, saturated with fuel, and ignited. Care should be taken to protect the hands and face.

d. Pattern Signals. The construction and use of pattern signals must take many factors into account. Size, ratio, angularity, contrast, location, and meaning are each important if the survivors' signals are to be effec-

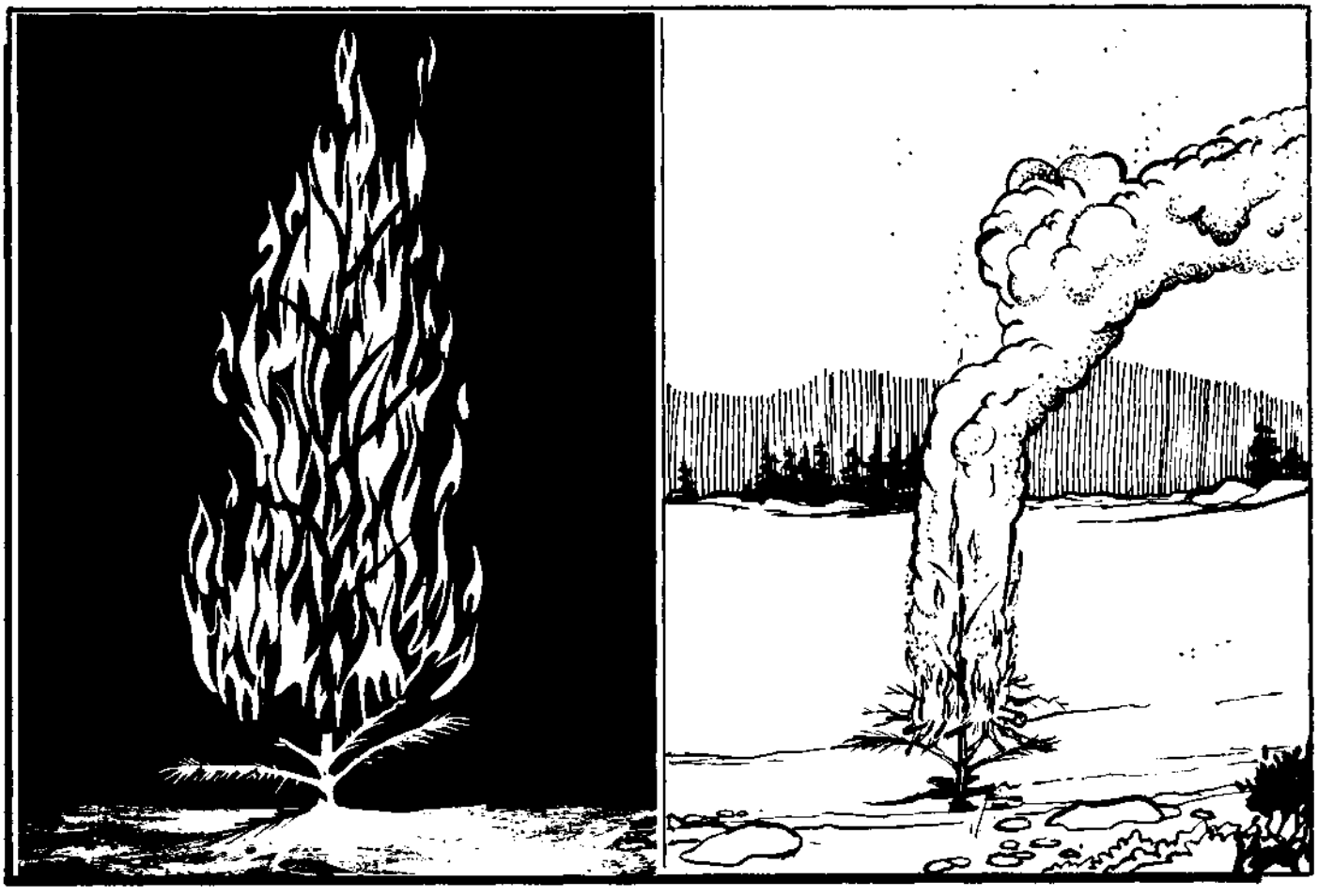

Figure 24-14. Tree Torch.

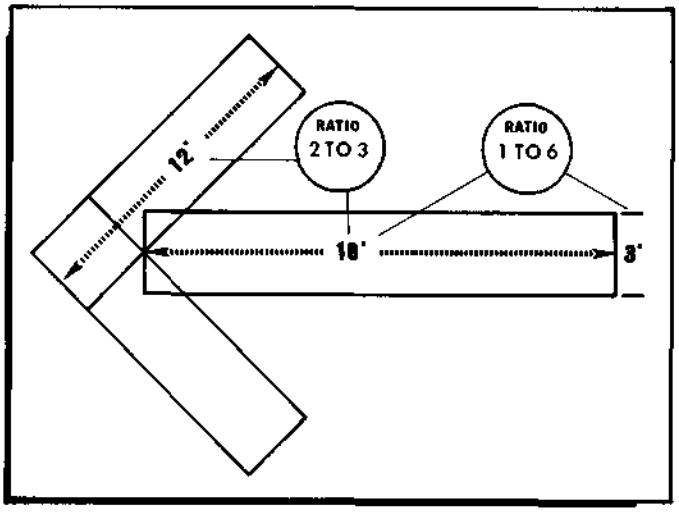

Figure 24-15. Pattern Signal Sizes.

tive. The type of signal constructed will depend on the material available to survivors. Not every crewmember will have a parachute, so ingenuity plays an important role in the construction of the signal. Survivors should remember to judge their signals from the standpoint of aircrew members who are flying over their location searching for them.

(1) Size. The signal should be as large as possible. To be most effective, the signal should have "lines" no less than 3 feet wide and 18 feet long (1:6) (figure 24-15).

(2) Ratio. Proper proportion should also be remembered. For example, if the baseline of an "L" is 18 feet long, then the vertical line of the "L" must be longer (27 feet), a 2 to 3 ratio, to keep the letter in proper proportion.

(3) Angularity. Straight lines and square corners are not found in nature. For this reason, survivors should make all pattern signals with straight lines and square corners.

(4) Contrast. The signal should stand out sharply against the background. The idea is to make the signal look "larger." On snow, the fluorescent sea dye available in the liferaft accessory kit can be used to add contrast around the signal. The survivor should do everything

Figure 24-16. Contrast (Snow).

Figure 24-17. Contrast.

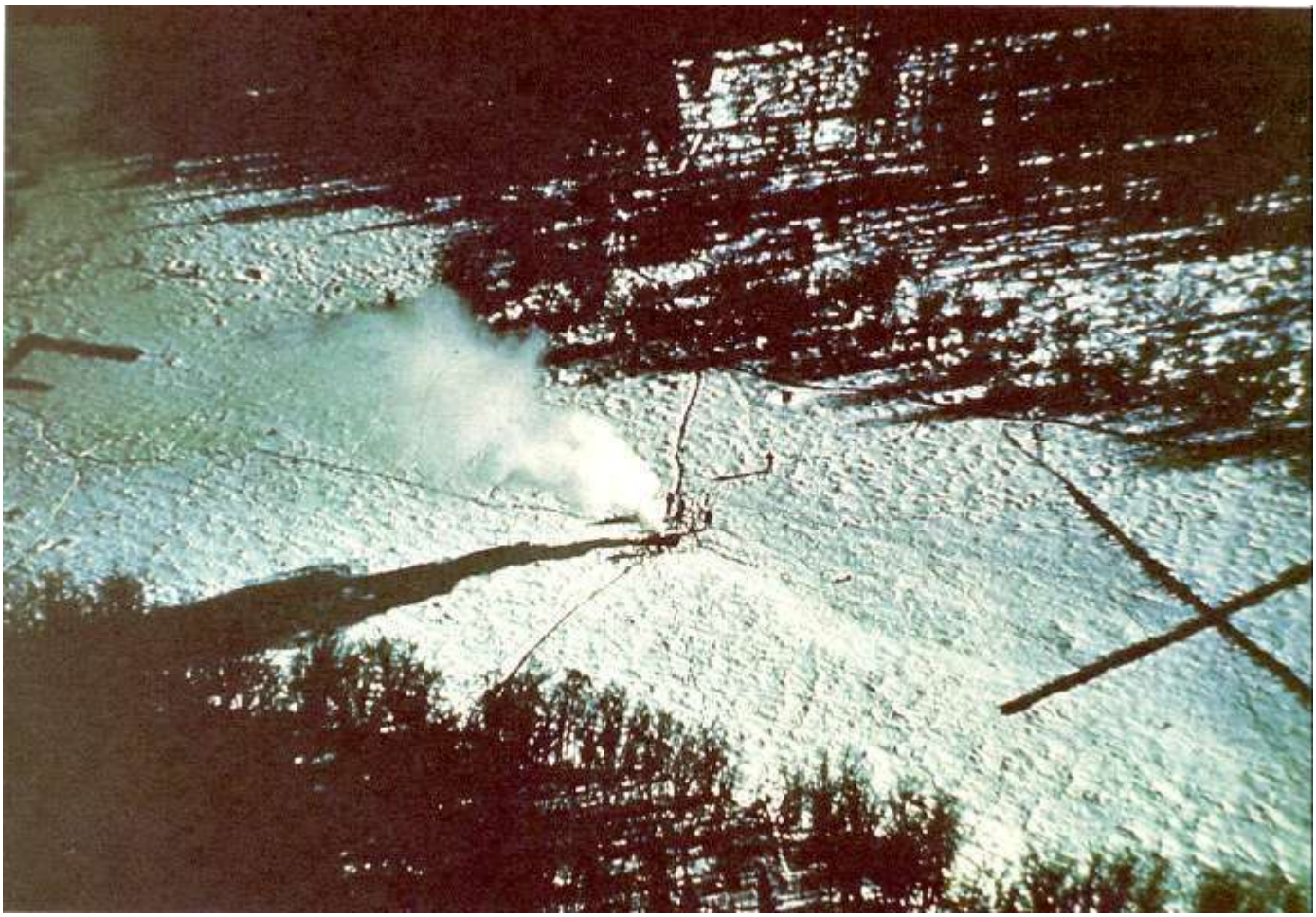

Figure 24-18. Location.

NO.	MESSAGE	CODE SYMBOL
1	REQUIRE ASSISTANCE	V
2	REQUIRE MEDICAL ASSISTANCE	X
3	NO or NEGATIVE	N
4	YES or AFFIRMATIVE	Y
5	PROCEEDING IN THIS DIRECTION	↑

Figure 24-19. Signal Key.

possible to disturb the natural look of the ground. In grass and scrubland, the grass should be stamped down or turned over to allow the signal to be easily seen from the air. A burned grass pattern is also effective. When in snow, a trampled out signal is very effective. Survivors should use only one path to and from the signal to avoid

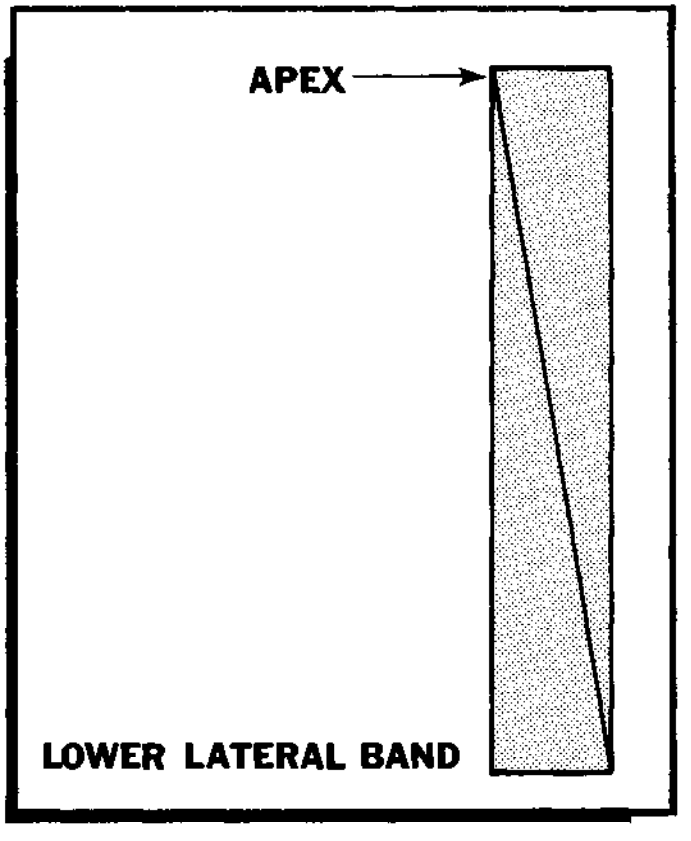

Figure 24-20. Parachute Strips.

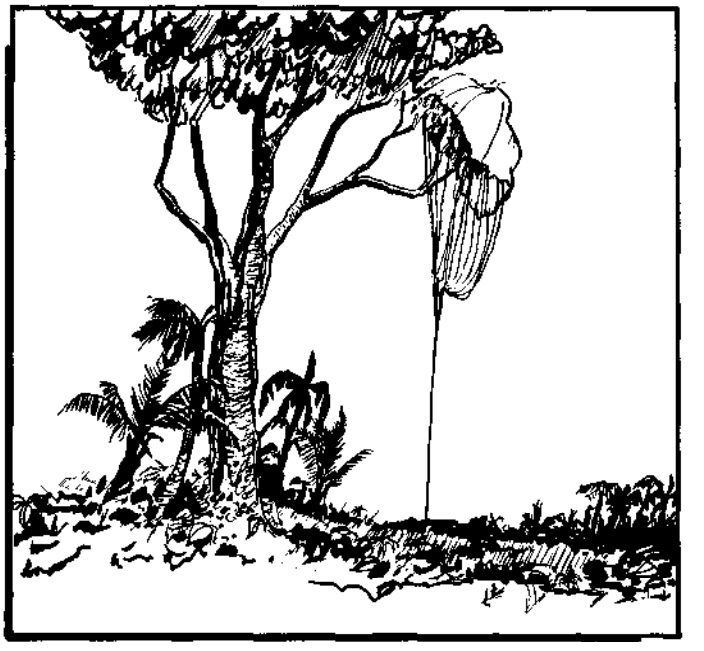

Figure 24-21. Parachute in Tree.

disrupting the signal pattern. Avoid using orange parachute material on a green or brown background as it has a tendency to blend in (figure 24-16). Contrast can be improved by outlining the signal with green boughs, piling brush and rocks to produce shadows, or raising the panel on sticks to cast its own shadow (figure 24-17).

(5) Location. The signal should be located so it can be seen from all directions. Survivors should make sure the signal is located away from shadows and overhangs. A large high open area is preferable. It can serve a dual function—one for signaling and the other for rescue aircraft to land (figure 24-18).

(6) Meaning. If possible, the signal should tell the rescue forces something pertaining to the situation. For example: "require medical assistance," or a coded symbol used during evasion, etc. Figure 24-19 shows the internationally accepted symbols.

e. Parachute Signals:

(1) Parachute material can be used effectively to construct pattern signals. A rectangular section of parachute material can be formed as shown in figure 24-20. When making a pattern signal, survivors should ensure the edges are staked down so the wind will not blow the panels away.

(2) A parachute caught in a tree will also serve as a signal. Survivors should try to spread the material over the tree to provide the maximum amount of signal (figure 24-21).

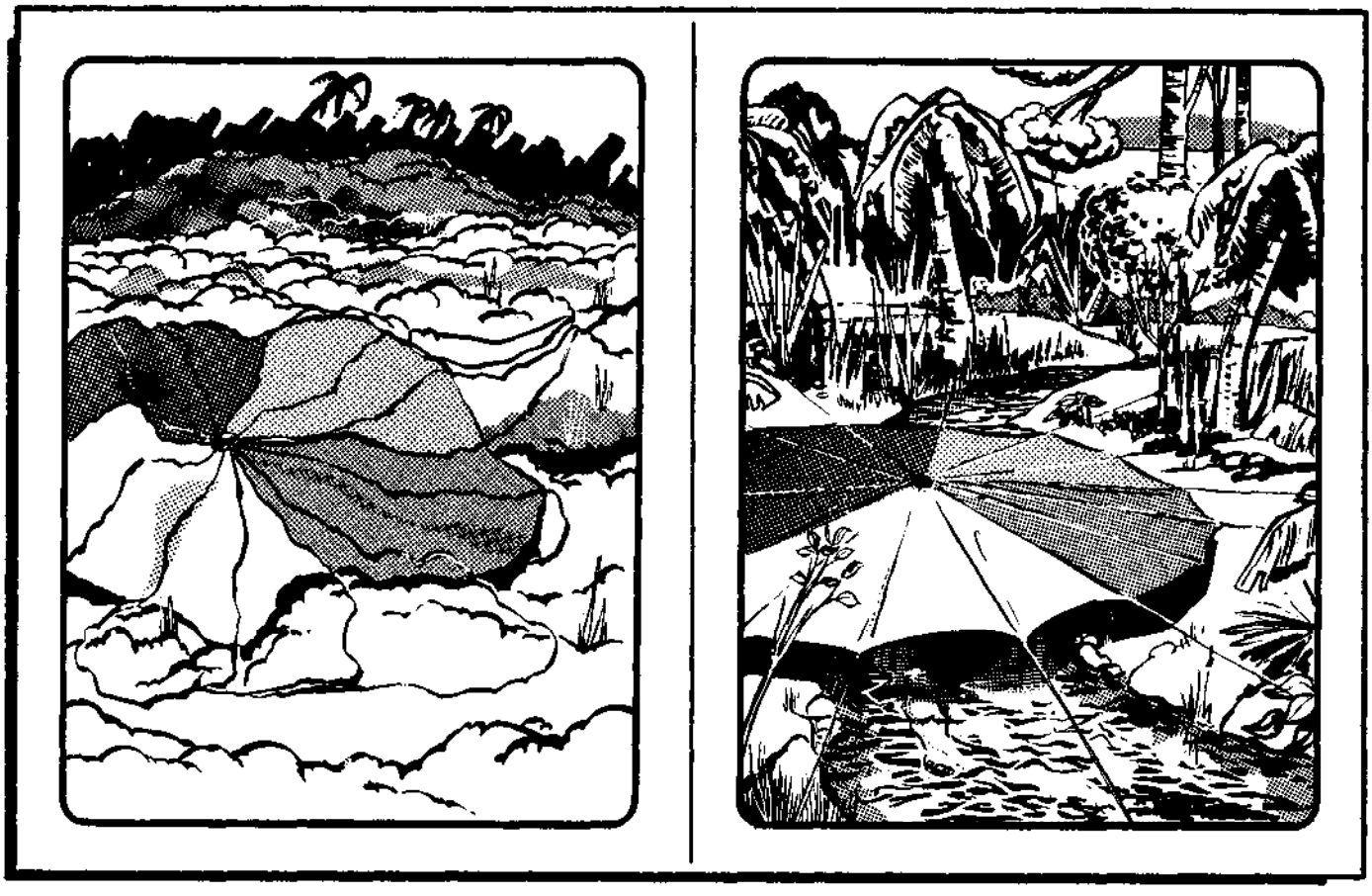

Figure 24-22. Chute Over Trees or Streams.

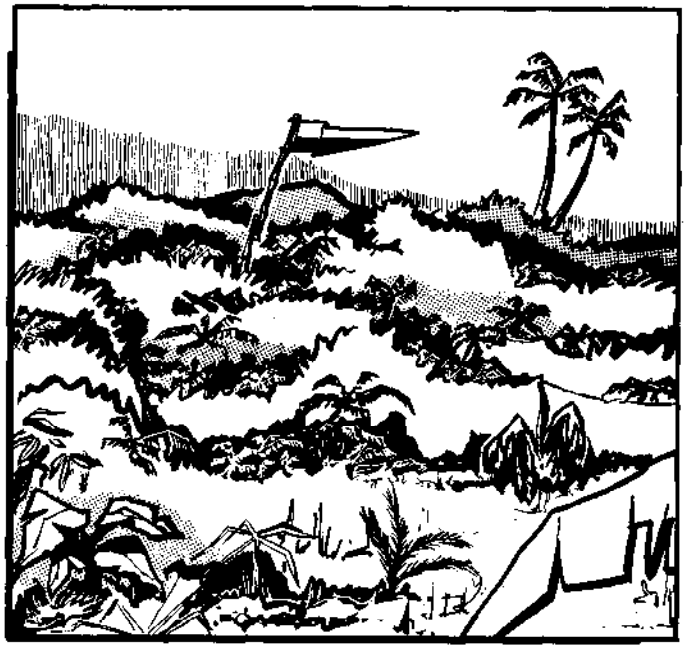

Figure 24-23. Pennants and Banners.

(3) When open areas are not available, survivors should stretch the chute over low trees and brush or across small streams (figures 24-22 and 24-23).

f. Shadow Signals. If no other means are available, survivors may have to construct mounds which will use the Sun to cast shadows. These mounds should be constructed in one of the international distress patterns. Brush, foliage, rocks, or snowblocks may be used to cast shadows. To be effective, these shadow signals must be oriented to the Sun to produce the best shadow. In areas close to the Equator, a north-south line gives a shadow at any time except noon. Areas farther north or south require the use of an east-west line or some point of the compass in between to give the best results.

g. Acknowledgements:

(1) Rescue personnel will normally inform the survivors they have been sighted by:

(a) Flying low with landing lights on (figure 24-24), and (or) rocking the wings.

(b) Emergency radio.

(2) Figure 24-25 depicts the standard body signals which can be used if electronic signaling devices are not available.

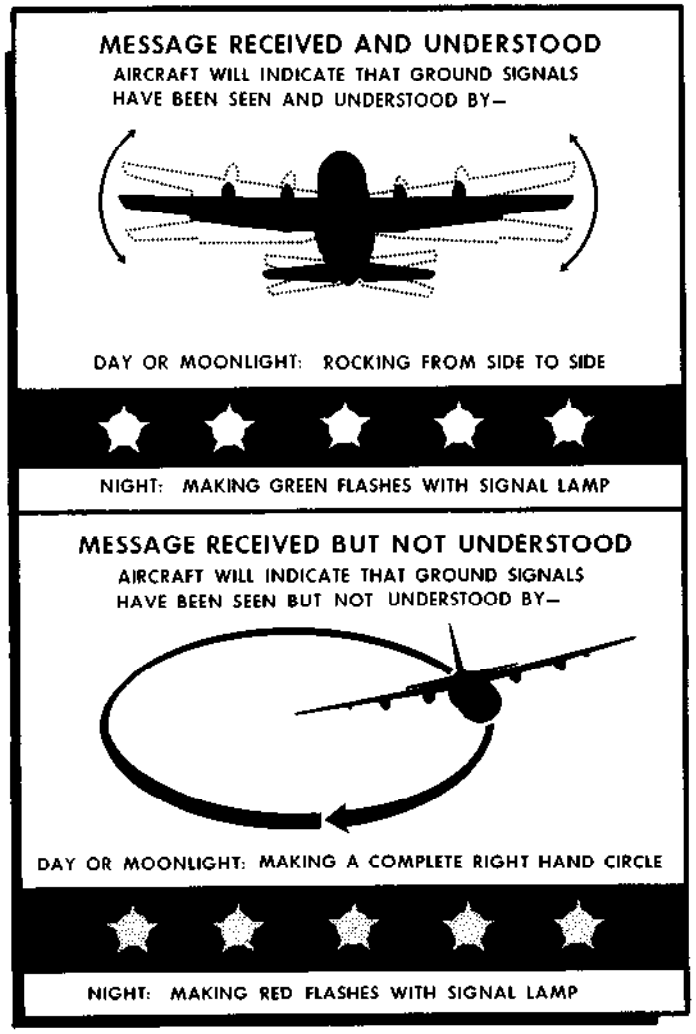

Figure 24-24. Standard Aircraft Acknowledgements.

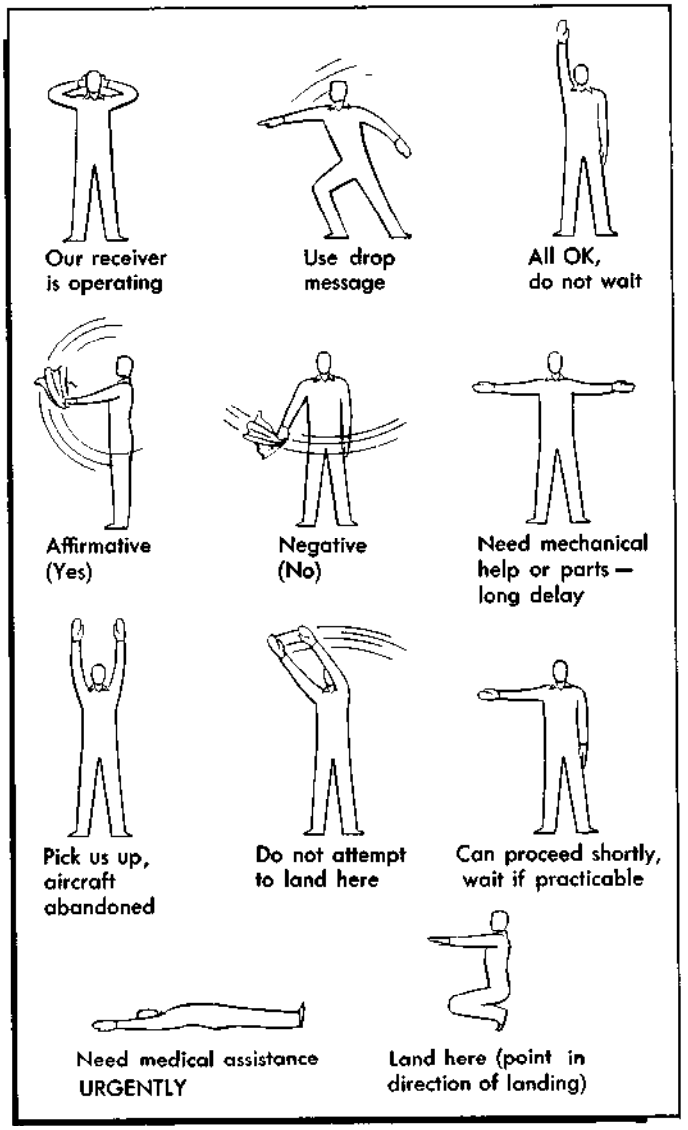

Figure 24-25. Close-In Visual Signals.

Chapter 25

RECOVERY PRINCIPLES

25-1. Introduction:

a. Receipt of a distress call sets a highly trained and well-equipped organization into operation; however, prompt and safe recovery is by no means ensured. The success of the rescue effort depends on many factors. Such factors as the availability of rescue forces, the proximity of enemy forces, and weather conditions can affect the success of the rescue. Above all, the survivors' knowledge of what to do in the rescue effort may make the difference between success and failure (figure 25-1).

b. The role of survivors in effecting their rescue changes continuously as aircraft and rescue equipment become more sophisticated. The probability of a downed aircrew member applying long-term survival training concepts under noncombat conditions continues to decrease while increasing under combat conditions.

c. There are several independent organizations engaged in search and rescue (SAR) operations or influencing the SAR system. The organizations may be international, federal, state, county, or local governmental, commercial, or private organizations. Survivors are responsible for being familiar with procedures used by international SAR systems in order to assist in rescue efforts. Some international organizations are:

(1) International Civil Aviation Organization (ICAO).

(2) Intergovernmental Maritime Consultive Organization (IMCO).

(3) Automated Mutual-Assistance Vessel Rescue (AMVER) System.

Figure 25-1. Recovery.

25-2. National Search and Rescue (SAR) Plan:

a. The National SAR Plan is implemented the instant an aircraft is known to be down. There are three primary SAR regions; they are the Inland Region, the Maritime Region, and the Overseas Region.

b. The Air Force is the SAR coordinator for the Inland Region, which encompasses the continental United States. The Coast Guard is the SAR coordinator for the Maritime Region, which includes the Carribean Area and Hawaii. The Third National Region is the Overseas Region. The Secretary of Defense designates certain Defense Department officers as United Commanders of specified areas where US Forces are operating. Wherever such commands are established, the Unified Commander is the Regional SAR Coordinator. Overseas regions are normally served by the Joint Rescue Coordination Center, operated under the Unified Action Armed Forces. Under the terms of the National SAR Plan, the "inland" area of Alaska is considered a part of the Overseas Region.

c. The National SAR Manual, Air Force Manual 64-2, provides a long-range rescue plan which personnel should study for additional information.

25-3. Survivors' Responsibilities:

a. The survivors' responsibilities begin at the onset of the emergency, with the dispatching of an immediate radio message. The radio message should include position, course, altitude, groundspeed, and actions planned. This information is essential for initiating efficient recovery operations.

b. Once recovery operations have been initiated, survivors have a continuing responsibility to furnish information. Both ground and radio signals should be immediate considerations.

c. If a group of survivors should become separated, each group member should, when contacted by rescue forces, provide information surrounding the dispersal of the group.

d. The greatest responsibility of aircrew members is to follow all instructions to the letter. The intelligence officer will brief aircrew members on procedures for tactical situations. These instructions must be followed

explicitly since it could mean the difference between life and death. When rescue personnel tell the survivor to unhook from the raft—it should be done immediately! If instructions are not followed, survivors could be responsible for causing their own death and (or) the death of rescue personnel.

25-4. Recovery Site:

a. Consideration must be given to a recovery site. The survivor's major considerations are the type of recovery vehicle carrying out the recovery and the effects of the weather and terrain on the rescue aircraft, such as updrafts and downdrafts, heat, wind, etc. Survivors should try to pick the highest terrain possible in the immediate area for pickup. When locating this rescue site, they should watch for obstacles such as trees, cliffs, etc., which could limit the aircraft's ability to maneuver. Overhangs, cliffs, or sides of steep slopes should be avoided. Such terrain features restrict the approach and maneuverability of the rescue vehicle and require an increase in rescue time.

b. Even though survivors should select a recovery site, it is the ultimate responsibility of rescue personnel to decide if the selected site is suitable.

25-5. Recovery Procedures:

a. Knowing Current Procedures. Since procedures involving recovery vary with changes in equipment and rescue capability, survivors must always know the current procedures and techniques. This is particularly true of the procedures used for wartime recovery, which are in AFR 64-3.

(1) In deciding whether or not supplies should be dropped, rescue forces consider such factors as the relative locations of the distress site to rescue unit bases, the lapse expected before rescue is initiated, and the danger of exposure. If a delay is expected, supplies are usually dropped to survivors to help sustain and protect them while they await rescue. The mobility of survivors on the land generally makes it possible to recover equipment dropped some distance away, but airdrops at sea must be accurate.

(2) Aircraft with internal aerial delivery systems, such as the HC-130, are the most suitable for delivery of supplies to survivors. Aircraft having bomb bays or exterior racks capable of carrying droppable containers or packages of survival requisites are the next most suitable for dropping supplies. However, these aircraft are not always available for supply dropping operations, so aircraft not specifically designed for this function may have to be used.

b. Rescue by Helicopter:

(1) Helicopters make rescues by landing or hoisting. Landings are usually required at high altitudes due to limitations of helicopter power for maintaining a hover. Hoist recovery is the preferred method for effecting a water rescue. Helicopter landings are made for all rescues when a suitable landing site is available, and danger from enemy forces is not a problem. Hovering the helicopters and hoisting the survivor aboard requires

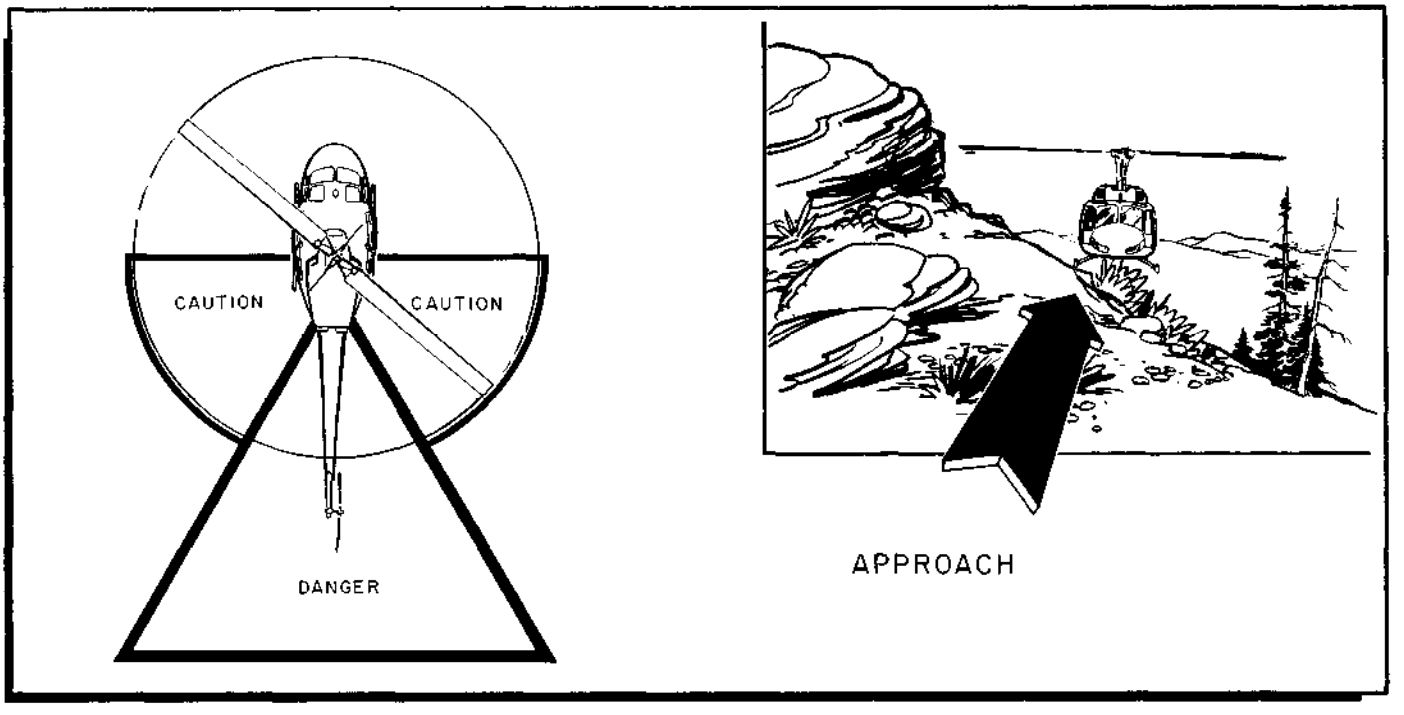

Figure 25-2. Approaching Helicopter.

more helicopter power than landing and presents a hazard to both the aircraft and the survivor. There is a danger if helicopters are operated close to collapsed parachutes. Parachute inflation by rotor downwash can cause the parachute to be sucked into the rotor blades of the helicopter.

(2) After landing, a crewmember will usually depart the aircraft. If for some reason this cannot be done, as in combat, the survivor should approach the helicopter from the 3 o'clock to 9 o'clock position relative to the nose of the helicopter and follow instructions (figure 25-2).

c. Rescue by Fixed-Wing Aircraft on Land:

(1) The most significant role played by fixed-wing aircraft in rescue operations is providing immediate assistance to survivors and serving as the "eyes" of approaching rescue units. This is done by pinpointing the survivors' position, orbiting the survivors, and dropping survival equipment. This type operation improves the morale of the survivors, fixes the survivors' location to prevent additional searching, and saves valuable time in getting the pickup unit on the scene.

(2) The role of fixed-wing aircraft in actually performing a rescue is limited to instances where there is a suitable runway near the survivor or where the aircraft is designed to operate from rough and improvised strips. Fixed-wing aircraft rescues have often been made in extremely cold climates where the aircraft have either used frozen lakes or rivers as runways or, when fitted with skis, have operated from snow-covered surfaces and glaciers. However, landing in unknown terrain under what appears to be ideal conditions is extremely hazardous.

d. Rescue by Ship:

(1) When a distress craft or survivors are a considerable distance from shore, rescue will normally be by long-range ships (specialized SAR ships, warships, or merchant ships). The rescue methods used by these ships vary considerably according to their displacement and whether the rescue is made in midocean or close to land. Weather, tides, currents, sea conditions, shallow water, reefs, daylight, or darkness may be important factors.

(2) Although it appears obvious that a marine craft should be used for rescue operations, it may be advisable to initiate an alternate method of recovery. For example, helicopters may be used to evacuate survivors picked up by marine craft in order to speed their delivery to an emergency care center.

(3) Removal of survivors from the water, liferafts, lifeboats, or other vessels to the safety of the rescue vessel deck may be the most difficult phase of a maritime search and rescue mission. In most cases, survivors will have to be assisted aboard. For this reason, all SAR vessels are usually equipped and prepared to lift survivors from the water without help from the survivors. There are numerous methods for rescuing survivors which may be used by SAR vessels. The most commonly used methods are listed in this chapter and are generally grouped as rescue of survivors in the water and rescue of survivors directly from their distressed vessel.

(a) When rescuing people from water, the following methods are generally used:

-1. Ship alongside/swimmer.
-2. Ship alongside/line thrower.
-3. Ship alongside/small boat.
-4. Ship circle/trail line.

(b) The most commonly used methods for rescuing personnel who are aboard distressed vessels are:

-1. Ship to ship/direct.
-2. Ship to ship/raft haul.
-3. Ship to ship/raft drift.
-4. Ship to ship/small boat.
-5. Ship to ship/haulaway line.

e. Rescue by Boat:

(1) When survivors are located on lakes, sheltered waters, rivers, or coastal areas, rescue will often be made by fast boats of limited range based close to the survivors or by private boats operating in the vicinity.

(2) Rescue boats are usually small and may not be able to take all survivors on board at one time; therefore, a sufficient number of boats to offset the rescue should be dispatched to the distress scene. When this is not possible, each boat should deploy its rafts so that those survivors who cannot be taken aboard immediately can be towed ashore or kept afloat while they are waiting. The boat crew should make sure any survivors who must be left behind are made as secure as circumstances permit.

(3) Assistance to an aircraft that has crashed or ditched on the water will usually consist of transferring personnel from plane to boat and picking up survivors from the water or liferafts. It may also include towing of an aircraft which is disabled on the water.

f. Coordinated Helicopter/Boat Rescues:

(1) Occasionally, boats and helicopters will be dispatched for a rescue operation. Generally the first rescue unit to arrive in the vicinity of the survivors will attempt the first rescue.

(2) If the helicopter arrives first, the boat will take a position upwind of the helicopter in the 2 o'clock position at a safe distance and stand by as a backup during the rescue attempt.

(3) If the helicopter must abort the rescue attempt, the pilot will depart the immediate area of the survivor and signal for the boat to move in and make its rescue attempt. Additionally, the helicopters may turn out the anticollision rotating beacon to indicate they require

boat assistance or are unable to complete the rescue. In certain operations where helicopter and boat coordinated rescue can be foreseen, specific signals should be prearranged.

(4) If the boat arrives first and makes the rescue, it will transfer the survivor to the helicopter to effect a rapid delivery to medical facilities.

25-6. Pickup Devices:

a. Assistance. When rescue forces are in the immediate area of survivors, they will, if conditions permit, deploy pararescue personnel to assist the survivors. Unfortunately, conditions may not always permit this, so survivors should know how to use different types of pickup devices.

b. Common Factors. Some common factors concerning all pickup devices are:

(1) The device should be allowed to ground to discharge static electricity before donning.

(2) To ensure stability, survivors should sit or kneel when donning a pickup device. Do not straddle the device.

(3) If no audio is available, survivors should visually signal the hoist operator when ready for lift-off—"thumbs up" or vigorously shake the cable from side to side.

(4) Most devices can be used as a sling (strop).

(5) Survivors must remember to follow all instructions provided by the rescue crew. When lifted to the door of the helicopter, survivors should not attempt to grab the door or assist the hoist operator in any way. They must not try to get out of the pickup device. The hoist operator will remove the device after the survivor is well inside the aircraft.

c. Rescue Sling. Before donning the rescue sling (strop), the survivor should face the drop cable and make sure the cable has touched the water or ground and lost its charge of static electricity.

(1) The most commonly accepted method for donning the rescue sling (strop) is the same as putting on a coat. After connecting the ring to form the sling (strop), the survivor's arms should be inserted one by one into the sling (strop) as it swings behind. The sling (strop) loop should be against the survivor's back with an arm around each side of the strop. The webbing under the metal ring can be held until tension is put on the cable. The survivor's hands may then be interlocked and rested on the chest. This tends to lock the survivor into the sling (strop) as upward pressure is applied (figure 25-3).

(2) Another way to enter the strop is to grasp the strop with both hands and lift it over the head to bring it down under the arms and around the body. Regardless of the method used, the survivor should remember the webbing and metal hardware of the device should be directly in front of the face.

d. Basket. If a basket is used, it will probably be accompanied to the water or ground by a member of the helicopter crew. The crewmember will assist survivors into the basket. There are two types of baskets: The litter type in which the person lies flat, and the seat type that survivors enter and sit down in as they would in a chair (figure 25-4).

e. Forest Penetrator:

(1) The forest penetrator rescue seat is designed to make its way through interlacing tree branches and dense jungle growth. It can also be used in open terrain or over water. The device is equipped with three seats

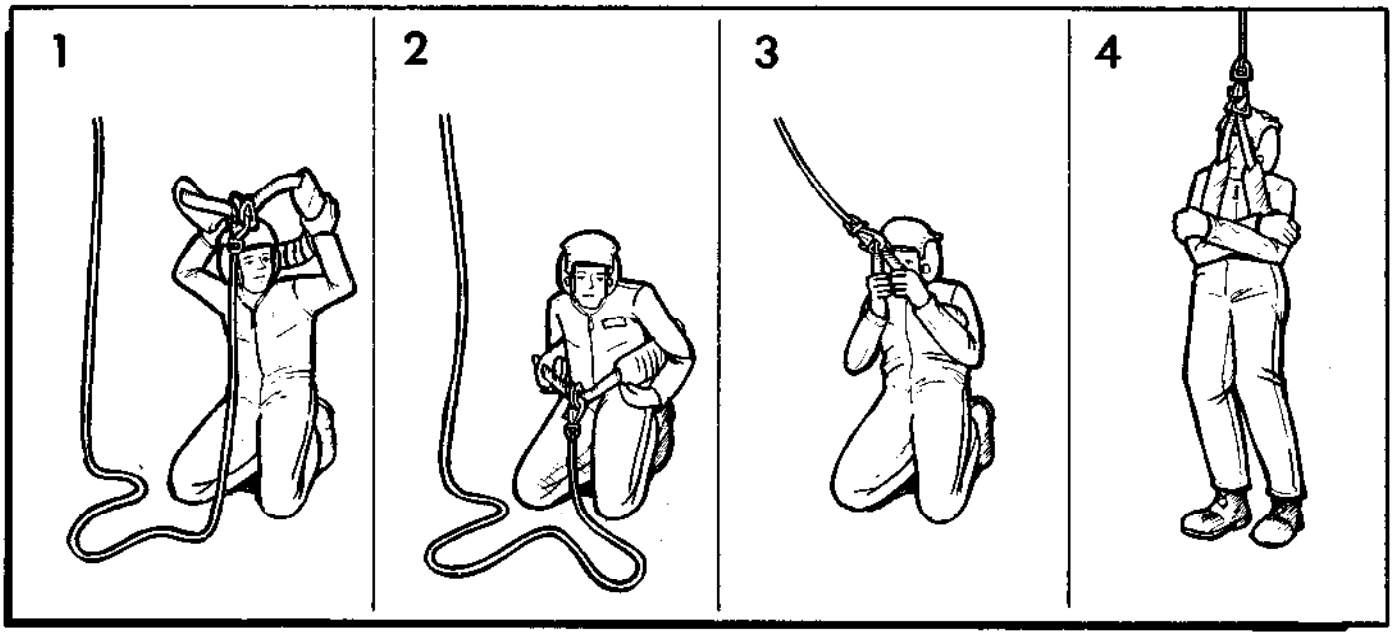

Figure 25-3. Horse Collar.

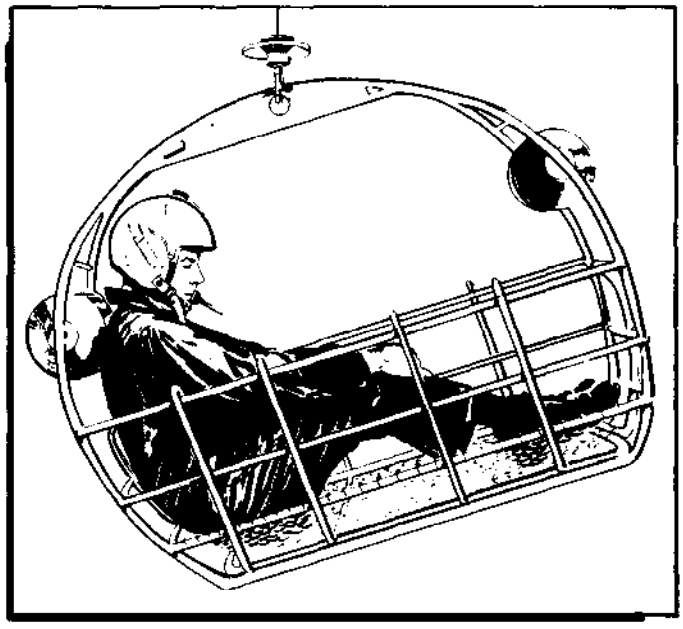

Figure 25-4. Basket.

which are spring-loaded in a folded position against the body or main shaft and must be pulled down to the locked position for use. On the main shaft of the tube, above the seats, there is a zippered fabric storage pouch for the safety (body) straps which are stowed when lowered to the survivor for a land pickup. The penetrator may also be equipped with a flotation collar. (NOTE: If the forest penetration is used for water pickup, it will be equipped with the flotation collar which enables the device to float with the upper one-third (approximately) of the device protruding above the water. Additionally, for water pickups, one strap will be removed from the stowed position, and one seat will be locked in the down position to assist the survivor in using the penetrator.)

(2) The safety strap is pulled from the storage pouch and placed around the body to hold the person on the penetrator seat. The strap should not be unhooked unless there is no other way to fasten it around the body. The survivor must make certain the safety strap does not become fouled in the hoist cable. After the strap is in place, the seat should be pulled down sharply to engage the hook which holds it in the extended position. The survivor can then place the seat between the legs. Then the survivor should pull the safety strap as tight as possible ensuring the device fits snugly against the body. The survivor must always keep the arms down, elbows locked against the body, and not attempt to grab the cable or weighted snap link above the device. After making certain the body is not entangled in the hoist cable, the signal to be lifted can be given (figure 25-5).

(3) In a combat area, under fire, survivors may be lifted out of the area with the cable suspended before

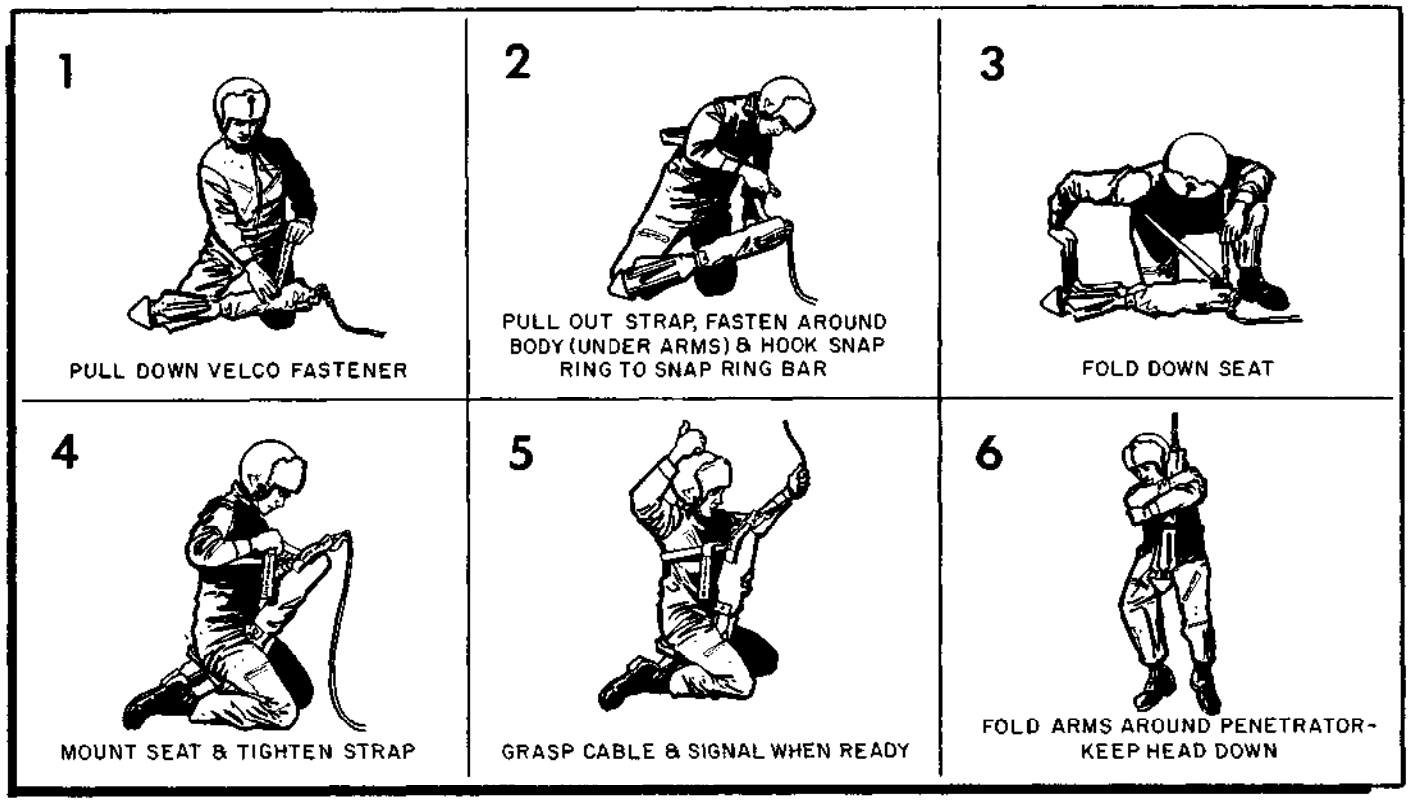

Figure 25-5. Forest Penetrator.

being brought into the helicopter. It is important to be correctly and securely positioned on the pickup device. The seat should always be held tightly against the crotch to prevent injury when slack in the cable is taken up. The hands should be kept below and away from the swivel on the cable with the arms around the body of the penetrator. Survivors should keep their head close to the body of the penetrator so that tree branches or other obstructions will not come between the body and the hoist cable.

(4) When survivors reach a position level with the helicopter door, the hoist operator will turn them so they face away from the helicopter and then pull them inside. The crewmember will disconnect the survivors from the penetrator once the device is safely inside the helicopter.

(5) The forest penetrator is designed to lift as many as three persons. When two or three survivors are picked up, heads should be kept tucked in and each individual's safety strap drawn tight. The penetrator can be used to lower a paramedic or crewmember to assist injured personnel, and both (survivor and paramedic) can be hoisted to the helicopter. If the forest penetrator seat blades have been lowered in a tree area, and if for any reason the pickup cannot be made, the blades should be returned to the folded position to prevent possible hangup on tree limbs or other objects while the device is being retracted.

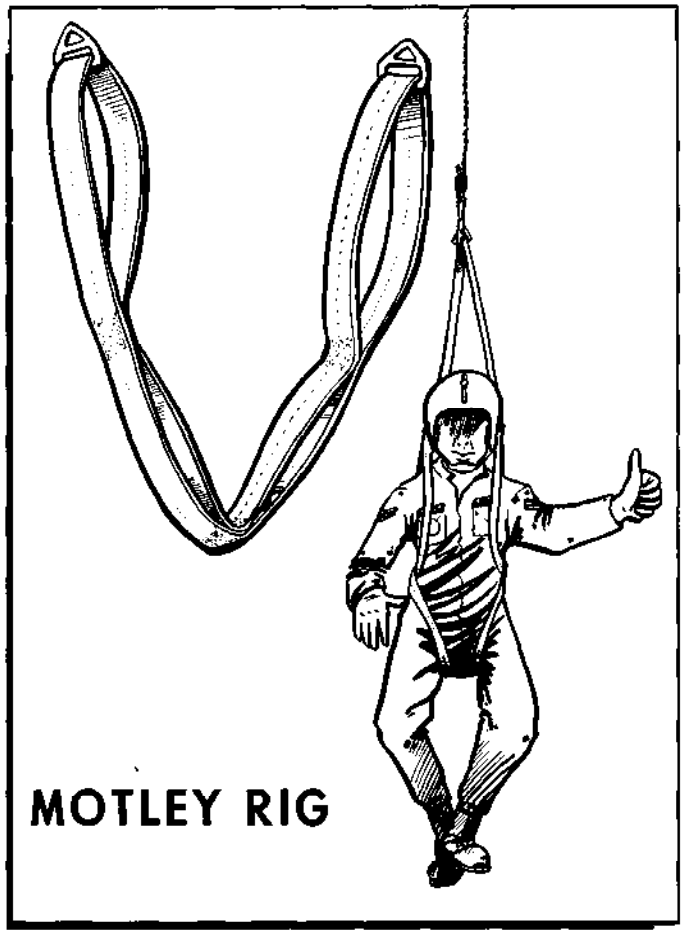

Figure 25-6. Motley Rig.

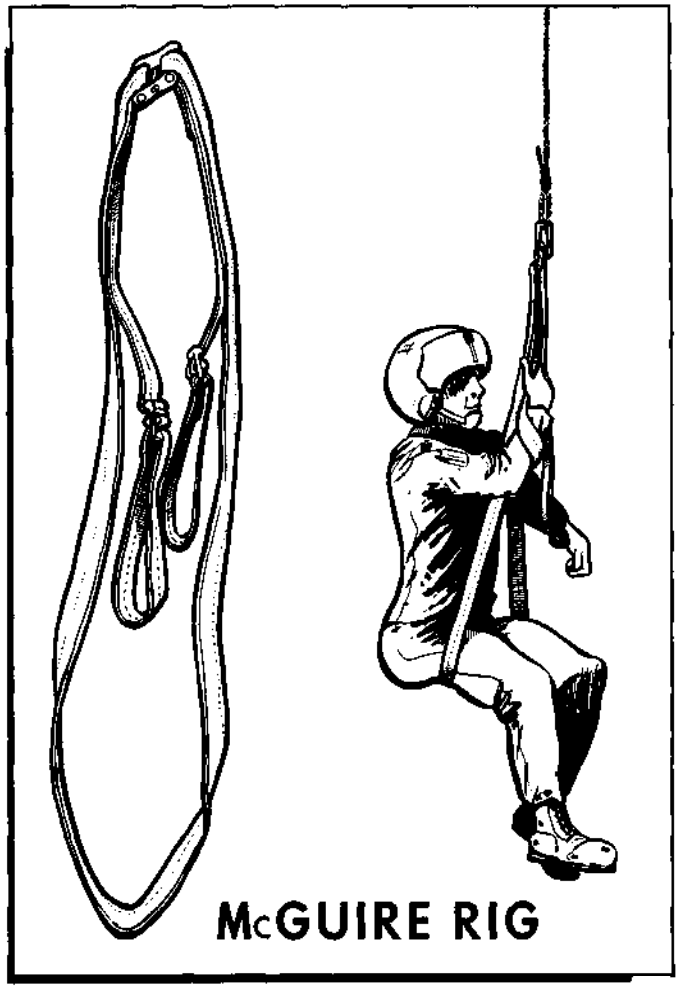

Figure 25-7. McGuire Rig.

(6) With all types of devices, it is necessary to watch the device as it is lowered. The devices weigh about 23 pounds. If the device were to hit a survivor, it could cause a serious injury or death.

f. Other Devices. There are other devices which could be used to pick up survivors. Some of them are the Motley and McGuire rigs (figures 25-6 and 25-7), the Swiss Seat and Stabo rig (figures 25-8 and 25-9), and the Rope Ladder (figure 25-10).

(1) Motley and McGuire Rigs. These devices may be carried by Army helicopters either designated as the recovery aircraft in assault or for use to insert or extract special ground forces. The device is normally packed in a weighted canvas container and dropped by rope. The device is dropped to the survivor, who is allowed time

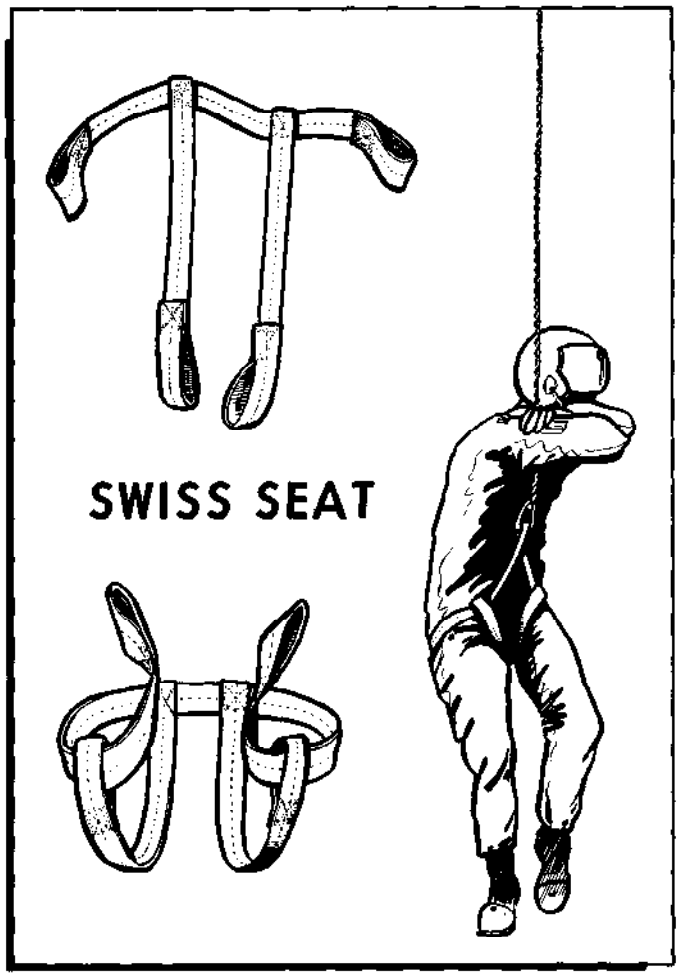

Figure 25-8. Swiss Seat.

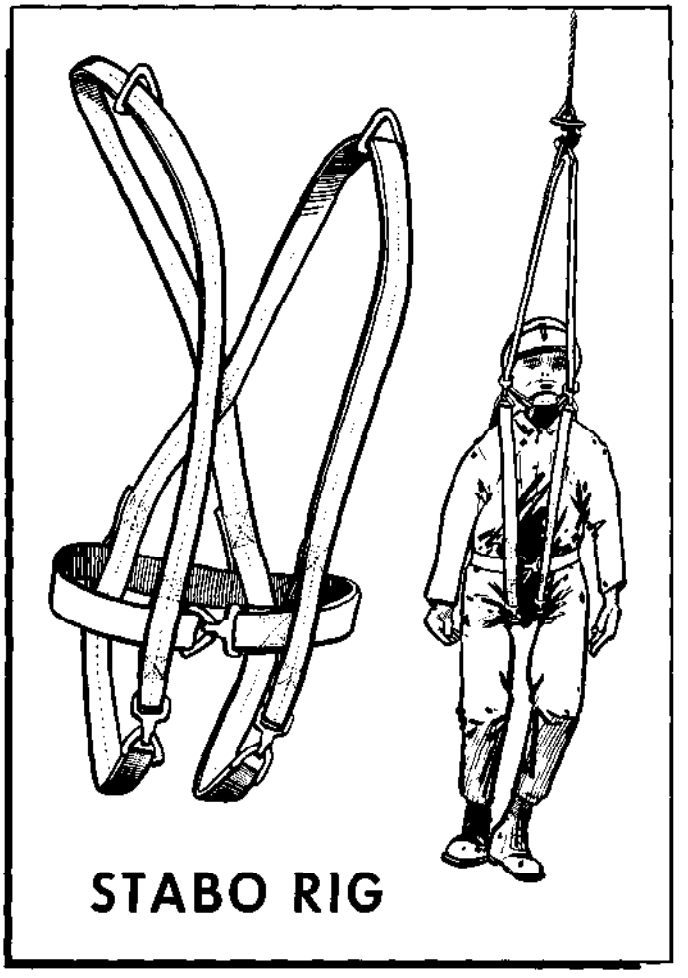

Figure 25-9. Stabo Rig.

for donning. The helicopter then returns trailing a rope which is then fastened to the device for pickup. Generally, the survivor is not hoisted into the helicopter; therefore, all safety straps should be securely fastened.

(2) Swiss Seat and Stabo Rig (figures 25-8 and 25-9). These devices are carried by special ground forces who may require instant extraction by helicopter. Special ground forces put their devices on and wait for the helicopter to drop ropes which are snapped into the devices for rapid extraction. Although not normally carried aboard the aircraft, the Army helicopter may supply one of these devices to the survivor. Again, the survivor would not be hoisted into the helicopter.

(3) Rope Ladder. This device is used primarily by the Army and special ground forces. If this device is used, it should be approached from the side and not the front. The survivor should climb up a few rungs, sit down on a rung, intertwine the body with rungs (figure 25-10). The survivor should not try to climb up the ladder and into the helicopter.

25-7. Preparations for Open Seas Recovery:

a. On sighting rescue craft approaching for pickup, (boat, ship, conventional aircraft, or helicopter), survivors must quickly clear any lines (fishing lines, desalting kit lines, etc.) or other gear which could cause entanglement during rescue. All loose items should be secured in the raft. Canopies and sails should be taken down to ensure a safer pickup. After all items are secure, the survivor should put on the helmet (if available). The life preserver should be fully inflated with the oral valve locking nut tight against the mouthpiece. Survivors should remain in the raft, unless otherwise instructed, and disengage all gear except the preservers. If possible, rescue personnel will be lowered into the water to assist survivors. The survivors should remember to follow all instructions given by rescue personnel.

b. If helicopter recovery is unassisted, the survivor will be expected to do the following before to pickup:

(1) Secure all loose equipment in raft, accessory bag, or in pockets.

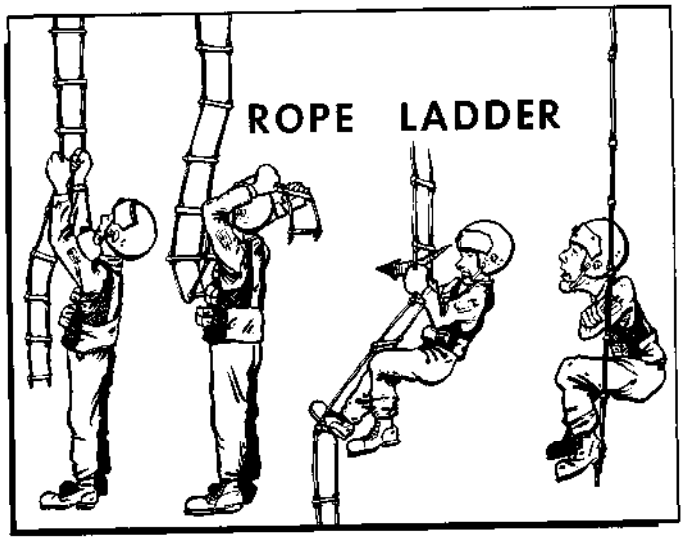

Figure 25-10. Rope Ladder.

(2) Deploy sea anchor, stability bags, and accessory bag.

(3) Partially deflate raft and fill with water.

(4) Unsnap survival kit container from parachute harness.

(5) Grasp raft handhold and roll out of raft.

(6) Allow recovery device and (or) cable to ground out on water surface.

(7) Maintain handhold until recovery device is in the other hand.

(8) Mount recovery device (avoid raft lanyard entanglement).

(9) Signal hoist operator for pickup.

Part Nine

EVASION

Chapter 26

LEGAL AND MORAL OBLIGATIONS

26-1. Introduction:

a. Aerial combat in the future, as in the past, will expose aircrews to possible ejection, bailout, or forced landings into enemy controlled territory. The aircrews who encounter such traumatic circumstances must be prepared to survive and evade the enemy to return to friendly control. An active commitment to solving problems and to individual survival (the "will to survive") is essential. Aircrew members must be prepared to exert extreme effort, both mental and physical, to successfully evade capture. In an excerpt from a debrief, one survivor describes what it was like coming to terms with these problems. "I thought to myself...'Well, I'm in a hell of a spot, what am I going to do about it?' The situation was such that I didn't do anything about it for quite awhile. I just sat there in the rain where I had landed and stared at the ground. It became colder, and after what seemed like hours, I lifted my gaze and sat staring now into the forest, looking at it without even really seeing it. I started thinking about the chain of events that had preceded my landing in this God-forsaken forest so far from my lines. My thoughts were as jumbled and unreal as the fog into which I had dived. Soon, it started to snow and I realized that I was in danger of being discovered or of freezing to death where I sat, and so, with a huge effort of will I forced myself to think coherently. I stood, and so began my journey." (See figure 26-1.)

b. This person had no survival training but nevertheless managed to successfully evade for miles back to his own lines with little more than a strong will to survive and common sense. Potential evaders must understand that evading capture presents a difficult challenge. They will meet and have to overcome a succession of obstacles, both manmade and natural. Knowledge gained from the experience of others, their own training, and prior preparation and planning will help them to overcome those obstacles.

c. This part addresses covert survival; that is, evasion. Areas to be covered include:

(1) Evader's moral obligations and legal status.

(2) Principles and techniques of evasion, camouflage, and travel (assisted or unassisted).

(3) The special aspects of food and water procurement.

(4) Combat signaling and recovery.

26-2. Definitions:

a. Evader. An "evader" (JCS Pub 1) is "a person who through training, preparation, and application of natural intelligence avoids contact with, and capture by, hostiles, both military and civilian."

b. Evasion. Evasion means all the processes involved in living off the land and, at the same time, avoiding capture while returning to friendly control. As used here, it includes all the techniques of evasion employed by those on foot in enemy territory.

26-3. Military Drive:

a. A crewmember becomes an evader when isolated in hostile areas, unable to continue the assigned mission, and when prevented from rejoining friendly forces. Definitions are useful but will not contribute to success in evading capture unless potential evaders understand what factors give direction and guidance to their efforts.

b. If the opportunity exists, evaders must be motivated to take advantage of it. The motivation to make a total effort to adhere to every evasion principle 24 hours a day may be personal or military. This strong central drive will give the survivor the necessary push to make these efforts.

c. Even if evasion is unsuccessful and the evader is captured, every hour spent eluding the enemy ties up enemy forces and lessens the evader's intelligence value.

d. In addition to the above reasons for evading capture, survivors also have moral and legal obligations to fulfull.

(1) Moral obligation is implied throughout the Articles of the Code of Conduct, specifically Article II. Article II states: "I will never surrender of my own free will. If in command, I will never surrender my men while they still have the means to resist." This Article of the Code should guide an evader's behavior during evasion just as it does in any other combat situation.

(2) The UCMJ continues to apply to evader's conduct during evasion or captivity. Particularly applicable are Article 99, Misbehavior Before the Enemy, Article 104, Aiding the Enemy, and Article 92, Failure to Obey a Lawful Order. Thus, one can be tried for misconduct as a combatant or as a noncombatant. A combatant is defined in AFP 110-31 as "a person who engages in hostile acts in an armed conflict on behalf of a party to the conflict." The combatant must conform to the standards established under international law for combatants, be authorized by his or her country to so act, and be recognizable as a combatant by uniform, insignia, or

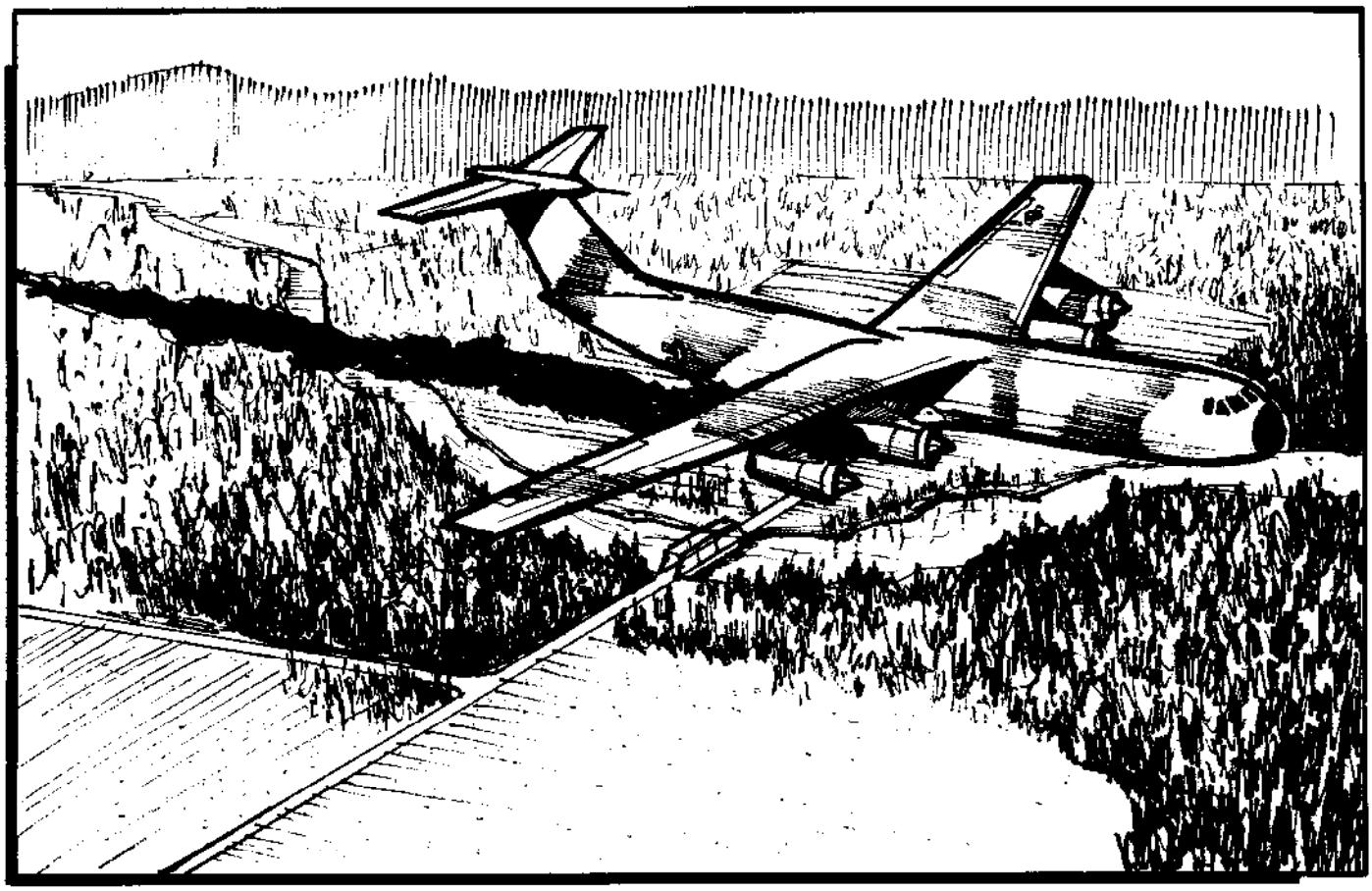

Figure 26-1. Evasion.

other sign. A noncombatant includes a wide variety of persons, including civilians, prisoners of war, sick and wounded persons, chaplains, medics, and other similar persons (AFP 110-31). Various countries around the world have developed written and unwritten laws of war. Four Geneva treaties were entered into by the United States and 60 other countries in 1949 and these treaties, as since amended, are in AFP 110-20.

(3) An "evader" is defined in international law as "any person who has become isolated in hostile or unfriendly territory and who eludes capture." An evader is a combatant and retains this status as a fighting person under arms according to international law until captured. Evaders are considered instruments of their government, under orders to evade capture, and never to surrender of their own free will. The evader is still militarily effective and may take such steps as necessary, within the rules of warfare, to accomplish the mission, which includes returning after striking an enemy target. While in combatant status, the evader may continue to strike legitimate military targets and enemy troops without being held liable to prosecution after capture for violation of the local criminal law. To do so while evading capture is the legal function of a combatant.

(4) Once captured, an evader becomes a noncombatant and will occupy the status of a prisoner of war (PW). A PW who kills or wounds enemy personnel of the Detaining Power in an attempt to escape or evade may be tried and punished for the offense. International law provides certain rights to a PW and requires issuance of an identity card showing name, rank, serial number, and date of birth. When questioned, a PW is bound only to provide this information. A PW who escapes remains a noncombatant until he or she rejoins the armed forces of his or her country or the armed forces of a friendly power. Once escape is completed, combatant status is regained and no punishment may be imposed by the Detaining Power in the event of subsequent recapture. A PW who attempts to escape and is recaptured before rejoining his or her armed forces is liable under the Geneva Convention only for disciplinary punishment in respect of this act, even if it is a repeated offense. However, special surveillance may be imposed.

(5) Disguise is a lawful means of evading enemy forces so long as a means of military identification is retained on the person. However, an evader while in disguise may not participate in or commit hostile acts

involving destruction of life or property. A person involved in such an offense will be classified as an "unlawful combatant," will not be entitled to PW status, and may be tried by the enemy and sentenced to imprisonment or execution under certain circumstances. A disguise also cannot be used by a military person for the purpose of gathering enemy military information or for waging war, and to do so will result in loss of PW status if captured.

(6) Evasion can be classified as either assisted or unassisted. Assistance can be defined as any help which is offered to an American aircrew member by any person. This help may include food, clothing, medicine, shelter, money, and even such a small item as a shoelace. Evaders should, in fact, consider they have been assisted even if, while evading through hostile territory, their presence has been ignored by indigenous personnel. The term unassisted evasion would then relate to the situation where the survivor, as an evader, must rely solely on his own knowledge and abilities to successfully emerge from an enemy-held or hostile area to areas under friendly control. The process of emerging may include aerial recovery, water recovery, or assisted evasion, but it is primarily an unassisted individual effort. Certain principles apply to all evasion situations; certain procedures should be followed; and certain techniques have widespread application. The beginning of an evasion experience is often the most critical phase. This is particularly true for a person who bails out over enemy territory during daylight hours and in sight of enemy personnel. The downed aircrew member can count on the enemy to make a determined effort to capture him if seen. Eluding the enemy is also a matter of effort and luck. Luck plays its part initially in establishing where a crewmember lands in relation to the location of any enemy personnel who may have seen the descent. For example, the downed crewmember who lands in a heavily populated area (city, military installation or combat area) may be taken prisoner immediately. Here the problem becomes one of early escape rather than one of evasion.

(7) An evasion situation should not be categorized in terms of length. History has proven that predicting the length of any specific evasion situation is practically impossible. All crewmembers should be prepared to evade until rescued, no matter how long the evasion experience might last. Emphasizing the advantages of "short-term" evasion over "long-term" may cause an overly optimistic, possibly even foolhardy, attitude toward evasion planning. Or, evaders may decide, if they are not rescured in a "short" period of time, it is no longer worth the effort; thereby, taking on a defeatist attitude.

Chapter 27

FACTORS OF SUCCESSFUL EVASION

27-1. Basic Principles. All potential evaders must have three things in their favor. These are the same three things needed by a potential escapee. The three factors which increase chances of successful evasion are preparation, opportunity, and motivation.

27-2. Preparation:

a. Preparation is one of the most important factors for successful evasion. The actions crewmembers take before the evasion episode can make the difference between being able to evade or being captured. In a hostile area the survivors should remember evasion is an integral part of their mission and plan accordingly. The enemy may make mistakes of every conceivable form and not suffer more than indignation, anger, and fatigue. The evader, on the other hand, must constantly guard against mistakes of any sort. Being seen is the greatest mistake an evader can make. The evader must prepare for this task (figure 27-1).

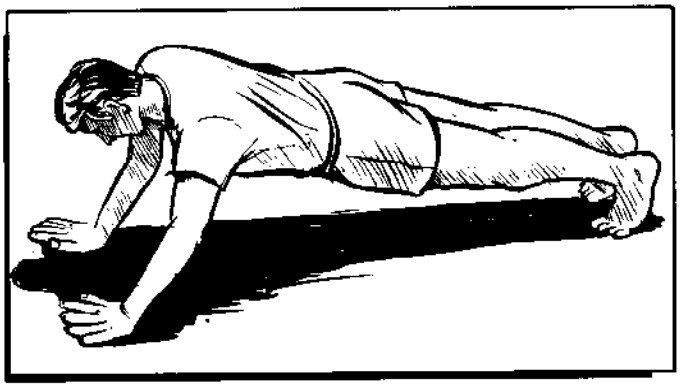

Figure 27-1. Preparation.

b. Three basic problems during evasion are:

(1) Evading the enemy.

(2) Surviving.

(3) Returning to friendly control.

c. Chances for a successful evasion are improved if evaders:

(1) Observe the elementary rules of movement, camouflage, and concealment.

(2) Have a definite plan of action.

(3) Be patient, especially while traveling. Hurrying increases fatigue and decreases alertness. Patience, preparation, and determination are key words in evasion.

(4) Conserve food.

(5) Conserve as much strength as possible for critical periods.

(6) Rest and sleep as much as possible.

(7) Maintain a highly developed "will to survive" and "can do" attitude. Evasion may require living off the land for extended periods of time and traveling on foot over difficult terrain, often during inclement weather.

(8) Study the physical features of the land. They should note the location of mountains, swamps, plains, deserts, or forests, type of vegetation, and availability of water.

(9) Consider the climate. Aircrew members should know the climatic characteristics and typical weather conditions of the area which may be flown over.

(10) Study ethnic briefs and survival, evasion, resistance, and escape (SERE) contingency guides before a mission and learn some of the customs and habits of the local people. Such knowledge will aid in planning missions and evasion plans of action. For example, it may give the evader the ability to avoid hostile people or groups or to identify and deal with "friendlies." This knowledge may also allow for blending into the local populace (figure 27-2).

(11) Know the equipment well! One must know the location of each item in the kit, its operation, and its value. An evader must preplan which equipment should be retained and which should be left behind.

d. Once in the evasion situation, planning for travel will be a consideration for evaders. They must have a definite objective and be confident in their approach and ability to achieve it. They will normally have several options with variations to choose from in selecting a plan of action or destination. The enemy force deployment, search procedures, terrain, population distribution, climate, distance, and environment (that is, NBC) will influence destination selection. Examples of options and destinations:

(1) Await SAR forces.

(2) Evade to a SAFE area.

(3) Evade to a neutral country. (NOTE: Border areas not disrupted by combat may have a security system intact.)

(4) If evaders are in the forward edge of the battle area (FEBA) and feel sure that friendly forces are moving in their direction, they should seek concealment and allow the FEBA to overrun their position. Evaders' attempts at penetrating the FEBA should be avoided. Evaders face stiff opposition from both sides.

e. The chances that one of these destinations may be close by will be determined by many things including the time and location of the bailout. Other determining factors are: the location and direction of movement of the FEBA, the presence or absence of willing assisters, and the knowledge of the evader's whereabouts possessed by rescue personnel. If the survivor does not land

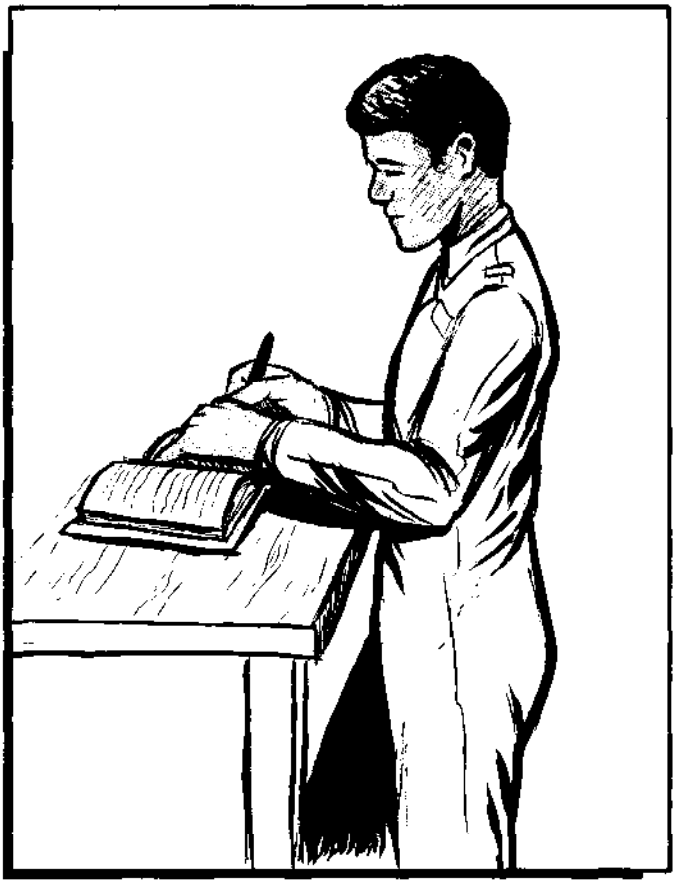

Figure 27-2. Study Ethnic Briefs.

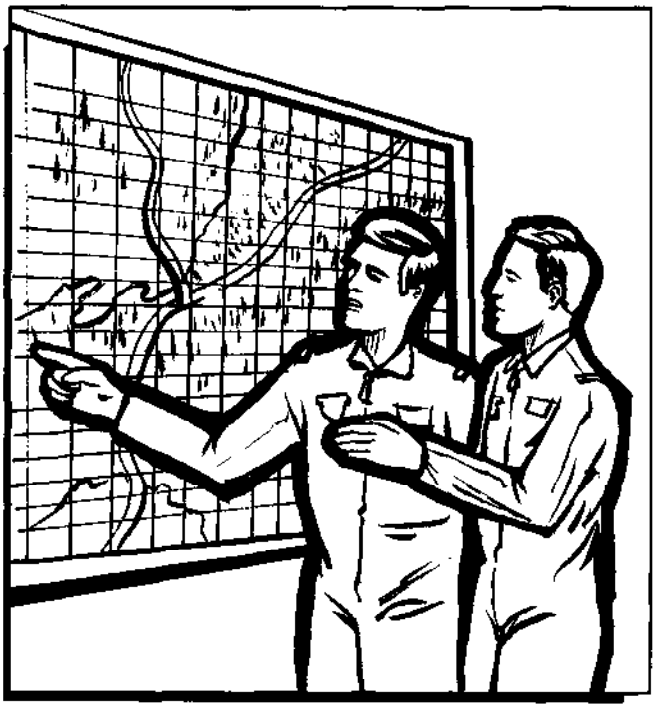

Figure 27-3. Planning for Travel.

close to one of the above areas or if the previously mentioned factors do not favor immediate air pickup, the survivor may have to travel some distance to reach one of the destinations (figure 27-3).

f. One consideration in choosing destination and direction of travel after bailout is whether one of these suitable areas for a pickup or contact with friendly forces exists and, if so, its location. Some preplanning should have been done before the mission. Information upon which to base a decision is derived from command area briefings, area studies, SERE contingency guides, and premission intelligence briefings.

g. Another consideration is physical condition. One's physical condition is the responsibility of the potential evader and has a great effect on the evader's ability to survive. Once on the ground, it is too late to get in shape. One more aspect is an aircrew member's personal habits. Upon first consideration, personal grooming habits might not be considered an important premission briefing item. However, using aftershave lotion, hair dressing, or cologne could add to the problems of an evader. The odor can carry for great distances and give away the evader's presence.

27-3. Opportunity. Potential evaders must take advantage of any and all opportunities to evade. This starts in the aircraft when an emergency is declared. Following current, approved emergency in-flight procedures for the theater of operations (when ejection, bailout, or ditching appears imminent), the aircraft commander will attempt to establish radio contact by first calling on the secure frequency of last contact; second, on an established common secure frequency; and third, on the international emergency frequency. When communication is established, the tactical call sign, type of aircraft, position, course, speed, altitude, nature of difficulties, and intentions will be transmitted. The identification friend or foe (IFF) should be set to the emergency position. When possible, ejection or bailout should be attempted over or near a SAFE area, lifeguard station, or submarine pickup point. This minimizes threat involvement for evaders and SAR forces alike. After ejection or bailout and during descent, the aircrew member must remain alert and steer the parachute away from potential threats (populated areas, gun emplacements, troop concentrations, etc.) or out to sea (feet wet). Once on the ground, the evader must be proficient in the use of the survival/evasion equipment to facilitate evasion (for example, use of the compass in conjunction with the survival radio to call in airstrikes on enemy forces threatening the evader). In addition to the opportunity to evade, motivation is essential to the evader's success.

Figure 27-4. Avoiding Detection.

27-4. Motivation. A strong, central drive will give the evader the necessary push to succeed. It may be personal, ranging in nature from a frame of reference gained through training to a desire to return to a family or loved ones. Motivation may be strictly military, involving one or all of the following reasons applicable to all military men:

a. To return and fight again.

b. To deny the enemy a source of military information.

c. To deny the enemy a source of propaganda.

d. To deny the enemy a source of forced labor.

e. To tie up enemy forces, transportation, and communications that otherwise might be committed to the war effort.

f. Return with intelligence information.

g. In addition, the Code of Conduct calls upon the military members not to surrender of their own free will.

h. Additionally, it is suggested that other personal reasons for being motivated to evade include: fear of death, pain, suffering, humiliation, degradation, disease, illness, torture, uncertainty, and fear of the unknown. From the evader's point of view, *evasion is far more desirable than captivity or death.*

27-5. Evasion Principles:

a. Besides the preparation, opportunity, and motivation factors important to evasion, there are other important principles. The evader should try to recall any previous briefings, standard operating procedures, or training. A course of action should then be chosen which has the greatest likelihood of resulting in the return to friendly forces.

b. Evader actions should be flexible. Flexibility is one of the most important keys to successful evasion. The evader, basically, must never be so firmly set in a course of action that a change is out of the question. The best thing an evader can do is to stay open to new ideas, suggestions, and changes of events. Having several backup plans of action can give the evader organized flexibility. If one plan of action is upset by enemy activity, the evader could rapidly switch to a backup plan without panic.

c. The evader is primarily interested in avoiding detection. Each evader should remember that people catch

people. If the evader avoids detection, success is almost assured. Evaders should:

(1) Observe and listen for sounds of enemy fire and vehicle activity during parachute descent and moving away from those enemy positions once on the ground. Flyers downed during daylight should assume they were seen during descent and expect a search to center on their likely point of landing.

(2) Be patient and determined while traveling.

(3) Use poor weather conditions as an aid in evading.

(4) Circumstances permitting, select times, routes, and methods of travel to avoid detection.

(5) Avoid lines of communication (waterways, roads, etc. (figure 27-4)).

d. The evaders' main objective is immediate recovery. In hostile areas or situations, survivors must sanitize all evidence of presence and direction of travel (figure 27-5.) Survivors may never be certain that rescue is imminent.

e. Although evaders would not normally move too far if rescue is imminent, in many situations they will have to leave the landing area quickly and travel as far as practical before selecting a hiding place. They should leave no sign which indicates the direction or presence of travel. All hiding places should be chosen with extreme care. The time evaders will remain in the first location is governed by enemy activity in the area, their physical condition, availability of water and food, rescue capabilities, and patience. It is in this place of initial concealment that the evaders should regain strength, examine the current situation, and plan for the evasion problems ahead (figure 27-6).

f. Once in a place of concealment, evaders should make use of all available navigation aids to orient themselves. After finding their location, evaders should also select an ultimate destination and any necessary alternate destinations. The best possible route of travel should then be decided upon. When the time comes to move, they should have a primary plan and alternate plans for travel that cover eventualities they may encounter.

(1) Evasion in a forward area has one great advantage which is not present further to the rear. Assistance may be close at hand. This assistance may come from serveral sources, each of which, under particular circumstances, may prove to be the most effective. These sources may be air cover by tactical fighter flights, helicopter recovery, and rescue by ground forces. Contact with friendly forces in forward areas requires extreme caution. Do not surprise them or move suddenly. They may mistake the evader as the enemy.

Figure 27-5. Sanitize Area.

Figure 27-6. Movement from the Area.

(2) The situation at the time of the emergency will determine the evaders' best course of action. High ground is normally the best position from which to await rescue; evaders may expect the best results from signaling devices, may observe the surrounding terrain, and may be kept under observation by friendly air cover. Whatever position is chosen, it must be clear of obstacles that would prevent a successful rescue.

(3) If not rescued immediately, the situation may compel evaders to move. Evaders must plan a course of action before leaving their position. When the evaders are certain their position is known to friendly elements, they might expect ground forces to attempt a rescue. They should remember their position may be detected as the enemy search parties approach. They must be prepared to evade to a new position.

(4) Evaders should remember that when traveling they are probably more vulnerable to capture. Once past the danger of an immediate search, evaders must avoid people. Inhabited areas should be bypassed rather than penetrated, even if it means miles of added travel. Many evaders have been captured because they followed the easiest and shortest route, or failed to employ simple techniques such as scouting, patrolling, camouflage, and concealment. As a rule, the safest route avoids major roads and populated areas, even if it takes more time and energy. Unaccompanied evasion requires self-reliance and independent action (figure 27-7).

Figure 27-7. Avoiding Populated Areas.

Chapter 28

CAMOUFLAGE

28-1. Introduction. Presence of evaders in an area controlled by the enemy may require the evaders to adopt and maintain camouflage to avoid observation. Camouflage consists of those measures evaders use to conceal their presence from the enemy. Camouflage is a French word meaning disguise, and it is used to describe action taken to mislead the enemy by misrepresenting the true identity of an installation, an activity, an item of equipment, or an evader. As a tool for evasion, it enables evaders to carry out life supporting activities and to travel unseen, undetected, and free to return to friendly control. Camouflage allows them to see without being seen. They should try to blend in with the surrounding environment. Effective individual concealment often depends primarily on the choice of background and its proper use. Background is that portion of the surroundings against which an evader will be seen from the ground and the air. It may consist of a barren rocky desert, a farm yard, or a city street. It is the controlling element in individual camouflage and governs every concealment measure. At all times, camouflage is the responsibility of the individual evader. In the event of group evasion, the group leader and each individual are responsible for the camouflage of the group. Evaders should remember that camouflage is a continuous, never-ending process if they want to protect themselves from enemy observation and capture (figure 28-1).

28-2. Types of Observation:

a. Of the five senses, sight is by far the most useful to the enemy, hearing is second, while smell is of only occasional importance. But these same senses can be of equal value to the evader and observer.

b. How useful these senses are depends primarily on range. For this reason, basic camouflage stresses visual concealment which is relatively long range. Most people are accustomed to looking from one position on the ground to another position on the ground.

Figure 28-1. Camouflaging.

c. Before evaders can conceal themselves from aerial observation, they should become familiar with what their activities look like from the air, both in an aerial photograph and from direct observation. The evaders must also have an understanding of the types of observation used by the enemy. There are two categories of observation—direct and indirect.

(1) Direct Observation:

(a) Direct observation refers to the process whereby the observer looks directly at the object itself without the use of telescopes, field glasses, or sniperscopes. Direct observation may be made from the ground or from the air. Direct aerial observation becomes more and more important because of the rapid changes in weapons and in tactical situations due to greater mobility of troops. Reconnaissance aircraft over enemy lines report locations of troops, vehicles, and installations (or shelter areas) as seen from the air-to-ground control stations. Reported targets can be immediately fired upon or troops can be sent in to investigate shelter areas or other suspicious areas (figure 28-2).

(b) The enemy may also use dogs, foot patrols, and mechanized units to patrol a given area. Such teams could physically search an area for signs of the passage of strangers, such as footprints, old campfires, discarded or lost equipment, and other "telltale" signs which would indicate that someone had been in the area.

(c) Observation by the local populace is also a possibility. Upon seeing an evader or "telltale" signs an

Figure 28-2. Direct Observation.

evader left behind, they may contact the local authorities, who initiate organized searches.

(2) Indirect Observation:

(a) Indirect observation refers to the study of a photograph or an image of the subject via photography, radar, or television. This form of observation is becoming increasingly more varied and widespread, and may be used from either manned or unmanned positions.

(b) Views from the ground are familiar, but views from the air are usually quite unfamiliar. In modern warfare, the enemy may put emphasis on aerial photographs for information. It is important to become familiar with the "bird's-eye view" of the terrain as well as the ground view in order to learn how to guard against both kinds of observation.

28-3. Comparison of Direct and Indirect Observation:

a. The main advantage of direct observation is that observers see movement of an evader without camouflage. An observation can be maintained over relatively long periods of time. The main disadvantage lies in human frailty. For example, the observer's attention may be diverted to another area, or the observer may be fatigued and unable to concentrate.

b. Indirect observation has many advantages. Indirect observation can be far-reaching, cover large areas, and be very accurate. It also produces a record of the area observed so that the recorded picture can be studied in detail, compared, and evaluated. The principal disadvantage is a photograph covers a very short period of time, making detection of movement difficult. This disadvantage can be partially overcome by taking pictures of the same area at different intervals and comparing them for changes (figure 28-3).

Figure 28-3. Indirect Observations.

28-4. Preventing Recognition:

a. Recognition is the determination (through appearance, behavior, or movement of the hostile or friendly nature) of objects or persons. One objective of camouflage concealment is to prevent recognition. Another objective is to deceive or induce false recognition. This implies that camouflage is not always designed to be a "cloak of invisibility." In some instances, camouflage is used to allow deception. The camouflaged object or person is then seen as a natural feature of the landscape.

b. Recognition through appearance is the result of conclusions drawn by the observer from the position, shape, shadow, texture, or color of the objects or persons. Recognition through behavior or movement includes deductions made from the actual movements themselves or from the record left by tracks of persons or vehicles or by other violations of camouflage discipline. Camouflage disciplines are those actions which contribute to an evader's ability to remain undetected. Proper use of camouflage discipline avoids any activity that changes an area or reveals objects to an enemy. Examples of common breaches of camouflage discipline include reflections from brightly shining objects (watches, glasses, rings, etc.) (figure 28-4), overcamouflaging, or using camouflage materials which are foreign to the area presently occupied by an evader. Evaders must also watch for signs that may reveal enemy camouflage efforts. Inadvertently walking into a camouflaged enemy position may result in capture.

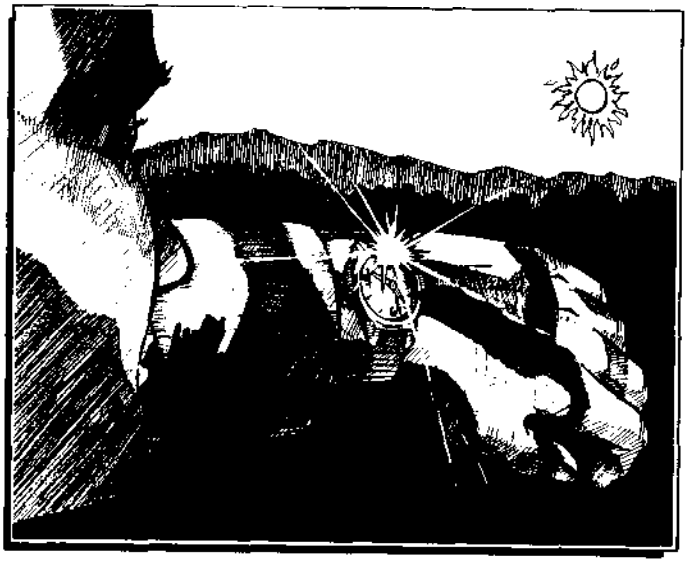

Figure 28-4. Reflections.

28-5. Factors of Recognition. Regardless of the type of observation, there are certain factors which help to identify an object. They are called the factors of recognition and are the elements which determine how quickly an object will be seen or how long it will remain unobserved. The eight factors of recognition are position, shape, shadow, texture, color, tone, movement, and shine. These factors must be considered in camouflage to ensure that one or more of these factors do not reveal the location of the evaders.

a. Position. Position is the relation of an object or person to its background. When choosing a position for concealment, a background should be chosen which will virtually absorb the evader (figure 28-5).

Figure 28-5. Position.

b. Shape. Shape is the outward or visible form of an object or person as distinguished from its surface characteristics and color. Shape refers to outline or form. Color or texture is not considered. At a distance, the forms or outlines of objects can be recognized before the observer can make out details in their appearance. For this reason, camouflage should disrupt the normal shape of an object or person (figure 28-6).

c. Shadow. A shadow may be more revealing than the object itself, especially when seen from the air. Objects such as factory chimneys, utility poles, vehicles, and tents (or people) have distinctive shadows. Conversely, shadows may sometimes assist in concealment. Objects in the shadow of another object are more likely to be overlooked. As with shape, it is more important to disrupt the shadow pattern than to totally conceal the object or person. The identifiable shadows can be broken up by the addition of natural vegetation at various points on the body. Wearing "shapeless" garments will also disrupt the outline. For example, a soft and shapeless field cap can be used instead of a helmet or flight cap (figures 28-7 and 28-8).

Figure 28-6. Shape.

d. Texture. Texture is a term used to describe the relative characteristics of a surface, whether that surface is a part of an object or an area of terrain. Texture affects the tone and apparent coloration of things because of its absorption and scattering of light. Highly textured surfaces tend to appear dark and remain constant in tone regardless of the direction of view and lighting, whereas relatively smooth surfaces change from dark to light with a change in direction of viewing or lighting. The application of texture to an object often has the added quality of disrupting its shape and the shape of its shadow, making it more difficult to detect and identify as something foreign to the surroundings in which it exists. As an example, a surface having the same color but with heavy "nap" or texture is tall grass. Each separate blade is capable of casting a shadow upon itself and its surroundings. The light reflecting properties have been cut to a minimum. It will look and photograph dark gray. Looking straight down, the aerial observer sees all of the shadows, whereas a person on the ground may not. The textured surface may look light at ground level, but to the aerial observer the same surface produces an effect of relative darkness. The material used to conceal a person or an object must approximate the texture of the terrain in order to blend in with the terrain. Personnel walking or vehicles moving across the terrain will change the texture by mashing down the growth. Therefore, this will show up clearly from the air as vehicle tracks or foot paths.

e. Color:

(1) Pronounced color differences at close range distinguish one object from another. The contrast between the color of the object and the color of its background

Figure 28-7. Shadow.

Figure 28-8. Shadow Breakup.

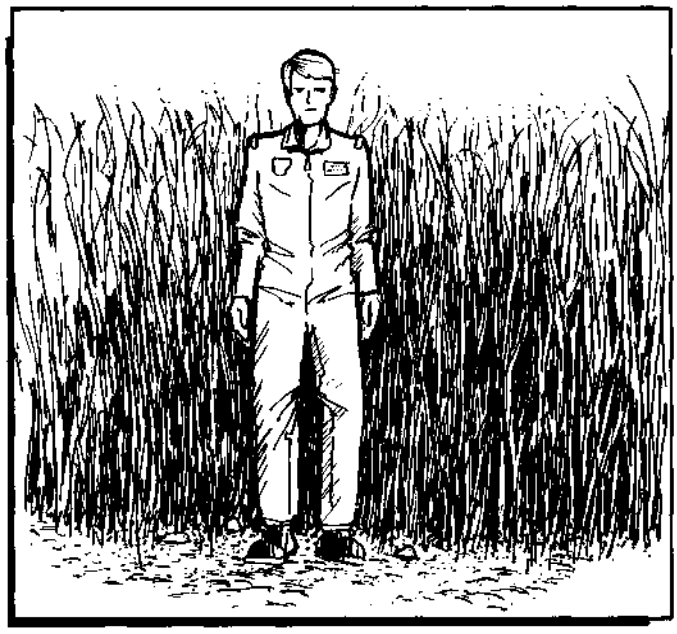

Figure 28-9. Contrast.

can be an aid to enemy observers. The greater the contrast in color, the more visible the object appears (figure 28-9).

(2) Color differences or differences in hue, such as red and green-yellow, become increasingly difficult to distinguish as the viewing range is increased. This happens because of atmospheric effects. Colors in nature, except for certain floral and tropical animal life, are not brilliant. The impression of the vividness of nature's colors results from the large areas of like colors involved and contrast of these areas with each other. The principal contrast is in their dark and light qualities. However, the dark and light color contrast does not fade out quickly and is distinguishable at greater distances. Therefore, as a first general principle, the camouflage should match the darker and lighter qualities of the background and be increasingly concerned with the colors involved as the viewing range is decreased or the size of the object or installation becomes larger. A second general rule to follow is to avoid contrasts of hues. This is especially true in areas with heavy vegetation. Light-toned colors, such as leaf bottoms, should be avoided as they tend to attract attention.

f. Tone. Tone is the amount of contrast between variations of the same color. It is the effect achieved by the combination of light, shade, and color. In a black-and-white photograph, the shades of gray in which an object appears is known as tone (figure 28-10). By adding texturing material to a smooth or shiny surface, the surface can be made to produce a darker tone in a photograph, because the textured surface now absorbs more light rays. Objects become identifiable as such because of contrasts between them and their background. Camouflage blending is the process of eliminating or reducing these contrasts. The principal contrast is that of tone; that is, the dark and light relationship existing between an object and its background. The two principal means available for reducing tone contrast are the application of matching or neutral coloration and the use of texturing to form disruptive patterns. Poorly chosen, disruptive patterns tend to make the object more conspicuous instead of concealing it.

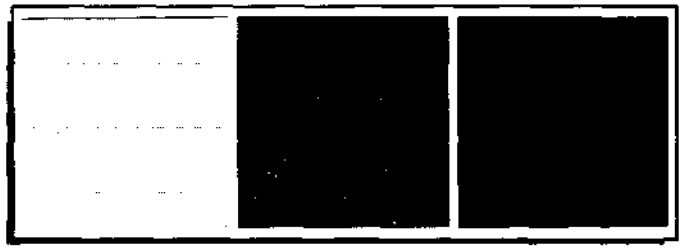

Figure 28-10. Three Half-Tone Blocks.

g. Movement. Of the eight factors of recognition, movement is the quickest and easiest to detect. The eye is very quick to notice any movement in an otherwise still scene. The aerial camera can record the fact that something has moved when two photographs of the same area are taken at different times. If an object has moved, the changed position is apparent when the two photographs are compared (figure 28-11).

h. Shine:

(1) Shine is a particularly revealing signal to an observer. In undisturbed, natural surroundings, there are comparatively few objects which cause a reflected shine. Skin, clean clothing, metallic insignia, rings, glasses, watches, buckles, identification bracelets, and similar items produce "shine." When light strikes smooth surfaces such as these, it may be reflected directly into the observer's eye or the camera lens with striking emphasis (figure 28-12).

(2) Such items must be neutralized by staining, covering, or removing to prevent their shine from revealing the location of evaders. This is especially true at night.

28-6. Principles and Methods of Camouflage:

a. No matter how applied, camouflage can be successful only by observing three fundamental principles. These basic principles of camouflage are choice of position, camouflage discipline, and camouflage construction.

b. When these factors have been considered, the evader is ready to begin application of various methods of deceiving the enemy. These methods are:

Figure 28-11. Movement.

(1) Hiding. The complete concealment of a person or object by physical screening.

(2) Disguising. Changing the physical characteristics of an object or person in such a manner as to fool the enemy.

(3) Blending. The arrangement of camouflage material on or about an object in such a manner as to make the object appear to be part of the background. To properly use these methods, three simple rules should be followed:

(a) First, the background should be changed as little as possible. When choosing a position to gain concealment, a background should be chosen that will visually absorb the elements of the position. Evaders should use a "natural" position if available. They should look for an existing position which can be used almost as is, such as a cave or thicket if there are many like it in the area. Isolated landmarks such as individual trees, haystacks, or houses should be avoided. They tend to attract attention and are likely to be searched first because they are so obvious. At times, by making use of background, complete concealment against visual and photographic detection may be gained with no construction. In terrain where natural cover is plentiful, this is a simple

Figure 28-12. Reflected Shine.

Figure 28-13. Background.

task. Even in areas where natural cover is scarce, concealment may be achieved through use of terrain irregularities. Regardless of the activity involved, evaders must always be mentally aware of their positions (figure 28-13).

(b) Secondly, the evader should use camouflage discipline. This means all of the factors of concealment are continuously applied.

-1. Daytime. Camouflage discipline is the avoidance of activity which changes the appearance of an area or reveals military objects to the enemy. A well-camouflaged position is only secure as long as it is well maintained. Concealment is worthless if obvious tracks point like directional arrows to the heart of the location or if signs of occupancy are permitted to appear in the vicinity. Tracks, debris, and terrain disturbances, are the most common signs of activity. Therefore, natural lines in the terrain should be used. If practical, exposed tracks should be camouflaged by brushing or beating them out. If leaving tracks is unavoidable, they should be placed where they will be least noticed and partially concealed (along logs, under bushes, in deep grass, in shadows, etc.). If tracks cannot be concealed, brushing them out will help them disintegrate quickly. Tying rags or brush to the feet will disguise boot prints and may help disguise them as refugee tracks (figure 28-14).

-2. Nighttime. Visual concealment at night is less necessary than in the daytime; however, noises at night are more noticeable. A simple act such as snoring may prove fatal. Calling to one another, talking, and even whispering should be kept to a minimum (figure

Figure 28-14. Camouflaging Tracks.

Figure 28-15. Sound.

28-15). But by far, the most important apsect of night discipline is light discipline. Lights at night not only disclose the evaders' position but also hinder the evaders' ability to detect the enemy. Even on the darkest nights, eyes grow accustomed to the lack of light in approximately 30 minutes. Everytime a match is lit or a flashlight is used, the eyes must go through the complete process of getting adjusted to the darkness again. Smoking and lights should be prohibited at night in areas in close proximity to the enemy because the light is impossible to conceal. Additionally, a cigarette light aggravates the situation by creating a reflection which completely illuminates the face. The smell of the evader's foreign tobacco would stand out even if the enemy is smoking.

-3. Evaders can lessen the effects of sound by simply taking precautions against sound production. They should avoid any sound-producing activity. Walking on hard surfaces should be avoided and full use should be made of soft ground for digging. Hand signals or signs should be used when possible during group travel. Individual equipment should be padded and fastened in such a manner as to prevent banging noises.

c. The evader should consider the following points regarding the use of camouflage:

(1) Take advantage of all natural concealment.

(2) Don't over-camouflage. Too much is as obvious as too little.

(3) When using natural camouflage, remember that it fades and wilts, so change it regularly.

(4) If taking advantage of shadows and shade, remember they shift with the Sun.

(5) Above all, avoid unnecessary movement.

(6) When moving, keep off the skyline; use the military crest (three-quarters' way up the hill).

(7) Do not expose anything that may shine.

(8) Break up outlines of manmade objects.

(9) When observing an area, do so from a prone position, while in cover.

(10) Match vegetation used as camouflage with that in the immediate locale, and when moving from position to position, change camouflage to blend with the new area's vegetation types.

28-7. Individual Camouflage:

a. At this point, with some of the general information about camouflage presented, it is time for a more detailed examination concerning individual camouflage.

b. Generally, individual camouflage is that personal concealment which evaders must use to deceive the enemy. Evaders must know how to use the terrain for effective concealment. Evaders must dress for the best concealment and carefully select their routes to provide for as much concealment as possible. All of the methods and techniques of camouflage addressed in this section have been successfully used by past evaders. If this information is learned and practiced by today's aircrew members (tomorrow's possible evaders), they will be more successful in evading and have a greater chance of returning to friendly forces.

c. Evaders should remember that in some areas they may have to engage in camouflage activities designed to deceive two types of enemy observation—ground and air. Many objects which are concealed from ground observation, may be seen from the air. This means the evader should camouflage for both types of observation.

d. Form is basic shape (body outline) and height. Three things which give an evader away in terms of form are to reveal outline of head and shoulders, to present straight lines of sides, and to allow the inverted "V" of the crotch and legs to be distinguishable. If staying in shadows, blending with background, adopting body positions other than standing erect, and other behavioral procedures are inadequate. They can be camouflaged by using "add-ons" such as branches or twigs to break up the lines. This addition of vegetation will also help an evader blend in with the background.

e. Effective concealment of evaders depends largely on the choice and proper use of background. Background varies widely in appearance, and evaders may find themselves in a jungle setting, in a barren or desert area, in a farmyard, or in a city street. Each location will require individual treatment because location governs every concealment measure taken by the individual. Clothing which blends with the predominant color of the background is desirable. There will be occasions when the uniform color must be altered to blend with a specific background. The color of the skin must receive individual attention and be toned to blend with the background.

f. There are certain general aspects of individual body and equipment camouflage techniques which apply almost anywhere. The evaders should take each of the following areas under consideration.

(1) Exposed Skin. The contrast in tone between the skin of face and hands and that of the surrounding foliage and other background must be reduced. The skin is to be made lighter or darker, as the case may be, to blend with the surrounding natural tones. The shine areas are the forehead, the cheekbones, nose, and chin.

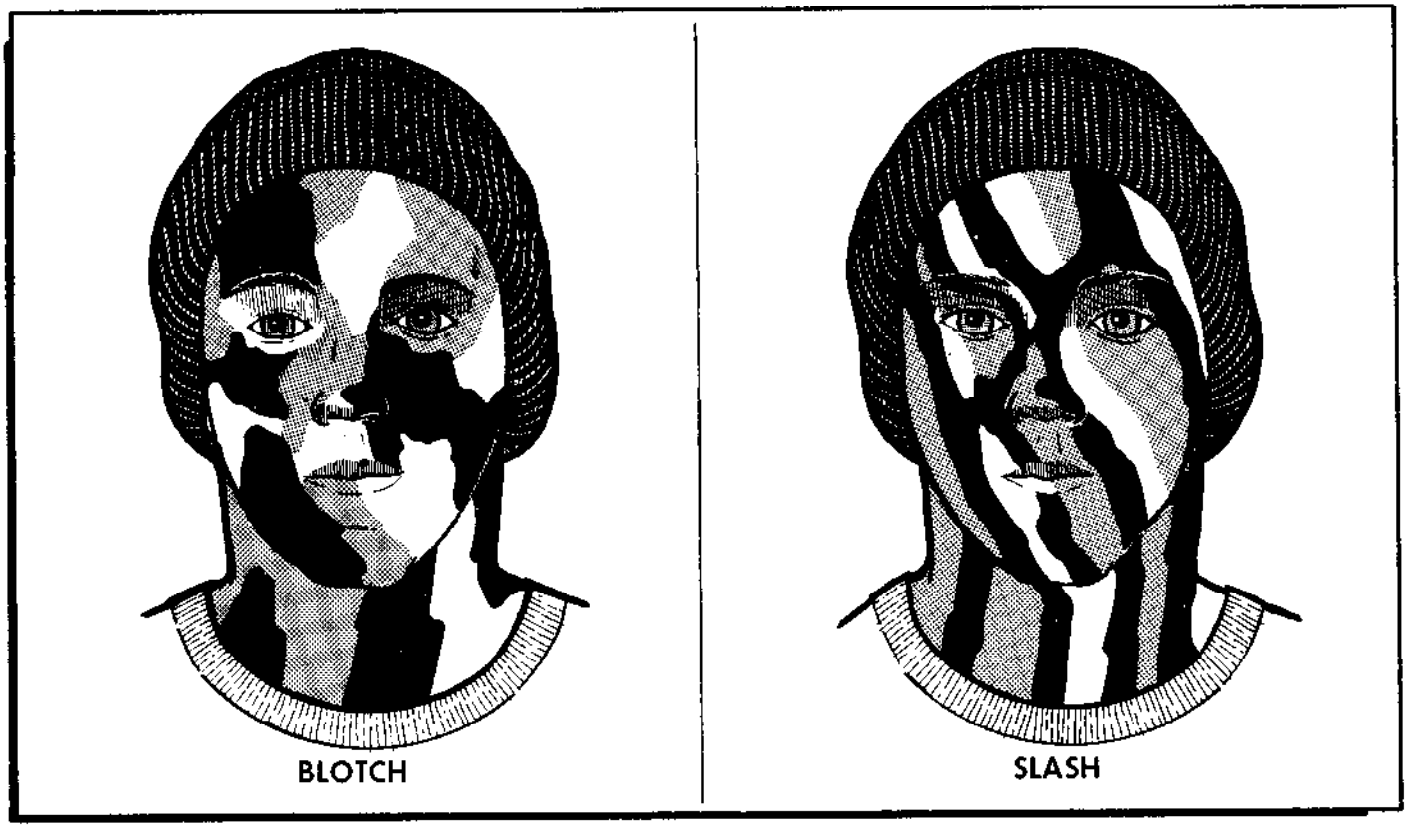

Figure 28-16. Camouflaged Faces.

These areas should have a dark color. The shadow areas such as around the eyes, under the nose and under the chin should have a light color. The hands, arms, and any other exposed areas of skin must also be toned down to blend with the surroundings. Burnt cork, charcoal, lampblack, mud, camouflage stick, berry stains, carbon paper, and green vegetation can all be used as toning materials.

(a) A mesh mosquito face net, properly toned down, is an effective method of breaking up the outlines of the face and ears.

(b) Two primary methods of facial camouflage have been found to be successful patterns. They are the "blotch" method for use in deciduous forests, and the "slash" method for use in coniferous forests (figure 28-16).

(c) Application of these two patterns are simply modified appropriately to whatever environment the evader is in. In the jungle, a broader slash method would be used to cover exposed skin; in the desert, a thinner slash; in barren snow, a wide blotch; and in grass areas, a thin type slash. To further break up the outline of the facial features, a flop hat or other loosely fitting hat may help. A beard that is not neatly trimmed may also aid the evader.

(d) When toning down the skin, evaders should not neglect to pattern all of the skin; for example, the back of the neck, the insides and backs of the ears, and the eyelids. Covering these areas may help somewhat, especially if there is a lack of other material to tone down the skin. Vegetation hung from the hat, collar buttoned and turned up, a scarf, or even earflaps may help. To cover the hands, evaders may use flight gloves, mittens, or loose cloth if unable to tone down wrists, backs of hands, and between fingers sufficiently. Evaders should watch for protruding white undergarments, T-shirt, long underwear sleeves, etc. They should also tone down these areas.

(e) Lack of hair or light colored hair requires some type of camouflage. This could include those applied to the skin or an appropriate hat, scarf, or mosquito netting.

(f) Odors in a natural environment stand out and may give evaders away. Americans are continually surrounded by artificial odors and are not usually aware of them. Human body odor would have to be very strong to be detected by ground troops (searchers) that have been in the field for long periods. The following odors should be of concern to the evader.

-1. Soaps and Shampoos. In combat areas, personnel should always use unscented toilet articles. Shaving cream, after shave lotion, perfume, and other cosmetics are to be avoided. The potential evaders should also realize that insect repellent is scented. They should try to use headnets, but if forced to use a repellent, the camouflage stick which has repellent in it is the least scented. Tobacco should not be used. The stain and odor should be removed from the body and clothing. Gum or candy may have strong or sweet smells—evaders should take care to rinse out their mouths after use. These odors, especially tobacco, can be detected at great distances.

-2. Smoke odors from campfires may permeate clothing, but if the potential searchers use fires for cooking and heat, they probably can't detect it on evaders.

(2) Clothing and Personal Items. These items require attention both before assignment to a combat mission and again if forced to evade. Prior preparation for the survivor may include:

(a) Ensuring that flight clothing does not smell of laundry products and is in good repair and not worn to the point where it shines or is faded.

(b) Checking zippers for shine and function.

(c) Checking rank insignia and patches for light reflection and color. Remove name tag, branch of service tag, and rank (whether they are stripes or metal insignia, bright unit patches, etc.) from the uniforms and place them either in the pack or in a secure pocket. Underwear should also be subdued for camouflage in the event that the outer layer is torn. Boots should be black but not shiny. Shiny eyelets should be repainted. Squeaky boots should be fixed or replaced. Sanitize pocket or wallet contents. Remove items which might aid in enemy exploitation attempts of the individual PW or PWs as a group; for example, credit cards, photographs, money, and addresses. Evaders should carry only those necessary pieces of identification which will prove a person is a US military member.

(3) Additional Clothing. Additional clothing may be desirable and located in flight clothing, such as hat, socks, scarf, and gloves. In an evasion environment, clothing and equipment need quick camouflaging attention. Anything to be discarded should be hidden at the initial landing site.

(4) Sanitizing Clothing. Clothing should be sanitized by removing anything bright or shiny. Evaders should consider camouflaging their clothing (to include boots) just as they would their skin. This is detrimental to the insulative qualities of their clothes but not as much as bullets or prison barbed wire if they are seen.

(5) Camouflaging Clothing:

(a) One principle of camouflage is to disrupt or conceal uniform color, straight lines, and squares—things rarely found in natural features. If ground-to-air signals are designed to exploit these visual characteristics, then evaders should certainly want to eliminate them.

(b) Evaders should reduce the tone of all equipment by smearing it with camouflage stick, mud, etc., or with whatever is available in a mottled pattern. In some instances equipment will have to be lighter in tone, in others, darker.

(c) To maintain the functional capability of clothing and equipment, it must be kept clean. In some areas, however, these items may be the only natural camouflage material survivors have to work with (such as desert regions). But, in areas where they have access to vegetation and the various dyes which can be made from vegetation, the vegetation should be used. In contrast to substances which soil the material and actually break down the fibers, dyes derived from grasses or plant sap (banana trees, ash trees, etc.) will offer the evader the toning material necessary to break up the solid green of a uniform and leave the fabric grit-free and still able to "breathe." The same saps which produce stains for cloth can be used to discolor metal objects. Banana tree sap, when left on the metal blade of a knife, will produce a blue-black stain which is a permanent discoloration. Trappers still boil their traps in ash tree chips and water to produce the blue-black, rust-inhibiting coloration to the tools of their trade.

g. All principles and techniques for care and use of clothing and equipment cannot be forgotten or ignored in an evasion situation, although some modifications may be necessary. A number of variables will influence what changes or omissions will be necessary.

(1) All cutting tools must be kept sharp. Evaders should try to coincide these noisy, yet essential, tasks with natural noise in the area (a downpour of rain for instance) or in a protected, noise-dampening area.

(2) Clothing must be kept clean if it is to protect a survivor from a harsh environment. Dirt-clogged, perspiration-soaked fibers will not give the insulating qualities of clean cloth. Clothing can be washed during the downpour of rain or possibly under the cover of darkness in a stream. Convenient, secluded puddles of water may afford the opportunity a survivor needs to clean clothing and equipment properly.

(3) Cooking and eating utensils must be kept clean on the inside to prevent dysentery and diarrhea. Simultaneously, the outsides can be toned down with soil, mud, etc., to camouflage them.

(4) Where metal pieces come into contact with one another, there should be padding between them so they will not inadvertently "clank" together. Evaders should place all items needed for environmental protection in the top of the pack where they will be most readily available. The rest of the gear can be used as padding around metal objects. In this manner, with everything stored inside the bundle (pack), it is secure from loss, damage, and enemy observation, as well as being readily available when needed. Evaders should also remove jewelry, watches, exposed pens, and glasses if possible. If glasses are required, hat netting or mask may help reduce shine.

(5) An evader's pockets should be secured and all equipment, including dog tags, arranged so that no jingle or rattle sounds are made. This can be done with cloth, vegetation, padding, or tape.

(6) Evaders should minimize the sound of clothing brushing together when the body moves. Moving in a careful manner can decrease this sound. Evaders should remember that camouflaged clothing and equipment alone won't conceal, but it must be used intelligently in accordance with the other principles of camouflage and movement. As one example, even if evaders are perfectly camouflaged for the arctic, there could still be problems. Because snow country is not all white, shadows and dark objects appear darker than usual. A snowsuit cannot conceal the small patches of shadow caused by the human figure, but that is not necessary if the background contains numerous dark areas. If the background does not contain numerous dark areas, maximum use is to be made of snowdrifts and folds in the ground to aid in individual concealment (figure 28-17).

(7) The concept of blending in with the background is indeed an important one for the evader to understand. One major point in blending with the background is not to show a body silhouette.

(8) Losing the body silhouette is done by making use of the shadows in the background. Evaders should be constantly aware of two factors—silhouette and shadow. From a concealment point of view, backgrounds consist of terrain, vegetation, artificial objects, sunlight, shadows, and color. The terrain may be flat and smooth, or it may be wrinkled with gullies, mounds, or rock outcroppings. Vegetation may be dense jungle growth or no more than small patches of desert scrub growth. The size of artificial objects may range from a signpost to a whole city block. There may be many colors in a single background, and they may vary from the almost black of a deep woods to the sand pink of some desert valleys. Blending simply means the matching with as many of these backgrounds as possible and avoiding contrast. If it is necessary for evaders to be positioned in front of a contrasting or fixed background, they must be aware of their position and take cover in the shortest possible time. The next point to which they will move for concealment must be selected in advance and reached as quickly as possible.

(9) As in the daytime, silhouette and background at night are still the vital elements in concealment (figure 28-18). A silhouette is always black against a night sky, and care must be taken at night to keep off the skyline. On moonlit nights, the same precautions must be taken as in daylight. It should be remembered that the position of the enemy observer, and not the topographic crest, fixes the skyline. At night, sound is an amplified, revealing signal. Movement must be careful, quiet, and

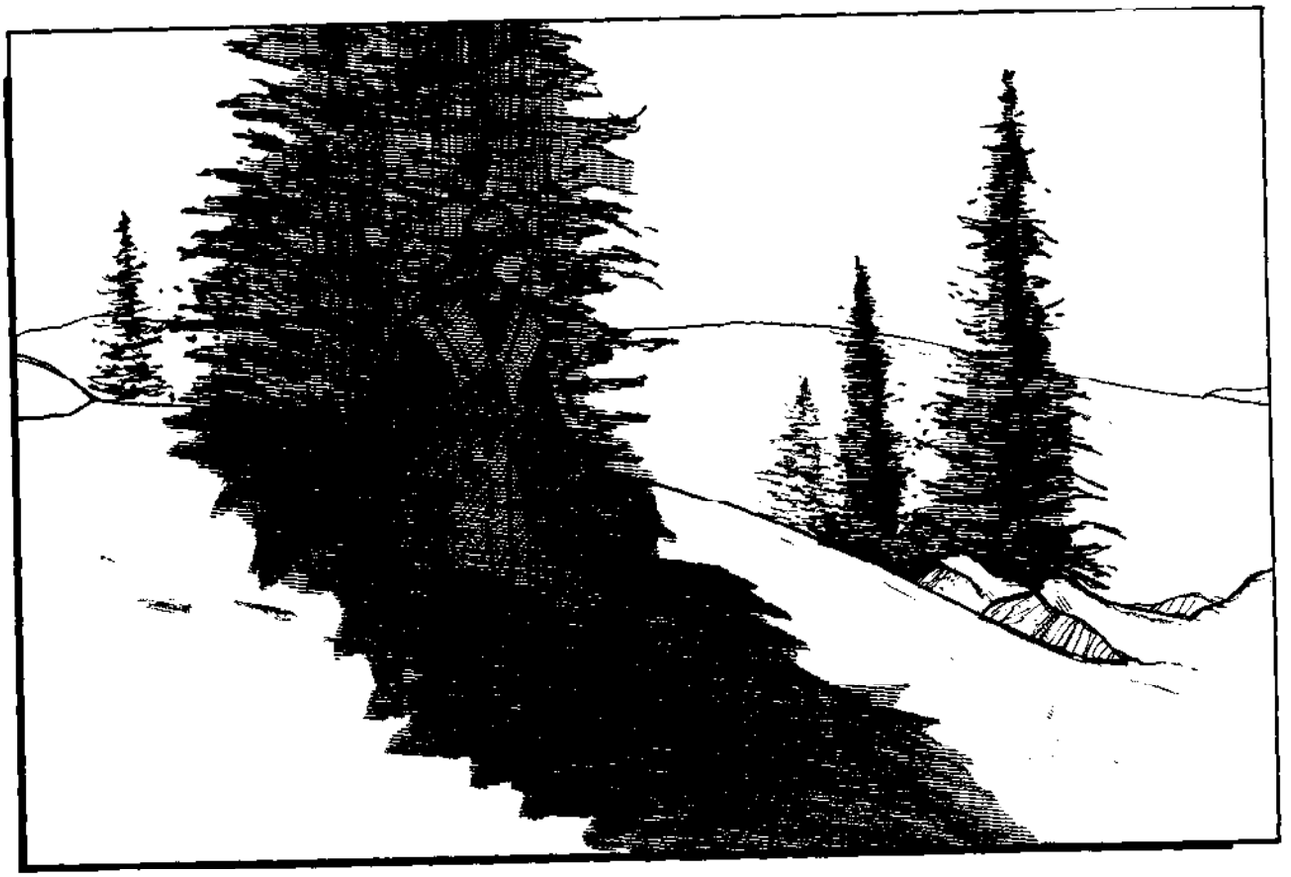

Figure 28-17. Arctic Travel.

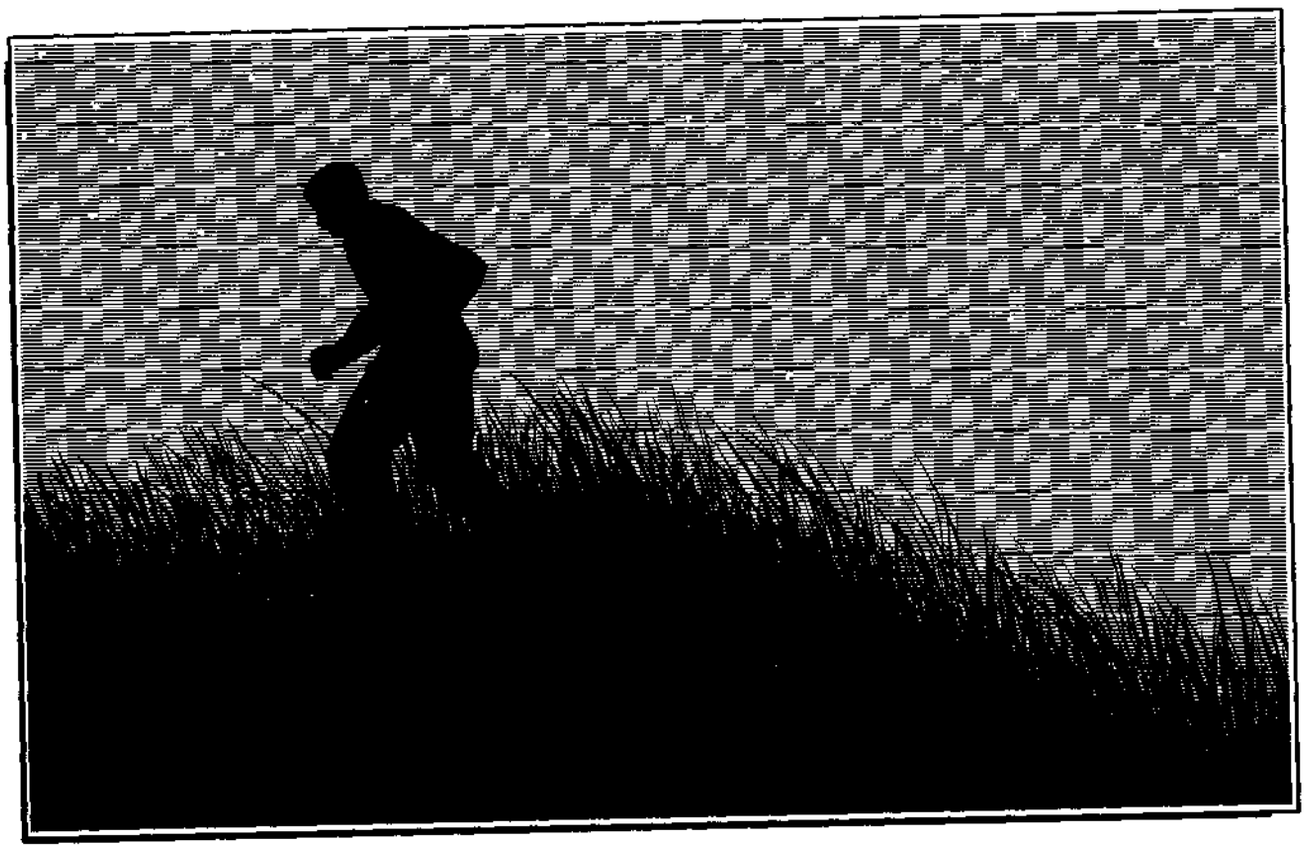

Figure 28-18. Silhouetting.

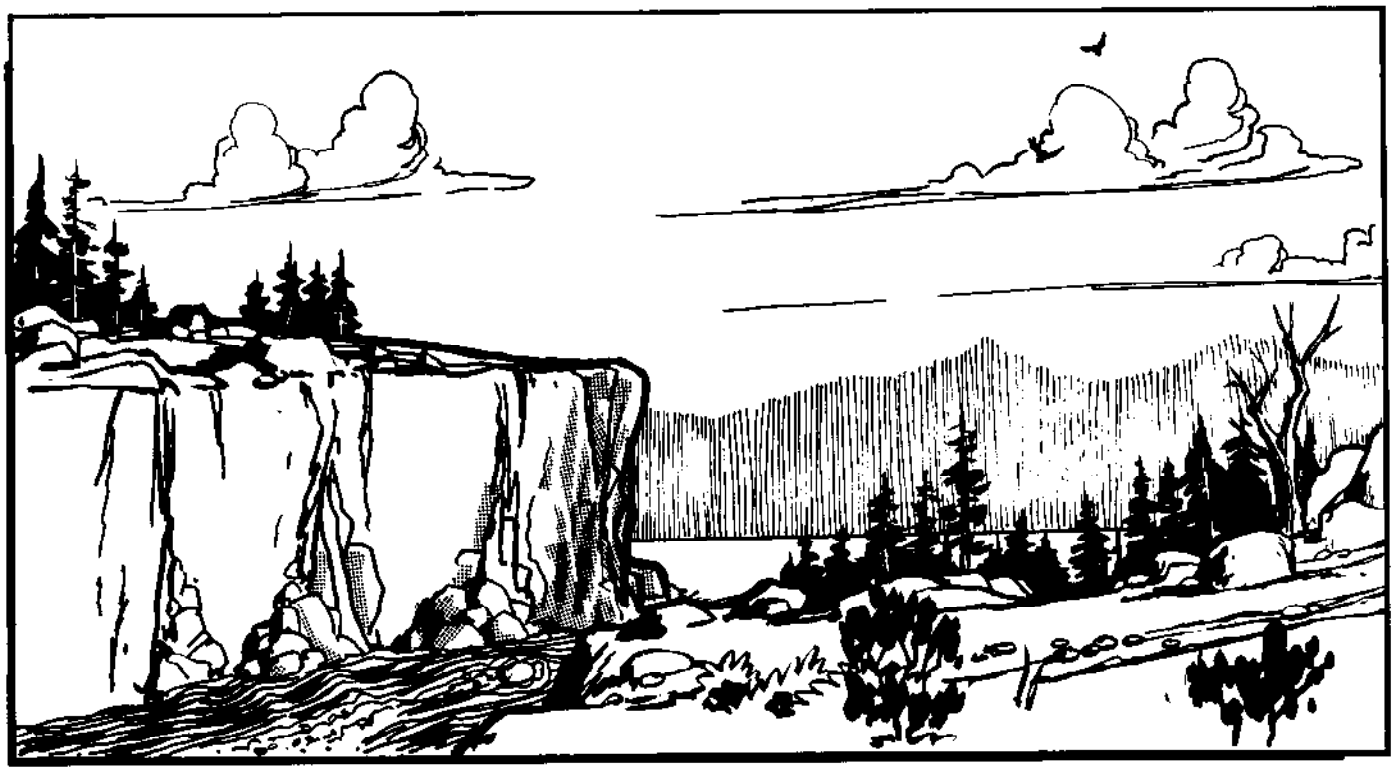

Figure 28-19. Natural Materials.

close to the ground. If the pop of a flare is heard before the illuminating burst, evaders must drop to the ground instantly and remain motionless. If they are surprised by the light, they must freeze in place with their faces down.

28-8. Concealment in Various Geographic Areas:

a. When not otherwise specified, temperate zone terrain is to be assumed in this section. Desert, snow, and ice areas are mostly barren and concealment may require considerable effort. Jungle and semitropical areas usually afford excellent concealment if the evader employs proper evasion techniques.

b. First, some general observations and rules regarding the addition of vegetation to the uniform and equipment. The cycles of the seasons bring marked changes in vegetation, coloring, and terrain pattern requiring corresponding changes in camouflage. Concealment which is provided in wooded areas during the summer is lost when leaves fall in the autumn. This will create a need for additional camouflage construction. Also, vegetation must be of the variety in the evaders' immediate location. It must be changed if it wilts or the evaders move into a different vegetation zone. Evidence of discarding the old and picking the new should be hidden. The vegetation should not be cut, this will give evidence of human presence.

c. Any type of material indigenous to the locality of the evaders may be classified as natural material. Natural materials consist of foliage, grasses, debris, and earth. These materials match local colors and textures and when properly used are an aid against both direct and indirect observation. The use of natural materials provides the best type of concealment. The chief disadvantage of natural foilage is that it cannot be prepared ahead of time, is not always available in usable types and quantities, wilts after gathering, and must be replaced periodically. Foliage of coniferous trees (evergreens) retains its camouflage qualities for considerable periods, but foliage that sheds leaves will wilt in a day or less, depending on the climate and type of vegetation (figure 28-19).

(1) The principal advantage in using live vegetation is its ability to reflect infrared waves and to blend in with surrounding terrain. When vegetation is used as garnishing or screening, it must be replaced with fresh materials before it has wilted sufficiently to change the color or the texture. If vegetation is not maintained, it is ineffective. Thorn bushes, cacti, and other varieties of desert growth retain growing characteristics for long periods after being gathered.

(2) The arrangement of foliage is important. The upper sides of leaves are dark and waxy; the undersides are lighter. In camouflage, therefore, foliage must be placed as it appears in its natural growing state, top sides of leaves up and tips of branches toward the outside of the leaves (figure 28-20).

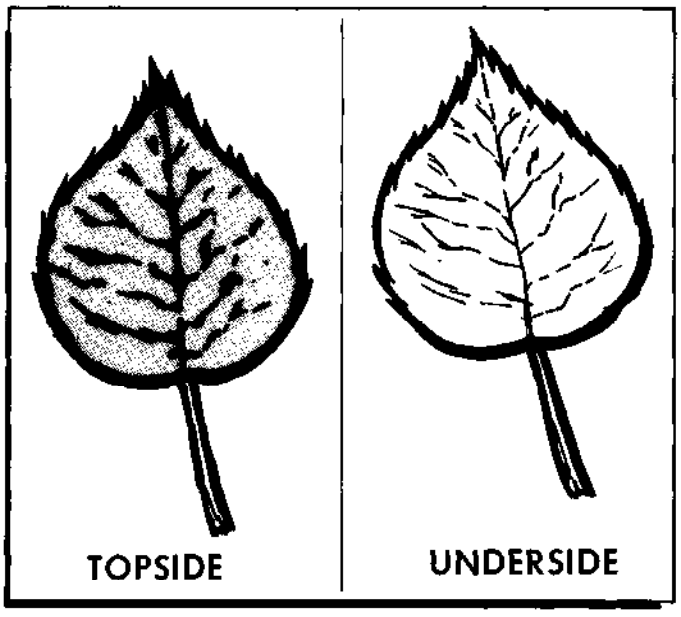

Figure 28-20. Dark and Light Leaves.

(3) Foliage gathered by survivors must be matched to existing foliage. For example, foliage from trees that shed leaves must not be used in an area where only evergreens are growing. Foliage with leaves that feel leathery and tough should be chosen. Branches grow in irregular bunches and, when used for camouflage, must be placed in the same way. When branches are placed to break up the regular, straight lines of an object, only enough branches to do this should be used. The evader must adopt principles that apply and know that the enemy also applies these principles.

(4) When vegetation is applied to the body or equipment of evaders, it must be secured to the clothing or equipment in such a way that:

(a) Any inadvertent movement of material will not attract attention.

(b) It appears to be part of the natural growth of the area; that is, when the evaders stop to hide, the light undersides of the vegetation are only visible from beneath. After evaders complete their camouflage, they should inspect it from the enemy's point of view. If it does not look natural, it should be rearranged or replaced.

(c) It does not fall off at the wrong moment and leave evaders exposed, or it should not show evidence of the evaders' passage through the area.

(5) Too much vegetation can give evaders away.

(6) If cloth material is used like vegetation to break up shape and outline and to help blend in with the environment, there are some points evaders should be aware of. Cloth can be used successfully when wrapped around equipment or designed into loose, irregular shaped clothing or accessories (figure 28-21).

(a) Some of the materials evaders may use are:

-1. The colors (green, brown, white) of parachute materials plus the harness.

-2. Excess clothing the evader may have prepacked (scarf, bandana, etc.).

-3. Burlap, when found. It is used in battle areas in the form of sandbags.

(b) Most artificial material is versatile, but it can have drawbacks:

-1. Parachute material, for example, tends to shine and the unraveled edges may leave fine filaments of nylon on the ground as evidence of evaders in an area. Parachute material is very lightweight. A sudden breeze might cause it to move when movement is not desirable.

-2. A white suit of parachute material is excellent for winter evasion over snow and ice environments. The time required to fabricate this suit should be considered. (NOTE: Shadows cast on the snow cannot be

Figure 28-21. Breakup of Shape and Outline.

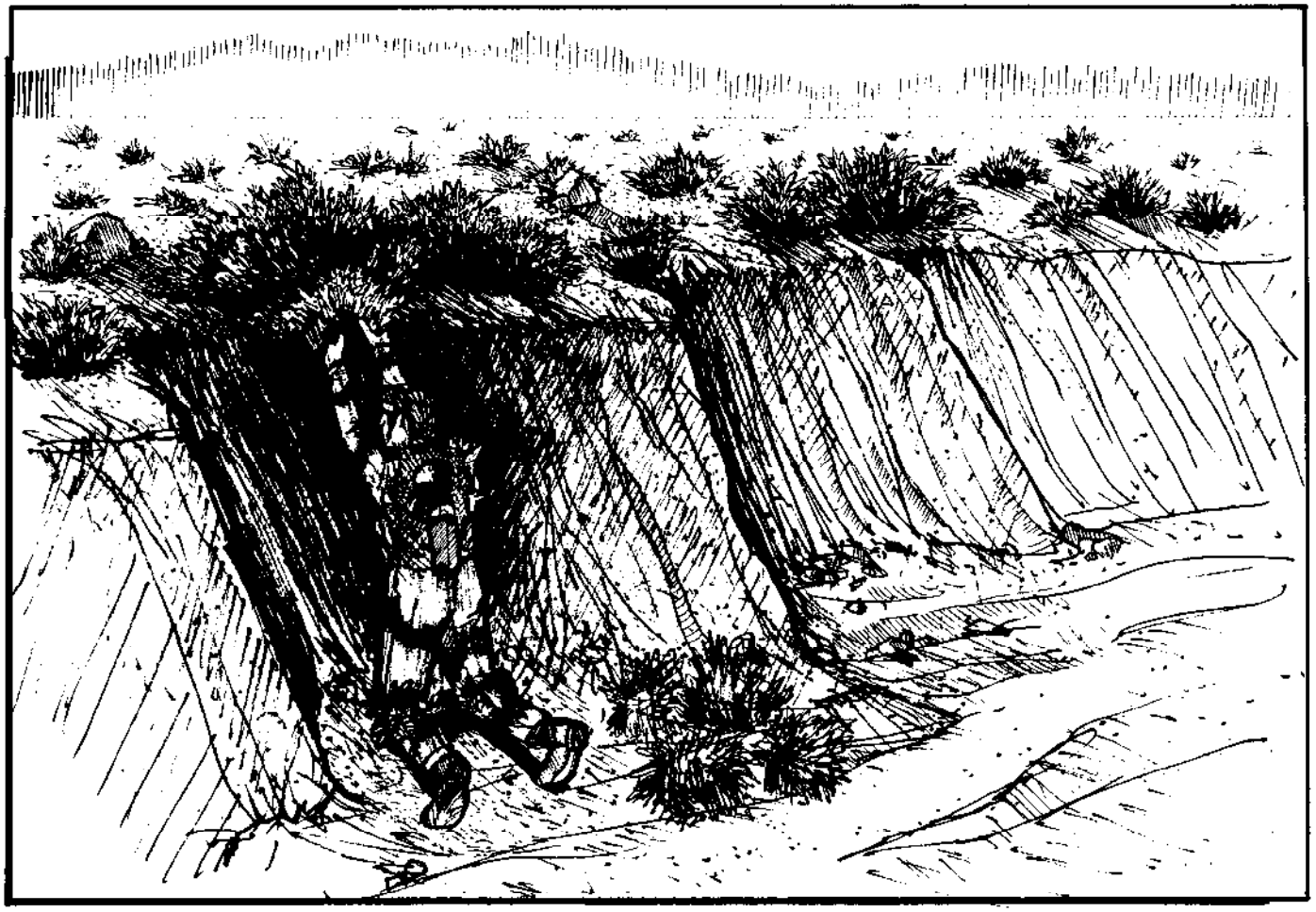

Figure 28-22. Concealment by Shadows.

Figure 28-23. Above Timberline.

camouflaged; they must be masked by terrain or other shadows.)

28-9. Concealment Factor for Areas Other Than Temperate:

a. Desert. Lack of natural concealment, high visibility, and bright tone (smooth texture) all emphasize the need for careful selection of a position for a campsite. Deep shadows in the desert, strict observance of camouflage discipline, and the skillful use of deception and camouflage materials aid in concealing evaders in a desert area.

(1) Deserts are not always flat, single-toned areas. They are sometimes characterized by strong shadows with heavy broken terrain lines and sometimes by a mottled pattern. Each type of desert terrain presents its own problems. When evaders and their shelters are located in the desert, their shadows are inky black and in strong contrast to their surroundings and are extremely conspicuous. To minimize the effect of these shadows when possible, use concealment that is afforded by the shadows of deep gullies, scrub growth, and rocks (figure 28-22).

(2) Many objects which cannot be concealed from the air can be effectively viewed from the ground. Even though these objects are observed from the air, lack of reference points in the terrain will make them difficult to locate on a map.

b. Snow and Ice. From the air, snow-covered terrain is an irregular pattern of white, spotted with dark tones produced by objects projecting above the snow, their shadows, and irregularities in the snow-covered surface such as valleys, hummocks, ruts, and tracks. It is necessary, therefore, to make sure dark objects have dark backgrounds for concealment to control the making of tracks in the snow and to maintain the snow cover on camouflaged objects.

(1) Mountain Areas Above Timberline and Arctic Areas. Common characteristics include an almost complete snow cover with a minimum of opportunities for concealment. Only a few dark objects protrude above the snow except for rugged mountain peaks (figure 28-23).

(2) Mountain Areas Below the Timberline and Subarctic Areas. Common characteristics of these areas are forests, rivers, lakes, and artificial features such as trails and buildings. The appearance of the area is irregular in pattern and variable in tone and texture (figure 28-24).

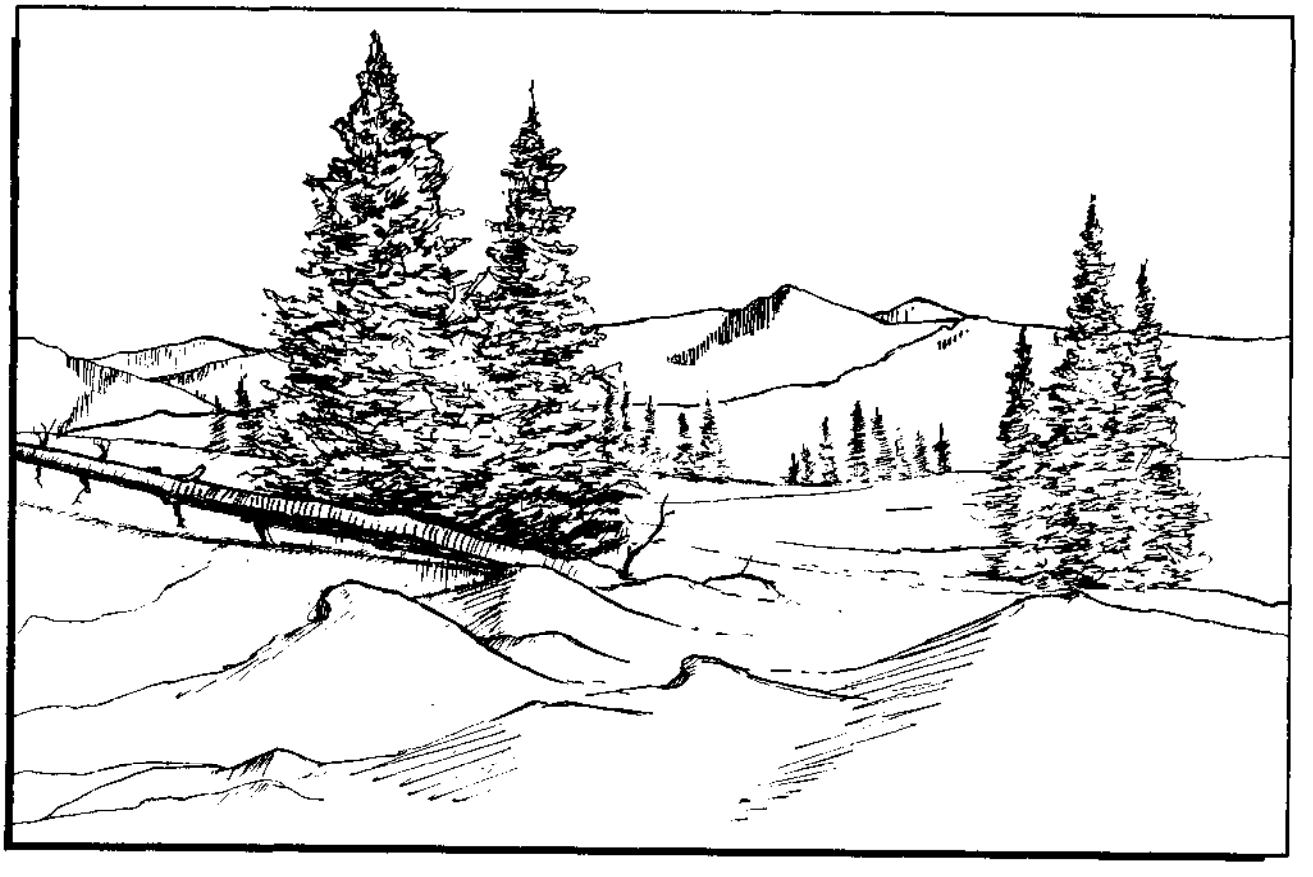

Figure 28-24. Below Timberline.

(3) Areas Between the Subarctic Zone and the Southern Boundary of the Temperate Zone. These have the same characteristics as mountain areas below the timberline and subarctic areas.

c. Blending With Background in Snow and Ice Terrains. No practical artificial material has yet been developed which will reproduce the texture of snow sufficiently well to be a protection against recognition by aerial observers. Concealment from direct ground observation is relatively successful with the use of white "snowsuits," white pants, and whitewash; these measures offer some protection against aerial detection. (White parachute cloth should also be considered.)

(1) People evading in snow-covered frozen areas should wear a complete white camouflage outfit. A white poncho-like cape can be made easily from parachute material (figure 28-25).

(2) A pair of white pants will normally be sufficient in a heavily wooded area. However, following or during a heavy snowfall when the trees are well covered with snow, the wearing of a complete white camouflage suit is necessary to blend in with the background. Other equipment, such as packs, should also be covered with white material.

d. Camouflage Checklist. The following checklist can be used to remember ideas concerning camouflage and to determine the completeness of individual camouflage application:

(1) Effective concealment is protection from hostile observation from the ground as well as from the air.

(2) Natural terrain lines are to be used for help in concealing the evader when possible.

(3) Every possible feature of the terrain should be used for concealment.

(4) A silhouette against the sky should be avoided.

(5) Every effort should be made to reduce tone contrast and eliminate shine.

(6) Evaders should be especially careful at night due to infrared and low light detection equipment which may be used by the enemy. Keeping close to the ground and using terrain masking for concealment provide the best protection.

28-10. Camouflage Techniques for Shelter Areas:

a. As used in this section, the word shelter refers to the concept of personal protection synonymous with the terms refuge, haven, or retreat. Readers should not visualize the word "shelter" when mentioned in this text to mean a dwelling traditionally occupied by survivors in nontactical situations; such as, tents, cabins, or other such places of habitation. While it is true that these structures can and do provide safety and relief, it must

Figure 28-25. Evading in Snow-Covered Areas.

Figure 28-26. Shelter.

be understood that where an evasion situation is concerned, concealment, not personal comfort or convenience, will be the primary concern for the evader who seeks "shelter" or sanctuary (figure 28-26).

b. Besides resting or sleeping, there are a number of reasons why survivors may need to conceal themselves for varying amounts of time. They may need to take care of problems concerning personal hygiene, adjustment of clothing, maintenance or alteration of camouflage, triangulation for position determination, food and water procurement, etc. Concealment cannot be overstressed in respect to the areas survivors (evaders) may select for shelter. One important reason evaders must be able to select secure areas for refuge is to avoid and prevent detection by the enemy. This is especially important if the haven selected is to be used for resting or sleeping. Anyone who is resting or sleeping will not be totally alert, and added precautions may be necessary to maintain security. Another factor to consider, since evaders are also survivors, is to protect themselves from the elements as much as possible.

c. At no time will evaders be able to safely assume they are free from the threat of either ground or aerial observation. Therefore, not only is the shelter area and type determined by the needs of the moment (enemy

Figure 28-27. Natural Shelter.

presence, etc.), but consideration must also be given to the terrain and climatic conditions of the area. Evaders must constantly be aware of how long they may have to remain in the area and, most importantly, what type of enemy observation may be employed.

d. The shelters may be naturally present, or they may be those which are "assembled" and camouflaged by the evaders. Full use must be made of concealment and camouflage no matter what types of shelter areas are selected. The use of the natural concealment afforded by darkness, wooded areas, trees, bushes, and terrain features are recommended; however, any method used for disguise or hiding from view will increase the chances for success. There is much for evaders to consider concerning the many facets of evasion shelter site selection if they expect to establish and maintain the security of their area (figure 28-27).

e. Evaders should locate their shelter areas carefully. They should choose areas which are least likely to be searched. They must be in the least obvious locations. The chosen areas should look typical of the whole environment at a distance. They should not be near prominent landmarks. Areas that look bland get a cursory glance. The areas should also be those least likely to be searched; for example, rough terrain and thickly vegetated areas. The shelter sites should also be situated so that in the event of impending discovery, the evaders will be able to depart the area via at least one concealed escape route. The shelter areas should never be in areas which may trap them if the enemy discovers the places of concealment.

f. Evaders should choose natural concealment areas—a "natural shelter." Examples include small, concealed caves, hollow logs, holes or depressions, clumps of trees, or other thick vegetation (tall grass, bamboo, etc.). The site should have as much natural camouflage as possible. There should be cover on all sides; this includes natural formations or vegetation which can also protect evaders from aerial observations. The site should be as concealed as possible with a minimum of work. Sites chosen this way will make concealment easier and require less activity and movement. This is most important if the evaders are close to population centers or if the enemy is present.

g. The evaders should attempt to stay as high as possible and to select concealment sites near the military crest of a hill if cover is available. Noises from ridge to ridge tend to dissipate. Whispers or other sounds made in a valley tend to magnify as listeners get further up a hill. Shelter areas located on a slope are subject to higher daytime and lower nighttime winds, thereby minimizing the chance of detection through the sense of smell.

h. If possible, evaders should be in such a position in the shelter area that shadows will fall over the side of the area throughout the day. This can best be done in heavy brush and timber.

i. Evaders should try to locate alternate entrance and exit routes along small ridges or bumps, ditches, and rocks to keep the ground around the shelter area from becoming worn and forming "paths" to the site. They should avoid staying in one area so long that it has the appearance of being "lived in." Evaders should try to stay away from and out of sight of any open areas; examples are roads and meadows. Several miles distance from those may be desirable.

j. Waterways such as lakes, large rivers, and streams, especially at the junctions, are dangerous places. Power and fence lines or any prominent landmarks may indicate places where people may be. Evaders will want to stay clear of these areas. The enemy may patrol bridges frequently. Evaders should avoid any areas close to population centers. The evaders should be able to observe the enemy and their movements and the surrounding country from this hiding area if at all possible. If any assemblage of camouflage is necessary at the shelter site, evaders should keep in mind that they should always "construct" to blend. They should match the shelter area with natural cover and foliage, remembering that over-camouflage is as bad as no camouflage. Natural materials should be taken from areas of thick growth. Any place from which materials have been taken should be camouflaged. The following is an easy to remember acronym (BLISS) for evasion shelter principles:

B - Blend.
L - Low silhouette.
I - Irregular shape (outline).
S - Small.
S - Secluded location.

k. Other facilities evaders may use, such as latrines, caches, garbage pits, etc., must be located and camouflaged in the same manner as the shelter sites. Evaders should avoid forming a line of installations which lead from point to point to their location. They should dogleg through ground cover to use concealment to its best advantage. A dry, level sleeping spot is ideal, but the ideal spot to provide nonvisibility and comfort may be difficult to find. Evaders must have patience and perseverance to stay hidden until danger has passed or until they are prepared and rested enough to safely move on. They must be constantly on the watch for shelter areas which need little or no improvement for camouflage, protection from the elements, or security.

28-11. Firecraft Under Evasion Conditions:

a. Whether or not to build a fire under evasion conditions is indeed a difficult decision evaders must make at times. Basically, fire should only be used when it is absolutely necessary in a life or death situation. Potential evaders must understand that the use of fire can greatly increase the probability of discovery and subsequent capture.

Figure 28-28. Fire.

b. If a fire is required, location, time, selection of tinder, kindling, and fuel, and construction should be major considerations.

(1) Evaders should keep the fire as inconspicuous as possible. The location of an evasion fire is of primary importance. All of the small evasion-type fires must be built in an area where the enemy is least likely to see them. If possible, in hilly terrain with cover, the fire should be built on the side of a ridge (military crest). No matter where the fire is built, it should be as small and smokeless as possible (figure 28-28).

(2) Fires are easier to disguise and will blend in better during the times of dawn and dusk and during times of bad weather. At these times, there is a haze or vapor trap that hinges in and around hills and depressions and is prevalent on the horizon. Any smoke from the fire will be masked by this haze in the early morning and at sunset. This is the time when the local populace is most likely to have their cooking or heating fires lit. Another method of disguising the smoke from a fire is to build it under a tree. The smoke will tend to dissipate as it rises up through the branches, especially if there is thick growth or the boughs are low hanging. If this is not possible, it will be helpful to camouflage the fire with earthen walls, stone fences, bark, brush, or snow mounds to block the light rays and help disperse the smoke.

(3) The best wood to use on an evasion fire is dry, dead hardwood no larger than a pencil with all the bark removed. This wood will produce more heat and burn cleaner with less smoke. Wood that is wet, heavy with pitch, or green will produce large amounts of smoke. When the wood has been gathered, evaders should select small pieces or make small pieces out of the wood collected. Small pieces of wood will burn more rapidly and cleanly thus reducing the chance of smoldering and creating smoke. The wood selected should be stacked so the fire gets plenty of air as ventilation will make the fire burn faster with less smoke.

c. One type of evasion fire which has the capability of being inconspicuous is called the Dakota Hole Fire (figure 28-29). After selecting a site for the fire hole, a "fireplace" must be prepared. This is done by digging two holes in the ground, one for air or ventilation and the other to actually lay the fire in. These holes should be roughly 8 to 12 inches deep and about 12 inches apart with a wide tunnel dug to connect the bottoms of the holes. The depth of the holes depends on the intended use of the fire. Place dirt on a piece of cloth so it can be used to rapidly extinguish the fire and conceal the fire

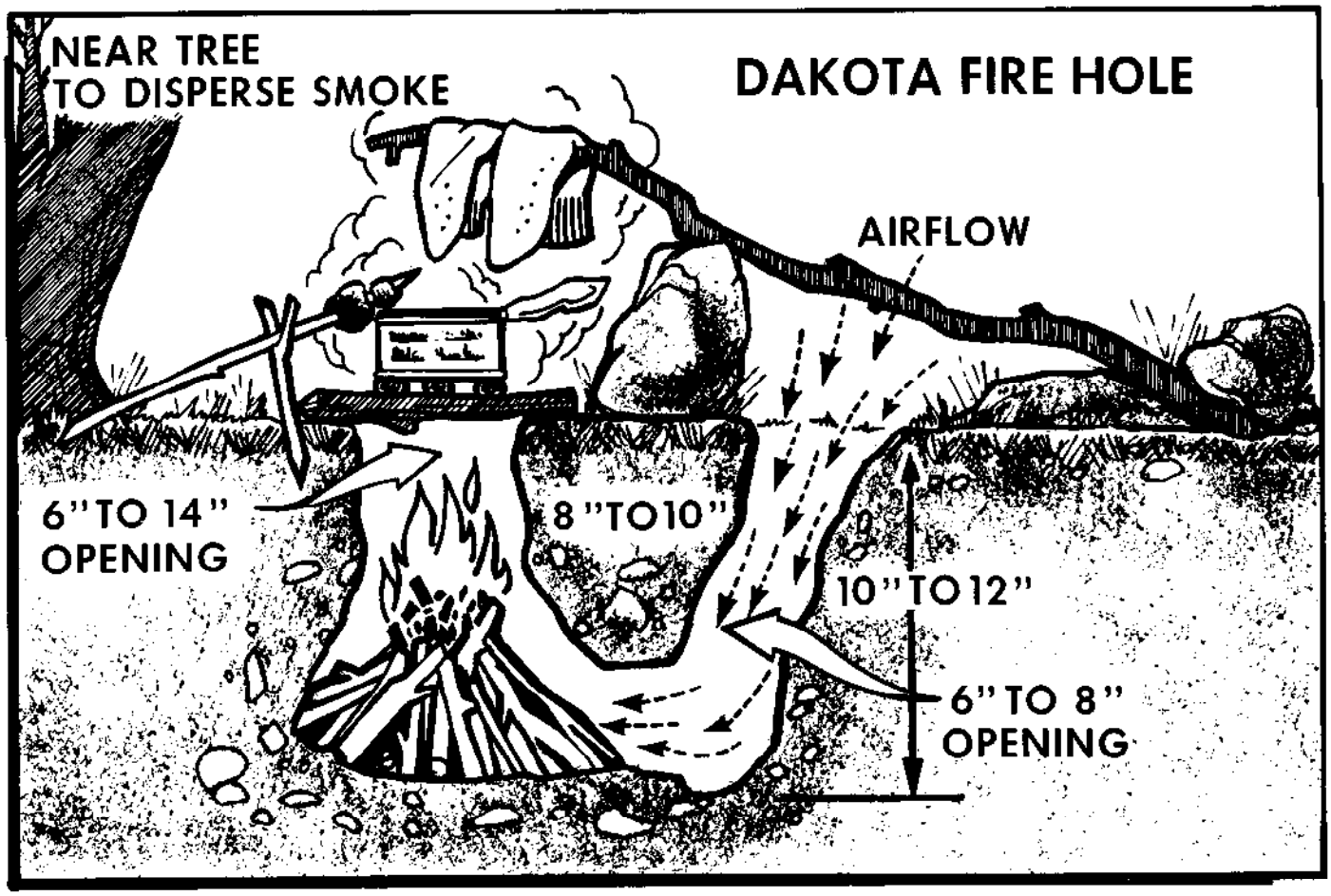

Figure 28-29. Dakota Hole.

site. Evasion fires should be small. Evaders who build these fires should strive to keep the flame under the surface of the ground. Initially the fire may appear to be smoking due to the moisture in the ground. At night the area may glow, but there should be no visible flame. A fire built in a hole this way will burn fast as all the heat is concentrated in a small area. This is a good type of fire for a single evader as opposed to many persons taking turns cooking over this one hole. Everything about the evasion situation will have to be examined before deciding which fire configuration would be most useful if a fire must be built.

d. Another good evasion fire is the trench fire. This fire is built by digging below the earth's surface 8 to 12 inches in an elongated pattern. The length depends on how many people need to use it. This is a fire more suited to meeting the needs of a small group. The fire should not crowd either end of the excavation, as it must be able to "draw" an adequate amount of air to help it burn hot and eliminate smoke (figure 28-30).

e. An evasion fire can also be built just below the ground cover (figure 28-31). Here the emphasis is on quick concealment of the area. It should be kept in mind, however, that some type of screen should be built

Figure 28-30. Trench Fire.

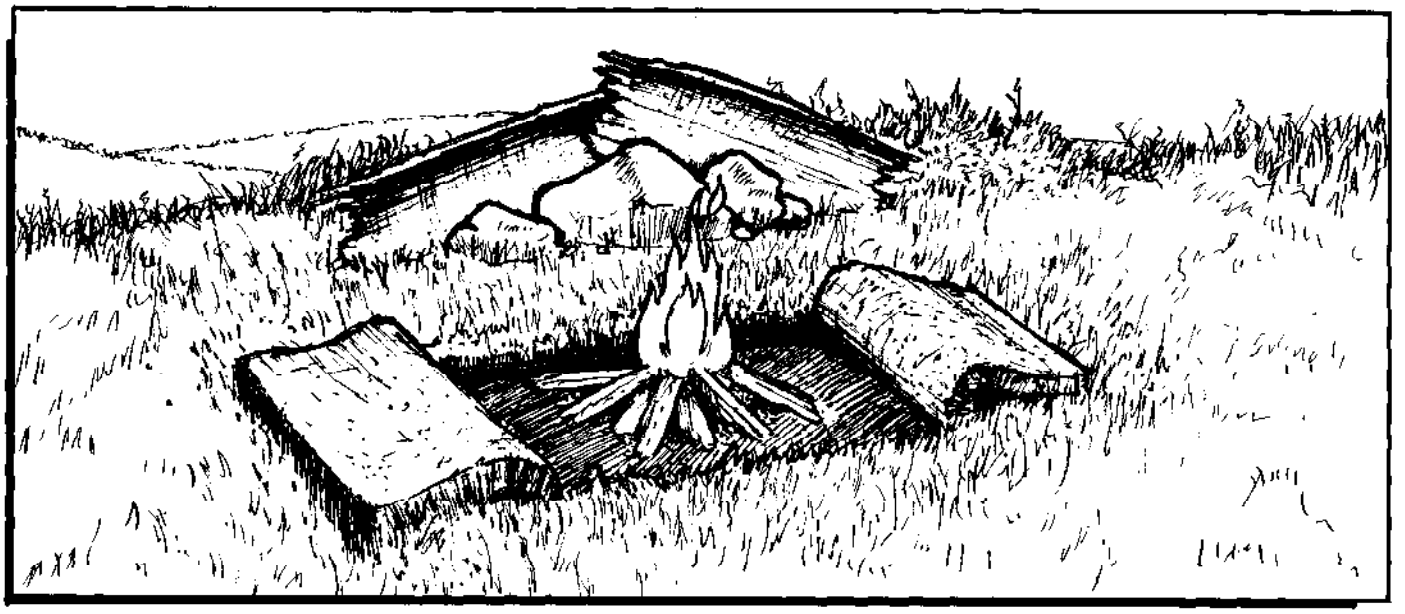

Figure 28-31. Fire Concealment.

to hide the flames. The small fire is built on the bare ground after a layer of sod or earth has been sliced and rolled back. After use, evaders simply scatter the fire remnants over the bare area and roll or fold the piece of ground back into place.

f. If evaders are in areas where holes can't be dug or sod can't be lifted, they will have to settle for some type of screen around the fire. They should also keep the fire small and finish using it quickly.

g. All traces of the fire should be removed. Unburned firewood should be buried. Holes should be totally filled in. Placing the soil on a holder, such as a map or piece of equipment, will aid in replacing it. This way there will be no leftover dirt patches on the ground after the holes were filled in. Once evaders feel all available measures have been taken to obliterate any leftover evidence, they should move out of the area if possible. Since evaders can never be positive the fire wasn't detected, they must assume it was spotted and take all necessary precautions.

28-12. Sustenance for Evasion:

a. As previously stated in this section, not only do evaders face the problem of remaining undetected by the enemy, they must also have the knowledge which will enable them to "live off the land" as they evade. They must be prepared to use a wide variety of both "wild" and domestic food sources, obtain water from different sources, and use many methods for preparing and preserving this food and water (figure 28-32).

b. No matter what the circumstances of the evasion situation are, evaders should never miss an opportunity to obtain food. Ordinarily food will be obtained from wild plant and animal sources. If possible, evaders should stay away from domestic plants (crops) and animals. Using wild animals and plant sources for food will reduce the probability of capture.

c. Animal foods are a prime source of sustenance for evaders, having more nutritional value for their weight than do plant foods. Evaders may obtain enough animal food in one place to last for several days while they travel.

d. There are several ways which evaders may procure animal foods (trapping or snaring, fishing, hunting, and poisoning). A few modifications should be considered regarding the use of traps and snares in a tactical situation even though the same basic principles apply in both tactical and nontactical environments.

(1) Because small game is more abundant than is large game in most areas, evaders should confine traps and snares to the pursuit of small game. There are other advantages to restricting trap and snare size. Evaders will find it is easier to conceal a small trap from the enemy. Small animals make less noise and create less disturbance of the area when caught.

(2) Conversely, there are also a few disadvantages pertaining to the use of snares or traps during evasion episodes. Two disadvantages are: Evaders must remain in one place while the snares are working, and there may be some disturbance of the area where materials have been removed.

e. Fishing is another effective means of procuring animal food. Fish are normally easy for evaders to catch, and they are easy to cool.

(1) There are several methods which evaders may use to catch fish. A simple hook and line is one of these methods; another is a "trotline." Evaders may construct a trotline by fixing numerous hooks to a pole and by

Figure 28-32. Sustenance for Evasion.

sliding it into the water from a place of concealment (figure 28-33).

(2) Nets and traps may also be used; however, they should be set below the water line to avoid detection. Spearing is another option. Here again, exposure in open waterways can be very dangerous to the evader.

(3) "Tickling" the fish (figures 28-34, 28-35, and 28-36) is also effective if evaders can remain concealed. This method requires no equipment to be successful. The main disadvantage to fishing is people live by water bodies and travel on them. This greatly increases the chances of being detected. (NOTE: Caution should be used when tickling fish in areas with carnivorous fish or reptiles.)

f. Weapons may be used by evaders to procure animals. The best weapons are those which can be operated silently, such as a blowgun, slingshot, bow and arrow, rock, club or spear. These should be used primarily against small game. One major advantage of using weapons is game can be taken while evaders travel. Because of noise, firearms should never be used in an evasion situation.

g. Plant foods are very abundant in many areas of the world. Evaders may be able to procure plant food types that require no cooking. One advantage of procuring plant foods during evasion is that by collecting natural fruits and nuts, evaders can remain deep in unpopulated areas. In some areas, it is possible to find old garden plots where vegetables may be obtained. When possible, select foods which can be eaten raw. (Refer to chapter 18 - Food.)

(1) The disadvantage of plant food procurement is that evaders may not be the only ones looking for food. The natives of the country could also be out looking for food. If natives know of a good area, they may visit that place many times. If evaders have been in the area, their presence could be discovered.

(2) Some other considerations and methods concerning plant food procurement are as follows. Evaders should:

(a) Never take all the plants or fruits from one area.

(b) Pick only a few berries off of any one bush.

(c) When digging plants, take only one plant, then move on some distance before digging up another.

Figure 28-33. Fishing.

(d) When digging plants from old garden plots, make sure the plot is old. In many countries the people plant their crops and do not return to the plot until harvest time.

(e) Camouflage all signs of presence.

(3) Most domestic foods must be procured by theft which is very dangerous. However, if proper methods are used and the opportunity presents itself, plants and animals may be stolen. The main reason thieves are captured is the boldness they display after several successful thefts. The basic rules of theft are:

(a) If at all possible, the theft should take place at night.

(b) Evaders should thoroughly observe the area of intended theft from a safe vantage point.

(c) Evaders should find the vantage points just before dusk and look the place over to make sure everything is the same as it was the last time a check was made.

(d) Evaders should check for dogs, which could be a big hazard. Barking draws attention; also, some dogs are vicious and can harm evaders. Besides dogs, other animals or fowl can alert the enemy to the evaders' position.

(e) Evaders should never return twice to the scene of a theft.

(f) Every theft should be planned, and after its accomplishment, evaders should leave no evidence of either their presence or the theft itself (figure 28-37).

(g) Only small amounts should be taken (figure 28-38).

(4) When evaders find it necessary to take cultivated plant foods, they should never take the complete plant. Taking plants from the inside of the field, not the edge, and leaving the top of plants in place may help conceal the theft.

h. The rules of theft also apply to taking domesticated animals. The evader should concentrate on animals that don't make much noise. If a choice has to be made as to which animal to steal, evaders should take the smallest one.

i. Water is very essential, but it can be difficult to acquire (figure 28-39).

(1) When procuring water, evaders should try to find small springs or streams well away from populated areas. The enemy knows evaders need water and may check all known water sources. No matter where water is procured, evaders should try to remain completely

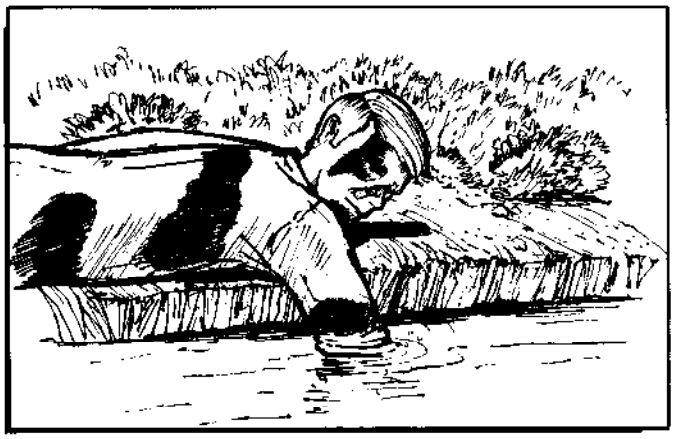

Figure 28-34. Fish Tickling - A.

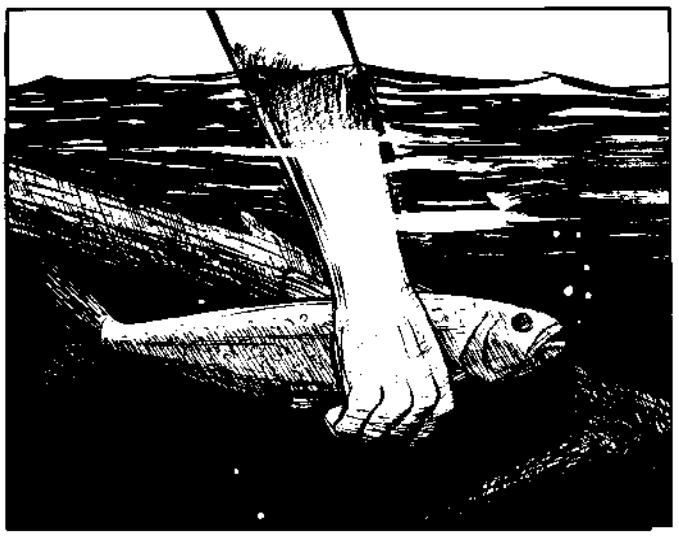

Figure 28-35. Fish Tickling - B.

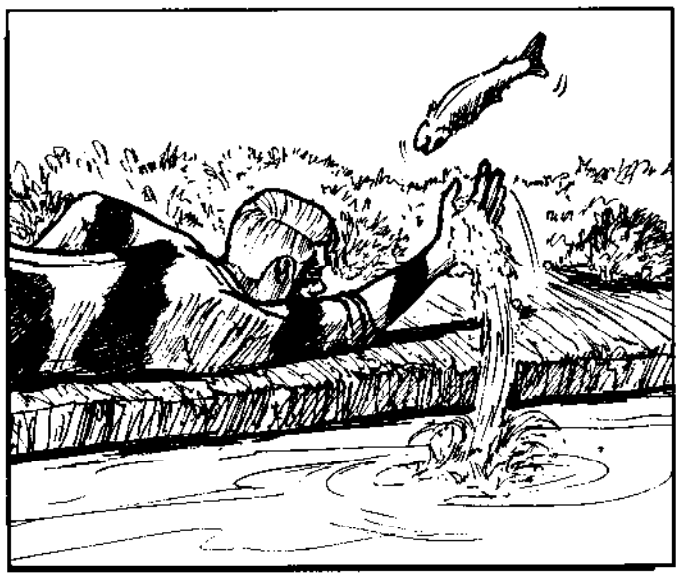

Figure 28-36. Fish Tickling - C.

concealed when doing so. The area around the water source should be observed to make sure it isn't patrolled or watched. While obtaining water and when leaving the area, evaders should conceal any evidence of their presence. Good sources of water include trapped rainwater in holes or depressions and plants which contain water.

(2) The preparation and purification of food and water by cooking is a precaution which should be weighed against the possibility of capture. It might be necessary to eat raw plant and animal foods at times. Some plant foods will require cutting off thorns, peeling off the outer layer, or scraping off fuzz before consumption. Raw animal foods may contain parasites and micro-organisms which may not effect evaders for days, weeks, or months, at which time, hopefully, they will be under competent medical care. If the environment in which the parasites live is altered by cooking, cooling, or drying, the organisms may be killed. Meat can be dried and cooled by cutting into thin strips and air-drying. Salting makes the meat more palatable. If the meat must be cooked, small pieces should be cooked over a small hot fire built in an unpopulated area. The best methods of preserving food during evasion are drying or freezing.

(3) In an evasion situation, boiling water for purification should not be used as a method except as a last resort. Iodine tablets are the best method of purification. If evaders do not have purification tablets, and the danger of the enemy detecting their fires is too great, evaders may have to forego purification. The only problem with this is if water is not purified, it may cause vomiting and (or) diarrhea. These ailments will slow down the evaders and make them susceptible to dehydration. Aeration and filtration may help to some degree and are better than nothing. If water cannot be purified, evaders should at least try to use water sources which are clean, cold, and clear. Rain, snow, or ice should be used if available.

28-13. Encampment Security Systems:

a. When evading for extended periods in enemy-held territories, it becomes essential for evaders to rest. To rest safely, especially if in a group, it is essential to devise and use some sort of early warning system to prevent detection and unexpected enemy infiltration. When establishing an evasion "shelter" area, there are certain things which should be done (day or night) for security purposes (figure 28-40).

b. Evaders should scout the area around their encampment for signs of people. They should pay particular attention to crushed grass, broken branches, footprints, cigarette butts, and other discarded trash. These signs may reveal identity, size, direction of travel, and time of passage of an enemy force. If large numbers of these signs are present, the evader should consider moving to a more secure area.

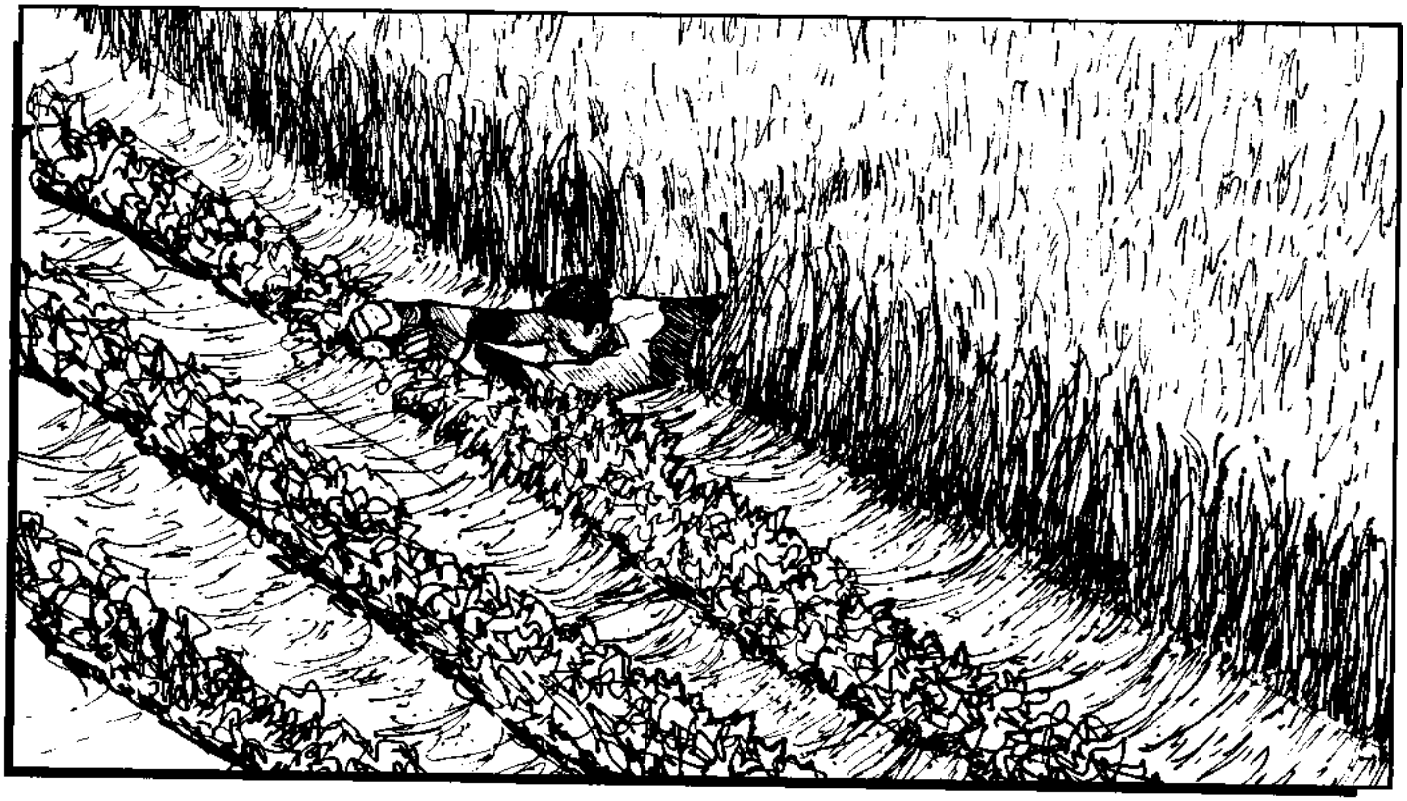

Figure 28-37. Stealing Vegetables.

Figure 28-38. Stealing Small Amounts of Food.

Figure 28-39. Procuring Water.

Figure 28-40. Rest Safely.

c. Once the camp area has been determined to be fairly secure, some type of alarm system must be devised. For a lone evader, this may consist of actually constructing wire or line with sound-producing devices attached. However, this system works for the enemy as well and may prove to alert enemy forces to the evader's presence. A lone evader should use the natural alarm system available. Disturbances in animal life around an evader may indicate enemy activity in the area. Group situations may allow for more security. Two or more evaders may use lookout(s) or "scout(s)" at observation posts strategically located around the encampment.

d. Readiness is another aspect of security. The evader should be aware that, at any time, the shelter area may be overrun, ambushed, or security compromised, making it necessary to vacate the area. If evaders are in a group and future group travel is desired, it is essential that everyone in the group knows and memorizes certain things; such as, compass headings or direction of travel, routes of travel, destination descriptions, and rally points (locations where evaders regroup after separation). Alternate points must be designated in case the original cannot be reached or if it is compromised by enemy activity.

(1) Once everyone in the evader's group has reached the final destination, alternate point, or rally point, a new emergency evacuation and rendezvous plan must be established.

(2) Evaders should always be aware of the next rally point, its location, and direction. These places, which provide concealment and cover, should be designated along the route in case an enemy raid or ambush scatters the group. There should be a rally point for every stage of the journey. Even when approaching the supposed "final destination" of the day, evaders should have an evacuation plan ready.

e. Maintaining silence is a very important aspect of security. It is essential to be able to communicate with individual group members and scouts so that everyone is aware of what is going on at all times during evasion. Hand signals are the best method of communication during evasion as they are silent and easily understood. Instructions and commands which must be conveyed throughout the entire group are: (See figure 28-41 for examples of hand signals.)

(1) Freeze.
(2) Listen.
(3) Take cover.
(4) Enemy in sight.
(5) Rally.
(6) All clear.
(7) Right.
(8) Left.

28-14. Evasion Movement:

a. During evasion travel, the evader is probably most susceptible to capture. Many evaders have been captured as a direct result of their failure to use proper evasion movement techniques. Evasion movement is the action of a person who, through training, preparation, and application of natural intelligence, avoids capture and contact with hostiles, both military and civilian. Not only is total avoidance of the enemy desirable for evaders, it is equally important for evaders not to have their presence in any enemy controlled area even suspected. A fleeting shadow, an inopportune movement or sound, and an improper route selection are among a number of things which can compromise security, reveal the presence of evaders, and lead to capture.

b. One evasion situation will not be identical to another. There are, however, general rules which apply to most circumstances. These rules, carefully observed,

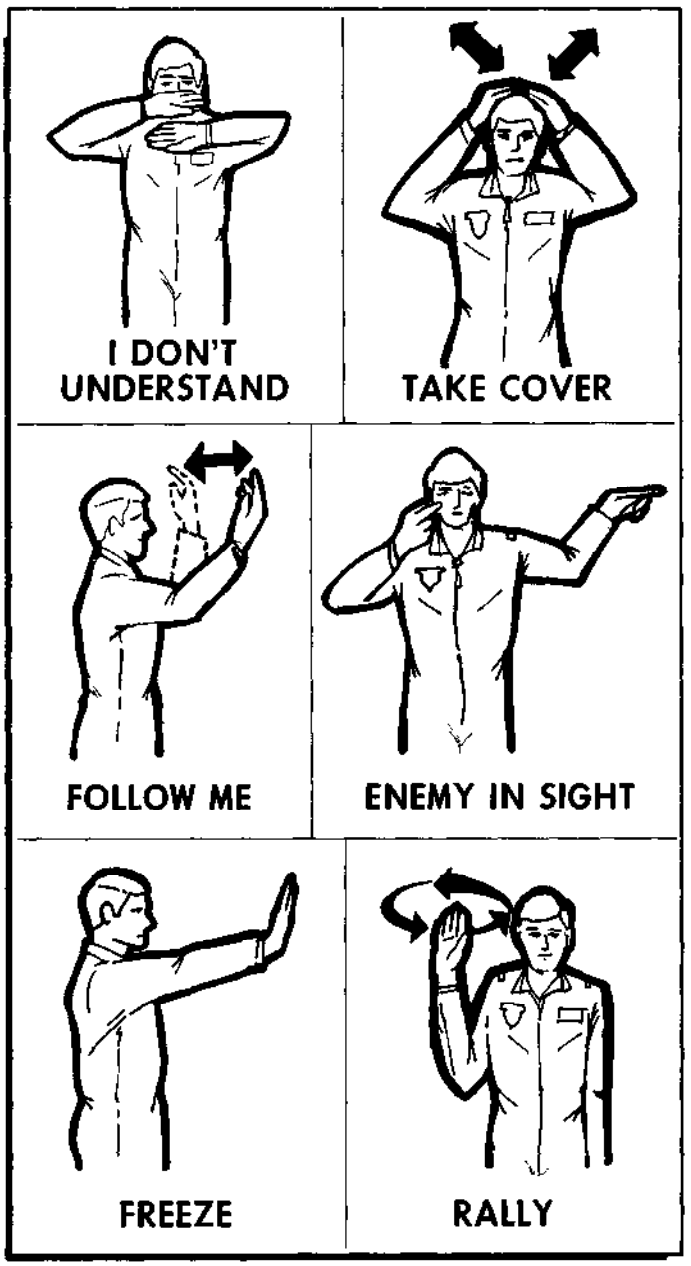

Figure 28-41. Hand Signals.

will enhance the evaders' chances of returning to friendly control.

(1) Evasion begins even before a crewmember leaves an aircraft over enemy territory. Two factors which are essential to successful evasion and return are opportuntiy and motivation. Pre-mission preparation and knowledge of areas of concern are very important to a crewmember. Pre-mission knowledge gained must be based on the most current information available through command area briefings, area studies, and intelligence briefings.

(2) Some areas of interest to the prospective evader are:

(a) Topography and Terrain. An aircrew member should know the physical features and characteristics, possible barriers, best areas for travel, availability of rescue, and the type of air or ground recovery possible. A future evader should also know the requirements for long-term unassisted survival in the area of operation.

(b) Climate. The typical weather conditions and variations should be known to aid in evasion efforts.

(c) People. A very critical consideration may be knowing the people in the area. From ethnic and cultural briefs read before the mission, crewmembers should familiarize themselves with the behavior, character, customs, and habits of the people. It may, at times during the evasion episode, be necessary to emulate the natives in these respects (figure 28-42).

(d) Equipment. Aircrew members who would hope to be successful evaders should be thoroughly familiar with all of their equipment and know where it is located. They should also preplan what equipment should be retained and what should be left behind under certain conditions of an evasion travel situation.

(3) Before addressing evasion movement techniques, it is appropriate to go over some of the factors which influence an evader's decision to travel.

Figure 28-42. Knowing the People.

(a) The first few minutes after entering enemy or unfriendly territory is usually the most critical period for the evader. The evader must avoid panic and not take any action without thinking. In these circumstances, the evader must try to recall any previous briefings, standard operating procedures, or training and choose a course of action which will most likely result in return to friendly territory.

(b) In those first few minutes, there is a great deal for the downed crewmember to think about. Quick consideration must be given to landmarks, bearings, and distance to friendly forces and from enemy forces, likely location for helicopter landing or pickup, and the initial direction to take for evasion. Knowledge of what to expect is important because when circumstances arise which have been considered in advance, they can be carried out more quickly and easily. Evaders should try to adapt this knowledge and any skills they have to their particular situation. Flexibility is most important, as there are no hard and fast rules governing what may happen in an evasion experience.

(c) In most evasion situations, evaders will be required to move if for no other reason than to leave the immediate landing area when pickup is not imminent. Because any movement has the inherent risk of exposure, some specific principles and practices must be observed. Periods of travel are the phases of evasion when evaders are most vulnerable. Many evaders have been captured because they followed the easier or shorter route and failed to employ simple techniques such as watching and listening frequently and seeking concealment sites.

28-15. Searching Terrain:

a. Evaders should visually survey the surrounding terrain from an area of concealment to determine if the route of travel is a safe one. Evaders should first make a quick overall search for obvious signs of any presence such as unnatural colors, outlines, or movement. This can be done by first looking straight down the center of the area they are observing, starting just in front of their position, and then raising their eyes quickly to the maximum distance they wish to observe. If the area is a wide one, evaders may wish to subdivide it as shown in figure 28-43. Now all areas may be covered as follows: First, by searching the ground next to them. A strip about 6 feet deep should be looked at first. They may search it by looking from right to left parallel to their front. Secondly, by searching from left to right over a second strip farther out, but overlapping the first strip. Searching the terrain in this manner should continue until the entire area has been studied. When a suspicious spot has been located, evaders should stop and search it thoroughly (figure 28-44).

b. The evader must question the movement:

(1) Is the enemy searching for the evader?

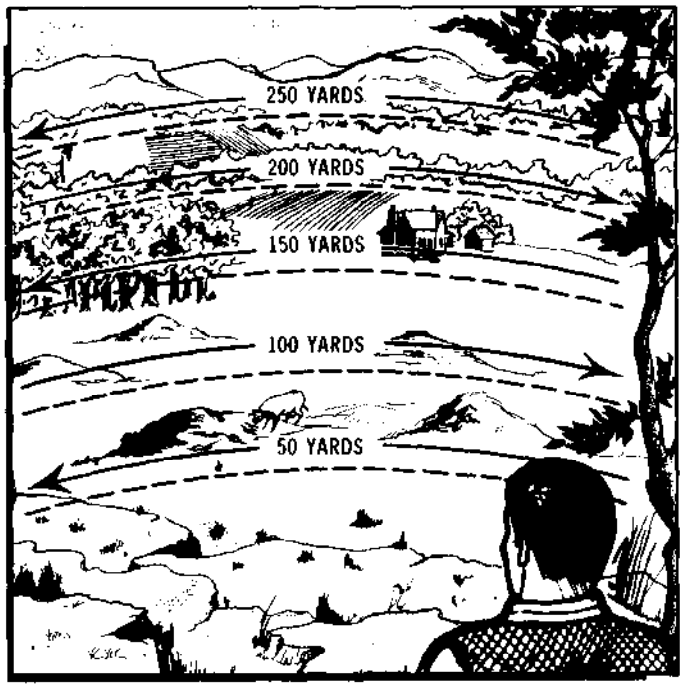

Figure 28-43. Viewing Terrain - A.

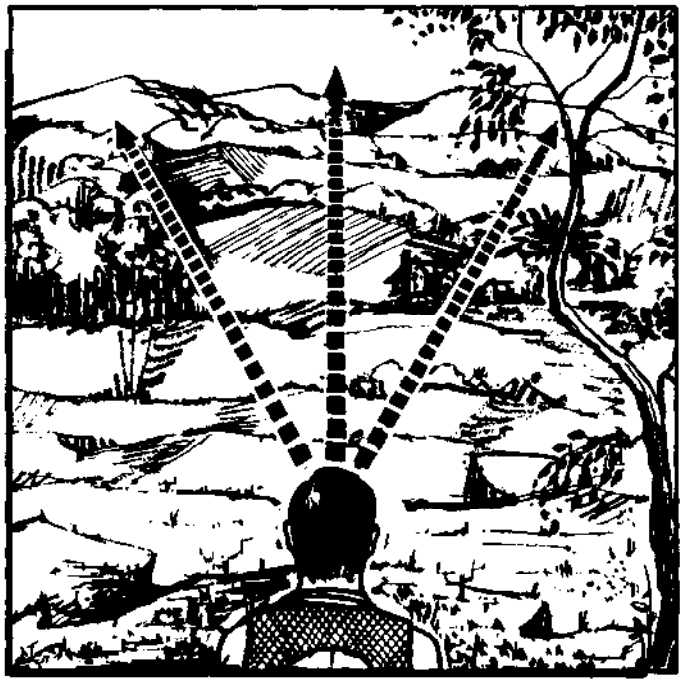

Figure 28-44. Viewing Terrain - B.

(2) What is the evader's present location?

(3) Are chances for rescue better in some other place?

(4) What type of concealment can be afforded in the present location?

(5) Where is the enemy located relative to the evader's position?

c. Having considered the necessity and risks of travel, evaders must:

(1) Orient themselves.

(2) Select a destination, alternates, and the best route.

(3) Have an alternate plan to cover all foreseeable events.

d. Cautious execution of plans cannot be overemphasized since capture of evaders has generally been due to one or more of the following reasons:

(1) Unfamiliarity with emergency equipment.

(2) Walking on roads or paths.

(3) Inefficient or insufficient camouflage.

(4) Lack of patience when pinned down.

(5) Noise or movement or reflection of equipment.

(6) Failure to have plans if surprised by the enemy.

(7) Failure to read signs of enemy presence.

(8) Failure to check and recheck course.

(9) Failure to stop, look, and listen frequently.

(10) Neglecting safety measures when crossing roads, fences, and streams.

(11) Leaving tracking signs behind.

(12) Underestimating time required to cover distance under varying conditions.

e. Evaders should understand progress on the ground is measured in stopover points reached. Speed and distance are of secondary importance. Evaders should not let failure to meet a precise schedule inhibit their use of a plan.

28-16. Movement Techniques Which Limit the Potential for Detection of an Evader (Single).

a. Evaders should constantly be on the lookout for signs of enemy presence. They should look for signs of passage of groups, such as crushed grass, broken branches, footprints, and cigarette butts or other discarded trash. These may reveal identity, size, direction of travel, and time of passage (figure 28-45).

(1) Workers in fields and normal activities in villages may indicate absence of the enemy.

(2) The absence of children in a village is an indication they may have been hidden to protect them from action which may be about to take place.

(3) The absence of young men in a village is an indication the village may be controlled by the enemy.

b. A knowledge of enemy signaling devices is very helpful. Those listed below are examples of some used by communist guerrillas in Southeast Asia.

(1) A farm cart moving at night shows one lantern to indicate no government troops are close behind.

(2) A worker in the fields stops to put on or take off a shirt. Either act can signal the approach of government troops. This is relayed by other informers.

(3) A villager fishing in a rice paddy holds a fishing pole out straight to signal all clear; up at an angle to signal the troops are approaching.

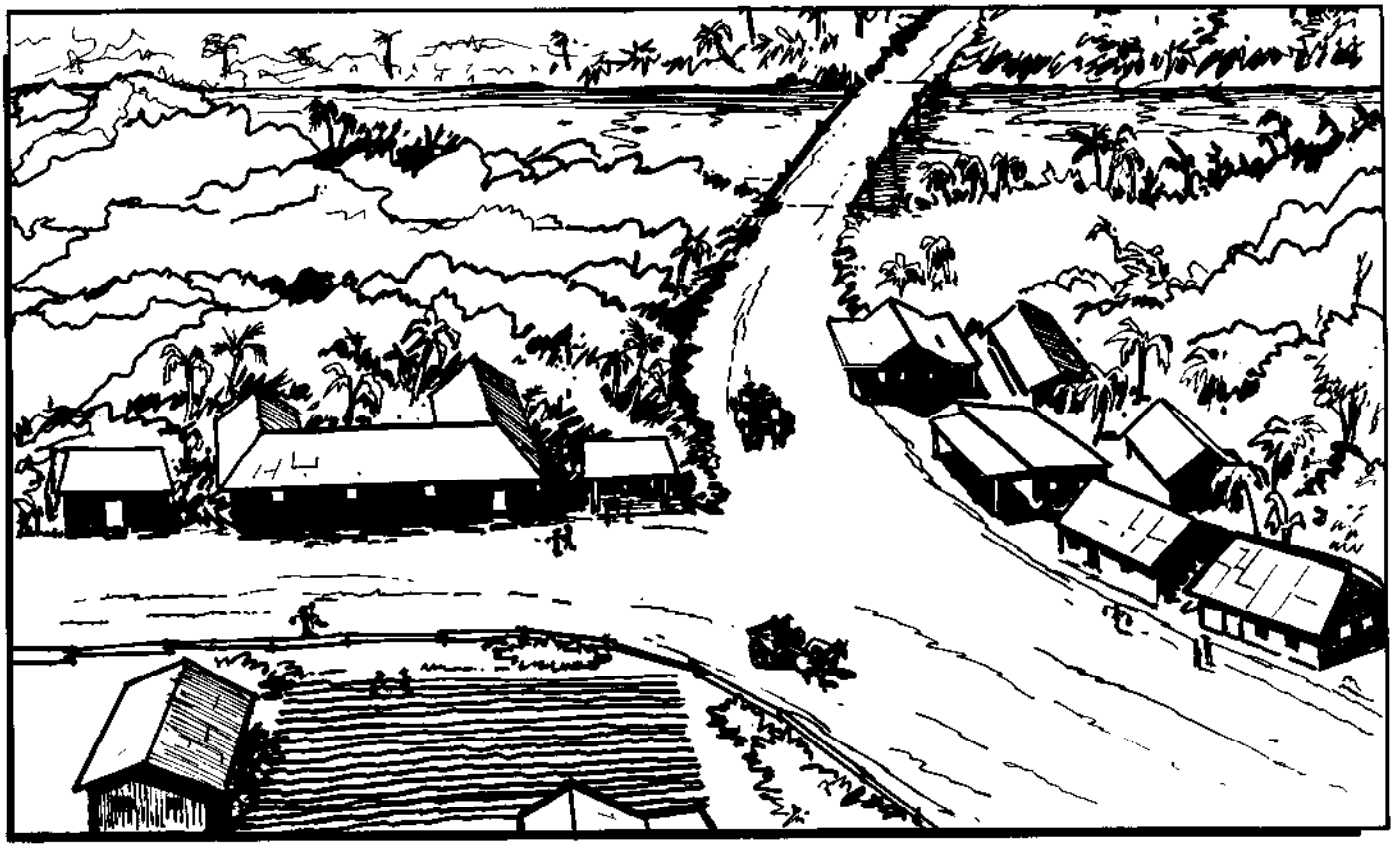

Figure 28-45. Signs of Passage.

c. The times evaders choose to travel are as critical as the routes they select. If possible, evaders should try to make use of the cover of darkness. Darkness provides concealment and in some cases there is also less enemy traffic. If, out of military necessity, the enemy is active during the hours of darkness, evaders may then find it wise to move in the early morning or late afternoon. Night movement is slower, more demanding, and more detailed than daylight movement; but it can be done. The alternative to night movement might be capture, imprisonment, and death. Evaders should then consider traveling under the cover of darkness first. However, if the enemy knows the evaders' position, or if the other factors dictate (terrain, vegetation, navigation consideration, etc.), other choices may have to be made. If travel is to be done during darkness, the terrain to be traversed should be observed during daylight if possible. While observing the area to be traveled, evaders should give attention to areas offering possible concealment as well as the location of obstacles they may encounter on their route. If the evader has a map, a detailed study of it should be made. However, it should be remembered that natural terrain features change with time of day and time of year. Certain features (ditches, roads, burned-off areas, etc.) may not be on the evaders' map if they are new. Such pretravel reconnoitering of an area will give evaders a head start on knowing how to adapt their travel movements and camouflage necessary from point to point (figure 28-46).

d. Evaders should try to memorize the routes they will take and the compass headings to their destinations. This information should not be written down on the map or on other pieces of material.

Figure 28-46. Pretravel Reconnoitering.

Figure 28-47. Military Crest.

e. If traveling through hilly country that provides cover, the military crest should be used as it may be the safest route. An evader's route should avoid game trails and human paths on the tops of ridges. The chance of encountering the enemy is greater on the tops of ridges. The risk of silhouetting during both day and night is also increased (figure 28-47).

(1) When it is necessary to cross the skyline at a high point in the terrain, an evader should crawl to it and approach the crest slowly using all natural concealment possible. How the skyline is to be crossed depends on whether it is likely the skyline at that point is under hostile observation. Evaders may never be certain any area is not under observation. When a choice of position is possible, the skyline is to be crossed at a point of irregular shapes such as rocks, debris, bushes, fence lines, etc.

(2) Another important point about moving along the military crest is that it is easier to evade the enemy on the side of the ridge than it is to lose sight of the enemy on the top of a ridge or in the valley below. Evasion along the side of a hill will afford a better chance for evaders to reach sites which are suitable for air recovery.

f. Evaders should move slowly, stopping and listening every few paces. Additionally, they should not make noise and should take advantage of all cover to avoid revealing themselves. If spotted, they should leave the area quickly by moving in a zigzag route to their goal if at all possible. If the enemy finds evidence of the evaders' passage, their route may help confuse the pursuer as to direction and goal. Background noise can be either a help or hindrance to those who are trying to move quietly—both the evaders and the enemy. Sudden bird and animal cries, or their absence, may alert evaders to the presence of the enemy, but those same signals can also warn the enemy of an approaching or fleeing evader. Sudden movement of birds or animals is also something to look out for.

g. The following are some techniques of limiting or concealing evidence of travel. Evaders should:

(1) Avoid disturbing any vegetation above knee level. Evaders should not grab at or break off branches, leaves, or tall grass.

(2) Glide gently through tall grass or brush. Avoid using thrashing movements. A walking stick may be used to part the vegetation in front, and then it can be

used behind to push the vegetation back to its original position. The best time to move is when the wind is blowing the grass.

(3) Realize that grabbing small trees or brush as handholds may scuff the bark at eye level. The movement of trees can be spotted very easily from a distance. In snow country it may mark a path of snowless vegetation that can be spotted when tracks cannot be.

(4) Select firm footing and place the feet lightly but squarely on the surface avoiding the following:

(a) Overturning duff, rocks, and sticks.

(b) Scuffing bark on logs and sticks.

(c) Making noise by breaking sticks.

(d) Slipping, which may make the noise of falling.

(e) Mangling of low grass and bushes that would normally spring back.

h. Evaders can mask their tracks in snow by:

(1) Using a zigzag route from one point of concealment to the next and, when possible, placing the unavoidable tracks in the shadows of trees, brush, and along downed logs and snowdrifts.

(2) Restricting movement before and during snowfall so tracks will fill in. This may be the only safe time to cross roads, frozen rivers, etc.

(3) Traveling during and after windy periods when wind blows clumps of snow from trees, creating a pockmarked surface which may hide footprints.

(4) Remembering that snowshoes leave bigger tracks, but they are shallower tracks which are more likely to fill in during heavy snowfall. Branches or bough snowshoes make less discernible prints.

i. Evaders' tracks in sand, dust, or loose soil should be avoided or else marked by:

(1) Using a zigzag route, making use of shadows, rocks, or vegetation to walk on to mask or prevent tracks.

(2) Wrapping cloth material loosely about the feet, this makes tracks less obvious.

j. By moving before or during wet or windy weather, evaders may find that their tracks are obliterated or worn down by the elements. Along roadways, evaders should be particularly cautious about leaving their tracks in the soft soil to the side of the road. They should step on sticks, rocks, or vegetation. They shouldn't leave tracks unless there are already tracks of natives on the road and their tracks can be made to look like the existing ones (small bare feet, tire sandals, or enemy footgear). Rolling across the road is a method of avoiding tracks. Walking in wheel ruts with the toes parallel to the road will help conceal tracks. If the road surface is dry, sand, dust, or soil tracks may be "brushed" out to make them look old or will help the wind erase them more quickly. This must be done lightly, however, so the tracks do not look as if they have been deliberately swept over. Mud will retain footprints unless the mud is shallow and there is a heavy rain. Evaders should try to go around these areas.

k. Many principles and techniques which work for the individual are also appropriate for use in groups of evaders. While it is not true that the only safe way to evade is individually, there is a certain danger in moving with a group.

(1) It is generally not advisable to travel in a group larger than three. If possible, the senior person should divide the group into pairs. Paired individuals must be compatible since any disagreement may prove fatal during the evasion process. Group travel can be advantageous because supplementary assistance is available in case of injury, in defense against hostile elements, in travel over rough terrain, and it provides moral support.

(2) In group travel, movement becomes critical. Movement attracts attention.

(3) The intervals and distances between individuals in the group should be made according to the terrain and the time of day. Intervals will probably be greater during the day. The natural but extremely dangerous tendency to "bunch up" is to be avoided when traveling in a group (figure 28-48).

(4) Under favorable conditions it is possible for the enemy to see 100 yards into open woods. If the undergrowth is light, the route must be farther into the woods, and the interval between evaders must be greater. Added consideration must be taken in deciding whether to travel at night or during the day. The leader will direct and guide the group to and from the best positions. All communications within the group should be made with silent signals only. Security in group evasion is of paramount importance. All members should stay alert. Security posts, lookouts, or guards should be designated for periods of rest or stopping.

(5) Various formations are available for use by the evading group during periods of travel. The group must be flexible and able to adapt to changes in the conditions of the situation. The type of formation may also change with the route. In choosing a formation, the following points should be considered:

(a) Group control and intercommunications.

(b) Security.

(c) Terrain.

(d) Speed in movement.

(e) Visibility.

(f) Weather.

(g) Enemy placement.

(h) The need for dispersion.

(i) Flexibility of change in speed and direction of travel.

(6) A "formation" is merely the formal arrangement of individuals within a group. This formal arrangement is designed to give the greatest dispersion consistent with adequate intercommunication, ease and speed of movement, and flexibility to change direction and speed of travel at a moment's notice; that is, close control. Any arrangement which provides the above advantages

Figure 28-48. Group Travel.

is satisfactory. The Army, which constantly moves groups of men of various sizes on various missions, has found the squad file, squad column, and squad line to be satisfactory.

(a) In a squad file, the personnel are arranged in a single file, or one directly behind the other, at different distances. The distance will vary depending on need for security, terrain, visibility, group control, etc. It is primarily used when moving over terrain which is so restrictive that the squad cannot adopt any other formation. It is also used when visibility is poor and squad control becomes difficult. When people are in a squad file, it is easy to control the group and provide maximum observation of the flanks. This is a fast way to travel, especially in the snow.

-1. However, there are disadvantages to this type of movement. A major one is that this formation is visually eye-catching. All the noise of the group is also concentrated in one place. This type of formation is easily defined and infiltrated. Depending on how many people are in the group, the area they walk through can become very packed down and easily detected by the enemy.

-2. If the group in the squad file is small, some members may have to double up on jobs; however, the "point" or lookout (scout) in the lead should only perform that job (figure 28-49).

(b) In the squad column, personnel are arranged in two files. The personnel are more closely controlled and yet maneuverable in any direction. There is greater dispersion with reasonable control for all-round security. It is used when terrain and visibility are less restrictive because it provides the best means of moving armed personnel into dispersed all-round security. It is easy to change into either the file or the line.

-1. There are many advantages to this type of formation, a major one being a greater dispersal of personnel. Visually, this way of moving is less eye-catching, and the sounds the group makes are less concentrated. With this formation, the rate of travel is reasonable, yet there is no well trampled corridor for enemy trackers to follow.

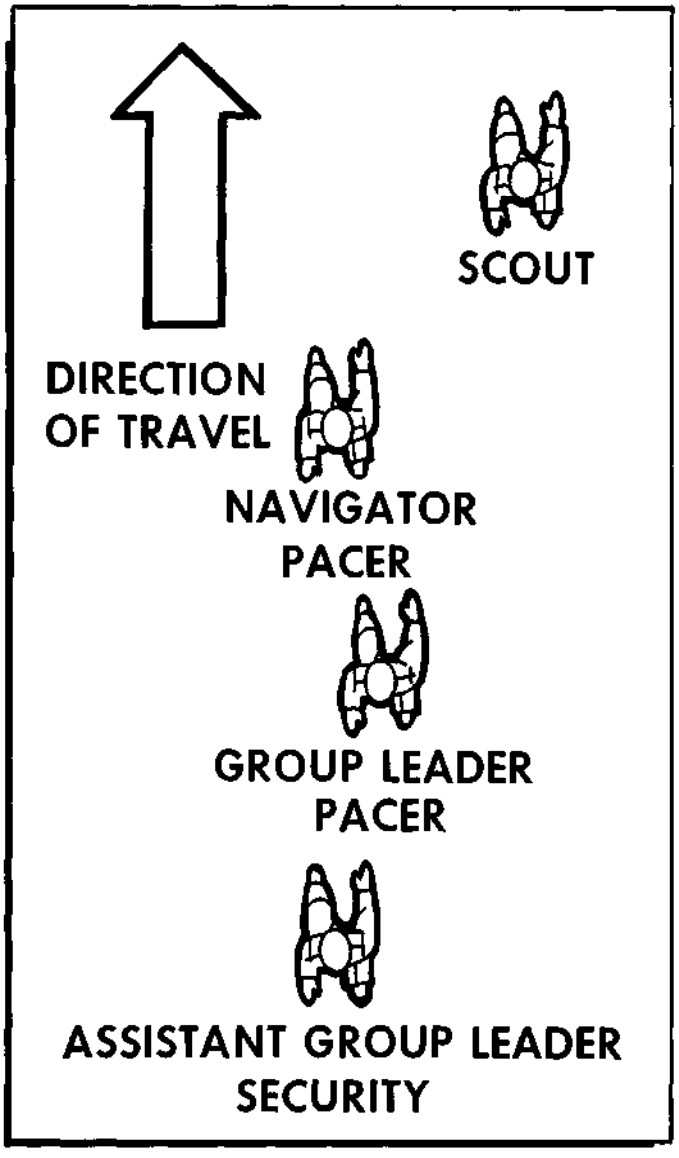

Figure 28-49. File Formation.

-2. There are a few disadvantages to movement in this manner. This formation of travel is hard to control in areas of dense vegetation with poor visibility and makes straying from the group likely. Although the paths may be faint, this mode of travel will also form a large number of trails. The rate of travel will be slower overall when traveling this way. An example of how personnel may be dispersed using the squad column methods is shown in figure 28-50.

(c) In the squad line, all personnel are arranged in a line. This formation is used primarily by the Army as an assault formation because it is best for short, fast movements.

-1. The advantage of this formation is that it is the quickest way to cross such obstacles as roads, fences, and small open spaces. It provides for tight control of individual movement while providing security for short-distance moves.

-2. The disadvantages are that there are extreme communication and control problems. Some personnel in this formation may be forced to traverse undesirable rough terrain in contrast to the other two formations. Figure 28-51 illustrates the organization of personnel in this type formation. The speed at which these formations will progress will vary with terrain, light, cover, enemy presence, health of personnel, etc.

l. Regardless of the formation used, the evader should pay particular attention to the technique used to travel. Knowing how to walk or crawl may make the difference between success and failure.

(1) Correct walking techniques can provide safety and security to the evader. Solid footing can be maintained by keeping the weight totally on one foot when stepping, raising the other foot high to clear brush,

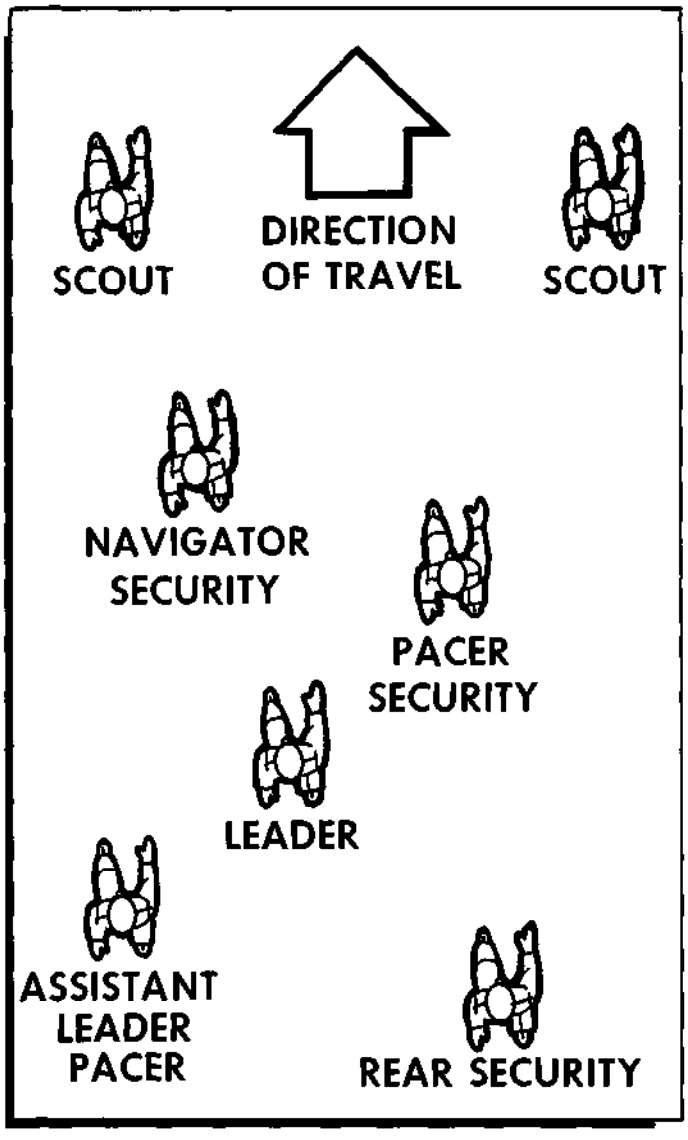

Figure 28-50. Squad Formation.

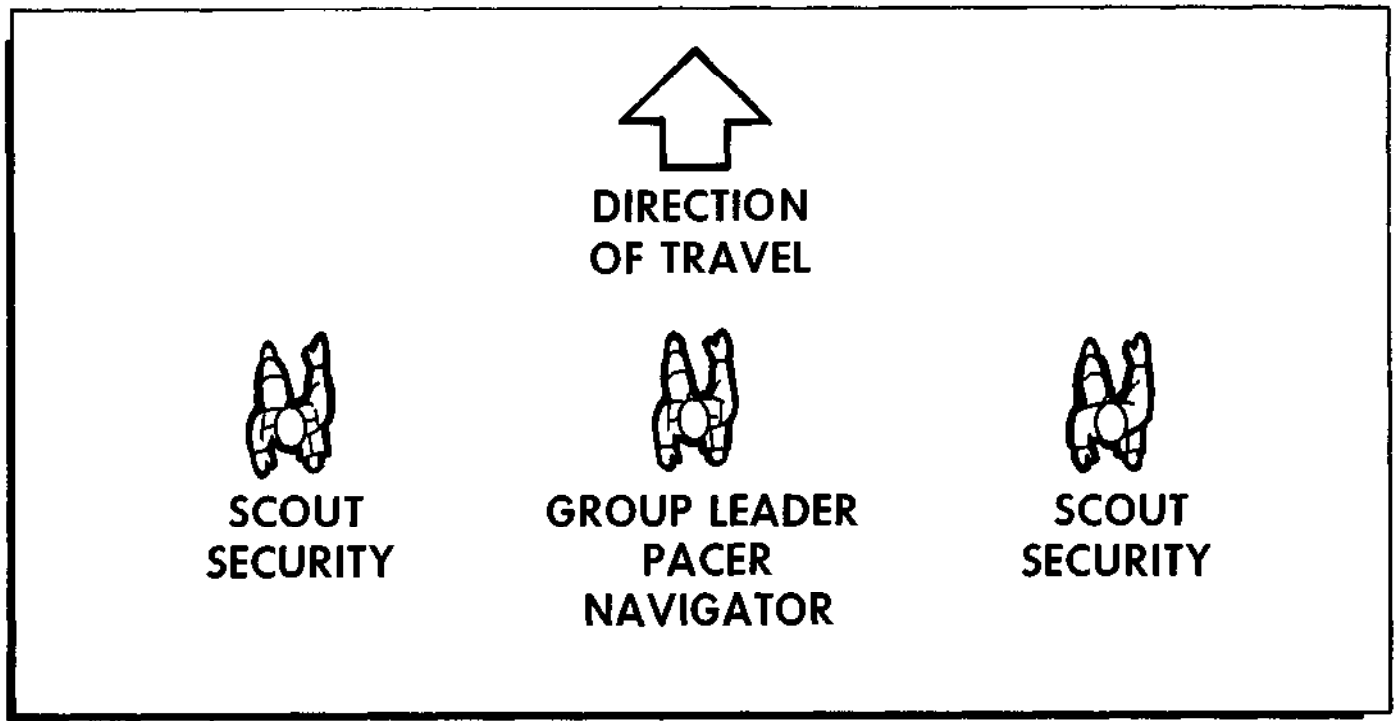

Figure 28-51. Squad Line.

grass, or other natural obstacles, and gently letting the foot down, toe first. Feel with the toe to pick a good spot—solid and free of noisy materials. Lower the heel only after finding a solid spot with the toe. Shift the weight and balance in front of the lowered foot and continue. Take short steps to avoid losing balance. When vision is impaired at night, a wand probe, or staff, should be used. By moving these aids in a figure-eight motion from near the ground to head height, obstructions may be felt.

(2) Another method of movement is by crawling. Crawling is useful when a low silhouette is required. There are times when evaders must move with their bodies close to the ground to avoid being seen and to penetrate some obstacles. There are three ways to do this: the low crawl, the high crawl, and the hands-and-knees position. Evaders should use the method best suited to the condition of visibility, ground cover, concealment available, and speed required.

(a) The Low Crawl. This can be done either on the stomach or back, depending on the requirement. The body is kept flat and movement is made by moving the arms and legs over the ground (figure 28-52).

(b) The High Crawl. This is a position of higher silhouette than the low crawl position, but lower than the hands and knees position. The body is free of the ground with the weight of the body resting on the forearms and lower legs. Movement is made by alternately advancing the right elbow and left knee, left elbow and right knee, elbows and knees laid flat on their inside surfaces (figure 28-53).

(c) Controlled Movement. The low crawl and high crawl are not always suitable when very near an enemy. They sometimes cause the evader to make a shuffling noise which is easily heard. On the other hand, carefully controlled movement can be made to be silent, and these techniques present the lowest possible silhouettes.

(d) The Hands-and-Knees Crawl. This position is used when near an enemy. Noise must be avoided, and a relatively high silhouette is permissible. It should only be used when there is enough ground cover and concealment to hide the higher silhouette involved (figure 28-54).

m. If evaders are moving as a group and are forced to disperse, being able to account for everyone after regrouping is important. Likely locations for rallying points are selected during map study or reconnaissance.

(1) Selecting a Rallying Point. The group leader must:

(a) Always select an initial rallying point. If a suitable area for this point is not found during map study or reconnaissance, the leader can select it by grid coordinates or in relation to terrain features.

(b) Select likely locations for rallying points en route.

(c) Plan for the selection and designation of additional rallying points en route as the patrol reaches suitable locations.

(d) Plan for the selection of rallying points on both near and far sides of danger areas which cannot be bypassed, such as trails and streams. This may be done

Figure 28-52. Low Crawl.

by planning that, if good locations are not available, rallying points will be designated in relation to the danger area; for example, "...50 yards this side of the trail," or "...50 yards beyond the stream."

(2) Use of Rallying Points. If dispersed by enemy activity or through accidental separation, each evader in the group should be prepared to evade, individually, to the regrouping (rallying) point to arrive at a predesignated time. If this is not possible, the individual will become a "lone evader." The group should not make any effort to locate someone not reaching the rally point. The group should formulate a new plan with new rallying points and clear the area.

(a) Rallying points should be changed or updated as they are passed. Points should not be directly on the line (route) of travel. By selecting points off line, the job of searchers (trackers) is made more difficult and the chance of being "headed off" or "blind stalked" by the enemy is reduced.

Figure 28-53. High Crawl.

Figure 28-54. Hands-and-Knees Crawl.

(b) If the group is dispersed between rallying points en route, the group rallies at the last rallying point or at the next selected rallying point. The group leader announces the decision at each rallying point as to which point the group will rally.

(3) Actions at Rallying Points. Actions to be taken at rallying points must be planned in detail. Plans for actions at the initial rallying point and rallying points en route must provide for the continuation of the group as long as there is a reasonable chance to evade as a group. An example of a plan would be for the group to wait for a specified period, after which the senior person present will determine actions to be taken based on personnel and equipment present. Even during movement phases, it is important to be able to check on the presence and status of all group members. A low toned, actual head count starting at the rear of the formation might be one way to do this; hand signals is another. One reason for keeping in touch with everyone is to establish new plans or adjust old ones while moving.

28-17. Barriers to Evasion Movement:

a. Obstacles. Throughout the evasion episode, many obstacles may be encountered which may impede evaders or influence the selection of travel routes. These barriers or obstacles can be divided into natural ones, such as rivers or mountains, and human ones, such as border guards or manmade fences or roads. Some of these obstacles may be helpful while others might be a hinderance.

(1) Rivers and Streams. When crossing rivers and streams, bridges and ferries can seldom be used since the enemy normally establishes checkpoints at these locations. This leaves a choice of fording, swimming, crossing by boat, or using some improvised method (figure 28-55).

(2) Mountains. In mountainous areas, survival may be the primary concern. It may be necessary for evaders

Figure 28-55. River Crossing.

Figure 28-56. Weather.

to remain in one location for an extended period of time, perhaps even waiting for the coming of spring before attempting travel. Many mountainous areas, however, are havens which afford cover, water, food, and low population densities. Also, the chances of receiving assistance from people in areas where homes and farms are separated by great distances are more likely. When traveling in mountainous regions, evaders should not forget to use the military crest if concealment is available. In plains areas, evaders should use depressions, drainages, or other low spots to conceal their movements. Route selection should be planned with the utmost care to avoid unnecessary delays caused by cliffs, large bodies of water, and flat areas.

(3) Vegetation. Some swamps, drainage areas, and thickets may be too thick for evaders to penetrate, and may require that a detour or alternate route be used. If the vegetation can be moved through, evaders should take care not to leave evidence of passage by disturbing the growth.

(4) Weather. Weather can sometimes be used to screen evaders from the enemy. Certain weather conditions mask the noise made by traveling (figure 28-56). Moving during a rainstorm may erase the footprints left by an evader; but after the rain the soft soil will leave definite signs of passage. Thunder may mask the sounds evaders make, but lightning may cause them to be seen. Snowstorms may be used to cover evaders' signs and sounds, but once the storm is over, evaders must use extreme care not to leave a trail.

b. Artificial Obstacles. Evaders may also encounter a wide variety of artificial obstacles while traveling within enemy territory or when attempting to leave a controlled area. As a general rule, evaders should *not* attempt to penetrate these obstacles if they can be bypassed. If an analysis of the situation reveals the obstacles cannot be bypassed, evaders must be skilled in the methods and techniques for dealing with specific artificial barriers to evasion. If possible, move to a less fortified (controlled) area or find a better damaged area in the barrier.

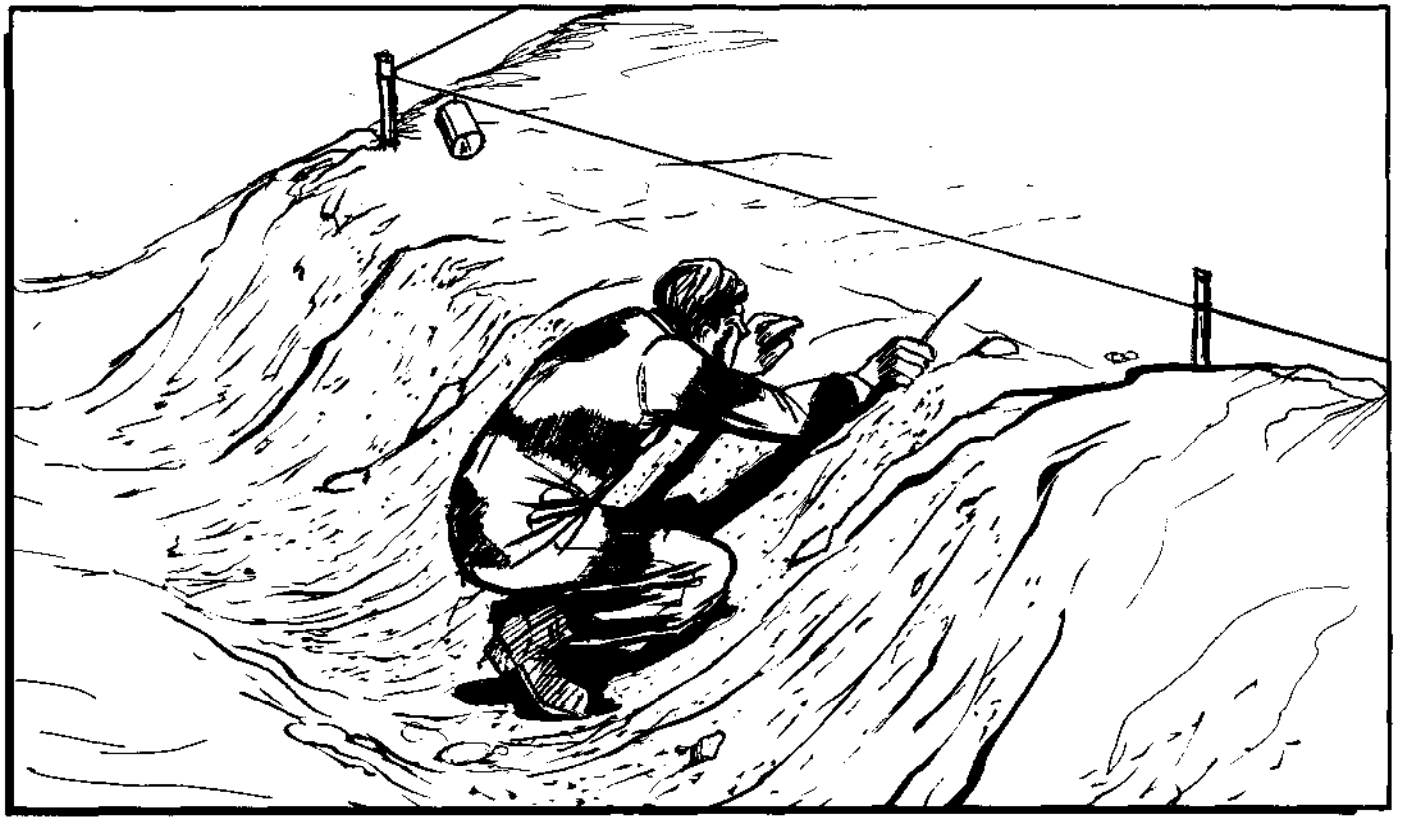

Figure 28-57. Trip Wires.

(1) Evaders may encounter trip wires. These wires may be attached to pyrotechnics, booby traps, sensor devices, mines, etc. These wires are normally thin, olive green (or other colors that blend with the environment), strong, and extremely difficult to see. A supple piece of wood can be improvised and used as a wand to detect these wires (figure 28-57).

(a) A tripwire may be set up to be from 1 inch to a number of feet above the ground and to extend any number of feet from the device to which it is attached.

(b) The pressure necessary to activate a sensing device, mine, pyrotechnic device, or other trap-associated device to which the wire may be connected can vary from a few ounces to several pounds. This means that the evader must be very careful when attempting to determine the presence of these devices.

(c) Once a tripwire has been detected, evaders should move around the wire if possible. If not possible, they should go either over or under it. They should not tamper with or cut the wire. If one device is discovered, be alert for "backup" devices.

(d) A number of devices activated by tripwires have a combination pressure-release arming mechanism. Cutting the wire or releasing the tension in the wire may activate the device. Some devices are electrically activated when there is a change in the current flowing through the attached wire—either because the wire is cut, in some way grounded, or otherwise altered. Evaders should take extreme care to avoid touching tripwires, but if contact is made, they must try to avoid sufficient pressure for activation.

(2) Illumination flares may also present a problem to evaders. These, of course, can be activated by evaders themselves by a tripwire, by the enemy in the form of electronically activated ground flares, or by flares dropped from an overflying aircraft. Other overhead flares may be fired by mortars, rifles, artillery, and hand projectors.

(a) The illumination flares may burn as bright as 20,000 candlepower and illuminate up to a 300-foot radius in case of a ground flare, or a much larger area in the case of an overhead flare which is lofted or dropped and burns high above the ground.

(b) If evaders hear the launching burst of an overhead flare, they should, if possible, get down while it is rising and remain motionless. If caught in the light of a flare when they blend well with the background, they should freeze in position and not move until the flare goes out. The shadow of a tree will provide some protection. If caught in an open area, they may elect to crouch low or lie on the ground and, as a general rule, should not move after the area is illuminated.

(c) However, if they are caught in the light of a ground flare and their position is such that the risk of

Figure 28-58. Electrified Fences.

remaining is greater than that of moving, they should move quickly out of the area. If within a series of obstacles or an obstacle system, evaders must remember running can be extremely dangerous because of the obstacles in the area and the fact that movement, especially fast movement, catches the eyes of an observer. If it is determined they cannot quickly move out of the area because of possible serious injuries due to existing obstacles or because they may be observed by the enemy, evaders should drop to the ground and conceal themselves as much as possible.

(d) Evaders should remember the light of a flare (either ground or overhead bursts) is temporarily blinding and the eyes should be covered to conserve night vision.

(e) If caught by a flare when actually penetrating an obstacle such as a concertina wire, evaders should get as low as possible, stay still, and cover their eyes. The light of a flare can act to an evader's advantage because the searching enemy will lose its night vision which may add to the evader's chance for success in departing the area after the flare has burned out. The light of a flare also creates very dark shadows which, under some circumstances, can afford good concealment from enemy observation.

(3) Various types of chain and wire fences may hinder the progress of evaders who are moving to the safety of friendly areas.

(a) For indications of electrical fences, evaders should watch for dead animals, insulators, flashes from wires during storms, and short circuits causing sparks (figure 29-58). A quick simple test can be conducted to determine if a wire is electrified. This test is made by carefully approaching the wire holding a stem of grass or a small, damp stick on the wire. If the wire is charged, a mild shock will result but will not cause injury.

(b) Evaders should use a wand to check for booby traps between strands of multi-strand barbed wire. Generally, they should penetrate the fence under the wire

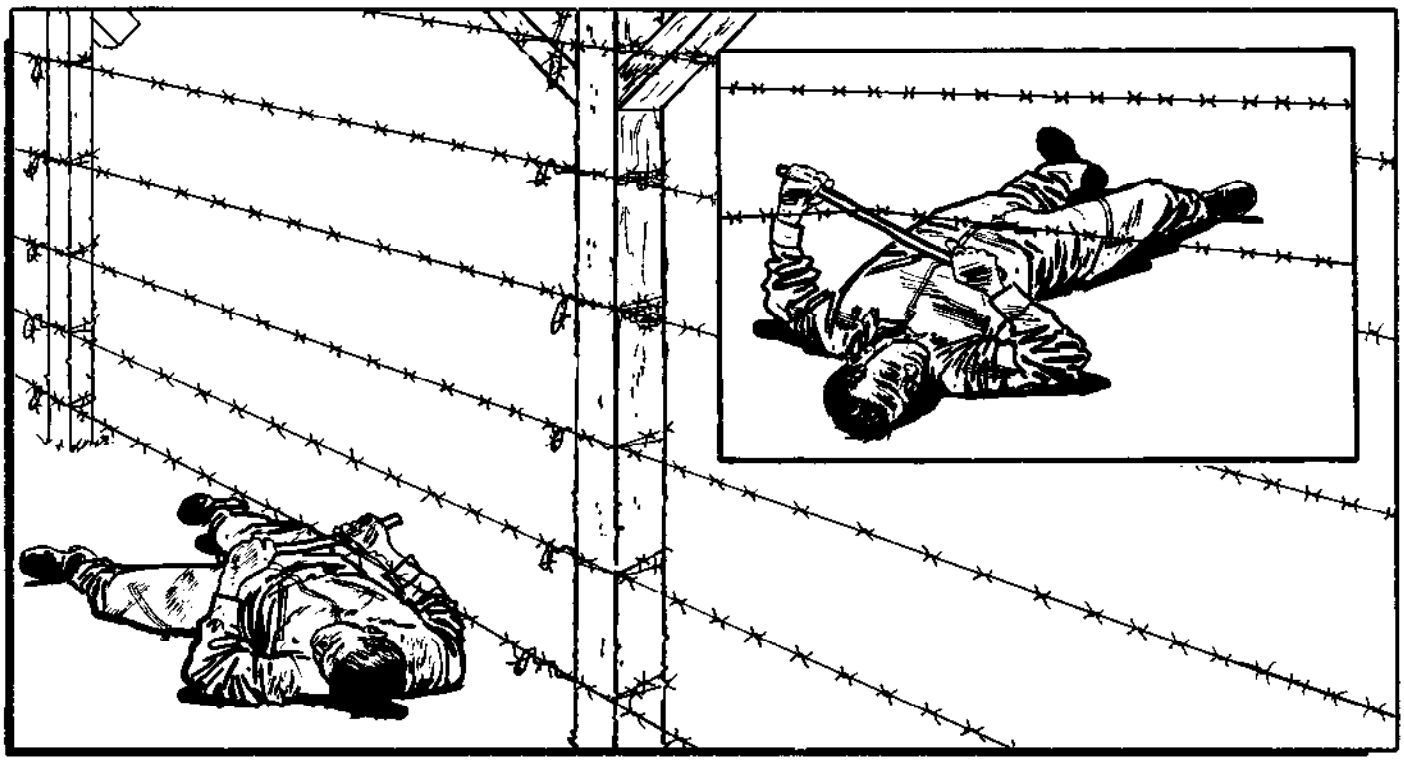

Figure 28-59. Penetrating Wire.

Figure 28-60. Penetrating Concertina Wire.

closest to the ground with the body parallel or perpendicular to the wire, depending on circumstances (figure 28-59). They should lie flat on their backs both to project the lowest possible silhouette and to provide good visibility of the wire against the sky. The probe can sometimes be used to lift the wire. If the lowest wire is close to the ground and is tight, evaders may have to modify their approach to the problem.

(c) The apron fence is penetrated in the same manner as any multi-strand fence—one wire at a time. Evaders should check the area between wires before proceeding.

(d) Concertina wire is penetrated with the body perpendicular to the wire using a probe to lift the wire if it is not secured to the ground (figure 28-60). If the wire is secured to the ground, the evader can crawl between the loops. If two loops are not separated enough, they may be tied apart using shoe laces, string, suspension line, or strips of cloth. After passing through, the ties should be removed for future use and to erase evidence of travel.

(e) Chain link fences should be avoided completely if at all possible. These fences are usually found only in highly sensitive zones. This means the area is probably more highly guarded and patrolled than other areas. There also may be other traps or devices installed. The fence may also be electrified. If the fence must be penetrated, the evader should go under it if possible (figure 28-61). If digging is required, the soil should be placed on the opposite side so it can be re-

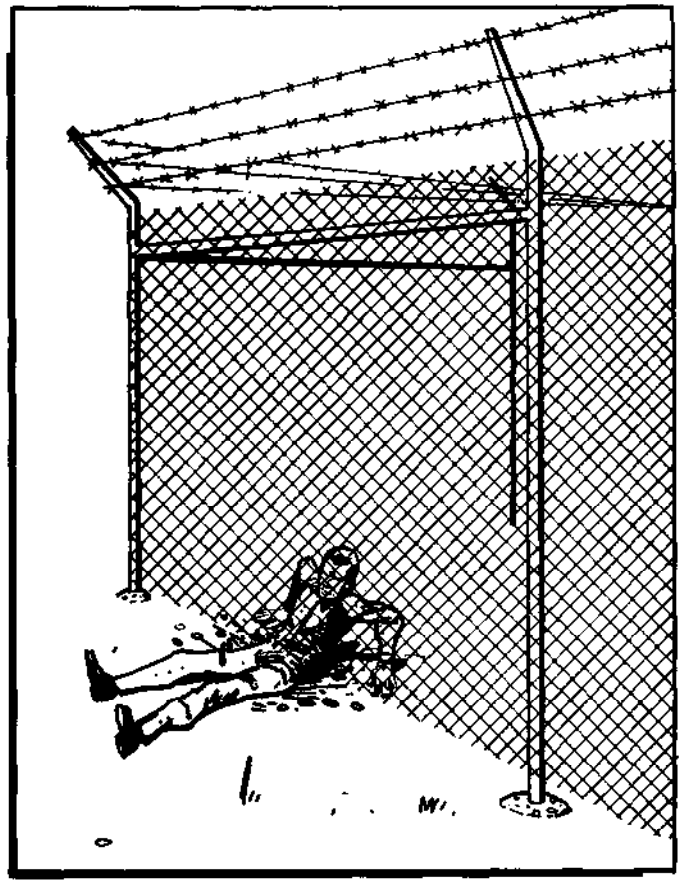

Figure 28-61. Penetrating Chain Link Fence.

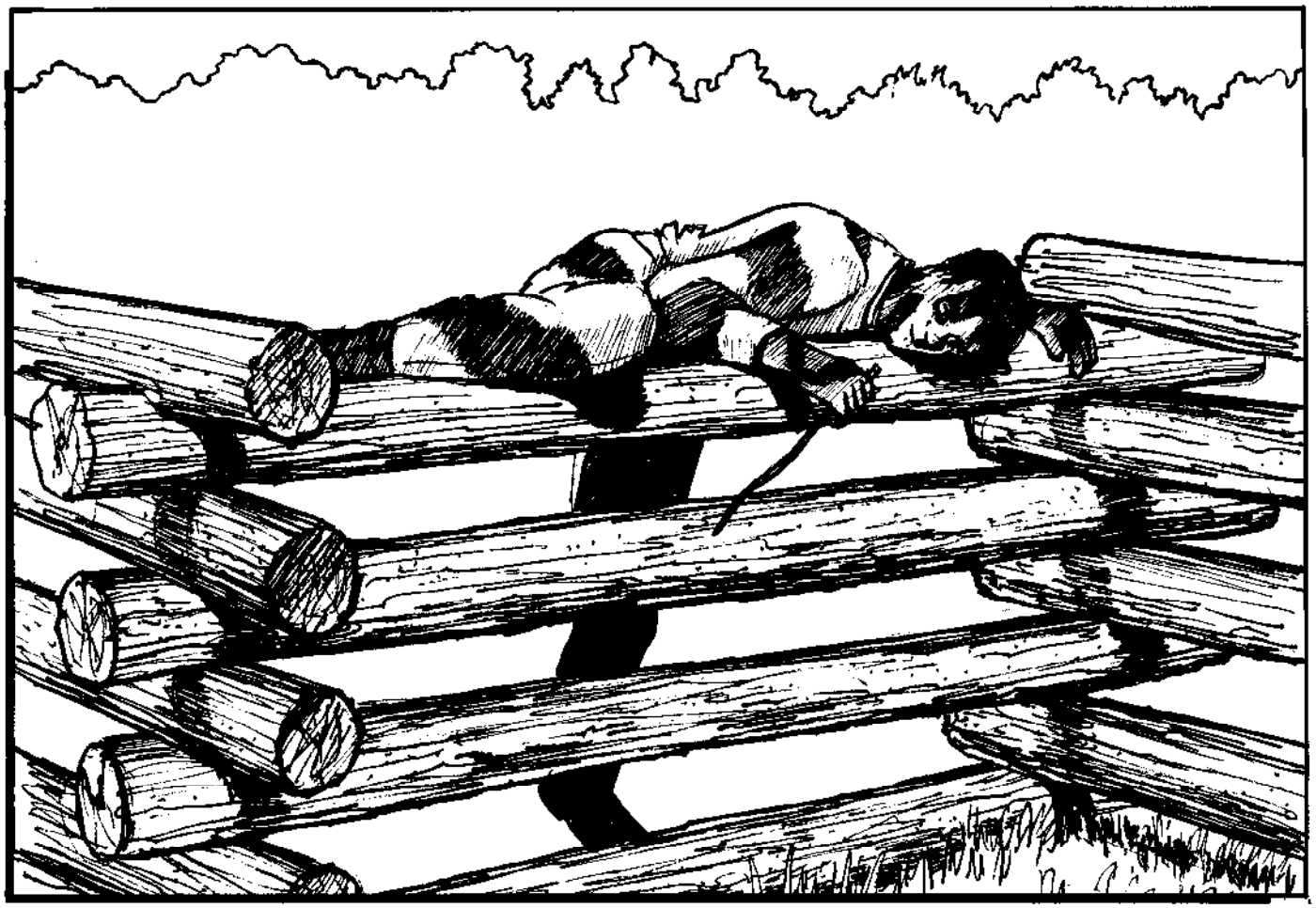

Figure 28-62. Penetrating Log Fence.

placed to remove evidence. Climbing the fence is recommended only as a last resort.

(f) Evaders may encounter rail and split-rail fences while evading or escaping. The fences are penetrated by going under or between lower rails if possible. If not, evaders should go over at the lowest point, projecting as low a silhouette as possible (figure 28-62). They should check between the rails and on the other side of the fence to detect tripwires or booby traps. Firmness of the ground should be checked on both sides of the fence. The body should be parallel to the fence before penetration.

(4) Raked or plowed areas may be found in areas of both low and high density security. If such an area is encountered, evaders should roll across the area, after making sure it is not a mine field, to avoid leaving footprints; or they may side-step, walk backwards, or brush out footprints. Any of the above may be done when it is a requirement not to leave clear-cut evidence of the direction of movement.

(5) Roads are common barriers to evasion and escape. When roads are encountered, evaders should closely observe the road from concealment to determine enemy travel patterns (figure 28-63). Crossing from points offering best concealment such as bushes, shadows, etc., is best (figure 28-64). Evaders should cross at straight stretches of road in open country and on the inside of curves in hilly or wooded areas. This allows the evader to see in both directions so the chances of being spotted or surprised in the open is minimized. Avoid leaving tracks both in the road and on the shoulder of the road.

(6) Culverts and drains offer excellent means of crossing a road unobserved (figure 28-65).

(7) Railroad tracks often lie in the path of evaders. If so, evaders should use the same procedures for observation as for roads. If it is determined that tracks are patrolled, a check should be made for booby traps and tripwires between tracks. Aligning the body parallel to

Figure 28-63. Common Barriers.

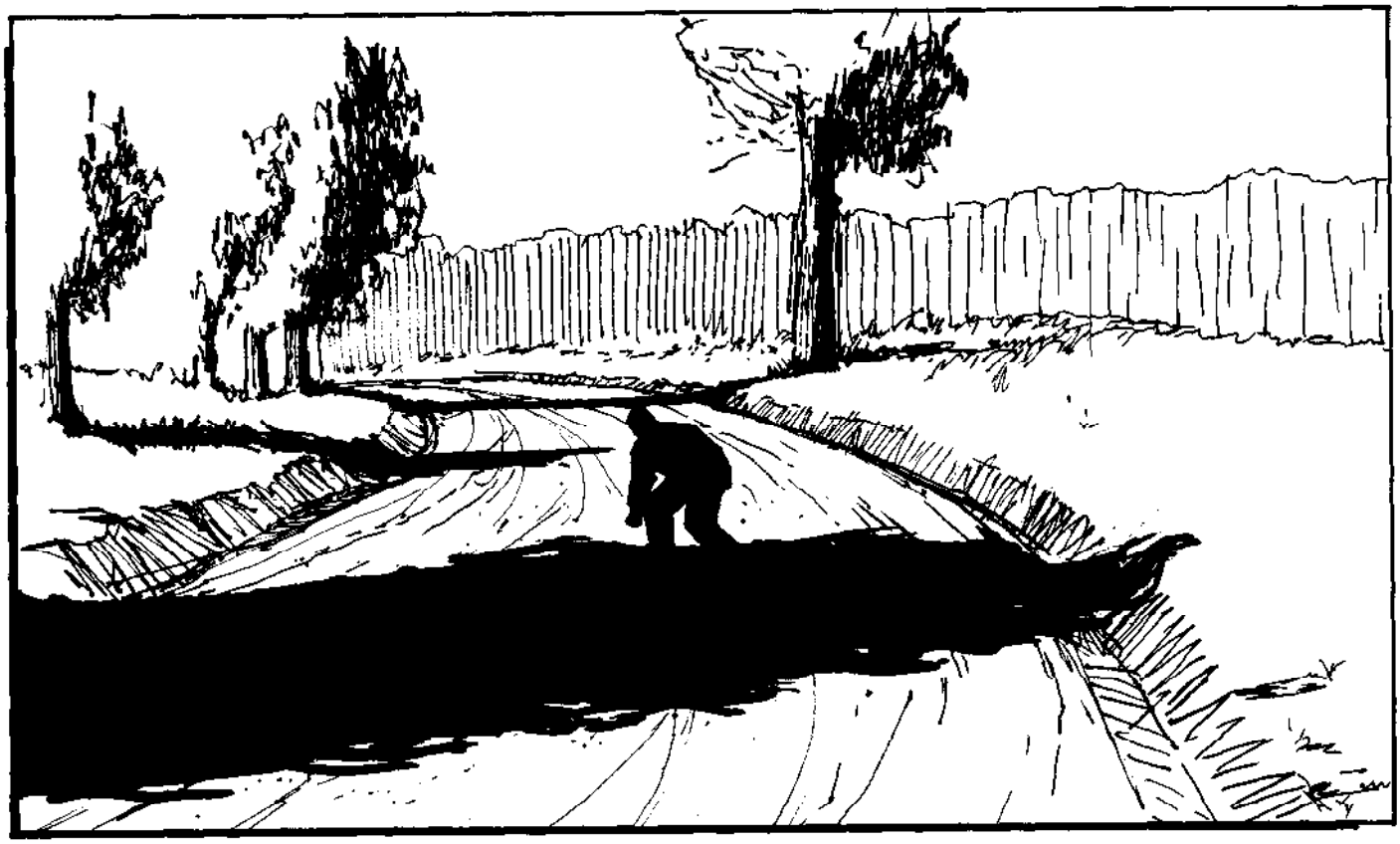

Figure 28-64. Crossing Roads.

Figure 28-65. Crossing at Culverts.

the tracks with face down next to the first track, evaders should carefully move across the first track in a semi-pushup motion, repeating for the second track and subsequent sets of tracks (figure 28-66). If there is a third rail, they should avoid touching it as it could be electrified. Sound detectors can also be attached to the tracks and can reveal any crossing if a track is touched. If determined from observation that the tracks are not

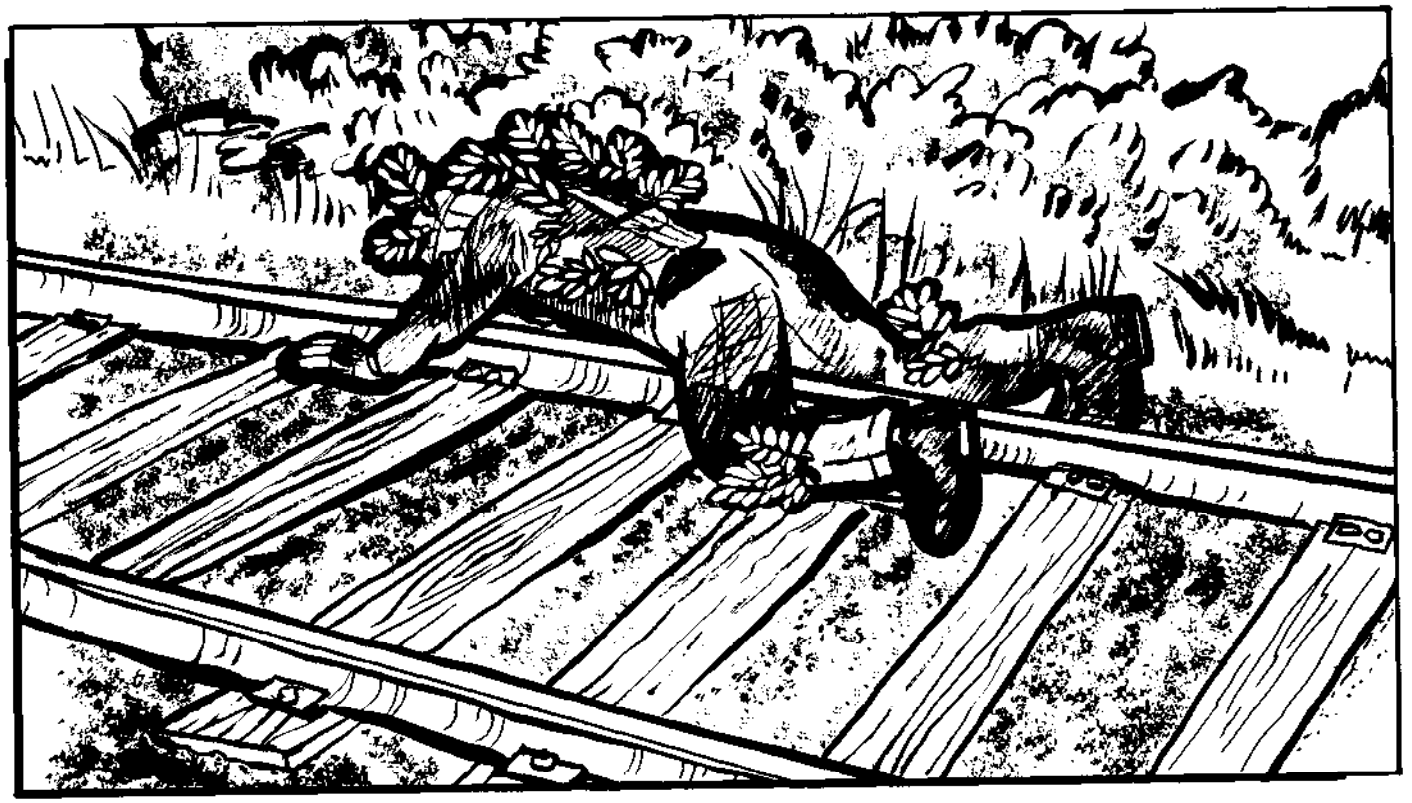

Figure 28-66. Crawling Over Railroad Tracks.

Figure 28-67. Climbing Down Cliff.

patrolled, evaders should cross in a normal walking or hands-and-knees manner, attempting to attract as little attention as possible. Evaders should try to keep their hands and feet on the railroad ties to prevent leaving foot or hand prints in the adjacent soil or gravel.

(8) Deep ditches (such as tank traps or natural drainages) may be obstacles with which evaders must deal. Ditches should be entered feet first to avoid injury to the head or upper torso should there be large rocks, barbed wire, or other hazards at the bottom (figure 28-67). Using a wand to detect tripwires and booby traps in the ditch and on the sides is highly recommended. Maintaining a low silhouette upon exiting the ditch is imperative.

(9) Open terrain complicated by guard towers or walking patrols is a definite hazard to evaders. These areas should be avoided if possible. If it becomes necessary to traverse open terrain or come near guard towers, evaders should stay low to the ground and, when possible, travel at night or during inclement weather. Use terrain masking since night vision devices may be used near border areas (figure 28-68).

(10) The problem of crossing areas which have been contaminated as a result of enemy or friendly NBC operations may arise. Chemical contamination should be suspected when the following are observed (figure 28-69):

(a) Shell craters with liquid in the bottom.
(b) Liquid droplets on vegetation.
(c) Water with "film" on the surface.
(d) Unexplained dead animals.
(e) Unseasonal discoloration of vegetation.

NOTE: Without protective clothing, mask, and accessories, evaders should bypass contaminated areas if possible.

(11) CAUTION: Stay away from borders unless absolutely necessary. The crossing of one or more borders presents a major problem. These areas may be located in any type of terrain.

(a) In areas where there is no well-defined terrain feature to indicate the border, artificial obstacles such as electrified or barbed wire fences, augmented with tripwires, anti-personnel mines, or flares may be encountered. Open areas may be patrolled by humans or dogs, or both, particularly during the hours of darkness. The enemy may also use floodlights and plowed strips as aids to detecting evaders (figure 28-70).

(b) The plan to cross a border must be deliberate and must be designed to take advantage of unusually bad weather (as a major distraction to the enemy force) or areas where security forces are overextended. These areas are usually found where there are natural obstacles.

(c) Crossings should be made at night, when possible, in battle-damaged areas. If it is necessary to cross during daylight hours, evaders should select a crossing point which offers the best protection and cover. They should then keep the area under close observation for several days to determine:

-1. The number of guards in the area.
-2. The manner of their posting.
-3. Aerial patrols and their frequency.
-4. The limits of the areas they patrol.
-5. Location of mines, flares, or tripwires.

(12) A difficult task in any situation is the attempt to cross the forward edge of the battle area. If unable to determine the general direction to friendly lines, evaders should remain in position and observe the movement of enemy military forces or supplies, the noise and flashes of the battle area, or the orientation of enemy artillery. After arriving in the combat zone, evaders should select a concealed position from which as much of the battle area as possible may be observed. They can then select a route and critical terrain features on which they can guide when infiltrating back to friendly positions under the cover of darkness. Several alternate

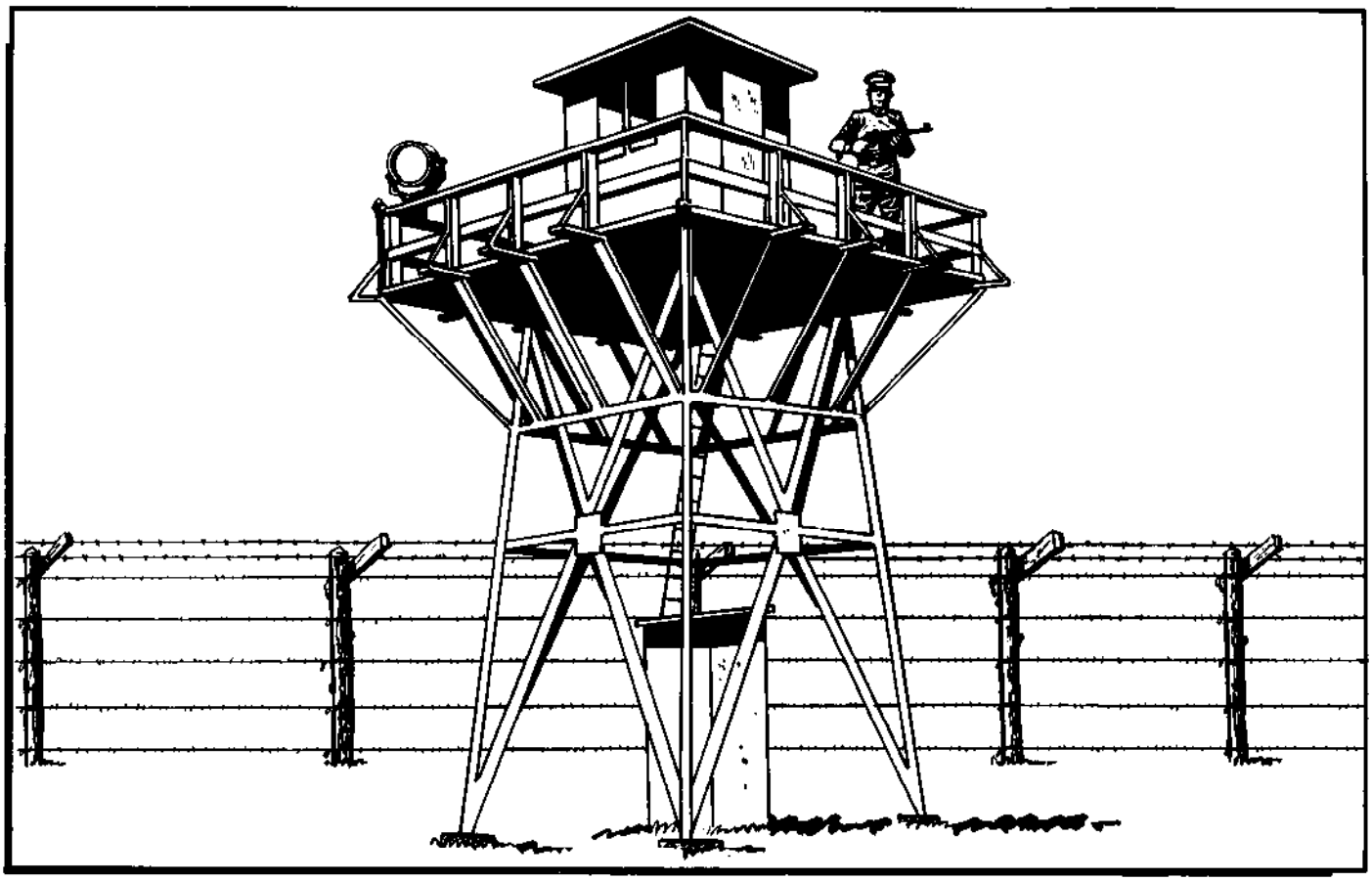

Figure 28-68. Open Terrain.

Figure 28-69. Contaminated Areas.

Figure 28-70. Border Crossings.

routes should be selected with care to avoid "easy" approaches to friendly lines which are more likely to be covered by friendly fire and enemy patrols. If in uniform, select exposure time during daylight hours and be close enough to be easily recognized by friendly troops.

(13) Evaders should also watch out for friendly patrols. If a patrol is spotted, evaders should remain in position and allow the patrol to approach them. When the patrol is close enough to recognize them, evaders should have a white cloth displayed before the patrol gets close enough to see the movement. A patrol may fire at any movement. Shouting or calling out to them jeopardizes both the patrol and the evader. Evaders should stand silently with hands over their head and legs apart so their silhouettes are not threatening. If evaders elect not to make contact with a patrol, they should, if possible, observe their route and approach friendly lines at approximately the same location. This may enable them to avoid mine fields and booby traps. (NOTE: The practice of following any patrol is extremely dangerous. The last persons in line are charged with security, and anyone following them would be considered hostile and eliminated.)

(14) If unable to contact a friendly patrol, the only alternative for evaders may be to make a direct approach of front-line positions. This will require them to crawl through the enemy's forward position near forward friendly elements. This action should be done during the hours of darkness. Once near friendly lines, however, evaders should not attempt to make contact until there is sufficient light for them to be recognized.

(15) In dogs, the ability to perceive odors is much greater than that of humans.

(a) In this portion, the term dog is meant to describe only the animals specifically trained in the areas of patrolling, guarding, and searching. Since dogs are basically odor-seeking animals, anything developed to work against their odor-seeking capabilities is worthy of experimentation for survivors (evaders).

(b) A problem which must be considered is evaders will not be working solely against a dog, but against a dog handler as well. There is no simple, sure method of evading a dog. Some possible means which could be tried by evaders are:

-1. Dogs detect the fatty acids in dead-skin cells that humans shed by the thousands every day. Scenting dogs may be distracted by scattering an irritant such as pepper behind the evader or traveling across an asphalt road on a hot day.

-2. Dogs track better when the weather is humid and the air is still—there is less evaporation and dissipation of odor.

-3. If evaders know they are being followed by dogs, either from the landing site or as a result from an escape, they should try to use water to conceal their tracks and eliminate their scent.

-4. If there is a choice of terrain and it is possible to travel on a hard surface, evaders should do so rather than travel on soft ground.

-5. Evaders should always attempt to move downwind of a dog. This should be attempted when they are traveling in open country or penetrating obstacles such as dog guard posts or border areas. If penetration of obstacles or escape is planned (after careful location and study of guard and dog areas of responsibility and their methods), evaders should select a time for movement when noise will distract the dog to a point away from the planned maneuver.

-6. If evaders are physically capable, they should attempt to maintain the maximum distance possible from dogs. Moving fast through rugged terrain will slow and probably defeat the handlers of dogs. Here evaders must choose between making mistakes in travel techniques while evading or being caught by dogs if they don't move fast enough.

28-18. Evasion Aids:

a. Survival Kits. Personnel may sometimes find it practical to devise and carry compact personal survival kits to complement issued survival equipment. If E&E kits are provided, potential evaders should be familiar with them, their uses, and their limitations.

b. Maps. Any maps of the area in an evader's possession should not be marked. A marked map in enemy hands can lead to the compromise of people and locations where assistance was given. Evaders should be wary of even accidentally marking a map; for example, soiled fingers will mark a map just as plainly as a pencil.

c. Pointee-Talkee. The "pointee-talkee" is a language aid which contains selected phrases in English on one side of the page and foreign language translations on the other side. To use it, evaders determine the question and statement to be used in the English text and then point to its foreign language counterpart. In reply, the natives will point to the applicable phrase in their own language; evaders then read the English translation.

(1) The major limitation of the "pointee-talkee" is in trying to communicate with illiterates. In many countries the illiteracy rate can be astoundingly high, and personnel have to resort to pantomime and sign language which have been relatively effective in the past.

(2) "Pointee-talkee" phrases are presented under the following eight subheadings:

(a) Finding an interpreter.
(b) Courtesy phrases.
(c) Food and drink.
(d) Comfort and lodging.
(e) Communications.
(f) Injury.
(g) Hostile territory.
(h) Other military personnel.

d. Barter Kits. Barter kits may be available in some commands. If not, crewmembers may elect to develop their own. Items for consideration should be selected from area studies. Some items to consider might be rings, watches, knives, local currency, coins, and lighters. Items should have no markings of personal significance or military value. Military items packed in kits may be considered if not essential to the evader. Flashing large amounts of cash or valuables can have negative results in a depressed, war-torn area. Show only small amounts and drive a hard bargain.

e. Other Evasion Aids. Information on other evasion aids and tools is available from unit intelligence officers.

f. Assisted Evasion. There may be people in a hostile nation or in an enemy-occupied country who are dissatisfied with existing conditions.

(1) History has revealed that in every major conflict there are groups of people in every country who will aid a representative of their government's enemies. The motivating force behind these groups may vary from purely monetary considerations to idealistic concepts of government reform. In many cases, their real objective will be the political advancement of their particular group. During WW II, many underground or resistance groups and movements were established in occupied countries. One of the major purposes or functions of these groups was the aiding of downed allied aircrew personnel. In most cases, the driving force behind these movements was patriotism and desire for political recognition for their cause.

(2) These circumstances favor active resistance movements. One of the functions of such movements may be the operation of escape and evasion (E&E) systems for the purpose of returning evaders to friendly territory.

(3) US Special Forces may also organize and operate E&E mechanisms (figure 28-71).

(a) E&E Organizations. Assistance may range from that rendered by a sympathetic individual to elaborate E&E nets organized by local inhabitants. E&E organizations may be limited in nature, such as providing assistance to reach a national frontier, or they may be linked to larger organizations capable of returning to friendly control.

(b) Acts of Mercy. These are usually isolated events during which evaders may be provided food, shelter, or medical attention for a brief period of time. Local people may find an exhausted or incapacitated

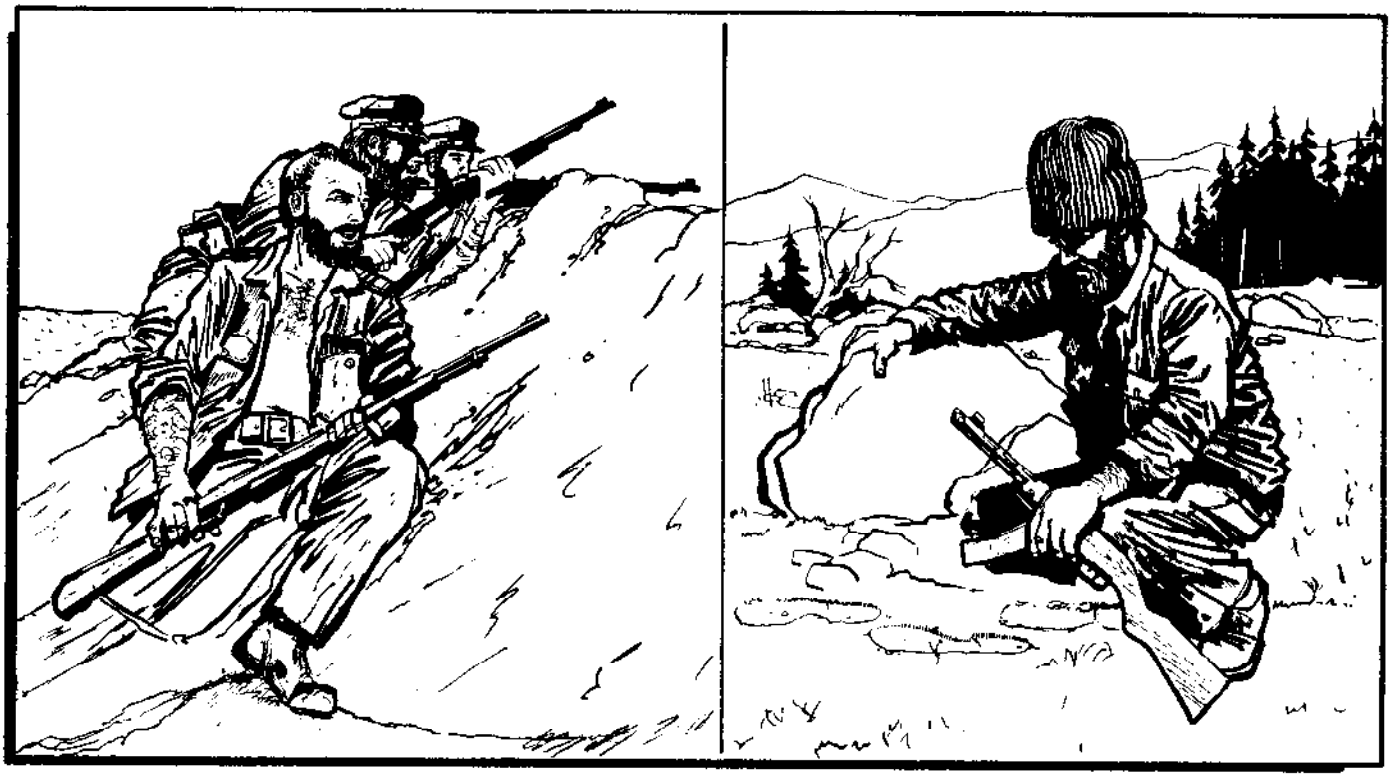

Figure 28-71. E&E Organizations.

evader and provide that evader with limited sustenance. This type of assistance is frequently offered with reluctance or under fear of reprisal because an act of providing comfort to the enemy would mean punishment. Unless an evader is in immediate need of medical attention, acts of mercy may consist of only an offer of food followed by an urgent plea that the evader leave the area immediately. If an evader is physically able to depart with a reasonable chance of evading capture, he or she should do so. An evader should not insist on receiving additional aid other than what is offered by the person who renders assistance only through human impulses (figure 28-72).

(4) Evaders must understand when dealing with any indigenous personnel while in enemy territory, their own actions will govern the treatment they will receive at the hands of these people. How evaders conduct themselves may also have much to do with getting back to their own forces should they fall into the hands of irregulars friendly to their (the evaders) own cause. The following list of suggestions may be a useful guide in dealing with these people or groups.

(a) Evaders should understand that failure to cooperate or obey may result in death.

(b) Evaders should try to avoid making any promises they cannot personally keep.

(c) Evaders should remember that the four conditions called for by the rules of land warfare must be met before members of an irregular group can be recognized as qualifying for PW status in the event of capture. These same rules will also apply to evaders who are involved with these groups.

(d) Prisoners of war, according to the current Geneva Convention, are persons belonging to one of the

Figure 28-72. Receiving Aid.

following categories, who have been captured by the enemy. Members of other militias and volunteer corps, including those of organized resistance movements, belonging to a conflicting Party and operating in any territory, even if this territory is occupied, provided that such groups fulfill the following four conditions of:

-1. Being commanded by a person responsible for subordinates.

-2. Having a fixed distinctive sign recognizable at a distance.

-3. Carrying arms openly.

-4. Conducting their operations in accordance with the laws and customs of war.

(5) If evaders join such a unit, the closest thing to a guarantee of treatment as a military person will be their uniforms if they are captured.

(6) If evaders have the opportunity to influence the group they are with, they should try to encourage them to abide by the four conditions mentioned.

(7) Evaders must avoid becoming associated in any way with atrocities these groups may commit against civilians, prisoners, or enemy soldiers.

(8) Evaders should not become involved in their political or religious discussions, take sides in their arguments, or become involved with the opposite sex.

(9) Evaders should show consideration for being allowed to share food and supplies. It may also be helpful if evaders understand and show interest in the assister's customs and habits.

(10) The overall best and safest course for evaders to follow is to exercise self-discipline, display military courtesy, and be polite, sincere, and honest. Such qualities are recognized by any group of people throughout the world. The impression left can influence the aid provided to future evaders.

g. E&E Lines. An E&E line is a system of one or more secret nets organized to contact, secure, and when possible, evacuate friendly personnel. Well-organized and supported lines normally can be expected to provide the following assistance:

(1) Temporary shelter, food, and equipment for the next phase of the journey.

(2) Clothing and credentials acceptable in the area to be traveled.

(3) Information concerning enemy security measures along the evasion route.

(4) Local currency and transportation.

(5) Medical treatment.

(6) Available native guides.

h. Conduct of E&E Lines. The success of an E&E organization depends almost entirely upon its security. The organization of a line includes much planning and work carried out under dangerous conditions. The security of the system often depends upon the evader's cooperation and working knowledge of how it functions, how it may be contacted, and what rules of personal conduct are expected of the evader. The following paragraphs summarize the major aspects of the operation of an E&E line.

(1) Contacting the Line. Premission briefings may inform evaders where to go and what actions to take to make contact with an E&E mechanism. After being picked up by an evasion and escape mechanism, evaders will be moved under the control of this mechanism to territory under friendly control or to a removal area, and arrangements may be made for air or sea rescue. The organizer of a line in friendly but enemy-occupied territory normally will have arranged a network of spotters who will be especially active when evaders are in the immediate area, but so will the enemy police and counterintelligence organizations. For this reason, certain precautions must be observed when making contact.

(2) Approach. Help may be refused by a person simply because he or she thinks someone else has seen the evaders approach to seek assistance. If captured with a local helper, an evader will become a prisoner, but the helper and perhaps an entire family may be more severely punished.

(3) Making Contact. Contacts with the natives are discouraged unless observation shows they are dissatisfied with the local governing authority, or previous intelligence has indicated the populace is friendly. Evaders should proceed to, and remain in, the nearest SAFE area where arrangements for contact can be developed. If the E&E system is operating successfully, the spotter will know evaders are present and will search the immediate area, making frequent visits to designated contact points. Identification signs and countersigns, if used, may be included in the preoperational briefing. It is seldom advisable to seek first contact in a village or town. Strangers are conspicuous by day, and there may be curfews or other security measures during the hours of darkness. The time of contact should be at the end of the daylight period or shortly thereafter. Darkness will add to the chance of escape, if the contact proves to be unfriendly, and may be advantageous to the contact in providing further assistance.

(4) Procedure After Contact. If contact is made, evaders may be told to remain in the vicinity where spotted, or, more likely, they will be taken to a house or other structure used by the E&E net as a holding area. It must be decided at this time whether or not to trust the contact. If there is any doubt, plans should be made to leave at once. It is also possible that the house may not belong to the E&E organization but rather to someone who will look after evaders until arrangements can be made for the line to identify and accept them in E&E net.

(5) Establishing Identity. Verification of identity will be required before anyone is accepted as a bona fide evader. The constant danger facing the operators of an

escape line is the penetration of the E&E system by enemy agents pretending to be evaders or escapees. Evaders should be prepared to furnish proof of identity or nationality. Since it may lead to later difficulties of identification, they should never give a false name—just their name, grade, service number, and date of birth. It is best for them to avoid talking as much as possible.

(6) Awaiting Movement on the Line. Delays can be expected while proceeding along the escape line. If the period of waiting is prolonged, frustration and impatience may become unbearable, leading to a desire to leave the holding area. This must not be done, because if seen by other people, the lives of the assisting personnel and the existence of the entire line itself may be endangered.

(a) Evaders should follow the orders of those assisting them. If kept indoors for any length of time they can keep fit by moderate physical exercise.

(b) The host should have a plan for rapid evacuation of the area if enemy personnel should raid the holding area; if not, evaders should have a personal plan to include measures for moving all traces of having occupied the area. If the net is being overrun and capture is imminent, evaders must be prepared to fend for themselves. The evader is the only one who knows more than one part of the net. The assisters may attempt to eliminate the security risk to the net.

(7) Traveling the Line. It would be a grave breach of faith and security for evaders to discuss, at any point on the line, the earlier stages of the journey. Evaders might be tested to see if they are trustworthy—they should discuss nothing of the net. For security reasons and to protect the compartmentalization of the line, no information should be revealed. It is also useless to ask where a line leads or how it will eventually reach friendly territory. Evaders should not try to learn or memorize names and addresses and, above all, they shouldn't put these facts or any other information in writing. Evaders should give the impression of having received no assistance from local inhabitants.

(8) Fellow Evaders. Caution is required in the case of fellow evaders on an escape line unless they are personally known. Even when it has been satisfactorily determined that another person is a genuine evader, no information should be given.

(9) Travel with Guides. If under escort, this fact should not be apparent to outsiders. In a public vehicle, for example, evaders should never talk to their guide or appear to be associated with the person in any way unless told to do so. This will lessen the possibility of both the evader and the guide being apprehended if one should arouse suspicion. It should always be possible for the guide to disown an evader if the guide gets into difficulty. When evaders are escorted, they should follow the guide at a safe distance, rather then walk right next to the person, unless instructions indicate the latter action is required (figure 28-73).

(10) Speaking to Strangers. Evaders should never speak to a stranger if it can be avoided. As a last resort, they should pretend to be deaf and dumb or even half-witted. This technique has often been successful. To discourage conversation in a public conveyance, they can also pretend to read or sleep.

(11) Personal Articles and Habits. Evaders should not produce articles in public which might show their national origin. This pertains to items such as pipes, cigarettes, tobacco, matches, fountain pens, pencils, and wristwatches. Evaders should also ensure their personal habits do not give them away; for example, they should not hum or whistle popular tunes or utter involuntary oaths. Again, in restaurants, imitating local customs in the use of knives and forks and other table manners is advisable. Study of the area before the mission may help evaders avoid making mistakes.

(12) Payment to Helpers. On an escape line, evaders should not offer to pay for board, lodging, or other services rendered. These matters will be settled afterwards by those who are directing and financing the line. If in possession of escape kits or survival packs, evaders should keep them as reserves for emergencies. If they

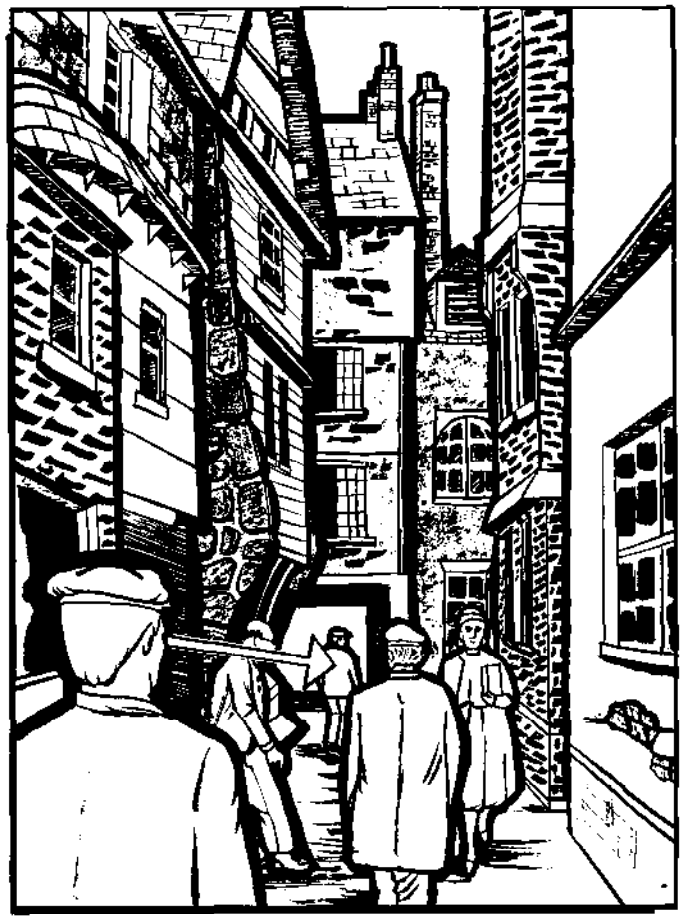

Figure 28-73. Travel with Guide.

have no food reserve, they should try to build up a small stock in case they are forced to abandon the line.

(13) Evaders Conduct:

(a) Be polite by local standards.

(b) Be patient and diplomatic.

(c) Avoid causing jealousy. Disregard the sex of the assisters.

(d) Avoid discussions of a religious or political nature.

(e) Eat and drink if asked, but don't over indulge or become intoxicated.

(f) Don't take sides in arguments between assisters.

(g) Don't become inquisitive or question any instruction.

(h) Help with menial tasks as directed.

(i) Write or say nothing about the other people or places in the net.

(j) Don't be a burden; care for self as much as possible.

(k) Follow all instructions quickly and accurately.

28-19. Combat Recovery:

a. Air recovery is one of several means of transportation for downed crewmembers in their quest for their final goal—returning to their own lines and units. How recovery will be effected will depend on a number of factors, among them are the following:

(1) Terrain.

(2) Capability of the rescue vehicle.

(3) Condition of the survivor (evader).

(4) Enemy activity in the area.

(5) Weather conditions.

(6) High or priority mission.

b. Even though a maximum effort will be made to recover downed aircrew members, survivors can jeopardize the whole rescue operation and the lives of rescue personnel in a combat zone by not taking a responsible role in recovery operations. The responsibilities are many and varied but essential to a successful rescue mission. Evaders should recall that even though they may have little experience in participating in actual rescues, they must nonetheless be very proficient in their actions. Evaders should always remember other people are endangering their lives in an attempt to retrieve them.

c. There are two basic types of air-recovery vehicles—rotary wing and fixed-wing aircraft. The rotary wing type can extract evaders from remote areas. With air-to-air refueling, the range of these aircraft has increased; they are limited only by the endurance of the crewmembers who operate them.

d. Survivors must be proficient in the operation of all the survival equipment at their disposal. For example, they must be able to switch off the beeper on the radio to receive verbal instructions from the rescue commander.

e. The initial contact with rescue or other combat aircraft in the area must be done as directed by authorities in that theater of operations; for example, in Southeast Asia, a contact method of transmitting 15-seconds beeper, 15-seconds voice (call sign), and 15-seconds monitoring was used. The method an evader should use to establish contact with rescue will be briefed before the mission.

f. One important aspect of the rescue process is evader authentication by rescue personnel. Survivors must be able to authenticate their identity through the use of questions and answers, responding as directed before the missions. Authentication in a combat area changes rapidly to reduce the chances of compromising the rescue efforts. Survivors (evaders) must keep up with these changes so their rescue can be made without undue danger to rescue personnel.

g. The purpose of filling out and using a personal authenticator card is positive verification of an evader's identity which is essential before risking search and rescue aircraft or the lives of assisters. The purpose of the photographs, descriptions, fingerprints, etc., on the card is to ensure all possible means are used for the proper identification of personnel. Intelligence personnel are responsible for ensuring that personnel fill out the card completely and that they are aware of the purpose and use of the information. When filled in, the personal authenticator card becomes classified Confidential, and is reviewed at least semiannually by both the crewmember and intelligence personnel. (NOTE: See AFR 64-3, Wartime Search and Rescue (SAR) Procedures.)

h. When selecting a site for possible evasion recovery, there is much for evaders to keep in mind. The area they choose could well decide the success or failure of the mission.

(1) The potential recovery site should be observed for 24 hours, if possible, for signs of:

(a) Enemy or civilian activity.

(b) Roads or trails.

(c) Farming signs.

(d) Orchards.

(e) Tree plantations.

(f) Domesticated animals or droppings.

(g) Buildings or encampment areas.

(2) The rescue site should be observable by aircraft but unobservable from surrounding terrain if possible. There should also be good hiding places around the area. The site should include several escape routes so evaders can avoid being trapped by the enemy if discovered. There should be a small open area for both signaling and recovery. It would be beneficial if the surrounding terrain provided a masking effect for rescue forces in order to avoid enemy observation and fire.

i. The size of evasion ground-to-air signals should be as large as possible but must be concealed from people passing by. Evaders should remember the six-to-one ra-

tio. The contrast these signals make with the surrounding vegetation should be seen from the air only. Any signal displays should be arranged so they can be removed at a moment's notice since enemy aircraft may also fly over the immediate area.

(1) All of the principles of regular (nontactical) signals should be followed by those building evasion signals. Crewmembers may be prebriefed as to the appearance of specially shaped signals (figure 28-74).

(2) Evasion signals, like all others, must be maintained to be effective. At times, evaders may be preinstructed to set out their signals according to a prearranged time schedule.

j. Whatever signal devices are available to evaders, they must be able to use them (mirror, flare, gyro-jet, etc.) with proficiency. These signals, like those used in nontactical situations, are to be used either to gain the attention of friendly aircraft or rescuers, or when directed by rescue. Extreme care must be taken to minimize or eliminate chances of enemy elements spotting the signals. For example, the strobe light and mirror can be directed and aimed instead of being used in an indiscriminate manner.

(1) If evaders are in or on the water, they should use lights, flares, dye, whistles, etc., with extreme care as they are readily distinguishable over water at great distances.

(2) In addition to knowing how to use radios correctly to effect their rescue, evaders should also be familiar with the special points of evasion radio procedures; for example, ensuring the radio transmits continuously during a night recovery operation using the electronic locator finder (ELF) system.

(3) It is also the evaders' responsibility to communicate all signs of enemy activity, such as:

(a) Locating anti-aircraft emplacements.

(b) Identifying when they are firing.

(c) Assisting strikes by spotting hits (high, low, or on target).

(d) Determining effectiveness of hits.

(e) Notifying personnel of changes in small arms positions, etc.

(4) Downed evaders can normally expect to be hoisted to a helicopter by one of five methods: basket, Stokes litter, bell, horse collar, or forest penetrator. Other pickup devices which may be used are the McGuire Rig, Swiss Seat, Motley Rig, Stabo Rig, rope ladder, or rope. Another method of recovery which may be employed to extract evaders is the Surface-to-Air-Recovery (STAR) System (figure 28-75). Evaders should remember whatever the pickup device, it should always be grounded before they grasp it.

(5) There is also fixed-wing capability of rescuing downed crewmembers. Evaders will be prebriefed as to which type rescue vehicle and systems to expect in their areas of operation. No matter which type is used, evaders must be capable of mounting and riding the rescue

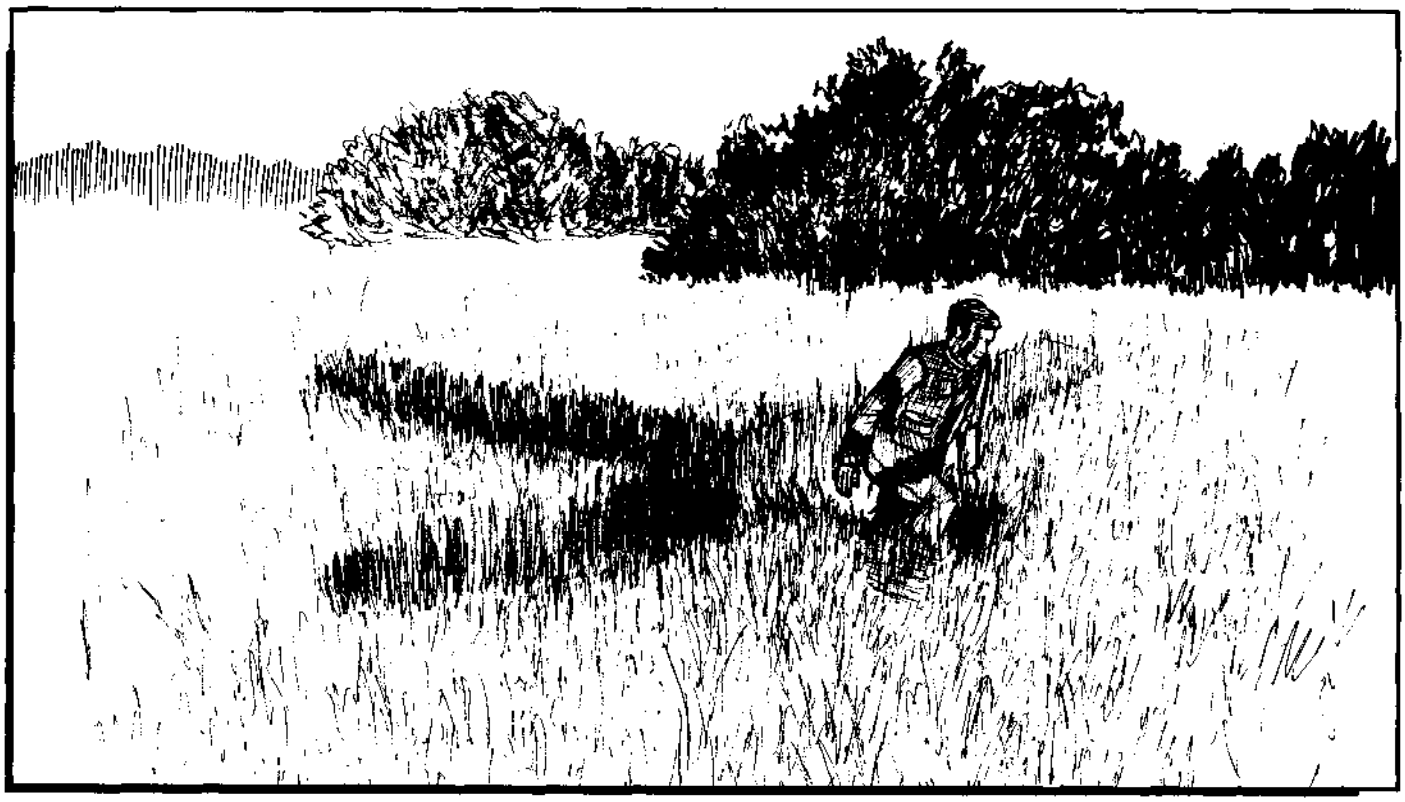

Figure 28-74. Evasion Signals.

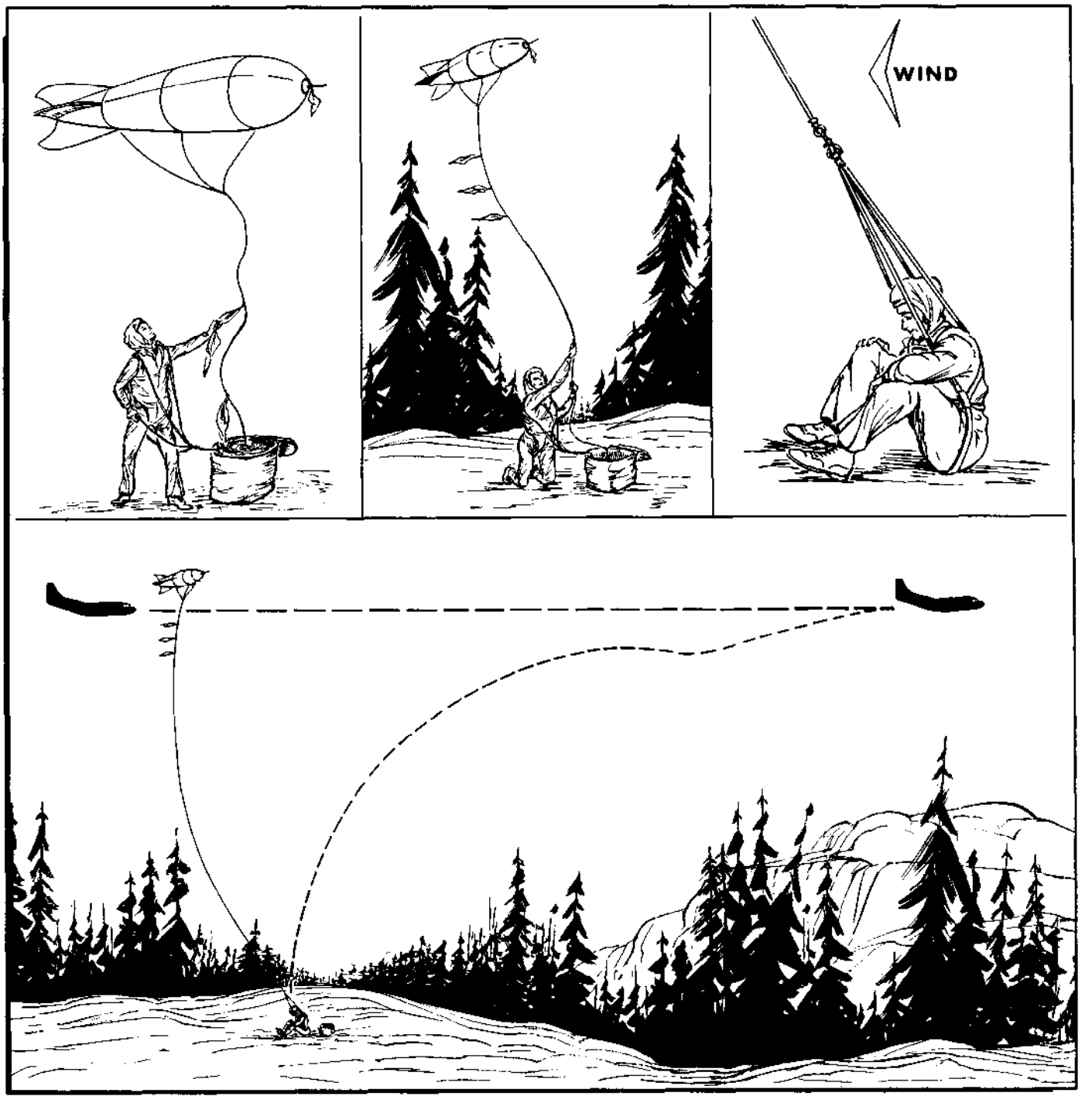

Figure 28-75. Surface-to-Air Rescue.

devices in a minimum amount of time, allowing the rescue craft to effect the rescue and depart quickly.

(6) As downed crewmembers, whether as a result of enemy action or of mechanical failure, it is important they all become familiar with the information that I day could prove instrumental in saving their lives. First, they must have prepared themselves to cope with the survival situation before an aircraft emergency. This can be done by including, as part of their preflight planning, a thorough inspection of survival equipment to determine its availability and serviceability. Second, they must realize there are many types of recovery vehicles available in the Air Force inventory, and a knowledge of all recovery techniques is a necessity. They should request detailed briefings from area or local rescue personnel who will explain the operational limita-

tions and recovery potential for each. Third, evaders must be capable of fulfilling their part in the recovery operation. This can be done by knowing when, where, and how to initiate communications and how to cooperate with the rescue aircraft crew.

(7) Other forms of assistance which may be available to evaders are Special Forces, Combat Control, RECON, SEALS, Riverine Operations, and Submarines. While aiding evaders is not the primary mission of these groups, as a secondary mission, they may travel in or near SAFE areas on their return journey to check for the presence of evaders.

Part Ten

INDUCED CONDITIONS — NUCLEAR, BIOLOGICAL, AND CHEMICAL (NBC)

Chapter 29

NUCLEAR CONDITIONS

29-1. Introduction. The possibility of "induced conditions" has served to intensify the difficulties of basic and combat survival because of the serious problems posed by nuclear, biological, and chemical warfare. Though the prescribed survival procedures recommended in other parts of this regulation are still applicable, a number of additional problems are created by the hazards of induced conditions (figure 29-1).

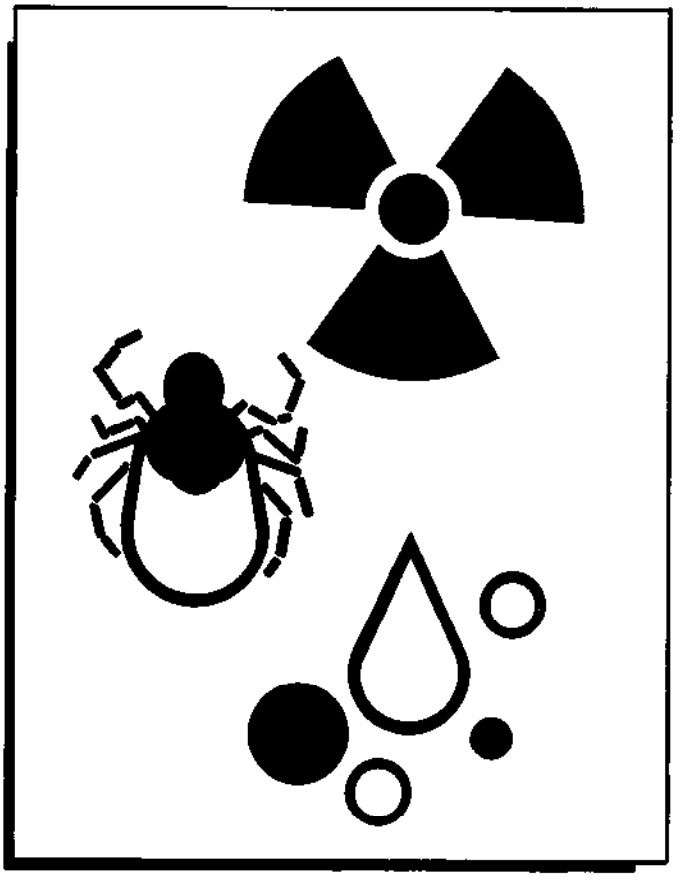

Figure 29-1. Induced Conditions.

29-2. Effects of Nuclear Weapons. Nuclear weapons cause casualties and material damage through the effects of blast, thermal radiation, and nuclear radiation. The degree of hazard from each of these effects depends on the type of weapon, height of the burst, distance from the detonation, hardness of the target, and explosive yield of the weapon.

a. Blast. The blast wave is the cause of most of the destruction accompanying a nuclear blast. After a nuclear detonation, a high-pressure wave develops and moves outward from the fireball. The front of the wave travels rapidly away from the fireball as a moving wall of highly compressed air. An example of the speed of the blast wave is: At 1 minute after the burst, when the fireball is no longer visible, the blast wave has traveled about 40 miles and is still moving slightly faster than the speed of sound. There are strong winds associated with the passage of the blast wave. These winds may have a peak velocity of several hundred miles an hour near ground zero. Ground zero is the point on the ground directly above or below the point of detonation. The overpressure, which is the pressure in excess of the normal atmospheric pressure, and the winds are major contributors to the casualty and damage-producing effects of the nuclear detonation. The overpressure can cause immediate death or injury to personnel and damage to material by its crushing effect. The high-speed winds propel objects, such as tree limbs or debris, at great speed and turn them into potentially lethal missiles. These winds can also physically throw personnel who are not protected, resulting in casualties. People both inside and outside of a structure may be injured as a result of blast damage to that structure; those inside by the collapse of the structure and by fire; and those outside by the flying objects carried by the winds (figure 29-2).

b. Thermal Radiation:

(1) Heat. Within less than a millionth of a second of the detonation of a nuclear weapon, the extremely hot weapon residues radiate great amounts of energy. This leads to the formation of a hot and highly luminous, spherical mass of air and gaseous residue which is the fireball. The heat radiated from the fireball contributes to the overall damage caused by a nuclear burst by igniting combustibles and thus starting fires in buildings and forests. These fires may spread rapidly among the debris produced by the blast. In addition, this intense heat can burn exposed personnel at great distances from ground zero where the effects of blast and initial nuclear radiation become insignificant. The degree of injury from thermal radiation becomes more marked with the increasing size of the weapon. The degree of injury from thermal radiation is also affected by weather and terrain. During periods of limited visibility, the heat effect will be reduced significantly. Additionally, since ther-

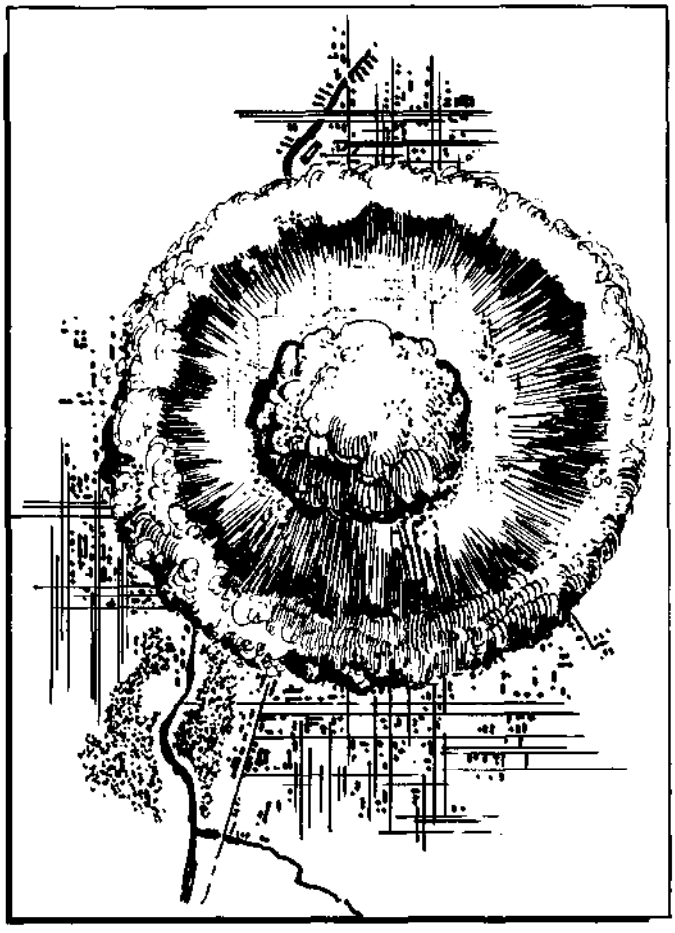

Figure 29-2. Blast.

mal radiation is primarily a line-of-sight phenomena, terrain masking can help reduce its effects (figure 29-3).

(2) Light. The fireball formed at the instant of a nuclear detonation is a source of extremely bright light. To an observer, 135 miles away from the explosion, the fireball of a 1-megaton weapon would appear to be many times more brilliant than the Sun at noon. The surface temperatures of the fireball, upon which the brightness depends, do not vary greatly with the size of the weapon. Consequently, the brightness of the fireball is roughly the same, regardless of the weapon yield. This light can cause injuries to personnel in the form of temporary or permanent blindness. Temporary blindness from a burst during daylight should be of very short duration and is not an important consideration for anyone other than aircrew members. At night, this loss of vision will last for longer periods because the eyes have been adapted to the dark. However, recovery should be complete within 15 minutes. The light flash can cause permanent injury to the eyes due to burns within the eye, but this is only likely to occur in personnel who happen to be looking directly at the fireball at the instant of explosion (figure 29-4).

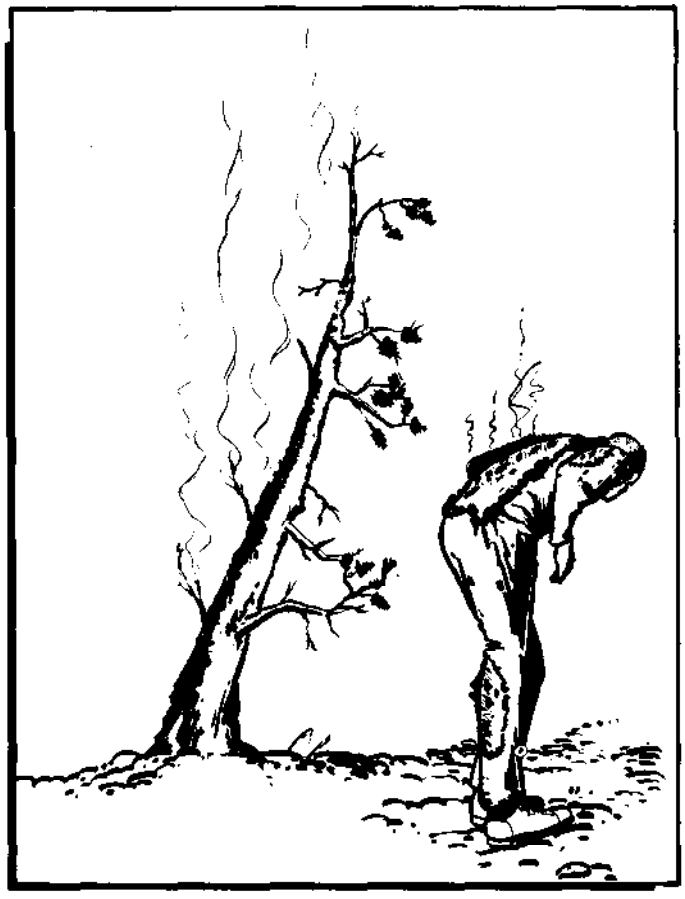

Figure 29-3. Thermal Radiation.

c. Nuclear Radiation:

(1) Initial nuclear radiation is the radiation emitted in the first minute after detonation. For practical pur-

Figure 29-4. Light.

poses, it consists primarily of neutrons and gamma rays. Both of these types of radiation, although different in character, can travel considerable distances through the air and can produce harmful effects in humans. Gamma rays are invisible rays similar to X rays. These penetrating rays interact with the human body and cause damage to tissues and the blood-forming cells. The effects of neutrons on the body resemble those of gamma rays. They are highly penetrating and are easily absorbed by human tissue. Neutron radiation can penetrate several inches of tissue. The neutron radiation produces extensive tissue damage within the body. The major problem in protecting against the effects of initial radiation is that a person may have received a lethal or incapacitating dose of radiation before taking any protective action (figure 29-5).

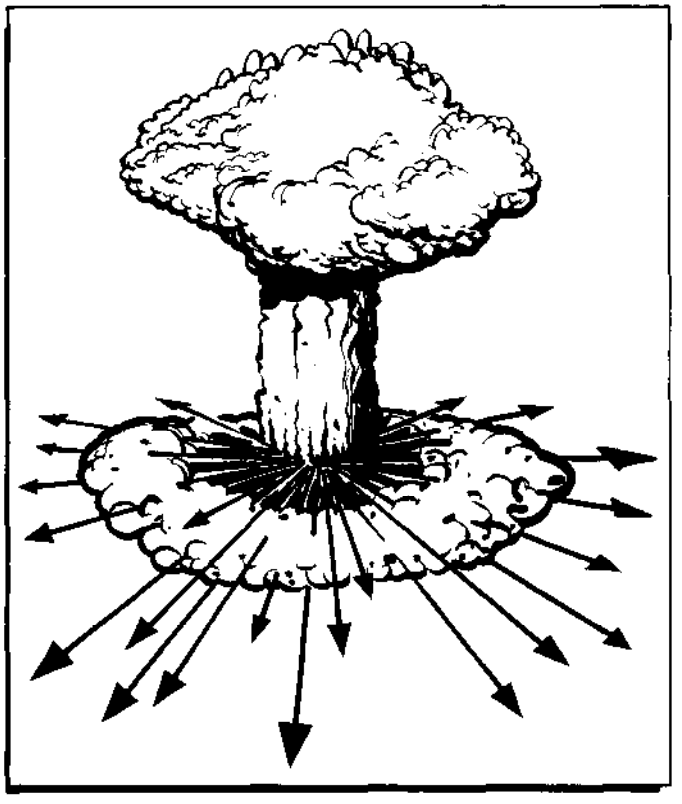

Figure 29-5. Nuclear Radiation.

(2) Residual nuclear radiation is that which lasts after the first minute and consists primarily of fallout and neutron-induced radiation.

(a) The primary hazard of residual radiation results from the creation of fallout. Fallout is produced when material from the Earth is drawn into the fireball, vaporized, combined with radioactive material, and condensed to particles which then fall back to the Earth. The larger particles fall back immediately in the vicinity of ground zero. The smaller particles are carried by the winds until they gradually settle on the Earth's surface. The contaminated areas created by fallout may be very small or may extend over many thousands of square miles. The dose rate may vary from an insignificant level to an extremely dangerous one for all personnel not taking protective measures.

(b) A secondary hazard which may arise is the neutron-induced radioactivity on the Earth's surface in the immediate vicinity of ground zero. The intensity and extent of the induced radiation field depend on the type of soil in the area around ground zero, the height of the burst, and the type and yield of the weapon. The only significant source of residual radiation from an airburst weapon is induced activity in the soil of a limited circular pattern directly beneath the point of burst (figure 29-6).

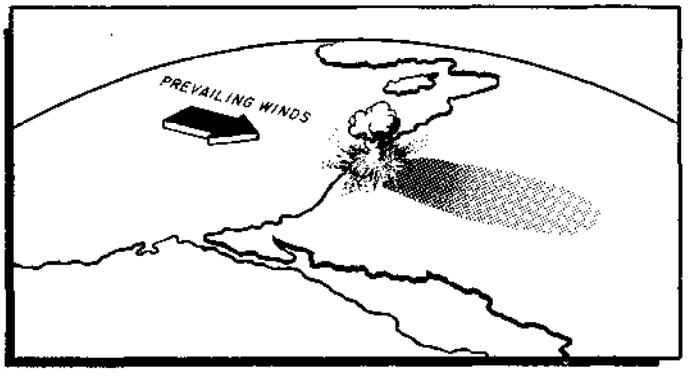

Figure 29-6. Residual Radiation.

29-3. Types of Nuclear Bursts. Nuclear bursts may be classified into three types according to the height of burst—airbursts, surface bursts, and subsurface bursts (figure 29-7).

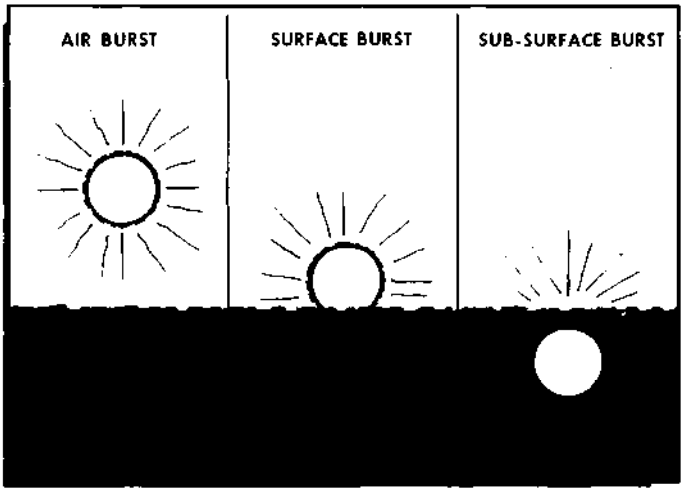

Figure 29-7. Types of Blasts.

a. Airburst. The detonation of a nuclear weapon at such a height that the fireball does not touch the surface of the Earth is called an airburst. Blast, thermal radiation, and initial radiation effects are increased in a low airburst. Fallout of radioactive material from an airburst is not of survival significance unless rain or snow falls through the radioactive cloud and brings the material to Earth. Neutrons from the detonation will cause induced radiation in the soil around ground zero. Except for very high airbursts, neutron-induced radiation in the area of ground zero will be of concern to survivors who are required to go into or across the area. Radiological monitoring will be required as units pass through such an area so that hazardous levels of radiation can be detected and avoided, if possible.

b. Surface Burst. The detonation of a nuclear weapon at such a height that the fireball touches the surface of the Earth or water is called a surface burst. Blast, thermal radiation, and initial nuclear radiation are not as widespread as from an airburst. Induced radiation is present but will be masked by residual radiation from fallout. The fallout produced by a surface burst is by far its most dangerous effect because the burst picks up a great deal more debris and radioactivates this debris; and, depending on the prevailing winds, the fallout covers thousands of square miles with high levels of radioactivity.

c. Subsurface Burst. The detonation of a nuclear weapon so that the center of the fireball is beneath the surface of the Earth or water is called a subsurface burst. If a fireball of this type breaks through the surface, fallout will be produced. Thermal radiation will not be a significant hazard since it will be almost completely absorbed by the soil. Blast effects will also be significantly reduced. Shock waves passing through the ground or water will extend for a limited distance. The range of the initial nuclear radiation will be considerably less than from either of the other two types of bursts because this will also be absorbed to a great extent by the soil. However, extremely hazardous residual radiation will occur in and around any crater. If the fireball does not break the surface, shock waves will pass through the ground and craters may result due to settling.

29-4. Injuries. The explosion of a nuclear bomb can cause three types of injures—blast, thermal radiation, and nuclear radiation. Many survivors receive a combination of two or all three of the above injuries. For example, an unprotected person could be killed by a piece of debris, could be burned to death, or could be killed by initial nuclear radiation if the person is within a few thousand yards from the center of the blast (figure 29-8).

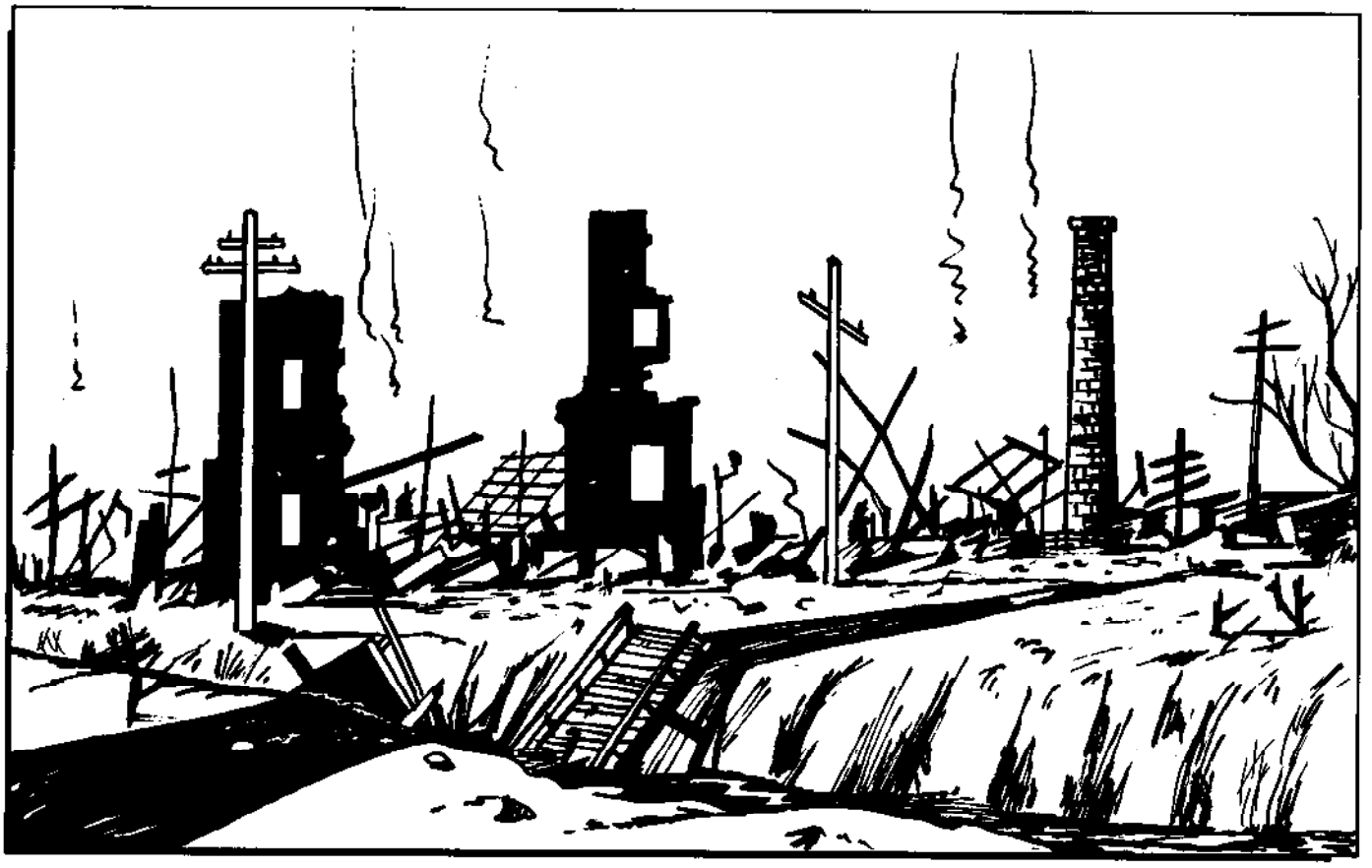

Figure 29-8. Injuries.

a. Blast Injuries. Direct blast can cause damage to lungs, stomach, intestines, and eardrums, or can cause internal hemorrhaging. However, the direct blast is not considered a primary cause of injury because those close enough to suffer serious injury from the direct blast will probably die as a result of initial thermal radiation, or they will be crushed to death. The greatest number of blast injuries are received as an indirect result of the blast from falling buildings, flying objects, and shattered glass.

b. Thermal Radiation Injuries. Burns are classified in degrees according to the depth to which the tissues are injured. In first-degree burns, the skin is reddened as in sunburn. In second-degree burns, the skin is blistered as from contact with boiling water or hot metal. In third-degree burns, the skin is destroyed or charred and the injury extends through the outer skin to deeper tissues. The degree of burn received from thermal radiation depends upon weather conditions, distance from the explosion, and available protection. Many thermal casualties are compounded by nuclear radiation and indirect blast injury. This makes it difficult to attribute casualties to thermal radiation alone.

c. Nuclear Radiation Injuries. The injurious effects of nuclear radiation from a nuclear explosion represent a new threat which is completely absent in conventional explosions. This does not infer that this source of injury is the most important in a nuclear explosion. Rays from radioactive material are not as great a hazard as people fear. The amount of danger from fallout depends upon where and how the nuclear bomb explodes and how well the person is protected. The greatest danger from residual radiation (fallout) comes from exposure for a long period of time to radioactive particles which are nearby, or from dust settling on the body or clothing. Since fallout (like X rays) can destroy living tissue, particularly in the blood-forming system, the exposure of persons working in a radioactive or "hot" area must be controlled so as not to exceed a safe limit. Although a person can become seriously ill and even die from breathing radioactive dust, there is less danger from this than when the whole body is exposed to fallout. Remember, all types of radiation are dangerous (nuclear, thermal, X ray, or even that from an infrared lamp).

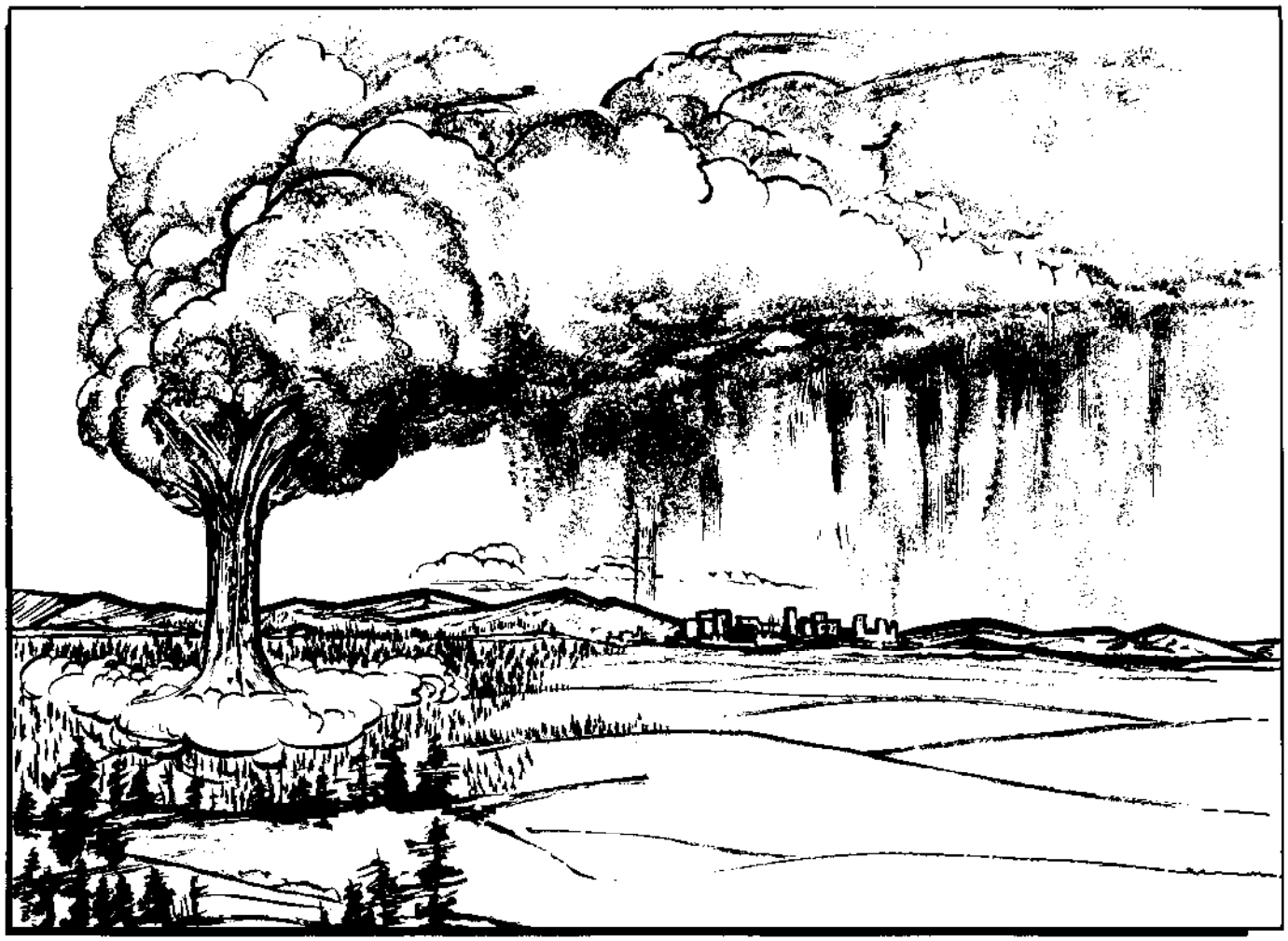

Figure 29-9. Fallout.

29-5. Types of Residual Radiation. The radioactive debris deposited on the surface as fallout contains three types of nuclear radiation—alpha particle emission, beta particle emission, and gamma radiation (figure 29-9).

a. Alpha Emitters. Alpha particles have low-pentrating abilities; therefore, survivors can easily shield themselves against these particles. Although alpha particles emitting (being given off) from radioactive elements will not penetrate the skin, alpha emitters present a serious hazard if ingested, inhaled, or allowed to enter the body by any other means. From a survival standpoint, alpha particles present the least danger of the three radiation hazards.

b. Beta Emitters. Basically, beta particles (radiation) are high-speed electrons which are only slightly penetrating; therefore, survivors can easily shield themselves against beta radiation by wearing materials of moderate thickness such as heavy shoes and clothing. Because serious skin burns may result from the direct contact of beta-emitting materials with the skin, survivors should take special care to brush themselves off, wash any previously unprotected areas of the body, or cover any exposed parts of the body. Since beta radiation is rapidly absorded by the air, distance will provide a good form of protection; in fact, 6 to 7 feet of atmosphere will stop most of the beta radiation resulting from fallout. In addition to presenting an external hazard, beta radiation will also cause serious internal effects. By using care in decontaminating foods and water and practicing good hygiene, survivors can greatly decrease the seriousness of this hazard.

c. Gamma Radiation. In contrast to either alpha or beta emitters, gamma radiation is highly penetrating. Gamma rays are similar to light rays and X rays, but are composed of shorter wavelengths and contain greater amounts of energy. Because of their penetrating abilities, they are the most hazardous of all types of external radiation.

(1) Fortunately, for the survivor faced with the important problem of obtaining immediate protection against gamma radiation, the amount of shielding is not as great as that required for the initial gamma radiation emitted during the fireball stage of a nuclear detonation. In addition to shielding, other methods of protection against penetrating gamma radiation, later addressed in more detail, include using the factors of time and distance. Because of the low absorption coefficient of gamma rays, their relative internal hazard is greatly reduced and, in this respect, they are far less dangerous than alpha and beta radiation. However, caution should be taken to ensure no radiation is absorbed—internally or externally.

(2) Since the fallout dust or debris contains particles emitting gamma rays, survivors must be especially careful to decontaminate themselves and their shelter area. Though radioactive fallout sometimes has the appearance of white ash or dust, it usually cannot be detected by the human senses. Survivors must assume any suspicious film of dust on water or plant life is radioactive, and they should apply decontamination procedures.

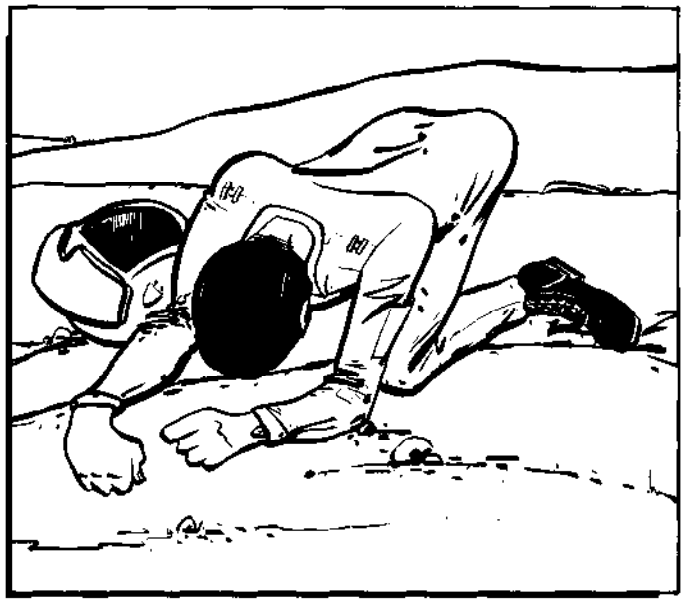

Figure 29-10. Effects.

29-6. Effects of Fallout on the Individual. The most harmful effects of fallout result from the changing of the blood cells. Because of this change in cells, some of the tissues which are essential to normal functioning of the human body are damaged or destroyed. The cells are unable to rebuild, so normal cell replacement in the organism is stopped. In addition, the formed products act as poisons to the remaining cells. The extent of damage to the body cells depends upon the dose received. Therefore, if the body receives a large dose of fallout radiation (gamma rays and possibly some beta particles), so many cells can be affected that survival is unlikely due to infection resulting from the loss of white blood cells (figure 29-10).

a. An overdose of radiation from fallout could be received if the survivor stays in the open and doesn't seek shelter. It is also possible to receive an overexposure from the fallout which settles on the clothing or body. Clothing does not stop gamma rays from penetrating and seriously injuring body tissues. Overexposure can also occur from remaining in a highly radioactive area too long. A survivor should not leave a shelter area unless absolutely necessary.

b. The first indication of an overdose of fallout radiation probably would not show up for several hours or possibly days. The survivor would then most likely become ill and begin to vomit. The time elapsing before the illness would depend on how large a dose was re-

EXPOSURE TYPE	GENERAL SYMPTOMS	ONSET OF SYMPTOMS	ABLE TO WORK	MEDICAL CARE	DEATHS	REMARKS
50-200r	Nausea, weakness, fatigue, possible vomiting, possible radiation burns	Within 24 hours	Yes	No	5%	
200-450r	Nausea, weakness, fatigue, vomiting, diarrhea, loss of hair, radiation burns, easy to bleed	3-6 hours	No	Yes	Less than 50%	Sick a few days, 1-2 weeks temporary recovery, then serious symptoms
450-600r	Same as 200-450r, plus hemorrhaging and infection	Within 3 hours	No	Yes	More than 50%	1 week temporary recovery, die without medical aid
600r+	Same as 450-600r, plus severe bloody diarrhea	Within 1 hour	No	Yes	100%	Death within 14 days
2000r+	Complete incapacitation	Within minutes	No	Yes	100%	Death within hours
Internal Radiation	Varies in proportion to internal dose	Varies	Varies	Varies	Varies	
Mild skin Injury	Itching and/or redness of skin, possible hair loss	1-2 days	Yes	No	None	
Severe Injury	Same as mild, plus radiation burns, & hair loss	1-2 days	Yes	Yes	Varies	Death depends on burn area

Figure 29-11. Radiation Sickness Symptoms.

ceived. When vomiting starts, it does not necessarily mean death will follow. For a few days, a survivor might feel far below par, but with proper medical care, complete recovery is possible (figure 29-11).

29-7. Radiation Detection. Since radiation is invisible and cannot be detected by the physical senses, detection instruments are used. This equipment includes devices for measuring the amount and intensity of area contamination and devices for determining the radiation dosage a survivor has received while in a contaminated area. A single radiation measurement usually has limited operational significance, except to personnel in the immediate area, since it gives information at the point of the reading only. However, a number of individual measurements considered together can give a picture of the radiation pattern over an entire area. A number of readings made at the same point over a period of time are required to determine the rate of decay of the fallout. Several different points for taking readings may be required in varying types of terrain.

a. Measurement of Radiation. Instruments developed for the detection and measurement of radiation are called radiac instruments. Radiac instruments measure the absorption of radiation, either in terms of dose-rate or dosage (figure 29-12).

(1) Dosage is the term applied to the total or accumulated amount of ionizing radiation (beta or gamma) received regardless of the time involved. Dosage is measured in terms of roentgens (r) or milliroentgens (mr). One milliroentgen is .001 of a roentgen. The total quantity of ionization received during a single radiation experience is called a dose. The radiation dose is also referred to as an exposure dose.

(2) When it is desirable to know how fast a certain dosage is being received, the term dose-rate is used to

Figure 29-12. Detection.

indicate this rate. Dose-rate is the amount of radiation being received per unit of time. Generally, the dose-rate is expressed in r/hr or mr/hr. The dose-rate is also used to indicate the level of radioactivity in a contaminated area (figure 29-13).

b. Radioactive Decay. The concept of radioactive decay is of vital importance in obtaining protection against nuclear fallout and in determining survival procedures. The debris from a nuclear explosion is made up of a mixture of radioactive materials of many kinds: unfissioned particles, fission products, and numerous other radioisotopes created by the neutron activation of inert material that takes place during the explosion. Fortunately, of the nearly 200 isotopes emerging, 70 percent are short-lived materials with a half-life of less than 1 day. Some elements, however, take a much longer time to decay; these are known as long half-life elements.

(1) Half-life is the time required for a radioactive substance to lose 50 percent of its activity through decay. Half-life for a mixture of isotopes (a term used to define precise species of elements) is not as simple as that described for a single radioactive isotope. The activity of the mixture diminishes very quickly after detonation, but as time passes, the longer-lived species become responsible for the major part of the radiation remaining, so that total radioactivity diminishes much more slowly.

(2) As a rule of thumb, radioactivity may be said to decrease in intensity by a factor of 10 for every sevenfold increase in time following the peak radiation level. Figure 29-13 illustrates the rapidity of the decay of radiation from fallout during the first 2 days after the nuclear explosion which produced it. Notice that it takes about seven times as long for the dose-rate to decay from 1000 roentgens per hour (1000 r/hr) to 10 r/hr (48 hours) as to decay from 1000 r/hr to 100 r/hr (7 hours). Only in high fallout areas would the dose-rate 1 hour after the explosion be as high as 1000 roentgens per hour.

(3) If the dose-rate 1 hour after an explosion is 1000 r/hr, it would take about 2 weeks for the dose-rate to be reduced to 1 r/hr solely as a result of radioactive decay. Weathering effects will reduce the dose-rate further; for example, rain can wash fallout particles from plants and buildings to lower positions on or closer to ground. Surrounding objects would reduce the radiation dose from these low-lying particles. Figure 29-13 also illustrates the fact that at a typical location where a given amount of fallout from an explosion is deposited later than 1 hour after the explosion, the highest dose-rate and the total dose received at that location are less than at a location where the same amount of fallout is deposited 1 hour after the explosion. The longer fallout particles have been airborne before reaching the ground, the less dangerous is their radiation.

(4) Two weeks after the last burst, the occupants of most shelters could begin working outside the shelters, increasing the number of hours each day. Exceptions would be in thermal damaged areas or in areas of extremely heavy fallout such as might occur downwind from important targets attacked with many weapons, especially missile sites and very large cities. To know when to come out safely, occupants would either need a reliable fallout meter to measure the changing radiation dangers, or they would need to receive information based on measurements made nearby with a reliable instrument, using the information in figure 29-14.

29-8. Body Reactions to Radiation. The effects of ionizing radiation upon the human body may be divided into two broad categories—chronic effects and acute effects.

a. Chronic Effects. Chronic effects are defined as those occurring some years or generations after exposure to radiation. Included in this category are the cancer-producing and genetic effects. While of concern because of their possible damage to future generations, these effects are of minor significance insofar as they may affect immediate survival.

b. Acute Effects. Acute effects are of primary significance in survival. Some acute effects appear within a few hours after exposure to radiation. These effects are the result of direct physical damage to tissue caused by

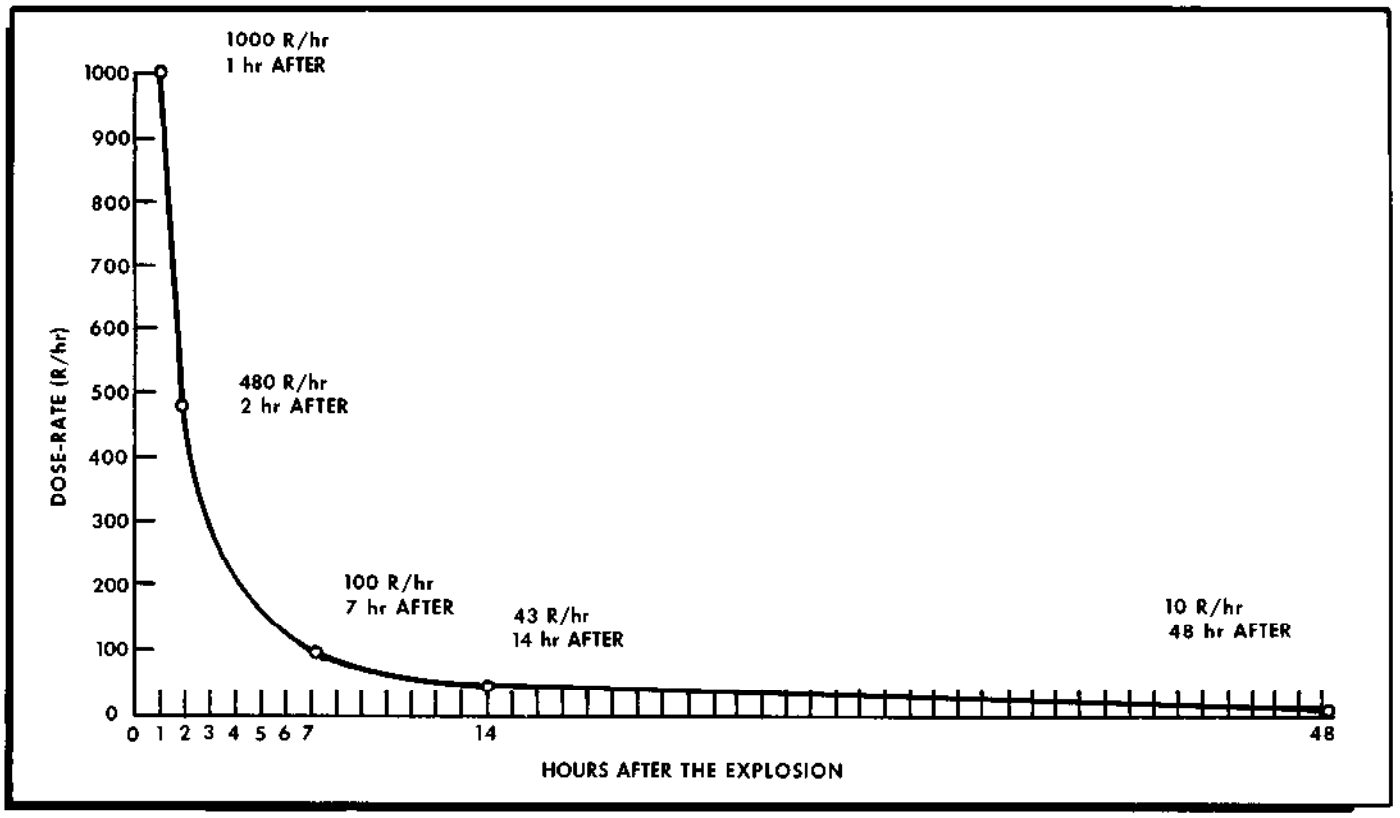

Figure 29-13. Decay Ratio.

radiation exposure; beta burns are examples of acute effects.

c. Body Damage. The extent of body damage depends to a large degree on the part and extent of the body exposed and the ability of the body to recover from radiation damage. Certain parts of the body, such as the brain and the kidneys, have little ability to recover from injury, while other portions of the body, such as the skin and bone marrow, can repair damage. The extent of body exposure to radiation is of great importance in determining the chances of subsequent recovery. If a dose of 350 roentgens was applied to just a small portion of the body, such as the hands or face, there would probably be little effect on overall health. Serious damage would, of course, be created in these exposed parts.

d. Hazards from Residual Radiation. There are two main hazards resulting from residue hazard, resulting from highly penetrating gamma radiation and less penetrating beta radiation which causes burns; and an internal hazard, resulting from the entry of alpha- and beta-emitting particles into the body.

(1) The external hazards result in overall irradiation and serious beta burns, while the internal hazards result in irradiation of the critical organs such as the gastrointestinal tract, thyroid gland, and bone.

(2) A very small amount of radioactive material can cause extensive damage in these and other parts of the body if allowed to enter the body by consumption of contaminated food or water or by absorption by the bloodstream through cuts or abrasions in the skin. By comparison, the material gaining entry by breathing presents a lesser hazard. Fortunately, most of the fallout particles are filtered out by the upper respiratory tract as seen in the observations of victims on the Marshall Islands who were accidentally exposed to fallout during the 1954 nuclear test. By using good hygiene and careful decontamination of food and water, a survivor can further reduce the internal hazard.

e. Degree of Radiation Damage. Because of the sensitivity of the gastrointestinal tract to radiation, the severity and the speed with which vomiting and diarrhea appear after exposure are good indicators of the degree of radiation damage. Almost everyone confined to immediate action shelters after a nuclear attack would be under stress and without clean surroundings. Many would also lack adequate water and food. Under these conditions, perhaps half the survivors who receive a whole-body dose of 200-450 roentgens would die within a few days. The human body can repair most radiation damage if the daily radiation doses are not too large.

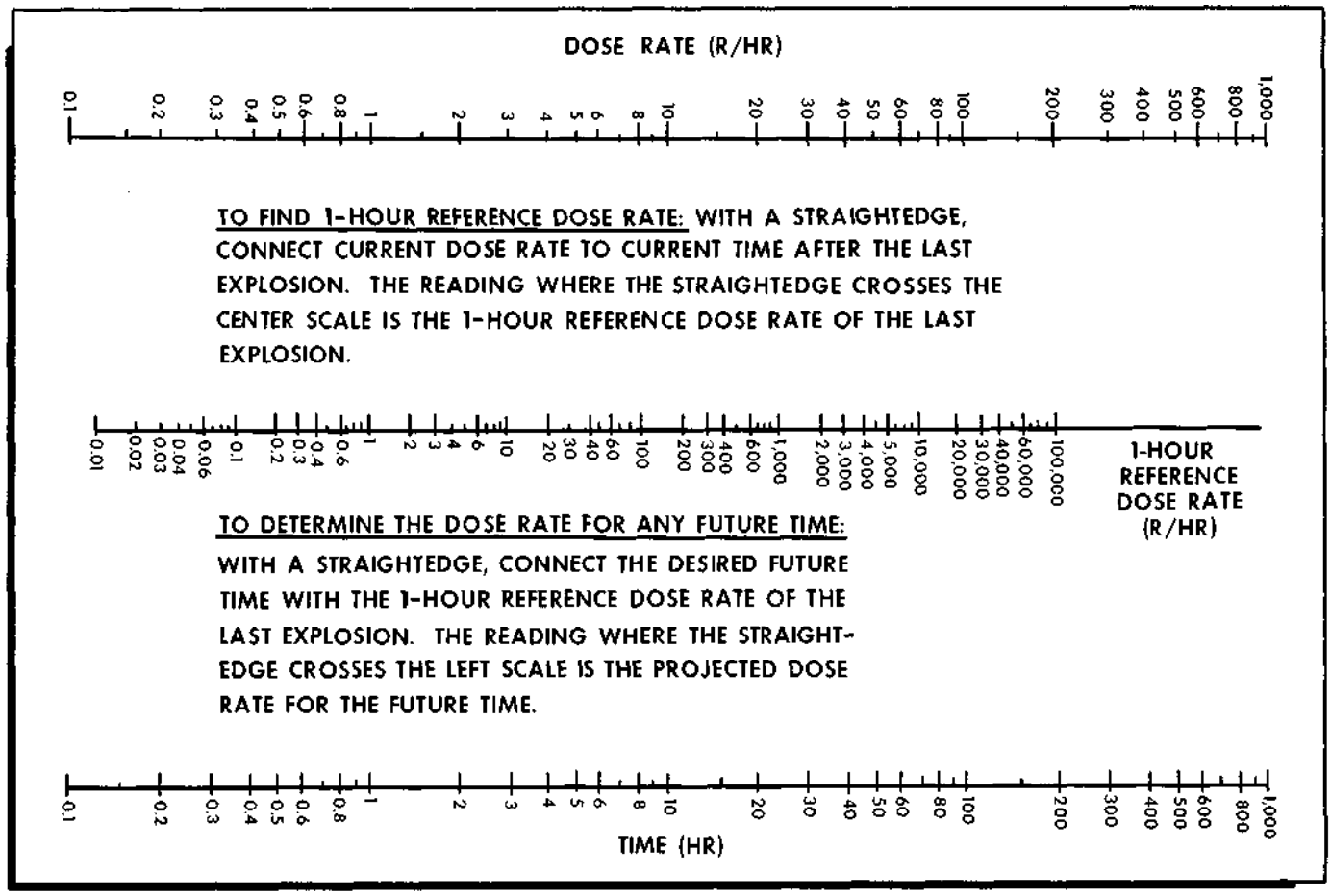

Figure 29-14. Dose Rate.

29-9. Countermeasures Against Penetrating External Radiation:

a. The means of protection from external radiation are threefold—time, distance, and shielding. By controlling the length of exposure time, by controlling the distance between the individual and the source of radiation, and most important of all, by placing some absorbing or shielding material between the survivor and the source of radiation, the survivor's level of radiation can be significantly reduced and thereby increase the chances of survival.

(1) Time. The effect of time on radiation exposure is easy to understand. Take a simple example such as the following: Assume a person is in an area where the radiation level from penetrating external radiation is 100 roentgens per hour and the dose-rate is constant; then in 1 hour, that person would receive 100 roentgens (or more properly, roentgen absorbed dose (rad)); in 2 hours, 200 rad; and in 8 hours, 800 rad. The important implication for a survivor is that, since radiation dosages received should be considered to be essentially cumulative, as little time as possible should be spent in an unprotected area, whether constructing a shelter or searching for water. Time is also of vital importance from another standpoint, that of radioactive decay. Knowledge of this characteristic can serve as one of the primary means of protection against radiation from fallout. The importance of time as a protective factor may be seen clearly by the following example. If survivors were to enter a high-intensity fallout area of 1000 roentgens per hour in which the radiation intensity had just peaked and remain there for 1 hour in the open, they would accumulate a biologically damaging dose of approximately 650 rad. This is a dose strong enough to kill 9 out of 10 people exposed to it. If, however, they entered this same area some 48 hours later and remained there for 1 hour, they would receive a dose of approximately 10 rad. Even if they were then required to remain in the open for 24 hours, their dose would only be 170 rad. Two weeks after the completed fallout, they would receive only one rad per hour in this same area. They could then remain in the area for several months, accumulating a dose of 180 rad, and not develop severe radiation sickness because of the rate at which the dosage was received (figure 29-15).

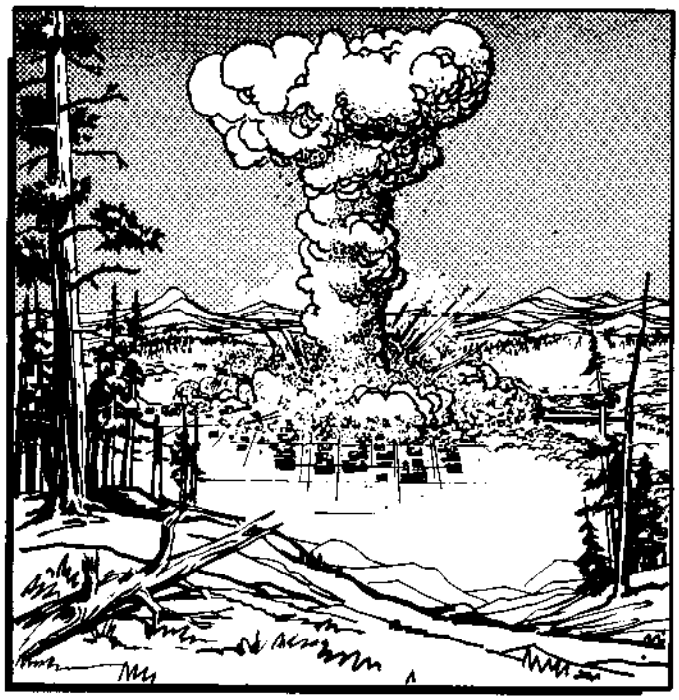

Figure 29-15. Countermeasures.

(2) Distance. Distance is effective as a means of protection against penetrating gamma radiation as shown by the inverse square law. This law states the intensity of radiation decreases by the square of the distance from the source. In other words, survivors exposed to 1000 units of penetrating external radiation at 1 foot from the source would receive only 250 units at 2 feet; when they double the distance, the radiation level is reduced by $(\frac{1}{2})^2$ or one-fourth the amount. When the distance is tripled, the dose is reduced to $(\frac{5}{8})^2$ or one-ninth of the original amount (or 111 units). At 10 feet, this is further reduced to $(\frac{1}{10})^2$ or one-hundreth of the radiation exposure at 1 foot. This relationship of distance to intensity of radiation exposure is dependent upon the distribution pattern of the radiation source. As just seen, if the radiation is concentrated into a very small area (referred to as a point source), the intensity is decreased to about one-fourth the original amount each time the distance is doubled. A complicated relationship, however, is obtained when the radiation source is not concentrated in a point, but is spread around in random patterns as in the case of surface contamination from fallout.

(3) Shielding. The third and most important method of protection against penetrating radiation is that of shielding. Since the damaging effects of penetrating radiation arise from the fact that the rays interact with electrons in the body, survivors must place dense material between themselves and the source of radiation. The more electrons there are in the makeup of the shielding material, the more gamma radiation will be stopped from entering the body. Lead, iron, concrete, and water are all examples of good shielding materials. Of the three described countermeasures against penetrating external radiation, shielding gives the greatest protection, is the easiest to use under survival conditions, and is, therefore, the most desirable. If shielding is not possible, however, the other precautions should be rigorously followed. The degree of protection they afford is significant and could provide the necessary margin of safety for survival (figure 29-16).

Figure 29-16. Shielding.

b. It is a common misconception that radioactive fallout can impart its radioactivity to the object with which it comes in contact, such as fruits and vegetables. The rays from radioactive fallout particles cannot make something radioactive merely by being in contact with it. For this reason, canned foods and smooth-skinned fruits and vegetables may be eaten once they have been decontaminated.

(1) Everyone entering a contaminated area should wear protective clothing (CW ensemble) to prevent exposure to radioactive dust. The main reason for this precaution is to shield the skin from beta particles. Beta particles won't penetrate far into the body, but they will harm the skin. The effect, a reddening and blistering of the skin, is called a "beta burn." Such damage may not appear until sometime after exposure since beta particles have a delayed action. Survivors may not know they have received skin burns from beta particles until it is too late. Therefore, protective clothing should be worn as instructed. Sleeves should be down and buttoned, wrist and ankle openings taped, and gloves and boots worn, if possible.

(2) Alpha particles are lowest in ability to penetrate, but they are most hazardous when taken into the body through contaminated food and drink. The best defense against alpha particles is to ensure all food is properly decontaminated before consumption.

(3) If at all possible, a survivor should always avoid "hot" areas. Sometimes, a survivor cannot avoid contact with this environment and, when entering a "hot" area, should keep radioactive dust off the body and out of the body's system. If survivors inhale radioactive dust, they can suffer serious internal injury. If there is radioactive dust in "hot" areas, a protective filter should be used over the mouth and nose.

c. The presence of fallout material in the area will require some minor modification of medical procedures. Wounds should be covered to prevent entry of active particles. Burns from beta activity are treated like any other burn, except the burned surface should be washed. Measures to prevent infection should be emphasized since the body will be especially sensitive to infections because of changes in the blood. Extra attention should be given to the prevention of colds and respiratory infections. The eyes should be covered to prevent entry of particles. Improvised goggles may be used for this purpose.

29-10. Shelter. A sufficient thickness of shielding material will reduce the level of radiation to negligible proportions. The thickness required to reduce gamma radiation from fallout is much less than that necessary against initial gamma radiation since fallout radiation has much less energy than the initial radiation from a nuclear explosion. Thus, a comparatively small amount of shielding material can provide good protection against residual gamma radiation.

a. The following table illustrates the thicknesses of various materials required to reduce the penetration from residual fallout gamma sources by 50 percent:

Iron and steel	.7 inches
Concrete	2.2 inches
Brick	2.0 inches
Dirt	3.3 inches
Ice	6.8 inches
Wood (soft)	8.8 inches
Snow	20.3 inches

b. The principle of half-value layer thickness is useful in understanding the absorption of gamma radiation by various materials. According to this principle, if 2 inches of brick reduce the level of gamma radiation by one-half, adding another 2 inches of brick will reduce the intensity by another half, namely to one-fourth the original amount; 6 inches will reduce the level of fallout gamma radiation to one-eighth its original amount; 8 inches to one-sixteenth; and so on. Thus in a shelter protected by 3 feet of earth, a radiation intensity of 1000 roentgens per hour on the outside would be reduced to about one-half roentgen per hour inside the shelter.

c. Areas where terrain provides natural shielding and easy shelter construction, such as the sides of ditches, ravines, rocky outcroppings, hills, and riverbanks, are ideal locations for an emergency shelter (figure 29-17). In level areas lacking natural protection, dig a foxhole or slit trench.

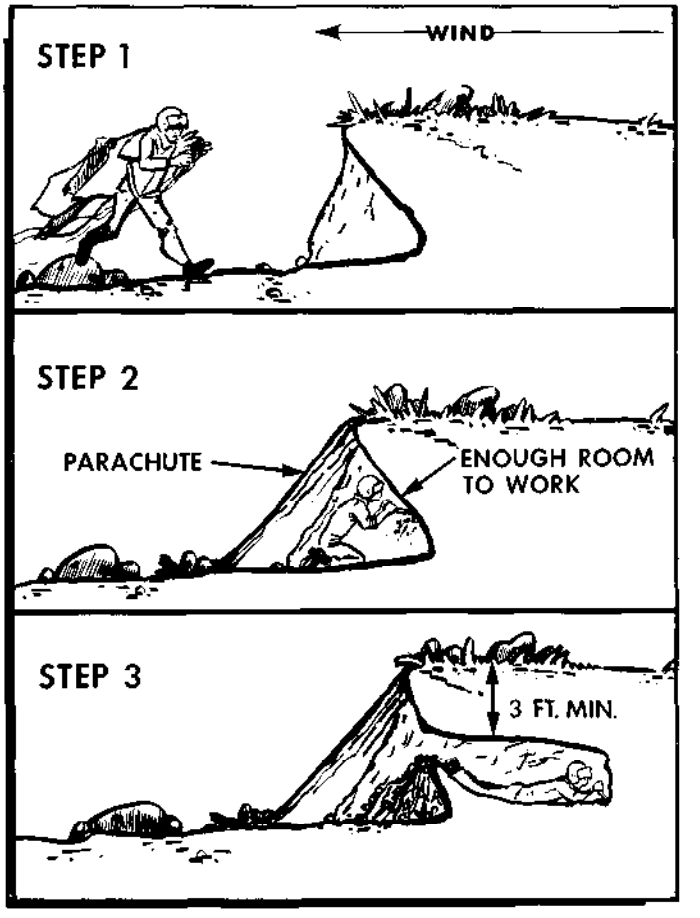

Figure 29-17. Immediate Shelter.

(1) In digging a trench, the survivor should work from the inside of the hole as soon as it is large enough to cover part of the body to prevent exposure of the entire body to radiation. In open country, an attempt should be made to dig the trench from a prone position, stacking the dirt carefully and evenly around the trench. On level ground, the dirt should be around the body for additional shielding. Depending upon soil conditions, the time to construct a shelter will vary from a few minutes to a few hours. Rapid shelter construction will limit the dosage received. Building a shelter under a tree is not recommended because cutting or digging out the numerous roots will be very difficult. Another disadvantage in making a shelter under trees is that more of the gamma rays from fallout particles on the leaves and branches would reach and penetrate the shelter than if

these same particles were on the ground. Many gamma rays from fallout particles on the ground would be scattered or absorbed by striking rocks, clods of earth, tree trunks, or buildings before reaching a below-ground shelter (figure 29-18).

(2) While an underground shelter covered by 3 or more feet of earth would provide the best protection against fallout radiation, the following additional unoccupied structures (in the order listed) offer the next best protection:

(a) Caves and tunnels covered by more than 3 feet of earth.

(b) Storm or storage cellars.

(c) Culverts.

(d) Basements or cellars of abandoned buildings.

(e) Abandoned buildings made of stone or mud.

d. Building a roof on the shelter should not be required. A roof should only be added if the materials are readily available to the survivor and will require only a brief exposure to the outside contamination. If the construction of a roof would require extended exposure to penetrating radiation, it would be wiser to leave the shelter roofless. The function of a roof is to lend distance from the source of fallout to the body, and unless dense roofing is used, a roof provides only scant shielding. A simple roof can be constructed out of a piece of parachute anchored down by dirt, rocks, or other refuse from the shelter. Large particles of dirt and other debris may be removed from the top of this parachute canopy by beating it at frequent intervals. Such a parachute cover will not offer shielding from the radiation emitted by active fission products deposited on the outer surface, but it will increase the distance from the fallout source and keep the shelter area covered from further contamination.

e. The primary criterion for locating and establishing a shelter site should be to obtain protection as rapidly as possible against the high intensity radiation levels of early gamma fallout. Five minutes to locate the shelter site is an excellent guide. Speed in obtaining shelter is essential. Without shelter, the dosage received in the first few hours will exceed that received during the rest of the week in a contaminated area; the dosage received in this first week will exceed that accumulated during the rest of a lifetime spent in the same contaminated area.

(1) Several initial actions should be kept in mind while seeking a shelter location. The survivor should:

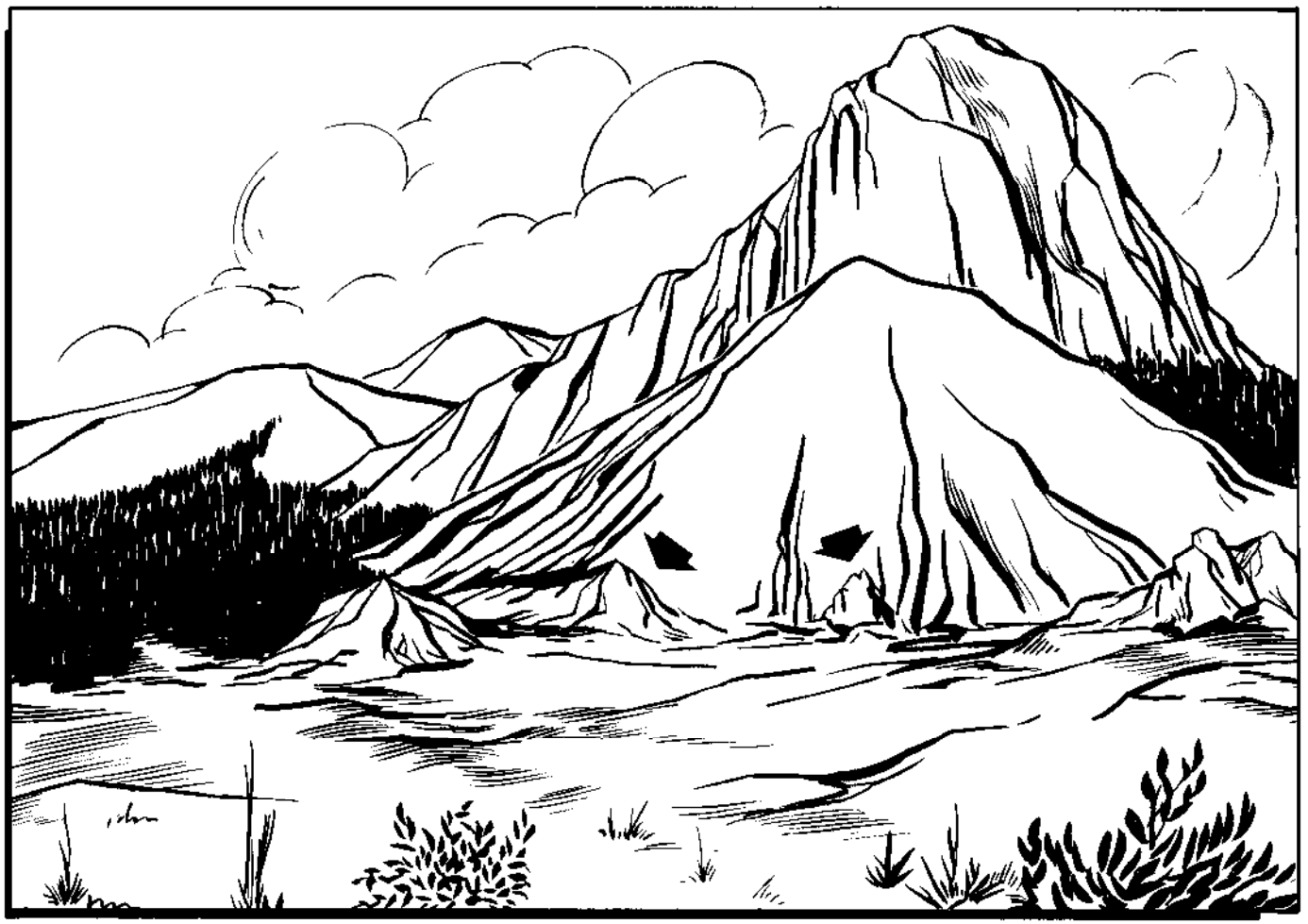

Figure 29-18. Terrain That Provides Natural Shielding.

(a) Where possible, attempt to remain with the aircraft until it is possible to eject or land in an area at least 20 miles upwind or crosswind from any known target. This action will ensure the landing area will have a minimum amount of fallout (figure 29-19).

(b) If it can be controlled, parachute opening should be delayed until a comparatively low altitude is reached in order to reduce the time of exposure to fallout during the descent.

-1. During the descent, select areas on the ground likely to provide good shelter.

-2. Immediately upon landing, take the parachute and survival kit and find cover and protective shelter.

-3. Take precautions to avoid detection and capture, but not at the expense of additional exposure to radiation which will lessen the chance of survival.

(2) In selecting and establishing the shelter, the survivor should keep the following additional factors in mind in order to reduce the time of exposure and the dosage received:

(a) Where possible, seek a crude existing shelter which can be improved. If none is available, dig a trench.

(b) Dig the shelter deep enough to obtain good protection, then enlarge it as required for comfort.

(c) Cover the top of the foxhole or trench with any readily available material and a thick layer of earth, if possible, without leaving the shelter.

(d) During the descent and while constructing a shelter, keep all clothing on, as well as a cap, scarf, and gloves to obtain protection against burns from beta radiation.

(3) The shelter location should be brushed or scraped clean of any surface deposit with a branch or some other object to be certain that contaminated materials are removed from the area to be occupied. The swept area should extend at least 5 feet beyond the area where the shelter is being dug. Any material brought into the shelter should be decontaminated. This includes grass or foliage that is being used as insulation or bedding material, outer clothing (especially footgear), and the parachute if it is to be used. If weather permits and the parachute and outer clothing are heavily contaminated, the survivor may want to remove them and bury them under a foot of earth at the end of the shelter. These may later (after decay factor) be retrieved when leaving the shelter. If these materials are dry, decontamination may be done by beating or shaking them outside the entrance to the shelter to remove the radioactive dust (figure 29-20). Any body of water, even though it may contain contaminated particles, may be used to rid materials of excess fallout particles by simply dipping

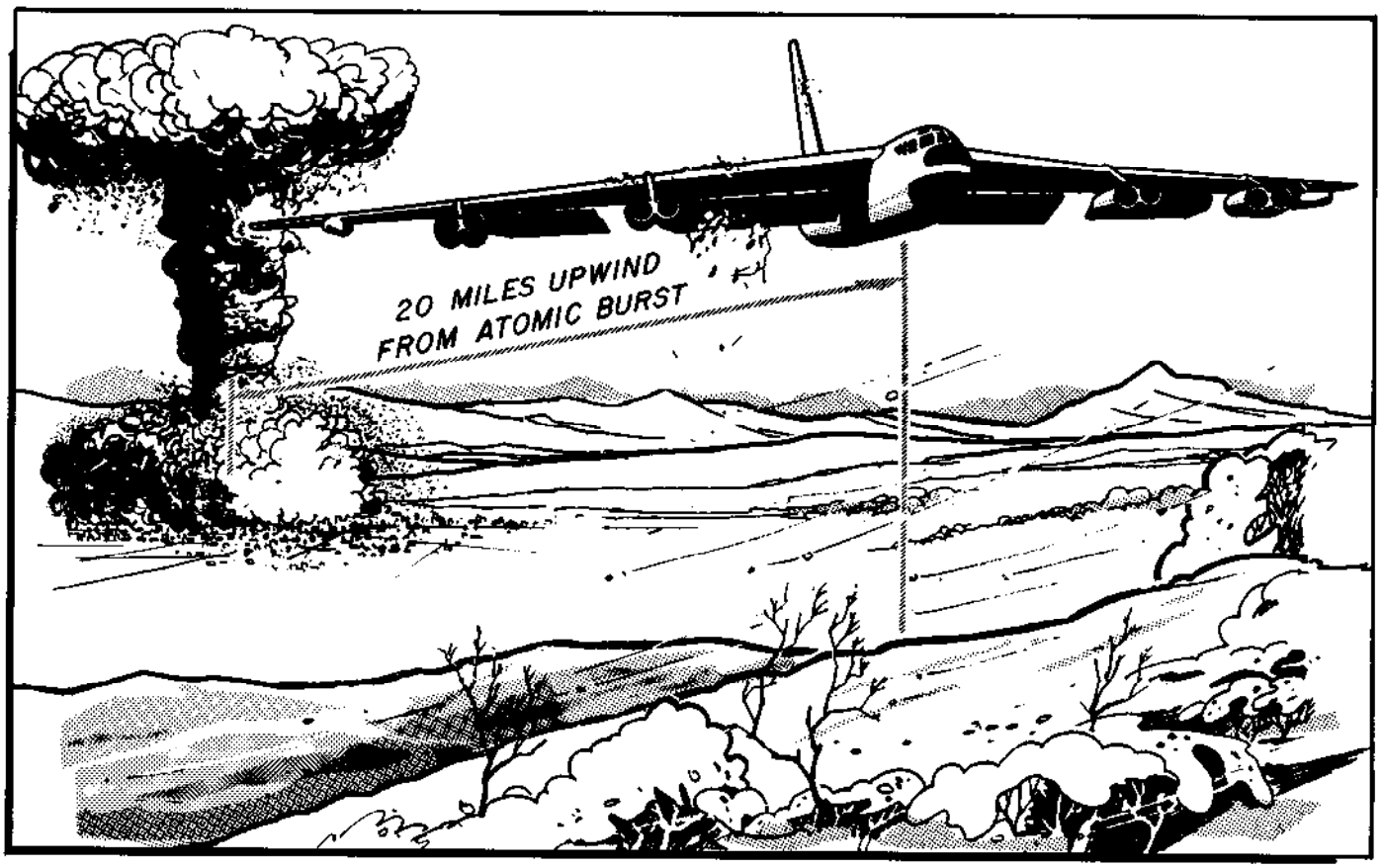

Figure 29-19. Route of Travel.

the materials into the water and shaking them to remove excess water. Do not wring out materials since this will trap the contaminated particles. If at all possible, wash the body thoroughly with soap and water without leaving the shelter. This washing will remove most of the harmful radioactive particles which are likely to cause beta burns or other damage. If water is not available, the face and any other exposed skin surfaces should be wiped to remove contaminated dust and dirt. This may be done with a clean piece of cloth or a handful of uncontaminated dirt obtained by scraping off the top few inches of soil and using the "clean" dirt.

f. Upon completion of the shelter, the survivor should lie down, keep warm, sleep, and rest as much as possible during the time spent in the shelter. There is no need to panic if nausea and symptoms of radiation sickness are experienced. Even small doses of radiation can cause these symptoms which may disappear in a short time. The following provides the time necessary to avoid serious dosage and still allow the opportunity to cope with survival problems:

(1) Complete isolation should be maintained from 4 to 6 days following delivery of the last weapon. A very brief exposure for procurement of water on the third day is permissible, but exposure should not exceed 30 minutes.

(2) On day 7, one exposure of not more than 30 minutes.

(3) On day 8, one exposure of not more than 1 hour.

(4) From day 9 through day 12, exposure of 2 to 4 hours per day.

(5) From day 13 on, normal operation, followed by rest in a protected shelter.

29-11. Water. In a fallout contaminated area, available water sources may be contaminated with radioactive particles. If possible, the survivor should wait at least 48 hours before drinking any water to allow radioactive decay to take place. Selecting the safest possible source of water will greatly diminish the danger of ingesting harmful amounts of radioactivity.

a. Although many factors, such as direction of wind, rainfall, and amount of particulate matter (clay, for example) in the water, will influence the choice in selecting water, the following priorities of water sources are

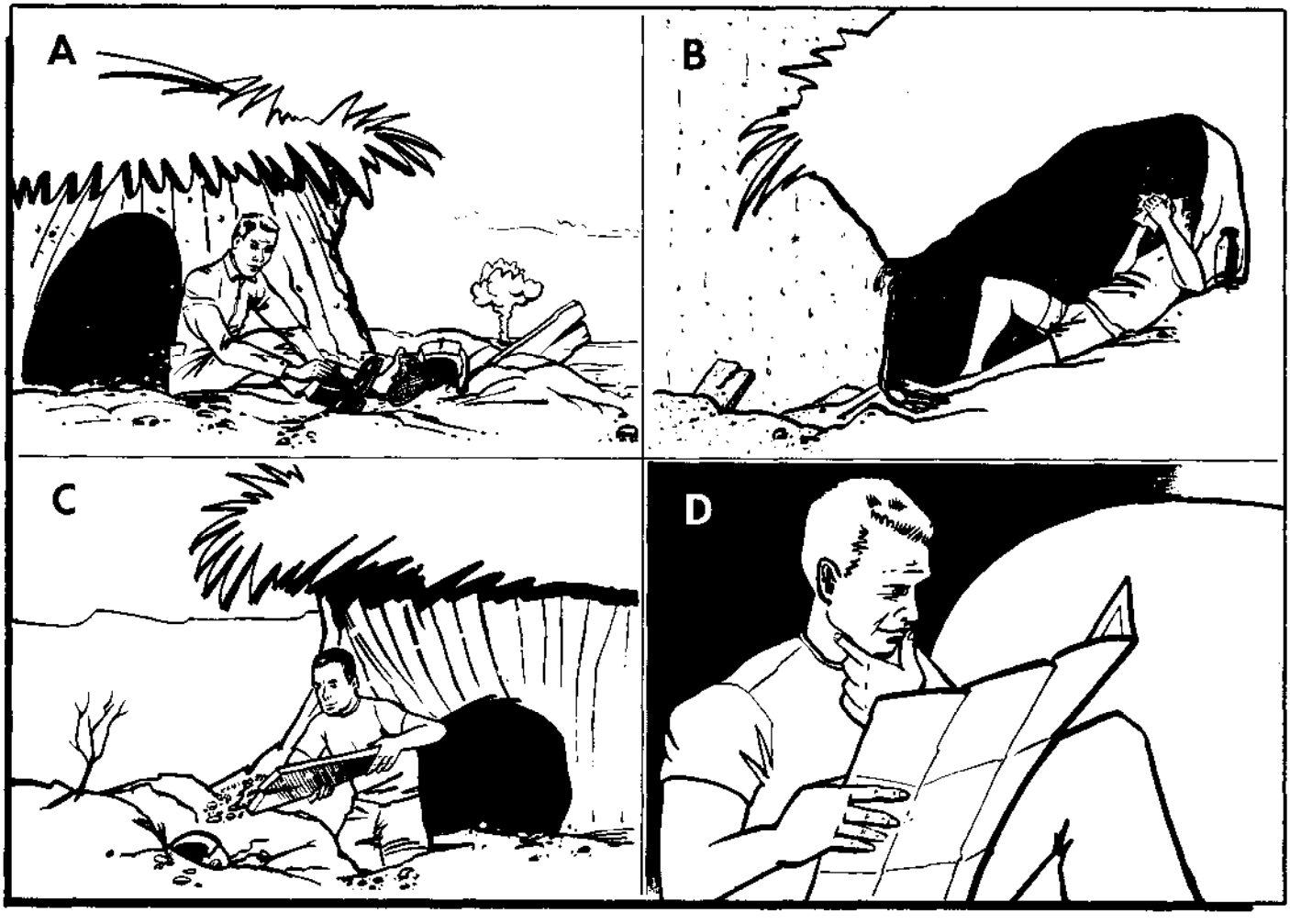

Figure 29-20. Decontamination.

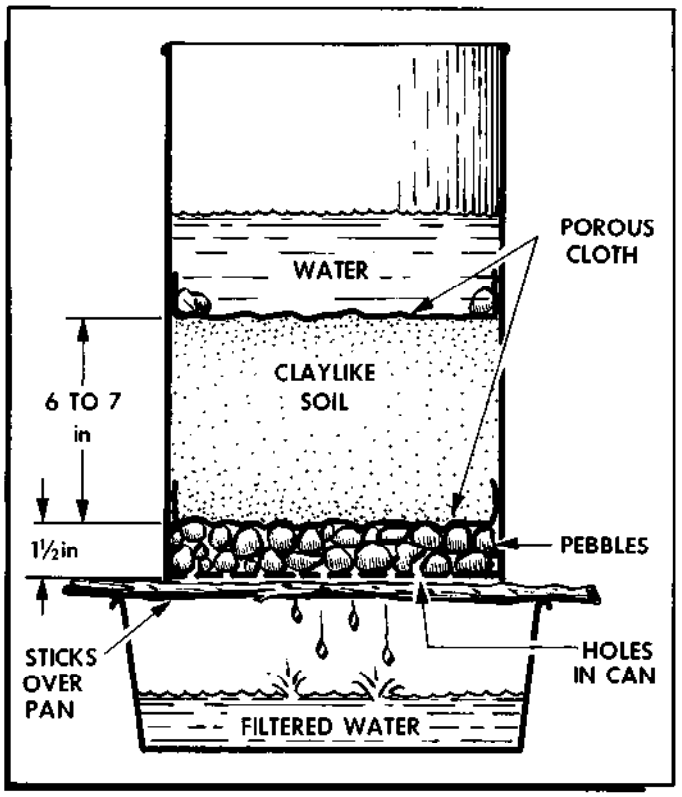

Figure 29-21. Filtering Water.

recommended: (As an additional precaution against disease, all water sources should be treated with water purification tablets from the survival kit or boiled for at least 10 minutes.)

(1) Water from springs, wells, or other underground sources having natural filtration will be the safest sources of water.

(2) Any water in the pipes or containers of abandoned houses or stores will also be free from radioactive particles and, therefore, safe to drink, although precautions will have to be taken against bacteria in the water.

(3) Snow taken from a level which was 6 or more inches below the surface during the fallout should be a safe source of water.

(4) Water from streams and rivers will be comparatively free from fallout within several days after the last nuclear explosion because of the dilution factor. If at all possible, such water should be filtered before drinking.

(5) Water from lakes, pools, ponds, and other standing sources is likely to be heavily contaminated, though most of the heavier, insoluble isotopes will settle to the bottom.

b. The degree of solubility of various isotopes varies. Some fission products are extremely water soluble, but most are relatively insoluble. Certain isotopes, in fact, have been found to be as much as 90 percent insoluble. The significance of this fact is that 99 percent of the radioactivity in water could be removed by filtering it through ordinary earth (figure 29-21). The best method of filtration is to dig sediment holes or seepage basins along the side of a water source. The water will seep laterally into the hole through the intervening soil which will act as a filtering agent and remove the contaminated fallout particles which have settled on the original body of water. It is important the hole be covered in some way (example, with a parachute) to prevent further contamination (figure 29-22).

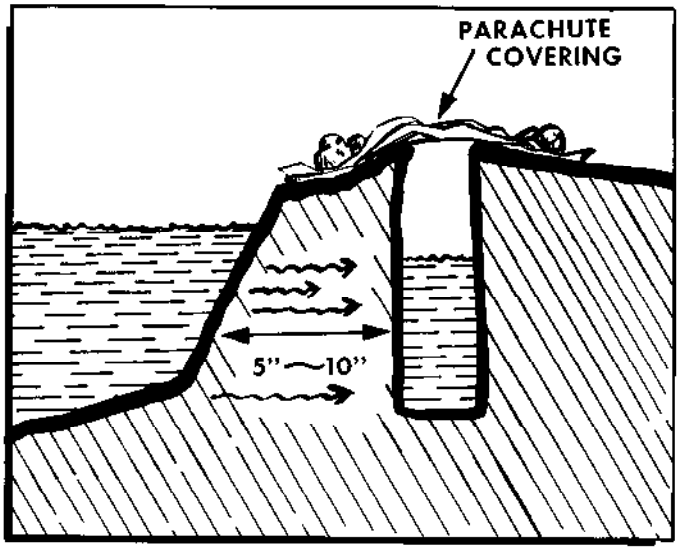

Figure 29-22. Settling Water.

c. Settling is one of the easiest methods used to remove most fallout particles from water. Furthermore, if the water to be used is muddy or murky, settling it before filtering will extend the life of the filter. The procedure is as follows:

(1) Fill a bucket or other deep container three-fourths full of the contaminated water.

(2) Dig dirt from a depth of 4 or more inches below the ground surface and stir it into the water. Use about a 1-inch depth of dirt for every 4 inches of water.

(3) The water is stirred until practically all of the dirt particles are suspended in the water.

(4) This mixture should be allowed to settle for at least 6 hours. The settling dirt particles will carry most of the suspended fallout particles to the bottom and cover them. The clear water can then be dipped or siphoned out and purified (figure 29-23).

29-12. Food. Obtaining edible food in a radiation contaminated area is possible. Survivors need to follow only a few special procedures in selecting and preparing rations and native foods for use. Since survival rations are protected by secure packaging, they will be perfectly acceptable for use after the ration containers are decontaminated, but survivors should supplement them with any food they can find on their trips away from the shelter. Any processed foods which may be found in abandoned buildings are acceptable for use once they are decontaminated. These include canned and pack-

Figure 29-23. Settling Pool.

aged foods after the containers or wrappers are discarded or washed free of fallout particles, food stored in any closed container, and food stored in protected areas (such as cellars). The containers should be washed before opening. For purposes of discussion, native food sources may be divided into two categories: plant food and animal food.

a. Animal Food. In the category of native animal food, survivors must assume all animals, regardless of their habitat or living conditions, will be exposed to radiation. Since the effects of radiation on animals are similar to those in humans, most of the wild animals living in a fallout area are likely to become sick or die from radiation during the first month following the nuclear explosion. Even though animals may not be completely free from radioactive materials, it may be necessary to use them as a food source. With careful preparation and adherence to several important principles, animals can be safe sources of food.

(1) If an animal appears to be sick, it should not be eaten. The animal may have developed a bacterial infection as a result of a radiation dose. Contaminated meat could cause severe illness or death if eaten, even if thoroughly cooked.

(2) All animals should be carefully skinned to prevent any radioactive particles adhering to the outside of the skin from gaining entry into the body.

(3) Meat around the bones and joints should not be eaten. A large percentage of radioactivity in the body of animals is found in the skeleton. The remaining muscle tissue of the animal, however, will be safe to eat. Before cooking the animal, survivors should cut the meat away from the bone, leaving approximately one-eighth of a inch of meat on the bone. All internal organs such as the heart, liver, and kidneys, normally used as survival food, should be discarded (figure 29-24) since they tend to concentrate beta and gamma radioactivity. All meat should be cooked until it is very well done. To be sure the meat is well done, it should be cut into pieces less

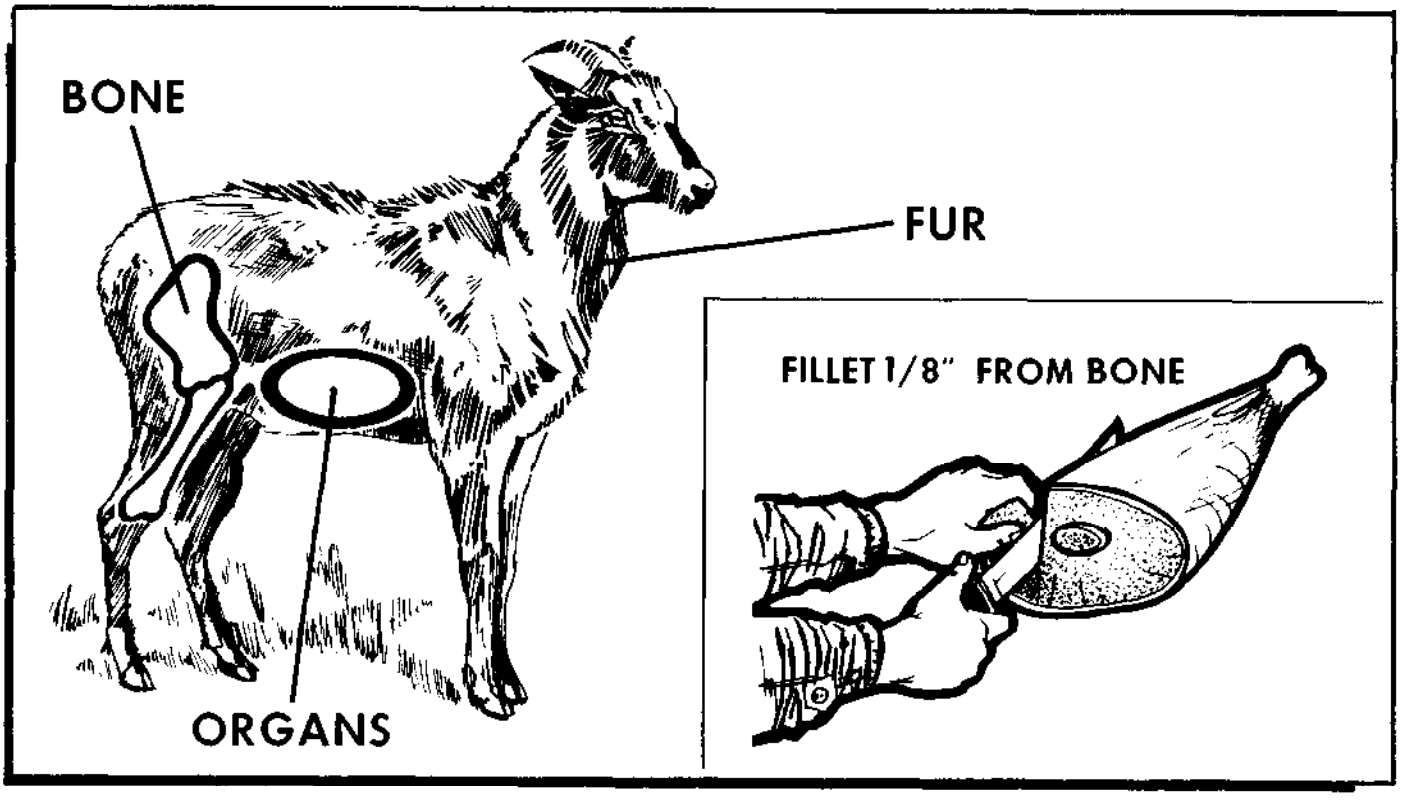

Figure 29-24. Meat Procurement.

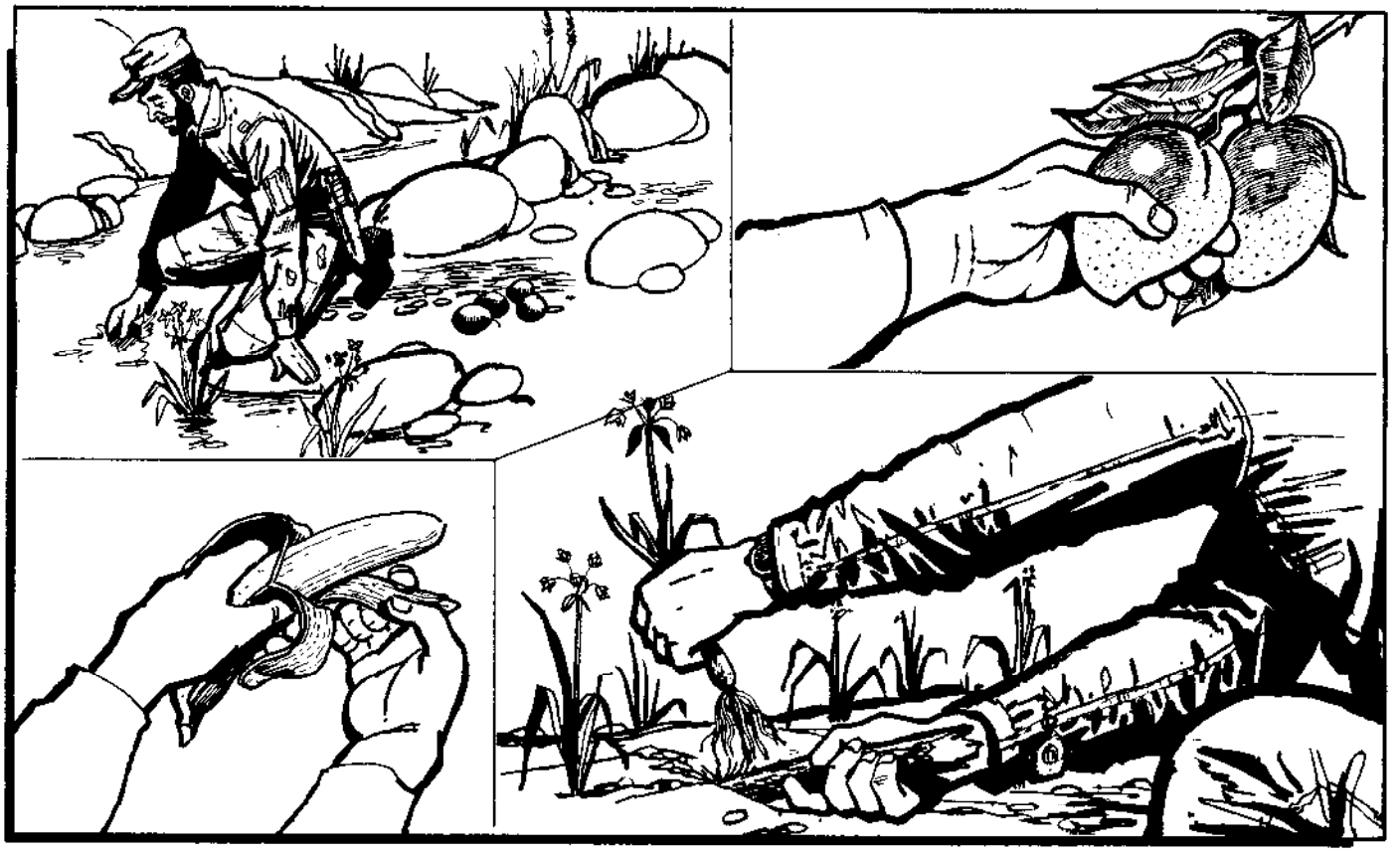

Figure 29-25. Collecting Food.

than one-half inch thick before cooking. This precaution also reduces cooking time and saves fuel.

(4) The extent of contamination of fish and clams following nuclear tests in the Pacific was found to be much greater than that of the land animals. On the basis of these findings and those of other tests showing the high concentration of radioactivity in aquatic plants and animals, especially near coastal areas, it is recommended that aquatic food sources be used only in an extreme emergency.

(5) All eggs (excluding the shell), even if laid during the period of fallout, will be safe to eat. Because animals absorb large amounts of radioactive strontium from the plants upon which they graze, milk from any animals in a fallout area should be avoided.

b. Plant Food. Plant foods are contaminated by the accumulation of fallout on their outer surfaces or by means of absorption through the roots.

(1) The survivor's first choice of plant food should be vegetables, such as potatoes, turnips, carrots, and other plants, whose edible portions grow underground. These are the safest to eat once they are scrubbed and the skins removed. Second in order of preference are those above-ground portions of the plant with edible parts which can be decontaminated by washing and peeling their outer surfaces; examples are bananas, apples, and other such fruits and vegetables. Other smooth-skinned vegetables, fruits, or above-ground plants, which cannot be easily peeled or effectively decontaminated by washing the radioactive particles off their surfaces, will be the third choice of emergency food (figure 29-25).

(2) The effectiveness of decontamination by scrubbing is generally inversely proportional to the roughness of the surface of the fruit. After the Marshall Islands incident, smooth-surfaced fruits were found to lose 90 percent of their radioactivity after washing, while washing rough-surfaced plants removed only 50 percent of the contamination. Plant foods, such as lettuce, having a very rough outer surface which cannot easily or effectively be decontaminated by peeling or washing, should be eaten only as a last resort. Other difficult foods to decontaminate by washing with water are dried fruits, such as figs, prunes, peaches, apricots, pears, and soybeans.

(3) Generally speaking, any plant food ready for harvest can be used for food if it can be effectively decontaminated. Growing plants, however, can absorb some radioactive materials through their leaves as well as from soil, especially if rains have occurred during or after the fallout period. With some elements, such as strontium, which is extremely soluble in water, data has shown greater amounts were taken up by plants through

EQUIVALENT RESIDUAL DOSE (ERD): The body can repair 90% of a dose of radiation damage. The unrepaired damage for any given day is the radiation dose multiplied by the ERD factor of the number of days after the exposure. NOTE: Each subsequent dose ERD Factor is based on the number of days after each subsequent dose.

DAY	FACTOR	DAY	FACTOR	DAY	FACTOR	DAY	FACTOR	DAY	FACTOR
1-4	1.000	10	.873	16	.764	22	.670	28	.590
5	.978	11	.854	17	.748	23	.656	29	.578
6	.956	12	.835	18	.731	24	.642	30	.566
7	.934	13	.817	19	.716	25	.629	31	.554
8	.913	14	.799	20	.700	26	.616	32	.543
9	.893	15	.781	21	.685	27	.602	33	.532

$$\begin{pmatrix}\text{1st}\\ \text{DOSE}\end{pmatrix}\begin{pmatrix}\text{ERD}\\ \text{FACTOR}\end{pmatrix} + \begin{pmatrix}\text{2d}\\ \text{DOSE}\end{pmatrix}\begin{pmatrix}\text{2d ERD}\\ \text{FACTOR}\end{pmatrix} + \ldots + \begin{pmatrix}\text{X}\\ \text{DOSE}\end{pmatrix}\begin{pmatrix}\text{X ERD}\\ \text{FACTOR}\end{pmatrix} + \begin{pmatrix}\text{SHELTER}\\ \text{DOSE}\end{pmatrix}$$

WATER SOURCES: Uncontaminated water--solar still or snow 6 inches below contamination level. Delay as long as possible from drinking from contaminated sources. PRIORITY: (1) underground, (2) running, (3) stationary. Filter sources 2 & 3 through 12 inches of earth and add purification tablets before use. BOILING WILL NOT REMOVE RADIOACTIVITY.

ANIMAL FOOD: DO NOT BUTCHER SICK ANIMALS. Discard internal organs and meat next to bone. Hides/feathers may be heavily contaminated. Eggs OK to eat.

SEA FOOD: Ocean sources OK, others on a risk basis. Fish from stationary bodies of water probably contaminated.

PLANT FOOD: Plants with edible portions below ground first choice. Smooth plant food easy to wash. Rough surfaced plants difficult to wash. Wash, pare, then wash again.

FIRST AID: NO FIRST AID FOR RADIATION SICKNESS. Infection, main danger. Personal hygiene important. Rest, avoid fatigue. Drink liquids.

BURNS: Normal first aid. Cool and cover burn. Do not use grease, etc. Treat for possible shock. Keep warm, lie with feet above head.

FRACTURES: Normal first aid. Immobilize and splint. Possible shock.

BLEEDING: Normal first aid. Apply pressure at break. Tight tourniquet can cause loss of limb--use only as last resort.

Figure 29-26. Equivalent Residual Dose Rate.

NUCLEAR EXPLOSIONS: FALL FLAT. Cover exposed body parts. Present minimal profile to direction of blast. DO NOT LOOK AT FIREBALL. Remain prone until blast effects over.

SHELTER: Pick ASAP, 5 minutes unsheltered max. PRIORITY: (1) cave or tunnel covered with 3 or more feet of earth, (2) storm/storage cellars, (3) culverts, (4) basements, (5) abandoned stone/mud buildings, (6) foxhole 4 ft deep--remove topsoil within 2 ft radius of foxhole lip.

RADIATION SHIELDING EFFICIENCIES: One thickness reduces received radiation dose by one-half. Additional thickness added to any amount of thickness reduces received radiation dose by one-half.

Iron/Steel	.7 in	Earth	3.3 in	Wood (Soft)	8.8 in
Brick	2.0 in	Cinder Block	5.3 in	Snow	20.3 in
Concrete	2.2 in	Ice	6.8 in		

SHELTER SURVIVAL: KEEP CONTAMINATED MATERIALS OUT OF SHELTER. Good weather; bury contaminated clothing outside of shelter--recover later. Bad weather; shake strongly or beat with branches. Rinse and (or) shake wet clothing--DO NOT WRING OUT.

PERSONAL HYGIENE: Wash entire body with soap and ANY water; give close attention to fingernails and hairy parts. No water: Wipe all exposed skin surfaces with clean cloth or uncontaminated soil. Fallout/dusty conditions, keep entire body covered. Keep handkerchief/cloth over mouth and nose. Improvise goggles. DO NOT SMOKE.

EXIT PLANNING: Explosion time and a dose-rate known, use rate decay nomograph.

Explosion time unknown: 1 Take dose-rate readings every hour. 2 When any reading is 1/2 of any previous reading, multiple time difference between two readings by 7/4. 3 Subtract the resultant from the time of first reading. 4 This new time is approximate time of explosion. Use new time with dose rate nomograph to determine safe exit time.

No rate meter, complete isolation first 4-6 days after last explosion.

Day 3/7: Brief exposure, 30 minutes MAX.

Day 8: Brief exposure, 1 hour MAX.

Days 9-12: Exposure of 2-4 hours per day.

Day 13 on: Normal movement.

MAXIMUM SURVIVAL DOSE/ERD: 200 roentgens

Figure 29-27. Nuclear Explosions.

their leaves than through their roots. In the tests conducted in the Marshall Islands, high levels of radioactivity were found in the water and on the external surfaces of the plants early after the detonation, but only small amounts of beta activity and no alpha activity were detected in the edible portions of the fruits and vegetables. High levels of contamination, similar in activity to that of the contaminated water nearby, were found in the sap of the coconut tree. From these and other tests, we know that small amounts of fission products may be immediately absorbed by plants growing on soil contaminated by fallout material.

(4) If the countermeasures recommended for obtaining protection against penetrating external gamma radiation and beta radiation are taken immediately, and the rules for constructing a shelter and selecting food and water are applied, a survivor's chances of surviving a nuclear attack are excellent. Figures 29-26 and 29-27 contain some summarized information applicable to a radioactive environment.

Chapter 30

BIOLOGICAL CONDITIONS

30-1. Introduction. Biological agents are viruses and micro-organisms, or their products, which are used to cause disease, injury, or death to people, animals, or plants (and, to a lesser extent, deterioration of material). Their use is an attempt to produce disease on a large scale. During war, these agents will probably be used primarily in a strategic role to attack rear area bases.

a. Most micro-organisms are harmless to humans, animals, or plants, and a few are helpful. Yeast is used in bread, beer, and cheese production. Some micro-oganisms, however, do produce disease. Biological agents could consist of:

(1) Fungi - mold, mildew, athlete's foot; histoplasmosis; and other pathogenic fungi.

(2) Bacteria - plague, tularemia, anthrax.

(3) Rickettsiae - Rocky Mountain spotted fever, typhus.

(4) Viruses - yellow fever, smallpox.

(5) Biotoxins - mushroom, algae, and bacterial poisons.

b. Many biological agents are living, require moisture, food, and certain temperature limits for life and growth. They are killed by simple acts such as boiling water, adding purification tablets to water, cooking food, exposing them to sunlight, and (or) using germicides. Biological agents enter the body through the nose, mouth, or skin; however, most will not penetrate intact skin. By preventing their entry into the body, a survivor is safe from biological agents.

30-2. Detection. There is no simple method of detecting biological agents. A person cannot see, feel, or taste these agents in a biological attack whether spread through conventional means or sabotage. Additionally, a person cannot taste the toxins in food. The basic methods of disseminating the agents are through the generation of an aerosol, the use of disease-carrying vectors, and food and water.

a. Aerosols are particles composed of many organisms, or a single organism, which are dispersed into the air and transported by air currents. Effective transmission as an aerosal requires that biological agents reach the target area with an effective percentage remaining alive and capable of causing disease. The appearance of certain clues may warn a survivor of an aerosol biological attack. They are:

(1) Aircraft dropping objects or spraying. Both enemy aircraft or friendly aircraft could be engaged in neutralizing or destroying opposing forces.

(2) Breakable containers or new and unusual types of shells and bombs, particularly those which burst with little or no blast.

(3) Smokes and mists of unknown origin.

(4) Unusual substances on the ground or on vegetation. Sick, dead, or dying animals.

b. Vectors, such as mosquitoes and ticks which carry disease, can be delivered to the target area in containers

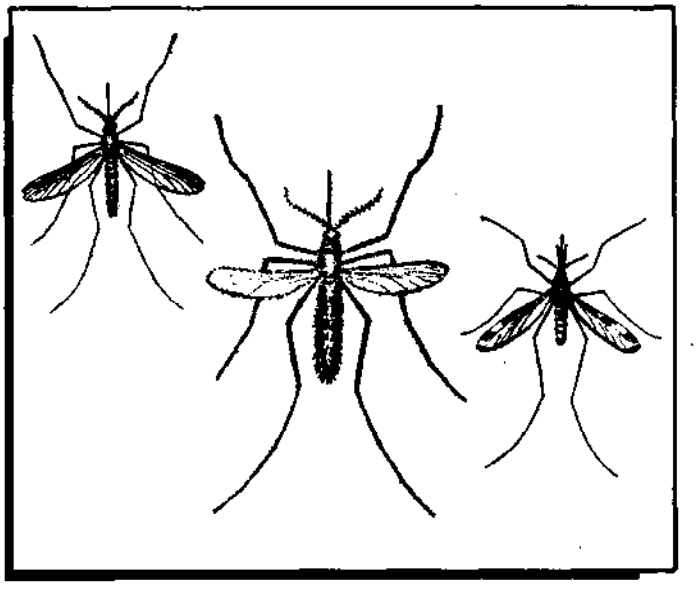

Figure 30-1. Vectors.

by aircraft or missiles. The containers rupture on impact and release the vectors. Some vectors need a "host" or carrier that can transmit the disease organism through the skin; others may be inanimate objects, such as contaminated food and water. Because disease organisms can infect personnel, the use of a "protective" mask may not help protect against viruses. The vectors can produce disease throughout their entire life spans, regardless of how far or where they travel. Furthermore, they may pass the disease to successive generations. Therefore, survivors must be extremely cautious when skinning wild game by wearing gloves and other protective clothing as the game may host fleas which carry many diseases. (NOTE: Contact with animals should be avoided unless they are to be used as food (figure 30-1).)

30-3. Climate. The various characteristics of aerosols and vectors will affect their utility in varying climates. While a survivor may encounter any form of biological agent once their use in combat occurs, the following factors may help a survivor assess the relative risk of various types of biological agents.

a. Aerosols are generally much more controllable in the area of application than are vector-borne agents. However, most aerosol agents deteriorate to some de-

gree when exposed to direct sunlight. They are more suitable for use against small area targets.

b. Strong winds are necessary to ensure maximum area coverage of both aerosols and vectors. However, the density of coverage may be decreased by extremely high winds.

c. Vectors are less controllable than aerosols in the area of application following release. They are more suited for use against broad area targets. Although vectors tend to last longer in humid climates, many potential vectors (flies, fleas, mosquitos, lice, etc.) will thrive in virtually any environmental area.

d. Generally, nighttime (1 hour before sunset to 1 hour after sunrise) is the best period to dispense vectors. Vector movement and activity is usually greater during the cooler hours of darkness.

30-4. Terrain. Biological agent aerosols tend to follow the contour of rolling terrain and valleys very much like airborne particles, such as fog. Vegetation can slow the downwind travel of agents by removing some particles from the air. Due to a lack of sunlight, densely vegetated areas, such as a jungle (warm humid), allow some agents to thrive for extended periods of time. Because most of the biological agents are more hazardous when inhaled then when directly exposed to the skin, contamination of the ground following an attack is less dangerous to the survivor than exposure to an aerosol attack.

30-5. Protection. Defense against biological warfare is neither simple nor easy. The best defense against these agents is the natural resistance of the survivor's body, a high standard of personal cleanliness, careful attention to sanitation, good nutrition, up-to-date immunization status, proper use of drugs, and immediate self-aid to any break in the skin or a punctured wound. Germs must actually get inside the body to cause disease.

a. A protective mask, properly fitted and in good condition, will greatly reduce the danger of inhaling germs. If a mask is not available, a handkerchief or parachute material over the mouth and nose will suffice to provide protection. Since survivors cannot detect the presence of biological agents, they should wear the mask or some other protective equipment over the mouth and nose until they are rescued, if possible.

b. Cuts and sores, in addition to the nose and mouth, are open doors to germs trying to enter the body. Wounds must be kept clean and protected with a bandage. Any type of clothing will give some protection. Fasten the shirt or jacket collar, roll the sleeves down and button the cuffs, wrap the trouser legs or tuck them inside the boots, and tie down all other clothes to stop the entry of germs which may be in the air or on the ground. If the survivor has a uniform used for protection against chemical agents, it should be worn because it gives a greater degree of protection against germs than ordinary clothing.

c. Survivors should always be careful about eating and drinking during and after a biological attack. One of the easiest ways for germs to enter the body is through food and water.

30-6. Tips for a Survivor:

a. Keep the body and living quarters clean.

b. Don't neglect preventive medicine. Keep the shot record up to date.

c. Keep alert for signs of biological attack.

d. Keep the nose, mouth, and skin covered.

e. Protect food and water. Bottled or canned foods are safe if sealed. If in doubt, boil the food and water for 10 minutes.

f. Build a shelter, if possible. Shelter should be located and constructed to minimize vector and aerosol access to the survivor; for example, shelter enclosed—entrance 90 degrees to the prevailing wind.

g. If traveling, travel crosswind or upwind.

Chapter 31

CHEMICAL CONDITIONS

31-1. Introduction. Chemical agents may be solid particles, liquids, or gases which are toxic (poisonous) chemicals. These agents produce poisonous gas, fire, and smoke. The poisonous agents may produce casualties or irritating effects and render material or areas unusable. The body is attacked by chemical agents which produce specific damage depending on the type and concentration exposure of the agent used. Survivors must have a thorough knowledge of how each of these agents affects the body.

31-2. Chemical Groups. Chemical agents are divided into seven groups—nerve, blood, blister, choking, vomiting, incapacitating, and riot control. These agents can be dispersed by artillery shells, mortar shells, rockets, aircraft spraying, and bombs (figure 31-1).

a. Nerve Agents. The nerve agents are among the deadliest of all chemical agents. They directly affect the nervous system and are highly toxic in both liquid and vapor forms. Examples of G-agents are tabun (GA), sarin (GB), and soman (GD). These nerve agents may be absorbed through any body surface. When dispersed as a vapor, they are absorbed through the respiratory tract or the eyes, but as liquid nerve agents, they can be absorbed through the skin. They are usually quick-acting casualty agents. Symptoms accompanying very small doses are headaches, dizziness, dimmed vision, and nausea. Large doses of nerve agents can interfere with breathing and may cause a tightness in the chest or convulsions, paralysis, and death. Symptoms of large doses of nerve agents are unpredictable, and circulatory collapse can occur without warning. The first effect of eye exposure to agents will probably be a dimming of vision caused by contraction of the eye pupils to pinpoint size. The pinpoint pupils will more noticeably affect vision in dim light. If the nerve agent contaminates the skin only, the pupils may remain normal or be only slightly reduced in size with other symptoms being first to appear. The injuries caused by nerve agents may range from mild disability to death depending on the dose received and the adequacy and speed of self-aid treatment. Nerve agents are odorless, unlike most chemical agents. A survivor must rely on observation of living things and detection devices to identify their presence.

b. Blood Agents (Cyanides). Blood agents produce their effects by interfering with some vital process within the body. The usual route of entry is inhalation. They prevent the body cells from using oxygen. Hydrocyanic acid (AC) and cyanogen Chloride (CK) are the important agents in the group. CK also acts as a choking agent (figure 31-2).

Figure 31-1. Chemical Conditions.

(1) Symptoms associated with blood agent contamination vary. One type of blood agent causes a marked increase in the breathing rate; whereas, another type

Figure 31-2. Symptoms.

causes a slow breathing rate, a choking effect, and a strong irritating effect. A slight exposure to still another type of blood agent causes headaches and uneasiness.

(2) Blood agents cause the skin to have a cherry-red color similiar to that seen in carbon monoxide poisoning. This symptom, by itself, may help identify the blood agents' poisoning. The symptoms produced by blood agents also depend upon the concentration of the

agent and the duration of exposure. If irreversible damage has not occurred, removal from the exposure to the agent may enhance recovery. Blood agents are used as quick-acting casualty agents. The speed in donning a protective mask is critical to survival in a blood agent attack.

c. Blister Agents (Vesicants). Blister agents were developed during World War I to circumvent the protective mask that had made chlorine gas obsolete. These agents are primarily designed to attack the body through the skin and eyes. They can also attack through the respiratory or digestive tracts and cause inflammation, blisters, and general destruction of tissue. Some examples of blister agents are mustard (HD), nitrogen mustards (HN), lewisite (L) and other arsenicals, mixtures of mustards and arsenicals, and phosgene oxime (CX). They are effective even in small quantities and produce delayed casualties.

(1) A drop of mustard-type agent the size of a pinhead can produce a blister the size of a quarter. Blister agents are more effective in hot weather than in cold weather. Vapors first affect the moist parts of the body (joints of arms and knees, armpits, and crotch). People who are sweating are especially sensitive to the agents. Blister agents are quickly absorbed through the skin. Reddening of the affected area may appear any time up to about 12 hours after exposure depending on the degree of contamination and the weather conditions. Blisters may appear in a day or less following the reddening. Healing time varies from about 6 days to as much as 8 weeks in severe cases. Since the damage is done during the first few minutes of exposure, speed in administering self-aid is essential.

(2) Damage to the eyes may be worse than the effects on the skin. Even as a liquid, the agent may only mildly irritate the eyes at first or there may be no pain at all. In a few hours, however, the eyes hurt, become inflamed, and are sensitive to light. Tears and great pain follow, and permanent injury can result. Some blister agents cause immediate pain in the eyes.

(3) When inhaled, blister agents inflame the throat and windpipe and cause a harsh cough. Cases of serious exposure may result in pneumonia and death. Immediate detection of blister agents and prompt protection against entry into the eyes and lungs and on the skin is vital.

(4) Blister agents may be absorbed by any material (wood, concrete, clothing, metal, plastics, or rubber). Direct skin contact with these objects can cause blistering. Liquid blister agents will eventually penetrate gloves and other garments. Immediate decontamination after exposure is essential to prevent delayed absorption.

d. Choking Agents (Lung Irritants). These agents cause irritation and inflammation of bronchial tubes and lungs but do not harm the skin or digestive system. They are usually disseminated as gases and inhaled into the body. The best known of these agents is phosgene. During and immediately after exposure, symptoms include coughing, choking, and a feeling of tightness in the chest, nausea, and occasionally vomiting, headache, and crying. If large amounts enter the lungs, they will fill with liquid causing death from lack of oxygen. A properly operating and well-fitted mask protects against all choking agents.

e. Vomiting Agents. These agents produce strong pepper-like irritation in the upper respiratory tract. Other symptoms include irritation of the eyes and uncontrollable tearing. Symptoms of these agents include a very stuffy nose, severe headache, intense burning in the throat, and tightness and pain in the chest. These are followed by uncontrollable sneezing, coughing, nausea, vomiting, and a general feeling of bodily discomfort. These agents are dispersed as aerosols and produce their effects by inhalation or by direct action on the eyes. If survivors inhale a vomiting agent before donning their masks, they may become ill after the respirator is on. As long as a vomiting agent is present, however, mask-wear is essential. The mask should be pulled away from the chin during actual vomiting, but not removed. If the survivor has vomited in the mask, caution should be taken to avoid inhaling or ingesting the vomit. Vomiting agents are not considered a major threat because of the comparative ease of protection against them and their lower toxicity unless used with other agents. Vomiting agents alone seldom produce death (figure 31-3).

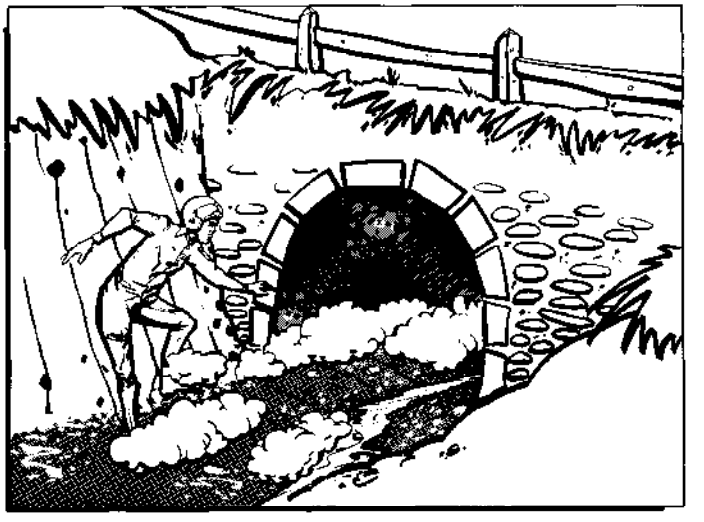

Figure 31-3. Avoiding Agents.

f. Incapacitating Agents. An incapacitating agent is any chemical which produces a temporary disabling condition which persists for hours to days after exposure to the agent has ceased (unlike that produced by riot control agents) and for which medical treatment, while not required, facilitates a more rapid recovery.

(NOTE: Symptoms of riot control agents may not be distinguished from other lethal agents; therefore, the survivor must be prepared to provide treatment for lethal agents.) In actual usage, the term "incapacitating agent" has come to refer primarily to those agents which:

(1) Produce their effects mainly by altering or disrupting the higher regulatory activity of the central nervous system (figure 31-4).

(2) Last for hours or days rather than the very short duration of riot control agents.

(3) Do not seriously endanger life (except when large doses are received) and produce no permanent injury.

Figure 31-4. Effects.

g. Riot Control (RC) Agents (Irritant Agents). RC agents are the least poisonous of the seven groups of chemical agents. They act primarily on the eyes, causing intense pain and tearing. Higher concentrations irritate the upper respiratory tract and the skin and sometimes cause nausea and vomiting. These agents may be dispersed as smoke or in solutions as droplet aerosols. Although they are used primarily in training and in riot control, some agents may be used in combat. When an unmasked person comes in contact with riot control agents, the effects are felt almost immediately. The effects begin in 20 to 60 seconds, depending upon agent concentration. Duration of effects lasts 5 to 10 minutes after removal to fresh air. There is usually no permanent damage to the eyes. For a short time (minutes), a person may be unable to see. If the mask is used before RC agents enter the eyes, increased protection is afforded.

31-3. Detection:

a. General Indications. Detection of a chemical agent requires the recognition of evidence gathered by direct or indirect means. Therefore, every survivor must be alert and able to detect any clues indicating chemical warfare is being used. General symptoms of chemical agents are tears, difficult breathing, choking, itching, coughing, and dizziness. In the presence of agents that are very hard to detect without the use of detection devices, survivors must watch their fellow aircrew members constantly for symptoms. Additionally, a survivor's surroundings may provide valuable clues to the presence of chemical agents; for example, dead animals, sick people, or people displaying abnormal behavior.

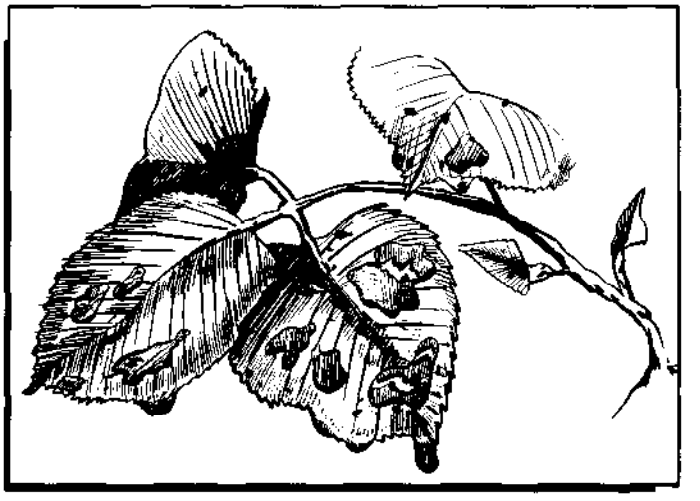

Figure 31-5. Detection.

b. Smell. Survivors cannot rely on the nose as a foolproof means of detecting chemical agents. Although some agents do have a characteristic odor, many others have little or no odor at all. An agent may smell quite differently to different individuals. A mixture of agents will have a different smell than any one agent by itself.

c. Sight. Since chemical agents are in one of three physical states—solid, liquid, or vapors—the sense of sight may help detect their presence. Most chemical agents in the solid or liquid state have some color. In the vapor state, some chemical agents can be seen as mist or thin fog immediately after bomb or shell bursts. Nerve agents are either a colorless liquid or a colorless vapor. Although survivors can't see nerve agents, their eyes may help by detecting the methods used to dispense the agents. Mustard gas, unless purified, is dark brown in its liquid state. As a liquid, it is easy to detect and would appear as oily, dark patches on leaves and buildings. However, liquid mustard changes slowly to a colorless gas. As a gas, it is still very toxic, but now the eyes will not be an effective aid to detection (figure 31-5).

d. Hearing. If survivors know the methods being used by the enemy to spread chemical agents, they can detect the sounds of the enemy's chemical munitions. For example, a bomb filled with an agent would probably cause only a muffled explosion; however, aircrew mem-

bers untrained in ordnance may have difficulty in making this distinction.

e. Feel and Taste. Irritation in the nose or eyes or on the skin is an urgent warning to protect the body from chemical agents. Additionally, a strange taste in food, water, or cigarettes may serve as a warning they have been contaminated.

31-4. Protection:

a. Protective Actions. Survivors should use the following steps, listed in the order of importance, to protect themselves from chemical attack:

(1) Use protective equipment.

(2) Give quick and correct self-aid when contaminated.

(3) Avoid the areas where chemical agents exist.

(4) Decontaminate equipment and the body as soon as possible.

b. Equipment. Survivors' masks are as vital to them as lifejackets are to sailors or as parachutes are to fliers. If properly adjusted, they protect the face, eyes, and lungs from chemical agents. Survivors are responsible for proper care of the mask and should inspect the masks frequently to ensure they are free from damage and in good condition. Aircrew members located in areas of potential contamination are issued protective clothing.

c. Self-Aid. Survivors must apply self-aid skillfully and promptly after exposure. Not only is it important for them to know what to do, but they must also know what not to do. It is evident from previous information in this chapter that each type of chemical agent produces certain conditions which require special treatment. However, there are certain essentials of self-aid which, if applied soon enough, give some relief and may prevent serious injury.

(1) Since there are definite time limits after which self-aid becomes useless, immediate self-aid or personal decontamination is all-important if survivors are exposed to liquid nerve or blister agents. Since they may not know whether the contamination is by liquid nerve agent or liquid blister agent, the following procedures are recommended:

(a) Don the mask and clear it.

(b) Contact with thickened (persistent) nerve or mustard agents requires the use of decontamination kit, or if not available, tear away the contaminated area of clothing and rinse immediately with water.

(c) Rinse contaminated areas with water (removes nerve agents).

(d) If effects of nerve agents become apparent, then and only then, use an antidote, realizing the antidote provides protection only from nerve agent (GA, GB, GD). It is also incapacitating and is not effective against V agents.

(2) Use the self-aid procedure given in the following paragraphs for specific agents if the agent has been identified.

(a) Nerve Agents. The protective mask and hood, if available, must be donned immediately at the first sign of a nerve agent in the air. Stop breathing until the mask is on and the facepiece cleared and checked. The mask should be worn constantly until the absence of the nerve agent in the air is indicated or the individual moves into a clean area (where there are live animals, etc.) If, after masking, the survivor has an unexplained runny nose, tightness of the chest, dimness of vision, or breathing difficulty, the use of the antidote should be considered. (Use of the antidote is moderately to severely incapacitating. The survivor should consider the severity of symptoms, availability of the buddy care system and requirements for rescue before injecting the antidote.) Exposure to high concentrations of a nerve agent may bring on incoordination, mental confusion, and collapse so rapidly that the survivor cannot perform self-aid. If this happens, a fellow aircrew member must administer first aid. Severe nerve agent exposure may rapidly cause unconsciousness, muscular paralysis, and breathing stoppage. Any remaining survivors should keep their masks on and move out of the area as soon as possible. The following precautions should be used when applying self-aid for nerve agents:

-1. Antidote should not be used until certain it is needed. Pinpointing of the eye pupils or blurred vision, tightness in the chest, and difficulty in breathing are signs it is needed. If certain nerve agents are inhaled, the antidote counteracts them and makes the survivor feel better.

-2. If survivors have inhaled a very large dose of nerve agent vapor, they may need more than one injection of the antidote to relieve their symptoms. If the symptoms are steadily becoming worse and the first injection does not relieve them, or if their mouth does not become dry, it may be necessary to use an extra dose. Inject the second dosage in a different muscle. (NOTE: Do not use your own combo pen to inject a victim; use theirs. Additionally, if you find a deceased aircrew member, remove the combo pens from the deceased and take them with you.)

-3. If the difficulty in breathing is not relieved by the second injection, one more dose may be administered. Dryness of the mouth is a good sign. It means they have had enough antidote to overcome the dangerous effects of the nerve agent.

-4. If a drop or splash of liquid nerve agent gets into the eye, instant action is necessary to avoid serious effects. The eye should be irrigated immediately with water by tilting the head back, looking straight upward. Slowly pour water into the contaminated eye. Hold the eye open with the finger if necessary. Pour the water slowly so that irrigation will last at least 30 seconds. Survivors must irrigate in spite of the danger of breath-

ing nerve agent vapor. Don the mask quickly after completing the irrigation. The pupil of the contaminated eye should be observed during the next minute, in a mirror if one is available or by someone nearby. If the pupil rapidly gets smaller, inject antidote intramuscularly at once. If the pupil does not get smaller, antidote is not needed.

-5. If liquid nerve agent gets on the skin or clothing, it should be removed. The liquid should be blotted off the skin with a handkerchief, a piece of cloth torn from the outer clothing, or personal decontaminating kits. Pinch-blotting the liquid won't spread the contamination. Contaminated clothing must be quickly removed and the skin washed with soap and water. In an emergency, the contaminated portion of the clothing can be cut away and the contaminated skin area flushed with water. The muscles under the contaminated area should be observed for any signs of twitching. If none develops in the next half hour and the survivor has no tightness in the chest, the decontamination was successful. If twitching of the muscles under the area of contaminated skin does develop, the antidote should be administered at once.

-6. Food and water which may be contaminated with nerve agents must be avoided. If a survivor has swallowed contaminated food or water and all of these symptoms occur—nausea, pains in the stomach, increased flow of saliva, and a tightness in the chest—the antidote must be administered.

(b) Blood Agents. If, during any chemical attack, a sudden stimulation of breathing or an odor like bitter almonds is noticed, the survivor must don the mask as quickly as possible. Speed is absolutely essential; this agent acts so rapidly that within a few seconds its effects will make it impossible for survivors to don the mask by themselves. The breath should be held until the mask is on the face, if at all possible. This may be very difficult since the agent strongly stimulates respiration.

(c) Blister Agents. The protective mask, hood, and clothing must be worn when liquid or vaporized blister agents are known to be present. There are two groups of blister agents, one called mustards and the other called arsenicals. Self-aid against mustards and arsenicals is the same. A liquid mustard in the eye will not hurt immediately. A liquid arsenical in the eye will sting and hurt severely.

-1. To remove a liquid blister agent from the eye, the eye is irrigated using the same procedure as for removing nerve agents. Speed in decontaminating the eye is absolutely essential. The self-aid procedure is very effective for mustard within the first few seconds after exposure but is of little value after 2 minutes.

-2. Generally, for any liquid blister agent on the skin, the survivor should:

-a. Pinch-blot to prevent the liquid from spreading, using cloth or any other absorbent material at hand. The used cloth or absorbent material should then be discarded.

-b. Scrub the skin with soap and water and rinse thoroughly with clean water. When scrubbing, special attention must be paid to areas not covered by clothing (neck and ears).

-3. Survivors must decontaminate or remove any clothing which is contaminated with liquid blister agent. Small areas on the clothing can be decontaminated by using soap and water. The contaminated parts of the clothing can also be cut out, thus making the clothing safe to wear.

-4. When mustard vapor is detected, a survivor must put on the mask and leave it on until clear of the area. There is no decontamination procedure of any value when a mustard vapor agent is used. The damage is done as soon as the mustard vapor strikes the eyes, although the full extent of the injury may not appear for several hours.

(d) Choking Agents. The protective mask should be donned immediately upon detection of any choking agents in the air by odor (like cut green corn or grass), irritation of the eyes, or change in the taste of a cigarette (smoking may become tasteless or offensive in taste). Survivors should hold their breath while masking. If an agent has been inhaled, normal survival duties should be continued unless there is difficulty in breathing, nausea and vomiting, or more than the usual shortness of breath from exertion. If any of the above symptoms occur, survivors should rest.

(e) Incapacitating Agents. The mask should be donned immediately. Complete cleaning of the skin with soap and water should be done at the earliest opportunity. The eyes should be flushed with clear water only. Survivors should shake or brush clothing, and when conditions permit, change clothing and thoroughly wash the contaminated clothing.

(f) Vomiting Agents. The protective mask must be worn in spite of coughing, sneezing, salivation, and nausea. The mask can be lifted from the face briefly if necessary to permit vomiting or to drain saliva from the facepiece. Carrying on duties as vigorously as possible will help to lessen and to shorten the symptoms. Survival duties can usually be performed despite the effects of vomiting agents.

(g) Riot Control Agents. After the protective mask has been donned and cleared, the eyes should be kept open as much as possible. When vision clears, activities can continue. The eyes should not be rubbed. If drops or particles have entered the eyes, the eyes can be flushed with water. Chest discomfort can usually be relieved merely by talking.

31-5. Avoiding Chemical Agents. If survivors are hit by a chemical attack, they may have to remain in a contaminated area. After an attack, they should not expose themselves to other enemy weapons and must seek ar-

eas which are less contaminated. If the attack is on a very small scale, they might seek an upwind area. Depending on the area and weather conditions, crosswind movement may be best. Chemical agent attacks may cover too large an area to permit area avoidance. Selecting routes on high ground may be advisable because gas is usually heavier than air and tends to settle in low places. Cellars, trenches, gullies, and valleys are examples of places to avoid when possible. Woods, tall grass, and bushes tend to hold chemcial agent vapors. (NOTE: Survivors have a better chance to avoid chemicals if they are familiar with the attack areas and if they know their personal location.)

31-6. Decontamination of Chemical Agents. Decontamination is removing, neutralizing, or destroying the agents. The purpose of personal decontamination is to remove agents from the body or personal equipment before serious injury occurs. An example of decontamination by removal is pinch-blotting the chemical agent from the skin. Neutralization makes the agent harmless. The contaminated cloth could also be buried. Common sense and quick thinking play a big role in personal decontamination. Survivors may have to rely on whatever they have at hand to remove chemical agents from the skin, eyes, or equipment. If liquid nerve or blister agents touch any part of the body, they must be removed as quickly as possible. If survivors are caught without soap and water, then anything which can dilute or remove the agents will have to be used; it may be mud, dirt, or urine. A crude remover may remove only two-thirds of the agent, but it is far better than leaving the agent in full concentration. It must be remembered that nerve and blister agents penetrate very rapidly. (NOTE: Use a scraping action to avoid pressing the agent into the skin.)

a. Soap is excellent for removing chemical agents. Cold water, while not as good as warm water, does dilute or weaken chemical agents. Hot, soapy water removes agents quickly. If the operational situation permits, a bath or a shower should be taken. The mask should be left on until after survivors have washed their hair and thoroughly scrubbed themselves while avoiding wetting the canister or cheek pads. Exposed areas and hairy regions of the body should be given extra attention.

b. When clothes have been exposed only to chemical agent vapors, airing usually decontaminates them (with the exception of mustard vapors which will absorb into the garment and requires washing for removal). If chemical agent droplets or liquid splashes are present, survivors will need detergent or soap and water. Wool clothes are best washed in soapy, lukewarm water. Cotton clothes can be boiled.

c. Boots or shoes can be scrubbed with soap and water and rinsed at least twice. If the survivors' choice is wearing contaminated clothes and shoes or nothing, decontamination of the material must be done the best way possible. Almost any effort will help the survival situation.

31-7. Tips for a Survivor:

a. Keep the body and living area clean.

b. Keep the nose, mouth, and skin covered.

c. If a protective mask is needed, but unavailable, improvise. The charcoal cloth from the CD suit or undergarment makes a moderately effective mask for short-term agent exposure. The use of the aircrew helmet, visor, and charcoal mask may provide a higher level of protection for the eyes and respiratory system.

d. Build a shelter or rest area in a clearing away from vegetation. Decontaminate the ground by removing the top soil. Keep the entrance closed and 90 degrees to the prevailing wind.

e. Do not use wood or vegetation from a contaminated area for a fire.

f. Look at the area around a water source and check for foreign odors (garlic, mustard, geranium, bitter almond), oily spots on the surface or nearby, and the presence of dead fish and animals. If they are present, do not use the water.

g. Keep food and water protected. Bottled or canned foods and water are safe if sealed, and the cans are decontaminated before opening.

h. If possible, obtain water from a closed source, precipitation (if there is no evidence of agent vapor in the air), and from a slow-moving stream after it has been filtered.

i. Do not use plants for food or water in a contaminated area.

j. Do not use sick animals as a food source. When skinning animals in a contaminated area, use protective clothing (gloves).

k. If traveling, travel crosswind or upwind.

MANUAL COMPLETO DE

SUPERVIVENCIA

Hugh McManners

EDITORIAL LA ISLA

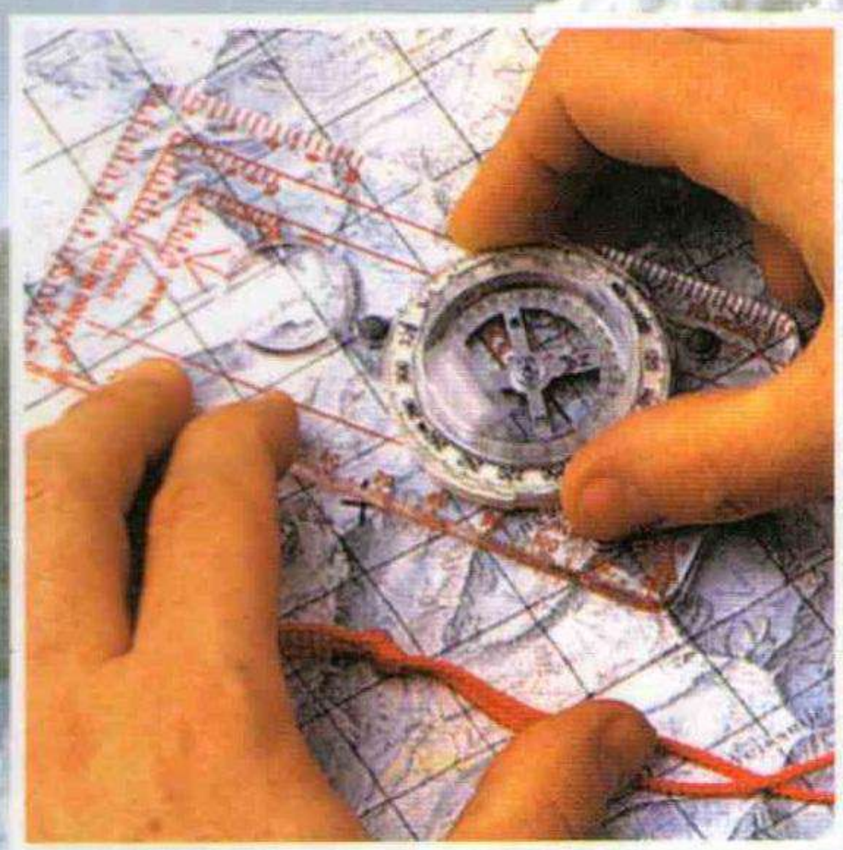

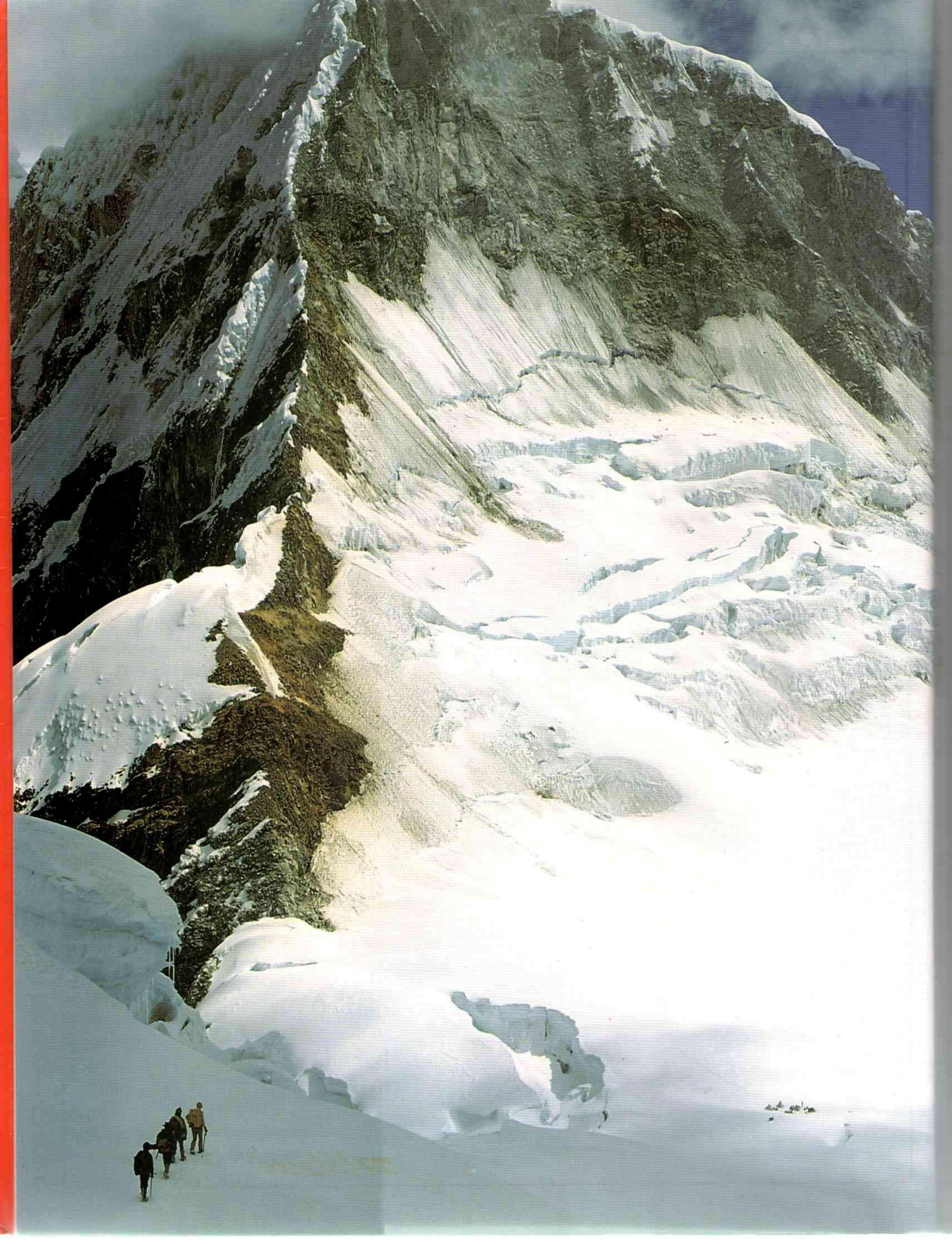

LA ISLA

A DORLING KINDERSLEY BOOK

Título original:
The Commando Survival Manual

Traducción:
Maite Rodríguez Fischer

Revisión científica y técnica de la edición en lengua española:
Ignacio Añibarro; Koldo Sansinenea; Javier Busselo; Pablo Sánchez; José M. Peña; Román Muñoz; Iñaki García; Cándido García; Asier Beltza; Giorgio Studer
Espeleólogos.
Profesionales expertos en vida en exteriores y acampada en diversos entornos.

Coordinación de la edición en lengua española:
Cristina Rodríguez Fischer

Primera edición argentina, versión en rústica, 1999
Primera edición argentina 1994

© 1994, 1999 La Isla, S.R.L., Buenos Aires
© 1994, 1998 Dorling Kindersley Limited, Londres
© 1994, 1998 del texto, Hugh McManners

I.S.B.N.: 950-637-105-9
Hecho el depósito que marca la ley 11.723
Impreso en Singapur

Todos los derechos reservados.
Queda prohibida la reproducción total o parcial de esta obra, sea por medios mecánicos o electrónicos, sin la debida autorización por escrito del editor.

Venta exclusiva en Argentina,
Chile, Paraguay y Uruguay

Nota Importante

Algunas de las técnicas descritas en este libro únicamente deben utilizarse en situaciones de emergencia grave, cuando la supervivencia de los individuos dependa de ellas. En consecuencia, los editores no pueden hacerse responsables de cualquier daño, pérdida o acción legal que pueda resultar del uso o abuso de la información contenida en este libro. No practique estas técnicas en propiedades privadas sin el permiso del propietario, y obedezca la legislación en lo que concierne a la protección de la tierra, la propiedad, las plantas y los animales.

Contenido

Introducción 6

Capítulo 1 ¿Es usted un superviviente?

Condición física 12
Necesidades nutricionales 14
La adaptación al medio 16
Estar en forma 18

Capítulo 2 Equipamiento básico

La ropa adecuada 22
Calzado 24
Equipo personal de campamento 26
Equipo básico de supervivencia 28
Camas y sacos de dormir 30
Herramientas 32
Cuerdas y nudos 34
El manejo de la cuerda 36

Capítulo 3 La vida en la naturaleza

La elección del campamento 40
La elección de una tienda 42
Montar una tienda de campaña 44
Desmontar una tienda de campaña 46
La construcción de un refugio 48
Refugios en la nieve 50
Refugios tropicales 52
Refugios permanentes 54
Un campamento seguro 56

El fuego en el campamento 58
Cómo preparar un fuego 60
Cómo encender un fuego 62
Levantar el campamento 64

Capítulo 4 En busca de agua

La importancia del agua 68
Recoger agua 70
Fuentes naturales de agua 72
La depuración del agua 74
Transporte y almacenamiento del agua 76

Capítulo 5 Encontrar y preparar alimentos

Alimentos para el viaje 80
Plantas comestibles 82
Plantas venenosas 84
Raíces, tubérculos y bulbos 86
Nueces y frutos 88
Hongos y líquenes 90
Las plantas como alimento 92
Animales invertebrados 94
Los animales invertebrados como alimento 96
Peces de agua dulce 98
Peces de agua salada 100
Equipo de pesca 102
Técnicas de pesca 104
Redes y trampas de pesca 106
El empleo de redes y trampas de pesca 108
Preparar y cocinar el pescado 110
Utensilios de cocina 112
Cocinas y hornos 114
Métodos de cocción 116
La conservación de los alimentos 118

Capítulo 6 En marcha

La planificación del viaje 122
La orientación 124
El empleo de mapas 126
Encontrar su posición 128
Principios de navegación 130
La evaluación del tiempo 132
Prepararse para el viaje 134
La marcha a pie 136
Caminar sobre hielo y nieve 138
Viajar sobre nieve 140
Cruzar el desierto 142
Viajar a través de la selva 144
Una escalada segura 146
Escalada de emergencia 148
Cruzar las aguas 150
Aguas peligrosas 152
Construir una balsa 154
El empleo de balsas y otras embarcaciones 156
El uso de vehículos 158
Otros medios de transporte 160

Apéndice A

Abandonar el barco 162
La supervivencia en el mar 164
Animales peligrosos 166
Peligros naturales 168
Meteorología extrema 170
Códigos y señales para comunicarse 172

Apéndice B

Primeros auxilios 174

Glosario 184
Lista de comprobación para supervivencia 187
Índice 188
Agradecimientos 192

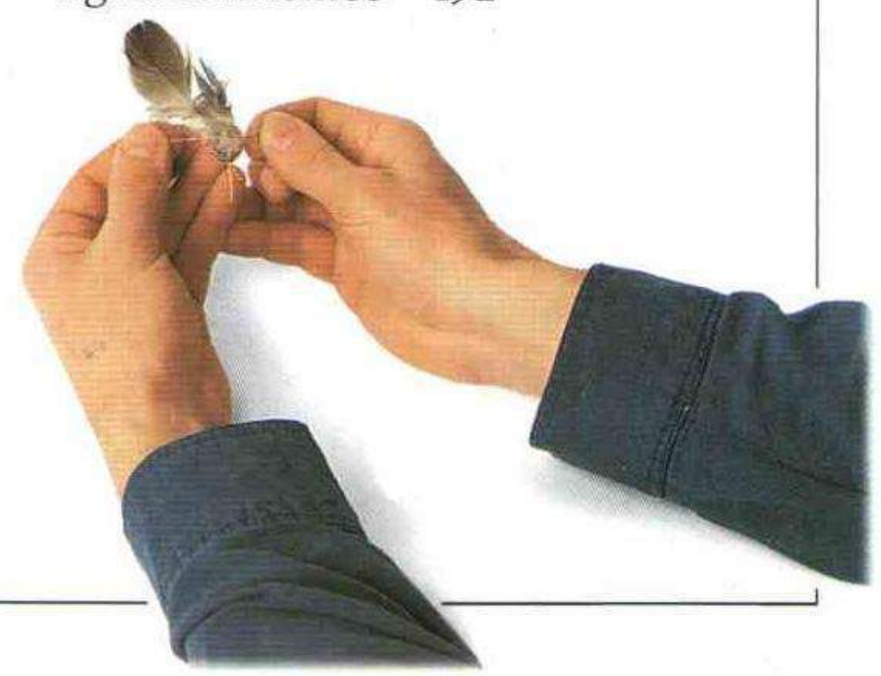

INTRODUCCIÓN

ES POSIBLE QUE la sentencia de los grandes filósofos «Cualquier tonto puede sentirse incómodo» la formulara después de un día de campo pasado por agua. Tanto si trata de secar su ropa en un campamento inundado de lluvia como si intenta sobrevivir a un desastre inesperado, el conocimiento de los principios básicos de la vida al aire libre le resultarán de una ayuda inmensurable. Estos principios incluyen la construcción de tiendas y refugios, la elección o confección de sacos de dormir o camas, encender un fuego, así como la prevención o cura de ampollas y de algunas enfermedades. El aprendizaje de algunos principios más especializados de supervivencia, como la navegación en cualquier tipo de terreno o la depuración de agua, es tan interesante y útil para los usuarios con fines recreativos, como vital para quienes desafortunadamente se convierten en «supervivientes».

A LA ESPERA DE LO INESPERADO

En la práctica, nuestro lema diario era «Espera siempre lo inesperado». Los largos años de entrenamiento en lugares lluviosos y potencialmente peligrosos valieron la pena; nuestro equipo de operaciones estuvo destacado en todo el mundo, a menudo trabajando en medios hostiles. La gente que trabaja en las fuerzas especiales es bastante realista sobre su trabajo, así que una vez se había comentado todo en el entrenamiento, no había mayores problemas al llevarlo a cabo en la realidad. Sin embargo, estábamos enfrentándonos a «lo inesperado».

Por el contrario, para los civiles que no se enfrentan a nada más peligroso que la posible avería del automóvil o al mal tiempo, el prepararse para alguna situación de emergencia puede parecerles ridículo. Sin embargo, la más inofensiva caminata familiar, particularmente en un terreno montañoso, puede convertirse rápidamente en una experiencia desagradable si los participantes ignoran los principios básicos de la predicción del tiempo, primeros auxilios, navegación, ropa y calzado. Únicamente conociendo la vida en la naturaleza evitaremos muchos desastres comunes, o por lo menos los identificaremos antes de que se conviertan en una amenaza seria para nuestra vida.

Prepararse para lo inesperado es un proceso constante de intentar entender cómo funcionan las cosas, y establecer líneas paralelas entre lo que forma parte de nuestra propia experiencia y entendemos, y las cosas extrañas y diferentes sobre las que no tenemos experiencia. Incluso cuando decidimos visitar un área desierta o un país subdesarrollado, permanecemos como espectadores, manteniéndonos alejados de la realidad de la vida diaria de aquellos lugares por la naturaleza misma de nuestra vida urbana, llena de tecnologías modernas. Pero debemos cambiar nuestras actitudes de comodidad si queremos entender cómo funciona la vida realmente. Si tuviéramos que establecernos en una remota isla tropical después de un accidente aéreo, por ejemplo, es muy poco probable que dispusiéramos de agua embotellada, de comidas precocinadas o preparadas, e incluso de gafas de sol, y los accesorios urbanos como las tarjetas de crédito serían absolutamente inútiles.

Así pues, es en nuestro propio beneficio que debemos hacer un esfuerzo serio para abandonar nuestras vidas cómodas y protegidas. Si podemos entender cómo encajamos en el mundo, reconociendo honestamente nuestras fuerzas y debilidades físicas y mentales, podemos dar el siguiente paso, convirtiéndonos en actores del mundo real, sin importar en qué parte del mismo actuamos.

CONOCIMIENTOS QUE PUEDEN SALVARLE LA VIDA

Es imposible predecir los desastres que pueden ocurrir, así que no podrá adquirir las habilidades específicas necesarias para cubrir todas las opciones. Entender los principios básicos de la supervivencia es mucho más importante que aprender técnicas de memoria, ya que la mayoría de las técnicas requieren alguna modificación cada vez que se utilizan. Nada es fácil, no importa lo bien entrenado que pueda estar, pero si aprende los principios básicos, será capaz de adaptarse a una gran variedad de situaciones diferentes.

La tecnología moderna enmascara a menudo los principios básicos que rigen todos los trabajos físicos. Creamos máquinas para cualquier trabajo, liberándonos de faenas que ocuparían la mayor parte de nuestro tiempo. Olvidamos los principios de la física que rigen desde el clima hasta las herramientas que improvisamos.

Es muy posible que hayamos aprendido y entendido muchos de estos principios en nuestras vidas, pero la mayoría de ellos están guardados en nuestras mentes, y no siempre somos capaces de aplicarlos a las situaciones de la vida real. La vida urbana tiende generalmente a hacernos inflexibles en la aplicación práctica de nuestros conocimientos, simplemente porque casi nunca nos encontramos en el tipo de situaciones en las que no hay otra opción que la improvisación para sobrevivir.

LOS PRINCIPIOS DE LA SUPERVIVENCIA

Cualquier tipo de conocimiento es importante. Incluso la física o la química casi olvidada del colegio, o una lección de biología, pueden ahorrarnos horas de experimentación. Cuando somos niños, construimos herramientas y otros artefactos continuamente, aprendiendo a aplicar los principios básicos, un proceso que se detiene cuando crecemos y empezamos a comprar artículos fabricados. En la naturaleza debemos volver a nuestra infancia en muchos aspectos, y volver a aprender a improvisar e inventar. Sin embargo, debemos trabajar como adultos, considerando los principios de ingeniería como el esfuerzo, la resistencia y la tensión, que deben combinarse teniendo una idea bien clara del propósito que nos guía. Este proceso está gobernado por la necesidad: si no lo logramos nos mojaremos, nos quedaremos hambrientos o padeceremos frío.

Al combinar paso a paso las instrucciones con fotografías e ilustraciones, este libro contiene información sobre una gran cantidad de técnicas prácticas, que muestran una transición fácil entre la vida urbana y la naturaleza en la que nuestros antepasados primitivos tuvieron que sobrevivir. Los conocimientos que contiene este libro podrían llegar a salvar su vida, si de repente se ve en la necesidad de tener que sobrevivir —incluso temporalmente—, como lo hicieron nuestros ancestros, por ejemplo después de un accidente, desastre o avería de su vehículo, o en condiciones climáticas adversas, cuando los servicios básicos se ven interrumpidos.

Por otra parte, puede que sea un aficionado a las caminatas o un amante de las excursiones en la naturaleza, para alejarse de las presiones y complicaciones de la vida urbana. Cualquiera que sea su situación, este libro le será de ayuda inestimable, ya que le proporcionará gran cantidad de información necesaria, al tocar una amplia gama de temas especializados.

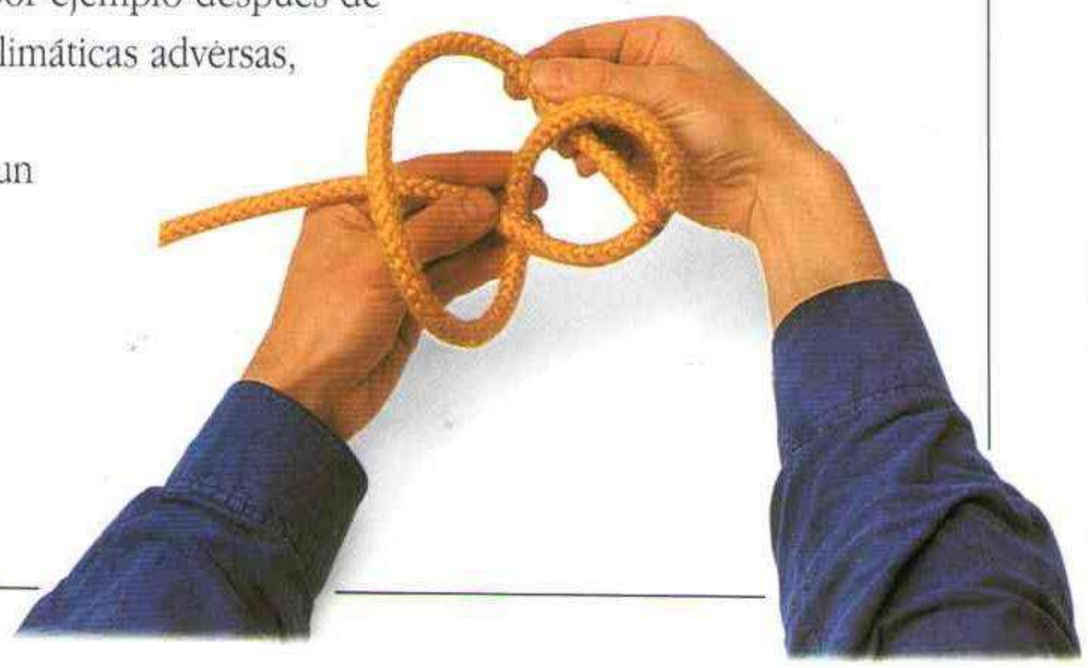

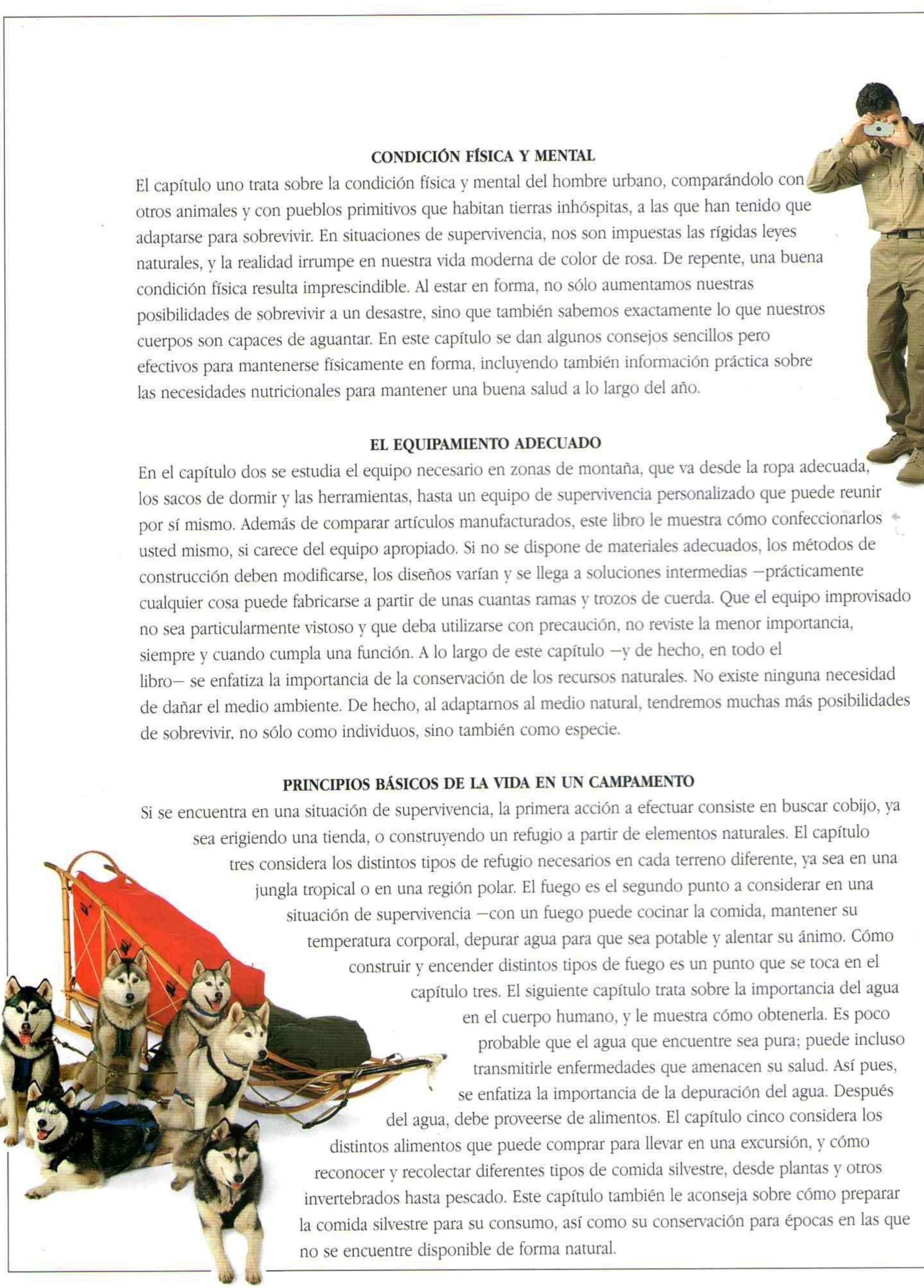

CONDICIÓN FÍSICA Y MENTAL

El capítulo uno trata sobre la condición física y mental del hombre urbano, comparándolo con otros animales y con pueblos primitivos que habitan tierras inhóspitas, a las que han tenido que adaptarse para sobrevivir. En situaciones de supervivencia, nos son impuestas las rígidas leyes naturales, y la realidad irrumpe en nuestra vida moderna de color de rosa. De repente, una buena condición física resulta imprescindible. Al estar en forma, no sólo aumentamos nuestras posibilidades de sobrevivir a un desastre, sino que también sabemos exactamente lo que nuestros cuerpos son capaces de aguantar. En este capítulo se dan algunos consejos sencillos pero efectivos para mantenerse físicamente en forma, incluyendo también información práctica sobre las necesidades nutricionales para mantener una buena salud a lo largo del año.

EL EQUIPAMIENTO ADECUADO

En el capítulo dos se estudia el equipo necesario en zonas de montaña, que va desde la ropa adecuada, los sacos de dormir y las herramientas, hasta un equipo de supervivencia personalizado que puede reunir por sí mismo. Además de comparar artículos manufacturados, este libro le muestra cómo confeccionarlos usted mismo, si carece del equipo apropiado. Si no se dispone de materiales adecuados, los métodos de construcción deben modificarse, los diseños varían y se llega a soluciones intermedias –prácticamente cualquier cosa puede fabricarse a partir de unas cuantas ramas y trozos de cuerda. Que el equipo improvisado no sea particularmente vistoso y que deba utilizarse con precaución, no reviste la menor importancia, siempre y cuando cumpla una función. A lo largo de este capítulo –y de hecho, en todo el libro– se enfatiza la importancia de la conservación de los recursos naturales. No existe ninguna necesidad de dañar el medio ambiente. De hecho, al adaptarnos al medio natural, tendremos muchas más posibilidades de sobrevivir, no sólo como individuos, sino también como especie.

PRINCIPIOS BÁSICOS DE LA VIDA EN UN CAMPAMENTO

Si se encuentra en una situación de supervivencia, la primera acción a efectuar consiste en buscar cobijo, ya sea erigiendo una tienda, o construyendo un refugio a partir de elementos naturales. El capítulo tres considera los distintos tipos de refugio necesarios en cada terreno diferente, ya sea en una jungla tropical o en una región polar. El fuego es el segundo punto a considerar en una situación de supervivencia –con un fuego puede cocinar la comida, mantener su temperatura corporal, depurar agua para que sea potable y alentar su ánimo. Cómo construir y encender distintos tipos de fuego es un punto que se toca en el capítulo tres. El siguiente capítulo trata sobre la importancia del agua en el cuerpo humano, y le muestra cómo obtenerla. Es poco probable que el agua que encuentre sea pura; puede incluso transmitirle enfermedades que amenacen su salud. Así pues, se enfatiza la importancia de la depuración del agua. Después del agua, debe proveerse de alimentos. El capítulo cinco considera los distintos alimentos que puede comprar para llevar en una excursión, y cómo reconocer y recolectar diferentes tipos de comida silvestre, desde plantas y otros invertebrados hasta pescado. Este capítulo también le aconseja sobre cómo preparar la comida silvestre para su consumo, así como su conservación para épocas en las que no se encuentre disponible de forma natural.

VIAJAR POR ÁREAS VÍRGENES

Es probable que no haya tenido la experiencia de cruzar un área totalmente virgen, pero mientras conozca los principios básicos de supervivencia en cada zona, sus oportunidades aumentarán enormemente.

El capítulo seis trata sobre la forma de viajar en áreas tan diversas como la jungla y el desierto, así como las regiones polares y las Montañas Rocosas. También se indican algunas técnicas para cruzar ríos y para caminar correctamente a fin de minimizar los accidentes, así como algunos métodos elementales de orientación y navegación que podrá utilizar en todo el mundo, y técnicas básicas de escalada en rocas, útiles si surge alguna emergencia. Además de los viajes a pie, este capítulo también considera el empleo de vehículos motorizados, trineos de perros, balsas, canoas y otras embarcaciones; e incluye consejos sobre la manera de confeccionar, cargar y utilizar este tipo de transportes.

PERMANEZCA TRANQUILO

Su supervivencia dependerá de que tome las decisiones adecuadas. Pero sólo entendiendo lo que puede pasar en una situación de supervivencia será posible tener la suficiente previsión para poder predecir algún desastre potencial, y así tomar la acción correctiva adecuada para prevenirlo. En las situaciones de supervivencia nos encontramos arrojados de repente hacia algo que no podemos evitar y en lo que muy pocas de nuestras habilidades normales son relevantes. Nuestro primer problema consiste en aceptar lo que ha sucedido. El pánico y la depresión son los mayores enemigos de la mente, e impiden que aflore en las personas el deseo innato de sobrevivir. Un reconocimiento lógico y franco de la situación, de los recursos y de las opciones, determinará los primeros pasos a tomar.

En el apéndice A se analizan los desastres y las situaciones peligrosas, así como sus causas, cómo predecirlas, prepararse para afrontarlas y sobrevivir a sus consecuencias. A continuación, se mencionan algunos ejemplos de situaciones peligrosas en las que podría llegar a encontrarse, como quedar a la deriva en una balsa salvavidas en mar abierto, enfrentarse a un animal peligroso, a los desastres naturales como los terremotos o a condiciones climáticas extremas como los huracanes. Este apéndice indica cuándo debe considerarse que tales desastres son inminentes y cómo superarlos si se enfrenta a ellos.

El apéndice B repasa las técnicas básicas de primeros auxilios, con un énfasis especial en las situaciones de supervivencia, en las que no puede simplemente llamar a una ambulancia por teléfono. Asimismo, le proporciona información sobre las enfermedades infecciosas en el mundo y cómo evitarlas.

DEPENDE DE USTED

Por encima de cualquier otra cosa, la supervivencia exige una gran fuerza mental y claridad de pensamiento. La evaluación lógica de todos sus conocimientos y un reconocimiento a conciencia de sus habilidades y de las de su equipo, pueden llevarle a una conclusión desagradable o a una decisión difícil. Sin embargo, a diferencia de lo que ocurre en la vida diaria, en una situación de supervivencia, usted determina su propio futuro, y de sus decisiones dependen su propia vida o muerte.

Aunque este libro le enseña cómo podrían hacerse las cosas, en una situación de supervivencia real debe hacerlas por sí mismo. Si las cosas no salen como se describe en estas páginas, tendrá que arreglárselas solo. Sin embargo, si sigue los detalles específicos de este libro, y adapta los principios básicos, aumentarán notablemente sus posibilidades de supervivencia, incluso en las condiciones más adversas.

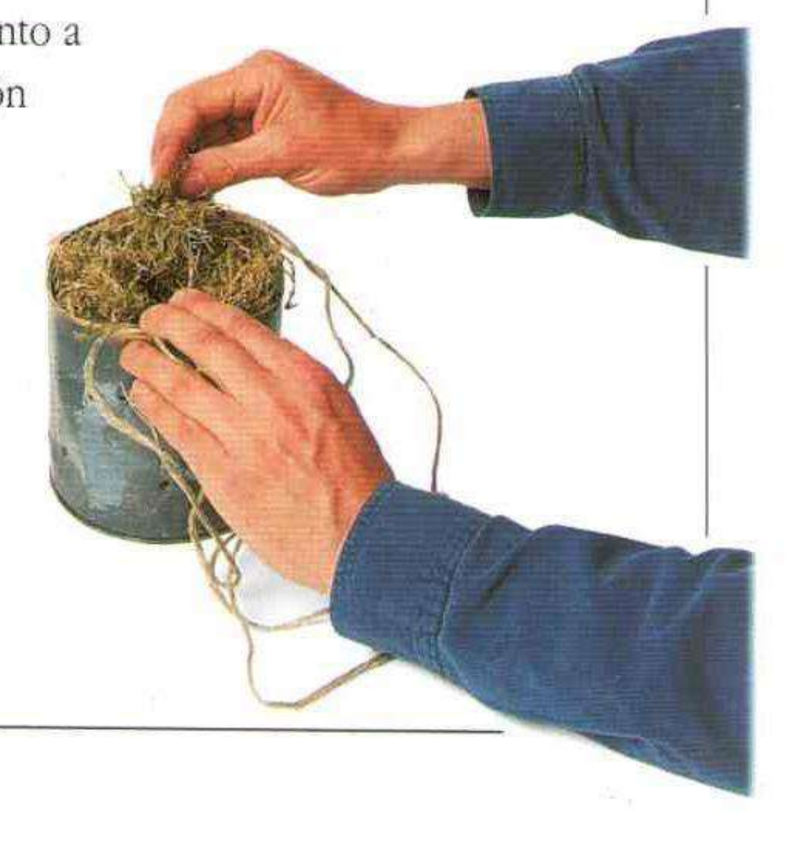

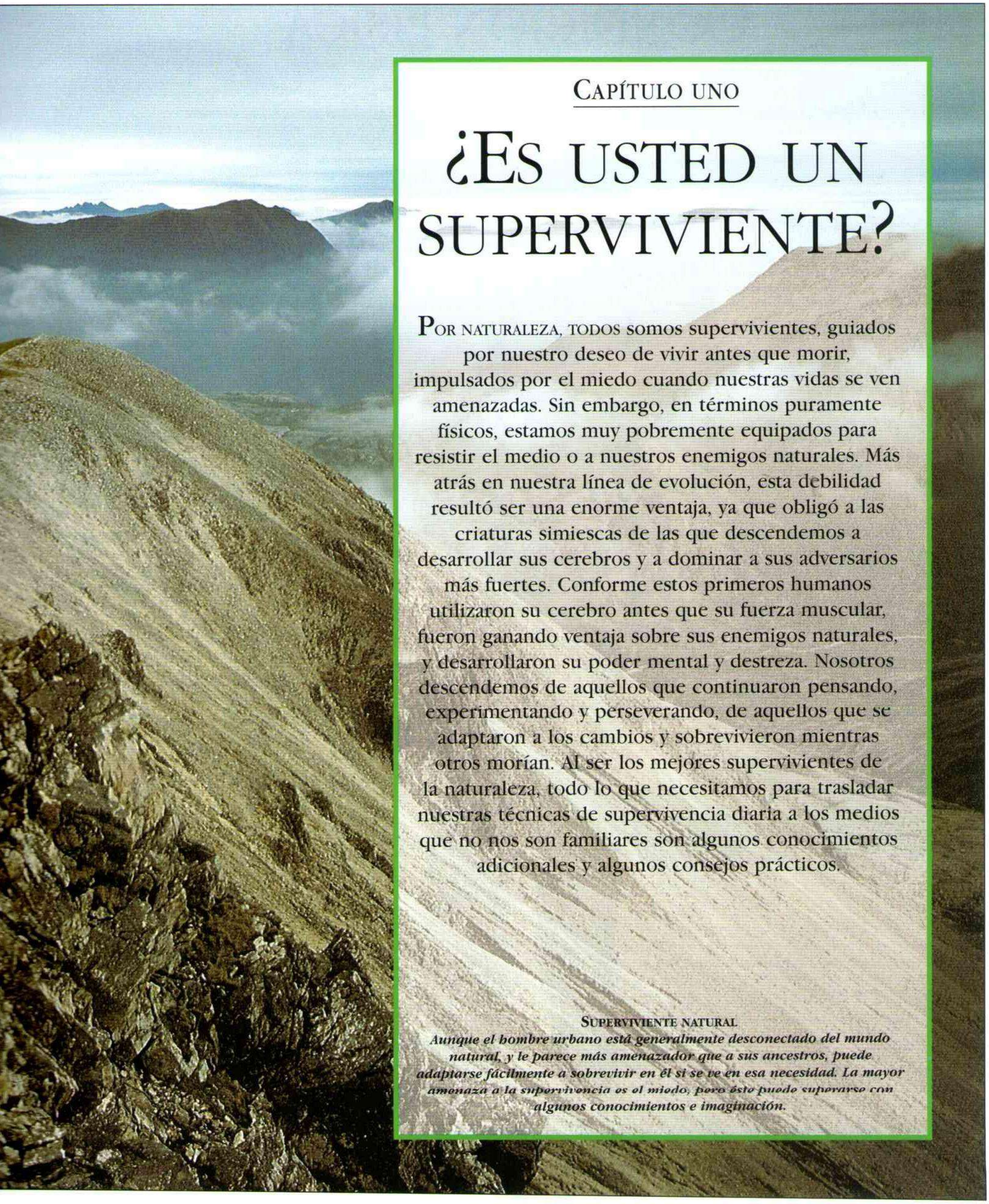

CAPÍTULO UNO

¿ES USTED UN SUPERVIVIENTE?

POR NATURALEZA, TODOS somos supervivientes, guiados por nuestro deseo de vivir antes que morir, impulsados por el miedo cuando nuestras vidas se ven amenazadas. Sin embargo, en términos puramente físicos, estamos muy pobremente equipados para resistir el medio o a nuestros enemigos naturales. Más atrás en nuestra línea de evolución, esta debilidad resultó ser una enorme ventaja, ya que obligó a las criaturas simiescas de las que descendemos a desarrollar sus cerebros y a dominar a sus adversarios más fuertes. Conforme estos primeros humanos utilizaron su cerebro antes que su fuerza muscular, fueron ganando ventaja sobre sus enemigos naturales, y desarrollaron su poder mental y destreza. Nosotros descendemos de aquellos que continuaron pensando, experimentando y perseverando, de aquellos que se adaptaron a los cambios y sobrevivieron mientras otros morían. Al ser los mejores supervivientes de la naturaleza, todo lo que necesitamos para trasladar nuestras técnicas de supervivencia diaria a los medios que no nos son familiares son algunos conocimientos adicionales y algunos consejos prácticos.

SUPERVIVIENTE NATURAL

Aunque el hombre urbano está generalmente desconectado del mundo natural, y le parece más amenazador que a sus ancestros, puede adaptarse fácilmente a sobrevivir en él si se ve en esa necesidad. La mayor amenaza a la supervivencia es el miedo, pero éste puede superarse con algunos conocimientos e imaginación.

CONDICIÓN FÍSICA

EL HOMBRE ES físicamente inferior a muchos otros animales, si no a la mayoría. Sin embargo, su gran ventaja es la falta de especialización. Tiene la inteligencia y la capacidad de desarrollar nuevas habilidades y de adaptarse a nuevas situaciones. En un medio urbano, el hombre no utiliza su cuerpo como lo haría en la naturaleza. El hombre primitivo es más dueño de su propio destino que su equivalente urbano.

EL CUERPO HUMANO

Olfato. *El hombre tiene un buen sentido del olfato, aunque no está tan desarrollado como el de otros animales, como el del perro. En la naturaleza, el olfato es útil para la identificación de alimentos, para advertir la presencia de otras criaturas y para prevenir peligros como los fuegos forestales.*

Oído. *Los animales salvajes están acostumbrados a identificar los peligros por su sonido. Sin embargo, el hombre urbano vive en un medio ruidoso que reduce la sensibilidad de sus oídos y su deseo de escuchar con atención.*

Gusto. *Las papilas gustativas en la lengua permiten al hombre la percepción de sabores agrios, dulces, salados y amargos. El cerebro interpreta la información recibida a través de las papilas y del olor del alimento para crear un sabor particular.*

Ritmo respiratorio. *El ritmo respiratorio del hombre depende de la cantidad de oxígeno que el cerebro cree que necesita el cuerpo. Este ritmo aumenta al realizar esfuerzos o al experimentar temor, y en las grandes altitudes, donde el aire contiene una menor cantidad de oxígeno.*

Columna vertebral. *El hombre aún está diseñado para caminar sobre cuatro patas, por lo que el dolor de espalda es el precio de permanecer erguido sobre dos piernas.*

Pulgares opuestos. *El hombre tiene pulgares y dedos que puede mover independientemente para coger los objetos. Esto le permite confeccionar y utilizar herramientas complejas, una habilidad que contribuye a su gran adaptabilidad.*

Articulaciones. *Algunas articulaciones, como las rodillas, son débiles y propensas a los accidentes. En la vida urbana, los humanos no las ejercitan lo suficiente. Éste es uno de los puntos más débiles de su diseño.*

Pies. *Como todo el peso del cuerpo descansa sobre los pies, las lesiones y otros problemas de pies son muy comunes, en especial en terrenos accidentados.*

Cerebro. *El cerebro grande y complejo del hombre compensa con creces la falta de adecuación de sus otros sentidos. Gracias a él, puede pensar con lógica, tener el sentido de la historia, una concepción del futuro y la capacidad de reflexionar sobre su lugar en el universo.*

Vista. *El hombre tiene un campo visual de unos 210 grados, de los cuales unos 120 grados se sobreponen, creando la visión estereoscópica. Esto le permite ver los objetos en tres dimensiones y apreciar las distancias. También puede distinguir los colores a la luz del día.*

Dientes. *Los dientes del hombre sirven tanto para cortar carne como para masticar vegetales. Así pues, los dientes están adaptados a una dieta omnívora.*

Corazón. *El corazón mantiene a los demás órganos funcionando al bombearles sangre rica en oxígeno. Sin un ejercicio regular, el corazón pierde su condición y deteriora el funcionamiento del resto del cuerpo.*

Piel. *La piel no sólo protege los órganos internos, sino que también controla la temperatura corporal gracias al sudor. Al haber perdido a lo largo de los siglos la mayor parte de su vello protector, la piel humana debe protegerse del sol, de la lluvia y del frío.*

El animal humano. *El hombre no puede equipararse a ningún otro animal. En la naturaleza, puede convertirse en presa de cualquier depredador, ya que es incapaz de defenderse con zarpas, garras, dientes o picos, o de protegerse con un caparazón, o de esconderse. Carece de la vista y del olfato de otros cazadores, así como de la velocidad suficiente para alcanzar a sus presas. En algunos ambientes, no puede mantener su temperatura corporal sin ropas especiales. Sin embargo, el hombre es un superviviente nato que, debido a sus debilidades físicas, se ha visto forzado a desarrollar su inteligencia y destreza —aptitudes fundamentales para sobrevivir.*

La adaptación de algunos animales

Lirón. *Pequeño y nervioso, con un metabolismo acorde, extremadamente rápido, el lirón hiberna durante el invierno cuando hay muy poco o ningún alimento disponible. Durante la hibernación, sus ritmos respiratorio y cardíaco disminuyen, y su cuerpo se alimenta de los depósitos de grasa que ha acumulado antes de su largo sueño.*

Ballena. *Aunque la ballena es un mamífero que necesita respirar aire, está bien adaptada a su medio marino: puede retener la respiración hasta 20 minutos para sumergirse en las profundidades en busca de alimento. Sin embargo, el estar tan bien adaptada es también una limitación, ya que es incapaz de escapar de ese medio.*

Zorro del Sáhara. *Las grandes orejas del zorro del Sáhara pueden girar para localizar con exactitud el menor de los sonidos, permitiéndole cazar de día o de noche. Su gran área superficial le permite disipar el excesivo calor corporal en su medio ambiente natural, el desierto.*

Musaraña. *Gracias a su pequeño tamaño y a su rapidez, la musaraña tiene una elevada actividad metabólica. Sin embargo, debe comer casi continuamente para permanecer viva.*

El efecto del medio ambiente

La naturaleza no es un jardín del Edén. Incluso las enfermedades simples que un hombre urbano considera normales, como la caries, la falta de higiene o las infecciones ordinarias, son potencialmente mortales. La vida urbana tampoco es mucho mejor, ya que a menudo crea tantos problemas de salud como los que resuelve. Éstos oscilan desde los dolores de cabeza causados por un ruido excesivo hasta el cáncer producido por un envenenamiento por radiación.

- Un bajo contenido de oxígeno en las grandes altitudes puede dificultar la respiración, y agravar los trastornos respiratorios como el enfisema.
- El humo del tabaco y la contaminación del aire pueden provocar cáncer de pulmón y otros trastornos respiratorios como el asma.
- El agua residual y las aguas contaminadas con organismos causantes de enfermedades pueden ser responsables de problemas digestivos.
- En los casos extremos, las sustancias químicas en los alimentos, así como los pesticidas en los productos agrícolas, pueden causar alergias, cáncer, daño a los ovarios, al hígado, a los testículos e hiperactividad.
- Los rayos ultravioleta del sol pueden causar cáncer y envejecimiento de la piel.
- El polvo y el polen en el aire son responsables de alergias como la fiebre del heno y de otros problemas respiratorios como el asma, la fibrosis y la neumonía.
- Los niveles extremos de ruido pueden causar sordera y zumbido, así como dolores de cabeza, estrés e irritabilidad.
- Los solventes industriales en contacto con la piel, o la inhalación de sus vapores, pueden causar reacciones alérgicas o dañar el hígado y los riñones.
- La radiación proveniente de fugas de máquinas o plantas nucleares puede causar daños en los espermatozoides y aumentar el riesgo de nacimiento de niños con leucemia y otros defectos congénitos.
- El estrés, resultado de una acelerada vida urbana, puede causar una serie de problemas físicos y mentales como la colitis y la úlcera gastroduodenal.

Cercano al hombre

Además de a los monos, el hombre es fisiológicamente cercano al cerdo. A semejanza del hombre, el cerdo puede alimentarse a base de carne y vegetales, y sus órganos y fisiología general son comparables a los del cuerpo humano, aunque tiene mayor grasa corporal y su piel es más gruesa que la del hombre. En el pasado, los caníbales de las islas del Pacífico hablaban de comer «cerdo largo», en referencia, quizá, a la similitud entre la carne humana y la de cerdo.

Cerdo doméstico. *El cerdo tiene una fisiología muy similar a la del hombre.*

La actitud mental

Sea cual sea la seriedad de una situación y la capacidad física de sobreponerse a ella de cada individuo, la actitud mental es la clave de la supervivencia. El hombre es el gran superviviente, y utiliza su cerebro para mantenerse por delante de sus competidores animales. La vida urbana sólo le obliga a utilizar una pequeña parte de sus atributos físicos y mentales. En una situación de supervivencia en la naturaleza, los hombres tienen que aprender de nuevo muchas habilidades manuales. El sentido común es una de las características humanas más valiosas, pero el deseo de sobrevivir en cualquier situación es mucho más importante.

NECESIDADES NUTRICIONALES

LAS CARACTERÍSTICAS FÍSICAS de los animales están determinadas por los alimentos que comen. Por ejemplo, los herbívoros tienen estómagos múltiples para digerir las resistentes hierbas, así como molares para pastar; los carnívoros tienen dientes para desgarrar carne, y cuerpos adaptados para moverse sigilosamente o desarrollar grandes velocidades. Sin embargo, el hombre es omnívoro, lo que significa que puede comer carne o prescindir de ella, y tiene dientes tanto para desgarrar carne como para triturar vegetales. Si sigue una dieta vegetariana, debe asegurarse de que le proporciona toda la gama de vitaminas y minerales que obtendría de cualquier otro animal.

LOS CIMIENTOS DE LA VIDA

Minerales. *Los minerales son necesarios para una gran variedad de funciones, que van desde el control del metabolismo hasta la osificación y la contribución al cierre de las heridas. Una deficiencia de minerales puede ocasionar calambres y convulsiones. Si no come carne, es vital que siga una dieta variada que incluya vegetales verdes, nueces y fruta para obtener todos los minerales, principalmente el hierro. Las setas también contienen minerales, así como las algas.*

Grasas. *Las grasas son una valiosa fuente de energía y son necesarias para generar y reparar las células del cuerpo. También contribuyen a la sensación de saciedad después de comer. Recuerde que, a fin de ser digeridas, las grasas requieren una gran cantidad de agua, y sólo deben consumirse cuando hay abundante agua disponible. Las nueces son ricas en grasas y en proteínas.*

Fibra. *La fibra facilita el funcionamiento correcto del intestino y la evacuación de residuos. Se obtiene de las frutas, las verduras y el pan.*

Hidratos de carbono. *Los hidratos de carbono desempeñan un papel vital en la nutrición al aportar energía al cuerpo. Se encuentran en forma de azúcares o almidones que se degradan y almacenan en el hígado hasta que son necesarios para liberar energía en forma de glucosa. El exceso de glucosa que no es requerido para aportar se convierte en grasa. La miel es excelente fuente de azúcar.*

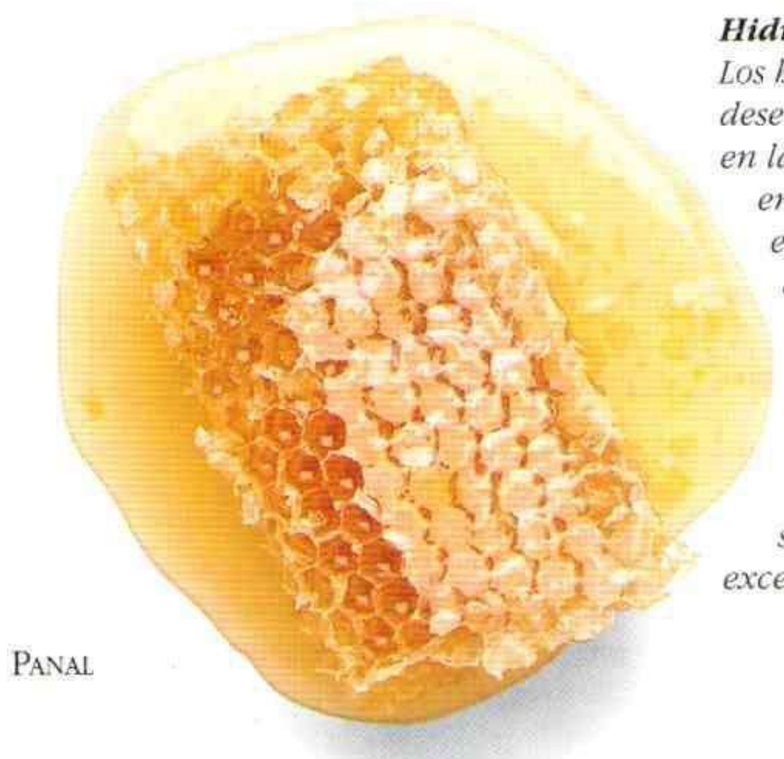

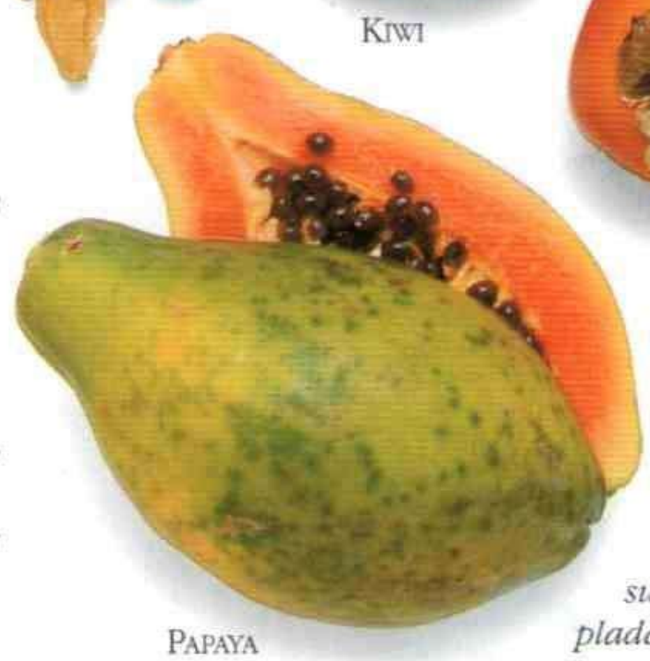

Vitaminas. *Las vitaminas controlan el crecimiento y el mantenimiento de los tejidos del cuerpo, y estimulan la producción de energía. Las frutas frescas son la mejor fuente de vitaminas. Durante las situaciones de supervivencia en regiones templadas, la fruta debe recolectarse y almacenarse para el invierno.*

Proteínas. *Las proteínas son imprescindibles en la formación de tejidos y para mantener el crecimiento. Algunas proteínas deben obtenerse de los alimentos, ya que proporcionan aminoácidos esenciales que nuestro cuerpo no puede sintetizar. La carne y el pescado proporcionan todos los aminoácidos necesarios y, adicionalmente, vitaminas y minerales. Algunas proteínas pueden obtenerse de productos vegetales.*

CALORÍAS NECESARIAS AL DÍA

La energía obtenida de los alimentos se mide en calorías. El hombre necesita un determinado número de calorías para subsistir. Incluso si permaneciera en la cama todo el día, quemaría cerca de 2.000 calorías. El número de calorías necesarias diariamente depende de varios factores, como la edad. Los supervivientes deben intentar equilibrar la entrada de calorías con el consumo de energía.

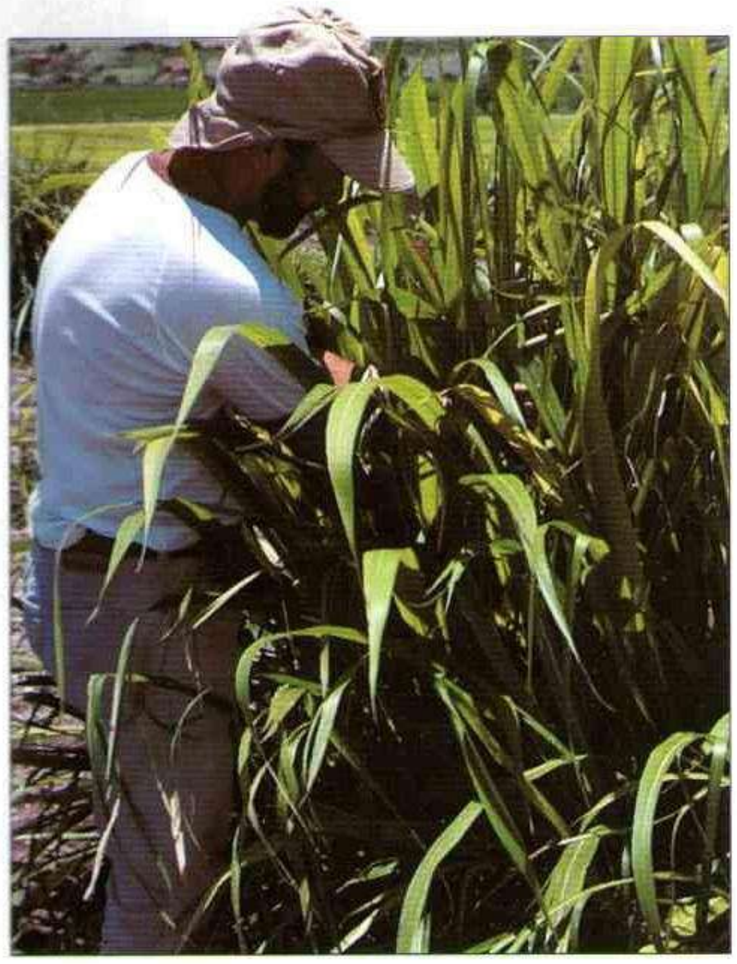

Científico tropical. *Las personas que viven en los trópicos requieren menos calorías que aquellas que viven en climas fríos. Las altas temperaturas provocan un sudor continuo, que produce una pérdida de minerales y sales, por lo que la dieta debe estar equilibrada.*

Leñador. *Una actividad física pesada, como la tala, requiere un mayor aporte diario de calorías que un estilo de vida sedentario, ya que se necesita más energía para mantener el cuerpo en movimiento y los músculos en uso. En las explotaciones forestales, estas calorías se obtienen gracias a comidas ricas y abundantes.*

Alpinista. *Las constantes temperaturas de congelación y la necesidad de estar alerta continuamente, convierten al montañismo en una actividad de gran consumo calórico. Son necesarias más de 5.000 calorías por día para que el cuerpo mantenga su temperatura interna. Pueden requerirse calorías adicionales para otras actividades como tirar de un trineo.*

CONSUMO DE ENERGÍA

Cada actividad requiere un distinto número de calorías, según el clima, la condición física y la cantidad de esfuerzo necesario para cada tarea en particular. Las siguientes cifras son una guía general para un hombre y una mujer «de actividad media». Debe recordar que sus necesidades calóricas particulares, y la cantidad de energía que requiere para cada actividad variarán según el clima y el tipo de terreno en el que trabaje.

Actividad	Calorías por hora
Descansar	70
Estar sentado	90
Encender un fuego	135
Caminar	180
Ir en bicicleta	240
Cortar madera	360
Esprintar	360
Correr	400
Nadar	500

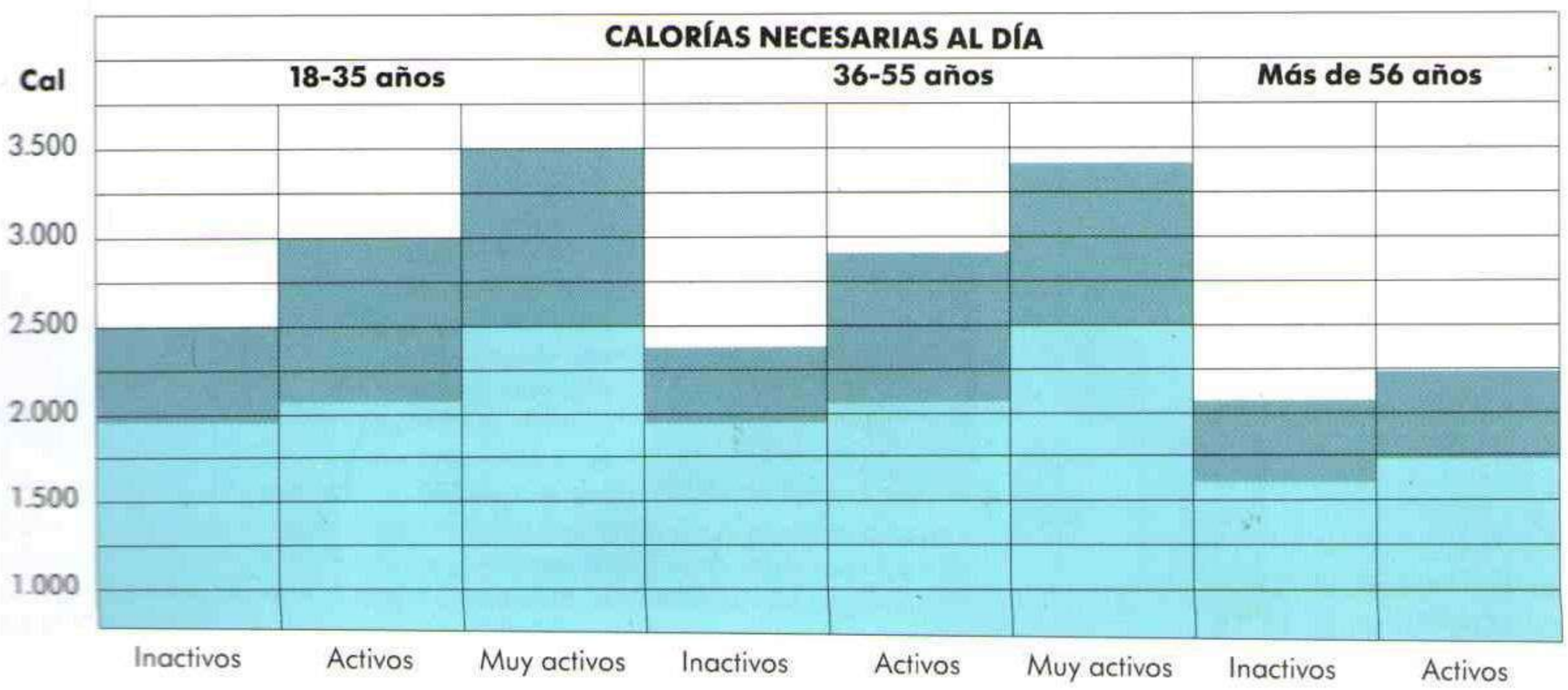

Necesidades calóricas. *El metabolismo de los hombres es menos eficiente que el de las mujeres; los hombres tienen menos aislantes naturales, por lo cual queman más calorías. Las personas jóvenes queman más calorías que las mayores, en parte porque son más activas, pero también porque la madurez tiende a disminuir el ritmo metabólico.*

HOMBRES

MUJERES

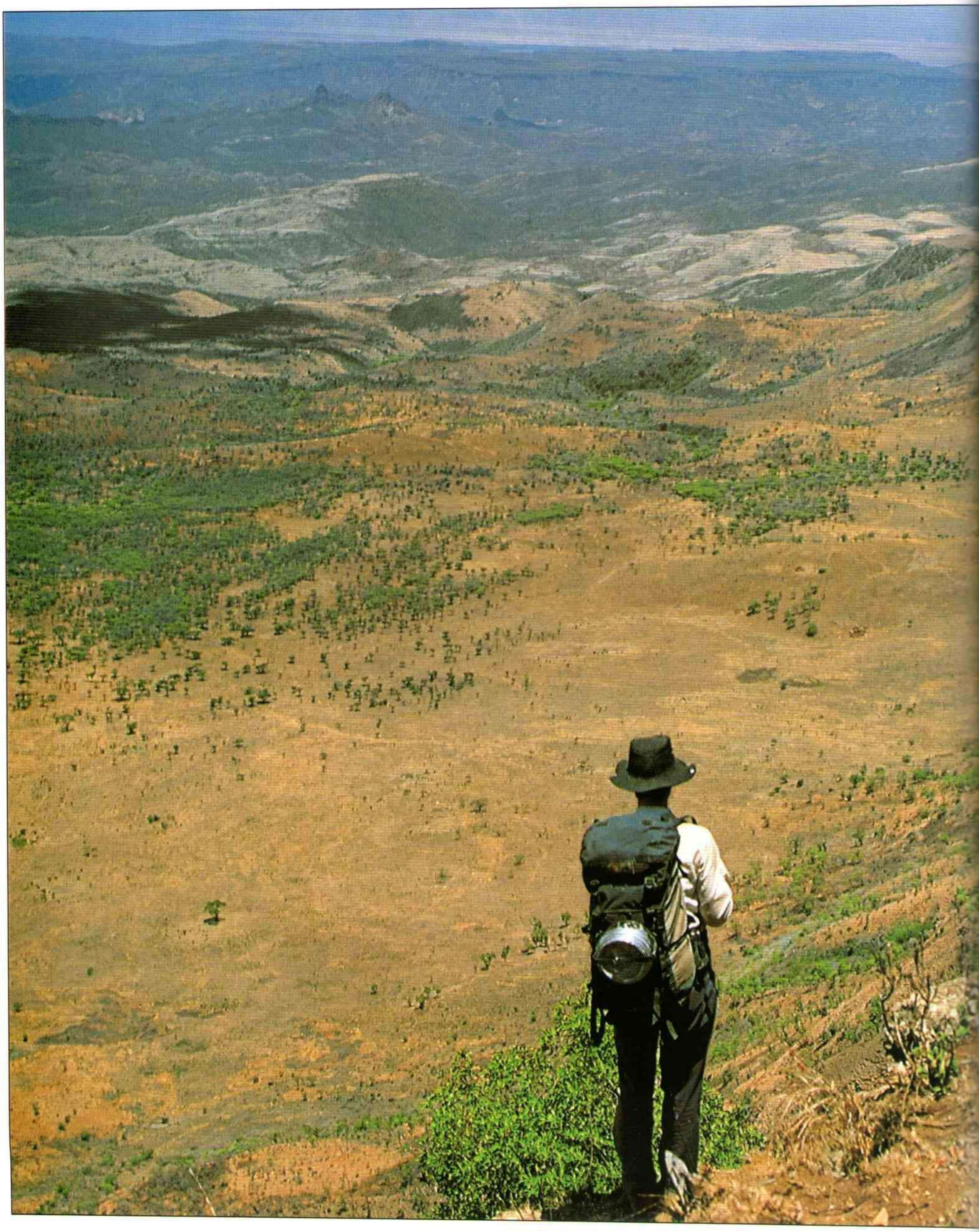

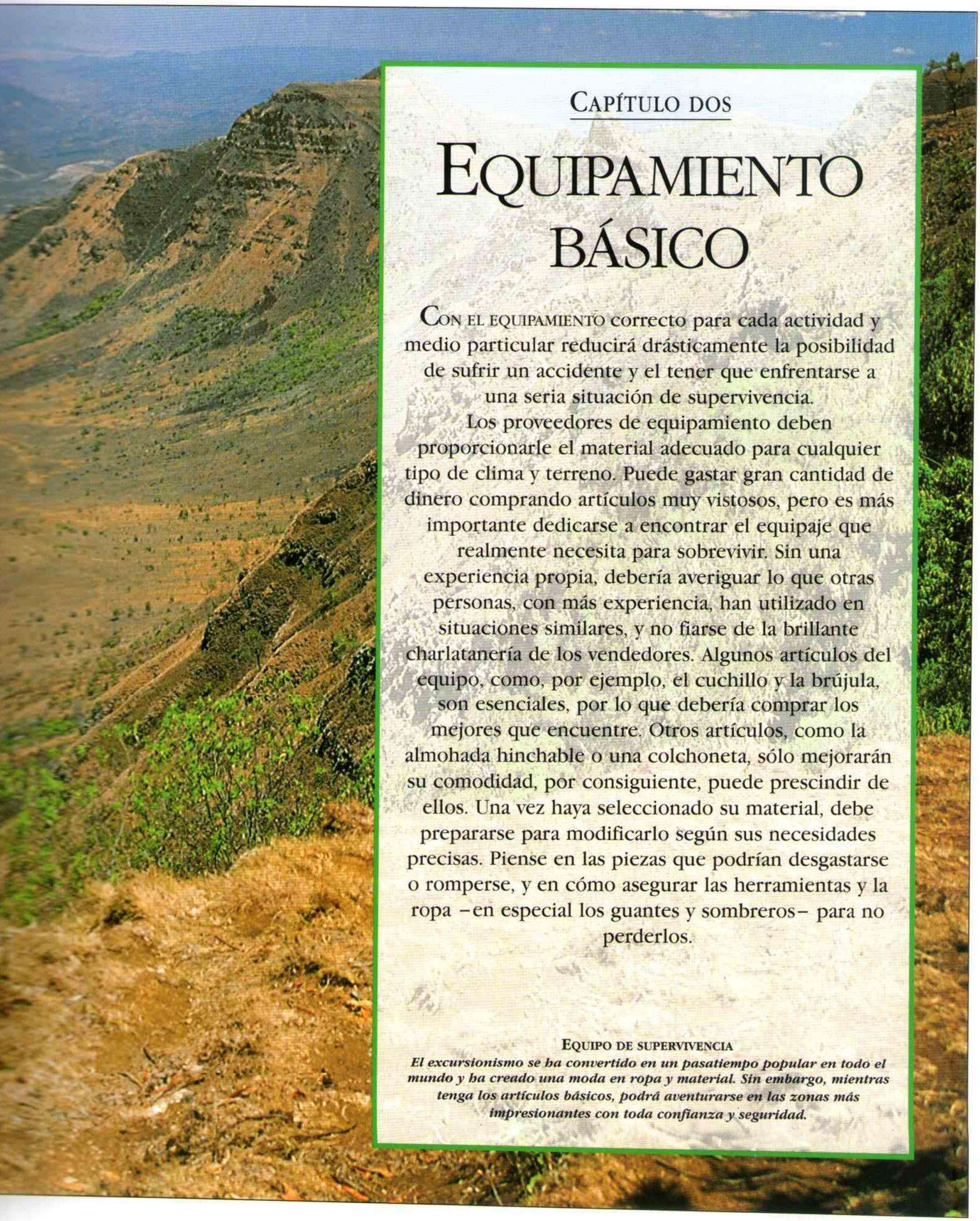

CAPÍTULO DOS

EQUIPAMIENTO BÁSICO

CON EL EQUIPAMIENTO correcto para cada actividad y medio particular reducirá drásticamente la posibilidad de sufrir un accidente y el tener que enfrentarse a una seria situación de supervivencia.

Los proveedores de equipamiento deben proporcionarle el material adecuado para cualquier tipo de clima y terreno. Puede gastar gran cantidad de dinero comprando artículos muy vistosos, pero es más importante dedicarse a encontrar el equipaje que realmente necesita para sobrevivir. Sin una experiencia propia, debería averiguar lo que otras personas, con más experiencia, han utilizado en situaciones similares, y no fiarse de la brillante charlatanería de los vendedores. Algunos artículos del equipo, como, por ejemplo, el cuchillo y la brújula, son esenciales, por lo que debería comprar los mejores que encuentre. Otros artículos, como la almohada hinchable o una colchoneta, sólo mejorarán su comodidad, por consiguiente, puede prescindir de ellos. Una vez haya seleccionado su material, debe prepararse para modificarlo según sus necesidades precisas. Piense en las piezas que podrían desgastarse o romperse, y en cómo asegurar las herramientas y la ropa –en especial los guantes y sombreros– para no perderlos.

EQUIPO DE SUPERVIVENCIA

El excursionismo se ha convertido en un pasatiempo popular en todo el mundo y ha creado una moda en ropa y material. Sin embargo, mientras tenga los artículos básicos, podrá aventurarse en las zonas más impresionantes con toda confianza y seguridad.

LA ROPA ADECUADA

RARA VEZ LLEVAMOS la ropa apropiada a las condiciones naturales en las que vivimos. La vida urbana nos acostumbra a llevar ropa para estar cómodos y a la moda, y no para mantener la temperatura corporal, y además nos aísla del medio con la calefacción artificial y el aire acondicionado. Incluso si compramos ropa diseñada para la vida exterior, ésta es más moderna que práctica. Puede gastar mucho dinero en conjuntos «de diseño» o comprar los artículos más esenciales, que cumplirán bien su función, aunque no sean del color de última moda.

EL PRINCIPIO DEL SISTEMA DE CAPAS

Un cierto número de capas de tejido mantendrán mejor su temperatura corporal que una sola prenda gruesa. Si tiene demasiado calor, puede controlar su temperatura corporal eliminando capas o «aireándolas» (abriendo cremalleras o botones para dejar escapar el aire caliente y permitir la entrada de aire fresco). Este principio se aplica tanto en los climas cálidos como en los fríos.

La capa interior. *La primera capa, en contacto con la piel, debe consistir en una camiseta de algodón o una camiseta térmica de manga larga, que se ajuste bien al cuerpo, pero sin que lo apriete. Debe estar hecha de un material que absorba la transpiración y que no se pegue a la piel (llevándola fuera del material). Esta capa debe mantenerse lo más limpia posible para evitar una acumulación de suciedad que llegaría a tapar sus poros.*

La segunda capa. *La segunda capa debe ir más suelta, pero con la posibilidad de proteger y mantener el calor en el cuello y las muñecas. Puede consistir en una camiseta con cuello tipo polo y cremallera, o en una camisa con cuello, mangas que puedan remangarse y puños con botones. En un clima cálido, esta capa puede ser la exterior, si bien pueda requerir una protección para el viento* (véase la página siguiente).

Calzoncillos. *Únicamente requerirá ropa interior larga y térmica en zonas con temperaturas bajo cero. En las regiones árticas, un refuerzo impermeable en la ingle protegerá del viento esa parte del cuerpo, y le será útil si tiene que esquiar. Un pantalón mojado se secará más deprisa si no lleva ropa interior larga, aunque puede combatir este problema llevando pantalones impermeables. En climas templados, use unos calzoncillos cortos de algodón.*

La tercera capa. *La tercera capa debería consistir en un jersey de lana o una chaqueta ligera de fibra polar. Si está en movimiento, incluso en el ártico, ésta es la capa que debe quitarse para evitar un acaloramiento excesivo. Puede airear su plumífero* (véase página siguiente) *si aún tiene calor. Cuando se detenga a descansar, vuelva a ponerse la capa intermedia antes de empezar a sentir frío. Esta capa puede sustituir a la exterior en regiones templadas cuando el clima es suave.*

La capa exterior. *La capa exterior debe ser una chaqueta impermeable o que le proteja del viento, o que cumpla ambas funciones, dependiendo del clima en el que se desenvuelva. En las regiones árticas requerirá un anorak acolchado que le proteja de los fuertes vientos. Debe tener la posibilidad de airear la chaqueta para evitar el exceso de calor. En áreas templadas el principal problema es la lluvia; puede llevar entonces una prenda impermeable sobre la chaqueta.*

Pantalones. *Los pantalones deben permitir la libertad de movimientos y deben estar hechos de un material que se seque rápidamente cuando se ha mojado. En condiciones muy húmedas, puede ponerse pantalones impermeables para proteger sus piernas de una lluvia intensa* (véase página siguiente), *aunque pueden causarle un calor excesivo. En condiciones de frío intenso deben utilizarse pantalones impermeables acolchados por encima de los pantalones y las botas, como protección adicional.*

CARACTERÍSTICAS DE LOS MATERIALES TEXTILES

Material	Ventajas	Inconvenientes
Lana	Esta fibra natural tiene propiedades aislantes incluso cuando está húmeda, manteniéndole cómodo a menos que esté totalmente mojada. Arde sin llama cuando se expone a una llama.	La lana es pesada cuando está mojada y tarda en secarse. Puede dar sensación de picor en contacto directo con la piel. Puede encogerse al lavar.
Algodón	El algodón es resistente, permite la «respiración» y absorbe la humedad. Es un material especialmente indicado para la ropa interior y para todo tipo de ropa que deba llevarse en contacto con la piel.	Es pesado cuando está mojado y puede encogerse si se seca a altas temperaturas. Puede rasgarse con facilidad. No corta el viento y arde fácilmente.
Fibra polar	Se utiliza como tercera capa para alejar la humedad del cuerpo al mismo tiempo que lo mantiene caliente. Pesa poco y es resistente. No absorbe la humedad.	Este tejido sintético no corta el viento. No se comprime fácilmente y puede producir «bolitas» en la parte exterior después de un uso prolongado.
Mat. sintéticos y transpirables	Permiten que el sudor se evapore al mismo tiempo que aíslan de la lluvia. Generalmente protegen del viento, y se utilizan para las prendas exteriores.	Las costuras en estos tejidos permiten el paso del agua. En condiciones muy húmedas, los poros del tejido pueden taparse, y en climas muy fríos la humedad puede condensarse en el interior. La evaporación del sudor puede dar como resultado una pérdida de calor.

La elección del material. *La ropa puede fabricarse con una gran variedad de materiales, desde la lana o el algodón hasta los tejidos sintéticos, algunos de los cuales permiten la transpiración, mientras que aíslan de la lluvia. Los materiales sintéticos no siempre son tan resistentes como los naturales. Las fibras naturales como el algodón y la lana pueden encogerse si se secan a temperaturas elevadas después de lavadas.*

ROPA PARA CLIMAS EXTREMOS

Si al elegir su ropa sigue el sistema de capas *(véase página anterior)* debe ser capaz de equiparse con las prendas adecuadas para cualquier tipo de clima. En una situación de supervivencia, puede que no disponga de las prendas fabricadas para un clima específico, pero puede utilizar el sistema de capas en combinación con refugios y fuegos para protegerse de los elementos adversos. En los climas extremos y en terreno hostil, la ropa debe protegerle de aquellas circunstancias que puedan dañarle o causarle la muerte –por ejemplo las que se encuentran en los desiertos *(véase página 142)*. El viento, el frío, el calor, el sol y la lluvia constituyen la principal amenaza. Algunas prendas que son específicas para un clima determinado –por ejemplo las botas para la selva– no suelen ser adecuadas para áreas menos extremas, pero, en general, la mayor parte de las prendas pueden adaptarse a sus necesidades.

Sombrero. *Un sombrero de ala ancha protege la cabeza y el cuello del sol. Los pequeños orificios de ventilación alrededor del ala contribuyen a mantener la cabeza fresca.*

Capa base. *Una camiseta de algodón de color claro debajo de la camisa absorbe el sudor de la piel y le mantiene fresco.*

Segunda capa. *Una camisa ligera actúa como capa principal. Las mangas deben llevarse bajas en situaciones de calor extremo, por ejemplo en el desierto* (véase página 142).

Capa exterior. *Una chaqueta ligera y aislante sobre la camisa actuará como capa final, protegiéndole del viento y del frío nocturno.*

Mantenerse fresco. *La ropa para el clima cálido debe ser ligera y suelta. Los colores claros reflejan el calor y contribuyen a mantenerle fresco. El principio de las capas también se aplica aquí, de la misma manera que en el clima frío.*

Pantalones. *Los pantalones no deben ser ajustados ni pesados, sino de un material ligero, como el algodón.*

Botas. *Las botas deben ser ligeras, aunque con una suela resistente para caminar* (véase página 24).

Protección para la cabeza. *Un pasamontañas cubre la cabeza, los laterales de la cara y el cuello, impidiendo la pérdida de calor. Puede llevar también una capucha aislante que le proporcione calor adicional.*

Capa base. *Una camiseta térmica y unos calzoncillos largos constituyen la primera capa. Deben absorber y eliminar la transpiración.*

Manopla intermedia. *Las manoplas de lana entre los guantes interiores y las manoplas exteriores le permiten coger los objetos sin perder el calor de las manos.*

Capa intermedia. *Un jersey de fibra polar o lana debería absorber el sudor de la capa base y mantener una capa de aire caliente alrededor de su cuerpo.*

Segunda capa. *Una camisa con cuello tipo polo debe cubrir el pasamontañas a la altura del cuello, y los guantes a la altura de los puños.*

Guantes interiores. *Evitan que la piel se adhiera a los objetos congelados, y permiten el uso de las manos durante periodos cortos sin necesidad de usar manoplas gruesas.*

Capa exterior. *Un anorak acolchado y con capucha es esencial. Debe tener un cierre frontal solapado y una capa impermeable en tejido transpirable.*

Pantalones. *Un pantalón de peto cubre la zona de la cintura, donde la ropa puede causar molestias al caminar; permite, al mismo tiempo, la aireación en hombros y pecho.*

Botas. *Las botas de nieve tienen una coraza plástica y forros térmicos que actúan como un par de botas internas. Deben llevarse dos pares de calcetines debajo de ellas.*

Manoplas externas. *El sistema de capas también se aplica a los guantes; es aconsejable llevar un par de manoplas gruesas encima de otras más finas.*

Mantenerse en calor. *En situaciones de frío extremo, deben cubrirse todas las partes del cuerpo expuestas; resulta vital el seguimiento del sistema de capas. Prepárese a airear o a quitarse alguna prenda si siente calor excesivo al moverse o trabajar. El sudor atrapado reduce las propiedades aislantes de la ropa y puede causar hipotermia* (véase página 163).

MANTENERSE SECO

Estar mojado constituye un grave problema de supervivencia. Debe ponerse el traje impermeable en el momento que empiece a llover, y quitárselo tan pronto cese la lluvia. El equipo de lluvia a base de materiales impermeables no es adecuado para los excursionistas ya que produce sudor –elija únicamente materiales transpirables. Utilice pantalones impermeables sólo cuando llueva y haga frío a la vez. Unas polainas mantendrán secos sus pies y pantorrillas, incluso al cruzar ríos.

Equipo para tiempo lluvioso. *El equipo impermeable debe cubrir todo el cuerpo, desde la cabeza hasta los pies.*

CALZADO

EL PESO DEL cuerpo humano está calculado para ser soportado por cuatro pies en lugar de dos, así que los pies están sometidos a una presión constante. Con el peso añadido que suponen las mochilas cargadas, cruzar un territorio escarpado o húmedo hace que la protección de los pies se convierta en la parte del equipo más importante a considerar. El calzado debe haber cedido antes de ser usado. Para ello, humedezca las botas de cuero y llévelas puestas hasta que se sequen. Una vez en camino, mantenga los pies limpios, lávelos cada día y aplíqueles polvos de talco. Use siempre calcetines limpios y secos, y cubra cualquier rozadura en los pies antes de que se conviertan en ampollas.

CALZADO LIGERO

***Botas de* trekking.** *Muchos montañeros llevan botas ligeras de marcha tanto en el campamento como durante la ascensión a una montaña elevada, y reservan sus pesadas botas de escalada para la nieve y el hielo. Las botas ligeras también pueden utilizarse para las caminatas sencillas y en distancias cortas, siempre y cuando el terreno no sea muy escarpado. Aunque ofrecen menos protección y apoyo que las botas de piel, las botas de marcha se secan rápidamente y dan a los pies una sensación de descanso después de llevar un calzado caluroso y pesado.*

Botas para el desierto. *Con suelas resistentes y fabricadas en cuero ligero, las botas para el desierto permiten que los pies respiren al mismo tiempo que evitan la entrada de arena caliente. Los laterales altos protegen los tobillos de la vegetación espinosa y les proporcionan apoyo. En condiciones húmedas, el cuero se vuelve todavía más poroso por lo que necesitan un tiempo prolongado para secarse.*

Zapatillas. *Una vez en el campamento, las zapatillas alivian los pies doloridos, y son excelentes para llevar en barcos o vehículos. Sin embargo, al ser cómodas y ligeras, no ofrecen tanta protección como las resistentes botas de lona. El uso constante de zapatillas puede restar resistencia a los pies y hacerlos más susceptibles a las heridas. Al caminar sobre un terreno húmedo pueden llegar a quedar absorbidas y perderlas.*

CALCETINES

Los calcetines gruesos, ya sean de lana o algodón, son indispensables. Los calcetines protegen el pie y evitan que las botas rocen la piel, al mismo tiempo que mantienen los pies cálidos y secos eliminando la humedad (trasladan la humedad hacia la capa exterior de las fibras, alejándola de la piel). En climas muy fríos es aconsejable llevar dos pares de calcetines.

CALCETINES INTERIORES CALCETINES EXTERIORES

CALZADO RESISTENTE

Botas de montaña. *Esta bota es un término medio entre peso, durabilidad y protección; indicada para todo tipo de actividades gracias a su suela gruesa y grabada, material impermeable y protección para los tobillos.*

La zona de los tobillos está acolchada para proporcionar un apoyo cómodo

Suela con grabado para agarrarse en terreno resbaladizo

Botas de jungla. *En la jungla se hace necesario proteger el tobillo de ramas y rocas. Con suela de goma, estas botas evitan la absorción de agua, al tiempo que permiten que los pies respiren a través de la lona de la que están hechas. Después de vadear un río, el agua es expulsada a través de ojetes especiales en el empeine. El dibujo de la suela, ancho y profundo, proporciona un buen agarre al suelo húmedo.*

La puntera de goma aísla del agua, y la parte superior de tela proporciona frescor al pie

Botas de plástico para nieve. *Diseñadas para mantener el pie rígido al utilizar garfios para un mejor agarre al hielo, estas botas son especialmente aislantes para ser utilizadas en los climas más fríos. Sin embargo, al igual que sucede con las botas para esquiar, caminar con ellas resulta incómodo. Tienen un forro interno térmico con la forma de la bota, que puede llevarse independientemente, por ejemplo dentro de una tienda de campaña.*

Suela con grabado antideslizante para caminar en la nieve

Los laterales rígidos proporcionan un apoyo excelente en terrenos rocosos

POLAINAS

Las polainas resultan indispensables, tanto en las zonas templadas para mantener secos los pies y los bajos del pantalón como en las regiones polares para evitar el contacto con la nieve. Al cruzar ríos poco profundos, las polainas impedirán que el agua entre en las botas. Son mejores que las botas de goma, ya que no provocan calor ni una transpiración excesiva del pie.

POLAINAS PARA BARRO POLAINAS PARA NIEVE

CUIDE SUS BOTAS

1 Retire los cordones y plantillas interiores y lave las botas para eliminar el barro. Es muy importante quitar toda la turba, cuyo contenido ácido puede dañar la piel de las botas.

2 Deje secar las botas completamente. Si es posible, manténgalas alejadas de calor directo del fuego o del sol, ya que pueden agrietar la piel.

3 Una vez las botas estén secas, frótelas con el dedo con un compuesto impermeable. Si no tiene previsto ponerse las botas enseguida, guárdelas en un lugar fresco y seco.

EQUIPO PERSONAL DE CAMPAMENTO

ÚNICAMENTE REUNIRÁ el equipo perfecto de campamento a base de intentos y equivocaciones. Antes de embarcarse en un viaje, considere la posibilidad de omitir cualquiera de los elementos de su equipo que no utilizó en su viaje anterior. Las modificaciones en el equipo vendrán por sí solas. Algunos artículos serán inamovibles, y pronto mostrarán las señales de su uso continuo. Llegará un punto en el que se encontrará tan identificado con su equipo que no querrá ni considerar la posibilidad de perder o romper cualquiera de sus elementos.

El mapa se dobla abierto mostrando el área en uso

Mapa. *Un mapa siempre debe mantenerse seco y bien doblado para mostrar el área en uso.*

Brújula. *Una brújula es una pieza básica del equipo, por lo que se debe guardar en un sitio seguro.*

Gire la parte superior para encender la linterna

Linterna. *Una linterna debe ser lo más pequeña posible y resistente al agua.*

Prismáticos. *Deben ser lo más pequeños y resistentes posible. Incluso si son estancos, guárdelos en una bolsa de plástico.*

Navaja del ejército suizo. *Es imprescindible una navaja con una cuchilla buena y sólida. Las herramientas adicionales de la navaja son también muy útiles, aunque deben utilizarse con cuidado.*

Escudilla. *Una buena escudilla es muy útil para cocinar en el campamento* (véase página 112). *Coloque cinta aislante alrededor de las agarraderas metálicas para evitar quemaduras.*

Botella de agua. *Una botella de metal puede agrietarse si el agua que contiene se congela. Una de plástico puede fundirse si se coloca cerca del fuego.*

Depurador de agua. *Se utiliza para depurar el agua natural antes de beberla y así prevenir infecciones* (véase página 75).

Cocina. *Elija una cocina sencilla* (véase página 114), *y manténgala limpia, además de protegerla de los golpes.*

Combustible para la cocina. *Asegúrese de que no tenga fugas, y de que esté limpio, filtrándolo si es necesario.*

PLATO LLANO PLATO HONDO

Platos y recipientes. *Utilice platos y recipientes de plástico o madera para evitar las quemaduras en la boca, que son dolorosas y pueden infectarse* (véase página 112). *Nunca coma de un cazo caliente.*

Taza. *Una taza de plástico es útil para contener bebidas frías o calientes* (véase página 112).

CUCHARA FUNDA PARA CUBIERTOS

CUCHILLO TENEDOR

Cubiertos. *Mantenga limpios todos los utensilios para evitar intoxicaciones. Si pierde o rompe sus cubiertos, puede tallar unos nuevos de madera* (véase página 113).

Protector labial. *El protector labial es fundamental en todo tipo de climas, ya que evita que los labios se corten. Utilícelo en cualquier momento del día, cuando los labios se le resequen.*

Protector solar. *Una crema protectora solar es indispensable en la nieve y en los climas soleados.*

Kit de supervivencia. *Contiene artículos básicos que no se encuentran en la naturaleza* (véase página 28).

Toalla. *Una toalla grande puede cortarse en pedazos más pequeños para secarse el sudor o humedecerse la cara. También puede utilizarse como pañuelo* (véase página 142).

Botiquín de primeros auxilios. *Es fundamental para el tratamiento de las heridas* (véase página 174).

Cerillas impermeables. *Tienen cabezas enceradas que las mantienen secas.*

Equipo de costura. *Utilícelo para remendar la ropa y las tiendas.*

Útiles de aseo. *Guarde todos los artículos de aseo personal en una bolsa especial. Cuelgue la bolsa de un árbol mientras se lave para evitar que algún artículo se pierda* (véase página 57).

CALIENTE Y SECO

La tienda y el saco de dormir serán sus únicas comodidades cuando se encuentre en plena naturaleza, aunque puede sobrevivir sin ellas. Siempre deben guardarse convenientemente, y los sacos de dormir no deben mojarse nunca.

TIENDA DE CAMPAÑA

SACO DE DORMIR

EQUIPO BÁSICO DE SUPERVIVENCIA

EN SITUACIONES DE supervivencia, algunos artículos pueden llegar a constituir la diferencia entre la vida y la muerte. Lleve siempre sobre su persona los artículos de mayor importancia, preferiblemente dentro de una lata pequeña que pueda meter en un bolsillo. Siempre que se aventure en la naturaleza, asegúrese de llevar otros utensilios que pudieran ser vitales para su supervivencia individual y afianzarlos de tal manera que la posibilidad de perderlos o dañarlos sea mínima. Algunos de estos útiles pueden colgarse de cuerdas resistentes alrededor del cuello; métalos debajo de la ropa para que no puedan desprenderse al engancharse con la vegetación. Coloque cada uno de los artículos en una capa diferente de la ropa para que no choquen entre sí y se rompan. Debe reunir un equipo de supervivencia acorde con sus necesidades personales.

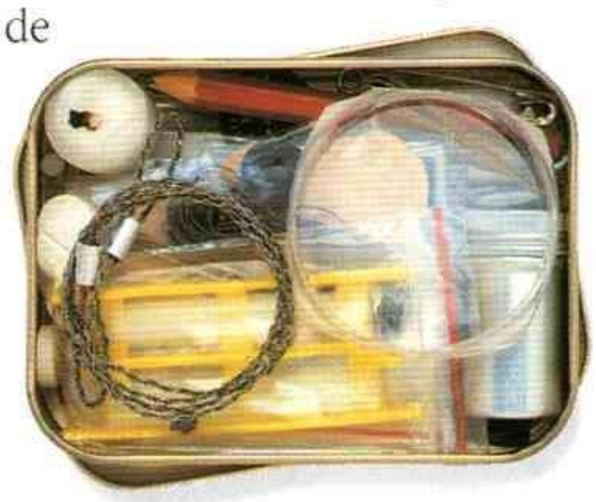

Equipo de supervivencia. *Los artículos más importantes de su equipo de supervivencia deben guardarse en una lata pequeña provista de una tapa que ajuste bien.*

Imperdibles. *Utilícelos para asegurar su ropa, o para arreglar su tienda o saco de dormir.*

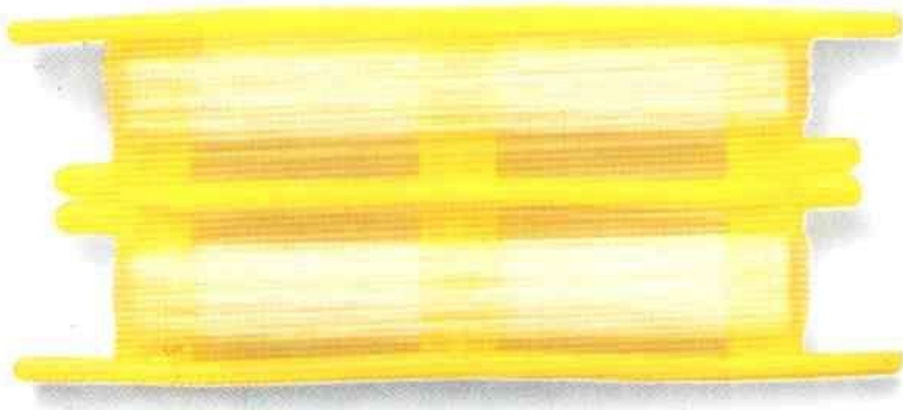

Hilo de pescar. *Lleve siempre consigo un hilo resistente para pescar o para atar herramientas.*

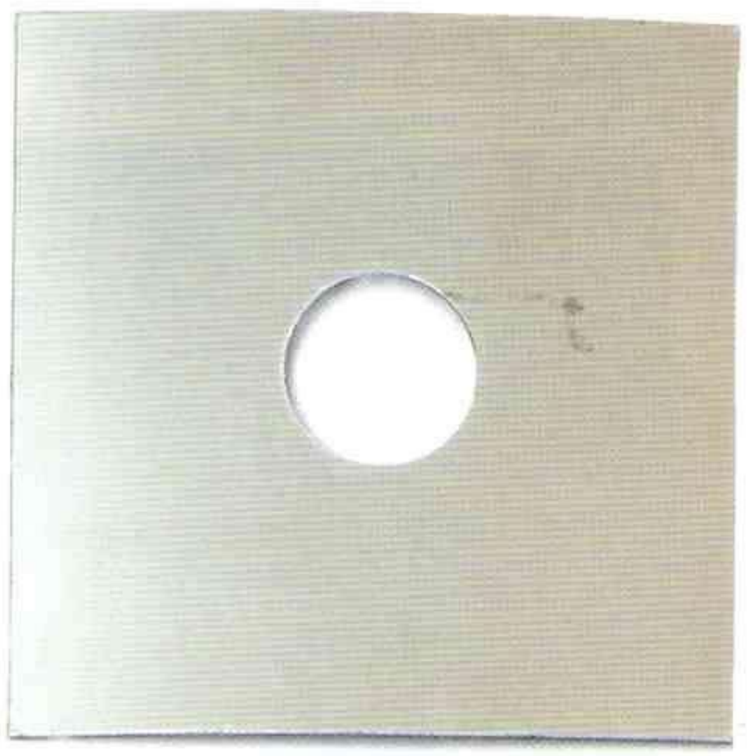

Reflector. *Puede utilizar la tapa brillante de la caja de su equipo de supervivencia, o bien un heliógrafo, para llamar la atención en una situación de supervivencia. Mirando a través del agujero en el centro del heliógrafo puede dirigir la luz reflejada hacia un avión.*

Alambre fino. *El alambre tiene muchas aplicaciones y puede reutilizarse varias veces. Debe ser lo suficientemente fino como para doblarse fácilmente, aunque no tanto como para que se rompa.*

Anzuelos y plomos. *Incluya una buena cantidad de anzuelos, corchos y plomos. Los anzuelos pequeños pueden utilizarse tanto para pescar peces grandes como pequeños.*

Lata de tabaco. *Debe guardar su equipo de supervivencia en una pequeña lata de tabaco provista de una tapa con buen ajuste. Selle la lata con cinta para mantener seco su contenido.*

Brújula. *Es esencial. La brújula ideal debería ser luminosa.*

Vela. *Utilícela para encender un fuego y no como lámpara. No utilice velas de sebo porque pueden pudrirse en climas cálidos.*

Cerillas. *Compre cerillas impermeables, o impermeabilícelas dejando caer una pequeña capa de cera de vela sobre sus cabezas y parte del cuerpo. Elimine la cera de cada cerilla antes de encenderla.*

Agujas, hilo y botones. *Incluya varias agujas en su equipo, todas ellas de ojo grande, así como una buena cantidad de hilo grueso, resistente y preferiblemente impermeable. Envuelva el hilo alrededor de las agujas y guárdelas en una bolsa de plástico. Los botones grandes son útiles para sujetar las pestañas de la tienda.*

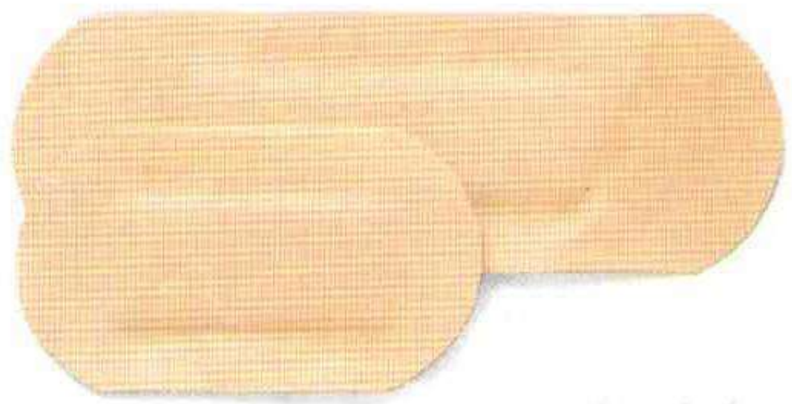

Tiritas adhesivas. *Debe incluir tiritas adhesivas de varios tamaños para evitar que alguna rozadura llegue a infectarse, o para proteger las ampollas del pie hasta que sanen.*

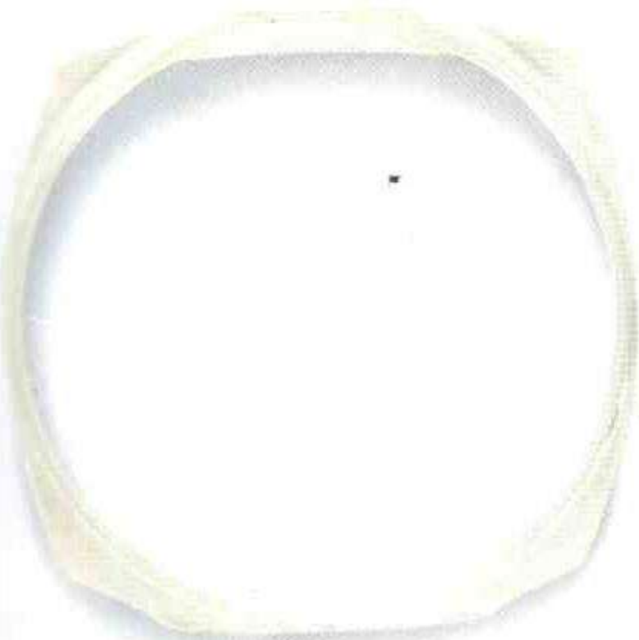

Lupa. *Una lupa puede hacer converger los rayos del sol en la madera seca, para hacerla arder y, de este modo, encender un fuego.*

Cable sierra. *Un cable sierra es de empleo fácil, aunque lento; con él podrá cortar la mayoría de los materiales. Guárdelo cubierto con una película de grasa dentro de una bolsa de plástico, para que no se oxide ni se rompa.*

Antibióticos. *No deben utilizarse hasta que sean verdaderamente necesarios.*

Pastillas esterilizadoras. *Guárdelas para una emergencia, cuando le sea imposible hervir agua que le parezca insalubre.*

Sal. *Es esencial para sobrevivir. Si no repone la sal perdida en el sudor y la orina, puede llegar a enfermar gravemente.*

Bolsa de plástico. *Una bolsa de plástico resistente tiene múltiples utilidades, como llevar agua desde un río, u obtenerla de la vegetación en una emergencia.*

Permanganato potásico. *Estos cristales son útiles para purificar el agua. Mezclados con agua obtienen un color rojo intenso; pueden utilizarse para el tratamiento de algunas infecciones por hongos, como el pie de atleta. Siga las instrucciones del envase.*

Bisturí. *Puede emplearse para muchos propósitos. Guárdela con la cuchilla engrasada en su envase original.*

Lápiz. *Un lápiz es de gran utilidad para anotar los sitios en los que ha estado, lo que es comestible, y para dibujar mapas.*

El equipo básico alrededor de su cuello

No debe prescindir nunca de las piezas más valiosas de su equipamiento, en particular si se adentra en tierra virgen. Debe tomar las precauciones necesarias para no perderlas o romperlas, ya que, alejado de la civilización, no podrán ser arregladas o reemplazadas. Ate cada uno de los artículos a un cordón diferente y cuélgueselos del cuello, de esta manera los tendrá siempre a mano cuando los necesite. Guárdelos entre su ropa cuando no los utilice para evitar que se rompan.

Equipo básico. *Manténgalo siempre al alcance de la mano.*

CAMAS Y SACOS DE DORMIR

EL DESCANSO ES tan importante para la supervivencia como el alimento. Un buen e ininterrumpido descanso nocturno puede aliviar todas las preocupaciones y tensiones y convertir la melancolía pesimista en un optimismo positivo. El estar abrigado y seco le proporciona una mayor sensación de seguridad. Incluso en el calor húmedo de la selva son necesarias la protección y el aislamiento de la humedad del suelo.

CARACTERÍSTICAS DE LOS SACOS DE DORMIR

Aunque hay muchas formas y estilos de sacos de dormir, un saco siempre debe estar lo suficientemente bien acolchado como para mantenerle abrigado incluso si carece de tienda. No debe permitir nunca que un saco de dormir se moje. Guárdelo siempre en una funda impermeable de un material poroso *(véase página 43)*.

Saco de extremo redondeado. *Diseño común para un saco de dormir moderno. La capucha evita que se escape el calor de la cabeza, el pecho y los hombros.*

Saco sin cremallera. *La abertura de este saco puede cerrarse con un cordel para mantener la cabeza y el cuello bien protegidos en climas fríos. Esto es muy importante, ya que puede perder la mitad de su calor corporal a través de la cabeza. Al carecer de cremallera, puede resultar difícil salir del saco rápidamente.*

Saco de forma rectangular. *Este saco es ideal para acampar si se dispone de un vehículo, y no tanto para dormir a la intemperie. Es muy ligero y, por tanto, no adecuado para climas muy fríos como las regiones polares.*

CONFORMACIÓN DE UN SACO DE DORMIR

Un saco de dormir relleno de plumón es ligero y proporciona abrigo. Sin embargo, el plumón puede perder su capacidad aislante al compactarse en bolas cuando se humedece o se almacena muy apretado. Los rellenos sintéticos son más abultados, pesados y carecen de comodidad. Sin embargo, le mantendrán abrigado aun si se mojan, y se secan rápidamente.

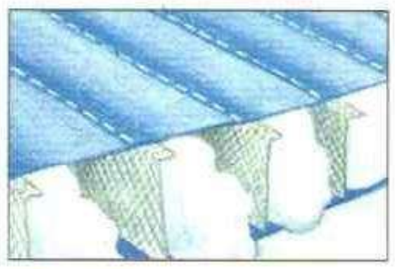

Tabiques. *En los sacos, el plumón se distribuye en «tabiques» de tela para reducir el apelotonamiento del relleno.*

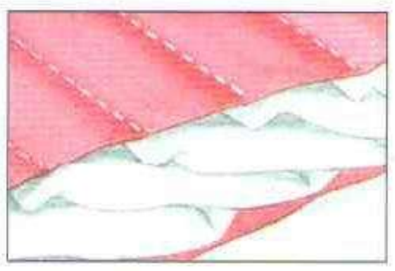

Tablillas. *Tienen capas inclinadas de fibras solapadas que se llenan con aire para proporcionar aislamiento térmico.*

Alvéolos. *El relleno se distribuye en alvéolos ovalados separados. Sin embargo, se pierde calor a través de las costuras.*

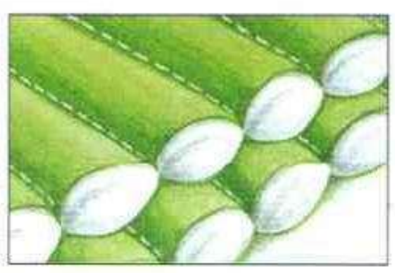

Doble barrera de alvéolos al tresbolillo. *Evita el apelotonamiento del relleno y la pérdida de calor a través de las costuras.*

ACCESORIOS PARA DORMIR

Funda. *Una funda de algodón forma una capa de aire aislante entre su cuerpo y el saco de dormir. Se lava con más facilidad que el saco.*

Aislante. *Es esencial para aislarse de la humedad y el frío del suelo. Debe colocarla debajo del saco de dormir.*

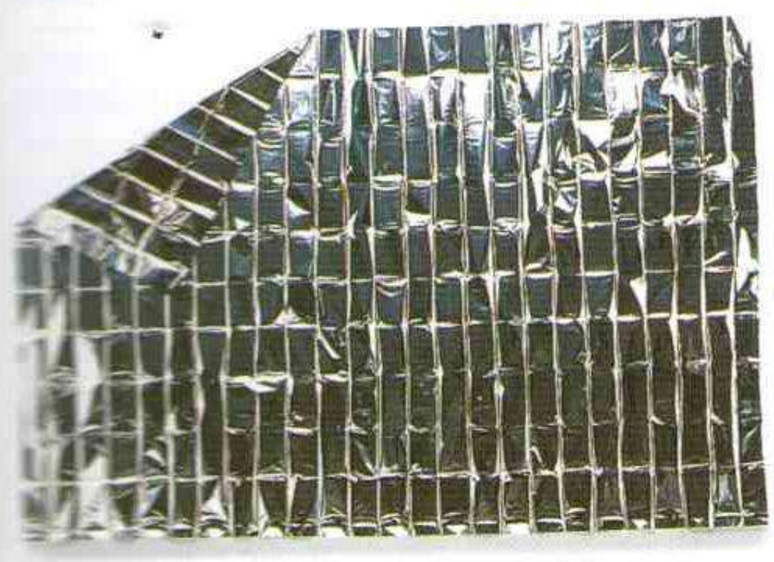

Manta de aluminio. *Se trata de una lámina de aluminio muy ligera para emergencias. El aluminio evita la pérdida de calor del cuerpo. También puede usar la manta como toldo para protegerse del sol, por ejemplo en un desierto.*

Almohada hinchable. *Aunque no es estrictamente necesaria, le proporcionará comodidad adicional en los viajes largos.*

CUIDADOS Y REMIENDOS

Sacuda su saco de dormir antes de utilizarlo para distribuir uniformemente el relleno, y airéelo después de su uso. Si se rasga, remiéndelo inmediatamente con aguja e hilo para evitar que el daño se extienda. Una cinta adhesiva ancha resulta útil para sellar desgarrones de forma temporal. Puede reparar su saco rasgado con un pedazo de tela de la bolsa que utiliza para guardarlo *(véase página 27)*, a menudo confeccionada con el mismo material.

Equipo de costura. Un equipo de costura resulta esencial para remendar los desgarrones en sacos de dormir, tiendas y en la ropa.

LA CONSTRUCCIÓN DE UNA CAMA

1 Si carece de un saco de dormir, debe construirse una cama levantada del suelo para asegurar un aislamiento adecuado y evitar los insectos y animales. Coloque varios troncos de longitud mayor que su estatura, uno junto a otro, en el suelo para formar la base de la cama.

2 Mantenga juntos los troncos y clave en el suelo unas estacas que habrá confeccionado a partir de ramas a las que previamente habrá afilado un extremo.

3 Cubra los troncos con una gruesa capa de hierba para esconder las irregularidades de los mismos debidas a la corteza y los nudos.

4 Añada más capas de ramas, helechos y follaje denso mezclado con hierbas y vegetación más fina para crear un espacio aireado y dar un efecto de profundidad. Debe construir un emparrado en el cual pueda recostarse cómodamente.

5 Continúe apilando capas de vegetación a modo de colchón hasta que sea suficientemente confortable. También puede confeccionarse una almohada a partir de un montón de hierba.

HERRAMIENTAS

AUNQUE ESTÉ ACAMPADO y tenga los últimos artilugios, o sobreviva a un desastre con un equipo mínimo, necesitará algunas herramientas con las que construir un refugio y cortar leña. En áreas vírgenes, no podrá reemplazar las herramientas rotas o perdidas hasta que vuelva a la civilización, por lo que debe cuidarlas. No haga mal uso de ellas de forma que pueda dañarlas, o lo que es más importante, que pueda dañarse a sí mismo. Puede improvisar algunas herramientas, aunque no se arriesgue a perder o estropear otra en el proceso. Trabaje siempre considerando las limitaciones de sus herramientas, así como su propia capacidad.

HERRAMIENTAS BÁSICAS PARA CORTAR MADERA U OTROS MATERIALES

La herramienta básica y esencial en un campamento y en una situación de supervivencia es un cuchillo. Aunque casi cualquier otra herramienta puede improvisarse, si no tiene un cuchillo le será muy difícil fabricar cualquier otro artículo.

Un hacha es muy valiosa para realizar trabajos pesados, pero debido a su peso es poco probable que lleve una.

Kukri. *Este cuchillo es empleado tradicionalmente por los gurkhas del Nepal. Está fabricado a partir de una ballesta trasera de un camión, de acero de alto grado de gran resistencia. En todo el mundo pueden encontrarse cuchillos similares. Si es posible, póngase unos guantes siempre que lo utilice para evitar heridas y rozaduras. Guárdelo en su funda de piel cuando no lo esté utilizando, para evitar herirse a sí mismo, o dañar el cuchillo.*

Navaja. *Una navaja resistente es la segunda herramienta en importancia después de sus dientes. Puede utilizarse en aquellos trabajos para los cuales su cuchillo es demasiado grande* (véase página siguiente). *Manténgala sujeta a su cuerpo con un cordel fuerte alrededor del cuello o de la cintura en todo momento. Manténgala siempre afilada.*

Cable sierra. *Un cable sierra consiste en varios hilos trenzados de alambre, con un par de lazadas como agarraderas. Puede incluirlo fácilmente en su equipo de supervivencia* (véase página 28). *Es una alternativa muy útil a un hacha pesada para cortar ramas, pero es delicado y debe utilizarse con cuidado.*

AFILAR UN CUCHILLO

1 Humedezca la piedra afiladera con agua. Deslice el cuchillo sobre la piedra, en dirección opuesta al filo de la hoja.

2 Después de afilar el cuchillo por una cara, toque la otra cara para hacer saltar las rebabas de metal provocadas por la abrasión.

3 Afile la otra cara de la hoja, realineando la rebaba hacia el centro. Puede necesitar más agua.

4 Suavice el cuchillo (deslizándolo arriba y abajo) sobre un cinturón de piel. Esto contribuirá a dar mayor uniformidad y a reforzar el filo.

EMPLEO DEL CUCHILLO

Cortar. *Corte con la sección principal de la hoja y mantenga sus brazos y piernas lo más alejados posible. Corte en dirección perpendicular a la fibra de la madera, con un ángulo de 45 grados para evitar que la hoja rebote y le biera.*

Corte siempre en dirección opuesta a su propio cuerpo para evitar accidentes

Tallar. *La talla es el arte de esculpir en madera pequeños objetos delicados, por ejemplo la punta afilada de un arpón para pescar* (véase página 104). *Para este tipo de trabajo, empuje la hoja cuidadosamente en sentido opuesto a su propio cuerpo con la yema del pulgar. Probablemente advertirá que a este efecto su pequeña navaja es mejor que el cuchillo.*

Cincelar. *El cincelado también debe realizarse en dirección opuesta a su cuerpo; tenga cuidado con sus dedos y miembros. Corte en el sentido de la fibra de la madera, dando pequeños golpes.*

EMPLEO DEL CABLE SIERRA

Haga las asas con una cuerda

1 Los lazos de metal pueden herir sus manos, por lo que debe añadir unas asas de cuerda o de tela a la sierra antes de utilizarla.

Sostenga el cable recto y tirante

2 Use ambas manos para mover la sierra y cortar la madera. Al cortar, mantenga el cable lo más recto posible, aunque cortará igual si lo curva alrededor de la madera. Mantenga un movimiento regular, pero ligero.

PRECAUCIÓN

Mantenga sus herramientas siempre afiladas, y tenga mucho cuidado al utilizarlas. Cuando saque un cuchillo de su funda, sosténgala por la parte correspondiente al lomo de la hoja. Un cuchillo afilado puede cortar la funda y seccionar los tendones de la mano. Un cable sierra también puede herirle si no lo utiliza adecuadamente.

HERRAMIENTAS IMPROVISADAS

1 Puede obtener una cuchilla natural dejando caer un trozo de pedernal sobre una roca. Al caer, se partirá, dejando los contornos afilados. Tenga cuidado con las astillas.

Un trozo de pedernal puede ser utilizado como cuchillo o como martillo

2 El borde afilado puede ser utilizado como una rudimentaria herramienta cortante. El otro extremo puede usarse como martillo. Cuando el borde pierda su filo, rompa otro de pedernal.

OTROS MATERIALES

Pala de bambú. *Del bambú puede obtenerse un canto muy afilado y fuerte. Construya una pala cortando una sección del tallo.*

Cuchillo de vidrio. *Un trozo de vidrio roto puede ser una cuchilla rudimentaria pero efectiva. Antes de usarlo, asegúrese de envolver uno de los extremos en tela a modo de empuñadura.*

CUERDAS Y NUDOS

LAS SOGAS Y cuerdas son imprescindibles para improvisar cualquier equipo, a menos que fabrique pinzas de madera o utilice alguna cola para unir las piezas de distintos materiales *(véase página 77)*. Si carece de cuerda o soga, puede fabricarse una a partir de materiales naturales. Utilizando los nudos adecuados, acordes a sus necesidades, es esencial –por su propia seguridad– que lo que haya hecho se mantenga entero, tanto si está usando una cuerda para cruzar un río como para asegurar el éxito de cualquier equipo o instalación que improvise. Si sólo es capaz de recordar un nudo, aprenda el nudo plano *(véase página siguiente)*.

LA CONFECCIÓN DE CUERDAS

Hay dos maneras principales de confeccionar una cuerda. La cuerda de tres cabos consta de tres hilos trenzados, mientras que el tipo *Kernmantle* tiene un núcleo central rodeado por una envuelta exterior tejida. Ambos tipos pueden fabricarse a partir de fibras naturales o sintéticas. El tipo de cuerda que elija dependerá del uso que pretenda darle. Es menos probable que se retuerza una cuerda tejida que una trenzada, mientras que una cuerda fabricada con materiales naturales puede pudrirse cuando se humedece. Las cuerdas de fibras naturales son más propensas a romperse que las sintéticas de nailon o poliéster.

Cuerda de tres cabos. *Una cuerda de tres cabos es más resistente que una cuerda* Kernmantle, *particularmente cuando se desgasta. Consta de tres hilos de fibra trenzados. Puede fabricarse con materiales naturales como el henequén, el cáñamo o la cáscara de coco, o con materiales sintéticos como el nailon. Las cuerdas de tres cabos se utilizan a menudo para amarrar o anclar barcos, debido a su resistencia.*

***Cuerda* Kernmantle.** *Los escaladores tienden a utilizar este tipo de cuerda, confeccionada a partir de un núcleo central de cuerdas finas de tres cabos entretejidas (el alma o* kern *en inglés), cubierto por una resistente envoltura exterior (la funda o* mantle *en inglés). Este tipo de cuerda es de fácil manejo, pero difícil de asir si está mojada o helada.*

CORDAJE IMPROVISADO

Para fabricar una cuerda a partir de tallos de ortiga, debe primero remojarlos para que adquieran mayor flexibilidad. Golpéelos después contra una piedra para separar las fibras de la médula, y deje secar las fibras. Podrá entonces entrelazarlos para formar cuerdas, con las que a su vez podrá formar un cordaje. También puede confeccionar cuerdas a partir de la corteza del sauce o del tilo.

1 Para trenzar las fibras, ate primero los cabos de tres hebras a una rama. Pase la hebra de la derecha por encima de la central.

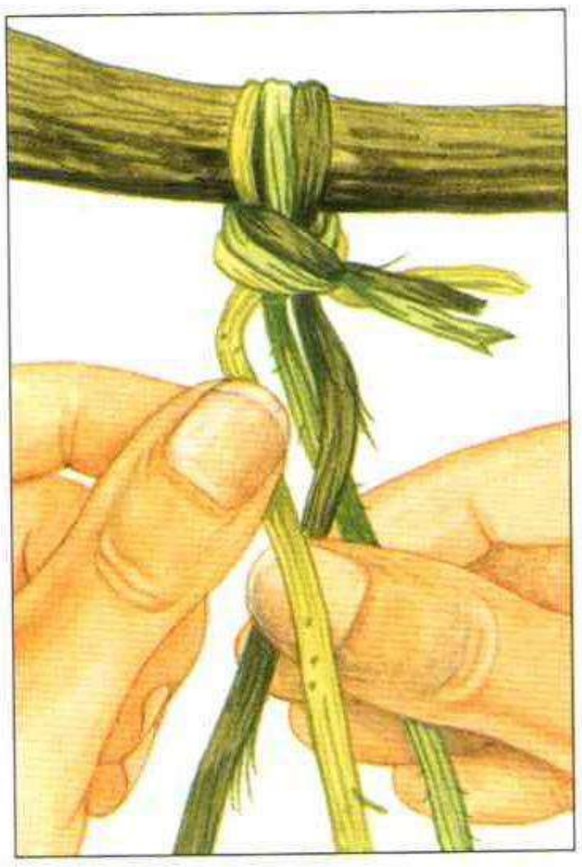

2 A continuación, pase la hebra izquierda por encima de la nueva hebra central, sin dejar de mantener tensas todas las hebras.

3 Pase la nueva hebra derecha por encima de la nueva hebra central. Con este paso terminará la primera sección de la trenza.

4 Repita los pasos hasta que la trenza tenga la longitud necesaria. Ate los cabos con fuerza para que no se deshaga la trenza.

NUDO PLANO

1 Para comenzar el nudo plano, pase el cabo derecho por encima y por debajo del izquierdo.

2 Pase el cabo izquierdo por encima y por debajo del cabo derecho, sujetando con fuerza ambos cabos.

3 Tire de ambos cabos para tensar el nudo. El nudo puede hacerse en el orden inverso (izquierdo sobre derecho, derecho sobre izquierdo), y puede deshacerse con facilidad.

BOULIN

1 Si se emplea para hacer un lazo, un boulin nunca se tensará o se aflojará. Haga una pequeña lazada y pase el cabo a través de ella por detrás.

2 Pase el cabo alrededor de la cuerda principal, y a continuación nuevamente a través de la lazada (la secuencia le resultará mucho más sencilla si recuerda: «el conejo sale de la madriguera, rodea el árbol y vuelve a entrar en la madriguera»).

3 Tire del cabo y de la cuerda principal para tensar el nudo. El boulin resulta útil, pero muchos escaladores prefieren el nudo en ocho para salvamento.

4 Para más seguridad, lleve el extremo sobre la parte derecha de la lazada y a través de ella, y a continuación nuevamente a través de la nueva lazada pequeña desde la parte superior.

NUDO EN OCHO

1 Este nudo es muy útil para hacer un lazo alrededor de un objeto. Forme una lazada doble de 1 m de longitud. Lleve el cabo libre por encima de la cuerda, y páselo a través de la lazada, formando la figura de un ocho.

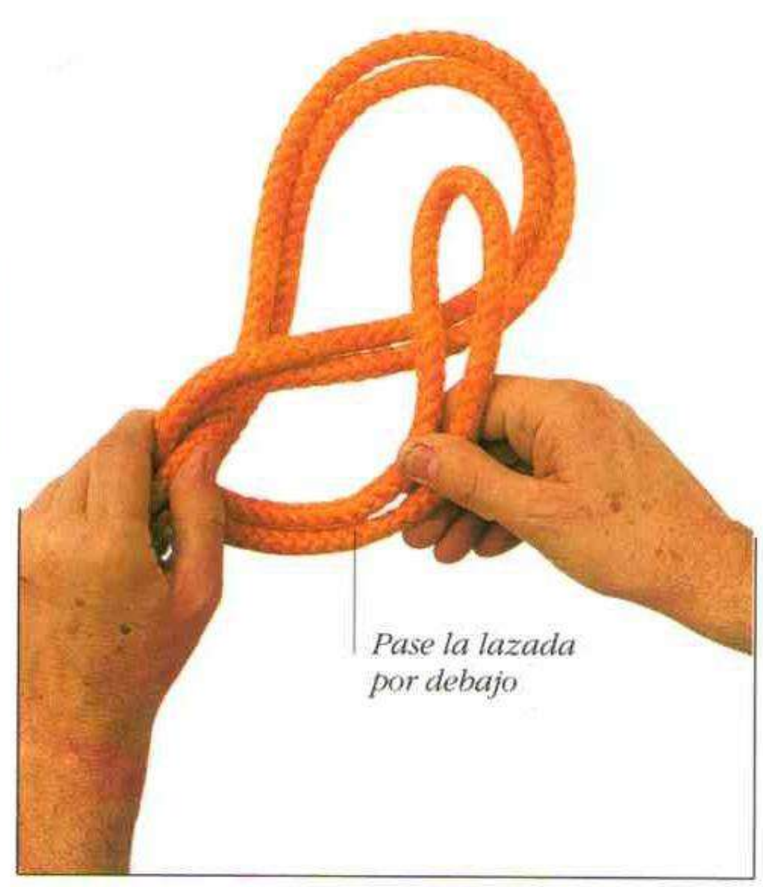

2 Tire de ambos cabos para tensar el nudo. Para formar el lazo, pase el cabo libre de la cuerda alrededor del objeto que quiera asegurar, y vuélvalo a introducir a continuación dentro del nudo en ocho.

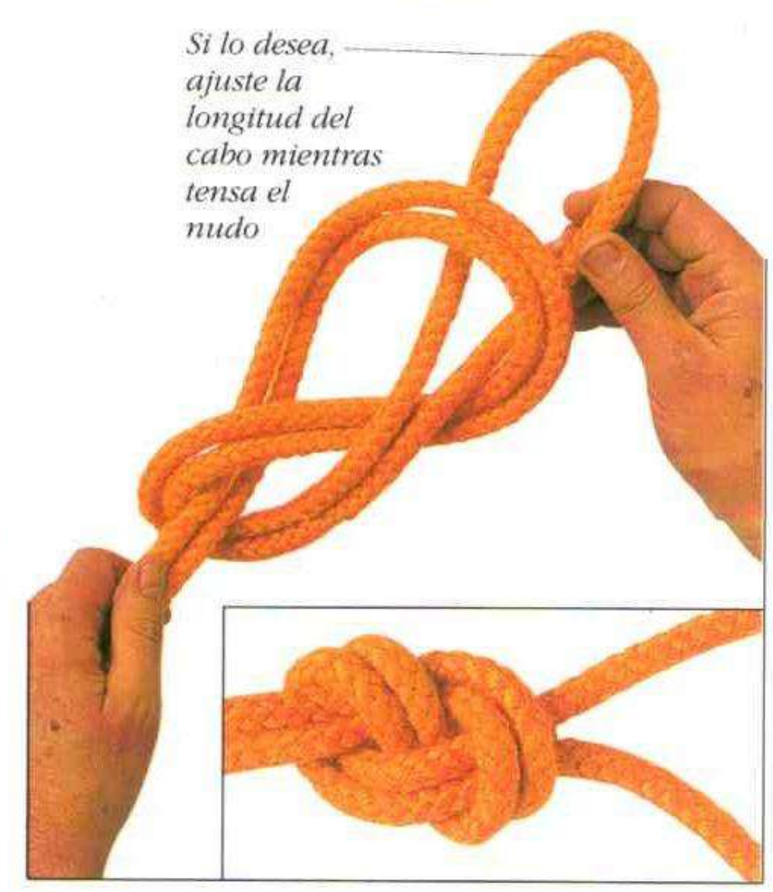

3 Tire con fuerza del lazo terminado y del cabo largo de la cuerda. Este nudo se emplea a menudo en escalada *(véase página 149)*. No se aflojará, aunque puede deshacerse con suma facilidad.

EL MANEJO DE LA CUERDA

UNA CUERDA PUEDE tener muchas aplicaciones, pero debe tratarla con cuidado: Su cuerda de escalar puede ser lo suficientemente resistente como para remolcar un vehículo averiado, pero perderá su capacidad de estirarse, y se ensuciará y estropeará en el proceso. Así pues, será menos probable que pueda salvarle la vida en un futuro, cuando la emplee para su propósito original. Utilice su cuerda uniformemente en sus dos cabos para no desgastarla de forma irregular. Nunca aplique una carga a una cuerda retorcida porque le causaría un daño importante. Mantenga limpia su cuerda lavándola con agua fresca y enrollándola cuando no la utiliza.

EL CUIDADO DE LA CUERDA

La humedad, la luz solar intensa y los insectos fitófagos pueden dañar las cuerdas de fibras naturales. Las fibras sintéticas pueden fundirse con el calor intenso, romperse al rozar con un canto bajo tensión, y son difíciles de asir cuando están húmedas. Las partículas de suciedad en las fibras de una cuerda pueden llegar a estropearla, así que manténgala siempre limpia y seca. Aclare las cuerdas húmedas o sucias en agua limpia y fresca. Cuelgue las cuerdas para que se sequen, en un sitio resguardado del sol y del fuego.

Daño en el alma. *Aunque la funda no está desgastada, esta cuerda no es segura, por lo que no debería utilizarse.*

Daño en la funda. *Corte las secciones dañadas, y utilice el resto para hacer ligaduras, ataduras, o empléelo en otros propósitos no tan vitales.*

CÓMO ENROLLAR UNA CUERDA

1 Siga el sentido natural de la cuerda al comenzar a enrollarla, y sacúdala para eliminar las zonas retorcidas o dobladas.

Forme círculos de unos 50 cm de diámetro

2 Utilice su pie para sujetar la cuerda mientras la enrolla. Si la superficie está sucia o polvorienta, enrolle la cuerda en la mano.

Asegúrese de mantener cada círculo junto al siguiente, sin enredarlos

Doble el cabo interior de la cuerda y forme una lazada de unos 30-35 cm de largo

3 Continúe formando círculos hasta que le quede cuerda para un círculo más. Para amarrar la cuerda, haga una lazada doblando el cabo del círculo interior.

Enrolle el cabo alrededor de los círculos para formar una ligadura

4 Enrolle la cuerda sobrante alrededor de los círculos. Pásela por encima de la lazada, pero deje libre su extremo.

Deje la lazada libre para la ligadura

5 Termine después de haber envuelto con la cuerda sobrante unos 10 cm de rollo, dejando sobresalir 3 cm de lazada.

6 Pase el extremo sobrante de la cuerda a través de la lazada y ténselo.

Tire de la lazada, por debajo de la ligadura

7 Tire del cabo de la cuerda que forma la lazada. De esta manera, tirará de la lazada por debajo de la ligadura y la tensará.

CÓMO UNIR DOS CUERDAS

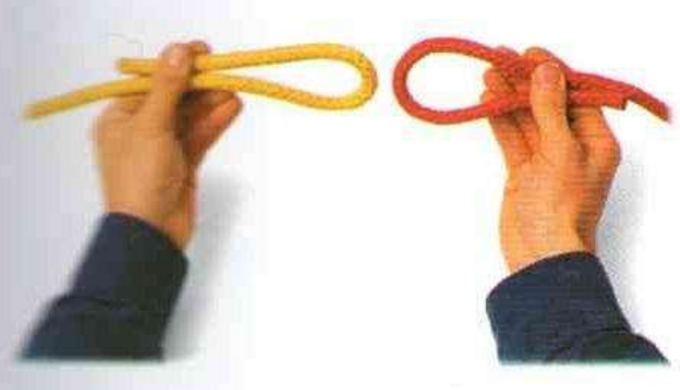

1 Si necesita unir dos cuerdas, átelas convenientemente para evitar que se separen. Comience por hacer una lazada con el cabo de cada cuerda.

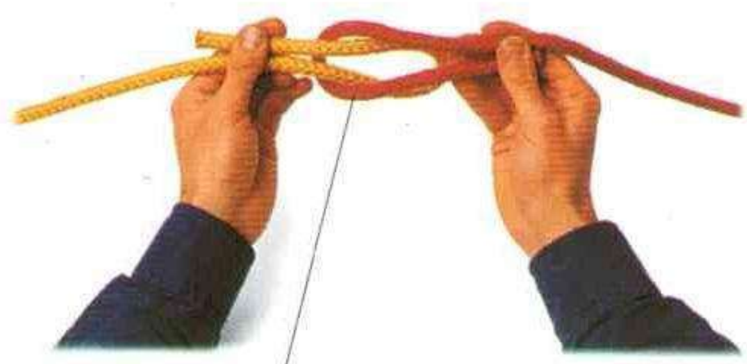

2 Acerque las lazadas hasta que una quede dentro de la otra. Podrá unir cuerdas iguales, o de materiales y grosores diferentes.

Una lazada entra en la otra

Sujete firmemente la primera lazada

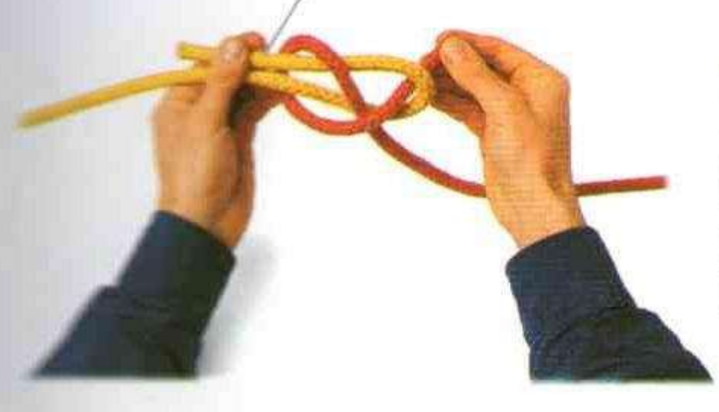

3 Tome el cabo suelto de la lazada exterior y páselo sobre sí mismo y después a través de la lazada interior. No suelte la primera lazada.

Deje que los cabos sean largos para evitar que se deshaga el nudo

4 Tire con fuerza de los extremos mientras sujeta la lazada interior. Si es posible, deje que los cabos sean largos para evitar que el nudo se deshaga.

ATADURAS

1 Para atar dos objetos —en este caso dos palos—, debe primero enrollarlos fuertemente con un cordel o una cuerda a fin de que queden bien unidos. Ate el cordel al primer palo.

El cordel se ata al primer palo para comenzar la atadura

2 Después de dar varias vueltas de anclaje alrededor del primer palo, pase el cordel firmemente alrededor del segundo.

3 Pase el cordel alrededor de ambos palos hasta unos 3 cm de altura, teniendo cuidado de no montar una vuelta sobre la anterior y manteniendo cada nueva vuelta fuertemente tensada.

La atadura debe ser tirante y esmerada

4 Pase el cordel entre los dos palos; páselo varias veces alrededor de la atadura para asegurarla.

5 Termine la atadura pasando el cabo del cordel bajo una lazada hecha alrededor de uno de los palos (ballestrinque), un par de veces. Tire del extremo con fuerza para asegurarla.

Pase el extremo del cordel por debajo de sí mismo para asegurar la atadura

6 Cuando la atadura haya quedado bien sujeta, mueva los palos hasta formar el ángulo deseado y utilícelos para formar varias estructuras, como un refugio con estructura en "A" *(véase página 52)*.

La atadura debe ser resistente, segura y flexible

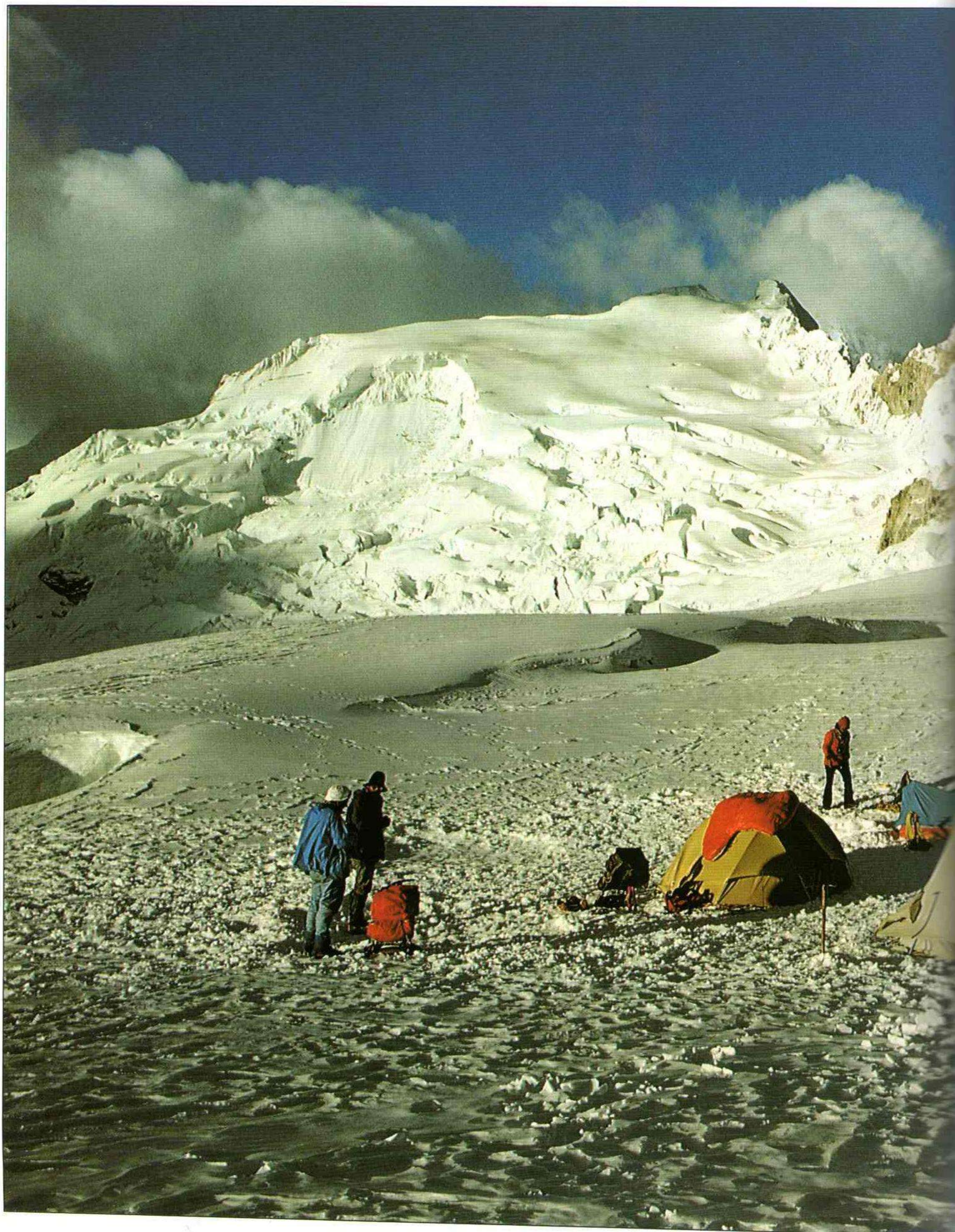

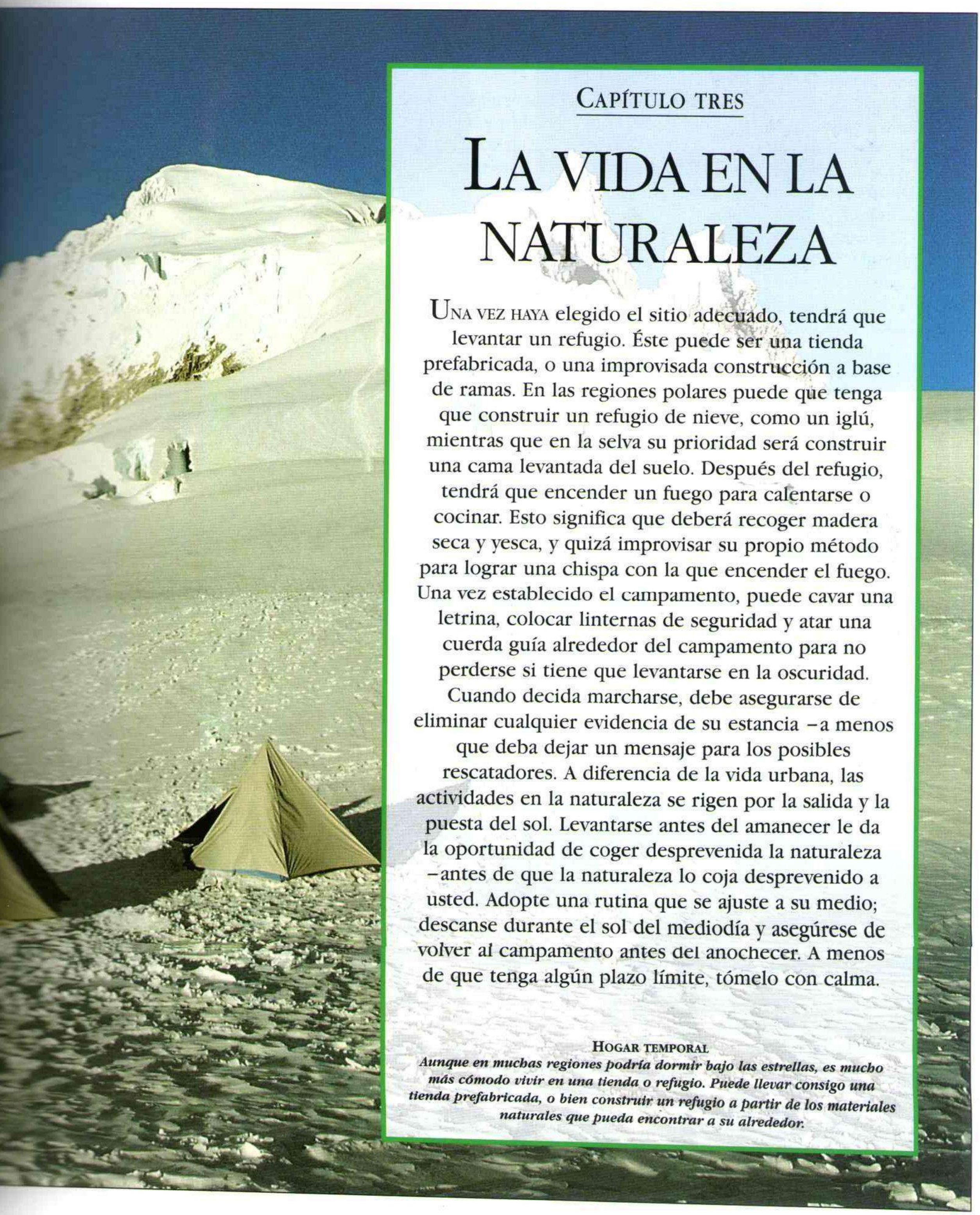

CAPÍTULO TRES

LA VIDA EN LA NATURALEZA

UNA VEZ HAYA elegido el sitio adecuado, tendrá que levantar un refugio. Éste puede ser una tienda prefabricada, o una improvisada construcción a base de ramas. En las regiones polares puede que tenga que construir un refugio de nieve, como un iglú, mientras que en la selva su prioridad será construir una cama levantada del suelo. Después del refugio, tendrá que encender un fuego para calentarse o cocinar. Esto significa que deberá recoger madera seca y yesca, y quizá improvisar su propio método para lograr una chispa con la que encender el fuego. Una vez establecido el campamento, puede cavar una letrina, colocar linternas de seguridad y atar una cuerda guía alrededor del campamento para no perderse si tiene que levantarse en la oscuridad. Cuando decida marcharse, debe asegurarse de eliminar cualquier evidencia de su estancia –a menos que deba dejar un mensaje para los posibles rescatadores. A diferencia de la vida urbana, las actividades en la naturaleza se rigen por la salida y la puesta del sol. Levantarse antes del amanecer le da la oportunidad de coger desprevenida la naturaleza –antes de que la naturaleza lo coja desprevenido a usted. Adopte una rutina que se ajuste a su medio; descanse durante el sol del mediodía y asegúrese de volver al campamento antes del anochecer. A menos de que tenga algún plazo límite, tómelo con calma.

HOGAR TEMPORAL

Aunque en muchas regiones podría dormir bajo las estrellas, es mucho más cómodo vivir en una tienda o refugio. Puede llevar consigo una tienda prefabricada, o bien construir un refugio a partir de los materiales naturales que pueda encontrar a su alrededor.

LA ELECCIÓN DEL CAMPAMENTO

SON MUCHOS LOS factores que influyen en la elección de un campamento, de los cuales la seguridad debe ser el más importante. Es posible que no sea capaz de identificar todos los riesgos potenciales o las desventajas de un sitio particular, pero para tener suerte, debe permitirse todo el tiempo necesario antes de la puesta del sol a fin de encontrar el sitio ideal donde acampar. El tiempo invertido en el reconocimiento nunca será tiempo perdido. Planifique sus actividades considerando la hora en que se pone el sol. Para entonces debe tener la tienda o el refugio montado y la comida casi a punto. Calcule una hora para establecerse, y al menos otra hora más para el reconocimiento. Así pues, si anochece a las 18.00 h, a las 15.00 h debe empezar a pensar en el campamento, y para las 16.00 h debe haberse detenido ya a buscar activamente el lugar adecuado.

Viento predominante. *Debe intentar identificar la dirección en la que sopla el viento predominante, y levantar su tienda de manera que la entrada se encuentre en la dirección opuesta. Debe cavar la letrina en la dirección del viento, más abajo de su campamento. Encienda el fuego de manera que el humo no entre continuamente en su tienda.*

CONSEJOS PARA LA VIDA EN UN CAMPAMENTO

- En tiempo de lluvia, cave un canal alrededor de la base de su tienda, que desemboque colina abajo para evitar las inundaciones. Si el viento es intenso, sujete las cuerdas de la tienda con piedras grandes.
- Deje todo guardado hasta el momento de utilizarlo. Así sabrá dónde están las cosas, será capaz de moverse rápidamente en caso de emergencia y disminuirá las posibilidades de perder algún equipo irremplazable.
- Nunca guarde comestibles dentro de una tienda. Cuelgue su mochila unos tres metros por encima del suelo y a un metro de los troncos de los árboles, de manera que los osos y monos no puedan cogerla.
- Deshaga su equipaje y remiende y airee su ropa y saco de dormir al sol antes de volverlos a meter en la mochila.

Secar las botas. *Deje secar sus botas en palos clavados en el suelo para evitar que entren animales en ellas.*

Árboles. *Aunque prefiera acampar cerca de los árboles a fin de disponer de leña abundante, o de madera para construir un refugio, tenga cuidado con los árboles muertos que puedan caer sobre su tienda, y con los animales peligrosos.*

Recodo del río. *Evite el recodo interior de un río, ya que en ocasiones el terreno es más bajo que en el recodo exterior, por lo cual es susceptible de inundarse. Las barreras de grava también se forman en los recodos interiores, donde el agua fluye más lentamente, por lo que se inundan con mayor facilidad.*

Peligro de avalancha. *Si acampa cerca de las montañas, no monte su tienda en el camino de una posible avalancha de rocas o nieve. Es más, el agua de los deshielos primaverales puede bajar por una ladera y causar graves inundaciones.*

La recogida de agua. *Recoja siempre el agua potable más arriba de su campamento y de los lugares en los que beben los animales.*

OTROS SITIOS

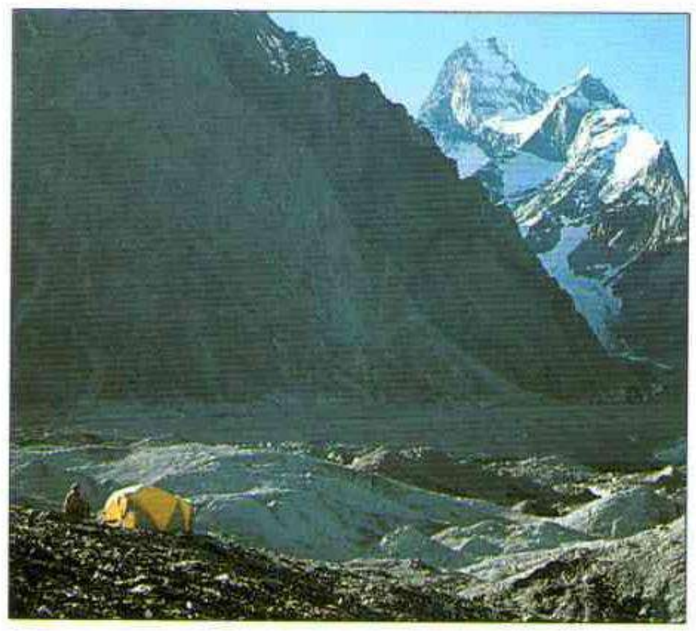

Al elegir un sitio, evite los frentes de aire frío y el suelo húmedo. Imagínese cómo fluirían el agua de una inundación y el aire frío, y elija un sitio por encima de estas líneas de flujo. Si le preocupan las picaduras de insectos, evite los suelos húmedos y busque zonas altas y barridas por el viento.

(OLLAS)

Lavar los utensilios. *Lave las perolas en un sitio intermedio entre donde recoge el agua para beber y el que usa para lavar la ropa. Limpie los restos de comida con arena o con un trapo antes de sumergir los utensilios en la corriente, ya que los restos de comida contaminarían el agua y atraerían a los animales. No use detergentes, podría envenenar a los peces.*

Tienda. *Plante la tienda lejos del río, donde el sonido de la corriente no encubra el ruido de los animales. Asegúrese de que la entrada no está situada de cara al viento ni cara a la letrina.*

Letrina. *Sitúe la letrina río abajo de la tienda. Asegúrese de que esté lo suficientemente lejos del río para no contaminar el agua.*

Fuego. *El humo del fuego contribuirá a alejar los insectos de la tienda; cuide que las llamas no queden tan cerca de la tienda como para quemarla.*

Lavar la ropa. *Lave su ropa en el río, más abajo de su tienda y de donde recoge agua para beber. Humedézcala primero en el agua y enjabónela en terreno firme. Aclárela en agua recogida en un cubo y no directamente en el río.*

El campamento ideal. *Bajo el abrigo de los árboles, pero lejos de cualquier árbol muerto que pudiera caer sobre la tienda, el campamento ideal está cerca del agua, pero lejos del peligro de inundaciones y de cualquier indicio de tipo animal. En la ladera de un valle, pero no a su pie, en un terreno nivelado y bien drenado, y protegido de los vientos predominantes en la zona.*

LA ELECCIÓN DE UNA TIENDA

LAS TIENDAS LE protegen del viento y de la lluvia y le proporcionan abrigo. La tienda exterior o doble techo debe ser resistente, impermeable y tensa, capaz de desviar incluso la lluvia más intensa y los vientos más fuertes. La tienda interior no debe estar tan tensada, y es de un material permeable para reducir al máximo el vaho al tiempo que retiene el calor, y posibilita la ventilación. Ambas partes de la tienda (interior y exterior) no deben tocarse; si lo hacen, dan lugar a la pérdida de calor y a la condensación de humedad en la parte inferior del doble techo, que puede filtrarse a la tienda interior. El suelo de la tienda interior es de una capa impermeable que la aísla de la humedad.

CARACTERÍSTICAS DE UNA TIENDA

Existen varios tipos de tienda, pero la mayoría siguen la línea básica de una tienda de dos pendientes. La tienda se sostiene gracias a palos hincados en tierra, y se sujeta al suelo mediante estacas que se introducen en los ojales del contorno inferior. Se mantiene tensa con los cables o vientos fijados al suelo con clavijas. Puede llevar una cubierta que la proteja de la humedad o del viento.

Tienda canadiense. *Una tienda canadiense es adecuada para acampar en cualquier sitio, desde una montaña hasta un jardín doméstico.*

SELECCIÓN DE LA TIENDA ADECUADA

Existen varios tipos diferentes de tiendas, pero todos ellos responden a un patrón parecido. Su aplicación precisa, así como el clima y el terreno de destino determinarán el tipo de tienda a comprar. También debería considerar el tamaño y el peso de la tienda ya que tendrá que cargarla si no utiliza un vehículo. Las tiendas individuales pueden resultar demasiado frías, por lo que son preferibles las tiendas para dos personas; resulta sorprendente la diferencia de calor que generan dos cuerpos.

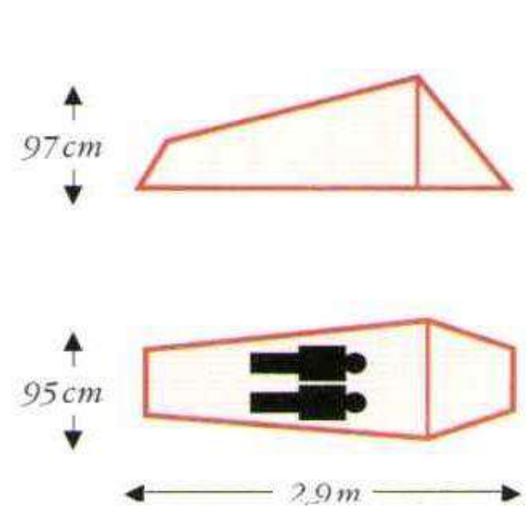

Tienda para dos personas. *Una tienda canadiense para dos personas es ligera; dispone de sitio para almacenar y cocinar debajo del doble techo y tiene el espacio justo para que dos personas duerman en su interior, una al lado de la otra.*

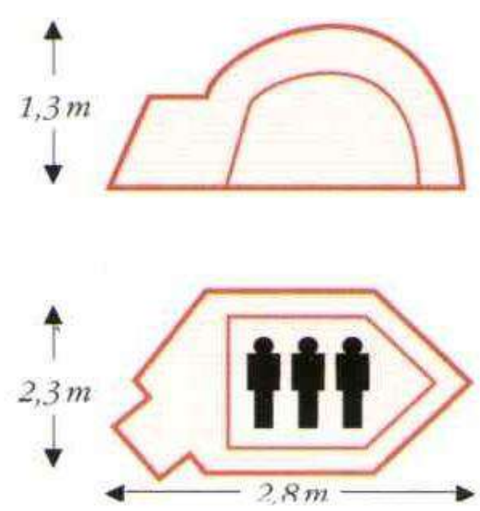

Tienda para tres personas. *Dentro de esta tienda en forma de cúpula hay espacio suficiente tanto para el almacenaje del equipo como para sentarse cómodamente. Tres palos flexibles proporcionan a la estructura una gran resistencia contra los fuertes vientos.*

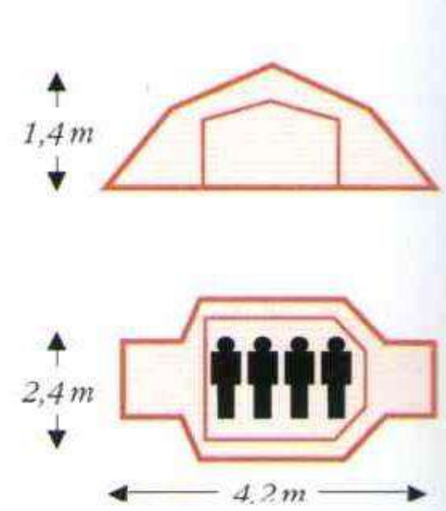

Tienda para cuatro personas. *Esta tienda, de ambiente agradable, tiene capacidad para cuatro personas. En una tienda compartida debe asegurarse de que la ventilación sea buena, y de que cada uno mantenga su equipo bien recogido para evitar un excesivo desorden.*

TIENDAS PARA CUALQUIER EQUIPO

Algunos tipos de tienda son más adecuados para un terreno particular que otros. Una tienda canadiense clásica puede utilizarse prácticamente en cualquier sitio, pero algunos modelos modernos, como las tiendas de cúpula geodésica, se adaptan mejor a los terrenos inhóspitos y a los climas extremos —por ejemplo, a los vientos intensos y a las densas nevadas de las regiones montañosas. Las tiendas con palos exteriores son más fáciles de montar en presencia de vientos fuertes que aquéllas con palos interiores. Las tiendas de dos aros son inestables en climas extremos.

Tienda de cúpula. *Es ideal para las condiciones extremas, como las que se encuentran en las regiones montañosas o polares, ya que puede permanecer estable si se entierra en la nieve. Sin embargo, debe limpiarse antes de que se acumule demasiada nieve. Enterrarla en un banco de nieve es la forma ideal de protegerla del viento.*

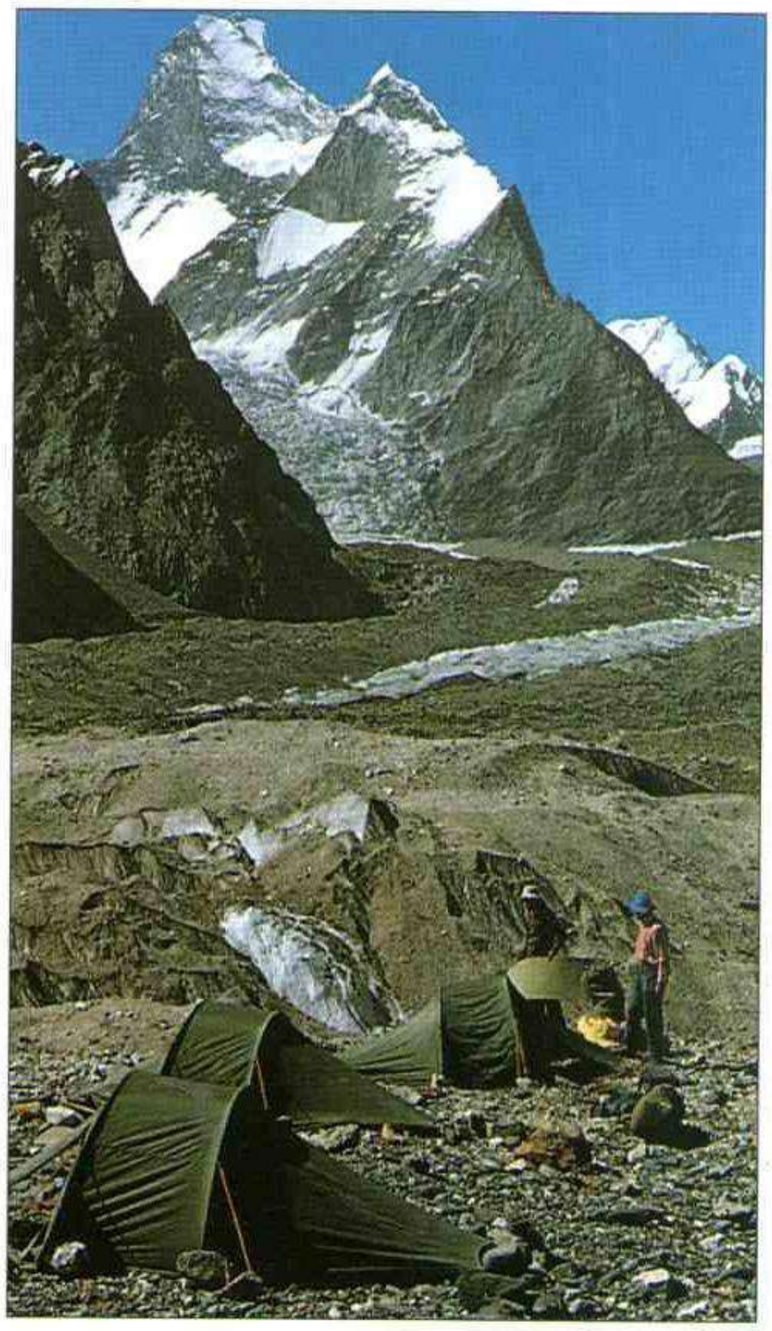

Tienda de túnel. *Una tienda de túnel es un híbrido entre una tienda de cúpula y una tienda canadiense. Las versiones individuales de este tipo de tienda son ideales para los excursionistas, ya que son ligeras y fáciles de levantar. Pueden utilizarse sobre hierba o en valles rocosos, y pueden desviar los vientos fuertes.*

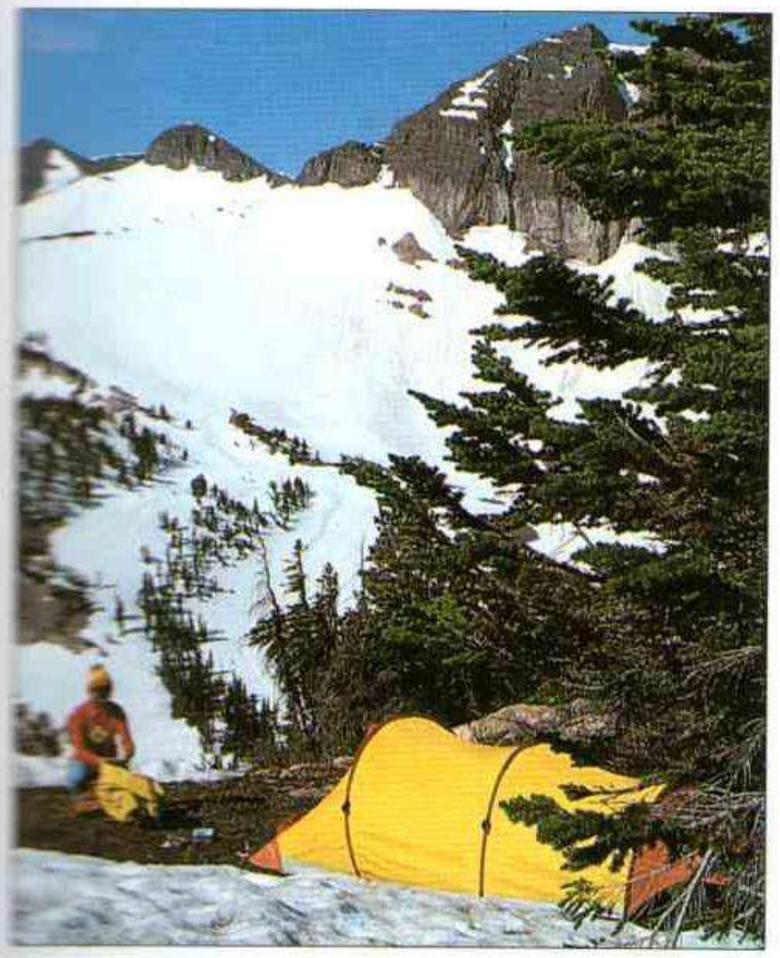

Tienda de túnel de tres aros. *Las tiendas de túnel pueden tener una estructura de hasta tres aros, dos pequeños en los extremos y uno mayor en el centro. Los túneles de tres aros pueden albergar dos o tres personas. Son relativamente estables y gracias a su forma no almacenan el agua de la lluvia. A menudo tienen dos entradas.*

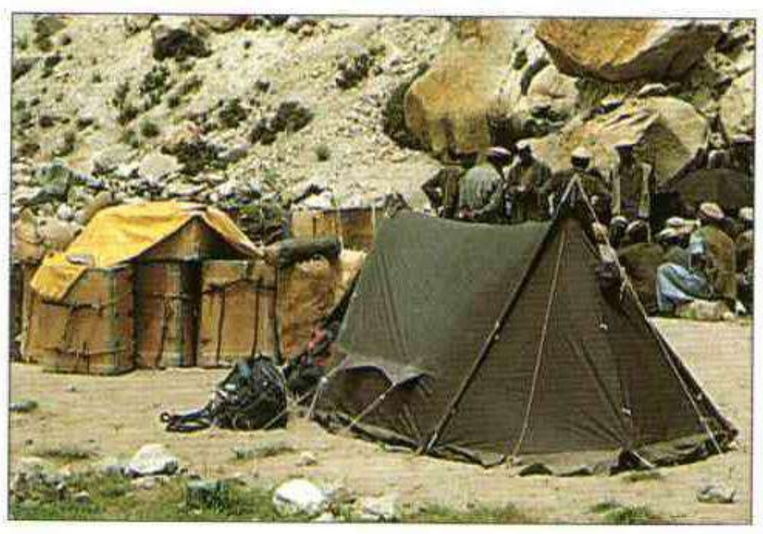

Tienda canadiense. *La tradicional tienda canadiense puede plantarse casi en cualquier sitio. Tiene una barra central horizontal, y dos barras de la misma altura en cada extremo. Este modelo de tienda se fabrica en cualquier tamaño, individual o con capacidad para varias personas.*

Tienda de cúpula geodésica. *Las tiendas de cúpula geodésica son resistentes, ligeras, fáciles de levantar, e igualmente fáciles de desmontar. En presencia de vientos fuertes deben sujetarse firmemente, pero la disposición de sus palos curvados y entrelazados las hace resistentes. Las tiendas de cúpula geodésica se sujetan gracias a unas varillas flexibles, y no a cables o vientos, que se cruzan a intervalos diferentes para mantenerla tensa. Gracias a su forma, proporcionan un espacio vertical considerable.*

FUNDA DE VIVAC

Una funda de vivac puede ser una simple cubierta impermeable para un saco de dormir; o si dispone de pequeños palos cortos, puede formar un porche y convertir el saco en una pequeña tienda. Debe fabricarse en un material transpirable para reducir al máximo el vaho, y puede utilizarse incluso bajo una lluvia intensa al aire libre —aunque resulta bastante ruidoso. Una funda de vivac puede utilizarse también como saco de supervivencia impermeable.

Cubierta para un saco de dormir. *Una funda de vivac constituye una excelente cubierta impermeable para un saco de dormir; puede utilizarse en lugar de una tienda.*

MONTAR UNA TIENDA DE CAMPAÑA

LA ELECCIÓN DEL campamento entraña la selección de la zona adecuada para su tienda con respecto a los aspectos que ha de incluir su campamento y considerando su seguridad y refugio. Debe decidir hacia dónde encarará la entrada de la tienda y cómo asegurará los vientos, así como tomar en cuenta cualquier peligro potencial *(véase página 40)*. Debe limpiar y nivelar la zona, eliminando o allanando cualquier protuberancia del terreno.

Al igual que cualquier trabajo en un campamento, intente establecer una rutina, realizando las mismas acciones en el mismo orden cada vez, de manera que se convierta en una costumbre. Esto le permitirá plantar su tienda en la oscuridad, en una ventisca o bajo una lluvia intensa. Asegúrese de que sabrá encontrar cada una de las piezas de la tienda cuando llegue el momento de montarla. También debe desmontarla sistemáticamente *(véase página 46)*.

1 Acostúmbrese a montar su tienda como si lo hiciera en medio de una fuerte ventisca. En primer lugar coloque la tienda interior en el suelo y fije inmediatamente las esquinas.

Sujete el ojal de la esquina con la clavija y déjelo recto y tenso

2 Monte los palos de la tienda. Tenga cuidado de unir las piezas correctas, o tendrá dificultades para separarlas posteriormente. Sujete el enganche posterior de la tienda en el palo más corto.

3 Levante el palo corto y fije su viento al suelo, tensándolo hasta que el palo permanezca vertical por sí mismo. Asegúrese de que la pared posterior de la tienda encara al viento.

4 Sujete el otro palo al ojal de la parte frontal superior de la tienda. Levántelo y tense el viento. Trabaje deprisa —si hace viento la tienda será poco estable, y si llueve se le mojará.

5 Coloque el doble techo sobre el palo posterior y el extremo más bajo de la tienda, y fíjelo al suelo.

6 Deslice el doble techo sobre el techo de la tienda enganchando el ojal en el mástil principal de la parte alta de la tienda. Sujete la parte frontal del doble techo al suelo con el viento. Es posible que tenga que mover el viento de la tienda interior hacia el palo para que el doble techo se ajuste bien.

PRECAUCIÓN

No utilice las cintas de goma de la tienda para sacar clavijas clavadas en un lugar equivocado porque las rompería. Para sacar las clavijas de un terreno duro, utilice otra clavija o un gancho especial para este fin.

DESMONTAR UNA TIENDA DE CAMPAÑA

CUANDO DESMONTE UNA tienda, guarde cada pieza de forma ordenada, para que pueda encontrarla rápidamente cuando sea necesario. Si está lloviendo, guarde todas las piezas de la tienda lo más rápidamente posible, en especial la tienda interior. Recuerde que usted es vulnerable si no tiene un refugio donde cobijarse, y que debe desplazarse hasta su siguiente campamento lo más rápidamente posible y levantar su tienda antes de que anochezca. Si el clima es agradable y seco, antes de guardar la tienda extienda todas sus partes al aire libre durante un rato para permitir que el viento y el sol sequen la humedad de la tienda interior y del doble techo. Esto evitará la aparición de moho durante su almacenamiento.

1 En primer lugar, saque todas las clavijas que sujetan la tienda interior y el doble techo. Si el viento es intenso, retenga montada parte de la estructura para evitar que el viento se la lleve. Tenga especial cuidado con las clavijas, nunca utilice los vientos o las cintas de goma para arrancarlas del suelo, y cuéntelas siempre.

Saque las clavijas del suelo teniendo cuidado de no doblarlas

2 Abra la cremallera de la puerta del doble techo y saque el palo frontal. Limpie la tierra o la humedad del palo, desmóntelo cuidadosamente y coloque las distintas partes directamente en su bolsa.

3 Saque el otro palo y desmóntelo. Límpielo y séquelo antes de guardarlo en su bolsa. Extienda la tienda sobre el suelo para poder recoger fácilmente los vértices del tejado.

4 La tienda interior y el doble techo deben enrollarse y guardarse separadamente para poder distinguirlos rápidamente al montar de nuevo la tienda. Para guardar el doble techo, levántelo de la tienda interior sujetándolo de los ojales en los vértices del techo. Tenga cuidado de no levantar también la tienda interior.

5 Sin dejar de sujetar el doble techo por los ojales, levántelo para que cuelgue según su forma natural. Después dóblela por la mitad para formar una gran «V». Asegúrese de que no esté mojada antes de doblarla.

6 Coloque el doble techo sobre un trozo seco de terreno y comience a enrollarlo lo más apretado posible. Doble los laterales conforme lo va enrollando, teniendo cuidado de no enredar los vientos.

7 Sacuda la tienda interior de manera que los vientos cuelguen libremente, colóquela sobre el suelo y enróllela lo más apretada posible. Doble los laterales conforme la enrolla.

8 Coloque los vientos cuidadosamente en el rollo antes de terminarlo, asegurándose de que no están enredados ni tienen nudos. Si éste es el caso, es el momento de deshacerlos.

9 Guarde las clavijas en una bolsa resistente para que no dañen la tienda. Guarde la bolsa con las clavijas dentro de la bolsa principal de la tienda. El barro o la humedad que pudiera quedar en las clavijas puede corroerlas o dañarlas.

10 Limpie los palos y métalos en una bolsa resistente, y después dentro de la bolsa de la tienda. Si los mantiene limpios, ajustarán fácilmente cuando los utilice.

11 Meta la tienda interior y el doble techo en la bolsa de la tienda. Es una excelente idea colocar antes cada artículo en una bolsa de plástico individual para dar una protección adicional.

PRECAUCIÓN

Asegúrese de no perder ninguna clavija, viento o palo al desmontar su tienda. También debe tener cuidado de no dañar ninguna de sus partes; si esto ocurriera, debe repararla antes de utilizarla de nuevo.

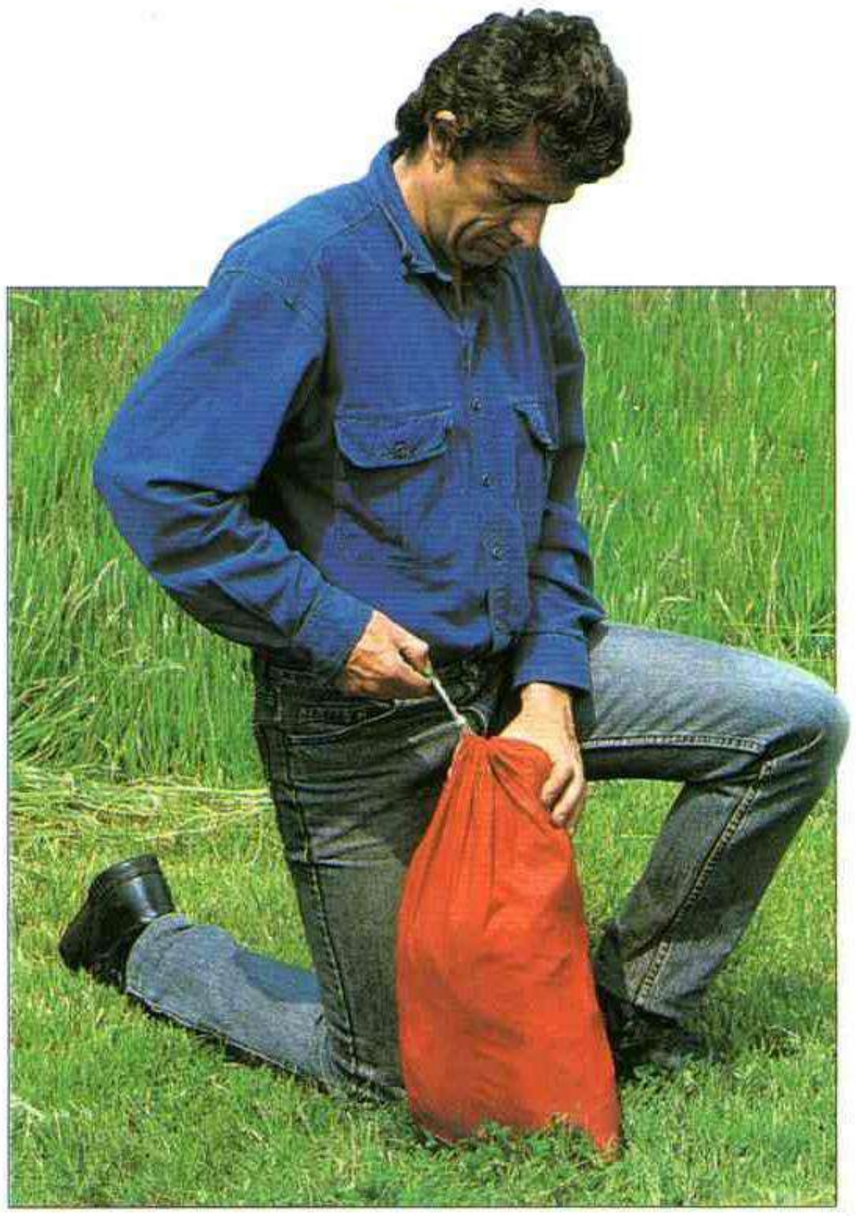

12 Antes de cerrar la bolsa de la tienda, compruebe que todo esté guardado, y que no ha olvidado ninguna clavija o palo. Si la hierba es alta es fácil extraviar alguno de estos artículos.

CUIDADO Y REPARACIÓN DE UNA TIENDA

- Su tienda debe estar totalmente seca antes de guardarla; además de ser más ligera, el tejido se deteriorará menos que si se almacena húmedo.
- Evite montar la tienda o desmontarla cuando hace mucho calor. El aislante impermeable puede derretirse con un calor excesivo.
- Compruebe todas las partes de su tienda. Si es necesario, repárelas inmediatamente.
- Antes de almacenar una tienda, lávela con agua fría. Cuélguela para que se seque antes de guardarla.

Parches. *Puede comprar parches especiales de vinilo con los que remendar temporalmente su tienda.*

LA CONSTRUCCIÓN DE UN REFUGIO

PROTEGERSE DEL VIENTO, de la lluvia y del sol para poder descansar y dormir resulta vital para sobrevivir. Cuanto mejor sea el refugio construido, más cómodo estará y mejor descansará. El estar descansado es esencial para su salud física y también para su estado psicológico, que determina la sensibilidad y la lógica de su pensamiento, y su decisión de sobrevivir. Un cobertizo es probablemente el tipo de refugio más sencillo y rápido de construir, y es, además, adecuado para la mayor parte de terrenos. Debería construirse en un lugar protegido, seguro y llano *(véase página 40)*, y de tal manera que el techo estuviera de cara al viento. La construcción de su refugio dependerá enteramente de los materiales de que disponga. Si dispone de una lona o de una cubierta de plástico, debe emplearla para construir el techo del refugio; en caso contrario, tendrá que improvisar con los materiales que pueda encontrar.

El travesaño debe ser más alto que el alcance de su brazo

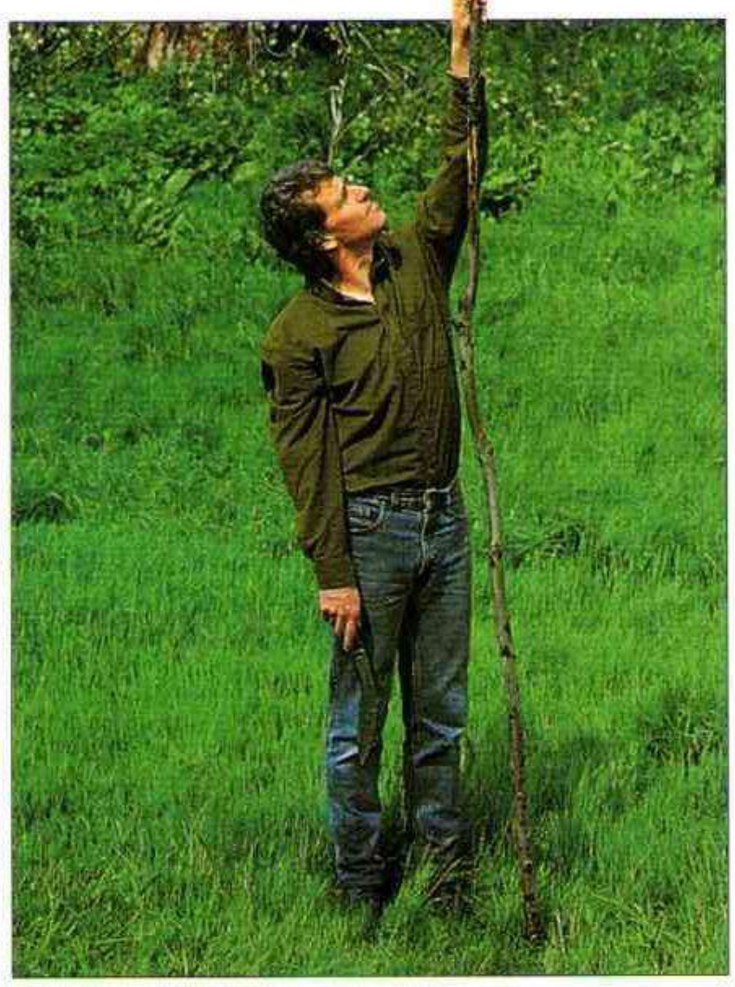

1 Para construir un refugio a un agua, comience cortando una rama larga que hará las veces de travesaño. Mida la longitud correcta de esta rama extendiendo su brazo y cortando unos 60 cm por encima de este punto. Intente que la rama sea lo más recta posible, y recorte todas las ramas que broten de ella. No utilice una rama muerta ya que un viento fuerte podría romperla.

2 Corte dos soportes en forma de «Y» cuya altura llegue hasta su pecho, y córtelos con una longitud adicional de 30 cm que se clavarán en el suelo. Afile los extremos inferiores.

Afile los extremos de los soportes para que puedan clavarse en el suelo

3 Con la ayuda de una piedra, clave los soportes en el suelo hasta que estén bien seguros y no se muevan. La distancia entre ellos debe ser de unos 60 cm menos que la longitud del travesaño.

4 Apoye el travesaño sobre los dos soportes en «Y», dejando la misma distancia en ambos extremos. El travesaño debe ser ligero, aunque lo suficientemente resistente para soportar el peso del techo terminado, así como la acción de la lluvia y el viento, sin derrumbarse.

5 Corte varias ramas resistentes lo suficientemente largas para inclinarlas contra el travesaño en un ángulo de 45 grados con el suelo. El techo inclinado permitirá el drenaje del agua de la lluvia y evitará su estancamiento. Colóquelas a unos 20 cm de distancia entre sí. Deben sobrepasar el travesaño unos 10 cm.

6 Corte varias ramas verdes rectas, más bien fuertes, dejando algunas hojas en ellas. Entretéjalas alternativamente por encima y por debajo de las ramas inclinadas del techo, hasta que todas ellas formen una red relativamente firme. Entreteja también una fila de ramas por encima y por debajo del travesaño para asegurarlo mejor.

Entreteja las ramas inclinadas con el travesaño para dar mayor estabilidad a la estructura y evitar que las ramas se muevan

7 Una vez formada la red soporte, entreteja ramas completas, incluyendo sus hojas, para evitar la entrada del viento y de la lluvia. Al emplear las ramas enteras, dará mayor fuerza al refugio, y el follaje permanecerá vivo un mayor tiempo, en especial si es caducifolio.

8 Continúe entretejiendo todo tipo de follaje hasta lograr la densidad deseada para su techo. Algunos tipos de hoja se secan más rápidamente que otras, y deberán ser reemplazadas transcurridos algunos días. Resulta mejor entretejer las ramas nuevas por encima de las viejas para ir construyendo un techo en varias capas. Son preferibles las ramas de árboles con hojas pequeñas, ya que las hojas grandes se rasgan, mueren más rápidamente, y son más difíciles de tejer para formar una superficie plana.

9 El refugio terminado puede requerir algunas mejoras: puede colocar rocas grandes a lo largo de la base para sujetarla en caso de vientos fuertes, o añadir paredes laterales. Encienda el fuego delante del refugio, pero mantenga las llamas alejadas del techo. Renueve las ramas cuando se marchiten.

El empleo de materiales artificiales

El empleo de materiales artificiales facilita la construcción de un refugio. Las láminas de polietileno, las bolsas de plástico y los cordeles para embalar son comunes entre la basura encontrada en el campo. Las hojas de hierro galvanizado y las cajas de madera pueden resultar útiles.

Otros materiales. *Los refugios pueden construirse a partir de material de desecho.*

REFUGIOS EN LA NIEVE

SI LAS TEMPERATURAS permanecen bajo cero, la construcción de refugios en la nieve es relativamente fácil. Protegerlos del viento es prioritario, ya que el viento puede disminuir drásticamente la temperatura *(véase página 141)*. Las temperaturas inferiores a −10 °C resultan desagradables, por lo que la construcción de refugios que puedan retener muy bien el calor se convierte en una necesidad. Estos pueden variar desde una simple trinchera en la nieve hasta un iglú, en cuya construcción bien puede emplear unas cuantas horas. En un refugio permanente, como un iglú, el aire frío y más denso se disipará del interior si cava un canal que desvíe el aire hacia la parte inferior del iglú y de allí hacia afuera. Una ventilación adecuada es imprescindible en cualquier refugio en la nieve para evitar la asfixia.

LA CONSTRUCCIÓN DE UN IGLÚ

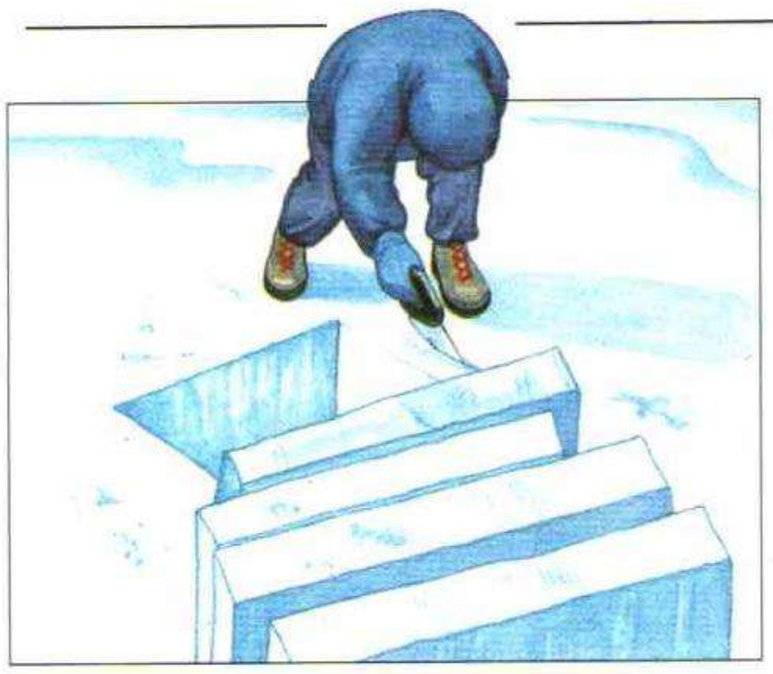

1 Corte bloques de nieve dura empleando una sierra para nieve o un cuchillo grande. Cada bloque debe tener aproximadamente 1 m de largo, 40 cm de alto y 20 cm de ancho.

2 Forme un círculo con los bloques alrededor del agujero que se ha formado en el suelo. Corte el círculo en espiral desde la parte superior del último bloque hasta el suelo. Así resultará más fácil construir la cúpula.

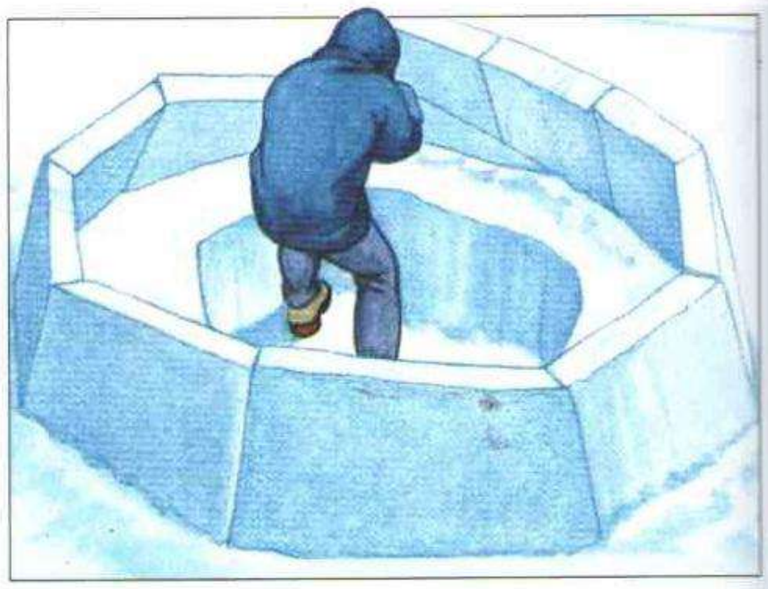

3 Construya las paredes solapando los bloques y colocándolos inclinados hacia el interior. Cave un agujero bajo la pared para la entrada y para disipar el aire frío. Coloque bloques en una pared como plataforma para dormir.

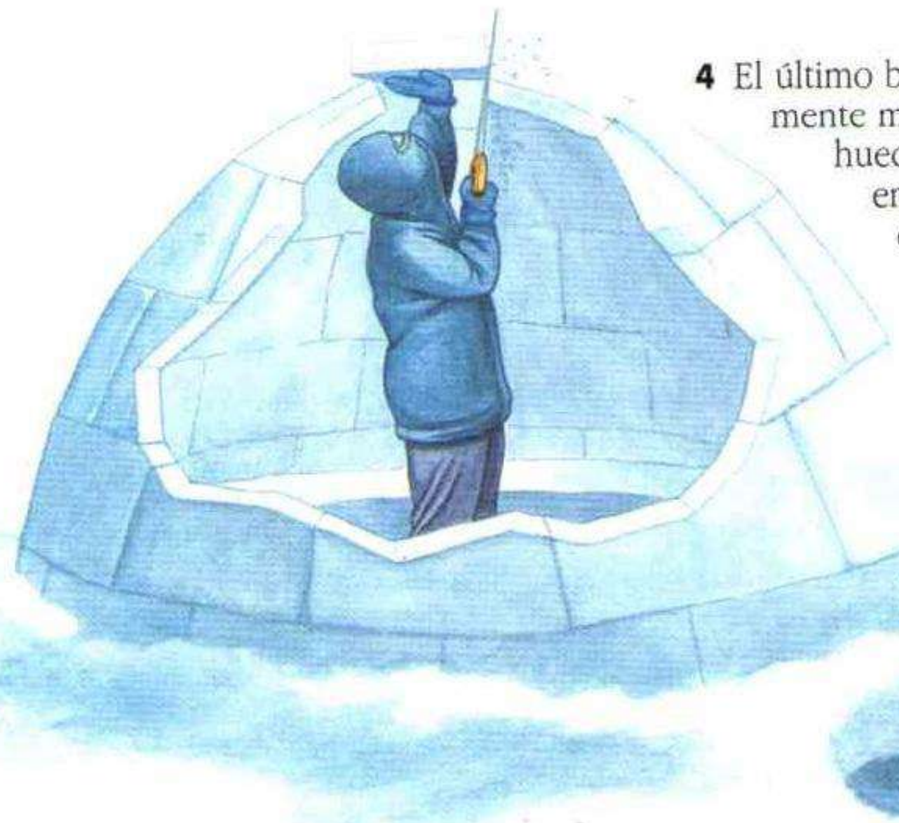

4 El último bloque debe ser inicialmente más grande que el hueco. Coloque el bloque encima del iglú y, desde dentro, déle la forma correcta y deslícelo para que encaje perfectamente en el hueco.

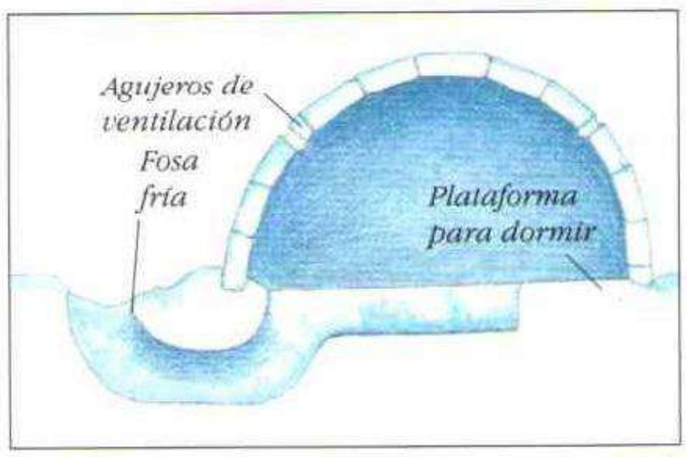

5 El aire caliente proveniente de su cuerpo y de la cocina sube, y queda atrapado por la cúpula. El aire frío baja al túnel y se disipa hacia el exterior. Los agujeros para la ventilación son indispensables; deben realizarse con un hacha especial para hielo.

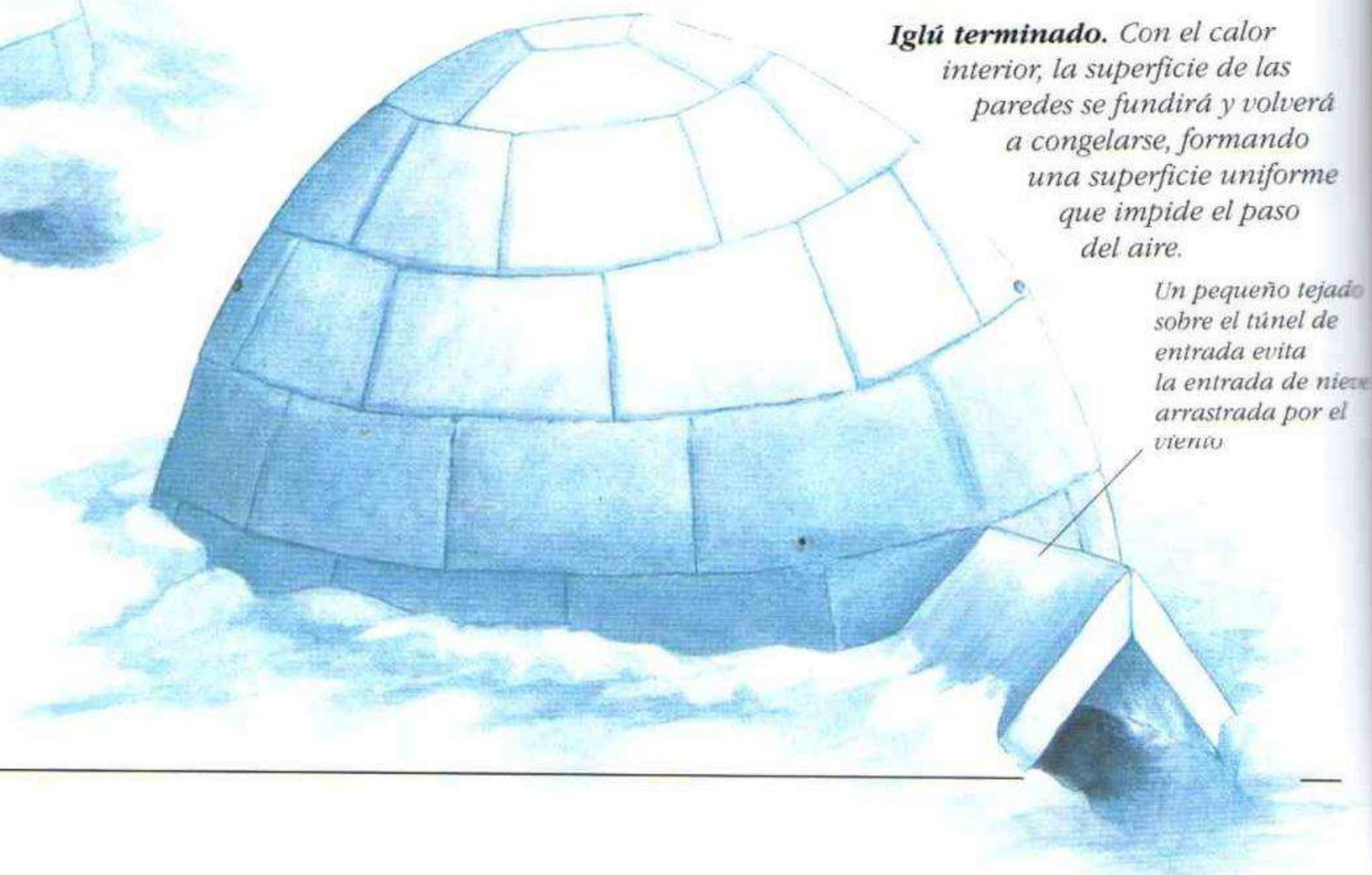

Iglú terminado. *Con el calor interior, la superficie de las paredes se fundirá y volverá a congelarse, formando una superficie uniforme que impide el paso del aire.*

ADVERTENCIA

Es fundamental perforar por lo menos un agujero de ventilación en el techo para evitar la asfixia. El interior del iglú se calentará con el calor de su cuerpo. Sin una ventilación adecuada puede acumularse monóxido de carbono, que es letal. Para construir un iglú se precisa de dos personas.

La construcción de un *quinze*

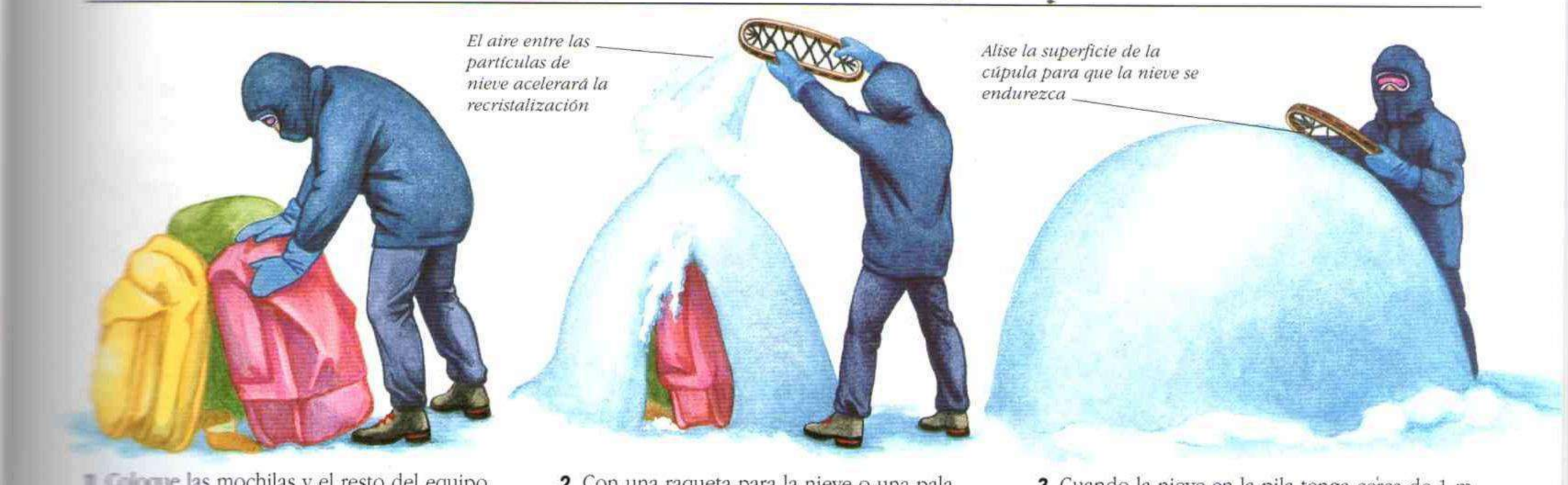

1 Coloque las mochilas y el resto del equipo formando un cono compacto. El equipo determinará la forma interior del refugio y reducirá la cantidad de nieve necesaria para construir el *quinze* (pronunciado «cuinsi»).

2 Con una raqueta para la nieve o una pala, apile nieve sobre las mochilas, y compáctela. Espere al menos 30 minutos a que la nieve se congele antes de añadir más nieve y lograr el grosor deseado para la cúpula.

3 Cuando la nieve en la pila tenga cerca de 1 m de grosor, alise la superficie y espere cerca de una hora para que se endurezca. Este período de espera es muy importante, ya que durante el mismo la nieve se recristaliza.

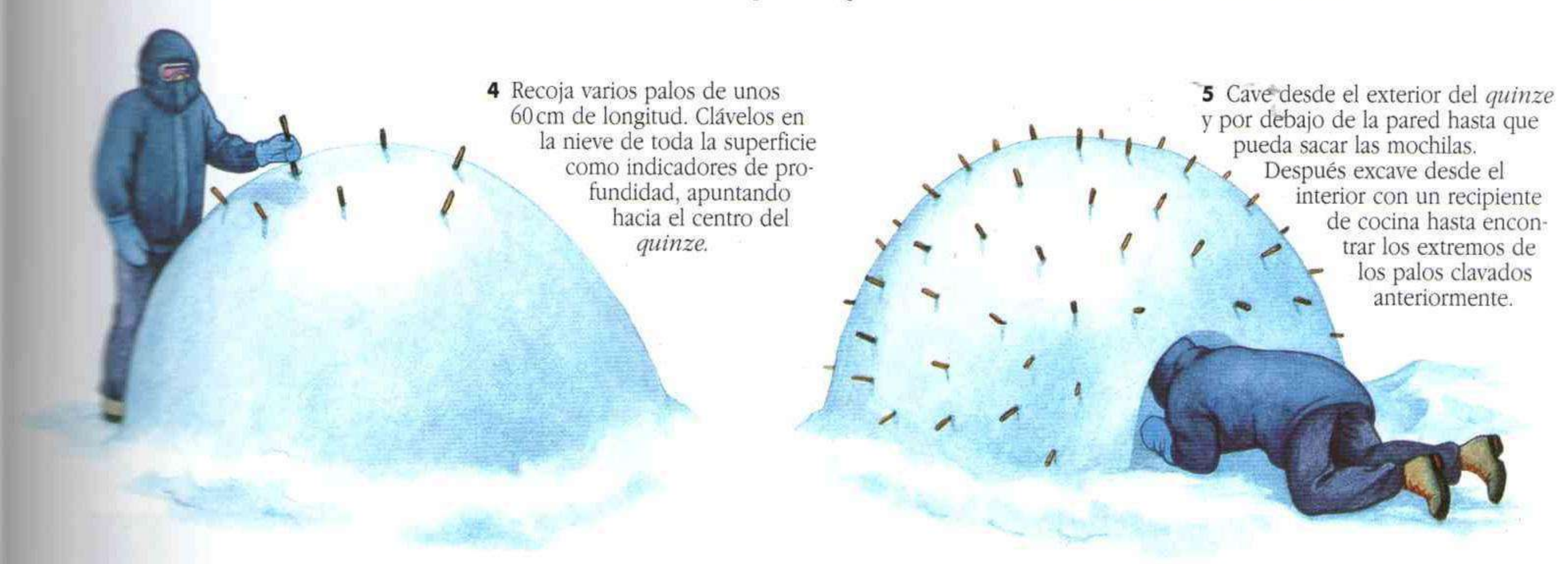

4 Recoja varios palos de unos 60 cm de longitud. Clávelos en la nieve de toda la superficie como indicadores de profundidad, apuntando hacia el centro del *quinze*.

5 Cave desde el exterior del *quinze* y por debajo de la pared hasta que pueda sacar las mochilas. Después excave desde el interior con un recipiente de cocina hasta encontrar los extremos de los palos clavados anteriormente.

Otros refugios en la nieve

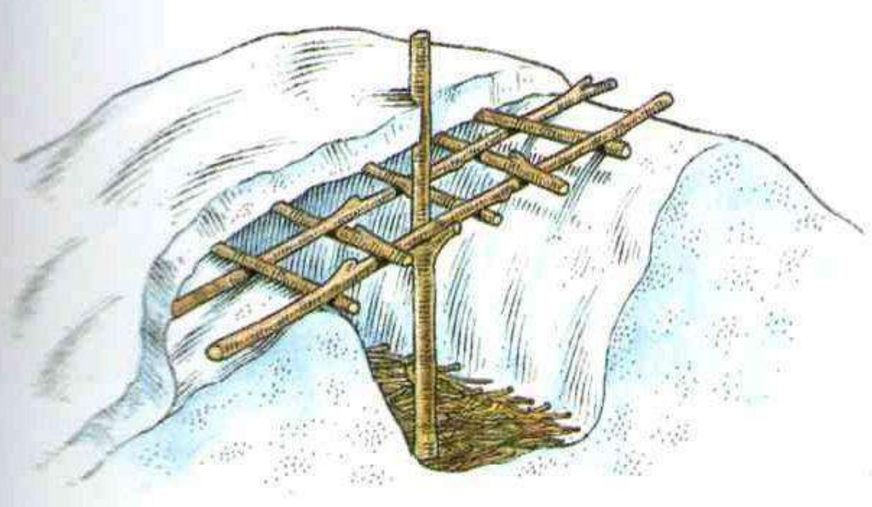

Trinchera. *Cave una trinchera y construya un techo entretejiendo ramas. Coloque nieve compacta sobre el armazón. No olvide colocar una rama vertical a través del techo para tener una ventilación adecuada en la trinchera.*

Cueva en la nieve. *En una emergencia, vacíe un banco de nieve y bloquee la entrada a su espalda con bloques de nieve. Haga un agujero en la pared para tener ventilación.*

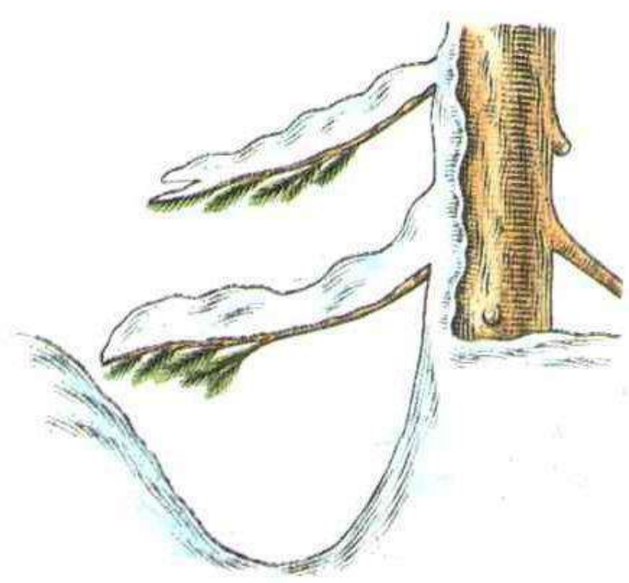

Resguardo natural. *Puede encontrar un refugio cómodo y casi listo para usar debajo de una conífera con ramas anchas y bajas. Tenga cuidado de no sacudir la nieve de las ramas.*

REFUGIOS TROPICALES

LA REGIÓN ENTRE el trópico de Cáncer y el trópico de Capricornio, unos 23 grados al norte y al sur del ecuador, se conoce como *los trópicos*. Esta región comprende selvas pluviales, pantanos y sabanas, que se caracterizan por su vegetación exuberante, las lluvias torrenciales y una abundante vida animal. Refugiarse de la lluvia, del viento y del sol resulta muy importante en todas las áreas tropicales. En las selvas pluviales montañosas, la noche puede ser bastante fría. En las selvas y pantanos, debe dormir a un nivel más elevado que el suelo, a ser posible lo suficientemente elevado para permitir el paso de animales pequeños por debajo, y para asegurarse de que la lluvia torrencial no le salpique al caer al suelo. Un techo impermeable y una mosquitera son también fundamentales. En las sabanas tropicales, dormir a un nivel más elevado que el suelo no es tan importante como en las selvas.

REFUGIO EN «A»

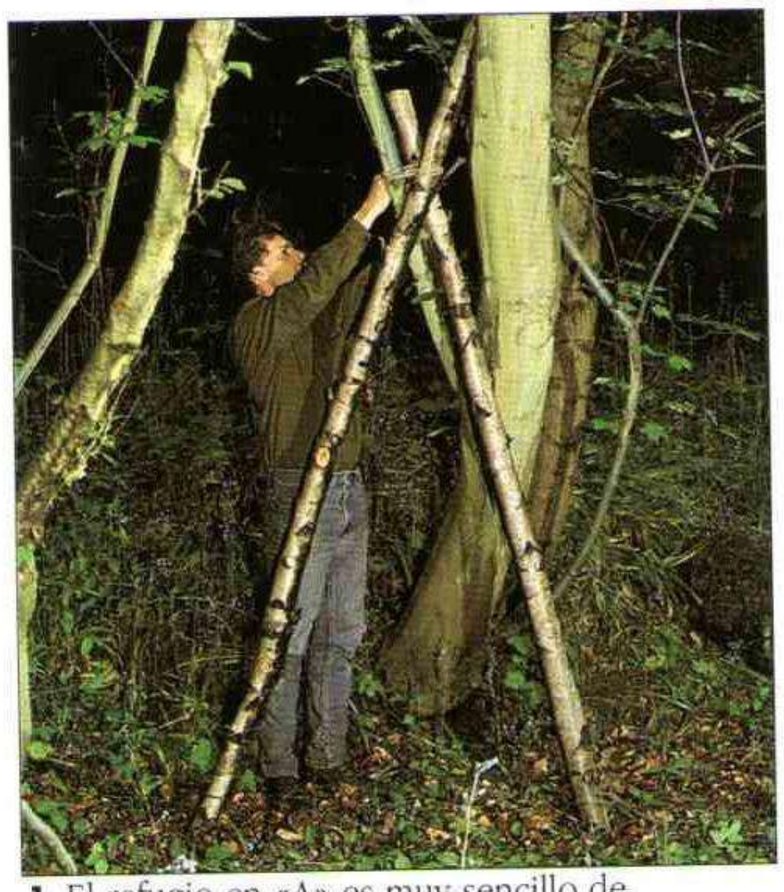

1 El refugio en «A» es muy sencillo de construir. Corte siete ramas largas y ate dos de ellas a un árbol.

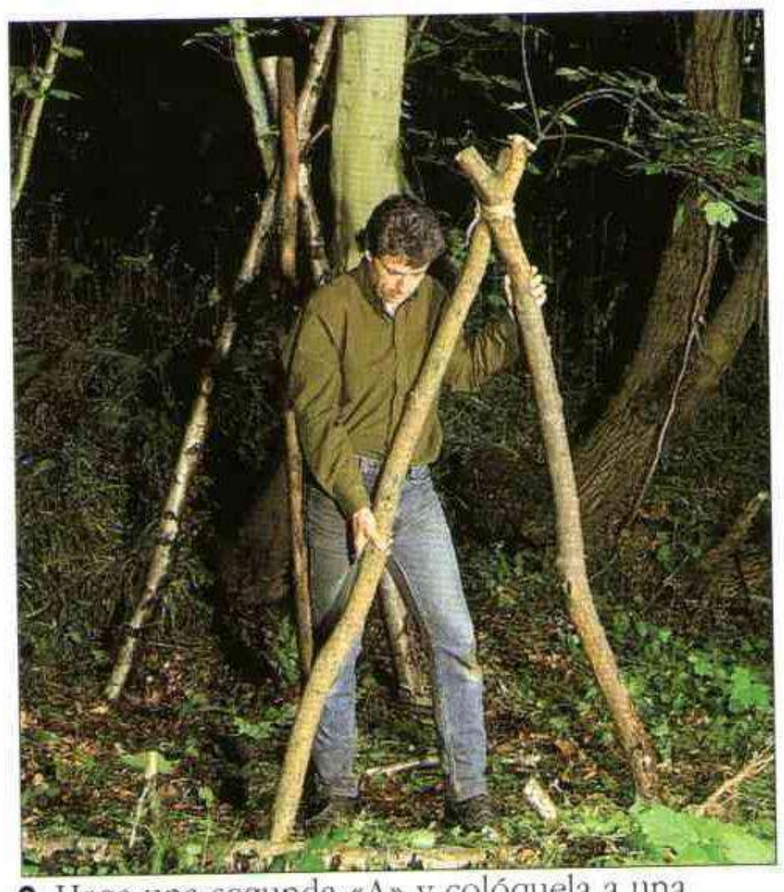

2 Haga una segunda «A» y colóquela a una distancia igual a la de su propia estatura más 60 cm, de la primera de ellas.

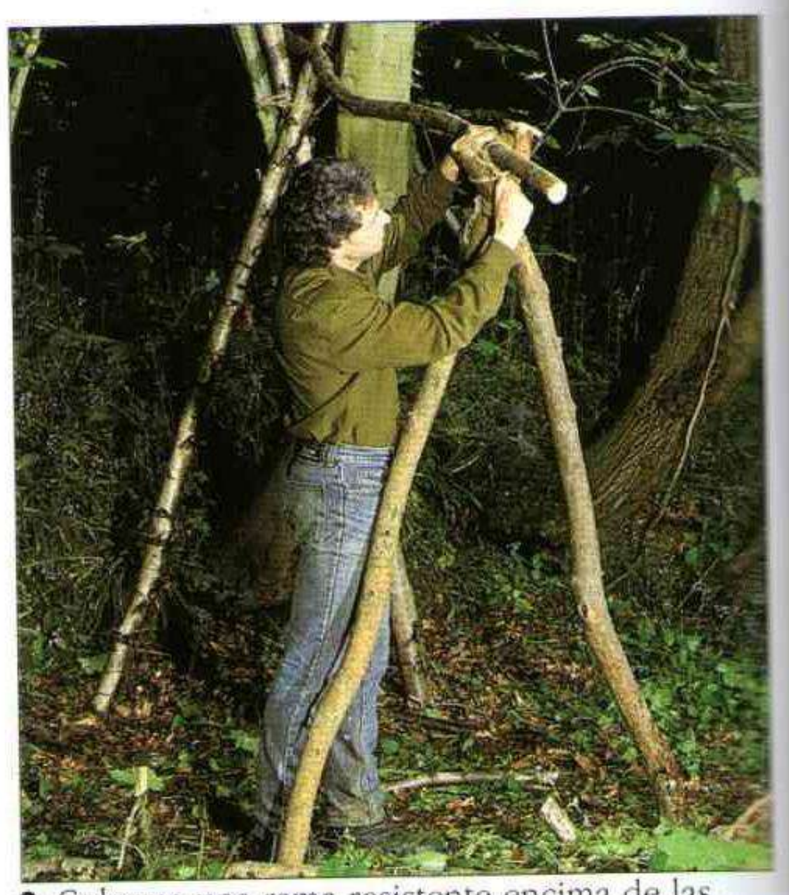

3 Coloque una rama resistente encima de las dos «V» superiores de cada estructura en «A». Esta rama actuará como travesaño.

4 Ate los lados de una tela impermeable o de una manta para formar un tubo. Inserte dos ramas largas y sepárelas hasta formar una «camilla». Encaje la camilla entre ambas estructuras en «A» con las ramas en el exterior.

5 Extienda una lona impermeable sobre el travesaño para construir el tejado. Ténselo a ambos lados y átelo a los árboles. El tejado debería protegerle de la lluvia y, al mismo tiempo, permitir la circulación del aire alrededor de su cabeza.

REFUGIO DE BAMBÚ

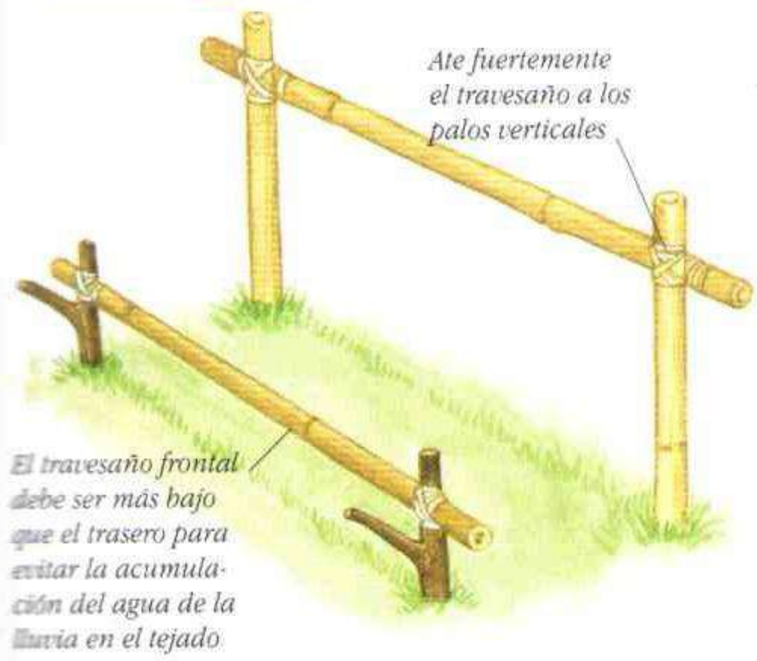

1 Este refugio puede construirse dondequiera que crezca el bambú. Coloque cuatro postes verticales gruesos en el suelo separados por una distancia igual a la de su estatura más unos 60 cm. Los postes frontales deben ser unos 30 cm más cortos que los traseros. Coloque los travesaños entre los postes verticales como se ilustra.

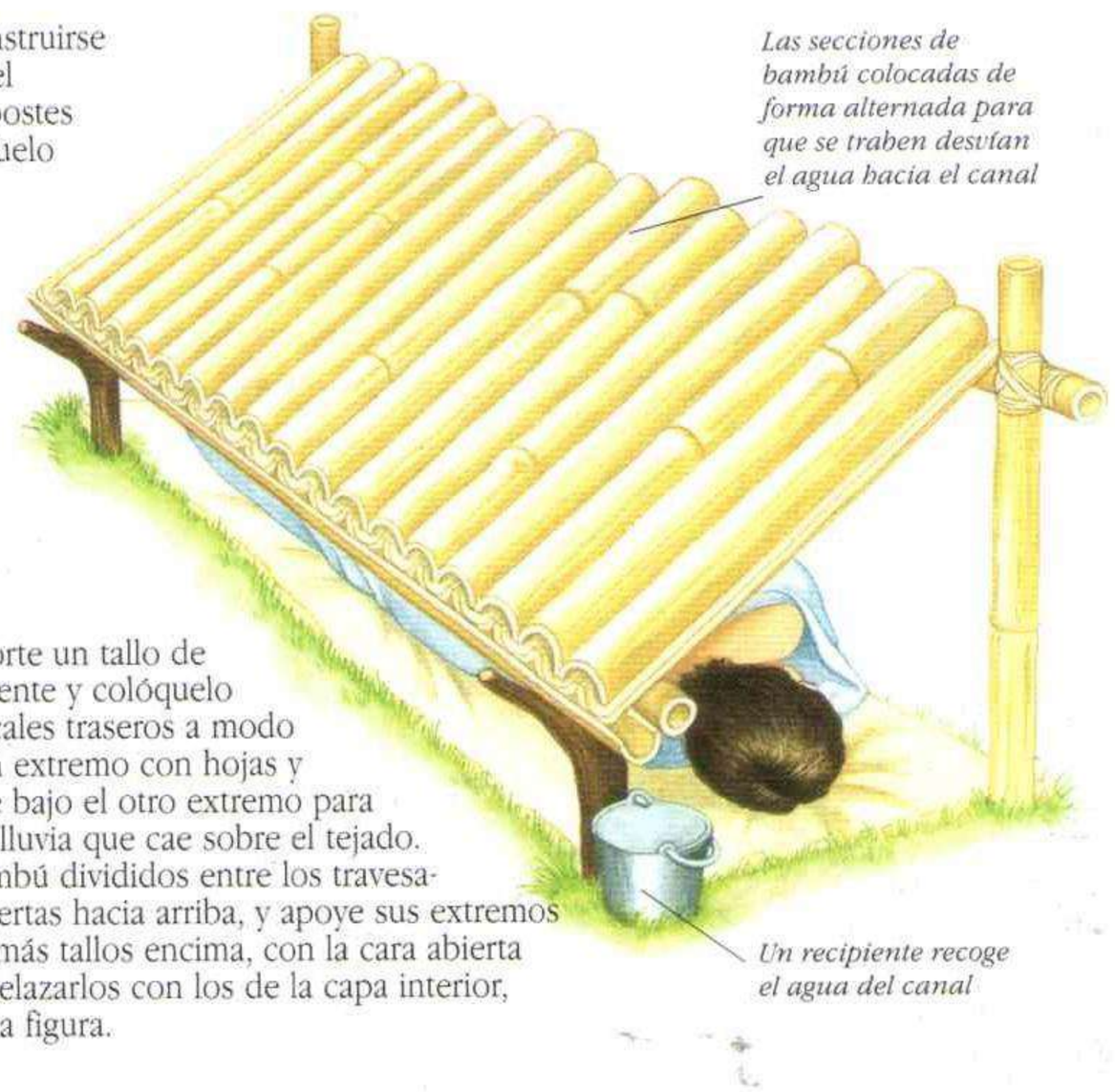

2 Con un cuchillo, corte un tallo de bambú longitudinalmente y colóquelo entre los postes verticales traseros a modo de canal. Obstruya un extremo con hojas y coloque un recipiente bajo el otro extremo para recoger el agua de la lluvia que cae sobre el tejado. Coloque tallos de bambú divididos entre los travesaños, con las caras abiertas hacia arriba, y apoye sus extremos en el canal. Coloque más tallos encima, con la cara abierta hacia abajo, para entrelazarlos con los de la capa interior, como se muestra en la figura.

PRECAUCIÓN

Los tallos de bambú pueden partirse fácilmente por el centro con un cuchillo, pero tenga cuidado al cortarlos para evitar la proyección de astillas, afiladas como cuchillas. Tenga cuidado al cortar el bambú, ya que algunos tallos pueden torcerse bajo presión, soltarse y cortarle.

REFUGIO EN LA SELVA

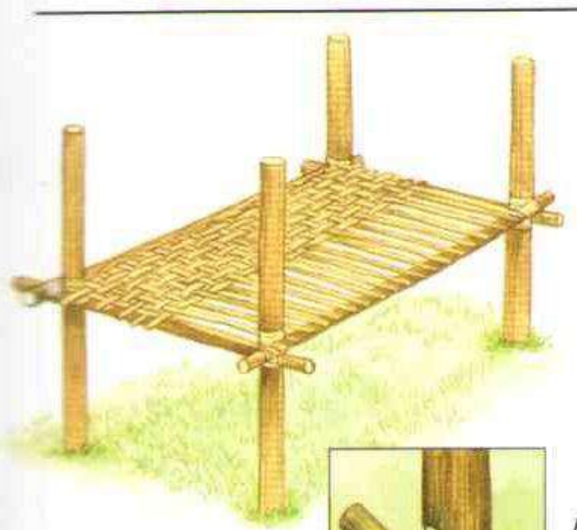

1 Para construir el armazón del refugio, clave cuatro troncos sólidos en el suelo, que harán de postes en las cuatro esquinas, separados por una distancia superior a la de su propia estatura. Coloque cuatro ramas horizontales, atándolas a la parte de los troncos verticales, a los que habrá hecho unas muescas para encajarlos mejor, y constituir así el armazón de la cama *(véase página 48)*.

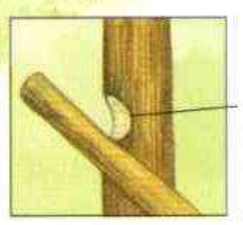

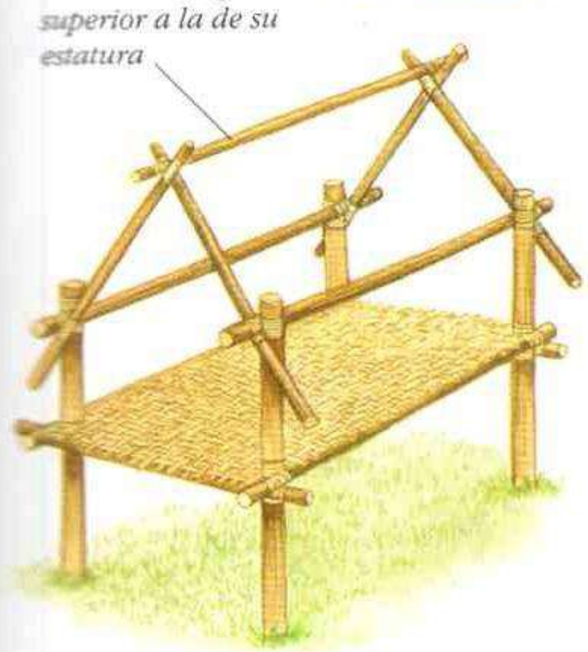

2 Ate siete ramas entre sí para construir la estructura del techo. El techo debe ser inclinado para no acumular el agua de la lluvia, y lo suficientemente resistente para soportar el peso de la cubierta. Átelo a los troncos verticales de las esquinas.

3 Ate más ramas entre los soportes del tejado y cúbralas con hojas grandes. Sobreponga las hojas contiguas y las de una fila con otra. Empezando por abajo, el pedúnculo de cada hoja debe apuntar hacia la parte superior, con el haz (cara brillante) hacia arriba para desviar el agua de la lluvia.

REFUGIOS PERMANENTES

EN UNA SITUACIÓN de supervivencia, desconoce el tiempo en el que permanecerá abandonado, o cuánto tiempo permanecerá en un mismo sitio. Por lo tanto, debe intentar construir el mejor refugio posible. Si se ve forzado a pasar todo un invierno, por ejemplo, necesitará un refugio más sólido, como una cabaña de troncos. Antes de dedicarse durante varios días, o incluso semanas, a construir un refugio permanente, debe asegurarse de que dispone del sitio más adecuado. Sus alimentos, combustibles y agua fresca deben encontrarse en las cercanías; el suelo, sólido, debe estar bien drenado; y el terreno protegido de cualquier peligro natural *(véase página 40)*.

CABAÑA DE TRONCOS

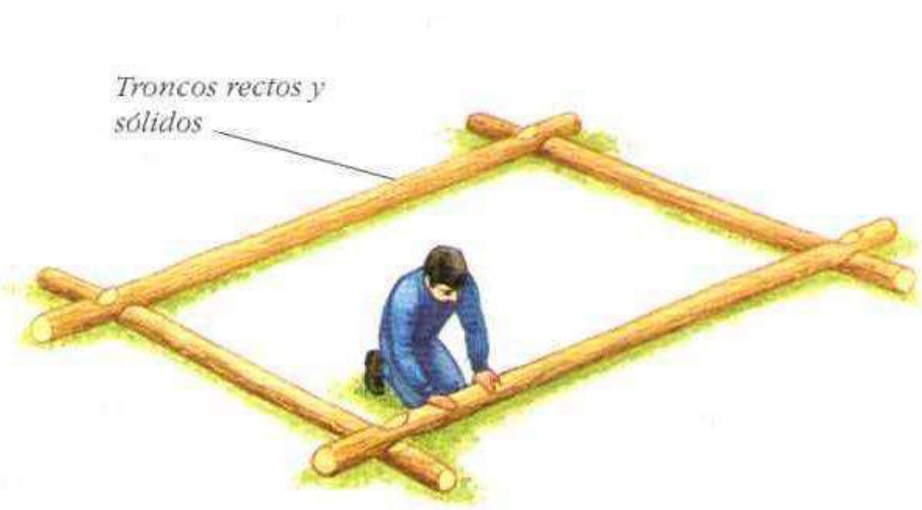

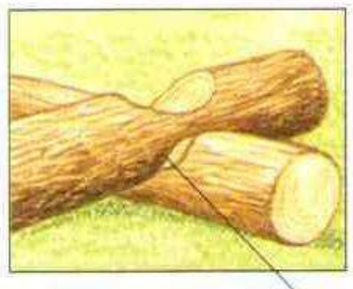

Para asegurar un buen encaje, debe hacer muescas a los troncos, en las puntas y a ambos lados

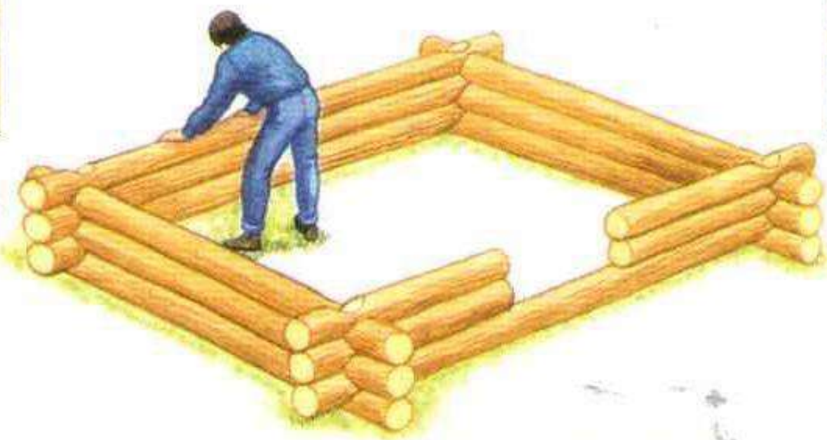

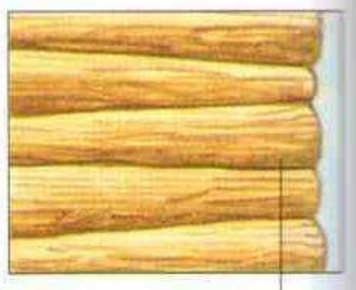

Coloque los troncos punta con base para que encajen y no dejen huecos

1 Limpie el suelo por lo menos hasta 1 m más allá de las paredes de la cabaña, y coloque los primeros troncos formando el contorno de la cabaña.

2 La clave de una buena construcción está en emplear troncos semejantes y regulares, y en hacer las muescas en los sitios correspondientes. Asegúrese de que con las muescas cada tronco queda bien colocado encima de los restantes, para que las paredes queden lo más perpendiculares posible, en ángulo recto con las esquinas. Construya las paredes gradualmente.

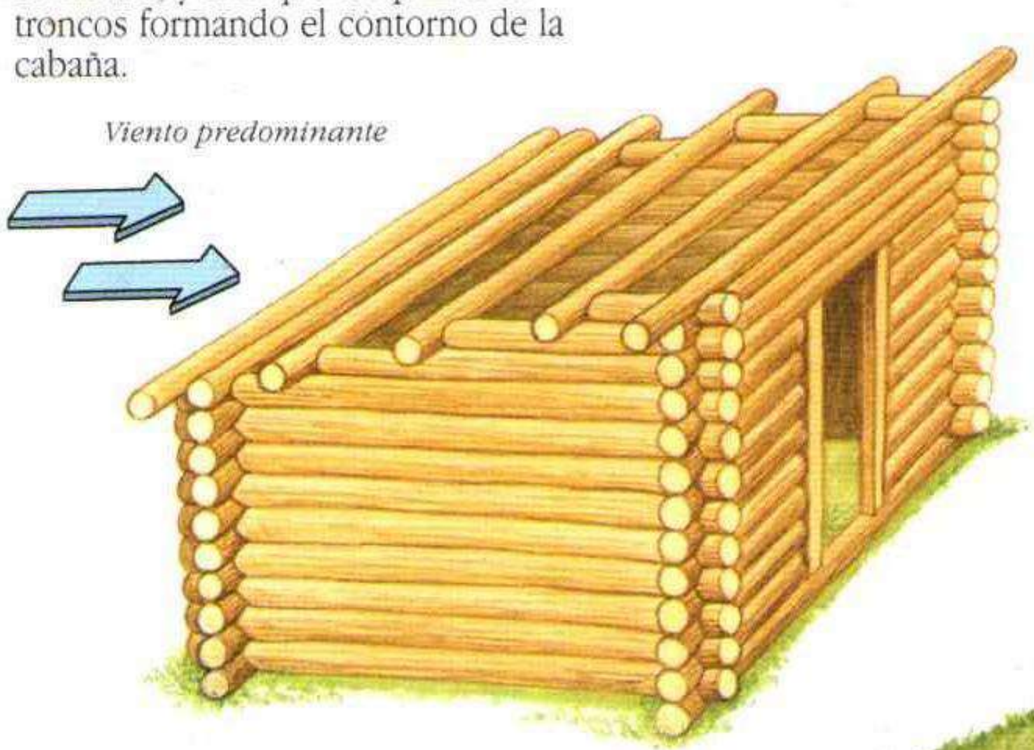

3 Las paredes traseras y la pendiente del tejado deben estar encaradas al viento y a las inclemencias del tiempo, mientras que la parte frontal con la entrada y cualquier ventana deben encararse hacia el sol. Incline el tejado de tal manera que la parte frontal sea más alta que la trasera, para que no se acumulen el agua de la lluvia o la nieve.

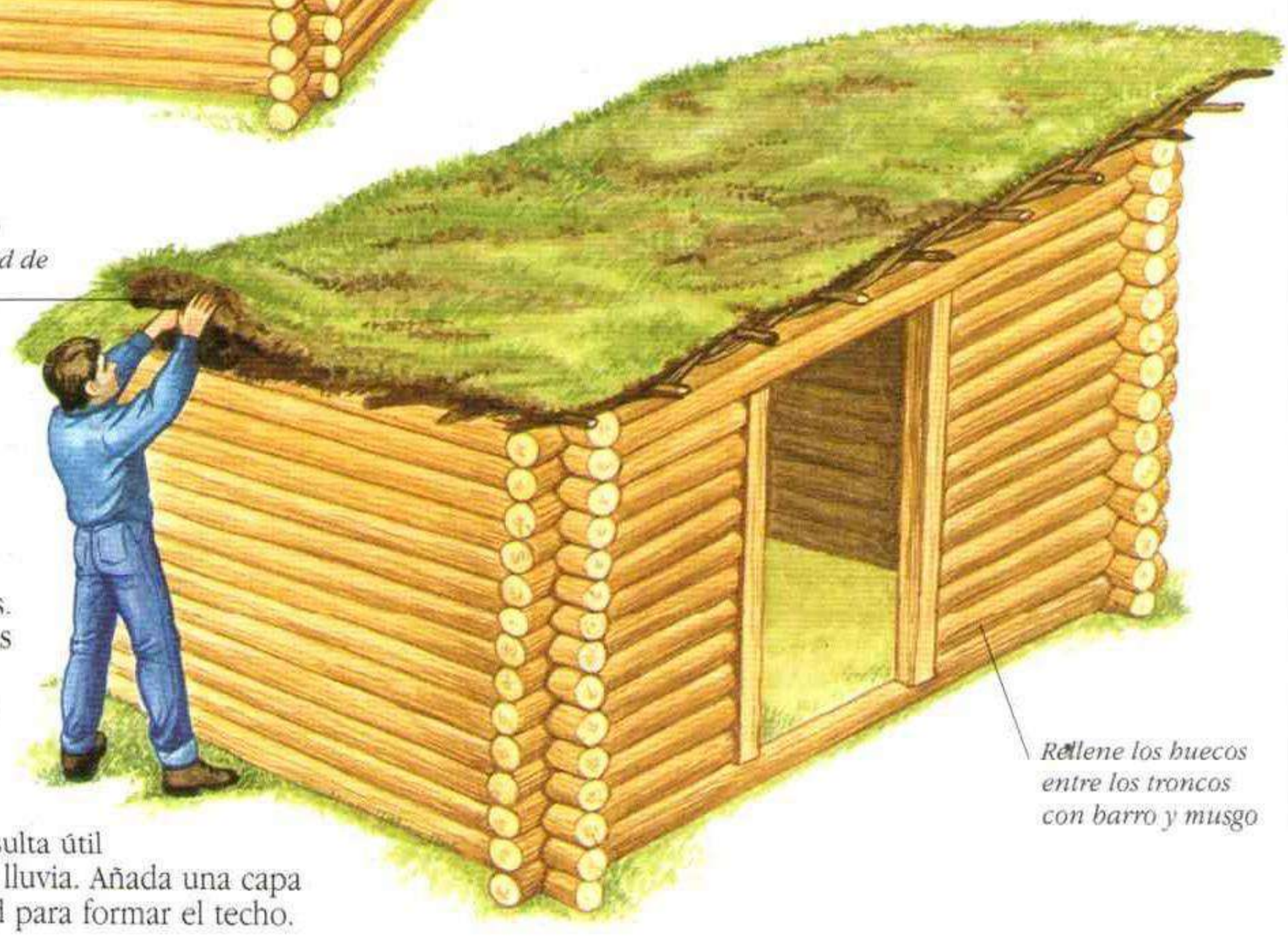

4 La cubierta del tejado puede ser de cualquier material. Coloque un aislante como helechos, hierbas, ramas de abeto, sobre un entramado de troncos. Si dispone de ellas, las bolsas de papel que contienen fertilizante son un aislante particularmente bueno; cualquier lámina de plástico resulta útil para protegerse de la lluvia. Añada una capa de bloques de césped para formar el techo.

UNA CHIMENEA DE PIEDRA

Debido al peligro de incendio, el hogar y la chimenea deben ser independientes y deben construirse de forma sólida, preferiblemente de piedra o ladrillo. La chimenea puede construirse fuera, con un agujero en la pared a manera de hogar, o bien en el interior, haciendo que la chimenea sobresalga del techo. Para construir la estructura, utilice rocas o ladrillos de barro.

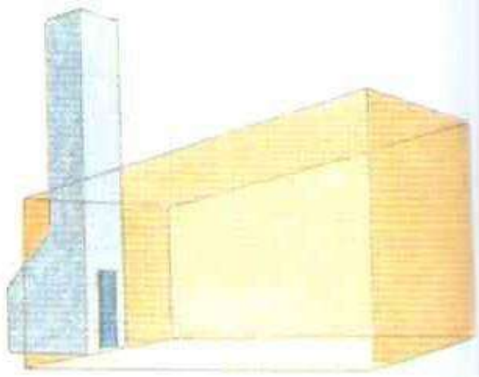

Chimenea de piedra. *Construya una chimenea de piedra fuera de las paredes de la cabaña para que el fuego no queme su refugio.*

CASA DE BLOQUES (TEPES)*

1 Los bloques de césped pueden utilizarse para construir como si se tratase de ladrillos sin secar, cortados y colocados sin mortero. Comience cortando una selección de bloques del suelo.

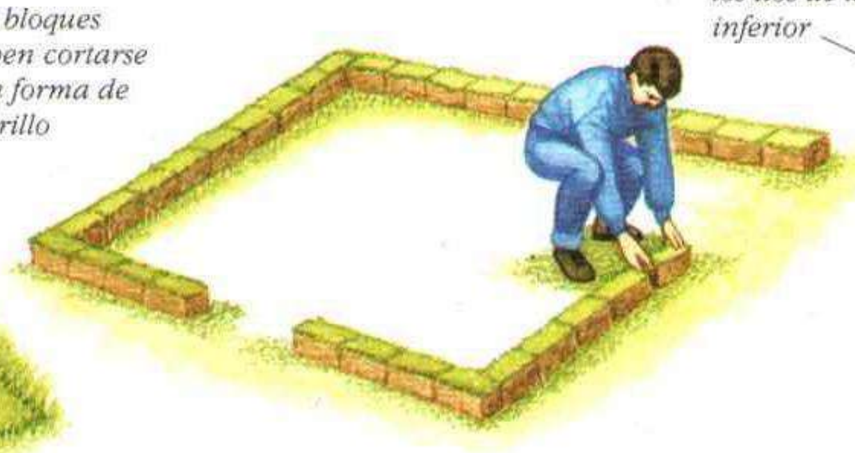

2 Coloque el contorno de la base de la cabaña sobre una porción limpia de terreno, dejando un espacio para la puerta. El suelo debe ser llano y sólido, y capaz de soportar el considerable peso de las paredes y del techo sin ceder. Comience a levantar las paredes.

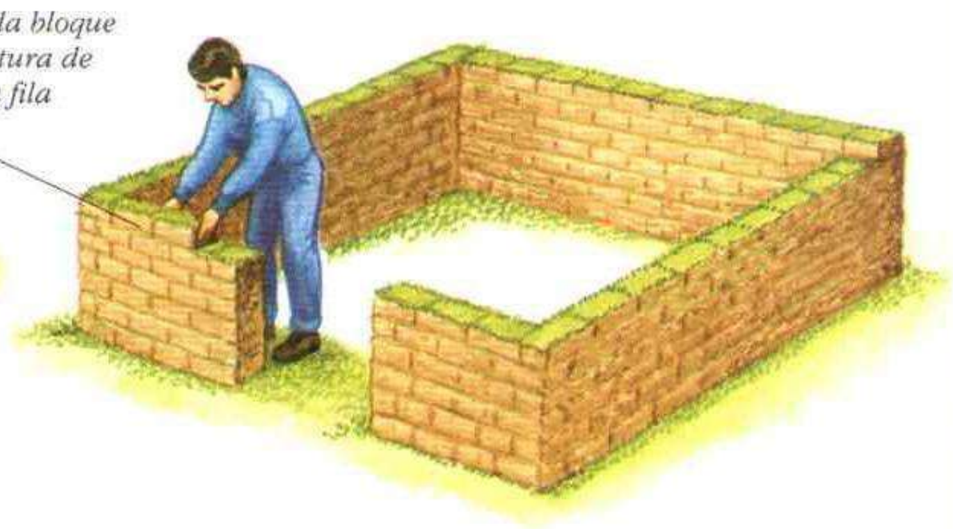

3 Construya gradualmente las paredes de la cabaña, solapando los bloques en las esquinas. Asegúrese de que cada nueva fila solapa la anterior para que las paredes sean firmes y no se derrumben. Intente hacer paredes rectas, a 90 grados una de otra, y perpendiculares.

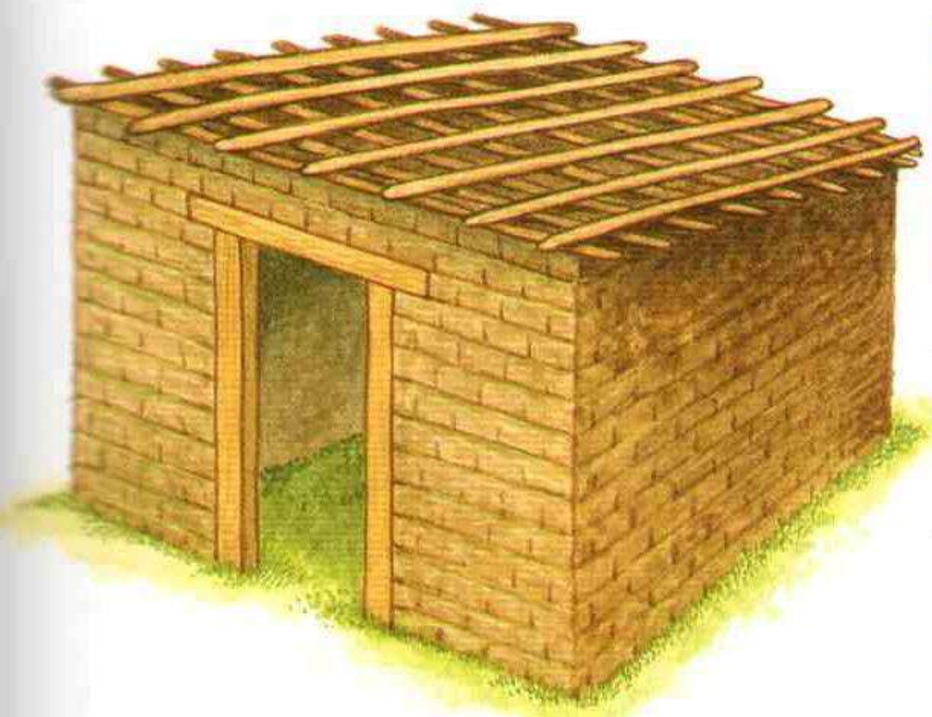

4 Construya la parte frontal más alta que la trasera, inclinando el tejado hacia donde sople el viento para evitar que se acumule el agua de la lluvia. Si es posible, construya una estructura de ramas para el techo.

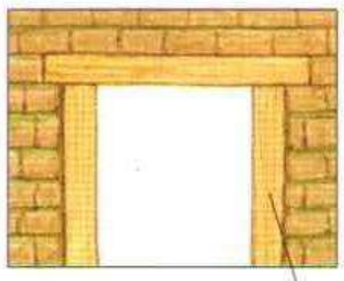

Añada un marco para la puerta hecho de madera, cuyo dintel debe ser lo suficientemente resistente como para soportar el peso de varias filas de bloques

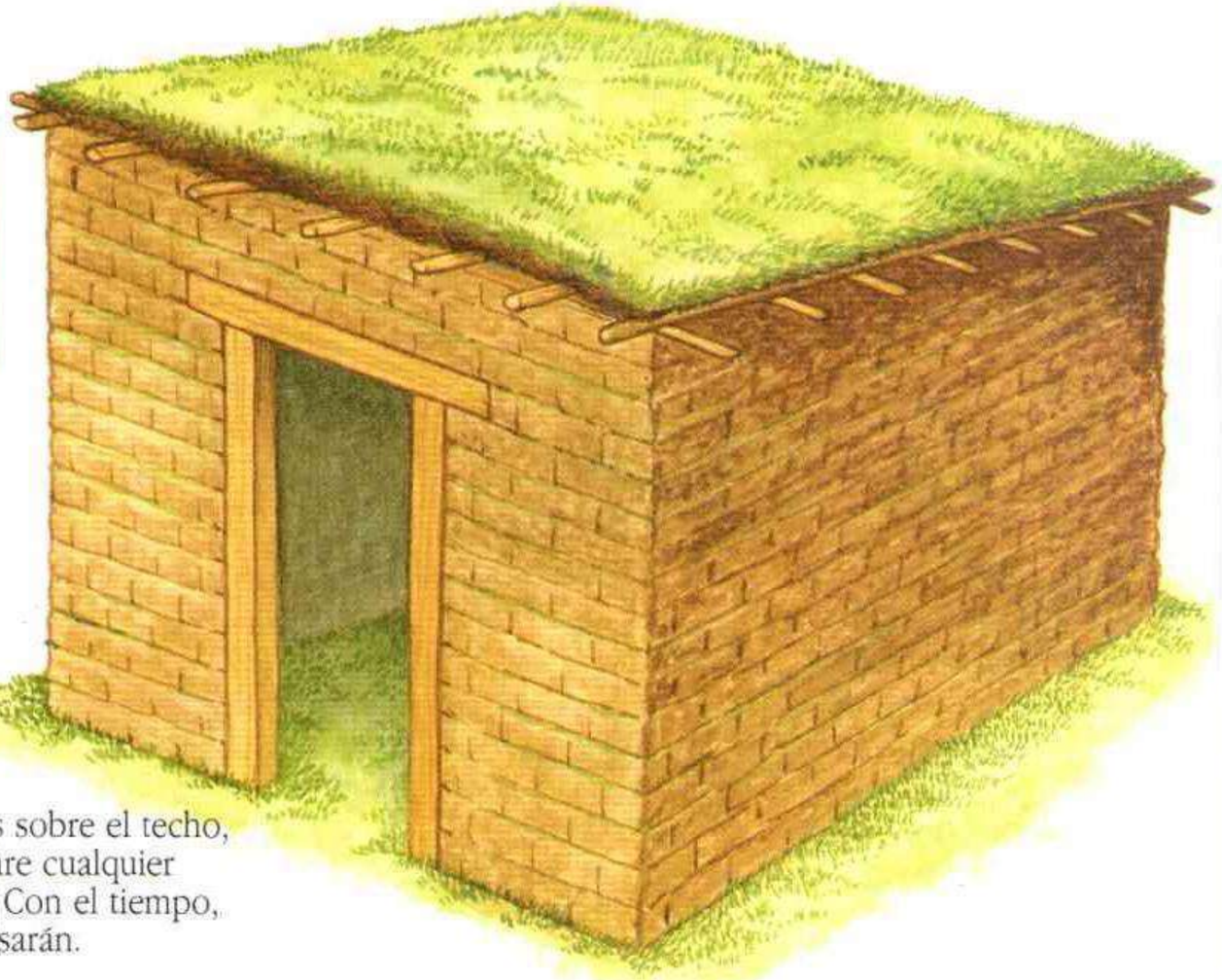

5 Coloque una capa de bloques sobre el techo, con el césped hacia arriba. Repare cualquier grieta en las paredes con barro. Con el tiempo, las paredes se endurecerán y alisarán.

OTROS REFUGIOS PERMANENTES

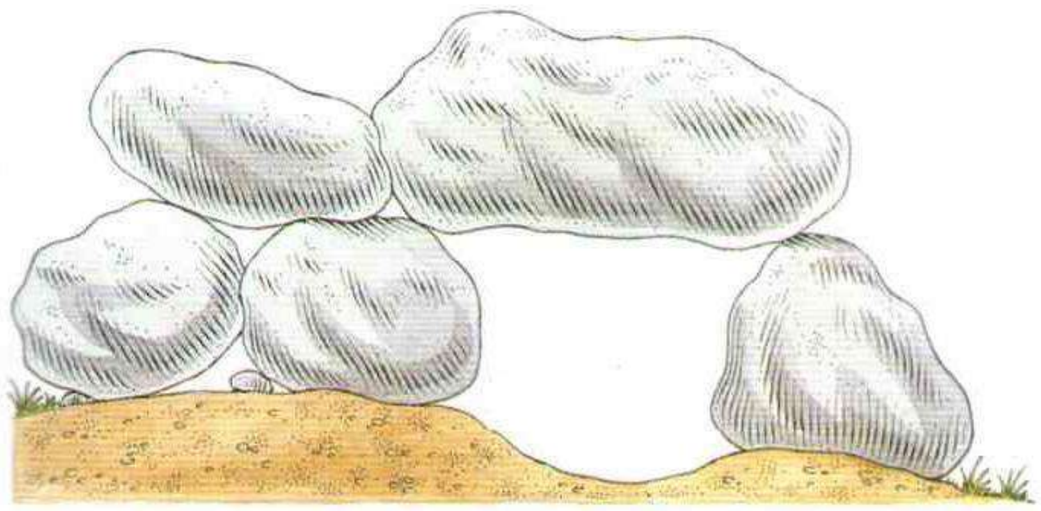

Huecos naturales. *En ocasiones pueden encontrarse refugios naturales bajo las grandes formaciones rocosas, pero debe tener cuidado y asegurarse de que la estructura es sólida e inamovible. Si excava el suelo por debajo, no toque los fundamentos de la estructura.*

Cuevas. *Cuando utilice una cueva como refugio, comience por explorarla en su totalidad. La entrada puede estar obstruida entre dos paredes. El fuego debe encenderse en el fondo de la cueva, de manera que el humo escape y no se acumule en el interior. Puede construir una pared de rocas o de troncos en el lado opuesto al fuego para reflejar nuevamente el calor hacia el interior de la cueva.*

* ***Advertencia:*** Este método destruye gran cantidad de cubierta vegetal, la cual tarda años en regenerarse. Por tanto, su construcción sólo debe emprenderse en caso de absoluta necesidad.

UN CAMPAMENTO SEGURO

UNA VEZ ELEGIDO el campamento, se deberán cumplir unas normas básicas de seguridad. En cada situación puede existir un peligro potencial, desde los fuegos y las cocinas sin ventilación hasta la falta de higiene básica. Nunca cocine dentro de una tienda, ya que puede incendiarse rápidamente y arder con fuerza. Cualquier tipo de llama dentro de una tienda reduce la cantidad de oxígeno disponible y puede generar gases venenosos, como el monóxido de carbono. Las tiendas y cualquier otro tipo de refugio deben estar bien ventilados *(véase página 114)*. Sin neveras y aseos, aumenta la posibilidad de sufrir una intoxicación por alimentos u otras enfermedades.

LA SEGURIDAD EN EL CAMPAMENTO

Si acampa por más de una noche, debe vigilar que su campamento sea lo más seguro posible. Los accidentes ocurren a menudo por la noche, por ejemplo, cuando las personas salen de su tienda para ir a la letrina y tropiezan con el fuego, caen al río o se pierden en la oscuridad sin ser capaces de volver a encontrar su tienda.

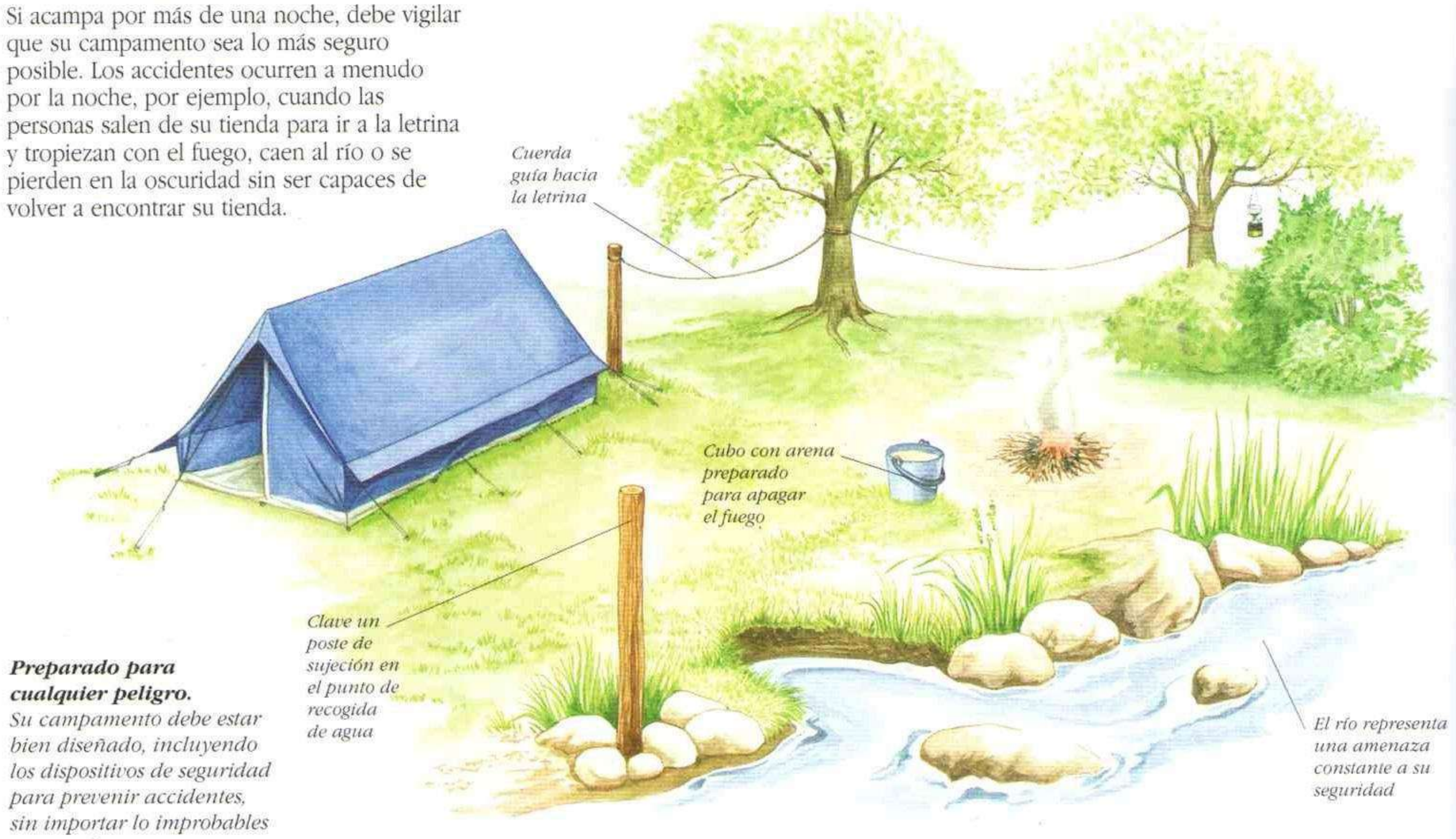

Preparado para cualquier peligro. *Su campamento debe estar bien diseñado, incluyendo los dispositivos de seguridad para prevenir accidentes, sin importar lo improbables que puedan parecer.*

LA CONSTRUCCIÓN DE UNA LETRINA

La letrina debe colocarse a favor del viento y más abajo de su tienda. Espolvoree un poco de tierra cada vez que la use. Nunca vierta sustancias químicas a una letrina, ya que evitan la descomposición natural de la materia fecal y la hacen oler mucho peor de lo que lo haría naturalmente. Construya un punto diferente para orinar, marcado por una estaca, a un lado. No se olvide de rellenar la letrina cuando marche *(véase página 64)*.

Inodoro de campamento. *Construya una letrina si debe permanecer en el campamento varios días, o si se trata de un grupo numeroso. Utilícelo únicamente para los desechos sólidos.*

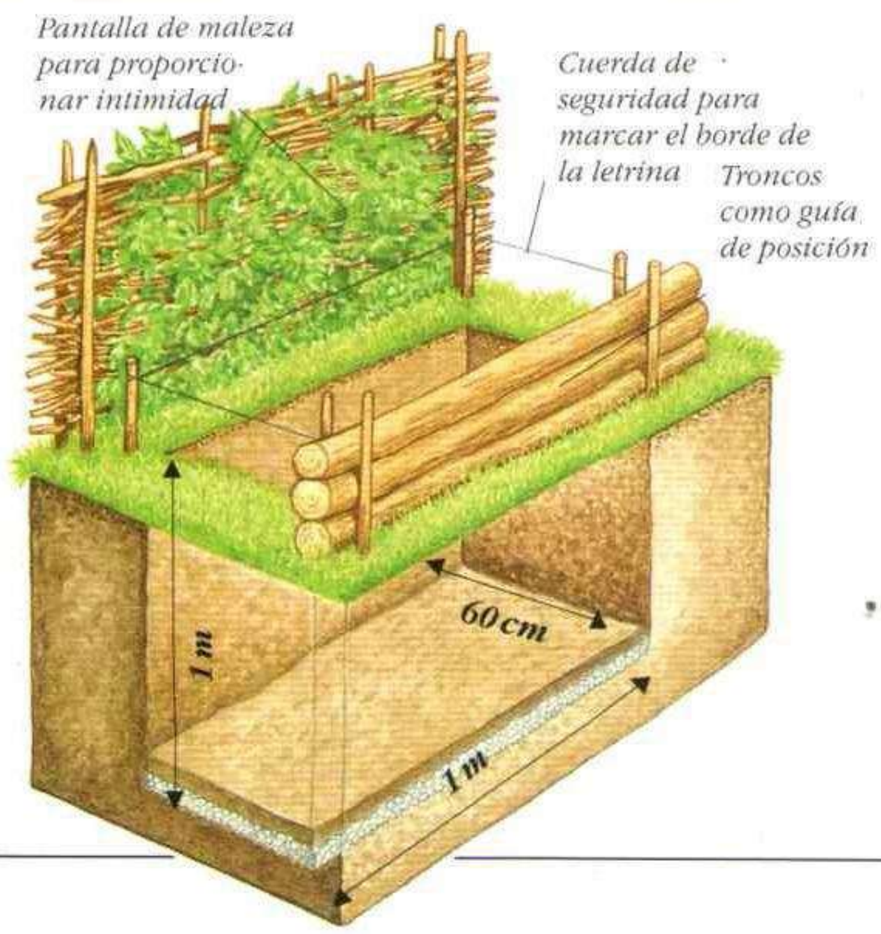

LATA COMO INODORO

Allí donde sea ilegal o resulte poco práctico enterrar desechos humanos, construya un inodoro portátil colocando una bolsa de plástico dentro de una lata de café con tapa. Eche un poco de lejía en la bolsa para reducir el olor. Utilice la lata para los desechos sólidos.

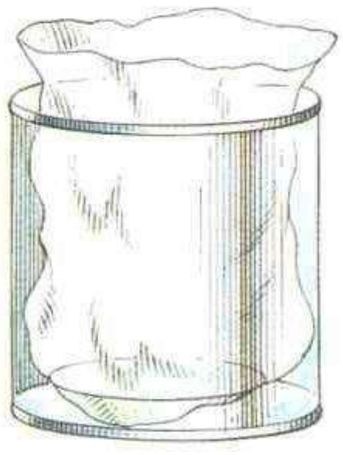

INODORO PORTÁTIL

El lavado de la ropa

Si su ropa está limpia, retendrá su capacidad aislante, y también contribuirá a su higiene personal. En las regiones tropicales, la ropa debe lavarse cada día, guardando una muda para llevar por la noche. En los climas fríos, únicamente debe lavar regularmente los calcetines y la ropa interior. La ropa debe remendarse tan pronto se rompa *(véase página 31)*, para que el roto no se haga mayor y la ropa resulte inservible. La ropa limpia y remendada ayuda a mantener la moral en las situaciones difíciles.

Higiene personal

Guarde sus artículos de aseo personal en un neceser impermeable para evitar que se pierdan. Si utiliza una bolsa también los mantendrá limpios. Mientras utilice sus artículos de aseo, cuelgue el neceser de un árbol para acceder rápidamente a cada artículo. Una toalla grande puede tener muchas aplicaciones. También puede cortarse en trozos menores para utilizarla de varias maneras *(véase página 27)*.

Neceser

Toalla

Limpiadores improvisados

Hoja del castaño de Indias. *Las hojas del castaño de Indias trituradas en agua caliente pueden utilizarse como jabón. También tienen un suave efecto antiséptico.*

Fresa. *Puede eliminar manchas de los dientes frotando fresas sobre ellos. La ceniza de madera también puede emplearse como pasta de dientes, o como jabón.*

La protección del medio ambiente

- Evite que la espuma del jabón, los restos de comida o cualquier otro contaminante lleguen a su fuente de agua potable. Recuerde que, aunque lave sus platos y se asee corriente abajo, puede haber otros campamentos río abajo.
- Evite dejar una tienda levantada en el mismo lugar durante un tiempo prolongado para permitir el crecimiento de la hierba.
- No corte los árboles y las plantas a menos que deba construir un refugio.
- Utilice los hoyos para fuego existentes cuando sea posible, antes de hacer uno nuevo, y tenga cuidado de no provocar un incendio.
- Queme toda la basura y llévese el residuo cuando se marche.
- Al abandonar el campamento, recuerde rellenar la letrina y el hoyo del fuego, y eliminar cualquier rastro de su estancia *(véase página 64)*.

Mantenga su comida a salvo

1 Cuando prepare la comida, quizá quiera que alguno de los ingredientes se mantenga fresco y alejado de los insectos y animales. Construya un contenedor temporal a partir de un rollo de tela, atando ambos extremos para cerrarlo. Coloque un plato en el interior, que hará las veces de soporte.

2 Cuelgue la tela de una rama. Guarde la comida dentro del contenedor durante un período corto. Los alimentos enlatados se mantienen frescos metiéndolos en la aguas poco profundas de los ríos.

EL FUEGO EN EL CAMPAMENTO

EL HECHO DE tener o no un fuego en el campamento puede determinar su supervivencia, ya que el fuego le proporcionará calor, un medio para cocinar sus alimentos, y le servirá también de estímulo. El fuego aleja a los animales salvajes y a los insectos; da calor, con lo que evita el consumo de calorías; convierte los alimentos incomibles en una deliciosa comida, y seca la ropa y las botas. Necesitará tres ingredientes para encender un fuego: yesca, astillas y leña. Un fuego también necesita oxígeno para quemar. La mejor fuente de combustible seco es la madera muerta que todavía no ha caído. Si debe usar leña caída, no recoja la que está en contacto con el suelo, sino las ramas que lo cubren.

TIPOS DE YESCA

Musgo. *El musgo verde y húmedo no sirve como yesca. El musgo muerto y seco es fino y denso, ideal para comenzar un fuego. Puede encontrar musgo junto a los troncos de los árboles o en terrenos pantanosos. Quizá deba secarlo antes de su uso.*

Hongos. *Algunos hongos que crecen en los árboles tienen una piel exterior impermeable y una carne suave, seca y esponjosa en el interior, que constituye una yesca excelente. Probablemente deba eliminar la piel exterior con un cuchillo para acceder al suave tejido interior.*

Corteza. *Incluso en climas húmedos, la corteza interior de los troncos caídos puede estar seca. Elimine los trozos húmedos o busque allí donde los insectos se han introducido produciendo en la madera un fino serrín, que es igualmente una buena yesca. La madera podrida también puede emplearse como yesca.*

Hojas muertas. *Siempre hay hojas secas y muertas en el suelo, incluso en los bosques más húmedos. Meta las hojas directamente en una bolsa impermeable según las vaya recogiendo. Puede utilizarlas enteras o estrujarlas para formar trozos pequeños antes de utilizarlas como yesca.*

Hierba seca. *Es muy fácil de recoger. Un manojo de hierba seca se encenderá rápidamente produciendo llamas* (véase inferior).

PREPARAR LA YESCA

Pulverizar. *Puede convertir los palos secos y los trozos de corteza en yesca en polvo cortándolos en trozos muy pequeños con un cuchillo. Corte los trozos lo más pequeños posible, y guárdelos en una bolsa.*

Cortar. *Haga una pequeña depresión en la superficie de la yesca, por ejemplo, en un trozo de hongo de árbol. Si deja caer un rescoldo incandescente en este hueco, éste encontrará una gran área superficial para calentar.*

Descarnar. *Para arder, incluso con la chispa más pequeña, la yesca debe ser tan fina como el algodón. Frote y retuerza el material entre sus dedos, o usando una piedra, para separar las fibras entre sí.*

LEÑA

Leña para calentar. *Las maderas blandas como la del fresno, el abeto, el manzano, el avellano y el acebo, arden rápidamente y con mucha luz, desprendiendo mucho calor. Sin embargo, también producen chispas. Las maderas que producen una mayor cantidad de chispas al calentarse son las del cedro y la cicuta. Al arder tan rápidamente, las maderas blandas son ideales para comenzar un fuego, y pueden resultar útiles para cocinar rápidamente o, por ejemplo, para calentar agua u otros líquidos. Sin embargo, se consumen muy rápido y dejan mucha ceniza en lugar de rescoldos o brasas que podrían emplearse para hornear o para otras formas de cocina lenta.*

ACEBO

ABETO

Leña para cocinar. *Las maderas densas y duras como las de la haya, el roble, el abedul, el alerce, el nogal y el sicómoro, arden lenta y uniformemente, desprendiendo mucho calor y produciendo brasas ideales para cocinar lentamente* (véase página 117). *El sabor de la comida cocinada sobre un fuego puede quedar afectado por el olor del humo. Cada tipo de madera produce un olor diferente al quemarse. Por ejemplo, el pino tiende a dar a los alimentos un sabor a resina, mientras que el manzano y el sicómoro pueden enriquecer su sabor.*

ROBLE

HAYA

Combustible de emergencia. *En una situación de supervivencia puede verse en la necesidad de tener que emplear cualquier cosa que encuentre para quemar. Los combustibles de emergencia pueden incluir excrementos animales, líquenes secos, musgo, brezo e incluso bloques de turba; todos deben secarse al sol antes de su uso. En la playa, puede servirse de algas secas. A veces, puede encontrar restos de carbón para quemar, mientras que el petróleo se filtra ocasionalmente formando charcos en la superficie del terreno. En ocasiones, la arena contiene petróleo, y puede quemarse. También puede utilizar la grasa animal como combustible.*

EXCREMENTOS SECOS

ALGAS

MADERAS QUE DEBE EVITAR

Algunas maderas resinosas estallan con fuerza al fuego, y deben evitarse dentro de lo posible. Éstas incluyen el pino y el endrino. Otras maderas como el aliso, el castaño, el álamo y el sauce no queman bien y se limitan a arder sin llama. El bambú estallará sin calentarlo, a menos que lo abra primero.

ÁLAMO

SAUCE

INGREDIENTES PARA UN FUEGO

Debe hacer el fuego gradualmente, comenzando por astillas de madera y progresando hacia las ramas mayores y los troncos. Clasifique su madera en yesca, un puñado de ramitas muy finas y una gran cantidad de ramas pequeñas, ramas grandes y troncos. Deje que cada etapa arda bien antes de añadir ramas más grandes. No utilice ramas verdes, ya que éstas tienden a doblarse.

Yesca. *Necesitará un montón de yesca, al menos del tamaño de un pomelo, cortada muy fina* (véase página anterior). *La yesca es la parte más importante de un fuego, ya que no puede encender un fuego comenzando por los troncos gruesos, a menos de que disponga de un encendedor que sustituya a la yesca* (véase página 62).

Astillas. *Una vez lograda la llama en la yesca, añada leña menuda seca en forma de pequeñas ramas y hojas. Asegúrese de que la leña está totalmente seca. Las ramas deben tener el grosor de un lápiz.*

Leña. *Cuando las astillas ardan correctamente, añada ramas del grosor de un dedo. Las ramas harán que el tamaño de las llamas aumente, desde su fase inicial hasta el tamaño normal para un fuego.*

Leña gruesa. *Las ramas grandes actúan como combustible principal. Deben ser más gruesas que un dedo, pero fáciles de romper. El tamaño de este combustible es el más habitual en los fuegos de campamento.*

Troncos. *Los troncos gruesos son ideales para mantener encendido un fuego estable durante toda la noche. Asegúrese de que están totalmente quemados antes de apagar el fuego.*

CÓMO PREPARAR UN FUEGO

EL SECRETO DE un buen fuego está en disponer de combustible seco y encenderlo poco a poco a partir de una llama pequeña. Debe escoger y construir su fuego con gran cuidado, eligiendo el más adecuado para sus propósitos, al tiempo que se adapta a las condiciones locales *(véase página siguiente)*. Los fuegos son potencialmente peligrosos –puede perder el control del mismo, incendiando la vegetación seca, la tienda o su ropa. Asimismo, pueden gastar todo el oxígeno disponible en un refugio cerrado asfixiando a sus habitantes. Un fuego también puede marcar el paisaje si no se apaga y limpia convenientemente *(véase página 65)*.

FUEGO SENCILLO EN ARMAZÓN DE TIPI

1 Levante un cuadrado de hierba y déjelo a un lado. Coloque una plataforma de ramas verdes en el hueco.

2 Comience a construir un armazón de tipi apoyando cuatro ramas verticales entre sí, haciendo que sus extremos superiores coincidan en un punto. Este armazón no tiene que ser muy grande.

PRECAUCIÓN

No construya un fuego donde la vegetación esté seca debido a la sequía. No construya un fuego de dimensiones mayores de las que pueda controlar, especialmente si hace viento. Tenga cuidado de no construir un fuego debajo de ramas bajas que podrían incendiarse.

3 Construya un tipi gradualmente, intentando hacerlo lo más robusto posible. Asegúrese de dejar el suficiente espacio debajo del tipi para colocar la yesca, y no olvide dejar una abertura en uno de los lados.

4 Coloque la yesca dentro del tipi terminado. Encienda la yesca *(véase página 58)*, y después añada hojas secas y pequeñas ramas para aumentar el tamaño de las llamas dentro del tipi. Conforme se acumula el calor, el tipi también comienza a quemarse, dando lugar a llamas grandes. El tipi se derrumbará sobre sí mismo, y obtendrá así un lecho de brasas al que puede añadir más leña, o utilizarlo para cocinar. La plataforma de madera también llegará a quemarse, originando más brasas.

Tenga ramitas y hojas a mano para alimentar el fuego

TIPOS DE FUEGO

Aunque puede utilizar el fuego sencillo en armazón de tipi en casi cualquier área, puede encontrarse con que deba adaptarlo a ciertas condiciones especiales. Por ejemplo, si hace mucho viento, puede construir el fuego en una trinchera para que esté protegido y arda correctamente. En la nieve, tendrá que escarbar hasta llegar al suelo sobre el que colocará la plataforma para construir el fuego.

Fuego de trinchera. *En condiciones de viento fuerte, cave una trinchera de unos 30 cm de profundidad. Los fuegos de trinchera son excelentes para cocinar, y tienen un bajo consumo de leña. No llegan a producir grandes llamaradas pero, a pesar de su construcción, disponen del aire suficiente para arder correctamente.*

Fuego en una pared rocosa. *Una pared rocosa protegerá el fuego del viento, impedirá las llamaradas peligrosas y limitará el consumo de leña. Elija rocas secas, no porosas, y evite la pizarra, que puede explotar al calentarla. Utilice las rocas más estables y de mayor tamaño que pueda encontrar. Puede rellenar los huecos con barro o arcilla para formar un área funcional para cocinar.*

La pared rocosa protege el fuego del viento

Fuego de estrella. *Una vez que su fuego dispone de un buen lecho de rescoldos o brasas, coloque cuatro troncos grandes en el centro. Conforme se van quemando, debe irlos empujando hacia el centro. Utilice este tipo de fuego únicamente en un campamento a largo plazo, para asegurarse de quemar los troncos en su totalidad antes de apagar el fuego* (véase página 64).

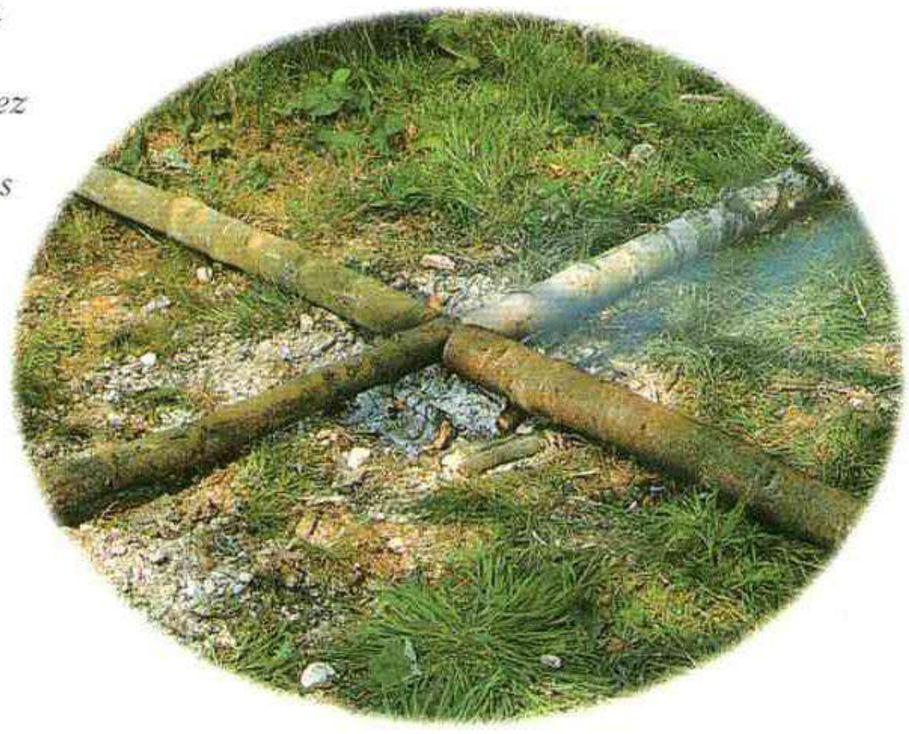

FUEGO DE LARGA DURACIÓN

Para que un fuego dure toda una noche, o todo un día, coloque tres troncos gruesos muy juntos sobre un lecho profundo de brasas. Si quiere calor adicional, puede construir una pared de madera u otro material detrás del fuego para reflejar el calor hacia usted.

Fuego de larga duración. *Un tronco largo puede arder toda la noche.*

CÓMO TRANSPORTAR EL FUEGO

1 Puede mantener vivos los rescoldos incandescentes varios días para comenzar un nuevo fuego dondequiera que esté. Anude un asa de cuerda o alambre a la lata.

Coloque el musgo seco en la parte superior

2 Llene la lata con musgo seco. Si únicamente puede encontrar musgo húmedo, puede utilizarlo, siempre que lo coloque en la parte inferior y lo cubra con musgo o hierba seca. El musgo húmedo se secará pronto.

3 Coloque unos rescoldos incandescentes en el nido de musgo y cúbralos con más musgo seco. Sople suavemente sobre los rescoldos cuando le parezca que vayan a apagarse.

Las brasas vivas hacen que el musgo se encienda

CÓMO ENCENDER UN FUEGO

JUNTO CON LOS primeros auxilios, saber encender un fuego es la habilidad de supervivencia más necesaria. Cuando la madera u otro combustible se calienta, despide gases que se encienden originando llamas. Los fuegos deben encenderse en pequeña escala y reforzarse gradualmente. Es muy importante disponer de leña seca, ya que la madera húmeda apagará prácticamente cualquier fuego. Los fuegos no sólo necesitan calor y combustible seco, sino también una gran cantidad de oxígeno, especialmente al encenderlos. Puede apagar un fuego cubriéndolo con tierra para impedir que las llamas tengan el suficiente oxígeno para continuar ardiendo.

MÉTODOS PARA ENCENDER UN FUEGO

Pedernal y acero. *Haga un nido de yesca junto al tipi. Frote el pedernal con el acero, encima de la yesca. Sople sobre las chispas para crear una llama. Una vez ha encendido la yesca, empújela dentro del tipi.*

Frote el pedernal con la sierra de acero dirigiendo las chispas hacia la yesca

CONSEJOS

- Asegúrese de que el fuego está protegido del viento *(véase página 60)*; recuerde que un fuego necesita oxígeno para arder.
- Mantenga una reserva de yesca seca.
- Reúna todos los materiales necesarios antes de encender un fuego.
- Impermeabilice las cerillas dejando caer la cera de una vela sobre sus cabezas. Elimine la cera antes de encenderlas.

Cerillas. *Encienda una cerilla protegiéndola del viento con las manos, y permita que la llama queme el palo. Acerque la cerilla a la yesca hasta que ésta arda. Deje caer la cerilla, añada más yesca y después leña.*

Proteja la cerilla del viento con las manos

Lupa. *Coloque un nido de yesca dentro del tipi, y enfoque los rayos del sol a través de la lupa de manera que el punto de luz más intenso caiga sobre la yesca. Cuando ésta comience a arder, sople ligeramente hasta que haga llama.*

Enfoque los rayos de sol sobre la yesca a través de la lente de aumento

ÚTILES PARA ENCENDER FUEGO

Hay muchas maneras diferentes de encender un fuego empleando utensilios diversos. Éstos varían desde el tradicional bloque de pedernal y la sierra de acero hasta las modernas pastillas de parafina y los palitos de astillas de madera tratadas con varios materiales químicos. Escoja el método que le parezca más sencillo de usar.

Cerillas impermeables. *Quite la cera de la cabeza de las cerillas antes de utilizarlas.*

Pedernal y acero. *Esta sierra de acero y el «pedernal» de aleación de magnesio producen chispas mayores que las del pedernal natural.*

Lupa. *Cuanto mayor sea la lupa, mejores resultados obtendrá, particularmente con el sol débil de las latitudes boreales.*

Algodón. *El algodón impregnado de parafina es una alternativa de peso ligero a la yesca natural.*

Pastillas de parafina. *Pueden utilizarse para encender ramas grandes sin yesca.*

Palos de fuego. *Los palos de virutas de madera tratada con materiales químicos pueden encenderse con una cerilla para prender fuego fácilmente, sin necesidad de yesca.*

FUEGO DE ARCO

1 Corte una rama de madera dura que haga las veces de eje *(véase página 59)*. Debe tener unos 40 cm de longitud y unos 4 cm de grosor. Redondee uno de los extremos y talle una punta con el otro.

2 Busque un trozo de madera suave, como de pino o de balsa, para utilizar como base, y corte una muesca en forma de «V» en uno de los lados. La fricción entre el eje y la base producirá cenizas.

3 Talle un agujero pequeño al lado de la muesca, en el que quepa la punta del eje. La ceniza caliente y negra creada por la fricción del eje girando en el agujero caerá en la yesca a través de la muesca.

4 Utilice un trozo de madera pesada como mango para sujetar y presionar el eje. Talle un agujero en su centro, para la parte superior del eje.

5 Corte una rama fuerte de madera dura para el arco, con una longitud de unos 60 cm y un grosor de 2 cm. Ate, sin tensar, un cordel o cuerda natural a ambos extremos, asegurándose de que está lo suficientemente suelta como para pasarla alrededor del eje.

6 Talle una muesca en el centro del eje y envuelva la cuerda alrededor del mismo. Coloque yesca cerca de la muesca, en la base. Coloque la punta del eje en la base y el mango en el extremo superior del eje. Hágalo girar tirando y empujando el arco, al mismo tiempo que empuja el mango con la otra mano. El humo no tarda mucho en producirse, junto con una gran cantidad de ceniza y chispas, que caerán sobre la yesca gracias a la muesca de la base.

VARIANTE MANUAL

Un método más sencillo que el fuego de arco es la rotación con las manos de un palo terminado en punta sobre una base. Sin embargo, ya que las manos crean menor fricción que el arco, este método únicamente funciona eficientemente en los climas cálidos, donde la madera esté muy seca. Empuje el eje hacia la base al girarlo, y hágalo rotar en impulsos cortos.

Técnica manual. *El empleo del eje manual es más sencillo, aunque no tan efectivo como el arco.*

LEVANTAR EL CAMPAMENTO

LEVANTAR EL CAMPAMENTO debe convertirse en una rutina en la que todos trabajan para lograr ponerse en marcha a una hora previamente acordada. Si las condiciones del tiempo son extremas, esperar a una persona puede ser muy molesto, e incluso peligroso. El segundo jefe –quien debe cerrar la marcha del grupo *(véase página 136)*– debe comprobar el campamento antes de ponerse en marcha. Es importante limpiar el lugar de manera que parezca que nunca haya estado allí –a menos que esté en una situación de supervivencia y deba dejar alguna señal para los rescatadores. Desmonte su tienda o refugio en último lugar, para poder aprovecharla durante el mayor tiempo posible.

LIMPIAR EL CAMPAMENTO

Debe dejar el campamento como si nunca hubiera estado en él: llévese toda la basura, reemplace cualquier bloque de césped que hubiera levantado y obedezca las normas locales. De esta manera, no sólo se asegurará de que la naturaleza permanezca inalterada, sino que también contribuirá al bienestar de los futuros visitantes a la zona. En una situación de supervivencia, si necesita dejar algún mensaje para los posibles rescatadores, puede hacerlo sin alterar el medio ambiente *(véase página siguiente)*.

Tienda. *Las tiendas deben desmontarse en último lugar, para que, en caso de mal tiempo, pueda refugiarse hasta el último momento. Debe acordarse un tiempo para el cual todo debe estar recogido y listo para marchar. Una vez guardadas las tiendas, abandone el lugar inmediatamente.*

Letrina. *La letrina debe rellenarse y volverse a cubrir de hierba; debe marcar la fecha de partida para información de los futuros campistas. En las zonas en las cuales el medio era demasiado sensible para cavar una letrina, debe llevarse todos sus desechos en bolsas de plástico.*

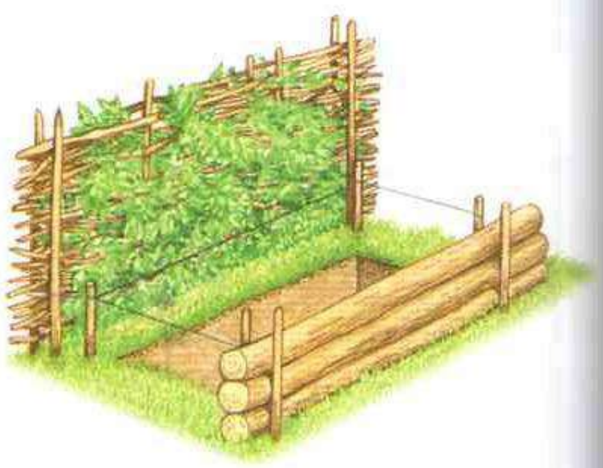

El hoyo del fuego debe cubrirse con el tepe levantado y esparcir hojas por encima

Guarde la basura en bolsas de plástico y llévselas consigo

Borre su presencia. *Es muy importante dejar el campamento como si nunca hubiera estado allí. Desmonte el fuego y disfrace el hoyo del fuego, llévese todos sus desechos, basura y residuos de comida y rellene la letrina. Si siempre actúa de esta manera al abandonar el campamento, no alterará la naturaleza en beneficio de la vida silvestre y de los futuros campistas.*

Fuego. *El fuego debe quedar completamente apagado. Las cenizas deben esparcirse y enterrarse en el suelo. Cualquier residuo sin quemar debe recogerse y guardarse para eliminarlo posteriormente.*

LIMPIAR EL FUEGO

Debe asegurarse de que el fuego esté totalmente apagado en el momento de abandonar el campamento. Incluso si ha cubierto el hoyo del fuego, éste puede contener rescoldos ardientes que podrían originar un incendio forestal. Utilice la menor cantidad de leña posible para preparar el desayuno el día de su marcha, y asegúrese de que toda la leña queda consumida.

1 Una vez el fuego deja de arder, remueva los restos hacia el centro hasta que queden reducidos a cenizas. Cuando las cenizas estén frías, espárzalas y entiérrelas en el suelo para dispersarlas.

2 Asegurándose de que no quedan cenizas en la superficie que puedan matar la hierba, tape el hoyo del fuego con tierra. Alísela con las manos. Después coloque el bloque de césped que había sacado previamente al construir el fuego. Si empleó rocas, dispérselas y camúflelas con mantillo.

3 Rellene los bordes del bloque de césped con tierra y hierbas para que no queden marcas de la unión. Esparza hojas y hierba encima del lugar para que se asemeje al área circundante. Guarde cualquier desperdicio en una bolsa de plástico y llévela consigo hasta que pueda quemarla o deshacerse de ella.

NO DEJE HUELLAS

Muchas áreas naturales quedan dañadas cuando los visitantes dejan basuras que estropean el paisaje y que pueden llegar a matar a los animales; alteran la vegetación y los árboles; causan incendios forestales, y contaminan los ríos. Debe dejar todas las áreas naturales como si nunca hubiera estado en ellas, poniendo especial cuidado en su campamento. En algunas áreas naturales vigiladas, está prohibido construir fuegos o cavar letrinas; respete estas reglas en beneficio del medio ambiente y de los futuros campistas. Si deja residuos de comida, los animales pueden convertirse en una molestia para los campistas, llegando incluso a poner en peligro no sólo la vida de los animales, sino también la de las personas.

DEJAR MENSAJES A LOS RESCATADORES

Cualquier signo fuera de lo común será avistado por aquéllos que le sigan. Indique la dirección de su marcha, así como el tiempo y la fecha, en cualquier señal que deje a los rescatadores. Estos signos permitirán trazar su ruta. Coloque los mensajes lo más alto posible, a salvo de los animales. También puede levantar un montón de piedras de modo que sea visible y colocar en él una nota.

Bandera. *Ate dos ramas para formar una cruz y cuelgue de ella una camiseta de un color vivo. Coloque la cruz en un claro. Si utiliza dos o más de estas señales, dispóngalas de tal manera que apunten hacia la dirección que haya tomado.*

Piña. *Coloque una piña en una rama dividida en el lado del camino que haya tomado. Cualquier suceso que no ocurra naturalmente será avistado por cualquiera que le busque. Tenga cuidado de no emplear algo que los animales puedan comerse.*

Flechas. *Colocadas en un claro, las piedras en forma de flecha indicarán la dirección de su viaje. La flecha también será visible desde el aire. Haga flechas nuevas a intervalos, y siempre en los puntos en los que el camino se divida.*

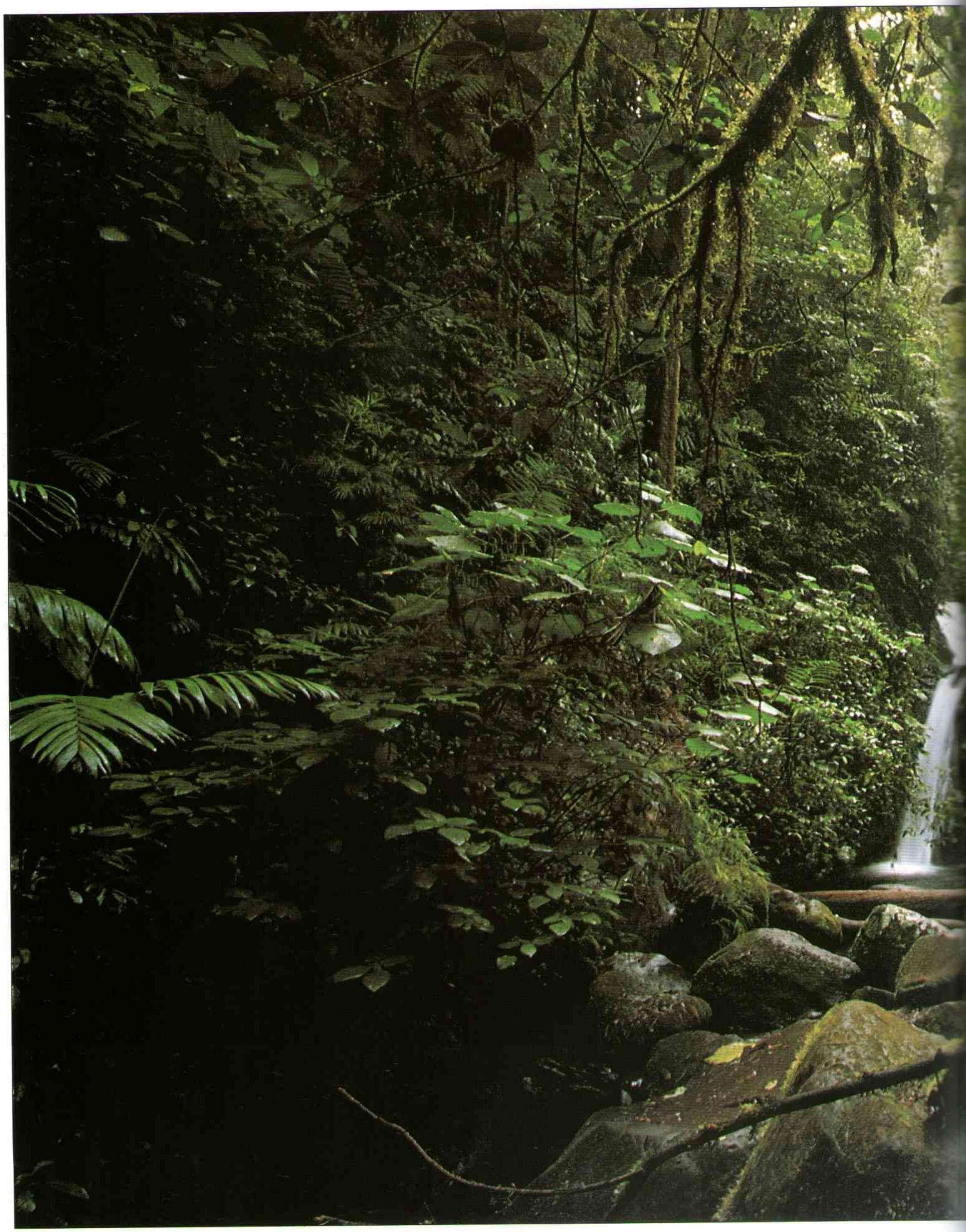

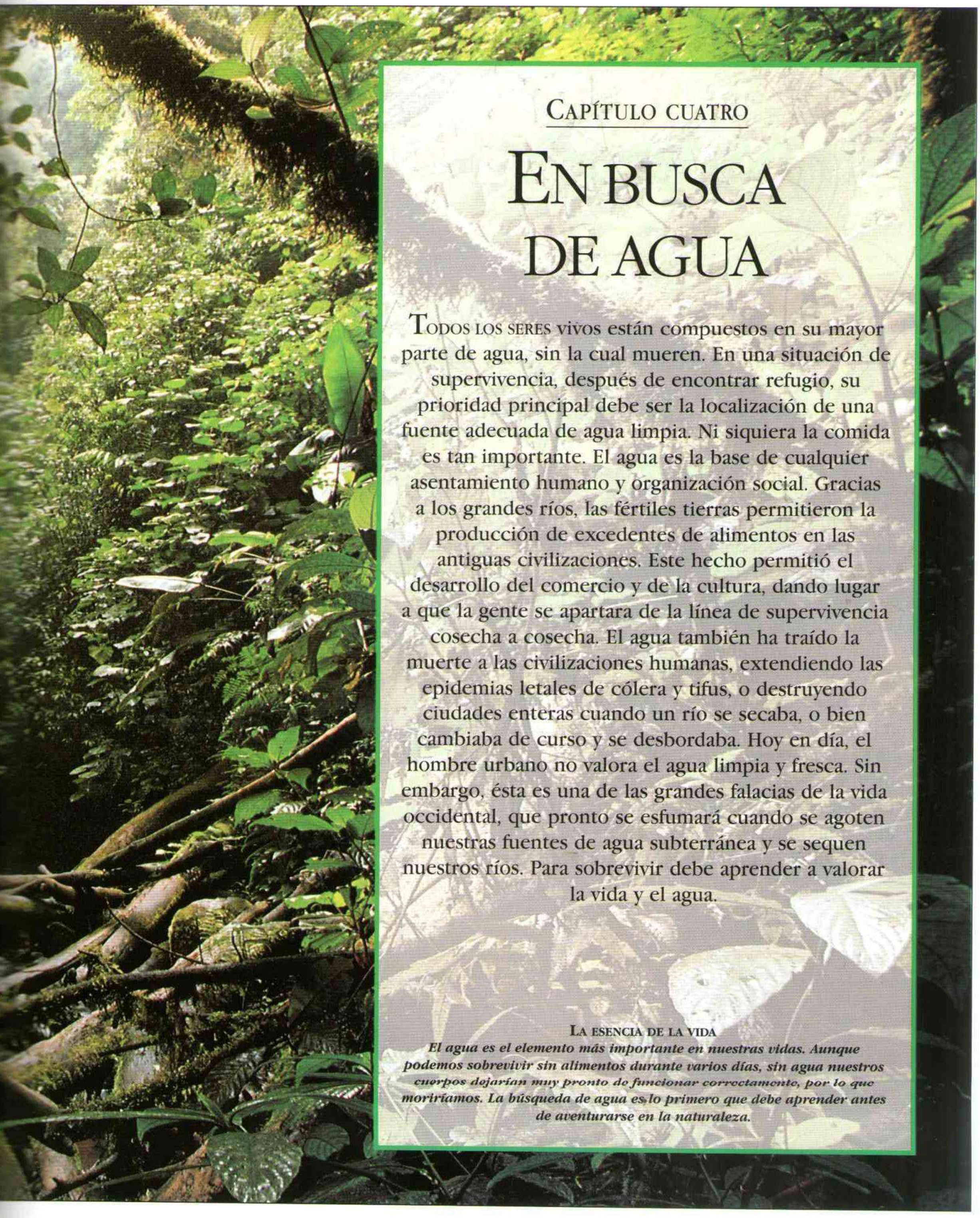

CAPÍTULO CUATRO

EN BUSCA DE AGUA

TODOS LOS SERES vivos están compuestos en su mayor parte de agua, sin la cual mueren. En una situación de supervivencia, después de encontrar refugio, su prioridad principal debe ser la localización de una fuente adecuada de agua limpia. Ni siquiera la comida es tan importante. El agua es la base de cualquier asentamiento humano y organización social. Gracias a los grandes ríos, las fértiles tierras permitieron la producción de excedentes de alimentos en las antiguas civilizaciones. Este hecho permitió el desarrollo del comercio y de la cultura, dando lugar a que la gente se apartara de la línea de supervivencia cosecha a cosecha. El agua también ha traído la muerte a las civilizaciones humanas, extendiendo las epidemias letales de cólera y tifus, o destruyendo ciudades enteras cuando un río se secaba, o bien cambiaba de curso y se desbordaba. Hoy en día, el hombre urbano no valora el agua limpia y fresca. Sin embargo, ésta es una de las grandes falacias de la vida occidental, que pronto se esfumará cuando se agoten nuestras fuentes de agua subterránea y se sequen nuestros ríos. Para sobrevivir debe aprender a valorar la vida y el agua.

LA ESENCIA DE LA VIDA

El agua es el elemento más importante en nuestras vidas. Aunque podemos sobrevivir sin alimentos durante varios días, sin agua nuestros cuerpos dejarían muy pronto de funcionar correctamente, por lo que moriríamos. La búsqueda de agua es lo primero que debe aprender antes de aventurarse en la naturaleza.

LA IMPORTANCIA DEL AGUA

EL HOMBRE DEBE beber un mínimo de tres litros de agua cada día. Así pues, una fuente constante de agua resulta vital para la vida. Viajar está, por lo general, determinado por la posibilidad de agua, ya que incluso los vehículos no pueden transportar más que la necesaria para unos cuantos días. Si viaja a pie, debe llevar suficiente agua al menos para un día, así como una reserva de emergencia. Resulta vital llegar a una fuente de agua al menos una vez al día. Saber dónde encontrar el agua y cómo obtenerla es, pues, esencial.

DÓNDE BUSCAR AGUA

Cuando la lluvia cae en las montañas, se recoge en pequeños riachuelos que se juntan formando torrentes mayores. En cuanto el agua alcanza las tierras bajas, estos torrentes se han convertido en grandes ríos, más lentos, que finalmente desembocan en el mar. El agua potable puede encontrarse en cualquier punto de este recorrido. Cualquier agua recogida en la naturaleza debe purificarse antes de beberse, para evitar posibles infecciones *(véase página 75)*.

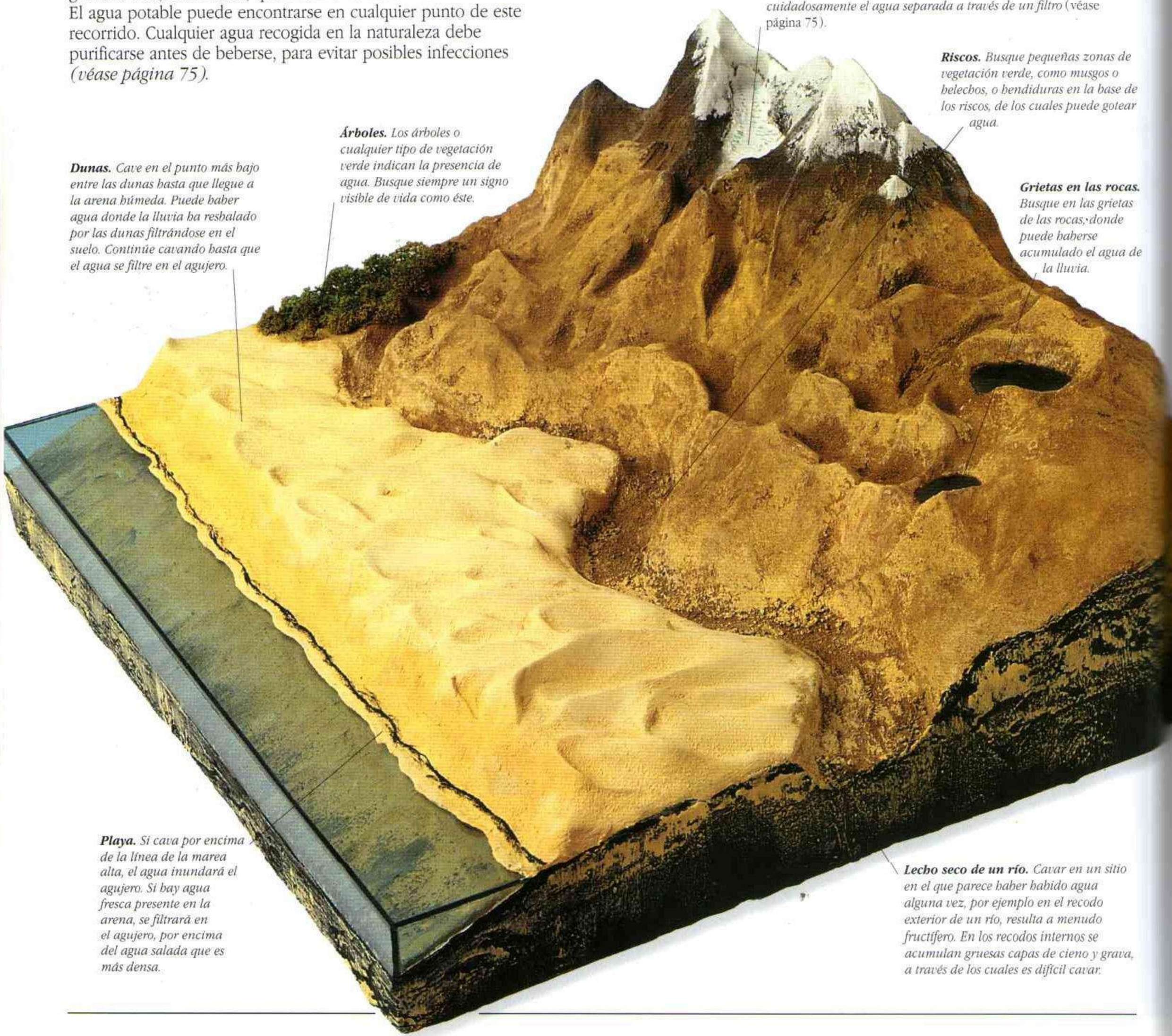

Glaciar. *Los glaciares a menudo generan vigorosos torrentes de agua de deshielos. Sin embargo, el agua de los glaciares contiene una gran cantidad de polvo de roca abrasivo, que puede causar diarrea. Antes de beber el agua, debe dejarla reposar una noche para que las partículas de mineral se asienten; vierta entonces cuidadosamente el agua separada a través de un filtro* (véase página 75).

Riscos. *Busque pequeñas zonas de vegetación verde, como musgos o helechos, o hendiduras en la base de los riscos, de los cuales puede gotear agua.*

Árboles. *Los árboles o cualquier tipo de vegetación verde indican la presencia de agua. Busque siempre un signo visible de vida como éste.*

Dunas. *Cave en el punto más bajo entre las dunas hasta que llegue a la arena húmeda. Puede haber agua donde la lluvia ha resbalado por las dunas filtrándose en el suelo. Continúe cavando hasta que el agua se filtre en el agujero.*

Grietas en las rocas. *Busque en las grietas de las rocas, donde puede haberse acumulado el agua de la lluvia.*

Playa. *Si cava por encima de la línea de la marea alta, el agua inundará el agujero. Si hay agua fresca presente en la arena, se filtrará en el agujero, por encima del agua salada que es más densa.*

Lecho seco de un río. *Cavar en un sitio en el que parece haber habido agua alguna vez, por ejemplo en el recodo exterior de un río, resulta a menudo fructífero. En los recodos internos se acumulan gruesas capas de cieno y grava, a través de los cuales es difícil cavar.*

INDICIOS DE AGUA

La presencia de agua está a menudo indicada por indicios de vida como vegetación verde, huellas animales o viviendas humanas. Sin embargo, hay otros indicios naturales que muestran la existencia de agua oculta. En un terreno aparentemente sin vida, busque signos de vida, aunque tenga en cuenta que algunas aves, como el buitre, pueden viajar a grandes distancias de sus fuentes de agua.

Árboles y otro tipo de vegetación. *Todos los árboles necesitan agua para vivir, aunque algunos tienen raíces muy largas que se extienden bajo el suelo en busca de agua. Las palmeras crecen, por lo general, donde hay algún tipo de agua cerca de la superficie, y algunas variedades almacenan agua en sus troncos o raíces* (véase página 72).

Huellas animales. *Los rumiantes necesitan beber por lo menos dos veces al día, por lo que se dirigen desde o hacia el agua. Busque el punto en el que convergen las huellas de los animales: puede ser un indicativo de una fuente de agua. Tenga cuidado si se acerca a una charca: cerca puede haber depredadores.*

Abejas y moscas. *Las abejas no se alejan, generalmente, más de 5 km de sus nidos y necesitan una reserva constante de agua. Observe la dirección en la que vuelan cuando abandonan el nido. Las moscas aún permanecen más cerca del agua —unos 100 m aproximadamente. Sin embargo, en el desierto parecen venir de ningún sitio, independientemente de donde esté la fuente más cercana de agua.*

Hormigas en movimiento. *Las hormigas dependen de una reserva constante de agua. Si observa algunas de ellas subiendo a un árbol, es muy probable que se dirijan hacia una reserva de agua de lluvia.*

LA DESHIDRATACIÓN Y LAS NECESIDADES DE AGUA DEL CUERPO

El agua es un constituyente primordial del cuerpo humano. No disponemos de ningún medio para almacenar agua en nuestros cuerpos, como hacemos con las grasas. Necesitamos una provisión constante de agua para controlar la temperatura corporal (transpiración), la eliminación de productos de desecho a través de los riñones y el funcionamiento del sistema nervioso. Sin alimentos podemos sobrevivir con nuestras reservas corporales durante unas tres semanas, pero si el agua perdida no se repone en cinco días, la muerte es inminente.

EFECTOS DE LA PÉRDIDA DE AGUA

Pérdida entre 1-5%	Pérdida entre 6-10%	Pérdida entre 11-12%
Sed	Dolor de cabeza	Delirio
Malestar	Mareo	Lengua hinchada
Letargo	Sequedad en la boca	Contracciones
Impaciencia	Dificultad para respirar	Sordera
Falta de apetito	Hormigueo en las extremidades	Visión borrosa
Enrojecimiento de la piel	Tono azulado de la piel	Falta de sensibilidad en la piel
Aumento del pulso	Dificultad en el habla	La piel comienza a arrugarse
Náuseas	Incapacidad para caminar	Incapacidad de tragar
Debilidad	Visión borrosa	Muerte

Necesidades de agua. *Cerca de un 75% del cuerpo humano es agua, imprescindible para la vida. Cuando el cuerpo la pierde a través de la transpiración y de la orina, debe ser reemplazada. Si no es así, se desencadenan diversos problemas de salud que se agravan hasta terminar en la muerte.*

PREVENIR LA PÉRDIDA DE AGUA

Incluso mientras descansa en la sombra, una persona normal pierde más de un litro de agua cada día debido a la respiración y a la orina. El ejercicio aumenta la pérdida de agua a través de la transpiración, especialmente en climas templados, así que, si su reserva de agua es limitada, debe descansar de día y trabajar únicamente de noche. Intente mantenerse fresco y a la sombra. Respire a través de la nariz para reducir la pérdida de agua, y no fume. No se recueste sobre el suelo caliente. Limítese a comer lo indispensable para mantenerse vivo. Omita los alimentos grasos en su dieta, y no beba alcohol —ambos requieren una gran cantidad de agua para ser digeridos. No espere a haber agotado su reserva de agua antes de comenzar a buscar más.

RECOGER AGUA

EN UNA SITUACIÓN de supervivencia, después de encontrar refugio, la búsqueda de agua debe convertirse en su siguiente prioridad, ya que sin agua no sobreviviría más que unos cuantos días. Es posible que no siempre tenga la fortuna de acampar al lado de una fuente de agua de confianza *(véase página 40)*. Sin embargo, puede sobrevivir en una situación de emergencia recogiendo el agua de lluvia o del rocío. El agua recogida de la atmósfera, de las plantas o del suelo, tiene una gran ventaja sobre las otras fuentes de agua naturales: su pureza *(véase página 74)*. El amanecer es la mejor hora para recoger agua.

RECOGER AGUA DE LA LLUVIA

Mientras permanezca en el campamento, ponga varios recipientes para recoger el agua de la lluvia que pudiera caer. También debe recoger el agua que fluye por el tejado de su refugio, improvisando un canal para encauzar el agua hacia un recipiente *(véase página 53)* o hacia un estanque *(véase página 77)*. Incluso si tiene un río o arroyo cerca, recoja el agua de la lluvia, que será pura y le permitirá ahorrar en medios de depuración químicos o le evitarán tener que hervirla *(véase página 75)*. Para recoger el agua de lluvia, extienda una tela impermeable tensa sobre un área extensa, preferiblemente en una pendiente. Sujete las esquinas al suelo con palos y recoja el agua en un recipiente.

Colector de lluvia. *Cualquier lámina impermeable recogerá la lluvia y la canalizará hacia un recipiente. Si la lluvia es intensa, vigile continuamente para reemplazar rápidamente el recipiente cuando se llene.*

RECOGER AGUA DEL ROCÍO

Cuando el aire se enfría por la noche, el vapor de agua que contiene se condensa en rocío sobre el terreno bajo, la vegetación y los vehículos. El agua se evapora rápidamente al salir el sol. Muchas plantas, insectos y animales dependen del rocío para sobrevivir; los hombres también pueden utilizar esta fuente natural de agua.

1 Puede recoger el rocío empapando un paño en la hierba larga y húmeda. La mejor hora para recoger el rocío es el amanecer, ya que después de la salida del sol se evapora rápidamente. Muchos pueblos indígenas de las tierras áridas utilizan este método regularmente para obtener agua pura.

2 Cuando el trapo esté empapado, exprímalo en un recipiente. Repita la operación hasta que tenga suficiente agua.

CAVAR EN BUSCA DE AGUA

A menudo el agua se filtra en un agujero excavado en un terreno pantanoso. Cave un agujero de unos 30 cm de profundidad. Recoja el agua que se acumula en el agujero. El agua estará ligeramente cenagosa las primeras veces que se llene el agujero, pero posteriormente subirá agua limpia que podrá depurar y beber.

Agua cenagosa. *Saque el agua lodosa de un agujero en terreno pantanoso hasta que el agua fresca aflore a la superficie.*

ADVERTENCIA

Antes de cavar un agujero de este tipo estudie el terreno. Nunca cave donde el suelo tenga un olor muy potente, o tenga limo verde en la superficie —cualquier agua en este lugar estará probablemente contaminada. No recoja agua donde encuentre animales muertos. Depure siempre el agua antes de beberla *(véase página 74)*.

CÓMO OBTENER AGUA DE LA VEGETACIÓN

Las hojas de las plantas desprenden vapor de agua. Ésta puede recogerse fácilmente envolviendo el follaje en una bolsa de plástico de manera que el vapor se condense en la bolsa en forma de pequeñas gotas.

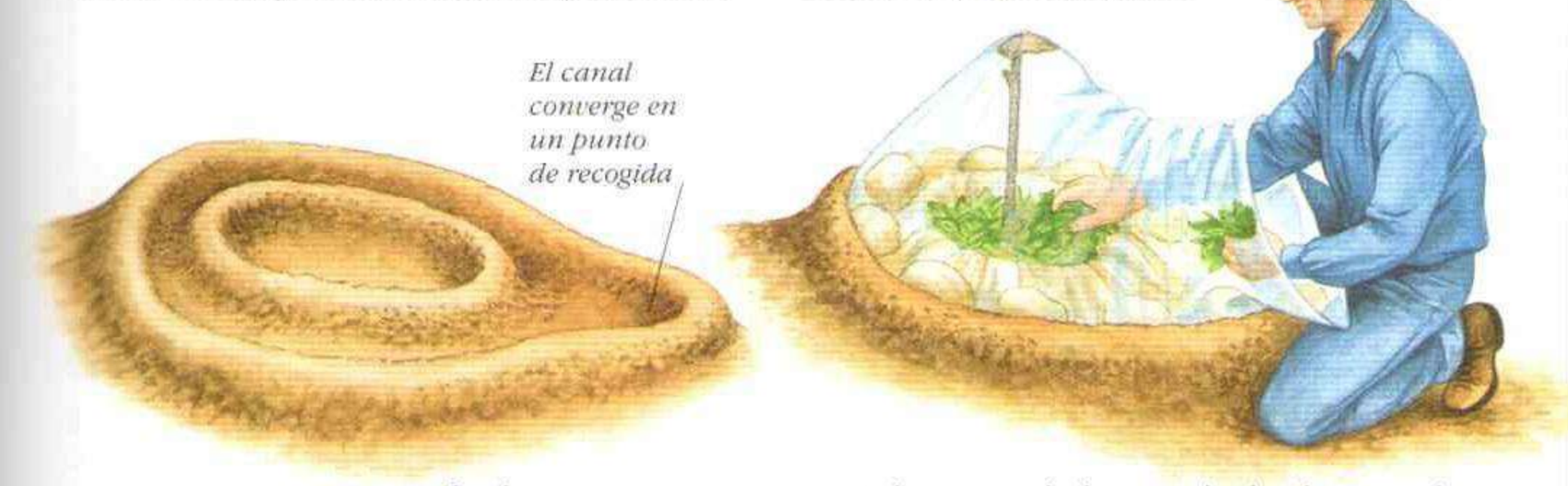

1 Cave un cráter poco profundo en una pendiente, rodeado por un cráter mayor canalizado hacia un punto de recogida. Los cráteres deben quedar separados por un caballón elevado de tierra.

2 Coloque una bolsa grande de plástico sobre ambos cráteres, alzada en el centro con una pequeña estaca y sujeta a los bordes interiores con piedras. Coloque hojas verdes y hierbas dentro del cráter central.

3 Cierre perfectamente el extremo de la bolsa, asegurándose de que la vegetación no toca el plástico. Conforme las hojas «sudan», el vapor de agua desprendido se calentará y condensará en el plástico. Se deslizará por los laterales de la bolsa hacia el canal, y después bajará hacia el punto de recogida. Vierta el agua en un recipiente.

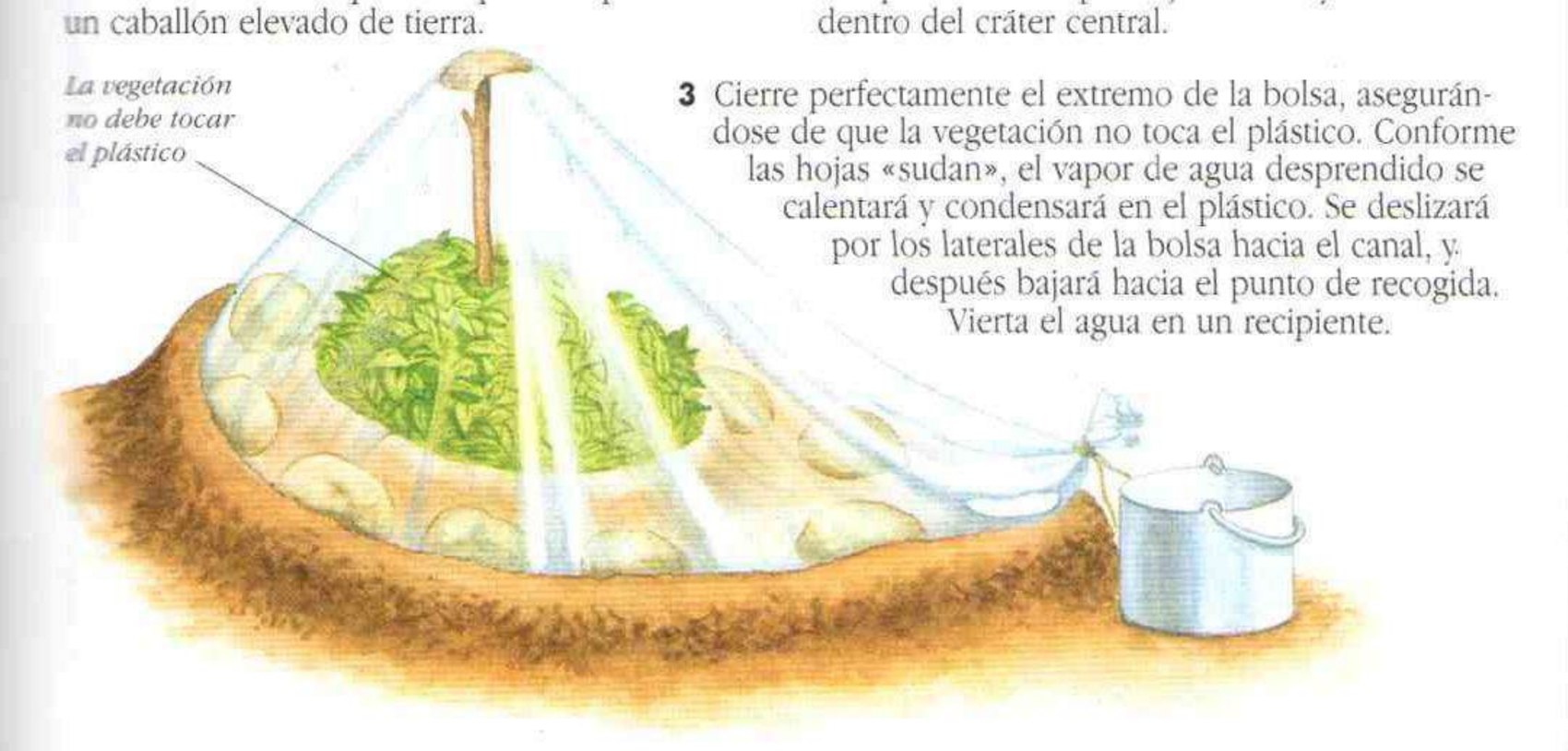

OTROS MÉTODOS

En ocasiones puede recoger pequeñas gotas de agua de la neblina o de una niebla muy densa colocando un trapo en el suelo, o colgándolo entre los árboles. En un caso de emergencia, coloque una lámina de plástico sobre la hierba durante la noche. Conforme el aire se enfría, el aire caliente del suelo se condensará en pequeñas gotas de agua en la cara inferior de la lámina. Aunque con este método no recogerá mucha agua, puede ser suficiente para mantenerle vivo hasta que encuentre una fuente mejor.

Empleo de una rama. *Ate una bolsa de plástico alrededor de una rama o una planta entera. El vapor de agua que desprende el follaje se calentará dentro del plástico y se condensará en agua en la superficie interna de la bolsa.*

CÓMO EXTRAER AGUA DEL SUELO

También puede extraer agua del suelo empleando un destilador solar. Mientras exista una diferencia de temperatura entre dos superficies, el aire entre ambas se calentará y saturará, condensando el agua en forma de pequeñas gotas sobre la superficie más fría.

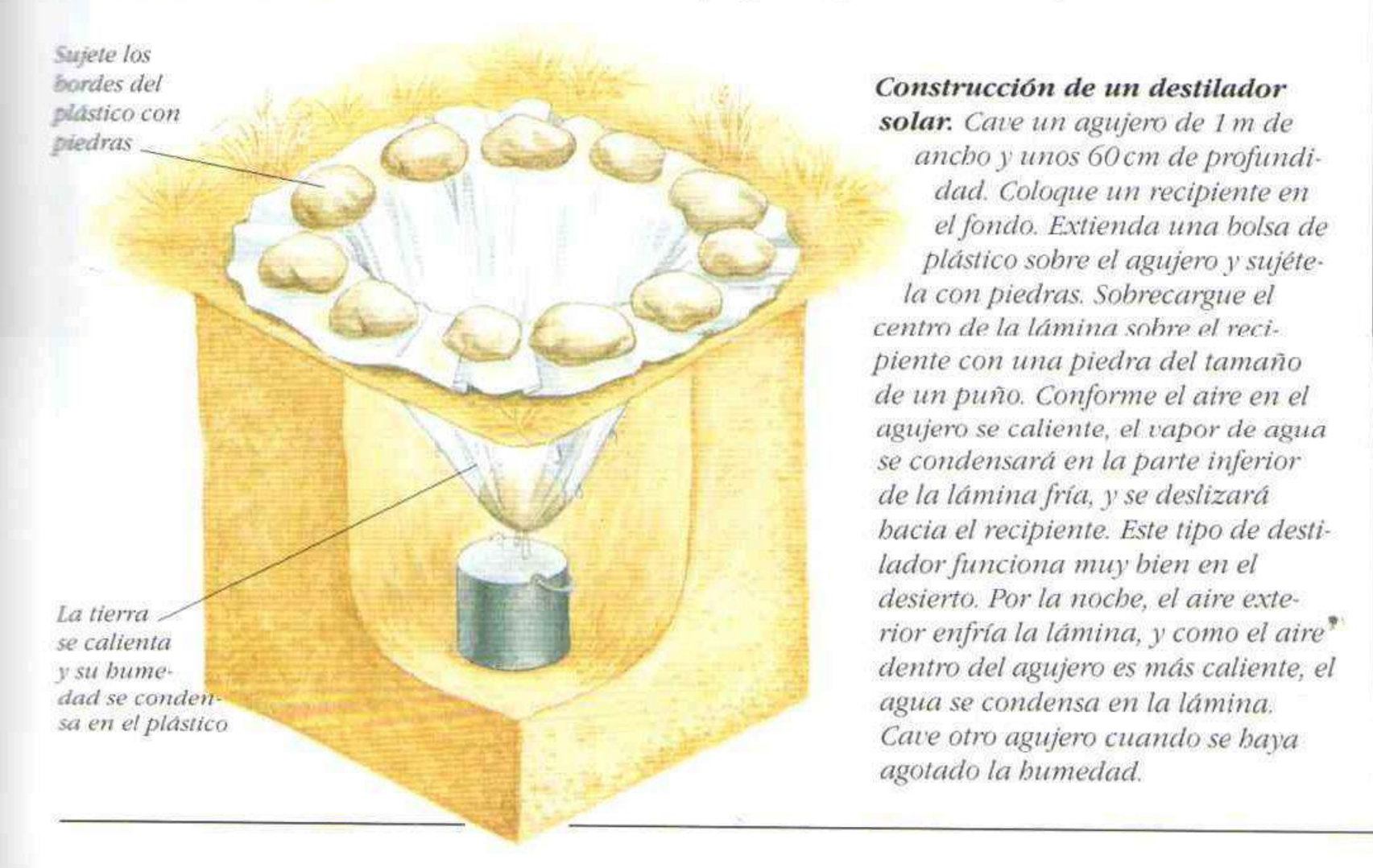

Construcción de un destilador solar. *Cave un agujero de 1 m de ancho y unos 60 cm de profundidad. Coloque un recipiente en el fondo. Extienda una bolsa de plástico sobre el agujero y sujétela con piedras. Sobrecargue el centro de la lámina sobre el recipiente con una piedra del tamaño de un puño. Conforme el aire en el agujero se caliente, el vapor de agua se condensará en la parte inferior de la lámina fría, y se deslizará hacia el recipiente. Este tipo de destilador funciona muy bien en el desierto. Por la noche, el aire exterior enfría la lámina, y como el aire dentro del agujero es más caliente, el agua se condensa en la lámina. Cave otro agujero cuando se haya agotado la humedad.*

CONDENSACIÓN DEL AGUA DE MAR

El agua de mar o la orina pueden condensarse en agua potable empleando el mismo principio del destilador solar. Coloque el agua salada en un recipiente, con una taza en el centro. Coloque una lámina de plástico sobre el recipiente y sujétela bien. Ponga una piedra en el centro para formar un cono sobre la taza. El aire bajo el plástico se calienta, y conforme se calienta el agua salada se condensa en agua dulce en la parte inferior de la lámina.

DESTILADOR DE AGUA DE MAR

FUENTES NATURALES DE AGUA

EXISTEN MUCHAS FUENTES naturales de agua. Por ejemplo, puede obtener agua de las plantas, de la sangre y de los ojos de los animales, y de algunas ranas que retienen la humedad en su piel mientras hibernan enterradas en el barro de las charcas secas. Los nativos saben dónde encontrar agua, que es su recurso más preciado. Al utilizar estas fuentes naturales de agua, puede estar destruyendo un ecosistema cuidadosamente mantenido, y puede levantar las objeciones de las gentes del lugar. Debe tener cuidado de no dañar las charcas o fuentes naturales; asegúrese de que puede recoger este agua, y no desperdicie ni una gota.

EJEMPLOS DE PLANTAS QUE ALMACENAN AGUA

Muchas variedades diferentes de plantas almacenan agua, bien en sus hojas o en sus raíces. Algunas recogen el agua de la lluvia para atrapar insectos o comida, mientras que otras segregan líquidos especiales para su propio uso, que pueden beberse en caso de emergencia. Abajo se muestran algunos ejemplos de plantas de todo el mundo que almacenan agua.

Estrella africana (Stapelia). *Al igual que otras plantas carnosas, la estrella africana almacena agua en el tallo. También puede masticar sus hojas carnosas para obtener humedad.*

Nepente o sarracenia (Nepenthes). *Esta planta atrapa insectos en un líquido acuoso en su «jarra». Puede extraer agua, pero debe filtrarla antes para separar los insectos (que puede comer también).*

Mimosa (Acacia). *Al igual que otros árboles del desierto, la mimosa almacena agua en sus raíces, justo por debajo de la superficie del suelo.*

Neoregalia. *A menudo las bromeliáceas recogen agua de lluvia en sus centros, y tienen hojas comestibles que pueden masticarse para extraer agua.*

Chumbera (Opuntia). *Algunos cactus, como la chumbera, tienen hojas carnosas que desprenden humedad al masticarlas.*

OTRAS FUENTES DE AGUA

Muchas palmeras contienen un líquido azucarado que puede extraerse de sus tallos floridos. Las raíces de las plantas del desierto son almacenes de agua, pero son difíciles de arrancar. Algunas ranas africanas y australianas almacenan agua en sus cuerpos para utilizarla durante su época de hibernación. En una emergencia extrema puede estrujarlas para extraer el agua.

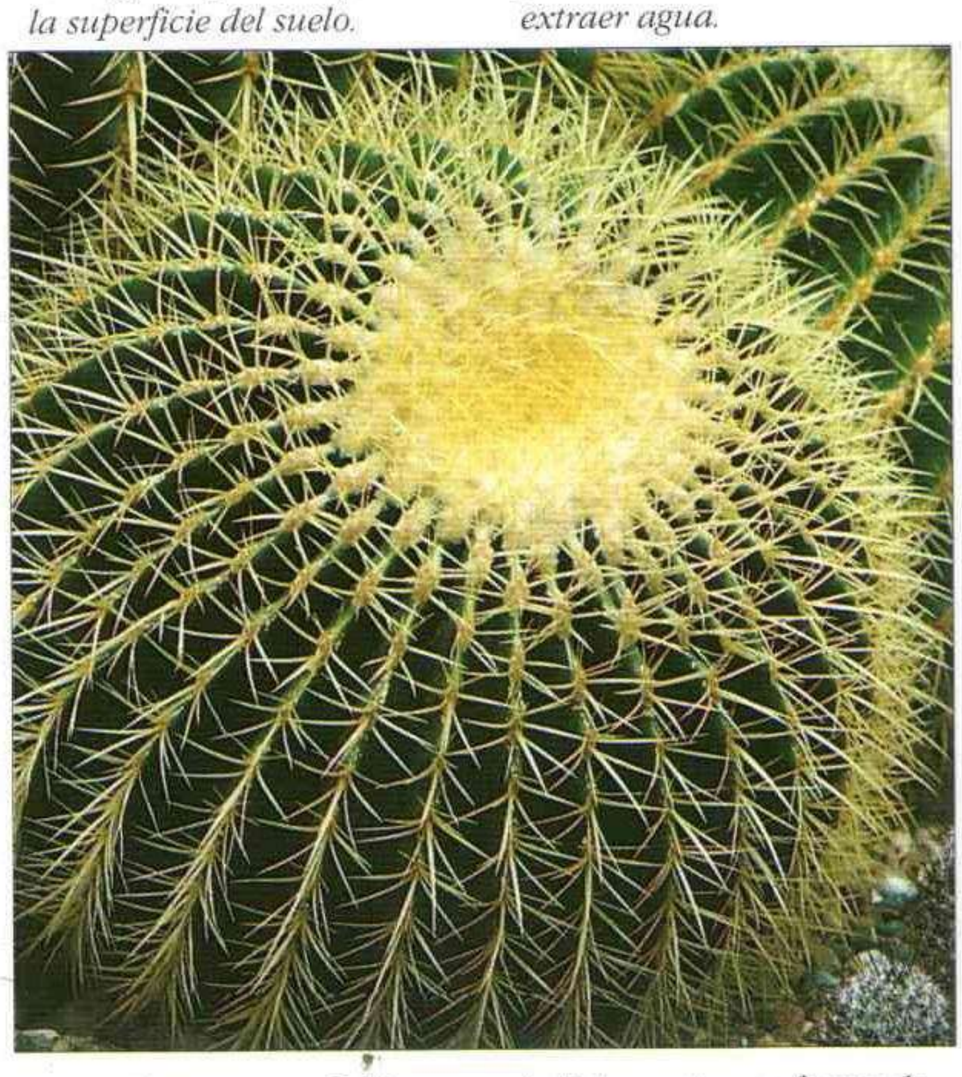

Cactus (Ferocactus, Echinocactus). *Estos cactus en forma de tonel son la única excepción a la regla de no beber ningún líquido lechoso proveniente de una planta* (véase página siguiente). *Hay varias especies de cactus de este tipo que pueden crecer hasta 1 m.*

CÓMO EXTRAER AGUA DE UN CACTUS

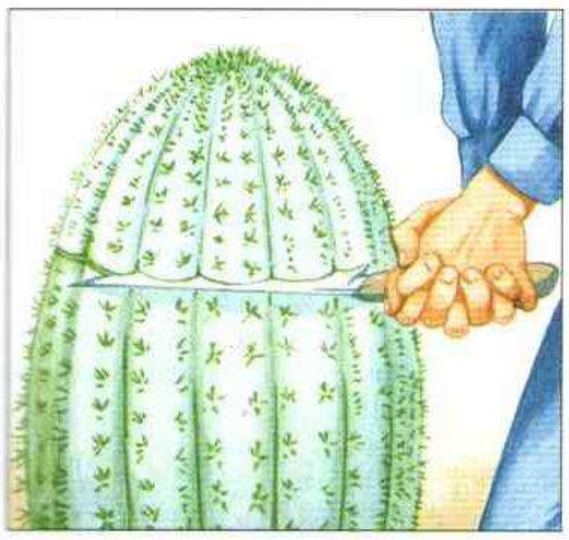

1 Para extraer agua de un cactus, corte cuidadosamente la parte superior con un cuchillo muy afilado o un machete, evitando las espinas.

2 Machaque la carne del interior del cactus con un palo hasta formar una pulpa. De esta manera separará el agua que contiene.

3 Chupe el jugo de la pulpa a través del tallo hueco de una hierba o caña. Únicamente debe utilizar un cactus para extraer agua en caso de emergencia, ya que su crecimiento es muy lento, y algunas especies podrían extinguirse en ciertas áreas si se utilizan en exceso.

Chupe el jugo de la pulpa utilizando una caña hueca

CÓMO EXTRAER AGUA DE UNA LIANA

1 Corte la liana en el punto más alto que pueda alcanzar, con un cuchillo largo o un machete. No corte primero la parte inferior, ya que el líquido fluiría hacia arriba por la acción capilar.

2 Tenga a mano un recipiente. Corte la liana en el extremo inferior y recoja el líquido que desprende. Corte otra liana si es necesario.

ADVERTENCIA

La savia lechosa de las plantas es generalmente venenosa, al igual que el líquido del cactus saguaro gigante *(Carnegiea)* del sudoeste americano. La leche de coco y los líquidos animales contienen proteínas, que requieren agua adicional para ser digeridos por el cuerpo. Además, la espesa y rica leche de coco maduro es un fuerte laxante, que podría causarle diarrea, cansancio y deshidratación.

CÓMO EXTRAER AGUA DE LAS RAÍCES DE LOS ÁRBOLES

Elija un bananero joven y corte el tronco justo por encima del nivel del suelo

1 Corte un banano joven a una altura de 8 cm por encima del suelo.

2 Al vaciar el centro del tronco, el agua se filtrará hacia el hueco proveniente de las raíces, y podrá extraerla.

AGUA DEL HIELO Y DE LA NIEVE

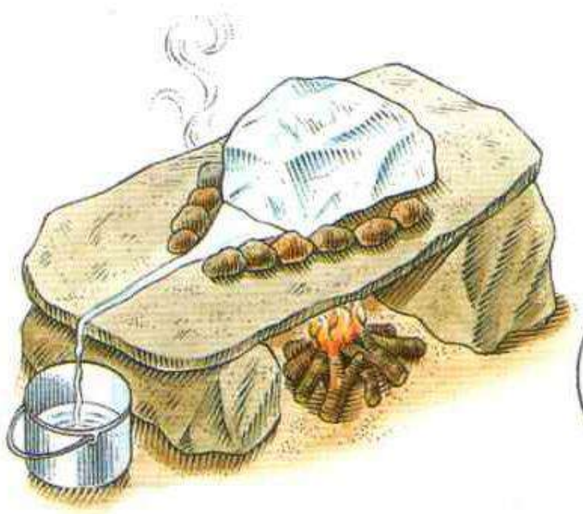

Fundir hielo. *Funda el hielo sobre una roca inclinada sobre un fuego. No utilice hielo marino recién congelado, ya que contiene sal. Utilice hielo viejo, de color azulado, que contiene mucha menos sal.*

Fundir nieve. *Utilice la nieve densa, extraída de la parte más profunda posible. Cuélguela en una tela sobre un recipiente cercano al fuego, y se derretirá gradualmente. No funda nieve en un recipiente sobre el fuego, ya que el agua producida será absorbida rápidamente por la nieve remanente y el recipiente se quemaría. Unos 40 cm de nieve producen 3 cm de agua potable.*

LA DEPURACIÓN DEL AGUA

EN LA NATURALEZA el agua rara vez es pura, y siempre debe depurarse antes de beberla, ya que puede contener microorganismos patógenos que pueden transmitir enfermedades. También deben separarse las partículas de limo y otros contaminantes. En los países occidentales, la gente supone que el agua siempre es potable, especialmente si sale de un grifo. Sin embargo, en muchas zonas, ni siquiera el agua del grifo es pura. Algunas personas toleran bien el agua ligeramente impura, pero también se acostumbran a sufrir toda una serie de problemas gastrointestinales.

Agua potable. *El agua natural debe filtrarse, por lo menos a través de un calcetín o un pañuelo. Asimismo, debe purificarse siempre para eliminar los organismos dañinos.*

MÉTODOS PARA DEPURAR EL AGUA

A excepción de la lluvia, incluso el agua con el aspecto más puro no es únicamente H_2O, sino que contiene sales, minerales y microorganismos. El agua de la naturaleza debe, en primer lugar, filtrarse para eliminar las partículas y, a continuación, debe depurarse para destruir los organismos patógenos hirviéndola por lo menos durante cinco minutos o tratándola con agentes químicos. Nunca debe beber orina o agua de mar en su estado natural, pero en caso de una emergencia pueden destilarse *(véase página siguiente)*. Existen muchos tipos de depuradores, la mayoría de los cuales filtran el agua y después la purifican por la acción de agentes químicos. Como forma alternativa, puede utilizar las pastillas depuradoras o hervir el agua después de filtrarla *(véase página siguiente)*.

Asegúrese de que el cuentagotas esté bien cerrado

Yodo. *El yodo puede ensuciarle al usarlo, y tiñe el agua de rosa. También da al agua un sabor particular.*

Permanganato potásico. *Ésta es una sustancia química de uso general para esterilizar el agua. Utilice el suficiente para teñir el agua de un ligero color rosado.*

Pastillas. *Estas pastillas contienen una base de cloro que da al agua un sabor que recuerda al de las piscinas, pero son muy seguras. Utilice una pastilla por cada medio litro de agua.*

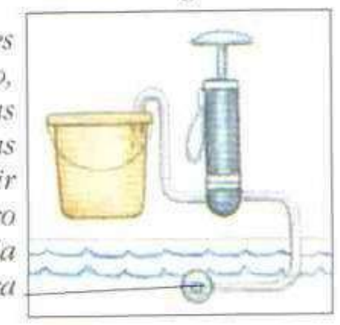

El agua a purificar es aspirada por el bulbo, y pasa a través de las sustancias químicas antes de salir purificada por el otro extremo de la manguera

La tapa del filtro sirve de taza

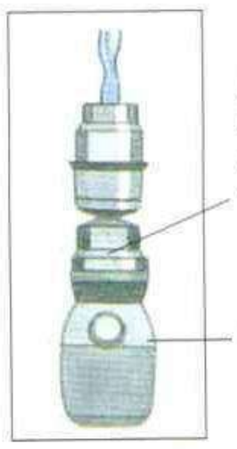

El filtro se ajusta a la cantimplora para filtrar pequeñas cantidades de agua

Cantimplora

Palanca de la bomba

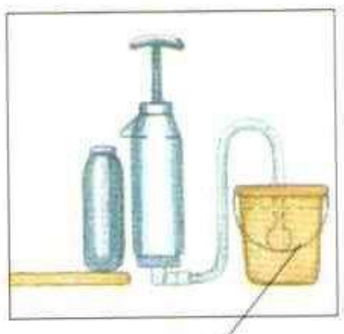

El agua se bombea desde el recipiente hacia la botella recolectora

El extremo de la manguera se sumerge en el agua a depurar

Mini filtro portátil. *El extremo de la manguera se sumerge en un recipiente con agua impura. El agua pura sale del pico de la bomba conforme el agua se bombea al subir y bajar la palanca, y se recoge finalmente en un recipiente.*

Las piezas del filtro encajan en una botella de agua

Lleva incorporado un filtro con sustancias químicas

Depurador con taza. *Éste puede guardarse en un bolsillo cuando se desmonta. El agua impura se vierte en el filtro, que se ajusta sobre la parte superior de una botella de agua. El agua pura caerá gota a gota en la cantimplora.*

La palanca de la bomba se mueve hacia arriba y hacia abajo

La bomba contiene sustancias químicas que depuran el agua

El bulbo se coloca en el agua a depurar

El agua pura sale por esta manguera

Depurador grande. *El bulbo se coloca directamente en una corriente de agua, y el otro extremo de la manguera dentro de un cubo. El agua de la corriente es aspirada por el depurador al subir y bajar la palanca.*

TRATAMIENTO IMPROVISADO DE AGUA

Filtro en un trípode. *Si no dispone de un depurador prefabricado, puede utilizar un calcetín para filtrar el agua antes de hervirla para depurarla* (véase página anterior). *Si utiliza un calcetín, rellénelo con otro calcetín mucho más fino, arena o un pañuelo, y después de cada sesión de filtrado déle la vuelta y aclárelo bien. Construya un trípode para sujetar su filtro* (véase página 119).

DESTILACIÓN DE AGUA SALADA

En un caso de emergencia puede obtener agua pura y potable a partir de agua del mar, por destilación. Hierva el agua y recoja el vapor en un trapo colocado sobre el recipiente. Utilice varios trapos a la vez, uno después de otro. Levántelos del recipiente con una rama para evitar quemarse. Una vez se han enfriado, exprima los trapos y recoja el agua.

LOS PELIGROS DEL AGUA EN LA NATURALEZA

Cuando los animales beben, transmiten al agua una variedad de microorganismos y parásitos intestinales muy perjudiciales para los seres humanos. Algunos de ellos necesitan un período de incubación, y otros tienen un efecto inmediato. Como el hombre urbano no ha desarrollado una inmunidad o tolerancia a estos contaminantes, se ve muy afectado por ellos. Aunque los arroyos de las montañas deberían ser puros, la gran variedad de sustancias químicas industriales utilizadas en la agricultura, así como la posible filtración en el suelo desde los asentamientos urbanos, hacen necesario depurar el agua recogida antes de beberla.

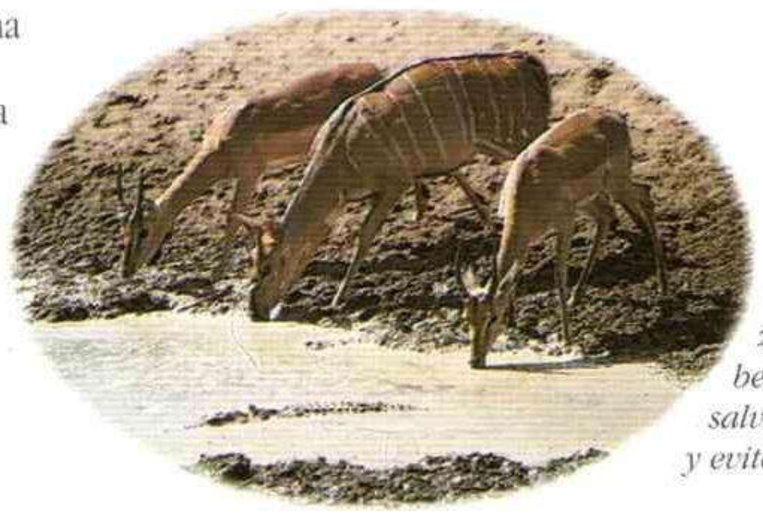

Peligro al acecho. *La recogida de agua en estado natural puede resultar peligrosa. Si utiliza charcas en las que beben los animales salvajes tenga cuidado y evite los depredadores.*

INDICADORES DE AGUA EN MAL ESTADO

Las calaveras y huesos de animales pueden no siempre estar presentes para indicarnos que el agua no está en buenas condiciones. Si el agua está contaminada por sustancias químicas, busque sedimentos polvorientos alrededor de las orillas. La ausencia de vegetación o numerosas algas verdes en la superficie indican que el agua no es potable.

Agua estancada. *Las espadañas y los juncos indican que el agua está estancada: debe evitarse.*

ENFERMEDADES PROPAGADAS POR EL AGUA CONTAMINADA

Enfermedad	Origen	Síntomas
Leptospirosis (en su forma severa = enfermedad de Weil)	Transmitida al ser humano a partir de animales infectados con la bacteria *Leptospirum*. Contraída por contacto con orina o líquidos fetales infectados. Penetra en el organismo a través de las heridas en la piel o las mucosas de la boca, nariz, garganta u ojos.	Provoca síntomas semejantes a los de la gripe (fiebre, escalofríos, dolor de cabeza, dolor muscular). En su forma más severa causa meningitis, ictericia, trastorno renal, hemorragias y problemas cardíacos.
Bilharziosis o *Bilharzia* (Esquistosomiasis)	Causada por un platelminto parásito de las corrientes de agua dulce y de curso lento. Penetra en la piel directamente, alojándose en el intestino. También se transmite por los parásitos de los caracoles de agua dulce.	Causa irritación de la piel, urticaria, ataques asmáticos, dilatación del hígado e irritación del tracto urinario.
Disentería amebiana	Contraída por beber agua contaminada con aguas residuales infectadas.	Causa diarrea con sangre y/o pus, e infección del colon. Una complicación de la infección conlleva hepatitis, abscesos en el hígado y los pulmones, y úlcera intestinal.
Anquilostomiasis	Las larvas del parásito penetran en el cuerpo humano al beber agua o directamente a través de la piel.	Los gusanos adultos se alojan en el intestino, provocando anemia y letargia. Las larvas en el torrente sanguíneo pueden provocar neumonía.
Giardiosis	Provocada por el parásito *Giardia* en agua contaminada con orina o heces infectadas.	Provoca diarrea y dolores abdominales. Cada vez más abundante en Norteamérica, África y Asia.

Cómo prevenir las enfermedades transmitidas por el agua contaminada. *La única manera de evitar contraer una enfermedad transmitida por el agua es purificándola. Si un miembro de su grupo enferma por beber agua contaminada, debe aislarlo junto con alguien que le asista. Debe tener un cuidado extremo con los excrementos y ropa de la víctima, ya que contendrán grandes cantidades del organismo patógeno. Tenga cuidado con la higiene del campamento para que nadie se contagie* (véase página 56).

TRANSPORTE Y ALMACENAMIENTO DEL AGUA

INCLUSO SI HA acampado al lado de un río, necesitará algún tipo de recipiente para contener el agua. Por ejemplo, querrá tener agua a mano para cocinar. También necesitará un recipiente para recoger agua del río. Como acampar junto a una fuente de agua no es siempre la opción más adecuada, el ir a buscar agua será una actividad que tendrá que desarrollar al menos una o dos veces al día. Además, al salir en busca de alimentos, en exploración o para trabajar, puede no encontrar agua cuando la necesite y tener que llevarla con usted.

Conservar el agua. *El agua es un bien muy preciado; nunca debe desperdiciarla. Puede que no le sea tan fácil renovar su reserva.*

ENVASES PARA AGUA

Hay muchos tipos de envases para transportar y conservar líquidos fríos o calientes. Varían desde las botellas y termos de plástico o acero hasta las bolsas en varios tipos de tejido que pueden doblarse cuando están vacías. Asegúrese de que puede diferenciar las botellas de agua de aquéllas en las que transporta el combustible para la cocina.

Termo. *Un termo metálico es prácticamente irrompible y puede conservar líquidos fríos o calientes.*

Botella de plástico. *Una resistente es ideal, pero se fundirá si se pone muy cerca del fuego.*

Botella de acero. *Es una botella resistente, pero pesada si está llena de agua.*

Bolsa para la cintura. *Una bolsa para la cintura le permite transportar agua mientras se desplaza, al tiempo que le deja las manos libres.*

Cantimplora con taza. *Ésta tiene una taza de plástico como tapa, igual que un termo.*

Pequeña bolsa para agua. *Puede doblarse cuando no se utiliza. No la deje en el suelo, ya que se rompe fácilmente.*

Cantimplora plegable. *Éste es un tipo de recipiente tradicional y muy popular. Tenga cuidado de no romper el asa que une el tapón con la botella, ya que podría perder el tapón.*

Bolsa grande de agua. *Muy útil para un campamento grande, esta bolsa de agua puede colgarse de un árbol desprotegido del viento para mantener frío su contenido.*

CÓMO TRANSPORTAR AGUA

Si ir a buscar agua es una actividad diaria, y si el campamento está a cierta distancia de la fuente diaria —como se aconseja para un campamento seguro—, llene grandes recipientes de agua para después transportarlos hasta el campamento. Lo mejor es llevar el envase a la espalda, dejando las manos libres para sortear los obstáculos. Si es posible, utilice una botella pequeña para llenar el recipiente. Esto le permite aprovechar mejor el agua.

Transporte de agua. *Construya un armazón* (véase página 135) *para cargar un recipiente grande.*

CONSTRUCCIÓN DE UN RECIPIENTE DE CORTEZA

1 Se puede improvisar un recipiente para agua a partir de corteza verde y flexible, como la del abedul. Corte un trozo en forma de rectángulo.

2 Sumerja las esquinas en agua para ablandar la corteza y doble los lados con suavidad.

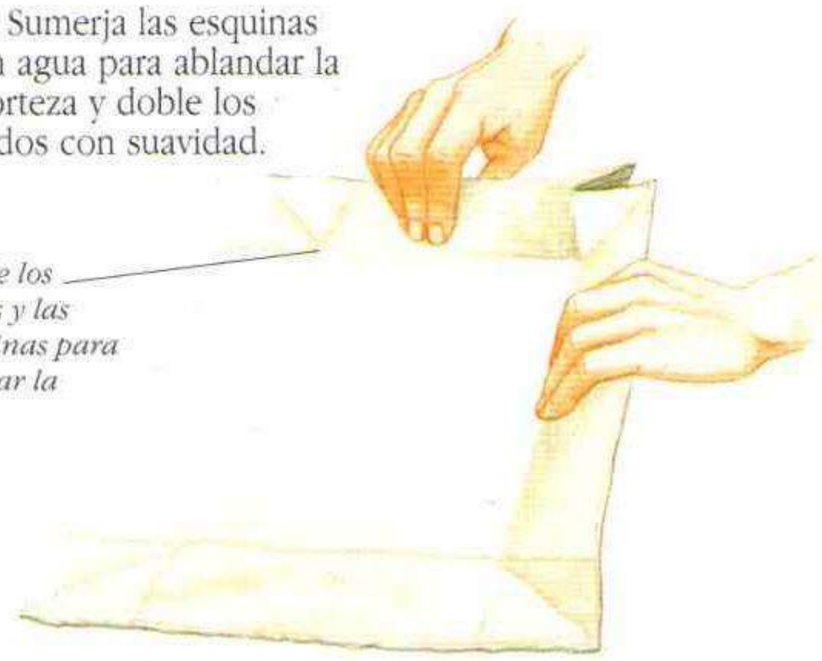

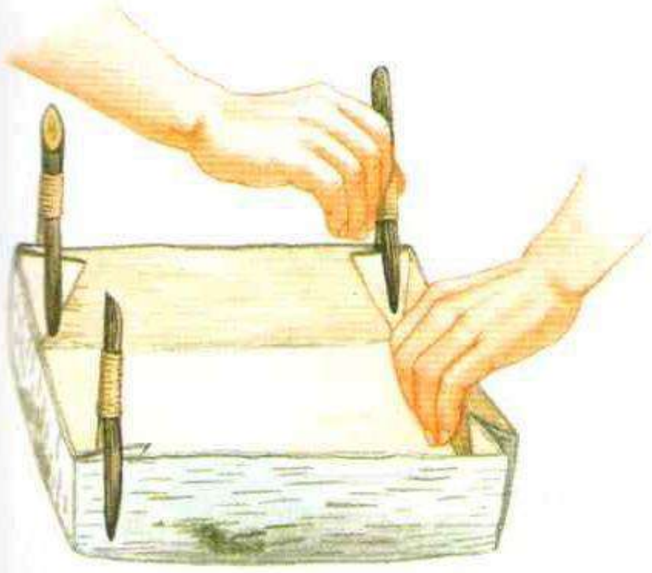

3 Construya unas pinzas a partir de ramitas partidas longitudinalmente y atadas con hierbas. Encole los lados de la caja con resina de pino, que puede desprender de la corteza de estos árboles. Deje las pinzas hasta que se seque.

4 Cuando el recipiente esté seco y retenga la forma, elimine las pinzas e impermeabilícelo con resina de pino empleando una rama astillada como pincel. Deje que la caja se seque antes de utilizarla.

RECIPIENTES IMPROVISADOS

Segmento de bambú. *Los tallos de bambú están formados por segmentos huecos divididos por «paredes» y constituyen excelentes recipientes para agua. Corte un tallo grueso de bambú unos 2 cm por debajo de una unión entre segmentos, y en el mismo lugar del segmento siguiente. Así obtendrá un recipiente hueco con un extremo abierto.*

Calabazas. *Las calabazas pueden vaciarse para emplearlas como recipientes para agua. Corte la parte superior, saque toda pulpa y deje las paredes limpias; utilice un cuchillo o una rama de punta afilada.*

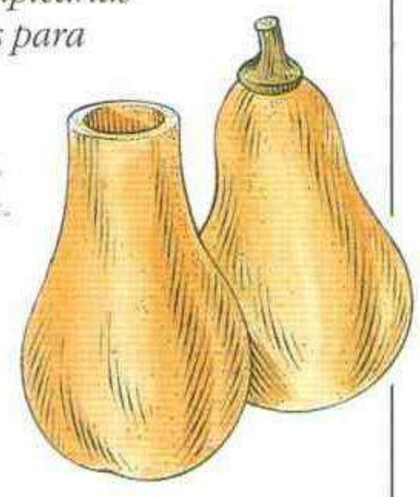

LAS REGLAS DEL AGUA

- Filtre y purifique siempre el agua recogida *(véase página 74)*.
- Intente recoger la cantidad suficiente de agua para cubrir sus necesidades y para guardar una reserva de emergencia para cuando no pueda encontrarla.
- Cuando llueva, disponga la mayor cantidad de recipientes posible para recoger el agua de la lluvia –nunca se sabe cuándo puede necesitarse.
- No la desperdicie nunca, ni cuando la tenga en exceso, ya que algún día le puede hacer falta.
- Nunca beba orina o agua de mar sin tratar. En caso de emergencia puede destilarlas *(véase página 75)*.
- Beba agua en abundancia en condiciones de frío intenso como lo haría en climas más cálidos, ya que caminar en la nieve o esquiar pueden hacerle sudar, y la falta de agua podría causarle deshidratación.
- No coma si no tiene una reserva adecuada de agua, porque necesitará agua para la digestión.

CÓMO ALMACENAR AGUA

1 Puede que tenga que almacenar agua durante la época lluviosa para emplearla en los meses de sequía, o porque necesite más agua que la que pudiera transportar desde la fuente. Escoja un lugar hacia el cual fluiría el agua naturalmente, por ejemplo una pequeña hondonada, y cave hasta encontrar arcilla o piedra.

2 Cubra el hoyo con una tela impermeable o recúbralo con arcilla húmeda, alisándola hasta que se convierta en una superficie hermética. Cubra el hoyo durante el día con una cubierta de vinilo o teja una estera de vegetación *(véase página 49)* para reducir al mínimo la evaporación. Retire la cubierta de noche y cuando llueva para que la lluvia y el rocío caigan en el estanque. Recoja el agua con regularidad, antes de que se evapore.

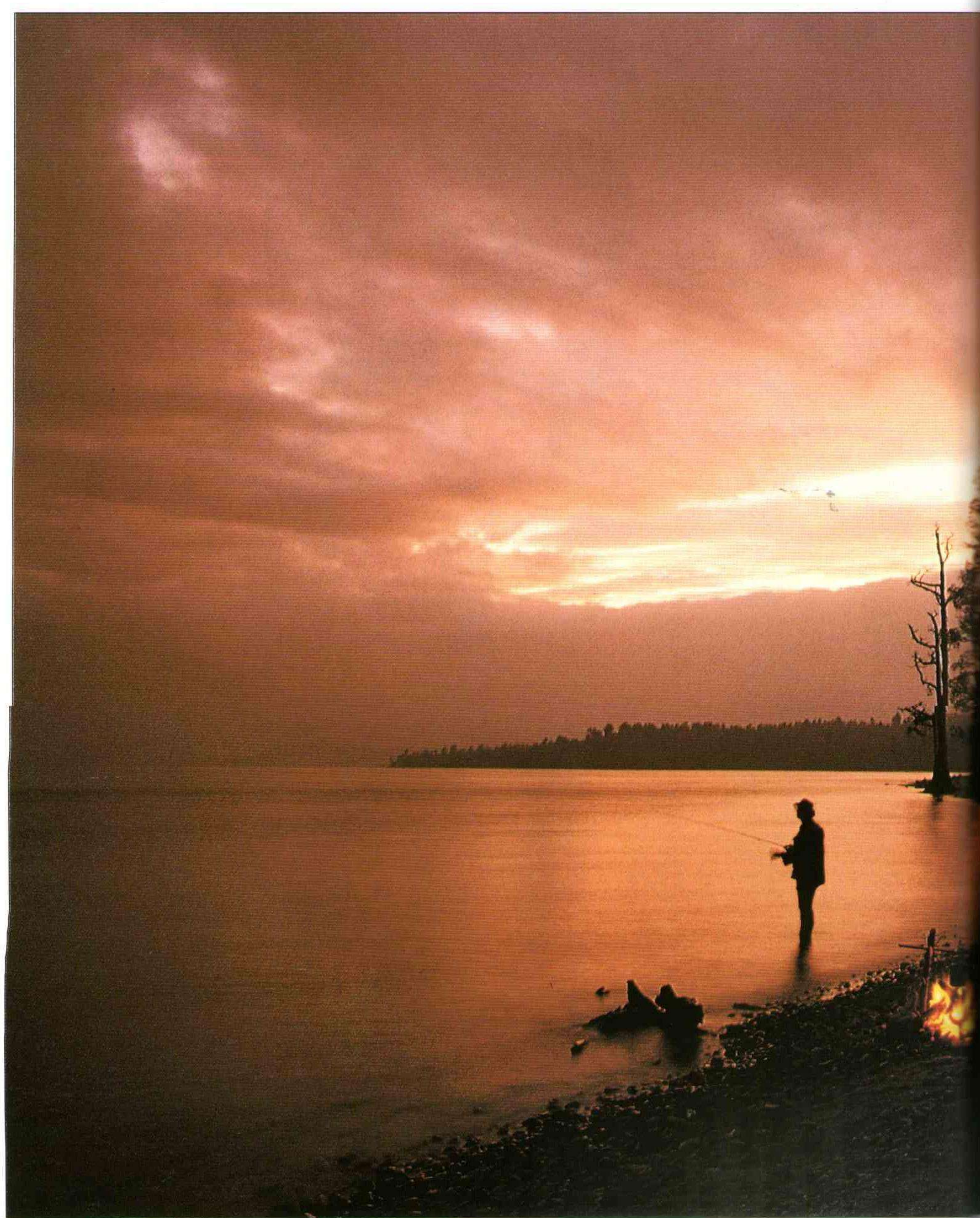

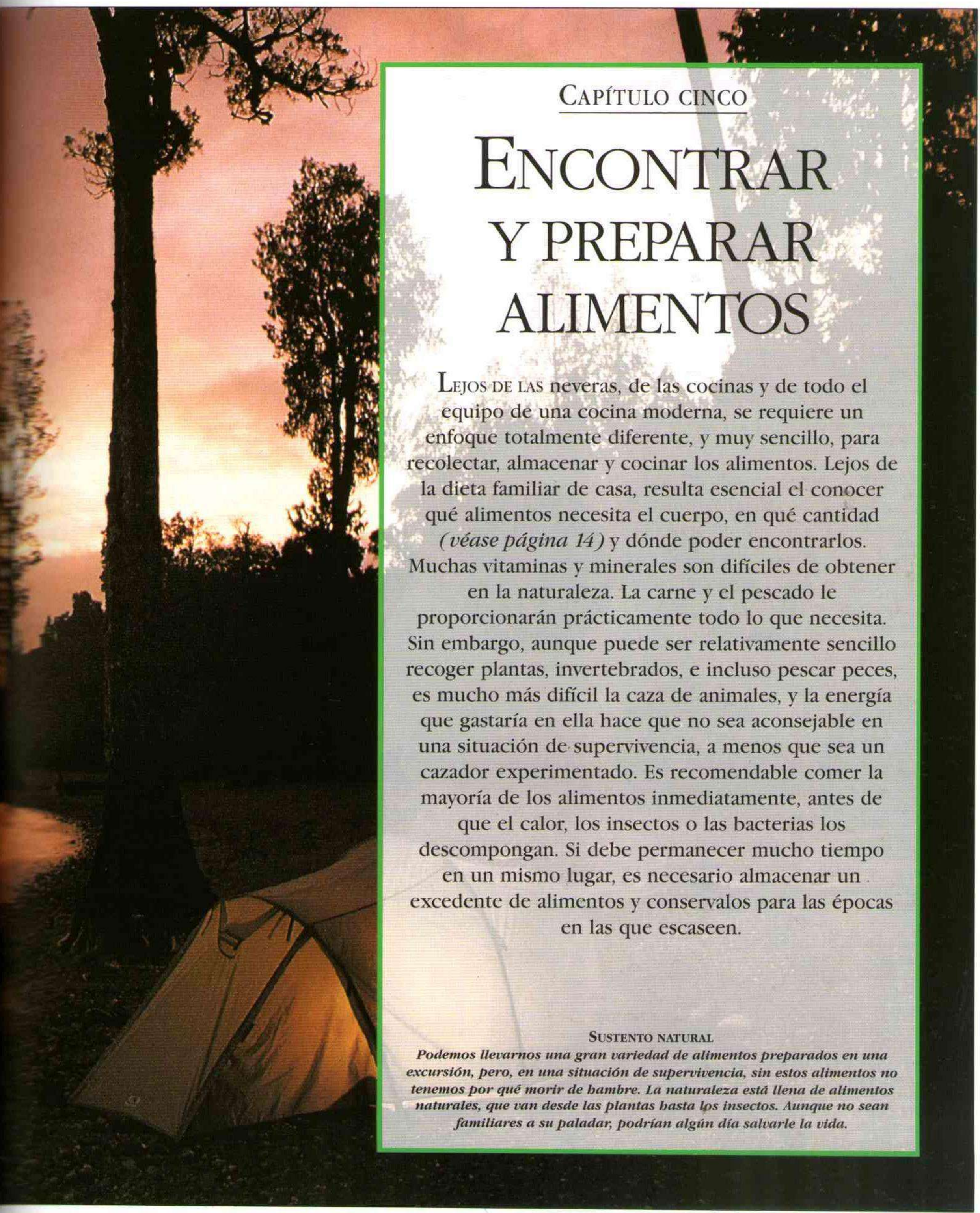

CAPÍTULO CINCO

ENCONTRAR Y PREPARAR ALIMENTOS

LEJOS DE LAS neveras, de las cocinas y de todo el equipo de una cocina moderna, se requiere un enfoque totalmente diferente, y muy sencillo, para recolectar, almacenar y cocinar los alimentos. Lejos de la dieta familiar de casa, resulta esencial el conocer qué alimentos necesita el cuerpo, en qué cantidad *(véase página 14)* y dónde poder encontrarlos. Muchas vitaminas y minerales son difíciles de obtener en la naturaleza. La carne y el pescado le proporcionarán prácticamente todo lo que necesita. Sin embargo, aunque puede ser relativamente sencillo recoger plantas, invertebrados, e incluso pescar peces, es mucho más difícil la caza de animales, y la energía que gastaría en ella hace que no sea aconsejable en una situación de supervivencia, a menos que sea un cazador experimentado. Es recomendable comer la mayoría de los alimentos inmediatamente, antes de que el calor, los insectos o las bacterias los descompongan. Si debe permanecer mucho tiempo en un mismo lugar, es necesario almacenar un excedente de alimentos y conservalos para las épocas en las que escaseen.

SUSTENTO NATURAL

Podemos llevarnos una gran variedad de alimentos preparados en una excursión, pero, en una situación de supervivencia, sin estos alimentos no tenemos por qué morir de hambre. La naturaleza está llena de alimentos naturales, que van desde las plantas hasta los insectos. Aunque no sean familiares a su paladar, podrían algún día salvarle la vida.

ALIMENTOS PARA EL VIAJE

EN UNA EXCURSIÓN puede llevarse comida enlatada, deshidratada en bolsas de aluminio, o en su estado natural en algún tipo de recipiente. Los alimentos enlatados son fáciles de cocinar y comer (fríos o calientes), pero son pesados. Los alimentos deshidratados son mucho más ligeros, pero necesitan una gran cantidad de agua para rehidratarlos, y en algunas regiones se vería obligado a transportar agua con este único propósito. Debe resistirse a la tentación de comer un alimento antes de rehidratarlo totalmente, ya que absorbería el agua de su cuerpo causando obstrucciones intestinales molestas o incluso peligrosas.

Alimentos altos en calorías. *Los dulces tienen un bajo contenido nutricional, pero son siempre bienvenidos. Los montañeros y exploradores polares obtienen a menudo la gran cantidad de calorías necesarias para la supervivencia de los dulces* (véase página 15).

CHOCOLATE EN POLVO

HELADO DESHIDRATADO

PREPARADO PARA FLAN DE FRESA

CARAMELOS CON SABOR A FRUTAS

Alimentos para el desayuno. *Los alimentos para el desayuno son fundamentales para obtener energía al comienzo del día. Los cereales como la avena y el* muesli *también aportan fibra que evita las obstrucciones del tracto digestivo.*

NUECES Y SIMILARES

MUESLI

FRUTOS SECOS

AVENA

Granos y legumbres. *El arroz proporciona los hidratos de carbono necesarios, pero puede ser difícil de preparar en una cocina de campamento. Las judías y las legumbres proporcionan fibra y proteínas. Deben estar bien remojadas antes de cocinarlas, en especial las judías. Siga las instrucciones del fabricante.*

ARROZ

LENTEJAS

JUDÍAS

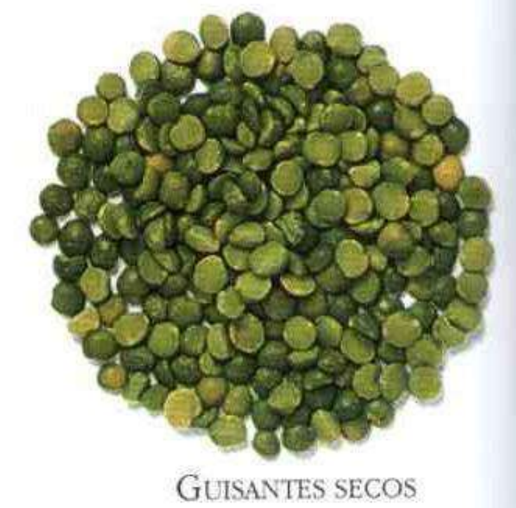
GUISANTES SECOS

Tentempiés. *Para mantener la energía, lo mejor es ir picando continuamente durante el día para evitar las punzadas de hambre, y hacer únicamente una comida importante al día, que pueda digerir totalmente durante la noche. Estos tentempiés incluyen caramelos, chocolate y galletas. Ellos aportan energía e hidratos de carbono.*

GALLETA DE MANTEQUILLA

GALLETA DE CHOCOLATE

GALLETAS DIGESTIVAS

BARRA DE GRANOLA

TABLETA DE CHOCOLATE

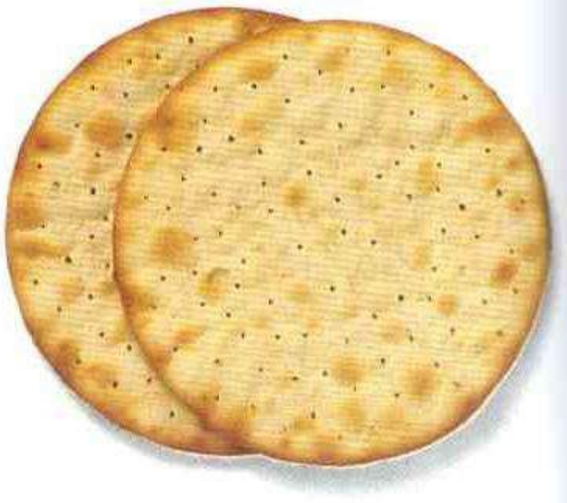
GALLETAS SALADAS

Bebidas. *Las bebidas son un lujo que aporta poco contenido nutricional. Sin embargo, proporcionan una sensación de calor y comodidad. La leche en polvo puede ser una fuente de calcio y da mejor sabor a las bebidas. El chocolate a la taza es una bebida deliciosa en una tarde fría.*

CAFÉ

TÉ

CHOCOLATE A LA TAZA

LECHE EN POLVO

Comidas principales. *La liofilización es un método de deshidratación que conserva intacta la estructura y la textura de los alimentos, en especial la de la fruta. Los alimentos deshidratados son ligeros y fáciles de transportar, pero deben rehidratarse en agua antes de cocinarlos.*

HARINA DE SOJA

COMIDA LIOFILIZADA

COMIDA DESHIDRATADA

COMIDA VEGETARIANA DESHIDRATADA

Condimentos. *La harina y la sal son artículos de primera necesidad; a partir de ellos pueden prepararse una gran variedad de platos* (véase página 117). *El sebo puede ser de gran valor si no puede obtener las grasas necesarias a partir de su comida. El azúcar hará más gustosa la comida silvestre.*

HARINA

SEBO

AZÚCAR MORENO

SAL

Condimentos. *Cualquier condimento que realce el sabor de una comida de campamento vale la pena de llevar* (véase página 93). *El curry en polvo da un sabor distinto a los alimentos, al igual que el ajo, las cebollas, las especias, la salsa de tomate y demás salsas.*

CUBOS DE CALDO DE CARNE

SALSA DE CEBOLLA

SOPA

SALSA DE TOMATE

CARNE Y PESCADO

Aunque la carne roja y el pescado proporcionan las proteínas necesarias, cualquier tipo de carne es difícil de conservar fresca. Lo mejor es llevarla enlatada o en conserva. Una vez las latas se han abierto, su contenido debe consumirse inmediatamente, por lo que debe comprar latas que se puedan consumir en una sola comida. Los alimentos enlatados pueden ser, sin embargo, pesados para llevar en una mochila. Las carnes curadas, como el salami, son más ligeras, pero su variedad es limitada. Puede conservar su propia carne o pescado de muchas maneras *(véase página 118)*. Una alternativa a los alimentos enlatados o curados es una selección de alimentos completos preparados a base de proteína de soja o *tofu*, así como la pasta y los vegetales.

SARDINAS

PASTA

SALCHICHAS Y JUDÍAS

SALAMI

PLANTAS COMESTIBLES

A MENOS QUE se encuentre en medio de un desierto árido, siempre habrá plantas a su alrededor, y muchas de ellas serán comestibles y nutritivas. El problema radica en conocer cuáles de estas plantas son nutritivas y en evitar las venenosas o las que podrían causarle alguna enfermedad. Algunas plantas nutritivas tienen finas vellosidades o aristas que producen irritaciones en la boca y en la garganta, o bien sus semillas son venenosas. Más aún, algunas plantas no son comestibles —o incluso son venenosas— en determinadas épocas del año. Algunas de ellas son comestibles pero no muy nutritivas; su recolección e ingestión consumen más calorías de las que realmente aportan.

PLANTAS DE CLIMA TEMPLADO

Evite las plantas y brotes viejos, que son fibrosos y duros, y pueden tener un sabor desagradable. Los brotes jóvenes son mucho mejores que los viejos, e incluso pueden ser comestibles sin necesidad de cocinarlos. También debe elegir las hojas jóvenes en lugar de las viejas, y debe cortarlas enteras. Aquí encontrará algunos ejemplos de plantas comestibles que crecen en climas templados.

Col marina (Crambe). *La col marina crece en las costas templadas. Sus gruesas hojas y tallos subterráneos pueden comerse hervidos; su sabor recuerda al de la col.*

Pimienta de muros (Sedum). *La pimienta de muros crece en rocas, paredes y guijarros en las regiones templadas. Tiene un fuerte sabor a pimienta. Sus hojas pueden comerse crudas como ensalada, o puede cocerlas para utilizarlas para dar sabor a una sopa.*

OTRAS PLANTAS DE CLIMA TEMPLADO

Algunas algas como las ovas *(Porphyria)* pueden ser hervidas, horneadas o secadas. El musgo de Islandia *(Chondrus crispus)* produce una especie de gelatina que puede utilizar para preparar un postre *(véase página 93)*. Las hojas del diente de león *(Taraxacum)* y de la acedera *(Rumex)* pueden cocinarse como espinacas, mientras que el ajo silvestre *(Allium)* puede utilizarse como condimento.

Bistorta (Polygonum). *Esta planta se encuentra en toda Europa, y algunas especies afines en Norteamérica. Crece hasta una altura de 60 cm en áreas herbosas o en bosques de zonas templadas, y es fácilmente reconocible por su larga cabeza floral de color rosa. Las hojas y brotes pueden comerse hervidos. Las raíces pueden comerse asadas, pero primero deben remojarse en agua.*

Berro terrestre (Barbarea). *Esta planta crece en Europa, Norteamérica y Nueva Zelanda. Las hojas pueden comerse crudas o hervidas.*

Saúco trepador (Aegopodium). *Esta hierba se encuentra a lo largo y ancho de Europa y Asia. Las hojas jóvenes y los tallos frondosos adquieren su mejor sabor cuando los brotes tienen unos 15 cm de longitud.*

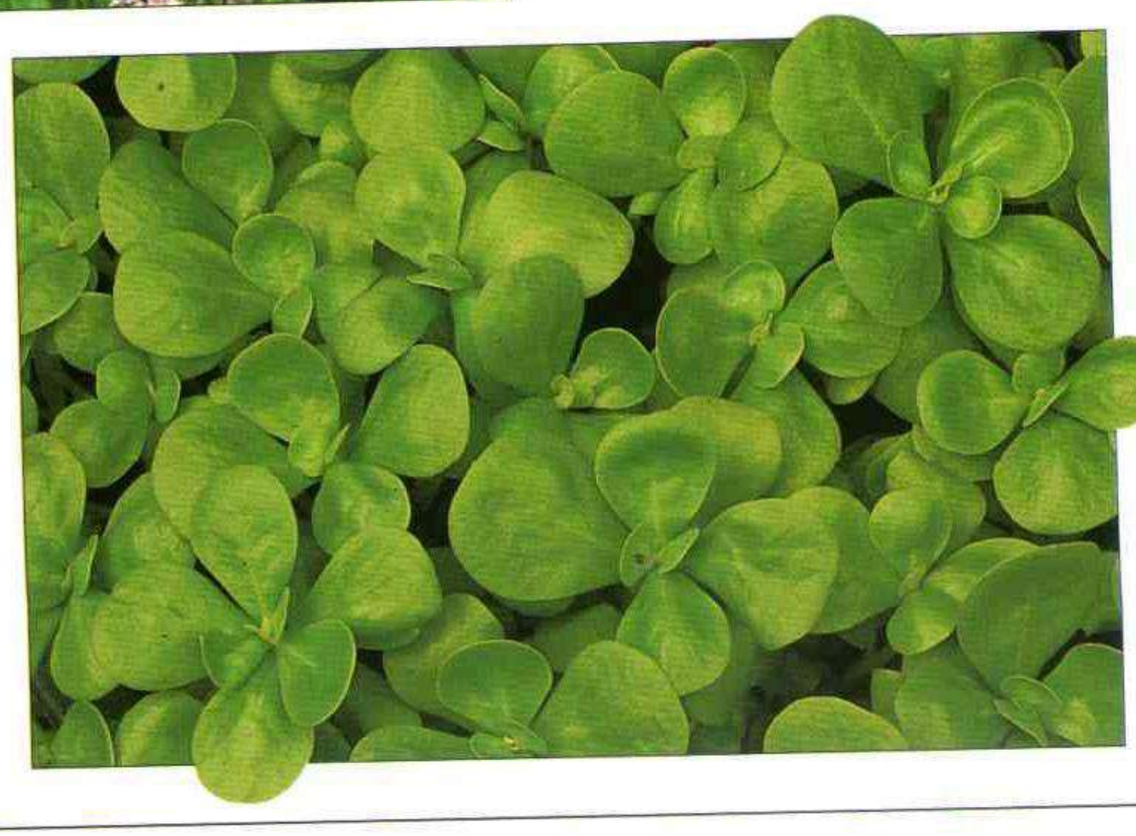

Verdolaga (Portulaca). *La verdolaga crece en climas templados, en las marismas debajo de la línea de la marea alta. Las hojas pueden recolectarse en cualquier época del año. Hiérvalas en agua y sazónelas con zumo de limón. Coma la verdolaga con otras plantas o alimentos, ya que si únicamente come un tipo de alimento, puede caer en un estado de desnutrición.*

PLANTAS TROPICALES

Con el calor y la humedad de los trópicos se desarrollan todo tipo de plantas. Con la ausencia de estaciones, crecen durante todo el año. Existen muchas variedades diferentes, siendo particularmente nutritivos muchos frutos y hortalizas. Pruebe las plantas que no le sean familiares antes de comerlas. Aquí se describen algunas especies comunes encontradas en zonas tropicales.

Bambú. (Pseudosasa). *Corte los tallos jóvenes de bambú desde la base de la planta y abra el revestimiento exterior con un cuchillo. La tierna pulpa interior puede hervirse en agua como si fueran espárragos. También puede comer las semillas. Existen muchos tipos diferentes de bambú, distribuidos en el mundo entero.*

Amaranto (Amaranthus). *Esta planta puede alcanzar hasta 1 m de altura. Corte las hojas y los tallos en pequeños trozos y hiérvalos en agua con sal como si fueran espinacas. Los brotes jóvenes pueden comerse crudos en ensalada.*

OTRAS PLANTAS TROPICALES

Las especies de plantas varían de región a región, y cientos de ellas son comestibles. Muchas tienen frutos deliciosos, aunque será incapaz de alcanzarlos si crecen en la parte más alta de una selva tropical. Las plantas que pueden alcanzarse incluyen la espinaca de Ceilán *(Basella)*, cuyos tallos y hojas son ricos en vitaminas, y la caña de azúcar *(Saccharum)*, cuyos tallos pueden masticarse crudos.

PRUEBA DE SABOR

En primer lugar triture una hoja. Si huele mal o a almendras, deséchela. Frote la savia en la parte interior de su brazo. Si no le produce irritación, coloque un trozo pequeño en sus labios, después en la comisura de la boca, a continuación en la punta de la lengua, y finalmente bajo la lengua; cada vez durante cinco segundos. Si no es urticante, trague una pequeña cantidad y espere cinco horas. Si no se presenta ninguna reacción desagradable, puede comerla.

Palmas. *Los brotes jóvenes de algunas palmas, como la del cocotero* (Cocos), *la palma de azúcar* (Arenga) *y la palma sagú* (Metroxylon) *son comestibles. La palma sagú también tiene una médula feculenta en el interior del tronco que puede hervirse para preparar un pudín de sagú, semejante al de arroz. Sin embargo, no todas las palmas son comestibles, debe aprender a reconocerlas.*

Corazones de palma. *La punta o «corazón» de algunas palmas es comestible, tanto cruda como cocida. No coma frutos de palma a menos que pueda identificarlos.*

ÁRBOLES

La mayoría de los árboles producen frutos, bayas o nueces de algún tipo, que pueden ser muy nutritivos *(véase página 88)*. Los brotes, yemas y corteza interior de muchos árboles también son comestibles, y la savia de algunos árboles puede beberse. Las agujas de algunos árboles perennes son ricas en vitamina C, y pueden remojarse en agua caliente para hacer té *(véase página 93)*.

Pícea (Picea). *La corteza interior de la pícea es rica en vitamina C, especialmente en la parte inferior del tronco, cerca de las raíces. Pele la corteza exterior y hierva la corteza interior hasta que esté tierna, antes de comerla. Utilice las agujas para preparar una bebida refrescante* (véase página 93).

Arce (Acer). *El arce tiene una savia azucarada que puede recogerse de las aberturas naturales, o realizando un corte en la corteza. Déjela hervir hasta que se espese para formar un jarabe rico en energía.*

OTROS ÁRBOLES

Las agujas jóvenes del pino *(Pinus)* tienen un buen sabor y pueden utilizarse para hacer té *(véase página 93)*. La corteza interior del álamo temblón *(Populus)* es muy nutritiva, mientras que la savia del abedul *(Betula)* es deliciosa. Recoja la savia cortando una «V» en la corteza, pero no la haga demasiado ancha. Otros árboles comestibles son el algarrobo *(Ceratonia)* y el tamarindo *(Tamarindus)*.

PLANTAS VENENOSAS

EL NÚMERO DE plantas venenosas es mucho menor que el de las comestibles, particularmente en las regiones templadas. Algunas son venenosas por contacto, y causan urticaria o graves irritaciones *(véase página 181)*. Otras son tóxicas por ingestión y pueden provocar vómitos, diarrea, e incluso la muerte. Algunas plantas venenosas se asemejan a las comestibles, así que la única manera de evitar un desastre consiste en aprender a diferenciar las plantas venenosas. Tenga siempre un cuidado extremo al probar las plantas *(véase página 83)*, particularmente en una situación de supervivencia, en la cual puede que no disponga de asistencia médica.

PLANTAS DE CLIMA TEMPLADO

Las hojas y tallos de algunas plantas son venenosas, así como algunas flores, raíces y savia. Incluso el roce con otras plantas puede inducirlas a liberar el veneno almacenado en vellos o poros. Existen muchas más plantas comestibles que venenosas en las regiones templadas, pero las plantas tóxicas pueden muchas veces ser mortales.

Dedalera (Digitalis). *Todas las partes de la dedalera contienen digitalina, un estimulante cardíaco altamente tóxico. La planta puede alcanzar una altura de 1,5 m, y tiene flores muy reconocibles, de forma acampanada y colores violeta, rosa, amarillo o blanco.*

Altramuz (Lupinus). *Esta planta crece en Norteamérica, Europa y Asia. Todas sus partes causan inflamaciones mortales del estómago y los intestinos. La planta tiene flores púrpuras, rosas, blancas o amarillas.*

OTRAS PLANTAS DE CLIMA TEMPLADO

Una de las plantas más tóxicas de las regiones templadas es el acónito *(Aconitum)*, con flores de color púrpura. La mortal cicuta *(Conium)* y la cicuta acuática *(Cicuta)* se reconocen por sus flores blancas, dispuestas en forma de paraguas. Evite cualquier planta umbelífera: la mayoría son peligrosas. El zumaque venenoso y el árbol de las pulgas *(Toxicodendron)* causan irritación por contacto.

Ricino (Ricinus). *Aunque esta planta es originaria de las regiones tropicales, ha sido introducida en las áreas templadas. Se cultiva a menudo por el aceite con efecto purgante obtenido de las semillas. Sin embargo, si estas semillas se ingieren en su estado silvestre, pueden ser mortales.*

ADVERTENCIA

Muchas plantas venenosas comunes de las regiones templadas se asemejan a alguna planta comestible, así que, a menos de que esté totalmente seguro de identificarla correctamente, evite comerla. Incluso un bocado de alguna de ellas, como la cicuta acuática (véase superior) puede causar la parálisis y la muerte en pocos minutos.

Zigadenus. *El bulbo letal de esta planta puede confundirse con el de una cebolla silvestre* (véase página 86). *Tiene largas hojas y flores blancas, y crece hasta unos 60 cm de altura. Se encuentra comúnmente en áreas boscosas y en pastos de Norteamérica.*

Hiedra venenosa (Toxicodendron). *La hiedra venenosa se encuentra en los bosques de Norteamérica. Puede ser reptante o de crecimiento vertical, como una hiedra común, pero se distingue por sus flores verdosas y sus bayas blancas. Produce urticaria al contacto con la piel.*

Botón de oro (Ranunculus). *Todas las especies de este género causan graves y dolorosas inflamaciones de los intestinos, en caso de ingestión. Se encuentran en todo el mundo, y todas las especies tienen al menos cinco pétalos amarillos, brillantes y superpuestos.*

PLANTAS TROPICALES

Las plantas tropicales son tan variadas y abundantes que, por su propia seguridad, sólo debe comer aquéllas que pueda identificar. Si debe considerar la opción de comer una especie que no le sea familiar, aplique en primer lugar la prueba de sabor *(véase página 83)*, y coma únicamente una cantidad pequeña. A continuación se describen algunos ejemplos de plantas venenosas.

Jatropha glandulosa. *Varias especies de plantas tropicales del género* Jatropha *tienen hojas, semillas, savia o frutos venenosos. Las semillas tienen un fuerte efecto purgante.*

Sapium insigne. *La savia de esta planta es lechosa y altamente tóxica. No la ponga en contacto con su piel ya que le produciría ampollas. Evite todas las plantas con savia lechosa, ya que es muy probable que sean venenosas.*

OTRAS PLANTAS TROPICALES

La ortiga *(Laportea)* crece cerca del agua en las regiones tropicales. Causa urticaria por contacto, y sus semillas son venenosas. La manzana de la playa *(Hippomane)* tiene frutos y savia venenosos. *Mucuna* crece en los bosques y matorrales, y sus vainas y flores irritan la piel al contacto. En contacto con los ojos, estas partes de la planta pueden causar ceguera.

Jatropha podagrica. *Esta planta tiene la base semejante a un nabo, pero todas sus partes son muy venenosas. La única manera de identificar las plantas desconocidas es reconociéndolas a partir de una fuente fidedigna, como una guía local, o a través de la gente que habita en la zona.*

Jatropha integerrima. *Esta planta se encuentra en las regiones boscosas de los trópicos. Todas sus partes pueden causar inflamación de los intestinos, además de diarrea y vómitos. Estas condiciones pueden provocar deshidratación* (véase página 181), *que podría amenazar su vida en una situación de supervivencia, si no es capaz de encontrar ayuda médica inmediata.*

ÁRBOLES

No existe regla alguna para distinguir los árboles venenosos de los que no lo son. Algunas partes de algunos árboles pueden comerse, mientras que otras son venenosas. Debe tener especial cuidado con los árboles tropicales, ya que las semillas y los frutos de algunos de ellos pueden ser mortales, mientras que la savia de otros puede causar ampollas o ceguera.

Mangle cegador (Excoecaria). *Este pequeño árbol se encuentra en los manglares y estuarios de Australasia, sudeste de Asia y África tropical. Debe evitarlo siempre que sea posible, ya que su savia produce ampollas en la piel, y ceguera al contacto con los ojos.*

Laurel de California (Umbellularia). *El laurel de California crece en Norteamérica. Sus resistentes hojas ovaladas y perennes son fuertemente aromáticas y venenosas. Tiene flores amarillas y bayas verdes o púrpuras.*

OTROS ÁRBOLES

Las vainas del codeso *(Laburnum)* son mortales, como lo son las bayas del tejo *(Taxus)*. El algarrobo negro *(Robinia)* de Norteamérica contiene sustancias tóxicas en la corteza, flores y vainas. Los árboles rhengas *(Gluta)* del sudeste asiático tienen una savia irritante, mientras que la savia del árbol arenero *(Hua)* de Sudamérica puede producir ceguera. Evite comer cualquier parte de un árbol a menos que pueda identificarlo con seguridad.

RAÍCES, TUBÉRCULOS Y BULBOS

LAS RAÍCES, TUBÉRCULOS y bulbos de algunas plantas son ricos en vitaminas, tienen un contenido especialmente elevado de fécula y pueden comerse en situaciones de supervivencia. Algunas raíces son tóxicas si se comen crudas, por lo que se recomienda cocinarlas bien antes de comerlas. Límpielas bien, hiérvalas después y termine asándolas al fuego *(véase página 116)*. La piel de muchas raíces es rica en vitaminas y minerales, así que evite pelar las raíces si las pieles están en buen estado. Sin embargo, algunas raíces deben pelarse para eliminar sustancias peligrosas. Aprenda a reconocer algunas raíces comestibles y deseche las demás.

RAÍCES COMESTIBLES

La única manera de determinar si una raíz o bulbo es comestible es aprendiendo a reconocer la planta que genera. Las plantas de diferentes regiones pueden parecerse, o tener bulbos similares, así que no suponga que porque un bulbo le parezca familiar es comestible; así, por ejemplo, el bulbo de la mortal *Zigadenus* parece el de una inofensiva cebolla *(véase página 84)*. Casi todos los bulbos y raíces deben hervirse antes de comerlos.

Salsifí (Tragopogon). *El salsifí tiene entre 60 y 90 cm de altura y crece en áreas secas. Tiene flores parecidas a las del diente de león. Sus hojas largas y el tubérculo parecido a la chirivía son comestibles, y llega a cultivarse.*

Castaño acuático (Trapa). *El castaño acuático crece en aguas dulces de Europa y Asia. Sus duras semillas son comestibles, tanto crudas como tostadas.*

Nenúfar (Nymphaea). *El nenúfar tiene hojas flotantes en forma de corazón, y crece en aguas dulces, tropicales y templadas, casi en todo el mundo. Sus raíces, tubérculos y tallos son comestibles, aunque las semillas son ligeramente amargas.*

OTRAS RAÍCES COMESTIBLES

Las raíces del diente de león *(Taraxacum)*, de la juncia *(Cyperus)*, y de algunas otras plantas pueden tostarse para obtener un sustituto del café. Las raíces de plantas como el aro *(Calla)* pueden molerse hasta obtener un polvo fino que puede utilizarse como harina *(véase página 92)*. Las raíces del palo dulce *(Astralagus)* pueden comerse crudas o cocidas, al igual que los tubérculos de la alcachofa de Jerusalén *(Helianthus)*, que crece silvestre en Norteamérica.

Saetas de agua (Sagittaria). *Pueden encontrarse cerca del agua dulce. Crece hasta 1 m de altura y tiene hojas erectas en forma de saeta, y pequeñas flores de tres pétalos. Sus tubérculos son comestibles en crudo, aunque son mejores cocidos.*

Ácoro o lirio dulce (Acorus). *El reconocible ácoro puede crecer hasta una altura de 1,3 m, y siempre crece junto a agua dulce. Sus raíces son fuertemente aromáticas y comestibles. Deben cortarse y hervirse hasta formar un jarabe, antes de comerlas.*

Cacahuete (Arachis). *Los cacahuetes se obtienen de las vainas que crecen bajo tierra, unidas a los tallos. La planta del cacahuete es pequeña y arbustiva, con hojas ovaladas y flores amarillas y planas.*

RAÍCES VENENOSAS

Hablando en términos generales, si las raíces de una planta son venenosas, sus hojas, tallo, flores y savia también lo serán. La celidonia menor es un buen ejemplo —sólo con tocar la planta aparecen ampollas en la piel. Sin embargo, hay excepciones a esta regla. Algunas raíces únicamente son comestibles una vez hervidas, y pueden ser terriblemente venenosas si se comen crudas. Tenga siempre mucho cuidado. A menos que esté seguro de haber identificado una planta correctamente, no la coma.

Mandioca (Manihot). *Los tubérculos de la mandioca o yuca son letales si se comen crudos. Deben remojarse durante 48 horas y hervirse a conciencia antes de comerlos.*

ADVERTENCIA

Ya que resulta tan difícil distinguir las raíces venenosas de las comestibles —a menos que la planta sea fácilmente reconocible— es mejor evitar comer raíces. Mientras pueda identificar con seguridad unas cuantas especies, aténgase a éstas e ignore las demás.

Narciso silvestre (Narcissus). *Es fácil de reconocer por su conocida flor, semejante a las variedades domésticas. Si no tiene flores, no toque la planta.*

Aro de los pantanos (Calla). *También conocida como cala silvestre, el aro de los pantanos tiene una espiga floral distintiva dentro de un capuchón. Crece cerca de agua dulce. Las raíces pueden comerse o ser molidas para obtener harina* (véase página 92), *pero deben hervirse antes. Todas las demás partes de la planta son venenosas, y deben evitarse.*

Ñame silvestre (Dioscorea). *Aunque algunas especies de ñame silvestre se cultivan en zonas tropicales, algunas de ellas son venenosas a menos que se pelen y se hiervan. Si no puede identificar las variedades, lo ideal es pelar y hervir los ñames silvestres. Tienen hojas grandes y crecen a menudo alrededor de los troncos de los árboles.*

Taro (Colocasia). *El taro crece en suelos húmedos en las regiones tropicales. La planta alcanza una altura de hasta 1,5 m y tiene grandes hojas verdes en forma de cuña, y una flor de color amarillo naranja. Sus tubérculos son venenosos si se comen crudos, pero pueden consumirse una vez cocidos. Tienen un sabor semejante al de la patata.*

OTRAS RAÍCES VENENOSAS

Los tubérculos de la patata silvestre *(Solanum)* son comestibles, pero deben hervirse antes. Sus frutos, que asemejan a un tomate, son venenosos. El tomate silvestre *(Lycopersicon)*, que es comestible, tiene un aspecto muy similar a la planta de la patata: es aconsejable evitar ambos. *Zigadenus* tiene un bulbo parecido a una cebolla, pero es letal *(véase página 84)*. Los tubérculos y las raíces de la celidonia menor o botón de oro *(Ranunculus)* y de la cicuta acuática *(Cicuta)* son mortales.

Vencetósigo (Vincetoxicum). *Esta planta tiene hojas apuntadas en forma de corazón y pequeñas flores de color blanco amarillento. Si se ingiere, sus raíces y vainas producen vómitos y pérdida de agua; su savia lechosa también es tóxica. Es común en regiones templadas.*

Matacandiles (Ornithogalum). *Esta planta propia de las regiones templadas crece hasta unos 30 cm de altura. Sus bulbos son comestibles siempre y cuando se hiervan, pero evite comer el resto de la planta.*

NUECES Y FRUTOS

TANTO LAS NUECES como los frutos constituyen la fuente más importante de alimento en una situación de supervivencia. En las regiones templadas, los frutos silvestres aparecen a partir de mediados de verano, mientras que las nueces aparecen más tarde, en otoño. La mayoría de los frutos silvestres son difíciles de almacenar a menos que haga una mermelada o seque las especies que lo permiten *(véase página 118)*. Recuerde que, al recoger frutos y nueces, está compitiendo con aves y mamíferos, particularmente cuando se trata de nueces, que algunos animales recogen como fuente principal de alimento para el invierno. No recoja frutos o nueces demasiado maduros o mohosos: podría enfermar.

NUECES COMESTIBLES

Las nueces proporcionan tanto proteínas como grasas. Con las nueces oleosas como las de la haya, puede separar su aceite y almacenarlo para cocinar. Hierva los hayucos lentamente en agua y separe el aceite de la superficie. También pueden ser molidas para obtener harina *(véase página 92)*. Las nueces son el alimento silvestre más fácil de almacenar, y se recomienda comerlas durante un traslado.

Nogal (Juglans). *Los nogales crecen en áreas templadas. Tienen una corteza especialmente nudosa. Las nueces tienen una cáscara verde y con pulpa, que se pudre después de caer del árbol. Las nueces tienen un alto contenido calórico.*

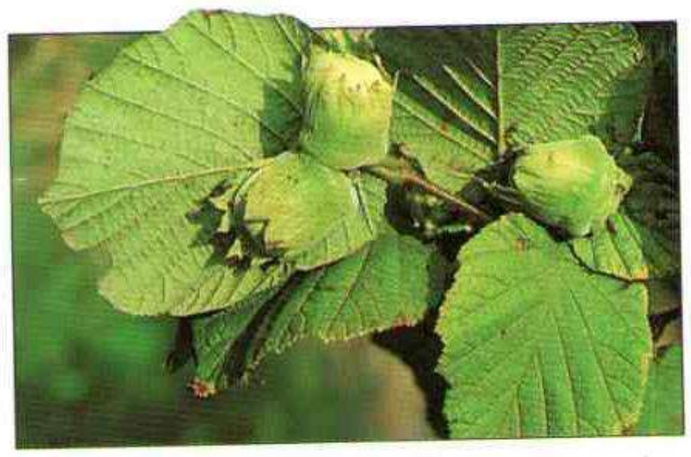

Avellanos (Corylus). *Las avellanas crecen en grandes arbustos y árboles de áreas templadas, de hojas serradas y acorazonadas, y amentos de color amarillo marrón. Las nueces crecen en cáscaras pilosas.*

Pistacho (Pistacia). *Los pistachos crecen en Europa, Asia y Australia. Sus nueces rojas se comen crudas o tostadas.*

OTRAS NUECES COMESTIBLES

La pacana *(Carya)* es relativamente común en las áreas húmedas de Norteamérica. Contiene más grasa que cualquier otra nuez, vegetal o fruto. Las nueces de macadán *(Macadamia)* de Australia también tienen un sabor agradable. Los piñones *(Pinus)* son comestibles.

Castaño (Castanea). *Los castaños tienen largas hojas serradas y cáscaras espinosas, dentro de las cuales se desarrollan las castañas. Para comer las castañas debe pelarlas, hervirlas y reducirlas a una pasta, o tostarlas al fuego.*

NUECES VENENOSAS

Pruebe las nueces con cuidado, y utilice la prueba del sabor *(véase página 83)*. Aunque hay pocas nueces venenosas, algunas pueden ser peligrosas a menos que las hierva, mientras que otras le causarán malestar si están mohosas. Si una nuez tiene un sabor fuera de lo común, dulce o semejante al de las almendras, deséchela porque puede contener ácido prúsico, que causa irritación.

Castaño de Indias (Aesculus). *El castaño de Indias, que crece en las regiones templadas, tiene grandes hojas palmeadas, flores pálidas y yemas pegajosas. Puede alcanzar los 30 m de altura en algunas áreas. Sus hojas venenosas tienen cáscaras menos espinosas que el castaño común, y las nueces dentro de las cáscaras son mucho mayores que las castañas comestibles.*

Anacardo (Anacardium). *Las nueces del anacardo son venenosas a menos que las pele y hierva. Tenga mucho cuidado mientras hierven, ya que los vapores pueden producir ceguera. El anacardo crece en áreas tropicales.*

OTRAS NUECES VENENOSAS

Aunque son comestibles, las almendras *(Prunus)* pueden contener ácido prúsico; tienen como resultado un sabor amargo. Las bellotas de los robles *(Quercus)* también son comestibles, pero deben hervirse o remojarse en agua fría durante varias horas, para después tostarlas a fin de eliminar su sabor amargo.

FRUTOS COMESTIBLES

Los frutos comestibles contienen muchos nutrientes esenciales, que incluyen azúcares y vitaminas A, B_2 y C. Las aves y los mamíferos también lo saben, por lo que tendrá que competir con ellos. Los insectos también atacan la fruta, pero puede beneficiarse de ello comiéndose cualquier gusano nutritivo que llegue a contener. No coma cantidades excesivas de cualquier tipo de fruta.

Camemoro (Rubus). *El camemoro crece en el ártico y en las regiones templadas del norte. Alcanza una altura de unos 30 cm, tiene flores blancas y frutos que se asemejan a las frambuesas.*

Fruta de la pasión (Passiflora). *La fruta de la pasión tiene largos tallos reptantes, hojas trilobuladas y flores de color blanco y violeta que crecen sobre zarcillos. Crece en áreas tropicales, sobre todo en Sudamérica. Los frutos comestibles son de color violeta, ovalados y jugosos.*

OTROS FRUTOS COMESTIBLES

Las uvas silvestres *(Vitis)* son comunes en las zonas cálidas. Sus hojas pueden comerse hervidas. Varias especies de fresa silvestre *(Fragaria)* se encuentran en los bosques y en zonas de hierbas secas. Son ricas en vitamina C, muy dulces y deliciosas.

Arándano (Vaccinium). *Crece en los brezales templados, en la tundra y en los bosques.*

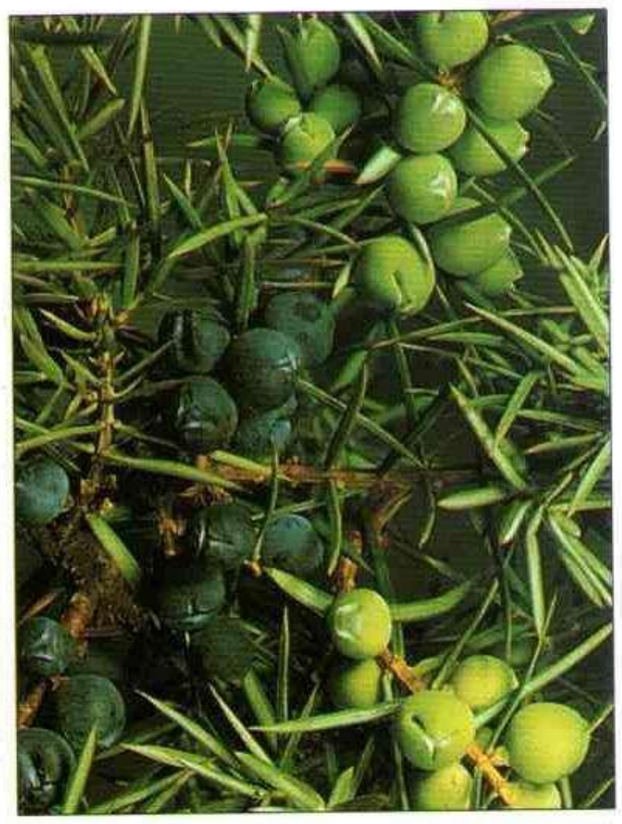

Enebro (Juniperus). *El enebro crece en las montañas templadas del norte, en forma de arbusto grande o pequeño, o en densos matorrales bajos. En algunas partes de Norteamérica se le conoce como* uvas de Oregón. *Las bayas maduras de color azul oscuro son comestibles, aunque un poco amargas.*

FRUTOS VENENOSOS

Ya que los frutos están diseñados para atraer a los animales y a los pájaros (los cuales, una vez han comido el fruto, diseminan las semillas en sus heces), muchas bayas venenosas se asemejan a las comestibles. Esto puede ser una amenaza para quienes intentan sobrevivir a base de alimentos naturales. Al igual que las demás plantas y setas, debe aprender a reconocer algunos frutos comestibles.

Enredadera de Virginia (Parthenocissus). *Las bayas de profundo color azul o púrpura de esta planta de clima templado se asemejan a las uvas. Ninguna baya azul o púrpura de una planta trepadora con zarcillos es comestible.*

Serbal. *Aunque los frutos de algunas especies de* Sorbus *son comestibles, en general las bayas blancas son siempre venenosas. No se arriesgue a comer ninguna baya de este color. Otras plantas con frutos blancos venenosos son la hierba de San Cristóbal y algunos tipos de belladona.*

Duchesnea. *Los frutos de* Duchesnea, *también conocida como* falsa fresa, *pueden ser mortales. La planta crece en las áreas tropicales de Asia y Norteamérica.*

OTROS FRUTOS VENENOSOS

Existen muchos frutos venenosos, en particular en las regiones tropicales. Algunos de ellos son sólo medianamente venenosos y pueden provocar diarreas y náuseas al ingerirse en grandes cantidades. Otros son altamente tóxicos y una pequeña cantidad basta para causar la muerte. En una situación de supervivencia, los riesgos se incrementan, especialmente si se encuentra débil. Algunos frutos venenosos comunes incluyen las bayas azules de la mortal belladona *(Atropha)*, que son letales si se ingieren. *Arum* tiene bayas rojas que pueden causar un grave malestar digestivo si se ingieren. La hierba de San Cristóbal *(Actaea)* tiene bayas blancas o negras que pueden producir vómitos e irritación interna. Los frutos de la manzana silvestre *(Malus)* pueden causar diarrea si se comen en grandes cantidades. El árbol tropical de la estricnina *(Strychnos)* tiene frutos que parecen naranjas y sus semillas son mortales.

HONGOS Y LÍQUENES

PARA MUCHAS PERSONAS en todo el mundo, la recogida de hongos y líquenes se ha convertido en una actividad habitual. Además de ser absolutamente deliciosos, los hongos frescos contienen minerales y algunas proteínas, y constituyen el elemento básico de muchos platos nacionales. Sin embargo, debe tener una gran precaución al recogerlos y prepararlos. Algunos hongos son letales, y muchos se asemejan a las variedades comestibles. Incluso los recolectores experimentados pueden equivocarse: en ocasiones, lo que no representa ningún peligro en un país se asemeja mucho a una variedad venenosa en otro. Consuma hongos únicamente si está seguro de poder identificarlos correctamente.

HONGOS COMESTIBLES

No existe una manera lógica de identificar los hongos comestibles. Recójalos únicamente si va acompañado de un experto y aprenda a identificar una o dos variedades comestibles, con referencia especial a sus hábitats. Aprenda a reconocer las variedades venenosas, en particular las que son mortales. Rechace cualquier hongo que no esté en perfectas condiciones. Aquí se muestran algunos comestibles.

Colmenilla redonda, morilla (Morchella esculenta). *Este hongo blanco crece en primavera en suelos arenosos o arenoso-arcillosos, bajo los árboles o al descubierto.*

Seta coliflor (Sparassis crispa). *Este hongo se encuentra en otoño en los bosques de coníferas; crece al pie de los árboles y sobre sus raíces. Su carne despide un característico olor a anís, y tiene un sabor similar al de una nuez.*

Bejín areolado (Calvatia utriformis). *Se encuentra en verano y en otoño en los campos y prados. Es preferible consumirlo joven, cuando la carne todavía se mantiene blanca.*

Peziza anaranjada (Aleuria aurantia). *Este hongo necesita luz solar para crecer, y se encuentra en los claros de las regiones boscosas y en áreas de pasto en otoño. A pesar de ser comestible, su carne no tiene un olor o sabor destacable.*

OTROS HONGOS COMESTIBLES

Las variedades *Boletus* tienen la forma tradicional de una seta, pero tienen poros en la parte inferior del sombrero, en lugar de láminas. Generalmente son deliciosos. Algunas especies son medianamente venenosas, pero pueden distinguirse de las comestibles ya que su sabor en crudo es desagradable, y tienen los poros y el pie de color rojizo o naranja. Elija únicamente las setas con poros amarillos o de color crema. Los hongos que crecen sobre los troncos de los árboles son generalmente comestibles, pero algunas variedades son excesivamente duras y con poco sabor. Todos los hongos silvestres deben hervirse antes de consumirse, para así destruir los compuestos ligeramente tóxicos. Recuerde, sin embargo, que no con sólo cocinarlos convertirá en comestible una especie venenosa. Los hongos difieren entre sí según la región, por lo que debe aprender a reconocer los de su entorno, o bien los de la región que pretenda visitar.

Políporo frondoso (Grifola frondosa). *Este hongo de árbol, que puede encontrarse desde la primavera hasta el otoño, crece en árboles de hoja caduca. Despide un característico olor a seta, y tiene un sabor dulce.*

Lengua de buey (Fistulina hepatica). *Este hongo es fácil de encontrar durante el otoño, sobre los robles. Su carne roja tiene un sabor amargo en crudo, por lo que debe remojarlo y cocerlo antes de consumirlo.*

HONGOS VENENOSOS

No hay características específicas que permitan distinguir los hongos venenosos. Las amanitas se encuentran entre los más peligrosos, y son las más fáciles de identificar, aunque algunas especies tienen un aspecto totalmente diferente. Peor aún, algunos hongos venenosos se asemejan a los hongos comestibles de otros países. Los síntomas de envenenamiento por hongos son muy variados. A menos que pueda identificar positivamente un hongo como variedad comestible, no lo toque. A continuación se muestran algunas de las especies más peligrosas.

ADVERTENCIA

Aunque los hongos del género *Amanita* incluyen algunos de los más mortales, no todos los hongos venenosos se les parecen. Siempre debe realizar una identificación positiva de una variedad comestible, rechazando las demás a menos que esté bien seguro de que son comestibles.

Amanita pantera (Amanita pantherina). *Suele ser mortal. Se encuentra en los bosques de hayas; tiene un sombrero marrón moteado de blanco, láminas blancas y anillos alrededor del pie.*

Matamoscas (Amanita muscaria). *Este hongo fácilmente reconocible crece durante el otoño, particularmente en los bosques de hayas y de pinos. Tiene un sombrero de color rojo brillante, moteado de blanco.*

Cicuta verde (Amanita phalloides). *La especie más venenosa de todos los hongos se encuentra en las regiones boscosas, bajo hayas y robles. Tiene el sombrero de color verde oliva, el pie blanco, láminas y carne blancas, y una gran volva.*

CÓMO RECONOCER UN HONGO VENENOSO

Los hongos venenosos del género *Amanita* son relativamente fáciles de identificar. Todos tienen una volva en la base, láminas blancas y anillos alrededor del pie.

Amanita. *Esta especie norteamericana de matamoscas se encuentra en los bosques de coníferas. Si se ingiere, provoca un sueño profundo y alucinaciones.*

Amanita maloliente (Amanita virosa). *Es mortalmente venenoso, y se encuentra en verano y otoño en los bosques. Su carne blanca tiene un olor dulce y nauseabundo. Los especímenes jóvenes pueden parecerse a los* Agaricus *comestibles.*

OTROS HONGOS VENENOSOS

Algunas variedades venenosas parecen idénticas a las especies comestibles de otras regiones. Por ejemplo, mientras que muchas especies del género *Agaricus* son comestibles, el agárico amarilleante *(Agaricus xanthodermus)* es muy venenoso. Puede reconocerlo por su intenso olor a yodo y la coloración amarilla que aparece al tocarlo. Otras variedades venenosas incluyen la cicuta blanca *(Amanita verna)*, de color blanco, y que se encuentra en suelos arenosos bajo árboles de hoja ancha, así como varias especies de *Cortinarius*, algunas de las cuales son mortales.

LÍQUENES COMESTIBLES

Los líquenes crecen sobre las rocas, a menudo en las regiones boreales o árticas. No hay variedades venenosas, pero todos deben remojarse una noche en agua y hervirse después a fin de eliminar cualquier ácido irritante, antes de ingerirlos.

Liquen de los renos (Cladonia rangiferina). *Esta planta resistente crece en las regiones árticas y su aspecto recuerda los cuernos de los renos, de ahí su nombre. Antes de comerlo debe remojarlo en agua durante varias horas y hervirlo bien para que resulte digerible.*

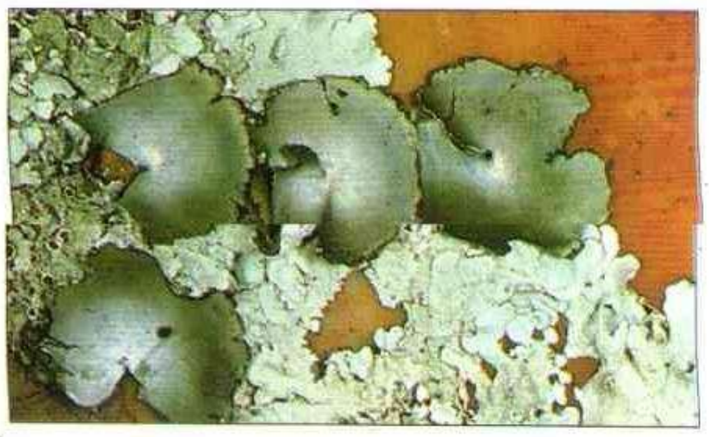

Callos de roca. *Estos líquenes nutritivos crecen en las regiones templadas del norte y en el ártico. Deben remojarse y hervirse antes de consumirse.*

LAS PLANTAS COMO ALIMENTO

LA RECOLECCIÓN NO consiste en recoger cualquier planta que parezca comestible. Debe reunir la cantidad suficiente de una o más especies para preparar una comida. Si se limita a recolectar unas cuantas especies, reducirá las posibilidades de incluir alguna especie no comestible o venenosa *(véase página 84)*. Al recolectar plantas, asegúrese de dejar una cantidad suficiente de cada especie para permitir su regeneración para, de este modo, poder abastecerse en el futuro. Las plantas cortadas se estropean muy rápidamente, por lo que es mejor guardarlas en una caja o bolsa.

RAÍCES Y TUBÉRCULOS

Algunas plantas almacenan almidón en sus raíces para utilizarlo durante el invierno. En primavera se convierte en azúcar para alimentar los nuevos brotes. De esta manera, las raíces son más nutritivas en otoño, y más dulces en primavera. Tenga cuidado al recoger las raíces en primavera, ya que algunas plantas tienen entonces algunos brotes y pueden ser difíciles de identificar.

1 Corte una rama gruesa de madera dura como el nogal, y afile uno de los extremos hasta darle la forma de un cincel. La punta puede endurecerse al fuego.

2 Cave profundamente al lado de la planta. Desprenda la tierra de alrededor de la raíz hasta que pueda sacarla de una sola pieza.

CÓMO PREPARAR LAS RAÍCES

Debe preparar y cocinar las raíces tan pronto las haya recogido para evitar que se estropeen.
Las raíces deben limpiarse cuidadosamente, aunque no debe pelarlas ya que la piel contiene a menudo muchas vitaminas.
Corte las raíces en trozos grandes y hiérvalos en agua. Así eliminará las sustancias químicas amargas que pudieran contener.
Los trozos deben hervirse hasta que estén lo suficientemente blandos como para poder pincharlos con un tenedor.

CÓMO OBTENER HARINA

La mayoría de las semillas no pueden digerirse bien a menos que se muelan para obtener harina. Puede entonces añadirla a un estofado o mezclarla con agua para hacer una masa para preparar una pasta o pan húmedo (*damper bread* es su nombre original) *(véase página 117)*. También puede obtener harina de las nueces *(véase página 88)*, así como de raíces o cabezuelas florales.

1 Para obtener harina a partir de semillas, comience por doblar la espiga de la planta que las contiene hacia una bolsa. Sacuda la espiga para desprender las semillas, o arránquelas con la mano. Deje secar las semillas al sol.

2 Trille las semillas frotándolas entre sí para desprender las cascarillas. A continuación, aviente las semillas agitándolas en un recipiente. Como las cascarillas son más ligeras, podrá separarlas de las semillas soplando.

3 Muela las semillas entre una piedra grande con una depresión en el centro y otra menor plana. Mantenga un movimiento continuo y circular. Tenga cuidado que el viento no se lleve la harina obtenida.

CÓMO PREPARAR LAS ALGAS

El musgo de Islandia *(Chondrus)* puede comerse fresco, seco o cocinarlo para obtener una especie de pudín de gelatina. Es rico en yodo, minerales y vitaminas A y B. Se encuentra en las rocas bajas de la playa, en primavera y a principios del verano. También puede hervir y comer las hojas de las ovas *(Porphyria)*, de la lechuga marina *(Ulva)* y de los varecs *(Alaria)*.

1 Recoja las algas desprendiéndolas de las rocas con su cuchillo. Extiéndalas al sol para que se sequen. Al secarse, las fibras palidecerán y adquirirán una textura correosa.

2 Después de aclarar las algas para eliminar los granos de arena, córtelas en pedazos.

3 Añada las algas al agua o a la leche y déjelas hervir lentamente al fuego hasta que la mezcla comience a espesarse.

4 Saque las algas y endulce la mezcla con bayas o azúcar.

5 La mezcla cuaja conforme se va enfriando; sáquela del recipiente con una cuchara. Tenga cuidado de no comerla hasta que esté completamente fría, o se quemaría la boca.

CÓMO PREPARAR BEBIDAS

Las agujas de los árboles perennes como el pino *(Pinus)* y las píceas *(Picea)* son ricas en vitamina A y C, y pueden utilizarse para preparar té. Recoja únicamente agujas frescas y verdes no decoloradas. Puede preparar un sucedáneo del café a partir de hojas de achicoria *(Chicorium)* o de bellotas *(Quercus)*.

1 Para preparar té a partir de agujas perennes, utilice dos cucharadas de té de agujas por taza de agua. Májelas ligeramente con una piedra grande.

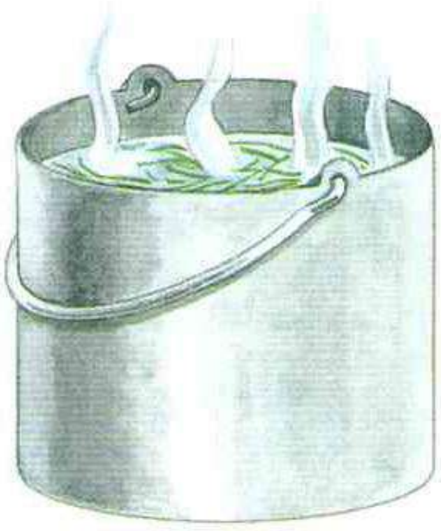

2 Sumerja las agujas en agua hirviendo. Mantenga el recipiente caliente sin permitir que el agua hierva, y deje remojar las agujas entre 5 y 10 minutos, removiendo de vez en cuando.

3 Separe las agujas del líquido filtrando el té a través de un paño sujeto sobre un recipiente. Puede endulzar el té con azúcar o miel.

OTRAS BEBIDAS

En Australia, las hojas del eucalipto *(Eucalyptus)* dan un peculiar sabor mentolado refrescante al té habitual. Esta infusión se conoce como *billy tea*. Las flores secas del trébol *(Trifolium)*, y las hojas de la ortiga *(Urtica)* y de la hiedra rastrera *(Glechoma)*, también pueden utilizarse para preparar té. La savia de los arces *(Acer)* puede beberse al mezclarla con agua.

CONDIMENTOS NATURALES

Las plantas silvestres tienen a menudo un sabor muy suave; quizá quiera realzarlo un poco con especias u otros condimentos. Puede llevar especias preparadas como curry en polvo, canela o clavo en su mochila para añadirlos a cualquier plato cuando sea necesario. Como alternativa, puede recoger especias en su estado silvestre y secarlas para utilizarlas más adelante. Un ejemplo de especia que puede encontrar en estado silvestre es la nuez moscada, la semilla de un fruto de una planta nativa de Indonesia Oriental *(Myristica)*. Una vez seca, la semilla puede rallarse o trocearse para dar sabor a los alimentos horneados, a cualquier pudín *(véase superior)*, o a otros platos preparados a partir de vegetales como la espinaca o la calabaza. También las hierbas como la albahaca o el tomillo pueden conferir sabores novedosos.

ANIMALES INVERTEBRADOS

LA IDEA DE comer una babosa, un caracol o un saltamontes no resulta demasiado apetecible, y nos hace pensar que preferiríamos pasar hambre. Sin embargo, en una situación de supervivencia es muy poco probable que una comida apetitosa sea fácil de obtener, y mientras espera se debilitará cada vez más hasta llegar al punto de no ser capaz de buscar fuentes alternativas de alimento. La experimentación es fundamental, pero para sobrevivir no debe rechazar ninguna fuente potencial de alimento, ni reparar en los escrúpulos que pueda sentir. Algunos invertebrados contienen incluso más proteínas que los vegetales, por lo que vale la pena no prescindir de ellos.

ANIMALES INVERTEBRADOS COMESTIBLES

Aunque muchos animales invertebrados son comestibles, no es tan sencillo meterse una babosa o un saltamontes en la boca, masticarlos y tragarlos. Únicamente algunas partes de estas criaturas son comestibles, y todas ellas deben cocinarse bien antes de consumirse. Algunas incluso deben prepararse de una manera especial *(véase página 96)*. A continuación se muestran algunos ejemplos.

Saltamontes. *Todos los insectos que saltan tienen grandes músculos en las patas que pueden comerse. Al asarlos destruirá los parásitos que puedan contener. Dan a la carne, según dicen los entendidos, un excelente sabor.*

Hormigas. *Las hormigas pueden picar o morder, pero son comestibles si se cogen con cuidado* (véase página 96). *Algunas hormigas colmeneras almacenan néctar y agua en sus abdómenes.*

Abejas. *Las abejas, sus crisálidas y larvas, son comestibles. Además, las abejas producen miel, una fuente instantánea de energía. Sin embargo, las picaduras de abeja pueden ser mortales, así que tenga cuidado al manipular estos insectos.*

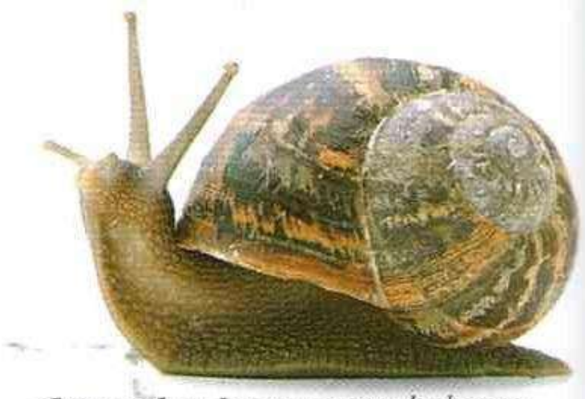

Caracoles. *Los gusanos, babosas y caracoles son una comida excelente, pero evite todos los caracoles marinos y cualquier caracol de tierra de colores brillantes, ya que podrían ser venenosos.*

Mariposas y polillas. *Las mariposas y polillas son comestibles, pero no aportan muchos nutrientes a menos que logre reunirlas en gran número. Las orugas son más fáciles de atrapar y constituyen una mejor comida.*

Estrellas de mar. *Pueden encontrarse adheridas a la madera a la deriva. Los cangrejos y las gambas también son comestibles y pueden encontrarse en las zonas rocosas cuando baja la marea.*

Erizos de mar. *Algunos equinodermos, como el erizo de mar* (Echinocyamus), *pueden desenterrarse de la arena de la playa para comerlos. Tenga cuidado al cogerlos, ya que algunos tienen largas púas.*

OTROS ANIMALES INVERTEBRADOS COMESTIBLES

Los insectos tienen un valor dietético superior al de los vegetales y son abundantes en cualquier región, aunque deben recolectarse en gran cantidad para obtener un aporte dietético adecuado. Las larvas de los insectos son particularmente jugosas. Para algunos pueblos indígenas constituyen un elemento primordial de su dieta. No olvide los insectos acuáticos, como los girinos *(Gyrinidae)*, aunque pueden resultar difíciles de atrapar a menos que se use una fina red de malla. Entre los invertebrados comestibles se incluyen también las termitas, la carcoma y el marisco. No es recomendable comer arañas, ya que muchas de ellas tienen vello o toxinas irritantes en la boca para matar a sus presas que, a su vez, pueden resultar dañinas para los humanos.

ANIMALES INVERTEBRADOS VENENOSOS

Debe aplicar el sentido común al recolectar animales invertebrados. Evite los insectos de colores brillantes, aquellos que comen carroña, los que parezcan enfermos o muertos, o los que tengan un olor fuerte, ya que podrían ser venenosos. Los animales invertebrados que comen otros animales pueden ser nocivos para los humanos. Tenga cuidado al recolectar animales invertebrados *(véase página 96)*.

Escarabajos rojos. *El color rojo o brillante de algunos insectos, como los escarabajos y gorgojos, advierte a sus depredadores que tienen un sabor desagradable, por lo que no debe cogerlos. Su toxina puede no ser mortal, pero sí puede provocarle una grave indisposición.*

Bonidos. *Hay cientos de tipos de caracoles marinos que disparan un aguijón que inyecta veneno. En algunas variedades resulta mortal. Estos caracoles se entierran en la arena, por lo que puede pisar uno accidentalmente.*

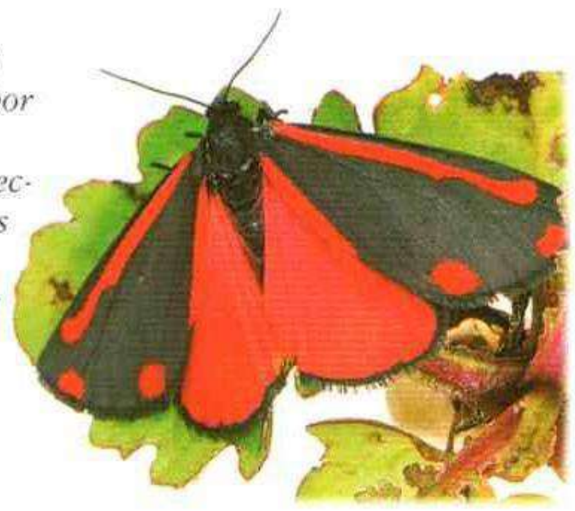

Polillas rojas. *La coloración roja de la polilla escarlata* (Callimorpha) *indica que tiene un sabor desagradable. Aunque algunas especies de insectos no tóxicos imitan los colores de las variedades venenosas, lo mejor es evitar aquellos que tengan colores brillantes, a menos que esté completamente seguro de que son inofensivos.*

ADVERTENCIA

Al igual que ocurre con la comida silvestre, es preferible aprender a identificar unos cuantos invertebrados comestibles. Si se encuentra en una situación de supervivencia, puede estar débil y ser más susceptible de lo habitual a las cantidades pequeñas de veneno.

OTROS ANIMALES INVERTEBRADOS VENENOSOS

Las orugas y los gusanos espinosos tienen sustancias irritantes en las púas que pueden ocasionar una erupción al tocarlos.
Los ciempiés, sobre todo las especies grandes, tropicales, tienen a menudo sustancias venenosas en la piel. Algunas especies de mariquita amarilla *(Coccinellidae)* también son venenosas. Aunque la mayoría de las especies de pulpo son comestibles, el pulpo de anillo azul de Australia *(Hapalochlaena)* tiene una picadura mortal, por lo que es preferible evitar cualquier especie a menos que pueda identificarla positivamente. También debe evitar las crisálidas enterradas, así como algunas especies de hormiga *(Formica)* que lanzan ácido fórmico a sus atacantes. Si no logra identificar algún invertebrado, no lo toque.

Hidrozoos marinos. *Algunos hidrozoos urticantes, como la hidra* (Hydra) *se encuentran en las corrientes cálidas. La fisalia* (Physalia) *puede tener tentáculos urticantes de hasta 12 m de longitud. Evite estos animales dentro de lo posible. También debe evitar las medusas, algunas de las cuales tienen una picadura venenosa* (véase página 180).

Avispas. *Son muy agresivas, y sus aguijones trabajan incluso después de muertas. Nunca intente atrapar avispas para comerlas. Si le pican en la cara, la hinchazón le impedirá respirar, lo que puede causarle la muerte.*

Crustáceos y moluscos. Las distintas especies de crustáceos y moluscos son en general comestibles una vez se han cocido, pero también son potencialmente peligrosos. Para alimentarse filtran grandes cantidades de agua y pueden retener abundantes sustancias tóxicas en sus cuerpos. Los moluscos deben cogerse vivos y comerse lo más pronto posible.

Anémonas marinas. *Las anémonas pueden encontrarse adheridas a las rocas descubiertas por la marea. La mayoría tienen tentáculos urticantes, por lo que debe evitarlas. La mayor parte de criaturas que matan a sus presas por medio de picaduras también pueden ser nocivas para los seres humanos: evítelas dentro de lo posible.*

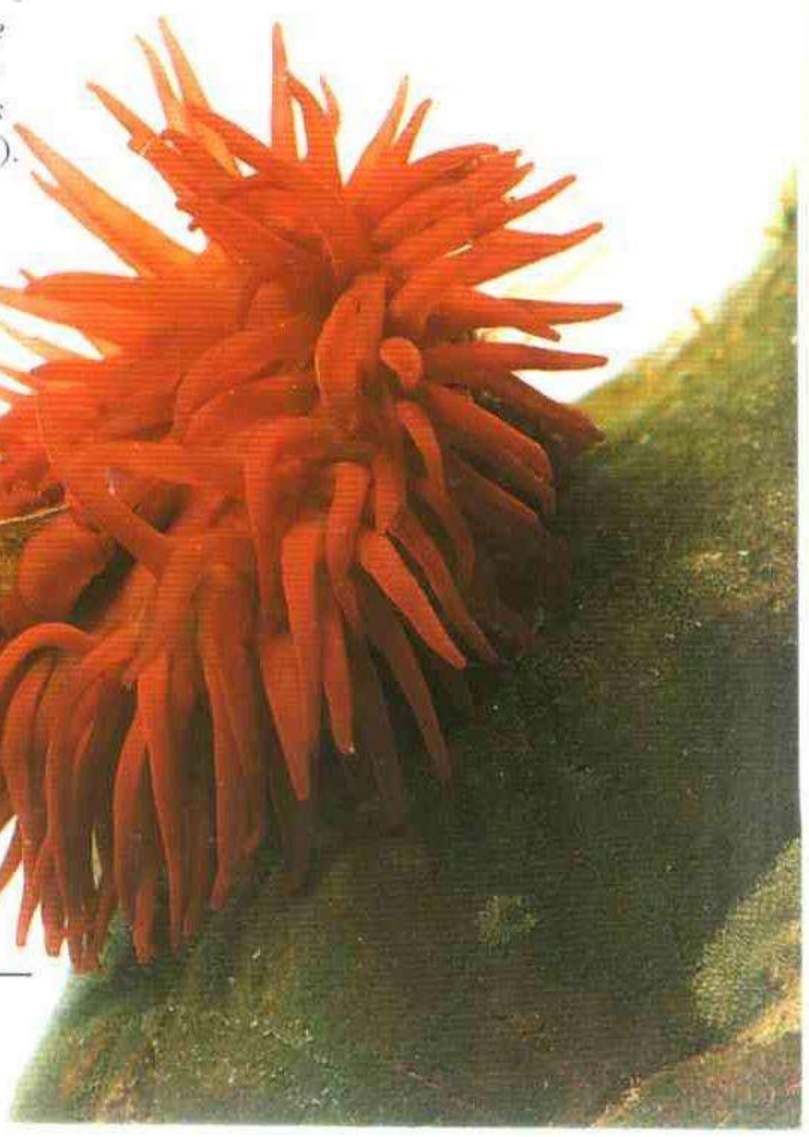

LOS ANIMALES INVERTEBRADOS COMO ALIMENTO

LOS ANIMALES INVERTEBRADOS son, con diferencia, los más abundantes del reino animal, y a menudo son las únicas criaturas que habitan en los medios y condiciones más extremos. Es más fácil atrapar animales invertebrados como los insectos que pescar, y en esa actividad se gasta menos energía. Sin embargo, los animales invertebrados parecen menos apetitosos que otras criaturas. Para superar este inconveniente, puede picarlos o molerlos, y después añadirlos a otros platos más sabrosos, antes que comerlos enteros.

INSECTOS VOLADORES

La mayoría de los insectos voladores son comestibles, pero las abejas productoras de miel proporcionan el mejor alimento. Tanto las abejas adultas como sus larvas son comestibles. Sin embargo, ya que las abejas pueden ser peligrosas, atrápelas únicamente si no le queda otra opción.

Abejas y miel. *Para atrapar abejas y recoger miel, acerque hierbas humeantes a la entrada de la colmena hasta que ésta se llene de humo. Selle el agujero y espere una noche. A la mañana siguiente, destápelo cuidadosamente con una rama y saque las abejas muertas y el panal.*

Voladores nocturnos. *De noche, extienda una tela de color claro entre dos ramas, ténsela y fíjela con estacas al suelo. Coloque una linterna detrás de la tela y un recipiente con agua delante de la misma. Las polillas se verán atraídas hacia la luz, chocarán con la tela y caerán al agua.*

CÓMO PREPARAR LOS INSECTOS

Elimine los aguijones de las abejas. Separe las alas y las patas de todos los insectos. Desde el punto de vista nutricional, los insectos deben comerse crudos, pero resultan más apetecibles hervidos o asados.

Molienda. *Los insectos pueden molerse entre dos piedras y añadirse a un estofado.*

INSECTOS REPTANTES

Las larvas de los escarabajos son muy nutritivas. Examine la madera podrida con ayuda de una rama para comprobar si hay alguna larva bajo la superficie. Las orugas se encuentran a menudo sobre las plantas que les sirven de alimento, y frecuentemente las encontrará en grandes colonias, pero recuerde que muchas tienen vellos urticantes en su piel, y pueden producir irritaciones al tocarlas.

1 Para atrapar termitas u hormigas, corte una rama larga y recta y elimine su corteza hasta dejarla lisa. Después introdúzcala lentamente en el nido de hormigas o termitas. Tenga a mano un recipiente para recoger los insectos sin que se le caigan al suelo.

2 Extraiga la rama lenta y suavemente. Las termitas habrán atacado la madera y aún estarán sujetas a la misma. Raspe la rama y hágalas caer al recipiente. Puede entonces prepararlas como se ha indicado anteriormente.

CÓMO ATRAPAR INSECTOS

Puede preparar una trampa para insectos. Cave un agujero con cabida suficiente para un recipiente, y coloque éste en su interior. Coloque una tabla de madera encima del agujero, sobre dos piedras, para dejar un pequeño espacio entre la madera y el suelo alrededor del agujero. Los insectos serán atraídos por la sombra que les proporciona la madera, y caerán en la trampa.

TRAMPA PARA INSECTOS

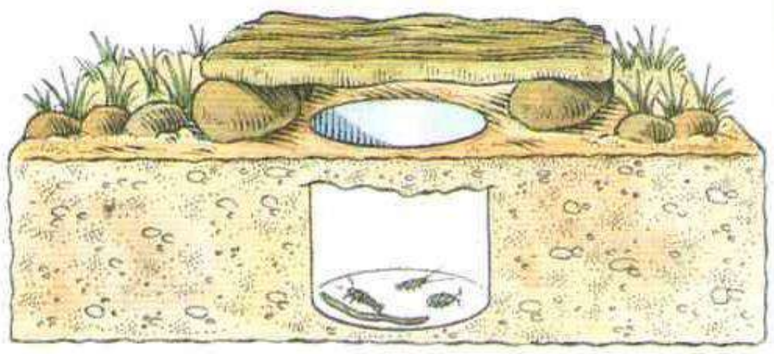

CRUSTÁCEOS Y MOLUSCOS

Los crustáceos y moluscos son fáciles de pescar en la mayoría de las playas, y pueden ser muy nutritivos. Sin embargo, debe asegurarse de que aún están vivos cuando los pesque, ya que su carne se deteriora rápidamente después de morir. Cocínelos siempre y consúmalos inmediatamente después de cogerlos. No los pesque cerca de cualquier posible fuente de contaminación.

En busca de navajas. *Las navajas se entierran verticalmente en la arena dejando agujeros para respirar en la parte superior de sus túneles. Puede atraparlas haciendo pasar un trozo de alambre o un palo a través de estos agujeros. Asimismo, puede espolvorear sal alrededor de la entrada del túnel y atraparlas cuando salen a la superficie.*

Bígaros y lapas. *Coja los moluscos pequeños, como las lapas adheridas a las rocas, cuando baje la marea. Tome únicamente aquellas que estén vivas —que estarán fuertemente adheridas.*

ADVERTENCIA

Nunca está de más advertir del cuidado que debe tener al manipular los crustáceos y moluscos. No debe comerlos nunca crudos, y deben hervirse o cocinarse al vapor por lo menos durante cinco minutos, y siempre inmediatamente después de pescarlos.

CÓMO COCINAR CRUSTÁCEOS Y MOLUSCOS

Hierva los crustáceos y moluscos en sus conchas por lo menos durante cinco minutos antes de comerlos. Deseche aquellos que no se abran en agua hirviendo, lo que significa que están en mal estado. Después de hervirlos, extraiga los bígaros y lapas de sus conchas con una ramita o un tenedor.

Estofado de marisco. *Hierva el marisco a fin de eliminar las sustancias nocivas.*

OTRAS DELICIAS INVERTEBRADAS

Los caracoles, las babosas y las lombrices son comestibles y pueden cogerse al amanecer. Mantenga los caracoles y las babosas vivos, con una dieta a base de hojas verdes durante 24 horas, antes de matarlos colocándolos en agua hirviendo. Puede hervir los caracoles, pero las babosas saben mejor asadas. Remoje las lombrices en agua salada durante 24 horas, y después extraiga sus partes blandas antes de añadirlas a un estofado o de secarlas.

Hormigas colmeneras. *Las hormigas colmeneras se alimentan de néctar y agua, y cada una de ellas se convierte en un auténtico almacén viviente que cuelga del nido.*

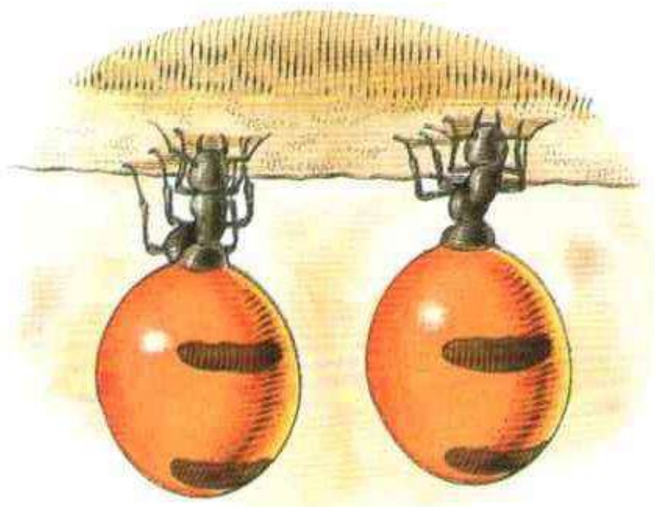

CÓMO PREPARAR UN CANGREJO

Cuando coja cangrejos, tenga cuidado con sus pinzas. Quizá se haga necesario atar las pinzas de las especies más grandes para evitar que se peleen y se hieran entre sí. Mátelos sumergiéndolos en agua hirviendo.

1 Un cangrejo debe prepararse con cuidado a fin de que todos sus órganos venenosos sean eliminados. Para preparar un cangrejo para comer, arranque primero las patas y las pinzas grandes. Todas ellas contienen carne comestible.

2 Para abrir el caparazón del cangrejo, inserte la punta del cuchillo entre las dos mitades y hágalo girar.

3 Elimine el estómago, cualquier sustancia verdosa y las dos branquias, ya que son venenosas. Separe la carne del caparazón. Rompa las pinzas y saque los músculos largos con el cuchillo.

PECES DE AGUA DULCE

LOS PECES DE agua dulce viven en ríos, corrientes y lagos, así como en los estuarios de los ríos más arriba de donde el agua dulce se sala y penetra en el mar. Algunas especies recorren los ríos en toda su extensión, mientras que algunos peces marinos, principalmente el salmón, migran hacia el interior para desovar en la relativa seguridad de las corrientes tierra adentro. Las condiciones y el tamaño de los peces de agua dulce varían de acuerdo a la cantidad y variedad de su alimento, y con las estaciones. Aunque los lagos y algunos ríos pueden tener tramos profundos lejos de la orilla, la mayoría de los peces de agua dulce viven en las aguas relativamente poco profundas de las orillas.

PECES COMESTIBLES

La mayoría de los peces que se encuentran en agua dulce pueden comerse, aunque algunas especies tienen un sabor más agradable, o son más fáciles de preparar que otras. Algunos peces tienen tal cantidad de pequeñas espinas que dificultan su preparación; estas espinas pueden ser peligrosas si se atraganta con ellas. A continuación se muestran algunos ejemplos de peces de agua dulce comestibles.

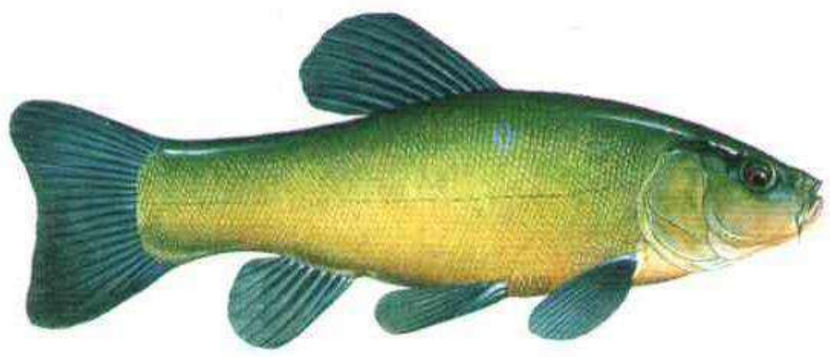

Tenca (Tinca). *La tenca vive en las aguas tranquilas de los tramos bajos de los ríos, antes de desembocar en el mar. Se encuentra en Europa, Norteamérica y Australia.*

Brema (Abramis, Pagellus). *La brema pertenece a la misma familia que los peces de colores de los acuarios. Se alimenta en los fondos de los ríos profundos y lentos en toda Europa y Asia. Algunos individuos alcanzan un tamaño considerable.*

Carpa (Cyprinus). *La carpa común crece hasta unos 60 cm de longitud y se alimenta en las aguas oscuras del fondo de los ríos, o en la superficie de las corrientes y ríos. Aunque la leyenda dice que puede vivir unos cuantos siglos, 40 años es una cifra más verosímil, así que no debe sentirse culpable si pesca un ejemplar grande.*

Trucha (Oncorhynchus). *La trucha se asemeja mucho al salmón* (véase página 100). *La trucha arco iris vive en las corrientes frescas y limpias, y en los lagos de Escandinavia, Norteamérica, Australasia, África e India.*

Escardinio (Scardinus). *El escardinio vive en aguas tranquilas con vegetación abundante, entre la que se oculta. Se encuentra en Norteamérica, Europa y Escandinavia. Se alimenta en la superficie, aunque algunos especímenes viejos y grandes a menudo descansan en el fondo.*

OTROS PECES COMESTIBLES

Existen muchas otras especies de agua dulce comestibles, en especial en los ríos y lagos tropicales. Las especies más comunes en todo el mundo incluyen el albur y el cacho *(Leuciscus)*, y el gobio *(Rutilus)*. La observación determinará qué especies se alimentan en el fondo o en la superficie, y la experiencia le dirá cuáles son las mejores para comer. Sin embargo, tenga cuidado con las especies desconocidas. Ante la duda, áselos a la parrilla, elimine la piel y las espinas, y después hierva la carne.

PECES PELIGROSOS

Los peces depredadores son más agresivos que sus presas, e incluso llegan a comerse a los miembros menores de su propia especie. Muchos de estos peces tienen aguijones afilados, que pueden estar conectados a glándulas venenosas. La mucosidad de su piel puede ser tóxica y causar erupciones en la piel. Incluso puede hacer que el pescado sea venenoso por ingestión a menos que se elimine la piel.

Lucio americano (Esox). *Éste es un miembro norteamericano de la familia de los lucios, y puede llegar a medir hasta 1 m de longitud. Prefiere las aguas tranquilas y claras con vegetación, y habita los ríos y corrientes que desembocan en el Atlántico. Tiene dientes afilados.*

Tandanus. *Un miembro de la familia de los siluros, habita en las aguas tranquilas de Australasia y de la región indopacífica. Otros siluros se encuentran en todo el mundo. La aleta dorsal o pectoral de los siluros puede tener una espina dura, en ocasiones en contacto con una glándula de veneno. Los siluros también están cubiertos por una mucosidad venenosa.*

OTROS PECES PELIGROSOS

Muchos peces comestibles de agua dulce pueden causar heridas con sus espinas o dientes afilados, o bien tener la piel revestida de una mucosidad venenosa. La conocida piraña *(Serrasalmus)* de la Sudamérica tropical es un ejemplo de pez peligroso y voraz, mientras que la anguila eléctrica *(Electrophorus)*, también originaria de Sudamérica, puede herir a una persona con una potente descarga eléctrica al nadar o vadear las aguas.

Siluro (Ictalurus). *El siluro tiene espinas venenosas y está cubierto por una mucosidad tóxica. Debe manipularse con cuidado y eliminar la piel antes de comerlo* (véase página 110). *El siluro come plantas e insectos, y se alimenta mayoritariamente en los fondos de las corrientes lentas y tranquilas, particularmente en Norteamérica y Europa. Es una especie de bagre, y tiene los bigotes característicos de esta familia.*

ADVERTENCIA

Una vez muerto, a menos que se conserve seco, ahumado o en escabeche (véase página 118), todo pescado se convierte en una fuente potencial de intoxicación debido a la proliferación de bacterias. A menos que le haya dado muerte usted mismo, es recomendable no comer ningún pescado muerto, independientemente de lo fresco que parezca.

Perca (Perca). *La perca es un devorador voraz que se desplaza en grandes cardúmenes que pueden llegar a mermar la población de otros peces de menor tamaño. Se encuentra en los ríos y corrientes tranquilas y lentas de Europa, Asia, Norteamérica y Australia. Tiene un buen sabor, pero su aleta dorsal es muy afilada.*

Lucio (Esox). *El lucio es un pez solitario y agresivo, con dientes muy afilados. En algunas áreas se le conoce como* lobo de río *debido a su apetito y poder de destrucción, ya que llega a pelear y comerse incluso peces de su mismo tamaño. A pesar de ello, tiene muy buen sabor. Se encuentra en Europa y Norteamérica, en la mayor parte de tipos de agua.*

Lucioperca (Stizostedion). *Tiene una gran similitud con la perca y el lucio, con quienes está emparentado. Se caracteriza por sus grandes ojos vidriosos, que brillan en la oscuridad reflejando la luz que incide sobre ellos. Se encuentra principalmente en Norteamérica y pertenece a la familia de las percas, por lo que tiene un sabor agradable. Sin embargo, tiene una peligrosa aleta dorsal; por lo tanto, tenga cuidado al manipularlo.*

PECES DE AGUA SALADA

COMO REGLA GENERAL, los grandes peces de agua salada viven en los mares profundos, mientras que los grandes cardúmenes de peces de menor tamaño viven en las aguas poco profundas cercanas a la costa. Los peces pueden llegar a migrar a grandes distancias en búsqueda de alimento, y procrear, asediados por multitud de predadores. Los peces grandes pueden llegar a acercarse lo suficiente a la costa para ser atrapados con una red o una caña, pero es más fácil pescarlos desde un barco mar adentro. Si carece de embarcación, puede vadear las zonas poco profundas de los mares o estuarios para pescar peces de agua salada, o bien adentrarse en los canales más profundos cuando baja la marea.

PECES COMESTIBLES

En realidad, existen muy pocos peces que no puedan comerse. La forma más adecuada y saludable de comer pescado es crudo, para conservar todos sus aceites naturales y su frágil valor nutricional. Para cocinarlo, prepárelo al vapor, a la brasa u horneado *(véase página 117)*. Sin embargo, deseche la piel porque puede estar revestida de una mucosidad tóxica, o tener espinas peligrosas.

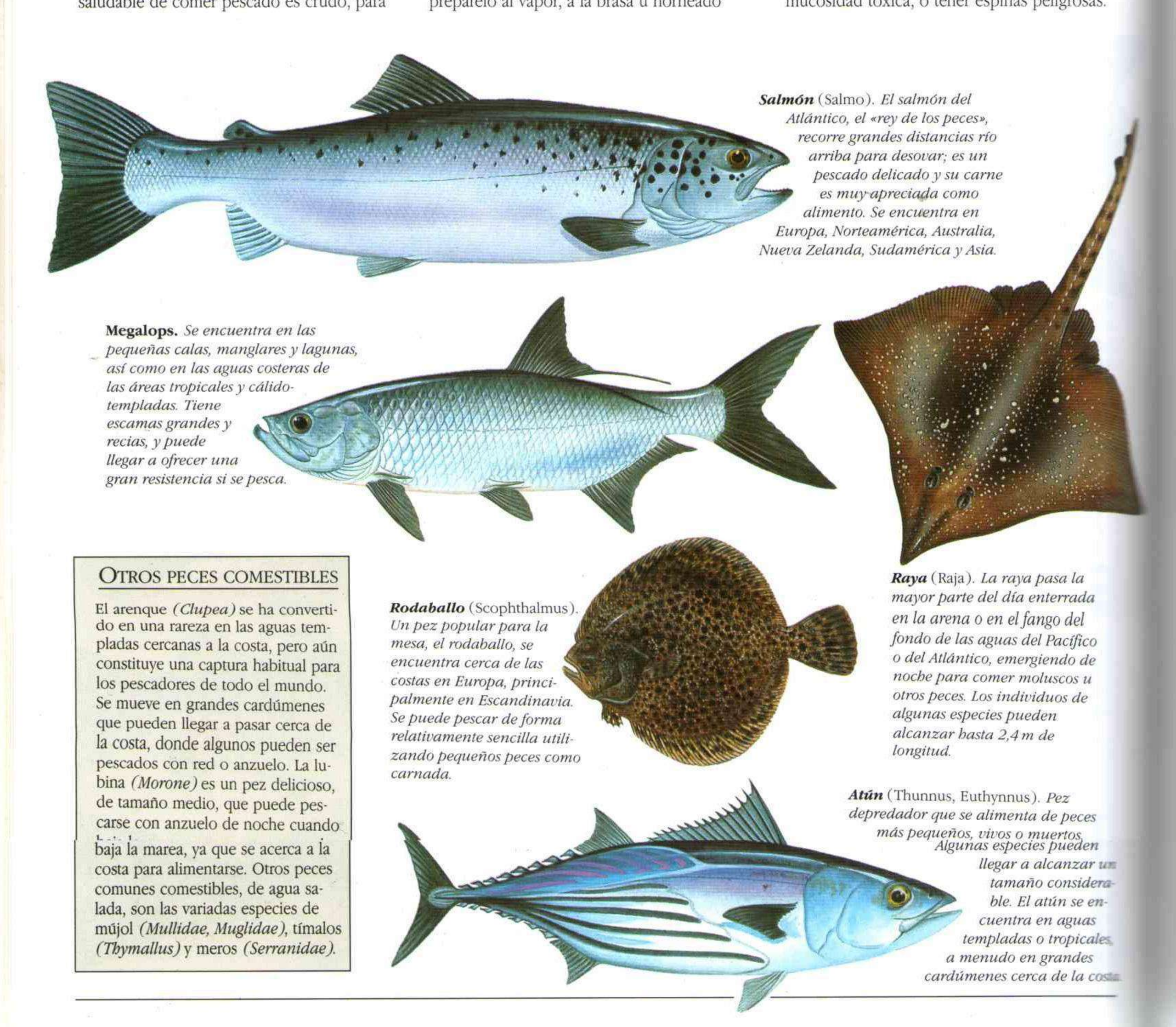

Salmón (Salmo). *El salmón del Atlántico, el «rey de los peces», recorre grandes distancias río arriba para desovar; es un pescado delicado y su carne es muy apreciada como alimento. Se encuentra en Europa, Norteamérica, Australia, Nueva Zelanda, Sudamérica y Asia.*

Megalops. *Se encuentra en las pequeñas calas, manglares y lagunas, así como en las aguas costeras de las áreas tropicales y cálido-templadas. Tiene escamas grandes y recias, y puede llegar a ofrecer una gran resistencia si se pesca.*

Raya (Raja). *La raya pasa la mayor parte del día enterrada en la arena o en el fango del fondo de las aguas del Pacífico o del Atlántico, emergiendo de noche para comer moluscos u otros peces. Los individuos de algunas especies pueden alcanzar hasta 2,4 m de longitud.*

Rodaballo (Scophthalmus). *Un pez popular para la mesa, el rodaballo, se encuentra cerca de las costas en Europa, principalmente en Escandinavia. Se puede pescar de forma relativamente sencilla utilizando pequeños peces como carnada.*

Atún (Thunnus, Euthynnus). *Pez depredador que se alimenta de peces más pequeños, vivos o muertos. Algunas especies pueden llegar a alcanzar un tamaño considerable. El atún se encuentra en aguas templadas o tropicales, a menudo en grandes cardúmenes cerca de la costa.*

OTROS PECES COMESTIBLES

El arenque *(Clupea)* se ha convertido en una rareza en las aguas templadas cercanas a la costa, pero aún constituye una captura habitual para los pescadores de todo el mundo. Se mueve en grandes cardúmenes que pueden llegar a pasar cerca de la costa, donde algunos pueden ser pescados con red o anzuelo. La lubina *(Morone)* es un pez delicioso, de tamaño medio, que puede pescarse con anzuelo de noche cuando baja la marea, ya que se acerca a la costa para alimentarse. Otros peces comunes comestibles, de agua salada, son las variadas especies de mújol *(Mullidae, Muglidae)*, tímalos *(Thymallus)* y meros *(Serranidae)*.

PECES PELIGROSOS

A pesar de que los peces de mayor tamaño tienden a habitar las aguas más profundas, las especies grandes y peligrosas también patrullan las aguas poco profundas en busca de comida fácil. Es bien sabido que los tiburones hambrientos llegan hasta las playas y atacan tanto a animales como a bañistas. Pueden detectar la sangre y el sonido de los peces o animales heridos incluso a grandes distancias. Algunos peces tienen espinas o púas que pueden causar un dolor muy intenso. Otros tienen dientes muy afilados.

Pastinaca (Urolophus, Dasyatis). *La pastinaca tiene aguijones venenosos y una cola que asemeja un látigo. Su picadura puede ser fatal si no se trata adecuadamente. Vive en el Pacífico Norte y en aguas tropicales.*

Acanthocybium. *Salta y sale del agua cuando ha picado el anzuelo, por lo que se hace difícil atraparlo. También tiene dientes muy afilados: puede ser peligroso incluso en tierra. Se encuentra en todo el mundo en las aguas cálidas y templadas.*

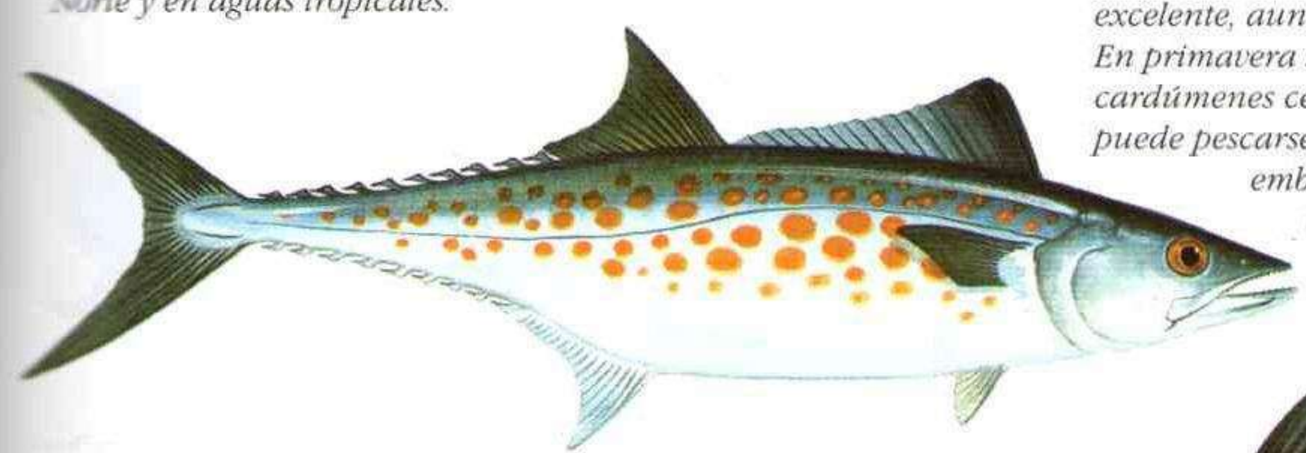

Caballa (Scomberomorus). *La caballa constituye un plato excelente, aunque un poco grasoso. En primavera se desplaza en grandes cardúmenes cerca de la costa, donde puede pescarse con redes. Sin embargo, es un nadador rápido, con dientes afilados, que ofrece gran resistencia s ser capturado. Se encuentra distribuida mundialmente.*

OTROS PECES PELIGROSOS

El pez ballesta *(Balistidae)*, que vive en los mares poco profundos y tropicales, tiene un único aguijón en el abdomen, a semejanza de un gatillo. Su carne es venenosa por ingestión. Tanto el pez piedra *(Synanceia)*, como el pejesapo *(Batrachoididae)* y la escórpora o pez cebra *(Scorpaenidae)*, tienen aguijones venenosos. El peje araña *(Trachinidae)* de las costas europeas y del occidente de África se entierra en la arena, y sus aguijones pueden causar un dolor muy intenso.

ADVERTENCIA

Cuando pesque con arpón, salga del agua tan pronto atrape su presa. El olor de la sangre y las vibraciones causadas por sus movimientos atraerán a los depredadores. Evite remover mucho el agua, ya que los depredadores interpretarán estas vibraciones como las de un pez herido.

Pez vela (Makaira, Tetrapturus). *El pez vela se encuentra en las aguas cálido-templadas y tropicales. Aunque de sabor agradable, es difícil de pescar ya que es un nadador rápido y ofrece una gran resistencia una vez pescado. También tiene la mandíbula superior fuerte y afilada. La forma más adecuada de pescarlo es con curricán, en la superficie o en aguas medias.*

Barracuda (Sphyraena). *La barracuda tiene dientes muy afilados y puede llegar a atacar a los bañistas. Las especies pequeñas, aunque muerden, son comestibles, y pueden encontrarse en aguas poco profundas. Asimismo, saltan una vez han picado, y pueden por ello ser difíciles de atrapar. La barracuda se encuentra en las aguas cálidas, templadas y tropicales de todo el mundo.*

Tiburón. *Hay cientos de especies de tiburones en todo el mundo, muchos de los cuales son agresivos y ofrecerán una gran resistencia a ser pescados. Algunas especies «hacen el muerto» una vez fuera del agua, y cogen a los pescadores desprevenidos. El tiburón leopardo* (Triakis) *vive en aguas poco profundas e interiores a lo largo de las costas del Pacífico de Norteamérica.*

EQUIPO DE PESCA

LA PESCA ES, en gran medida, cuestión de tiempo y paciencia. Sin un equipo prefabricado y resistente, los peces pueden picar y escaparse. Los peces pueden pescarse con anzuelo, trampas, redes, arpones, o incluso puede capturarlos con las manos *(véase página 104)*. Los anzuelos, cebos y corchos pueden construirse a partir de cualquier material tanto natural como artificial. Fabrique un sedal a partir de hierbas naturales *(véase página 34)*, o bien utilice el hilo de pescar de su equipo de supervivencia *(véase página 28)*. Puede improvisar una caña a partir de una rama. Tenga cuidado de no romper su equipo improvisado cuando los peces piquen, y asegúrese de tener preparado algún método alternativo de pesca (redes) para cerciorarse de que no se le escapa ninguno.

ANZUELO DE ESPINAS

1 Corte una sección de un tallo de zarza, de unos 2,5 cm de longitud, que tenga una espina grande y fuerte.

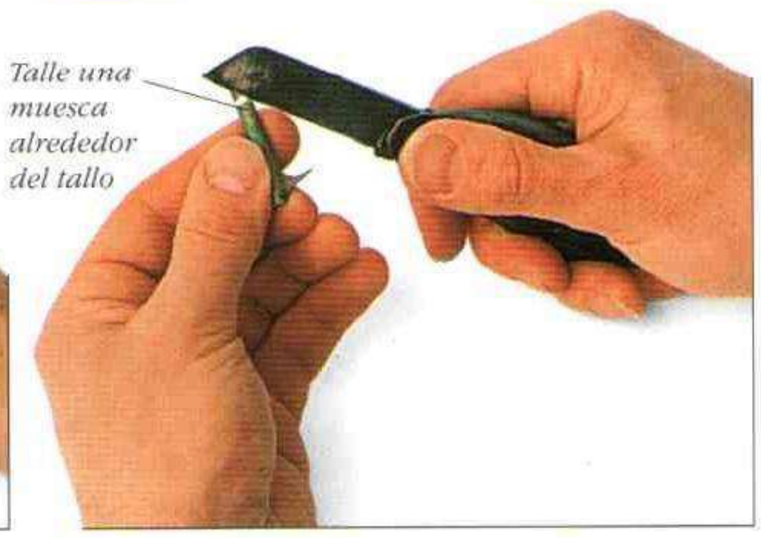

2 Con una navaja, talle una muesca en el extremo del tallo para evitar que el sedal resbale cuando lo ate alrededor del mismo.

3 Ate el sedal alrededor de la muesca en el tallo dando varias vueltas.

Anzuelo terminado. *Utilícelo para pescar peces que se traguen la carnada entera, como las anguilas y los bagres. El anzuelo también puede engancharse en las agallas o bocas de otros peces.*

UN CLAVO COMO ANZUELO Y UNA PLUMA COMO CEBO

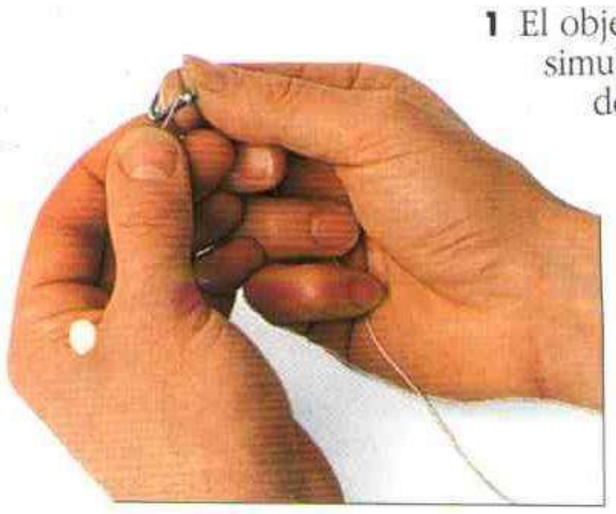

1 El objetivo de un cebo es simular la carnada, a fin de atraer al pez y para que muerda y se trague el anzuelo. Estudie lo que comen los peces en su hábitat y construya un cebo adecuado. Para construir un anzuelo simple, doble un clavo y átele el sedal.

2 Ate una pluma encima del anzuelo para que actúe como señuelo. La pluma, al simular un insecto, atraerá a los peces. Roce la superficie del río o corriente con el cebo y después sumerja repetidamente el anzuelo en el agua a fin de atraer a los peces para que muerdan el anzuelo.

Cebo terminado. *Asegúrese de que el anzuelo está bien sujeto al sedal para no perderlo cuando pique el pez.*

UN ESCARAMUJO COMO CORCHO Y UN IMPERDIBLE COMO ANZUELO

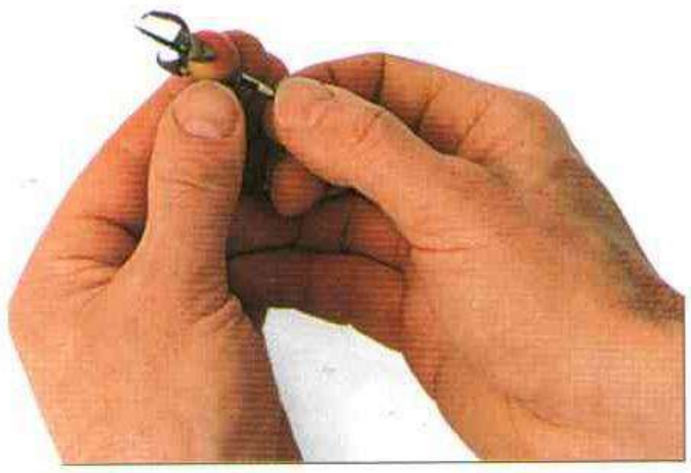

1 Puede emplear materiales naturales para construir un corcho. Si utiliza un escaramujo, agujeréelo con un cuchillo, un palo afilado, o una espina.

2 Pase el hilo de pescar o un cordel a través del escaramujo. La función del corcho consiste en mantener el anzuelo a la profundidad adecuada para el pez que se pretenda pescar.

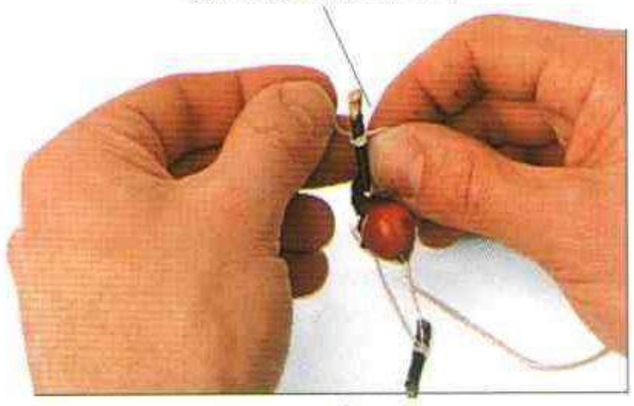

3 Ate ramas pequeñas por encima y por debajo del corcho para que actúen como estabilizadores y para evitar que el corcho se deslice a lo largo del sedal.

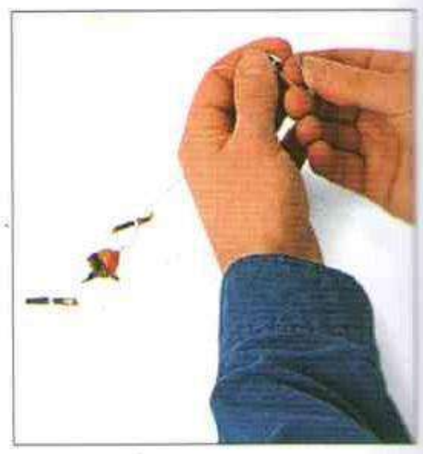

4 Ate el anzuelo (en este caso un imperdible) por debajo del corcho a la profundidad deseada.

CEBO

El cebo puede extenderse en la superficie del agua, sujetarse a un anzuelo o colocarse dentro de una trampa. Los insectos, los frutos pequeños, el pan, la carne cruda y los trozos de pescado son buenos cebos. La carnada viva atrae a los peces con sus movimientos, ya que indican que está herida y que es una comida fácil. El secreto de una carnada adecuada está en tentar al pez con su dieta habitual.

ARPÓN CON TRAMPA

1 Este tipo de arpón se emplea principalmente en aguas profundas *(véase página 104)*. En primer lugar, corte los vástagos laterales para obtener una rama recta y esbelta.

2 Afile el extremo más grueso de la rama hasta lograr una punta afilada y larga. Puede endurecer la punta pasándola cuidadosamente entre las llamas de un fuego. Talle los laterales de la punta para formar una superficie lisa a la que atara la trampa utilizando una cuerda.

3 Ate dos ramas resistentes a cada lado de la punta del arpón para que se cierren por encima y más allá de ese punto, a fin de dificultar la liberación del pez atrapado. Estas ramas se mantienen separadas por medio de un palo recto que se soltará al empujarlo, permitiendo que ambas ramas se cierren con una pinza.

ARPÓN CLAVETEADO

1 Haga muescas en uno de los extremos de una rama larga y recta para poder atar las puntas al arpón.

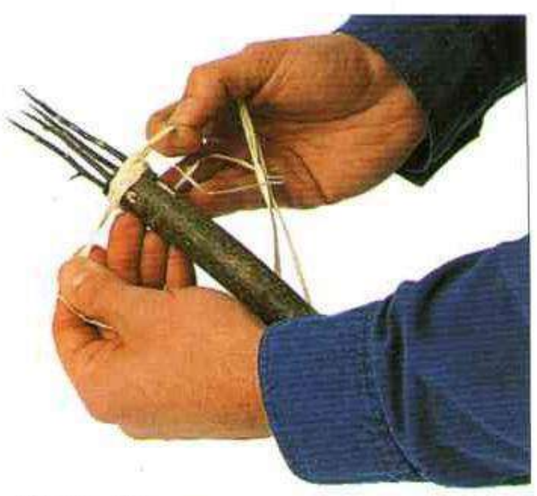

2 Ate firmemente pequeñas ramas espinosas alrededor del astil. Recuerde que tendrán que soportar parte del peso del pescado.

Arpón terminado. *Tenga cuidado de no romper las puntas contra las rocas o el fondo del río. Es conveniente utilizar un salabardo para sacar los peces del agua.*

TOME PRECAUCIONES AL PESCAR

Tenga mucho cuidado al utilizar equipos de pesca improvisados, ya que pueden romperse con facilidad y hacerle perder la presa, o el equilibrio. Si cae al agua, además del riesgo de ahogarse, puede contraer neumonía o alguna infección transmitida por el agua *(véase página 75)*. No entre en un río turbio sin una rama como soporte que a su vez le sirva para tantear el terreno. Recuerde que otros animales —entre ellos los depredadores— pueden encontrarse pescando o bebiendo en las inmediaciones

A la caza del pescador. *Busque las huellas de los animales que se dirigen a beber, y asegúrese de no convertirse en el cebo para un depredador.*

TÉCNICAS DE PESCA

PARA UN PESCADOR deportivo moderno, estas técnicas tradicionales pueden parecer muy rudimentarias. Sin embargo, el éxito del pescador y el suyo propio dependen, no del equipo, sino del conocimiento de la presa. Tómese el tiempo necesario para observar dónde van los peces durante las distintas horas del día, cuándo y dónde se alimentan, y qué comen. A continuación, diseñe la estrategia para pescarlos. Cuando utilice un equipo de pesca improvisado, tenga cuidado de no romperlo o perderlo, y piense siempre primero en su propia seguridad. Las técnicas descritas aquí pueden utilizarse en ríos y en corrientes de agua dulce o salada. Algunas de estas técnicas son ilegales en muchos lugares del mundo, y únicamente debe emplearlas si se encuentra en una situación de vida o muerte.

EL EMPLEO DE ANZUELOS Y SEDALES

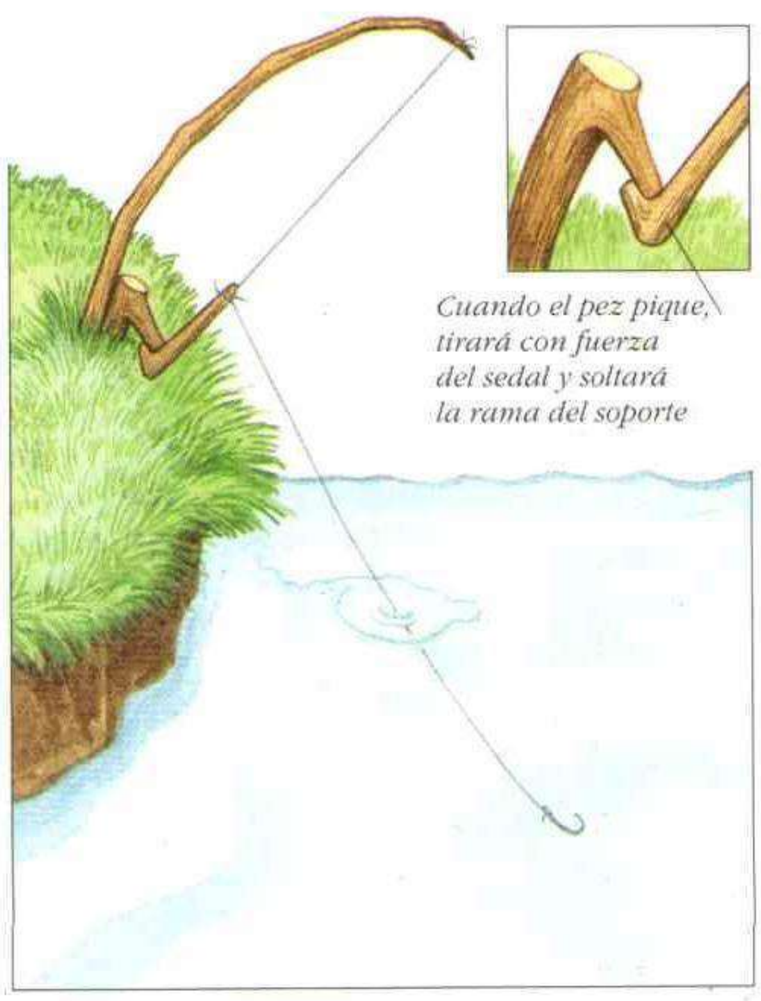

Sedal autónomo. *Un sedal de este tipo funcionará solo. Cuando el pez pique, el sedal se verá desplazado del soporte que lo sujeta y la caña volverá a extenderse; el anzuelo se clavará entonces en la boca del pez.*

Sedal múltiple. *Cuantos más anzuelos tenga en el agua, mayores serán sus posibilidades de pescar. Con este método, los peces picarán más rápidamente que con un solo anzuelo. Utilice este método únicamente en una situación de supervivencia, ya sea en agua dulce o en una ensenada de agua salada.*

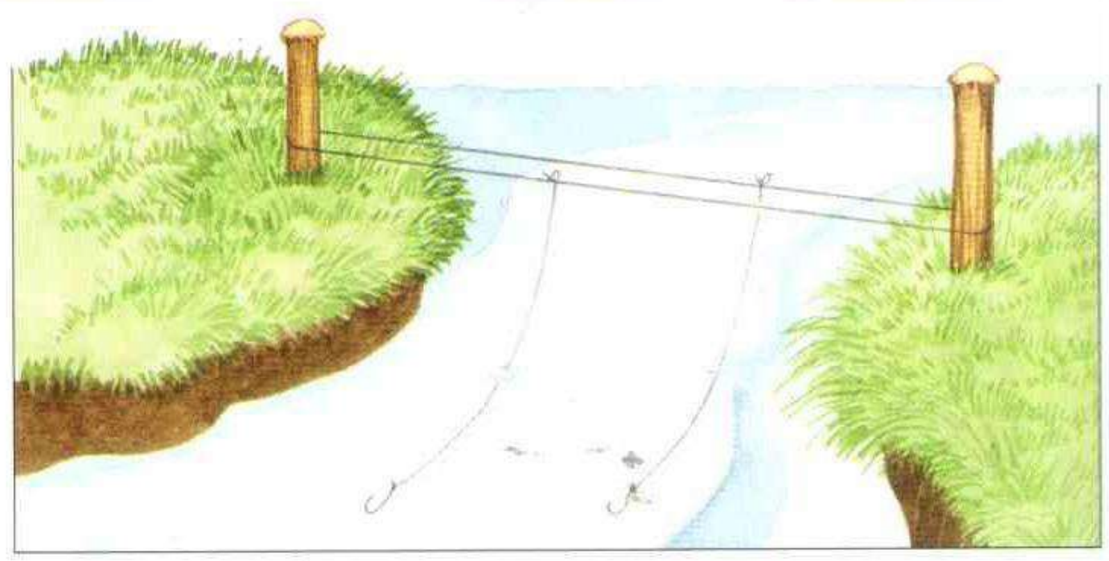

PESCA EN EL HIELO

Con una sierra o cuchillo para hielo, corte un pequeño agujero en el hielo, cerca de la orilla, a fin de comprobar que tenga como mínimo 5 cm de espesor y para poder soportar su peso. Corte un agujero de unos 30 cm de diámetro. Procure no romper el hielo porque podría ceder y provocar su inmediata caída al agua. Ate el sedal a dos palos en forma de cruz a los que previamente habrá atado una bandera.

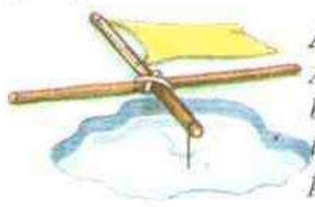

Anzuelo con cebo. *Ate el sedal a una bandera para evitar tener que sacarlo para comprobar si algún pez ha mordido el cebo.*

El pez ha picado. *Si la bandera está en posición vertical significa que algún pez ha picado.*

PESCA MANUAL

1 Este método es ilegal en la mayoría de los países. Es más adecuado para las aguas resguardadas. Tantee a lo largo de la orilla con las palmas de las manos extendidas hacia arriba.

2 Deslice lentamente sus manos por debajo de un pez que permanezca quieto. Coja el pez rápidamente y láncelo a tierra.

PESCA CON ARPÓN

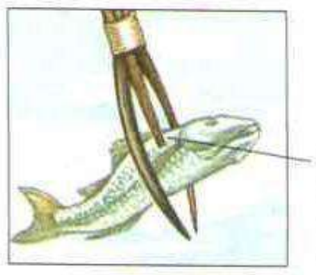

Arpón con trampa. *Este tipo de arpón atrapa al pez en su interior, evitando que se escape.*

El empleo del arpón. *La pesca con arpón requiere tiempo y mucha paciencia. Tenga cuidado de no romper sus arpones contra las rocas o el fondo del río. Si mantiene la punta del arpón en el agua, evitará las salpicaduras. Aseste el golpe rápidamente una vez haya divisado la posible presa.*

CÓMO SACAR LOS PECES DEL AGUA

1 Una vez el pez ha picado, tire con fuerza de la caña hacia arriba. Tenga cuidado de no romper su equipo, en especial al acercar el pez a la orilla.

2 Proceda lentamente a sacar el pez del agua. Utilice la red tan pronto le sea posible para liberar el peso de la caña; entre en el agua si es necesario. No se deje llevar ni por su entusiasmo ni por la corriente. Su seguridad debe prevalecer por encima de todo.

EL EMPLEO DE UN GARFIO

Garfio. *Como alternativa al empleo de una red, saque el pescado con un garfio. No levante la pieza en el aire.*

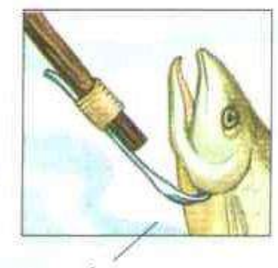

El garfio construido con un tenedor sujeta el pescado

DÓNDE BUSCAR LOS PECES

Un regalo escondido. *Encontrará peces en cualquier lugar donde puedan alimentarse sin sentirse inseguros; éste varía según la especie, el medio y la hora del día. Las aguas muertas (cuando un río se inunda), las orillas salientes, los troncos grandes y las rocas ofrecen buenos refugios, y en días cálidos y soleados, los peces buscarán el agua bajo la sombra o los lugares más profundos. Si el clima es frío, busque los peces en aguas poco profundas, por ejemplo en las orillas de un lago, en una zona calentada por el sol.*

Un árbol caído ofrece un buen refugio

Los árboles que sobresalen crean zonas de sombra donde los peces descansan en los días cálidos

En aguas rápidas o de inundación, compruebe los cantos rodados río abajo, o los recodos interiores, donde el río fluye más lentamente

Compruebe las orillas socavadas por el río. Hágalo en silencio sin molestar a los peces

Compruebe las zonas de vegetación flotante

Los peces pueden encontrarse descansando en aguas tranquilas, en la parte que da río abajo de las graveras, donde el agua rápida se arremolina para formar un pozo

Las burbujas u ondas en la superficie del agua indican que hay peces comiendo

REDES Y TRAMPAS DE PESCA

EN UNA SITUACIÓN de supervivencia, las redes y trampas de pesca son más efectivas que los anzuelos y los sedales sencillamente porque no requieren una atención continua. Las redes pueden fabricarse a partir de cordeles, cuerdas o fibras naturales *(véase página 34)*, utilizando ramas de árboles jóvenes como asas. También puede fabricar una red con ropa que no necesite. Las trampas pueden construirse a partir de ramas y cuerdas, pueden tejerse con hierbas resistentes, e incluso improvisarlas a partir de materiales fabricados por el hombre, como las botellas. El equipo improvisado únicamente debe usarse en caso de emergencia *(véase página 108)*.

CONSTRUCCIÓN DE UN SALABARDO

CON UNA CAMISETA

1 Pase las ramas de un árbol joven a través del dobladillo de una camiseta, haciendo dos pequeñas aberturas en la costura.

2 Cuando las ramas se encuentren, haga una nueva abertura en la camiseta en este punto y pase ambas ramas a través de ella. Ate ambos extremos entre sí utilizando una cuerda, y vuélvalos a introducir en el dobladillo.

3 Ate la camiseta por encima de las mangas y el cuello. Puede cortar la tela debajo del nudo, o bien invertir la red para que el exceso de tela no se arrastre en el agua cuando la utilice.

CON CUERDAS

1 Ate los extremos de las ramas de un árbol *joven para formar un círculo.*

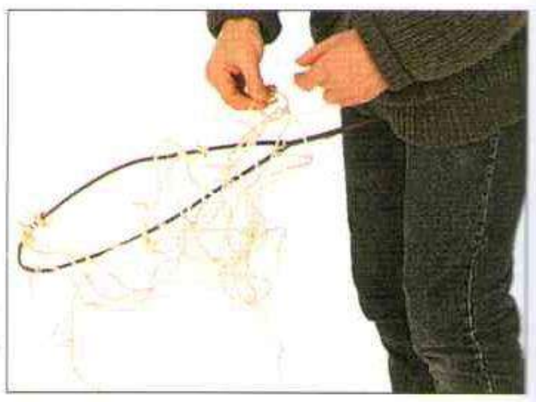

2 Ate uno de los cabos de la cuerda en el círculo. Ate el otro cabo en el lado opuesto del círculo formando aros.

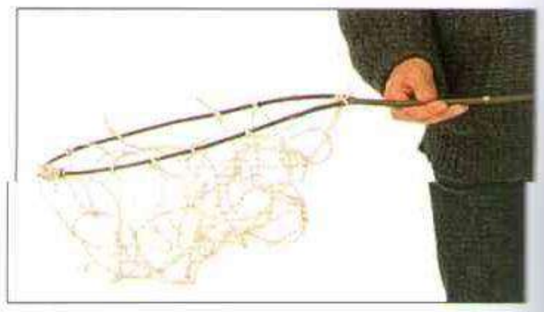

3 Entreteja más cuerdas entre los aros que cuelgan del círculo formado por las ramas, hasta formar una red.

CONFECCIÓN DE UNA RED VERTICAL (PARA PESCA POR ENMALLE)

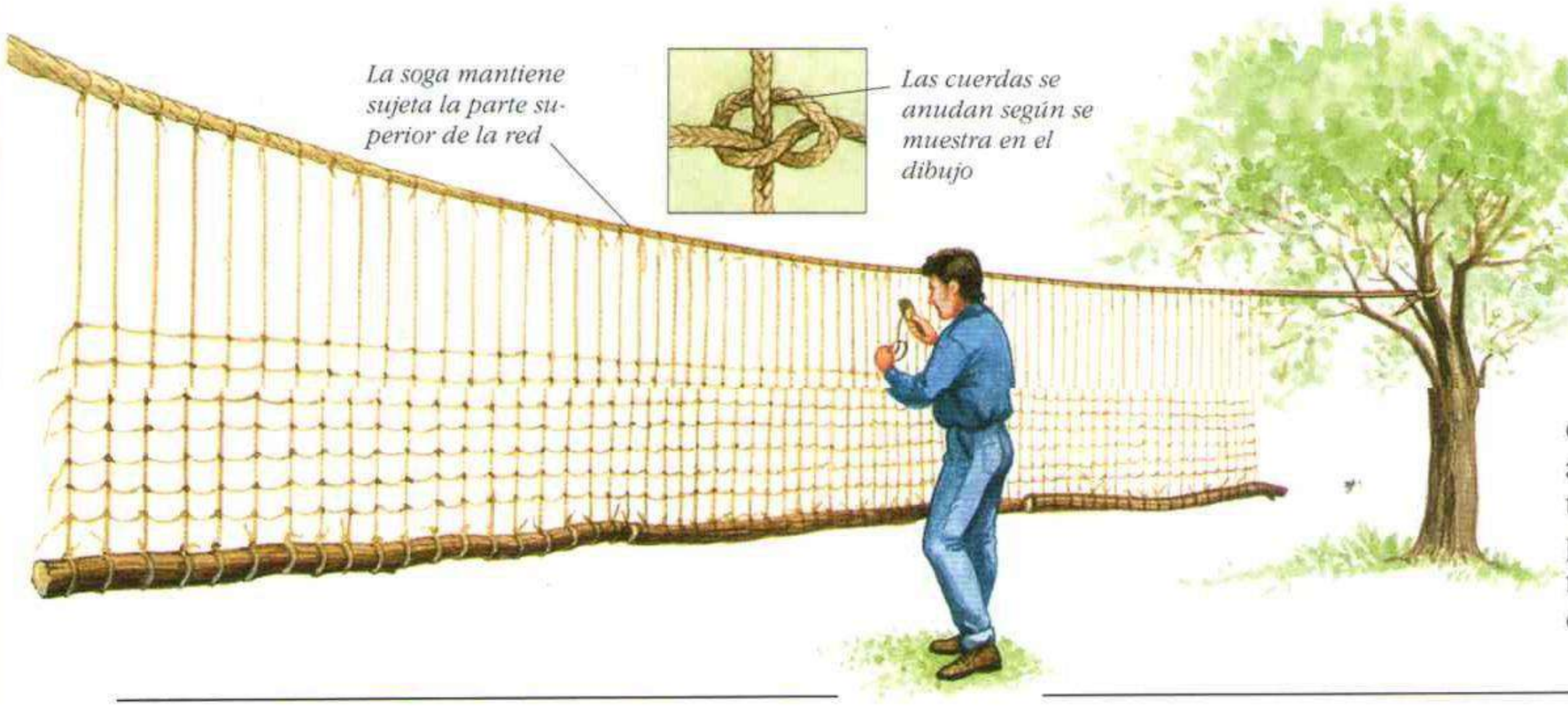

Las redes verticales (para pesca por enmalle) pueden ser de cualquier tamaño. Ate una soga entre dos árboles y anude los cabos de varios cordeles a intervalos regulares a lo largo de la misma. Mantenga estos cordeles tensos atando sus cabos inferiores a una rama. Entreteja más bramantes entre las cuerdas colgantes, haciendo un nudo alrededor de cada una para asegurar la malla. Puede ajustar el tamaño de malla al tipo de pez que pretenda pescar. Una vez terminada la red, córtela para liberarla de la soga y de la rama.

CONSTRUCCIÓN DE UNA NASA

1 La nasa puede utilizarse en un río o en una cala sujeta al efecto de la marea. Ate una gavilla de ramas o árboles jóvenes por sus extremos más gruesos.

2 Forme un aro atando los extremos de una rama joven.

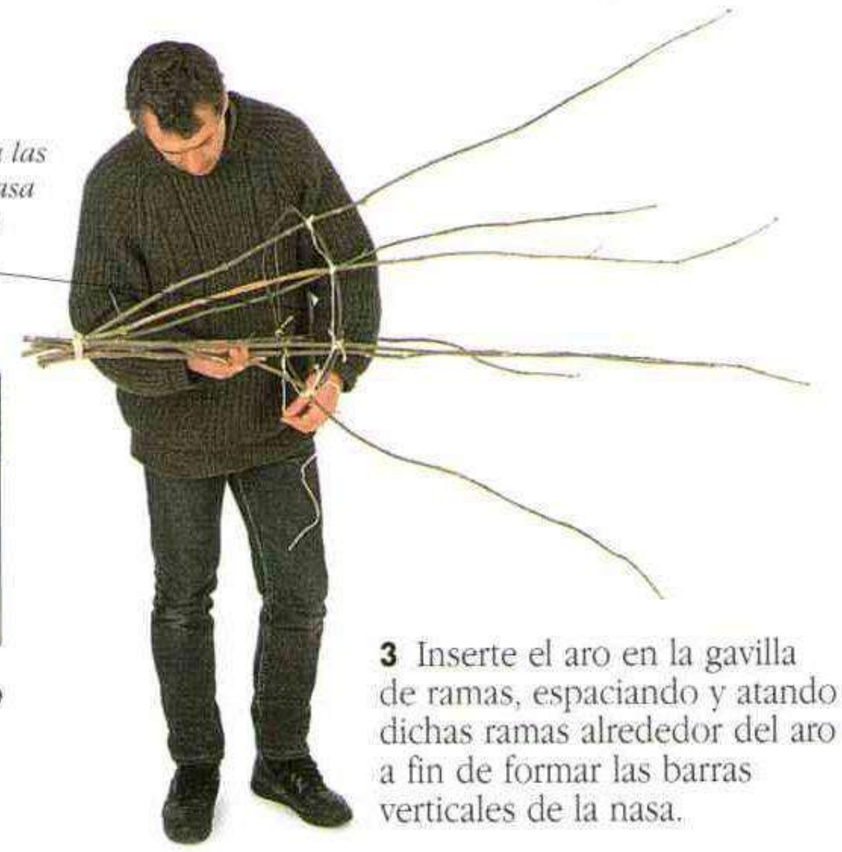

3 Inserte el aro en la gavilla de ramas, espaciando y atando dichas ramas alrededor del aro a fin de formar las barras verticales de la nasa.

4 Forme un segundo aro en el extremo opuesto para construir la cámara de la trampa. Puede hacer la trampa tan larga y ancha como desee.

5 Para construir la entrada de la trampa encaje dos aros, uno grande y uno pequeño, sobre un manojo de ramas en forma de embudo, con una salida ancha.

6 Invierta el embudo en el extremo abierto de la nasa y átelo en esta posición. Comience a entretejer cuerdas alrededor de las barras para construir los laterales de la nasa.

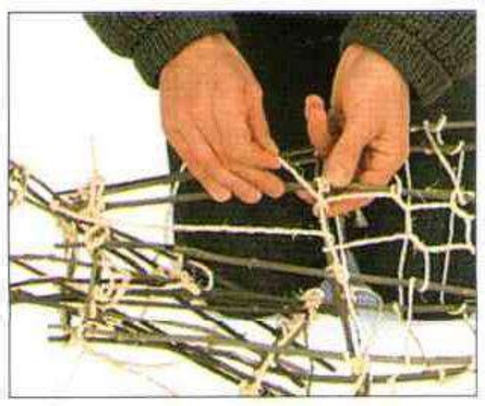

7 Rellene los laterales de la nasa con cordeles. Los peces entrarán en la trampa pero no podrán salir *(véase página 109)*.

OTRAS TRAMPAS

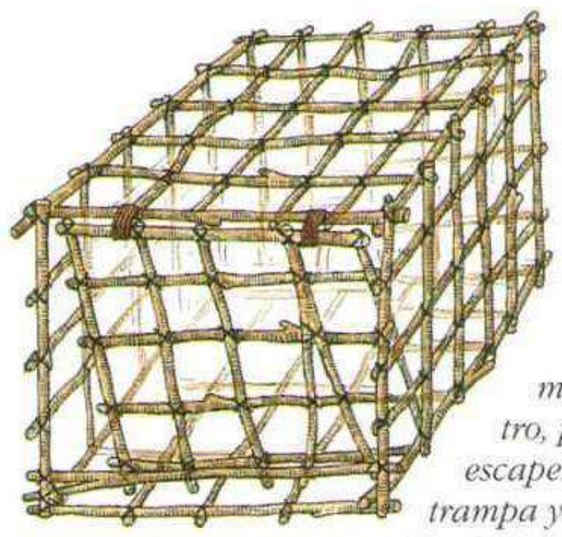

Trampa para cangrejos. *Esta trampa es muy sencilla de construir. Ate varias ramas entre sí para formar una caja. Ajuste el tamaño de la caja a la especie que pretenda capturar. La puerta de la trampa únicamente se abre hacia adentro, para evitar que la presa escape. Coloque un cebo en la trampa y lástrela con una roca para que permanezca en el fondo del río o del mar. Coloque la trampa en mares poco profundos para pescar cangrejos y langostas, o en aguas dulces para pescar cangrejos de río.*

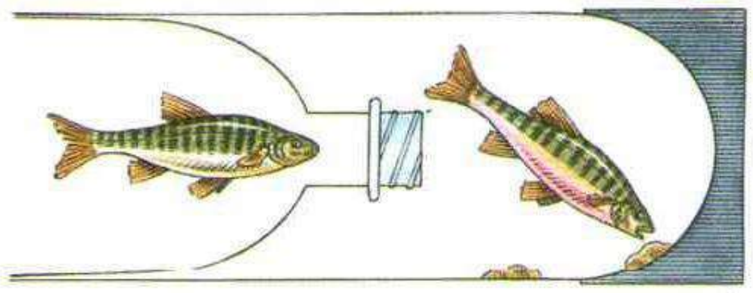

Trampa con botella. *Corte la parte superior de una botella de plástico, invértala e insértela en la base. Los peces entrarán en la botella para comer el cebo colocado en el fondo, pero no podrán encontrar la salida. Coloque la trampa en la corriente y compruébela a intervalos regulares.*

EL EMPLEO DE REDES Y TRAMPAS DE PESCA

PUEDE DEJAR UN sedal con anzuelo durante toda una noche, pero a menos que lo compruebe al rayar el alba del día siguiente, cualquier pez que haya picado habrá sido engullido por un pez mayor, o se habrá soltado. A diferencia de los anzuelos, las trampas pueden dejarse puestas durante varios días. Las redes no deben utilizarse de la misma manera: deben colocarse durante un período corto de tiempo, o bien utilizarse como redes de arrastre. Los peces de cualquier tamaño pueden quedarse atrapados en una red vertical, y si la deja extendida a lo ancho de un río, puede matar o herir a cualquier especie. Por esta razón, las redes verticales son ilegales en muchas partes del mundo. Estas redes únicamente deben emplearse en una situación de supervivencia en caso de emergencia.

EL EMPLEO DE REDES VERTICALES

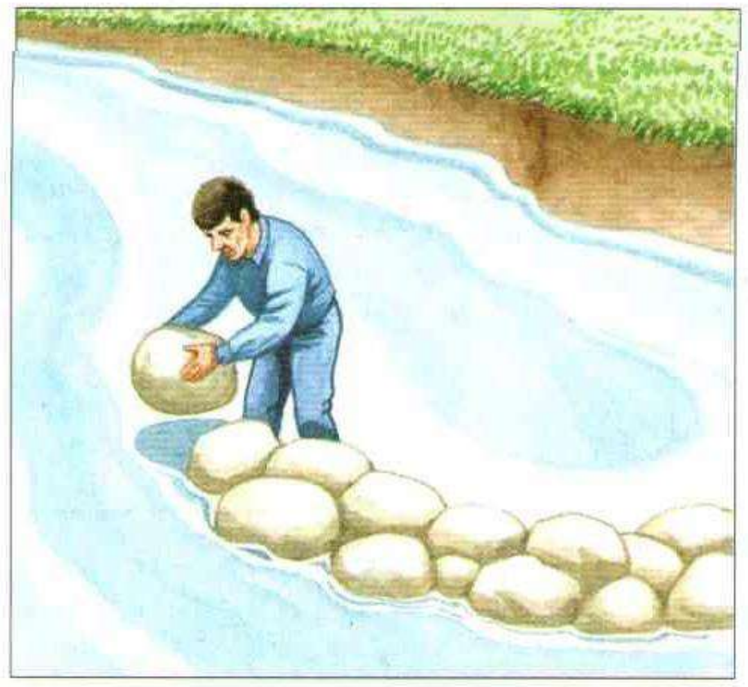

1 En una situación de supervivencia puede que tenga que construir un dique en un río a fin de pescar la suficiente cantidad de peces como para preparar conservas para el invierno. Construya una pared con rocas en el centro de una corriente, escogiendo un recodo, donde el agua fluye más lentamente.

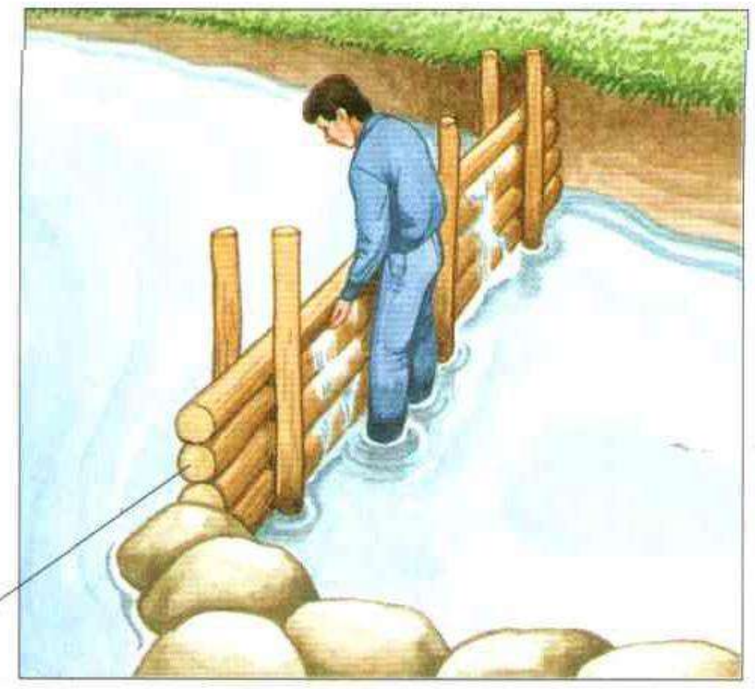

2 Clave tres postes sólidos en el lecho del río separados unos 30 cm entre sí. Colóquelos en diagonal para que el agua no derribe la pared de troncos una vez terminada. Coloque un par de troncos contra los postes en el lado superior de la corriente. Clave otros tres postes como retención. Después añada más troncos para construir la pared.

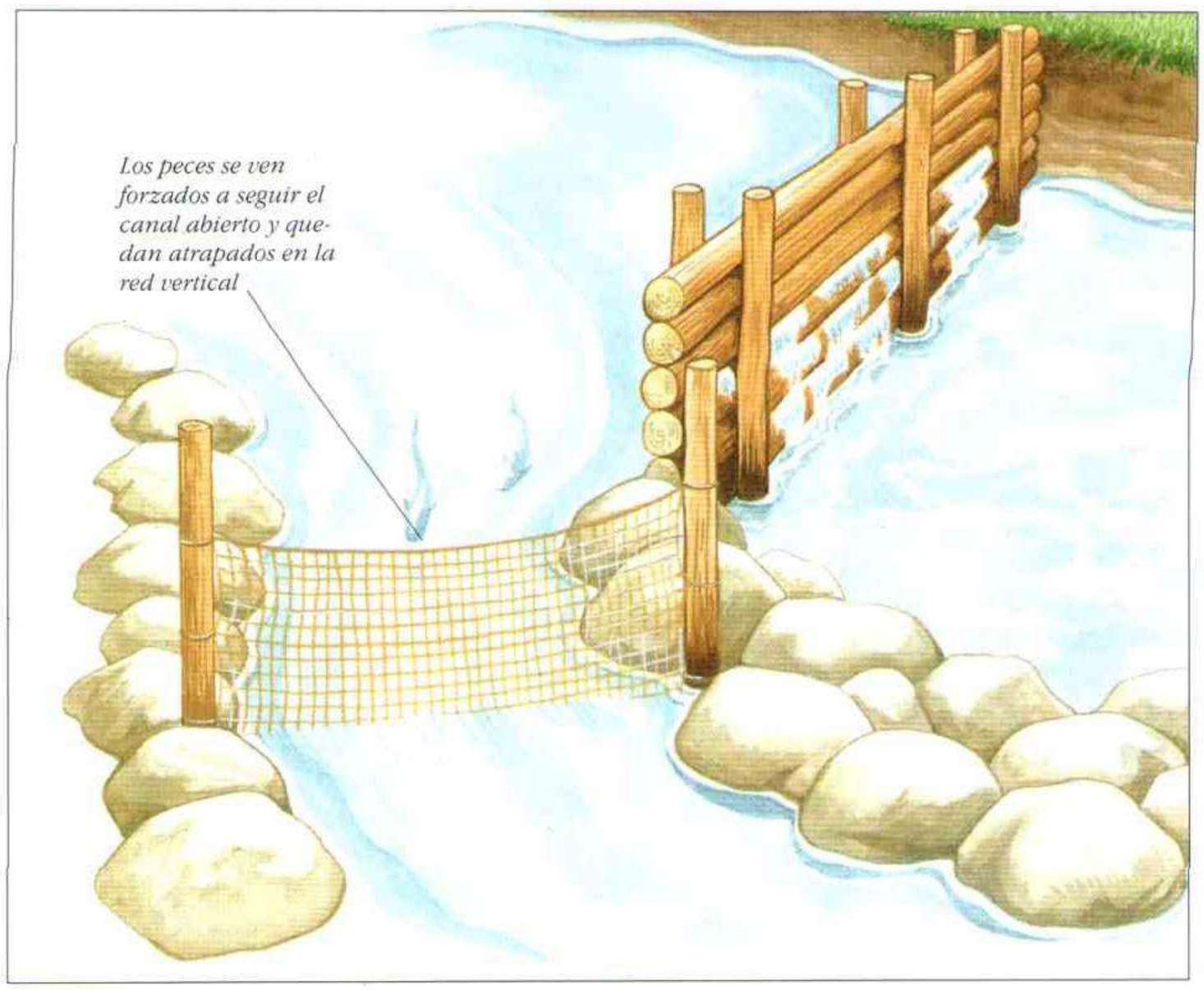

3 Clave dos postes a cada lado del canal cerca de la pared y ate a ellos su red vertical *(véase página 106)*. La presa forzará a los peces a seguir por el canal y quedarán atrapados en la red. Cuando dé por finalizada la pesca, retire la red y los postes horizontales de la pared para no atrapar o herir peces sin necesidad.

MANGA DE PESCAR

Una persona toma el extremo de la red, lo coloca sobre sus hombros y se adentra en el agua, mientras la otra permanece en la orilla. La primera persona camina hasta estirar la red y después describe una curva volviendo hacia la playa. Al llegar a la orilla, cualquier pez dentro de los límites de la red quedará atrapado. La red se lleva entonces a la playa y se recogen los peces.

Pesca en la playa. *Dos personas pueden pescar cardúmenes enteros de peces en las aguas poco profundas utilizando una red grande.*

TRAMPAS CON PAREDES DE ROCAS

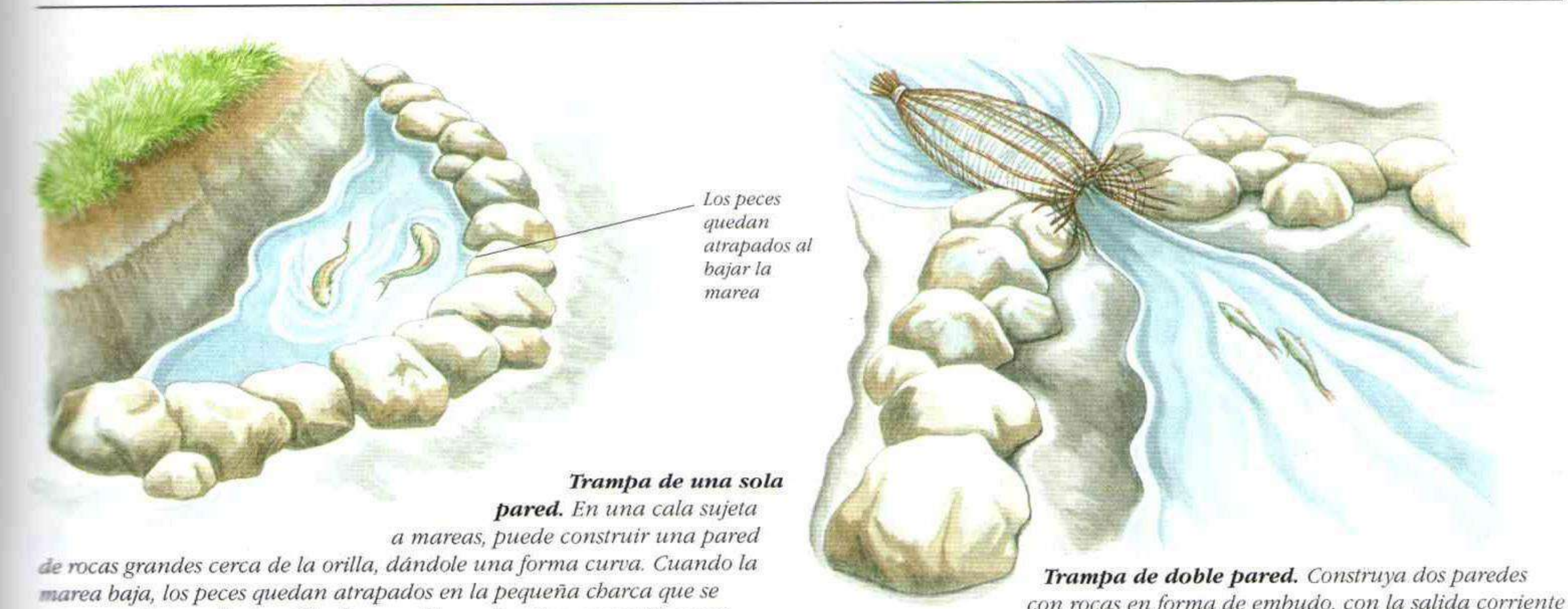

Trampa de una sola pared. *En una cala sujeta a mareas, puede construir una pared de rocas grandes cerca de la orilla, dándole una forma curva. Cuando la marea baja, los peces quedan atrapados en la pequeña charca que se crea entre la pared y la orilla. En una playa, si no hay una orilla o un arrecife, puede seguir el mismo principio, pero debe construir un círculo completo. Cualquier hueco entre las rocas grandes puede taparse con piedras pequeñas para evitar que los peces se escapen.*

Trampa de doble pared. *Construya dos paredes con rocas en forma de embudo, con la salida corriente abajo. Los peces que naden corriente abajo se dejarán guiar por las paredes hacia el embudo. Coloque su trampa o red en la angostura que forman las dos paredes* (véase página 106).

EL EMPLEO DEL SALABARDO

1 A los peces les desagrada el agua lodosa, ya que dificulta la respiración al disminuir el nivel de oxígeno. Remueva el lodo del fondo de un lago con sus pies. Los peces se acercarán a la superficie para respirar.

2 Utilice la red lenta y cuidadosamente, pero una vez que el pez esté dentro, levántela rápidamente del agua, y cierre la parte superior de la red, como si de una bolsa se tratara, para evitar que el pez salte fuera.

EL EMPLEO DE REDES CON CEBO

Algunos peces pueden pescarse con una red con cebo. Utilice una red fabricada con una camiseta *(véase página 106)*, o confeccione una con un calcetín. Pase una rama bífida a través del dobladillo del calcetín, que le servirá de asa. Coloque excrementos o despojos de pescado como cebo, y muévala en el agua.

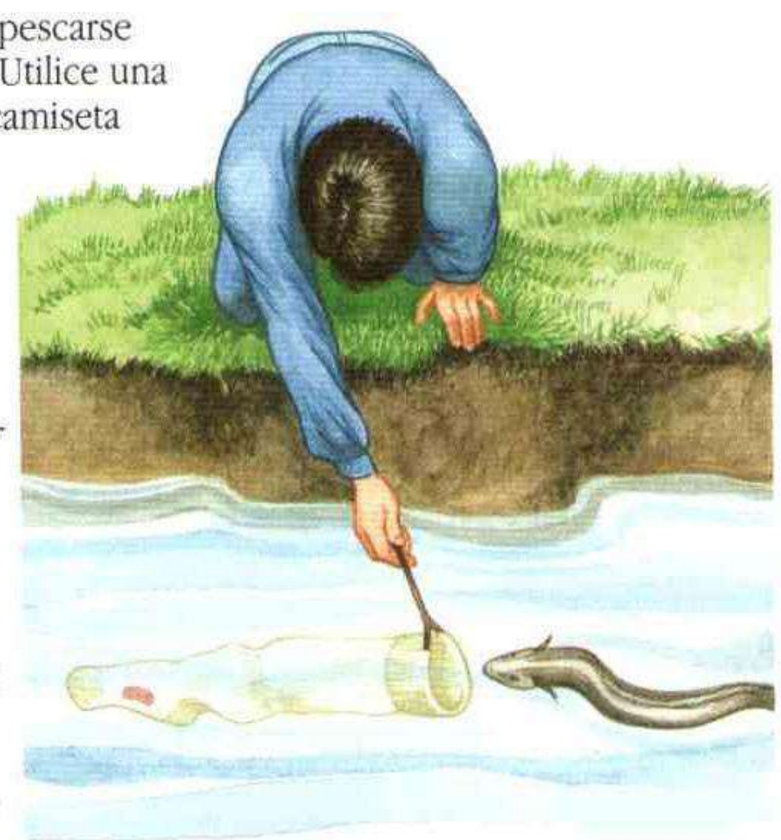

Cebo para anguilas. *Las anguilas se sienten atraídas hacia una red con despojos o excrementos animales.*

PUNTOS QUE DEBE TENER EN CUENTA AL PESCAR

- El agua siempre es peligrosa, fluya de forma rápida o lenta, sea poco o muy profunda. Incluso una caída sobre las rocas húmedas puede ser fatal.
- Considere concienzudamente todos los aspectos sobre su técnica de pesca, intentando imaginar los peligros antes de que ocurra un accidente; tome siempre las precauciones necesarias.
- Los peces tienen tendencia a volver al agua. Piense cómo llevará los peces a tierra una vez los ha pescado.
- Los peligros submarinos varían desde las algas en las que puede enredarse, o las criaturas peligrosas como los cocodrilos, hasta los pozos profundos y las fuertes corrientes. No entre en un río o corriente si puede evitarlo, aun si puede ver el fondo.
- Nunca salte al agua, y menos de cabeza. El reconocimiento de la zona resulta vital antes de tomar un baño con fines recreativos.
- Tenga mucho cuidado al utilizar equipos improvisados de pesca como el arpón, ya que podría romperse y hacerle caer al agua.

PREPARAR Y COCINAR EL PESCADO

INMEDIATAMENTE DESPUÉS DE llevar un pez a tierra, mátelo asestándole un golpe seco en la parte posterior de la cabeza, utilizando un palo o una piedra. Debe limpiar, cocinar y comer el pescado lo más pronto posible, o bien conservarlo para su posterior consumo *(véase página 118)*. En los climas cálidos, basta una corta exposición al sol para que el pescado se estropee, ya que su piel húmeda es un excelente foco de cultivo para las bacterias y las moscas. Una vez descamado, desollado, cortado y limpio es menos probable que se descomponga. En los climas fríos, puede esperar entre seis y doce horas antes de limpiar y cortar el pescado, lo cual permitirá que todas las terminales nerviosas mueran, y que el proceso sea mucho más sencillo que con una pesca reciente. Las espinas, piel y cabeza del pescado pueden ser hervidas en agua para hacer un caldo —siempre y cuando lo haga inmediatamente. Debe darse al caldo el mismo tratamiento que al pescado sin conservar: manténgalo fresco y consúmalo tan pronto como sea posible.

CÓMO LIMPIAR EL PESCADO

La limpieza elimina las partes del pescado con mayor tendencia a la descomposición, dejando la mayor parte de carne posible. Además, el pescado puede alojar en sus órganos internos sustancias contaminantes. Por ejemplo, los metales pesados como el mercurio, presentes en el mar en pequeñas cantidades, se concentran en el hígado y pueden resultar mortales para los hombres si se consumen en grandes cantidades. Las sustancias peligrosas no tienden a acumularse, por lo general, en los músculos del pescado.

1 La mayoría de los pescados no necesitan ser desollados antes de cocinarse. Sin embargo, deben eliminarse las escamas. Raspe las escamas desde la cola hacia la cabeza con el cuchillo, deslizando el cuchillo en dirección opuesta a su propio cuerpo.

2 Inserte la punta de un cuchillo afilado en el orificio anal. Tenga cuidado de cortar únicamente la parte ventral externa, abriendo el pescado justo hasta las agallas. No corte los músculos de la cola ni perfore los órganos internos.

3 Saque los órganos internos con cuidado. Extienda el pescado para comprobar que ha eliminado todos sus órganos. Guarde las huevas (huevos en las hembras, o lechaza en los machos) para comer, y después lave el pescado por dentro y por fuera.

4 Corte la cabeza, la cola y las aletas. Abra el pescado y separe las espinas de la carne con la punta del cuchillo. Puede ser más fácil eliminar las espinas una vez el pescado esté cocido.

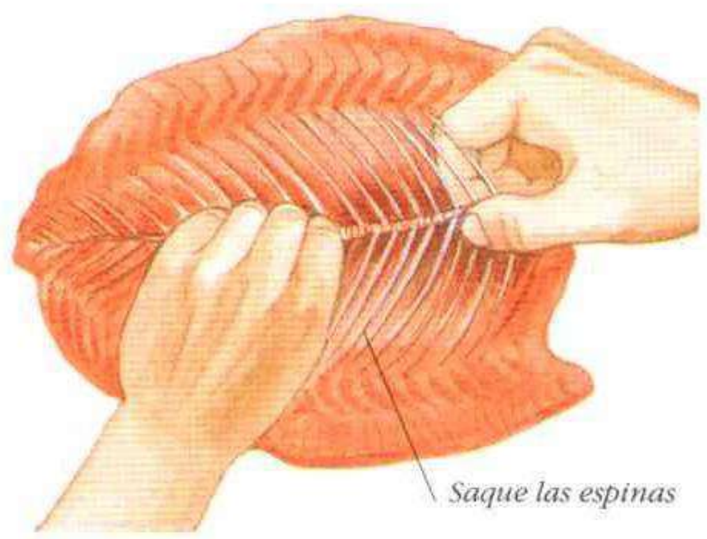

5 Separe la cabeza de la espina de la carne con la punta de su cuchillo. Después, despacio pero con firmeza, separe las espinas de la carne en una sola pieza. Si las espinas no salen bien limpias, tendrá que limpiar la carne de ellas con el cuchillo.

CÓMO DESOLLAR EL PESCADO

No todas las especies de pescado necesitan ser desolladas antes de comerse, en particular las más pequeñas. Sin embargo, algunos peces segregan en su piel mucosidades urticantes o venenosas. Otros, como los bagres o los tiburones, tienen pieles muy duras. Es pues recomendable eliminar y desechar la piel antes de comer el pescado.

Cómo desollar un filete. *Separe la piel de la carne. Sujétela con una mano y separe la carne de la piel con el cuchillo con un movimiento de vaivén.*

CÓMO ESPETAR EL PESCADO PARA COCINARLO

Es recomendable comer el pescado inmediatamente después de ser pescado. Los métodos más sencillos y rápidos de cocción son los que proporcionan el pescado con mejor sabor. El pescado no requiere mucha cocción. Áselo sobre un fuego de ascuas o empálelo en un espetón *(véase página 116)*. Los peces grandes pueden freírse en una sartén, o cocerse al vapor o en un horno de barro *(véase página 117)*.

1 Abra el pescado limpio, y ensarte dos palos delgados como soporte a lo largo de ambos extremos. La madera verde no se quemará con tanta facilidad. No utilice madera con gran contenido en savia porque podría chisporrotear.

2 Parta una rama joven longitudinalmente hasta la mitad. Inserte el pescado limpio en la hendidura y extiéndalo, separando ligeramente la parte central del filete para que se cueza uniformemente.

3 Ate la rama dividida para evitar que el calor del fuego la separe del todo. Clave la rama en ángulo con el suelo para que se incline sobre las brasas de un fuego. El pescado estará cocido cuando la carne parezca separarse.

CÓMO COCINAR PESCADOS PEQUEÑOS

Fría o ase los pescados pequeños enteros sin limpiarlos –no es necesario limpiar cualquier pieza de menos de 15 cm de longitud. Una vez limpios, clávelos en un palo verde y áselos sobre las ascuas o en una sartén. No elimine las cabezas y colas, ya que ayudan a conservar el pescado entero.

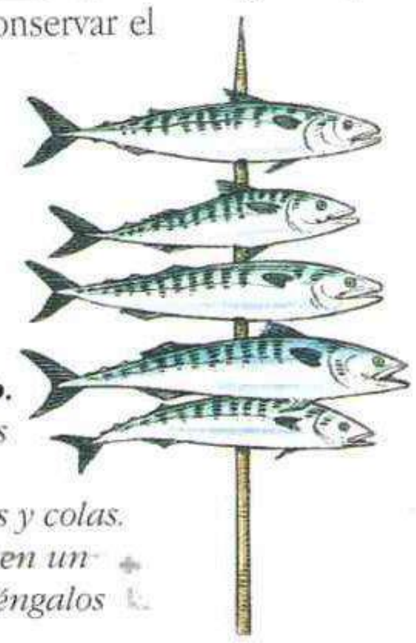

Pescado espetado. *Limpie los pescados pequeños, pero no elimine las cabezas y colas. Clave los pescados en un palo. Después, sosténgalos sobre un fuego.*

CÓMO DESOLLAR UNA ANGUILA

Las anguilas constituyen un plato excelente, pero deben destriparse, desollarse y limpiarse antes de cocinarlas. Su longitud y su piel viscosa y escurridiza hacen que el desollarlas sea muy difícil, a menos que efectúe esta operación inmediatamente después de pescarlas. Limpie una anguila utilizando el método mostrado en la página anterior y desóllela como se muestra a continuación.

1 Pase un palo resistente a través de las agallas de la anguila, detrás del cráneo. Apoye este palo entre otros dos, verticales y bífidos.

2 Sujete la anguila con un trapo y corte la piel alrededor de la cabeza. (Aunque la anguila esté muerta, aún puede llegar a moverse.)

3 Separe la parte superior de la piel de la carne. Después, con un movimiento firme y constante, deslice la piel hacia abajo, en dirección a la cola. Es muy probable que tenga que usar ambas manos.

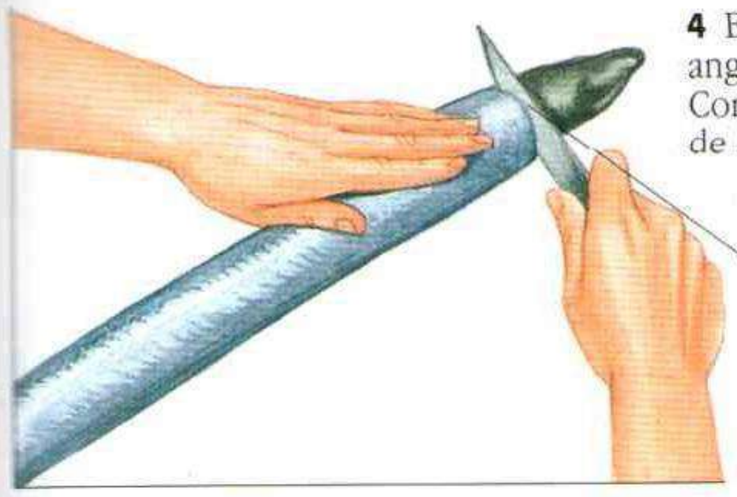

4 Elimine la piel y quite la anguila del palo que la sujetaba. Corte la cabeza y la parte final de la cola.

5 Sujetando la anguila con un trapo, recorra con un cuchillo afilado la columna vertebral de forma paralela a la superficie sobre la cual tiene apoyada la anguila. No corte demasiado cerca de la columna.

UTENSILIOS DE COCINA

LOS UTENSILIOS BÁSICOS de cocina en un campamento son un cazo, una cocina o fuego donde guisar los alimentos, y una cuchara de madera. Al cocinar, tanto el cazo como la cocina o el fuego deben protegerse del viento, y es recomendable que el cazo disponga de una tapa. De esta manera economizará en combustible. La higiene y el orden al cocinar resultan fundamentales. Lave sus manos como primer paso, y después saque de su mochila únicamente lo indispensable para preparar la comida. La mayoría de los alimentos para mochileros es preferible cocinarla y comerla en un solo guiso, preparada con mucha agua y en un mismo cazo.

Los mochileros experimentados tienden a mezclar la fruta, el postre y los platos principales. De esta manera ahorran combustible y lavan menos platos.

Cocina al aire libre. *Mantenga la cocina resguardada del viento y asegúrese de que no pueda volcar. Tenga cuidado de no tocar un cazo caliente con las manos desnudas.*

CAZOS Y SARTENES

Si come directamente de un cazo, en especial si cocina en un fuego al aire libre, el riesgo de quemarse los dedos, los labios, la boca o la lengua es mayor. El empleo de vasos y platos de plástico evita este riesgo. Resulta innecesario cargar con una gran cantidad de recipientes para cocinar, a menos que viaje con un grupo numeroso o en un vehículo. Elija el tipo de cazo que utilizará con mayor frecuencia, y cuide de él.

Tazón. *Un tazón de plástico resistente es de gran utilidad. Debe tener asimismo un asa resistente.*

Plato. *Un plato permite separar una porción de comida mientras el resto se mantiene caliente dentro del cazo.*

Escudilla. *Una escudilla puede utilizarse para comer los cereales del desayuno, unas gachas o un estofado.*

Sartén antiadherente plana. *Una sartén antiadherente es de gran ayuda en un campamento, pero tenga cuidado de no dañar la capa que la recubre, ya que entonces los alimentos se pegarían.*

Sartén antiadherente honda. *Una sartén honda es de utilidad al cocinar para más de una o dos personas.*

Hervidor. *Un hervidor de poco peso le permite utilizar el resto de sus recipientes para cocinar, y no para calentar el agua para las bebidas calientes.*

Juego de cazos. *Un juego de cazos de buena calidad tiene poco peso y contiene un conjunto de cazos, y en ocasiones incluso un hervidor.*

Sartén. *Se pueden encontrar sartenes ligeras especiales para campamentos. Tenga cuidado de no quemarlas.*

Cazo de aluminio. *Los cazos de aluminio calientan la comida rápidamente, pero pueden quemar los alimentos si no se vigilan continuamente.*

CUBIERTOS

Recuerde que los cubiertos deben ser de uso particular y que debe mantenerlos escrupulosamente limpios. Los cubiertos sucios pueden provocar úlceras bucales y otras infecciones. Para limpiarlos, sumérjalos en agua hirviendo siempre que sea posible. Si pierde o rompe su cuchillo, tenedor o cuchara, puede tallar uno nuevo en madera *(véase página siguiente)*. Puede improvisar los cubiertos y la vajilla a partir de ramas, corteza, hongos de árbol y huesos.

TENEDOR CUCHILLO CUCHARILLA

ESTUCHE PARA CUBIERTOS

CÓMO HACER UN RECIPIENTE DE BARRO

1 Moldee una base circular que deberá ser más gruesa que las paredes para dar resistencia y estabilidad al recipiente.

2 Construya las paredes del recipiente con rollos de arcilla previamente moldeados.

3 Suavice la boca del recipiente con el pulgar una vez obtenga la altura deseada.

4 Alise el recipiente con agua. Una vez esté seco, póngalo al fuego para que se endurezca.

CÓMO TALLAR UNA CUCHARA DE MADERA

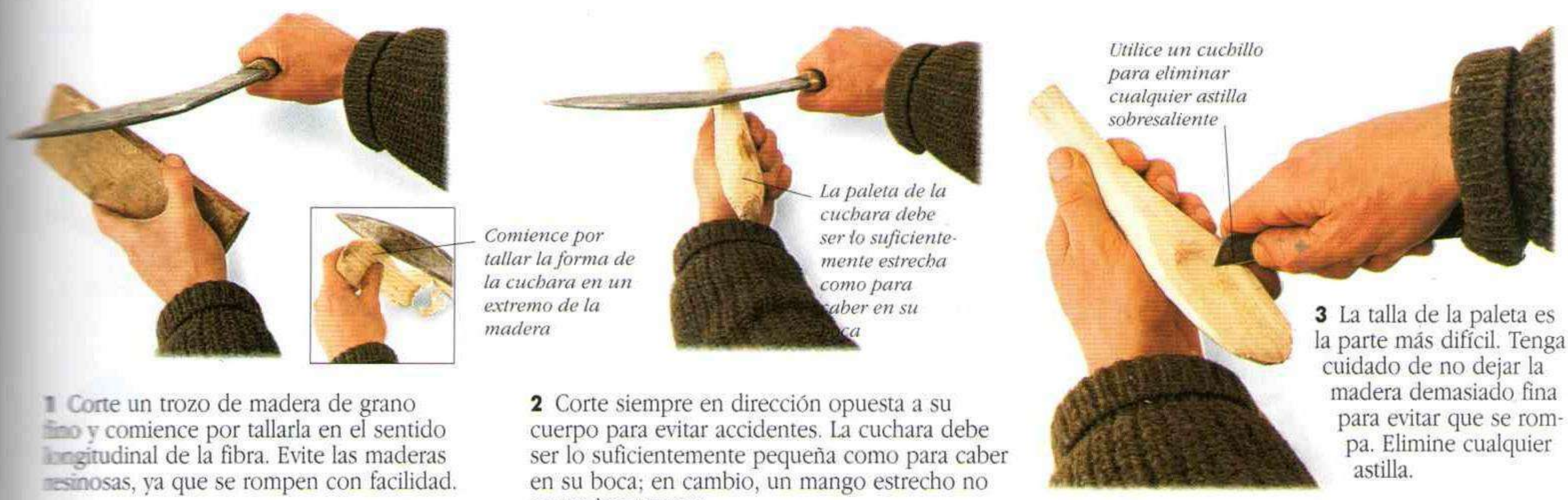

1 Corte un trozo de madera de grano fino y comience por tallarla en el sentido longitudinal de la fibra. Evite las maderas resinosas, ya que se rompen con facilidad.

2 Corte siempre en dirección opuesta a su cuerpo para evitar accidentes. La cuchara debe ser lo suficientemente pequeña como para caber en su boca; en cambio, un mango estrecho no es tan importante.

3 La talla de la paleta es la parte más difícil. Tenga cuidado de no dejar la madera demasiado fina para evitar que se rompa. Elimine cualquier astilla.

CÓMO CONSTRUIR UN CUENCO DE MADERA

1 Talle un cuenco a partir de un bloque o un nudo de madera. Los nudos de madera pueden ser extremadamente duros, por lo que requerirá largo tiempo para tallarlos. El empleo cuidadoso de las ascuas de un fuego puede ayudarle a acelerar el proceso.

2 Retire el carbón de la madera quemada con el cuchillo, y vuelva a aplicar las brasas, soplando suavemente para quemar en el lugar adecuado.

3 En las últimas etapas, tenga cuidado de no dejar la madera demasiado delgada, o de perforar el cuenco con el cuchillo o las brasas.

CAZOS Y SARTENES IMPROVISADOS

Cazo de lata. *Construya un cazo atando un asa de cuerda o alambre a una lata vacía. Tenga cuidado de mantener el cazo por encima de las llamas del fuego, para evitar que el asa de cuerda se queme.*

Sartén de lata de sardinas. *Ate una lata de sardinas vacía a una rama viguda verde, empleando un cordel natural o artificial. Utilícela únicamente sobre un fuego lento, a pesar de que el mango de madera verde soportará mucho calor antes de quemarse.*

COCINAS Y HORNOS

LA FORMA MÁS sencilla de cocinar es en un fuego al aire libre, o con bloques de combustible sólido que arden con una llama regular durante el tiempo suficiente para preparar una bebida caliente. Si no puede construir el fuego con facilidad, o que sea seguro, necesitará una cocina con una llama controlable, que se encienda fácilmente y que queme el combustible con buen rendimiento. Para una cocina más elaborada, puede construir un horno o un foso de vapor. Guarde la cocina, el combustible sólido, las cerillas y el cazo con su tapa en un bolsillo lateral de su mochila, para poder encontrarlos fácilmente cuando los necesite.

Recargar una cocina. *Elija la cocina que más se ajuste a sus necesidades culinarias, y en función de la posibilidad de encontrar gas o combustible para recargarla. Al recargar una cocina, asegúrese de que la llama esté realmente apagada.*

COCINAS DE CAMPAMENTO

Mini-cocina. *Esta cocina ultraligera requiere una mezcla de butano y propano. No es adecuada para soportar cazos grandes.*

Cocina para uso general. *Esta cocina rápida es de uso general, y tiene soportes plegables para cocinar con cazos grandes.*

Cocina de gasolina sin plomo. *Esta cocina requiere gasolina sin plomo y tiene un accionamiento de purga para limpiar el generador.*

Cocina de combustible múltiple. *Es una de las cocinas más populares del mundo; funciona con gasolina, parafina o queroseno.*

ADVERTENCIA

Cualquier cocina debe tratarse con sumo cuidado. Recuerde que siempre existe el riesgo de explosión. Las áreas de cocina deben estar bien ventiladas, ya que las cocinas consumen oxígeno y despiden monóxido de carbono tóxico. Cuando se apaga la llama de una cocina con gas a presión, el vapor se almacena y cualquier llama o chispa dará lugar a una gran explosión.

Cocina no presurizada. *Esta cocina de uso popular funciona con alcohol de quemar, y lleva incorporado un paraviento, así como una batería de cocina. Debido a su forma, el riesgo de rotura de algún componente es mínimo; es bastante estable, pero no calienta con demasiada rapidez.*

COMBUSTIBLE

Las botellas de combustible deben distinguirse fácilmente de aquellas que contienen agua, para evitar confundirlas. No deben presentar fuga alguna que pudiera contaminar los alimentos y deteriorar las ropas y el equipo. La disponibilidad de combustible debe preverse antes de viajar, así como la eliminación de los recipientes usados.

BOTELLA DE GASOLINA

BOTELLA DE PARAFINA

COMBUSTIBLE SÓLIDO A BASE DE ALCOHOL

BOMBONA DE BUTANO/PROPANO

HORNO DE BARRO

1 Cave una zanja estrecha de unos 20 cm de profundidad. Coloque una olla con tapa de unos 23 cm de profundidad sobre uno de sus costados. Clave un palo largo en el suelo justo en la base de la olla.

2 Cubra la olla completamente con tierra o arcilla húmedas, moviendo el palo para dejar libre el agujero de la chimenea. Compacte la tierra para que la cúpula sea lisa y estable.

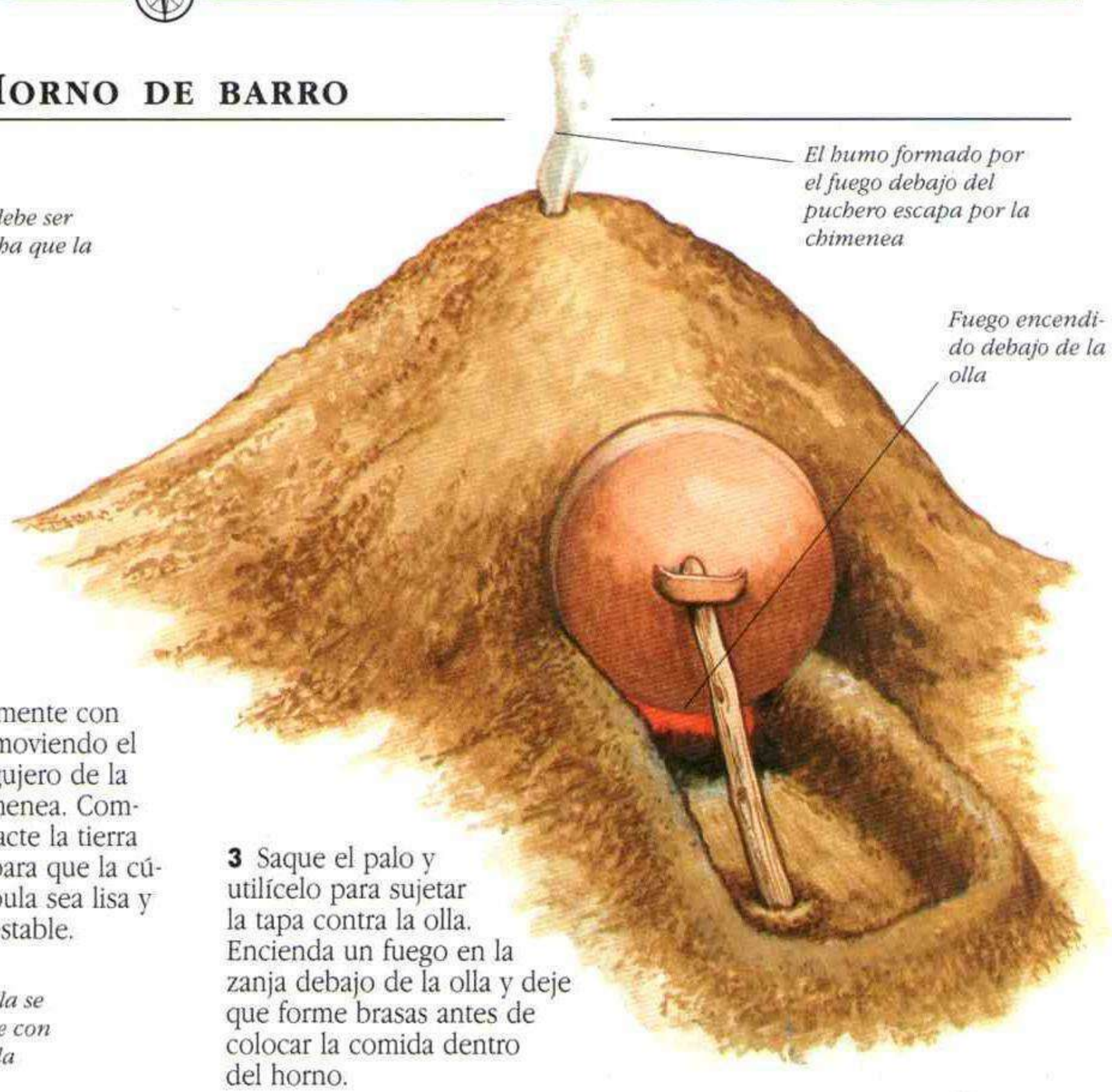

3 Saque el palo y utilícelo para sujetar la tapa contra la olla. Encienda un fuego en la zanja debajo de la olla y deje que forme brasas antes de colocar la comida dentro del horno.

FOSO DE VAPOR

1 Encienda un fuego sobre una plataforma de ramas encima de una zanja *(véase superior)*. Coloque piedras pequeñas sobre la plataforma.

2 Una vez la plataforma se ha quemado y las piedras han caído a la zanja, clave un palo vertical en la zanja. Lleve las ascuas a un lado. Envuelva el alimento en hierbas y colóquelo en la zanja sobre una capa gruesa de hierba.

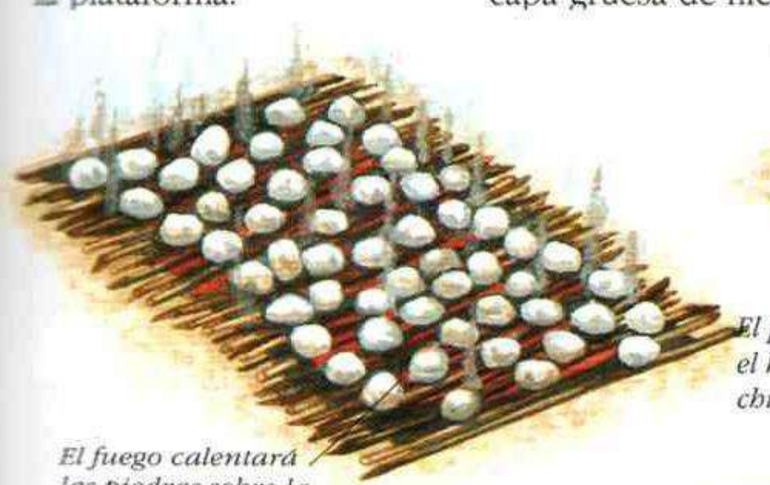

3 Coloque la carne en el centro del foso, donde se acumula la mayor parte del calor. Extienda una capa gruesa de hierba sobre el alimento. Cubra el foso con tierra.

4 Saque el palo y vierta un poco de agua en el hueco para generar vapor.

5 Selle el foso con tierra, compactándola hasta dejarla lisa. El alimento estará cocido transcurridas unas cuatro horas.

HORNO DE YUKÓN

Se trata de una chimenea encima de un fuego, que quemará incluso leña húmeda. Su parte superior constituye una cocina muy efectiva. Construya primero un fuego de tipi *(véase página 60)*, pero no lo encienda. Levante una pared de piedras alrededor del tipi, utilizando arcilla como mortero, y selle el exterior con arcilla. La forma cónica de la estructura incrementa el flujo de aire alrededor del fuego, creando más calor. Encienda el extremo de un palo largo y empújelo hacia el fuego desde arriba. Utilice madera seca para encender el fuego. Una vez esté bien encendido puede secar la madera húmeda sobre el horno.

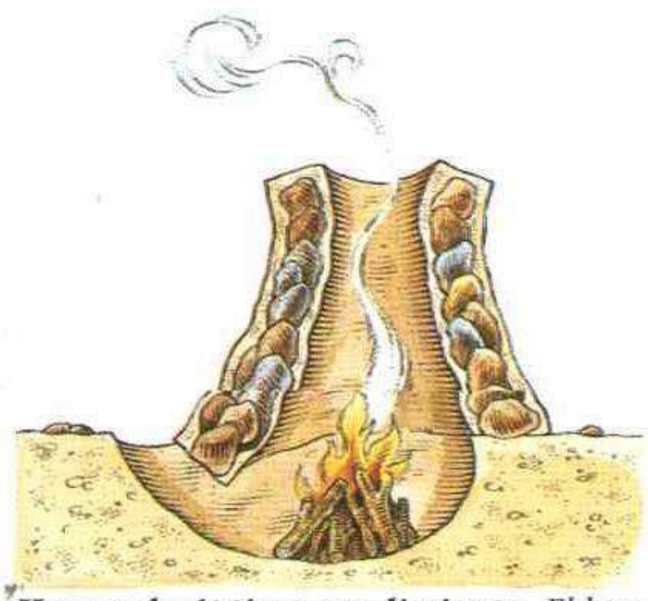

Horno de óptimo rendimiento. *El horno de Yukón consume poco combustible y desprende gran cantidad de calor.*

MÉTODOS DE COCCIÓN

LA FORMA MÁS básica de cocción consiste en asar los alimentos empleando, por lo general, un espetón sobre las llamas o las brasas de un fuego. En una situación de supervivencia es importante cocinar todos los alimentos que den lugar a dudas, a fin de matar cualquier bacteria peligrosa, posibles parásitos o sustancias químicas. Cocinados los alimentos son más apetecibles y digeribles que en su estado natural. Además, una comida caliente levanta el ánimo a cualquiera.

Cocinar sobre un fuego al aire libre. *Puede preparar un estofado en un cazo suspendido sobre el fuego, o puede esperar hasta que las llamas se conviertan en brasas para utilizarlas en otras modalidades de cocina* (véase página siguiente). *Si dispone de grandes cantidades de agua, tenga siempre a mano agua hervida para preparar bebidas calientes.*

SOPORTES PARA CAZOS

Espetón. *Prepare un espetón clavando dos palos ahorquillados en el suelo a ambos lados del fuego y apoyando un tercer palo entre ellos. Evite que el fuego arda demasiado y envuelva el cazo. Utilice un palo para empujar el cazo hacia el extremo del espetón cuando quiera servirse la comida.*

Soporte oscilante. *Resulta muy útil al cocinar estofados que deban apartarse del fuego para ajustar su temperatura. Clave un palo ahorquillado en el suelo y equilibre una rama larga sobre la horquilla para que de un extremo cuelgue el cazo sobre el fuego. Ate el otro extremo al palo vertical. El cazo puede apartarse del fuego con un palo.*

Trípode. *Este método resulta muy estable y muy seguro. Construya un trípode* (véase página 75), *y cuelgue una rama en forma de «U» de su centro, atada con un cordel. Cuelgue el cazo de la rama. Esto le permitirá levantar o bajar el cazo sobre el fuego y controlar así la temperatura de la comida. Permita que el fuego baje antes de colocar el trípode sobre los rescoldos.*

Una rama en forma de «U» permite ajustar la posición del cazo

El fuego debe quedar en rescoldos antes de comenzar a cocinar

EL EMPLEO DEL BAMBÚ

Al ser resistente y hueco, el bambú constituye un excelente recipiente para cocinar. Con una rama terminada en punta, haga un agujero en cada una de las paredes que separan los compartimentos huecos, dejando intacta la de un extremo. Vierta agua en el tallo hasta llegar al primer anillo. Apoye el tallo sobre un fuego sujetando el extremo superior con una rama en forma de «U». Caliente el tallo para hervir el agua y cocinar al vapor los alimentos colocados en la parte superior.

Recipiente natural. *Un tallo de bambú hueco puede emplearse como recipiente para cocinar alimentos sobre un fuego abierto.*

COCINAR SOBRE PIEDRAS CALIENTES

1 Construya un lecho denso de piedras grandes y planas. No utilice pizarra ni otro tipo de roca estratificada porque podría estallar en pedazos al ser calentada por la acción del fuego.

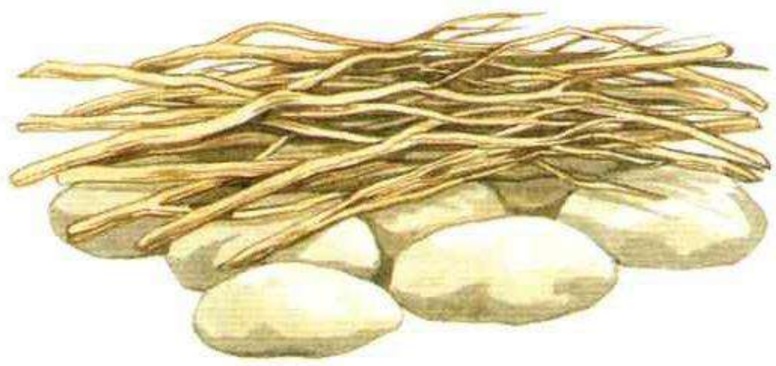

2 Coloque yesca sobre las piedras y cúbrala con ramas secas Después encienda la yesca. Puede dejar que el fuego se reduzca a cenizas mientras prepara la comida *(véase página 110)*.

3 Con un manojo de ramas verdes, limpie las cenizas y los rescoldos de la superficie de las piedras. Tenga cuidado de no tocarlas con las manos porque estarán extremadamente calientes.

La comida se cuece directamente sobre las piedras calientes

4 Coloque la comida encima de las piedras. Los alimentos que precisen una cocción prolongada y lenta deben colocarse a los lados, donde las piedras están más frías que en el centro. Continúe colocando comida, hasta que las piedras se hayan enfriado. Las patatas sin pelar pueden enterrarse bajo las piedras calientes y dejar que se horneen.

HORNEAR EN BARRO

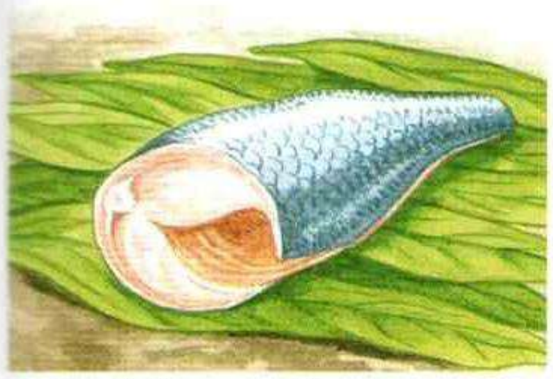

1 Puede hornear alimentos como el pescado envueltos en barro al fuego. Recoja un manojo grande de hierbas y largas hojas.

2 Envuelva el alimento en las hojas y ate el envoltorio con hebras trenzadas de hierbas.

3 Cubra la totalidad del paquete con barro. Asegúrese de que queda completa y uniformemente cubierto, y de que esté bien cerrado.

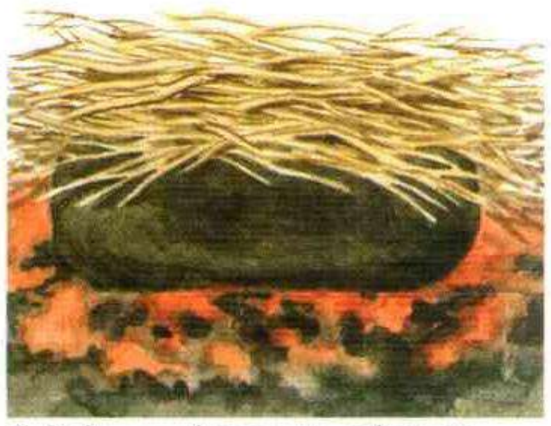

4 Coloque el paquete sobre un lecho de rescoldos calientes y construya un fuego en la parte superior. El pescado puede quedar cocido en una hora.

CONSEJOS CULINARIOS

- Las hierbas y especias son de gran valor al preparar la comida ya que la hacen más apetecible. También son útiles el ajo fresco, las cebollas y los cubos de caldo.
- La forma más sencilla de cocinar es hirviendo agua y utilizándola para hacer un estofado «todo en uno». La mezcla de una conserva de pescado con los bizcochos y los frutos secos parece desagradable, pero con la adición de curry en polvo y una gran cantidad de aire fresco, le parecerá una comida deliciosa.
- Si hierve sus alimentos en agua, es importante beberla también, para aprovechar al máximo el valor nutricional del alimento (a menos que lo hierva para eliminar las sustancias tóxicas).

PAN HÚMEDO *(DAMPER BREAD)*

Dé a la masa la forma de un salchichón

1 Este manjar típico de un campamento australiano es fácil y rápido de preparar. En primer lugar, prepare una masa elástica a partir de harina y agua, y déle la forma de un salchichón largo y grueso.

La espiral de masa se desprenderá del palo cuando esté hecha

2 Enróllelo en forma de espiral alrededor de un palo de madera y sujételo sobre los rescoldos de un fuego mientras se cuece. El pan se desprenderá del palo cuando esté listo.

LA CONSERVACIÓN DE LOS ALIMENTOS

UNA VEZ RECOLECTADOS, los alimentos comienzan a deteriorarse. Sin refrigeración, la carne fresca y el pescado son de consumo dudoso un día después de ser capturados en cualquier clima, a excepción de las regiones polares, mientras que las frutas y las verduras pierden sus valores nutricionales sin un almacenamiento adecuado. Para una situación de supervivencia a medio o largo plazo, debe recolectarse una gran cantidad de alimento en la estación de mayor abundancia, y conservarlo para las épocas de escasez, cuando resulta difícil seguir una dieta equilibrada. Recorrer los alrededores en busca de alimento para un solo día requiere una gran cantidad de calorías. En cualquier situación de supervivencia, a menos que sea muy corta, debe comenzar a conservar y almacenar alimentos inmediatamente, y a guardar sus provisiones, como las conservas, como raciones de emergencia real.

CÓMO SECAR LOS ALIMENTOS

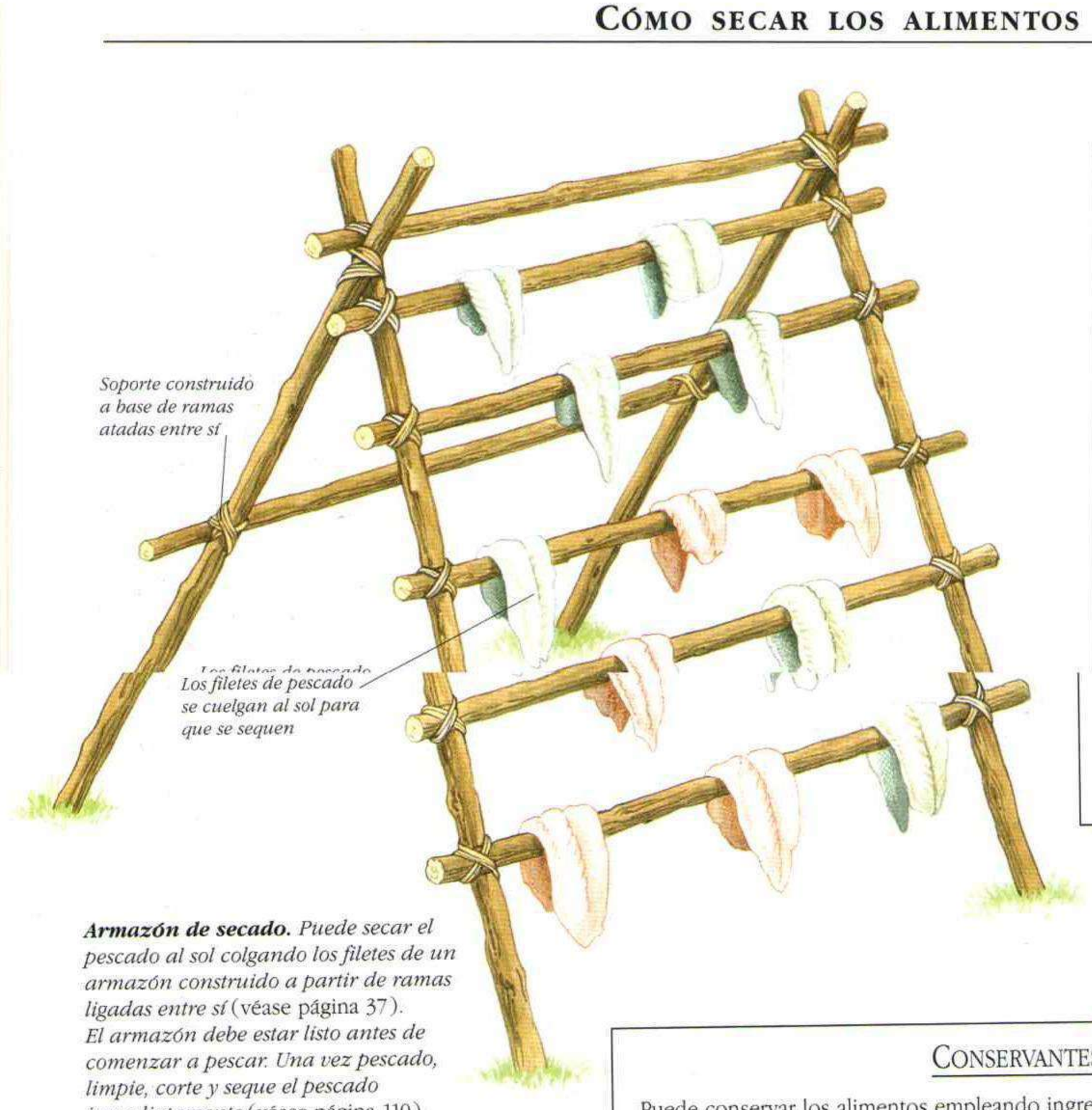

Armazón de secado. *Puede secar el pescado al sol colgando los filetes de un armazón construido a partir de ramas ligadas entre sí* (véase página 37). *El armazón debe estar listo antes de comenzar a pescar. Una vez pescado, limpie, corte y seque el pescado inmediatamente* (véase página 110). *Es necesario un sol intenso para que el pescado se seque antes de que empiece a estropearse. Coloque el armazón en dirección hacia donde sople la brisa para alejar las moscas. Cuelgue los filetes de pescado del armazón dejando mucho espacio entre las piezas para permitir la circulación del aire. Los filetes estarán listos cuando toda el agua que contenían haya sido eliminada y muestren un aspecto quebradizo. Cuando los filetes de pescado estén suficientemente secos, puede almacenarlos para consumirlos más adelante, o puede molerlos entre dos piedras y utilizarlos para espesar sopas o estofados, o para preparar* pemmican (véase página siguiente).

OTROS MÉTODOS DE SECADO

Secado natural. *En los climas cálidos y secos, puede poner a secar setas y plantas sobre una roca. Para secar hierbas, cuélguelas por los tallos, boca abajo.*

Secado sobre el fuego. *Para secar frutas, córtelas en rodajas, y cuélguelas de un espetón sobre un fuego humeante. Tenga cuidado de que el fuego no se avive demasiado y cocine la fruta.*

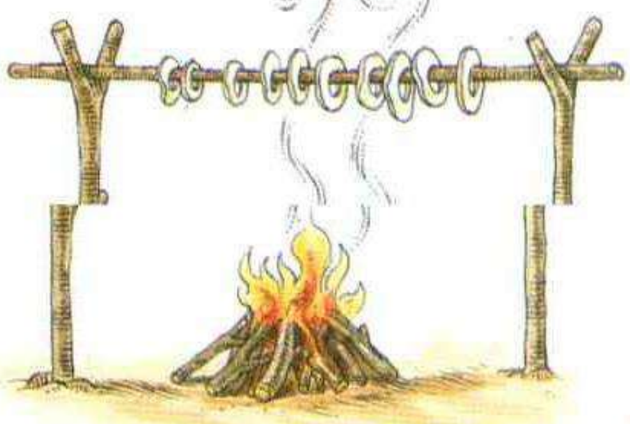

CONSERVANTES NATURALES

Puede conservar los alimentos empleando ingredientes naturales. Sale el pescado y sumerja los vegetales en jugo de lima o de limón, o hiérvalos en agua que contenga especias para encurtir. El pescado salado se conservará bien a temperaturas de hasta 11 °C.

SAL — ESPECIAS PARA ENCURTIR

LIMA

LIMÓN

CONSTRUCCIÓN DE UN TIPI PARA AHUMAR

1 Si no hay suficiente sol para secar el pescado, puede ahumarlo en un tipi especialmente construido para este fin. Ate tres ramas verticales por un extremo con un cordel *(véase página 37)*.

2 Separe las patas unos 60 cm y ate unas ramas horizontales entre ellas para formar el soporte de un entramado o plataforma de secado.

3 Asegúrese de que el triángulo de ramas que forma la plataforma de secado está por lo menos a unos 30 cm del suelo, para poder construir un fuego debajo de ella.

4 Coloque ramas más pequeñas a través del triángulo para formar un entramado sobre el cual ahumar los alimentos.

5 Deje suficiente espacio entre las ramas para permitir que el humo entre en contacto con el alimento sobre el soporte. Coloque el ahumador sobre un fuego hasta que el pescado esté seco.

CÓMO PREPARAR *PEMMICAN*

1 Puede conservar el pescado cocido preparando *pemmican*. En primer lugar, trabaje con las manos una porción de grasa de pescado para ablandarla.

2 Corte o ralle el pescado cocinado en trozos muy pequeños y mézclelos con la grasa. Puede añadir condimentos, si así lo desea.

3 Añada bayas para dar mejor sabor, y forme porciones del tamaño de una pelota de golf.

4 Dé a cada porción la forma de una salchicha y colóquela al sol para que se endurezca. El *pemmican* se conserva durante años y es muy nutritivo.

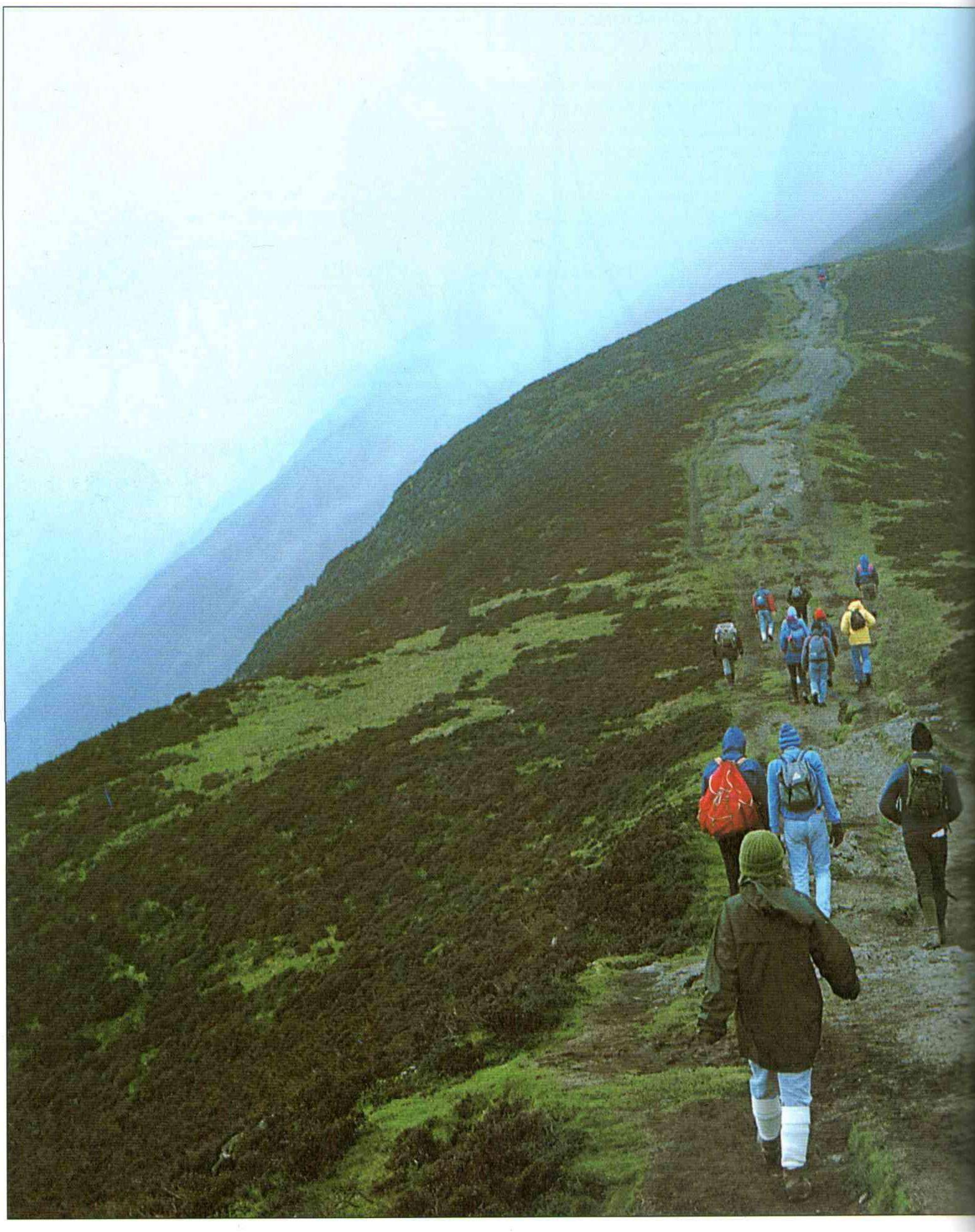

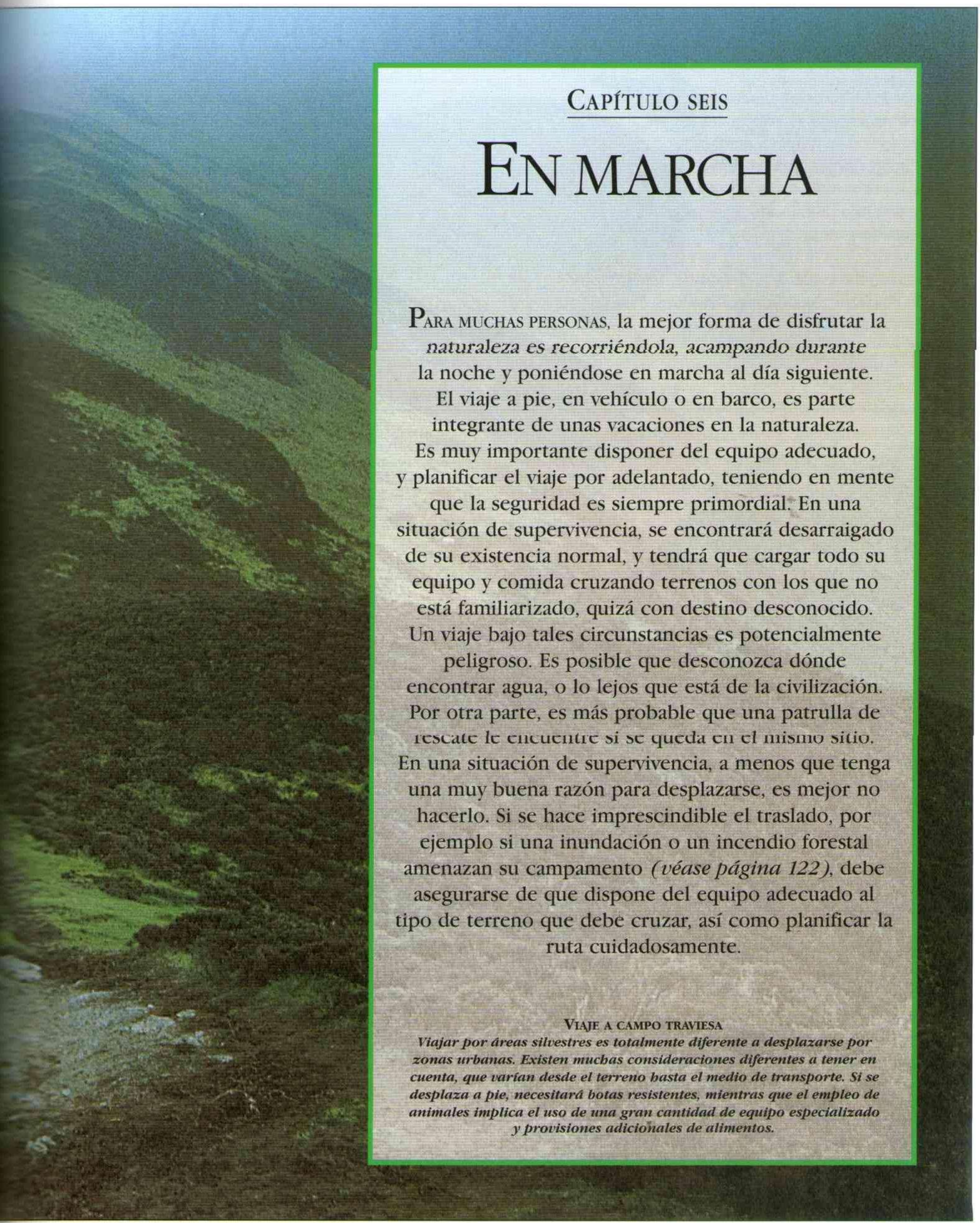

CAPÍTULO SEIS

EN MARCHA

PARA MUCHAS PERSONAS, la mejor forma de disfrutar la *naturaleza es recorriéndola, acampando durante* la noche y poniéndose en marcha al día siguiente. El viaje a pie, en vehículo o en barco, es parte integrante de unas vacaciones en la naturaleza. Es muy importante disponer del equipo adecuado, y planificar el viaje por adelantado, teniendo en mente que la seguridad es siempre primordial. En una situación de supervivencia, se encontrará desarraigado de su existencia normal, y tendrá que cargar todo su equipo y comida cruzando terrenos con los que no está familiarizado, quizá con destino desconocido. Un viaje bajo tales circunstancias es potencialmente peligroso. Es posible que desconozca dónde encontrar agua, o lo lejos que está de la civilización. Por otra parte, es más probable que una patrulla de rescate le encuentre si se queda en el mismo sitio. En una situación de supervivencia, a menos que tenga una muy buena razón para desplazarse, es mejor no hacerlo. Si se hace imprescindible el traslado, por ejemplo si una inundación o un incendio forestal amenazan su campamento *(véase página 122)*, debe asegurarse de que dispone del equipo adecuado al tipo de terreno que debe cruzar, así como planificar la ruta cuidadosamente.

VIAJE A CAMPO TRAVIESA

Viajar por áreas silvestres es totalmente diferente a desplazarse por zonas urbanas. Existen muchas consideraciones diferentes a tener en cuenta, que varían desde el terreno hasta el medio de transporte. Si se desplaza a pie, necesitará botas resistentes, mientras que el empleo de animales implica el uso de una gran cantidad de equipo especializado y provisiones adicionales de alimentos.

LA PLANIFICACIÓN DEL VIAJE

AUN FORMANDO PARTE de una expedición bien planificada, el viaje puede exponerle, junto con su grupo, a muchos riesgos imprevisibles. En una situación de supervivencia, estos riesgos son aún más pronunciados. Es importante estar preparado y tener un plan de acción. Si espera ser rescatado debe quedarse donde está. Sin embargo, puede verse forzado a abandonar su campamento por varias razones. Los animales pueden acabar con sus provisiones, o un desastre natural, como una inundación o un incendio forestal, pueden destruir la vegetación que le rodea. Antes de emprender el viaje, debe reconocer cuidadosamente el terreno que pretende cruzar.

Inundaciones. *Las inundaciones destruyen la vegetación y dejan una profunda capa de lodo que puede dificultar la búsqueda de plantas comestibles o de madera con la que construir refugios o encender fuego.*

EL RECONOCIMIENTO DEL TERRENO

Montañas. *Las montañas sin árboles proporcionan poco refugio, y existe el peligro continuo de desprendimientos, nevadas, hielo, además de un clima variable. Puede requerir equipos especiales y habilidades que sólo posee un montañero para superar las pendientes más peligrosas* (véase página 146).

Bosques templados. *En un bosque persiste el peligro de caída de árboles muertos. Los animales peligrosos, como los jabalíes, habitan estas áreas* (véase página 166). *Aunque la búsqueda de alimento o de materiales para construir un refugio resulta fácil, desplazarse no lo es tanto.*

Desierto. *El agua es la consideración clave en el desierto* (véase página 142). *Sin nubes que retengan la humedad, las temperaturas suben alarmantemente durante el día para descender rigurosamente durante la noche. Si tuviera que cruzar un desierto, hágalo de noche, de una fuente de agua a la siguiente.*

EVALUAR A SUS COMPAÑEROS

Un grupo de supervivientes debe elegir a un líder, capaz de evaluar las capacidades de cada individuo antes de establecer un plan de viaje. El líder debe considerar, con cuidado y en privado, los puntos fuertes y débiles de cada miembro del grupo, y así establecer un plan de acción basándose en la capacidad de cada uno. El liderazgo en esta situación suele ser muy difícil, debido a la presión de una situación de vida o muerte y, especialmente, cuando algún individuo no acepta someterse a la dirección del líder.

Trabajar en equipo. *Las posibilidades de supervivencia para el grupo serán mayores si permanece unido y mantiene su integridad gracias a un líder íntegro y dedicado.*

Estar preparados. *Antes de desplazarse, debe hacerse con la mayor cantidad de información posible acerca de la región que pretende cruzar. No suponga siempre que los mapas son exactos* (véase página 126), *en particular cuando planifique buscar agua. Vale la pena enviar una serie de expediciones cortas, con equipo ligero, para reconocer la zona antes de emprender cualquier desplazamiento.*

EL EMPLEO DE PRISMÁTICOS

Unos prismáticos le ahorrarán muchos kilómetros. Ajuste la imagen hasta que sea nítida y observe el terreno. Explore la zona desde distintos puntos de observación, e intente descubrir por dónde fluyen las corrientes o los ríos, o la situación de los montes.

Escrutar. *Recorra el paisaje con la mirada, observando cuidadosamente, y no simplemente mirando los alrededores.*

Ríos. *Los ríos son peligrosos, y debe acercarse a ellos con cuidado. Para cruzar un río, puede que tenga que buscar troncos caídos para emplearlos como puente, o buscar zonas más angostas y poco profundas para vadearlos* (véase página 150). *También pueden estar habitados por animales peligrosos, como los cocodrilos* (véase página 75).

Selvas. *El calor y la humedad de la selva pueden llegar a ser muy molestos, además de poco saludables para los humanos, pero permiten el desarrollo de los insectos y las bacterias. Las heridas más pequeñas se infectan, la comida se descompone y la ropa está continuamente empapada debido a la lluvia o al sudor, o a ambos factores. En la selva, los desplazamientos se hacen particularmente difíciles.*

Sabana. *El agua es un factor importante a considerar en este ecosistema. Puede tener problemas para encontrar refugio* (véase página 53). *Recuerde que también estará expuesto al peligro que suponen los depredadores que cazan los grandes rebaños de herbívoros que pastan habitualmente en estas enormes áreas.*

Mar y playa. *No subestime nunca el mar. Siempre es peligroso. En las zonas de mareas es mucho más fácil caminar por la arena durante la marea baja que superar las dunas o los acantilados. Sin embargo, debe vigilar continuamente para asegurarse de que sabe en qué dirección viene la marea. Puede encontrarse con criaturas peligrosas en la playa o en las aguas poco profundas* (véase página 109).

LA ORIENTACIÓN

LA ORIENTACIÓN ES el aspecto más importante de la navegación, así que una brújula fiable se convertirá en su herramienta más importante. Un reloj o cronómetro le será de gran utilidad, ya que puede indicar tanto la dirección como el tiempo recorrido, con lo que podrá estimar la distancia recorrida *(véase página siguiente)*. Existen muchos tipos diferentes de brújula que funcionan atraídos por los polos del campo magnético de la Tierra. Si no dispone de una brújula, podrá orientarse observando la posición del sol y de las estrellas, los animales y las plantas.

BRÚJULAS

Las agujas de una brújula señalan el norte magnético y el sur magnético, los polos del campo magnético natural de la Tierra. Imagínese que la Tierra contiene un gran imán vertical, y que los polos se encuentran en la parte superior e inferior del imán, de los cuales emana la carga magnética. Hay muchos tipos diferentes de brújula. Si únicamente pretende dar un paseo, basta con una brújula sencilla para un caso de emergencia *(véase página 28)*, pero utilice una brújula como transportador con fines de navegación y orientación básica *(véase inferior)*. Las brújulas prismáticas incluyen un prisma a través del cual se toman los rumbos, y son resistentes y precisas. Algunas de estas brújulas incluyen un espejo además de, o en lugar del prisma. Sin embargo, son bastante caras. Una brújula Silva es lo único que necesita realmente para una navegación precisa, y le evitará complicaciones que pueden provocar errores.

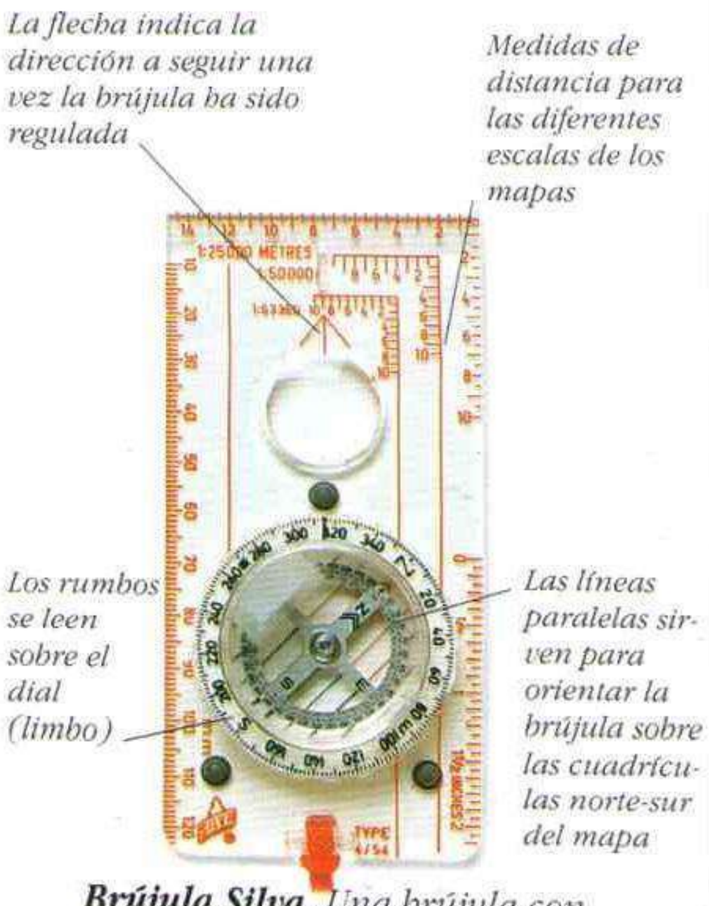

Brújula Silva: *Una brújula con transportador le permite medir rumbos sobre un mapa sin tener que mover la brújula* (véase página 130). *De esta manera evita el tener que usar un transportador aparte y tener que hacer cálculos aritméticos mentales para determinar la ruta.*

CÓMO FABRICAR UNA BRÚJULA

Si se suspende libremente una pieza de metal ferroso (que contiene hierro) imantado, girará hasta orientarse en un eje norte-sur. Esto se debe a que se ve atraída por el campo magnético de la Tierra. Aprovéchese de este fenómeno para construir una brújula, imantando un trozo de metal.

En lugar de utilizar un paño de seda para imantar una aguja, puede utilizar un imán, que hará la misma función pero más deprisa

1 Magnetice una aguja frotándola repetidamente en una dirección sobre un paño de seda, como su pañuelo *(véase página 142)*. Esto genera una carga de electricidad estática.

Permita que la aguja se mueva libremente haciéndola flotar en un cuenco lleno de agua

2 Haga flotar una brizna de hierba en un cuenco lleno de agua, y coloque la aguja sobre ella. La aguja se orientará por sí misma apuntando en dirección norte-sur.

3 Una vez la aguja esté fija, compruebe la posición del norte utilizando otros indicadores *(véase página siguiente)* y marque el extremo norte de la aguja. Ahora puede utilizarla como brújula.

UTILICE UNA CUCHILLA DE AFEITAR COMO BRÚJULA

En un caso de emergencia, puede construir una brújula con una cuchilla de afeitar en lugar de una aguja, si dispone de ella. Sin embargo, tenga la precaución de desgastar la hoja sobre una piedra, y tenga mucho cuidado al imantarla.

Cuchilla de afeitar como brújula. *Imante una cuchilla de afeitar sin filo, y suspéndala de un hilo para que cuelgue libremente y apunte hacia el norte.*

CÓMO ORIENTARSE A TRAVÉS DEL SOL

Como el sol siempre sale por el este y se oculta por el oeste, puede ser utilizado como indicador de dirección. Mientras pueda verlo y sepa la hora, podrá orientarse en relación al sol. También puede dar con la dirección aproximada construyendo un reloj de sol, o empleando un reloj como transportador.

Hemisferio norte. *Apunte hacia el sol la manecilla horaria. Imagine una línea que divida en dos el ángulo entre la manecilla horaria y las 12.00 h. Esta línea apuntará hacia el sur.*

Hemisferio sur. *Si se encuentra en el hemisferio sur, apunte las 12.00 h del reloj hacia el sol. El norte se encuentra en el punto medio entre las 12.00 h y la manecilla horaria.*

RELOJ DE SOL

Clave una rama vertical en el suelo. Por la mañana, observe dónde cae la sombra y marque el punto exacto en el suelo (será el oeste en el hemisferio norte). Ate una cuerda de la misma longitud de la sombra a un palo pequeño y dibuje un arco alrededor de la rama. Marque el punto en el que la sombra toca el arco por la tarde (el este).

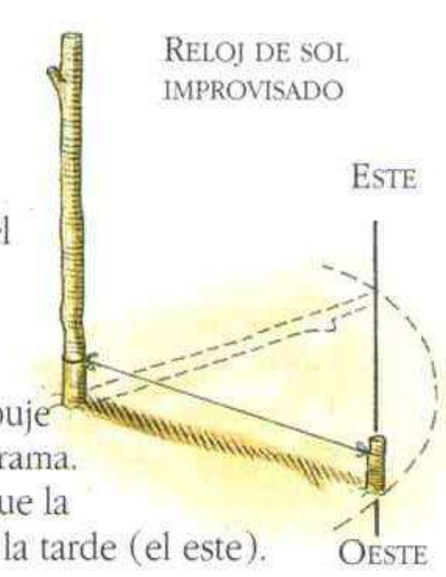

CÓMO ORIENTARSE A TRAVÉS DE LAS ESTRELLAS

Las estrellas permanecen siempre en la misma posición. Sólo parecen moverse en el cielo debido al movimiento de la Tierra. Únicamente hay una estrella que aparenta no moverse: la Estrella Polar. Esta estrella puede utilizarse en el hemisferio norte para encontrar el norte. En el hemisferio sur, debe localizar la Cruz del Sur para encontrar el sur.

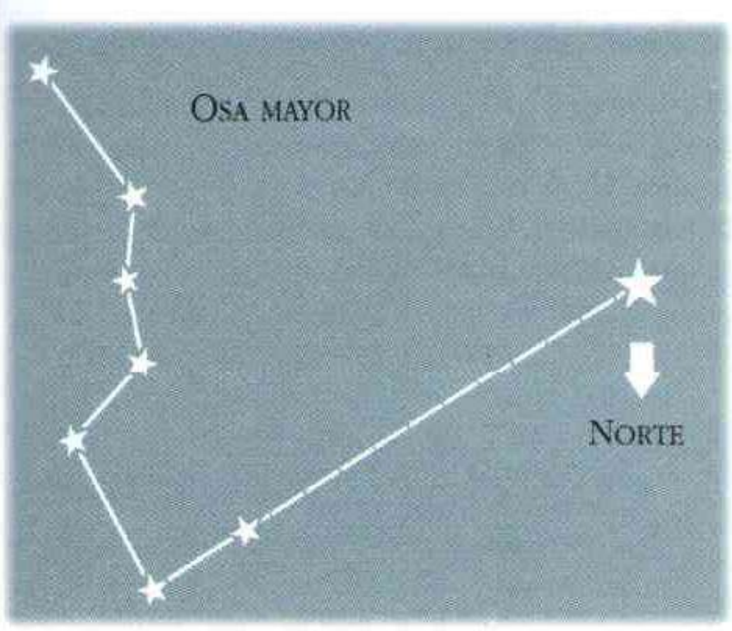

Hemisferio norte. *Mirando hacia la Osa Mayor (el Carro), trace una línea imaginaria entre las dos estrellas que forman la parte frontal de la Osa y prolónguela multiplicada por cuatro hasta encontrar la Estrella Polar. Esta brillante estrella está situada al norte en el horizonte.*

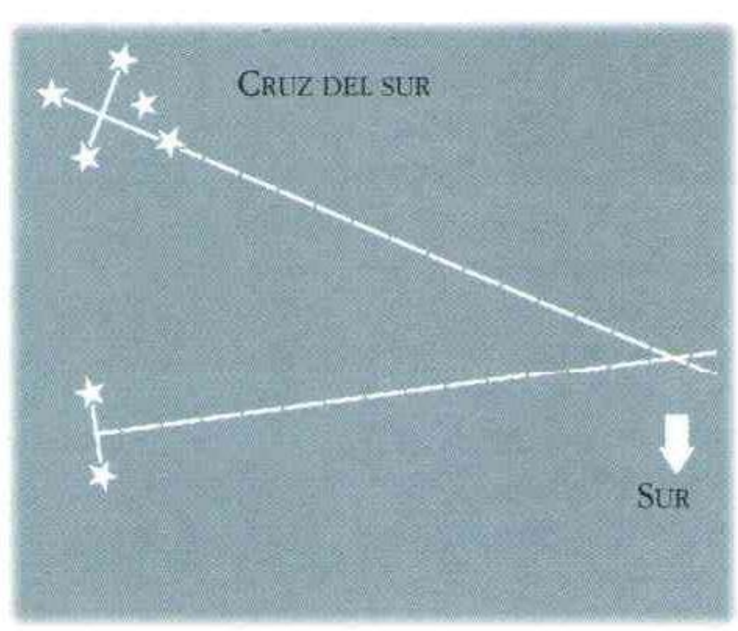

Hemisferio sur. *Utilice la Cruz del Sur para saber de forma aproximada dónde está el sur. Trace una línea imaginaria desde el travesaño, aproximadamente unas 4 veces y media su longitud. El sur debe estar en el horizonte debajo de este punto. Dos estrellas brillantes debajo de la Cruz le ayudarán a encontrar el punto correcto.*

INDICADORES DE DIRECCIÓN NATURALES

La naturaleza está mejor orientada y adaptada al campo magnético de la Tierra que los humanos. Las plantas siempre crecen en dirección al sol —hacia el sur en el hemisferio norte y hacia el norte en el hemisferio sur. Los girasoles siguen el movimiento del sol cruzando el cielo. La planta denominada Polo Norte en Sudáfrica se inclina hacia el norte. Los anillos de los árboles son a menudo mucho más anchos en el lado en el que da el sol.

Nidos de termitas. *Algunos nidos de termitas están construidos siguiendo un eje norte-sur. De esta manera reciben el máximo calor durante la mañana y la tarde, y están protegidos del sol al mediodía.*

Pájaros tejedores. *Algunos pájaros pueden utilizarse como indicadores de dirección básicos. Por ejemplo, algunas especies de pájaros tejedores construyen sus nidos únicamente en la cara oeste de los árboles. Si se encuentra en una zona donde viven estos pájaros —por ejemplo, en Sudáfrica— le servirán como indicativo de dirección bastante fiable.*

EL EMPLEO DE MAPAS

EXISTEN MUCHOS TIPOS diferentes de mapas, cada uno diseñado para un propósito diferente. La elección del mapa adecuado resulta fundamental si se adentra en un territorio desconocido. Necesitará mapas a gran escala de la totalidad del área que pretenda visitar, así como mapas detallados de aquellas zonas que recorrerá a pie. Los mapas a escala 1:50.000 son ideales para traslados a pie (esta escala significa que, por cada unidad de medida en el mapa, la distancia real es 50.000 veces mayor). Por lo general, tienen una cuadrícula de 2 cm, cada una de las cuales representa 1 km, y están marcados con dos coordenadas para cada línea, que va de 0 a 99. Por muchas razones, algunos mapas pueden ser bastante inexactos. Modifíquelos sobre la marcha y aprenda a sortear esta falta de exactitud.

CÓMO DESCIFRAR UN MAPA

El agua determina el paisaje, ya que lo cruza formando corrientes y ríos, creando a su paso colinas, montañas y valles. Las construcciones humanas, como las carreteras, pueden cambiar rápidamente. Para entender las características de un terreno utilizando un mapa, identifique en primer lugar los ríos, luego los valles y las colinas, que apenas varían durante la vida de un mapa.

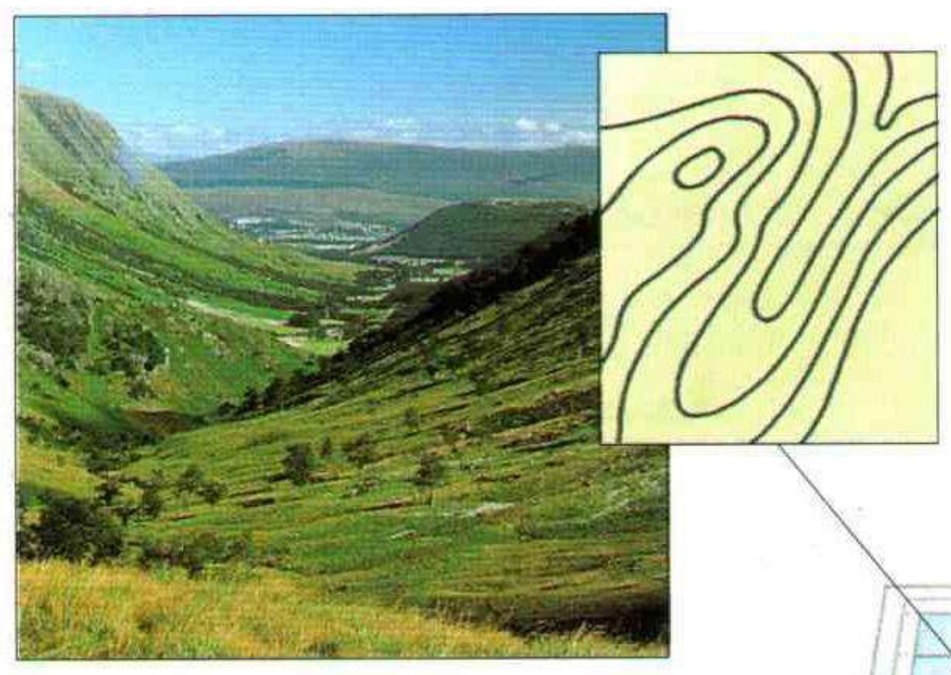

Valle. *Un valle, y el río que lo ha formado, aparecen en un mapa como un remolino complicado de líneas de igual altura, mostrando su forma exacta, en particular las zonas que no son visibles para el observador que se acerca a ellas desde tierra. La altura sobre el nivel del mar se marca con números a lo largo de las líneas.*

Montaña. *Un mapa muestra un plano completo de la montaña, la mayor parte de la cual no es visible desde cualquier posición en tierra. Debe imaginarse la parte que no ve desde el punto en que esté situado. La altura se indica en el mapa como si estuviera escrita sobre la ladera.*

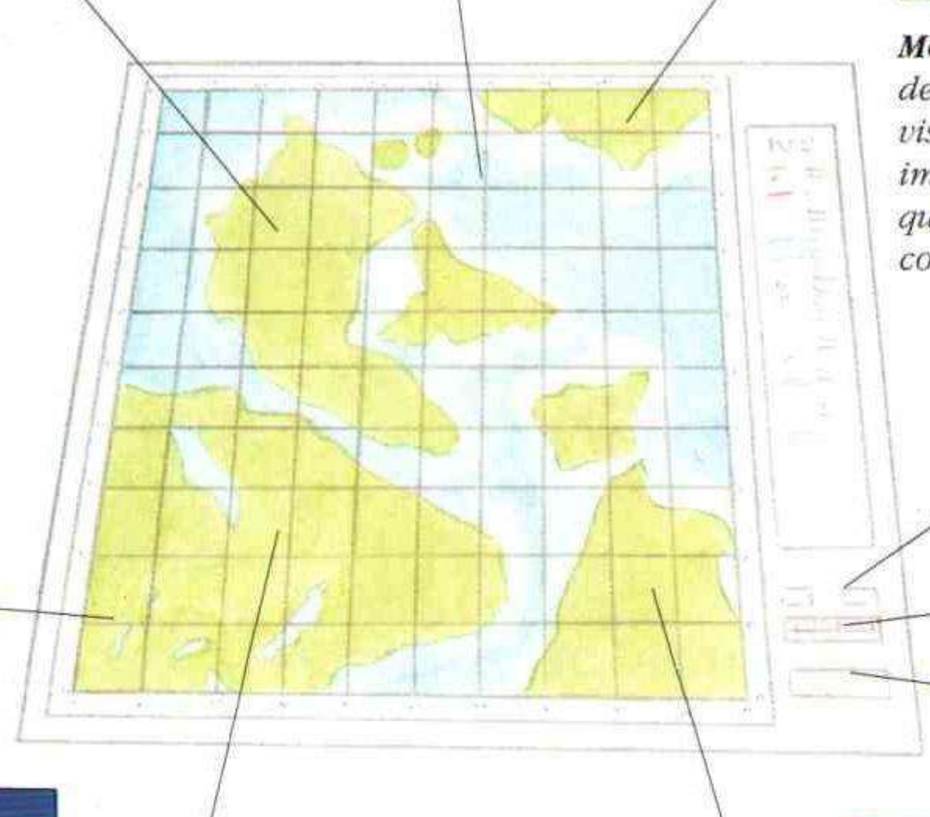

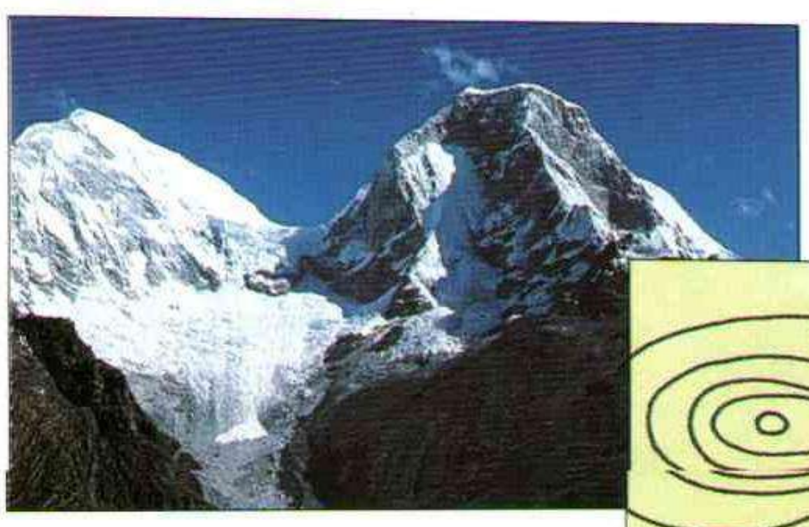

Collados. *Un collado es una depresión en una cordillera entre dos montañas o picos. A nivel del suelo, si mira desde un extremo, únicamente verá una de las montañas. En un mapa, un paso aparece representado por dos círculos enlazados por curvas de nivel.*

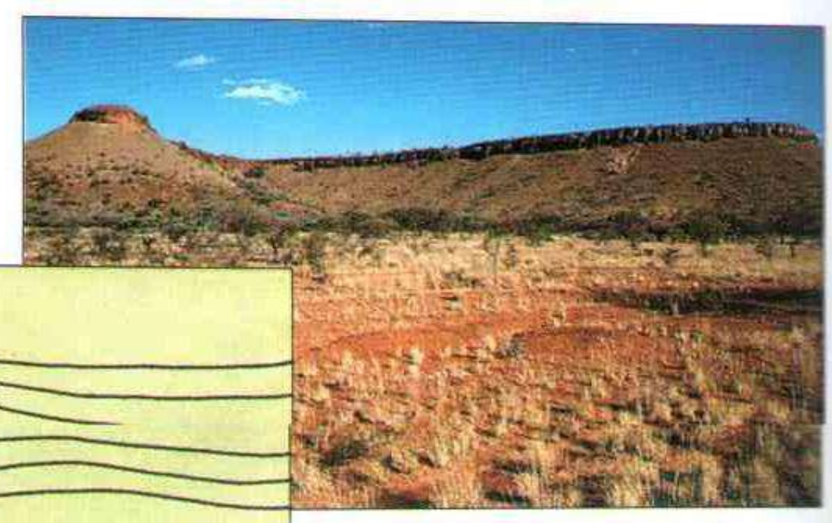

Cresta. *Los ríos siempre están separados por crestas. Éstas se dibujan en un mapa como los dedos de una mano, sobresaliendo de alguno de los lados de una montaña. A menudo es más fácil caminar a lo largo de la parte superior de la cresta que por el valle, en particular al cruzar una selva densa.*

GRADIENTES

Las curvas de nivel en un mapa unen puntos de igual altitud. Las alturas de algunas de las curvas de nivel están indicadas en los mapas. Por lo general, se escriben encarando la parte inferior de la pendiente, como si estuvieran escritas sobre la ladera de la montaña.

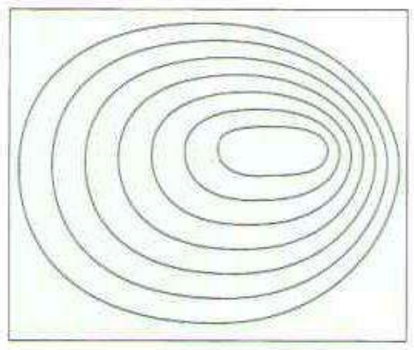

Pendiente pronunciada. *Generalmente existirá una pendiente gradual entre curvas de nivel. Unas líneas muy contiguas indican que la pendiente es muy pronunciada.*

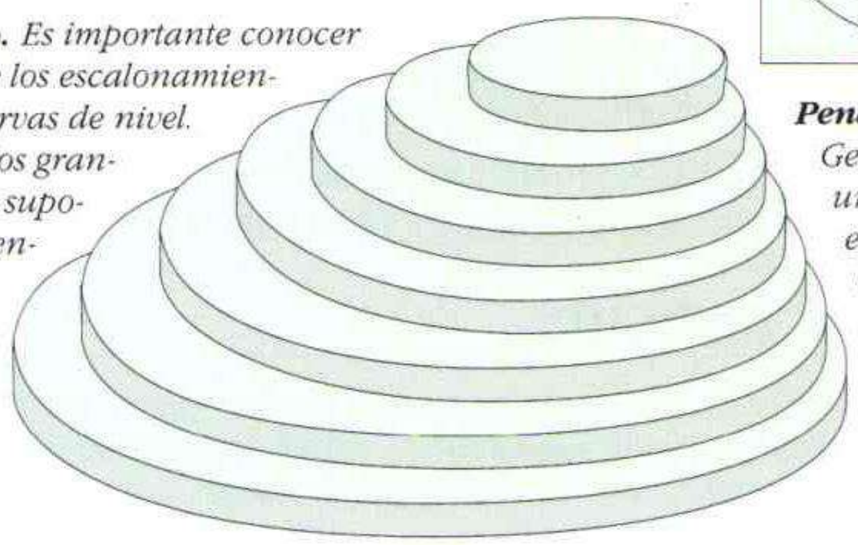

Acantilado. *Es importante conocer la altura de los escalonamientos entre curvas de nivel. Los intervalos grandes pueden suponer la existencia de un acantilado.*

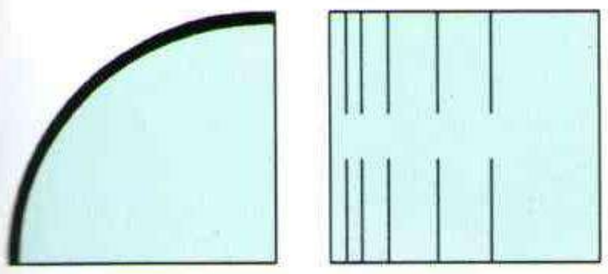

Pendiente convexa. *No es posible vislumbrar la cima de una pendiente convexa si se encuentra a su pie. Las curvas de nivel se acercan mucho en la base de la pendiente y se alejan en la cima.*

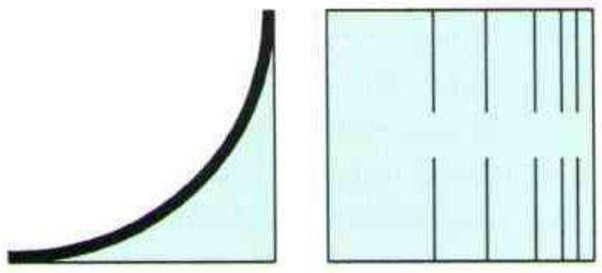

Pendiente cóncava. *Puede ver la cima de una pendiente cóncava desde su base, ya que presenta un declive suave en la parte inferior. En un mapa, sus curvas de nivel se juntan en la cima.*

REFERENCIAS EN CUADRÍCULA

Una cuadrícula le permite encontrar con precisión una posición en el mapa. Las líneas verticales se llaman *meridianos*, y las horizontales *paralelos*. Al dar una posición siempre se hace referencia primero a los meridianos. También pueden darse letras como referencia para una sección del mapa. Al determinar una referencia en la cuadrícula, siga la línea vertical más cercana a la posición en el pie del mapa para encontrar su coordenada —por ejemplo, 04. Estime el número de décimas desde la línea de cuadrícula hasta la posición —en este caso, 5. Repita la operación con la línea horizontal más cercana a la posición (410). Añada las letras de referencia, si es necesario.

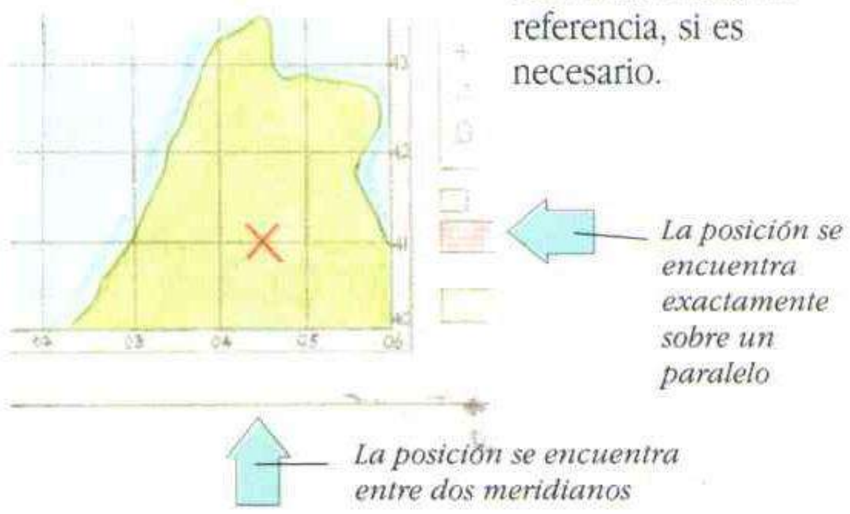

La posición exacta. *El método antes descrito proporciona una referencia de seis cifras sobre una cuadrícula (045410). Si su mapa tiene letras puede insertar las que sean relevantes para su posición anteponiéndolas a los números de referencia.*

PROYECCIONES

Un mapa intenta mostrar, con un grado elevado de exactitud, los detalles superficiales de la corteza terrestre, que es una esfera irregular. La proyección de la misma sobre una superficie cilíndrica o cónica crea muchas imprecisiones y distorsiones, que deben ajustarse al imprimir un mapa. Debido a este fenómeno, la mayoría de los mapas no son del todo precisos, particularmente en las latitudes mayores. La mejor proyección conocida es la de Mercator, empleada universalmente en el mar.

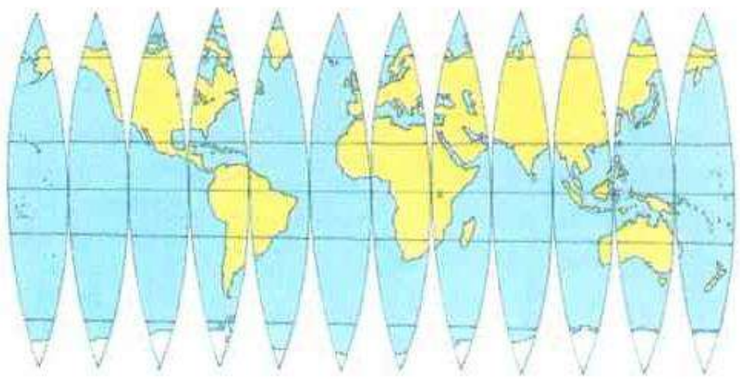

«Pelar» la Tierra. *Si «pelamos» la Tierra como una naranja, su piel quedará plana, pero habrá grandes huecos entre los segmentos en las latitudes más altas. Si se rellenan para dibujar un mapa plano, se introducen grandes distorsiones.*

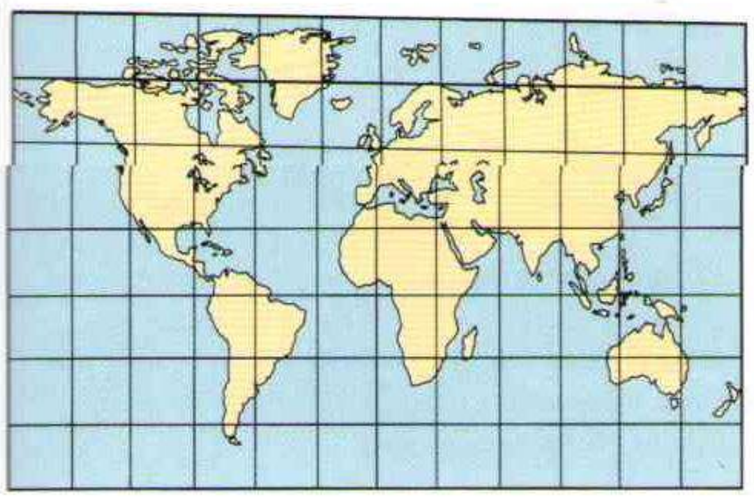

Proyección cilíndrica. *La proyección de Mercator muestra el mundo entero como si se hubiera proyectado sobre un cilindro de papel. Al desenrollar el cilindro, el mundo puede verse sobre una superficie plana. Aunque es rectangular por convención, y los meridianos y paralelos se cruzan en ángulos rectos, al igual que lo hacen en el globo, la Proyección de Mercator hace que los países de las latitudes polares aparezcan mucho mayores* (véase superior).

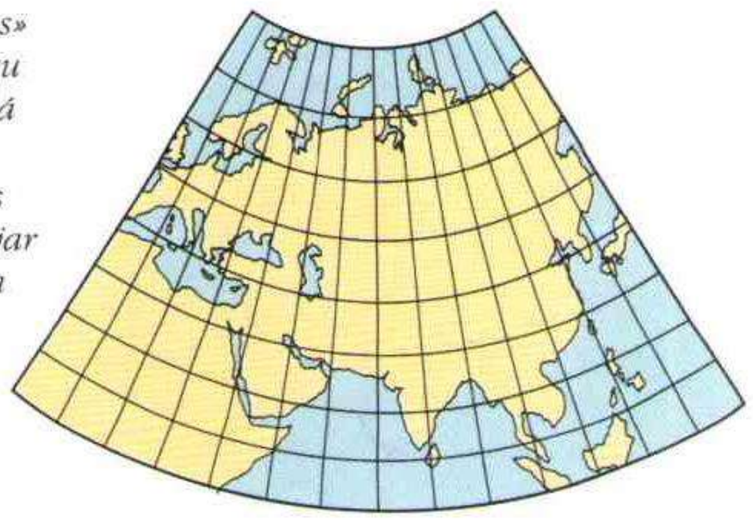

Proyección cónica. *Existen muchas clases de proyecciones, cada una con propiedades diferentes. El mapa de la ilustración se forma proyectando el mundo sobre un cono y desenrollándolo. La proyección cónica se utiliza a menudo en los atlas de mapas regionales. Aunque existen todavía distorsiones, particularmente en la base del cono, éstas no son tan importantes como en otras proyecciones.*

ENCONTRAR SU POSICIÓN

ES IMPOSIBLE ESTAR totalmente perdido. Incluso las víctimas de un accidente aéreo conocen el país o la zona en que han caído. Averiguar el lugar exacto en que se halla es un proceso que consiste en ir descartando opciones hasta poder determinar un punto en un mapa. El ser capaz de dar una referencia precisa sobre un mapa *(véase página 127)* por radio, le asegura un rescate inmediato. Si conoce su posición sobre un mapa particular, es difícil que esté perdido. Determinando las características del terreno y encontrando alguna referencia geográfica prominente y relacionándola con el mapa, el proceso de descartar opciones no le tomará mucho tiempo. Es posible que tenga que dibujar su propio mapa por varias razones: puede que no disponga de un mapa, o que el que tenga no sea preciso. Si dibuja su propio mapa, se asegurará de poder volver a encontrar su campamento, el agua y las fuentes de comida.

DETERMINACIÓN DE LA POSICIÓN

Para determinar su situación utilizando un mapa y una brújula, debe relacionar su posición con las características del terreno que pueda ver e identificar en el mapa. Un rumbo le dará la dirección desde alguna referencia hasta su punto de observación. Si toma estos rumbos desde una o más referencias, podrá delimitar su posición de forma bastante exacta.

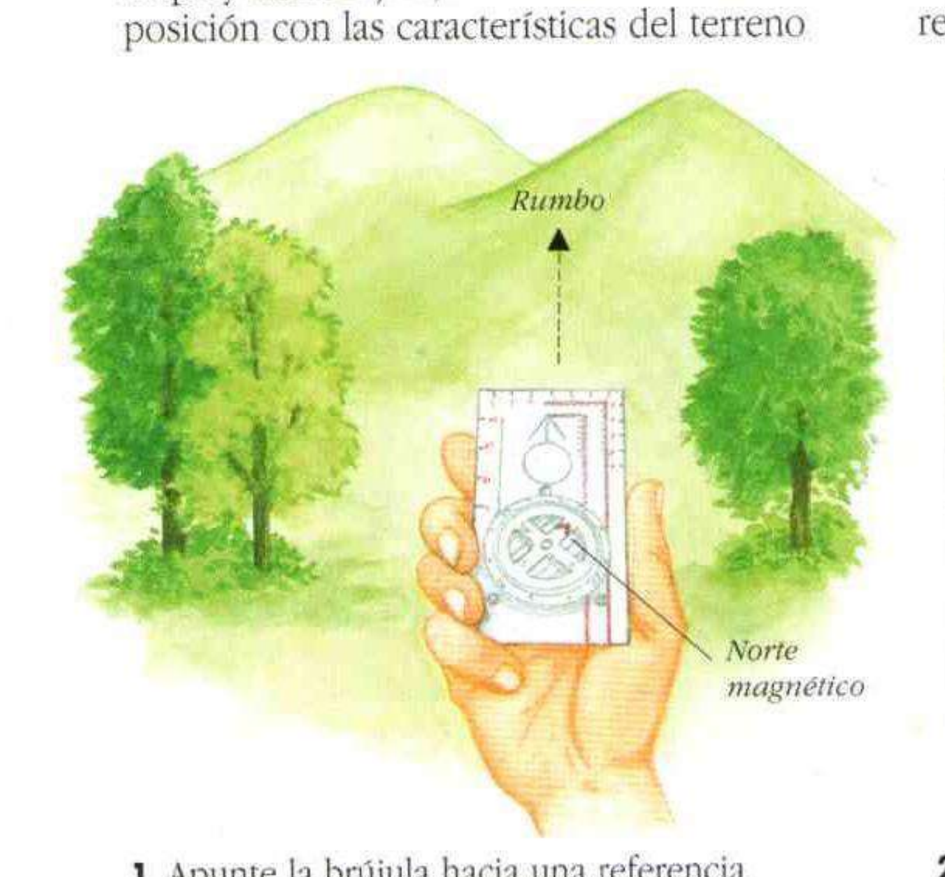

1 Apunte la brújula hacia una referencia destacable a su lado derecho, por ejemplo un paso entre dos montañas. Gire el dial (limbo) de su brújula hasta que la flecha norte del dial apunte en la misma dirección que la aguja magnética.

2 Coloque la brújula sobre el símbolo del collado dibujado en el mapa. Gire la brújula hasta que la flecha del dial apunte hacia el norte en el mapa. Dibuje una línea desde el collado a lo largo del rumbo (rumbo inverso).

3 Apunte la brújula hacia otra referencia que pueda ver con claridad. Esta segunda referencia debería estar situada en un ángulo de unos 90 grados de la primera, y por lo menos a 1 km de su posición actual. Lo ideal es la cima de una montaña bien definida.

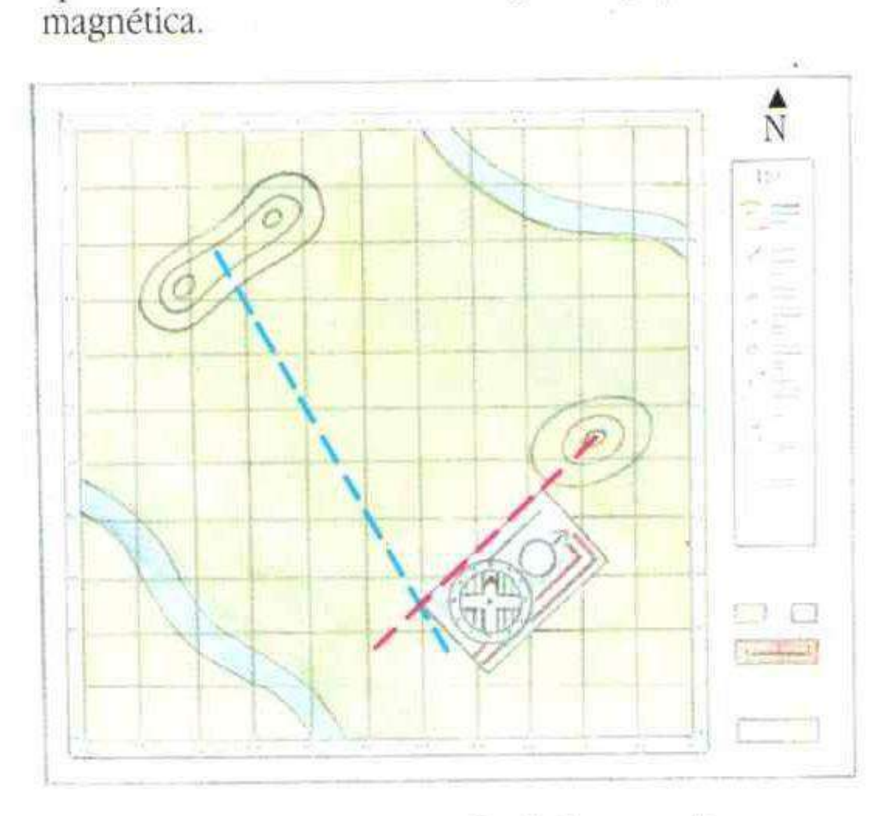

4 Trace un rumbo inverso desde la segunda referencia sobre el mapa, siguiendo la metodología antes indicada. Su posición se encuentra en la intersección de las líneas. Si así lo desea, puede tomar un tercer rumbo para obtener mayor precisión.

EL EMPLEO DE UN GPS

El GPS (siglas en inglés correspondientes a *Global Positioning System* = Sistema de Situación Global) ha revolucionado la navegación. Este sistema utiliza una constelación de 24 satélites, que transmiten señales de radio hacia la Tierra. Empleando un receptor, puede interceptar estas señales para determinar su posición de forma relativamente exacta. Los transmisores fijos calculan la posición de los satélites en órbita. La navegación en un desierto sin apenas referencias geográficas, o en el mar, se hacía antes a base de determinar rumbos con la brújula y una confirmación de la posición con un sextante y un cronómetro. Los métodos tradicionales son un refuerzo al GPS.

Enlace por satélite. *Las terminales en tierra emiten y reciben señales de los satélites en órbita. Estas señales pueden ser recogidas por unos receptores especiales que son de gran ayuda en la navegación.*

CÓMO DIBUJAR SU PROPIO MAPA

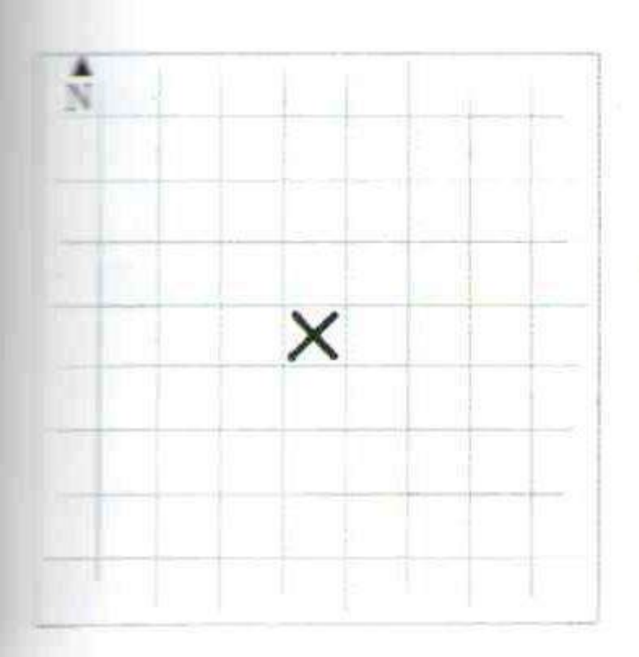

1 Dibuje una cuadrícula, cada cuadrado representará 1 km². Trace una X en el centro que representará su posición actual.

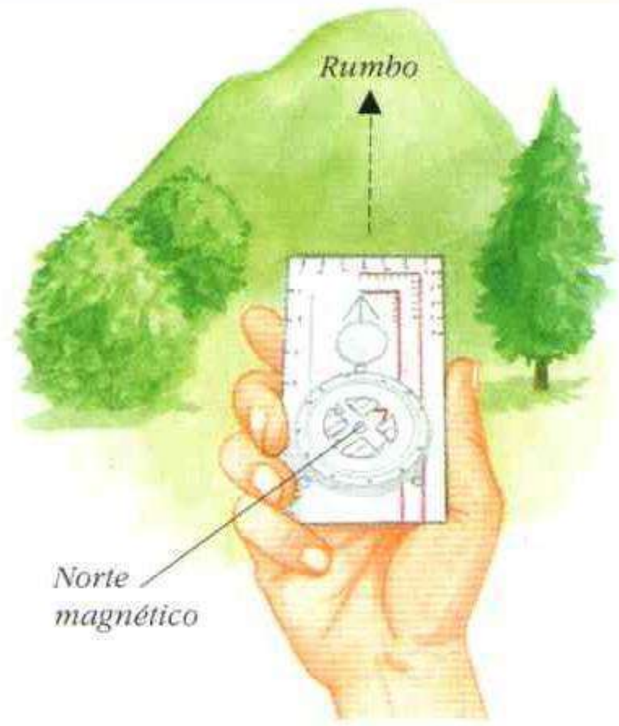

2 Desplácese hasta una posición ventajosa. Tome un rumbo hacia una referencia del terreno y estime la distancia a la misma *(véase inferior)*.

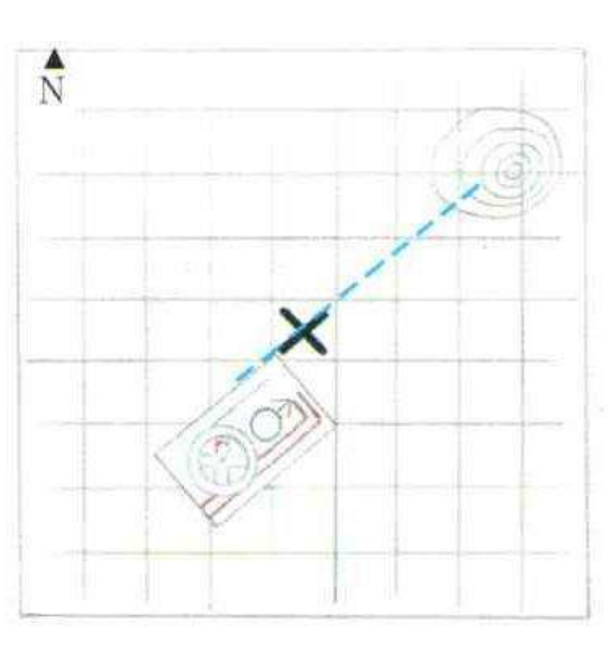

3 Señale el rumbo en su mapa, alineando la cuadrícula en dirección norte-sur. Dibuje la montaña a la distancia estimada.

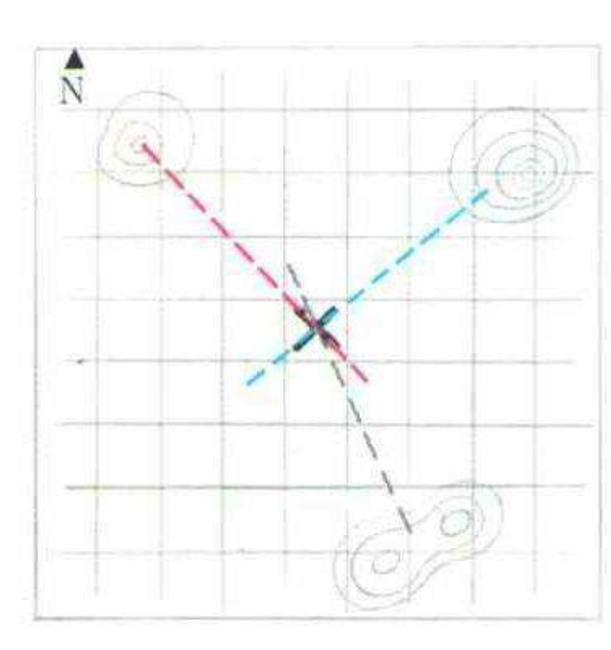

4 Tome rumbos a otras dos referencias del terreno y dibújelas en su mapa. El punto en el cual se cruzan es su posición actual.

LA MEDICIÓN DE LA DISTANCIA

Para dibujar una referencia del terreno sobre su mapa, debe, en primer lugar, determinar la distancia a la que se encuentra. Puede emplear rumbos inversos para hacerlo, pero para ello debe moverse de «X» a otra posición, que puede marcarse con toda precisión en el mapa. El rumbo de su nueva posición a la referencia le dará una intersección con el primer rumbo –que podrá marcarse entonces en el mapa. Su distancia puede calcularse con bastante precisión.

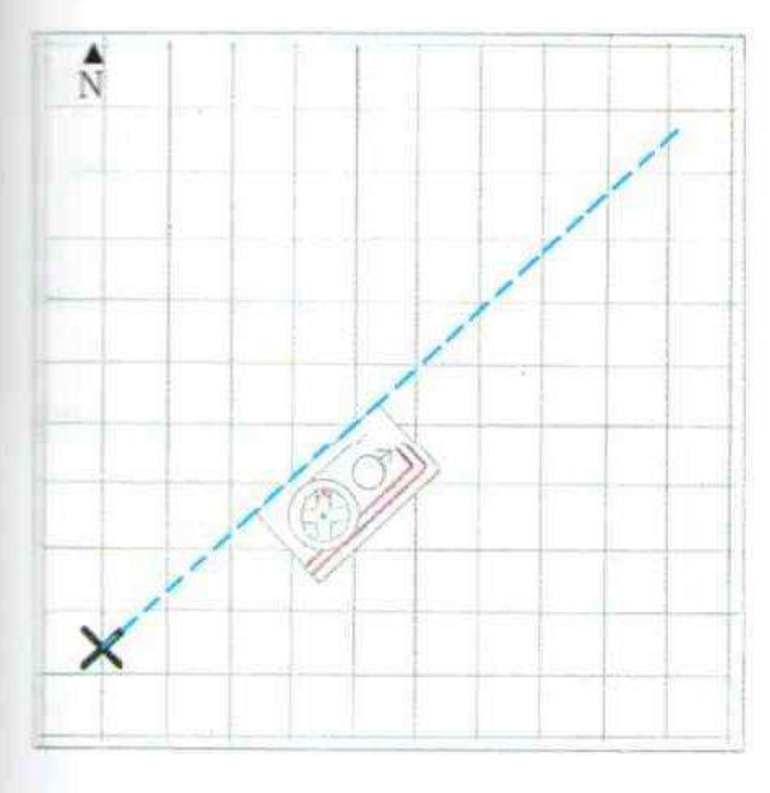

1 Puede medir la distancia de su posición a una referencia, tanto sobre un mapa impreso como en uno confeccionado por usted. Si ha dibujado una cuadrícula, los cuadrados deben ser exactos. En primer lugar, tome un rumbo a una referencia, por ejemplo una montaña. Dibuje el rumbo como una línea en la cuadrícula desde la cruz (su posición actual) en dirección a la referencia.

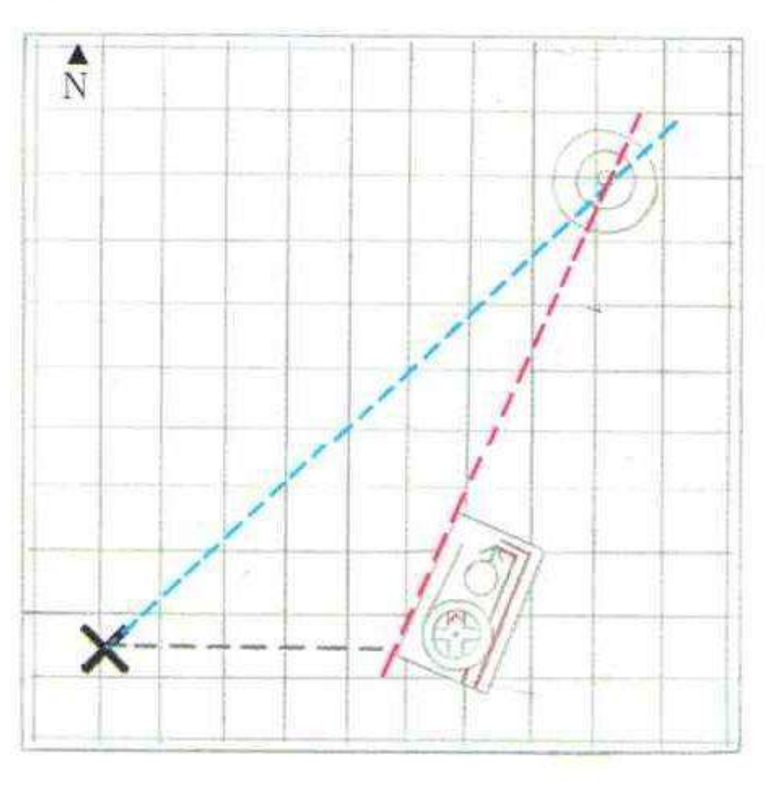

2 Camine con un rumbo de 90 grados con respecto al norte durante una distancia conocida *(véase página 131)* hasta que su siguiente rumbo a la montaña difiera en, al menos, 30 grados con respecto al primero. Señale en el gráfico su nueva posición, así como el nuevo rumbo. El punto en el que ambos se intersectan es la situación de la montaña. Puede determinar la distancia hasta la montaña contando las cuadrículas.

VARIACIÓN MAGNÉTICA

La orientación puede resultar confusa al existir tres nortes ligeramente diferentes. En la mayor parte del globo terráqueo, esta variación magnética es lo suficientemente pequeña como para que la mayoría de los caminantes no tengan que preocuparse por ella. Sin embargo, en otras zonas, como en las altas latitudes, la variación magnética se incrementa de forma significativa, llegando incluso a resultar inútil el uso de una brújula. Es, pues, vital conocer la variación magnética del área en que se navega. Ésta debe añadirse o sustraerse del rumbo, dependiendo de su localización. La variación magnética del área en cuestión se indica al pie de la mayoría de los mapas, utilizando tres flechas.

Tres nortes. *El norte magnético es aquél al que apunta la brújula; el norte geométrico es el norte marcado en el mapa; y el norte verdadero es el polo geográfico astral real. La variación magnética es la diferencia entre el norte geométrico y el norte magnético.*

ALGUNOS CONSEJOS PARA NO PERDERSE

- La regla de oro es: «Confíe en su brújula». Aun así, muchas personas se pierden al confiar antes en su propio sentido de la orientación que en su brújula.
- Mida siempre el rumbo y las distancias.
- Asegúrese de comprender cómo se indican los distintos tipos de terreno en los mapas, localizando los valles, las montañas y las crestas.
- Contraste el mapa utilizando la brújula *(véase página 130)*, busque con la vista las referencias geográficas más relevantes, e identifique su posición sobre el terreno antes de tomar un rumbo.

PRINCIPIOS DE NAVEGACIÓN

LA NAVEGACIÓN EN tierra consiste en una combinación de trabajos sobre el mapa y con la brújula, con el propósito de desplazarse por un terreno determinado de la forma más segura y sencilla. De ninguna manera es tan precisa como la navegación en el mar, donde las únicas variaciones son las mareas y el viento. La velocidad de desplazamiento se ve sobremanera afectada por el tipo de terreno, las rocas y el suelo, el clima, el estado del tiempo y la vegetación.

La navegación en tierra requiere un seguimiento constante de la ruta, no sólo para evitar los peligros, sino también para no perderse.

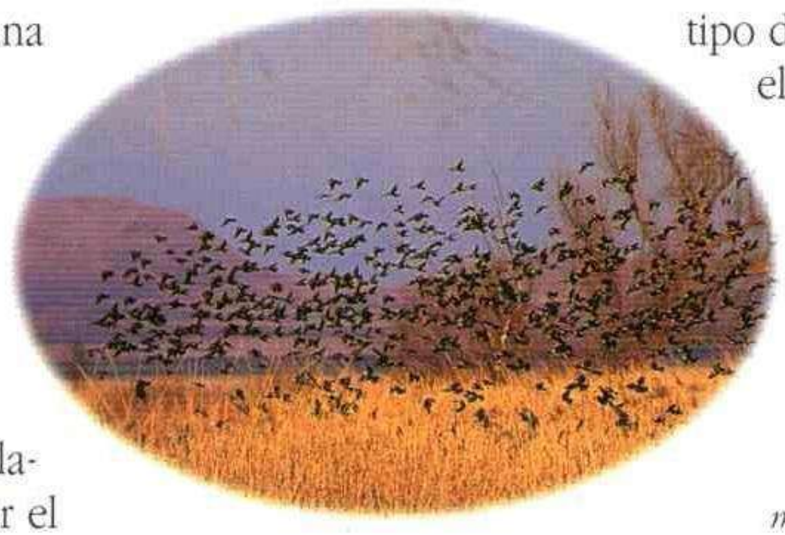

La migración. *Muchas aves han desarrollado una capacidad de navegación tal, que les permite regresar desde el otro lado del Globo a los mismos lugares de cría, un año tras otro.*

CONTRASTAR SU MAPA

Antes de emprender un recorrido, debe contrastar su mapa con la brújula, determinando dónde se encuentra su posición actual y su destino en el mapa. Además de trazar el rumbo, debe asimismo saber a qué distancia se encuentra del lugar de destino. Puede determinar dicha distancia utilizando la escala del lateral de su brújula, o bien la escala de la parte inferior de su mapa. Mientras mide esta distancia, compruebe en el mapa el tipo de terreno que espera cruzar.

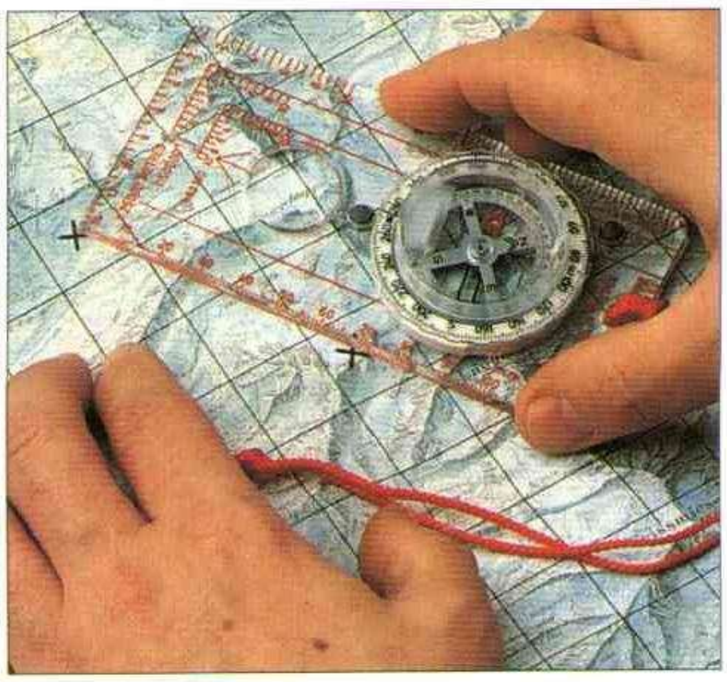

1 Para determinar el rumbo del punto A al punto B, coloque la brújula sobre el mapa entre A y B, y apunte la flecha de dirección del extremo de la brújula hacia la dirección deseada. Lea la distancia entre A y B en la escala del lateral de la brújula y compárela con la escala del mapa.

2 Sin mover la brújula, gire el dial (limbo) central hasta que las líneas paralelas norte-sur estén alineadas con las líneas de la cuadrícula del mapa. La flecha norte (roja) del dial (limbo) debería apuntar hacia el norte geométrico *(véase página 129)*. De esta manera, establecerá el rumbo (el ángulo entre la línea A-B y el norte magnético) con la brújula.

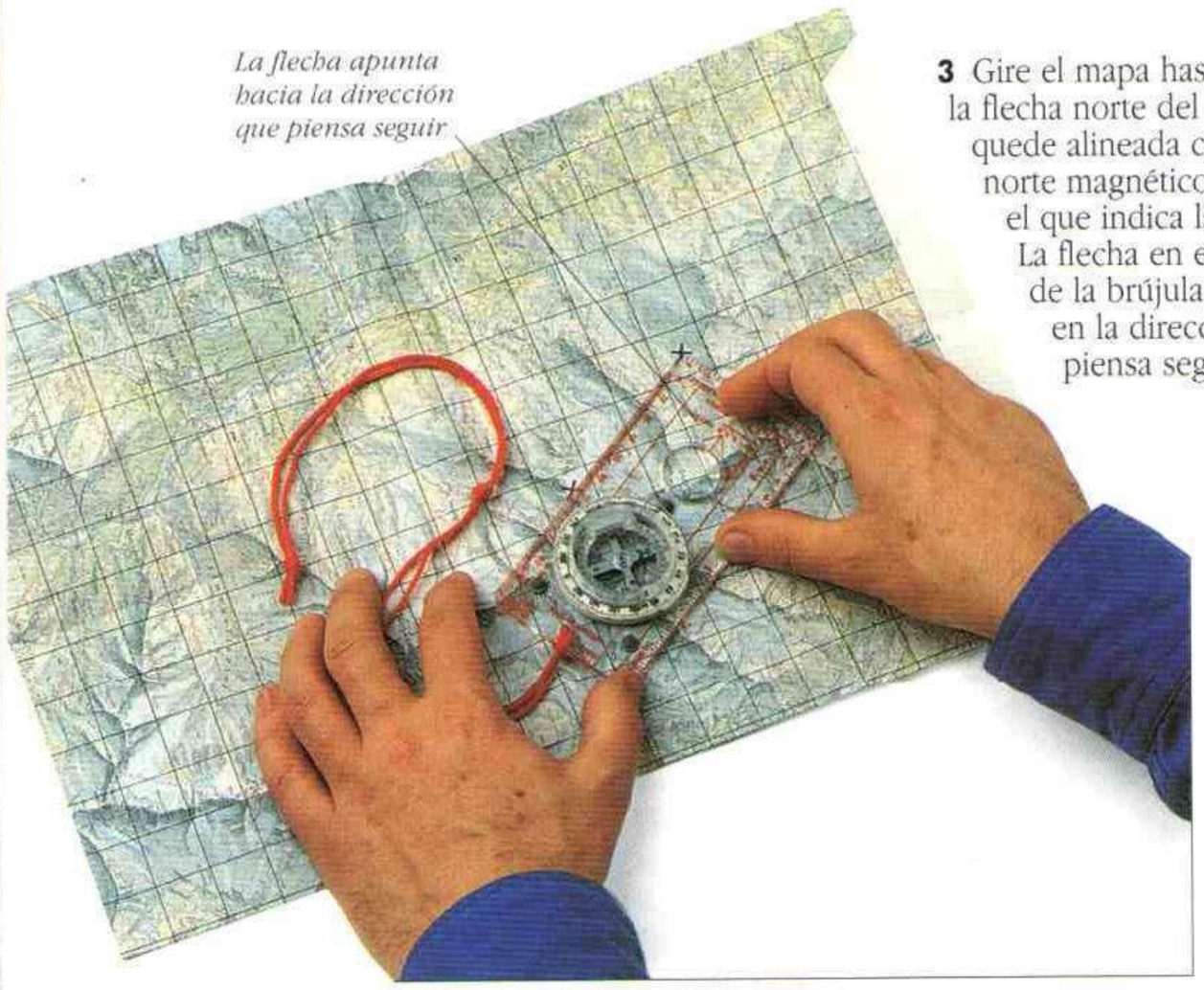

3 Gire el mapa hasta que la flecha norte del dial quede alineada con el norte magnético, que es el que indica la aguja. La flecha en el extremo de la brújula apuntará en la dirección que piensa seguir.

4 Ahora puede sujetar la brújula con la mano y seguir la flecha de dirección. Asegúrese de mantener alineadas la flecha norte del dial y la aguja magnética. Al seguir un rumbo, mantenga la brújula en posición horizontal.

TÉCNICAS DE NAVEGACIÓN

En tierra no siempre es posible o recomendable caminar en línea recta siguiendo un rumbo. Los siguientes «trucos del oficio» le mostrarán cómo determinar el rumbo a seguir sorteando los terrenos difíciles y siguiendo la ruta más sencilla sin perder tiempo para comprobar la situación, y sin perderse. Es importante recordar que siempre debe haber establecido un rumbo en su brújula *(véase página anterior)*, aunque sea únicamente para emplearlo como referencia.

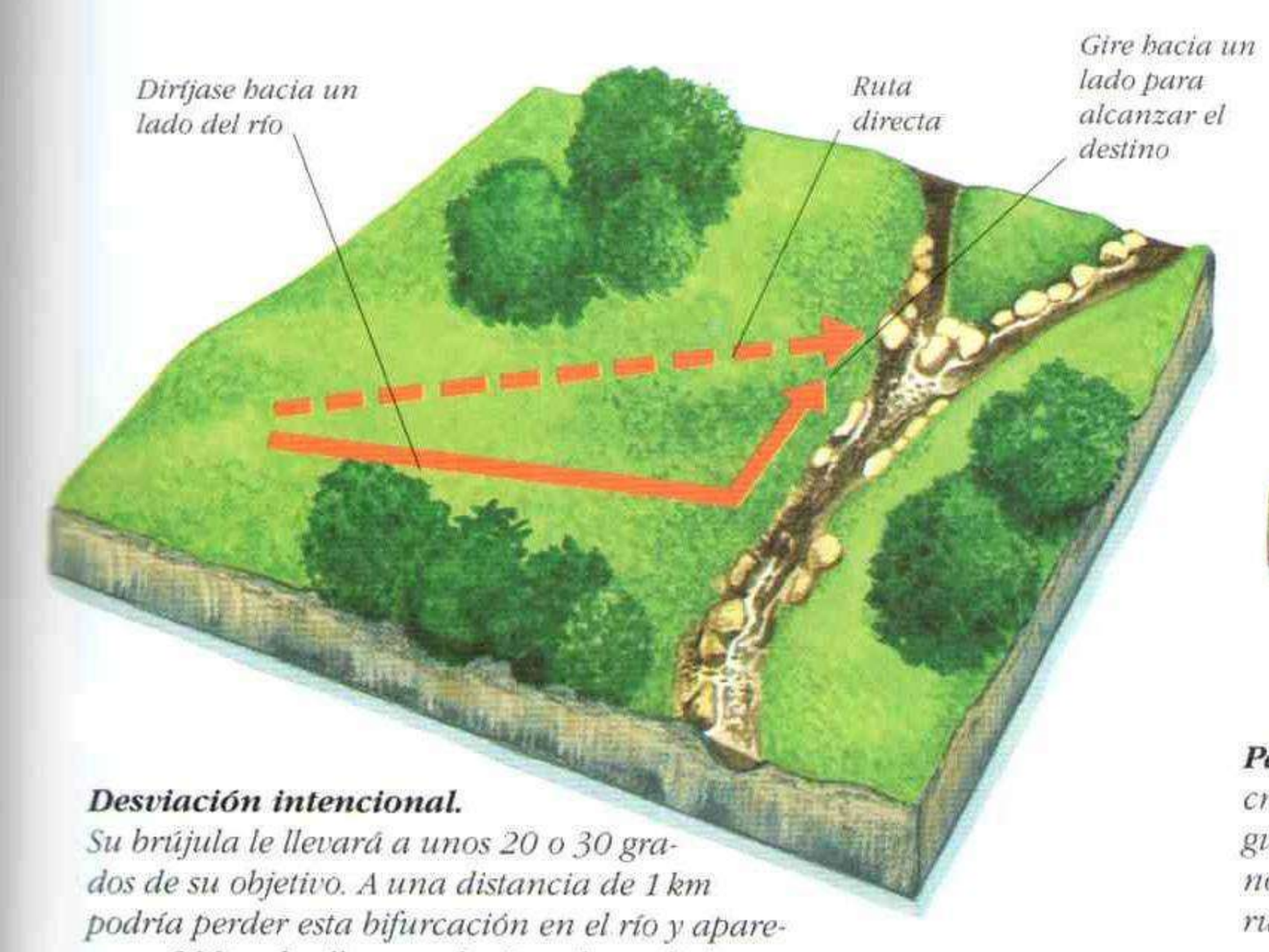

Desviación intencional. *Su brújula le llevará a unos 20 o 30 grados de su objetivo. A una distancia de 1 km podría perder esta bifurcación en el río y aparecer a 300 m de ella a cualquiera de sus lados. Más aún, no sabría en qué dirección buscarla. Para evitarlo, desvíese intencionalmente hacia uno de los lados. Cuando alcance el río, sabrá si girar hacia la derecha o a la izquierda de la bifurcación.*

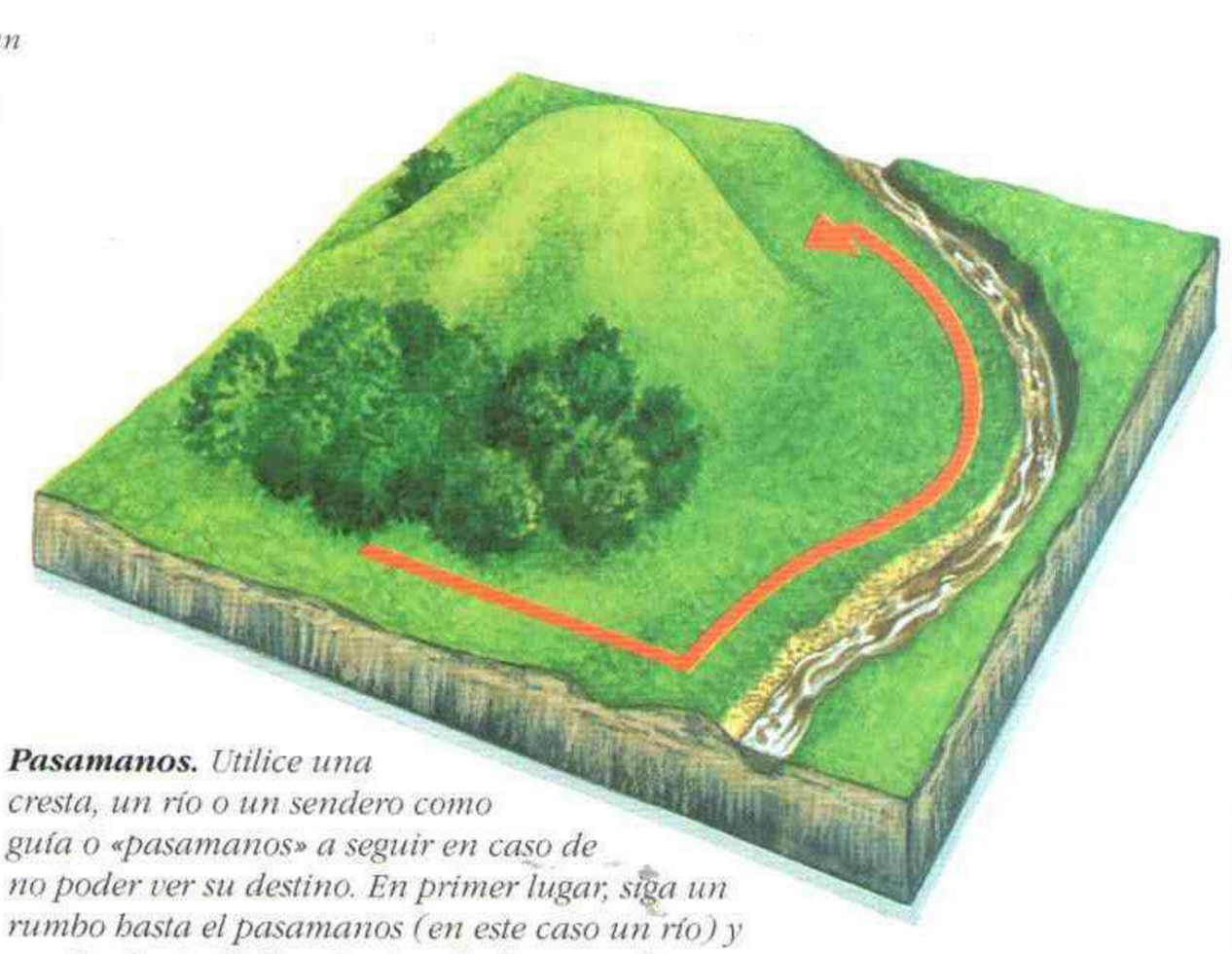

Pasamanos. *Utilice una cresta, un río o un sendero como guía o «pasamanos» a seguir en caso de no poder ver su destino. En primer lugar, siga un rumbo hasta el pasamanos (en este caso un río) y camine hasta él. En este ejemplo, la montaña se encuentra a unos 250 grados del punto más cercano al río. Siga el río hasta que aparezca la montaña (a 250 grados) y gire entonces a la izquierda para seguir ese rumbo.*

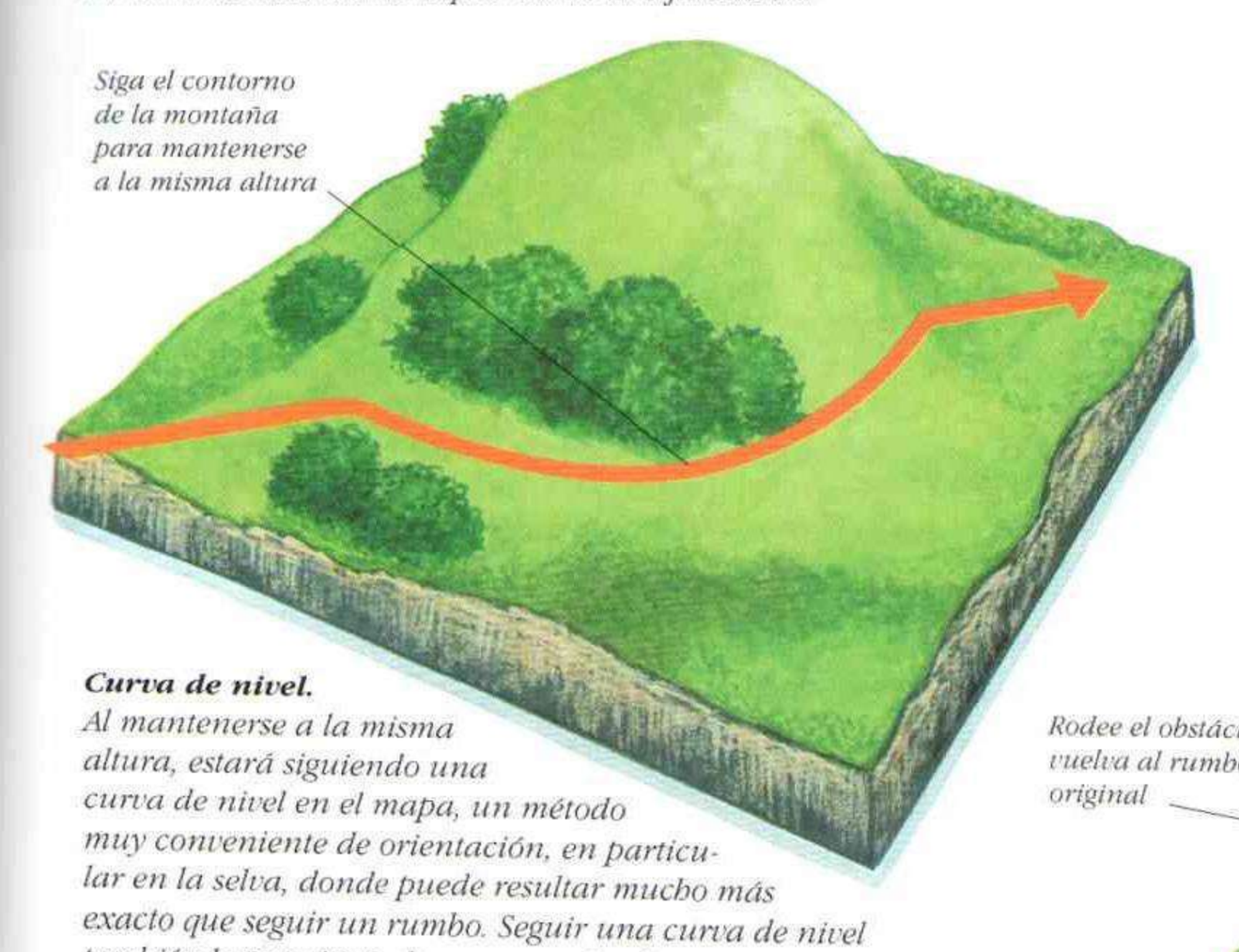

Curva de nivel. *Al mantenerse a la misma altura, estará siguiendo una curva de nivel en el mapa, un método muy conveniente de orientación, en particular en la selva, donde puede resultar mucho más exacto que seguir un rumbo. Seguir una curva de nivel también le permitirá ahorrar energía al no perder ni ganar altura.*

LA ESTIMACIÓN DE LA DISTANCIA RECORRIDA

Al moverse a pie, puede contar sus pasos (los que da con el pie izquierdo) o utilizar el reloj. La marcha varía de acuerdo con el terreno y el ritmo de viaje. El tiempo es una forma más adecuada de medir la distancia recorrida. Deténgase después de los primeros 10 minutos para determinar en el mapa la distancia que ha recorrido, y entonces, una vez que ha determinado su velocidad, puede detenerse cada hora para comprobar el mapa y el equipo. Aprenderá a determinar la velocidad de su marcha con una exactitud sorprendente. En un terreno escabroso, y con un equipo pesado, una velocidad de 4 km/h es muy buena.

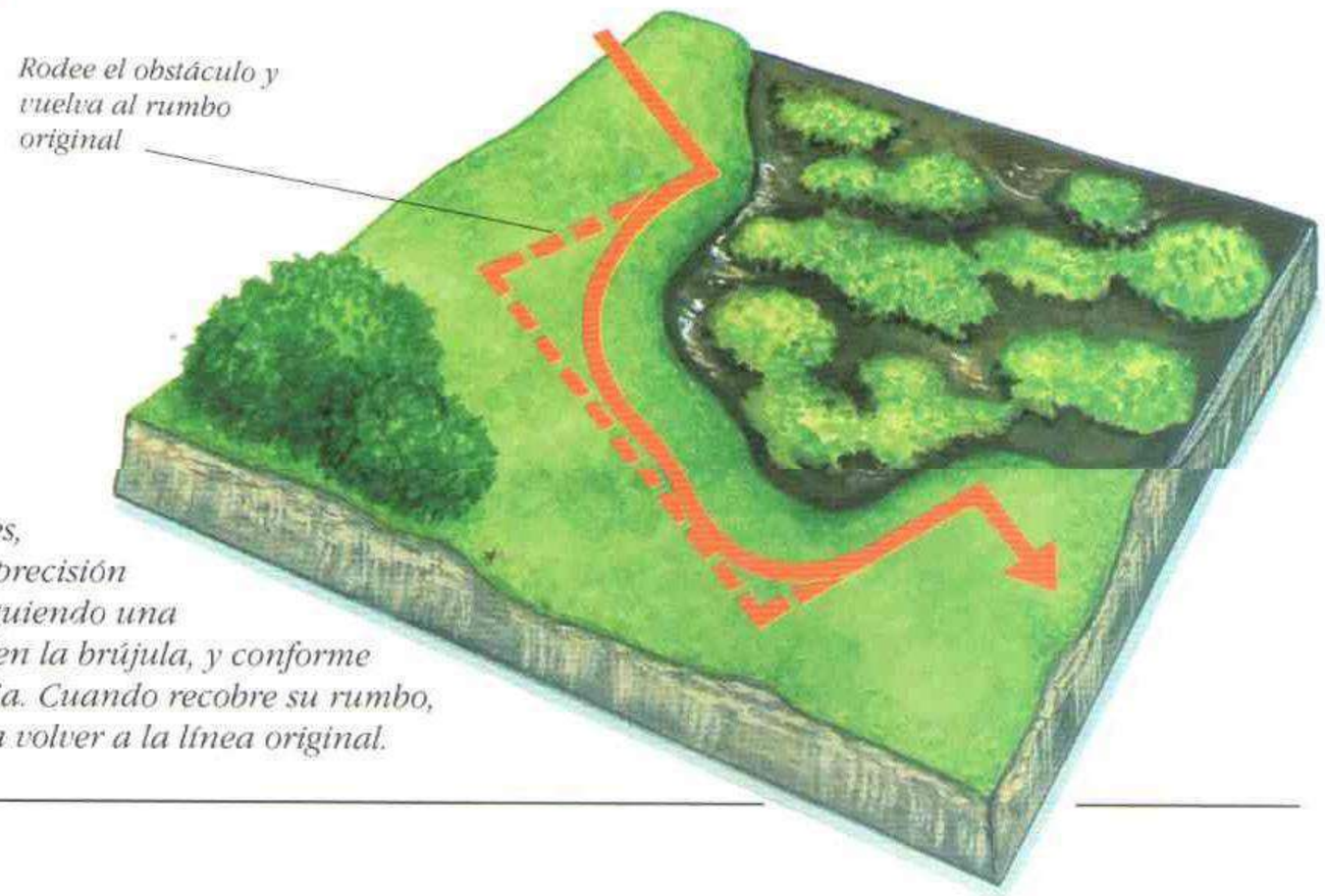

Rodeo / Desviación. *Los obstáculos importantes, como una ciénaga, no siempre se marcan con precisión en los mapas. Al rodearlos, al igual que hace siguiendo una curva de nivel, siga la línea recta de su rumbo en la brújula, y conforme va rodeando el obstáculo, determine la distancia. Cuando recobre su rumbo, sabrá la distancia que tendrá que recorrer para volver a la línea original.*

LA EVALUACIÓN DEL TIEMPO

EL TIEMPO ES el día-a-día del clima. Los cambios locales son causados por las diferencias en la presión atmosférica y la temperatura. Al pasar por encima de los diferentes tipos de terreno, los movimientos de aire localizados producen vientos y lluvias. Las áreas costeras tienen, por lo general, vientos que soplan hacia el interior durante el día y hacia la costa durante la noche. En algunos lugares, como en el interior de los continentes, el tiempo es relativamente estable, mientras que en otros el clima es muy variable. Los climas realmente ecuatoriales no tienen distinción entre estaciones, sino una rutina diaria de lluvia y sol.

El tiempo en el mundo. *Un tipo de tiempo o clima puede localizarse en un área específica, o bien, puede desplazarse alrededor del mundo, afectando a zonas muy alejadas entre sí —como es el caso de los huracanes.*

LEER LAS NUBES

Las nubes son grandes masas de vapor de agua en condensación. Su presencia, tipo y tamaño indican la temperatura y la presión de las masas de aire, permitiendo la predicción del tiempo. Cuanto más altas sean las nubes, mejor y más estable será el tiempo. Las nubes tormentosas son generalmente oscuras, bajas y agrupadas en grandes masas, mientras que las nubes que presagian buen tiempo son altas y blancas. Las nubes muy bajas pueden envolver las tierras altas en la niebla.

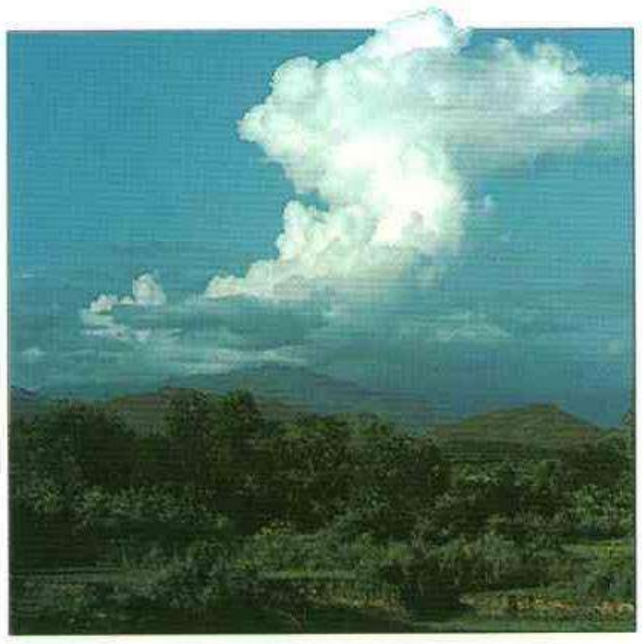

Cumulonimbos. *Los cumulonimbos traen consigo granizo, vientos y tormentas. Las formaciones de estas nubes oscuras y densas pueden alcanzar un gran tamaño y un desarrollo vertical considerable, tomando formas extrañas. En ocasiones, anuncian tormenta eléctrica* (véase página 171).

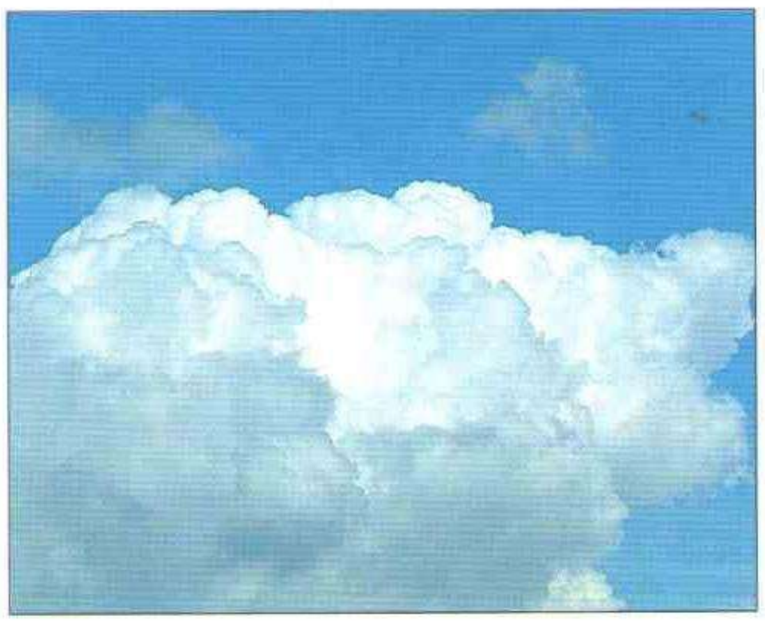

Cúmulos. *Los cúmulos de aspecto algodonoso indican buen tiempo, pero si las nubes están muy juntas y se tornan de un color oscuro, indican lluvia. Sobre mar abierto, pueden indicar la proximidad de tierra* (véase página 165).

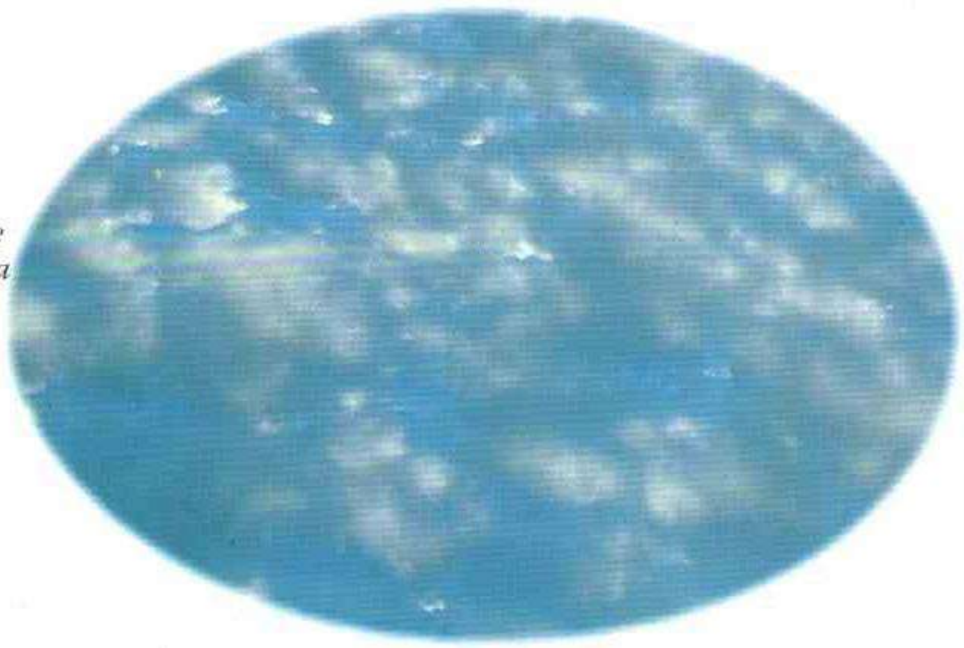

Estratocúmulos. *Los estratocúmulos se forman encima de los cúmulos, extendiéndose para formar una gruesa capa. Los estratocúmulos pueden producir lluvias ligeras que se dispersan al atardecer.*

Cirros. *Los cirros, también conocidos como penachos, son nubes muy blancas en forma de filamentos. Tienden a formarse en las grandes alturas si el tiempo es bueno. Como la atmósfera es tan fría a esas alturas, estas pequeñas nubes están formadas totalmente por cristales de hielo.*

CÓMO AFECTA EL PAISAJE AL TIEMPO

Cuando las cordilleras o las montañas obstaculizan el paso de los vientos cargados de humedad provenientes del mar, les obligan a elevarse, formando nubes y perdiendo la humedad en forma de lluvia. Esta lluvia cae por la ladera de la montaña hasta el mar, haciendo que la franja costera presente una intensa precipitación pluvial. En el otro lado de la montaña, los vientos, ahora secos, descienden hacia el interior. Al no contener lluvia, se crea una «sombra de lluvia» que deja el interior seco (como sucede en la costa oriental de Australia).

Forzar la lluvia. *Las montañas obligan a los vientos húmedos a ascender y a perder su humedad en forma de lluvia.*

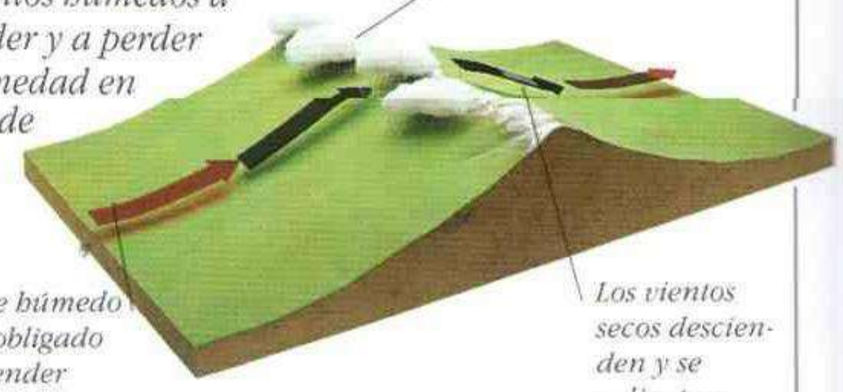

La predicción natural del tiempo

Muchos de los llamados *cuentos de viejas* están basados en la observación. Por ejemplo, la antiquísima rima inglesa «Cielo rojo al atardecer —delicia del pastor; cielo rojo al amanecer —alarma del pastor» puede resultar ser una predicción precisa. El comportamiento animal también puede indicar cambios en el tiempo, al igual que algunos fenómenos naturales (arco iris).

Arco iris. *Un arco iris se origina cuando la luz del sol se refracta y refleja en las pequeñas gotas de vapor de agua en el aire después de la lluvia. Las gotas actúan como prisma, dividiendo la luz del sol en los diferentes colores del espectro. La presencia de un arco iris indica que el tiempo será bueno, particularmente si aparece durante la tarde.*

Animales en migración. *Los animales detectan los cambios de la presión atmosférica que preceden cualquier cambio en el tiempo. Cuando se aproxima una tormenta o nevada, muchos animales gregarios que viven en las grandes alturas se reúnen y bajan a las tierras más bajas y seguras.*

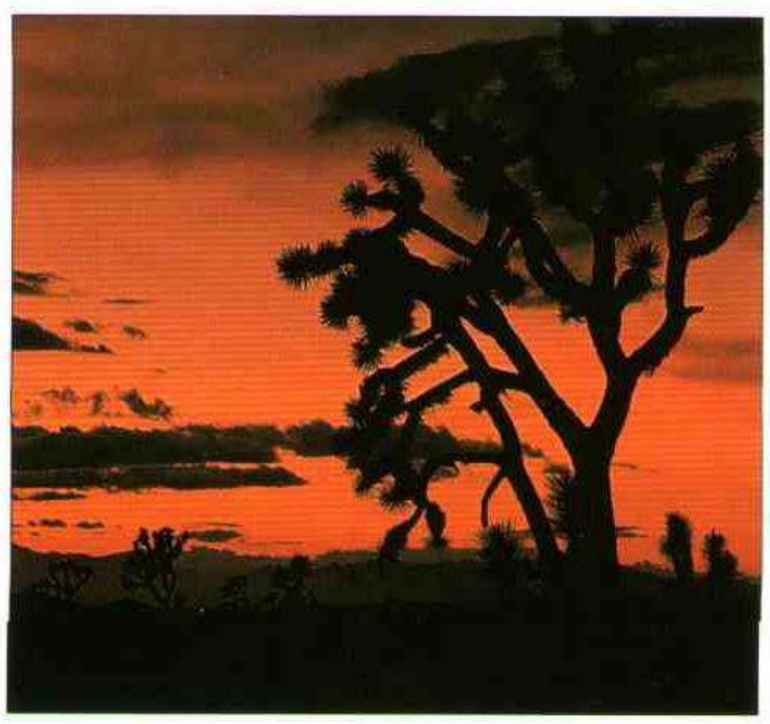

Amanecer rojizo. *Un cielo rojo al amanecer a menudo indica que hay una gran cantidad de humedad en el aire, ya que el sol se refleja en las nubes. Esto significa, generalmente, que se aproxima una tormenta, como bien dice una antigua rima* (véase superior).

Los distintos signos de la naturaleza

Un cambio en el tiempo puede ser indicado por un cambio en la dirección o en la fuerza del viento, o por cambios en la formación de las nubes. Los vientos predominantes traen, por lo general, un tiempo particular en cada ocasión. Un viento seco y continuado que cambia de dirección o amaina suele preceder a la lluvia. El rocío matinal o la niebla indican un tiempo estable, pero el viento —especialmente junto con niebla en las montañas bajas— puede traer lluvia. Un cielo claro al anochecer indica que la noche será fría y que probablemente helará, ya que no hay nubes que retengan el calor. Antes de llover, el aumento de la presión atmosférica incrementa los dolores reumáticos, hace que las plantas abran sus poros y produce la dilatación de los objetos de madera. Asimismo, los sonidos parecen viajar más lejos.

Piñas de los pinos. *Las piñas de los pinos también reaccionan ante la humedad. Si el aire es seco, las glumas de la piña se encogen y ésta se abre. Poco antes de la lluvia, las glumas absorben la humedad y la piña recobra su forma.*

Tiempo húmedo

Tiempo seco

ESCALA DE BEAUFORT PARA MEDIR LA VELOCIDAD DEL VIENTO

Fuerza	Descripción	Velocidad (km / h; m / s)	Efectos
0	Calma	0-1; menos de 0,3	El humo sube verticalmente
1	Ventolina	1-6; 0,3-1,5	El humo, aunque no el viento, indican la dirección
2	Viento flojito o brisa	6-12; 1,6-3,3	El viento se nota en la cara; las hojas revolotean
3	Viento flojo	12-20; 3,4-5,4	Hojas en continuo movimiento; el viento hace ondear las banderas
4	Viento bonancible	20-29; 5,5-7,9	Se levantan el polvo y los papeles sueltos; se mueven las ramas pequeñas
5	Viento fresquito	29-39; 8,0-10,7	Se inclinan los árboles pequeños con hojas
6	Viento fresco	39-50; 10,8-13,8	Las ramas grandes se mueven; los paraguas se usan con dificultad
7	Viento frescachón	50-62; 13,9-17,1	Los árboles se mueven; dificultad para caminar
8	Viento duro	62-75; 17,2-20,7	Se rompen las ramas de los árboles; la circulación es prácticamente imposible
9	Viento muy duro	75-92; 20,8-25,4	Ligeros daños en las construcciones; se levantan las tejas de las chimeneas
10	Temporal	92-102; 25,5-28,4	Árboles desarraigados; daños estructurales considerables
11	Borrasca	102-117; 28,5-32,6	Daños generalizados
12	Huracán	Más de 117; más de 32,6	Zona devastada

Escala de Beaufort para medir la velocidad del viento. *Los cambios de tiempo están anunciados por un cambio en la velocidad del viento. Utilizando la escala de Beaufort, puede evaluarse la velocidad del mismo. En esta escala, la velocidad del viento se divide en 12 fuerzas, desde la fuerza 0 o calma hasta la fuerza 12 o huracán.*

PREPARARSE PARA EL VIAJE

UNA VEZ SE ha decidido a viajar, es fundamental prepararlo todo con antelación. Lave, repare y guarde su equipo. Evalúe el terreno que planea cruzar *(véase página 122)*, y planifique su ruta con la mayor exactitud posible. Reúna reservas de agua y de alimento, y lleve consigo todo aquello que pudiera serle útil en el camino o en su próximo campamento, como los materiales para el refugio. En una situación de supervivencia, si se ve obligado a cambiar de campamento, asegúrese de que tanto usted como sus compañeros tienen la condición física necesaria para realizar el viaje. Puede ser necesaria la construcción de algún equipo antes de viajar *(véase página siguiente)*.

EL USO DE LA MOCHILA

Su mochila debe estar preparada de tal manera que pueda acceder rápidamente a la comida y a la ropa durante la marcha, además de llevar el resto de su equipo de campamento con toda comodidad. No ate ningún artículo a la parte exterior; podría dañarse o perderse. Coloque todo aquello a lo que deba acceder rápidamente en los bolsillos exteriores. En el interior, dentro de una bolsa grande, resistente e impermeable, debe guardar los artículos más ligeros y voluminosos en el fondo, y los más pesados encima, para mantener el centro de gravedad de la mochila lo más alto posible. Dentro de su mochila debe guardar la ropa, la comida, los artículos de limpieza, un botiquín de primeros auxilios y otros artículos, en bolsas separadas e impermeables. Mientras permanezca en el campamento, mantenga todo lo que no necesite en la mochila para reducir al mínimo el desorden, y para que, en caso de necesidad, pueda abandonar el campamento rápidamente.

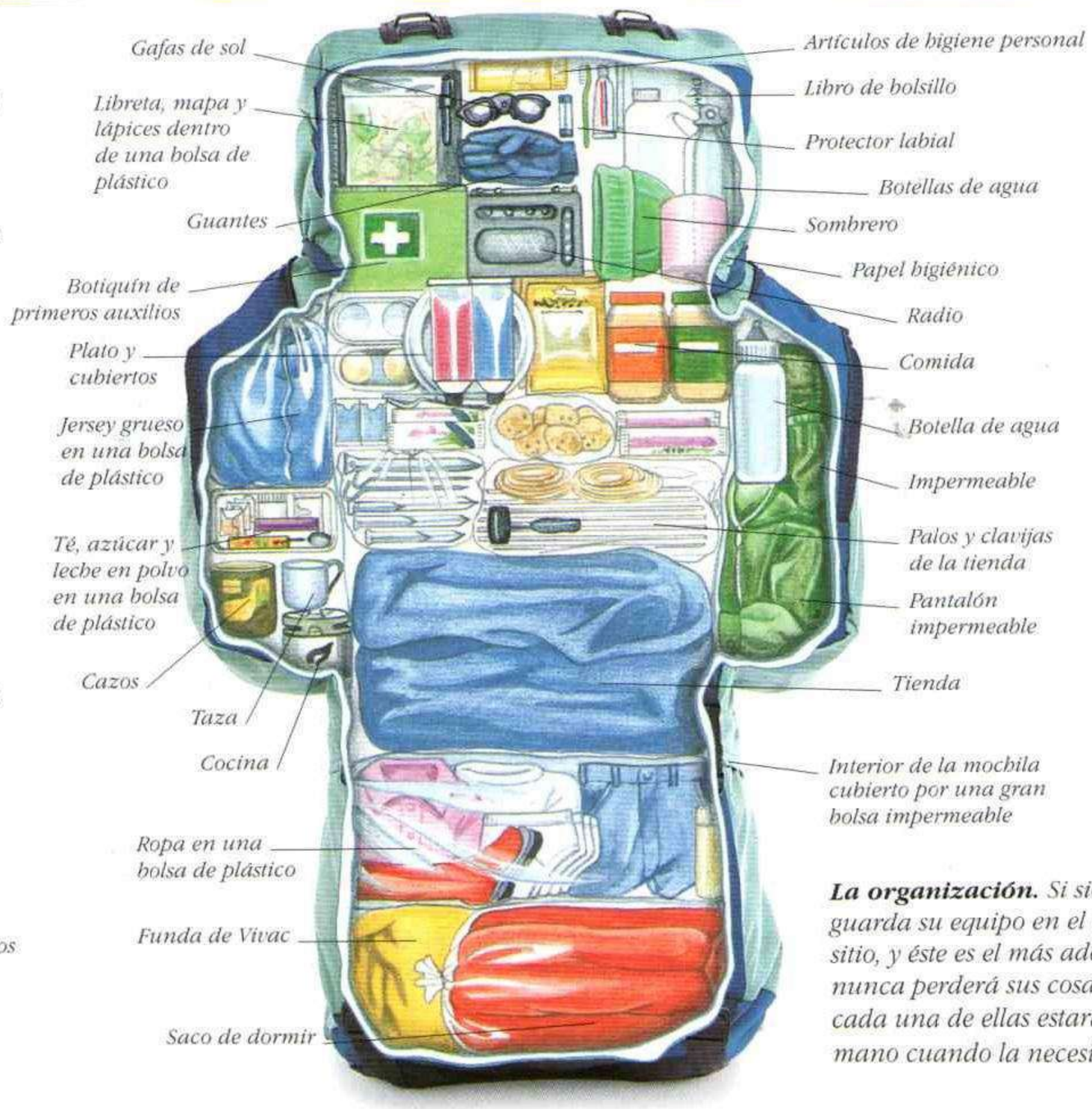

La organización. *Si siempre guarda su equipo en el mismo sitio, y éste es el más adecuado, nunca perderá sus cosas, y cada una de ellas estará a mano cuando la necesite.*

Llevar una mochila. *Mantenga el centro de gravedad de la mochila lo más alto posible, para evitar que tire de usted hacia atrás. Sus piernas deben realizar el trabajo de cargar la mochila. Un cinturón acolchado para la cadera permite que una parte del peso sea transferida de los hombros a la pelvis, liberando una parte del peso de la columna y de los hombros. Las correas reparten una parte del peso al pecho. Al ser ajustables, permiten que las mochilas puedan llevarse en la parte superior de la espalda.*

BOLSAS Y MOCHILAS

Para un viaje corto, es mucho más cómodo llevar una bolsa pequeña que contenga alimento, ropa y provisiones, que toda la mochila. Los artículos de higiene personal pueden llevarse en un pequeño neceser, mientras que los documentos de viaje pueden guardarse en portadocumentos impermeables.

MOCHILA

NECESER

PORTADOCUMENTOS

CONSTRUCCIÓN DE UN ARMAZÓN PARA LA MOCHILA

1 Corte una rama ligera unos 30 cm por debajo de una bifurcación, dejando cerca de 1 m de horquilla. Elimine cualquier posible nudo.

2 Talle muescas en los tres extremos de la rama. Ate una cuerda o un cordel doble alrededor de las muescas, que servirá como correa para los hombros.

3 Ate los otros cabos de las cuerdas alrededor de la muesca en la rama central. Asegúrese de que las correas se ajustan bien a sus hombros, pero recuerde que no deben apretar.

4 Envuelva su equipo en una manta o lona impermeable y átela lo más fuerte posible al armazón por el lado opuesto a las correas.

5 No sobrecargue su mochila improvisada —aunque el armazón soportará bien un peso considerable. Ate las dos correas para los hombros por delante para evitar que las cuerdas finas causen rozaduras en sus hombros. Puede acolchar las correas con vegetación.

CONSTRUCCIÓN DE UN REMOLQUE O *TRAVOIS*

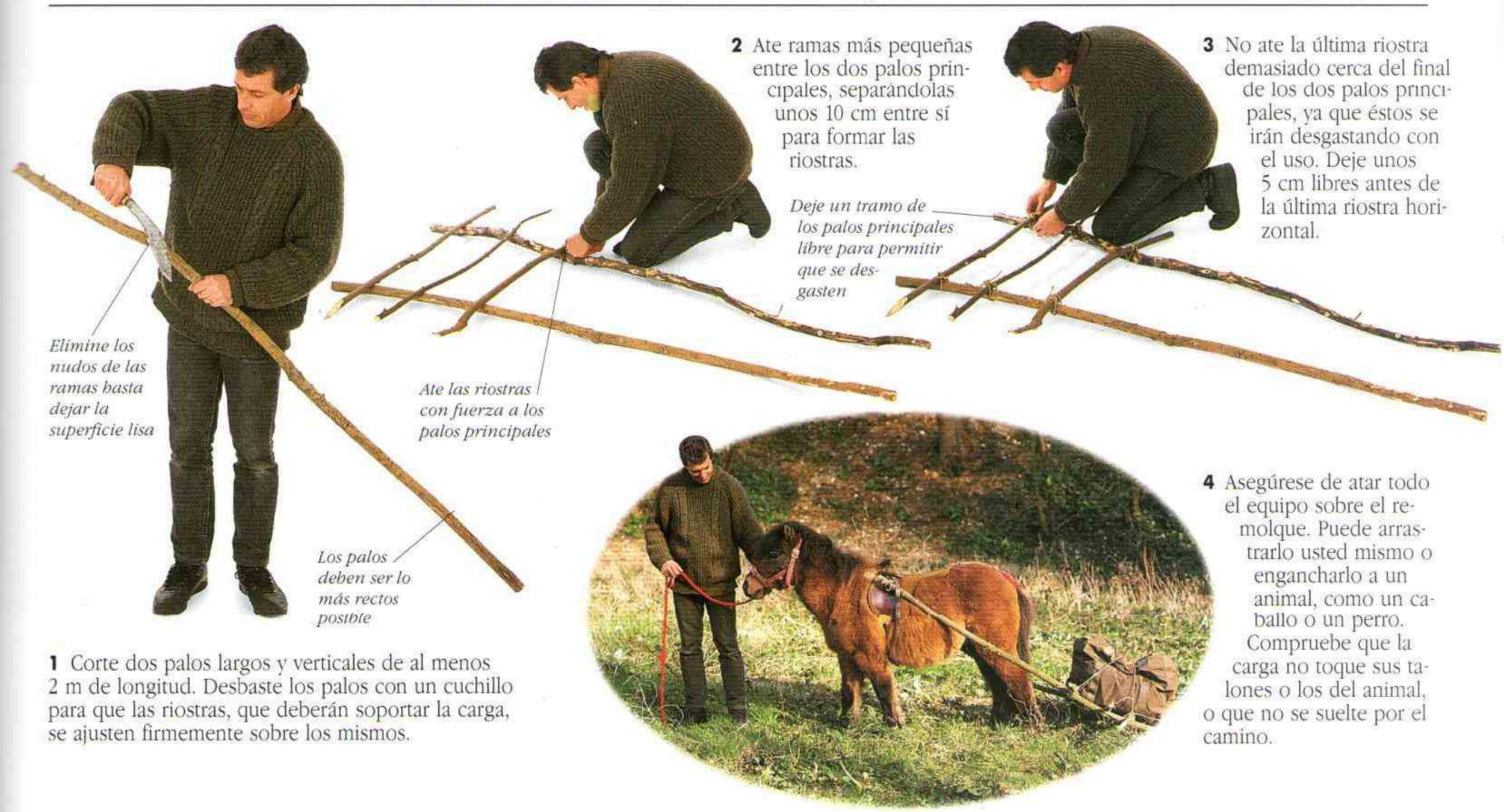

1 Corte dos palos largos y verticales de al menos 2 m de longitud. Desbaste los palos con un cuchillo para que las riostras, que deberán soportar la carga, se ajusten firmemente sobre los mismos.

2 Ate ramas más pequeñas entre los dos palos principales, separándolas unos 10 cm entre sí para formar las riostras.

3 No ate la última riostra demasiado cerca del final de los dos palos principales, ya que éstos se irán desgastando con el uso. Deje unos 5 cm libres antes de la última riostra horizontal.

4 Asegúrese de atar todo el equipo sobre el remolque. Puede arrastrarlo usted mismo o engancharlo a un animal, como un caballo o un perro. Compruebe que la carga no toque sus talones o los del animal, o que no se suelte por el camino.

LA MARCHA A PIE

EL IR A pie llevando un equipo pesado nada tiene que ver con nuestros paseos diarios urbanos. Al caminar en grupo hay que seguir el ritmo del miembro más lento, y se requiere una planificación, organización y previsión muy detalladas. Una vez comenzado el viaje, hay que detenerse después de 10 minutos para ajustar los calcetines, botas, vestimenta y equipo, comprobando que se mantiene la dirección correcta y que todos se encuentran bien. A partir de ese momento, hay que seguir un ritmo, aminorando la marcha al ir cuesta arriba o cuesta abajo.

TÉCNICAS DE MARCHA

Camine lenta y regularmente, balanceando los brazos para mantener el impulso y el equilibrio, y no fuerce los pasos. Intente mantener un paso uniforme y continuado –aquél que pueda seguir durante un período prolongado sin perder el aliento o lastimarse. Intente detenerse a intervalos regulares para descansar y comprobar su equipo.

Caminar cuesta arriba. *Disminuya la longitud de su paso al ir cuesta arriba, pero mantenga el mismo ritmo. Inclínese hacia adelante y coloque el pie plano sobre el suelo.*

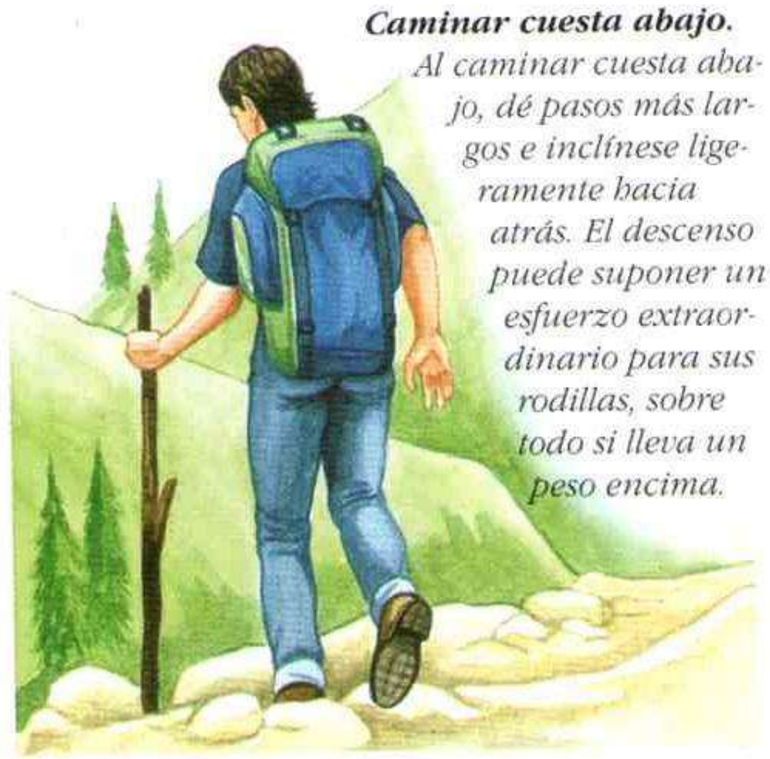

Caminar cuesta abajo. *Al caminar cuesta abajo, dé pasos más largos e inclínese ligeramente hacia atrás. El descenso puede suponer un esfuerzo extraordinario para sus rodillas, sobre todo si lleva un peso encima.*

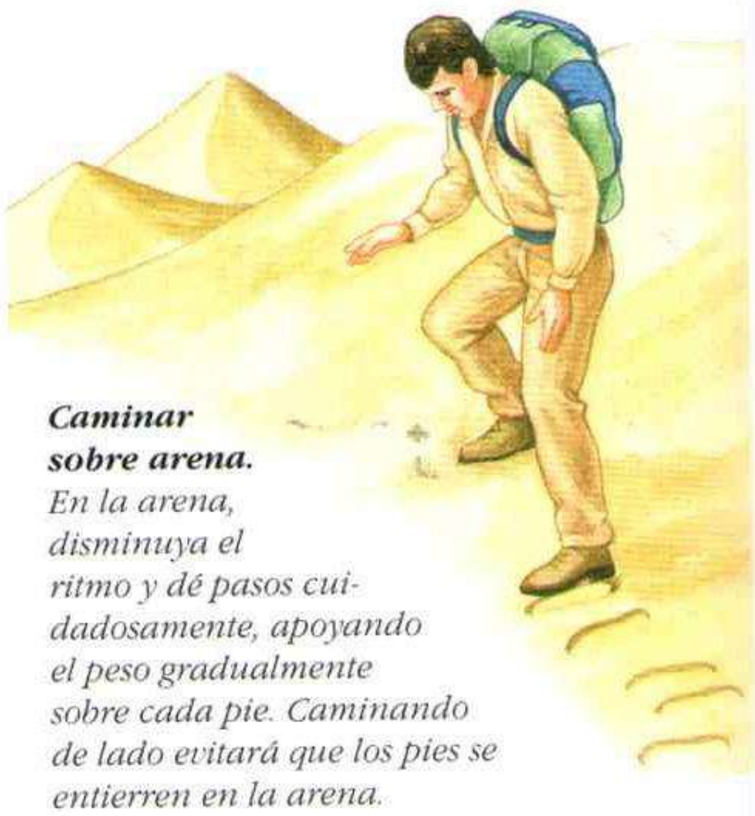

Caminar sobre arena. *En la arena, disminuya el ritmo y dé pasos cuidadosamente, apoyando el peso gradualmente sobre cada pie. Caminando de lado evitará que los pies se entierren en la arena.*

CAMINAR EN GRUPO

Al caminar en grupo, envíe una avanzadilla de personas en buena forma física para comprobar la ruta, encontrar un sendero y buscar la manera de sortear los obstáculos. Al final de la jornada, estos mismos miembros del grupo pueden encontrar también un sitio donde acampar, levantar las tiendas o refugios, recoger agua y encender el fuego. Sin embargo, deben permanecer con el grueso del grupo la cantidad suficiente de personas en forma para ayudar a aquellos que no puedan caminar tan deprisa como los demás, y para ayudar a cualquier persona enferma o herida.

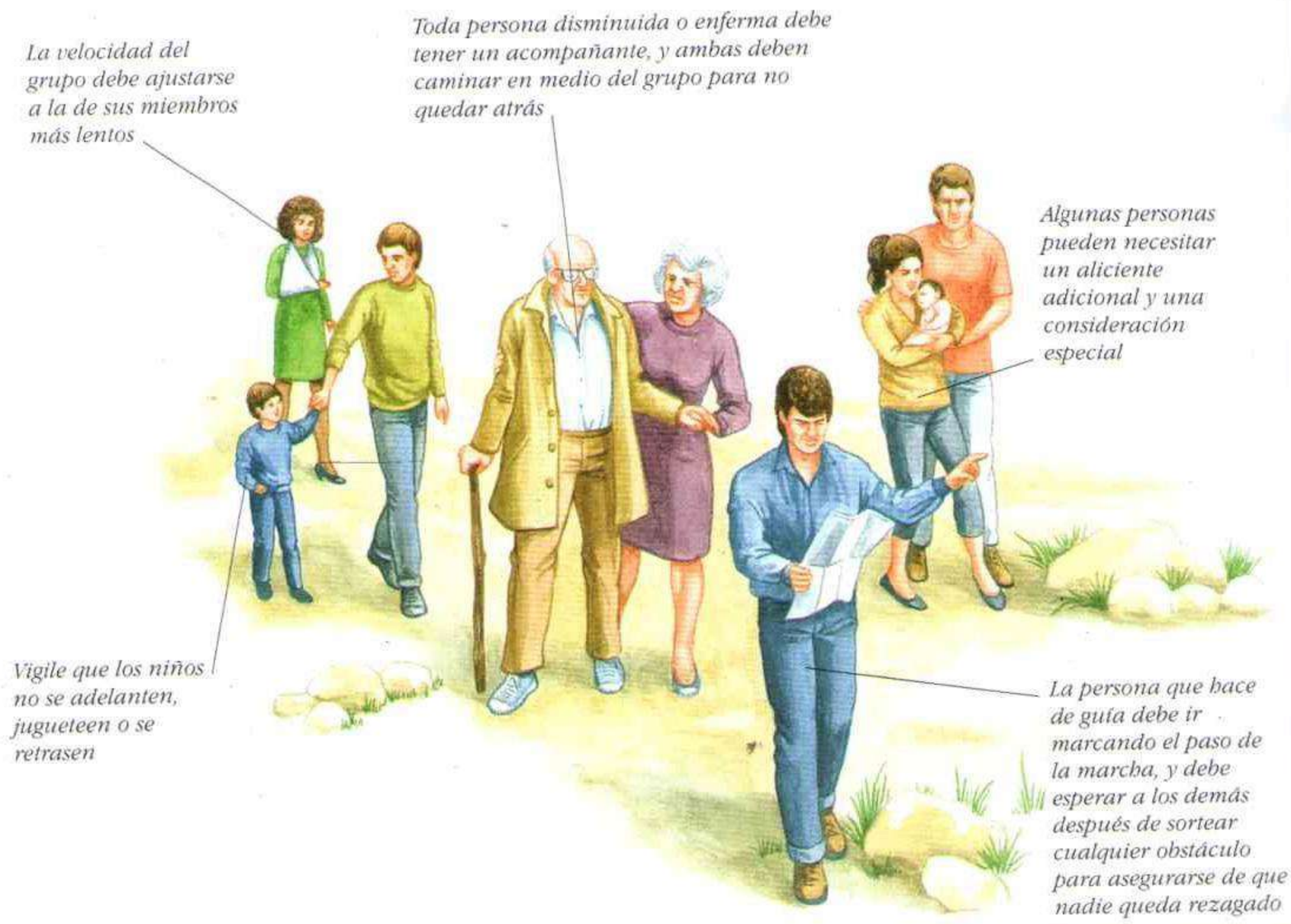

Mantenerse juntos. *Es importante que un grupo se mantenga unido –a excepción del grupo de avanzadilla–, en especial en una situación de supervivencia. No deje que los niños o las personas heridas se retrasen. Mantenga un paso que todos puedan seguir.*

PLANIFICAR LA RUTA

La mejor forma de planificar su ruta es recorriéndola primero, aunque esto es generalmente poco factible, en especial en una situación de supervivencia. Antes de ponerse en marcha, busque un punto elevado, como una loma o un árbol al que pueda trepar fácilmente, y utilice los prismáticos y un mapa, así como cualquier otro elemento disponible, para determinar la ruta más fácil *(véase página 122)*. En la selva, las crestas siempre tienen menos vegetación que el fondo de los valles, y son más fáciles de seguir. Sin embargo, puede que existan senderos marcados a lo largo de los cursos de agua, creados por los habitantes del lugar. Intente seguir siempre la ruta más fácil, particularmente si se encuentra en una situación de supervivencia.

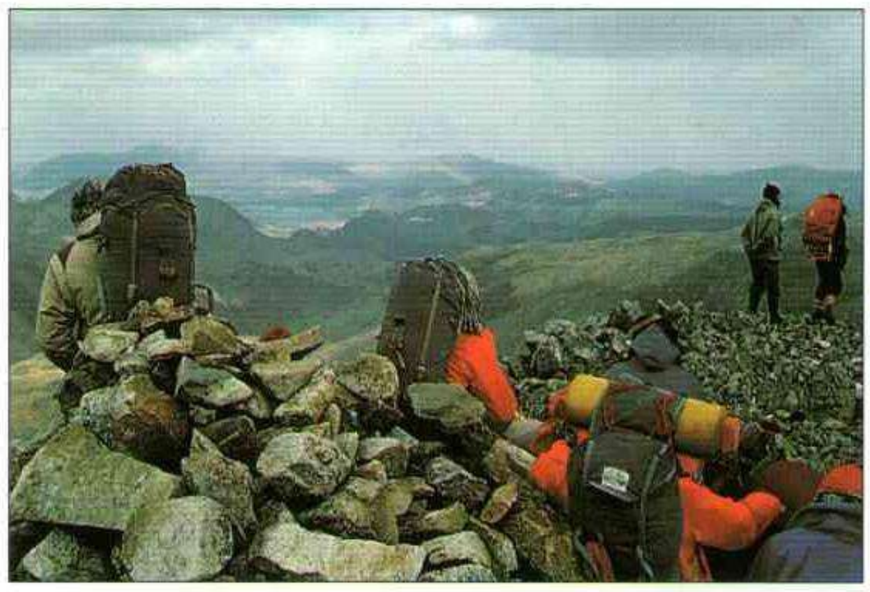

Parada de descanso. *Las paradas para descansar son esenciales, en particular si camina en grupo. Intente no prolongarlas más de 10 minutos en cada ocasión, pero no empiece a contabilizar los minutos hasta que todos los miembros hayan llegado y estén sentados. Mientras el grupo descansa, la avanzadilla puede ir comprobando la ruta.*

CAMINAR DE NOCHE

Camine lentamente durante la noche, comprobando a cada paso el estado del terreno, antes de apoyar todo el peso sobre el pie. Aproveche la luz de la luna y las estrellas, y camine por terrenos abiertos, alejados de los árboles. Agáchese y mire hacia arriba para comprobar el contorno del camino contra el cielo. Mire hacia los lados de los objetos y no directamente a los mismos. Deténgase a intervalos regulares, permaneciendo completamente en silencio, y escuche con atención.

Conserve la visión nocturna. *Nuestros ojos necesitan unos 30 minutos o más para adaptarse a la oscuridad. Para mirar el mapa y la brújula con la linterna, cierre un ojo. De esta manera, recobrará más rápidamente la visión.*

LOS PELIGROS DE LA MARCHA

Existen muchos peligros al caminar en la naturaleza, que varían desde las caídas hasta las mordeduras de insectos y serpientes. Por ejemplo, hay varias especies de mosca negra urticante en los bosques y tierras húmedas de Norteamérica. Las garrapatas también constituyen otro peligro, y pueden pasar accidentalmente de los arbustos a su ropa *(véase página 180)*. Tendrá que adaptar su estilo de marcha al tipo de terreno que atraviese para evitar lesionarse *(véase página 145)*. Otro peligro es el agotamiento por calor. Asegúrese de que su ropa permite la transpiración para refrescarse sin pasar frío *(véase página 23)*.

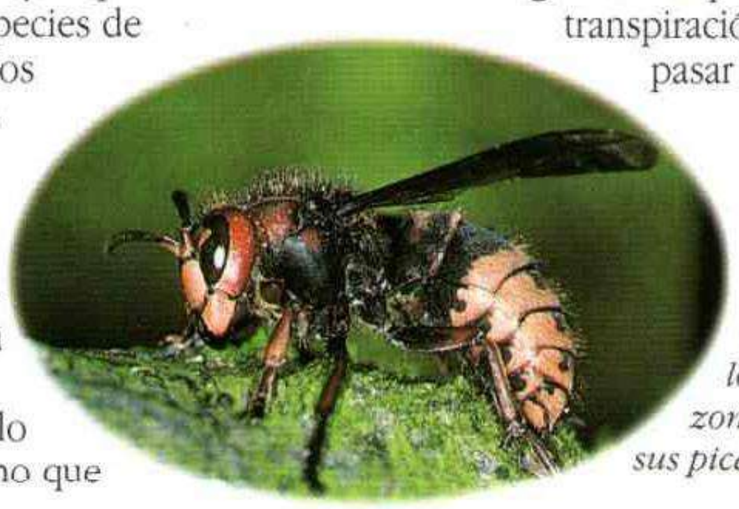

Avispón. *Varias especies de avispones* (Vespa) *pueden constituir una amenaza para los caminantes en las zonas templadas, debido a sus picaduras muy dolorosas.*

MOCHILA «HUDSON BAY»

1 Si no dispone de una mochila, puede llevar su equipo en esta mochila improvisada. En primer lugar, envuelva una piedra en cada una de las esquinas de una lona o manta impermeable para proporcionar un buen punto de sujeción.

Ate una esquina de la lona alrededor de una piedra

2 Extienda su equipo sobre la lona y enróllela con la ropa dentro haciendo un hatillo en forma de salchicha.

3 Ate el hatillo con fuerza utilizando una cuerda; compruebe que todo su equipo esté bien enrollado en el interior y que no pueda perderlo. Ate una cuerda de un extremo al otro, para que haga de correa.

El hatillo se lleva cruzado sobre la espalda, como una mochila

4 Cuelgue el hatillo cruzándolo por su espalda, pasando la cuerda por el pecho. Puede acolchar su hombro con vegetación para evitar que la cuerda le roce la piel.

CAMINAR SOBRE HIELO Y NIEVE

LOS PELIGROS DE caminar sobre el hielo y la nieve radican tanto en la naturaleza del terreno como en el efecto del frío sobre el cuerpo humano. En las regiones polares, mantener el calor no es tanto un problema como el exceso de calor seguido de sudor al trabajar o viajar. Debe tener una excelente condición física para recorrer un territorio polar, además de disponer del equipo adecuado, sin el cual no podría sobrevivir durante mucho tiempo. Caminar sobre el hielo y la nieve puede ser muy peligroso en las áreas templadas, y no debe intentarlo sin una planificación apropiada.

INDUMENTARIA ADECUADA PARA LAS REGIONES POLARES

El intenso frío de las regiones polares requiere una indumentaria especial. Para caminar en estas zonas necesita ayudas especiales como las que le proporcionan las raquetas de nieve, los crampones o los esquís, mientras que para cruzar un área de hielo, como los glaciares, requerirá un piolet. Para evitar un sobrecalentamiento, debe ser capaz de ventilar su vestimenta, sea abriendo cremalleras o eliminando capas. Es fundamental seguir el sistema de capas *(véase página 22)* y llevar prendas resistentes al viento como capa exterior. Las temperaturas muy bajas y los vientos helados exigen cubrir incluso las mínimas áreas de piel expuestas al aire para evitar la congelación *(véase página 141)*. También son esenciales unas gafas para impedir la ceguera producida por el reflejo de la nieve.

EQUIPO INDISPENSABLE

En las regiones polares, unas gafas pueden evitar la ceguera producida por el reflejo de la nieve. Los crampones evitan que resbale sobre el hielo, mientras que las raquetas le permiten caminar sobre la nieve profunda. Un piolet para apoyarse en las pendientes heladas y unos palos de esquí le ayudan a mantener el equilibrio.

Capucha. *La capucha debe extenderse hacia adelante para cubrir la cara. Debe llevar un gorro debajo de la misma.*

Anorak. *Un anorak debe ser resistente al viento y fácil de ventilar, de la talla adecuada para permitir llevar muchas capas de ropa por debajo. Asimismo, debe ser aislante y de peso ligero.*

Piolet. *Un piolet es útil para construir escalones en las pendientes heladas. También puede utilizarlo como freno en caso de resbalar y caer.*

Guantes. *Son necesarios tres guantes: unos finos para evitar que la piel se adhiera al metal, unos intermedios de lana y unas manoplas exteriores resistentes al viento e impermeables. Los guantes deben sujetarse al cuerpo y entre sí por una cinta que suba por los brazos y pase por los hombros; así evitará perderlos.*

Pantalón. *Un pantalón o peto (pantalón con cintura alta y tirantes) debe tener una capa exterior resistente al viento y una aislante en el interior. Los pantalones deben tener cremalleras traseras para permitir realizar las funciones naturales sin tener que quitarse el anorak o exponer parte del cuerpo al aire frío.*

Martillo. *Se utiliza para clavar clavijas y asideros en el hielo al escalar* (véase página 146).

Mantener el calor. *La indumentaria para las regiones frías debe componerse de varias capas que retengan el aire caliente entre ellas y de esta manera aíslen el cuerpo* (véase página 23).

Botas. *Las botas constan de una cubierta exterior de plástico aislante y una bota interior gruesa, con una suela adecuada para el uso en las tiendas y en el campamento. Deben llevarse encima de varias capas de calcetines, finos y gruesos.*

RAQUETAS IMPROVISADAS

1 Curve una rama joven y verde ayudándose con la rodilla y dóblela gradualmente hasta que sea flexible.

2 Con un cuchillo, elimine cuidadosamente toda la corteza en la parte interior de la curva. De esta manera, facilitará el doblamiento de la madera.

3 Corte un lado de ambos extremos de la rama hasta que quede encajado uno en el otro.

4 Ajuste los extremos del aro terminado. Ate los extremos con un cordel.

5 Ate dos palos cortos entre sí por el centro, con una cuerda. Repita la operación con dos pares más de palos.

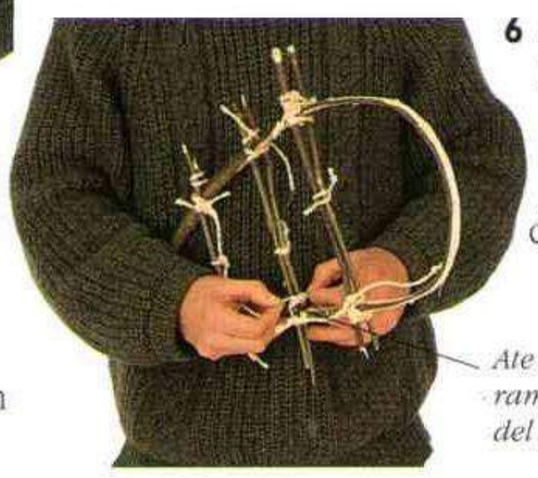

6 Ate tres pares de ramas pequeñas de un lado a otro del aro. Éstas servirán de soporte al pie.

Ate cada par de ramas a lo ancho del aro

7 Entreteja un cordel entre las ramas y alrededor de las mismas para formar la base de la raqueta.

8 Ate las raquetas terminadas a sus botas con un cordel. Las raquetas reparten el peso del cuerpo sobre un área mayor que la que proporcionan los pies.

TÉCNICAS PARA CRUZAR HIELO Y NIEVE

Nieve profunda. *Para cruzar una gran extensión de terreno nevado en grupo, lo ideal es caminar en fila india, dejando poca distancia entre cada miembro del grupo. Esto evita que alguno se pierda en una ventisca, y permite que, en el caso de que alguien cayera, los demás pudieran socorrerle rápidamente. Una capa de hielo sobre la nieve profunda puede soportar su peso, pero tenga mucho cuidado. Utilice el piolet como soporte para caminar.*

Hacer frente a las pendientes. *Cruzar pendientes heladas o lagos congelados requiere el mayor de los cuidados. Debe atarse a sus compañeros y emplear su piolet para tantear el camino por si encontrara grietas ocultas, o para sujetarse en caso de resbalar. Si el hielo se cuartea, reparta su peso en la mayor superficie posible, y utilice el piolet para empujarse fuera. No camine nunca encima de un embalse congelado. Si éste se ha vaciado, podría existir un hueco de varios metros debajo del hielo, y si el hielo se rompiera no podría volver a la superficie. Si resbala sobre una pendiente helada, clave la punta de su piolet en el terreno y apoye todo su peso sobre ella. Puede actuar como freno y detenerle en la caída por la pendiente.*

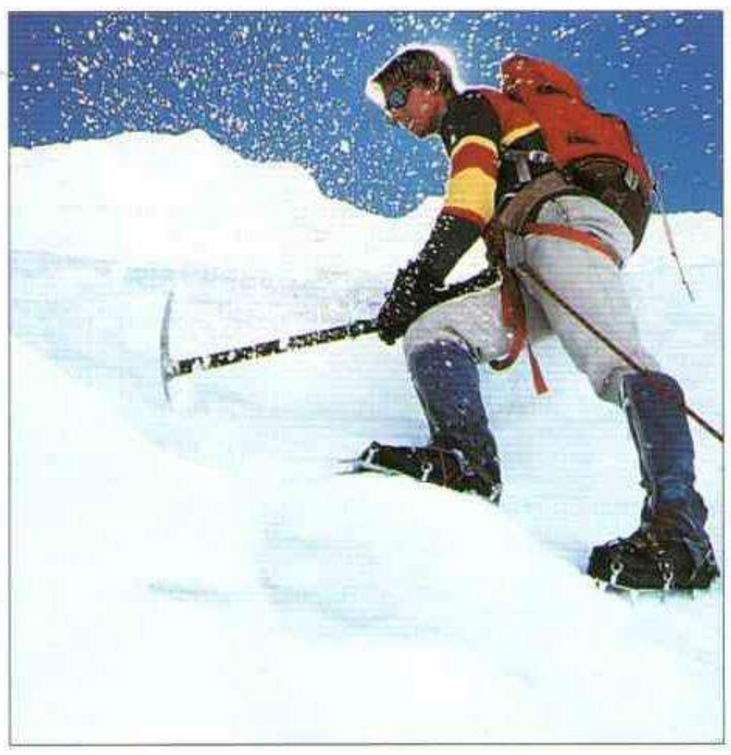

ESQUÍ DE FONDO

Los esquís empleados en el esquí de fondo son largos y estrechos. Están encerados para que se deslicen y tengan un buen agarre a la vez, permitiendo un movimiento frontal y evitando los deslizamientos hacia atrás. El impulso proviene del empuje de la pierna que queda atrás, mientras que la otra se desliza hacia delante. Los esquís de fondo son muy recomendables para desplazarse sobre la nieve en polvo muy profunda, donde resulta imposible caminar.

Deslizamiento silencioso. *Los esquís de fondo permiten desplazarse de forma eficaz y casi silenciosa en zonas silvestres.*

VIAJAR SOBRE NIEVE

CON PATINES Y esquís especiales, podrá desplazarse rápida y eficientemente sobre el hielo y la nieve profunda. Sin embargo, cruzar una extensión helada, o incluso con una pequeña cantidad de nieve, puede suponer un enorme esfuerzo. Únicamente las personas experimentadas y con un buen entrenamiento, y dotadas del equipo adecuado, deberían emprender viajes a las regiones polares, donde las bajas temperaturas embotan la mente, las noches invernales casi interminables deprimen a la persona más alegre, e incluso los vientos moderados aumentan drásticamente el riesgo de congelación.

Vehículos para la nieve. *Con fines deportivos o prácticos, las motonieves han reemplazado a los trineos de perros en la vida cotidiana de las regiones árticas.*

TRINEOS DE TRACCIÓN HUMANA

Los trineos de tracción humana son muy eficientes para su empleo sobre hielo o nieve helada, en especial si lleva esquís. Esquiar sobre nieve blanda requiere un mayor esfuerzo, y resulta más adecuado usar raquetas, aunque el esfuerzo sigue siendo considerable. Debe tener mucho cuidado de no acalorarse y sudar, porque el sudor mojaría su ropa. Asegúrese de que puede ventilar su ropa para liberar el exceso de calor.

Tirar de un trineo. *Los trineos evitan el tener que cargar un peso y se deslizan bien sobre el hielo y la nieve dura. Sin embargo, resultan difíciles de frenar cuando van cuesta abajo, en cuyo caso es imposible hacer un viraje. Asegúrese de poder liberarse del arnés rápidamente, pero mantenga siempre su trineo bajo control, deteniéndose después de cada tramo cuesta abajo, aunque esto no siempre sea sencillo.*

ADVERTENCIA

Los dedos, la nariz, los pies, las orejas y la cara son las partes del cuerpo que primero se congelan. Cuando la piel comienza a congelarse da una sensación de hormigueo. Después aparecen unas zonas entumecidas y de aspecto ceroso. Si no vuelven a calentarse, estas zonas adquieren un aspecto granuloso, después se enrojecen, forman ampollas, y mueren antes de desprenderse totalmente.

TRINEO DE PERROS

Transporte tradicional. *Los trineos de perros fueron utilizados tradicionalmente para el transporte polar, y algunos exploradores aún lo consideran el mejor método para el transporte de provisiones a través de los terrenos con hielo o nieve. Increíblemente resistentes, capaces de dormir en medio de una ventisca y de desplazarse rápidamente en casi todo tipo de terreno, los perros son más fáciles de mantener y son una compañía mucho más agradable que las motonieves.*

Trabajo en equipo. *El equipo trabaja en conjunto como si se tratara de una manada, siendo el perro guía su jefe. Éste considera al conductor el guía de la manada.*

EL EFECTO REFRIGERANTE DEL VIENTO

Velocidad	Temperatura (°C)														
Calma	-4	-7	-9	-12	-15	-18	-20	-23	-26	-29	-31	-34	-37	-40	-43
	Temperatura equivalente de congelación (°C)														
8 km/h	-7	-9	-12	-15	-18	-20	-23	-26	-29	-32	-34	-37	-40	-43	-45
16 km/h	-12	-15	-18	-23	-26	-29	-32	-37	-40	-43	-45	-51	-54	-57	-59
24 km/h	-18	-20	-23	-29	-32	-34	-40	-43	-45	-51	-54	-57	-62	-65	-68
32 km/h	-18	-23	-26	-32	-34	-37	-43	-45	-51	-54	-59	-62	-65	-70	-73
40 km/h	-20	-26	-29	-34	-37	-43	-45	-51	-54	-59	-62	-68	-70	-76	-79
48 km/h	-23	-29	-32	-34	-40	-45	-48	-54	-57	-62	-65	-70	-73	-79	-82
56 km/h	-23	-29	-34	-37	-40	-45	-51	-54	-59	-62	-68	-73	-76	-82	-84
64 km/h	-26	-29	-34	-37	-43	-48	-51	-57	-59	-65	-70	-73	-79	-82	-87

Las condiciones son relativamente agradables si se toman las precauciones normales

Las condiciones son muy desagradables; se requiere una vestimenta exterior térmica

La piel comienza a congelarse cuando se expone al aire libre durante un período prolongado

Un desplazamiento al aire libre puede ser peligroso; la piel expuesta puede congelarse en un minuto

Condiciones extremadamente peligrosas; la piel expuesta puede congelarse en 30 segundos

Vientos fríos. *El viento puede llegar a matar en las regiones del norte, ya que puede enfriar el aire por debajo de la temperatura ambiental. Conforme aumenta la velocidad del viento, su efecto sobre la temperatura se multiplica.*

CÓMO CONSTRUIR UN TRINEO

1 Puede construir un trineo a partir de una rama de árbol ahorquillada. Ate los extremos de la rama a la curva que forma la horquilla para tensarla.

2 Cualquier cuerda que toque el suelo se desgastará rápidamente, por lo que debe atar ramas de refuerzo a los patines para construir un soporte para la plataforma.

Construya un asa con una rama y una cuerda

Trineo cargado. *Envuelva su equipo en una manta o tela impermeable y átelo con fuerza al trineo.*

3 Ate algunas ramas sobre las piezas de refuerzo para formar la plataforma principal del trinco.

Las ramas están atadas a los refuerzos para soportar la carga

4 Ate la mayor cantidad posible de travesaños a los refuerzos para formar una plataforma segura sobre la cual sujetar su equipo.

Un transporte fiable. *Gracias a su grueso pelaje y al pelo que protege sus patas, los huskies están perfectamente capacitados para viajar por las regiones polares.*

Los tirantes unen los arneses de todos los perros entre sí para que trabajen al unísono

El grueso pelaje les protege del viento y la nieve

CRUZAR EL DESIERTO

LOS DESIERTOS ESTÁN generalmente deshabitados porque no existe la suficiente cantidad de agua ni de vegetación para el sustento de una comunidad permanente. Son lugares con climas muy extremos: a menudo hace mucho calor durante el día y el frío es extremo por la noche. Las condiciones climáticas también pueden variar rápidamente debido a las violentas tormentas de arena y a las lluvias torrenciales. El desierto es un medio inhóspito, en el que el agua es una necesidad esencial y en el que únicamente sobreviven las criaturas más resistentes.

INDUMENTARIA ADECUADA PARA EL DESIERTO

El objetivo principal de la indumentaria para el desierto es mantener constante la temperatura corporal. La ropa debe proteger al cuerpo del calor del sol durante el día, y retener el calor corporal durante la noche, cuando las temperaturas bajan drásticamente. Todas las partes del cuerpo deben estar cubiertas y protegidas del sol y de la incomodidad que supone una tormenta de arena. En los desiertos no hay prácticamente humedad; esto hace que el calor sea más soportable que en otras regiones cálidas. Como la ropa para el desierto no debe protegerle de la lluvia, puede ser cómodamente porosa, permitiendo la ventilación entre las capas, y proporcionándole así un buen aislante durante la noche.

Sombrero. *Un sombrero de ala ancha y color claro refleja la caliente luz solar, y aísla la cabeza del frío nocturno.*

Gafas de sol. *Las gafas de sol con filtros protegen de la ceguera causada por el sol.*

Camisa de algodón. *Una camisa ligera de algodón debe ser de manga larga para protegerle los brazos del sol. Debajo de la camisa debe llevar una camiseta que controlará la evaporación del sudor y ayudará a mantenerle fresco.*

Chaqueta. *Una chaqueta resistente al viento protege de la vegetación espinosa, del sol, del viento y del frío nocturno.*

Cinturón. *Un cinturón tejido o trenzado no sólo evitará que la arena entre en sus ropas, sino que también podrá utilizarlo como cabestrillo en caso de emergencia* (véase página 179).

Manos. *Proteja sus manos de las quemaduras solares con cremas protectoras con filtro solar.*

Pantalón de algodón. *Un pantalón de algodón de tejido cerrado es resistente al viento y protege las piernas del sol, de la arena y de la vegetación espinosa.*

La ropa para el desierto. *La vestimenta para el desierto debe ser ligera y estar confeccionada con un material poroso pero resistente. Debe protegerle de los fuertes rayos ultravioleta y del calor del sol durante el día, así como del aire frío de la noche.*

Botas. *Las botas para el desierto tienen suelas gruesas y aislantes, y están confeccionadas con ante poroso que permite la respiración de los pies.*

IMPROVISAR EL EQUIPO

Como el desierto puede ser un lugar muy peligroso, es importante contar con el equipo adecuado, en especial con ropa apropiada que le proteja del calor y de los rayos ultravioleta del sol. Puede improvisar unas gafas de sol a partir de película fotográfica, un pañuelo a partir de una toalla para proteger su cuello, o hacerse unas botas con tela.

Gafas. *Debe proteger sus ojos del sol en todo momento. Improvise unas gafas cortando ranuras muy finas en una tira de película de fotografía, que atará alrededor de su cabeza con la ayuda de un cordel.*

TIPOS DE DESIERTO

Una zona se clasifica como desierto si recibe menos de 25 cm de lluvia al año. Los desiertos pueden ser muy diferentes, según su localización y clima *(véase página 16)*. Pueden ser áreas rocosas, enormes extensiones de dunas de arena, montañas rocosas con matorrales o simplemente una tierra desolada que no puede soportar una vida animal o vegetal especializada.

Desierto rocoso. *Las áreas llanas y rocosas fueron en su día los lechos de los antiguos mares. Estos desiertos, que se encuentran en todo el mundo, son muy inhóspitos debido a la falta de refugios y a los vientos constantes.*

Desierto de cactus. *Algunos desiertos pueden soportar algún tipo de vida especializada, como los cactus, que retienen agua en sus tejidos. Estos desiertos se encuentran en Norteamérica, Sudamérica, Australia y Asia.*

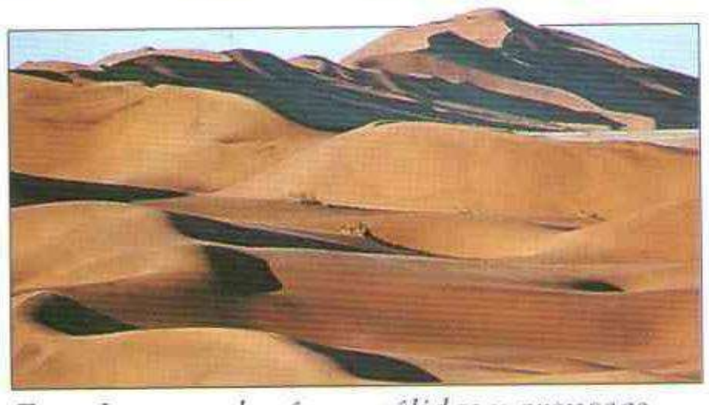

Erg. *Las grandes áreas cálidas y arenosas con dunas se conocen como* ergs. *Se encuentran en Australia, en el norte de África y en Asia.*

Semidesierto templado. *A pesar del crecimiento de algunos arbustos leñosos, es demasiado seco para soportar el crecimiento de hierbas. Los semidesiertos templados se encuentran en Norteamérica, Sudamérica, Asia central y Australia.*

CÓMO PROTEGERSE EN UNA TORMENTA DE ARENA

1 Las tormentas de arena son muy desagradables y pueden llegar a alterar el aspecto de un desierto. Debe marcar el sentido de su marcha con piedras o ramas.

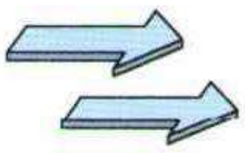

2 Refúgiese donde pueda, con el viento a su espalda. Cubra su cuerpo totalmente, protegiendo su cara y cuello de la arena. Espere hasta que la tormenta de arena haya cesado por completo antes de reemprender la marcha.

EL VIAJE POR EL DESIERTO

En una situación de supervivencia en el desierto, no debe moverse a menos que sea imprescindible. Si debe desplazarse, por ejemplo para buscar agua, o porque no existe la esperanza de que pueda ser rescatado desde donde se encuentra, hágalo únicamente de noche, cuando refresca, y si es posible bajo la luz de la luna. Como es muy difícil determinar dónde se encuentra en un desierto sin referencia alguna, la busqueda de un oasis (suponiendo que sabe dónde se encuentran) está lejos de ser fácil.

Sombra improvisada. *Puede preparar un espacio sombreado colgando una manta sobre una cuerda atada entre dos estacas. Cave una depresión en la arena para tener espacio suficiente para moverse bajo el tejado.*

TIEMPO DE SUPERVIVENCIA EN EL DESIERTO CON RESERVA LIMITADA DE AGUA

°C	SIN AGUA	2 LITROS	10 LITROS	°F
Descansando en la sombra				
50	2-5 días	2-5 días	3-5 días	122
30	7 días	5-8 días	14 días	86
20	12 días	14 días	23-25 días	68
Caminando de noche				
50	1 día	2 días	3-5 días	122
30	4 días	5 días	5-7 días	86
20	9 días	10-15 días	5-15 días	68

Supervivencia sin agua. *Los seres humanos necesitan un mínimo de 4-5 litros de agua por día para ser capaces de viajar a pie con un mínimo grado de seguridad. En el desierto, la conservación del agua de su cuerpo es primordial, por lo que únicamente debe trabajar de noche, hacer los mínimos esfuerzos, e intentar evitar la transpiración. A temperaturas de 50°C, sin agua para beber, únicamente sobrevivirá cinco días, si se limita a descansar en la sombra. Este tiempo se reducirá a la mitad si camina, incluso de noche. En una situación de supervivencia, debe cuestionarse seriamente la necesidad de cualquier tipo de movimiento a menos de que disponga de una gran reserva de agua.*

VIAJAR A TRAVÉS DE LA SELVA

A PESAR DE SER cálida, húmeda y estar llena de insectos, la selva es, en muchos aspectos, un jardín del Edén. Los alimentos son abundantes, no hay estaciones de carencia, y el agua fresca cae del cielo en grandes cantidades. Sin embargo, para aquellos que no estén familiarizados, la selva es húmeda, oscura y ruidosa: un lugar incómodo y amenazador. La humedad y el intenso esfuerzo que supone viajar a través de la selva hacen que la insolación sea un peligro constante, por lo que debe llevar consigo una reserva importante de agua para evitarla.

INDUMENTARIA ADECUADA PARA LA SELVA

Debe cubrirse completamente cuando viaje a través de la selva, y llevar ropa ligera y resistente que se seque rápidamente. Es posible que su ropa esté húmeda la mayor parte del día; deberá lavarla en agua fresca cada día para evitar que se pudra. No tiene sentido llevar ropa impermeable porque únicamente le haría sudar. Un jersey ligero resulta útil en las noches frescas. Tenga siempre a punto una muda de ropa limpia y seca para dormir. Guarde todo su equipo a menos que lo esté utilizando. Los artículos que no estén bien sujetos pueden engancharse en la vegetación, así que sujete sus objetos valiosos con cuerdas resistentes alrededor del cuello.

Sombrero. *Un sombrero de algodón debe tener un ala lo suficientemente ancha como para evitar que una rama, una hoja, o cualquier tipo de vegetación lastime sus ojos.*

Artículos importantes. *Los artículos más importantes de su equipo deben quedar sujetos a una cuerda alrededor del cuello para asegurarse de no perderlos* (véase página 28). *Al viajar, guárdelos debajo de la camisa.*

Red para mosquitos. *La red para mosquitos únicamente debe utilizarse al atardecer, cuando estos insectos son particularmente molestos. No debe utilizarla mientras viaja, ya que la malla, aunque fina, obstruye la visibilidad.*

Camisa. *La camisa debe ser de algodón y de manga larga. Asimismo, debe tener el cuello que pueda levantarse y abotonarse para protegerle la garganta.*

Cantimplora. *Una cantimplora es absolutamente necesaria. Llévela sujeta al cinturón para no perderla. Recuerde filtrar y depurar toda el agua que recoja antes de beberla* (véase página 74).

Botiquín de primeros auxilios. *Lleve siempre consigo un botiquín de primeros auxilios, sea en el cinturón o en un bolsillo.*

Tiras repelentes de insectos. *Estas tiras que se llevan en las muñecas y tobillos están impregnadas con repelente de insectos para alejar a los mosquitos y otros parásitos.*

Cuchillo. *Su* kukri *(cuchillo) o machete debe estar bien protegido; guárdelo dentro de una funda en su cinturón. Llévelo en la mano únicamente mientras lo esté empleando. Compruebe constantemente que no lo ha perdido.*

Pantalón. *El pantalón debe ser holgado, confeccionado con material resistente pero ligero, y debe ir ceñido a los tobillos, sea con cordones incorporados o con cinta elástica.*

Equipo para la selva. *Lleve siempre las mangas de la camisa bajas cuando se adentre en la selva. Proteja sus manos, cuello y cara con repelente de insectos. No se ponga repelente en la frente; si lo hace, el sudor y el repelente resbalarán hacia sus ojos.*

Botas. *Las botas deben tener suelas resistentes, con protección especial para los dedos y el empeine* (véase página 25). *La parte superior debe estar confeccionada en tela, para que se seque rápidamente, y deben estar provistas de válvulas de paso único para evacuar el agua al caminar.*

BOTIQUÍN PARA LA SELVA

No permita que las pequeñas heridas lleguen a infectarse. Al tener que abrirse camino entre una vegetación espesa y espinosa, las heridas en las manos son inevitables. Los antibióticos en polvo son ideales para prevenir infecciones. Cuando se abra paso entre la densa vegetación, tenga cuidado y evite herirse con el machete, porque podría dejarle imposibilitado o causarle una herida que difícilmente no se le infectaría. Necesitará una gran cantidad de tiritas impermeables para cubrir las pequeñas heridas y cortes.

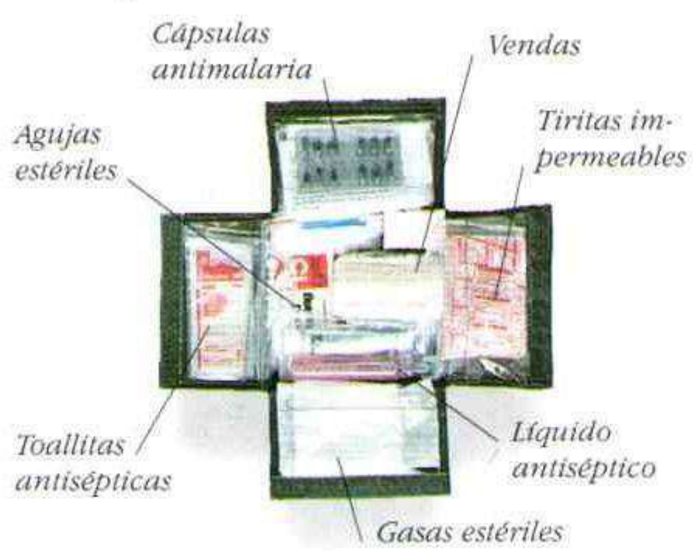

Medicación tropical. *Existen botiquines especiales para las regiones tropicales. Contienen artículos comunes para primeros auxilios y medicamentos para contrarrestar las enfermedades tropicales.*

CARACTERÍSTICAS DE LA SELVA

Bóveda verde. *En una selva primaria, los árboles pueden crecer hasta una altura de más de 60 m, creando una cubierta densa que oculta el sol. Esta bóveda va siguiendo la forma del suelo que hay debajo, pero desde el aire, lo único que puede verse es el denso techo de vegetación. La mayoría de los animales que habitan este medio lo hacen en la copa de los árboles.*

Vegetación exuberante. *La selva secundaria crece allí donde los hombres quemaron árboles para obtener tierras que cultivar. Después de la cosecha, abandonaron estos claros permitiendo que el sol llegara al nivel del suelo y estimulara el enorme crecimiento. Este tipo de selva es difícil de cruzar. Debe abrirse un sendero a través de la vegetación, y éste es un trabajo extenuante.*

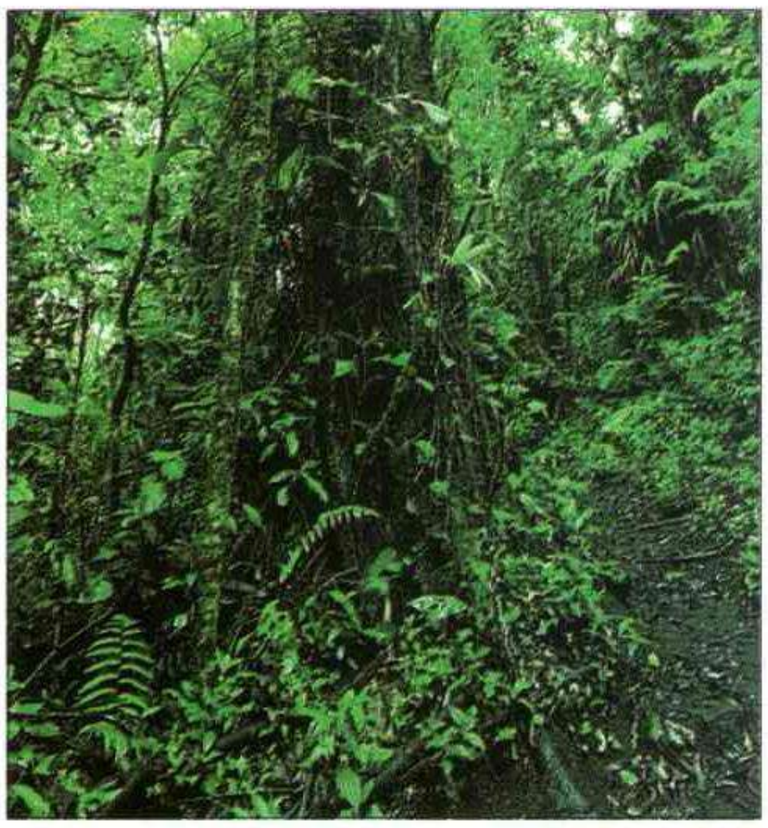

Selva umbría. *Debajo de la bóveda, el suelo de la jungla permanece relativamente libre de vegetación. Las lianas gigantes crecen hacia abajo, y una gran variedad de flora y fauna habita los distintos niveles de este medio húmedo y sombreado.*

DESPLAZARSE A TRAVÉS DE LA SELVA

Aunque es fácil desplazarse en la selva primaria, la selva secundaria es densa, llena de obstáculos y desagradable. Cortar la vegetación es un trabajo árduo que hace muy lento el avance. Abundan los insectos y las hierbas urticantes, muchas de las cuales tienen espinas peligrosas. Si queda atrapado por una de ellas, la mejor manera de liberarse por sí mismo es moviéndose hacia atrás, invirtiendo el proceso por el que quedó atrapado.

Aguarda un momento. *Algunas especies de* rattan, *conocidas como «aguarda un momento» pueden convertir su ropa en jirones. Si queda atrapado, espere pacientemente a que un compañero le libere.*

LA RUTINA EN LA SELVA ECUATORIAL

Conforme se acerca al ecuador, los días se hacen más regulares. En las áreas ecuatoriales, el amanecer y el atardecer tienen lugar a la misma hora cada día, y a menudo llueve a la misma hora en punto —por lo general, justo al atardecer. No hay una puesta de sol gradual, como ocurre en las latitudes más altas, sino una oscuridad repentina y total. La rutina diaria consiste en levantarse al amanecer, trabajar hasta el mediodía y descansar durante las horas de mucho calor. Si se está desplazando, comience a buscar un lugar para acampar no más tarde de las 15.00 horas.

Eliminar las sanguijuelas. *No se limite a arrancar las sanguijuelas, ya que las partes bucales podrían quedar dentro de la herida y causar infecciones. Tóquelas con una rama ardiendo o empápelas en repelente de insectos para que suelten su presa, y quítelas entonces.*

Toque las sanguijuelas con una rama ardiendo

LOS PELIGROS DE LA SELVA

La selva tiene muchos peligros, que van desde los animales salvajes hasta enfermedades como la malaria. La constante humedad puede causar infecciones fúngicas en los pies, por lo que es recomendable aplicar polvos de talco regularmente *(véase página 24)*. Sacuda la ropa y las botas antes de ponérselas, en caso de que se hayan convertido en refugio de arañas o serpientes venenosas.

Asegúrese de vacunarse contra la malaria, y tome la adecuada medicación antimalaria antes de adentrarse en la selva. Utilice los repelentes de insectos y cuelgue una mosquitera durante la noche para evitar sus picaduras.

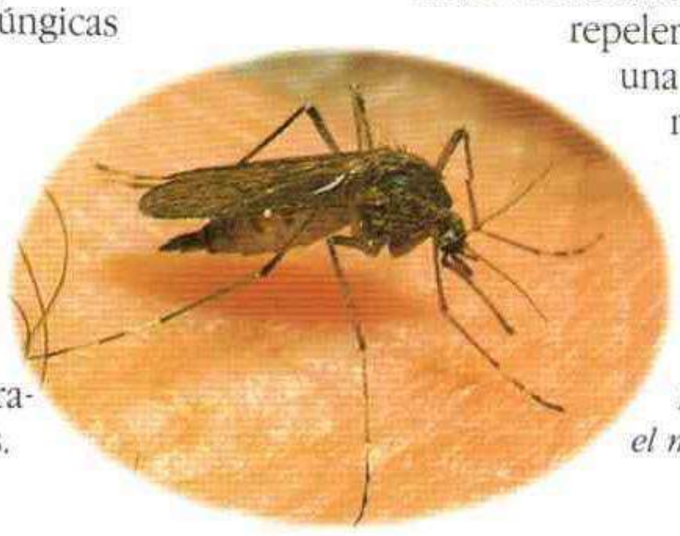

Enemigo mortal. *La hembra del mosquito* Anopheles *puede transmitir el mortal virus de la malaria.*

UNA ESCALADA SEGURA

ÚNICAMENTE DEBE EMPRENDER una escalada si no hay otra manera de sortear un obstáculo, y nunca si está solo. Si cae o se queda atrapado, nadie podrá ir en busca de ayuda. La regla de oro de todo escalador es mantener tres puntos de contacto con la roca, emulando la estabilidad de un trípode. Piense cuidadosamente cada movimiento, intentando mantener tres partes de su cuerpo (los dedos de los pies, los dedos de las manos, los pies o las manos —nunca las rodillas—) en contacto con la roca en todo momento. Tenga cuidado al evaluar el estado de la roca y la estabilidad de las presas para manos y pies. Utilice siempre una cuerda, si dispone de ella.

PRESAS PARA MANOS Y PIES

Las presas para manos y pies son la base de la escalada en roca. Antes de poner el pie o la mano en una presa, póngala a prueba con una patada o con un golpe de mano. No intente asir una muy lejana, o ponerse de puntillas para alcanzar una muy alta. Utilice sus piernas, más que las manos, para soportar su peso y empujarse hacia arriba. No se levante con los brazos.

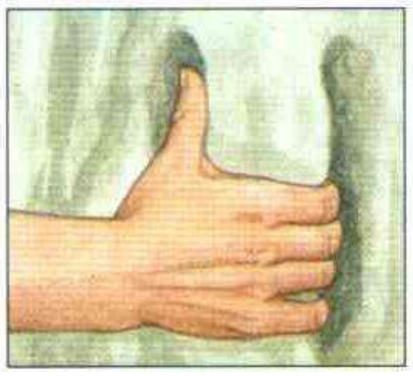

Presa de oposición. *Puede servirse de presas laterales para equilibrarse. Las presas de oposición son ideales si puede utilizar el pulgar para presionar hacia el lado contrario.*

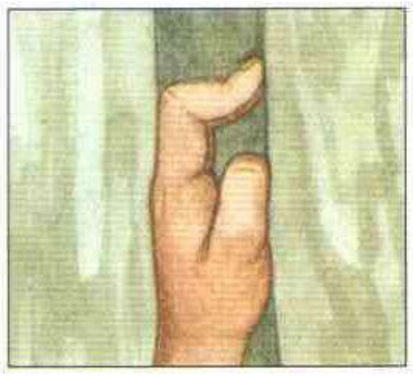

Empotramiento de mano. *Puede encajar su mano o puño en una pequeña grieta, e impulsarse hacia arriba cuando no exista una presa de otro tipo.*

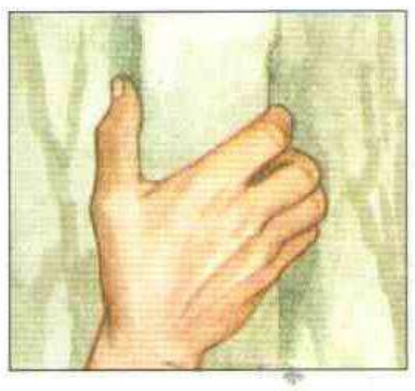

Presa de pellizco. *Cuando no pueda aferrarse bien a una presa, agárrela como si se tratase de un ladrillo. Intente pellizcar la roca.*

Empotramiento de pie. *Utilice una hendidura de la roca para apoyar la parte interior del pie. Deslice la suela a lo largo de la hendidura, en un ángulo de 90 grados con la pared.*

Presa grande. *Los agujeros en la roca se conocen como* presas grandes *y constituyen apoyos de pie ideales. Equilíbrese para que sus brazos y piernas soporten su peso por igual.*

Presa de pie en pendiente. *Intente colocar la totalidad de la suela sobre la presa. Si únicamente puede apoyar la punta del pie, doble el tobillo para mejorar la adherencia.*

ESCALADA BÁSICA

Si es usted un novato, no debería escalar sin estar sujeto a una cuerda asegurada por un escalador experimentado. Sin embargo, en una situación de supervivencia, puede no tener otra opción. Antes de escalar, busque con la vista el recorrido más fácil y seguro, y continúe mirando mientras va escalando. Si tiene cualquier duda sobre la ruta, retroceda y empiece de nuevo.

Ascenso. *Planee varios movimientos con antelación, y esté preparado para desplazarse, no sólo verticalmente, sino también lateralmente. Utilice los músculos de las piernas para impulsarse hacia arriba, y no dependa sólo de los brazos.*

Descenso. *El descenso le parecerá poco natural y las presas serán difíciles de ver. Si la pendiente es suave, baje mirando hacia afuera* (véase página 136). *Descienda de lado conforme aumenta la pendiente, hasta llegar a encarar la roca. Busque las presas con los ojos, no tantee con los pies.*

ASEGURADO

El seguro es el sistema por el cual un escalador se sujeta a otro que está escalando con una cuerda para parar una posible caída.

El asegurador debe sujetarse a la pared rocosa («anclaje») antes de comenzar. Para sujetar a alguien que desciende, o si su compañero necesita más cuerda para ascender, vaya soltando cuerda en la forma inversa a como la recogería.

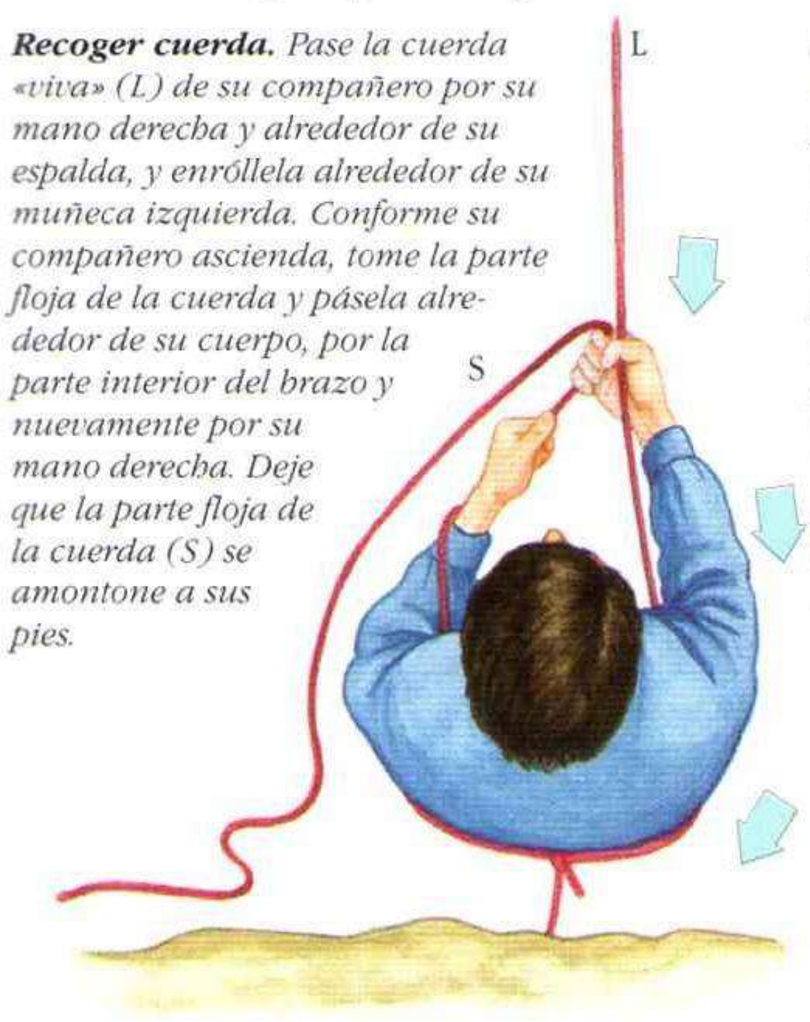

Recoger cuerda. *Pase la cuerda «viva» (L) de su compañero por su mano derecha y alrededor de su espalda, y enróllela alrededor de su muñeca izquierda. Conforme su compañero ascienda, tome la parte floja de la cuerda y pásela alrededor de su cuerpo, por la parte interior del brazo y nuevamente por su mano derecha. Deje que la parte floja de la cuerda (S) se amontone a sus pies.*

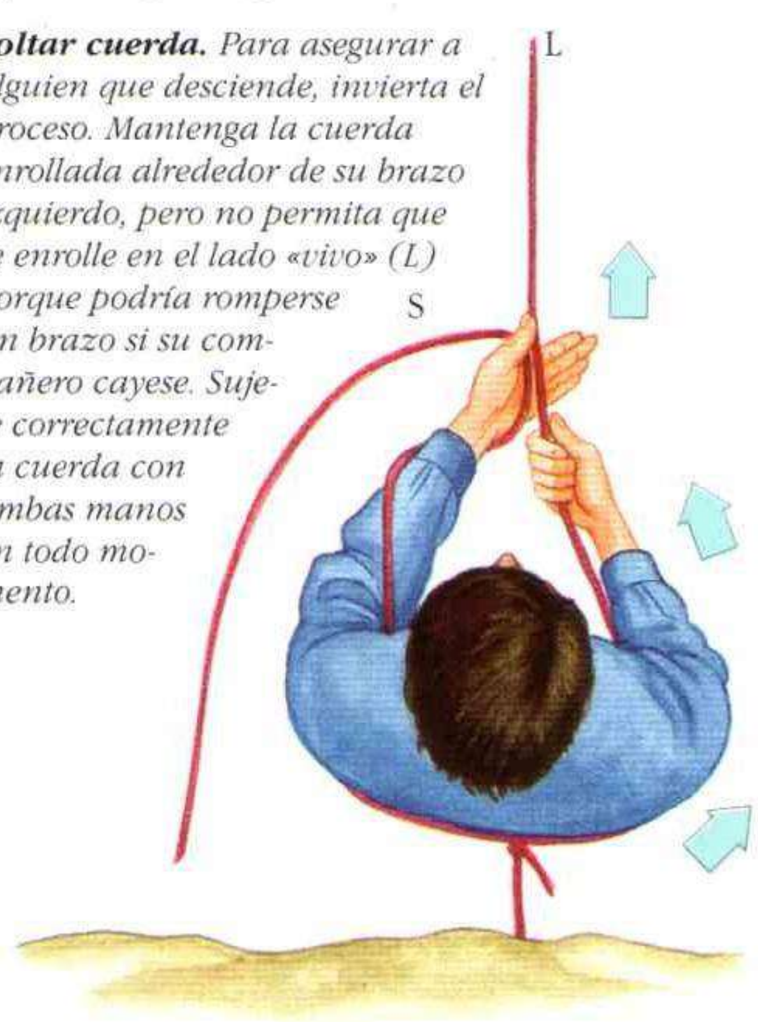

Soltar cuerda. *Para asegurar a alguien que desciende, invierta el proceso. Mantenga la cuerda enrollada alrededor de su brazo izquierdo, pero no permita que se enrolle en el lado «vivo» (L) porque podría romperse un brazo si su compañero cayese. Sujete correctamente la cuerda con ambas manos en todo momento.*

AGUANTAR UNA CAÍDA

Esté siempre preparado para aguantar a su compañero por si se cae. Para parar una caída, cruce su brazo izquierdo sobre la parte delantera de su cuerpo sujetando con fuerza su antebrazo contra el pecho. Reclínese en el punto de anclaje y haga fuerza.

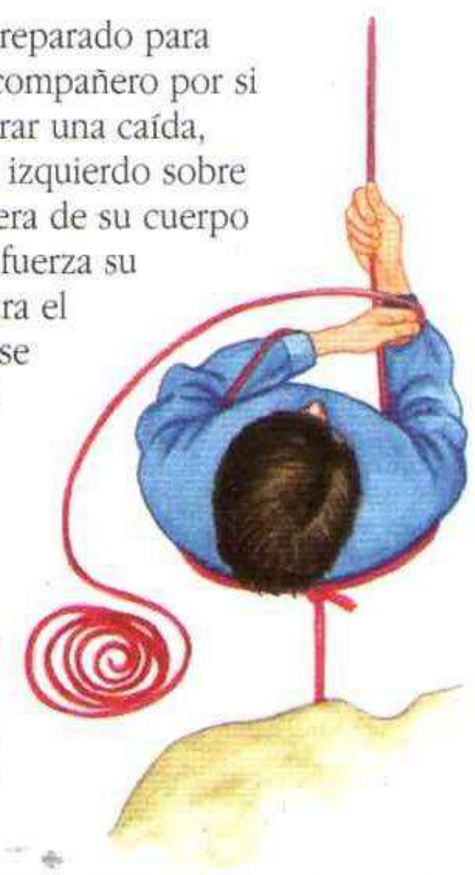

Bloqueo. *Esté preparado para sujetar a su compañero si cayese.*

RAPPEL ESPAÑOL O *DÜLFER*

El *rappel* español o *dülfer* es una técnica rápida y sencilla de descenso con cuerda. Pase la cuerda alrededor de un punto de anclaje seguro y suéltela. Colóquese frente al anclaje, ponga la cuerda entre las dos piernas, pásela alrededor de la cadera de modo que cruce el pecho hasta el hombro contrario; pasela por el hombro y déjela caer por la espalda. Agárrela con la mano que está del mismo lado que la cadera rozada por la cuerda. La otra mano aguanta la cuerda por arriba para mantener la posición vertical. Descienda caminando de espaldas, dejando que la cuerda se deslice por sus manos y alrededor de su cuerpo.

Dülfer. *Controle su velocidad con la mano derecha.*

ASCENSOS EN VARIOS LARGOS

Si ha de ascender por paredes de longitud superior a la de su cuerda, hágalo en varios largos, deteniéndose en puntos de reunión para asegurar a los demás miembros del grupo antes de continuar el ascenso.

1 El escalador que ascienda primero deberá buscar el recorrido más sencillo y seguro. Si la pared es muy empinada y peligrosa, puede ir insertando empotradores o clavijas conforme asciende, a fin de sujetar anillas de seguridad que aguanten la cuerda. El segundo escalador va soltando cuerda mientras permanece convenientemente anclado a la pared.

2 Una vez alcanzado el punto de reunión, se ancla a la pared y asegura al segundo escalador en su ascenso. Si el primer escalador ha empleado empotradores, el segundo deberá retirarlos conforme ascienda.

ESCALADA DE EMERGENCIA

AUN LA ESCALADA más sencilla puede complicarse. Incluso escaladores experimentados no pueden seguir, caen o tienen que ser rescatados. Los expertos utilizan avanzadas técnicas para librarse de algún problema. Son peligrosas y no tan fáciles como parecen en las ilustraciones. En una situación de supervivencia, los problemas más graves deben solucionarse utilizando los recursos disponibles. Atienda a estas técnicas por si tuviera que emplearlas. Si intenta probarlas antes, consulte un manual de escalada y pida el consejo de un escalador con experiencia.

OPOSICIONES

El avance en oposición es una técnica de escalada que consiste en escalar por el interior de grietas lo suficientemente anchas para el cuerpo humano. Debe prever la forma en que saldrá una vez llegue a la parte superior. Es posible que entonces no sepa cómo salir.

1 Compruebe si hay un recorrido alternativo antes de ascender por una grieta. Manténgase cerca del exterior de la grieta en una roca seca. Sujétese con ambas piernas y manos empujando contra las paredes.

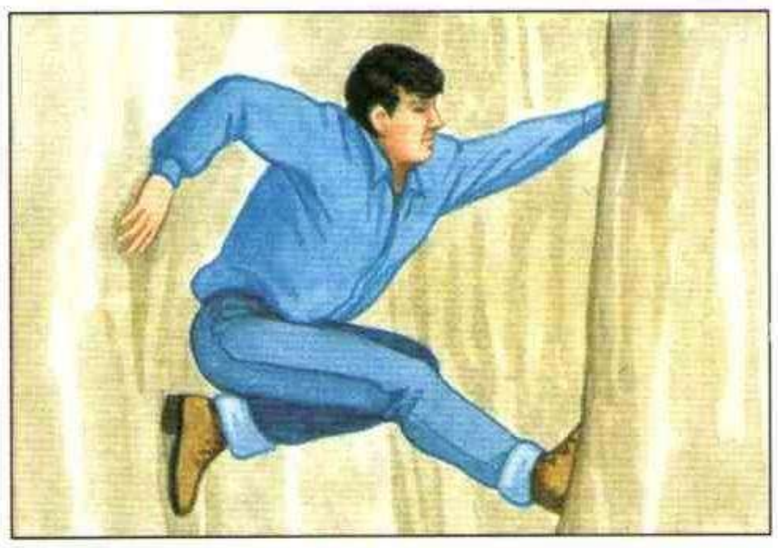

2 Presione un pie contra la pared trasera y la mano de ese mismo lado contra la pared delantera. Deslice la espalda hacia arriba, empujando con las piernas.

TOMARSE UN DESCANSO

El avance en oposición supone un esfuerzo muscular constante. Descanse a intervalos regulares apoyando la espalda contra la pared y manteniendo los brazos y piernas rectos.

DESCANSANDO

TÉCNICA DE CUBIERTA

Desde abajo, los resaltes en la pared de roca pueden parecer imponentes, en particular cuando no puede ver más allá de las presas. La técnica de cubierta se utiliza para superar estos resaltes. Únicamente debe intentarlo si no existe otra manera de superar un obstáculo. Es una técnica muy difícil, que requiere una gran condición física y habilidad.

1 Escale hasta que sea capaz de colocar ambos codos sobre el resalte, soportando su peso en los pies.

2 Apóyese sobre un codo, manteniendo el pecho y la cabeza lo más cerca posible de la roca. Apoye un tobillo sobre el resalte.

3 Empujándose con ambas manos, mueva la rodilla hacia el resalte. Levántese con cuidado y suba al resalte.

NUDO PRUSSIK*

Un nudo prussik es un lazo de tipo corredizo que se aprieta bajo un peso, pero que puede desatarse fácilmente. Los nudos prussik constituyen una forma segura de atarse a una cuerda, y pueden utilizarse para trepar por una cuerda, quizá después de una caída, o para subir un saliente.

1 Haga un aro en su mano con una cuerda (aquí coloreada de amarillo) y dóblela sobre otra cuerda (coloreada en rojo). Pase ambos cabos de la cuerda amarilla a través del aro amarillo.

2 A continuación, enrolle la cuerda amarilla alrededor de la cuerda roja, igual que hizo en el paso 1, para obtener cuatro vueltas.

3 Tire de la cuerda amarilla y junte las vueltas para tensarlas sobre la cuerda roja y hacerlas uniformes. Tenga cuidado de no superponerlas.

4 Para una mayor tirantez, repita los pasos 2 y 3 hasta obtener cuatro vueltas ajustadas alrededor de la cuerda roja. Retoque el nudo hasta que las vueltas sean uniformes y tensas.

MOSQUETÓN

Un mosquetón es un sistema de enganche útil para unir cuerdas o equipos *(véase página 152)*. Los mosquetones con abertura de muelle son más rápidos y fáciles de utilizar que aquéllos con cierre de seguridad (con tuerca), pero estos últimos son los más recomendables para el asegurado, el *rappel* y otras situaciones peligrosas *(véase página 147)*.

Mosquetón con cierre de seguridad. *Utilícelos para las técnicas que requieran una conexión muy segura, como el asegurado y el* rappel.

ESCALADA CON NUDOS PRUSSIK

Con dos cuerdas cortas y un arnés, puede ascender por una cuerda fija y salvar una situación difícil. En una emergencia, si no se dispone de un arnés apropiado, puede improvisar uno con cuerda. Haga un lazo en el que se pueda sentar y asegúrelo a una cuerda alrededor de su cintura con un nudo doble de tejedor *(véase página 184)*.

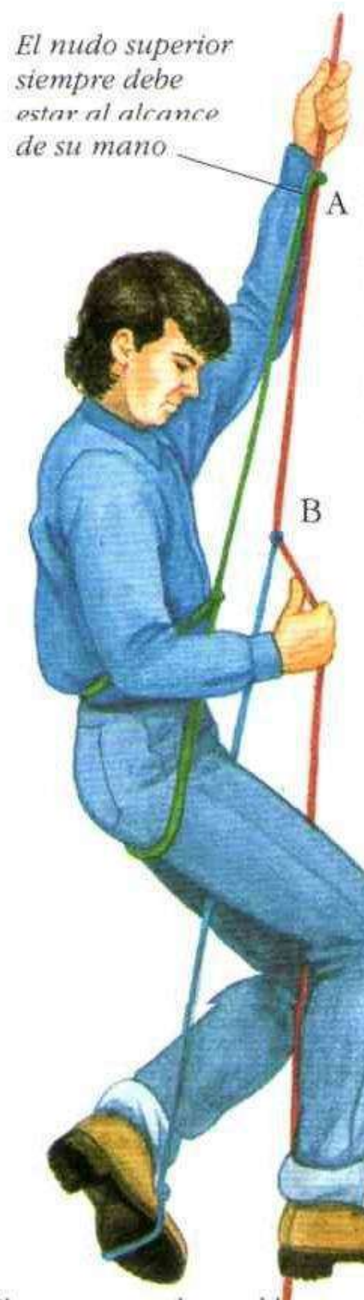

1 Asegure la cuerda de su arnés (aquí, coloreada de verde) a la cuerda fija, utilizando un nudo prussik (A). Sujete una segunda cuerda corta (aquí, coloreada de azul) a la cuerda fija con un nudo prussik (B). Haga una lazada en el extremo inferior de la segunda cuerda corta e introduzca en ella el pie.

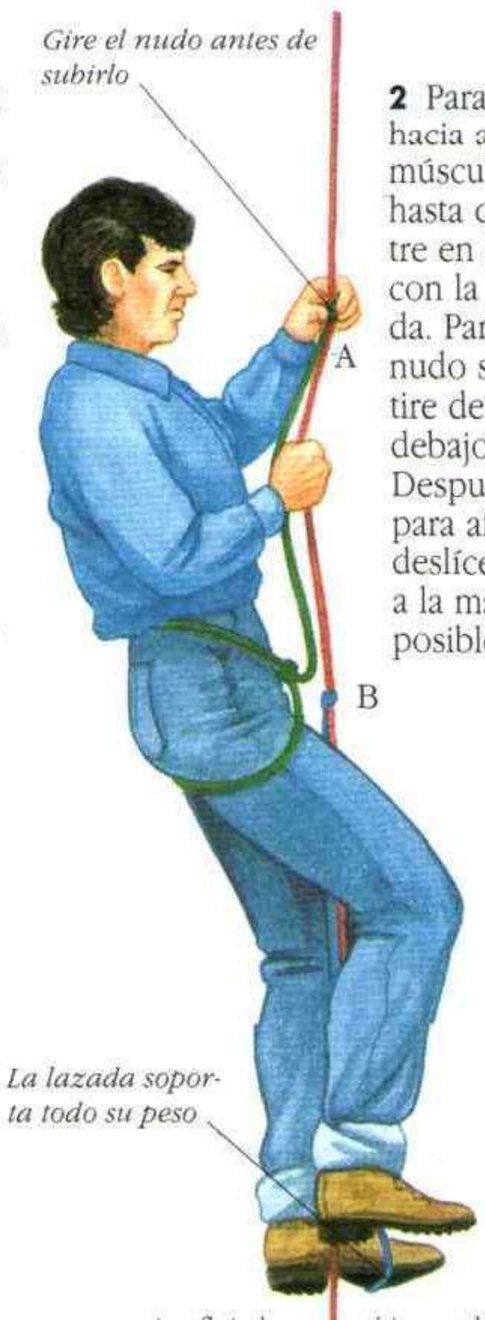

2 Para subir, empuje hacia arriba con los músculos de la pierna hasta que se encuentre en posición vertical con la pierna extendida. Para mover el nudo superior (A), tire de la cuerda fija debajo del nudo. Después gire el nudo para aflojarlo y deslícelo hacia arriba a la mayor distancia posible.

3 Muévase lentamente hasta sentarse y deje que el nudo doble de tejedor en el arnés soporte su peso. Al tensar la cuerda fija bajo el nudo más bajo (B), puede mover ese nudo hacia arriba, dejándolo listo para volverse a incorporar y ascender un nuevo tramo.

* El prussik no es un nudo autobloqueante. De la misma manera que es necesario aflojarlo para subir, puede ser necesario apretarlo para que soporte el peso sin deslizarse hacia abajo. El diámetro de la cuerda que forme el prussik debería ser inferior al de la cuerda fija.

CRUZAR LAS AGUAS

LOS RÍOS Y las corrientes son siempre peligrosos, desde sus cabeceras, o fuentes, donde son rápidos, angostos y poco profundos, hasta las profundidades tranquilas de sus cuencas más bajas. Aun si el agua aparenta ser tranquila y de movimiento lento, poco profunda y segura, suponga siempre que hay algún peligro oculto. Ni en las aguas más claras podrá ver todo lo que hay debajo de la superficie. Las aguas de cabecera son más fáciles de cruzar que las aguas profundas, pero tenga cuidado al atravesar aguas rápidas. Las aguas tranquilas son generalmente más profundas que las rápidas, y pueden contener algas traicioneras, bancos de fango y otros obstáculos ocultos muy peligrosos. No vadee ni nade si existe alguna otra opción más segura, y pregúntese siempre si es realmente necesario cruzar.

DÓNDE CRUZAR

El lugar más indicado para cruzar cualquier corriente de agua es un puente, un pontón o un transbordador. Así pues, antes de mojarse, compruebe las inmediaciones, tanto río arriba como río abajo. Es posible que encuentre un puente, o un tramo ancho y regular donde el lecho del río sea firme, y por donde pueda cruzar con toda seguridad. Compruebe que la orilla opuesta no sea demasiado escarpada.

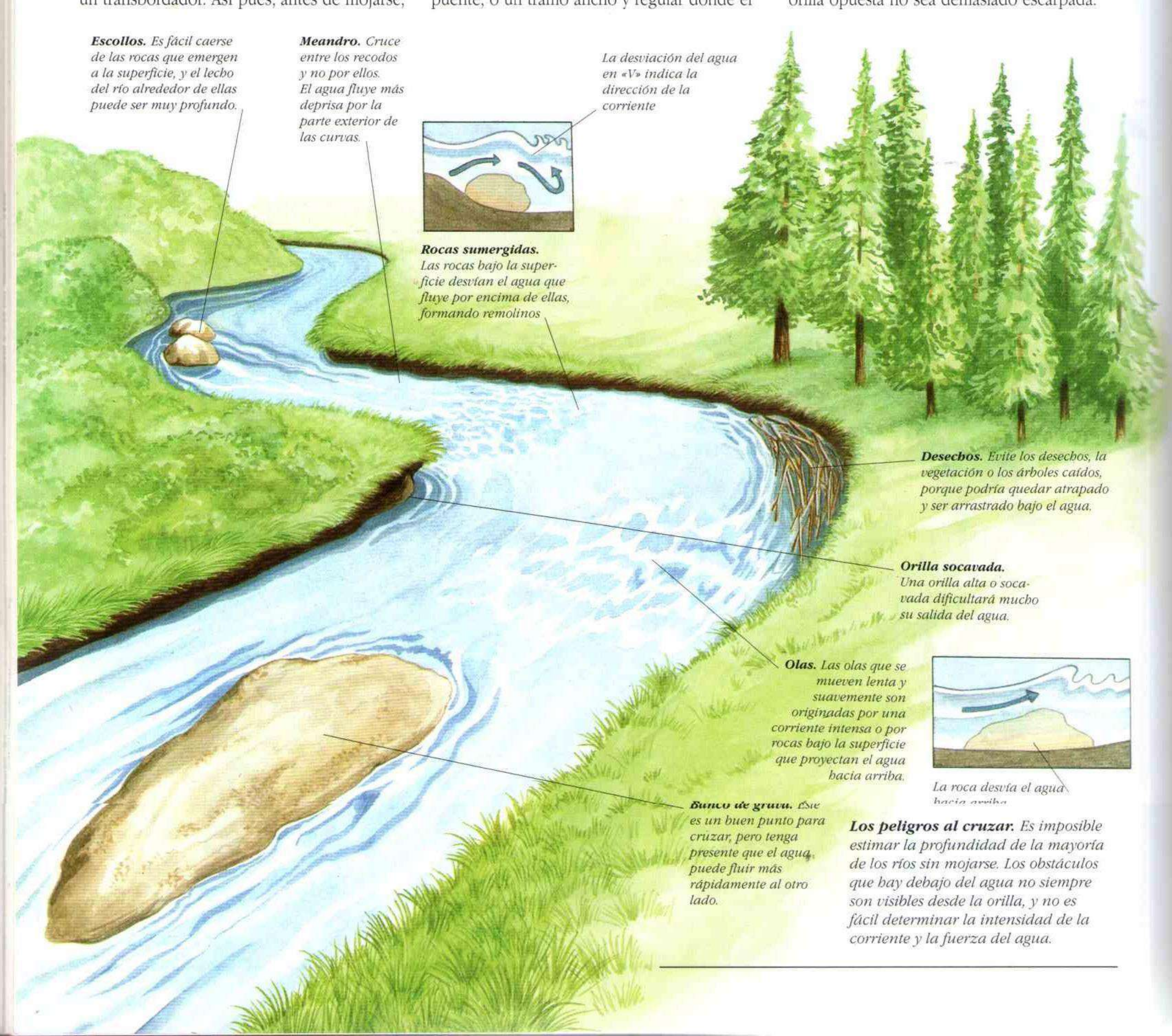

Los peligros al cruzar. *Es imposible estimar la profundidad de la mayoría de los ríos sin mojarse. Los obstáculos que hay debajo del agua no siempre son visibles desde la orilla, y no es fácil determinar la intensidad de la corriente y la fuerza del agua.*

VADEAR

Lleve algún tipo de calzado para proteger sus pies y tener un mejor punto de apoyo. Esté preparado para afrontar un fondo lodoso, con vegetación, o un cambio repentino en la profundidad del río. Observe el comportamiento de las olas y si hay algún signo de obstrucción bajo el agua *(véase página anterior)*. Cruce siempre muy despacio.

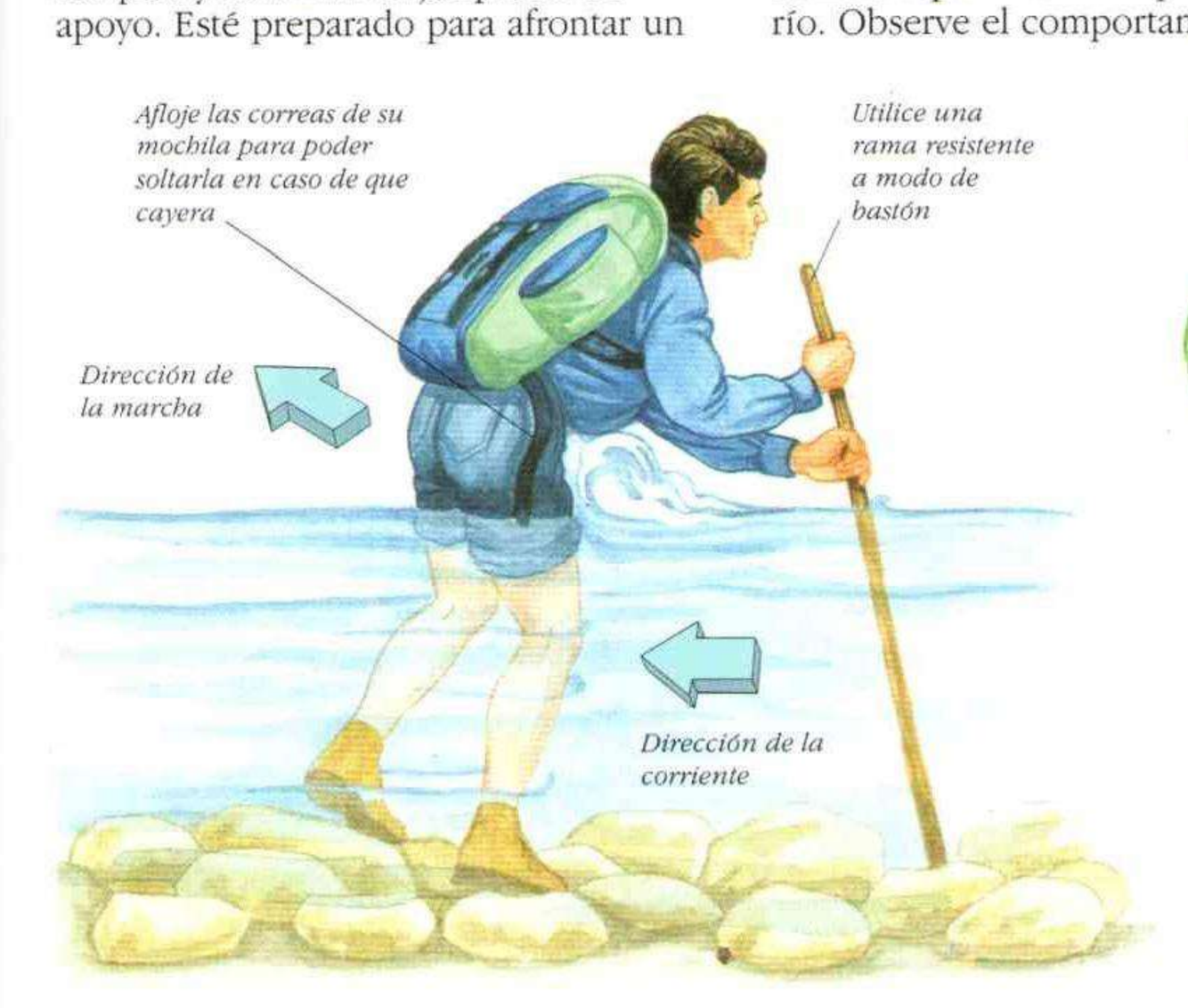

Cruzar solo. *Utilice un palo a modo de sonda y, una vez en el agua, como si se tratase de una tercera pierna, para mantener una forma de «T» extendida, como un trípode. Coloque el bastón corriente arriba y apóyese en él al dar un paso lateral con respecto a la corriente, volviendo a pisar firmemente el lecho del río. Dé pasos cortos, arrastrando los pies, para asegurar que la corriente no empuje su pierna hacia atrás y le haga caer.*

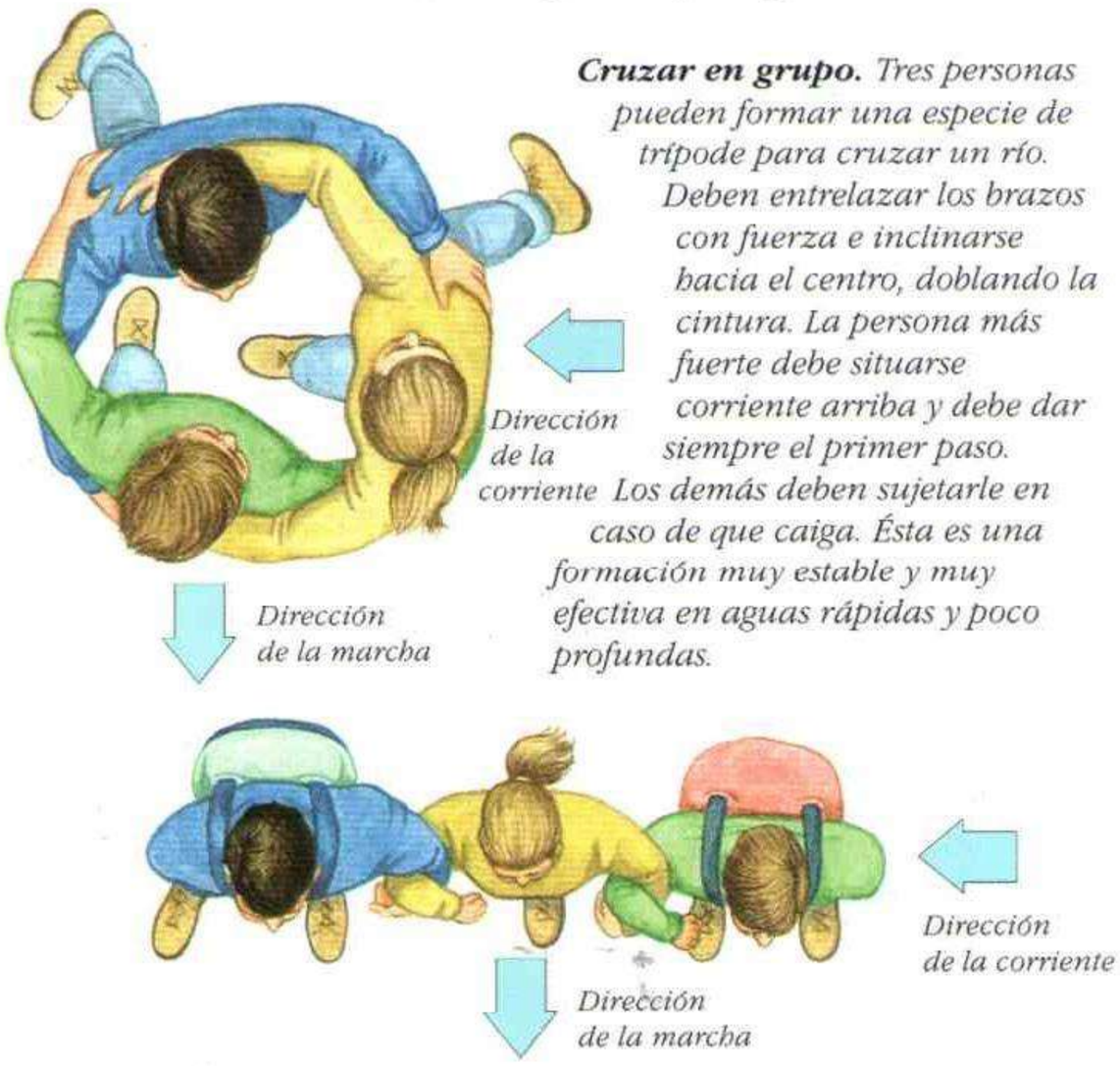

Cruzar en grupo. *Tres personas pueden formar una especie de trípode para cruzar un río. Deben entrelazar los brazos con fuerza e inclinarse hacia el centro, doblando la cintura. La persona más fuerte debe situarse corriente arriba y debe dar siempre el primer paso. Los demás deben sujetarle en caso de que caiga. Ésta es una formación muy estable y muy efectiva en aguas rápidas y poco profundas.*

Cruzar en línea. *Un grupo de personas puede cruzar la corriente en línea. La persona más fuerte debe colocarse corriente arriba, y las restantes, dando estabilidad y sujetando a cualquiera que pudiera caer. El líder debe decidir por dónde cruzar y dar los primeros pasos. Los demás deben entrelazar sus brazos con los de él; la persona más débil y menos pesada debe colocarse en el centro. Hay que cruzar despacio y con cuidado. Debe mantener el equilibrio y moverse paso a paso.*

NADAR

Si el agua es muy profunda y no puede vadearla, será necesario cruzar a nado. Fabrique un flotador que le ayude. Antes de entrar en el agua, busque un sitio adecuado para desembarcar en la orilla opuesta.

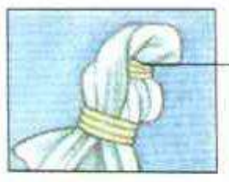

Retuerza la parte superior de la bolsa dejándola bien apretada, dóblela sobre sí misma y átela con fuerza

1 Quítese la ropa para que se mantenga seca, y colóquela junto con su equipo en una bolsa impermeable.

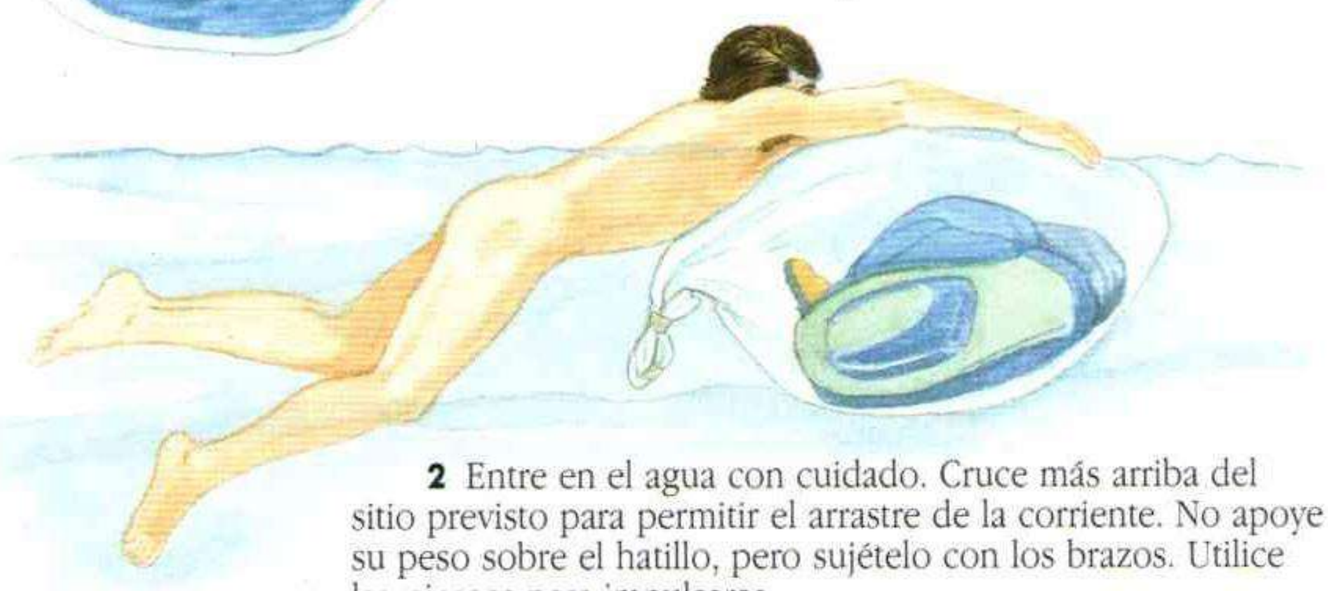

2 Entre en el agua con cuidado. Cruce más arriba del sitio previsto para permitir el arrastre de la corriente. No apoye su peso sobre el hatillo, pero sujételo con los brazos. Utilice las piernas para impulsarse.

HIPOTERMIA

Desnúdese (salvo el calzado) antes de entrar en el agua. Evalúe la conveniencia de llevar alguna prenda impermeable o resistente al viento para darle algo de protección. El frío le producirá letargo. Inmediatamente después de cruzar, séquese y vístase con ropa de abrigo seca y, si es posible, prepárese una bebida caliente y azucarada. Auxilie a quien esté temblando. Trabaje en parejas. El líder debe comprobar si en el grupo hay síntomas de hipotermia, y debe estar preparado para dar el tratamiento adecuado. Si quien la sufre no recupera el calor inmediatamente, puede llegar a morir *(véase página 163)*. La mejor manera de proporcionar calor a una persona es metiéndola en un saco de dormir con otra persona. Puede darse a la víctima una bebida caliente y ropa de abrigo, además de palabras tranquilizadoras, para que recupere el calor corporal. Los síntomas de hipotermia aparecen habitualmente en el siguiente orden:

- Temblores, carne de gallina, piel pálida y entumecida.
- Apatía, confusión, comportamiento irracional, amnesia, incoherencia y agresividad.
- Letargo seguido de arranques de energía frenética.
- Lapsos de consciencia con respiración lenta y poco profunda, arritmia cardíaca.
- Pulso lento y débil. Puede llegar a producir un paro cardíaco *(véase página 176)*.

AGUAS PELIGROSAS

CRUZAR CUALQUIER TIPO de corriente de agua debe considerarse una acción peligrosa. Las fluctuaciones en las precipitaciones pueden cambiar la seguridad de cualquier punto de cruce. La única manera de evaluar la profundidad y la intensidad de la corriente es enviando a una persona a cruzarla primero, sujetándola con una cuerda de salvamento atada a la orilla y controlada por un miembro del grupo. Si se ve obligado a nadar, utilice siempre un flotador, y en los rápidos, zonas rocosas o de aguas muy turbulentas, lleve un casco protector de color brillante.

EL EMPLEO DE CUERDAS DE SEGURO

Si está solo y tiene que cruzar un río peligroso, únicamente puede emplear una cuerda de seguro pasándola alrededor de una roca o árbol y sujetándola con nudos fáciles de aflojar. Con dos o más personas pueden utilizarse las cuerdas de seguro para disminuir los posibles peligros. Utilice siempre una cuerda de seguro, y asegúrese de que las personas que estén en el agua no se enreden en la parte floja de la cuerda.

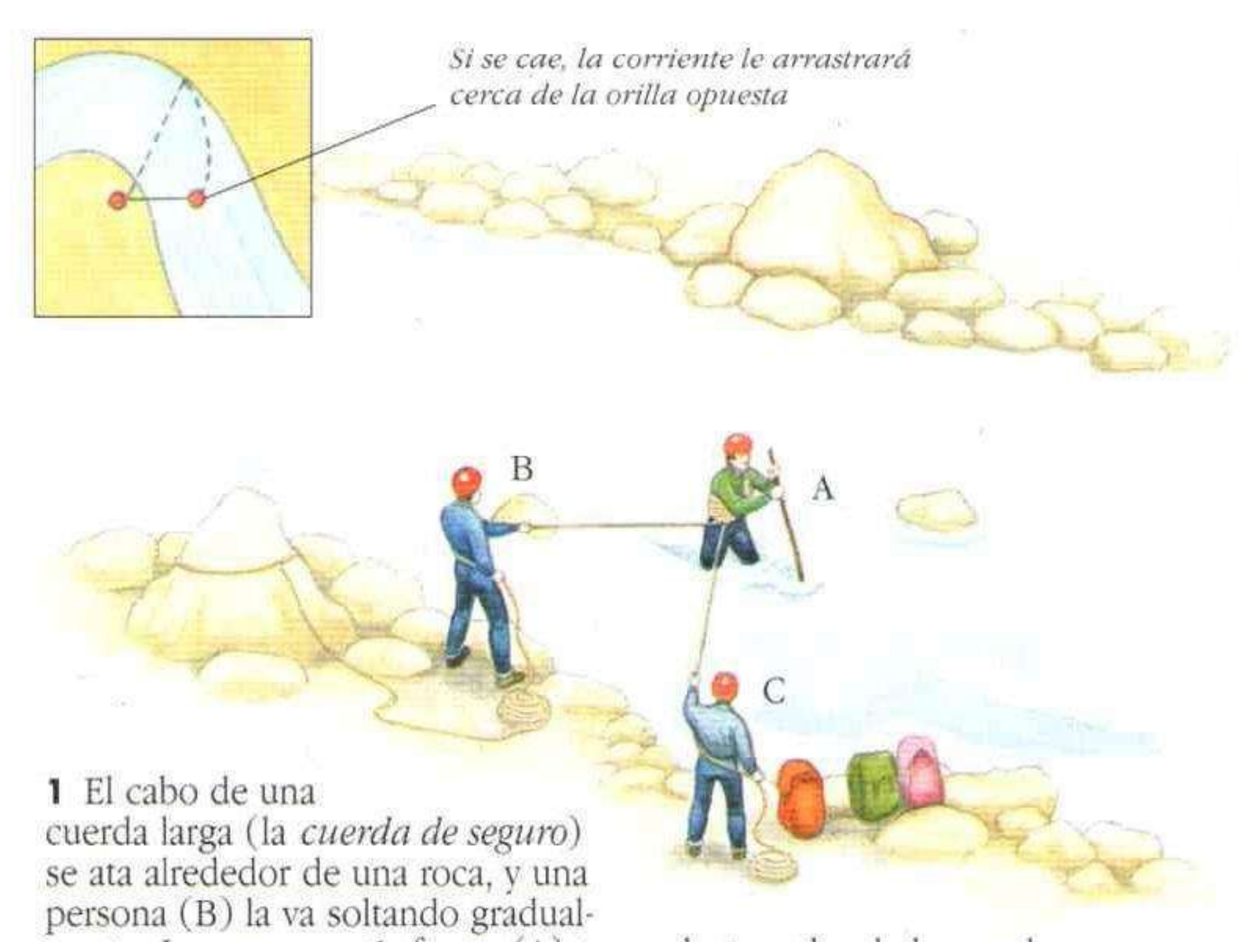

1 El cabo de una cuerda larga (la *cuerda de seguro*) se ata alrededor de una roca, y una persona (B) la va soltando gradualmente. La persona más fuerte (A) toma el otro cabo de la cuerda y cruza el río. Asimismo, lleva consigo el cabo de una segunda cuerda (la *cuerda de cruce*), cuyo otro extremo está sujeto por una tercera persona (C). Esta cuerda lleva un mosquetón en su centro *(véase página 147)*. La persona A lleva un casco y botas, y utiliza un palo para sondear el camino.

2 Cuando la persona A llega a la orilla opuesta, ata el cabo de la cuerda de seguro a una roca. Sujeta el mosquetón a la cuerda de seguro y ata el cabo de la cuerda de cruce alrededor de sí mismo para sujetar a la siguiente persona en cruzar. El mosquetón se lleva a la primera orilla, y la persona B lo sujeta a su cinturón. Agarrándose a la cuerda de seguro, cruza lentamente en el lado que da río arriba, utilizando el palo como sonda. La persona C va soltando la cuerda de cruce, mientras que la persona A va tirando del otro lado.

3 Conforme cada persona va llegando a la orilla opuesta, desprende el mosquetón de su cinturón y lo sujeta a la cuerda de seguro. A continuación se recoge la cuerda de cruce hacia la orilla de partida. Las mochilas se llevan a la orilla opuesta de la misma manera, sujetando las correas al mosquetón. Es posible que tenga que tensar la cuerda de seguro para evitar que las mochilas se mojen.

4 La última persona en cruzar el río (C) desata la cuerda de seguro de la roca. Ata el cabo de esta cuerda alrededor de sí mismo y cruza el río, tanteando el terreno con un palo. Los miembros restantes del grupo sujetan la cuerda de cruce, preparados para una eventual caída.

CÓMO CONSTRUIR UN PUENTE

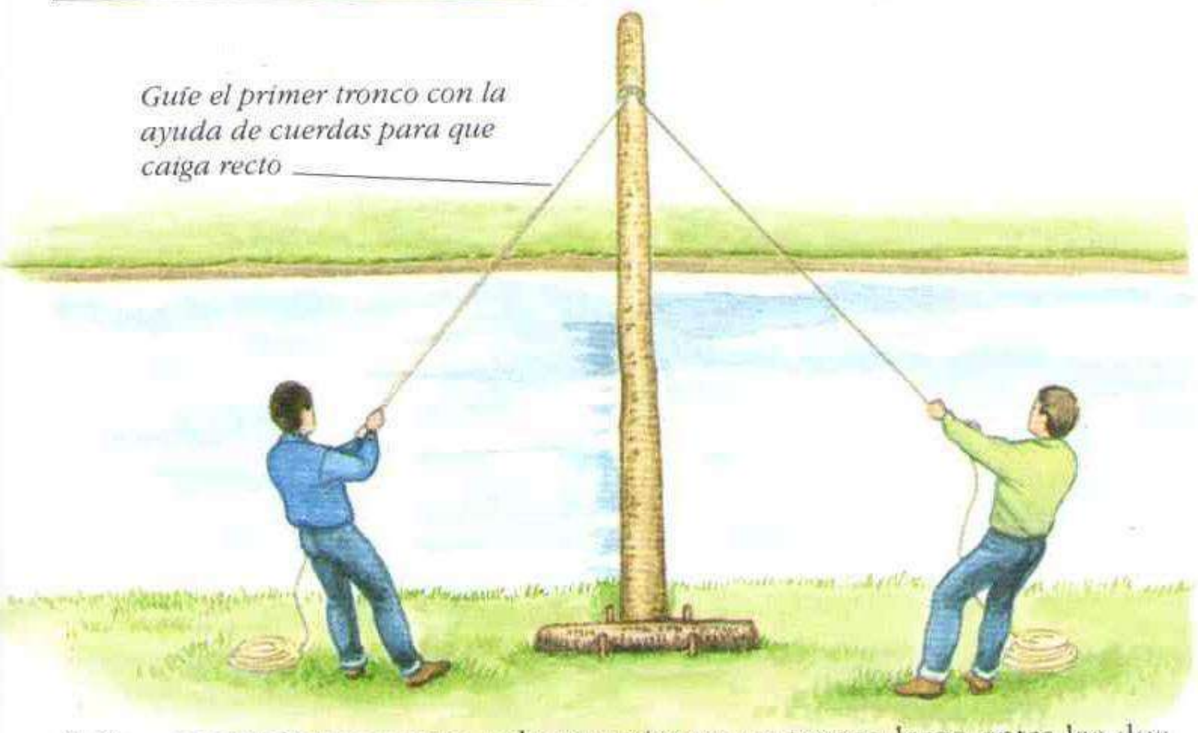

1 Para construir un puente, coloque primero un tronco largo entre las dos orillas con la ayuda de dos cuerdas. El extremo del tronco que permanece en la primera orilla debe asegurarse a otro tronco más corto, que se mantiene en su sitio gracias a unas estacas.

2 Una vez ha colocado el primer tronco, deslice un segundo tronco por encima del primero hasta que el extremo alcance la orilla opuesta. Haga caer el segundo tronco junto al primero y manténgalos juntos.

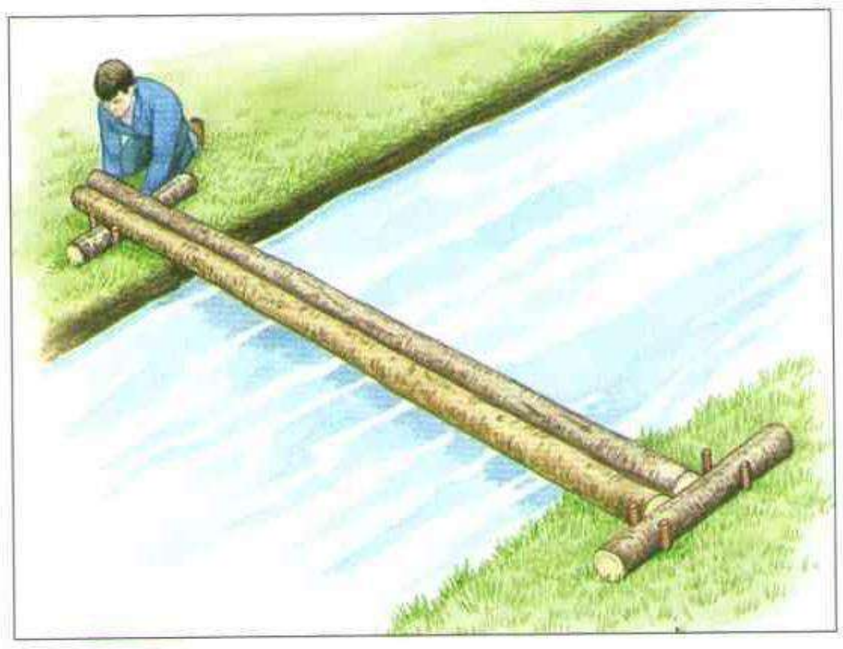

3 Cruce al otro lado sobre los dos troncos y coloque un tronco más corto bajo sus extremos para evitar que se entierren en el suelo. Fije los troncos con estacas para que no se separen.

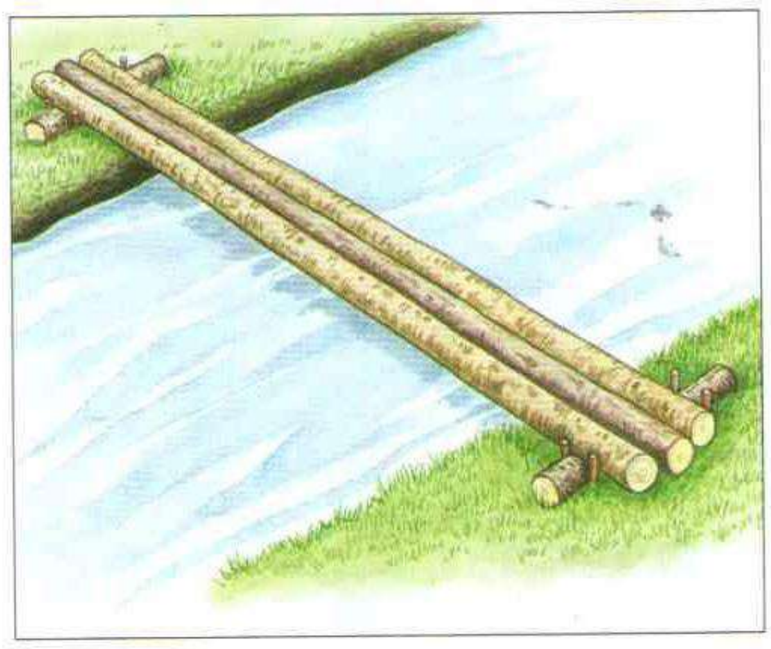

4 Deslice un tercer tronco junto a los otros dos. Coloque el tronco corto bajo sus extremos en la primera orilla y fíjelos con estacas como hizo en el paso 3. Puede hacer un puente tan ancho como desee, aunque con tres troncos suele ser suficiente.

COMO CRUZAR UN TERRENO PANTANOSO

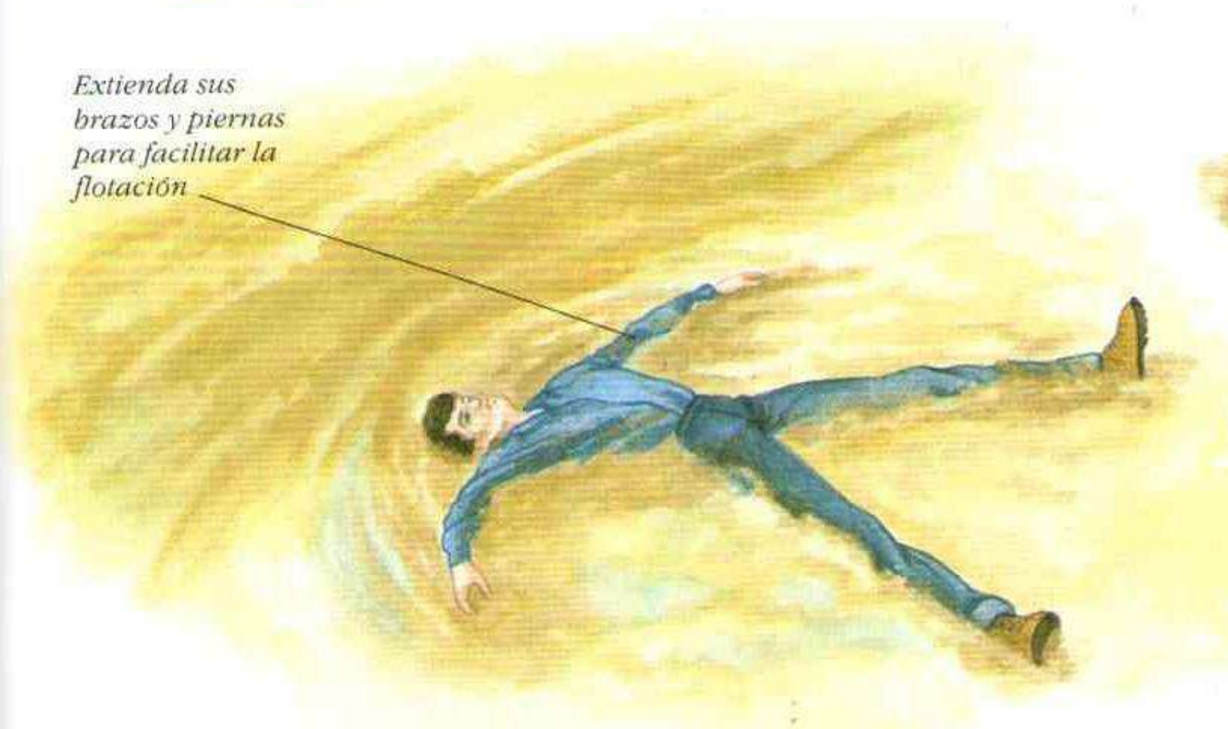

Si está solo. *Las arenas movedizas y los pantanos no son más que suelos con agua estancada, por lo que si cae en ellos debe ser capaz de flotar o cruzarlos a nado, con tal de que logre repartir su peso sobre la mayor superficie posible y así evitar hundirse. Trate primero de cruzarlos corriendo, pisoteando las matas de hierba. Si comienza a hundirse, suelte la mochila y flote sobre su espalda. Extienda los brazos y las piernas y nade lentamente hacia la orilla* (véase página 163). *No se asuste ni forcejee.*

Con ayuda. *Si no está solo cuando comienza a hundirse en las arenas movedizas o en un pantano, otra persona puede recostarse a orillas de la arena, pasarle el extremo de una rama larga y ayudarle a salir tirando de ella. Nuevamente, es importante que no forcejee ni se asuste, ya que los movimientos rápidos agitarán la superficie y le harán hundirse más rápidamente que si se limita a flotar o a moverse lentamente. Si se apoya sobre su mochila, repartirá su peso sobre una mayor superficie y evitará hundirse más.*

CONSTRUIR UNA BALSA

SI DISPONE DE provisiones y equipo abundante y tiene que cruzar un río, o si quiere navegar río abajo y no dispone de un barco, una balsa es la embarcación más sencilla que puede construir. Los troncos son los materiales tradicionales empleados en la construcción de una balsa, aunque puede utilizar también cañas de bambú o bidones de aceite. El principio básico de la construcción es hacer la menor cantidad posible de cortes en la madera de la embarcación para que permanezca a flote, y para que sea lo más estable posible. Una balsa nunca debe utilizarse en un río con rápidos, porque podría romperse y usted resultar herido. Si elige construir la balsa con maderos, es recomendable cortar árboles inclinados, ya que puede estimar con certeza dónde caerán. Si es posible, pruebe la balsa en aguas poco profundas antes de botarla en un río profundo.

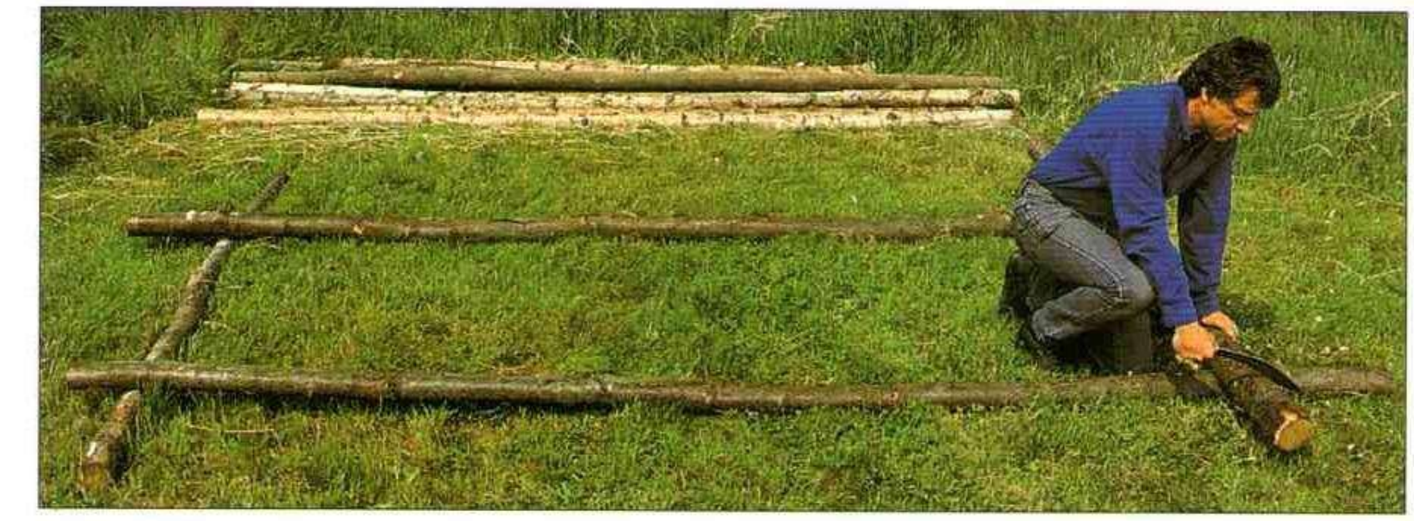

1 Para formar la cubierta, corte entre 12 y 14 troncos resistentes de la misma longitud. Corte seis troncos más, cada uno unos 30 cm más largo que la anchura prevista para la balsa. Éstos actuarán como retenes, manteniendo unida la balsa. Antes de unir los maderos de la cubierta, coloque dos troncos largos debajo de dos de los maderos más cortos para actuar como dispositivo de botadura *(véase página 156)*.

2 Con un cuchillo grande, corte una muesca a todo lo largo de cada uno de los troncos de retén, dejando unos 30 cm en cada extremo. Las muescas deben tener una profundidad de casi medio tronco. Tenga cuidado de no cortar con demasiada profundidad porque el tronco de retén podría ceder.

3 Asegúrese de que todas las muescas tengan la misma longitud y sean uniformes. Los maderos largos que formarán la cubierta se apoyarán sobre estos troncos de retén; la cubierta debe ser lo más plana y uniforme posible.

4 Coloque dos troncos con muesca, con la cara tallada hacia arriba, en extremos opuestos al dispositivo de botadura. Éstos actuarán como soporte base. Comience a colocar los maderos largos entre estos dos soportes para formar la cubierta de la balsa.

5 Coloque el extremo de cada uno de los maderos en las muescas de los troncos de soporte, trabajando desde el exterior hacia el centro de la balsa. Alterne el grosor de los maderos para que su peso esté bien repartido, alineándolos de manera que se ajusten bien entre sí.

6 El último tronco debe entrar bien ajustado, apretando a los otros troncos entre sí para que la balsa se mantenga unida. Una vez en el agua, tendrá que pisar los troncos para mantenerlos en posición, añadiendo ligaduras adicionales si es necesario. Es normal que algo de agua se filtre entre los troncos.

7 Coloque los dos últimos troncos con muesca, con la cara tallada hacia abajo, encima de los maderos de la cubierta. Éstos actuarán como retenes. Recuerde que cuando la balsa esté en el agua, los troncos flotarán hasta estos retenes, que servirán para mantener unidas todas las partes de la balsa.

8 Los dos troncos de retén en cada extremo, por encima y por debajo de la cubierta, deben atarse firmemente entre sí, tensándolos tanto como le sea posible. Mantenga los nudos a la vista para que, una vez la balsa esté en el agua, pueda comprobarlos y ajustarlos si es necesario.

9 Ate los troncos de retén a ambos lados de los troncos de la cubierta utilizando nudos planos *(véase página 35)*.

10 Cerciórese de que las ligaduras estén bien sujetas y correctamente anudadas. Si dispone de tiempo, puede hacer más ligaduras entre cada uno de los troncos.

11 Ate dos ramas de madera en forma de cruz. Debe ser lo suficientemente resistente para soportar el peso del timón.

Construya un soporte para el timón

12 Encaje la cruz en la parte posterior de la balsa, entre dos de los troncos de la cubierta, y asegúrese de que queda sujeta. Ajuste las ligaduras si es necesario. No la ate aún en esta etapa.

13 Construya un remo *(véase página 157)*. Éste actuará como timón y le permitirá guiar la balsa. Si quiere mantenerse de pie mientras gobierna, la balsa tendrá que alargar la caña del «timón» (el mango de la pala) añadiéndole una rama larga.

14 Compruebe que el timón se ajuste al soporte y átelo firmemente. Para evitar que el timón y el soporte se rompan al botar la balsa, retírelos de la misma. Sujételos en su sitio cuando la balsa esté en el agua. La balsa debe ser lo suficientemente grande como para llevarle a usted y su equipo al otro lado del río, o río abajo. Compruebe si la balsa flota antes de cargarla.

LAS BALSAS Y SU FORMA DE NAVEGAR

- En lugar de madera, puede utilizar gruesos tallos de bambú para construir la cubierta de la balsa. Deberá unir dos capas de bambú porque una sola no soportaría su peso. En lugar de atar retenes por encima y debajo de la balsa, puede hacer agujeros a ambos lados de los tallos y pasar unas ramas finas o cuerda a través de ellos para unir los tallos entre sí y formar la cubierta. Pase una cuerda alrededor de cada uno de los tallos como medida de seguridad adicional.
- No sabrá cómo flota su balsa hasta que la bote. Con algunos tipos de madera, una balsa cargada flotará ligeramente por debajo de la superficie del agua, aunque esto no constituye un problema importante.
- Una balsa virará de forma relativamente lenta. Téngalo en cuenta al maniobrar.
- Evite las colisiones con la orilla o los objetos grandes, ya que pueden romper las ataduras y desmontar la balsa.

EL EMPLEO DE BALSAS Y OTRAS EMBARCACIONES

EN MUCHAS ZONAS silvestres, la vegetación densa y el terreno escarpado hacen recomendable el viaje por río, lago o a lo largo de la costa. Sin embargo, recuerde que aventurarse en el agua puede resultar peligroso, en especial si lo hace en una barca improvisada. Debe asegurarse, por lo tanto, de que su embarcación sea la adecuada para las aguas que pretende cruzar, sea improvisada o no. El estado del agua puede variar rápidamente, convirtiendo una travesía segura en peligrosa. Por esta razón, no debe aventurarse a navegar a menos que tenga experiencia o no le quede otra opción. Lleve siempre algún tipo de chaleco salvavidas.

CÓMO BOTAR UNA BALSA

Una balsa sólo deberá utilizarse en aguas tranquilas. Constrúyala cerca del agua, ya que podría resultar ser muy pesada para arrastrarla hasta la orilla. Antes de meter la balsa en el agua, átela a un objeto inamovible, dejando suficiente cuerda libre para permitir botarla.

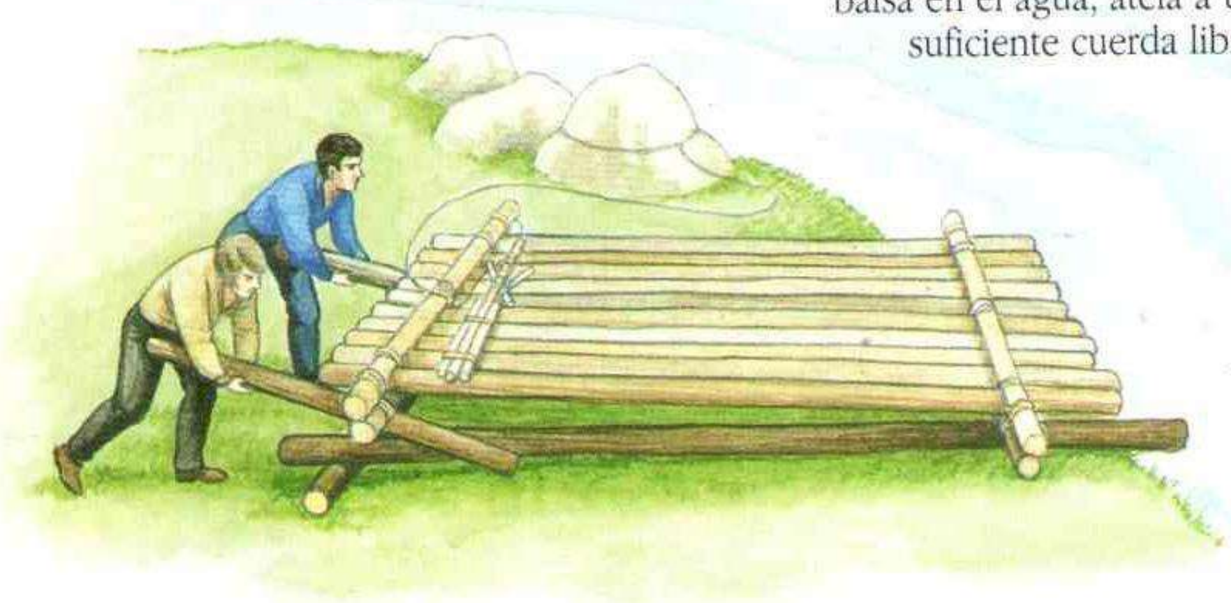

Haciendo palanca. *Bote la balsa haciendo palanca con unos troncos en uno de los lados hasta lograr deslizarla al agua.*

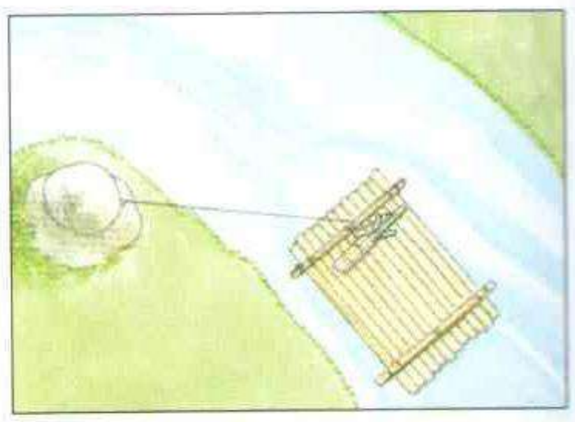

A flote. *Una vez ha comprobado que la balsa no se hunde en el agua, puede sujetar el timón y comenzar a cargar su equipo en la parte central.*

APAREJAR UNA BALSA

Para travesías prolongadas a lo largo de un río recto, levante un mástil y una vela para aprovechar el viento al máximo —siempre y cuando el viento sople en la dirección adecuada. Sin una quilla, el timón únicamente controlará la dirección si el viento es muy suave, pero le permitirá mantener la balsa recta. Si la proa se hunde en el agua, reduzca la superficie de la vela. Únicamente podrá moverse río abajo siguiendo la corriente.

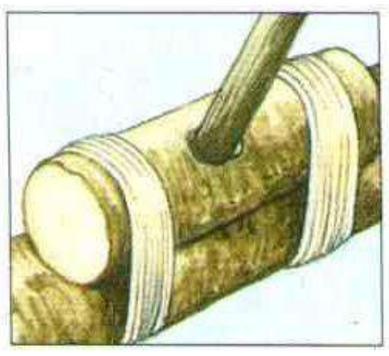

Pequeños troncos de madera en los cuatro costados de la balsa permiten la talla de muescas para insertar palos de refuerzo

Un cabo sujeto al mástil permite izar o arriar la vela según el viento que sople

Los cabos evitan que la verga baja gire

Utilice una lona o tela impermeable como vela

Los postes mantienen el mástil vertical

Un remo sujeto a la popa hace las veces de timón

Los soportes del mástil se insertan en las muescas y se sujetan con fuerza

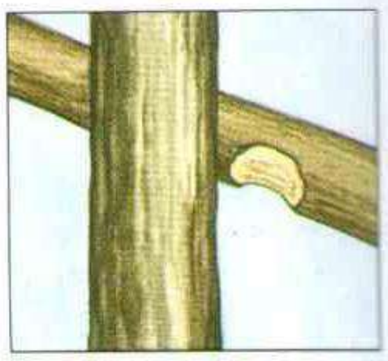

La verga alta tiene una muesca que permite subirla o bajarla empleando el cabo que pasa por encima del mástil, y así reducir la vela

Levantar un mástil con vela. *Sujete un mástil largo entre cuatro troncos y asegúrelo con fuerza. Refuerce con cuatro palos, cada uno ligado a la cubierta. Utilice una manta o una lona impermeable como vela.*

La elección de la embarcación

Cualquier tipo de embarcación menor tiene sus ventajas y sus inconvenientes, pero nunca debe utilizarse para un propósito inadecuado, en condiciones que vayan más allá de sus posibilidades. Algunas embarcaciones, como el coracora, son aptas para las aguas tranquilas, mientras que otras, como las neumáticas modernas, pueden utilizarse en condiciones inadecuadas para otro tipo de embarcaciones.

Neumática. *Una neumática puede utilizarse en condiciones relativamente adversas, y es la embarcación preferida por muchos científicos y exploradores. Resulta manejable incluso en mares agitados. Sin embargo, el hielo superficial y otros obstáculos pueden perforar los compartimientos de flotación. Cuando utilice una neumática deberá mantener el peso en la proa para evitar que la barca se levante hacia atrás.*

Coracora.
Un coracora requiere una gran habilidad para su construcción y manejo. Consiste en un armazón de maderas flexibles en forma de media esfera, cubierto con piel o lona impermeable. Un coracora únicamente es recomendable para las aguas muy protegidas y tranquilas.

Motor fuera borda. *Una embarcación de fondo plano con un motor fuera borda es ideal para los ríos anchos de flujo uniforme. Es capaz de cubrir grandes distancias. Un motor fuera borda sobrecarga la popa de una embarcación pequeña, en particular durante el avance, por lo que debe equilibrarse con un peso en la proa. La longitud del eje de cola reducirá el calado del barco hasta cierto punto.*

Kayak. *El tradicional kayak esquimal ha servido como modelo para la construcción de las resistentes canoas modernas capaces de sortear las aguas más rápidas. Las embarcaciones de piel o tela no son tan robustas como las versiones en fibra de vidrio, pero son más ligeras y pueden gobernarse fácilmente por los rápidos demasiado difíciles de pasar en barca. Se necesita una gran habilidad para controlar un kayak, en particular cuando va cargado con el equipo, o cuando se utiliza en aguas bravas. Al carecer de quilla, vuelcan fácilmente.*

Consejos para ir en barca

- Lleve siempre un chaleco salvavidas, o improvise uno llevando botellas de agua vacías en sus bolsillos.
- Ate cualquier artículo de su equipo a la embarcación con una cuerda.
- No sobrecargue la embarcación.
- Explore el río que pretende recorrer antes de botar su embarcación.
- Si tiene un motor fuera borda, acelere hasta que la barca adquiera una posición horizontal, y reduzca después para mantener la marcha.

Cómo construir un remo

1 Para construir un remo, elimine todos los ramos de una rama fuerte y verde con un cuchillo. Talle uno de los extremos en forma de cuña.

2 Ate dos ramas más cortas y rectas a ambos lados de la cuña, que actuarán como pala del remo.

3 Ate una tercera rama en la parte central, forzándola en el hueco formado por las otras dos, y apoyándola en la cuña al final del mango.

Remo terminado. *Es posible que tenga que volver a atar el remo cuando se humedezca, ya que la cuerda puede aflojarse.*

A Guide to Medicinal Plants

An Illustrated, Scientific and Medicinal Approach

Koh Hwee Ling
Chua Tung Kian
Tan Chay Hoon

World Scientific

1. *Abrus precatorius* L. (Leguminosae)

Rosary Pea, Indian Licorice, Precatory Bean

Seeds of *Abrus precatorius*

Description: *Abrus precatorius* L. is a perennial climber with a slender stem. Leaves are pinnate and 5–8 cm long. Leaflets are rhomboid, numbering 20–24 or more, opposite and are 1.2–1.8 cm long. Leaf margin is entire. It bears pink flowers arranged in dense axillary racemes. Pods are oblong, cylindrical, inflated, 5–6 cm by 1 cm and contains 3–6 round, glossy, black and red seeds.[1–3]

Origin: Native to Pakistan, India, Ceylon and tropical Africa; and introduced widely in the New and Old World.[4]

Phytoconstituents: Abrectorin, abricin, abridin, abrins A–D, (+)-abrine, abruslactone A, abrusgenic acid, abrusogenin, abrusoside A–D, precatorine, abruquinones, abraline, abrusic acid, abruquinone G and others.[2,5–18]

Traditional Medicinal Uses: A decoction of the leaves has been prescribed for scurvy, cough, bronchitis, sprue and hepatitis and as a refrigerant. They are also

b716_Ch_01-10.indd 2 1/8/2009 10:08:49 AM

applied on painful swellings, eye inflammation, cancer, syphilis and on leucodermic spots.[19] The leaves are also effective in the treatment of coryza, cough, fever, and jaundice resulting from viral hepatitis and intoxications.[6] The seeds have been used to treat fever, malaria, headache, dropsy and to expel worms.[3] A decoction of the seeds is applied for abdominal complaints, conjunctivitis, trachoma and malarial fever.[2] Central Africans use powdered seed as an oral contraceptive.[5] It is also used to lower high blood pressure and relieve severe headache.[5] The seeds are very toxic and can be applied externally to treat bacterial infection and accelerate the bursting of boils and to cure mastitis and galactophoritis.[6] The seed has purgative properties and is used as an emetic, tonic, aphrodisiac, and for nervous disorders. The poultice can be used as suppository, abortifacient, or tonic for pregnant women and children and to treat severe headaches.[19] Water from the boiled roots is used to cure cough, bronchitis, sore throat and also applied as an emetic agent.[2,5]

Pharmacological Activities: Antibacterial,[20] Anthelmintic,[21] Antiviral,[17] Anti-inflammatory,[10,22,23] Anticancer/Antitumour,[11,24–30] Antiplatelet,[23] Antiprotozoal,[31] Immunomodulatory,[27,32–35] Antioxidant,[23] Antiplasmodial,[17] Antitubercular[17] and Molluscicidal.[36]

Dosage: In Central Africa, 200 mg powdered seeds are used as an oral contraceptive which can last 13 menstrual cycles.[5] Approximately 14 g of powdered seeds are used as tonic for pregnant women and children.[19] About 5–7 of seed grains are prescribed for pertussis.[37]

Adverse Reactions: Ingestion of *Abrus* seeds resulted in pulmonary oedema and hypertension.[38] Abrine can cause coma, confusion, convulsions, dehydration, gastroenterosis and hypotension.[37]

Toxicity: The LD_{50} of abrin in mice is 0.02 mg/kg body weight.[2] Ingestion of seeds causes severe stomach cramping accompanied by nausea, severe diarrhoea, cold sweat, fast pulse, organ failure, coma and circulatory collapse.[2,5,39] Oral administration of a 50% ethanol extract of *A. precatorius* seeds (250 mg/kg) in albino rats for 30 and 60 days induced infertility in males which was reversible.[40] Dose-dependent degenerative changes in the testicular weights, sperm count, spermatogenesis and Leydig cells were observed in testes of rats treated with steroidal fraction of the seeds.[41] The methanol extract of the seeds caused a concentration-related impairment of percentage human sperm motility with an EC_{50} of 2.29 mg/ml.[42]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

2. *Adiantum capillus-veneris* L. (Polypodiaceae)

Black Maidenhair Fern, Southern Maidenhair Fern,
Venus Hair Fern

Adiantum capillus-veneris fern

Description: *Adiantum capillus-veneris* L. is a perennial fern with short creeping stems. Leaf blades are lanceolate, pinnate, 10–45 cm by 4–15 cm and glabrous. Ultimate segments are various but generally cuneate or fan-shaped to irregularly rhombic, about as long as well as broad with its base broadly to narrowly cuneate. Plant is delicate, brittle and has dark stalks.[1–6]

Origin: Native to America, Mexico, West Indies and South America. It can also be found in temperate regions of Eurasia and Africa.[6]

Phytoconstituents: Adiantoxide, adiantone, isoadiantone, isoadiantol, hydroxyadiantone, capesterol and others.[1,4,7–11]

b716_Ch_01-10.indd 4 1/8/2009 10:08:52 AM

Traditional Medicinal Uses: The fern is considered an astringent, demulcent, depurative, diaphoretic, diuretic, emmenagogue, emollient, expectorant, laxative, stimulant, sudorific and a tonic.[1] It is also used for alopecia, asthma, bladder ailments, catarrh, chest ailments, chills, cold, dropsy, dysmenorrhoeal, fever, head ailments, hepatosis, labour, lung ailments, respiratory problems, rheumatism, sclerosis, snake bite, sores, sore throat, splenosis, stones and other urinary calcification.[1,4] The plant expels worm, induces vomitting and relieves fever. Used externally, it is poulticed on snakebites and as a treatment for impetigo.[1,3] The Russians use the herb for rhinitis while the French, with orange flowers and honey, uses it for pulmonary catarrh.[1] A handful of leaves is made into a tea and drunk as an expectorant, astringent, tonic for coughs, throat afflictions and bronchitis.[1] The plant is also used as a hair wash for dandruff and to promote hair growth in Latin America.[1,3] The ashes mixed with vinegar and olive oil is rubbed into the scalp to cure alopecia.[5] In Traditional Chinese Medicine, the leaves are used for bronchial diseases and as an expectorant;[3] and in Africa, the leaves are smoked to prevent head and chest colds.[1]

Pharmacological Activities: Antibacterial.[12]

Dosage: Taken internally as a tea prepared from powdered dried fronds. The standard single dose is 1.5 g of drug to 1 cup of fluid per dose.[13]

Adverse Reactions: No known side effects with therapeutic dosages.[13]

Toxicity: No information as yet.

Contraindications: Should not be used during pregnancy.[13]

Drug-Herb Interactions: No information as yet.

3. *Allamanda cathartica* L. (Apocynaceae)

Allamanda, Common Allamanda, Golden Trumpet

Allamanda cathartica flowers and leaves

Description: *Allamanda cathartica* L. is a woody shrub that can grow up to 4 m tall. The stems exude milky white sap when incised. Leaves are simple, exstipulate, glossy, leathery and glabrous. Leaf blade is oblong-lanceolate, 8–15 cm by 4–5 cm and arranged in opposites of 3–5 sessiles. Flowers are large, tubular, bright yellow and 4–5 cm long.[1–3]

Origin: Native to South America; cultivated in China for medicine.[4]

Phytoconstituents: Allamandin, plumericin, plumieride, ursolic acid and others.[3,5–10]

Traditional Medicinal Uses: The plant has been used as a purgative to induce vomiting at low dosage.[3] Its leaves are cathartic and the bark is used as a hydragogue for ascites.[11] In Surinam's traditional medicine, its roots are used against jaundice, for complications with malaria and enlarged spleen.[11]

b716_Ch_01-10.indd 6 1/8/2009 10:08:53 AM

FA

Pharmacological Activities: Anthelmintic,[12] Antifungal,[7,13–15] Antineoplastic,[16] Antivenom[17] and Wound healing.[18]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: Every part of the plant was reported to be poisonous.[19] The sap of the plant was reported to cause mild and occasional oral irritation and slight nausea when sucking cut stems. Rash or dermatitis were also reported when sap was in contact with sensitive skin.[19]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

b716_Ch_01-10.indd 7 1/8/2009 10:08:57 AM

4. *Aloe vera* Mill. (Aloaceae)

Aloe, Lidah Buaya

Aloe vera

Young and old plants

Description: *Aloe vera* Mill. is a short-stemmed, up to 50 cm, succulent herb with thick green leaves that have a sharp and pointed apex, 15–50 cm by 4–7 cm, arranged in a rosette around the short stem. Blade is green to variegate with small white or glaucous dots, irregular bands, lanceolate, tapering from base to apex, glabrous with green and spiny-toothed margins. The leaves contain a thick colourless juice. Flowers are yellow, orange or red, crowded into a rosette and in panicles.[1–3]

Origin: Native to North Africa; cultivated in China for medicine and widely used as indoor ornamental plants.[3,4]

Phytoconstituents: Aloin (barbaloin), arabinose, aloe-emodin, aloetinic acid, emodin, aloeresin A–C, aloesone, aloeride and others.[2,5–12]

Traditional Medicinal Uses: The plant has been used in cosmetic preparations for the treatment of pimples, acne and mouth ulcers.[2] It has also been used to control bleeding, itching of piles, and relief from arthritic pains.[2] The Chinese uses the plant juice as a mild laxative, wash for piles, abscesses and scabies. In the Philippines, it is used to treat dysentery and pain in the kidneys.[2] The plant has been found to treat bacterial infection, as a cathartic, emmenagogue, purgative and vermifuge. It can be used in the treatment of burns, oedema, pain, swellings and wounds; treatment of leukemia, lung cancer; treatment of constipation, eczema, piles and pertussis.[6,13–15] The whole

b716_Ch_01-10.indd 8 1/8/2009 10:08:57 AM

plant has also been used for the treatment of rectal fissures and piles while the root is used to treat colic.[16] The juice from the leaves is used to increase menstrual flow.[13] The jelly is used as aperient, for wounds, applied on the abdomen in fever, after confinement, on swelling and is especially useful in correcting constipation due to intake of iron medication.[17,18] Fresh juice of the leaves is cathartic and cooling and used for various eye diseases.[16] The dried juice is applied with lime juice for reducing swellings and promoting granulation in ulcers. In Malaysia, it is used for treating wounds, fever, swellings and put on the abdomen of women after confinement. The mixture of sugar with sap obtained from heated leaves is taken for asthma. The mucilaginous flesh and the sap are used for burns. The watery extract is used as a hair tonic. It is also used in cosmetics for decreasing wrinkles and other skin problems. It is mixed with milk and given for dysentery and pains in the kidney.[15] It is used in Ayurveda to alleviate pain and is also mentioned in folk medicine of Arabian Peninsula for the management of diabetes.[19]

Pharmacological Activities: Angiogenic,[20] Antifungal,[21–23] Antidiabetic,[24,25] Anti-inflammatory,[12,26–30] Anticancer,[31–38] Antimicrobial,[39] Antioxidant,[11,40–45] Antiproliferative,[12,38] Chemopreventive,[35,46] Gastric mucosal protection,[47,48] Hepatoprotective,[19,49,50] Neuroprotective,[51] Hypolipidaemic,[52] Immunomodulatory,[11,33,53–57] Immunostimulatory,[58] Antimutagenic,[31] Alloantigenic,[59] Antileishmanial,[60,61] Prevention of kidney stones,[62,63] Radioprotective[64,65] and Wound healing.[66–69]

Dosage: Single dose of powdered Aloe, 50–200 mg at bedtime; tincture BPC 1949 (1:40, 45% ethanol), 2–8 ml. Aloes should only be taken for short periods, maximum 8–10 days.[70] Doses of 10–30 mg act as a bitter stomachic; 60–200 mg as a laxative and 300–1000 mg as a purgative.[18,71] A dose of 1 teaspoon after meals, or otherwise advised by manufacturers and practitioners has been reported.[72] To prevent kidney stones, a dose of 2 to 3 tablespoon daily is reported. As a laxative, the recommended dose is 500 to 1000 mg daily. For burns or wound healing, fresh gel from plant may be applied topically and liberally. For haemorrhoids, as a stool softener, 0.05 to 2 g of dry aloe extract is administered. In the treatment of HIV, 800 to 1600 mg of acemannan daily (equivalent to 0.5 to 1 L of *Aloe vera* juice) is administered. To relieve constipation, 20 to 30 mg hydroxyanthracene derivatives daily, calculated as anhydrous aloin is prescribed.[73]

Adverse Reactions: Barbaloin was shown to have a laxative effect.[70] Ingestion of *A. vera* is associated with diarrhoea, electrolyte imbalance,

FA

kidney dysfunction, and conventional drug interactions; episodes of contact dermatitis, erythema, and phototoxicity have been reported from topical applications.[74] *A. vera* could also induce acute liver damage.[75]

Toxicity: Severe gastrointestinal cramping can occur if the latex (which is just below the leaf surface) is taken internally.[72] Toxic doses cause severe haemorrhagic diarrhoea and kidney damage, and sometimes death. The lethal dose of the dried plant extract is stated to be 1 g/day taken for a period of several days.[76]

Contraindications: Contraindicated in intestinal obstruction, acute inflammatory intestinal diseases (e.g., Crohn's disease, ulcerative colitis), appendicitis and idiopathic abdominal pain. Should not be used during pregnancy. Should not be given to children below 12 years of age.[77] Should avoid application of *Aloe* topically on deep, vertical wounds. Contraindicated in menstruation and if the person has kidney complaints.[72]

Drug-Herb Interactions: Increase the actions of cardiac glycosides and antiarrhythmic drugs (chronic use of aloe causes potassium loss), thiazide diuretics, loop diuretics, licorice and corticosteroids.[77] Aloe gel, when taken orally, can reduce the absorption of many medications. Thus, it should be taken two hours apart from all medications.[72] A study reported that *Aloe vera* preparations improved the absorption of both vitamins C and E.[78]

b716_Ch_01-10.indd 10 1/8/2009 10:09:02 AM

5. *Andrographis paniculata* (Barm.f.) Nees (Acanthaceae)

Hempedu Bumi, Sambiloto, Chuan Xin Lian

Andrographis paniculata plants

Andrographis paniculata flowers

Description: *Andrographis paniculata* (Barm.f.) Nees is an annual herb that grows up to 1 m in height. Stems are glabrous and articulated. Leaves are simple, opposite and exstipulate. Blade is dark green, bitter, glossy, simple, lanceolate, opposite and 4–8 cm by 1.3–2.5 cm. Its small and white flowers grow in terminal or axillary panicles. Both the bracts and the 5-lobed calyx are small. Fruits are upright, fusiform, capsular and contains 2–4 seeds.[1]

Origin: Native to Indian subcontinent and cultivated elsewhere.[2]

Phytoconstituents: Andrographolide, andropanolide, andrographic acid and andrographidine A, andrographatoside, andropaniculosin A and andropaniculoside A and others.[3–14]

Traditional Medicinal Uses: The plant is used orally to prevent and treat common cold, influenza, pharyngotonsilitis, allergies and sinusitis. Traditionally, it is used for many conditions including anorexia, atherosclerosis, insect and

b716_Ch_01-10.indd 11 1/8/2009 10:09:02 AM

snake bites, bronchitis, prevention of cardiovascular disease, diabetes, hypertension, cholera and as a tonic.[1,15,16]

Pharmacological Activities: Antiapoptotic,[17] Antibacterial,[18–20] Antifungal,[18] Anticancer,[21–30] Antidiabetic/Hypoglycaemic,[31–36] Antifertility,[37] Anti-inflammatory,[38–48] Antioxidant,[45,49–51] Antiplatelet,[52,53] Antiprotozoal,[54–56] Antiviral,[57,58] Cardioprotective,[59] Chemopreventive,[60] Hepatoprotective,[61–63] Hypotensive,[64] Immunomodulatory,[35,42,65,66] Psychopharmacological activities,[67] Vasorelaxant[68] and Cytotoxic.[12]

Dosage: For decreasing symptoms of common cold, doses of 400 mg of standardised andrographolide are required three times daily; for preventing common cold, a dose of 200 mg daily for 5 days in a week; for relieving fever and sore throat in pharyngotonsilitis, doses of 3 g and 6 g daily were used.[16] Use 6–9 g for influenza with fever, sore throat, ulcers in the mouth, acute or chronic cough, colitis, dysentery, urinary tract infection, carbuncles, sores and venomous snake bite.[69]

Adverse Reactions: Orally, large doses of Andrographis may cause gastrointestinal distress, anorexia, emesis and urticaria. Androgapholide taken orally at 5 mg/kg three times a day may cause headache, fatigue, rash, abnormal taste, diarrhoea, itching, lymphadenopathy, anaphylactic reactions, etc.[16]

Contraindications: Contraindicated in pregnancy, likely to be unsafe due to abortifacient effect.[16]

Toxicity: No toxic effect was observed after administration of a decoction of *Andrographis paniculata* leaves to rabbits.[70] LD_{50} of androgapholide in mice through oral route is > 40 g/kg body weight, which indicates low toxicity.[71]

Drug-Herb Interactions: Simultaneous application of *A. paniculata* and warfarin did not produce significant effects on the pharmacokinetics of warfarin, and practically no effect on its pharmacodynamics.[72]

6. *Ardisia elliptica* Thunb. (Myrsinaceae)

Mata Pelanduk/Ayam, Sea-Shore Ardisia, Shoebutton Ardisia

Ardisia elliptica trees

Flowers of *Ardisia elliptica*

Ardisia elliptica fruits

Description: *Ardisia elliptica* Thunb. is a small shrub that can grow up to 10 m tall. The leaves are obovate, 6–9 cm long with smooth margins. They have an acute apex and a cuneate leaf base. The leaves have a leathery texture. The plant bears whitish pink axillary inflorescences and the drupes are globular, 1–1.2 cm in diameter and grow in clumps, pale red when immature, and turning dark purplish upon maturity.[1–3]

Origin: Native to tropical and temperate Asia.[4]

b716_Ch_01-10.indd 13 1/8/2009 10:09:06 AM

Phytoconstituents: Rapanone,[5] bauerenol, α- and β-amyrin,[6,7] syringic acid, isorhamnetin and quercetin,[8] bergenin,[9] 5-(*Z*-Heptadec-4′-enyl)resorcinol and 5-pentadecylresorcinol.[10]

Traditional Medicinal Uses: The decoction of the leaves is used by the Malays to treat pain in the region of the heart.[10] The *Kadazan Dusun* tribes in Malaysia used paste made from the leaves of *Ardisia elliptica* to treat herpes and measles.[1,11] The fruits are used in Thai Traditional Medicine to cure diarrhoea with fever.[8]

Pharmacological Activities: Antiplatelet[10,11] and Antibacterial.[8]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

7. *Areca catechu* L. (Palmae)

Betel Nut Palm, Areca Nut, Pinang

Fruit of *Areca catechu*

Areca catechu tree

Description: *Areca catechu* L. is a tall, slender palm that can grow up to 10 m. Leaves are dark green, pinnate and up to 1.2–2 m long. Inflorescence is branched and male flowers grow in one row surrounding the female flower at the base of branch. Fruit is a one seeded ovoid berry about 5 cm long.[1-4]

Origin: Originate from the Philippines.[5]

Phytoconstituents: Arecoline, arecaidin, arecaine, catechin, glucides, guvacine, guvacoline, arecolidine, isoguvacine, nicotine and others.[1,6–8]

Traditional Medicinal Uses: In Irian Jaya, parts of this tree are used on wounds, swellings and other skin afflictions. The pericarp is effective in the treatment of flatulence, oedema, dysuria and hyperemesis of pregnancy. On the Finschhafen coast, Papua New Guinea, the inner seed is chopped, heated over fire and pressed on sores caused by sea urchins. Chewing the betel nut with lime and the leaves of catkins or *Piper betel* gives a stimulant effect as well as an attributed sedative effect.[9] This may also be used to soothe a mad person. The red mixture is applied to ulcers in New Britain and to treat sores

b716_Ch_01-10.indd 15 1/8/2009 10:09:10 AM

caused by venereal disease in Northern Province; whereas in New Ireland, the scraped bark is mixed with sea water with a leaf of *Indocarpus fagiferus* and drunk to treat asthma.[9] The Malays use a decoction of the leaves to treat diarrhoea in children.[1] The dried ripe fruits have been used by the Chinese to expel tapeworms and roundworms, treat diarrhoea, indigestion, lumbago, urinary problems and increase menstrual flow.[1,10] The kernel of the fruit is chewed as a narcotic, fresh or cured with slacked lime and betel leaves.[2] The fruits are also used for beriberi, dysentery, dyspepsia, dysuria, oedema and malaria.[6] The fruit can also be applied on venereal sores (ground fresh nut with betel leaf, *Nigella sativa* and roots of *Gymnema hirsutus*, cooked in mustard oil or butter and applied).[11]

Pharmacological Activities: Analgesic,[12] Anthelmintic,[6,13] Antibacterial,[14] Anticancer/Antineoplastic,[15,16] Anticonvulsant,[17] Antidepressant,[18,19] Antihypertensive,[20,21] Antimitotic,[22] Antioxidant,[23] Apoptotic,[24] Hypocholesterolaemic,[25] Immunomodulatory,[26,27] Immunostimulatory,[28–31] Antihyaluronidase,[32] Antivenom,[33] Cell growth inhibitor,[34] Molluscicidal[35] and causes periodontitis.[36]

Dosage: For the treatment of diarrhoea, 30 g areca powder in 200 cm^3 water, simmered for 1 hour is taken before breakfast. If expulsion does not take place within 9 hours, 50 cm^3 of 50% magnesium sulphate solution may be taken.[11] A decoction of the pericarp has been prescribed in a daily dose of 6 to 12 g to treat flatulence, oedema, dysuria and hyperamesis during pregnancy. To treat diarrhoea and dysentery, a daily dose of 0.5 to 4 g of the kernel has been used.[7] For sore throat, the pressed juice is used as a gargle. 2 g of fresh nut can be chewed for 15 min or more before spitting it out. Another reported usage is rolling the leaves and placing them between teeth and gums/lips.[37]

Adverse Reactions: Excessive chewing can cause dizziness, nausea, vomiting, diarrhoea, and seizures.[38]

Toxicity: A dose of 8–10 g is toxic to humans.[6,39] Heavy consumption may cause the development of cancer in the upper and middle third of the oesophagus respectively[40] and chronic kidney disease.[41] Betel quid (a mixture of areca nut and flavouring ingredients with or without processed tobacco leaves) chewing resulted in a statistically significant increase in the risk of total and cerebrovascular deaths in the elderly population.[42] At higher doses, bradycardia, reflex excitability, tremor, spasms and eventual paralysis may occur. Long term effect as stimulant causes malignant tumours of oral cavity

through formation of nitrosamines.[39] Betel quid chewing during pregnancy has a substantial effect on a number of birth outcomes, including sex ratio at birth, lower birth weight and reduced birth length.[43] It is toxic during pregnancy[44,45] as it also possesses cytotoxic and genotoxic activities.[46,47]

Contraindications: Should not be used during pregnancy and lactation. Should not be given to children. Patients with oral or oesophageal cancers, ulcers, oesophagitis, or renal disease should avoid its use.[37]

Drug-Herb Interactions: Decrease action of antiglaucoma agents. Increase action of beta-blockers, calcium channel blockers, cardiac glycosides (digoxin, digitoxin). For neuroleptics, extrapyramidal symptoms can occur.[37] Avoid taking alcohol and atropine.[38]

8. *Asplenium nidus* L. (Aspleniaceae)

Bird's Nest Fern

Asplenium nidus growing on the ground

Asplenium nidus on a tree

Description: *Asplenium nidus* L. is a common epiphytic fern found growing on trees. Fronds are long, simple, green and grow from a central rhizome attached to the tree branch in the shape of a nest. Parallel lines of spores are found on the undersides of the fronds and radiate away from the midrib towards the leaf margin.[1]

Origin: Native to tropical Africa, temperate and tropical Asia and Australasia.[2]

Phytoconstituents: Kaempferol-3-*O*-gentiobiosie-7,4′-bisglucoside, kaempferol-3-*O*-diglucoside, kaempferol-3,7-diglycoside and kaempferol-3-*O*-vicianoside.[3]

Traditional Medicinal Uses: *A. nidus* is regarded as depurative. Infusion of the fronds is used to ease labour pains by Malaysia native tribes. The Malays pound the leaves in water and apply the resulting lotion to feverish head.[4] Two young fronds are eaten when they are still coiled, just after menstruation, in the morning as a contraceptive. Tea made from the fronds is recommended for general weakness.[5,6]

Pharmacological Activities: Oxytocic activity.[6]

Dosage: No information as yet.

b716_Ch_01-10.indd 18 1/8/2009 10:09:14 AM

FA

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

b716_Ch_01-10.indd 19 1/8/2009 10:09:16 AM

9. *Aster tataricus* L.f. (Compositae)

Tatarian Aster, Tatarian Daisy

Aster tataricus flowers

Aster tataricus shrub

Description: *Aster tataricus* L.f. is a small shrub with abundant fibrous roots. Leaf blades are oblanceolate to lanceolate, margins serrate or entire, 4–18 by 1–5 cm and acute. Flower heads are in large bunches with white petals and yellow centre.[1–3]

Origin: Native to Siberia.[3]

Phytoconstituents: Shinone, friedelin, epifriedelinol, shinoside A–C, asterinin A–F, astins A–E, astertarone A&B and others.[4–11]

Traditional Medicinal Uses: The underground rootstock is used as a purgative, treats colds, coughs with excessive sputum or with blood and painful menstruation.[2] It is also used as a bechic-expectorant.[12] Used as an aromatic tonic in chronic gastroenteritis.[13]

Pharmacological Activities: Anticancer[14,15] and Antioxidant.[4]

Dosage: No information as yet.

b716_Ch_01-10.indd 20 1/8/2009 10:09:16 AM

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

10. *Azadirachta indica* A. Juss. (Meliaceae)

Neem

Fruits of *Azadirachta indica*

Azadirachta indica tree

Description: *Azadirachta indica* A. Juss. is a tall evergreen tree, growing up to 30 m in height. Leaves are pinnate with opposite or alternate, lanceolate, serrated and glabrous 8–16 leaflets, 20–32 cm long. Flowers are yellowish white. Fruits are small, ellipsoid, about 5 cm long and green.[1–3]

Origin: A native of India and China, cultivated and naturalised throughout India, Malay Peninsula, Indonesia and Pakistan.[4]

Phytoconstituents: Azadirachtin O–Q, nimbin, deacylnimbin, salanin, nimbidin, nimbinin, nimbidol, azadirone, melianol, meliacinol, nimbothalin, nimonol, azharone and others.[3,5–28]

Traditional Medicinal Uses: It is used for the treatment of a variety of human and veterinary ailments including head lice, mange, fleas, fever, convulsions, leprosy, scrofula, rheumatism, asthma, worm infestations, treat bacterial infection, insecticide, local application for indolent ulcer and consumed as tonic after childbirth.[8] It is used for boils, heart disease, fever,

b716_Ch_01-10.indd 22 1/8/2009 10:09:21 AM

tuberculosis, diarrhoea, jaundice, dysentery, to promote healing, measles, smallpox, sores, inflamed gums, syphilis, leprosy, piles, urinary diseases, to expel worms, purgative, emollient, local stimulant, treat fever (crushed leaves added to lemon), disinfectant (oil from nuts), astringent, contraceptive and tonic.[29,30] It is also used for dermatological problems in Nigeria.[31] Neem has also been used to protect crops, stored grains and library books from insects. Neem leaves buried in grain bins are used to keep the stored crops insect-free. Crushed seeds soaked in water produce a potent pesticide that does not harm mammals, birds, earthworms and bees.[3]

Pharmacological Activities: General review.[32] Antibacterial,[33–37] Anticancer,[38–46] Anticarcinogenic,[47–55] Anticonvulsant,[56] Antifertility,[57,58] Antifungal,[35,59–61] Anti-inflammatory,[62] Antimalarial,[63–68] Antimicrobial,[69] Antioxidant,[70–74] Antiproliferative,[75,76] Antipyretic,[77] Antiviral,[78,79] Gastroprotective,[80–82] Hepatoprotective,[83–85] Hypoglycaemic,[86,87] Hypotensive,[88–90] Immunostimulatory,[38,91–93] Neuroprotective,[94] Anthelmintic,[95] Antihaemorrhagic,[37] Antileishmanial,[96] Antimutagenic,[97–99] Molluscicidal,[100] Insecticidal and Insect repellent.[101–110]

Dosage: Approximately 100 g of bark is soaked in 1 L of water and approximately 3 ml of this infusion is consumed daily for one month as a male contraceptive.[30]

Adverse Reactions: Nausea, vomiting, anorexia, hypersensitivity, Reye's syndrome (infants).[111]

Toxicity: Mice injected with the tetranortriterpenoid fractions (>86 mg/feeding) into the tail vein died within 24 hours.[8] Various acute, subacute and chronic toxicity tests of extracts of neem have been reported, especially on human and animal fertility.[112] Non-aqueous extracts appeared to be the most toxic neem-based pesticide products.[112]

Contraindications: Should not be used during pregnancy and lactation, and in children. Should not be used in persons with hypersensitivity to neem.[111]

Drug-Herb Interactions: No information as yet.

[**Authors' Note:** An extensive review[112] on the safety of neem derived pesticides concluded that the use of neem derived pesticides as an insecticide should not be discouraged.]

11. *Barringtonia asiatica* L. (Lecythidaceae)

Beach Barringtonia, Fish-killer tree, Putat Laut

Barringtonia asiatica fruit

Developing fruits of *B. asiatica*

Barringtonia asiatica tree

Description: *Barringtonia asiatica* L. is a large tree bearing large simple leaves, 20–30 cm long which taper to the leaf base. Flowers are large and white with several white stamens. The flowers are actinomorphic and have four petals. The fruit is oblong, green, large, 8–10 cm across and contains one seed.[1–3]

Origin: Native to Africa, temperate and tropical Asia and Australasia.[4]

Phytoconstituents: A_1-barrinin, ranuncoside VIII, A_1-barrigenin and others.[4,5–7]

Traditional Medicinal Uses: The plant is used to treat fungal infections,[8] burns and wounds.[9] The leaves are heated and used to treat stomachache and

b716_Ch_11-20.indd 24 1/8/2009 10:15:22 AM

rheumatism in the Philippines. Its fruits are used as a fish poison and the fruit juice for controlling scabies while the seeds are used for the expulsion of intestinal worms and also as a fish poison.[10] They are also used to treat sores, cough, influenza, sore throat, diarrhoea, swollen spleen after malaria.[11] In other provinces of Vietnam, the fresh nut is scraped and applied to sores; dried nut is ground into a powder, mixed with water and drunk to cure coughs, influenza, sore throat, bronchitis, diarrhoea and swollen spleen.[12] The bark is used in the treatment of tuberculosis.[11] In Yambio (Sudan), the inner bark is crushed and mixed with water and drunk to ease the aching associated with malaria. It is also used in combination with other plants as a medicine to treat tuberculosis in New Ireland and the Solomon Islands.[12]

Pharmacological Activities: Insect repellent,[5] Antibacterial and Antifungal.[13]

Dosage: To relieve the aching associated with malaria, inner bark is crushed and mixed with water and drunk, 2 cups per day for 2 days.[12]

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

12. *Barringtonia racemosa* (L.) K. Spreng
(Lecythidaceae)

Putat Kampong, Samundrapandu

Barringtonia racemosa flowers

Barringtonia racemosa fruits

Barringtonia racemosa tree

Description: *Barringtonia racemosa* (L.) K. Spreng is a small tree with large, simple leaves about 20 cm long. Flowers are large with numerous pinkish stamens. The fruit is ellipsoid, green, 8 cm long, and turns red upon maturity.[1,2]

Origin: Native to Africa, temperate and tropical Asia, Australasia and the Pacific.[3]

Phytoconstituents: Nasimalun A and B, barringtonin, R_1-barrigenol, R_2-barrigenol, barringtogenol, barringtogenic acid and others.[4–9]

b716_Ch_11-20.indd 26 1/8/2009 10:15:29 AM

Traditional Medicinal Uses: The plant is widely used in the form of a decoction in Sri Lankan traditional medicine.[10] The leaves and bark are used for rat and snake bites, rat poisoning and on boils.[10,11] The fruits are used for cough, asthma and diarrhoea.[11,12] Kernels of the drupe are mixed with milk to treat bilious diseases and jaundice whereas the seed has been used as an insecticide and tonic.[11] The seeds along with other ingredients are employed in preparations for the treatment of itch, piles and typhoid fever, while the bark is also used for gastric ulcers.[10] Its roots act as a coolant and deobstruent.[11]

Pharmacological Activities: Antinociceptive,[10] Antibacterial,[12] Glucosidase and Amylase Inhibition,[13] Anticancer [14] and Cytotoxic.[15]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: Aqueous extracts (500, 750, 1000 or 1500 mg/kg) of *B. racemosa* bark in male rats did not produce any unwanted side effects or toxicity or alter fertility, gestational length, peri- and neonatal development and appeared to be non-teratogenic.[10] *B. racemosa* seed extract was found to be devoid of acute and short-term toxicity to mice, when administered daily, intraperitoneally for a fortnight up to a dose of 12 mg/kg. The treated mice showed conspicuous toxic symptoms only at 24 mg/kg. The LD_{50} of male mice for a single i.p. dose is 36 mg/kg.[14]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

13. *Bauhinia purpurea* L. (Leguminosae)

Butterfly Tree

Leaves of *Bauhinia purpurea*

Flower of *Bauhinia purpurea*

Bauhinia purpurea tree

Description: *Bauhinia purpurea* L. is a deciduous tree. Leaves are simple and stipulate. Blade is butterfly shaped, coriaceous with 9–11 pairs of secondary veins. Flowers are showy, pink and arranged in axillary. The fruits are darkish pods, 1.5 by 15 cm and woody.[1,2]

Origin: Native to India and grown in the Asia-Pacific as ornamental plants.[2]

Phytoconstituents: Bauhiniastatins 1–4, bauhinoxepin C–J, bauhibenzofurin A, bauhispirorin A, bauhinol E and others.[3–9]

Traditional Medicinal Uses: The root is grated with water and is drunk to treat common fever. The flowers are used as laxative and leaves applied to sores and boils, and for cough treatment.[10]

b716_Ch_11-20.indd 28 1/8/2009 10:15:30 AM

Pharmacological Activities: Antibacterial,[9,11] Antifungal, Antimalarial, Cytotoxic,[9] Anticancer,[8] Anti-inflammatory[9,12] Antinociceptive, Antipyretic[12] and Thyroid hormone regulating.[13]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

14. *Bixa orellana* L. (Bixaceae)

Annatto, Lipstick Tree

Fruits of *Bixa orellana*

Bixa orellana flower

Bixa orellana tree

Description: *Bixa orellana* L. is a small tree with simple and spiral leaves, 10–20 cm by 6.3–12.5 cm, dark green, ovate, acuminate, truncate at the base and glabrous. Flowers are large, 5 cm in diameter, pinkish or white, arranged in terminal panicles. Fruits are dehiscent, ovoid capsules containing 15–20 trigonous seeds in bright red pulp.[1–3]

Origin: Originate from tropical America. Cultivated Pantropically.[4]

Phytoconstituents: Bixin, valencene, β-elemene, β-selinene, copaene, δ-cadinene, spathulenol, γ-cadinene, δ-elemene, ledol, α-muurolene, α-cadinol and others.[5–10]

b716_Ch_11-20.indd 30 1/8/2009 10:15:35 AM

Traditional Medicinal Uses: The leaves have been used to treat snakebites and jaundice and the seed is considered a good cure for gonorrhoea.[11] The bark of the root is used to treat fever and as an aperient.[11] In Cambodia, the leaves are a popular febrifuge while in Indonesia, water in which the leaves are rubbed is poured over the head of children with fever. In Malaysia, the leaves are used in a postpartum medicine and in the Philippines the leaves are pounded in coconut oil and heated, then applied to the abdomen to relieve tympanites. Pastes of the fresh leaves are rubefacient and used in dysentery. In Vietnam, lotions or baths of leaves are used during fever. Its unripe fruits are emollient in leprosy.[3] Alcoholic extracts of seed coat are taenifuge and laxative.[3] Decoctions of barks are used for catarrh. Infusions of seeds are used to treat asthma and excessive nasopharynx mucus production.[3] Traditionally, it is also used as a gargle for sore throats and oral hygiene.[12] In Trinidad and Tobago, the leaves and roots are used for hypertension, diabetes and jaundice.[13] Leaves and seed pods are used as a female aphrodisiac.[14]

Pharmacological Activities: Antibacterial,[12,15–18] Anticancer,[19] Anticonvulsant,[16] Antidiarrhoeal,[16] Anti-inflammatory,[19,20] Antioxidant,[3,16] Antiplatelet,[21] Hypoglycaemic,[3,22] Immunostimulatory,[3] Sedative,[16] Antigenotoxic and Antimutagenic,[23] Antifungal, Antileishmanial[24] and Radioprotective.[3]

Dosage: Approximately 9 seed pods are boiled in 3 cups of water for 10 mins and drunk as a diuretic before each meal. 10 g powdered seed/40 ml oil for topical pastes.[25] As a female aphrodisiac, 3 leaves in 0.5 L of water and red paste of seed pods.[14]

Adverse Reactions: Urticaria and angiooedema are possible adverse reactions with annatto dye. A patient developed these symptoms and hypotension within 20 min of ingestion of annatto containing fibres.[3]

Toxicity: Toxic to dogs dosed with 60 mg/kg trans-bixin.[26] However, annatto containing 5% bixin was non-genotoxic and non-carcinogenic to rat livers even at the highest concentration tested at 1000 ppm (4.23 bixin/kg body weight/day).[27] Annatto given through gavage to Wistar rats on days 6–15 of pregnancy showed no adverse effect on the mothers and foetus.[28]

Contraindications: Trans-bixin is hyperglycaemic and should not be ingested by patients with diabetes mellitus.[25]

Drug-Herb Interactions: No information as yet.

15. *Calophyllum inophyllum* L. (Guttiferae)

Indian Laurel, Penaga Laut, Borneo Mahogany

Flowers of *Calophyllum inophyllum*

Calophyllum inophyllum tree

Description: *Calophyllum inophyllum* L. is a large tree with broad, glossy, leathery, elliptic-oblong leaf blades, 8–16 cm by 4–8 cm, and with numerous parallel side veins. The tree bears sweetly scented white flowers in erect racemes. Fruits are globose, 2 cm across and are green in colour.[1,2]

Origin: It is found in Africa, tropical and temperate Asia, Australasia and the Pacific.[3]

Phytoconstituents: Inophynone, canophyllol, canophyllic acid, calophyllolide, inophyllolide, inophyllum B, C, P, and E, jacareubin, (+)-calanolide A, inocalophyllins A and B, calophinone, calophyllumin C, inophyllin A and others.[4–28]

Traditional Medicinal Uses: The whole plant is used as a crude drug for curing rheumatism and skin affections in South India.[14] Its juice is a purgative and the seed oil is specific for rheumatism and various skin diseases (i.e., scabies, ringworm and dermatosis). The bark is used for internal haemorrhage and as an astringent.[29] In Buso, Papua New Guinea, the milky latex from the leaves is diluted with water and the solution is applied to irritated eyes.[30] The gum is emetic and purgative.[29]

b716_Ch_11-20.indd 32 1/8/2009 10:15:37 AM

Pharmacological Activities: Antibacterial,[31,32] Anticancer/Antineoplastic,[19] Anti-inflammatory,[33] Antiplatelet,[34,35] Antipsychotic,[33] Antiviral,[15,18,36–38] Photoprotective,[39] Molluscicidal[40] and Piscicidal.[5]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: The unrefined oil is toxic.[41] There was a significant difference in the plasma cholesterol levels of the rats fed with *C. inophyllum* oil when compared with the control. Mild, focal to severe and widespread lesions were found in the kidneys, hearts and livers of rats fed with *C. inophyllum* seed oil.[42]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

16. *Cananga odorata* (Lam.) Hook. f. & Th. (Annonaceae)

Kenanga, Ylang-Ylang

Flowers of *Cananga odorata*

Cananga odorata shrub

Description: *Cananga odorata* (Lam.) Hook. f. & Th. is a shrub which grows to a height of 4 m. Its bark is grey and smooth. Leaves are simple, alternate, exstipulate, oblong to broadly elliptic, large, 3–6.5 cm across, 8.5–29 cm long, with distinct venation pattern. Flowers are fragrant, in clusters on older branches. Fruit turns black on maturity and has many seeds in two rows.[1–3]

Origin: Originates from Indochina, Malesia and tropical Australia.[4]

Phytoconstituents: Acetogenin, aporphine, liriodenine, canangone, α-humulene, β-cubebene, germacrene D, cananodine, γ-eudesmol and others.[3,5–7]

Traditional Medicinal Uses: It is used for asthma, malaria, fever, cholera, typhoid, scabies, dermatitis, ulcer and wounds.[8] The seeds are used for stomach complaints with fever and in Indonesia, the bark is used for scabies.[3] In Malaysia, a paste of fresh flowers is applied to the chest for asthma and to treat malaria. In Solomon islands, a paste of fresh flowers is applied to boils while in India, the essential oil from the flowers makes an external remedy for cephalgia, ophthalmia and gout.[3,9]

b716_Ch_11-20.indd 34 1/8/2009 10:15:43 AM

Pharmacological Activities: Antibacterial,[10,11] Antifungal,[11,12] Antihypertensive,[3,13] Antioxidant,[10] Antineoplastic[7,12,14,15] and Antiprotozoal.[12,16]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: A 50% ethanolic root bark extract administered orally to male albino rats at the dose of 1 g/kg body weight/day for 60 days resulted in decreased epididymal sperm motility and sperm count, and morphological abnormalities in the sperms.[17] However, it is non-toxic at the current level of intake as a food ingredient (0.0001 mg/kg/day). Although sometimes Ylang-Ylang oil has been reported to cause dermal sensitisation reactions in animals and humans, it is unclear what constituent(s) within the essential oil comprise the offending agent.[18]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

17. *Capsicum annuum* L. (Solanaceae)

Chilli, Red Pepper

Capsicum annuum fruits

Capsicum annuum plant

Description: *Capsicum annuum* L. is a small herb that can grow up to 1 m tall. Leaves are oblong-ovate, ovate, or ovate-lanceolate, 4–13 cm by 1.5–4 cm with entire margin. Flowers are small, white or tinged purple. Fruits are mostly red, but can be green, orange, yellow and can grow up to 15 cm. Seeds are pale yellow, discoid or reniform and 3–5 mm.[1–5]

Origin: Native to Mexico and South America, widely cultivated throughout the world.[5]

Phytoconstituents: Capsaicin, capsicosides E–G, capsianosides 1–4, capsianosides VIII, IX, X, XIII, XV and XVI, solanidine, solanine, solasdine, scopoletin and others.[6–15]

Traditional Medicinal Uses: The leaves are used to treat toothache. The fruits are used to stimulate gastric activities and increase blood circulation.[4] It is also a stimulant, carminative, and used locally for neuralgia and for rheumatism.[15] Uterine pain associated with childbirth is treated with soup

b716_Ch_11-20.indd 36 1/8/2009 10:15:49 AM

containing the fruit.[16] The Commision E approved *Capsicum annuum* for painful muscle spasms in areas of shoulder, arm and spines. Preparations are used to treat arthritis, neuralgia, lumbago and chilblains.[17]

Pharmacological Activities: Antibacterial,[18,19] Antifungal,[20,21] Anticancer,[22,23] Antioxidant,[10,24–26] Antiprotozoal,[27] Hypocholesterolaemic/ Hypolipidemic,[19,28] Immunomodulatory,[29] Antimutagenic[30,31] and Pesticidal.[32]

Dosage: Liquid extract is prepared by percolating 100 gm of the plant extract with 60 mg of ethanol, to be used as an antirheumatic. External daily dose of semi solid preparations containing maximum of 50 mg of capsaicin in 100 gm neutral base is also used as an antirheumatic and applied to the affected area not more than 3 or 4 times daily.[33]

Adverse Reactions: Internally, it may cause gastrointestinal cramping, pain, and diarrhoea. Topically, it may cause painful irritation of mucous membrane.[34]

Toxicity: Oral LD_{50} values for capsaicin are 161.2 mg/kg (rats) and 118.8 mg/kg (mice), with haemorrhage of the gastric fundus observed in some of the animals that died. However, capsaicin is considered to be safe and effective as an external analgesic counterirritant.[35] Rabbits fed with *C. annuum* powder at 5 mg/kg per day in the diet daily for 12 months showed damaged liver and spleen. A rabbit skin irritation test of *C. annuum* fruit extract at 0.1% to 1.0% produced no irritation but caused neoplastic changes in the liver and intestinal tumours were observed in rats fed red chili powder at 80 mg/kg per day for 30 days.[35] High doses administered over extended period of time can cause chronic gastritis, kidney damage, liver damage and neurotoxic effects.[33]

Contraindications: Should not be used during pregnancy and lactation, in people with hypersensitivity and in children. Should not be used on open wounds or abrasions, or near the eyes.[34]

Drug-Herb Interactions: Reported with concomitant administration with aspirin and salicylic compounds.[33,36] Decrease the actions of α-adrenergic blockers, clonidine (anti-hypertensive) and methyldopa (antihypertensive). Hypertensive crisis with monoamine oxidase inhibitors.[34]

18. *Cassia fistula* L. (Leguminosae)

Golden Shower Tree, Indian Laburnum, Purging Cassia

Cassia fistula fruits and flowers

Cassia fistula tree

Description: *Cassia fistula* L. is a large tree, which grows to 10 m tall. Leaves are alternate, pinnate, 3–8 pairs of leaflets, broadly ovate and pointed. Flowers are dense, bright yellow and about 4–5 cm across. Fruit pods are long, 30–60 cm, cylindrical, brown in colour and contains many seed.[1–3]

Origin: Native to India and Sri Lanka.[2,4]

Phytoconstituents: Fistucacidin, chrysophanic acid, chrysophanol, clitorin, sennosides A and B, chrysophanein and others.[2,4–9]

Traditional Medicinal Uses: The whole plant is used for anthrax, burns, cancer, constipation, convulsions, delirium, diarrhoea, dysentery, dysuria, epilepsy, fever, influenza, gravel, haematuria, pimples, syphilis, tumours and worms.[4,10] The leaf is used for skin diseases (juice), healing ulcers, for ring-worm and irritation of skin (juice of young leaves), facial paralysis and rheumatism (paste).[10] The raw black pulp found between the seeds is a

b716_Ch_11-20.indd 38 1/8/2009 10:15:51 AM

popular remedy for constipation.[3,4] It is also used as a cathartic, for rheumatism and snakebite (pulp), treats bacterial infections (pulp mixed with leaves of *Cassia angustifolia*), liver complaints, heart disease, reduce fever, as abortifacient, demulcent and is useful in liver, throat, eye diseases, convulsions and sores. The seed is an emetic, carminative, appetiser, and is used for constipation, jaundice, cancer on face and syphilis.[10] The roots act as a purgative while the rootbark is used for cleansing wounds.[4] The root is also used as an astringent, tonic, febrifuge, for skin diseases, leprosy, tuberculous glands, syphilis and epilepsy.[10]

Pharmacological Activities: Antimicrobial,[11–14] Anticholinergic,[15] Antifertility,[16] Anti-inflammatory,[17] Antineoplastic,[18] Antioxidant,[19–22] Depressant,[23] Hepatoprotective,[24–26] Hypocholesterolaemic,[27] Anti-leishmanial,[28] Larvicidal[29] and Wound healing.[30,31]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: In cases of overdose or prolonged administration, loss of electrolytes, especially potassium ions, aldosteronism, albuminuria, haematuria, inhibition of intestinal motility and muscle weakness may occur. Rarely, cardiac arrhythmia, nephropathy, oedema, and accelerated osteoclasis may arise.[32]

Contraindications: Contraindicated with acute inflammatory diseases of intestine and appendicitis. Should not be used during pregnancy and while nursing. Should not be used in children under 12 years of age.[32]

Drug-Herb Interactions: Interaction with anthranoid laxatives.[33] Enhancement of effects of cardioactive steroids may occur. Effects of antiarrhythmics may also be affected.[32]

19. *Catharanthus roseus* (L.) G. Don
(Apocynaceae)

Madagascar Periwinkle, Rose Periwinkle

Catharanthus roseus flower

Catharanthus roseus shrub

Description: *Catharanthus roseus* (L.) G. Don is a herb, up to 80 cm tall. Stem is woody, slightly branched and all parts contain white milky latex. Leaves are simple, dark green, glossy, obovate-elliptic, 4–5 cm by 2–3 cm with prominent lateral veins on the abaxial surface. Flowers are bisexual, white, purple, pink, red or white with a red or pink centre. Fruits consist of pairs of greenish succulent follicles, 2–3 cm long and contain small oblong seeds.[1,2]

Origin: Native to Madagascar. Cultivated or naturalised in all tropical countries.[2,3]

Phytoconstituents: Vinblastine, vincristine, leurosine, akuammicine, carosine, catharanthine, catharicine, catharine, catharosine, cathovaline, catharanthiole, vindoline, vindolinine, vincaleucoblastine, secologanin, mauritianin, rosicine and others.[4–11]

Traditional Medicinal Uses: The plant is used as a remedy for diarrhoea, malaria, diabetes, astringent, diaphoretic, bechic, emmenagogue, menstrual pain, hypertension, insomnia and depurative after parturition in Indochina,

b716_Ch_11-20.indd 40 1/8/2009 10:15:55 AM

the Philippines, Jamaica, West Indies, South Africa, Southeast Asia, India and Queensland.[2,5–7,12,13] The plant is also used for cold, cough, fever and bronchitis.[14] In Malaysia, the crushed leaves are applied to scalds, burns, sores, mumps, swollen neck, tonsillitis and insect bites.[13] In Puerto Rico and Cuba, the flowers are decocted and used as an eyewash.[12]

Pharmacological Activities: Anticancer/Antineoplastic,[5,12,15–19] Antioxidant,[20] Antiangiogenic,[21] Chemopreventive,[22] Hypoglycaemic[23–28] and Wound healing.[29,30]

Dosage: Leaves are useful in treating oliguria, haematuria, diabetes mellitus, and menstrual disorders in a daily dose of 4 to 8 g as a decoction or liquid extract.[7] A decoction of 30 g of the plant is taken for diabetes, dysentery, enteritis, menstrual pains, hypertension, insomnia and cancer in Malaysia.[13] For the treatment of cold and sore throat, tea is made from 9 pink flowers in 1 pint of water and sipped throughout the day.[31] Dilute infusions of roots are used to treat diabetes.[32]

Adverse Reactions: *Catharanthus roseus* pollen can trigger IgE-mediated respiratory allergy in the people living in close proximity.[33]

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

[**Authors' Note:** Clinically, vinblastine and vincristine are administered by intravenous injection as solutions of their sulphate salts.[34]]

20. *Celosia argentea* L. (Amaranthaceae)

Feather Cockscomb, Red Spinach

Flowers of *Celosia argentea*

Celosia argentea plant

Description: *Celosia argentea* L. is an annual tropical herb, up to 1 m tall. Stems are cylindrical and the aerial part is branched. Leaves are simple, small, spirally arranged, about 5–8 cm by 1–3 cm, alternate and exstipulate. The blade is lanceolate and ovate. The apex is acuminate. It bears several pinkish or white flowers which are minute. The fruits are globose and seeds are black.[1–5]

Origin: Native to India.[6]

Phytoconstituents: Celosian, nicotinic acid, celogenamide A, celogentin A–D, H, J and K, moroidin and others.[3,5,7–11]

Traditional Medicinal Uses: *C. argentea* is used internally for haematological and gynaecologic disorders and externally to treat inflammation and as a disinfectant. The whole plant is used to treat dysentery and dysuria, and used externally as poultices for broken bones.[5] The plant is used for eye

b716_Ch_11-20.indd 42 1/8/2009 10:15:56 AM

and liver ailments in Yunnan, China[12] and also for the treatment of mouth sores and blood diseases and used as an aphrodisiac.[13] The petioles are used to treat sores, wounds, boils and swellings.[4,12] The seeds are used for the treatment of conjunctivitis and hypertension. In China, the seeds are used for haemorrhage, menorrhagia and opthalmia.[5,12,14] In Indonesia, the flowering tops are used for bleeding lungs whereas in Malaysia, the red flowering tops are prepared as decoctions which are given in cases of white discharges, excessive menstruations, haematuria, dysentery, proteinuria, bleeding piles and bleeding nose.[5]

Pharmacological Activities: Antibacterial,[15] Antimitotic,[16] Antineoplastic,[17] Diuretic,[5] Hypoglycaemic,[18] Hepatoprotective,[11,19] Immunomodulatory,[5,17,20] Cytoprotective[21] and Wound healing.[22]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: Leaves should not be eaten by menstruating women.[23]

Drug-Herb Interactions: No information as yet.

21. *Centella asiatica* (L.) Urban (Umbelliferae)

Indian Pennywort, Asiatic Pennywort

Top view of *Centella asiatica* herb

Description: *Centella asiatica* (L.) Urban is a creeping, perennial herb with long slender horizontal stolons, characterised by long internodes. Leaves are green, fan-shaped or round-reniform, 1–4 cm by 1–7 cm with a crenate or dentate margin. Flowers are umbels with white or light purple-to-pink petals and bear small oval fruit.[1,2]

Origin: Native to India, found in tropical America, Africa, West Pakistan, China, Japan and the Pacific.[3]

Phytoconstituents: Asiaticoside, asiatic acid, brahmic acid, brahmoside, centellic acid, centellose, indocentelloside, madecassic acid, madecassoside, thankuniside, vellarin, bayogenin, centellin, asiaticin, and centellicin.[4–15]

b716_Ch_21-30.indd 44 1/8/2009 10:17:28 AM

Traditional Medicinal Uses: The plant is used in cooling drinks when boiled, for diarrhoea, diuretic, gravel, leprosy treatment, stones, wound healing and as a tonic.[2,4] It is also part of a mixture to treat colic in Indonesia.[16] The Chinese uses it to improve appetite, aid digestion, treat sores and ulcers. In India, it is used to treat skin, nervous system and blood diseases.[2,7,17] The plant is used to treat fever and rheumatism, as a detoxicant, sedative and peripheral vasodilator.[7,18] The plant is reported as treatment for cancer, circulatory stimulant, hypotensive, stimulant, tonic, cicatrizant (leaf), used for the treatment of haematuria, gonorrhoea, peptic ulcer and sore throat.[5] It is also used for tuberculosis, headache (decoction), dysentery (decoction with other ingredients), boils and tumours (paste applied), leucoderma, anaemia, urinary discharges, bronchitis, insanity, leprosy (decoction), mental deficiency, dysentery (juice or paste on empty stomach for 2–3 days), cough (decoction with ginger and black pepper), cooling (paste with pepper and salt), tonic (juice with palm jaggery given to women after childbirth), elephantiasis, enlarged spleen (ointment from leaves), cures stuttering or stammering, diuretic, small pox and as a local stimulant.[19] External application in the form of poultices is prescribed for contusions, closed fractures, sprain and furunculosis.[18]

Pharmacological Activities: Antibacterial,[20] Antidepressant,[21] Antiemetic,[22] Antineoplastic,[8,23–26] Antioxidant,[27–32] Antithrombotic,[33] Anxiolytic,[34,35] Gastroprotective,[9,36] Immunomodulatory,[37,38] Antigenotoxic,[39] Nerve-regenerative,[40] Radioprotective[41,42] and Wound healing.[43,44]

Dosage: Approximately 0.6 g dry weight of whole plant taken three times a day (condition not indicated).[45] Larger amounts of fresh leaves are sometimes eaten as vegetable and dietary supplement.[45] 60 mg of *C. asiatica* extract given daily for varicose veins.[46] 30 or 60 mg of Total Triterpenic fraction of *Centella asiatica* (TTFCA) three times a day for improving venous hypertension.[47] Titrated extract of *C. asiatica* given 60 or 120 mg daily for chronic venous insufficiency.[48] For the treatment of fever, measles, haematemesis, epistaxis, diarrhoea, dysentery, constipation, leucorrhoea, jaundice, dysuria, furunculosis, dysmenorrhoea, varices, daily dose of 30 to 40 g of fresh plant in the form of extracted juice or decoction is taken.[18]

Adverse Reactions: Allergic contact dermatitis.[49] Sedation, increased blood glucose and cholesterol levels.[50]

Toxicity: Hepatotoxic.[51]

Contraindications: Should not be used during pregnancy and lactation and not to be given to children.[50]

Drug-Herb Interactions: Reported with ephedrine, theophylline, atropine and codeine.[52] Decrease effectiveness of antidiabetic and antilipidemic drugs.[50]

[**Authors' Note:** Topical creams containing the active component asiaticoside is available commercially for wound healing.]

22. *Cerbera odollam* Gaertn. (Apocynaceae)

Pong Pong Tree, Indian Suicide Tree, Sea Mango

Fruit of *Cerbera odollam*

Cerbera odollam leaves

Cerbera odollam tree

Description: *Cerbera odollam* Gaertn. is a medium-sized tree with smooth and grey bark. Leaves are simple, few, without stipules, 12–16 cm by 3–5 cm, arranged in a spiral. The blade is succulent, dark green, glossy and lanceolate. Flowers are white, large and bisexual with a yellow eye in the throat of the corolla tube and arranged in terminal. Fruits are round, waxy surfaced, large, 5–10 cm across, with fibrous husk covering the single seed. The fruit turns from green to reddish brown upon maturity.[1–4]

Origin: Native to Indian subcontinent, Indochina, Malesia and the Pacific.[5]

b716_Ch_21-30.indd 47 1/8/2009 10:17:40 AM

Phytoconstituents: Cerberin, cerleaside A, 17α-neriifolin, 17β-neriifolin, thevetin B, acetyl-thevetin B, diacetylneriifolin, cerberoside, odollin and others.[3,6–12]

Traditional Medicinal Uses: In Malaysia, rheumatism is treated with embrocations of the fruits.[11] The seeds are poisonous and have been used to poison rats and dogs.[3,11] The seeds are also narcotic. In Indonesia, oil obtained from the seeds is rubbed on the body as remedy for colds, scabies, and rheumatism. In the Philippines, oil of the seeds is also used to treat rheumatism.[11] It is used in Burma as an insecticide or insect repellent when mixed with other oils.[13] The bark, latex and roots are used as purgatives and emetics in India.[3,11,13]

Pharmacological Activities: Antineoplastic[8,10] and Antipsychotic.[14]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: Seeds are poisonous and have been used to poison rats and dogs.[3,11] The seeds have a long history as an ordeal poison in Madagascar, due to the highly toxic cardiac glycosides they contain. The kernel contains cerberin, cerberoside and odollin which are toxins.[15] Humans poisoned by oral consumption of half to one seed kernel would result in sinus bradycardia, wandering pacemaker and second-degree sino-atrial block and nodal rhythm.[15,16] Other symptoms include nausea, retching and vomiting.[16] The poisoning may result in hyperkalemia, which can cause death.[17]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

23. *Cissus quadrangularis* L. (Vitaceae)

Grape Leaf, Veld Grape

Flowers of *Cissus quadrangularis*

Cissus quadrangularis plant

Description: *Cissus quadrangularis* L. is a herbaceous plant with a thick, quadrangular, succulent stem and is constricted at regular intervals. It has long, slender and simple tendrils. Leaves are ovate, entire, crenate-serrate, 3–5 cm by 5–13 cm and glabrous on both sides. Flowers are pink and white, 2 mm long, while the berries contain one or two seeds.[1,2]

Origin: India, Pakistan, Indonesia, Malaysia, East Africa, Sri Lanka and Arabia.[2]

Phytoconstituents: Quadrangularins A–C, δ-amyrin, δ-amyrone, resveratrol, piceatannol, pallidol, parthenocissine A and others.[3–7]

Traditional Medicinal Uses: Whole plant is used for urinary *schistosomiasis* in Mali.[8]

b716_Ch_21-30.indd 49 1/8/2009 10:17:46 AM

Pharmacological Activities: Antibacterial and Antioxidant,[9] Analgesic and Anti-inflammatory,[10] Antimalarial,[11] Gastroprotective[12–16] and Antiosteoporotic.[17]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

24. *Cocos nucifera* L. (Palmae)

Coconut Palm, Kelapa

Cocos nucifera leaves

Fruits of *Cocos nucifera*

Cocos nucifera trees

Description: *Cocos nucifera* L. is a tall palm with a ringed stem that can grow up to 30 m tall. Leaves are pinnate, 2–6 m long with numerous pairs of narrow leaflets. Flowers are arranged in large panicles among the leaves with female flowers near the base of the inflorescence. The fruit is symmetrical, ovoid and about 20–30 cm across. Thick fibrous husk encloses the hard shell (endocarp) in which the fleshy pericarp adheres.[1,2]

Origin: Native to the Pacific.[3]

b716_Ch_21-30.indd 51 1/8/2009 10:17:49 AM

Phytoconstituents: Trans-zeatin, dihydrozeatin, dihydrozeatin-*O*-glucoside, *meta*-topolin riboside, N^6-isopentenyladenine, N^6-benzylaminopurine and others.[4,5]

Traditional Medicinal Uses: Its fruit juice is used to treat poisoning, cholera and is a diuretic. Fresh coconut juice with rice flour is poulticed onto carbuncles, gangrenous sores and indolent ulcers.[6] Coconut water is also used for fever, urinary complaints and to stop vomiting.[7] Juice of the green, unripe coconut is boiled and drunk to relieve diarrhoea. In Somoa, coconut oil is used as a laxative and to relieve stomach ailments.[8] The decoction of husk fibre has been used in northeastern Brazil traditional medicine for the treatment of diarrhoea and arthritis.[9] Coconut water has also been used as short-term intravenous hydration and resuscitation fluid in emergencies.[10] The rootbark is used as an astringent, styptic and in haemorrhages.[6] It is also used for uterine diseases, bronchitis, liver complaints and dysentery.[7]

Pharmacological Activities: Analgesic,[11] Antibacterial,[9,12] Antifungal,[13] Antineoplastic,[14,15] Antioxidant,[11,16] Antiprotozoal,[17] Antiviral,[9] Hypoglycaemic,[18] Hypolipidaemic,[19,20] Hypotensive,[21] Immunomodulatory[22] and Antitrichomonal.[23]

Dosage: No information as yet.

Adverse Reactions: A case report of occupational allergic conjunctivitis due to coconut fibre dust has been reported.[24]

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

25. *Coix lacryma-jobi* L. (Gramineae)

Job's Tears, Adlay

Fruits of *Coix lacryma-jobi*

Coix lacryma-jobi plants

Description: *Coix lacryma-jobi* L. is a small herbaceous plant which grows up to 2.5 m. Leaf sheaths are glabrous and the leaf blades are narrowly lanceolate, 20–50 cm by 1.5–4 cm. The midrib is prominent. Fruit is tear-shaped, 8 mm by 1.1 cm, with glossy berries which turn black upon maturity.[1–3]

Origin: Native to tropical and temperate Asia.[4]

Phytoconstituents: Coixol, coixenolide, α-coixins, γ-coixins, syringaresinol, mayuenolide, coixan A–C, coixic acid and others.[5–9]

Traditional Medicinal Uses: A decoction is believed to benefit the blood and breath and used to wash newborns to prevent diseases.[10] The kernels are used to treat lung and chest complaints, rheumatism, dropsy and gonorrhoea.[3] Fruit is used for intestinal or lung cancers and warts. The fruit is also used as a vermifuge and for hypertension. As the seed is diuretic and refrigerant, a decoction is used for appendicitis, arthritis, beriberi, bronchitis, cancer, diarrhoea, dryskin, dysuria, oedema, hydrothorax, inflammation, pleurisy, pneumonia, pulmonary abscesses, rheumatism and tuberculosis.[10,11] Seeds are also used for the treatment of enteritis, persistent diarrhoea in children, urinary lithiasis, rheumatism and acrodynia.[12] An infusion of the seeds is prescribed for bronchitis, pulmonary abscess, pleurisy and hydrothorax.[11] The root is given along with roots of long pepper and other herbs for fever

b716_Ch_21-30.indd 53 1/8/2009 10:17:57 AM

with drying of saliva and intense thirst, for dysentery, diarrhoea and puerperal fever.[13]

Pharmacological Activities: Anti-inflammatory,[14,15] Anticancer/Antineoplastic,[16–21] Antioxidant,[8,18] Hypoglycaemic,[9,22] Hypolipidaemic[22–24] and Hypotensive.[25]

Dosage: For the treatment of enteritis, persistent diarrhoea in children, oedema, urinary lithiasis, rheumatism and acrodynia, doses of 10 to 30 g daily in the form of powder or decoction are given.[12]

Adverse Reactions: No information as yet.

Toxicity: Embryotoxicity in pregnant rats was observed. Oral administration of 1 g/kg body weight of water extracts caused an increase in foetal resorptions and postimplantation mortality.[26]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

26. *Crinum asiaticum* L. (Amaryllidaceae)

Crinum Lily, Spider Lily, Bawang Tanah

Crinum asiaticum flowers

Leaf blade of *Crinum asiaticum*

Crinum asiaticum plant

Description: *Crinum asiaticum* L. is a bulbous herb with a flowering stalk in the centre of the plant. The leaves are narrowly lanceolate, acuminate, 0.5–1.5 m long, greenish and have a hairy texture. The flowering stem is about 1–1.2 m. Flowers are white, 6–12 in an umbel. Filaments are reddish. The fruit is irregularly globose, 4–5 cm across, with one or a few large green seeds.[1–4]

Origin: Native to Tropical Asia, introduced in Northern America.[5]

Phytoconstituents: Crinamine, lycoricidine, hamayne, isocraugsodine, palmilycorine, lycoriside, ambelline, crinasiatin, hippadine, bakonine, pratorimine, crinine, powelline, ungeremine, criasbetaine, crinasiatine, phenanthridone II and others.[6–17]

Traditional Medicinal Uses: In Southeast Asian countries, *C. asiaticum* has a considerable medicinal reputation as a potent folk medicine in the treatment

b716_Ch_21-30.indd 55 1/8/2009 10:18:02 AM

of injury and inflamed joints.[18] The plant has been used for carbuncles and cancer.[7] In Indonesia, the oiled and heated leaves are useful to treat wounds by poisoned arrows, bites and stings. In Malaysia, poultices of the leaves are applied to swellings, swollen joints, lumbago, pains and in cases of headache and fever. The leaves are also an emollient. In Northwest Solomon Islands, the leaves make a topical treatment for inflammation.[19] In Malaysia, the leaves are used as a rheumatic remedy and to relieve local pain.[18] On Karkar Island and in Simbu, Papua New Guinea, the latex from the leaves is applied to cuts. In India, the leaves are applied to skin diseases and inflammation.[20] The crushed leaves are used to wash piles or mixed with honey and applied to wounds and abscesses.[3] Its seeds are considered purgative and emmenagogic.[7] In the Trobriands, Papua New Guinea, the stem fibres are used to stop bleeding and in New Ireland, the milky sap from the stem is used for stonefish wounds.[20] The bulb is an emetic and counter-irritant. In Papua New Guinea, juice obtained from bulb is ingested regularly for 2 months to treat gonorrhoea. In the Philippines, the bulbs are crushed and applied as an ointment.[19] Juice from the fresh bulbs, taken several times per month induces vomiting. It is also instilled in the ear to treat otitis.[21] The root is also an emetic, diaphoretic and nauseant when fresh.[22] In a Finschhafen area village, Papua New Guinea, the cut root is cooked in a banana leaf, then cooled and placed on an aching tooth. Roots are used in New Caledonia, Indonesia, and Malaysia in a poultice for wounds, ulcers and swellings.[20]

Pharmacological Activities: Analgesic and Antibacterial,[19] Anti-inflammatory,[18] Antiviral,[18,23] Anticancer,[13,24] Antitumour,[15] Mast cells degranulation,[10] Antimitotic and Membrane stabilising.[8]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

27. *Cymbopogon citratus* (DC.) Stapf. (Gramineae)

Lemon grass

Cymbopogon citratus plants

Description: *Cymbopogon citratus* (DC.) Stapf. is a grass composed of dense leafy clumps that grows up to 1.8 m tall. Leaf sheath is tubular and acts as a pseudostem, 12–25 cm long, thickening towards the base and 1–2 cm in diameter. Leaf blades are sessile, simple, green, linear, glabrous, 60–90 cm by 1–2 cm, and possess parallel venation. Inflorescence is a raceme.[1–3]

Origin: Native to South Asia, Southeast Asia and Australia.[3]

Phytoconstituents: Citral, citronellal, cymbogonol, α-terpineol, citronellic acid, α-camphorene, geranial, isoorientin, isoscoparin and others.[1,4–8]

Traditional Medicinal Uses: The plant is used to treat digestive problems and relieve cramping pains.[4] A decoction of the plant is used by the Chinese to treat coughs, colds and blood in sputum.[2,9] The entire plant is used to treat

b716_Ch_21-30.indd 57 1/8/2009 10:18:09 AM

bacterial infection and possesses fever-reducing and stomachic properties.[5] In many Asian countries, leaves in water provide a bath to reduce swelling, to remove body odour, improve blood circulation, treat cuts, wounds, bladder problems and leprosy.[9] Its oil is used as an insect repellent, for aerosols, deodorants, floor polishes and household detergents.[4,5] The oil is carminative for cholera, and is prescribed for dyspepsia, vomiting, fever, and headache,[10] and used externally to treat eczema.[5] Its roots are taken to induce sweating, increase flow of urine,[9] treat coryza and influenza.[5]

Pharmacological Activities: Analgesic,[11,12] Anthelmintic,[13] Antibacterial,[14–18] Antifungal,[16,19–25] Anticancer/Antineoplastic,[26,27] Antimalarial,[28] Antioxidant,[8,29,30] Antiplatelet,[31] Hepatoprotective,[32] Hypoglycaemic,[33] Sedative,[34] Vasorelaxant,[35] Antimutagenic,[36] Insecticidal[37] and Radioprotective.[38]

Dosage: A dose of 10 to 20 g of roots is used for treating coryza, influenza and fever. It is also prescribed for dyspepsia and vomiting and as a carminative, by using 3 to 4 drops of the essential oil diluted in water.[5]

Adverse Reactions: Application of thick ointments with the volatile oil on the skin has led to rare incidence of allergy.[39]

Toxicity: Alveolitis occurred as a result of inhalation of the volatile oil.[39] It is also reported to be cytotoxic and genotoxic.[40]

Contraindications: Should not be used during pregnancy.[41]

Drug-Herb interactions: No information as yet.

28. *Dolichos lablab* L. (Leguminosae)

Lablab Bean, Hyacinth Bean

Dolichos lablab leaves

Dolichos lablab flowers and beans

Description: *Dolichos lablab* L. is a woody climbing herb which can reach a length of 5 m. Leaves are pinnate and generally 3-foliolate. Leaflets are acute, entire, 6–12 cm by 5–9 cm. Flowers are white or purplish pink. Fruits are green pods, 6 cm long by 2 cm wide, flattened, contain 4–5 seeds and turn light brown when mature.[1–4]

Origin: Native to Africa.[5]

Phytoconstituents: Dolichin, arabinogalactan 1 & 2, lablabosides A–F, phytin, pantothenic acid, saponin I, putrescine, spermidine, spermine and others.[3,6–14]

Traditional Medicinal Uses: The plant is decocted for alcoholic intoxication, cholera, diarrhoea, globefish poisoning, gonorrhoea, leucorrhoea and nausea.[15] Its seeds are used to stimulate gastric activities, as antidote against poisoning, to treat colic, cholera, diarrhoea, rheumatism and sunstroke.[4,16] The juice from the fruit pods are used for inflamed ears and throats. The fruit is also astringent, digestive, stomachic and used to expel worms. The seeds

b716_Ch_21-30.indd 59 1/8/2009 10:18:14 AM

are reportedly alexiteric, aphrodisiac, febrifuge, stomachic and used for menopause and spasms.[15] The flowers are used to treat dysentery, inflammation of uterus and to increase menstrual flow.[4,15] They are also used for leucorrhoea, menorrhagia, and summer heat disorders, as they have alexiteric and carminative properties.[15]

Pharmacological Activities: Antifungal,[6,17] Antiviral[6] and Haemagglutinating activities.[18–20]

Dosage: The reported dose of the seeds is 8 to 16 g daily in the form of powder or a decoction for the treatment of nausea, vomiting, diarrhoea, enteritis, abdominal pains and alcoholism.[21]

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

29. *Elephantopus scaber* L. (Compositae)

Elephant's Foot, Tutup Bumi, Tapak Sulaiman

Elephantopus scaber flowers

Elephantopus scaber herbs

Description: *Elephantopus scaber* L. is a tropical herb that can grow up to 30–60 cm tall. Leaves are simple, without stipules and when fully developed, form a rosette on the ground. Blades are obovate or oblong obtuse, hairy, large, 5–10 cm by 1.2–3 cm. The petioles are short, hairy, often crowded at the base of stem. Flowers are small and whitish pink.[1–3]

Origin: Native to Africa, tropical and temperate Asia and Australasia.[4]

Phytoconstituents: Molephantin, crepiside E, deoxyelephantopin, stigmasterol, stigmasteryl, scabertopin, lupeol and others.[5–9]

Traditional Medicinal Uses: In Vietnam, the plant is considered diuretic and administered at parturition.[10] A decoction of whole plant is bechic, used to treat pulmonary disease and scabies. In India, it is used as a tonic, laxative, analgesic, aphrodisiac and to treat inflammation. In the Philippines, the plant is febrifuge, diuretic and emollient.[10] In Malaysia, it is used as a preventive medicine after childbirth, to expel intestinal worms, for coughs and venereal diseases. In Chinese medicine, it is used to cure "dumpheat", which includes indigestion, swollen legs and loss of appetite.[1] In Indonesia, the plant is prescribed when there is a yellowish discharge from the vagina.[1] The plant is also considered as diuretic, emollient, used for relief of anuria and blennorrhoea,

b716_Ch_21-30.indd 61 1/8/2009 10:18:18 AM

remedy for leucorrhoea and anaemia, to treat fever, inflammation and as a disinfectant.[6] The whole plant is decocted for abscesses, cold, dysentery, oedema, gastroenteritis, gonorrhoea, influenza, pharyngitis and snakebite.[11] In Burma, a decoction of the stem and leaves is used in cases of menstrual disorders.[10] The roots are used for fever in children, on pimples, wounds of cattle, as an abortifacient, for urinary complaints, amoebic dysentery and other digestive problems.[12]

Pharmacological Activities: Antibacterial,[13] Anticancer/Antineoplastic,[14,15] Anti-inflammatory,[16] Antiviral[17] and Hepatoprotective.[18,19]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: Molephantin isolated from plant demonstrated cytotoxic activity.[5]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

30. *Euphorbia hirta* L. (Euphorbiaceae)

Asthma Weed

Euphorbia hirta herb

Description: *Euphorbia hirta* L. is an annual herb, which can grow up to 30 cm tall, and branched near the base. Leaves are elliptic-oblong, 2–3 cm by 0.8–1.5 cm. Inflorescence is small, in axillary dense clusters, each with one female flower and 4–5 male flowers inside.[1]

Origin: Native to the Pantropic.[2]

Phytoconstituents: Euphorbon, euphosterol, camphol, leucocyanidol, xanthorhamnin, taraxerol, taraxerone, myricitrin, euphorbianin and others.[3–6]

Traditional Medicinal Uses: The whole plant is decocted for athlete's foot, dysentery, enteritis, fever, gas, itch, and skin conditions.[3] It is also regarded as anodyne, depurative, diuretic, lactogogue, purgative, and vermifuge. The plant is used for asthma, bronchitis, calculus, colic, cough, dyspnoea,

b716_Ch_21-30.indd 63 1/8/2009 10:18:19 AM

eruptions, excrescences, influenza, fractures, gonorrhoea, headache, hypertension, measles, nausea, ophthalmia, sores, splinters, stomachache, tumours, urogenital ailments, warts and wounds.[3,7] In Central Province of Papua New Guinea, the plant is boiled and the solution is taken by patients who pass blood in the urine. The Chinese use the plant to treat fever, dysentery and skin conditions. In the Philippines and Indonesia, the plant is used to treat bowel problems.[8] The latex is used on warts and abscesses.[7]

Pharmacological Activities: Analgesic,[9] Antibacterial,[10–14] Anti-diarrhoeal,[6,15] Anti-inflammatory,[9,16,17] Antiplatelet,[18] Antiprotozoal,[19,20] Antipyretic,[9] Anxiolytic,[21] Diuretic,[22] Sedative,[23] Antianaphylactic[24] and Molluscicidal.[25,26]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: Toxic to brine shrimp.[11]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

31. *Eurycoma longifolia* Jack (Simaroubaceae)

Tongkat Ali, Ali's Umbrella, Pasak Bumi

Eurycoma longifolia leaves

Eurycoma longifolia fruits

Eurycoma longifolia tree

Description: *Eurycoma longifolia* Jack is a small tree with compound leaves on branches that can grow up to 1 m long. The numerous leaflets are opposite or subopposite, lanceolate to ovate-lanceolate, 5–20 cm by 1.5–6 cm, with smooth margins. Flowers are tiny, reddish, unisexual and are densely

b716_Ch_31-40.indd 65 1/8/2009 10:19:52 AM

arranged. The drupes are ovoid with a distinct ridge, 1–2 cm by 0.5–1.2 cm and they turn dark reddish brown when ripe.[1–3]

Origin: Native to Malesia and Indochina.[4]

Phytoconstituents: Eurycomalactone, eurycomanol, eurycomanone, eurylactone, eurylene, laurycolactone A and B, longilactone, pasakbumins A to D, eurycomalide A and B, piscidinol A and others.[3,5–19]

Traditional Medicinal Uses: The plant is used to cure indigestion and lumbago. It is used as a tonic after childbirth, to relieve pains in the bone and for treatment of jaundice, dropsy, cachexia and fever.[3,20] Tongkat Ali is one of the most well known folk medicines for intermittent fever (malaria) in Southeast Asia.[16] Decoction of the leaves is used for washing itches, while the fruits are used in curing dysentery.[3] Its bark is used as a vermifuge.[3] The taproots are used to lower high blood pressure, while the root bark is used for the treatment of fever and diarrhoea.[18] The roots of this plant are used as folk medicine for the treatment of sexual insufficiency, aches, persistent fever, malaria, dysentery, glandular swelling and also as health supplements.[18]

Pharmacological Activities: Antianxiety,[21] Antibacterial,[22] Anticancer,[6,8,17,18,24–26] Antitumour,[23,27] Antimalarial/Antiplasmodial,[5,13,18,23,28–32] Antischistosomal,[23] Antiulcer,[33] Aphrodisiac[15,18,34–38] and Plant growth inhibitor.[39]

Dosage: 1 g daily is recommended to be the maximum dose for supplemental use.[39]

Adverse Reactions: No information as yet.

Toxicity: One animal study found that the LD_{50} in mice was 1500–2000 mg/kg of the alcohol extract and 3000 mg/kg of the water extract. A subacute toxicity study with the alcohol extract indicated that 600 mg/kg daily was associated with signs of toxicity while 200 mg/kg daily was not, and another study found no toxic effects at 270–350 mg/kg daily but toxic effects were observed at 430 mg/kg daily.[39] Eurycomanone was identified as the most toxic component from its butanol extract.[40]

Contradindications: Should be used with caution and preferably not for extended periods without taking periodic breaks from use when it is used as a supplement. Tongkat Ali should not be taken by methods other than oral administration.[39]

Drug-Herb Interactions: No information as yet.

32. *Hibiscus mutabilis* L. (Malvaceae)

Cotton Rose, Chinese Rose

Hibiscus mutabilis flower

Hibiscus mutabilis tree

Description: *Hibiscus mutabilis* L. is a small tree that can grow up to 5 m tall. Leaf blades are heart-shaped, broadly ovate to round-ovate or cordate, 5–7-lobed, 10–15 cm in diameter, and papery. Abaxially, they are densely stellate and minutely tomentose; adaxially they are, sparsely stellate and minutely hairytoothed, 8–15 cm wide. Flowers are solitary and with multi-petals, white colour in the morning, changing to pink in the afternoon.[1–3]

Origin: Native to China.[4]

Phytoconstituents: Isoquercitrin, hyperoside, rutin, quercetin, naringenin, tetracosanoic acid, daucosterol, salicylic acid, quercimeritrin, meratrin and others.[5–10]

Traditional Medicinal Uses: The plant is used for fistulae, pustules and tumours.[11] The leaves and flowers are used as an analgesic, to expel phlegm,

b716_Ch_31-40.indd 67 1/8/2009 10:19:59 AM

treat excessive bleeding during menstruation, painful urination, inflammation and snake bites.[3] The leaves and flowers are also used as demulcent, diuretic and treat bacterial infection. They are used to treat boils, particularly on the chin, in the form of a poultice made of powdered dried leaves and flowers mixed with concentrated tea infusion which makes the boils burst earlier and less painfully. They are also used for treating impetigo, prurigo, metritis, leucorrhoea, mastitis, nephritis, cystitis, dysuria and infections.[12] The leaf is applied to swellings, crushed and compressed and applied onto abscesses, burns, and ulcers. It is also used as anodyne, alexipharmic, demulcent, expectorant, and refrigerant.[11] The flowers are used for lung ailments, with leaves for burns, inflammation, and snake bite. They are also prescribed for cough, dysuria and menorrhagia.[11]

Pharmacological Activities: Anti-inflammatory.[13]

Dosage: For the treatment of impetigo, prurigo, metritis, leucorrhoea, mastitis, nephritis, cystitis, dysuria and infections, a dose of 5 to 20 g of leaves and flowers is taken daily in the form of a decoction.[12]

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

33. *Hibiscus rosa-sinensis* L. (Malvaceae)

Hawaiian Hibiscus, China Rose, Bunga Raya

Leaves of *Hibiscus rosa-sinensis*

Hibiscus rosa-sinensis flower

Hibiscus rosa-sinensis shrub

Description: *Hibiscus rosa-sinensis* L. is a small evergreen perennial tree that can grow up to 3.6 m tall. It produces flower all year round. The large glossy leaves are ovate, alternate and vary in colour from pale green to dark green, with serrated edges. The flowers have five petals, and their colour varies but are mostly red.[1,2]

Origin: Native to South Eastern Asia and cultivated throughout the world as decorative plants.[3]

Phytoconstituents: Gossypetin, anthocyanin, myristic acid, palmitic acid, ambrettolide, campesterol, methyl sterculate, malvalate and others.[4–6]

Traditional Medicinal Uses: In Suquang, Papua New Guinea, the whole plant juice is applied directly to sores, to treat headaches and irregular

b716_Ch_31-40.indd 69 1/8/2009 10:20:04 AM

periods.[7] In Central Province, the plant is used to treat eye sores whereas in Northern Province and North Solomons Province, it is used to induce labour. In Indonesia, it is used as a purgative, an abortifacient and to regulate menstruation.[7] The juice also provides a soothing effect on mucous membrane that line the respiratory and digestive tracts.[5] The flowers and leaves are used to treat skin diseases, mumps, to relieve fever,[2,4] as well as to be used as emollient, anodyne and laxative.[8] The flower is also used as an astringent,[5] for excessive menstruation, fever and skin diseases. Its roots are used to treat gonorrhoea.[8] In Finschhafen, the roots and leaves are crushed and the juice is drunk to treat diarrhoea.[7]

Pharmacological Activities: Antianxiety,[9] Anticancer,[10] Anticonvulsant,[9] Antifertility,[11–14] Antioxidant,[15] Hepatoprotective,[16] Hypoglycaemic,[17–19] Hypolipidaemic[18,20] and Wound healing.[21]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: Not to be taken by small children, and during pregnancy and lactation.[22]

Drug-Herb Interactions: No information as yet.

34. *Hibiscus tiliaceus* L. (Malvaceae)

Linden Hibiscus, Sea Hibiscus, Mahoe

Leaves of *Hibiscus tiliaceus*

Hibiscus tiliaceus flowers

Description: *Hibiscus tiliaceus* L. is an evergreen coastal tree that can grow up to 10 m tall. Leaves are heart-shaped, 10–15 cm long and wide, leathery and green. The large showy flowers are yellow with maroon centre. Fruit is capsular, subglobose to ovoid, about 2–3 cm long and wide, surrounded by the calyx. Seeds are reniform, smooth and glabrous.[1–4]

Origin: Native to the Pantropic.[4]

Phytoconstituents: Hibiscones A–D, hibiscusin, hibiscusamide, friedelin, epifriedelanol, pachysandiol A and others.[2,5–7]

Traditional Medicinal Uses: In Malaysia and Indonesia, the leaves are considered cooling and are used to treat fever. They are also used as a soothing agent and to remove phlegm from the respiratory passages.[2] The flowers are widely used as a means of birth control in traditional medicine in Asian and African countries.[8] In the Philippines, the mucilaginous water obtained by soaking the fresh bark in water is prescribed for dysentery[2] while the root is an ingredient of embrocations and is used as a febrifuge.[8]

Pharmacological Activities: Tyrosinase inhibitory,[9] Antioxidant[10,11] and Cytotoxic.[5]

Dosage: No information as yet.

b716_Ch_31-40.indd 71 1/8/2009 10:20:10 AM

FA

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

b716_Ch_31-40.indd 72 1/8/2009 10:20:15 AM

35. *Impatiens balsamina* L. (Balsaminaceae)

Balsam Plant

Impatiens balsamina plants

Description: *Impatiens balsamina* L. is an annual, erect, succulent plant, which is about 30–90 cm tall. Leaves are simple, alternate, 15 cm long, serrated and arranged spirally. Flowers are white, pink, purple or variegated and usually in leaf axils. Fruits are capsular, tomentose and contain several seeds.[1–5]

Origin: Native to India.[5]

Phytoconstituents: Impatienol, pelargonidine, delphinidine, cyanidine, balsaminones A and B, hosenkosides F–O and others.[4,6–10]

Traditional Medicinal Uses: A lotion of fresh leaves is used to treat eczema, itches and insect bites.[4] In Vietnam, decoctions of leaves are used to stimulate growth and to wash hair.[4] The juice is also used for warts, cancer treatment and expectorant.[11] A decoction of flowers is taken for infections,

b716_Ch_31-40.indd 73 1/8/2009 10:20:17 AM

vomiting, urine retention and as a tonic. In India, flowers are regarded as cooling, tonic and useful when applied to burns and scalds.[4] The flowers are also used for lumbago and intercostal neuralgia, snakebite, improves circulation and relieves stasis.[11] In Japan, juice squeezed from the white flower petals are applied on the skin to alleviate dermatitis.[6] In China, the seeds are prescribed for difficult labour, puerperal pains, difficult menstruation, cough, hiccups and poisonings. The seeds are mixed with arsenious acid for removing teeth.[4] In Malaysia, the seeds are taken for gastrointestinal tract cancer, and to dislodge fish or chicken bones in throat.[4]

Pharmacological Activities: Antibacterial,[12–15] Antihypotensive,[16,17] Antifungal,[18] Anti-inflammatory[4–8] and Antipruritic.[6,19]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

36. *Imperata cylindrica* (L.) P. Beauv. (Gramineae)

Lalang, Alang-alang, Speargrass

Imperata cylindrica

Inflorescences of *Imperata cylindrica*

Imperata cylindrica fruits

b716_Ch_31-40.indd 75 1/8/2009 10:20:21 AM

Description: *Imperata cylindrica* (L.) P. Beauv. is a perennial erect grass about 0.6–1.5 m tall. Leaf blades are linear to lanceolate, long, 0.5–1 cm by 15–30 cm, narrow, with sharp margins and prominent nerves. Inflorescences are white and fluffy. Rhizome is hard, coriaceous and deep within the soil.[1]

Origin: Native to China and Africa but this widespread weed can be found throughout the world.[2]

Phytoconstituents: Arborinione, arundoin, anemonin, isoorientin, imperanene, cylindol A and B, graminones A and B, cylindrene, isoarborinol, impecyloside and others.[3–10]

Traditional Medicinal Uses: Rotted grass from thatch boiled with wine is used to treat bug bite, haemoptysis, severe constipation and vaginismus. It is also used for drug withdrawal symptoms.[3] Besides being used to quench thirst, its flowers and rhizome are also used to treat blood in the sputum, nose bleeds, lung and kidney diseases, jaundice,[11] haemorrhage, wounds, haemoptysis, epistaxis, haematemesis, haematuria, nephritic oedema, high fever, and urinary tract infections.[3] The roots are used to treat fever, cough with phlegm,[11] asthma, cancer, dropsy, epistaxis, haematuria, jaundice, nephritis, diarrhoea, gonorrhoea[3] and dysuria.[12]

Pharmacological Activities: Antidiuretic,[12] Anti-inflammatory,[6] Neuroprotective[9] and Antibacterial.[13]

Dosage: Rhizome is used for the treatment of urodynia, pollakiuria, haematuria and fever. The recommended usual daily dose is 10 to 40 g in the form of a decoction, to be administered orally.[1]

Adverse Reactions: Allergic contact dermatitis[14] and sensitivity to grass pollen (Type I allergy).[15]

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

37. *Ipomoea pes-caprae* (L.) Sweet (Convolvulaceae)

Beach Morning Glory, Goat's Foot Creeper, Bayhops

Ipomoea pes-caprae plants

Ipomoea pes-caprae flower

Description: *Ipomoea pes-caprae* (L.) Sweet is a seaside trailing herb. The stems are long and the roots are found at the nodes. Leaves are simple, ovate, quadrangular or rounded, 2.5 cm by 10 cm, with slender petioles. Cymes are 1–3 flowered; the 5 sepals are ovate and blunt. The flower funnel is about 3–5 cm long. The plant bears dehiscent capsules (1.2 cm) containing hairy seeds that break up easily.[1–4]

Origin: Native to the Pantropic.[5]

Phytoconstituents: Pescaproside A & B, pescapreins I–IX, stoloniferin III, E-phytol, β-damascenone and others.[4,6–12]

Traditional Medicinal Uses: The juice squeezed from the plant is used by the Malays to treat fish stings.[3] In Burma, infusions of the plant with rusted iron are used to cure gynaecologic haemorrhages. [4] In India, the plant is known to be cooling, astringent and laxative. It is used internally as an emollient in Vietnam.[4,13] The plant has also been administered for headache.[14] In Thailand, it is used to treat dermatitis caused by jellyfish (e.g., Portuguese man-of-war).[15] The leaves are used in Indonesia to hasten the bursting of boils and used externally for injured feet. Sap from the young leaves are boiled in coconut oil and used to treat sores and ulcers. In the Philippines, the

b716_Ch_31-40.indd 77 1/8/2009 10:20:27 AM

boiled leaves are used to treat rheumatism and fungal infection.[3,4,13] In Papua New Guinea, decoctions of the leaves are applied to sores. On the Solomon Islands, preparations of the leaves are ingested for digestive troubles.[4] The seeds are chewed with areca nut to sooth abdominal pains and cramps or consumed alone as a purgative.[3,13] Decoctions of the tubers are also ingested to treat bladder problems.[4]

Pharmacological Activities: Analgesic,[16] Antidiabetic,[17] Anti-inflammatory,[1,16,18] Antiplatelet aggregation,[14] Antiviral,[4] Antivenom,[4,19] Collagenase inhibitory activity,[12] Muscle relaxant[4,15] and Prostaglandin synthesis inhibition.[9]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

38. *Ixora chinensis* Lam. (Rubiaceae)

Chinese Ixora

Ixora chinensis flowers

Ixora chinensis shrub

Description: *Ixora chinensis* Lam. is a small, dense shrub that can grow up to 2 m tall. Leaves are short-stalked, obovate-oblong, waxy, 6–10 cm long. Flowers are densely arranged, with 4 petals in bright red.[1]

Origin: This common garden flower is native to China and Thailand.[1]

Phytoconstituents: Ixoric acid, ixoroside, ixoside, geniposidic acid and others.[2–4]

Traditional Medicinal Uses: It is used for hypertension.[5] The whole plant is used to treat rheumatism, abscesses, bruises, and wounds. It is also used as an anodyne, and resolvent. It is beneficial to bone marrow and pregnant uterus as well.[6,7]

Pharmacological Activities: No information as yet.

Dosage: No information as yet.

b716_Ch_31-40.indd 79 1/8/2009 10:20:31 AM

FA

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

b716_Ch_31-40.indd 80 1/8/2009 10:20:37 AM

39. *Jatropha curcas* L. (Euphorbiaceae)

Barbados Nut, Physic-nut

Flowers of *Jatropha curcas*

Jatropha curcas leaf

Jatropha curcas shrub

Description: *Jatropha curcas* L. is a woody shrub, 2–5 m tall. Stems are smooth and leaves are ovate, 5–15 cm long, green and glabrous. The blade is thin and 5-lobed-cordate. The flowers are 6 mm by 7 mm, hermaphrodite, and with fragrance. Fruit is 3-lobed fleshy capsule of about 2.5 cm in length, containing black seeds.[1,2]

Origin: Native to Mexico,[2] widely cultivated and naturalised in New and Old World Tropics.[3]

Phytoconstituents: Curcain, curcasin, curcacycline A, jatropholone A and B, heudelotinone, nobiletin, jatrocurin, curcin, curcusone B and others.[4–14]

b716_Ch_31-40.indd 81 1/8/2009 10:20:37 AM

Traditional Medicinal Uses: The leaves are used to treat scabies, parasites and as a rubefacient for paralysis and rheumatism.[15] The fruit is used for dropsy and anasarca.[16] The seed oil is emetic, laxative, purgative and for skin ailments.[4] The latex is applied to bee and wasp stings, to dress ulcers, sores and inflamed tongues. It is also used as a haemostatic agent,[15] and to treat whitlow, carbuncle and sores in mouth.[16] The roots are used in the form of a decoction as mouthwash for bleeding gums and toothache.[15]

Pharmacological Activities: Antidiabetic,[17] Anti-inflammatory,[18] Anticancer,[13,14,19] Antiprotozoal,[20] Antiviral,[21,22] Coagulant,[23] Abortifacient,[24] Haemolytic,[25] Lipolytic,[26] Insecticidal,[27–29] Molluscicidal[30] and Wound healing.[31]

Dosage: Two seeds have been used as a purgative.[32]

Adverse Reactions: No information as yet.

Toxicity: Human poisoning may lead to amnesia, convulsions, delirium, diarrhoea, nausea, vertigo, visual disturbances, acute abdominal pain, nausea, depression and collapse may also occur.[32] Intake of 4–5 seeds may cause death. It is toxic to rats, chicks and human.[32–40]

Contraindications: No information as yet.

Drug-Herb Interactions: May have interactions with other drugs that are metabolised by cytochrome P450 enzymes.[20]

[**Authors' Note:** *Jatropha curcas* is also cultivated as a source of biofuel.]

40. *Juniperus chinensis* L. (Cupressaceae)

Chinese Juniper

Leaves of *Juniperus chinensis*

Juniperus chinensis tree

Description: *Juniperus chinensis* L. is conical shaped tree that can grow up to 25 m tall. Leaves are both scale-like and needle-like. Needle-like leaves are present on both young and adult plants, decussate or in whorls of three, loosely arranged, ascending, nearly lanceolate whereby scale-like leaves are present only in the adult. Seed cones are brown when ripe, usually glaucous, subglobose and 4–9 mm in diameter.[1]

Origin: Native to Myanmar and temperate Asia.[2]

Phytoconstituents: Chinensiol, yatein, podophyllotoxin, thujopsenal, widdrol, sandaracopimaric acid, hinokiic acid and others.[3–11]

b716_Ch_31-40.indd 83 1/8/2009 10:20:40 AM

Traditional Medicinal Uses: The leaves are used as a tonic to treat bleeding resulting from coughs.[12] It is also used for cold and haemorrhage. Liquor brewed from fresh leaves is used as a tonic and to treat hemoptysis. Others include treatment for convulsions, excessive sweating and hepatitis.[3] Its roots are used on burns, scalds and to promote hair growth on scars.[3]

Pharmacological Activities: Antibacterial,[13] Antifungal,[13,14] Anti-neoplastic,[15] Antioxidant,[6] Hypoglycemic[16] and Hypolipidaemic.[8,16]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

41. *Kaempferia galanga* L. (Zingiberaceae)

Galangal, Sand Ginger, Kencur

Kaempferia galanga herbs

Description: *Kaempferia galanga* L. is a small herb with short underground stems. Leaves are usually in pairs, oval, glabrous, pointed, 6–15 cm long, and spread out above ground with prominent veins. Flowers are in short stalked spikes. The corolla is white or pinkish, with violet spotted lip.[1]

Origin: Native to tropical Asia.[2]

Phytoconstituents: Ethyl cinnamate, 1,8-cineole, δ-3-carene, α-pinene, camphene, borneol, cyene, α-terpineol, α-gurjunene, germacrenes, cadinenes, caryophyllenes and others.[3–8]

b716_Ch_41-50.indd 85 1/8/2009 10:23:14 AM

Traditional Medicinal Uses: The whole plant is used as a postpartum protective medicine, treatment for stomachache, diarrhoea, dysentery, treatment for rheumatism, swellings, fever, coughs, asthma and as a tonic/lotion.[8] In Malaya, the leaves and rhizomes are chewed to stop cough. In Indonesia, it is used for abdominal pain, for swelling and muscular rheumatism. In the Philippines, the rhizome is used for boils, chills, dyspepsia, headache and malaria. The Indians also use the rhizomes as lotions, poultices for fever, rheumatism, sore eyes, sore throat and swellings.[3] The rhizomes are stimulant, used to treat toothache, chest pains and constipation.[9] They are also used as carminative, prophylactic, stomachic, for dandruff and scabs. A decoction of the rhizome is used for cholera, contusion, dyspepsia, headache, lameness, lumbago, and malaria. It is also roasted and applied to rheumatism and tumours.[3] To facilitate delivery during birth, it is mixed with the juice of *Curcuma montana, C. aromatica* and ginger rhizomes and consumed.[10]

Pharmacological Activities: Antibacterial,[11,12] Antifungal,[13] Antihypertensive,[14] Anti-inflammatory,[8] Antineoplastic,[15–18] Antioxidant,[19] Antiprotozoal,[20] Depressant,[21] Immunomodulatory,[21] Vasorelaxant,[22] Antiallergy,[23] Insect repellent,[24,25] Insecticidal[25–28] and Wound healing.[29]

Dosage: Oral doses range from 3 to 6 g of the rhizome per day, administered in the form of decoction, powder or pill for the treatment of pectoral and abdominal pains, headache, toothache and cold.[30]

Adverse Reactions: No information as yet.

Toxicity: The ethanolic extract injected intraperitoneally in increasing doses of 25, 100, 250 and 800 mg/kg body weight resulted in a decrease in motor activity, respiratory rate, loss of screen grip and analgesia in rats. A dose of 2000 mg/kg body weight was observed to be lethal. Acute and subacute oral toxicity test of *Kaempferia galanga* produced neither mortality nor significant differences in the body and organ weights between controls and treated rats.[21]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

42. *Lantana camara* L. (Verbenaceae)

Common Lantana, Bunga Tahi Ayam, Wild Sage

Lantana camara flowers

Description: *Lantana camara* L. is a small shrub with long, weak branches. Leaves are opposite, ovate, 5–12 cm long with a pungent scent. Flowers are in dense spikes, with salver-shaped corolla, 1–1.2 cm long, orange, red, pink or variegated. Fruit is bluish, globose and 4 mm in diameter.[1]

Origin: Native to Mexico and Southern America.[2]

Phytoconstituents: Lantadene A and B, lantanic acid, lantanilic acid, icterogenin, lantanose A & B, lamiridoside, geniposide, δ-guaiene, camarinic acid and others.[3–14]

Traditional Medicinal Uses: The whole plant is used as a bath for scabies and leprosy.[3,15] The entire plant is also a carminative, diaphoretic, vulnerary, used for fistulae, pustules, tumours, treating malaria, rheumatism, tetanus and spasms.[3] In Budiope county, Uganda, the leaves are used for treating malaria.[16] The root and leaves are decocted to treat fever, mumps, neurodermatitis and traumatic injuries.[3,16]

b716_Ch_41-50.indd 87 1/8/2009 10:23:20 AM

FA

Pharmacological Activities: Anthelmintic,[7,17] Antimutagenic,[13] Antibacterial,[13,18–21] Anticancer/Antineoplastic,[22,23] Antifertility,[24] Anti-inflammatory,[25] Antimicrobial,[26–28] Antimotility,[29] Antioxidant,[30] Antithrombotic,[31] Antiviral,[32] Antifilarial,[33] Insect repellent,[34,35] Insecticidal,[36–39] Larvicidal,[40] Molluscicidal[10,41] and Spermatotoxic.[42]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: Ingestion of green berries had been reported to cause human fatalities.[43] *L. camara* is harmful to animals[44–58] and causes teratogenicity in rats.[59]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

b716_Ch_41-50.indd 88 1/8/2009 10:23:26 AM

43. *Lonicera japonica* Thunb. (Caprifoliaceae)

Japanese Honeysuckle, Jin Yin Hua

Lonicera japonica shrub

Lonicera japonica leaves

Description: *Lonicera japonica* Thunb. is a climbing shrub having tomentose young leaves and stems. Leaves are simple, opposite and exstipulate. Blade is elliptic, 3–8 cm by 2–3 cm, truncate at base, obtuse and chartaceous. Flowers are axillary, white, and turns yellow upon maturity. Fruits are globose and black.[1]

Origin: A native of East Asia, widely cultivated and naturalised throughout the world.[2]

Phytoconstituents: Linalool, luteolin, geraniol, aromadendrene, eugenol, loniceroside A, B, C, L-phenylalaninosecologanin, (Z)-aldosecologanin, (E)-aldosecologanin and others.[3–8]

Traditional Medicinal Uses: In China, the flowers are used for influenza, boils and carbuncle.[3] In Malaysia, decoctions of dried flowers are used for cooling, flu, fever, headache, and boils. Distilled flowers are used to produce

b716_Ch_41-50.indd 89 1/8/2009 10:23:26 AM

a medicine for treating postprandial stomachaches.[3] Flower tea is prescribed to treat fever, sore throat, mouth sores, headache, conjunctivitis, keratitis, corneal ulcers, breast infections, muscle and joint pain, stomach problems, diarrhoea, and painful urination.[9] They are used in the treatment of arthritis and inflammation.[10] Flower buds are used in infusions for cutaneous infections, scabies, as diuretic and treat bacterial infection. Decoction is used for bacterial dysentery, cold, enteritis, infected boils, laryngitis, lymphadentitis, rheumatism and sores. The flowers and stems are regarded as cooling and are used to treat aching bones and boils. Other uses include intestinal inflammation, stomach ulcers, painful haemorrhoids, sore throats and intoxication.[3]

Pharmacological Activities: Antibacterial,[11,12] Anticancer/Antineoplastic,[12–15] Antifungal,[16] Antihypertensive,[17] Anti-inflammatory,[9,18–25] Antioxidant,[7] Antiplatelet,[26] Antiviral,[27] Hepatoprotective[28] and Antiatherogenic.[29]

Dosage: In China, 10–60 g of dried floral buds are used for decoction. A combination of 10 g of honeysuckle, 10 g of forsythia, a little mint and bamboo leaf is a prescription for a bad cold from a drugstore in China.[30] About 9–15 g dried flowers has been used in decoction, pills, powder or poultice of the powder.[31] The recommended daily dose is 4 to 8 g of flowers or 10 to 20 g of stems and leaves in the form of a decoction, infusion, extract or alcoholic maceration for the treatment of boils, impetigo, urticaria, allergic rhinitis, fever, malaria, erythema, measles, diarrhoea, dysentery, syphilis, rheumatism and lichen tropicus.[32] For carbuncles, boils, erysipelas, acute dysentery, pharyngitis, upper respiratory infection and epidemic febrile diseases, 6–15 g of dried flowers are used.[33]

Adverse Reactions: Linear, itchy, raised blisters on the skin may occur on contact.[34]

Toxicity: No sign of acute and subacute toxicity was observed.[35]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

44. *Mangifera indica* L. (Anacardiaceae)

Mango, Mangga

Mangifera indica tree

Fruits of *Mangifera indica*

Description: *Mangifera indica* L. is a fruit tree which grows up to 8 m high. Bark is grey and fissured. Leaves are simple, 12–30 cm by 4–9 cm, narrowly elliptic, pointed, and slightly leathery with wavy edges. Flowers are very small, greenish-yellow or white, fragrant and arranged in panicles. The fruit is kidney-shaped and the yellow to orange flesh is edible. Seed is elongated, fibrous and flattened.[1]

Origin: Native to India and Indochina.[2]

Phytoconstituents: Mangiferin, ambolic acid, ambonic acid, arabinan, mangiferonic acid, quercitin, violaxanthin and others.[3–7]

Traditional Medicinal Uses: The leaves are used in the form of ashes for burns and scalds; chewed to strengthen the gums, while the burning leaf smoke is inhaled for hiccups and other throat ailments. They are also used for skin ailments, asthma and cough.[4] Paste of leaves is applied to warts and used as styptic ointments.[3] The leaves are also used in the treatment of malaria in Budiope county, Uganda.[8] The fruits are used to treat pain in abdomen, diarrhoea, and to quench thirst (with Aegle and salt).[9] Pulp of the

b716_Ch_41-50.indd 91 1/8/2009 10:23:29 AM

fruit is used in China to promote blood circulation, the fruit rind as tonic in Burma and dried slices of young fruits are used for septicaemia in Palau.[3] Its seeds are used for stubborn colds, coughs, diarrhoea, vermifuge and menorrhagia.[4] They are also used for asthma.[9] In the Philippines, raw seeds are used to expel worms and the roasted ones are for diarrhoea.[3] The bark is used for the treatment of fever or sunstroke, cholera, rheumatism, sty in eye, ulcerated tongue (with roots of *Ichnocarpus* and bark of *Zizyphus rugosa*), haemiplegia, diarrhoea (with bark of *Streblus asper*, roots of *Oroxylum indicum* and *Helianthus annuus*), dysentery (with bark of *Streblus* and *Spondias pinnata*), poisoning, uterine haemorrhage and jaundice.[9]

Pharmacological Activities: Analgesic,[10] Anthelmintic,[11] Antibacterial,[12–14] Anticonvulsant,[15] Antidiarrhoeal,[13] Antifungal,[16] Anti-inflammatory,[10,11,17–19] Anticancer,[20-23] Antioxidant,[24–38] Antiviral,[39,40] Gastroprotective,[41,42] Hepatoprotective,[43,44] Antiprotozoal,[15] Hypoglycaemic,[45–47] Hypolipidaemic,[46,48] Immunomodulatory,[49] Larvicidal[50] and Radioprotective.[51,52]

Dosage: As an anthelmintic, it is recommended to take 20–30 g of powdered seed.[53]

Adverse Reactions: Anaphylactic reaction following ingestion of fruit.[54]

Toxicity: Intraperitoneal LD_{50} (50% ethanolic extract) is > 1000 mg/kg in mice.[55]

Contraindications: No information as yet.

Drug-Herb Interactions: Modulation of P450 isozymes.[56,57]

45. *Manihot esculenta* Crantz (Euphorbiaceae)

Tapioca, Cassava

Leaves of *Manihot esculenta*

Manihot esculenta shrub

Description: *Manihot esculenta* Crantz is an erect woody shrub that can grow up to 5 m tall. Leaf blades are palmately 3–9-lobed, 5–20 cm. Lobes are oblanceolate to narrowly elliptic, entire, 8–18 cm by 1.5–4 cm. Root tubers beneath the ground yield tapioca, which is fleshy, starchy and edible.[1,2]

Origin: Native to Brazil, cultivated throughout the tropics.[2]

Phytoconstituents: Linamarin, esculentoic acid A and B, esculentin, esculin, scopoletin, scopolin, oxalic acid, saponins and others.[3–8]

Traditional Medicinal Uses: Rhizome is made into a poultice and applied to sores.[5]

Pharmacological Activities: Antifungal,[6] Antineoplastic,[7] Hepatoprotective,[9] Hypercholesterolaemic,[8] Antithyroidal,[10] Neurotoxic[11] and Superoxide dismutase inhibition.[12]

Dosage: No information as yet.

b716_Ch_41-50.indd 93 1/8/2009 10:23:37 AM

Adverse Reactions: Several minor acute intoxications were seen, with complaints of nausea, vomiting, abdominal pain and headache following a meal of cassava.[12]

Toxicity: Hydrocyanic acid — oral, human LD_{50} 570 μg/kg; Oxalic acid — oral, human LD_{50} 700 mg/kg; Saponin — oral, mouse LD_{50} 3000 mg/kg; Tryptophane — oral, rat TD_{50} 1100 mg/kg. Bitter cassava juice can cause death due to cyanide poisoning.[3] Haemorrhage, necrosis, fibrosis and atrophy of the acinar tissue and fibrosis of the islets of Langerhans of the pancreas occurred in dogs fed on cassava.[13] Epidemic spastic paraparesis (konzo) is known to be due to long-term intake of cassava (*M. esculenta*).[10] Tropical ataxic neuropathy, a polyneuropathy with sensorineural hearing loss and optic atrophy can also result from intake of cassava in humans.[14]

Contraindications: No information as yet.

Drug-Herb Interactions: Co-administration of cassava rich diet and alcohol is found to reduce the alcohol induced toxicity in rats.[7]

46. *Melaleuca cajuputi* Roxb. (Myrtaceae)

Gelam, Paper-bark Tree, Kayu Puteh

Leaves of *Melaleuca cajuputi*

Melaleuca cajuputi flowers

Melaleuca cajuputi tree

Description: *Melaleuca cajuputi* Roxb. is a medium sized tree with an often twisted trunk when the tree is very old. Bark is light brown, flaky and peeling. Leaves are simple, slightly curved, with 5–7 longitudinal veins. Flowers are white and borne in dense spikes. Seeds are borne in a capsule.[1–3]

b716_Ch_41-50.indd 95 1/8/2009 10:23:41 AM

Origin: Native to tropical Asia and Australasia.[4]

Phytoconstituents: Cajeputol, cineole, β-pinene, eugenol, phellandrene, α-terpineol, eugenetin, isoeugenetin and others.[2,5,6]

Traditional Medicinal Uses: In Malaysia, it is used for the treatment of colic and cholera. It is also used externally for thrush, vaginal infection, acne, athlete's foot, verruca, warts, insect bites, cold sore and nits.[5] Cajuput oil is distilled from the leaves and used by the Burmese to treat gout. The Indochinese uses cajeput oil for rheumatism and pain in the joints and as an analgesic.[2] The oil is used externally in Indonesia for burns, colic, cramps, earache, headache, skin diseases, toothache and wounds. When administered internally, it can induce sweating and act as a stimulant and antispasmodic. In the Philippines, the leaves are used to treat asthma.[2,6]

Pharmacological Activities: Antibacterial.[6]

Dosage: For treatment of coryza, influenza, cough, asthma, dyspepsia, earache, toothache, rheumatism, osteodynia, neuralgia, wounds, burns, post partum haematometra. The reported dose for the above ailments is 20 to 40 g of fresh leaves or 5 to 10 g of dried leaves in the form of a decoction or infusion.[7]

Adverse Reactions: Contact dermatitis may occur. Glottal spasms or bronchial spasms or asthma-like attacks may occur if the oil is applied to the facial areas of infants and small children.[8]

Toxicity: Due to the high cineole content, life-threatening poisonings can occur with overdoses of cajuput oil (more than 10 g). Symptoms of poisoning are reduction in blood pressure, circulatory disorders, collapse and respiratory failure.[8]

Contraindications: Do not consume orally in severe liver diseases and presence of inflammatory condition of gastrointestinal tract or of the biliary ducts. Cajuput oil preparations are not to be applied to the faces of infants or small children.[8]

Drug-Herb Interactions: No information as yet.

47. *Melastoma malabathricum* L. (Melastomaceae)

Sendudok, Singapore Rhododendron

Melastoma malabathricum flower

Fruits of *Melastoma malabathricum*

Melastoma malabathricum shrub

Description: *Melastoma malabathricum* L. is a small shrub that can grow up to 1 m tall, densely clothed with tiny scales. Leaves are simple, ovate, elliptic, 4–14 cm by 1.7–3.5 cm, and stiffly papery. Flowers are purple with five petals and fruits are urceolate-globular, 6–15 mm by 6–12 mm and succulent.[1,2]

Origin: Native to tropical and temperate Asia and the Pacific Islands.[2]

Phytoconstituents: Malabathrins A–F, nobotanins B, G, H and J, kaempferol and others.[3–9]

Traditional Medicinal Uses: In Indonesia and Malaysia, the leaves and roots are used as a remedy for diarrhoea and dysentery. The leaves are also used to

b716_Ch_41-50.indd 97 1/8/2009 10:23:48 AM

wash ulcers, to treat piles and for small pox pustules to prevent development of scars.[10] Its bark is used to treat skin diseases.[11]

Pharmacological Activities: Analgesic,[12] Anti-inflammatory,[13] Antineoplastic,[14] Antinociceptive,[13] Antiviral,[14,15] Antioxidant[15] and Antiplatelet.[16]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: *M. malabathricum* was reported to be an aluminium (Al) accumulator (with up to 10 g Al/kg of dry weight in old leaves and up to 7 g Al/kg dry weight in new leaves) and could lead to accumulation of aluminium in the bone and brain, leading to neurotoxicity, when consumed in large amount.[17]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

48. *Mimosa pudica* L. (Leguminosae)

Touch-me-not, Sensitive Plant, Rumput Simalu

Mimosa pudica leaves and flowers

Description: *Mimosa pudica* L. is a prickly, herbaceous weed. Leaves bipinnate, very sensitive, fold together when touched, in rain or at night. Leaflets are 15–20 pairs, small oblong, nearly sessile. Flowers are pink and fruits are flat pods covered with bristles.[1–3]

Origin: Native to Brazil.[3]

Phytoconstituents: Mimosine, 2-Hydroxymethyl-chroman-4-one and others.[4–6]

Traditional Medicinal Uses: The plant is used on cuts and wounds.[4,7] It is also used for childbirth and infertility in Trinidad and Tobago.[8] Bath with plant decoction relieves insomnia.[4] The leaves are used for hydrocele, dressing for sinus, sores, piles and swelling of feet. In Mexico, aqueous extracts

b716_Ch_41-50.indd 99 1/8/2009 10:23:56 AM

from dried leaves are employed to alleviate depression.[9] In the Philippines, the leaves soaked in coconut oil is used for ulcers.[4]

Pharmacological Activities: Anthelmintic.[10] Antibacterial,[11] Anticonvulsant,[12] Antidepressant,[9] Antifertility,[13] Antifungal,[5] Hyperglycaemic,[14] Antioestrogenic[15] and Antivenom.[16,17]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

49. *Mirabilis jalapa* L. (Nyctaginaceae)

Four O'Clock Flower, Bunga Pukul Empat

Flowers of *Mirabilis jalapa*

Description: *Mirabilis jalapa* L. is an erect herb that can grow up to 1 m tall. Leaves are simple, heart-shaped, 3–12 cm long, opposite, tapering to a pointed end. Flowers are bisexual, red, pink, yellow or white, with perianth distinctly constricted above, and they bloom late in the afternoon. Fruits are black and globose, 5–8 mm in diameter.[1,2]

Origin: Native to tropical America, introduced in China and in many tropical areas.[2]

Phytoconstituents: Trigonellin, 2′-*O*-methylabronisoflavone, 6-methoxy-boeravinone C and betaxanthins.[3,4]

Traditional Medicinal Uses: Its leaves are used as a decoction for abscesses, juice for wounds and cooked with pork as tonic.[5] The leaves are also placed

b716_Ch_41-50.indd 101 1/8/2009 10:24:03 AM

on boils, blisters, and to relieve urticaria.[6] In Indochina, the seeds are used as a purgative.[3] The flowers release a strong odour at night which will stupefy or drive away mosquitoes.[7] The roots are used as a laxative in a decoction, with or without pork, for colds, inflammation and leucorrhoea.[5] In Malagassy, Madagascar, they are used to treat intestinal pains. In South Africa, the roots are used as purgative agents.[7]

Pharmacological Activities: Antibacterial,[8,9] Antifungal,[4,10] Antineoplastic and Abortifacient.[11]

Dosage: Approximately 8–10 g of roots are taken as a purgative.[12]

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

50. *Morinda citrifolia* L. (Rubiaceae)

Mengkudu, Indian Mulberry, Noni

Flowers and fruits of *Morinda citrifolia*

Morinda citrifolia tree

Description: *Morinda citrifolia* L. is a tree that grows up to 9 m. Leaves are simple, large, 10–15 cm by 20–30 cm, decussate and stipulate. Blade is broadly elliptic to obovate, glossy, soft and succulent. Flowers are small, sessile, white, 1.5–2 cm across, terminal and axillary. The fruit is succulent, oval, 5–7 cm across, light greyish green and turns yellow upon maturity.[1–5]

Origin: Native to tropical and temperate Asia and Australasia.[6]

Phytoconstituents: Morintrifolins A and B, morindin, morindone, rubiadin, morindadiol, morindicone, morinthone, morindicinone, morindicininone, noniosides E–H, morinaphthalenone and others.[1,4,7–28]

Traditional Medicinal Uses: The whole plant is used to treat aching bones and arthritis. In Malaysia, the heated leaves are applied to the chest and abdomen to treat coughs, nausea, colic and enlarged spleen. In Japanese and Chinese medicine, *M. citrifolia* is used to treat fever and as a tonic whereas in Indochina, the fruit is prescribed for lumbago, asthma and dysentery.[1] A decoction of the leaves taken orally is effective for the treatment of fever,

b716_Ch_41-50.indd 103 1/8/2009 10:24:09 AM

dysentery and diarrhoea. Poultice of fresh leaves can be used to cure furunculosis. The fruit when consumed together with a little salt is stomachic, aperient, and active on dysentery, uterine haemorrhage, cough, coryza, oedema and neuralgia. The root bark has beneficial effects in hypertension, osteodynia and lumbago.[7] The fruit is also used for throat and gum complaints, dysentery and leucorrhoea while the root is used as a cathartic and febrifuge.[29] The processed fruit juice is also in great demand for various ailments such as diabetes, high blood pressure, headaches, heart disease, AIDS, cancers, gastric ulcers, mental depression, senility, poor digestion, atherosclerosis and drug addiction.[30]

Pharmacological Activities: Analgesic,[30] Anthelmintic,[30,31] Antiangiogenic,[32] Antibacterial,[11,30,33] Anticancer/Antineoplastic,[34–37] Antitumour,[30] Antihypertensive,[30] Antiviral,[30] Anti-inflammatory,[24,25,38] Antioxidant,[14,16,39–44] Antiprotozoal,[45] Antidiabetic,[26] Sedative/Anxiolytic,[25] Chemopreventive,[24] Insecticidal,[9] Wound healing[46] and Hepatoprotective.[47]

Dosage: For treatment of hypertension, osteodynia and lumbago, 10 to 20 g of rootbark is prescribed daily as a decoction or alcoholic maceration of torrefied material.[7]

Adverse Reactions: Sedation, nausea, vomiting, anorexia, hypersensitivity, hyperkalemia may occur.[48, 49]

Toxicity: The lyophilised aqueous extract of *M. citrifolia* roots did not exhibit any toxic effects but did show a significant, dose-related, central analgesic activity in the writhing and hotplate tests.[50] Chemical analysis and genotoxicity tests revealed that noni juice does not have a genotoxic potential and that genotoxic anthraquinones do not exist in noni juice.[51] However, according to some, anthraquinones are the most likely hepatotoxic components found in *M. citrifolia*.[52] Hepatitis induced by Noni juice had been reported.[53,54]

Contraindications: Should not be used during pregnancy and lactation. Contraindicated in people with hyperkalemia or with hypersensitivity to *M. citrifolia*.[48]

Drug-Herb Interactions: No information as yet.

51. *Nelumbo nucifera* Gaertn. (Nymphaeaceae)

Sacred Lotus, East Indian Lotus, Oriental Lotus

Nelumbo nucifera plants with flowers

Description: *Nelumbo nucifera* Gaertn. is an aquatic plant that grows in shallow waters. Leaves are green, round, 30–60 cm across and with long petiole. Flowers are pink, white or red, 10–30 cm and solitary. Fruits are non-edible and non-fleshy.[1]

Origin: Native to tropical and temperate Asia, Australia and Eastern Europe.[2]

Phytoconstituents: Nuciferin, nornuciferin, nelumboroside A & B, nelumstemine, dotriacontane, ricinoleic, roemerin, liensinine, neferine, lotusine, liriodenine, asimilobin, pronuciferine and others.[3–8]

Traditional Medicinal Uses: The leaves are used to treat sunstroke, diarrhoea, dysentery, fever, dizziness and vomiting of blood.[9] The plant is used as an antidote for mushroom poisoning[9] and for smallpox.[10] In Ayurveda,

b716_Ch_51-60.indd 105 1/8/2009 10:27:29 AM

the plant is used to treat cholera, diarrhoea, worm infestation, vomiting, exhaustion and intermittent fever.[3] The fruits are used in decoction for agitation, fever, heart and haematemesis while the stamens are used to "purify the heart, permeate the kidneys, strengthen virility, to blacken the hair, for haemoptysis and spermatorrhoea".[10] They are also used to treat premature ejaculation,[9] as astringent for bleeding,[3] excessive bleeding from the uterus,[9] abdominal cramps, bloody discharges, metrorrhagia, non-expulsion of the amniotic sac,[10] and as cooling agent during cholera.[11] The seeds are believed to promote virility, for leucorrhoea and gonorrhoea.[10] Powdered beans are used in treating digestive disorders, particularly diarrhoea.[3] They are also used as a tonic,[9] for enteritis, insomnia, metrorrhagia, neurasthenia, nightmare, spermatorrhoea, splenitis and seminal emissions.[10] The roots are for the treatment of diarrhoea, dysentery, dyspepsia, ringworm and other skin ailments and as a tonic as well.[10,11]

Pharmacological Activities: Antianxiety,[12] Antiarrhythmic,[13] Antibacterial,[4] Anticonvulsant,[14] Antidiarrhoeal,[15] Anti-inflammatory,[16] Hepatoprotective,[18] Antioxidant,[8,17–20] Antiplatelet,[21] Antiproliferative,[22] Antipyretic,[23,24] Antiviral,[25,26] Hypoglycaemic,[27,28] Hypolipidaemic,[29] Immunomodulatory[30] and Insecticidal.[4]

Dosage: A daily dose of 10 to 30 g of the ripe seeds as a decoction or powder is used for the treatment of neurasthenia, spermatorrhoea and metrorrhoea. The pericarps should be removed before using the seeds. Decoctions of 15 to 20 g and 2 to 4 g of dried leaves and seed cores respectively have been used for treating insomnia, haemorrhage and haematemesis. 6 to 12 g of plumules, 5 to 10 g of filaments, or 15 to 30 g of the receptacles in the form of a decoction, are used in the treatment of bloody stools, haematuria, uterine haemorrhage and haematemesis.[31] A daily dose of 5 to 8 g of rhizome or seeds taken by mouth is recommended.[32]

Adverse Reactions: Stamen, receptacle, rhizome node, leaf, and embryo can be safely consumed with proper usage.[33] No known side effects with appropriate therapeutic dosages.[3]

Toxicity: No information as yet.

Contraindications: Seed is contraindicated in constipation and stomach distention.[33]

Drug-Herb Interactions: No information as yet.

52. *Nephelium lappaceum* L. (Sapindaceae)

Rambutan, Hairy Lychee

Nephelium lappaceum fruits

Nephelium lappaceum tree

Description: *Nephelium lappaceum* L. is an evergreen tree that can grow up to 10 m tall. Leaves are pinnate, 25–40 cm long, compound with 2–4 pairs of leaflets. Blades are elliptic or obovate, 6–18 cm by 4–7.5 cm, thinly leathery, glabrous, with 7–9 pairs of lateral veins. Flowers are greenish white, fragrant and with no petals. The fruit is hairy, turning from green to red when mature and contains one hard seed.[1–3]

Origin: Native to Indonesia, Malaysia, Philippines and Thailand.[3]

Phytoconstituents: Type II and III cyanolipids, β-caryophyllene, monoterpene lactones 1 & 2, paullinic acid and others.[4–7]

Traditional Medicinal Uses: The fruits are used to treat diarrhoea and fever,[2] and also for the treatment of dysentery and dyspepsia.[8] Its seeds are narcotic[8] and the bark is used to treat disease of the tongue while the roots are used for fever.[2]

Pharmacological Activities: Antiviral,[9] Antifungal,[6] Antibacterial,[10] Antioxidant[10–13] and Cytotoxic.[12,13]

Dosage: No information as yet.

b716_Ch_51-60.indd 107 1/8/2009 10:27:37 AM

FA

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

b716_Ch_51-60.indd 108 1/8/2009 10:27:41 AM

53. *Nerium oleander* L. (Apocynaceae)

Oleander

Flowers of *Nerium oleander*

Nerium oleander shrub

Description: *Nerium oleander* L. is a small shrub up to 2 m high. Leaves are very narrowly elliptic, 5–21 by 1–3.5 cm, dark green, without stipules, leathery and arranged in whorls of three. Flowers are showy and fragrant. Sepals are narrowly triangular to narrowly ovate, 3–10 mm. Corolla is purplish red, pink, white, salmon, or yellow. Fruits consist of cylindrical follicles, 12–23 cm. Seeds are oblong, coma, about 0.9–1.2 cm.[1,2]

Origin: Native to southern Europe,[2] and widely cultivated and naturalised in Asia, Europe and North America.[1]

Phytoconstituents: Oleandroside, kaneroside, neriaside, nerigoside, neriumoside, neridiginoside, nerizoside, neritaloside, proceragenin, neridienone A, cardenolides N-1 to N-4 and others.[3–11]

Traditional Medicinal Uses: The plant is used in Ayurveda to treat scabies, eye disease and haemorrhoids.[3] It is used to treat parasitic infection in Calabria (Southern Italy).[12] Leaf decoction is used to treat diabetes in

b716_Ch_51-60.indd 109 1/8/2009 10:27:41 AM

southeastern Morocco.[13] Bark, leaf, flower are used medicinally as a cardiotonic and diuretic.[14]

Pharmacological Activities: Analgesic,[15] Anti-inflammatory,[8,9,15] Antibacterial,[16] Anticancer/Antineoplastic,[8,9,17–22] Antifungal,[23] Depressant,[24,25] Antimitotic,[26] Insecticidal,[27] Larvicidal,[28] Muscle stimulatory[29–30] and inhibits Nuclear factor-kappa B (NF-κB) activation.[31]

Dosage: No information as yet.

Adverse Reactions: Depression, dizziness, stupor, headache, nausea, vomiting, anorexia, abdominal cramps, spontaneous abortion, hypersensitivity, contact dermatitis, hyperkalemia and tachypnea.[32]

Toxicity: Toxic to humans[33–39] and animals.[40–48] The plant contains numerous toxic compounds, many of which can be deadly to people.[33–39] Ingestion can cause both gastrointestinal and cardiac effects and also affect the central nervous system.[33–39]

Contraindications: Should not be used during pregnancy and lactation, in children and in persons with hypersensitivity to oleander.[32] Should not be taken internally.[49]

Drug-Herb Interactions: Fatal digitalis toxicity can occur with concurrent usage of cardiac glycosides such as digoxin and digitoxin.[32] Concurrent use of quinidine, calcium salts, saluretics, laxatives or glucocorticids increases efficacy as well as side effects.[3]

54. *Ophiopogon japonicus* Ker-Gawl. (Liliaceae)

Dwarf Lilyturf, Mondo Grass, Mai Men Dong

Ophiopogon japonicus herb

Description: *Ophiopogon japonicus* Ker-Gawl. is an evergreen, stemless, rhizomatous herb. Leaves are sessile, long, 10–50 cm by 2–4 mm, linear, grass-like, 3–7 veined, and have pointed tips. Flowers are either solitary or paired.[1,2]

Origin: Native to Japan and Korea.[2]

Phytoconstituents: Ophiopogonin D and E, ophiopogonin C′ and D′, bornanol, ophiopogonanone A, C, E and F, ophiopojaponin D and others.[3–9]

Traditional Medicinal Uses: The roots are used to cool the body system, as a tonic, purgative, thirst quencher, treatment for sore throat, cough and fever.[10] In China, the roots are also used to treat bronchitis, cold, dysuria, haemoptysis, laryngitis, restlessness, thirst, tuberculosis, stress, as an aphrodisiac,

b716_Ch_51-60.indd 111 1/8/2009 10:27:47 AM

promoting fertility and memory, also as a sialogogue, to treat cancer and frequently included in polyherbal prescriptions for diabetes mellitus. The Indochinese uses the rhizomes to treat fever and inflammation, as a febrifuge, galactagogue, and also for intestinal, kidney and liver ailments.[3]

Pharmacological Activities: Antiarrhythmic,[11] Anti-inflammatory,[12,13] Antithrombotic,[14,15] Immunomodulatory,[16] Immunostimulatory,[17,18] Cardioprotective[19,20] and Chemoprotective.[21]

Dosage: The daily dose of tuberous roots of 6 to 20 g in the form of a decoction, pills or syrup consumed as an expectorant or antitussive.[22]

Adverse Reactions: No information as yet.

Toxicity: LD_{50} value of *O. japonicus* in mice was more than 2 g/kg intraperitoneally.[12]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

55. *Peltophorum pterocarpum* Backer ex K. Heyne (Leguminosae)

Jemerlang Laut, Yellow Flame, Yellow Flamboyant

Peltophorum pterocarpum tree

Peltophorum pterocarpum flowers

Fruits of *Peltophorum pterocarpum*

b716_Ch_51-60.indd 113 1/8/2009 10:27:50 AM

Description: *Peltophorum pterocarpum* Backer ex K. Heyne is a large tree with dome-shaped crown that can grow up to 24 m tall. The main rachis are 15–30 cm long and the pinnae are 6–20 paired with each pinna having 20–30 pairs of oblong leaflets. Flowers are yellow and in large bunches. The tree bears pods which are oblong, 5–10 cm long, flat and thin.[1–3]

Origin: Native to Malaysia, Ceylon, the Andamans and North Australia.[3]

Phytoconstituents: Rhamnetin, hirusitidin, bergenin and others.[4]

Traditional Medicinal Uses: The bark extract is used internally to cure dysentery and externally as a lotion to treat sprains, muscular aches, ulcers, and as an eye lotion, gargle and tooth powder.[2]

Pharmacological Activities: Antibacterial[5–7] and Antifungal.[7]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

56. *Persicaria hydropiper* L. (Polygonaceae)

Water Pepper, Laksa Plant

Persicaria hydropiper herbs

Persicaria hydropiper top view

Description: *Persicaria hydropiper* L. is an erect herb that can grow up to 30–50 cm tall. Leaves are oblong-lanceolate, short petioled, tapering to a pointed end, ciliate on the under surface. Flowers are white, in sparse, thin and with hanging false ears. Fruit is black, nut-like and has a flat and domed side.[1]

Origin: Native to Northern Africa, tropical and temperate Asia, Australia and Europe.[2]

Phytoconstituents: Rhamnazin, hydropiperoside, polygoidal, warburganal, isopolygodial, isodrimeninol, drimenol, confertifolin and others.[1,3–5]

Traditional Medicinal Uses: The liquid extract of the plant is used as a contraceptive and a haemostatic.[6] The plant is also used alone or with other herbs decocted for diarrhoea, dyspepsia, dysentery, enteritis, diuretic, expelling worms, heat stroke, itching skin, haemorrhage, jaundice and cancer as well.[7] In folk medicine, it is used internally for uterine bleeding, menstrual bleeding, bleeding of haemorrhoids, gastrointestinal bleeding, rheumatic pain, as a diuretic, for bladder and kidney disease, and gout. It is used externally for poorly healing wounds, sprains and contusions.[1] The leaves are

b716_Ch_51-60.indd 115 1/8/2009 10:27:53 AM

pounded and applied to skin diseases, for uterine disorders,[7] while its seeds are used as carminative, diuretic and stimulant.[7]

Pharmacological Activities: Analgesic,[8] Anthelmintic,[9] Antifertility,[10,11] Antifungal,[12–14] Antineoplastic,[15] Antioxidant,[12] Antimutagenic[16] and Insect repellent.[17,18]

Dosage: A tea prepared by pouring 0.25 L of hot water over 1 heaped teaspoon of the dried plant extract and strained after 10 min is to be drunk 3 times a day.[1] Homeopathically, it is used to treat varicose veins, 5 drops, 1 tablet or 10 globules are to be taken every 30 to 60 min in acute cases, or 1 to 3 times daily for chronical cases.[1]

Adverse Reactions: No known side effects with therapeutic dosages. Larger amounts can cause gastroenteritis. Skin irritation may occur if applied externally.[1]

Toxicity: Toxic to animals[19–21] and the LD_{50} for its chloroform leaf extract was 758.58 mg/kg in male albino mice.[22]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

57. *Phyllanthus amarus* Schum. & Thonn.
(Euphorbiaceae)

Pick-a-back, Carry Me Seed, Ye Xia Zhu

Phyllanthus amarus herbs

Description: *Phyllanthus amarus* Schum. & Thonn. is a small herbaceous annual plant that can grow up to 60 cm tall. Leaves are simple, alternate, green and stipulate. The blade is 3–8 mm by 2–4.5 mm and oblong-elliptic. The fruits are green, depressed globose in shape, 3-lobed and smooth. Both the flowers and fruits are borne under the branches.[1,2]

Origin: Native to Mexico and South America.[3]

b716_Ch_51-60.indd 117 1/8/2009 10:27:58 AM

Phytoconstituents: Phyllanthusin D, geraniin, corilagin, elaeocarpusin, amariin, amariinic acid, amarosterol-A and B, phyllantin, hypophyllantin and others.[4–11]

Traditional Medicinal Uses: The aerial part of the plant is used for various conditions. In Chinese medicine, the plant is made into a tea to cure kidney problems, venereal diseases, stones in the kidneys and bladder. The Malays use it to increase menstrual flow, reduce fever and cure colic. It is used by the Indians as a fish poison.[5] Indians also use the plant as liver tonic to treat liver ailments, ascites, jaundice, diarrhoea, dysentery, intermittent fever, conditions of the urogenital tract, eye disease, scabies, ulcers and wounds.[12] In Vietnam, it is used to induce sweating, and increase menstrual flow. It is also prescribed for toothache, muscle spasms and gonorrhoea. It is considered a diuretic, colic remedy and abortifacient in Southeast Asia.[5] It is also commonly used in Benin, Africa, as folk medicine against malaria.[13]

Pharmacological Activities: Analgesic,[14] Antibacterial,[15,16] Antidiarrhoeal,[17] Antifertility,[18] Antifungal,[19] Anti-inflammatory,[20–24] Antineoplastic,[25–29] Antioxidant,[30–32] Antiplasmodial,[13,33] Antiviral,[34–39] Diuretics,[40] Hepatoprotective,[41–43] Hypoglycaemic,[11,30,44] Inhibition of gastric lesion,[22] Antimutagenic,[27,45] Insecticidal[46] and Radioprotective.[47]

Dosage: A decoction may be prepared with 10 plants in 1 L of water.[12]

Adverse Reactions: No known side effects with therapeutic dosages.[12]

Toxicity: Non-toxic to mice at a dose of 100 mg/kg body weight[18] but at doses of 400 mg/kg, 800 mg/kg and 1000 mg/kg body weight in rats.[48]

Contraindications: No information as yet.

Drug-Herb Interactions: An alcoholic extract of *P. amarus* was found to inhibit cytochrome P450 enzymes both *in vivo* as well as *in vitro*.[48]

58. *Piper nigrum* L. (Piperaceae)

Pepper, Lada

Leaves and fruits of *Piper nigrum*

Piper nigrum herbs

Description: *Piper nigrum* L. is a perennial woody vine with aerial roots at the stem nodes. Leaves are shiny, broad, alternate and heart-shaped. Flowers are small and white. The fruits are globular berries, about 5 mm in diameter, green when unripe but turn bright red upon maturity and each bears a single seed.[1]

Origin: Native to Southeast Asia.[2]

Phytoconstituents: Piperine, sabinene, nigramides A–S, pipertipine, pipercitine; pellitorine, guineensine, piperettine, pipericine, dipiperamides A–C, pipnoohine, pipyahyine and others.[3–10]

Traditional Medicinal Uses: The plant is used in many Asian countries as a stimulant, for the treatment of colic, rheumatism, headache, diarrhoea, dysentery, cholera, menstrual pains, removing excessive gas and increasing the flow of urine.[11] It is also used in folk medicine for stomach disorders and digestive problems, neuralgia and scabies. In Ayurveda, it is used for arthritis, asthma, fever, cough, catarrh, dysentery, dyspepsia, flatulence, haemorrhoids, urethral discharge and skin damage. In Chinese medicine, it is used for

b716_Ch_51-60.indd 119 1/8/2009 10:28:00 AM

vomiting, diarrhoea and gastric symptoms. Homeopathically, it is used for irritation of mucous membrane and galactorrhoea.[3] A heavy dose of pepper with wild bamboo shoots is said to cause abortion.[4] In Assam, a method of birth control include *Cissampelos pareira* in combination with *Piper nigrum*, root of *Mimosa pudica* and *Hibiscus rosa-sinensis*.[12] The fruits are used to remove excessive gas in system, increase flow of urine, treat colic, rheumatism, headache, diarrhoea, dysentery, cholera, and menstrual pains.[13] White pepper is used for cholera, malaria, stomachache while black pepper is used for abdominal fullness, adenitis, cancer, cholera, cold, colic, diarrhoea, dysentery, dysmenorrhoea, dysuria, furuncles, headache, gravel, nausea, poisoning due to fish, mushrooms or shellfish.[4]

Pharmacological Activities: Antibacterial,[14–16] Antidiabetic,[17] Antifungal,[18] Anti-inflammatory,[19] Antineoplastic,[20] Antioxidant,[21–26] Gastroprotective,[27] Hepatoprotective,[28,29] Hypolipidaemic,[30,31] Antimutagenic,[32] Antithyroidal[33] and Insecticidal.[5,6,9,34–38]

Dosage: A single dose ranges from 300–600 mg of the berries is taken internally for stomach disorder.[3] Homeopathic recommendations for irritation of mucous membranes and galactorrhoea are 5–10 drops, 1 tablet or 5–10 globules 1–3 times daily.[3] For the treatment of haemorrhoids, 5–15 whole peppercorns are recommended to be taken.[39] For congestion, cold, head cold, chicken soup with black pepper can be taken.[40] The average daily dose of the berries is stated to be 1 to 3 g as a decoction, powder or pills, for the treatment of dyspepsia, vomiting, diarrhoea and colic resulting from cold.[41]

Adverse Reactions: No known side effects with appropriate therapeutic dosages.[3] Hypersensitivity may occur in certain people. The fruit can be safely consumed with proper usage.[42] However, when taken in large amounts in children, apnoea can occur.[43]

Toxicity: *P. nigrum* did not cause any adverse effects when tested with rats at doses 5 to 20 times normal human intake.[44]

Contraindications: Not to be taken during pregnancy and lactation or by children and people who are allergic to it.[43]

Drug-Herb Interactions: *P. nigrum* extracts inhibits the activity of cytochrome P450 enzymes.[8,45–49] Piperine enhanced the bioavailability of beta lactam antibiotics, amoxycillin trihydrate and cefotaxime sodium significantly in rats.[50] Piperine potentiated pentobarbitone sleeping time in a dose dependant manner, with peak effect at 30 min.[19] Piperine enhanced the

FA

bioavailability of phenytoin significantly, possibly by increasing the absorption.[51] Piperine from black peppers increased the AUC of phenytoin, propranolol and theophylline in healthy volunteers and plasma concentrations of rifampicin in patients with pulmonary tuberculosis.[52] Avoid concurrent use with drugs metabolised by cytochrome P450 enzymes.[43] Avoid concurrent use with anticoagulant agents as well.[53]

b716_Ch_51-60.indd 121 1/8/2009 10:28:03 AM

59. *Piper sarmentosum* Roxb. (Piperaceae)

Wild Pepper, Kadok, Sirih Tanah

Piper sarmentosum herb

Close up of *Piper sarmentosum* flowers and fruits

b716_Ch_51-60.indd 122 1/8/2009 10:28:03 AM

Description: *Piper sarmentosum* Roxb. is a climbing herb that can grow up to 10 m long with long runners that can develop into plantlets. Leaves are alternate, simple, 7–14 cm by 6–13 cm, heart-shaped and young leaves have a waxy surface. Flowers are bisexual or unisexual, in terminal or leaf opposite spikes. Fruit is small, dry, with several rounded bulges. Plant has a characteristic pungent odour.[1–5]

Origin: Native to Cambodia, India, Indonesia, Laos, Malaysia, Philippines and Vietnam.[5]

Phytoconstituents: Sarmentosine, sarmentine, (+)-sesamin, horsfieldin, brachystamide B, sarmentamide A, B, and C, (+)-asarinin, methyl piperate and others.[6–11]

Traditional Medicinal Uses: The whole plant can be used to treat fever and aids digestion.[3] The fruit is used as an expectorant[6] while the roots are used to treat toothache, fungal dermatitis on the feet, coughing, asthma and pleurisy.[6,12]

Pharmacological Activities: Antibacterial,[8,10,13] Antineoplastic,[6] Antiprotozoal,[10,14,15] Antipsychotic,[16] Hypoglycaemic,[17] Antifungal,[10] Insecticidal,[18,19] Antituberculosis,[11] Antimalarial[14] and Antioxidant.[20]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

60. *Plantago major* L. (Plantaginaceae)

Common Plantain, Whiteman's Foot, Daun Sejumbok

Plantago major herbs

Description: *Plantago major* L. is a small perennial herb. Leaves are nearly all basal, exstipulate, lanceolate to ovate, 5–20 cm long and rosette. Flowers are small, white, in dense spike-like inflorescence. Sepals are broadly elliptic, oblong to rounded obtuse or subacute and corolla are greenish or yellowish, with four lobed and imbricate. Seeds are dull black and endospermous.[1,2]

Origin: It is found in Europe, Northern and Central Asia, and introduced all over the world.[2]

Phytoconstituents: Aucubin, catalpol, scutellarein, nepetin, chlorogenic acid, neochlorogenic acid, hispidulin, homoplantaginin, nepitrin, ursolic acid and others.[3,4]

b716_Ch_51-60.indd 124 1/8/2009 10:28:07 AM

Traditional Medicinal Uses: The Greeks and Romans used it as an astringent, to heal wounds, asthma, fever and eye disorders.[5] In Brazil, it has been used to treat skin ulceration (cutaneous leishmaniasis) caused by *Leishmania braziliensis*.[6] *P. major* has been used in Turkey in the treatment of ulcers by taking the powdered dried leaves together with honey daily before breakfast.[7] Infusion of the leaf has been taken for diarrhoea, ulcers, bloody urine, digestive disorders, and excess mucous discharge. The American Indian groups make use of a poultice of the leaves for pain, swelling, cuts, wounds, sores, infections, blisters, insect bites, snakebites and haemorrhoids.[5] Its seeds are used to induce sweating, increase flow of urine, treat diarrhoea, dysentery, rheumatism, malaria, asthma, kidney problems, bladder diseases, gonorrhoea and piles.[3] Its roots are used to treat fever, respiratory infections and constipation.[5] The Commision E approved the internal use of plantain for catarrhs of the respiratory tract and inflammatory alterations of the oral and pharyngeal mucosa while its external application is approved for inflammatory reactions of the skin.[8]

Pharmacological Activities: Analgesic,[9] Antibacterial,[10,11] Anti-diarrhoeal,[12] Anti-inflammatory,[4,13–15] Anticancer,[10,16–18] Antioxidant,[19] Antiprotozoal,[20,21] Antiviral,[22] Immunomodulatory,[23,24] Immunostimulatory,[25,26] Proliferative,[10] Antiulcerogenic,[6] Antimutagenic,[27,28] Uterotonic[29] and Wound healing.[30]

Dosage: A daily dose of 8 to 16 g of the whole plant or seeds in the form of a decoction or extract is used to treat oedema, dysuria, haematuria, persistent cough, bronchitis and ophthalmia.[31] Approximately 2 to 4 ml of the fluid extract taken orally three times daily serves for general well being.[32] As a rinse or gargle, 1.4 g of cut herb is immersed in 150 ml of cold water for 1 to 2 hours.[8] For internal use, 1.4 g of herb is immersed in 150 ml of boiled water for 10–15 min, drunk as infusion, for 3 to 4 times daily.[8]

Adverse Reactions: Nausea, vomiting, diarrhoea, anorexia, bloating, hypersensitivity and dermatitis may arise. Life threatening anaphylaxis may occur in more serious cases.[32]

Toxicity: The 70% ethanol extract was found to be toxic to shrimps[33] but *P. major* possesses a low toxicity in rats after oral and intraperitoneal administration.[34]

Contraindications: Should not be used during pregnancy and lactation. Should not be used in persons with intestinal obstruction or those who developed hypersensitivity to plantain.[32]

Drug-Herb Interactions: Decreases the effects of carbamazepine and enhances the effects of cardiac glycosides, β-blockers and calcium channel blockers.[32]

61. *Punica granatum* L. (Punicaceae)

Pomegranate, Delima

Punica granatum fruit

Punica granatum tree

Description: *Punica granatum* L. is a small woody shrub that can grow up to 3–4.5 m tall. Leaves are opposite, oblong-lanceolate, 3–8 cm and branches are spiny. Flowers are large, trumpet-shaped, and bright orange-red in colour. Fruit is globose, with waxy surface, tough leathery skin, turns deep pink or red upon maturity, and contains numerous seeds with fleshy covering.[1]

Origin: Native to temperate and tropical Asia.[2]

Phytoconstituents: Punicafolin, punicalagin, friedelin, betulic acids, estrone, estradiol, piperidine, pomegranatate, pseudopelletierine and others.[3–7]

Traditional Medicinal Uses: The plant is used in folk medicine to expel worms, for diarrhoea, dysentery, as an abortifacient, astringent, for haemorrhoids (when used externally) and as a gargle for sore throat. In Ayurveda it is used for diarrhoea, dysentery, vomiting, and eye pain. Used in Chinese medicine for chronic diarrhoea and dysentery, it is also used to treat blood in

b716_Ch_61-70.indd 127 1/8/2009 11:42:34 AM

stools, worm infestation, and anal prolapse. Homeopathically, it is used for gastrointestinal disturbances.[5] The leaves are used to relieve itch. The pericarp is used to treat diarrhoea, rectocele and to expel pinworms.[8] The dried pericarp is decocted with other herbs for colic, colitis, diarrhoea, dysentery, leucorrhoea, menorrhagia, oxyuriasis, paralysis and rectocele,[3] to treat piles, yellowish discharge from the vagina, sore throat, bad breath and nose bleed.[8] The stem bark is used as an emmenagogue and for dysentery.[3]

Pharmacological Activities: Anthelmintic,[3,9,10] Antibacterial,[11–21] Anticancer/Antineoplastic,[22–36] Antidiabetic/Hypoglycaemic,[37–41] Antidiarrhoeal,[42] Antifertility,[43] Antifungal,[44,45] Anti-inflammatory,[46,47] Antimalarial,[48] Antioxidant,[34,48–62] Antiviral,[63–65] Gastroprotective,[66,67] Hepatoprotective,[68] Hypolipidaemic,[41,69] Immunomodulatory,[70,71] Neuroprotective,[72–75] Antiatherogenic,[76–78] Wound healing,[79,80] Larvicidal[81] and Molluscicidal.[82,83]

Dosage: Daily dose for the treatment of tapeworm is one part pericarp, root, or stem bark boiled with 5 parts of water. Bark juice extract is recommended for tapeworm at a single dose of 20 g. Otherwise, 250 parts powdered bark are boiled in 1500 parts water for 30 minutes.[5] Doses of 4–5 g of powdered dried flower are used in haematuria, haemorrhoids, dysentery, chronic diarrhoea, and bronchitis. Either 1.5–3 g of root and bark powder, 100–200 ml of bark decoction (for children, 28–56 ml) is used for anthelmintic purposes.[84] 7 g flower in 300 ml water is prescribed for inflamed mouth and throat; 5–12 g root or stem bark boiled in 240 cm^3 water until 1/3 has evaporated and taken on a 3 hourly basis on empty stomach 2 hours after taking 40 ml castor oil.[85] A daily dose of 20 to 50 g of dried root bark or stem bark as a decoction is used in the treatment of taeniasis (tapeworm infection). Daily dose of 15 to 20 g fruit rind in the form of a decoction is taken to treat dysentery and diarrhoea.[86]

Adverse Reactions: No known side effects with appropriate therapeutic dosages. In some cases, gastrointestinal disturbances may occur due to high tannin content.[5] Pomegranate fruits have rarely been reported to cause immediate hypersensitivity.[87]

Toxicity: The bark[5,9,88,89] and peel[90,91] are toxic.

Contraindications: Contraindicated in the elderly, children and pregnant women.[92] Pericarp is contraindicated in diarrhoea. Should not be taken with fats or oils when it is used for killing parasites.[88] Should not be used by patients with hepatic diseases or asthma, and who are hypersensitive to pomegranate.[91]

Drug-Herb Interactions: Inhibition of cytochrome P450 enzymes,[93,94] interaction with carbamazepine[95] and interaction with tolbutamide.[96]

62. *Rhodomyrtus tomentosa* (Ait.) Hassk
(Myrtaceae)

Rose Myrtle

Rhodomyrtus tomentosa plants

Flowers of *Rhodomyrtus tomentosa*

Description: *Rhodomyrtus tomentosa* (Ait.) Hassk is a bushy shrub with woolly young parts. Leaves are simple, oblong, 4–10 cm, opposites with three main longitudinal veins. Flowers are large, showy and pink, 3–4 cm wide, solitary and auxillary. Berries are oblong, 1–1.25 cm long, dark purple, fleshy and aromatic.[1–3]

Origin: Native to tropical and temperate Asia.[4]

Phytoconstituents: Rhodomyrtone, casuariin, castalagin, friedelin and others.[5–11]

Traditional Medicinal Uses: The leaves are used by the Chinese as a pain-killer. They are also used in Indonesia to heal wounds.[1] In Malaysia, the plant, including the roots, are decocted to treat diarrhoea and heartburn.[1,5] The buds and young leaves are used for treatment of colic, diarrhoea, dysentery, abscesses, furunculosis and haemorrhage. Concentrated decoction of leaves is used as a disinfectant for wounds, impetigo and abscesses.[12] The seeds are used as a digestive tonic and to treat snake bites.[1]

Pharmacological Activities: Antibacterial.[6]

b716_Ch_61-70.indd 129 1/8/2009 11:42:35 AM

Dosage: Daily dose of 10 to 30 g of fresh buds or young leaves in the form of an extracted juice or dried for use as a powder or in a decoction is used for the treatment of colic, diarrhoea, dysentery, abscesses, furunculosis and haemorrhage.[12]

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

63. *Rhoeo spathacea* (Sw.) Stearn (Commelinaceae)

Purple-leaved Spider Wort, Moses-in-the-Cradle, Oyster Plant

Rhoeo spathacea plants

Flowers of *Rhoeo spathacea*

Description: *Rhoeo spathacea* (Sw.) Stearn is a short-stemmed herb with underground rhizome that can grow up to 90 cm tall. Leaves pointed upwards, fleshy, hard, long, narrow, 15–25 cm by 3–5 cm, pointed, dark green on the adaxial surface and purple on abaxial surface. Flowers are small, white, and concealed within a boat-shaped envelope of two bracts.[1,2]

Origin: Native to Central America and West Indies, widely cultivated as ornamental plants in most tropical countries.[2]

Phytoconstituents: No information as yet.

Traditional Medicinal Uses: They are used to treat intestinal bleeding, blood in the sputum and dysentery.[3]

Pharmacological Activities: Anti-inflammatory,[4] Antiadrenergic[5] and Uterostimulatory.[6]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

b716_Ch_61-70.indd 131 1/8/2009 11:42:37 AM

64. *Ricinus communis* L. (Euphorbiaceae)

Castor Oil Plant, Castor Bean

Ricinus communis leaf

Fruits of *Ricinus communis*

Ricinus communis plant

Description: *Ricinus communis* L. is an erect herb, growing up to 3.6 m high, having pinkish succulent stem and large alternate palmate leaves that are green or reddish brown. Leaves are lobed, consisting of 6–8 radiating leaflets with serrated edges and prominent central veins. Flowers are green, pink or red and inconspicuous, with no petals. The fruits are capsular, with three lobes, prickly and green, containing three seeds.[1]

Origin: Native to Africa, naturalised throughout tropics and subtropics.[2]

b716_Ch_61-70.indd 132 1/8/2009 11:42:41 AM

Phytoconstituents: Ricin, ricinoleic acid, ricinine, p-coumaric acid, ferulic acid, o-coumaric acids, syringic acid, cinnamic acids, stigmasterol, fucosterol and others.[3–8]

Traditional Medicinal Uses: Its leaf poultice is applied to boils and sores in India; to treat headaches and fever in Hawaii.[9] The leaves and roots are used in a decoction for anal prolapse, arthritis, constipation, facial palsy, lymph-adenopathy, strabismus, uteral prolapse, cough, and also as a discutient and expectorant. The heated leaves are applied to gout and swellings as well.[5] The leaves and oil are used for dermatological purposes in Nigeria.[10] Its seeds are used to treat abscesses and skin eruptions, deafness, headache, skin problems, bleeding, constipation, boils, piles and to promote labour.[4] They are rubbed on the temple for headache, powdered for abscesses, boils, and carbuncles. The plant is also used for dogbite, scrofula and several skin infections. The Chinese rub the oil on the body for skin ailments. The seeds are crushed and made into a pulp and rubbed into the palms for palsy, introduced into the urethra in stricture and rubbed on the soles of feet of parturient woman to hasten birth or expulsion of the placenta. The seeds are also used to treat colic, diarrhoea, dysentery, enteritis, acute constipation, for itching, ringworms, warts, dandruff, hair loss and haemorrhoids. It is also used as a laxative before X-ray examination of bowels.[3,5] Midwives sometimes use castor oil to induce labour.[11]

Pharmacological Activities: Antifertility,[12–18] Antioxidant,[19] Antipsychotic,[20] Antiviral,[21] Anti-inflammatory,[19] Convulsant,[20] Hepatoprotective,[22] Filaricidal,[23] Haemaglutination[6] and Insecticidal.[24]

Dosage: For internal use in the treatment of constipation or to expel worms, at least five 2 g capsules or ten 1 g capsules are recommended.[3] Externally, a paste made of ground seeds may be applied to affected skin areas 2 times daily, up to 15 days.[3] Homeopathically, 5 drops, 1 tablet or 10 globules are recommended every 30–60 min for acute treatment or 1–3 times daily for chronic cases.[3] Standard adult dose is reported to be 5.0–20.0 ml/dose, not exceeding 60.0 ml per day. Standard child dose is reported to be 4.0–12.0 ml/dose.[11] The seed oil is recommended for use as a laxative in a dose of 2 to 5 ml and as a purgative in a dose of 20 to 30 ml. 15 seeds may be crushed and applied to the plantar surface of the feet for the induction of labour at term and placenta delivery in cases of placenta retention.[25] The woman's feet are advised to be thoroughly washed after

delivery of foetus and placenta.[25] For constipation, it is reported that 15 to 60 ml of castor oil may be taken orally a day.[26]

Adverse Reactions: Gastric irritation, nausea, vomiting, colic, and severe diarrhoea may occur. Long term use may result in loss of electrolytes, especially potassium, and causes intestinal motility inhibition.[3]

Toxicity: Seeds are highly poisonous due to the toxic lectins, principally the albumin and ricin.[27] The seeds are toxic to both animals and humans.[3,12,26–35]

Contraindications: Not to be taken by children under 12 years of age, patients with acute inflammatory intestinal diseases, inflammatory bowel disease and appendicitis.[3] contraindicated in pregnancy, intestinal obstruction and abdominal pain.[3,11] Not to be given with potentially toxic oil-soluble anthelmintics.[36]

Drug-Herb Interactions: Effect of cardioactive steroids may be increased.[3] To prevent decreased absorption of castor, it is advised not to take within one hour of antacids, other drugs and milk.[37]

65. *Ruta graveolens* L. (Rutaceae)

Herb of Grace, Common Rue

Ruta graveolens herb

Ruta graveolens leaves

Description: *Ruta graveolens* L. is a glabrous herb with stem that can grow up to 14–45 cm. Lower leaves are more or less long-petiolate with ultimate segments 2–9 mm wide, lanceolate to narrowly oblong. Inflorescence is rather lax; pedicels are as long as or longer than the capsule; bracts are lanceolate, leaf-like. Sepals are lanceolate and acute. Petals are oblong-ovate, denticulate and undulate. Capsule is glabrous; segments somewhat narrowed above to an obtuse apex.[1]

Origin: Native to Europe.[2]

Phytoconstituents: Rutoside, rutaverine, arborinine, rutin, elemol, pregeijerene, geijerene, furocoumarins, bergapten, xanthotoxin, fagarine, graveolinine and others.[3–5]

Traditional Medicinal Uses: It is frequently used to treat worm and parasitic infection.[6] It has been commonly used for the treatment of psoriasis and vitiligo due to the psoralens and methoxypsoralens present.[7] It is also used to relieve muscle spasms, as carminative, emmenagogue, haemostat, uteronic, vermifuge, to treat hepatitis, dyspepsia, diarrhoea, bug bite, cancer, cold, fever, snakebite, earache, toothache and as an antidote especially in malarial poisoning.[3,4,8] It is also used as an abortifacient to terminate pregnancy.[9] The plant has been used for pain relief in Mexico.[10]

b716_Ch_61-70.indd 135 1/8/2009 11:42:44 AM

Pharmacological Activities: Analgesic,[11] Antibacterial,[12] Anticancer,[13–17] Antifertility,[18–20] Anti-inflammatory,[5,21] Antimicrobial,[22–24] Antiprotozoal[25] and Cytotoxic.[24]

Dosage: For the treatment of earache, the oil is poured on cotton and inserted into the affected ear. For oral administration, 1 capsule or 0.5 to 1 teaspoon of extract is taken three times daily with meals. The topical cream may be applied to the affected area when necessary.[26] For delayed menstruation, 2 cups of infusion per day is used.[8]

Adverse Reactions: Hypotension, hypersensitivity, rash, erythema and blisters may occur when applied topically.[26] Therapeutic dosages could bring about melancholic moods, sleep disorders, tiredness, dizziness and spasms.[27] Misuse as an abortive during pregnancy can lead to vomiting, epigastric pain, kidney damage, depression, sleep disorders, feelings of vertigo, delirium, fainting, tremor, spasm and sometimes may end up with fatal outcome.[8]

Toxicity: Toxic to humans[27–33] and animals.[34]

Contraindications: Should not be used during pregnancy.[9] Should not be used during lactation, in children and in person with hypersensitivity to rue. Patients with heart diseases should use it with caution.[26] Contraindicated in patients with poor renal function.[9]

Drug-Herb Interactions: Concurrent use of cardiac glycosides (e.g. digoxin, digitoxin) with rue may cause increased inotropic effects. Concurrent use of antihypertensives with rue may cause increased vasodilation.[26]

66. *Saccharum officinarum* L. (Gramineae)

Sugarcane, Tebu

Saccharum officinarum plants

Saccharum officinarum stems

Description: *Saccharum officinarum* L. is a perennial plant that forms tall clumps that can grow up to 6 m tall. Stems are greenish, yellowish or dark purplish and juicy. Leaf blades are broadly linear, glabrous and 80–150 cm by 4–6 cm. Inflorescence is a large silky panicle.[1,2]

Origin: Native to Southeast Asia and Pacific Islands, widely cultivated elsewhere.[2]

Phytoconstituents: Octacosanol, policosanol, orientin, tricin-7-*O*-glycoside, palmitic acid, oleic acid and linolenic acid and others.[3–9]

Traditional Medicinal Uses: Consumption of beets (*Beta vulgaris*) combined with molasses from *S. officinarum*, is used by Dominican healers to shrink fibroids or to "strengthen and fortify the uterus after the fibroid had been drained from the body".[10] The cane juice promotes expulsion of phlegm from the respiratory passages, stimulates gastric activities, treats wounds, ulcers and boils.[11]

b716_Ch_61-70.indd 137 1/8/2009 11:42:49 AM

Pharmacological Activities: Analgesic,[3] Anticancer,[6] Anti-inflammatory,[3,9,12] Antiosteoporotic,[5] Antioxidant,[6,13,14] Antiplatelet,[15–17] Antithrombotic,[18] Hypocholesterolaemic,[19–24] Immunomodulatory,[25,26] Immunostimulatory,[27] Antiatherogenic[28] and Myocardial protective.[4,17]

Dosage: Short-term studies have shown the efficacy and tolerability of policosanol at 10 mg/day on hypercholesterolaemia in obese patients with Type 2 diabetes.[19,20]

Adverse Reactions: Pollen extract of *S. officinarum* showed strong sensitising potential which induces allergy.[29]

Toxicity: Studies on the toxicity of higher aliphatic primary acids (D003) and higher aliphatic primary alcohols (policosanol) isolated from *S. officinarum* did not show any toxic effects.[30–33]

Contraindications: No information as yet.

Drug-Herb Interactions: Pretreatment with high doses of policosanol significantly increased propranolol-induced hypotensive effects, while the effects of nifedipine remained unchanged. Policosanol does not antagonise the hypotensive effect of β-blockers but it can increase the hypotensive effect of beta-blockers without modifying cardiac frequency.[34] Pretreatment with single doses of policosanol significantly increased the nitroprusside-induced hypotensive effect.[35] Warfarin alone and the combination of policosanol and warfarin induced a moderate, but significant prolongation of the bleeding time.[36]

[**Authors' Note:** Freshly squeezed sugarcane juice is commonly consumed and sugar is commercially obtained from sugarcane.]

67. *Sauropus androgynus* (L.) Merr.
(Euphorbiaceae)

Sweet Leaf Bush, Cekup Manis, Daun Katuk

Leaves of *Sauropus androgynus*

Sauropus androgynus plant

Description: *Sauropus androgynus* (L.) Merr. is a small shrub that can grow up to 3 m tall. Leaf blade is ovate-lanceolate, oblong-lanceolate, or lanceolate, 3–10 cm by 1.5–3.5 cm and thinly papery. Flowers are small and borne in axillary clusters. Mature fruit is about 5 cm wide. Young branches and leaves are used as vegetable.[1,2]

Origin: Native to China and tropical Asia.[3]

Phytoconstituents: Sauroposide, (–)-isolariciresinol, corchoionoside C and others.[4–10]

Traditional Medicinal Uses: The fresh leaves or roots possess uterotonic activity and are used for the treatment of retained placenta. A mouth wash

b716_Ch_61-70.indd 139 1/8/2009 11:42:54 AM

made from the juice of the fresh leaves and honey, and applied to the tongue and gums cures thrush of the tongue in infants. The leaves are used for erythema, measles and dysuria. Its roots serve as a diuretic and relieve congestion.[4] The root decoction is used for fever and urinary bladder complaints.[11]

Pharmacological Activities: Anthelmintic.[12]

Dosage: For the treatment of retained placenta, a dose of 40 g, in the form of an extracted juice is administered in 2 subdoses at 10 minutes interval.[4]

Adverse Reactions: Excessive consumption of the leaf reportedly caused dizziness, drowsiness and constipation.[13]

Toxicity: Reported to cause bronchiolitis obliterans,[14–16] ischaemic bronchi necrosis,[17] irreversible obstructive ventilatory defect,[18,19] breathing difficulties[20,21] and is cytotoxic.[22]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

68. *Sesbania grandiflora* Pers. (Leguminosae)

Scarlet Wisteria Tree, Red Wisteria, Daun Turi

Sesbania grandiflora leaves

Fruit of *Sesbania grandiflora*

Sesbania grandiflora tree

Description: *Sesbania grandiflora* Pers. is a tree that can grow to 8–10 m in height. The compound leaves are about 30 cm long with 12 to 20 pairs of rounded, narrow, oblong leaflets, 3–4 cm by 1 cm. Flowers are 5–10 cm by 3 cm, in pale pink, red, purple or white. The pods are 25–50 cm, slender, and cylindrical with many light brown to red brown seeds.[1]

Origin: Native to Malesia and cultivated in the tropics.[2]

Phytoconstituents: Grandiflorol, (+)-leucocyanidin, oleanolic acid, lutein, beta-carotene, violaxanthin, neoxanthin, zeaxanthin and others.[3–9]

b716_Ch_61-70.indd 141 1/8/2009 11:42:57 AM

Traditional Medicinal Uses: In the Philippines, the plant is used for its hypotensive properties.[10] It is used in Indian folk medicine for the treatment of liver disorders.[11] The juice of the leaves and flowers are popularly used for nasal catarrh and headache when taken as snuff. Various leaf preparations are used to treat epileptic fits. Applied externally for treatment of leprous eruptions. A poultice of the leaves is used for bruises. The leaf juice is mixed with honey for congenital bronchitis or cold in babies.[12]

Pharmacological Activities: Antibacterial,[13] Anticonvulsant,[14] Anti-inflammatory,[15] Anxiolytic,[14] Depressant,[15] Diuretic,[15] Hepatoprotective,[11] Hypoglycaemic,[16] Hypotensive[15] and Haemolytic.[17,18]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

69. *Solanum nigrum* L. (Solanaceae)

Black Nightshade, Terong Meranti, Poison Berry

Fruits of *Solanum nigrum*

Solanum nigrum flower

Solanum nigrum shrub

Description: *Solanum nigrum* L. is a small herb, up to 1.5 m tall. Leaves are ovate, ovate-oblong, glabrous, hairy, 1–16 cm by 0.25–12 cm. Inflorescence of 2–10 in an extra-axillary cluster, with white or purple corolla and yellow central protrusion. Fruit is globose, black in colour but is green when immature, 0.5 cm in diameter, with many seeds.[1]

Origin: Native to Southwest Asia, Europe, India and Japan.[2]

Phytoconstituents: Solanidine, α-,β-,γ-chaconine, desgalactotigonin, α-,β-solamargine, diosgenin, solanadiol, α-,β-,γ-solanines, soladulcidine, solanocapsine, α-,β-solansodamine, solasodine, α-solasonine, tigogenin, tomatidenol, uttronins A and B, uttrosides A and B, solanigroside A–H and others.[3–9]

b716_Ch_61-70.indd 143 1/8/2009 11:43:04 AM

Traditional Medicinal Uses: The stem, leaves and roots are used as a decoction for wounds, tumours and cancerous growths, sores and as an astringent.[3,5] They are also used as a condiment, stimulant, tonic, for treatment of piles, dysentery, abdominal pain, inflammation of bladder, relief of asthma, bronchitis, coughs, eye ailments, itch, psoriasis, skin diseases, eczema, ulcer, relief of cramps, rheumatism, neuralgia and expulsion of excess fluids. The roots are used as an expectorant.[5] The plant has yielded medicines for sore throats, coughs and digestive problems. It has also been used as an agricultural insecticide.[10] Europeans in Africa used the plant to treat convulsions. It is used by the Africans for treating headache, ulcers and as a sedative. The whole plant is used for the treatment of dermatitis, inflammation, heavy female discharge, diarrhoea and dysentery.[11] It is also used as a diuretic and febrifuge. Whole plant is decocted for abscesses, cancer of the cervix, inflammation, leucorrhoea and open sores. Young shoots are consumed as virility tonic for men and to treat dysmenorrhoea in females.[3] In Indochina, the leaves are used as purgative and high blood pressure lowering agents while the fruits are used as laxatives.[12]

Pharmacological Activities: Antibacterial,[13] Anticancer/antineoplastic,[8,14–33] Antiulcerogenic,[36,37] Antinociceptive,[15] Anti-inflammatory,[15] Antioxidant,[34,35] Antiviral,[38] Depressant,[39] Hepatoprotective,[40–43] Hypolipidaemic,[44] Antimutagenic,[45] Enzyme modulation,[46] Larvicidal,[47] Molluscicidal[47–49] and Parasiticidal.[50]

Dosage: 10 drops of extract is taken internally 2 to 3 times a day or 5 to 10 g of tincture may be taken daily for gastric irritation, cramps and whooping cough. For external use as a rinse of moist compress, it is boiled in 1 L water for 10 minutes before usage for psoriasis, haemorrhoids, abscesses, eczema and bruising.[4]

Adverse Reactions: No known side effects with appropriate therapeutic dosage.[4]

Toxicity: Harmful to rats.[51] It is toxic to cattle as it can cause acute nitrate toxicity which leads to death in cattle. In chronic cases, decreased milk yield, abortion, impaired vitamin A and iodine nutrition can occur. The proposed LD_{50} for nitrate toxicity is 160–224 mg/NO_3/kg for cattle.[52] Overdoses can lead to headache, queasiness and vomiting, due to high alkaloid content. Mydriasis may also occur, although rare.[4,52] Solanine, in doses of 200–400 mg, may cause gastroenterosis, tachycardia, dyspnea, vertigo, sleepiness, lethargy, twitching of the extremities and cramps. It is also teratogenic.[52]

Contraindications: No information as yet.

Drug-herb Interactions: No information as yet.

70. *Swietenia macrophylla* King (Meliaceae)

Honduras Mahogany, Broad-leaved Mahogany

A fruiting *Swietenia macrophylla* tree

Seeds and fruits of *Swietenia macrophylla*

Description: *Swietenia macrophylla* King is an evergreen tree, up to 30–35 m tall. Bark is grey and smooth when young, turning dark brown, ridged and flaky when old. Leaves are up to 35–50 cm long, alternate, glabrous with 4–6 pairs of leaflets. Each leaflet is 9–18 cm long. Flowers are small and white; and the fruit is dehiscent, usually 5-lobed capsule, erect, 12–15 cm long, grayish brown, smooth or minutely verrucose. The seed is woody, glossy and possesses wing-like structure at the base that aids its dispersion by wind.[1]

Origin: Native to South America, cultivated in the Asia-Pacific and the Pacific for its quality wood.[1]

Phytoconstituents: Swietenine, swietenolide, andirobin, khayasin T, swietemahonins E–G, swietenine acetate, swietenolide tiglate and others.[2–7]

Traditional Medicinal Uses: The seeds of *Swietenia macrophylla* are traditionally used in several indigenous systems of medicine for the treatment of various ailments such as hypertension, diabetes and malaria.[9] The local folks of Malaysia believe that the seeds are capable of "curing" hypertension and diabetes. The seeds are usually consumed raw by chewing.[8] A decoction of seeds of *Swietenia macrophylla* is reported to treat malaria in Indonesia.[9] Among the Amazonian Bolivian ethnic groups, the seeds are traditionally

b716_Ch_61-70.indd 145 1/8/2009 11:43:08 AM

used to induce abortion by drinking a decoction of the seeds and to heal wounds and various ailments of the skin via external application of the mashed seeds.[10]

Pharmacological Activities: Antimalarial,[11] Antihypertensive[8] and Antidiarrhoeal.[12]

Dosage: No information as yet.

Adverse Reactions: No information as yet.

Toxicity: Uterine haemorrhage.[13]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

71. *Terminalia catappa* L. (Combretaceae)

Indian Almond, Katapang

Fruits of *Terminalia catappa*

Terminalia catappa tree

Description: *Terminalia catappa* L. is a tall tree, up to 25 m tall. Branches are horizontally whorled, giving it a pagoda shape. Leaves are shiny, obovate, 10–25 cm long, tapering to a short thick petiole. Leaves are yellow that turn red before shedding. Flowers are small and white. Fruits have smooth outer coat, 3–6 cm long, flattened edges, with a pointed end. Pericarp is fibrous and fleshy.[1–3]

Origin: Native to tropical and temperate Asia, Australasia, the Pacific and Madagascar.[4]

Phytoconstituents: Catappanin A, chebulagic acid, 1-desgalloylleugeniin, geraniin, granatin B, punicalagin, punicalin, tercatain, terflavins A & B, tergallagin, euginic acid and others.[2,5–13]

Traditional Medicinal Uses: *Terminalia catappa* has been used to treat dysentery in a number of Southeast Asian countries. In Indonesia, the leaves are used as a dressing for swollen rheumatic joints while in the Philippines, they are used to expel worms.[2] In Karkar Island, New Guinea, juice from

b716_Ch_71-75.indd 147 1/8/2009 11:44:15 AM

the squeezed leaves is applied to sores and the sap from the white stem pith is squeezed and drunk to relieve cough. In Nasingalatu, Papua New Guinea, the flower is crushed, mixed with water and drunk to induce sterility. In New Britain, the old yellow leaves are crushed in water and drunk to sooth sore throat. In Bougainville, the leaves are heated and placed on pimples and the bark is applied to sores. In Tonga, the juice from pounded leaves and bark is applied to mouth sore. In Irian Jaya, the leaves are applied to wounds and burns while in Somoa, it is used to cure cough and sore throat.[14] The fruits are used after childbirth to strengthen the back. An enema made from the crushed fruit mixed with *Trigonella foenum-graecum*, animal fat and warm water is administered to the new mother after childbirth.[15] The leaves are used for the treatment of scabies and skin diseases while the juice is used to treat headache and colic.[13,14] The bark is used as a diuretic, cardiotonic and for dysentery.[16] The leaves of this plant have also been used as a folk medicine for treating hepatitis in India and Philippines.[5]

Pharmacological Activities: Antimicrobial,[17–19] Anticancer/Antineoplastic,[20,21] Anti-inflammatory,[5,22] Hypoglycaemic,[8] Anti-oxidant,[6,10,11,23,24] Hepatoprotective,[7,11,25–29] Antiviral,[30] Chemopreventive,[31] Aphrodisiac[32] and Antimutagenic.[33]

Dosage: 2 tablespoons of a decoction (few leaves in 200 ml water) is given every 2 hours to stop diarrhoea.[34]

Adverse Reactions: No information as yet.

Toxicity: Preliminary oral LD_{50} doses of petroleum ether, methanol and aqueous extracts of *T. catappa* in mice were found to be 343, 195, and 210 mg/kg respectively.[8] Rats fed on *T. catappa* diet maintained their body weight but suffered from stomach, small intestine and pancreas hypertrophies as well as spleen atrophy.[35] Larger doses enhanced liver damage.[36]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

72. *Thevetia peruviana* (Pers.) K. Schum.
(Apocynaceae)

Yellow Oleander, Trumpet Flower

Thevetia peruviana flower and fruit

Thevetia peruviana shrub

Description: *Thevetia peruviana* (Pers.) K. Schum. is a shrub, up to 6 m tall. All parts contain highly poisonous milky latex. Leaves are simple, few, exstipulate and spirally arranged. Blade is linear, 7–13 cm by 0.5–1 cm and glossy. Flowers are large, yellow, 5 cm across, gathered in few flowered terminal cymes. Fruits are green, shiny, globose, 4–5.5 cm across with 4 or less poisonous seeds.[1]

Origin: Native to Central and South America.[2]

Phytoconstituents: Thevetins A and B, thevetosides, acetylperuvoside, epiperuviol, perusitin, theveneriin, thevebioside, thevefolin, pervianoside I–III and others.[3–6]

Traditional Medicinal Uses: Used as an abortifacient, to treat congestive heart failure, malaria, leprosy, indigestion, ringworm, venereal disease and

b716_Ch_71-75.indd 149 1/8/2009 11:44:23 AM

even as a suicide instrument.[7] Used in India as an astringent to the bowel, useful in urethral discharge, worms, skin diseases, wounds, piles, eye problems and itch. Used in continental Europe and is considered particularly useful in mild myocardial insufficiency and digitalis intolerance.[5] Its bark is used as an emetic, febrifuge, insecticidal, poison and for reviving patients with heart failure.[4]

Pharmacological Activities: Antiarrhythmic,[5] Antifungal,[8] Hepatotoxicity and Nephrotoxicity,[5] Larvicidal, [9] Molluscicidal,[10] and Cardiotonic.[5]

Dosage: No information as yet.

Adverse Reactions: Vomiting, dizziness, and cardiac dysrhythmias such as conduction block affecting the sinus and AV nodes may occur.[11]

Toxicity: Toxic to humans,[3,11–16] fishes[17–19] and animals.[5,20] Ingestion of half to fifteen seeds may cause a varying degree of vomiting, palpitation, arrhythmia including sinus bradycardia and could also lead to death.[21] The kernels of about 10 fruits may be fatal to adults while the kernel of one fruit may be fatal to children.[22]

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

[**Authors' Note:** The purified glucoside thevetin, extracted from the seeds, is prescribed as a cardiotonic drug in a 0.1% solution orally, in a dose of 1 to 2 ml daily, or in 1 mg/2 ml ampoules parenterally. 1 to 2 ampoules have been given daily for its cardiotonic property.[23] A dose of 2.4 mg followed by average maintenance dose of 600 μg has been shown to be effective in 22 patients with congestive heart failure.[5] The herb has been administered as a thevetin tablet of 0.25 mg/tablet. A maintenance dose consists of 0.25 mg/day. For injection administration, an injection ampoule is prepared from 0.25 mg/ml diluted in a 5% glucose solution.[24] However, note that thevetin is not conventionally used in clinical practice and its dosage is not found in pharmacopoeial monographs.]

73. *Tinospora crispa* (L.) Diels (Menispermaceae)

Akar Putarwali, Batang Wali

Tinospora crispa shrub

Flowers of *Tinospora crispa*

Description: *Tinospora crispa* (L.) Diels is a woody climber with numerous protrusions on the stem. Leaves are oblong-ovate, cordate, 8–9 cm by 7–8 cm and tapering to a pointed end. Flowers are small, with 6 petals, 2 mm in length and 8–27 cm racemes. Male flowers have yellow sepals whereas female flowers have green sepals. Drupelets are red, juicy and 7–8 mm long.[1,2]

Origin: Native to Malesia, Indochina, Indian subcontinent and China.[3]

Phytoconstituents: Boropetol B, borapetoside B, C & F, jatrorhizine, magnoflorine, palmatine, protoberberine, tembolarine, diosmetin, cycloeucalenol, cycloeucalenone and others.[4–8]

Traditional Medicinal Uses: It is used for hypertension, diabetes mellitus,[9] to treat malaria, remedy for diarrhoea and as vermifuge.[6] In Malaysia, *T. crispa* extract is taken orally by Type 2 (non-insulin-dependent) diabetic patients to treat hyperglycaemia.[10]

Pharmacological Activities: Anti-inflammatory,[11] Antioxidant,[7] Antimalarial,[12,13] Antiprotozoal[14] and Hypoglycaemic.[10,15,16]

b716_Ch_71-75.indd 151 1/8/2009 11:44:28 AM

Dosage: No information as yet.

Adverse Reactions: The plant may result in an increased risk of hepatic dysfunction due to marked elevation of liver enzymes but is reversible upon discontinuation of *T. crispa*.[17]

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: *T. crispa* extract showed 61.3% increase of cytochrome P450 3A4 (CYP3A4) enzyme inhibition after 20 min of preincubation with human liver microsomes.[18]

[**Authors' Note:** In a randomised double blind placebo controlled trial[17] to determine the efficacy of *T. crispa* as an additional treatment in patients with Type 2 diabetes mellitus who did not respond to oral hypoglycaemic drugs and refused insulin injection, 20 patients received *T. crispa* powder in capsule form at a dose of 1 gram three times daily for six months. Twenty patients received a placebo. The results showed no evidence to support the use of *T. crispa* for additional therapy in such patients. Two patients had elevated levels of liver enzymes, which was reversible on discontinuing *T. crispa*. Patients receiving *T. crispa* had significant weight reduction and cholesterol elevation. Note that the study only included patients that did not respond to oral hypoglycaemic therapy.]

74. *Vitex rotundifolia* L. f. (Verbenaceae)

Round Leaf Chastetree, Beach Vitex

Vitex rotundifolia leaves

Vitex rotundifolia tree

Description: *Vitex rotundifolia* L. f. is an evergreen woody tree, densely covered with short hairs. Leaves are opposite, simple, ovate, broadly oblong-elliptic, 2–5 cm long by 1.5–3 cm wide, rounded or abruptly acute at the base. Inflorescence panicles are at the terminal, densely flowered, 4–7 cm long with purple corolla. Fruits are globose, 5–7 mm.[1]

Origin: Native to Temperate and Tropical Asia, Australasia and Pacific.[2]

Phytoconstituents: Rotundifuran, prerotundifuran, vitexilactone, previtexilactone, vitexicarpin, vitricine, vitetrifolins D–G, vitexifolins A–E, isoambreinolide and others.[1,3–9]

Traditional Medicinal Uses: In Malaysia, various parts of the plants are considered panacea for illnesses ranging from headache to tuberculosis.[10] In China, the plant has been used for the treatment of cancer.[11] A poultice of the leaves is used to treat rheumatism, contusions, swollen testicles and as a discutient in sprains. In Indonesia, leaves have been used in medicinal baths, as

b716_Ch_71-75.indd 153 1/8/2009 11:44:32 AM

a tincture or for intestinal complaints.[10] In Papua New Guinea, sap from crushed heated leaves is diluted with water and drunk to relieve headaches. The fruits are used to expel worms and in Vietnam, a decoction of dried fruits has been used to treat colds, headache, watery eyes and mastitis.[10] In Thailand, fruits have been used for asthmatic cough and haemorrhoids.[10] Infusion of the boiled roots is regarded as diaphoretic and diuretic, drunk for fever, after child-birth and for liver diseases.[10]

Pharmacological Activities: Analgesic,[12] Antibacterial,[13,14] Antifungal,[15] Anti-inflammatory,[16] Antineoplastic,[11,15,17–20] Antinociceptive,[8] Antioxidant,[6] Antiprotozoal,[21] Hypotensive,[22] Immunomodulatory,[17] Antimutagenic[23] and Insect repellant.[15]

Dosage: No information as yet.

Adverse Reactions: *V. rotundifolia* may trigger various allergic reactions such as sneezing, respiratory problems, dizziness, headache and nausea.[10]

Toxicity: No information as yet.

Contraindications: No information as yet.

Drug-Herb Interactions: No information as yet.

75. *Zingiber officinale* Roscoe (Zingiberaceae)

Common Ginger

Zingiber officinale rhizome

Zingiber officinale plant

Description: *Zingiber officinale* Roscoe is an herbaceous plant that grows up to 1.2 m high and with an underground rhizome. The stem grows above ground and leaves are narrow, long, lanceolate, with distinct venation pattern and pointed apex. Flowers are white or yellowish-green, streaked with purple and fragrant.[1]

Origin: Originate from tropical Asia, widely cultivated in the tropics.[2]

Phytoconstituents: Gingerol, zingiberene, farnesene, camphene, neral, nerol, 1,8-cineole, geranial, geraniol, geranyl acetate and others.[3–7]

Traditional Medicinal Uses: Ginger is the folk remedy for anaemia, nephritis, tuberculosis, and antidote to *Arisaema* and *Pinellia.*[8] Sialogogue when chewed, causes sneezing when inhaled and rubefacient when applied externally. Antidotal to mushroom poisoning, ginger peel is used for opacity of the cornea. The juice is used as a digestive stimulant and local application in ecchymoses.[8] Underground stem is used to treat stomach upset, nausea, vomiting, nose bleeds, rheumatism, coughs, blood in stools, to improve digestion, expel intestinal gas, and stimulate appetite.[9] The rhizomes are

b716_Ch_71-75.indd 155 1/8/2009 11:44:38 AM

used to treat bleeding, chest congestion, cholera, cold, diarrhoea, dropsy, dysmenorrhoea, nausea, stomachache, and also for baldness, cancer, rheumatism, snakebite and toothache.[8] It is also used as postpartum protective medicine, treatment for dysentery, treatment for congestion of the liver, complaints with the urino-genital system/female reproduction system and sinus.[10] Besides that, it is used to alleviate nausea, as a carminative, circulatory stimulant and to treat inflammation and bacterial infection.[3] The Commision E approved the internal use of ginger for dyspepsia and prevention of motion sickness.[11] The British Herbal Compendium indicates ginger for atonic dyspepsia, colic, vomiting of pregnancy, anorexia, bronchitis and rheumatic complaints.[12] European Scientific Cooperative on Phytotherapy (ESCOP) indicates its use for prophylaxis of the nausea and vomiting of motion sickness and to alleviate nausea after minor surgical procedures.[13]

Pharmacological Activities: Analgesic,[14–16] Anthelmintic,[17] Antiarthritic,[18] Anticancer,[19–33] Antidiabetic,[34,35] Antidiarrhoeal,[36] Antiemetic,[37–53] Antihyperlipidaemic,[54] Antihypertensive,[55] Anti-inflammatory,[14,15,56–66] Antimicrobial,[67–78] Antioxidant,[6,7,79–82] Antiplatelet,[83–87] Antispasmodic,[88] Antiulcer,[89,90] Antiviral,[91–93] Anxiolytic,[94] Hepatoprotective,[93–97] Hypocholesterolaemic,[57,98,99] Hypoglycaemic,[14,100,101] Hypolipidaemic,[102] Hypotensive,[103] Immunomodulatory,[104–106] Neuroprotective,[107] Insect repellent[69,108] and Radioprotective.[109–111]

Dosage: A tea is prepared by pouring boiling water over 0.5 to 1 g of the coarsely powdered ginger for 5 min and passing through a tea strainer, taken to prevent vomitting.[112]

Adverse Reactions: Fresh rhizome can be safely consumed with proper usage.[113] Contact dermatitis of the fingertips has been reported in sensitive patients.[114]

Toxicity: It is nontoxic when tested in rats[115] but overdose may cause cardiac arrhythmia and CNS depression.[116]

Contraindications: Consult physician before using ginger preparations in patients with blood coagulation disorders, taking anticoagulant drugs or with gallstones. Avoid dried rhizomes during pregnancy.[113]

Drug-Herb Interactions: Interacts with anticoagulants such as heparin, warfarin, drugs used in chemotherapy and ticlopidine. Ginger taken prior to 8-MOP (treatment for patients undergoing photopheresis) may substantially reduce nausea caused by 8-MOP.[117] Ginger appears to increase the risk of

bleeding in patients taking warfarin.[118] However, ginger at recommended doses does not significantly affect clotting status, the pharmacokinetics or pharmacodynamics of warfarin in healthy subjects.[119] Ginger also significantly decreased the oral bioavailability of cyclosporine.[120]

[**Authors' Note:** Ginger is widely eaten as a food ingredient and used in many different cultures as traditional medicine. Ongoing scientific research has shown diverse pharmacological activities.]

Navigation Down Under

For
Northern-Hemisphere Tourists
Walkers and Scouts
Boating and Outdoor Orientation

1001 Ways to Extend

Your Navigational Know-How

2010

CONTENTS

Section page

1 BIG-PICTURE TOP-DOWN————————3
Global Perspective
Sun
Stars
Earth
Moon
Wind
Clockwise Navigation
Anticlockwise Systems
Revision, Overview, Explanations

2 THE INSIDE STORY—Mental Manoeuvres—34
Learn
Think Interpret
Look
Integrate
Orientation
Disorientation
Reorientation
Lost?
Miscellaneous: Angle; Night-time

3 REALLY INTERESTING STUFF
—Bush Navigation——————————56
Nature's Norths
Memory & Naming Skills
Lie of the Land
Tracks and Bends

4 DISTANCE——————————————90
Progress
Pace
Range
MAP and/or COMPASS
Map Use
Magnetic Compass Use
Mapmaking

5 Navigational Mnemonic————————110

6 Appendices——————————————116
Tide and Moon Dial
Star Charts; Zodiac
Morse Code & Rhythm Mnemonics
Semaphore Code & Mnemonics
Symbols for Track-Notes & Mud-maps
A List of Star Coordinates
Great Circles; Calendars in the Sky
Tropical Countries

7 Contents/Index——————————128

Big-Picture Rule One: Locate a Distant Reference Direction First

Until you Orient yourself to the bigger picture, by finding an unchanging direction to latch on to,

•Small-scale pictures and large-scale mental layouts cannot come together. •And the local features won't fit into the bigger perspective.

•Your mental photographs will float around in chaos, like so many loose jigsaw pieces, or letters in an alphabet soup.

In choosing a Reference Direction, your first choice should be...

•ASTRONOMICAL — Sun, Moon, stars, planets, satellites—e.g. Locate the direction of the North Pole Star. There is an astronomical overview p32f

Then •GLOBAL — Compass or Botanical or Geomorphological North e.g. find Canada's North Magnetic Pole.

Next •CONTINENTAL — Weather Systems, radio transmitters, powerline networks, highways, sea-coast—e.g. notice the Sydney to Hobart flight path.

Then •DISTANT landforms — Lie, Lay, Lines, Hills; Lights, etc.—e.g. the direction of the main river valley draining to the sea.

Last of all comes: •What you can see in front of your nose! I.e. don't let your mind latch on to some unreliable or meaningless reference like how the car is parked, or where the wind is, or the initial direction of the track/the drive in/the main street, or uphill, or the way you got out of bed…

•If you can't see a thing beyond a few metres, and have no compass, you will need to "keep a straight line" (see p. 87b) by lining up markers.

Now that you have a reference direction, you can implement the other "Rule One"s...

•Match what you see in front of your nose to the reference;

•Predict your overall trend before you start; •Notice your start-off direction as you start.

Visualisation is your Key Skill

This is most especially relevant to orienting your nose, in your mind's eye, to the Bigger or Invisible Pictures, but is needed at all scales.

Imagine	where the Southern Cross is during the day;	or	where the cars are, now; how you got here.
Visualise	the Earth's invisible spin-axis	and	the weather map you saw, yesterday
Mentally picture	the Galactic spiral! — which clock-sense is it?	or	the turnoff you need — will it be uphill?
Remember	the tropical countries under the Sun	and	the way back
Describe	the path of a shadow-tip in mid-Winter	or	what makes this place recognisable
Picture	sunrise in Antarctica tomorrow morning	or	old blazes, in your mind, so you don't miss any
Interpret numbers, labels and words into reality	e.g. what the shape of the night-sky has to do with your global position		
Express visually	what each clue means	e.g.	what "high noon" means to you right now
Mentally photograph	the map as you study it beforehand	and	how the climb opened out onto the plateau
Mentally match	the invisible landscape trends to the reference direction and to what you see in front of your nose		

At night, and in the rain, navigation gets to be 'all in the mind'.

By recreating the reality you can't see, you can navigate what is in front of your toes, by intelligent imagination.

Big-Picture Orientation is a Mental Layout Acquired Beforehand

Because the invisible picture is so large, you can do your global and continental orientation at home before you start.
Travelling all day to go walking somewhere else, will hardly alter it.

You use: a globe; an atlas; a roadmap of the continent — to learn the lay beforehand. These are layout maps... The relative positions shown are fixed.
But layout views — as seen from above — will hardly ever be seen in your reality, first-hand.
(In contrast, when you come to putting yourself in the middle of the big-picture, you see the landscape first-hand, and you will need "Nose Navigation" — as seen from the ground — which is all about the apparent relative directions.)
The global constants won't change as you walk, e.g. the behaviour of the Sun and stars, because everything is so distant.

You need extra care to visualise the Big Picture, and its meaning, because it is (a) Round (global) (b) Layout (i.e. mental only) (c) Invisible (d) Big!
Do this before you attempt to nose your way through the local scenery.

Smaller-scale clues can then 'fit-in', sensibly, to the larger context.

Compass Navigation Robs You of a Global Perspective

Flat maps must distort the true picture. Only a globe is true. Mercator-projection maps (diagram 1) distort land areas and longitude lines, in order to "preserve the straightness" (!) of a compass course, which never was straight!

If you glue your nose to a compass needle, to follow a compass-course, the course which looks straight on a flat Mercator map will actually take you on a spiral, into either the South or North Pole (as when you keep any landmark at a diagonal).

The sense of the spiral (turning to your right/left) changes as you cross the Equator...

Because you do need a big-picture approach, **you need a global perspective**.

(Here's a good place to set out the basics...

^
Width or latitude on the map's side
V

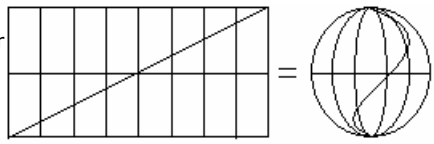

A 3D spiraling helix

Remember: "Latitude **equates** to the distance from the Equator" — the short dimension.

"Longitude turns upon the time — how **long** has the globe turned since midnight") < long for longitude >

Here's a test: Imagine the midsummer sunset from well Down Under, looking even further South. Do you think of a tropical country in that direction?

The following section — about Sun Moon Earth and Stars — aims to restore a true perspective to people who can hardly think except with a flat diagram, and can hardly visualise anything without divorcing it from the real landscape in front of their nose. (There is always a tropical country under the Sun.)

Think of this 'global orientation' section as a global **Re**-Orientation and don't forget, that... **Global means Round!**

The Big Picture is Invisible — Your Job is to Visualise It!

As a matter of principle, we should work top-down
i.e. from a framework down to the detail, from general to specific, large to small, overall to local, unchanging to changeable.

As a matter of fact, we are lazy. The point of doing "useless" grand-scale orientation exercises is to:

make global orientation second nature — easier next time around;
to make you forget flat pictures of a round Earth and to imagine it properly;
to give your struggling memory some comprehensive and relable framework to work within;
to integrate what you can't see with what you can, e.g. to link different walks and areas;
to orient to, and to understand, the Sun's movement, and to orient to the Moon and stars and planets as a bonus;
to make extra clues and techniques available, e.g. to open up astro-navigation techniques;
to work "top-down" — to avoid jumbling up your mind.

The idea is to read the following double-pages horizontally, as one topic, i.e. Sun-Sun; Stars-Stars; Clockwise-Anticlockwise.
To follow one topic through, flip over the pages and look in the corresponding positions.

Global

Global Re-orientation — Light-Half/Dark-Half

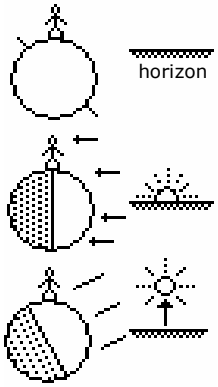

horizon Step 1. **Always Position Yourself on Top of the World.** ..whether in a diagram or mentally or in reality.
This makes your reorientation realistic, doesn't it?
Look at your feet and imagine you are standing on a huge basketball in the sunlight.
To visualise the light-half/dark-half of the globe, carry a real ball with you, preferably a globe of the world.
Imagine that everything which you see, this side of the horizon, is the tiny top-fraction of a huge ball.

Step 2. Face the Sun, but imagine it low down, setting, with horizontal rays coming to you.
Notice the circle which divides day from night, passing to left and to right underneath your feet.
"He divided light from darkness by a circle drawn on the face of the sea." Job 26:10

Step 3. 'Raise' the Sun, and its sunbeams, to their actual height.
And as you do, let the dark-half fall back behind you by the same amount
i.e. degree for degree around a round globe—one degree of fall-back for one degree of Sun-rising.

After this you can do your best to answer: (Help will follow soon) What countries are seeing the Sun as horizontal?
i.e. Which countries does the 'circle' pass through at the moment? Which ones are seeing the Sun rise?
Which ones see it setting? Which parts see a 24-hour sunset, with the Sun skimming the horizon? (Use a globe)

Sun

Your Best Reference Direction: Imagine the Sun at High Noon — Always

(This won't work so well in the tropics)

Night or day, your mental reference direction is North or South by the Sun. So imagine the Sun North of you (Southern Hemisphere) rain or shine.
It is the natural interpretation… of looking at your watch, and
of looking at the Sun, when it is not North of you. (Warning: Don't look directly at the Sun! It will blind you!)
of looking at the compass. (You don't want to locate 'Canada's magnetic pole', after all!)
of judging botanical North/South (p.74a);
of using a map in conjunction with the landscape,
and even when using the Moon

Imagine the *high-noon* Sun in a real landscape — this one here!
— at all times of day or night.
— at the correct height, on its circular path,
— which is tilted at its rise-set angle.
Warning:
You will get disoriented
when you cross the Equator!

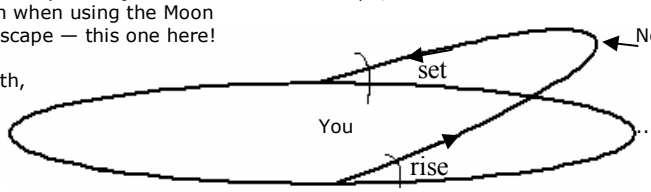

Stars

The 4 Bright Outer Stars of Orion Point out Where the Pot is Going

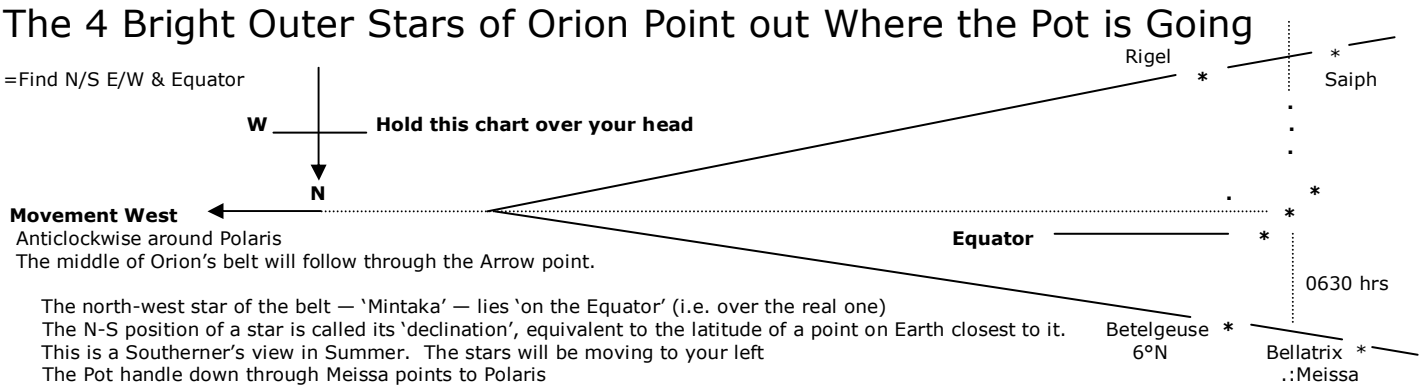

The north-west star of the belt — 'Mintaka' — lies 'on the Equator' (i.e. over the real one)
The N-S position of a star is called its 'declination', equivalent to the latitude of a point on Earth closest to it.
This is a Southerner's view in Summer. The stars will be moving to your left
The Pot handle down through Meissa points to Polaris

Moon

Look for the Line Joining Sun, Moon, Planets & Zodiac Stars—the Ecliptic

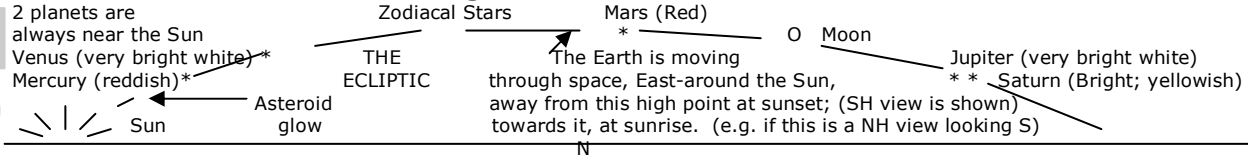

The solar system — Sun and planets — is flat, and the Zodiac marks that plane, by design. It's **the 'edge-on' view, which we see, from within**.
The Moon orbits within 5° of that plane, and so is never far off line. **The line is called the Ecliptic** — a great circle.
(It looks as straight as the horizon, unlike the diagram. See 'Optical illusion' in the Moon box, p.8d)
During the day, you can orient, not by the Sun only or the Moon only, but by the line joining them. Sometimes Venus may be visible too.
You have to imagine the rotation of that line (more on this soon). It does not move in the same direction in which it points, but its path is up to 23°½ tilted from that.

Clockwise

Mnemonic: "Clock-wise Northern Hemisphere **SWiNES**"

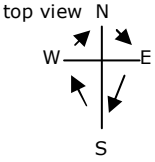

Europeans make chronometers to follow their Sun.
They see, looking South (diagram top right)… rise in the East ; South at noon ; set in the West.
That's 'clockwise' (seen from above; East to South to West) for them
The swines! They force us, down under, to place our clocks on the southern wall and think behind us!
(We see, looking North: rise East, to North at noon, to West i.e. SH; anti-clockwise)……

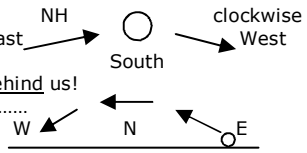

clockwise top view

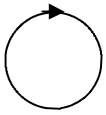

Anyway, changing your direction "Clockwise", as seen from above, is turning to your right, and, **i**gnoring the **i**, the mnemonic is:
South, to **W**est, to **N**orth, to **E**ast, to **S**outh or "clockwise: **SWiNES"**

A mnemonic for global orientation: **"True Polarity"**

"**True**" courses refer to the Earth's **Poles** — i.e. the axis of rotation of the Earth — **the single most directional thing on our spherical Earth**.
"True" bearings measure the angle between the course and the longitude line — a North/South line which connects the poles, and which passes through your location.

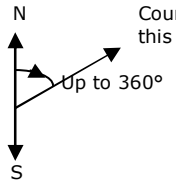

The angle is measured **clockwise from 'North'**
e.g. 065°T
"Oh Six Five degrees True (-polarity)"!

Remind yourself, whenever you use "True", by saying "True Polarity"!

Mnemonic: "**N**oon Sun Passes **N**orth of My **N**ose Down Under"— It Rhymes in ' **n** ' (dow-nun-der)
(Now, Keeping your **n**ose to **N**orth, say…) (The second parts work for both hemispheres provided you face North, not noon)
"Sun **Ri**ses on My **Ri**ght (to the right of North)" —It Rhymes in ' **ri** '
"Sun S**e**ts on My L**e**ft (in the W**e**st)" —It Rhymes in ' **e** '
"That's Anticlockwise around the North Pole Star, if you are facing North from the SH; [Clockwise around the South Pole Star", if facing S from NH]
"The **SH** ho**ri**zon **shi**fts to the **ri**ght"
"The **NH** ho**ri**zon **N**ever **H**eads **ri**ght, it moves left"

For Southerners, looking at the noon Sun i.e. from South of the tropics

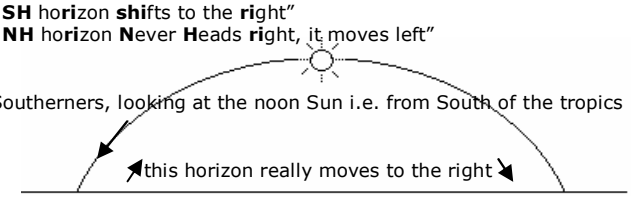

West/Set/Left **North**/Nose/**Noon**/**Down Under** East/Rise/Right

Paper Diagrams aren't Enough.
Imagination isn't enough.
Go outside and take a look.
(You will need to shield your eyes from the Sun, e.g. with your hand)
Let the reality sink in.

The Two Brightest Stars Point out the **South** Pole Star

Canopus and Sirius line up almost South/North.
Use the star next to Sirius ('Mirzam') and the line-up is perfect.
The 'Dog Star' Sirius is the brightest star in the sky.
The Big Dog seems to have a front leg and back leg and a tail...

Sirius 17 degrees S. * . Mirzam

A GOOD SOUTHERN SUMMER SIGN

A**l**udra, in the tai**l** η *
Wezen, where he wees from δ * * α
Ad**h**ara, the **h**ind leg ε * *β Mirzam
This is the nearest bright star to Sirius

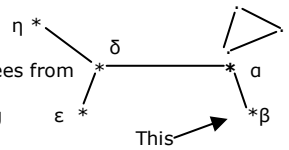

The Big Dog

0620 RA
('Right Ascension' or star-time)

* Canopus 53 degrees S.
(Go another 37 degrees to South polestar)

Mnemonic: "Wind? Weather Map!"

For navigating by the wind, don't just leave it at the level of "The wind is from over there".

* Look at the skyline "from Bruce Valley"
* Look at the Sunshine "from 30° to the left of the Sun at the moment" (Don't look at the Sun!)
* Look at the Sun-path "a 'sunset' breeze" i.e. from sunset (Simply imagine its rise-set path.)
* Imagine the continent beyond the skyline "from the West Coast"
* Meld that reality to the weather map you remembered.

Don't stop interpreting the wind until you can fit it into a Big Invisible Picture connected to that skyline over there.

Mnemonic: Down Under, The Sun Moves **ANTI**-BigB**ENW**i**SE** **i.e. anti-clockwise**

Look at the 'ENWiSE' in the mnemonic.
Anticlockwise is **E**ast to **N**orth to **W**est to **S**outh to **E**ast.
Ignore the **i**. Look down from above onto the horizontal directions involved…
Anticlockwise is turning to your left. " Anti-clockwise is ENWISE "

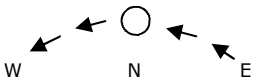

(If you put a clockface on the north wall and look at the Sun's movement, (Don't look at the Sun!),
you will see that the Southern Hemisphere's second-hands tick backwards! i.e. against the Sun's movement.
So put it on the south wall facing North, and the clock hands will move 'with' the Sun.)

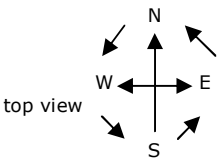

In the tropics, the Sun will pass somewhere near overhead. To make it seem "anti-clockwise" anyway, face North.

Don't get mixed up and think of Big Ben wise as clockwise. Down here our Sun moves anti- , and that is ENWiSE.

Earth Sun Stars Wind Anticlockwise

Global

Mnemonic: "Time Tells Turning"

What's the use of looking at your watch and spouting numbers! At least, glance at the Sun, or rather at its shadows (don't look directly at the Sun) and then, imagine the Sun's direction at high noon as a reference direction.

That's what time tells you! — the Sun's position relative to its noon direction, North/South, called the "meridian".

A.m. says "ante meridiem" i.e. "before noon", and p.m. says "post meri-diem" i.e. "after mid-day"

Midday = 12 noon (but only in an ideal world, because noon is only approximately at 12)

One pm = one hour past 12md, and so on.

24-hour time tells you how far around the Earth has turned, since the Sun was opposite you i.e. since midnight.

(And then don't forget that the local significance of the time is: progress — distance made good — "watch" your pace)

Mnemonic:

"Watch" the Sun (without looking at it) whenever you watch your watch
(and "watch" your steps too, for progress)

Sun

Notice these, Yesterday — to Guess the Time Today

- Sunrise time (and direction) ..If you know the time, you can guess at the directions.
- Sunset time (and direction) ..These first 2 are symmetric around North.
- The horizontal angle of separation between sunrise and set
- Midday time (and direction, and height) ..This is halfway between. Memorise how short your shadow is.
- When did it get light enough? To see. To walk.
- When did it get too dark? To cook. To walk.
- How long is the usable daylight? ..The day-length, between dawn and dusk.
- How many hours of dark?
- When did the Sun reach due East and due West? (How high?) ..This won't happen in winter. They are symmetric around midday.
- Times for the Sun at magnetic East, noon, West? ..These don't depend on guessing or calculating the magnetic variation.

Update yourself every week.

Stars

The Spinning Earth Makes **Circular Paths** for the Sun, Moon, Etc.

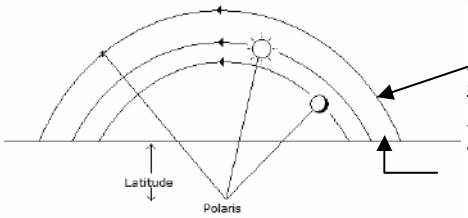

Sun, Moon, stars, planets — they all move only slightly in relation to each other during 1 day/night — but they all zoom around at about 15° per hour from East to West.

Although I can draw this on flat paper as a circular path, the Equatorial path actually looks as straight as the horizon, in reality! (Southern Hemisphere view)

You can now guess at time or direction by your familiarity with the circular 'orbits' e.g...
The highest point of the path will be North-South of you, if you can guesstimate it
Latitude determines the rise/set angle of the Equatorial path.
Rising and setting places are symmetrical around North.

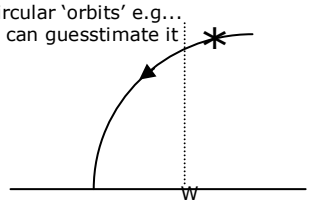

Although individual rising and setting is symmetrical about North,
whole constellations that rise vertically in mid-latitude may set horizontally!
and vice versa e.g. The Big Dog or Scorpio rise flat
but dive vertically into the western horizon.

Moon

The Moon; Rule 4: Look for It

You have a 1 in 4 chance of finding it during the day. Provided you remember to look!

Rules 1 and 2 (like so many other navigational clues) are...

- PREDICT when and where it should be visible
- GUESS where to look And Rule 3 goes with the guess:
- LEARN from that mistake.

Keep looking for it. It's easy to miss. E.g. through the tree canopy; or rising through the horizon haze. Polaroid glasses help; you need to twist them

You will then have clues to… North, East, and West
Time of night (i.e. later)
Ocean tides
Tomorrow's Moon — Phase; rise, set, direction, height
Upper-level winds, if there are clouds about.

Hint: Measure the Sun-Moon separation with a piece of string (Don't look directly or indirectly at the Sun!) so you know exactly where to look later.

Clockwise

"VEERING" with the Sun is "Clockwise" only in the Northern Hemisphere
"BACKING", against the Sun's movement, is "Anticlockwise"
— but only in the Northern Hemisphere.

(The terms refer to horizontal directions changing either with or against the Sun's movement)

"Veering and Backing" are Eurocentric images, so they can be very confusing Down Under. Don't use those words in the Southern Hemisphere!

Say "Clockwise" and "Anticlockwise" instead

..for both hemispheres, as a matter of habit. Although these meanings too are Eurocentric, they are not confusing!

(the terms refer to a bird's-eye view looking down from above as if onto a clock-face)

Meteorologists and mariners still use the terms 'veering' and 'backing' and they do mean 'clockwise' and 'anticlockwise'

But the Sun-based image is backwards, south of the tropics.

The sky turns **15° per hour** — Use your **thumb and forefinger**

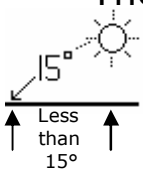

The Earth rotates once a day. Yes? That's 360° in 24 hours, yes?
Well that works out at 15° per hour — the sky spins 15°/hour to the West.
That calls for **recognising 15°** E.g. (diagrams left, and right) "one hour of sunlight left" (SH)

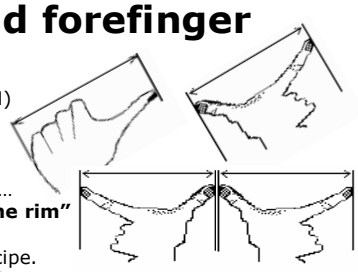

For Accuracy:
(0) Measure each handspan along the Sun's 'circular' path, not horizontally.
(1) Use constant stretch — of arms and fingers — 'just strained' — for consistency.
(2) Use constant chest-arm geometry. It is best to use two hands squarely in front of you…
(3)**Test for 15°, horizontally,** against the full 360° horizon — **"24 spans go around the rim"**
Span off, 24 times, from point-to-point right around the horizon.
Adjust your hand configuration until you find one that fits the "24 Hours" recipe.
(4) Try gripping one hand with the other, or touching one, at a specified point, with the other
, or, (5) Look along your shoulder, sideways, at one hand.

My recipe for accuracy fits my arms:"4 spread fingers, wrist bent back fully (at arm's length in front=60cm or 1cm/degree")
What do your arms require for 24 spans = 360°?

Earth

Mnemonic: The Southern Summer Sun Sets Somewhat South

Surprising perhaps. (If you equate North with 'hot' you get caught on this.) (for Northern Hemisphere substitute 'North' for 'South')
When the Sun comes South for summer its sunrise and sunset places also shift South with it.
Memory hexagon for Mid-Winter and Mid-Summer Down-Under at 37° S…
(top view) N. (the Sun moves anti-clockwise through North)

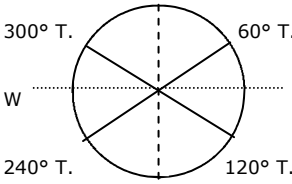

… In Mid-Winter, Down-Under, at 37°S., sunrise is at 60°T. and sunset is at 300°T. … A 9½ hour day.
In Winter the Sun will never show East or West of you.

At Equinoxes — Mid-Autumn, and Mid-Spring — rise and set are East and West… a 12 hour day.

… In Mid-Summer, Down-Under, at 37°S., sunrise is at 120°T., and sunset is at 240°T. … A 14½ hour day.
The Sun will be East and West at **mid**-morning and **mid**-afternoon respectively, and **mid**-high — about 45°.
Your main aim is to get familiar with the Sun's circular path.

Sun

Stars Leave Star-Trails ..if you photograph them with time-exposure...

Star-trails are bits of circles, because… ..stars circle around the North and South Pole-Stars...
i.e. the Earth rotates on an N-S axis, which points to the Pole-stars. The pole star stays fixed in position all night (if you do).
Southern Stars Circle Clockwise, once per night-and-day, around the South Pole Star.
Northern Stars Circle Anticlockwise, around the North pole star, Polaris. These clock senses are the same in both hemispheres.
But **Note**: low Southern stars will move anti-clockwise around you, (i.e. to your left around the horizon, anti-clockwise as seen from above you). (Clockwise — to your right in the NH) (More on this later)

It helps you with global orientation to use 'star compasses'.

One of the clearest compasses is to blur the star circles into full circles then take vertical tangents to those circles.
In the following diagram, the left-hand direction is "30°" to the left of the Earth's axis (or to 60° maximum latitude).
The right-hand one is "20°" to the right (or to 80° maximum latitude). They are global.
If you follow these star directions, they will take you right around the earth in a 'straight' line — a big circle.
You will pass through places equal in latitude to the star which is overhead (p.11c), so you can already *see* the maximum latitude that such a 'great circle course' would take you to; then it takes you back towards the equator.
HINT: You won't get far in star-navigation, without a detailed star list and star chart. See p. 123 for a list.

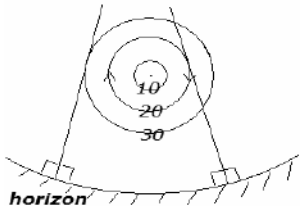

Stars

Mnemonic: "Wind in Your Face"

The Southern Hemisphere version of Buys Ballot's Law… [Try to use unobstructed wind i.e. not in a valley nor in a mountain eddy]

- "**Face** the Wind [N.H. rule uses 'Back to the wind, Sun-twist…]
- Sun-twist yourself 15°-30° [i.e. anti-clockwise in S. H.]
  — to counteract landscape friction on the surface winds.
  [When the winds are impeded, they slow down, and so stop circling around the pressure centres, and start to cross the isobars, toward the low pressure (a 'sink' hole) at an angle of 15°-30°, less at sea and over flat land; more over rough terrain.]
- Rhyme: **"Low on My Left** &/or **High on My Righ-t"**

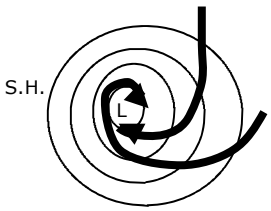

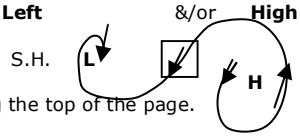

Top view of you, in the square, facing the top of the page.

Wind

Southerners Need **Anti**clockwise Systems

Our clocks seem to run backwards, compared to the Sun.
Our compass-bearings are measured 'backwards'. (see p.10e)
'Veering' and 'Backing' seem to mean the wrong clock-sense.
When the Sun moves forwards, all the other systems seem to push backwards.
We lose the global perspective.
And this section is all about aiming to put that right.
Consequently, we need a Southerner's viewpoint…

I hope you find that the anticlockwise systems which will follow soon are much more workable than Eurocentric systems.
Some of the other systems given on this right-hand side of the page, are universal — they can be used in either hemisphere.

Anticlockwise

Global

Global Orientation: Latitude and Polar Distance

Step 1. Stand up on the globe again (in your mind and in reality).
Step 2. Put your left shoulder to the Equator i.e. Face East (sunrise).
(If you are in the northern hemisphere: face west, sunset.)
Step 3. Mnemonic: EQUATOR on My Left equals My LATITUDE is on My LEFT,
RISEN POLE on My RIGHT equals My POLAR DISTANCE is on My RIGHT.

Backviews...
Facing E.

Equator | Latitude on left | P.D. on your right

Since the South Pole is 90° latitude, it is further South from the Equator than you are, and will be on your Right, as you face East.
Step 4. Measure the degrees of distance around the Earth's circumference as an angle at the centre of the Earth (by slicing the globe in two).
The two angles mentioned must add up to 90° i.e. they are "complementary" angles (spelt with an 'e').
If you are, say, 45° south of the Equator, then the closest pole must be 45° further south. The 'distance' to the Pole is 45°.
If you are 30° south, then the South Pole will be 60° away from you. Your '**colatitude'**—or *90° minus latitude*—or "Polar Distance"—is 60°.
The goal is to imagine those two angles — Latitude and Polar Distance — at the Globe's centre beneath your feet, and around the circumference.

Sun

Mnemonic: "Sun at High Noon — How High?"

When you navigate by the Sun, think: "Sun": the mnemonic is: "Sun at high noon — how high?".
You look: North, there, then up there, that high.— Southern Hemisphere (cf Noon will be South in NH) (Don't look at the Sun, if it is already there!)
It might be morning : "It is going up there"
It might be noon : "See how high it gets!"
It might be afternoon: "It came from up there"
Visualising the height and path of the Sun also helps you to interpret plants for 'bush noon' (p.74a).
"Walking Stick North" depends on knowing the Sun's maximum height. (See page 30b)

On Overcast Days, Average the Brightest Sky over 1/4 hour

If the clouds are moving fast enough, and not entirely uniform, sit still for long enough to see many clouds pass over. Keep pointing at the brightest patch of sky using some reference like a tree, and watch to see whether you can confidently locate the average position of brightness.

Stars

RISE and SET Directions for a (**Local**) STAR COMPASS

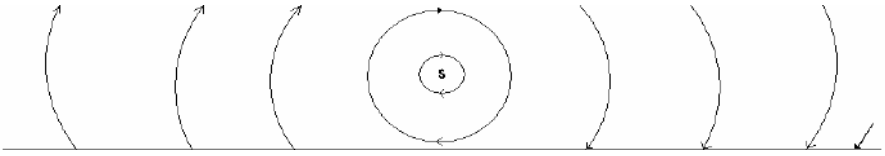

The star directions don't change night by night or with longitude, but they do change with latitude.

The circles in the diagram will rise from the horizon as you travel further south, and so change their intersections with the horizon.

This star rises here every night for the seasons when it is visible

You need to know your stars and visualise their circular paths

Stars which share the same star path indicate the same direction at staggered times during the night

Sun-set and Moon-set (examples only) can be fitted in to the scheme to span the day

Just make sure to mark sunset to a landmark on the skyline horizon and carry it over into night-time.

(SH view)

(Some stars stay up all night every night; these don't move far either, and so are good orientation value too)

Moon

From the Moon, Locate the Sun-Moon Direct Line, Carefully.

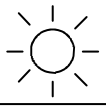

The line is the Approximate Ecliptic, and being a great-circle is a straight line.
OPTICAL ILLUSION!
The Moon seems to be lit from above the 'direct' line to the Sun.
But that may only prove that the Sun is "up" and shining down.

Bust the illusion with a piece of string…
Caution: Don't look directly at the Sun!
horizon

Prove it in 3 dimensions, by using a torch and a ball.
When the ball, in line with the Moon, looks right, the torch direction mimics actual sunlight.
Below horizontal = at night

By locating the Sun, you can guess the time of night, and guesstimate North more reliably, because the path of the Sun is more predictable and familiar than that of the Moon.

Clockwise

Mnemonic: (when you say) "COMPASS" (say) "CANADA"

Compass needles point to Northern Canada. That's where the North Magnetic Pole lives. Lat. 75° N.; long. 100° E.
E.g. from Australia, Canada lies to the East of North…

Check it out on a globe. Locate the Queen Elizabeth Islands.

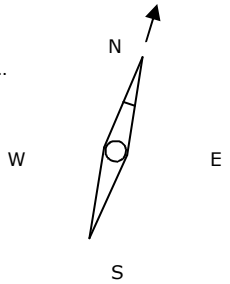

To find the Southern Magnetic Pole, (but this doesn't work so well)
locate the Antarctic coast south of Adelaide, at Lat. 65° S.; long. 140° E.

The Angle of Rising and Setting — is Determined by Your Latitude

To post-dict sunrise or pre-dict sunset direction, you can use your latitude to visualise the tilt at which the Sun **rises** (or the rise-angle of a near-Equatorial Star/Moon/planet) and the deviation from straight down, at which it **sets**.
Remember to use vertical, not horizontal as the expected norm. (To judge vertical correctly, see p19c. To practise, see 30b)

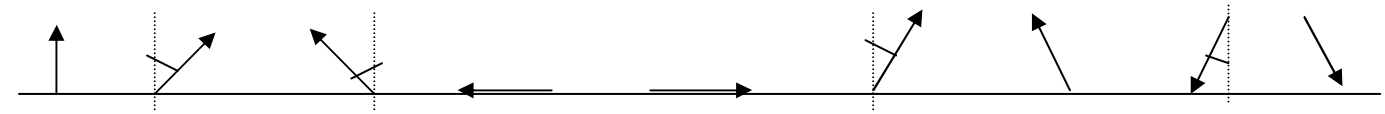

The Sun's path is very nearly a straight line, despite the illusion of being circular—see p.15a.
Hint: Whenever you see the shadow of your head, provided the Sun is not too high, you can make the "anti-Sun" rise or set positions obvious, by following up from that shadow, to the horizon at the correct angle. In midwinter the anti-Sun will be in a midsummer position and vice-versa.

High Noon — How High?

Answer one: the same as last week. Take a look. It changes only slowly.
Answer two: **within 23½ degrees of your co-latitude (or 'polar distance') high**
i.e. 90° – Lat., plus or minus 23½ degrees, depending on the season.
Mid-Summer… co-lat+23.5° (very high in the sky)
Mid-Winter… co-lat–23.5° (very low in the sky)
Equinoxes… co-lat (average height).
Why? At 0° Latitude, i.e. at the Equator, the average Sun is 90° high at noon i.e. overhead. Yes?
At the Poles, 90° Lat., the Sun skims the horizon at equinoxes. Yes?
If you need further proof (that its height is 90° minus Lat.) draw it.

See how the height of the average Sun is 'latitude degrees below vertical'.
That's an angle of elevation of 90° minus lat, isn't it? Now adjust for the Sun's seasonal movement which is 23½ degrees each side of that.

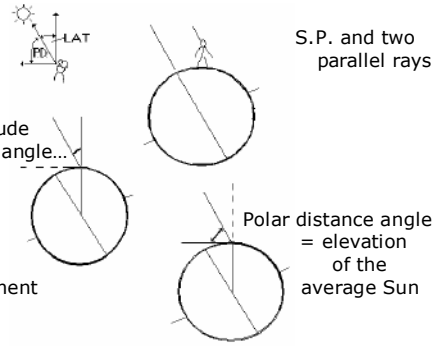

Global Star-Compasses e.g. 'Perpendicular to' rise and set

This is one step harder and three steps better. It allows 'great-circle' course keeping.
A global Star-compass yields 'direct', global, 'straightest path' courses — ones which take you around the globe and back to your starting point.
E.g. You take a rise or set compass-point and head at right angles to it, and the new horizontal direction is global—no longer tied to one latitude.
That's easier to imagine by visualising the light half/dark half of the Earth: All who see the Sun on the horizon could point to it, but each would be pointing in a different "direction". But if they all pointed horizontally at right angles to the Sun, they would all be pointing in truly the same way —around the Great Circle which divides day from night. The same goes for any horizontal star. Imagine the star as a pole star, and use its "equator"
There are many alternative star appearances, besides rise or set & **you don't have to 'go at right angles to' these to set your course...**
E.g. aim at **one star above another**
The midpoint of One star **level with another**
One star **rising or falling vertically**.
Toward a star which has a 2nd star off to one side **at right angles** i.e. lower. You would pass *under* the point you kept aiming for.

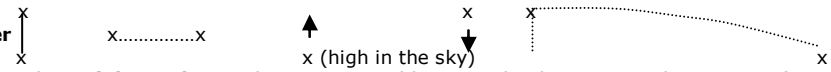

You don't need a clear sea horizon — the stars are high, and a plumb-line will suffice.
These guides occur every night, in the same global direction, for several seasons, then disappear for a while due to daylight, but reappear.
They will occur at the same times next year. **They are always worth learning more of**, adding to your familiarity of favourite 'Star directions'.
These global directions are labelled by saying *how close to the poles* they would take you, and you can see or guess that from the stars, see p. 7c.

'Low' Clouds Come Clockwise Down Under

In the Southern Hemisphere

- Low-**pressure-system winds** spiral **clockwise** around. I.e. into your face from your left as you face their winds...
- 'Low **clouds**' (i.e. below freezing level) follow the surface level isobars i.e. the weather-map lines.

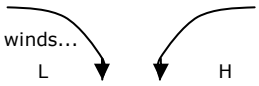

The double-meaning for ' Low ' in the mnemonic reminds you of this.
'Lows' cause more clouds than 'Highs', so you focus on 'Lows'.
I.e. when you see the low clouds; you say 'low clouds come clockwise' and automatically imagine…… '**L**ow On My **L**eft' Plus 'H**igh** On My R**ight**'
The Alternative Southern Hemisphere version of Buys Ballot's law. This **new version goes like this** :
Face the low clouds (unaffected by friction—winds at 600m high are considered unimpeded),
then rhyme: L— L ;igh—igh. (Don't Suntwist yourself.)

[Cf "Face the wind, (In the northern hemisphere you have to say "back to the wind")
Sun-twist yourself up to about 30° to counteract friction (which makes the surface winds spiral inward to the low centre)]
Surface winds could also be expected to be half the speed of unimpeded winds, less impediment over sea, more over land—about one third.

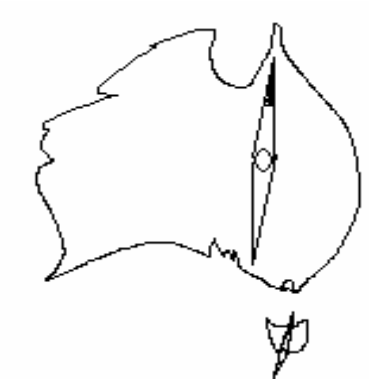

A Compass-Needle Continent—Swinging the Continent

Mentally glue a miniature continent under the compass needle, to swing with it.

The idea is twofold:
1. To reprogram your mind — "The continent is not swinging, so... **I** must be!"
2. To orient yourself meaningfully — to the invisible picture
3. To remind yourself of compass variation

When using a compass you could: just stare at the compass-needle, like a lone night sailor in a fog.
Or: look at what the compass is pointing to — that bump!
Even better: also 'see' the needle lying across the whole continent
Keep going: pointing to Northern Canada
Don't forget: the map you are using — the real one, not this imaginary one — should also line up properly — twist it around until it does.

Global

Global Orientation: Point to the South Pole… ..half its polar distance below horizontal

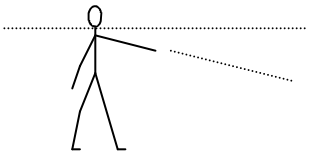

If you had to point "South" you need to point horizontally.
But when you have to point out "the South Pole", you must point below horizontal.
(It is similar when Northerners use the North Pole)
Imagine the true scale… you are on the circumference, not above it...

The question is: "What is the exact angle below horizontal?"
The answer is : "The South Pole is **HALF its polar distance** below South."
Example: at 45° S., you point 22½ degrees below the horizon. (90-45) /2
Example: latitude 30°; colatitude = 90–30 = 60°; half that equals 30°
Example: latitude 60°; polar distance 30°; 15° below horizontal.

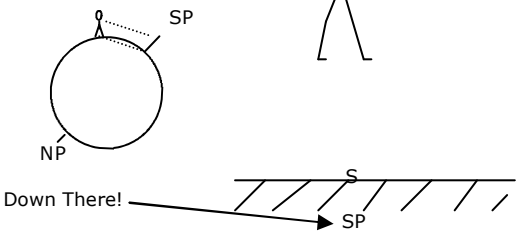

Sun

Optical Illusion — Beware! — Use Your Shadow Not the Sun, For Direction

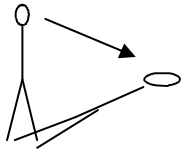

When the Sun is directly behind you, it seems to be behind on the left when you look over your left shoulder,
and behind on the right when you look over your right shoulder! And when the Sun is exactly to the side,
it nonetheless seems to be ahead of you. (Every sports coach knows that when you turn your head, your body follows it.)
Don't rely on where the Sun seems to be… USE A SHADOW INSTEAD! It is also easier on your eyes.

And then don't get fooled by the shadow of your finger, when you throw a shadow onto your map to find the Sun's direction.
It may look straight but it doesn't point to the Sun, until you line it up with your finger**tip** while looking from directly above.
The shadows will always be parallel, but the Sun's rays will always join object to image, fingertip to fingertip-shadow…
The first diagram is illuminated from behind you……
To avoid looking at the Sun, try this:
Point your walking stick from somewhere beneath your eyes, outwards, and rotate yourself until it is in line over its own shadow.

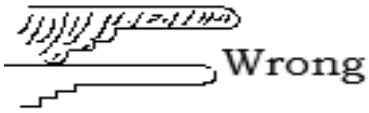

Stars

Right Around the Horizon, Stars Move to Your **Left**, Down South

Looking South, all the lower stars move left.

It is only the high Southern stars which move to your right.

Looking East all the stars rise to the left
(i.e. we see the bottom half of the circles down at horizon level, looking South)

Looking West all the stars set to the left

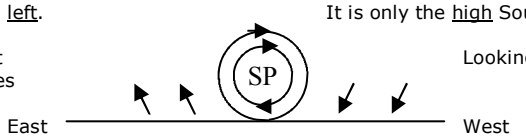

Looking North, all the stars move left.
(we see the top of the circles)

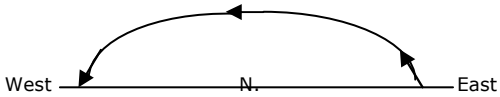

The same goes for Sun, Moon, Planets.

Looking directly overhead, or behind your head, will not work!

(it is all opposite in the northern hemisphere)

Moon

Mentally Rotate the Sun-Moon Line Around Polaris

A whole line through the sky gives a better 'feel' for direction than a single point — than either the Sun or the Moon alone —
So join the two and watch them rotate together during the rest of the day.

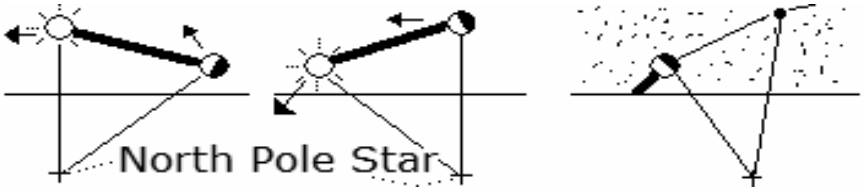

(Southern Hemisphere view)

When you can see both, and you know where North is, you watch them.
When you see only the one, you look for the other, and so find north more exactly than by using either one's path by itself.
Add in any planets to continue the process at night, then from planet to planet or to Zodiac stars, after moonset.

Clockwise

"Bears Circle Clockwise"
Compass Bearings are **C**lockwise, A Full **C**ircle, and Euro**c**entric, **C**ee?

In Europe the Sun moves clockwise… from North (Where the Sun is at midnight—0 hours) = 0° as far as the horizontal angle involved is concerned.

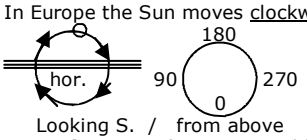

East is at about 06 hours (sunrise) or 90° True — Sun-wise or Clock-wise from North
South at about 12 hours or 180° T (noon)
West at about 18 hours or 270° T (sunset) — so you see, the bearings circle clockwise, from North, 360°.

A **Southerner**, of course, would prefer to go 360° **anti**-clockwise, from **South**, but they didn't get a vote.
Everyone in the world is expected to submit to the (NH) rule: compass bearings should be measured in degrees, up to 360°, from North, clockwise around through East — like the ancient Babylonians; like the Northern temperate Sun.
Not many people consider using 400 grads (4 right-angles of 100 grads make up the circle), or consider going left-around, through West,
Or even think of starting at South. So don't you dare (unless you **don't** call it "a compass bearing").

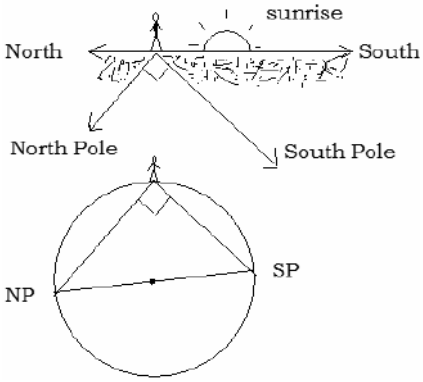

The North Pole is *Perpendicular to the South Pole*

Since the poles are 180° apart — opposite — you see them as 90° apart — half the arc distance.
The numbers can't work out any differently.

Example: 10° S. Lat. = Equator is 10° North of you. South Pole is 80° away — the "Polar Distance".
So S.P. = 40° below S.
North Pole is another 90° past the Equator.
So N.P. = (90 + 10)/2 = 50° below N. And the gap is 90°.

A simple way of looking at it is to say
"Diameters yield right-angles at the circumference"

Now go outside and point to the Poles.

Earth

The Dark-Blue Polarisation Band Perpendicular to the Sun

90° from the Sun there is a circle of darker sky. Bees navigate by this clue. The whole sky shows degrees of polarisation.
You can see it naked-eye, but since it is due to polarisation, it is much clearer through Polaroid sunglasses or a camera filter.
Point the top of your polaroid sunglasses toward the Sun for best effect.
The dark blue band is narrowest highest up (in the direction of your shadow, and perpendicular to that sunbeam) and varies with visibility and haze.
The sky is darkest away from the Sun and away from the horizon haze and 90° from the Sun.
You can always guess where the Sun is, from the gradations of blue. After sunset the band is caused by the brightest part of sky.
Example: You are in a forest at dusk. The sky is blue overhead, but the sunset direction is not visible. Find the Sun from the blue!
See the "Sunny Skies System" (p.19e) for how to use the ends of the dark blue band, and its trend.

The Centre of the Rainbow is Down-Sun

The shadow of your head coincides with the centre of the rainbow's circle. (Pilots often see the full circle of a rainbow)
The radius of this circle is two wide hand-spans — spread out your 10 fingers fully, thumbs touching.
Whenever you see a rainbow, treat it as a clue to 'shadow-South' (p.14b) i.e. relate its position to the time of day and to South.
The shadow of your head is on the end of the shadow of you, of course, but when you can't see it due to clouds, you might be able to see a rainbow.

Sun

Identify the Stars That Pass Directly Overhead

Each star has its own latitude, called "declination", which is very nearly constant.
So the same series of stars of the same latitude will pass overhead that latitude, each night, each in turn, each at some fixed star-time.
This is true anywhere on the globe.
Use a plumb-line. The top of the plumb-line points 'up' to overhead (when looked at from any side, so) where the 'up' lines cross is 'overhead'.
Look past it from 2 or 3 directions to find overhead, especially looking North/South for longitude-timing, and East/West for latitude-positioning.
Then use a star-chart to identify "exactly overhead", latitude-wise, at least. If you can do this, even without a watch, this will fix your **latitude**.
Plus or minus a few kilometres is the expected maximum accuracy from using a plumb-line.
For **longitude**, you need to determine the star-time at which one of these stars is overhead, or at least passing from East of you to West of you.
Greenwich (absolute) star-time, not Local (relative) star-time, that is. To the nearest 20 seconds is okay. See p30c. A 4 min error is 1° error in long.
Alternatively, set your reference star-time to some other longitude, e.g. home, or start, or destination, or a round number.
This time & position data is life-long, give or take a few kilometres of predictable precessional drift per decade.

You can get back to that place later — Match the appearance and timing. It's easy to tell which way to move in order to "match up" (p22a,c; 23c).
You need to know a star's longitude, so you need an almanac or list. Look down the list for bright stars with your latitude, and go look for them.
You will quickly learn about having to wait for the right 'star-time'!

Stars

Weather Fronts Distort The (Circular) Wind Circulation

Buys Ballot's law is only approximate, because of the sudden wind-shift at a cold front.
Cols, Ridges and Troughs are likewise noncircular. E.g. 'centres' will sometimes split into two…

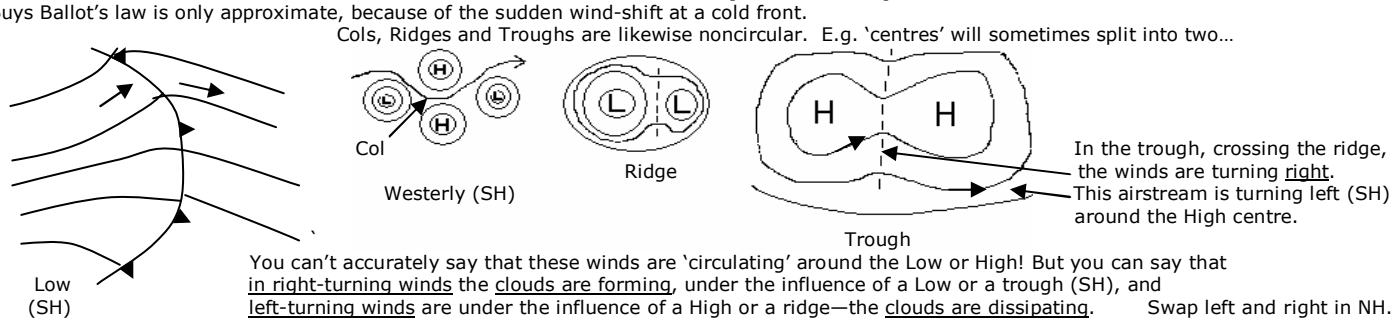

You can't accurately say that these winds are 'circulating' around the Low or High! But you can say that
in right-turning winds the clouds are forming, under the influence of a Low or a trough (SH), and
left-turning winds are under the influence of a High or a ridge—the clouds are dissipating. Swap left and right in NH.

Wind

Mnemonic: (Interpret a Compass Needle as) "RED HOT — WHITE FROST"

Southerners are looking North to the "red, hot, tropics, dripping with sweat, in the Sun"…

When southerners look South, it is to the "white, icy, Antarctica, frosty cold, in the shadow"…

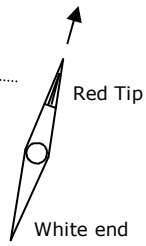

So when you notice the red point of the compass needle you say to yourself "red-hot!"
and try to feel boiled on that side (as if by a bon-fire), and then visualise the invisible tropics North of you.
And when you notice the white point (e.g. during a back-bearing) you say to yourself "white-icy cold-frosty-Antarctica"
And then try to feel frozen on your 'Shadow-South side', and imagine the South Pole, e.g. point (down) to it.

You can also remember that the Canadian flag is red (but it is not hot in Canada).

Anticlockwise

Global

Global Orientation: the South Pole Star, and Polaris

(In the Southern Hemisphere…)
Polaris is <u>below</u> the northern horizon, by exactly as much as your South latitude.
E.g. at 45°S. lat. Polaris is below North by 45° Around it the northern stars circle anticlockwise……

The equivalent "South Pole Star" (a dim pinprick in the midst of dim pinpricks) is
<u>above</u> the southern horizon by exactly as much as your South latitude. E.g. 30° S. = 30° above South
Around it the Southern stars circle clockwise, once per day-night.
Locating the South Pole Star, helps you in the following...

- You can use the bright stars — Cross, Pointers, Achernar — as hour-hands on a 24-hour clockface.
- Star-paths Circle around the Pole Stars
- So do the horizon, the Sun, the Moon and the planets — You can predict all their positions.

Finding the Pole Stars is Fundamental to Visualising Rotation.

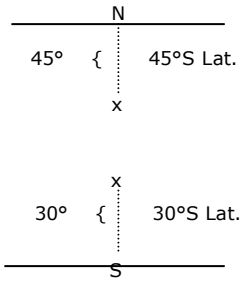

Sun

- Your Shadow Rotates with the **same <u>clock-sense</u>** as the Sun — not surprisingly!
- Shadow-Rise and Shadow-Set is **opposite <u>direction</u>** to the Sun — not surprisingly!

<u>Together</u> they <u>can</u> surprise you!

- Mnemonic: (Face the opposite direction as for the Sun mnemonic p.5b)

<u>Face **Sh**adowy **S**outh</u> for **S**outhern **H**emisphere (Face Su**nn**y South for **N**ortherners)

"Shadow **Ri**se is on the **Ri**ght; S**e**t is on the L**e**ft" (Same rhyme as for sunrise and sunset, but facing South, both hemispheres)
Your Shadow Passes **South** of Southerners At Southern Noon (compare: 'The Sun is North of Your Nose at Noon')
(North of Northerners at Noon)
Your Shadow Moves **Anti**clockwise — around to Your **Left** — Down Under
(Clockwise — Around to Your Right — in N. H.)
(The Southern Hemisphere Horizon still shifts to the right, in real terms. The Northern Hemisphere Horizon still Never Heads Right')
In addition, notice: **The shadow *<u>tip</u>* moves due East, at around noon,** and in a much straighter line than you expect, all during the day, as the shadow itself changes, from pointing to opposite sunrise, then to pointing to South of the object, then to lying opposite sunset.

Stars

15 Ways to Locate the **South** Pole Star Or the N./S. Meridian

1. Halfway between Achernar and Epsilon Centauri (located from the six bright stars near it)
2. Rigil Kentaurus to S., at right angles to Hadar……
3. Beta Crucis to S., parallel to gamma-to-alpha Crucis, 5 Cross lengths … Use 2. and 3. together
4. Beta Hydri to alpha Crucis, as far as the beta-to-gamma-Hydri length … most accurate
5. Mirzam through Canopus
6. Suhail through the top half of the False Cross through Miaplacidus
7. Make a triangle out of the two Magellanic clouds toward R. K. — very crude but useful
8. Gamma Hydri to beta Triangulum Australe — the "back" star of the Triangle
9. Alpha Hydri to Hadar
10. Arcturus to between the two 'Pointers' (R. K. & Hadar) } These are used in reverse
11. Spica to between the Cross and Pointers } — to <u>locate</u> Arcturus
12. Middle of Corvus through middle of Cross } Spica and Corvus
13. Atria to the star between Canopus and Achernar, alpha Dorado
14. Peacock to Avior
15. 'Your Latitude' degrees above horizontal, due South

("Crucis" says "of the Crux" i.e. of the Cross)
(gamma looks a bit like a thin alpha twisted 90°)

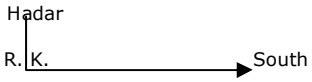

Moon

"Cusp North" — Approximate Direction from the Moon

The Sun-Moon line marks the approximate Ecliptic.
The "centre" of that circle of Sun-planets-asteroids is located in "Draco",
quite near "Polaris", the North Pole Star. For a start, ignore the difference…

- Join the Cusps, tip to tip
- Drop the line diagonally
- Keep going down to (near) Polaris (latitude degrees below horizontal)
- Come vertically back up to "Moon North" (in the <u>worst</u>-case +/-60° "accuracy"!)

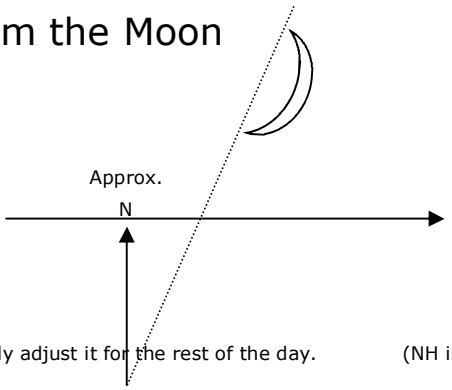

You have to do this only once or twice a day. Compare it with True North and automatically adjust it for the rest of the day. (NH is similar)

Clockwise

Mariner's "Compass Points" work on Successive Halvings

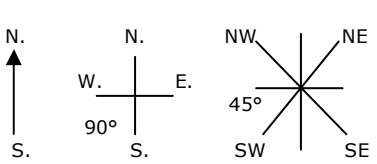

180° difference

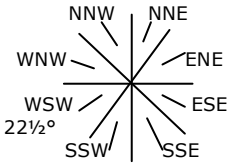

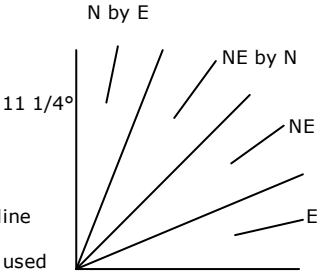

Start with the nearest major point; add 'by'

'by' means 'come 11 1/4° towards' (or 'one point')
There are 32 points altogether on the compass.

One Point is approximately One Palm wide
(32 around the horizon—test it)

Make your own emergency protractor (e.g. as a compass, sextant or plotter) by folding any piece of paper, and folding it again and again, matching line to line so as to exactly halve it with every fold.
One more halving and you get 64 divisions, each of "100 mils" (so named and used because **1mil is approximately 1 in 1000 e.g. 1 m in 1 km**) i.e. 6400 mils around, clockwise, 17.7778 mils per degree.

Pole Star Locations — "(Latitude)° above and below, South and North"

Why? Think: at 0° Lat., they are 0° high. Yes? At 90° Lat., they are 90° above and below horizontal. Yes?
If you need further proof: Draw a Diagram of two parallel rays to South Pole Star...

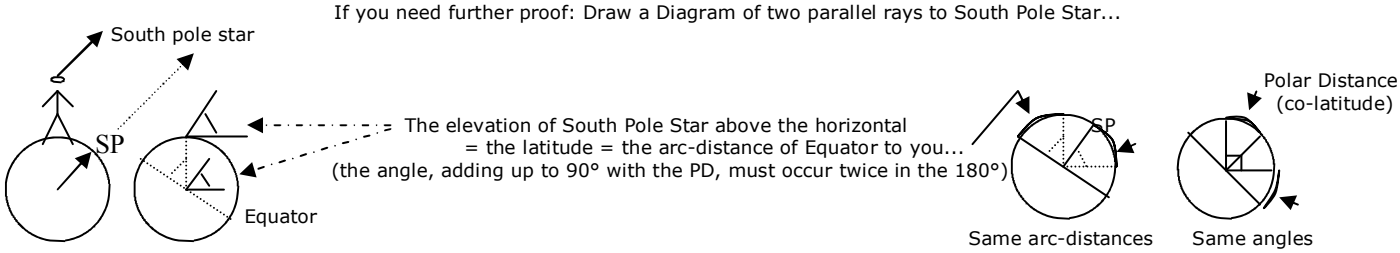

The Sun Passes Quickly Past North-South Nearer the Tropics

This pattern shows
the hourly change
in Sun or shadow direction
at 30° latitude, Mid-Summer...

The Sun is East/West
at (noon –/+ 2 hrs 44½ m.)
at 52.9° high.

This shows the
seven hours' sunlight
(one side of North/South).

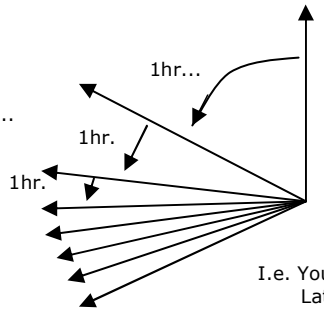

Don't get caught misjudging the Sun's horizontal direction near midday.
When it is high, it moves more than 15° per hour in horizontal direction.

At other latitudes the pattern will be different — more uniform
away from the Equator, and with a different day-length,
depending upon the season.
I.e. You need to calculate each particular pattern from the
Latitude and declination and season (as if calibrating a vertical stick sundial).

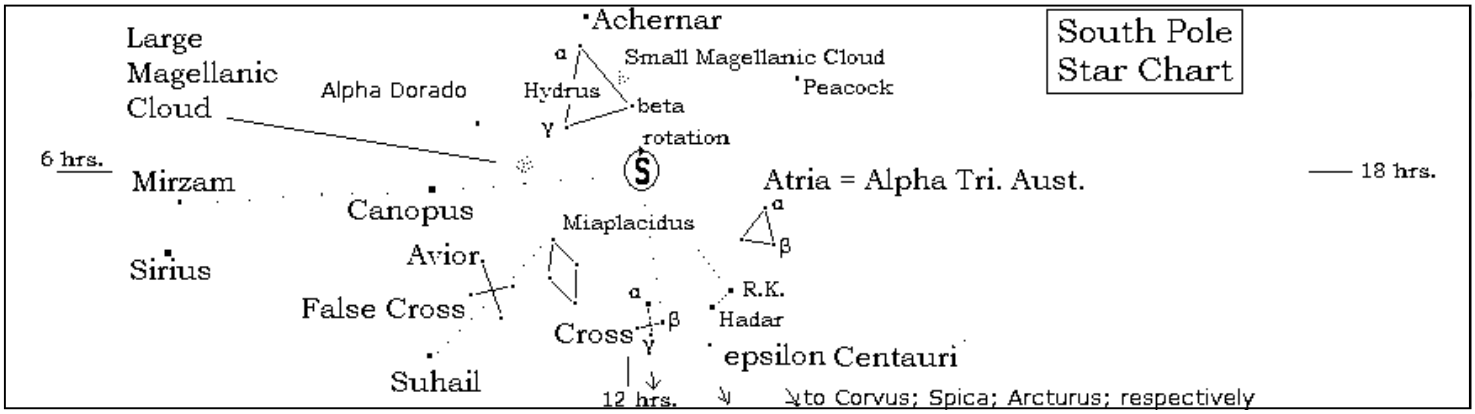

Wind-Direction-Change Rule for Buys Ballot's Law

If "the Low is on Your Left Or the High is on Your Right"...**and** the wind is gradually **changing**: Add in this new information as follows...
Imagine how the Low and High are moving across the ocean/continent as you **reenact the wind change**...
Exaggerate the change as you *twist yourself to follow the speeded-up movement* of the wind change in your face.
Point to the two choices (left & right) while you twist, then ask yourself which of your hands is moving to the east.
Usually only one of the two options above — 'Low' or 'High' — will make sense of the general movement of the weather systems to the East.
Temperate systems usually do move to the East from the West. (Both Hemispheres)
..From 0 knots (stationary highs) to 15 (average) up to 25 knots (Lows in winter).
Tropical cyclonic hurricanes move Westerly initially at 25-40 knots, then recurve to the pole, but the "re-enact" rule still works to show up its movement.
In this way you can decide: "These winds are from a Low" or "from a High".
Don't, of course, confuse local winds, e.g. a sea breeze, with the more general 'synoptic' circulation due to the large-scale pressure systems.
To further help you decide where the wind has travelled from —from your right or from your left— check out the temperature and absolute humidity — i.e. the air-mass characteristics. Is the air hot and dry, or cool and moist, or warm and humid — which one is supported by the geography?

Four Reference Directions — Name Them: "**Rise, Noon, Set, Midnight**"

..not E N W S

The trouble with "North" is that it is sometimes behind you... then you have to think backwards.
Use all four and you always have one in front of you... N., S., E., W.
Give them meaningful names as in the title... and you can picture the meaning.
Base it on the Sun's daily movement and the effect is vivid.
To flesh out the meaning even more, **add in a second clue: "Rise/Start; Noon/Hottest; Set/End; Midnight/Coldest"**.

What this does is to take the outside picture inside with you, when you enter a new building, a hospital complex, a shopping centre.
E.g. "I'm now looking toward the mid-morning Sun direction, i.e. 'after rise' or 'before noon', which is 'warming up' ".

It is then up to you to
1) get to know the Sun's daily movement.
2) locate the Sun's path outside, once.
3) *keep track* of your twists and turns when the Sun is not visible e.g. at night.

Ignore the time of day or night—it is the directions which you are imagining. You use the real Sun only as a start-up hint.

Global

Point to 2 Places on the Equator — Halfway between the Poles

You can find the 2 Poles, apparently 90° apart, so that means you can point 45° — halfway — between the Poles, to find the Equator, then 90° to that for the opposite Equator (the "45°" is because the Equator is 90° from each Pole and we halve that arc).
It locates (1) the furthest point on the Equator from you e.g. Brazil then...
(2) the nearest point on the Equator from you e.g. New Guinea—ish.

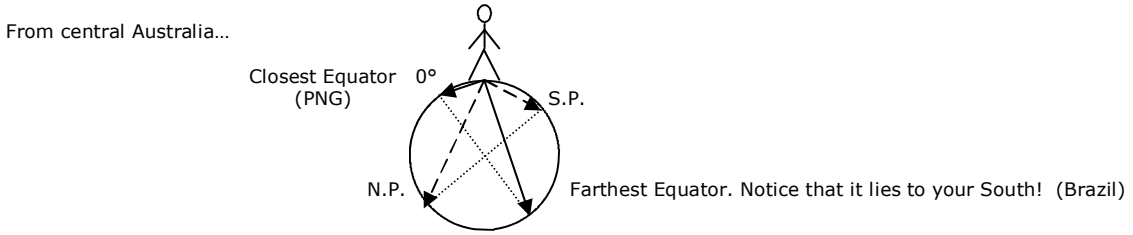

Sun

Shortest Shadow South — Opposite Highest Noon. Always Imagine "Shadow-Noon" on your "Shadow South" Side

i.e. Your shortest shadow is South of you at noon. Memorise how short it is (this week).

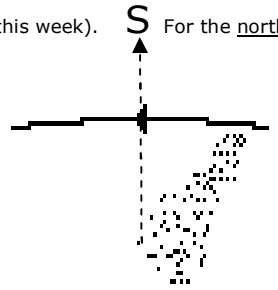

For the northern hemisphere use "North" ("shadowy-North")

(Shadow Noon is Opposite Direction to "Noon-North") ..."Noon-South"
Imagine your shadow moving left around… ..right-around

In the morning: South must be "ahead" of my shadow ..North
At midday : Look — see how short it is! … **Due South at noon** **..North**
After noon : My shadow has "passed" South ..North

To guess North from your own shadow, see "Walking stick North". (p.30b)
Try to feel the icy South, freezing that side of you! ..North
(As on a cold night facing away from a campfire).
Mentally **visualise**: Antarctica below horizontal; an aurora above that horizon; Southern stars behind that blue sky. ..Arctic; Aurora Borealis
In the tropics, you'll have to be more observant, to know your shadow's expected movements more accurately. ..Northern stars

Stars

Find the Polar Axis by Star-Trail Bisection

You need to imagine the centre of the star-circle.
A Hypothetical Example: Lost in a stormy night! You want to find North-South.
You catch the occasional glimpses of unknown stars.

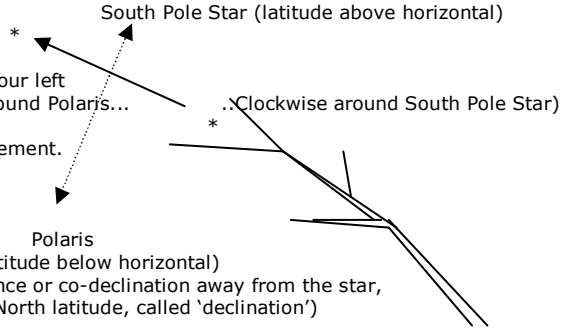

Lie down and line up one such star with a tree branch.
Maintain your exact position and wait for another glimpse.
It may only need a couple of minutes to see its direction of movement.
You can also point something straight at the star and fix it in place, to see the star's movement.
Bisect the result. It points to Polaris, or to the South Pole Star.
You now need to know your latitude, roughly, to locate Polaris or South Pole Star.
The higher the star the clearer the result — it will be moving West.
The only ambiguity is a low star moving horizontally (to your left).
The Southern stars will be moving slowly; northern ones, quickly
— so you may be able to guess which is which.
Northern Hemisphere is similar — but Polaris is visible.

Moon

The Rabbit Looks up Ahead at Full Moon Using the Ecliptic

When the 'cusps' disappear at full Moon, the Rabbit ears and face appear,
and let you know which way the Moon is moving.
These views are for the Southern Hemisphere — where the 'Rabbit' is upright.
In the first diagram, the cusps are difficult to distinguish, and so is the Rabbit.
In the second diagram, you don't need the cusps, because the ears show.
The dashed arrows point down to near Polaris.
As an alternative to cusp-North, you can use the Ecliptic perpendicular to the cusps.
The other, solid arrow, falls roughly along the Ecliptic, with the light side of the Moon
pointing to the Sun and the dark side pointing to the anti-Sun.
Therefore you can judge approximate sunrise or anti-sunrise, sunset or anti-sunset,
just as soon as you see the Moon, either from its cusps, or from the Full Moon.
Find the nearest horizon on the Ecliptic and picture which of those 4 directions it is.
A setting Moon in early daylight will show anti-sunrise, but after dusk will show up sunset.
An eastern Moon before dawn indicates sunrise, but in the p.m. will point to anti-sunset.

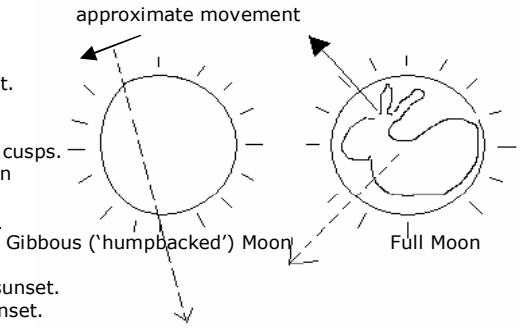

Clockwise

"E. W. N. South" **Of What**!? Rule: **Say** What

- Say 'True' and you mean 'degrees East of True North' 270° T = due W.
- Write "Gee" and you mean: relative to the map **grid**. 270°G ≠ 270°T Both are clockwise around, but they start off from different directions
- S20°E means '(Start at)South(;come)20°(toward the)East' e.g.S20°E=20° to the right of the white end of the compass needle. (old system)
  Clockwise/to the Right/East-around-from-North is standard if not otherwise specified.
- 270° M says 'Magnetic' — what a good compass should read.
- 270° C says 'Compass' — what this compass does read.

Note that these are standard: 'G' does not mean 'Geographical' or 'Great Circle' or 'Grads' but is recognised as 'Grid'.
'M' does not mean 'Map', 'Map grid' etc.
xyz° can not mean anti-clockwise.
The remaining problem is how to shift between different methods… See section 4: "East Least West Best" rule. (p.107b)

Perspective Distortion — The Sun Moves in a Straight Line!

The 'horizon' (at sea) seems "straight", yes? but joins to itself! This is because you see it from the centre of a circle. That's no problem yet.
Two different great-circles circling around you do cut twice, but *it certainly looks strange*, since looking at the gap between them,
you see "parallel straight lines"; Yet to the left **and** right of that you will see them both join!
Great Circles Intersect Twice. **Nowhere does either circle seem to 'curve', yet they manage to meet.**
The Sun 'comes up and goes down' but is actually 'moving' in a nearly straight line—as straight as the horizon…

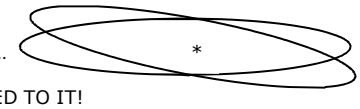

PUT YOUR HEAD INTO A HOOP; GET OUTSIDE; CHECK IT OUT; GET IT INTO YOUR HEAD: GET USED TO IT!
Caution: Don't look directly at the Sun. Watch your shadow and you will see that it moves in a nearly straight line from East to West.
The Sun's average path — the 'Equator' up in the sky — where the 'Saucepan' (Orion; the Pot) moves — is a great circle.
It intersects the horizon due East and due West. At times of year other than at Equinox, the Sun's path is still almost a great circle.
The Milky Way is another maximum sized circle; and the Zodiac (i.e. the Ecliptic) is another such 'Great Circle'.
They hit the horizon twice, in opposite directions (not E/W).
2 other 'great circles' you need to visualise are: Satellite paths; and the Dark Blue Band. (Star-trails are usually 'small' circles, and do look curved.)

Earth

Locate Polaris (and Find North) by Imagining the Centre of Circling

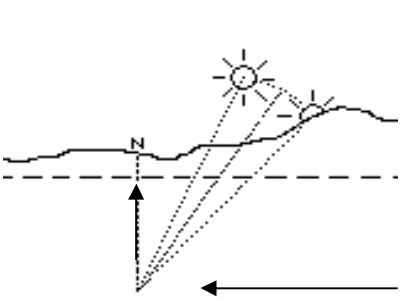

This works for the Sun, Moon, planets or stars.
You *find the centre of their circular paths by bisecting the chord*.
(Northerners locate the South Pole Star, South)

The **perpendicular bisector** of an arc or chord
passes through the centre of a circle...

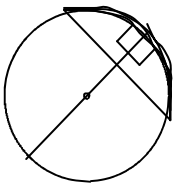

Moonrise varies quickly from day to day
— make sure you check its exact position.

Don't look directly at the Sun to do this!

Polaris is, in degrees, at 'your south latitude below horizontal' (not the horizon) — judge it carefully.
and/or, 'polar distance (co-declination) away from the Sun': 90°-N dec, e.g. dec=10°S 90-(-10)=100°

Sun

The Eagle Clips the Equator, South of Altair

Find the star 40% along the southwest side of the rhombus…

Finding the Equator helps you

find the Ecliptic (not shown)
and find where the Milky Way crosses the Equator…

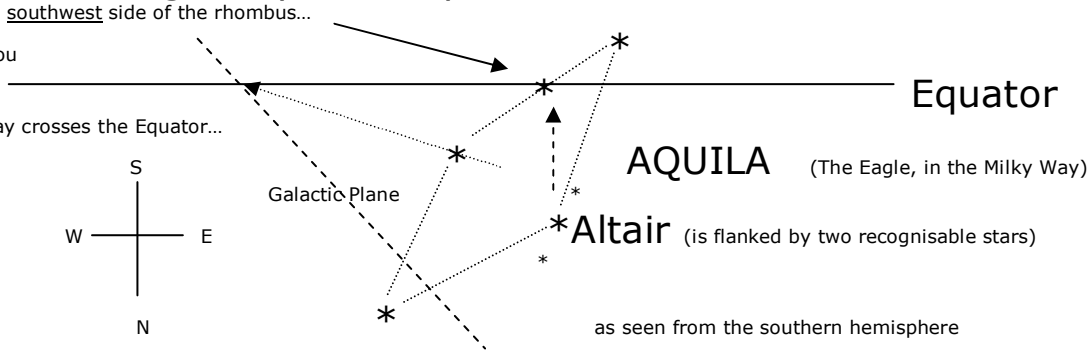

Stars

Steering Winds

Mid-to-high-level (upper) winds tend to push sea-level pressure-systems along
at **half the speed of the upper winds, and in the same direction**.

This rule is said to work best with small, weak, less well-developed systems; and with hurricanes.

E.g. If the low is on your left And if the upper winds are coming from your left
Then the storm centre might pass straight over you.

Since the upper level winds often oscillate in the 'Rossby Wave', and tend to generate high and low pressure areas beneath them
(because the 'wave' causes the upper winds to slow down or speed up) the High-pressure systems will on average be steered slightly toward the Equator, and the Lows will go slightly toward the Poles, as the systems move to the East — a rule of thumb worth remembering.

But this extra steering-wind rule will alert you to non-average situations.

Wind

"Rise-Noon-Set-Midnight" carries its meaning across the Equator! N.S.E.W. doesn't.

Name the directions according to the average outside picture…
This is a visual system and addresses the disorientation of turning around.
For back-bearings i.e. when you turn around: replace 'rise' with 'set',
'noon' with 'midnight' (start with end, & hottest with coldest)
e.g. if the old direction was "after setting" the new one will be "after rising".
When you turn right or left adjust by one of the four notches
e.g. 'after midnight' becomes 'after rising' if you turn 'with' the Sun.
Note: 'after the setting' (or 'before midnight') is a direction of the evening Sun
— S.W. — and can be bright sunlight, in midsummer !
And 'after midnight' ('before East') is the direction of the South-East Sun.
At Mid-Summer sunrise it can similarly be very much bright !
Example: You enter a hospital with "setting" (i.e. due West) on your right, so you start
at "midnight" (SH). You then navigate through look-alike wings, look-alike wards, and
look-alike staircases, remaining oriented, and finally emerge, expecting "rise" on your right — having not looked out of a window for hours!

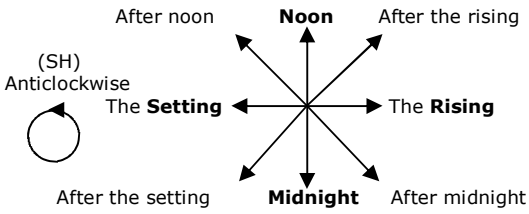

(The Northerner's version needs to go clockwise!)

Anticlockwise

Global

Global Orientation: Half the Arc-Distance below Horizontal

Point out any country or place on Earth, by first finding the (arc-)distance there, and then pointing below the horizon…
The method works **from any place on Earth, in any direction, to any place on Earth**.
Step 1. Measure the angle around the circumference to it. E.g. guess from a globe of the world (This is an angle at the centre of the globe)
Step 2. **Halve** it. That is how far below horizontal it is.
E.g. South Pole from 30° S. Point 30° below since it is 60° away
E.g. Perth to Peking (which is directly North) is 32 S. to 40 N. Therefore the 'arc' to Peking/Beijing is 72°…
Try this (outside): Go through the motions of pointing out a whole circle of real places on earth,
from the horizon ahead of you, to the horizon behind you, using your index finger.
E.g. point North, then down a bit to below horizontal e.g. New Guinea, now onto the North Pole.
Continue down until pointing down underneath you e.g. mid North Atlantic.
At this stage you have pointed through 90°, but have pointed out 180° around the Earth.
Turn around to continue, (Brazil, Antarctica) bringing your finger higher and higher until it is back to horizontal
pointing at the hill on the Southern skyline.
Your index finger has pointed out a 360° circle (as measured from the centre of the Earth) by traversing only 180°!
Get Hold of a Globe. Go Outside. Point Out A Few Countries.

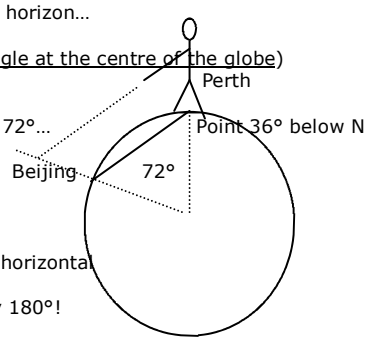

Sun

" Noon " is Not at 12 O'clock

' High Noon ' is at 'midday' i.e. in the middle, between sunrise & sunset. And that is **not '12 noon' or '12 midday'** but at North — Noon North. (SH)

- In summer-time 'high noon' may be near 1 or 1:15 p.m. or so. If you are on Summer Time, true noon is one hour late.
- Even in winter-time it may be over half an hour away from 1200 hrs. This is because we use a one-hour time zone to cover 15° of longitude.
  i.e. you may be near the next time-zone, longitude-wise, and therefore be up to one half an hour 'out'.
  You can check this up if you have a map with longitude on it. The middle of the time zone is at 7½°, 22½°, etc, in steps of 15°.
  Then it will take 4 minutes to travel an extra degree westwards. Work out how early or late the Sun will be in your longitude.
- Plus; the Sun speeds up in its 'orbit' 'around' the Earth when it is close in the Southern Summer, getting up to 20 minutes 'fast', by the end of summer; likewise it can be up to 20 minutes 'slow' at other times, in addition to the time zone effect. See the Equation of Time prediction opposite.

The best thing to do about this, to predict when, exactly, it will be at 'midday' today, is to look at the Sun yesterday.
(Don't! Use a shadow instead. Notice when it is due South.) Or see 26a.

Stars

Equatorial Stars Rise and Set Due East and West

Always

Globally

- The Sun at Equinox — about March 21st, September 22nd
- The Pot in Orion — Mintaka, the north-west star of Orion's belt — the one that sets first — is right on the Equator
- Various dull stars — like the one south of Altair
- A Half Moon at solstice — give or take 5° or so — It must be 90° from the Sun which is at the solstice position, so that the Moon is at the ecliptic equinox (Equatorial) position.
  Other phases can be worked out, e.g. Full Moon at Equinox, but see p. 29d on how to predict an equinoxial Moon every fortnight.
- Planets 90° from the Sun-at-solstice
- The equinoctial points, if you can identify them yet. One is between Spica and Regulus; the other is between Diphda and Markab.

Moon

The Moon Cusps Push and Pull the Hour Hand **Backwards**, Down South

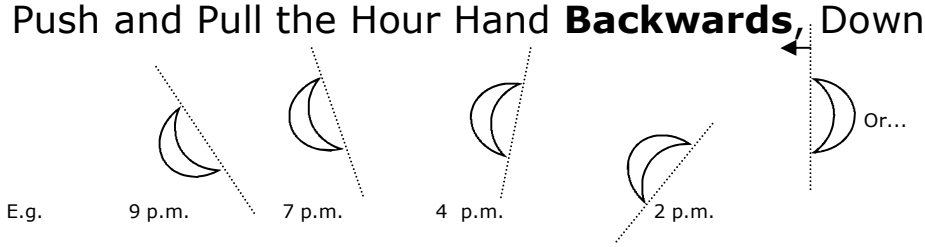

Watch the Moon rotate during the day, and night. Tell the time from it. Update yourself once per day.
The Moon rotates like this about 14½ degrees or 15° per hour. So it is like a 24-hour clock hour-hand running "backwards".
The correct vertical to use as a reference is not at the Moon, but at the Pole Star, if you want to be accurate.

(NH — the Moon seems to turn clockwise)

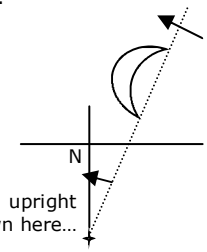

Clockwise

Back-Bearings Killed Sleepy Pilots

Back-bearings are opposite directions — "going back home".
The Babylonian system of 360° in a circle has a lot to answer for, when you want to turn around!

What is required, is to add or subtract 180° (groan)!

But after an all-night mission, tired navigators can't be relied upon to come up with the right answer.
In World War II, many air-crews perished from following the wrong back-bearing back.
Happy, relaxed, but lost.
It was a problem requiring a solution, and a simple solution was found. What is it? (page 18e)

Half the Arc-Distance below Horizontal — 2 Explanations Why

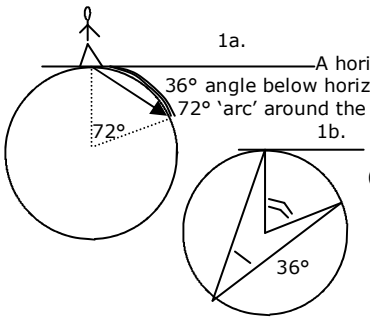

1a.

A horizontal tangent

36° angle below horizontal

72° 'arc' around the circumference — twice the angle from the tangent — but why?

1b.

The angle at the centre (used for 'arc-distance') is always double the angle at the circumference...

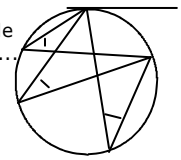

1c.

..It doesn't matter which point on the circle you use, as long as it is on the one side —on the larger arc here— you get the same angle each time.

The tangent point is one such acceptable point — the last one.

2. (Isosceles triangle angles... (180 – 2a)/2) = 90 – a

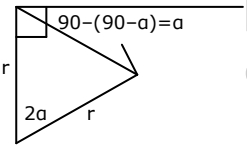

1d.

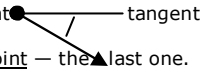

The "Equation of Time" Diagram

The real Sun moves quicker through the sky near Christmas because it is closer to the Earth (i.e. we orbit it quicker, elliptically!).
I.e. The Sun 'gets ahead of' its predicted average position, according to our clocks, in the New Year,
And doesn't come back to average until about the end of August.
After that, during September to December, it will be behind average. But then, in addition, the details are complicated by how it also zips quickly past (gaining on) our longitude lines near solstice — summer or winter — when it is near +/–23½° latitude.
Here the longitude lines crowd together, if only slightly, giving the maximum effect in June & December, super-imposing upon the preceding scheme.
The **overall result** may be graphed... Oct.-Nov. +16 1/3m (both hemispheres)
See the mnemonic on p. 86b

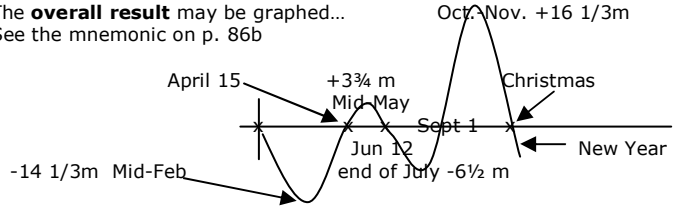

Note that when the astronomers say '+', the Sun is behind itself; we look at the position of the Sun (not directly at it!); it comes up 'late' and we have to add (about a ¼hr) to our guess at the apparent time, to find the clock-time. We say our clocks are 'fast' at predicting the Sun's position, but it is really the Sun which is lagging!
cf Australia is in a "**+**10 hr time-zone" = 10hrs ahead, or fast on Greenwich

Rigel-to-Capella is a good S./N. Pair

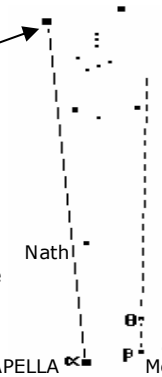

This pair points to Polaris (and hence to North, *when they are 'Upright'*) (this view is for the Southern Summer)

You can also see that the Pot handle points down through the middle of Orion's belt to Meissa and on to Polaris. The Pot 'handle' is actually the scabbard of Orion's sword—seen upside down (SH)

If you cannot see Capella, then Bellatrix to Elnath, right alongside, is also a good guide to Polaris.

Auriga, The Charioteer

....**The six o'clock stars are a useful N./S. pair** — they are upright at 6 a.m. local star time.

'Right Ascension', or star-time, in hours of longitude East of the first point of Aries

Atmospheric Pressure Changes

The air pressure falls before a front, and rises reasonably rapidly as the front passes. It falls in a trough; rises in a ridge; falls in a heatwave inland; falls with altitude (about one millibar per 10m).
The atmospheric pressure normally fluctuates by a couple of millibars every day in a twice-daily cycle tied to the Sun (3mb in the tropics) and the wind speed may vary with it. The daily pressure on average peaks around 10 a.m. and 10 p.m., being lowest at 4 a.m. and 4 p.m.; This on average, is said to vary the wind speed by a couple of knots, lowest at 9 a.m., peaking at 3 p.m., but all of this is usually masked by other weather changes in the temperate latitudes. Pressure fluctuation is much more significant in the tropics, with any fall (or flattening) more than a few millibars during summer/autumn warning you of a possible hurricane (You then look for long wavelength ocean swells preceding the storm centre). Because we can't feel atmospheric pressure changes, it is very helpful to use a barometer, to detect the longer-term pressure change trend. Test your barometer to see that it does not respond to temperature changes!

Watch the Clouds Cross the Moon

You can detect the upper wind direction most easily by watching clouds move past some fixed point in the sky.
Look during the night but don't forget to look during the day also.
The stars, high buildings, high trees and mountain tops also work well.

The **0, 1, 2, 3, 4** Reference Directions

Put these numbers onto the Rise-Noon-Set-Midnight System, so as to follow the Sun.

0 = midnight, South (Southern Hemisphere)
1 = rising, East
2 = noon, North (use South for northern hemisphere)
3 = setting, West
4 = midnight, South — use cycle 4 i.e. call it "0"

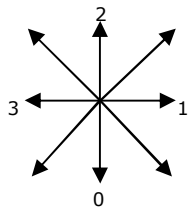

This corresponds to measuring anticlockwise! from South! in decimal degrees, that is, in grads!
100 grads = 90° so "1" = 100 grads; "4" = 400 grads or zero grads — add or subtract "4" at any time.
"0.1" in this system = 10 grads; 1/10 of a right angle; **one fist at arm's length**.
Example: In the thick bush, you see your destination but know you will soon lose sight of it. You can't keep a straight line, so you decide to keep track with the 0—1—2—3—4 system and your fist. With each change of direction you update e.g. "0.6, 1.6, 1.0, 0.9..."
For a **back bearing** — when you turn around — **add or subtract 2**.
Add 1.0 when you **turn with** the Sun (**left**), or **subtract 1.0** when you **turn against** the Sun (**right**). (S.H.)

Global

Example: Point to Tahiti

It's 9 a.m. summer-time, the Sun is, say, 20° high and over Tahiti (see time-zone chart).
Firstly, point horizontally toward the Sun (not directly at the Sun!) — that is the direct path to Tahiti.
Secondly, point below the horizon, to locate the actual direction to Tahiti…
Step1. Make a perpendicular from a sunbeam. Perpendicular to the Sun is 'horizontal' in Tahiti…
Step2. Subtract its (present) height (20°) from 90°
The arc from you to Tahiti is 70°
Explanation:
(90° - the height of the Sun), the 'co-height', must be the arc-distance from you to underneath the Sun. Think of it like this...
If the Sun is 0° away from you, it will be 90° high — you are underneath it.
If the Sun is 10° away from you, it will be 80° high, overhead a point 10° away from you, and so on.
Step3. halve the remainder (70°/2=35°) Now add back on the 20° above the horizon…
"Tahiti is 55° below the Sun" from your point of view. You didn't need a globe or the arc-distance to find it.

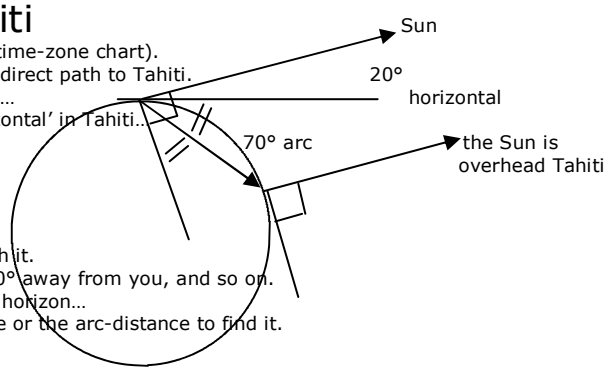

Sun

The Sun Oscillates Seasonally between +/- 23½° Latitude

To approximate the Sun's "Declination"—or North Latitude—you approximate its variation as sinusoidal.
Using a 365 day cycle: You approximate 360° as 1° per day, so for ease and accuracy of calculation you can and should approximate the day-number as 30 days per completed month + day of the month.
E.g. **March 21st** is 2 lots of 30 days plus 21 = "81" days or degrees.

Then to use a sinusoidal curve, you need to run from Equinox to Equinox — i.e. starting at March 21, Day 81
Declination = 23.45° x sin (day number of the year – 81)
E.g. 0° in **September 21**, six months or 180 days or degrees later; uses day no. 261.
From solstice to solstice you can use a minus cosine curve...
Declination = 23.45° x –cos (day number of the year + 9)
..since the solstice is around 10 days earlier than the New Year.
Dec 21 solstice, is day no. 11x30 + 21 = 351 "–23°½ in December"
E.g. + 23½ degrees on **June 21**, day no. 171.
From this approximation you can predict noon-height (and, with the right formulae, sunrise & sunset time & direction and E./W. height, of the Sun).

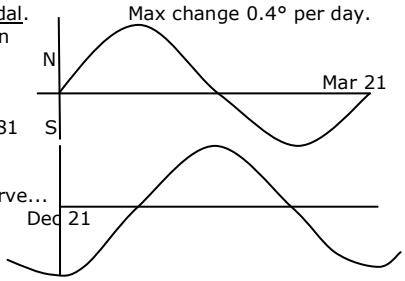

Stars

Use the Southern Cross as an Hour Hand

You can visualise the rotation of the Southern Cross as a huge 24-hour clock hand.
It is correct on April Fool's Day i.e. upright at midnight (at 183°T), then it gains two hours a month, or 30°. (Use the correct vertical! — South.)
As a learning aid, mark the 'Southern Cross time' on your calendar every month
e.g. "two hours fast" at the start of May; and "four hours fast" at the end.
Then, anytime you are out at night, you can ask: **"What's the time — by the Cross?"**

Picture the Southern Cross during the Day

Yes it is there behind the blue sky… "If there were an eclipse now, it would show… there!"
When you are navigating by the 'spin' of the Earth, and by the Big Picture, the proper reference direction is the North-South rotation axis,
Which means: Polaris + South Pole Star.
In practice, what we Southerners see, is the Southern Cross.
So if you are imagining: sunrise — noon — set, or Moon position, or a star-compass course, or a shadow-south direction,
you ought to be thinking: "North — Polaris down there; South — South Pole Star up there", then you can picture the Southern Cross as well.

Moon

The Moon is Lit By the Sun

When the Moon is **opposite** the Sun, we must see its fully lit face… See diagram of the first man looking at the Moon, below.
e.g. a full Moon rising at Sunset; Setting at Sunrise; Overhead at Midnight or at least highest in the sky.
(A fortnight later…)
When the Moon is **near** the Sun, we see only its dark backside or at most a thin crescent… See the middle man below.
e.g. a crescent Moon is only ever seen near sunrise or sunset. Wait until the Sun is below the horizon — Don't look at the Sun!
At other times the Sun has set, taking the Moon with it, or has risen, outshining the pale crescent nearby.
(A week later…) When the Moon is **halfway** between, we see half of each side… See the third figure...
e.g. half a disk overhead — only at sunrise or sunset; half a disk setting — only at midday or midnight.
A half full Moon is misleadingly called a Quarter Moon… since it takes four weeks to move between the four "quarters".

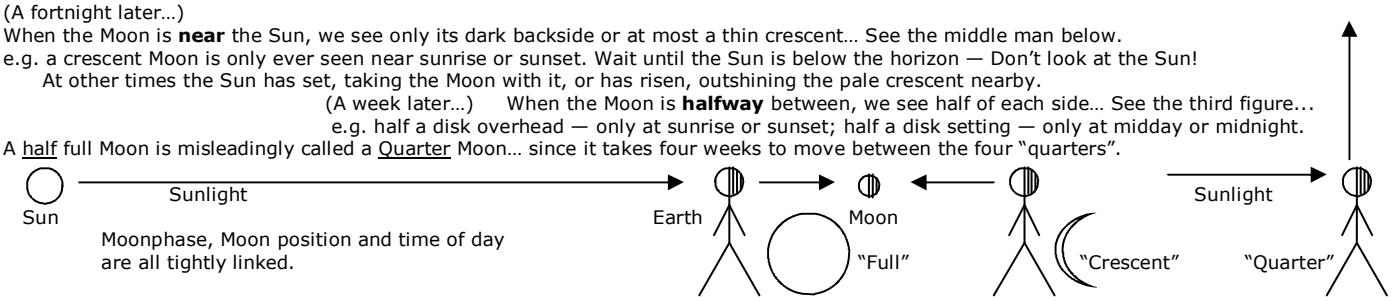

Moonphase, Moon position and time of day are all tightly linked.

Clockwise

Back-Bearing: **Plus or Minus 200° Then Minus or Plus 20°**

(Or — for the very advanced sleepy student: adjust by 20° first (+/-20°) then -/+200°)

Examples: 10° is small, so make it big, 210°, adjust back to 190°
290° is big, so make it small, 090°, then back up to 110°
Advanced: 190° is middling, so boost it first, by 20°, to 210°, then remove 200° and use 10°
210° is middling, so try +20°, then –200° = 030°
When you end up with numbers greater than 360°, remove 360°. When less than zero, add 360°.
Difficult: -10° is small; first convert it, using +360° (= 350°)
350° is big so –200° +20° = 170°
185° + 200° -20° = 365°, so adjust it back to 005°

There is another solution to back-bearings — decimal degrees — but it hasn't taken on yet.

a simpler solution with a protractor compass, is to...

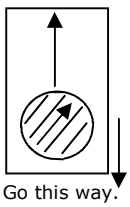

(see page 109b)

Use **Latitude** for a **Distance** Scale (Not 'Long'-itude, Despite the Name)

(Especially on marine charts, atlases and old maps—Newer maps are likely to have the mostly meaningless but well-scaled international map grid.)
Latitude means "distance" from the Equator. Remember the polar "distance" 90° complement to latitude? which measured arc-distance, in degrees!
1° of latitude = 60 nautical miles = 111 km The world is 360° around; or 360x60 = 21,600 nautical miles; or 40,000 km
The circumference of the Earth determines how long a kilometre or a nautical mile is — at least in the somewhat outdated original definitions.
(The N/S circumference of the Earth has proven to be closer to 40,007 km, and the E/W circumference is bloated by the Equatorial bulge.)
One 'kilometre', comes from the decimal system: whereby one circle equals 400 'grads'; one grad has 100 divisions… 40,000 km
Use the latitude scale on the left & right sides of a Mercator map to measure length or distance — but only at that latitude on the map.
But do not use the longitude scales running across the top and bottom of the map — since the spacing of longitude lines should vary with latitude.
Longitude lines must converge towards the Poles, to nothing. They span a variable distance = 1° x cos (latitude), but a Mercator map will (falsely) show a constant separation for longitude lines (this is to make compass courses look straight). A Mercator projection will also show a variable separation for latitude parallels, even though they 'should' be constant. This is to preserve the shape of land areas—The latitude stretching must match the longitude stretching. Therefore any scaling shown on the E/W spacing is only valid for the latitude at which it is subdivided.
When a latitude scale says ½° or 30 n ml or 55 km, it means it, at that latitude, but when a longitude scale says ½°, the distance meaning is missing

Earth

Using the Declination of the Sun

- **Noon Height** = 90° – latitude + declination = (from horizontal) Convert "N." to + ; "S." to – e.g. 30° S. = –30
  E.g. Lat 30° S.; Decl. 20° N : Height = 90+30+20=140° or 40° (the two (opposite) heights must add up to 180°)
  i.e. The more the Sun 'declines' away from the Equator, and the more you decline on the other side of the Equator, the more the Sun drops in the sky. Alternatively, the colatitude (90–L) is the maximum height of the sky Equator... Then adjust for the declination.
- **Sunrise/Sunset** — time and direction prediction
  Cosine (angle from North/South) = sine (declination) / cosine (latitude) Try dec =23°.5; lat =37°.1; 55°.67; 62°.6, for a different hexagon
  Cosine (time in degrees from noon) = –Tan (Lat.) x Tan (dec.) 1° of longitude = 4 minutes of time.
  E.g. measure the time between sunrise & set; divide by 2; solve for Lat. (make sure to use horizontal, not horizon.)
- **East/West Sun** — Time and height prediction
  Cosine (E-W time in degrees, from noon) = Tan (dec.) / Tan (Lat.) e.g. 4 hrs each side of noon @ Lat 41° (Northern Tasmanian Summer)
  Sine (E-W height above horizontal) = sin (dec.) / sin (Lat.) e.g. 45° high E./W. from Lat 34 1/3°; +/– 3h21m; Mid-Summer

e.g. take the time when the Sun is E. & W. (opposites — you need to remember to carry 60, not 10, & to interpolate carefully) and then solve for Lat.
See p. 86b for a memorable laugh. Many more formulae and techniques are available, from books on astro-navigation.

Sun

Perspective Distortion — Use the Correct Vertical

Vertical is vertical, right? Wrong! To tell the time by the sky, take care.
The higher a pair of points (e.g. stars) gets in the sky, the more you have to choose which vertical is needed for a reference, to judge 'vertical' by.
Each different reference direction will give a different answer to "How vertical are those two points up there?" e.g. Moon cusps, Southern Cross
A side view makes this clearer....
..a straight line which looks horizontal when looking south,
looks diagonal in the East

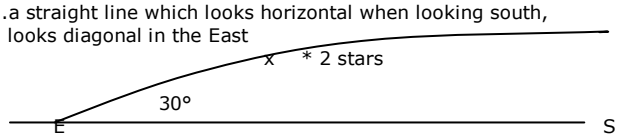

The top view also makes it very clear… a line XY high up in the sky cuts each different vertical at a different angle..

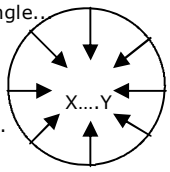

Don't gauge the angle of rising/setting (p9a) at anywhere other than at the rise/set horizon in the rise/set direction.
Don't use the nearest horizon, as your reference for which way is up.
To visualise a constant 'rotation' of the stars, use Polaris or the South Pole Star as the 'centre' of rotation.
I.e. Use N-S longitude lines in the sky rotating in relation to vertical at North/South.

Stars

To Determine Upper Wind Direction from the Clouds

Walk in the reverse direction, while keeping a chosen cloud lined up
on a chosen high point. You must walk across level ground to be accurate.
Mark your starting point, then drag a line in the ground to mark your walking direction.
This method works well for very slowly moving clouds.

Other methods: Watch cloud shadows.
Use the Moon, Venus or the Sun (only when it is very well obscured).
Look in different directions at the moving clouds
until their motion seems neutralised—they are moving directly toward you.
Use binoculars, to magnify the movement (but never look towards the Sun).

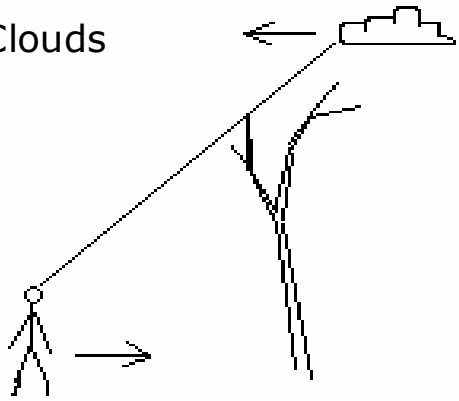

Wind

The Sunny Skies Quadrant System

To orient yourself on a clear day...
Use the two ends of the dark blue band — where it hits the horizon — and the trend of the middle bit of the band,
in conjunction with the Sun (but don't look directly at the Sun!) and your shadow.

- No matter which way you are facing, you can detect one or two signs to orient yourself by,
  Or at least imagine them if you can't see the Sun just now.
  You don't have to look over your shoulder, which is misleading anyway.

The middle of the blue band is above your shadow, and at right angles to the sunbeams near you.
The whole system rotates, during the day.
If the **Moon** is up, add it into the system,
and **Moon-North** too, which is off to one side of North, and changes slowly during the day.
Combine this system with the fixed systems, of 0, 1, 2, 3, 4 and m.n., rise, a.m., m.d., p.m., set.
E.g. In the diagram, the Sun is to the left of North, so "the Sun is 'afternoon', the Moon is 'before noon'."

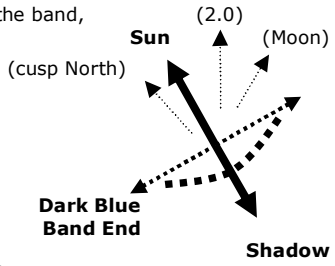

Anticlockwise

Global

The Horizon Rotates to Your Right in the Southern Hemisphere

1.Face the High Noon Sun. (Imagine it, if you have to, when it is not midday.) It moves horizontally from Right to Left (S. H.)
2.Reverse your perspective — let the Sun stand still and the horizon move, from Left to Right.
Imagine the horizon sliding horizontally beneath the Sun (and you shifting sideways with it).
3.Now imagine the Sun rising from the eastern horizon — it moves up to your left — then reverse perspective...
Imagine the eastern horizon moving *diagonally down and away* from the Sun, twisting down on the right as it goes.
Face North. Grip the whole horizon on the East and on the West, and imagine a tilted steering-wheel, and rotate it around the Pole Stars. Imagine a "steer right" motion.
The invisible steering wheel faces you diagonally upwards, while the horizon, as if fixed (diagonally) to the steering wheel, starts off horizontal and then rotates out of its horizontal.
4. Keep rotating the horizon, sliding the northern horizon horizontally underneath the high noon Sun, and beyond, until the western horizon comes up diagonally to meet your now "setting" Sun.
You need to do this outside to appreciate it in three dimensions, to have any chance of visualising it easily.
So put your clock back on the Northern wall (see 5e) and say "See, we Southerners have it right, after all!
— the real movement — of the solid stuff — is clockwise"

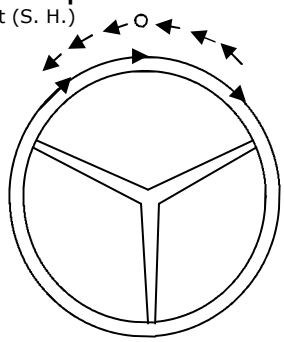

Sun

Tell the Time by the Compass Protractor — Tilt It Toward the Equator

Take a **tilted bearing** of the Sun. The answer comes back as so many degrees to one side of noon-North (or noon-South in the N. H.).
You then use 15° per hour to convert that to hours before or after noon-time ('noon' will be nearer one o'clock during daylight savings time).

Step one: align the compass to True **North**
Step two: **tilt** it upward to the 'Equator' in the sky (the Sun's average path)
Step three: rotate the compass arrow to point toward the Sun
Step four: make the edge of the compass cast no shadow
This avoids looking at the Sun
Step five: measure the degrees away from high-noon North, on a tilt

It is more accurate than using a finger-span to measure 15° per hour
i.e. Use the compass protractor to make a 24-hour clock face
To use the compass protractor as a 12-hr clockface, see the box opposite.

1. Top View 2. Side View

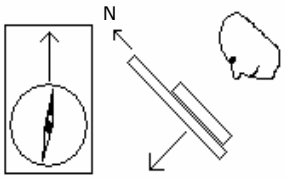

The reason for tilting is because a level sundial/upright gnomon will not give a 15°/hr linear reading (See "The sun passes quickly")

Stars

The Southwest Corner of the Great Square is N. S. E. W.

The Great Square of Pegasus only has one really good right angle but it shows N. S. E. W. perfectly.
In the Southern Hemisphere, it is the first star of the Square to rise.

Note: When you hold a star chart over your head 'East and West' reverse!

This is the view for the Southern Hemisphere…

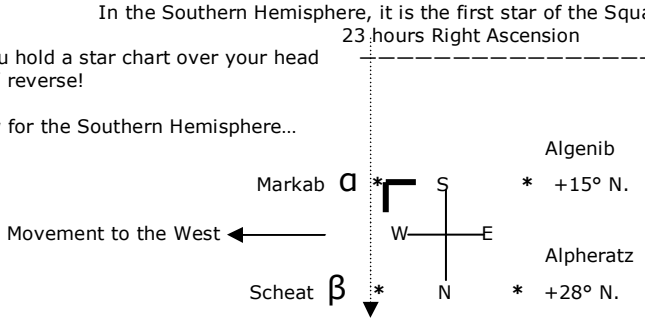

(Southern Hemisphere view)

Moon

The Moon Rises and Sets 50 Minutes Later Each Night

The Moon moves quickly to the West, but…
It progresses Eastward through the night skies, night by night, by 12½ degrees a day — it 'lags' the Sun or 'moves backwards' during the month.
So that after 29½ days it is back in the same place. It moves about 27 diameters' worth East through the stars, per 24 hours or about a diameter an hour to the east in relation to the Sun. One 'diameter' is half a degree, and the same as the Sun's.
Sun, Moon, Planets and Stars rise in the Eastern half and set in the Western half, but some are quicker than others. The Moon is the slowest of all.
E.g. (twice-daily) tides get 25 minutes later each tide, on average — half a day + 25 mins apart.
E.g. the 12 month lunar year gains 11 days on the 12-calendar-month year, or "one day per month except for February, on average".
(a calendar month equals 30.4 days; a lunar year equals 354 days. Months 1,3,5,7,8,10,12 have 31 days—1½ extra; ½ extra in 30 day months)
The explanation for the Moon's "slowness" in rotating westwards with all the rest of the sky, i.e. from moonrise to moonset (it rotates only 14½ degrees per hour, not 15° like the Sun) is that it slowly revolves around the Earth to the East, in the same clock-sense as the Earth revolves around the Sun,
and so must complete 13 revolutions star-wise, to the East while we see only 12 Moon months.
By the next full Moon, its backside is toward a more easterly star…
For an anti-clockwise rotation system, viewed from the North...

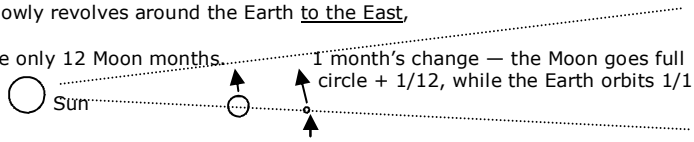

1 month's change — the Moon goes full circle + 1/12, while the Earth orbits 1/12.

Clockwise

Turn 90°? "+/–100° –/+10°" !

Turning right is turning clockwise; turning left is turning anti clockwise.
Clockwise **increases** the count of degrees East of North. So that requires **+100°** -10°. E.g. 257° changes into 347°
Anticlockwise, when turning left needs **–100°** +10°. E.g. 257° changes into 167°
Or use +/-10° first, then -/+100° if that is easier (near 0°). Adjust by 360° if necessary.
There is no convenient choice as there was with a back-bearing, so you must remember whether to use + or – 100°.
Remember to look at the Sun's clock-sense to judge 'anti-clockwise'
— anticlockwise in the southern hemisphere — to your left. 'With' the Sun implies – 90°
— clockwise — to your right — ('With' the northern hemisphere Sun) implies + 90°.

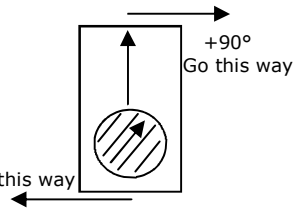

Hint: When sailing, you can always temporarily head either towards or away from, or side-on to, on the left or the right, a landmark from which you want to take a bearing, without having to change tack. Therefore you can use +/–180° or +/–90° from the boat's compass-heading. In heavy weather this means that you do not need to leave the helm to use a hand-bearing compass, but merely sight along the lines of the boat, or cross-wise along the lines of the cockpit, stern or cabin. Simply remember that the more clockwise heading is bigger than the left-hand direction

Global Perspective — Rotation of the Horizon

Each point on the horizon rotates…

- In a circle
- Around in an Eastwardly sense, like all the real matter in our anticlockwise spiral galaxy
- Around the North-South Earth rotation axis
- Along a fixed sky-latitude circle (i.e. declination circle)
- At a fixed angle of rising/setting (one or the other) in relation to the fixed stars
- Once per day

Example: The "Sugarloafs" were on the left of the Sun at sunrise, but six hours later they had moved to be underneath it, by rotating around Polaris and now, at the end of the day, they have moved way over to its right, and will keep going around for the next 12 hours of night-time.

To further help you visualise the motion, remember that the ground under your feet is twisting you clockwise in relation to the sky, in the SH, so that the stars, etc, near the horizon all move only to the left as they rise or fall. For the Northern Hemisphere use "anti-clockwise… right".

Or Judge the Time By Sight — by Mirroring the Sun Around North (To Make an Hour Hand on a 12 Hr Clockface)

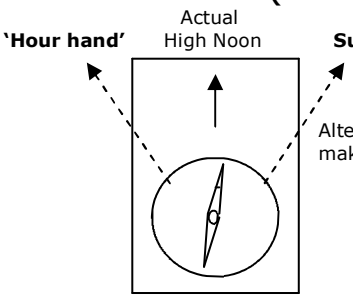

Sun = 'Noon' position <u>on the Clockface</u> (i.e. ignore real Noon-North after you align the compass properly) (SH)
(Here, the 'hour hand' is showing 2½ hours until 'noon', judging it by sight)
(For the Northern Hemisphere, you have to reverse things… hour hand — Sun — South see p.28b)
Of course this method is really just helping you to GUESS!

Alternative: Line up a pencil tip with Polaris, while making the pencil cast no shadow. With your shadow-side hand, make your index finger ("The Hour Hand") touch the pencil tip, and look identical on the <u>other</u> side of Polaris.
You can make it work on the Moon or a star or a planet if you know when they are North/South of you.
You can make it work for magnetic noon, to avoid using magnetic variation.
It works because the Sun moves at half the speed of an hour hand,
but the <u>doubled</u>-angle shown, <u>doubles</u> that rate of change, to match a clock's hour-hand.
Don't forget to **tilt the compass**, so that its axis lines up with the Earth axis.

The Southern Summer Rectangle is Worth Looking for — since it has four N. S. E. W. guides

1. North: from Procyon, down through the Twins
2. North: from Bellatrix down through Nath
3. West to East: between Bellatrix and Betelgeuse to between Procyon and Gomeissa
4. West to East: Nath to between Pollux and Castor, closer to Pollux

(To remember the two Twins, say "Castor & Pollux, α & β, are North & South")

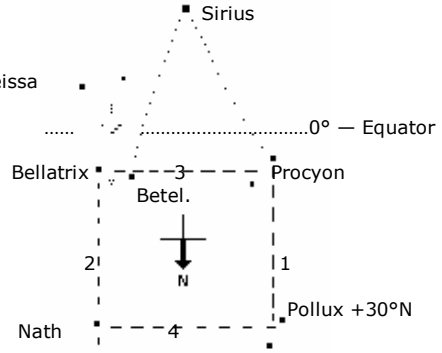

And a Triangle Pointing South

Betelgeuse, Procyon and Sirius.

Clouds and Con-trails and Their Winds

- Fog, which is cloud clinging to the ground, usually oozes downhill. Now, working upwards...
- Topographic cloud banks — may indicate e.g. the plateau edge or a mountain range. I.e. They cling to the topography.
  Also look for eagles soaring along otherwise invisible cliff edges.
- Low clouds — they follow the weather map isobars. Also look for sideslip of flying birds and aeroplanes.
  As a rule of thumb (clenched fist!) low cloud puffs are at least three knuckles wide, alto puffs are two, and cirro puffs are one, from sea level
- Middle ('alto') cloud — the puffs look smaller, the shading, less dense, but no 'iciness' seems apparent. The clouds obscure the Sun's outline somewhat. Their wind-direction is usually different from both higher and lower clouds. This wind will 'steer' low level pressure systems.
  (An 850 mb level chart helps) When a change is due, the middle clouds often betray what low wind direction to expect soon.
- High ('Cirrus') cloud — thinner; smaller puffs if puffy; icier i.e. fleecier, silky, filamentous; transparent, with halos & no shading.
  The Sun's outline remains crisp when seen through cirrus. Don't focus <u>on</u> the Sun!
  The cirrus-level cloud-direction in the subtropics is usually westerly +/–45°.
  Be sure to test this in <u>your</u> latitude. (A 500 mb chart helps) Try to catch glimpses of upper clouds through gaps in the low clouds.
- Jet Con(-densation) Trails — are usually in the high band. Look for them according to a timetable, each day.

The "SWEETHEART" Method of Orientation

We are <u>human</u>, and 'reference directions' are not. So <u>make</u> them more human, more personal, more emotional.
"Home" is not bad. "Sweetheart" is better.
Find something lovely about each cardinal direction. E.g. 'our favourite holiday spot on the coast'
Something you love or hate, want, or fear, about each reference direction…

You will find it easier to maintain awareness of bends
and to interpret meaning onto the skyline — what lies beyond and behind that horizon!

During one walk, these highlights are unlikely to change direction much.

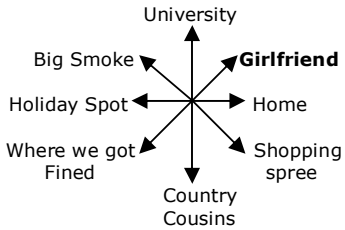

Global

How to Shift Your Horizon — Its Timing and Height in relation to the sky.

You can navigate by the "shape" of the sky, back to a spot where you have been before, by noticing the "position" of the moving horizon.
On a round earth you can shift your position and horizon 1°, or 360°, by travelling 1°, or 360°, around the circumference. (111 km, or 40,000 km)
The line where the horizon cuts across the sky, varies with the latitude & longitude & the time of day. Using the Sun as an example…
Note: The "1° for 1°" rule for your own movement, only applies after you neutralise the other motions, so as to be as if the sky is standing still.
E.g. You need to neutralise the Sun's daily motions, by looking at the same time of day — sunrise, noon, 6 a.m., or 12:22 a.m., or whatever.
E.g. You also need to neutralise the Sun's seasonal motion by looking at the same date every year — Christmas morning, June 25, or whatever.
(Leap years can only make a difference of up to 1° — you could improve accuracy by looking only every four years.)
Rule one: for West/East travel, remember that events like Moonrise seem to happen **later to an observer further west**, and…
Earlier in the **E**ast; **four minutes of time (early) for every 1° of longitude (East).** (these rules work in both hemispheres)
E.g. "Christmas Day Sunrise at camp 4 was at 6 a.m. last year, but now I see the Sun already up at 6 a.m., so to get to where we spent Christmas last year, we would have to go further west to make the Sun lower (the horizon higher) at 6 a.m." (Assume you are at the correct latitude)
Rule two: The horizon drops to reveal new sky as you travel onwards — **1° for every 1° of travel** — this works in any direction.
E.g. "The Sun rose to 75° high, but here it reaches only 74½ degrees — so I would need to travel a half-degree toward the Equator for it to be 75° again." (This example has deliberately chosen a North South change.)
Then you can shift your horizon, its timing and its height, by moving around the earth, in order to match a previous timed snapshot of the sky.

Sun

Walking in Phase with the Sun, Wind and Moon

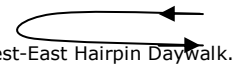
Simple West-East Hairpin DayWalk.

- In a hot climate, walking into the **Sun** is unpleasant, hot and glary, especially in the afternoon, especially if you wear glasses, so...
- The morning **wind** can be very uncomfortably cold, and it is best to plan to walk away from it, too.
- Beach walks are more pleasant when you don't have to battle back into the afternoon sea breeze or prevailing wind; so assess the **forecast** before you choose a start point.
- Night walks and late camps are best done in bright **moonlight** — simply predict the phase.
- Spotlighting is the easiest on a waning Moon — since the **dark period** is in the evening, not before dawn.
- Spectacular pink full-Moons only rise at sunset, so it is best to predict such a **moonrise** and plan an early camp with a view.

E.g. Up in the afternoon **shade**; down in the morning warmth…
Similar considerations apply to sailing—e.g. aim to make landfall in the pre-dawn. You can identify the harbour, or the coast, by its light-houses, then soon after be able to enter the harbour in daylight.
E.g. Aim never to cross the sandbar on an outgoing tide—with steep breaking waves, in shallow water, getting shallower!
Coordinate Your Route, Pace, Campsites & Timetable To the Prevailing Conditions.

Stars

Navigating by the Shape of the Night Sky

Since the star-map — the background to our view of the sky — is unchanging, the 'shape' of the sky involves only the **position of the horizon** across the background of fixed stars, and the **rotation** of that horizon with time.
The sky gives a false impression that the sky itself is spinning, but it is we who are turning upside down! It is our horizon which is changing.

Just as you can neutralise the Sun's motion by looking at the *same time every day* (which will show up only its seasonal and longer term variations) and just as you could neutralise the Moon's motion, by looking at the same phase and Moontime every month (e.g. full moon rise), to show up its unpredictable motion, so too with the stars. You must **look at the same star-time every night**. If, in addition, you look at the stars from the same spot you will notice no change at all in the 'shape of the night sky' or the position of the horizon — star motion is all predictable, with no significant variations.
This makes the stars excellent for navigation — you simply locate the horizon in terms of the heights of the stars, the tilts of star groups, and the star-time of night. E.g. what star appears overhead and at what star time.
Thus, to find that spot again, even much later, you need to make your horizon look identical at the identical star time.
CAUTION: Don't go star-navigating; rather only imagine doing so, until you become an expert. Understanding it is more useful than doing it!
All figures, formulae and theories should be checked before you trust your safety to them.

Moon

The Moon is "in Phase" with the Sun — Watch for It

The Moon progresses (12½ degrees per day) contrary to the general rotation, so…
A **new** Moon, first seen after sunset, must grow from a crescent, in the **after sunset** part of the sky, as it separates from the Sun, Eastwards, night by night, more and more so after the new Moon. I.e. a 'waxing' (growing) Moon means Moonlit evenings. After full, the Moon continues Eastwards…
A **full** Moon must diminish, to a crescent, in the **before sunrise** time of night. I.e. a 'waning' Moon leaves the evenings dark, until Moonrise, predawn.
In other words, the predawns are moonlit for the second half of the Moon month, after Moonrise. In the first half, the predawns are dark, after Moonset.
The whole night is dark in the 'no Moon' changeover of months.
During the day, you ask "Is the Moon ahead of or behind the Sun?" I.e. Is the Moon predawn (waning) or post-sunset (waxing)?
Evening spotlighting — is best in the dark before the waning Moon rises — late in the Moon month.
Night-walking—is best in Moonlight after Sunset i.e. on a full Moon, or a waxing, almost full Moon. (early-mid month)
(Think of the Moon as prouder of its nighttime role as it moves away, and ashamed as the Sun gets closer)
Highest and lowest ('Spring') tides occur on or soon after a new or full Moon.
'Neap' tides occur (when the Sun and Moon pull at right-angles to each other) at "half full Moon".
(Shallow waters introduce a lag of a day or three, due to friction and resonance effects.)

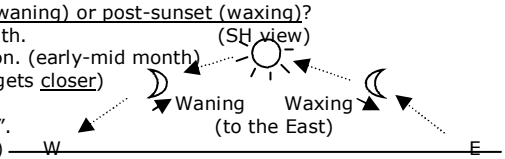

Make it your aim to see these things all happen in phase with the Moon month.

Clockwise

The Clock-Face Method for Direction

Quicker than 360° — it only counts up to 12! — it communicates relative directions easily.
E.g. "3 o'clock high" = "on our right — up in the sky" "9:30 from Rigel" = to the left and just above (that star)
Or, for horizontal directions, " 5 o'clock from the Sun " = looking down-Sun, and slightly to the left "six o'clock true" = South
At 30° per "hour" it is easily compatible with 360° — both are Babylonian systems.
(But don't confuse it with a 24-hour clock face, which is only 15° per hour. Unfortunately our clocks do not use a 24-hour system, which would make them nicely travel around once per day — as fast as the Sun.)
For shifting between clock face directions and 360° directions…
— First convert to or from the hour: Multiply hours by 30, for degrees. Use 12hrs=0 : 0 30 60 **90**; 120 150 **180**; 210 240 **270**; 300 330 **360**
— "5° extra" = "add 10 minutes" e.g. 255° = 0830 hrs (240° = eight hours; 15° = +10+10+10)
— "1° extra" = "add 2 minutes" e.g. 258° = 0836 hrs
When using the clock-face for compass directions, for NE, SE, SW, NW use 1:30, 4:30, 7:30, 10:30.
Translate 1, 2 o'clock as "(roughly) NNE, ENE," etc.
Don't use the minute marks on the clock-face because every 6° is not very user-friendly

Latitude without Numbers

When all you want is the latitude, the shape of the sky can tell you, without numbers, without calculations, without a watch.
E.g. 'Canopus overhead', or 'the Pot sets at 30° off vertical'.

If all you want is to recognise appearances locally, you don't need to set your Star watch to Greenwich Time, or to know a star's longitude relative to Greenwich.
E.g. 'I set my watch to the Dog Star passing North, from home last month'.

If you don't want calculators and trigonometry, you don't need to use them.
E.g. 'The sky is behaving the same as last year, we must be within sight.'

I am about to give you the star time and star longitude information, only for those who wish to work with the easiest benefits from them.
E.g. 'That must mean that our longitude is 150° E.'

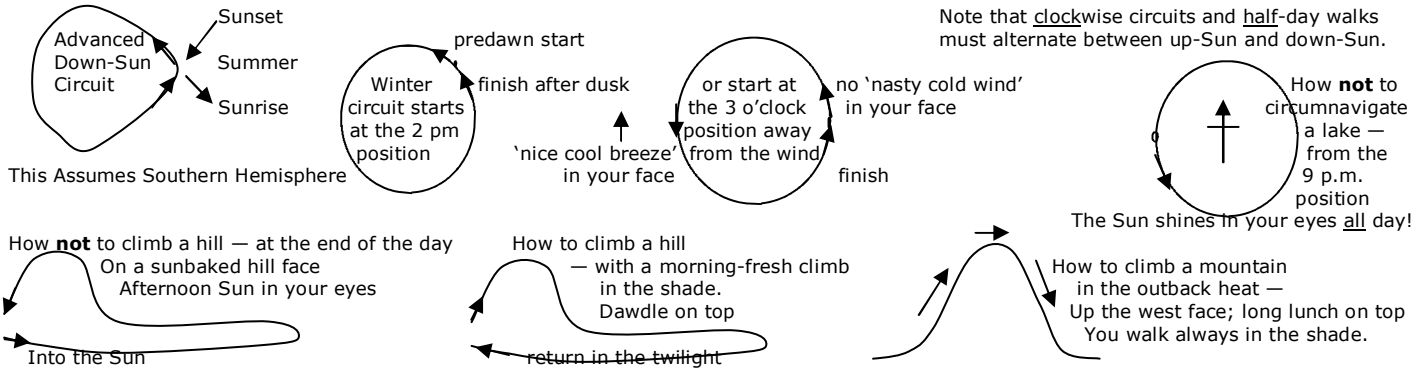

Matching Sky to Sky — Move Only Your Horizon

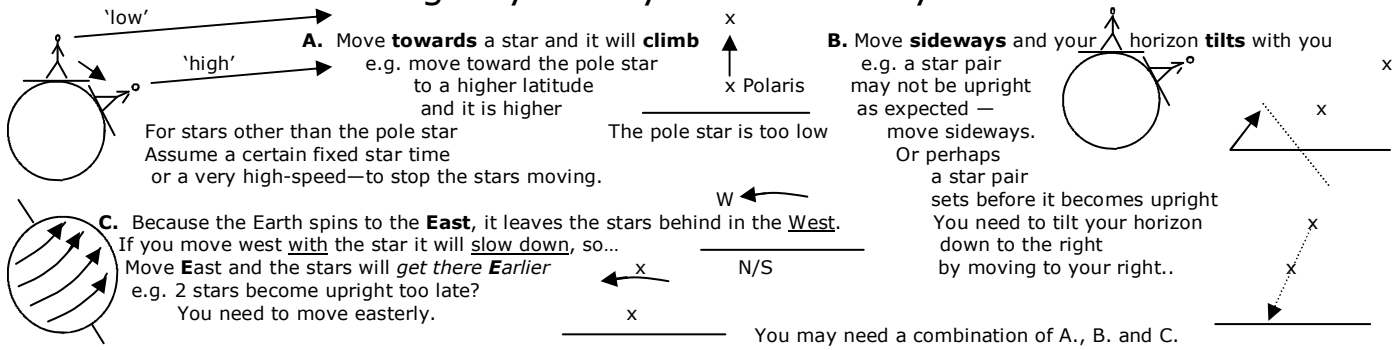

Prevailing Winds on Creek Banks

One of the two Creek banks will often have a mud- or rock- cliff due to prevailing wave-action, with the other bank being a mud flat.
E.g. If prevailing winds are westerly, the west bank will show less erosion.
For more on finding the prevailing wind direction when the wind is not blowing, see section 3: "Nature's Norths".
Because prevailing wind is almost continental in scale, navigating by it is a sort of Big Picture technique.

Tips for calm weather

1. Listen for the wind! Sounds will carry better downwind e.g. surf or traffic. Remember this, at night when all else is still and quiet.
2. Watch the angle of mist droplets falling. (Stop still for this one) Or smoke rising.
3. Huff into the air to make a fog.
4. Watch the flimsiest vegetation.
5. Look at the leaves which dangle down from the topmost leaf clusters of a tree.

Mnemonic: "North; Coast"

That's what you should say, whenever you say 'North'.
"North (Coast) is over there" — it reminds you to think,
or rather to picture, the invisible picture.
You should make it your business to name the North-Coast town north of you,
before you go, by consulting a roadmap
Then of course, there is the East Coast, and the South Coast and the West Coast.
Next, of course, the same must go for "North-West"
— think of the North-West Coast township, and so on.
You are not functioning properly unless you mean something when you talk!

Go Overseas!

This is a one-off effort, so go to the trouble of getting a globe,
to sort out what is truly "East" and "West" of you,
in a Great Circle straight line.
It will surprise you what East and West should mean to you.

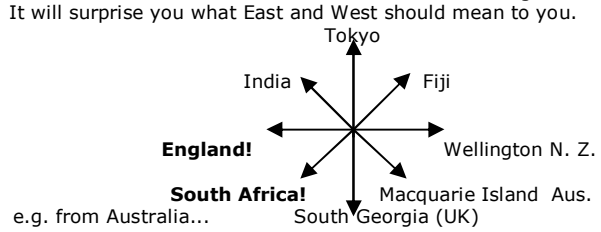

Global

Mnemonic "The Sun Travels the Tropics"

I.e. Imagine the **tropical countries and oceans** which are **'underneath' the Sun** as it moves East to West during the day — The Equatorial band.
For Australia that means from Central America Across the Pacific Ocean Past New Guinea
On to the Indian Ocean And on toward Africa
When we in Australia see the Sun rise, it is past noon in the Americas. At our sunset, it is approaching noon in Africa.
See the Reference List of Tropical Places (for The Eastern Australian Time Zone) opposite. As a help to global orientation, make it your aim to memorise these time zones in sequence. Check it out with an atlas. See the tear-out clockface diagram in the appendix.
The same sequence holds for any planet, or star near the Equator e.g. for the stars listed 'opposite the Sun' (below), but with different timing.
You can also imagine which country the Sun is overhead while it is night-time near you,
or which country the star opposite the Sun is overhead while it is daytime near you.
Now, with this in mind name the general North/South directions, according to the height of the Sun: To the Sun = "Hot"; To noon = "Hottest";
and the Shadow direction = "Cold"; Shortest shadow = "Coldest".
•Although the Moon, by comparison with the Sun, has a "cold" feel to it, make sure you apply the same list of hot tropical countries to it also.
When you see any Zodiacal body, you should think: "That is a Hot direction; It is over a Hot country." and try to name the country.

Sun

Tell the Time of Night by the Star **Opposite the Sun**

Such a star will be **up all night** — whether the night is long in winter, or short in summer — look for it. Write them on your calendar.
This yields the solar time, i.e. with respect to local noontime, but not the daylight-savings or summer-time, so adjust by one hour in summer.

January	Sirius	*Castor, Pollux*, Procyon					
February	Suhail, Miaplacidus	Alphard	*Regulus*				
March	Pointers (Dubhe-Merak)	*Denebola*	Phecda	Southern Cross			
April	Southern Cross	Alioth	*Spica*	Arcturus, Pointers (R.K./Hadar)			
May	*Zubenelgenubi*, Kochab	Alphecca	*Antares*				
June	*Antares*	Atria	Rasalhague	Scorpion Sting	Kaus Australis	Vega	
July	Vega	*Nunki*	Altair (Aquila)	Peacock	Deneb		
August	Deneb	Enif (ε Pegasus)	Alnair				
September	Fomalhaut	Middle of Great Square					
October	Diphda	Mirach	Achernar	Hamal-Almach			
November	Mirfak						
December	*Alderbaran*	Capella-Rigel	Saucepan	Six O'clock Stars; Canopus			

This list is arranged time-wise
the earlier stars are correct
at the beginning of the month,
the last stars, near the end.
Use each as a 24 hour-hand around the South Pole Star
The Zodiacal stars
(the ones on the Ecliptic)
are in italics.
Only they are truly opposite the Sun.
This chart is **true for**
both Hemispheres, permanently.

Stars

Two Different Back-and-Forth-Star-Directions Give You a Position Fix

..Yes, it's just like two bearings (which are not opposite each other) from two landmarks. But because these 'landmarks' are so distant,
the accuracy is only to within seeing distance. For better accuracy you would need a theodolite or sextant.
E.g. when two stars look upright, that locates a global star-direction; a second upright pair gives you the second direction, and a position fix.
To fix your latitude. you have to **measure the time-lapse between the two star pairs**. (Simultaneous is simplest.) So you need a plumb-line and also a watch. And for accurate (Greenwich) longitude fixing, you need to work in (Greenwich) **star-time** — otherwise you have only latitude which is fixed. So you need either an almanac, or calculator, or star-watch, or a work-it-out-at-home-later approach.
This is equivalent to fixing which star is exactly overhead and when. To see this, look up past the top of the plumb-line to 'overhead' from the two different directions—one for each pair—to see that they must cross exactly overhead.
The 'fix' is repeated accurately over years, and every night at the same star-time, with some very small predictable drift in location and timing.
No one else on earth can see exactly what you see right now. That's what gives you a unique position fix.
People at the same latitude see something very similar earlier or later than you do.
The best two directions to cross are Northwest/Southeast and Northeast/Southwest. Top-view...
This is because East/West star directions are time-wastingly difficult to determine.
This method is most helpful, of course, at sea, or in the desert, or as a hobby, or when lost.

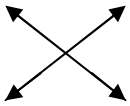

Moon

Predict the Full Moon Path According to the Time of the Year

The full Moon is **opposite the Sun**, and is near the **Ecliptic**, i.e. in the Zodiac (the Sun's path through the stars) so that…
When the Sun is in midsummer position—high, and up for a long time—the full Moon will be in a midwinter position—low, and setting again quickly
When the Sun is in the midwinter position—short days—the full Moon will be high and up for a long time — long nights.
The two complement each other, since when one comes up the other goes down, and when one goes down the other comes up.
Both Sun and Moon are on the Ecliptic—not usually on the Equator—but, at the Equinoxes, the Sun is on the Equator, so the full Moon is too.
Near Equinox, of course, the Sun and Full Moon will rise and set in almost the same positions — East and West — and follow similar paths — about 12 hours each, along the Equator. The Equator and the Ecliptic intersect in two opposite points — the equinoctial points.
In other words, you can **use the full Moon to visualise the Ecliptic**, and to guess where the Sun is, i.e. to guess the time of night,
Or you can use the Sun to visualise the Ecliptic, and to guess where the full Moon will be, how high, and for how long.

Finding North from the Moon, by its height & phase

The Moon will follow a Midsummer Sun's path, or a Midwinter Sun's path, or an average height path, according to where on the Ecliptic it lies.
Because the phase of the Moon is determined by its relative position to the Sun, the combination of the time of day, the phase, and cusp-Ecliptic gives you an unmistakable sense of where North should be, when you see the Moon. When you see the Moon surprisingly low in the sky, and the cusps confirm that the Ecliptic lies fairly horizontal, you know that it is somewhere near North (SH), following a low and horizontal trajectory, rather than a Moon 'just now rising' on a high trajectory, like a summer Sun in the East.

Clockwise

One Degree is about One in Sixty

1 in 60

And **n° is about n in 60**. e.g. **10° is 1 in 6** (p96a), and **6° is 1 in 10** (3 fingers wide, or 3½ knuckle-centres, at arm's length, or 'blink your eye').
These are handy rules of thumb for navigators. If you use an isosceles triangle, **the rule is remarkably accurate for most acute angles**.
Example: "Oh I forgot magnetic variation! I've been **5° off**!! Oh well, I will simply adjust sideways. Isosceles
I have come 12 km; 1/60 of that = 0.2; times five = **1 km off**"
Actually 1° is 1 in 57.3 (1.75:100), predictably slightly more than 1 in 60, but as the angle increases so does the accuracy of the 1:60 rule.
The rule (n° = n across at 60 away) is based on: the third side of the **isosceles triangle** (is twice the sine of half the angle).
(If you don't like numbers, try this other "handy" rule...) 6 (equilateral) lots of 60°
Rule of Finger: point your index finger at something — the fingernail is 1°.
Why is 1° slightly more than one in 60?... 60° r r r
Six equilateral triangles go all around the circle, but, as you know, 6 1/4 radii go around the circle.
This is because the circumference, longer than the hexagon, is 2pi times the radius, or 6.283r one radian=57.3° r 1c r
So one 'radian' of arc, which is one radius long, but curved, will be slightly less than 60° i.e. 57.3°.
So if the radius is 60, one sixtieth of a radian will be "one in 60", and slightly less than 1°. 1/60 of 1^c = 1 in 60 r 1^c
377 of them go around the circle. For 360 degrees around the circumference, the radius needs to be **57.3**.
The equivalent is '5/3 across at 100' (5 at 300, 167 at 1000; above 10°) with the more exact figure for up to 10° = 7/4 in 100

Tropical Countries and Places Under the Sun — Arranged in Time Zones Relative to Eastern Australian Time — 10 hrs ahead of Greenwich

-12hrs. The mid North Atlantic (is opposite Australia)
-11hrs. Brasilia 16° Sao Paulo/Río de Janeiro 23½° S.
-10 hrs. Trinidad and Tobago 10° N.
-9 Cuba/Haiti 20° N.; Bogotá 5° N.
— Panama 8½ hrs ahead of Australia
-8 Galápagos 0°; Guatemala 15° N.
-7 Gomez 26° S.; Mexico 23½° N.
-6 Pitcairn Island — Easter Island 25 to 27° S. (L.A. 34°N).
— Tip of Californian Peninsula 23½°N; IDR Gigego 20°N
-5 Mangareva Tropic of Capricorn
-4 Tahiti 17°½ S.
—Cook Islands 11° S.
-3 Hawaii Island, Tropic of Cancer
-2 Fiji 15° S.
-1 Noumea 21° S. See pp 132f

0 Rockhampton, Tropic of Capricorn 23½° S. (Canberra 36° S.)
1 Alice Springs, Tropic of Capricorn
2 Sulawesi 0°; Manila 15° N.; Port Hedland 20° S.
3 Christmas Island 10° S.; Jakarta 7° S.; Singapore 1°N.
— Cocos Islands
4 Calcutta Tropic of Cancer 23½° N.
5 Maldives Equator
6 Gulf of Oman Tropic of Cancer
— Seychelles
7 Port of Aden 12° N.; Mogadishu 3°; Madagascar 23½° S.
8 The north of South Africa on the Tropic; Lake Victoria 0°; Sudan/Egypt on the Tropic; Aswan High Dam
9 Namibian Desert coast on the Tropic; Lake Chad 14° N.
10 Timbuktu 17° N.; Accra 5° N.
11 West Africa coastal bulge 15° N.

Earth

Tell North from The Time By Using a Protractor

Reverse the method for telling the time from the compass protractor and North. To make the zero point North...
Make the sundial (protractor) show the correct time (one side of 'noon' time). Simultaneously make sure the sundial face is tilted to be parallel to the Equatorial Plane… I.e. For best results the tilt of the protractor according to the latitude, needs to be accurate.
Consequently, you may find it an interesting exercise to mockup a large (bigger = better) cardboard multi-protractor — e.g. by marking the sides of a cardboard carton.
Use a plumb-bob to **accurately tilt the protractor, according to your latitude**. E.g. pin the plumb-line to the top, side corner, and mark 90° on the flat face…

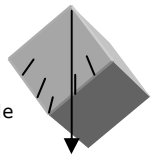

This instrument may be used to find North if you know the local time, or to find the local time if you know North.
One side of the carton is used for the latitude protractor in conjunction with the plumb-bob, a second side can be marked as a sun-dial protractor.
E.g. use another (long, straight, perpendicular) pin at the top corner of the side facing you, to cast a shadow across a 90 degree scale (— this is for accuracy, rather than make a full 180° or 360°).
Morning and afternoon then require separate scales; Winter-time requires that you face away from the Sun.

Sun

For Land-Navigation, Choose the Plumb-Line Star-Pair Method

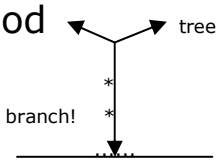

- Use a smooth constant diameter line, preferably thick. Make the plumb-line tall and stand far back. Make the cord either white or black to stand out against the sky, then maybe flash a torch onto it.
- Pour sand around the plumb, to anchor it to the ground. Avoid wind, especially if you hang the line from a tree branch!
- " Blink " the stars "on" and "off" behind the plumb-line, both on its left and on its right, by moving your head.
- Average the "start" & "finish" times. I.e. "It seems to line up now for the first time"… "definitely finished now". One accurate set of measurements is better than several rough ones. One star will gradually move over the top of the other star, and you have to pick exactly when.
- Choose well-separated stars (e.g. Rigel & Capella)
- If you want the timing to give you your longitude immediately, choose a North/South pair (e.g. Rigel & Capella) — (you need to know the star's longitude or its timing elsewhere, i.e. from a known longitude).
- An East/West pair will give you your latitude (without the need for a watch). The observation here is not the time but **whether or not** the given pair will attain to uprightness, after the first glimpse or before the last glimpse. Beyond a certain latitude, a given star pair will always remain tilted, less than vertical. This latter process can be time-consuming. Use a hand-held plumb-line occasionally, to check the progress.

Stars

Predict the Moon Phases by " N minus M "

" N " is a number (you calculate) for one particular year (e.g. For 2004 it was "24", in Australia's time zone, for the New Moon)
" M " is the month number (e.g. April is "4")
" N – M " is the day of the month for a certain fixed phase (e.g. in April 2004: 24 – 4 = 20th April 2004 for the New Moon)
April 20 was dark all night. New Moon seems to show the most regularity. The 'new Moon' is more accurately 'no Moon'.
Other phases happen one week later, two weeks later, three weeks later — 'quarter' by quarter — 29½ days for the whole cycle.
(The number, N, in 2004, would be 2; 9; 17; 24, 32 approximately, for 1st; 2nd; 3rd; final and 1st quarters respectively.) You can add 29½ to N.
Then in 2005, the number, N, for the new Moon was 13... (i.e. in April 2005, 13 minus four equals… the ninth of April — no Moon).
And in 2006 I used N = 32; 21 in 2007. The number N for the New Moon must be close to any **New Moon date + M**, e.g. December New Moon+12
Check your diary's calendar each year for the Moon phases; test and tweak a number N which works to within plus or minus one or two days.
The prediction is fairly crude, but with trial and error you can come up with *the best number and phase combination* for that year. For 2010, it is 18.
—usually **11 days less than the year before** (12 for leap year) because the lunar year of 354 days is 11 days shorter than the ordinary solar year
I.e. a good mnemonic is: **the Moon gains one day a month for every month except February** (which has only 28 days).
Hints: *Subtract " M " twice for January and February,* for more accuracy on non-leap years. Alternatively, to cope better with leap years...
Start the year in March (after the short month, February). Call January & February "months 13 & 14", but subtract 1 more for their accuracy.
Note: The "Full Moon" is when the Moon is nearest to "opposite the Sun" and it may not be exactly on the horizon when the Sun is.

Moon

Walk across Australia!

(Substitute whatever continent is appropriate)

Mentally line up a wall-poster size map of the whole continent (or bigger) with the track.
When you approach a bend, place the map on the bend, then walk all over it!
It's hard to miss the meaning of the bend if you bother to do this.
Your mind may not believe that it is not still going 'straight ahead' until you force it to catch up with the facts!

English!

"East-around" from North, is West-around from South, in a horizontal clock-sense. An "Easterly" course means East-wards — moving **to** the East.
An "Easterly" motion means East-wards. E.g. an "Easterly spin"-motion = "orbiting in an Easterly sense" = moving E around the N/S spin-axis **But**…
An "East(erly)" **wind** means flowing Westwards, **from** the East. An "Easterly" **current** might be flowing Westwards, from the East, but an easterly tidal "set" is an offset to the East. (Set is how far the tide "sets" you off track. The speed is called the tidal "drift".)
" Due " East means "exactly" East, as when a loan is "due" — "right now!" **Always double-check what someone means**!

Anticlockwise

Global

North/South and Noontime by **Bisection (of Rise and Set)**

Rise and set are **symmetrical around North/South**, provided you use horizontal, and not the horizon.
In hilly country, instead of 'rise and set', you should use **equal height** of the Sun (e.g. in NE and NW).
"Equal height" translates as *equal length shadows* of an *upright*, *straight* stick or cord, onto *flat* ground……
Or, better, use a bush sextant, p.27c, set to a fixed reading.
There are four ways to bisect two directions to find North/South — so don't muck it up by 90° or 180°!
The second diagram will work in both hemispheres — but double check the results for sense.
There are also two ways to **bisect the time** — so don't mix up 'just before midnight' with 'just after noon'.
It is also hard to 'bisect the time' to find noon, which of course, is not at 12 noon exactly, nor at 1300 hrs exactly in summer.
To work with hours, you need to **carry 60 minutes**, not 10 or 100! So be careful and crosscheck your figures.
Be accurate or your results will not be worth the effort.
E.g. rise: 0529 hrs; set: 1820 hrs Add… = 2349 hrs Halve… = 1100+(**60** +49)/2 = 1154½ hrs = noon.
Check: total elapsed time: 1820 –0529 = 1251 = 2x0625½; 0529 + 0625½ = 1154½ = correct
1820 – 0625½ = 1154½ = correct

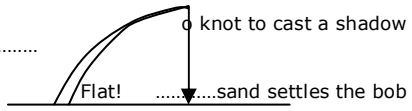

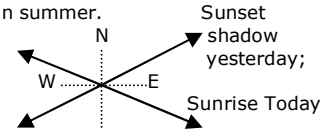

Sun

Longitude and Latitude from the Sun

Longitude...
Try timing the Sun for **equal height** each side of noon e.g. NE & NW (SH) to calculate the time of noon.
This can tell you your Greenwich longitude, provided you know the Sun's longitude, that is, its G. H. A.— from an almanac or calculator.
The Sun shifts predictably at a rate of near 15°/hour westerly. Almanacs make it their business to predict its West longitude exactly.
Of course the Sun's longitude at the time of your noon is your longitude too — since it is north of you.

Now try timing the Sun in two **opposite directions** at the same heights i.e. East and West.
You could try using the shadow of a plumb-bob on flat and level ground. Mark the shadow carefully.
This will be most accurate in the season when the Sun reaches high — in the summer sky.
By a series of early observations you can guess at when to make a final observation.

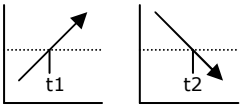

Using a theodolite or bubble sextant allows you to graph the rise & fall of the Sun (or a star) and see when 'same height again' happens.
See the 'Bush Theodolite' for an alternative measuring method.
Latitude (There are easier ways!)...
You can now solve the E/W formulae for your latitude — e.g. try the formula that predicts the timing of East and West for the Sun.

Stars

Equal-Height Stars Can Be Useful When the Sea Horizon is Visible

(or with a bubble sextant)

For direction, use the mid-point, of two stars which appear to be of equal height. — a great circle direction
For longitude, the timing is the important bit. — these stars should be approximately north or south of you
For latitude, the test is whether or not the stars do attain to equal height. — these stars will be approximately East or West of you
For a position fix, try to get two pairs, one north-west, the other north-east of you, for instance.
A **bush quadrant** is easy to make with a string tied to a stick! — just pull the string to length to measure or compare star heights.
Rest your hand and chin on something solid if possible. Use your full arm's length for best accuracy.

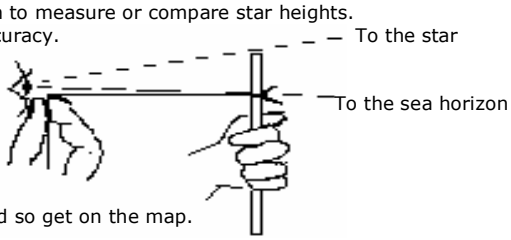

You can fix your latitude by measuring the **timing** of **two** separate star-pair events
— the relative timing — e.g. '5¼ minutes apart' (is unique to that latitude).
Simultaneous events are simplest, but separated events are easier for one person to time.
You don't 'get on the map' with this simple method unless you have been there before.
Unless you want a lot of trigonometry no calculations are required,
only the observation of re-recognisable appearances.
You can fix your longitude, too, by measuring any absolute timing — an exact star time — and so get on the map.

Anticlockwise, to the East, is Your Best Spin-Reference Direction

Because we use our North-South axis as our reference direction, we have to learn the *spin-sense* as something *separate*. See 'Orientation' p 35a.
The standard viewpoint for any spiral system (e.g. a weather system, a clock-face) is 'from the top', "**as seen from above**". Horizontal directions are measured "clockwise around" as you look *down* on them. For our spiral galaxy & solar system, the correct viewpoint is therefore 'from the North'
Looked at 'from the top', i.e. from the North, our Galaxy spins anti-clockwise. Or more simply, 'to the East'. 'Up North' is 'on the left' of East.
This means that the real motions are East-around the North/South spin axis—opposite to the familiar idea that everything sets in the West.
Because Northerners have their back to the North Pole when they see the Sun Down South of them, they see their horizon spin anti-clockwise
— that is, the hills and trees on the horizon move to the left of the Sun. They are looking "from the top" i.e. from the North further South.
Their Sun only appears to move clockwise (to their right, to the West), because it is their horizon moving—anti-clockwise (to their left, to the East).
To avoid confusion, always make the distinction between real mass movement and apparent virtual rotation. And think: We spin East.
Because we think of 'clockwise' and 'right-hand' motion as 'standard', and setting "in the West" as 'normal', it's a bit of a shock, to rethink!

When you are Down Under, on the South side, looking North at the noon Sun, you see the bottom of our Galaxy, and Solar System, and Earth.
The three 'North' spin-axes are all very roughly aligned together, all lying in our "Northern" Hemisphere — not too much tilted from each other or from our own spin-axis. You mentally lift yourself further out (not "up") into space to look at (but not really 'down' on) the spiral motions
Remember that from a 'standard' (Northerner's) point of view, the Southerner's view is really "a bottom view looking upwards" — from the South half, and you have moved out further South, for a better view — with the South Polestar behind you. You should see clockwise real movements.
Turn around and look Southwards at the Southern sky and you will see as if 'from the top'—or at least 'from within', but looking South. Anticlockwise
Remember to rethink "Up" as "to the *North* Pole of the spin".

Clockwise

- It should now be easy for you to visualise, in either hemisphere, your Southern horizon spinning anticlockwise.
- Look also at the Moon—visualise its orbit and its spin—'anticlockwise' from N.H., but clockwise-around if you are seeing it from e.g. Australia.
- Find the bright planets — can you 'see' them moving 'anticlockwise around the Sun' in orbit? The Ecliptic marks their path.
- And, go further, visualise the real Sun spinning much like the Earth does, on its own N-S axis. The Sun is North of you in the Southern hemisphere, spinning on its own axis 'clockwise from below', or in the Northern hemisphere, South of you 'anti-clockwise from above'. Use the Ecliptic's N-S spin axis as a good guide for the Sun's, or simply use the Earth's own axis as a rough guide — Don't look at it, but imagine its sunspots moving.
- Keep going: The Earth hurtles through space in orbit around the Sun 'to the left, while looking to the Sun, around a N-S axis of spin' (Down Under, looking North) 'to the right, relative to the Sun' (NH) '**anticlockwise**, looking South' (NH)—'with' the apparent movement of the Sun, in both cases. That makes the noon horizon move in opposition to the movement of the whole Earth through space. Like the slow 'inside' of a spinning ball circling.
- Imagine the whole Earth-orbit in the sky. It's easy when you get the hang of it. New Zodiac stars seem to emerge (invisibly) on the Sun's West side, in seasonal procession, because "the Sun is edging Eastward along the Ecliptic in a yearly cycle" i.e. we move to the West side of the Sun and see behind it to the background Zodiac stars. Don't look at the Sun, but you should be able to 'see' this happening as if there were a star chart behind the blue sky! Try just before dawn.
- Now, at night-time, find the centre (thickest part) of the Milky Way near Scorpio/Sagittarius and visualise the whole Solar System orbiting it perpendicular to that, the Sun/Earth/Moon all moving as a unit, to your left (SH) or right (NH) towards Suhail, away from Deneb.

The Milky Way marks the plane of the spiralling galaxy (from which you can determine its N-S axis, tilted differently from ours).

Finding Magnetic Variation by Bisecting Rise and Set

Accuracy is needed.

Measure the compass directions of the morning and evening Sun at equal heights e.g. rise & set on a flat Horizon
Bisect the result. E.g. Rise: 079° Compass Set: 271° Compass
Add: 350° Halve: 175° Use the Opposite: +200–20=355° (See p.18e)
So the Compass North must be 5° to the East of geographic North. **Draw diagrams to avoid getting confused.**
Check all numbers! E.g. 355°–271°=084° 355° to 360°=5° + 079° more=84° Correct.
Or 271 – 79 = 192° spanning = 2x96°
79+96 = 175°
= opposite to (375 – 20)°
Correct.

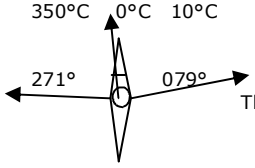

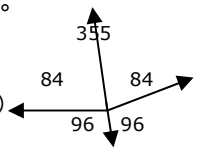

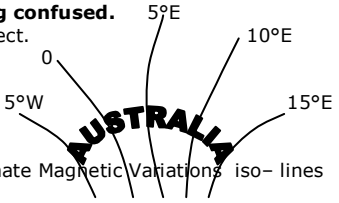

The compass readings are "Least", by 5° (see p 107b)
E.g. 079°C is really 084°T. (see p 14e)
Approximate Magnetic Variations iso– lines

A Bush Theodolite for Overnight Equal Height Observations

This is for accuracy in calculating... True **North**, Noon **time** and Compass **variation**
And for any equal-height paired observations — Sun/stars/etc.
Find a large bent stick. (Or bend it with a bowstring)
Hold it up by a loop-knot e.g. hang it on a twig.
(If it is early or late i.e. a long shadow, the stick will not need to be bent much.)
Late afternoon, at camp,
Mark where the shadow of the knot (or the other end of the stick) falls on the stick;
take the time; line up the landmark it points to; and find its compass direction.
Early morning, next day,
Hold it up (in reverse); wait until the shadow again falls in its proper place;
get the time, direction and compass reading again.
For a **Bush Bubble Sextant**, arrange a water filled clear tube on a triangle of lashed walking sticks so that there is a trapped bubble in a tube section with an ever so slight convex-upwards bend.
Make the bubble always stay at one marked place when you are using it.

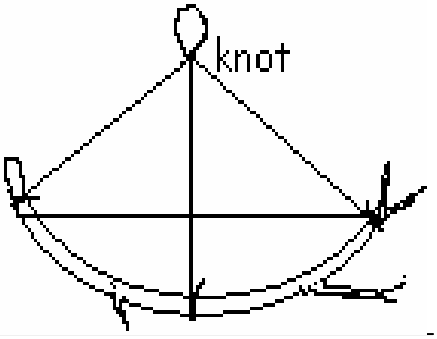

Great-Circle Reference Star Directions

You need some familiar reference points. It's a simple matter of going out at night, waiting around, and tracking down a few of these...
Look for **stars which are closer to the pole than you are**, and see where they **rise and fall vertically**. It happens once or twice a night, per star.
This gives you a series of great-circle star directions, already labeled, because **the 'maximum latitude' is equal to that star's 'declination'**.
E.g. Acrux is about 30° from the pole, or "60° latitude". If you steered towards one of its vertical-motion directions you would end up underneath it.
So as it rises or falls vertically (once each per night, if visible at the time), as it circles the South Pole Star, it lures you to 60°S latitude.
E.g. A star overhead, as the extreme limiting case, is moving East to West, rising vertically from the East and falling vertically to the West, meaning that you are *already at* the maximum latitude of that course, if you had been heading towards that star guide, now travelling E-W also.
You then have to head directly *away from* it (e.g. the next night) to further pursue that course.
Heading away from that star either East or West would take you eventually back towards the Equator, like the way you travelled down there.
The **stars further from the pole, than you are from the Equator,** make a complementary set of reference directions, by using the horizontal direction 'at **right-angles** to where stars **rise and set**'. **The declination is a measure of *how close to the pole* you would pass**.
The stars close to the pole are unavailable. The co-declination labels 'the maximum latitude'.
E.g. Sirius is 17° South, so rising from the horizon, somewhere to the South of East, it would send you to within 17° to the left of the South Pole.
Turn around and it would guide you (part way) to cross the Equator and to within 17° of the left of the North Pole.

Make a Bush Psychrometer

..for absolute humidity and dew-point-temperature measurements

Carry a wettable piece of material — dip it in water — hang it up to dry (in the shade out of the breeze) — time how long it takes.
Measure the air temperature at the same time, and cross correlate the two, to find the absolute humidity. If you calibrate the method at home, by graphing the results for a whole range of different temperature and humidity conditions, you will get to recognise a dry air mass from a moist one.
The piece of material needs to be non-porous, non-rotting and dark coloured — to show the drying process clearly, and to yield repeatable results.
Night-time minimum temperature more or less bottoms out as soon as an air mass starts to dew. The ground then wets down further instead of cooling down further. The normal method of measuring absolute air humidity—the percentage of water vapour—is by evaporatively cooling a water drop in it. It requires wetting a thermometer bulb, fanning it vigorously (trying hard not to break it) and measuring how far you can make the "wet bulb temperature" "dip" below the "dry bulb (air-)temperature". The water temperature won't get all the way down to "dew point"—the temperature at which dew or frost will form in that air mass—unless it already is dripping wet outside. It may get a bit more than half-way, and then only when only dew is expected—DP above 3° C. **If frost is possible,** i.e. when the dew point temperature is 3° or less, **it will dip at most half way down to the dew point**. Why 3°? Well, when the air is very dry, dewing will further reduce the water vapour content of still, cold air, from a dew-point temperature of a few degrees above zero, down to 0°—resulting in 'frozen dew'. **Rule of thumb** for calculating dew-point from dip:
When the DP is 40°C/30°/20°/10°/5°/0° then add to the 'dip' 0/1/2/3/4/5 quarters of the dip, to get down to the real dew point.
If the night is long enough and calm enough to cool down enough, & the sky is clear enough, a DP near zero will result in a frost.
E.g. 15° to 10° = 5° dip, 5°DP. 15° to 13° = 2°dip, 11°DP. 19° to 18° = 1°dip, 17°½DP. (Remember that you can't go far wrong anyway.)

South is a Triply Better Reference Direction than North, in the South

1. The closest geographic Pole in the Southern hemisphere is just over the horizon in Antarctica. Globally seen, all directions pointing South are convergent, whereas directions pointing North are initially divergent, and lengthy and less meaningful... Travelling North you first cross the Equator before it gets colder — so the gut-feel meaning is less clear. The meaningful direction, when you are in South latitudes, is to go further South to get much colder. (It is the same as in the Northern hemisphere, where going further North makes it much colder.)
You may find that in a place like Tasmania for instance, the maps seem to be upside down, because although North is at the top of the map, the logical reference direction is toward the bottom of the map, where the cold place gets colder. You don't think: "This is a warm place. In which direction does it get warmer?".
As a result it is easy to get disoriented by 180°, if you insist on using North as your reference direction. In addition…
2. When you put South at the top, and look South, then the Sun will move to your right, as if seen from the N.H., across the map as time passes, for both hemispheres—and that will make the clock-to-Sun link obvious (like putting a Southern clock on the S. wall to match the Sun's motion).
Northerners especially, would find that a more natural way to view any map—as if from further North, looking South, to the Sun moving right.
3. Putting South at the top, with the Sun & Tropics behind you, looking down onto a map, is the mathematically standard viewpoint
— like looking from the top of the Galaxy — from North to further South — the land South of you moves Anti-clockwise, to your left, to the East.
You can still orient your map normally and correctly, but there is no law against habitually viewing it from its Northern edge, looking South. Try it.

Global

Find East/West By Delayed Bisection of the Sun and the Full Moon At Rise or Set

To find due East (+/–10°): **Bisect the sunrise and the full Moon rise** — since they are of opposite declination.

For due West (+/-10°): **Bisect the sunset and the full Moon set** — since they set each side of due West.

Obviously, you need to wait about 12 hours for this method — it is really just for your background comfort level about where things are.
There is a slop of 5°, since the Moon's plane of orbit is about 5° away from the Ecliptic.
There is also a slop of about 24 hours, since the full Moon occurs at some time other than Moonrise or Moonset,
plus it is hard to pick an exactly full Moon, and the Moon varies in declination, i.e. North/South position, from day to day.

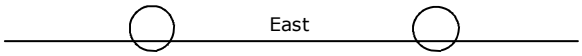

Sun

Find North by your (non-digital) Watch-Hands

Point the noon on the watch (**not the 12!**) to the Sun.
Tilt the watch axis toward Polaris as best you can judge.
This tilt matters a lot, but is not usually mentioned!

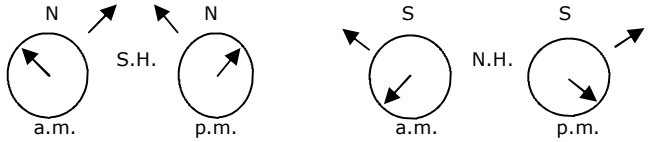

Southern hemisphere: **North is in between the Sun and the hour hand**.
Northern hemisphere: The Sun is in between the hour hand and South.
Why?: Clocks Go Clockwise; at 30° per hour — twice around per day.
For the Southern Hemisphere that means: — twice as fast as the Sun and opposite.
So 'in between' moves neither clockwise nor anticlockwise — it stays N/S. At real noon-time, the hour hand and the Sun cross over.
For the Northern Hemisphere, the hour hand overtakes the Sun at noontime, at South.
It's easy to visualise in fast motion, once you catch the idea — it's like the hare overtaking the tortoise.
Remember to tilt the watch-face up to be parallel to the Sun's orbit, not horizontal, because a sundial shadow is not linear.
You could use the middle of sunrise/sunset times to find the local actual high-noon midday time, which is needed for "point noon at the Sun".
The same method can be applied to the stars, the Moon or the planets — provided you know when the one you want to use is north of you.

Stars

Leo the Lion Looks like a Lion or Sphinx Squatting It Has Four Lines which Show N./S. E./W.

Two in the 'hindquarters'; one along the 'back'; one as the stem of the 'question mark' (Leo's mane).
(You have to turn the '?' around.) For the Southern Hemisphere, turn it all upside down.

This is a good sign in Northern spring/Southern autumn.

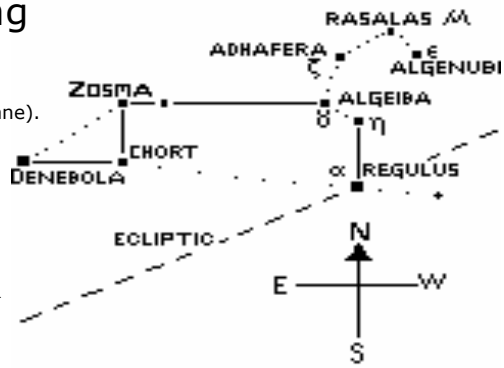

(NH view — LEO is upright)
Direction of movement →

Moon

To Predict the Moon's Path, Visualise the Ecliptic

For more clues on how to visualise the Ecliptic, first see the overview at the end of this section.
The Moon stays within 5° of the Ecliptic. That helps you see where the Ecliptic should be, or alternatively,
where the Moon should be. The Full Moon is easy, and has been covered on p. 24d.
A **half-Moon,** for example, is 90° from the Sun; so if one is on the Equator the other is not and vice versa.
E.g. a half Moon in mid-winter will rise and set due E./W. +/- 5° or so.
Secondly, recall that the Earth's spin-axis is tilted 23½ degrees from the solar system.
This is how it would look from Sydney (*), at midday, in Midsummer, i.e. at Southern Summer Solstice…
The two great circles intersect at 23°½. The whole double-circle rotates West over 24 hours. i.e. sunset is at X…
To visualise the Ecliptic (see also p. 30d & 33b), from the Sun...
..Since the Sun moves eastwards seasonally through the stars you can know that
the part of the Ecliptic which the Sun was recently on must 'lead' the Sun to the West.
E.g. after solstice in Southern Summer, the most Southerly part of the Ecliptic is ahead of the Sun
to the West, and so a waning crescent Moon (to the West of the Sun) will be higher in the sky, on average, than the Sun. (SH)
E.g. A crescent Moon at sunset at spring equinox will look extraordinarily high over the sunset glow, because of the 23°½ tilt.

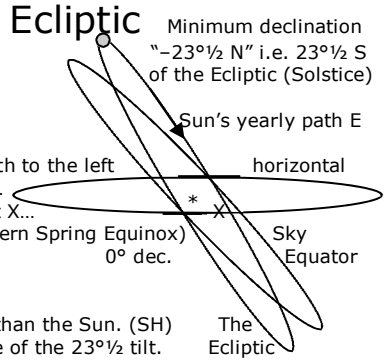

Clockwise

Bearings & Tracks: "If it is Right, it is Big, in my sight"

= Anything on my **Right is Bigger than 'ahead'**. I.e. If I am heading 100°, something on the right has a bearing bigger than 100°, e.g. 200°.
And "Any track on my **Left is too Little to look at**" = Anything on my **Left is Less** than ahead, e.g. 20°.
These mnemonics compare fixed tracks or directions, and their 'Big and Little' measure in degrees, East-around-from-North i.e. clockwise bearings.

[The changeover from 360° to 0° can give you trouble. Where 'ahead' was say 355°, 15° bigger should be 'big' e.g. 370° but will look really small, 10°. Because it is nowhere near 355 in number, there should be no possible confusion, provided you think of 10° as 370°.
'Little' may look really big, if 360° gets in between. E.g. where 'ahead' was 005°, 15° smaller would be –10°, which is the same as 350°, which is nowhere near 5 in number.]

Furthermore, if A sees B on his right, relative to a fixed direction, B sees A on his right,
provided that they use **acute** angles, or they look in **opposite** directions e.g. facing each other.
This could all get **confusing if you don't remember those two provisos**. See p108c for how to put this mnemonic into practice.

Find North from the Combination of the Sun and the Moon

When you are familiar with the path of the Sun and of the Moon, and with their seasonal movements, and with the Equator and Ecliptic, then whenever you see the Moon up during the day, imagine where the Sun's path lies in relation to the Moon's, and vice versa.

This gives you two concentric circles to visualise, which of course allows you to locate Polaris (at the centre of those paths), and from there, North.

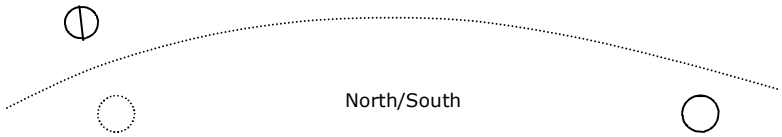

See box d/e below, to know where the Moon should be.

Find Any Other Direction Too — Using the Hour Hand

This method works, in theory (there are easier methods!), for any other body — star, Moon or planet — or for any other direction.
For instance, you may notice when the Sun is due East, and how high it is; or again, when it is in the direction in which you wish to head.
Or again, you may be following a star, and of course it will move and of course you will wish to compensate for its movement.
In these cases you substitute the time of that body being in the chosen direction for "noon", and substitute the chosen direction for "North".
Keep the watch axis pointed to Polaris as best you can judge it.
For example: you know the Sun is West at 4 p.m. At midday it will look like this… (SH version)

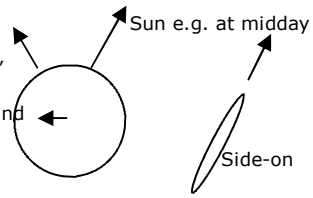

.."Point the four to the Sun; West (or the chosen direction) is between the Sun (or star etc) and the hour hand"
The catch here is that you must imagine the height of the body accurately, before the direction can be found.
This is because the method only shows the rotational position of the body — longitude wise, not its height declination wise — and with the clockface being tilted, the horizontal direction being pointed out is not obvious until you choose a height.
Of course, the accuracy of the method depends also upon aligning the clockface to the Equatorial plane.

String North

Carry a piece of string, about one metre long, to **measure the angular height of the Moon, planet or star** which you intend to use later.
You measure it once, as a matter of preparation, then use it many times. The Moon goes out of date the most quickly, but the stars are permanent

Put a loop knot at one end, to hold the bottom knot, level with the horizon or horizontal. You may have to wait, until the height is clear.
Hold the string at arms' length in a vertical plane to the North-South, which is where the star will reach its highest point.
Put a second knot to show the maximum height. This is equivalent to the 'noon' height for the Sun.

For the Sun you could use the shadow, and hold the string *down* from the horizontal. It will be accurate for a week.
Better still, use the shadow methods soon to be described, to avoid the temptation of looking at the Sun.

Now at *other* times of the day/ night, first hold the string in a vertical plane to reconstruct the expected height, then hold the string horizontally from that same high angle above you. Lastly, twist and turn, so that the horizontal string apparently points to the body used — Moon, star, planet.
You have to **twist yourself until the string is *horizontal*, and at the right *height*, and *points to* the chosen body**. Then you are facing N-S.
Actually the method approximates a slightly curved small-circle star-path with a great-circle straight course.
You can find approximate North with the string, then adjust it slightly by imagining the actual star-path as a small circle around the Pole Star.
The straight string will then appear to point to just above the star, but in your imagination, it curves down to match the guesstimated star-path.
This is the direct sighting equivalent to the 'shortest-shadow South' methods on the next page.

A 4 Week *Position Cycle* of Moonrise/set, to Find East/West

The Moon, like the Sun, lies close to the Ecliptic (in the Zodiac), and completes a cycle along it (within 5°) in 27¼ days (not quite 28 days). So...
The Moon changes position in **a four-week cycle, like *a speeded up year***. It spends a week in each 'seasonal' position: High; mid; low; mid; hi...
The Moon crosses the Equator every fortnight, rising due East and setting due West that day. Find it! Use p9c to find E/W on those days/nights.
In alternate weeks it will rise & set in the Midsummer then Midwinter positions *for the Sun*. Get to know them, and expect each one every 4 weeks.
• Its change of rise/set position day to day is most sensitive near the equinox position, due E/W (be careful), and is least critical a week later.
In between times, the rise/set **positions can be guessed by the day of the week**, by interpolation, as a good direction guide on any day.
The cycle runs fast by only 1 day in 37 1/3 days. So it is a continuous guide to E/W, **for a month at a time**, as it moves through 13 Zodiac signs.
• The best way to **set your Moon-watch is by observation**. If Moon-rise/set is too late at night, then look for Moon-set/rise during the day.
Sunset Full Moon is a good starting point for estimating the nightly-cycle; Sunrise New Moon, for the daytime. Midsummer or Midwinter or Equinoctial full Moon are very good starting points, because the Moon will be in the anti-Sun position. Visualise the Ecliptic (through the Sun and perpendicular to the Moon cusps joined) and its shape in relation to the Equator, and use a compass. Say the full Moon is opposite the Sun on Saturday, but by Monday the Moon sets due West. Every Monday for *up to* 5 weeks the Moonset will be in a recognisable position. E.g. W, N-W, W, S-W, W.
On a later Monday, you will need to step back one day and say "every Sunday, for five weeks" the Moon will repeat those four positions. Adjust yourself by one day, *as it shows itself necessary*. You can simply say "It is a Thursday Moon at the moment", expecting to have the change to "a Wednesday Moon" in a fortnight or 2. Keep your finger on the Moon-pulse to keep up-to-date. E.g. Say: "It's a Spring-type Moon next Tuesday".
• To keep track of the weeks, here is a recipe: **Call the new year e.g. 'a Friday Winter Moon'** and adjust 1 day every 37 days, 10 days over 1 full year, or by **4/5 of the completed months**. 7, 8, 9, 10 days for *after* 7-, 8-, 9-, 10-ber, using the Latin names; 0-7 days while *in* the rest, 1-8.
• Perhaps you lose track of a 'Wednesday Moon' for a couple of months, and try to start again. Track the Moon's Motion along the Ecliptic.
The Moon overtakes the Sun at No Moon, gets one season ahead of the Sun within a week, at 'the first quarter', is opposite it in another week ('full'), and is one season behind the Sun in the 3rd quarter. There are 12 and a third months in a year, so the Moon is moving 12 1/3 times more quickly around than the Sun, in its seasonal cycle. **What takes the Sun one month will take the Moon only 2½ days**. (31=2.5x12.3).

Step 0: Visualise the cycle. — I.e. imagine where the **lines** of the Ecliptic and Equator must be in relation to the Sun and Moon.
Step 1: Place the *Sun* on the Ecliptic — I.e. estimate the time of year, and from that, the time before/after the next/last **solstice/equinox**.
Step 2: Try to **visualise where on the Ecliptic the *Moon* must be**, according to its phase.
Step 3: **Interpolate for the exact phase**, and label it a 'Monday Moon' or whatever. Check your guess with a compass if possible.

E.g. There was an eclipse of the Sun on Tues July 22 2009, 1 month after solstice. The 'Winter Moon' must have been 'Sat night'; and for 2009 'Mon'
E.g. 23 November is one month before the SH Summer solstice (2½ days worth of Moon motion through the stars). You can work out, for any phase, which phase and position to expect in 2½ days. For instance, Nov 23 is late Spring, almost Summer, Southern hemisphere (with the Sun South of the Equator and moving further South). The Sun is high and getting higher. Opposite that, is a *Full* Moon position (North of the Equator moving further N). It stays low in the N, getting lower. Or if it is a '*First–Quarter* half-moon' at that time, it will be in the 'late-Summer-almost-Autumn' position (approaching the Equator from the S, midhigh in the sky). Or if the Moon is a *third quarter* half-moon, it will be in an 'almost Spring' position (approaching the Equator from the N). You would therefore know that, taking the last example, over the next 3 weeks, a Waning Crescent/ New Crescent/ Waxing Gibbous/ Past-Full Moon, will rise E/SE/E/NE, and set W/SW/W/NW, in approximately 2½/9/16/23 days.

Global

Sun Path Norths — Summary

- Noon-time: You look at your watch to see how far the Sun is away from apparent-noon-time, then look at how far your shadow must be from shadow-South. Near noon, the (untimed) shortest-shadow methods below aren't accurate, but this timed method can fill in the gap.

- Chord bisection: This works for any body e.g. the Moon.
- Sundial methods: You set the sundial to the right time. E.g. tilt the compass protractor to the equatorial plane.
- Shadow bisection: This can work for the Moon too. Just remember to use a vertical stick on level ground, e.g. a dry puddle.
- Equal-height: You can use an improvised sextant/quadrant/bubble sextant. Don't forget to bisect both the time and the direction, carefully.
- Sun at East/West: You need to be familiar with the height &/or the time when the Sun is East or West.

- Equinoctial sunrise/sunset: You keep track of how far off East/West the sunrise/sunset is, at this time of year.
- Equinoctial rise/set angle: You can extend the Sun's path to noon, rise or set, in a straight line, using the latitude angle.
- Equinoctial East: Where the shadow moves to, with slight adjustments at other times of year.
- The Noon-shadow path is always to due East. This is useful for anytime *near* noon, say +/- 2 hours.
- When the Sun is *very high* near noon, you *need* to see the direction of movement of any point shadow on flat ground — West to East.

Sun

Locating Anti-sunrise or Anti-sunset

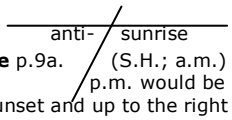

Imagine a line from your head-shadow, *opposite the Sun's path*, cutting the horizon at the correct rise/set angle p.9a.
(The Sun's path is approximately a great-circle, so now imagine a plane, edge-on along that line, cutting the level ground.
Look for where the shadow would be closest to you. This approximates the Shortest Shadow South…)

Shortest Shadow Souths ..from calculating, memorising or knowing the angle of high noon shadows.

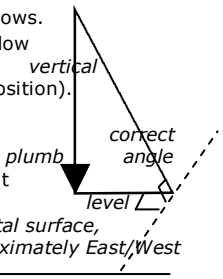

• Walking-Stick North: You lean your walking-stick, so that at noon it would cast no shadow, and so that right now it casts a shadow at right angles to itself, a direction which is nearly West-East. Adjust (all these methods) slightly for the season, afterwards.
• Body-shadow South: You use your own shadow tip joined at right angles to your shortest-shadow tip (guess at its length and position). This works best in the early morning and late afternoon, where a guess as to your shortest-shadow is sufficient.
• Cardboard Triangle East/West: You cut it to point to high-noon height, like the walking stick & body methods combined.
• Plumb-line North: You tie a string triangle instead, to the right shape, and maybe attach it to your walking stick.
• Notebook North: You flip open one page from horizontal, so that it casts no shadow, and also points straight to high-noon height simultaneously. I.e. the page should be held at the correct angle, mimicking the Sun's maximum height.
You can discover the correct angle by observation, yesterday,
or you can calculate that angle as "latitude degrees minus declination degrees away from vertical"

Stars

"Star-Time" Gains Four Minutes a Day

3 minutes 56 seconds actually, by appearances; (3m 56.555s if you use the almanac, and 'the first point of Aries', because "Aries" as it is known for short, is not a real star, and it drifts slowly through the Zodiac)
Stars complete one extra orbit — 360° — than the Sun in one year — 366¼ in 365¼ days — so that's about 1° per day which is 4 minutes of time.

By "time" we mean Sun-time "15° per hour on average", in ordinary usage — a slightly varying rate in practice.
By "star-time" we mean "15 and a bit degrees per hour spin rate — a constant rate"

The stars outrun the Sun, night by night, by 1° on average, so the Sun seems to migrate 'backwards'—to the east—through the Zodiac belt of stars.
You could say: **"The Sun drags its future after it!"**, meaning that the Sun sets in the West, but its future star-position shows up just to the East.

You can zero your star-time, by timing when a bright star (preferably one near the Equatorial belt) passes overhead, or north or south of you on a particular night from a known particular longitude. You then calculate star-time from "Sun-Time plus 3m56s per day since then" including part-days in your detailed calculation, at a rate of 10 seconds' time per extra hour — 1 second per six minutes.
Your reference longitude becomes like Greenwich to you — or, alternatively, you can adjust everything back to Greenwich longitude and Greenwich star time. You can also buy quartz controlled star-watches or calculators.

Moon

Adjusted Moon-Cusp North: To the Left or Right, By Moving the Sun

"Cusp-North" can be improved, by mentally moving the Sun away from the Ecliptic *onto the sky Equator*. Do the adjustment once a day. Guess the declination of the Sun, from the season; then shift it onto the nearest bit of the Equator and the Sun-to-Moon illumination angle may tilt a little.
Instead of pointing to the pole star for the Ecliptic, in Draco, the new cusp-North will point more to the pole star for the Equator — Polaris.
E.g. the midwinter Sun needs to climb 23½ degrees higher — which would tilt the cusps of the Moon (due to the shift in illumination). Diagram 1.
Caution: The illumination angle of the Moon does not change much when the *Moon* is shifted onto the Equator, and not at all for a half-Moon.
E.g. a half Moon (90° from the Sun) at Equinox could be adjusted by 23½° to bring it onto the Equator (It is like the Sun at Solstice), (diagram 2) but cusp North does not change. Being 90° from the Sun the cusps outline one and the same great-circle.

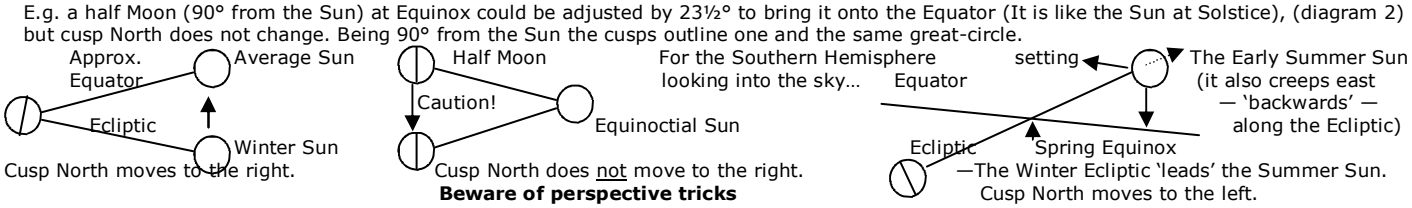

Clockwise

"By the Nose" Navigation

Our Mental Map is a Nose-Eye View — not at all like the bird's-eye-view given by a layout map. (See "Strip Map Navigation" p.95c)
Oh yes, our mind eventually fills out a (spatially-solid) layout map, but initially we collect a set of (spatially confused) mental photographs from all different perspectives taken from ground level. Our mind doesn't deal in map symbols or contour lines!
Rather it has search images, image enhancement, image recognition, three-dimensional images, and memorised **relative angles**.
Consequently, a more natural method of finding your way is what I call 'Nose Navigation' — almost opposite to mapwork.
Nose navigation uses 'strip maps' (Section 4), relative angles, and a couple of special mnemonics…

"Because (my shadow) seemed to shift right just now; Therefore *I* must have turned left"

Now substitute in the (…): Ayer's Rock/Uluru; That distant cloud; The Sun; The Moon; Compass Needle; East Wind; The RDF Null; That bright star
In other words: "My Nose Turns EVERYTHING Away!" i.e. when I turn my nose, everything else seems to turn oppositely.

This simple rule attacks head-on your insistent illusion that you have been walking "straight" ahead!

And your illogical belief that "Ayer's Rock has moved"! Neither has your shadow 'turned right' just now, simply because your nose 'hasn't moved'!
It is not as easy as you think to apply, e.g. when you are reenacting where you thought you had gone, in your (already confused) head.
The second nose mnemonic is on the opposite page.

Seasonal Adjustments For Equinoctial East

Shadows move due East at Equinox, all-day. At other times of year, when sunrise/sunset is not due East/West, the Shadow path curves slightly at an angle away from sunrise/sunset, an angle which you should become familiar with each week. Start by assuming that the shadow is moving "East"… Verify this **Rule: Shadow Paths Bulge toward the Hot Pole**, even in the tropics, in both hemispheres. (The Sun crosses at Sept 22 & March 21.)
Adjustment: The first guess at **"Hot Pole" Needs to be Adjusted Slightly,** less near noon, little or none nearer equinox, to "un-bulge" the path.

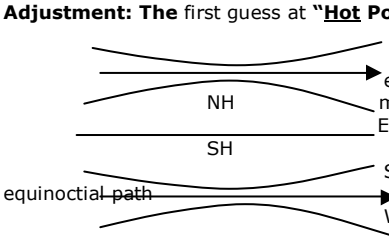

Remember that "the Southern Sun Sets Somewhat South in Southern Summer", so that opposite-Sunrise & opposite-Sunset need to come back South to point out E & W correctly
E.g. You guess East (or West in the a.m.) from the shadow's movement over 1 minute,
(It will roughly be opposite sunset/sunrise direction if you are late/early in the day)
then twist your guess a little back to the true hot pole, and to due East (or West).
Let's say it is midwinter in the Southern Hemisphere, so your shadow is somewhere S of you. The 'hot pole' is the North Pole, so twist 'North', as gauged from your shadow tip, away from the shadow, back toward you (i.e. North, to the hot North pole)
Visualise the bulge, then straighten it out at each end, back toward the hot pole.

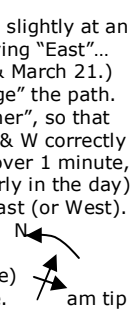

North-Finding from Two Paper Triangles

or even from only one

Cut out a back-to-back triangle shape and fold it along the join. Line it up as shown and 'N' will point to Polaris…

'1' is your South **latitude** (For the Northern Hemisphere use the North Latitude and swap N & S)
'2' is the **polar distance** of the Sun (from the 'depressed' pole, N)
i.e. its co-declination, or 90°-N dec. See p. 18b.
You can set it by observation or calculation; and can fold it to the correct angle, rather than cutting it.
('3' is a right-angle, but '1' and the horizontal is what really matters)
This two-triangle method works best away from noon, but nearer noon simply find the way shadows move
The 'N-S' edge is parallel to the Earth's axis and acts as a hinge around which the Sun revolves.
Sun and shadows move at right angles to that axis.
You only need **one** of the two angles if you find the direction of shadow movement
and swivel 'N-S' to be perpendicular to that. (You don't even need one angle, near noon)
E.g. Poke a straight stick into the ground, pointing at the Sun, casting no shadow; (Or use any point-shadow rotating around its point)
By waiting a short while, find the direction of movement at a half-way point, in mid-air, i.e. maintaining an equal distance to the tip,
and manoevre your 'N-S' axis (index finger), to be perpendicular to how shadows move, then at the correct angle to the Sun and/or Horizontal.

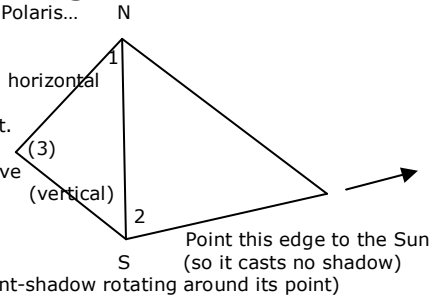

Working the Star-Longitude Numbers

The reference star point, a non–star, is 0°N/S;0°E/W, and is called "The First Point of Aries" or "the Vernal Equinox" where the 'northern spring' Sun crosses the ecliptic (on March 21). Other stars are measured from that (imaginary) star, like latitude and longitude are measured from 0,0 a nameless spot of ocean — each star has an almost fixed longitude in relation to Aries.
N/S 'latitude' is called +/- ' Declination '
E/W 'longitude' is called (a) Right Ascension — R.A. — or 'Star Time' — It is measured **Easterly**, in hours at 15° per 'hour', from "Aries"
Or (b) Hour-Angle — H.A. — It is measured **Westerly**, in degrees, like longitude, but 360° around, from "Aries".
The first, RA, is designed to tell us the time — **when** the star will show — since all the stars move around westward with Aries.
The second, HA, is designed to track the West **longitude** of the star in reference to the Earth, as the star progresses West around.
We need them both, but they are entirely equivalent to each other.
(The only trouble is that a star 'hour' of 15° is slightly faster than a Sun hour of 15° of progress to the West.)
Next you need to know that the first point of Aries was at 100° W. Earth longitude at the beginning of the millennium,
and at every turn of leap-year thereafter, at January 1.0 (midnight). Then it gains at a rate of 1.002737909, or about 1° per day on the Sun, (3m 56s time), so that the stars turn full circle during four seasons. "1°" is adequate accuracy in the short term for us but not for yacht navigators.
E.g. January 31.5= 100+30°+0.5° W., then you add the H.A. (or subtract 15 x the R.A. hours) for a particular star — that's the longitude where the star will be overhead. **For exact numbers consult an almanac**, but I am concerned here with giving you an overview.

Topographic-Level Winds

This is what you feel in your face at ground level, or close to it.
Fog; Smoke; Valley Winds; Sea Breeze; Eddies; Catabatic Winds; Forest Winds; Scree-Slope Winds; Mountain Winds; Glacier Winds; Lake Winds.
The various thermal winds listed will of course be very variable. The thermal winds flow from cooler areas, before rising in the warmer region.
Nevertheless they may betray the slope of the land, for instance at night. For example, fog forms on calm cold nights then drifts slowly downslope.
Being dense, fog will drift down-valley, down-stream, and off-shore. You can see the location of creek outlets from the early morning beach fog.

The ground-winds due to pressure systems, are affected by overall friction and therefore average out slower and less Sun-twisted than upper-winds.
Ground-level winds form waves, foam 'streets', sand ripples & dunes, drag marks, snow ripples & drifts, which all may persist beyond the winds,
and they cause lopsided bushes and salt-spray-'cut' bushes. See Section 3 for more details.

Gusts are at different directions to lulls since upper winds penetrate to lower layers.

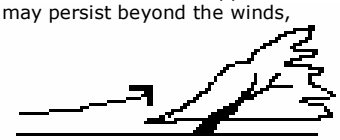

"Nose-Navigation Needs an 'Anti' Mnemonic": "Nor-West, Not-Least"

N.O.R. West: "North **on Right** = Going West" or "Needle on Right = Going **West**" i.e. going into the westerly **half** (top diagram)
N.O.T. L.East: "North **on the Left** = Going East" or "Needle on the Left = Going **East**" = to the easterly **half** (bottom diagram)

This is always true when you are relying on "Needle North". **You don't have to think any further.** But,
It is equivalent to ANTI-CLOCKWISE BEARINGS, which you can't read from a clockwise-marked protractor!
The following numbers refer to an ordinary compass — marked clock-wise, and set to 0° as straight ahead
(not North) so North swings left and right instead…
Example: in the rain or fog; the needle swings meaninglessly
Example: deep in a cave network
Example: on a yacht, with a fixed hand compass
Example: with a fixed car-compass, not marked anticlockwise
Example: making a strip map; or using one
Example: using a compass-holder, which fixes 0° as straight ahead

The compass needle might point to "045° NE"
but your nose is pointing NW
The compass protractor numbers say "North-West, 315°"
but your nose is pointing NE
Interpret "Anti-clockwise, West from 0"

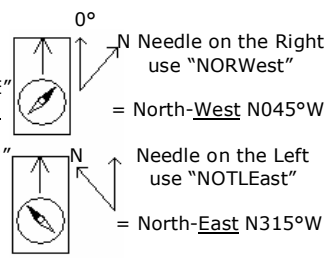

Revision, Overview and Explanations

Global

Top Down Orientation

The Big Bang **Universe** bloated over billions of years; and broke into bubbles of turbulence; all bits in each bubble orbiting each other; in millions of galaxies, with 20 or so galaxies in our little **local cluster**;
Local galaxies, like the Magellanic Clouds or Andromeda, glow milky (just as 'galact-ose' says 'milk-sugar').
Our galaxy is a great spiral, swirling slowly, anti clockwise from above i.e. Eastwards around its North-South spin-axis
— flat like a fried egg — a 'great circle', glowing as the Milky Way. The thickest milkiness, in Sagittarius, marks the Galaxy Centre.
The Milky Way is like the 'Equatorial plane' for the whole spiral, seen from within, with its northern pole star in 'Berenice's Hair' (28° N. 168° W.) and its Southern pole star in 'Sculptor' (28° S. 348° W. 012° E.) and our milky galaxy is, in its own turn, full of spinning, gyrating gyroscopes.
You know how a spinning top will pirouette with a slow-motion gyration when it tilts slightly? Well, precession, as it is called, causes slow changes.
Stars are Suns, especially shiny in our own spiral arm of the galaxy (i.e. closest) and seemingly set solidly in space by their extremely slow motion.
So the constellation shapes are chartable for centuries. Their apparent positions are gridded out in relation to the first point of Aries, which in turn is fixed each year by our Northern Spring Equinox—not a star in the sky, but a point which creeps along the Ecliptic, to the West.
Hint: When you are wondering whether a rule holds in both hemispheres or not, ask yourself whether you are down at the level of the spinning Earth or up beyond that in a larger scale. Try to imagine looking at the whole globe from an outside perspective. E.g. the appearance of the Moon — its phase, eclipses, tidal forces, etc — is mostly independent of where we look at it from Earth — and will be the same for both hemispheres.
But it's apparent path — 'noon' direction; rise/set angle; etc — is seen down at Earth level — it changes with the hemispheres.

Sun

Closer to Home

The solar system is also spinning flat, anticlockwise, but slanted — tilted like a top, in relation to the Galaxy, and faster.
Its 'top' star is currently in Draco, in the North. Its bottom star, in the South, is the larger Magellanic cloud.
Its 'Equator' is called the 'Ecliptic' — marked by the stars in the Zodiac, and the planets. In our lifetime we won't notice the Ecliptic stars change.
Our Sun circles the centre of the galaxy every 225 million years—heading toward Suhail.
To our sight, because we orbit it Eastwards, i.e. around its N/S axis, not ours, it swings seasonally, and also edges Easterly along the Ecliptic circle, revealing new zodiac stars to the West of it, and placing itself diametrically opposite a series of anti-Solar stars,
so giving us a star calendar to look for.
It also spins easterly, on its own axis, showing its sunspots for some days at a time.
The planets lie planar, 'wandering' (as the Greek word plan-et says), in the Zodiac plane, circling easterly around the Sun, over a timescale of decades, and only ever seen very close to the Ecliptic.
Mnemonic: Son, my very earnest mother 'as just served us 19 platefuls. Sun, Mercury, Venus, Earth, Mars, asteroids, Jupiter, Saturn, Uranus, Neptune, Pluto.
The earth is an inner planet, **circling the Sun**, easterly, so that the stars seem to out-run the Sun in setting West — star-time spins more speedily than solar time. Planets closer to the Sun than we are, are never seen far from the Sun, but outer planets can move right away from it.

Stars

The Moon

..moves mostly on the Ecliptic but might miss that mark by as much as 5°. This 5° slant causes a precessional change taking 19 years to cycle.
The Moon moves around the Earth monthly in an easterly ellipse,
and ever eyeing us, each full Moon, is the Rabbit, because the Moon also spins easterly, on its own axis, exactly once a month, so it stays facing us.
The Moon moves slowest to set Westwards, since it migrates eastwards in the sky each day. 12°½. Or **50 minutes later per day**.
And as it pulls the tides with it, they too are later each day, almost an hour, on top of normal locally fixed delays due to friction.
A full Moon rises and sets fully opposite the Sun, give or take 5°.
The Moon is most responsible for tidal forces, but the Sun plays some part too, reinforcing or diminishing the Moon's effect and changing the timing.
Tidal force attracts the nearest water the most, and the farthest water the least, resulting in two tidal bulges...
When the Moon is eccentric from our Equator, there is one big Northern bulge every 24 hours,
and a smaller high tide every 24 hours, as the Earth spins through the bulges.
There are equal bulges every 12 hours on the Equator. In between are the low tides, every 12 hours.
You should be able to tell which bulge is 'nearest' the Moon despite delays due to friction.
Southern latitudes will have one big bulge every 24 hours, and one 'small high tide' every 24 hours, offset by 12 hours from the NH.
Try to correlate the Moon's phase, declination and height with the local tides—timing, inequality, range—e.g. "The Full Moon rises with the tide"

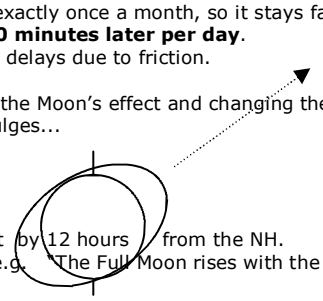

Moon

The Moving Horizon

Since it is the Earth which is spinning, not the Universe, it is **the horizon** which rises, in the West — where it eclipses the Sun,
so that the Sun only seems to set in the West.
The horizon slumps in the East, so that all the sky things — Sun, Moon, planets, stars — seem to 'rise' in the East.
Any spots on the Northern half of the horizon, circle to the east around Polaris — clockwise through the sky — i.e. 'seen from below'. Or, seen from above, that is from the North, looking South, spots on the Southern horizon circle anticlockwise around South Pole Star i.e. seen as if 'from the top' looking down on the Earth and its South Pole star. They hit the same stars each night, at the same angle, which is related to latitude,
so that the fixed stars none-the-less seem to trail 'small circle' tracks through our sky, and they always hit their favourite horizon spots when they set, always at the same angle,
and those same stars will rise later in mirror image to that around North/South, giving rise to **north-south-east-west finding by bisection**.
And Equatorial stars will rise/set due East/West.
The southern-hemisphere horizons move to the right, the northern-hemisphere ones to the left. Looking South though, from either hemisphere, it will be seen as the horizon circling anti-clockwise around the South-pole star. How come? Well, the southerners see the bottom of the star-circles, around an elevated pole-star, the northerners see the top of the same circles, around a depressed pole-star.

Clockwise

The Spinning Earth

The earth is **spinning like a top**, on its own axis, easterly, half as fast as a '12 hour' clock.
This gives us **the single most dominant direction clue on the otherwise uniformly round Earth—the North-South spin axis of the Earth.**
Spot its Equatorial Plane by the Pot. The 'top' is always pointed to Polaris. Spot the South Pole Star by the Seven Superstars surrounding it.
The spin generates a magnetic core more or less North/South — Find magnetic North in far northern Canada.
The magnetic field poles generate auroral displays, anywhere near the poles.
The earth is also tilted like a top, at 20 odd degrees.
The tilt gives rise to seasonal Sun height shifts, somewhat like a sinewave (just imagine the Earth's tilted spinaxis not changing as we orbit the Sun) and it turns on un-twinned top-of-the-tide bulges, twice in 24 hours. The Pole Stars are latitude° down/up @ 0°/180° T.
Any tilted top will precess the equinoxes — this generates a precessional gyration of 26,000 years,
but this time, retrogradely, i.e. to the West, along the Ecliptic (3° in 200 years) — i.e. the stars 'move' to the East through our star-map-grid.
Longitude has a lot to do with spin-time, seen from above the spin-axis. E.g. West longitude is how long after Greenwich, a place meets the Sun.
When we want to mark the apparent place of the Sun in the sky, or anything else, at any particular time,
we can just record the latitude and longitude of the point on Earth directly 'under' it i.e. closest to it.
But visualising that in reverse, we simply **project the Equator and poles, latitude and longitude up onto the heavens to grid them out**.

Revision, Overview and Explanations

Time is counted in cycles

The larger the distance scale of any motion — pulsation or precession, orbit or rotation — the larger the time scale of the cycle of change.
The Universe has been expanding its size and condensing its matter for many billions of years (and perhaps it will not 'cycle') It appears static to us.
Geological history spans billions of years, like the existence of our spiral arm. In our lives we see only a frozen frame of our spiralling Galaxy. It cycles, but we can't see it. The timescale of the rotation of the galaxy is in hundreds of millions of years.
In our lifespan we see only the slightest change in relative star positions, through telescopes, due to the 'proper' motions of the nearest stars.
It is in the nature of spiral turbulent motion that the larger rotation replicates itself in the smaller scale, but at a faster rate.
Each of the components of our solar system are also rotating and orbiting in the same rotation sense as the Galaxy. (Uranus is an exception)
For example, the Earth is rotating eastwards, orbiting the Sun eastwards, and has a Moon rotating eastwards and orbiting us eastwards.
"Anticlockwise, as seen from the North" is the single best spin you can put upon all that you see—including the N/S Earth spin-axis as 'the single best direction clue on our spherical Earth'. The tilt of the Earth's spin-axis gives us the year, and the Moon's orbit gives us the month, independent of our rotation. The Earth's rotation gives us our day. Star time outstrips solar time slowly. The Southern Cross clocks it correctly at the close of March, and the Centaur soon follows. The first point of Aries gains a touch over the stars, migrating 50 seconds of arc along the Ecliptic per year. The vernal equinox arrives 59 seconds late for its 366th "orbit", after 365 days, requiring 24 leap years in 100 years. 365.242 days per year.
Atomic time, used for astronomy, is independent of all these tilted rotations, orbits and precessions, which are all independent of each other too.

The Precession of the Equinoxes

The actual or 'proper' motions of the stars are only of cosmological interest, because they change relative positions only over centuries.
But history, actually our slowly gyrating Earth-spin axis, rewrites our star charts (**im**properly, as it were) with precessional revisionism, so that stars 'move' slowly easterly through our star-map-reference-grid—an Earth-centred way of seeing the sky — with 26,000 years needed to complete a full circle precessionary Earth-wobble. This artificial motion is called "the precession of the equinoxes"— which means that our reference grid readjusts just as quickly as our earth axis of spin precesses — by five arc minutes in six years. That's why we need an almanac to predict the 'fixed' star positions!! A good analogy would be the slow drift of the north magnetic pole, **if** we set our latitude and longitude by it — that would yield a false or 'improper' 'continental drift' in addition to the proper continental drift. See the appendix at p. 123.
To visualise the Ecliptic, consider the Equator in the heavens to be geostationary, although invisible, but the Ecliptic to be like a sinewave moving daily to the West along it, on the video display unit of the sky—as on a cathode ray oscilloscope—with a period of 24 hrs & an amplitude of 23°½.
The Sun appears fixed somewhere along the Ecliptic, creeping 'backwards' on it (against the setting sky)—to the East—during the whole year
To visualise the precession of 'the Equinoxes', imagine that *the points where the Ecliptic sinewave crosses the* Equator (the x-axis), move 'forwards' ('with' the setting sky) — to the West, very very slowly — during the millennium. At the moment these 'equinoctial points' are between Regulus & Spica (the Autumnal Equinox), and between Diphda & Markab (the Vernal or Spring Equinox, or the 1st point of Aries, (0,0) for the stars).
You might say that the Sun and the Moon and the outer planets all move West through our sky but East through the starry heavens.

Global Weather Patterns

The spinning globe makes the climate zones, and weather belts, follow the latitude bands. When the wind tries to blow, from a High, into a Low, in the Northern Hemisphere/Southern Hemisphere you see huge anticlockwise/clockwise spirals instead, by Coriolis force. E.g. the wind, tries to go straight ahead from the hot spin-equator to the cold spin-pole while the spinning Earth slows down beneath it (it moves slower nearer the Poles).
In the North the wind therefore seems to "move to the right" as we see it, and in the Southern hemisphere "to the left". Circulation is the result.
Note: Winds around a Low go against the Coriolis force. To visualise the pressure system winds at low level, imagine placing two flat-headed screws pointing up, one on Japan and one on Sydney, an ordinary right-handed screw for Japan but **a left-handed spiral screw for the Southern Hemisphere**. Run your fingers over the threads and a right-hand thread is like screwing a screw in or out, a left-hand thread is opposite.
High pressure system winds screw downwards and outwards; Low pressure system winds reverse that—up, converging.

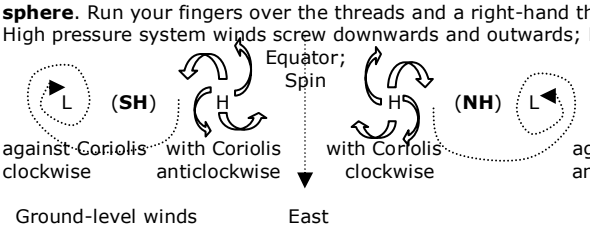

Higher level wind circulation reverses itself from that because the convergence/divergence is reversed.

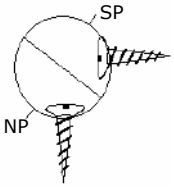

The Spherical Earth

The earth is curved round, like a globe, 360° round — 40,000 km around (by definition of a kilometre) 21,600 nautical miles, by definition.
◇ at 400 grads in a circle that's 100 km/grad ◇ at 360° in a circle that's 111.1 km/degree.
The Equator is a 'great circle'. Longitude circles are 'Great Circles' — passing through opposite poles, and perpendicular to all latitude lines.
Longitude counts 360° around East or West, or 180° East and 180° West from Greenwich.
Its North/South dividing lines mark out 24 one hour time zones of 15° each, and those North-South lines must converge on the poles, so that longitude degrees are of unequal arc-length at different latitudes.
Latitude has 90 equal degrees from the Equator to pole around the surface, which, by definition of the nautical mile, as one minute of arc, makes 60 nautical miles per degree of 'arc-distance'. 21,600 nautical miles must be the same as 40,000 km, so that's 10 km to 5.4 nautical miles. (2 to 1)
Latitude circles are 'small circles'.
We use them to divide the polar regions from the temperate regions (with the Arctic Circle, for instance, marking the limits of 24-hour nights);
and to divide the temperate regions from the tropics (for instance with the Tropic of Capricorn, which marks the limit of 'high noon' fully overhead).
'Polar Distance' complements the 'Latitude' and reminds us that arc-distance is real distance.

Moving the Horizon

At the spin-equator, on a globe, the polestars must lie horizontally, and the Equatorial spin-plane must pass East/West overhead.
But move north/south to another latitude, and the earth being round, the horizon — which follows you, like 'up' changing with you — must fall and rise latitude-degrees, so that one pole star falls latitude-degrees out of sight behind you and the other one, ahead, elevates latitude-degrees above the horizon. The Equatorial plane must also fall latitude-degrees, behind you as you move, from being overhead. Other star-latitude (declination) circles in the sky (e.g. star-trail circles) must also rise or fall the same amount — latitude degrees.
Now this same thing happens if you move in **any** direction, say ex degrees away, but for this let's just temporarily freeze the spinning of the globe.
'Horizontal' and 'up' must follow with you exactly ex degrees, and so the horizon now blocks out ex degrees of sky behind you, and reveals ex degrees of new sky ahead of you.
Looking side on, you will see the horizon **tilt** by ex degrees in relation to the old skyscape, lifting behind and falling ahead, of your direction of travel
Looking at the same star-time any night, will "temporarily freeze the nightly spinning of the Earth" for you, for star navigation.

Spherical geometry also means that if a place is 2° or 20° away from you, then it must also be 1° or 10° below your horizon — half the arc distance
E.g. the North Pole on Earth is ½ x (90° minus latitude°) below horizontal, and at 0° True.

END OF SECTION ONE

Navigation is More Than Navigating

Properly done, navigation is your chance to fit yourself in to the experience of walking through the land.
Why are you there walking?... Navigation zeroes in on the land. Do you want to see the landscape?... Navigation opens your eyes.
Don't you want to appreciate ecology and geography?... Navigate by them! Navigation stops you "drifting off" into mindless incompetence.
Do you want to remember the views and special places?... Navigation forces you to improve such memories.
So, achieve confidence when you are alone; tackle more difficult trips; lead others; call yourself "experienced"— navigation is the key.

Common Sense

Navigation skills invite you into the wilderness. The wilderness is a wild place! — a world of dangers. **Don't go**. But if you do...
..at least make sure your navigation skills are up to the task.
Likewise: Don't venture to sea with approximate star navigation techniques until you know exactly what you're doing.
Learn much more about standard astronavigation, before you experiment with new techniques.

Your Subconscious is Superior to Science!

Your natural abilities will do a better job at keeping you oriented. Your brain is a supercomputer.
That makes it subject to the "rubbish in; rubbish out" truism. Your job is to feed relevant data to your subconscious, not to feed, say, a calculator.

As far as important detail goes, your mind will 'go for the jugular'. Exact science will go for unimportant decimal places.

E.g. I can immediately guess "about 8 km away" far more quickly than any triangulation.
E.g. Will any numbers ever tell you how to recognise where you are? Or how to imagine what that peak might look like from behind?

So feed clues into your brain. It is portable, waterproof and teachable.

Don't leave your mental map behind.

Visual Assessment is Your Main Clue — Rely on It More

Since noone navigates well when they "can't see where they are going", we must be relying on visual assessment most of the time.
The idea is to run with that fact of life, and make sure you have a good look when you can
(e.g. from the lookout; while driving up; at the crest of a ridge; at the waterfall; on the bridge; in the riverbed)
Further, bring to conscious focus what is normally subconscious, in what you see.
And aim to remember what you saw. (Draw it; take notes; talk it out; use visual memory skills)

A What-to-Look-for Checklist: How far is it to the skyline? Where are the Sun, sunrise, Moon, wind and North in relation to the skyline?
How does the weather interact with the land — wind, cloud base, Sun shadow, wind shadow, cloud shadows, haze?
Talk the skyline through — name its features. What do you recognise already? What does it remind you of?
What landmarks are visible? What distances to landmarks? What sizes of hills? What is the strike or trend of the ridges, ranges & valleys?
What angles between landmarks? What alignments? What valleys will your current path take you through?
What lies in between? What slopes & shapes are showing, especially for your route? Can you see the path ahead? Is it curved?
What difficulties lie ahead? What easier paths are there? What clear lines of sight are there? What alignments will become visible?
Can anyone see where you have come from? Could you recognise 'here' again? Can you see your destination or departure point?
How many different vegetation types? How high is it to the skyline in each direction? How does what you see mesh with the map?

Integrate: Jetstream, Earth Rotation & Heaven Rotation

Imagine your global motion. Tilt yourself and try (futilely) to oppose the supersonic speed of the Earth's rotation, as if walking upon a treadmill.
Aim for a Big Picture 'feel' for each local direction.
The idea is to go outside and walk through each of these motions until you **know** what it feels like.

(Both hemispheres)

When you are.....	**standing still**	**or moving Westwards**	**or Eastwards**	**or Northwards**	**or Southwards**
(read down these lists, each in turn)					
The **Earth** rotates	Eastwards	rotates against you	rotates with you	to your right	to your left
The **Heavens** move	Westwards	slowly forward	quickly upwards	to your left	clockwise
The **high clouds** oppose the heavens outrun the Earth	jet Eastwards	zoom toward you	creep ahead of you	move to your right	to your left
You are being **tilted**	to the East	backwards	forwards	to your right	to your left
Lean oppositely	to the West	forwards	backwards	to your left	to your right

---

Measuring angles at 1 cm per degree

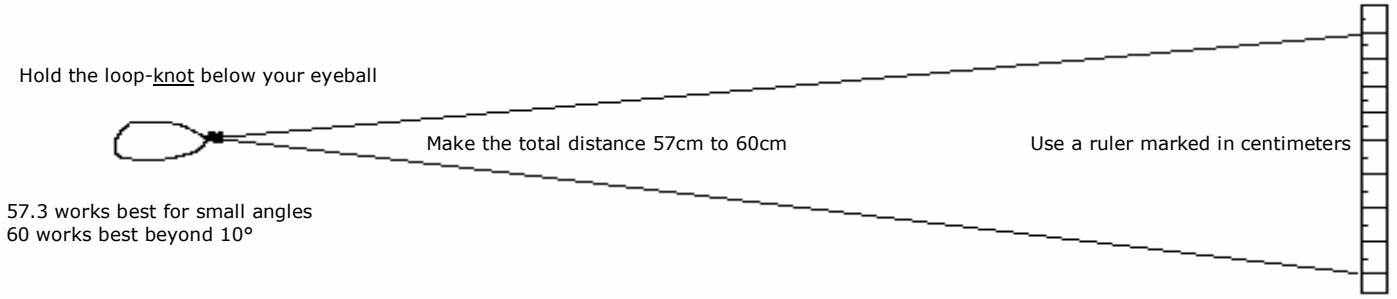

Orientation = Direction Sense

Oriental means "the rising" i.e. in the East (i.e. East from Europe!). Occidental means "the setting" i.e. in the West.
Because most ancient temples faced the midyear sunrise and there were no North-pointing magnetic compasses, and since all the world can agree on where equinoctial sunrise is, most of the rest of the world has used East, for most of Time, as their method of "orientation" — as their cardinal direction. Imagine putting East at the top of maps, so the Sun "falls down" the page.
Before magnets, East was the sensible choice for a standard direction.
This is no coincidence: "To the East" is also the watchword for the spin-sense of all our solar system parts. "North" and/or "South" is inferior
The direction of the *axis* of rotation doesn't tell you the *sense* of rotation; But the *direction of rotation*, gives you the axis of rotation *as well.*
East was the more sensible choice. You didn't lose orientation when you crossed the Equator; the Sun still rose in the East.
I recommend that you concentrate on this when changing hemispheres, and even now, to maintain your global 'orientation'.
Orientation means direction sense, direction finding, direction recognition, placing yourself in relation to over-the-horizon landmarks,
Or twisting anything (e.g. your map) to the right orientation.
Then again, it means starting off by getting your bearings straight.
Orienteering means athletes running through the bush.
Disorientation means... possible death.

The Navigator's Top 10 Problems; and Band-aid Solutions

1.	Not staying **alert** to navigation; not doing all you need to do……	**Talk** navigation; love navigation; practise navigation; **interpret** all clues
2.	Few **Clues** — E.g. Rain/Fog/Blizzard/Nighttime/Tall grass or dense vegetation/Look-alike landscapes	**Sit** still, until you can see; **go prepared** e.g. with a contour map; a compass; learn how to keep consistent courses
3.	Getting **Lost**	Remember the **start direction**
4.	Getting **Disoriented**	Be able to reorient; imagine the **Sun** in all situations; work top-down
5.	**Forgetting** the Way, the bends, etc.	Take a **notebook** and pencil; talk it out loud
6.	Losing track of **direction**	**Predict** the trend; use a **compass**; a notebook; integrate all direction clues
7.	Losing track of **distance**	Predict the timing; use a **watch**; use the map scale; and notebook
8.	Inadequate **Map**-Work	**Learn** how to use a map; study it first
9.	Not learning how to navigate	Take this booklet. **Practise.**
	Not knowing what to look for e.g. bush north	Practise being observant, at the expense of other preoccupations.
	Not **understanding** e.g. the lie, contours, heavens	**Teach** what you do know
10.	Choosing a wrong or difficult **route** E.g. a long "shortcut" across a dense gully	Get **directions** Don't 'shortcut'.

Reorientation

Reorientation is not an automatic body sense — it is an ongoing action.
"Orientation" in navigation is a misnomer, since it suggests some sort of initialisation, not to be repeated, rather than a continuous reprocessing.
"Re-orientation" is a more accurate description of what the navigator is involved in. Continuous reorientation.
Whether it is subconscious or conscious, it is something which must be done and redone, or else you are simply not oriented, now.
It is not good enough simply to be "good at navigation (theory)" — you must be putting it into practice, continuously, or else each bend will disorient you.
The essence of reorientation is conscious interpretation of the meaning of what you are doing.
This is, at the best of times, a bit difficult — like mental arithmetic — but it must be done, or else your brain will remain confused.
You may hope that direction sense would "become automatic with practice", but the last step of conscious interpretation must remain conscious.
And that is the process which should become second-nature. You must be aware, or you are fooling yourself into a false sense of security.
Delegating to your subconscious, is abdicating your responsibility. So bring yourself up-to-date, continuously, and...
Do not be overconfident. You are not oriented when you think you know where you are and where other things are—but only when you are right!
There are so many ways to make silly mistakes: confusing alternative systems; mistakes in numbers; forgetting your lessons; optical illusions; instrument errors; flawed theories; trusting approximations; misidentifying a star; making a logical blunder; and, of course... being overconfident.

Lost? Rule One: Stop! ..getting More Lost

STOP — so that you don't get deeper into trouble. ("When you get yourself into a hole, stop digging")
DON'T PANIC — Look for reasons not to panic "I can always walk all night, or sleep under that ledge,
or wait out the weather, or go hungry, or enjoy it…"
PRAY — He really does want to help the helpless.
SIGNAL — Shout 'Help!', 'Cooee', Whistle, Smoky fire, Heliograph, 3 of anything, SOS, CQ, Fire at night, Flashlight
LISTEN — for a reply.
THINK p. 37d
RETHINK — your journey
SIT It Out — = The Simplest Way to Beat the Fog. Gather fuel for a fire. Don't try to move — you may not win.
CAMP — choose a place that will be a good 'base' for further exploration.
SLEEP ON IT
STAY — in the area, where you can be found, e.g. on the track, at a hut or intersection.
PLAN — Implement a pre-**thought-out plan** e.g. Stop, Think, Navigate, Camp, Retrace, EPIRB, Stay.
For a Start: Slow down — almost to a stop. Be slow, cautious and thoughtful, not impulsive.

360° SOS
Three quick turns
three slow turns
three quick turns
(with the stick vertical)

Point and Blink = 10% Shift

In the field you simply point and blink—If you focus in the distance, the shift is about 6 degrees, which is 10% sideways.
6° or 1 in 10 is a very useful angle to guesstimate. Point with one eye and finger, then swap eyes, to see what a 6° shift looks like in the distance.
For repeatability you should touch your two index fingertips as you point.
For accuracy, use a pencil tip, and know the exact figure. The instructions below will show you how to hold the pencil to get an exact result.
To put the Point and Blink method on an accurate footing, measure your Inter-Pupillary Distance at Arm's Length.
Stand behind a window, pointing through it at some distant mark on the horizon, with both your finger-tips just touching the windowpane.
Use one eye and the same-side finger; Press your fingerprint onto the window; Swap eyes and fingers, back and forth, moving only your eyelids;
Check your alignment—both your fingertips should appear to exactly overlap when focussed on the distant object. Press your second fingerprint onto the glass; Measure the separation carefully.
Because the pupil is not pinpoint, it matters that you measure the practical separation at the working distance. Measure that working distance carefully— from the corner of your eye to the window when your pointing fingers just touch it.
Compare the two distances; i.e. divide the interpupillary distance by the armslength; **expect about 1:10** or 6°; Find the arcsine of that ratio for better accuracy. Try it out on the length of the Southern Cross. Compare it with 3 fingers, and 4 knuckles, and your binoculars' field of view.
Measure again on a middle-ground object, as when measuring the sideslip of a yacht by pointing at the wake. Measure out the distances as a check

The "Good" Navigator. You.

If you like exploring, you can use that to advantage. Or maybe you are a good photographer, well then you will understand 'mental photos'.
Use your creative writing streak, in creating memorable names, or else, appreciate the action 'direct', carefree and spontaneously.

Perhaps you are the obsessive type — well then, it's map and compass detail which will suit you.

So whether you are the artistic type, using your visual dominance, or the cool calculating logical sort,
or maybe a 'thinker' doing a good job at planning, or a 'doer' doing, you already have many useful skills for navigation.
Don't talk yourself down.

Those handful of useful traits which are already within you, you can develop into an orientation sense.

Think out Loud

Your subconscious will be telling you "That's odd".
Your conscience will be whining "Something is wrong".
Your mind will notice, before you do, "That's new".

When you notice something talking inside you, you need to talk back. Out loud. Don't leave it 'inside'.

Internal Problem	Antidote
"Shouldn't we be there by now?"	Ask the others.
"Have we missed the junction?"	Discuss it.
"Why is the Sun over there now!"	Name the lurking enemy!

You Need Excuses to Stop and Have a Good Look

- If it is a local peak, say "Let's look for a minute while we have the opportunity"
- Declare a chocolate stop!
- Take off your pack, adjust the padding. Adjust your socks, air your feet, put on a Band-Aid.
- Get out the maps; Identify the landmarks. Draw the relative angles to other lookouts, for future reference.
- Get out the camera and tripod. Unpack your pack on any pretext — noone can move until you repack!
- Make a sketch, compose a poem. Start an interesting conversation.
- Go for a swim or wash some clothes.
- Cook your main meal at midday & keep the biscuits for the late meal.
- Camp there.

You Can't Easily Revisit Places Left Behind — unless you first spend memorable times there.

Visualise Familiar but Invisible Landscape in Relation to the Sun

You need to 'fill in the void' between the very distant reference, and what you can see, with what you know, but can't see.
I.e. When you get a clear view, mentally plot and remember the sky-to-land layout for later use. "If the hill disappeared, the hut would be *there.*"
Mentally reconstruct any view which is helpful. E.g. as seen from camp, or from the start, or the peak…
Concentrate on knowing and revisualising landscape direction trends. "If I were at home, I'd see the coast… *across there*"
Do this at various times of day, **paying particular attention to direction *trends* in relation to the Sun's path**.
After all, that's what our direction sense consists of—a confidence in the real directions just out of sight, based on what we can see.
Then, when you are 'bushed in' and you see the Sun through the canopy you can say…
"Ah! the road is running sideways to the Sun, the valley lies entirely off to the right, I must have swung around left, parallel to the ridge…"

Section 1 dealt with the Sun/noon/North in the global abstract—unrelated to the local detail.
Section 3 deals with the local detail in relation to the Sun.
"Land to Sky" (soon, p 38d) is specially for one time of day, and concerned mainly with particular landmarks, not whole trends.
But now catch up with the overall game—**the Sun's path, the local skyline, and the local layout are one piece for a substantial local area**,
and you need to mentally integrate them, and carry them with you, to avoid disorientation.

Guesstimating Angles at 1 degree per cm

First find how to make your **arm's-length 57 cm**, e.g. by touching your fists together in front of you.
Next, start measuring in **centimetres**.
Find a 1cm wide fingernail—1°. Check that your fingers are 2° wide, for 2°, 4°, 6°. Are your knuckles 2 cm apart? 2½° for the big ones.
A small fist, with the thumb tucked in is 9°; 10° with the thumb showing; a palm is one 'point'.
15° and 16° may require experimentation. 20° can be spanned with the fingers.

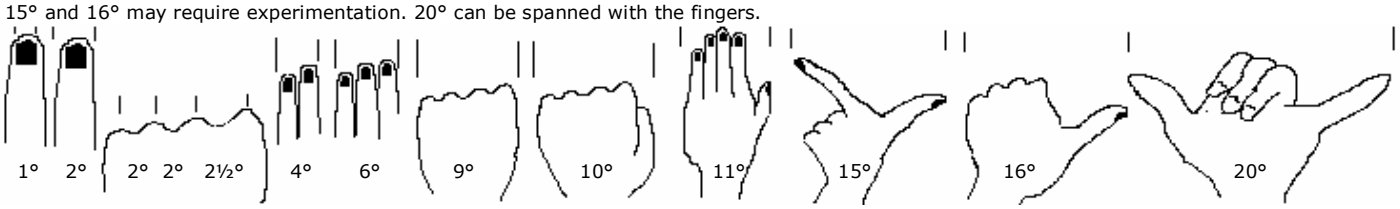

The 'Blink of eye' method also yields 6° — point to something with your finger at arm's length or with something narrower, with one eye, and blink.
Larger horizontal angles can be visualised from between your feet of from overhead.

Initial Orientation is the Single Fundamental — The Get-Back Path

"Which way do I go?" is more important than "Where am I?/How far?/Where is North?"
"How do I get back to a familiar path home?" is the real need. That means you need to orient yourself **before** getting lost.
The single most important direction is **Start to Finish** i.e. Which way did you **leave** the car park? **From** what recognised point?
"Which way have I been **going**?" (if I had to backtrack)
Which direction did you **head off** in — **toward** what identifiable point?
E.g. "Downhill". The direction need not be horizontal. We function on 'paths' more than on straight lines.
E.g. "to the lake". The direction need not be relative to North; It does need to be something you can reverse.
E.g. "with the Moon ahead on the left" "the long way around" "to the opposite side" "away from the mid-morning Sun"

Taking a map and compass is good; A watch is nice; But the only really necessary bit of navigation you must take with you on a walk, to get you back when lost or disoriented, is... what direction you have travelled out.

Initial Orientation is the Single King Principle — and the rest may not be much help without it. See the mnemonic p. 111c for some help.
Which way you went is also the single biggest help for searchers—**tell someone** where you are going!

How I Manage to Get Lost or Disoriented

… **starting off** without paying attention … blindly following the track or the leader … taking a minor junction
… choosing a hard route — thick vegetation … going without a compass, or map … trapped by a peak or clearing (p.45b)
… falling asleep during a journey — a gap in my reckoning … waking up in a new place … not noticing the "end of track" marker
… inside a hospital complex, or a shopping centre; coming out … in look-alike landscape e.g. suburbs; losing the car on city streets.
… getting out of bed the wrong way or entering a town by a last-minute bend … talking; inattention; not noticing bends & junctions; losing track
… misinterpreting "directions"; forgetting them; being given wrong directions … hurrying; "short-cuts"; leaving the track
… on river-flats near river junctions and bends … preoccupied — with map-making! or any other task
… inadequate memory e.g. in caves, the way back, the many bends … stubborn psychological disorientation syndrome
… miscommunication between parties e.g. about how & when to meet … confusion; blunders (see list p. 39b)
… not enough knowledge of what to look for & how to interpret what you do see … visiting the Northern Hemisphere
… minimal input technique of navigation e.g. 'just follow the signs'; overconfidence … misinterpreting the map—especially the scale
… new tracks, not on a map … false identification of landmark features … losing confidence just short of my goal
… unsure of distance covered … navigating by creek junctions or minor dirt roads … blinkered by fog, darkness or thick forest
… wrong turnings; walking in an unintentional circle; overlooking junctions; not looking back … waking up on a dark night, 180° disoriented.
… too much new territory too quickly … negotiating thick bush without keeping orientation

---

Voice Your Thoughts — Don't Suppress Important Clues

When your silent auto-pilot notices something and tries to tell you… Say it out loud. That's the trick to it.
Tell someone. If you are alone, yes, talk to yourself. Bring it to light.
You can't remember the jigsaw bits later, unless you look at them now! **Subconscious thoughts flit away like dreams**, so express them first.

"Oh I didn't expect to reach this so soon!" ? Misidentified
"North seems to have changed. It shouldn't be out there!" Disoriented!
"Hasn't the wind swung? Does anybody agree?" ? Are the others asleep!!
"The shape's not quite right" Most probably because it's wrong!
"The markers have stopped" ? Off the main track

The same goes for run-of-the-mill clues. A forgotten clue is a missing jigsaw piece when you need it, so say them all out loud.
"We're crossing a divide — see — we are now going downhill"
"Creek number... four" "Marker!... Blaze!... Cut Branch!..."

Back-track in Your Mind before You Back-track on Your Steps

Lost?
"Don't Be Afraid to Think"
Spend time rethinking.
Try to reconstruct where you might have erred. Have I overrun my goal, or am I not there yet? Or am I simply off course to left or to right?
Did anything 'not make sense' on the way out?

- Work backwards from now, before you forget.
- Assess how successful back-tracking might prove.
- Don't ask "Where do you think…" but shoot for the details — ask "Why do you think…"
- Each of you should draw a mud-map separately, then come together, to compare recollections. Argue out the sequence.
- Back-track on your presumptions.

Detail the facts with a cool head and many problems will dissolve.

Distance using Apparent Size & 6°

The visual size of people, cars, houses, as measured at arm's length, varies with distance. You know how big they are already.
Blink your eye or otherwise guesstimate what fraction of 6° it makes, and multiply its size up by 10, plus for the fraction, to get the distance to it.
E.g. something one knuckle space wide is 3 times further away than if it were 6° big, and therefore is 30 times further away than its absolute size.

Visual Angle, Size and Distance

Find the angle, and multiply the size up by 60 in 1-per-degree. E.g. If something 1m wide looks 5° across, = it is 12m away. 0.5° means 120m.
Find the distance from the map, and work out the size from the apparent angle, by 1 in 60. Something 10km away, 10° wide is 10/6 km across.
Using 7/4 in 100, for accuracy because the angle is small, something 10km away and 10° across should be 1750m wide.

Side-wise Offset

Blink your eye at something in the distance, and it will shift across the background skyline.
Walk sideways, pacing out the distance, until it lines up with the new skyline mark. Multiply by 10 for the distance to the something.
Alternatively, mark the new position alongside the object, which the 'blink' indicates, then walk up to it and pace out the offset distance & ten-tuple

Don't Just 'Get By' and 'Make Do'

We all have a handful of techniques we feel comfortable with. We all get lost.

Most "How to" books necessarily are telling you how to battle uphill — against your natural incompetence.
This is doubly true for a "How to Improve" (your navigation) book!
E.g. 'Improve your memory skills'! Easily said; hard to digest.
Get used to the idea, and therefore embrace it — the idea that you, as a good navigator, should want to extend your familiar range.

Navigation is inherently a problem (of not being able to see around corners; of not having a magnetic sense; of not having excellent memory; etc.).
However humans inherently do have the potential skills to cope. But do you have the guts?

Memories are Made of Landscapes, Not Compass Needles!

What to look for is detailed later, but for now consider...
How to 'look' …

Use your compass and map vaguely! Approximately! Lightly!
It is safer to be aware of their meaning, rather than to be accurate but disoriented.

Imagine you had no map, no compass. No signposts. No highways.
"How did the pioneers do it!?" You learn how to not get lost; how to 'see' what you are looking at.

Once you get oriented, then you can bury yourself in the instructions on decimal places, for accuracy when you need it.
Otherwise numbers and recipes have an ability to preoccupy your mind and numb your thinking.

Treat your map and compass as shy — self-deprecating — and turn your piercing stare away from them to what they are pointing out to you.

You Need a Road-Map To Climb a Hill!

It is a waste of effort to realise your mistakes at the top of a hill.
"I wish we had brought binoculars" is all too common a regret.
"We must remember next time to bring all our maps"

You have few such opportunities to positively identify landmarks,
and so you should **take the chance to familiarise yourself with what each hill looks like from every other hill.**

But often, it is only your smallest-scale map — the roadmap — which will show you the big layout.

"Didn't anybody bring the compass??"

It's up to you to plan ahead, to plan on making the most of your efforts.

'Land to Sky — Sky to Land' Navigation

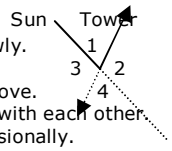

Step Zero: **Leave yourself out**, and your compass, of the scheme of things — so that the whole lifeless picture only changes slowly.
The result is simplicity itself — You see **how the Sun shines across the landscape**. Remember the angles.
Step One: "The Tower is ↖ by angle, from the Sun as I see it" Both change only slowly (or rather, seem to) as you move.
The Big Bonus is: You can keep track of all your bends, twists & turns, by the Sun & Land working in conjunction with each other.
You can locate a landmark easily through the vegetation — you know where to expect it when it shows itself occasionally.
i.e. You can always easily locate the tower/peak/etc, from kilometre to kilometre, in relation to the (now slightly shifted) Sun (or Moon, wind, cloud)
Step Two: "The Tower is ↗ from my shadow" This allows you to walk away from the Sun, yet still stay oriented to the landmark.
Step Three: "**The Anti-Tower** is ↖ ..from the Sun" These allow you to walk away from the landmark,
Step Four: ↙ ..from my shadow" yet still "see" where it is.
Step Five: Mnemonic: "The Landscape on the **Left** Lingers Longest in the Light" — its apparent movement as you walk, follows the Sun
& the Shadow's anticlockwise shift. (SH only) i.e. you don't need to worry much about time passing, and angles changing.
"But Landscape on the Right… soon goes wrong" — it 'moves' against the clock-sense of the Sun. (Southern Hemisphere rules)
i.e. you need to **refresh your angles** more often during the day. (NH: Left Loses the Light; Right is all right)
Step Six: Rule of Thumb: "**15 km away** changes at 15°/hour (i.e. at about the same rate as a low midday Sun) as you walk past it at **4 kph**".
The Sun shifts at less than 15°/hr when it is rising or setting — so try "20 km away/12°" to match it; and "10 km/25°" if the Sun is very high.
Yes, you should convert the angle to "ESE" or rename it "just S. of sunrise" for later, but **the idea is to use the Sun as it is now, not North.**

Grid-line Walk — Explore in a Square Grid Pattern

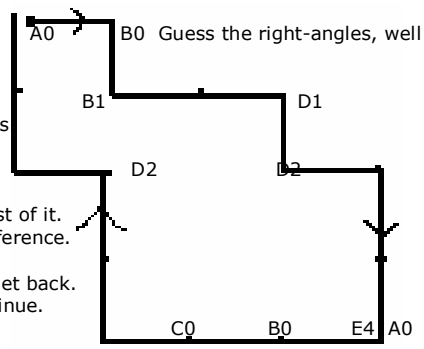

This only adds 20% walking effort, but lets you know where you are, and how to get back,
and along a route of your choice, among many choices.
You need to judge a given distance, well, e.g. 100 double-paces; one minute; or 100m; or by pedometer.
Choose a **reference direction** as 'Up' or 'ahead'. Sun-North; Compass Needle North.
Up-weather by clouds; down–clouds. A strong wind. Distant Hill. If needed, keep a straight line by marks
Choose a **scale** to suit your ambitions, e.g. 100m squares (so that each 10m is an extra 0.1).
Try to cover a whole square at a time. Use 0.5, if needed, or you can cross at a diagonal, 140m long.
Count **forward and back progress** in whole chunks, as A-Z. G.5 means 7½ squares.
After Z comes AA, AB, AC. See p84b for making suitable mnemonics. **Don't lose count!** That's the worst of it.
Count **right and left progress** as +/- 0-10, etc. Minus simply means 'backwards' or 'to the left' of the reference.
When you have definitely decided to head back to base, remember the grid-reference. Stop counting.
Reset the count to Zero. This is easier than trying to count backwards, using subtraction, to get back.
Reverse the reference direction, e.g. to downwind, shadow South, compass South, and continue.
When you arrive at the original grid-count, you should be within sight of home-base.

Disorientation is More Dangerous than Being Lost

…because the same disorientation that will get you lost, may then get you more lost.
Orientation is primarily a mental picture — therefore **— Disorientation is a wrong mental picture.** It is one and the same principle.
So Disorientation really extends to cover more than direction sense, into bewilderment and logical debilitation!
You can be disoriented, deluded, bewildered and irrational about more than direction or position. E.g. You may misjudge which way is up/out/back.
E.g. you may be mistaken about the time of day, or the existence of a road, or whether the promised weather front has passed you or not.

'Lost' is 'I don't know where I am' which is different from disorientation, since position sense is independent of direction sense.
Again, you can know where you are on the map, but still be bewildered on the ground — not 'fitting in' even if direction-wise oriented.
Then again, you can be oriented, but lost — yet not necessarily worried or in difficulties — because you are oriented.
You can be disoriented unawares — even when you know where you are — i.e. not necessarily aware of the danger of loss of direction sense.
To be both disoriented and lost is most dangerous, but **it is orientation which stops you from getting lost**.

Orientation keeps your finger on the pulse. It stops you getting disoriented. In many ways it gives you a position sense too e.g. that you are in a particular catchment, downwind of a known landmark, up-Sun of the overall course. Orientation of the map puts the map onto the landscape—aligned to reality—so that the map can point out its riches to you. Orientation lets you know where you are, recognise places, & know the way back

---

Logical Blunders — Examples

Mistakes of category: "The last stream was flowing to the left" (Oops!)
"We have crossed the main road" (Now, the main access road is dirt, whereas the main through road is bitumen, but you muddle them up. Or maybe you are unaware of a fork in the main road.)
Mistakes of number: "We have passed two lakes" (which were joined behind the intervening vegetation).
Mistakes of quantity: "We haven't changed direction" (except for one hairpin bend!)
Ironstone may have swung the compass, when you put the map and compass on the ground 'for extra certainty'!
Mistakes of quality: "From the peak to the lake was 090°" — but the false peak you were on, gave you a false sense of accuracy, in the mist!
"See those houses 'across' the river?" — Rather, see the bend on the left — they are on this side!
"The island should poke out from behind the far point. Why isn't it there?" (It is 'the far point', of course)
"But we have crossed the River" No, only a main tributary, or only onto an island.
Gaps in knowledge: "We haven't seen the junction" — because you short-cut past it; around one bush is enough!
Unexpected: E.g. there are sometimes two parallel roads — the old & the new. E.g. a stream sometimes goes underground "Mole Creek".
E.g. A water canal may cross catchments. E.g. "Remember the creek junction at the base of the hill" — The 'hill' was only an island.
Assumption: "That will be the ridge we want up ahead, let's turn off here" — Unproven!
"Ah, we are back to the road!" Yes, but not the same road!

Trust the Untrustworthy

It's like people. 'You can't trust anybody', but you have to trust your life to them sometimes.

A compass has dozens of potential faults, but in a fog, you trust it, implicitly.
You yourself are less trustworthy than a compass.
When all you have outside your own opinions are reasonably reliable external clues (like where the Sun is!) you had better trust your life to them, not to yourself.
For instance, sometimes that 'external clue' is someone else telling you that you are disoriented, when you don't think so — you had better listen!

A disoriented person may not want to trust the 'untrustworthy'. Don't let that be you.
E.g. You may find yourself twisting the map to make it line up with a mountain which doesn't really line up, except to your expectations.
You prefer to believe that the map may be distorted!
It takes practice to believe. I once spent a whole week wondering why all the trees were pointing too much to the sunset side of noon.
I was trusting myself, not the trees, and was simply wrong about where North was. I didn't realise which was the more trustworthy.

Lost? Rule 2: Start! Getting Unlost. Find yourself. Regain confidence.

(After you STOP getting lost — Rule 1) It is time to **START NAVIGATING** — Rule 2. (You can practise this ahead of time, without getting lost.)
E.g. What can I see & recognise? E.g. Can I interpret the wind or sky or trees or drainage for direction?

- First: Mark your spot e.g. tie up a hanky. Mark your Trail—so you can get back here, as 'base'. Retrace your steps, if that is the best option.

Act out a Search & Rescue, looking for your 'lost' starting point. (If you can see your surroundings, you don't have to feel 'lost'.)...

- Explore in all directions systematically.
- Make a map. 'You' in the centre. Locate yourself in the landscape. Climb a high point, to orient yourself; at night too, for lights.

Let others find you...

- Leave notes about your plans. Especially at 'findable' places e.g. a hill top.
- Leave clues so that searchers will not give up for 'lack of clues'. Make some marks visible from the air.

Search out a track...
Ask: "Where should there be a track?" E.g. to... that peak?, a fishing spot?, any low pass through the range?, the main access valley?

Mental Grid Map

This is the square grid method, where you simply walk, along a path of choice, and rather than following the grid lines,
guesstimate your decimal position in the grid scheme.
The beauty of it is that it constantly forces you to orient yourself to a reference *direction*, to notice *distance*, and to be aware of *position*,
yet it takes nothing more than mental equipment. You should aim to practice it properly, once, and you may not want to discard it.
It causes less mental strain if there are two or more of you. **Delegate** the left/right to one person; the forward/back to another.
The one person guesstimates distances by one method, their way; The other gauges perhaps a different distance and with different accuracy.
It doesn't really matter if your guesstimation mismatches, provided you remain consistent and imagine a rectangular grid.
In addition you need a feel for triangle lengths, for a 1 in 10 slope, up to a 10 in 10 slope, so you can guesstimate decimals.

~You need to divert 5 in 10 to add more than **10%** to the length walked. E.g. from Z0 to A0.5 takes 112m across 100m squares.
~Maximum extra length is a 45 degree diagonal, at **141%**. E.g. from Z0 to A1 takes 141m. These two set the feel for guessing distances
1 in 9 adds one third. For 6 in 10 to 8 in 10, triple the first figure as the extra %. E.g. from A0 to B0.7 takes 122m.
If in doubt draw a 10x10 square, draw in the diagonals to each of 18 opposite points, and label the lengths as the square root of $100+n^2$
This lets you see the direction of each n in 10 slope, and know its length.
~You could use degrees, up to 45, and call them percent extra distance, with fair accuracy, but there's no need for such figures.

When I simply guessed, by eyeballing distances and directions, it brought me back home without noticeable error, after a few kilometres.

Walk Where the Navigators Talk — if you want to learn navigation

Not out of earshot.
Navigation is a learning experience.
It works better when you can share the process of learning with someone.

Explore

Explore the unexpected — the side tracks. You walk to explore, usually, not to exercise.

It's No Fantasy

Navigational know-how is not "If I were lost I could do this…" but is "What I am doing right now, all the time, will stop me getting lost".
Learning new skills and how to interpret all the clues is the practice of navigation here and now.
Continuously pointing out new clues to yourself, and to others willing to listen, is the process of not losing yourself.

Feed Your Subconscious through Your Conscious Focus

E.g. To improve distance guesstimation, focus on the finest visible distant detail. Notice the haze level, the shadow hue, the sizes you can see.

To train your direction 'sense' (which is entirely lacking in humans) feed it all the clues you can —
Sun & shadows, wind & weather, compass & landmarks, trends & bends… Don't just look at them! Put words and meanings onto what you see.

To extend your memory of a sequence of turnings… use pencil and paper 'memory'.

A working rule of thumb: **Bring to conscious note**
what you hope to keep track of subconsciously.
Don't expect to be able to remember what you haven't noticed or thought about consciously.
So learn what to look for, and how to see it.

Look into the Far Distance

..for Trig. Points; Towers. They are there to be seen, by definition. ..for Roads; Houses
..at night, for City glow; Auroral Lights ..for the dip and strike of the underlying geology
Assess the difficulties ahead, whenever you glimpse them e.g. a steep bank on the other side of a lake. Use binoculars.

Look through the Haze

What are you looking for? You want to pick: intervening ridges, hidden gorges, distant cliffs, overhangs, false peaks, mountains or islands, the Moon, and the trail ahead. Those cloudy lumps are not always clouds. Those invisible peaks are there to be found.

If you don't see them right away… **Wait** ..it is the **changing light conditions** which show up the contours.

E.g. Look at night, when the whole emphasis shifts, and light sources show up.
E.g. Wait for the haze to shift — it moves with the Sun.
E.g. Wait for the haze to thicken — to show up the valleys — weather is changeable
A rain shower may show up a valley; a cold front might clear the air of haze.
E.g. Wait for a cloud shadow or patch of sunlight to drift across — it will disappear into any hollows, and highlight any foreground.

Join the Invisible Big Picture To the Visible Landscape

The Big Picture might be invisible — e.g. the South Pole Star is "Over the South Pole in Antarctica" — but geographic "South" is quite different and is visible — "that knoll there!". Join the two!
The first principle to practise, is to work top-down; but the second and more important need, is to integrate everything.

When you put a continental or global picture onto the landscape, you have a full-scale map of the world in front of you.
e.g. Point just beneath 'that Knoll' and put your imagination into gear...Down over that horizon are icebergs separating from glaciers with a huge splash! Can't you feel the chill? Up there is the night-time aurora. Down there, anti-polar bears look straight up and see the Southern Cross whirling around overhead! And we must be on the south side — the cold dark side — of this hill behind us.

You say: "The destination lies beyond that valley." "Mount Saddleback marks sunrise"
"The track winds around behind this hill." "The Sun skims those treetops"
"Porpoise Bay must be yonder; Melbourne that way; and my girlfriend, 50km in that direction…"
Integrate as you walk — in real time and place

Cyclical Motion Divided into 12ths

Any back-and-forth motions or cycles which are caused by circular orbits or sinusoidal oscillations are easily understood and followed *in units of 30°*.
This is because 30° of rotation will swing *half way across* a radius from top-dead-centre, or half-way up/down from right/left extremities.
E.g. quick rising for 4 twelfths, then *2 twelfths* in the *top half* of the swing, 4 quickly reversing itself, and wallowing for 2 at the bottom.
You may not *see* the 'circular' nature of a one-dimensional oscillation, yet it will still conform to this rule.

Tidal Height	Sun's Seasonal Latitude (or Moon's monthly position cycle; use 2½ days)
1 hour falling slowly from the top	1 month at Northern solstice slowly coming South
2 hours dropping rapidly to midway	2 months rapidly moving back to the Equator — to the South
2 hours dropping rapidly from midway	2 months rapidly moving South of the Equator
1 hour dropping slowly to minimum height	1 month slowly approaching Southern solstice
1 hour at bottom rising slowly	1 month slowly turning back to the North
2 hours rising rapidly to midway	2 months rapidly returning to the Equator — toward the North
2 hours rising rapidly from midway	2 months rapidly moving North of the Equator
1 hour rising slowly to top	1 month slowly approaching the Northern solstice

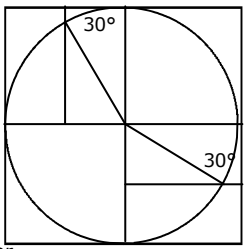

You can simplify that into sixths, saying: "1 sixth *at the extreme*; 2 *on each change*; 1 at the other extreme"
Or say: "*twice as long on the change*". That makes 'the change' sound 'slower' but it actually goes *faster* and *for longer*.

Which Way Did I Come In? Which Way Do I Go on?

Have you ever interrupted your pleasant journey at a roadside picnic area and driven out the wrong way — back the way you came — and not onwards? Or you turn left instead of right when you drive "out" of a local dead-end network, into another dead-end maze?
Or you sat down for a rest on the track and when it's time to go, both directions look identical to you?
It's embarrassing on the road. It's nerve-racking in the bush! You feel struck with insanity. But it's normal enough.

Or when you come back to visit a familiar place, it's on the 'wrong' side of the river! Every time you come back?!

An extra external clue, like the Sun, is needed, and needed to be believed, but it is really needed before you make a fool of yourself.

And you may have to undergo 're-orientation' to dis-illusion yourself.

Don't let it happen to you deep inside a cave system!

Logical Presumption Leads to Utter Delusion

Think about those example 'logical blunders', about your own fallibility, about the dependence of logic upon its basic assumptions.
Being twitted by a logical mistake leaves you with a horrible 'lost' feeling, plus a pathetic 'helplessness' about how to reorient.

These are not simply mistakes of guesstimation, subject to fine-tuning!
Once you subconsciously make a decision about how the world is, then your mind will close to the possibility of error.
False assumption is presumption, so I say "**Presumption Means Delusion**"
Such Left/Right, Yes/No, crossed/uncrossed, 2, or was it 1? -type mistakes lead to utter bewilderment, so I add
"and errors involving discrete categories can be especially indiscreet!"

Being aware of the phenomenon of logical blunders, is half the battle in overcoming them, when they bite.
Keep an Open Mind about the truth of your assumptions!

Trust Less and Prove More

Don't trust yourself At numbers
Or remembering accurately
At judging time or distance
Or walking in a straight line
At identifying landmarks properly,
Or walking back upstream, and so on.

Don't trust A compass to be dead accurate
A map to show North reliably
The weather forecast
Old track markers
A navigational 'fix'
Complicated or finely adjusted instruments. They can easily go out of adjustment, without letting you know.

The point is not so much to **double-check** and **cross-check**, as to: **back-up** with an alternative; **interpret** for sense; understand; and **be wary**.

The Simplest Way to Get Unlost is to Retrace Your Steps

The sooner you go back to where you went wrong, the sooner you can relax. Yes that might be uphill. Yes it might be a long way.
"Oh I wish I had retraced my steps earlier, while I had the chance!"
You still have the chance, you probably just don't want to.
For instance, if you have been pushing on, hoping to find the way, you should be doubly aware of exactly where you have come,
since you doubly anticipate the need to retrace, if you fail.
"Short" cuts have a habit of surprising you at just how long they can turn out to be, repeatedly.
A track naturally avoids gullies and obstacles, so a 'shortcut' is almost guaranteed to hit one square on!

Retracing your steps may be something you can't short-cut. It is the shortcut. It short-circuits many other possibilities.

Caution: Sometimes retracing your mistakes will only compound them—more of the same!
You may sort yourself out better if there is an alternative system to try e.g. to walk back upridge if you got lost going downstream.

Approximate Sine-Wave Values 97% accurate

(From school you could have memorised that sin 30° is 0.5 and sin 60° is 0.866, and perhaps sin 45° =0.707 ???)

The first twelfth of an oscillation takes something half-way (0.5) to its peak amplitude (1.0) from the middle ground (0), by the rule of twelfths.
~*The motion is virtually linear for that first 30 degrees*. So to convert those *first* 30 degrees to a sine-value, **Divide by 60**.
I.e. add at a rate of 0.5 in 30, or 1 in 60. E.g. 25°/60 = 5/12 (or 0.417). Sin 25° is actually 0.423, which means we had a –0.006 error.
Sin 15° is 0.259, but we get 0.25 (–.009 error). Sin 20° is 0.342, but we would guess 0.333. That's about the maximum error, 0.01 low

The next twelfth carries the motion from half-way to 86.6%, in a curve; The extra is 36.6% or a slope of 1.22 per degree. Let's use 87,36,1.2
A linear fit cuts the curve twice, and so is close to most of it. If we could add 36%, over 30°, to 0.51, to arrive at 0.87, we get +/– 0.01
~To the ***extra*** degrees, **Add 20% ; Add 1; Call it a percentage** (divide by 100). (*To add 20%, add 1 tenth, 2 times*, e.g. 55 + 5.5 + 5.5 = 66)
E.g. 60 degrees = 30 plus an extra 30; so start with 0.5 and add 0.3 then 0.03, 0.03, 0.01 to get 0.87, which is 0.004 too much.
E.g. 40 gives 0.5 + 0.1 + 0.02 +0.01 = 0.63, instead of 0.643, or a –0.013 error. Sin 45 is about 0.5 plus 0.15 + 0.04 =0.69 or –0.017
To make the next approximation, use 0.87 for the first 60 degrees, then worry about any extra degrees.

The last 12th is the flattish top of the curve, building up to the crest. It climbs only 13.6% from 0.866 to 1.00 in 30 degrees, in a curve.
We start at 0.87 and finish at 1.02 cut the curve twice and best-fit to within 2%.
~For the last 30° degrees **Divide by 2, then 100**, to gain 15%, from 0.87 to 1.02.
E.g. 80 degrees moves to 0.5 then to about 0.87, then an extra 20 degrees or 0.1, yielding 0.97, 0.015 below the real value, 0.985.
75 yields 0.5 + 0.37 + 0.075 = 0.945, 0.021 below the true value 0.966. 70 gives 0.87 + 0.05 = 0.92, 0.02 below 0.940. I.e. +/– 0.02

Self-Motivation is Fundamental

For a sense of direction to become "second nature" you need to **train** your second nature.
Improved navigational ability boils down to… *putting new ideas into practice*.
(Not perfecting old mistakes! Nor leaving the right technique on the shelf)

Don't just know about a good idea; make it yours.
For this you need to want to improve your skills.
to force yourself to practise your weaker points.
to push yourself to develop new habits.
Which is fun.

Enjoy the trip!

How to Feed Your Brain

When and where does the Sun get up?
Get up and look (not at it!). Don't make it an academic question.
You don't need an academic answer, either. You measure it with your eyeball. Perhaps draw it.
That doesn't mean "ignore the details" but "memorise the feel".
You do 'measure' it, but not with numbers.

There! This early!

Star navigation? It doesn't have to be a "trigonometrical nightmare".
For us it means: Find the Southern Cross.
Guess the time of night.
Find the Pot. Get to know the stars.
Project their movement forward into the daytime.

Let your brain do the calculations. Show it what to work on.

Wait for a Change in Perspective...

..to reveal the hidden shapes and layout.

As you drive up, keep **re**-examining the changing skyline. Continue the process as you walk. Look for ridges & valleys showing up through the vegetation as well as on the skyline.
E.g. Use parallax to reveal a deeper horizon behind a hidden valley...
Keep your eye on particular bumps on the skyline.
If there is a deeper background, i.e. more distant, then a lump on it will show up by moving with you —
more quickly than a 'foreground' horizon. You may have to look through the deep haze to even locate particular bumps.

Climb a tree, or a hill.
Use binoculars. Look especially for the trail ahead, in relation to the terrain.

It is the **Changing Parallax** which **Shows Up the Landform** and its Relative Distances.

Landmark to Landmark Network Navigation

This is a permanent, unchanging, top-down, whole-region method, **prepared at home**, for a large area, and for particular walks.
1. Choose only **prominent landmarks** from the map. e.g. peaks with transmitters.
Add in major ports, mouths & coastal extremities, lighthouses, around the coast.
Draw the network of triangles between them. ("Which one am I inside?")
Draw the circumscribed circle for each triangle. ("Am I inside or outside?")
Label each line in **kilometres**. ("What does 20km look like?")
Label each horizontal **angle** in degrees. ("Does it look larger (inside the circle) or smaller (outside)?")
Label each line with its **trend**-direction. ("Am I to left or to right of **track**? Which is this river **parallel** to?)
On the right of a line you will be looking to the left of its trend to see one of the landmarks.
2. Keep a **notebook** for local peaks — smaller but locally prominent lookouts.
The idea here is to navigate only by landmarks, not by (favourite) low spots.
You don't have to worry about whether one landmark is visible from another.
The idea is that **even when you can't see a landmark** (as well as when you can),
you can know where it is and how it lies in the land layout, how other trends line up with it and with other visible lines, and *how you fit in*.

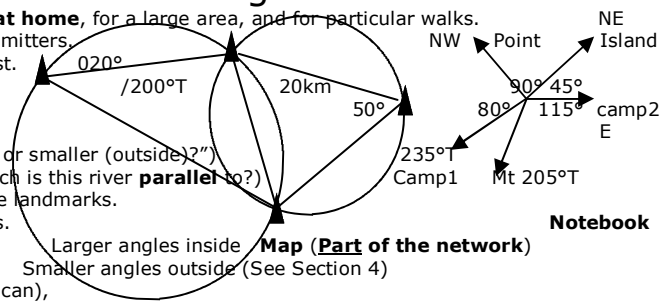

Notebook
Larger angles inside **Map** (**Part of the network**)
Smaller angles outside (See Section 4)

Name Each Clue — with solid ground

"That's Pacific Sunlight shining" — a mid-morning Sun, for Australia
"I have an Anti-American Shadow at the moment" — pointing away from North America
"It's a West Coast Cloud Flow" — you can't lose your orientation
"I love a Tropical Moon" — when the Moon is North of you (SH)
"That wind would blow straight in our back door" — 'Back-Door Blow'
"Downstream from here is uphill from Port Stevens" — 'Steven's Stream'
"This cliff sees the Southern Lights" — 'Cold Cliff'
"Topple-Tree Trunk would take us to True North"

The next step is to add something in line with that — see the next page.

Disorientation — A Killer Syndrome

Pilots have been known to fly upside down, or rigidly leaning to one side according to their mistaken belief about "up"!
If even our natural senses are unreliable we certainly can't afford to believe in any 'sixth sense' of direction, or on any unproven magnetic inbuilt direction-sense in humans! A trudge can dull & distort our senses. Your mind tries valiantly to maintain mental orientation—that is the problem.
Several times each year pilots die, scuba divers get lost in sea caves, cavers fail to find the way up and out, boatmen falsely navigate the fog, and bushwalkers get lost in the bush, fatally. They were not just careless; they were probably trying hard, but in the wrong direction, stubbornly.
Belief is an irrational master. Its force is stronger than reason. The conflict, combined with physical tiredness and sensory deprivation, is stressful.

We prefer to get lost rather than to rethink wrong assumptions!
We become willfully blind to strong contrary clues in view (including signposts, instruments, buoy colours and numbers!).
We magnify weak clues, to justify our delusion. This heightens the stress, because our sub-subconscious knows that we are wrong.
We actually lie to ourselves about what the clues do say! (e.g. we change the number, colour, reading or wording in our mind!)

Such irrational behaviour is explained as defending our psyche from hurt pride and from being devastated by confusion, fear and panic.
Don't laugh it off as "Oh, that's only for the worst case" because it is the less obvious self-delusion which will likely trap you.
Being human includes you — you definitely are vulnerable to being even less sensible than trained pilots!

Casual Outings Catch You Out

..unprepared e.g. without a map or compass or weather forecast.
Casual outings also have a habit (because exploring is interesting) of "I'll go just a little bit further"; and turn into longer than expected full-scale safaris. Exploring is built into the genes of a small fraction of any animal population—it's a group-survival insurance. But don't let it catch you out.

Serious trips go more reliably because you know what is needed, but you rarely prepare well for a casual trip.
e.g. you have no pedometer or notebook and so don't keep a good enough log of distance or direction.
So **Lesson Zero: Make some precaution** (to avoid getting lost) **every time** you wander off.
Another unexpected casual mistake is to decide to leave the track without thinking what precautions may be required.
Try this. Look up and repeat after me: "God, have I forgotten anything? Is there anything I should do first?"
Let someone know the single biggest clue: "I'm going exploring that way",
but then don't forget the unobvious catch—that person might also go exploring & not tell anyone where you went.
(Being disabled by an injury is one thing, but dying of exposure or thirst or wild animals, simply because no-one knows where you are, is another whole added dimension—that is probably the strongest reason for bothering to let someone know.)
So before you go off, pass on any vital information about others, as well as about yourself.

Trust Yourself Less

..when the compass and you disagree, for example.
When the map shows the junction there, but you reckon it should be here, for some reason such as 'but we've come too far!'
Or similar cases, of confusion.

Your reasoning sense has holes in it, big enough to outsmart yourself.

In the end you will usually conclude "Well, I was wrong, due to this very good reason(!), and it was how the map and compass told me it was."
But you don't want to admit it, at the time, until you are proven wrong, yet again.
Your reasoning is not, despite your worship of it, the highest god in the Universe. You need to be able to criticise it freely.

Unreliable as map and compass can be, at times, it is better to trust them more and yourself less.

More is Better — Thought, Time Care

A simple principle to apply is to do a better job at what is marginal.

For instance, "flat" ground nevertheless does have a slope which you can determine with care.
The wind may be gentle but a little extra time could make the difference between success or failure at determining its direction.
You can always double-check a casual decision, e.g. about which way the dry creek-bed flows.
Instead of wandering in an unrecorded meandering fashion, you can mentally log your changes in direction by lining up pairs of marks.
Just before you leave the track, a quick checklist can bring a few necessaries to mind, but only if you take the trouble.

A little extra thought; an adequate time-allocation to each task; more faith; and that bit of extra care—will solve a lot of problems.

Moveable Position vs Fixed Orientation

• It helps to park the car in a cardinal direction, or pitch the tent North/South, since we all rely on such clues to revisualise what we left behind.

• But when you do try to orient yourself to something invisible such as how the car was parked when you walked off, there is a simple trap to avoid

Don't imagine the car/tent as 'behind' you and twist its imagined *orientation* to suit the Sun or compass;
Instead change its *position — move* it to somewhere other than 'behind'.
Mentally manipulate its placement, not its orientation only. Or else the 'correct' orientation will confuse you — nothing will make sense.

Don't let the Sun 'change its direction' — that is a tell-tale that you are keeping the reference position fixed *behind* you rather than *moving* it as you swing to left and right.
Because we have no good tools to log changes of position or distance, we tend to neglect that half of the story in favour of maintaining orientation.

• As another hint on how to use a fixed orientation, e.g. North-South, so as not to let it 'change' in your mind, imagine it as a long, huge, high, iron girder bridge overarching you and your car/campsite, in the appropriate direction. Whichever way you imagine it, e.g. through-bolted to the continent, or from one peak to another, make it so solidly fixed to the ground and impressively immovable that you feel so insignificant yourself that you feel *you* must *turn yourself around* in order to pass under it.

Routine Mind-Shut-Down Procedures

You know how you become dull and passive — when being over-taught, lectured at, just entertained, occupied in repetitive tasks, sitting in front of the television, failing to solve too-difficult maths problems, or being proved to be a perverse idiot against your own opinion…
Automatic self-defence from constant attack... involves shut-down — into a semi-hypnotic trance.

Similarly,
Walking becomes a "trudge" after about an hour without a break, depending on the load, the heat, and how many trudges have preceded on that day, and previous days.
Pain-relieving opiate levels rise. Conversation takes a dive. The path sliding under your feet gets stared at a lot.
Your mind is passive, plus your body is on auto. It is physiologically involved in the hypnotism.

Mental Torpor is a navigator's constant enemy.
Keep the Different-to-Same-Ratio up, for both yourself and your companions, so that you can stimulate each other into full consciousness!

Feed Your Subconscious Subconsciously

E.g. Open your Atlas to a map of the world. Orient it and leave it open, oriented, and in view.
E.g. Place a magnetic compass somewhere in view at home.
E.g. Develop a habit to Orient your watch N-S each night when you take it off, wherever you happen to sleep.
E.g. Sleep with your head to the north each night if possible.
E.g. Clearly visualise what direction your feet are facing before going to sleep each night.
E.g. Imagine your body lying across a map of the continent.

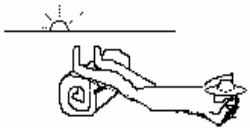

New Habits Open up New Horizons.

"Retracing is Re-preciating"

..This memory-aid is something you have to tell yourself — to avoid blinding yourself and missing the goodness of the walk back.

If you are only titillating your senses for the passing instant, by constantly seeking "new" experiences, you will develop the mind-deadly "Bin there — Dun that" attitude
Familiarity breeds contempt… "I know the way" doesn't guarantee that you know much else about the way.
The return journey degenerates into a dull, silent, often over-fast, boring, mindless trudge, even if you drove all day to come and see this top-quality location! In effect, you have walked the track twice, but looked at it only once.

Two solutions:
Take an interest in your surroundings, to overcome passive mindlessness.
Questions on your mind overcome boredom. If you have asked questions about a new place, you should be able to answer them on the way back.

Remind yourself of the opportunity to "re-preciate".

Aim to Integrate…

- ..Integrate daytime clues with night-time orientation—use the Moon, the place of sunset, the sunset glow later (which is further South—SH), the twilight Blue Band, Venus, the Ecliptic, the skyline — anything to carry you over from daytime to night-time, during that awkward twilight changeover phase.
- ..Integrate invisible landscape with visible landscape—work out what lies behind each hill, to picture the outside world against that skyline.
- ..Read the "map check-list" and specify what each item means from your point of view—if you didn't have a map, for example.
- ..Integrate the changing, with the unchangeable. E.g. the Sun's daily rotation needs to be followed, by naming its current direction. Same with the track; and your nose; and the wind.
- ..Integrate one walk with another — don't just carry two separate mental maps, add a third overall view.
- ..Integrate each system with each other one. E.g. don't let the map be one "world", while the walk is another, and the night sky a third. Join them into one whole.
- ..Integrate each direction clue with each other direction clue — E.g. "The tower is just before magnetic noon from here" (3 pictures in 1) E.g. the boat-radar may fail in the heaviest weather, unable to penetrate the rain squalls—you need a boat-compass too.

It is the **combination** of clues which stops you getting disoriented — if you mistake one, the others will alert you.

Notice and Match Up

'Notice the clues' is the easy half of navigation.
Sometimes it is as simple as seeing that the track is curved.
However, a slight curve due to a river bend can turn you in a hairpin without you realising, even if you *do* notice the curve.
Until you also notice the *extent* of the curve, you are at the mercy of disorientation. Something more than looking is needed.
You need something to orient to, like the wind or clouds or landscape.
Again, it does you little good to have *seen* the clouds steaming by if you forgot to see where they came from.
'Notice *and Interpret*' is the catch-cry, which never stops.
But when the wind stops, will you know what direction you have been going? And 'heading for the mountain' is fine, unless you lose sight of it.
Something more than single-clue orientation and interpretation is needed.

Do Notice; Do Interpret; Do Orient; Then ***Match up***; Link it to something else. And to something else. E.g. the clouds are from the mid-morning Sun; The mountain is on my shadow's port bow; The river has bent from 'towards the coast' around to 'away from the coast'.
For special examples, see '3 in line', opposite

Un-Dis-Oriented

You don't need to be perfectly and continuously oriented!
In fact it is quite hard to maintain an exact direction-sense, as soon as you turn your head, move a few steps, divert a few degrees, etc.**
You only need to be realistically confident of position & direction.
Hence **it is better to be roughly aware of <u>many</u> clues, than to aim for perfection, say, in location or direction**.
E.g. 'Moon North' is approximate, +/-60°, but awfully comforting! — It adds to a complete picture.
E.g. You may not even look at your map again, after studying it once, if by now you already know what it is telling you.
Don't aim to be perfectly oriented — just undisoriented.

E.g. Hold the conversation on how to judge range; what's behind the hill; where to find moss; why you can't yet interpret lichens; and so on…
and you can hardly miss seeing the walk.
E.g. Run through the Start—Stop—Rest Mnemonic, p.110, as a memory challenge to all… "What's next in it?"…
which has the benefit of practising all its clues too.
**As an exercise, guess, from a distance, diagonally, where the side of your house or car points to on the horizon, then go and line it up to find out
Or guess which point is opposite that hillock or tree, then establish how accurate your guess was. My results: +/– 10°; +/– 30° when not trying.

Trapped! By Not Looking Back

A fish funnelled into a fish trap may never again find the narrow entrance hidden somewhere in the large perimeter.
Converging tracks and sudden clearings trap bushwalkers too, when they walk through an irreversible change without noticing.
"Where did our small track join onto the main route?" "Where did we hit the fence/road/pipeline/track, and turn along it?"
E.g. <u>stepping into the creekbed, to explore upstream, is a one-way move</u>—you may not recognise it coming back, and could overshoot.

- **Beware of <u>all</u> sudden clearings**.
  "<u>Where did we emerge</u> from the bushes: out into the clearing, onto the peak, onto the riverbank, out onto the beach"?
  Look for footprints — anyone's, but especially your own.
- Remember the phenomenon — and mark those junctions in your mind! Or even mark the junction <u>on the ground</u>. E.g. place a rock there.
- "Which tributary did we follow down?"
  "Which ridge did we come up on?" More on p.63b
- Log the distance and directions involved at such changes/junctions/clearings, before you walk on and 'lose track'.

Disoriented? Stop. Say So. Reorient!

Be prepared to voluntarily switch from 'certainty' to uncertainty.
Learn to live with fallibility.
As soon as you suspect a disorientation episode, admit it. Say so, out loud. Don't miss the chance.
Tell your brain off!
"You <u>know</u> you're just guessing! You made some mistake and won't admit it! Let's work this thing out together"

If you <u>don't know</u> that you are disoriented, that can be worse than being lost but oriented.
You don't even have to be lost — Even if you know where you are now, if you are disoriented, you may walk on <u>into</u> trouble, rather than be looking to walk <u>out</u> of trouble.

Recall how often you have experienced a hard-to-shake feeling of disorientation.

Lost? Rule 3: Do a Good Job — of walking out or staying put

It may not be safer to stay. Well, if you decide not to stay, don't be aimless in what you <u>do</u> decide to do…

- E.g. Retrace your steps.
- Keep a straight course — e.g. downstream; along the road; toward the mountain; to night lights; to the traffic noise; a chosen direction
- Be single-minded — purposeful, but not stubborn
- Leave a Trail — you may need to retrace. Yes, it consumes valuable time, but "don't get more lost"
- Leave clues; notes — for others
- Walk quickly if you must, to cover sufficient ground, but 'rescuers' may not be able to keep up!
- Yes, your exit strategy may need to be aborted, but at least nominate an exit strategy for the exit strategy beforetime.

In-Line Integration

Put your preparation into something useful by using the '**3-or-more-in-line**' interpretation method…

"The wind is blowing *from* Port Bluster, *over* Hot Hill, *past* the carpark, *to* us, and on *towards* the Sun.
"The Pacific Moon is shining parallel to the highway"
"The Mountain has Montague behind it, that cleared paddock in front of it, and that hillock in the opposite direction"
"The compass needle points past the Northern airport to the mainland"
"My anti-American shadow is pointing past Penstock to Point West at the moment"
"The stream is flowing from Mt. Big, down over Lake Low towards Mersey Mouth"
"That Tropical Sun is over New Guinea, shining Past Alice Springs, over the intersection to me, and into the teeth of the wind"
"The Nor'Easter clouds are flowing from Noumea, over the Main range, to here"
"That persistent pigeon is cooing from behind camp"

The effect is to turn your reference orientation directions into fixed tracks (see p71bi), making interpretation easy.

Navigation is Mental Work — Stay Alert

The tools to work with may be "map and compass" "line and direction" "time and distance"

But any workman needs, and needs to practise…

Attention Observation Analysis Interpretation Preparation Concentration Memory Visualisation

Specialist knowledge Deft application Reliability Right attitudes A store of safe and useful habits Coordination

In other words: **you get lost by lapses in applying your skills, not through lack of skills**.

Mental Equipment is Very Light to Carry — You already have it; Don't forget to use it!

Instruments Instruct the Intelligent

Don't let the instruments — compass, clock, protractor, map — do the navigating for you.
Rather, let them teach you, how to get by without them.

E.g. practise finding North by everything except the compass; then check your guesses with the compass. Try bush North repeatedly.
E.g. Guess at "adjusted Moon North", then consult the compass.
E.g. with a watch you can practise telling the time by the Moon.
E.g. before you use a protractor, guess the angle.
E.g. gauge the distance by eye, then check it on the map.
E.g. attach a sundial compass to your magnetic compass, for the sole purpose of practising & checking your use of the sundial.

In general, predict and guess before observing, and learn from the results.
You should **always check** an observation for good sense, and check any instrument, or any calculation — but prediction forces you to think **first** (rather than the go-to-sleep-first alternative — which is to always check afterwards).
Control your instruments — don't put yourself onto auto-pilot — or you will hypnotise yourself into blindness.

Look Back Regularly

Look At how you got here — so that you can recognise the way back if you need it — and can make sense of your walk.
At the change from one terrain type to another i.e. at the link between natural sections of a walk — after you walk through the change.
At the directionality of the sky haze down low — the change in thickness and colour.
At the colours of the leaves, lichens & flowers — they change dramatically upsun/downsun, and therefore forwards/backwards.

Look Back… At tracks joining in at an acute angle — which otherwise you won't see until you come back — which will confuse you.
At the bending of the track.
At what you may have dropped or left behind! (Check each other. Anyone carrying loose items should walk in front.)
At the distance come, compared to the distance yet to go.
At the weather clouds stalking up on you from behind.

Can you see your starting point, and get a bearing on it? What is opposite that — i.e. the landmark which now shows you your exact trend to date?

Look Around

Since turning your head while concentrating forwards, taking your eye off the ball or turning a smooth corner are all enough to thwart your direction sense, the anti-dote is to deliberately turn your head! Keep looking around at the wider context, including behind, so you don't lose sight of it.

Imagine an Aerial Photograph of the Continent

Visualise exaggerated shadows lying parallel across your imagined photograph — connect those with your shadow and tree shadows.
Visualise the Sun shining like a flashlight upon your imagined continent map — connect that Sun with the Sun you see (Don't look at it!).
Visualise the cloud shadows moving across it — there's one! See?
Draw the isobars across it, lined up with the low clouds wind direction — feel it in your face.
Superimpose a huge compass needle across the continent map — look at your needle. You should be aligned.

Usually you avoid using any small-scale map, since it is the real and full-scale world you want to see or visualise.
But sometimes, imagining a map helps, e.g.…

1. Your navigational map (i.e. remember what you studied!)
2. The mental continent glued to your compass needle, to help you '(not) swing the map'
3. The continental map you walk over at bends, to help you take in the new orientation
4. The bird's eye/satellite view e.g. when reconstructing where you might have gone wrong and when re-enacting it in miniature
5. Sleeping on a map, with your atlas open, and your compass needle showing
6. The Road Map marked out in Regional Triangles from peak to peak.

Making a M.U.D. M.A.P.-m.a.p.-m.a.p.

For a more accurate job, see Mapmaking p101c,103c,106ab,109c.

Mudmap your trip onto paper, because paper remembers better than you do. Use any scrap of paper. Take a pencil and draw you meanderings.
Orient the paper to your expected trend (p51a Rhumb Line & p111c S.T.A.R.T.), so as to give you most room to map your journey.

Mark a dot where you start from — probably on the edge closest to you.
Up-arrow for a reference **D**irection; draw in a long one, Up-Sun is good. e.g. Up-wind; Up-shadow; Up-peak to a prominent landmark; Up-Weather **D**irection. (the Cloud flow); Up-North, by compass needle, or Up-Moon. *Interpret* the reference-orientation-**D**irection(s) before you leave.
U-p**D**-ate the arrow direction(s) or interpretation as needed, e.g. as the Sun shifts slowly, in relation to the wind, or as you walk past the peak.
Maintain a trend, a straight-line 'leg' p73b. Use the reference direction if possible. E.g. use the anti-peak.
Or pick a distant mark and walk up to it somehow; Or line up two points and keep them in line (p85e,87e,89e); Or just guess.
Align the **A**rrow when you stop (See Map Alignment 97a-105a, 103b). Turn **A**round to see where you have come from, to map the directions well.
If all other direction references fail you, **Align the next leg carefully to the last leg**. Be careful, because this will accumulate errors.
A-lignment is the magic ingredient. But it is the extra reference A-rrow which keeps you straight and avoids those accumulated errors.
Interpret the new alignment before you disorient yourself: "Where then is the starting-point now?" — or else you have a map but no sense.
Don't forget to use the map on the way back keeping track of the bends and where the start point is, and the finish point.
Pencil in the **P**rogress **P**arallel to the last leg; Pencil in the last leg *distance*. Turn around again and pencil in the *next* leg *direction.*
Guess at the distance, or pace it as you go, or use a pedometer, or a **stopwatch**. Add helpful notes as needed — p95c, 122 might help.
Repeat the m.a.p. for each new leg.

Avoid Disorientation: OverNavigate

Your aim is surely not: 'how to read a compass' etc. But: 'how to avoid a disorientation episode'
Being 'lost' is no panic, although you try to avoid that. Being 'disoriented' is panic (even when there is no real need).
To avoid the psychological syndrome and the real dangers of being disoriented, and lost, the most sensible attitude is…

Avoid getting disoriented in the first place — Maintain orientation
Navigate carefully e.g. learn 'how to read your compass' etc.! — Overnavigate — don't rely on one theory, one skill, one faculty
Don't give in to panic, or even to anxiety.

Over-navigating when you can, makes it easier when you can't, e.g. sea-sick or injured, bushed-in, exhausted or confused.

It is a rule in life (Murphy's Law) that things go wrong very quickly, not just one thing at a time, but all at once, compounding the difficulty.
As soon as you realise you are disoriented, you'll probably also realise that you also don't know the time, the position of the sun, the direction of the wind, the direction of the clouds, or whether you can interpret the bushes for noon. Then it begins to rain, go dark, and out come the leeches, and you sprain your ankle too. You could learn this law the hard way, or take my advice and do your 'emergency' navigating before the emergency!

How to Get Lost — Follow the Track— A Track is a Trap-Door

A track will lure you into the middle of the wilds. You end up depending upon it completely.
I suppose tracks know where they are going, but you may not, and they can't think. They are mute guides of the blind.

Antidote: Over-navigate. Do more than you need to. **Don't just follow the track mindlessly. Don't let it do the navigating for you.**

Keep your finger on the pulse e.g. How far have I come? What speed am I averaging? What landmarks can I see?
Have I checked the compass lately? The weather?

STAY ALERT
STAY SKEPTICAL

Reorientation; And a Fresh Start

Step 0. Admit it: "I am dis-oriented!". Say So. Stop. Decide to 'reorient'. Discuss it out loud.
Step 1. RETHINK where you might have gone wrong.
2. IMAGINE the true situation.
3. IMAGINE A BIRD'S EYE VIEW of how it 'must' be. Lay it all out, on the ground, in your re-enactment.
4. USE THE MAP; and all the clues — the time, the clouds, the distance…
5. ACT IT OUT on foot, in miniature. "I left this way, turned…"
6. SAY OUT LOUD a) Well that's where I thought it was b) But it is really over there (e.g. that Sun there).
7. Imagine the stars circling where they ought to; the coasts trending the way the compass tells you; etc.
8. Aim for a NEW START "I don't know what happened, but now I divorce the past".

Only when 'the ground shifts under your feet', do you know that your brain has got hold of the new situation.

How to Leave a Message

- Put it in a prominent place (e.g. a hilltop; a campsite; a rock face) with a prominent marker — bright and colourful for instance.
- Make it helicopter-visible e.g. in a clearing; large
- Use international distress symbols e.g. 'V'; Fire; 3 lots. At sea, a white flag or white stroboscopic (flashing) light.
- Any bit of metal may help you scratch out a message e.g. a zipper runner; a coin; a key; e.g. onto a tree trunk.
  If not, try a rock, or stick, to scratch with. Next time, have paper and pen with you.
- Use spittle, or urine, to make mud, then cover the mud message, but prominently — e.g. when searchers remove the coat from the tepee they see the message. E.g. make a flat mud pie; press in twigs to form letters; keep the rain off it.
- A freshly dead leaf is like paper, to write on.
- A piece of charcoal makes a handy black chalk.
- Brittle Yakka gum makes a bright orange powder splotch, to attract attention, and can also be used as a somewhat waterproof crayon.

Disorientation in new territory

Familiar territory does *not* disorient you, so you must *have* the mental faculties to 'know where you are' when you can 'see where you are'.
The overload of *novel* surroundings must overwhelm your normal abilities to mentally keep track of small diversions and accumulated progress.
These suggestions aim to make an exploration more 'familiar' and less 'new'. They come from all over this book. I suppose the message is:
'Don't just read about the methods, take extra time and care to put them into practice; extra 'bother'. Slow down; try harder. Remember more'.
Look back *very* regularly, to increase your familiarity of already explored territory. If you had eyes in the back of your head, it should *all* look 'familiar'. Walk back over that territory before continuing, p 68b. It will feel like old familiar ground if you bother. You could find your way through, and back through, an obstacle course at night without a torch, *if* you went forward three legs and back two as a matter of habit or necessity. That's 5 whole journeys. Talk to the obstacles along the way, and use a mnemonic to recognise them, and what you did about them, and how you turned.
Leave visible markers behind you, so looking back is more helpful to orient you. Leave a trail. It boosts your confidence, even if you don't use it.
Look through the foreground into the distance to **maintain contact with reference directions**. Fix what you can see, especially ahead, in your mind, in relation to the Sun, shadow, wind, etc, until you can point to them from behind visual obstacles. Remember more of what you see. Interpret each clue for its relationship with your memories of how things looked. Integrate the bits into a whole. Use some distance measure.
Maintain a straight line if convenient, so the interpretation of where you are is transparent. Use long straight 'legs' so you can interpret your accumulated path more easily. Notice and remember the angles *between* legs especially. Create a mudmap, so you remain oriented to your start position, and to your current position, and to landmarks sighted along the way. Practise these and other techniques, when it doesn't matter, for later.

To Work Harder, Rest Harder

"Mental alertness" may involve

- A siesta nap!
- An early night (not a late-night campfire party or sing-song).
- Regular rests during the walk.

Tiredness, Exhaustion, Constancy — these stop you thinking efficiently, and affect your memory badly too. So do nicotine, alcohol and drugs.

- When you are in difficulties, sleep on the problem, or take your mind off it with a diversion, occasionally.
- For the sake of having stimulating conversation, the whole party ought to guard their mental alertness too.
- Navigation games do more than orient you, they wake you up.

Be alert to your biorhythms—daily, monthly, seasonal—they affect your alertness or make you depressed—all your mental faculties can be affected.

Minds Work Best with Intelligent Questions on Them

- Am I judging the scale on the map properly? How high and steep is the next climb?
- How many creeks have we crossed? How many more to come? What is our next position-check landmark? How soon?
- What direction will the next creek be flowing in? Could we see Mount X from here?
- What is the nearest escape route? To where? What side-trips would people make?
- What does this place remind me of? Why is it sandy soil?
- Will it frost tonight or not? How fast are those cloud shadows going?
- Where would the animals drink, and shelter?
- What is this place's own special uniqueness, and why?

Anticipate, with questions, the things you want answers to.

Look Through the Vegetation, to the Contours

- Bushes cloud your vision. If, for example, you want to 'stick to the ridge', or want to memorise a junction, then look for the more **permanent** clues…
  ..like slope, creeklines, direction trends, terrain, soil & rock types, alignments, distances, context.
  Of course, look at the vegetation too, but not only at the vegetation.
- Trees cloud the skyline. But there are gaps and lines of sight that open up as you pass by. Look through the forest for the ridgelines, slopes, and horizon. As you move sideways watch the parallax—to reveal otherwise hidden contours.
- Look through the trees, to the **most distant**, largest, tallest features to navigate by — you will undoubtedly catch glimpses of these much more often than you think, if you are looking. They may be other trees, so get used to recognising individual trees.
- Examine the height of the sky beneath the canopy.
  E.g. the lowest sky is 'downhill', but **the lowest sky uphill** is toward a pass or gap, & **the highest sky downhill** shows up the ridge.
- Forest-bound **river beds** are often clear of trees and give you a rare distant view.
  Stop in the middle of that rope bridge! And hang around, to look around.

Connections to the Outside World — You Don't Need to Feel Isolated

Ruins connect you to the past. Earthworks and mineworkings also. Even the landscape engraves its history for you to see.
A **GPS** connects you to satellites; a **radio** to a radio transmitter direction; a TV antenna points to the transmitter; a **phone** connects to the grid.
The Sun, Moon, stars, planets + time + vertical can give you an **astro-fix** to connect you to latitude and longitude — globally.
The **magnetic field** connects you to magnetic poles, as do the aurorae. The deep blue polarisation band finds the Sun for you. So does a rainbow.
Navigational antennae link to flight paths. Flight paths indicate airports; pipelines indicate water supplies; electricity lines go to substations.
Paths point to points of interest. Recent tracks betray traffic. Track-markers put you on track; tracks connect to roads; roads to highways; to cities
Nightlights betray townships and cities; traffic noise links you to highways; surf noise, to the beach. Tides tell you about the Sun and the Moon.
River gravel shows up the upstream geology. East West gully shape shows up North; **rivers** connect you to the coast; **ridges** to peaks.
Air pressure here depends upon altitude; and upon the polar front and the pressure systems. Air temperature relates to the size of the landmass.
Absolute humidity tells about the proximity of oceans. **Cirrus clouds** show up the (westerly) upper winds. **Low clouds** and winds remind you of the isobars on the pressure map. Ground winds telegraph the topography. The radio links you to **weather information**. Waves, to distant storms.
Fence lines connect you to farms; boundary lines to survey grids; trig points to **maps**. Smells to camp fires, marshes, farms, abattoirs, toilets, seaweed, water. Stock trails lead to water; seed eating birds lead to waterholes; the plant and insect types are tied to their range and the region.
Vegetation is tied to altitude, latitude and aspect. **Bush north** indicates noon. Bush slant indicates the **prevailing wind**. Recent wind leaves tracks of yesterday. Deep soil temperature is related to altitude. Why feel "lost"?

OverNavigate: Equals Over-Familiarise

Overfamiliarise yourself with your route as you walk it, the first time you walk it.

Familiarity is the navigator's best friend, next time.
I.e. Years later you can more easily retrace your steps. Even when the track has overgrown.
It is an investment opportunity now, not to be wasted.
The extra time and effort will pay dividends, but like all good investments you must have an interest in the future.

E.g. Notice how and where you went wrong — mentally photograph it.
E.g. Lay to heart the most permanent features — the landscape and trends — not the track, its signposts, or bushes, which may change with time, or with a bulldozer.
E.g. Mentally photograph the lie of each junction, as if a fire might burn up the signpost.
E.g. Describe out loud your short-term memory about how you got here, lest that familiarity flits with the passing of the day.
E.g. Have your own 'track-notes' book.

How to Get Lost — Follow the Leader

A leader is like a track. You can follow him mindlessly. He ends up being a seeing guide for the willfully blind.
Ask yourselves, often, this hypothetical question: What would happen if the navigator dropped dead?
Could anyone else navigate competently?
Do others know where we are? On the map?
Could I find my own way back? Where are the cars?
If the 'leader' himself gets disoriented or lost, noone will know where they are!

Antidote: Teach. Learn. Test each other. Involve each other.
Demystify the magic art. Share it. Talk about it.
Listen to the debate. Look at the landmarks. Contribute.
Don't walk fourth in line. (Rotate regularly)

"Here, you find the way — I already know it"

"As Solid As the Rocks in My Head"

"Who am I to argue with the Sun? to say to it 'You should be over there — not where you are!' "

Telling your head off is the sort of attitude you need… when you find yourself walking uphill to the campsite which you know is down by the river!
… when the clues are shouting at you "Something is wrong here!!"

Navigators who are disoriented prefer to believe in such remote but real possibilities as a compass error or a mistake on the map, signpost or buoy, or to believe in the majority opinion being wrong, rather than to believe in the closer and more real possibility: "Perhaps it is ME who has made some mistake".

True, compasses & maps are notoriously prone to error, but your internal feelings, vague recollections, position/time/direction senses are more notoriously unreliable.

Lost? Rule 4: DON'T Do It Again!

Make resolutions. E.g. 'I will always take a compass.' (I tie a key ring compass to my hat string and I always take my hat.)
(I also tie a whistle to my keyring)
E.g. 'I will buy a map/mobile-phone/EPIRB'
E.g. 'I will always leave my itinerary with friends'
E.g. 'I will learn more about navigation; take pencil and paper; practise'
E.g. 'I will never go on a walk unless I have asked: "Which way am I leaving; Going in what direction?" '

Even better… Don't wait to be forced into a sensible habit. Declare your present attitudes to be 'not yet fit for human navigation' and put it right, now.

A Three-Track Mind

Here is a trick to avoid attention-deficit syndrome in your navigating:
Practise concentrating on two things at once.
Your feet find the path by themselves.
Your mouth talks about how good the meal will be.
Your mind sees the scenery without missing it.

Dual tracking your mind is like many other good things — just a little bit too hard to persevere with under normal circumstances, but possible to achieve, with extra dedication.

Don't automatically get lost into only one train of thought.
Slate this down for practice, for the rest of your life.

Interpret!

E.g. If you wear a face mask in these days of extreme UV radiation, or sunglasses, you become aware of when you are walking into the Sun.
The obvious interpretation, hanging on to this piece of otherwise subconscious information, is: "I must be travelling toward the Equator".
It's a chance to orient yourself to the new orientation. Can you see your shadow ahead of you?—you are getting closer to the pole.
E.g. You might be in a deep dark forest and see a rare bit of sun-dapple piercing the dense canopy.
That means you can go over there and see the sunbeams for a direction check.
E.g. The 'crunch crunch' underfoot changes to a 'clop clop' — "Oh, if this is rocky, maybe I've passed a break of slope."
E.g. Exotic plants betray human occupation — you could look for an old access track.
As a general rule: **Look for any asymmetry**, and interpret it — usually for direction. E.g. When swimming with goggles, twinkly bubbles mean "upSun" — dull ones, "down-Sun". The wave and ripple pattern indicate wind direction. Sunbeams show up the down-Sun direction.
E.g. Scuba-diving: The anchor rope shows up-/down- current; alternating current is the swish/swash from the longest wavelength waves—note their direction. Direct current equates with "away from my starting point". Bubbles show "up" (with some adjustment for current if needed).
Interpret the whole story of settlement: Valleys attract farms, dams, fences, boundary roads, houses, access roads. I.e. you see a lonely fence post on a mud flat, and you say "once several paddocks probably converged onto this old watering place—look for old tracks near the fence post". Look for old nail-holes in the post, to find out the direction of the fenceline.

Play the Field-Scientist

Vary your **focus-distance**.
A herpetologist will see snakes at his feet & frogs at 10 metres.
A geologist will see the skyline, and the creekbed outcrops.
An ornithologist will look upwards, a botanist downwards.

Scientists see more, because they know what to look for, and how.
The more they look, the more they see, observe, discover, and learn i.e. the more they learn what to look for.
You can begin the self-reinforcing cycle in yourself by: wanting an explanation, by trying an interpretation.
And, of course, don't forget to look into the topic, into books about it and field guides for instance.
Be interested in all levels, in all scales and directions.

Focussing keenly will improve your navigation.

Orientation is More Fundamental Than Position Sense

We all carry mental recollections of recognisable spots—like a mental photo album.

But try to mentally picture whole stretches of a country—roads, coast, range, river—with a sense of direction & relative location & linkage.

Better is to carry a library of whole mental maps—walks, paths & layouts—not just photos of isolated spots or isolated linear features.

But Best is to keep all those maps in right relation to each other in your mind.

As your ability improves, you focus less on spots and more on the links between spots;
less on position fixing and more on trend analysis;
less on separate techniques, more on integration.
Direction clues are only helpful if you can coordinate them with what else you need to locate.
Positions only make most sense if you can say how they lie, in distance & direction, in relation to other positions.

Think in Pairs, in lines, trends and tracks

Two landmarks lie on a line, of known direction, and fixed track.
We tend to see things individually, and not see the join between them.
So try instead to get to know
the *direction* of 'camp to that moored boat',
the *bearing* of 'Mt Big to Tower Hill',
the *angle between* 'sunset and the headland'
the *trend* of 'Razorback Ridge'
the *pointing* of 'tree one to tree two'
the *line* from 'start to finish'
the *track* from 'Lake Low to River Junction'
the *tilt* of a star-pair
the *join* of two subsections

Needle-North, -South, -East & -West. Identify them on the Landscape

One of the major problems of navigating in bush, is keeping log of estimated side-slip, distance-wise, even if you are oriented, direction-wise. But you can see your sideways movement, see p53a, if you plot known directions onto known points on the skyline, (E.g. Magnetic N,S,E,W require no numbers; E.g. your initial forward bearing is already there waiting to be used). Or better, use points in the middle foreground landscape — about as far away as you intend to walk. E.g. local hillocks, towers, houses—things which will be visible from time to time during the whole walk, the closer the better. Make a note of their directions. Use these spots later to find home or your starting point. When you see them later, in relation to compass, you can visualise how much distance out there you have moved and in what direction. Page 106c "Tom Thumb Parallax" shows you how.

Orientation to Your Start-Finish (Rhumb) Line—by sight or compass

Work out the start-to-finish visual bearing to the destination if it is prominent, or to a landmark or point on the skyline beyond the destination. Whenever you can see that point, you can see your left-to-right sideways positioning, in relation to the fixed track from start to finish.
A simple way to keep track of your sideways movement, at least in open country, is to choose an approximate rhumb line compass course (from start to finish) and watch it move across the landscape in front of you. E.g. "The follow-me arrow is now pointing more to the *left* of that gulch".

Orientation to Your Finish Point, by the Parallax Method

If you are making a one-way journey, use the same "parallax method" as above, to keep a log of the estimated sideslip, but use your **destination**, not your start point, for your reference set of bearings. As the walk progresses, you can see yourself closing in on the correct destination.

Exponential Anxiety — the 11:59 syndrome

Your fear of being 'lost' rises in inverse proportion to the need for fear, as you approach your goal in unfamiliar terrain.
Will it be there as expected? Why haven't I got there already? Surely it should have shown by now!
Have I overrun it?? Maybe I just didn't recognise it? Has it disappeared?
When will this track ever end? Will it end at all?? Will I have to walk all the way back! And so on.

Now, obviously, if you have done your homework well, it will be there, where you thought it was, but maybe not exactly when you thought, so optimistically, you 'ought to' arrive.
Distance guesstimation is our weak point.

Don't add needless anxiety to your background worries. Wait and see it sort itself out.

Anxiety, fear and panic work against good navigation.

Explore Junctions, While You Can

It only adds a few minutes' pleasant diversion to travel the branch tracks which you don't want. (Especially if you can dump a heavy pack)
You may never get to find out about them any other way.
This includes natural junctions—of creeks & ridges—and canal junctions.

When you are exploring, you may never retrace those exact steps, so don't say "I'll explore that on the way back".
The other reason not to say that, is that you will be even more tired later.
When you are uncertain of your position, such exploration ought to be considered "essential", now — not later when you realise your mistake.
Later you may speculate about what it might lead to, but now you can prove it.

In any case, as a matter of principle, explore far enough so that you don't have to repeat the foray
E.g. walk left, until you are sure you should not go left.

"Meet back here in 10 minutes" is one way to investigate in all directions at once.

Separated? An Agreed Plan Helps

Example: Where did I park the car? Is it stolen or am I lost? Are the others worried yet? What will they expect me to do?
Example: Our driver has driven off! Why? What will we do? What will he do?
Example: She has not returned on time! She was exploring that way.

Try these out...
"We will, for at least one hour and at most two hours, try to rejoin each other before changing to panic-plan"
"Where we were last together, is the agreed rendezvous now"
"After that we check our intended destination for messages"
"We know to communicate via a 3rd party—a 3rd person, a 3rd place, or a message bank… "
"..the nearest post office/petrol bowzer/pub/police station/emergency service/Fire Brigade/bank/Red Cross/Salvation Army"

Concentrate Outwards

Both introversion and extroversion belong elsewhere.
If you cogitate inwardly, you become blind to the surroundings.
When you chatter away about other things, that's equally distracting.

Don't spoil the special time of being outdoors.

While it is there in front of you, look to see what is to be seen — and talk about that — about the history of the place, or something about it that helps to keep you oriented, to its unique nature.

Learn not to be distracted from the main game.

SPELL Out Numbers — Into Words

The idea is to express, in words, what numbers measure. Humans hate hearing numbers! So translate mathematics into user-friendly words.
e.g. 270° is "West". That's much more meaningful.
10 past One — "Oh, it's midday!" (in your locality, for summer-time)
45° South — "sub-tropical, even temperate; halfway to the Pole"
10.5 km — "almost half-way"
020°T — just to the right of magnetic North.

PAINT your Words — Into Pictures

E.g. To make it even more meaningful, "West" = "Sunset!" (a word-picture)…or even better……
"019°" = "Where the Sun is just before it reaches magnetic North"
(This translates numbers into a familiar and visual time of day.)
"Midday" = "High noon!" (How high?) "Half-way" = "Within sight of the Gorge"
"Halfway to the Pole" = "The Sun rises and sets at 45°. The pole star is halfway up"
'Meaning' usually means a 'picture'
Words without meaning are like looking at the brushstrokes of an oil painting through a magnifying glass, missing the painting!

Here!
(Picture it;
Don't look at it!)

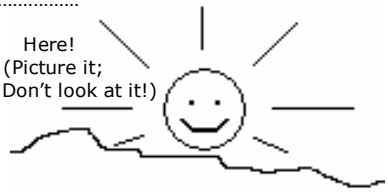

Look at the Ground

- Soil types; rock types (p. 79a)
- Animal runs; animal footprints
- Fallen track markers; toppled-over cairns
- Human footprints e.g. yours, on the way back
  Examine each other's boot treads, then the boot prints. Talk them through — expressing the recognisable peculiarities.
  Practise also on car-tyre treads and prints. (p. 66b)
- If you see a deep trench or a dry mud-hole in Tasmania, that may mean 'an old track' i.e. where constant traffic had refused to let the vegetation grow back.

Look for Foreshortening & Overlap & Parallax

Looking flat on to a shoreline e.g. the other side of the lake, means that any slight inlet or headland is probably a major and very sinuous feature. Focus sharply on it, looking for clues to depth. Don't just run on past it. Move sideways e.g. 20m, to show up the depth, by parallax.

Make Common Sense of the Various Direction Clues

Maybe the navigator has his eyes glued to the compass — but you, you can say "Oh we're just going upwind".
Perhaps you can see six landmarks — but the sensible way to see them is e.g. "We are heading halfway between Mount William and our shadows".
Cloud direction is a good clue, but changeable from hour to hour, so there is no loss in saying "Diagonal to the Sun" which is also changeable.

Don't Degrade the Point of a Clue or System

E.g. The point of a watch, is to tell you where the Sun is. (Don't look at the Sun!)
The purpose of a mnemonic is to remind you to think, not just to remember or to recite.
The reason for a "midnight-rise-noon-set" system is to work with the full-circle—even when you turn corners or turn around. It images the sky
The benefit of a skyline method, is for when you can't see the skyline. You will glimpse it 10 times more often, if you are trying to use it.

The aim of a practical method, is to practise it, not just to know about it!

The Aurorae

The Southern and the Northern Lights are especially frequent during high sunspot years (maximum every 11 years.... 2000, 2011, …).
They are most prominent after midnight. They are best seen in 'high' latitudes (> 40°). One to three days after a solar magnetic storm is best.
The most common aurora is the most easily overlooked — it is dismissed as a 'pale glow' in the North or South, low above the horizon.
But it is not dawn — that is East. It is not from a city — because it changes. Sometimes the sky will be pale green below and rosy above (between the clouds) and pierced by 'searchlights' which move left and right, slowly but visibly. Sometimes 'curtains' appear, shimmer and disappear slowly.
All the rays seem to emanate from somewhere near the North or South Magnetic Pole, and it spans 30° each side or more.
It is an interesting direction guide.
You can estimate current magnetic activity by analysing sunspot activity. **Never look at the Sun**. Instead, throw only an **image** of the Sun onto white paper, as follows... Point your binoculars at the Sun, **without looking through them**! Then, as if you were throwing a minimum shadow of one binocular barrel onto paper, and as if you will **let the paper look at the Sun through the binoculars**, from a distance of say half a metre from the eyepiece, point the big lens at the Sun. Adjust its alignment until the sunbeams go right through it and come out at the other end where the smaller eyepiece lens is. Steady the binoculars properly, and focus the bright, circular, white, image of the Sun so produced, carefully on the paper, by using the focusing mechanism and/or altering the paper distance, so as to reveal tiny pairs of black dots, on the paper. The larger the dots the bigger the storm. Caution! Looking at the Sun will damage your eyes and even looking at the bright image on white paper requires sunglasses.

Orientation to Your Start Point, by the Parallax Method

Step 1: **Start off Aligned**: Make a note of Needle-NSEW, bearings of any other recognisable points, initial-bearing landmark, rhumb-line landmark.
Step 2: **"The Direction Has Changed Its Mountain!"** Instead of saying: "The landmark has changed its direction", first say: "The direction has changed its landmark". I.e. retain in your mind the initial bearing, then, whenever you glimpse that landmark again, you can see how far sideways you have moved off that track—by locating which landscape point that direction now points to. You estimate the sideways disparity across the land, because your compass needle now points to a certain distance to one side of the original mark. Your job is to estimate that distance out there.
Step 3: **"I'm moving through the landscape, out there".** Imagine a thousand of you dotted all over the landscape, all moving parallel, and at the same speed, say, at 4 km per hour. All your initial bearings to all those recognisable spots out there, are moving with you, leaving those spots behind at a rate of 4 km per hour, and now those bearings point to new spots a calculable distance ahead. In reverse, this lets you see how far you have moved sideways to any direction. Bring those distances back to your feet, and see how far away from "start" you are, and how to get back. (At 15 km range, if you are walking at 4 km per hour, you will leave the landscape each side of you, behind at a rate of 15° per hour, but...) From the satellite's point of view, you and your directions are (all) simply moving through the whole landscape at 4 km per hour, no matter at what distance. An alternative viewpoint is to say that every bit of the landscape always moves "equal, parallel and opposite" to your motion, relatively speaking.
Step 4: **Estimate the gap between parallel tracks**—See "Advancing Position Lines" p.108a. I.e. from the initial bearing (a track) to the new one. —by taking a new bearing & using sin(change in direction) x new range. This converts the distances you can see, to a number of kilometres.
The answer, if you want an answer in numbers, is how far you have to travel at right-angles to the original bearing to restore the original bearing.

Nervous Nellie Navigation — Not Necessary

Anticipate a problem. Do something crisp about it. Clear your mind.

Examples: Continually looking at a compass needle — can be short-circuited — by dividing the route into short legs, p. 73b.
Continually gauging the distance covered — can be short-circuited — by making one prediction before you start, and waiting for it, p92b
Getting lost trying to keep track of many bends — is a symptom of not knowing the overall trend in the first place.
Continually guessing where that junction might have led to, is avoided, by exploring it when you meet it.

You never know when you may be totally distracted from your normal pattern of thinking — it is then that you will need a cleared mind.
Plus, if you habitually allow your mind to move on to the next problem, by dealing with the last problem, you avoid ongoing confusion.
Forethought can see the normal problems coming, and pre-empt them.

Look for the Essentials — the 'Directions'

You can't afford to miss them!
E.g. 'Turn left at the Y-junction' means someone has to be looking for it!
E.g. 'Head uphill after you cross the creek' means: notice the creek bed!
E.g. 'If you hit the main road you have gone too far' means that you were talking about lunch instead!

Directions are worth writing down as soon as you hear them: "First right, third left, 6 km, look for the track"
Try to give directions more sensibly: e.g. "Up Fish River Valley from the Park Boundary"
Try to use fixed directions e.g. "North West" "from the lake to the Peak" "between sunrise & Mount William"
Not variable or relative ones e.g. 'turn left'

A Whistle-Code for Contact Calls

Example: The track splits. It's probably just a detour. "I'll go left; You go right". You keep in contact by whistling, cheerfully & tunefully.
Use your favourite tunes agreed between you...

To signal:	"It's me"	Try:	"I did it My Way"
	"You Come to me"		"Oh Come All Ye Faithful"
	"I'm coming"		"I'm Coming; I'm Coming… ..Old Black Joe"

			The rhythm is...
Radio **Morse**:	CQ = Please Contact Me	= " Dah-dit-dah-dit Dah-dah-dit-dah "	"C is Seasick, Q Forms a Queue" "Calling Calling, All radiOs"
	R = Roger — Received	= " Dit-dah-dit "	"An R dits"
	Wait	= " Dit-dah-dit-dit-dit "	"aWait f'r a bit"
S&R:	Come Back i.e. Recall of Party	= Dah — Dah — Dah — Dah	"Ree-Call-All-Souls"
Other:	International Distress	= XXX = Dah-dit-dit-Dah — 3 times.	"all X's hurt, all ex's hurt, all ex's hurt"
	SOS — real emergency — di-di-di-dah-dah-dah-di-di-dit — repeated		"A bit of Save Our Souls isn't it?"
			See appendix for the full code.

City Glow

From out in the bush you can usually see the direction of about three cities or towns, by their night-time glow.
This works best in hazy atmosphere — virtually any night! — and especially on overcast nights and when no skylight will show.

You need to check a map before you go — for the nearby townships — then check the theory against the facts,
then you can find it useful when you need it.
E.g. you can hold a straight line at night — relative to any recognised glow.

Don't confuse it with aurorae, sunset afterglow, predawn Sunglow, Milky Way, asteroid belt (Zodiacal lights), headlights, local floodlights, bushfires.

Just remember to get up and look for what you normally don't notice.

Come to your Senses

..And it will be surprising what comes to your senses.
Smell the air — for seaweed, marshes, farms, abattoirs, smoke, smog, civilisation, scent trails, fresh wombat dung, humid air, perfume.
Listen occasionally. To help you hear, cup your hands to your ears. Selectively tune out the high-frequency noises to listen to the low frequencies; and vice versa. Listen through the white noise, and through the silence of night.

For occasional vehicles, trailers bumping, boats thumping, jets overflying, thunder rumbling
For constant traffic rumble of a highway
For ocean rollers thumping on shore. You can also often hear the difference between high and low tide, from the wave noise.
For running water, frogs
For wind noise
For generators, pumps, machinery
For farm stock
For the crunch under your feet, changing

Look **carefully**. "It is amazing what you can see, if you look". It is a most important habit for the navigator and is very rewarding.
E.g. you read what others have scratched over the sign that says "5 hours" and discover "8 hours"; or find old track markers almost overgrown.

TRANSFER Your Pictures into Reality

E.g. "Sunset" = "THERE!" even after sunset!

Find the red glow afterwards, further to the left e.g. at 8:30 p.m.

"High noon" = Look at how short your shadow is.
"Within sight" = "Well, can you see it yet?"
"The Pole Star is halfway up" = Do you see how the Cross doesn't set?

Living in the realm of theory is like an absent-minded professor — who can't find his way home, or remember which way he was travelling.

Put That Experience in Your Memory Bank

E.g. "The Pot will go down behind "Pot Hill"."
"I didn't realise how high the Sun can get at 45° latitude!"
"You can see the Gorge from the halfway mark, provided you climb the ridge"
"The Southern Cross keeps correct time at the end of March" — the beginning of Winter. See opposite for the next step…

Look, with Vengeance Aforethought

A bird observer doesn't just stumble into birds, he runs them down.
A native tracker doesn't simply 'notice' the clues, he searches for them.
Your task is to track down navigational clues, then bail them up.

Your enemy is: losing conscious grip on your task i.e. when the trudge gets the better of your brain.

For Practice: Express the unexpressed. It's no good to be only half aware of what is going on.
Interpret the meaning of each clue. It is no good staring at a compass needle, unless you realise where you are going!
Look outwards, to see what you can see.
Change 'a cursory look' into a specific search.
But, what exactly should you be looking for (once you have exhausted your repertoire of clues)??
Answer: Find out! I.e. **Teach yourself**. Don't wait until someone tells you.

Preparation is the Key to Difficult Navigation

Marine navigators in rock-strewn waters, or battling seasickness in heavy weather, or short-handed, know that the only way to cope with seat-of-the-pants reflex navigation is to have **already prepared clear instructions** about danger-bearings; back-bearings; lines of sight and distances to run, so that they don't need to leave the helm.

It is the total picture, beforehand, which orients you, later.

They need to plan the entire passage (so as to make landfall just before dawn, for instance) according to the tidal streams and ocean currents that they expect to meet (since it is the current much more than the winds which affect boat-speed-made-good, even for yachts). Their key desire is to sail with the current, and preferably with the wind and the weather systems, avoiding the storm-season. See "Aiming-off" p96a.

An hour seem such a long time to spare when you are busy at home; But when you get out on foot, relaxed, with hours to walk but no opportunity for desk-work, it seems that every hour not spent in preparing what could have been prepared, was a wasted hour, as far as being useful now.

Helpful Symbols for Standardising your Chart–Work

"The Deduced Reckoning Position"
from course and distance information
("Ded. reckoning, or D.R.)

"The Estimated Position"
taking into account currents and leeway
to distinguish it from simple D.R.

A known position, or a Fix

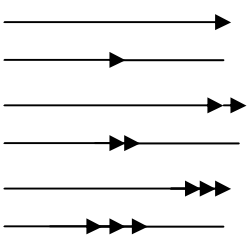

Lines of possible position should have the arrow on the end
e.g. pointing to a landmark
A central arrow means a course. i.e. travelling, not pointing.
One Arrow signifies the course steered through the water
A Transferred Position Line, i.e. advanced across the chart

Two arrows is the course made good over the ground

(A doubly transferred position line, for a three-way fix)
Three arrows is for the stream or the current.

Look Backwards; Think Forwards.

You sometimes, with only a mud-map to go by, have to follow a straight line by compass or landscape, to a low-lying, invisible feature, e.g. a tarn, hoping to come across it without bypassing it. Well, success will often make you feel that that is the end of the matter; "Mission Accomplished". Until next time!
But look back, if you can, to a recognisable starting point, then take a compass or landscape bearing, backwards to the start.
You then convert that to a forward bearing for next time. Place 2 pebbles on the ground in line, turn around and line up the landscape mark needed
The result is a "track", not just a direction — see p.71b — a 'can't–go-wrong' guide.
If your beginning point is not visible, turn around and take a forward bearing on the skyline feature you think you should have headed towards (and probably did). E.g. "One span to the left of Mt. Visible". The result is only a direction from the start—a guessed-at track—but nearly right.

On p.42d I said only to use bearings from high points to high points, for a general navigational framework.
But add to that, a mental or notebook entry of the bearing from a low destination point back to visible possible starting points, when you arrive successfully, or of the forward course you used — preferably by the landscape itself. They are "How-to-get-to (a low point)" tracks.

(Interpret)

Extend Visible Reality into the Invisible Picture

E.g. "It's boiling hot in Baghdad"
"The Sun is over Townsville"
"From the ridge the water flows down to the Gorge and eventually out to the Murphy Mouth"
"The Earth must be rotating Easterly, from West"

Don't be satisfied with half a meaning.
Until the interpretation leaves you oriented, and able to remember something, you haven't gone far enough.

See p. 71b(1) for the last word on interpreting 270T.

To Get There Quicker, Go Slower

We all prefer to run ahead of necessary preparations, i.e. we prefer instant gratification to wisdom.
We also prefer to get something unpleasant, finished, and quickly, once we have started, e.g. "can we make this bad "shortcut" work out?".

Don't rush! "More haste, less speed" = More care, less fluster.
Slight panic makes you hurry. Hurry is a recipe for carelessness in navigation, and for blunders.
E.g. over-running the junction
E.g. missing most of the clues
E.g. taking a foolish "shortcut"
Common causes for hurry, other than anxiety, include…
"Let's beat the rain/the dark"
"Let's get back on time/for lunch"
"Let's sort this out quick-smart as to where we are"
"Let's explore one more kilometre before we admit we're lost and turn back"

Search and Rescue Signals — Noises Repeated Regularly

One = We are searching for someone. (Is it you?)

Two = I acknowledge your signal. "Okay; Roger received-and-understood; I heard that"

Three = "Search for me. I need help."

Four = Recall of Search Party

E.g. rifle shots; loud whistles or horns; two boards clapped together

---

End of Section 2.

Nature's Norths

Learn these when the sky — the astronomy Norths — is clear, so that you can **lean** on them later. (Assume S. Hemisphere 1,2,3,5,6)

1. "Dew" North — Dew, moist soil & roads stay wetter in the shade. Frost will persist in the shadiest places, e.g dew South-West.
2. Snow South — Snow remains higher on the shady side of an obstacle.
3. Mud North — Half-dried mud-crusts betray the hot, wind-exposed cracked edges, with the damp mud in the shadowed places.
4. Rock Noon — a whole mountain, or a crack in a rock, can show up the sunny and the shady sides. (See next box)
   — Also, in the evening, after a sunny day, you can feel the colder side of a boulder or gully.
5. **Gully South** — The shady bank of a creek is steeper than the sunny side, at least on average, all around the world.
   The side facing the noon Sun is flatter. The side facing the pole is steeper. This results in roads, railways and buildings etc, being more common on one side—the flatter side—of an East-West river than the other. In other words, creeks tend to cling to the Pole-facing slope of a ridge, as if they migrate equator-wards. Perhaps this is a remnant of a past geological age when the erosion was more marked e.g. hot days eroding the sunny side flat, with ultra-violet light; and constant moisture and icy nights eroding the shady base steeply; therefore suitable mnemonics are: "lying down flat sunbaking" or "steeped in ice" (i.e. South-facing in the S.H.). Lakeshores have steeper or flatter banks according to the same principle.
6. Animal Orientation — "**Barnacle South**" (SH), on dark intertidal rocks — it shows up the wetter i.e. shadier side.
   'Anthills' (Termite mounds) sometimes orient to avoid the heat, so the length of them is North-South.
   Insects prefer the warmer side of a trunk in winter e.g. spider cocoons, ant colonies, boring grubs.
   Some spiders spin their orbs to face noon, like satellite dishes but more vertical; others face sunrise.
   Migratory flight-paths may be North/South &/or upslope/downslope. Animals flee in predictable directions, e.g. inland, uphill, North, upwind e.g. when faced with a fire. They often lie with their backs to a cold wind.
   Siesta South: Farm animals find the midday shade in a paddock. You can smell the urea. You can see the dung piles.
   (7—10 will be expanded soon)
7. **Wind West** — part of the W in the 'WEATHER' mnemonic. It is designed for windswept places when the wind is not blowing.
   The prevailing wind leaves its signature in the visible growth history of the plants, and in the landscape.
8. Wave West — if, in the Roaring Forties, you can see the accumulated effect of ocean rollers.

9. Botanical Asymmetry — where you sort out the competing effects of wind, light, salt, exposure.
10. **Bush Noon** — you never need to be disoriented in the thick bush.
    This works best in temperate latitudes — where the sunlight is directional.

Of course, anything to do with nature, refers to **averaging continual observations**. The most variable signs are the least reliable.
Less Natural Norths

11. Magnetic North; And don't forget magnetic dip — the downward tilt of the magnetic lines of force, varying with latitude.
12. Solar Panel Noon — Observe how we set up solar panels, and verandahs. Nature gives us the Sun, and we all follow it.
13. Satellite Dish Noon — Geo-stationary satellites are, by nature, Equatorial.
14. Microwave dish to microwave dish — Peaks 'naturally' point to each other! See "landmark to landmark network navigation"p42d

Instead of feeling bored, or getting lost, or being 'bushed in', you can have the valuable bonus of feeling at home in the bush, having a hobby, and being oriented to the slope, Sun, vegetation, ecology, weather patterns…
Bush skills are a good investment for future bushwalks. So go for some experimental walks in the bush on overcast, foggy, rainy days.

Don't Forget to Remember, Your Way!

Long-term memory is your best friend — so feed it. You aim for recognition, years later, through familiarity now
e.g. Have you walked the track at night? both ways? twice?
e.g. Collect as much first-and-second-hand information as you can.
e.g. Reread your track-notes occasionally before you lose the memories.
Photographs, Maps, Track guides, Track notes, etc are all permanent on paper! but you aim for permanent memories, not records.
I.e. Mental photographs, Memorised records, What the locals call "knowledge".
Use memory methods so you don't waste your efforts. Use the paper records to help you consolidate your local knowledge.
To that end: Make drawings. Make your own map. Keep a diary. Make your own track directions, names & comments. Tell somebody.
Use biro not pencil. Make a neat copy of your notes and file it.
"Look at…" means "Notice…" means "Take note" means "Take notes" i.e. Take a notebook.

Unlock Memorable Chapters in Your Life

Memory is fundamental in navigation (alongside direction, distance, position sense, visualisation and alertness).
Medium-term memory is how you find your way back, but aim also for long-term recall.
We remember best in definite 'bites' — location by location.
"Right, team; We have to remember this. Help me. We need to discuss it, describe it, mentally photograph it, draw its skyline, make it a memorable stop, time it, note it, put it in its logical place in sequence."

Go for the Jugular. Zero In on What Matters

Concentrate on the crucial and take-away information

— the crucial "directions"	E.g. the sequence of junctions
— the details that have direct orientation value	E.g. the visual angle between North and Start-off
— which give relevant memories	E.g. the hill-top view laying it all out for you
— track notes e.g. the bearing to follow, the "directions"	
— the overall trend — top-down, top priority pieces of information	
— the perspectives that fit everything into the context.	E.g. the surrounding trends in relation to Sun-North.

E.g. A mental map	is better than a 'position' on paper.	A map diverts you from the real picture.
Recognition	outvotes names;	Zero in on the most peculiar visual peculiarity
names	outvote position fixes.	

The overall details — What you need to know in three years' time, when you try again,
are the **time, distance and difficulty**.

The Sequence (of subsections & features) ought to be remembered...	..rather than e.g. 'bearing 035° magnetic'
Recognising Sections of the walk...	..is more relevant than having a series of map positions.
Surrounding landmarks last longer...	..than "turn left at the red tin-can-marker".

Appreciate the plants and animals; the soils, geology and climate. Instead of "Bin there; Dun that" you aim for familiarity.

The "Lie" of The Land

Think of this as the up-down part of geography. Think "Lie-Down" (versus "Lay-Out" i.e. a map layout).
Think of — Drainage, Watersheds, Catchments & Divides—the accumulated results of Ages of uplift, drainage, erosion & deposition.
— Geomorphology — it determines the shapes
— Ecology — it responds to the shapes, the slopes & drainage
— and how they all fit into each other.
E.g. "This north-facing slope supports warmth-loving vegetation"
"These foothills of the main range drain Westerly"
"The basalt uplift has generated these clayey mud flats"
"This is an isolated hillock within the main valley"
Think of the 'lie' as 'down toward what general direction?'
and 'rise' as 'up toward what encircling range?'
I.e. Put a general trend on major rivers and ridges. Imagine the coast, where all the (now nearly parallel) ridges & rivers run offshore.

Lift your eyes to the "Big Picture" scenery. Interpret it; Memorise its detail; Put it to good use.

Listen to the Pioneers

I.e. listen to the place-names

"The Great Dividing Range"	I suppose this one is obvious
"Blackbird Flat"	This means there is a river nearby
"Stockyard Creek"	A flat clearing is promised
"North Ridge"	This is related to the North Arm of the River
"Huon Valley"	It houses the longest tributary
"Boggy Plains"	Down on the flats!
"Eliza Plateau"	Up on the boggy flats!
"Bright Pass"	Means a change of catchment
"East Pillinger"	It ought to be East of West Pillinger
"Pelion Gap"	A saddle — between catchments
"StrathGordon"	It says "the Valley of" Gordon River
"-Dale"	Also says "Valley"
"Mole Creek"	Burrows underground

Lie Has To Do with Water Drainage

Look for dips in the skyline.
If you find a 'soak'—where water soaks out—look for a ridge-bound marsh uphill—where water soaks in.
Carry an artificial horizon, to reveal slight down-slopes, e.g. a flat plastic jar lid suspended on three strings, adjusted to horizontal.
Or lay your clear plastic drink bottle on the ground. Maybe even pour out some water to see where it runs off to. Listen for frogs!

Stick to the Track

- If you are waiting for the rescue party
- When a "shortcut" looks inviting
- If you have told someone exactly where you are going
- When you are nursing an ailment
- When sacred or heritage sites are at risk from tramping and erosion, rubbish, pollution, desecration, etc.
- When you are one more tourist in the constantly invading hordes, and "Please keep to the track" is aimed squarely at you
- When you are in a greedy hurry for more of the same elsewhere
- If you can't find your way out of a wet paper bag.

Reasons for Leaving the Beaten Track

- Tracks often fail to meet the needs of birdwatchers, botanists, photographers and bushwalkers. Remember why you are there. Talk to the ranger about whether the "No Camping" sign refers to drunken car-based parties, or to scientists and bushwalkers too.
- You sometimes want to search out the worst routes! Easy slopes have one sort of ecosystem/niche/microclimate; Difficult terrain has the more unusual biology.
- Interaction with a bare-dirt strip is not exciting. Fan out, Tread Carefully, Go Slowly, Keep your party size small — then you can appreciate the flowers, lizards, lichens. Each step is watchful — which forces you to see these things.
- A fast-track through a nice place robs you of appreciation-time. The experience passes just as quickly as the tramping to get there!
- Designed tracks will often steer you clear of exactly what makes a place interesting to see — cliff-edge views, water-edges, botanical hot-spots, refuge habitats,…

Isolated places often don't have tracks, and you will need to choose, navigate and negotiate your own route.

Cross-Country Routes

All routes are 'cross-country' but some don't have tracks to follow.
There are two sorts of tracks — Un-natural ones—artificial paths (man & animal made) which form a simple network,
and often stick close to the contours — gently upslope or gently downslope
And — natural routes (following the geomorphology) which cross the contours at right angles and
which form a double network.
I.e. Natural routes are determined by an interleaved network of ridges and rivers...
You are well advised to stick to these routes if only for ease of route finding.
Then there are three sorts of off-track navigation to consider…
— departing from the artificial paths
— departing from the natural routes (to follow a trend to a destination) and
— trying to keep a dead straight line cross-country...

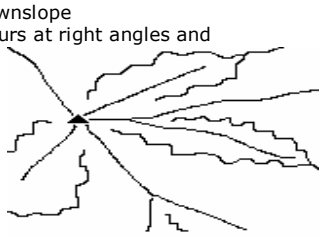

Rock Noon

Rocks blister and crack in the heat. In addition, ultra-violet light hastens all other erosion processes. The top sunny sides erode easily. Conversely, the shady side is subjected to lichen growth, icing, and moisture, especially at the base.
Differential erosion results in tilted rocks, in asymmetry of cracks and gullies, and in various light and vegetation effects.

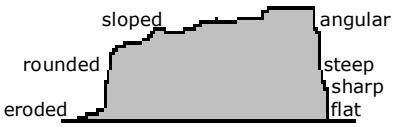

On the hot, rounded side, look for fissures, cracks, furrows, blisters and pock-marking — as signs of heat stress resulting in more rapid erosion.
e.g. Blistered flakes leave a patch of bare rock with no lichens on it.
Such patches are hard to find on the shady side.
Algae grow in the protected fissure under an almost blistered-off flake of rock.
The green stain on the underside of the flake is greener on the sunny side!

Caution: This is **unreliable** unless the rocks show a clearly interpretable pattern.
Of course, **use bedrock** or larger boulders, since they are going to be in their original position — not shifted about regularly.

Examine the cracks — The shady side is sharp and steep; the sunny side is rounded.
Just like "Gully South" but in miniature.

Examine the steps — sharp step-downs indicate a 'dip' in the layers down to noon.

— rounded step-downs show a geological dip away from noon.

Examine the colours — looking down-Sun you see pale, bare rock (where it has recently blistered).
— looking up-Sun you see multicoloured lichens on a darker background.
— diffuse lichen grows better on crumbly rock. The sunny side will grow such lichens.
— thicker-bodied, textured lichens grow older on the shady, non-eroding side.

Feel the topside vs the backside for a difference in temperature.
Look for green grass and moss in the shade of the rock. These are often far more reliable than the rocks themselves.
Beware that a fire-storm will kill lichens, moss & grass, and will blister the rock, crumble it and split chunks off.
i.e. some of the effects you will notice may be due to a recent fire, e.g. a firestorm from a particular direction.
Rock noon only needs to be reliable enough to tell you which end of your improvised compass points North.

Don't Try to Memorise Details without an Overall Framework

It's a recipe for confusion — to try to keep track of every bend, every hillock, every weather change, every creek bed, every channel.
There are too many of them. They all look alike in close-up.
So start with the map: Learn the **major** rivers and the divides first.
Go for the big names, the overall trends.
Fit the details in by retrospect, after the big picture shows itself.

Don't Try to Remember Anything You Haven't Memorised

i.e. put the effort in first—don't fool yourself. There is a process to memorising something, e.g. establishing a framework first.
And...
Don't try to remember anything you haven't made memorable. Don't battle uphill—make it enjoyable or emotional and it will be easier.
And...Don't write down anything you haven't memorised. First make it memorable, then write it down.
Notes are not an excuse to forget to memorise something!
Like most short-cuts, they are an enticing alternative, but they, like shortcuts, short-change you as well. Memories are precious.
And...Don't try to remember anything you haven't expressed or named. It is a fact of life that we refer to memories by name, in words
And...Don't expect to remember details you haven't understood, or organised, structured, or reduced to a bare minimum—a prompt.
It always helps to interact with, analyse and process anything you want to remember. E.g. investigate it with questions on your mind; criticise it; teach it. It also helps to visually set it out in your notebook with colour, shape, balance and symmetry.

If you relax, because there is a track, or because you have a map, or you have 'directions' for the trip,
and maybe someone else is with you who is familiar with the area, then you cease to build up your own memories and familiarity.
Again you have short-changed yourself.

Write down What You Memorise!

Does that sound like over-kill? It isn't. It's just good sense.

- Memory isn't perfect. So help it along. Do so before you forget.
- You will want to write down your track notes, for years later, anyway.
- Writing helps your memory — it "involves" you more — it stimulates your expression, revises your experience, adds time spent thinking about it, involves your physical participation and your verbal and visual faculties.
- You never know when slight stress (I'm lost!) will distract and divert you for an unknown time.
- To write it all down clears your mind of the previous task — you can concentrate elsewhere.
- On the way back you can easily glance through your notes and then know what to look for.

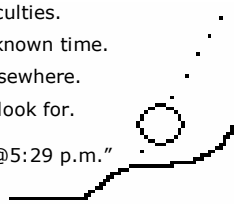

E.g. DRAW a picture of where the Sun sets on the skyline. LABEL it e.g. "Sunset @239°T @5:29 p.m."
Or was it 229°C @5:39? That's why you write it down.

Set aside a notebook for a "TRACK DIARY" (SH)

Major Rivers & Ridges, vs Minor Ones

Obviously, all water, falling on one side of a catchment-dividing ridge, runs to the sea, via a single river mouth; Conversely, river-mouth headlands are dividing ridges. Followed inland, all headlands must reach right around the river catchment—they are the two ends of the dividing ridge. The rivers and ridges form a sort of geometrically inverse pair of systems.
The catchments with the shortest creeks occur between named River mouths, and are called 'coastal'.
I.e. All other coastal land rises to some different, more local, dividing ridge; you may have to undulate up and down along it on the way
Not all creek-lines from a 'major ridge' (a catchment divide), descend all the way i.e. although the water may go all the way to the sea, the creek carrying it may not: the creek might only be a tributary, ending where it empties into a river. Similarly not all ridges running off the catchment-dividing-ridge reach all the way to the coast—some are "dead ends"—they end in a "shoulder", cut off between two creeks joining, or at a river bend. All valley-dividers are ridges, but not all ridges divide whole valleys, or even minor catchments. Those that come to a dead-end before they reach the coast are not dividing ridges, and may be called "minor".
All creek lines do not necessarily start on a dividing ridge: The creek itself might attach to a minor ridge, within the valley area.
Those ridges and rivers that transect or divide whole catchments are called 'major'. The named river ought to be a major line.
Unless you look for them on the map you may not be able to notice anything special about them.
Major ridges & rivers do however tend to stay higher and lower for longer, and often provide you with the **gentlest climbs**,
And they attract roadways.
Main ridges also tend to run parallel to the main valleys and vice versa.

Sort it out on the Map

- You cannot see catchment layouts from ground level!
- If you walk around a slope, half-way up, e.g. following an aqueduct, then you cannot easily know when you cross a major catchment. There are many minor ridges to cross and any one of them could turn into a saddle which divides valleys.
- Around peaks and plateaus, valleys spawn in all directions — it is hard to identify the right ones. High marshes may drain in several directions.
- Near a major river junction, travelling along the dividing shoulder can give a confusing alternation of views, between one valley and then the other.

Catchment Divisions Snake from Coast to Coast

Ridgetops fit parallel to and between the two highest contours marked on your map, and can be very sinuous.
They mark neighbouring catchment edges; & so theoretically you can cross the continent on highest ground without crossing any creek.
Major catchment divisions must reach the coast at a point each side of the river mouth (except, of course, for inland salt lakes).

A Catchment-Division* Separates Whole Valleys

*called a "watershed" in Australia, unlike in USA

Hint 1: Mark catchment divisions on your map with a pink highlighter — pink is opposite to blue-green for creeklines.
Hint 2: Now forget about those new pink delineations on your map. They refer only to the layout, as seen from above.
You need to imagine real, high ridges inbetween deep valleys.
Hint 3: Now forget about those razor-back ridges in your mind. You need to know that they could be remarkably flat paddocks.
— not very eye-catching — look for them carefully.
A catchment division will follow the highest flattest ground and consequently have a few marshes and puddles on it
A metre-high bump in a paddock might separate two raindrops into two different 1000 km journeys in each direction.
So you can't simply "walk downstream" — not until you know where you might end up!

Tracks are like Naughty Children

If you take your eye off them for a moment, they will split as soon as they are out of your sight.

A shortcut is often the result of a junction.
This is because tracks often bend at a T or Y junction.
Shortcutting is constantly encouraged at any bend,
So there will likely be an inviting well-worn shortcut there.
You can always split the party to explore both routes.

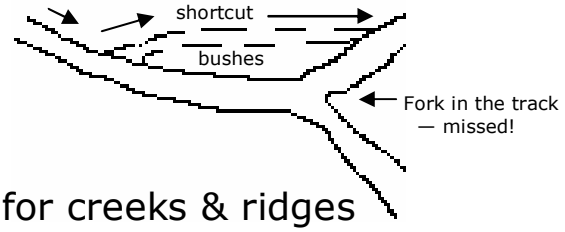

The same is true for creeks & ridges

Unless you stick to the bank of the riverbed, you cannot count on seeing the tributaries, and if you do stick to one side you miss the other side tributaries. Also, where a stream is "braided" into parallel interconnecting branchlets, and you follow one gutter, you are easily blinded to its tributaries joining in from side-valleys.
If you don't stick to the main gutter or crown, you miss seeing junctions on the far side of the river or ridge and can follow a wrong one. But side ridges are harder to see from the crest, just like tributaries are hard to see from the central channel!

Explore Branches While You Can

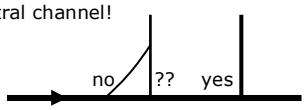

As a matter of principle, it is easier to explore something if you are there, than if you are not!
Be quick to remember that before you walk on past.
Remember… that you are there to enjoy the country. So put down your pack & amble off for a minute or delegate someone to explore.
You may surprise yourself by discovering a helpful signpost at a second nearby junction or a new access point and car park,
Or even, that it is the track you should have taken just now. Bends Breed Shortcuts: A shortcut may look like 'the junction we don't take' which means that you may mistake the real junction anyway.
Avoid endlessly having to wonder later… "But if that other junction turned east, then it might have been the link…" etc.
When you are there, find out for certain.

Convergent Tracks Lead to Civilisation

For example, you get lost, stumble across the track and meet a T-junction — which branch do you take?
You can examine any Y-junction for signs of splitting into two or more minor branches.
The most acute angle of a Y-junction usually points to a focus point (e.g. downstream; car-park)

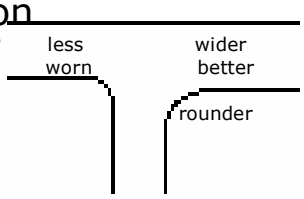

Down-Track vs Up-Track

This one is not obvious, and so it may trap you.
You seem to be following a well-worn path down, until you turn around and try to find your way back up.
It is then that you discover a multitude of other tracks joining into yours at an acute angle!
You see, people walking downhill, downstream, get channeled into the paths of least resistance, converging just like the creek lines do.
But people walking upstream or uphill are faced with choices and tend to spread out into all possibilities, until they find the right one.
The best path is worn in from down traffic, but that may not make it obvious to up-traffic.

Allow extra time when going up-stream, or climbing a steep peak, for eliminating false leads.

Wind West

Assume you are in a wind-swept place in medium to high latitudes e.g. southern Australia, where strong Westerlies prevail.

'W' in the 'WEATHER' mnemonic (p.114b) then reminds you
of the prevailing **W**esterlies in the high 30s, the Roaring 40s, the Furious 50s and the Screaming 60s.

'W' for '**Weather** systems come from the **West'** in temperate latitudes.
and '**Wind** in your face' locates these weather systems for you.
and 'Upper-level **Westerly Wave'** steers these systems to the E/NE/SE in a waving pattern.
A common pattern is to have a warm nor-westerly storm preceding a strong cold sou-westerly change (SH).
and '**Wind-West'** reminds you to look for the lower-level **prevailing Westerlies**, and their effect on the bushes, landscape, etc.

You don't have to be in the Roaring 40s to find a windy place and look for tell-tale signs of the prevailing wind.
— Onshore winds near the coast, being laden with salt, leave unmistakable clues imprinted in the bushes, and on the seashore.
— High plateaus and wide open clearings are usually windswept.
— Gullies produce up-gully breezes and down-gully gusts.
— Hilltops are windy. To get an undistorted picture of the prevailing wind direction, go to a hillock.

To navigate by the local prevailing wind may be more preferable locally, than to navigate by bush north.
The prevailing wind is not as universal or reliable as bush-noon, but it can be easily visible, audible and tangible.

In general, the **tallest trees** are the best tell-tales.
Look for wind-induced asymmetry**, especially at the tree tips** and the **oldest** tree trunks, when you are in a wind-sheltered zone.

Think Wind

Distinguish in your mind the various causes of wind damage…

- Storm Winds — brief but furious, as a weather change hits. They may uproot trees. They are often hot and dry.
- Fire-storms—from the hot dry inland, or sweeping uphill. They can leave a trail of dead black fallen trees lying parallel.
- Cold Change Winds — Cold, Gusty, Strong, Persistent for a day or three; Usually south to southwesterly (S. H.).
- Persistent Prevailing Winds — Westerlies 'Waving' between northwest and southwest in between 'changes'.
- Salt Laden Onshore Winds — They 'cut' the exposed vegetation back to size.
- Cold Blast Winds — These kill off the less hardy competitors.
- Hot Dry Blast Winds — These leave only the waxy/succulent/woody species.
- Sand Blast Winds — They remove vegetation, carve out sandstone sculptures, and remove lichens from the exposed side.
- Windy Season Winds — Ask the locals what to expect at each time of year.

Anything is Memorable

Make it personal	"Where I found the pocket knife!"
anticipated	Study the map beforehand to imagine the sections
active	Do something; involve yourself
important	Draw attention to why it should be remembered
associated	It looks like a … ; sounds like my middle name ; uphill from the mine
creative	Make up an enjoyable name
emotional	Expect to notice, then remember, decision points & anxious times
visual	e.g. symmetric, colourful, shapely
geometric	This triple-trunk matches the triangle of paths here
noticed	Draw attention to it "See how we have just entered tea-tree ?"
peculiar	That tree looks like a traffic policeman!
prominent	The tallest tree here, see?
liked	What aspect about it interests you?
named	e.g. "Midnight Junction"
fitting	"Power Point" where the electricity line crosses the lake
funny	'Willy Willy Waterhole' where we saw the dust-devil
human	"Betty's"

Make Individual Memories Memorable!

The problem is an excess of information, with a shortage of enough meaning to go with each detail.
e.g. hillocks and creeks have endless detail — you can't hope to remember all.
Visual memory is not the problem — pictures are perfect — but recall is difficult.
(And sorting pictures out later, into the original sequence is also hard.)
You need to add memory handles, worthwhile meanings, specific focus, etc.
Hint: Do something memorable there. Dance, Sing, Jump, Shout, Throw stones into the puddle.
Make a cairn. Sketch a tree. Have a meal — Chewing is said to aid memory!
Leave something of 'yours' there — something peculiar — e.g. lean a stick against a tree trunk; scratch the ground.
Insert it into a larger scheme — time & sequence; lie & surroundings; an acrostic poem.
Make a game of it. E.g. "Who can say what's peculiar about this creek?"
Make it 'yours' e.g. Discover it. Climb it. Explore it. Study it. Explain it. Draw it. Map it. Show it to someone. Name it.
Spend time there, looking.

Remembering Trees

Simply spend time looking. It may only be 10 seconds, but that is 2000% more than the usual ½ second!
Look at the branches, the shape, the trunk, until something strikes you
E.g. "Oh, these two branches grow across each other"
E.g. "The bark is right-hand spiral"
E.g. "See that ant-highway going up into the hollow—it might have water in it"
Walk around it, to see the shape from side-on. Characterise it & its context
E.g. "entirely shorter than that old stag's first branch"
E.g. "three tufts on top"

You get better with practice.

Ridges

Ridges have gentler slopes up/down ridge i.e. provided you stick to the crown.
The slope 'makes' rapidly on each flank. It is only the ridge-line which is flatter.
Ridges have nicer views. But these views are easily halved unless you walk right on the crest.
have nice breezes, provided you are right on the crest.
have more open and stunted vegetation plus shallower soil and more exposure to the elements.
are wider and flatter than a creek bed, and their rocks are firmer underfoot and dry.
Their run-off slides off downhill to each side.
These advantages are also lost as soon as you slip off the 'crown' of the ridge.
Off the ridge is steep, vegetated, narrow, loose, windless, hot, humid.
The steepest but best way up a ridge is... up the ridge because (1) it gets even steeper off the ridge
& (2) contouring is 'easier' but it doesn't get you 'up' but only into more difficult terrain.

Ridge Names

- Mountain to Mountain — e.g. Mount Ossa to Legges Tor (trace it on the map)
- Mountain to Mouth — e.g. Mount Ossa to Arthur River South Head
- Mountain to Headland — e.g. Mount Wellington to Wrest Point
- Catchment Division — e.g. The Arthur-Pieman Divide
- Tributary Divide — e.g. The South Arm Ridge

A ridge may not be "major" but if it divides your walk, you do want to know about it—so make up an appropriate name for it.

Mnemonic: "Ridges Rise"

Whenever you cross a ridge don't forget to ask: "Which way does this ridge rise, in trend terms — to the left or to the right?"
It is like asking which way water is flowing, but the truth can be obscured by local undulations in the ridge.
Ask: "Does this ridge, if followed to the (right), end further inland i.e. upstream, or further downstream?"
It must correlate with the overall corresponding drainage.

Uphill — Which Hill? Upstream — to which Divide?

To Which ridge? Which plateau? Which highest ground? Just as surely as rivers reach the sea, so ridges reach a peak.
Which peak? is the question which orients you.
Until you know which catchments you are climbing out of and in to, you don't have a grasp of where you are.

Sympathy and Understanding with Track Markers

Take time out to explore (e.g. leap-frog fashion) from one track marker to another.
Always guess where a track ought to go.
When you come across a marked trail—maybe poorly marked and/or difficult—the question arises: 'Should we follow it or not?'
Intelligent guesswork helps here

- Where would *I* put a track?
- What is the direction trend of the new trail?
- What is the slope trend of the trail?
- What tracks & routes might you expect?
- What alternative access routes exist?

E.g. uphill, heading West, may well mean: 'to Lake West from the lower car park'

If you leave the track behind, because it is unclear, you may never get back onto it.
But if you work with what you have (a bad track) it may be quicker in the long run.

Old Track Markers

Practise noticing them, before you need to.

- Old axe-cut blazes on tree-trunks — from the original surveyors
- Very old paint flecks in the blazes — search for them
- Blazes and markers on the far side of the tree — look back
- Old nail holes or nails or rust stains — the marker has fallen off
- Fallen markers — on the ground
- Fallen rock cairns — the rocks blend in, but show their pale underbellies
- Arrows on wet rock-faces — the wetness obscures the scratches
- Sawn branches, trunks and logs from when the path was cleared
- A clear line-of-sight through the trees
- Thick young growth — an unused road
- Smooth or shiny rocks with less lichen — worn smooth by feet
- Fallen trunks with a gutter worn across them — betraying foot traffic
- Broken branches of varying ages; bent stems e.g. bracken fronds
- Stumps cut off at ground- or eye-level.

Be Wind-Wise: Listen

Listen directionally — the wind-noise will be mainly from up-wind.
When you think you are navigating by the beach noise, you are probably navigating by the onshore wind, and should keep track of it.
Listen **through** the noise — You may distinguish three frequency bands, as the rumbles emerge from the hiss.
Listen through the **silence** — of night, in the stillness. Up-wind sounds carry a long way.
Listen to the **sounds** (birds, traffic, house, factories) — They carry clearly from up-wind only, especially the high frequencies (whistles)
Listen to the **beach** noise — The 'hiss' will be from a direction more up-wind, and closer, location-wise.
The 'rumble' will be from further away from you and further away from up-wind direction.
Listen to the **thunder** — Air temperature zones create directional effects, so that approaching storms, which are preceded by low level cold bursts of air, have their noises refracted up and out of earshot.
(See "Tips for Calm Weather" p.23d & "Topographic Level Winds" p.25d)

Obvious Tell-Tales: Landscape

- E/W coastal asymmetry e.g. Temperate West Coast ocean beaches might have a kilometre of surf, but East Coast beaches are narrower and calmer. Trade-wind East coasts are surf beaches, but headlands with West facing shorelines will be sheltered.
- Bays end in exposed points. Look for tall, steep, bare sandhills on the weather side. If there are tall sandhills at a narrow point, you will see sandhills from both sides, so ask 'am I seeing the green (sheltered) backside?'
- Sand dunes. There are variations, but bare dunes tend to be cross-wind and vegetated dunes tend to trend down-wind.
- Puddles, Lagoons, Lakes, Creeks, Beaches — Look for wind & wave action effects e.g. dark mudflats up-wind & white sand-steeps downwind. Shallow puddles often occur on the downwind end of a flat drainage.
- Sand ripples show steep backsides downwind, as the wind driven grains topple over the eroding crest.
- Clumps in sand-stretches show a long tail downwind. E.g. behind dead seaweed deposited on the dry beach.
- Grass-clump erosion is the reverse of sand ripples — look for steep bare dirt and exposed roots on the exposed side.
- Camps & campsites, picnic tables & fireplaces, bird nests & animal shelters all avoid the un-sheltered side of vegetation.
- Wind-shadow growth—plants grow taller in the lee of a mountain range, or of a pebble, sand ridge or house.

Obvious Tell-Tales: Windblown Objects

- Flotsam collects on the downwind side of a lake (but on the <u>sheltered</u> side of a coastal point!).
- Trunks fall downwind in storms; branches fall downwind in gales; leaves, nuts and flowers fall downwind in strong winds. Look for ribbon bark hanging in a tree downwind from its source.
- Rubbish (old dead twigs & leaves) collects <u>behind</u> a clump.
- Dead long leaf blades lie cross-wind when blown <u>against</u> an obstacle e.g. a clump...
- Dead long leaf blades lie down-wind when blown <u>from</u> a source e.g. a tussock...
- Broken off branchlets lie with their <u>big</u> spread-out twiggy end downwind. Meadow-grass tufts pulled up by grazing animals also.
- Feathers catch in bushes, usually on <u>downwind</u> branches, and mat against the <u>up</u>-wind side of those branches.

The Difference between Looking at, and Seeing

..has a lot to do with involvement.
e.g. a seed-collector will see the vigour of the plants and their stage of fruiting. Would you even see the plants!
e.g. a hunter will see a route for stalking from behind cover.
e.g. a photographer will see the contrast in light.

Would you as a navigator care to notice footprints, or their absence? What about the flight-paths & contrails, or gems in the river gravel?
Navigation is your opportunity to see more.

Search-Image

It is possible to stare at something obvious and not see it! e.g. the pattern of wind-blown grass stems, or the peculiar geometry of a tree. But after some <u>practice</u>, you can't miss it. So train your brain to the search images <u>before</u> you need them. When you are flustered and hurried you will not feel confident in your abilities to see something difficult, and will likely give up too soon.

We Recognise '<u>Sameness</u>' But Need to Describe the <u>Differences</u>

A gum-tree is a gum-tree; but <u>this</u> gum tree has three trunks!
A 'bend on the road' is just a bend; but <u>this</u> one bends away from sunrise.
A 'hillside' looks much like all hillsides; but <u>this</u> one is regrowth with wattles dominant.

Don't just gaze at "the view" or "the landscape" or "vegetation". Look <u>past</u> the fuzzy 'sameness' to the <u>unique</u> peculiarities.

Unfortunately that often means: 'Focus <u>on</u> the boring <u>detail</u>'
e.g. 'This hillock has five dead tree trunks, totalling nine main branches' which makes this hillock recognisable and unique.
So. Hint: **When all else fails, COUNT!** (The same goes for botanical species identifications.)

There is often no way around <u>a detailed look</u>. There will not be anything unexpected to hit you in the face — you must keep looking until the interesting peculiarities emerge.
E.g. "No other dirt road junction I have seen has an Acacia sapling growing in each corner."

"Shoulders Stop"

A huge mountain range may end in a "shoulder" i.e. a 'shoulder' amounts only to a 'minor ridge' stopping within the valley.
You will not get to the coast by following it down. (If you are on the very last ridge next to the coast it would be called a headland).

There may also be isolated high ground within a valley, a lot like an island, and this too is doomed to end in only 'shoulders'.
Such high ground is drained by one river only, but via several tributaries.
It will have one, often low, flat connection (a saddle) 'uphill' to the main Divide.

Shoulders are Not Easy (like Ridges) To Walk on

1. They end a spur, and so are likely to be steep, as the ridgeline plunges down to the creek bend or junction below it.
2. They are wetter than ridge tops. Although they shed water, being convex, they are still fed with run-off from above, as on a hillside. So the vegetation can easily become very thick.

Saddle — A Flat between Hills, Sitting Astride a Ridge

They are so named for their resemblance to the two-bumps-two-gutter horse-saddle shape, designed for two legs astride a sway-back.
When a flat area is drained by three creeks and so has three ridges lifting from it, it is called a "monkey saddle".
The third gutter accommodates the tail. This can cause uncertainty-of-choice for navigators!

A "col" is a saddle.
A "pass" is usually a low point in a high ridge i.e. it is a saddle.
A "gap" however may simply be a river gorge.

Every undulation and flattening of a ridge leaves a saddle. In forest, there will be a puddle. It may also attract a campsite or a road.
A saddle may look like a valley crossing—a muddy dip in the path—but the two sides both drop away, if you look carefully through the wall of vegetation.

Weather maps show pressure 'saddles' between two High's and two Low's—p.11d.

You can also picture a saddle as a U-shaped valley going up and over a gap between two hills and continuing down the other side.
Saddles make creek sources, but not every creek source begins as a flat saddle—any ridge or slope may spawn drainage gutters without necessarily flattening out.

Track Markers

— Decode the colour-code. Separate tracks often have separate colours. Reverse track markers may be of a different colour.
— Successive generations of track marking may be evident, or several colour-coded tracks may share one leg of the journey.
— Triple marking, often means "End of Markers" or "End of Track".
— Big cairns mark junctions, campspots, features. Cairns with a stick may also indicate a junction.
— Side-tracks are marked with something extra or different.
— Sticks over the track, or especially crossed sticks say "No Entry. Wrong Way. Go Back".
At an unexpected junction, if you have to step over some branches you have probably missed the intended message.
— Un-needed signposts — often mark the entrance of disused tracks. They may (or may not) be at an unexplained bend in the route.

When You Lose the Track-Markers

Be suspicious! Start by over-navigating beforehand — search for all track markers as a habit, even when the track is well-worn and plainly visible. It gives you the practice for when the track marking does get patchy.
Also, only then will you notice whenever the track takes an unofficial shortcut (e.g. following vehicle tracks, not the walking track).
Or maybe you have taken a wrong turning off the marked track and need to realise it, no matter how good the new track is.
Whenever you lose the markers or whenever you lose the track, explore in these directions: **straight ahead and straight behind**.

One or two markers may simply have fallen off — so maintain the trend
You may be approaching a main track junction with the first few sidetrack markers removed to conceal the sidetrack you are on.
An unexplained bend, which deteriorates, may be explained by a missed junction/bend...

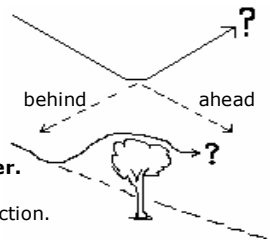

Where a tree has fallen nearby, try to pick up the trend of the old track, on the far side...
Backtrack to where you went wrong — to the last bend & from the last certain marker.
Next, explore the unexpected directions.
Very often a marker will be placed at a bend, because of the bend, or even because of a Y junction.
Then explore in all directions — "Meet back here in five minutes" might save 50 minutes.

Notice (with Alarm!) When...

..The track itself **deteriorates**. You may be on a false track, made by people doing just what you have done
e.g. you overran a junction or bend; you sidestepped a fallen tree without noticing.
..Boot prints, that were there before, cease. Especially on a 'through' track.
..The track gets **better**! This means that in back-tracking you may miss your small track and follow the other track which joins in there
..The country abruptly **clears** e.g. a beach, a peak, a clearing, a road. There may be no way to find the single small track you used.
By the time you remember to look back, you may have lost the way back out. There may be too *many* paths to choose from.
..Any **barrier** e.g. a beach, a road, a fence, a riverbank — clear or not.
When you hit the barrier you do a right-angled bend, but later you may not remember exactly where you met that barrier.
..Any **catchment change** i.e. crossing a ridge, or even a creek crossing—into "another part of the catchment".

Name the enemy: Be the first to say out loud **"Abaracadabara! Peek behind you! because it's a...(fill in the blank)"** A **B**arrier; A **R**idge/River/Road Junction; A **C**learing; a **D**own-track (or focus point); A **B**etter/worse track; A **R**idge/River crossing; A **Peak**.

Light & Colour Effects

This is one of the delights of nature! Chances are that you will be startled, even if you know what to expect, but only if you look for it. You need to wade out into a uniform reed-bed, preferably when the Sun is side-on to the prevailing wind, then look up- & down- wind. Up-wind, the tussock tips are pointing toward you. You then see between them into the deeper shaded body of the plant.

- The view up-wind is dark & colourful. The best effect is from rich brown or dark green, drooping, chest-high reeds.
- Down-wind the tips lie flat and reflect the sky shine at you — faded & bright.
  The prevailing wind may result in permanent sunbleach on one side of the grass blades.
- On a windy day you will see gusts causing bright patches in a dark background, due to the same flat blade effect.
- To neutralise the directional lighting of the Sun, look down-Sun — to your left & right will show pale vs colourful.
- The combined effect of the Sun and the wind result in four quadrants — darkest up-wind; shiniest downwind; greyest up-Sun.
- This happens even with ordinary green plants — look for it in any low, flat, wind-swept vegetation.
- The prevailing wind doesn't need to be blowing to see its permanent effect on the plants.
- At sunrise in the reed-bed, you should notice a 'funnel-web' effect on your right — like the circular scratches showing up through a sunlit window.
- At other times you may simply notice that any shiny leaves on the left of the Sun tilt left, and on the right of the Sun tilt right.

- Looking up-wind anywhere is greener. Downwind looks more bleached — due to leaf-tip scorch, leaf fading and leaflessness.
- Weather-side sandhills are white. Lee-side sandhills are greener, due to thicker growth.
- Exposed lake banks are steeper & eroded. So, sandy lagoons show white sandy vs dark muddy ends.
- In sandy heaths, downwind shows up as windblown sand patches of erosion. Up-wind shows up as continuous plant cover.
- Rocks in sandy places are sandblasted pale on the up-wind side and darkened with lichens on the downwind side.

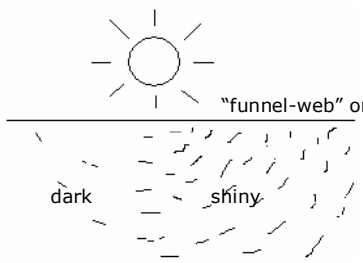

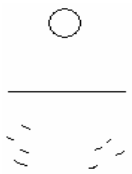

Shiny leaves

Expect some variations

Place Recognition
You Recognise What You See, By What You Can't See

Every locality fits into a context — the invisible surroundings.
How long did I take to get here? How far from the turnoff? In what direction from home-base? Uphill or down or contour?
How many creeks have we crossed up to now?

To know where you are, you can't ignore where you are not.
e.g. 'It was an all-day slog to get there, we finished the day worn out'
is part of where you were, and fundamental to your memory of it.

Landscape recognition is relative.
That involves your senses of time, direction, distance, difficulty, slope, topography, layout, number and network.

Aspects, invisible to the eye, which surround a place, live in your mind's eye.

Description Requires Words
Can You Describe All That You Can Recognise?

E.g. a magpie. What pattern is on the head? What colour legs?
E.g. your brother — what colour eyes? what voice accent?

"Recognition" means "I need to see it before I can recognise it"
"Description" is "I can recall it to my mind and describe it"

If you don't deliberately memorise something you can only recognise it but can't describe it.
So thoughtful examination, notes and practice, can fill in a lot of blanks.
Repeated visits build up a subconscious picture, but you can speed up that familiarisation process by **deliberate scrutiny**.
Expressing the description in explicit words, forces your subconscious 'recognition' powers, into a new dimension.

Hint: Open your mouth; say: "The place where… " and see what comes out spontaneously.
try: "In between…" "Running alongside the… "
"You probably would remember it as the time when… "

Notice When You Cross a Catchment!

It is a major method of navigation
e.g. at night; or in a blizzard; or when you are lost; or in a forest; or descending from high ground; or exploring upstream.
But you need to get used to noticing ridges and to counting creek crossings and to matching the catchments to the map.
Notice also, when you cross from one side to the other side of a catchment.
Ask: Is this riverbed likely to be the lowest, largest channel in the whole catchment?
Look for creek-line vegetation, as sometimes the only clue to the drainage pattern.
Look for any darker patch in the road or track or beach, as a tell-tale sign of a dry creek bed.
Look at the track ruts. If the rocks are exposed, water flows over it. If the ruts flatten out, water drains away from the track there.
A subtle rise, dip or bend in the track can be traced to an obvious ridge or drain crossing it. Look sideways to appreciate the lie.
Assess any puddle, marsh or mud hole — Is it a creek line or a saddle top? — Is it in a dip or on a crest? — Can you trace a gutter?;
Notice any crest; Trace ridges to a nearby peak. I.e. Don't get caught whingeing about the government, instead of noticing the lie.
Ridges divide: they shed water. Creeks collect: they catch water. Valleys drain: they empty into the sea.
Rivers cut the catchment in two.
See "Understanding Contours" for some hints — later in this section. See also p.83b.

The Giveaway! A Change of View — Mnemonic: "Catch the View"

The new panorama is the tell-tale sign that you have changed catchments.
Perhaps you are contouring around a headland and catch the new view around the corner, or you are sidling around a ridge, or you are breasting the saddle, or reaching the peak, or cresting a paddock… It is the new vista that greets your eyes which tells you that something has changed. It is also the loss of the old valley view.
Well, maybe you can't see the view, but you may now be able to see the skylight or even blue sky down through the trees below you.
When climbing, you find yourself looking down instead of looking up. Your muscles, of course, tell you the same story.
The deeper the new valley, the more far-reaching, on average, the new view.

You Can't Cross a Creek Twice from the Same Side

If you make two successive creek crossings in the same flow-sense, you have crossed a creek divide into a different creek catchment.

Alternating crosses=same creek...

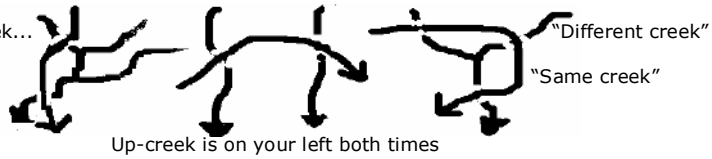

Up-creek is on your left both times

The same rule can hold for ridges, but only those which slope uniformly down from high ground.
Otherwise an unexpected crest slope, instead of being a 'different ridge', may alert you to 'a local up-trend' on a down-sloping ridge.

Mark Your Trail

a pair of stones...

Two principles are needed 1. Trail marking takes extra **time** — budget for it.
And this one is widely applicable… 2. It's easy when you know how, so **think** how!

a gumnut on a rock

Mark Your Trail Non-Permanently

E.g. when you could otherwise get lost

- Use what is available — pebbles, leaves, tussocks — throughout the area, for ease and consistency.
  e.g. "always look for a pair of stones"; "tie a grass blade on the top shoot of any shrub"
- Leave your mark in a visible place. E.g. "look for freshly broken twigs at the base of each trunk on the path";
  "check the topmost leaves of anything chest-high for a mark"
- Leave your mark in an expected place. E.g. cross-country: Zig-zag in straight lines between prominent features, so coming back you know exactly where to look for the next markers, and searchers do too.
  E.g. hillock, to tall tree, to clearing, to outcrop.
- Do something only slightly unnatural: (see diagrams) A gumnut on a rock; a twig through a leaf; a pair of torn leaves; grass stems bent in a consistent direction; or tied in a knot; or crumpled; a scissor-cut dead leaf; a scissor-cut fresh leaf; bark hanging from a barkless branch e.g. gum bark in a Banksia; a twig drilled into the ground; a leaf wedged into bark at eye level; a triangle of twigs; a half broken bent twig hanging in a shrub; an opposing pair of gum leaves; a trail of sticks; windblown debris which is reversed; upturn a rock onto leaf litter
- Deliberately plant your boot print when you can. Step on top of older footprints, deliberately.
- Use fresh material so that your trail may be 'aged' according to the fading or drying e.g. of green leaves or mud.
- Position markers so that time, animals, wind & rain will disrupt them (at a recognisable rate). E.g. a teepee of sticks.
- Try a snow cairn; chalk marks on rocks; cotton tied around a branch; a stick dragged in the dirt; a scratch on a rock; balls of mud; charcoal or charcoal marks; tiny shreds of paper.
- Choose one method and stick with it, with a backup method in reserve.
- Double-mark a trail e.g. when doubling back, by duplicating it.

a cut leaf…

dead twig…

Tracking Someone Else

- Look for the **absence** of: spider webs across the path; dew on overhanging branches. To discern a foot-path look for the absence of fallen branches; leaf litter; lichen & moss; old wombat dung on prominent rocks; unbroken sticks; un-trampled seedlings; undisturbed soil
- Look for the **presence** of freshly disturbed stones/soil crust/puddles, plus footprints.
- Examine puddles: for wet edges — indicating a recent forward swash; examine for muddy water; look for muddy water in the deepest layers — indicating disturbance some time ago; for wet stems above the water — they have been trod-den in and have sprung back
- Overlapping fronds across a path — they 'remember' which way they were last pushed through.
- Don't step on the boot-prints or tracks you are trying to decipher—you often need to look at them again.

Botanical Tell-Tales

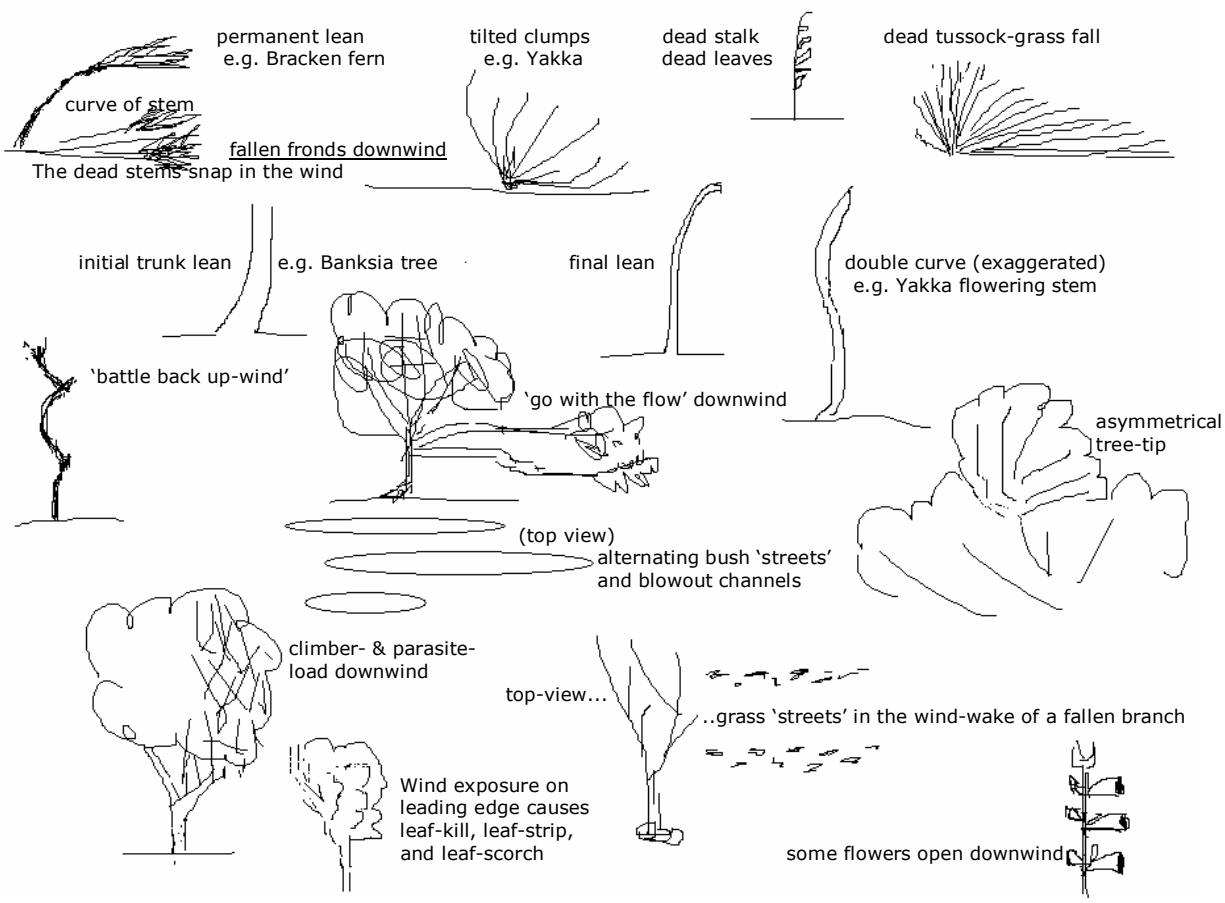

Discuss It

Let others open your eyes: "What can you see about this?" "Where does it remind you of?"
Ask Questions: "How do you know where you are?" "Could you get back here again next year?"
Make a game of it: "What else can you notice from here?"
Express the specifics: e.g. "That boot-print has zed-shaped ripples on the right and ess-shaped ripples on the left.
The heel patten is a smaller version of the front half, as per normal."
e.g. "This tributary lies northwest/southeast and points to the left of the peak."

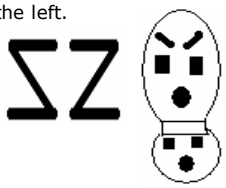

Many aspects of navigation yield to these simple tricks…

discuss it express it out loud make it the topic of conversation
delegate the tasks between yourselves
keep yourself alert
make up your own names
collect the details integrate them

Collect the Details

Identify a Specific Place by its Specific **Collection** of Details

E.g. "A low conical hill in flat terrain south of Mount Cameron"
That is not 'interesting'. The bits are not 'peculiar'. But the collection of the bits is unique.
E.g. Plant and insect species are often so similar to other family members that no one character is 'characteristic' enough.
The species will be identified by a set of characters.
E.g. Is this our old road?? Does anybody recognise it from last year?? but it has…

trend e.g. toward sunrise
slope e.g. mostly uphill
straightness e.g. can't see for more than 20m mostly!
traffic intensity e.g. many fallen trees across it, and encroaching bushes
length e.g. it goes on and on and gets nowhere!
vegetation e.g. set amongst old gums with yakkas under
ends e.g. it came from the saddle and ends up behind the peak

If you think you recognise it, by the combined characters, you will probably be right.
E.g. Make one name for the big tree on the crest; another to name that crest; another to describe its relative position…
and soon you can't mistake it…"the big tree with the south facing hollow on the port-side Dam-wall Ridge downstream of Tree-fern Crk.
The add-on bits are part of the name of each bit i.e. "the Big-tree Crest" (downstream of Tree-fern Creek) (near the dam wall).

The smallest collection is two, but that is a powerful combination, e.g. a peculiar log, plus a recognisable hollow trunk, is almost unique.
The combination is the peculiarity to remember — a third peculiarity.
The geometry of the combination is a fourth and even more powerful memory aid… "This log points to that hollow".

Include What You Can't See

For instance, have you noticed that there is no bare ground showing? That you can't see farmland? That there are no intervening ridges?

Break of Slope/Make of Slope

The 'make' of slope is the **steepening** as you go downhill ie the "slope" is a down-slope. The best views are at the most abrupt 'make's
The 'break' of slope is the first **flattening** out as you descend. The thickest soils occur below the 'break'.
The make and the break refer to points which separate different 'slopes'.
There is a difference between a making, or eroding slope
and a breaking, or sedimenting slope…
A ridge-top is a make of 2 slopes, in cross-section, but any abrupt steepening
can also be called a make of slope.

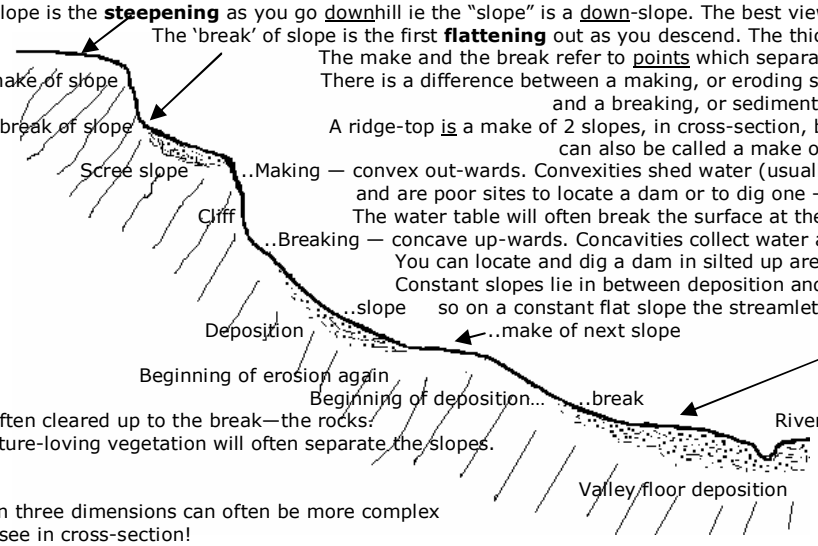

What you see in three dimensions can often be more complex
than what you see in cross-section!
— When convexity changes to concavity there is a **momentary straightening** — a 'break of slope'.
This is also called a 'point of inflection' in the curve.
It's not just a meaningless mathematical quirk; it betrays a set of real changes, resulting from the erosion history.

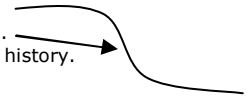

Scree Slopes — Tread Carefully!

The boulder-field at the base of a cliff, gives you a false sense of security!
Each rock seems so solidly jammed in place, but the whole field could 'slump' and slide all the way down, if given too much of a shake.
That is why there are no old trees growing in landslide-prone slopes!

The essence of a scree slope is, like a pile of gravel: It is in a continual state of critical instability, as evidenced by the straight-edged slope (flat but not level), about 36°. 36° doesn't sound like much, but up a sandhill, it puts you on all fours!
The rock surfaces look 'fresh'. Soil doesn't accumulate.
Tread carefully. Distribute your weight on all fours. Tread on bigger rocks.
Choose the cliff-base — it has hand-holds — or the scree-base — it has a solid bottom.

Mark Your Trail More Permanently and Ostentatiously

E.g. when you are lost or in trouble or in doubt

- A teepee
- Mud upon a tussock — quite unnatural; A tussock upon a tussock!

- Hang a piece of clothing from a prominent tree limb
- Dig a hole and make a hump

(Direction of travel is ⟶)

Mark Your Trail Invisibly

E.g. so that other walkers who are not lost, don't disrupt your markers, out of a sense of bush etiquette.

- "One mark every 50 paces". E.g. one pointed pebble. You may like to use double-paces.
  It is almost impossible to notice accidentally, but impossible to miss if you look.
- "Every 10 paces, as an option" added to your scheme — to cope with junctions and bends.
- Place a pebble on the ground, on the onward side of an existing cairn. You can recognise it, and scatter it, when you come back through the fog among a maze of similar cairned routes.
- Put a small mark at the base of each track marker (e.g. under blazes-on-a-trunk or ribbons-on-a-branch)
  — you can easily find it, and know that you have been along that track.
- "Half way between (2 markers)" or "10m North (of a juncture)", can remove the marks from obvious view or significance.

A pebble Code — Keep It Simple

All codes have to be memorable to all users — complications never survive the years of disuse.
I.e. use 'one' to 'four' only, preferably only 'two' and 'three'.

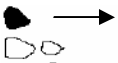

Tread the pebbles into the ground for clarity & permanency.
Upturn the pebbles for visibility, and **point the stones** appropriately.

Use the smaller pebbles at the front.

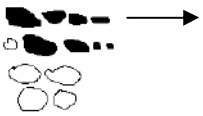

One: "I woz 'ere, going that way"
Two: "Me go that way"
Three: "I'm lost, exploring that way"
Four: "Look for a message, there"
Five: "Someone has seen this message"

Square: "Don't go this way"
Circle: "Oh" for "I'm OK"
Bull's-eye: "I will be back here soon"
Junction: "I have turned off (and returned, then continued)"

Double marking: Close the pebbles up the second time past, until they touch.

Combined Effects of Wind, Salt and Light

When assessing wedge-shaped shrubs, leaning trunks, tilted canopies, etc., try to distinguish wind-cut from salt-cut and from light-lean. But beware, in any natural system, it would be easy to over-generalise from a few known specifics, or falsely extend an inadequate interpretation.

Here is a typical wind-ramp. The leading edge consists of dead twigs.
The vegetation gets greener and thicker and taller downwind.

But you cannot easily tell whether it is the result of salt-laden wind, or strong wind only.

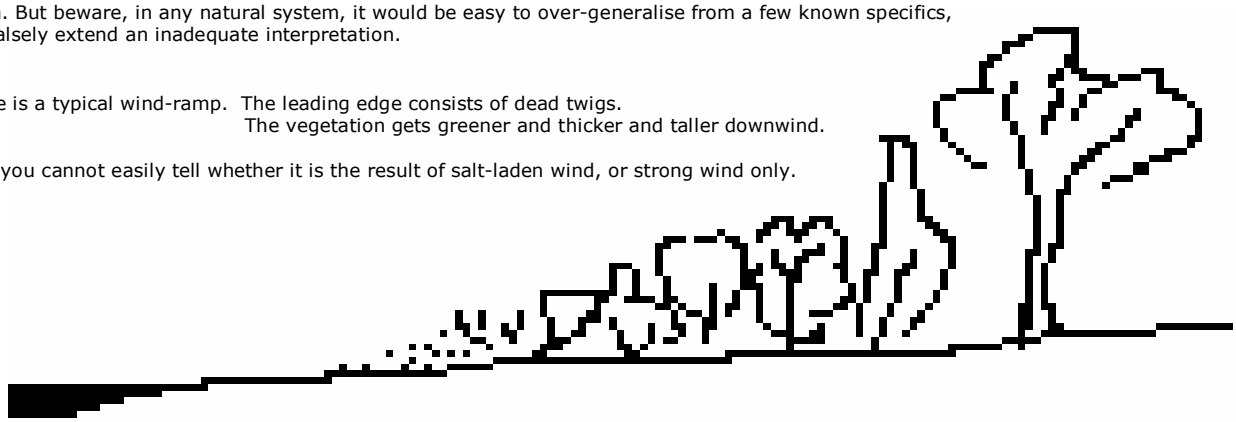

Salt-cut will show you the direction of the prevailing onshore wind only.
Inshore, wind-ramping will show you the direction of the prevailing wind, which may be significantly different in direction.

Botanical Asymmetry

Plants must respond to the wind.
If they did not deliberately push up-wind, we would be walking through a landscape of leaning-over plants.
Therefore it must be possible to 'read' the effect of strong wind written into the branching history of a plant.

The branches of a bush may grow according to the strong wind, but the leaves will still respond according to the asymmetrical light.
The two effects may or may not be independent and you will need to examine them carefully to see the difference.
Look for asymmetry in the green parts of a plant to show up bush-noon and look for asymmetry in the other parts for bush-West.
If there is a macro response to the wind, which will confuse the macro light-response, look for micro signs of a light response, on individual branchlets.

You don't "need to know your plants" beforehand. Instead you examine the species which present themselves.
You assess what responses they make to light, wind, shelter, etc.
E.g. Do the stems colour up in sunlight? If so keep looking at that.
Do the branches throw out more leaves on one side? If not look for something else.

Extend Your Familiar Range Systematically

Explore in every direction, in ever longer legs, radially. Leave recognisable marks alongside recognisable things.
Then explore in concentric circles to make sure that you can recognise individual trees and rocks.
You know you can explore further within the endless sea of scrub, if you have 'familiar friends' which you have seen repeatedly.

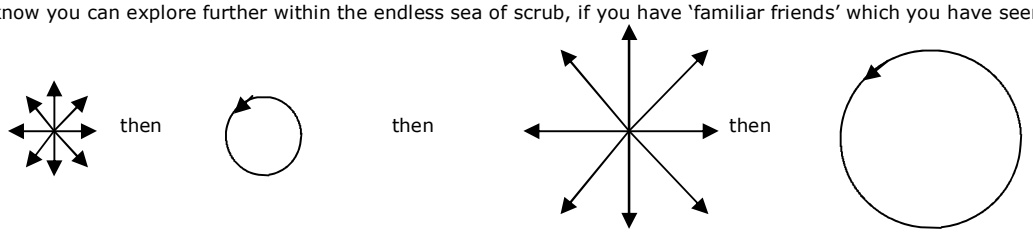

Markers for future recognition e.g. on your own property…

Plant a tree Graft two branches together Suspend a forked branch into a forked trunk
Arrange logs e.g. a parallel heap Arrange rocks e.g. a cairn, or a rock on a stump, or a rock 10 m south of the biggest tree
Dig a trench Etch a mark Paint a rock Leave something solid behind
A stake A signpost A numbered metal tag A path A Campfire-place or Barbecue

Hints

It's the Campsites You Don't Use, that are Easiest to Forget

I'm forever asking: "Can anyone remember" whether we can camp between here and there?
Take notes of potential campsites. It makes your trip-planning more flexible, next time.

In general, take the opportunity to spy out and note down potential future trips — side-tracks, ridges, climbs…

Leap-Frog Walk in New Territory

I like to walk a new track twice, anyway. It is no big difficulty to take your time, once, and achieve better familiarity, by leap-frogging.
Explore 2 legs forward then one leg back, and you cover the ground 3 times on the way out, once more on the way back.
4 legs forward, 2 legs back, and it's like walking a return track 3 times, but it gives you a much more solid handle on each section.
You gain what you want, appreciation and re-preciation, and what you need, familiarity and sequence memory.

The best way is to combine this with a mnemonic for each leg. See AA to ZZ for an excellent method.
Its also a good time to mark each new leg discretely.

Break of Slope — And Route-Finding

- The first signs of an approaching ridge-top when climbing include: the change from deposition (breaking) to erosion (making); i.e. onto outcrops; into exposure to the wind.
- Take a rest at the break of slope — before climbing upwards.
- Get your bearings at the most abrupt make of slope, where the view is clearest.
- Look up- and down-hill for the lowest skylight beneath the trees, to reveal the change of slope.
- Avoid contouring around the break of slope — it is tempting, but — it is steep and unstable.
- In rocky gorges, or on any creek-bank, or around any permanent water, look for a break of slope just beneath water-level. It may be safer, than rock-climbing higher up, to get your feet wet, and use such ledges. The slope near the water may be less steep, at least on average on jagged slopes; you can't fall as far; your landing is softer (wet) — if you 'keep as low as possible'.
- On steep gravelly roads, stick to the 'crown', or the 'gutter', to avoid slipping. The crown will snake from side to side as the road turns.
- Slopes are awkward to walk on—steeper than the ridge or creek, so, when you drift up onto the sides of the creek you want to follow, you will be forced further up — onto the ridge. When you drift off the crest of the ridge you want to follow, you will be forced further down — into the creek bed. In this way 'minor' detours (e.g. around an outcrop, on the ridge or in the creek bed) may result in major redirection orders!

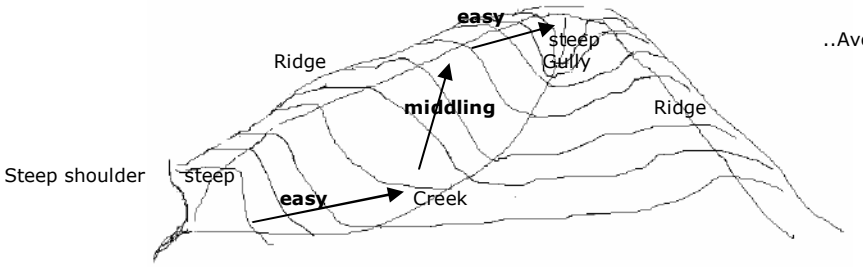

..Avoid the steep bits.

Keeping Track of the **Bends**

Don't just follow your nose — follow the bends!
There have already been some hints on keeping direction sense:"Walk across Australia!" & "Integrate the Jetstream and Earth rotation"
Plus the "Sky to Land" and "Land to Land" navigation systems (p.29e,34d,38d,42d).
Following a track can be mindless. But following the bends in the track means that you understand
— that you have changed direction, and by how much.
We have an ingrained bad habit — that we walk 'away from' the starting point, which is therefore 'behind' us.
Only when we 'turn around' is it 'ahead' of us on the track.
We usually only notice junctions, features & hunger, not distance or direction or exact time.
The commonest mistakes of guesstimation, are failing to notice a track curve,
and **underestimating** exactly how far a gentle but insistent bend will **turn** you around i.e. simple disorientation.

Hints: Notice the bends, with your eye on the sky or the skyline. Establish the habit of looking back.

Desired Equipment

- Compass — look at it often and interpret it
- Polaroid sunglasses — look for the dark blue band
- Radio — use it to keep orientation if you don't have a compass
- Map — orient the map to the landscape, and it acts as a rough compass
  — lay a rhumb-line through the bends, as an average course, as a reference 'track'…

Logging Constant Diversions

A stick or string helps you measure direction changes…
Held at arm's length, you add or subtract, as you decide to change course left and right.
Thus you can revert to an intended course after a short 'blind' section,
and average out the course changes even without a compass, or being able to see the Sun.
You could use the skyline as a reference direction, or lineup two tree trunks for each change.

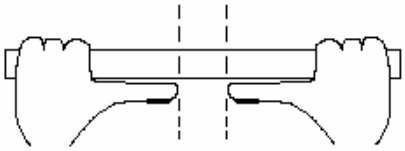

A rougher method is to counterbalance your diversions.
If you go left-around the rock; next time, go right-around the tree.

Keeping track of countless bends is one of the few major problems in navigation
— like keeping a log of distance covered, having no location sense, and trying to stay alert.
So aim for an independent cross-check e.g. a landscape fix or a distance log.

Think like a Plant

E.g. Imagine you are a small ground-cover plant. In a desperate attempt to not be eaten or trampled to death you will probably branch and leaf out in all directions to simply survive... So don't expect much botanical asymmetry underfoot.

The idea is to become an amateur ecologist. For instance, you could work out the pollinators available, the dispersal mechanisms used, the biogeographical zones and overlaps, and the associations between species.
Therefore think about the geology, geoform, latitude, altitude and soils.

Temperature & Relative Humidity explain a lot of botany.
E.g. very cold morning winds are also very drying when they warm up later in the day.

Altitude, for instance, affects the species mix, so that with practice you can gauge your elevation above sea level.
You notice when sub-alpine species kick in e.g. Telopea (Waratah), then when true alpine species begin e.g. Richea (Scoparia).
Individual members of genera have altitude preferences and so are helpful to distinguish altitude
e.g. various Coffee berries (Coprosma) or Pink Berry (Cyathodes).
N.B. Since the prime effect of altitude is to drop the temperature, this is really a latitude-dependent temperature effect.
I.e. Going further polewards, the altitude preferences will change.

'**Aspect**' (i.e. South sloping or North facing) will be important in interpreting the presence of particular species.

Living stresses explain simple botanical systems.
E.g. In the desert, late afternoon is the baking-hot-sunlight side (cf noon in forest). Trees will be shiny leafed — to shed excess light (cf searching for more).
So consider the prevailing winds, salt spray, max. and min. temperatures, wet season, disturbance, trampling, fire, clearing, introduced competition or diseases, grazing, salinity, mineralisation.
Pioneer Species — These are the easiest to identify, because they are lonely.
Their preference is the saltiest, hottest, driest, highest, coldest, poorest, windiest, sand-blown, most flood-prone, …
If life becomes too easy, other species will smother them out. Hence, look for those that don't mind the hot baking Sun, etc.

There is a height-**succession** of plantforms grading from **lowest to tallest and deepest rooted**.
Ground lichen; moss; grass; herbs; creepers; shrubs; trees; ' emergents '
The depth of soil is crucial. The taller ones shade out the humbler ones, and their deeper roots seem to starve out the shallow-rooted ones from below. E.g. mosses need no soil and not much light, but will quickly be overgrown if these are provided.
Other factors keeping plants humble are trampling & grazing, fire frequency and dry season. Grasses survive all these stresses well.

Competition between species explains the more diverse vegetation.
E.g. Lichens overrun each other, and leave the rock surface, to become 'leafy' or fibrous or pendulous.
Lush growth betrays the most comfortable living conditions. E.g. rainforest; the underside of a rock in the desert; sheltered microclimates — crowded with species. This is opposite to pioneer species.
E.g. Where excess sunlight is the 'living stress', lush growth will show up on the shady side.

Draw the Skyline, the Ridge-lines, the Outlines
Label it with the 8 main directions

Talk the panorama into words. "Nor-Norwest Hillock, in front of a larger hill on the left, runs down into a gully behind NE– Plateau.
East Mountain is behind NE Plateau and drops away behind the vegetation south of us.
Distant Peak pokes up behind in the SW, and the closer skyline reappears in the West,
falling steadily to the NW Dip before the larger hill."

That took one minute. You do have time to look, notice, talk it through, and even draw it.
Marine navigators know that they have to.
Don't just imagine doing it, or do it silently, or do it half-heartedly — actually do it.
Unlike an avigator who uses a layout map to recognise roads, rivers, ranges, coastlines, etc, from the air,
a walker needs a nose-eye view…
MAPS DON'T DRAW YOU A NOSE-EYE VIEW — SO YOU HAVE TO!

Examine the Skyline BEFORE You Walk up to It

That might mean looking behind you, six hours ago, or to left or right, right now,
because the skyline ahead of you rapidly disappears behind foothills, just when you want a clear view of it!

A clear view of the skyline is uncommon, so you need to take advantage of any skyline view while you have it
— and that is often anywhere else except where you want to look — ahead.

Circuit walks and return trips mean that, if you haven't habitually looked at the surroundings abreast of you & behind you,
you won't be able to recognise what you are walking up to later.

One way to recognise the skyline later, is to use it
e.g. use it to keep track of bends; to keep track of progress; to disclose hidden valleys (by parallax or haze).

Another way is to name what you use e.g. "the hill behind the dam wall from camp 2—Dam Hill"

Mnemonic: "Down in the Mouth"

"Downstream" is "Down to the Coast". "Which coast? Which river mouth?"
Until you have answered that, you are disoriented.

You pass a creek-bed or a flowing stream: "Which way is it flowing?" is only half a question. "Where is it going?" is the other half.
You turn and head 'downhill': You think 'Down' — then say: 'Down in the Mouth — Which mouth?'

If you also ask "Upstream? To which ridge?" there are many possible answers because upstream is divergent.
Nonetheless it is good to know whether a river reaches to the very edge of the valley — whether it is a 'major' tributary.

Which way is the flow?

In dry, flat creek-beds, examine the tilt of the rocks… If any rock ever tilted up into the flow, it would be flipped over.
Examine the silt build up — on the sheltered downstream side of any obstacle.
Look for flood debris — matted against the upstream side of branches.
Consult your artificial horizon, if you have one.

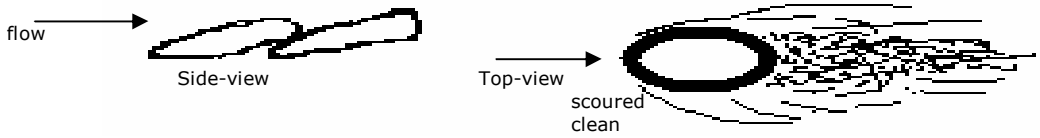

(This is much like grass tussocks in the wind.)

Hint

Don't think "this side" & "the other side". It only makes sense for the one main lowest riverbed.
Think: Up-catchment/down-catchment/cross-catchment, overall, & Port half/Starboard half of the whole catchment in question.
(as seen from the coast, looking upstream)
Think plural: "the other sides"/ "another side" in the network…

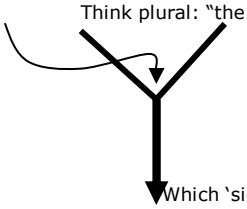

Which 'side' of the river are you now? "I have crossed to another part of the drainage system"
Another possible answer: Examine the map for the contour heights of each channel for the lowest one.
"We are now in the starboard half of the catchment"

The Difference between 'Direction' and 'Track'

'North' is a general orientation reference **direction**.	But	'North from here to there' is a specific route — a 'bearing', 'line' or **'track'**.
'Stick to a direction'/a 'heading' means e.g.'toward sunset'		'This road points toward that Gap' will never change — it is a fixed 'track'.
It can start from any point,	cf.	'Stay 5 miles off the lighthouse' is a (curved) line of position — a track.
and sidesteps around obstacles, without bothering to return to track		'From this stump, past that tree, to the rock' — sticks to a fixed straight line, despite detours.

So the final practical 'interpretation' of, say 270°T.,
is not 'West', not 'sunset', nor 'Pot-set', not 'THERE!', not a drawing in your notebook, Not 'How hot it must be in Baghdad!'
but is a navigational track... "From this place to… that particular tall tree" "Then we will get a line on another point directly West"

I.e. Interpret a 'compass course' , finally, into accurate, achievable 'legs'—'on track'.

Continuous bends

Riverside tracks are especially sinuous, yet slowly so, so that they are deceptive.
Beaches near headlands, and lake edges, are especially insistent in their constant bending.
Contours, constant horizontal angles, and constant distance-off, are all continuously curved. Aqueducts follow contours.
Consequently we are necessarily **continually disoriented**, in following them.
No sooner do we get our bearings straight, than we lose them.
Logically then, you need to orient, not by a direction e.g. the Sun, but if at all possible, to **orient by a 'track'**
i.e. a particular path on the ground — it doesn't have to coincide with your own track, but it often will
e.g. from headland to headland. (The beach will then lie wholly on one side of this shortest distance 'track'.)
A visible, noticeable, landscape track is best e.g. the biggest river valley around.
You now can feel continuously oriented to something fixed, something which you can see & 'feel in your bones'.
As a second best choice, choose a fixed compass course e.g. a rhumb line, and watch where it intersects the landscape.

Follow the bends by anticipating the trends

Before you drive off, or plunge into the twists and turns of endless undulations, examine the map for your average course
— to give you some overall reference track ahead of you. It may be mental, or visual, or by compass.
You can see when you turn to the left or right of it, direction-wise. You can feel it when you drift too far to one side, track-wise.
Interpret each leg meaningfully. E.g. "toward the morning Sun"
(the time of day doesn't matter).
E.g. "Keep Mt Wellington on my port quarter".
E.g. "From the pine trees to the river mouth".
Divide the journey into recognisable sections, if it helps…
Make this a first priority, before studying any details.
Preferably, mark them lightly on the map.

"Mosses Mop up Moisture"

◦ "Mosses grow best on the shaded side of a tree" or a rock,
◦ where there is little soil to hold the moisture,
◦ and it doesn't dry out
◦ where there is not enough light for competitors
◦ and often there is too much moisture for other plants.

Mosses actually monopolise the moisture

In hot places it is the late afternoon which is driest, so it is **the sunrise side, in a sunny climate**, which is the moister side.

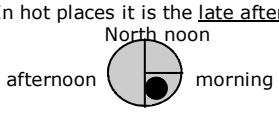

(SH)

Note: Dead moss is black and brittle.

Examine fallen logs — the shady side has a sheltered dark crevice...

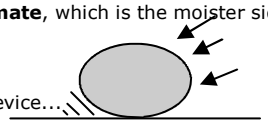

In a hot climate, it may be more accurate to say,
"Mosses grow best on the sheltered side"

Examine ditches — One side might be scorched, the other, lush.
Examine bumps in the ground — One side is greener, with more moss and taller grass.
Examine underneath shrubs — for the moister, mossier, greener side under the bush.
Examine intertidal rocks — for moss-like seaweed growths on the shady, wetter side.

"Lichens Like Light"

Lichens don't mind being dried out, so sometimes they are opposite to the mosses. However, not all lichens are so robust.
Lichens are fungi which live off algae entrapped and protected within ("Fungi Feed on Food"). They may live on mineralised substrates.
They are therefore very colourful. So the best way to learn lichen-likes is to open your eyes to the colours of each direction…

You may see white blotch lichens looking downSun but not upSun.
Yellow blotches may all be only looking uphill, not downhill. Leafy lichens like the shady enclaves, not the Sun-drenched flats.
Crusty greys and fibrous green might cohabit the underside of branches. "White paint" may only grow an the top of any rocks.
Orange granite might be just above high tide level, etc.
Diffuse lichens colonise the crumbly side of a sunbaked rock. Flat-topped rocks are trampled and won't have thick lichens.
Look closely at surface colours of rocks and barks and soil-crusts— they may be due to diffuse lichens.
Look on the branches of shrubby growth — the lichens will grow best on one side or the other.

The idea is to start looking locally, before you need them. It's a nice hobby. And hobbies keep you awake and observant.

"Fern Fronds Face the Front" i.e. the sunny side

Looking down-Sun, you'll see an array of green topside fronds flat on.
Looking up-Sun, you'll see more undersides & yellowness & edge-on fronds. It may not always show, but is clear when it does.
Tree ferns retain their lower branches longer on the sunlit side, and so seem to tilt toward the light.
(But in a windy place you will notice that the permanent lean, curve & frond-fall of ferns are downwind.)

Make a Mental Movie

Memorise a spectacular cliff looming closer overhead.
Watch the shape change. See the foreground shift. Notice the colours and contrasts. Look at the landscape flow.
Describe the silhouette. Peer into the deep shadows. Scan the skyline, looking at each interruption or change in trend.
Appreciate the hazy distance.
Follow a complete outline around, naming its shape.

Don't forget to "record" the sound-track!

Initially try a full production only for beautiful memorable scenes.
With practice you may find that your memories stick for more and more locations.

Take Mental Photographs

Familiarity and Recognition are primarily visual. Your visual impression is adequately vivid, but your recall of it is faulty.
So put handles on the memory and you will be able to revisualise e.g. "where we found the weather balloon".
To give your memory something to recall, take a detailed look.
The tricks are simply to give it a go, and to notice **the specifics**:

The colours, shapes, skyline, peculiarities, sequences — Talk about them.
Play a game: "Your turn: what else can you add to the picture"
Overlay some geometry on what you see — e.g. the log is parallel to the path and as far from it as the log itself is long.

Now close your eyes and see what you can reconstruct. Open them again and notice where you were wrong or were lacking detail.
Repeat this a few times until you could picture it in bed the next night.

Mental photographs are, after all, **your take-away memories**.
They are what names name; what "directions" talk about, what 'recognition' and 'familiarity' is based upon.

Take a Mental Photo of Each Change of Track

Where one subsection ends and another starts
— Look forward onto it, with the old terrain in the foreground and the new terrain joining on to it.
Now walk past the join, Look back, take in another view, with the old terrain now in the background.

This adds visual sequencing information to individual memories.

The Inside of a River Bend is Flatter Ground

The inside of a meander may be wider, sandier, marshier or of different vegetation.
On the outside of a meander, cliffs form, through ongoing undercutting. See also p.56a about the steeper North bank of a river (SH).
Consequently plan ahead and work out when to swap sides to find the easiest walking.

If the track becomes steeper underfoot, that would mean that you could be on the outside of a bend.
Check to see whether your heading has changed with the bending river.
If the track becomes flatter underfoot, this may indicate an inside bend (or a river junction)

Rivers Bend! Don't Forget to Look Backwards

Obstacles cause creeks to change direction. Outcrops of bedrock and creek junctions, for instance, cause bends, and this may be shown up ahead of time, by ridgelines pushing in closer, as you follow a riverbed.
Expect to meet a new creek clinging to the base of any ridge or shoulder which crosses your path, and also expect to change trend.

In addition, to that cause of bending, rivers follow contours and so usually are on a curved course anyway.
Furthermore, in the flats, rivers never follow a straight course but will 'meander'.

So if, while following a creek valley, you notice the slope underfoot changing, or the vegetation altering, watch out that the creek doesn't leave you, by bending away from you!

If you find yourself in the middle of a larger than normal river flat, suspect a river junction, and be very careful interpreting the trends.
E.g. you may meet 'the river' 'coming back toward you' when you were trying to go downstream!

So you can easily lose orientation while the winding river-track follows the lie of the land and gradually wends its way around corners!

Look ahead to catch the trend — and interpret it.
Look back, to see the extent of the curve.

Rivers Meander. Ridges Run Straighter

A ridgetop route will often prove shorter than following a winding river.

Keeping a Compass Course, by Legs

You can't see far ahead, but you can line up a rock or a bush which is in the right compass direction, then walk up to it.
It doesn't matter how you get there, as long as you get there.
Repeat the procedure, even if you can see far ahead — because, if you are merely following a compass needle or a distant landmark, you can "side-slip" off track.
Your eyes are freed from the compass needle while you walk.

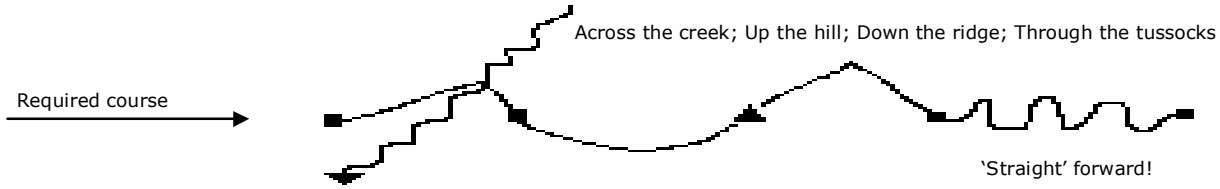

Turn Smooth Bends into Sharp Jumps

Your brain wants something to "latch on to". A smooth bend is just the opposite! So make a series of straight-line "legs".
You "let go" of the last leg, and make a definite jump, to "latch on to" the next leg, and your subconscious smiles, and remembers.

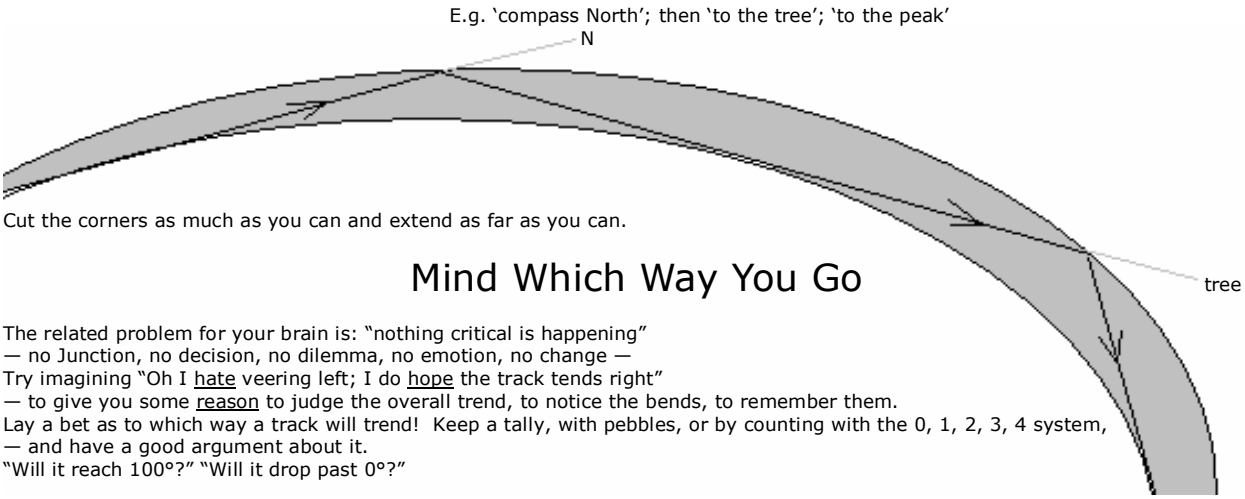

Cut the corners as much as you can and extend as far as you can.

Mind Which Way You Go

The related problem for your brain is: "nothing critical is happening"
— no Junction, no decision, no dilemma, no emotion, no change —
Try imagining "Oh I hate veering left; I do hope the track tends right"
— to give you some reason to judge the overall trend, to notice the bends, to remember them.
Lay a bet as to which way a track will trend! Keep a tally, with pebbles, or by counting with the 0, 1, 2, 3, 4 system,
— and have a good argument about it.
"Will it reach 100°?" "Will it drop past 0°?"

Bush Noon— The Sun is Always out, If You Know How to Look

Trees & shrubs live in the light, day in, day out. You would expect that they should respond to the light, and lean or reach out to the strongest light, and throw out more branches and more leaves to the bright side.
The disposition of branches and leaves stays in view even when the Sun isn't out, and it lasts longer than a cloudy spell.
Although the Sun is a huge factor in explaining foliage asymmetry, be aware of other factors…

- A lake or sea can double the light intensity on one side, pulling the foliage toward it.
- A clearing will let more light in on one side.
- Down slopes allow the trees to see more skylight on the down-side, and less skylight on the up-side.
- Other trees can shield a tree from the Sun. Look especially for recently **fallen** ones on the ground, to explain standing branches
- The tree itself shields its own branches, so that most growth is simply outwards.
- Unrelated factors like gravity, wind, and wood collecting, alter the shape of a tree.

However, if you **average out** all the other influences, you should be left only with the noon-side Sun's influence — "Botanical Noon".
With adequate practice you should never again be "lost in the bush" because it is the thick bush itself which will tell you where North is.
You may have an initial hopeless feeling of "I can't see any meaningful asymmetry!" but the feeling will be temporary (eventually!).

Of course, this works best only where the Sun is always one-sided—in nontropical latitudes.
Hint: If a tree confuses you by pointing in the wrong direction, go over to it to learn why.

False Asymmetry

The slope of the ground
may be forcing the asymmetry,
rather than the differential in the lighting.
In the second diagram, there is asymmetry:
A is ungrassed; B is grassed.

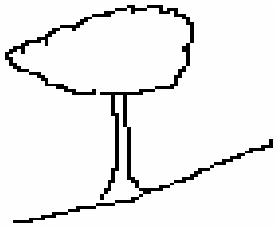

False ("Chalk and Cheese") comparisons abound in Nature. True asymmetry is there too
Your task is to sort it all out.

E.g. Shaded foliage sees mostly blue skylight, which should make it long and lanky, compared to the red-end light from the Sun

False Symmetry

C and D are both grassed.

E.g. the West side shadow is frostier for longer and the East side shade is cooler in the p.m.

True Symmetry

B and C are both grassed.

E.g. the North side soil will be warmer deeper than the South side of a thick hedge

True Asymmetry

A and D are differentiated from each other.

Dirt road wheel ruts...

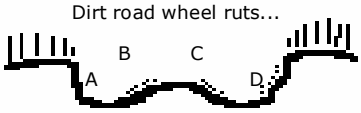

Look on the ground beneath a shrub, or log, for asymmetrical shade and shadow effects.

Remembering the Sequence, Needs More Than Logic

Logic is logical and helpful e.g. The bridge was after lunch because we were worried about the crossing over lunch.
The red rock was after that, because the second creek was dry — it had all those beautiful red lichens in it.
Then the funny tree was where we had chocolate, i.e. 3 o'clock stop.
But logic by itself is not 'sexy' enough to be emotive and memorable.
So deliberately add some emotion…
E.g. "lunch", "worried", "beautiful", "red", "funny", "chocolate"
You may need to add a song and dance, a gimmick, a clue to the time or sequence, distance or linkage.
The trouble with a memory gimmick, is that you focus on the gimmick, not on the memory itself.
E.g. If you use an acrostic, remember to feel the sequence of it, as well.

DO Something NUMERICAL

E.g. Lean one stick against "Walking-Stick Cliff" at one o'clock.
Sing two songs at "Sing-Song Pond" at two o'clock.
Make three cairns at three o'clock at the "T(hr)ee Junction"
Make four arrows on "Four Arrow Slope" during four o'clock.
Do it. Don't just imagine yourself doing it.
E.g. Scratch numbers in the ground, in sequence, as you pass points of interest.
e.g. "5FF" (the 5th stop, with the flat floor); "6SS" (six sizzling sausages—the fireplace).

Make a SHAPE; Match the SHAPE

Every or location in a forest track or in open country can be recognised from its **peculiar shapes**. Every bend or fork can be **matched to** the forest, or the shrubs and stones, so both the bend or clearing and the surroundings can be remembered, with joy.
Locations: What does this fallen log **point to**? What is **parallel to** it. *Place* something so it *does* **make a shape**.
How many shrubs do these three trunks enclose? What **face** or **picture** can you make out of the bushes, rock, clearing and the creek?
Do you see that **line** of tussocks; that **clump** of trees half-way up the slope; those parallel gutters; that oval of sand-dunes? Where does it drain? What **direction** between those **paired** stags?
Bends: ~Look for a tree trunk **leaning** over at the same angle as the bend of the track, preferably leaning over the track or bend and in the right direction. The tree stands out from all the others because of that. (Any lean at any angle in any direction will suffice.)
~Look for a branch **pointing** in the new direction, or curling, or **bending** in a way which mimics the track-bend.
~Look for a log **lying parallel** to the new direction, or any ditch, or line of rocks, or outcrop. (It doesn't have to match.)
~Look for a large rock nearby which **fits the shape** of the bend, or orient it until it does. (Matching shapes are always the best)
~Look especially at the tree trunk, or the shrub which blocks the straight line of sight ahead, or the shape of the fallen-over trunk in the way. E.g. "I'll fall into that black hole in the trunk if I keep going straight ahead! I'll have to swerve". "I need to go around that round stump". A tree **right at the bend** will be visible from both directions.
~Look for a forked branch which **matches the angle of a fork** in the track. **A series of parallel branches** at the fenceline. A couple of trunks **leaning towards each other** over the gateway. A tri-parallel set for a creek with two banks.
~Look for a branch which **points downstream**. Watch it for the next 50 years as it grows. A tree recognised is a friend for life.
~At a noticeable uphill pitch, find a leaning tree trunk and imagine climbing up it.
~Some leaning trunks can be interpreted in **reverse**… from top to bottom, to indicate the new direction.
The more you latch on to a peculiar shaped feature and the more you study its peculiarities, the better your recognition will stick.
Try to **double up** on the first peculiarity to clinch the recognisability. A **matching pair** of branches. A second confirmatory log.
You will have to **look backwards** at each bend if you want it to work well for the return trip.
Every bend and feature is a good chance to recognise another couple of trees. You stop at the bend and look around for your "friends".

Land-Form

Work downwards from a 'Big Picture' understanding. Start with a small-scale geology map, showing overall elevations.
Continental-scale domes (anticlines) & depressions (synclines) explain the overall drainage and watersheds.
Fault lines — shown by escarpments & cliffs — and individual ranges, interrupt the overall scheme.
The geological layers may be visible connecting across large distances.
E.g.

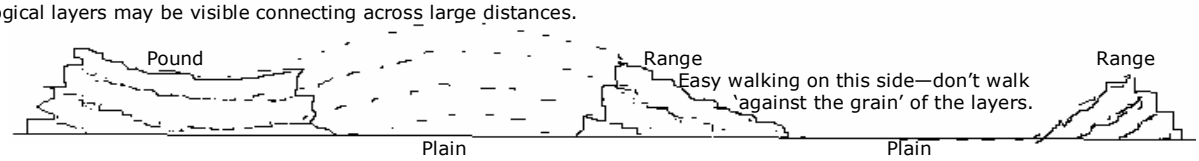

Layers become hardened and resistant to erosion when they are bent concavely upwards. They are compressed rather than fissured.
Convex layers are fractured and easily eroded, because they have been stretched. They will have been eroded away.
The present landform is a result of the erosion history on the previous landforms.

Hint

Most peaks have a steep side and a gentle side. Don't forget to make a mental note of that trend
in relation to the rest of the land and the other visible peaks
— you may notice a pattern, or simply keep your overall bearings better, or recognise the peak later, from afar.

What Makes a Cliff and an Overhang?

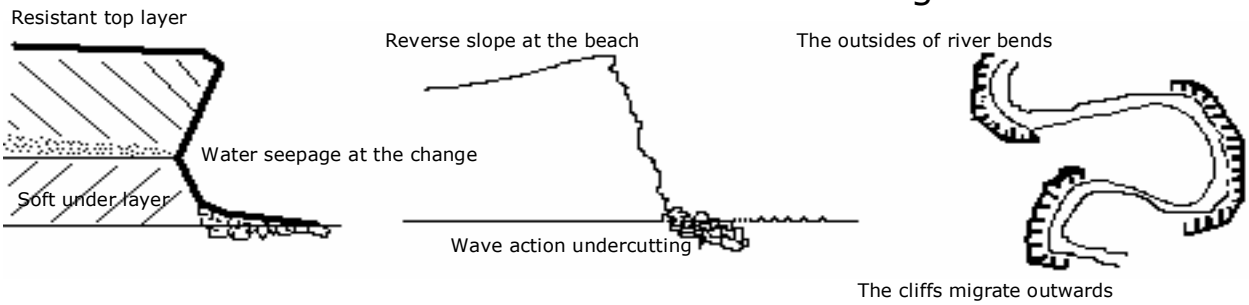

Waterfalls form at hard layers, at fault lines and at escarpments.
Look for outcrops striking across the land.

"End the Bend" i.e. Notice the Un-Bends

A bend is not simply a sharp corner, not just a change in direction: 'end of story!'.
It has a start, a curve, and **a finish, where it straightens out again**.
The curve needs to be **named** as a whole, **measured** as a whole, and thought of as **a whole bend**.
Look for the dividing points between successive curves

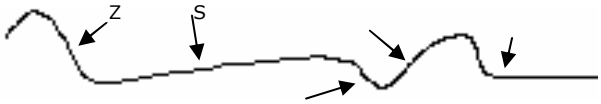

The arrows show where a curve to the right changes into one to the left,
(a Z bend), & vice-versa (an S bend). These are "points of inflection"

— a straightening out, and remaining straight, is the other possibility.
Incorporate half of the straight section into each adjoining 'bend'.
At the end of each bend, reorient yourself — look behind, at the compass, at the Sun and your shadow. Point to 'home'.

Name the Bend, By the Trend, of its End

Trend:	Left, toward Mt Michael	e.g. "Viewmont Bend"
	Right away from camp	e.g. "The Long Right Leg"
	Left into Lake Jordan	e.g. "Left it down the Drain"
Handedness:	Hard left around the spur	e.g. "The West Bank Left Bank"
	Half left from the creekbed	e.g. "The Left Contour"
	"Clockwise Climb"	
	"Right around the Bend"	Bends will often alternate.
Direction:	"Ending with Sunset in our face — Sunset Slap"	
	"Turning until the Full Moon Rises on our Right — Werewolf Way"	
Compass:	"North Now, to Alaska"	
	"Left-Sou-West"	

Add some description
to seal the uniqueness e.g... Skyline: "Lone Pine Ridge" "Three Tree Hill"
Detail: "Uphill; On the Left Bank; Just past the river crossing; at 10 a.m."

Notice and Name Perverse Bends

Unexplained bends, or ones which buck the trend, deserve a name.
Most bends are predictable — around obstacles — at creeks — around shoulders & outcrops — shortcutting
— contouring around sloping ground — at Y-junctions — joining with disused overgrown tracks
E.g. Seaside roads will hug the coast, go diagonally downhill but upstream (into a valley) and climb diagonally out of it, downstream
— as a compromise between shortness and flatness.
E.g. For the same reason, tracks will often change direction at a creek crossing.

So name the curves that go against the flow…
E.g. "Sinister Switch" (Latin for left) "A Dextrous Diversion" (Latin for right) "Unorthodox Twist" (Greek: not "right" = left)
"A Wandering a Way" "Aimless Avenue" "It shouldn't've gone left!!" "The I'm-going-on-strike Uphill Bend"
"Mind Bender" "Morally Bent"

Higher Density of Leafiness on the Noon Side

From underneath, next to the trunk, you can look up in all directions and see which side is leafier. Looking side-on you may see one solidly green side and the opposite side patchier in its cover. Trying to look through the foliage to the back of a shrub, you will have most difficulty on the North.

The Lowest Leafy Branches are on the Noon-Side

Looking from the sunny side you see a lower branch or even a wall of greenery, right to the ground. Low branches on the shady side are often short, dead or only thinly covered with foliage. Looking toward noon, you see more trunks, dead wood & grey-brown sticks.

"It's the Shady Side that Shrubbery Shows up"

In practice you don't look for the noon side, but for the absence of leaves low on the pole side. Looking toward noon, you see ill-developed shapes. Many plants will throw out branches and leaves in profusion in all directions, but you seek out the one direction which is different — the one less densely green — where the light is in short supply. It is **the 'light null' for solar direction finding**. Look for where no vegetation leans: "No stems, plants, branches point in that direction".

Tall, Steep and Ragged Backside

Shade-side branches often reach for the sky, and are more vertical on that side, making it steep. The noon side will be more of a smooth continuous wall of greenery from low-down upwards.

Wedge-Shaped

Looking away from noon, you see smooth and continuous foliage, slanted down toward you, often clinging right to the ground, like a wedge.
Looking toward noon you see flat tall backsides of the bushes.

Hollow under the Shady Side

You can stand closer to the trunk, or at least see more of the trunk more clearly from one side. Down low on the shady side is where the light matters most. There may be space enough for a branch but not light enough. On the sunny side you will see leafiness deep into the low recesses.

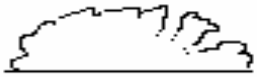

Imperfect Shape on the Shady Side

Many shrubs seem to adopt a perfect dome shape, until you look more closely.

Gaps on the Shady Half.

The backside of shrubs is often tall, untidy, and lacking leafy branches, close to the ground especially.

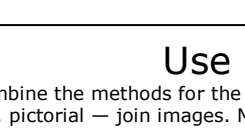

Dead Branches

The overall shape may be perfect, but one side is alive and green, while the other side lacks leaves. **"Dead" South**. There is a scale-effect to dead-south: The larger bushes have more depth of foliage. The more foliage up-sun, the darker down-sun. Only the larger or thicker bushes work best.

Use a Type of Memory which you are Good at

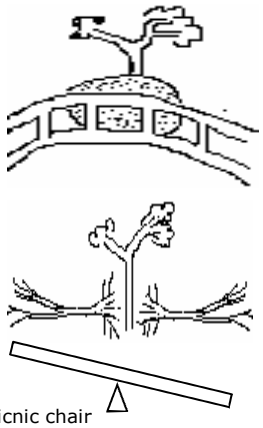

Combine the methods for the most memorable effect.
E.g. pictorial — join images. Make them touch or glue onto, squash or hold hands, follow logically or overlap — use a stack of pictures to put incidents into sequence.
"After the bridge we came to a red rock and then sat down for a snack at the funny tree"
E.g. ditties — A Capital Tee of Tall Trees at Three O'clock.
E.g. invented "associations"— a vivid or absurd image like a Tasmanian Tiger eating that hill.
Paint it in partly related colours, with shapes and other senses involved, to give extra "handles"/clues.
E.g. Imagine successive scenes in an invented journey/building/furnished room/street scene/pattern points. Emphasise movement, and links between stages. Make it visual, simple, exaggerated and appropriate. e.g. "I went inside the cave and talked to the sleepy bear (Ranger) about the bearing to follow, and he said "Seek the morning sun", so coming back out I turned to the east until I met the keeper of the bridge who said "over my dead body lies the path to right; and left". So climbing up over him...
E.g. numbers — One Wooden Bridge, Two Red Rocks, Three Funny Tree… to reflect the hours of the day
Make up a story with rhyme, from e.g. 1, bun for lunch, 2, through the dew, 3, see the sea… 7 rhymes with leaden/Heaven/redden, 11 with Aladdin/pelican/caravan/a raven, 12 with elves/shelve/delve/selves
E.g. shapes — Make a 10-rock pyramid at the 10 o'clock stop; an 11 stick teepee at 11; a 12 leaf hexagon — e.g. see-saw on a log at 7 and draw a see-saw in your notes—it has 7 corners.
E.g. You can weave a story, of invented images to go with the times of day.
e.g. each digit looks like a... 0 hole/ball/eye, 1 stick/pen, 2 cobra/sphinx/swan, 3 bosoms/camel, 4 yacht/picnic chair 5 measuring scoop/wheelchair, 6 bomb with a fuse/yo-yo, 7 arrow/step, 8 shotgun/glasses, 9 balloon/comma, 10 ball bouncing off a wall/golfball off its tee, 11 jail/stump/goalpost, 12 jail-break/caged lion.
E.g. acrostic — 7 a.m. Z. for Zero Rule Zero: Don't forget to pack the lunch!....7:15 A Bend A in the Creek
7:30 B Bridge across the River at Bend B. 7:45 C We didn't Cee the third bend on the map!
Using this method, elapsed hours happen at Z,D,H,L,P,T,X i.e. every four letters.

Use the Name of the Track for the Acrostic

L ying down tree
E ee! There's a bull!
E e-normous bracket fungus
S wim

P addled Across
A lternative Detour
D
D
O
C
K
S

*Every quarter of an hour, look for things beginning with that letter to remember that kilometre by.
*four letters make up one hour *Use stepped word pairs, 'Lying East-west; Exciting Exit; Extra Searches
*Do Something Alphabetical
E.g. Lie down at L; Scream at E; Examine the fungus at E; SPlash at SP
*Make all your names begin with that letter e.g. Pot-holes, Pines, Possum trees, Possible campsite, Photo
*Write them all down, but try to memorise the best one.
*You can attempt to improve the sequencing, by taking the letters two or more at a time e.g….
LE In the Lee of the lying down tree
EE EEE!
ES E.S.P. told us to find the track next to the E-normous bracket fungus
SP Splash Splash. Had a Swim
*Try also to relate picture to picture in sequence e.g….
Look at the Long horns on the Lying down tree; Enormous bull!; Enormous bracket fungus, dripping with water… Swim!

Acrostics combine individual memories with a clear sequence — firstly by letter, and secondly, by overlapping one pair of letters with the next, and thirdly by overlapping one picture with the next.
You have already memorised many long sequences of letters, e.g. the words of a song, which could be used as a unique acrostic.

Headlands, Bays, Prevailing Winds, Longshore Drift of Sand, Creeks

Headlands are ridges. Bays are valleys.
The ridges determine the headlands, bays, creeks, and — look for this — the position of offshore rocks and islands, especially where a 'point' marks a change in the trend of the coastline.

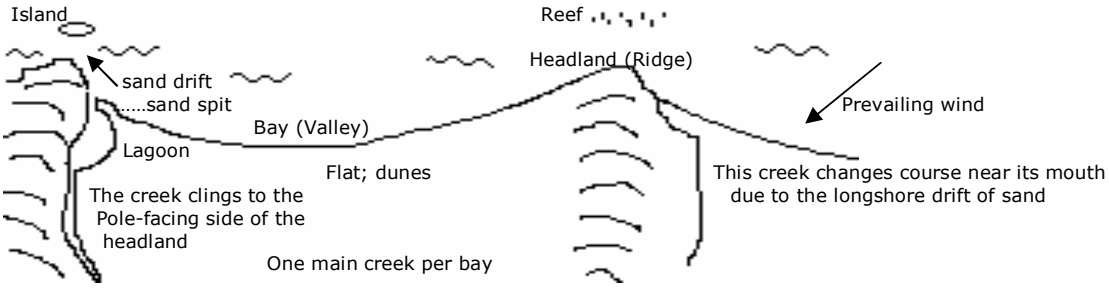

Beaches — Slope, Particle Size, and Wave Energy

Energetic erosion washes away any small particles.
A steep beach is a high energy beach with big particles.
When it is deep offshore the waves break with high energy.

A flat beach reduces the energy of breaking waves.
A low energy beach is flat with small particles.
Sandbars and shallow sandy offshore seabeds make the waves break over a long-distance.

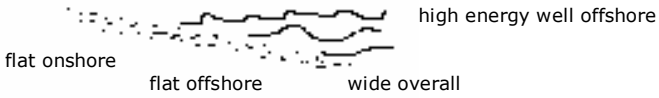

Much the same seems to happen on land slopes — the flatter slopes have fewer rocks. I.e. the accumulated silt which is not washed away by surface water, covers up the larger rocks.

Sea-levels have been much higher in the past, so that there will almost always be buried boulders just above the present seashore.

Look Through the Trees, to the Sky and Skyline

This is a habit which stops you being befuddled by bends. You focus on the most distant thing you can see…

- E.g. the farthest tree trunk
- E.g. the skyline — ridges, gullies, peaks, light/dark division
- E.g. the clouds — but do stop to examine their movements occasionally!
- E.g. the Sun, the Moon — obviously the best, but they move too, and… don't look at the Sun!
- E.g. the dark blue polarisation band — Use Polaroid sunglasses if you are walking a forest track and can see only the sky.

Don't forget what is abreast of you e.g. a particularly tall tree. If you need to, look behind for some reference point.

Habits take practice to develop. This one is easy, because once you appreciate it, you won't want to give it up.
— It turns many small bends into a few major trends —

Periscope

Climb a tree. Climb a hill. Stand on somebody's shoulders. Or try this…
Attach a mirror (bigger = better) at 45° (try to be exact) to one end of a long stiff stick (e.g. a fishing rod)
— then scan the tree top horizon for landmarks.

Navigating by Trees

Tall trees on high ridges stand out against the sky and are visible for miles around.
Get to know the biggest, tallest, most visible ones.
Individual trees, gum trees especially, are wonderfully fingerprinted, by their branching pattern, outline, position & orientation.
E.g. look for the direction between the two largest branches. This stands out like a weathercock whenever you glimpse that tree again.
E.g. locate the relative positions of a handful of the best trees, so as to cover a wide area.
E.g. try to note the relative direction of pairs of trees.
Most big trees will have large branches to noon. So look for big branches which are *not* pointed to noon.

If you keep looking through the immediate green barrier for more distant (unmoving) skyline trees — you will notice them and recognise them surprisingly often.

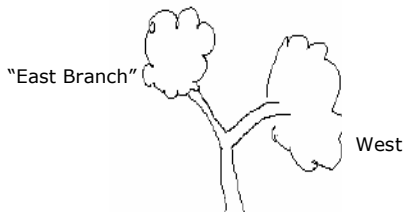

'Solar-Panel' Tilting of the Canopy Top

Some trees may be 'flat-topped', but they are not 'level'. They tilt toward noon. Best seen from side on.

Wedge Shaped

Branches seem to know when to stop growing, and so manage to display
a fairly **uniform**, slanted **flat wall of leaves** toward the outside and especially toward the light.

Lopsided Lean

Branches on the backside usually reach higher to bask in the skylight.
Branches on the front side are more horizontal, lower and longer.
Some trees don't seem to try to compensate and simply leave their trunk behind, on one side.

Half-Dome

— an extreme case of a steep back side

NOTE that all these signs work best in open woodland, not in forest understorey, but not always...

One-Sided Leafiness

E.g. Banksia trees
Such extreme asymmetry often only occurs in the understorey — where light is limited.
E.g. After a bushfire a gum tree revegetates from the trunk — most leafiness survives on the sunny side.

Individual Branchlets

— show similar shapes to whole trees
— tilted, asymmetrical twig patterns, one-sided leafiness, etc
The effect will show up best on <u>the east and west sides of a plant, where the lighting is most asymmetrical</u>.

Bare Backbone

Look for it on individual branchlets, on the rear-side of main branches, and sometimes on whole trees.
You could run your finger up the entire length on one side of many shady-side branches.

Group Several Related Images into One

This is a standard method of making your memory ability efficient.
You have <u>less items to remember</u>—they become a few <u>memory prompts</u>, each to a <u>handful</u> of items.
E.g. 4 groups of 4… 6,7,8,9 a.m. / 10,11,12,1 / 2,3,4,5 / 6,7,8,9 p.m.
Write each group out <u>visually shaped</u> e.g. at the points of a square for a group of four
with 2 opposite squares for a group of eight
inside an infinity sign for a pair of items
List the hardest to remember items, <u>first and last</u>.
Group the shapes i.e. the groups, into another <u>visual structure</u>.
Use symmetry wherever possible—it makes it easy to recall.

The Start-Stop-Rest mnemonic (see Section 5. p.110) is a good example of this "clumping" technique—never more than a handful of things to remember, yet it leads you into hundreds of points.
Put the S for Sky at the top, and use a hexagon for the six start up points. Use a clock-face for TIME, and the compass-rose for NSEW.
Balance it all on a triangle, for the 6 a.m. start

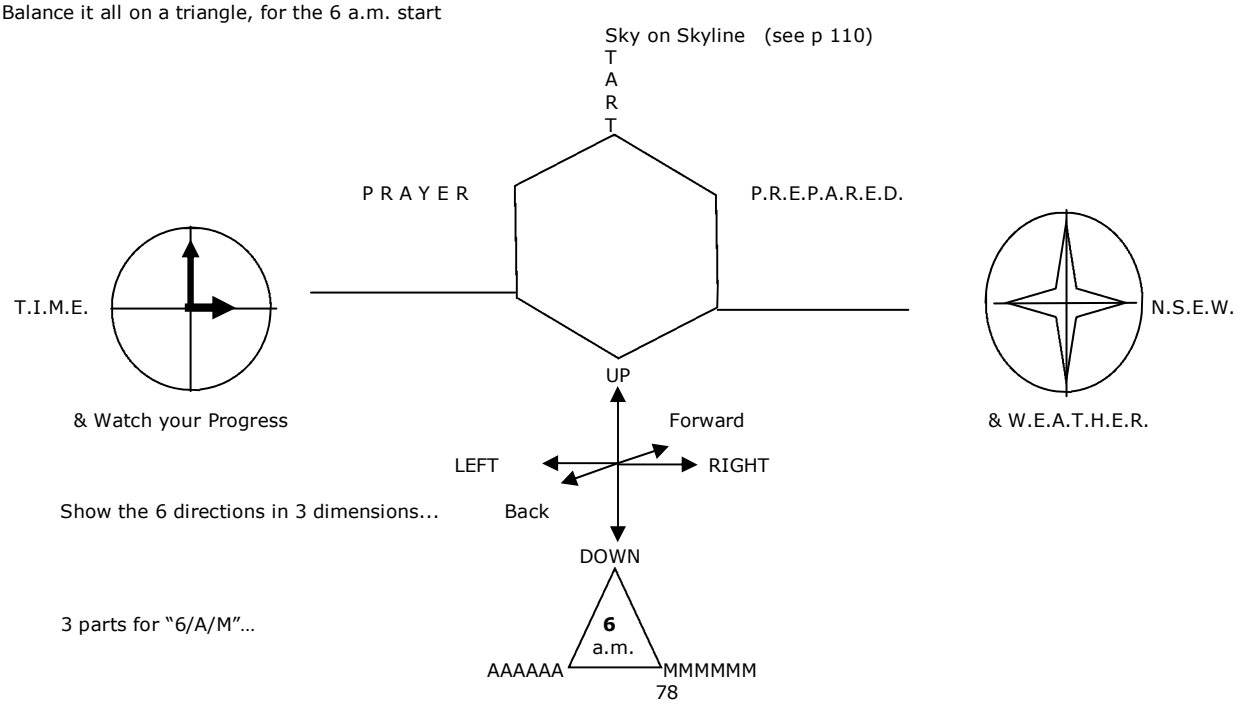

Rock Typing

Hardness — how easily eroded. Keep trying to scratch it until you find something just harder than it e.g. your pocket knife.
Colour — what colour silt. Scratch it against harder rock and look at the colour of the crumbs.
Sedimentary — the components are not of local origin. Layers — sedimentary origins. Melted sediments — crystalline bands
Sand content — grano-diorite — pale & silica rich. Clay content — basalt — dark & mineral rich. Mineral crystals in it.
Crystallisation — small crystals mean rapid (surface) cooling; large crystals means slow (deep earth—'plutonic') cooling.
Mineral content — what sort of soils does it produce? Flakiness. Behaviour in fire (Caution!)— it cracks, blisters or explodes! Melting pt.
Density, Magnetism, Fracture Pattern, Transparency, Permeability, Acid Solubility—evaporites and marine deposits may be limey.
Understanding these details is a key memory aid, since it gives you more meaning to look for, to memorise and recognise.
E.g. in basalt areas expect a more diverse & rich flora & fauna.
E.g. in the sandy plains, expect rivers to run underground, trees to have deep tap-roots and to line only the dry watercourses.
E.g. in dry sedimentary areas, look for a harder 'evaporite' layer underneath.
You don't have to identify the mineral, but you should characterise the rock type.

Soil Typing

1. Climate drives soil typing. What is the climate?
2. Latitude drives climate. What is the latitude?
3. High altitude mimics high latitude. What is the altitude?
4. How is the local climate modified? Look especially for temperature and rainfall factors.
   E.g. ranges upwind? A cold landmass? Warm ocean currents up-weather?
5. Evaporation produces salty limey evaporite deposits, whenever the evaporation rate exceeds the rainfall rate. High-temperature and wind, speed up evaporation. Flat inland areas receive less rain. Where rainfall exceeds evaporation, the soils are acidic, leached, and deep. Does the evaporation rate exceed the rainfall rate annually?
6. Climate drives vegetation as well as soil types. Soil types drive vegetation. The three go together. What is the vegetation type?
7. What is the parent bedrock? Granodiorite produces pale, sandy, friable, well-drained, mineral poor soils.
   Basalt produces dark, clay, dense, waterlogged, mineral rich soils.
   Warmth & rainfall aid chemical weathering and erosion rates of the parent bedrock, and so produce deep soils.
8. What are the relative constituents? Shake some soil and water together in a glass jar and leave it to stand until it is layered. Organics & charcoal float. Heavy minerals, pebbles & sand sink. Silts lie on top of them; then clays. Colloids remain suspended. Organic remnants accumulate in soil when it is warm enough to grow plants, but cold enough not to rot them away quickly. The decay of organic materials then provides acids for leaching, provided that high rainfall doesn't dilute the organic acids. Can you see from the sort of soils how cold and wet it has been?
9. What are the depths & layers of soil? It is leaching which transports poorly soluble salts to lower layers, until the water table dilutes the acidity. Subsoil clays also accumulate where leachates are deposited. Evaporites accumulate just below the surface and above the water table. Dry areas have poor soil depth due to slow bed-rock weathering and erosion rates.
10. What is the soil colour? Red indicates partly leached soil — iron and aluminium salts remain. Paler, ash-coloured soils are left behind after further acid leaching. Black soils result from poorly drained unleached soils. The importance is that leaching reduces fertility due to the lack of soluble plant foods. Leaching is aided by warmth (except in the case of limestone which dissolves better in colder climates).

Bush networks

The bush roadmap is made up of a network of interleaved ridges and rivers.
Instead of a major road or a minor road, you have a main ridge or river, shorter ones, and dead-end ones e.g. a 'shoulder'.
A 'cross roads' is where two creeks & two ridges meet — a saddle.
The seashore corresponds to a boundary fence; Peaks and islands act like townships — they are a focus of 'roads' and ring-routes.

Don't Navigate by Networks Only

Networks are fractal — big **creeks** have little creeks running into them; **walking tracks** have side tracks branching off from them; big ridge lines have little **ridge lines** running off from them;
dirt roads have smaller dirt roads which have still smaller dirt roads joining in.
That is why it is difficult to recognise the scale of any map just by looking at the networks on it. The more detailed the map, the more detail there is to show up. The smaller scale the map, the more detail has to be left off.

So there will always be small creeks, unmarked roads, minor tracks and minor ridges to tempt you into thinking
"This one is what is marked on the map".
You will likely find one "exactly" where you look for it, but it may not be the right one.
Don't navigate simply by expectations of network junctions.
E.g. the next creek-line has a 50% chance of being "just as I expected it — a creek-line coming in from the right"
or if not expected, the 50% chance is that you will say "probably just not marked on the map"!
If by chance it is actually correct, it will only encourage you to continue navigating by network,
but you won't always pick the correct one.

Antidotes For Guessing at Network Junctions

- **Distance** covered. Accurate measurements of progress can eliminate some minor junctions which are "almost" where you expect the desired junction to be.
- Trend — of the junction branches. I.e. the **directions** must match your expectations too. E.g. "Each side of North"
- Landmark **bearings**. Try to separate identical looking junctions by cross-checking their position in relation to skyline features.
- Care.
- **Count**. Look-alike branches are nonetheless unique, when numbered.
- Recognition, from last time, if you had bothered to identify its peculiarities. I.e. bother now, for next time.

Branching Patterns — Examine Them for thickness, length, angle, curvature

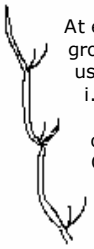

At each node some branches grow better than others usually toward the light i.e. stretching outwards or up-wind upwards or downwind down low. Check the **spacing.**

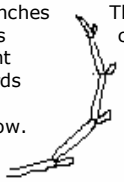

The choice of branchlet can **curve** a whole branch or stiffen it against the wind…

bottom view looking up…

Side branches can **curl** around toward noon. Examine the ends of the branches.

This enhances the 'bare backbone' appearance of the dark side of a trunk or branch.

A well-leafed branchlet will usually hormonally dominate one without buds, by active biochemical suppression. In this way growth can be regulated.

E.g. horizontal decurvature can search out extra light underneath the canopy

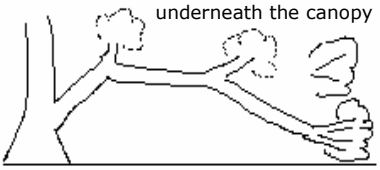

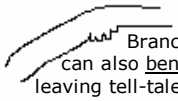

Branches can also bend, leaving tell-tale 'wrinkles' in the skin.

Double curvature can fight gravity or wind.

Gum trees show asymmetry especially easily, since they do not try hard to be symmetrical in the first place!

Examine the main trunk itself. It may curve away from the light to act as a counter**balance** to the heavy light-side branches.

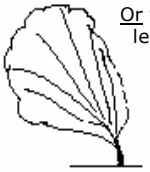

Or the trunk may be overbalanced, by its branches lengthening, thickening and curling toward the light.

Length:
Some trees 'reach' for the light becoming lopsided

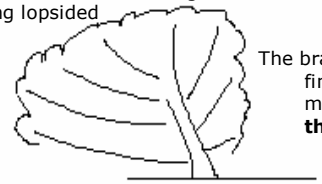

The branches that find the most light **thicken** most…

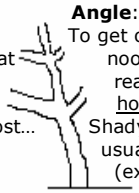

Angle:
To get out from under the canopy, noon-side branches often reach out a long way horizontally...
Shady side branches usually reach upwards (except maybe in a windy place).

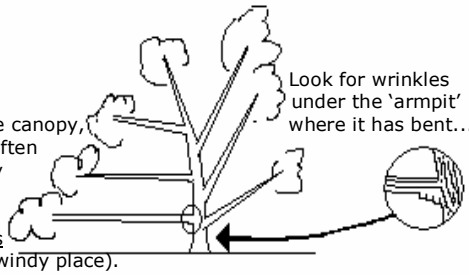

Look for wrinkles under the 'armpit' where it has bent...

Working Names — Invent Them

E.g. Past the 'Little Castle Rock'
Up the 'Yellow-Crowned Tree' trail
Under the U-shaped rock face
Between the Big-Boulder Bluffs

"40 minutes in" is a handy name — when you recognise it, coming back out.

Stupid gawky names help you remember, but even better is to do something stupid to match.
E.g. Gallop like a galah, through the pass, and call it "Gallop Gap with Galahs Galore!"

Standard Name Plus Personal Name

Actual names last longer and are widely known. Use them.
But make up a personalised name too.

E.g. "Oh it is actually called 'Boomer Hill', but we call it Porridge Hill, because it lies halfway between Oatlands and the Table Tier!
E.g. We call Wollongong "Kaus Australis" 'coz that's where we watched Kaus Australis go straight overhead.
E.g. Kaus Australis becomes "The Wollongong Star"

Remembering Names

To help you later, when the name is 'on the tip of your tongue',
Count the syllables, consonants, short vowels, long vowels, now.
Rhyme it with something appropriate e.g. Condominium Creek has a pond of minimum seep.
Look at the spelt visual shape… for ascenders, descenders, double letters, symmetry.

Then later run through the alphabet for a guess at the first letter, then at the second, and so on.
Think about anything related, like why you went, what you ate, who you met, what you saw, the weather, the season, the news at the time…
Ask each person for these recollections. Recreate the full scene.

Contour Lines

Contour lines mark horizontal ground.
A simple trick to match up the map contour lines and the landscape, is to **point horizontally** parallel to the ground.
E.g. if you are in a valley, point along both creekbanks, using both hands. The steeper the valley sides, the narrower the contour angle.

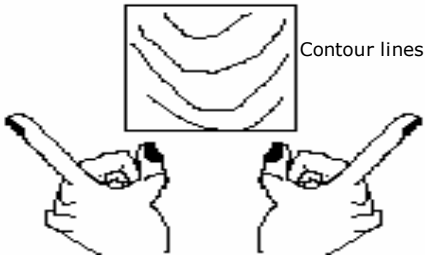

Contour lines are like water-marks on the bank of a reservoir, lake or swimming hole, visible after the water-level has dropped.
But bigger like a global flood drying up 10m at a time, leaving water-marks.
like a photograph of it, from above.
like horizontal slices progressively taken off from the top of the mountains, as when you slice clay with a thin wire and you look at what is left.
like building up a model landscape one layer of cardboard at a time, from below, with progressively smaller shapes cut out according to the map contours.
Then you draw the results on a map.
They show the path you have to walk to go 'around the contour' so that you neither climb nor descend.

Contour Navigation: **Practise** It, **Before** you Need It

At night, or in fog, or in a blizzard, you may have no other clues, but the contours, on the map, and underfoot. Try to match them up.
Hint: Don't expect every little bump to show up on the map, but only the average smooth slope.
Hint: Keep good track of compass direction, time & distance as your best crosscheck. Assess the wind. Use an altimeter (,GPS, radar…).
Hint: **Delegate** tasks — map reading; course holding; distance logging; slope assessment; exploration; memory…
Hint: To go quicker, go slower — so as to be sure. Explore the slopes to left and to right. (But don't get lost or disoriented!)
Hold regular conferences to check your judgments.
Hint: Assess slopes by looking sideways, across them. They always seem so steep when you look up!

Boat navigators in a fog, can sail toward land until they reach a predetermined depth, then sail along parallel to the coast, using the depth contours marked on their chart, and some method of finding the depth. Marine depth sounders rely on a good echo, but you can always toss a weighted line over the side to find the actual depth. Make sure to make corrections for the depth of the depth-sounder, the depth of your keel, and the height of the current tide!

Sticking to the Bush Network

- Walking **up-ridge** may perch you on an isolated hillock.
- Walking **down-stream** must get you to salt water; or at least to water, civilisation, flats, roads; but also to thicker vegetation and a loss of view; and sometimes to impassable waterfalls and gorges.
- Walking upridge, uphill, upstream lands you in steeper, drier, more jagged terrain with more open vegetation, a better view and maybe an along-ridge track.
- Walking down-ridge will "take you to the lake", but simply walking downhill may lead you into a heavily vegetated creek-valley which then drains into the lake.

Natural Routes Naturally Converge

That means that, in reverse, they must also **diverge** — which is dangerous, open to mis-navigation.

- Provided you don't cross catchment divides, walking downstream or up-ridge is **convergent** and so is safe for a one-way trip, but **un-safe as a means of retracing your steps back to base**. E.g. it is easy to climb a hill, hard to retrace exactly.
- Walking down-ridge is **divergent — deceptive and dangerous**.
  You cannot easily see ridges peeling off below, and the map won't mark all the creeks you can see.
  If you are trying to retrace your steps or navigate from the map, you will have to be very careful.
- Ridges converge to hilltops — so walking **downridge** leads you widely astray.
- **Upstream** is divergent — you don't know where you will end up, but at least you can find your way back, because creeks converge toward the river mouth.
- Tracks & trails converge to: popular features, bridges, car parks, waterholes, homesteads, campsites, clearings, etc.
  So **leaving a focal point faces you with multiple choices** "How on earth do I get outta here!"

Coping with Divergence

The result of "directions" such as "You can't go wrong. You can't miss it" is that the truer that is, the easier it is to go wrong — after you have found it. Retracing your steps, or finding the desired way out, onward, can become a nightmare.

- Ask directions for the onward bit, not just how to get there.

For backtracking…

- Pay attention on the way in, despite your excitement at arriving. Note the direction of your arrival. Look back regularly. Mark the junctions you used, especially from minor out onto major paths. Retrace your own tracks.

Reverse logic

Be like yachtsmen, who should always explore harbours, shallows and tributaries at low tide "in case we do get stuck".

- Explore upstream — a divergent labyrinth — so that you can walk downstream to return.
  Provided you don't cross water-divides, or overshoot your starting point, "you can hardly go wrong".
  Ahah! There's the hint: Before you set off upstream, peek behind you, to see when to stop "coming back downstream"!
- Explore downridge, and climb back upridge later. Only don't climb down anything you can't climb back up.
- Explore away from a focus, so that you can converge back to it.

Isolated Vegetation Shows the Clearest Asymmetry

E.g. Look in any **clearing** for downwind debris/erosion,
and for up-Sun/down-Sun asymmetry at the edges.
E.g. an isolated tree in a paddock...
E.g. an open creek will give the bushes a clear view across the water...
E.g. a roadway will exaggerate the light asymmetry on one side
and suppress it on the other...
E.g. Look in the wheel ruts of bush tracks for asymmetries due to wind & light…

Examine **the tallest trees — The emergent ones** have a clear view of the Sun.

At the **edge** of a clearing, subtract the effect of wind…

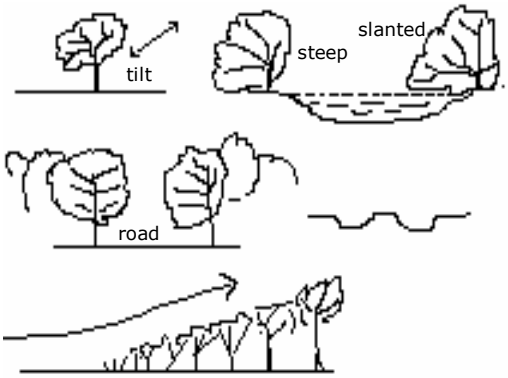

Closed Canopy — Look for Gaps

Each tree canopy might look uniform, but be blown back on the upwind side,
leaving a gap there, with dead branches, where it *has tried* to fill the gap.
But look for a gap without dead branches and without other explanation
— It might be 'shady south', which gives the tree little incentive to cover.
The tree *could have* spread that far by now, if it had wanted to, but hasn't.

Tall and quick growing saplings **show the most asymmetry...**

They can be the best tell-tales of all, and the worst!
Because they respond to clearings, slope, water-shine, nearby trees — be careful!
Take extra care to <u>average</u> out their message.
Any one tree gives you <u>too</u> definite an impression!
Back-to-back saplings may show <u>opposite</u> asymmetry, by shading each other...
Treat the overall shape as <u>one</u> tree.
The sapling effect, arises from young forest trees <u>shedding</u> branches more often
and more easily than mature trees.

Roadsides & Campsites Reverse Some Effects!

Examine trunks for broken or lopped boughs before interpreting!
The bigger lower branches are easiest and best firewood pickings!

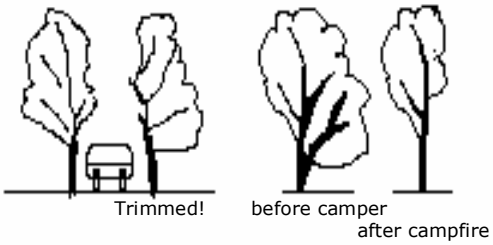

Name the Subsections

Heart-Attack Hill	Bust-me-gall Bend	Starve-gut-Ridge			: Be Vivid
On-the-Double Decline	Cross-the-Creeks Roller-Coaster	Blackberry Bank			: Be Descriptive
Left-hand Loop	Clockwise Climb	Right-around Contour	The Zed Bend		: Be Cheiral
N-E Straight	Into the Sun Slog	West Bank	Downwind Dunes		: Be Directional
Long Leg	Short Stretch	The Middle Mile			: Tell the Distance
Last-legs Leg	Dinnertime Creek	Sunset Strip			: Tell the Time
On the Left of Third Hill		Nine Mile Beach	Five Ways	Fourth Island	: Count
The Bush Blocks	The Rock Band	Turning Point	The Up-Part	The Trip down	: Be Clever
Rubble Ramble	Possum Passage	Flower Flats			: Alliterate
The Birdwatching Bit	The Deer Park	The Gloriously Green Glade			: Be Observant
The Monkey Saddle (three creeks)		The Big Bulge	Round Shoulder	Tight Turn Tongue	: Be Shapely
The Breakaway	The Hither Side of the Hill		Round the Rim	Unto the Brink	: Be Literary
Brown Beach Bay	Wobbly Shacks (= Mount Wobbly behind the shacks)			Zigzag Track	: Be Visual
Perverse Reverse	Inland Cusp	Round the Back	Out to Sea and Back		: Name Whole Curves

Name the Sequence

Fit Each Subsection into a Larger Scheme.

- Name the Links : The Cuvier Track The Mount Barker Road The East West Ridge
- Name the Junctions : Point Pass Steppes Junction Tee -tree Junction
- Contextualise : Beyond Burrowing Crayfish Ridge Cross Catchment Climb
- Overlap Sections : Deer Sally to Sandy's Shore to All-Terrain Trek } these are simultaneous
  Sherwood Forest to Bumpy Flat to Uphill Haul } and overlapped 50%
  I.e. halfway through one section, think about how it is changing character
  into the next obvious section, and anticipate it.
- Time-Distance Clues : Lunchtime Lookout The Approach Thither Heather Tea Junction Ten-Mile Dam
- Lie-Linkage : Downstream from Dusty Waterfall Wind-Shadow Plateau
- Numbers : Fourth Bump Fifth Creek Tee Junction # Three
- Acrostic : Short Steep Step in the Southern Subdivision after the Shallow Segment

Contour Curves: Your Job is to Imagine the Shapes into Life And to Match the Lines to the Landscape

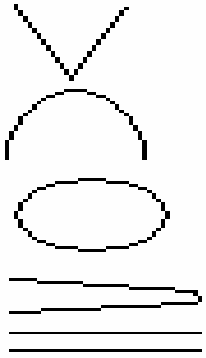

Deep Vee 'Ditches' — they gather water. The **sharp** points, point **up**-river. (sharp/up rhymes in p)

Round 'Shoulders' — they shed water. The **round** points point **down**-ridge. (round/down rhymes)

— being higher ground, they separate gullies

Peak or Depression — you can't easily tell which. Peaks are much more common.

Hairpin 'Ridges' — they can look parallel, like 'a slope', if the edge of the map cuts them off.
Ridges usually show much rounder than this.

Parallel 'Slopes' —be careful about which way is 'up'. Check that they are not the same height—on a ridge or river

Here is a typical confusing array of parallel lines.
The first two are ambiguous, but seem to lead down into a valley…
The valley is shown as lines 3, 4 & 5…

The three corners here are all sharper than the corresponding ridges.
Downstream is to the left…
Imagine a deep-vee ditch.

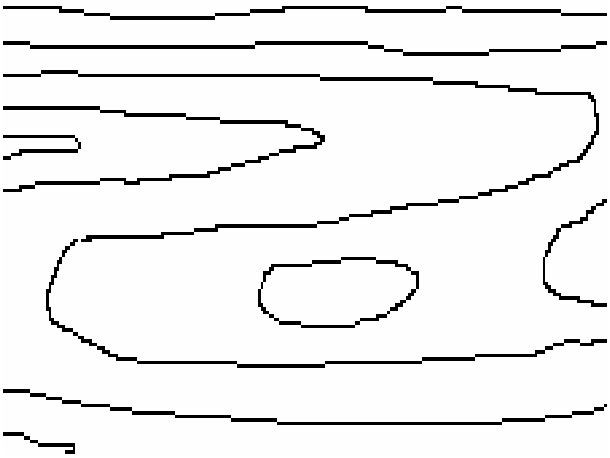

The shoulders and peaks shown here are all much rounder…
..Uphill is → peak ← saddle → ridge or peak …….
Imagine a circular mound with a spur down to left and right.

Down here, upslope to higher ground is…↑

Trackless Navigation

When you leave the bush network to cut across the natural routes, you still navigate by them, by noticing when you cross them.
You choose a route in relation to the framework — of ridges, rivers, and the contours that cross them at right angles.
E.g. "I am now going diagonally right, uphill, on the left bank, to the Ridge, then…"
You describe the sort of route you are following, and **start a new subsection whenever that sort of route changes** in quality.
E.g. A dry shallow creekbed marks a change from downhill to uphill.
E.g. The major break of slope—on the "hillside"—marks the change between being on the "hill/crest/ridge" or in the "valley".
Don't forget to ask, as you cross each 'road', what that ridge or creek is doing, to the left & to the right e.g. joining, splitting, rising, falling, dividing major valleys.

The method of Names

The point of this method is to bring your normally subconscious memory into conscious focus.
You put a **name** i.e. an essential description, onto each subsection. This is in addition to naming the creeks and ridges you cross.
That is equivalent to supplying your own signposts & flashing lights to the landscape. But write down the other details also.
As with memorising any journey there are a handful of steps: Notice; Recognise; Describe; Name; Remember; Put it into Sequence.

Bush Networking Essentials

Answer up to six backbone questions about each leg, before deciding on an appropriate name for that subsection…

1. Hill-wise — What Peak? — Is your direction trend in relation to it Uphill or Downhill; towards or away?
2. Hillside-wise — Slope — Choose between Upslope/Downslope/Flat/Contouring.
3. Ridge-wise — Rise — Upridge/Downridge/Crossridge?
4. Drainage-wise — Lie — Upstream/Downstream/Crossstream?
5. Side-wise — using the standard names* — Left bank/Right bank (below the break of slope); Right flank/Left flank (above).
   *As a yachtsman going **up-**stream sees it — Port and Starboard. Use these terms in your names to remind you.
   As a traveler going **up-**ridge sees it.
   I.e. the standard directions are **up**-stream; **up**-ridge; **up**-north; **up**-wind; **up**-hill, **up**-slope.
   Be careful! The 'right bank' is the 'left flank'. You can climb to the right on the left flank (of a ridge which falls to your right).
   Walking diagonally 'downslope' can be 'upstream'.
6. Hand-wise — Tack — Clockwise around or Left around/On the left diagonal or Right diagonal?

Mnemonic: (With your right index finger) touch your **head** (the peak);
forehead (the slope);
bridge of your nose (the break of slope);
down the ridge of your **nose** (the ridge);
nostrils (the drainage);
right **cheek** (the side);
and now look at your **hand** in wonderment (What is my tack?)
Then **point** out some peculiarities...
Now **add the non-essential** — exact measurements e.g. direction/length/time/steepness/count —and descriptions, sequence, context.
(Remember: don't navigate only by networks.) See opposite, for descriptive names.

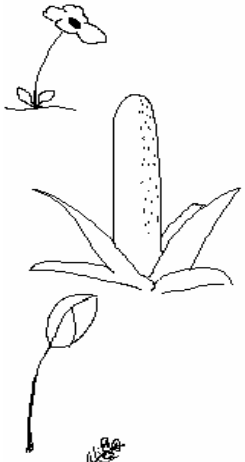

Flowers Respond to Sunlight

Examine the stalk-lean and flower-head tilt
e.g. of daisies.
Look up-Sun and down-Sun to see the striking difference, whenever the Sun is low.

Look for the first flush of flowering on the sunny side of a flower spike
E.g. Yakka (Xanthorrhoea). The mid-morning sunlight direction gets the earliest flowers.

Examine closed flowers to see whether they betray where the Sun last came from.
Many flowers 'follow' the Sun around.

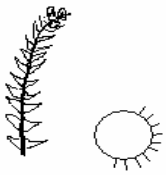

Some leafy stems seem to tilt to the noon, perhaps to point their terminal flowers into the Sun.
E.g. Sea Spurge

(Top view)
Banksia pistils sprout earlier on the sunny side.

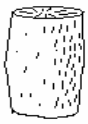

Banksia stamens deepen in colour on the sunny side.

Fruit Distribution and Ripening

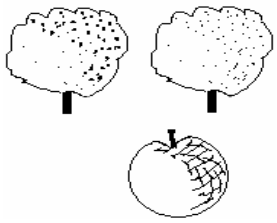

Some bushes will fertilise more flowers in the sunlight because pollinators can see them better.

Fruit are fertilised flowers. So some species 'set' more fruit on the sunny side.
Look in all directions to see which aspect presents the most flowers or fruits to view.

'Ripe' colours are a result of sunlight e.g. apple, peach.
Sun-ripened colours aren't confined to fruits.
Mushrooms are often old enough to have seen some sunlight.
One mushroom in the open may not help you much, but they do come in flushes, so look for a *pattern*.

Remember the Sequence from AA to ZZ

— A Ready Made Acrostic of memorable unique items, with endless extension if needed. I recommend it as the pick of the mnemonics.

You may need to remember each tree in a forest walk, or every twist in a path, or every choice in a maze, or every square in a grid.
AA to ZZ supplies 676 *differently named and sequenced letter pairs*.

If you turn the letter pairs into appropriate *two word phrases*, there is no limit to the useful memory aids.
E.g. I could call this system the Alphabetic Zygote (of 'letters paired' A-Z).
As with all gimmicks, it only works best if you *study and* ***memorise*** **the object** ***receiving*** **the label**.
Examine the tree for anything which could attract the label DF (damp foliage; dendritic form; downpointing flowers),
or mentally *photograph* the MS munch-stop, or *actually* bird-watch at BW.

It's up to you to use the letter pair in a memorable way. Don't forget to use visual, emotional, sequenced and appropriate words.
Consistent rules help.
On a pure sequence walk, begin with double letters, rather than with A-Z single letters. This makes two word phrases the consistent rule.
In fact begin with *any* double letter, but especially **the initials of the walk**, e.g. BL for Blackmans Lagoon.
Two word Phrases e.g. AA = Ask Again, avoid the ambiguity about which other letter in the word or phrase might be emphasised.
E.g. is Aeroplane, AE or AR or AP or AN? Was 'Anabolic steroids' A or AS?
Add a third word to the phrase as needed, to spell out what it means, e.g. Backwater Island turnoff.

The trick is to make the mnemonic **match** the feature.
E.g. if there is a plague of mosquitoes which you attack with a stick, Aaron's Almond-tree walking stick might be appropriate for AA.
Then the need is to somehow link the previous with the next.
E.g. AB Aaron's Budded walking stick (if you know that story), might focus your attention on the buds in the next gum-tree.
Without a '**theme**' it is difficult to pluck out your memory what 'DF' might have been. But 'Doctor's Emergency' where you stubbed your toe, leads on to 'Doctor's Ferari' when you walk on quickly anyway, and on to 'Doctor's Gasoline' where you topped up with lunch.

If you do a 'grid-line walk', you use A-Z to begin with, and say you move from A0 to C4; you might go through A1, A2, B2, B3, C3, C4.
The memory aid goes A0, then Ahead on 1 leg, hopping, then Across, on 2 legs, Ban-tu, Ban-three, See the three cattle?, Count to 4...
'Minus', in cases of ambiguity, might come out as 'which was 4 metres deep', or 'where I left 4 pebbles behind', or 'and Not 4 blisters' or 'at anything but 4pm'.
When you get into AD you look for a-d **peculiarities** to remember like 'Add 4+6 wattles (to get 46)' or '33AD crucifixion tree' or 'Adder-Death, two score of them'.
As an alternative, name all grid squares with *letters*, e.g. D-G, and look for things to name, such as Dark Green; Discarded Gaiter.

In a maze, you could choose Long words for turning Right; Short words for turning Left, a mixed phrase for Straight Ahead. These match the meaning-words in length of sound.
Maybe you need some **variations**:
One, two, three or four syllables for left, right, through, reverse.
Long first vowel sound vs short vowel sound.
Begin with one *letter* and *end* with the other *sound*, e.g. Amanda, Abib, Ace, Add, Abbey…
Let sounds take priority where needed. E.g. few words start with K, but many start with a hard C. X is hard unless you allow Ex-.

Memorised sequences are most effective if revised almost on the spot, e.g. via a leap-frog walk (2 forward, one back), and on the return.
Equally, other methods are enhanced by adding this mnemonic, such as when marking a trail, writing track notes, or making a mudmap.

Contour Heights: Read the Numbers!

Determine **the Contour Interval** — and interpret which way is 'up' from the **numbers**… Note: On maps, the numbers are printed to be read as if looking **uphill.**

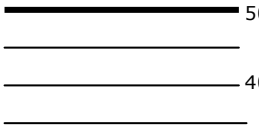

This first diagram is "every 50 m". And 'up' is toward the top.

Zero means "sea level"
Read the absolute numbers!

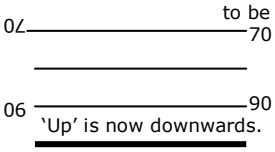

In the second diagram, the contour interval is every '10 m'

'Up' is now downwards.

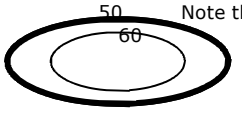

Note that **major** contours may be marked as thicker e.g. every 50 or 100 m

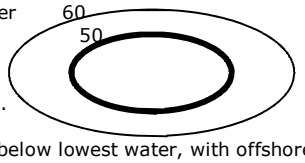

..The first is a peak; the second is a hollow...

Marine charts show spot-depths, 'soundings', in metres and decimeters, below lowest water, with offshore depth-contours.
If the "depth" is underlined, it refers to a "drying height" of intertidal rocks.
At sea, keep a record of the soundings as you sail in to land, preferably in a straight line. Plot them onto tracing paper or mark them on the edge of a ruler, and try to match the pattern to charted depths.

Contour Spacing: Close = Steep Slope

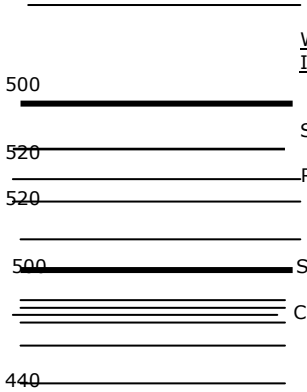

Wide spacing = flat land Examples: mountain plateaus; valley floors; paddocks; dunes
Irregular contour shapes (meandering all over the map; not shown here) = very flat land

Slope

Ridge — carefully notice when numbers are identical.

The rule "Closer Equals Steeper" only holds for different height levels, not for hairpin contours around ridges or rivers.
In fact, "close" might equal "flat" i.e. on a ridgetop

Slope

Cliff — Very close = Very steep

Slope; have you worked out downhill?

Walking "Straight" through the Bush

- 'Straight' downstream; upstream; down-ridge; up-ridge; uphill; downhill
- 'Straight' around the contour
- 'Straight' along the path of least resistance
- 'Straight' through the obstacle course, dodging to left and right
- The Line 'of Best fit' — to a chosen trend cross-country

None of these are straight, but most are easier than 'straight across a slope' — crossing innumerable ups and downs!
Most are predictable.
The navigator simply goes last in line with a compass and logs the twists and turns, whatever course that takes.

Why last in line? It is a fact of life that following your nose through the bush quickly disorients the leader of the party, while those following behind maintain their bearings easily.
Also, to clearly show up the trend of a sinuous course, the last person can drag a long smooth string loosely behind him.

Straight around

At a…

- Constant Angle—to the (changing) Sun, Moon, star, planet, landmark e.g. keep the peak "ahead on your left" — a spiral course.
- Constant Distance Off — as judged by the rule of thumb — you keep a prominent landmark appearing to be a constant size — from base to peak. The course is circular, concentric to the landmark. E.g. offshore, around a lighthouse.
- Constant Horizontal Angle — between two prominent features. This gives a cusp-like circular course.

You can retrace these courses, but the going may well be difficult — crossing many streams and ridges.

How to Walk in an Unintentional Circle

Haven't you heard of people walking in a large circle when they are lost?
They thought that they were walking in a straight line, but some hours later they recognise the same scenery!
I've done it; it's no myth. Discover your natural drift...

Blindfold yourself. Try to walk straight across a flat. I bet you can't.
Most people have one slightly shorter leg, usually their left leg.

Now try to walk across a slope, blindfolded of course, in a straight line. I bet you drift.
But once you realise it, you are just as likely to head uphill too much, in overcompensation.

Don't trust your senses in a fog. Find out by trials at home — before you find out the hard way — how far 'off' you are.
Also find out how to properly walk in a straight line in a fog!

Stem and Leaf Colouring

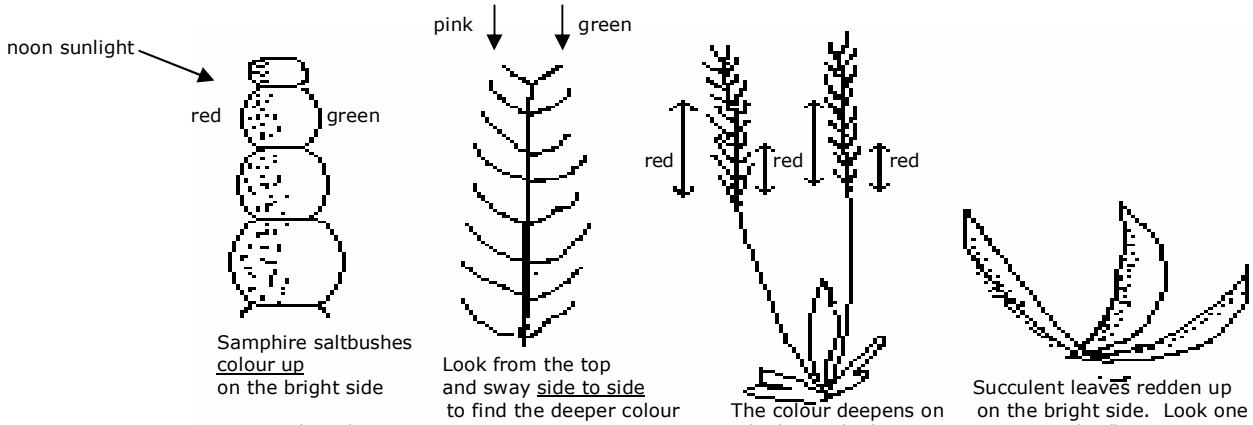

Samphire saltbushes colour up on the bright side

Many other plants show red stems or leaves etc Look for a paler side.

Look from the top and sway side to side to find the deeper colour

The colour deepens on the bracts higher up on the sunny side

Succulent leaves redden up on the bright side. Look one way over the flats & it is pinker than the other way, especially just after sunset!

N.B. Older stems may show a reverse trend! The colour fades on the sunny side after having deepened more while young.

Beware: When flower-stalks twist as they grow, the colouring or fading effect is unreliable. You can detect the twisting *by* the fading.

Example: Blackberry stems and leaves redden in sunlight, staying green on the shady side.

New Growth is Colour Differentiated

Looking up- or down-Sun (on a cloudy day) will show up new growth as yellowish or brownish or bluish (depending on the species) in one direction more than the other.
The sunny side will 'green up' the new leaves more quickly and will look more normal coloured.
But test it out first — it may go the other way — sunlight might promote new growth, before it colours it up.
When new growth is distributed throughout the shrub, the sunny side will often show deeper green (i.e. more of the same pigment) and be more vari-coloured e.g. some leaves take on a red blush or red margins in response to the Sun.

Revise your Memories

..after a 10 minute distraction
..after an hour, a day, a week, a month, a year e.g. for an acrostic which you wish to remember permanently
..and whenever after, when you begin to have trouble remembering
Choose undistracted, mindless times
e.g. while trudging or being driven, waiting or relaxing, or last thing before sleep, to run over things in your head.
Overpractise your mistakes. Out loud is always better. Take any opportunity to tell someone.

To Remember Formulae such as those at p. 19b

(1) Make up memorable ditties, and refresh your memory, as above, before you lose them.
(2) 'Read' the formulae *for sense* — They are just shorthand for how things *must* behave. If you don't like maths, it's probably because you never understood the common sense behind it, and that is probably because you never had the opportunity to see it, or developed the skill to read formulae.
E.g. '**Sin, Sin, Sin! It's High Latitudes which Declines it**, for both Easterners and Westerners'
for: the three sine ratios in the formula… 'sin (**heigh**t E-W) sin (**latitude**) = sin (**declin**ation)'
i.e. as the latitude increases, the height must decrease, and vice-versa, to preserve any sin (dec) figure as constant.
The formula should return zero height for zero declination, and 90° height for dec=lat.

This next one assumes that you measure the angle away from East-West (rather than from the 'elevated pole', as on p 19b).
You can use 'sine of the variation from E-W' instead of 'cosine of the angle from the pole', because 'co-sine' says 'sine of the co-angle'.
'The sin of the increasing the sunset angle, by cause of increasing latitude, is a sign of decline'.
sin (sunset angle) (multiplied by) cos (lat) = sin (dec)
Read this as: For any specific declination, as lat increases (=the cos reduces), the angle must increase, to preserve a constant sin (dec).
You can also put in dec=0 and expect a 0° answer no matter what the lat.
At the Equator, declination directly determines the direction of set N or S of E-W.

'Why bother with the E-W Sun? Be**cause of its E-W *timing*!** It's **today or too late**' t for tan; d for dec; late for lat;
90° = 6hrs each side of noon cos (time of E-W) = tan D / tan L, *'or' for "o'er" for 'over'*
Lat = 0 throws the calculation into a spin, because any declination other than 0 can't be E or W. Dec = 0 must return 90° for any lat.
Dec = 90 must be impossible (it's N or S!), but so must any dec > lat, since cosine doesn't go above 1. Those stars stay polewards.

'One New Day at a Time? Not quite, **'cause for tomorrow's sunrise time, today's time's not too late**!'
It changes only slowly, as dec changes. cos (time of sunrise) = tan D x —tan L 15° = 1 hr. each side of noon
0,0 must return 90°. Opposite hemispheres must return a large positive cos and hence a smaller angle, and a short day.
Same hemispheres must return a negative cos, which means a long day.
Too much lat, and the tropical bodies disappear altogether, as expected. Too much dec, like a circumpolar star, and it won't even set.

Try **the Equation of Time**, p.17b.
The First FFive months, From FFebruary, Fence-sit Four Features (halF-each side). The next 7 *finish* at four more features.
−1**4**,0,**4**, 0, -**6**,0,1**6**, 0 Makes nice mirror patterns of the two sets of four features. Minus means a fast Sun, up early; a 'slow' clock.
FFFebruary Founders at FFFourteen; Marches up again; April crosses (its hot-cross buns); May turns 4; June double-crosses;
July *finishes* work at six; August rises augustly, to zero; SSSeptember and October get up to SSSixteen;
November and **Dec**ember **dec**line a lot; December crosses at Cross-mass; January Sun decends from the cross, fast, for a new year.

Contour Slopes: Work Out 1 mm and 1 cm Contour Spacing

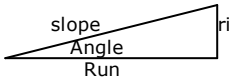

Example: a Rise of 10 m in a Run of 100 m is a Slope of one in 10 and an angle of 5.7°
[Tangent(angle)=rise/run Angle=Arctan(rise/run)]
rise/run is technically called the "gradient" i.e. an up-slope.

1) Example of 1 mm contour spacing: A 1:100,000 map-scale says: 1mm equals 100m (See 'map scaling factors' p.92a)
with 50m contours the slope must be 50m every 100m or 1 in 2 or 26°½.
2) Example, with 1 cm contour spacing: 10 times less slope, in the same example, or one in 20, or 3°

Using these 2 reference spacings (**work them out again for every new map**) fit the actual contour spacing (eg 3mm) between them.
E.g. map scale: 1:25,000 (1mm = 25m); for 10m contours, 10m rise in the 1mm run, on the map (=25m in reality) is,
"1mm contour spacing = 1 in 2.5"; (10mm contour spacing) = 1 in 25; (20mm contour spacing) = 1 in 50;
So… 3mm means 1 in 7.5

Reference Slopes — To Imagine the Climb with!

One in one; 45°; One up / one across; On all fours!
One Degree is 'one in 57' (or approximately 1 in 60) So 10° is 'one in 5.7' — maximum for public **roads**
One in Four is 14° — a neat coincidence to memorise. So One in 40 is 1.4° — maximum for **railways**
36° slope — 8 in 11 — more or less maximum for sand, gravel, scree, or any loose material.
Small slopes may be multiplied with approximate linear accuracy, but in an **inverse** fashion…
Examples: 2° is one in 30 — from the one in 60 rule
1 in eight is 7° — from the one in four rule
5° is one in 11 — either from 57/5 or from 5.7 x 2 or 1/8 of 36° see also "sine-cos chart" p.106b

The Beaufort Wind Scale — Modified for Bushwalkers

One in 40	contour lines absent or rare; pleasant ambling; wind whistles through the lips; grey hair relatively common
One in 30	car engine quieter than road noise; perambulators plentiful; locomotives absent.
One in 20	car engine noise louder than road noise; walkers plentiful
One in 15	skin moist despite cold conditions; heavy breathing is heard above background noise
One in 10	eyes sting from sweat; bicycles experience difficulty staying upright
One in eight	flies are constant companions; car smoke exits parallel to the ground; mountain bikers walk
One in six	cigarette butts become less frequent; 2 wheel drives mostly absent
One in five	walkers experience difficulty making forward progress
One in four	cannot hear the sound of a cigarette being lit; leaf litter easily lifted off ground
One in three	sturdy walkers stagger and sway; frequent rests; slow progress; loose soil swept away
One in two	large trees tilt at a noticeable angle to the ground; rocks experience difficulty staying in place; walkers remain upright only by zigzagging
One in one	toenails begin to bleed; trail bikes begin to bleed
Two in one	fingernails begin to bleed; small trees are pulled up by their roots
Three in one	contour lines begin to merge; danger of severe falls; you are in danger of being completely… winded!

Walking an Exact Straight Line in Open Country

Assume the visibility is limited, or the terrain difficult, or the vegetation is closed in…

- Lengthen the Whole Party—In Single File. Shout to the leader, if he wanders to left or right. Stop still if you lose sight of the person behind you, i.e. do something to make sure that no one gets lost in the fog.
- Drag a flat object (or a person) on a long string. Tie your end of the string onto a stick, and hold it forward above your head. Alternatively, pivot the stick in your hand at its halfway point, and it will point in the right direction.
(Watch that the wind doesn't blow it sideways; Stop the flat object rolling; Stop it sliding downhill; Don't let it drag at an angle)
Try just a slippery fishing line with no object.

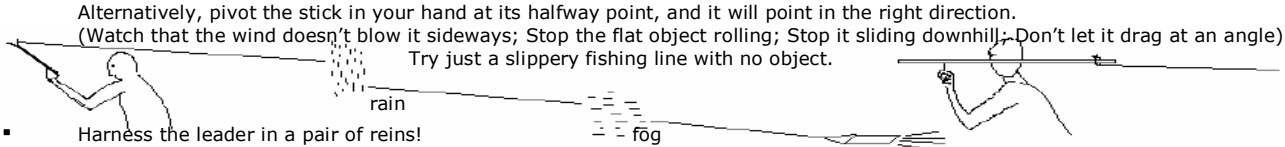

- Harness the leader in a pair of reins!
Tie one string to either arm of the leader, and steer him properly.
- Hold the rear end of the string over a compass.
- Leave a trail of visible markers lined up exactly.
You need to look back frequently and drop a new marker.
— stones, sticks, stumps, tree trunks
Leapfrog with only three marker sticks — take them with you.
Leapfrog people!—leave someone behind; send someone ahead; line them up with each other & with a compass if available.

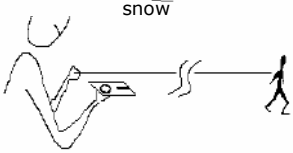

- Scratch a line into the dirt. Look back frequently. Readjust yourself
- Use the radio for radio direction finding.
- Orient by Sunset glow, the wind on your cheeks, the blue band, town lights, distant noise, etc.

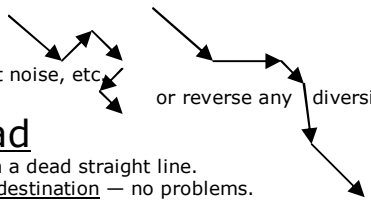

To keep to a fixed "track", manoeuvre around obstacles in a rectangular fashion, or reverse any diversion.

- # Leap-frog a Straight Line by Looking Ahead

A. A straight line course is easily achieved visually — like keeping a hundred fence posts in a dead straight line.
In clear open country, you simply walk toward a new mark halfway between you and your destination — no problems.
Keep the 2 lined up. When you get halfway choose another halfway point in line. If you can keep sight of the more distant destination, this method means you only use the magnetic compass once, to locate the distant destination—then keep a fixed 'track' visually.
When obstacles force you sideways off track, you can easily regain track, visually.
B. In more difficult circumstances, you cannot see far ahead and usually rely on looking back, as above.
But leaving markers behind you slows you down and keeps you looking around.
This method does not require you to look back, but requires a simple discipline — leapfrogging visible marks exactly ahead of you.
Step one: Choose two distinct marks exactly in line with your desired direction
Step two: **On reaching the first mark, choose a third, still in line. And so on.**
You can do this while following a formed track. The track will bend from the average straight line, but you'll notice that immediately.
I.e. when you don't get to the third mark, and have to start again, you are well aware of the change in trend.

You

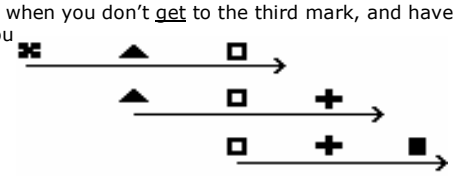

The end result accumulates errors
(despite the averaging process).

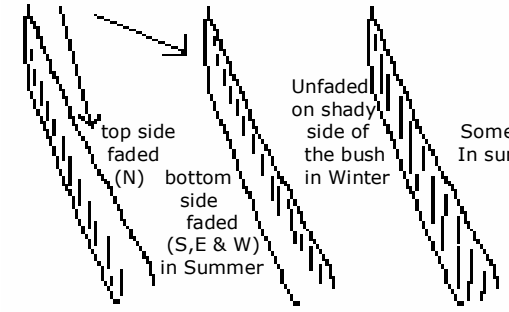

Noon-Sunlight is More Vertical

Examine needle bush (e.g. Hakea) needles, one by one.

Some effects will be more noticeable in winter, and in higher latitudes.
In summer, the 'shady' side is actually bathed in horizontal sunlight twice a day!

The Shedding of Bark on the Sun-Baked Side

The patchwork of coloured fresh new bark shows that
the bark strips an extra layer deep on the sunny side.
Look also at the coloration of the freshly exposed bark, for asymmetry.
The bark 'stocking' clings higher on the shady side.
It is older, deeper, darker and allows more and older lichens.
Look for the 'curl' of bark ribbons — the pipe-like shapes peeling off —
as there may be more curling shown up on the dry side,
but it depends on the species.

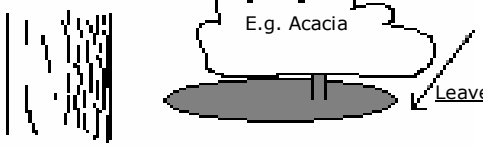

Where to Look for Fading or Lack of it

****Turn leaf blades around** to compare the greenness in each direction.
Find a pair facing the same direction, and twist one of them, then make a comparison to see which side is more faded.

****Twist an *untwisted vertical reed stem***, to detect the more faded side.

Look at stumps and old posts, tree trunks and fallen sticks.

E.g. Sunlight bleaches the sunny side of bark and dries out any secretions

E.g. The dark side of the trunk will often be darkly stained
e.g. with cider-gum secretions.

Leaves dropped into the shady zone will not fade and dry out so quickly.
The shady zone is offset toward the pole,
because the sunlight penetrates deeper on the noon side.

The broken faces of lying timber may be more coloured
and less bleached in the shade.

Bumps and slopes, diggings and burrows, may show dead dark moss on the hot side, or grass which has dried up.

Look for a very subtle pink glow on the shady side of dead timber, which is absent or greyer and paler on the exposed side.
Once your brain sees a hint of it, it becomes easier to see—so persevere.

Bracken fronds die off from their tips, but the fronds that retain their greenness the longest are on the lowest, shady side.
Dead thistle stems fade on the Sunny side.

The topsides of grass blades fade more due to being blown over flat by strong **wind**s, and so are exposed to sunlight, wind & drying.
Upwind edges of green bushes often show a brown wind-scorch.

Fresh diggings (echidna, bandicoot, wombat) expose soil to bleaching. Look at the steepest sides for the least faded soil.
Some fading is due to **oxidization** (exposure to air), some to **dessication** (loss of humidity), some to sun-**bleaching** (exposure to ultra-violet). Again it will pay to study the various effects before you need to rely on them.
Also be aware that diffuse lichens will give a covering colour to dirt, wood, rocks, etc, which may complicate the appearances.

Button-grass flower-stalks fade but darken on their exposed side, perhaps due to fungus. One side is dark grey; the other pale yellow.

Make a Clinometer — a Slope Measurer

One person holds the loop knots apart under slight tension. He barely touches both plumb-bobs onto the ground, to settle them.
Their identical lengths ensure a parallelogram.
The second person alters the middle string until it just touches the plumb-line.
The identical lengths ensure an isosceles triangle.
The 5 cm marks represent whole degrees of slope. Use 6 m of non stretchy cord. Measure and constru .

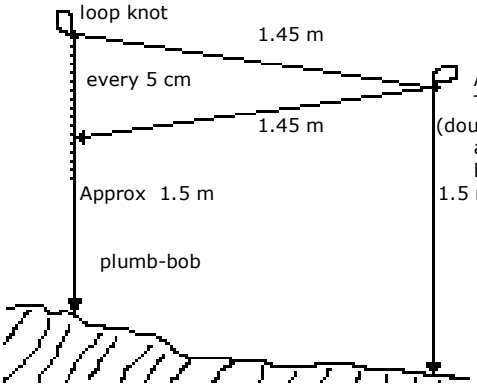

E.g. the slope is 11°½ here…

Accurate from 0° up to 20° or so
The method is based on the isosceles triangle rule
(double the half-angle sine; nearly linear for small angles)
and on 1 in 56. You can use 1.40 to 1.50,
but the lengths must be exactly equal.

The one-man version has two foot-loops.
Spread you feet and hold up the plumb bob.
See which side the plumb leans to.

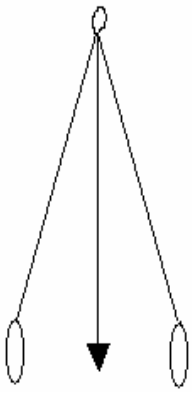

Use it occasionally, to refresh your judgment of slight slopes.
It may also be used to gauge the height of landmarks or stars. (Hang the second loop knot from a sky hook; Sight along the top string; Steady the bobs). Used horizontally, it yields horizontal angles. (Complete the string parallelogram; Sight along the two 'top' strings; and double the reading)
Of course this is only for a hobby, for interest's sake, for mathematical curiosity, not for absolute accuracy!
The more you play, the more you learn.

Just Draw It

On the map… Peak, 500 m elevation

High ground, 300 m

Camp, 200 m elevation

The question is…
Can I see the peak from camp?
Draw three lines of 5, 3, 2 units…

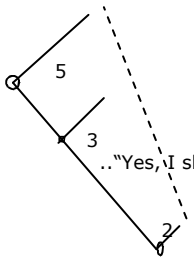

..."Yes, I should see over the ridge to the peak"

Radio Direction Finding — As Good As a Compass

AM stations — Amplitude Modulated — the older style, for medium distance transmissions
— also called Medium Wave; Medium Frequency — MW; MF

Your portable radio has an **internal aerial** for AM stations, which is direction sensitive.
It usually lies lengthwise inside the radio, as a long **ferrite rod** wrapped in a wire coil.
To make that ferrite rod horizontal, you probably need to lay the radio down flat.
Now twist the radio to change its horizontal direction.

When the internal ferrite rod points at the medium wave transmitter,
the reception drops off — you have detected a "null".

This occurs at two opposite orientations — **pointing away from the transmitter as well**.
(Good reception is by pointing the radio 90° away from that null direction.)

FM stations — Frequency Modulated — the newer style, of shorter distance transmission
— Very high, Ultrahigh frequencies — VHF; UHF — It does not use a ferrite rod.

With a whip aerial, or an extensible aerial, or a long wire aerial,
you can detect a null by pointing the aerial at the transmitter.
This is for horizontally polarised transmissions. Some stations are polarised vertically.

Shortwave stations — Amplitude Modulated — higher frequencies, for longer distance transmission
— SW; High-Frequency, HF — It usually requires an external aerial.

Use a **long wire aerial**; keep the ferrite rod perpendicular to the long aerial.
The null occurs when the long aerial does not point to the transmitter — but it is perpendicular to that direction. (Opposite to FM)
Point the long-wire aerial perpendicular to the transmitter, for a 'null'.

Weak AM stations can also be used, provided you have a long wire aerial to pull them in. The exact length sometimes matters—you will notice the volume rise and fall in cycles as you wind the aerial back in.

THE RADIO "NULL" ACTS AS A COMPASS NEEDLE — it keeps a constant angle to the distant transmitter tower or repeater.
You swing the radio aerial each side of the null to keep checking its direction.
To avoid having to give the radio constant attention, break your course into separate lines of sight ('legs') as with a compass.
Marine navigators know that radio signal direction can be distorted by mountains, or by the signal crossing the coast at an angle, or by reflections off the rigging, but bushwalkers seldom have to worry about that, because they only want a constant reference direction, not an exact bearing.

You could consider setting up a radio beacon so that you can always say: e.g. "The car is that way" or "Home-base is yonder"

End of Section 3.

Distance Guesstimation is a Major Problem for Walkers

Because distance measurement is a major problem, we find it easier to ignore.
We resort to some other method of navigation.
If we simply, out of laziness, give up on logging our pace & distance-already-come, we waste a sensible cross-check.

The first of these three sub-sections will look at the amount of distance you have covered — progress — distance made good…

- Elapsed distance — puts you on the map precisely — unless you are not very sure of your progress! or the map scale!
- Map scale — "It looked so easy on the map!" — Judging time, distance & difficulty is a question of reading the map-scale properly. "How far could we get by lunchtime?". Guessing at a curved distance is always a bit of a bluff. If you hear yourself saying "On the map that's **about** 5 km" stop and measure from the map-scale more accurately!
- Sideways drift — is a matter of trigonometry, when you discover an error in your direction heading
- Up/down progress — is in another dimension. You don't need an altimeter, but it helps. (If you do take a portable altimeter/barometer, make sure to test it to see whether or not it reacts falsely to temperature change—put it in the fridge then in a warm place; then don't forget to keep adjusting it according to the atmospheric pressure changes)

The second sub-section is about the rate of covering distance — pace — and distance-yet-to-go.

The third sub-section is about the distance you can see, but don't cover — estimating the 'range'. How Far to That Landmark?
E.g. You can identify a peak on the map — but only if you are certain enough of the distance from you to it.

Distance = Speed x Time

E.g. 4 km per hour x 3 hours = 12 km
Hint: Measure only elapsed, 'walking' time; Delegate someone to keep track of accumulated time out for stops and diversions.
Hint: Mentally bring forward your starting time e.g. "(as if) we started at 10a.m." (You started at 9, but had two, ½ hour stops)
E.g. foam takes ½ a second to pass your 10 m yacht; speed equals 10/.5 = 20/1= 20 m per second.
E.g. I seem to have drifted 12 km in one day = ½ kph current. E.g. The cloud shadows cover 5km in 2minutes = 150kph wind.

Keep a Log!

That's how mariners keep track—they write it down in the special book. Do you carry a notebook?

Adding Two Speeds

Convert speed to a length (e.g. 4 kph = 4 cm) and direction to an arrow.
Add the two arrows, tail to tip, in either order.
Complete the triangle or quadrangle tail-to-tip.

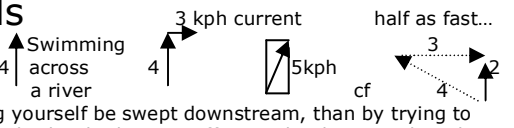

Note that it is faster to swim a river by pointing straight across and letting yourself be swept downstream, than by trying to stem the flow by pointing upstream. Only when the two speeds are perpendicular do they not affect each other e.g. the 4 km per hour is not affected by the current in the first example. Or e.g., If your boat speed is 4kph ahead, and the tidal flow is 5kph against you, your actual movement is backwards! (with any sideways component added in independently).

Example: Three air-speeds 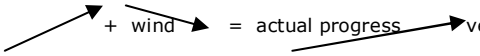 + wind = actual progress very fast across the ground, slightly to your right.

I.e. Your speed & direction through the medium (air/water), relative to the ground + the medium's movement, over the ground (the medium carries you when it moves) = your total speed and direction, over the ground.
Lastly, *to analyse apparent wind speed, to avoid confusion, draw only the three winds involved*. I move this way...
First convert your progress over the ground (top arrow) to an opposing apparent wind (right-hand arrow opposing it).
Don't try to add wind to your motion—that's not a sensible question/answer. The medium (air) doesn't carry you when it moves. Answer: I feel the actual ground-wind, a cross-wind (⟶) in my face, diagonally as shown…

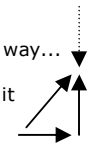

Range: Rule One: Guess

Then don't trust any calculation which doesn't agree with your initial guesstimation.
But remember that distance is deceptive over featureless terrain e.g. water/beach — we underestimate it.
Birds bob their heads to see how much an object shifts against foreground and background. It's worth some practice.

"Range" of Visibility — Two Halves to the Picture

"Extreme visual range" must be calculated in two bites — So Don't Forget To Add!
— e.g. when asking "Should we be able to see Ben Nevis from Brady's Peak?"
— e.g. at sea when you can first see the lighthouse lens, not just its loom.
From a kilometre high you can see 1° of latitude
before the curvature of the earth cuts your line of sight.
— e.g. you can see a 1 km high mountain from two degrees away
if you are on a 1 km high mountain — see diagram…

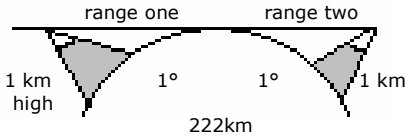

"Range Squared is Dee Height"

This is an easy mnemonic. "Dee" is dee Earth's Diameter i.e. Extreme Range = √(D x H) see p.92c
H is your height above the intervening horizon e.g. the wave tips, or a 250m plateau.
Since the Earth is 40,000 km around, divide by Pi to get **12,700 km diameter**.

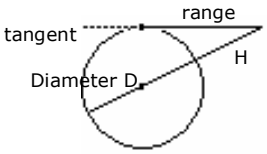

The formula is a theorem in geometry,
about the geometric mean (the tangent) of an extended diameter and the extension length (H).
The formula states: the square (range x range) = the rectangle (not shown: H x D+H).
(i.e. the range substitutes for both parts of the product, 'geometric-' or area-wise)
but since the height H, in reality, is very small, therefore **range^2 = D x H**

To "Read" a Map — Rule One: Don't Lose Your Place!

Get the geometry fixed in your vision — of where you are on the map e.g. "in the top right grid square, next to that lake".
And then don't keep losing your place every time you look at the map again!
..Whether the map is now upside-down, side-on, or whatever.

Rule Two: Imagine Arrows, for the Sun, & Your Movement

Again, visualise them geometrically, in relation to the whole map shape, from one corner of the map, for the Sun, and from where you are on the map, for your movement. When your map is properly oriented, the movement arrow will point in front of you, ahead, matching the trend of your movement onwards.

Rule Three: Get Used To Any New Map Scale

Maps are "Fractal" They show "self-similarity across scale"
i.e. A river system will look the same on a small-scale or large-scale map — there is always finer detail to show.
A road network is the same — the smaller the scale, the more roads are left out, leaving a similar looking network.
A coastline shows such self-similarity. Magnify the details and it still looks like 'a coastline'.

New Map-Scales are Hard to Adjust to
It takes time and testing to get used to a map scale, because they all look alike no matter what the scale.
To get used to the new 'feel' of a map, check up regularly on your map-progress, until you are no longer surprised by how far you have or haven't come.

Rule Four: Don't Give Up!

Don't let the combination of difficulties bluff you!
E.g. You misjudge the scale, misjudge the Sun's direction by optical illusion, you see turns & bumps not marked on the map, doubt the accuracy of the map, fail to assess the slope accurately,… and so eventually… give up in confusion!
Problems are normal, and **solving them by the map is normal too — don't give up trying**.

—Also See "Map Checklist" p.113—

"L.M.N.Or P."

L for Landscape; M for Map; N for Compass Needle: Or for Either/Or (the big point of the mnemonic)**; P for protractor**
Either/Or means: Each component of map&compass work can be used alone and independently for some tasks, if desired.
Plotting compass bearings can seem complicated. This mnemonic is designed to smooth the way.
A navigational compass has a protractor independent of its needle. The trick is to remember that!
The protractor has 2 parts too (p.93b). Note that some compasses have the magnet one-piece—glued onto—the protractor card.
If you get used to using each component separately, you will find 'map and compass work' to be simpler.
The trick is to isolate each of "LMN Or**P" in turn, to avoid confusion.** That includes the 2 moving parts of the protractor.
It is hard to use more than two at a time! so know which two you need to use at each stage.
E.g. A 'prismatic' compass will let you see both **L**and & numbers (**P**) at the same time—to take an accurate bearing e.g. at sea.
E.g. You can align the **M**ap to the **L**andscape, without ever using a number. Or you can point the **N**eedle to the horizon likewise.
E.g. The **P**rotractor may be used to assess slopes, to measure the Sun's height, to take a map bearing, to draw an angle on the map, to take a horizontal angle between two landmarks, to sight a vertical angle like a sextant (this may need two people), to find North from Sun-time, or to set a course to the left of the Sun.
So don't forget to use the **P**rotractor to its full potential, quite independently of that funny little piece of metal rattling around inside it. (**N**).
E.g. An emergency 'compass' is just a floating sewing **N**eedle, or a suspended magnet. So the protractor has little to do with orienting the map by the magnetic field, or with heading North by the compass needle.

If your navigational compass has a magnifying glass, a set-square and a graduated ruler, you should hardly get confused by thinking they had something to do with the magnetic field! — so don't worry when you see some numbers marked around your compass needle — I think you learnt how to use a protractor in elementary school, didn't you?

Working from land to compass is called "Taking a Bearing" (L+P1,N+P2) You use the protractor to put a number on the direction
..from map to compass is called "Getting a Heading". (M+P). You then use the needle (P+N) to find then follow that heading (L).
Working from compass to map is called "Plotting a Fix". (P+M). I.e. You use the protractor, ruler and set square to draw angles.

Maps Mean: Imagine Me! — Making Mental Maps

The meaning of using a map is that **it lines up**
— with what you see; with what you saw; with your mental maps; with what you know; with what you can't see & don't know.
E.g. You glimpsed the coastline as you drove in, the lie of the valleys, the drainage of the creeks, and of course you know the highway that you drove in on! You can no longer see them, but it is normal to orient yourself by imagining the invisible.
And then, your mental map is still incomplete, so that you need to use the map to add to your imagination…

"Map To Mind" Navigation

Transfer from the map both visible and invisible features back into your mental model and mental layout.
1. Start with the Invisible Big Pictures — visible on the map but not to you.
Look at the sky and the skyline as you interpret the map for what lies out of sight.
2. Add in the local visible features — visible both on the map and to you.
These two steps work together like lock and key — each is not very useful without the other.
Yet both steps are often neglected and seldom put together as a whole.

"The cars are now behind that notch in the skyline"
"The highway runs behind that hill then turns South-East"
"That low spot marks the river we want"
"Our camp tonight is in line between Mount Socks and the pine plantation yonder"

When you become temporarily bushed in, or blinkered by the rain, or shut in by the dark — then you see whether you can maintain proper orientation — by the invisible — or whether you just feel lost — "I can't see where I am!".
If you do not make a mental map and orient yourself by imagining the invisible—you will be disoriented, even if you "can see where you are".

♦ Part of the job is to look at the map in order to join your several local mental maps into one regional map.

Remember that a map is a great memory aid, or a great memory bypass — take your choice.

♦ Rather than "put yourself on the map", put the map in you!

Use the Map Scale — Don't Just Look at It!

Many errors of judgment result from guessing at the map distance, even when the map scale is clearly marked.
The scale is not much good down in the bottom corner of the map, fixed in place! So copy the map scale onto cardboard.
Somehow attach it to the map e.g. into a separate paper-pocket; onto a string; in the map bag.
The string is useful for measuring distance accurately. You place it over a curved path then stretch it out straight to measure it.
Point the corner of the cardboard at the starting position; position the string from there; then stretch the result along the scale.

The Jargon

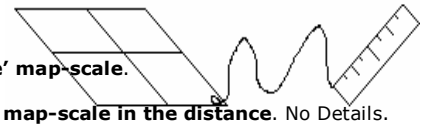

"Large-Scale" is like a "Full-Scale Model" It shows all the details.
Like a 'large-scale simulation' — it leaves out nothing. **Picture a magnified 'Large' map-scale**.
"Small-Scale" is like a "Small-Scale Model". You can't see the small details.
It is like a 'small pilot study' — not full scale. **Imagine a very hard to read small map-scale in the distance**. No Details.
"Full-Scale" is "1 to 1" like the Cosmos. "Half-Scale" is "1 to 2". *"Small-Scale" is* "1 to a million" *like Cosmological diagrams.*
The bigger the second number, the smaller the fraction. The bigger the big number, the bigger the span on the map.

Map-Scaling-Factors are Meant to be Interpreted

1:100,000 means "One unit represents 100,000 units"
Split this into two parts: (a) 100 & (b) 1000 and **use centimetres i.e. 1/100 m** — 1 cm: 100,000 cm
Say (a) 1 cm represents 100 cm i.e. 1 m — **drop off two zeros and call centimetres 'metres'**
And (b) 1 m, from part a, represents 1000 m i.e. 1 km — **drop off another three zeros and call metres 'kilometres'**
So the answer is: "1 cm to 1 km"
1: 1,000,000 1 mm = 1 km I.e. **You can also drop off three (or 6) zeros and call millimetres 'metres' (or km)**
50,000 units makes you calculate '1/50th of a metre' (=2cm), so 1:50,000 must be '2cm to 1km'. Or multiply to 2:100,000.
1: 250,000 Turn this into 4:1,000,000 by multiplying both sides by four. 4mm to a km (since 1mm is 1/1000m)
or (a) ÷100: cm to m (i.e. 4cm to 10,000m no longer cm to cm) (b) ÷10: cm to lots of 10m (c) ÷1000: cm to 10km
the grid should be marked in **10 km squares, of 4 cm each**
Your compass will be marked in centimetres, but perhaps also in special scales e.g. 1 inch to a mile (1: 63,360)

Pace Your Progress — Pre-Dict It, Don't Post-Dict It!

To avoid looking at the map continually, to locate yourself continually, predict the estimated time of arrival **once**, and relax.
We are normally lazy about looking at our watch often, poor at calculating, on the hop, distance covered, and lazy about locating ourselves on the map often. Mental arithmetic is clearly noxious, so we try to avoid it. Working abstractly from time to distance to map, after the event is too hard for us. But we usually only resort to the watch after we begin to doubt our position. (See "the 11:59 syndrome" — we do doubt.)
Something needs to be done about this!

Consequently, the antidote to us refusing to log our distance, is to anticipate the timing instead
i.e. to work more concretely, from map (I can see that) to distance (I can visualise that) to time (I can calculate that).

- Examine the map beforehand — as a first step not a last step. You will have to do the calculation later anyway, on the trot! Why not do it now, calmly.
- Guesstimate the time needed to cover the distance ahead — to the next recognisable milestone. E.g. use 4kph, ¼hr/km
- Look at your watch, as the last step, when you near your goal.

In this way the **landscape 'means' time** to you; "When we reach the bridge, we turn left — and that should be in 25 minutes"
Your guesswork will get better too.
You will find yourself looking at your watch more frequently, with the question: "Should I be there yet?".

Anticipation Is Exciting. Back-Calculation is as exciting as the washing-up.

"Range Squared is: Dee Height" "Dee" is Diameter

i.e. Extreme Range of Visibility = √(D x H) on this side of the horizon. **Don't forget to add the 'beyond the horizon' range**.
Five round-number rules are highlighted. By playing around with some approximations, we can find some rules of thumb…
127 is close to 128 and 125 and 11.3^2; Pi (3.14) is close to √10 (3.16);… I hope these tricks will make the square roots easier

40,000/pi x 40/pi km =	**400** km range from 12.7 km	Notice the two related doubling sequences I have given.
40,000/pi x pi km high =	**200** km range from 3.14km	You can only see twice as far from four times the height
12,500 km x 4/5 km =	**100** km range from 800 m high	125 is five cubed
113^2 x 9^2 / 10^2 =	**100** km range from 810m	34 km from 90 m; 11.2 km from 10 m — 1/3 from 1/9 height
12,500 km x 1/5 km =	**50** km range from 200 m high	"Rule of (lesser) fives: **50 km from 1/5 km**";
12,500 km x 1/20 km =	**25** km range from 50 m high	Multiply or divide H by four to double or halve the range
	12½km range from 12½m	e.g. from the top of a beach

12,700kmx.0127km=**12.7km from 12.7m—**from a ship's deck, to the horizon, when estimating distances across open water.

12,700 km x 1.27 km =	**127** km range from 1.27 km high	1/10 from 1/100. Remember the number 127 must be 400/Pi;
100,000/8 x 10/8 km =	**125** km range from 1.25 km	0.125 is 1/8

12,800 km x 1/2 km =	**80** km range from 500 m high	or "Rule of (greater) fives: **50 miles from 500m**"
12,800 km x $1/2^3$ km =	**40** km range from 125 m high	You can also derive 160 km from 2 km; 800 km from 50 km
40,000/√10 x 0.4/√10 =	**40** km range from 127 m high	= "400km from 12.7 km" = 1/10 range from 1/100 the height.
12,800 km x $1/2^5$ km =	**20** km range from 31 m high	= 200 from 3.14 above.
12,800 km x $1/2^7$ =	**10** km range from 7.8 m high	..if you bother to halve 1000m seven times!
12,800 km x 1/500 =	**5 km range from 2m high**	**From a dinghy/beach**. 1/5 range from 1/25 height cf 50m high.

21,600 nautical ml/pi =	**150** nautical miles from Pi n.ml high	approximately, because 150 squared is 22500.
83^2 n.ml x 1 n.ml. =	83 n.ml. **150km from 1 n.ml**.	60 x 360 / Pi is approximately 83 squared nautical miles
12,300 km x 1 km =	**111 km** range from 1 km height	**1° of latitude, from 1 km high**, approx. "**55** from 1/4" = 250m
111^2 km x 1/9 km	**37 km** range from 111 m high	1/3 of the range from 1/9 of the height

Look Carefully at the Fine Detail on the Map

E.g. a kink in the track, just before a junction, may make it look "not like the map!" until… you re-examine the map…
"Oh yes, it is marked like that after all, **I just didn't look closely enough**"
E.g. that maze of fine blue watercourses marked — don't be a-maze-d by it — look at its detail!
There is usually a magnifying glass on the compass.
For a stronger **magnifying glass**, reverse your binoculars. For an even stronger magnifier, remove the eyepiece and use it.

Look Carefully at the Fine Print

E.g. date of compilation ..1864?
E.g. "adjust all longitudes by + 2 minutes of arc"
E.g. yearly magnetic drift
E.g. Contour interval — 10 m or 100 m?

Some Map Details are Not to Scale!

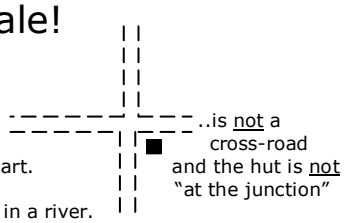

E.g. Two wheel ruts might look like a 100 m wide road on a map (if you took it seriously).
Consequently… ..may be a 100 m Zig-zag! And…

E.g. A "cascade" symbol… ..might equally mean two huge waterfalls, 300 m apart.

E.g. 'Smooth' bends ..probably hide many small extra sinuous ones. E.g. in a river.

Moral: Keep an Open Mind. **Keep your eyes open** for alternative interpretations.

Using the Protractor by Itself

The protractor has two sets of lines — the lines on the movable 0-360 protractor-circle, and the fixed "follow me" arrow, taken together with the edges. Line up one set with horizontal, or North, or whatever; line the other up with the slope or angle to be measured. The angle in degrees clockwise is the space between the zero and the "follow me" arrow, not the other way around.

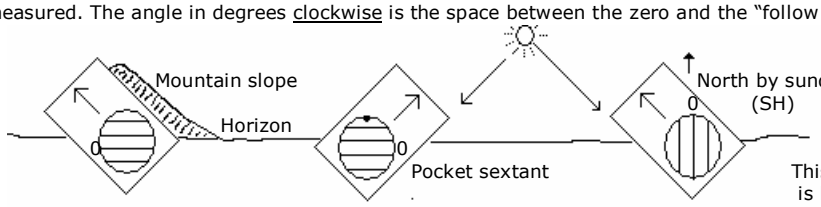

When the angle of the Sun (it moves around at 15°/hr) corresponds with the time of day, you must be facing North/Polaris.
This may be used when the magnetic field is locally anomalous or untrustworthy.
(Tilt the compass axis to Polaris)
Reverse this to get the time from North.

Point the edge of the compass body directly at the Sun until it casts no shadow. (Don't look at the Sun)
If your compass fluid has a bubble in it, it can be used as an artificial horizon.

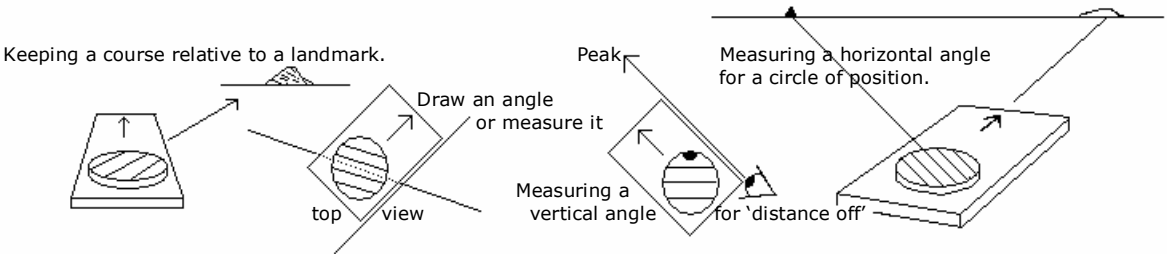

Mental Networks

Our mental maps seem to run on networking — junctions and joins; choices and destinations; routes and 'directions'
Consequently mental maps are prone to logical-type blunders

- Missed junctions e.g. of creeks or paths
- Misidentifications e.g. of segments, features or forks
- Poorly judged measurements of lengths or direction e.g. hairpin bends which aren't noticed
- Mistakes of number

If you reread the section "Logical Blunders" (p.39b) you will notice that most can be seen as a **wrongly imagined network**.
That's one more reason to study the real map, in order to bring your mental map up to speed.
When we function on short term memory of mental branching directions, any shock or daze can be crucially disorienting.

The Most Valued Map

..Is one which you have made yourself. A do-it-yourself job makes it "yours". Practise with a MUDMAP, p46e.

When you copy important detail from other maps onto your map e.g. peak elevations; track notes; waterholes, then your map becomes more valuable than a replacement map.
You add extra details e.g. campsites; walking times; side-tracks & old tracks and suddenly you prefer the old dog-eared version to a brand-new map!
Sometimes it is even better to start with a blank sheet of paper.
But since it does you no good to pinpoint your position on a blank map, unless you are at sea, you proceed to add in the important local details so that you can tell where you are…

To copy a map: **Copy the Grid First**. ..if there is one. Add in many diagonals as guidelines.
The next rule is: **Get the Landmarks in their correct relative position first** **i.e. work top-down.**

The Psychological 'Half-Way Point'

You have to judge and juggle several factors before you decide that you are 'half-way'.

E.g. The "Unknown Territory" Scare-Factor (Ambition versus Safety Margin). In planning your time-budget, you can't relax in a one-way journey through unknown territory, until you know you have a reasonable safety margin in hand So 2/3 the distance along, might be 'half-way through the test'. Other examples...

The "Point of No Return" — You tend to speed up, once committed to the next goal.
The "Downhill Run" — It is the quicker half.
The "Back-Track Return" — tends to be quicker i.e. the turnaround point can be after lunch.
The Morning Spurt-Afternoon Hurt Syndrome — puts 'halfway' before lunch.
The Side-Track Scheduling — if you don't explore side tracks on the way in, when you are fresh, you may not want to tackle them on the return journey.
The "My Feedbag is That Way" Home-run spurt on the last day.

When You Meet Other Walkers

"When did you leave this morning?" = how far to your next camp.
"Anything about the track that we should know?" E.g. slow sections
"Any deep water crossings" or other hold-ups

Other questions, less related to walking progress and pace, are…

Anyone else ahead of us? (Will there be room for us all?)
Do you carry a track guidebook? (Can I read it?) Do you know the forecast?
Any good water available on the way? Can you recommend a camp site?
Did you pass our tents? (Are they still there!)

Pluck up your courage to ask what you want to know, and be prepared to share what you know.

Advice about Walking Times — Double & Halve!

E.g. Brochures seem to add "return" or "one way" by guesswork sometimes.
So "6 hrs" means 3 to 12 hours until proven otherwise.
E.g. Fit young rangers saying "six hours" may mean "for me; double it for you" or "for you; halve it for me"
E.g. "walking time" may mean "excluding stops, rests and meals" — you could find yourself pushed for time.
E.g. "Not far, you can't miss it, it's easy all the way" could prove disastrous. E.g. ('provided you find the right track')

20 Minutes per Kilometre — Double & Halve — Easygoing Maths

It means **3 km per hour** —an easy-going overall walking pace when you include short rests, difficult vegetation (within reason) and lack of any hurry — tourists on track, walkers off track.
Obviously you need to multiply and divide by 20 to convert kilometres to minutes, or minutes to kilometres.

If the number is small — kilometres — multiply it up to the larger number — minutes — **double** it and add a zero.
If the number is big — minutes — divide it down to the smaller number — kilometres — **halve** it and subtract a zero.
You could also express it as a "half a kilometre every 10 minutes" but the "20" makes the maths easy;
And even simpler, to cope with the zero, express yourself in 'lots of 10 minutes'.
E.g. 5½ kilometres will take about… (double it, to get many minutes)… 11 **lots of 10 minutes**
E.g. "We've been going 50 minutes — (five lots of 10 minutes) — ; that's about… (halve it, to get a few km)… 2½ kilometres"

4 km per hour — Normal Walking Pace — Double and Halve!

You can easily double or halve your pace. Start with 4kph as a sustainable rate of progress, and use the doublings 1,2,4,8…
8 km per hour is someone tall walking in a hurry. 2 km per hour is someone small dawdling. 1kph is for very difficult going.

Coincidentally 4 km per hour is 10 miles in four hours or 2½ miles per hour.
Incidentally also, at 4 kph, a landmark 15 km on your left will change direction at 15° per hour as you walk past it.

Double, the Acute Angle, off Your Nose — For the Distance Off

This double-angle method puts numbers on a very "natural" method of gauging distance…
In the first diagrams you walk from the bottom left to the top, and watch a landmark, ahead on the right, pass you by.
The distance walked = the distance off if you walk far enough to double the original angle (away from straight ahead).
The initial angle must be acute. You need to maintain a straight course throughout.
The simplest case uses the diagonal of a square. 45° changes into 90°
The next simplest example is the equilateral triangle. 60° changes to 120°
Followed by the 30° isosceles triangle. 30° changes to 60°

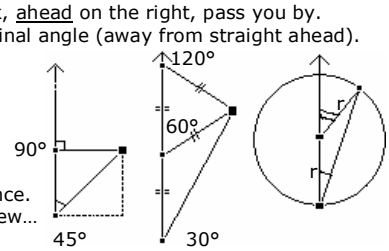

But the general case is made clear by the geometry of two equal radii in a circle.
It is because the angle at the centre of a circle is double the angle at the circumference.
The angle at the centre is the final angle—double the initial one. top view…

Ahead of you, the angle is always less than 90°.
In the second diagrams, below, the landmark may not be ahead of you and the initial angle may not be acute...

If you walk backwards, from top to bottom, using the top method in reverse, you get...

Halve, Any Angle, from your Stern—For Distance Off,

..i.e. In Retrospect The initial distance off, not the final distance off is what must be travelled.
The initial angle is now the double angle, and need not be acute.
I.e. turn the bow-angle diagrams upside-down…
E.g. 90° becomes 45°; or 120° becomes 60° becomes 30°…

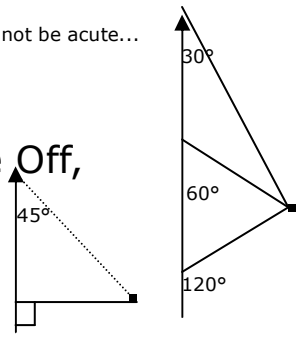

Permanent Landscape is More Trustworthy Than Surface Detail!

When what you see "doesn't match what is on the map" have the sense to distinguish between 100 m deep contours which would take an atom bomb to shift, and walking-track junctions which a fire could reshape.
Professional map-makers and satellite photos rarely make mistakes — in landscape positioning.
But when they add the local details, it is another story.
Time changes such details also. New roads, cleared bushland, second bridges — these will mislead and confuse you.
When your mind says: "The map is wrong!" remember that what the map says — about landscape — is <u>not</u> wrong.

Details Added on to a Map <u>Are</u> Prone to Error

Notorious for errors are: Map North; Map Scale; Roads.
So beware of anything other than a good-quality official government map.

Expect The Map To Be Unhelpful

Many important details will not be shown on the map — you just have to expect that.
You can often trust what is on the map; You can often trust that what you want, won't be on the map.

Maps Do Give You 10 Times More Detail Than You Can Notice

<u>After</u> you get lost and finally sort it all out on the map, you notice… "Oh yes, silly me, those contours are **up**slope not downslope! And they **do** show a creek-line, by those sharp bends! Oh, of course, there is **a peak** on the right, and that fenceline opposite — I missed both of those when I first studied the map. You **can** see Mount Bruce, lined up with the lake, from there, but not at all from over there. And that squiggle in the blue line, is the hair-pin bend that tricked me, but I should have realised that all rivers must drain to the north-west eventually…"

Try to reduce your ten-fold folly to five-fold folly for next time!

Using the Needle Alone — Emergency Magnetic Compass

- Magnetise your repair-kit sewing needle; It will probably already be magnetised somewhat. Stroke it with a South Pole toward the sharp end, so that the North-seeking pole is at the pointy end.

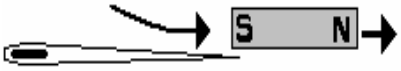

- Float the needle in calm water, on the surface tension (no detergent!)
  (a) lay the needle on toilet paper
  (b) lay both on a bowl of clean water

  The paper will wet and sink, leaving the needle afloat. You could also try sticking it through a piece of foam.

- Hang any magnet horizontally, from hair or fine thread. It helps if you know which end is North-seeking.
  In the field, test it against a Natural North, eg, Bush Noon. At home, test it on a compass. Like poles repel…

Hint Understand that **magnetic poles should be called 'North-Seeking' and 'South-Seeking'** and that **our "North Pole" is a South-seeking pole!**
(so that North-Pole <u>seeking</u> poles point to it — <u>Un</u>like poles attract)

The Earth's internal magnet is as shown (It <u>is</u> slightly off-line with the rotation axis!)

A compass needle on the Earth's surface has a red end labeled 'N' i.e. North-Seeking

The white end is labeled 'S'

Strip Map Navigation — Follow-Your-Nose Navigation

Our mind definitely prefers a strip map — a <u>sequence</u> of "directions" to follow. (None-the-less, make a MUDMAP in conjunction)
Our eyes definitely prefer a strip map — "up" is "straight ahead"! (See "N.O.R.West/N.O.T.Least" "Nose Navigation" p31e)
A strip-map is simply a way of <u>formalising your track notes</u>—every navigator really ought to do it this way…
There are two good ways to make a strip map as you go: with a sheet of graph paper, working from bottom to top, and
in a track notebook, working from top to bottom, but…
Both assume that **progress arrows are up-page**. You could turn your notebook side-on or upside-down to write in it.
In either case you set aside several **vertical** lines for each **type** of information,
and set aside one **horizontal** line for each **subsection** of the journey,
and a second line for descriptions and for <u>comments</u>,
and a line for <u>overall information</u> e.g. trend, drainage, season; totals: time, climbing and distance.
The trick is to use symbols. One vertical line for one sort of symbol e.g. compass needle direction
(See the **Symbol appendix** p 122) Another line for another sort of symbol e.g. slope-direction, and steepness
This way you can cram a lot of information into a small space, in a consistent and regulated manner, and be able to <u>read</u> it!
e.g. …

track feature	compass needle	up slope	landmark direction	time check	leg time	leg distance	distance check	comments
Climb			Lake	1:20			6km	Peak by lunchtime
				1pm	20m	1km	5	cliffy
			Peak	12½	30m	1km	4	trend right after creek
				10½		½km		munch stop overlooking creek

e.g... Saddle
Crest
Stop
Turn
End Bend
Start Bend

AIMING OFF by 10° — 10° is **175m per kilometre**

Distance

1° is one in 57.3, so one in 5.73 is about **10°**, (and 1 / 5.73 = .175) or **175 in 1000**. (For bigger angles use the 1 in 60 rule)
Alternatively, 2sin5° = 0.174 isosceles chord, yielding "0.174 per one" or 174 per 1000. (167m per 1000m / degree for >10°)
This is easier than working in sixes using "one in 60". **Hint: Express any ratio as "something in 1" or "in 10, 100, 1000"**
E.g. '1 in 50' becomes '2 **in 100**' or '200 m in 10 km' (.4 in 20 etc)

Example: "The T-junction is somewhere ahead — about 017°, 5 km. But so that we can't miss it, we must aim off to the right e.g. 027°. We therefore expect to hit the road 875m to the right of the Junction" (5 lots of 175m).
This is how ocean navigators find an island in an ocean: Step 1: sail N/S to the right latitude. Step 2: sail E/W to the right long. Don't just head straight toward something. If it doesn't show, you won't know which way to turn!
Remember that a 10° error is fairly easy to make when using a compass over rough terrain (or trackless shifting ocean)!

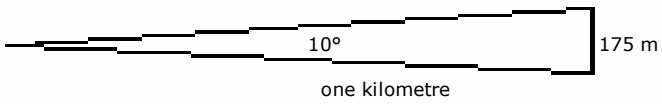

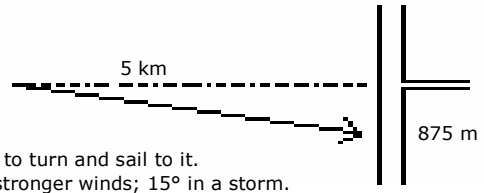

Aiming-off at Sea…

Aim-off to arrive up-current & up-wind of your destination—to make it easy to turn and sail to it.

- Aim-off 6° to counteract leeway when sailing into the wind; 10° in stronger winds; 15° in a storm.
  Measure the appropriate angle from the wake e.g. 6° is how much your pointing-finger-tip shifts when you change eyes
- Aim-off 6° (1 in 10) to counteract a 10% (1 in 10) cross-current. This refers to the perpendicular component, compared to your speed. The rule works +/–1° up to 36°—up to 60% cross-current. To avoid judging components...
- Aim-off 1°,2°,3°,4°,5°,6°, to counteract a 10% current at 10°,20°,30°,45°,60°,90° to course, either ahead or behind.
- Aim-off to avoid GPS-induced collisions caused by very accurate course-holding. You should plot a course off to one side of the obvious line on the chart, to avoid vessels coming the other way on the same line!

Aim Off 6° or 10%

Blink your eyes at the destination, and when you get there, pace off the 6° offset. Ten-tuple the result for the distance there.

---

1 Metre per Second is 3.6 kph

This is merely a logical reference point for your calculations. (And slightly faster than "20 minutes/kilometre" (3 km per hour))
60 minutes x 60 secs per min is **3600 seconds in one hour**, at 1 m per second, is 3600 m in one hour, or 3.6 km per hour.

One metre is Up to your belly button A child's double pace One strong fast stride

One second is the time it takes to pronounce "Alligator One" in a normal fast voice—don't let it drag. Follow it with "Parramatta two", "Oodnadatta three", (or "a thousand and four", but "alligator" keeps its speed more consistently when drowsy).
Get used to the rhythm of tapping your finger in time with the seconds — your mind subconsciously remembers the timing.
Take a Watch, Perhaps a Stopwatch
1 m per second is very nearly 2 knots.

"1 Hour Extra per 5 Hours"

You can normally walk at 5 km per hour, on a track, on a short walk.
But you will sustain an average 4 km per hour, when you include fatigue, and rest stops.

"1 Hour Extra per 500m easy climbing"

Gentle up-climbing is tiring—allow this extra time. (We are not concerned here with how long it takes to rock-climb!)
Down-slope walking may speed you up, or slow you down, depending on the slope.
Dense tussock country (e.g. snow grass; button grass) is the equivalent of a climb, in its up-down-up steps. Crossing the drainage can also be as tiring as climbing, in its regular climb-descend-climb-descend path.

Distance by Optical Rangefinder — Accurate up to a Kilometre or Three

We are woeful at guessing distance in that range, so there are focussing devices available which are designed for it.

Range by Parallax and a Piece of String — DIY Rangefinding

Animals shift their heads to see how much an object shifts. Blink you eyes to see the effect; walk sideways to practice it.
For more distant objects, the very slight sideways offset can be magnified then measured...
The landmark you are measuring appears in a slightly different direction from each of two different viewpoints.
Pack a long non-stretch string e.g. a polymer kite-string. Know its length. Mark it every 5 m before you go.
Use it as a giant pair of eyes—a long baseline—& to mark out a large square or rectangle on the ground, one side at a time.
Guess at the right-angles — they don't need to be accurate. Just make sure to use equal tension when you stretch the string.
But use a plumb-bob to mark the corners of the square on the ground accurately — with a pebble.
Use the plumb-line also for accurate sighting of the landmark. (You can pack up the string once you have placed the pebbles.)
Twice as big = four times as accurate (in measuring d) If you have enough room to move, aim for over 10 m sides.
Range = ab/d or a^2/d if a = b; from the similar triangles involved. Measure the shortfall, d, at the front, or the overfall, d, at the baseline.

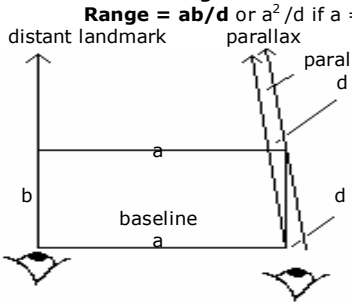

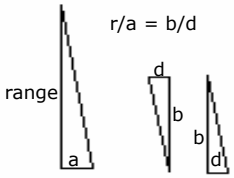

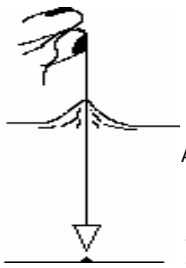

An easy way to remember the formula is to say that the area R d = the area a b because if you shear the ab rectangle, it becomes a long thin parallelogram R long, and d wide, preserving its area.

As a result, the formula becomes...
Range = Area / parallax offset, d
km sq. m mm

The Range=ab/d formula also *makes sense*
The range *should* so depend on a & b and d

"Maps Grip the Universe"

When you swing the map in your hands, around with you as you turn, the whole Universe doesn't follow it! — it doesn't twist!
Instead your map gets disoriented.
Leave the map behind, as you twist! — leave it lined up.
The map is designed to lock into the landscape, to 'map onto' it, and should be used like that, not twisted out of alignment.

You need to swing the map an *equal and opposite* amount to neutralise what you do.
It's called "swinging the map" (when you leave it locked onto true North while you swing to left and to right).

It should not be called that!

It should be called "**NOT Swinging the Map** with you"

Map work includes 'not swinging the map'. See "Magnets Grip Maps". The most common method is to match the map to a compass needle p.103b

Maps mean landmarks — those landmarks there! Open your eyes — and line the map up with the visible landscape. p101a.
i.e. orient the picture of the land — the map — to map onto the actual land.
Line the Grid North up with North e.g. by compass or by Sun.

You may have to twist yourself, or put 'North' at the bottom left, but to always put 'North' up the top, is to divorce your eyes!
Yes, you may have to read names upside down! Yes, it is a bother, to keep changing its 'up' every time you turn.
But it means that you can just lift your eyes from the map to see what it points out. And can know what lies beyond that.
If your map is not consciously oriented, how can you be oriented, to other than what you can see in front of your nose?

Think "+/- 5°" For Compass Bearings; "+/- 10°" For Following a Compass; "+/- 15°" For Using a Compass!

Compass bearings "+/– 5°"
Test yourselves: Ask each person in turn to nominate where "magnetic North" is on the skyline
using one particular compass, handheld. **My results:+/- 5° "User Inaccuracy"**
Now try putting that compass on the ground. **My results: +/- 3° "Needle Instability, friction" etc**
Try stretching a fishing line over the needle. **My results: +/- 1° "We still can't agree!"**
Next try that best method for different compasses. **My results: +/- 3° "Calibration and Zero Errors"**
Now try **not** being very careful, with an **untested** compass, at night, when you are tired!

Following a compass blind: Walk on a fixed compass bearing for 100 m or so in a fog i.e. eyes glued to the compass.
Repeat it e.g. turn around & see where you come back to. Give the next person a try at doing better.
I think it matters which eye you use and which hand you hold the compass in.
We don't necessarily walk 'straight ahead' when we try. We don't necessarily point the 'follow me' arrow straight ahead.
We don't do very well at integrating and averaging out any diversions around obstacles.
Slopes play havoc with your sense of a 'straight' line ahead.

Try this…

The front person finds the route.
The back person guides the route-finder.
The navigator logs the route directions taken.

Using a compass: You may plot a single line on the map, or as +/- 5°, but mentally allow an extra +/- 10°, i.e. think +/- 15° (for accumulated errors, possible blunders about magnetic variation, etc.). Keep the possibility of error in the back of your mind

---

Caving Strip-Maps — Come in Books

A "cave-directory" booklet records **each separate section of a cave on a separate page**.
E.g. Make your own…

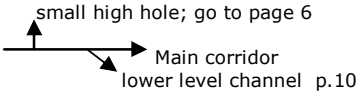

You may need separate booklets for different levels of the cave system.

Use the same symbols as before, for needle, slope, turns, etc.

Use Abbreviations: E1, E2, E3… for Exits; L1, L2, L3… for Light sources; Parentheses (L1) = "Can see L3 from here"

Make up names for each section e.g. "Chamber of Horrors"

Add copious notes e.g. "water level up to ½m deep"

Forethought and cooperation is required to quickly and effectively map a warren or a labyrinth.
Initially you make a muddy map, as you go, but then you add a neat copy into the book later.
(But keep your muddy copies, for next time.)

Distance

Up-down Progress

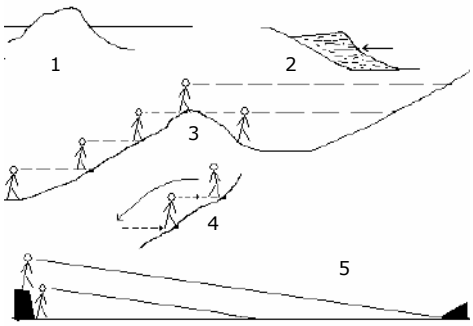

In the first diagram you can see that you are halfway up the mountain alongside of you — **use the horizon to gauge your altitude**.
Note: the horizon drops below true horizontal as you go higher.
In the second diagram your shadow is obviously halfway up the shadow of the hill you are on — **use the shadow of the mountainside.**
To find your tiny shadow:
Close one eye — to line up more accurately.
Look past the shadow of your head falling onto your hand
Look for a 'halo' — a bright spot directly down-Sun on the plain
Wave your raincoat around
Well, you know where the shadow should be by now!
So you don't even need to see your shadow!

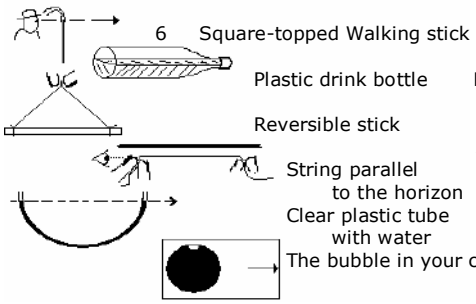

Keep count of height in body-lengths
(3) Walk up to a stone which was horizontal from you before.
(4) Downhill can be counted by leapfrogging with two people.

Half-Way across? Provided you stay in one position...
Looking parallel from half the height gives you half the distance
(5) E.g. when someone is swimming across a lake

Artificial horizons allow you to maintain parallel gaze (6)
E.g. "The edge of the lake is 'two fingers low'

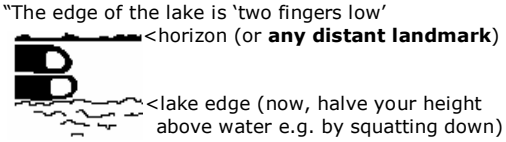

Adjust Your Expected Pace

! Three hours up, two hours down — are typical proportions, except in very steep or loose descents — 6 hours up, 6 hrs down.
! You walk more briskly while fresh, in the morning, than after a whole morning's walk. Pace yourself with this in mind.
! Groups travel slower. Allow the slowest person to set the pace, because stragglers may battle or even panic if left behind.
! Retracing your steps is quicker, since it has fewer diversions, less interest, and no unknowns. Allocate less time to return.
! Carrying packs is not necessarily slower, on flats. If your pack is uncomfortable, you won't dawdle, or even want to sit down, and can't stop to look at insects or flowers or to chase birds.
! River crossings chew up enormous amounts of "fussing around" time — organising and reorganising afterwards.
! Beware of smoothed out "routes" marked on maps. The actual terrain may involve many obstacles and consequent slow going and detours. The words "route only" are trying to tell you that.
! Your pace quickens surprisingly when the party falls "into the groove" of walking (without talking, noticing things or relaxing often).
! Path finding can make you nervous especially in unfamiliar territory, which tends to quicken your pace without you knowing it. The same person, following a leader, is relaxed — and slows down.
So when you take the lead and you wonder why the others seem to fall behind, realise that you are too keen.

Don't Overcompensate for Slopes And Meanders

The worst case for slopes is about 36° for rock scree. But that adds only a quarter to the distance shown on the map. It is far more relevant to say "It's the climbing and the awkwardness which will slow us down more".
The worst case for obstacles is grass-tussock negotiation — but that only adds a maximum of 50% to your distance covered.
An ordinary zigzag may only add 10%

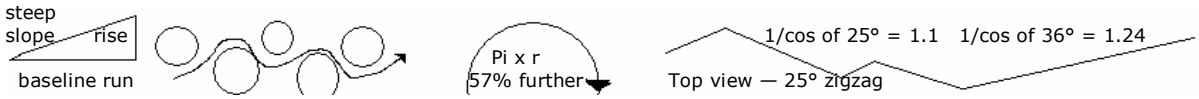

Distance by averaging Maximum & Minimum Guesses

Yes it is worth doing, because it reduces the slop in your individual guesses.
E.g. "More than a kilometre, for sure. Less than 3 km surely! Split the difference & say 2 km

Distance by Comparison with Some Length You Know

E.g. The central business district—1 square mile or whatever
E.g. old-fashioned 'chains' i.e. cricket pitch lengths
E.g. 100 m race; playing fields
E.g. "a day's walk"
E.g. that island which you can see, which the map says is 1 km long.

Distance by Cloud Shadows — Look for Them

- Notice how quickly the cloud shadows move here (use distance/time), then look how slowly they are moving over there
- Time how long they take to get there.
- See how large they look here, and how large there.
- Look how high the cloud base is here, then compare the size of hills in relation to that ground-to-cloud gap, over there.

Hint: To see which shadow attaches to which cloud, hold a string, to pass from Sun to shadow (past its cloud) but… Don't look directly at the Sun!

Sound Travels Three Seconds to the Kilometre

This tells you how close the lightning storm is. "(Flash), Alligator One, Alligator Two, Alligator Three, Boom!, One Kilometre!"
If you see a flash to ground, you can find the distance there.
Otherwise, the distance is to the nearest part of the lightning flash overhead.

Align the Track

Line the map up with what you see! Often that is the road ahead, or a well-marked track.
This is a great trick for the car navigator when using a road map — it's so easy to see the line!
It's equally helpful when following any marked track on the map with the map aligned to the landscape — it locates you on the map — on that bit of the map which lines up with the track.

If you were landing a fighter jet onto an aircraft carrier in the fog, you would not want to miss the runway!
Well, think of the real track as the runway on the aircraft carrier, the land as the sea around it, & the tiny map-picture of the track as the jet plane.
Twist and turn your 'jet plane' until it can come in to land properly. Even when you are in the fog!

Assuming you have a properly oriented map…

You can find where you are on it… me!

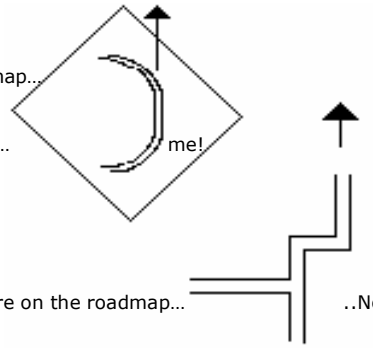

If you know where you are on the roadmap… ..Now the map is properly oriented.

Compass Errors — Be Impressed! Be Careful!

Electrical currents swing the needle e.g. reading the compass by torchlight; untwisted pairs of DC wires e.g. marine compass lights; Mobile phones; GPS; radio. "Impressed (DC) current" anti-corrosion systems in marinas.
Magnets e.g. other compasses, motors, loudspeakers, clocks, ironstone. Note: Iron-rich dolerite doesn't look rusty.
Iron & Steel e.g. your pack, pocket knife, keys, wrist watch, cutlery, steel-capped boots, glasses, geology hammer, batteries, torch, studs, zips, pens, trig points, shed, fence, car, ironstone, buried machinery, coils of old fencing wire half-buried in the ground, tools, fuel stove, internal springs, reinforced concrete, railway lines, power pylons, pipelines. Navigating in a mineshaft. Many of these might have permanent magnetism too. Stainless steel is often magnetic.
User error
Parallax — viewing the needle from one side; viewing the scale from side-on.
Misreading — e.g. any scale marked every 2° is easily confused. Mis-setting the movable scale.
Miscalculations — especially applying the variation the wrong way — adding it instead of subtracting. Confusing left and right.
Out of date chart of deviation. Deviation is the general name for any compass error—apart from expected, normal variation.
Quality of Construction
Needle instability; under-damping; over-damping; motion. Let it settle. Put it on the ground.
Vibration — has unpredictable effects e.g. twisting the needle clockwise. Walking provides regular slow vibration! Motors.
Sticky pivot — Tap the compass Un-centred pivot
Zero-error — the needle doesn't match the 'zero' Mis-calibration of the protractor
Needle mis-magnetisation — not quite longitudinal; amateur attempts at re-magnetisation.
Reversed polarity — Poor quality needles can pick up remagnetisation from nearby strong magnets without you knowing it!
Variation The difference between True North and Magnetic North changes with time and place. The map is out of date.
Magnetic drift of the magnetic pole — we only have 'predictions' or 'trends' of an unpredictable variation. Are you up to date?
Latitude change — alters the magnetisation of surrounding fixed iron e.g. a car body or a boat's motor or hull.
Local anomalies e.g. Tasmania is full of ironstone dolerite. Many rocks preserve ancient magnetism.
External magnetic fields e.g. which way a boat moors, or who moors next to it. Competing fields change or capture each other.
Lightning strike after-affects.
Correcting mechanism out of **adjustment**.

Map Folding and Map Holding

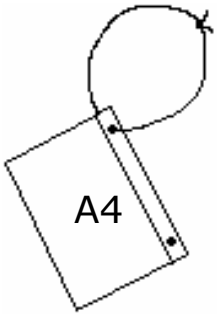

Maps aren't forever — they blow over waterfalls, tear & get unreadable.
Pack a spare
Cover the whole surface with (non-shrinking) clear plastic film
With good-quality laminating plastic & cutting, taping and sealing, maybe it will last forever!

Folds and Corners lose details, and wear through.
Cut along the fold lines cleanly e.g. to A4 size

Use a photocopy of your own map to use in the field, to keep your original map in good condition.
Wet maps lose details when rubbed.
Get an A4 plastic sheet protector for the map. Tape it closed. Hang it around your neck.

Mapwork requires bearing-lines to cross the whole map.
Tape the (A4 or other) sections back together with wide clear tape, leaving wide gaps, as shown.
— you can butt the edges closely when needed; the tape will crease out of the way
— you can refold it at will, as often as you like, and show the two most useful sides out of the many sections.

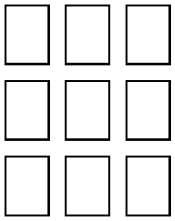

large gaps

Distance by Double Pace — Test Yours

A 'mile' (or "thousand" in Latin) is a thousand double-paces, on forced march i.e. 5.28 feet (Romans were small people)
My dawdle double-pace is 5.25 feet. What is yours? In metres?
Answer: Test it. Test it again and again. Average it. Each person is different. Relax, don't push yourself.
Test it again uphill, downhill, laden, unladen, tired, in a hurry.
Test it at one double-pace per second — walking by the clock.
It helps to make a 10 m cord. Use a tape measure to make the cord, then use the cord.
E.g. I know I use 12.5 **paces (single steps) per 10 m** i.e. alternating 12 & 13 double places per 20 m.
and I walk at 4.5 kph; 1 1/4 metres per second; eight **seconds per 10 m.** The timing varies more easily than the distance
While watching the seconds tick over on my watch, and striding out in time with them, I manage 1.5m/sec, get tired, and step further out.
You Can Be As Accurate As You Have Been Careful. If you are full-grown, working this out once can set you up for life.

Tie Your Ankles Together!

For a well-measured pace (heel to heel) use a 50 cm cord, for instance.
Count double-paces, and don't lose count — see 'finger math'

Test your pedometer

Pedometers work in different ways, so test yours on different terrain. Your knee-lift is higher on uphill sections, and your pace is shorter in rocky terrain. You can make your own pedometer from an old odometer, working off the flex of your boot, your hip, your walking stick, or whatever, or you can roll a small bicycle wheel along the ground, fitted with an odometer.

Don't Lose Count! Use Finger-Math

Finger counting is a silent and **number-free** counting method — you only pronounce the number after you have finished counting. Count automatically, by placing your fingers down, each in turn, onto a hard surface.
A thumb counts for five fingers, and the **left-hand counts 10 per finger**.
When you run out of digits you have 99, so you next say aloud "A hundred" and start over again with no fingers down.
Place each finger down onto a surface, in a **definite sequence**: Right Index ("1"); Middle (two fingers down); Ring (3); Little (4); Lift all four and put down the thumb ("5"); Leave the thumb down and put and leave down the index finger (6); middle (7); ring (8); little (9); Lift all five and put down the left index finger ("10"); Then put down the right index finger again (11), and so on, until you need to put down the left middle finger in place of the other five (two 'tens' down).
With practice, you can count, add, subtract, multiply and divide — almost thoughtlessly — as quickly as with an abacus.
We are concerned here with **counting single or double paces**, e.g. in a fog, and doing so **in lots of 10m** (or 20m or 40m for double paces). I.e. you count up to about 12 or 25 single or double paces mentally, and add "one more lot" to your finger tally.
If you are in a group, you can check each other's counting by stopping where you think every hundred metres is.

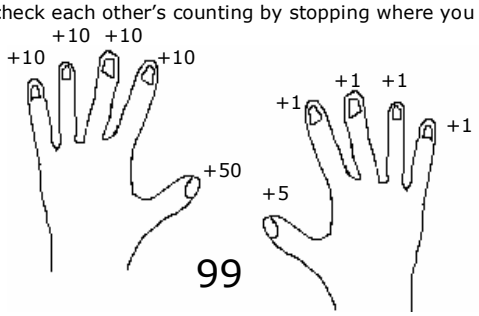

Lift all your fingers off

and nod your head down

for '100', then start again.

100 lots of 10m equals 1km
100 lots of 20 m equals 2 km
100 lots of 40m equals 4 km

or 100 double paces is 1/10 of your mile

Lighting Conditions & Eyesight

Make allowances for optical effects…
Looking up-Sun, especially in the afternoon, you see haze (backlit dust & insects) which gives a false impression of 'distance'.
Looking down-Sun on a clear morning after rain or after a cold change (i.e. minimal haze) things 'look close enough to touch'.
As your eyesight begins to fail in mid-life, everything seems 'distant' but turns out to be surprisingly close
i.e. you have been judging distance by crispness of detail all your life…

Visual Acuity is about One Minute of Arc

..which means that you can spot a 1 m object at 3.6 km on a clear day (since one minute of arc is (1/60)° or '1 in 3600').
I.e. You can see details several thousand times smaller than the distance to them.

This is how seamen can judge distances over open water, to land, even though there are no other clues to go by.
They ask: "Can I see trees? What about tree trunks? Or branches? Or leaves!"

Lighting, Contrast, Haze and Eyesight play havoc with accuracy; But experience is a great teacher.
Feed your mind and memory by deliberately looking through the haze to the finest details you can pick up.
Use binoculars as a cross-check. Have competitions with your family.
Children have better eyesight, but the idea is to judge the distance according to how it looks to you.

Align the Map to the Landmarks

Make the Map Match the Marks — if you know where you are on the map you need only one landmark…

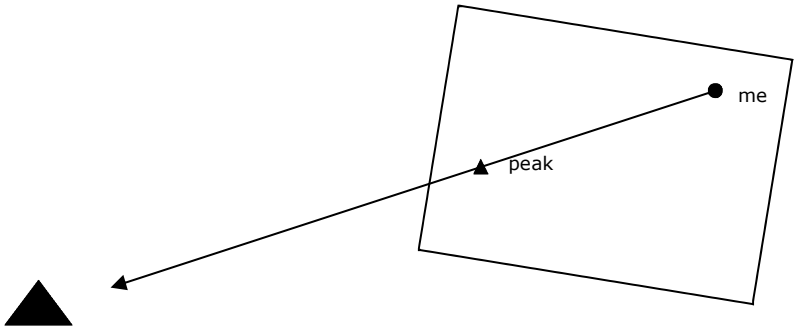

The easiest way to keep the map aligned to the landscape is by then using the direction of the Sun now, a very visual method.
The commonest method uses a magnetic compass to find North.
This one-landmark method can be used whenever you can see something in the distance.

If you don't know where you are on the map you can still orient it to match the landscape, but you will need two or more landmarks identified to do this.

Compass Use — Master It, but Be a Good Master.

- Use it to orient yourself, then put it away!
- Use it to orient the map or to choose a route, then put it away!
- Use it as a protractor. Use it as a ruler. Use it as a map scale measurer. Use it as an artificial horizon. As a set-square.
- Use the set square for a back-course or a side-course without altering the forward-set-course.
- Don't rely on it — use other methods of direction-holding.
- Let it point your eyes outwards, to the horizon.
- Use NORWEST/NOTLEAST, if you have to stare at it — e.g. in a cave.
- Hold it 1½m above rocks; 10m away from cars, sheds, etc; several item 'lengths' away from small items — test it.
- Let it point out a foreground mark as a destination, then follow a visual line of sight
  e.g. 'leapfrog' or manoeuvre by 'legs'
- Don't consult it again until you need to check or choose another point.
- Don't trust it, but trust yourself to it, when you need to.
- CHECK ITS ERRORS. Check it against the map by taking a bearing from a known point to a distant landmark
- Take a pair of forward and reverse bearings—to and from a chosen mark, to check for local magnetic field anomalies.
- Check your errors in using it, e.g. on the move; against someone else.
- Interpret it for orientation value, not just for route-finding.
- Consult it more often than you want to, when you are 'not using it'.
- Check your guesswork as to e.g. 'Gully North' or 'Wind-West' by it.
- Take it.

Layout Map — Blind

- This project requires a pedometer to measure every hundred, metres or paces or something. A stop-watch could do.
- Use graph paper, so that each square on the graph paper represents one 'leg' of 100 (or '1 minute's walking').
  (Hint: Graph paper is a handy thing when making a casual mud-map of your progress, for this very reason.)

Try it at least once. Without triangulation and cross-checks you can build up the layout just by reading the compass.
i.e. Beware: This is a bottom-up method, prone to inaccuracy. After one accurate practice-run, you can do a better casual job.
I achieved over 90% efficiency at coming back to my starting point after a 24 km circuit on my first try!
(In fact I know that the 8% error was mainly only due to a few blunders in marking the wrong direction.
Also, I only measured 'direction' to +/–15° since I was also testing the 'clockface method' of direction — to the nearest ½hr.)

When the compass read "2:30", I needed a reverse clockface on the map, and I needed to mark…

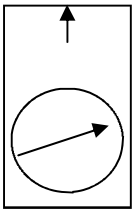

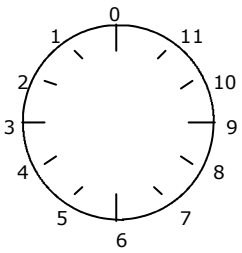

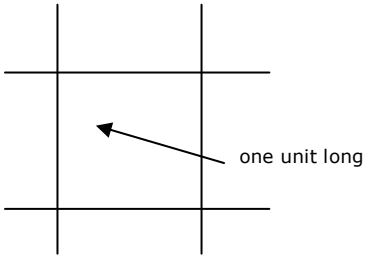

(Hence I made a few left-right blunders which I had no record of!)
You don't need a map for position-sense, if you make your own as you go. Position sense is hard to come by, otherwise.

Line of Sight Can Be Extremely Accurate

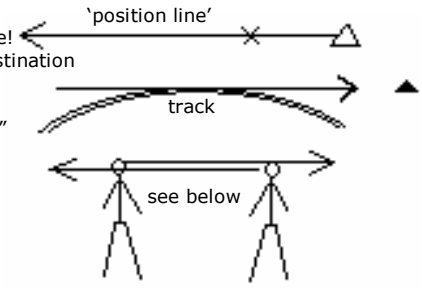

..But is often neglected, for far inferior methods.
E.g. locate yourself on a 'position line' as a means of marking your progress…
E.g. keep a straight line by walking to a point ½ way between you and your next destination
E.g. 100 fence posts can be positioned to look like one when seen end on.
E.g. fishermen can **locate** their offshore fishing spot again
— "The jetty is underneath Mount Lofty, and the tower is behind the Town Hall"
You can sometimes consult a map to fix your position...
E.g. You may notice that **"this part of the track runs directly toward"** a peak.

Un-mappable lines abound. Choose the **nearest** objects in front for best accuracy.
E.g. "That far rock lines up with this bush. **Remember** them!"

Use mapped features and the accuracy is still extraordinary.
E.g. "The peak lines up with the intersection, but if you step only three paces left you can see the difference!"
The alignment doesn't need to be accurate, provided you can memorise the exact inaccuracy e.g. 1° to the left of the hillside.
Two such alignments are needed for a fix. Look around: "What is in front of Mount William? What is behind the lighthouse?"

Two Opposite Landmarks

It **more than doubles** the alignments available to you if you have eyes in the back of your head!
With two people, separate some distance, for accuracy, then **look past each other's heads**.
Walk slowly around in a circle looking at the horizon for possible line-ups.
With one person, **place a rock on the ground then circle it**, scanning the horizon. Place a pebble to mark each feature.
The good thing about this method is that one landmark cannot 'blot out' another — all the visible features are available to you.
You may use this method to stick to a direct 'track' between two landmarks.

A Slow Kilometre Cancels out Three Quick Ones

Let's try to keep up an average of 4 km per hour…
First we travel 2 km in half an hour, but then we hit a slow section and manage only 1 km in the next half-hour (2 kph).
To counteract this we would need to triple our slow speed to 6kph for the next half-hour!
So we then decide to stop and catch our breath for half an hour.
To counteract that we would need to double our desired average speed (to 8kph) for the next ½hr!! Or 6kph for 6km.
i.e. a wasted hour takes several hours of catch-up, to regain.
(2kph for ½ hr + 6kph for ½ hr = 4kph; 0kph for ½ hr + 8kph for ½ hr = 4kph; 0kph for ½hr+ 6kph for 1hr= 6km in 1½hr)

Slow sections chew up the time available, so don't dawdle.
Quick sections can't easily repair the damage, because they don't last long.
You may try to be time-efficient and walk more quickly on the easy sections i.e. hurry along the flats,
but you can't gain much time that way.

To Progress Faster More Easily
Pace Yourself — Slow Down, but Don't Stop

i.e. Pace yourself so that you don't need rest stops, since they ruin your time budget, and it is time which determines pace.
It is better for your overall pace to keep walking slowly than to stop completely.
Go slowly on the hard sections. Pick up your pace, maybe, on long flat easy sections.
Shorten your rest stops e.g. don't remove your packs.
The old rule to cover a lot of ground quickly is: run 50, walk 50 (paces or double paces).
Catch the paradox of it — to keep going faster, go easier.

Approximate Figures for Visual Acuity

At 1:1,000 "I can just count my grandchildren"
At 1:2,000 you can count individual items.
At 1:3,000 you **can't count** them. E.g. people/shrubs/birds
At 1:4,000 you can **see** items, but can't count them.
At 1:5,000 you **can't see** them individually.
At 1:6,000 you can **read unfamiliar letters**
At 1:7,000 you can't **read unfamiliar letters**.
At 1:8,000 you can **read familiar signs**.
At 1:9,000 you can't read familiar signs
At 1:10,000 you cannot see that there are individual letters.
At 1:15,000 you can see single objects e.g. trig points against the sky.

You need to supply figures which work for you.
Multiply the size of the smallest objects you can **count** by 2,000
Multiply the size of the largest objects you **can't** count by 3,000
Multiply the smallest items you can **see** by 4,000.
Multiply the largest details you **can't** see by 5,000.
Multiply the letter-stroke widths you can **read** by 6,000.
Multiply the letter-stroke widths you **can't** read by 7,000.

Of course it's rough! I told you so in Rule 1: Guess!
Adjust the numbers to your eyesight e.g. "I can only count to 1500".
The idea is to build up your subconscious guesswork.
E.g. "I **can count** to 3000 and can count house windows" try 1 m x 3,000 = 3 km away
"I **can't quite see** heads" try 25 cm x 4,000 = 1 km away

Test Your Own Visual Acuity

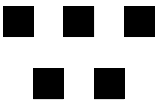

Walk up to a signpost until you can read it. Pace the remaining distance.
Measure the size of its letter strokes.
When do you lose sight of a dollar coin?! Can you count heads in a crowd at 600 m?
Set up a 1 cm test pattern of squares (for others) to count. Try yellow squares too.
Test yourself in different conditions — on an overcast day, in direct sunlight, in glare, on a hazy day, and at a large distance.

Align the Map to Identify Landmarks

Lift your eyes from the properly aligned map back onto the landscape… "There should be a peak in the distance over that way." "Look for an outstation in this direction" "Can you see the transmitter structure??"

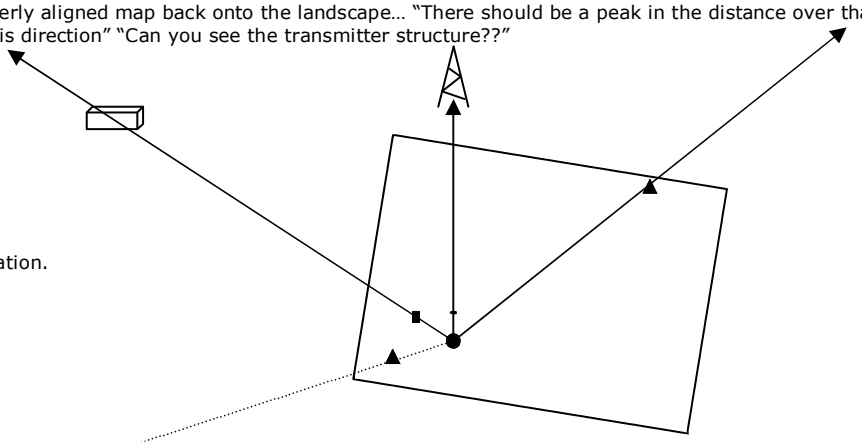

This is "Map To Land" navigation.

If you are on a peak, you may want to keep a copy of these relative directions in a notebook (especially the ones which go off your map).

"Magnets Grip Maps"

This mnemonic is meant to remind you to twist the map to the north — the map should be one piece with the compass needle.

- Place the compass on the map
- **Grip both the map and the compass in one hand**
- Turn the whole map, or yourself, or both — so that the magnetic needle points to the map's magnetic North

The result is that the map is aligned to the landscape.
Keep it that way!
Even while you walk, and twist, and turn… This is called "Swinging the Map".
A one-off effort is called "Orienting the Map".
Swinging all three of you is called "Orienting yourself" e.g. to North.

The Magnetic North marked on the map is crying out to be matched up with…

Align Your Map in the Dark

..the magnetic needle but check that the angle is drawn to scale!

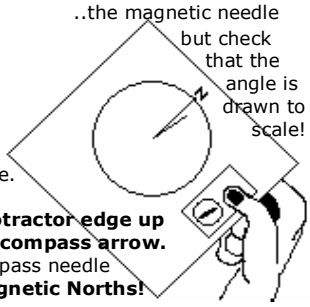

The commonest way to align the map to the landscape is by a magnetic compass.
The magnet — the compass needle — must 'grip the map', magnetic North to magnetic North, in order to keep the map oriented.
One method is to set the compass protractor to zero — true North — and
set the compass edge to grid North/true North (see diagram) and then
set the magnet needle to magnetic North — as marked on **the compass rose on the map**
Another method is to line the edge of the compass up with the magnetic North mark on the rose.
Then you simply keep the needle point pointing to zero on the protractor.
A third method is to set the compass to the correct compass heading then line the protractor edge up with the desired course on the map, and the needle to the protractor, then follow the compass arrow.
A fourth method is to set the compass to the correct true course on the map, then line the compass needle up with magnetic North on the map, then follow the arrow. But **always line up the two magnetic Norths!**

A Compass Holder

Rather than hanging the compass around your neck, where you hardly ever see it, you can make it constantly visible.
For map making, it is indispensable.

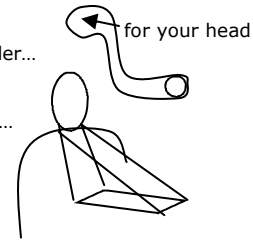

Version 1: Aluminium wire around your neck and around the compass, like a bib or harmonica holder…
The compass circle is grasped by a circle of non-magnetic wire.

Version 2: Hang a flat map board from your neck, and blue-tack the compass on to the map board…
Drill holes in the four corners of a light board.
One string loop goes around your neck and through the two close corners.
Another string loop goes around your neck and through the other two corners and has an adjusting knot in it (rolling hitch).
The compass protractor is set to 0° i.e. Use Magnetic North
or it is set to x° i.e. Use True North
To orient the map, or yourself, turn around and face North.
For a back-bearing, turn around and use the white end of the needle to read your direction — where you have come from.
Use the "NORWEST; NOT LEAST" mnemonic to keep you in touch with normal orientation. See p.101c & 106b for how to build up a map, or keep track, blind, using such a compass holder.

Distance

"Changeover Lines" of Progress — Relative Apparent Sizes

All visual elements which surround you have an apparent size, each of which changes as you progress.
E.g. "The hut behind us seems so tiny now, poking out from the sea of snow"
E.g. "That hill now seems to loom **high** above the plain" (a vertical angle).
Two apparent sizes will, at some stage, become equal, at e.g. as you move away from one object, which shrinks, and toward another, which expands. Simply notice "the changeover" and you have 'a line of position' to log your progress.
"The hill is the **same size as** the leaning over gum tree" (a vertical angle, a diagonal angle, and a changeover line)
You don't need to remember what size you measured — just compare them for equality!
"The width of the patch of forest is much less than the island is **across**" (horizontal angles)
"But the island alongside us is about the **same width** as the **gap** between the two" (another "changeover point")
You don't get a single point fix, only a line of separation. The third and fourth diagrams each show an example.
A measurement without numbers is like saying "We are halfway there" instead of saying "4.5 km on"
or like waiting until you are "Exactly East of Mount Big" rather than measuring "079°C" at any time
or somewhat like "These two line up, if you put a hand span in the gap"
The closer the objects, the better. The more distant, the more careful you need to be!

Use the Landscape Itself to Measure & Mark Progress

Example: "Where I came out onto the beach,
Sloop Rock was one island to the right of Gannett Island"
A top view shows that this is simply a line of sight, locating a point next to the island.

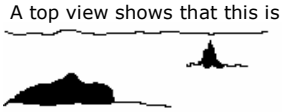

Island = gap;
An apparent-size
changeover line

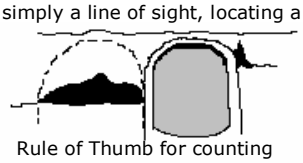

Rule of Thumb for counting

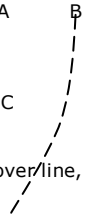

horizontal angles: AB>BC AB<BC

changeover line, AB=BC

°

Distance-Off by Vertical Angle

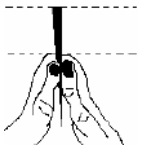

The simplest range finder is a piece of grass held at arm's stretch, (if your fingernails are too small).
Measure the apparent height of something which you can recognise later,
then measure how many millimetres of grass was needed!
You can relocate that distance-off later by its apparent height e.g. distance from a prominent peak;
from a lighthouse seen from offshore; from any recognisable tree; or from a cliff with a clear base to measure up from.
This is how our mind analyses the skyline heights — it is a very 'natural' method.
Yacht navigators will find out the height of a lighthouse from a chart,
measure the angle with a sextant, and calculate the range by trigonometry…

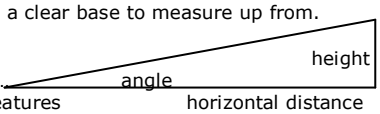

To fix your position requires two separate measurements e.g. two different vertical features
In theory, the plotting looks like two circular arcs intersecting, because any distance-off must give you a circle of possible positions.
The closer the features, the smaller the circles, and the more accurate the fixing of the location.
Confuse your friends with: "1 inch away from the windmill, and 13 mm from the pine tree"…

Safe Distance Off

You might be sailing around a headland or reef, and want to maintain safety…
Without needing to know where you are you can maintain a minimum distance off.
E.g. "no higher than 1° apparent height by vertical angle"
E.g. "keep the horizontal angle acute, between those two points"
E.g. "stay outside the apparent radar range of 5 km"
E.g. "deeper than 20 m"

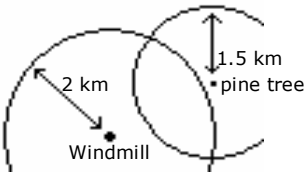

Discard one of the two possible positions.

Circle-of-Position by Horizontal Angle

Here you keep track of the horizontal space between two prominent points; or any noticeable gap; or the apparent width of an island. Use a string stretched across your line of sight, or a stick gripped between two fists.

In this horizontal case though, unlike the vertical angle, the 'circle' of position is really only a cusp — part of a circumference, which must pass through the two points, and your position.
If you are closer — a wide angle — the arc is flatter; a narrower angle means you are further away — a larger part of the circle lies on your side of the landmarks.

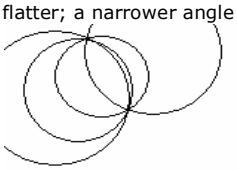

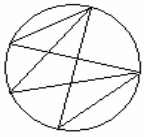

There is a constant angle
at anywhere on the circumference
when looking at the two points.

You don't need to draw this though.
Only memorise or note the angle.
To fix your position make two separate measurements e.g. one vertical angle and one horizontal angle too; or two horizontal angles—using three landmarks (or two vertical angles; or a radar range & direction from a single landmark; or use any other line(s) of position e.g. a compass bearing or line of sight or an advanced position line).
Now you can relocate that buried treasure, that fishing spot, waterhole, rare plant, good campsite… simply through making a couple of notches in your walking stick.

The best range-finder

..is a theodolite/sextant — designed to measure angles very accurately. You can use them horizontally.

Align the Map to Fix Your Position

Match the map to the landscape visually.
(The first result is to show you geographical (true) North.)
(The second result is to show up other landmarks.)
The third result is to show up **where you must be**.
(The fourth result will be to indicate which way to continue.)

I.e. Work **backwards** — from Land to Map.
One method is as follows…
Place a small pebble on the ground.
Place other pebbles in a circle around that,
each one lining up with some landmark.
Now place the map over that centre pebble.
Twist and shift the map until it lines up.
Get down behind the map and visually sight over it.
When everything lines up you must be at the centre of the circle
— over the central pebble.

Your position should **look** obvious!

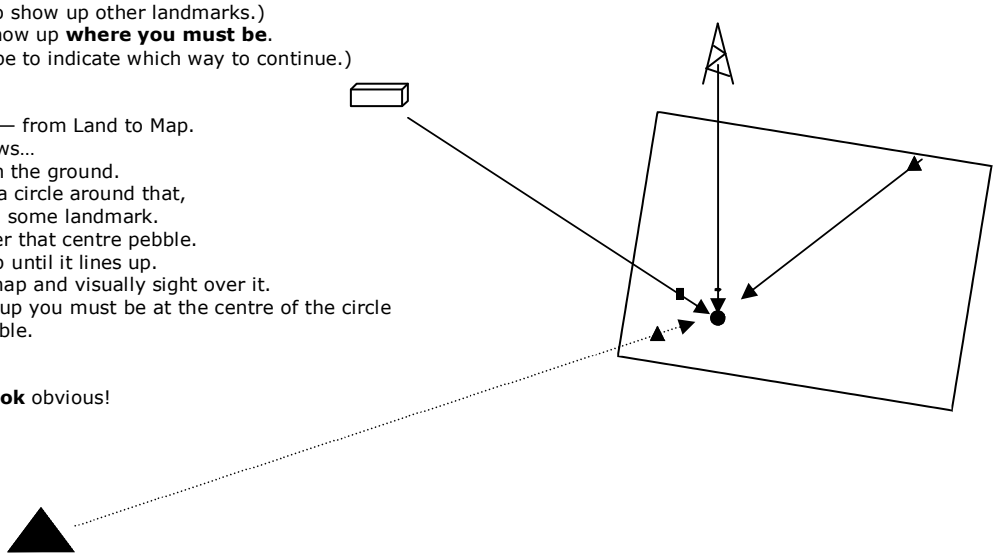

A Horizontal Angle between Features Yields a Position Circle

There is no need to draw this position circle (diagram 1), and it is somewhat awkward to construct, but there are two cases which are well worth drawing (permanently) on the map…

Diagram 2. A right-angle between two peaks — it means that those two peaks mark the **diameter** of the circle — easy to draw.
Diagram 3. The 'circumscribed' circle of three peaks — you can <u>see</u> whether you are inside or outside of that circle.
To find the centre of the circle, use perpendicular bisectors.

The angle seen from the circumference is constant.

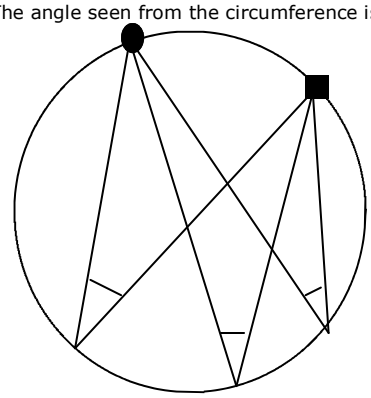

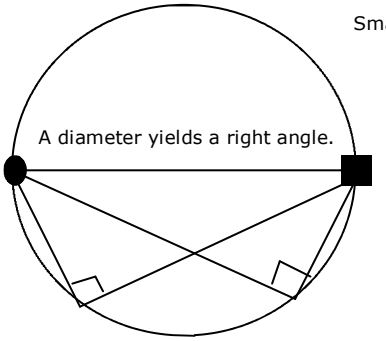

Horizontal angles are **larger** on the **inside**.

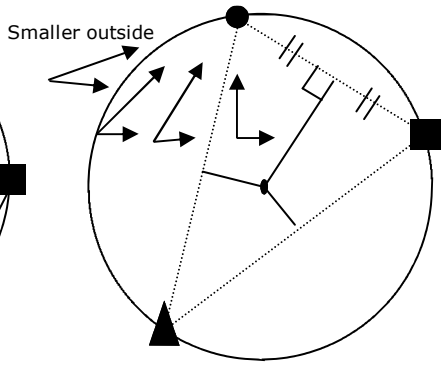

'Two Horizontal Angles' Usually Fixes Your Position

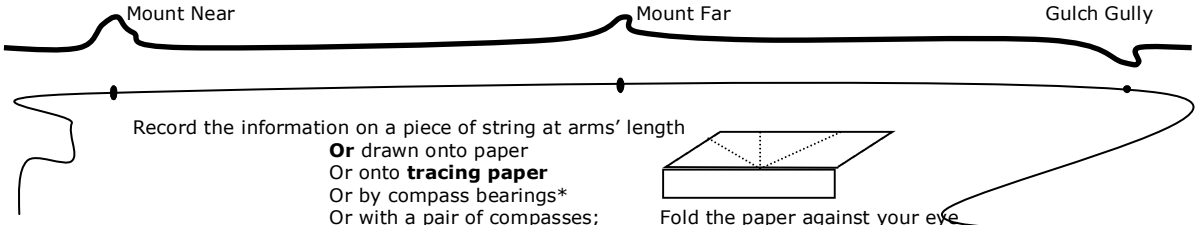

Record the information on a piece of string at arms' length
Or drawn onto paper
Or onto **tracing paper**
Or by compass bearings*
Or with a pair of compasses;
Or by angles cut into a leaf; drawn on the ground; folded into paper; etc
*you could plot back-bearings, or simply record the three directions for reference.

Fold the paper against your eye

This is a quick, natural, accurate method which uses no numbers or circles. It is how our mind judges the horizontal layout.
Only one position on the map will produce such a pair of horizontal angles.
If you draw the angles onto tracing paper, you can **place it directly down over the map, to fit it on properly**.
If you make a three-arm trio of compasses with a locknut at the pivot, you can also place this directly onto the map.
Coastal yacht navigators find the method useful, and to measure angles very accurately they use a sextant horizontally.
In theory it plots as two intersecting circles — only one intersection makes sense.
The nearer the landmarks, the smaller the circles, and the better the fix.
Just don't choose three landmarks which are already on one circle along with you!

Place your eye directly over the pivot

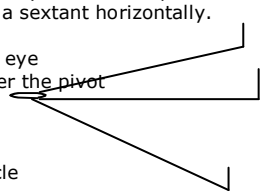

Wrong

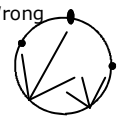

Right

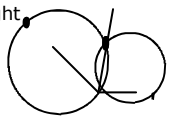

You don't have to draw these circles
Also, if you want to find an angle from the map
you can just read it from the map without drawing any circle

Angle

Distance

The Sine-Cosine Tally Method of Keeping Track, Blind

This project is only "for experience" — i.e. you are unlikely to want to do it twice! — but you should do it once.
The idea is to **keep a running total of your 'Northing' and 'Easting'** from a starting point.
I scored 97% efficiency in getting back home, on an 11 km circuit, on my first try, with an unreliable home-made pedometer!
It is a 'bottom-up' method — for when you are walking blind — when you cannot work down from the Big Picture.
With pen and paper you record your pedometer readings & directions, and update the numbers, say every hundred double paces.
E.g. "Current Pedometer Reading: 1400; (time-check 1 1/4 mins; pacing check 140 m) efficiency: 90% (slightly indirect path)
Recent Direction 005°; **adjustment needed +90, +8**; new tally: 419, -45; (comments: crossed a wet ditch)"

The chart of 'adjustments' follows this box, with an example.
200% efficiency would mean "I missed the last 100 reading"
50% = "I had a diversion" 110% ="I stopped at 110 — make the next reading 90% efficient"

- Set it all out with headings and long columns.
- Your judgment and arithmetic don't need to be accurate, because the errors are likely to cancel (!)
- Check your pedometer for each type of terrain, by time and by pacing.
- Avoid blunders e.g. check that the direction and the adjustments are similar between two legs.

It is quite pleasant to stop every minute or so — at every 100 on the pedometer — It keeps you alert, rested and oriented.
Making a full map (see 'Layout Map—Blind) is more satisfying and just as quick.

Mariners can (either draw their tacks, or, using this method) calculate an overall course-made-good from their many compass legs, by first adding up (distance x sin true course) East, and then adding up (distance x cos true course) North, from each individual distance and course.
Their overall distance is $\sqrt{(sum1^2+sum2^2)}$ at an angle of arc-tan(sum1/sum2). Then correct for magnetic variation.

Rise and Run Adjustments — Sine and Cos — North and East

Caution: These are COUNTER-Clockwise Bearings! — as seen from a compass holder...
I.e. The needle swings to left and to right while you point straight ahead.
E.g. when the needle points to "20° East of 0°", as in the diagram,
"N(orth) O(n) R(ight)-West" reminds you that you are heading NNW-ish,
not to the East. N20°W is your **anti-**clockwise heading.
Don't let this unusual method of measuring bearings confuse you about normal clockwise bearings.
There are really only 10 pairs of adjustments used in the whole table
I have made East positive, only to match normal bearings and
Map grids, which use East as positive (West as negative).
Here, **West is negative; East is Positive.**
South is negative; North is positive
Examples: North East is + + N315°W;
South West is – – N135°W;
North West is +– N45°W;
South East is – + N225°W
The + and – signs tell you how to change your tally of North and East.
E.g.
100 paces to the NNW ish, as in the diagrams is...
94 paces of Northing and 34 of Westing
Or +94, –34

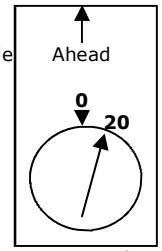

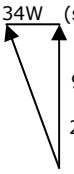

34W (sin20 = 34 per 100)

94N (cos20 = 94 per 100)

20° to the West of North

					Change N	Change E
Actual direction…	(NE	SE	SW	NW)		
Adjustments…	+ +		– –		cos	sin
		– +		+ –		
Protractor Readings…	360	180	180	**000**	**100**	**0**
N nnn° W…	355	185	175	**005**	**100**	**9**
	350	190	170	**010**	**98**	**17**
	345	195	165	**015**	**97**	**26**
	340	200	160	**020°**	**94**	**34**
	335	205	155	**025**	**91**	**42**
	330	210	150	**030**	**87**	**50**
	325	215	145	**035**	**82**	**57**
	320	220	140	**040**	**77**	**64**
	315	**225**	**135**	**045**	**71**	**71**

The next figures rerun the previous ones...

310	230	130	050	64	77
305	235	125	055	57	82
300	240	120	060	50	87
295	245	115	065	42	91
290	250	110	070	34	94
285	255	105	075	26	97
280	260	100	080	17	98
275	265	095	085	9	100
270	270	090	090	0	100

e.g. As a different example, a tally of –640 +770 means that you are 640 paces South and 770 paces East of your starting point, which means that you are now N230°W from home, according to the table.

Rule of Fingernails

Your pointing fingernail should be about 1°, one in 60, when you stretch your arm.
Your thumbnail at arm's length, held upright, might be one in 50……
A pencil in your fist will be about one in 100.
Check it out with a tape measure and calculator. E.g. 1.2 cm (width) / 69 cm (distance from eye) = one in 57 = 1°
Adjust how you hold your arm (e.g. square in front, diagonally- or fully- on the side) and which fingernail, etc, until you find suitable combinations which you can remember.
But the trick to making this useful in range guesstimations is to guess the distance away first, then estimate the height of a mountain, as follows… 60 km away — use one in 60. One full nail means 1 km showing;
30km away—use 1 in 60 =½km showing; 50km away—use 1 in 50 = 1km showing; 10km away—use 1 in 100 = 100m showing
Reverse logic can give you the distance away, if you can see the base clearly e.g. from sea, and know the full height.

Rule of Tom Thumb

Tom Thumb

At various times you have to judge parallax.
And the trick is to project the needed sideways movement onto the intervening landscape.
Imagine yourself over there on that ridge — Are you as thick as that tree trunk? Are you as tall as that bush?
To make the two landmarks line up, imagine the distance over there in the distance.
"I would have to walk that far sideways over there" = "I would have to walk the same distance sideways over here"

Apparent Size

The size of people, cars, houses, as measured at arm's length, varies with distance. You know how big they are already.
To get a good gauge of very small angles, compare their width & height with a key or knife-blade or toothpick, held edge-on.

Which Way is North?

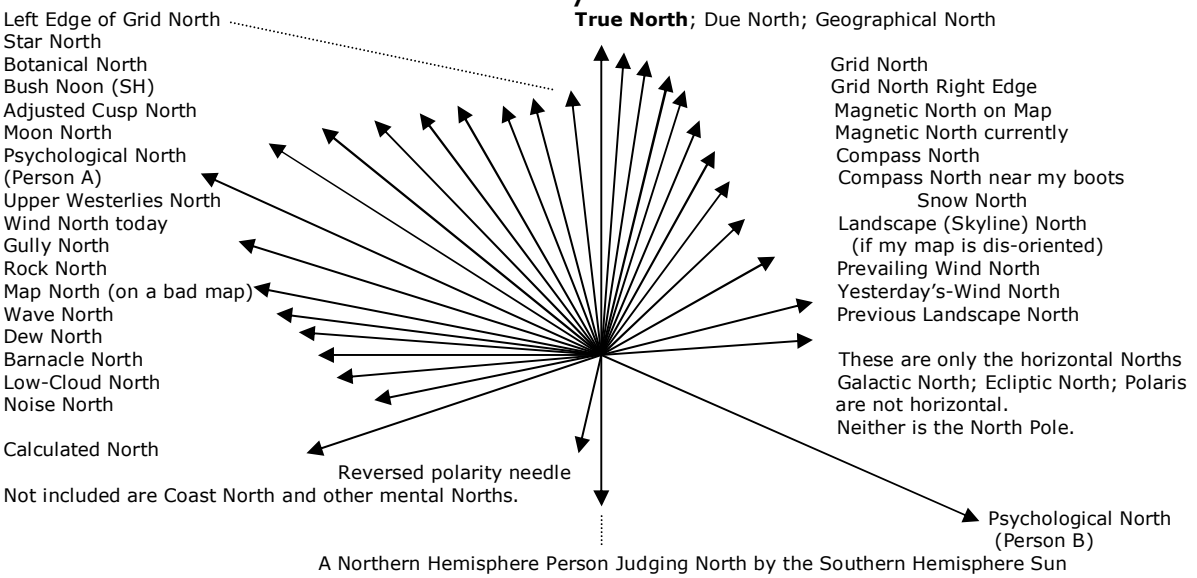

'East-least; West-Best' Rule — Tells You Which Bearing is 'Best'

This rule works only for horizontal directions, measured clockwise.
It is for <u>shifting between</u> True, Grid, Magnetic & Compass bearings 0-360°.
"Least" really means "Less than the other alternative number" **"Best" really means "More than the other choice"**
The following wording ("the <u>North</u> which is") is important, since you need to choose correctly between a right and wrong choice!
***<u>The "North" which is East</u>* (clockwise) of another "North"** ***<u>gives the Least bearing</u>*.**
Any "*<u>North</u>" to the West* (anticlockwise) of an alternative "North" yields the *<u>Best</u>* & Biggest numbers.
Specifically, it is most used to shift back and forth between magnetic North and true North bearings.
The wording may then be rendered: "Compass error West; compass heading best". Memorise that wording.
Because this mental arithmetic is a difficult ask, you need to draw a diagram to make the conversion clear.
Learn this reminder "Check-up-on-yourself!" diagram… And <u>visualise</u> it, on the real landscape in front of you.

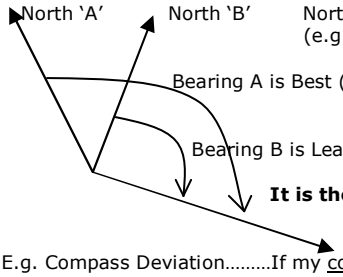

North 'A' is **West** of North 'B': it yields the **Best** of the two bearings.
(e.g. if A is true North, and B is magnetic North, the 'magnetic variation' is 'To the East' and "the compass heading is Least")

Bearing A is Best (It yields a bigger number — East-around)
(e.g. if the magnetic bearing is 085°, the true bearing is bigger… 120°)

Bearing B is Least North 'B' is **East** of North 'A': it gives the **Least** clockwise angle.

It is the *North* which is "East" or "West"; The *bearings from it* are "least" or "best".
Numbers are meant to make sense — your job is to look for that sense, <u>visually</u>.

E.g. Compass Deviation……If my <u>compass</u> *North* points too far <u>East</u> e.g. by 10°, 050°M = 040°<u>Compass (-*bearing* is Least)</u>
E.g. Grid Convergence Correction…If Grid N is 2° tilted to E of <u>true</u> N, 070°T = 068°G (Grid is Least, since Grid N is East)
E.g. Map to Compass…… 090° on the map means: set your compass to 100°M, when magnetic variation is 10° West (Best)
E.g. Compass to Map…… 205°M as a compass bearing = draw 220°T on the map, when magnetic variation is 15° Least
E.g. *Add* the GridtoMagnetic Convergence clockwise *E* angle to <u>magnetic</u> N to get *Grid* N "MAG=**M**agnetic, **A**dd(the GMC)=**G**rid"

String Map

This is a fun-project. As you un-do each knot later, you relive the journey!
A suitable string is builders' twine — long, thin & strong.
An extraordinary number of code variations are possible, provided you can "read" the different knots afterwards!
Keep it simple, for a start.
Example: Start with a slip knot. Half-hitch it every hundred paces, to keep track of distance.
Tie a special knot at any change — to remember it by.
E.g. a strangler hitch around some bark for "thick forest"!; a thumb knot around a feather for "Water Bird Lake"
Chain the slip loop to keep it going. Or thumb knot the loop to lock it all up.

A simple numerical code is: slip knot; + n half-hitches; chain on; m half-hitches; chain on; ongoing tally…
n might mean something like 'direction' by clockface
m could be 'terrain type'

Here are some variations…

right and left half-hitches — multi-twisted hitch — twisted hitch — crossed hitch — figure-of-eight loop — multi-turn hitch

slip knot; two half-hitches; chain-on — reef knot chain-on — thumb knot

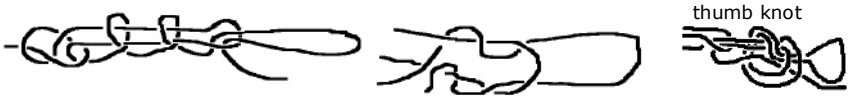

Distance

Progressing A Position Line across the Map

Situation: You catch sight of one landmark; plot a back bearing; then travel — say for two hours, West — before you catch sight of a cross-bearing. You can still plot a 'fix'!
Example (fictitious): You lose sight of Mount Connor in the Northeast; and later see Ayer's Rock. Your pace is 4kph. From the bearing of Mount Connor, you plot a position line (by a back bearing running back from Mount Connor). You must label this '10 a.m.'
Now **shift the whole position line** 2 hours to the West i.e. 8 km, and label it '12md'.
To do this choose any two points on it; shift those points 8 km West;
draw **a new line parallel to the old**. You are now ready to plot another back bearing — this time from Ayer's Rock — to cross the advanced position line, for a 'pinpoint fix'.
Alternatively, you could 'shift' only Mt Connor on the map, first.

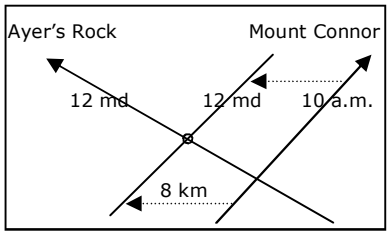

Pinpoint "Fixes" are a Myth!

(Rule: Don't suppress uncertainty, nor reject any unwelcome information!)

Any 'line' of position looks exact. But it comes with unavoidable uncertainty
and should be plotted (mentally at least) as a wide band of 'likely position'.
Potential inaccuracies abound. E.g. lining up the compass;
reading the compass; plotting that angle;
grid north variation across the map; map distortions;
errors in magnetic variation; protractor inaccuracy;…

You can now advance your "area of uncertainty"
across the map, expanding it with each blind move!
'Deduced Reckoning' navigation at sea (i.e. "intelligent guess")
is more a matter of keeping track of **accumulated errors**
due to tides, leeway, inaccurate course-holding/recording/plotting, inaccurate instruments. So too on land.

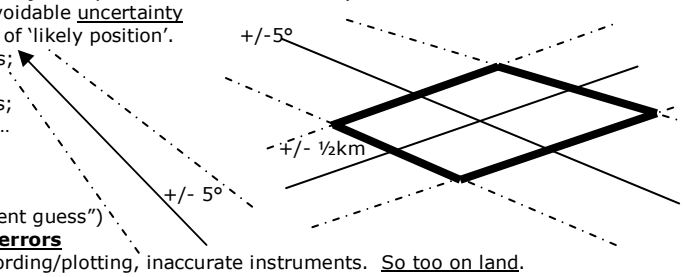

Getting Started in the Dark

The task is to know what direction to head off in, when you can't necessarily see the landscape…
You need a Map & Compass.

TAKE A MAP-BEARING
— Find the desired angle, on the map, in relation to the North-South grid……
— "About 60°, as a **guess**"
— Use the edge of the protractor
— Measure the angle — only the protractor is needed for this!
Line up both sets of parallel lines — mapgrid lines,
and the lines on the needle-holder-circle.
— The protractor will display the angle between the two sets of its parallel lines
"55° E. of N." (It is marked clockwise)
(This is Called Working from **Map to Compass**)

SET THE COMPASS
— Using 5° East magnetic variation, and ignoring compass deviation…
Compass bearings should be 'Least'
i.e. 050°C
— adjust the 'Follow Me' Arrow to point to 50° (to the East of North)
Twist the circle on its mounting.
— Don't alter the protractor again!

FOLLOW THE COMPASS
— Now turn your attention to the needle
— Line it up with the 'zero' on the protractor, by twisting the whole instrument,
And by using the grid lines on the movable circle as a guide.
— Twist yourself this time, to line up with the 'follow me' arrow. And off you go!

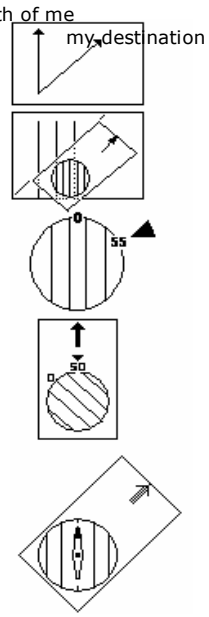

Mnemonic: **Map to Compass** (= Map to Compass Course) Means **M**ap Grid to Compass Zero;
Angle the **A**rrow to the course target;
Put on the correction; **To** Get it, use the East Least West Best Rule;
Compare the Compass Zero to the Compass Needle; **Pas**t the Follow Me Arrow is the Course Target.

Thinking Backwards

"Danger bearings" allow you to stay safe, without knowing your position accurately, but you need to think accurately…
As seen from the lighthouse, you must stay on the right of 315° (and to the left of 060°). That much is clear from the chart.
Anything 'to the right', is clockwise from the observer, so for the keeper to keep something (e.g. you) 'on the right' means to keep its bearing bigger than ahead. 'Ahead' is parallel to the 315° track shown, since the lighthouse keeper is wishing to see you 'on its right'. He will be happy to see you on any bearing bigger than 315° and less than 060° i.e. 315 to 360, 0 to 60.
From your point of view, that means, that you must *use back bearings,* of 135° and 240°, and then keep those tracks on your right and on your left respectively, as you look in. 'Ahead' is parallel again, but opposite to the track shown on the chart, if you are trying to keep the track 'on your right'. Looking in, moving yourself to the left will place the track further to the right, and make bearings to things on it bigger. That means, in numbers, the lighthouse will bear more than 135° and less than 240°, i.e. It is safe to approach at **anything between 135° and 240°**. Notice that the more/less sense is preserved with such use of backbearings.

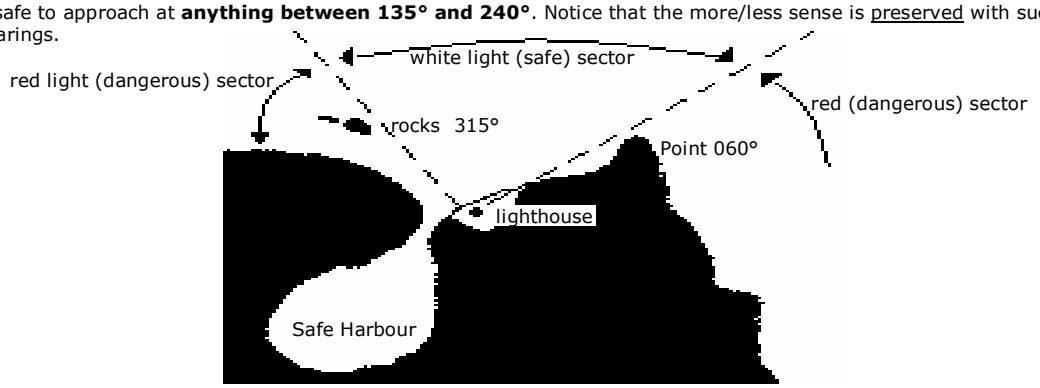

Taking a Bearing from the Landscape

- **Start with a guess** e.g. "That hill is about 060°T from me"
- Point the 'follow me' arrow to the landmark
- Grip and twist the movable protractor circle until the grid marks on it — i.e. the zero — line up with the compass needle
- Read the resulting angle off — between the zero (which is "North" —Compass North in this case), and the landmark.

The result should agree with your guess e.g. "The landmark lies at 065°C" i.e. East of Compass North i.e. clockwise from.
Take a B.E.A.Ring Mnemonic: **B**earings are Tracks; **E**xpect to have to plot a back-bearing;
Arrow to Landmark; **Ring** to Needle and **R**ead It off. (Think: "Arrowing" to remember A)

If it helps you sight the landmark, hold your compass string vertically upwards from underneath the tip of the follow me arrow.
The 'Zero' is meant to be used for 'North'! But if you point the follow-me arrow to North and the 'zero' to the landmark, then (a) it is more awkward and (b) it is an anti-clockwise "bearing".

If you are about to use a map, you must now convert the compass bearing to a 'True bearing'.
Firstly mentally change the compass bearing to a magnetic bearing if you are aware of any compass deviation i.e. compass errors. (You usually do not worry about compass deviations when bushwalking.)
Next mentally convert the magnetic bearing to a true bearing using the East-least, West-best rule.
Example: "6° variation West" (The Magnetic *North* is West of True) "Compass Best" so… 065°C = "059° True"
Lastly, physically change the protractor to read 059° when you want to draw on the map, at an angle to the map's True North.

See opposite & above

Plotting a ('Compass-to-Map') Back-Bearing

Draw a line on the map from the landmark in an opposite direction. Use the protractor to get the angle right.
If you see a landmark North of you, you must be South of it.
Or, what looks 'South-West of me' must mean that you are 'North-East of it'.
Think backwards! See "Back bearings: +/-200° then -/+20°" p.18e E.g. 059°T from me to it…
Note that I have already converted Compass to True. 239°T from it to me.

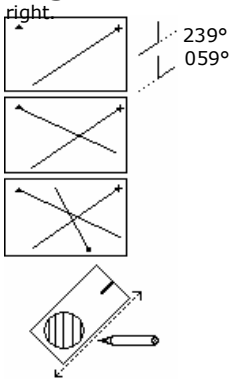

One bearing yields **one** line of position (from a back-bearing plot — diagram 1)…
Two bearings from two different directions cross for a 'fix' — diagram 2…
Three bearings don't cross. They form a 'cocked hat' and keep you humble — diagram 3…
In this case, you could be anywhere nearby (which is also true for any "Fix").
Expand your possible positions, don't try to pinpoint a "better" one.
Take more care next time, especially on the closer i.e. the more reliable landmarks.
Check your compass for fixed errors e.g. "always 3° too much".
Choose directions which are quite different from each other, to get the clearest crossing.
Compass to Map Mnemonic: (Compass B.E.A.Ring, then) **Compass to** Grid (or True) Correction;
Make it a Back Bearing; **A**ngle to Grid; **P**encil to Paper—Draw It in. (Think: "Mangle" to remember A)

No-Number Back-Bearings

Method 1. Use the white-tipped (South) end of the compass needle, when you take the bearing.
Method 2. Point the "follow me" arrow back toward you, when you take the bearing or follow the compass.
Method 3. Line the protractor zero up with Map South, when you plot the bearing.
Method 4. Point the compass edge to the landmark on the map and draw your back-bearing line backwards from the landmark.
You don't need to alter the compass setting at all (except for the magnetic variation or grid/magnetic convergence angle).
Warning 1. Don't mix back-bearing shortcut-methods, or you may get a 'double negative'!
Warning 2. Check all you do, against reality, for sense.
Warning 3. Don't try to shortcut Compass-to-True conversions at the same time as shortcutting the back-bearing, or you will get confused.

Parallax While You Walk — JUST DRAW IT!

Most **hand-made maps** start with **a known baseline** scale — a straight segment of a known distance.
I.e. "I walked from here to there in a straight line for 1 km" — for this you may need a pedometer.
Or if you are Captain Cook or Matthew Flinders "We sailed on (this) bearing at (this) speed for (this amount of) time"

You can then **add cross-bearings at any time** to fix the position of distant landmarks
— obviously at the beginning and end of the baseline are the preferred places,
perhaps combined with one from the middle…
The longer the baseline the more accurate the fix on distant marks.

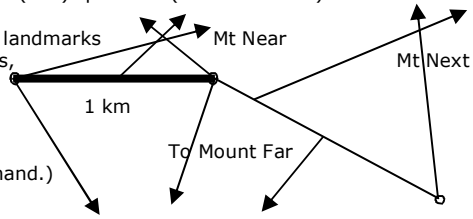

You really only need a protractor for this **running base-line triangulation**.
(And on a Mudmap, you just eyeball the direction and draw an arrow in freehand.)

Here is an example of the easiest sort, and the reverse of a double-angle…
After a week at sea, with no sight of the sky to navigate by, I was understandably over-anxious about my position, and "needed" a "fix"! I decided to sail toward an isolated lighthouse-island, in the hope of passing close enough to see it, perhaps, if we weren't too far off course. The winds were hurricane force, of course, and the sky was filled with salt-spray lifted off from the water's surface. It was dusk. Suddenly, above the mist, dead ahead, the lighthouse turned on! Alarmingly high, and close. Too close!
Were we "found" or about to be "lost"!? (See "Aiming Off"!) After the cheers: "What do we do? How far off are we?!"
"Turn at right angles to it; Maintain that course; Take regular bearings until it is diagonal; Keep track of the distance run. Then resume your course. That way we can tell how far away it is."
"How can you do that!" said the bewildered, inexperienced navigator.
I replied: "I'll **just draw it**"

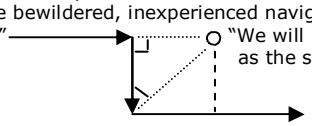

"We will pass as far away as the sideways distance we have run"

A "Start-Stop-Rest" Navigation Mnemonic Overview

S T A R T
"Start" has six points, as illustrated in the diagram below…
These mnemonics, e.g. S.T.A.R.T. or S.T.O.P., will be expanded in the rest of this section—they summarise the book.

- Start with a **Prayer** — "What more do I need to do to avoid getting dead?"
- **P.R.E.P.A.R.E.D.** — What have I forgotten to do or to take, to practise or remember? Start off oriented before you go!
- **S.T.A.R.T.** — with your start-off-trend-direction, which is the single most important navigational item. Start off aligned!
- Start on **T.I.M.E.** — Read your watch, and benchmark your trip. Say: "Time & Tide" as well, because sometimes tide levels are crucial.
  — Check pace & progress. Estimate the Range, because time is all about distance.
- Start time is **6 a.m.** — It's always time to **look** in **6** directions, plus 2—Time to Look **A**ll **A**round, **A**ppreciatively, plus time to look at the **M**ap
- Start with **N.S.E.&W.** —4 sets of 4 directions. The final W is for **W.E.A.T.H.E.R.**

The final R signals that you are **Ready** to go. Don't start until you **R**!

S.T.O.P. A **S**top is a **S**ignal—Interpret the Why? of your stop: Does it **S**ignal a change in watershed, catchment or view;
in **T**rend or **T**errain or '**T**ime & **T**ired';
in **O**rientation (embarrassment—if lost, rule one: Stop);
in **P**rogress/**P**osition/**P**lacement—campsite, junctions, features?

R E S T is best remembered as a mnemonic of 5 RE's…
R.E.-O.R.I.E.N.T. — using these letters as a mnemonic. E.g. **RI** re-Visualise the Invisible Big Picture, and **ENT** Relate yourself to Nature.
RE-RE.L.A.T.E. — re-interpret all directions to relate to each other and to the skyline, as explained below. Now don't forget...
RE-MEMBER — the WAY and the SEQUENCE and especially to NOTE down (in the NOTEBOOK) the NAMES, and DRAW what you need to.
RE-LAX — Play the Navigational Game—teach it—and 5 other games.
There is a final '-art' to 'stop and rest-'
RE-ST-ART 1.1 **re-**pray; 1.2 **re-**prepare; 1.3 **re**-s.t.a.r.t; 1.4 **re-**t.i.m.e.; 1.5 **re**-look (6 a.m.); 1.6 **re-**direct (NSEW), (which includes **re**-weather)
... e.g. 1.3 Re-S.T.A.R.T. after a rest, on the next leg—to relate it to the last & to the *start*-finish trend.

Summary: **"Start"** 6 times**; "Stop, and Re-st-"** (with 4 new **"Re-"**s)

(The Start mnemonic is a good example of the "clumping" technique
—never more than a handful of things to remember, yet it leads you into hundreds of points.)
Put the S for Sky at the top, where it should be.
Use a hexagon for the 6 START up points.
Use a clock-face for TIME,
and the compass-rose for NSEW.
Show the 6 directions in 3 dimensions.
Balance it all on a triangle
—for the "6 a.m. start" , since it has 3 parts.

Sky and Skyline
Toward what Target; through what Terrain
Away in what Anti-direction
Rain catchment; Region; Ranges
Trends

See the boxes to follow

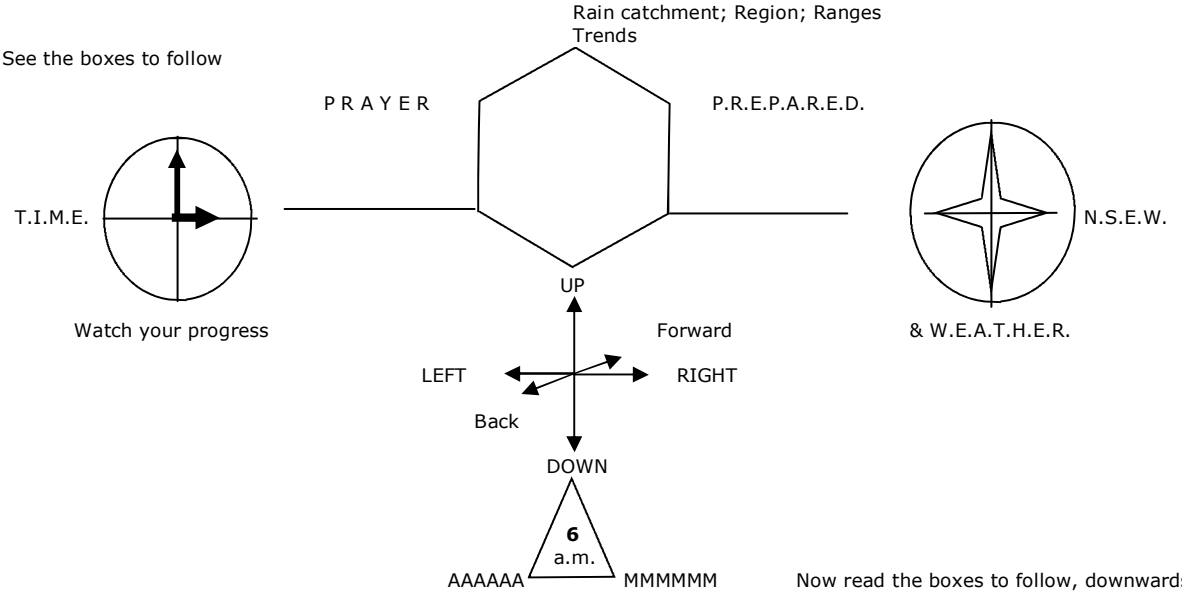

Now read the boxes to follow, downwards...

P.R.E.P.A.R.E.D.—A Before-You-Go Checklist ..so that most of your navigating is already done, from home!

P.ractise Global orientation. Get a blow-up globe. Teach others. Learn (what to learn, look for, do etc.). E.g. predict the Sun.
Learn the Mnemonics, Codes & standards e.g. S&R signals, compass directions
Pace your paces — metres per minute; steps per 10 metres. Left-right bias (blindfold).
R.esearch History. Ruins. Artifacts. Mines. Geology. Get a geological map from the library. Landform. Soils. Climate. Windy season. Prevailing winds. Wet season. Vegetation analysis. Rare species. Biogeography. Geography — global neighbours; time zones; local coasts & river mouths. Catchments. Cardinal features — places N,S,E,W from you. Land lines on map. Place names — their meanings and history.
E.quipment See the next box.
P.lace names See the box after that.
A.stronomical orientation Sleep on a map, and with a compass nearby. Point the map North. Times and directions and heights of Sunrise & set, Moonrise & set. Southern Cross. Pot. Sun at noon, at magnetic North/South, when East, when West, at magnetic East and West. What landscape markers (from the map) for these directions? The star opposite the Sun. Time zone. Local noon — its exact time. The Sun's coordinate predictions. Your longitude. The "Equation of Time". Moon-phase, -number and –day of week. Tide times.
R.eminder Note your intentions—and tell someone where you are going and when you expect to be back.
E.xpectations of the Weather Barometer reading. Get the Weather forecast. Follow the Changes. Buys-Ballot's law. Steering winds.
D.irections & Familiarity Ask someone who has been there. Read track guides. Look at an aerial photograph. Go beforehand! Find your old track notes. Look closely on the way in. Choose a route. Plan your strategies. Coordinate the cars, the car keys, any food drops, shuttles or codes

Equipment — A What-You-Might-Like-to-Take Checklist

(E. in PR**E**PARED, for "Equipment"…)

- **A Map** **A compass** **A clock**
- A notebook and pen — for track notes. A track guide. An aerial photograph. A navigator's handbook.
- For map making — graph paper, paper, pencil, a compass-holder. A map-holder (string around neck to a…), clipboard, stopwatch.
- Polaroid glasses — to orient by the dark blue sky band. Radio — for RDF, weather. Mobile phone. GPS. Camera.
- Binoculars. Pedometer. Rangefinder. Calculator. Altimeter. A sine-cosine table.
- A map cover. A spare map. A navigational protractor. A magnetised needle. A sundial compass. A two-flap cardboard compass.
- Marker tape; Chalk; Scissors (for cutting leaves!). A clear plastic tube. A clear plastic drinking bottle (for an artificial horizon).
- A long piece of string (e.g. for a range-finder). Short pieces of string (e.g. for a plumb-bob). A walking stick. Someone who knows the way.

(P in PRE**P**ARED) **P**lace-Names —Learn Them—You Navigate by the Local Features!

"Look. That's Big Split Rock and that's Little Split Rock. Wild Dog Tier is that range over there, and this plain drains out over there, into the Ouse River You can find your way anywhere around here from those."

Suddenly I was unforgettably oriented, just like he said he was, when his dad told him.
For me, the funny names on the map now meant something worth knowing.

Place-names are not just names you don't know; they name places, which you ought to know.

More accurately, place names 'place' you in the landscape, as well as 'orient' you to it.
A sense of location is almost independent from your sense of direction, and is rarer to come by.
Direction-sense places you in relation to very distant background references i.e. it doesn't 'place' you at all!
Location-sense places you in relation to the immediate foreground — visible landmarks, catchment basins, fixed 'tracks', etc.
Place Names Place You

S.T.A.R.T. off Oriented

— **S**tart-off Direction? (e.g. "upstream") — the **S**ingle most important navigational knowledge — it tells you the return path, and the search path.
Sky/**S**kyline i.e. especially according to the **S**ky (e.g. "toward the Moon; into the sunset") and related to the **S**kyline.
Start-off directions for the parallax method, pp. 51a,53a, e.g. to magnetic NSEW skyline or recognisable middle-ground points, or rhumb-line point and another set of such bearings, taken from the map, as seen from the finish, if you are on a one-way trip.
— **T**oward what **T**arget? Specify a
Track (e.g. "downhill to the lake") or a **T**rend. A rhumb-line bearing is best—from start to finish.
Through what **T**errain? (e.g. especially see what the land-shapes would look like without vegetation)
— **A**way from What Point? (e.g. "from the lookout carpark), and in what
Anti-direction (e.g. toward that other peak"—a skyline direction)
— **R**ain Catchment — In what major
Region? — as can be gauged by the surrounding
Ranges. (e.g. "on the main spur dividing North from South arm, downstream of…")
— **T**rend: **R**ead the **T**rend off the map, for starters; then continue to **R**ead the **T**rends & bends and the progress as you go on.
Read (your compass; map; watch) in **R**elation to the **T**rend, i.e. in order to follow the trend, which is the focus of this mnemonic.
Plus: **R**ead any other **T**rends, from any other places, from memory, in relation to the sky. Be able to revisualise them at will.
— Re-s.t.a.r.t. on the next leg, to relate it to the start-finish trend, and to relate it to the previous leg.

T.I.M.E. & Distance — that's the point

TIME to Look at your Watch — PACE YOUR PROGRESS — the point of this T.I.M.E. mnemonic is distance.
— "Watch" your steps—i.e. calculate your pace from the elapsed time;
— "Watch" the Sun without looking at it, because your watch tells you where it should be.
— TIDE times may be necessary.

Instrument Check — watch, stopwatch, pedometer, odometer, rangefinder — for progress
— compass: Double the bow angle, or halve the stern angle p.94c; parallax readings on landmarks.
— (and other instruments — barometer, altimeter, GPS, radio null, radar, depth-sounder)

Map-Scale Check — are you judging the scale well? — What is the next goal? (This M is not for Map in general, but only for distances.)

Estimate the Time of **E**xpected Arrival — guess each milestone ahead of time to set you free from a nervous TIME. Nominate an E.T.A.

Distance

PROGRESS: can also be gauged by lines of sight and changeover lines. Use a stopwatch, or pedometer. Read the scale.

PACE: distance divided by time; 20min/km, double and halve = 3kph; 1mps = 3.6kph; 4kph normal walking pace;
finger-math helps; pace yourself evenly.

RANGE: Use the letters of RANGE & FACULTIES as a mnemonic to remember the different ways of gauging distance in the distance...
Range squared = dee height; Area divided by offset; New direction parallax; Grass-leaf vertical angle, horizontal angles too; Educated guess
Familiar lengths, and Fingernail Rule; Acuity of vision and Apparent sizes; Cloud behaviour; Upper and lower guesses, averaged; Landscape lengths;
Tom Thumb Parallax; Intervening detail & haze alters your perception of distance; Extreme range comes in two bites; Sound travels at 1 km/3secs and
Sidewise offsets with blinking.

150km from 1 nautical mile high; 1° latitude from 1km; 50 miles from 500m; 50km from 1/5 km; 5km from 1/500 km (2 m); 12.7km from 12.7m.

Look in These 6 Directions — in 6 ways — Up and Down (Part of "Start at **6** a.m.")

6 Directions to Look in

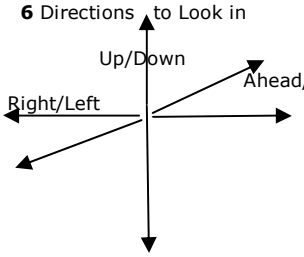

UP — **Astronomy** — Sun Moon Calendar-Stars
— **Sky** — Clouds, Contrails — Blue Band
— What's up with the Weather?
then DOWN —Geology — **Soil** Type; Rock Type
—Clues — Animal **Trails**; Bootprints; Fallen markers

— Up-**Sun** — haze & glare
— Over what tropical country? Then — **Down**-Sun — haze, shadows, colour changes

— **Wind**-wise — Up-Wind — Where on the skyline? — Down-Wind — Around what obstacle?
— Around what pressure system? (Wind in your face)
— From what coast/ocean? Listen; Smell
— Up-Breeze — To what cold surface? — Down-Breeze — To what warm surface?
— What's up-Weather? Where are the steering winds coming from?

— **Land**-wise
— Up-Hill — To what dominant peak?
— Point to the dominant peak, even when it is invisible,
— Down-Hill — antidirection?
turn around to see what lies away from the peak — the "anti-peak",
especially in relation to your shadow — this anti-direction has a lot to
do with a compass back-bearing—a visible line of position, you see
— Up-Ridge — To what hill/hillock? — Down-Ridge — Down which ridge? — To what headland?
— what offshore/underwater topography e.g. shallow reefs; islands?
Since this mnemonic is about direction, try to point to them.
— Up-Slope — To what ridge? — Down-Slope — Into what valley? Name them.
At what angle across the gradient?
— Where is the lowest upslope skyline, under the canopy? — Where is the highest downslope skyline, under the canopy?
— Up-**Stream** — To what distant divide? — **Down**-Stream — What stream? To what distant mouth?
— to what sedimented shelf; or deep-water channel?

Summary: **Above & Below** (twice each) + **Sun & Wind** (above eye-level) + **Land & Water** (below eye-level) = 6 ways for "Up & Down"

---

4 More Slopes

1 in 4, is 14°;

10° (aim off, 1 in 6) is 175 m per km;

1 in 60, is 1° (1 in 57.3 is the exact figure, but it doesn't multiply up linearly after a few degrees, whereas 1 in 60 does well after a few degrees)
& n° is n in 60
(1 in 6 is 10°; 1 in 10 is 6°);
The equivalent is 5/3 in 100, 5 in 300, 167 in 1000, for use with multiplying up with the isosceles triangle angle method.

36°, 8 in 11, is critically unstable.

Multiply them up, inversely.

Look in These 6 Directions — in 6 ways — Ahead and Behind

Look Back Regularly: Is bad weather sneaking up behind you? (p.46c) "Abracadabara Peek behind you" (p.63b)
How much has the track bent around behind you? Any acute track junctions? Can you remember this way back?
Compare Forwards with Backwards: Do you need a back bearing? Are you half-way? Is your starting point visible, to get an anti-track forwards?
Are there any changes in the terrain, to photograph, in your mind? Have you memorised the transition, from the far side?
Look forward — in **time** — the forecast! What is the Estimated Time of Arrival? Look Back — in time — How far have I come?
Look ahead and behind in **U**pper case **Time** — Could you recognise this place much later, in hindsight? — What is the history of this place?

Look AHEAD — Has the **view** changed — into a new catchment?
— at the terrain, the track showing up ahead, and the difficulties
— Look for track **markers**, and for trails ahead
— Line up two (new) marks exactly ahead to keep your direction sense.
Is it time to choose another leg, and a **pair of marks** lined up to it?

Look BEHIND — Have you left the old view behind — crossed a divide?
— Do you remember the **way**? Could you recognise it again?
— Look behind for hidden, reverse track markers.
— Look for the old lined up mark left behind.
Do I need to find new ones behind me?

Mnemonic: R,S,T-U,V,W,X-Y-Z **R**egularly **R**everse your nose. **S**tart/anti**S**tart bearing & **S**napshot-**S**equence your transitions
time (+**U**pper case Time); The **V**iew; **W**ay (remember it?); **XYZ: X** marks the Track (Markers); the (**Y**&**Z**) Leapfrog Marks =6 ways for Ahead & Behind.

Look in These 6 Directions — in 6 ways — Right and Left

Look RIGHT Through!
1. — the bushes, to the land-form underneath; to the contours & bedrock
2. — the canopy, to the skylight beneath it
3. — the trees, to the most distant marks, including other trees
4. — the treetops, to the sky — the Moon, clouds and the blue band
5. — the foreground, to the most distant skyline horizon, whenever you can glimpse it
6. — the haze, to the skyline details beyond the land horizon—use binoculars too.
E.g. to barely visible mountain peaks, or a second skyline & its parallax, or to islands in the haze, or the rising Moon.

Look LEFT
1. — to the most distant thing you have LEFT BEHIND on each side! i.e. Have you left it behind yet? Is it time to choose another one?
2. — Have you left behind your: handbook; air photos; track guide; track notebook? — No, well, don't just carry them — Look at them!
3. — Have you left anything behind at the rest stop? It's an important habit, to ask this each time you move.
4. — Look to the right and left, for the contours, the natural 'tracks', and the behaviour of the network you are crossing and leaving behind.
5. — Things on the right have big bearings; things on the left are lesser; than straight ahead, even less as you leave them behind.
6. — NORWest & NOTLEast — North needle 'on the left', means you are heading roughly East. (N. On your Right = West-ish)

Summary: **Right Through & Left Behind**.
In summary for all the 6 directions make sure you **Check Each One Out—Thoroughly!** (6 ways each)

..Plus in 2 More Directions — (1.) All, Around, Appreciatively, And At 6 L's

1. **All** your senses: Listen, Smell. Look around with your other senses.
2. Look 360° **A**round for the best & worst showings of colour, shine, flowers, greenery, new growth, fern fronds, lichens, trunks, branches, dirt.
   Compare Ahead with Behind, & Left with Right, & upSun with down-Sun, & up-wind, uphill, etc. for direction-dependent light and colour effects.
3. **A**PPRECIATE — the scenery; the colours; the contrasts; the shapes, outlines and silhouettes; the foreground movements.
4. **A**n Environmental Notice Time (see later box E.N.T. in REORIENT)
5. **A**ny LOCAL PECULIARITIES to remember.
6. **A**nd **A**t — **the six L's...** — **Land**marks — distances, sizes (apparent Largeness), altitudes Use binoculars.
   — Are any familiar, from other walks you have done elsewhere?
   — Do the place names make any sense? — interpret them!
   — **Lay**out — relative angles, compass bearings. Am I inside or outside of major landmark circles/triangles?
   — **Lie** — slope and terrain, catchment and drainage, contours and altitude
   — **Lines** Catchment Dividing Ridges and Rivers; Contours at right angles to the slope; Flight Tracks
   — **Ligns** — 'Same direction' Lines of Sight, or nearly so; and 'Opposite directions' alignments: Am I in between?
   — the fixed A-lignments . Am I left or right of major 'ligns'?
   — **Sky-Line** — talk it through; draw it. Cross-relate all directions to each other, and to the skyline

Summary: 'A' for 'a.m.': 6 A's, with 6 L's

..(2.) Look at **the Map**: a Systematic Checklist. (What 'System'??)

The system is to examine **top-down**. From general to specific detail. From large-scale to small-scale. Look for the downward progression in these…
3-D location: Latitude, Longitude, Elevation. Centre of the time-zone. The inset map showing where you are on the continent. Scale & size. The overall width- and breadth- distance of the map coverage. Grid size e.g. 10 km squares. Where you are on the map, geometrically too. The Starting Point. The Relative Location—over the horizon landmarks; coastal towns N.S.E.W. of you; something NE,NW,SE,SW of you. Radio stations. Highways. Light sources. Powerlines. Pipelines. Place names. Visible Landmarks. False peaks. Lines of the land — between landmarks. Recognisable hill-hill reference distances. Recognisable elevations e.g. towers. Grid North; True North; Magnetic North; Magnetic Variation; Yearly Drift in Variation.
Catchments & Divides; Catchment Size; Adjacent Catchments; River Drainage. Land shapes i.e. interpret the slopes, contours, the contour interval.
Trends and Bends — of the coast; rivers; roads; ridges; ranges; the track. Notice any creeks which flow backwards i.e. against the trend, and why.
What rising ground will be visible in each direction? Perimeter features surrounding your walk e.g. a river, a range, a road, the coast, a forest. Escape routes — unplanned exits. What are the ridge-line limits of the sub-catchment which most encloses your possible location? What high points are on it?
Possible routes. Clockwise or anticlockwise? Distances. Overall distance of your route. Milestone features. " Wake-Up " landmarks.
Initial bearings to landmarks. Initial angles to landmarks. Initial horizontal angles to landmarks. Final bearings. Rhumb–line trend from start to finish
Around what Obstacles? Fine details of the route. Don't forget the fine print — the date and reliability of all the map details; the grid datum for GPS use; the legend of symbols; colour codes; contour line thicknesses.

Mark the Lines-of-the-Land On Your **Map**

6 M's for Map:
Map Checklist (above); Map-Lines (this box); (all below...) Map Alignment; East-Least West-Best Mnemonic; Map & Compass; Map-Making Methods.

Pre-prepare your map — **Mark in the navigational lines** which the mapmaker has left out. Most of these are permanent.

A. Cardinal Tracks — NSEW lines. Across the whole map, from the very prominent peaks; or simply covering a local area, from the local peaks
B. Alignments — e.g. from Peak to Peak; Tower to Town; Point to Point — You may see two in line, or walk in between two landmarks.
C. Directions to major distant features — (for orientation purposes) — e.g. capital cities; coastal towns
D. The Overall 'Track' — the RhumbLine — from start to finish. Mark this lightly; &/or two or three obvious overall 'legs'
E. Circles of Distance Off — e.g. 5, 10, 15 km from Mt. X.
F. Circles of 90° horizontal angle — i.e. make two landmarks the diameter of a circle
G. Circumscribed circles around (i.e. which just include) the triangles in the triangular grid of landmarks
H. Transmitter to Transmitter networks — i.e. a triangular network of major peaks; include the distances and bearings on each line.
I. Flight Paths — if you can get hold of an aeronautical chart — it forms a second triangular network.
J. Catchment divisions.

Mnemonic: Four Straight Lines; Three Circles; Two Sets of Triangles; And One Very Sinuous Set.

Mnemonics for **Map** and Compass Work

◦ Map Alignment

A.L.I.G.N. the **M.A.P:**
A… Align the map to Sun with an imagined **A**rrow, and to your course with **A**nother **A**rrow;
L… to the **L**andscape;
I… To **I**dentify Landmarks and to get a Fix;
G… **G**rip the Magnet to the Map, and the Map to the Universe;
N… **N**ot "Swinging the Map" As You Turn;
Magnetic Needle to **M**agnetic North (in four variations, "0°" to North or arrow to course, Magnetic or True);
Angles Horizontal, fit over the map;
P… Align the **P**ath like landing a **P**lane, or align the map by aligning the **P**ath.

◦ Mnemonic East least, West best

East is Least and West is Best, tells you the North which gives the bearing that is best. i.e. the biggest in number; or lesser in number.

◦ Map & Compass

LMNOrP: Landscape; **M**ap; **N**eedle; **P**rotractor; either one **O**r two at a time, but not all three.

M.A.P. To Com.Pass, = Map to Compass Course, Means…
Map Grid to Compass Zero; **A**ngle the **A**rrow to the course target; **P**ut on the correction;
To Get it, use the East Least West Best Rule;
Compare the Compass Zero to the Compass Needle; **Pas**t the Follow Me Arrow is the Course Target.

Take a B.E.A.Ring: Bearings are Tracks; (**E**xpect to have to plot a back-bearing); **A**rrow to Landmark; **R**ing to Needle and **R**ead it off.

Compass to M.A.P.; Compass bearing, then **Compass**-**to**-True Correction; **M**ake it a Back Bearing; **A**ngle to Grid; **P**encil to Paper—Draw it in

◦ MUDMAP:

Mudmap on paper; **M**ark the start point on one edge; **U**p-arrow for orientation (**U**pdate it); by a Reference **D**irection (Up**D**ate it);
Maintain a trend, a long straight leg; **A**lign the **A**rrow when you stop; **P**encil in the **P**rogress, and a new leg direction.

◦ Other **M**apmaking **M**ethods — Use the letters of COMPASSES:

Copy a Map Top-Down; Own Map is Valued; Mental Map—Map to Mind Navigation; Parallax onto a Running Baseline; Accumulate the Layout Blind; Sine-Cosine Tally of Accumulated Eastings and Northings; String Map; Exploring a Cave System Needs a Book, of Strip Maps.

"N.S.E.& W." — the four cardinal directions, is a mnemonic of fours

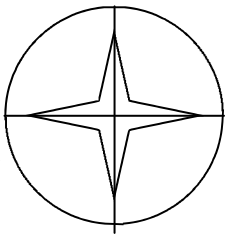

N for "**N**orth Coast" — all about visualising the invisible directions…

Name which coastal towns are N. S. E. W.
"**S**weetheart" features Name your favourite spots NSEW of you
Encircling horizon features — **Horizon** North; **Skyline** South; **Environmental** East; **Landscape** West.
Which-way-am-I-going trend (rhumb-line), from the map, then marked onto the terrain—e.g. to behind that hillock

S for "**S**hortest **Shadow S**outh" — all about the path of the **Sun** (and Moon)—the 1234 system
— and the **S**unny **S**kies quadrant system
— How the Sun, Blue Band trend & ends, and your Shadow, currently lie in relation to…

N Highest **Noon** (How High?)
S Shortest-**Shadow** Souths (SH) & the two-flaps South. Coldest South is the best reference in the SH. String South using the Sun's Shadow.
E Sunrise (Where?) Anti-sunrise from a string-line to the horizon.
W Sunset (When?); Bisect them. Anti-sunset from a string-line in the afternoon.

E for **E**quinoctial **E**ast—i.e. other things astronomical, **E**tc.

Needle-North-Seeking-Northern-Canada—over Red Hot tropics!
South Pole Star (How High?) & Southern Cross (Where?) & Star-Chart octants & Star Compasses & Star Fixes
Ecliptic & Equator & Equinoctial Precession
W for the motion of the Pot, s**e**tting in the W**e**st (rhyme), at the latitude angle

W for **W**ind **W**est, i.e. Natural Directions.

N for Nature's Norths: Rock Noon; Botanical Noon; Solar Panel Noon; Satellite Dish North
S for Snow South; Frost South-East; Steep Gully South; Mud-in-the-Shadow South; Dead South; Siesta South.
E for Moss East-South-East
Wind: Upwind: scorch, erosion & flats, Downwind: shelters, debris & steeps; Noise & colour from up-wind; Rumble & skyshine from downwind.
& **W**eather Directions (& see below)—Where is Wind West? What is Upwind? Upper Winds Steer the Weather (from the West?) What is Upweather?

Summary: 4 directions with 4 variations; Then finish with **W**.E.A.T.H.E.R. After a rest, re-direct yourself.

W.E.A.T.H.E.R. — Check for Change

W. for Wind. Has it changed — in strength, direction, gustiness?
for Wind — in Your Face — Low on Your Left. Low winds bring the rain, clockwise.
for "Weather systems come from the West" — in temperate latitudes — can you make a prediction?
for the upper level steering Westerly-Wave Winds — are they deviating much from average? Are they steering a storm toward you?
for "Wind-West" showing up in the vegetation, for when the prevailing wind is not blowing.
for the prevailing Westerlies (or whatever).
E. for Ephemerals — i.e. the clouds and contrails (don't forget this E.!)
Have they changed? — in height, type, wateriness, thickness, darkness — in directions — steering wind directions? In relation to the Sun?
— Examine them carefully for dissipation (i.e. thinning, soft edges, layering) or for building up (hard edges and heaping).
A. for Atmospheric Pressure Change — imagine the isobars across the land from the forecast.
Sometimes you can see a lowering atmospheric pressure when marsh gas starts to bubble in cold weather i.e. not due to warming.
T. for Temperature — of the air-mass
H. for Humidity — the absolute one (measure the dew point — will it frost?). For the relative humidity try huffing to produce a fog, or onto metal.
Does the air mass humidity & temperature, speed & turbulence let you know which direction the wind must be from?
E. for Expectations — i.e. the weather forecast — test it — is it coming true?
R. for Rain — Will it? R.ight? Are you **R.eady** to move? Grab your R.aincoat and let's go… Keep looking until it is time to S.T.O.P. for a rest…

Oriented? No?
Rule one: Stop! getting more lost.
Rule two: Start! getting unlost. Start navigating!
Rule three: Stay, or Steer a Straight course.
Rule four: Don't do it again!

Position/**P**lacement/**P**rogress—has it changed?
Have you arrived somewhere?
Is it a named feature? A campsite? A junction?

Have you changed view?
Have you changed catchment?
Interpret the Why of your stop...

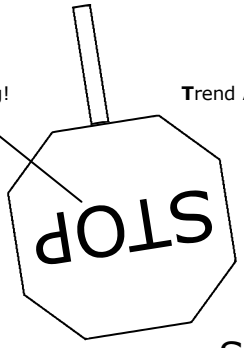

Trend And **T**errain—Has it changed?
Time & Tired—Does that explain your stop?

Stop -Signal: Why have you stopped?

REST and R.E.O.R.I.E.N.T yourself

Take 5; a 5 minute rest; REST is a mnemonic of 5's.
RE-RELATE; **RE-ORIENT; RE-member**; **RE-lax;** (= the 4 new RE-'s before you) **RE-start**. = 5 RE's

RE-RELATE to the skyline.
RE-RE for Re-Read the map,
L for Re-Late the Map to the Landscape
A for Astronomy—relate the Sky to the Skyline
T for Trends—imagine the Trend of invisible landscapes, in relation to the Sun, and relate your current trend to a point on the skyline
E for Each direction clue Re-lated to Each other direction clue, especially to the landscape/skyline shapes.

RE.O.RI.ENT — **RE** for Re-Orient the Map to the landscape,
— **O.** for **O**ops! Don't make logical blunders; Have questions on your mind; notice; speak up; integrate; connect; interpret.
Observe the **Overall** trends and bends.
— **R.I.** for **R**e-Visualise the **I**nvisible (Next checklist; and see Section 1 Astronomical Overview; Include the clock-sense mnemonic here)
Re-late yourself to the Big Picture
— **E.N.T.** for an **E**nvironmental **N**otice **T**ime — Re-late yourself to Nature — coming up.

R.I. For Revisualise the Invisible (and then Relate It to the skyline!)

Work down from the Biggest Picture. Work most on the ones you can't yet visualise.

Astronomy — Visualise the paths of Sun/Pot/Moon/Stars; Point to the South Pole Star; North Pole Star; the **Clock-sense** of stars around each Pole
— The star opposite the Sun during the Day & during the night; the Southern Cross during the day; the Ecliptic; the Milky Way
— Earth rotation; Horizon rotation; horizon trails; clock-sense of Horizon rotation; **clock-sense** of low stars' movement around horizon
— the Moon-Sun, and Ecliptic line; The Moon phase explaining the present tidal force. See the overview at the end of section 1.

Global — Latitude; Polar distances; Equators; Earth Poles; Seasonal Sun height; Shadow South — visualise the aurorae; Antarctica;
— North Magnetic Pole in Canada, as a 'South-seeking' Pole; Aurorae; Unequal Tidal bulges. "Go Overseas" p.25e.
— Tropical countries — corresponding to sunrise, noon, noon height, sunset, the country under the Sun now, the time zones
— Light & Dark halves — Sunrise-/sunset-circle countries; Half the arc-distance; How the horizon is shifting due to your movement.

Continental — Weather map isobars—visualise them across the continent — the one in front of your eyes, and on a continental aerial photo.
— Latitudinal wind system and pressure system helices.
— North Coast; N-S-E-W townships; NW-NE-SW-SE features. "Walk across the Continent" p.29e. "Compass-needle Continent" p.9e.

Local — Catchments; Divides; Mouths; Mounts; Radio Transmitters; Night Lights
— Nearby landmarks & places & highways; Home; Car
— Perimeter features; Escape targets & Routes — Bird's-eye layout of your route.

Clock-Sense Mnemonic

13 pairs of points

0 The standard viewpoint is "seen from above"; ◦ Clockwise is to your right; Compass Bearings Circle Clockwise;
1 Wrong Way Go Back! **Anti clockwise is ENWiSE** ◦ ◦ **SWiNES** That's how the Sun moves in the northern hemisphere
2 Unscrewing is an anticlockwise motion (on a right-hand helix) ◦ ◦ Screwing in is a 'right-hand' motion (for a right-hand screw)
3 24 hour clockfaces have only 15° per hour, like the Sun ◦ ◦ one hour = **30° clockwise**, on the 12 hr clock face method
4 High pressure systems circle **anti**clockwise in the SH ◦ ◦ **'Low' winds** come clockwise, in the SH.
5 Things on the **Left** have a **Lesser Bearing** ◦ ◦Things on the **Right** have a **Bigger Bearing**
6 All the solids rotate **anti-clockwise,** seen from the **N.** ◦ ◦ **Low stars** all around the horizon follow the Sun-sense
The ground under your feet shifts anti-clockwise in NH ◦ ◦ The ground under your feet turns clockwise (SH)
7a&b So the ho**RI**zon **N**ever **H**eads **RI**ght in the **NH** ◦ ◦ The ho**RI**zon **SH**ifts **RI**ght in the **SH**
8 Wind-shift due to ground friction opposes the Sun-sense ◦ ◦ Coriolis forces follow the Sun sense
9 **North lies between the SH Sun and the hour hand** ◦ ◦ The NH Sun is between the hour-hand & South
(actual noon points to Sun; clockface tilted parallel to Equator) (noon-time to Sun; axis to Polaris)
10 The SH Sun seems to move anticlockwise ◦ ◦ The Moon goes backwards, night by night (clockwise, in SH)
11 Nose Nav'n turns everything away **NOTLEast** Nth on the left=E ◦ ◦ **NORWest** — Needle-North on the right = heading West
12 **Shadows** Follow the Sun-Sense. + You can't cross a stream twice in the same flow-sense

---

E.N.T. for Environmental Notice Time —Orient yourself to Nature

Take **a top-down approach**...

Geology; geo-trends; geo-morphs i.e. landforms; geo-Norths
Dividing ridges; dominant peak; upstream catchment size; wind shadow; rain shadow
Season; climate type; prevailing wind; average rainfall
Rocks; bedrock; outcrops; parent rocks e.g. upstream
Soil types; nutrient availability
Soil Drainage — good or poor; Water Sources; groundwater? salinity
Aspect i.e. which direction does the slope face, in relation to high noon?
Living stresses? — salt, minerals, wind, exposure, frost, heat, low humidity, fire, flood, wet roots, pests, diseases, competition, grazing, trampling, disturbance, rock instability.
Vegetation type; the overall density of green, and of cellulose, and the height; dominant species; associations of plant types; diversity; individual species; individual plants. Pollinators available; seed dispersal mechanisms operating.
Mammals; Birds; Reptiles; Insects. Drinking water availability, cover for animals, hunting pressure
The 'Natural Norths' — Gully North; rock North; Bush North; insect North; Dew North; snow North; ripple North; prevailing wind North

---

RE-MEMBER to Note down the Names in the Notebook

Remember the **Way**, and the **Sequence**.
Remember to take a Notebook. Make a neat permanent copy of your notes i.e. in ink, maybe sometime later on, in your master track notebook.
Be Bit-wise!

Name the subsections, junctions & links. Name the peculiarities. In the name, try to include…
The Why, When & Where of your stop. Link & Sequence information — time, distance, context
Lie of the Land Words to indicate the direction trend, & curves — left or right
Visual words — colour, shape, alignments, curves Descriptive words. Collect the details. Enlist the others to describe and name it.
Give yourself something to remember — create an acrostic; craft a creative name; sit down and write up your track notes
Update any acrostic alliteration. And the count e.g. of the creek crossings. Do something numerical. Do something alphabetical.
Be vivid, clever, memorable. Do something Shapely or Geometrical there Do something Memorable there. Do something there.

Draw the shapes, outlines & skylines. Draw the directions of other landmarks from any high viewpoint. Take reference photographs.

Remember most of all to do it all — don't just recite the reminders to do it all!
Mnemonics of Fives: Remember the Way; the Sequence; Make Notes; Invent Names; Draw What You Need to.

Re-lax, and Play — Navigation Games and Activities (5 pairs to recall)

Walk & Talk — a night walk — a fog walk — a walk off the track — climb a hill with a small scale map & binoculars & compass.
"Track My Trail" ("If you want lunch, you'll have to find me — give me a quarter of an hour, then track me down").
Keep a compass course, blind — keep a straight course, blind — on a slope? "Let's all make an acrostic to remember this track by".
Try Orienteering — Try making a map of a maze of trails — then a string map — a strip map — or a sine/cosine tally map.
Look & Learn — "Go to the front of the line if you spot a track marker which the others have missed" In a windy place look for wind clues, etc.
Take it in turns to name a navigational or orientation clue which is visible, until you have to drop out.
"I notice… (fill in the blank). Your turn. What extra can you add (to the description, or to the clues)?"
Guess & Test — Each of you estimate the time of arrival; the direction to the car; the distance to a landmark — then check the map.
Guess the time from the Sun, Moon, Southern Cross, or other stars, then check your watch. Use your compass to check your guesstimation of North from the bush, gullies, rocks, a star-trail, a star-set etc. Predict the tide, or the time & direction of moonrise — then sit down and wait for it. Predict a star-set position. Find your 'shortest shadow' angle today & use it tomorrow.
Make & Use — an odometer from a wheel, a pedometer, a sundial/declination compass, a portable sundial, a string rangefinder, a slope-meter.
Measure — Try some star navigation observations — Rise & set bisection; Use a plumb-line; a bush quadrant; a bush sextant.
Test how accurately you can each use a compass. Test each compass against the others.
& Memorise — yesterday's acrostic. Remember all the mnemonics! Can you add to any of my checklists? Have a star identification evening.

Moon-Phase & Tide-Time Dial

This is a handy tide-predictor, based on the assumption that tides will follow the moon. It shows you where the moon will be, at high tide. It also tells you what the moon is doing if you know the date but haven't seen the moon for a while. The deluxe version can predict tides and phases years ahead of time. It consists of a series of dials rotating on a single pin.

◇ The idea is that many of these concentric cardboard/plastic dials rotate together once properly adjusted, and as quickly as the Moon orbits the Earth. They show the Moon Phase, the Sun, the Earth and the resulting astronomical tide-producing forces.

Each disk is loosely fixed to a neighbour with friction (e.g. a smear of dried silicone glue) or a clip or tape or a pin, to allow adjustment.

Set it up by observation, or by tide-table predictions. Add notes, as you accumulate information about locations, tidal range, etc.
It can be especially useful for travellers, e.g. when fishing an unknown coast.
If the dials are correct for any one tide (spring-tide is best), they should be <u>correct for all the tides thereafter</u> *
* give or take a couple of hours for the smaller effect of the position of the Sun.
The effect is 43% of that of the Moon, and can be estimated with practice—change the hour reading to <u>toward</u> the Sun, or the anti-Sun, a little, if the Moon is <u>near</u> the Sun-anti-Sun line (no more than 45° from). Local shallow water resonance effects will also alter the actual effect of the Sun. When in doubt, around neaps, try averaging the tide-time for one or two days ahead and 1 or 2 days after. The theory is that slosh resonance, or the Moon alone, or the Sun alone, might predict the difficult tides better.

◦ The <u>hanger</u> (outline not fully shown) is hung from the wall and also folds down behind to receive the pin on which the dials rotate.
It doubles as an indicator window, and should mark five consecutive high tides (ignoring a.m./p.m.).
◦ The <u>clockface</u> is 24-hours i.e. 15° per hour clockwise, (markings not fully shown). Think "a.m./p.m." for the <u>two</u> high tides per day, no matter what the reading says.
Each <u>new location</u> visited may need the clockface to be twisted slightly to allow for <u>local tide delays</u>. Keep a note of such settings.
Daylight saving requires a +/– 1hr change — to relocate the Sun forward in summer, back to normal in winter.

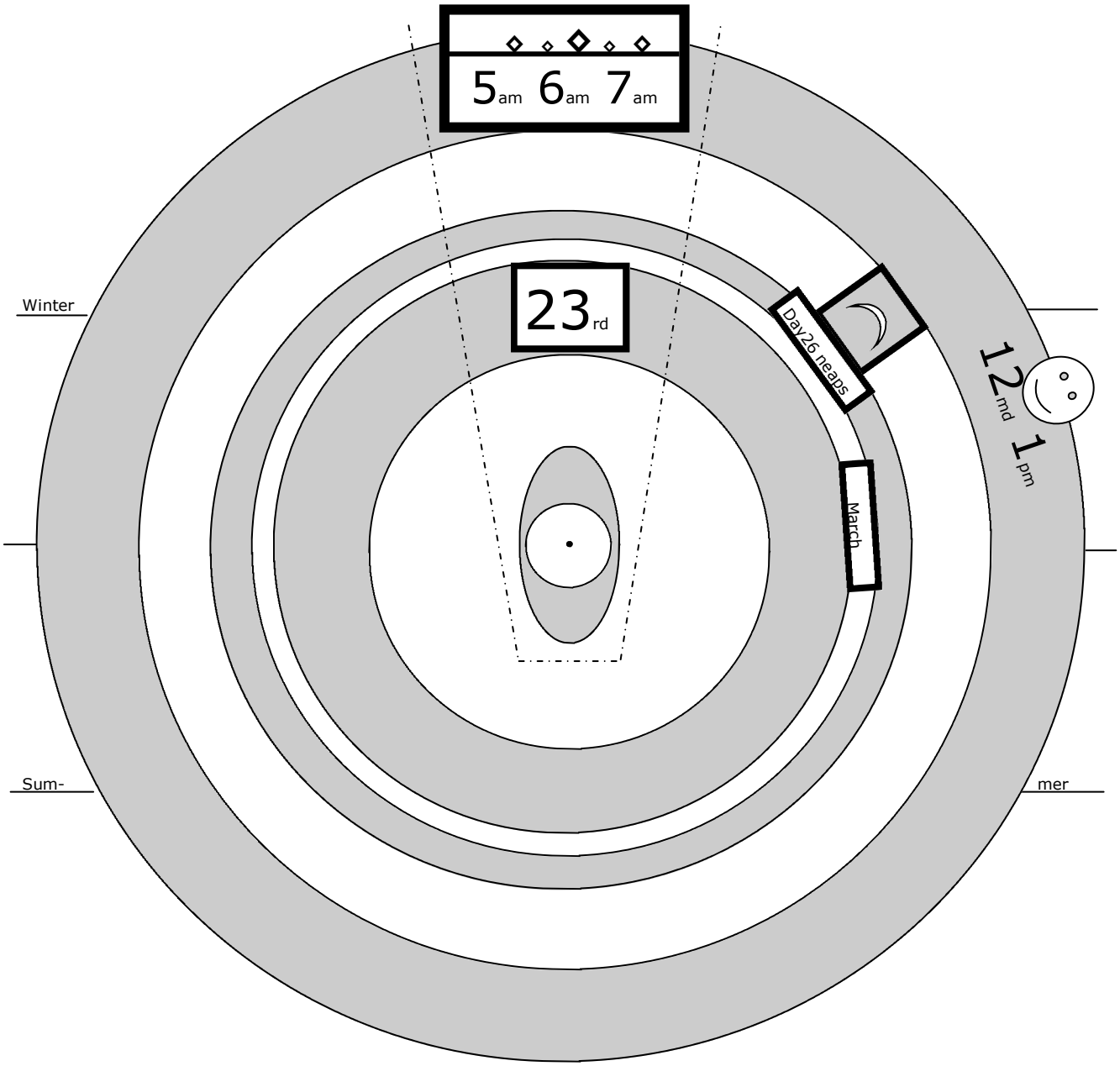

◦ Friction-fix the Moon-window dial (white with grey window) to the hanger, so that it maintains its position. It covers over all the other Moon phases.
◦ The Moon-phase diagrams (29 of them) underneath that, on a separate dial of the same diameter, should look right for the Southern Hemisphere—i.e. the Sun rising in the East, illuminating the Moon; all the phases parallel. Make the Moon-shape highly visible from a distance. Attach the Sun to the No-Moon phase, Day 0, via a transparent link; then set the Sun to the Clockface, to indicate true noon—when the Sun is exactly North.
◦ The optional tide-type ring (showing "neaps"), set to the Moon-phases, allows you to adjust the delay after new/full Moon, before spring tides.
If you also indicate which high tide is dominant—the one nearest the Moon or the one farthest—then that will need adjustment when the Sun crosses the Equator. I suggest that you set the top tide for the tide when the Moon is visible, to give you a visual feel for the tide producing forces.

◦ The date-dial is 29.5(30526805) days around—the complete lunar month—and marked clockwise from 1 to 32 (only one date is shown in the diagram). **Dial up the date, by rotating the Sun**. (For a 12-hour clockface, you would need to double the 12.19077473° per day). Since the next high tide will (if you are looking in the daytime) be on average that afternoon, I suggest you think of the date number as "the middle of the 23rd"—to avoid adjusting the date hand minutely for how far through the day the next tide will be.

◦ The month window is attached to the date dial, at Day 1 of the month, if you include a month setter.
◦ The optional but useful month-setter dial (markings not all shown) starts at March 1, to avoid leap-year adjustments; It makes March 32 = April 1, April 31 = May 1, and so on—spanning approximately 11 days for the year. Restart every March 1, and reset it every month. Calculate the monthly positions by the day number since March 1, avoiding an off-by-one error
(17.9°;23.6;41.6;47.3;65.2;83.1;88.8;106.7;112.5;130.4;148.3). Set the month name to the first day of the date dial.
You can also calculate and indicate a one-year increment, 129°.6327779, a leap year, 141°.82355, and a four-year increment, 170°.7218865. To make an adjustment, make the old date indicated by one mark on the month dial, now line up with the increment mark selected, so as to indicate a date in a new month/year/leap year cycle.

◦ An optional Earth ring in the very centre, can show the entire Southern Hemisphere map (not illustrated, or NH on the reverse side), and be set to the right longitude. It is fixed to the hanger, and to the tide-bulge ring...
◦ A second ring just behind the earth can indicate two tidal bulges, one pointing upwards (or otherwise if you want to indicate the local tidal delay compared to the actual ocean bulge).

Mock it up in paper, until you understand the functioning.
Glue those paper discs onto cardboard to trial the thing for a while.
After deciding on improvements, make it neatly, to last for decades.
E.g. Try getting the computer to print the dial markings onto acetate, and glue that to sheets of stiff plastic.

You can experiment with different dial orders, e.g. the date-dial on the outside of the phase-dial.

You can make room for comments,
e.g. by expanding the dial area showing, to say when the moon rises and sets; e.g. "light from midnight onwards"—beneath each Moon-phase diagram
e.g. when and where the Sun rises and sets; Mark three horizontal lines in the background for Summer, Equinoxes and Winter.
I.e. mark them on the backfold of the hanger piece, or on the wall. Try to match midwinter sunrise time to where the "winter horizon" cuts the clockface, which will also suggest the approximate direction of sunrise.
e.g. when and where the first, second and third quarters set in each season; Full Moon opposite the Sun, i.e. 6 months different; Half Moon Waxing is one season ahead of the Sun; Half Moon Waning is one season behind the Sun (as far as which horizon line to use).

You will discover that an invisible connection needs to be made between the outer top window and the inner window and dials, so that they do not rotate with the other dials. One way to do this is to push the pin through a piece of rubber eraser, and glue that rubber to the dial. Another method is to have a square pin, and key the inner dials to it, while using round holes for the rotating dials. A third and simpler method is to have a transparent link from the centre, over the top of the other dials. You do not want the hanger to obscure any diagrams or dials which need to be visible.

The Time and Date dials are the indispensable ones. Instead of dialing up the date, you can use a normal clockface (12md always on top) and an hourhand to point out the next tide; you point the pointer at the date. On a 24hour clock-face it can be made like a + sign, to point out all 4 tides that day.

Here is the simpler alternative version, showing an 11:30 high tide on the 19th…

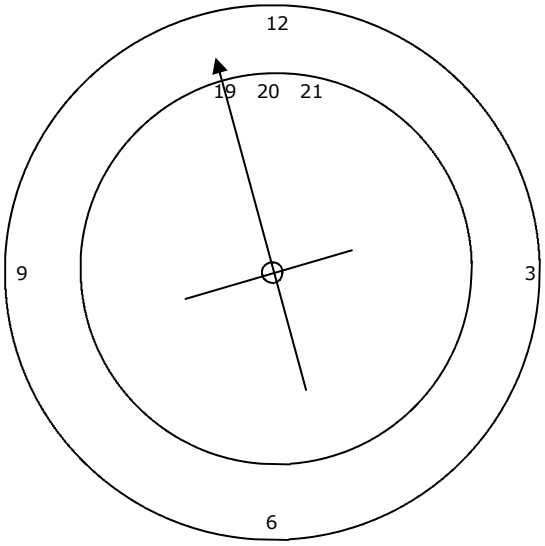

Divide the Star Globe into 8 Octants—and Learn the Stars in Each

The three dividing great circles are the Equator, the Ecliptic, and the Milky Way Galactic Plane.
A star may be North or South of each line, or on one or more of them. The lines on the backside of the globe are shown dashed.
The labels refer to the three lines in the order given. Six of the octants show in the front view with the two in parentheses on the back.
Verify each label before you start. E.g. NSN = North of the Equator, South of the Ecliptic, North of the Milky Way.
Compare with a good star chart e.g. out of a large atlas, as this is only a memory aid.

Polaris

The Milky Way Crosses the Equator
at the centre of
Procyon-Sirius-Betelgeuse

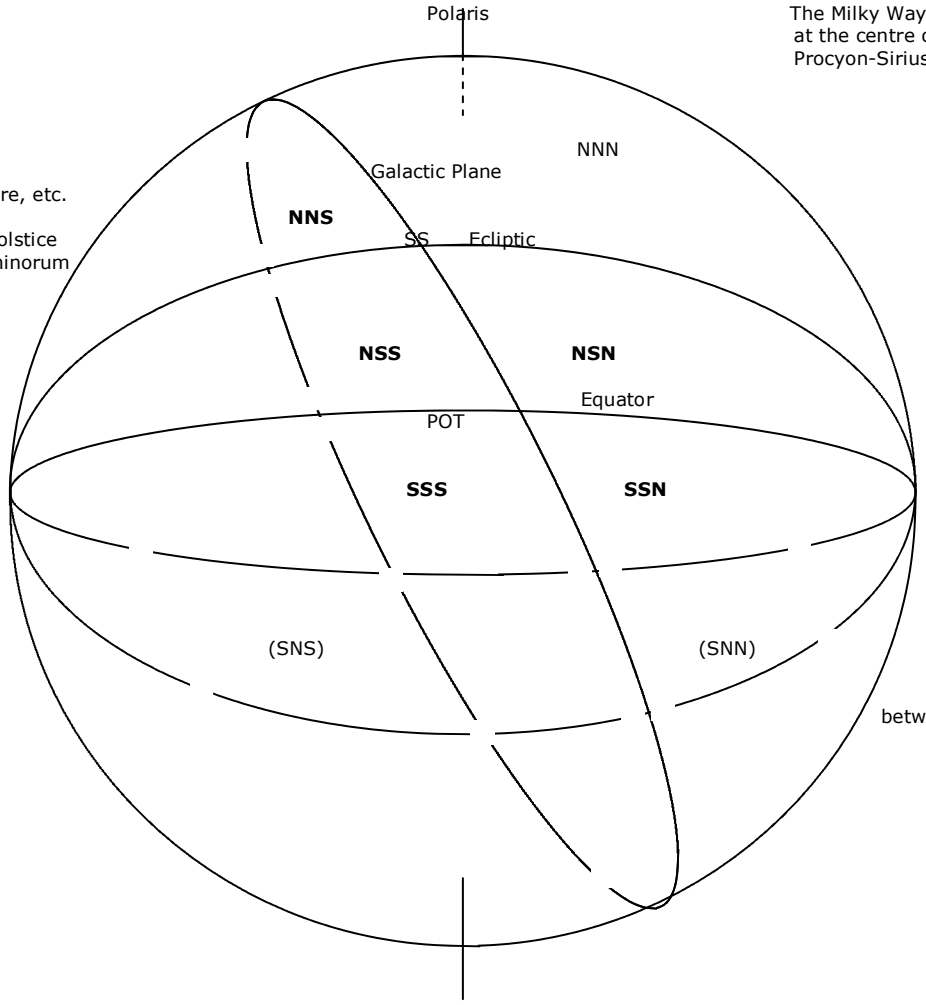

NNS—The Great Square, etc.

Summer (Northern) **S**olstice
Elnath to Gamma Geminorum

NSS
Aldebaran
Betelgeuse
Bellatrix
Menkar

SSS
Pot
Big Dog
Canopus
Diamond Cross
False Cross
Acamar
Achernar (central)
Magellanic Clouds
South Galactic Pole
South Pole Star
Diphda
Ankaa
Fomalhaut
The Crane
Peacock
The Archer
Scorpion's Tail
Atria

NSN
Procyon
Gamma Geminorum

SSN
The Centaur
Antares
Spica
Corvus
Suhail
Alphard

between SSS & SSN
Southern Cross
Southern Pointers

South Pole Star

The Southern Figure of Eight—Two Star-Circles—joined by the backbone of Canopus—Achernar—Fomalhaut, a straight line
(not to scale; consult a star chart to verify each star, then find them in the sky)

Fomalhaut
Achernar
Canopus
Sirius |
Orion's belt | these 3 are also straight
Aldebaran |
Pleiades
Aries
Great Square

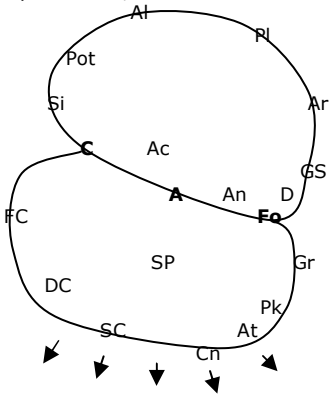

Canopus
Achernar (find Acamar on the way)
Fomalhaut (find Ankaa on the way)
Great Square (find Diphda)
Aries, the Ram
Pleiades, The Seven Sisters
Aldebaran, in the Bull
The Pot, Orion's belt
Sirius, in the Big Dog

Milky Way
Cn=Pointers (alpha and beta Centauris)
Southern Cross
Diamond Cross
False Cross

Around the South Pole Star
Canopus (find Acamar next)
Achernar (find Ankaa next)
Fomalhaut
Grus, the Crane, with Alna'ir
Peacock then Atria (find Scorpio)
Centaur, the two Pointers (find Arcturus)
(find Spica)
Southern Cross (find Corvus, the Crow)
(find Leo)
Diamond Cross, False Cross, Canopus

Ach.-Fom.-Pck form a triangle to find AlNair & Ankaa

Visible in its entirety in October plus or minus a few months.

Divide the Star Globe into 8 Octants—and Learn the Stars in Each

The smaller groups, and the crossover points, are worth learning, because they show you the three lines best.

NNN
Alkaid (central)
Capella
Twins
North Galactic Pole
Big Dipper
Little Dipper
Polaris
Arcturus
Alphecca
Rasalhague
Draco
Vega
Deneb
Most of Leo
(Regulus adjoins NSN)

Summer **S**olstice
Between Aldebaran & Pollux

Autumnal **E**quinox
Between Spica
and Regulus

NNS
Altair
Enif
Great Square
Mirach
Almach
Algol, Mirfak
Aries
Pleiades
Elnath
(Caph, Schedar, Ruchbah)

Winter **S**olstice
Nunki to Sabik

SNN
Zubenelshemali
Sabik
(Zubenelgenubi adjoins SSN)

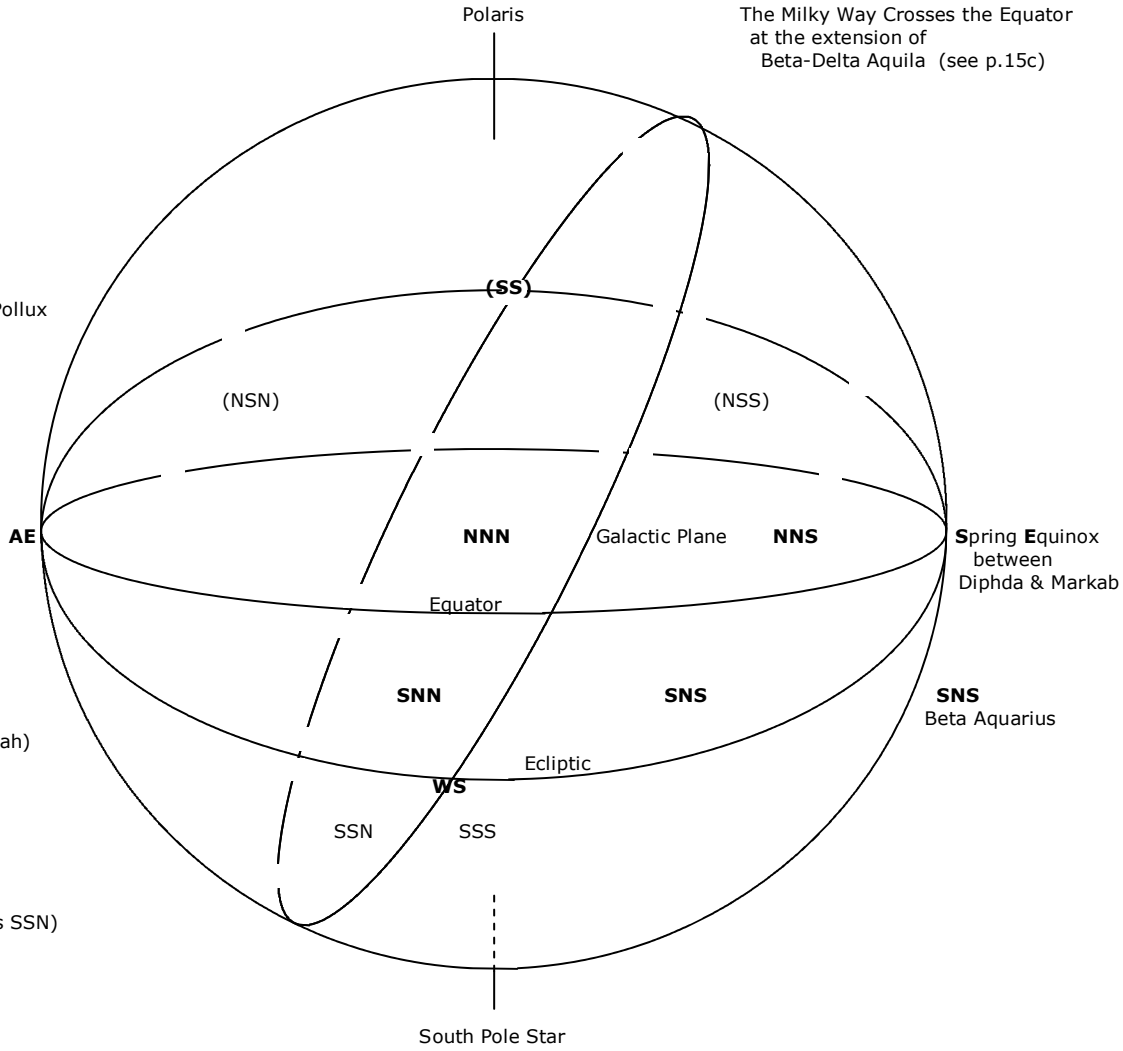

The Zodiac i.e. the constellations along the ecliptic. In order of appearance. Not all of them are bright or obvious in shape to look at.

Mnemonic:	Are	Tall	Twins	Cancerous	Leo?	Virginal	Liberated	Scorpions	Shoot	Capricious	Aquarium	Fishes.
i.e.:	Aries,	Taurus,	Gemini,	Cancer,	Lion,	Virgo,	Libra,	Scorpio,	Sagittarius,	Capricornus,	Aquarius,	Pisces
	Ram	Bull		Crab			Scales		Archer	Goat	Water-Carrier	

Morse Code & Rhythm Mnemonics

Dit/Dah/Pause is an extremely flexible system. Adapt it to torchlight, sounds, sewing stitches, body movements… A lot of fun for all the family.
if dit = one time unit; then Dah = three dits long; the di-Dah separator = one time unit; letter separator = three units long; word-separator = 6

	Say…		**Write…**	**Signal…**	Special Meanings (only for reference)
A	an A		. _	di Dah	
B	B bi-bi-bips		_ . . .	Dah di di dit	Fog Horn: I am under tow
C	C is Sea-sick		_ . _ .	Dah di Dah dit	Between ships manoeuvring: I a**c**knowledge your signal
	(Capital letters for a Dah syllable; lower case for a dit)				i.e. You may overtake on the side you indicated
D	Dee did it		_ . .	Dah di dit	Fog Horn: restricted manoeuvrability; deep draught; sailing; etc.
E	egg		.	Dit	I am altering my course to... **Starboard**
F	f' f' Forfeit	f' f' Fore foot	. . _ .	di di Dah dit	
G	Gee Gee git	Go Go get	_ _ .	Dah Dah dit	I intend to overtake you, to Starboard
H	h'-h'-h'-hit		. . . .	di di di dit	I intend to turn completely around to…
					Fog Horn: Pilot vessel on duty
I	is it		. .	di dit	I am altering my course to... **Port**
J	a J Goes Jay		. _ _ _	di Dah Dah Dah	
K	K for Kay		_ . _	Dah di Dah	
L	an ELL leph'nt /in ELLement /inELLegant		. _ . .	di Dah di dit	(draw out the Ell sounds)
M	My My		_ _	Dah Dah	Fog Horn: "under power, not making way"
N	No en		_ .	Dah dit	
O	Oh Oh Oh		_ _ _	Dah Dah Dah	
P	a P Goes pip		. _ _ .	di Dah Dah dit	
Q	Q Forms a Queue		_ _ . _	Dah Dah di Dah	(to the tune of "Here Comes the Bride"—Queue up to get married!)
R	an R dits		. _ .	di Dah dit	
S	s-s-s	ess ess ess	. . .	di di dit	I am under Astern propulsion
T	Tea		_	Dah	Fog Horn: "under power, making way"
U	it is U		. . _	di di Dah	
V	v' v' v' Vee		. . . _	di di di Dah	(to the tune of Beethoven's V)
W	a White Wine	a World War	. _ _	di Dah Dah	
X	EX mimics P	All ex's Hurt	_ . . _	Dah di di Dah	(drawl the EX, to lengthen it in the first, not in the second)
Y	Why is Y Y?		_ . _ _	Dah di Dah Dah	(say the 'is' quickly)
Z	Zee Goes zi-zit	Zees Go zi-zit	_ _ . .	Dah Dah di dit	I intend to overtake you, to Port
Ü	it is U too	(this pairs to Zed)	. . _ _	di di Dah Dah	(U umlaut or U diaresis. Invent a meaning for it.)

Hint: For Learning, repeat each letter thrice.
By the time the other person recognises the first, the second has finished and he is waiting to check it out against the third.

0	No No No No No		_ _ _ _ _
1	it Carts Four More Dahs		. _ _ _ _
2	it is Two Two Two		. . _ _ _
3	di di di Makes Three		. . . _ _
4	di di di di Four	pitter patter 4 (your puppy's feet)	. . . . _
5	fit 'em in a mit	(your five fingers, that is)	. . . . .
6	Five si– si– si– six		_ . . . .
7	Five Seeks seven dits	(=567dits drawled on the 6)	_ _ . . .
8	8 8 8 bi bits	(8 binary bits in a byte)	_ _ _ . .
9	9 9 9 9 nips		_ _ _ _ .
	Long dash	Daaaaaaaash	_____ e.g. (Fog Horn or car horn) before you enter a blind bend!
	Error	it isn't a bit of it	. . . . . Between yachts: "I do not understand/agree with your intention/action"
:	Colon	Now Here Comes in a list	_ _ _ . . .

Learn the next ones in <u>pairs</u> i.e. swap dit for dah

.	a STOP a STOP	(i.e. don't pronounce 'stop' as short)	. _ . _ . _
;	SEMi SEMi SEMi	(i.e. drawl the first syllable)	_ . _ . _ .

These 2 shapes are convex to the right

,	Go Go comma Go Go	Dah Dah di di Dah Dah	_ _ . . _ _
?	did he Dah Dah did he?	did 'e Blah Blah did 'e	. . _ _ . .

These 2 come in pairs normally

""	a Quote is a Speech bit	di Dah di di Dah dit	. _ . . _ .
()	Here's the Blah Blah bit Here	There's the Blah Blah bit There	_ . _ _ . _

These 2 are in the middle of a word ..hyphen and apostrophe

-	'Aitch in the middle Daash	('Hyphen' starts with an 'aitch)	_ . . . . _
'	it Saves All Those Hard bits	di Dah Dah Dah Dah dit	. _ _ _ _ .

/	Slaash t' diVide it	(Drawl the first vowel) Slaash is a Slaash bit	There was the Whole bit _ . . _ .
Á	an A Sharp an A	You could use it for "Repeat":	re Peat Whole re Peat . _ _ . _
É	it is E a-cut(e)	. . _ . .	cute is cut short (You could use it for "Bad Reception": cannot Hear the bits)
Ñ	Nya Nya ny' Nya Nya	_ _ . _ _	(with thumb to nose, fingers waving) "ny" sound (I use enya for "Do You reCeive Me")

SOS	a bit of 'Save Our Souls' isn't it?	(no letter breaks; repeat regularly)	. . . _ _ _ . . . . . _ _ _ . . . (Life & death only)
XXX	Come quickly Come, Come quickly Come, Come quickly Come		All ex's Hurt All ex's Hurt All ex's Hurt (When you Need urgent help)

Signals procedures, developed by hard experience, short circuit a world of miscommunications — stick to them. I.e. learn <u>them too</u>.

I'm Looking For Talk; CQ CQ CQ;	Calling Calling All Radios; Hear ye Hear ye All AuditOrs; Tune in Tune in Lend Me y'r Ears
From; D, E; De = French for 'from'	Heard of it ; fr'm Dah-di-dit dit
Go Ahead; invitation for <u>anyone</u> to transmit; K	Call in Now
Break In B, K	Break in to it; Make a Break ("BK DE (identify yourself)" = "let me use this channel")
Starts; Attention; Message Coming	Start it Start it Now Start of My bit Now _ . _ . _
End of Paragraph; Pause; new thought; double dash	Pause if it is New _ . . . _
Wait a Wait f'r a bit . _ . . .	You could invent "Speed up All Your Work" to pair with it.
Over; invitation for <u>your</u> contact to transmit	O ver Your Turn stop _ . _ _ . (to start a person-to-person contact: ###DE### Over)
End of Message in traffic; not end of contact; not clear of station	the End of My bit; con Clude con Clude it . _ . _.
R; Roger; Received and Understood <u>all</u> of Message	
End of Work; but still on air; clear of contact	it is the End of Work (to the tune of… here we go Out to Play) . . . _ . _
Clear of <u>all</u> contacts; shutting down; Clear of Station; C, L	Clear it Clear it, b' Clear of it

Add your own signals for procedural or other matters. E.g. "Please repeat my message back to me for a check" "Adjust the beam left"

Ä	an A an A	. _ . _	an A 'm Laut It pairs to C Invent a use for it like "i Don't a-Gree" to be opposite to C, I Agree
Å	an A an A too	. _ . _ _	an A an Aang Strom
Xhi	Khi Khi Khi Khi	_ _ _ _	K-H-Like-H soft Kh sound (I use it for "to" i.e. T+O, after DE "from") cf Ree Call All Souls
Ö	Oh Oh Oh dit	_ _ _ .	Goe– Goe– Goethe It pairs to V You could use it to say Sorry

North American Morse is somewhat similar. There is also a Japanese version. Morse is no longer supported as International.

Semaphore Signalling Code & Mnemonics for it

Semaphore is a clockwise system, as viewed by the decoder, and it starts from below.
When the two hands can't be seen, that is the neutral position, used for a… **Pause**
The black square represents the signaller's body; the line(s) are his arm(s).
The numbers 1-9 and 0 are the same as A-I and J. The first 7 single-arm positions are shown…
You can tell the difference, because the signal "numbers coming" precedes a number, and a "pause" follows the end of the numbers.

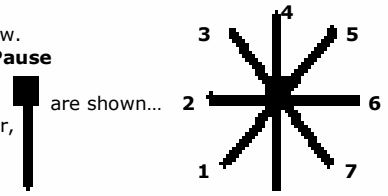

One arm, pointing up, therefore is either D or 4,
Think of God as the 4th Dimension to your life, and remember 'up' as "**4D**"
Or if it sticks in your memory, think perversely, that "**Up is fo(u)r D.own**!?"
Learn the letters in symmetrical pairs.

Here is a "**B F**lat" mnemonic. One arm flat is either a B (-flat) or an F (-flat)
"2 be or not 2 B?" and a "**F Sixth** chord" might help you remember the numbers.

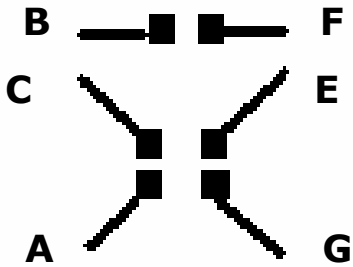

Wave the C to say "Replying". Wave the E to signal "Error"
(cf Morse "C" for "Acknowledge")
The *odd number* positions are diagonals.
1,3,5; 7… **One is an ACE** isn't it? And Dia**g**onal has a **G** in it doesn't it?
a G7 chord to follow the F **F**lat minor **6**
ABC = 123 is easy enough to see, and **5E follows 4D** makes the E position easy

2 Hands…

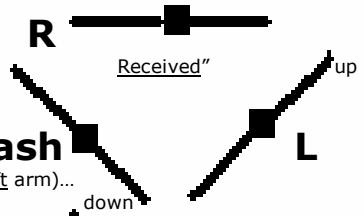

"Roger" is simply an indispensable signal to learn… it means "Yes, I understand, OK, Received" up
For Learning: Acknowledge every letter understood, with a Roger

2 Hands Straight:
Think **"Long**, then **Left down = Slash down** or **L**eft **up = L**ift **up"** (referring to **their** Left arm)… down

2 hands in a catching formation, above the head…
Shout **"Catch iT…"** and think **"..Tee—numbers coming"**

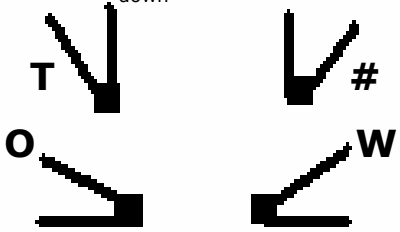

2 Hands in a slip-catching formation, to one side…
Shout: **"HOWZ"**at!

You are meant to wave black and white flags around,
but usually you make do with opening your palms.
Only be careful to position yourself against a good uniform background.

'aitch rhymes with eight

2 hands at a right-angle…
"Ya gotta have the **Right P**y**J**amas" ("PJ"s)
J, the tenth letter, is used for Zero ABCDE FGHIJ

PJ is not in alphabetical order, because in the Semaphore Alphabet J comes after #

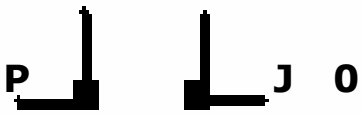

2 hands at a right angle, Vee-shaped, up or down…
Can you see the **shapes** of the **lower case letters n and u** in them?

To attract attention wave U, to say "Message Coming"

2 hands, like a Roman Numeral Five on its side…
I and **X** are **Roman Numerals too.** The I is used for numeral **9 (IX)** in semaphore
The X looks like **the right-half of the X**

So it sort of goes, left to right "nine, ten"

2 hands flapping and banking. I.e. a shallow Vee formation…
"MY Queue**S"** is the clockwise mnemonic
But YQ is not in alphabetic order (like PJ, the only two cases).

You can use Q for "Query" - What did you say/mean?

From the sky, down to earth come… **"K**ilo**V**olts" **(**or **"K**iller**V**olts"**)**

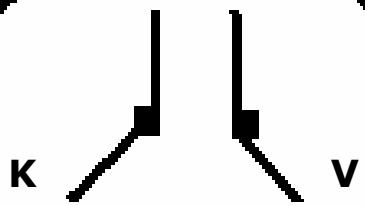

You can follow the clockwise logic of Semaphore if you leave out the JVWXZ till last
Pause, ABCDEFG, HIKLMN, OPQRS, TUY/, #JV, WX, Z

Make each in turn to see the progression.

Appendix: More symbols for map-making Useful for mud-maps, or for track notes.

The first symbol in a strip map column should show the *reason* for the stop. This saves a lot of words and is instant to read.

In turn these are:

A Context Entry; Surroundings; Overall Perspective. E.g. catchment and drainage. It's a chance for a general note.
A Stop. Expect a reason like 'Lunch'.
A Location. Expect a description.
Crossing. This might be a crossing of a ridge, river, road, canal, boundary. Expect notes on what is to left and right.
Leg. The first of the 2 main reasons for an entry: either a new subsection (leg) or the location *between* two sections in its own right (next symbols).
Overlap-section between legs. This is a memory aid, to sequence subsections, to remind you how individual sections join to each other.
Change of legs; new subsection; = notes on *the changeover*, rather than on the next section.
Junction (Y for Y junction, but used also for other junctions). This might be a 'natural' junction between ridges or creeks; or an intersection of paths.
Viewpoint; Change of view; First glimpse; Last glimpse (I for 'Eye')
Corner
Bend
Curved section beginning; Arrow shows clock-sense. This one is a Right-hand bend.
Curved section ending; Arrow should show same clock-sense as previous one. This one ends a left-hand bend.
Change from left-curve to right-curve
Break of slope
Feature
Special feature
Ridge; Crest
Saddle;
Valley; creek-bed; Gutter
Tent; Campsite
Peak
Fix
Cliff
Creek
New grid square

Boundary walks: fenceline; railwayline; road; path; circumnavigation; shoreline
Use symbols which work for you, e.g. Rise or Set direction. Add your own, e.g. Hut

The second symbol ought to show the main direction clue, followed by subsidiary ones as a third symbol

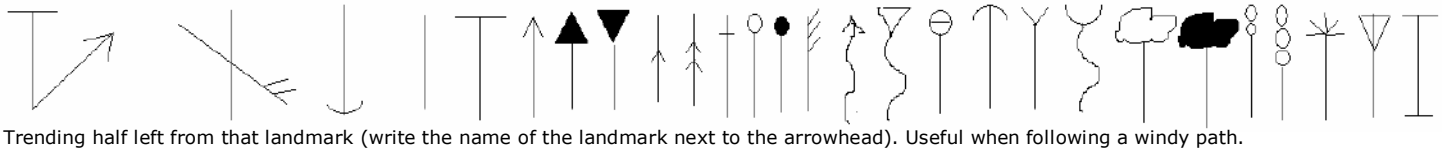

Trending half left from that landmark (write the name of the landmark next to the arrowhead). Useful when following a windy path.
Maintaining half right of downwind;
Walking the crest down ridge — Progress is assumed to be 'Up'wards on the page.
Leg.
Trend direction (for when the terrain is up and down, left and right) ('T' for Trend).
Named landmark; General purpose arrowhead; add the name and the bearing
Peak
Anti-peak direction (e.g. as mentally marked on the skyline opposite the peak)
Up-Slope.
One arrow for 'below average'; Two arrows for 'average'; three arrows for 'above average' — low; medium; high. Use this for other things too.
Compass needle North, modelled after the symbol for North
Up-Sun
Shadow; Down-Sun
Wind source; Upwind. Use the 1-2-3 code for strength. This one is 'above average'.
Down-stream river flow
Up-stream (to a peak) (useful when 'downstream' is behind you)
Moon direction
Up-Ridge; The direction of the highest skyline through the vegetation, if that is all you can see.
Down-valley. The direction of the lowest skyline when that is all you can see.
General Drainage trend (down the valley to the river mouth)
Up-Weather; Low cloud source direction; Cloud Source (when only one set of clouds is visible)
Down-clouds
Alto Clouds source direction (make the circles fluffier than I can show)
Cirro-clouds source direction
Glow at night
Radio-null (from the symbol for an antenna)
Trend of some feature like the coast — write in the details.

1,2,3 for size and importance of junctions…

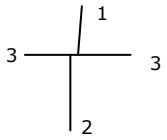

The notebook might be set out as follows: The first line across a narrow notebook is in informal columns; the second is for written notes.

Time ;	**Reason** ;	**Trend** ;	**Directions** ;	**Times** ;	**Distances** ;	**Grid**
Fill in a **number** if you don't have a watch Sequence is a primary need	Use the top **symbols** Start with an 'Overall'	Use the main clue	make a rose of others	elapsed time leg time lost time total time out net walking time	total elapsed last leg length lost distance total diversions net progress	e.g. A5

(Use abbreviations, e.g. NWT, or dedicated columns)
(Do your calculations here at leisure)

Now write *across* the next line(s) the **name; mnemonic; description; comments**
The description of a cross-country leg could include the side or flank, hand, clock-sense, break, steepness, difficulty, curviness, hilliness, perversity, terrain, contours, topography, view, what you can't see, skyline, traffic intensity, track-markers, their age, rock type, soil type, vegetation, biology, colours, peculiarities, reference to another page of the notebook to a diagram.
The more you include in the symbols, the less you need to say, and the easier it is to read later.

Appendix: A List of Star Coordinates.

100 brightest stars
Declination is equivalent to Latitude. –nn = South; +nn = North
Siderial Hour-Angle is equivalent to Longitude West of Zero.

Common Name (s)		Dec.	S.H.A.	Constellation
Miaplacidus		-69.72	221.70	b Carinae
Atria		-69.03	107.83	a Trianguli Australis
Acrux		-63.10	173.35	a Crucis
Rigil Kentaurus/Toliman		-60.84	140.10	a Centauri
Hadar	/Agena	-60.37	149.04	b Centauri
Mimosa		-59.69	168.07	b Crucis
Avior		-59.51	234.37	e Carinae
Aspidiske	/Tureis	-59.28	220.72	i Carinae
Achernar		-57.24	335.56	a Eridani
Gacrux		-57.11	172.21	g Crucis
Peacock		-56.74	053.60	a Pavonis
k Velorum		-55.01	219.46	k Velorum
d Velorum		-54.70	228.82	d Velorum
e Centauri		-53.47	155.02	e Centauri
Canopus	2nd brightest	-52.70	264.01	a Carinae
Muhlifain		-48.97	169.62	g Centauri
Regor		-47.34	237.61	g Velorum
Alna'ir		-46.96	027.95	a Gruis
b Gruis		-46.88	019.34	b Gruis
Suhail	/Al Suhail	-43.43	223.01	l Velorum
Akrab/ Sargas	/Rastaban Sc.	-43.00	095.67	theta Scorpii
Ankaa		-42.31	353.43	a Phoenicis
eta Centauri		-42.16	141.12	eta Centauri
Naos		-40.00	239.10	z Puppis
k Scorpii	(not Akrab)	-39.03	094.38	k Puppis
Shaula		-37.10	096.60	l Scorpii
pi Puppis		-37.10	250.71	pi Puppis
Menkent	(not Menkar)	-36.37	148.33	th Centauri
Kaus Australis		-34.38	083.96	e Sagittarii
e Scorpii		-34.29	107.46	e Scorpii
Ascella /Nushaba		-29.88	074.36	z Sagittarii
Fomalhaut		-29.62	015.59	a Piscis Austrinis
Aludra		-29.30	248.97	eta Canis Majoris
Adhara		-28.97	255.34	e canis Majoris
Antares		-26.43	112.65	a Scorpii
Wezen		-26.39	252.90	d Canis Majoris
Nunki		-26.30	076.18	s Sagitarii
Dschubba		-22.62	119.91	d Scorpii
Diphda	/Deneb Kaitos	-17.99	349.11	b Ceti
Mirzam	/Marzim	-17.96	264.33	b Canis Majoris
Arneb	/Arnebo	-17.82	276.82	a Leporis
Sirius	/Dog (brightest)	-16.72	258.72	a Canis Majoris
Sabik		-15.72	102.41	eta Ophiuci
Spica		-11.16	158.70	a Virginis
zeta Ophiuci		-10.57	110.72	zeta Ophiuci
Saiph		-09.67	273.06	k Orionis
Zubenelschemali		-09.38	130.75	b Librae
Alphard		-08.66	218.10	a Hydrae
Rigel	not Rigil K.	-08.20	281.37	b Orionis
Alnitak	S. belt	-01.95	274.80	z1 Orionis
Alnilam	Mid belt	-01.20	275.94	e Orionis
Mintaka	N. belt	-00.30	277.00	d Orionis

Common Name (s)		Dec.	S.H.A.	Constellation
Polaris	N. Pole Star	+89.26	(322.06)	a Ursa Minoris
Kochab		+74.16	137.32	b Ursa Minoris
Alderamin		+62.59	040.35	a Cephei
Duhbe		+61.75	194.07	a Ursa Majoris
Navi	/Cih	+60.72	345.82	g Cassiopeiae
Ruchbah	/Ksora	+60.24	338.55	d Cassiopeiae
Caph		+59.15	357.70	b Cassiopeiae
Schedar	/Schedir	+56.54	349.87	a Cassiopeiae
Merak		+56.38	194.53	b Ursa Majoris
Alioth		+55.96	166.50	e Ursa Majoris
Mizar	(+Alcor nearby)	+54.93	159.01	z Ursa Majoris
Phecda		+53.69	181.54	g Ursa Majoris
Eltanin	/Etamin	+51.49	090.86	g Draconis
Mirfak		+49.86	308.92	a Persei
Alkaid	/Benetnash	+49.31	153.12	eta Ursa Majoris
Capella		+46.00	280.83	a Aurigae
Deneb	(not Denebola)	+45.28	049.64	a Cygni
Menkalinan		+44.95	270.12	b Aurigae
Almak	/Alamac	+42.33	329.02	g Andromedae
Algol		+40.96	312.96	b Persei
Sadar	/Sadir /Sadr	+40.26	054.45	g Cygni
Vega		+38.78	080.76	a Lyrae
theta Auriga		+37.22	270.07	th Aurigae
Mirach		+35.62	342.57	b Andromedae
Aljanah		+33.97	048.45	e Cygni
Castor	N. twin	+31.89	246.34	a Geminorum
Alpheratz	/Sirrah	+29.09	357.90	a Andromedae
Elnath	/Nath	+28.61	278.43	b Tauri
Scheat	/Schert	+28.08	014.06	b Pegasi
Pollux	S. twin	+28.03	243.67	b Geminorum
Izar		+27.07	138.75	e Bootis
Alphecca	/Gemma	+26.71	126.33	a Coronae Borealis
Hamal		+23.46	328.20	a Arietis
Sheratan	/Sheratain	+20.81	331.33	b Arietis
Zozma	/Zosma	+20.52	191.47	d Leonis
Algeiba		+19.85	205.00	g1 Leonis
Arcturus		+19.18	146.08	a Bootis
Aldebaran	(not Alderamin)	+16.51	291.01	a Tauri
Alhena		+16.40	260.56	g Geminorum
Markab		+15.21	013.82	a Pegasi
Denebola		+14.57	182.73	b Leonis
Rasalhague		+12.56	096.27	a Ophiuci
Regulus		+11.97	207.90	a Leonis
Enif		+09.88	033.96	e Pegasi
Altair		+08.87	062.31	a Aquilae
Betelgeuse		+07.41	271.20	a Orionis
Bellatrix		+06.35	278.71	g Orionis
Procyon		+05.22	245.17	a Canis Minoris
Vernal Equinox, defined as fixed		00.00	000.00	'the first point of Aries'

Figures are from Epoch 2000.0.

Precession of the equinoxes:

Stars drift about 1° of position relative to the (0,0) reference star, in **72 years**, because it moves to the **WNW along the Ecliptic**. See p 33b
The third figure will therefore become off by one after 10-20 years. E.g. Mimosa used to be –59.59; Atria used to be 108.3W
One degree of latitude or declination = 60n.ml. =111km; so a maximum change in **0.1° over 18** years represents 11km, or about 0.6km per year, max.
(The actual change in Declination depends on its relationship to the Northern Spring Equinox point — see below.)
One degree of SHA = 60n.ml. at the Equator; half that at 60° Lat. ; **1° reduction in 80 years** is about 3 seconds later in timing per year, or 1.4 km max, at the Equator.

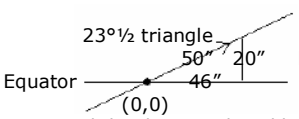

Ecliptic The first point of Aries moves along it at about 50 seconds of arc per year (5/6° in 60 Years)
Declination change of 1/3 minute of arc per year (20 secs of arc)
SHA change of 46 seconds of arc per year $20^2 + 46^2 = 50^2$

If you set your star watch by the actual visible stars each year, you eliminate E-W 'precession' component, and are left with only the drift *in declination*.
So it is of some worth to be able to predict how that rate of change in declination depends on SHA.
Obviously Stars to the North of Aries lose Northern dec, at the full rate. Stars to the South must *gain* Southern Dec.
Stars *opposite* Aries (SHA=180°) do the opposite. Stars exactly in between are *unaffected* (SHA=90° or 270°).
The rate for all cases follows **the cosine of the SHA**. E.g. stars at 60° SHA will have 50% of the full dec adjustment (of 1/3 n. ml. per year).
Take care with Polaris, since it is too close to the Pole.

Appendix: # Drawing Great Circles

First, bisect a globe.
Two equal halves.

Put the halves back together.
Notice the ellipse.
The backside is dashed.

Draw it symmetrically around the centre.
Use *flat* ends to the ellipse.

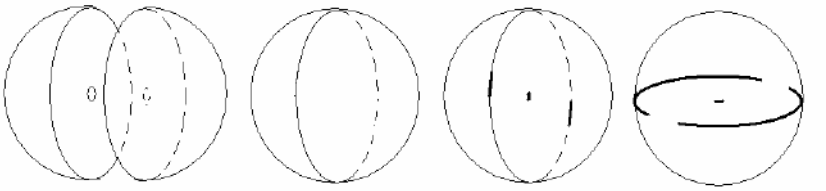

Choose the clearest perspective
E.g. 1. Rising/ setting is 90° away, so put yourself in the centre, surrounded by a 90° circle. Z is used for your Zenith — overhead.
E.g. 2. Look at the Equator side-on to make it and the latitude circles straight.
E.g. 3. Centre on the pole, to show up the polar distance. P is used for the Pole.
E.g. 4. Curve away from the centre. If in doubt, draw the whole ellipse and discard the unwanted parts.
5. A Great-circle can be drawn as a straight line. X is used to mark the position of the Star/ Sun/ Moon/ Planet
6. Draw the 'Vertex' of the course as a tangent to the maximum latitude small-circle.
7. The vertex can be drawn side-on.

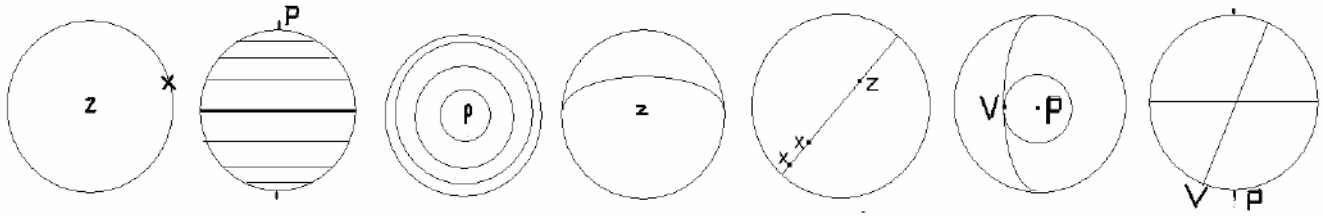

The PZX spherical triangle

8. Deliberately avoid straight lines to show up the curve of the triangles.
8a. ZX is the angular distance from you to the star, or the co-height.
8b. PX is the polar distance or co-declination to a point directly underneath the star.
8c. PZ is the polar distance or co-latitude of your position, 90° — lat.
8A. The angle ZPX, or 't', is a longitude difference — between the SHA of the star and your West longitude. A longitude difference is a time delay.
8B. The angle PZX is a horizontal direction relative to P, towards X. Technically called an 'azimuth'.
8C. The angle PXZ is similar but only meaningful if there were someone under the star, looking toward you.
9. A, B, C and a, b, c are used to label angles and sides of any spherical triangle; It's up to you to know which ones label what.
10. Draw a full 180° to show up the supplementary angles.
11. Draw the Equator cutting East and West of you, 90° away. Mark any right angles, and the complementary angles.

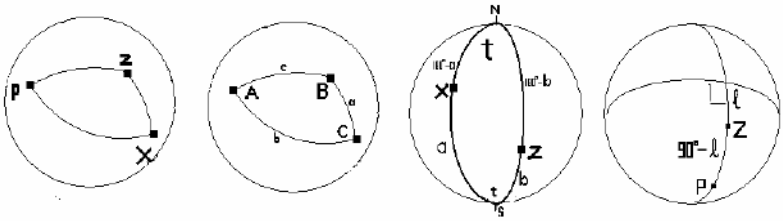

The PZX navigational triangle conforms to a handful of standard **fundamental spherical trigonometrical formulae**…

The Angle **Sum**: a+b and A+B are both greater than, or both smaller than, or both equal to, 180°
The **Sine** Rule: sina/sin A = sin b/sin B = sin c/sin C = constant. Ambiguity (about obtuse angles) can be resolved by the angle sum rule.
The **Polar** Triangle Rule: abcABC can all be swapped for the supplement of the opposite members.
E.g. 'a' becomes '180°—A'; 'B' becomes '180°—b', and the resulting formula is still valid. E.g. the sine rule simply inverts.
The **Cosine** Rule: cos a = cos b cos c + sin b sin c cos A
The Polar Cosine Rule: cos A = — cos B cos C + sin B sin C cos c
The **Four Parts** (in a row) Rule: e.g. AbCa or aBcA in the triangle aBcAbC or aCbAcB. You can find the 4th (outer) part from the first 3.
cos (**inner** side) **cos** (**inner** angle) = s**in** (**in**ner side) c**ot** (**ot**her **side**) — s**in** (**in**ner angle) c**ot** (**o**u**t**er **angle**)
If you want an inner part, use the sine rule and angle sum rule to find the 5th part, then go to the half-angle rules.
The **Five Parts** Rule: sin 1 cos 2 = sin 3 cos 5 + sin 5 cos 3 **cos** 4, where 4 is a side. It gives you any but part number 3, from the other 4.
The Polar Five Parts Rule: sin 1 cos 2 = **—** sin 3 cos 5 + sin 5 cos 3 cos 4, where 4 is an angle
The **Half-Side** Rule, to find the *middle* of 5 parts, when the middle part is a side: halve each side and angle first:
tan (half side) = tan (half sum) cos (half sum) / cos (half difference) ; tan refers to sides, sin and cos to angles
= tan (half difference) sin (half sum) / sin (half difference) but this introduces order-ambiguity because of the sines.
The **Half-angle** Rule to find the middle of 5 parts when it is an angle
cot (half angle) = tan (half sum) cos (half sum) / cos (half difference)
= tan (half difference) sin (half sum) / sin (half difference)

Make North and West positive, and East and South negative, and the formulae automatically work on PZX for angles up to 360°.

The formulae simplify where a side or an angle is 90°, since sin 90°=1 and cos 90°=0.
Napier's Rules…
The rules for **right-angled** triangles and **right-sided** triangles are simplified when the 5 other parts of the triangle laid out **in a pentagon**, and the 3 parts not adjacent to the right angle are changed to their **complement**.
In addition the part opposite a 90° *side* must have an extra **negative** sign.
E.g. if A is the right-*angle*, use (90°—a) as part 3, (90°—B) and (90°—C) for parts 2 and 4; if a is a 90° *side*, use A-90° for part 3.
Then… Sin of any one part = t**a**n t**a**n of **a**djacent parts or = c**o**s c**o**s of **o**pposite parts (in the pentagon of changed parts).
Ambiguity must still be resolved by special rules: Problems occur when there are 2 or more right-angles in the triangle.
'A' and 'a' (opposite parts in the triangle, with matching letter names) are both less than or both more than 90°. Likewise for Bb and Cc.
If one part and any part not opposite it (i.e. not A and a, as above) are on the *same* side of 90°, so are all the other mismatched pairs.
If one part and any part not opposite it (i.e. not A and a) are *not* on the same side of 90°, neither are any of the other mismatched pairs.
These refer to the triangle parts, not the altered pentagon parts.

Appendix: Calendars in the Sky — the first glimpse and last glimpse of stars

~ Firstly, What's wrong with **sunrise/sunset time**?

- To know what time of year it is, you could note the ordinary standard Sun time of sunrise/set at *that* place (e.g. from home).

—But this is insensitive (maximum shift is about 1 minute a day), useless around solstice, latitude & longitude dependent & requires a watch

~ You need to use apparent *star* time; no watches, to **locate the Sun's position on the Zodiac relative to the stars**.
— i.e. 'what stars are up' at midnight, or at twilight, or what stars are invisibly up at mid-day or sunrise or sunset.

- The Sun creeps Eastward about 1° per day along the Ecliptic (through the middle of the Zodiac belt).
- That quadruples the sensitivity to 4 minutes change a day, steady year-round.
  It is only affected by the Equation of Time, which is a minor variability, 1/4 min a day, max.
- The Sun will 'blot out' each Ecliptic star for about a month at a time.
  That's a rough way of finding the month: 'I can't find The Twins' (July).
- 'Nautical twilight' is when the Sun is 10° below the horizon (about 1 hour, in Tasmania), dark enough for the bright navigational stars to be visible while the horizon still shows clearly enough to be used with a sextant;

So bright stars 10-15° away from the Sun begin to be visible. That means that a bright Ecliptic star might only be invisible for 3 weeks.
[Civil twilight is 5° below, requiring lighting; and astronomical twilight is 15° below, dark enough to be called 'night', and to use a telescope]
E.g. 'What stars are up at midnight', uses the 'apparent star time' of the anti-Sun, as it were. You need to bisect sunrise and sunset, and be up then

- Find which star is N-S of you (the 'local apparent star time') at 'local apparent midnight' (= opposite noon — i.e. the anti-Sun stars).

But that is difficult — you need to know the apparent noon (which is longitude dependent), not the clock noon — and it still requires a watch.
So how can you do better?

- You could note **the clock-time of any particular star event**. E.g. follow 'Rigel over Capella' during Summer. E.g. Antares at 10° high.

This is *simple, and accurate,* but it would only work for that location, no other, since longitude alters the clock-time of appearances.

- You could note **the clock-time *delay*** between any particular star-event and either sunrise or sunset. **This is a good method**, since it pinpoints sunrise or sunset more accurately than saying 'At twilight, when the stars come out, or when they fade out'.

For best results, choose which star-event directions to use, as explained below, so that people at all latitudes see the same appearances.

- This pinpoints the star-time—it tells you where the stars are—in relation to the apparent Sun-time (by tying it to sunrise or set).
- For Rise/Set observations, you need a good horizontal horizon (or other fixed height-of-star) for star-rise/set to be latitude-independent, and location independent; but at home you simply need a repeatable point of view of a given horizon.

~ But to get back onto the main topic: The use of a watch can be eliminated by using twilight 'glimpses'...

- For Twilight observations, you need dependable eyesight; & a clear sky, with no Moon or light pollution, or haze, to be repeatable.

The idea of a 'glimpse' is that it is positively there or not. And it doesn't need a watch. There are 10 or so varieties...

- If you see a star *before dawn*, it's positively there, or if you can't see it *after dark*, it's not there.
  - Look to the East before dawn and you will see **a new star appearing** for the first time in the season — a first glimpse.
    You will then lose sight of it due to daylight. This is the standard star-calendar event.
  - Look to the West after sunset and a star will have **disappeared** for the rest of the season — a last sighting.
    You glimpse it briefly, before it moves into the sunset glow, and is lost to sight for weeks.
  - Examples: The last glimpse of Spica is just after the September Equinox; Regulus is first seen just ahead of the September Equinox;
    The Equinox is between the two, and both are on the Ecliptic, about 25° each side. That's 25 days for the Sun's movement.
    The first glimpse of Antares is around Christmas time, since it is about 20° ahead of the solstice;
    Aldebaran reappears around the winter solstice; The Twins disappear soon after it.
    The solstice is between these two, 25° away from each.

You need to check the exact dates of appearance and disappearance because they vary with the exact latitude—this is a semi-local calendar.

- If *a star-pair* has tilted *out* of alignment by the time the stars come out, you simply *don't* see it occur;

or if it doesn't align before dawn, you can't see it 'happen'.

- Look pre-dawn for the first date on which you can glimpse **a star pair reach its proper alignment**, before dawn takes over.
- Look post-sunset for the *last* date on which you can see a star pair attain its right alignment, after which it is lost for a month or so.

- Each day a Western star gets closer to the Western horizon from above. Each day a rising star gets earlier to the Eastern horizon, from below.
  - Look for the first glimpse of a star clearly **rising or setting** *pre-dawn*
  - Look for the last glimpse of a star clearly **rising or setting** *post-sunset*.

If you see it setting, you know it; if you see it rise, you have seen it happen.

- Wait to see whether a star reaches a known fixed **height** (e.g. not too far above the horizon), to make observation easier.
  - Looking at a reflection of a star in smooth water or oil, will *double* its height 'above horizontal', if you want a method to work anywhere.
- Wait until a star coincides with a known fixed **direction** (e.g. 'over the tower'). This makes it location-specific, but may be useful.

These all have the effect of describing the height or tilt or direction of the horizon 'at twilight'.

- A star **overhead** at twilight will show up as a first/last glimpse. You look for it, but it may not 'happen'. This is latitude-specific.

~ You could now combine the 'glimpse' methods and 'crisp timing' to avoid the indistinctness of twilight (but this is more latitude dependent).
Look 'exactly 1 hour from sunset/rise' to crisp up the timing of the 'events'. Any exact time-span at least that long, would suffice.

~ To clinch the star-time, in a latitude-independent way, **use Great-Circle type directions** ('at right-angles to sunset/sunrise' 'a star near the Sun at twilight' 'opposite it') to ensure that people over a wide range of latitudes more or less agree in their observations.
The standard direction is not sunrise/sunset itself but 'where the Sun is *at twilight*' which is 1 hour to the left (SH); *the brightest part* of the horizon.

- identify star-pairs ***at right-angles to the Sun's*** **direction at twilight**.
- or examine the rise/set of **stars *near the Sun***, especially those ahead or behind it, in its rising and setting path.
- or look for stars rising, setting or simply being visible, ***opposite*** **the Sun** at twilight.
- You can use any other 'star direction', but it will only apply to that latitude. The best great-circle directions are between NW/SE & NE/SW.

~ But the sensitivity can be further improved, by **choosing *which* twilight** gives the most daily change…

- The nightly change (averaging 4 mins) is not the same amount through the year.
  It can be reduced to 3 or increased to 5 in a place like Tasmania, as the seasonal day-length varies by several hours.
  The change in sunrise or sunset time by up to a minute a day is overlaid onto the steady star progression.

Look at the time of day when the most rapid change is taking place, since you want the most accurate discrimination:

After mid-summer, look for 'first glimpses' in the pre-dawn ;

- Nights lengthen after the summer solstice — more stars rise in the East than set in the West as the night-time gradually dominates

At dawn, the Sun lingers a little longer under the horizon allowing more stars to rise, in their creeping ahead of the Sun.
But at sunset, the Sun goes down earlier, like the stars do, and you don't notice much change as to which stars are up.

- So when the nights lengthen, look **before dawn** and try to see any first glimpse this season.
  Luckily the pre-dawn in Summer/Autumn is mild, but it can be quite early. Use the clock-time delay method 'from sunset', not 'to sunrise'.

After mid-winter, look for 'last glimpses' post-sunset ;

- Conversely, after the winter solstice, the increasing day length dominates over the night sky — more stars set and fewer stars rise.

You will lose sight of stars setting in the West, quite easily, as the Sun will linger before setting, because of the season,
but find it hard to see a new star rising in the East for the first time that year, because the Sun pops up early.

- So when the days lengthen, to pinpoint the time of year, look **after sunset** for any last glimpse.
  That's not too late at night in Winter/Spring, and it should not be too cold. Or use the clock-time 'until sunrise', to maximise the daily gap.

The overall idea works in both hemispheres, but the optimal choices of when to look are tied to the seasons, not the date.
Use either method, when near to the solstice, or near the Tropics, or at any time you want to.

- Around mid-winter and mid-summer, you can happily use any of the 10 glimpses, or any of the clock-time methods.
- Near the Tropics, also, there is little differential advantage, compared with temperate latitudes, since day-length is fairly constant

~ Keep a special perpetual calendar (a 365 day chart with no days of the week specified) to note down your observations; with room for related seasonal changes, like bird migrations, frosts, plantings.
All this has little to do with navigation, directly, but gives you practice in *following the seasonal movements* of Sun and stars.

About the Author.

I have a big bushy beard from too much bushwalking and ocean sailing, but now that it has grown too grey, it is time to hang up my boots and write it all down.
My dad was a ’plane navigator during World War 2 and as a child I studied his manuals. Then at University I spent more time in studying naval navigational manuals and in bushwalking off track, finding my way over the snow or canoeing across lakes, than in studying for my exams. I used the library to learn spherical trigonometry, then, as soon as I finished Uni, set out across the trackless ocean. I soon found myself inventing techniques to plug the holes in my normal abilities. I returned home years later, not knowing what I was meant to do in life, but as I lay there on my old bed, I knew that if I was to be a navigator I would like to be one like Captain Cook. He experimented with new techniques, and with old — he described native sea-navigation expertise. How *did* they do it? Even now, no-one seemed to know. I immediately had a vision of one star above another, which kept me busy for years on great-circle navigation methods.
I had honed my skills trying to teach boaties and would-be ocean sailors, but my biggest challenge was to teach my wife and children. We became a feral family for a while and walked every track we came across, here and overseas. Only then did I appreciate how much more there is to orienting yourself *to your surroundings*, than simply having a compass.
My next inspiration came while walking the forest on a foggy day. I simply asked how the trees themselves would hold the secret to which way is which. Years of observation and testing later again, and this is the result. I don’t want it to die with me.
I hope you too can now enjoy the magic of the science and love the art of exploring, even teaching yourself, before you too have to stop still, inside the bounds.
Learn a little before it is too late.

Outline

“Navigation Down-Under” is your complete manual of bush navigation and outdoor orientation for when you go off the well-worn track. A free pocket guide to take with you bush-walking is part of the pack.
Do you know how to predict and find direction from the Moon by memorising just two simple numbers? Can you interpret the Australian bush plants for direction? You will find some never-before published new techniques, as well as all you need to know about traditional methods like Lie-of-the-Land.
If you come from the Northern hemisphere, you’ll need the top-down global approach, and can try some special Southern Hemisphere methods of direction keeping.
Sailors will find here a valuable beginning into celestial navigation.
A lot of it is about how to navigate without a compass, how to stay oriented and how not to get lost. It is an essential course in basic bush awareness and orientation, for surety and safety, plus it also throws out the challenge that you can improve your already existing competence very quickly and easily.
For Scouts or emergency search-and-rescue workers, it is a thorough course, advancing beyond the map and compass. Or perhaps you are a tour-guide and need more than course and distance — You’ll get something more, something more like biogeography or ecology.
Each page is a carefully arranged lucky dip of interesting tid-bits.

Keywords
Navigation; Orientation; Celestial Navigation; Bushwalking; Map and Compass; Without a Compass

Section 1. BIG-PICTURE TOP-DOWN pp 3-33

General Introductory Points Big-Picture Rule One:
3 Locate a Distant Reference Direction First
3 Visualisation is your Key Skill
3 Big-Picture Orientation, a Mental Layout Acquired Beforehand
3 Compass Navigation Robs You of a Global Perspective
3 The Big Picture is Invisible — Your Job is to Visualise It!
Global Perspective; Earth
3 Compass Navigation Robs You of a Global Perspective
4 Global Re-orientation — Light-Half/Dark-Half
5 A mnemonic for global orientation: "True Polarity"
6 Mnemonic: "Time Tells Turning"
7 The sky turns 15° per hour —Use your thumb and forefinger
8 Global Orientation: Latitude and Polar Distance
9 The Angle of Rising and Setting—is Determined by Your Lat.
10 Global Orientation: Point to the South Pole…
..half its polar distance below horizontal
11 The North Pole is *Perpendicular to the South Pole*
12 Global Orientation: the South Pole Star, and Polaris
13 Pole Stars—"(Latitude)° above and below, South & North"
Point to 2 Points on the Equator — Halfway between the Poles
15 Perspective Distortion — The Sun Moves in a Straight Line!
16 Global Orientation: Half the Arc-Distance below Horizontal
17 Half the Arc-Distance below Horizontal—2 Explanations
18 Example: Point to Tahiti
19 Use Latitude for a Distance Scale (Not 'Long'-itude)
20 The Horizon Rotates to Your Right in the S. Hemisphere
21 Global Perspective — Rotation of the Horizon
22 How to Shift Your Horizon —Its Timing and Height
23 Latitude without Numbers
24 Mnemonic "The Sun Travels the Tropics"
25 Tropical Countries and Places Under the Sun
— Arranged in Time Zones Relative to Eastern Australian Time
— 10 hrs ahead of Greenwich
26 North/South and Noontime by Bisection (of Rise and Set)
27 Finding Magnetic Variation by Bisecting Rise and Set
Find E/W by Delayed Bisection of Sun & Full Moon At Rise/Set
28 Find North from the Combination of the Sun and the Moon
29 Eastern Australian Standard Time Zone
10 hours ahead of Greenwich
Showing Tropical countries and places underneath the Sun
32,33 Overview
128, 129 Tropical countries and places underneath the Sun
Sun Your Best Reference Direction:
4 Imagine the Sun at High Noon — Always
5 Mnemonic: **N**oon Sun Passes **N**orth of My **N**ose Down Under
6 Notice these, Yesterday — to Guess the Time Today
7 Mnemonic: The Southern Summer Sun Sets Somewhat South
8 Mnemonic: "Sun at High Noon — How High?"
8 On Overcast Days, Average the Brightest Sky over 1/4 hour
9 High Noon — How High?
10 Optical Illusion — Beware!
— Use Your Shadow Not the Sun, For Direction
11 The Dark-Blue Polarisation Band Perpendicular to the Sun
11 The Centre of the Rainbow is Down-Sun
12 Your Shadow Rotates with the same clock-sense as the Sun
Shadow-Rise and Shadow-Set is opposite direction to the Sun
12 Mnemonic: Shadow Rise is on the Right; Set is on the Left
12 Mnemonic: Your Shadow Passes South of Southerners
at Southern Noon
13 The Sun Passes Quickly Past North-South Nearer the Tropics
14 Shortest Shadow South—Opposite Highest Noon.
Always Imagine "Shadow-Noon" on your "Shadow South" Side
15 Locate Polaris (& Find N) by Imagining the Centre of Circling
16 " Noon " is Not at 12 O'clock
17 The "Equation of Time" Diagram
18 The Sun Oscillates Seasonally between +/- 23½° Latitude
19 Using the Declination of the Sun
20 Tell the Time by the Compass Protractor—Tilt to the Equator
21 Or Judge the Time By Sight—by Mirroring the Sun Around N
(To Make an Hour Hand on a 12 Hr Clockface)
22, 23 Walking in Phase with the Sun, Wind and Moon
24 Tell the Time of Night by the Star Opposite the Sun
25a Eastern Australian Standard Time Zone 10 hours ahead
25 Tell North from The Time By Using a Protractor
26 Longitude and Latitude from the Sun
27 A Bush Theodolite for Overnight Equal Height Observations
28 Find North by your (non-digital) Watch-Hands
29 Find Any Other Direction Too — Using the Hour Hand
29c String North
30 Locating Anti-sunrise Shortest Shadow Souths
30 Sun Path Norths — Summary
31 The 2 Flaps Method for N.—Lat. & Dec. Angles Separately
31 Seasonal Adjustments For Equinoctial East
32, 33 Overview
132, 133 Eastern Australian Standard Time Zone
10 hours ahead of Greenwich
Showing Tropical countries and places underneath the Sun
Stars
4 The 4 Bright Stars of Orion Point out Where the Pot is Going
4d The Ecliptic
5 The Two Brightest Stars Point out the South Pole Star
6 The Spinning Earth Makes Circular Paths for the Sun, Moon
7 Stars Leave Star-Trails
8 RISE and SET Directions for a (Local) STAR COMPASS
9 Global Star-Compasses e.g. 'Perpendicular to' rise and set
10 Right Around the Horizon, Stars Move to Your Left, Down S.
11 Identify the Stars That Pass Directly Overhead
12 Ways to Locate the South Pole Star Or the N./S. Meridian
12a&13a Global Orientation: the South Pole Star, and Polaris
13 Pole Stars—"(Latitude)° above and below, South and North"
13 South Pole Star Chart
14 Find the Polar Axis by Star-Trail Bisection
15 The Eagle Clips the Equator, South of Altair
16 Equatorial Stars Rise and Set Due East and West
17 Rigel-to-Capella is a good S./N. Pair
18 Use the Southern Cross as an Hour Hand
18 Picture the Southern Cross during the Day
19 Perspective Distortion — Use the Correct Vertical
20 The Southwest Corner of the Great Square is N. S. E. W.
21 The Southern Summer Rectangle And a Triangle Pointing S.
22 Navigating by the Shape of the Night Sky
23a Latitude without Numbers
23 Matching Sky to Sky — Move Only Your Horizon
24b Stars Opposite the Sun
24 2 Different Back-and-Forth-Star-Directions for a Position Fix
25 For Land-Navigation, the Plumb-Line Star-Pair Method
26 Equal-Height Stars When the Sea Horizon is Visible
27 Great-Circle Reference Star Directions
28 Leo the Lion Looks like a Lion or Sphinx Squatting
It Has Four Lines which Show N./S. E./W.
29 String North Latitude without Numbers
30 "Star-Time" Gains Four Minutes a Day
31 Working the Star-Longitude Numbers
32, 33 Overview
118,119 Appendices: Divide the Star Globe into 8 Octants
—and Learn the Stars in Each
The Southern Figure of Eight
The Zodiac
122 A List of Star Coordinates
124 Drawing Great Circles
125 Calendars in the Sky
Moon
4 Look for the Line Joining Sun, Moon, Planets & Zodiac Stars
— the Ecliptic
6 The Moon; Rule 4: Look for It
8 From the Moon, Locate the Sun-Moon Direct Line, Carefully.
10 *Mentally Rotate the Sun-Moon Line Around Polaris*
12 "Cusp North" — Approximate Direction from the Moon
14 The Rabbit Looks up Ahead at Full Moon Using the Ecliptic
16 The Moon Cusps Push and Pull the Hour Hand Backwards,
Down South
17 Watch the Clouds Cross the Moon
18 The Moon is Lit By the Sun
20 The Moon Rises and Sets 50 Minutes Later Each Night
22 The Moon is "in Phase" with the Sun — Watch for It
24a The Moon Moves Over the Tropics
24 Predict the Full Moon Path According to the Time of the Year
24d Finding North from the Moon, by its height & phase
25 Predict the Moon Phases by " N minus M "
28a Find E/W By Bisection of Sun & Full Moon At Rise or Set
28 To Predict the Moon's Path, Visualise the Ecliptic
29a Find North from the Combination of the Sun & the Moon
29 A 4 Week *Position Cycle* of Moonrise/set, to Find E/West
29e Keeping Track of the Moon's Motion along the Ecliptic
30 Adjusted Moon-Cusp North:
To the Left or Right, By Moving the Sun
32 Overview: The Moon
116,117 Appendix: Moon-Phase & Tide-Time Dial
Wind See also pp. 60a-68a
5 Mnemonic: "Wind? Weather Map!"
7 Mnemonic: "Wind in Your Face"
9 'Low' Clouds Come Clockwise Down under
11 Weather Fronts Distort The (Circular) Wind Circulation
13 Wind-Direction-Change Rule for Buys Ballot's Law
13 Steering Winds
15 Atmospheric Pressure Changes.
17 Watch the Clouds Cross the Moon
19 To Determine Upper Wind Direction from the Clouds
21 Clouds and Con-trails and Their Winds
23 Prevailing Winds on Creek Banks Tips for calm weather
27 Make a Bush Psychrometer
31 Topographic-Level Winds
33 Overview: Global Weather Patterns

52e The Aurorae
53e City Glow

Clockwise See also Map & Compass Work
4 Mnemonic: "Clock"wise Northern Hemisphere SWiNES
6 "VEERING" with the Sun is "Clockwise"
only in the Northern Hemisphere
"BACKING", against the Sun's movement, is "Anticlockwise"
— but only in the Northern Hemisphere.
8 Mnemonic: (when you say) "COMPASS" (say) "CANADA"
10 "Bears Circle Clockwise" Compass Bearings are Clockwise,
A Full Circle, and Eurocentric, Cee?
12 Mariner's "Compass Points" work on Successive Halvings
14 "E. W. N. South" Of What!? Rule: Say What
16 Back-Bearings Killed Sleepy Pilots
18 Back-Bearing: Plus or Minus 200° Then Minus or Plus 20°
20 Turn 90°? "+/-100° -/+10°" !
22 The Clock-Face Method for Direction
24 One Degree is about One in Sixty
26 Anticlockwise is Your Best Spin-Reference Direction
28 Bearings & Tracks: "If it is right, it is big in my sight"
30 "By the Nose" Navigation
32 Overview
Anticlockwise
5 Mnemonic: Down Under, The Sun Moves ANTI-BigBENWiSE
i.e. anticlockwise
7 Southerners Need Anticlockwise Systems
9 A Compass-Needle Continent—Swinging the Continent
11 Interpret a Compass Needle as "RED HOT—WHITE FROST"
13 Four Reference Directions—Name Them "Rise, Noon, Set, Midnight"
15 "Rise-Noon-Set-Midnight" carries its meaning across the Equator!
N.S.E.W. doesn't.
17 The 0, 1, 2, 3, 4 Reference Directions
19 The Sunny Skies Quadrant System
21 The "SWEETHEART" Method of Orientation
23 Mnemonic: "North; Coast"
23 Go Overseas!
27 Walk across Australia! English!
29 South is a Triply Better Reference Direction than North, in the Sth
31 Nose-Navigation Needs an 'Anti' Mnemonic: "Nor-West, Not-Least"
32 Overview
Revision, Overview and Explanations pp. 32,33
Top Down Orientation
Time is counted in cycles
Closer to Home
The Precession of the Equinoxes
The Moon
Global Weather Patterns
The Moving Horizon
The Spherical Earth
The Spinning Earth
Moving the Horizon

Section 2. THE INSIDE STORY—Mental Manoeuvres pp.34-55

Learn
34 Navigation is More Than Navigating Common Sense
36 The "Good" Navigator. You.
38 Don't Just 'Get By' and 'Make Do'
40 Walk Where the Navigators Talk — if you want to learn navigation
40 Explore It's No Fantasy
42 Self-Motivation is Fundamental
44 Routine Mind-Shut-Down Procedures
46 Navigation is Mental Work — Stay Alert
48 To Work Harder, Rest Harder
50 A Three-Track Mind
52 Concentrate Outwards
54 Come to your Senses
Think; Interpret
34 Your Subconscious is Superior to Science!
36 Think out Loud
38 Memories are Made of Landscapes, Not Compass Needles!
40 Feed Your Subconscious through Your Conscious Focus
42 How to Feed Your Brain
44 Feed Your Subconscious Subconsciously
46 Instruments Instruct the Intelligent
48 Minds Work Best with Intelligent Questions on Them
50 Interpret!
52 SPELL Out Numbers — Into Words
52 PAINT your Words — Into Pictures
54 TRANSFER Your Pictures into Reality
54 Put That Experience in Your Memory Bank
55 Extend Visible Reality into the Invisible Picture
Angle:
34 Measuring angles at 1 cm per degree
35 Point and Blink = 10% Shift
36 Guesstimating Angles at 1 degree per cm
37 Distance using Apparent Size;
37 Visual Angle, Size and Distance & 6°; Side-wise Offset
(38, 39 Grid-line Walk & Mental Grid Maps)
40 Cyclical Motion Divided into 12ths
41 Approximate Sine-Wave Values
Look
34 Visual Assessment is Your Main Clue — Rely on It More
36 You Need Excuses to Stop and Have a Good Look
38 You Need a Road-Map To Climb a Hill!
40 Look into the Far Distance Look through the Haze Wait
40 Wait for a Change in Perspective...
42"Retracing is Re-preciating"
44 Look Back Regularly. Look Around
46 Look Through the Vegetation, to the Contours
48 Play the Field-Scientist
50 Look at the Ground
52 Look for Foreshortening & Overlap & Parallax
54 Look, with Vengeance Aforethought
Integrate
3 Visualisation is your Key Skill
34 Integrate: Jetstream, Earth Rotation & Heaven Rotation
36 Visualise Familiar Landscape Trends, Under the Sun
38 'Land to Sky — Sky to Land' Navigation
40 Join the Invisible Big Picture To the Visible Landscape
42 Landmark to Landmark Network Navigation
44 Aim to Integrate…
46 Imagine an Aerial Photograph of the Continent
48 Connections to the Outside World
—You don't need to feel isolated
50 Orientation is More Fundamental Than Position Sense
52 Make Common Sense of the Various Direction Clues
52 Don't Degrade the Point of a Clue or System
54 Preparation is the Key to Difficult Navigation
42e Moveable Position vs Fixed Orientation
43e Name Each Clue — with solid ground
44e Notice and Match Up
45e In-Line Integration
50e Think in Pairs, in lines, trends and tracks
Orientation 35 Orientation = Direction Sense
37 The Get-Back Path is the Single Fundamental
38e Grid-Line Walk
39 Disorientation is More Dangerous Than Being Lost
39e Mental grid Map
41 Which Way Did I Come In? Which Way Do I Go on?
43 Disorientation — A Killer Syndrome
42e Movable Position vs Fixed Orientation
45 Un-Dis-Oriented;
47 Avoid Disorientation: OverNavigate
46e Making a M.U.D. M.A.P.-m.a.p.-m.a.p.
49 OverNavigate: Equals Over-Familiarise
51 Needle N, S, E & W. Identify them on the Landscape
51 Orientation to Start-Finish Rhumb Line—by sight or compass
51 Orientation to Your Finish Point, By the Parallax Method
53 Orientation to Your Start Point, By the Parallax Method
55 Look Backwards, Think Forwards
Disorientation
35 The Navigator's Top 10 Problems; and Band-aid Solutions
37 How I Manage to Get Lost or Disoriented
39 Logical Blunders — Examples
41 Logical Presumption Leads to Utter Delusion
43 Casual Outings Catch You Out
45 Trapped! By Not Looking Back
47 How to Get Lost—Follow the Track—A Track is a Trap-Door
47e Disorientation in new territory
49 How to Get Lost—Follow the Leader
51 Exponential Anxiety — the 11:59 syndrome
53 Nervous Nellie Navigation — Not
Reorientation
35 Reorientation
37 Voice Your Thoughts — Don't Suppress Important Clues
39 Trust the Untrustworthy
41 Trust Less and Prove More
43 Trust Yourself Less
45 Disoriented? Stop. Say So. Reorient!
47 Reorientation; And a Fresh Start
49 "As Solid As the Rocks in My Head"
51 Explore Junctions, While You Can
53 Look for the Essentials — the 'Directions'
55 To Get There Quicker, Go Slower
Lost?
35 Lost? Rule One: Stop! ..getting More Lost
37 Back-track in Your Mind before You Back-track on Your Steps
39 Rule 2: Start! Getting Unlost. Find yourself. Regain confidence
38e Grid-line Walk — Explore in a Square Grid Pattern
39e Mental Grid Map
41 The Simplest Way to Get Unlost is to Retrace Your Steps
43 More is Better — Thought, Time Care
45 Lost? Rule 3: Do a Good Job—of walking out or staying put
47 How to Leave a Message
49 Lost? Rule 4: DON'T Do It Again!
51 Separated? An Agreed Plan Helps
53 A Whistle-Code for Contact Calls
55 Search and Rescue Signals — Noises Repeated Regularly
120 Appendix: Morse Code and Rhythm Mnemonics for it
121 Appendix: Semaphore Code and Mnemonics for it

Section 3. REALLY INTERESTING STUFF
—Bush Navigation pp56-89 See Also 'Wind'

Nature's Norths
56 Nature's Norths
58 Rock Noon
60 Wind West See also Wind, Section 1
62 Think Wind
64 Be Wind-Wise: Listen
64 Obvious Tell-Tales: Windblown Objects
64 Obvious Tell-Tales: Landscape
66 Light & Colour Effects
68 Botanical Tell-Tales
70 Combined Effects of Wind, Salt and Light
70 Botanical Asymmetry
72 Think like a Plant
72 "Mosses Mop up Moisture" "Lichens Like Light"
72 "Fern Fronds Face the Front" i.e. the sunny side
74 Bush Noon— The Sun is Always out, If You Know How to Look
74 False Asymmetry & Symmetry True Symmetry & Asymmetry
76 Higher Density of Leafiness on the Noon Side
76 The Lowest Leafy Branches are on the Noon-Side
76 "It's the Shady Side That Shrubbery Shows up"
76 Tall, Steep and Ragged Backside Wedge-Shaped
76 Hollow under the Shady Side & Imperfect Shape
76 Gaps on the Shady Half Dead Branches
78 'Solar-Panel' Tilting of the Canopy Top
78 Wedge Shaped Lopsided Lean Half-Dome One-Sided Leafiness
78 Individual Branchlets Bare Backbone
80 Branching Patterns
—Examine Them for thickness, length, angle, curvature
82 Isolated Vegetation Shows the Clearest Asymmetry
82 Closed Canopy — Look for Gaps; Saplings
82 Roadsides & Campsites Reverse Some Effects!
84 Flowers Respond to Sunlight Fruit Distribution and Ripening
86 Stem and Leaf Colouring New Growth is Colour Differentiated
88 Noon-Sunlight is More Vertical
88 The Shedding of Bark on the Sun-Baked Side
88 Where to Look for Fading or Lack of it

Memory & Naming Skills See Sect.5, Navigation Mnemonic
56 Don't Forget to Remember, Your Way!
56 Unlock Memorable Chapters in Your Life
56 Go for the Jugular. Zero in on What Matters
58 Don't Try to Memorise Details without an Overall Framework
58 Don't Try to Remember Anything You Haven't Memorised
58 Write down What You Memorise!
60 Anything is Memorable
60 Make Individual Memories Memorable!
60 Remembering Trees
62 The Difference between Looking at, and Seeing
62 Search-Image
62 We Recognise 'Sameness' But Need to Describe the Differences
64 Place Recognition
64 You Recognise What You See, By What You Can't See
64 Description Requires Words
64 Can You Describe All That You Can Recognise?
66 Discuss It
66 Collect the Details
66 Identify a Specific Place by its Specific Collection of Details
66 Include What You Can't See
68 Extend Your Familiar Range Systematically Hints
68 Leap-Frog Walk in New Territory
70 Draw the Skyline, the Ridge-lines, the Outlines
70 Label it with the 8 main directions
72 Examine the Skyline BEFORE You Walk up to It
72 Make a Mental Movie
72 Take Mental Photographs
72 Take a Mental Photo of Each Change of Track
74 Remembering the Sequence, Needs More Than Logic
74 Do Something NUMERICAL
74 Match the SHAPE
76 Use a Type of Memory which you are Good at
76 Use the Name of the Track for the Acrostic
78 Group Several Related Images into One See also p. 110
80 Working Names — Invent Them
80 Standard Name Plus Personal Name
80 Remembering Names
82 Name the Subsections
82 Name the Sequence
84 Revise your Memories To Remember Formulae
86 Remember the Sequence — from AA to ZZ

Lie of the Land
57 The "Lie" of The Land
57 Listen to the Pioneers Lie Has To Do with Water Drainage
59 Major Rivers & Ridges vs Minor Ones Sort it out on the Map
59 Catchment Divisions Snake from Coast to Coast
59 A Catchment-Division Separates Whole Valleys
61 Ridges Ridge Names Mnemonic: "Ridges Rise"
61 Uphill — Which Hill? Upstream — to which Divide?
63 Shoulders Stop. Shoulders are Not Easy (like Ridges) to Walk on
63 Saddle — A Flat between Hills, Sitting Astride a Ridge
65 Notice When You Cross a Catchment!
65 The Giveaway! A Change of View—Mnemonic: "Catch the View"
65 You Can't Cross a Creek Twice from the Same Side
67 Break of Slope/Make of Slope Scree Slopes — Tread Carefully!
69 Break of Slope — And Route-Finding
71 Mnemonic: "Down in the Mouth" Which way is the flow?
71Hint: Don't think "this side" & "the other side".
73 The Inside of a River Bend is Flatter Ground
73 Rivers Bend! Don't Forget to Look Backwards
73 Rivers Meander. Ridges Run Straighter
75 Land-Form. Hint. What Makes a Cliff and an Overhang?
77 Headlands, Bays, Prevailing Winds, Longshore SandDrift, Creeks
77 Beaches — Slope, Particle Size, and Wave Energy
79 Rock Typing Soil Typing

Contours and Slopes
81 Contour Lines
81 Contour Navigation: Practise It, Before you Need It
83 Contour Curves: Your Job is to Imagine the Shapes into Life
And to Match the Lines to the Landscape
85 Contour Heights: Read the Numbers!
85 Contour Spacing: Close = Steep Slope
85 Contour Slopes: Work Out 1 Mm and 1 Cm Contour Spacing
85 Reference Slopes — To Imagine the Climb with!
85 The Beaufort Wind Scale — Modified for Bushwalkers
87 Make a Clinometer — a Slope Measurer
89 Just Draw It

Tracks and Bends
57 Stick to the Track Reasons for Leaving the Beaten Track
57 Cross-Country Routes
59 Tracks are like Naughty Children
59 The same is true for creeks & ridges
59 Explore Branches While You Can
59 Convergent Tracks Lead to Civilisation
59 Down-Track vs Up-Track
61 Sympathy and Understanding with Track Markers
61 Old Track Markers
63 Track Markers When You Lose the Track-Markers
63 Notice (with Alarm!) When...
65 Mark Your Trail Mark Your Trail Non-Permanently
65 Tracking Someone Else
67 Mark Your Trail More Permanently and Ostentatiously
67 Mark Your Trail Invisibly
67 A pebble Code — Keep It Simple

Bends
69 Keeping Track of the Bends Desired Equipment
69 Logging Constant Diversions
71 The Difference between 'Direction' and 'Track'
71 Continuous bends
71 Follow the bends by anticipating the trends
73 Keeping a Compass Course, by Legs
73 Turn Smooth Bends into Sharp Jumps
73 Mind Which Way You Go
75 "End the Bend" i.e. Notice the Un-Bends
75 Name the Bend, By the Trend, of its End
75 Notice and Name Perverse Bend
77 Look Through the Trees, to the Sky and Skylines
77 Periscope Navigating by Trees
79 Bush networks Don't Navigate by Networks Only
79 Antidotes For Guessing at Network Junctions
81 Sticking to the Bush Network Natural Routes Naturally Converge
81 Coping with Divergence Reverse logic
83 Trackless Navigation The method of Names
83 Bush Networking Essentials
85 Walking "Straight" through the Bush Straight around
85 How to Walk In an Unintentional Circle
87 Walking an Exact Straight Line in Open Country
87 Leap-frog a Straight Line by Looking Ahead
89 Radio Direction Finding—As Good as a Compass

Section 4a. DISTANCE pp. 90-108 left-hand pages
See also 'Angle'
Progress
90 Distance Guesstimation is a Major Problem for Walkers
92 Use the Map Scale—Don't Just Look at It The Jargon
92 Map-Scaling-Factors are Meant to be Interpreted
94 Psychological 'Half-Way Point' When You Meet Other Walkers
96 AIMING OFF by 10° — 10° is **175m per kilometre**
96 Aiming-off at Sea; Aim Off 10% or 6°
98 Up-down Progress
100 Distance by Double Pace—Test Yours Tie Your Ankles Together
100 Don't Lose Count! Use Finger-Math
100 Test your pedometer
102 Line of Sight Can Be Extremely Accurate
102 Two Opposite Landmarks
104 "Changeover Lines" of Progress — Relative Apparent Sizes
104 Use the Landscape Itself to Measure & Mark Progress
106 The Sine-Cosine Tally Method of Keeping Track, Blind
106 Rise and Run Adjustments — Sine and Cos — North and East
108 Progressing A Position Line across the Map
108 Pinpoint "Fixes" are a Myth!
Pace
90 Distance = Speed x Time Keep a Log! Adding two Speeds
92 Pace Your Progress — Pre-Dict It, Don't Post-Dict It!
92 Advice about Walking Times — Double & Halve!
92 20 Minutes per Kilometre — Double & Halve — Easygoing Maths
92 4 km per hour — Normal Walking Pace — Double and Halve!
94 1 Metre per Second is 3.6 kph "1 Hour Extra per 5 Hours"
94 "1 Hour Extra per 500 m easy climbing"
96 Adjust Your Expected Pace
96 Don't Overcompensate for Slopes And Meanders
98 Don't Lose Count! Use Finger-Math
100 A Slow Kilometre Cancels out Three Quick Ones
100 To Progress Faster More Easily Pace Yourself
— Slow Down, but Don't Stop
102 Distance-Off by Vertical Angle Safe Distance Off
104 Circle-of-Position by Horizontal Angle
105b A Horizontal Angle between Features Yields a Position Circle
105c 'Two Horizontal Angles' Usually Fixes Your Position
106 Rise and Run Adjustments — Sine and Cos — North and East
Range
24e One Degree is about One in Sixty
34-36e Estimating angles
37e Distance using Apparent Size & 6°
37e Visual Angle, Size and Distance Side-wise Offset
51a & 53a Parallax methods
90 Range: Rule One: Guess
92 "Range" of Visibility — Two Halves to the Picture
92 "Range Squared is Dee Height"
92 "Range Squared is Dee Height"
94 Double, the Acute Angle, off Your Nose — For the Distance Off
94 Halve, Any Angle, from your Stern—For Distance Off
96 Distance by Optical Rangefinder
96 Range by Parallax and a Piece of String DIY Rangefinding
98 Distance by averaging maximum & minimum Guesses
98 Distance by Comparison with Some Length You Know
98 Distance by Cloud Shadows — Look for Them
98 Sound Travels Three Seconds to the Kilometre
100 Lighting Conditions & Eyesight
100 Visual Acuity is about One Minute of Arc
102 Approximate Figures for Visual Acuity
102 Test Your Own Visual Acuity
104 Circle-of-Position by Horizontal Angle
105b A Horizontal Angle between Features Yields a Position Circle
105c 'Two Horizontal Angles' Usually Fixes Your Position
104 The best range-finder
106 Rule of Fingernails Rule of Tom Thumb Apparent Size

Section 4b. MAP and/or COMPASS pp. 91-109 right-pages
Map Use
54e Helpful Chart Symbols 122 Appendix: Symbols
38,39e Grid-line Walk & Mental Grid Maps
46e Making a M.U.D. M.A.P.-m.a.p.-m.a.p.
91 To "Read" a Map — Rule One: Don't Lose Your Place!
91 Rule Two: Imagine Arrows for the Sun, & Your Movement
91 Rule Three: Get Used To Any New Map Scale
91 Rule Four: Don't Give Up!
93 Look Carefully at the Fine Detail on the Map
93 Look Carefully at the Fine Print
93 Some Map Details are Not to Scale!
95 Permanent Landscape is More Trustworthy Than Surface Detail!
95 Details Added on to a Map Are Prone to Error
95 Expect The Map To Be Unhelpful
95 Maps Do Give You 10 Times More Detail Than You Can Notice
97 "Maps Grip the Universe"
99 Align the Track
101 Align the Map the Landmarks
103 Align the Map to Identify Landmarks
105 Align the Map to Fix Your Position
107 Which Way is North?
Magnetic Compass Use See also "Clockwise", Section 1
37e Chart symbols
91 "L.M.N.Or P."
93 Using the Protractor by Itself
95 Using the Needle Alone — Emergency Magnetic Compass
95 Hint Understand that magnetic poles should be called
'North-Seeking' and 'South-Seeking'
97 Think "+/- 5°" For Compass Bearings "+/- 10°"
For Following a Compass; "+/- 15°" For Using a Compass!
99 Compass Errors — Be Impressed! Be Careful!
101 Compass Use — Master It, but Be a Good Master.
103 "Magnets Grip Maps"
103 Align Your Map in the Dark
107 'East-least; West-Best' Rule—Tells You Which Bearing is 'Best'
108b Getting Started in the Dark
108c Thinking Backwards
109a Taking a Bearing from the Landscape
109 Plotting a ('Compass-to-Map') Back-Bearing
109 No-Number Back-Bearings
124 Appendix: Map Symbols
Mapmaking
38e Grid-line walk
39e Mental Grid Map
46e Making a M.U.D.M.A.P.-m.a.p.-m.a.p.
91 Maps Mean: Imagine Me! — Making Mental Maps
91 "Map To Mind" Navigation
93 Mental Networks
93 The Most Valued Map
95 Strip Map Navigation — Follow-Your-Nose Navigation
97 Caving Strip-Maps — Come in Books
99 Map Folding and Map Holding
101 Layout Map — Blind
103 A Compass Holder
106a The Sine-Cosine Tally Method of Keeping Track, Blind
106b Rise and Run Adjustments—Sine and Cos—North and East
107 String Map
109 Parallax While You Walk — JUST DRAW IT!

Section 5. Navigational Mnemonic pp. 110-115
78b Group Several Related Images into One
110 A "Start-Stop-Rest" Navigation Mnemonic
110 P.R.E.P.A.R.E.D.— A Before-You-Go Checklist
111 Equipment—A What-You-Might-Like-to-Take Checklist
111 Place-Names; Learn Them; You Navigate by Local Features
111 S.T.A.R.T. off Oriented
111T.I.M.E. & Distance — that's the point
111 Distance
111 Progress; Pace; Range
112 Look in These 6 Directions — in 6 ways — Up and Down
112 4 More Slopes
112 Look in These 6 Directions — in 6 ways — Ahead and Behind
112 Look in These 6 Directions — in 6 ways — Right and Left
113 ..Plus in Two More Directions
113 (1) All, Around, Appreciatively, And At 6 L's
113 (2) Look at the Map: a Systematic Checklist. What 'System'?
113 Mark the Lines-of-the-Land On Your Map
113 Mnemonics for Map and Compass Work
114 N.S.E.& W.—the 4 cardinal directions, is a mnemonic of fours
114 W.E.A.T.H.E.R. — Check for Change
114 Stop-Signal: Why have you stopped?
114 REST and R.E.O.R.I.E.N.T yourself
115 RI For Revisualise the Invisible & then Relate It to the skyline
115 Clock-Sense Mnemonic
115 ENT for Environmental Notice Time—Orient yourself to Nature
115 REMEMBER to Note down the Names in the Notebook
115 Re-lax, and Play — Navigation Games and Activities

Miscellaneous pp. 34e - 55e
52e The Aurorae
53e City Glow
54e Helpful Symbols for Standardising your Chart–Work

Appendices pp. 116-121
116 Moon-Phase & Tide-Time Dial
118 Divide the Star Globe into 8 Octants
—and Learn the Stars in Each
118 The Southern Figure of Eight
119 Divide the Star Globe into 8 Octants
—and Learn the Stars in Each
119 The Zodiac
120 Morse Code & Rhythm Mnemonics
121 Semaphore Signalling Code & Mnemonics for it
122 Symbols for Track-Notes and Mud-maps
122 A List of Star Coordinates
128, 129 Tropical countries and places underneath the Sun
124 Drawing Great-Circles. The PZX navigational triangle
125 Calendars in the Sky
126,127 About the Author; Outline
Index pp. 128-129
Contents p. 2

Eastern Australian Standard Time Zone

10 hours ahead of Greenwich
Showing Tropical countries and places underneath the Sun

p24a,25a

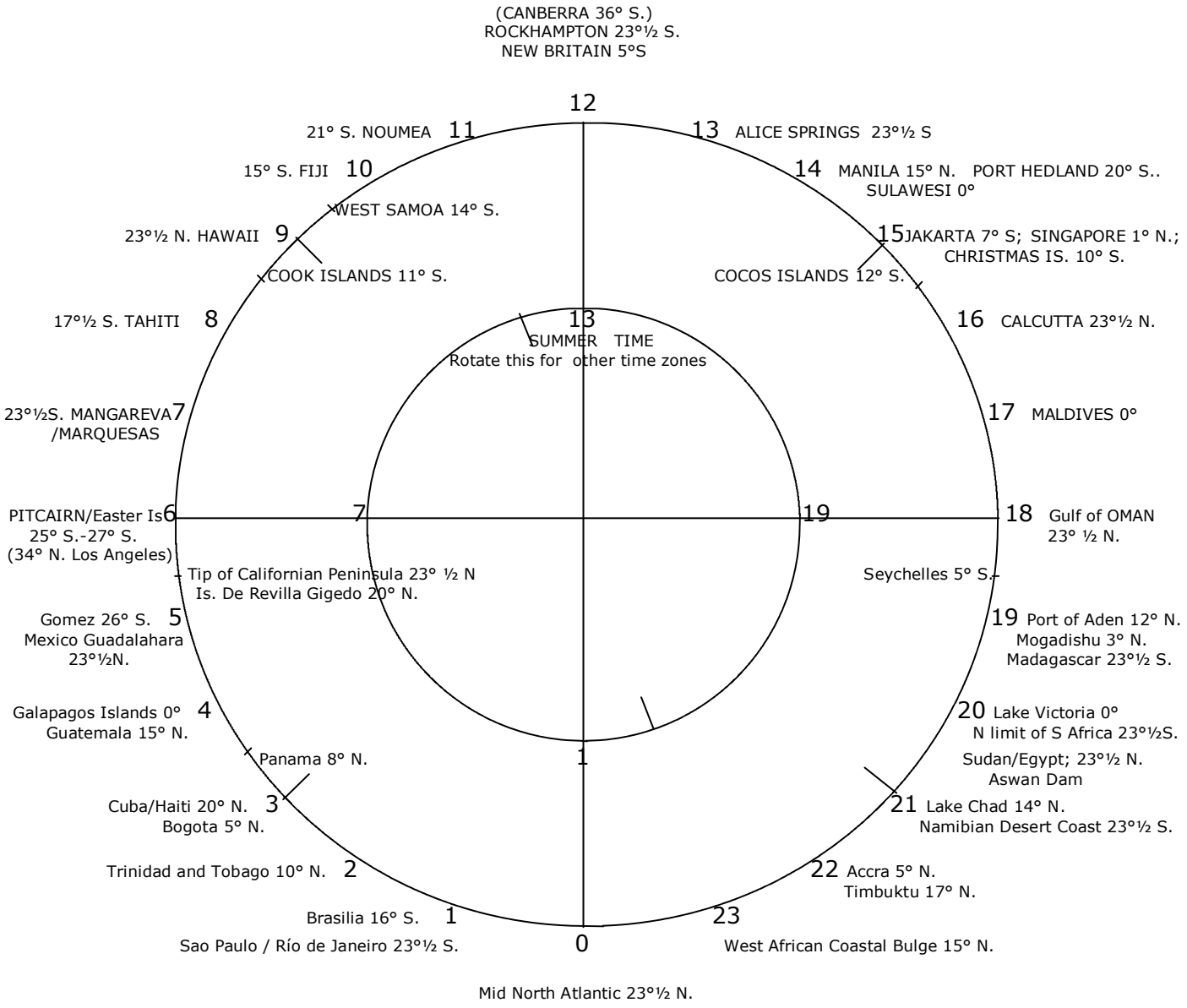

Places in capitals are less than 6 hours away from Australia. The others are on the other side of the world.
View this as if from above the North Pole—the Sun 'moves' clockwise, but the Earth spins anticlockwise.

Eastern Australian Standard Time Zone

10 hours ahead of Greenwich
Showing Tropical countries and places underneath the Sun

Cut this one out and post it on your wall until you have learned most of it.

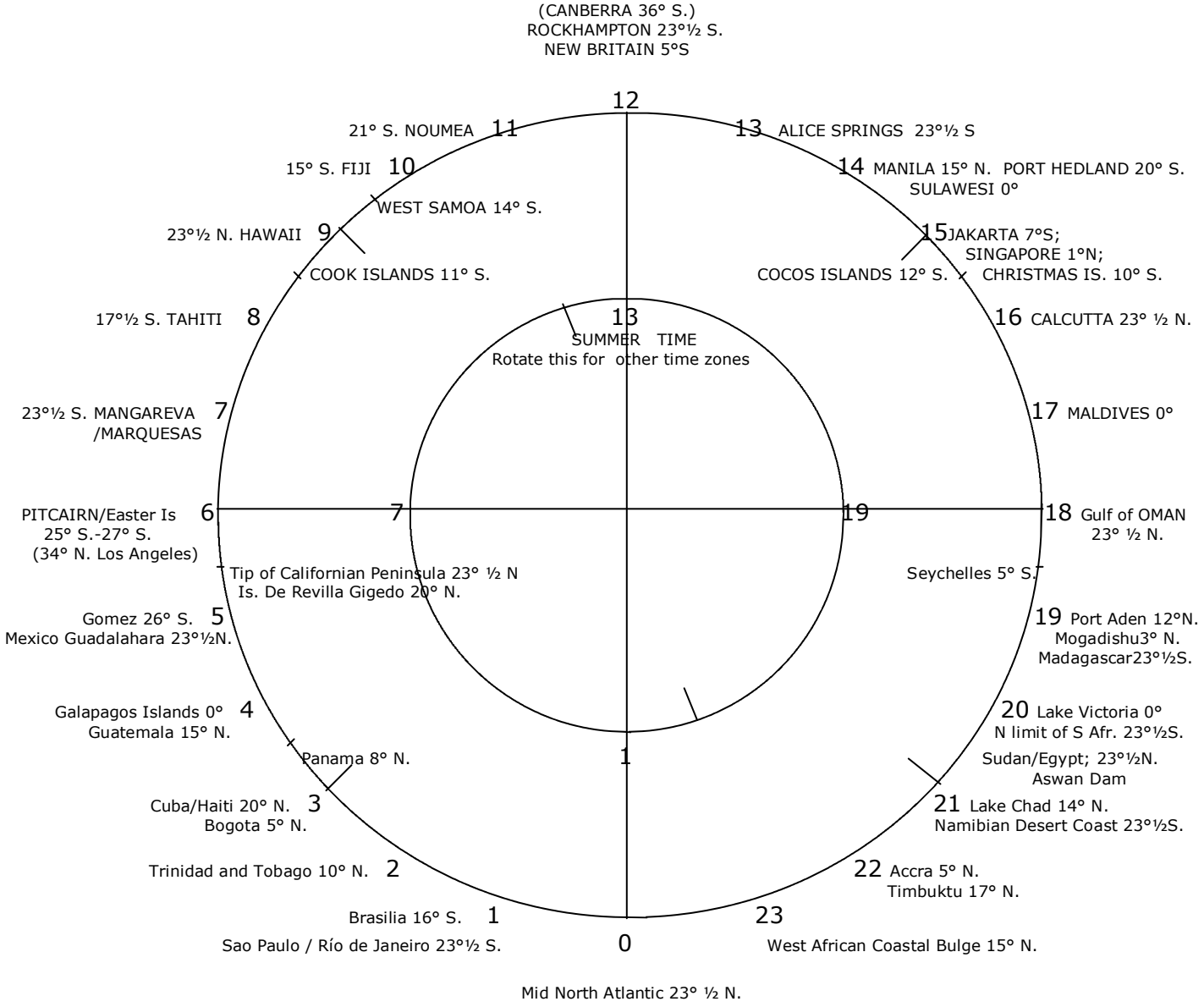

Places in capitals are less than 6 hours away from Australia. The others are on the other side of the world.
View this as if from above the North Pole—the Sun 'moves' clockwise, but the Earth spins anticlockwise.

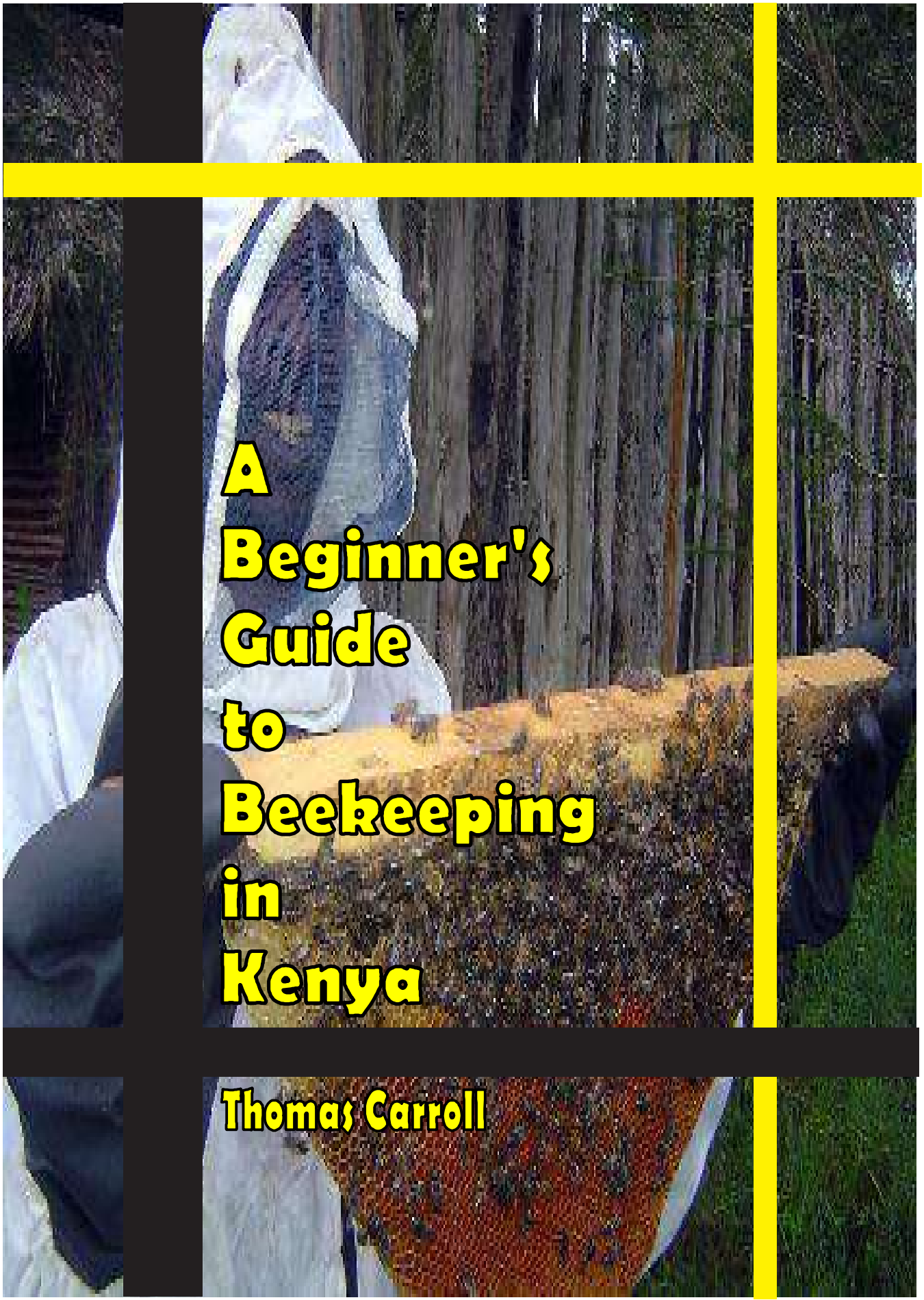

A Beginner's Guide to Beekeeping in Kenya

Thomas Carroll

Contents

ILLUSTRATIONS (ix)

PREFACE (xi)

PART ONE INTRODUCTION TO BEEKEEPING (1)
What is Beekeeping? (1) Getting to Know Bee Species and Races (1)
What Bees Do (2) All about Honey (3) Information on Beeswax (3)
What is Propolis? (3) Residents of the Beehive (3)

PART TWO GETTING STARTED: THE BASICS (6)
Types of Hives (6) Apiary Location (17) Hive Placement (19)
General Apiary Management (22) Beekeeping Equipment and How to Use Them (23)
Handling Bees (25) What If No Bees Enter the Hive? (28)
Causes of Absconding and Its Control (30)

PART THREE THE BEEKEEPING YEAR (31)
Conditions for Maximum Honey Production (31)
Beekeeping Management during the Year (38)
Swarm Prevention and Control (39) Making a Division (40) Feeding (40)
Pests and Diseases and Their Control (41) Major Nectar-bearing Plants in Kenya (43)

PART FOUR THE HONEY HARVEST (47)
Time of Harvesting (47) Harvesting Procedure (47)
Honey Refining (50) Extraction of Beeswax (52)

PART FIVE THE MARKET FOR BEE PRODUCTS (54)
Honey (54) Beeswax (56) Propolis (56)

PART SIX STARTING A BEEKEEPING ENTERPRISE: COSTS AND PROFIT (57)

GLOSSARY OF TERMS (61)

ANNEXES (65)
ANNEX 1 Current Status of the Kenyan Beekeeping Industry (65)
ANNEX 2 More Information on Stingless Bees (67)
ANNEX 3 Hive Recordkeeping (68)
ANNEX 4 Recipes/Uses for Beeswax, Propolis, and Honey (70)
ANNEX 5 Approaches to Beekeeping Extension (74)
ANNEX 6 Sources of Further Information (76)
ANNEX 7 Baraka Agricultural College Short Courses on Beekeeping (81)
ANNEX 8 Beekeeping Resources on the Internet (83)
ANNEX 9 Recommended Reading/References (84)

Illustrations

FIGURES

1. Types or Castes of Bees (**4**)
2. Bee Egg (**5**)
3. Bee Larva (**5**)
4. Bee Pupa (**5**)
5. Construction Drawing of KTBH (**13**)
6. Assembly of KTBH (**14**)
7. Hanging KTBH Posts (**20**)
8. Hanging Patterns for KTBH (**20**)
9. Conditions Required for Maximum Honey Yields (**31**)
10. Seasonality of Honey Production in Kakamega (**35**)
11. Candle Making (**72**)

PHOTOS

1. Wild colony of bees (**6**)
2. The pot hive (**7**)
3. Traditional log hive (**7**)
4. Ugandan basket hive (**8**)
5. Examples of KTBH and Langstroth hive (**8**)
6. Crudely constructed box hive (**9**)
7. The Langstroth frame hive (**10**)
8. Beekeeper in Kitui District (**11**)
9. The Kenya top bar hive (**12**)
10. Top bar (**12**)
11. Working with bees in a KTBH (**15**)
12. Mud/stick hive inside a bee house (**16**)
13. Bee house at Baraka Agricultural College (**17**)
14. Apiary fenced with off-cuts (**18**)
15. KTBH hanging in foreground (**19**)
16. Newly constructed bee house (**21**)
17. Water for bees (**22**)
18. Good gloves (**23**)
19. Farmer in Nakuru wears an inexpensive bee suit (**23**)
20. Smoker (**24**)
21. Lighting the smoker (**24**)
22. Lighting a tin can smoker (**24**)
23. Bee brush and hive tool (**25**)
24. Mick, my son (**25**)
25. Catcher boxes for trapping passing bee swarms (**29**)
26. A tree in flower (**32**)
27. Hive infested with wax moths (**41**)
28. Tree encircled by *mabati* (**42**)
29. Citrus fruit in bloom (**44**)
30. Banana in flower (**44**)
31. Acacia *senegal* in bloom (**45**)
32. Acacia *persiciflora* in flower (**45**)
33. Kei-apple (**46**)
34. *Cordia africana* (**46**)
35. Harvesting white combs (**48**)
36. Harvesting traditional log hive in Tabora, Tanzania (**49**)
37. Harvested honey (**49**)
38. Simple honey refining method for small-scale beekeeper (**50**)
39. Centrifuge (**51**)
40. Uncapping frame full of honey (**51**)
41. Solar wax melter used to extract beeswax (**53**)
42. A sampling of honey products (**54**)
43. Simple box hive for stingless bees (**67**)
44. Hive with shaded glass cover for stingless bees (**67**)
45. An exhibition stand of the Kenyan National Beekeeping Station (**70**)
46. Cake of beeswax for making candles (**71**)
47. Beautiful beeswax candles made using plastic pipe mould (**72**)
48. Bee forage tree seeds (**75**)
49. The teaching apiary at Baraka Agricultural College (**82**)

TABLES

1. Duration of Development Stages (**5**)
2. Beekeepers' Preferences for Different Hive Types (**34**)
3. Plants Suitable for Bees in Molo (**36**)
4. Plants Suitable for Bees in Transmara (**37**)
5. Costs of Starting a Beekeeping Business (**57**)
6. Gross Margin Analysis of Beekeeping (**58**)
7. Gross Margin for a Hectare of Maize (**59**)
8. Hive Recordkeeping (**68**)
9. Marketing Recordkeeping (**69**)

Preface

This book is about beekeeping as practised in Kenya. Wikipedia (a free online encyclopaedia) defines it as "the practice of intentional maintenance of honeybee colonies, commonly in hives, by humans. A beekeeper (or apiarist) may keep bees in order to collect honey and beeswax, or for the purpose of pollinating crops, or to produce bees for sale to other beekeepers. A location where bees are kept is called an apiary". We also know beekeeping as apiculture, from the Latin *apis* or bee.

The purpose of beekeeping is to produce and sell as much honey as possible. Beekeeping is a means to expand livelihoods and make the best use of available resources, in this case, floral nectar and pollen. A diversified livelihood is a more secure one. Beekeeping also has immense benefits in terms of provision of pollinators, which enhance crop yields. Scientists estimate that one in every three bites of food we eat is a result of active pollination of plants in which bees play a very important role. Bees are environment-friendly and so are beekeepers. Many beekeeping groups initiate tree nurseries as they realise the importance of enhancing the environment to improve honey yields. Beekeepers are also the most active in safeguarding natural environments, as they know honey crops are dependent on local vegetation, mostly trees, as their source of nectar. **Teach people to keep bees and they will plant trees themselves.**

Better beekeeping requires good management of bees and hives. Anyone—young or old, men or women—can keep bees. Beekeeping needs very little land, time, or equipment. It is an ideal activity for small-scale farmers in Kenya and other African countries.

Farmers who want to keep bees generally have two challenges:

1. LACK OF INFORMATION ABOUT BEEKEEPING

This is a common problem: where can beekeepers in Kenya and other African countries get relevant information? Information from other continents such as Europe and America is useful but not always applicable. African bees and conditions are different and require appropriate information on local beekeeping constraints as well as opportunities.

2. A FEAR OF BEES

Many people start beekeeping only to run into problems with excessive stings that result in the farmer getting rid of the bees. This is also a common difficulty in Kenya as Kenyan bees are unpredictable and sometimes very aggressive. Knowledge on how to keep and handle bees properly can overcome a fear of bees.

This book assists with information on beekeeping, which helps overcome fears. It also directs to sources of further information. All beekeepers need to keep learning and improving their craft. Learn from bees themselves, neighbours and fellow beekeepers, the local beekeeping officer/extension worker, and all the books and other resources obtainable.

A Glossary is available after Part Six to assist you with definitions for the more important terms used in this guide.

(N. B. *On exchange rates: KSH refers to the Kenyan shilling. Exchange rates at the time of writing are USD1=KSH72; GBP1=KSH136; EUR1=KSH92.*)

Part One
INTRODUCTION TO BEEKEEPING

WHAT IS BEEKEEPING?

Beekeeping is the art of managing honeybees in order to obtain honey, beeswax, and other bee products for food, income, and sometimes, even medicine. Men and women of any age can carry out beekeeping, but there are taboos against women handling bees in some Kenyan communities.

It is an ideal income-generating activity for groups such as those for women, youth, men, church, etc. We have seen a number of very active youth groups involved in beekeeping. They can generate income without owning land, a major constraint to other lucrative endeavours. Beekeeping requires little space; it complements other farm activities. Beekeeping does not need good soil.

GETTING TO KNOW BEE SPECIES AND RACES

A. HONEYBEES

There are many different species of bees in the world, most of them solitary. A few species of bees are kept to produce honey. In Africa, America, Australia, and Europe, we have *Apis mellifera*, which is a very important species for honey production. Recent research has shown that all races of *Apis mellifera* have origins in Africa. In Kenya, we also have Apis mellifera or the species most familiar to everyone: the honeybee. This book is mostly about the honeybee. Within this species, there are a number of bee races in Kenya that have their particular characteristics and adaptations to our environmental conditions: *Apis mellifera scutellata*, *Apis mellifera monticola*, *Apis mellifera yeminitica* (*nubica*), and *Apis mellifera littorea*.

1. *Apis mellifera scutellata*

These are small with relatively short tongues. The bee is highly aggressive and has great tendency to reproduce (swarm) and abscond (migrate). It is found in plains. Massive flowering attributes to their high reproductive rate, which occurs in plains just before rains. Scutellata is able to nest in a broad range of sites: from cavities to open nests.

2. *Apis mellifera monticola*

This bee is called the mountain bee. *Monticola* are a large dark gentle race with longer hairs than other African bees. The bee inhabits places where clouds, mist, and nocturnal ground frosts obscure the sun. It is the largest bee in Africa. It

has a tendency to reduce brood rearing at the first sign of forage decline and may not migrate. It is less productive and less vicious than other species. *Monticola* is found in Meru and Mt Elgon.

3. *Apis mellifera yemenitica* (formally, *Apis mellifera nubica*)

This is the smallest race in Africa. It has the most slender abdomen and the largest yellow abdominal colour band of all African races. It withstands and survives drought conditions by excessive migration. It is found mostly in the northern parts of Kenya.

4. *Apis mellifera littorea*

The bee inhabits lowlands of the Kenyan coast. It does not migrate as much as *scutellata*. It has a tendency to rear brood throughout the year due to availability of forage along the coast.

Bee races I am most familiar with are *A. m. monticola* and *A. m. scutellata*. *A. m. monticola* tend to be in the highlands and more docile. We probably have these around Molo. *A. m. scutellata* are smaller and more aggressive bees found in the lowlands of Kenya. In my immediate vicinity, we have them in Nakuru and Baringo Districts. We may also have hybrids of these two types. Much research has yet to be done in Kenya and Africa as a whole on mapping various bee races and their characteristics.

(Reference: *Beekeepers' Guide Book*, National Beekeeping Station, Nairobi)

B. STINGLESS BEES

There are also species of stingless bees in Kenya. Beekeeping with stingless bees is called meliponiculture. These bees produce honey prized as medicine. The honey is very sweet and liquid compared to *Apis mellifera* honey. Stingless bees may be kept in small hives but are not yet kept commercially in Kenya. There is renewed interest in these bees with the discovery of a new species in Kakamega by a scientist working for the National Museums of Kenya. Beekeepers in other African countries keep stingless bees. There is even a very fascinating webpage about someone in Moshi, Tanzania, who keeps them. Stingless bees are not, however, the main focus of this book; honeybees are.

(N. B. *See Annex 2 for more information on stingless bees in Kenya.*)

WHAT BEES DO

- Bees forage on floral nectar and pollen. They don't compete with livestock for food such as grass.
- Bees help the pollination of flowers, plants, and crops.
- Bees increase the quantity and quality of flowering crops, e.g., coffee, papaw, banana, avocado, macadamia, mangoes, etc.
- Bees fly three to five kms from the hive to collect nectar, so you benefit from other people's flowers.

- Bees produce honey, beeswax, and propolis (used in medicines) as well as other products such as royal jelly, pollen, bee venom, and bee brood. Honey, beeswax, and propolis are the three products currently exploited in Kenya. They are discussed in this book because of their importance.

ALL ABOUT HONEY

- Honey contains sugars (80-85%) that are easily absorbed by the body—good for young, old, and sick. It is made from nectar, a sugary secretion of flowers. Nectar contains 70-80% water. To make honey, bees add enzymes and reduce the water content to that of honey. (Good honey contains less than 19% water.)
- Honey is very good energy food. Use it as sweetener for food (cake, chapati, bread, etc.) and drink. Try some in your tea.
- Honey has medicinal properties. Use it for coughs, ulcers, wounds, and sore throats. (*See recipe for honey cough syrup in Annex 4.*)
- Honey has a high market value. It is used to make local beer and is also a food preservative.
- Honey is always in demand. It is a good source of cash.

INFORMATION ON BEESWAX

- Beeswax is the substance bees use to make their combs. Worker bees secrete it from special wax glands on the abdomen (underside).
- Beeswax is used in cosmetics, soaps, leather and wood preservation, candles, ointments, batik, shoe polish, and foundation sheets for Langstroth hives.
- You can turn honeycombs into wax. Simple instructions and recipes for using the wax are found in Annex 4.
- With a little training, you can make and sell candles, body cream, and shoe polish made with beeswax. Farmers trained by the Baraka Agricultural College/SHDI beekeeping outreach project sell homemade body creams and generate additional income for themselves.

WHAT IS PROPOLIS?

- Propolis is a resin that bees collect from plants. It is black and sticky.
- Bees use propolis to cover the inside of the hive and fill in the cracks.
- It has medicinal value, e.g., as an antibiotic.
- If you chew propolis, it tastes bitter but is good for the throat and chest.
- There is an export market for propolis. Annex 4 has a recipe for making propolis ointment.

RESIDENTS OF THE BEEHIVE

In the beehive, there are three types or castes of bees. It is important for all beekeepers to recognise them. The first caste is called the Queen: a sexually mature female. The second caste is called the Drone: male bees. The third caste is the Worker: an immature female. Details of all three castes plus the brood are given below.

Figure 1: Types or Castes of Bees
Only the worker bee does all the work in the hive!

1. The Queen Bee

The queen is a sexually mature female. There is only one queen in the hive. Her job is to lay eggs that hatch into other bees. She can lay up to 2,000 eggs per day. Queens can live for up to five years, but the most productive period of their life is the first two years.

The queen may be hard to find in the hive, but you can recognise her by her length. She is long and slender and her wings reach only halfway down her back. She is far bigger than the numerous workers but must not be confused with drones, which have a square body and very large eyes.

2. The Drone

There can be several hundred drones in a hive depending on the time of year. In times of food shortage, workers throw drones out of the hive. When a new queen starts life, she mates only once with up to ten drones outside the hive, high up in the air.

Drones have very large eyes used to spot the queen during mating. Their major task is to mate. They have no sting. Drones look large and square. They make a loud buzzing noise when they fly.

3. The Worker

Most bees in the hive are worker bees. As their name suggests, they do all the work. Workers feed the young, feed the queen, guard the hive, and collect honey/pollen/ propolis and water. Workers undertake different tasks in the hive depending upon their age after hatching. Worker bees have a sting and use it to defend their nest. Workers can number up to 60,000 in a very strong colony of bees. The lifespan of a worker bee depends on colony activity: during times of high activity, e.g., when collecting nectar to make honey, they can wear out in a few weeks. In the dormant season when there isn't much work, they can live for up to six months.

4. The Young or Brood

The queen lays an egg that looks like a grain of rice. To see eggs, hold combs from the centre of the brood nest (centre of the hive) up to the light. Look carefully at seemingly empty cells; they usually have eggs at the bottom.

The eggs develop into larvae that look like white maggots. The larvae are fed on brood food (produced in glands on the head of nurse worker bees) and some pollen/

honey. (Pollen is the powdery substance produced by the anther of flowers and is rich in protein. It is the coloured substance seen on bees' legs and stored in combs.) Larvae are later sealed in the comb where they turn into adult bees. During the changing process, they are called pupae. Thus there are three stages to adulthood: egg, larva, and pupa.

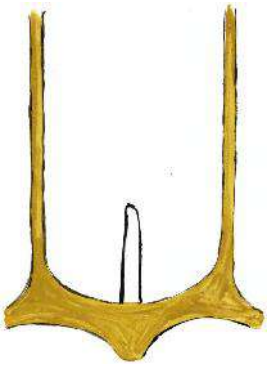

Figure 2: Bee Egg

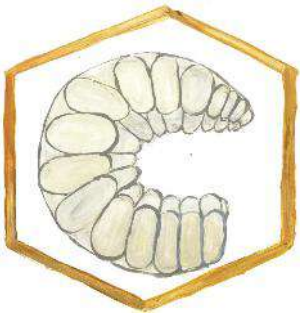

Figure 3: Bee Larva

Figure 4: Bee Pupa

Bee Caste	Eggs-Larvae (# of days)	Larvae-Pupae (# of days)	Pupae-Adults (# of days)	Total Days (Egg to Adult)
Queen	3	4.5	7	14.5
Worker	3	4.25	11.25	18.5
Drone	3	6.5	14.5	24.0

Table 1: Duration of Development Stages
Table shows time spent in each stage of development for African bees, specifically *Apis mellifera scutellata*.
(Source: Kigatiira, K. I., *Beekeeping for Beginners*. See Annex 9 for more details.)

Part Two
GETTING STARTED: THE BASICS

TYPES OF HIVES

A hive is the box or some other container where bees live. Without a hive, you cannot be called a beekeeper!

Photo 1: Wild colony of bees
Wild colony of bees nests under a manhole cover.

Different types of hives are used in beekeeping. Some are:

- Pot hives or clay pots (See Photo 2.)
- Log hives (See Photo 3. They are made from hollowed-out tree trunks.
- Basket hives (See Photo 4. They are made from woven sticks smeared with mud.)
- Kenya Top Bar Hive or KTBH (See Photo 5.)
- Langstroth hives (See Photo 5. These use frames.)

Photo 2: The pot hive
The pot hive has a super or honey chamber on top.

Photo 3: Traditional log hive
A traditional log hive located in Koibatek District.

Photo 4: Ugandan basket hive
The lid is removed to harvest honey. Care in harvesting yields top quality honey.

Photo 5: Examples of KTBH (left) and Langstroth hive (right)
Hives in Turkana District. Beginners should go for the simpler and cheaper KTBH.

In this book we will give you measurements of the KTBH and a cheaper mud/stick hive, which also has top bars. These two hives are relatively easy and inexpensive to make. They are very good hives to start with in beekeeping in Kenya. The Langstroth-type frame hive is more complicated and expensive to make/purchase. It has been promoted in Kenya the past few years. Each type of hive has advantages and disadvantages.

ADVANTAGES OF TRADITIONAL HIVES

(e.g., logs, pots, and baskets)

- They are relatively easy and inexpensive to make.
- You can use local materials and traditional knowledge for making them.
- Bees often prefer these hives to more 'modern' types.
- With care in harvesting and handling, honey quality can be very good.

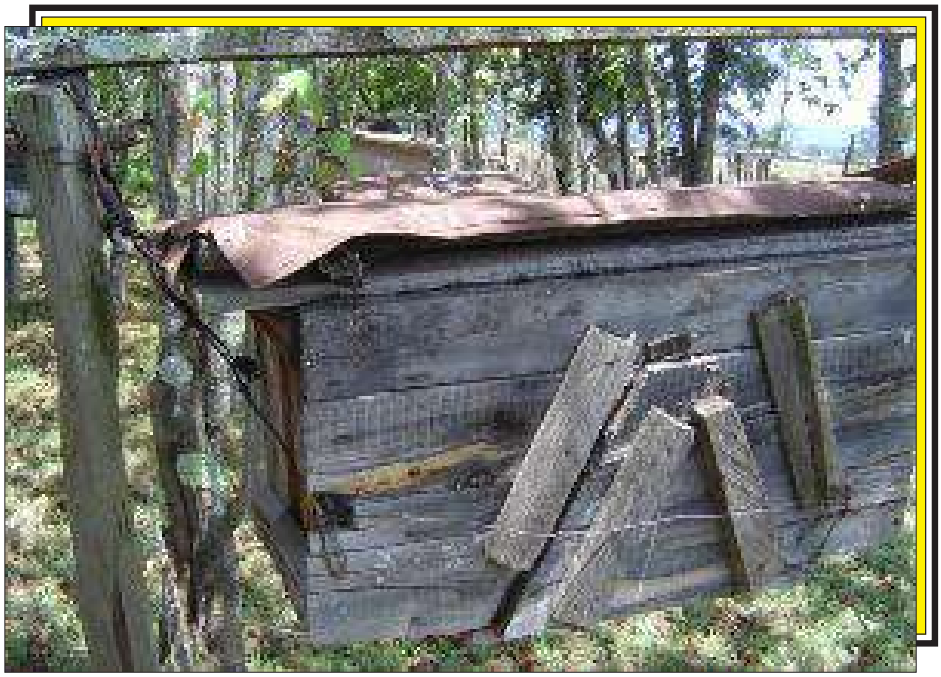

Photo 6: Crudely constructed box hive
This box hive is located in Londiani, Kericho District.

DISADVANTAGES OF TRADITIONAL HIVES

- Bee management is difficult.
- Logs are no longer readily available to make hives in Kenya.
- They can be difficult to harvest.
- Yields and quality of honey can be lower than top bar and frame hives.

ADVANTAGES OF TOP BAR HIVES

- Bee management is possible, making it easy to harvest and manipulate colonies to maximise honey production.
- They are easier and cheaper to make/buy than frame hives.
- They are easier to harvest than traditional hives, creating the potential for better honey quality.
- Honey extraction is easy compared to frame hives, as there is no need for specialised equipment.

DISADVANTAGES OF TOP BAR HIVES

- Combs in the top bar hive are not supported. They can break if not handled carefully.
- Wax is harvested with honey forcing bees to build more wax to replace harvested combs, which results in lower honey yields but more harvested wax.
- The volume of the KTBH, like the traditional hive, is fixed, so the hive can fill very quickly in the honey season. An overcrowded hive can swarm leading to reduced honey crops.

ADVANTAGES OF FRAME HIVES

(e.g., Langstroth)

- It has frames that make the combs very strong especially when transported.
- The honey is extracted by centrifuge, returning wax to bees with potentially much greater honey yields.
- The space of the hive can be added to easily by adding more supers. Additional space is important during honey flow. Supers can be removed when the honey flow is over to contract hive size.
- Using a queen excluder between the bottom brood box and the honey supers means honey is separated completely from the brood and is of high quality.

Photo 7: The Langstroth frame hive

DISADVANTAGES OF FRAME HIVES

- Frame hives are more expensive to make than traditional or top bar hives.
- They are more complicated to make and manage, as they require more spare parts such as frames, foundation starter sheets, and supers, which are costly and not readily available to beekeepers in Kenya.
- Extracted/unextracted honey supers are prone to damage by wax moths, ants, and rodents.
- Frame hives need a centrifugal extractor to extract honey. Extractors are not readily available in Kenya and are expensive to purchase. Beekeepers are often forced to cut honey from frames just as in top bar hives.
- Frame hives require a greater level of investment (time and money) as well as beekeeping skills to make them pay off.

In recent years, Langstroth frame hives have been promoted in Kenya. There has been considerable hype around their use with many people automatically believing they are the best hives. Frame hives are not necessarily better than either traditional or top bar hives. Frame hives have the potential to be better in terms of higher yields and better quality honey **if** managed well! Unfortunately in many cases, too much emphasis is placed on the technology and not on building beekeeping skills and capacities.

Photo 8: Beekeeper in Kitui District
Beekeeper poses in front of Langstroth hives.
Donor support provides many such hives in Kenya.

As a beginner, you can easily start beekeeping in Kenya using the intermediate technology top bar hive such as the KTBH—unless someone donates a Langstroth to you! Once you have beekeeping experience, understand why you should purchase a frame hive, and have money to do so, and then use the frame hive by all means, but you should be aware that there are no automatic increased yields from having a Langstroth. The hive is only as good as you manage it. I have produced up to 40kgs per KTBH while many people who purchased Langstroths have produced little or nothing! So understand bees and bee management and you will produce good crops of honey no matter the type of hive you have. When you have better skills and knowledge, move to the more expensive hive, which can then give better yields.

This book deals primarily with top bar hives because they are the easiest for beginners in Kenya who often do not have a lot of money to spend for starting up. The principles of beekeeping are, however, the same for both top bar and frame hives and issues of bee management are very similar.

THE KENYA TOP BAR HIVE

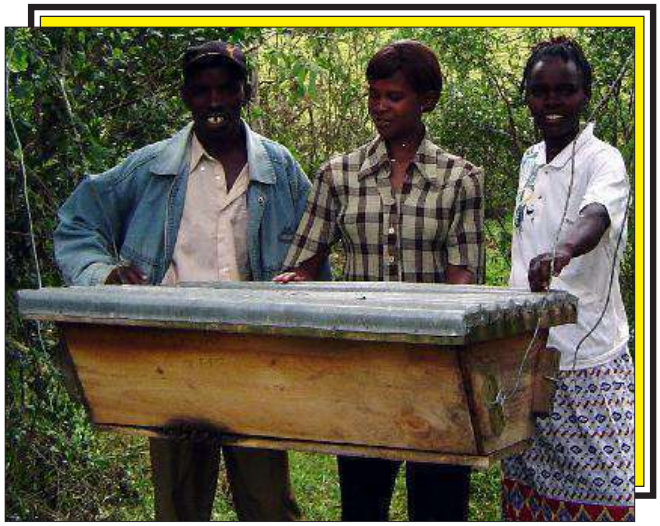

Photo 9: The Kenya top bar hive
The Kenya top bar hive or KTBH was developed in Kenya and is now used around the world.

The Kenya Top Bar Hive (KTBH) was developed in Kenya and is a good option for many Kenyan beekeepers and other beekeepers from around the world. KTBH are also used outside Africa, e.g., the United States of America. Its big advantage is its simplicity and low cost. You also get to produce beautiful comb honey for home consumption or for sale. Figure 5 shows the construction drawing of a KTBH. Remember to follow the measurements—the top bars, in particular!

Very important: You must copy **exactly** the measurements we give you. When it comes to bee management, it is important that all your hives are the same size so you can move top bars from one hive to another. This allows better bee management. In particular, watch the dimensions of the top bar: 3.2cm wide and 48.3cm long. The ridge at the centre of the top bar is coated with beeswax to give bees a line to follow when they build their comb. The smell of beeswax also attracts bees to enter a new hive.

Photo 10: Top bar
There are two views provided for the top bar: top view (top) and side view (bottom). Notice the ridge protruding from centre in side view.

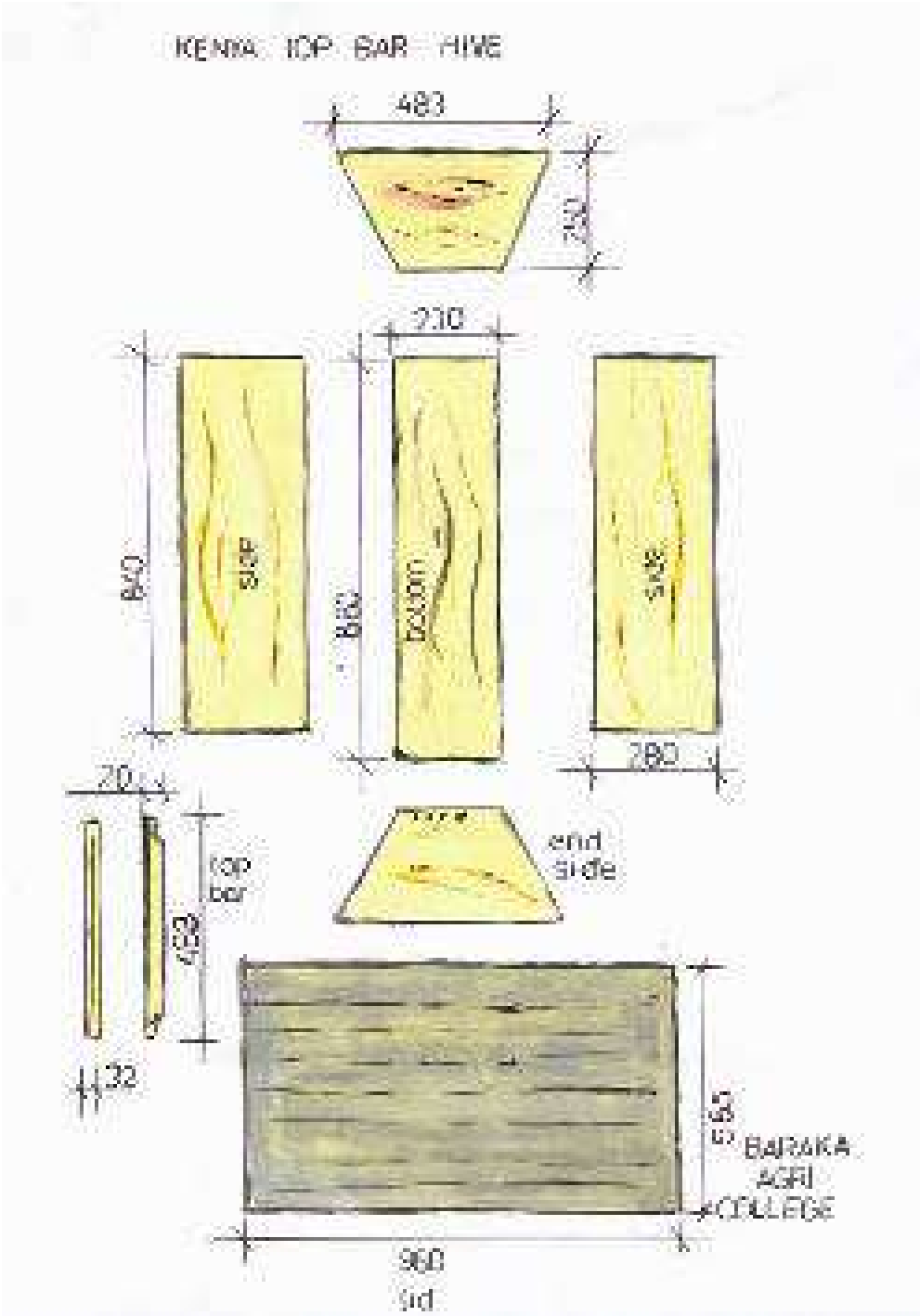

Figure 5: Construction Drawing of KTBH
Measurements are in millimetres.
(Source: Baraka Agricultural College Workshop)

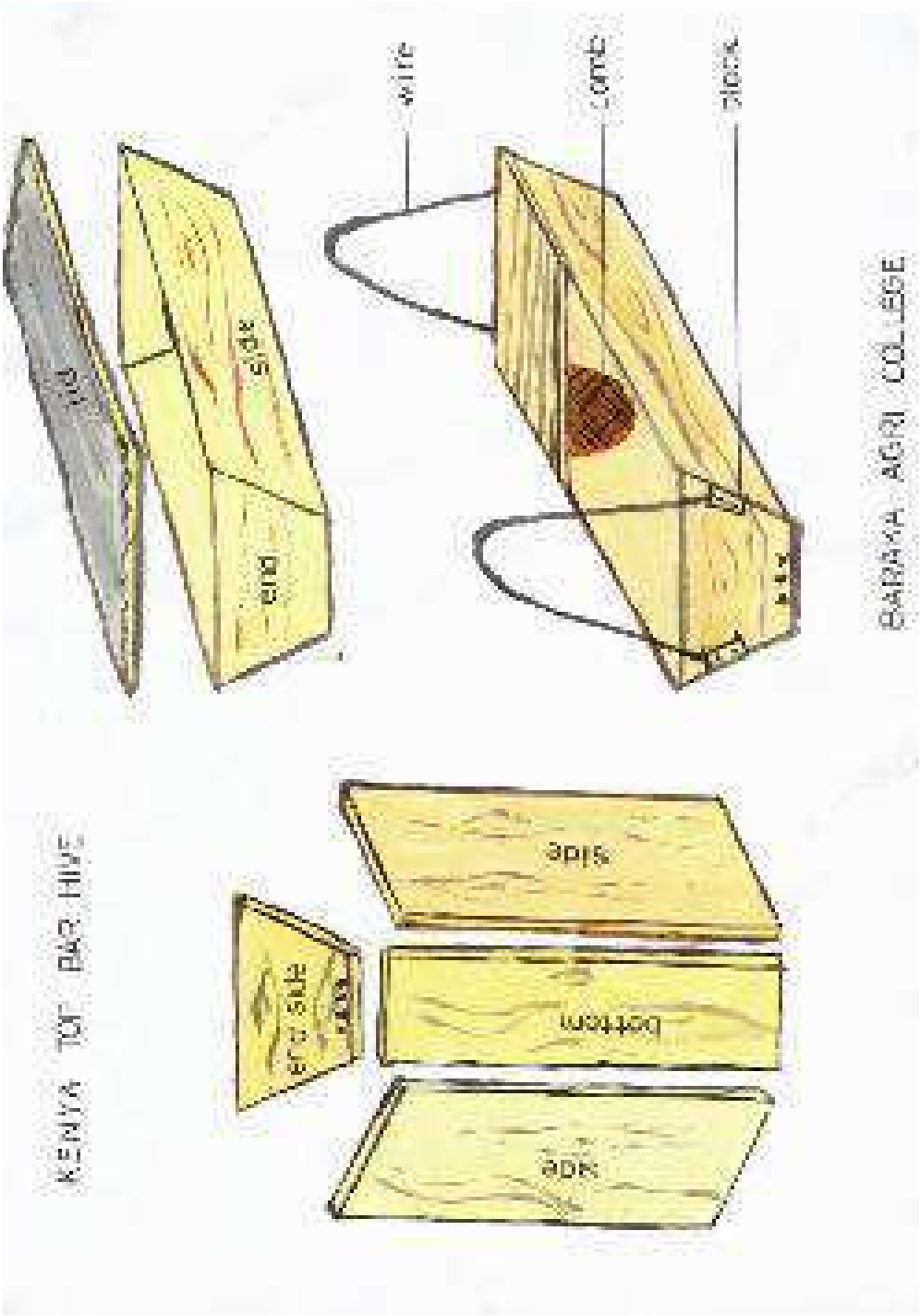

Figure 6: Assembly of KTBH
(Source: Baraka Agricultural College Workshop)

TIPS FOR MAKING A GOOD KTBH

- Use timber ¾ inch thick (19mm) to make the hive.
- Utilise dry timber to prevent cracking and warping.
- Make use of whatever timber is inexpensive and available locally. Pine, cypress, and *Grevillia robusta* are commonly used, but if you have your own wood, use it.
- Get the right size for the hive body of the KTBH though some drawings show a slightly longer hive. The measurements given are those used at Baraka Agricultural College.
- Follow the measurements of the top bars exactly. It is **very important**. Top bars are made a specific measurement to ensure that bees build one comb per top bar making combs easy to inspect.
- Top bars should be the right width and fit the hive body well.
- Don't forget to wax top bars to attract bees. Use a paintbrush to apply melted beeswax onto protruding ridge at the centre of the top bar. The strip of wax guides bees to build straight combs that are easy to inspect and harvest.
- All parts of the hive should fit together properly. There should be no holes.

Photo 11: Working with bees in a KTBH
Notice one comb per top bar, which allows modern bee management by the ability to move combs without breaking them.

THE MUD/STICK HIVE

A mud/stick hive is the answer to all of you who say you have no money to start beekeeping. No excuses allowed. Beekeeping is about skills and knowledge. Even with very little money, you can start beekeeping by making your own equipment.

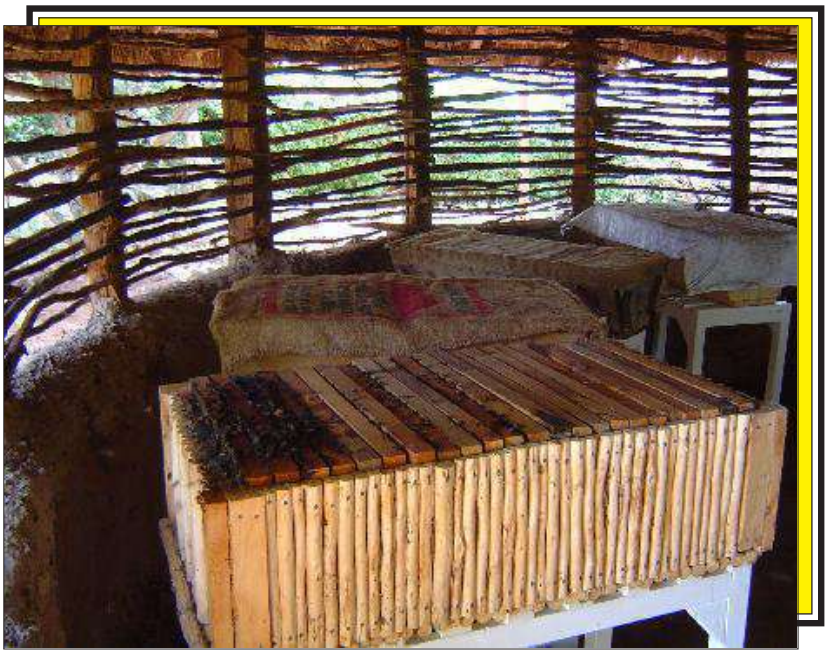

Photo 12: Mud/stick hive inside a bee house
The bee house is also made from mud and sticks.
The hive body is made from sticks with gaps between them filled with mud.

For the mud/stick hive, use the same dimensions (length, width, and height) as the body of the KTBH, but make the body square. For the body, use about 50 long and 30 short straight sticks. First, make a frame of sticks for the bottom. You can also make a timber frame and add sticks as shown in Photo 11. To this frame, tie/nail on other sticks. Use wires or nails to do this. If you bind them tightly, you can have a very strong frame. Plaster the inside of the hive with mud and cow dung and leave to dry. Top bars fit along the top of the hive just like the KTBH. Remember to make some small holes at one end of the hive as the entrance. As a roof, you can use *mabati* (tin). The mud/stick hive works best, however, under shelter, e.g., a bee house, because of potential damage from heavy rains. If inside one, just use a sack for a cover.

Photo 13: Bee house at Baraka Agricultural College
Baraka Agricultural College is in Molo.

TIPS FOR MAKING A GOOD MUD/STICK HIVE

- Shelter hive from heavy rains.
- **Important**: make top bars the same measurement as that of the KTBH, which allows transfers from one type of hive to the other.

The two hive designs mentioned are an improvement of the traditional log and basket hives because with top bars, combs are moveable, which allows management of bees. Top bar hives are not as expensive or difficult to manage as Langstroth hives making them more suitable for use by small-scale farmers.

You can also be innovative and use other materials to make the hive body. In Zambia, beekeepers use a concrete KTBH hive, which works very well. The body is made from concrete and set on a stand. Top bars are added on top of the hive and covered with sacking material. The Zambian concrete hive is placed under grass shade cover to keep out rains and the hot sun.

APIARY LOCATION

An apiary is a place where hives are kept. Keep a limit of not more than 20-25 hives per apiary depending on availability of bee forage. Bees forage in a radius of about three to five kilometres from the apiary, so if you want to keep more

than 20 hives, find another site three or more kilometres away from the existing one. Plant a good high hedge around your apiary using a shrub such as kei-apple. As you wait for the fence to grow, you can use off-cuts—waste timber from sawmill—to make a fence. The hedge separates bees from people and animals, which is important in Kenya as our bees can be aggressive. Apiary siting is very important on small farms to ensure safety and comfort living with bees.

Photo 14: Apiary fenced with off-cuts
The fence, made of waste timber, is to screen bees away from people and animals. (Nakuru)

- Choosing a good site to hang your hives is very important.
- If you choose a poor site, bees may sting people and animals.
- If the site is insecure, hives and honey may be stolen.
- If you live in a hot area, your hives will need shade as well as water.
- If you live in a cool area such as the highlands, only minimal shade is required or the bees will be cold and damp.
- Get a good balance between light and shade. A bright apiary without direct sunshine appears to be best. Bees are most active then.

An ideal site to establish an apiary would be:

- Away from humans and livestock, dwelling areas, roads, and public areas.
- Safe from strong direct sunshine, windy areas, and theft.
- Accessible to nearby water and bee plants, i.e., provide water and bee forage shrubs, trees and/or crops.

- Away from swampy areas and smelly places.
- Near a good source of nectar, e.g., forests, trees, or nectar-bearing crops.

So now you have made a hive or two. Did you follow instructions carefully? Are your top bars the right measurement? Have you chosen the site for the apiary? If you are still not sure, ask someone who knows. Remember: once bees enter hives, it will be more difficult to change things. The bees may become aggressive or the hive will be heavy with brood and honey. We now come to the procedure for putting hives in the apiary.

HIVE PLACEMENT

A. HANGING HIVES

(N. B. *Not recommended for mud/stick hive, as it is not strong enough. See below for placement.*)

Photo 15: KTBH hanging in foreground
Another one is on a stand, background left.

- Use two strong and heavy posts, each of about 2.5 to 3 metres long.
- Dig two holes about ¾ metre deep and 2 metres apart—or take two strides.
- Pack soil and stones around posts. Make sure posts are very firm. Think ahead to when the hive will be heavy with honey. If they are not firm, they will fall over later.
- **Remember**: once bees enter the hive, it will be difficult to make changes!
- Now that holes are dug and posts are in position, use wires to hang the hive between the two posts.
- **Remember**: put wire around the back of the posts as shown in Figure 7.
- **Remember**: hang the hive at waist height and keep it level to ease work and not strain your back.

Take time to do things properly and you will be happy later on!

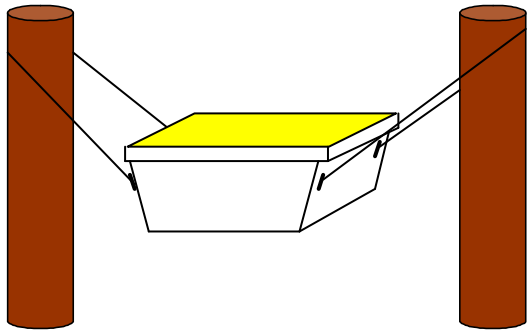

Figure 7: Hanging KTBH Posts
The posts are two metres apart and the hive about one metre from the ground.

You can also hang hives between two trees or a post and a tree. Follow the same instructions, but remember to use strong heavy nails. When hanging more than one hive, you can follow some of the patterns below:

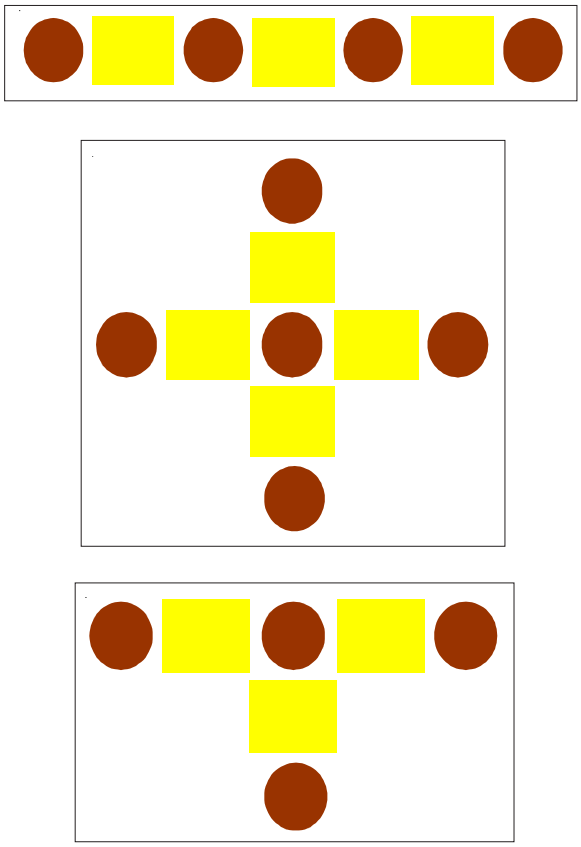

Figure 8: Hanging patterns for KTBH
Above are the line pattern (top), cross pattern (middle), and T-pattern (bottom).
(□ = hive, ○ = hanging post)

B. HIVE PLACEMENT

- If you don't want to hang hives, you can place them on stands. It is best to use this method for the mud hive, as it is not strong enough to hang.
- **Remember**: the stand should be sturdy and high enough for the hive to be at waist level.
- Put legs of the stand in cans of used engine oil to prevent pests from getting into the hive.
- Hives on stands are more prone to attack by the honey badger. (See list of pests and diseases in Part Three.)

(N. B. *Putting a number of hives under a shelter or into a bee house is also an option. Just make sure that bees have a way of getting in and out of their hives. A bee house is useful because you can lock it to prevent thieves from stealing honey or children from interfering with hives. A bee house can be a simple mud hut with holes in the walls for bees to come and go. From my experience with bee houses, get a balance between shade and light: no strong direct sunshine but also plenty of light. The balance of light and shade will depend where you are. In a cold highland area, you need less shade than in the hot lowlands.*)

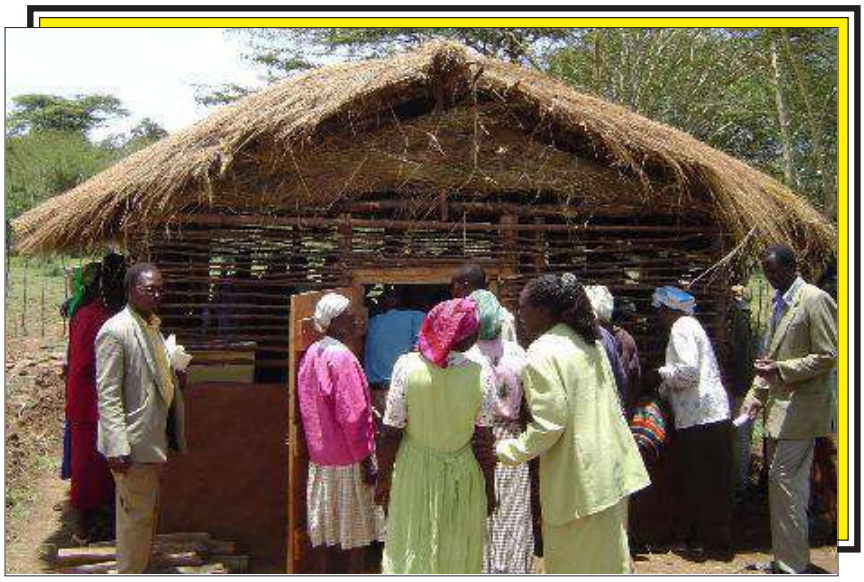

Photo 16: Newly constructed bee house
Farmers in Koibatek and Ministry of Livestock staff inspect the structure.

GENERAL APIARY MANAGEMENT

- Keep apiary clean: cut grass short and trim branches that reach hives to prevent pests, e.g., ants, from crawling into hives and disturbing bees.
- Grease wires holding hives to keep off crawling pests, particularly ants.
- Do not make noise around the apiary as it can agitate bees.
- Supplement the bee's source of nectar by cultivating certain plants around the apiary, e.g., bananas, sunflower, mangoes, citrus, coffee, eucalyptus, pawpaws, passion fruits, croton, acacia, bottlebrush, *Calliandra*, etc. Trees will also give shade to bees and provide a screen between bees and nearby people and animals.
- Remove old combs from the hive to prevent wax moth damage.
- Do not spill honey near the apiary. It will attract pests and cause bees to fight with their neighbours, i.e., bees from adjacent hives.
- Inspect hives regularly to monitor bee progress and ensure no pests are attacking them. (See list of pests and diseases in Part Three.)
- When bees are making honey, inspect hives once a week by opening them up. This is important, so you ensure bees have enough space to store honey. With the KTBH hive, remove honey to make space; with the Langstroth, add an extra super. At other times, inspect hives once a month. You can also simply look at hives almost daily without opening them.
- Observing bees for a few minutes can tell you a lot. Are the bees collecting pollen? If they are, it indicates that bees are feeding the young. It also shows they are healthy and have a laying queen.
- Provide a permanent source of water for bees near the apiary. This is important as lack of water can contribute to absconding in the dry season.

Photo 17: Water for bees
Provide water slightly away from apiary and not in their flight path to avoid bees defecating in the water, which spreads disease. Place sticks for bees to stand.

BEEKEEPING EQUIPMENT AND HOW TO USE THEM

1. Hive

The hive is for housing bees. (KTBH and other types have already been mentioned.) It is the most important piece of beekeeping equipment. Without a hive, you cannot be a beekeeper!

2. Bee Suit

The bee suit consists of the following items that give protection from stings:

- Veil: for covering head and face
- Overall: for covering the rest of the body
- Gloves: for covering hands
- Gumboots: for covering feet

Photo 18: Good gloves
Good gloves are essential when working African bees.

(N. B. *Farmers are able to improvise their own bee suits and make them out of sacking material, which cost about KSH300 or US$4.17 to assemble.*)

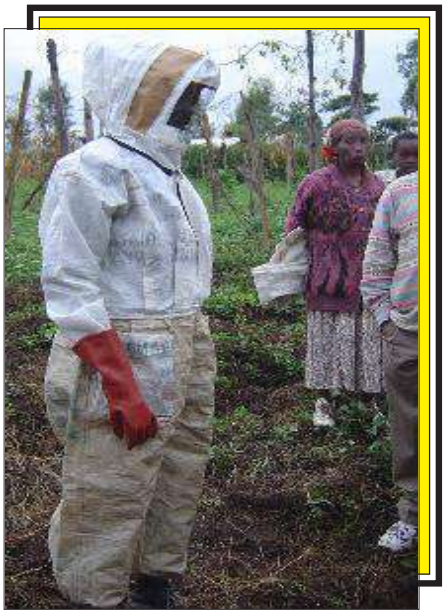

Photo 19: Farmer in Nakuru wears an inexpensive bee suit
It is made from used sacks combined with rubber gloves.

3. Smoker

The smoker is a vital piece of equipment as smoke is used to control bees and keep them from becoming aggressive when handled. When lighting a smoker, use dried grass to light it. When burning well, put in some damp grass, so that it produces a thick cool smoke. Do not allow the smoker to spark or flame, which will burn bees. An alternative smoker fuel is dry cow dung. Be careful when extinguishing the smoker that you do not start a fire!

Photo 20: Smoker

(N. B. *A simple smoker can be made from a tin can punched with holes and with a wire handle attached. The beekeeper blows smoke onto bees with his/her breath.*)

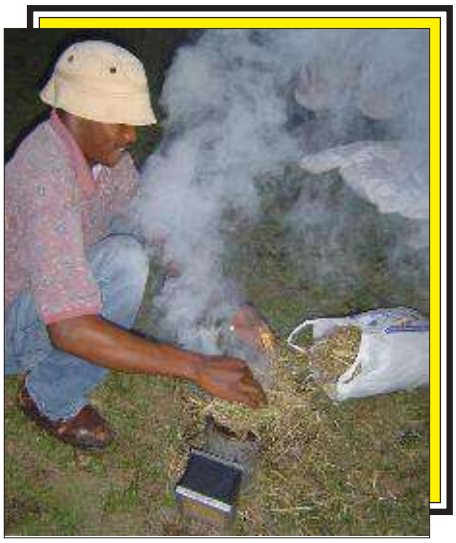

Photo 21: Lighting the smoker
Use cool smoke from damp grass. Do not use toxic materials, e.g., rubber or plastics.

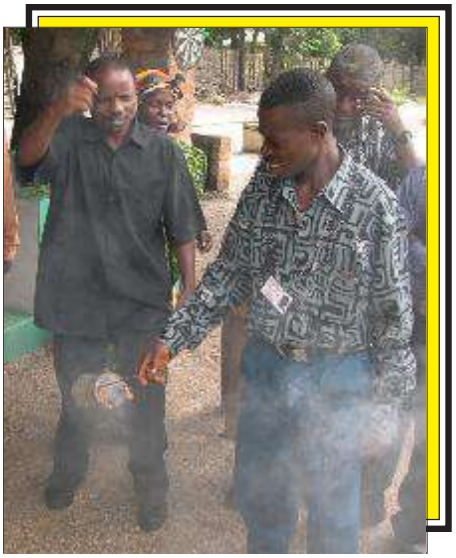

Photo 22: Lighting a tin can smoker
Beekeeping trainees in DR Congo light an alternative smoker. In beekeeping, all equipment can be improvised at low cost.

The hive, bee suit, and smoker are the most important pieces of equipment. The following items can also help:

4. Catcher Box
It is used for catching swarms and transferring bees from one place to another. It is a miniature hive.

5. Hive Tool
It is for lifting top bars. You can improvise the hive tool by using a strong knife.

6. Bee Brush
It is for brushing bees from combs when harvesting honey or at any other time you need to remove bees. You can use a feather as an alternative.

Photo 23: Bee brush (left) and hive tool (right)

A harvesting bucket is also essential as well as a knife for cutting combs when harvesting honey.

HANDLING BEES

Photo 24: Mick, my son
My eight-year-old son works African bees in Nakuru. He wears a full suit, gloves, and gumboots. If a young boy can handle bees without fear, adults should be able to do the same. When you understand bees, fear goes away!

Many people are afraid of bees because they sting. Many of us in Kenya hear of stories where bees attack and even kill people and livestock. (African bees can be very aggressive. Handle them carefully.) Like other forms of livestock, bees must be handled with respect and care. If managed properly, bees will not cause any problems.

Some bee types tend to sting less than others. A beekeeper who frequently inspects his/her hives can easily tell those bees, which are better and more docile. He/she can then eliminate more aggressive bees and breed from docile ones. (Not all African bees are aggressive. I have worked bees in Northern Somalia, which were very docile and didn't require a veil! Our bees in Molo are also generally docile.)

With frequent handling, bees appear "used" to inspection so less aggressive. A beekeeper will come to know the character of his bees. If you have many hives, number them and keep records of the bees' behaviour at each inspection. Eliminate aggressive colonies.

When handling bees:

- Always wear a bee suit and take the time to put it on properly. Many people do not do this and get stung when hives are open. If you don't know how to put on a bee suit, get someone to help you—maybe your experienced friend.
- Avoid wearing woollen clothes as they agitate bees, which can be stuck in them.
- Avoid drinking alcohol and using strong smelling soaps or sprays all of which may aggravate bees.
- Always use a smoker when handling bees. Smoke makes bees suck honey from the combs. It calms them down and masks alarm smells the bees use for communication. Do not let the smoker go out during the operation or the bees can become aggressive. Keep plenty of smoker fuel handy as you work. It is always better to have two smokers alight in case one goes out.
- Start with the least aggressive colonies always. This allows you to work in peace with the pleasant colonies first.
- Work gently and quietly. Do not knock or bang the hive as this makes bees angry.
- Always handle bees in the evening between 5.30pm and darkness. (Be careful initially. Once you know your bees and their aggressive tendencies, you will know the best time to handle them.) If bees become aggressive during evening hours, they have a chance to cool down before the following morning. They seem to be less aggressive in the cool of evening. For bees you've never handled before or for very aggressive bees, take extra precaution in handling them at dusk. Use a torch to see them.
- When handling, avoid crushing bees or making sudden movements. Work carefully and with confidence. Remain calm even if bees become aggressive. If bees appear to be getting out of control, close up the hive and try another day. If bees get into your veil, remain calm. Walk to a safe distance before rectifying the problem.

- Work hives with two or more people at a time. One person can lift out combs while the other uses the smoker. This allows better control of bees.
- Do not stand in front of hive entrance when examining hive. Bees flying in and out may become agitated to find their way blocked. Always cut down disturbance to bees in every way you can.
- Advise any onlookers to move away quietly if stung, covering their eyes. No running about waving arms as this can annoy bees.
- Remove bee stings from skin as soon as possible using a hive tool or your nail to scrape it off. Trying to pull it out tends to squeeze in more venom. Use smoke to cover the scent of a sting. When a bee stings, the scent attracts other bees to sting you again if you do not use smoke.
- On finishing the job, close up the hives. **Do not go directly to where you are to remove your bee suit.** Take a route via bushes or tall maize, sugarcane, etc., if these are in the vicinity. Rub against leaves to rid yourself of bees, which might be following or on you.
- If you are with a fellow beekeeper, check each other and see that there are no bees on yourselves before removing bee suits.

As you gain beekeeper experience, you will be able to judge the mood of bees more accurately and handle them calmly. For the first few times, it is better to get a friend to help you. I have a short video online to demonstrate how I handle bees in Nakuru. You can view it from the following address: www.apiconsult.com/video.htm.

Warning: Some people get an allergic reaction to bee stings. The normal response in most people is localised swelling and pain. Bee sting allergy symptoms often begin with a dry cough. The eye area begins to itch and swell. As symptoms progress, the patient begins to sneeze and wheeze and develops itchy skin irritation. These symptoms are warning signs of a dangerous condition called anaphylaxis. Anaphylactic symptoms include sudden anxiety and weakness, breathing difficulty, chest tightness, very low blood pressure, loss of consciousness, and shock. Anaphylactic shock can occur within minutes and result in death. A patient in shock needs immediate medical treatment.

People who are allergic to bee stings should avoid situations in which they are likely to get stung. Anaphylactic shock is treated with an injection of epinephrine (also called adrenaline), a hormone that stimulates the heart and relaxes the airways. You may combine it with an injection of antihistamine, which counteracts the histamine produced by immune cells during an allergic reaction. Those known to have severe bee sting allergies should carry a self-injection kit, including antihistamine tablets, for emergency treatment; however, they should still seek medical care after self-medication.

For most people, a few stings may actually be beneficial. Bee venom is used to treat arthritis and can be extracted from bees commercially. Contact ICIPE Nairobi for further information on bee venom. (See address in Annex 6.)

(N. B. *The health information on bee sting allergies is intended to supplement the care provided by your doctor. It is not intended or implied to be a substitute for professional medical advice.*)

WHAT IF NO BEES ENTER THE HIVE?

It often happens that bees do not enter the hive for quite a while. The new hive you have worked so hard to build stays empty. An empty hive does not produce any honey! A survey of beehives in Kakamega and Kirinyaga revealed that hive occupation varied between 40 to 70%. A recent survey in Kitui by DANIDA (2005) showed occupation rates of between 30 to 75%. At a farmers' research meeting at Baraka Agricultural College in 2006, three different groups of farmers working separately said low hive occupation rates was their biggest threat to beekeeping.

When this occurs, check the following:

- Have you waxed your top bars?
- Is the hive clean and pest-free?
- Are there any ants or rats nesting in it? If so, clean it out and re-grease wires. Bees do not like a dirty hive.
- Is it the swarming season? Ask a friendly beekeeper in your area. Swarming occurs when bees reproduce themselves and fly out of the old hive looking for a new home. During the swarming season, it is easier for hives to be occupied quickly. In Kerugoya town in 1995, bees occupied two hives at our extension office before we even had time to hang them.
- Have you chosen a good site for hanging the hive?

What will happen if it were not the swarming season? If it was but bees still did not enter the hive, what could you do? The beekeeper can put bees into the hive. There are different ways to do this:

A. CATCH A WILD SWARM OF BEES

For this operation, it is best to seek help from an experienced friend. The swarm has a better chance of staying in its new hive during a nectar flow. Don't waste time with swarms smaller than a person's head as they are more trouble than they are worth.

1. First, prepare your hive by smearing it with some melted beeswax, so it smells nice for bees.
2. Use your smoker and bee suit. Some swarms may be hungry and difficult to manage. Always be careful with strange bees!
3. Look for a swarm of bees clustering on a branch where you can catch them. Wear your bee suit and smoke bees very gently so as not to disturb them. (Don't smoke if you can avoid it.) Shake bees into a catcher box or similar container, e.g., cardboard box. If the queen falls into the box, the rest of the bees will follow. Wait 20 minutes or so. If bees return to the original site, try again.
4. Once you have the bees, leave the box in a shady place until evening. Make sure they do not become hot in the container, so covering with a damp cloth helps.
5. When evening comes, take home the bees. Shake them into your empty hive.
6. If you have other hives, give bees a comb with some uncapped honey and a brood comb with eggs from another hive to encourage them to stay.

(N. B. *At any stage in this operation, the bees may decide they don't like being disturbed and fly away. African bees frequently abscond.*)

Photo 25: Catcher boxes for trapping passing bee swarms
This location trapped 17 swarms in 1 season alone. (Nakuru District)

B. TRANSFER BEES INTO YOUR HIVE

You can transfer bees into a top bar hive from a wild nest or from a traditional hive with fixed combs, which are not moveable. To do this, cut out combs one by one and tie them to the top bars of your hive. Get help on this one as it is not easy for the beginner.

C. MAKING A DIVISION OF AN EXISTING COLONY

You can make a division of an existing colony, explained later under the section, "The Beekeeping Year".

(N. B. *In many countries, you can purchase a starter bee colony called a nucleus or package of bees: a queen and worker bees with or without wax combs on frames. In the USA, you start most new beehives by ordering a package of live bees, usually in a three-pound size [1.3kgs]. Bee breeders supply these packages. In Kenya, we currently don't have bee breeders or suppliers of nucleus hives so the beekeeper is on his/her own at the moment in getting hives occupied.*)

CAUSES OF ABSCONDING AND ITS CONTROL

Absconding happens when bees leave the hive completely. It is very frustrating for the beekeeper that has done so much to get the hive occupied.

BELOW ARE SOME CAUSES OF ABSCONDING AND ITS PREVENTION

- Lack of food and water for bees. Leave food for bees when harvesting. If necessary, feed them with sugar syrup. Provide a source of water in the apiary.
- Frequent attacks by pests. Check hives frequently. Grease hanging wires to control ants.
- Unfavourable weather conditions, e.g., cold and rain. Make sure the bees have enough food.
- Careless handling by beekeeper, e.g., breaking combs, over-smoking, etc. Be more careful in handling them.
- Genetic character of bees. Some bees are prone to absconding, so don't breed those with this trait.
- Excessive heat on the hive—no shade in very hot sun. Put hives in a well-shaded place when in hot areas.
- Extreme cold and damp. Don't put hives under dense trees in high altitude areas. Give them a sunny place with minimal shade.

Part Three
THE BEEKEEPING YEAR

CONDITIONS FOR MAXIMUM HONEY PRODUCTION

We have seen how to get started in beekeeping: how to make the hive, choose the apiary site, hang the hive, and get bees. Now let's look at the bigger picture: how bees change throughout the year. You will get to know about the bee colony cycle. It is important for the beekeeper to understand this, so he/she can manipulate the bee colony to produce larger amounts of honey.

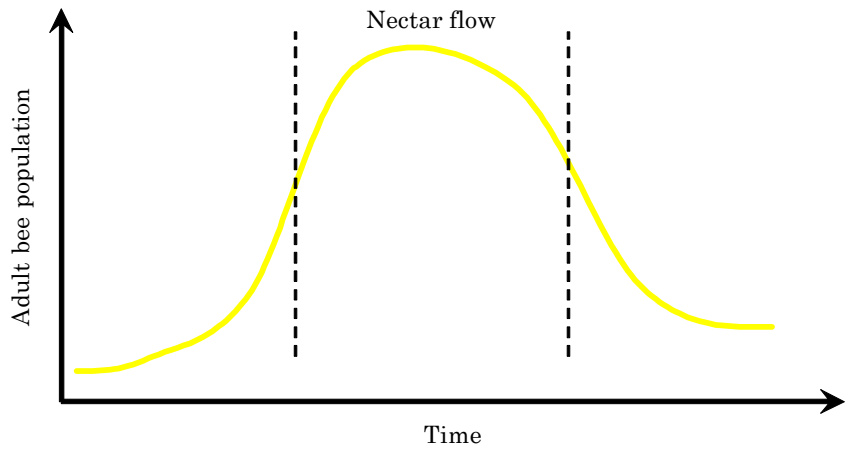

Figure 9: Conditions Required for Maximum Honey Yields

The chart shows that, if peak colony strength occurs at the time of nectar flow, a good honey crop results, providing nectar flow is good and weather conditions are right.

Your chief aim as a beekeeper is to get peak colony strength at the right time, i.e., during main nectar flow. Imagine this: peak strength of the colony comes when flowers are finished, which means bees will sit in the hive with little work to do. Manipulating the colony to be at peak strength at the right time is fundamental to good beekeeping. Good flying conditions, i.e., dry weather, for bees are also important during nectar flow to get maximum yields. This factor is, however, beyond your control as beekeeper.

Figure 9 also shows that colony size changes over time. When food is plentiful, there are more bees because workers feed the queen bee more food and she lays more eggs. When food is scarce, workers feed the queen less and the number of bees in the colony shrinks.

The bees' behaviour is very sensitive to their environment. When there is plenty of food, bees make more honey for eating when there is less food. The beekeeper shares in this stock of food. When harvesting, always remember to leave some for the bees. Don't be too greedy!

EXAMPLE ONE: KIRINYAGA

As with most parts of Kenya, Kirinyaga District has huge variation in climate. It stretches from the snow-capped peaks of Mt Kenya to the semi-arid Mwea Division. In between, there are tundra, bamboo, montane forest, tea, and coffee zones. Conditions for bees vary widely in all these areas. Things as climate and vegetation can change totally over a few kilometres. As with the changing environment, yield potentials and harvesting seasons for bees differ broadly. It is, therefore, very important for a beekeeper to be knowledgeable of the environment where he keeps bees because circumstances change dramatically over short distances.

In Kirinyaga, there are two harvesting seasons: March/April and September/ October. The yield of honey obtained in Kerugoya averaged 7.5kgs/hive per harvest or 15kgs/hive per year. The main nectar-bearing trees in the area are coffee (*Coffea spp*), bananas (*Musa spp*), *Grevillia robusta*, avocados (*Persea americana*), macadamia (*Macadamia tetraphylla*), mangoes (*Mangifera indica*), croton, pawpaw (*Carica papaya*), and beans (*Phaseolus vulgaris*). Flowering maize is an important source of pollen.

Hive occupation can vary depending on the time of year. Swarming season was observed to be in July. Bee swarms have been noted to migrate from Mwea Division to areas of higher altitude and vice versa, based on the annual period.

Mwea Division

Mr. Kariuki is new to Mwea Division and has bought three acres of land. He has never owned a beehive. At one of the beekeeping field days, Mr. Kariuki observed that most apiaries were sited under trees. He said he had no trees on his new farm but asked which tree species he should plant to act as bee forage. He was given the following advice:

Grevillia robusta: *mukima* (*Gikuyu*), silky oak (English). This is the most abundant tree in the district. It grows very fast and produces a lot of nectar. It flowers from March to April thus ensuring honey flow in May and June.

Photo 26: A tree in flower

It is also a source of fodder and timber as well as coffee shade, mulch and green manure. Wildings and seedlings propagate the tree.

Acacia mellifera: *muthigira* (*Gikuyu*), hook thorn (English). It is a low shrub propagated by direct sowing, seedlings, coppicing, and wilding. It flowers before the rains. The flowers produce quality honey. Other uses are as fodder (pods, twigs, leaves, and flowers), live fence, and fuel wood.

Others of the same genus are *A. lahai* (*Mugaa*), *A. seyal* (*Mugaa*), *A. abyssinica* (*Mugaa*), *A. brevispica* a.k.a. wait-a-bit-thorn (*Mwikunya*), and *A. Gerardo* (*Muthi*).

Eucalyptus saligna: *muringamu*, Sydney blue gum. The tree grows at an altitude of 1200 to 2400m except in arid areas and those infested with termites. Seedlings and direct sowing propagate it. It flowers in June and July. You should not plant the tree near crops but in a woodlot.

Other useful trees and shrubs are:

- *Croton megalocarpus* (*mukinduri*): flowers in March to April, useful for boundary marking, timber, and shade.
- *Azadirachta indica* (*mwarubaini*): flowers in March to April with seeds in June; useful as medicine, insecticide, shade, fertiliser, and de-wormer.
- *Calliandra calothyrsus* (*calliandra*): used for shade, windbreaks, and fodder. The shrub flowers in February to March.
- *Callistemon citrinus* (bottlebrush): very fast growing; both useful and ornamental.
- *Cajanus cajan* (njugu): drought-resistant food crop with other uses as nitrogen fixation agent and fodder (leaves and pods).
- *Kigelia africana* (*muratina* or sausage tree): flowers in January to February. The unripe fruit is poisonous. It does well in arid areas. The Kikuyu do not plant it near homesteads as they used it to mark gravesites. It is a slow-growing tree, a non-prolific seeder with poor germination rate.

Cultivated bee-forage trees include:

- Papaw (*Carica papaya*)
- Bananas (*Musa sp*)
- Beans (*Phaseolus sp*)
- *Muembe* (*Mangifera indica*)
- *Mubera* (*Psidium guajava*)
- *Mukandamia* (*Macadamia tetraphylla*)

EXAMPLE TWO: KAKAMEGA

The following is a brief description of beekeeping practices, potential yields, flowering trees, and seasonality of honey production around Kakamega.

There are three different types of hives used in the area, namely:

- Traditional log hives
- Kenya top bar hives (KTBH)
- Basket hives

Log hives are made from trees species such as *munga*, *likoro*, *munuku*, and *mukomari* (Luhya terms).

KTBH are reputed to be very productive if managed well, but they may get too hot in the dry season and too cold in the wet season, so you need to insulate them. Using grass for heat retention in the wet season works well.

Basket hives are least favoured by beekeepers.

Factors	Basket	KTBH	Log
Quantity Harvested	3	2	1
Ease of Harvest	3	1	2
Heat Properties	2	3	1
Least Expensive	1	3	2
Rain Resistance	3	1	2
Durability	3	2	1
Total Score	**15**	**12**	**9**

Table 2: Beekeepers' Preferences for Different Hive Types
(The ranking was as follows: 1 = best, 2 = second best, 3 = last.)

(N. B. *Table 2 shows a good lesson for those of us who promote improved hive designs. The log hive comes out the best with the KTBH second and the basket hive last. If the KTBH is not managed to maximise its advantages [e.g., moveable combs], it can be outperformed by the log hive. Advantages of the KTBH are many, e.g., ability to manipulate bees for increased honey production and ease of harvesting that allows collection of comb honey, which obtains a significantly higher price in the market. But if you don't manage a KTBH or Langstroth and it produces crude honey, it is actually worse than a log hive because of its higher purchase cost.*)

The swarming season in Kakamega is reported to be in November.

Hives are normally harvested only once a year. Bees produce honey from November to April, but production is highest between December and March. (See Figure 10.) You can get 30 to 40kg/hive of crude honey harvesting in February/March. Harvesting can take place more than once during this period. As the weather becomes cooler when rains start in March/April, bees become net consumers of honey. If you don't harvest before the rains, yields will go down.

Honey production is reliant on the flowering of forest trees and other plants, including *Leucaena leucocephala*, bananas, *isungusa* (Luhya), *isirimoi* (Luhya), and *iludolio* (Luhya). The flowering of *Croton megalocarpus* (*musine*, Luhya) is an indicator of when to harvest.

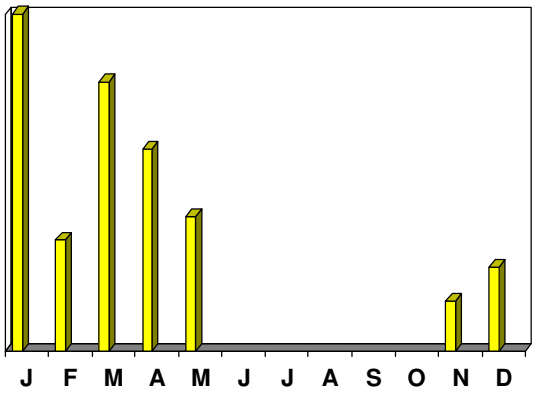

Figure 10: Seasonality of Honey Production in Kakamega

(Reference for Kakamega section: *Beekeeping in Kakamega Core Target Area*, ABLH Report No.15 of 1996.)

EXAMPLE THREE: MOLO

Molo is an example of beekeeping in a highland area where potential is lower than other areas due to high altitude. With good management of bees combined with good weather conditions, however, yields of up to 20kg/hive are achievable—very good yield for this area. In poor seasons and under poor management, 3kg/hive yields have been obtained.

Beekeeping in the area has lagged behind for many years. The following are major problems:

- There are few flowering plants and shrubs to produce nectar.
- Vast acreages have been turned into large-scale farming, thus reducing natural flowering vegetation.
- There is lack of knowledge on the management of beehives.
- Cold climate in the area makes bees abscond the hives.
- Farmers are reluctant to get into beekeeping probably because of the above factors.

Plant Name	Flowering Time
Dombeya goetzentii	September/October
Bottle brush (*Callistemon citrinus*)	June/July/August
Sagebrush	August/September/October
Dahlia	February to April
Maize	August to September (pollen only)
Eucalyptus saligna	February to April
Vernonia	July/August/September
Pawpaw	August/September
Wild rape	July/August
Croton megalocarpus	March/April
Fuchsia (ladies' eardrop)	year-round except April/May and June
Daisy	February/March and August/September

Table 3: Plants Suitable for Bees in Molo

Please note that the above table is only a guide. Depending on the season, plants may flower late or not at all. The honey harvest season in Molo is towards end of September/October from the *dombeya* tree, which flowers around this time. Beekeeping at Baraka Agricultural College in Molo has improved since I came in 1994. We planted extensively a fuchsia hedge known as ladies' eardrop around the college grounds. Over the years, the problem of absconding bees has dropped as bees now have forage all year round. Honey yields from Baraka hives have also increased. This proves that planting bee forage has a very positive impact on beekeeping. In particular, if communities can come together and plant collectively, the impact can be much greater.

EXAMPLE 4: TRANSMARA

This is a good beekeeping area. There are many trees and shrubs, which have not been cultivated, that grow in the district. Natural vegetation remains relatively undisturbed. Farmers are organised into groups and they own most of the hives, not individuals. An umbrella body called Transmara Association of Beekeepers coordinates them.

Beekeepers usually harvest three times a year: March, July, and November. From each harvest, about 10kgs/hive of honey is expected. The main honey flow is in November. Apart from these times, you may also harvest honey in other months.

Plant Name	Flowering Time
Oskett	May to September
Ogilai	May to September
Olmolongina	May to September
Oltorioi	June to August
Orkikeorgos	June to August
Olkinyei	November to February
Okilenyai	November to February
Olkel	November to February
Olchartuyian	June to August/November to February
Olmesuli	June to August/November to February
Olea africana (wild olive)	June to August/November to February
Thurnbegia atela	June to August/November to February
Scutia myrtina	June to August/November to February
Cardia moncica	June to August/November to February
Acacia seiberiana	June to August/November to February

Table 4: Plants Suitable for Bees in Transmara
Maasai names may not be spelled correctly.

Thanks to Baraka Agricultural College for the Molo and Transmara sections.

EXAMPLE 5: NANDI HILLS

Some of the major flowering/honey-producing trees and plants in Nandi Hills are:

- Avocado: March/April
- *Dombeya*: April, September to December
- *Grevillia robusta*
- Banana: all year round
- Pawpaw: all year round
- Beans: April/May and September/October
- Weeds: always
- Coffee
- Sydney blue gum
- *Tebesuet* (Nandi)
- *Croton macrostachyus*: November to February

Honey Harvesting

April to June is the best season. It is also possible to harvest honey from October to February.

A Sample Harvest Season
(*J. Muemah,* beekeeper)

"In April, I harvested 33kgs of crude honey from one hive. After two months, in June, I harvested a further 15kgs of crude honey from the same hive, which gave a total of 48kgs of crude honey. After refining the honeycombs, I got 24kgs refined honey and 1.5kgs of beeswax. From the 24kgs of honey, I filled 48 500g jars for sale. I sold each of the 48 jars at KSH120, which gave me revenue of KSH5,760 (US$80) from just one hive. This yield per hive is, however, not always obtained.

"Another time, I harvested 16 colonies and only got 67kgs of crude honey. Eight of the colonies were strong and eight were weak. This is why it is very important to manage bees properly. With proper knowledge and management of all colonies, it is possible to have them give maximum yield".

(N. B. *The above example is an interesting one. Eight colonies in the apiary were strong and eight were weak. This situation is common. In many apiaries, only a few colonies out of many give significant yields. Beekeepers should aim to ensure that all colonies are strong and producing honey if the enterprise is to maximise profits. Get rid of weak and unproductive colonies and divide productive colonies to replace them. Combine this selection of the best bees with good bee management.*)

For more information on beekeeping in Nakuru District, see www.apiconsult.com, our website where you can find regular updates on what is happening in my apiary through past editions of the Apiconsult e-newsletter. Click on the news and articles section.

BEEKEEPING MANAGEMENT DURING THE YEAR

From the above examples, we realise that conditions for bees vary widely throughout the country. You can also see that management of bees depends on where you find them. It is very important that the beekeeper gets to know the area in which he/she lives. What plants and trees do bees use? When do they flower? When are the swarming seasons? Which trees/plants give the best honey and which give bitter honey? When are the correct times of year to expect honey? Which are signs of harvesting? What factors such as rainfall and temperature affect plant flowering and nectar secretion?

Collect and write down information. It is very important in helping you become a good beekeeper. Be observant always of your bees and the environment in which you live. Such knowledge will allow you to make timely and sound decisions on bee management. In essence, all beekeepers in Kenya must become their own researchers! Research (systematic investigation to establish facts) should not only be seen as the preserve of academics but also that of ordinary people,

including beekeepers. See Annex 5 for more information on Participatory Action Research.

Most beekeepers in Kenya do not manage their bees at all. Every time a beekeeper is seen to go near a hive with a bee suit, he/she is said to be going for harvest. "*Unakwenda kuvuna?*" (Are you going to harvest?)

In this book are highlighted simple and achievable management practices. If you have invested money in more expensive equipment, e.g., Langstroth, then you also need to manage bees better to get good returns on your investment.

Traditional hives with a fixed comb make management very difficult. A top bar hive makes management far easier as does the frame hive Langstroth. Seek further advice on management of bees and read as much as you can on the subject. Outlined below are some simple practices that you can carry out to help you improve your honey yield.

SWARM PREVENTION AND CONTROL

Swarming is what happens when bees divide themselves to reproduce. If swarming occurs when bees are building up in numbers or during nectar flow (when there is an abundance of flowers around), the beekeeper will not get much honey because the bee population in the hive will have reduced dramatically when they are required to collect nectar to make honey. (Look again at Figure 9). The beekeeper should try, therefore, to keep his/her bees from swarming. It is impossible to prevent all swarming, but with proper management of bees, it can be minimised and controlled.

You can take some simple steps to help stop swarming:

If possible examine the bees every 7 to 10 days during the honey period, when flowering season is about to start, until after harvest. Make sure that the queen has enough room to lay eggs by making extra space around the brood nest. If the queen runs out of cells to lay eggs in, the bees will want to swarm. Make extra space by exchanging empty combs with those filled with honey around the brood nest. This is important as African bees build up quickly and can easily become crowded in the hive and want to swarm. Give the queen plenty of space to lay eggs and space for bees to store honey.

If the bees are building queen cells (long thumb-shaped cells protruding from the edge of combs) to make a new queen then you know they are going to swarm. To prevent swarming you can do one of the following:

- Destroy all queen cells in the colony wanting to swarm and switch hive locations with a weaker colony. Foraging bees of the strong colony will return to the original site of the hive and strengthen the weaker colony. As the weaker colony becomes stronger, the strong colony becomes weaker. This may control the swarming urge of the strong colony.

- Destroy all queen cells and give brood combs without bees to the weak colony. This may get rid of the swarming urge in the strong colony.
- Artificially swarm the bees. (See "Making a Division" below.)

MAKING A DIVISION

Break down all queen cells except one.

Transfer comb with the queen cell, one other comb of brood, and two combs of food (honey and pollen) into a catcher box/new hive. Include bees on all combs. Shake in bees from other combs as well. Remember to put brood combs in the middle and honeycombs on either side to insulate the brood nest.

If possible move the catcher box/new hive to a new site at least two kilometres from the old site. (It is not always practical to move the new colony two kilometres away. The division can still work if you move it a short distance, but you can expect adult foragers to return to the old site and weaken the new colony. Give extra bees to compensate for those that will return to the old site.) These bees will become a new colony. The rest of the bees will continue working and a new queen will hatch out in the new colony.

Most of the adult bees will remain in the old hive and continue making honey; however, avoid making divisions during honey season because it reduces honey production. Make divisions after honey flow to increase colony numbers. You can make a division without queen cells as long as the new colony has eggs in the combs transferred. They will be able then to make a queen cell and raise a new queen. Bees can make a queen from worker eggs or worker larvae less than three days old (i.e., very small larvae).

FEEDING

Feeding bees is not recommended in Kenya unless you are sure as to why you are feeding them and the benefits you are likely to get. You waste money to buy sugar for feeding bees when you need it for yourself and your family. People often feed when it is unnecessary and they get no return for their money. Bees can also abscond and your money is wasted.

If want to feed bees, however, the best time to do so is:

- Before nectar flow (stimulatory feeding) when the queen bee will lay eggs and the colony will build up in numbers before honey flow. Start six to eight weeks before the date of flowering because it takes about 40 days from egg to adult foraging bee.
- In times of food shortage, e.g., drought or an excessively wet and cold period.

Feeding at such times may prevent bee migration and keep bees alive.

To feed bees, dissolve white sugar in water by heating gently in a *sufuria* or pot, stirring all the time.

To stimulate bees: use 1:1 solution by volume of sugar and water.
To feed bees to store: use 2:1 solution by volume of sugar and water.

Place the sugar syrup in a feeder box in the hive when it cools. Place some twigs/ grass on the surface of the floating syrup to prevent bees from drowning.

Warning: Never spill sugar syrup or honey around the apiary as it can start robbing: bees attacking each other to raid stores of honey. Weak colonies are prone to robbing by strong ones.

PESTS AND DISEASES AND THEIR CONTROL

For the most part, beekeepers in Kenya need only worry about the following major pests:

- Wax moths
- Safari and other ants
- Honey badgers
- Human beings

(N. B. *Other pests such as birds and wasps can be problematic in some areas of Kenya.*)

Wax Moth

- Destroys wax comb and damages hive parts.
- Control by having a strong colony.
- Remove old combs that bees are unable to cover in time of food scarcity when colony size shrinks.
- Avoid holes and cracks in top bars and hive body where wax moths can lay eggs.

The wax moth is a very difficult and destructive pest to control particularly with frame hives when supers have to be stored after honey extraction.

Photo 27: Hive infested with wax moths
Wax moths infested this hive after bees absconded during the dry season.

Safari and Other Ants

- Hang hives and grease hanging wires regularly.
- Spread ashes around posts holding hives.
- Keep grass short and branches from touching hives.
- Make hive parts fit together without gaps.
- If hives are on stands, place legs of stands in tins of old engine oil.

Termites

- Treat posts used for hanging.

Hive Beetles

- Use holes instead of slits for the entrance.
- Keep strong colonies.
- Hand pick or destroy them if found in hive. African hive beetles, particularly the small hive beetle (*Aethina tumida*), are not serious pests in Kenya but are now causing problems for beekeepers in the USA, Europe, and other parts of the world where they have been introduced accidentally.

Birds (Honey Guides)

- Don't leave brood combs exposed. Scare birds away.

Honey Badger

- Hang hives securely to prevent the badger from knocking them down. Hives placed on stands are prone to attack by the honey badger.
- In Transmara, log hives are suspended two metres from the ground by means of a twisted bark rope to prevent damage from the honey badger.
- In Kitui, beekeepers put *mabati* around tree trunks bearing hives to prevent the badger from climbing them.

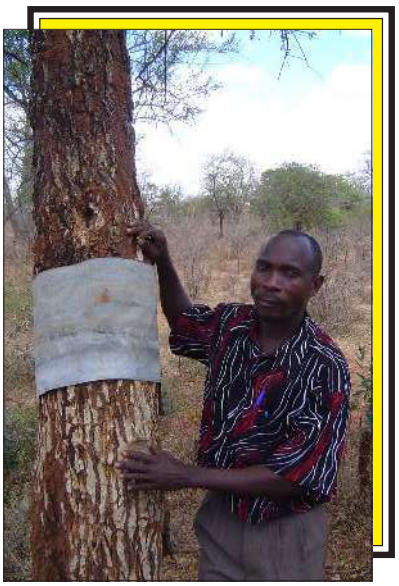

Photo 28: Tree encircled by *mabati*

The *mabati* is to prevent honey badgers from climbing and attacking bee colonies. When the badger encounters the tin, he cannot grip and falls down!

Pirate Wasp

- Attacks and eats bees.
- Place a dish of water below hive entrance, which has a mirror in the bottom. Wasps fly in and drown.

Bee Louse

- Negligible damage.

Man

In parts of Kenya, people cause a lot of damage to hives. They often steal hives and honey, making it difficult to keep hives in areas where security is poor. Children often antagonise bees by throwing stones at beehives, which often results in bees stinging people and livestock. Always keep your hives where you can supervise them. The problem is worse with modern hives kept close to the ground.

Note on Varroa Mites (*Varroa destructor*)

Varroa is a parasitic mite, which causes bee colonies to die out. This pest causes havoc to bee populations in many parts of the world, including Africa. Varroa mites are now in many southern African countries and may spread to Kenya in the future. Kenya currently is varroa-free and we should try to keep it that way. Bee importations into Kenya are prohibited and should never happen to prevent the accidental introduction of this destructive mite or other bee pests and diseases into our country.

MAJOR NECTAR-BEARING PLANTS IN KENYA

I have written about many trees, shrubs, and crops useful for bees in Kenya elsewhere in this guide, but below is a consolidation of the major ones. Planting multipurpose bee forage trees is a good idea for farmers as you can benefit from bee forage plants in other ways such as fencing posts and firewood. Suitability of tree species for planting varies widely in Kenya due to variation in climate. To find out which are the best species to plant, seek advice from your local Beekeeping Officer or Forest Officer.

Major nectar-bearing plants in Kenya: acacia spp, banana, citrus, dombeya, eucalyptus (gum), croton spp, jacaranda, kei-apple (*Albera caffra*), sunflower, vetch, rape, and coffee.

It helps if you have what people call a flowering calendar for the area. Make one by noting the major nectar-bearing trees in your area and the times they flower. Record this information carefully. It is then easy to predict the flowering of trees and the dates. (They may vary from year to year.)

Photo 29: Citrus fruit in bloom

Photo 30: Banana in flower
It is a good source of nectar in banana-growing areas.

Photo 31: Acacia *senegal* in bloom
Acacia species are very important nectar sources in arid areas of the country.

Photo 32: Acacia *persiciflora* in flower
This is in Nakuru.

Photo 33: Kei-apple
It makes a great fence and is a good forage plant for bees.

Photo 34: *Cordia africana.*
It is very attractive to bees.

Part Four THE HONEY HARVEST

TIME OF HARVESTING

Harvest honey during dry spells, i.e., January, February, March, July, August, September, November, and December. The harvest time in each area differs, so check the right time in your area. In areas where there are dominant bee plants like coffee, sunflower, etc., you should harvest after the flowers wither.

Regular inspection of hives during nectar flow will ensure that the beekeeper harvests as soon as honey is ready. Do not harvest unripe honey. (See below.)

Ideal harvesting time of the day is from 5.30pm to 7.30pm.

HARVESTING PROCEDURE

- Ensure you are sting-proof by putting on protective clothing.
- You will need a smoker in good working condition and a plastic bucket with lid—must be clean and dry—for storing honey.
- Smoke the entrance of the hive with about eight to 10 puffs and then gently lift the lid and smoke again. Leave hive for a minute or two before opening lid to allow smoke to affect bees. Smoke causes bees to engorge themselves with honey making it difficult for them to bend and sting. (They become too full!)
- Very gently tap top bars with a hive tool. A hollow sound will indicate where there is no comb.
- Remove top bar from the hive that has no comb attached, so you can examine the rest of bars in the hive. Honeycombs are usually at the end of the hive opposite the entrance. Select combs that are ¾ or more sealed or capped full of honey. (These combs are said to be ripe or have a low moisture content, <19%, which ensures that honey will not ferment later when bottled. Leave combs with brood and pollen for future production of honey.)

Photo 35: Harvesting white combs
Harvest white combs that are at least ¾ capped, as shown above, for the best honey.

When harvesting a comb:

- Brush bees gently from the comb using a bee brush. You can cut the harvested comb from the top bar to fall into the bucket.
- Replace lid of bucket to prevent bees from entering with the honey.
- Return top bar, minus comb, to the hive.
- As an alternative, place the whole comb and top bar (after brushing bees off) in another empty hive or catcher box where you can take it away later for comb honey. Fix spare top bar in place of the one removed.

Gentle smoking is a continuous process during harvest time to control bees. It is important to avoid smoking the honey directly or excessively because it can damage honey flavour. After harvesting, replace the first bar and cover hive with the lid. Make a final smoke before you leave to keep bees away from the harvester and to prevent them from following him/her all the way home. Remember to move through a bushy area first to get rid of bees.

Photo 36: Harvesting traditional log hive in Tabora, Tanzania
The hive is lowered to the ground for harvest and the two halves of the log separated.

(N. B. *Two people are better than one when harvesting or carrying out any bee operations. Two smokers in operation are also better than one to ensure a continuous supply of smoke for subduing bees. Sufficient smoke is very important when working aggressive bees.*)

Photo 37: Harvested honey
Honey harvested from the log hive is placed in clean bucket to ensure its high quality.

HONEY REFINING

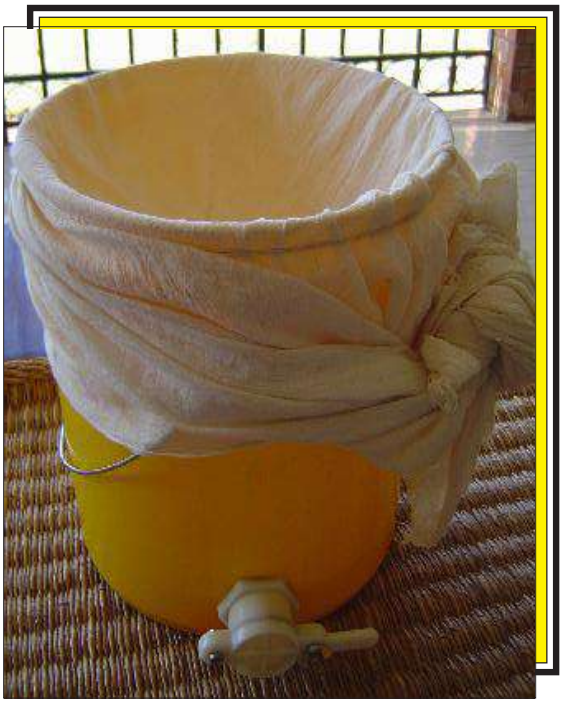

Photo 38: Simple honey refining method for small-scale beekeeper

STEPS

1. Remove wax capping from combs using a knife to cut off the capping.
2. Break combs into smaller particles and sieve them through a net or nylon fabric into a plastic container (See Photo 36.) The sieving process can take a few days. Cover sieved honey with a lid and keep in a dry room away from bees.
3. You can place the bucket in sunshine for two or three hours to heat honey gently so it flows freely.

A word of warning: **NEVER** boil honey as this destroys its flavour and medicinal characteristics!

Once honey has drained through the cloth and settled at the bottom of the bucket (usually after 2 days or so), use a jug to pour it into honey jars for sale. For wax remaining behind on the straining cloth, squeeze out any remaining honey and process leftover wax. (See below.)

Photo 39: Centrifuge
Centrifuge that handles honey from the KTBH (in wire baskets) as well as frames from the Langstroth hive.

Photo 40: Uncapping frame full of honey
It is done with an uncapping knife in preparation for extraction with a centrifugal extractor.

EXTRACTION OF BEESWAX

Beeswax is a product of the beehive. (For every 100kgs of honeycomb, you can get about eight to 10kgs of beeswax.) Honeybees secrete beeswax in the form of thin scales. 12- to 17-day-old worker bees produce them through glands on their ventral (stomach) surface. Honeybees use beeswax to build honeycomb cells in which they raise their young and store honey and pollen.

To produce wax, bees must consume about eight times as much honey by mass. Estimates are that bees fly 150,000 miles to yield one pound of beeswax or 530,000 km/kg.

PROCEDURE TO EXTRACT WAX FROM COMBS

1. Mix combs and water in a *sufuria* (aluminium pot) and heat. Wax melts at about 62 to 64° Celsius, so there is no need to boil. Boiling damages the wax and can be dangerous. Overheated wax can burst into flames. Do not use iron, brass, zinc, or copper containers for heating wax as it can discolour the finished product.
2. Pour melted combs and water into an extraction bag. You can use cotton for sieving. (You can also use the small bags maize seeds come in after you clean thoroughly.)
3. Smear sides of a second *sufuria* with soapy water to prevent wax from sticking to its sides.
4. Filter wax into the second *sufuria*. Use two sticks (such as two top bars) to squeeze the bag containing melted combs to extract wax. The yellow wax will come out along with water; waste will remain in the filter bag. If the combs contained bee brood, you can feed these to poultry as they would be cooked by then.
5. After filtering, wax separates from water and floats to the top.
6. Remove wax after leaving it to cool in the *sufuria*, with lid on to keep away dust, for 12 hours.
7. Scrape dirt from the bottom of wax cake when cooled.
8. Store wax blocks in a cool dry place. Never store near pesticides/chemicals as it may absorb them.

Your wax block is now ready for sale or for further use. Wax currently sells in Kenya at a price of about KSH100 to 150 per kilogramme or more depending on the demand.

Photo 41: Solar wax melter used to extract beeswax
It is most efficient for use with new combs/cappings instead of brood comb.

Part Five
THE MARKET FOR BEE PRODUCTS

Photo 42: A sampling of honey products.
Most people enjoy honey!

HONEY

There is currently a very high demand for honey both in Kenya and overseas. Kenya is now licensed to export honey to the European Union. Market opportunity exists for beekeepers but lack of production is at present the main constraint in Kenyan beekeeping. Good quality honey can fetch a high price. Any farmer who has the good fortune and skill to obtain honey should have no trouble selling it. In general, if you present honey to the consumer in a clean, unspoiled condition, the price will be higher. Beekeepers should aim for the highest grade of honey to maximise returns from beekeeping. Honey in Kenya is sold in the following grades:

CRUDE HONEY

This is a mixture of ripe and unripe honey. At harvesting time, the wax, honeycomb, and bee-and-brood comb are all mixed into one container. This container is often an old tin. Crude honey is used mainly for brewing local beer because quality requirements are not very strict. The demand for this type of honey is high.

SEMI-REFINED HONEY

Semi-refined honey is generally the liquid honey that remains when you skim wax off the top of crude honey. Honey sinks to the bottom as it is heavier. Semi-refined honey still contains particles of wax and other debris such as bees' legs. It can be stored for the beekeeper's own use or it can be refined further and packed for sale. It gains a higher price than crude honey.

REFINED HONEY

Refined honey is clean. You strain it to remove all particles of beeswax and other materials. Remember: refined honey is unchanged, it is only strained. Nothing else is added so it is still the pure honey that bees made in the hive.

CHUNK HONEY

Whole combs of capped honey can be harvested carefully from the beehive. You can cut up pieces of the comb and put them into jars of liquid honey. This gives the consumer a feeling that the honey is real and not adulterated with sugar. Chunk honey can fetch a higher price than refined honey.

COMB HONEY

Honeycombs of capped honey that have a nice white capping can be cut up, placed on small trays, and covered with cling film. These are very marketable in Kenya and command a very high price in the market, particularly in affluent Nairobi suburbs and other towns. This product should be the ultimate aim of all beekeepers with access to these markets. This product is priced per gram.

PACKAGING AND LABELLING

Package honey in either plastic or glass jars, which should be clean and dry. 454g jars are available in Kenya from Nairobi suppliers as are plastic trays for selling comb honey. (See Annex 6 for contacts of suppliers.)

Labelling of honey must include:

- net weight of honey,
- name and address of producer,
- country of origin (Kenya), and
- description of contents, e.g., "pure honey"

Initially you can make a simple label by hand or on a computer, which you can then photocopy. You can have a more attractive label printed later, but this is costly and requires many jars to spread the cost.

Some people use Kasuku/Kimbo plastic containers or tree top bottles (700ml) to sell honey. Metal paint tins (*mikebe ya rangi*) are not good. Honey is acidic and can eat the metal and spoil the honey.

WHERE TO SELL HONEY

You can sell your bottled and labelled honey directly to shops. Build up a name for supplying the right quality, in the right quantity, at the right price, on time. Many beekeepers develop a reputation for beekeeping and people flock to their homes to buy honey directly. If you are enterprising, you can even set up a shop of your own selling bee products. You could also buy and sell honey from other beekeepers.

A shop or kiosk located in the right place such as the market or along a busy route can attract many customers.

You can also consider the option of forming a beekeeping association to market products of farmers in your area. Examples are the Kakamega Forest Beekeepers Association and the Transmara Association of Beekeepers. You can also consider the option of starting a cooperative. An example of a successful beekeeping cooperative is Ruai Beekeeping Cooperative Society in Naru Moru, which markets eight tonnes of honey per annum collectively. (See Annex 6.)

The advantage of farmers marketing together is fixed costs, e.g., transport, can be spread over a larger quantity of honey reducing cost. Collective selling gives farmers access to higher priced markets such as Nairobi.

Whatever type of honey you produce, you can easily find a market for it in Kenya and the price you receive will compare favourably with that available elsewhere in the world.

BEESWAX

Most people in Kenya throw away wax combs upon harvesting or after honey extraction. Beekeepers do not know its value. Local villages use beeswax in very limited ways, e.g., for shoe repairs by cobblers. Some companies such as Bata (shoe company) and Kiwi (shoe polish company) purchase beeswax, which they often obtain from *miti ni dawa* (honey beer) brewers. You leave the wax after brewing beer from crude honey, which contains honeycombs.

Baraka Agricultural College buys clean beeswax cakes from other buyers. (See Annex 6 for college contacts.) You can also use your beeswax to make some of the recipes in Annex 4.

PROPOLIS

Export markets for propolis exist. People use it as medicine, selling it as capsule, ointment, or tincture (dissolved). You can chew propolis raw as medicine for the throat. It is on sale in this form in very limited quantities in Nairobi. When harvesting, simply scrape off propolis and store in an airtight container. You can also try making medicine from propolis, e.g., propolis ointment. Try the recipe in Annex 4.

Part Six
STARTING A BEEKEEPING ENTERPRISE: COSTS & PROFITS

I have based these figures on 20 KTBH hives per farmer. On the other hand, I have calculated equipment costs from the Baraka Agricultural College Workshop bee equipment price list. All figures are valid for 2007.

The farmer can add significant value if he/she sells honey in jars or bottles. This is realistic in many parts of Kenya where there is a very strong local demand for honey with high prices. In other parts of Kenya where local production is high, bulk sale of semi-refined honey at a lower price to middlemen is the norm. In remote areas of the country, honey prices tend to be low but hive numbers are much higher, particularly in semi-arid areas where individuals can own hundreds of traditional fixed comb hives. An average production figure of 20kgs per occupied hive is achievable for most farmers; however, this can increase to 30kgs or more depending on the area and the management of bees, especially if you control swarming and maximise honey production. An occupation rate of 80% means 16 hives out of 20 are in production, which is a good occupation rate by Kenyan standards. That said, there is no reason why the farmer cannot achieve 100% occupation rates with some effort.

In this example, we take an average farmer with 20 KTBH hives, with an 80% occupation rate, producing an average of 20kgs per occupied hive. The farmer harvests and sells honey in bulk at a price of KSH100/kg with no value addition. This example is not the best or worst but an average situation.

Expenditure	Cost (KSH)
20 Kenya top bar hives @ 2,000/hive	40,000
Bee suit @ 3,270	3,270
Smoker @ 800	800
25 Hanging posts @ 150/post	3,750
Grease and miscellaneous items	500
Total	**48,320**

Table 5: Costs of Starting a Beekeeping Business

Estimated honey production from one colony = 20kgs
Total honey production from 16 occupied hives = 320kgs
Estimated price of 1kg honey in 2007 = KSH100/kg
Estimated income from 16 occupied hives = KSH32,000
Net income year one = KSH32,000 – 48,320 = (16,320)

A negative of KSH16,320.

Equipment will have a lifespan of up to ten years or more depending on how one cares for them, so in subsequent years, capital costs will be paid off and net income will be much higher. The question is how do poor farmers finance capital costs to get started? The important thing in beekeeping is that a farmer can start by using simple hives made from local materials. It is not necessary or advisable for you to take a loan to get started unless you are absolutely sure you can make the hives produce. This certainty comes from experience. Start slowly. Learn as you go along. Once you understand what you are doing, expand. Many farmers in Kenya who purchased expensive frame hives on loan had ended up repaying through other means as the hives failed to produce anything—very demoralising. In particular, with the recent promotion of the Langstroth hive, people are misled by the belief that it will automatically produce honey and lots of it. **Be warned: this is not the case!** It will only yield well with good management.

GROSS MARGIN ANALYSIS

The following is an analysis of beekeeping versus a major source of livelihood in Kenya: maize production. The table below looks at the same beekeeping enterprise above minus capital costs. It also looks at the enterprise at different production levels: low, medium, and high. Low is 10kgs per occupied hive/annum. Medium is 20kgs per occupied hive/annum. High is 30kgs per occupied hive/annum. The sale price of honey is estimated at KSH100/kg. Through packaging, one can easily add value and increase the price of honey. Many farmers in Kenya can sell honey in raw form for KSH150/kg or more to their neighbours, e.g., beekeepers in the Lare area of Nakuru District.

Description	Production Level		
Production	Low (10kgs)	Medium(20kgs)	High(30kgs)
Yield per hive x 16	160	320	480
Price per kg honey	100	100	100
Gross Output (a x b)	16,000	32,000	48,000
Variable Costs			
Grease and miscellaneous	500	500	500
Total Variable Costs	**500**	**500**	**500**
Gross margin/16 occupied hives	15,500	31,500	47,500

Table 6: Gross Margin Analysis of Beekeeping

Description	Production Level		
Production	Low	Medium	High
Yield of maize (90kg-bag)	20	50	87
Price of maize	1,300	1,300	1,300
Gross Output (a x b)	26,000	65,000	113,100
Variable Costs			
Total variable costs	18,000	32,500	60,714
Gross margin/hectare	8,000	32,500	52,386

Table 7: Gross Margin for Hectare of Maize

(Source: Baraka Agricultural College)

From the above comparison, a beekeeping enterprise of 20 KTBH compares very favourably with returns from a hectare of maize, particularly at low levels of efficiency in maize production. In addition, a beekeeper has the option of adding value to his/her honey to increase returns while a maize farmer is more at the mercy of market forces where gluts in production frequently depress prices. The bee enterprise takes up very little space in comparison to maize and the time input is much less. The most important but often overlooked inputs for successful beekeeping are skills and knowledge, so develop them very well.

GLOSSARY OF TERMS

Abdomen - third posterior section of a bee's body.

Absconding - when a colony of bees abandons the hive or nest site due to infestation of pests or lack of food. (It happens with African bees).

Allergic - in beekeeping, refers to the severe bodily reaction to bee venom following a sting.

Anther - the part of a flower's stamen that produces pollen. The male part of the flower.

Apiary - bee hives grouped together.

Bee space - gap large enough for bees to walk and work, e.g., the space between two parallel combs or between a comb and the wall of the hive.

Beeswax - material produced by honeybees for building combs.

Brood nest - the centre area of the combs of a colony where brood is reared.

Capped honey - ripe honey in the comb, sealed with a wax capping.

Cappings - thin wax covers of cells full of honey, particularly after they have been removed from the surface of the comb prior to honey extraction.

Caste - different types of mature honeybees (and other social insects).

Comb - beeswax structure of two plates of cells arranged back to back on a midrib and used to rear brood or store honey and pollen.

Drone - male honeybee. Drones undertake no work within the hive. Their sole function is to fertilise the queen.

Feeder box - box used to feed bees artificially with sugar syrup.

Foundation - thin sheet of beeswax usually embossed with a hexagonal pattern of cells placed in each wooden frame that serves as a base upon which honeybees build their comb.

Frame hive - hive containing frames. The honeybees are encouraged to build their combs within these frames. The frames enable combs to be lifted from the hive for examination.

Gland - an organ that produces and releases one or more substances.

Hive - artificial structure for housing a colony of bees. It is also a verb to denote the process of introducing a swarm into a hive to house it permanently.

Langstroth hive - a design of frame hive. The inventor, Rev L. Langstroth, recognised the importance of bee space and this allowed him to design the movable frame hive.

Larva - an insect in the first stage of development after hatching from the egg. (plural: larvae)

Migration - seasonal movements of whole honeybee colonies, leaving no brood behind in the nest.

Nectar - Sweet liquid secreted by flowers. Watery solution of various sugars.

Participatory Action Research - PAR has emerged in recent years as a significant methodology for intervention, development, and change within communities and groups. It is now promoted and implemented by many international development agencies and university programmes as well as countless local community organisations around the world. PAR involves people working together to solve their problems through learning in a systematic way. As the name implies, the output of the research is action.

Pollen - fine dust-like substances, which are the male reproductive cells of flowering plants collected by bees as a source of protein.

Propolis - an aromatic, resinous, antibiotic plant material collected by bees for blocking holes, waterproofing, and generally protecting the hive from intruders and diseases.

Pupa - final stage of development of the bee in which it is sealed into its cell and from which it will emerge it as an adult. (plural: pupae)

Queen - female parent of the colony. The only sexually developed female.

Race - biological definition of race is a categorisation of organisms with differing characteristics while maintaining enough similarities to be a part of a common genus and species. In this context, it is synonymous with subspecies.

Ripe honey - capped honey in the comb, i.e., bees have added enzymes and dehydrated it to an acceptable moisture level below 19% before sealing.

Species - in biology, a species is a group of related organisms that, more or less, share a distinctive form and are capable of interbreeding.

Stamen - male reproductive organ of a flower. It consists of a stalk on the end of which is the anther.

Starter - narrow strip of wax foundation fastened in a frame or under a top bar for the bees to start building comb in the desired position.

Super - any hive box placed above the brood nest or chamber. Usually contains combs in which bees will store honey.

Sustainable livelihood - occupation or employment enabling someone to provide for his/her basic needs and to be secure in its continuation in the future.

Swarming - when a honeybee colony becomes large enough to divide into two.

Top bar hive - intermediate technology hive in which bees are encouraged to build their combs suspended from bars placed across the top of the hive.

Unripe honey - honey in cells not fully processed by hive bees, i.e., when still thin and not capped.

Worker bee - female honeybees that make up the bulk of the colony and undertake all the work of the colony except for mating and egg laying. Workers are sterile females.

ANNEX 1
Current Status of the Kenyan Beekeeping Industry

BRIEF HISTORY OF BEEKEEPING IN KENYA

In the late 1960s, the Kenyan Ministry of Agriculture, the International Bee Research Association (IBRA), and Oxfam collaborated on the development of an appropriate hive for Kenya. The result of this joint effort is the Kenyan Top Bar Hive (KTBH). In the 1970s, the Ministry of Agriculture began introducing the KTBH to Kenyan farmers supported by the Canadian International Development Agency (CIDA). Beekeeping cooperatives were set up in many parts of Kenya to facilitate marketing of bee products. All of these marketing structures collapsed later. Many have made numerous attempts to develop beekeeping in the country since the 1960s, but production of honey is still very low in relation to its potential. In recent years, people have promoted Langstroth frame hives as the answer to Kenya's beekeeping problems but with limited success. Neither the Langstroth frame hive nor KTBH have had much impact on Kenyan beekeeping, as about 80% of all beehives in Kenya are still traditional type hives. In spite of the production potential, Kenya does not export honey. The European Union has licensed Kenya to export to their region.

BACKGROUND TO THE KENYAN BEEKEEPING INDUSTRY

Kenya relies heavily on agriculture like other East African countries. 75% of its people live in rural areas and 60% of these live in absolute poverty. Kenya is a nation of small holders with over five million small-scale farmers and pastoralists. Cutbacks in public services and the free market philosophy of recent years have hit rural communities very hard. As this is unlikely to change, the future of such rural communities will depend on developing their capacities from within to meet the challenge. Beekeeping is an opportunity to harvest and add value to a local resource (floral nectar) to generate wealth and employment and beat poverty.

80% of Kenya's land area is either semi-arid or arid land. Beekeeping is suitable for these areas due to prevalent plant species, e.g., acacia trees, and also due to limited livelihood options available in these areas. About 80% of Kenyan honey is produced in arid and semi-arid areas. (Muya 2004)

Majority of Kenyan beekeepers still use traditional systems of beekeeping: simple fixed combs, mostly hollow log hives, in spite of over 30 years of beekeeping extension carried out by the government and non-governmental organisations (NGOs) to promote improved hives, e.g., the Kenya Top Bar Hive, which is an intermediate technology hive.

In the last five years, there has been a major push by some NGOs and private companies, supported by major donors, to introduce the Langstroth frame hive used in Europe and America. There is conflicting information between the actual impact of these hives and the claim of success by its promoters. Information collected in a number of recent beekeeping studies indicates limited impact on enhanced production of bee products and on improvement of livelihoods. Conflicting information on different production technologies results in confused farmers who receive different messages from different extension agents.

In relation to bee product marketing, research indicates that the Kenyan honey market is underdeveloped due to low volumes. Volumes and quality have not yet been reached for export though Kenya has been licensed to export honey to the European Union since 2003. The shortage of bee product production also means that local demands are not met sufficiently. Significant quantities of imported honey, mostly from Tanzania, abound in the Kenyan market, but some people repackage them as Kenyan honey. Opportunities exist for the development of fair trade and organic honey export markets due to Kenya's largely pollution-free environment and disease-free bees.

Some quarters report that local production of honey has improved in recent years due to reduced official honey imports; however, one can explain it by an increased cross-border trade in honey marketed as Kenyan honey in local markets.

Other issues of importance affecting the Kenyan beekeeping industry include:

- Lack of a clear government beekeeping policy but which is now under development. Different agencies take different approaches to beekeeping with no national coordinating body resulting in a disorganised and fragmented industry.

- Environmental degradation, mostly forest destruction/charcoal burning that results in reduced beekeeping potential in Kenya.

- Little or no beekeeping research in Kenya.

- Fragmentation and disorganisation of beekeepers. Beekeepers are difficult to reach due to poor infrastructure resulting in difficulties in bulking honey for marketing.

OPPORTUNITIES IN KENYAN BEEKEEPING

- There is a strong local demand for honey with current production unable to meet local demand.
- There is a ready export market with a strong international demand for honey and bee products. The European Union allows Kenya to export to their region.
- The bee population is relatively disease-free.
- There are considerable existing skills and knowledge in beekeeping.
- Kenya is a big country with space for a much larger beekeeping industry.

ANNEX 2
More Information on Stingless Bees

Apart from honeybees (*Apis mellifera*), there are other types of bees in Kenya that collect nectar and make honey. These are stingless bees, which usually live in holes in the ground. There are many species of stingless bees (*Meliponinae*) in Kenya and these are yet unexploited on a commercial basis. In Brazil and India, these bees are utilised because of their ability to make good quality medicinal honey. People prize this honey as a cure for baldness and impotence. These bees can yield up to five kgs per season. Stingless bees are the best pollinators as they penetrate deeper into the flower. Baraka Agricultural College in Molo and the National Beekeeping Station at Lenana in Nairobi have started keeping stingless bees. Beekeepers can keep them in small hives. They store honey in small pots that you can harvest daily with a clean syringe!

Photo 44: Hive with shaded glass cover.
Bees require darkness.

Photo 43: Simple box hive for stingless bees.
(Baraka Agricultural College, Molo)

ANNEX 3
Hive Recordkeeping

Hive No.	Occupation Date	Date of Inspection and Comments	Harvesting Date	Amount Crude Honey (kgs)	Amount Refined Honey (kgs)	Amount Beeswax (kgs)	Remarks

You can copy the previous table in a hardback notebook, which you should fill out after each hive inspection. You can also photocopy it. In the remarks section, you can specify the inspection date and any information of use such as the temperament of the bees, i.e., if they are docile or difficult to handle. This information is useful when selecting bees in the future. You can record any management activities here such as: (a) amount of food fed to bees (if any), (b) the cost, (c) action taken to control swarming, etc.

Below is a table for marketing records.

Date	Amount of Honey Sold (kgs)	Price per kg of Honey	Total Revenue

Table 9: Marketing Recordkeeping

ANNEX 4
Recipes/Uses for Beeswax, Propolis, and Honey

Photo 45: Kenyan National Beekeeping Station exhibition stand
The stall displayed many excellent bee products.

BEESWAX FURNITURE/WOOD POLISH

Ingredients
200g Beeswax
100g Turpentine
50g Baby oil

Procedure
Grate beeswax into flakes. Gradually add turpentine to soften wax. Add oil and mix. Store in a tin with a tight-fitting top or in a jar.

Tip for sewing shoes, leather, and other thick materials:

Pull thread through small block of beeswax. The wax stiffens and smoothes the thread making it easier to sew.

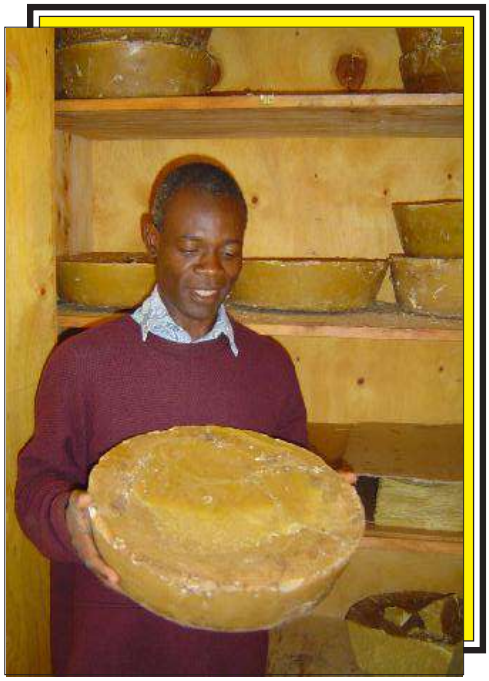

Photo 46: Cake of beeswax for making candles
(Baraka Agricultural College, Molo)

BEESWAX CANDLES

Requirements
Beeswax
Wick (use cotton thread)
Bottle top
Plastic pipe that fits into bottle top
Cooking oil

Procedure
Smear inside of plastic pipe with soapy water or cooking oil. Thread wick through pipe. Punch hole in the middle of the bottle top. Thread wick through hole and knot it.

Melt beeswax indirectly, i.e., use a double pan (one *sufuria* sitting in hot water inside another *sufuria*). Do not boil the wax.

Cover bottom end of pipe to prevent wax flowing away by pushing it down into bottle top. You can also stick end of pipe in the ground. Tie wick to suspension wire. Pour melted wax into mould making sure wick is in the middle of the candle. Wait for wax to cool before filling the rest of the mould with molten wax. Leave candle to solidify. (It should come out easily from mould because you had greased it with cooking oil.)

Candle is now ready for use. Vary wick to get a good candle that burns well. Use many plastic pipes to make many candles at the same time.

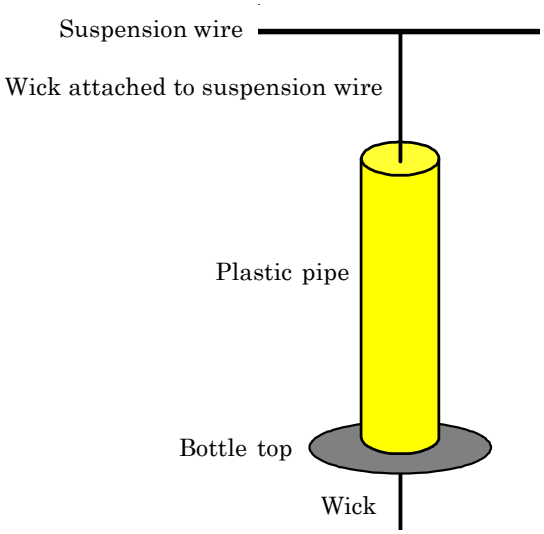

Figure 11: Candle Making

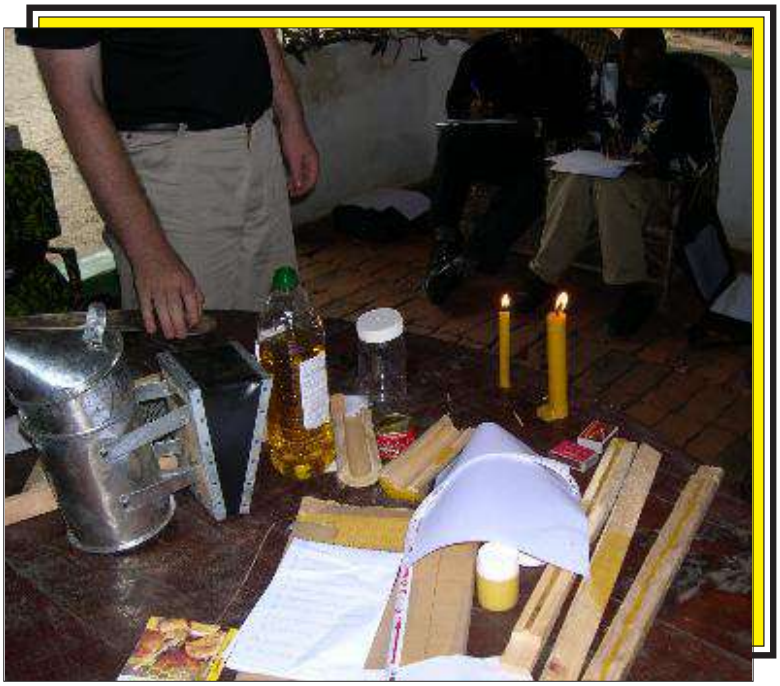

Photo 47: Beautiful beeswax candles made using plastic pipe mould

PROPOLIS HEALING OINTMENT

The first stage is to make propolis granules:

Step 1
1. Scrape off the propolis from hives putting scrapings in a tin.
2. Add enough water to cover scrapings and heat.
3. Wax will melt and float to surface.
4. Propolis will stick to bottom of can.
5. Stir often to help release wax.
6. Remove from heat and let cool.

Step 2
1. Remove waxy layer formed on water surface.
2. Pour off water and save coloured mass beneath.

Step 3
1. Place can in cool place to harden propolis andmake it brittle. Freezer is best.
2. Dry and store in sealed container.

Ingredients
1 tablespoon Beeswax
4 tablespoons Mineral oil (e.g., baby oil)
1 tablespoon Propolis granules
1 tablespoon Honey

Procedure
Melt ingredients in double boiler. Stir until it cools. It should have the consistency of cream. You can pack in small containers and sell. Use cream as a skin ointment for sores, rashes, blisters, etc. Many farmers' groups we work with are now making and selling beeswax and propolis creams and making good money from this activity. For more information, contact Baraka Agricultural College. (See contacts in Annex 5.)

HONEY COUGH SYRUP

Ingredients
6 tablespoons Pure honey
Juice of 2 Lemons

Mix ingredients together well before taking in spoonfuls for coughs.

Warning: Be careful when working with hot wax and other ingredients used to make these recipes. I cannot be held responsible for adverse effects resulting from the use of methods or ingredients mentioned in this publication.

ANNEX 5
Approaches to Beekeeping Extension

In this section, we outline experiences of Baraka Agricultural College that has been involved in beekeeping extension services to farmers since 2002. An Irish development agency, Self-Help Development International (SHDI) funded the extension project. Since 2002, the college has been working with beekeeping groups to develop beekeeping mostly in Rift Valley Province but also in other parts of Kenya/Eastern Africa.

INNOVATIVE ASPECTS OF THE BAC/SHDI PROGRAMME

1. Participatory Action Research (PAR) Approach to Beekeeping

As a country, Kenya covers a very large area presenting a huge diversity of climates and vegetation within its borders. Beekeeping extension workers can only make general recommendations. There is enormous need for beekeepers to adapt and try out ideas for themselves at local level. The basis of the PAR approach is the importance of learning from what we do in a systematic way. Every aspect of life can benefit from this kind of approach. The Baraka/SHDI project encourages farmers to innovate/solve their own problems in beekeeping. This approach comes in light of an inadequate government beekeeping extension service as well as little or no formal research in Kenya beekeeping. In November 2006, farmers researched reasons why they have low hive occupation rates, so they can make decisions on what to do about their problem.

The PAR approach to beekeeping development takes into account and builds on farmers' indigenous technical knowledge.

Photo 48: Bee forage tree seeds
BAC/SHDI farmer researchers collected and documented these seeds.

2. Bee Multiplication

There is considerable shortage of bees and many hives are unoccupied. The project works with farmers to ensure bees occupy hives. This is done through training and support in colony division and swarm trapping. The project has started training farmers in queen rearing in response to one of Kenya's biggest beekeeping problems: low hive occupation rates.

3. Farmer-to-Farmer Trainers

The project works with beekeeping groups in different geographic areas. A cluster of groups has a farmer beekeeping extension worker that the project trained and supported. In turn, he/she trains other farmers in beekeeping. This has proved a very effective beekeeping promotion method and builds local capacity in beekeeping.

4. Artisan Training

The Baraka Agricultural College/SHDI project trains local carpenters and tailors in making beekeeping equipment. This has assisted beekeeping development as beekeepers can access beekeeping equipment at affordable prices. This approach also encourages self-reliance amongst local communities.

(N. B. *For further information on the Baraka Agricultural College/SHDI project, please contact the college. See Annex 6.*)

ANNEX 6
Sources of Further Information

KENYA

You can contact the following organisations as a starting point for further information on beekeeping in Kenya. Many others are involved in beekeeping; the number changes all the time. This list is not exhaustive and is just to establish a place for further networking.

Apiculture Division
Ministry of Livestock and Fisheries Development, Hill Plaza Building
P.O. Box 34188–00100
Nairobi
Telephone: (254) (0)20 2722601/2722637
Fax: (254) (0)20 2728609
Remarks: Contact for information and advice or get in touch with your local District Beekeeping Officer. There's one in every district.

Baraka Agricultural College
P.O. Box 52
Molo
Telephone: (254) (0)51 721091
Email: baraka@sustainableag.org
Website: www.sustainableag.org
Remarks: Contact for bee equipment, advice, training courses, honey, and beeswax marketing

The Kenya Honey Council
P.O. Box 271-00606
Sarit Centre, Nairobi
Telephone: (254) (0)20 4183120
Email: info@kenyahoneycouncil.org
Website: www.kenyahoneycouncil.org
Remarks: An umbrella body representing different stakeholders in the Kenya beekeeping sector.

National Beekeeping Station
Apiculture and Emerging Livestock Division
Ministry of Livestock and Fisheries Development Lenana
P.O. Box 34188-00100
Nairobi
Telephone: (254) (0)20 564302
Remarks: Check out for its library, bee equipment, and advice.

The Kenya Beekeepers Association
c/o The National Beekeeping Station
P.O. Box 34188-00100
Nairobi

Self-Help Development International Kenya
2nd Floor, Catholic Diocese of Nakuru Building
Stadium Road off Kenyatta Avenue
Telephone: (254) (0)51 2212291
Email: kenya@shdi.org
Remarks: The agency currently promotes beekeeping in the Gilgil area.

Ruai Beekeeping Cooperative Society
P.O. Box 8
Naru Moru
Remarks: This is a great example of collective marketing of bee products in Kenya. It markets up to 8 tonnes of honey per annum.

ICIPE (International Centre for Insect Physiology and Ecology)
P.O. Box 30772-00100
Nairobi
Telephone: (254) (0)20 8632000
Fax: (254) (0)20 8632001/8632002
Remarks: It has a commercial insect section dealing with beekeeping
E-mail: dg@icipe.org

African Union
Maendeleo Ya Wanawake House
Nairobi
Remarks: It has an interesting collection of beekeeping books in its library.

General Plastics Limited
P.O. Box 10032
Nairobi
Enterprise Road off Mombasa Road
Industrial Area near Hillock Inn
Telephone: (254) (0)20 530032/3/4/5
Remarks: It supplies plastic jars and lids for packing honey.

Honey Care Africa Limited
Muringa Avenue, Jamhuri Park
Nairobi
Telephone: (254) (0)20 574448
Remarks: It promotes Langstroth hives. It also buys and markets honey.

African Beekeepers Limited
Industrial Area
P.O. Box 3752-00506
Nairobi
Mobile: (254) (0)722 700226
Remarks: It manufactures equipment and markets honey.

Kerio Valley Development Authority
KVDA Plaza
P.O. Box 2660
Eldoret
Telephone: (254) (0)53 2063361
Email: kvda@kenyaweb.com

Strengthening Informal Sector Training and Enterprise (SITE)
Telephone: (254) (0)20 2718155
Jabavu Road, Nairobi
Remarks: It deals in training and support to beekeeping and bee equipment.

REST OF AFRICA

Bee Natural Products Ltd
PO Box 5318
Pan Africa House
Plot 3 Kimathi Avenue
Kampala, Uganda
Telephone: (256) (0)41 234676/ 253844
Mobile: (256) (0)77 907045
Fax: (256) (0)41 259052
Website: www.beenaturalproducts.com

Tanzanian Ministry of Natural Resources and Tourism
P.O. Box 9372
Dar-es-Salaam, Tanzania
Telephone: (255) (0)22 130091
Email: fordev@africaonline.co.tz
Remarks: It has very good publications on beekeeping in English and Swahili.

The National Honey Show Limited
P.O. Box 71326
Dar-es-Salaam, Tanzania
Telephone: (255) (0)22 2126280
Email: epictanzania@email.com
Remarks: It organised the first Tanzanian honey show in Dodoma in 2006.

Tawiri Njiro Wildlife Research Centre
Ministry of Natural Resources and Tourism, Forestry and Beekeeping Division
c/o P.O. Box 246
Dar-es-Salaam, Tanzania
Email: fordev@africaonline.co.tz
Remarks: Tawiri runs beekeeping courses and carries out beekeeping research.

Alternative Contacts:
P.O. Box 661
Arusha, Tanzania
Telephone: (255) (0)27 7677
Telefax: (255) (0)27 2548240
E-mail: tawiri@habari.co.tz or info@tawiri.org

Plant Protection Research Institute South Africa
Private Bag x134
0001 Pretoria, South Africa
Website: www.arc.agric.za
Remarks: It has a very good book, *Beekeeping in South Africa*.

TUNADO (The Ugandan National Apiculture Development Organisation)
c/o Bee Natural Products Limited (See above.)

INTERNATIONAL

Bees for Development
Troy, Monmouth NP25 4AB
United Kingdom
Telephone: (44) 016007 13648
Fax: (44) 016007 16167
Website: www.beesfordevelopment.com
Remarks: It produces a very nice magazine. People may pay for subscriptions with beeswax. It also sells beekeeping books.

Bee Support
Email: micha_el@dds.nl
Remarks: It undertakes promotion of beekeeping for development.

Hives Save Lives
Saltdean, Brighton BN51 9AB
United Kingdom
Website: www.hivessavelives.com
Remarks: This organization has current projects on Uganda.

Bees Abroad UK
Website: www.beesabroad.org.uk
Remarks: A UK-based charity supporting beekeeping development. Bees Abroad also has projects in Kenya.

International Bee Research Association (IBRA)
18 North Road, Cardiff CF1 3DY
United Kingdom
Website: www.ibra.org.uk
Remarks: It has a good library on beekeeping around the world and publishes the *Journal of Apicultural Research* and *Bee World*.

CTA/Agromisa
P.O. Box 41
Wageningen, The Netherlands
Remarks: It has booklets on agriculture and bees.

Bee Vital
Website: www.BeeVitalPropolis.com
Remarks: A UK-based company that purchases propolis.

ANNEX 7
Baraka Agricultural College Short Courses on Beekeeping

The following one-week short courses in beekeeping are currently on offer from Baraka Agricultural College. It offers courses both in Swahili and English, depending on the needs of the group. For further information on course timings and costs, please get in touch with the Short Courses Coordinator at Baraka Agricultural College. See contact details below.

An Introduction to Beekeeping (Establishing a Beekeeping Enterprise)

This course is about starting a beekeeping enterprise in Kenya. The college aims the course at beekeeping beginners, but experienced beekeepers can also benefit by attending. The course covers an introduction to bee biology and behaviour (African bees, *Apis mellifera scutellata*), beekeeping equipment, apiary establishment, bee and hive management, harvesting, and handling and marketing honey.

Processing of Bee Products

This course covers in detail how to process and add value to bee products, including honey, beeswax, propolis, bee brood, and pollen. Local and international markets/ marketing for these products is also covered. One spends much time in practical lessons, i.e., making a variety of value-added recipes, which one can market and sell in Kenya and overseas. Recipes include a variety of candles and creams, polishes, and medicinal/nutritional products from the beehive.

Making Bee Equipment

Making bee equipment is a specialised course aimed at artisans and those interested in making their own beekeeping equipment to quality standards. Both beekeeping staff and Baraka Agricultural College workshop staff, who manufacture beekeeping equipment, facilitate the course. The course covers making a Kenya Top Bar Hive, Langstroth hive, smoker, bee suits, hive tools, and bee brushes. They also give details on how to make a simple honey extractor. Much of the course held at Baraka Agricultural College workshop is actually making the equipment.

Bee Breeding and Multiplication

This is a new course on offer for 2007. The course is designed to assist beekeepers/ beekeeping extension workers occupy the many empty beehives in Kenya. Research shows hive occupation rates as low as 30% in some areas of the country. The course covers details of bee genetics, selection, and breeding. Simple methods of bee multiplication such as swarm attractants, making divisions, and queen rearing are covered in detail, including the Demaree, Miller, Jenter, and Grafting methods of rearing queens.

Further details of Baraka Agricultural College beekeeping courses are available from the college website: www.sustainableag.org. You can also contact the Short Courses Coordinator at telephone number: (254) (0)51 721091; or send email to shortcourses@sustainableag.org.

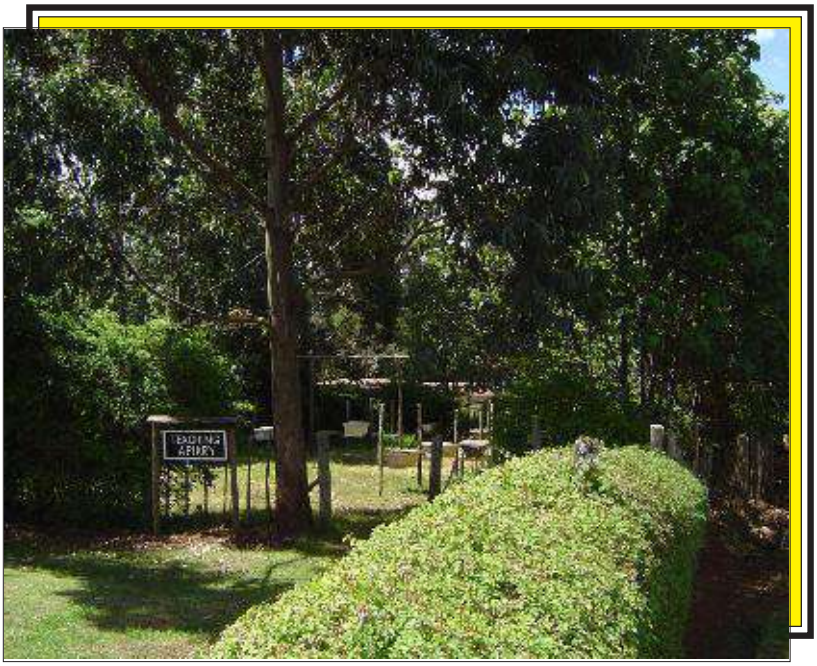

Photo 49: The teaching apiary at Baraka Agricultural College

ANNEX 8
Beekeeping Resources on the Internet

There are a number of websites, which are of interest to beekeeping in Kenya and Africa. They offer a large variety of resources and information. It is well worth checking out internet resources. There is a list of beekeeping resources suitable for Africa on the links page of Apiconsult. See www.apiconsult.com/beekeeping-links.htm. This page is updated regularly.

In addition, the Apiconsult website has a discussion forum on African beekeeping where beekeepers on the African continent share information. Please join the discussion and share your experiences.

As of November 2006, I consider the following websites useful:

www.apiconsult.com

www.apiservices.com (loads of information on world beekeeping)

www.beesfordevelopment.org (UK-based site promoting beekeeping in development)

www.gsu.edu/~biojdsx/main.htm (beekeeping in top bar hives)

www.honeybadger.co.za (South African site on beekeeping)

www.honeycareafrica.com (beekeeping with Langstroth hives in Kenya)

www.topbarhive.com (information on top bar hives in the USA)

en.wikipedia.org/wiki/Top-bar_hive (online encyclopaedia with information on top bar hives and general beekeeping to which you can edit and add your own ideas and information to the articles presented)

There are many more.

A good way to generate more beekeeping links is to search with a search engine such as Google (www.google.co.ke) on “beekeeping Africa” or “top bar hives”, etc.

Happy surfing!

Encyclopedia of INSECTS

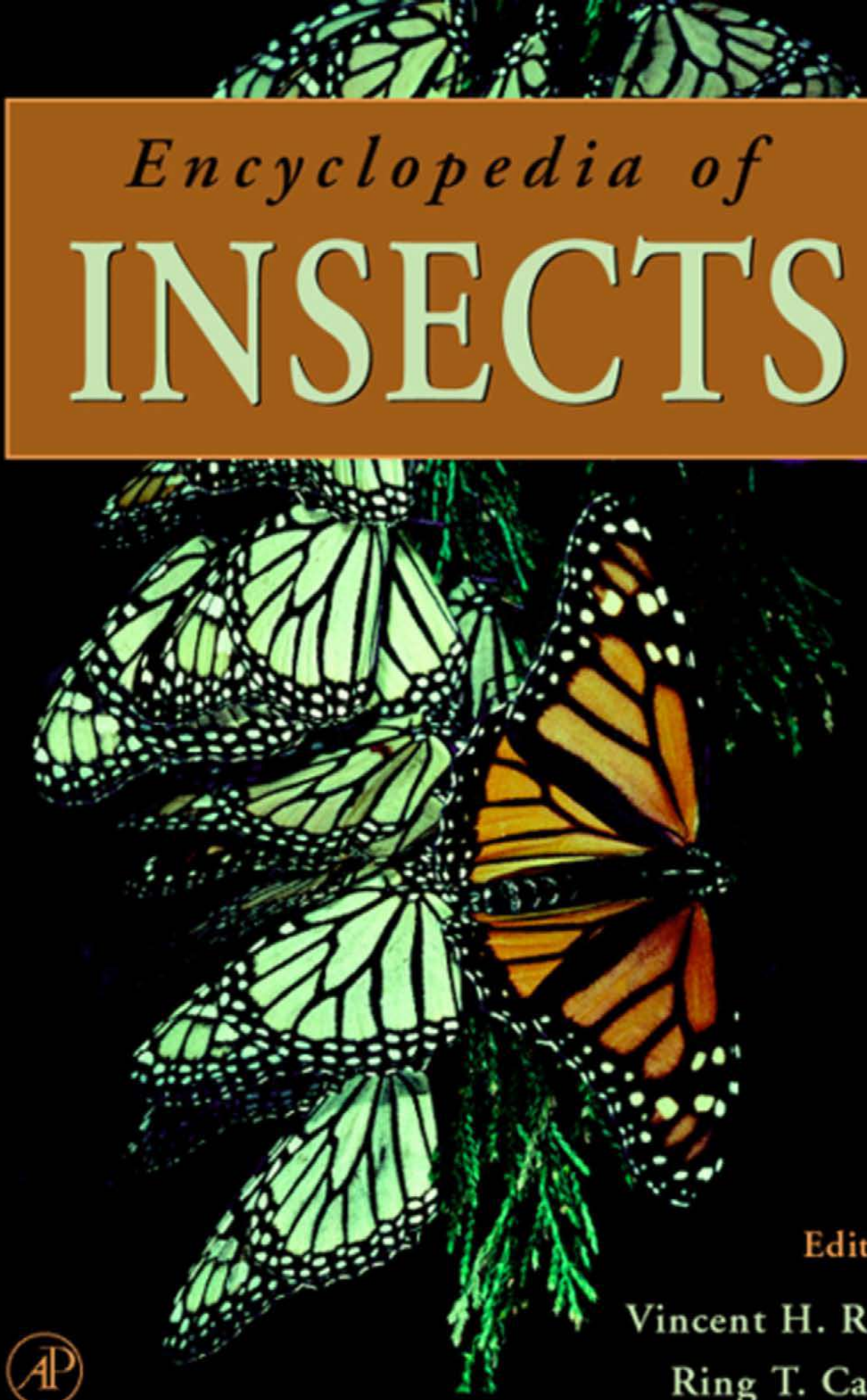

Editors

Vincent H. Resh

Ring T. Cardé

A

Acari

see *Mites; Ticks*

Accessory Glands

Diana E. Wheeler
University of Arizona

The accessory glands of reproductive systems in both female and male insects produce secretions that function in sperm maintenance, transport, and fertilization. In addition, accessory glands in females provide protective coatings for eggs. Accessory glands can be organs distinct from the main reproductive tract, or they can be specialized regions of the gonadal ducts (ducts leading from the ovaries or testes). Typically, glandular tissue is composed of two cell types: one that is secretory and the other that forms a duct. The interplay between male and female secretions from accessory glands is a key element in the design of diverse mating systems.

ACCESSORY GLANDS OF FEMALES

Management of Sperm and Other Male Contributions

Sperm management by females involves a wide range of processes, including liberation of sperm from a spermatophore, digestion of male secretions and sperm, transport of sperm to and from the spermatheca, maintenance of stored sperm, and fertilization.

Accessory gland secretions can have digestive functions important in sperm management. First, digestive breakdown of the spermatophore can free encapsulated sperm for fertilization and storage. Second, male contributions can provide an important nutritional benefit to their mates. Female secretions can digest the secretory components of male seminal fluid to facilitate a nutritive role. In addition, females can digest unwanted sperm to transform it into nutrients. Third, female secretions in some species are required to digest sperm coverings that inhibit fertilization.

Transfer of sperm to and from the spermatheca is generally accomplished by a combination of chemical signals and muscular contractions. Secretions of female accessory glands in some species increase sperm motility or appear to attract sperm toward the spermathecae. Transport of fluid out through the wall of the spermatheca may also create negative pressure that draws in sperm.

Sperm can be stored for some length of time in spermathecae, with the record belonging to ant queens that maintain sperm viability for a decade or more. Secretions of spermathecal glands are poorly characterized, and how sperm is maintained for such extended periods is not known. Spermathecal tissue seems to create a chemical environment that maintains sperm viability, perhaps through reduced metabolism. A nutritional function is also possible.

Transport of sperm out of storage can be facilitated by the secretions of the spermathecal gland, which presumably activate quiescent sperm to move toward the primary reproductive tract. One potential function of female accessory glands that has been explored only slightly is the production of hormone-like substances that modulate reproduction functions.

Production of Egg Coverings

Female accessory glands that produce protective coverings for eggs are termed colleterial glands. Colleterial glands have been best characterized in cockroaches, which produce an oothecal case surrounding their eggs. Interestingly, the left

and right glands are anatomically different and have different products. Separation of the chemicals permits reactions to begin only at the time of mixing and ootheca formation. Other protective substances produced by glands include toxins and antibacterials.

Nourishment for Embryos or Larvae

Viviparous insects use accessory glands to provide nourishment directly to developing offspring. Tsetse flies and sheep keds are dipterans that retain single larvae within their reproductive tracts and provide them with nourishment. They give birth to mature larvae ready to pupate. The gland that produces the nourishing secretion, rich in amino acids and lipids, is known as the milk gland. The Pacific beetle roach, *Diploptera punctata*, is also viviparous and provides its developing embryos with nourishment secreted by the brood sac, an expanded portion of oviduct.

ACCESSORY GLANDS OF MALES

Accessory glands of the male reproductive tract have diverse functions related to sperm delivery and to the design of specific mating systems.

Sperm Delivery

Males of many insects use spermatophores to transfer sperm to females. A spermatophore is a bundle of sperm contained in a protective packet. Accessory glands secrete the structural proteins necessary for the spermatophore's construction. Males of the yellow mealworm, *Tenebrio molitor*, have two distinct accessory glands, one bean-shaped and the other tubular (Fig. 1). Bean-shaped accessory glands contain cells of at least seven types and produce a semisolid material that forms the wall and core of the spermatophore. Tubular accessory glands contain only one type of cell, and it produces a mix of water-soluble proteins of unknown function. Spermatophores are not absolutely required for sperm transfer in all insects. In many insects, male secretions create a fluid medium for sperm transfer.

Effects on Sperm Management and on the Female

The effects of male accessory gland secretions in the female are best known for the fruit fly, *Drosophila melanogaster*, in which the function of several gene products has been explored at the molecular level. Since insects have a diversity of mating systems, the specific functions of accessory gland secretions are likely to reflect this variation.

In *Drosophila*, the accessory glands are simple sacs consisting of a single layer of secretory cells around a central lumen (Fig. 2). Genes for more than 80 accessory gland proteins have been identified so far. These genes code for hormonelike substances and enzymes, as well as many novel

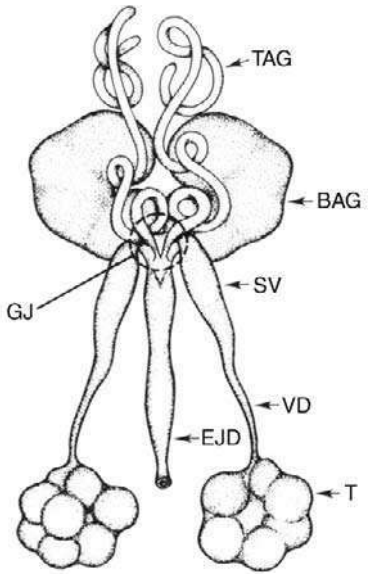

FIGURE 1 Male reproductive system of *T. molitor*, showing testes (T), ejaculatory duct (EJD), tubular accessory gland (TAG), and bean-shaped accessory gland (BAG). [From Dailey, P. D., Gadzama J. M., and Happ, G. M. (1980). Cytodifferentiation in the accessory glands of *Tenebrio molitor*. VI. A congruent map of cells and their secretions in the layered elastic product of the male bean-shaped accessory gland. *J. Morphol.* **166**, 289–322. Reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.]

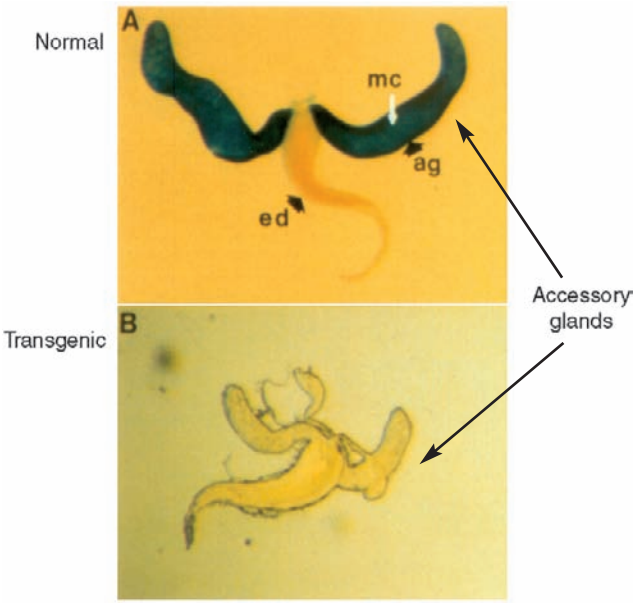

FIGURE 2 Accessory gland of *D. melanogaster*. (A) The cells in this normal accessory gland express b-galactosidase driven by a promoter of a gene for an accessory gland protein. (B) A transgenic accessory gland, cells expressing the gene have been selectively killed after eclosion. These flies were used to explore the function of accessory gland secretions. In transgenic males, accessory glands are small and translationally inert. [From Kalb, J. M., DiBenedetto, A. J., and Wolfner, M. F. (1993). Probing the function of *Drosophila melanogaster* accessory glands by directed cell ablation. *Proc. Natl. Acad. Sci. USA* **90**, 8093–8097. Copyright 1993, National Academy of Sciences, U.S.A.]

proteins. The gene products or their derivatives have diverse functions, including an increased egg-laying rate, a reduced inclination of females to mate again, increased effectiveness of sperm transfer to a female's spermatheca, and various toxic effects most likely involved in the competition of sperm from different males. A side effect of this toxicity is a shortened life span for females. Other portions of the reproductive tract contribute secretions with diverse roles. For example, the ejaculatory bulb secretes one protein that is a major constituent of the mating plug, and another that has antibacterial activity.

See Also the Following Articles

Egg Coverings • Spermatheca • Spermatophore

Further Reading

Chen, P. S. (1984). The functional morphology and biochemistry of insect male accessory glands and their secretions. *Annu. Rev. Entomol.* **29**, 233–255.

Eberhard, W. G. (1996). "Female Control: Sexual Selection by Cryptic Female Choice." Princeton University Press, Princeton, NJ.

Gillott, C. (1988). Arthropoda—Insecta. *In* "Accessory Sex Glands," (Adiyodi and Adiyodi, eds.). Vol. 3 of "Reproductive Biology of Invertebrates," pp. 319–471. Wiley, New York.

Happ, G. M. (1992). Maturation of the male reproductive system and its endocrine regulation. *Annu. Rev. Entomol.* **37**, 303–320.

Wolfner, M. F. (2001). The gifts that keep on giving: Physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila. Heredity* **88**, 85–93.

Acoustic Behavior

see *Hearing*

Aedes Mosquito

see *Mosquitoes*

Aestivation

Sinzo Masaki
Hirosaki University

Aestivation is a dormant state for insects to pass the summer in either quiescence or diapause. Aestivating, quiescent insects may be in cryptobiosis and highly tolerant to heat and drought. Diapause for aestivation, or summer diapause, serves not only to enable the insect to tolerate the rigors of summer but also to ensure that the active phase of the life cycle occurs during the favorable time of the year.

QUIESCENCE

Quiescence for aestivation may be found in arid regions. For example, the larvae of the African chironomid midge, *Polypedilum vanderplanki,* inhabit temporary pools in hollows of rocks and become quiescent when the water evaporates. Dry larvae of this midge can "revive" when immersed in water, even after years of quiescence. The quiescent larva is in a state of cryptobiosis and tolerates the reduction of water content in its body to only 4%, surviving even brief exposure to temperatures ranging from +102°C to –270°C. Moreover, quiescent eggs of the brown locust, *Locustana pardalina,* survive in the dry soil of South Africa for several years until their water content decreases to 40%. When there is adequate rain, they absorb water, synchronously resume development, and hatch, resulting in an outburst of hopper populations. The above-mentioned examples are dramatic, but available data are so scanty that it is difficult to surmise how many species of insects can aestivate in a state of quiescence in arid tropical regions.

SUMMER DIAPAUSE

Syndrome

The external conditions that insects must tolerate differ sharply in summer and winter. Aestivating and hibernating insects may show similar diapause syndromes: cessation of growth and development, reduction of metabolic rate, accumulation of nutrients, and increased protection by body coverings (hard integument, waxy material, cocoons, etc.), which permit them to endure the long period of dormancy that probably is being mediated by the neuroendocrine system.

Migration to aestivation sites is another component of diapause syndrome found in some species of moths, butterflies, beetles, and hemipterans. In southeastern Australia, the adults of the Bogong moth, *Agrotis infusa,* emerge in late spring to migrate from the plains to the mountains, where they aestivate, forming huge aggregations in rock crevices and caves (Fig. 1).

Seasonal Cues

Summer diapause may be induced obligatorily or facultatively by such seasonal cues as daylength (nightlength) and temperature. When it occurs facultatively, the response to the cues is analogous to that for winter diapause; that is, the cues are received during the sensitive stage, which precedes the responsive (diapause) stage. The response pattern is, however, almost a mirror image of that for winter diapause (Fig. 2). Aestivating insects themselves also may be sensitive to the

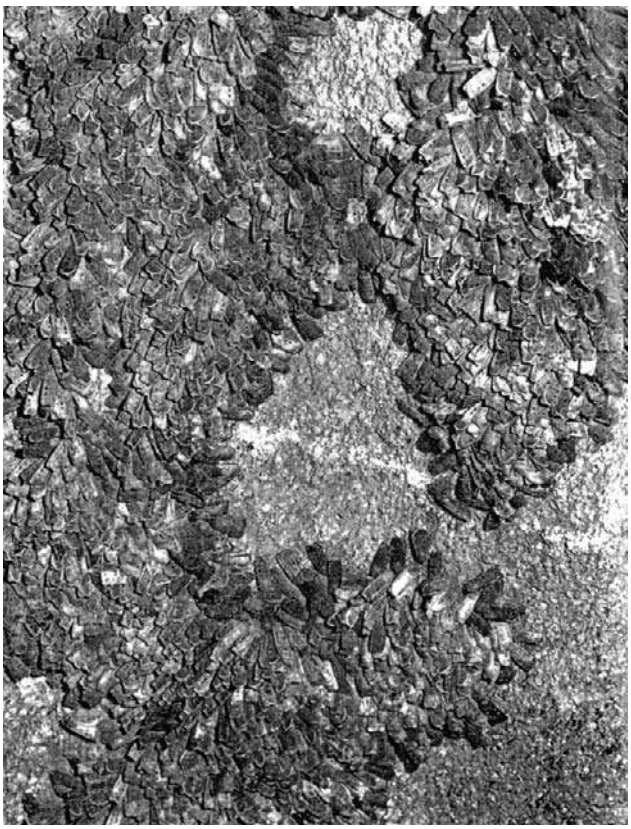

FIGURE 1 Bogong moths, *Agrotis infusa,* aestivating in aggregation on the roof of a cave at Mt Gingera, A. C. T., Australia. [Photograph from Common, I. (1954). *Aust. J. Zool.* **2,** 223–263, courtesy of CSIRO Publishing.]

seasonal cues; a high temperature and a long daylength (short nightlength) decelerate, and a short daylength (long nightlength) and a low temperature accelerate the termination of diapause.

The optimal range of temperature for physiogenesis during summer diapause broadly overlaps with that for morphogenesis, or extends even to a higher range of temperature. Aestivating eggs of the brown locust, *L. pardalina,* can terminate diapause at 35°C and those of the earth mite, *Halotydeus destructor,* do this even at 70°C. The different thermal requirements for physiogenesis clearly distinguish summer diapause from winter diapause, suggesting that despite the superficial similarity in their dormancy syndromes, the two types of diapause involve basically different physiological processes.

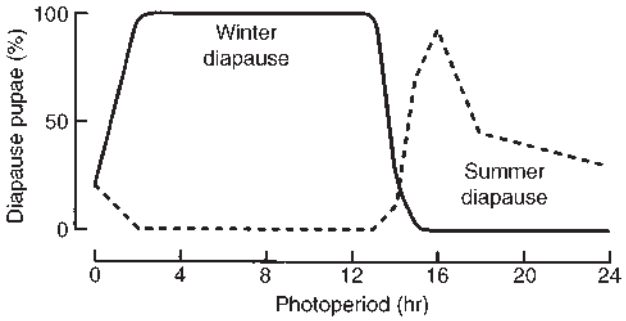

FIGURE 2 Photoperiodic response in the noctuid *M. brassicae* controlling the pupal diapause at 20°C. Note the different ranges of photoperiod for the induction of summer diapause (dashed line) and winter diapause (solid line). [From Furunishi *et al.,* 1982, reproduced with permission.]

See Also the Following Articles

Cold/Heat Protection • Diapause • Dormancy • Migration

Further Reading

Common, I. F. B. (1954). A study of the biology of the adult Bogong moth, *Agrotis infusa* (Boisd.) (Lepidopera: Noctuidae), with special reference to its behaviour during migration and aestivation. *Austral. J. Zool.* **2,** 223–263.

Furunishi, S., Masaki, S., Hashimoto, Y., and Suzuki, M. (1982). Diapause response to photoperiod and night interruption in *Mamestra brassicae* (Lepidoptera: Noctuidae). *Appl. Entomol. Zool.* **17,** 398–409.

Hinton, H. E. (1960). Cryptobiosis in the larva of *Polypedilum vanderplanki* Hint. (Chironomidae). *J. Insect Physiol.* **5,** 286–300.

Masaki, S. (1980). Summer diapause. *Annu. Rev. Entomol.* **25,** 1–25.

Matthée, J. J. (1951). The structure and physiology of the egg of *Locustana pardalina* (Walk.). *Union S. Afr. Dept. Agric. Sci. Bull.* **316,** 1–83.

Tauber, M. J., Tauber, C. A., and Masaki, S. (1986). "Seasonal Adaptations of Insects." Oxford University Press, New York.

Africanized Bees

see *Neotropical African Bees*

Agricultural Entomology

Marcos Kogan
Oregon State University

Ronald Prokopy
University of Massachusetts

The study of all economically important insects is the object of the subdiscipline "economic entomology." Agricultural entomology, a branch of economic entomology, is dedicated to the study of insects of interest to agriculture because they help increase crop production (e.g., pollinators); help produce a commodity (e.g., honey, silk, lacquer); cause injury leading to economic losses to plants grown for food, feed, fiber, or landscaping; cause injury to farm animals; or are natural enemies of agricultural pests and, therefore, considered to be beneficial. Study of all fundamental aspects of the ecology, life history, and behavior of insects associated with agricultural crops and farm animals falls within the realm of agricultural entomology. These studies provide the foundation for the design and implementation of integrated pest management (IPM) programs (Fig. 1).

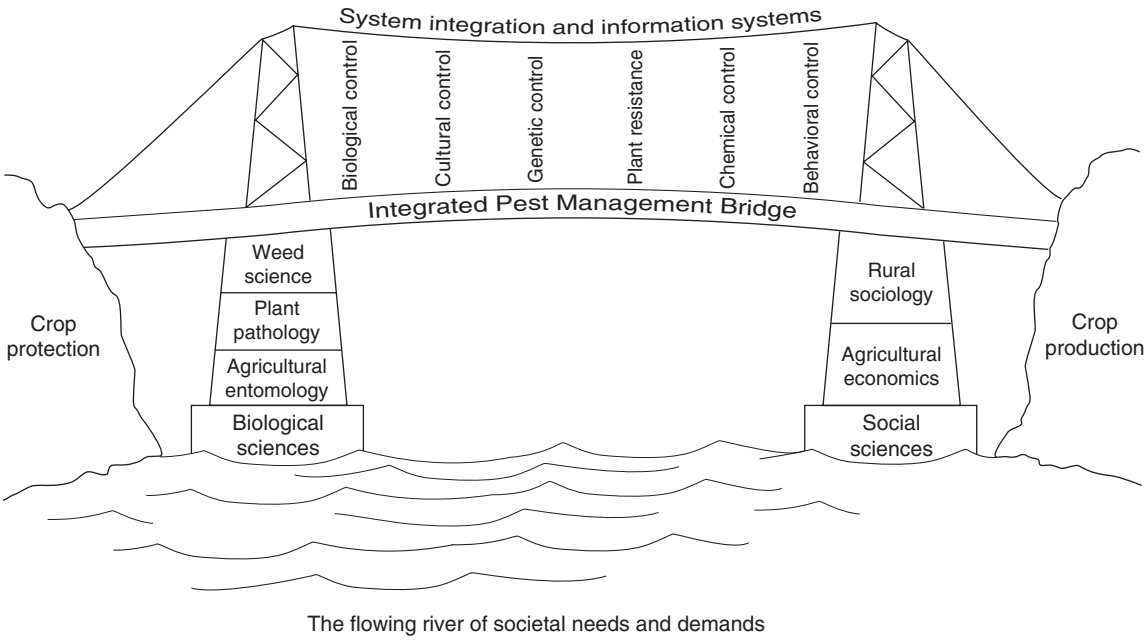

FIGURE 1 A bridge metaphor: agricultural entomology is conceived as one of the main pillars, together with plant pathology and weed science, of supporting the "integrated pest management bridge." The bridge connects two-way "traffic" between crop production and crop protection. The other pillar is provided by the social sciences of economics and sociology. The main tension cables, which are system integration and information systems, hold the vertical lines that together give stability to the bridge; these are the tactical components of IPM. Under the bridge runs the "river" of ever shifting societal needs and demands.

ECONOMIC ENTOMOLOGY

Insects are regarded by some as the main competitors of humans for dominance on the earth. Humans depend on insects for pollination of many crops, for production of honey and silk, for the decomposition of organic matter and the recycling of carbon, and for many other vital ecological roles. But it is the negative impact of insect pests that has been of greatest concern to humans. There are no reliable estimates of aggregate losses caused by insects as vectors of pathogens and parasites of humans and domestic animals, as agents causing direct damage to dwellings and other human-made structures, and as pests of crop plants and farm animals, but the amounts run to probably hundreds of billions of dollars annually. Losses caused by insects and vertebrate pests worldwide in the production of only eight principal food and cash crops (barley, coffee, cotton, maize, potato, rice, soybean, and wheat) between 1988 and 1990 have been estimated at $90.5 billion.

In the late 1800s and early 1900s, entomology became established in many academic and research institutions as a discipline equal in rank with botany and zoology. The diversity of insects and their economic importance was the justification for ranking the study of a class of animals (Insecta) as being equivalent to the study of two kingdoms of organisms (plants and animals other than insects). Through the first half of the twentieth century, there was a schism between basic and applied (or economic) entomology. Since then, common use of the expression "economic entomology" has declined, being replaced by designations of its principal branches, such as agricultural entomology, forest entomology, urban entomology, and medical and veterinary entomology. A detailed historical account is beyond the scope of this article, but Table I provides a chronology of some landmarks in the development of agricultural entomology through the ages.

The realm of agricultural entomology includes all basic studies of beneficial and pest insects associated with agricultural crops and farm animals. This article deals mainly with crops, but the general principles and concepts are equally applicable to farm animals. The starting point of such studies is a correct identification of the insect species, in accordance with the science known as biosystematics.

BIOSYSTEMATICS

Scientific nomenclature is a powerful tool for obtaining information about the basic biology of closely related species within a genus. When systematic studies have been extended beyond the naming of species (taxonomy) and contain detailed information on geographic distribution, host records, and biology of one or more species in a genus, it is often possible to extrapolate the information to other closely related species of that genus. Although details of the biology must be ascertained for each individual species, biosystematics offers a blueprint to follow when dealing with a new pest. For example, the genus *Cerotoma* (Coleoptera: Chrysomelidae) contains 10 to 12 species distributed from southern Brazil to the northeastern United States. All seem to be associated with herbaceous plants in the family Fabaceae (bean family). The biology of two of the species, *C. trifurcata* in North America and *C. arcuata* in South America (Fig. 2), has been studied extensively. Based

TABLE I Some Landmarks in the Historical Development of Agricultural Entomology[a]

Significant events	Years ago from 2000	Date
Beginnings of agriculture	10,000	8000 B.C.E.
First records of insecticide use	4,500	2500 B.C.E.
First descriptions of insect pests	3,500	1500 B.C.E.
Soaps used to control insects in China	900	1100
Beginning of scientific nomenclature—10th edition of Linnaeus, *Systema Naturae*	242	1758
Burgeoning descriptions of insects	100–200	18th and 19th centuries
First record of plant resistance to an insect	169	1831
Charles Darwin and Alfred Wallace jointly present paper on the theory of evolution	142	1858
First successful case of biological control: the cottony cushion scale, on citrus, in California, by the vedalia beetle	112	1888
First record of widespread damage of cotton in Texas by the cotton boll weevil	106	1894
First record of an insect resistant to an insecticide	86	1914
First edition of C. L. Metcalf and W. P. Flint's *Destructive and Useful Insects*	72	1928
Discovery of DDT and beginning of the insecticide era	61	1939
First report of insect resistance to DDT	54	1946
Term "pheromone" coined by P. Karlson and P. Butenandt, who identified first such substance in the silkworm moth	45	1959
First edition of Rachel Carlson's *Silent Spring*	48	1962
Expression "integrated pest management" first appears in the press	32	1968
Rapid development of molecular biology	20	1980s
Release of Bt transgenic varieties of cotton, corn, and potato	5	1990s

[a] Based in part on Norris *et al.* (2003).

on information for these two species, it is possible to infer that the other species in the genus share at least some of the following features: eggs are laid in the soil adjacent to growing leguminous plants; larvae feed on nitrogen-fixing root nodules and pupate in soil inside pupal cases; first-generation adults emerge when seedlings emerge, and second-generation adults emerge when plants are in full vegetative growth, feeding first on foliage and, later on, switching to feeding on developing pods. The biosystematic information on the genus allows students of agricultural entomology in South, Central, or North America to understand, at least in general terms, the role of any other species of *Cerotoma* within their particular agroecosystem.

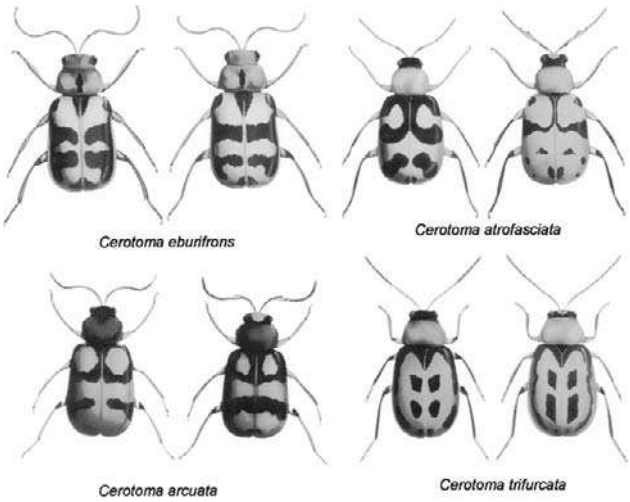

FIGURE 2 Morphological diversity and biological similarities in the genus *Cerotoma*: four of the dozen known species are illustrated by male and female specimens. The species are clearly distinguishable by morphological characters, but they have similar life histories and behaviors. (From unpublished drawings by J. Sherrod, Illinois Natural History Survey.)

The flip side of this notion is recognition that closely related and morphologically nearly undistinguishable (sibling) species may have many important biological differences. Examples of the critical need for reliable biosystematics studies are found in the biological control literature. The present account is based on studies conducted by Paul DeBach, one of the leading biological control specialists of the twentieth century. The California red scale, *Aonidella aurantii*, is a serious pest of citrus in California and other citrus-producing areas of the world. Biological control of the red scale in California had a long history of confusion and missed opportunities because of misidentification of its parasitoids. The red scale parasitoid *Aphytis chysomphali* had been known to occur in California and was not considered to be a very effective control agent. When entomologists discovered parasitized scales during foreign exploration, the parasitoids were misidentified as *A. chrysomphali* and therefore were not imported into California. It was later discovered that the parasitoids were in fact two different species, *Aphytis lingnanensis* and *A. melinus*, both more efficient natural enemies of the California red scale than *A. chrysomphali*. Now *A. lignanensis* and *A. melinus* are the principal red scale parasitoids in California. Further biosystematics studies have shown that what was once thought to be single species, *A. chrysomphali*, parasitic on the California red scale in the Orient and elsewhere, and accidentally established in California, is in fact a complex including at least seven species having different biological adaptations but nearly indistinguishable morphologically.

Knowledge of the name of a species, however, is not an indication of its true potential economic impact or pest status. A next important phase in agricultural entomology is, therefore, the assessment of benefits or losses caused by that species.

PEST IMPACT ASSESSMENT

The mere occurrence of an insect species in association with a crop or a farm animal does not necessarily mean that the species is a pest of that crop or animal. To be a pest it must cause economic losses. The assessment of economic losses from pests is the subject of studies conducted under conditions that match as closely as possible the conditions under which the crop is grown commercially or the animals are raised. Much of the methodology used in crop loss assessment has been established under the sponsorship of the Food and Agriculture Organization (FAO) of the United Nations as a means of prioritizing budget allocations and research efforts. Key data for these studies relate to the determination of the yield potential of a crop. The genetic makeup of a crop variety determines its maximum yield in the absence of adverse environmental factors. This is known as the attainable yield. To determine the attainable yield, the crop is grown under nearly ideal conditions; the actual yield is what occurs when the crop is grown under normal farming conditions. The difference between attainable and actual yields is a measure of crop loss (Fig. 3).

To assess crop losses and attribute the losses to a specific cause (e.g., the attack of a pest) requires setting up experiments to isolate the effect of the pest from all other constraints. Methodologies vary with pest category—whether the pests are insects, vertebrates, plant pathogens, or weeds, for example. The quantitative relationship between crop losses and pest population levels is the basis for computing the economic injury level for the pest. The economic injury level is a fundamental concept in IPM.

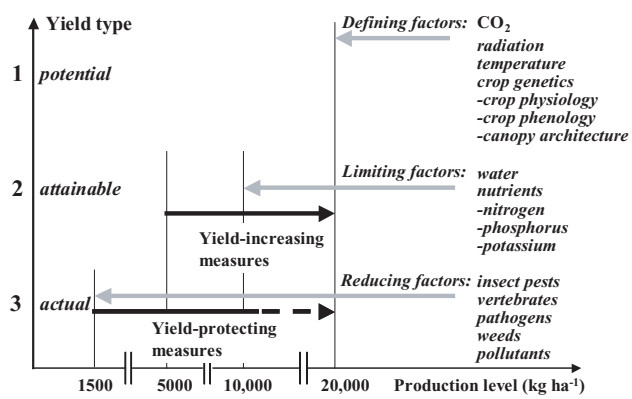

FIGURE 3 Factors impacting the yield potential of a generic crop. (Adapted from information on a Web site originated at IMI/University of Miami, Summer Institute.)

LIFE HISTORY AND HABITS

Once the identity and pest status of a species have been well established, it becomes essential to extend the informational base on the life history and habits of the species to the conditions under which the crop is grown. Economically important life history traits include information on developmental threshold temperatures and temperature-dependent developmental rates. These data are used in modeling the phenology of the pest. Other essential studies include the orientation, feeding, host selection, and sexual behavior of the species. Many of these studies provide the foundation for strategic planning in IPM and for the development of target-specific control tactics. For example, the study of sexual behavior involves the definition of the role of pheromones in mating and the identification of those pheromones. These, in turn, may be used for monitoring pest incidence and abundance or in mating disruption, both valuable components of IPM systems for many crops. The study of host selection behavior often leads to the identification of kairomones, equally important in IPM development.

PHENOLOGY

The life cycle of different insect species varies greatly, although all insects undergo the basic stages of development from egg to reproductive adult (or imago). Depending on the length of the life cycle, there is considerable variation in the number of generations per year, a phenomenon called voltinism. A univoltine species has one generation per year; a multivoltine species may have many generations per year. The range of variation in the Insecta is evident when one considers that the 17-year periodical cicada has one generation every 17 years, whereas whiteflies or mosquitoes may complete a generation in about 21 days. Under temperate climate conditions, generations often are discrete, but under warmer subtropical conditions they frequently overlap. The definition of temporal periodicity in an organism's developmental cycle is called phenology. The relationship between the phenology of the crop and the phenologies of its various pests is of interest in agricultural entomology. Figure 4 shows an example of such a relationship for soybean grown under conditions typical for the midwestern United States.

POPULATION AND COMMUNITY ECOLOGY

Population and community level studies are within the scope of insect ecology. Although the species is the focal biological entity for agricultural entomology, for management purposes it is essential to understand population and community level processes. Populations are assemblages of conspecific individuals within a defined geographical area (e.g., a crop field, a river valley, a mountain chain). Many insects have a large reproductive capacity. As calculated by Borror, Triplehorn, and Johnson, a pair of fruit flies (*Drosophila*), for

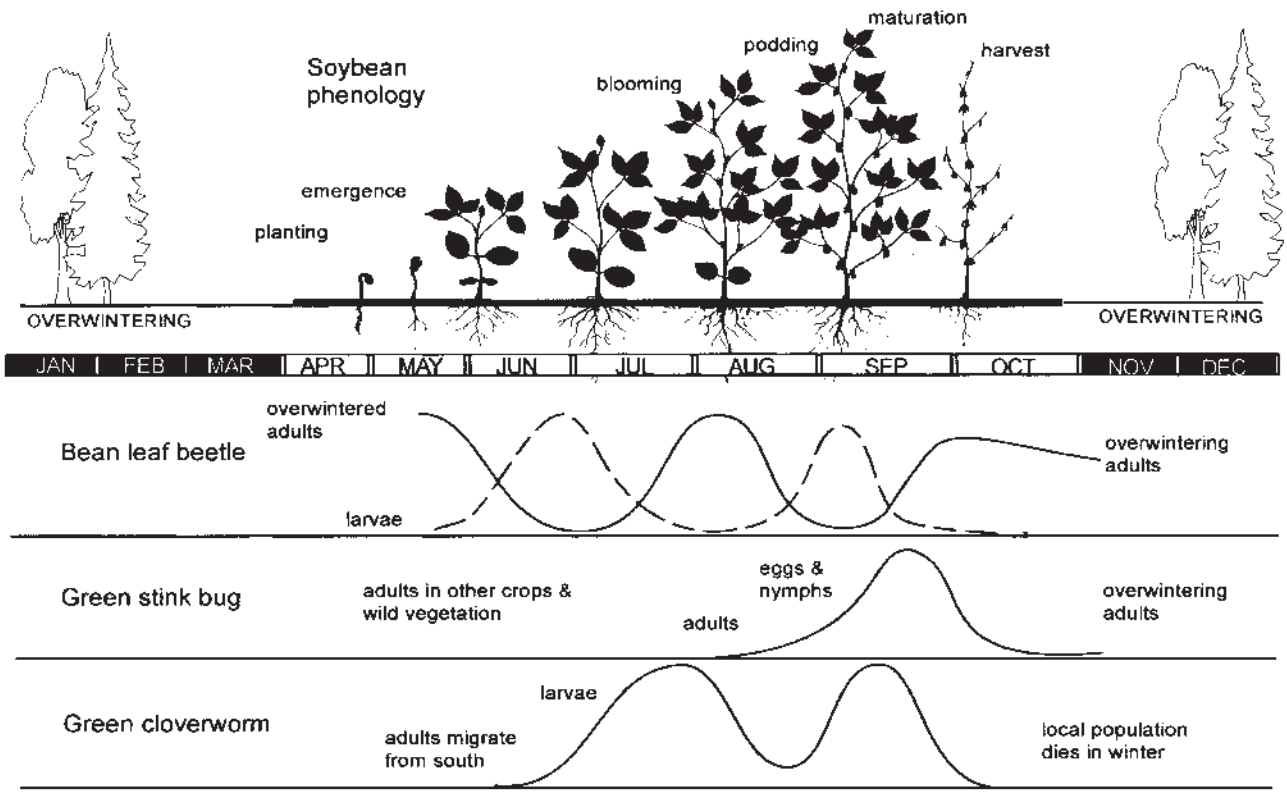

FIGURE 4 Crop phenology and pest phenology: relationship between the phenology of soybean in the midwestern United States and three of its most common insect pests, the bean leaf beetle, *C. trifurcata* (Coleoptera: Chrysomelidae); the green stink bug, *Acrosternum hilare* (Hemiptera: Pentatomidae); and the green cloverworm, *Hypena scabra* (Lepidoptera: Noctuidae).

example, produces 100 viable eggs, half of which yield females that in turn will lay 100 eggs and so on for 25 possible generations in one year; by the end of the year, the 25th generation would contain 1.192×10^{41} flies, which, if packed tightly together, 60,000 to a liter, would form a ball of flies 155 million km in diameter or a ball extending approximately from the earth to the sun. Obviously, such unlimited population growth does not occur in nature. Normally, populations are regulated by the combined actions of both physical (or abiotic) and biological (or biotic) factors of the environment. An understanding of the mortality factors that help regulate insect populations is one of the most active areas of research in agricultural entomology.

The set of species coexisting in an area and interacting to varying degrees form what is known as an ecological community. In a crop community, the crop plants and the weeds that persist within the crop field or grow along the borders are the primary producers. The animals within the crop community maintain dynamic trophic relationships: some feed on living plants, others on the decaying plants, and still others on animals. Those that feed on the plants are the herbivores, or primary consumers. Pests are primary consumers on the crop plants. Parasitoids and predators are the secondary consumers. Those that feed on the pests are beneficial natural enemies. Finally, decomposers and detritivores feed on decaying organic matter. All biotic components of the community are interconnected by "food webs." An understanding of food webs and trophic interactions in crop communities is important because it provides a basis for interpreting the nature of disturbances in crop ecosystems. Disturbances in trophic relations may lead to outbreaks of pest organisms and the need for control actions.

LINKS TO IPM SYSTEMS DEVELOPMENT

With the advent of integrated pest management and its success in the last third of the twentieth century, it has become difficult to separate agricultural entomology from IPM. In entomology, the two fields of endeavor are inextricably interconnected. A reliable database of biological information provides the means to design and develop IPM strategies. For example, there is growing interest in methods of enhancing biological control through habitat management. The technique requires information on source–sink relationships among pests and natural enemies across crop plants, neighboring crops, natural vegetation, and especially managed vegetation in the form of cover crops and field hedges. Theoretically, diversification of the crop ecosystem leads to an increase in natural enemies and to greater stability of the system. The complexity of interactions, however, makes it difficult to interpret conflicting results of experiments designed to test working hypotheses. The analysis of

within-field and interfield movement, the host selection behavior of phytophagous and entomophagous insects, multitrophic interactions among community members, and the dynamics of populations, all under the scope of agricultural entomology, are only a few of the many components of the knowledge base necessary to develop advanced IPM systems.

The advent of the World Wide Web has had a major influence on accessibility to basic information on agricultural entomology. Most major agricultural research centers have developed Web pages that organize information and make it available to students worldwide. More importantly, the dynamic nature of the Web offers the opportunity to provide weather-driven modeling capabilities that greatly increase the scope and applicability of studies about the phenology and population dynamics of major pest organisms. Two sites that offer such capabilities are http://www.orst.edu/Dept/IPPC/wea/ and http://www.ipm.ucdavis.edu/PHENOLOGY/models.html.

Entomologists in the late 1800s and early 1900s studied the biology of insect pests in great detail. Articles and monographs published during that period remain valuable sources of information. These early entomologists recognized that deep knowledge of the life history of an insect and its habits could provide insights useful for the control of agricultural and other pests. The advent of organosynthetic insecticides in the mid-1940s created the illusion that pest problems now could be solved forever. Many entomologists redirected their efforts to testing new chemicals and neglected basic insect biology studies. The failure of insecticides to eradicate pests and the environmental problems engendered by the misuse of these chemicals led to the advent of IPM. For IPM to succeed, entomologists have had to return to the basics and again refocus their efforts on the study of insect biology. Agricultural entomology has come full circle as new generations of entomologists endeavor to refine knowledge of the group of animals that remain humans' most serious competitors.

See Also the Following Articles

Biological Control • History of Entomology • Integrated Pest Management • Phytophagous Insects • Plant–Insect Interactions • Population Ecology

Further Reading

Borror, D. J., Triplehorn, C. A., and Johnson, N. F. (1992). "An Introduction to the Study of Insects." Saunders College Publishers and Harcourt Brace College Publishers, Philadelphia and Fort Worth.

DeBach, P. (1964). "Biological Control of Insect Pests and Weeds." Reinhold, New York.

Evans, H. E. (1984). "Insect Biology: A Textbook of Entomology." Addison-Wesley, Reading, MA.

Huffaker, C. B. and Gutierrez, A. P. (1999). "Ecological Entomology." Wiley, New York

Jones, D. P. (1973). Agricultural entomology. *In* "History of Entomology." (R. F. Smith, T. E. Mittler, and C. N. Smith, eds.), pp. 307–332. Annual Reviews, Palo Alto, CA.

Norris, R. F., Caswell-Chen, E. P., and Kogan, M. (2003). "Concepts in Integrated Pest Management." Prentice Hall, Upper Saddle River, NJ.

Oerke, E. C. (1994). "Crop Production and Crop Protection: Estimated Losses in Major Food and Cash Crops." Elsevier, Amsterdam and New York.

Price, P. W. (1997). "Insect Ecology." Wiley, New York.

Schowalter, T. D. (2000). "Insect Ecology: An Ecosystem Approach." Academic Press, San Diego, CA.

Alderfly

see *Megaloptera*

Amber

George Poinar Jr.
Oregon State University

Amber is a fossilized resin ranging from several million to 300 million years of age. This material is a gold mine for the entomologist because it contains a variety of insects preserved in pristine, three-dimensional condition. Fossils in amber provide evidence of lineages dating back millions of years (Table I). External features are preserved so well that taxonomists can make detailed comparisons with living taxa to follow evolutionary development of genera and even species. Amber has a melting point between 200 and 380°C, a hardness of 2 to 3 on the Moh's scale, and a surface that is insoluble to organic solvents. Aside from providing direct evidence of an insect taxon at a particular time and place, amber insects give clues to past distributions and phylogeny, as well as indirect evidence of plants and vertebrates and the establishment of symbiotic associations, and clues for reconstructing ancient landscapes.

TABLE I Significant Amber Deposits in the World

Deposit	Location	Approximate age (million years)
Baltic	Northern Europe	40
Burmese	Burma (Myanmar)	100
Canadian	Alberta, Manitoba	70–80
Chinese	Fushun Province	40–53
Dominican	Dominican Republic	15–45
Hat Creek	British Columbia, Canada	50–55
Lebanese	Middle East	130–135
Mexican	Chiapas	22–26
New Jersey	Northeastern United States	65–95
Siberian (Taimyr)	Russian arctic	78–115
Spanish	Alava, Basque country	100–115

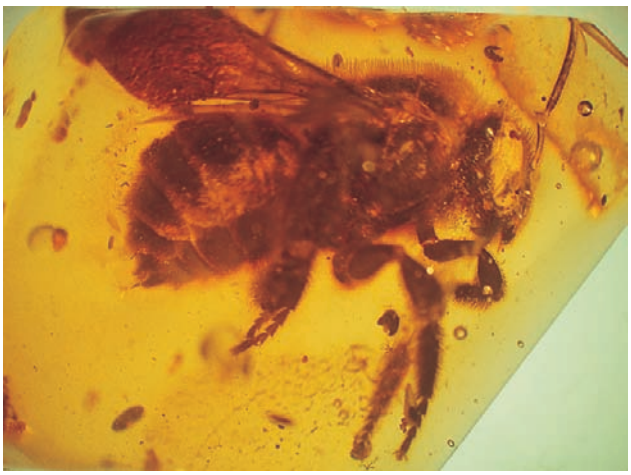

FIGURE 1 Origins of the honey bee lineage are provided by this primitive bee, which possesses characters of both modern *Apis* and extinct apids. Its presence in Baltic amber suggests a European origin of *Apis,* thus challenging the current view that honey bees originated in Asia.

USE OF AMBER IN TRACING INSECT LINEAGES

As a result of the excellent preservation of amber insects, specific genera (the majority, if not all, amber insects are extinct at the species level) can be recognized and compared with modern ones. In this way, lineages can be traced back tens of millions of years. An example is a small parasitic wasp of the genus *Aphelopus* (Hymenoptera: Dryinidae) trapped in Lebanese amber. This genus is still extant, and the fossil demonstrates a lineage that has survived for 130 to 135 million years.

The origin of genera can also be obtained from amber insects owing to their high degree of preservation. A recent example from Baltic amber, which contains a variety of bees, deals with the origin of the common honey bee. This fossil contains basic features characteristic of the genus *Apis* as we know it today, including pollen-collecting apparati on the hind legs and a barbed stinger (Fig. 1). This appears to be one of the most primitive *Apis* ever discovered, thus indicating a time (40 mya) and place (northern Europe) for the origin of the honey bee.

PROVIDING INDIRECT EVIDENCE OF OTHER ORGANISMS

There are size and habitat limitations to the types of organism that can be trapped in amber. For example, many plants would not normally leave flowers or leaves in the resin, and when they did, the remains would likely be difficult to identify. Vertebrates might leave hairs, feathers, or scales but these structures would also be difficult to identify. However, arthropods that are specific to certain hosts (e.g., ticks and mammals) can provide clues to other organisms that existed at that time. This use of fossils relies heavily on the principle of behavioral fixity, which asserts that, at least at the generic level, the behavior of a fossil organism would have been similar to that of its present-day descendants.

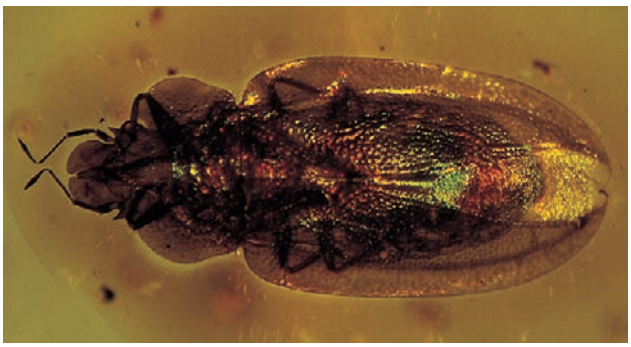

FIGURE 2 The unique morphological features (smooth, flattened body) of this fossil palm bug (*Paleodoris lattini*) in Dominican amber not only characterize it systematically but also provide clues to its lifestyle of living in confined spaces between the unopened fronds of pinnately leafed palms.

Many insects form specific associations with plants. Such associations can often be deduced by the morphological features of the insect (functional morphology). One extremely flattened hemipteran in Dominican amber (Fig. 2) that was identified as a palm bug displayed characters similar to those of an existing species in the same subfamily. The extant species lives between the closed leaves of royal palms (*Roystonea* spp.) in Cuba. This fossil provided indirect evidence that pinnately leafed palms, quite likely an extinct species of *Roystonea,* existed in the original amber forest. Other plant-specific insects, such as fig wasps and palm bruchids, provide evidence of figs and palms in the original ecosystem.

Insects that require a blood meal to complete their development can also be used as indirect evidence of a vertebrate group. Evidence of birds in the original Dominican amber forest is implied by the presence of a female *Anopheles* mosquito in amber because extant species of this subgenus normally attack birds. The presence of other vertebrate groups is implied by fleas (Siphonaptera), horseflies (Diptera: Tabanidae), biting midges (Diptera: Ceratopogonidae), and other bloodsucking arthropods such as ticks.

PROVIDING INDIRECT EVIDENCE OF SPECIFIC HABITATS

Amber insects can provide evidence of specific habitats. Diving beetles (Coleoptera: Dytiscidae), caddisflies (Trichoptera), and damsel flies (Odonata) all provide evidence of aquatic habitats. The *Anopheles* mosquito belongs to a group that normally oviposits in ground pools. Other insects can provide evidence of phytotelmata (standing water in plant parts), wood, moss, bark, and detritus.

PALEOSYMBIOSIS

Because of the sudden death of captured organisms in amber, symbiotic associations may be preserved in a manner unlikely to occur with other types of preservation. Also, the fine details of preservation may reveal morphological features characteristic

FIGURE 3 Documentation of paleophoresis is provided by a pseudoscorpion grasping the tip of the abdomen of a platypodid beetle in Dominican amber. Similar rider–carrier associations occur today, suggesting that this behavior is mandatory for survival of the pseudoscorpion.

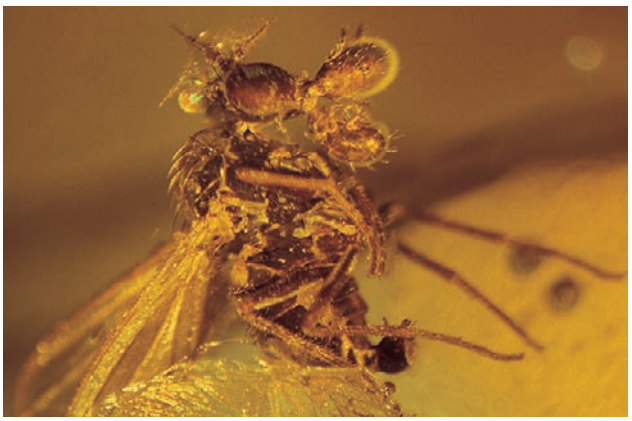

FIGURE 4 Paleoectoparasitism is shown by two thrombidid mites attached to the mouthparts of a long-legged fly (Diptera: Dolichopodidae) in Baltic amber.

of symbiotic associations. Cases of paleosymbiosis in amber include inquilinism, commensalism, mutualism, and parasitism.

Paleoinquilinism involves two or more extinct organisms living in the same niche but neither benefiting nor harming each other. Numerous insects form inquilinistic associations under tree bark, and many pieces of amber contain flies and beetles common to this habitat.

Phoresis (one organism transported on the body of another organism) is probably the most typical type of paleocommensalism in amber. This usually involves mites and pseudoscorpions being carried by insects. The arachnid benefits by being conveyed to a new environment, where the food supply is likely to be better than the last one. The carrier generally is not harmed and only serves as a transporting agent. An example of this category in Dominican amber consists of pseudoscorpions being carried by platypodid beetles (Coleoptera: Platypodidae) (Fig. 3). The method of attachment of the pseudoscorpion to the beetle was the same then as it is today. In fact, these ancient records lead scientists to believe that such behavior is mandatory for the survival of the pseudoscorpions that live in beetle tunnels and require effective dispersal mechanisms for survival.

In paleomutualism, both organisms benefit and neither is harmed. Amber bees carrying pollen provide evidence of insect–plant mutualism in which the bee obtains a food supply and the plant is pollinated. An example of insect–insect mutualism is demonstrated by a rare fossil riodinid butterfly larva in Dominican amber. Specialized morphological features of this *Theope* caterpillar indicative of a symbiotic association are balloon setae and vibratory papillae in the neck area, and tentacle nectary organ openings on the eighth abdominal tergite. Extant caterpillars in this genus have similar features and are associated with ants. The tentacle nectary organs provide nourishment for the ants, whereas the vibratory papillae (which beat against the head capsule and make an audible sound) and balloon setae (which emit a chemical signal) are used to attract ants when the caterpillar is threatened by an invertebrate predator or parasite. This fascinating association between butterfly larvae and ants was established at least 20 mya.

Paleoparasitism is very difficult to verify in the fossil record. There are many records of amber insects (especially wasps and flies) whose descendants today are parasitic on a wide range of organisms, but to discover an actual host–parasitic association is quite rare.

Paleoectoparasitism is the most obvious of all parasitic associations found in amber. The ectoparasite is often still attached to its host, and systematic studies can be conducted on both organisms. In amber, ectoparasites are usually parasitic mites, such as the larvae of Thrombididae attached to the mouthparts of a fly in Baltic amber (Fig. 4). These larval mites were feeding on the host's hemolymph, and their mouthparts are still in place. After molting to the nymphal stage, the parasites would leave the fly and become free-living predators. Large infestations could kill the host. These mites are not to be confused with phoretic ones, which are simply carried around by insects.

Paleoendoparasitism is extremely difficult to verify because internal parasites are rarely preserved as fossils. However, some parasites attempt to leave their hosts when they encounter resin. Mermithid nematodes (Mermithidae: Nematoda) and hairworms (Nematomorpha) that have nearly completed their development and are almost ready to emerge from their host will often reveal their presence (Fig. 5). Under normal conditions, they would enter soil or water and initiate a free-living existence.

BIOGEOGRAPHICAL STUDIES

In many instances amber insects provide evidence of a more extensive distribution in the past for various insect genera and families as well as indicating a warmer climatic regime in many parts of the world. Perhaps the most spectacular examples of this phenomenon are insects discovered in amber sites located far from their descendants' current habitat. Examples from

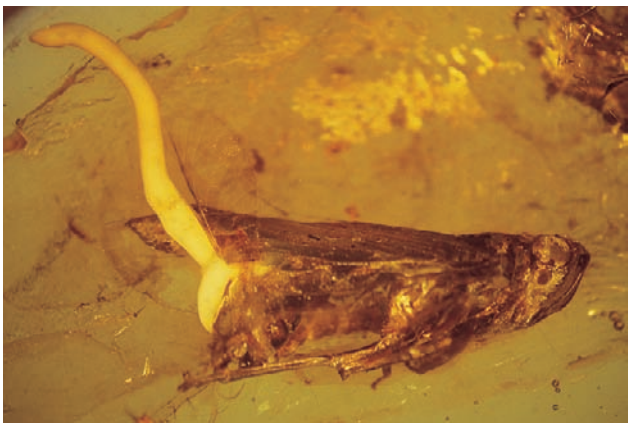

FIGURE 5 Paleoendoparasitism in amber is exemplified by a mermithid nematode (Nematoda: Mermithidae) emerging from the body of a planthopper (Homoptera: Fulgoroidea) in Baltic amber. Such records set minimum dates for the establishment of host–parasite associations.

Dominican amber include *Mastotermes* termites (Isoptera: Mastotermitidae) and *Leptomyrmex* ants (Hymenoptera: Formicidae) that obviously were part of the insect fauna some millions of years ago in the Caribbean but occur nowhere in the New World today. Both genera are represented today by a single relict species in the North and East Australian Region.

A North American example is the presence of the tropical arboreal ants of the genus *Technomyrmex* in Hat Creek amber in British Columbia, Canada, living 50 mya, hundreds of kilometers north of their present-day range. These tropical ants in Eocene Hat Creek amber provide evidence that the climate in that region of the world shifted from tropical to temperate. Other examples of past distributions involve the palm bug shown in Fig. 2, which has no present-day descendants in the Dominican Republic, with only a single living Cuban species in the subfamily. Similarly, there are no members of the genus *Theope* in the Dominican Republic or the Greater Antilles today, all living representatives being restricted to Mexico, and Central and South America. Further evidence of climatic shifts over time are clear with many of the Baltic amber insects, many of whose descendants occur in the Old World tropics today. The primitive honey bee shown in Fig. 1 evolved under subtropical conditions that characterized most of northern Europe in the Eocene. Thus, it is not surprising that most of the species and varieties of the genus *Apis* live only under tropical conditions today.

RECONSTRUCTING ANCIENT LANDSCAPES

Every amber fossil tells a story and is a piece of a jigsaw puzzle that can be used to reconstruct the natural environment at the time the amber was being produced. The challenges are to identify the inclusions, determine their biology and ecology by researching the habits of their extant descendants, and then make inferences regarding the original environment. There will always be gaps in the puzzle because there are many life-forms that are too large to become entrapped in amber or have a lifestyle that does not normally bring them into contact with the sticky resin. However, the habitat that existed in that ancient world can, in large part, be reconstructed by studying select insects that can be typified as phytophagous, soil-loving, bark inhabitants, or parasites, and identifying the associated predators, vertebrates, and special habitats.

See Also the Following Articles

Biogeographical Patterns • Fossil Record

Futher Reading

Boucot, A. J. (1990). "Evolutionary Paleobiology of Behavior and Coevolution." Elsevier, Amsterdam.

Poinar, G. O. Jr., (1992). "Life in Amber." Stanford University Press, Stanford, CA.

Poinar, G. O., Jr., and Milki, R. (2001). "Lebanese Amber. The Oldest Insect Ecosystem in Fossilized Resin." Oregon State University Press, Corvallis.

Poinar, G. O., Jr., and Poinar, R. (1994). "The Quest for Life in Amber." Addison-Wesley, Reading, MA.

Poinar, G. O., Jr., and Poinar, R. (1999). "The Amber Forest." Princeton University Press, Princeton, NJ.

Ross, A. (1998). "Amber." Harvard University Press, Cambridge, MA.

Weitschat, W., and Wichard, W. (2002). "Atlas of Plants and Animals in Baltic Amber." Pfeil, Munich.

Anatomy: Head, Thorax, Abdomen, and Genitalia

David H. Headrick
California Polytechnic State University

Gordon Gordh
U.S. Department of Agriculture

Anatomy is a subdiscipline of morphology concerned with naming and describing the structure of organisms based on gross observation, dissection, and microscopical examination. Morphology and anatomy are not synonyms. Morphology is concerned with the form and function of anatomical structure; because anatomy is an expression of organic evolution, morphology seeks to investigate possible explanations for organic diversification observed in nature. Before 1940 insect morphology focused on naming and describing anatomical structure. The need for this activity has not diminished, as much about insect anatomy remains to be revealed, described, and understood. This article focuses on the anatomical structures of the three major tagmata of the insect body: head, thorax, and abdomen, and on the external genitalia. A hypothetical ground plan for major structures is given, followed by themes in anatomical variation based on adaptation observed in the Insecta.

CONTEXT OF ANATOMICAL STUDY

Terms of Orientation and Conventions

Terms to describe orientation are not intuitive for insects. Most orientation terms are derived from the study of the human body—a body that stands upright—and their application to insects causes confusion. Some standard terms used with insects include anterior (in front), posterior (behind), dorsal (above), ventral (below), medial (middle), and lateral (side). Anatomical description usually follows in the same order, hence, we begin our discussion with the head, move on to the thorax and then the abdomen, and finish with the genitalia. Description of the relative placement of anatomical features can be cumbersome, but they are critical elements in the study of anatomical structure because relative position is one of the three basic tenets of homology, including size and shape, and embryology.

Measures of Success

The design of the insect body can be described as successful for many reasons: there are millions of species, they range in size over four orders of magnitude, their extensiveness of terrestrial and aquatic habitat exploitation (the diversity of resources), and once a successful form has been developed, there appears to be relatively little change over evolutionary time (Fig. 1). The basic insect design allows for adaptation to a variety of environmental requirements. The success of the design is rooted in the nature of the main material used for its construction.

The Building Material

When we look at an insect, it is the integument that we see. Structurally, the integument is a multiple-layered, composite organ that defines body shape, size, and color. The ultrastructure of the integument is composed of living cells and the secretory products of those cells. Each layer is of a different thickness and chemical composition, and each displays physical properties different from those of the surrounding layers. Perhaps more importantly, the integument also is the organ with the greatest diversity of structure and function.

FIGURE 1 Fossil insects are easily recognizable today, indicating an early establishment of a successful design. Left to right: *Heplagenes* (Late Jurassic 150 mya, Liaoning, China); cricket (Eocene, 50 mya, Green River formation, Utah); fulgorid (Eocene, 50 mya, Green River formation, Utah).

There are two common misconceptions about the integument. First, some believe chitin is responsible for integument hardness. Actually, there is proportionally more chitin found in the soft and flexible membranous parts of the integument than in the hard, sclerotized plates. Integument hardness is attributed to an increased number of cross-linkages between protein chains contained in the integument layers. Second, some believe that the integument is rigid and that growth is incremental and limited to expansion during molting; yet some endopterygote insects are able to grow continuously between molts.

The integument determines the shape of the insect body and its appendages. One of the most captivating features of insects is their seemingly infinite variation in body shapes—everything from a simple bag (Hymenoptera grub) to a mimic of orchid flowers (Mantidae). Similarly, appendage shape is exceedingly plastic. Terms such as "pectinate," "flabbate," and "filiform" are among more than 30 terms taxonomists have proposed to describe antennal shapes. Leg shapes are similarly highly variable and express functional modifications. Among these shapes are "cursorial," "gressorial," "raptorial," "fossorial," and "scansorial." Again, these modifications of shape reflect the function of structure. Finally, wing shapes are highly variable among insects and are determined by body size and shape as well as by aerodynamic considerations.

Tagmata

Most people recognize the three tagmata—head, thorax, and abdomen—as characteristic of insects. The way they appear is rooted in a division of responsibilities. The head is for orientation, ingestion, and cognitive process; the thorax for locomotion; and the abdomen for digestion and reproduction. But even casual observations reveal further divisions of these body regions.

Segmentation of Tagmata

Two types of segmentation are evident among arthropods, primary and secondary. Primary segmentation is characteristic of soft-bodied organisms such as larval holometabolans. The body wall in these organisms is punctuated by grooves or rings that surround the anterior and posterior margin of each somite. These rings represent intersegmental lines of the body wall and define the limits of each somite. Internally, the grooves coincide with the lines of attachment of the primary longitudinal muscles. From a functional standpoint, this intrasegmental, longitudinal musculature permits flexibility and enables the body to move from side to side.

More complex plans of body organization exhibit structural modifications. Secondary segmentation is characteristic of hard-bodied arthropods, including adult and nymphal insects. Secondary body segmentation is an evolutionarily

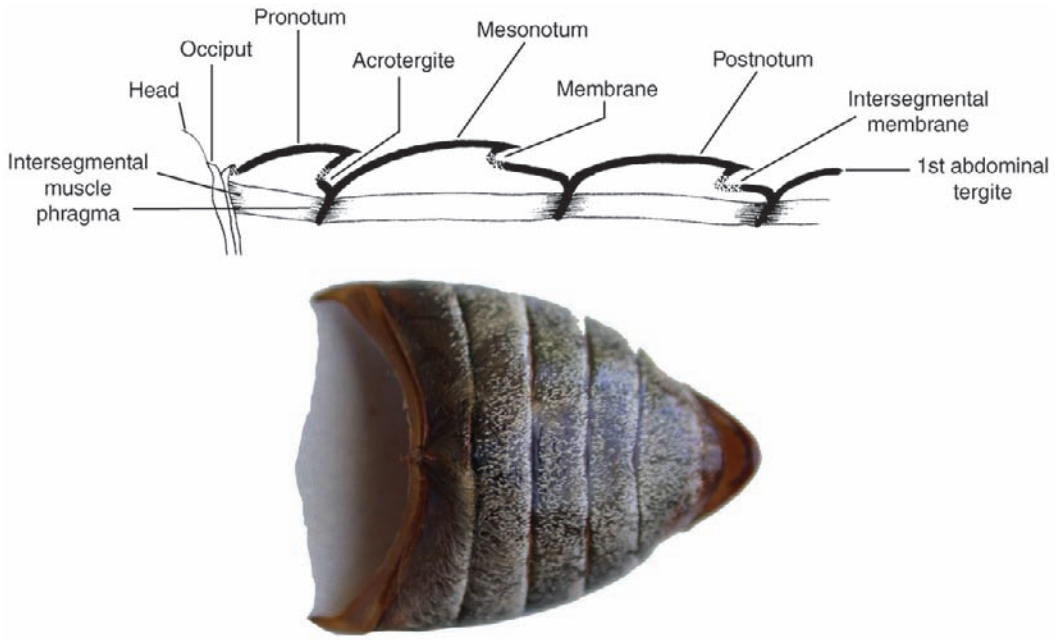

FIGURE 2 Secondary segmentation. Top: diagram of sagittal section of dorsal sclerites of thorax. Bottom: ventral view of abdominal sternites showing overlap due to secondary segmentation (Coleoptera: Scarabaeidae).

derived anatomical feature. The musculature we see in secondary segmentation is *inter*segmental, or between segments (Fig. 2). The acquisition of secondary segmentation represents a major evolutionary step in the development of the Arthropoda. The soft-bodied arthropod has primary segmentation and muscles that are *intra*segmental, or within each segment. Movement of the body and its parts is relatively simple because the body wall is flexible. However, when the body wall becomes hardened, flexibility is restricted to the articulation between hardened parts or the extension provided by intersegmental membranes. The arthropod is, in a metaphorical sense, clad in a suit of armor; most movement is possible only if soft and flexible membranes are positioned between inflexible (hardened) body parts. Exceptions may be seen in the indirect flight mechanism of pterygote insects.

In all probability secondary segmentation evolved many times, and it probably continues to evolve in response to specific problems confronting insects today. Secondary segmentation is most evident and most readily appreciated in the insect abdomen. It is less apparent in the thorax and almost totally obscured in the head.

Sclerites

The hardening of the body wall contributes significantly to the external features observed in insects. Sclerites are hardened areas of the insect body wall that are consequences of the process of sclerotization. Sclerites, also called "plates," are variable in size and shape. Sclerites do not define anatomical areas and do not reflect a common plan of segmentation. Sclerites develop as *de novo* hardening of membranous areas of the body wall, as *de novo* separations from larger sclerotized areas of the body, and in other ways.

The hardened insect body displays many superficial and internal features that are a consequence of hardening. Understanding the distinction between these conditions and the terms applied to them is critical in understanding insect anatomy and its application in taxonomic identification. These features are of three types. First, sutures (Latin, *sutura* = seam), in the traditional sense of vertebrate anatomists, provide seams that are produced by the union of adjacent sclerotized parts of the body wall. On the insect body, sutures appear as etchings on the surface of the body and form lines of contact between sclerites. Second, sulci (Latin, *sulcus* = furrow) represent any externally visible line formed by the inflection of cuticle. Biomechanically, a sulcus forms a strengthening ridge. In contrast, lines of weakness are cuticular features that are used at molting. Lines of weakness are frequently named as if they were sutures, but they should not be viewed as such. For instance, the ecdysial cleavage line is a line of weakness that is sometimes considered to be synonymous with the epicranial suture. The two features are similar in position and appearance, but structurally they may have been derived from different conditions. Finally, apodemes (Greek, *apo* = away; *demas* = body) are hardened cuticular inflections of the body wall that are usually marked externally by a groove or pit. Structures called apophyses (Greek, *apo* = away; *phyein* = to bring forth) are armlike apodemes. Apodemes have been defined as a hollow invagination or inflection of the cuticle and an apophysis as a solid invagination. Functionally, apodemes strengthen the body wall and serve as a surface for muscle attachment.

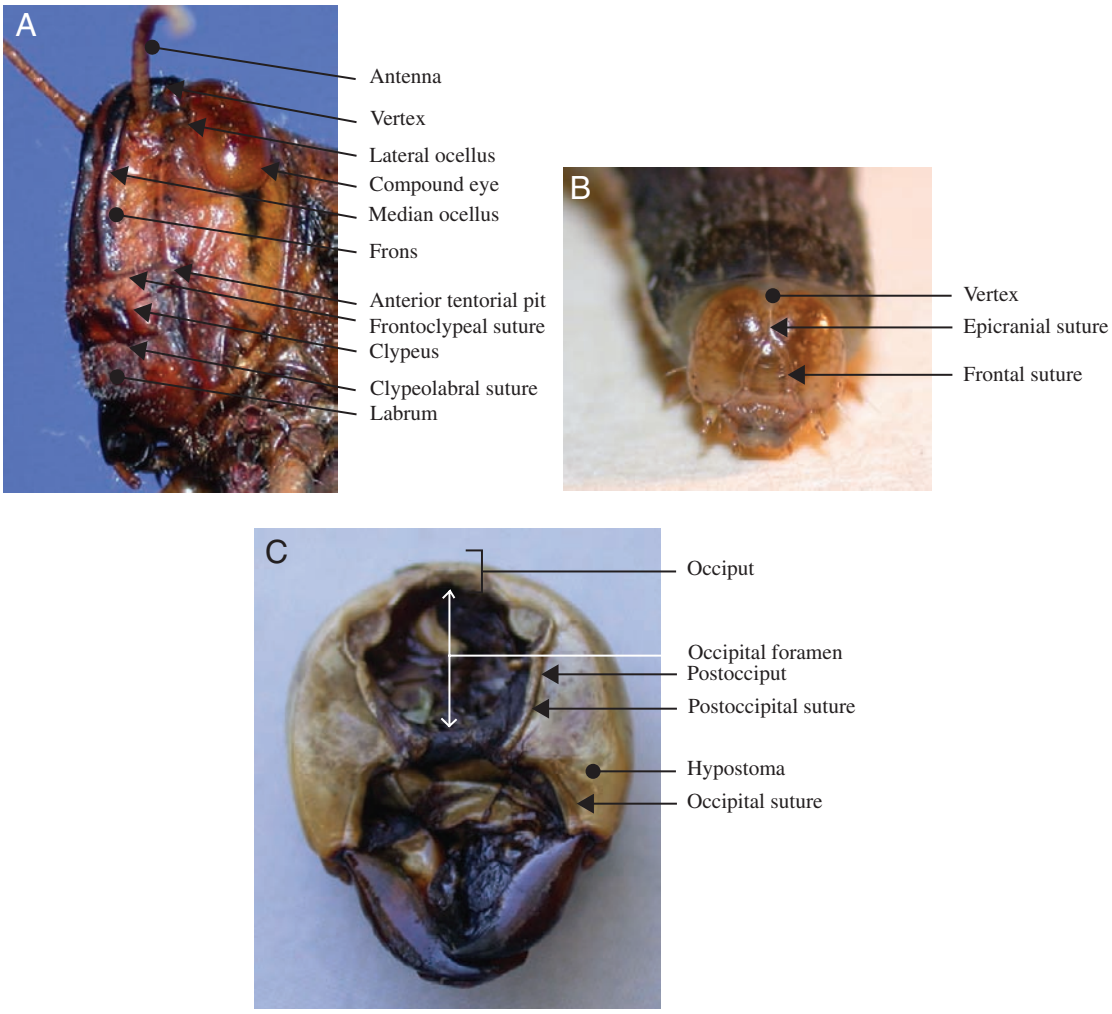

FIGURE 3 (A) Anterior view of the head of a grasshopper (Orthoptera: Acrididae). (B) Larval pterygote head showing epicranial and frontal sutures (Lepidoptera: Noctuidae). (C) Posterior aspect of the head (Orthoptera: Stenopelmatidae).

Sclerites receive different names depending upon the region of the body they are located. Tergites (Latin, *tergum* = back) are sclerites that form a subdivision of the dorsal part of the body wall (tergum). Latrotergites are sclerites that form as a subdivision of the lateral portion of the tergum. Sternites (Latin, *sternum* = breast bone) are sclerites that form as a subdivision of the ventral part of the body wall (sternum), or any of the sclerotic components of the definitive sternum. Pleurites (Greek, *pleura* = side) are sclerites in the pleural region of the body wall that are derived from limb bases.

HEAD

The head is a controversial area for anatomical nomenclature, but it provides some of the best examples of evolutionary trends in anatomy. Most insect morphologists believe that the head of modern insects represents the fusion of several segments that were present in an ancestral condition. However, the number of segments included in the ground plan of the insect head has been a contentious issue among morphologists for more than a century. Any argument that attempts to explain head segmentation must take into account comparative anatomical, embryological, and paleontological evidence, and must examine modern forms of ancestral insects.

Ground Plan of the Pterygote Head

Given the difficulty in homologizing anatomical features of the head, we describe regions associated with landmarks of a ground plan or an idealized hypognathous insect head. In terms of modern insects, the Orthoptera probably come closest to displaying all the important landmark sutures and sclerites that form the head (Fig. 3A).

The vertex (Latin, *vertex* = top; pl., vertices) is the apex or dorsal region of the head between the compound eyes for insects with a hypognathous or opisthognathous head. This definition does not apply to prognathous heads because the primary axis of the head has rotated 90° to become parallel

to the primary axis of the body. The vertex is the area in which ocelli are usually located. In some insects this region has become modified or assumes different names.

The ecdysial suture (coronal suture + frontal suture, epicranial suture, ecdysial line, cleavage line) is variably developed among insects. The suture is longitudinal on the vertex and separates epicranial halves of the head (Fig. 3B). Depending on the insect, the ecdysial suture may be shaped like a Y, a U, or a V. The arms of the ecdysial suture that diverge anteroventrally, called the frontal sutures (frontogenal sutures), are not present in all insects (Fig. 3B). Some of these complexes of sutures are used by insects to emerge from the old integument during molting.

The frons is that part of the head immediately ventrad of the vertex (Fig. 3A). The frons varies in size, and its borders are sometimes difficult to establish. In most insects the frons is limited ventrally by the frontoclypeal suture (epistomal suture), a transverse suture located below the antennal sockets. As its name implies, the suture separates the dorsal frons from the ventral clypeus (Fig. 3A).

The face is a generalized term used to describe the anteromedial portion of the head bounded dorsally by the insertion of the antennae, laterally by the medial margins of the compound eyes, and ventrally by the frontoclypeal suture. In some insects the area termed the face is coincident with some, most, or all of the frons.

The clypeus (Latin, shield) is a sclerite between the face and labrum (Fig. 3A). Dorsally, the clypeus is separated from the face or frons by the frontoclypeal suture in primitive insects. Laterally, the clypeogenal suture demarcates the clypeus. Ventrally, the clypeus is separated from the labrum by the clypeolabral suture (Fig. 3A). The clypeus is highly variable in size and shape. Among insects with sucking mouthparts the clypeus is large.

The gena (Latin, cheek; pl., genae) forms the cheek or sclerotized area on each side of the head below the compound eye and extending to the gular suture (Fig. 3). The size of the gena varies considerably, and its boundaries also often are difficult to establish. In Odonata the gena is the area between compound eye, clypeus, and mouthparts. The postgena (Latin, *post* = after; *gena* = cheek; pl., postgenae) is the portion of the head immediately posteriad of the gena of pterygote insects and forms the lateral and ventral parts of the occipital arch (*sensu* Snodgrass) (Fig. 3). The subgenual area is usually narrow, located above the gnathal appendages (mandible and maxillae), and includes the hypostoma (Figs. 3 and 4) and the pleurostoma. The pleurostoma is the sclerotized area between the anterior attachment of the mandible and the ventral portion of the compound eye. The hypostoma is posteriad of the pleurostoma between the posterior attachment of the mandible and the occipital foramen. The subgenal suture forms a lateral, submarginal groove or sulcus on the head, just above the bases of the gnathal appendages (Fig. 4). The subgenal suture is continuous anteriorly with the frontoclypeal suture in the generalized pterygote head. Internally, the subgenal suture forms a subgenal ridge that presumably provides structural support for the head above the mandible and maxillae. In some instances, the subgenal suture is descriptively divided in two. The part of the suture that borders the proximal attachment of the mandible to the head (Fig. 4) is called the pleurostomal suture (the ventral border of the pleurostoma). The posterior part of the subgenal suture from the mandible to the occipital foramen is called the hypostomal suture (the ventral border of the hypostoma).

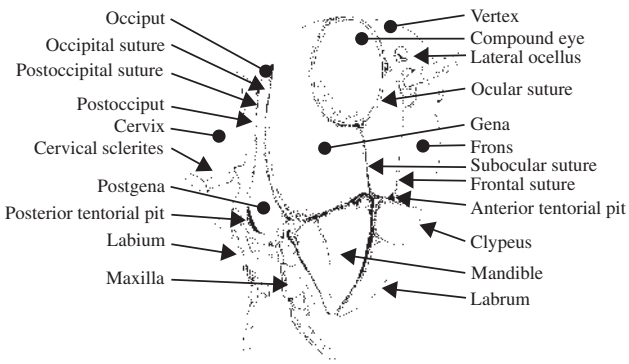

FIGURE 4 Generalized view of an insect head. [This and other line drawings after Snodgrass, R. E. (1935). "Principles of Insect Morphology," McGraw-Hill Co.].

Head Size and Shape

The size and shape of the head and its appendages reflect functional adaptations that can be used to explain biological details of the insect—the realm of morphology as opposed to anatomy.

SIZE Upon casual observation, the size of any given insect's head appears to be in proportion to the size of its body. A head that is disproportionately small or large relative to body size suggests that some adaptation has taken place that serves a functional need. Proportional head size varies considerably in the Insecta. Some fly families have very tiny heads in relation to their body size (e.g., Diptera: Acroceridae). Among Orthoptera, grass-feeding species typically have larger heads than herbaceous-feeding species. The large head is filled with powerful adductor muscles because grasses (monocots) are more difficult to chew than dicotyledonous plants. Furthermore, the postseedling stages of grasses are nutrient poor, meaning that more grass must be bitten, chopped, or ground to provide adequate nutrition.

SHAPE Head shape varies considerably among insects. Many unusual shapes seem to be influenced by behavior and may be used to illustrate examples of structural form and function. The functional importance of head shape may be difficult to determine in preserved specimens. A few hours of observation with living insects can provide considerable insight into the importance of shape. Globular heads are seen in some insects, including the burrowing crickets (e.g.,

stenopelmatines and gryllids). This form of head is adapted for pushing soil. Hypercephalic heads are seen in the males of some Diptera (Sepsidae, Diopsidae, Drosophilidae, and Tephritoidea) and Hymenoptera (Pteromalidae and Eurytomidae); the broad heads of the males are featured in various aspects of courtship behaviors.

Topographical Features of the Head

Morphologists experience considerable difficulty in defining regions and determining homologies of structure on the insect head. We cannot unambiguously characterize topographical features of the insect head because more than a million species are involved in the definition, and they show incredible diversity in head anatomy. Shape alone is not adequate or suitable because there are many head shapes, and often a head shape can be derived independently in several unrelated lineages. Some head shapes are influenced by behavior.

AXIAL POSITION The posture or orientation of the head in its resting position relative to the long axis of the body can be important in providing definitions of the anatomical features of the head. Axial position in insects typically falls into three basic categories: hypognathous, prognathous, and opisthognathous.

In general zoological usage, the word "hypognathous" (Greek, *hypo* = under; *gnathos* = jaw) serves to designate animals whose lower jaw is slightly longer than the upper jaw. In entomological usage, "hypognathous" refers to insects with the head vertically oriented and the mouth directed ventrad. Most insects with a hypognathous condition display an occipital foramen near the center of the posterior surface of the head. The hypognathous condition is considered by most insect morphologists to represent the primitive or generalized condition. The hypognathous position is evident in most major groups of insects and can be seen in the grasshopper, house fly, and honey bee. Other conditions are probably derived from ancestors with a hypognathous head.

In general zoological usage "prognathous" (Greek, *pro* = forward; *gnathos* = jaw) refers to animals with prominent or projecting jaws. In entomological usage, the prognathous condition is characterized by an occipital foramen near the vertexal margin with mandibles directed anteriad and positioned at the anterior margin of the head. When viewed in lateral aspect, the primary axis of the head is horizontal. Some predaceous insects, such as carabid beetles and earwigs, display the prognathous condition. In other insects, such as cucujid beetles and bethylid wasps, the prognathous position may reveal a solution to problems associated with living in concealed situations such as between bark and wood or similar confined habitats.

In general zoological usage, "opisthognathous" (Greek, *opisthos* = behind; *gnathos* = jaw) refers to animals with retreating jaws. In entomological usage, the opisthognathous condition is characterized by posteroventral position of the mouthparts resulting from a deflection of the facial region. The opisthognathous condition is displayed in many fluid-feeding Homoptera, including leafhoppers, whiteflies, and aphids.

SUTURES OF THE HEAD Head sutures are sometimes used to delimit specific areas of the head, but there are problems. Establishing homology of sutures between families and orders is difficult. From a practical viewpoint, standards have not been developed for naming sutures among insect groups. Some names are based on the areas delimited (e.g., frontoclypeal suture); other sutures are named for the areas in which the suture is found (e.g., coronal suture). Sutures frequently have more than one name (e.g., frontoclypeal suture and epistomal suture are synonymous).

The compound eye is an important landmark on the insect head. An ocular suture surrounds the compound eye and forms an inflection or an internal ridge of the integument (Figs. 3, 4). The ocular suture is not present in all insects and is difficult to see in some insects unless the head is chemically processed for microscopic examination. When present, the ocular suture probably provides strength and prevents deformation of the compound eye.

A subocular suture extends from the lower margin of the compound eye toward the subgenal suture. In some species the subocular suture (Fig. 4). may extend to the subgenal suture; in other species it may terminate before reaching another landmark. This suture is straight and commonly found in the Hymenoptera, where it may provide additional strength for the head.

POSTERIOR ASPECT OF THE HEAD The entire posterior surface of the head is termed the postcranium (Fig. 3). The surface may be flat, concave, or convex, depending on the group of insects. The occiput (Latin, back of head) of pterygote insects is the posterior portion of the head between the vertex and cervix (Latin, neck). The occiput is rarely present as a distinct sclerite or clearly demarcated by "benchmark" sutures. When present, the occiput signifies a primitive head segment. In some Diptera the occiput forms the entire posterior surface of the head. In other insects it forms a narrow, horseshoe-shaped sclerite.

The occipital suture (hypostomal suture *sensu* MacGillivray) is well developed in orthopteroids, but it is not present in many other groups of pterygote insects (Fig. 3C). When present, the occipital suture forms an arched, horseshoe-shaped groove on the back of the head that ends ventrally, anterior to the posterior articulation of each mandible. Internally, the occipital suture develops into a ridge, providing strength for the head.

The postoccipital suture is a landmark on the posterior surface of the head and is typically near the occipital foramen (Fig. 3C). The postoccipital suture forms a posterior submarginal groove of the head with posterior tentorial pits marking its lower ends on either side of the head. Some morphologists regard this suture as an intersegmental boundary (labium) between the first and second maxillae.

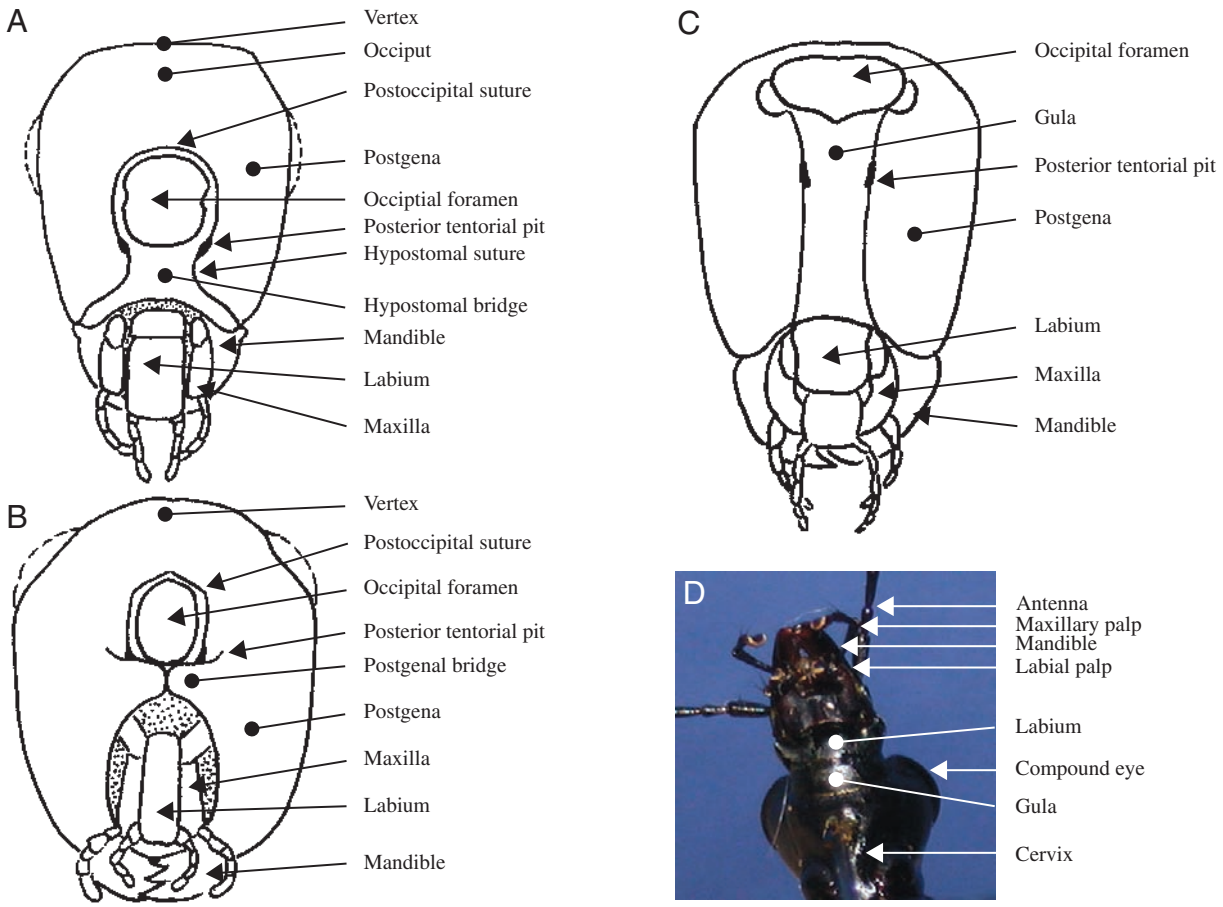

FIGURE 5 Occipital closures (diagrammatic): (A) hypostomal bridge, (B) postgenal bridge, (C) gula and (D) ventral view of gula of adult ground beetle (Coleoptera: Carabidae).

Internally, the postoccipital suture forms the postoccipital ridge that serves as an attachment for the dorsal prothoracic and cervical muscles of the head. The absence of the postoccipital suture in pterygote insects is a derived condition.

The postocciput of pterygotes forms the extreme posterior, often U-shaped sclerite that forms the rim of the head behind the postoccipital suture. The postocciput is interpreted as a sclerotic remnant of the labial somite in ancestral insects.

In pterygotes such as Orthoptera the occipital foramen and the mouth are not separated. More highly evolved insects have developed sclerotized separations between the mouthparts and the occipital foramen. At least three types of closure have been identified (Fig. 5): the hypostomal bridge, the postgenal bridge, and the gula. An understanding of these structures provides insight into the operation of the head and suggests evolutionary trends in feeding strategies.

The hypostomal bridge is usually developed in adult heads displaying a hypognathous axial orientation. The bridge is formed by medial extension and fusion of hypostomal lobes (hypostoma) (Fig. 5A). The hypostomal bridge is the ground plan condition of closure for the posterior aspect of the head, but it is not restricted to primitive insects. The hypostomal bridge is found in highly developed members of the Heteroptera, Diptera, and Hymenoptera. In Diptera the hypostomal bridge also has been called the pseudogula.

The postgenal bridge is a derived condition from the hypostomal ground plan and is developed in adults of higher Diptera and aculeate Hymenoptera. The bridge is characterized by medial extension and fusion of the postgenae, following a union of the hypostoma (Fig. 5B). The posterior tentorial pits retain their placement in the postoccipital suture.

The gula (Latin, gullet; pl., gulae) is developed in some Coleoptera, Neuroptera, and Isoptera. Typically, the gula is developed in heads displaying a prognathous axial orientation and in which posterior tentorial pits are located anteriad of the occipital foramen. (Fig. 5C, D). The median sclerite (the gula) on the ventral part of a prognathous head apparently forms *de novo* in the membranous neck region between the lateral extensions of the postocciput. The gula is a derived condition that is found in some but not all prognathous heads.

Endoskeletal Head Framework

Although the hardened integument of the head forms a structurally rigid capsule, this design is insufficient to solve the problems associated with muscle attachment and

maintaining structural integrity during chewing. Thus, insects have evolved a tentorium (Latin, tent; pl., tentoria): a complex network of internal, hardened, cuticular struts that serve to reinforce the head. The tentorium forms as an invagination of four apodemal arms from the integument in most pterygote insects. The tentorium strengthens the head for chewing, provides attachment points for muscles, and also supports and protects the brain and foregut.

Anatomically, the tentorium consists of anterior and posterior arms. In most insects, the anterior arms arise from facial inflections located just above the anterior articulations of the mandibles. Externally, the arms are marked by anterior tentorial pits positioned on the frontoclypeal or subgenal (pleurostomal) suture (Fig. 4). Internally, the anterior region may form a frontal plate. Posterior arms originate at the ventral ends of the postoccipital inflection. They are marked externally by posterior tentorial pits (Fig. 4). The posterior arms usually unite to form a transverse bridge or corpus tentorii (internally) across the back of the head. Dorsal arms (rami), found in many insects, arise from the anterior arms. They attach to the inner wall of the head near antennal sockets. The dorsal arms are apparently not an invagination of cuticle, because pits do not mark them externally.

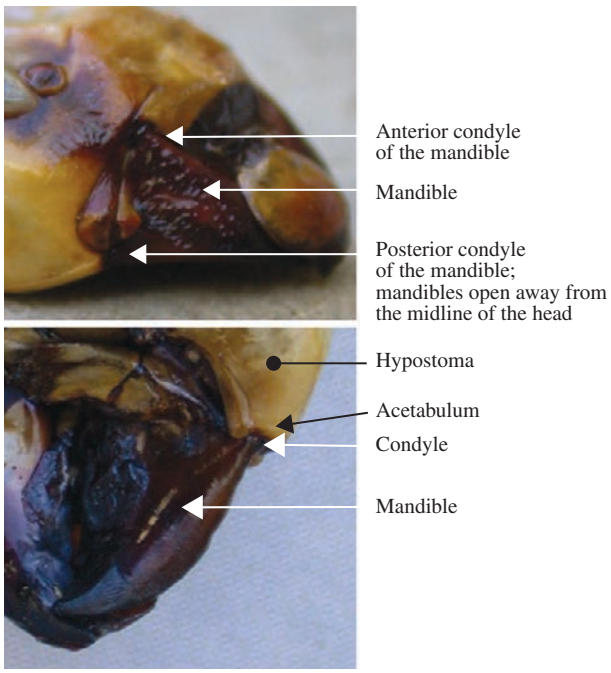

FIGURE 6 Mandible articulation. Top: lateral view; bottom: posterior view (Orthoptera: Stenoplematidae).

Mandible Articulation and Musculature

The hypothetical ancestor of insects is thought to possess a mandible with one point of articulation. Later, insects acquired a second point of articulation. The basis of this assumption comes from a survey of the Hexapoda in that the modern Apterygota have a monocondylic mandible and the Pterygota have a dicondylic mandible.

The term condyle (Greek, *kondylos* = knuckle) refers specifically to a knoblike process. For the mandible, the condyle is the point of articulation with the head. On the head itself is an acetabulum (Latin, *acetabulum* = vinegar cup), a concave surface or cavity for the reception and articulation of the condyle (Fig. 6).

The dicondylic mandible is the derived condition and is found in the Lepismatidae and Pterygota. The dicondylic mandible has secondarily acquired an articulation point anterior to the first point in the monocondylic mandible. These attachments form a plane of attachment. In the monocondylic mandible there is no plane of attachment, and the mandibles move forward or rearward when the muscles contract. The two points of articulation create a plane of movement that restricts the direction of mandible movement.

THORAX

The thorax represents the second tagma of the insect body. The thorax evolved early in the phylogenetic history of insects. In most Paleozoic insects the thorax is well developed and differentiated from the head and abdomen, and the three distinct tagmata probably developed during the Devonian.

In terms of insect phylogeny, the thorax of Apterygota is strikingly different in shape compared with the head or abdomen. Of modern apterygotes only the Collembola display taxa in which thoracic tagmatization and segmentation are not obvious.

Apparently, the primary, functional role of the thorax has always been locomotion, since the primary modifications of the thorax have been for locomotion (first walking, and then flight). Modification for locomotion probably developed before other morphological adaptations, such as metamorphosis. Diverse independent and interdependent mechanisms for locomotion have evolved throughout the Insecta, including walking, flight, and jumping. Active participation in flight by insects is unique among invertebrates.

Anatomy of the Thorax

The cervix is the connection between the head (occipital foramen) and the anteriormost part of the thorax (pronotum) (Fig. 2). Typically, the area between the head and pronotum is membranous. The ground plan for the insect cervix contains two cervical sclerites on each side of the head that articulate with an occipital condyle of the head and the prothoracic episternum. Musculature attached to these sclerites increases or decreases the angle between the sclerites, and creates limited mobility of the head.

The thorax of modern insects consists of three segments termed the prothorax, mesothorax, and metathorax. The last two collectively are called the pterothorax (Greek, *ptero* = wing or feather) because extant insects bear wings on these segments only. The individual dorsal sclerites or terga of the thoracic

segments are also known as nota (Greek, *notos* = back; sing, notum). The nota of Apterygota and many immature insects are similar to the terga of the abdomen with typical secondary segmentation. The nota of each thoracic segment are serially distinguished as the pronotum, mesonotum, and metanotum.

The size and shape of the prothorax are highly variable. The prothorax may be a large plate as in Orthoptera, Hemiptera, and Coleoptera, or reduced in size forming a narrow band between the head and mesothorax as in Hymenoptera. The prothorax is usually separated or free from the mesothorax. The sclerites are separated by a membrane that may be large and conspicuous in more primitive holometabolous insects such as Neuroptera and Coleoptera, or reduced in size in more highly evolved holometabolous insects such as Diptera and Hymenoptera.

The pterothorax includes the thoracic segments immediately posteriad of the prothorax. In winged insects the relationship between thoracic segments involved in flight can be complicated. In contrast, the thorax of larval insects and most wingless insects is relatively simple. The mesothorax and metathorax of these insects are separated by membrane. Adult winged insects show a mesothorax and metathorax that are consolidated (i.e., more or less united) to form a functional unit modified for flight.

The development of the pterothoracic segments varies among winged insects. When both pairs of wings participate equally in flight, the two thoracic segments are about the same size. This condition is seen in Odonata, some Lepidoptera, and some Neuroptera. When one pair of wings is dominant in flight, the corresponding thoracic segment is commensurately larger and modified for flight, whereas the other thoracic segment is reduced in size. This condition is seen in Diptera and Hymenoptera, where the forewing is large and dominant in flight. The reverse condition is seen in the Coleoptera, where the hind wing is large and dominant in flight.

In more closely related insect groups, such as families within an order, that are primitively wingless or in which wings have been secondarily lost in modern or extant species, many modifications to the thorax occur. Many wingless forms can be attributed to environmental factors that promote or maintain flightlessness. For instance, island-dwelling insects are commonly short winged (brachypterous), or wingless, whereas their continental relatives are winged, presumably because for island species, flight increases the likelihood of being carried aloft, moved out to sea, and subsequently lost to the reproductive effort of the population. The anatomical consequences of flightlessness can be predictable; in the Hymenoptera, short wings bring a disproportionate enlargement of the pronotum and reduction in size of the mesonotum and metanotum.

Sutures and Sclerites of Wing-Bearing Segments

The wing-bearing segments of the thorax are subdivided into a myriad of sclerites that are bounded by sutures and membranous areas. These sutures and sclerites are the product of repeated modification of the thorax in response to various demands placed on the insect body by the environment. Similar modifications have occurred independently in many groups of insects; some modifications are unique. Generalizations are difficult to make, given the large number of sutures and sclerites, coupled with the number of insects that there are to consider.

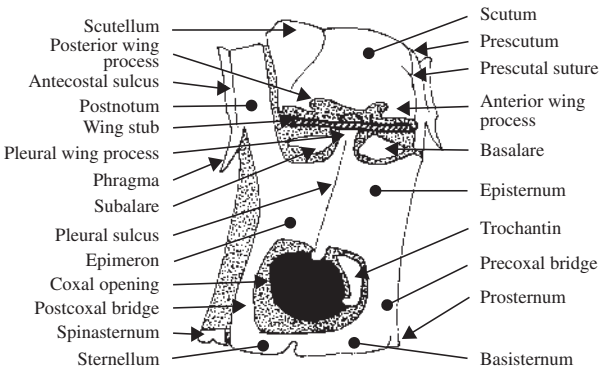

FIGURE 7 Diagram of the pterygote pterothorax.

Dorsal Aspect

The nota of the pterothorax are further subdivided into the prescutum, scutum, and scutellum; again, serially distinguished as mesoscutum and mesoscutellum, and metascutum and metascutellum (Fig. 7). Additionally, there are sclerites anterior and posterior to the notum, as discussed shortly.

The prescutum is the anterior portion of the scutum, laterally bearing prealar bridges separated by the prescutal suture from the mesoscutum. The scutum is the largest dorsal sclerite of the notum and is bounded posteriorly by the scutoscutellar suture, which divides the notum into the scutum and scutellum. The scutellum is generally smaller than the scutum. In Heteroptera it is a small triangular sclerite between the bases of the hemelytra. In Coleoptera the scutellum is the small triangular sclerite between the bases of the elytra. In Diptera and Hymenoptera the scutellum is relatively large, forming a subhemispherical sclerite, sometimes projecting posteriad. The posteriormost sclerite of the notum is the postnotum, separated from the notum by secondary segmentation. In some insects there is a postscutellum (metanotal acrotergite) that forms the posteriormost thoracic sclerite of the metanotum, or the posteriormost sclerite of the thorax. In Diptera the postscutellum appears as a transverse bulge below the scutellum.

The acrotergite and postnotum deserve further explanation. Again, the anteriormost sclerite is an acrotergite, the anterior precostal part of the notal plate. The postnotum is an intersegmental sclerite associated with the notum of the preceding segment. The postnotum bears the antecosta, a marginal ridge on the inner surface of the notal sclerite

corresponding to the primary intersegmental fold. The postnotum also usually bears a pair of internal projecting phragmata. The antecostal suture divides the acrotergite from the antecosta, the internal ridge marking the original intersegmental boundary. Thus, when the antecosta and acrotergite are developed into larger plates and are associated with the notum anterior to them, they are referred to as a postnotum. The final structure associated with the dorsal aspect of the pterothorax is the alinotum (Greek, *ala* = wing; *notos* = back; pl., alinota). The alinotum is the wing-bearing sclerite of the pterothorax.

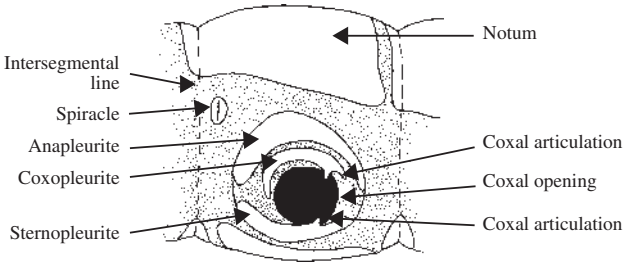

FIGURE 8 Pleural aspect of the apterygote thorax: diagrammatic.

Wing Articulation

The thoracic components necessary for wing movement include the prealar bridge, anterior notal wing process, and posterior notal wing process. The components of the wing itself that articulate with the thoracic components are the humeral and axillary sclerites; they form the part of the wing closest to the body and are not treated in this article.

The prealar bridge is a heavily sclerotized and rigid supporting sclerite between the unsclerotized membrane of the pterothorax and the pleuron; it supports the notum above the thoracic pleura. The prealar bridge is comprosed of cuticular extensions from the anterior part of the prescutum and antecosta. The anterior notal wing process is the anterior lobe of the lateral margin of the alinotum supporting the first axillary sclerite (Fig. 7). The posterior notal wing process is a posterior lobe of the lateral margin of the alinotum that supports the third axillary sclerite of the wing base (Fig. 7).

Lateral Aspect

The pleuron (Greek = side; pl., pleura) is a general term associated with the lateral aspect of the thorax. Adults, nymphs, and active larvae all display extensive sclerotization of the pleural area. Sclerites forming this part of the body wall are derived from the precoxa, subscoxa, or supracoxal arch of the subcoxa.

PLEURAL REGIONS OF THE THORAX

Apterygota and Immature Plecoptera The anapleurite is the sclerotized area above the coxa (supracoxal area) (Fig. 8). The coxopleurite is a sclerotized plate situated between the coxa and the anapleurite (Fig. 8). It bears the dorsal coxal articulation, the anterior part of which becomes the definitive trochantin. The sternopleurite, or coxosternite, is the definitive sternal sclerite that includes the areas of the limb bases and is situated beneath the coxa (Fig. 8).

Pterygota The basalare is a sclerite near the base of the wing and anterior to the pleural wing process (Fig. 9). The basalare serves as a place of insertion for the anterior pleural muscle of the wing. The subalare is posterior to the basalare and the pleural wing process (Fig. 9). It too serves as a place for insertion of the wing's posterior pleural muscle. The tegula is the anterior most independent sclerite associated with the wing base. The tegula is typically scalelike, articulates with the humeral sclerite, and protects the wing base from physical damage. The tegula is absent from Coleoptera and from the metathorax of most orders. The pleural wing process is located at the dorsal end of the pleural ridge and serves as a fulcrum for the movement of the wing (Fig. 9). The parapteron is a small sclerite, articulated on the dorsal extremity of the episternum just below the wings (Fig. 7).

The pleural suture is an easily visible landmark on the pterothoracic pleura (Fig. 9). It extends from the base of the wing to the base of the coxa. The pleural ridge is formed internally by the pleural suture and braces the pleuron above the leg. The episternum is a pleural sclerite anterior to the pleural suture and sometimes adjacent to the coxa (Fig. 9); the episternum is typically the largest lateral thoracic sclerite between the sternum and the notum. The epimeron is the posterior division of a thoracic pleuron adjacent to the coxa and posterior to the pleural suture (Fig. 9); it is typically smaller than the episternum and narrow or triangular. The episternum and the epimeron of many insects have become subdivided into several secondary sclerites bounded by sutures. The simplest condition shows the episternum divided into a dorsal anepisternum and a ventral katepisternum (Fig. 9). Similarly, the epimeron is divided into an anepimeron and katepimeron. The trochantin is a small sclerite at the base of the insect leg of some insects (Figs. 7, 9). Some workers theorize that the trochantin may have developed into the pleural wall. The trochantin is often fused to the episternum or absent.

The precoxal bridge is anterior to the trochantin, usually continuous with the episternum, frequently united with the

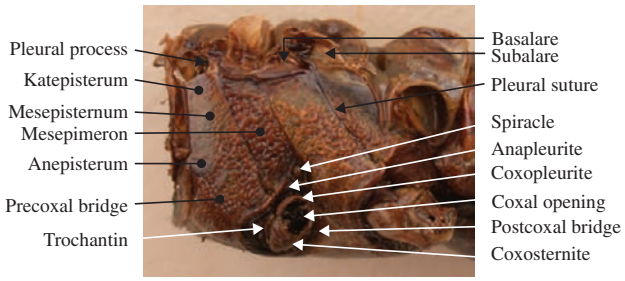

FIGURE 9 Lateral aspect of the pterygote thorax (Orthoptera: Acrididae).

basisternum, but also occurs as a distinct sclerite (Fig. 9). The postcoxal bridge is the postcoxal part of the pleuron, often united with the sternum behind the coxa (Fig. 9). The sclerite extends behind the coxa and connects the epimeron with the furcasternum. The meron is a lateral, postarticular basal area of the coxa and is sometimes found disassociated from the coxa and incorporated into the pleuron. The meron is typically large and conspicuous in panorpid and neuropteran insects. In Diptera the meron forms a separate sclerite in the thoracic pleuron.

Ventral Aspect

The ground plan of the sternum (Greek, *sternon* = chest; pl., sterna) consists of four sclerites, including an intersternite (spinasternite), two laterosternites (coxosternites), and a mediosternite (Fig. 10). The mediosternite and the laterosternite meet and join, and the line of union is called the laterosternal sulcus (pleurosternal suture) (Fig. 10).

Paired furcal pits are found in the laterosternal sulcus (Fig. 10). A transverse sternacostal sulcus bisects the ventral plate and thereby forms an anterior basisternite and posterior furcasternite (Fig. 10). The basisternite (basisternum) is the primary sclerite of the sternum (Fig. 10). It is positioned anterior to the sternal apophyses or sternacostal suture and laterally connected with the pleural region of the precoxal bridge. The furcasternite (furcasternum) is a distinct part of the sternum in some insects bearing the furca (Fig. 10). The spinasternum is a "spine-bearing" intersegmental sclerite of the thoracic venter, associated or united with the preceding sternum. The spinasternum may become part of the definitive prosternum or mesosternum, but not of the metasternum. The sternellum is the second sclerite of the ventral part of each thoracic segment, frequently divided into longitudinal parts that may be widely separated (Figs. 7, 10).

ABDOMEN

The abdomen is more conspicuously segmented than either the head or the thorax. Superficially, the abdomen is the least specialized of the body tagma, but there are notable exceptions such as the scale insects. The abdomen characteristically lacks appendages except cerci, reproductive organs, and pregenital appendages in adult Apterygota and larval Pterygota.

Ground Plan of the Abdomen

The ground plan abdomen of an adult insect typically consists of 11 to 12 segments and is less strongly sclerotized than the head or thorax (Fig. 11). Each segment of the abdomen is represented by a sclerotized tergum, sternum, and perhaps a pleurite. Terga are separated from each other and from the adjacent sterna or pleura by a membrane. Spiracles are located in the pleural area. Modification of this ground plan includes the fusion of terga or terga and sterna to form continuous dorsal or ventral shields or a conical tube. Some insects bear a sclerite in the pleural area called a laterotergite. Ventral sclerites are sometimes called laterosternites. The spiracles are often situated in the definitive tergum, sternum, laterotergite, or laterosternite.

During the embryonic stage of many insects and the postembryonic stage of primitive insects, 11 abdominal segments are present. In modern insects there is a tendency toward reduction in the number of the abdominal segments, but the primitive number of 11 is maintained during embryogenesis. Variation in abdominal segment number is considerable. If the Apterygota are considered to be indicative of the ground plan for pterygotes, confusion reigns: adult Protura have 12 segments, Collembola have six. The orthopteran family Acrididae has 11 segments, and a fossil specimen of Zoraptera has a 10-segmented abdomen.

Anamorphosis is present among some primitive ancestral hexapods such as the Protura—they emerge from the egg with eight abdominal segments and a terminal telson. Subsequently, three segments are added between the telson and the last abdominal segment with each molt. In contrast, most insects undergo epimorphosis in which the definitive number of segments is present at eclosion. Given the extent of variation in abdominal segmentation, morphologists conventionally discuss the abdomen in terms of pregenital, genital, and postgenital segmentation.

Abdominal Anatomy

Typically, the abdominal terga show secondary segmentation with the posterior part of a segment overlapping the anterior

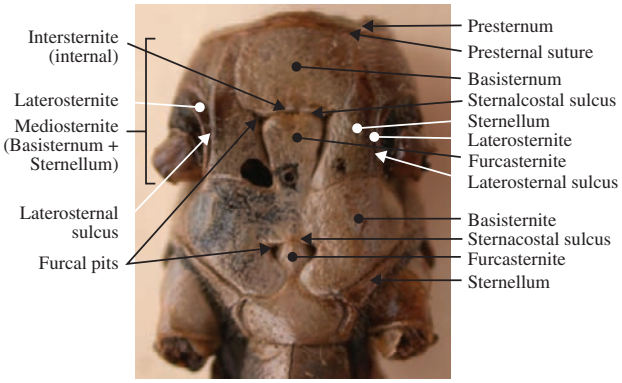

FIGURE 10 Ventral aspect of the thorax (Orthoptera: Acrididae).

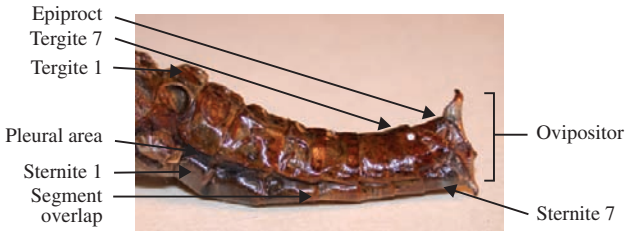

FIGURE 11 Insect abdomen (Orthoptera: Acrididae).

part of the segment behind it (Fig. 11). Such overlap prevents damage or injury to the animal while it moves through the environment, particularly in confined spaces.

The pregenital segments in male insects are numbered 1 through 8; the pregenital segments in female insects are numbered 1 through 7 (Fig. 11). Among the Apterygota, male genitalia in Collembola are positioned between segments 5 and 6 and in Protura between segments 11 and the paraproct. Genital segments of Pterygota include segment 9 in males and segments 8 and 9 in females. Postgenital segments of pterygote insects are 10 and 11 in females and 9 and 10 in males.

In general there is little modification of the pregenital sclerites. A notable exception is found in the Odonata. Male Odonata do not have an intromittent organ on segment 9. Instead, the male moves the abdominal apex forward and deposits sperm in a reservoir along the anterior margin of the third abdominal sternum. Other modifications of the pregenital sclerites are not related to sexual behavior. Some of these modifications are glandular.

Modification of the genital sclerites from the ground plan is frequently observed among insects. Adult Pterygota are characterized by a well-developed reproductive system, including organs of copulation and oviposition. This duality of function has resulted in considerable differentiation of associated segments and contributed to difference of opinion regarding homology of genitalic parts. Among pterygote insects the male genitalia are generally positioned on segment 9. The ninth sternum is called a hypandrium (Greek, *hypo* = beneath; *aner* = male; Latin, *-ium* = diminutive) in many insects, including Psocoptera. In Ephemeroptera, the tenth sternum is called a hypandrium. Fusion of segments 9 and 10 in Psocoptera results in a structure called the clunium (Latin, *clunais* = buttock).

The gonopore (Greek, *gone* = seed; *poros* = channel) of the female reproductive system serves as the aperture through which the egg passes during oviposition. The gonopore usually is located on segment 8 or 9. Enlargement of sternum 8 in some female insects is called a subgenital plate.

Modification of postgenital sclerites is frequently observed and seems to be a functional response to adaptations associated with copulation and oviposition. Some modifications include fusion of the tergum, pleuron, and sternum to form a continuous sclerotized ring. The phenomenon is notable in apterygota and pterygote insects.

The eleventh abdominal segment forms the last true somite of the insect body. Frequently, this segment is found in embryonic stages of primitive insects even when it cannot be observed in postemergent stages. When the eleventh segment is present, it forms a conical endpiece that bears an anus at the apex, flanked laterally by cerci (Greek, *kerkos* = tail) (Fig. 12). The dorsal surface of the eleventh segment is called an epiproct (Greek, *epi* = upon; *proktos* = anus); the ventrolateral surface is called a paraproct (Greek, *para* = beside; *proktos* = anus) (Fig. 12). A longitudinal, medial, membranous area connects

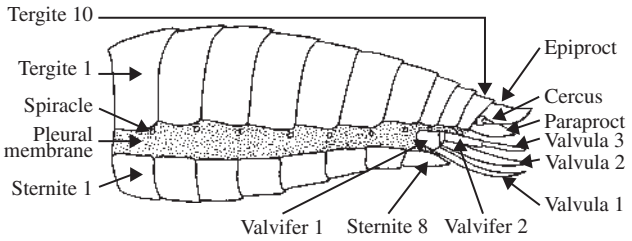

FIGURE 12 Abdominal segmentation: diagrammatic.

the paraprocts ventrally. Primitive groups of extant insects such as Thysanura and Ephemeroptera, and some fossil groups such as Paleodictyoptera, display a conspicuous, long, median filament that apparently projects from the apex of the epiproct. This is called the appendix dorsalis or caudal style. The appendage appears annulated and similar in shape to the lateral cerci, but the function of the appendix is unknown. The twelfth abdominal segment is called the periproct in Crustacea, and it forms a telson in some embryonic insects. The periproct appears in adult Protura and naiadal Odonata.

Abdominal Appendages

Presumably, the hypothetical ancestor of the Insecta was a myriapod with one pair of appendages for each body segment. Among contemporary insects the head appendages are represented by the antennae, mandibles, and the first and second maxillae. Thorax appendages are represented by legs, whereas the wings are considered to be secondary in origin. In most Apterygota, paired abdominal appendages are apparent. In most true insects embryological appendages are formed and lost before eclosion. The appendages found in embryos apparently represent ancestral conditions that are not expressed in postembryonic stages of modern insects. In modern insects, most pairs of appendages have been lost, and the irregular distribution of the remaining appendages makes a summary evaluation difficult. Abdominal appendages do not resemble the structure of thoracic legs of any insect.

Appendages are common among some entognathous hexapods, and some ancestral forms display unique abdominal appendages. Collembola are highly specialized entognathous Hexapoda. The abdomen of Collembola bear saltatorial appendages, which gives the group its common name of springtail, and a ventral tube, the collophore, which is the basis of the ordinal name.

The collophore (Greek, *kolla* = glue; *pherein* = to bear) is found on the first abdominal segment of Collembola. The collophore forms a ventromedial tube that is eversible with hydrostatic pressure and is drawn inward with retractor muscles. Some morphologists believe the collophore represents the fusion of paired, lateral appendages of an ancestor. An early explanation of the collophore function noted it was an organ of adhesion. The collophore also is used as a grooming organ in some Collembola. The collophore is

connected to secretory glands in the head, and the median longitudinal channel on the venter of the thorax extends from the head to the base of the collophore.

OTHER APPENDAGES Protura maintain short, cylindrical appendages on each of the first three abdominal segments. Each of these arises from membranous areas between the posterolateral angles of the terga and sterna. The position suggests a pleural origin.

APPENDAGES OF PTERYGOTA The aquatic neuropteran larva *Sialis* has long, tapering, six-segmented appendages on each of the first seven body segments. These appendages articulate to pleural coxopodites. Similar appendages are found on the abdomen of some aquatic coleopteran larvae.

The tenth abdominal segment is present in most larval and adult Holometabola. As noted earlier, it is sometimes fused with segment 11. Segment 10 displays paired appendicular processes called pygopodia in Trichoptera, Coleoptera, and Lepidoptera. Pygopods form terminal eversible appendages in some beetle larvae. Pygopodia are bilaterally symmetrical, with eight podia, or feet, per side. Control of the podia is apparent because they are not always everted or inverted. Podia are withdrawn into the segment and have a common or median stalk. Each podium has several rows of equally spaced acanthae that apparently serve as holdfasts. Functionally, the acanthae enable the larvae to attach to and move on different substrates. When the larva walks on a flat substrate, the pygopodia are retracted into the body. When the larva walks on the edge of a leaf, the pygopodia are everted and used as holdfasts.

The larval prolegs of terrestrial Lepidoptera and Symphyta are not well developed, but they are adapted to grasping substrates. These structures are considered to be serially homologous with legs, but they also are referred to by some as adaptive structures with no relation to legs.

The adult pterygote abdomen has appendages that are not generally observed. These appendages are grouped for discussion based on the segments of the abdomen on which they are found.

Pregenital appendages are rare among insects. Adult whiteflies have a curious structure on sternum 8 that propels honeydew away from the body. Genitalia are segmental appendages and are treated in the next section. Postgenital appendages include cerci (Latin, circle), which are thought to represent primitive appendages because they are found in the Apterygota (except Protura) and many Pterygota. Cerci originate on abdominal segment 11 in a membranous area between the epiproct and the paraproct (Fig. 12). In insects that have lost segment 11, the cerci appear to originate on segment 10. Cerci occur in all orders among the Hemimetabola except for hemipteroids; among the Holometabola, they are found only in the Mecoptera and Symphyta.

Cerci are highly variable in size and shape and function. They are longer than the body in Thysanura, and in some Orthoptera cerci may be indistinct. Cerci resemble forceps in Japygidae and are annulated in Dictyoptera. In Dictyoptera they detect air currents, are sensitive to sound, and may be chemoreceptive. Some Ephemeroptera use cerci to propel themselves through water. Japygidae and Dermaptera probably use cerci to subdue prey. In some groups such as Embioptera and Orthoptera, cerci are sexually dimorphic and may serve a role in copulation.

There are some features on the insect body that appear as appendages but are not. Urogomphi (Greek, *oura* = tail; *gomphos* = nail; sing., urogomphus) are fixed or mobile cuticular processes on the apical abdominal segment of some coleopteran larvae. They may or may not be homologous with cerci, or other true appendages.

GENITALIA

The examination of the reproductive anatomy of different insect orders helps to develop an appreciation for the evolutionary trends in the formation of the external genitalia. The male genitalia are derived from the ninth abdominal segment. The female genitalia are derived from the eighth and ninth abdominal segments. In the female, the aperture through which the egg passes is called a gonopore. The gonopore serves as a boundary between the external and internal genitalia and is usually independent of the anus. Exceptions include some flies, such as the Tephritidae, where a common lumen termed a cloaca serves for excretion, copulation, and oviposition.

There is usually a single, medially located gonopore. The Dermaptera and Ephemeroptera are ancient groups of hemimetabolous insects. Both orders display a condition in which the lateral oviducts do not combine to form a median oviduct. Instead, the lateral oviducts independently connect with paired gonopores on the conjuctival membrane along the posterior margin of the seventh abdominal segment.

Many female insects with a genitalic opening on the posterior margin of the eighth abdominal segment display an appendicular ovipositor (Fig. 13). The ovipositor is a structure that develops from modified abdominal appendages or segments. It functions in the precise placement of eggs. It is commonly assumed that insects that do not show an ovipositor have ancestors that had an ovipositor. Thus, the structure has been lost during the course of evolutionary adaptation to a particular lifestyle.

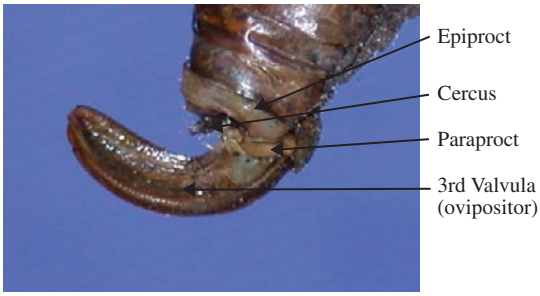

FIGURE 13 Appendicular ovipositor (Orthoptera: Tettigoniidae).

Female insects with a genitalic opening on the posterior margin of the ninth abdominal segment typically display a rudimentary or suppressed appendicular ovipositor. These insects lack special provisions for egg placement, but sometimes they reveal other abdominal modifications intended to facilitate oviposition.

Female Genitalia

Morphologists often use the Thysanura as a starting point for developing a generalized model to explain the evolution of the external reproductive system of pterygote insects. The thysanuran abdomen has basal sclerotized plates called coxopodites on which styli are attached. These plates are serially homologous along the abdomen, and the pregenital plates are regarded as identical with the genital plates. The plates located on segments 8 and 9 are considered to be genital plates. The styli associated with these segments are called gonapophyses. There are four gonapophyses on segments 8 and 9 (i.e., a pair of styli on each segment). The gonapophyses are medially concave and directed rearward. The basal sclerite is called a gonocoxa, and in some Thysanura it may be fused with the style.

The primitive pterygote with a gonopore on segment 8 has an appendicular ovipositor that consists of three components. A basal apparatus corresponds to the basal plate or primitive gonocoxite of the thysanuran abdominal appendage. The second part is the first valvifers (on the eighth sternum), and second valvifers (on the ninth sternum) are responsible for providing support and points of articulation for the tube through which the egg passes (Fig. 14). Interpolated between the first and second valvifers is a small sclerite called a gonangulum, which articulates with the second gonocoxite and tergum 9. The gonangulum is present in Odonata and Grylloblatoidea. It apparently is fused with the first valvifer in Dictyoptera and Orthoptera. In the remaining orders these structures are highly variable.

The shaft of the ovipositor consists of two pairs of elongate, closely appressed sclerites called the first and second valvulae (Fig. 14). The first pair of valvulae is positioned on the eighth abdominal sternum. The second pair of valvulae is located on the ninth abdominal sternum and is dorsal in position. Third valvulae are positioned on the posterior end of the second valvifers. These valvulae usually serve as a sheath for the shaft of the ovipositor (Figs. 13, 14).

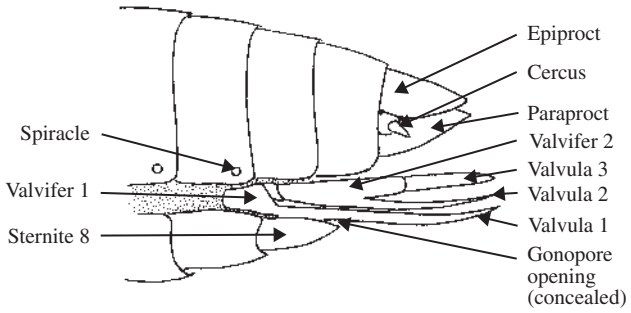

FIGURE 14 Female genitalia (diagrammatic), based on orthopteran female.

Male Genitalia

The primary function of the male genitalia in insects is insemination of the female. Methods of achieving insemination that involve special functions of the external genitalia include clasping and holding the female, retaining the connection with the female gonopore, the construction of spermatophores, and the deposition of spermatophores or semen into the female genital tract; in some insects the injection of semen takes place directly into the female body (traumatic insemination of some Hemiptera). Other functions of the male genitalia include excretion and various sensory functions.

The genitalia of male insects exhibit such an enormous variety of shapes and constituent parts, often further complicated by structural rotation or inversion of all or some of the parts, that determination of a ground plan is virtually impossible. Examination of ancient orders shows highly variable and specialized conditions. In general, the coxites of the eighth segment in most apterygotes are reduced and without gonapophyses, and they are absent altogether in the Pterygota. Thus, the male external genitalia are derived from the ninth abdominal coxites.

Again, the Thysanura have genitalia that closely resemble that of the pterygote orders: a median intromittent organ or phallus, and paired lateral accessories (the periphallus of Snodgrass). The phallus is a conical, tubular structure of variable complexity (Fig. 15). Primitive insects may not display differentiated parts, and the entire structure may be long, sclerotized, and tapering apicad. In a ground plan

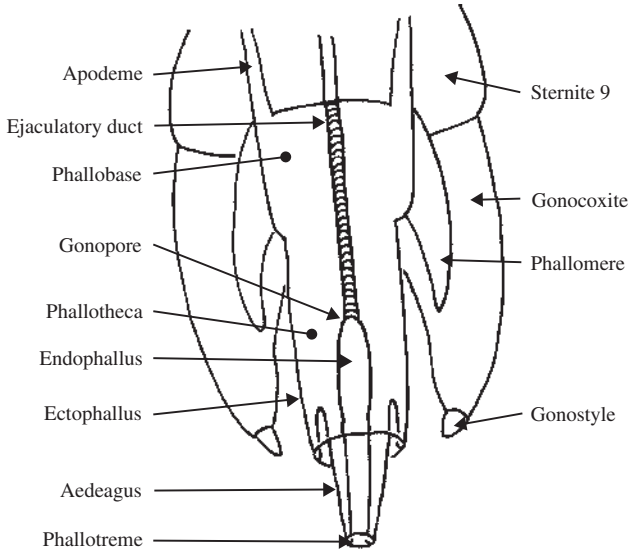

FIGURE 15 Male genitalia (diagrammatic).

condition for pterygote insects, there is a sclerotized basal portion termed the phallobase and a distal sclerotized portion called the aedeagus (Fig. 15). The phallobase in insects is characterized by highly variable development: sometimes sclerotized and supporting the aedeagus, sometimes forming a sheath for the aedeagus. The phallobase often contains an apodeme, which may provide support or a point for muscle attachment. The phallobase and aedeagus are joined by a membranous phallotheca (Fig. 15). The external walls of the phallobase and aedeagus are called the ectophallus (Fig. 15). The gonopore is positioned at the apex of the ejaculatory duct and is concealed within the phallobase. The gonopore is connected to the apex of the aedeagus via a membranous tube called the endophallus (Fig. 15). In some insects the endophallus may be everted through the aedeagus. The circular aperture at the apex of the aedeagus is called the phallotreme (Fig. 15). In some insects the endophallus and the gonopore may be everted through the phallotreme and into the female's bursa copulatrix. Genital lobes referred to as phallomeres form at the sides of the gonopore in the ontogeny of some insects. Usually the phallomeres unite to form the phallus.

See Also the Following Articles

Body Size • Integument • Legs • Mouthparts • Muscle System • Segmentation

Further Reading

Chapman, R. F. (1982). "The Insects. Structure and Function." 3rd ed. Hodder & Stoughton, London.

DuPorte, E. M. (1957). The comparative morphology of the insect head. *Annu. Rev. Entomol.* **2,** 55–77.

Gordh, G., and Headrick, D. H. (2001). "A Dictionary of Entomology." CAB Internation, Wallingford, Oxon, U.K.

Hinton, H. E. (1981). "The Biology of Insect Eggs." 3 vols., Pergamon Press, Oxford, U.K.

Matsuda, R. (1969). "Morphology and Evolution of the Insect Abdomen with Special Reference to Developmental Patterns and Their Bearings upon Systematics." Pergamon Press, Oxford, U.K.

Snodgrass, R. E. (1935). "Principles of Insect Morphology." McGraw-Hill, New York and London.

Tuxen, S. L. (1970). "Taxonomist's Glossary of Genitalia in Insects." 2nd ed. Munksgaard, Copenhagen.

Anopheles Mosquito

see *Mosquitoes*

Anoplura

see *Phthiraptera*

Antennae

Catherine Loudon
University of Kansas

Antennae are segmented appendages that function primarily as chemosensory and mechanosensory structures. An insect typically has a single pair of antennae located on its head. Antennae in juvenile insects are often very different in morphology from antennae in adult insects, typically being larger or more elaborate in the adult stage. Adult antennae may be sexually dimorphic, appearing very different in the males and females. Antennae are absent in the wingless insects belonging to the order Protura and may be extremely reduced in size in some holometabolous larvae.

STRUCTURE

The overall shape of most insect antennae is elongate and cylindrical (Fig. 1, top), although elaborations into plumose, lamellate, or pectinate forms have arisen many times in different insect lineages (Fig. 1, bottom). An elongate, cylindrical morphology, probably the ancestral condition for insect antennae, is found in fossil insects and many other arthropods. There are three parts to an insect antenna: the scape, the pedicel, and the flagellum. The scape is the first segment (most proximal) of the antenna, and it is attached to the head by a rim of flexible, intersegmental cuticle. Thus, the scape (and the rest of the antenna) can move with respect to the head. All the antennal segments are similarly joined to each other by thin, flexible cuticle.

The movements of an antenna are controlled in part by one or two pairs of muscles that attach inside the head (such

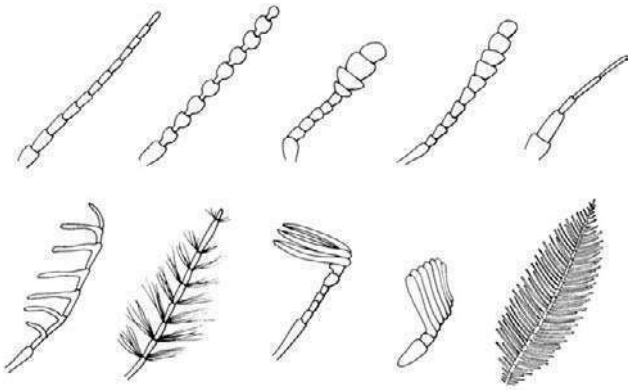

FIGURE 1 Insect antennae exhibit a variety of shapes including elongate morphologies (top) and those with lateral elaborations (bottom). [After Romoser, W. S., and Stoffolano, J. G., Jr. (1998). "The Science of Entomology," WCB/McGraw-Hill, Boston, and Loudon, C., *et al.* (1994). *J. Exp. Biol.* **193,** 233–254, published by McGraw-Hill, with permission of the McGraw-Hill Companies.]

as on the tentorium) with the other end attached inside the scape. An additional pair of muscles runs from the scape to the next segment of the antenna, the pedicel. The combined action of these two sets of muscles is capable of moving an antenna in almost any direction. The final (most distal) segment of the antenna, the flagellum, is the most variable in morphology among insects. The only insects that have intrinsic muscles in the flagellum (joining adjacent segments) are members of the wingless orders Collembola and Diplura. In all other insects (the majority), there are no muscles in the flagellum. Many specialists prefer "annulus" or "subsegment" to "segment" for an individual part of a flagellum in this latter group of insects, because "segment" is reserved for parts with their own musculature. Movements of an annulated flagellum without intrinsic musculature may still occur, such as the spreading and closing of the lamellae or lateral extensions in an antenna (Fig. 1, bottom), but these movements are driven by changes in the pressure of the hemolymph (blood) inside the antenna and thus are hydraulic rather than muscular.

In most insects, circulation of hemolymph through an antenna is facilitated by muscular pumping by an accessory heart located in the head near the base of the antenna. This antennal heart pumps the hemolymph into a blood vessel that discharges the hemolymph at the distal end of the antenna. The return flow of the hemolymph back to the head (and the general open circulatory system of the insect) is not inside a blood vessel. The lumen of an antenna also contains tracheae and nerves, which branch into any lateral extensions of the flagellum. Sensory neurons that send action potentials in response to chemical or physical stimuli sensed by the antennae terminate in the deutocerebrum of the brain. The deutocerebrum is also the site of origin for the motor neurons that stimulate the muscles associated with the antennae.

GROWTH AND DEVELOPMENT

Antennal growth and development in holometabolous insects (those that undergo complete metamorphosis) differs greatly from that in other insects. In holometabolous insects, adult antennae form from imaginal disks, which are clumps of undifferentiated cells that will develop into adult structures. The antennal imaginal disks may appear in the embryonic (fly) or late larval (moth) stage of the immature insect. Properties of the antennal imaginal disks determine to a large extent the chemical stimuli to which an adult will respond, as is seen from experiments in which antennal imaginal disks were cross-transplanted between larvae, which were then reared to adulthood and assayed.

In hemimetabolous and apterygote (wingless) insects, the nymphs are very similar in overall form and habit to the adults, and their antennae resemble smaller, shorter versions of the adult antennae. As with all external structures that are replaced at each molt, a new antenna is formed inside the old antenna. The primary morphological change that occurs at each molt is that the flagellum lengthens with the addition of more segments or annuli, either at the distal end (orders Collembola and Diplura), the proximal end (most other insects), or along the length of the flagellum (some members of the orders Orthoptera and Odonata).

Antennae are serially homologous to mouthparts and legs, reflecting the ancestral condition of a single pair of appendages per body segment shared by arthropods and related groups. Common developmental features between legs and antennae can be seen, for example, in the action of the homeotic gene called *Antennapedia,* which results in the substitution of leglike appendages for antennae on the head when expressed ectopically in mutant *Drosophila.* Leglike appendages appearing in the antennal location in adult insects have also been observed after regeneration of antennae following injury during the larval stage (Fig. 2).

FUNCTION

The primary function of antennae is the assessment of the chemical and physical characteristics of the environment. Detection is made with innervated chemosensory and mechanosensory organs that are arrayed on the antennae. A single antenna usually has sensory organs of several types, with different properties. Most of the chemosensory organs are located on the flagellum and often take the form of microscopic chemosensory hairs (sensilla) each only 1 or 2 μm in diameter. Some antennae, such as the feathery pectinate antennae of silkworms (*Bombyx mori*), have tens of thousands of sensilla, which are capable of very thoroughly sampling the air that passes in the small spaces between them. A cockroach antenna may have hundreds of thousands of sensilla. The chemicals that may be detected by chemoreceptors on the

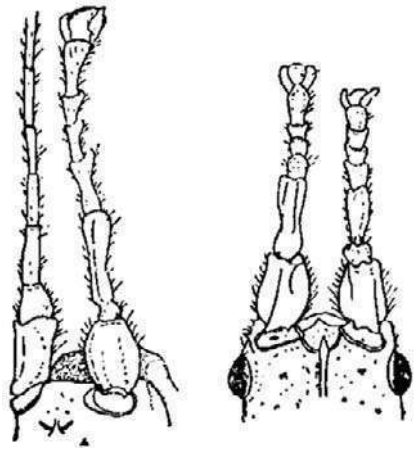

FIGURE 2 Left: head of an adult Indian stick insect (*Carausius morosus*) with a normal antenna on the left and a regenerated antenna with leglike morphology on the right. Right: head of adult *C. morosus* with two regenerated antennae with leglike morphology. [After Fig. 78 in Wigglesworth, V. B. (1971). "The Principles of Insect Physiology." Chapman & Hall, London, with the kind permission of Kluwer Academic Publishers.]

antennae are usually biological in origin and airborne (volatiles), although (depending on the insect species) the sampled chemical compounds are sometimes in a liquid or associated with a solid surface. The chemicals intercepted by antennae may alert the insect to the presence of prospective mates, food, suitable places to lay eggs, or predators.

The physical stimuli detected by mechanoreceptors on the antennae may be used by the insect to indicate air speed during flight, to detect vibrations of the air, or to detect solid boundaries in its environment by touch. While a single mechanosensory hair will send information to the brain about the local physical conditions existing at its microscopic location, an antenna also has mechanosensory organs that evaluate the physical forces acting on the antenna as a whole. These mechanosensory organs, located near the base of the antenna, include Johnston's organ, Böhm bristles, hair plates (groups of mechanosensory hairs), and campaniform sensilla (thin flexible patches of cuticle that are innervated). Johnston's organ is located in the pedicel and responds to changes of location or vibrations of the whole antenna. In contrast, the Böhm bristles, located near the scape–pedicel boundary, send information to the brain about the antennal position, rather than its movements. The variety of mechanosensory organs associated with the first two segments of the antennae are believed to act together to inform a flying insect about its air speed, because greater flying speed will cause greater deflection of the antennae by the air rushing past. Contact chemosensory hairs, so called because the chemical compounds are usually detected when the insect is touching a liquid or solid surface with the antennae, often have mechanosensory capabilities as well and are usually located near the distal ends of antennae.

The function of the antennal sensory organs will be affected by their arrangement on the antennae. For example, sensory organs on the distal tip of a very long antenna will permit chemical or physical sampling of the environment far from the body of the insect. Close packing of sensory hairs will decrease the airflow in their vicinity, and hence will modify both the chemical and physical sampling of the environment by those hairs. The function of the antennae will also be dependent on the behaviors of the insect that will affect the airflow around the antennae, such as flying, wing fanning, postural changes, or oscillating the antennae. A structure projecting into the environment is liable to collect debris that might interfere with its sensory function; both antennal grooming behaviors and modifications of leg parts against which an antenna is scraped are common in insects (Fig. 3). In some insects, antennae are modified for nonsensory functions such as clasping mates during copulation (fleas and collembolans), holding prey items (beetle larvae), or forming a temporary physical connection between an underwater air reservoir and the atmosphere (aquatic beetles).

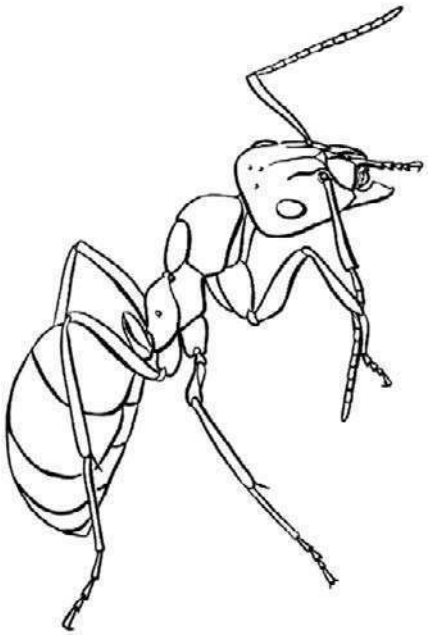

FIGURE 3 A worker ant (*Formica polyctena*) cleans one of its antennae by dragging it across the specialized comb of right foreleg. (Reprinted by permission of the publisher from THE INSECT SOCIETIES by Edward O. Wilson, Cambridge, MA.: The Belknap Press of Harvard University Press, Copyright 1971 by the President and Fellows of Harvard College. Original drawing by Turid Hölldobler.)

See Also the Following Articles

Chemoreception • Imaginal Discs • Mechanoreception • Pheromones

Further Reading

Hansson, B. S., and Anton, S. (2000). Function and morphology of the antennal lobe: New developments. *Annu. Rev. Entomol.* **45,** 203–231.

Heinzel, H., and Gewecke, M. (1987). Aerodynamic and mechanical properties of the antennae as air-current sense organs in *Locusta migratoria.* II. Dynamic characteristics. *J. Comp. Physiol. A* **161,** 671–680.

Kaissling, K. E. (1971). Insect olfaction. *In* Handbook of Sensory Physiology (L. M. Beidler, ed.), Vol. IV of "Chemical Senses," Part 1, "Olfaction," pp. 351–431. Springer, Verlag, Berlin.

Keil, T. A. (1999). Morphology and development of the peripheral olfactory organs. *In* "Insect Olfaction" (B. S. Hansson, ed.), pp. 5–48. Springer-Verlag, Berlin.

Loudon, C. and Koehl, M. A. R. (2000). Sniffing by a silkworm moth: Wing fanning enhances air penetration through and pheromone interception by antennae. *J. Exp. Biol.* **203,** 2977–2990.

Pass, G. (2000). Accessory pulsatile organs: Evolutionary innovations in insects. *Annu. Rev. Entomol.* **45,** 495–518.

Schneider, D. (1964). Insect antennae. *Annu. Rev. Entomol.* **9,** 103–122.

Schneiderman, A. M., Hildebrand, J. G., Brennan, M. M., and Tumlinson, J. H. (1986). Transsexually grafted antennae alter pheromone-directed behaviour in a moth. *Nature* **323,** 801–803.

Steinbrecht, R. A. (1987). Functional morphology of pheromone-sensitive sensilla. *In* "Pheromone Biochemistry" (G. D. Prestwich and G. J. Blomquist, eds.), pp. 353–384. Academic Press, London.

Zacharuk, R. Y. (1985). Antennae and sensilla. *In* "Comprehensive Insect Physiology Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert eds.), Vol. 6, pp. 1–69. Pergamon Press, New York.

Ants

Nigel R. Franks
University of Bristol

The ants comprise a single family, the Formicidae, within the superfamily Vespoidea and the order Hymenoptera. There are 16 extant subfamilies of ants with a total of 296 extant genera. Some 9000 to 10,000 species of ants have been described, and it is estimated that there may be 15,000 species of ants alive in the world today. The earliest known fossil ants are from the Cretaceous (ca 105–110 mya), but ants probably did not become common until the Eocene (ca 45 mya).

EVOLUTION AND ECOLOGICAL SUCCESS

Ants are now extremely successful ecologically. There may even be an equal biomass of ants and humanity in the world today. They dominate, at their size scale, many terrestrial ecosystems from latitudes north of the boreal tree line to such southern climes as Tierra del Fuego, Chile. In certain tropical forests the contribution of ants to the biomass is spectacular. In Brazilian rain forests, for example, the biomass of ants has been estimated as approximately four times greater than the biomass of all of the vertebrates combined.

One of the reasons ants are so successful is that their colonies have extremely efficient divisions of labor: they evolved factories millions of years before we reinvented them. Another reason is that they can modify their immediate environment to suit themselves, much as we do. Leafcutter ants (*Atta*), for example, evolved agriculture tens of millions of years before humanity developed agronomy. Furthermore, leafcutter ants also use antibiotics and symbiotic bacteria to protect the crop of fungi they grow on the leaves they collect. By contrast, weaver ants (*Oecophylla*) fashion homes from living leaves by sowing them into envelopes, using their larvae as living shuttles and the silken thread they produce as glue. Ants can also dominate areas by mobilizing large numbers of well coordinated foragers; indeed, an ant colony's foragers can be so numerous and well organized that they give the impression of being everywhere at once.

Ants can also be important as seed distributors and as seed harvesters, in the turnover of soils, and in the regulation of aphid numbers and the minimization of outbreaks of defoliating insects. Economically important pest species include the imported fire ant (*Solenopsis invicta*) in North America and leafcutter ants (such as *Atta*) in the neotropics. There are also many ecologically destructive "tramp" ants or invasive species that have been distributed to alien habitats by human commerce.

Ants and plants often have closely coupled ecological relationships. Certain plants even encourage ants by producing rewards such as energy-rich elaiosomes on their seeds to encourage seed dispersal, nutritious Beltian bodies and extrafloral nectaries to entice ants to visit their leaves and shoots (hence to remove the plant's natural enemies while there), or even by supplying preformed homes (domatia) to invite ants directly to inhabit and thus better protect them. Although many ants are hunter-gatherers, very many species tend aphids for the excess honeydew they excrete. By "milking" aphids in this way, ants can in effect become primary consumers of plant products and by thus operating at a lower trophic level they can build up a larger biomass than obligate carnivores would be able to do. Yet most ants mix their diet by also consuming animal protein; for example, they will devour their own aphid milk cows if the latter become sufficiently abundant.

Arguably, the best evidence of the ecological success of ants is that their worst enemies are other ants.

EUSOCIALITY, SOCIAL ORGANIZATION, AND SOCIAL DIVERSITY

Except for a few species that have secondarily lost the worker caste, all ants are eusocial: they have an overlap of adult generations, cooperative brood care, and reproduction dominated by a minority of the colony's members. Typically, an established ant colony consists of one or more queens (each of which may have mated with one or more winged males on a nuptial flight), an all-female set of wingless workers, and the colony's brood of eggs, larvae, and pupae. The majority of queens mate only before they establish a colony. Thereafter, they store the sperm they have received.

All ants have haplodiploid sex determination. This property probably had a major role in the evolution of their eusociality through kin selection. Males are haploid, having only a single set of chromosomes, and thus the sperm that individual males produce is genetically homogeneous. Hence, the (diploid) daughters of the same mother and father are unusually closely related to one another, a circumstance likely to have favored the evolution of female workers. Nevertheless, there can be continuing conflicts within colonies between the workers and the queen (or queens) over the sex ratios they produce and which colony members produce the males. Queens can choose to produce either unfertilized (haploid) eggs destined to become males or fertilized (diploid) eggs. The latter may develop into workers or potential new queens (gynes) generally depending on how much food they receive as larvae. The workers may or may not be sterile. Fertile workers produce viable (unfertilized) haploid eggs that can develop into males. Hence, there can be conflict both among the workers and between the workers and the queen over whose sons the colony produces. Indeed, in many species of ants with only small numbers of workers in their mature colonies, there are dominance hierarchies among the workers, who fight one another over egg production. Sometimes the queen moves with active aggression against the most dominant worker to curtail its production of sons in favor of her own. In addition,

even when workers are sterile and serve one, singly mated queen, they may prefer to raise more of the queen's daughters, to whom they are more closely related, than the queen's sons. For all these reasons, the study of ants has had a major impact in recent pioneering evolutionary biology because these insects provide test cases by which the evolutionary resolution of the tension between cooperation and conflict can be explored. It is clear, though, that the apparent social cohesion of ant colonies is often partly an illusion.

Among ants, there is a diversity of mating systems and social organizations. So even though it is tempting to think of the typical ant colony as having a single, singly mated queen and occupying a single nest site, the diversity of social systems among the ants is in fact huge. For example, many ant species consist of facultatively multiqueened (polygynous) colonies. Indeed, roughly half of European ant species exhibit polygyny, and there seems to be no reason to regard this as an unusual proportion. Some ant colonies are founded by solitary queens; some by groups of unrelated queens that may later fight over who will be the one to succeed. Other colonies simultaneously occupy multiple nests (polydomy), a habit often associated with polygyny, while others exhibit colony fission, with both daughter colonies usually being monogynous. Most persistent polygyny is associated with the secondary adoption of queens. Unusual social systems include queenless ants, workerless ants (inquilines), and slave-making ants. In certain queenless species, the workerlike females produce other diploid females through a parthenogenetic process called thelytoky. By contrast, certain inquilines have dispensed with the worker caste, and queens infiltrate and exploit established colonies of other species. Slave making may occur both intraspecifically and interspecifically. Interspecific slave making is also associated with nonindependent colony foundation in which slave-maker queens infiltrate established colonies of their host species, kill the host queen or queens, and produce workers that are reared by currently available host workers. The slave-maker workers raid other neighboring host colonies to capture large larvae and pupae. Such raids thus replenish the stocks of slave workers, which do all the foraging and brood rearing for the slave makers. There are also ant species in which there are polymorphic queens, others in which there are polymorphic males, and many in which there are polymorphic workers.

One of the outcomes of eusociality is that established colonies can be well defended by the workers against enemies. Thus, ant colonies are relatively K-selected; that is, they are selected to hold onto resources and to persist for long periods rather than being ephemeral, here-today-gone-tomorrow, r-strategists. Associated with this trait is the extreme longevity of ant queens. It is estimated that they can live 100 times longer than other solitary insects of a similar size. Worker populations in mature, well-established monogynous colonies range from a few tens of millions to 20 million, and certain so-called supercolonies consist of a huge network of linked nests each with many queens. One supercolony of *Formica yessensis* in Japan may have as many as 300 million workers. Given such longevities and densities, it is clear that ants may also prove to be important model systems for understanding the spread of disease or the evolution of mechanisms to minimize the spread of disease among viscous populations of close kin. It is even possible that polygyny and multiple mating (polyandry) have evolved, at least in part, to promote genetic heterogeneity within colonies and thus help to minimize disease risks.

DIVISION OF LABOR

The relatively large biomass of ants in many ecosystems can be attributed not just to the way in which the ants interact with other organisms but to the way in which they interact with their nestmates in general and, in particular, to efficiencies that accrue from divisions of labor. One of the most dramatic traits associated with the division of labor among the workers is physical polymorphism, which is the presence of different physical worker forms within the same colony. In the African army ant, *Dorylus wilverthi,* for example, the smallest workers at 0.12 mg dry weight are only 1% of the dry weight of the largest workers (soldiers), and this relatively great size range is exceeded in certain other species (e.g., in *Pheidologeton diversus,* the smallest workers have a dry weight that is about 0.2% that of the largest majors). It is not just the size range that is impressive in such species but also the degree of polymorphism among the workers. Darwin, writing in *The Origin of Species,* seemed well aware not only of the phenomenon but also of its implications. Indeed, one of Darwin's most penetrating insights in his 1859 masterpiece was his suggestion that sterile forms evolved in social insects because they are "profitable to the community" and that "selection may be applied to the family, as well as to the individual." He further suggested that once such colony-level selection had begun, the sterile forms could be molded into distinct castes "Thus in [the army ant] *Eciton,* there are working and soldier neuters, with jaws and instincts extraordinarily different" (Fig. 1a, b).

Such worker polymorphism is now known to be associated with the differential growth rates of different putative tissues and body parts during the preadult stages. Indeed, the study of ants made a major contribution to the development of the concept of allometric growth (Fig. 1c, d). Notably polymorphic genera include the army ants *Eciton* and *Dorylus,* leafcutter ants (*Atta*), carpenter ants (*Camponotus*), and members of the genera *Pheidole* and *Pheidologeton.* Indeed, *Camponotus* and *Pheidole* are the two most species-rich ant genera.

However, genera with polymorphic workers are in the minority. Approximately 80% of ant genera consist entirely of species with monomorphic workers, most of the remaining genera consist of species in which there are at most only two easily recognizable worker morphs, and only about 1% of genera have species in which three or more worker morphs can be relatively easily recognized within colonies.

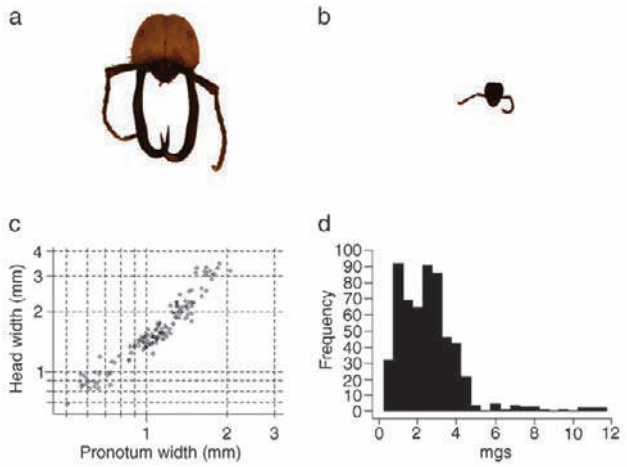

FIGURE 1 The army ant, *Eciton burchelli.* (a) Head of major worker. (b) Head of minor worker. (c) Head width vs ponotum width allometry for workers. (d) Frequency–dry weight histogram for a large sample of workers. The allometrical relationship has a slope greater than 1, so larger workers (such as majors) have disproportionately large heads. The size frequency distribution is skewed to the right so relatively few of these very large majors are produced. (Drawings © Nigel R. Franks.)

Polymorphism among the workers is mostly associated with extreme physical specialization. Thus, *Eciton* majors have ice-tong-like mandibles and are specialist defenders of the colony against would-be vertebrate predators or thieves (Fig. 1a). It has been shown that colonies of *Pheidole pallidula* can produce more defensive majors in response to stresses induced by conspecific competitors. Majors are not always for defense: large-headed majors in *Pheidole* and *Messor* serve as specialist grinders of harvested seeds. Even among such polymorphic species, however, the majority of workers belong to castes of generalists, which give their colonies an ability to respond rapidly to changes in the environment. Such generalists show behavioral flexibility not possible with the extreme morphological specialization of certain physical castes. Nevertheless, divisions of labor also occur within the majority generalist caste. Such workers typically specialize in different tasks at different times during their lives. This is known as temporal polyethism, in contrast to physical polyethism.

The sophisticated divisions of labor in monomorphic ants are being investigated. In *Leptothorax albipennis,* the workers show very little size variation, and colonies consist of, at most, a few hundred such workers living in flat crevices between rocks. Such crevices can be only 2 or 3 cm wide and deep and may have an internal cavity height of only 1 or 2 mm. Individual workers could easily roam all around such nests within a minute, but instead they have spatial fidelity zones; that is, they remain faithful to certain parts of the nest and the segregated tasks within such areas for months on end. The workers can even reconstruct their own spatial fidelity zones relative to one another if, and when, their colony is forced to emigrate to a new nest site because of the destruction of the old site. In this (and many if not all) ant species, younger workers tend to work deep within the nest at its safe center, tending the queen and the eggs. As they get older, workers tend to move progressively out from the center of the nest, and toward the end of their lives they eventually engage in the most dangerous task of foraging in the outside world, where they are likely to meet predators and other hazards. However, the correlation between age and task is often very weak, and in an increasing number of species it has been shown that the division of labor among monomorphic workers is extremely flexible. Workers can respond to the removal of other workers by reverting to tasks that they did earlier in their lives or, if need be, they may begin foraging even when they are very young. Thus, though age may influence what workers do, it is unlikely to be the organizing principle of the division of labor in many species. Rather, it seems that workers are continuously monitoring their workloads and the delays they experience while waiting to interact with their nestmates and will flexibly change their tasks accordingly to maximize their productivity.

COMMUNICATION AND PHEROMONES

Ants have diverse systems of communication, but by far the most important medium for signaling involves the chemicals known as pheromones. Ants can deposit chemical trails to recruit nestmates to discoveries of food. Many ants can also produce highly volatile chemicals to signal alarm when they encounter dangerous predators or other hazands. Different ants in different subfamilies use a remarkable diversity of glandular structures even just to produce recruitment pheromones. These may be produced from cloacal glands, Dufour's glands, the hindgut, poison glands, pygidial glands, rectal glands, sternal glands, or even tibial glands on the back legs. Furthermore, many pheromones appear to be complex mixtures of many chemical compounds.

Pheromones can be effective in minute quantities; it has been estimated that one milligram of the trail substance of the leafcutting ant, *Atta texana,* if laid out with maximum efficiency, would be sufficient to lead a colony three times around the world.

Nestmate recognition is another important aspect of communication in ants. A pleasing metaphor for the ant colony is a factory inside a fortress. Ant colonies are dedicated to the production of more ants; but workers need to "know" that they are working for their natal colony, and colonies also need to be well defended against other ants and against infiltration by other arthropods, which might tap into their resources. Ant colonies employ colony-specific recognition cues as one of their defense systems. These are often in the form of cuticular hydrocarbons that can be spread throughout the colony both by grooming and trophallaxis (the latter is usually associated with liquid food exchange). Slave-making ants circumvent the recognition cues of their slaves by capturing them as larvae and pupae—these captives are not yet imprinted on their natal colony odor but later become imprinted on the odor of the

FIGURE 2 Scanning electron micrograph of a worker of *Lasius flavus* with a kleptoparasitic mite, *Antennophorus grandis,* gripping on its head. The mite steals food when two workers exchange nutritious liquids during trophallaxis. (Photomicrograph © Nigel R. Franks.)

colony that kidnapped them after they have metamorphosed into adult workers. Sometimes colony-specific odors also can be influenced by chemicals picked up from the colony's environment. Nevertheless, countless species of arthropods from mites to beetles have infiltrated ant colonies. For example, more than 200 species of rove beetle (Staphilinidae) are associated with New World army ants alone, and other groups such as mites are probably even more species rich. Often these infiltrators are called "guests" simply because their relationships with their host ant colony and to its resources are unknown (Fig. 2).

SELF-ORGANIZATION, COLLECTIVE INTELLIGENCE, AND DECISION MAKING

A rapidly developing approach to the study of ants and other social insects is the application of self-organization theories. Here self-organization can be defined as a mechanism for building spatial structures and temporal patterns of activity at a global (collective or colony) level by means of multiple interactions among components at the individual (e.g., worker) level. The components interact through local, often simple, rules that do not directly or explicitly code for the global structures. The importance of studies of such self-organization is that they can show how very sophisticated structures can be produced at the colony level with a fully decentralized system of control in which the workers have no overview of the problems they are working to solve.

A simple and very intuitive example of how ants use self-organization is found in their ability to select short cuts. Certain ants can select the shortest paths to food sources. Indeed, where there is a short and a long path to the same food source, the decision-making mechanism can be surprisingly simple. The ants that happen to take the shorter path get there and back more quickly than the ants that happen to take the longer path. All the ants lay attractive trail pheromones, and such pheromones are reinforced more rapidly on the shorter path simply because that path is shorter and quicker. In such cases, individual ants do not directly compare the lengths of the two paths, but the colony is able to choose the shorter one. Sometimes the shorter path is used exclusively, while at other times a small amount of traffic may continue to use the longer path. Having some traffic that continues to use the longer path is likely to be costly in the short term, but it may represent a beneficial insurance policy if the shorter path becomes blocked or dangerous. Self-organization also has a major role in such phenomena as brood sorting, rhythms of activity within nests, and building behaviors. This new approach may help to answer, at least in part, the age-old challenge of how ant colonies are organized.

See Also the Following Articles

Caste • Colonies • Nest Building • Pheromones • Sex Determination • Sociality

Further Reading

Bolton, B. (1994). "Identification Guide to the Ant Genera of the World." Harvard University Press, Cambridge, MA.

Bourke, A. F. G., and Franks, N. R. (1995). "Social Evolution in Ants." Monographs in Behavioral Ecology. Princeton University Press, Princeton, NJ.

Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraulaz, G., and Bonabeau, E. (2001). "Self-Organization in Biology." Princeton University Press, Princeton, NJ.

Detrain, C., Deneubourg, J.-L., and Pasteels, J. (1999) "Information Processing in Social Insects." Birkhäuser Verlag, Basel, Switzerland.

Grimaldi, D., and Agosti, D. (2000). A formicine in New Jersey Cretaceous amber (Hymenoptera: Formicidae) and the early evolution of the ants. *Proc. Natl. Acad. Sci. USA* **97,** 13678–13683.

Hölldobler, B., and Wilson, E. O. (1990). "The Ants." Belknap Press, Cambridge, MA.

Keller, L., and Genoud, M. (1997). Extraordinary lifespans in ants: A test of evolutionary theories of ageing. *Nature* **389,** 958–960.

Passera, L., Roncin, E., Kauffmann, B., and Keller, L. (1996). Increased soldier production in ant colonies exposed to intraspecific competition. *Nature* **379,** 630–631.

Schmidt-Hempel, P., and Crozier, R. H. (1999). Polyandry versus polygyny versus parasites. *Philos. Trans. R. Soc. (Lond) (B)* **354,** 507–515.

Sendova-Franks, A. B., and Franks, N. R. (1994). Social resilience in individual worker ants and its role in division of labour. *Proc. R. Soc. (Lond) (B)* **256,** 305–309.

Aphids

John T. Sorensen

California Department of Food and Agriculture

Aphids are remarkable, evolutionarily exquisite creatures, and among the most successful insects. Aphid evolution

has been shaped through nutrient-driven selection and by the host plants on which they feed, and aphids have responded by developing intricate life cycles and complex polymorphisms. These sap-feeding hemipterans have coped with a hostile world through developing an exceptionally high reproductive rate and passive wind-borne dispersal, a strategy in which individuals are quite expendable, but survival and prosperity of their genes are guaranteed. Because of their intriguing evolutionary adaptations, aphids were among our most worthy competitors as humans entered the agricultural era.

MAJOR GROUPS AND HOST AFFILIATIONS

Aphids, as the superfamily Aphidoidea, belong to the Sternorrhyncha within the Hemiptera, a group they share with Aleyrodoidea (whiteflies), Psylloidea (jumping plant lice), and Coccoidea (scale insects and mealybugs). Aphidoidea has three families: Adelgidae (adelgids), Phylloxoridae (phylloxorids), and Aphididae (aphids), although some workers place the Adelgidae and Phylloxoridae in a separate superfamily, Phylloxoroidea. Adelgids and phylloxorids are primitive "aphids" and older groups, each with about 50 species. They differ from Aphididae by having an ovipositor and by reproducing by means of ovipary. Adelgids are restricted to conifers (Pinaceae), and some form characteristic galls (e.g., *Adelges piceae,* balsam woolly adelgid). Phylloxorids, which may also form galls, occur on plants of the Salicaceae (willow family), Fagaceae (oak family), Juglandaceae (walnut family), and Rosaceae (rose family). An exceptional species, *Daktulosphaira vitifolae,* grape phylloxera, feeds on grapes (Vitaceae), damaging European grape cultivars unless they are grafted to resistant rootstocks developed from American grape species.

Aphids originally evolved on woody plants in the Northern Hemisphere and are functionally replaced by whiteflies and psyllids in the Southern Hemisphere. As a group, they evolved and began their diversification with angiosperms, over 140 mya during the lower Cretaceous. While most fossil aphid groups became extinct during the Cretaceous–Tertiary boundary, most modern aphid groups radiated during the Miocene. Aphids have siphunculi, which vary by group from being mere pores on the abdominal surface to being very elongate tubes. They also have a cauda, which varies by group from rounded and hardly noticeable to knobbed or long and fingerlike. Aphids lack an ovipositor and are viviparous, bearing young parthenogenetically.

Aphid taxonomy is difficult; their subfamily classification has been argued and confused with nearly as many classifications as aphid taxonomists. Remaudiere and Remaudiere's 1997 classification, followed here, recognizes about 25 aphid subfamilies, with tribal groupings for about 600 genera and 4700 species of aphids. Many aphid lineages coevolved with, and radiated among, their host plant groups. Often during their phylogenetic history, however, aphid groups opportunistically switched to radically unrelated host groupings, driven by developmental requirements but tempered by evolutionary constraints.

Many aphid subfamilies are small, but several are larger and important: Chaitophorinae (e.g., *Sipha flava,* yellow sugarcane aphid), on Salicaceae and Gramineae (grass family); the closely related Myzocallidinae (e.g., *Therioaphis trifolii* f. *maculata,* spotted alfalfa aphid), Drepanosiphinae (e.g., *Drepanaphis acerifoliae,* painted maple aphid), and Phyllaphidinae (e.g., *Phyllaphis fagii,* beech aphid), often considered to be one subfamily and usually on dicotygledonous trees, but also Fabaceae (legume family) and bamboo; Lachninae (e.g., *Essigella californica,* Monterey pine aphid), mostly on Pinaceae, but also Fagaceae, Rosaceae, and Asteraceae (composite family) roots; and Pemphiginae (e.g., *Pemphigus bursarius,* lettuce root aphid), often on roots and host alternating to dicotyledonous trees forming galls. Other noteworthy subfamilies include Pterocommatinae, on Salicaceae; Greenideinae, on Fagaceae; Mindarinae, on Pinaceae; and the host-alternating Anoeciinae and Hormaphidinae, the latter causing galls.

The largest and most evolutionarily recent subfamily, Aphidinae, has two large, diverse, and agriculturally important tribes. The first tribe, Macrosiphini (e.g., *Aulacorthum solani,* foxglove aphid), is diverse in genera, which often lack attendance by ants but may alternate hosts. The second tribe, Aphidini, is diverse in species but less so in genera; these are often attended by ants. Tribe Aphidini has two important subtribes. Subtribe Rhopalosiphina (e.g., *Rhopalosiphum padi,* bird cherry–oat aphid) host alternates between Rosaceae to Gramineae or Cyperaceae (reed family). Subtribe Aphidina (e.g., *Aphis fabae,* bean aphid) host alternates mostly among Rosidae and Asteridae and is home to genus *Aphis,* which alone contains well over 1000 species.

NUTRITION-DRIVEN EVOLUTION: LIFE CYCLES AND POLYMORPHISM

Aphid life cycles are complex and may be either monoecious or dioecious, involving holocycly or anholocycly. Because of this, aphids have evolved many specialized morphs; a multitude of confusing, often synonymous names have risen among aphid biologists, but these are minimized here.

In the simple and generalized monoecious holocyclic aphid life cycle (Fig. 1A), a single host plant species is used throughout the year and sexual morphs are produced in the fall, usually in response to decreasing daylength. The males and oviparae mate, producing genetically recombinant eggs that overwinter on the host plant and often experience high mortality. In the spring, the fundatrix emerges from the egg, matures parthenogenetically, and gives live birth to nymphs that become viviparae and continue in that reproductive mode through the summer. If the aphid group produces plant galls, the fundatrix is responsible for their production.

The viviparae may be apterae or alatae (Fig. 2), but in some groups (e.g., Drepanosiphini, some Myzocallidinae) all viviparae are alatae. The parthenogenetic reproduction of

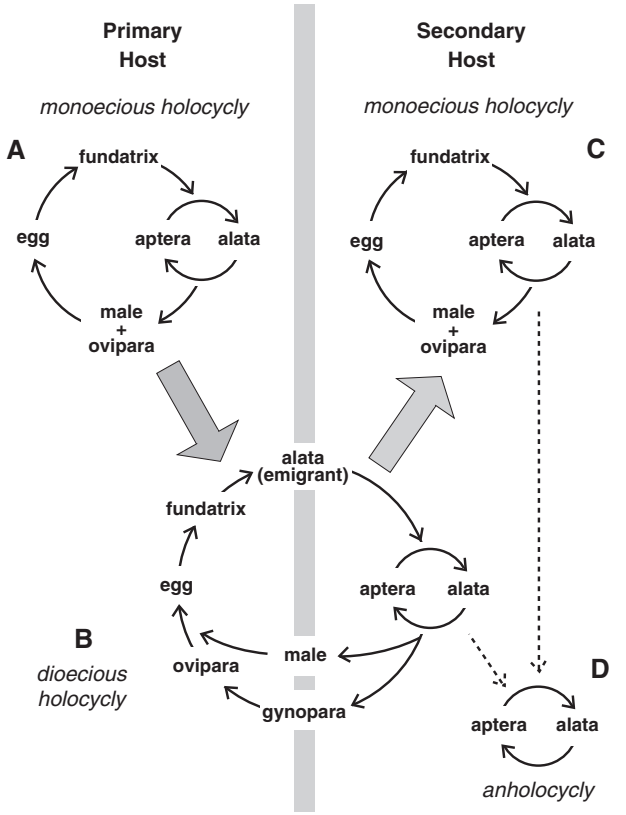

FIGURE 1 Evolutionary development of generalized aphid life cycles. Initially, aphids developed monoecious holocycly (A) on an ancestral woody primary host, where aestivation occurred because sap amino acids were unavailable during summer growth cessation. Next, multiple subfamilies independently evolved dioecious holocycly (B), where viviparae moved to summer-growing herbaceous secondary hosts but returned to their ancestral host in autumn. In some aphids, secondarily monoecious holocycly (C) developed on the secondary host when the primary host was lost. Often in warm areas, where selection for an overwintering egg is not imposed, some populations of dioecious and secondarily monoecious holocyclic aphids may lapse into facultative anholocycly (D) on their secondary hosts; this condition may become obligate anholocycly if the ability to produce sexuals is lost.

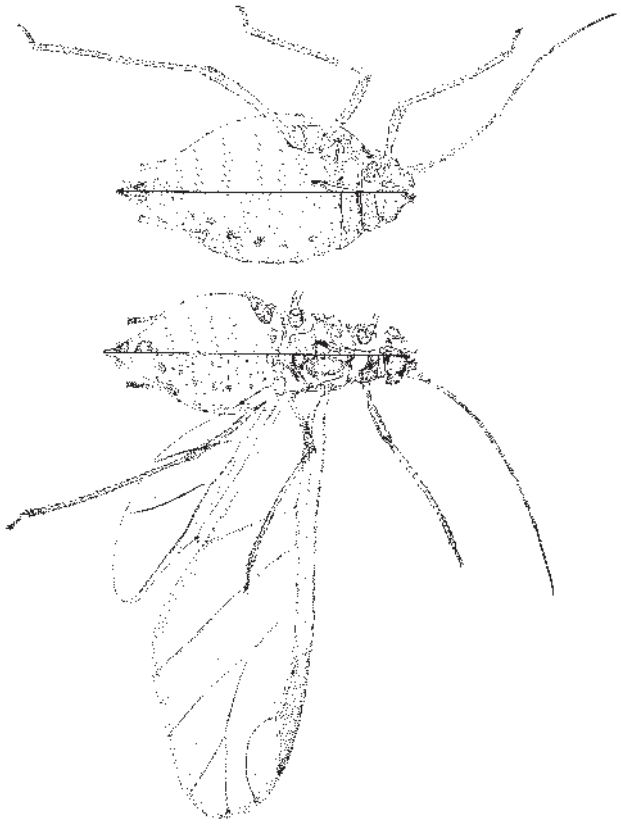

FIGURE 2 Aptera (top) and alata (bottom) viviparae of *M. persicae.* Shown in split images with ventral (upper half) and dorsal (lower half) aspects with heads to the right. [Drawings by Tokuwo Kono, modified from Kono, T., and Papp, C. S. (1977). "Handbook of Agricultural Pests, Aphids, Thrips, Mites, Snails, and Slugs." California Department of Food and Agriculture, Sacramento.]

viviparae allows very rapid buildup of numbers and collapse of generation time. When a viviparous nymph is born, it has the embryos of both its daughters and granddaughters within it, creating a "telescoping" of generations. Apterae have lost their wings and associated musculature to optimize reproduction. They produce more offspring per female than do alatae, which must invest resources in their flight apparatus. However, alatae produce progeny earlier in life than do apterae, giving their relatively reduced number of offspring a better generational turnaround time than apterae mothers can. Apterae are selectively produced when the host plant is a good source of nutrients. Once an aphid population has built, either inducing a crowding effect among apterae or stressing its host to the level of impacting nutrient levels, the population usually switches to produce alatae, which migrate to better situations. However, the risks of successful migration are great, especially for monophagous aphids that feed on uncommon hosts, because the flight of alatae is wind-borne and relatively passive. Alatae can be blown over 1600 km, often across an ocean, and survive the trip. Upon successfully alighting on their proper host and feeding for a short time, alatae begin autolysis of their flight musculature, precluding further flight but self-cannibalistically providing nutrients for their offspring. The production of viviparae continues until fall conditions trigger production of the sexuals.

A second, more complicated dioecious life cycle (Fig. 1B) has independently evolved among several different aphid groups that show seasonal alternation between differing hosts. This dioecious cycle probably evolved in response to the seasonally inadequate supply of nitrogen-based nutrients, especially amino acids, on their primary host. The phloem sap that aphids feed on has limited nitrogen availability, and nitrogen is the limiting nutrient in aphid development. Woody deciduous plants normally translocate amino acids in quantity only during the spring, when they are foliating, and in the fall, when leaf senescence breaks down leaf protein and nitrogen is translocated to the roots for overwinter storage. Aphids groups evolving on and restricted to such plants face

a nitrogen deficit during the summer, when active plant growth ceases and phloem sap is low or devoid of nitrogen. Such groups (e.g., *Periphyllus* spp.) may develop an aestivating nymph that halts growth until fall. Other aphid groups (e.g., Aphidinae) whose ancestors originated on deciduous woody plants, have evolved to leave those primary hosts during the late spring, after the nitrogen flush associated with foliation has ceased. In doing so, their spring alatae, as emigrants, migrate to herbaceous secondary hosts that actively grow and transport nitrogen during the summer. In the fall, however, as their secondary hosts die back, the aphids return to their woody primary host by producing migrating males and gynoparae. There, the aphid's sexuals, its males and oviparae, capture that host's fall nitrogen flush and mate to lay their overwintering eggs in anticipation of the spring nitrogen flush. Depending on the aphid or its group, the secondary hosts may vary from quite specific to a broad number of botanical groups; but the primary hosts are often specific to a plant genus. Most aphid lineages have adapted specific types of secondary hosts, such as grasses (e.g., *Metopolophium dirhodum,* rose-grain aphid), roots (e.g., *Smynthurodes betae,* bean root aphid), other woody plants (e.g., *Hormaphis hamamelidus*), or herbs (e.g., *Macrosiphum rosae,* rose aphid). Some aphids specialize on secondary hosts of a particular environmental ecotype; for example, *Rhopalosiphum nymphae,* waterlily aphid, uses aquatic plants in many plant families.

Some aphid lineages (e.g., *Schizaphis graminum,* greenbug) have evolved beyond dioecious holocycly, entirely leaving their primary host to remain on their secondary host, in secondarily monoecious holocycly (Fig. 1C). However, an important form of year-round residence on the secondary host occurs in warmer climates, where populations do not require an egg for overwintering survival. Under such conditions, otherwise holocyclic dioecious or monoecious populations may lapse facultatively into anholocycly on their secondary hosts (Fig. 1D). If such populations remain anholocyclic long enough, they may eventually evolve into obligate anholocycly by losing the ability to produce sexual morphs, despite undergoing environmental conditions that normally trigger their production. Depending on the aphid lineage and its adaptation to its host(s) or their alternation, nearly all aphid morphs may be winged or wingless, but the morph's wing condition is specific to the aphid group.

APHID BEHAVIOR

Aphids feed by inserting their rostrum-borne stylets into a plant and ratcheting them between plant cells, seldom penetrating any until they enter the phloem sieve tubes and extract sap. Stylet advancement is lubricated by saliva containing a pectinase that loosens the bonding between plant cells. The saliva forms a stylet sheath that is left in the plant when the stylets are withdrawn. To cope with a sap diet, aphid guts have specialized groups of cells, mycetomes, containing rickettsia-like symbiotic bacteria, mycetocytes, which aid in synthesis of nutrients. These bacteria, which are passed from mother to embryonic daughter, have coevolved with aphid lineages, differing among them.

Whereas aphids largely rely on a high reproductive rate and great dispersive ability to maximize survival in a hostile environment, morph-specific behaviors exist to promote genetic survival of the individual or its clone. Behavior of alatae optimizes dispersion and finding a successful host. When alatae initially take to flight, they are attracted to the short wavelengths of light that predominate in a clear or cloudy sky, and fly up toward them. After flying a while, however, they come to prefer the longer light wavelengths reflected from plants, and they descend, moving to them. In some species, alatae have shown a preference for colors characteristic of their host plant's leaves. Generally alatae are attracted to yellow, a predominant hue in growing or senescent plants, which are better nitrogen sources. Upon alighting on a plant, they briefly probe below the epidermis with the rostrum to locate specialized secondary plant compounds that are of no nutritional value but are specific to the aphids' given host. If these feeding triggers are not found, the alatae move on.

In contrast, apterae usually move only when necessary to procure a better feeding site or if a predator or parasite molests them. Ants tend apterae in many aphid groups in a form of facultative mutualism; in some relationships, ants actively "farm" their aphid "cattle" by moving them among locations. Generally, however, aphid groups with elongate siphunculi are less likely to be tended by ants. In exchange for the aphid's sugary honeydew waste, the ants protect them from predators, such as coccinellid, lacewing, and syrphid fly larvae, or specialized aphid parasites, such as chalcidoid and braconid wasps. When stroked by the ant's antennae, the aphid will raise the tip of its abdomen, extruding a honeydew drop, which may be retracted if not accepted by the ant. If an ant does not accept honeydew after a while, the aphid will revert to its normal behavior of flicking the honeydew drop away with its hind leg or cauda, to prevent an accumulation of honeydew from fouling the aphid colony.

Aphids communicate by chemicals and sound. Parasites and predators are often foiled by the use of an aphid alarm pheromone, such as *trans*-β-farnesene. When molested, aphids exude microdroplets of alarm pheromone from their siphuncular pores, and in response adjacent aphids quickly withdraw their stylets from their host and drop to the ground. Shortly thereafter, the fallen aphids visually orient to vertical lines or structures and move toward them in an attempt to climb the plant stem. Aphid sexual pheromones are also used as male attractants by oviparae, being released from specialized pores on their hind tibiae. Sound communication is used by *Toxoptera* spp., which have a stridulatory mechanism consisting of a row of short pegs on the legs, which are rubbed against filelike ridges on the lower epidermis of the abdomen, just below the siphunculi. When disturbed, colonies of *T. aurantii* emit an audible high piercing stridulatory sound, to which their apterae respond.

Fundatrices of gall-forming species use species-specific patterns of feeding or probing behavior to induce characteristically shaped galls on their specialized hosts, in which their progeny can safely develop. The fundatrix of *P. bursarius* climbs the developing leaf petiole on *Populus nigra* about halfway and probes its rostrum around the petiole to create an array of punctures oriented perpendicularly to the petiole shaft. This induces a swollen globular gall with a slit oriented perpendicularly to the petiole shaft. In contrast, on the same host, a *Pemphigus spyrothecae* fundatrix probes the petiole shaft in an upward spiral array of punctures, yielding a corkscrew-shaped petiole gall. Not only do plant galls provide a protective encasement for aphid development, but aphids of even nongalling species do better on galled tissue, probably because of a local increase in plant nutrients in that tissue.

Many aphid species have some lower degree of sociality, especially among apterae, which is expressed as a gregariousness within colonies and probably confers better protection or response to attacks by natural enemies. Alatae of *Drepanosiphum platanoides,* sycamore aphid, are more likely to be distributed in a clustered manner among sycamore leaves, in groups in which the tips of their antennae and legs touch among the aphids. Some aphids have evolved a higher degree of sociality, however. The tribe Cerataphidini of the Hormaphidinae has genera in which species produce a soldier morph with enlarged forelegs, which defend their relative clones differentially. Soldiers discriminate between soldiers and nonsoldiers but do not attack soldiers of their own species. The investment in soldier production by the colony is related to areas needing defense, such as a gall's surface.

AGRICULTURAL IMPORTANCE

While aphids are among the most serious agricultural problem insects, only about 250 species are considered to be agricultural pests. Pest aphids may affect only a very specific host (e.g., *Brachycorynella asparagi,* asparagus aphid), or group of related hosts, such as crucifers (Brassicaceae) (e.g., *Brevicoryne brassicae,* cabbage aphid). Some, however, are quite polyphagous (e.g., *Aphis gossypii,* cotton aphid; *Myzus persicae,* green peach aphid), with an extremely wide host range. Some common polyphagous pest aphids represent sibling species complexes that are morphologically identical but differ in karyotype. They comprise anholocyclic clones, or biotypes, that differ in host preferences, ability to transmit diseases, or resistance to pesticides.

Aphids cause damage in several ways. They can build to high population densities and damage plants directly, by removing enough sap to cause withering and eventual plant death. If not washed off, aphid excrement, or honeydew, can build up enough on plants to serve as a medium for the growth of sooty molds, impairing photosynthesis and plant development, and eventually promoting other fungal diseases. Salivary secretions of some aphids are phytotoxic, causing stunting, leaf deformation, and gall formation. Even if the feeding effects of aphids are not apparent, they may affect plant hormone balances, changing host metabolism to their advantage, thus essentially hijacking the plant's physiological functions.

The aphid vectoring of stylet-borne and circulative plant viruses is the most serious problem to agriculture posed by aphids. Stylet-borne viruses occur on the aphid's epidermis and are not aphid specific. These viruses are acquired quickly and transmitted during the aphid's probing of the plant's epidermis. They are nonpersistent, however, and the aphid's infectiousness is lost upon molting. Circulative viruses, in contrast, live internally in the aphid's gut. The aphid must feed for a while to acquire these viruses, which require an incubation period before they can be successfully transmitted. They are persistent, however, and once infected, the aphid remains a vector throughout its life. The virus–aphid–plant linkage is fairly specific for circulative viruses, and a given virus is transmitted by only one or few aphid species. Virus-infected plants often show an aphid-attractive yellowing and have increased free amino acids, so aphids benefit by virus transmission.

APHID CONTROL IN AGRICULTURAL CROPS AND HOME GARDENS

Agricultural control of aphids best uses an integrated pest management (IPM) strategy, where species are identified and tactics reflect the allowable tolerance level on a crop. Within fields, aphids may be monitored by means of yellow water pans or sticky traps, which attract them. In some agricultural regions, especially seed-growing areas with plant virus sensitivities (e.g., the Netherlands, Idaho), specialized agencies run aerial trapping networks in which large suction traps are used to detect alates and forecast population levels. Proper aphid IPM emphasizes sustainable control, maximizing organically compatible methods to minimize effects on nontarget species, such as biological control agents, or vertebrates. IPM tactics include cultural control methods, such as minimizing weed or ant populations that promote aphids, using ultraviolet-light-reflecting or colored films near plants to repel alates, or interplanting pollen and nectar plants among crop rows to promote aphid natural enemies. Biological control agents include small wasps (e.g., *Aphidius* sp.) that parasitize aphids and disperse well within populations. Predators, which as immatures voraciously consume aphids, can be released. These include lacewings (e.g., *Chrysopa* spp.), aphid midges (e.g., *Aphidoletes* spp.), and ladybird beetle larvae (e.g., *Hippodamia convergens*). Predators may, however, disseminate when released as adults. One can apply entomopathic fungi (e.g., *Beauveria bassiana*), whose spores attach to the aphid's exoskeleton, penetrate it, and kill the aphid. Insect growth regulators applied by spray act through various means to prevent maturation of aphids. These may act in conjunction with biological control agents if the latter fail to provide adequate control. Use of chemical poisons in aphid IPM should be minimized because of the effect on nontarget species. While poison use may sometimes be necessary, heavy usage promotes insecticidal resistance in aphids, as well as secondary resurgence of aphid populations,

once biological control agents have been hampered. Chemical poisons range from less toxic pyrethroids to more toxic organophosphates. They may be applied directly as contact insecticidal sprays or dusts, or indirectly as plant systemic insecticides that are ingested with the plant's sap. Cultivation of aphid-resistant crop varieties is also important.

In home gardens and yards, nontoxic controls should be emphasized. Aphid detection involves inspection of buds, stems, fruits, and the underside of leaves, where the insects are most likely to congregate. Effective control can simply involve frequently hosing aphids off plants with water, being careful to hit the leaf undersides. Spray applications of a mixture of garlic and water may repel aphids. Sprays of cuticle-disrupting insecticidal soaps, which cause fatal desiccation, often give control. Under overhanging trees, problems from aphid sooty molds on driveways, patios, and walkways are best controlled by hosing the surfaces. Control for aphid galls or leaf distortion on deciduous trees can be problematic, sometimes requiring the winter application of a dormant oil to kill overwintering eggs. Ultimately, elimination of the tree to may be required to solve problem, so tree species in yards should be carefully selected and placed, in view of their potential aphid pests.

See Also the Following Articles

Ants • Biological Control • Rostrum • Sternorrhyncha

Further Reading

Blackman, R. L., and Eastop, V. F. (1984). "Aphids on the World's Crops: An Identification Guide." Wiley, New York.

Blackman, R. L., and Eastop, V. F. (1994). "Aphids on the World's Trees: An Identification and Information Guide." CAB International, Wallingford, U.K.

Dixon, A. F. G. (1985). "Aphid Ecology." Blackie & Son, Glasgow, U.K.

Minks, A. K., and Harrewijn, P. (1987). "Aphids, Their Biology, Natural Enemies and Control," Vols. A, B , C. Elsevier, Amsterdam.

Moran, N. (1992). The evolution of aphid life cycles. *Annu. Rev. Entomol.* **37,** 321–348.

Remaudiere, G., and Remaudiere, M. (1997). "Catalogue of the World's Aphididae." INRA Editions, Paris.

Stern, D. L. (1995). Phylogenetic evidence that aphids, rather than plants, determine gall morphology. *Proc. R. Soc. Lond. B.* **260,** 85–89.

Stern, D. L., and Foster, W. A. (1996). The evolution of soldiers in aphids. *Biol. Rev. Cambridge Philos. Soc.* **71,** 27–79.

Apis Species (Honey Bees)

Eva Crane

International Bee Research Association

Honey bees (genus *Apis*) are social insects in the family Apidae, order Hymenoptera; they are among the Aculeata (i.e., those having stingers). They evolved after the separation of the Americas and Australia from Eurasia/Africa and are native only in the Old World. The genus *Apis* probably first appeared in the Eocene, about 55 mya. Tropical species *A. dorsata* and *A. florea* existed by the end of the Oligocene 25 mya, and cavity-nesting *A. mellifera* and *A. cerana,* which can also live outside the tropics, were separate species by the end of the Pliocene about 2 mya. Therefore, the highly advanced cavity-nesting species have existed only perhaps a tenth as long as the open-nesting species, which were confined to the warmer tropics. The most important species to humans is *A. mellifera,* which has been introduced all over the world for use in beekeeping.

THE GENUS *APIS*

Known Species

The genus *Apis* contains 11 known species. *A. mellifera* (Fig. 1) is the source of most of the world's honey. It is native throughout Africa, the Middle East, and Europe except for the far north regions. All other *Apis* species are native to Asia. *A. cerana,* which is kept in hives in the temperate zone as well as the tropics, is smaller than *A. mellifera,* and it makes smaller colonies. Other Asian species that build a multiple-comb nest in a cavity are *A. koschevnikovi* and *A. nuluensis* reported in Borneo, and *A. nigrocincta* in Sulawesi.

Other *Apis* species native in parts of the Asian tropics build a single-comb nest in the open. The most important to humans is *A. dorsata,* a bee much larger than *A. cerana. A. laboriosa,* which is even larger, lives in parts of the Himalayas too high for *A. dorsata.* Much smaller than *A. cerana, A. florea* is widespread below around 500 m and can live in drier areas than *A. dorsata.*

Mating, and How Reproductive Isolation Is Achieved

Honey bees mate in flight; the process has been studied in detail in *A. mellifera,* and involves three stages. A queen flies out when only a few days old, and drones that are flying in the area, attracted by the pheromones she produces, follow her. If a drone succeeds in clasping the queen with his legs,

FIGURE 1 Worker honey bees *(Apis mellifera)* on honeycomb. (Photograph courtesy of P. Kirk Visscher.)

his endophallus is everted and mating occurs. When they separate, part of his genitalia remains in the queen, and he falls away and dies. She may mate more than once (usually on the same flight), and the semen she receives is stored in her spermatheca for use throughout her egg-laying life.

The main component of the pheromone attracting drones to the queen seems to be the same for all *Apis* species (9-oxo-*trans*-decenoic acid). In an area with more than one species, reproductive isolation can be achieved if the drones of different species fly at different times of day.

APIS MELLIFERA

Colony Life

The reproduction of individual bees takes place in the colony, and each colony normally contains a single mated female (the queen), many nonreproductive females (workers) and, during the reproductive season, a smaller number of reproductive males (drones). Colonies reproduce by swarming during a season when much food is available. The workers rear several young queens, each in a special queen cell. The old queen and perhaps half the workers of the colony leave as a swarm, which finds a new nest site. One of the young queens mates and heads the parent colony; the others are killed.

Many aspects of the beekeeping cycle and social behavior of honey bee colonies have been studied in detail (see Further Reading).

In the tropics, temperatures are never too low for plants to flower or for bees to fly, and colony activity is governed by rainfall rather than temperature. There are two seasonal cycles in the year, so colonies do not grow as large, or store as much honey, as they do in temperate zones. If the stores of a colony of *A. mellifera* become low in a dearth period, the colony may leave its hive and fly to a nearby area where plants are starting to bloom, rebuilding its combs in a nest site there. Such movements are referred to as absconding or migration, and preventing them is an important part of beekeeping in tropical Africa.

Subspecies and Their Distribution

During the Ice Ages, geographical features in Europe such as mountains confined *A. mellifera* to several separate areas, where they diversified into a number of subspecies or races. The most important in world beekeeping, and their native areas, are *A. mellifera ligustica* (Italian) in northwestern Italy south of the Alps, *A. mellifera carnica* (Carnolian) in the eastern Alps and parts of the Balkans, *A. mellifera caucasica* (Caucasian) in Georgia and the Caucasus mountains between the Black Sea and the Caspian Sea, and also *A. mellifera mellifera* north of the Alps.

The first introductions of *A. mellifera* from Europe to new continents, after 1600, enabled future beekeeping industries to build up and flourish in many countries. Some of the subsequent introductions of *A. mellifera* carried diseases or parasites not previously present, and these caused much damage. From the late 1800s, after the movable-frame hive was devised, there was great interest in breeding more productive honey bees, and colonies of many races were transported from the Old World to other continents. Italian bees, especially, could store much honey in warm regions with consistently good nectar flows. During the 1900s, scientists introduced exotic species and races of honey bees into Europe for experimental purposes, but none is known to have survived in the wild.

Moving honey bees to new areas in tropical or subtropical environments can have wide-reaching effects. In 1956 a number of honey bee queens were transported from southern Africa to Brazil in an attempt to improve the beekeeping in that South American country, where bees of European origin performed poorly. Through an accident, a few of the African queens escaped with swarms, and this led to hydridization with bees of European origin. The consequent "Africanized" bees had characteristics that enabled them to become dominant over the "European" bees already in the American tropics, and they spread rapidly, reaching the Amazon by 1971, the north coast by 1977, Mexico by 1986, and then several southern U.S. states.

In warm regions, many native plants may be pollinated by small bees (Apoidea) whose populations are reduced if colonies of the larger *A. mellifera* are introduced, which in turn can endanger the reproduction of such native plants. This problem has been reported in Australia and Brazil.

A. mellifera is now used in beekeeping in almost every country in the world.

APIS CERANA AND RELATED SPECIES

Of the subspecies of *A. cerana,* the Asian hive bee, *A. cerana indica* is present from Yunnan in China through India to the Philippines. *A. cerana cerana* is in much of China, also the Himalayas, Afghanistan, and the Russian Far East, and *A. cerana japonica* in Tsushima Island and Japan.

After *A. mellifera* was introduced in eastern Asia, *A. cerana* became restricted to areas with native flora. Then in 1985–1986 colonies of *A. cerana* were taken from one of the Indonesian islands to Irian Jaya, also Indonesian but part of New Guinea. The bees reached Papua New Guinea by 1987 and islands in the Torres Strait by 1993. By 2000 they were found (and killed) in Brisbane, Australia, and strenuous efforts are being taken to prevent any further entry and spread of this bee in Australia because it would probably carry the varroa mite, a pest that is of serious economic importance to beekeepers.

APIS DORSATA AND RELATED SPECIES

The large single-comb nests of the giant honey bee, *A. dorsata,* built in the open, are still the most important source of honey in India and some other tropical Asian countries. *A. dorsata* is present in most of the Indo-Malayan region, from the Indus River in the west to the eastern end of the

Indonesian chain of islands, and from the Himalayas to Java in the south. *A. breviligula* is in the Philippines, and *A. binghami* in the Celebes. In the high Himalayas, *A. laboriosa*—a species even larger than *A. dorsata*—nests up to 3000 m, whereas *A. dorsata* rarely nests above 1250 m.

APIS FLOREA AND RELATED SPECIES

The area of the smaller *A. florea* extends as far northwest as Iran. It has also been reported around the Persian Gulf in Iran and Iraq, and in the Arabian peninsula. It reached parts of this last area, and also Sudan, by the aid of humans, and people may also have facilitated its spread along the coast west of the Indus Valley. It is characteristically found in hot dry areas at altitudes below 500 m; in some localities it is the only honey bee that could survive. Its eastern range does not extend as far as that of *A. dorsata,* possibly because *A. florea* could not cross wide sea channels.

A. andreniformis, rather similar to *A. florea,* occurs in southern China, Myanmar (Burma), Palawan in the Philippines, Thailand, Indonesia, Laos, and Vietnam. It is likely that some early statements about *A. florea* in these areas refer instead to *A. andreniformis.*

See Also the Following Articles

Beekeeping • Caste • Hymenoptera • Neotropical African Bees • Pollination and Pollinators

Further Reading

Crane, E. (1990). "Bees and Beekeeping: Science, Practice and World Resources." Heinemann Newnes, Oxford, U.K.

Crane, E. (1991). "The World History of Beekeeping and Honey Hunting." Duckworth, London.

Graham, J. M. (ed.) (1992). "The Hive and the Honey Bee." rev. ed. Dadant and Sons, Hamilton, IL.

Hepburn, H. R., and Radloff, S. E. (1998). "Honeybees of Africa." Springer-Verlag, Berlin.

Michener, C. D. (2000). "The Bees of the World." John Hopkins University Press, Baltimore, MD.

Otis, G. W. (1991). A review of the diversity of species within *Apis. In* "Diversity in the Genus *Apis*." (D. R. Smith, ed.), Ch. 2. Westview Press, Boulder, CO.

Rinderer, T. E. (ed.) (1986). "Bee Genetics and Breeding." Academic Press, New York.

Ruttner, F. (1988). "Biogeography and Taxonomy of Honeybees." Springer-Verlag, Berlin.

Aposematic Coloration

Mathieu Joron

Leiden University

Insects attract collectors' attention because they are extremely diverse and often bear spectacular colors. To biologists, however, bright coloration has been a constantly renewed puzzle because it makes an insect a highly conspicuous prey to prospective predators. Charles Darwin understood that bright colors or exaggerated morphologies could evolve via sexual selection. However, he felt sexual selection could not account for the conspicuous color pattern of nonreproductive larvae in, for example, *Pseudosphinx* hawk moth caterpillars (Fig. 1). In a reply to Darwin about this puzzle, Alfred R. Wallace proposed that bright colors could advertise the unpalatability of the caterpillars to experienced predators. Indeed, prey that are not edible to predators are predicted to gain by exhibiting conspicuous and very recognizable colors; experienced predators can then correctly identify and subsequently avoid attacking such prey. E. B. Poulton later developed this idea, expanded it to other warning signals (i.e., sounds or smells), and coined the term "aposematism" to describe this phenomenon (from the Greek "away" and "sign").

FIGURE 1 *Pseudosphinx tetrio* hawk moth caterpillar from the Peruvian Amazon, showing a combination of red and black, classical colors used by aposematic insects. These larvae feed on toxic latex-sapped trees in the Apocynaceae. Length 14 cm. (Photograph © M. Joron, 1999.)

Aposematic color patterns are found everywhere throughout the insects, from black- and yellow-striped stinging wasps to black and red, bitter-tasting lady beetles, or brightly colored, poisonous tropical butterflies. Although warning coloration has involved fascination, empirical and theoretical studies for some time, the puzzle of aposematism still motivates much debate today. First, although there is little doubt that bright coloration is often an antipredatory strategy, how aposematism evolves is far from clear. This is because brightly colored mutants in a population of cryptic (camouflaged) prey are more exposed to predators. How can a warning coloration evolve in a prey if the very first mutants exhibiting such coloration in the population are selected against? Second, the reasons for the brightness and conspicuousness of warning colors are not always clear and may be multiple. Are aposematic colors "road signs" that help predators learn better to differentiate inedible from edible prey, or are bright colors more easily memorized and associated to bad taste by predators? Did yellow and red colors, often borne by poisonous insects, evolve because of innate biases against these colors in the predators' brains, or are more complex cognitive, behavioral, frequency-dependent, or coevolutionary mechanisms involved in the evolution of warning patterns?

Finally, why are warning patterns highly diverse in the insect world, whereas all toxic prey would gain by bearing the same color, thus reducing the probability of being sampled by a naïve predator?

WHAT TO ADVERTISE

"Aposematism is quite simply the correlation between conspicuous signals, such as bright coloration, and prey unprofitability," Candy Rowe wrote in 2001. But why should some prey become unprofitable in the first place, while others do not?

Unprofitability is difficult to define, and even more difficult to measure. It is certainly contextually defined, because the propensity of an animal to eat something is highly dependent on its level of hunger and its ability to use the prey for energy once eaten. Palatability (i.e., the predator's perception of prey profitability) greatly determines whether the predator will or will not eat the prey. Predator–prey coevolution led predators in part to rely on proximal perception to gauge the prey profitability. In particular, taste sensitivity may well have evolved in predators as an assessment of food toxicity: indeed, predators usually consider toxic chemicals to be distasteful. Some insects have external defenses such as horns, or spines, many of which cause irritation. Such physical defenses may be coupled to venom, as with the irritant hairs of many caterpillars or hymenopteran stingers. These insects may be otherwise perfectly profitable, and some predators evolve ways around the physical defenses, such as bee-eaters that are able to remove a bee's stinger and venom sac. Other insects have passive chemical defenses that predators discover upon consumption, such as chemicals in the hemolymph or sequestration glands of lubber grasshoppers or monarch butterflies. Such insects usually develop extra signals such as powerful smells, at least when handled, to advertise their toxicity before being consumed.

Toxicity is not the only way an insect can be unprofitable to predators. Difficulty in capturing prey (due to fast escape, erratic flight, breakable wings, etc.), or difficulty in handling prey (due to toughness or a hard cuticle) are other ways that insects can bring no net reward to the predators that spend energy chasing them, even if the chase results in the prey being seized. However, multiple unprofitability traits might be important in the evolution of warning signals.

Predators can have three kinds of response to a prey depending on their perception of prey profitability. If consuming a prey leads the predator to be more likely to attack similar prey in the future, perhaps even using the prey's appearance as a search image, the prey is called palatable. In feeding experiments, this usually leads birds to attack nearly 100% of the palatable prey offered. Of course, the predator may satiate after consuming a number of prey, and consequently the propensity to attack may decrease at high prey densities. In contrast, if experienced predators are less likely to attack similar prey, the prey is called unpalatable. Of course, predators' memorizing capacity, and the strength of the prey unpalatability, may all influence how fast information regarding prey is acquired and how long it is retained. However, a distasteful prey will inevitably lower the predators' instantaneous propensity to attack this prey further, an effect analogous to an immediate satiation. Finally, eating the prey may have no effect on the predator's subsequent behavior, which means that the prey is effectively neutral. This category is mainly derived from theory; there is little evidence that it exists in nature.

Variations in unpalatability among prey species, along what is called the "(un)palatability spectrum," affect the rate at which predators modify their behavior with experience. Predator's perceived toxicity is likely to be a sigmoid function of actual toxin concentration per unit prey mass, meaning that little of the palatability spectrum may fall into intermediate perceptions between "unpalatable" and "fully palatable." Although how predators learn is still under debate, experiments and theory suggest that they respond to a large extent to the (perceived) concentration of nasty chemicals they can tolerate per unit time.

The distastefulness of insects is generally linked to the host plants they utilize. Indeed, many distasteful or defended insects are herbivorous; most defended nonherbivorous insects are Hymenoptera. Some plant families, like the Solanaceae and the Passifloraceae, which are hosts to many chemically defended insects, contain alkaloids and cyanogens, respectively, as secondary metabolites. Some insects, like monarch butterflies (*Danaus plexippus*) that feed on *Asclepias* plants (milkweeds, Asclepiadaceae), sequester the compounds of such plants and store them; these insects thus avoid the toxic effects of the toxic compounds altogether. In soft-bodied insects (e.g., larvae), toxins are usually stored near the teguments or in special glands, ready to release their contents upon handling. The toxicity of insects that extract and sequester plant chemical compounds is dependent on the concentration of these compounds in the host plant. Sawfly larvae (Hymenoptera: Tenthredinidae), for example, reflex-bleed drops of hemolymph when touched; the unpalatability of such larvae is shown to be directly dependent on the glucosinolate concentration of their crucifer host plant over 24 h before "bleeding."

Other insects, however, synthesize their toxins *de novo,* like many chrysomelid beetles; they probably use the same enzymatic machinery that serves (or has served, in their ancestors) to detoxify the plant's secondary compounds. Although many of these species still use precursors derived from their food plant, these insects are usually less dependent on the plant's toxicity to develop their own noxious compounds. Some groups like ithomiine or heliconiine butterflies also get toxin precursors in their adult diet.

Whatever route to distastefulness is taken, we observe a general correlation between clades of distasteful insects and toxicity in host plant families. In butterflies, the distasteful Troidinae (Papilionidae) tend to feed on Aristolochiaceae, monarchs (Nymphalidae: Danainae) usually feed on milkweeds (Asclepiadaceae), longwing butterflies (Nymphalidae: Heliconiinae) feed on Passifloraceae, and clearwings (Nymphalidae: Ithomiinae) mainly on Solanaceae and Apocynaceae. In

contrast, butterfly clades feeding on chemical-free monocotyledonous plants, like browns (Nymphalidae: Satyrinae) on grasses, or owl-butterflies (Nymphalidae: Brassolinae) on palms or Marantaceae, did not evolve distastefulness. Thus toxicity in insects may frequently have evolved as a mere by-product of adaptation to utilize new kinds of food, particularly toxic plants. The costs of detoxification or toxin production could be covered by the benefits of invading competition-free hosts, perhaps assisted by the increased survival afforded by chemical protection.

DISGUSTING, BRIGHT, SIMPLE, AND CONTRASTED: WHY AND HOW TO ADVERTISE

Why should unprofitable prey advertise? Instead of parading with gaudy colors, why should all prey not try to escape predators' detection altogether through camouflage? Although the initial steps to aposematism are not obvious, the advantage of aposematic signals once established is clear. Indeed, numerous studies have shown that most predators are able to learn and recognize, and subsequently avoid, prey they associate with a bad experience.

Some distasteful prey, such as the transparent ithomiine butterflies found in the forest understory in tropical America, are not particularly conspicuous. The rampant mimicry found in this group of inconspicuous butterflies demonstrates that predators are able to learn and avoid such prey (although other stimuli, e.g., motion, might also be important). Still, most distasteful insects are brightly colored. Why should aposematic signals usually be conspicuous, and use simple color patterns of red, yellow, or black? Most of the answer is likely to be found in the cognitive behavior of the predators that selected for such colors. Several hypotheses have been put forward to explain the correlation between bright colors and unprofitability in insects. Bright contrasted colors are thought to be (1) easier to learn, (2) more difficult to forget, and (3) as different as possible from edible prey, thereby facilitating the avoidance of recognition errors. All these mechanisms are supported by experimental data to some extent (e.g., Fig. 2). Because both predators and aposematic prey benefit from correct identification, aposematic colors are believed to take advantage of any bias in the predator's cognition system. Likewise, predators in turn gain in being biased in the same direction as that taken by the prey. Therefore, prey signaling and predators' cognition are likely to have coevolved, which, incidentally, makes experimental evidence for any of the foregoing hypotheses generally hard to establish independently. Hypothesis 3 is the most likely to involve interactions between perception and cognition in the predators, leading to fast evolution of the prey's signals.

Many aposematic insects simultaneously send signals of different kinds, and some argue that such "multimodal" warning signals may reveal unconditioned biases that are absent when each sensory modality is examined alone. Assuming that predators would rely solely on color and not behavior, motion,

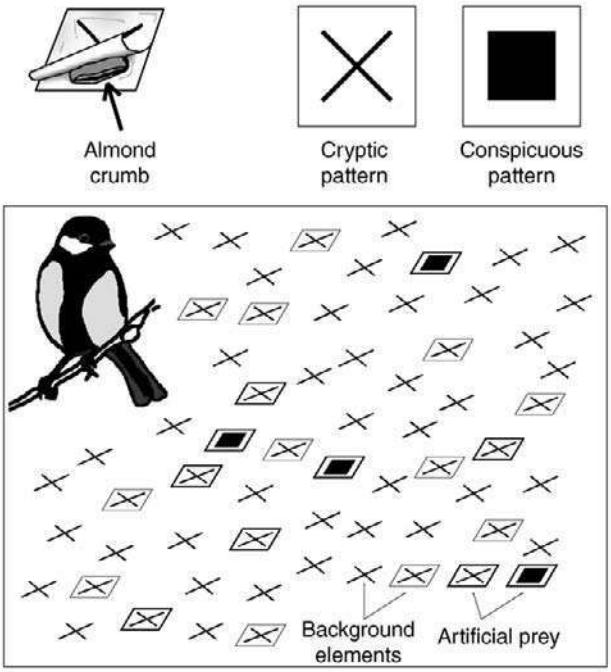

FIGURE 2 Schematic outline of the "novel world" experimental design developed by R. Alatalo and J. Mappes (University of Jyväskylä, Finland) to study the mechanisms of predator's selection on conspicuous prey. Great tits, *Parus major,* are trained to forage in a room covered with small black symbols (e.g., crosses) on a white background, some of which are actual prey. Almond crumbs are placed between two 1-cm^2 pieces of paper glued together that bear a symbol on the outside. Black-squared prey items stand out conspicuously on the black-crossed background and represent potential warning signals, whereas black-crossed prey items are cryptic. The novelty of all symbols ensures that innate or previously learned prey recognition does not interfere with the predator's response during the experiment. This setup also partly resolves one drawback of garden experiments, where the local food abundance for predators is artificially increased, making the searching costs, search images, and other predatory behavior unrealistic. By playing on the palatability of the prey items, it is possible to monitor how the birds learn to avoid the conspicuous signal. Mimics can also be incorporated in the environment at varying frequencies to study the dynamics of Batesian mimicry.

or sounds is perhaps simplistic, and it is sometimes argued that multiple signals could even be a prerequisite for the evolution of warning coloration. In fact, the reason for the apparent importance of multimodality probably lies again in the coevolutionary history of predators and their prey, which shapes innate biases. Predators are generally good entomologists for the potential prey they encounter often, and predators integrate various sensory modalities to make decisions regarding a particular action. Most aposematic insects are mimicked by edible species (Batesian mimics) that parasitize the warning function of the signal. The presence of these Batesian mimics reduces the reliability of the warning signal and means lost prey for the predators. Model species may therefore escape being mimicked by evolving new dimensions for signaling, in addition to the established one (i.e., in different sensory dimensions).

Many warningly colored insects live in groups, which enhances the warning function of their signal for three reasons (Fig. 3). First, predators tend to associate and retain noxiousness

FIGURE 3 Examples of gregarious warningly colored insects. (A) Gregarious *Chromacris* lubber grasshoppers nymphs (Orthoptera: Romaleidae) feeding on a toxic *Solanum mite*. Although it is not a bright color, black is often used as a warning color by insects, presumably because it increases conspicuousness by contrast against green foliage. (B) Gregarious *Morpho* sp. caterpillars spending the day in a dense cluster. They disperse at night for feeding. Clusters of aposematic prey often create an emergent, enhanced pattern presumably perceived as a supernormal stimulus by the predators and therefore better memorized. Caterpillars about 10 cm long. (Photographs from the Peruvian Amazon, © M. Joron, 1997.)

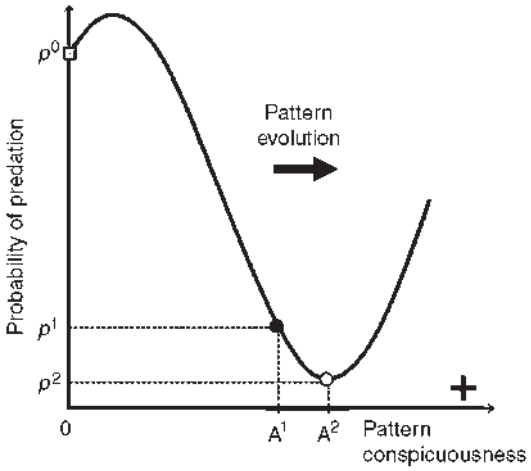

FIGURE 4 Peak shift mechanism applied to prey conspicuousness. The curve describes a fitness function that relates pattern conspicuousness to the probability of suffering predation. Predators are assumed to have knowledge of prey ranging from crypsis (open square: zero conspicuousness, tastefulness, high probability of predation) to aposematism (solid circle: high conspicuousness, A^1; distastefulness, low probability of predation, p^1). From this knowledge, predators extrapolate the palatability of prey with higher conspicuousness they have not yet experienced, hence the curve beyond the solid circle. In particular, slightly more conspicuous prey (open circle) elicit a stronger rejection response than the known aposematic prey (solid circle), and therefore suffer an even lower probability of predation, p^2. Consequently, selection should cause the conspicuousness of the distasteful prey to shift gradually from A^1 to A^2. In contrast, cryptic distasteful prey cannot drift away from crypsis (open square) toward higher levels of conspicuousness because they are more likely to be preyed upon when they become more conspicuous. An initial phenotypic jump is necessary to get to levels of conspicuousness where birds can categorize the prey as warningly colored, avoiding confusion with the normal staple of cryptic tasteful prey.

and a warning signal more quickly when presented with multiple copies of the same signal. Second, all unpalatable prey get an individual advantage in living in groups by the immediate avoidance, similar to the immediate satiation discussed earlier, caused in the individual predator that avoids the group altogether. Third, at the population level, clustering spatially also reduces the number of individual predators the population of prey is exposed to and has to educate, which again enhances the benefit of signaling. In short, it pays to be sitting right next to the toxic individual that is being sampled by a predator, because this is where the probability of predation is lowest, hence the advantage to living in groups. However, many solitary prey also exhibit warning coloration, and gregariousness does not necessarily evolve prior to aposematism.

One common mechanism leading to fast divergence in signals throughout the animals, known as peak shift, hinges here on the coupling of prey coloration, predator experience, and predator innate aversion. Prey can be placed on a conspicuousness axis (Fig. 4), with cryptic edible prey at one end (close to 0 conspicuousness) and incipient aposematic prey or new color pattern mutants at some distance down the axis. More conspicuous prey usually elicit stronger aversion in educated predators, which extrapolate the idea that stronger conspicuousness should mean stronger noxiousness. Therefore, stronger signals (away from edible prey appearance) lead to a supernormal response in the predators that thereby select for increased conspicuousness in the prey. This mechanism is a special case of a runaway process and could be an important route to the evolution of aposematic prey that are bright and

contrasted. It is thought that the coupling of such cognitive biases with the ability to learn leads to the selection of more strongly exaggerated warning colors and patterns in noxious prey than is expected in purely nonlearning predators.

EVOLUTIONARY ROUTES TO APOSEMATIC COLORATION

The Problems

There are obvious benefits to bearing warning colors in a population of warning-colored prey. As noted by early naturalists like A. R. Wallace and later E. B. Poulton, experienced predators avoid warningly colored prey, and presumably the number of prey killed during the predators' education is lower than in the absence of signaling. These benefits are clear at the group level but are not so clear at the individual level, because the first warningly colored individuals in a population of cryptic (and noxious) prey suffer strongly increased predation. Indeed, novel warningly colored prey not only suffer increased detection by prospective predators, but also elicit *no* avoidance in the predators. Consequently, there is strong positive frequency dependence, putting novel rare warning signals at a disproportionate disadvantage against an established strategy (crypsis, or another already established warning signal).

How could warning signals evolve at all if the first mutants using this strategy are killed? Laboratory experiments using the "novel world" design (Fig. 2) show rather unequivocally that aposematic patterns cannot evolve gradually in unpalatable prey. Indeed, small increases in visibility in cryptic prey increased attack rates without enhancing learning. Similarly, deviant phenotypes in established warning patterns suffered stronger predation. Finally, rare conspicuous prey suffered disproportionate predation, even when presented in groups. Therefore, a gradual increase in conspicuousness towards aposematism seems unlikely. This means that the evolving population must undergo a sudden jump, both in phenotype (to get a pattern that predators categorize as a different item) and in numbers beyond a threshold frequency (to allow the local predators to learn about the new pattern). Once the new pattern has achieved the minimum frequency and phenotypic thresholds, positive frequency dependence helps the new mutant to spread in the population. Peak shift or other processes can then occur, increasing the conspicuousness or adding other components to the signal. How can these evolutionary leaps be achieved—or circumvented—by an incipient aposematic prey?

Deterministic Evolution via Immediate Benefits

NEOPHOBIA A new aposematic form could in theory escape the disadvantage of being rare and novel by causing immediate avoidance without having to be tasted at all by the predator. Indeed, predators are somewhat reluctant to sample novel-looking prey, particularly if novelty is associated with bright colors. This phenomenon is called "neophobia," a kind of diet conservatism in predators. Neophobia could arise from various foraging biases, such as the formation of search images in the brains of predators as they search for edible-looking prey and ignore other prey, or via cultural inheritance, as with nestlings that tend to prey upon what they were fed by their parents. Neophobia is sometimes presented as a potential route toward aposematism. However, it does not really resolve the frequency dependence problem, because it is essentially a transient phenomenon involving no information acquisition by predators. Therefore as soon as numbers grow, however slightly, neophobia tends to vanish. Neophobia should best be classified as a predator's bias, like other innate biases against colors, smells, or sound, evolved by predators in response to their prey environment. Such biases are likely to channel the ultimate form taken by the aposematic signal (to the benefit of both preys and predators), but it is unlikely that they cause its evolution in the first place.

INDIVIDUAL ADVANTAGE One obvious way around initial obstacles is not to be killed by predators' attacks. Then, prey could both educate the predators and be avoided in subsequent encounters. Indeed, most birds taste-test their prey before ingesting them, and many aposematic prey have noxious compounds in their outer parts, making it possible to be tasted but not injured by predators. For instance, ithomiine and danaine butterflies concentrate alkaloid in their wings. Day-flying pericopine moths let a voluminous and bitter hemolymph froth out of their body, likely tasted (or smelled) by a predator before it has profoundly injured the moth. Moreover, most unpalatable butterflies have very elastic bodies, which resist crushing. Strong smells that predators take as a warning for bad taste or toxicity, like those of stinkbugs, are another way by which prey can gain immediate advantage without having to be effectively tasted by the predators.

PREY ALREADY CONSPICUOUS Another way by which prey can overcome the difficulty of evolving conspicuous color is not to suffer any cost (i.e., avoid the necessity of a phenotypic leap) as a result of increased conspicuousness. Indeed, most flying insects are rather conspicuous in flight and rely on their difficulty of capture to escape predation. They may not suffer any cost to bearing conspicuous colors, and indeed many butterflies, if not most, irrespective of their palatability, display bright patches of colors on the upper side of their wings, visible in flight, while having cryptic underwings making them inconspicuous when sitting. Such bright dorsal colors might initially evolve as sexual signals in male–male or male–female interactions long before unpalatability evolves. Once noxiousness has evolved, predators can learn an already conspicuous pattern without making recognition errors because of the resemblance to the palatable prey they have as search images. In a way, conspicuous flying insects can be said to be "preadapted" to evolve warning colors. But such patterns can then also change or drift according to predators' biases. In particular, already bright color patterns

can be enhanced toward brighter coloration through processes like peak shift, as described earlier. According to James Mallet, examples of this mechanism are the unpalatable *Taenaris* and *Hyantis* (Nymphalidae: Morphinae), which have evolved strikingly conspicuous warning spots via the enhancement of some of the less conspicuous eyespots that are still found on the undersides of their palatable relatives, the well-known blue *Morpho* butterflies.

MÜLLERIAN MIMICRY The easiest way to avoid the cost of rarity and conspicuousness altogether is to jump to an aposematic pattern already present in the habitat and known by the local predators. The shared appearance between several defended prey is called Müllerian mimicry, and it is likely that most aposematic species evolved via this route. Indeed, mimicry rings usually include a large number of Müllerian species (all of which are noxious). Of these, only one evolved the pattern first, followed by the other species that colonized an already protected pattern. This pattern of evolution is detectable by examining the biogeography and phylogeny of the species in question. For example, *Heliconius erato* and *H. melpomene* are Müllerian mimics throughout their distribution range. However, the *H. melpomene* was shown to have much younger color pattern races, with a clearly distinct genealogy, than *H. erato,* suggesting that *H. melpomene* is a Müllerian mimic that adopted the established color patterns of *H. erato.*

Population Processes: Kin Selection, Drift, and Shifting Balance

Because many unpalatable prey are indeed gregarious, it is easy to conclude that gregariousness allows the evolution of aposematism. The evolution of aposematism through gregariousness relies on the predator rejecting the whole group after sampling only one or few individuals. This extrapolation from one prey to the whole group is analogous to a superfast learning in the predator, which can be enhanced by conspicuousness. However, it also pays for aposematic preys to live in groups, thereby increasing their apparent density to the local predators. It is therefore not clear whether gregariousness or aposematism should evolve first to trigger the evolution of the other. Groups of gregarious larvae (Fig. 3) are usually family groups, suggesting that kinship might allow a new mutation quickly to get to a locally high frequency in such little-dispersing insects through kin selection. However, one should be aware that relatedness per se is not what favors the local rise in frequency of the gene here, but simply the local founding event by one or few family groups.

In fact, many adult aposematic adult insects are either not gregarious at all or do not aggregate in family groups. Besides, some of the most gregarious insect larvae come from the joint oviposition of several unrelated females. Although these examples could have arisen after the initial evolution of warning color through kin selection, it is more parsimonious to infer that non-kin-selection arguments can also explain the evolution of aposematism. Drift alone, particularly, followed by positive frequency dependence, is a good candidate mechanism (and in fact kin founding is only a special case of genetic drift). Indeed, when the ratio of predators to prey decreases in a locality, selection for antipredatory strategies is greatly diminished, allowing the exploration of other color pattern possibilities by the local population. Using release–recapture techniques of different warningly colored forms of *H. cydno* in Ecuador, D. Kapan showed that selection was relaxed when the butterflies were released in larger numbers. Therefore, the prey population could move via genetic drift above the required threshold, after which the new warning color invades the population. Positive frequency dependence has the interesting property that although it hinders the initial evolution of new patterns, it hinders the removal of any pattern once it has been established. If genetic drift in prey populations matches the fluctuations of selection pressures in time and space, new local aposematic patterns can be established frequently in different locations. These are essentially the first and second steps of the shifting balance theory of S. Wright. Competition between geographically adjacent warning color types then allows one pattern to spread to neighboring populations, like the traveling waves of color races documented in South America for *H. erato* or *H. melpomene.*

CONCLUSIONS

Despite the advantages of bearing a warning coloration established in the locality, the evolution of aposematism is not straightforward because proximal mechanisms seem to represent obstacles to its initial evolution. However, aposematic patterns are extremely diverse at all geographical and taxonomic levels, and this major discrepancy between theory and nature clearly suggests that positive frequency-dependent arguments are not as restrictive against the rise of novel warning colors. Similarly, predator generalization, which should not allow gradual shift of cryptic prey toward bright warning colors, does not seem to be efficient in restricting the rise of new conspicuous patterns.

In fact, both population dynamics and psychological arguments might well explain such spectacular diversification. First, positive frequency dependence would allow new local forms to be established through drift, relayed by other processes involving predator's cognitive biases. Second, the initial steps toward warning color are determined largely by which cognitive biases in the predators are exploited. That is, the initial pathway taken toward the evolution of warning coloration probably profoundly affects the aposematic phenotype that eventually evolves. Similarly, positive frequency dependence prevents deviations from the evolutionary pathway that is taken. In short, although aposematism is not expected predictably to evolve via Fisherian selection, it is such a powerful strategy once evolved that it is possibly inevitable in a contingent and varying world, where the nature and the height of the initial obstacles to its evolution fluctuate. It may thus follow a

ratchetlike pattern of evolution, where more routes may lead toward aposematism than routes away from it.

See Also the Following Articles

Chemical Defense • Crypsis • Mimicry • Monarchs

Further Reading

Alatalo, R. V., and Mappes, J. (1996). Tracking the evolution of warning signals. *Nature* **382,** 708–710.

Edmunds, M. (1974). "Defence in Animals. A Survey of Anti-predator Defences." Longman, New York.

Endler, J. A. (1988). Frequency-dependent predation, crypsis, and aposematic coloration. *Philos. Trans. R. Soc. Lond. B* **319,** 505–524.

Guilford, T. (1988). The evolution of conspicuous coloration. *Am. Nat.* **131,** S7–S21.

Lindström, L., Alatalo, R. V., Lyytinen, A., and Mappes, J. (2001). Strong antiapostatic selection against novel rare aposematic prey. *Proc. Nat. Acad. Sci. U.S.A.* **98,** 9181–9184.

Mallet, J., and Joron, M. (1999). Evolution of diversity in warning color and mimicry: Polymorphisms, shifting balance and speciation. *Annu. Rev. Ecolo. System.* **30,** 201–233.

Mallet, J., and Singer, M. C. (1987). Individual selection, kin selection, and the shifting balance in the evolution of warning colors: The evidence from butterflies. *Biol. J. Linn. Soc.* **32,** 337–350.

Poulton, E. B. (1890). "The Colours of Animals." Trübner, London.

Rowe, C. (ed.). (2001). Warning signals and mimicry. Special issue of *Evolutionary Ecology* [1999, vol. 13, no 7/8]. Kluwer, Dordrecht, The Netherlands.

Sillén-Tullberg, B. (1988). Evolution of gregariousness in aposematic butterfly larvae: A phylogenetic analysis. *Evolution* **42,** 293–305.

Sword, G. A., Simpson, S. J., El Hadi, O. T. M., and Wilps, H. (2000). Density dependent aposematism in the desert locust. *Proc. R. Soc. Lond. B Biol. Sci.* **267,** 63–68.

Wallace, A. R. (1879). The protective colours of animals. *In* "Science for All" (R. Brown, ed.), pp. 128–137. Cassell, Petter, Galpin., London.

Apterygota

Apterygota is a subclass of the class Insecta in the phylum Arthropoda. It contains two orders, the Archaeognatha and the Thysanura.

Aquatic Habitats

Richard W. Merritt
Michigan State University

J. Bruce Wallace
University of Georgia

Less than 3% of the world's total water occurs on land, and most of this is frozen in polar ice caps. Streams and rivers are one of the more conspicuous features of the landscape; however, their total area is about 0.1% of the land surface, whereas lakes represent about 1.8% of total land surface. Some authors have questioned whether insects have been successful in water because aquatic species represent only a small portion of the total hexapod fauna. However, 13 orders of insects contain species with aquatic or semiaquatic stages, and in five of these (Ephemeroptera, Odonata, Plecoptera, Megaloptera, and Trichoptera) all species are aquatic with few exceptions (Table I). Few aquatic insects spend all of their life in water; generally any insect that lives in water during a portion of its development is considered to be "aquatic." Usually, but not always, for most "aquatic" species, it is the larval stage that develops in aquatic habitats, and the adults are terrestrial (Table I). The pupae of some taxa undergoing complete metamorphosis (i.e., holometabolous) remain within the aquatic habitat; in others the last larval instar moves onto land to pupate, providing the transition stage from the aquatic larva to the terrestrial adult.

The success of insects in freshwater environments is demonstrated by their diversity and abundance, broad distribution, and their ability to exploit most types of aquatic habitat. Some species have adapted to very restricted habitats and often have life cycles, morphological, and physiological adaptations that allow them to cope with the challenges presented by aquatic habitats. One aquatic environment in which insects have not been very successful is saltwater habitats, although some 14 orders and 1400 species of insects occur in brackish and marine habitats; only one group occurs in the open ocean. One of the most widely accepted attempts to explain why more insects do not live in marine environments is that successful resident marine invertebrates evolved long before aquatic insects and occupy many of the same niches inhabited by freshwater insects. Thus, marine invertebrates, such as crustaceans, have barred many insects from

TABLE I Occurrence of Life Stages in Major Habitat Types for Aquatic and Semiaquatic Representatives of Insect Orders (A, adult; L, larvae; P, pupae)

Order	Terrestrial	Freshwater
Collembola	A, L	A, L
Ephemeroptera	A	L
Odonata	A	L
Heteroptera	A	A, L
Orthoptera	A, L	A, L
Plecoptera	A	L
Coleoptera	A, L, P	A, L
Diptera	A, P	L, P
Hymenoptera	A	A, L, P
Lepidoptera	A	L, P
Megaloptera	A, P	L
Neuroptera	A, P	L
Trichoptera	A	L, P

Modified from Ward, J. V. (1992). "Aquatic Insect Ecology," Vol. 1, "Biology and Habitat." Wiley, New York.

marine habitats by competitive exclusion. Problems with osmoregulation have been given as another reason for the paucity of saltwater species; however, one of the two multicellular animals found in the Great Salt Lake is a member of the order Diptera (see later: Unusual Habitats), providing evidence that some insects display a strong ability to osmoregulate.

The first aquatic insects are believed to have inhabited flowing water as early as the Permian and Triassic. It was not until the late Triassic and early Jurassic that evidence of abundant lentic, or still-water, fauna arose, accompanied by rapid diversification of water beetles, aquatic bugs (Heteroptera), and primitive Diptera. On the basis of several lines of evidence including osmoregulation, fossil evidence, secondary invasions to water of many taxa, and great variation in gill structure among and within orders, some authors have suggested that the first insects may have lived in water rather than in terrestrial habitats. However, the general consensus is that an aquatic origin for insects seems unlikely and that aquatic insects may not have shown up until 60 to 70 million years later than their terrestrial counterparts.

Freshwater systems are often divided into standing (lentic) and flowing (lotic) waters. Although such a division is useful for indicating physical and biological differences, habitat diversity can vary tremendously within these two broad categories, and some of the same taxa may be found in both lentic and lotic habitats, depending on the physiological constraints of a given habit. Many factors influence successful colonization of aquatic insects to a given habitat; however, most of these would fall under four broad categories: (1) physiological constraints (e.g., oxygen demands, respiration, osmoregulation, temperature effects), (2) trophic considerations (e.g., food acquisition), (3) physical constraints (e.g., coping with harsh habitats), and (4) biotic interactions (e.g., predation, competition). However, these categories are so interrelated that detailed analysis of each factor separately is very difficult.

HABITAT, HABIT, AND TROPHIC CLASSIFICATION SYSTEM

The classification system used here for lotic and lentic habitats stresses the basic distinction between flowing water (i.e., streams, rivers) and standing water (i.e., ponds, lakes, swamps, marshes) habitats (Table II). This separation is generally useful in describing the specific microhabitats (e.g., sediments, vascular hydrophytes, detritus) in which aquatic insects may be found. Both stream/river currents and lake shoreline waves often create erosional (riffle-type) habitats and may resemble each other in their physical characteristics, whereas river floodplain pools and stream/river backwaters create depositional (pool-type) habitats that may resemble lake habitats as well (Table II). Within a given habitat, the modes by which individuals maintain their location (e.g., clingers on surfaces in fast-flowing water, sprawlers on sand or on surfaces of floating leaves, climbers on stem-type surfaces, burrowers in soft sediments) or move about (e.g., swimmers, divers, surface skaters) have been categorized (Table III). The distribution pattern resulting from habitat selection by a given aquatic insect species reflects the optimal overlap between habit and physical environmental conditions that comprise the habitat, such as bottom type, flow, and turbulence. Because food in aquatic habitats is almost always distributed in a patchy fashion, the match between habitat and habit is maximized in certain locations. This combination will often result in the maximum occurrence of a particular species.

In view of the complex physical environment of streams, it is not surprising that benthic invertebrates have evolved a diverse array of morphological adaptations and behavioral mechanisms for exploiting foods. Throughout this article we will follow the functional classification system originally described by K. W. Cummins in 1973, which is based on the mechanisms used by invertebrates to acquire foods (Table IV). These functional groups are as follows:

- **Shredders,** which are insects and other animals that feed directly on large pieces of organic matter (e.g., decomposing leaves and fragments of wood >1 mm in size) and their associated fungi and bacteria, and convert them into fine particulate organic matter (FPOM) through maceration, defecation, and physical degradation;
- **Collector-filterers,** which have specialized anatomical structures (e.g., setae, mouth brushes, fans, etc.) or silk and silklike secretions that act as sieves to remove fine particulate matter less than 1 mm in diameter from the water column;
- **Collector-gatherers,** which gather food, primarily FPOM, that is deposited within streams or lakes;
- **Scrapers,** which have mouthparts adapted to graze or scrape materials (e.g., periphyton, or attached algae, and the associated microbes) from rock surfaces and organic substrates;
- **Predators,** which feed primarily on other animals by either engulfing their prey or piercing prey and sucking body contents.

These functional feeding groups refer primarily to modes of feeding or the means by which the food is acquired, and the food type per se (Table IV). For example, shredders may select leaves that have been colonized by fungi and bacteria; however, they also ingest attached algal cells, protozoans, and various other components of the fauna along with the leaves.

LOTIC HABITATS

Streams vary greatly in gradient, current velocity, width, depth, flow, sinuosity, cross-sectional area, and substrate type, depending on their position in the landscape with respect to geology, climate, and the basin area they drain. Anyone who has spent much time around or wading in streams is aware that these can be extremely diverse habitats, often manifesting great

TABLE II Aquatic Habitat Classification System

General category	Specific category	Description
Lotic–erosional (running-water riffles)	Sediments	Coarse sediments (cobbles, pebbles, gravel) typical of stream riffles.
	Vascular hydrophytes	Vascular plants growing on (e.g., moss, *Fontipalis*) or among (e.g., pondweed, *Potamogeton pectinatus*) coarse sediments in riffles.
	Detritus	Leaf packs (accumulations of leaf litter and other coarse particulate detritus at leading edge or behind obstructions such as logs or large cobbles and boulders) and debris (e.g., logs, branches) in riffles.
Lotic–depositional (running-water pools and margins)	Sediments	Fine sediments (sand and silt) typical of stream pools and margins.
	Vascular hydrophytes	Vascular plants growing in fine sediments (e.g., *Elodea,* broad-leaved species of *Potamegeton, Ranunculus*).
	Detritus	Leaf litter and other particulate detritus in pools and alcoves (backwaters).
Lentic–limnetic (standing water)	Open water	On the surface or in the water column of lakes, bogs, ponds.
Lentic–littoral (standing water, shallow-water area)	Erosional	Wave-swept shore area of coarse (cobbles, pebbles, gravel) sediments.
	Vascular hydrophytes	Rooted or floating (e.g., duckweed, *Lemna*) aquatic vascular plants (usually with associated macroscopic filamentous algae).
	Emergent zone	Plants of the immediate shore area (e.g., *Typha,* cattail), with most of the leaves above water.
	Floating zone	Rooted plants with large floating leaves (e.g., *Nymphaea,* pond lily), and nonrooted plants (e.g., *Lemna*).
	Submerged zone	Rooted plants with most leaves beneath the surface.
	Sediments	Fine sediments (sand and silt) of the vascular plant beds.
	Sediments	Fine sediments (fine sand, silt, and clay) mixed with organic matter of the deeper basins of lakes. (This is the only category of "lentic–profundal.")
Lentic–profundal (standing water, basin)	Freshwater lakes	Moist sand beach areas of large lakes.
Beach zone	Marine intertidal	Rocks, sand, and mud flats of the intertidal zone.

After Merritt, R. W., and Cummins, K. W. (1996). "An Introduction to the Aquatic Insects of North America." Kendall/Hunt, Dubuque, IA.

TABLE III Categorization of Aquatic Insect Habits: That Is, Mode of Existence

Category	Description
Skaters	Adapted for "skating" on the surface where they feed as scavengers on organisms trapped in the surface film (e.g., Heteroptera: Gerridae, water striders).
Planktonic	Inhabiting the open-water limnetic zone of standing waters (lentic; lakes, bogs, ponds). Representatives may float and swim about in the open water but usually exhibit a diurnal vertical migration pattern (e.g., Diptera: Chaoboridae, phantom midges) or float at the surface to obtain oxygen and food, diving when alarmed (e.g., Diptera: Culicidae, mosquitoes).
Divers	Adapted for swimming by "rowing" with the hind legs in lentic habitats and lotic pools. Representatives come to the surface to obtain oxygen, dive and swim when feeding or alarmed; may cling to or crawl on submerged objects such as vascular plants (e.g., Heteroptera: Corixidae, water boatman; Coleoptera: adult Dytiscidae, predaceous diving beetles).
Swimmers	Adapted for "fishlike" swimming in lotic or lentic habitats. Individuals usually cling to submerged objects, such as rocks (lotic riffles) or vascular plants (lentic) between short bursts of swimming (e.g., Ephemeroptera: Siphlonuridae, Leptophlebiidae).
Clingers	Representatives have behavioral (e.g., fixed retreat construction) and morphological (e.g., long, curved tarsal claws, dorsoventral flattening, ventral gills arranged as a sucker) adaptations for attachment to surfaces in stream riffles and wave-swept rocky littoral zones of lakes (e.g., Ephemeroptera: Heptageniidae; Trichoptera: Hydropsychidae; Diptera: Blephariceridae).
Sprawlers	Inhabiting the surface of floating leaves of vascular hydrophytes or fine sediments, usually with modifications for staying on top of the substrate and maintaining the respiratory surfaces free of silt (e.g., Ephemeroptera: Caenidae; Odonata: Libellulidae).
Climbers	Adapted for living on vascular hydrophytes or detrital debris (e.g., overhanging branches, roots and vegetation along streams, submerged brush in lakes) with modifications for moving vertically on stem-type surfaces (e.g., Odonata: Aeshnidae).
Burrowers	Inhabiting the fine sediments of streams (pools) and lakes. Some construct discrete burrows that may have sand grain tubes extending above the surface of the substrate or the individuals may ingest their way through the sediments (e.g., Ephemeroptera: Ephemeridae, burrowing mayflies; Diptera: most Chironominae, Chironomini, bloodworm midges). Some burrow (tunnel) into plants stems, leaves, or roots (miners).

After Merritt, R. W., and Cummins, K. W. (1996). "An Introduction to the Aquatic Insects of North America." Kendall/Hunt, Dubuque, IA.

TABLE IV General Classification Systems for Aquatic Insect Trophic Relations

Functional group[a]	Subdivision of function group: Dominant food	Subdivision of function group: Feeding mechanism	Examples of taxa	General particle size range of food (μm)
Shredders	Living vascular hydrophyte plant tissue	Herbivores—chewers and miners of live macrophytes	Trichoptera: Phyrganeidae, Leptoceridae	$> 10^3$
	Decomposing vascular plant tissue and wood—coarse particular organic matter (CPOM)		Trichoptera: Limnephilidae	
			Plecoptera: Pteronarcyidae, Nemouridae	
			Diptera: Tipulidae, Chironomidae	
Collectors		Detritivores—filterers or suspension feeders	Trichoptera: Hydropsychidae	
	Decomposing fine particular organic matter (FPOM)		Diptera: Simuliidae	$< 10^3$
		Detritivores—gatherers or deposit (sediment) feeders (includes surface film feeders)	Ephemeroptera: Ephemeridae, Baetidae, Ephemerellidae	
			Diptera: Chironomidae	
Scrapers	Periphyton—attached algae and associated material	Herbivores—grazing scrapers or mineral and organic surfaces	Trichoptera: Glossomatidae	
			Coleoptera: Psephenidae	$< 10^3$
			Ephemeroptera: Heptageniidae	
Predators (engulfers)	Living animal tissue	Carnivores—attack prey, pierce tissues and cells, suck fluids	Heteroptera: Belostomatidae, Odonata,	$> 10^3$
	Living animal tissue	Carnivores—ingest whole animals (or parts)	Plecoptera: Perlidae, Perlodidae	$> 10^3$
			Coleoptera: Dytiscidae, Megaloptera	
			Trichoptera: Rhyacophilidae	

After Merritt, R. W., and Cummins, K. W. (1996). "An Introduction to the Aquatic Insects of North America." Kendall/Hunt, Dubuque, IA.
[a]General category based on feeding mechanism.

differences over short distances. In the upper reaches of a catchment or drainage basin, small streams often display a range of habitats characterized by areas that are shallow, with fast flow over pebbles, cobbles, and boulders. There are also areas with steep gradients, cascades, or waterfalls when the underlying substrate is bedrock. There also may be areas of slow velocity in pools of deeper water.

In many streams draining forested watersheds, pools are found. Pools are depositional areas during normal flow as organic and inorganic particles settle to the substrate, and a similar settling process often occurs in side channels or backwater areas of streams. Pools are also created upstream of large instream pieces of wood, which may form obstructions known as debris dams. Because pools are generally characterized by reduced water velocity, many of the small particles normally suspended in fast flows settle to the bottom. In many low-gradient streams, including large rivers, bottom substrate often consists of silt, sand, and gravel-sized particles that are frequently moved by the force of the flowing water. In such systems, large pieces of woody debris entering the river from bank erosion or from adjacent floodplain or upstream areas may represent an important habitat for invertebrate colonization.

Substratum characteristics are often perceived as a major contributor to the distribution of many invertebrates; however, many other factors, including water velocity, food,

feeding habits, refuge, and respiratory requirements, can be associated with specific substrates. Substratum particle size is influenced by several items, including geology, physical characteristics of the rock, past and present geomorphic processes (flowing water, glaciation, slope, etc.), climate and precipitation, and length of time over which the processes occur. These in turn influence landform, which exerts a major influence on various hydrological characteristics of aquatic habitats. Unlike many lentic environments, in lotic systems the velocity of moving water is sufficient to pass the water around the body of an insect and turbulence provides reaeration; thus, dissolved oxygen is rarely limiting to stream inhabitants. Local transport and storage of inorganic and organic materials by the current may be either detrimental (e.g., scouring action) or beneficial (as a food source). For example, most aquatic insects in flowing waters are passive filter feeders and depend on the water current for delivery of their food. Scouring flows may therefore remove from the streambed large organic particles (e.g., leaves) as well as smaller ones, creating temporary reductions in food supplies. In contrast, moderately rapid flows may facilitate feeding of some scraper or grazer insects by preventing excessive sedimentation buildup on the surfaces on which they feed.

Some Insects and Their Adaptations to Erosional Habitats

Adaptations of aquatic insects to torrential or "rapid flow" habitats include the dorsoventral flattening of the body, which serves two purposes: it increases the organism's area of contact with the surface substratum, and it offers a mechanism by which animals can remain in the boundary layer when water velocity diminishes, thereby reducing drag under subsequent exposure to high velocities. However, this second idea may be an oversimplification. Indeed, some authors have suggested that the dorsoventrally flattened shape may actually generate lift in the insect. Examples of animals inhabiting stones in torrential habitats include a number of mayflies (Ephemeroptera) belonging to the families Heptageniidae (Fig. 1A) and Ephemerellidae; some Plecoptera, such as Perlidae (Fig. 1G); some Megaloptera (i.e., Corydalidae) (Fig. 1D); and caddisflies (Trichoptera), such as Leptoceridae (*Ceraclea).*

In addition to body shape, many mayflies and stoneflies have legs that project laterally from the body, thereby reducing drag and simultaneously increasing friction with the substrate. Most of these taxa are either scrapers or gatherers on surfaces of stones or predators on other aquatic insects. Undoubtedly, the diverse physical forces encountered in aquatic environments, especially streams, influence the array of morphologies found among aquatic insects.

In some caddisflies (e.g., Glossosomatidae), the shape of the case rather than the insect is modified. The larvae of Glossosomatidae in their tortoiselike cases are frequently seen grazing on the upper surfaces of stones in riffle areas. Another

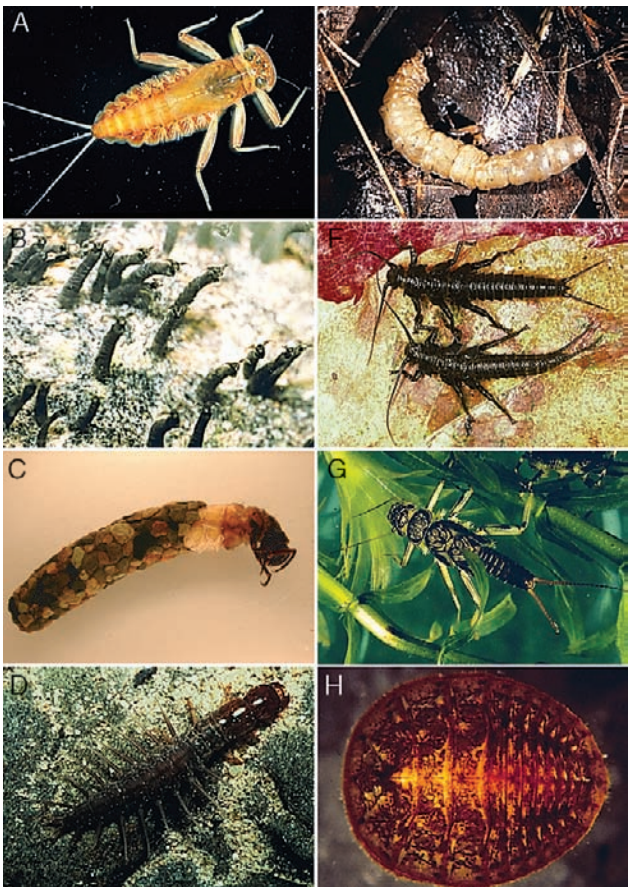

FIGURE 1 Typical insects inhabiting lotic environments. (A) Ephemeroptera: Heptageniidae (*Rhithrogena*). (Photograph by H. V. Daly.) (B) Diptera: Simuliidae (*Simulium*), (C) Trichoptera: Limnephilidae (*Dicosmoecus*), (D) Megaloptera: Corydalidae (*Corydalus*), (E) Diptera: Tipulidae (*Tipula*), (F) Plecoptera: Pteronarcyidae (*Pteronarcys*), (G) Plecoptera: Perlidae, (H) Coleoptera: Psephenidae (*Psephenus).*

curious caddisfly grazer on stone surfaces is *Helicopsyche,* whose larvae construct coiled cases of sand grains shaped like snail shells. Both glossosomatids and helicopsychids reach their greatest abundances in sunny cobble riffles, where they feed on attached periphyton or algae. Another lotic insect that relies on a rather streamlined case is the limnephilid caddisfly *Dicosmoecus* (Fig. 1C).

Larvae of the dipteran family Blephariceridae are unusual in that they possess hydraulic suckers. A V-shaped notch at the anterior edge of each of the six ventral suckers works as a valve out of which water is forced when the sucker is pressed to the substrate. The sucker operates as a piston with the aid of specialized muscles. In addition, a series of small hooks and glands that secrete a sticky substance aid sucker attachment as the larvae move in a zigzag fashion, releasing the anterior three suckers, lifting the front portion of the body to a new position, and then reattaching the anterior suckers before releasing and moving the posterior ones to a new position. These larvae are commonly found on smooth stones in very rapid velocities and are usually absent from stones covered

with moss and from roughened stones that interfere with normal sucker function. Several other aquatic insects have structures that simulate the action of suckers. The enlarged gills of some mayflies (e.g., *Epeorus* sp. and *Rhithrogena* sp.: Fig. 1A) function as a friction pad, and *Drunella doddsi* has a specialized abdominal structure for the same purpose. Some chironomids have "pushing prolegs" represented by circlet of small spines that function as a false sucker when pressed to the substrate. Mountain midge larvae (Deuterophlebiidae) possibly use a similar mechanism to attach their suckerlike prolegs. Most of these animals are primarily grazers on thin films of epilithon (algae, associated fine organic matter, and microbes) found on the surface of stones.

Flowing water usually carries many organic (and inorganic) particles and a number of insects exploit these suspended particles. Filter-feeding collectors (Table IV) exploit the current for gathering food with minimal energy expenditure. For example, certain filtering collectors exploit locations where flows converge over and around substrates, thus allowing the animals to occupy sites of greater food delivery. Examples include caddisfly larvae belonging to the families Hydropsychidae and Brachycentridae. Silk is used for attachment by a number of caddisflies (e.g., Hydropsychidae, Philopotamidae, and Psychomyiidae), which build fixed nets and retreats (Fig. 2A). Although the Philopotamidae are found in riffle habitats, their fine-meshed, tubelike nets are usually found in crevices or undersides of stones in low velocity microhabitats (Fig. 2C). The nets of the caddisfly, *Neureclipsis,* are limited to moderately slow (< 25 cm s^{-1}) velocities and the large (up to 20 cm long), trumpet-shaped nets (Fig. 2D) are used to capture small animals drifting downstream. *Neureclipsis* larvae are often very abundant in some lake outflow streams where drifting zooplankton are abundant.

Some case-making caddisflies (e.g., *Brachycentrus* sp.) also use silk for attaching their cases to the substrate in regions of moderately rapid flow. Many chironomid larvae construct fixed silken retreats for attachment or silken tubes that house the larvae, with a conical catchnet spun across the lumen of the tube. Periodically, the larva devours its catchnet with adhering debris that has been swept into the burrow by the water currents. Meanwhile, other chironomid larvae such as *Rheotanytarsus* spp. construct small silk cases that are attached to the stream substratum with extended hydralike arms. The arms project up in the current and are smeared with a silklike secretion to capture particles.

Larval blackflies (Simuliidae, Fig. 1B) use a combination of hooks and silk for attachment. The thoracic proleg resembles that of chironomids and deuterophlebiids, described earlier, and the last abdominal segment bears a circlet of hooks, which it uses to anchor itself to substrates. The larva moves forward, inchwormlike, spins silk over the substrate, and attaches the proleg and then the posterior circlet of hooks to the silken web. Most blackfly larvae possess well-developed cephalic fans, which are used to filter small particles from suspension. These attached larvae twist

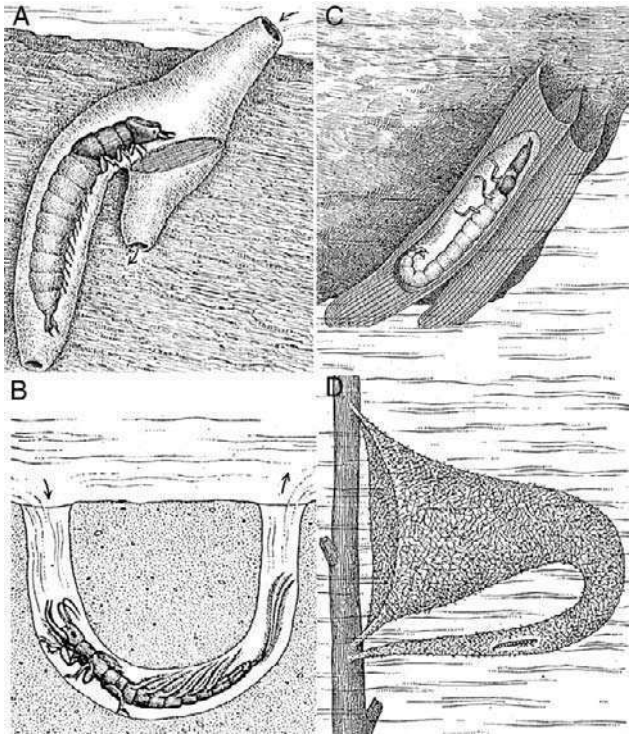

FIGURE 2 Representative lotic insects in their environment: (A) Caddisfly larva (*Macrostenum*) in its retreat grazing on materials trapped on its capture net, (B) mayfly larva of *Hexagenia* (Ephemeridae) in its U-shaped burrow, (C) tubelike nets of philopotamid caddisfly larvae (Philotamidae) on the lower surface of a stone, (D) the caddisfly larva and cornucopia-shaped net of *Neureclipsis* (Polycentropodidae). [Habitat drawings modified and taken from Wallace, J. B., and Merritt, R. W. (1980). Filter-feeding ecology of aquatic insects. *Annu. Rev. Entomol.* **25,** 103–132, (B); Merritt, R. W., and Wallace, J. B. (1981). Filter-feeding insects. *Sci. Am.* **244,** 131–144 (A, C, D).]

their bodies longitudinally from 90° to 180° with the ventral surface of the head and fans facing into the current. The fusiform body shape of blackfly larvae reduces turbulence and drag around their bodies, which are often located in regions of relatively rapid flow. Blackfly pupae are housed in silken cases that are attached to the substrate.

Although unidirectional current is the basic feature of streams, most lotic insects have not adapted to strong currents, but instead have developed behavior patterns to avoid current. Very few lotic insects are strong swimmers, probably because of the energy expenditure required to swim against a current. Downstream transport or drift requires only a movement off the substrate to enter the current. Streamlined forms, such as the mayflies *Baetis* spp., *Centroptilum, Isonychia* spp., and *Ameletus* spp., are capable of short rapid bursts of swimming, but most lotic insects move by crawling or passive displacement. One characteristic of these latter mayflies is the possession of a fusiform, or streamlined, body shape: examples include several Ephemeroptera such as *Baetis, Centroptilum,* and *Isonychia,* as well as a number of beetle (Coleoptera) larvae. A fusiform body shape reduces resistance in fluids,

and within the mayflies the shape is often associated with excellent swimming abilities.

The benthic fauna in streams often can be found in cracks and crevices, between or under rocks and gravel, within the boundary layer on surfaces, or in other slack-water regions. Another method of avoiding fast currents is living in debris accumulations consisting of leaf packs and small woody debris. This debris offers both a food resource and a refuge for insects and contains a diverse array of aquatic insects including stoneflies such as Peltoperlidae and Pteronarcyidae (Fig. 1F), caddisflies such as Lepidostomatidae and some Limnephilidae, as well as dipterans such as chironomids and tipulid crane flies (Fig. 1E).

In some streams with unstable sandy or silt substrates, woody debris can represent a "hot spot" of invertebrate activity. Wood debris provides a significant portion of the stable habitat for insects in streams when the power of the flowing water is insufficient to transport the wood out of the channel. In addition to the insect component using wood primarily as a substrate, there is often a characteristic xylophilous fauna associated with particular stages of wood degradation. These include chironomid midges and scraping mayflies (*Cinygma* spp. and *Ironodes* spp.) as early colonizers, and larvae and adults of elmid beetles. In western North America, an elmid (*Lara avara*) and a caddisfly (*Heteroplectron*) are gougers of firm waterlogged wood, chironomids are tunnelers, and the tipulids, *Lipsothrix* spp., are found in wood in the latest stages of decomposition. Woody debris is most abundant in small, forested watersheds, but it is also an important habitat in larger streams with unstable beds. In the southeastern coastal plain of the United States and in low gradient mid- and southwestern streams and rivers with unstable bottom substrate, woody debris or "snags" often represent the major habitat for aquatic insect abundance and biomass. High populations and biomass of filter-feeding animals such as net-spinning caddisflies (*Hydropsyche* spp., *Cheumatopsyche* spp., and *Macrostenum*) (Fig. 2A), and blackflies occur in these streams and rivers. In addition to filter feeders, other groups such as odonates, mayflies, stoneflies, elmid beetles, nonfiltering caddisflies, and dipteran larvae can be locally abundant on large pieces of woody debris. Invertebrate shredders and scrapers promote decomposition of outer wood surfaces by scraping, gouging, and tunneling through wood. In fact, wood gouging habits of some net-spinning caddisflies have been blamed for the failure of submerged timber pilings that had been supporting a bridge!

Sand and silt substrates of rivers and streams are generally considered to be poor habitats because the shifting streambed affords unsuitable attachment sites and poor food conditions. An extreme example of this instability is the Amazon River, where strong currents move the bedload downstream, resulting in dunes of coarse sand up to 8 m in height and 180 m in length, thus largely preventing the establishment of a riverbed fauna. However, sandy substrates do not always result in poor habitat for all aquatic insects: some sandy streams are quite productive. Blackwater (i.e., high tannic acid concentrations from leaf decomposition) streams of the southeastern United States have extensive areas of sand, with some of insects, such as small Chironomidae (< 3 mm in length), exceeding $18{,}000/m^{-2}$ in abundance. Their food is derived from fine organic matter, microbes, and algae trapped in the sandy substrate. Numerically, the inhabitants of sandy or silty areas are mostly sprawlers or burrowers, with morphological adaptations to maintain position and to keep respiratory surfaces in contact with oxygenated water. At least one insect, the mayfly *Ametropus,* is adapted for filter feeding in sand and silt substrates of large rivers. *Ametropus* uses the head, mouthparts, and forelegs to create a shallow pit in the substrate, which initiates a unique vortex (flow field in which fluid particles move in concentric paths) in front of the head and results in resuspension of fine organic matter as well as occasional sand grains. Some of these resuspended fine particles are then trapped by fine setae on the mouthparts and forelegs. Many predaceous gomphid (Odonata) larvae actually burrow into the sediments by using the flattened, wedge-shaped head and fossorial (adapted for digging) legs. The predaceous mayflies *Pseudiron* spp. and *Analetris* spp. have long, posterior-projecting legs and claws that aid in anchoring the larvae as they face upstream. Some mayflies (e.g., Caenidae and Baetiscidae) have various structures for covering and protecting gills, and others (e.g., Ephemeridae, Behningiidae) have legs and mouthparts adapted for digging. The predaceous mayfly *Dolania* spp. burrow rapidly in sandy substrates and have dense setae located at the anterior–lateral corners of the body as well as several other locations. The larva uses its hairy body and legs to form a cavity underneath the body where the ventral abdominal gills are in contact with oxygenated water.

Dense setae also are found in burrowing mayflies belonging to the family Ephemeridae that are common inhabitants of sand and silt substrates. They construct shallow U-shaped burrows and use their dorsal gills to generate water currents through the burrow (Fig. 2B), while using their hairy mouthparts and legs to filter particles from the moving water. Hairy bodies seem to be a characteristic of many animals dwelling on silt substrates, which include other collector mayflies such as *Caenis, Anepeorus,* and some Ephemerellidae.

Many dragonflies (e.g., *Cordulegaster* spp., *Hagenius* spp., and Macromiidae) have flattened bodies and long legs for sprawling on sandy and silty substrates. Some caddisflies, such as *Molanna,* have elongate slender bodies but have adapted to sand and silt substrates by constructing a flanged, flat case. They are camouflaged by dull color patterns and hairy integuments that accumulate a coating of silt. The eyes, which cap the anterolateral corners of the head, are elevated over the surrounding debris. The genus *Aphylla* (Gomphidae) is somewhat unusual in that the last abdominal segment is upturned and elongate, allowing the larvae to respire through rectal gills while buried fairly deep in mucky substrate. Some insects burrow within the upper few centimeters of the substratum in depositional areas of streams. This practice is

found among some dragonflies and a number of caddisflies, including *Molanna,* and various genera of the families Sericostomatidae and Odontoceridae.

Specialized Flowing Water Habitats

The hyporheic region is the area below the bed of a stream where interstitial water moves by percolation. In gravelly substrates or glacial outwash areas, it may also extend laterally from the banks. In some situations an extensive fauna occurs down to one meter in such substrates. Most orders are represented, especially taxa with slender flexible bodies or small organisms with hard protective exoskeletons. Some stoneflies in the Flathead River of Montana spend most of their larval period in this extensive subterranean region of flow adjacent to the river. Stonefly larvae have been collected in wells over 4 m deep, located many meters from the river. Rivers draining glaciated regions where there are large boulders and cobble appear to have an exceptionally well-developed hyphoreic fauna.

Other specialized flowing water habitats include the madicolous (or hygropetric) habitats, which are areas in which thin sheets of water flow over rock. These often approach vertical conditions (e.g., in waterfalls) and have a characteristic fauna. Among common animals in these habitats are caddisflies, including several microcaddisflies (Hydroptilidae), Lepidostomatidae, beetles such as Psephenidae, and a number of Diptera larvae belonging to the Chironomidae, Ceratopogoniidae, Thaumaleidae, Tipulidae, Psychodidae, and some Stratiomyiidae.

Thermal (hot) springs often have a characteristic fauna, which is fueled by algae and bacteria adapted to high temperatures. The common inhabitants include a number of dipteran families such as Chironomidae, Stratiomyiidae, Dolochopodidae, and Ephydridae, as well as some coleopterans. A number of these survive within rather narrow zones between the thermal spring and cooler downstream areas.

LENTIC HABITATS

Lentic or standing-water habitats range from temporary pools to large deep lakes and include marshes and swamps, as well as natural (i.e., tree holes, pitcher plants) and artificial (i.e., old tires, rain barrels) containers. The available habitats and communities for insects in a pond or lake were defined in Table II. These habitats include the littoral zone, which comprises the shallow areas along the shore with light penetration to the bottom and normally contains macrophytes (rooted vascular plants). The limnetic zone is the open-water area devoid of rooted plants, whereas the deeper profundal zone is the area below which light penetration is inadequate for plant growth, water movement is minimal, and temperature may vary only slightly between summer and winter. The aquatic and semiaquatic insect communities inhabiting these zones are known as the pleuston (organisms associated with the surface film), plankton and nekton (organisms that reside in the open water), and benthos (organisms associated with the bottom, or solid–water interface). Nektonic forms are distinguished from plankton by their directional mobility, and the latter are poorly represented in lentic waters by insects; the majority of insects found in standing-water habitats belong to the benthos. Their composition and relative abundance is dependent on a variety of factors, some of which are integrated along depth profiles. The overall taxonomic richness of benthic insect communities generally declines with increasing depth.

Among the aquatic communities of lentic habitats, the following orders of aquatic and semiaquatic insects are commonly found within the littoral, limnetic, and profundal zones: the springtails (Collembola), mayflies (Ephemeroptera), true bugs (Heteroptera), caddisflies (Trichoptera), dragonflies (Anisoptera) and damselflies (Zygoptera), true flies (flies, gnats, mosquitoes, and midges) (Diptera), moths (Lepidoptera), alderflies (Megaloptera), and beetles (Coleoptera). Not all these groups occur in lakes, and many are associated with ponds or marshes; examples of typical lentic insects are shown in Figs. 3 and 4.

The Pleuston Community

The unique properties of the water surface or air–water interface constitute the environment of the pleuston community. The Collembola, or springtails, are small in size, have a springing organ (furcula), and a water-repelling cuticle that enables them to be supported by and move across water surfaces. Among the true bugs, the Gerridae (water striders) and related families, the Veliidae (broad-shouldered water striders) and Hydrometridae (water measurers), are able to skate across the water. Adaptations for this habit include retractable preapical claws to assist in swimming, elongate legs

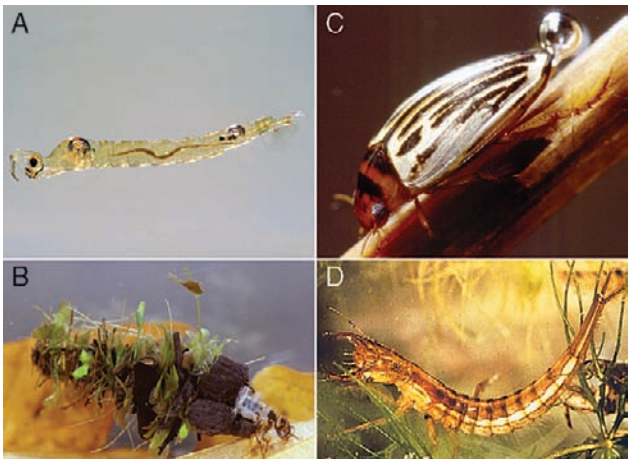

FIGURE 3 Typical insects inhabiting lentic environments. (A) Diptera: Chaoboridae (*Chaoborus*), (B) Trichoptera: Limnephilidae (*Limnephilus*), (C) Coleoptera: Dytiscidae (*Agabus*), (D) Coleoptera: Dytiscidae. (Photographs in A, B, and C by M. Higgins.)

FIGURE 4 Typical insects inhabiting lentic environments (A) Coleoptera: Hydrophilidae (*Hydrochara).* (Photograph by M. Higgins.) (B) Diptera: Chironomidae (*Chironomus*), (C) Odonata: Libellulidae (*Pantala).* (Photograph by M. Higgins.)

and body to distribute the insect's weight over a large area of the surface film, and hydrofuge (nonwettable) hairpiles for support on the surface. Some gerrids also are capable of detecting surface vibrations caused by potential prey. Adult whirligig beetles (Gyrinidae) live half in and half out of water with each eye divided into upper and lower halves, permitting vision simultaneously in both the air and the water; glands keep the upper portion of the body greased to repel water. The middle and hind legs of adult gyrinids are paddle shaped, enabling them to be one of the most effective swimming invertebrates. Among the Diptera, only the mosquitoes (Culicidae) may be considered to be permanent members of the pleuston of lentic waters. The larvae and pupae of most species use the underside of the surface film for support. Larval *Anopheles* lie horizontally immediately beneath the air–water interface, supported by tufts of float hairs on each. Larvae of most other genera (*Aedes, Culex, Culiseta*) hang upside down, with an elongated terminal respiratory siphon penetrating the surface film. Feeding adaptations associated with pleuston specialization include predation by the Hemiptera and Coleoptera, to practically all functional feeding modes by different mosquito larvae, including collecting-filtering and gathering, scraping, and shredding (Table IV).

The Nekton and Plankton Communities

The nekton are swimmers able to navigate at will (e.g., Coleopera, Hemiptera, some Ephemeroptera), whereas plankton are floating organisms whose horizontal movements are largely dependent on water currents. The phantom midge *Chaoborus* sp. (Chaoboridae) (Fig. 3A) is normally regarded as the only planktonic insect and is abundant in many eutrophic (nutrient-rich) ponds and lakes. The tracheal system in these larvae is reduced to kidney-shaped air sacs that function solely as hydrostatic organs, and the larvae slowly descend or rise by adjusting the volume of the air sacs. *Chaoborus* remains in benthic regions during the day but moves vertically into the water column at night. These journeys are dependent on light and oxygen concentrations of the water. The larvae avoid predation by being almost transparent, and they have prehensile antennae that are used as accessory mouthparts to impale zooplankton (Fig. 3A). The only other group of insects that may be considered to be planktonic are the early chironmid instars, which have been reported in the open water column.

Among the Heteroptera, nektonic species are in the Notonectidae (back swimmers), Corixidae (water boatman), and Belostomatidae (giant water bugs), all of which are strong swimmers. Many of these rise to the water surface unless continously swimming or clinging to underwater plants. Notonectids have backs formed like the bottom of a boat and navigate upside down. They hang head downward from the surface or dive swiftly, using their long hind legs as oars. On the underside of the body, they carry a silvery film of air, which can be renewed at regular intervals, for breathing while submerged. Two genera of backswimmers (*Anisops* and *Buenoa*) use hemoglobin for buoyancy control, and this adaptation has enabled these insects to exploit the limnetic zone of fishless lentic waters, where they prey on small arthropods. They have been considered for use as biological control agents for mosquito larvae in some areas of North America. In contrast to notonectids, corixids always swim with the back up, using their elongate, flattened oarlike legs. Although some water boatmen are predators, they are the only group of semiaquatic Heteroptera that have members that are collectors, feeding on detritus and associated small plant material. The Belostomatidae are strong swimmers, but probably spend most of their time clinging to vegetation while awaiting prey, rather than actively pursuing their food in the open water. They are masters of their environment and capture and feed on a variety of insects, tadpoles, fish, and even small birds. The eggs of many belostomatids are glued to the backs of the males by the females and carried in this position until nymphs emerge, a remarkable adaptation for protection of the eggs.

Although most aquatic beetles (Coleoptera) are associated with the substrate, members of the Dytiscidae (predaceous diving beetles) and the Hydrophilidae (water scavenger beetles) are often found swimming in the water column and together constitute the majority of all species of water beetles. The dytiscids are mainly predators in both the adult and larval stage (Fig. 3C, D), while adult hydrophilids are omnivorous, consuming both living and dead materials. The larvae of hydrophilids are predaceous. (Fig. 4A). To respire, hydrophilid adults, having their largest spiracles on the thorax, break the surface film with their antennae; dytiscids, having their largest spiracles on the abdomen, come up tail-end first, as do the larvae of both families. Overall, there are actually few truly nektonic insects, and most of them pass through the limnetic zone when surfacing for emergence. This may be, partly, because with no resting supports in the limnetic zone, maintaining position requires continuous swimming or neutral buoyancy. The vast majority of lentic insects occur in shallow water with emergent plants and are considered to be part of the benthos.

The Benthos Community

Benthos, derived from the Greek word for bottom, refers to the fauna associated with the solid–water interface and includes insects residing on the bottom or associated with plant surfaces, logs, rocks, and other solid substrates. In lentic habitats, many insects fall into this category as mentioned earlier, particularly the Chironomidae, which often represent over 90% of the fauna in the profundal (deep-water) zone of lakes and ponds. These inhabitants are mostly burrowers that feed on suspended or sedimented organic materials and are capable of tolerating low dissolved oxygen or even anaerobic conditions. Chironomid larvae build U- or J-shaped tubes with both openings at the mud–water interface. Body undulations cause a current of water, providing conditions under which oxygen and particulate food can be drawn through the tube. Some midge larvae found in sediments (mainly *Chironomus* sp.) are bright red and are known as bloodworms (Fig. 4B). The red color is caused by the respiratory pigment hemoglobin, which enables a larva to recover rapidly from anaerobic periods because the pigment takes up oxygen and passes it to the tissues more quickly than is possible by diffusion alone.

Other members of the benthos of deeper waters include the mayfly, *Hexagenia* (Ephemeridae), which inhabits the silt and mud of nearshore lake bottoms and has legs modified for digging to construct U-shaped burrows (Fig. 2B). Mayfly numbers have been increasing because of improved water quality standards for lakes and streams. Exceptions to the main constituents of the profundal zone are some immature mayflies, stoneflies, and caddisflies that have been collected at depths from 30 to 100 m in Lake Superior, Michigan. Also, a stonefly, *Utacapnia lacustra* (Capniidae), occurs at depths of 80 m in Lake Tahoe, Calfornia–Nevada, and completes its entire life cycle at this depth, never needing to surface.

Several orders of aquatic insects reach their greatest abundance and diversity in the shallow littoral zone of ponds and lakes as benthos typically associated with macrophytes (macroalgae and rooted vascular plants). The occupants are burrowers, climbers, sprawlers, clingers, swimmers, and divers (Table III) and include the Ephemeroptera, Heteroptera, Odonata, Trichoptera, Megaloptera, Lepidoptera, Coleoptera, and Diptera. The same groups occupy marshes and some swamps, which generally tend to be shallow, with an associated plant zone across the entire surface. Mayflies belonging to the families Baetidae and Siphlonuridae are generally swimmers, clingers, and climbers in vegetated ponds and marshes and mainly feed by means of collecting-filterering or -gatherering (Table IV). Heteroptera include the water scorpions (Nepidae), which have long slender respiratory filaments and are well concealed by detritus and tangled plant growth because of their sticklike appearance. These sit-and-wait predators capture organisms that frequent their place of concealment. Other families of Heteroptera adapted for moving through vegetation in ponds are the Pleidae or pygmy back-swimmers and creeping water bugs, the Naucoridae.

The Odonata, particularly the Gomphidae, are all predators and usually conceal themselves by either burrowing in substrate, sprawling among fine sediment and detritus, or climbing on vascular plants. Sprawlers are more active hunters and include the Libellulidae (Fig. 4C) and Corduliidae. Numerous setae give them a hairy appearance to help camouflage the larvae, and color is protective in patterns of mottled greens and browns. Most Zygoptera (damselflies) and the dragonfly (Anisoptera) family Aeshnidae are mainly climbers or clingers, lurking in vegetation or resting on stems of aquatic plants. The larvae stalk their prey, and both dragonfly and damselfly larvae have a unique lower lip (the labium) armed with hooks, spines, teeth, and raptorial setae that can extend to seize prey and then bring it back into the mouth, holding the food while it is being eaten. The food of larval odonates consists of other aquatic insects such as midges, semiaquatic bugs, and beetles, as well as small fish. Predators of larval odonates include aquatic birds, fish, and large predaceous insects.

In the order Megaloptera, which includes the hellgrammites or dobsonfly larvae of streams, only the predaceous larvae of the alderfly (*Sialis*) is common in ponds and lakes. They are generally found in sand or mud along the margins, but occasionally in deeper water, and they prey on insect larvae and other small animals. The only aquatic family in the related order Neuroptera is the Sisyridae (the spongilla flies), and these are found feeding on freshwater sponges that occur in some streams and the littoral zones of lakes and ponds. The larvae, which occur on the surface or in the cavities of the host, pierce the sponge cells and suck the fluids with their elongated mouthparts.

Although most caddisflies are observed living in lotic waters, several families of caddisflies are either associated with temporary ponds in the spring, aquatic vegetation in permanent ponds, lakes and marshes, or wave-swept shore lines of

lakes. The Hydropsychidae (net spinners), Helicopsychidae (snail case makers), Molannidae, and Leptoceridae are often found along wave-swept shorelines of lakes, and their feeding habits range from those of scrapers and collector-filterers to predators. The Phryganeidae and several genera within the Limnephilidae are climbers, clingers, and/or sprawlers among vegetation in temporary and permanent ponds and marshes; generally, they are shredders of vascular hydrophytes and other decaying plants. The cases of lentic caddisfly families vary with the environment they are found in. Some cases consist of narrow strips of leaves put together in spiral form around a cylinder (Phryganaeidae: *Phryganea* sp.), others consist of plant materials such as leaves and bark arranged transversely to produce a bulky cyclindrical case (Limnephilidae: *Limnephilus*) (Fig. 3B).

Both aquatic and semiaquatic moths (Lepidoptera) occur in lentic habitats, and several genera form close associations with vascular hydrophytes. Larvae of the family Pyralidae (*Parapoynx* sp.) spend the first two instars on the bottom and feed on submerged leaves of water lilies, whereas older larvae generally become surface feeders. Silk spun by the caterpillars is often used to build protective retreats, and pupation usually takes place in silken cocoons or silk-lined retreats. Larval habits of aquatic and semiaquatic moths include leaf mining, stem or root boring, foliage feeding, and feeding on flower or seed structures. One semiaquatic lepidopteran called the yellow water lily borer (the noctuid *Bellura gortynoides*), mines the leaves as a young caterpillar and then bores into the petioles of lilies as an older caterpillar. Within the petiole, larvae are submerged in water and must periodically back out to expose the posterior spiracles to the air before submerging again. The larvae swim to shore by undulating their bodies and overwinter under leaf litter in protected areas.

In addition to the water scavenger and predaceous diving beetles that may occur as nekton swimming through the water column, larvae and adults of other beetles are considered to be part of the benthos of ponds and marshes. These include the Haliplidae (crawling water beetles), which are clingers and climbers in vegetation, and the Staphylinidae (rove beetles), which are generally found along shorelines and beaches, as well as in the marine intertidal zone. The Scirtidae (marsh beetles) are generally found associated with vascular hydrophytes but also are a prominent inhabitant of tree holes. The aquatic Chrysomelidae (leaf beetles) occur commonly on emergent vegetation in ponds, especially floating water lily leaves. The larvae of one genus, *Donacia,* obtain air from their host plant by inserting the sharp terminal modified spiracles into the plant tissue at the base of the plant. Water lilies can be heavily consumed by larvae and adults of the chrysomelid beetle, *Galerucella* sp., and some of the aquatic herbivorous beetles belonging to the family Curculionidae (weevils) include pests of economic importance such as the rice water weevil (the curculionid *Lissorhoptrus*).

The Diptera is clearly one of the most diverse aquatic insect orders, inhabiting nearly all lentic habitats and representing all functional feeding groups and modes of existence. Although the benthic Chironomidae may reach their highest densities in the profundal zone of eutrophic lakes and ponds, they also are largely represented in the littoral zone associated with submergent and emergent plants, where they often graze on the algae attached to leaf surfaces or are vascular plant miners. Other dipteran families that occur in the littoral or limnetic zone, along with their specific habitat, habit (mode of locomotion, attachment, or concealment), and functional feeding mode are summarized in Table V. Among these, a few are of particular interest because of their high diversity and/or abundance in these habitats, namely the crane flies (Tipulidae), the shore and brine flies (Ephydridae), and the marsh flies (Sciomyzidae). The Tipulidae, the largest family of Diptera, are found along the margins of ponds and lakes, freshwater and brackish marshes, and standing waters in tree holes. A few littoral species inhabit the marine intertidal zone. To these are added the large numbers of species that are semiaquatic, spending their larval life in saturated plant debris, mud, or sand near the water's edge or in wet to saturated mosses and submerged, decayed wood. Ephydridae larvae have aquatic and semiaquatic members and occupy several different lentic habitats ranging from salt water or alkaline pools, springs, and lakes to burrowers and miners of a variety of aquatic plants in the littoral margins of these freshwater lentic habitats. All larvae utilize a variety of food, but algae and diatoms are of particular importance in their diet. The Sciomyzidae share some of the same habitat with the shore and brine flies, particularly fresh- and saltwater marshes, and along margins of ponds and lakes among vegetation and debris. The unique aspect of their larval life is that they are predators on snails, snail eggs, slugs, and fingernail clams. The aquatic predators float below the surface film and maintain buoyancy by frequently surfacing and swallowing an air bubble. Prey may be killed immediately or over a few days.

MARINE HABITATS

As noted earlier, insects have been largely unsuccessful in colonizing the open ocean, except for some members of the heteropteran family Gerridae. Most marine insects live in the intertidal zone (i.e., between high and low tide marks), especially on rocky shores or associated with decaying seaweed on sandy beaches (Table I). Although several orders have representatives in the intertidal zone, only a few orders, notably the Diptera, Coleoptera, and Collembola, have colonized these habitats in any numbers. The harsh physical environment of this area has forced these groups to occur buried in sand or mud and to hide in rock crevices or under seaweed.

UNUSUAL HABITATS

Because of adaptive radiation over evolutionary time, insects have colonized virtually every aquatic habitat on earth. Therefore, it is not surprising that these organisms are found

TABLE V Summary of Ecological Data for Benthic Aquatic and Semiaquatic Diptera Larvae Inhabiting Lentic Habitats

Family	Habitat	Habitat	Functional feeding mode
Ceratopogonidae (biting midges, "no-see-ums")	Littoral zone (including tree holes and small temporary ponds and pools)	Generally sprawlers, burrowers or planktonic (swimmers)	Generally predators some collector-gatherers
Chironomidae (nonbiting midges)	All lentic habitats including marine, springs, tree holes	Generally burrowers, sprawlers (most are tube builders); some climber-clingers	Generally collector-gatherers, collector-filterers; some shredders and scrapers
Corethrellidae	Limnetic and littoral margins	Sprawlers	Predators
Psychodidae (moth flies)	Littoral detritus (including tree holes)	Burrowers	Collector-gatherers
Ptycopteridae (phantom crane flies)	Vascular hydrophytes (emergent zone), bogs	Burrowers	Collector-gatherers
Tipulidae (crane flies)	Littoral margins, floodplains (organic sediment)	Burrowers and sprawlers	Generally shredders, collector-gatherers
Dolichopodidae	Littoral margins, estuaries, beach zones	Sprawlers, burrowers	Predators
Stratiomyidae (soldier flies)	Littoral vascular hydrophytes; beaches (saline pools, margins)	Sprawlers	Collector-gatherers
Tabanidae (horseflies, deerflies)	Littoral (margins, sediments and detritus); beaches, marine and estuary	Sprawlers, burrowers	Predators
Canacidae (beach flies)	Beaches—marine intertidal	Burrowers	Scrapers
Ephydridae (shore and brine flies)	Littoral (margins and vascular hydrophytes)	Burrowers, sprawlers	Collector-gatherers, shredders, herbivores (miners), scrapers, predators
Muscidae	Littoral	Sprawlers	Predators
Scathophagidae (dung flies)	Vascular hydrophytes (emergent zone)	Burrower-miners (in plant stems), sprawlers	Shredders
Sciomyzidae (marsh flies)	Littoral—vascular hydrophytes (emergent zone)	Burrowers, inside snails	Predators or parasites
Syrphidae (flower flies)	Littoral (sediments and detritus), tree holes	Burrowers	Collector-gatherers

in the most unusual of aquatic habitats. The title of most versatile aquatic insect must be shared among members of the dipteran family Ephydridae, or shore flies. Shore flies can breed in pools of crude petroleum and waste oil, where the larva feed on insects that become trapped on the surface film. Other species of this family (*Ephydra cinera*), known as brine flies, occur in the Great Salt Lake, Utah, which has a salinity six times greater than that of seawater. Larva maintain water and salt balance by drinking the saline medium and excreting rectal fluid that is more than 20% salt. Another related family of flies, the Syrphidae, or "rat-tailed maggots," occur in sewage treatment lagoons and on moist substrates of trickling filter treatment facilities. Both families have larvae with breathing tubes on the terminal end, which permits the larvae to maintain contact with the air while in their environment. Some Stratiomyiidae, or soldier flies, live in the thermal hot springs of Yellowstone National Park with temperatures as high as 47°C! Other members of this family inhabit the semiaquatic medium of cow dung and dead corpses. A few species of insects have invaded caves and associated subterranean habitats, as mentioned earlier (see Lotic Habitats).

Another unsual aquatic habitat that several insect orders occupy is referred to phytotelmata or natural container habitats and include tree holes, pitcher plants, bromeliads, inflorescences, and bamboo stems. Synthetic container habitats, such as old tires, cemetery urns, rain gutters, and similar natural habitats such as hoofprints also harbor similar insects. Some of these habitats are extremely small and hold water only temporarily, but nevertheless can be quite diverse. The most common order found in these habitats is the Diptera with more than 20 families reported. Over 400 species of mosquitoes in 15 genera alone inhabit these bodies of water and some of these species are important vectors of disease agents.

Insect communities inhabiting pitcher plants (*Sarracenia purpurea*) in North America are exemplified by a sarcophagid or flesh fly (*Blaesoxipha fletcheri*), a mosquito (*Wyeomyia smithii*), and a midge (*Metriocnemus knabi*). The relative abundance of these pitcher plant inhabitants is related to the age, inasmuch as each of the three species consumes insect remains that are in different stages of decomposition. Specifically, the larvae of the flesh fly feed on freshly caught

prey floating on the pitcher fluid surface. The mosquito larvae filter feed on the decomposed material in the water column, and the midge larvae feed on the remains that collect on the bottom of the pitcher chamber. Temporary habitats are important because they are populated by a variety of species, often with unique morphological, behavioral, and physiological properties.

See Also the Following Articles

Cave Insects • Marine Insects • Mosquitoes • Soil Habitats • Swimming

Further Reading

Allan, J. D. (1995). "Stream Ecology." Kluwer, Dordrecht, the Netherlands.

Brönmark, C., and Hansson, L.-A. (1998). "The Biology of Lakes and Ponds." Oxford University Press, Oxford, U.K.

Cummins, K. W. (1973). Trophic relations of aquatic insects. *Annu. Rev. Entomol.* **18,** 183–206.

Cushing, C. E., and Allan, J. D. (2001). "Streams: Their Ecology and Life." Academic Press, San Diego, CA.

Hynes, H. B. N. (1970). "The Ecology of Running Waters." University of Toronto Press, Toronto, Ont., Canada.

McCafferty, W. P. (1981). "Aquatic Entomology." Jones & Bartlett, Boston.

Merritt, R. W., and Cummins, K. W. (eds.). (1996). "An Introduction to the Aquatic Insects of North America." Kendall/Hunt, Dubuque, IA.

Merritt, R. W., and Wallace, J. B. (1981). Filter-feeding in aquatic insects. *Sci. Am.* **244,** 131–144.

Resh, V. H., and Rosenberg, D. M. (1984). "The Ecology of Aquatic Insects." Praeger Scientific, New York.

Ward, J. V. (1996). "Aquatic Insect Ecology," Vol. 1, "Biology and Habitat." Wiley, New York.

Williams, D. D. (2002). "The Ecology of Temporary Waters." Blackburn Press, Caldwell, NJ.

Williams, D. D., and Feltmate, B. W. (1992). "Aquatic Insects." CAB International, Wallingford, Oxon, U.K.

Arachnida

see *Scorpions; Spiders*

Archaeognatha
(Bristletails)

Helmut Sturm
University Hildesheim, Germany

The Archaeognatha (Microcoryphia; part of the subdivided order Thysanura) are apterygote insects with a body size between 6 and 25 mm and a cylindrical shape (Fig. 1). The eyes are large and contiguous, and there are two lateral and one median ocelli (small single eyes with a single beadlike lens). The flagellate (whiplike) antennae extend one-half to three times the length of the body. The mouthparts are ectognathous (freely visible) and the mandibles are linked with the head by a monocondylic joint (i.e., one point of attachment). Some authors believe that this feature distinguishes the Archaeognatha from all other ectognathous Insecta. The seven-segmented maxillary palps are longer than the legs. The thoracic tergites II + III are in lateral view strongly arched, and the two or three tarsal segments of the legs are rigidly united. Some taxa have additional scopulae (dense brushes of specialized hairs) on the distal end of the third tarsal segment. There are mostly pairs or double pairs of eversible vesicles on the coxites of the abdomen (Fig. 3). On each of the abdominal coxites II to IX, styli (pointed, nonarticulated processes) are present (Fig. 3).

FIGURE 1 A male archaeognathan (*Machilis germanica*), body length ca. 12 mm, lateral view; for details see Fig. 2.

Females have two long gonapophyses on each of the abdominal segments VIII + IX, forming the ovipositor. The penis of the males on abdominal segment IX varies in length, and in Machilidae it is fitted with paired parameres on abdominal segments IX or VIII + IX. The three filiform and scaled caudal appendages (one long filum terminale and two laterally inserted cerci) are directed backward. Tergites, cerci, and coxites are always scaled. The molts continue in adult stages. Many species are petrophilous (living on and under stones). The order comprises about 500 species in two families (Machilidae and Meinertellidae).

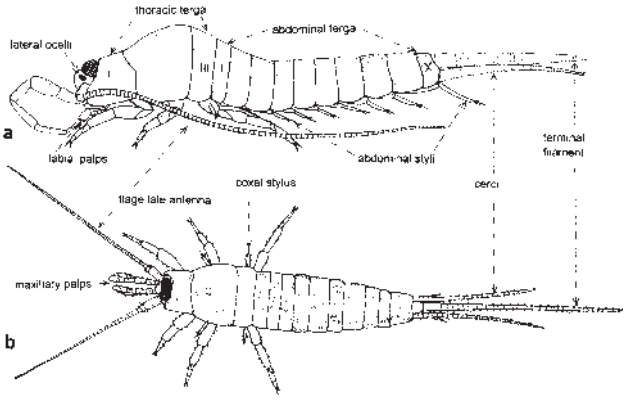

FIGURE 2 General structure of Archaeognatha, semidiagrammatic. (a) Lateral view. (b) Dorsal view, color pattern of dorsal scales intimated. (Reprinted from *Deutsche Entomologische Zeitschrift* **48,** p. 4, © 2001 by Wiley-VCH, with permission.)

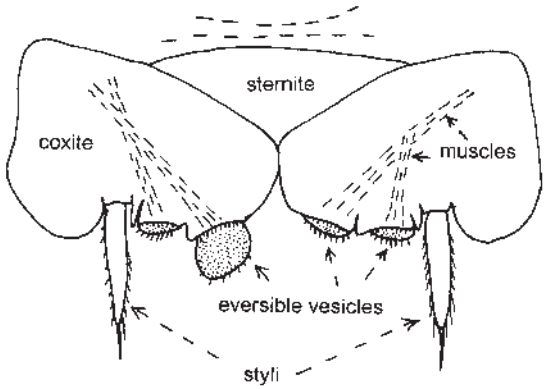

FIGURE 3 Abdominal coxites III of *M. germanica,* ventral view. The eversible vesicles can be exserted by increasing the inner pressure and retracted by muscles.

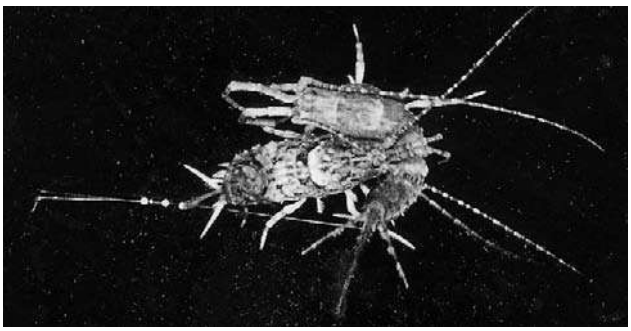

FIGURE 4 Mating position of *M. germanica,* dorsal view. The male has drawn out a secreted thread, deposited three sperm droplets on the thread, and taken up a U form. The ovipositor of the female is touching one of the sperm droplets.

FOSSIL RECORD, SYSTEMATICS, AND BIOGEOGRAPHY

The fossil record of Paleozoic and Mesozoic apterygotes is poor, and many of the fossils of ectognathous representatives cannot be clearly assigned to extant orders. From the Mesozoic, the only archaeognathan fossil is *Cretaceomachilis libanensis* from the lower Cretaceous of Lebanon. For the Cenozoic period, there are many fossils of Archaeognatha, most being amber inclusions. For example, from Baltic amber (ca. 35 mya) seven species of Machilidae are known. All extant forms and the fossils from the Cretaceous and the Tertiary can be included in the superfamily Machiloidea. This group includes two families, the more primitive Machilidae (46 genera and some 325 species), with three subfamilies (Machilinae, Petrobiinae, Petrobiellinae), and the more derived Meinertellidae (19 genera and some 170 species).

The Machiloidea are distributed worldwide. Only the Meinertellidae occur in South America, the Caribbean, South Africa, Australia, and Melanesia. Both Machilidae and Meinertellidae occur in the United States.

BEHAVIOR AND ECOLOGY

The mating behavior of archaeognathans is unique. There are three different modes of sperm transfer. In the most widespread and unique mating behavior, a carrier thread is used. In *Machilis germanica,* for example, the male approaches the female and drums on her with his long maxillary palps. The female shows "willingness" to mate by moving toward the male and bending up the tip of her abdomen. The male then attaches a secreted thread to the ground with his parameres. While the thread is being drawn out, the male secretes three to five sperm droplets onto the thread (Fig. 4). The male curves simultaneously around the female, preventing her from moving forward. Finally, the male places the sperm droplets onto the ovipositor of the female. The indirect transmission of sperm droplets, which are deposited on a thread stretched between parameres and the ground, is unique within the animal kingdom. At least two other possibilities of indirect sperm transfer are known. In *Petrobius* (Machilidae), the sperm are moved directly from the penis onto the ovipositor of the female and in all Meinertellidae sperm are transferred by stalked spermatophores deposited on the ground.

Archaeognatha are found in habitats with very different climates. Representatives of the genus *Allopsontus* (Machilidae) are found up to 5000 m in the Himalayan region. In contrast, two meinertellid species live in the Amazonian forest. Some genera (e.g., *Petrobius*) are found near the seacoast. In tropical forests, meinertellid genera are found on the leaves of bushes and trees. Most Archaeognatha feed on green algae, lichens, and dead leaves. Several species of spiders probably are their principal predators.

Their protection against enemies is probably provided by three main factors: (1) the presence of long appendages with sensilla (filum terminale, cerci, antennae); (2) a dense scale cover on the relatively thin and flexible tergites and coxites; and (3) their ability to jump, which is fully developed in all free-living stages and in all recent representatives, and probably is their most effective defense.

PHYSIOLOGY

The cuticle of the Archaeognatha is generally thin and flexible and bears a multitude of sensory setae and other sensory organs. The hypopharynx (a mouthpart between maxillae and labium) is well developed. The coxal vesicles absorb fluids; their number varies from none to two within a coxite. In all machilid males, coxite IX bears the penis (length from $\frac{3}{4}$ of the coxite to $1\frac{1}{2}$ in Machilidae, in Meinertellidae ca. $\frac{1}{2}$ of the coxite length). In females, the paired gonapophyses on coxites VIII and IX form the ovipositor. It is of different length and has generally a specific chaetotaxy in each species.

DEVELOPMENT

The eggs (diameter 0.7–1.3 mm) are deposited into crevices of rocks or bark, and this stage lasts 60 to 380 days. Developmental

stage I (first free-living larva) has a distinct prognathy, a strong forward projection of parts of the maxillae (laciniae), which aids in emergence from the egg, and rod-like setae on the head and the terga. These features disappear after the first molt. From developmental stage III onward, a scale cover is present in all terga and on abdominal coxites and caudal appendages. The molts continue during adult life.

See Also the Following Article

Zygentoma

Further Reading

Bitsch, C., and Bitsch, J. (1998). Internal anatomy and phylogenetic relationships among apterygote insect clades (Hexapoda). *Ann. Soc. Entomol. France (N.S.)* **34,** 339–363.

Mendes, L. F. (1990). An annotated list of generic and specific names of Machilidae (Microcoryphia, Insecta) with identification keys for the genera and geographical notes. *In* "Estudios, Ensaios e Documentos." *Centro Zool. Inst. Invest. Tropicall,* Lisbon.

Sturm, H., and Machida, R. (2001). Archaeognatha. *In* "Handbook of Zoology," *Vol. IV,* Part 37. Walter de Gruyter, Inc., Berlin/New York.

Wygodzinsky, P. W., and Schmidt, K. (1980). Survey of the Microcoryphia (Insecta) of the northeastern United States and adjacent provinces of Canada. *Ann. Mus. Novitates* **2701,** 1–17.

Arthropoda and Related Groups

James H. Thorp
University of Kansas

More than 75% of all described species in terrestrial, freshwater, and marine ecosystems are in the phylum Arthropoda. No other invertebrate phylum, with the possible exception of the nematodes, approaches their economic and ecological importance. This article briefly reviews all subphyla and classes of Arthropoda, with its nearly one million described species, and provides more details about major arthropod taxa not covered in other entries in the Encyclopedia. Aspects of arthropod evolutionary relationships, diversity, anatomy, physiology, and ecology are discussed.

OVERVIEW OF THE PHYLUM ARTHROPODA

The Arthropoda is a phylum more diverse than any other living or extinct animal taxon. Counted among this immense assemblage are beetles, butterflies, silverfish, centipedes, scorpions, mites, sea spiders, crabs, sow bugs, and barnacles, and many other common names too numerous to mention. Arthropods are the numerically dominant metazoan on land and rank among the most prominent benthic (bottom-dwelling) and planktonic members of freshwater and marine ecosystems. They colonize virtually every conceivable habitat—from the equator to the poles, from high mountains to deep ocean trenches, and from rain forests to deserts and hot springs—and fill all trophic niches above the level of primary producer. Parasitism, especially ectoparasitism, is common in some groups, but most species are free-living. They range in size from tiny gall mites (80 μm) to Japanese spider crabs with leg spans of 3.6 m. While some arthropods are vectors for human disease-causing organisms and others are major agricultural competitors with humans, they are also vital to the functioning of most ecosystems and a boon to humans in many ways. In addition to deriving nutrition from some arthropods (e.g., directly or indirectly from bees, crabs, lobsters, and shrimp), humans probably could not survive without arthropods.

The name "Arthropoda" is from the Greek, meaning "jointed foot." The presence of jointed appendages is the primary feature distinguishing arthropods from other phyla. Advantages provided by these appendages, a metameric or segmented body, and a hard skeleton are the three most important reasons for the phylum's success. Arthropods are segmented like annelid worms, but the evolutionary trend has been to fuse several metameres into body regions (tagmata) with specialized functions. Spiders have two tagmata, insects have three, and many crustaceans have two; however myriapods (millipedes and centipedes) lack tagmata. Arthropods have chitinous and proteinaceous exoskeletons that are frequently strengthened with calcium salts. A modest, nonchitinized endoskeleton of inwardly projecting apodemes aids muscular attachment. To allow for continued somatic growth, the exoskeleton is shed periodically during ecdysis, a relatively strenuous and often dangerous process. Modifications of the exoskeleton have permitted arthropods to fly, swim, run, and burrow effectively.

Except for the molluscan cephalopods (e.g., the octopus), arthropods surpass all invertebrates in internal organ complexity. Although they are a phylum with a coelom, this structure no longer serves as a hydrostatic skeleton (as in annelids) but persists only as a cavity surrounding reproductive and/or excretory organs. The principal body cavity is instead the hemocoel, which is derived from the circulatory system. The open circulatory system consists of a dorsal heart, blood sinuses, and one or more discrete vessels. Hemocyanin and hemoglobin are the principal oxygen-carrying blood pigments. Respiration is achieved through the skin surface in some small species, but with gills in most aquatic organisms and tracheae and/or book lungs in terrestrial species. Excretory and osmoregulatory organs vary in type in accordance with the typical environmental moisture and salt content, as do the primary excretory products (ammonia in water and usually either uric acid or guanine on land). Cilia are absent externally and internally. The neural system is highly developed, with brain centers and complex sensory organs; indeed, next to vertebrates and cephalopods, the arthropod brain is the most complex on earth.

Most species reproduce sexually and are primarily dioecious (i.e., with an individual being a single gender), although

parthenogenesis occurs in many taxa. Courtship and brood care are uncommon but are found in some members of all subphyla. Rather than possessing the spiral cleavage typical of many other protostomates, arthropods usually develop by superficial cleavage of a cytoplasmic layer above a yolky sphere. Larvae or discrete juvenile stages are common in terrestrial and aquatic taxa, but aquatic larvae never resemble the trochophore larvae that characterize related phyla.

EVOLUTIONARY RELATIONSHIPS WITH OTHER PHYLA

Arthropods were traditionally linked with the phylum Annelida in the phylogenetic clade Articulata because both are metameric phyla, but more recent molecular analyses provide no support for a particularly close relationship. Instead, arthropods seem to be linked with other phyla that must shed their cuticle during ecdysis to grow. This clade of "Ecdysozoa" also includes the phyla Tardigrada, Onychophora, Nematoda, and Nematomorpha, along with the more distantly related Priapulida and Kinorhyncha. At a greater phylogenetic distance from Ecdysozoa is the other major protostomate clade consisting of the phyla Rotifera, Annelida, Mollusca, Bryozoa (Ectoprocta), Brachiopoda, and Phoronida. Pentastomida is listed by some authors as a separate phylum with links to Arthropoda through the clade Ecdysozoa, but here it is included as a class within the arthropod subphylum Crustacea.

Within Ecdysozoa, the three closest phyla are Arthropoda, Tardigrada, and Onychophora. This conclusion is based in part on molecular studies using 18*S* rRNA. Fossil evidence from the mid-Cambrian (~520 mya) suggests that onychophoran-like limbs developed in aquatic invertebrates and may have served as a preadaptation for terrestrial life. Similarities in morphology and physiology also seem to link these phyla. For example, tardigrades possess striated muscles, paired ventral nerve cords, and a large hemocoel. In addition to these characteristics, onychophorans have the following arthropod-like features: a tracheal respiratory system, mandible-like mouth appendages, cardiac ostia, an excretory system comparable to the green gland of crustaceans, one pair of antennae, and similar defensive secretions produced by repugnatorial glands. Both Onychophora and Tardigrada, however, have some decidedly non-arthropod-like characteristics (e.g., nonjointed legs). Onychophora, Myriapoda, and Hexapoda are grouped by some systematists into Uniramia, a single phylum of arthropod-like animals having a single branch (ramus) of body appendages. According to this theory Uniramia is phylogenetically isolated from Crustacea and Chelicerata, but all are in the superphylum Arthropoda.

TAXONOMIC DIVERSITY AND INTRAPHYLETIC AFFILIATIONS

Arthropoda is treated here as a monophyletic clade of genetically diverse but evolutionarily linked species. Some zoologists, however, maintain that this alleged phylum is actually an artificial, polyphyletic grouping of similar taxa evolving multiple times from different prearthropod ancestors. Much of this debate has centered on evolutionary relationships between the phyla Arthropoda and Onychophora. Classified within Arthropoda are one extinct subphylum (sometimes called super class), the Trilobitomorpha (trilobites), and four living subphyla: Chelicerata (spiders, mites, horseshoe crabs, and sea spiders), Myriapoda (millipedes and centipedes), Hexapoda (springtails, bristletails, beetles, flies, true bugs, etc.), and Crustacea (crayfish, barnacles, water fleas, pill bugs, etc.). Sometimes the number of extant subphyla is reduced to three (Chelicerata, Uniramia, and Crustacea) or even two groups (Chelicerata and Mandibulata).

Molecular studies of arthropod phylogeny present a reasonably clear picture of relationships among three of the four living subphyla. Chelicerates are evolutionarily distinct from insects and crustaceans, and they differ from all other living arthropods in lacking a tagma for either a "head and trunk" or a "head, thorax, and abdomen." Instead, they possess an anterior prosoma without a distinct head and a posterior opisthosoma. Another major clade evident from gene sequences is Mandibulata, composed of the other three extant subphyla. Morphological observations of appendages would seem then to link Myriapoda and Hexapoda into a group (Uniramia) of taxa with only one branch to each appendage and distinct from the biramous Crustacea, but molecular evidence is inconclusive on this point. In some gene sequence trees, myriapods are tightly linked with insects, while other molecular analyses show the millipedes and centipedes as deeply entangled within other genetic branches.

Accurate estimates of both relative and absolute diversities of arthropods are often problematic because of the enormous species richness, large number of unexplored habitats, greater emphasis on studies of economically important taxa, and increasingly serious lack of qualified taxonomists. For those reasons, the literature is replete with divergent estimates of the total number of species in most groups, especially the insects and mites. Table I lists the classes of Arthropoda and includes estimates of taxonomic diversity.

SUBPHYLUM TRILOBITOMORPHA

Trilobites probably played a crucial role in the evolution of living arthropods. Members of the now extinct subphylum Trilobitomorpha began roaming primeval seas in the Precambrian, reached their zenith in the late Cambrian with 4000–10,000 species, and then slowly went extinct around 230 to 275 mya. They were flattened, bilaterally symmetrical arthropods with bodies divided by longitudinal and transverse grooves unlike living arthropods (Fig. 1). Most adults were 2 to 7 cm in length, but giants of 50 cm are known. Their organ systems probably resembled those in modern arthropods. Smaller species were probably planktonic suspension feeders, whereas most species and all larger species were probably benthic deposit feeders or facultative predators. Unlike most other arthropods, trilobites

TABLE I Estimates of Arthropod Diversity, with Comments on Certain Arthropod Taxa and Two Related Phyla[a]

Taxon	Estimated number of species (families)	Biological features
Phylum Arthropoda	1–5 million	
Subphylum Trilobitomorpha	4,000–10,000	Extinct marine trilobites
Subphylum Chelicerata	77,000–1 million	Originally marine, but subsequent evolution has primarily been in terrestrial mites
Class Merostomata	5	Marine horseshoe crabs and extinct sea scorpions
Class Arachnida	~76,000 (550)	Spiders, scorpions, and mites
Class Pycnogonida	1,000 (8+)	Sea spiders
Subphylum Myriapoda	13,000 (140+)	Terrestrial millipedes, centipedes, and others
Class Chilopoda	3,000 (20)	Predaceous centipedes
Class Symphyla	160 (2)	Small (1–8 mm), mostly herbivorous; live in forest litter
Class Diplopoda	10,000 (120)	Millipedes
Class Pauropoda	500 (5)	Minute (< 1.5 mm) dwellers in forest litter
Subphylum Hexapoda	1–4 million	Insects, springtails, bristletails, etc.
Class Ellipura	6,000+	Wingless, entognathous (orders Protura and Collembola, or springtails)
Class Diplura	800 (9)	Blind, wingless inhabitants of forest litter; entognaths
Class Insecta	1–4 million species	Winged and wingless insects; all adults with six pairs of legs
Subclass Apterygota	600 (5)	Primitive, wingless insects (order Thysanura with bristletails and silverfish)
Subclass Pterygota	~1 million+ (906)	Mostly winged insects (grasshoppers, true bugs, beetles, flies, butterflies, ants, etc.)
Subphylum Crustacea	~50,000	Shrimp, crabs, waterfleas, barnacles, copepods, etc.
Class Cephalocarida	9 (2)	Primitive; live in soft marine sediments
Class Malacostraca	29,000 (103)	Crabs, water scuds, isopods, mantis shrimp, etc.
Class Branchiopoda	1,000 (29)	(= Phyllopoda) water fleas and brine, clam, pea, and tadpole shrimp
Class Ostracoda	6,650 (46+)	Seed shrimp enclosed in a bivalved chitinous carapace
Class Mystacocarida	11	Interstitial species living in shallow water or intertidally
Class Copepoda	8,000 (97)	Dominant crustaceans in zooplankton; a few parasites of marine fish and invertebrates
Class Branchiura	125 (1)	Fish lice (ectoparasites)
Class Pentastomida	100 (7)	Highly modified parasites of tetrapod vertebrates
Class Tantulocarida	10+ (9)	Deep-water parasites of crustaceans; some sources estimate diversity up to 1200 species
Class Remipedia	9+	Ancient, vermiform crustaceans found in marine caves; some estimate diversity as high 1200 species
Class Cirripedia	1,000 (20)	True barnacles and small groups of parasitic taxa
Phylum Onychophora	70 (2)	Velvet worms; most confined to tropical habitats
Phylum Tardigrada	800 (17)	Water bears in aquatic and moist terrestrial habitats

[a]Estimates of species richness are for living taxa only, except for the subphylum Trilobitomorpha. The reliability of these estimates varies widely among taxa.

FIGURE 1 *Asaphiscus wheeleri,* an extinct trilobite in the subphylum Trilobitomorpha. (Photograph courtesy of Sam M. Gon, III.)

proceeded gradually through a life with three larval stages, 14 or more juvenile steps in the first year of life, and multiple adult stages lasting a maximum of 3 more years.

SUBPHYLUM CHELICERATA

Few invertebrates on land or sea are so often miscast in a sinister role as the spider, which along with mites, sea spiders, and horseshoe crabs comprises the 77,000 described species in the subphylum Chelicerata. Indeed, most arachnids and all other chelicerates either are harmless to humans or are actually quite helpful in their roles as predators and parasites of insects or as decomposers of terrestrial litter. Although the subphylum evolved in the sea, the ocean now supports only the five species of horseshoe crabs (class Merostomata), a thousand or so species of sea spiders (class Pycnogonida), and a few mites.

Class Merostomata: Horseshoe Crabs

During the full or new moons of late spring and early summer when tides are the highest, vast numbers of horseshoe crabs

FIGURE 2 Representative of the subphylum Chelicerata: ventral view of an adult horseshoe crab, *Limulus polyphemus* (class Merostomata).

(all *Limulus polyphemus;* Fig. 2) come ashore in eastern North America to breed in bays and estuaries. The five Asian and North American species of this class are remnants of this strictly marine class. Although formidable looking and up to 60 cm in length, horseshoe crabs are harmless to humans. When not breeding, these chelicerates reside on or in soft bottoms in shallow water. They are scavengers and predators on clams and worms. Although they can swim weakly by flapping their book gills (modified abdominal appendages), their primary locomotion is walking. Extinct merostomates (order Eurypterida) may have been the largest arthropods ever to have evolved (nearly 3 m long) and seem to have given rise to terrestrial arachnids. Despite their name, horseshoe crabs are not closely related to true crabs (subphylum Crustacea).

Class Arachnida

During the middle Paleozoic, chelicerates made the rigorous transition from water to land; only later did over 5000 species of arachnid mites adopt a secondary aquatic existence. Arachnids then rapidly radiated in form and species richness in association with their predaceous and parasitic exploitation of insects. One crucial factor in this success has been the diverse uses of silk by spiders, pseudoscorpions, and some mites. Although some degree of "arachnophobia" afflicts many people, relatively few of the 76,000 described species are directly harmful to humans because of their venom, link with diseases and allergies, or competition for plant resources. More than balancing their negative attributes is the substantial role in biocontrol of insect pests.

SCORPIONS, SPIDERS, AND HARVESTMEN

True Scorpions The 1500 to 2000 species of true scorpions (order Scorpiones) are elders of the arachnid clan (Fig. 3). In addition to the true scorpions, several other arachnid orders are called "scorpions": false scorpions (Pseudoscorpiones with 2000 species), wind scorpions (Solifugae with 900 species), whip scorpions (Uropygi with 85 species), and tailless whip scorpions (Amblypygi with 70 species). All are much smaller than true scorpions, but are also typically carnivorous.

FIGURE 3 Desert hair scorpion, *Hadrurus arizonensis.* (Photograph by Jim Kalisch, courtesy of University of Nebraska Department of Entomology.)

True Spiders (Order Araneae) Most arachnids lack biting mouthparts and must, therefore, partially digest prey tissue before sucking it into their bodies. Prey are subdued with poison injected by fangs present on each chelicera. Arachnids reproduce with indirect fertilization (without a penis), often after elaborate courtship rituals. Their leglike pedipalps are used by males to transfer spermatophores. Eggs are wrapped in a protective silken cocoon, and brood care is common. Silk is produced normally by caudal spinnerets and by a small platelike organ (cribellum) in cribellate spiders only. Uses for this silk include cocoons, egg sacs, linings of retreats, and capture webs. Locomotion is typically by walking or jumping, but aerial dispersal through the process of "ballooning" with long silken threads is common in most spiderlings and adults of some smaller taxa. Most spiders are terrestrial and are found anywhere insects are located. All are carnivorous, and ecological divergence in prey type and capture method has led to the wide evolutionary radiation. In addition to insects, spiders attack other spiders, small arachnids, and a few other prey taxa including small vertebrates. Several spiders are poisonous to humans, such as the black widow and the brown recluse.

Harvestmen The order Opiliones includes arachnids known as "daddy longlegs," a name reflecting its enormously long walking legs. They are also called "harvestmen" because some species undergo a seasonal population explosion each autumn around the farm harvest. They have "repugnatorial glands" that produce an acrid secretion to repel predators. The 5000 species are more closely related to mites than to true spiders. Most are tropical, but taxa are known from colder subarctic and alpine zones. Opilionids frequent humid forest floors, being less arboreal than true spiders. Although carnivory on small arthropods and worms is common, harvestmen are notable as the only arachnids other than mites that consume vegetation.

MITES AND TICKS At least 30,000 species of arachnid mites and ticks have been described in the order (or subclass) Acari. The major habitat of mites is on land, where they are

either free-living or parasites of plants and animals, but lakes, streams, and even hot springs support 5000 taxa, with the marine fauna being less diverse.

Many acarines are ectoparasitic in larval and/or adult stages. Animal parasites attack mammals (including humans and domestic animals), birds, reptiles, amphibians, fish, aquatic and terrestrial insects, other arachnids, and some other invertebrates, including echinoderms, mollusks, and crustaceans. They are vectors for several human diseases, and some (e.g., chiggers) have annoying bites. Many people develop allergies to mites living on household dust. On the other hand, microscopic mites commonly consume dead tissue and oily secretions on human faces, and they are used to control harmful insects and mites. Their feeding habits and role in spreading viruses make them severe pests of natural and agricultural plants. A great diversity of mites, however, are free-living, mostly in forest and grassland litter, where they feed directly on litter or on microorganisms decomposing detritus. Many mites prey on other mites, nematodes, and small insects.

Class Pycnogonida: Sea Spiders

The body shape and gangling legs of sea spiders call to mind their terrestrial namesakes. Most of the thousand species live in shallow benthic zones at higher latitudes. They are predominantly predators of hydroids, bryozoans, and polychaetes, but some consume microorganisms, algae, and even detritus. Food is either macerated with chelae or externally predigested and then sucked into the digestive tract with a proboscis.

Sea spiders have a barely perceptible head and a body comprising four pairs of long legs joined by a narrow, segmented trunk. Most are small (1–10 mm), but some deep-water behemoths reach 6 cm in body length with a 75-cm leg span. Their eyes are mounted on an tubercle to give them a 360° arc of vision.

SUBPHYLUM MYRIAPODA

Myriapoda ("many feet") is a subphylum of elongate arthropods with bodies divided into a head and trunk with numerous segments, most of which have uniramous appendages; no pronounced tagmatization is evident. Myriapods range in length from 0.5 to 300 mm and are primarily terrestrial. Most live in humid environments, commonly in caves. Some have invaded arid habitats, but few are aquatic. Four classes are recognized: Diplopoda (millipedes), Chilopoda (centipedes), Pauropoda, and Symphyla, with 10,000, 3000, 500, and 160 species, respectively. The last two are minute dwellers of the forest floor that consume living or decaying vegetation. Symphylans look somewhat like centipedes but the adults have 14 trunk segments and 12 pairs of limbs; the posterior end of the trunk has two conical cerci and spinning glands. Members of the class Pauropoda are soft-bodied, blind myriapods with 9 to 11 leg-bearing trunk segments and branched antennae.

Class Chilopoda: Centipedes

Centipedes are dorsoventrally flattened with 15 to 173 segments, each with one pair of legs (Fig. 4A). Poisonous forcipules (fangs) enable centipedes to kill and consume insects, other centipedes, annelids, mollusks, and sometimes small vertebrates; under most circumstances, the poison is not lethal to people. The body is partially hung beneath the legs to increase stability and to allow hind legs to step over front ones, which allows the insects to run swiftly in search of prey or to escape predators. Centipedes are found in most terrestrial environments including the desert fringe; the latter is surprising given their chitinous, noncalcified exoskeleton, which is relatively permeable to water.

Class Diplopoda: Millipedes

Millipedes have a somewhat cylindrical body with 11 to 90 segments (which are really fused "diplosegments") and two pairs of legs per segment. Segmental plates are constructed to prevent "telescoping" as the body bulldozes through forest litter, while still allowing the animal to roll up or coil when threatened (Fig. 4B). Millipedes are slow moving and herbivorous by nature, eating decaying leaves and wood. They lack poisonous fangs and instead repel predators with volatile poison produced by repugnatorial glands. Millipedes are relatively long-lived, with some surviving 7 years.

SUBPHYLUM HEXAPODA

Hexapoda ("six feet") includes a tremendous diversity of winged insects (class Insecta, subclass Pterygota) and many fewer wingless insect (subclass Apterygota) and noninsect classes (Diplura and Ellipura, orders Collembola and Protura).

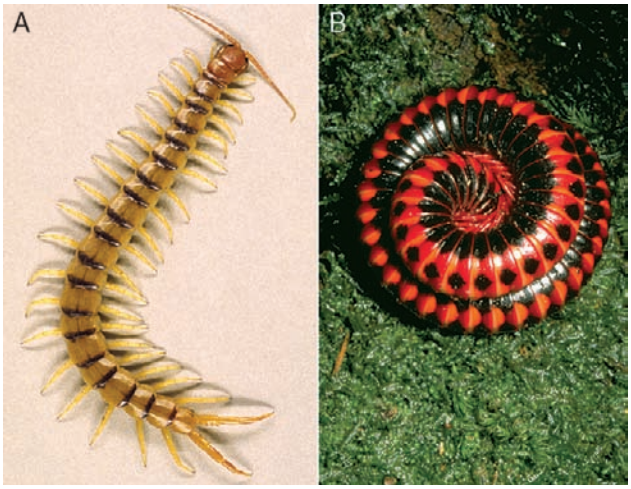

FIGURE 4 Members of the subphylum Myriapoda. (A) Centipede (class Chilopoda). (Photograph by Jim Kalisch, courtesy of University of Nebraska Department of Entomology.) (B) Millipede (class Diplopoda), coiled in a defensive posture. (Photograph by D. R. Parks.)

Hexapods have three major body regions (head, thorax, abdomen) and six thoracic legs.

Entognathous Hexapods: Collembola, Protura, and Diplura

Entognathous hexapods include two small taxa (class Diplura and Ellipura, order Protura) living in moist forest litter and a large group of springtails (class Ellipura, order Collembola) with at least 4000 species in terrestrial and semiaquatic environments. Most springtails live in moist terrestrial environments, but some colonize the surface film of quiet fresh and marine waters. They occur at densities much higher than almost any other invertebrate in soil litter. Unlike insects, springtails have only six abdominal segments, and cleavage of their eggs is total. Their name is derived from their ability to spring forward several centimeters when a forked structure (the furcula) flexed under the abdomen is rapidly uncocked. They have indirect fertilization, the young closely resemble adult Collembola, and adults continue molting throughout their lives (2–50+ molts). Springtails feed on decomposing organic matter or on microorganisms at the water surface. Proturans are completely terrestrial, their antennae have nearly atrophied away, and their front legs function somewhat like antennae. Diplurans are primitive hexapods whose ancestors may have given rise to both Protura and Collembola, and they are more closely related to insects than are ellipurans. Diplurans are blind and have two prominent abdominal cerci.

Class Insecta: Winged and Wingless Insects

The million or so species in the subclass Pterygota include all winged invertebrates and some insect species that have secondarily lost wings during evolution. They include two orders of ancient winged insects (Ephemeroptera and Odonata) and some 25 to 30 (depending on the classification system) orders of modern folding-wing insects. Most have 11 abdominal segments. The head features two antennae and compound eyes. Respiration is generally with internal tracheae, but aquatic species may use external, tracheate gills or other means to obtain sufficient oxygen. Fertilization is usually direct, distinct developmental stages are common, and molting generally stops with attainment of reproductive maturity. Their most prominent features are two pairs of wings, but a great many insects (e.g., fleas) lack wings or have dispensed with either the hind (e.g., flies) or fore pair (beetles). Among their beneficial attributes are pollination of most flowering plants, production of honey and silk, predation on harmful insects, decomposition of animal wastes and carcasses, and facilitation of ecological processes at all trophic levels above primary producer. Negative attributes include transmission of diseases, annoying bites, and damage to crops, stored food, ornamental plants, forests, and wooden structures.

All insect species that did not evolve from a winged hexapod and whose adults all lack wings are in the subclass Apterygota, order Thysanura. This small group of 600 or so primitive species includes bristletails, silverfish, and rock jumpers. These are small to medium-sized insects (5–25 mm) without compound eyes. They have an 11-segmented abdomen with a prominent caudal filament between two terminal cerci. Fertilization is indirect, and molting continues after the reproductive state has been reached (unlike insects). No pronounced metamorphosis is evident from subadult to adult stages. Thysanurans are swift, agile runners (probably to avoid predators) and are omnivorous scavengers of animal and plant matter. Most live in litter of forests and grasslands, but silverfish also infest houses, where they can extensively damage clothing and books.

SUBPHYLUM CRUSTACEA

Crustaceans surpass all other invertebrates in their direct contribution to human diets (from crabs, shrimp, lobsters, and crayfish) and are vitally important to many ecosystems, especially planktonic food webs. Unfortunately, they also foul boat hulls (barnacles) and destroy wooden piers in coastal waters (burrowing isopods).

Distinguishing characteristics of adults include the following: five-pairs of cephalic appendages (two mandibles, four maxillae, and two antennae), two to three tagmata, a chitinous cuticle often elaborated as a shieldlike carapace, more than 11 abdominal segments, and jointed, biramous appendages. Evolutionary trends involved specialization of mouthparts, body segments, and appendages for locomotion, sensory reception, and reproduction. Respiration is typically with gills, and hemocyanin is the principal respiratory pigment. Excretion of ammonia generally occurs through modified nephridia. Sexes are mostly separate, but hermaphroditism is common. Development always includes triangular nauplius larvae (with six appendages and a median eye), which are commonly planktonic. Many crustaceans have a relatively sophisticated behavioral repertoire and communicate visually, tactilely, and chemically.

Most of the roughly 50,000 species are marine, but crustaceans are ubiquitous in freshwater habitats and a few species have colonized saline lakes and terrestrial environments. Crustaceans are most often scavenging predators or have a generally omnivorous diet. They range in length from minute to truly gigantic (0.25 mm to 360 cm).

Class Malacostraca: Shrimp, Crayfish, and Others

About 60% of crustacean species are malacostracans, including all species consumed by humans (mostly decapods) and many important benthic crustaceans in marine and freshwater ecosystems (e.g., amphipods and isopods).

Most orders and the 29,000 species in this class are dominated by crustaceans that live near the bottom. Ecologically important exceptions include planktonic krill (order Euphausiacea), a vital prey of baleen whales. Parasitism is relatively rare and is confined mostly to ectoparasites of fish, crustaceans

(both mostly in the large order Isopoda), and marine mammals (small component of the large order Amphipoda).

Class Branchiopoda: Water Fleas, Brine Shrimp, and Others

Nearly 25% of all freshwater crustaceans are branchiopods ("gill feet"), and almost all crustaceans in saline lakes are in this class; only 3% of the species occur in oceans. The class contains about 800 species of water fleas (in four orders of cladocera), 200 species of fairy and brine shrimp (order Anostraca), and a few taxa of clam (order Spinicaudata), pea (Laevicaudata), and tadpole shrimp (Notostraca). Most taxa other than cladocerans are restricted to ephemeral pools, and most branchiopods are suspension feeders. Branchiopods are noted for producing dormant embryos resistant to adverse environments.

Class Ostracoda: Seed Shrimp

The 6650 species of seed or mussel shrimp are minute crustaceans characterized by a protective, dorsal, bivalved carapace of chitin heavily impregnated with calcium carbonate. Freshwater species rarely exceed 3 mm in length, but marine taxa can reach 30 mm. Ostracodes are especially prevalent in freshwater habitats (particularly benthic areas) but are common in marine environments. A few genera have adapted to damp humus habitats of the forest floor. Ostracods are typically suspension feeders on benthic and pelagic detritus, and almost all are free-living.

Class Copepoda

In relatively permanent freshwater and marine environments, the 8000 species of copepods are the most important planktonic crustaceans, and other species contribute to the benthic fauna. Herbivory on microalgae prevails, but raptorial feeding on other zooplankton is common. Although freeliving copepods predominate, bizarre forms, barely recognizable as copepods, have evolved as parasites of fish and invertebrates.

Class Cirripedia: Barnacles

The name "barnacle" evokes a rugged image of the sea in the minds of many people, but few recognize this taxon's kinship with familiar crustaceans like shrimp and crabs because the barnacle's body is hidden inside calcareous plates and free-living species are sedentary. The 1000 species in the class Cirripedia include free-living species that live directly on a hard surface or are raised on a stalk (peduncle). This surface may be an inanimate object (e.g., rocks, floating refuse, ship hulls) or the outer layer of a living whale, turtle, invertebrate, or other larger organism. Barnacles usually cement the head to hard surfaces and employ setose legs to capture plankton. Other species, barely recognizable as barnacles, are ecto- and endoparasites of echinoderms, corals, and other crustaceans, especially crabs.

Pentastomida and Other Crustaceans

The remaining 264 to 1500 species of crustaceans are divided among six classes. Of these, only the ectoparasitic fish lice (class Branchiura), with 125 species, were recognized more than a few decades ago. Three classes closely resemble copepods—Cephalocarida, Mystacocarida, Tantulocarida—with the first two living in marine sediments and the last being ectoparasites of deep-water crustaceans. The class Tantulocarida is usually listed with 10 to 20 species, but some scientists believe that the true diversity is greater than 1000. Members of the class Remipedia are presently restricted to tropical underwater caves. Their long bodies with abundant lateral appendages call to mind segmented polychaete worms.

CLOSELY RELATED PHYLA: TARDIGRADA AND ONYCHOPHORA

Tardigrada: Water Bears

Tardigrades are called "water bears" because of their slow lumbering gait and relatively massive claws on lobopodous legs. The permeability of the cuticle limits tardigrades to aquatic habitats (often interstitial), the surface film of terrestrial mosses, and damp soil. About 800 species have been described, but many more undoubtedly exist in unexplored habitats. Faced with inhospitable microhabitats from environmental changes, both terrestrial and aquatic species may undergo cryptobiosis, where the body becomes dehydrated and metabolism is greatly reduced until favorable conditions return. Some water bears have been "resuscitated" from this state after decades! These normally dioecious organisms can also reproduce by parthenogenesis. Tardigrades typically feed on plants cells pierced by a pair of mouth stylets and sucked into the gut, but a few species are carnivorous.

Onychophora: Velvet Worms

Velvet worms are giants compared with tardigrades, for some individuals reach a length of 15 cm, but they share many characteristics with this sister phylum of Arthropoda. They are generally nocturnal and move by extending their legs, with hydrostatic pressure generated by muscular contraction within the legs.

Though most are herbivores or omnivores, many species consume small arthropods in a rather unique manner. They attack their prey and also defend themselves from predators by expelling a sticky, proteinaceous substance that entangles the target. This glue is produced by slime glands within oral papillae. The onychophoran then bites and secretes toxins to kill and partially liquefy the victim.

See Also the Following Articles

Biodiversity • Fossil Record • Insecta, Overview

Further Reading

Anderson, D. T. (ed.) (1998). "Invertebrate Zoology." Oxford University Press, Melbourne, Australia. (See especially Chaps. 10–14, 18.)

Bliss, D. E. (ed.-in-chief). (1982–1985). "The Biology of Crustacea," Vols. 1–10. Academic Press, New York.

Parker, S. P. (ed.). (1982). "Synopsis and Classification of Living Organisms," Vol. 2, McGraw-Hill, New York.

Pechenik, J. A. (2000). "Biology of the Invertebrates." 4th ed. McGraw-Hill, New York (Particularly Chaps. 14 and 15.)

Thorp, J. H., and Covich, A. P. (eds.) (2001). "Ecology and Classification of North American Freshwater Invertebrates." 2nd ed. Academic Press, San Diego, CA. (See especially Chaps. 16–23.)

Auchenorrhyncha

(Cicadas, Spittlebugs, Leafhoppers, Treehoppers, and Planthoppers)

C. H. Dietrich
Illinois Natural History Survey

The hemipteran suborder Auchenorrhyncha is the group of sapsucking insects comprising the modern superfamilies, Cercopoidea (spittlebugs, Fig. 1), Cicadoidea (cicadas, Fig. 2), Membracoidea (leafhoppers and treehoppers, Fig. 3), and Fulgoroidea (Fig. 4) Together, these groups include over 40,000 described species. Morphologically, Auchenorrhyncha differ from other Hemiptera in having the antennal flagellum hairlike (aristoid), the rostrum (modified, beaklike labium) arising from the posteroventral surface of the head, a complex sound-producing tymbal apparatus, and the wing-coupling apparatus consisting of a long, downturned fold on the

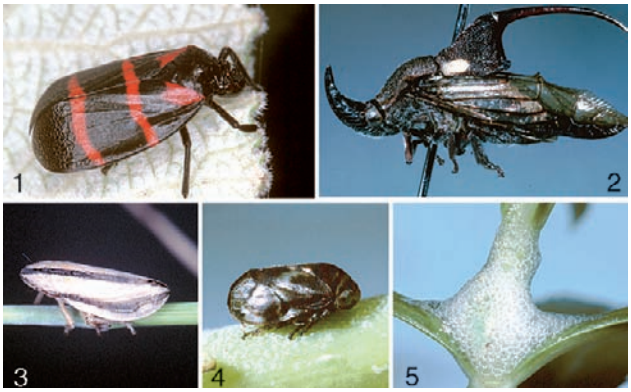

FIGURE 1 Cercopoidea: spittlebugs and froghoppers: (1) *Tomaspis* sp. (Cercopidae), Mexico, (2) *Machaerota* sp. (Machaerotidae), Vietnam, (3) *Paraphilaenus parallelus* (Aphrophoridae), Kyrgyzstan, (4) *Clastoptera obtusa* (Clastopteridae), Illinois, U.S.A., (5) spittle mass of *P. spumarius* nymph, Illinois, U.S.A.

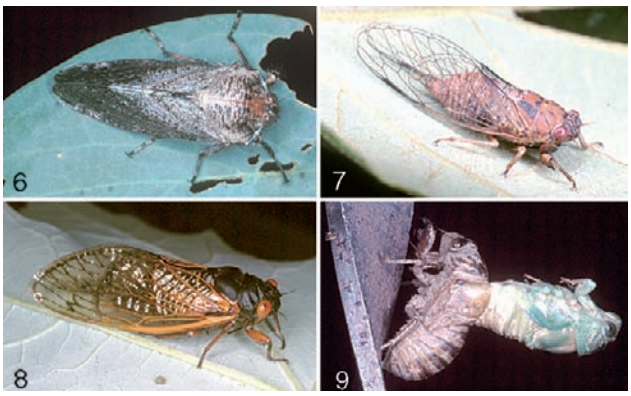

FIGURE 2 Cicadoidea: cicadas: (6) a hairy cicada, *Tettigarcta crinita* (Tettigarctidae), Australia, (7) *Melampsalta calliope* (Cicadidae), Illinois, U.S.A., (8) a periodical cicada, *Magicicada cassini,* with a 13-year life cycle, Illinois, U.S.A., (9) a dog day cicada, *Tibicen* sp., molting into the adult stage, Illinois, U.S.A.

FIGURE 3 Membracoidea: leafhoppers and treehoppers: (10) a brachypterous, grass-feeding leafhopper, *Doraturopsis heros,* Kyrgyzstan, (11) *Pagaronia triunata* (Cicadellidae), California, U.S.A., (12) *Eurymeloides* sp. (Cicadellidae), Australia, (13) fifth instar of *Neotartessus flavipes* (Cicadellidae), Australia, (14) female *Aetalion reticulatum* (Aetalionidae) guarding egg mass, Peru, (15) ant-attended aggregation of treehopper adults and nymphs (Membracidae: *Notogonia* sp.), Guyana.

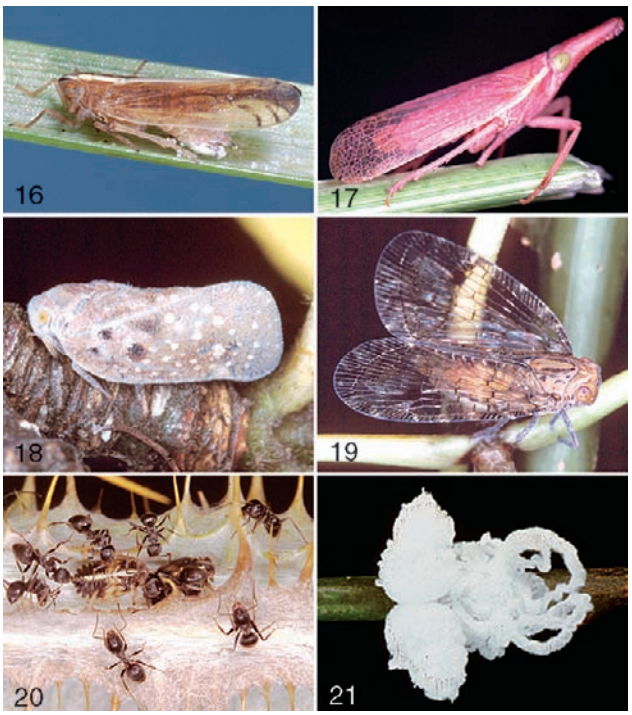

FIGURE 4 Fulgoroidea: planthoppers: (16) female *Stenocranus* sp. (Delphacidae) covering oviposition site with wax, Illinois, U.S.A., (17) *Chanithus scolopax* (Dictyopharidae), Kyrgyzstan, (18) *Metcalfa pruinosa* (Flatidae), Maryland, U.S.A., (19) *Biolleyana* sp. (Nogodinidae), Mexico, (20) *Tettigometra* sp. (Tettigometridae) nymphs tended by ants, Greece, (21) unidentified planthopper nymph completely covered with wax filaments, Guyana.

forewing and a short, upturned lobe on the hind wing. Auchenorrhyncha are abundant and ubiquitous insects, distributed worldwide in nearly all terrestrial habitats that support their host plants, but they are particularly diverse and speciose in the tropics. Some are important agricultural pests, injuring plants either directly through feeding and oviposition, or indirectly through the transmission of plant pathogens.

PHYLOGENY AND CLASSIFICATION

Nomenclature

The monophyly of the four existing superfamilies of Auchenorrhyncha has long been accepted, but controversy persists regarding the relationships of these lineages to each other and to various other fossil and extant hemipteran lineages. Consequently, no single classification scheme has gained universal acceptance, and the nomenclature of the various groups is presently unstable. Traditionally, Auchenorrhyncha were treated as one of three suborders of the order Homoptera. Fossil evidence, as well as phylogenetic analyses based on DNA sequences of extant taxa, suggest that Heteroptera (true bugs; Hemiptera, *sensu stricto*) arose from within Homoptera and, possibly, from within Auchenorrhyncha. Thus, many recent workers have combined Homoptera and Heteroptera into a single order. This order is usually referred to as Hemiptera (*sensu lato*), but some entomologists advocate using the ordinal name Rhynchota to avoid confusion with the more restricted definition of Hemiptera (Heteroptera) widely used in the literature. Some recent workers have further proposed dividing the Auchenorrhyncha into two suborders: Clypeorrhyncha for the lineage comprising Cicadoidea, Cercopoidea, and Membracoidea, and Archaeorrhyncha for Fulgoroidea. The older names Cicadomorpha and Fulgoromorpha, respectively (usually treated as infraorders within suborder Auchenorrhyncha), are more commonly used for these two groups. For convenience, and because the phylogenetic status of the group has not been elucidated satisfactorily, Auchenorrhyncha is retained here as the subordinal name with the caveat that this group may represent a paraphyletic assemblage rather than a monophyletic group. The current classification of families is presented in Table I.

TABLE I Classification of the Hemipteran Suborder Auchenorrhyncha (synonyms and common names in parentheses) Excluding Extinct Taxa

- Auchenorrhyncha (Cicadinea)
  - Infraorder Cicadomorpha (Clypeorrhyncha, Clypeata)
    - Superfamily Cercopoidea (spittlebugs, froghoppers)
      - Aphrophoridae
      - Cercopidae
      - Clastopteridae
      - Machaerotidae
    - Superfamily Cicadoidea (cicadas)
      - Cicadidae (Platypediidae, Plautillidae, Tettigadidae, Tibicinidae)
      - Tettigarctidae (hairy cicadas)
    - Superfamily Membracoidea (Cicadelloidea)
      - Aetalionidae (Biturritiidae)
      - Cicadellidae (Eurymelidae, Hylicidae, Ledridae, Ulopidae, leafhoppers)
      - Melizoderidae
      - Membracidae (Nicomiidae, treehoppers)
      - Myerslopiidae (Cicadellidae, in part)
  - Infraorder Fulgoromorpha (Archaeorrhyncha, planthoppers)
    - Superfamily Fulgoroidea
      - Acanaloniidae
      - Achilidae
      - Achilixiidae
      - Cixiidae
      - Delphacidae
      - Derbidae
      - Dictyopharidae
      - Eurybrachidae
      - Flatidae
      - Fulgoridae (lanternflies)
      - Gengidae
      - Hypochthonellidae
      - Issidae
      - Kinnaridae
      - Lophopidae
      - Meenoplidae
      - Nogodinidae
      - Ricaniidae
      - Tettigometridae
      - Tropiduchidae

Fossil Record

Auchenorrhyncha arose in the Paleozoic, first appearing in the fossil record in the Lower Permian (280 mya) and, judging from the abundance of forms described from Permian strata, they diversified explosively. These early auchenorrhynchans had adults with well-developed jumping abilities and somewhat resembled modern leafhoppers and spittlebugs, but nymphs (juveniles) associated with these insects were bizarrely flattened or biscuitlike, with short legs, foliaceous lobes on the head, thorax, and abdomen (similar to those of some modern Psyllidae) and elongate mouthparts, suggesting a sessile, cryptic lifestyle. The fulgoromorphan and cicadomorphan lineages (Table I) apparently diverged by the middle Permian. By the late Permian, Fulgoroidea appeared and Cicadomorpha (*sensu lato)* had diverged into the Pereboreoidea, comprising three extinct families of large cicada-like insects, and the smaller Prosboloidea, from which the three modern cicadomorphan superfamilies apparently arose. Cicadomorphans with a greatly inflated frontoclypeus (Clypeata in the paleontological literature = Clypeorrhyncha) did not appear until the Mesozoic. Prior to that, the head of Cicadomorpha resembled that of modern Psyllidae in having the frontovertex extended ventrad on the face to the antennal ledges and the lateral ocelli situated close to the eyes. This change in head structure is thought to have been associated with a shift from phloem to xylem feeding. Xylem feeding was apparently the predominant feeding strategy of the group throughout the Mesozoic, but in the late Cretaceous or early Tertiary the major lineages of phloem-feeding leafhoppers and treehoppers, which predominate in the recent fauna, arose. In these insects, the frontoclypeus became more flattened, probably because of the reduction in size of the cibarial dilator muscles. This was presumably in response to a shift from feeding on xylem, which is under negative pressure, to phloem, which is under positive pressure. Cicadoidea and Cercopoidea first appeared in the Triassic, and Membracoidea in the Jurassic. With the exception of Tettigarctidae, which arose in the late Triassic and is now confined to Australia, extant families of these groups do not appear in the fossil record until the Cretaceous or early Tertiary. Most Auchenorrhyncha from Baltic and Dominican amber of the Tertiary age are virtually indistinguishable from modern forms.

LIFE HISTORY

Courtship

Adult male and female Auchenorrhyncha locate each other by means of species-specific acoustic courtship signals. These signals are produced by specialized organs at the base of the abdomen called tymbals, present in both sexes (except female cicadas). A few cicadas and planthoppers are also able to use the stridulatory surfaces of their wings to produce sound. The loud, sometimes deafening, calls of many male cicadas are well known. In noncicadoids, the courtship calls are usually inaudible, being transmitted through the substrate, and distinct tympana are absent. The calls of some leafhoppers and planthoppers, audible only with special amplifying equipment, are among the most complex and beautiful of any produced by insects. Males move from plant to plant, signaling until they receive a response from a female. In addition to intensification of the vibrational signals, precopulatory behavior in some species may involve the male buzzing or flapping the wings, tapping the female with the legs, or repeatedly walking around or over the female. Copulation involves insertion of the male aedeagus into the female vulva at the base of the ovipositor and may last from a second or less to several hours, depending on the species. Females of most species seem to mate only once, while males often mate several times.

Oviposition and Nymphal Development

Females lay eggs singly or in batches, usually either by inserting them into plant tissue or by depositing them on plant surfaces [Figs. 3(14), 3(15), and 4(16)]. In some groups, eggs are deposited in the soil or in litter. Egg batches may be covered with plant debris, wax filaments, or secretions produced by various internal glands. Eggs may or may not undergo diapause depending on the species and climate. After hatching, the juveniles [nymphs, Figs. 2(9), 3(13), 3(15), 4(20), and 4(21)] undergo five molts prior to reaching the adult stage. In most species the nymphs feed on aboveground parts of host plant, but in cicadas, Cercopidae, a few fulgoroid families, and a few leafhopper genera, the nymphs are subterranean root feeders. Formation of galls, common among aphids and psyllids, is known in only one Auchenorrhyncha species (a leafhopper). Nymphal development requires from a few weeks to several years (in cicadas), depending on the species. Some species exhibit parental care behavior (see later).

BEHAVIOR AND ECOLOGY

Feeding and Digestion

Adult and nymphal Auchenorrhyncha feed by inserting the two pairs of feeding stylets (modified mandibles and maxillae) into the host plant tissue, injecting saliva, and ingesting fluid. Unlike Sternorrhyncha, in which the stylets pass between the cells of the host tissue (intercellular feeding), Auchenorrhyncha stylets usually pierce the cells (intracellular feeding). After selecting an appropriate feeding site based on visual and chemical cues, the insect presses the tip of the labium onto the plant surface and inserts the feeding stylets. Just prior to, and during probing of the plant tissue with the stylets, the insect secretes sheath saliva that hardens on contact with air or fluid to form an impervious salivary sheath surrounding the stylets. The sheath forms an airtight seal that prevents leakage of air or fluid during feeding. Stylet probing continues until a suitable tissue is found (xylem, phloem, or mesophyll, depending on the

species), after which feeding can commence. During feeding, watery saliva is injected into the plant to aid digestion and to prevent clogging of the stylet opening. This is also the mechanism by which the insect may infect the plant with pathogens (see later). Feeding may last from a few seconds to many hours at a time, depending on the auchenorrhynchan species and the quality of the plant tissue. During feeding, droplets of liquid excretion are ejected from the anus, several droplets per second in some xylem feeders.

Plant sap is a nutritionally imbalanced food source; phloem is high in sugar and xylem is, in general, nutrient poor and extremely dilute. Auchenorrhyncha have acquired various adaptations that enable them to convert the contents of plant sap into usable nutrients. Most Cicadomorpha have part of the midgut modified into a filter chamber that facilitates rapid removal of excess water. Fulgoroidea lack a distinct filter chamber but have the midgut tightly coiled and partially or completely enclosed in a sheath of specialized cells that apparently absorb solutes from the gut contents. A broad array of transovarially transmitted (i.e., from the mother through her eggs to her offspring) prokaryotic endosymbionts have also been identified in various Auchenorrhyncha species. The roles of these endosymbionts have not been fully elucidated, but presumably they function in the conversion of the nutritionally poor plant sap on which the insects feed into essential vitamins, amino acids, and sterols. The symbionts are housed either intracellularly in specialized fat body cells called mycetocytes, intracellularly in the fat body, or in the gut epithelium. Several distinct mycetomes, consisting of groups of mycetocytes, are often present. In Cicadomorpha, each mycetome may house up to six different kinds of endosymbiont. In Fulgoroidea, only a single kind of endosymbiont is housed in each mycetome.

Host Associations

Nearly all Auchenorrhyncha are plant feeders; the few known exceptions (e.g., Fulgoroidea: Achilidae and Derbidae) feed on fungi as nymphs. Auchenorrhynchans use a wide variety of plants including mosses, horsetails, ferns, cycads, conifers, and angiosperms, but the vast majority of species feed on flowering plants. Most species appear to be restricted to a single genus or species of plants. Many species, particularly among the xylem-feeding groups, normally use a few or a single plant species but are capable of feeding and developing on a variety of alternate hosts if the preferred host is not available. A few xylem-feeding species have extremely broad host ranges. For example, the meadow spittlebug, *Philaenus spumarius,* with over 500 documented food plants, has the broadest known host range of any herbivorous insect. Phloem- and mesophyll-feeding species, comprising the majority of Auchenorrhyncha, tend to have narrower host ranges than xylem feeders, and many species appear to use a single plant family, genus, or species. Host associations appear to be conservative in some auchenorrhynchan lineages. Delphacidae and Cicadellidae (Deltocephalinae) include large numbers of grass- and sedge-specialist species and are among the dominant herbivores in grasslands. Most of the major lineages of Auchenorrhyncha do not exhibit a distinct preference for any particular plant taxon and usually include both host-generalist and host-specialist species. Some species alternate hosts during different stages in the life cycle or in different seasons. For example, nymphs of many leafhoppers and treehoppers develop on herbs, but the adult females oviposit on a woody host.

Migration

Most species of Auchenorrhyncha are relatively sedentary, completing their life cycle within a small area. Although most species have well-developed wings and are strong fliers, few seem to move more than a kilometer from their birthplace. Many species, particularly those inhabiting grasslands and deserts, are submacropterous or brachypterous [short winged, Fig. 3(10)] and, thus, incapable of sustained flight. Some of these species occasionally produce macropterous (long-winged) females that move to new patches of suitable habitat. Other species produce both short- and long-winged forms either simultaneously or in alternate generations. The proportion of macropterous to brachypterous forms often varies in response to population density. Some Auchenorrhyncha species undergo annual migrations that may cover hundreds of kilometers. Not coincidentally, many of these accomplished migrants are important agricultural pests. Among the best studied of these are the brown planthopper (*Nilaparvata lugens)* and the potato leafhopper (*Empoasca fabae).* Neither of these species can normally overwinter in high latitudes. Populations build up in the tropical or subtropical parts of their range and migrate to higher latitudes each spring. They are assisted in their migratory flights by convection and favorable winds, and the initiation of migratory behavior is apparently triggered by favorable atmospheric conditions. Sporadic incidents of very-long-range migrations have also been documented. In one such incident in 1976, swarms of *Balclutha pauxilla* (Cicadellidae), probably originating from a source population in Angola, descended on Ascension Island, 2700 km away in the mid-Atlantic.

Thermoregulation

Most Auchenorrhyncha species appear to regulate their body temperature behaviorally, by seeking out microhabitats in which the ambient temperature remains within a narrow range and moving among alternate microhabitats as conditions change. In some cicadas, physiological mechanisms are also involved. Some species are facultatively endothermic, producing metabolic heat to facilitate calling, courtship, and other activities. This is usually accomplished by vibrating the flight or tymbal muscles until the body temperature rises to

an optimal level. Some desert cicadas cool themselves by evaporation of excess water released through pores on the thorax and abdomen. In this way they are able to remain active at ambient temperatures that would kill other insects.

Defense and Escape

Because they are among the most abundant phytophagous insects in many habitats, Auchenorrhyncha are an important food source for numerous vertebrate and invertebrate predators (see next section: Natural Enemies). Species of Auchenorrhyncha exhibit myriad strategies for avoiding predation. These range from relatively simple behaviors, such as dodging around to the opposite side of a leaf or branch as a predator approaches, or hiding under a leaf sheath, to complex mutualistic associations and mimicry. Adults of many species are strong flyers and nearly all (except cicadas) are also excellent jumpers. Juvenile (nymphal) cicadas, spittlebugs, treehoppers, and some planthoppers are incapable of jumping and have adopted other strategies for avoiding predators. All cicada nymphs and many spittlebug and planthopper nymphs are subterranean; thus, their exposure to most predators is minimal. Spittlebug nymphs live within masses of froth and machaerotid nymphs live in calcareous tubes cemented to the host plant. The free-living nymphs of most other auchenorrhynchans appear to rely on cryptic coloration and body forms to escape detection by visual predators such as birds. For example, many treehopper nymphs are strongly flattened with the ventral surfaces of the body concave, enabling them to lie flat against the bark or leaf surfaces of their host plant. Others resemble plant parts such as bud scales or leaflets. Many planthopper nymphs secrete copious quantities of wax [Fig. 4(21)], with which they coat themselves and, often, surrounding parts of their host plants. The wax may prevent parasites and predators from grasping the nymphs, allowing them to leap away. Adults of some species mimic various venomous arthropods such as ants, wasps, robber flies, assassin bugs, and spiders. Some bear horns or spines on the pronotum [Membracidae, Fig. 3(15)] or scutellum [Machaerotidae, Fig. 1(2)] that make them physically difficult for some vertebrate predators to swallow. Many adult cercopids and membracids have conspicuous (aposematic) color patterns, presumably indicating that they are unpalatable. Others have the forewing apices marked with false eyespots, and a few (e.g., Fulgoroidea: Eurybrachidae) have prolongations resembling antennae; the head and thorax of such species often bear transverse lines resembling abdominal segmentation. Adults of various planthopper species mimic lizards, flowers, and lichens. Another strategy involves complex mutualistic associations with ants and other social hymenopterans. Ant mutualism has been documented in numerous lineages of Fulgoroidea and Membracoidea and occurs universally in some groups [e.g., tettigometrid planthoppers Fig. 4(20) and eurymeline leafhoppers]. In such groups, the nymphs usually form aggregations that are tended by ants. The aggressive worker ants drive off predators and receive gifts of honeydew, a sugary excretion, from the nymphs. Ant mutualism may have facilitated the development of subsocial behavior in some groups (see Membracoidea section under Diversity).

Natural Enemies

Auchenorrhyncha are preyed upon by insectivorous vertebrates such as birds and lizards, as well as by invertebrate predators such as spiders, ants, assassin bugs, wasps, and robber flies. Auchenorrhyncha are also attacked by various parasitoids such as dryinid and chalcidoid wasps, epipyropid moths, pipunculid flies, strepsipterans, and nematodes. Because they feed on plant sap, cicadomorphans are not usually susceptible to infection by viral, bacterial, or protozoan pathogens. Thus, entomopathogenic fungi, which do not need to be ingested to infect insects, are the most important pathogens of Auchenorrhyncha.

Economic Importance

Although the vast majority of species of Auchenorrhyncha are benign, the group contains some of the most destructive pests of agriculture. Among the most important are the brown planthopper, sugarcane planthopper (*Perkinsiella saccharicida*), corn planthopper (*Peregrinus maidis*), meadow spittlebug, beet leafhopper (*Neoaliturus tenellus*), potato leafhopper, corn leafhopper (*Dalbulus* spp.), African maize leafhopper (*Cicadulina* spp.), green rice leafhopper (*Nephotettix* spp.), and various grape leafhoppers (*Arboridia* and *Erythroneura* spp.).

Auchenorrhyncha injure plants directly through feeding or oviposition or, more often, indirectly through the transmission of plant pathogens. Economic injury to plants involving cicadas, which occurs rarely, is mainly due to oviposition, although some species occasionally inflict feeding damage (e.g., on sugarcane). Spittlebugs injure plants primarily through feeding and through transmission of xylem-limited bacterial pathogens. Species of Cercopidae are the most significant pests of forage grasses in pastures in Latin America and are also destructive of sugarcane. Interestingly, much if not most of the economic damage done by spittlebugs is due to native spittlebug species colonizing nonnative hosts (e.g., introduced forage grasses, clovers, etc.). Presumably, such plants lack natural resistance to spittlebugs and are more susceptible to injury.

Leafhoppers and planthoppers are among the most significant groups of vectors of plant pathogens, transmitting viruses, bacteria, and mycoplasma-like organisms. Over 150 species are known vectors of economically important plant pathogens. The insects usually acquire the pathogen by feeding on an infected plant, but some pathogens may be transmitted transovarially from mother to offspring. Phloem-limited viral and mycoplasma-like pathogens typically multiply within the vector and enter the plant when the insect injects saliva during feeding. Some xylem-limited bacterial pathogens (e.g., *Xylella*) are apparently unable to travel from

the gut to the salivary glands and require regurgitation from the foregut during vector feeding to infect the plant. Annual losses to maize, rice, and sugarcane attributed to pathogens spread by leafhoppers and planthoppers are estimated in the hundreds of millions of dollars. Xylem-feeding cicadelline leafhoppers are also the main vectors of *Xylella fastidiosa,* which causes X diseases of stone fruits (*Prunus* spp.), Pierce's disease of grape, citrus variegated chlorosis, and alfalfa dwarf.

Some Auchenorrhyncha species are considered to be beneficial. Cicadas are used as food by several human cultures. The use of Auchenorrhyncha in biocontrol of weeds has also begun to be explored. For example, a Neotropical treehopper species (*Aconophora compressa)* has been introduced into Australia for control of *Lantana* (Verbenaceae).

Control

Control of auchenorrhynchan pests has traditionally involved the use of conventional contact insecticides, but overuse of chemical insecticides has led to the development of resistance in many pest species and has suppressed populations of their natural enemies. Modern integrated pest management has promoted greater use of resistant plant varieties, cultural control (e.g., removal of litter to reduce numbers of overwintering individuals), and biological control by means of parasitoids and pathogens, as well as more judicious use of pesticides.

CAPTURE AND PRESERVATION

Auchenorrhyncha are most commonly collected by sweeping vegetation with a heavy canvas net. Many species are also attracted to lights. Vacuum collecting is effective for collecting from dense grassy vegetation where many species reside. A gasoline-powered leaf blower fitted with a vacuum attachment can be used to suck the insects from dense vegetation. A fine-mesh insect net bag taped to the end of the intake nozzle will capture the specimens. Other effective collecting methods include Malaise trapping and insecticidal fogging of forest canopy. Auchenorrhycha may be killed in a standard insect killing jar containing potassium cyanide or ethyl acetate, or by freezing.

Specimens for morphological study are usually mounted dry on pins or point mounts. Point mounts should be glued to the right side of the thorax. To identify the species of a specimen, it is often necessary to examine the male genitalia. To do this, the abdomen is removed and soaked in 10% potassium hydroxide solution for several hours (or boiled in the same solution for a few minutes) to clear the pigment. The abdomen is then rinsed in clean water containing a small amount of glacial acetic acid, rinsed again in pure water, and immersed in glycerine. After examination, the cleared abdomen is stored in a glass or plastic microvial pinned beneath the rest of the specimen. Auchenorrhyncha may also be preserved indefinitely in 80 to 95% ethanol, but this causes some green pigments to fade to yellow.

DIVERSITY

Cercopoidea

Cercopoidea (froghoppers and spittlebugs, Fig. 1) are characterized by the following combination of morphological characters: head with frontoclypeus inflated; median ocellus absent; ocelli on crown distant from margin; pronotum extended to scutellar suture; body clothed with fine setae; hind coxae conical, tibia without rows of setae but often with one or more conspicuous spines; male subgenital plate present. The superfamily comprises four families Aphrophoridae, Cercopidae, Clastopteridae, and Macherotidae. The first Cercopoidea (Procercopidae) appear in the fossil record during the Lower Jurassic. These insects retained a median ocellus and apparently lacked the dense setal covering of modern cercopoids. Aphrophoridae and Cercopidae did not appear until the middle Cretaceous; Clastopteridae and Machaerotidae apparently arose during the Tertiary.

Approximately 2500 species and 330 genera of Cercopoidea have been described. The classification has not been revised in over 50 years, and the phylogenetic status of most cercopoid genera and higher taxa remains unknown. Cercopidae [Fig. 1(1)], the largest family, differs from Aphrophoridae [Fig. 1(3)], the next largest, in having the eyes slightly longer than wide and the posterior margin of the pronotum straight (instead of emarginate). The small families Machaerotidae and Clastopteridae differ from other Cercopoidea in having a well-developed appendix (distal membrane) on the forewing. Machaerotidae [Fig. 1(2)] differ from Clastopteridae [Fig. 1(4)] in having two or more r-m crossveins in the forewing and in lacking an outer fork on the radial vein of the hind wing.

Production of "spittle" is a unique characteristic of Cercopoidea [Fig. 1(5)]. Nymphs of Machaerotidae produce the froth during molts, while in other families nymphs live permanently surrounded by the froth. The lateral parts of nymphal abdominal segments are extended ventrally into lobes, which form an open or closed (in machaerotids) ventral cavity, filled with air. The nymphs introduce bubbles of air into their liquid excretion by bellowslike contractions of this device; periodically the tip of the abdomen is extended through the surface of spittle mass to channel air into the cavity. The same air supply is used for breathing via spiracles that open into the ventral cavity. The froth is stabilized by the action of the secretory products manufactured in the highly specialized Malpighian tubules of the nymphs and mixed into the main watery excreta. Wax secreted by plates of epidermal glands on the sixth through eighth abdominal terga (Batelli glands) may also help stabilize the froth.

The function of the spittle mass is not completely understood. It is usually assumed that it protects the insect from predators and desiccation. Cercopoid nymphs are sessile and live within the spittle mass (or, in Machaerotidae, inside fluid-filled tubes). In some species, nymphs tend to aggregate, forming large spittle masses containing hundreds

of individual nymphs. Nymphs of Cercopidae apparently feed on roots, whereas aphrophorid and clastopterid nymphs occur on aboveground parts of their host plants. Nymphs of the Machaerotidae live immersed in liquid inside tubes cemented to the twigs of their host plants. The tubes are constructed from calcium carbonate and other salts secreted by the midgut and an organic matrix secreted by the Malpighian tubules. Adult cercopoids do not produce spittle and are free living. They cannot run, and often use only the front and middle legs to walk, dragging the extended hind legs. Consequently, they rely mostly on their strong jumping and flying abilities for movement.

Species of Cercopoidea are often restricted to particular habitats, but many if not most seem to be capable of utilizing a variety of host plants. Many species seem to prefer actinorrhizal and other nitrogen-fixing hosts, presumably because the xylem sap of such plants contains more amino acids and is more nutritious. Cercopoidea is a predominantly tropical group, occurring mostly in wet and mesic habitats. Nevertheless, the genus *Clastoptera,* has radiated extensively in north temperate North America, and *Aphrophora* comprises numerous arboreal species throughout the Holarctic. Cercopidae are primarily grassland insects, feeding on grasses and other herbs. The family Aphrophoridae includes both grass-feeding and arboreal species. Machaerotidae and Clastopteridae are primarily arboreal.

Members of the superfamily Cercopoidea occur worldwide. Cercopidae and Aphrophoridae are pantropical in distribution, with relatively depauperate faunas in the Holarctic. Machaerotidae are restricted to the Oriental and Australian regions. Clastopteridae are mostly New World animals, but one small genus, possibly misclassified, occurs in the oriental region. Most tribes are restricted to either the New or the Old World, and phyletic diversity seems to be highest in the oriental region. A few genera (e.g., *Philaenus* and *Aphrophora*) are widespread, partly as a result of human activities, but most are restricted to a single biogeographic realm.

Cicadoidea

Cicadoidea (cicadas, Fig. 2) are distinguished from other extant Auchenorrhyncha in having fossorial front legs (in nymphs) and three ocelli grouped in a triangle on the crown of the head; in addition they lack the ability to jump. They are conspicuous insects because of their large size (1.5–11 cm) and the loud courtship calls of the males. Most authorities recognize two families: Cicadidae and Tettigarctidae. Tettigarctidae [Cicadoprosbolidae in the paleontological literature, Fig. 2(6)], which differ from Cicadidae in having the pronotum extended to the scutellum and lacking distinct tympana, are a relict group with two extant species in southern Australia and Tasmania and several fossil taxa dating to the Lower Jurassic. Cicadidae [Fig. 2(7–9)], which do not appear in the fossil record until the Paleocene, comprise two main (possibly polyphyletic) groups, those with the tymbals (sound-producing organs) concealed and those with exposed tymbals. These two groups are sometimes given status as separate families, Cicadidae (*sensu stricto)* and Tibicinidae, respectively. Together these groups comprise approximately 1300 extant species. Phylogenetic analyses of the major lineages are in progress and it is likely that the classification of the superfamily will be substantially revised in the near future.

Although cicadas almost always lay eggs on aboveground parts of their host plant, the nymphs drop to the ground soon after hatching and use modified (fossorial) front legs to burrow into the soil, excavating a subterranean feeding chamber adjacent to a root. They feed on the xylem of the roots of perennial plants, coating themselves and lining their burrows with "anal liquid" that appears to be similar to that produced by cercopoid nymphs. Development in most species requires from 2 to 6 years (13 or 17 years in the periodical cicadas of temperate North America). Larger nymphs of some species inhabiting wet habitats construct towers of mud that facilitate aeration of the burrow. Mature nymphs emerge from the ground and climb onto a vertical surface prior to molting into the adult stage [Fig. 2(9)]. As far as is known, all cicadas feed on xylem sap; hence the frontoclypeus is strongly inflated owing to the presence of strong cibarial dilator muscles. Like the Cercopoidea, cicadas do not walk or run well; instead they rely on flight to move over distances greater than a few centimeters. In some cicada species, males are sedentary, often forming aggregations and calling loudly in choruses to attract females. In others, the male calls are less audible, and males fly frequently from place to place in search of females. Male and female cicadas have auditory organs (tympana) at the base of the abdomen. Unlike other Auchenorrhyncha, female cicadas (except *Tettigarcta*) do not produce acoustic signals. Tettigarctidae differ from other cicadas in producing only substrate-borne signals (in males and females).

Cicadoidea are the most ecologically uniform of the Auchenorrhyncha superfamilies. Nymphs of all species are subterranean root feeders, and adults feed on the aboveground parts of their host plants. Most cicada species tend to be associated with particular habitats, and many seem to be host plant specific. Sympatric species often call at different times of day or mature during different seasons, thus temporally partitioning their habitat. The cicada faunas of deserts and savannas are particularly rich in genera and species, but tropical rain forests also harbor a great diversity of species.

Cicadoidea occur worldwide but, like the other two cicadomorphan superfamilies, are largely a tropical group. A few genera (e.g., *Cicada, Cicadetta*), occur on several continents, but most are restricted to a single biogeographic realm. Most species appear to have fairly narrow geographic ranges. The high degree of endemism in many groups has proven useful in studies of biogeography, particularly in the geologically complex island areas of the oriental and Australian regions.

Membracoidea

Membracoidea (leafhoppers and treehoppers, Fig. 3), by far the most speciose of the auchenorrhynchan superfamilies, are characterized morphologically by the narrow costal space of the forewing, the large, transversely articulated metathoracic coxae, the elongate hind femora, the longitudinal rows of enlarged setae on the hind tibiae, and the presence of scutellar apodemes. The superfamily includes Cicadellidae (leafhoppers), a paraphyletic taxon that apparently gave rise to a lineage comprising the three currently recognized families of treehoppers (Melizoderidae, Aetalionidae, and Membracidae). A fifth family, Myerslopiidae, consists of two genera of small, flightless, litter-dwelling insects found only in New Zealand and Chile and thought to represent a distinct, relatively primitive lineage. Together, these groups comprise nearly 25,000 described species, currently grouped into about 3500 genera.

Membracoidea first appeared in the Jurassic, represented by the extinct family Karajassidae. These early membracoids were leafhopperlike insects with inflated faces (indicative of xylem feeding), and they retained a median ocellus and more primitive wing venation (forewing with CuA_1 free distally), but nevertheless had acquired the rows of enlarged setae on the hind tibia characteristic of modern leafhoppers. The first Cicadellidae appeared in the Lower Cretaceous. Treehoppers (Aetalionidae and Membracidae) make their first appearance in Tertiary age Mexican and Dominican amber.

The largest family, Cicadellidae [Fig. 3(10–13)], is characterized by the presence of four rows of enlarged, spine-like setae on the hind tibia, a peg-and-socket joint between the hind coxae, and the production of brochosomes. Membracidae [Fig. 3(15)], the next largest family, differ from Cicadellidae in having three or fewer rows of enlarged setae on the hind tibia, the male genital capsule with a lateral plate, and the pronotum enlarged, usually extended posteriorly over the scutellum and frequently bearing spines, horns, or other ornamentation. Like Membracidae, Aetalionidae [Fig. 3(14)]. have three or fewer setal rows on the hind tibia but differ in having the front femur fused to the trochanter, in having the scutellum completely exposed, and in having digitiform processes on the female genital capsule. Melizoderidae also resemble Membracidae but differ in having parapsidal clefts on the mesonotum. Myerslopiidae, thought to be the most primitive membracoid family, are bizarre, flightless insects with elytra-like forewings, vestigial ocelli, and a triangular mesocoxal meron resembling that of Cercopoidea. The phylogenetic status and relationships among the major lineages are only beginning to be understood.

Cicadellidae are unique among insects in producing brochosomes, which are minute proteinaceous granules synthesized in a specialized segment of the Malpighian tubules. After each molt, leafhoppers spread brochosomes over external surfaces of the body in an act known as anointing. Rows of modified setae on the legs of leafhoppers are used to distribute the brochosomes during anointing and subsequent acts of grooming. The brochosome coating of nymphal and adult leafhoppers makes the integument extremely hydrophobic and protects leafhoppers from becoming entrapped in drops of water and their own often copious excreta.

Ant mutualism and parental care behavior are widespread among treehoppers [Membracidae and Aetalionidae, Fig. 3(15)]. Females of many species guard their eggs [Fig. 3(14)] and sometimes remain with the nymphs throughout their development. In the treehopper tribes Hoplophorionini and Aconophorini, ant mutualism was lost but parental care was retained. In these groups, females are often able to drive off invertebrate predators by buzzing the wings and/or using the hind legs to kick the intruder off the plant. Acoustic alarm signals produced by the nymphs trigger the mother's defensive response. Female *Aconophora* coat the stem of the host plant on either side of their egg masses with a sticky secretion that traps predators and parasitoids.

Most species of Membracoidea seem to have fairly narrow host and habitat requirements, and this has probably contributed to their remarkable diversity. Particularly notable are the large leafhopper faunas of temperate and tropical grasslands, where they are, by far, the most speciose component of the grass-feeding herbivore fauna. Many leafhopper species in deserts and dry grasslands are flightless or only occasionally produce winged individuals. This trait has presumably reduced gene flow among populations and facilitated speciation in some lineages. In temperate forests of the Northern Hemisphere, the leafhopper subfamily Typhlocybinae has diversified extensively through specialization on individual tree genera and species. In tropical forest canopies, the treehopper family Membracidae and the leafhopper subfamilies Idiocerinae and Typhlocybinae are particularly diverse. In Australia, the endemic fauna has radiated extensively on *Eucalyptus.* The North American treehopper tribe Smiliini has radiated extensively on oak (*Quercus* spp.).

Membracoidea are distributed worldwide. Among the five currently recognized families, Cicadellidae and Membracidae occur on all continents except Antarctica. Aetalionidae have a disjunct neotropical/oriental distribution, Melizoderidae are restricted to South America, and Myerslopiidae occur only in New Zealand and Chile. Most species and genera are restricted to a single continent; many tribes and subfamilies are also restricted to particular continents.

Fulgoroidea

Fulgoroidea (planthoppers, Fig. 4) differ from other Auchenorrhyncha in having the frons occupying most of the facial part of the head and usually with distinct longitudinal carinae, tegulae usually present at the base of the forewings, the second segment (pedicel) of the antenna enlarged and (usually) bearing conspicuous placoid sensilla, the forewing

anal veins confluent basad of the claval margin, and longitudinal carinae usually present on the head, pronotum, scutellum, and legs. Most have two ocelli dorsolaterally on the head, anterad of the compound eyes, but some Cixiidae also have a medial ocellus on the face. Fulgoroidea first appear in the fossil record in the middle Permian, and Cixiidae appear in the Jurassic. Other modern fulgoroid families apparently arose during the Cretaceous or early Tertiary. Twenty families are currently recognized, comprising approximately 1400 genera and 12,000 species. Fulgoroid families are distinguished from each other based mainly on the shape of the head, the spination of the hind tarsi, and the venation of the forewing. Fulgoroidea are the most morphologically variable of all auchenorrhynchan superfamilies, ranging from 1 mm to over 9 cm in length and exhibiting extensive variation in head shape, wing venation, and genital morphology.

Unlike Cicadomorphans, nymphs of Fulgoroidea apparently do not coat themselves with specialized Malpighian tubule secretions. Instead, they produce wax from specialized glands on the abdominal terga and other parts of the body. The wax forms a hydrophobic coating and may conceal some insects from predators. Adult females of many fulgoroid families also produce wax, with which they coat their eggs [Fig. 4(16)]. In certain tropical fulgoroid species, adults of both sexes produce strands of wax up to 75 cm in length. Aggregation behavior with or without ant mutualism has been documented for nymphs and adults in a few fulgoroid families, but egg guarding is known only in Tettigometridae.

In contrast to the ecologically similar Cicadoidea, the Fulgoroidea are the most ecologically diverse superfamily of Auchenorrhyncha. Nymphs of Derbidae and Achilidae live under bark or in litter, feeding on fungi, while nymphs of Cixiidae, Hypochthonellidae, and Kinnaridae are subterranean root feeders. At least four families include cavernicolous (cave-dwelling) species. Ant mutualism has been documented in several fulgoroid families and seems to occur universally among Tettigometridae [Fig. 4(20)], nymphs of which usually inhabit ant nests. Nymphs of most remaining families and nearly all adults feed on the aboveground parts of vascular plants and most seem to be host specialists. Planthopper species usually feed on woody dicotyledonous plants, but most Delphacidae are grass or sedge specialists. Several species of Delphacidae feed on emergent plants in marshes and are capable of walking on the surface of the water. Delphacidae primarily inhabit temperate and tropical grasslands, and diverse faunas of Issidae, Dictyopharidae (Orgeriinae), and Tettigometridae occur in deserts.

Fulgoroidea occur throughout the temperate and tropical regions of the world but are most diverse in the tropics. The Old World tropics harbor the greatest numbers of described families, genera, and species, but the neotropical fauna is less well studied and may be comparable in diversity. The holarctic fauna is rich in Delphacidae and Issidae, but most other families are poorly represented or absent. Tettigometridae, Ricaniidae, Gengidae, Hypochthonellidae, and Meenoplidae are apparently restricted to the Old World. Some genera, particularly in Cixiidae and Delphacidae, are also cosmopolitan in distribution, but most appear to be restricted to a single biogeographic realm.

See Also the Following Articles

Host Seeking, for Plants • Phytophagous Insects • Phytotoxemia • Plant Diseases and Insects • Prosorrhyncha • Sternorrhyncha

Further Reading

DeLong, D. M. (1971). The bionomics of leafhoppers. *Annu. Rev. Entomol.* **16,** 179–210.

Denno, R. T., and Perfect, T. J. (eds.) (1994). "Planthoppers: Their Ecology and Management." Chapman & Hall, New York.

Hamilton, K. G. A. (1982). "The Insects and Arachnids of Canada," Part 10, "The Spittlebugs of Canada, Homoptera: Cercopidae." Agriculture Canada, Ottawa, Ontario.

Marmarosch, K., and Harris, K. F. (eds.). (1979). "Leafhopper Vectors and Plant Disease Agents." Academic Press, New York.

Metcalf, Z. P. (1960). "A Bibliography of the Cercopoidea (Homoptera: Auchenorhyncha [sic])." Waverly Press, Baltimore.

Metcalf, Z. P. (1960–1962). "General Catalogue of the Homoptera," fascicle VII, "Cercopoidea." North Carolina State College, Raleigh.

Metcalf, Z. P., and Wade, V. (1966). "A Catalogue of the Fossil Homoptera (Homoptera: Auchenorhyncha [sic])." Waverly Press, Baltimore.

Moulds, M. (1990). "Australian Cicadas." New South Wales University Press, Kensington, NSW, Australia.

Nault, L. R., and Rodriguez, J. G. (1985). "The Leafhoppers and Planthoppers." Wiley, New York.

Oman, P. W., Knight, W. J., and Nielson, M. W. (1990). "Leafhoppers (Cicadellidae): A Bibliography, Generic Check-list and Index to the World Literature 1956–1985." CAB International Institute of Entomology, Wallingford, Oxon, U.K.

Ossiannilsson, F. (1949). Insect drummers. *Opusc. Entomol. Suppl.* **10,** 1–145.

Shcherbakov, D. E. (1996). Origin and evolution of Auchenorrhyncha as shown by the fossil record. *In* "Studies on Hemiptera Phylogeny" (C. W. Schaefer, ed.). Entomological Society of America, Lanham, MD.

Sogawa, K. (1982). The rice brown planthopper: Feeding–physiology and host–plant interactions. *Annu. Rev. Entomol.* **27,** 49–73.

Wood, T. K. (1993). Diversity in the New World Membracidae. *Annu. Rev. Entomol.* **38,** 409–435.

Autohemorrhage

Many insects voluntarily discharge blood in response to a threat. This behavior, autohemorrhaging, may serve as a physical deterrent (e.g., by enveloping a potential predator or by exposing the predator to noxious substances). Many beetles in the Chrysomelidae, Meloidae, and Lampyridae are well known for this behavior, which is also called reflex bleeding. For example, the blood of blister (meloid) beetles exposes potential predators to the noxious substance cantharidin. Species apparently release blood through an increase in

hydrostatic pressure. When the hydrostatic pressure returns to normal levels, much of this blood is withdrawn into the insect's hemocoel, and so little is actually lost from the insect.

Autotomy

Autotomy is a defensive response to attack involving the amputation or active breaking of a body part along a breakage plane and usually involves loss of a leg. Many invertebrates (e.g., crayfish, daddy-long-legs), including insects such as crickets, grasshoppers, and walkingsticks, and many vertebrates (e.g., salamanders) exhibit this ability. For example, walkingsticks (Phasmida) have weakened areas at the trochanter that break under stress, such as when an appendage is grasped by a predator. If the insect is not an adult, regeneration occurs at the next molt. The amputated leg of the walkingstick twitches after being detached, which may divert the predator's attention away from the attacked insect. A grasshopper, when held by a hind leg, can voluntarily discard that limb by intense muscular contraction and rupture it at the trochanter–femur junction; autotomy can also be induced by mechanical pressure or electrical shock. The individual may benefit from the loss of limb by surviving the potential predator but loss of balance, reduced ability to forage for food, and reduced ability to escape from the next predator result as well.

Sting autotomy, the self-amputation of the stinger and its glands, occurs in many social Hymenoptera as part of colony defense, especially against vertebrates, and may occur because of the size and shape of the sting barbs. Chemical cues released by the detached venom apparatus may enable other attacking individuals to orient themselves to the predator.

See Also the Following Article

Defensive Behavior

Further Reading

Personius, K. E., and Chapman, R. F. (2002). Control of muscle degeneration following autotomy of a hindleg in the grasshopper, *Barytettix humphreysii*. *J. Insect Physiol.* **48,** 91–102.

B

Bee

see *Apis Species*

Beekeeping

Eva Crane
International Bee Research Association

Beekeeping is the establishment and tending of colonies of social bees of any species, an activity from which the beekeeper obtains a harvest or reward. This reward is usually honey, but it may be some other bee product, or bees themselves (e.g., queens, or colonies for pollination). In beekeeping, each colony is usually in a hive, but some beekeeping is done with honey bees that build their nests in the open. Beekeeping is also done with certain nonsocial bees that are reared for pollinating crops.

TECHNIQUES OF MODERN MOVABLE-FRAME HIVE BEEKEEPING WITH *APIS MELLIFERA*

Most of the world's beekeeping is done with *A. mellifera.* In past centuries, these bees were kept primarily for the production of honey and beeswax. Beekeeping is still done mainly to produce honey, but there are also other specialized types of operation. These include the rearing of queens or package bees for other beekeepers who are producing honey. Another type of beekeeping provides colonies of bees to pollinate crops, since in many areas of large-scale agriculture the native pollinators have been destroyed. Since the 1950s specialized beekeeping has also been developed for the production of royal jelly, pollen, and bee venom.

Each type of beekeeping requires the management of colonies to stimulate the bees to do what the beekeeper wants—for instance, to rear more young house bees to produce royal jelly, or more foragers to pollinate crops. During the 1900s, effective methods were developed for the commercial production of substances other than honey: bee brood, bee venom, beeswax, pollen, propolis, and royal jelly.

Honey Production

A colony of honey-storing bees collects nectar from which it makes honey. Nectar is not available continuously, and to store much honey a colony of bees needs many foraging bees (over, say, 10 days old) whenever a nectar "flow" is available within their flight range. Bees may fly 2 km if necessary, but the greater the distance, the more energy they expend in flight, and the more nectar or honey they consume. Thus, it is often cost-effective for the beekeeper to move hives to several nectar flows in turn during the active season.

Figure 1 shows a movable-frame hive with two "deep" boxes. The hive in Fig. 2 also has two deep boxes for brood (i.e., immature bees, eggs, larvae, and pupae), and a shallow box for honey that is less heavy to lift. Any number of honey boxes (also called supers) may be added to a hive, but these are always separated from the brood boxes by a queen excluder, to keep the honey free from brood. Some empty combs in these supers may stimulate honey storage, but supers are not added far in advance of their likely use by the bees.

It is essential that hives and frames have standard dimensions and that an accessory (spacer) be used to ensure that frames are always exactly the correct distance apart.

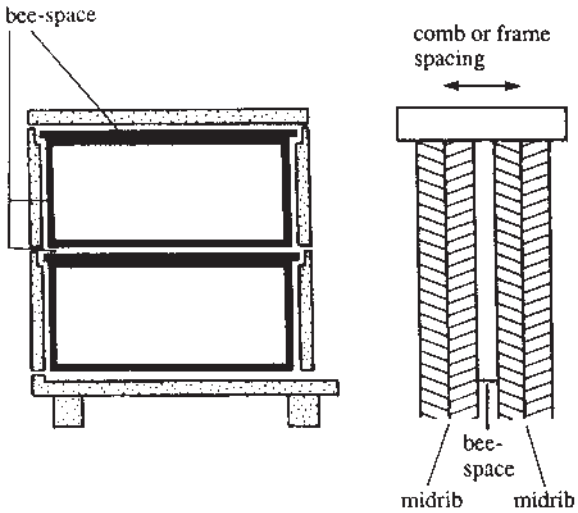

FIGURE 1 Vertical section through a movable-frame hive, showing a brood comb in each box, and the bee spaces. [After Crane, E. (1990). "Bees and Beekeeping: Science, Practice and World Resources." Heinemann, London.]

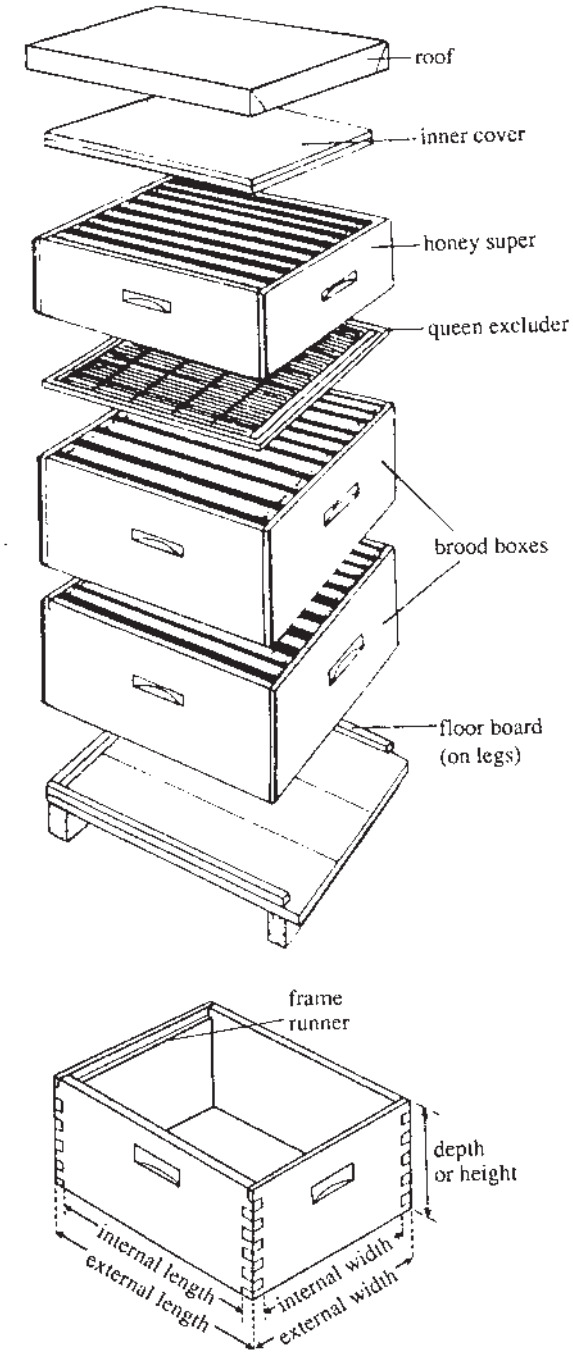

FIGURE 2 Top: exploded view of a movable-frame hive showing the component parts. Bottom: empty hive box showing one of the frame runners. [After Crane, E. (1990). "Bees and Beekeeping: Science, Practice and World Resources." Heinemann, London.]

Queen Production

Large-scale operations are done in five steps, which provide specific conditions for the successive developmental stages of the immature bees that will develop into mated and laying queens.

1. The larvae from which queens will be reared are taken from worker cells of a colony that is headed by a breeder queen selected for chosen genetic characters.
2. Very young larvae are transferred into cell cups mounted mouth down on wooden bars in a "cell-starter colony" that has been queenless for 2 to 4 h. This colony is made up of many (young) nurse bees, and little or no other brood for them to rear. Its bees build the cell cups further and feed the young larvae, and the colony can "start" the rearing of 45 to 90 queen larvae a day.
3. As the larvae grow larger, they receive more food and are better cared for if the number of nurse bees per larva is high. So it is usual to put about 15 cells in each of a number of colonies, where they are separated from the colony's queen by a queen excluder.
4. When the bees have finished feeding the larvae, they seal each immature queen in her cell. The only requirements of an immature queen during the next 7 days are appropriate conditions of temperature and humidity, and these are provided in an incubator. Each queen must emerge from her cell as an adult in a separate cage, for protection from attacks by other queens already emerged.
5. Finally each queen is placed in a "mating hive" containing a few hundred or more workers but no other queen. These hives are taken to a mating apiary, which contains a few strong colonies that include many drones (i.e., males) of a selected strain of honey bees. The apiary is located as far as possible from hives that might contain other drones; a distance of 15 km is likely to be safe, but it varies according to the terrain. When the queen is a few days old, she flies out and mates with drones, and a few days later starts to lay eggs. This shows that she is ready to head a colony.

Package Bee Production

The term "package bees" is used for a number of young worker bees (usually approximately 1 kg) hived with a newly reared and mated queen; these bees together have the potential to develop into a honey-producing colony.

Package bees are normally produced at relatively low latitudes where spring comes early, and are sold at higher latitudes where it is difficult to keep colonies over the winter; many northern beekeepers find it more cost-effective to kill some or all of their colonies when they harvest the season's honey, and to buy package bees next spring. (If they overwinter no colonies, they can follow another occupation for 6 months or more; at least one beekeeper spends the Canadian winter beekeeping in New Zealand, where it is then summer.) The site where the packages are produced should be earlier weatherwise, by 2 months or more, than the site where the bees are used. A package bee industry is most likely to be viable where a single country stretches over a sufficient north–south distance (at least 1000 km, and up to 2000 or even 2500 km). But in New Zealand, package bees are produced at the end of the bees' active season and sent by air to Canada, where the season is just starting.

Package bees are prepared as follows. First, all the bees are shaken off the combs of three or four colonies into a specially designed box, taking care that the queens are left behind. The bees are then poured through the "spout" of the box into package boxes, each standing on a weighing machine, until their weight is either 1 or 1.5 kg, as required. Each box is given a young mated queen in a cage, and a can of syrup with feeding holes. (Enough bees are flying around to return to their hives and keep the colonies functional.) For transport, the package boxes are fixed by battens in groups of three or four, slightly separated; they may travel 2400 km, and the truck needs special ventilation. Air transport, though possible, presents various difficulties.

Crop Pollination

Colonies taken to pollinate crops should be strong, with many foraging bees, and also much unsealed brood (to stimulate the bees to forage for pollen), and space for the queen to lay more eggs. Hives should not be taken to the crop before it comes into bloom, or the bees may start foraging on other plants and continue to do so when the crop flowers. If the hives are in a greenhouse, four to eight frames of bees in each may be sufficient, but the beekeeper must check regularly that the bees have enough food; alternatively, each hive may be provided with two flight entrances, one into the greenhouse and one outside. Beekeepers who hire out hives of bees for crop pollination need to have a sound legal contract with the crop grower; they should also be aware of the risks of their bees being poisoned by insecticides.

In addition to honey bees, certain native bees are especially efficient in pollinating one or more crop species, and several species are managed commercially for pollination. The following are quite widely used for the crops indicated: *Andrena* spp. for sarson and berseem in Egypt and India; *Bombus* spp. for tomato and red clover in Finland and Poland; *Megachile* spp. for alfalfa in Chile, India, South Africa, and the United States; *Nomia melanderi* for alfalfa in the United States; *Osmia* spp. for alfalfa in France; and *Xenoglossa* spp. for apple in Japan, Poland, and Spain, also cotton and curcurbits in the United States.

Special Features of Beekeeping in the Subtropics and Tropics

The subtropics (between 23.5° and 34°N, and 23.5° and 34°S) include some of the most valuable world regions for honey production. Like the temperate zones, they have an annual cycle with a distinct seasonal rhythm and a well-marked summer and winter; however, the climate is warmer and the winters are mild, so the bees can fly year-round. All the major honey-exporting countries include a belt within these subtropical latitudes: China, Mexico, Argentina, and Australia.

Between the Tropics of Cancer and Capricorn (23.5°N and S), the situation is different. The seasons (and honey bee colonies) undergo two cycles in the year because the noonday sun is overhead twice a year. So colonies do not generally grow as large as at higher latitudes, nor do they store as much honey. When forage becomes scarce, a colony may cease brood rearing, then fly as a unit to a nearby area where plants are coming into bloom; this flight is referred to as absconding or migration. So one beekeeper may lose colonies, while beekeepers in the other area put out bait hives to receive the swarms.

Beekeeping in the tropics using traditional hives has been well studied, and many development programs have been carried out to introduce more advanced methods. Francis Smith pioneered successful movable-frame hive beekeeping in tropical Africa.

In the tropics, bee diseases are of less importance than at higher latitudes, but bees in torrid zones may be subject to attack by more enemies, certain birds, mammals, and insects. Tropical honey bees therefore defend their nests more vigorously than temperate-zone honey bees. For instance, tropical African honey bees *(A. mellifera)* are easily alerted to sting and, as a result of rapid pheromone communication between individuals, they may attack *en masse.* People in tropical Africa have grown up with the bees and are accustomed to them. But after 1957, when some escaped following introduction to the South American tropics, they spread into areas where the inhabitants had known only the more gentle European bees, and those from tropical Africa were given the name "killer bees." But once beekeepers in South America had learned how to handle the new bees, they obtained much higher honey yields than from the European bees used earlier.

OTHER ASPECTS OF MODERN HIVE BEEKEEPING

World Spread of *A. mellifera*

In the early 1600s the bees were taken by sailing ship across the Atlantic from England to North America. They would have been in skeps (inverted baskets made of coiled straw), which were then used as hives. The first hives were probably landed in Virginia. The bees flourished and spread by swarming, and other colonies were taken later. By 1800 there were colonies in some 25 of the areas that are now U.S. states, and by 1850 in a further 7. The bees were kept in fixed-comb hives (skeps, logs, boxes).

The bees may possibly have been taken from Spain to Mexico in the late 1500s, but they reached other countries later: e.g., St. Kitts-Nevis in 1720, Canada in 1776, Australia in 1822, and New Zealand and South America in 1839. They were taken later to Hawaii (1857) and Greenland (1950).

In Asian countries where *A. cerana* was used for beekeeping, *A. mellifera* was introduced at the same time as movable-frame hives. Some probable dates of introduction were 1875–1876 in Japan, 1880s in India, 1896 in China, and 1908 in Vietnam.

Between 1850 and 1900 there was widespread activity among beekeepers in testing the suitability of different races of *A. mellifera* for hive beekeeping. The most favored race was Italian *(A. m. ligustica),* named from Liguria on the west coast of Italy, south of Genoa.

Origination and World Spread of Movable-Frame Beekeeping

The production of a movable-frame hive divided the history of hive beekeeping into two distinct phases. This new hive type was invented in 1851 by Reverend Lorenzo Lorraine Langstroth in Philadelphia. He was familiar with the Greek movable-comb hive (discussed later under Traditional Movable-Comb Hive Beekeeping) and with some rectangular hives devised in Europe that contained wooden frames for the bees to build their combs in. These hives, however, had only a very small gap between the frames and the hive walls, and the bees built wax to close it. In 1853 Langstroth described how he had often pondered ways in which he "could get rid of the disagreeable necessity of cutting the attachments of the combs from the walls of the hives." He continued, "The almost self-evident idea of using the same bee-space [as between the centerlines of combs in the frames] in the shallow [honey] chambers came into my mind, and in a moment the suspended movable frames, kept at a suitable distance from each other *and from the case containing them,* came into being" (author's italics). Framed honey combs were harvested from an upper box, and the brood was in the box below. A queen excluder between the boxes prevented the queen from laying eggs in the honey chamber.

The use of hives based on Langstroth's design spread rapidly around the world, dimensions often being somewhat smaller in countries where honey yields were low. Some dates for their first known introduction are 1861, United Kingdom; 1870, Australia; 1878, South Africa; 1880s, India; and 1896, China.

Beekeeping with *A. cerana* in Movable-Frame Hives

Bees of most races of *A. cerana* are smaller than *A. mellifera;* they also build smaller colonies and are less productive for the beekeeper. Unlike *A. mellifera, A. cerana* does not collect or use propolis. *A. cerana* was the only hive bee in Asia until *A. mellifera* was introduced in the late 1800s; it had been kept in traditional hives (logs, boxes, barrels, baskets, pottery) since the first or second century A.D. in China and probably from the 300s B.C. in the upper Indus basin, now in Pakistan.

The movable-frame hives used for *A. cerana* are like a scaled-down version of those for *A. mellifera.* Colony management is similar, except that the beekeeper needs to take steps to minimize absconding by the colonies. In India 30 to 75% of colonies may abscond each year. To prevent this, a colony must always have sufficient stores of both pollen and honey or syrup, and preferably a young queen. Special care is needed to prevent robbing when syrup is fed. Colonies must also be protected against ants and wasps.

The bees at higher latitudes are larger, and in Kashmir (altitude 1500 m, and above) *A. cerana* is almost as large as *A. mellifera* and fairly similar to it in other characteristics; for instance, the colonies do not abscond.

Honey Bee Diseases, Parasites, Predators, and Poisoning

The main brood diseases of *A. mellifera,* with their causative organisms, are American foulbrood (AFB), *Paenibacillus larvae;* European foulbrood (EFB), *Melissococcus pluton;* sacbrood, sacbrood virus (Thai sacbrood virus in *A. cerana*); and chalkbrood, *Ascosphaera apis.* Diseases of adult bees are nosema disease, *Nosema apis;* amoeba disease, *Malpighamoeba mellificae;* and virus diseases. Parasites are tracheal mite, *Acarapis woodi;* varroa mites, *Varroa jacobsoni,* and *V. destructor;* the mite *Tropilaelaps clareae;* bee louse (Diptera); *Braula* spp.; and the small hive beetle, *Aethina tumida.*

Disease or parasitization debilitates the colonies, and diagnosis and treatment require time, skill, and extra expense. Most of the diseases and infestations just listed can be treated if colonies are in movable-frame hives, and in many countries bee disease inspectors provide help and advice. Colonies in fixed-comb hives and wild colonies cannot be inspected in the same way, and they can be a long-term focus of diseases. But by far the most common source of contagion is the transport into an area of bees from elsewhere.

The parasitic *Varroa* mite provides an example. It parasitized *A. cerana* in Asia, where the mite and this bee coexisted. In the Russian Far East, it transferred to

introduced *A. mellifera,* whose developmental period is slightly longer, allowing more mites to be reared. Because colonies could then die from the infestation, the effects were disastrous. In the mid-1900s, some infested *A. mellifera* colonies were transported to Moscow; from there, mites were unwittingly sent with bees to other parts of Europe, and they have now reached most countries in the world.

Since the 1950s it has been increasingly easy to move honey bees (queens with attendant workers, and then packages of bees) from one country or continent to another. One result has been that diseases and parasites of the bees have been transmitted to a great many new areas, and to species or races of honey bee that had little or no resistance to them.

The development of large-scale agriculture has involved the use of insecticides, many of which are toxic to bees and can kill those taken to pollinate crops. In California alone, insecticides killed 82,000 colonies in 1962; in 1973 the number was reduced to 36,000, but in 1981 it had risen again, to 56,000. More attention is now paid to the use of practices that protect the bees, including selecting pesticides less toxic to beneficial insects, using pesticides in the forms least toxic to honey bees (e.g., granular instead of dust), spraying at night when bees are not flying, spraying only when the crop is not in flower, and using systemic insecticides and biological pest control. Possible actions by the beekeeper are less satisfactory: moving hives away from areas to be treated, or confining the bees during spraying by placing a protective cover over each hive and keeping it wet to reduce the temperature.

By 1990, legislation designed to protect bees from pesticide injury had been enacted in 38 countries, and a further 7 had established a code of practice or similar recommendations.

TRADITIONAL FIXED-COMB HIVE BEEKEEPING

A. mellifera in the Middle East, Europe, and Africa

Humans have obtained honey and wax from bees' nests in the Middle East, Europe, and Africa since very early times. Beekeeping with *A. mellifera* was probably initiated in an area when the human population increased so much that it needed more honey or wax than was available at existing nest sites, or when some change occurred that reduced the number of nest sites—for instance, when trees were felled to clear land for agriculture.

In the Middle East, population increase was linked with the development of civilizations. The earliest known hive beekeeping was done in ancient Egypt, and similar traditional beekeeping is still carried out in Egypt. In Abu Ghorab, near Cairo, an Old Kingdom bas-relief from around 2400 B.C. shows a kneeling beekeeper working at one end of hives built into a stack; smoke is used to pacify the bees, and honey is being transferred into large storage pots. Over time, the use of horizontal cylindrical hives spread throughout the Mediterranean region and Middle East, and also to tropical Africa, where hollow log hives were often fixed in trees, out of reach of predators.

In the forests of northern Europe, where honey bees nested in tree cavities, early humans obtained honey and wax from the nests. When trees were felled to clear the land, logs containing nests were stood upright on the ground as hives. As a result, later traditional hives in northern Europe were also set upright. In early types such as a log or skep, a swarm of bees built its nest by attaching parallel beeswax combs to the underside of the hive top. If the base of the hive was open as in a skep, the beekeeper harvested honey from it. Otherwise harvesting was done from the top if there was a removable cover, or through a hole previously cut in the side.

Skeps used in northwestern Europe were made small so that colonies in them swarmed early in the active season; each swarm was housed in another skep, and stored some honey. At the end of the season, bees in some skeps were killed with sulfur smoke and all their honey harvested; bees in the other skeps were overwintered, and their honey was left as food during the winter.

A. cerana in Asia

In eastern Asia the cavity-nesting honey bee was *A. cerana,* and it was kept in logs and boxes of various kinds from A.D. 200 or earlier. But farther west in the upper Indus basin horizontal hives rather similar to those of ancient Greece are used, and it has been suggested that hive beekeeping was started in the 300s B.C. by some of the soldiers of the army of Alexander the Great, who settled there after having invaded the area.

Stingless Bees (Meliponinae) in the Tropics

In the Old World tropics, much more honey could be obtained from honey bees than from stingless bees, and the latter were seldom used for beekeeping. But in the Americas, where there were no honey bees, hive beekeeping was developed especially with the stingless bee, *Melipona beecheii,* a fairly large species well suited for the purpose. It builds a horizontal nest with brood in the center and irregular cells at the extremities, where honey and pollen are stored. The Maya people in the Yucatan peninsula in Mexico still do much beekeeping with this bee. The hive is made from a hollowed wooden log, its ends being closed by a wooden or stone disk. To harvest honey, one of the disks is removed to provide access to honey cells; these are broken off with a blunt object, and a basket is placed underneath the opening to strain the honey into a receptacle below. Many similar stone disks from the 300s B.C. and later were excavated from Yucatan and from the island of Cozumel, suggesting that the practice existed in Mexico at least from that time.

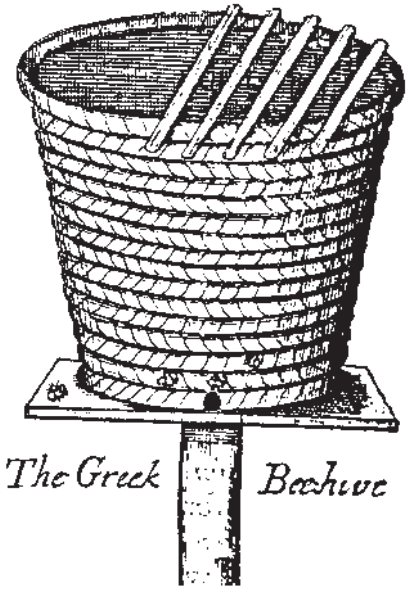

FIGURE 3 Sir George Wheler's drawing of a Greek top-bar hive. [After Wheler, G. (1682). "A Journey into Greece." W. Cademan and others, London.]

Nogueira-Neto in Brazil developed a more rational form of beekeeping with stingless bees. In Australia the native peoples did not do hive beekeeping with stingless bees, but this has recently been started.

TRADITIONAL MOVABLE-COMB HIVE BEEKEEPING

Movable-comb hive beekeeping was a crucial intermediate step between fixed-comb beekeeping, which had been done in many parts of the Old World, and the movable-frame beekeeping used today.

In a book published in 1682 in England, Sir George Wheler recounted his journeys in Greece and provided details of the hives he saw there (Fig. 3). He described the wooden bars shown lying across the top of the hive as "broad, flat sticks" and said that the bees built a comb down from each top-bar, which "may be taken out whole, without the least bruising, and with the greatest ease imaginable." So it was a movable-comb hive. The Greek beekeepers must have placed the bars at the bees' natural spacing of their combs. They made a new colony by putting half the bars and combs from a hive into an empty one; the queen would be in one of the hives, and the bees in the other would rear a new queen.

In the mountain range that separates Vietnam from China, some of the native peoples use a movable-comb hive for *A. cerana;* it is not known how old this method of beekeeping is. The bars are fitted across the top of a log hive at the correct spacing for *A. cerana.* This bee builds small combs without attaching them to the hive sides, and the combs can be lifted out by their bars. There seems to have been no development of a movable-frame hive from this movable-comb hive for *A. cerana.*

TRADITIONAL BEEKEEPING WITHOUT HIVES

A. dorsata

In tropical Asia, a nest of the giant honey bee, *A. dorsata,* which is migratory, can yield much more honey than a hive of *A. cerana.* In a form of beekeeping with *A. dorsata* practiced in a few areas, people use horizontal supports called "rafters" instead of hives. (A "rafter" is a strong pole, secured at a height convenient for the beekeeper by a wooden support, or part of a tree, at each end.) At the appropriate season, beekeepers erect rafters in a known nesting area for migratory swarms of the bees. Sheltered sites with an open space round one end are chosen, which the bees are likely to accept for nesting. After swarms have arrived and built combs from the rafters, the beekeeper harvests honey every few weeks by cutting away part of the comb containing honey but leaving the brood comb intact. When plants in the area no longer produce nectar, brood rearing ceases and the bees migrate to another site.

A. florea

The small honey bee, *A. florea,* builds a single brood comb perhaps 20 cm high, supported from the thin branch of a tree or bush. It constructs deeper cells round the supporting branch and stores honey in them. The whole comb can easily be removed by cutting through the branch at each side, and in some regions combs are then taken to an apiary where the two ends of each branch are supported on a pile of stones or some other structure. This is done, for instance, in the Indus basin near Peshawar in Pakistan, and on the north coast of Oman.

RESOURCES FOR BEEKEEPERS

There are various sources of information and help for beekeepers. Many countries publish one or more beekeeping journals, and have a beekeepers' or apiculturists' association with regional and local branches. Apimondia in Rome, Italy (http://www.apimondia.org) is the international federation of national beekeepers' associations.

In many countries, the ministry of agriculture or a similar body maintains a bee department that inspects colonies for bee diseases and often also provides an advisory service for beekeepers. Research on bees and/or beekeeping may be carried out under this ministry or by other bodies.

The International Bee Research Association in Cardiff, U.K. serves as a world center for scientific information on bees and beekeeping, and publishes international journals, including *Apicultural Abstracts,* which contains summaries of recent publications worldwide. Information about access to the Association's data banks can be obtained from its Web site (http://www.ibra.org.uk), which is linked to Ingenta.

See Also the Following Articles
Apis Species • Honey • Rearing of Insects • Royal Jelly

Further Reading

Connor, L. J., Rinderer, T., Sylvester, H. A., and Wongsiri, S. (1993). "Asian Apiculture." Wicwas Press, Cheshire, CT.
Crane, E., (ed.) (1976). "Apiculture in Tropical Climates." International Bee Research Association, London.
Crane, E. (1978). "Bibliography of Tropical Apiculture." International Bee Research Association, London.
Crane, E. (1990). "Bees and Beekeeping: Science, Practice and World Resources." Heinemann Newnes, Oxford, U.K.
Crane, E. (1999). "The World History of Beekeeping and Honey Hunting." Duckworth, London.
Crane, E., and Walker, P. (1983). "The Impact of Pest Management on Bees and Pollination." Tropical Development and Research Institute, London.
Crane, E., and Walker, P. (1984). "Pollination Directory for World Crops." International Bee Research Association, London.
Delaplane, K. S., and Mayer, D. (2000). "Crop Pollination by Bees." CAB International, Wallingford, U.K.
FAO. (1986). "Tropical and Sub-tropical Apiculture." Food and Agriculture Organisation of the United Nations, Rome.
Graham, J. M. (ed.) (1992). "The Hive and the Honey Bee." rev. ed. Dadant and Sons, Hamilton, IL.
Langstroth, L. L. (1853). "Langstroth on the Hive and the Honey-bee, a Beekeeper's Manual." Hopkins, Bridgeman, Northampton, MA.
McGregor, S. E. (1976). "Insect Pollination of Cultivated Crop Plants." U.S. Department of Agriculture, Washington, DC.
Morse, R. A., and Nowogrodzki, R. (1990). "Honey Bee Pests, Predators and Diseases." 2nd ed. Cornell University Press, Ithaca, NY.
Nogueira-Neto, P. (1997). "Vida e Criação de Abelhas Indígenas sem Ferrão." Edição Nogueirapis, São Paulo, Brazil.
Smith, F. G. (1960). "Beekeeping in the Tropics." Longmans, London.
Webster, T. C., and Delaplane, K. S. (2001). "Mites of the Honey Bee." Dadant & Sons, Hamilton, IL.

Bee Products

Eva Crane
International Bee Research Association

Honey and beeswax are the main bee products used by humans. Bee brood has been eaten by humans since ancient times in some Asian countries, but until the 1900s only honey and beeswax were produced commercially. Then in the 1950s the price of honey on the world market was depressed by surplus production, and beekeepers in certain technologically advanced countries, seeking ways of diversifying the sources of income from their bees, explored the commercialization of royal jelly, bee venom, pollen, and propolis.

BEE VENOM

Bee venom is a secretion from the venom glands of the worker or queen of a species of honey bee *(Apis);* it is not produced by stingless bees (Meliponinae). The main components of commercial freeze-dried venom from *A. mellifera* worker bees include 15 to 17% enzymes, including phospholipase and hyaluronidase; 48 to 58% small proteins, including especially mellitin; 3% physiologically active amines, including histamine; 0.8 to 1.0% amino acids, and numerous minor components. Queen venom differs somewhat from worker venom in its composition and its pattern of change with the age of the bee. A few studies have been made on the venom of other *Apis* species; for instance, toxicity has been reported to be similar in venoms from *A. mellifera* and *A. dorsata,* less in *A. florea* venom but twice as high in *A. cerana* venom.

Bee venom is by far the most pharmacologically active product from honey bees. The general mechanism of its action on humans who are not hypersensitive is as follows. Hyaluronidase breaks down hyaluronic acid polymers that serve as intercellular cement, and the venom spreads through the tissue. (Protective antibodies that develop in the serum of most beekeepers can effectively neutralize hyaluronidase, preventing the spread of the venom.) A protease inhibitor prevents enzymatic destruction of the hyaluronidase. Simultaneously, the mast cell degranulating peptide penetrates the membrane of the mast cells, creating pores. This releases histamine, which (in combination with some small molecules of the venom) contributes to the swelling and flare, and the local itching and burning sensation. As venom penetrates blood vessels and enters the circulatory system, phospholipase A and mellitin (as a micelle, a colloidal-sized aggregate of molecules) act synergistically to rupture blood cells.

When only a few stings are received, the action just described is mostly localized, and actual toxic effects are insignificant. After massive stinging (or injection of venom directly into the circulatory system), the action may become widespread and toxic effects severe, particularly when significant amounts of venom enter the circulatory system. Apamine acts as a poison to the central nervous system, and both mellitin and phospholipase A are highly toxic. Large concentrations of histamine are produced and contribute to overall toxicity. The role of other components is unknown.

Only a very small number of people are allergic (hypersensitive) to insect venom, between 0.35 and 0.40% of the total population in one U.S. survey. In a person allergic to bee venom, the hyaluronidase may participate immediately in an antigen–antibody reaction, triggering an allergic response; both mellitin and phospholipase A can also produce allergic reactions. There may be antigen–antibody reactions to any or all of the components mentioned. Severe reactions can result in death from anaphylactic shock.

Antihistamines can give some protection to a moderately hypersensitive person if taken before exposure to stings. Systemic reactions following a sting should be treated immediately with adrenaline; extremely prompt medical treatment is essential for acute anaphylaxis.

Some allergy clinics provide carefully regulated courses of venom injection, which can decrease sensitivity to the venom; various types of immunotherapy (desensitization) have been used, involving the application of a series of graded doses of pure venom, and these can be effective in 95% of cases. If a

beekeeper or another member of the household develops serious hypersensitivity to bee stings, an allergy specialist may be able to recommend a course of desensitization that will allow the beekeeper to continue.

Germany was probably the first country to produce bee venom commercially. Between 1930 and 1937, girls stationed in front of hives would pick up one worker bee at a time and press it so that it stung into a fabric tissue that absorbed the venom; the venom was extracted from the fabric with a solvent (distilled water), which later was removed by freeze-drying, leaving the venom as a crystalline powder.

A more recent method is to use a bare wire stretched to and fro across a thin membrane mounted on a horizontal frame placed directly in front of a hive entrance. When a low voltage is applied to the ends of the wire, a few "guard" bees are shocked; they sting into the membrane and also release alarm pheromone that quickly alerts other bees to sting into the membrane as well. The bees can withdraw their stings and are unharmed, and the drops of venom released are removed from the underside of the membrane; in hot weather they dry and can be scraped off.

BEE BROOD

Bee brood (immature bees) was probably a useful source of protein to hunter-gatherers in many parts of Asia and Africa, and honey bee larvae have now produced commercially, and marketed either raw or cooked. Mature *A. mellifera* larvae have been found to contain about 60% as much protein as beef and about 30% more fat (fresh weight). Pupae contain somewhat more protein and less fat. Both larvae and pupae contain vitamins A and D. Such bee brood is eaten in parts of Asia (e.g., Korea, China, Japan, Laos, Malaysia, Thailand, and Vietnam) but not in India, Pakistan, or Bangladesh. Some eastern Mediterranean religions forbade the eating of certain insects because these were regarded as unclean. One of the Dead Sea scrolls, from about 200 to 100 B.C., had the prohibition: "Let no man defile his soul with any living being or creeping thing by eating of them, from the larvae of bees [in honey] to all the living things that creep in water." (The digestive system of any animal was considered to be unclean, and it was impractical to remove these organs from individual bees.)

POLLEN

Protein is required by young adult honey bees, and it is an important component of the food they give to larvae. It is obtained from pollen (microspores of seed plants) that older bees collect from flowers and store in the nest. In one study on *A. mellifera* in the United States, bee-collected (air-dried) pollens contained 7 to 30% crude protein and 19 to 41% carbohydrates (mostly sugars from honey that bees mixed with the pollen). Pollen also contains minerals (it has an ash content of 1–6%), vitamins, enzymes, free amino acids, organic acids, flavonoids, and growth regulators.

When a worker honey bee moves past the anthers of flowers, pollen becomes trapped by her body hairs. She leaves the flowers and, with special movements of her legs, passes the pollen backward to bristles on the tibiae of her hind legs. She packs it into a "pollen load" on each of these legs, moistening it with a little nectar or honey in the process. The pollen loads carried by a foraging bee have a variety of colors, which provide clues to the plant sources.

It is relatively easy for a beekeeper to collect the pollen being brought into hives by bees: a pollen trap, fixed over the hive entrance, incorporates a grid (or two grids) through which incoming bees must push, and while they do this most pollen loads are knocked from their hind legs and drop into a tray below, although some bees get through the trap with their pollen loads. The beekeeper needs to ensure that the colony always has enough pollen to rear sufficient brood to maintain its population. (A colony can be made to collect more pollen by giving it extra combs of young brood to rear.) In 1990 pollen was known to be produced commercially in Europe (seven countries), the Americas (five), Asia (four), and Africa (one), and also Australia, where Western Australia alone produced 60 to 130 tonnes a year.

Pollen is used as a dietary supplement for humans and domestic animals, as well as for feeding to a honey bee colony to increase its brood production. Pollen from specific plant species (or cultivars) is also used for fruit pollination, in plant breeding programs, and in the study and treatment of allergic conditions such as hay fever.

PROPOLIS

Propolis is the material that honey bees and some other bees can collect from living plants, which they use alone or with beeswax in the construction and adaptation of their nests. Most of the plant sources are trees and bushes. The material collected may be a wound exudate (resin and latex) or a secretion (lipophilic substances, mucilage, and gum). Propolis thus has a much more varied origin than any other material collected by honey bees. Analyses of various samples (mostly of unknown plant origin) have shown the presence of over 100 compounds, including especially flavonoids.

A bee that collects propolis carries it back to the nest on her hind legs. She goes to a place in the hive where propolis is being used and remains there until her load is taken from her by bees using it. The propolis is mainly collected in the morning and used in the hive in the afternoon.

Where propolis is available, *A. mellifera* uses it for stopping up cracks, restricting the dimensions of its flight entrance, and other minor building works. Observations on both tropical and temperate-zone *A. cerana* indicate that this species does not collect or use propolis, even in a region where *A. mellifera* does, but uses beeswax instead. Propolis is sometimes used by *A. dorsata* to strengthen the attachment of the comb to its supporting branch. It is probably essential to *A. florea* for protecting its nest from ants. These bees build

two rings of sticky propolis round the branch that supports the nest, one at each end of the comb attachment, and may "freshen" the propolis surface so that it remains sticky and ants cannot cross it.

To collect propolis from a hive, the beekeeper inserts a contrivance, such as a flat horizontal grid having slits 2–3 mm wide that will stimulate the bees to close up the gaps with propolis. On removal from the hive, the contrivance is cooled in a freezer. The propolis then becomes brittle, and a sharp blow fractures it off in pieces, which can be stored for up to a year in a plastic bag.

The total commercial world production of propolis may be between 100 and 200 tons a year. China produces more propolis (from hives of introduced *A. mellifera*) than any other country; some South American countries are next in importance. Most importing countries are in Europe.

Propolis has various pharmacological properties, partly from its flavonoid content. It is used in cosmetic and healing creams, throat pastilles, and chewing gum. A few people (in the United Kingdom about one beekeeper in 2000), are allergic (hypersensitive) to propolis, and contact with it leads to dermatitis.

Stingless bees mix much propolis with the wax they secrete before they use it in nest construction; the mixture is called cerumen.

See Also the Following Articles

Honey • Royal Jelly • Venom

Futher Reading

Crane, E. (1990). "Bees and Beekeeping: Science, Practice and World Resources," Chap. 14. Heinemann Newnes, Oxford, U.K.

Crane, E. (1999). "The World History of Beekeeping and Honey Hunting," Chap. 51. Duckworth, London.

Ghisalberti, E. L. (1979). Propolis: A review. *Bee World* **60**, 59–84.

Hocking, B., and Matsumura, F. (1960). Bee brood as food. *Bee World* **41**, 113–120.

Riches, H. R. C. (2000). "Medical Aspects of Beekeeping." HR Books, Northwood, U.K.

Stanley, R. G., and Linskens, H. F. (1974). "Pollen: Biology, Biochemistry, Management." Springer-Verlag, Berlin.

Beeswax

Eva Crane
International Bee Research Association

Beeswax is secreted by workers of most Apidae, which use it to build combs of cells in their nests, for rearing brood, and for storing food. Workers are female members of a colony of bees, active in foraging or nesting, but laying no eggs or only a few compared with a queen. The term "beeswax" is commonly used for the wax from honey bees (*Apis*), especially that from *A. mellifera*, which is the basis of the world's beeswax industry.

PRODUCTION, SECRETION, AND USE OF BEESWAX BY *A. MELLIFERA*

Beeswax is secreted by four pairs of wax glands situated on the anterior part of the worker's last four normal sternal plates (i.e., the ventral portion); the secreted wax hardens into thin scales. In *A. mellifera* workers, the glands increase in secretory activity during the first 9 days or so after the adult bee has emerged from her cell. They usually start to regress at 17 days of age, but may be regenerated later if the colony needs new comb. Honey bees construct their combs of beeswax and also use this substance with propolis to seal small cracks in their nest structure or hive. The requirements of the colony largely determine the amount of wax secreted by its bees. Calculations have shown that an *A. mellifera* worker is likely to have the potential to secrete about half her body weight in wax during her lifetime.

COMPOSITION AND PROPERTIES OF BEESWAX

The major components of *A. mellifera* beeswax include monoesters, diesters, hydrocarbons, and free acids, which together make up more than half the total weight. Over 200 minor components have also been identified. Of the physical properties of beeswax, its thermal properties are of special practical importance, particularly the wide temperature range between its becoming plastic (32°C) and melting (61–66°C). Its relative density at 15 to 25°C is 0.96 and its refractive index at 75°C is 1.44.

Many pesticides used to control mites in the hive can contaminate beeswax.

HARVESTING AND PROCESSING

In the hive, the purest beeswax is that which has recently been secreted: in "cappings" with which cells have been sealed, and in recently built comb. Wax scraped from hive walls or frame bars may be mixed with propolis. Old, dark combs in which brood has been reared are of least value.

When a beekeeper harvests combs of honey from the hives, the honey is first extracted from the combs. Then the wax is melted and the liquid wax separated from any contaminants. On a small scale, clean wax from hives may be melted and strained through cloth, or a "solar wax extractor" may be used, in which the wax pieces are spread out on a sloping metal base in a shallow container with a double glass top, to be melted by radiation from the sun. The liquid wax flows into a container; any contaminants settle at the bottom, and clear wax flows out through an outlet near the top.

In some commercial wax extractors the wax is heated with water, floats to the top, and flows out through an appropriately placed opening. More efficient devices use a steam

press. The percentage of beeswax extracted from the initial material varies according to the source of the wax and the method of extraction.

USES

Beeswax has a very rich history, with a far wider range of uses than any other bee product. In the past, beeswax was especially valued for candles, because it has a higher melting point than many other waxes, and so the candles remain upright in hot weather. Beeswax was also used for modeling and for casting. Some of the world's finest bronze statues and gold ornaments have been made by the lost-wax process, in which a beeswax model is made and encased in mud or plaster that is allowed to dry; the whole is then heated, the molten wax allowed to escape, and molten metal poured in. The metal solidifies in the exact shape of the original beeswax cast, and the casing material is then broken away.

In the batik method of dyeing cloth, and in etching on a glass or metal surface, beeswax can be used as a "resist," applied to certain areas of a surface to protect them from reaction during a subsequent process.

One of the most important current uses of beeswax is in ointments, emollient skin creams, and lotions. It also is still used in polishes and other protective coatings, and as a lubricant in the armament and other industries. Its dielectric properties have led to its use in electrical engineering.

WORLD PRODUCTION AND TRADE

Beekeeping with modern movable-frame hives aims to maximize honey production, and wax production is suppressed by providing the bees with sheets of ready-built wax comb foundation in frames. In experiments in Egypt, wax production in modern hives was only 0.4 to 0.6% of honey production, whereas in traditional hives it was 9 to 11%.

Bees secrete beeswax more readily in hot than in cold climates, and most surplus beeswax is produced in those tropical regions where traditional hives are still used. According to export figures published in 1990, relating to the preceding decade, the three regions producing most beeswax annually were Asia, Central America, and Africa (15.9, 10.5, and 8.7×10^3 tonnes, respectively). Major importing countries (in 1984) were France, German Federal Republic, United States, and Japan.

WAX FROM OTHER BEES

Because the waxes of different species of social bees differ slightly, if *A. mellifera* wax is mixed with that of other bees, its characteristics are altered. Melting points have been reported as follows for wax from other species of honey bees: *A. dorsata,* 60°C; *A. florea,* 63°C; *A. cerana,* 65°C; stingless bees, Meliponinae: *Trigona* spp. (India), 66.5°C; *T. beccarii* (Africa), 64.6°C; *T. denoiti* (Africa), 64.4°C; and bumble bees, *Bombus,* 34–35°C. The temperature in a bumble bee nest is much lower than that in a honey bee nest.

See Also the Following Articles

Beekeeping • Commercial Products from Insects

Further Reading

Coggshall, W. L., and Morse, R. A. (1984). "Beeswax: Production, Harvesting, Processing and Products." Wicwas Press, Ithaca, NY.

Crane, E. (1990). "Bees and Beekeeping: Science, Practice and World Resources," Chap. 13. Heinemann Newnes, Oxford, U.K.

Crane, E. (1999). "The World History of Beekeeping and Honey Hunting." Duckworth, London.

Hepburn, H. R. (1986). "Honeybees and Wax." Springer-Verlag, Berlin.

Michener, C. D. (1974). "The Social Behavior of the Bees: A Comparative Study." Belknap Press, Cambridge, MA.

Beetle

see *Coleoptera*

Biodiversity

Nigel E. Stork

Cooperative Research Centre for Tropical Rainforest Ecology and Management at James Cook University, Australia

Biodiversity is a term created in the mid-1980s to represent the variety of life. It is a contraction of "biological diversity" and came into common usage following the signing in 1992 of the Convention on Biological Diversity at the United Nations Conference on Environment and Development, in Rio de Janeiro. The convention defines biodiversity as "the variability among living organisms from all sources including terrestrial, marine and other aquatic ecosystems and ecological complexes of which they are a part: this includes diversity within species, between species and of ecosystems."

However, biodiversity encompasses not just hierarchies of taxonomic and ecological scale but also other scales such as temporal and geographical scales and scaling in the body size of organisms. Biodiversity represents different things to different people. To those working in museums and herbaria it perhaps represents a new thrust for efforts to describe Earth's fauna and flora. To ecologists it may represent a growing concern about the balance of nature and how well ecosystems can function as biological diversity decreases. To economists and politicians it may represent a new and largely untapped source of needed income for developing nations.

To entomologists biodiversity *is* insects because more than half of all described species on Earth are insects.

Biodiversity is crucial to the planet's survival because as a result of it people have food, construction material, raw material for industry, and medicine, as well as the basis for all improvements to domesticated plants and animals. Biodiversity helps maintain ecosystem functions and evolutionary processes, and stores and cycles nutrients essential for life, such as carbon, nitrogen, and oxygen. Biodiversity absorbs and breaks down pollutants, including organic wastes, pesticides, and heavy metals. It also recharges groundwater, protects catchments, and buffers extreme water conditions.

The "ownership" of biodiversity and who should pay for its conservation are emotive subjects particularly in developing countries. These and other issues that relate to the sustainable utilization of biological and nonbiological resources and the maintenance of well-nurtured populations of humans throughout the world, are extremely complex.

TABLE I Comparison of the Estimated Number of Species for Vertebrates and the Four Most Species-Rich Orders of Insects, Description Rates, and Publication Effort

Order	Described species	Average description rate (species year^{-1})[a]	Publication effort[b]
Invertebrates			
Coleoptera	300,000–400,000	2308	0.01
Lepidoptera	110,000–120,000	642	0.03
Diptera	90,000–150,000	1048	0.04
Hymenoptera	100,000–125,000	1196	0.02
Vertebrates			
Birds	9,000	5	1
Mammals	4,500	26	1.8
Amphibians and reptiles	6,800	105	0.44
Fish	19,000	231	0.37

[a]Average for 1977–1988.
[b]Number of papers per number of species per year.

GENETIC DIVERSITY

The individuals that make up a population are rarely identical. Such variation in the outward appearance of individuals (i.e., in their phenotype) results from the interaction of their individual inherited genetic makeup (genotype) with their surrounding environment. Most natural populations maintain a high level of such genetic diversity. This inherited genetic variation is the basis upon which evolution operates, and without it adaptation and speciation cannot occur. Genetic diversity fundamentally occurs in the form of nucleotide variation within the genome, which originates by mutation (changes in the nucleotide composition of genes, in the position of genes on chromosomes, and in the chromosome complement of individuals) and is maintained both by natural selection and by genetic drift. Other forms of genetic diversity include the amount of DNA per cell and chromosome structure and number. It is estimated that there are 10^9 genes in the world, although some of the genes for key processes vary little across organisms.

The long-term survival and success of a species depends to a large extent upon the genetic diversity within species, which makes possible both a degree of evolutionary flexibility in response to long-term climatic and other environmental change and a dynamic ecological community. The long-term aim of any conservation effort must be to maintain a self-sustaining dynamic ecological community, with the minimum of human intervention. This objective cannot be attained without recognition of the genetic diversity of the member species of the community.

SPECIES DIVERSITY

In spite of immense efforts by 19th- and 20th-century taxonomists to describe the world's fauna and flora, the true dimensions of species diversity remain uncertain. Understanding is hampered by lack of a consensus about the total number of species that have been named and described, with estimates ranging from 1.4 to 1.8 million species. This probably represents less than 20% of all species on Earth, and with only about 20,000 new species of all organisms being described each year, it seems that most species will remain undescribed for many years unless there is a rapid increase in species descriptions (but see http://www.all-species.org).

About 850,000 to 1,000,000 of all described species are insects. Of the 30 or so orders of insects, four dominate in terms of numbers of described species, with an estimated 600,000 to 795,000 species: Coleoptera, Diptera, Hymenoptera, and Lepidoptera (Table I). There are almost as many named species of beetle as there are of all other insects added together, or all other noninsects (plants and animals).

There is no complete catalog of names for all organisms, and for many groups it is often difficult to know what has or has not been named and described. It can sometimes be difficult for taxonomists to determine whether a series of individuals constitutes one or several species, or whether a new individual is the same species as others that have been described. On the other hand, a species may be described more than once. A taxonomist in one part of the world may not realize that a given species has already been described from elsewhere. Some species are so variable that they are described many times. For example, the ladybeetle, *Adalia decempunctata* has more than 40 synonyms. This species has many color morphs, and at various times during the last 200 years different taxonomists have given names to the color morphs without realizing that they were all one species. The level of such synonymy in some groups of organisms may be extremely high: (e.g., 80 and 35% synonymy for Papilionidae and Aphididae, respectively).

The question of how many species in total there are on Earth, including undescribed species, also remains a mystery. In 1833 the British natural historian John Westwood estimated that there might be some 20,000 species of insects worldwide. Today it is recognized that there are about this number of insect species in Britain alone. Estimates for how many species there are on Earth have continued to rise, and still it seems that the answer cannot be provided to within a factor of 100. Groups such as birds, large mammals, and some woody plants are well known, and estimates of their global numbers of species can be made with a fair degree of confidence. However, the scientific rationale for almost all estimates of global numbers of species for the remainder taxa, including insects, is surprisingly thin. Although estimates for global numbers of all species, from bacteria to vertebrates, vary from as low as 2 million to more than 100 million, much evidence seems to support estimates on the lower end of this scale: 5 to 15 million species.

Much of the recent literature on global species estimates has focused on insects and in particular on tropical forest insects. Until the 1980s most entomologists thought that there might be about 2 to 5 million insect species on Earth. However, Terry Erwin of the Smithsonian Institution in 1982 calculated that there are 30 million species of tropical arthropods alone, based on his knockdown insecticide fogging samples of beetles from the canopy of Central American tropical forests. He sampled 1200 species of beetles from the canopy of a single species of tree in Panama and suggested that 13.5% of these (162) must be specific to that tree. He arrived at his total of 30 million by suggesting that (1) all 50,000 species of tropical tree had the same level of insect host specificity, (2) beetles represented 40% of canopy arthropods, and (3) the canopy is twice as rich in arthropods as the ground. Others have since criticized all the steps in Erwin's calculation, suggesting that he overestimated the relative proportion of ground to canopy species, the relative proportion of beetle species to other groups of insects and, perhaps most important of all, the number of species that are host specific to a given species of tree. Another argument Stork and others have proposed is based on well-known insect faunas such as those for Britain and for butterflies. There are some 22,000 insect species in Britain and 67 of these are butterflies. It is also estimated that there are 15,000 to 20,000 species of butterflies in the world. Therefore, if the ratio found in Britain of butterfly species to all other insect species is the same for the whole world, this would indicate that there are 4.9 to 6.6 million species of insects on Earth. These and other analyses indicate that lower estimates (5–10 million insect species worldwide) may be realistic.

One of the reasons so few species have been described is that there are few taxonomists, and most of these are in the developed world. For example, 80% of insect taxonomists are found in North America and Europe. Another critical factor is that most of the type specimens on which species names depend are found in European and (to a lesser extent) North American museums.

It may seem that a great deal is known about the biology, distribution, and threatened or nonthreatened status of insects. In practice, this is far from the truth. For well-known insect faunas, such as those of Britain and other areas of Europe, virtually all species (but, surprisingly, not all) have been described. Even so, distribution maps for these species are often extremely poor, and the data used are often based on records more than 50 years old. For other parts of the world, particularly tropical regions, knowledge of the biota is largely nonexistent. Rarely are there even species lists for some of the better known groups, let alone taxonomic keys and field guides to identify these and other less well-known insects.

Much of the information on the distribution and biology of species is housed in the museums, herbaria, and libraries of developed countries. Some of this information is on index cards. There is now a growing effort to place information associated with specimens in the collection into electronic databases and to make this information readily available. Similarly, the biology and conservation status of the vast majority of insect species remain unknown. For this reason the International Union for the Conservation of Nature's (IUCN) Red Data Books on the threatened status of organisms are mostly limited to groups of large vertebrates and higher plants.

THE EVOLUTION AND EXTINCTION OF BIODIVERSITY

Evolution, simply speaking, is change through time. In genetic terms, evolution is an alteration in the frequency with which different genes are represented in a population, and it results primarily from the processes of natural selection and random drift. Natural selection operates through differential survival and reproductive success of individuals in a population, which determines their contribution to the genetic composition of the next generation. Natural selection acts on individual phenotypes best suited to the environment.

There has been life on Earth for at least 3.5 billion of the 4.6 billion years that the planet has existed. Multicellular plants and animals have evolved in just the last 1.4 billion years. The earliest fossil insect, or insect relative, is a hexapod, the collembollan *Rhyniella praecursor,* from the Lower Devonian (about 380 mya) from Scotland. It is unlikely that insects existed before the Devonian, and there was extensive radiation during the Carboniferous. There are fossils from 300 mya of several nonextinct groups, such as Paleodictyoptera, Meganisoptera, Megasecoptera, and Diaphanopterodea. The only extant orders represented by Carboniferous fossils are Ephemeroptera, Blattodea, and Orthoptera. Orders of modern insects, except Hymenoptera and Lepidoptera, appear to have been established by the Triassic (225 mya), and some of the early groups had disappeared by the late Permian. The massive explosion of insect diversity appears to coincide with that of the flowering plants (angiosperms) in the Cretaceous (135–65 mya).

Numerous studies have shown that there have been periods of rapid evolution of biodiversity and even more dramatic periods of extinction. Four of the five big episodes of extinction in the last 500 million years of the fossil record saw the removal of approximately 65 to 85% of the animal species in the ocean that are preserved as fossils, and the fifth resulted in the loss of 95% or more. In spite of these huge losses, it is now estimated that through subsequent rapid evolution, the present-day diversity of organisms, at both the species level and higher taxonomic levels, is greater than at any other time. Some suggest that present-day diversity may represent roughly 1% of all the species that have ever existed.

There have been many attempts to estimate the life span of species in the fossil record and these range from 0.5 to 13 million, although a few species present today appear to be unchanged in the fossil record for up to 50 million. Some data suggest that the average life span for species is 4 to 5 million.

The extinction of species, just like the evolution of species, is a natural process, and thus the extinction of existing species should occur at the same time as the evolution of new ones. The current list of all plants and animals that are recognized as having become extinct in the last few hundred years is relatively short. In total this amounts to just 600 plant and 491 animal species, and of these only 72 are insects. It is not surprising therefore that the fate of many thousands of threatened species of insects, other invertebrates, and fungi is almost completely overlooked. The death of the last passenger pigeon, "Martha," in 1914 is well known to many conservation biologists, yet the coextinction of two species of lice (*Columbicola extinctus* and *Campanulotes defectus*) that were host specific to this bird went unheralded. Some estimated extinction rates would indicate that most insect species are more likely to become extinct than to be named by taxonomists

Of the 72 species of insects listed on the IUCN's Red Data List as extinct, more than 40 are from Hawaii, and many of the others are from other islands. Proving that a species as small as an insect has become extinct can be very difficult, and indeed one of the largest species of insects that was thought to be extinct, the 15-cm-long Lord Howe Island stick insect (Phasmatodea), was discovered surviving in a remote part of this small island 80 years after its extinction had been declared. Of the insect species that no longer exist, most were driven to extinction by the introduction of other animals such as rats or invasive insects, whereas the demise of most extinct species of birds and mammals resulted from overhunting or loss of habitat.

It seems that there is a genetic or population threshold below which the survival of a species diminishes rapidly. For some species this "minimum viable population" may be 10 individuals and for others, hundreds or thousands. Such species with numbers of individuals below this threshold, the "living dead," although not presently extinct, appear to be doomed to extinction in the near future. A critical factor in the long-term survival of a single species or group of species is the maintenance of the intricate web of interacting species that are important in some way or other for each other's survival. For example, the Brazil nut tree, *Bertholletia excelsa,* relies on euglossine bees for pollination and seed setting, whereas the bees rely on the availability of other resources in the forest to complete their life cycle. Loss of these resources through forest fragmentation or disturbance could lead to the loss of the bees. The Brazil nut tree, however, might survive for many years before becoming extinct. This is just one example from the continuum of cosurvival of species, from those that are entirely dependent on the existence of one other species to those that are only in part dependent on one or a number of species. In this way, the survival or extinction of species or groups of species is linked to the survival of whole habitats or ecosystems.

THE DISTRIBUTION OF BIODIVERSITY

Life-forms of one kind or another are to be found in almost all parts of the surface of Earth, and insects are known to exist in most of these environments except the marine ecosystem. Clearly there is a strong latitudinal gradient in biodiversity, with few species occurring in higher latitudes and most species occurring in the tropics, peaking in tropical rain forests and coral reefs. Freshwater systems occupy a very small part of Earth's surface. Only 2.5% of all water on Earth is nonmarine, and most of this is unavailable to life; 69% of all fresh water exists as ice, principally in the polar regions, and another 30% is present underground. Just 0.3% of Earth's fresh water is freely available in rivers, streams, lakes, and freshwater wetlands, taking up only about 1% of the planet's surface! Although occupying a tiny percentage of Earth's surface, freshwater ecosystems support a rich and varied insect fauna. For some groups, the number of freshwater inhabitants is seemingly out of proportion to the representation of existing freshwater systems.

Of the world's open forest and shrubland, 75 and 42%, respectively, lie within tropical boundaries. At least two-thirds of all plant species are tropical, and thus, 6 to 7% of Earth's surface may contain 50 to 90% of all species of plants and animals. The high species richness of tropical forests is illustrated by La Selva forest of Costa Rica, 13.7 km^2 of which harbors 1500 species of plants, more than the total in the 243,500 km^2 of Great Britain. This Central American area also contains 388 species of birds, 63 of bats, and 42 of fish, as well as 122 reptile species and 143 butterfly species. A single site in southeastern Peru has yielded more than 1200 species of butterflies—almost a quarter of the 5000 species thought to be found in South America.

Two strata in forests are particularly noteworthy, both for their important roles in the functioning of animal and plant communities and for their high insect species richness: the canopy and the soil. The canopy of trees has been called by some the "last biotic frontier" because of the immense diversity of insects, plants, and fungi found there. Forest canopies came to the attention of biologists largely through

the work of entomologists using knockdown insecticides to collect insects from the tops of trees. In 1982 Stork used knockdown insecticides released by a fogging machine hoisted in to the canopy of a 75-m-high rain forest tree in Borneo to collect canopy insects. When the collection had been sorted by taxonomists at the Natural History Museum in London, there were more than 1000 species, and yet the area of collecting sheets on the ground was only 20 m^2. In total, 4000 to 5000 species of insects were collected and sorted in a similar way from just 10 Bornean trees. For one group, the Chalcidoidea wasps, 1455 individuals were collected, but after sorting it was found that this represents 739 species. Because fewer than 100 chalcid species had been recorded before from Borneo, this indicates how little is known about the diversity of insects in some ecosystems.

Elsewhere, 43 species of ants were collected by canopy fogging from a single tropical tree in Peru, a number approximately equal to the ant fauna of the British isles. Tropical forests may cover only a small percentage of Earth's surface, but they are vital for the global cycling of energy, water, and nutrients. Most terrestrial life is found in temperate and tropical forests and grasslands. Some other vegetation types, such as the fynbos of South Africa, are also extremely species rich. This system supports more plant species per square meter than any other place on Earth, with more than 8500 species in total, 68% of which are endemic.

Perhaps less attention has been paid to the diversity of life in soils and associated leaf litter and dead wood. It is probable that there are at least as many species of insects specific to the soil as to the canopy. The diversity of soil organism assemblages and their importance in ecosystem functioning is just beginning to be understood. Relatively obscure groups such as fungi, springtails (Collembola), mites, and nematodes are all rich in species in the soil and are extremely important in ensuring that organic material is broken down and the resulting nutrients made available for the growth of plants. Earthworms in temperate regions and termites in tropical regions are critical for the production, turnover, and enrichment of the soil. They also help to aerate the soil and increase the through flow of water, hence reducing water runoff and soil erosion.

FIGURE 1 Insecticide fogging being released from a knockdown insecticide fogging machine in a hardwood plantation in Cameroon. Note the circular catchment trays suspended above the ground to catch the falling insects released by the insecticide. (Photograph by N. Stork.)

THE ROLE OF SYSTEMATICS IN BIODIVERSITY ASSESSMENT

Systematics is the part of comparative biology that tracks the diversity of organisms with regard to specified relationships among those organisms. It is the branch of biology responsible for recognizing, comparing, classifying, and naming the millions of different sorts of organism that exist. Taxonomy is the theory and practice of describing the diversity of organisms and the arrangement of these organisms into classifications.

Widely accepted as the most basic of natural taxa is the *species.* However, there is still some argument over what exactly a species is. A major problem stems from variation observed among individual organisms, and the species question is largely one of how biologists attempt to classify individual organisms, all of which differ to a greater or lesser extent when compared with one another, into discrete groups or taxa. There is a range of definitions that largely reflects the various theories of the origin of diversity. When biological classification was first developed, organisms were considered each to have a fundamental design and the task of the taxonomist was to discover the essential features of these "types." Even after the publication of Darwin's theory of biological evolution, this concept did not change.

It was only with the emergence of a reliable theory of inheritance, and the development of the disciplines of genetics and population biology, that biologists began to develop rational explanations for the origin of diversity and then apply this knowledge to the species concept. The initial step forward was the recognition of geographical variation, first as "varieties," then as subspecies. This led to the concept of the species as a group of populations that reflected both common ancestry and adaptation to local conditions. In turn, this view was developed into the biological species concept, which defined the species as "groups of interbreeding natural populations that are isolated from other such groups." This species concept is perhaps the most widely accepted today, but it applies only to sexually reproducing species.

After the recognition of species, the next step in taxonomy is to classify the relationship of these species. A number of methods have been developed by which phylogenetic relationships can be estimated. Of these, cladistic analysis is now widely acknowledged as the best. Cladistic analysis rests upon three basic assumptions: features shared by organisms (termed homologies or apomorphies) form a hierarchic pattern; this pattern can be expressed as a branching diagram (cladogram); and each branching point symbolizes the features held in common by all the species arising from that node. Cladograms are the most efficient method for representing information about organisms, hence are the most predictive of unknown properties of those organisms.

Once a cladogram of taxa has been established, the next stage is to formally recognize and name the species and higher taxa. Names are assigned to these taxa according to a system based upon that first developed by the Swedish naturalist Linnaeus in the mid-18th century. Species are grouped into genera, and these in turn are grouped into families, orders, classes, phyla, and kingdoms. The ultimate goal of this nomenclature is to produce a universal system of unambiguous names for all recognized taxa. Animals, plants, and bacteria each have a separate set of rules or codes, which are applied voluntarily by taxonomists and are designed to promote stability and consistency in taxonomic nomenclature, and thus to biological science in general. Traditionally, life-form have been grouped into two kingdoms, Animalia (including the insects) and Plantae, but in the last few decades this view has been questioned by experts, with other kingdoms being recognized. Recent work using analyses of ribosomal RNA sequences has shown that the total genetic diversity of the traditionally well-known groups such as fungi, plants, and animals is only a tiny proportion of the genetic diversity shown by microorganisms. The term "domain" has now replaced "kingdom," with the higher organisms (fungi, plants, and animals) being grouped in the domain Eukarya and a variety of microorganisms being included in two further domains, Archaea and Bacteria.

THE ECOLOGICAL CONTEXT

Ecology is the study of the relations between organisms and the totality of the physical and biological factors affecting them or influenced by them, or more simply, as the study of patterns in nature. Ecologists investigate the biology of organisms, looking for consistent patterns in their behavior, structure, and organization. Although a relatively new field in comparison to systematics, ecology has already provided considerable insights into the organization of taxa.

Ecosystem function refers to the sum total of processes operating at the ecosystem level, such as the cycling of matter, energy, and nutrients. The species in a community influence its productivity, nutrient cycling, and fluxes of carbon, water, and energy. Ultimately, species may be responsible for such factors as the maintenance of atmospheric composition, the dispersal and breakdown of waste material, the amelioration of weather patterns, the hydrological cycle, the development of fertile soils, and even the protection of many coastal areas.

Biogeochemical cycling is the movement of materials including carbon, nitrogen, phosphorus, and calcium through an ecosystem as individuals of different trophic levels are consumed by others at higher trophic levels. These nutrients are returned eventually to the abiotic "nutrient pool," where they are again available to primary producers.

Some of the important roles played by different species in biochemical cycling can be outlined briefly.

By their photosynthetic activity, plants play a fundamental role in the carbon cycle, introducing carbon into the food web. Microorganisms are also crucial. It is estimated that algae and cyanobacteria are responsible for 40% of the carbon fixed by photosynthesis on Earth. At the other end of the process, wood-decaying fungi release approximately 85 billion metric tons of carbon into the atmosphere each year as carbon dioxide. Termites also play an important role in global carbon cycling (hence, potentially, in global climate change) through their production of methane. Earth's nitrogen cycle is dependent on bacteria for nitrogen fixation and the release of nitrogen by denitrification. The microbial community thus controls the amount of nitrogen available to an ecosystem, determining ecosystem productivity in areas where nitrogen is limiting. By absorbing water from soils or other surrounding media, plants have a fundamental effect on the water cycle.

There is an ongoing debate between those who believe that all species in a given ecosystem are important and those who say that some are "functionally redundant." That is, if a species is removed from an ecosystem, can other species fulfill the same role? Two factors influence the importance of a species in ecosystem functioning: the number of ecologically similar species in the community and the extent to which a species has qualitative or quantitative effects on the ecosystem.

SPECIES INTRODUCTIONS

The introduction of exotic species has been responsible for great perturbations in many ecosystems. The arrival of predators, competitors, pests, and pathogens has caused decreases in populations of native species in many areas. Native or endemic species often occupy narrow ranges, have small population sizes, and lack defenses, all of which make them vulnerable to species introductions. The arrival of alien species is generally a more serious problem on islands, especially remote islands than in continental areas. For example, dramatic changes have occurred on the Hawaiian islands as a result of species introductions since human colonization. Although introductions may increase local diversity, most colonizers are cosmopolitan and are not endangered, whereas many endemic species are potentially threatened. Ultimately, many local ecosystem types may be lost worldwide, leading to a more homogenous global biota.

See Also the Following Articles

Conservation • Endangered Insects • Genetic Variation • Introduced Insects • Island Biogeography and Evolution

Further Reading

Ehrlich, P. R., and Ehrlich, A. (1981). "Extinction. The Causes of the Disappearance of Species." Random House, New York.

Groombridge, B., and Jenkins, M. D. (2000). "Global Biodiversity: Earth's Living Resources in the 21st Century." World Conservation Press, Cambridge, U.K.

Gullan, P. J., and Cranston, P. S. (1994). The insects. *In* "An Outline of Entomology." Chapman & Hall, London.

Hillis, D. M., and Moritz, C. (1990). "Molecular Systematics." Sinauer, Sunderland, MA.

Lawton, J. H., and May, R. M. (eds.). (1995). "Extinction Rates." Oxford University Press, Oxford, U.K.

McNeely, J. A. (1988). "The Economics of Biological Diversity." International Union for Conservation of Nature, Gland, Switzerland.

Mawdsley, N. A., and Stork, N. E. (1995). Species extinctions in insects: Ecological and biogeographical considerations. *In* "Insects in the Changing Environment." (R. Harrington and N. E. Stork, eds.), pp. 321–369. Academic Press, London.

May, R. M. (1992). How many species inhabit the earth? *Sci. Am.* October, 18–24.

Stork, N. E. (1988). Insect diversity: Facts, fiction and speculation. *Biol. J. Linn. Soc.* **35,** 321–337.

Stork, N. E. (1995). The magnitude of global biodiversity and its decline. *In* "The Living Planet in Crisis: Biodiversity Science and Policy" (J. Cracraft and F. T. Grifo, eds.), pp. 3–32. Columbia University Press, New York.

United Nations Environmental Programme. (1995). "Global Biodiversity Assessment." Cambridge University Press, Cambridge, U.K.

World Conservation Monitoring Centre. (1992). "Global Biodiversity. Status of the Earth's Living Resources." Chapman & Hall, London.

Wilson, E. O. (1993). "The Diversity of Life." Allen Lane, Penguin Press, London.

Wilson, E. O., and Peter, F. M. (eds.). (1988). "Biodiversity," pp. 3–18. National Academy Press, Washington, DC.

Biogenic Amines

Arnd Baumann
Forschungszentrum Jülich, Germany

Wolfgang Blenau
Universität Potsdam, Germany

Joachim Erber
Technische Universität Berlin, Germany

Biogenic amines are important messenger substances and regulators of cell functions. In insects, these small organic compounds act as neurotransmitters, neuromodulators, and neurohormones. Biogenic amines control endocrine and exocrine secretion, the contraction properties of muscles, the activity of neurons, and the generation of motor patterns. In addition, certain biogenic amines are involved in learning and the formation of memory. Biogenic amines mediate these diverse cellular and physiological effects by binding to specific membrane proteins that primarily belong to the superfamily of G-protein-coupled receptors.

Specialized Terms

Ca^{2+} signaling Change in the intracellular Ca^{2+} concentration, through the release of Ca^{2+} ions from intracellular stores by the opening of ligand-gated ion channels or the entry of Ca^{2+} ions into the cell through different types of Ca^{2+}-selective channels located in the plasma membrane, that plays a role in the regulation of various cellular processes, including cell metabolism, gene expression, cytoskeletal dynamics, and neurotransmission.

cyclic AMP Cyclic derivative of adenosine monophosphate that is synthesized from ATP by adenylyl cyclase. Intracellular second messenger involved in the regulation or modulation of ion channels, protein kinase activity, and gene expression.

G-protein-coupled receptors Integral membrane proteins that constitute a large family of neurotransmitter, hormone, or olfactory receptors. Characterized by seven transmembrane regions. When agonists bind to these receptors trimeric GTP-binding (G) proteins are activated that then regulate the activity of intracellular secondary effectors, which change intracellular concentrations of second messengers or ion channel activity.

neurohormone Small organic or peptidergic substance that is produced in neurosecretory cells. Released into the hemolymph at special regions called neurohemal organs. Transported to target tissues with the hemolymph.

neuromodulator Neuroactive substance that is released by synaptic terminals. Simultaneously acts on large numbers of cells in the proximity of the releasing cell and modifies the properties of synaptic transmission and the properties of target cells.

neurotransmitter Chemical substance that is released from the presynaptic endings of a neuron. Transmits information across the synaptic cleft to specific receptors located on the surface of postsynaptic cells.

phosphorylation Transient, reversible posttranslational modification of proteins in which the terminal phosphate group of ATP is transferred to specific residues of a polypeptide by kinases and often alters the properties of the protein.

second messenger Intracellular substance, such as Ca^{2+}, cyclic AMP, inositol-1,4,5-trisphosphate, that modifies or modulates cellular responses. Concentration changes in response to activation of G-protein-coupled receptors.

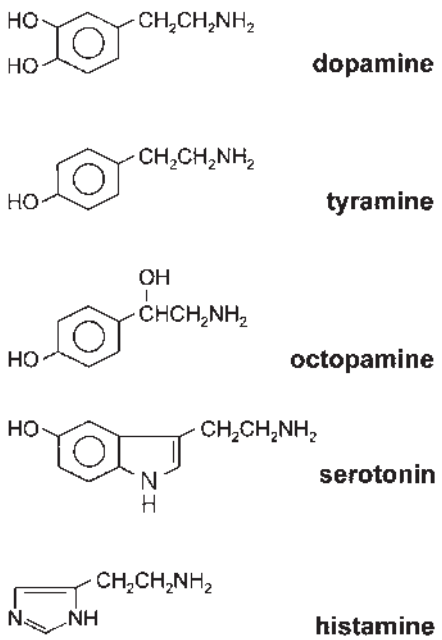

FIGURE 1 Biogenic amines of invertebrates. In insects five substances have been identified as biogenic amines: dopamine, tyramine, octopamine, serotonin, and histamine.

BIOSYNTHESIS OF BIOGENIC AMINES

Biogenic amines are involved in a variety of regulatory functions. Five primary amines are considered biogenic amines in invertebrates: histamine (HA), serotonin (5-HT), dopamine (DA), tyramine (TA), and octopamine (OA) (see Fig. 1). These small organic compounds are synthesized from three different amino acids by single to multistep enzymatic reactions.

LOCALIZATION OF BIOGENIC AMINES

Aminergic systems in insects and vertebrates are quite different. In insects, OA and TA are present in relatively high concentrations, whereas they appear to have only minor significance in vertebrates. In contrast, the catecholamines norepinephrine and epinephrine are important chemical messengers in vertebrates, whereas in the insect nervous system they are detected only in very low concentrations if at all. Several additional catecholamines are involved in the process of cuticle tanning, hardening, and sclerotization in insects. These catecholamines are cross-linking reagents for cuticle proteins and chitin.

Considerable physiological, biochemical, and histochemical evidence suggests that HA, 5-HT, DA, OA, and TA act as transmitters or modulators in the central and peripheral nervous systems of insects. Antisera to HA, 5-HT, DA, and OA often label interneurons that have wide branching patterns within the central nervous system, sometimes innervating neuropils bilaterally. The dorsal and ventral unpaired median neurons, which can contain OA, are well-known examples of such large-field cells. Amine-containing neurons with large arborizations are well suited to act on large groups of other neurons simultaneously. In addition to these large-field cells there are small-field aminergic neurons, especially in the central complex and in the optic lobes. The neuroanatomy of these cells suggests that they communicate with a limited number of target cells.

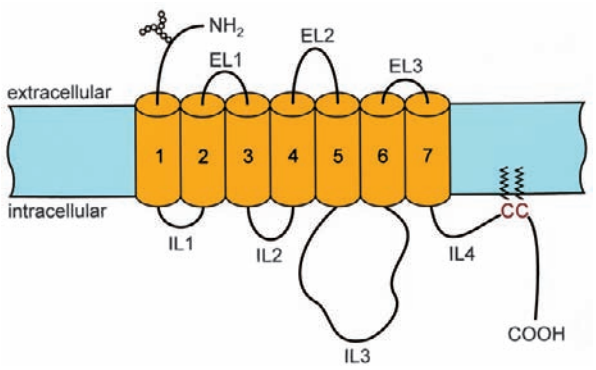

FIGURE 2 Transmembrane topography of G-protein-coupled receptors. The polypeptide spans the membrane seven times. The transmembrane regions (TM 1–7) are depicted as cylinders. The N-terminus (NH_2) is located extracellularly and often contains glycosylated residues (o). The C-terminus (COOH) is located intracellularly. The membrane-spanning regions are linked by three extracellular loops (EL1–EL3) that alternate with three intracellular loops (IL1–IL3). Posttranslational palmitoylation of cysteine residues (C) in the cytoplasmic tail creates a fourth intracellular loop (IL4).

BIOGENIC AMINE RECEPTORS

Biogenic amines bind to specific integral membrane receptors belonging predominantly to the superfamily of G-protein-coupled receptors. Physicochemical, biochemical, and immunochemical investigations show that these polypeptides share the common motif of seven transmembrane (TM) segments (Fig. 2). The N-terminus is located extracellularly, whereas the C-terminus is located intracellularly. The N-terminus is the target of a common posttranslational modification. In this part of the polypeptide consensus sequence motifs are often glycosylated. The membrane-spanning regions are linked by three extracellular loops (EL) that alternate with three intracellular loops (IL). Cysteine residues in the C-terminus of the polypeptides are the target of posttranslational palmitoylation. This modification creates a fourth intracellular loop.

A receptor is activated after binding of the specific biogenic amine in a binding pocket formed by the TM regions in the plane of the membrane. Individual residues in TM3, TM5, and TM6 were shown to participate in ligand binding. Once the ligand is bound, the receptor changes its conformation. This structural alteration usually is registered by intracellular trimeric GTP-binding proteins (G proteins). Residues that reside in close proximity to the plasma membrane in IL2, IL3, and IL4 of the receptor proteins determine the specificity and efficacy of the interaction between receptor and G protein.

GENERAL FUNCTIONS OF BIOGENIC AMINES

Biogenic amines have diverse functions controlling all phases of the life cycle of an insect. They are important chemical messengers during embryonic and larval development and they participate in the synaptic organization of the brain in the adult. As neuroactive substances they act on sensory receptors, inter- and motoneurons, and muscles and other peripheral organs (fat body, firefly lantern, salivary glands, corpora allata and corpora cardiaca, oviduct, etc.). Biogenic amines can initiate or modulate different types of behavior and they are involved in learning and the formation of memory in insects.

The effects of biogenic amines in the insect central nervous system are studied with the techniques of electrophysiological recordings, primary cell cultures, microinjections of amines and receptor ligands, and behavioral assays. Often the physiological responses to biogenic amines last for many minutes, which suggests that they can also act as neuromodulators. Biogenic amines modulate neuronal activity and the efficacy of synaptic transmission in all parts of the nervous system. The huge projection fields of many aminergic neurons support the idea of parallel modulation of entire neuronal circuits by just a few aminergic cells. In addition to synaptic neurotransmission, some aminergic neurons release the amine into the hemolymph. The substances are transported throughout the body and may thus have hormonal functions in specific target tissues.

The physiological role of OA at different levels of the organism is well documented. As a stress hormone in the periphery and in the central nervous system OA prepares the animal for energy-demanding behaviors. This monoamine stimulates glycogenolysis, modifies muscle contraction, supports long-term flight, and regulates "arousal" in the central nervous system. OA and OA agonists can enhance behavioral responses, like escape or aggressive behavior in crickets and sucrose responsiveness in honey bees. Injection of OA can elicit flight motor behavior in locusts, even in isolated thoracic ganglia. It is assumed that in insects OA has functions similar to those of the adrenergic system in vertebrates.

Both OA and 5-HT can modulate sensory receptors and receptor organs in insects. In many cases the sensitivities of the receptors are enhanced. Different funcions of OA and 5-HT at the sensory periphery are not very well understood, because the two amines often differ only in the degree of modulation. The increased sensitivity of sensory receptors due to the action of OA can modify behavior and is part of the "fight or flight" function. Studies on the *Drosophila* tyramine receptor mutant *hono* suggest that TA can also modulate the sensitivity of olfactory receptor cells, thus modulating behavioral responses to olfactory repellents.

The modulation of interneurons or effector neurons by biogenic amines is another level of modifying signal processing. OA and 5-HT can have functional antagonistic effects in a number of different systems. In these systems OA usually enhances the sensitivity or activity of single neurons and 5-HT usually has the opposite action. These effects, which can be measured at both the behavioral and the single-cell level, are dependent on the state of the insect. OA can induce a state of "arousal" in inactive animals and has only minor effects on very active animals, whereas 5-HT shows the largest effects in active animals.

In addition to modulatory functions during the adult life of an insect, DA and 5-HT have important functions during development. In *Drosophila*, high DA concentrations coincide with larval and pupal molts. Reduced levels of DA during larval stages lead to developmental retardation and decreased fertility in adults. 5-HT similarly acts as a chemical signal during larval development in *Drosophila*. Impaired 5-HT synthesis can lead to abnormal gastrulation movements, cuticular defects, and even embryonic death.

The neurotransmitter HA is released from photoreceptors in the compound eyes and ocelli in response to illumination. HA has also been detected in mechanosensory cells in *Drosophila*.

FUNCTIONS IN LEARNING AND MEMORY

Biogenic amines are involved in different forms of learning and memory formation in *Drosophila* and honey bees. However, it has not been unequivocally proven that the same biogenic amines serve identical functions in both species. Research on the neuronal and molecular bases of learning and memory over the past two decades in insects has focused on the mushroom bodies and antennal lobes of the brain. These two structures are involved primarily in processing of olfactory stimuli. Experimental evidence suggests that DA signals the presence of reinforcers and modulates intrinsic mushroom body neurons during conditioning in *Drosophila*. Thus DA could trigger signaling cascades that affect the storage of information about the conditioned stimulus.

In the honey bee, OA appears to be the modulatory transmitter which conveys information about rewarding sucrose stimuli and induces medium- to long-term modifications in interneurons during associative olfactory learning. Electrical stimulation of an identified octopaminergic cell, the ventral unpaired median VUM_{mx1} neuron, can substitute for the sucrose reward during olfactory conditioning. This neuron has extensive arborizations in different brain regions, including the antennal lobes and the mushroom bodies. Microinjections of OA into these two neuropiles of the bee brain confirmed that OA in fact induces associative learning.

See Also the Following Articles

Brain and Optic Lobes • Chemoreception • Learning

Further Reading

Blenau, W., and Baumann, A. (2001). Molecular and pharmacological properties of insect biogenic amine receptors: Lessons from *Drosophila melanogaster* and *Apis mellifera*. *Arch. Insect. Biochem. Physiol.* **48**, 13–38.

Davis, R. L. (1996). Physiology and biochemistry of *Drosophila* learning mutants. *Physiol. Rev.* **76**, 299–317.
Erber, J., Kloppenburg, P., and Scheidler, A. (1993). Neuromodulation by serotonin and octopamine in the honeybee: Behavior, neuroanatomy and electrophysiology. *Experientia* **49**, 1073–1083.
Hammer, M. (1997). The neural basis of associative reward learning in honeybees. *Trends Neurosci*. **20**, 245–252.
Homberg, U. (1994). Distribution of neurotransmitters in the insect brain. In "Progress in Zoology" (W. Rathmayer, ed.), Vol. 40, VCH, Stuttgart.
Kutsukake, M., Komatsu, A., Yamamoto, D., and Ishiwa-Chigusa, S. (2000). A tyramine receptor gene mutation causes a defective olfactory behavior in *Drosophila melanogaster. Gene* **245**, 31–42.
Monastirioti, M. (1999). Biogenic amine systems in the fruit fly *Drosophila melanogaster. Microsc. Res. Tech.* **45**, 106–121.
Nässel, D. R. (1999). Histamine in the brain of insects: A review. *Microsc. Res. Tech.* **44**, 121–136.
Osborne, R. H. (1996). Insect neurotransmission: Neurotransmitters and their receptors. *Pharmacol. Ther.* **69**, 117–142.
Roeder, T. (1999). Octopamine in invertebrates. *Prog. Neurobiol.* **59**, 533–561.
Stevenson, P. A., and Spörhase-Eichmann, U. (1995). Localization of octopaminergic neurons in insects. *Comp. Biochem. Physiol.* **110A**, 203–215.
Vanden Broeck, J. J. M. (1996). G-protein-coupled receptors in insect cells. *Int. Rev. Cytol.* **164**, 189–268.
Wright, T. R. F. (1987). The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster. Adv. Genet.* **24**, 127–222.

Biogeographical Patterns

Peter Zwick
Max-Planck-Institut für Limnologie

Biogeography, which deals with the description and interpretation of plant and animal distributions, is linked with other sciences, especially ecology and (paleo-) geography; zoogeography is the branch addressing animal distribution.

Most animal species inhabit restricted ranges, and only relatively few are cosmopolitan. A comparison of the areas inhabited by different species reveals common distributional patterns that are complex reflections of the ecology and of active and passive animal dispersal, but also of the evolutionary history of both the species and the earth's surface. Zoogeography was sometimes divided into different disciplines, descriptive as opposed to causal zoogeography; the latter was then subdivided into ecological and historical zoogeography. Although studies may differ in their emphasis, the interrelations among these disciplines are too close for a formal division. This article describes the major zoogeographical patterns and uses selected examples from among the insects to highlight the significance of some of the factors just mentioned.

Insects are of great geological age, and most orders existed and were diverse when familiar vertebrates were only beginning to appear. Therefore, the distribution of most insect orders dates back much further than the distributions of many birds and mammals.

Insects are generally absent from some habitats. For example, with the exception of a few littoral specialists for unknown reasons, the only insects in the sea are some high ocean surface skaters among the bugs. Therefore, marine distribution patterns need not be considered here. The salt content of seawater is not the cause of this absence; insects are well represented in epicontinental waters of all kinds: fresh, brackish, and even hypersaline. Aquatic insects played an important role in the development of modern insect zoogeography. Because of their specific habitat ties, aquatic insects are easily collected, and the distributions of many are exceptionally well documented. Their distributions resemble those of terrestrial insects, in part because most aquatic insects have terrestrial adults that disperse over land.

The early explorers were struck by overall differences between the faunas of the lands they visited. The recognition of distinct faunal regions on a global scale thus has a long tradition and is briefly presented as an introduction. In addition to landmass topography, ecological conditions provide the basic setting for animal distributions; a brief outline of the major bioregions with similar overall ecology is therefore also presented.

Reproductively isolated species are the only naturally defined animal taxa; subspecific taxa can interbreed, whereas supraspecific taxa such as genera or families are human abstractions that change with conventions. It is convenient to use extant species to explain some concepts related to ranges and to discuss insect dispersal. Next, distribution patterns shaped by Pleistocene events are used to illustrate the importance of ecological change. The final focus is on disjunct (discontinuous, divided) distributions of monophyletic taxa that can best be explained by much older events, particularly continental drift.

ZOOGEOGRAPHICAL REGIONS

The major faunal regions (or realms) of the world only partly coincide with major landmasses (Fig. 1). Each region has a characteristic fauna distinguished by the particular combination of endemic taxa that exist in only this one region and those occurring also elsewhere. This early descriptive approach has long dominated zoogeography.

The Holarctic region is the largest region and is composed of the Palearctic and Nearctic regions, with many animals distributed over all the entire Holarctic region. Although a narrow land bridge (i.e., Central America) connects the Nearctic with the Neotropical region, the faunal change is pronounced. This land bridge is recent and was available only intermittently in the past. The Sahara Desert separates the Palearctic region from the Ethiopian (or Afrotropical) region, which includes the Arabian peninsula; Madagascar is now recognized as a distinct subregion. In Southeast Asia, climatic

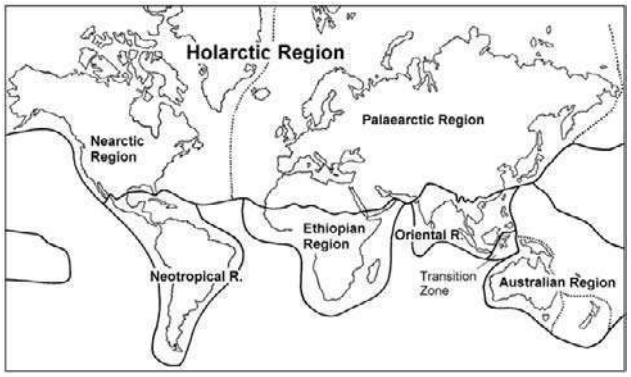

FIGURE 1 Terrestrial zoogeographical regions. [Modified from DeLattin, G. (1967). "Grundriss der Zoogeographie." Gustav Fischer Verlag © Spektrum Akademischer Verlag, Heidelberg.]

and other ecological differences cause a rather abrupt change of the biota south of the Himalayas, where the Palearctic and Oriental regions meet. In the southeast, the Oriental region is in contact with the Australian region, which includes New Guinea, New Zealand, New Caledonia, and the Oceanic subregion. The Australian region is most distinct, but the change toward the Oriental region is nevertheless not abrupt. Depending on the animal group studied, different variants of a border line (named Wallace's line after Alfred Russel Wallace, the earliest observer) were proposed in the past. The Oriental–Australian transition zone is sometimes called Wallacea.

There are puzzling resemblances among the faunas of different zoogeographical regions that are not in physical contact, and related animals may live on widely separate continents. Examples are provided by similarities between the faunas of eastern South America and West Central Africa, between Madagascar and India, or between Andean South America and the Australian region. Also, the fauna of eastern North America has resemblances to the European fauna, and the Far East Asian fauna to that of northern and western North America, despite the intervening oceans. On the other hand, insects in western North America are more distinct from those in the east, and those in Europe differ more from those in Asia, than one would expect in view of the continuous landmasses. These inconsistencies cannot be explained from present geography or ecology but reflect histories of ancient landmasses.

BIOREGIONS OR BIOMES

Seashores, glaciers, high mountains, and deserts pose obvious physical limits to animal distribution. Even in the absence of physical barriers, however, most species inhabit only part of a major landmass, because of ecological constraints. It is rare that a single ecological factor, or a precise combination of factors, limits an insect's distribution. However, most ranges can readily be assigned to a particular biome or bioregion, that is, a large landscape with characteristic overall ecological conditions. Biomes can conveniently be described by general landscape physiognomy, mainly by reference to plant cover, which, among other things, determines the microclimate the insects experience. Biomes do not coincide with zoogeographic regions and each biome comprises separate areas on different continents. A particular biome may harbor animals that look similar or behave similarly but are not necessarily closely related. Instead, they may be characteristic life-forms exhibiting certain traits evolved independently, in response to similar ecological conditions; desert beetles offer examples. Authors differentiate and subdivide biomes to different degrees; some clearly distinct and almost universally recognized biomes are briefly discussed here (Fig. 2).

Arboreal Bioregions

The arboreal biome includes the areas supporting forests, as opposed to only individual trees. Patches of meadows, rocky outcrops, or swamps may occur because of local ecological conditions, and although they are mostly treeless, they still form part of the arboreal biome. Temperature and humidity mainly determine the type of forest occuring in an area. Only the large zonal types are briefly characterized; most are more or less disjunct today.

HYLAEA A name orginally proposed for the Amazonian rain forest, Hylaea is now widely used to designate all tropical evergreen rain forests—dense, multi-storied forests, with little light reaching the forest floor. Animals and plants are adapted to favorable conditions such as temperatures, precipitation, and air humidity that are continuously high. Biodiversity is very high, probably partly because of the presumedly continuous existence of tropical rain forests over exceptionally long periods of time. Processing of shed plant material is fast, and little detritus accumulates on the forest floor. Recent studies using fumigation techniques have shown that most insects inhabit the tree crowns. From an amazingly large number of undescribed species discovered by this method, the total number of existing insect species would be 35 million; more broadly based estimates range from 10 to 30 million species of insects.

The climate supporting the Hylaea is basically nonseasonal. However, seasonal snowmelt in the Andean headwaters of the Amazon results in a seasonal discharge regime that leads to months-long seasonal flooding of vast rain forest areas and drastic seasonal changes of conditions for all life. Similar situations may occur elsewhere. Evergreen tropical rain forests exist in parts of South America, in Central America, in a discontinuous belt across equatorial Africa, and in parts of Southeast Asia, from whence they extend into tropical northeastern Australia, where only small remnants remain.

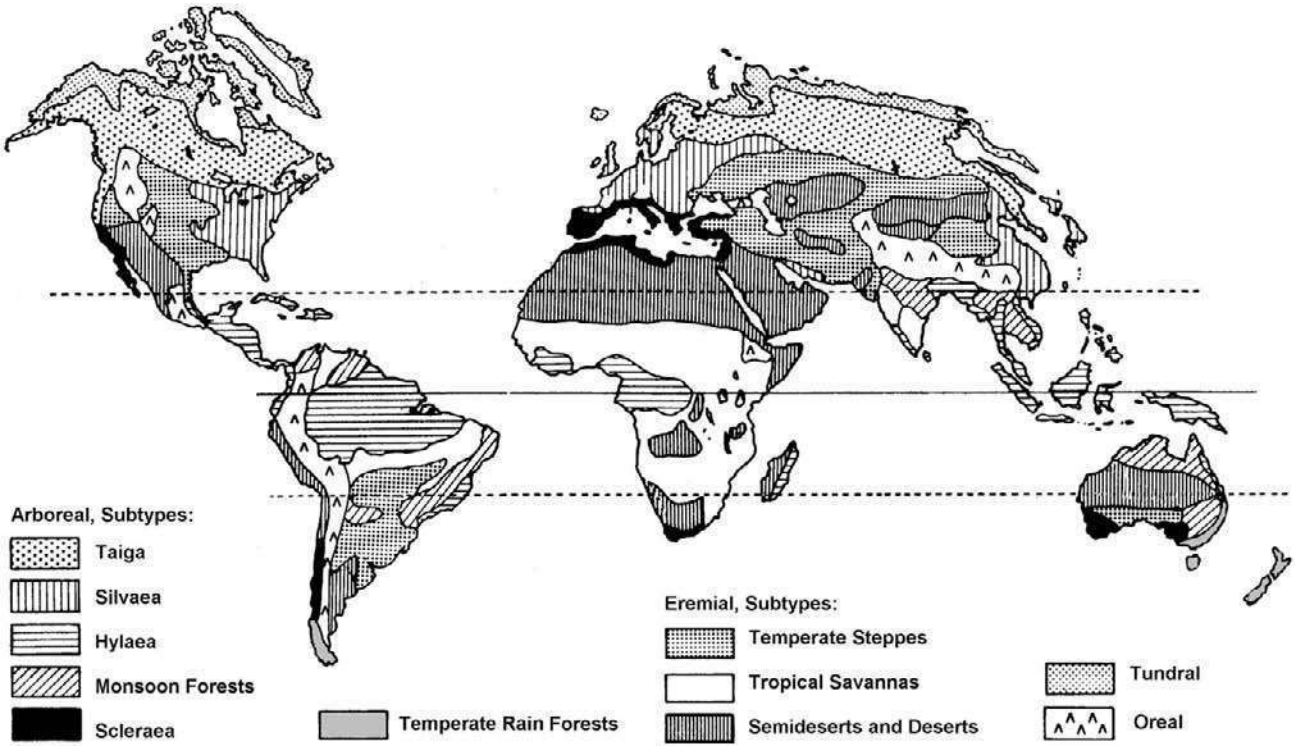

FIGURE 2 Main biomes. [Modified from a map of vegetation zones in Cox, B. C., and Moore, P. D. (1985). "Einführung in die Biogeographie." UTB 1408. Gustav Fischer Verlag © Spektrum Akademischer Verlag, Heidelberg.]

SILVAEA The term "Silvaea" refers to summer green deciduous broad-leafed forests, which, like the Hylaea, were much more widespread in the Tertiary than they are today. They occur in oceanic-to-suboceanic subtropical-to-temperate areas, mainly in eastern North America, in central Europe and the northern portion of southern Europe and Asia Minor, and in eastern China, Korea, and Japan. Adequate humidity is permanently available, and the species-rich vegetation offers protection against wind and radiation. During the vegetation period, the biota experiences favorable temperatures. In autumn, insects withdraw and most are inactive during winter, which may be frosty, although the soil does not freeze to a great depth. Strictly seasonal leaf shedding provides enormous amounts of dead plant material. Because low temperatures reduce production less than decomposition, much detritus accumulates, which provides habitat and food for many specialists among the diverse arthropod and insect fauna.

The evergreen temperate rain forests are located in restricted areas of the Southern Hemisphere, especially in Patagonia, southeastern Australia (including Tasmania), and New Zealand. They are in many ways similar to the two previously mentioned types but geographically disjunct, except along the Australian east coast where tropical and temperate rain forests meet and intergrade. Large tropical and subtropical areas with monsoon climate support forests that are only seasonally green (Fig. 2).

SCLERAEA Hard-leafed trees and shrubs dominate in the Scleraea in subtropical areas with rainy, mild winters and hot, dry summers. This evergreen forest type occurs not only along the western borders of the large landmasses, mainly in California and the European Mediterranean region, but also in middle Chile, the southwestern Cape of Africa, and southwestern Australia. Thick bark, as well as hard, reflectant, and often wax-covered leaf surfaces, or feltlike, often rolled-in leaf undersides and other modifications protect the deeply rooted plants against summer heat and dryness; winters tend to be wet and cool, but frost is rare. Life cycles are fitted to this pronouncedly seasonal regime, and some specialized insects also move to protected underground habitats where there are minute, blind, soil-inhabiting ground beetles and rove beetles (Carabidae and Staphylinidae). Plant cover is diverse and sufficiently dense to provide food and protection so that overall conditions for insect life are good; insect diversity is therefore very high.

TAIGA A small number of conifer species form a belt of northern evergreen forests known as taiga. The taiga ranges from North America through northern Asia to Scandinavia. Summer temperatures during the short vegetation period may be high, but the duration and severity of winters, and also the relative monotony of vegetation, limit the number of insect taxa. During the Pleistocene, the taiga was displaced southward; the more southern montane conifer forests in the

Northern Hemisphere date back to this period. Biodiversity in the taiga is generally low; a few species dominate and may inhabit vast areas.

Eremial Bioregions

The large generally treeless arid areas on earth, mainly steppes, semideserts, and deserts, are collectively called "eremial." An almost continuous belt extends through North Africa and Asia, from Mauretania to eastern Mongolia. The western portion, including the Indian Thar Desert, is hot and dry; the more northeasterly areas experience extreme winter cold. The Ethiopian eremial center and the Kalahari were originally separate but became more or less connected to Arabia and the northern Palearctic eremial belt via dry savannas in central and East Africa, after the Pleistocene. This permitted some exchange of eremial fauna, but the eremial centers on other continents remain isolated from each other. Therefore, the faunas are phylogenetically different, but all must be tolerant of lack of cover, as well as dryness, strong solar irradiation, extreme diel temperature changes, and often strong winds. Compared with other biomes, insect biodiversity is low, but even the most extreme, vegetation-free types of desert are not totally void of insects; several darkling beetles (Tenebrionidae), for example, survive on wind-transported organic material.

Oreal and Tundral Bioregions

High mountains above the tree line constitute the oreal biome, and treeless areas close to the poles form the tundral biome. The oreal and the tundral biomes are ecologically similar and often are considered together as oreotundral. This classification is particularly justified in view of contacts between the two realms during the Pleistocene. Treelessness is caused by cold temperatures and the short vegetation period, and sometimes also by exposure to wind. The Arctic tundral is a large zonal biome, and large areas have permafrost soil. During the short summer, the deeper soil never thaws; thus meltwater remains on the surface, leading to the establishment of extensive swamps and bogs. Low temperatures impede the rotting of dead plants, which are largely mosses, and peat formation is therefore common. Tundral areas in the Southern Hemisphere have no permafrost soil; they are highly fragmented and essentially restricted to the subantarctic islands. Insect biodiversity is low, especially in the tundral, and less so in the oreal.

Dinodal Bioregions

The overall agreement of the distribution of freshwater insects with terrestrial fauna is in line with the universal experience that freshwater bodies in many ways mirror the conditions in their catchments. This applies particularly to running waters that are largely allotrophic (i.e., depend on inputs of organic material from the environment), most obviously in that members of the shredder functional feeding group depend on dead leaves and other coarse organic material that only terrestrial inputs make available in streams. However, not all trophic groups are equally dependent on the terrestrial environment. Based on the presence of several endemic stream caddisflies in formerly glaciated or severely impacted areas in central Europe, swiftly and vehemently flowing streams (usually assigned to the arboreal biome) are believed to constitute a separate new biome, the dinodal, which would be largely independent of the biome in the catchment. (See work by Malicky, 1983.)

INSECT RANGES

The geographical area in which a species regularly occurs and maintains itself through natural reproduction is called its distribution area or range. Species ranges differ in size from individual small islands (or some other island habitats, like an isolated mountaintop, a particular lake, or some individual cave) to entire continents to the entire Northern or Southern Hemisphere or even to almost global distributions. The term "range" is also used to describe the distribution of geographical races (subspecies) and supraspecific taxa, such as genera and families.

Within their ranges, insects are not randomly or evenly distributed. Specimens are usually clumped and restricted to habitats fulfilling the species' particular ecological requirements. Abiotic factors, or the presence of particular food or host plants, but also the absence of predators, parasites, competitors, and others, may be important in determining their occurrence. Where suitable habitats are at some distance, more or less isolated subpopulations, which only occasionally interbreed, result. However, as long as gene flow is not completely disrupted there is a single, continuous range. Within-range aspects, such as fine-scale patterns of distribution and clinal or discrete variation in morphology, physiology, or other characters across ranges, occur but are not considered here. Similarly, seasonal or diurnal differences of specimen distribution, habitat changes between different life stages, and other intrinsic details occur but are not dealt with here.

Species that occur in several spatially separate, reproductively isolated populations are called disjunct. Disjunctions arise because ranges experience extensions and restrictions that are often induced by a combination of factors. Active and passive animal dispersal, changing ecological conditions, and changes of the earth's surface—for example, by sea transgressions (e.g., level changes), orogenesis (e.g., mountain building), or continental drift—may be involved. Because of this complexity, size and shape of ranges are not generally related to insect size and mobility. To provide some examples from butterflies that are strong flyers, the birdwing, *Ornithoptera aesacus,* is endemic to small Obi Island in the Maluku Islands, whereas the peacock, *Inachis io,* is endemic to Eurasia.

Dispersal

ACTIVE DISPERSAL Random movement of individuals in a growing population leads by itself to some peripheral range extension until eventually the entire inhabitable space is occupied. Small-scale ecological change and normal insect activity lead to range extensions, or to restrictions, if conditions deteriorate. Although some active dispersal is involved in all range extensions, the term is most often applied to long-distance movements. These are often observed in migratory species, but only the area in which a species regularly reproduces is called its range. The area in which it appears only during migration is separately recorded. Long-distance migrations may be by single specimens or by large numbers and may or may not lead to temporary or lasting range extensions.

Under favorable conditions, the population density of some insects such as in the notorious migratory locusts, can become high enough to induce emigration of large numbers. Similar situations occur in dragonflies, for example, the European fourspotted chaser *(Libellula quadrimaculata),* and in other species. However, migrants usually move within the general range of the species, attaining only temporary and marginal range extensions. In Europe, some butterflies—for example, the painted lady and the red admiral (*Vanessa cardui* and *V. atalanta*)—regularly migrate to north of the Alps, and some Mediterranean moths (e.g., the death's head hawk moth, *Acherontia atropos,* the convolvulus hawk moth, *Agrius convolvuli,* and *Daphnis nerii*) do the same in warm summers. However, photoperiodic cues or winter temperatures do not permit lasting establishment in central Europe. Seasonal mass migrations are performed by the monarch butterfly, *Danaus plexippus,* and long-distance dispersal of individual butterflies is often observed. *Danaus* established itself in New Zealand, Australia, and elsewhere but only after humans introduced milkweeds, the food plants that had originally been absent from these areas.

Active dispersal is most easily noticed in spectacular forms such as those mentioned earlier. In most insects, the numbers moving and the distances traveled remain unknown but may be important. For example, many insects, such as hoverflies (Diptera: Syrphidae) and moths (Noctuidae) but also large numbers of dragonflies, were observed migrating across some high alpine passes in Switzerland. On most days, thousands were trapped in malaise traps or light traps.

PASSIVE DISPERSAL Passive dispersal or transport of insects occurred naturally long before the involvement of human traffic. For range extensions, passive dispersal may be equally or more important than active movements. Transport in the pelt or plumage of larger animals occasionally occurs and may be important for the colonization of, for example, isolated ponds. Flooding streams and rivers move huge amounts of riparian organic debris plus the associated fauna downstream, sometimes over large distances. In large, tropical rivers, floating trees with a diverse fauna or entire vegetation islands have been observed traveling substantial distances, eventually also over sea. There is now also widespread agreement that the post-Pleistocene (re)colonization of parts of Scandinavia, Iceland, and Greenland was through drifting ice carrying soil and associated biota from refugial areas in western Europe.

Species with limited flight capacity and strong flyers alike are exposed to air transport. Collections made on ships stationed on the open sea or from airplanes or the outfall on high mountains show that amazingly large numbers of insects and spiders travel as aerial plankton. These air-transported species are mainly small organisms and not only those spending part of their life in some resistant inactive state, such as rotifers and tardigrades. Apparently through passive aerial dispersal, some of the smallest animals have some of the largest ranges, and some very small species are even of global distribution. Air transport ("ballooning") may form part of distributional strategies, for example, in spiders and also some first-instar Lepidoptera that produce silk strands, facilitating being caught and carried by moving air. Examples can be found in the arrival and partly the subsequent establishment of a number of butterflies and probably also other insects in New Zealand during the last 150 years.

However, the importance of passive transport has sometimes been overestimated. For example, at a time when no other explanations seemed to exist, transport by westerly storms encircling the southern end of the world in the "Roaring Forties" latitudes was thought to have caused continental disjunctions that can today more convincingly be explained by continental drift.

Various quarantine measures are presently taken against unintended human transport of insects, but the problem is an old one. For example, when the Vikings came to North America, they contracted human fleas. Preserved fleas were found in Viking settlements on Greenland whence they were apparently carried to Europe, where human fleas first appeared around the year 1000. Several soil-dwelling beetles were introduced to North America with ship ballast collected in Europe at sites where the particular beetles abound. Their survival today indicates suitable ecological conditions in America, but only some of the beetles dispersed widely. Others spread easily, such as the Colorado potato beetle *(Leptinotarsa decemlineata),* which expanded its restricted natural range over much of North America and over Europe, where it was also introduced. As with the monarch butterfly, the intentional introduction of the food plant had prepared its way.

LARGE-SCALE ECOLOGICAL CHANGE: EFFECTS OF THE PLEISTOCENE

The ecological relations of extant (i.e., existing) insects are assumed to have been the same in the past as they are now; if different assumptions are made, they must be explained. The drastic climatic changes since the end of the Tertiary and especially during Pleistocene glaciations profoundly

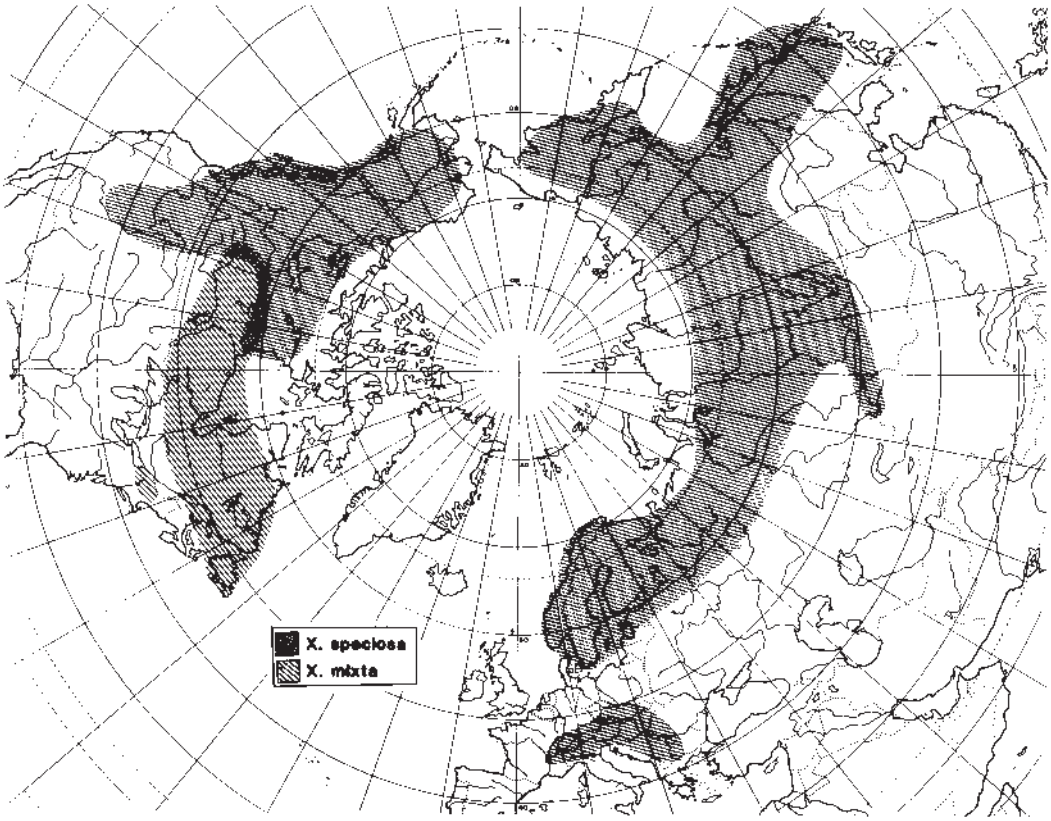

FIGURE 3 Transberingian and arctoalpine disjunctions in circumpolar Noctuidae of the genus *Xestia;* the Old World *X. speciosa* is represented in western North America by the subspecies *aklavicensis;* the origin of the eastern North American *X. mixta* is thought to predate the Wisconsin glaciation. [From Mikkola, K., Lafontaine, J. D., and Kononenko, V. S. (1991). Zoogeography of the Holarctic species of the Noctuidae (Lepidoptera): Importance of the Beringian refuge. *Entomol. Fenn.* **2,** 157–173.]

impacted insect distributions. The north polar ice expanded south, leading to a southward displacement of zonal biomes in the Holarctic. The tops of major mountains further south also acquired ice caps. The Silvaea and its fauna were driven into southern refugia, whereas cold-adapted oreotundral insects became established in their former place. The process was reversed when the ice began to retreat. The zonal biomes and their associated biota then shifted again northward; they probably continue to do so today.

The essentially north–south orientation of major mountain ranges in North America facilitated the displacements. In Europe, where major mountains run mainly from east to west, there remained an ice-free corridor between the polar ice front and the glaciated Alps, inhabited by a mixed fauna of northern and alpine origin, respectively. Many animals were driven further southwest, toward the Pyrenees, or southeast into the mountains of the Balkan peninsula. About 18,000 years ago, the ice began to retreat and cold-adapted insects followed it. There was a partial faunal exchange, and a number of fracturing of ranges, or disjunctions, resulted. Today, representatives of several insect orders exhibit boreoalpine or boreomontane (or arctoalpine and arctomontane, respectively) disjunctions (Fig. 3), but only rarely has this led to perceptibly divergent evolution, or even speciation.

Because of European topography, Pleistocene refugia of the south-retreating Silvaea and its fauna were mainly on the three large Mediterranean peninsulas. Many of the present central European species can clearly be assigned to one particular refugium because their ranges tend to coincide and occupy all of the former refugium, even though postglacial climate change and human impact strongly fragmented the Mediterranean deciduous forests. In contrast, at the northern range limits the individual species returned, variably far into once devastated areas. Postglacial recolonization by European insects was apparently fast because not all insects were affected by barriers like the English Channel, which formed about 8000 years ago, or the Baltic Sea straits separating Jutland from Scandinavia, which definitely formed some 6400 years before the present. The foregoing scenarios, initially inferred from distribution patterns, were later backed up by fossils, especially well-sclerotized and easily preserved beetles. Today, molecular genetic studies in a new line of research, phylogeography, provide support to these historical reconstructions.

The binding of much water as ice during the glacial periods lowered the sea level by about 100 meters during the last glacial period, which made important land bridges available. Tasmania was connected to Australia, which in turn established contact with New Guinea, which itself had

ties with the Oriental region. The latter included a large continuous landmass, the Sunda plate, where numerous separate islands remain today. Japan was connected to the Asian, and England to the European mainland.

The Bering bridge connected East Asia and western Alaska, which were covered by tundra. Thus, there is a fair number of shared species or pairs of sibling species in East Asia and northern North America, mainly among tundral insects. Numbers of terrestrial as well as aquatic insect species are of circumpolar distribution (Fig. 3). Ecological demands seem to mostly prevent a southward spread of these northern species.

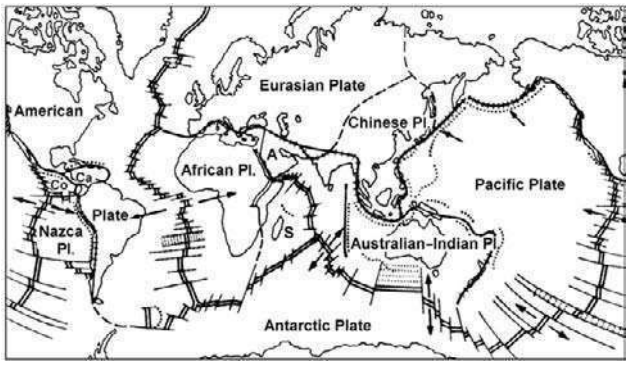

FIGURE 4 Midoceanic ridges and continental plates: A, Arabic plate; Ca, Caribic plate; Co, Cocos plate; S, Somalian plate. Arrows indicate directions of plate movements. Double lines represent midoceanic ridges; transverse lines across them are faults. Subduction and compression zones are mainly along deep-sea rift valleys (dotted lines) or mountain chains (crosses). Figures and stippling identify million years of seafloor spreading. [Modified after Thenius, E. (1979). "Die Evolution der Säugetiere." UTB 865. Gustav Fischer Verlag © Spektrum Akademischer Verlag, Heidelberg.]

ANCIENT DISTRIBUTION PATTERNS AND CHANGING EARTH SURFACE

Changing Concepts

Widely disjunct distributions of older origin than discussed so far are revealed when taxa of higher rank, for example, families, are considered. Mainly in the Southern Hemisphere, ranges of close relatives may be separated by wide oceans. Explanations proposed for these patterns changed in accordance with the developing understanding of animal evolution and of changes occurring on the surface of the earth.

First, scientists proposed the former existence of numerous land bridges in early times to explain disjunctions. Most of the proposed land bridges never existed, but a few indeed did occur beyond those entirely caused by Pleistocene sea level fluctuations. Greenland, for example, was long connected with North America; the so-called DeGeer route connected northern Greenland with the extreme northwest of the then larger European continent. To the south, the Thule bridge connected Iceland, the British Isles, and the rest of Europe.

Later, and for as long as continents were believed to have been stable, insect dispersal was the favorite explanation for disjunctions. Routes and corridors along which animals might have moved, and bottlenecks or filters allowing only the passage of selected taxa, were discussed. Dispersal over long distances and across major hazardous obstacles along so-called sweepstake routes, where only few taxa would succeed, and largely by chance, also was considered. Actually, some dispersal can never be excluded, but the phenomenon is by itself insufficient to explain many insect distributions. This is also obvious from the studies of L. Croizat in 1958, who developed a method that he called pan-biogeography. Croizat connected ranges of related animals by lines (or tracks) and observed similar patterns in quite different groups of animals, suggesting the existence of "generalized tracks." Some of these tracks connected landmasses across oceans that all the many taxa with different dispersal capacities could not possibly have crossed; however, no general explanation of wide transoceanic disjunctions was offered.

Major advances in both animal systematics and in the earth sciences profoundly changed the situation, and this occurred only a few decades ago. On the zoological side, the work of Willi Hennig was instrumental in the development of modern zoogeography. Henning showed how the degree of phylogenetic relationship, or closeness of common ancestry (as opposed to some vague relatedness) between taxa can be recognized and reflected in the animal system. He also explained that postulating former land connections is logically justified only if sets of phylogenetically related taxa (i.e., branched sections from the hierarchical animal cladogram) exhibit similar disjunctions. Otherwise, relic distribution from once wider ranges or, alternatively, chance dispersals, are no less probable, even among widely disjunct individual sister taxa.

In the earth sciences, Alfred Wegener in 1912 suggested that continent positions are not stable but change over time. Evidence presented in support of continental displacements included the good fit of continental shelf lines, as well as observations of areas with particular deposits or minerals, traces of paleozoic glaciations, and particular mountain chains on separate continents. However, as long as no mechanism providing the power for movements of continents could be identified, this evidence remained unconvincing.

The situation changed a few decades ago when midoceanic ridges on the seafloors (Fig. 4) were recognized as sources of magma from the fluid interior of the earth; ridges form a network delimiting the continental plates. As magma appears at the surface, it pushes sideward, and the seafloor is spreading. Magnetic particles in the magma become uniformly oriented in the global magnetic field. This orientation is preserved when the magma cools and hardens. Bands of seafloor differing in magnetic orientation (or in paleomagnetism) extend parallel to the midoceanic ridges; evidently, the global magnetic pattern is at times reversed. Several centimeters of new seafloor is produced per year. In combination with measurements of paleomagnetism, the age of seafloors was estimated and found to increase with distance from midoceanic ridges, from contemporary at the ridge to only

about 65 mya at the distant sites. Seafloors are generally young, the most ancient ones are only about 200 mya old. Seafloor spreading provides the power that shifts the continental plates, which because of overall differences in elemental composition, are of lower specific weight than seafloor. Therefore, most of the continental material remains afloat while essentially excess seafloor is subducted back into the fluid center of the earth. Such subduction zones occur in deep-sea valleys, mostly along continental edges. Subduction zones coincide with arcs of major vulcanism and earthquake activities. Floating continents may slide past each other along friction zones, continents may collide and cause upfolding of mountains, or they may merge or break up.

Overview of Continental Drift Pattern

Once the mechanism driving continental movements had been recognized, continental drift was widely accepted. Continental plates are moving, merging, and breaking up since their formation. Using all available evidence, paleogeography can describe past changes of the earth's surface in fair detail (Fig. 5). The origin of life in general and also of several insect orders dates back farther than the formation of a single supercontinent, Pangaea, but the methods of zoogeography cannot provide insight into earlier events. Zoogeography deals mainly with subsequent changes, in particular, the breakup of Pangaea into the northern and southern continents, Laurasia and Gondwanaland, respectively, and their further fates.

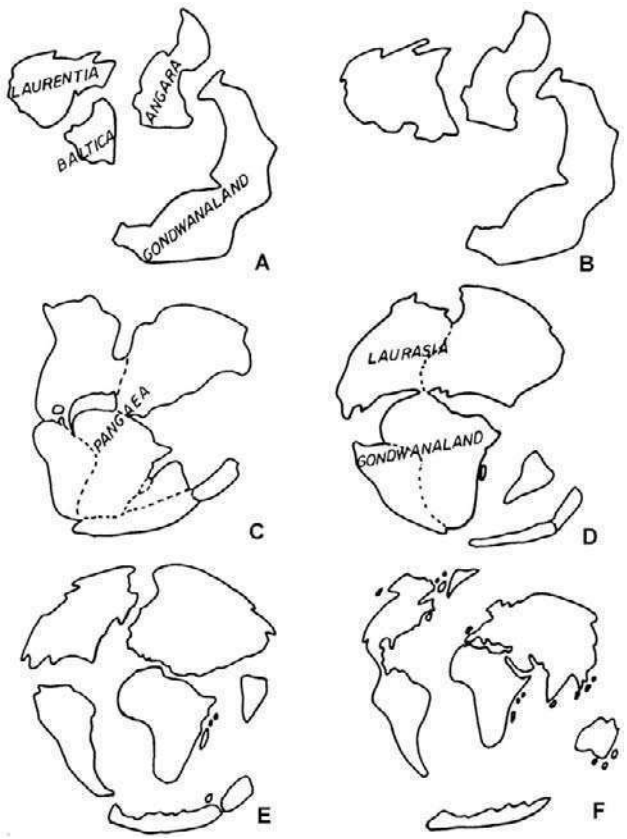

FIGURE 5 Pictorial summary of continental drift, from approximately 330 mya to present times: (A) 330 mya, (B) 300 mya, (C) a single landmass, Pangaea, 280–200 mya, (D) separation of Gondwanaland and Laurasia (180 mya), (E) 40 mya, (F) the Americas reunited, 25 mya. [From Vickery, V. R. (1989). The biogeography of Canadian Grylloptera and Orthoptera. *Can. Entomol.* **121,** 389–424.]

Consequences for Biogeography

Knowledge of seafloor spreading, plate tectonics, and continental drift has profoundly affected zoogeography and increased the relative importance of its alternative approaches. The so-called vicariance approach, based on evolutionary theory and on phylogenetic systematics in combination with information on continental drift, gained great explanatory power. Closely related taxa inhabiting separate ranges are called vicariant. Splitting up of populations with disruption of gene flow, which are called vicariance events, becomes the starting point of divergent evolution, eventually leading to differences in species. Over geological time, different vicariant sister clades may evolve from the ancestral species.

Today, one can understand how the breakup of supercontinents led to wide disjunctions and induced separate evolution of related taxa, on separate continents. Continental drift actually provided for means of transport, and insects can now be seen rafting on drifting continents instead of dispersing between them, across wide oceans. Vicariance biogeography seeks for congruences between the evolution of landmasses and the evolution of animals living on them and envisions the first process driving the second. There are now elaborate methodological considerations as well and they are described in works by Humphries and Parenti, and Wiley.

Ancient Disjunctions in Northern Hemisphere

Intra-American faunal differences provide evidence that contemporary insect distributions are almost always the result of a variety of causes that were effective at different times. Most disjunctions between related groups inside North America occur along a line that runs through the central plains, from northwest to southeast. This separation line results from present ecological differences between the mountains that support mainly the arboreal biome and the essentially eremial plains, from the past existence of a midcontinental seaway in the area of the present plains, and from past affiliations of the mountainous eastern and western halves of North America with other continents.

The areas adjacent to the present Bering Strait support tundras, and so did the ice-free areas on and around the former Bering bridge, but forest-dwelling insects had no access to this land bridge. Nevertheless, among the more southern arboreal insects, genera are often shared between eastern Asia and North America; species tend to differ between continents (Fig. 6). These disjunctions date back much further than the

Pleistocene. Because the Angara shield was separated from the Canadian shield only about 2 mya, by the opening of the northern Pacific Ocean, range disjunctions at the specific or generic levels between the Asian Far East and western North America are observed in many orders.

The northern continents formed through fusion of the ancient Canadian (or Laurentian), Fennoscandian (or Baltic), and Angara continental shields that were subsequently again divided and reunited, until the present pattern appeared. North America and Europe were connected until the opening of the Atlantic Ocean, through seafloor spreading, about 70 mya ago. Close phylogenetic relations between various insect groups in eastern North America and Europe are evidence of the past unity (Fig. 7). Europe was to the east long separated from Siberia by the Turgai Strait east of the Ural Mountains, which explains differences in the European and Asian faunas, despite the present continuity of land (Fig. 6).

Southern Hemisphere Case Study

The stepwise disintegration of Gondwanaland caused some of the most striking disjunctions and long unexplained "transantarctic" or "amphinotic" relations between animals living in Andean South America, Australia, and New Zealand. These landmasses are now known to have long remained connected with, or closely adjacent to, the then forested and inhabitable Antarctic continent.

The fundamental change in views, from dispersalism to continental drift, is recent; the Plecoptera (or stoneflies) provide an example. In 1961 stonefly evolution was still explained entirely by long-distance dispersal involving transgressions of the equator, two in each of the two then recognized suborders. An initial movement of primitive taxa from south to north was assumed, followed by the return of evolutionarily advanced forms to the Southern Hemisphere. In 1965 elements of cladistics and continental drift were added to this scenario. A few years later, a cladistic approach to Plecoptera systematics led to a widely accepted revised system, suggesting that continental drift steered stonefly evolution. The breakup of Pangaea into Laurasia and Gondwanaland seems to have caused the separation into distinct Southern and Northern Hemisphere suborders, the Antarctoperlaria and Arctoperlaria, respectively. When Gondwanaland fell apart, the ranges of the suborder Antarctoperlaria and its families became disjunct, distinct representatives that evolved on each of the distant landmasses (Fig. 8).

However, continental drift alone can probably not explain all of the present Plecoptera distribution. Ecology and also dispersal remain important. Antarctoperlaria must have been present on Gondwanaland before Africa and India broke away from it. Ecological change, perhaps past dryness, is thought to have caused their disappearance from these lands. More difficult are two arctoperlarian families of which subordinate endemic groups are present also in the Southern Hemisphere. Contrary to widespread belief, not all Plecoptera are cool adapted. The Australian Gripopterygidae, Eustheniidae, the arctoperlarian Leuctridae, and the Nemouridae include many tropical species; they are most numerous in the large family Perlidae. The many *Neoperla* in the Ethiopian region are clearly of northern origin, but the origin of the diverse South American Perlidae is uncertain. Most problematic, however, is the so-called family Notonemouridae. Its monophyly is doubtful; it may represent independent early branches of the Nemouridae. Nevertheless, all notonemourids live in temperate parts of South America, Australia, New Zealand, South Africa, and Madagascar, but not in India. Dispersal seems to have contributed to these distributions that, admittedly, remain essentially unexplained. A practical test using the methodological refinements of vicariance biogeography proposed by Humphries and Parenti would require a better understanding of phylogenetic relationships among Plecoptera than is presently available.

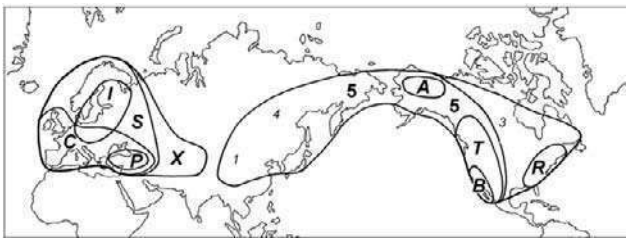

FIGURE 6 America–Asian relations, and the distinctness of the European fauna: distribution of the genera of Chloroperlinae (Plecoptera: Chloroperlidae). The ranges of five genera (*Alloperla, Haploperla, Plumiperla, Suwallia,* and *Sweltsa*) indicated by **5** in northeastern Asia and northwestern America, largely overlap. Numbers of genera of this group decline east- and westward; figures in italics are numbers present in the respective areas. The other genera are *A, Alaskaperla; B, Bisancora; C, Chloroperla; I, Isoptena; P, Pontoperla; R, Rasvena; S, Siphonoperla; T, Triznaka; X, Xanthoperla.*

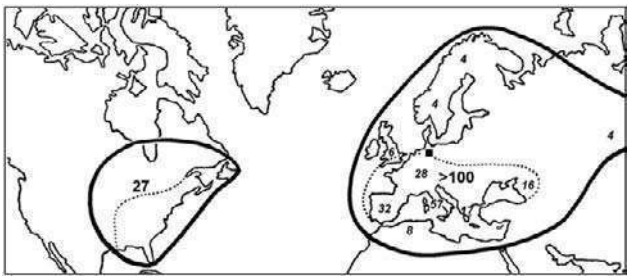

FIGURE 7 American–European disjunctions. Solid lines: range of the stonefly genus *Leuctra* (Plecoptera: Leuctridae); bold figures are total numbers of species per continent, figures in italics are regional numbers of species. A single species, *L. fusca,* occurs all over Europe and extends through Siberia to the southern portion of the Russian Far East. Broken lines: ranges of the extant ants *Ponera pennsylvanica* (America) and *P. coarctata* (Europe); *P. atavia* (black square) is an amber fossil (Hymenoptera: Formicidae). [Range information after Noonan, G. R. (1988). Faunal relationships between Eastern North America and Europe as shown by insects. *Mem. Entomol. Soc. Can.* **144,** 39–53.]

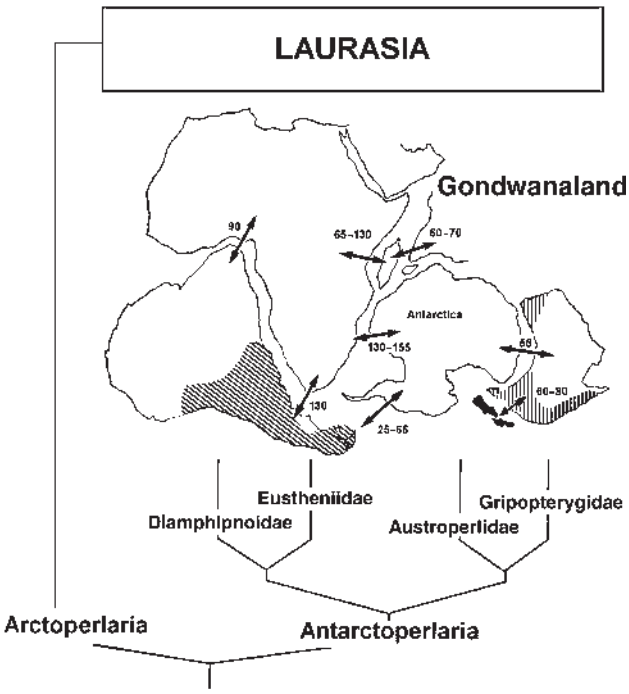

FIGURE 8 Phylogenetic system and distribution of the Plecoptera. The range of the family Gripopterygidae is shown on a map of Gondwanaland at the end of the Cretaceous, with dates of last possible faunal exchange (mya). Different shading indicates that each of the disjunct areas has an endemic fauna; no genus is shared. The Eustheniidae and Austroperlidae are distributed in the same way but have narrower ranges, the five species of the Diamphipnoidae are all South American. [Map based on Crosskey, R.W. (1990). "The Natural History of Blackflies," copyright The Natural History Museum, London.]

It is not common for the first step in the breakup of Pangaea to be clearly reflected in the phylogenetic system; otherwise, however, the Plecoptera have many parallels among other insects. Southern Hemisphere disjunctions suggesting a Gondwanian origin are widespread among aquatic (e.g., Ephemeroptera, Odonata, various dipteran midges) and terrestrial insects, for example, in the Hemiptera, Neuroptera, Mecoptera, and Coleoptera, to name a few. Some of these disjunct groups also comprise African representatives. The phylogenetic relationships within several of these groups of insects appear to reflect the proposed sequence of the disintegration of Gondwanaland. The breakup provided series of vicariance events enabling phylogenetic divergence.

See Also the Following Articles

Biodiversity • Fossil Record • Introduced Insects • Island Biogeography and Evolution • Population Ecology

Further Reading

Briden, J. C., Drewry, G. E., and Smith, A. G. (1974). Phanerozoic equal-area world maps. *J. Geol.* **82,** 555–574.

Cranston, P. S., and Naumann, I. D. (1991). Biogeography. *In* "The Insects of Australia," Vol. 1, pp. 180–197. Melbourne University Press, Melbourne.

Croizat, L. (1958). "Panbiogeography," Vols. 1, 2a, 2b. Caracas.

Downes, J. A., and Kavanaugh, D. H., eds. (1988). Origins of the North American insect fauna. *Mem. Entomol. Soc. Can.* **144,** 1–168.

Hennig, W. (1960). Die Dipteren-Fauna von Neuseeland als systematisches und tiergeographisches Problem. *Beitr. Entomol.* **10,** 221–329.

Humphries, C. J., and Parenti, L. R. (1986). "Cladistic Biogeography." Oxford Monographs on Biogeography 2. Clarendon Press, Oxford, U.K.

Malicky, H. (1983). Chorological patterns and biome types of European Trichoptera and other freshwater insects. *Arch. Hydrobiol.* **96,** 223–244.

Morgan, A. V., and Morgan, A. (1980). Faunal assemblages and distributional shifts of Coleoptera during the late Pleistocene in Canada and the northern United States. *Can. Entomol.* **112,** 1105–1128.

Morrone, J. J., and Crisci, J. V. (1995). Historical biogeography: Introduction to methods. *Annu. Rev. Ecol. Syst.* **26,** 373–401.

Platnick, N. I., and Nelson, G. (1978). A method of analysis for historical biogeography. *Syst. Zool.* **27,** 1–16.

Raven, P. H., and Axelrod, D. I. (1974). Angiosperm biogeography and past continental movements. *Ann. Missouri Botan. Garden* **61,** 539–673.

Taberlet, P., Fumagalli, L., and Wust-Saucy, G. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mole. Ecol.* **7,** 453–464.

Tarling, D. H., and Tarling, M. P. (1975). "Continental Drift." 2nd ed. Doubleday, Garden City, NY.

Wiley, E. O. (1988). Vicariance biogeography. *Annu. Rev. Ecol. Syst.* **19,** 513–542.

Zwick, P. (2000). Phylogenetic system and zoogeography of the Plecoptera. *Annu. Rev. Entomol.* **45,** 709–746.

Biological Control of Insect Pests

R. G. Van Driesche
University of Massachusetts, Amherst

M. S. Hoddle
University of California, Riverside

Biological control is a form of pest control that uses living organisms to suppress pest densities to lower levels. It is a form of ecologically based pest management that uses one kind of organism (the "natural enemies") to control another (the pest species). Types of natural enemies vary with the type of pest. For example, populations of pest insects such as scales are often suppressed by manipulating populations of parasitoids, which are insects that develop in or on the pest insects they attack and kill. Populations of plant-feeding mites, such as the common twospotted spider mite (*Tetranychus urticae*) are often limited by predators, especially mites in the family Phytoseiidae. Populations of weeds can be suppressed by specialized herbivorous insects that feed on them. Finally, many insect populations have pathogens (e.g., bacteria, viruses, or fungi) that infect them. Such pathogens, whether they occur naturally or are applied artificially as microbial pesticides, can locally and temporarily suppress a pest's numbers.

Biological control is thus about the relative numbers of pests and their natural enemies. Predators (or pathogens, parasitoids, herbivorous insects) increase in number over time and feed on the pest, whose population then declines, because of higher mortality or lower birthrates caused by predation by natural enemies. Biological control agents are living organisms that increase in number through reproduction in response to pests that are used for nutrition. Biological control, at least in some of its forms, has the potential to be permanent in its action, through the reproduction and spread of the natural enemies as they track target pest populations.

There are four broad ways in which people have manipulated natural enemies to enhance their action: natural enemy importation, augmentation, conservation, and application of microbial pesticides. Each of these approaches has its own rationale, history, and level of past successful use. Biological control has important advantages compared with other methods of pest control. Moreover, because pest control is relatively permanent and requires no further capital input, biological control through self-sustaining forms (natural enemy importation and conservation) is often cheaper than use of pesticides. Although there may be significant initial costs (especially for projects that import new species of natural enemies from the pest's country of origin), costs drop to low or even zero levels in later years, whereas the benefits of the pest control achieved continue to accrue for years. For other forms of biological control (natural enemy augmentation and use of microbial pesticides), control is not permanent and costs recur annually, as with pesticides. For the latter two approaches, biological control may be either more or less expensive than other approaches depending on details such as the cost of natural enemy production by commercial insectaries that sell beneficial organisms, and the efficacy of other control tactics. In all four forms, biological control has the advantage of being virtually harmless to people and vertebrates, whereas pesticides must be actively managed for safe use to mitigate harm to humans and other nontarget organisms.

Biological control as a scientific endeavor has a history of about 125 years of effective use (beginning in the 1880s), over which time new information, techniques, and technologies have increased humankind's ability to use biological control agents with increasingly greater understanding and effectiveness. Before this period of active use, there were several centuries during which ideas about predators, parasitoids, pathogens, and their links to pest populations evolved. These ideas had to be developed before biological control as an applied pest management activity could be conceptualized.

HISTORICAL DEVELOPMENT

Biological control is the deliberate attempt by people to make practical use of the capacity of predation, parasitism, herbivory, and disease to restrain the growth of plant and animal populations. The ability to make practical use of these processes depends on an understanding of how pest population densities are controlled by natural enemies. Biological control also requires detailed knowledge of the biology of pests and their key natural enemies, because such knowledge often provides the means for their practical manipulation.

Predation

Predation by vertebrates on other vertebrates has long been part of human knowledge. Predation as a force affecting pest insects was recognized when people first saw individual acts of predation taking place on their crops. A close observer of an aphid colony, for example, cannot help but see the predatory action of ladybird beetles and cecidomyiid midge larvae as they devour their aphid prey. The predaceous effect of some species of ants on pest insects associated with citrus was recognized thousands of years ago by farmers in Yemen and China, who used the knowledge to suppress these pests by moving ant colonies into new orchards. In Europe during the Renaissance, the emergence of natural history as a subject worthy of observation and thought led keen-eyed naturalists to arrive at similar findings. The father of the classification of plants and animals, Caralus Linnaeus, observed in 1752 that "Every insect has its predator which follows and destroys it. Such predatory insects should be caught and used for disinfesting crop-plants." By the early 1800s, such observations led naturalists such as Erasmus Darwin and American entomologists such as Asa Fitch to suggest that predaceous insects should be used to suppress pest insects by making releases of the predators in places where they were lacking. These suggestions formed the fundamental basis for the modern use of augmentative biological control in greenhouses, vegetable production, and various outdoor crops.

Parasitism

The action of insect parasitoids on their hosts has no direct analogue among animals that people could easily observe before the invention of magnifying lenses. Consequently, the concept of parasitism took longer to become recognized. Because many parasitoids feed inside their hosts, their presence was not easily recognized, and the detection of parasitoids required that insects be either reared or dissected. By the 1600s, European naturalists were noticing the occurrence of parasitoids. Aldrovandi, for example, in 1602, reared tiny parasitoid wasps from the pupae of a nymphalid butterfly and recorded what he saw in woodcut print. Because he misunderstood the process, however, he wrongly concluded that the tiny wasps were an alternate adult form to the usual butterfly. The first person to publish a correct interpretation of insect parasitism was the English physician Martin Lister, who in 1685 noted that the ichneumon wasps seen emerging from a caterpillar were a distinct kind of insect that originated from eggs inserted into the caterpillars. No

one thought of any way to make practical use of such parasitoids, however, until 1855 when Asa Fitch proposed the importation of parasitoids from Europe to America to help suppress a nonnative invasive pest of wheat, the wheat midge *Sitodiplosis mosellana.* Fitch's ideas provided a clear plan for the modern practice of biological control through natural enemy importation, but they were not acted on for nearly 30 years. The first importations of exotic species of parasitoids between continents occurred in the 1880s, when *Cotesia glomerata* was brought to the United States from Europe to suppress *Pieris rapae,* a European pest of cabbage that had invaded North America in 1860.

Insect Diseases

The study of the diseases of insects started not for purposes of killing pest insects, but rather for protecting economically important species such as silkworms and honey bees. In the mid- to late nineteenth century, microscopes made it possible to observe bacteria and microscopic fungi, and the study of these organisms as pathogens of domesticated insects initiated insect pathology. The infectious nature of insect diseases was first demonstrated by Agostino Bassi of Italy, who in 1835 studied a fungal disease of silkworm larvae caused by the fungus *Beauveria bassiana.* Louis Pasteur continued work on silkworm diseases in France in the 1860s. The first attempt to use pathogens to destroy pest insects was made in 1884 by the Russian entomologist Elie Metchnikoff, who reared *Metarhizium anisopliae,* a fungal pathogen, and attempted to suppress the sugar beet curculio, *Cleonus punctiventris*, with application of the fungal spores. In 1911 the German scientist Berliner observed a bacterial disease of larvae of the flour moth, *Anagasta kuehniella,* and by 1938 this bacterium, *Bacillus thuringiensis,* was being marketed as a microbial pesticide for control of some species of caterpillars. These early efforts established the concepts that insects were subject to infectious diseases and that the causative agents could be reared in quantity artificially. Technical methods to use reared pathogens to reliably infect insects in crops, hence to achieve biological control, came later.

Use of Insects to Suppress Weedy Plants

Although humans have known for millennia that insects damage and even kill plants, the idea that specialized herbivorous insects could be manipulated to suppress plants considered to be weeds is a relatively new concept. The first person to suggest such use was Asa Fitch. In 1855, Fitch noted that some European plants that had invaded North America, such as toadflax (*Linaria vulgaris*), had no American insects that fed on them. He suggested that importation of insects from Europe might help suppress these invasive plants. In 1863 this concept was implemented when a scale insect was moved from northern to southern India for the purpose of damaging an invasive nonnative species of cactus (*Opuntia vulgaris*). Two related cacti, *O. stricta* and *O. inermis,* were introduced as ornamentals to Australia and became highly invasive and damaging. By 1925, these plants occurred in dense stands over approximately 20 million hectares of land. The Australian government began a survey of South America (the home of these cacti) looking for specialized insects attacking the ornamentals and a moth, *Cactoblastis cactorum,* was released into Australia in 1926. By 1932, the cacti were killed over most of the infested area and native vegetation and crops were able to reclaim the cleared ground.

These early suggestions and projects laid out the concepts that many plants are limited in number by specialized insects and that plants moved to new regions often become separated from these specialized insects because they are not moved along with the seed or nursery stock used to import the plant.

IMPLEMENTATION METHODS

The observation that nonnative insects and plants could be suppressed by importing missing specialized natural enemies from their homelands led to the first successful method for practical use of biological control. This approach is called classical biological control (because it was the first deliberate, successful application of biological control as a technology), or importation biological control, or simply natural enemy introduction.

After World War II, the chemical industry began the rapid development and marketing of chemicals to control pest insects by poisoning them. Pesticides became very popular and were used on a large scale in the second half of the twentieth century, such that the frequent application of insect-killing poisons to crops became routine. Widespread pesticide use led to a substantial reduction in the level of natural control provided by predators and parasitoids of pest insects, necessitating the further use of pesticides to suppress pest insect populations. However, many pests became resistant to one or more pesticides. This resistance sparked an interest in restoring natural control by reducing the use of insecticides in crops and making their use less damaging to natural enemies by manipulating their timing, placement, or formulation. The effort to restore natural control while making judicious use of pesticides formed the basis of the integrated pest management (IPM) movement in the late 1950s. Efforts to restore and protect natural controls by removing damaging influences such as pesticides are referred to as conservation biological control. A more recent, and less successful, mode of conservation biological control has been the attempt to increase natural enemy numbers by actively providing them with better food sources or habitats. Ideas that have been investigated include a variety of vegetation manipulations in or near crop fields, including ground covers between crop rows and unmowed field borders, where flowering plants provide nectar and pollen for natural enemies.

In the 1970s farmers in Europe producing vegetables in greenhouses were also interested in enhancing natural control

of pests such as whiteflies in tomato and cucumber crops, because frequent development of pesticide resistance had rendered pesticides alone unreliable. The desire to reduce or even completely avoid the use of insecticides in tomato and cucumber crops was further stimulated in the 1980s when growers began the practice of placing colonies of bumble bees inside their greenhouses for crop pollination. Inside greenhouses, however, there were very few natural enemies because natural enemy immigration from outdoors is difficult in the indoor, sealed crop environment. Placing whitefly parasitoids, such as *Encaria formosa,* or predatory mites, such as *Phytoseiulus persimilis,* in the crop shortly after planting allowed natural control to develop and act on incipient pest populations. Once there was demand by growers for natural enemies, it became possible for specialized businesses (insectaries) to rear and sell natural enemies. This kind of pest management is called augmentative biological control, because the goal is to augment or initiate a natural enemy population.

In some crops, pests that feed directly on the edible part of the plant may not be adequately controlled by natural enemies before damage occurs. For example, cabbageworm larvae feed on cabbage heads and codling moth (*Cydia pomonella*) larvae burrow into apples. A faster acting form of biological control that can be applied when and where needed may be necessary in these situations. Microbial pesticides can be used in this way and were developed to meet immediate needs without resorting to disruptive chemical pesticides. The most successful of these products are those containing *B. thuringiensis,* a bacterium that produces toxic proteins that kill insects within a few days of ingestion. There are many subspecies of this bacterium that can be used to control the larvae of some moths, butterflies, beetles, and flies.

CLASSICAL BIOLOGICAL CONTROL

Ecological Justification

People routinely move species such as crop plants and ornamentals across natural barriers such as mountain ranges or oceans that would otherwise limit their spread. These plants may carry with them small, unrecognized infestations of pest insects. In some cases, the plants themselves may spread and become damaging. Both invasive plants and insects often escape their specialized natural enemies when they cross geographic barriers and establish in new locations. This allows these species to reach abnormally high densities and become damaging pests. Classical or importation biological control is based on the premise that the pest was originally limited to lower densities in its area of origin by specialized natural enemies, that these control agents are missing in the invaded area, and that densities of the pest in the invaded area can be reduced by importing the missing specialized natural enemies. Two recent biological control success stories from Africa, the control of cassava mealybug (*Phenacoccus manihoti*) and water hyacinth (*Eichhornia crassipes*) illustrate these processes.

CASSAVA MEALYBUG Cassava (*Manihot esculenta*) is a tropical shrub that produces starchy tubers used much like potatoes, as a staple food source. The crop is a native of the Americas, but it is now a basic crop in all tropical countries, from Asia to Africa. In the 1970s, an unknown species of mealybug appeared on cassava in West Africa and spread rapidly throughout the cassava belt of tropical Africa. In this region, cassava was a basic food for some 200 million people. Within a few years, cassava crops began to fail as plants suffered extreme damage from high-density mealybug populations. Because the pest was clearly an exotic invader in Africa, importation biological control was seen as a means to suppress it. Furthermore, this method was chosen because it offered the possibility of providing permanent control that would not require the region's cash-poor farmers to repeatedly buy expensive pesticides and application equipment.

Cassava mealybug was believed to be from the Americas, the area of origin of the crop plant. The pest, however, was initially an unknown species. Therefore, no one knew where it could be found in South or Central America. With international funding, a cassava mealybug control project was organized. Crop protection laboratories in Africa (the International Institute of Tropical Agriculture in Benin) and South America (Centro International de Agricultura, in Colombia) worked with the Commonwealth Institute of Biological Control in Trinidad (now CABI-Bioscience, a private biological control organization in the United Kingdom) to find the pest, locate specialized natural enemies attacking it, import natural enemies to quarantine laboratories in the United Kingdom, and ship pure cultures of natural enemies on to Africa for release and evaluation in the effort to control the cassava mealybug.

Initial efforts were frustrated by an inability to find the pest in the Americas. Eventually, cassava mealybug and its parasitoid, the encyrtid wasp *Epidinocarsis lopezi,* were found in Paraguay. Upon release of this parasitoid, control was rapidly achieved. The parasitoid has spread (both naturally and from releases made by entomologists) throughout the cassava region, covering more than 26 countries. In 95% of the region, this single parasitoid has achieved stable, permanent control of this pest.

The net result of this project has been to increase food security in a region that frequently experiences food shortages. A pest has been controlled permanently (for nearly 20 years now in some areas), at no recurring cost, with no use of contaminating pesticides, and no damage to native plants or wildlife.

WATER HYACINTH *E. crassipes* is both a plant used in ornamental fish ponds and the world's worst aquatic weed. Its beautiful lavender flowers have led people to take it far

from its native range in the Amazon basin of South America. Wherever water hyacinth has been introduced into subtropical or tropical climates, it has escaped into the wild, forming gigantic mats that clog rivers and cover over bays and ponds. Among the many places invaded by water hyacinth is Lake Victoria in East Africa. The pest was first recorded there in 1980 and by the mid-1990s some 12,000 ha of weed mats had clogged bays and inlets around the lake. Economic losses resulted for fisheries (the mats impede the launching of boats and the use of nets) and for waterworks and hydroelectric power plants. Ecologically, the weed threatened one of evolution's greatest products—the radiation of cichlid fishes in the lake, some 200 to 400 species of endemic fish that have evolved in the lake. These fish, often separated by mating habits based on bright colors, were threatened by hybridization among species induced by low light under weed mats, where color-based visual recognition mating systems could not be sustained.

Controls efforts recommended to the governments of the affected countries (Uganda, Kenya, and Tanzania) included applying herbicide to the mats, using harvester boats to cut the mats, and releasing specialized herbivorous insects. Two weevils, *Neochetina eichhorniae* and *N. bruchi,* known to be specialists on water hyacinth from earlier work in Florida, were chosen for release. In 1995 Uganda was first to release biological control insects against the weed, followed by the other two countries in 1997. On the Ugandan shore, weed mats began to show damage and disappear by late 1998. By 1999 some 75% of the mats had died and sunk into the lake. *Neochetina* weevils also produced dramatic results on a water hyacinth infestation in Kenya in only a few months in 1999 (Figs. 1 and 2).

Description of the Process

The following steps are typical of importation biological control projects.

1. Choice of the target pest. There should be broad social agreement that the species chosen as targets of importation biological control are pests and need to be reduced in density. Targets should be species that are strongly regulated by natural enemies in their native ranges, and these species should be missing in the areas invaded by the pest.

2. Pest identification and taxonomy. Correct identification of the target pest is essential. Mistakes at this stage cause project delays or failure. If the pest is an unknown species, its nearest relatives need to be identified, for this information can provide clues to the pest's likely native range.

3. Identification of the native range. The region in which the pest evolved needs to be identified to facilitate the search for specialized natural enemies that evolved with the pest. Several criteria can be used, including the center of the geographic range of the pest, the area where the principal host plant of the pest evolved, regions where the pest is recorded to occur but remains at low densities, and regions with the largest numbers of species closely related to the pest.

4. Surveys to collect natural enemies. Natural enemy collection, or foreign exploration, needs to be done extensively over the range of locations and habitats where the pest is found naturally, and in the proper seasons. Surveys of natural enemies in the invaded area are unlikely to locate effective natural enemies but are needed to identify any natural enemies that may already be present because of their own natural invasion of the region.

FIGURE 1 Water hyacinth infestation at a yacht club in Kisumu, Kenya, May 6, 1999. (Photograph courtesy of Mic Julien.)

FIGURE 2 Reduction of the water hyacinth infestation by *Neochetina* weevils at the yacht club in Kisumu, Kenya, December 16, 1999. (Photograph courtesy of Mic Julien.)

5. Importation to quarantine. Promising natural enemies collected in surveys need to be shipped to quarantine laboratories, where they can be colonized and maintained on the pest for further study.

6. Host specificity and biology studies. To promote selection of safe species for importation, the biology and degree of host specificity of each candidate biological control agent must be determined through a mixture of field observations in the area of origin and laboratory studies in quarantine before release into a new area is approved.

7. Release and colonization in the field. Releases need to be made at numerous locations where the target pest is present, and over extended periods, until efficient means to establish the natural enemies in the invaded area have been discovered or until it is clear that the agents are unable to establish. Once established, natural enemies are further redistributed throughout the range of the pest.

8. Evaluation of efficacy. Field experiments in the invaded area comparing pest density in plots having and lacking the introduced natural enemy are needed to measure the degree to which the natural enemy is able to reduce the density of the pest.

9. Documentation of benefits. Economic and ecological consequences of the project need to be recorded and published.

Extent of Successful Use

Following introductions of natural enemies, pest densities may be reduced, sometimes by 90 to 99% or more. This has been achieved for a variety of kinds of pest insects, including caterpillars, sawflies, aphids, scales, whiteflies, and mealybugs. Over the past 125 years, some 1200 projects of insect biological control have been attempted. Of these, 60% have resulted in a reduction of the pest's density. In 17% of projects, no further controls were needed and control was complete. Introductions of specialized herbivores have been attempted against about 133 species of invasive plants and, of these, 41 species (31%) have been completely controlled.

Economics

Importation biological control is an activity conducted by governments for the benefit of society. Funds for such work are typically provided by governments but may come from grower organizations representing particular crops in a region. Costs of projects are concentrated at the beginning of the work, as costs to search for and study new candidate natural enemies are high. Use of biological control agents of proven value in new locations (where need arises because of the continued spread of the pest into new regions) is cheaper, as much of the initial work need not be repeated and known natural enemies can quickly be introduced. Benefits of successful projects accrue indefinitely into the future, and benefit-to-cost ratios of past projects have averaged 17:1, with some projects reaching as high as 200:1. In successful programs, control is permanent and does not require continued annual investments to sustain the benefits, in contrast to other forms of pest control (e.g., pesticide applications). This makes the method particularly attractive for the protection of natural areas and of crops in countries with resource-poor farmers. Biological control also promotes good environmental stewardship of farmlands in developed countries.

Safety of Natural Enemy Importations

Insects may be released as natural enemies of either invasive plants or invasive insects. Both biological weed control and biological insect control show a very high level of safety to human health and to the health of all other vertebrates. There are three safety issues when insects (herbivores, predators, or parasitoids) are imported to a new region: identification of unwanted contaminants, recognition of organisms damaging to other biological control agents, and potential damage to nontarget species (e.g., native insects or plants) in the area of release by natural enemies with broad host ranges.

The first two safety concerns are addressed by the use of quarantine facilities, which are designed to prevent the unintentional release of new species into the environment following importation. In quarantine, desired natural enemies are separated from miscellaneous insects that might have been accidentally included in the package by the collector, as well as from extraneous plant materials and soil inadvertently sent along.

A taxonomist then confirms the species identification of the organism and ensures that all individuals collected are the same species. Voucher specimens are deposited with an entomological museum for possible future reference. Natural enemy identification indicates either the name of the organism or, sometimes, that it is a species new to science and has not yet been described. New species can usually be placed in a known genus, for which some biological information may exist. A sample of the natural enemies is also submitted to a pathologist to determine whether they carry any microbial or nematode infections. If they do, they are either destroyed or, if possible, treated with antibiotics to cure the infection. This group of field-collected, healthy individuals is then bred in the laboratory on the target host. This series of steps eliminates any undesirable parasitoids (for herbivores attacking weeds) or hyperparasitoids (for insect agents) that might exist in the collected material and, if established, could damage the biological control project by reducing the efficacy of imported natural enemies. For insect parasitoids, rearing for one generation on the target host excludes the possibility that a hyperparasitoid has been obtained by mistake, since such agents typically do not breed on the host itself because they use the natural enemy as nutrient source.

The third safety concern—potential attack on nontarget species after release—requires that scientists estimate the host range of the natural enemy proposed for release and that this information be carefully evaluated as part of the decision whether to release the species from quarantine. For both weed and insect biological control agents, estimation of an agent's host range is based on several sources of information, including the hosts known to be attacked by the agent in the region from which it is collected, any species of interest that occur with the agent in its home range but are not attacked, and data from laboratory tests. For herbivorous insects released for weed biological control, these laboratory tests include studies of the adult's preference for where it lays its eggs, the immature feeding stages' preferences to eat various plants, and the ability of these plants to sustain normal growth of the agent's larvae to maturity. Similar tests can be applied to the study of parasitioids (i.e., both oviposition preferences and survival of the immature stages on a given host). For predators, oviposition preferences may sometimes exist; feeding preferences of both adults and larvae must be measured.

Estimation of host ranges of herbivorous insects used against weeds began in the 1920s, evolving from initial testing of local crops only to a phylogenetically based attempt to define the limits of the host range by testing first plants in the same genus as the target weed, then plants in the same tribe, and finally plants in the same or other families. This process has been highly successful in avoiding the introduction of insects whose host ranges are wider than initially thought. Attacks of introduced herbivores on nontarget plants have largely been limited to other species in the same genus. Also, some attacks were forecast by quarantine studies and judged acceptable by agencies granting permission for release, rather than being unforeseen attacks. Of 117 species introduced into North America, Hawaii, or the Caribbean for biological weed control, only one species (the lacebug *Teleonemia scrupulosa,* introduced into Hawaii in 1902 against the shrub *Lantana camara*) has attacked nontarget plants that were neither in the same genus as the target weed, nor a very closely related genus (for the lacebug, the native shrub *Myoporum sandwicense*).

Estimation of host ranges of parasitoids and predators introduced for biological control of insects began in the 1990s, in response to changing views on the ecological and conservation value of native nontarget insects. Techniques for making estimates of arthropod natural enemy safety are less well developed than those for herbivorous biological control agents. A few examples of harm from parasitoids or predaceous insects to nontarget insects have been reported. Importation of generalist species that have broad host ranges should be avoided because of such potential to harm native insects.

Laws governing biological control importations exist principally in New Zealand and Australia. Laws in the United States regulate importation of herbivorous insects used against weeds but do not currently regulate importation of parasitoids or predators.

NATURAL ENEMY CONSERVATION

Concept of Natural Control

All insects and plants, to various degrees, are attacked by natural enemies independent of any deliberate manipulations by people. Such natural control is rarely sufficient to suppress an invasive species: the local natural enemies lack specialized relationships to the invader, since by definition, the new pest is outside the evolutionary experience of the prey species. For native species, however, natural control may suppress plants and insects below pest levels. Also, for invasive species against which specialized natural enemies have been imported and established, the latter become part of the fauna, providing naturally occurring control. Thus, for all species, apart from invaders not yet subject to natural enemy importations, natural control exists and may be sufficient to suppress such pests adequately for human needs. However, in crops and other artificial landscapes, people can disrupt natural control, particularly with the application of pesticides that kill, sterilize, or repel important natural enemies. Conservation as a form of biological control aims to avoid this loss of natural control either from the use of pesticides or habitat simplification. Sometimes active intervention on behalf of natural enemies to provide them with key missing foods or hosts is necessary.

Effects of Pesticides on Natural Enemies

Before 1947, few synthetic pesticides were used in crops. Most available materials were stomach poisons based on heavy metals such as lead and arsenic, which kill only if eaten. Some botanical extracts, such as rotenone and pyrethrum, both of which quickly degrade in the environment, were also used. After World War II, a business revolution occurred when it became recognized that a variety of compounds that could be artificially synthesized in laboratories were highly effective in killing insects by mere physical contact. Beginning with DDT in 1947, many types of chemicals were marketed to kill insects. One of the undesirable consequences of this change in farming practice was the mass destruction of beneficial insects in crops, resulting in a substantial decrease in natural control. Indeed, insecticides often killed natural enemies more efficiently than they killed the target pest. This unintended consequence was due to the smaller body size, greater relative surface area, and lower levels of detoxification enzymes possessed by parasitic Hymenoptera and other natural enemies, compared with herbivorous pests.

PEST RESURGENCE Occasionally, farmers found that pests for which they applied pesticides were, within a few months, more numerous than they had been before the

application of insecticide. This population rebound has been termed pest resurgence. The steps in resurgence are as follows:

1. The pest population is reduced by the insecticide.
2. The same insecticide application destroys most of the natural enemies that were partially suppressing the pest before the application.
3. Natural enemies are slower to increase in number than the pest after the pesticide residue from the application has degraded to levels unable to kill insects.
4. In the absence of the pesticide and with few remaining natural enemies, the survival and reproductive rates of the pest population increase, leading to higher densities.

In rice crops in Asia, outbreaks of a sucking insect called rice brown planthopper (*Nilaparvata lugens*) were rare before the 1960s. In the 1970s, outbreaks occurred with greater frequency and intensity, as insecticide use increased to control this pest. Research conducted at the International Rice Research Institute in the Philippines demonstrated that this was a classic example of pest resurgence and that pesticide applications were destroying spiders and other generalist predators that were otherwise usually able to suppress rice brown planthopper. As farmers used pesticides more often, outbreaks became larger and more frequent. This phenomenon led to a pesticide treadmill for rice brown planthopper control. A program of grower education was supported by the Food and Agriculture Organization of the United Nations to help rice farmers to understand pest resurgence, natural enemy recognition, and the beneficial role of natural enemies in rice paddies. This outreach program successfully reduced pesticide use on rice crops in Asia, ending a cycle of damaging pesticide use and crop loss.

SECONDARY PEST OUTBREAK A related population process occurs when insecticides applied to suppress a primary pest induce a different species, formerly not damaging, to become a pest. This is called a secondary pest outbreak. In apple crops in the eastern United States, growers must control two serious direct pests of the fruit, apple maggot (*Rhagoletis pomonella*) and plum curculio (*Conotrachelus nenuphar*). These species are most often controlled by repeated application of insecticides to foliage with chemicals that have long periods of residual activity. These applications destroy the parasitoids of leafminers and predators associated with spider mites. Outbreaks of these two foliar pests later in the summer are a direct consequence of grower efforts to control these two key fruit pests.

SEEKING PESTICIDES COMPATIBLE WITH NATURAL ENEMIES To reduce the destruction of natural enemy populations caused by insecticides, there are two potential solutions: using pesticides that have intrinsically selective action or using application systems that are ecologically selective.

Selective Pesticides Three kinds of insecticide have shown the greatest compatibility with natural enemies: stomach poisons, systemic pesticides, and insect growth regulators. Stomach poisons are materials that must be ingested to kill. Materials such as the microbial pesticide. *B. thuringiensis* and some mineral compounds such as kryolite are examples. Pests eating foliage with residues of these materials are killed, but natural enemies walking on treated foliage are not affected.

Systemic pesticides are materials that enter plant tissues and are translocated through the plant. These compounds may be applied to soil and absorbed by roots, or they may move translaminarly into leaves after application to the foliage. Because residues are available only to insects that feed on the crops' tissue or sap, natural enemies resting or walking on plants are not affected.

Insect growth regulators are chemicals that mimic or disrupt insect hormones, preventing normal molting. These compounds kill only when the insect tries to molt. Such materials can be selective if only the pest is likely to be exposed in a susceptible stage. In principle, screening programs could identify specific insecticide–natural enemy combinations in which any contact pesticide might turn out to be selective relative to some particular natural enemy. However, because such materials tend to be rare and screening trials to discover them are costly, only a few are available.

Ecologically Selective Methods of Pesticide Use Manipulation of a pesticide's formulation, timing, or method of application is another method for achieving selectivity in control. Granular formulations of pesticides that fall to the soil, for example, are unlikely to damage natural enemies that forage for hosts or prey on the foliage. Thus, a granular material may be applied at transplant into a cabbage field to protect the roots of young plants from feeding in the soil by larvae of cabbage maggot (*Delia radicum*) without injuring the braconid parasitoids that search the leaves to find and parasitize cabbage aphids (*Brevicoryne brassicae*). More complex methods of separating the pesticide from the natural enemies exist, such as monitoring the emergence of key natural enemies and applying pesticides either earlier or later than the peak activity period of the natural enemy. However, methods that require effort on the part of growers, or are at all complex, tend not to be used.

Loss of Natural Control through Simplification of Crop Fields

Natural control of pest insects and mites in crops has also been reduced by habitat simplification and physical changes in crop plants used in commercial, large-scale agriculture. To sustain their populations, parasitoids need hosts, carbohydrates, and secure places to live that are not subject to insecticide application or physical destruction by plowing, flooding, or fire. Predators need prey and can benefit from or even subsist on alternative nonprey foods such as pollen.

Natural control in crops can be maintained or improved by considering the degree to which these basic necessities of natural enemies are provided within or adjacent to the crop. A few examples illustrate the process.

ADDING POLLEN TO ENHANCE PREDATOR MITES Phytoseiid predatory mites are often important in control of pest spider mites. In some crops, numbers of such phytoseiids may be too low to provide effective control. One approach to increasing phytoseiid numbers is to provide pollen as an alternative food, especially for periods when spider mite densities are low. Levels of pollen on foliage of citrus and other orchard crops may be increased by use of species of trees in windbreaks around orchards or species of grasses as ground covers within orchards that are prolific pollen producers. Effective use of this approach has been made in South African citrus orchards for control of citrus thrips (*Scirtothrips aurantii*) with the phytoseiid *Euseius addoensis addoensis.*

KEEPING USEFUL STRUCTURES ON PLANTS Many plants, such as cotton, have sugar-secreting glands called nectaries both inside and outside of flowers. Many species of natural enemies feed on these sugars. Plant breeding has made it possible to eliminate such nectaries in some crops, and this is sometimes done to deny pests access to the carbohydrate resources. The decision to eliminate or retain necataries needs to be based on studies of the net benefit to pest control of these structures. Plants (e.g., grapes) also often have on their leaves pits or pockets, called domatia, that provide physical refuges for phytoseiid mites. Varieties with domatia often have higher phytoseiid densities and fewer pest mites. Retention of such structures in new crop varieties may be important and should be an explicit part of plant breeding.

ENHANCING SPACES BETWEEN CROP ROWS OR AROUND CROP FIELDS AS REFUGES Natural enemies of some species remain tightly linked to the plant and are little affected by the larger environment. Parasitoids of scales on citrus trees, for example, have all their needs met on citrus trees, provided insecticides are not used and some scales are present year-round for parasitism, host feeding (feeding on host body fluids), or production of honeydew (a sticky carbohydrate waste product produced by homopterans that parasitoids use for food). Other species of natural enemies move about more, passing through the spaces between crops rows, or moving back and forth between crops and noncrop vegetation in uncultivated borders. Species such as spiders and carabid beetles are generalist predators of value in vegetable plantings. However, bare or plowed soil between rows often becomes too dry and hot to favor these predators. Reduced tillage, through greater use of herbicides, or use of cover crops between rows, can enhance populations of these predators. Plants between crop rows, however, must not compete with the cash crop for water or nutrients, or crop yield may be reduced. In cereal crops in the United Kingdom, populations of ground beetles, generalist predators that eat cereal aphids, can be increased by leaving low dikes through fields that are not plowed. These dikes produce perennial grass and herb communities that act as refuges for carabid beetles, which then forage in the cereal plots and consume aphids. Also, in crops that are sometimes treated with pesticides, nontreated patches of noncrop vegetation along crop borders can act to reinoculate crops with natural enemies.

Extent of Successful Use of Natural Enemy Conservation

Natural control is ubiquitous and contributes extensively to pest control in most settings. Conservation of natural enemies through reduction of conflicts with pesticides is a major focus of integrated pest management (IPM) philosophy and practice, and many studies have been conducted that have led to better conservation of natural enemies in crops such as citrus, avocados, apples, and greenhouse tomatoes. Because it is often associated with reductions in out-of-pocket costs, this form of conservation is particularly acceptable to growers, who often are asked to reduce or stop altogether a costly practice (such as applying a pesticide). In contrast, practices that require positive action, such as providing a resource or manipulating vegetation in or near the crop, have been adopted much less often. To be valued by growers, such measures must clearly produce pest control benefits that significantly exceed the costs of undertaking them. Practical use of these ideas presently is limited to organic growers and others who wish to produce crops with little or no use of synthetic pesticides.

Safety

Conservation biological control is universally considered to be a very safe activity. Measures to reduce insecticide use, or to convert to selective or compatible materials, both reduce risks to people working on or living near farms and minimize environmental contamination.

BIOLOGICAL CONTROL THROUGH AUGMENTATION

Pros and Cons of Augmenting Natural Enemies

Entomologists and farmers, working together, have developed methods to rear some species of predators and parasitoids that attack pest insects. This approach of deliberately rearing natural enemies and releasing them against target pests has been applied against insects and mites of both greenhouse and outdoor crops.

The use of this practice in greenhouse-grown tomatoes was begun in the 1920s with the rearing by English growers of *Encarsia formosa,* a parasitoid of the greenhouse whitefly

(*Trialeurodes vaporariorum*). This control program died out because of grower use of pesticides. Biological control was revived in the 1970s by Dutch greenhouse tomato growers because whiteflies had developed resistance to pesticides. In greenhouses that are closed up against the cold early in the crop cycle, natural enemies may be scarce or absent. Augmentative biological control was seen as a way of correcting this natural enemy absence. Natural enemy rearing for the greenhouse industry started when one grower began producing natural enemies for his own use, but soon he was selling surplus parasitoids or predators to other growers, and the operation became a separate business (an insectary). From 1970 to 2000, the number of commercial insectaries grew from just a few to several dozen firms, which collectively produce about 100 species of natural enemies for sale. A few species (mainly the parasitoid *E. formosa* and the predatory mite *Phytoseiulus persimilis*), however, make up most of the sales. Today, a variety of natural enemies are used in indoor settings that include greenhouses, plant conservatories, mushroom houses, and animal holding buildings such as dairies, hog-rearing facilities, poultry barns, and zoos.

Outdoor releases of several species of predators and parasitoids are regularly made by growers in various countries. Egg parasitoids in the genus *Trichogramma* (Hymenoptera: Trichogrammatidae) have been used extensively throughout the twentieth century to suppress pest weevils and caterpillars in cotton, corn, and sugarcane, especially in China, Russia, and tropical sugar-producing countries. Predators of mealybugs for release on citrus crops in parts of California have been reared by a growers' cooperative since 1926. One of the more common current uses of augmentative biological control on outdoor crops is the release of various species of predatory phytoseiid mites for control of pest spider mites, an approach that has been used most often with strawberries and with foliage plants grown outdoors in shade houses.

There are two different approaches to augmentative biological control. Most indoor releases of natural enemies intend only to seed the crop with a founding population of the natural enemy, which then reproduces and eventually suppresses the pest after its numbers have increased naturally in the crop. This approach is called inoculative biological control. Cost of this approach is minimized because smaller numbers of the natural enemy are needed. In contrast, with inundative biological control, an attempt is made to release enough natural enemies to control the pest immediately. Because much higher numbers are released, this approach is economical only against natural enemies with very low production costs, and use has been most successful on crops with a high cash value per hectare.

How Insectaries Turn Natural Enemies into Mass Market Products

To profitably market a natural enemy, an insectary must succeed in a series of activities.

1. Find a suitable natural enemy. Commercial augmentative biological control starts with the discovery of a natural enemy that research suggests may be effective. The natural enemy must attack an important pest efficiently, be able to be reared under mass production conditions, be easily harvested and able to survive transit stress, and be competitive in price with other forms of pest control available to growers.

2. Develop a mass rearing system. To commercially produce a natural enemy, insectaries must be able to make a financial profit on the species. Successful production systems vary. For some species, such as whitefly parasitoids, production can use natural hosts on their favored plants. *E. formosa,* for example, is reared in greenhouse whitefly produced on tobacco plants. Similarly, the important predatory mite. *P. persimilis* is grown on the spider mite *Tetranychus pacificus* on bean plants in greenhouses. In other examples, costs of production or the scale of production are improved by rearing species other than the target pest. Most *Trichogramma* wasp species are grown on the eggs of moths that feed on stored grain, rather than on eggs of the target moths themselves, because colonies of grain-feeding moths can be reared much more cheaply, allowing the production of *Trichogramma* in huge numbers at low cost.

3. Develop harvest, storage, and shipping methods to get the product to customers. Most predators and parasitoids must be used within a few days or weeks of production. For some species, induction of an arrested state called diapause can be used to store immature parasitoids inside parasitized hosts for months. Shipping to customers must use rapid transport (1–3 days) and avoid delays at international borders. Longer delays invariably result in the deaths of natural enemies due to heat, desiccation, continued development, or starvation.

4. Provide clear instructions on effective release methods and rates for customers. The final step in the effective use of natural enemies reared in insectaries is their release by the farmer at the right rate and in the correct manner. Effective rates are discovered by controlled trials in universities and government laboratories, and by ascertaining the experience of growers who have used products in accordance with advice from producers.

Extent of Successful Use

INDOOR CROPS The use of augmentative biological control has become widespread in greenhouses in northern Europe and Canada that produce vegetables, with over 5000 ha using *E. formosa* for whitefly control and over 2800 ha using *P. persimilis* for spider mite control. These amounts, however, still represent only a small percentage of the world's protected culture because these biological control agents are used much less often in southern Europe and Japan, areas with extensive greenhouse vegetable production but with differences in temperatures and open rather than closed greenhouses. Similarly, use of biological control is very

limited in greenhouses producing bedding plants or floral crops, the major focus of greenhouse production in the United States.

OUTDOOR CROPS The scientific use of augmentative natural enemy releases in outdoor crops is best established in northern Europe for control of European corn borer (*Ostrinia nubilalis*) in corn. Use is greatest in Germany and France, where over 3200 ha is protected annually with *Trichogramma* releases. This fraction is, however, small compared with the total corn acreage in Europe, and use of biological control is concentrated principally where pesticide use is not allowed because of concern for health of people living near cornfields. Natural enemy releases for mite control have been successful in strawberries in California, Florida, and the northeastern United States, and in outdoor shade houses used for production of foliage plants in Florida. In Mexico, Russia, China, and other countries, large-scale releases of *Trichogramma* spp. have been made for a variety of moth and beetle pests of corn, sorghum, and cotton, but the efficacy of these releases has not been well demonstrated. Some of these activities have been state supported, and their actual economic value for pest control is not clear.

Safety

Release of parasitoids and predators replaces pesticide application and thus enhances human safety. For workers in insectaries, handling of large quantities of insects or mites constitutes an allergy risk. Where problems arise, risk can be reduced through air exchange or filtration to reduce concentrations of airborne particles and through use of gloves and long-sleeved shirts to reduce skin contact with arthropod body fragments. Risk to native species posed by releases of nonnative natural enemies can be of concern, as well. Generalist, nonnative species released in large numbers may establish outdoors and attack or suppress populations of native species, or they may reduce densities of native natural enemies through competition for resources. Consequently, some governments, such as those of Hawaii, Australia, and New Zealand, restrict importation of natural enemies used in augmentative biological control. For example, importation of North American green lacewing species (Neuroptera: Chrysopidae, *Chrysopa* spp.), used in greenhouses as predators of aphids, might lead to establishment of such species in the wild, increasing competition with the endemic native lacewings in Hawaii, which have conservation value as unique native wildlife.

MICROBIAL PESTICIDES

Using Microbes as Tools

Insects suffer from diseases caused by pathogens of several kinds, including bacteria, viruses, fungi, nematodes, and protozoa. Sometimes natural outbreaks of disease occur that locally and temporarily influence the density of pest populations. Microbial control seeks to use pathogens as tools to suppress pest insects. This process involves finding pathogens able to kill pest species of concern, followed by development of methods to rear pathogens economically. Methods must be developed to store the pathogen's infective stages without loss of viability and to apply the pathogen to the target in ways that result in high rates of infection and thus control. Details of the biology of each pathogen and the effects of environmental conditions on infectivity after application are crucial in determining whether any given pathogen can be used effectively as a microbial pesticide.

Bacteria

Many of the bacteria that infect insects are lethal only in stressed insects because the bacteria, lacking effective means of escaping from the host's gut after ingestion, are unable to enter its body cavity. Species in the genera *Bacillus* (*B. thuringiensis, B. sphaericus,* and *B. popililae*) and *Serratia* (*S. entomophila*) are the main bacteria that have been used as microbial pesticides. Of these, only *B. thuringiensis* has been widely successful. This species produces toxic crystalline proteins inside its spores. Crystals from different strains of this bacterium vary in their ability to bind to the gut membranes of different species, thus shaping the host ranges of each subspecies of the pathogen. If crystals are able to bind to the gut membranes, these tissues are degraded, allowing bacteria to penetrate the body cavity and kill the host.

Strains discovered in the 1920s infected only some species of caterpillars. Later, new strains were discovered that were able to infect mosquito larvae, chrysomelid beetle larvae (such as the Colorado potato beetle, *L. decemlineata*), and scarabs (such as the Japanese beetle, *Popillia japonica*). Commercial use of this pathogen is possible because it can be successfully mass-reared in fermentation media without any use of living hosts. This makes its production inexpensive. Applications of *B. thuringiensis* have advantages for use in forests, where residues of conventional pesticides are objectionable because of potential harm to native wildlife, and in IPM programs in crops where conservation of natural enemies is desired. *B. thuringiensis* is compatible with most natural enemies because it must be ingested to have any effect and because its toxic proteins are selective in their gut binding properties. Genes from *B. thuringiensis* that code for toxic proteins have been isolated and inserted into plants where they are expressed and produce insecticidal proteins in plant tissue and pollen. Transgenic varieties of such major crops as corn, soybeans, and cotton exist and are widely planted in the United States.

Other species of bacteria have had limited commercial use. *B. sphaericus* is formulated for use against some species of mosquito larvae. *B. popilliae* was once commercially produced for use against larvae of Japanese beetle (an important

pest of turf), but this bacterium must be reared in living host larvae, which has made its production uneconomical. In New Zealand, *S. entomophila* causes an infection known as amber disease in a native turf grub (*Costelytra zealandica*), and its commercial use is being promoted. As with *B. thuringiensis,* the ability of *S. entomophila* to be reared in fermentation media apart from living hosts has been a key feature in promoting its commercial use.

Fungi

Species of Deuteromycotina fungi in several genera, including *Beauveria, Metarhizium, Verticillium,* and *Hirsutella,* infect insects and can be grown on fermentation media in solid culture. The spores of these species, when applied, adhere to the bodies of insects, and special hyphae from the spores use enzymes and mechanical pressure to penetrate through the insect's cuticle to cause infection. Infection requires spore germination, a process that is sensitive to environmental conditions. In general, many fungal strains or species require a minimum number of hours (often 12–24) of high relative humidity (often above 80%) to germinate. However, these requirements vary within species and among isolates from different locations and hosts. Spore germination requirements, if not met, can lead to control failures. Successful commercial use of entomopathogenic fungi has focused on ways to either meet these requirements by manipulating the formulations of the product applied (e.g., adding oils when used in arid climates), using these products in inherently favorable climates (e.g., greenhouses), using them in favorable habitats (e.g., soil), or finding strains or species with less stringent environmental requirements for spore germination. Commercial use of these fungi is also affected, but not prevented, by the inability of most species of fungi to produce spores under water. This prevents the use of liquid culture methods, requiring the use of solid media (like boiled rice) or a diphasic system in which mycelial growth takes place in liquid culture, followed by plating out of fungi on solid media for spore production as a second production step.

Successful use of microbial pesticides based on fungi has been achieved by an international consortium (LUBILOSA) in Africa, which has developed the fungus *Metarhizium anisopliae* var. *acridum* (Green Muscle) for control of locusts in Africa. This locust control project is highly beneficial to the environment because this selective, naturally occurring fungus replaces the use of highly toxic, often persistent, pesticides such as dieldrin. Field trials in a number of African countries have demonstrated both high levels of efficacy and costs competitive with the use of conventional pesticides (about U.S. \$12 ha^{-1}). Success in this effort involved screening over 160 isolates of fungi to find the best fungus and the development of formulations for both storage (without refrigeration in hot climates) and application. Field trials demonstrated high initial levels of mortality and pathogen recycling, leading to persistence of suppression. Similar success of fungal pesticides in general agriculture has not yet occurred. In the United States, for example, only one species, *Beauveria bassiana,* is commercially available, and its use is extremely limited.

Nematodes

Nematodes are multicellular organisms as opposed to unicellular microbes, but they are formulated and applied like microbial pesticides. Nematodes in more than 10 families infect insects, but only those in the families Steinernematidae and Heterorhabditidae have been commercially employed for insect control. These species, unlike those in other families, can be reared in fermentation media apart from living hosts. Techniques for large-scale production in liquid broths containing ingredients from dog food can be used to rear about six species in these families. Entomophagous nematodes actively penetrate insect hosts through the insect integument or natural body openings (spiracles, mouth, anus). Once inside the insect body cavity, the nematode defecates specialized bacteria that it carries symbiotically. These bacteria (in the genera *Xenorhabdus* and *Photorhabdus*) quickly kill the host with toxins. Nematodes then reproduce as saprophytes in the decaying host tissues. Entomopathogenic nematodes are sensitive to desiccation, which has limited their use in pest control. Applications made to dry foliage are ineffective because nematodes usually die before encountering hosts. Successful use of these nematodes has been limited to control of pests in moist habitats, such as fungus gnats and scarab grubs in soil, and lepidopteran borers in plant stems.

Viruses

Insects are subject to infections by viruses in a number of families. However, only those in the highly specialized Baculoviridae have been considered for use as microbial pesticides. Viruses in this family infect only insects and are very safe to people and wildlife. However, all viruses are obligate parasites of living cells, and none can be grown in fermentation media. Currently, they are produced in live host insects, which themselves must be mass-reared. This makes viral products relatively expensive, although the governments of some countries, notably Brazil, have promoted their use. A further aspect of the biology of viruses is their high level of host specificity. Extreme specificity of viruses reduces the economic value of products because they kill very few species of pests. Because of these economic factors, no virus products have been economically successful in the United States or Europe, although a few have been developed and briefly marketed.

Safety of Microbial Pesticides

In the United States and many other countries, microbial preparations (but not nematodes) sold for pest control are considered to be pesticides that require government approval and product registration before sale. Requirements for

registration have been modified to reflect differences between chemical and microbial pesticides. Manufacturers are required to specify the exact identity of the microbe in their products, the production process, including controls to prevent contamination, and safety data on infection and allergenic properties of the pathogen and the product as a whole. The safety record to date suggests that risks from such products are either nonexistent or too low to detect.

Degree of Use

Except for genetically transformed plants that express the *B. thuringiensis* toxin (which are not microbial pesticides, but a related development), microbial pesticides are niche products. In no control programs have microbial pesticides widely displaced synthetic pesticides from pest control markets. *B. thuringiensis* is the most widely used organism, but *B. thuringiensis* products represent 1 to 2% of the pesticide market. These products do, however, have important value as pesticides because they are more readily incorporated into IPM programs that include natural enemies.

FUTURE USE OF BIOLOGICAL CONTROL

Biological control can be implemented through four different approaches: conservation of existing natural enemies, importation of new species for permanent establishment, temporary natural enemy augmentation, and use of microbial pesticides. The first two methods are most widely applicable and have produced the greatest benefits. Conservation biological control is the foundation of all insect control. Importation biological control is the method that is appropriate to combat exotic invasive pests (whose numbers are large and increasing). Augmentative biological control is limited by cost factors and largely restricted to high-value crops in greenhouses. Microbial pesticides are niche market tools useful in IPM programs but are limited by high production costs or the narrow host ranges of the pathogens. Biological control's greatest strengths are in public sector applications (conservation, importation) rather than private sector approaches (augmentative, microbial pesticides). Expanded use of biological control will require increased commitment of public resources and recognition that publicly supported programs are more effective for biological control implementation.

See Also the Following Articles

Agricultural Entomology • Genetically Modified Plants • Host Seeking by Parasitoids • Integrated Pest Management • Physical Control of Insect Pests • Predation

Further Reading

Barbosa, P. (ed.) (1998). "Conservation Biological Control." Academic Press, San Diego.

Bellows, T. S., and Fisher, T. W. (eds.) (1999). "Handbook of Biological Control: Principles and Applications of Biological Control." Academic Press, San Diego.

Clausen, C. P. (ed.) (1978). "Introduced Parasites and Predators of Arthropod Pests and Weeds: A World Review." Agricultural Handbook 480. U.S. Department of Agriculture, Washington, DC.

DeBach, P., and Rosen, D. (1991). "Biological Control by Natural Enemies." Cambridge University Press, Cambridge, U.K.

Follett, P. A., and Duan, J. J. (eds.) (2000). "Nontarget Effects of Biological Control." Kluwer, Boston.

Gaugler, R., and Kaya, H. K. (eds.) (1990). "Entomopathogenic Nematodes in Biological Control." CRC Press, Boca Raton, FL.

Gurr, G., and Wratten, S. (eds.). (2000). "Biological Control: Measures of Success." Kluwer, Dordrecht, The Netherlands.

Jervis, M., and Kidd, N. (eds.) (1996). "Insect Natural Enemies: Practical Approaches to Their Study and Evaluation." Chapman & Hall, London.

Julien, M. H., and Griffiths, M. W. (eds.) (1998). "Biological Control of Weeds, a World Catalogue of Agents and their Target Weeds." 4th ed. CAB International, Wallingford, U.K.

Pickett, C. H., and Bugg, R. L. (eds.) (1998). "Enhancing Biological Control: Habitat Management to Promote Natural Enemies of Agricultural Pests." University of California Press, Berkeley.

Tanada, Y., and Kaya, H. K. (1993). "Insect Pathology." Academic Press, San Diego.

Van Driesche, J., and Van Driesche, R. G. (2000). "Nature Out of Place: Biological Invasions in a Global Age." Island Press, Washington, DC.

Van Driesche, R. G., and Bellows, T. S. (1996). "Biological Control." Chapman & Hall, New York.

Van Driesche, R. G., and Hoddle, M. S. (2000). Classical arthropod biological control: Measuring success, step by step. *In* "Biological Control: Measures of Success." (G. Gurr and S. Wratten, eds.), pp. 39–75. Kluwer, Dordrecht, The Netherlands.

Bioluminescence

James E. Lloyd and Erin C. Gentry
University of Florida

Light that is produced in a chemical reaction by an organism is called bioluminescence. This "living light" is most commonly produced in tissues or organs within and shines out of the emitter's body, but luminous secretions are produced by some organisms and oozed or squirted out, even smeared on attackers. Chemiluminescence is but one of several forms of light emission collectively known as luminescence, which occurs when atoms of a substance emit photons (packets of light energy) as their electrons return to their stable state after being lifted to a higher and unstable energy level by input energy.

The best-known insect bioluminescence is that of beetles of the family Lampyridae. They are known as fireflies, lightningbugs, blinkies, and many other local and colloquial names around the world.

TAXONOMIC OCCURRENCE (PHOTIC BIODIVERSITY)

Bioluminescence occurs "everywhere" among organisms, with self-lighting species appearing in all kingdoms of a

four-kingdom classification—Monera, Fungi, Plantae, and Animalia—in 11 of 29 phyla. In the phylum Arthropoda, luminescent forms also are found among the sea spiders, crustaceans, millipedes, and centipedes. Self-luminescent insects occur in the families Poduridae and Onychiuridae (Collembola), Mycetophilidae (Diptera), and coleopterans Lampyridae, Phengodidae, Elateridae, and Staphylinidae and, possibly, the Telegeusidae.

CHEMISTRY AND EVOLUTIONARY ORIGIN

Bioluminescence chemistry varies widely among organisms. Bacteria use riboflavin phosphate, the sea pansy uses diphosphoadenosine, and fireflies use adenosine triphosphate (ATP) in the oxidative decarboxylation of substrates generically known as luciferins, with enzymes termed luciferases. The present, cautious conclusion would be that bioluminescence has evolved from many separate biochemical origins.

Molecular structures and their alterations along light-producing pathways of some systems are illustrated in general references, but many systems have not been investigated. Many use oxidative mechanisms that involve two major stages: the first creates an energy-rich molecule ("large energy quantum") by combining molecules, the second then excites a luminescent molecule that unloads this energy as a photon of light when it returns to its stable state. Among insects, photons range in color from an unbelievable bright, ruby red in the headlight of the railroadworm (*Phrixothrix tiemanni,* Phengodidae, Coleoptera) to the demure blue of glowing Appalachian glowworm larvae (*Orfelia fultoni,* Mycetophilidae, Diptera). In twilight-active fireflies, longer wavelengths (orange-yellow), with appropriate filters in the eyes, may be connected with enhancing signal reception against (noisy) backgrounds of green foliage. The different colors are caused by alterations in the amino acid composition of the luciferase that shift the emission peaks.

ANATOMY OF EMISSION

The ultrastructure of the flashing lantern was first seen in the 1960s, when the electron microscope revealed that a miniature and new type of structure, the tracheal end organ, occurred throughout the flashing lantern, and that each microunit was obviously involved in controlling the photocytes associated with it. The light-emitting layer of a flashing lantern is organized into a sheet of rosettes, each with a central channel (cylinder), through which air-supply tubes and nerve trunks pass, and surrounding photocytes, which abut the photocytes of neighboring rosettes (Fig. 1).

The flashing lantern of adult fireflies does not develop (ontogenetically) from the glowing lantern of juveniles. The difference in the light output of these two lantern types is remarkable. On the one hand, larval lanterns require perhaps a full second to reach their full but much lesser intensity, and in

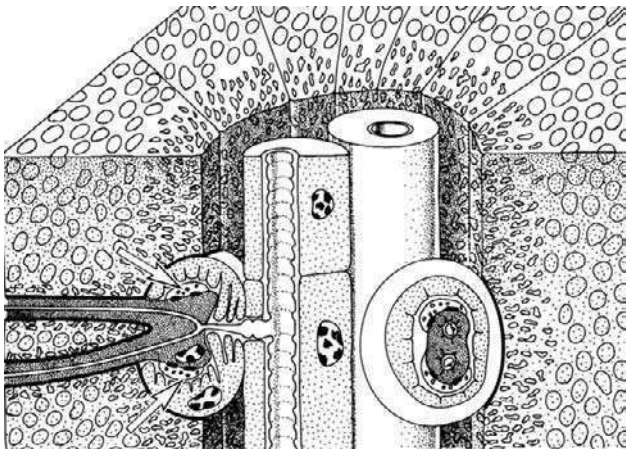

FIGURE 1 Diagram of part of a single rosette ("unit") in the flashing lantern. The central channel (cylinder) is ringed by photocytes, which are differentiated into inner and outer regions. Within the cylinder are two tracheal trunks and two tracheal end organs, one cut longitudinally, one transversely. Note within the end organ the intimate contact of the air supply, nerve ending (arrows), tracheal end cell, and tracheolar cell (the last two indicated by lightly and densely stippled structures, respectively). Original "magnification" approximately 1800×. (Reproduced, with permission, from H. Ghiradella, 1983, Permeable sites in the firefly lantern tracheal system: Use of osmium tetroxide vapor as a tracer. *J. Morphol.* **177**, 145–156. See Ghiradella 1998.)

an array of lantern types their behavior is little removed from the granules in subdermal fat cells or excretory tubules that glow continuously or in a simple circadian (daily) rhythm *(Keroplatus, Orfelia),* which is perhaps controlled by changes in hormone level. On the other hand, a flashing lantern is capable of photic finesse that can be appreciated only with electronic detector systems—the flashes of male fireflies of some species have very sharp *on* transients, and field recordings of flying males reveal that a flash can reach its bright peak in 20 ms (Figs. 2A and 2B), the flicker signal of a Florida *Photuris* species is modulated up to 45 Hz (Fig. 2C), and the four subliminal peaks of what appears to the eye to be a single flash of an Andean Mountain *Photinus* occur at 25 Hz (Fig. 2D).

The triggering of such light emission is currently thought to be connected with the release or gating of oxygen into the photocytes. This occurs in response to patterned volleys from the central nervous system, delivered by neurons that connect to or are closely associated with other key elements within the tracheal end organ (Fig. 1). A recent study suggests that nitric oxide gas plays a key role in the release of oxygen into the photocytes.

DESCRIBING/QUANTIFYING EMISSIONS

Humans have observed and written about light-emitting insects for more than 2 millennia, but early in the 20th century they began to give careful scientific attention to different colors and forms of firefly emissions, noting glows, flashes, flickers, tremulations, scintillations, and so on, and they borrowed descriptive terms from other senses, such as

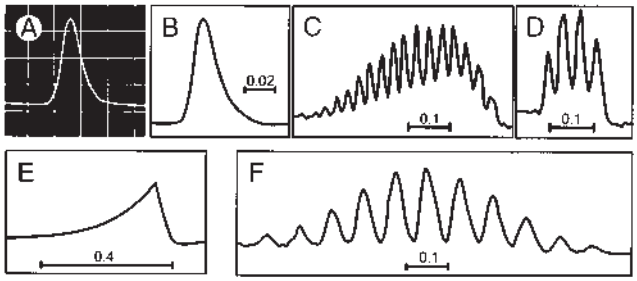

FIGURE 2 Flash patterns of fireflies as displayed for examination and measurement from electronic (frequency-modulated) field recordings, as described in the text, with standard display convention: horizontal axis, time; vertical axis, relative intensity (not total photon flux, which would be virtually impossible to measure from the entire surface of a flying light organ, but intensity change through time as detected from a single position in space, as visible to another firefly). Bars show time scale in seconds; horizonal grid in A, 0.02 s. (A) Nearly symmetrical flash of *Photuris salina,* photographed from the CRT screen of a storage oscilloscope. (B) Same flash scanned from ink tracing of a chart recorder. (C) Crescendo flash of *Photuris* unnamed species "D", with subliminal (to human eyes) modulations at the light organ. (D) Flash pattern of a Colombian *Photinus,* with subliminal modulations. (E) Crescendo flash of *Photuris cinctipennis.* (F) Visibly flickering flash pattern of a New Guinea *Luciola* species.

"crescendo" and "ramp" (Fig. 2E)—both describing a flash that begins dimly and gradually rises in intensity.

The simplest form of light emission is a glow of indeterminate length, as produced by fungus gnats, some Collembola, beetle larvae, adult fire beetles (elaters), phengodid beetle larvae and adult females, and lampyrid glowworm firefly females. A useful description of glows requires only a statement of apparent brightness (distance visible) and, cautiously, apparent color. In contrast, an adequate description of many adult lightningbug emissions often requires a chart, such as first published by Frank McDermott in 1914, with relative-intensity/time on the axes and notes of variations among flash patterns, from pattern to pattern, and at different ambient temperatures.

In the 1930s electronic technology, with photocells, string galvanometers, and ocillographs, made it possible to electronically chart firefly flashes in the laboratory. A generation later photomultiplier-tube systems permitted recording of the flashes of fireflies flying in the field. With today's miniaturized solid-state, digital circuits, detectors, and tape recorders flash detecting/recording systems can be hand-held. The flash patterns shown in Fig. 2 were all recorded in flight in the field, with a photomultiplier tube as detector, whose output was frequency modulated (fm) to encode intensity information; the fm conversions were recorded on magnetic tape. In the laboratory the recorded fm patterns were demodulated and fed into an oscilloscope (Fig. 2A) and chart recorder (Figs. 2B to 2F), for display and measurement. Video cameras add yet another dimension to observation and flashing behavior analyses.

There is no evidence that bioluminescent insects make use of color discrimination—it should be noted that the color a human observer perceives in the field often errs because of the intensity of the luminescence, its background (sky, street) lighting, and the degree of dark adaptation of the viewer's eyes. For example, dim green light may appear white because color vision apparatus (cone vision) is not stimulated, and the yellow flashes of roadside *Photinus pyralis* may appear green when under a sodium-vapor streetlight.

BEHAVIORAL ECOLOGY

Behavioral ecology is the analysis of behavioral features of the phenotype as ecological and reproductive adaptations. Insect bioluminescence offers remarkable opportunities for applying the "adaptationist's program" of behavioral ecology, observation, speculation, systematic observation, and experimentation, in the laboratory and field. The experimenter can enter these informational transactions with a penlight or computer-driven light-emitting diodes. Further, interactions often occur quickly and can be photographed, videotaped, and electronically recorded for precise analysis.

The mating signals of lightningbug fireflies are the most commonly seen example of insect luminescence, but others are easily found if sought in their habitat: prey-attracting glows of larval Appalachian glowworm flies *(O. fultoni)* in beds of impatiens at roadside springs and under overhanging mossy banks of streams along dark mountain roads; glowing *Arachnocampa luminosa* larvae hanging from ceilings of New Zealand caves, attracting midges from streams below and tourists from around the world; and prey-attracting glows of larval termitophageous click beetles *(Pyrearinus termitilluminans),* which make termite mounds look like high-rise buildings of a metropolis seen from the air, in the dry-scrub region (open-formation cerrado) of northeastern Brazil.

The significance of many luminosities remains problematic: why do sparkling, galaxy-like arrays of flashing Collembola result when rotting forest litter under damp logs is scratched with a hand cultivator? Is each individual, when stimulated by our invasive touch, warning relatives, or a predator, say, a firefly larva? Several firefly knowns and unknowns are illustrated in Fig. 3, in which black circles, teardrops, stripes, and beads represent emissions of different forms. Coordinates on the axes guide attention to specific locations in the scene—near 3M an armadillo views a flashing firefly under its nose, perhaps retrieving memory data that flashing lights taste terrible or, previously when eaten, vomiting followed. The blood of some fireflies has been found to have cardioglycosides that can be deadly for some animals. (Recently, to the regret of pet owners, several exotic lizards died after eating North American fireflies.) This fact makes a warning (aposematic) function of luminosity a testable explanation.

The most often seen and best understood bioluminescent emissions are the mating flashes of lightningbug fireflies. Nearly all flying emissions seen afield are the mating signals of males, signaling over and over, advertising, with their species' sexual-recognition flash patterns. The male flash patterns of many species are distinctive and diagnostic in a

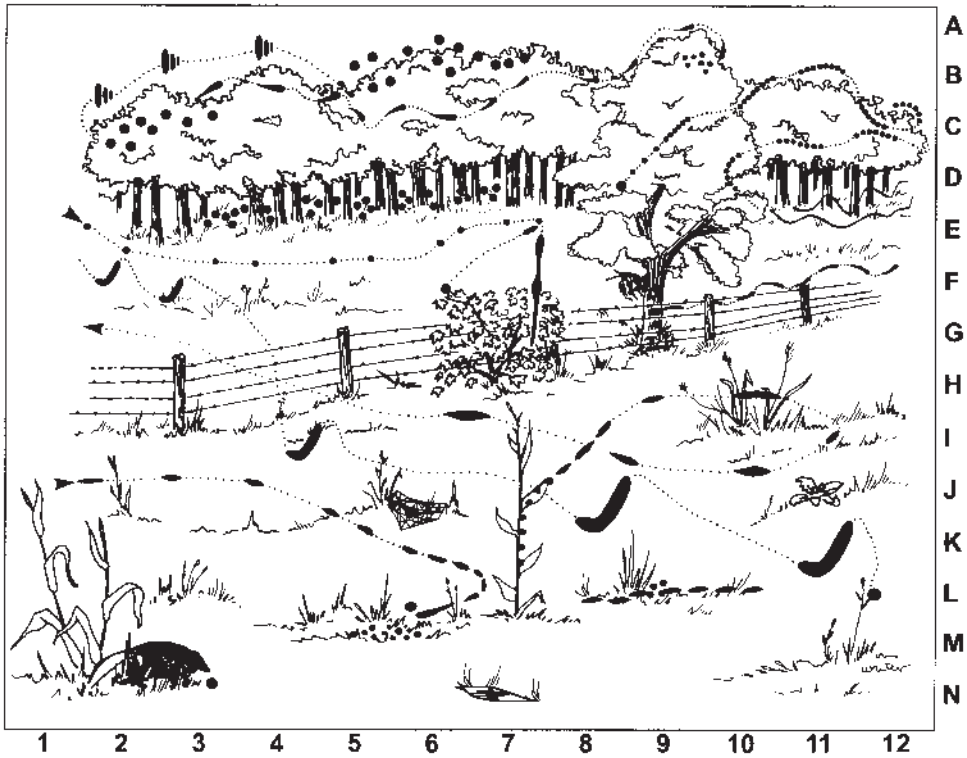

FIGURE 3 A graphic scene of firefly emissions, with alphanumeric coordinates to index text discussion. The diagram illustrates flash patterns, response flashes, a warning flash (3N), an attack flash (7E), illumination flashes, a flash pattern default (9D), some (probably?) meaningless flashes of "stressed" fireflies (6J, 7N, 9F), and a *Photuris* larval glow (8L–10L).

given location. In Fig. 3 several examples are illustrated. A male *Photinus macdermotti* flies from 2E to 7E, repeating his two-flash pattern every few seconds; a *Photinus collustrans* presents several of his low arcing flash patterns behind the fence 9F–12F; *Photuris frontalis* males fly low in the woods beyond the fence at 3E–7D; a *Photinus pyralis* repeats his J-stroked flash pattern diagonally across the view between 1F and 11K, and three or possibly four *Photuris* species cruise over the forest canopy where high-flying one-short-flash patterns are often difficult to identify without attracting the emitters to view them in the hand, but the crescendo flash at 3B–9A is that of *Photuris lucicrescens.* Low in the woods at the upper right (10E–12D) two glowing males of the glowworm firefly *Phausis reticulata* cruise low over the ground emitting their green light continuously.

In the signal system used by most North American lightningbugs, perched females flash responses to the flash patterns of males of their own species. The response signals of females are single flashes in most cases, emitted after a slight delay (<1 s) and, after a brief dialogue of flash patterns and responses, males reach and mount answering females. However, the female delay in the common *Photinus pyralis* is distinctive and varies between 2 and 4 s, depending upon temperature. A female *Photinus pyralis* is answering her male from a perch up a spike of grass near 12L.

After mating, males return to mate searching and females turn to ovipositing and/or hunting *(Photuris);* this explains why males nearly always greatly outnumber females during their mating-time window, although the sex ratio at egg laying and at adult eclosion is expected to be 1:1. The operational sex ratio is reversed in some species toward the end of the season. Under such circumstances males theoretically are expected to become the more discriminating (i.e., the choosier) of a courting pair. Individuals may receive information that influences their flashing and mating behavior from the number of other flash patterns they observe around them.

Among the dangers that flying, signaling males encounter in the dark are the predaceous firefly females of the genus *Photuris.* These versatile *femmes fatales* mimic the flash responses of females of other species, attract males, and eat them (Fig. 4; termed variously aggressive, Batesian–Wallacian,

FIGURE 4 A predatory female of the genus *Photuris* (Florida member of the *Photur. versicolor* complex) devouring a male *Photinus tanytoxus* she has attracted with false mating signals. When hunting females of this species are presented with simulated "hovering" male flash patterns they sometimes launch aerial attacks.

or Peckhammian mimicry); they also launch aerial attacks on flying, flashing males, aiming at their luminescent emissions. In Fig. 3, the flash pattern of the *Photinus macdermotti* male is answered (6F) but he flies on, then is attacked in the air (7E) by the *Photuris* female. Such predators may be expected to have had a strong evolutionary impact on both the signal coding (countermeasure trickery) of prey species and the flight paths and bioluminescent signals of mate-seeking males. Predaceous females obtain the defensive steroidal pyrones called lucibufagens from some firefly prey, which they incorporate into their own and their eggs' defenses.

The hawking *Photuris* female (7E) flashes just as she reaches the male, and they fall glowing into the shrub; possibly illuminating her attack so she could seize him. *Photuris* females use what appear to be illumination flashes in other situations: as they approach the ground or vegetation for landing (2J–6L); as they climb vegetation and take flight, when their flashes gradually become less frequent, then stop altogether as the females disappear into the night (7L–6I); and as they walk around in tangles of vegetation on the ground (5M–6M, 9L), perhaps seeking oviposition sites. Other flashes of unlikely if any behavioral-ecological significance are those emitted by fireflies in spider webs (6J), water puddles (7N), and tangles of Spanish moss (9F).

Males of many *Photuris* species use two or more flash patterns during mate search. Several species in one species group use a flicker with modulations timed exactly like those in the flash pattern of *Pyractomena angulata* (similar to the flicker shown in Fig. 1F), in addition to their species' own identification flash pattern. For example, *Photuris tremulans* males usually emit a single short flash every 2 s (20°C; Fig. 3, 2C–3C), but occasionally several or all males in a local population will emit the described *Py. angulata* flicker pattern (except it is green, not amber). When a short-flashing *Photuris tremulans* male is answered with a short flash, he approaches the respondent (female or penlight), maintaining a dialogue, and lands near; but, when a flickering male is answered, he switches (defaults) to his short flash pattern and then approaches as described (11C–9D).

When males of a related *Photuris* species receive an answer to their *Py. angulata* flicker (supernumerary) pattern, they default to their species' identification pattern, which is a pulsed pattern like the one seen above the trees at 2B–4B, or a variation of it. This firefly demonstrates another twist in the use of supernumerary flash patterns that will be important for understanding bioluminescent communication. For a few minutes at the onset of evening activity all males begin with the pulsing pattern but soon some, then more and more of them use their *Py. angulata* flicker pattern, until in an hour or so a peak of 50 to 80% flickering is reached; then, the proportion of males using the flicker pattern gradually decreases across the midnight hour.

Contrasting behavior of *Photuris lucicrescens* may be helpful for understanding signals of other *Photuris* species. They also use two flash patterns, a short flash and a crescendo flash (5B–7B), but they apparently do not switch patterns (default) during approach, and a nocturnal changeover has not yet been recognized. There is much to be learned about the behavioral ecology of *Photuris* fireflies and, in particular, the sexual selection aspects of their bioluminescent signals (Fig. 5). Note that the supernumerary patterns of several *Photuris* species are copies of flash patterns of known prey of the males' own females. In other words, pattern-copying *Photuris* males use a signal that hunting conspecific females will answer. This suggests a possible evolutionary origin of these prey-mimicking mate search patterns. Without this information the behavioral ecologist cannot understand selection pressures that have influenced the evolution of the bioluminescent signals we see today.

FIGURE 5 Male fireflies of two *Photuris* species flashing over an oldfield in Connecticut seeking mates. One species is responsible for nearly all of the patterns seen here.

HUMAN APPLICATIONS OF FIREFLY CHEMISTRY

Bioluminescent insects are choice subjects for behavioral studies, but of equal or greater significance is the human-serving use that molecular and cellular biology and medicine have made of firefly bioluminescence chemistry. As one of the few known instances in which ATP is involved in light production, firefly light has been used as a research tool for understanding the machinery of cellular energy conversion and a variety of other applications, including medical diagnoses. ATP, the energy currency of life, is produced, stored, and used by living cells, and each photon of firefly light requires the energy released "from" (the terminal phosphate of) one ATP molecule. Light is easily quantified in the laboratory at extremely low levels by photo detectors (luminometers) and recorders that do not intrude into an ongoing living or chemical systems. Thus, photon monitoring (of ATP) in real time can reveal even subtle variations in reaction rates and enzyme kinetics under different experimental conditions.

Glow intensity of an extract or culture will vary directly as ATP is produced or used ("up") by a cellular reaction and can reveal the diurnal rhythmicity, quiescence or torpor, and health and vigor of living tissue, the last being of particular interest when exploring a bacterial culture's sensitivity to antibiotic substances. Firefly genes whose products are involved in this reaction have been put into the genomes of bacteria, mammals, and plants.

CULTURE AND LITERATURE

The Aztecs combined science and humanities with a firefly: "'A firefly in the night' the Nahuas called their songs: a tiny light in a great darkness, a little truth within the ignorance surrounding them." It is a culturally impoverished American who has not heard of Wah-wah-taysee, the firefly in a memorized passage from Longfellow's *Song of Hiawatha* or does not know where to find reference to a glowworm in *Hamlet,* or what Robert Frost said about fireflies, or what a glowworm did for the Mills Brothers. There are firefly books in English for children, but nothing to compare with the literature available in Japan, especially in connection with the Yokosuka City Museum and their dynamic firefly program.

See Also the Following Articles

Chemical Defense • Coleoptera • Mating Behaviors • Mimicry

Further Reading

Eisner, T., Goetz, M. A., Hill, D. E., Smedley, S. R., and Meinwald, J. (1997). Firefly "femmes fatales" acquire defensive steroids (lucibufagins) from their firefly prey. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 9723–9728.

Ghiradella, H. (1998). The anatomy of light production: The fine structure of the firefly lantern. *In* "Microscopic Anatomy of Invertebrates," Vol. 11A, pp. 363–381. Wiley–Liss, New York.

Harvey, E. N. (1952). "Bioluminescence." Academic Press, New York.

Harvey, E. N. (1957). "A History of Luminescence." Am. Philos. Soc., Philadelphia.

Herring, P. (ed.) (1978). "Bioluminescence in Action." Academic Press, New York.

Lloyd, J. E. (1979). Sexual selection in luminescent beetles. *In* "Sexual Selection and Reproductive Competition in Insects" (M. Blum and A. Blum, eds.). Academic Press, New York.

Lloyd, J. E. (1983). Bioluminescence and communication in insects. *Annu. Rev. Entomol.* **28,** 131–60.

Lloyd, J. E. (1997). Firefly mating ecology, selection and evolution. *In* "Mating Systems in Insects and Arachnids" (J. C. Choe and B. J. Crespi, eds.), Chap. 10. Cambridge University Press, Cambridge, UK.

Sivinski, J. (1981). The nature and possible functions of luminescence in coleoptera larvae. *Coleoptera Bull.* **35,** 167–180.

Wickler, W. (1968). "Mimicry in Plants and Animals." McGraw–Hill, New York.

Wood, K. V. (1995). The chemical mechanism and evolutionary development of beetle bioluminescence. *Photochem. Photobiol.* **62(4),** 662–673.

Biotechnology and Insects

Bryony C. Bonning

Iowa State University

Biotechnology can be broadly defined to include all practical uses of living organisms. As such, biotechnology has been practiced since the beginning of recorded history through endeavors such as fermentation of microorganisms for production of beer, selective breeding of crops, beekeeping for the production of honey, and maintenance of silkworms for the production of silk. Laboratory techniques developed within the last 20 years that enable transfer of genes from one organism to another have resulted in tremendous scientific and commercial interest and investment in biotechnology. The word "biotechnology" is now commonly used to refer to manipulation of organisms at the molecular level. This article reviews insect-derived tools used for biotechnological research and the use of recombinant DNA technology for management of insect pests and insect-borne disease.

INSECT-DERIVED TOOLS USED FOR BIOTECHNOLOGICAL RESEARCH

Expression of Foreign Proteins in Insect Cells

Production of large amounts of a particular protein is extremely valuable for both research and industrial purposes. Baculoviruses, which are double-stranded DNA viruses that infect mainly insects, have been developed as baculovirus expression vectors (BEVs) by genetic modification to include a gene of interest. BEVs can replicate in lepidopteran cells and larvae, thereby efficiently transferring foreign genes into eukaryotic cells. The foreign gene is usually under transcriptional control of a viral promoter so that the gene is transcribed by the virus, but translated by the host cell biosynthetic machinery. The BEV system is one of the best tools for recombinant protein expression in a eukaryotic host and has been used for the production of many different proteins for research purposes. The BEV system also has potential industrial application for the production of proteins used in vaccines, therapeutic agents, and diagnostic reagents. Advantages of this protein production system include production of large quantities of foreign protein, and eukaryotic protein processing allowing production of more authentic eukaryotic proteins. The BEV expression system is only transient, however, because the baculovirus ultimately kills the host cells. Baculoviruses do not infect vertebrates and therefore provide relative safety for laboratory manipulation. The use of a baculovirus for production of a foreign protein was first demonstrated by expression of human β-interferon and *Escherichia coli* β-galactosidase.

Insect cells can also be engineered directly to express the recombinant protein, without the baculovirus expression vector intermediate. Such insect cells are stably transformed to constitutively express a foreign gene. Expression levels are usually lower than for the BEV system, but stably transformed cells produce recombinant proteins continuously and process them more efficiently than infected cells.

Insect-Derived Genes Used in Biotechnology

Reporter enzymes allow monitoring of gene expression in living tissues and cells. The gene encoding the reporter enzyme is typically inserted under control of the promoter of the gene

of interest, and production of the enzyme is monitored by means of an enzyme assay. Luciferases belong to a unique group of enzymes that produce light as an end product of catalysis. The luciferases derived from the North American firefly *Photinus pyralis* (Coleoptera) and the Jamaican click beetle *Pyrophorus plagiophthalamus* (Coleoptera) have been used as genetic reporter enzymes in virtually every experimental biological system, including prokaryotic and eukaryotic cell cultures, transgenic plants and animals, and cell-free expression systems. These luciferases, which evolved for the nocturnal mating behavior of the beetles, use ATP, oxygen, and D-luciferin as substrates in the catalysis of a light-producing reaction. The ease and reliability with which luciferase can be assayed, combined with the sensitivity of the technique, has made this enzyme a highly valuable research tool.

USE OF BIOTECHNOLOGY FOR MANAGEMENT OF INSECT PESTS IN AGRICULTURE

The ability to move genes from one organism to another has enabled scientists to develop insect-resistant transgenic crops and insect pathogens with enhanced insecticidal properties. The technology also has the potential to protect beneficial insects from chemical pesticides.

Insect-Resistant Transgenic Plants

Despite the progress made in recent years, a significant proportion of the world's food supply is lost to the activities of insect pests. The deleterious impact of chemical pesticides on the environment, combined with the emergence of technologies enabling plants to be transformed with foreign genes, has driven the seed industry to develop transgenic plants as novel, environmentally benign means of pest control. Insect-protected crops were among the first products of biotechnology to have a significant impact on crop protection, and at times their use has resulted in decreased application of classical chemical pesticides.

The bacterium *Bacillus thuringiensis* (Bt) *kurstaki* has served as a microbial insecticide for many years, but widespread use was limited by its instability when exposed to ultraviolet light and its poor retention on plant surfaces in wet weather. The high toxicity of the Bt toxins to a variety of insect pests, and the ease with which the gene could be isolated from bacterial plasmids, made it an obvious choice for development of the first insect-resistant transgenic plants. The active Bt toxin binds to a receptor in cells lining the insect gut and creates a channel allowing free passage of ions. The cells lining the gut die, and very soon, the insect dies, too. Different strains of Bt contain plasmids encoding different toxins with different specificities of action against insects. A particular toxin is generally effective against only a limited range of closely related species. Bt toxins are used in a variety of transgenic crops in the United States, including cotton, for protection against various lepidopteran pests, corn (maize), for protection against the European corn borer *Ostrinia nubilalis* (Lepidoptera), and potatoes, for protection against the Colorado potato beetle *Leptinotarsa decemlineata.*

Industry has expended enormous effort to identify new isolates of Bt. with different specificities and increased virulence for development of insect-resistant crops. Other bacteria also provide a resource for identification of insect-specific toxin genes such as those derived from *Bacillus cereus* and the entomopathogenic nematode-associated bacterium *Photorhabdus luminscens.*

Plants have a variety of strategies to avoid or survive attack by insects, and genes encoding endogenous plant defensive compounds are also candidates for enhancing the resistance of crops to insect pests. Such factors include inhibitors of digestive proteinases that disrupt digestion by phytophagous insects. However, expression of serine protease inhibitors rarely results in high mortality of the insect pest, and the levels of mortality achieved were below what is required for commercial viability. A variety of lectins that bind specifically to carbohydrate residues have also been expressed in transgenic plants for protection against insects. A gene encoding the snowdrop lectin has been engineered into transgenic plants, and it confers protection against a variety of pests, including aphids and planthoppers. As with proteinase inhibitors, the levels of protection conferred by the foreign lectins are not sufficient for commercial viability.

Transgenic Arthropod Natural Enemies

Recombinant DNA methods may be applied to produce improved strains of natural enemies such as predatory arthropods and parasitoids, but techniques are in the early stages of development. For example, the western predatory mite, *Metaseiulus occidentalis* (Acari), is among a group of mites that are mass-reared for the control of spider mites. However, pesticides applied for control of other pest species often wipe out the predatory mites. Engineering beneficial insects such as the western predatory mite with insecticide resistance genes would in theory provide protection from chemical sprays applied for control of insect pest species.

Engineered Insect Pathogens for Pest Control

Insect pathogenic bacteria, viruses, fungi, and nematodes have been used for the management of insect pests in various niche markets. However, each agent suffers from at least one major limitation, such as susceptibility to environmental stresses, temperature extremes, desiccation, or solar radiation. Most work has been done on the genetic enhancement of bacteria and viruses, in part because of the relative ease of genetic manipulation of these organisms. Genetic engineering to enhance the insecticidal properties of entomopathogenic nematodes and fungi is in its infancy.

Genetic engineering has been used to enhance the insecticidal efficacy of various strains of Bt. by increasing virulence,

extending host range, and increasing field stability, and by introducing alternative toxins to facilitate resistance management. Techniques have been developed for production by genetic means of new strains of Bt. with new combinations of toxin genes.

Considerable progress has been made toward optimization of entomopathogenic viruses at the genetic level. The baculoviruses are arthropod-specific viruses that have been studied extensively both as protein expression vectors and as insect pest control agents. These viruses have been genetically engineered with genes encoding insect-specific toxins that are active within the hemocoel of the insect. Upon infection of the insect host, the toxin is produced as the virus replicates, and the infected insect dies from the effects of the toxin delivered by the virus. Baculoviruses have been engineered with genes encoding a variety of insecticidal proteins and peptides. The most effective enhancement of insecticidal properties of the virus was achieved by introduction of genes into the virus genome that encode insect-specific neurotoxins derived from scorpion venom. Recombinant baculovirus insecticides have been developed that now approach the efficacy of the classical chemical insecticides.

USE OF BIOTECHNOLOGY FOR MANAGEMENT OF INSECT PESTS AND INSECT-BORNE DISEASE

Transgenic Insects

Transposable elements are mobile segments of DNA that can move from site to site within a genome and can be used for delivery of foreign DNA into the genomes of insects. Although the vinegar fly *Drosophila melanogaster* (Diptera) was, in 1982, the first organism to be transformed, leading to tremendous advancements in genetics research, the application of this technology to other insects has been slow. Recent successes however indicate that stable transformation of insects may become more routine in the foreseeable future. Transformation using transposable elements has been achieved for relatively few species, mostly within the Lepidoptera and Diptera (Table I). Other gene transfer systems using viruses or gene expression from transformed bacterial endosymbionts (so-called paratransgenesis) have been used for some species that are not amenable to direct transformation. The genomes of bacteria and viruses are also significantly easier to engineer than eukaryotic genomes. Bacteria and viruses have been used as vectors for both transient and stable foreign gene expression in insects. For example, the bacterial symbionts of the kissing bug, *Rhodnius prolixis,* were successfully engineered to reduce the quantity of *Trypanosoma cruzi,* the parasitic protozoan that causes Chagas disease and is carried by this vector. The bacterial endosymbionts were engineered to express an antimicrobial peptide or antibodies that specifically target the parasite. Similar methods are being developed to prevent transmission of the malaria parasite *Plasmodium* by its mosquito vectors.

GENETIC APPROACHES FOR MANAGEMENT OF INSECT PEST POPULATIONS

The sterile insect technique (SIT) relies on release of large numbers of sterile male insects that mate with wild females, thereby reducing reproductive potential or, if sufficient numbers of males are released over time, resulting in eradication of the pest population in a given area. Successful SIT programs have been conducted against the screwworm, *Cochliomyia hominivorax,* the Mediterranean fruit fly, *Ceratitis capitata,* and the tsetse fly, *Glossina* spp. One of the problems associated with SIT is that laboratory rearing and sterilization of males results in reduced fitness of the insects.

Alternative genetic control systems include use of natural sterility such as cytoplasmic incompatibility induced by infection with the bacterium *Wolbachia,* and conditional lethal traits. For a conditional lethal release, insects are engineered to carry a lethal trait that is active only under certain conditions, such as certain temperatures, or at diapause. Since the trait is not lethal immediately, it can spread in a population. Genetic techniques have also been developed that allow induction of female-specific lethality. These autocidal control strategies have been demonstrated only in the model organism *Drosophila* thus far. The ability

TABLE I Genetic Transformation of Nondrosophilid Insects

Order	Species transformed	Common name	Pest status
Diptera	*Anopheles stephensi*		Disease vectors
	A. albimanus		
	Aedes aegypti	Yellow fever mosquito	
	Culex quinquefasciatus	Southern house mosquito	
	Musca domestica	House fly	
	Stomoxys calcitrans	Stable fly	
	Ceratitis capitata	Mediterranean fruit fly	Horticultural pests
	Bactrocera tryoni	Queensland fruit fly	
	B. dorsalis	Oriental fruit fly	
	Anastrepha suspensa	Caribbean fruit fly	
Coleoptera	*Tribolium castaneum*	Red flour beetle	Stored-product pest
Lepidoptera	*Bombyx mori*	Silkworm	None
	Pectinophora gossypiella	Pink bollworm	Cotton pest

to insert the desired genes into insect genomes will be critical to the success of these genetic approaches for management of insect pests in the future.

See Also the Following Articles

Genetically Modified Plants • Genetic Engineering • Pathogens of Insects • Sterile Insect Technique • Wolbachia

Further Reading

Atkinson, P. W., Pinkerton, A. C., and O'Brochta, D. A. (2001). Genetic transformation systems in insects. *Annu. Rev. Entomol.* **46,** 317–346.

Durvasula, R. V., Gumbs, A., Panackal, A., Kruglov, O., Aksoy, S., Merrifield, R. B., Richards, F. F., and Beard, C. B. (1997). Prevention of insect-borne disease: An approach using transgenic symbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.,* **94,** 3274–3278.

Durvasula, R. V., Gumbs, A., Panackal, A., Kruglov, O., Taneja, J., Kang, A. S., Cordon-Rosales, C., Richards, F. F., Whitham, R. G., and Beard, C. B. (1999). Expression of a functional antibody fragment in the gut of *Rhodnius prolixis* via transgenic bacterial *Rhodococcus rhodnii. Med. Vet. Entomol.* **13,** 115–119.

Estruch, J. J., Warren, G. W., Mullins, M. A., Nye, G. J., Craig, J. A., and Koziel, M. G. (1996). Vip3A, a novel *Bacillus thuringiensis* vegetative insecticidal protein with a wide spectrum of activities against lepidopteran insects. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 5389–5394.

ffrench-Constant, R. H., and Bowen, D. J. (2000). Novel insecticidal toxins from nematode-symbiotic bacteria. *Cell Mol. Life Sci.* **57,** 828–833.

Gatehouse, J. A., and Gatehouse, A. M. R. (2000). Genetic engineering of plants for insect resistance. *In* "Biological and Biotechnological Control of Insect Pests" (J. E. Rechcigl and N. A. Rechcigl, eds.), pp. 211–241. CRC Press, Boca Raton, FL.

Handler, A. M. (2001). A current perspective on insect gene transformation. *Insect Biochem. Mol. Biol.* **31,** 111–128.

Harrison, R. L., and Bonning, B. C. (2000). Genetic engineering of biocontrol agents for insects. *In* "Biological and Biotechnological Control of Insect Pests" (J. E. Rechcigl and N. A. Rechcigl, eds.), pp. 243–280. CRC Press, Boca Raton, FL.

James, A. A., Beerntsen, B. T., Capuur, M. D., Coates, C. J., Coleman, J. Jasinskiene, N., and Krettli, A. U. (1999). Controlling malaria transmission with genetically-engineered, *Plasmodium*-resistant mosquitoes: Milestones in a model system. *Parasitologia* **41,** 461–471.

Jarvis, D. L. (1997). Baculovirus expression vectors. *In* "The Baculoviruses" (L. K. Miller, ed.), pp. 389–431. Plenum Press, New York.

Naylor, L. H. (1999). Reporter gene technology: The future looks bright. *Biochem Pharmacol.* **58,** 749–757.

Thomas, D. D., Donelly, C. A., Wood, R. J., and Alphey, L. S. (2000). Insect population control using a dominant, repressible, lethal genetic system. *Science* **287,** 2474–2476.

Blattodea

(Cockroaches)

Donald G. Cochran

Virginia Polytechnic Institute and State University

Cockroaches are an ancient and highly successful form of insect life. They were among the groups of insects that evolved during the first great radiation of insects and have been in existence for at least 350 million years, or since early Carboniferous times. They seem to have achieved an optimum body form and other features early in their evolutionary history. Fossil specimens are relatively abundant; some that are at least 250 million years old are easily recognizable as cockroaches and could pass for modern species. Among the features that allowed them to escape the extinction that claimed many of the earlier insect groups was the ability to fold their wings over the body. This allowed them to more easily hide from predators and escape other dangers. They also evolved early in their existence an ootheca that could be hidden, hence offering some measure of protection for their eggs.

Cockroaches are referred to as generalized orthopteroid insects, which classifies them with the true Orthoptera (crickets, katydids, grasshoppers, locusts), Phasmatodea (walkingsticks), Mantodea (praying mantids), Plecoptera (stoneflies), Dermaptera (earwigs), Isoptera (termites), and a few other minor groups. The phylogenetic relationships among all these groups are not firmly established, although several theories exist. The closest relatives of cockroaches are believed to be the mantids, and some modern taxonomists prefer to place these two groups, as well as termites, in the order Dictyoptera. Indications are that termites evolved out of the cockroach stem or that cockroaches and termites both evolved from a common ancestor. One family of cockroaches (Cryptocercidae) and one extant relic species of termite *(Mastotermes darwiniensis)* have certain characteristics in common. Among them are the segmental origin of specific structures in the female reproductive system and that both deposit their eggs in similar blattarian-type oothecae. They also share a system of fat body endosymbiotic bacteria that is common to all cockroaches but is unique to *Mastotermes* among the termites.

THE SPECIES OF COCKROACHES

Between 3500 and 4000 species of cockroaches have been identified, with one relatively simple classification scheme dividing this group into five families as follows:

- **Cryptocercidae** is the most primitive family and consists of one genus with fewer than 10 species. These cockroaches live as isolated family groups in decaying logs and occur in the United States, Korea, China, and Russia. They are large, reddish brown insects that are wingless at adults.
- **Blattidae** is a diverse family with many genera and hundreds of species. Those classified as *Periplaneta* and *Blatta* are widely distributed, while other genera are more regional. They are large insects that tend to live outdoors. Several species are referred to locally as palmetto bugs.
- **Blattellidae** is also a diverse family with many genera and around 1000 species. These cockroaches are widely distributed in the world but are concentrated in the tropics and subtropics. Blattellids are mostly small outdoor

cockroaches, including those called wood cockroaches. The genus *Blattella* contains the German cockroach.

- **Blaberidae** is the largest family of cockroaches, with dozens of genera and more than 2000 species. These insects are widely distributed outdoors in tropical and subtropical regions. Some members of the genus *Blaberus* are extremely large, reaching more than 80 mm in length. These are the most highly evolved cockroaches, having developed the ability to incubate their eggs internally and, in some species, to nourish them.
- **Polyphagidae** is a small family with only a few described genera and 100 to 200 species. Females of most species are wingless. These cockroaches are widely distributed in harsh environments, such as deserts and other arid climates. Some members of the genus *Arenivaga* have evolved structures that can absorb moisture from humid air.

Other, more complex, classification schemes for cockroaches also exist, indicating that the subject is not entirely settled. In addition, insect collections from formerly unexplored locations often include many undescribed cockroach species. Thus, it is likely that the total number of extant species is much higher than the figures just given. Indeed, the vast majority of cockroaches live in the tropical regions of the world, many of which have not been adequately assessed to establish the diversity of insect life, including cockroaches, occurring there.

Cockroaches, being hemimetabolous insects, have egg, nymph, and adult stages and grow through a series of molts. They vary greatly in size, ranging from a few millimeters to over 100 mm in length. Many cockroaches are dark brown, but some are black or tan, and others show a surprising amount of color variation and cuticular color patterns. Most species have four wings as adults, and some are capable of rapid, sustained flight; others are wingless or have wings that are variously reduced in size. The majority of species are either nocturnal (or are hidden from view because of where they live), but some are diurnal. Cockroaches occupy diverse habitats, such as among or under dead or decaying leaves, under stones or rubbish, under the bark of trees, under drift materials near beaches, on flowers, leaves, grass, or brush, in the canopy of tall trees, in caves or burrows, in the nests of ants, wasps, or termites, in semiaquatic environments, and burrowing in wood. Thus, the common view of cockroaches as pests is not representative of the group as a whole.

COCKROACHES AS PESTS

The most important of the several reasons for considering some cockroaches to be pests is based on the species that invade people's homes and other buildings and become very numerous. Most people find such infestations to be objectionable, in part because the important pest species also have an unpleasant odor and soil foods, fabrics, and surfaces over which they crawl. However, on a worldwide basis less than 1% of all known cockroach species interact with humans sufficiently to be considered pests. The actual number varies depending on location, because some pest species are greatly restricted in their global distribution. It is also true that more pest species are encountered in tropical locations than in the colder parts of the world. Of the 25 to 30 species that can be a problem, more than half are only occasionally of importance and should be rated as minor or even incidental pests. Of the remaining species, only four or five are of global importance as pests, with the other nine or ten species having regional significance only.

FIGURE 1 German cockroach. From left: adult male, adult female, nymph, ootheca.

The most important pest species is the German cockroach, *Blattella germanica* (Blattellidae) (Fig. 1). It has a worldwide distribution and can survive well in association with any human habitation that provides warmth, moisture, and food. It is small, measuring 10 to 15 mm in length. Adults are yellowish-tan but nymphs are black, with a light-colored stripe up the mid-dorsum. There are two longitudinal, black, parallel bands on the promotum of both nymphs and adults. The wings cover most of the body in adults of both sexes. This is a nocturnal species that lives mainly in kitchen and bathroom areas. When a person enters the kitchen of an infested house at night and turns on a light, the cockroaches scurry out of sight—a startling experience that adds to the desire to eliminate them. There are three or four generations per year. Each egg mass (ootheca) contains from 30 to 50 eggs, and each female can produce three to six oothecae. The potential for rapid population expansion is obvious.

The oriental cockroach, *Blatta orientalis* (Blattidae) (Fig. 2), the next most important pest species, is restricted to the more temperate regions of the world. It is large, measuring 20 to 27 mm in length. All stages are dark brown to black. Females are essentially wingless, but in males the wings cover about two-thirds of the abdomen. This cockroach frequents basements and crawlspaces under buildings, where temperatures are cooler, and often lives outdoors. It is a long-lived insect and may require 1 to 2 years to complete its life cycle. The ootheca contains 16 eggs, and one female may produce eight

FIGURE 2 Oriental cockroach. From left: adult male, adult female, nymph, ootheca.

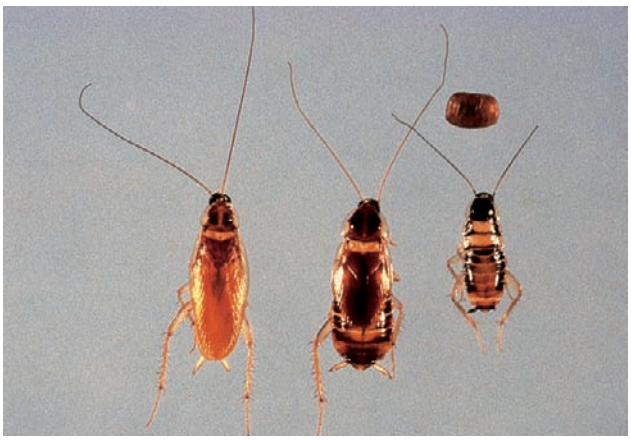

FIGURE 4 Brown-banded cockroach. From left: adult male, adult female, nymph, ootheca.

or more oothecae. Under favorable conditions *B. orientalis* can become very numerous.

There are five or six species belonging to the genus *Periplaneta* (Blattidae) (Fig. 3) that are important pests. The American cockroach, *P. americana,* is the most notorious. It measures 35 to 40 mm in length and is a chocolate-brown color in all stages. Adults of both sexes are fully winged and may undertake a weak flight. *P. americana* is widely distributed around the world but does not extend into the temperate zones as far as does the German cockroach. It requires 6 to 9 months to complete its life cycle. Among other *Periplaneta* species of importance are the Australian cockroach, *P. australasiae,* the smoky-brown cockroach, *P. fuliginosa,* and the Japanese cockroach, *P. japonica.* Each has a more restricted distribution, with *P. japonica,* for example, being found in Japan and China. They are all large cockroaches with a long life cycle but can become numerous under certain conditions. Although tending to be outdoor cockroaches, they often occupy buildings in which food is stored, prepared, or served.

The brown-banded cockroach, *Supella longipalpa* (Blattellidae) (Fig. 4), is a nearly cosmopolitan pest. It is small, measuring 10 to 14 mm in length. As its common name indicates, there are two dark, transverse stripes or bands on the dorsum. The pronotum lacks the two black bands found on the German cockroach. Nymphs are light colored. Females produce numerous oothecae and glue them in inconspicuous places. Each one has about 16 eggs. The life cycle requires approximately 3 months to complete. This cockroach occupies homes and other buildings but unlike the German cockroach is not restricted to the kitchen and bathroom.

FIGURE 3 American cockroach. From left: adult male, adult female, nymph, ootheca.

Other pest species include the Turkistan cockroach, *Blatta lateralis* (Blattidae), two species in the genus *Polyphaga* (Polyphagidae), the Madeira cockroach, *Rhyparobia maderae* (Blaberidae), the lobster cockroach, *Nauphoeta cinerea* (Blaberidae), the Suriname cockroach, *Pycnoscelus surinamensis* (Blaberidae), the Asian cockroach, *Blattella asahinai* (Blattellidae), the harlequin cockroach, *Neostylopyga rhombifolia* (Blattidae), and the Florida cockroach, *Eurycotis floridana* (Blattidae). Most of these species are of regional concern as pests.

COCKROACHES AND HUMAN HEALTH

Cockroaches harbor many species of pathogenic bacteria and other types of harmful organism on or inside their bodies, but they do not transmit human diseases in the same manner as do mosquitoes. They acquire the harmful organisms because of their habit of feeding on almost any type of organic matter, including human and animal wastes. These cockroach-borne organisms can remain viable for a considerable period of time. If the cockroach next visits and soils food intended for human consumption, it is likely that harmful organisms will be deposited on the food. Consuming such food can lead to gastroenteritis, diarrhea, and intestinal infections and pathogenic conditions of other types.

Cockroaches have been shown to harbor pathogenic bacteria belonging to the genera *Mycobacterium, Shigella, Staphylococcus, Salmonella, Escherichia, Streptococcus,* and *Clostridium.* They also harbor pathogenic protozoa in the

genera *Balantidium, Entamoeba, Giardia,* and *Toxoplasma,* and parasites in the genera *Schistosoma, Taenia, Ascaris, Anclyostoma,* and *Necator.* From these lists of organisms it is clear that cockroaches are important as potential disease vectors.

Cockroaches are also important because people can become allergic to them, especially under conditions of constant exposure. These reactions usually involve the skin and/or respiratory system. Studies have shown that people who exhibit skin or bronchial responses to cockroaches have elevated levels of cockroach-specific antibodies. These responses can be severe and may require treatment. The species most commonly involved in producing allergic reactions are the German and American cockroaches.

COCKROACH CONTROL

There are several considerations that come into play in any discussion of cockroach control. The first is understanding how infestations arise. The usual modes of entry for German and brown-banded cockroaches are through infested parcels containing food or other materials, and by movement from abutting dwellings. Most of the other, larger pest species tend to live outdoors and can move from one building to another. They can also be introduced in parcels. Thus, the next consideration is prevention of entry. All entering parcels should be inspected to be sure they do not contain cockroaches, and dwelling defects should be corrected to exclude invaders. Finally, human living space should be kept free of clutter, which can act as hiding places for cockroaches, and food left on dishes, in sinks, or on floors, which can feed a population of cockroaches, should be disposed of properly.

When infestations occur, there are two main methods of control. Nonchemical methods include trapping and vacuuming cockroaches, both of which can significantly reduce the size of an infestation. In addition, freezing, overheating, or flooding structures with a nontoxic gas can be used to kill the pests. Some of the latter procedures require specialized equipment and are best done by professional pest control operators.

The most common method of control is the use of chemical poisons. A large variety of insecticides exist that will kill cockroaches. Some of them are contact poisons that are absorbed as cockroaches walk over treated surfaces. The most common of these belong to the chemical classes called pyrethroids, organophosphates, and carbamates. They kill by disrupting the insect's nervous system, each in a specific manner. Other insecticides are administered in bait formulations that must be eaten by the cockroach. Among them are avermectin and fipronil, which also attack the nervous system, hydramethylnon, which disrupts cellular respiration, and boric acid, which destroys the cells lining the insect gut wall. Each of these materials, as well as others not mentioned, has its own chemical characteristics and must be used in accordance with label instructions.

New insecticides are regularly being introduced that can kill cockroaches, and older ones are being phased out. A critical goal is to develop safer chemicals and safer methods of applying them. For example, the older practice of applying insecticides to surfaces over which cockroaches are expected to crawl is being used less frequently and, as a consequence, the organophosphate and carbamate insecticides especially are being phased out. The practice of dispensing chemicals as baits has largely replaced the surface application method. With baits, the insecticide is more confined and the safety (of humans and pets) is thereby enhanced. The use of baits has become practical in recent years because some of the newer chemicals are highly palatable for cockroaches in bait formulations.

Cockroach control in the future will likely depend on the availability of new insecticides as well as the development of better methods of applying them. Among the approaches that are possible is searching for chemicals that act on sites not previously exploited. For example, a combination of two chemicals is known that prevents cockroaches from producing uric acid. Previous research has shown that storing and recycling the chemical constituents in uric acid is critical to the survival of cockroaches. The functioning of this system is dependent on the fat body endosymbiotic bacteria, mentioned earlier. Other points of metabolic vulnerability will also probably be found in the future.

Another reason for the need for new chemical approaches is that the most important cockroach pest, *B. germanica,* has become resistant to many of the older insecticides. When this occurs, either the effectiveness of those chemicals is greatly reduced or they become useless against resistant populations. With continued use of the newer chemicals, resistance to some of them will probably develop. A steady supply of new chemicals with new modes of action will greatly alleviate this problem and facilitate continued control.

See Also the Following Articles

Isoptera • Medical Entomology • Orthoptera • Phasmida • Urban Habitats

Further Reading

Baumholtz, M. A., Parish, L. C., Witkowski, J. A., and Nutting, W. B. (1997). The medical importance of cockroaches. *Int. J. Dermatol.* **36,** 90–96.

Bell, W. J., and Adiyodi, K. G. (1982). "The American Cockroach." Chapman & Hall, New York, NY.

Cloaric, A., Rivault, C., Fontaine, F., and LeGuyader, A. (1992). Cockroaches as carriers of bacteria in multi-family dwellings. *Epidemiol. Infect.* **109,** 483–490.

Cochran, D. G. (1999). "Cockroaches: Their Biology, Distribution, and Control." World Health Organization/CDS/CPC/WHOPES/99.3. WHO, Geneva, Switzerland.

Cornwell, P. B. (1968). "The Cockroach," Vol. 1. Hutchinson, London.

Helm, R. M., Burks, W., Williams, L. W., Milne, D. E., and Brenner, R. J. (1993). Identification of cockroach aeroallergins from living cultures of German and American cockroaches. *Int. Arch. Allergy Appl. Immunol.* **101,** 359–363.

Labandeira, C. C., and Sepkoski, J. J., Jr. (1993). Insect diversity in the fossil record. *Science* **261,** 310–315.

McKittrick, F. A. (1964). Evolutionary studies of cockroaches. Cornell University Agricultural Experimental Station Memorandum 389. Cornell University, Ithaca, NY.

Nalepa, C. A., and Lenz, M. (2000). The ootheca of *Mastotermes darwiniensis* Froggatt (Isoptera: Mastotermitidae): Homology with cockroach oothecae. *Proc. R. Soc. Lond. B* **267,** 1809–1813.

Roth, L. M. (1989). *Sliferia,* a new ovoviviparous genus (Blattellidae) and the evolution of ovoviviparity in Blattaria (Dictyoptera). *Proc. Entomol. Soc. Wash.* **91,** 441–451.

Rust, M. K., Owens, J. M., and Reierson, D. A. (eds.) (1995). "Understanding and Controlling the German Cockroach." Oxford University Press, New York.

Thorne, B. L., and Carpenter, J. M. (1992). Phylogeny of the Dictyoptera. *Syst. Entomol.* **17,** 253–268.

Blood

see *Circulatory System*

Blood Sucking

M. J. Lehane
University of Wales, Bangor

Of the 1 million described insect species, only 300 to 400 species feed on blood. The best known groups of blood-sucking insects are the lice, fleas, mosquitoes, sand flies, black flies, and bugs. But there are also several lesser known groups such as the nycteribiids and streblids, two families of cyclorrhaphous flies found exclusively on bats; the Rhagionidae or snipe flies, a little-studied group of brachyceran flies; some lepidopterans (e.g., *Calpe eustrigata*); and even some coleopterans (e.g., *Platypsyllus castoris*) that appear to have started on the evolutionary road to hematophagy.

Blood-sucking insects are of immense importance to humans, primarily because of the diseases they transmit. They also cause huge losses in animal husbandry because of disease transmission and because of direct losses linked to the pain and irritation they cause to animals. The most spectacular example of this agricultural loss is the prevention of the development of a cattle industry worth billions of dollars a year through much of sub-Saharan Africa because of tsetse fly-transmitted trypanosomiasis, although some argue that this has been Africa's savior because it has preserved wildlife and prevented desertification. Blood-sucking insects also cause serious losses in the tourist industry in areas as diparate as the French Camargue, the Scottish Highlands, and the state of Florida. We cannot ignore the sheer annoyance they can cause to us all.

EVOLUTION OF THE BLOOD-SUCKING HABIT

Although blood-sucking insects are poorly represented in the fossil record, it seems probable that they emerged along with the first nesting or communal dwelling vertebrates (reptiles) in the Mesozoic era (65–225 mya). Evolution of the blood-sucking habit probably occurred in two main ways. The first route involved the attraction of insects to vertebrates, with the attraction being either to the protection of the nest environment or for the utilization of vertebrate-associated resources such as dung. The second route involved morphological preadaptations that permitted the rapid adoption of the blood-sucking habit.

Many insects would have been drawn to vertebrate nests because of the protected environment and abundance of food there. Gradually, some would have progressed to feeding on cast skin or feathers. Phoresy also would have permitted easy travel from one nest to another. Once phoresy was adopted, the insects may have begun to feed directly on the host animal and thus established an even more permanent association with the host; mallophagan lice make a good example of this type of association. Regular accidental encounters with blood may then have led rapidly to the evolution of the blood-sucking habit because of the highly nutritious nature of blood compared to skin, fur, and feather.

Other insects are attracted to vertebrates outside the nest situation to utilize other vertebrate-associated resources, notably dung. Dung is used by a wide variety of organisms and there is strong competition to be the first to lay eggs in it. So, for example, the female horn fly *Haematobia irritans* lays its eggs in dung within 15 s of its deposition. To do this, the insect must remain permanently with the vertebrate; to do that, it must feed on the vertebrate. The high nutritional content of blood will then make hematophagy a favored evolutionary route.

Some insects also had morphological preadaptations for piercing surfaces, facilitating the relatively easy switch to blood feeding. Entomophagous insects (those that feed on other insects) and plant-feeding insects are prime candidates. For example, the Boreidae are a group of small apterous scorpion flies who are capable of jumping. They live in moss and feed on other insects by piercing them with their mouthparts. They are commonly found in nests because of the moss content and abundance of insects found there. It is easy to imagine such a lineage developing into fleas.

Insects that feed on plants may also have switched to the blood-feeding habit. An unusual example is a blood-feeding moth, *C. eustrigata.* This moth belongs to a group of noctuids having a proboscis that is hardened and modified to allow them to penetrate fruit rinds. *C. eustrigata* has used the morphological preadaptation to feed on vertebrate blood.

HOST CHOICE

The question of host choice is an extremely important one because it defines patterns of disease transmission and economic damage caused by blood-sucking insects. Blood-sucking insects in general feed on a range of different hosts, including birds, reptiles, mammals, and amphibians. Even invertebrates such as annelids, arachnids, and other insects

are sometimes included in the diet. But any particular blood-sucking insect generally feeds only from a small segment of the available hosts. This segment of choice is preferred but it is not immutable. This can be clearly seen around zoos where the exotic animals are quickly incorporated into the diet of the local blood-sucking insects.

The determinants of host choice are complex, but probably one of the most important factors is simply host availability. Changes in host availability because of more intensive animal husbandry, coupled with decreasing rural populations of humans and improved, mosquito-free housing, were a major factor in the disappearance of autochthonous malaria from Northern Europe in the past century. Despite our poor understanding of the factors determining host choice, there is a direct relationship between the number of hosts that blood-sucking insects utilize and the insects' locomotory abilities (which is often reflected in the amount of time they spend with the host). Thus, ectoparasites (which have poor locomotory abilities and usually remain permanently on hosts) are often restricted to a single host species. A good example is the louse *Haematomyzus elephantis,* which is restricted to elephants. At the other extreme, those flying blood-sucking insects such as mosquitoes that make contact with the host only long enough to take a blood meal often display a very catholic host choice. For example, a sample population of the mosquito *Culex salinarius* was shown to take 45% of its blood meals from birds, 17% from equines, and 15% from canines; moreover, 13% of the meals was a mixture of blood from more than one host!

In general terms, the most common hosts chosen are large herbivores. Large, social herbivores present an abundant, easily visible food source that is reliable and predictable from season to season. Carnivores in comparison are fewer in number, often solitary, and range unpredictably over wide areas. Another reason large herbivores are chosen is that they are poor at defending themselves from attack compared to small, agile animals that will often kill and/or eat attacking blood-sucking insects.

HOST LOCATION

For lice and other blood-sucking insects that are permanently present on the host, finding a new host is simply a matter of moving from one to the other when the hosts are in bodily contact. For blood-sucking insects that are only in temporary contact with the host, finding a host is a more difficult proposition. The following host-seeking behaviors are not rigidly patterned but they probably typically follow one another in a loose sequence. For most blood-sucking insects, olfactory stimuli are the first host-related signals perceived, and visual signals from the host probably are apt to become important at a later stage in host location. Blood-sucking insects make use of this predictability by permitting the current behavioral response to lower the threshold required for the next host signal to elicit the next behavioral response in the host location sequence. The increasing strength and diversity of host-derived signals that the blood-sucking insect receives as it moves closer to the host are thereby enhanced.

Host location is usually restricted to particular times of the day for each species of blood-sucking insect. Thus, tsetse flies tend to be crepuscular, *Anopheles gambiae* (the most important vector of malaria) is a night feeder, and the stable fly *Stomoxys calcitrans* bites during the day. As hunger increases, bouts of host location behavior intensify. For many blood-sucking insects such as the tsetse, the first behavior is often to choose a resting site where they have a good chance of encountering a host-derived signal and once there to remain motionless and wait for a host-derived signal. This strategy combines minimal energy usage with a good chance of encountering a host. Other blood-sucking insects use more active strategies. If a gentle wind is blowing from one direction the optimum strategy can be to fly across the wind so that the probability of contact with a host odor plume is enhanced.

Host-derived olfactory clues used include carbon dioxide, lactic acid, acetone, octenol, butanone, and phenolic compounds found in urine. These are probably used in combination by each insect's sensitivity to different combinations of smells. For example, we can look at the power of phenolic components found in bovine urine to draw tsetse flies to a bait. Used singly, 3-*n*-propylphenol draws roughly equal numbers of *Glossina pallidipes* and *G. m. morsitans.* In contrast, when 3-*n*-propylphenol is used in combination with 4-methylphenol catches of *G. pallidipes* increase 400%, whereas catches of *G. m. morsitans* decrease. The explanation for this may be that first, mixtures of odors are a stronger guide to the presence of a host than a single odor alone and so will minimize energy consumption from chasing false trails. Second, mixtures may help in host choice by guiding blood-sucking insects to particular host species.

Tracking the source of an odor plume while in flight is a major task. It is believed that many blood-sucking insects achieve this by using upwind optomotor anemotaxis. During flight, insects will be blown off course by any wind that is blowing. They can use this fact to determine wind direction. To do this, they observe the perceived movements of fixed objects on the ground and by comparing this to the direction in which they are trying to fly determine wind direction. The suggestion is that the insect flies across wind until an odor plume is encountered, when it turns upwind. If the odor plume is lost, it recommences flying across wind until it refinds the odor and turns upwind once more. This proceeds until the insect comes into the immediate vicinity of the host. It is believed that hosts can be detected by odor at about 90 m by tsetse flies and at 15 to 80 m by some mosquitoes.

Vision is also used in host location by the majority of blood-sucking insects and is used most extensively by day-feeding insects in open habitats. In general, blood-sucking insects can detect and discriminate between objects on the basis of color contrast, relative brightness (intensity contrast), movement, and shape. Insects are quite sensitive to movement and their color vision stretches up into the UV but not down to the red.

Night-feeding blood-sucking insects have relatively better intensity contrast than color contrast, whereas for day-biting blood-sucking insects movement perception and color contrast may be particularly important. Large individual herbivores (as opposed to herds) are thought to be detected by vision at about 50 m by tsetse flies and at 5 to 20 m by some mosquitoes.

Once the blood-sucking insect is in proximity to the host, heat and humidity become important factors in location in addition to the continuing importance of vision and odor. Temperature is probably a useful guide from about 5 cm to a meter or so from the host depending on insect species. Even when they have contacted a host, blood-sucking insects will imbibe a blood meal only if it provides the correct biochemical characteristics (i.e., taste).

THE BLOOD MEAL

Blood-sucking insects take huge meals. Temporary ectoparasites such as the tsetse fly typically ingest more than their own unfed body weight in blood. The reasons are twofold. First, taking a blood meal is a very dangerous activity and taking huge blood meals minimizes the number of times an insect must associate with the host. Second, locating the host is often difficult and huge blood meals are a way of making the most of each encounter. Mouthparts are adapted to the blood-feeding habit. Typically, they are either of the piercing kind seen in mosquitoes, bugs, lice, and fleas or the cutting kind seen in tabanids, black flies, and biting flies.

The host usually responds to feeding activity, particularly the injection of saliva, by mounting an immune response that includes pruritis (itching). Typically this begins to occur about 3 min after feeding commences. Thus, there is a selective advantage in completing the blood meal within this 3-min "safety period" after which the host will be alerted to the presence of the insect. To help achieve this, blood-sucking insects have produced a range of antihemostatic molecules in the saliva, one of the major functions of which is to minimize host contact time.

Antihemostatic molecules produced by the blood-sucking insect include anticoagulant molecules working variously, for example, on thrombin or factors VIII and X. However, platelet plugging of small wounds is probably of more importance to blood-sucking insects than blood coagulation. Consequently, they also produce anti-platelet aggregating factors such as apyrase. These are used to impede the plugging of the penetration wound in capillaries and to prevent clogging of the insect mouthparts. The insect saliva also contains powerful vasodilatory substances to increase blood flow to the wound and anti-histamines that will minimize inflammation and itching, possibly extending the "safe period." Salivary components are also important as they can facilitate the transmission of arthropod-borne pathogens. For example, the production of *Leishmania*-enhancing factor in the saliva of the sand fly *Lutzomyia longipalpis* enhances the establishment of the parasite *Leishmania major* in the vertebrate host. It has also been shown that such effects may be limited to naïve hosts, suggesting that the history of exposure to vector saliva may influence the outcome of potentially infectious inoculations. Parasites can also manipulate the salivary glands to their own advantage. Thus, malaria sporozoites damage the salivary glands of mosquitoes, reducing antihemostatic effectiveness, and thus extend probing time and increase the chances they will be transmitted to a new host.

Some blood-sucking insects feed only on blood during their entire life. Examples include the tsetse flies, streblids, hippoboscids and nycteribiids, triatomine and cimicid bugs, and lice. Blood is deficient in certain nutrients such as the B-group vitamins and pantothenic acid, and the insect cannot make these itself. To make up for this deficiency, these obligate hematophages harbor symbiotic microorganisms that produce these extra nutrients. These symbionts are often housed in a specialized body compartment, traditionally called a mycetome or, more recently, a bacteriome. For example, the tsetse fly *Glossina* harbors three symbiotic microorganisms, including *Wigglesworthia glossinia,* which is from the γ-subdivision of the Proteobacteria, in the bacteriome of the anterior gut.

HOST–INSECT INTERACTIONS

There are several evident morphological adaptations for a blood-sucking life. Piercing or cutting mouthparts are the clearest example. In addition, many periodic and permanent ectoparasites such as fleas and lice are laterally or dorsoventrally flattened and are wingless, which are adaptations allowing them to move easily through the pelage or feathers and to avoid being groomed by permitting them to flatten themselves against the skin. Most of these ectoparasites also have cuticular extensions in the form of spines and combs. These are longer and spinier in bird-infesting forms than in those found on mammals. The combs in particular are found covering weak spots in the body such as the articulations between body segments. The spacing of the tips of the combs correlates well with the diameter of the hairs on the body of the host. This suggests that these combs have a dual function: protecting the body from abrasion and acting as an anchoring device for the ectoparasite.

The host regulates the numbers of permanent ectoparasites by grooming, usually with both the toes and the teeth. This often limits ectoparasite distribution on the host to those areas the host can groom least efficiently, such as the head and neck. The immune response mounted against these blood-sucking insects is often very localized. It makes feeding on these protected areas of the skin difficult, with the result that the insects feed less well or move to less affected areas of the body where they are more easily groomed. The result is that the host regulates ectoparasite numbers.

The host also shows behavioral defenses to temporary ectoparasites such as mosquitoes. The level of defensive behavior is usually density dependent and thus can have important consequences for disease transmission. For example, the arbovirus eastern equine Encephalitis (EEE),

which is naturally found in birds, is transmitted in the United States by the mosquito *Culiseta melanura.* During spring and early summer, these mosquitoes feed almost exclusively on passerine birds, transmitting the virus among them. Later in the season, as mosquito numbers increase, bird defensive behavior increases and mosquitoes are more willing to feed on other vertebrate hosts. This is when EEE is transmitted to other vertebrates including horses and humans.

See Also the Following Articles

Medical Entomology • Mosquitoes • Phthiraptera • Siphonaptera • Tsetse Fly • Veterinary Entomology

Further Reading

Beaty, B., and Marquardt, W. (eds.) (1996). "The Biology of Disease Vectors." University of Colorado Press, Boulder.

Braks, M. A. H., Anderson, R. A., and Knols, B. G. J. (1999). Infochemicals in mosquito host selection: Human skin microflora and *Plasmodium* parasites. *Parasitol. Today* **15,** 409–413.

Charlab, R., Valenzuela, J. G., Rowton, E. D., and Ribeiro, J. M. C. (1999). Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis. Proc. Nat. Acad. Sci. U.S.A.* **96,** 15155–15160.

Clements, A. (1999). "The Biology of Mosquitoes." CABI Int., Oxon.

Dye, C. (1992). The analysis of parasite transmission by bloodsucking insects. *Annu. Rev. Entomol.* **37,** 1–19.

Hurd, H., Hogg, J. C., and Renshaw, M. (1995). Interactions between bloodfeeding, fecundity and infection in mosquitos. *Parasitol. Today* **11,** 411–416.

Lane, R., and Crosskey, R. (eds.) (1993) "Medical Insects and Arachnids." Chapman & Hall, London.

Lehane, M. J. (1991). "Biology of Blood-Sucking Insects." Chapman & Hall, London.

Sutcliffe, J. F. (1986). Black fly host location: A review. *Can. J. Zool.* **64,** 1041–1053.

Titus, R. G., and Ribeiro, J. M. C. (1990). The role of vector saliva in transmission of arthropod-borne disease. *Parasitol. Today* **6,** 157–160.

Vale, G. A., Hall, D. R., and Gough, A. J. E. (1988). The olfactory responses of tsetse flies, *Glossina* spp. (Diptera: Glossinidae), to phenols and urine in the field. *Bull. Entomol. Res.* **78,** 293–300.

Waage, J. K. (1979). The evolution of insect/vertebrate associations. *Biol. J. Linn. Soc.* **12,** 187–224.

Body Size

Christer Björkman and Mats W. Pettersson

Swedish University of Agricultural Sciences

Because of its structure, the environment offers many more niches for small organisms than for large ones. The relatively small size of insects, which is one of the reasons for their success, has therefore made them very diverse, a characteristic that has resulted in a high number of species. In spite of their generally small size, limited by their method of gas exchange, insects show as large a range in size as other groups of organisms.

SIZE VARIATION

Size varies tremendously among and within orders, families, and species of insects. The smallest extant insects known are about 0.2 mm in length and can be found among beetles of the family Ptiliidae and wasps of the family Mymaridae, which are egg parasitoids. Insects of this minute size are smaller than the largest one-celled protozoans. The largest extant insects are phasmid (walkingstick) species (up to 30 cm long), sphingid moths (wingspan of up to 30 cm), and some beetles of the genera *Megasoma, Dynastes,* or *Goliathus* (up to 100 gr). The high number of possible runners-up for the title "largest insect on earth" highlights another difficulty when dealing with size: What measure should be used?

Although insects are not as large as some other organisms, the range in size among insects is almost as large. For example, the difference in volume between the largest mammal (the blue whale) and the smallest (a minute shrew) is about $1{:}2 \times 10^8$. This is comparable to $1{:}1.5 \times 10^8$ for the two extremes among insects (ptiliid beetle and Goliath beetle).

HOW TO MEASURE INSECTS

Body length (measured from tip of head to tip of abdomen) is probably the most often used measure of size and the easiest to comprehend. Weight (or biomass) is a measure of size that interests many ecologists because it often correlates well with fecundity. Fecundity, in turn, is often a major fitness component and may be a key feature in population dynamics. For practical reasons, traits that are correlated with weight are used instead, such as hind tibia length, front wing length, elytra length, head capsule width, or body volume. The ratios between these and other morphological traits are often used in taxonomic descriptions of insect species. These ratios give descriptions of shape. The main objective in describing size is to identify traits that are easy to measure (e.g., does not have a curved shape) and are stable on prepared specimens.

SOME EXPLANATIONS FOR VARIATIONS IN SIZE

The size of an insect individual is determined by its genes and by the environment in which it grows. Temperature, crowding, food quantity, and food quality are examples of environmental factors that affect size, but insects may make up for such effects by compensatory feeding.

The size of female insects often determines their fecundity, which may be manifested in giving birth to many small offspring or a smaller number of large ones. To be able to produce many large offspring, which may be adaptive for survival in a harsh environment, the female herself must be large. The importance of size for female fecundity can often be seen in the sexual dimorphism of insect species, in which males typically are much smaller than females.

Although the primary role of male insects is to fertilize the eggs, males may benefit from being large because they

contribute to the realized fecundity of females by providing resources through their ejaculate or in competing with other males to obtain mates. An example of the latter characteristic is provided by some species of digger wasps (Sphecidae), in which males compete intensely with each other for females only half their size.

Sometimes, adult or larval foods come in packages or shapes that allow only very small insects to use them. Such foods include very small items like seeds and insect eggs, or very thin items like pores of fungi. Many insect families that use these foods [e.g., bruchid beetles, mymarid parasitic wasps, and nanosellin (Ptiliidae) beetles, respectively] have been adapted to and have radiated into several species under such living conditions.

Insects smaller than 1 mm operate in a world where gravity and molecular forces are in the same order of magnitude. This can be advantageous when, for example, insects find it easier to climb vertical surfaces. However, it can also lead to problems when, for example, an insect is trapped in a drop of water by the water's surface tension.

FACTORS THAT LIMIT SIZE

The smallest insects will have difficulty making room for the internal organs that are necessary for their existence. For example, some ptiliid beetles can lay only one egg at the time because their eggs may be up to 0.7 times the size of the whole insect.

The largest size an insect can reach is limited by the tracheal system. In insects, gas exchange with air is mediated directly to the tissues by a highly branched system of chitin-lined tubes called tracheae. No cells in the insect body are more than 2 to 3 μm from a tracheole. Diffusion along a concentration gradient can supply enough oxygen for small insects, but forms that weigh more than about a gram, or are highly active, require some degree of ventilation. Most insects have ventilating mechanisms to move air in and out through the tracheal system, but the need to allow enough oxygen to reach the tissue by diffusion imposes limits on tracheal length. Most large insects present today have long slender bodies, a trait that also limits tracheal length. Furthermore, elaborations of the tracheal system could not be made without destroying the water balance in large insects. However, there are exceptions: some of the heaviest extant beetles have bulky bodies, but these insects are not (or do not have to be) very quick and do not fly.

CHANGE OVER TIME

It has been suggested that organisms increase in size over an evolutionary time scale. However, there is no evidence to support this suggestion, and perhaps natural selection acts on correlated traits that constrain the evolution of increased size. In fact, fossils reveal that some insects in the past were much larger than their extant relatives. For example, many winged Carboniferous and Permian insects, existing about 300 mya, had wingspans exceeding 45 cm; the largest was the Permian dragonfly *Meganeuropsis schusteri,* which had a wingspan of 71 cm. These insects certainly also had long, narrow bodies, to reduce the length of the trachea. During these prehistoric times, the atmospheric oxygen concentration was much higher (up to about 35%) than the present level (20.9%), which may have allowed sufficient oxygen to reach the innermost tissues of very large insects. However, such an oxygen-rich atmosphere also would have augmented aerodynamic properties in early flying insects. It has been suggested that later appearing insects could not evolve to a large size because of competition for niches with birds and other later appearing animals.

OTHER RELATIONSHIPS

Ectotherms, including insects, in contrast to endotherms, seem to follow the converse of Bergmann's rule: that is, they are smaller toward higher latitudes and altitudes. A high degree of genetic determinism seems to underlie this pattern. A possible reason could be an evolved response to geographic patterns in season length.

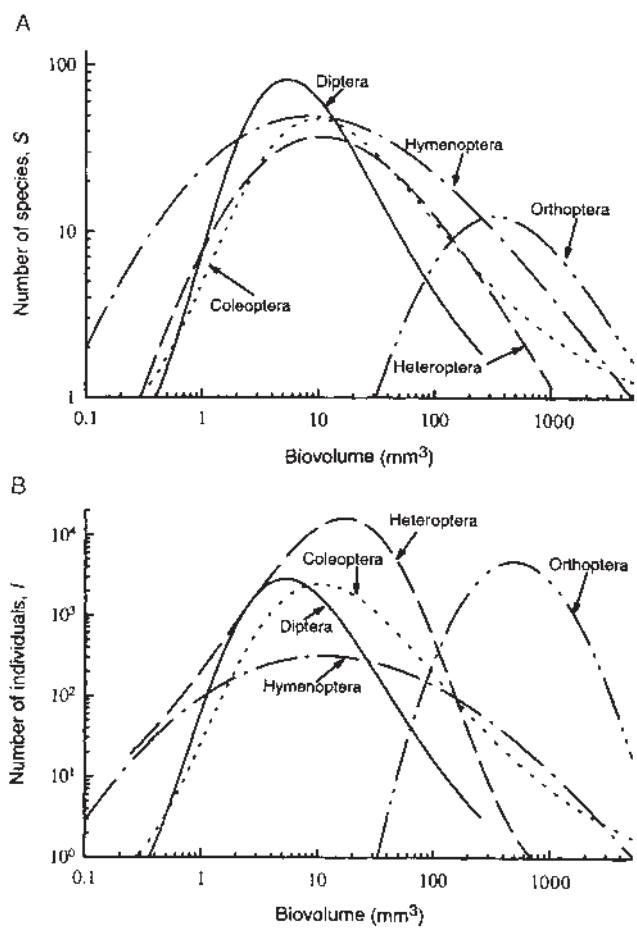

FIGURE 1 Relationship between insect size (body volume) and (A) species richness and (B) number of individuals for five insect orders. [Redrawn from Fig. 2 in Siemann, E., Tilman, D., and Haarstad, J. (1996). Insect species diversity, abundance, and body size relationships. *Nature* **380,** 704–706, with permission from *Nature.*]

Insect assemblages are thought to be structured by competition, with most of the insects found in medium-sized classes. Thus, the size of a particular insect is governed by its living habits and its feeding guild, in which competition with similar insects has forced some to evolve a larger or smaller body size. Empirical data show that species diversity in any taxonomic group of insects peaks at some intermediate body size (Fig. 1). For some authors, this implies that there may be fewer undescribed small insect species than previously suggested, which in turn, suggests that global biodiversity probably is lower than the highest estimates (30–50 million species). However, it is not clear whether such a pattern results from biological processes or from statistical or sampling properties. The size distribution of cars parked at Heathrow Airport also shows a peak in diversity at an intermediate size. Nevertheless, more knowledge about the causes behind size distribution patterns among insects and other organisms may provide key information in the effort to preserve biodiversity.

See Also the Following Articles

Biogeographical Patterns • Growth, Individual • Tracheal System

Further Reading

Conner, J., and Via, S. (1992). Natural selection on body size in *Tribolium:* Possible genetic constraints on adaptive evolution. *Heredity* **69**, 73–83.

Dudley, R. (1998). Atmospheric oxygen, giant Paleozoic insects and the evolution of aerial locomotor performance. *J. Exp. Biol.* **201,** 1043–1050.

Gaston, K. J., Blackburn, T. M., and Lawton, J. H. (1993). Comparing animals and automobiles: A vehicle for understanding body size and abundance relationships in species assemblages. *Oikos* **66,** 172–179.

Mousseau, T. A. (1997). Ectotherms follow the converse to Bergmann's rule. *Evolution* **51,** 630–632.

Price, P. W. (1984). "Insect Ecology." Wiley-Interscience, New York.

Siemann, E., Tilman, D., and Haarstad, J. (1996). Insect species diversity, abundance and body size relationships. *Nature* **380,** 704–706.

van Voorhies, W. A. (1997). On the adaptive nature of Bergmann size clines: A reply to Mousseau, Partridge and Coyne. *Evolution* **51,** 635–640.

Boll Weevil

R. Nelson Foster
United States Department of Agriculture

The boll weevil, *Authonomus grandis grandis,* is a beetle belonging to the family Curculionidae. A native of Mexico or Central America, it was first described in 1843 by the Swedish entomologist Boheman. Boll weevil is a new world pest of cultivated cotton and occurs in all Central America countries where cotton is cultivated, Mexico, the United States, Cuba, Hispaniola, Colombia, Venezuela, Brazil, and Peru. Where established, it is probably the most serious pest of cotton. Since its arrival in the United States, it is estimated that actual damage loss combined with control costs attributed to boll weevil exceed $16 billion (consumer-price-index-adjusted value of $91 billion). The boll weevil is important to examine in detail because of its destructive impact in cultivated cotton, use of aggregation pheromone, overwintering diapause, existence as three different but difficult-to-distinguish forms of the species, and status as the target of successful, area-wide eradication efforts in the United States.

MIGRATION AND DISPERSAL IN THE UNITED STATES

Boll weevil is a migratory pest, and its movement is largely dependent on wind direction and speed. It has been known to travel as far as 272 km and can hitchhike on cars, trucks, and trains. Boll weevil was first reported in the United States in the fall of 1894 from Brownsville, Texas, and may have been established as early as 1892. By 1895, the weevil had spread north to San Antonio and eastward to Wharton, Texas. The weevil reached Louisiana in 1903, Mississippi in 1907, and Georgia by 1916; by the 1920s it had infested cotton throughout the Mississippi Delta and the southeastern United States, and by 1922 had become established almost to the northern limits of cotton production. Northern and western portions of Texas became infested as a result of a sequence of expansions of the pest range between 1953 and 1966. Arizona was plagued with problems from the boll weevil beginning in the late 1970s, and in 1982 the weevil was detected in the southern desert valleys of California. The weevil became established in New Mexico in the early 1990s.

DESCRIPTION

Eggs of the boll weevil are pearly white, usually elliptical, and approximately 0.8 mm long by 0.5 mm wide. Legless white grubs hatch from eggs; they have light brown heads approximately 1 mm long. The larva, as it feeds and grows for 7 to 14 days, exhibits a ventrally curved, crescent form, with the dorsum strongly wrinkled and the venter smooth. The pupa is white at first but turns brown as it develops for 3 to 5 days. The adults are initially light in color but darken with age to colors from reddish brown to gray; depending on abundance and nature of food, they range in size from about 2.5 to 7 mm. The dark snout of this weevil is about half as long as the body (Fig. 1). Three forms (Mexican boll weevil, southeastern boll weevil, and thurberia weevil) have been separated by using characters ranging from morphological profemoral width/length ratios and spermatheca shapes to analysis systems relying on molecular biology.

LIFE HISTORY

The boll weevil belongs to a family of insects that is strictly phytophagous. This group is highly host specific and

FIGURE 1 Boll weevil on cotton leaf.

generally prefers flower buds for feeding. Cotton (*Gossypium* spp.) is the principal host plant of boll weevil, but it also develops on certain species of the related genera *Thespesia, Cienfuegosia,* and *Hampea.* The boll weevil passes the winter in diapause in the adult stage sheathed beneath brush and ground litter and in other protected locations in or around cotton fields. In arid areas, overwintering sites may be associated with increased moisture habitats such as near irrigation canals and rivers. Winter survivors emerge from overwintering sites in the spring and begin feeding on the tips of cotton seedlings and squares (i.e., the cotton flower buds). Weevils that emerge before cotton plants have begun to form squares, feed on leaf buds and growing terminals, and live for only a week or two; those that emerge later produce eggs for 3 to 6 weeks. Later generations survive the winter in a diapause state. The female deposits eggs singly in the bottom of punctures she makes in the cotton squares and later in the season in bolls. Overwintered females produce fewer than 100 eggs, but later generations produce 300 or more eggs. The average female's rate of reproduction is 5 or 6 eggs a day. Depending on the temperature, larvae hatch in 3 to 4 days. Larvae feed for 7 to 14 days and pupate. Adults emerge 3 to 6 days later.

A sex pheromone facilitates mating, after which the females begin laying eggs in 3 to 5 days. Two to seven generations can occur in a season. However, as many as 10 generations may develop under favorable conditions. Late in the season as cotton ceases to produce fruit, boll weevils move in large numbers from cotton fields to overwintering sites. Only 1 to 20% of weevils survive the winter. Reduced survival is seen after unusually cold winters, and unusually dry summers also cause some mortality through loss of moisture in overwintering sites.

CONTROL AND ERADICATION

Sound cultural practices combined with chemical control relying on numerous compounds has been the traditional practice for combating boll weevil. Coordinated eradication efforts in the United States have been quite effective. The eradication effort in the United States is based on three major activities: mapping all cotton fields, evaluating weevil presence in each field with pheromone traps, and applying control treatments. The program consists of a series of sequential expansions and usually lasts 3 to 5 years in any particular area. The program relies on intensive, carefully coordinated, ground and aerial treatments of ultra low volume malathion (almost exclusively), concentrated over one to three seasons, in response to predetermined numbers of weevils caught in pheromone traps. The program started in 1983 in the Carolinas and has expanded to parts of all of the cotton-growing states in the country. Active eradication has been completed in Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Arizona, California, and in some parts of the other cotton-growing states. It is projected to be completed in 2005. In the southeastern states, where active eradication has been completed, a remarkable increase in cotton production has occurred. When completed, nationwide eradication will result in substantial economic and environmental benefits throughout the areas once plagued by the boll weevil.

See Also the Following Articles

Agricultural Entomology • Coleoptera • Migration • Regulatory Entomology

Further Reading

Burke, H. R. (1997). Early boll weevil fighters. *Southwest Entomol.* **22,** 248–277.

Burke, H. R., Clark, W. E., Cate, J. R., and Fryxell, P. A. (1986). Origin and dispersal of the boll weevil. *Bull. Entomol. Soc. Am.* **32,** 228–238.

Haney, P. B., (exec. ed.) (2001). "History of Boll Weevil Eradication in the United States." National Cotton Council Series. Herff-Jones, Montgomery, AL.

Pfadt, R. E. (1985). Insects pests of cotton *In* "Fundamentals of Applied Entomology. 4th ed. (R. E. Pfadt, ed.). Macmillan, New York.

Bombyx mori

Satoshi Takeda
National Institute of Agrobiological Sciences, Japan

The silkworm, *Bombyx mori,* is used for sericulture and is one of the most economically important insects in the world (Fig. 1). The species of silkworm usually raised by sericulturists is *B. mori* (Lepidoptera: Bombycidae). A closely allied species is the mulberry wild silkworm, *B. mandarina.* Its morphology does not differ markedly that of from *B. mori,* and hybrids are highly fertile.

Along with *Drosophila melanogaster, B. mori* larvae have been used as a model for various biological studies for many years. These two insect species have greatly contributed to the progress of research in several scientific fields.

FIGURE 1 Larva of the silkworm, *B. mori.* Larva at 4 days of fifth instar.

LIFE CYCLE OF *B. MORI*

Silkworms undergo complete metamorphosis (Fig. 2). Larvae feed on the leaves of the mulberry (family Moraceae, genus *Morus*); they will consume other genera in this family, but growth rate is reduced. At 23 to 25°C the five instars require 25 to 30 days to hatch. Toward the end of fifth instar, *B. mori* spin a cocoon over a 3-day period and pupate within that cocoon; the pupal stage lasts for about 10 days. After molting to the adult has occurred inside the cocoon, the moth emerges. The moth softens the cocoon by orally excreting a special enzyme, cocoonase, and then emerges from the end of the cocoon. The female moth mates the same day as emergence and begins egg deposition. One generation of *B. mori* spans 40 to 45 days.

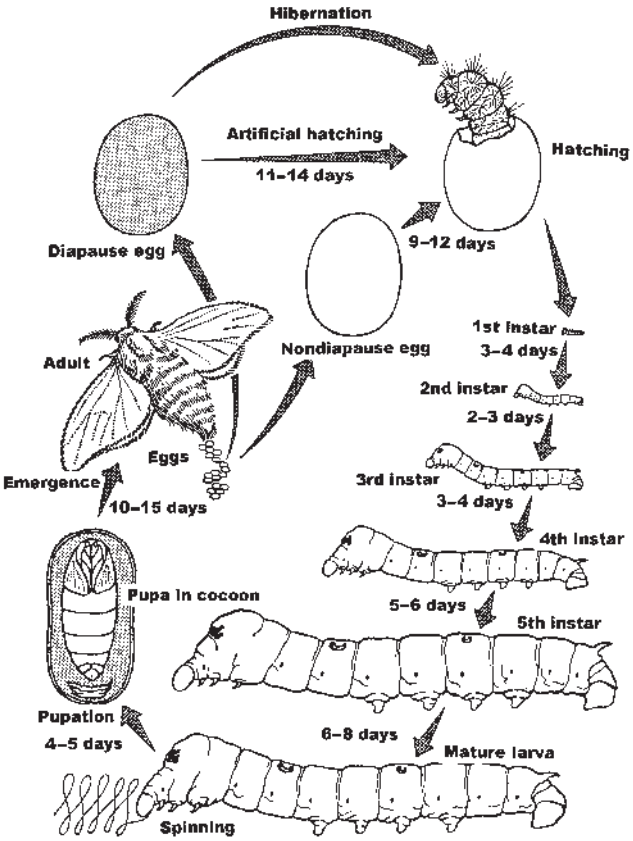

FIGURE 2 Life cycle of *B. mori* reared at 23 to 25°C. [From Mori, T. (1970). Life cycle of *Bombyx mori. In* "The Silkworm—New Experimental Tool in Biology" (T. Mori, ed.), p. 17. Sanseido Press, Tokyo.]

Voltinism (i.e., the number of generations occurring in a single year) has a genetic basis, but it is also strongly affected by environmental factors. Some races are univoltine (only one generation a year); others are bivoltine (two generations a year) or polyvoltine (three or more generations a year, as seen in tropical silkworms that do not undergo diapause). Voltinism is closely tied to the geographic distribution of silkworms, which are divided into Japanese, Chinese, European, and tropical races. Typically, the silkworms distributed in the cold regions are univoltine, those distributed and adapted to warm regions are bivoltine, and those in tropical regions are polyvoltine.

EMBRYONIC DIAPAUSE OF *B. MORI*

Silkworms undergo diapause, during which the embryo within the egg stops growing. In the *B. mori* embryo, diapause is primarily determined by the temperature, light, nutrition, and other conditions of the lifetime of silkworms. Of the various factors, temperature and light during the egg (embryo) stage have the greatest influence. When the eggs of bivoltine silkworms were incubated at relatively high temperatures (25°C) with long days (> 13 h of light), all silkworms that grew from these eggs laid diapausing eggs. When the eggs were incubated at 15°C with short days (< 13 h of light), adults of all these silkworms subsequently laid nondiapausing eggs.

The subesophageal ganglion, located just below the brain, secretes a peptide hormone that induces the embryonic diapause of *B. mori.* In 1951, K. Hasegawa and S. Fukuda, in separate studies, demonstrated that the subesophageal ganglion plays role in induction of diapause. The diapause hormone is composed of 24 amino acids and is produced and secreted by six pairs of neurosecretory cells of the subesophageal ganglion.

SILK PRODUCTION BY *B. MORI*

B. mori larvae have a unique metabolic system for producing a large amount of cocoon protein and efficiently using dietary nitrogen. Both male and female silkworms digest and absorb about two-thirds of the nitrogen in the mulberry leaves they consume, and high percentages of the digested and absorbed nitrogen (66% in females and 70% in males) are utilized in the production of cocoon protein.

During the last larval stage (fifth instar), the silk gland produces the silk for the cocoon from a pair of curved glands found on the ventral side of the digestive tube. The weight of this organ accounts for about 25% of the weight of larvae in the late fifth instar.

The silk gland can produce massive amounts of fibroin and sericin, the proteins constituting silk. Sericin surrounds a fibroin core. The ratio of fibroin to sericin is approximately

3:1. Fibroin is rich in four amino acids: glycine (Gly), alanine (Ala), serine (Ser), and tyrosine (Tyr). The fibroin molecule contains repeats of a section composed of a regular arrangement of three amino acids, Gly, Ala, and Ser. Major amino acids constituting sericin are Ser, aspartic acid, glutamic acid, and glycine. A characteristic of sericin is that, unlike fibroin, it is soluble in hot water. Therefore, when cocoon threads are reeled, most of the sericin is removed, and the remaining raw silk is composed of fibroin alone.

Studies on *B. mori* greatly contributed to early discoveries in insect endocrinology and to the isolation and analysis of insect peptide hormones in 1980s and 1990s. The large size of these insects made experimental morphological studies easier, and because of the importance of this species to the sericultural industry, large quantities of materials for hormone extracts were made available.

In silkworms, larval ecdysis is induced by a molting hormone secreted by the prothoracic gland, which is located inside the first thoracic spiracle. The role of the prothoracic gland in ecdysis was discovered in 1944, and the molting hormone, ecdysone, was structurally determined in 1954 in studies that used large amounts of silkworm pupae as material. Ecdysone was the first hormone to be isolated from an insect species. In addition, the function of the corpora allata in Lepidoptera was also first discovered in silkworms in 1942. The corpora allata, which are small organs located adjacent to the brain, secrete juvenile hormone, which controls silkworm development together with molting hormone. Among the peptide hormones, the molecular structures of prothoracicotropic hormone (initially named "brain hormone"), which controls the secretion of molting hormone, and the diapause hormone, which induces silkworm egg diapause, were elucidated by using silkworms.

TRANSGENIC SILKWORMS AND INSECT FACTORIES

The National Institute of Sericultural and Entomological Science in Japan created transgenic silkworms with a jellyfish fluorescent protein gene as a marker. Prior to this, there were only a few transgenic insects in the Diptera (e.g., *Drosophila, Aedes*). This success resulted from development of a microinjector to introduce DNA into silkworm eggs, and the use of an effective transposon vector. The transgenic silkworms are expected to provide new opportunities for silk production.

A promising use of silkworms outside the clothing industry is in so-called insect factories, where silkworms are used for biological production of peptides or proteins useful for humans. Silkworms are infected with nuclear polyhedrosis virus (one species of baculovirus) to enable them to produce useful substances on a large scale.

See Also the Following Articles

Biotechnology and Insects • Diapause • Ecdysteroids • Lepidoptera • Sericulture • Silk Production

Further Reading

Maeda, S., Kawai, M., Obinata, H., Fujiwara, T., Horiuchi, T., Saeki, T., Sato, Y. and Furusawa, M. (1985). Production of human α-interferon in silkworm using a baculovirus vector. *Nature* **315,** 592–594.

Mori, T. (1970). Life cycle of *Bombyx mori. In* "The Silkworm—New Experimental Tool in Biology" (T. Mori, ed.), In Japanese. Sanseido Press. Tokyo.

Tajima, Y. (1978) "The Silkworm: An Important Laboratory Tool." Kodansha, Tokyo.

Book Louse

see *Psocoptera*

Borers

Timothy D. Paine
University of California, Riverside

Insects that are borers belong to a wide range of taxonomic groups, but they all share a common life history trait: they spend all or part of their larval life feeding within the tissues of their host plant. Some borer species deposit eggs within host plant tissues, whereas other species oviposit on the external surface and the larvae bore into the plant. Although there may be some feeding activity within the phloem and cambial tissues the larvae typically excavate feeding galleries within the woody tissues of perennial plants, within the stems of annual plants, and within the stalks or stems of grasses. Adult borers are free-living outside the host plant.

TAXONOMIC AFFILIATIONS

Insect orders that include species commonly referred to as borers include Lepidoptera, Coleoptera, Hymenoptera, and Diptera. The most varied and numerous representatives are among the moths and the beetles. The primitive wasps, which include the horntails (Siricidae), wood wasps (Xiphydriidae and Syntexidae), and sawflies (Tenthredinidae and Cephidae), are the only representatives of the Hymenoptera. The horntails and wood wasps prefer to colonize weakened hosts and the larvae construct feeding galleries in the wood. Larvae of boring sawfly species often feed within the center of tender shoots, twigs, and stems of their host plants. There are a small number of fly species in the family Agromyzidae with life history strategies that leave injury patterns that could be characterized as boring, but the larvae actually mine the cambial tissue, and the trees overgrow the galleries and the mine remains in the wood.

There are many moth families in the Lepidoptera that include species of larval borers. Among the most important families comprising only species that have a boring life history or having large numbers of boring species are the Hepialidae (ghost moths or swifts), the Sesiidae (clearwing moths), the Cossidae (carpenterworm and leopard moths), and the Tortricidae (leafroller and olethreutine). Larvae of the Hepialidae and the Cossidae tunnel extensively into the wood of their host plants and can cause substantial damage. Larvae of many species of Tortricidae bore through the twigs and tender terminals of vigorous trees and shrubs, whereas the life histories of species of Sesiidae can be highly variable and may include boring in bark, cambium, wood, roots, or gall tissues. Other families of Lepidoptera with at least some species that can be characterized as borers include the Agonoxenidae, Argyresthiidae, Gelechiidae, Momphidae, Nepticulidae, Noctuidae, Pterophoridae, Pyralidae, and Thyrididae. Larvae of one important agricultural pest, the European corn borer, *Ostrinia nubialis* (Lepidoptera: Crambidae), bore into the stems of woody host seedlings as well as the stems of grasses, grains, and herbaceous host plants.

The Coleoptera include a large number families composed exclusively of species with larvae that are boring or having very few representatives that have evolved alternative life history strategies. Although not an entirely comprehensive list, the families include Anobiidae (deathwatch and drugstore beetles), Bostrichidae (false powderpost beetles), Brentidae (brentid beetles), Buprestidae (metallic or flatheaded wood borers), Cerambycidae (longhorned or roundheaded wood borers), Lyctidae (powderpost beetles), Lymexylidae (timber beetles), Platypodidae (ambrosia beetles), and Scolytidae (bark beetles). The Curculionidae (snout beetles or weevils) is a very diverse family that includes a number of species with larvae that bore into plant tissues.

FEEDING STRATEGIES

Many different plant parts serve as sites of insect feeding activity. In general terms, borers can be distinguished from miners. Typically, larvae of miners feed within plant foliage, whereas larval borers may feed within other plant tissues, including roots, stems and twigs, meristems, fruit, conductive tissues, galls, and bark. The variety of plant tissues that are used by borers also spans an array of plant groups that range from the ferns and gymnosperms to the grasses and dicotyledonous angiosperms.

All plant tissues may be subject to borer colonization. Larvae of a number of families (e.g., Cerambycidae and Hepialidae) may construct feeding tunnels, or galleries, within the large roots of broadleaf trees and conifers, which may weaken the trees directly or provide entry points for invasion by pathogenic fungi. At the other extreme, there are many species of insects that colonize the meristematic tissues at branch terminals, tips, twigs, and canes. Some of these insects feed in the phloem tissues girdling the twigs, whereas larvae of other species burrow through the growing tips and into the elongating stems. These types of larval feeding can reduce plant growth, apical dominance, and plant form.

FIGURE 1 Adult pairs of the European elm bark beetle *S. multistriatus* excavate parental galleries in the cambium and phloem of host elm trees. Larvae eclose from eggs laid in niches cut into the margins of the parental galleries and construct feeding galleries that extend laterally into the same host tissues.

Between the twigs or apical tips and the roots is the main stem or trunk of the tree. The woody xylem tissues, cambial layers, phloem tissues, and bark may all have different groups of specialist borers. For example, larvae of a few species of clearwing moths feed within the bark of their host plant. Scribble-barked gums are species of *Eucalyptus* in Australia that derive their common name from the twisting galleries constructed in the outer bark by lepidopteran larvae. In different feeding strategies, bark beetle larvae feed within the cambial and phloem tissues of their hosts (Fig. 1), whereas larvae of many species of longhorned and flatheaded borers feed in the outer layers of phloem and cambium but then bore deep into the wood to pupate. Alternatively, many other cerambycid and buprestid species feed almost entirely within the wood of their host trees. Larvae of cossid moths also feed entirely within woody tissues and may take several years to complete their larval development. Woody plant tissues are not as rich in nutrients as the cambial tissues and the quality deteriorates as the tissues age. Consequently, many wood borers may have prolonged larval development and long generation times.

Not only the larvae bore into woody tissues but also the adults in a number of species within a variety of families (e.g., some species of Scolytidae and Platypodidae) bore into the plant. The larvae of ambrosia beetles are found in galleries excavated within the wood, but they feed on a fungus inoc-

ulated into the tissue by the parental adults rather than on the plant itself. The fungi acquire nutrients from a large volume of plant tissue as the hyphae ramify throughout the wood.

PLANT CONDITION

Plants in a wide range of physiological conditions may be subject to colonization by borers. Although some species of borers use healthy hosts or healthy host tissues, plants that are suffering from some type of stressful condition either attract or inhibit further dispersal behavior by many other borer species. Insects that bore into tender tips and stems frequently colonize young and vigorously growing plants. Consequently, younger plants may suffer more damage than mature plants. Open wounds or stressed, damaged, or weakened plant tissues may be subject to invasion. Weakened or stressed host plants may result from chronic growing conditions (poor-quality site) or from acute detrimental changes (e.g., fire, flood, drought, lightning strikes). Infections by pathogens, particularly plant pathogenic fungi, nematodes, and parasitic plants, weaken host plants and increase their susceptibility to subsequent borer infestations. In addition, previous infestation by other insect herbivores may weaken the host plant and increase susceptibility to subsequent borer colonization. Recently killed and dying trees are particularly suitable for colonization by a range of borers. For example, there are several species of wood wasps and flatheaded wood borers that are attracted to trees that have been recently killed by fires. A wide range of borers have developed complex relationships with tree-killing pathogens and are responsible for transmitting the pathogens into the host trees. Scolytid bark beetles transmit species of pathogenic *Ophiostoma* and *Ceratocystis* fungi into a variety of hosts (e.g., *Scolytus scolytus, S. multistriatus,* and *Hylurgopinus rufipes* transmit *Ophiostoma ulmi,* the causal agent of Dutch elm disease). Cerambycids in the genus *Monochamus* are responsible for transmitting the nematode, *Bursaphelenchus xylophilus,* the pathogen causing pine wilt disease, into susceptible host pines. Females of the European woodwasp, *Sirex noctilio,* inject a phytotoxin and spores of the pathogenic fungus *Amylostereum areolatum* into susceptible host trees during oviposition. Borers that are adapted to colonize the woody tissues of dead or dying plants may also colonize trees that have been cut during commercial logging or even timber that has been milled into lumber. It is not unusual for adult borers to emerge from products or materials constructed from infested wood that has not been kiln dried or otherwise treated to kill the infesting insects.

See Also the Following Articles

Forest Habitats • Integrated Pest Management • Plant Diseases and Insects

Further Reading

Creffield, J. W. (1996). "Wood Destroying Insects: Wood Borers and Termites," 2nd ed. CSIRO Australia, Collingwood.

Furniss, R. L., and Carolin, V. M. (1980). "Western Forest Insects." U.S. Department of Agriculture Forest Service, Washington, DC. [Miscellaneous Publication 1339].

Hanks, L. M. (1999). Influence of the larval host plant on reproductive strategies of cerambycid beetles. *Annu. Rev. Entomol.* **44,** 483–505.

Johnson, W. T., and Lyon, H. H. (1988). "Insects That Feed on Trees and Shrubs." Cornell University Press, Ithaca, NY.

Paine, T. D., Raffa, K. F., and Harrington, T. C. (1997). Interactions among scolytid bark beetles, their associated fungi and live host conifers. *Annu. Rev. Entomol.* **42,** 179–206.

Solomon, J. D. (1995). "Guide to the Insect Borers of North American Broadleaf Trees and Shrubs." U.S. Department of Agriculture Forest Service, Washington, DC. [Agricultural Handbook 706]

U.S. Department of Agriculture Forest Service. (1985). "Insects of Eastern Forests." U.S. Department of Agriculture Forest Service, Washington, DC. [Miscellaneous Publication 1426]

Brain and Optic Lobes

Nicholas J. Strausfeld
University of Arizona

Authors variously use the term brain either to include all neuropils located within the head capsule or, restrictively, to refer to only those neuropils (called preoral neuropils) that lie dorsal to the esophagus. These are considered to lie anterior to the mouth. Preoral neuropils are also known as the supra-esophageal ganglion, which comprises three fused ganglia: the protocerebrum, deutocerebrum, and tritocerebrum. The preoral brain of the larger Hymenoptera, such as the predatory wasp *Pepsis thisbe,* can contain well over a million neurons, with more than a third of a million neurons in each mushroom body. The extreme density of neurons packed into a small volume, and the likelihood that single nerve cells can be functionally divided into several discrete elements, suggests that the largest insect brains have impressive computational power.

The first definition of the brain includes neuropils of the subesophageal ganglion, which is composed of the fused ganglia from three postoral segmental neuromeres. These are located ventrally with respect to the digestive tract, as are ganglia of the thorax and abdomen. In most hemimetabolous insects, and in many paleopterans, the subesophageal ganglion is connected by paired circumesophageal commissures to the supraesophageal ganglion. In many crown taxa (those representing more recent evolved lineages) the subesophageal and supraesophageal ganglia are fused, as is the case in honey bees or the fruit fly *Drosophila melanogaster,* which is the taxon here used to summarize the major divisions of the brain (Figs. 1–6). A consequence of fusion is that tracts of axons that would otherwise form the circumesophageal commissures are embedded within a contiguous neuropil.

In insect embryos, the three preoral segmental neuromeres providing the proto-, deuto-, and tritocerebrum are contiguous

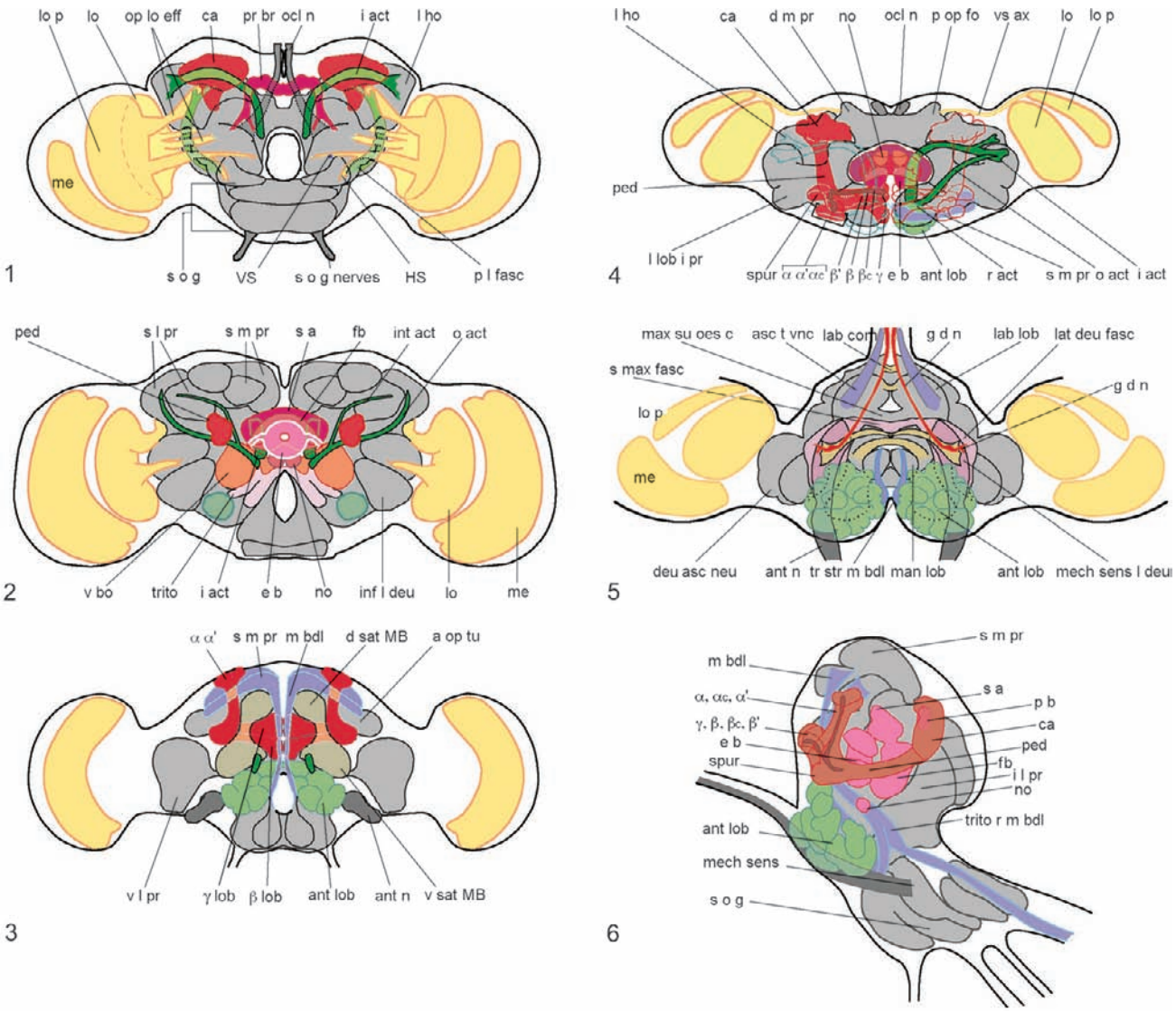

FIGURES 1–6 Summary diagrams of the brain of the fruit fly *D. melanogaster.* The reader is referred to these searchable figures in the atlas of the *Drosophila* brain at http://www.flybrain.org/Flybrain/html/atlas/schematic/index.html. The higher centers of the mushroom bodies and central complex are shown in reds and oranges. Optic lobe regions are yellow. Antennal lobes are light green and their axon projections are dark green. The median bundle is shown in light blue. Other neuropil areas are gray/pink. (1) Posterior aspect, vertical section. According to the neuraxis (see Fig. 7), up is rostral in Figs. 1–3. (2) Middle aspect, vertical section, at the level of the central body and associated regions. (3) Frontal aspect, vertical section, at the level of the antennal lobes (green) and mushroom body lobes (red). Dark green profiles in Figs. 1–3 are the antennocerebral tracts. (4) Top-down view, looking onto the mushroom bodies and central complex. One mushroom body only is shown to the left, with the antennocerebral tracts from the antennal lobes to the lateral protocerebrum shown to the right. The front of the brain is down, the rear of the brain is up. (5) Top-down view of the deutocerebrum/tritocerebrum and the root of the ventral nerve cord. (6) Side-on view of the brain, emphasizing the ascending tracts (blue) from the subesophageal ganglion to the rostral protocerebrum via the medial bundle. Note the disposition of the mushroom body and central complex. Abbreviations used: **α, α′, αc, (β, β′,βc, γ) lobes,** subdivisions of the mushroom body medial (β, β′,βc, γ) and vertical (α, α′, αc) lobes; **ant n,** antennal nerve; **ant lob,** antennal lobe; **a op tu,** anterior optic tubercle (optic glomerulus); **asc t vnc,** ascending tracts of ventral nerve cord; **ca,** calyx of mushroom body; **deu asc neu,** deutocerebral neuropil receiving ascending terminals; **d m pr,** dorsal median protocerebrum; **e b,** ellipsoid body of the central complex; **fb,** fan-shaped body of the central complex; **g d n,** giant descending neuron (typifies descending pathways); **inf l deu,** inferior lateral deutocerebrum; **i act,** inner antennocerebral tract; **int act,** intermediate antennocerebral tract; **l lob i pr,** lateral lobe of the inferior protocerebrum; **lat deu fasc,** lateral deutocerebral fascicle; **lab lob,** labral lobe; **lab com,** labral commissure; **lo,** lobula; **lo p,** lobula plate; **l ho,** lateral horn; **max su oes c,** maxillary subesophageal connective; **me,** medulla; **mech sens l deu,** mechanosensory neuropil of the lateral deutocerebrum; **mech sens,** mechanosensory strand and neuropil supplied by the antennal nerve; **m bdl,** median bundle; **no,** noduli of the central complex; **op lo eff,** optic lobe efferents; **ocl n,** ocellar nerve; **o act,** outer antennocerebral tract; **pr br,** protocerebral bridge of the central complex; **p l fasc,** posterior lateral fascicle; **p op fo,** posterior optic focus (glomerulus); **r act,** root of antennocerebral tract; **s a,** superior arch of the central complex; **s o g,** subesophageal ganglion; **s l pr,** superior lateral protocerebrum; **s m pr,** superior median protocerebrum; **s o g nerves,** nerve bundles of subesophageal neuromeres; **spur,** spur of mushroom body; **trito,** tritocerebrum; **tr str m bdl,** tritocerebral strand of the median bundle; **trito r m bdl,** tritocerebral root of the median bundle; **VS, HS,** axons of giant vertical and horizontal cells (movement sensitive neurons); **vs ax,** visual interneuron axons; **v bo,** ventral body (also known as lateral accessory lobes); **v sat MB, d sat MB,** ventral and dorsal satellite neuropils of the mushroom bodies.

with the three postoral neuromeres that will give rise to the subesophageal ganglion. These neuromeres are, in turn, contiguous with fused neuromeres of the thorax and abdomen. In many species of hemimetabolous insects, such as locusts and cockroaches, the sub- and supraesophageal ganglia separate postembryonically and are connected by paired tracts. In cockroaches, each segmental ganglion is separate from the next, except for the last three abdominal ganglia, which are specialized to serve receptors of the cerci and contain the dendrites of giant ascending neurons and local networks of interneurons that mediate escape reactions. However, in many holometabolous insects there are various degrees of ganglion fusion, one of the most extreme being in certain Heteroptera such as the water strider *Gerris* sp. In *Gerris,* the supraesophageal, subesophageal, and thoracic–abdominal ganglia comprise a contiguous mass perforated by the gut. This arrangement is reminiscent of the nervous systems of another group of arthropods, the chelicerates. In adult cyclorrhaphan flies the three thoracic ganglia and all abdominal ganglia are fused into a single mass connected to the sub- and supraesophageal ganglion by long neck connectives (this has also been achieved in the nervous systems of crabs).

The subesophageal ganglion, which comprises the mandibular, maxillary, and labial neuromeres, has a ground pattern organization comparable to that of the thoracic and abdominal ganglia. The roots of motor neurons (the exit point of motor neuron axons) are generally dorsal with respect to incoming sensory axons. This arrangement is the opposite of that in the vertebrate spinal cord.

The names of the subesophageal ganglia reflect the appendages that their motor neurons control and from which they receive sensory supply. However, this relationship is not a strict one. For example, in flies mechanosensory neuropil extending into the subesophageal ganglion also receive afferents from mechanosensilla on the head, including between facets of the compound eyes, around the margin of the eyes, the frons, between and flanking the ocelli, and at various positions on the rear of the head capsule. As on the thorax and abdomen, or on the limbs, wings, and halteres (modified wings in Diptera that are organs of balance), sensilla on the head provide receptor neuron axons to defined locations in their target ganglia. Principles underlying the development and organization of the central representation of sensilla are best known from Walthall and Murphey's 1988 studies of cricket cerci or studies on the central projections of receptors to discrete regions of the thoracic ganglia of dipterans, also by Murphey and colleagues in 1989. In flies, groups of receptors encoding different modalities at a segment supply axons to modality-specific regions within the ganglion. In such regions, the peripheral locations of receptors within a sensory field can be represented as a map of axon terminals onto the dendritic trees of postsynaptic neurons. Burrows and Newland have shown that such maps play important roles in the activation of the postsynaptic elements that participate in circuits controlling limb actions and position.

THE PREORAL BRAIN

Although there has been in the past endless debate about how many segments contribute to the head and to the brain, expression of homeobox genes now confirms three embryonic brain segments only. The main issue of contention focused on the neuromeric identity of the optic lobes, which were claimed by some to have a distinct segmental origin. Developmental studies, which are summarized by Meinertzhagen and Hanson's 1993 review, showed that the inner optic lobe neuropils in the adult insect brain develop from a lateral outgrowth of the protocerebrum. George Boyan and his colleagues at the University of Munich have provided crucial evidence supporting a three-neuromere origin of the supraesophageal ganglion from studies of the segment polarity gene *engrailed,* which is expressed in cells (including neuroblasts) in the posterior compartment of each segment. The expression of *engrailed* in the first wave of neuroblast generation shows the delineation of the tritocerebrum from the first (maxillary) subesophageal neuromere, as well as the delineation between the tritocerebrum and the deutocerebrum and the delineation between the deutocerebrum and the protocerebrum, the last being the most rostral segmental neuromere. Crucially, the expression of *engrailed* shows the latter to be segmentally indistinct from the developing optic lobes.

THE PROTOCEREBRUM

The ground structure of the protocerebrum suggests its ancestral affinities with segmental ganglia. In the protocerebrum, as in postoral ganglia, ascending sensory interneuron tracts enter it ventrally, whereas premotor interganglionic interneurons exit dorsally. Afferents (here the optic lobe output neurons; see later) distribute to local interneurons in a manner reminiscent of sensory afferents within postoral ganglia.

Despite its basic similarities with segmental ganglia, the protocerebrum contains neuropils that are not normally found in other segments and appear to have no counterparts in other ganglia, unless generated ectopically by genetic manipulation. Unique protocerebral neuropils comprise: (1) the central complex and (2) the mushroom bodies and some satellite neuropils belonging to both of these. A midline indentation between the two protocerebral lobes, called the pars intercerebralis, with its accompanying populations of neuromodulatory neurons, may also be unique to the protocerebrum. But without the relevant developmental studies on thoracic ganglia, it is not clear whether any of the unique clusters of neurons at their dorsal midlines are segmental counterparts of neurons at the pars intercerebralis.

The structure of the protocerebrum is best approached by understanding the basic organization of major axon tracts that extend between its two halves. Studies of *Drosophila* embryos show that major cerebral tracts appear early in

development and pioneer the trajectories of interneurons linking later developing neuropil regions. Again, research by Boyan and colleagues on the development of locust embryos provides important insights into early brain development and demonstrates that neurons developmentally ascribed to the protocerebrum can actually end up distant from it in the adult. For example, studies of *engrailed* expression show that in locusts the first episode of neuroblast generation in the protocerebrum includes three neuroblasts that migrate caudally to lie beside the glomerular antennal lobes, which are structures usually ascribed to the deutocerebrum. The segmental origin of these three neuroblasts, which contribute neurons to the antennal lobe system, cautions against uncritically ascribing segmental identities to neurons in the adult brain.

Nevertheless, many of the tracts and neuropils described from the adult have been both ascribed to one of its segments and named, even though only a few are yet understood at a functional and developmental level. The reader is referred to two brain atlases, one by Strausfeld and the other an electronic publication, FLYBRAIN (www.flybrain.org), both of which focus on the adult structure of dipteran brains (the housefly *Musca domestica* and the fruit fly *D. melanogaster*). The basic divisions of the *Drosophila* brain are shown in Figs. 1–6. The following neuropils, or neuropil groups, comprise the salient regions of the adult protocerebrum.

The Central Complex

Insect and crustacean protocerebra contain unique midline neuropils and, in more advanced taxa, satellite neuropils associated with them. These structures and their interrelationships have been described by several authors, one cardinal study being by L. Williams in 1975. This and other studies summarized here are described in some detail in a recent article by Strausfeld published in 1999.

The midline component of the central complex, called the central body, is similar to a unique midline neuropil in the brains of chilopods, branchiopod crustaceans, and archaeognathan insects. Comparative studies suggest that these neuropils have become elaborated through time. In the flightless Zygentoma (e.g., "silverfish") as well as in the Palaeoptera (e.g., mayflies and dragonflies), several paired satellite neuropils are reciprocally connected to two midline neuropils: the columnar ellipsoid body and, above it with respect to the brain's neuraxis, the fan-shaped body, which is usually recognized by its scalloped profile. Further elaboration has occurred in the Neoptera, in which a distinct stratum called the superior arch is attached to the fan-shaped body anteriorly. A bridge of neuropils, called the protocerebral bridge, connects the two protocerebral lobes and provides axons that extend into the fan-shaped body and to the ellipsoid body behind it. In many taxa (e.g., locusts, flies, wasps) the protocerebral bridge is divided into 16 discrete modules, 8 each side of the midline. These connect to the 16 modules of the fan-shaped body and ellipsoid body. The most lateral module on one side of the bridge is linked to the most medial module of the other side. The next most lateral module is linked to the next most medial one, and so on. These connections provide an elaborate pattern of chiasmata between the bridge and the fan-shaped and ellipsoid bodies. Some of these neurons also extend to a pair of ball-like structures, called the noduli, situated caudally with respect to the fan-shaped and ellipsoid bodies. Two synaptic zones in the noduli, a core and an outer layer, receive connections from the fan-shaped body such that one-half is represented in the core of the contralateral nodulus, whereas the other half is represented in the outer layer of the ipsilateral nodulus. A recent account by Renn and colleagues (see FLYBRAIN database) uses genetic markers to dissect these various components and trace their development.

The protocerebral bridge receives a system of elongated fibers from the medial protocerebrum, which is itself supplied by terminals of ascending interganglionic interneurons that originate in thoracic ganglia. These elongated protocerebral fibers extend through dendritic trees that contribute to the modules across the bridge and are assumed to provide inputs to their dendrites, although this awaits confirmation.

The superior arch appears to be distinct from protocerebral bridge inputs and is connected heterolaterally to neuropils of the protocerebral lobes, themselves receiving terminals from the median bundle, a midline tract originating from the subesophageal ganglion and ventral cord and ascending along the midline of the ventral surface of the protocerebrum. The superior arch shares local interneurons with the fan-shaped body and ellipsoid body.

The fan-shaped body and ellipsoid body each receive fan-like terminals from axons that originate at dendritic trees in various lateral neuropils of the protocerebrum. Both the ellipsoid and the fan-shaped bodies supply outputs that extend to lateral protocerebral neuropils, particularly a ventrocaudal region called the ventral bodies, known also as the lateral accessory lobes. These lobes are invaded by the dendritic collaterals of many of the descending neurons leaving the brain for thoracic and abdominal ganglia.

The central complex is strictly a higher center that is distant from sensory inputs. Dye fills fail to demonstrate any sensory interneuron inputs to central complex neuropils nor do the antennal lobes or optic lobes provide direct connections to the central complex. The central complex has no direct connections with the mushroom bodies. Instead, various regions of the protocerebrum that are connected to the central complex are also connected to the mushroom bodies and to higher level sensory neuropils, such as the lobula of the optic lobes and antennal and vertical lobes of the deutocerebrum (see later).

As described in Nässel's 1993 review, the central complex is richly supplied by peptidergic neurons that originate from the pars intercerebralis. The pars also provides a wealth of peptidergic neurons whose axons leave the brain for the retro-

cerebral complex (the corpora allata and corpora cardiaca) via the corpora cardiaca nerve NCC1. Other neurosecretory cells sending axons out of the brain lie lateral and rostral with respect to the protocerebrum (NCC2) and in the lateral tritocerebrum (NCC3). An exquisitely detailed enhancer trap analysis of these systems has been published by Siegmund and Korbe. The central complex is implicated in the control of motor actions, although exactly what it controls is not yet known. Studies of motor-coordination-defective *Drosophila* show that certain behavioral mutants have midline lesions of their protocerebral bridge or the fan-shaped bodies. Roland Strauss, at the University of Würzburg, has shown that these mutant flies are incapable of adjusting step length during turning. A similar disruption across the midline occurs in nature: certain rowing Heteroptera, such as the water strider *Gerris,* have split protocerebral bridges, minute noduli, and reduced modules in their fan-shaped and ellipsoid bodies. In contrast, insects that show sophisticated asymmetric but highly coordinated limb actions, such as are employed in grooming, object manipulation, or cell construction, possess elaborately modular central complexes and complete protocerebral bridges.

Prominent connections between the fan-shaped and the ellipsoid bodies with the lateral accessory lobes of the protocerebrum are of functional interest. The lateral accessory lobes are visited by dendritic collaterals from many of the interganglionic descending interneurons that send axons from lateral protocerebral regions to neuropils of the thoracic and abdominal ganglia. One interpretation of this organization is that the central complex plays a role in gating outgoing information from the brain.

The central complex is richly supplied from protocerebral regions involved in sensory discrimination. This organization, with the elaborate arrangement of repeat units (modules and chiasmata) within and between the protocerebral bridge and midline neuropils, might suggest that the central complex assesses the context and occurrence of sensory stimuli around the animal and that this plays a crucial role in modifying descending information to motor circuits.

Mushroom Bodies

The mushroom bodies, discovered by Félix Dujardin in 1850, were the first brain centers to be recognized as distinct entities. Dujardin's suggestion that mushroom bodies supported intelligent actions was with reference to social insects, in which mushroom bodies are largest and most elaborate. Since his 1850 paper, the mushroom bodies have been considered to be centers crucial to learning and memory.

The mushroom bodies are paired lobed neuropils. Comparative studies by Strausfeld, Ito, and others have identified mushroom bodies in all groups of insects except the archaeognathan. In zygentoman and palaeopteran insects, mushroom bodies comprise two sets of lobes, one set extending medially toward the midline (medial lobes), the

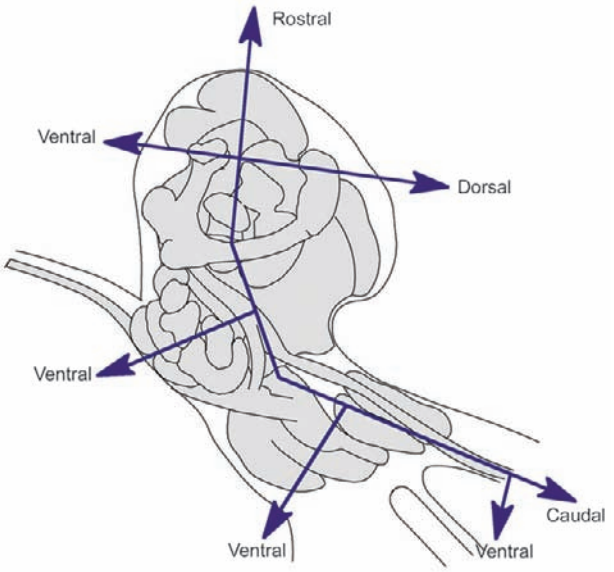

FIGURE 7 Neuraxis. Many descriptions in the literature rarely make the point that the brain's orientation is not that of the body. During postembryonic development, the brain undergoes morphogenic movements, tilting upward and back. This brings the dorsal surface of the brain to face caudally with respect to the body's axis. The front of the brain is its ventral side according to the neuraxis. The top part of the brain is rostral. Likewise, the antennal lobes are ventral, not frontal.

other extending ventrally, with respect to the neuraxis. However, because most brains are tilted upward (Fig. 7), these lobes can point forward or even upward. They are thus collectively referred to as the vertical lobes.

The mushroom body lobes comprise many thousands of approximately parallel-running processes. These originate from clusters of minute globuli cells situated dorsorostrally in the protocerebrum's cell body rind. In neopteran insects, these neurons have distal dendritic trees that contribute to rostral neuropils called the calyces. Each mushroom body has a pair of calyces, each of which is divided into two halves. A crucial study by Kei Ito and colleagues demonstrated that each half is generated by one of a quartet of embryonic neuroblasts. The four half-calyces are supplied by four lineages of globuli cells, all of which provide dendrites in the calyces and long axon-like processes in the lobes. These "intrinsic neurons" of the mushroom bodies are known as Kenyon cells, named after their discoverer. Lineage analysis of the *Drosophila* mushroom bodies has shown that each of the four neuroblasts generates the same sequence of Kenyon cells, certain types of which differentiate before others. Different types of Kenyon cells contribute to different and discrete subdivisions of the lobes. Farris's studies on the mushroom bodies of the cockroach *Periplaneta americana* and the worker honey bee *Apis mellifera* have shown that the sequence of Kenyon cell production and segregation to subdivisions is similar to that in *Drosophila* described by Lee *et al.* in 2001.

Observations of the cockroach and honey bee calyces show these neuropils as organized into nested zones, each of which is defined by the types of afferents supplying it. The

most up-to-date study on the honey bee demonstrates that an outer region called the lip comprises three zones, each of which receives axon collaterals from neurons that project from glomeruli of the antennal lobes (see later) to regions of the lateral protocerebrum. A second region of the calyx, which comprises the collar, is further divided into discrete zones, each of which is defined by visual and other afferent endings, such as from gustatory neuropils of the subesophageal ganglion. However, in many other insect orders, the calyces receive sparse inputs, if any, from the visual system. A central region of the calyx called the basal ring is similarly divided into modality-specific zones.

Kenyon cells having dendrites in one of these zones send their axon-like processes into a specific stratum that extends all the way through the vertical and medial lobes. Each stratum thus represents a zone of the calyces. However, a special class of Kenyon cells that is generated earliest in development supplies axon-like processes to a separate division of the mushroom bodies, called the gamma lobe. Depending on the taxon, this lobe lies parallel to the medial (flies), vertical (honey bees), or both (cockroach) lobes. Important studies by Zars and Heisenberg on gene expression in different parts of the mushroom body of *Drosophila* have implicated the gamma lobe in supporting short-term memory.

One pervasive misconception is that the calyces are the "input region" of the mushroom bodies, whereas their lobes are their output regions. This view of the mushroom body does much to confuse and mislead theoretical considerations about how the mushroom bodies might work. Palaeopteran insects lack calyces supplied by sensory interneurons, yet their lobes both receive afferent endings from other protocerebral neuropils and provide efferents that extend back to protocerebral neuropils. In neopteran insects, the lobes likewise receive inputs and provide outputs, with the axon-like processes of Kenyon cells providing local circuits between them. However, in neopterans, Kenyon cells also supply calyces with dendrites that are visited by sensory interneurons. The role of the calyces is not fully understood. Possibly, afferents ending on Kenyon cell dendrites serve to modify the activity of local circuits in the lobes that are supplied by the processes of Kenyon cells, thereby providing sensory context dependence to computations that occur via Kenyon cell processes between inputs and outputs at the mushroom body lobes. It is also possible that inputs to the calyces provide persisting memory-like alterations of groups of Kenyon cells. Peptidergic and other modulatory neurons (e.g., octopaminergic, dopaminergic) associated with the mushroom bodies have been implicated in memory formation, and genetic disruption of vesicle recycling in a modulatory neuron of the *Drosophila* mushroom body shortens memory. It is still somewhat of a mystery why there are two sets of lobes, with most Kenyon cell processes dividing into each of them. However, as shown by Pascual and Préat, working at the CNRS in France, if the vertical lobes are absent, as in one type of *Drosophila* mutant, then long-term memory cannot be established. A role for the mushroom bodies in learning and memory has also been suggested by chemical ablation of the mushroom body neuroblasts, and a consequent lack of the mushroom body abolishes olfactory associative learning by the adult fly. However, such ablations also remove a set of local circuit neurons in the antennal lobes, complicating the interpretation of such experiments. A further complication in interpreting the mushroom body's role in memory acquisition is Dubnau's recent finding that synaptic transmission by mushroom body neurons is necessary only for memory retrieval and not for memory formation.

The Rostral Lateral Protocerebrum and Lateral Horn

The protocerebrum is composed of many discrete centers, most of which do not have obvious order and neat geometries, as in the mushroom bodies and central complex. Nevertheless, each protocerebral center is a unique entity and specific centers can be identified across different species. It is likely that studies of enhancer trap lines, as well as genetic labeling of clonally related neurons, developed by Liqun Luo and his colleagues, will in the near future reveal many new features of the cellular organization of the protocerebrum. But, so far, few studies have been done on these neuropils even though they together impart great complexity to the brain. This section focuses on just two neighboring regions, the lateral protocerebrum and lateral horn, which are now known to be second-order olfactory neuropils.

Antennal lobe projection neurons relay information from olfactory glomeruli to various areas of the brain, via three axon bundles called the inner, intermediate, and outer antennocerebral tracts. Axons of the inner antennocerebral tract provide axon collaterals to the mushroom body calyces. However, olfactory projection neurons providing input to the calyces do not terminate there but end in a region of the protocerebrum called the lateral horn and, caudally adjacent to it, the lateral protocerebrum. Axons of the intermediate and outer antennocerebral tracts also invade these neuropils, which therefore must be considered second-order olfactory processing centers of the brain. In honey bees, certain axons of the intermediate tract also target some neuropils that lie in front of and beneath the calyces as well as neuropils enwrapping the vertical lobes. Studies from Liqun Luo's laboratory at Stanford University have now shown that discrete fields of endings in the lateral horn and lateral protocerebral areas lying immediately caudal to it are supplied by specific groups of antennal lobe glomeruli, thus showing that the olfactory map that occurs among antennal lobe glomeruli is partially maintained within this lateral protocerebral area. With the exception of the calyces, neuropils targeted by antennal lobe projection neurons are second-order olfactory centers. These neuropils are not, however, unimodal olfactory centers as they also receive inputs from the optic lobes via large ascending fascicles.

The lateral horn and lateral protocerebrum give rise to systems of local interneurons as well as long-axoned interneurons, certain of which terminate in the mushroom body lobes. However, the relationship of the lateral protocerebrum with descending pathways is not yet known. A further area of ignorance is its relationship with the central complex.

OPTIC LOBES

Studies on the fly visual system dominate the literature on insect vision, and descriptions of visual regions have been mainly from dipteran insects, whose structure and physiology have been described in great detail in numerous papers.

Retinotopic Organization

The optic lobes of palaeopteran and neopteran insects consist of three retinotopic neuropils. These are the lamina, medulla, and lobula complex. In certain orders of insects (e.g., Diptera Lepidoptera, Coleoptera) the lobula complex is divided into two separate neuropils: a lenticular lobula that is mainly composed of columnar neurons and a tectum-like lobula plate that is hallmarked by wide-field tangential neurons. However, in insects with an undivided lobula, deeper layers comprise tangential neurons that probably have the same functions as tangential neurons in the lobula plate. Connections between the medulla and the lobula plate in Diptera are homologous to connections between the medulla and the deep lobula layers in honey bees. The lobula plate or its equivalent supports achromatic motion vision, whereas the lobula is thought to support object and color vision. The lobula plate sends axons to dorsal neuropil of the lateral and medial protocerebrum from which descending neurons supply neuropils belonging to the neck and flight motor systems. The lobula supplies bundles of axons to discrete glomerular-like neuropils of the ventrolateral protocerebrum, certain of which retain retinotopic organization. These optic glomeruli (also called optic foci) are invaded by processes of local interneurons and relay neurons.

The Lamina

The lamina is the first neuropil of the optic lobes and the best known with respect to its cellular organization, synaptology, and development. It consists of relatively few types of neurons whose relationships achieve surprising complexity. These have been reviewed by several authors, including Nässel and Strausfeld and Meinertzhagen and Sorra. The following summary focuses on the fly's visual system because of its preeminence in vision research.

Each ommatidium of the compound eye contains eight (in honey bees nine) receptor cells. As summarized in Hardie's 1986 review, the axons (called short visual fibers) of six blue-green sensitive photoreceptors (termed R1–R6) end in the lamina, whereas the axons (long visual fibers, termed R7 and R8) of the other two photoreceptors (blue and UV sensitive) terminate in the medulla. In insects equipped with apposition eyes or neural superposition eyes, a set of six short visual fibers terminate at each columnar subunit of the lamina, called an optic cartridge. The six endings belong to six photoreceptors that share the same optical alignment and thus "look" at the same restricted area of visual space even though, in neural superposition eyes, the optically coherent photoreceptors are distributed among six ommatidia, a discovery made by Kuno Kirschfeld in 1967 and elegantly explained from a developmental view point by Meinertzhagen in his 2000 review article. Nicholas Franceschini has termed an optically coherent set of receptors a visual sampling unit or VSU.

Each VSU is associated with several classes of interneurons, the most prominent of which are cells known as large monopolar cells or LMCs. In Diptera, LMCs include the pair of radial monopolar cells (also called L1 and L2) that are postsynaptic to all six terminals of a VSU. A third monopolar cell (the brush or unilateral monopolar cells, also called L3) is postsynaptic to all six receptors of a VSU but has about two-thirds fewer dendrites than either L1 or L2. This triplet of neurons sends axons to the medulla, alongside the axons of the R7 and R8 receptor neurons that originate from the optically relevant ommatidium. In insects with superposition eyes, the lamina is not obviously divided into cartridges and the dendrites of second-order neurons (monopolar cells) extend across many photoreceptor terminals.

In addition to the L1–L3 monopolar cells, the lamina contains two types of amacrine neurons whose processes provide local interactions between receptor terminals of several cartridges and the dendrites of other efferent neurons that leave each cartridge. These are the types L4 and L5 monopolar cells and the basket cell, referred to as the T1 efferent neuron. Amacrine cells, which are local interneurons that lack axons and provide local circuits, are postsynaptic to receptors from many VSUs and are presynaptic onto the dendritic trees of T1 and dendrites of L4 neurons. Each L4 neuron contributes axon collaterals to a rectilinear network of connections beneath the lamina. These collaterals are presynaptic to the axons of L1 and L2 and reciprocally pre- and postsynaptic to the L4 neuron collaterals of neighboring cartridges. A fifth species of efferent neuron, called the midget monopolar cell, or L5, has one or two minute tufts of dendrites that are visited by a second species of amacrine cell. A comprehensive review by Strausfeld and Nässel demonstrates that similar types of neurons have been identified in crustaceans and in honey bees.

Four distinctive types of centrifugal cells visit the lamina. Two are associated with each cartridge. One is the type C2 centrifugal cell, which is presynaptic onto the L1 and L2 monopolars at a level above their dendrites. The second is the type C3 centrifugal cell, which is presynaptic to the L1 and L2 monopolar cells at the level of their dendrites. Both C2 and C3 are GABAergic and both have dendrites at various levels

in the outer and inner medulla. Two wide-field centrifugal cells, called the types 1 and 2 lamina tangential cells, have bistratified and concentric dendritic fields in the medulla and send axons back out to the lamina, where their terminals provide isomorphic plexi of presynaptic endings.

The Optic Chiasmata

Axons originating from the front of the lamina end in the back of the medulla; axons from the back of the lamina end in the front of the medulla, so reversing the horizontal order of vertical rows of VSUs. The order is rereversed in the lobula by a second chiasma beneath the medulla.

THE MEDULLA The arrangement of retinotopic columns in the medulla is defined by pairs of long visual fiber endings, each pair accompanied by the sextet of endings of neurons leaving the optically corresponding optic cartridge (L1–L5, T1). These endings terminate at specific levels of the medulla where they coincide with the processes of amacrine cells, the dendrites of relay neurons, and dendrites and terminals of tangential neurons. The medulla is immensely complicated. Each column may contain as many as 40 relay neurons and many strata of amacrine cells intersect these. Thus, the medulla has one of the highest densities of nerve cells outside the mushroom bodies.

A broad layer of incoming and outgoing axons belonging to tangential neurons separates the outer two-thirds of the medulla from its inner third. Tangential cells have dendritic trees or terminal fields that extend across many retinotopic columns. Axons of centripetal tangential cells project centrally via the posterior optic tract, to reach ipsi- and or contralateral regions of the dorsal protocerebrum. Some tangentials connect the medullae of the two lobes. Tangential endings in the medulla derive from dendrites within the medial protocerebrum. These neurons carry processed information about motion stimuli back out to peripheral layers of the visual system.

Retinotopic columns may each consist of as many as 40 different morphological types of axonal neurons. These neurons have dendrites at characteristic levels, spreading through defined fields of neighboring columns. Each morphological cell type sends its axon to a characteristic deeper level of the optic lobe. The class of neurons called transmedullary cells (Tm cells) sends its axons to various depths of the lobula. A subset of transmedullary cells, whose dendrites are restricted to within a retinotopic column, supplies a special sheet of synaptic neuropil over the surface of the lobula where they end among quartets of bush-like dendritic trees called T5 cells. These neurons were shown by J. K. Douglass to be the first in the system to exhibit directional- and orientation-selective responses to moving visual stimuli. Their axon terminals segregate to four levels in the lobula plate where they end on layered systems of tangential neurons. Pioneering studies by Klaus Hausen, at the Max Planck Institute in Tübingen, demonstrated that these large-field neurons respond selectively to wide-field directional motion across the retina, relaying this information to the midbrain and to the contralateral optic lobes so that signals from both eyes can be integrated.

A second class of transmedullary cells (called intrinsic transmedullary neurons) serves to link the outer layer of the medulla with its inner layer. A third class of transmedullary cells, called Tm Y cells, consists of neurons whose axons branch, one tributary reaching the lobula plate and the other reaching the lobula. Retinotopic neurons also originate from the inner layer of the medulla. These must derive their inputs from other transmedullary cells because their dendrites lie beneath afferents from the lamina. Neurons from the inner medulla send axons to the lobula or provide bifurcating axons to the lobula plate and lobula.

The lobula contains ensembles of tree-like neurons organized retinotopically. The spacing of their axons coarsens the original retinotopic mosaic so that an oval ensemble of retinotopic inputs from the medulla visits the dendritic tree of a single lobula neuron. The size of the dendritic fields of these columnar neurons varies, however, depending on the cell type. The smallest lobula neuron has a dendritic field equivalent to an approximately oval array of nine VSUs. However, it cannot be assumed that such fields are the physiologically receptive fields because the functional organization among their inputs is not known.

Columnar neurons in the lobula are likely to be tuned to highly specific visual features although few recordings have been made from medulla neurons supplying the lobula. Those that have been recorded suggest that the lobula receives information about orientation but not about wide-field motion. However, the presence in the lobulas of male flies of sex-specific neurons that respond to movement of small objects in the visual field must imply that a class of directionally selective neurons from the medulla supplies at least a part if not all of the lobula.

Each ensemble of identically shaped columnar neurons provides a coherent bundle of axons that targets a circumscribed region of neuropil in the lateral protocerebrum. Such neuropils are called optic foci and are reminiscent of and may be functionally equivalent to olfactory glomeruli of the antennal lobe.

OUTPUTS FROM THE OPTIC LOBES The now classic studies by Walter Gehring and his colleagues on the ectopic expression of compound eyes in *Drosophila* by genes controlling eye formation have demonstrated that supernumerary eyes are formed on limbs. If the ancestral origin of the compound eye was a limb, it would follow that the arrangements of sensory neurons leading centrally from the compound eye neuropils (comprising the optic lobe) should reflect arrangements of sensory neurons and interneurons that, in other ganglia, serve the appendages. Cell tracer studies indeed suggest that central projections from the optic lobes

are reminiscent of sensory-interneuron arrangements in thoracic ganglia. Afferents from the optic lobes do not appear to terminate directly onto descending neurons, as was once assumed. Rather, optic lobe outputs end on systems of local interneurons and some appear to map the retinotopic mosaic into optic glomeruli, suggesting further high-order visual reconstruction. Certain local interneurons connect optic glomeruli. Others have long axons that connect glomeruli on both sides of the brain. Yet others project anteriorly, via a thick fascicle of axons, to end rostrally in the lateral protocerebrum where they meet the terminals of olfactory projection neurons.

Optic glomeruli are also associated with neurons that reach the fan-shaped and ellipsoid bodies of the central complex or extend to the mushroom body lobes. Thus, optic lobes supply protocerebral centers from which interneurons extend to higher centers. Descending neurons mainly receive their optic lobe inputs via intermediate local interneurons. Thus, organization in the protocerebrum between sensory inputs (that is, lobula and lobula plate outputs) and interneurons is reminiscent of sensory-to-interneuron arrangements in a segmental ganglion.

Exceptions to this general arrangement include the giant vertical motion-sensitive neurons of the lobula plate, which, with neurons from the ocelli, establish mixed electrical and chemical synapses onto descending neurons that are involved in the stabilization of roll and pitch during flight and the control of visually induced head movements.

ACCESSORY MEDULLA AND CIRCADIAN RHYTHM The optic lobes also support pathways involved in circadian rhythms. The most important of these, described from *Drosophila* and the locust by Helfrich-Forster and others, involve systems of neurosecretory pacemaker cells associated with a small satellite neuropil in the optic lobes, called the accessory medulla.

DESCENDING OUTPUTS FROM THE BRAIN Dye fills into the ventral nerve cord of flies reveal large numbers (>200) of cell bodies on each side of the brain. These belong to descending neurons that carry information from the brain to thoracic ganglia circuits that control flight, walking, and other motor actions. Descending neurons that have their dendrites in dorsal neuropils of protocerebrum and receive inputs from the lobula plate terminate dorsally in thoracic ganglia where they contribute to the visual and mechanosensory stabilization of flight. Descending neurons with dendrites in more ventral protocerebral neuropils terminate in ventral thoracic and abdominal regions.

Electrophysiological recordings from locusts, dragonflies, and flies have all shown that descending neurons are multimodal, carrying integrated information about visual, olfactory, mechanosensory, and acoustic stimuli. Descending neurons are activated by correlative information from different modalities. Thus, descending neurons involved in controlling the direction of flight, and which respond selectively to panoramic movement around the vertical axis of the body (yaw motion), also relay information about the corresponding displacement of head hairs that occurs when the body undergoes a yaw displacement.

Among descending neurons are systems of axons that provide extremely fast motor actions in response to defined visual stimuli. These comprise the class of "escape circuits," the best known of which is the lateral giant motion detector and descending contralateral motion detector system of the locust and the Col A–giant descending neuron system in flies. In both locusts and flies optic lobe neurons (LGMD, Col A) provide mixed chemical and electrical synapses onto large axon diameter descending neurons (DCMD, GDN). In flies the paired GDN are electrically coupled and are electrically contiguous with the tergotrochanteral motor neurons that provide sudden midleg extension. At the same time, the direct (power) muscles of the wings are activated by electrically coupled local interneurons relaying signals from the GDN terminal to axons of motor neurons supplying the longitudinal flight muscles. A comparable connection between the DCMD and the hindleg extensor muscles is found in locusts.

ASCENDING SUPPLY TO THE BRAIN Sensory afferents supplying segmental ganglia distribute outputs to intersegmental interneurons. Many of these interneurons extend only locally, between neighboring or next-to-neighbor ganglia, and serve functions in regulating leg movements and posture. However, a large number of ascending neurons have axons that ascend through ganglia to reach the brain via ventrally disposed tracts of axons. Dye filling these axons demonstrates their terminals in ventrolateral and ventromedial protocerebral neuropils and in neuropils of the deutocerebrum, tritocerebrum, and subesophageal ganglion. Ascending axons do not, however, terminate in any of the mushroom body neuropils nor in neuropils of the central complex.

Functional studies of ascending pathways, exemplified by the recent studies of Nebeling, have mainly focused on ascending acoustic interneurons in crickets, which terminate in specific caudolateral and caudomedial protocerebral neuropils. However, the distribution of terminals in many protocerebral neuropils, including those receiving inputs from the optic lobes and antennal lobes, adds weight to the idea that a large volume of the protocerebrum is involved in multimodal integration. That such areas provide inputs to, and receive outputs from, the mushroom bodies adds credence to these paired centers being higher integration neuropils. Likewise, the relationship of the fan-shaped and ellipsoid bodies with protocerebral neuropils also suggests their crucial role as a higher integrator for the control of motor actions.

THE DEUTOCEREBRUM

The second preoral neuromere is called the deutocerebrum. It consists of sensory and motor neuropils and is constructed along the ground pattern typical of postoral ganglia.

Ventrally its neuropils comprise two paired sensory centers: the antennal lobe and, caudal to it, the vertical lobe. Three recent reviews by Vosshall, Hildebrand, and Hansson provide useful summaries of how olfactory receptors wire into the antennal lobes, how the lobes themselves are structured, and how these structures relate to the rest of the brain. Antennal lobe neuropils consist of discrete islets, called glomeruli. Most of these receive mainly olfactory receptor endings from the antenna's funiculus, although certain glomeruli receive inputs from the maxillary palps. Studies on *Drosophila* and the moth *Manduca sexta,* which are reviewed, have demonstrated that as a rule glomeruli have unique identities and positions in the antennal lobe.

The vertical lobes receive mainly mechanosensory terminals from the scapus and pedicellus. Some glomeruli situated caudally in the antennal lobe, bordering the dorsal lobe, receive inputs from antennal thermoreceptors, receptors responding to water vapor, and some mechanoreceptors. Motor neurons controlling antennal musculature originate lateral to and dorsal to these sensory regions. However, there is little information about the arrangements of interneurons and relay neurons supplying antennal lobe motor neurons.

Glomerular antennal lobes appear to be typical of neopteran insects. The antennae of primary apterygotes and palaeopterans supply columnar and layered neuropils of the vertical lobes, as do mechanosensory axons from the first two antennal segments (scapus and pedicellus) in neopterans.

Cellular arrangements in the antennal lobes are reminiscent of arrangements in malacostracan crustaceans and in vertebrates. Whether these similarities are a consequence of convergent evolution is debatable. However, in both phyla olfactory receptor neurons tuned to a specific odor molecule converge to the same address (glomerulus) in the antennal lobe. These addresses are then represented as a coarser map in the neuropils of the lateral protocerebrum by the axons of projection neurons originating in glomeruli. Antennal lobes also integrate olfactory information by virtue of complex connections provided by local inhibitory local interneurons. Sexual dimorphism also occurs in the antennal lobes, particularly in the Lepidoptera, in which receptors encoding components of the female pheromone blend send their axons to discrete glomeruli of the male-specific macroglomerular complex.

THE TRITOCEREBRUM

The tritocerebrum is the third and structurally the most discrete of the three preoral neuromeres, with sensory motor connections with the third metamere of the head. An analysis using expression of the segmentation border gene *engrailed* has demonstrated the appendage-type identity of the labrum, which supplies a major input into the tritocerebral ganglion, thus establishing its sensory supply and metameric relationships. The tritocerebrum also supplies motor neurons to labral muscles. The tritocerebrum receives a substantial supply of terminals from interganglionic interneurons with dendrites in the anterior medial protocerebrum (according to the neuraxis), which itself receives inputs from the median bundle originating in the subesophageal ganglion with some additional elements recruited from the thoracic ganglia. The tritocerebrum gives rise to descending neurons to the thoracic and abdominal ganglia, as does the deutocerebrum. However, even though Rajashekhar and Singh have described its general architecture and relationships with the protocerebrum, relatively little is known of the tritocerebrum's relationships to other brain regions or about its intrinsic connections. In all neopteran insects it is the protocerebrum that shows the greatest elaboration.

THE POSTORAL BRAIN (SUBESOPHAGEAL GANGLION)

Various functions have been ascribed to the subesophageal ganglion, including arousal prior to motor actions and sensory convergence from the brain. Generally, the three neuromeres of the subesophageal ganglion relate to the metameric identity of the mouthparts. Edgecomb and Murdock describe from flies that the labial neuromere of the subesophageal ganglion receives sensory axons from the dorsal cibarial organ, labellar sensilla, and labral sense organs and possibly some intersegmental inputs arrive from the tarsi. Stocker and Schorderet have supplied evidence from *Drosophila* that mechanosensory and gustatory chemo- (taste) receptors segregate out in the labral neuromere into discrete modality-specific zones. In honey bees, inputs from the mandibles invade the mandibular neuromere, which has been suggested to be absent in flies. However, it is unlikely that an entire neuromere has been eliminated and anatomical evidence for it is indeed present. In 1992 Shanbhag and Singh made the attractive suggestion that taste receptors segregate into chemospecific zones, because horseradish peroxidase uptake by the tips of functional species of receptors identified seven arborization areas in the subesophageal ganglion. Shanbhag and Singh suggested that these areas correspond to seven types of gustatory sensilla. These findings have not been contradicted and in larvae such receptor-specific zones appear to be substantiated by genetic markers of specific chemosensory axons to target neuropils. What is uncontroversial is that, at any neuromere of the subesophageal ganglion, neuropils are divided into a number of discrete synaptic regions. Some of these clearly belong to interneuron–motor neuron assemblages. In flies, local interneurons responding to sucrose reflect the discrete partitioning of lateral subesophageal neuropils.

See Also the Following Article

Eyes and Vision

Further Reading

Bacon, J. P., and Strausfeld, N. J. (1986). The dipteran 'giant fibre' pathway: Neurons and signals. *J. Comp. Physiol. (A)* **158,** 529–548.

Boyan, G. S., and Williams, J. L. (2001). A single cell analysis of *engrailed* expression in the early embryonic brain of the grasshopper *Schistocerca*

gregaria: Ontogeny and identity of the secondary headspot cells. *Arthropod Struct. Dev.* **30,** 207–218.
Burrows, M., and Newland, P. L. (1993). Correlations between the receptive fields of locust interneurons, their dendritic morphology, and the central projections of mechanosensory neurons. *J. Comp. Neurol.* **329,** 412–426.
Douglass, J. K., and Strausfeld, N. J. (1998). Functionally and anatomically segregated visual pathways in the lobula complex of a calliphorid fly. *J. Comp. Neurol.* **396,** 84–104.
Douglass, J. K., and Strausfeld, N. J. (1996). Visual motion-detection circuits in flies: Parallel direction- and non-direction-sensitive pathways between the medulla and lobula plate. *J. Neurosci.* **16,** 4551–4562.
Dubnau, J., Grady, L., Kitamoto, T, and Tully, T. (2001). Disruption of neurotransmission in *Drosophila* mushroom body blocks retrieval but not acquisition of memory. *Nature* **411,** 476–480.
Edgecomb, R. S., and Murdock, L. L. (1992). Central projections of axons from taste hairs on the labellum and tarsi of the blowfly, *Phormia regina* Meigen. *J. Comp. Neurol.* **315,** 431–444.
Farris, S. M., and Strausfeld, N. J. (2001). Development of laminar organization in the mushroom bodies of the cockroach: Kenyon cell proliferation, outgrowth, and maturation. *J. Comp. Neurol.* **439,** 331–351.
Franceschini, N. (1975). Sampling of the visual environment by the compound eye of the fly: Fundamentals and applications. *In* "Photoreceptor Optics" (A. W. Snyder and R. Menzel, eds.), pp. 98–125. Springer-Verlag, Berlin.
Gilbert, C., and Strausfeld, N. J. (1991). The functional organization of male-specific visual neurons in flies. *J. Comp. Physiol. [A]* **169,** 395–411.
Halder, G., Callaerts, P., and Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila. Science* **267,** 1788–1792.
Hansson, B. S. (2002). A bug's smell—Research into insect olfaction. *Trends Neurosci.* **25,** 270–274.
Hardie, R. C. (1986). The photoreceptor array of the dipteran retina. *Trends Neurosci.* **9,** 419–423.
Hausen, K. (1984). The lobula complex of the fly: Structure, function, and significance in visual behavior. *In* "Photoreception and Vision in Invertebrates" (M. A. Ali, ed.), pp. 523–599. Plenum Press, New York.
Heisenberg, M. (1983). What do the mushroom bodies do for the insect brain? An introduction. *Learn. Memory* **5,** 1–10.
Helfrich-Forster, C., Stengl, M., and Homberg, U. (1998). Organization of the circadian system in insects. *Chronobiol. Int.* **15,** 567–594.
Hildebrand, J. G. (1996). Olfactory control of behavior in moths: Central processing of odor information and the functional significance of olfactory glomeruli. *J. Comp. Physiol. [A].* **178,** 5–19.
Ito, K., Awano, W., Suzuki, K., Hiromi, Y., and Yamamoto, D. (1997). The *Drosophila* mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. *Development* **124,** 761–771.
Lee, T., Lee, A., and Luo, L. (1999). Development of the *Drosophila* mushroom bodies: Sequential generation of three distinct types of neurons from a neuroblast. *Development* **126,** 4065–4076.
Ludwig, P., Williams, L., and Boyan, G. (2001). The pars intercerebralis of the locust brain: A developmental and comparative study. *Microsc. Res. Tech.* **56,** 174–188.
Marin, E. C., Jefferis, G. S., Komiyama, T., Zhu. H., and Luo, L. (2002). Representation of the glomerular olfactory map in the *Drosophila* brain. *Cell* **109,** 243–255.
Meinertzhagen, I. A. (2000). Wiring the fly's eye. *Neuron* **28,** 310–313.
Meinertzhagen, I. A., and Hanson, T. E. (1993). The development of the optic lobe. *In* "The Development of *Drosophila melanogaster*" (M. Bate and A. Martinez-Arias, eds.), pp. 1363–1491. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Meinertzhagen, I. A., and Sorra, K. E. (2001). Synaptic organization in the fly's optic lamina: Few cells, many synapses and divergent microcircuits. *Prog. Brain Res.* **131,** 53–69.
Mitchell, B. K., Itagaki, H., and Rivet, M. P. (1999). Peripheral and central structures involved in insect gustation. *Microsc. Res. Tech.* **47,** 401–415.
Murphey, R. K. Possidente, G., Pollack, D., and Merritt, D. J. (1989). Modality-specific axonal projections in the CNS of the flies *Phormia* and *Drosophila. J. Comp. Neurol.* **290,** 185–200.
Nässel, D. R. (1993). Neuropeptides in the insect brain: A review. *Cell Tissue Res.* **273,** 1–29.
Nassif, C., Noveen, A., and Hartenstein, V. (1998). Embryonic development of the *Drosophila* brain. I. Pattern of pioneer tracts. *J. Comp. Neurol.* **402,** 10–31.
Nebeling, B. (2000). Morphology and physiology of auditory and vibratory ascending interneurones in bushcrickets. *J. Exp. Zool.* **286,** 219–230.
Pascual, A., and Préat, T. (2001). Localization of long-term memory within the *Drosophila* mushroom body. *Science* **294,** 1115–1117.
Rajashekhar, K. P., and Singh, R. N. (1994). Neuroarchitecture of the tritocerebrum of *Drosophila melanogaster. J. Comp. Neurol.* **349,** 633–345.
Rowell, C. H., and Reichert, H. (1986). Three descending interneurons reporting deviation from course in the locust. II. Physiology. *J. Comp. Physiol. [A]* **158,** 775–794.
Scott, K., Brady, R., Jr., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C., and Axel, R. (2001). A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila. Cell* **104,** 661–673.
Siegmund, T., and Korge, G. (2001). Innervation of the ring gland of *Drosophila melanogaster. J. Comp. Neurol.* **431,** 481–491.
Shanbhag, S. R., Singh, K, and Singh, R. N. (1995). Fine structure and primary sensory projections of sensilla located in the sacculus of the antenna of *Drosophila melanogaster. Cell Tissue Res.* **282,** 237–249.
Stocker, R. F. (1994). The organization of the chemosensory system of *Drosophila melanogaster:* A review. *Cell Tissue Res.* **275,** 3–26.
Strausfeld, N. J. (1976). "Atlas of an Insect Brain." Springer-Verlag, Berlin.
Strausfeld, N. J. (1999). A brain region in insects that supervises walking. *Prog. Brain. Res.* **123,** 274–284.
Strausfeld, N. J. (2002). Organization of the honey bee mushroom body: Representation of the calyx within the vertical and gamma lobes. *J. Comp. Neurol.* **450,** 4–33.
Strausfeld, N. J., and Nässel, D. R. (1980). Neuroarchitectures serving compound eyes of Crustacea and insects. *In* "Handbook of Sensory Physiology VII/68" (H. Autrum, ed.), pp. 1–132. Springer-Verlag, Berlin.
Vosshall, L. B. (2001). The molecular logic of olfaction in *Drosophila. Chem. Senses.* **26,** 207–213.
Walthall, W. W., and Murphey, R. K. (1986). Positional information, compartments and the cercal system of crickets. *Dev. Biol.* **113,** 182–200.
Zars, T., Fischer, M., Schulz, R., and Heisenberg, M. (2000). Localization of a short-term memory in *Drosophila. Science* **288,** 672–675.

Bristletail

see *Archaeognatha*

Bubonic Plague

Irwin W. Sherman
University of California, Riverside

Bubonic plague, a devastating bacterial disease most commonly transmitted by fleas, has produced profound changes in human societies throughout history.

THE DISEASE IN HUMANS

During the last 2000 years, three great plague pandemics, including the Black Death of the 14th century, have resulted in social and economic upheavals that are unmatched by armed conflicts or any other infectious disease. Plague, caused by the rod-shaped, gram-negative bacterium *Yersinia pestis* (*Pasteurella pestis* was the name used before 1970), is a zoonotic infection, transmitted by any one of several species of fleas, that predominantly affects small mammals such as rodents; humans actually are accidental hosts.

At present, most human cases of the plague are of the bubonic form, which results from the bite of a flea, usually the common rat flea *Xenopsylla cheopis,* that has fed on an infected rodent. The bacteria spread to the lymph nodes (armpits and neck but frequently the area of the groin) that drain the site of the bite, and these swollen and tender lymph nodes give the classic sign of bubonic plague, the bubo (from the Greek word *boubon* meaning groin). Three days after the buboes appear, there is a high fever, the infected individual becomes delirious, and hemorrhages in the skin result in black splotches. Some contend that these dark spots on the skin gave the disease the name Black Death, whereas others believe "black" is simply a mistranslation of "pestis atra" meaning not black, but a dark or sinister disease.

The buboes continue to enlarge, sometimes reaching the size of a hen's egg, and when these buboes burst there is agonizing pain. Death can come 2 to 4 days after the onset of symptoms. Sometimes, however, the bacteria enter the bloodstream. This second form of the disease, which may occur without the development of buboes, is called septicemic plague. Septicemic plague is characterized by fever, chills, headache, malaise, massive hemorrhaging, and death. Septicemic plague has a higher mortality than bubonic plague.

In addition, the bacteria may move via the bloodstream to the alveolar spaces of the lungs, leading to a suppurating pneumonia or pneumonic plague. Pneumonic plague, the only form of the disease that allows for human-to-human transmission, is characterized by a watery and sometimes bloody sputum containing live bacteria. Coughing and spitting produce airborne droplets laden with the highly infectious bacteria, and by inhalation others may become infected. Pneumonic plague is the rapidly fatal form of the disease, and death can occur within 24 h of exposure. It is likely that this form of transmission produced the devastating Black Death. The nursery rhyme "Ring around the rosies, a pocket full of posies, Achoo! Achoo! We all fall down" refers to plague in 17th-century England: the rosies are the initial pink body rash, posies the perfumed bunches of flowers used to ward off the stench of death, "achoo" is the coughing and sneezing, and death is signified by "we all fall down."

Y. pestis is one of the most pathogenic bacteria: the lethal dose that will kill 50% of exposed mice is only a single bacterium that is injected intravenously. Typically, *Y. pestis* is spread from rodent to rodent by flea bites, but it can also survive for a few days in a decaying corpse and can persist for years in a frozen body.

THE DISEASE IN THE FLEA

The disease in fleas also has a distinctive pattern. Small mammals, such as urban and sylvatic (or wood) rats, as well as squirrels, prairie dogs, rabbits, voles, coyotes, and domestic cats, are the principal hosts for *Y. pestis.* More than 80 different species of fleas are involved as plague vectors. Fleas are bloodsucking insects, and when a flea bites a plague-infected host (at the bacteremic/septicemic stage) it ingests the rod-shaped bacteria; these multiply in the blood clot in the proventriculus (foregut) of the flea. This bacteria-laden clot obstructs the flea's bloodsucking apparatus and, as a consequence, the flea is unable to pump blood into the midgut, where normally it would be digested. The flea becomes hungrier and in this ravenous state bites the host repeatedly; with each bite, it regurgitates plague bacteria into the wound. In this way, infection is initiated. *Y. pestis* can also be pathogenic for the flea, and fleas with their foregut blocked rapidly starve to death. If the mammalian host dies, its body cools down, and fleas respond by moving off the corpse to seek another live warm-blooded host. However, if there is an extensive die-off of rodents, the fleas move on to less preferred hosts such as humans, and so an epidemic may begin.

HISTORICAL

Plague outbreaks occurred prior to the current era (i.e., 2000 years ago), but the numbers affected and the societal impact of the events remain unrecorded. During the current era, however, there have been three well-documented plague pandemics. The first, the plague of Justinian, arrived in 542 and raged intermittently until 750. It came to the Mediterranean region from an original focus in northeastern India or central Africa and was spread by infected rats hitchhiking on ships. It is estimated that a million people died. In Constantinople, the capital of the Roman Empire in the east, plague contributed to Justinian's failure to restore imperial unity because of a diminution in resources, which in turn prevented Roman and Persian forces from offering more than token resistance to the Muslim armies that swarmed out of Arabia in 634.

In the year 1346, the second pandemic began, and by the time it disappeared in 1352 the population of Europe and the Middle East had been reduced from 100 million to 80 million people. It is estimated that in cities such as Siena, Marseilles, and London, at the height of the pandemic, approximately 1500 people died each day from the plague. This devastating phenomenon, known as the Black Death, the Great Dying, or the Great Pestilence, put an end to the rise in the human population that had begun in 5000 B.C., and it took more than 150 years before the population returned to its former size.

Providing the source of the second pandemic were bacteria from the first pandemic that had moved eastward and remained endemic for seven centuries in the highly susceptible black rats *(Rattus rattus)* of the Gobi Desert. Plague-infected rats and their fleas moved westward along the Silk Road, the caravan routes between Asia and the Mediterranean; plague traveled from central Asia around the Caspian Sea to the Crimea. There the rats and their fleas boarded ships and moved from port to port and country to country, spreading plague to the human populations living in filthy rat-infested cities. Indeed, the story of the Pied Piper of Hamlin may have had its roots in the plague-ridden cities of Germany.

By 1347, there were plague outbreaks in Kaffa, Constantinople, and Genoa. By 1348 it had spread via North Africa to Spain, and it was also present in France, Germany, Switzerland, and Great Britain. In 1349 a ship from London carrying its crew, wool cargo, and infected rats landed in Bergen, Norway. In this way plague came to Scandinavia. In 1351, it was in Poland and when, in 1352, it reached Russia, the plague had completed its circuit.

Because no one in medieval times knew that microbes cause infectious diseases, any public heath measures were crude and generally ineffectual: ships were restricted in their entry into ports and sailors had to remain on board for 40 days while their vessels were tied up at the dock, a practice that gave rise to the term quarantine (from *quarant,* meaning 40 in French). But the disease continued unabated because flea-bearing rats left the ships by means of docking lines. Cordon sanitaires (i.e., quarantine zones) may have had some effect, but oftentimes infected individuals were shut up in their homes with the uninfected members of the family and the flea-infested rats, conditions that actually led to higher mortality. More effective measures included the burning of clothing and bedding, and the burying of the dead as quickly as possible.

The public, unable to identify the real source of the plague, used Jews, prostitutes, the poor, and foreigners as scapegoats. The Black Death led to societal and religious changes: feudal institutions began to break down; the laboring class became more mobile; merchants and craftsmen became more powerful; and guild structures were strengthened. There was also a decline in papal authority, and people lost faith in a Catholic Church that was powerless to stem the tide of death. The horrors of the plague during this time are depicted in Pieter Brueghel's 1562 painting *Triumph of Death* and graphically described in the introduction to Giovanni Boccaccio's classic collection of short stories, the *Decameron.* "Plague doctors" who ministered to the dying wore special costumes depicted in drawings and engravings (Fig. 1), as seen in popular movies such as *The Seventh Seal,* directed by Ingmar Bergman.

Though the Black Death was undoubtedly the most dramatic outbreak of plague ever visited upon Europe, it did not disappear altogether after 1352. Between 1347 and 1722, plague epidemics struck Europe at infrequent intervals and occurred without the introduction from caravans from Asia. In England, the epidemics occurred at 2- to 5-year intervals between 1361 and 1480. In 1656–1657, 60% of the population of Genoa died, half of Milan in 1630, and 30% in Marseilles in 1720. In the Great Plague of 1665, which was described in the diary of Samuel Pepys (and fictionalized in Daniel Defoe's *Journal of the Plague Year*), at least 68,000 Londoners died.

FIGURE 1 During the plague physicians wore protective clothing in an attempt to avoid acquiring the disease from their patients. The beaklike mask was supplied with aromatic substances and perfumes to ward off the stench of death, the stick was used to touch the afflicted. To prevent the disease vapors from entering the body of the "plague doctors," a hat was worn, as well as a coat impregnated with a waxy material. (Illustration from *Der Pestarzt Dr. François Chicoyneau*. © Germanisches National Museum, Munich. Reproduced with permission.)

The third and current pandemic began in the 1860s in the war-torn Yunnan region of China. Troop movements from the war in that area allowed the disease to spread to the southern coast of China. Plague-infected rodents, now assisted in their travels by modern steamships and railways, quickly spread the disease to the rest of the world. By 1894, plague had arrived in Hong Kong, and there Alexander Yersin (1863–1943) and Shibasaburo Kitasato (1852–1931), by taking material from buboes, independently discovered the occurrence of the bacillus in humans. Yersin also isolated the same bacterium from dead rats, thus demonstrating the importance of these rodents in transmission. Four years later, during the phase of the epidemic that swept over India, Paul-

Louis Simond and Masanori Ogata independently determined that the flea was the vector of plague.

THE BACTERIUM, *Y. PESTIS*

Y. pestis has been subdivided into three phenotypic biovarieties—Antiqua, Medievalis, and Orientalis. Based on epidemiological and historical records, it has been hypothesized that Antiqua, presently resident in Africa, is descended from bacteria that caused the first pandemic, whereas Medievalis, resident in central Asia, is descended from bacteria that caused the second pandemic; those of the third pandemic, and currently widespread, are all Orientalis. It is believed that *Y. pestis* probably evolved during the last 1500 to 2000 years because of changes in social and economic factors that were themselves the result of a dramatic increase in the size of the human population, which was coincident with the development of agriculture.

This increase in food supply for humans allowed rodent populations to expand as well. Increased numbers of rodents coupled with changes in behavior (i.e., living in and around sylvatic rodents and human habitation) triggered the evolution of virulent *Y. pestis* from the enteric, food-borne, avirulent pathogen *Y. pseudotuberculosis.* This occurred by means of several genetic changes. For example, development of a gene whose product is involved in the storage of hemin resulted in the ability of the bacteria to block the flea proventriculus, enhancing flea-mediated transmission. Other gene products (phospholipase D and plasminogen activator) facilitated blood dissemination in the mammalian body and allowed for the infection of a variety of hosts by fleas.

DIAGNOSIS, TREATMENT, AND VACCINE

Mortality and morbidity from plague were significantly reduced in the 20th century. However, the disease has not been eradicated. Plague remains endemic in regions of Africa, Asia, and North and South America. From 1983 to 1997, there were 28,570 cases with 2331 deaths in 24 countries reported to the World Health Organization (WHO). In 1997 the total number of cases reported by 14 countries to the WHO was 5419, of which 274 were fatal. Epidemics occurred in Madagascar in 1991 and 1997, in Malawi, Zimbabwe, and India in 1994, and in Zambia and China in 1996. In contrast, there were four cases and only one death in the United States in 1997.

Though human disease is rare, a feverish patient who has been exposed to rodents or flea bites in plague endemic areas should be considered to be a possible plague victim. Diagnosis can be made by Gram stain and culture of bubo aspirates or sputum. The bacteria grow aerobically and form small colonies on blood and MacConkey agar.

Unless specific treatment is given, the condition of a plague-infected individual deteriorates rapidly and death can occur in 3 to 5 days. Untreated plague has a mortality of more than 50%. A variety of antibiotics including streptomycin, sulfonamide, and tetracycline are effective against bubonic plague. Tetracycline can be used prophylactically, and chloramphenicol is used to treat plague meningitis. No antibiotic resistance has been reported.

Two plague vaccines have been approved for use in humans. One is a formaldehyde-killed, whole-cell vaccine first used in 1942, and the other is a live vaccine used in the former Soviet Union since 1939. A new subunit vaccine that uses the bacterial capsular antigens F1 and V for immunization is under development.

See Also the Following Articles

Blood Sucking • Medical Entomology • Siphonaptera • Zoonoses

Further Reading

Achtman, M., *et al.* (1999). *Yersinia pestis,* the cause of plague is a recently emerged clone of *Yersinia pseudotuberculosis. Proc. Natl. Acad, Sci. U.S.A.* **96,** 14043–14048.

Carniel, E. (2000). Plague. *In* "Encyclopedia of Microbiology," (J. Lederberg, ed.), Vol. 3, pp. 654–661. Academic Press, San Diego, CA.

Herlihy, D. (1997). "The Black Death and the Transformation of the West." Harvard University Press, Cambridge, MA.

Hinnebusch, B. J. (1997). Bubonic plague: A molecular genetic case history of the emergence of an infectious disease. *J. Mol. Med.* **75,** 645–652.

McNeill, W. (1998). "Plagues and People." Anchor Books, New York.

Perry, R., and Fetherston, J. (1997). *Yersinia pestis*—etiologic agent of plague. *Clin. Microbiol. Rev.* **10,** 35–66.

Zeigler, P. (1969). "The Black Death." Harper & Row, New York.

Bumblebee

see *Hymenoptera*

Butterfly

see *Lepidoptera*

C

Caddisfly

see *Trichoptera*

Caste

Sean O'Donnell
University of Washington, Seattle

The term "caste" refers to long-term, stable differences among insect colony members that affect the roles played by individuals in their social group. It was the existence of distinct morphological castes in insect colonies that led Charles Darwin to identify social insects as a major challenge to his theory of evolution by natural selection. Few topics are more central to the study of social insect biology than caste. It is ironic, then, that few topics have generated more controversy and debate among social insect biologists. Beyond the deceptively simple definition just offered, there is little agreement on how the term should be defined, or on what characteristics should be used to identify the castes of individual insects. In part, the controversy exists because published definitions of caste are often not operational. For example, some definitions do not specify whether caste differences must be developmentally fixed and permanent or, alternatively, whether individuals can exhibit caste flexibility. As a result, "caste" has been applied to a wide array of physiological and behavioral phenomena.

The diversity of caste systems between and within evolutionary lineages of social insects may preclude a simple, universal definition of caste. Rather than advocate a single definitional point of view, this article explores the diversity of ways in which the concept has been used and the array of important phenomena caste encompasses in different insect societies.

REPRODUCTIVE CASTES

Differences in Reproductive Function

A fully social or eusocial group is generally understood to exhibit reproductive division of labor. This means that eusocial groups must include some individuals that forgo direct reproduction and instead aid the rearing of the offspring of others in their group. In eusocial insects, the helpers comprise the worker caste and reproductive females are referred to as queens. Termite colonies possess long-lived royal couples (a queen and a king), whereas in eusocial Hymenoptera, males are sometimes referred to as drones. Males in the order Hymenoptera (bees, ants, and wasps) rarely work for their colonies and typically die soon after mating. In contrast, male eusocial thrips (Thysanoptera) and termites (Isoptera) comprise part of the worker force and participate fully in colony labor.

Social insect species vary according to whether the group's members are permanently relegated to reproductive versus worker roles and in the degree of fecundity differences between reproducers and workers. There is a general evolutionary trend toward increased reproductive caste specialization as more complex, larger societies evolve from smaller, simpler ones. In some ants, workers lack reproductive organs and are permanently sterile. In most species, however, workers can achieve limited direct reproduction under some conditions.

Morphological Differences

Some species are reproductively monomorphic, and reproductives do not differ significantly in body structure from workers. Many sweat bees and bumble bees, some paper wasps, and even some primitive ants are examples of reproductively monomorphic species. Workers in monomorphic species are

often smaller than reproductives, but there can be considerable overlap in body size distributions among the reproductive castes. In some cases, clear physiological differences distinguish workers from reproductives when morphology does not. For example, temperate *Polistes* paper wasp colonies produce gynes (potential female reproductives) at the end of the summer. Gynes possess enlarged, nutrient-laden fat bodies, not present in female workers, that permit them to overwinter in a quiescent state.

In contrast, consistent reproductive caste differences in body size and shape have evolved in several lineages of social insects. Most eusocial insects with wingless workers, such as ants and termites, retain a morphologically distinct reproductive caste with wings. In species with flying workers, developmental allometry can still result in the production of distinct, nonoverlapping body forms for reproductives and workers. Morphologically discrete reproductive castes are found among honey bees, stingless bees, and some paper wasps. Reproductives are often larger than workers, but also differ in body proportions (hence shape) in ways that suggest specialization in egg laying, such as relatively enlarged abdomens. The degree of morphological differentiation between reproductive castes probably evolves in response to a complex array of natural selection pressures. For example, the degree to which the colony occupies a defensible, long-lasting nest site may in part determine the whether queens can afford to adopt relatively immobile body forms.

Caste Determination: Immature Development and Adult Interactions

Other than an interesting exception in the ant *Harpagoxenus sublaevis,* there are no well-documented cases of genetic differences that affect reproductive caste differentiation. Often caste differentiation must depend in part on differential patterns of gene expression during development, particularly in species with distinct caste morphology.

Differences in environmental conditions during immature development can have strong effects on an individual's caste. Nutritional effects on reproductive caste have been documented in numerous taxa and appear to be widespread, if not universal, among eusocial insects. Differences in the amount of food provided to larvae may underlie many of the differences between reproductives and workers, especially in species exhibiting the common pattern of larger body size for reproductives. However, differences in food quality, possibly including the addition of glandular secretions and pheromones, cannot be ruled out. Especially interesting in this regard are those eusocial wasps whose reproductives are smaller than workers (genus *Apoica*) or identical in size but different in shape (genus *Pseudopolybia*).

Social interactions among adults may also influence reproductive caste, particularly in species without apparent morphological caste differences. For example, dominance interactions among paper wasp (*Polistes*) females, which often cooperate to start new colonies, determine which female acts as the sole reproductive. Subordinate *Polistes* females function as workers.

CASTES IN THE WORKER FORCE

Morphological Castes

DISCRETE WORKER MORPHOLOGY In all termites and in approximately 10% of ant species, workers exhibit developmental allometry resulting in body shape variation within the worker caste. Interestingly, this type of morphological caste has not been documented in social insects with flying workers, such as bees and wasps. There is typically some association between a worker's body form and the tasks that she performs. One of the most common types of morphological specialization is the assignment of large workers, called soldiers, to the special role of colony defense. When the colony is threatened by an animal, the soldiers advance and attack, while other workers flee. Often the soldiers uniquely possess heavily armored exoskeletons and some type of weaponry, including enlarged muscular heads, long, piercing mouthparts, or glands that produce defensive chemicals. In other cases, worker body shape variation affects the performance of more mundane tasks such as food collection. In army ants (*Eciton* spp.), longer-legged workers select larger food items to carry back to their colonies. In leafcutter ants (*Atta* spp.), the largest workers are soldiers, the medium-sized workers cut and transport leaves, and the smaller workers usually remain in the nest to tend the colony's fungus garden. An ant worker's body size and shape are fixed upon adult emergence; further growth is not possible. In contrast, some termite workers (*Zootermopsis* spp.) exhibit considerable caste plasticity, potentially molting among different body forms, and even switching from soldier to nonwinged reproductive castes under certain conditions.

BODY SIZE EFFECTS Even in monomorphic species, body size differences can influence the tasks that workers perform. In some species, larger-bodied workers dominate their smaller nestmates (*Polistes metricus, P. fuscatus,* and *P. dominulus*), and dominance status in turn affects the tasks a worker performs. In some bumble bees (*Bombus* spp.), however, larger workers are more likely to perform certain tasks such as foraging to collect food for the colony, independently of obvious worker aggression.

Behavioral Castes

Workers can be assigned to behavioral castes when they specialize on a subset of the tasks that the colony needs. In some eusocial insect species such as *Bombus* and *Polistes,* workers exhibit a great deal of flexibility, switching among tasks often, and behavioral castes are weakly defined. In honey bees (*Apis mellifera*) and swarm-founding paper wasps (*Polybia* spp.), on the other hand, workers specialize more consistently.

AGE OR TEMPORAL POLYETHISM Changes in task specialization as workers age are among the best-studied factors that influence workers' behavioral caste. "Age" or "temporal polyethism" refers to an ordered, predictable sequence of task specializations through which an adult worker passes as it ages. Typically, species with temporal polyethism exhibit centripetal development: workers begin by working deep inside the nest, close to the queen(s) and brood; they later perform tasks at the nest periphery; and they finally move further out to perform risky tasks such as foraging and nest defense (Fig. 1). This centripetal pattern of development is remarkably similar among the diverse eusocial insect species that exhibit well-developed temporal polyethism. Workers usually follow the same sequence of task specializations, but individuals vary in their rate of passage through the sequence. Changes in hormone titers, such as juvenile hormone, have been implicated in determining the rate of temporal polyethism in *Apis* and *Polybia.*

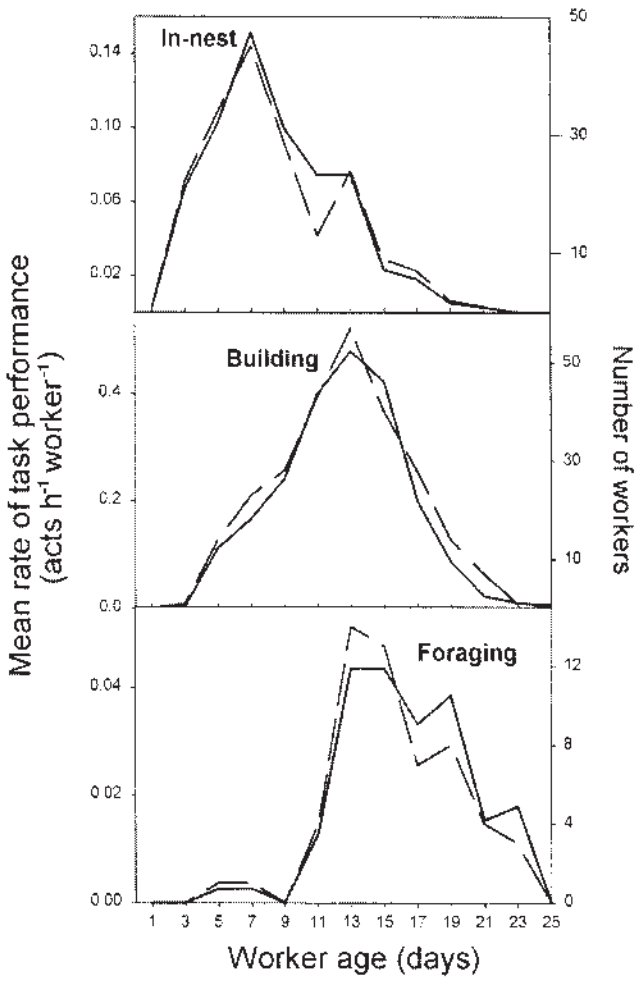

FIGURE 1 Temporal or age polyethism in the paper wasp *Polybia aequatorialis.* Data were collected on 130 individually marked, known-age workers. Two measures of worker activity at three task sets are plotted against worker age (solid lines: mean rate of task performance; dashed lines: number of workers performing the task). Note the typical centripetal developmental sequence: in-nest tasks (mostly nest cleaning) are followed by building on the nest exterior, and later by foraging (leaving the nest and returning with food and building materials).

Workers' relative age influences social status and task performance in some species. In the paper wasps *P. exclamans,* the first-emerging (and consequently the oldest) workers in the nest tend to socially dominate their nestmates, a pattern referred to as gerontocracy, which is independent of body size variation. In this case, age influences workers' behavioral caste in a static way, rather than in a dynamic way as in temporal polyethism.

INDIVIDUAL DIFFERENCES AND SPECIALIZATION Superimposed on broader patterns of division of labor, such as body size or age effects, workers sometimes exhibit idiosyncratic specialization on tasks. For example, *Apis* and *Polybia* foragers often specialize by collecting one of the several materials their colony needs to function. Such specialization may be benefit the colony by increasing the efficiency or reliability of task performance.

Genetic Effects Genetic relatedness among the offspring in the worker force is highest when the workers are born to a single reproductive female, which has mated with a single male. Some social insect species exhibit mating behavior or social structure that decreases the genetic relatedness among the offspring workers within colonies. When queens mate with several different, unrelated males (polyandry), or when several reproductive females are present in the colony (polygyny), workers can find themselves sharing a nest with a combination of more closely and more distantly related individuals. In a number of polyandrous and polygynous species, including *Apis* spp. and stingless bees (genus *Partamona*), several species of ants, and *Polybia* spp., workers that are more genetically similar have been found to specialize on similar tasks.

Experience and Learning A predicted benefit of task specialization is that workers can improve performance as they gain experience. There is evidence that some insect workers learn to perform tasks more effectively with experience. Bumble bees collecting nectar and pollen from complex flowers learn to do so more rapidly after repeated attempts to handle a given type of flower. *Polybia* foragers are less likely to return from foraging trips empty-handed as they gain foraging experience.

COLONY-LEVEL INVESTMENT

Investment in Growth and Maintenance versus Reproduction

One of the major challenges that faces growing organisms is the developmental decision of how many resources to invest in growth and how many to invest in reproduction. Insect colonies can be treated as organisms in this sense, since each colony must decide how much it will invest in different castes (i.e., in workers vs reproductives). To the extent that colonies

are reproductive units, optimality theory predicts that natural selection will favor colonies that allocate their limited resources efficiently into different castes. Many insect societies segregate the production of workers (early in colony development) from the production of new queens and males (later in colony development).

Worker Caste Ratios

Insect colonies appear to behave in an adaptive manner by adjusting their worker caste ratios to meet current colony needs. Production of different worker castes reflects a trade-off between the costs and benefits of producing and maintaining workers of different kinds. As ant colonies with morphologically specialized workers grow in size, their amount of investment in large-bodied workers increases, and many eusocial insects produce tiny nanitic workers early in colony development. Colonies of the ant *Pheidole pallidula* increase their rate of production of soldiers when exposed to potential competitors. Similar colony flexibility is apparent in age–caste distributions. In honey bees and paper wasps, if the level-of-colony need for foragers changes, some workers accelerate or reverse their behavioral development, performing the age-atypical tasks that are in greatest demand. Identifying the mechanisms that link individuals' developmental plasticity with the level of colony need remains as a central challenge in the study of caste.

See Also the Following Articles

Colonies • Division of Labor • Hymenoptera • Isoptera • Juvenile Hormone • Sociality

Further Reading

Evans, J. D., and Wheeler, D. E. (2001). Gene expression and the evolution of insect polyphenisms. *Bioessays* **23,** 62–68.

Hölldobler, B., and Wilson, E. O. (1990). "The Ants." Harvard University Press, Cambridge, MA.

Hunt, J. H. (1994). Nourishment and social evolution in wasps sensu lato. *In* "Nourishment and Evolution in Insect Societies" (J. H. Hunt and C. A. Nalepa, eds.), pp. 221–254. Westview Press, Boulder, CO.

O'Donnell, S. (1998). Reproductive caste determination in eusocial wasps (Hymenoptera: Vespidae). *Annu. Rev. Entomol.* **43,** 323–346.

Oster, G. W., and Wilson, E. O. (1978). "Caste and Ecology in the Social Insects." Princeton University Press, Princeton, NJ.

Peters, C., and Ito, F. (2001). Colony dispersal and the evolution of queen morphology in social Hymenoptera. *Annu. Rev. Entomol.* **46,** 601–630.

Caterpillars

Frederick W. Stehr
Michigan State University

The larvae of butterflies, skippers, and moths of the order Lepidoptera are generally known as caterpillars. Caterpillars come in a diversity of sizes, shapes, and colors. The most common form has a conspicuous head, a thorax with three pairs of legs, and an abdomen with five pairs of prolegs that bear crochets (hooks) (Fig. 1) that enable the caterpillar to cling tightly to or wedge itself between materials. In fact, some of the giant silk moth caterpillars (Saturniidae) can cling so tightly to a twig that a proleg can be ripped from the body if they are pulled too hard. A few other orders of insects contain larvae that are caterpillar-like, but only the larvae of the leaf-feeding sawflies (Hymenoptera) are commonly encountered. They are easily mistaken for caterpillars, but they usually feed in groups (as do some caterpillars), rear up when disturbed, have more than five pairs of prolegs on the abdomen, and never have crochets on the prolegs.

FIGURE 1 Caterpillar of the polyphemus moth, *Antheraea polyphemus,* showing the five pairs of prolegs bearing crochets (hooks). (Photograph by Joseph L. Spencer, Illinois Natural History Survey.)

FIGURE 2 Twig-mimic inchworm caterpillar of a moth (Geometridae). (Photograph by Fred Stehr, Department of Entomology, Michigan State University.)

FIGURE 3 "Bird dropping" mimicry by the caterpillar of the orangedog, *Papilio cresphontes.* (Photograph by J. Mark Scriber, Department of Entomology, Michigan State University.)

FIGURE 4 A slug caterpillar, *Euclea delphinii* (Limacodidae), showing the lack of prolegs and the sucker-like discs. (Photograph by Jane Ruffin, Rosemont, Pennsylvania.)

BIOLOGY AND ECOLOGY

Caterpillars are commonly encountered because many are leaf-feeders and are not concealed feeders, although they may be cryptically shaped or colored. The realistic twig mimicry (Fig. 2) and behavior found in some inchworms (Geometridae) are complete with body markings that resemble bark irregularities, scars, and stipules. Another excellent mimic is the caterpillar of the orangedog, *Papilio cresphontes,* whose black and white larvae resemble bird droppings in all instars (Fig. 3). Some other swallowtail caterpillars resemble bird droppings in the early instars, whereas the later instars change to a greenish color (with black and orange, or yellow, markings) that camouflage them on their host plants. One of the most unusual diet-induced camouflage systems is that of the geometrid moth, *Nemoria arizonaria.* This moth has two generations, the first feeds on oak catkins in the spring and takes on the shape and colors of the catkins; the second generation hatches when the catkins are gone and only leaves are available for food, resulting in the caterpillar being a twig mimic.

Nearly all species of plants are fed upon by caterpillars, and many caterpillars are important pests. They also form galls; are scavengers on dead plant materials, fungi, and animal materials such as feathers, wool, or fur (clothes moths, Tineidae); and some are important pests of stored products (meal moths, Pyralidae). Some larvae of the family Pyralidae are truly aquatic and have gills, and some caterpillars of other families feed in or on aquatic plants. A few caterpillars are predators such as the lycaenid butterfly, *Feniseca tarquinius,* which feeds on and among clusters of wooly aphids. Some tropical lycaenids are predators on ant larvae in their nests. The ants benefit from this arrangement by obtaining sweet secretions from the dorsal abdominal glands of the caterpillars. For other lycaenid caterpillars that feed on plants above ground, the relationship is clearly symbiotic, with the larvae providing secretions for the ants and the ants protecting the larvae from predators and parasitoids. Probably the most unusual predators are the Hawaiian geometrids in the genus *Eupethecia,* whose larvae strike backward with their enlarged thoracic legs, seizing any prey that touches their rear end.

Most caterpillars move by a wavelike movement of the legs from rear to front. However, those with reduced numbers of prolegs or none at all proceed in a different manner. The inchworms (Fig. 2), which usually have two pairs of prolegs, and the looper cutworms of the Noctuidae, which have only three or four pairs of prolegs, move by pulling the hind end close to the thoracic legs and then looping the body forward. Caterpillars of the Limacodidae (Fig. 4), the slug caterpillars, have no prolegs and move on sucker-like discs, gliding along in a manner similar to slugs.

Caterpillars may be aposematically colored to advertise that they are distasteful or poisonous. They may also bear diverse

FIGURE 5 The ill-smelling, eversible, dorsal, prothoracic glands of a tiger swallowtail caterpillar, *Papilio glaucus.* (Photograph by J. Mark Scriber, Department of Entomology, Michigan State University.)

FIGURE 6 The snake-mimicking behavior of the spicebush swallowtail larva, *Papilio troilus.* (Photograph by J. Mark Scriber, Department of Entomology, Michigan State University.)

FIGURE 7 Tent of the western tent caterpillar, *Malacosoma californicum* (Lasiocampidae). (Photograph by Fred Stehr, Department of Entomology, Michigan State University.)

lobes, spines, horns, knobs, and urticating hairs or spines that may or may not be irritating in some way. Most caterpillars that feed in protected or concealed locations do not possess such structures or defenses. The ill-smelling secretions of the eversible, dorsal prothoracic glands (Fig. 5) of swallowtail butterfly caterpillars (Papilionidae) are well known, as are the stinging properties of saddleback caterpillars (Limacodidae). Some papilionid and sphingid caterpillars have large eye-like spots on the thorax and rear up their head and expanded thorax, thereby mimicking small snakes (Fig. 6). The distastefulness to birds of many monarch larvae and adults that feed on toxic species of milkweeds is common knowledge, but monarch larvae that have fed on the relatively nontoxic species of milkweeds are quite edible, as are their adults.

Caterpillars spin silk through a conspicuous labial spinneret. Species use silk in diverse ways, from webbing together leaves or other materials (Tortricidae, Pyralidae, and many others) to constructing webs [fall webworm (Arctiidae)], large "tents" of the tent caterpillars (Fig. 7) (Lasiocampidae: *Malacosoma*), or silken tube shelters (some Pyralidae, Tineidae). Some, such as the casebearers (Coleophoridae), the bagworms (Psychidae), and the sackbearers (Mimallonidae), make tough, complex shelters that they carry with them. The "sack" of the mimallonid caterpillars is particularly interesting because it is extremely tough, with an opening at both ends that can be blocked by the caterpillar's head or its hard anal plate, either of which fits tightly against either opening.

Many caterpillars spin cocoons in which they pupate, ranging from the tight and commercially valuable cocoons of the silkworm, *Bombyx mori,* to those of the gypsy moth, *Lymantria dispar,* which are at best a loose net of silk (even though it was originally brought to the United States because it was believed to be a good prospect for silk production). Other common caterpillars, such as those of sphinx moths, cutworms, and some saturniids, pupate in the ground or litter where they form cells that can be silk lined.

Caterpillars are occasionally an item of commerce. Canned caterpillars (gusanos) are eaten in (and sometimes imported from) Mexico. Tequila worms, usually in the family Megathymidae (giant skippers), were formerly imported in bottles of tequila from Mexico. If the tequila worm was well preserved, presumably the alcohol concentration was satisfactory.

See Also the Following Articles

Cocoon • Larva • Lepidoptera

Further Reading

Stehr, F. W. (ed.) (1987). "Immature Insects," Vol. 1. Kendall/Hunt, Dubuque, IA.

Cat Fleas

Nancy C. Hinkle
University of Georgia

Michael K. Rust
University of California, Riverside

Fleas are small (2 mm), dark, reddish brown, wingless, bloodsucking insects. Their bodies are laterally compressed (i.e., flattened side to side) and covered with many hairs and short spines directed backward, permitting forward movement through hairs on the host's body. Their hind legs are long and well adapted for jumping. Adult flea mouthparts are equipped for sucking blood from the host.

In North America, the cat flea, *Ctenocephalides felis felis,* is the most common ectoparasite of dogs and cats. In tropical areas it is a year-round pest, whereas in temperate climates its

season varies. Because they attack a range of warm-blooded hosts, including humans and pets, cat fleas are both a veterinary problem and a household pest.

SIGNIFICANCE OF FLEAS

Fleas are important to humans because of their potential as disease vectors, in addition to the annoyance they produce merely by biting. Pathogen transmission is facilitated by their habit of feeding sequentially on several hosts. The best known disease associated with fleas is bubonic plague; the plague bacterium, *Yersinia pestis,* is transmitted almost exclusively by rodent fleas. Murine typhus is another disease for which cat fleas have been implicated in the transmission cycle. The flea-borne typhus causative agent, *Rickettsia typhi,* is transmitted from its rodent reservoir by several flea species, including *C. felis.* Fleas probably play a role in maintenance and transmission of several other disease organisms such as *Bartonella henselae,* causing cat scratch disease. The cat flea is the intermediate host for the dog tapeworm, *Dipylidium caninum,* which can affect small children as well as dogs and cats.

Pets infested with fleas bite and scratch themselves repeatedly. In situations in which flea numbers are high, veterinarians occasionally see kittens and puppies near death from flea-produced anemia. Sensitized people suffer from flea bites, which can cause intense itching, with scratching opening the skin to infection.

Flea allergy dermatitis (FAD) is a severe condition found primarily in dogs, but also occasionally seen in cats. In a flea-allergic animal, flea salivary antigens initiate a cascade of symptoms, resulting in intense pruritus accompanied by scratching, biting, and self-inflicted trauma. An affected animal typically displays obsessive grooming behavior, with accompanying depilation, leaving the skin with weeping sores, often resulting in secondary infection. FAD is treated with corticosteroids, which possess undesirable side effects, especially when continuous use is required as in chronic FAD cases. Until development of FAD immunotherapy, successful treatment involves flea elimination from the animal's environment and flea bite prevention.

Fleas and their associated diseases can constitute over half a veterinary practice's caseload in some areas of the country. More energy and money are spent battling these insects than any other problem in veterinary medicine.

LIFE HISTORY

The cat flea is a cosmopolitan, eclectic species, having been recorded from more than three dozen species, including opossums, raccoons, kangaroos, and even birds. This wide host range explains this flea's ability to repopulate domestic animals after suppression efforts. Because it lacks host specificity and tends to feed on humans, the cat flea is a pest of both companion animals and humans with whom they share their abode.

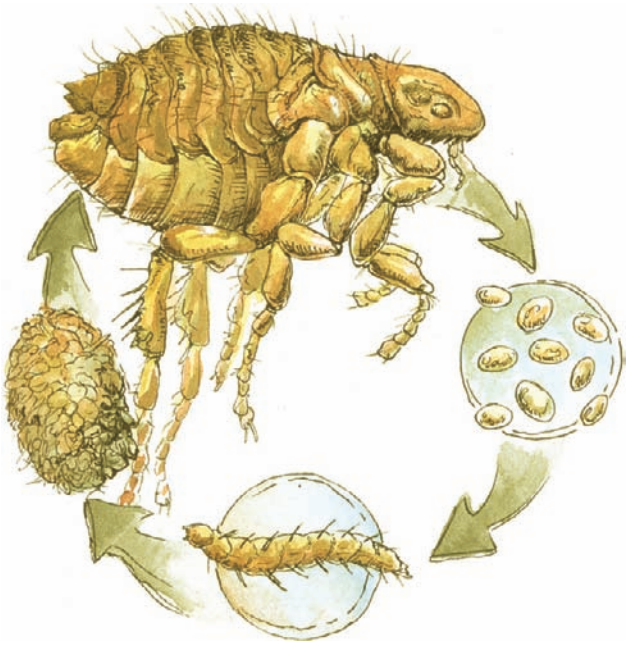

FIGURE 1 Cat flea life cycle: adult (top), eggs, larva, and cocoon (enclosing pupa). (Illustration by Cara J. Mitten.)

Adults

Once adult cat fleas locate a host, they tend to remain on that animal unless dislodged. They feed readily and mate on the host. Female fleas lay eggs while on the host and because the eggs are not sticky, they readily fall off into the host's environment, with large numbers accumulating in areas frequented by the animal. Each female flea can produce more than two dozen eggs per day. Adult fleas are about 1 to 4 mm in length and are strongly flattened from side to side. They are equipped with relatively long legs armed with strong outwardly projecting spines. Cat fleas have a collar of spines (ctenidium) on the back and another row of spines above the mouth. These characteristics allow for rapid movement through the host's hairs and also serve to resist removal from the fur.

Once the adult flea finds a host, it begins to feed. Typically the female mates and begins oviposition within a couple of days. On the host, a female flea averages about one egg per hour and, as a female flea can live on the host for several weeks, potential production can amount to hundreds of eggs in her lifetime. Only the adult stage is parasitic; all other life stages develop off the host (Fig. 1).

Eggs

Cat flea eggs are approximately 1 mm in length, with little surface structure other than aeropyles (permitting gaseous exchange for the developing embryo) and micropyles (for sperm entrance during fertilization). Typically, the larvae hatch within 24 to 48 h after oviposition, with more rapid hatching at warm temperatures.

Larvae

From the eggs emerge small, white, eyeless, legless larvae with chewing mouthparts. Because they seldom travel far from where they hatch, cat flea larvae are usually found in furniture, carpeting, or outside in areas frequented by flea hosts. Flea larvae have three instars that, under favorable conditions, can be completed in as little as 10 days. Larvae will develop only in protected microhabitats in which the relative humidity exceeds 75%. Cool temperatures, food shortages, or other unsuitable environmental conditions may extend larval developmental time to several weeks or a month. The third instar voids its gut approximately 24 h before initiation of cocoon construction. The white prepupa wanders until it locates an appropriate site for pupation and then begins to spin a silk cocoon. Frequently, environmental debris is incorporated into the cocoon, adhering to the sticky silk fibers, so that the cocoon may appear as a small dirt clod or lint ball.

Pupae

Within its cocoon, the prepupa molts to the pupa and continues metamorphosis to the adult flea within about 4 days, under favorable conditions. Length of the preemerged adult stadium is the most variable in the flea life cycle, ranging from less than a day to several months (or perhaps over a year). The mechanisms are not completely understood, but it appears that some individuals are programmed to delay emergence. Likely this is an evolutionary strategy whereby offspring emerge over an extended interval, ensuring that some successfully achieve hosts. Stimuli such as pressure, carbon dioxide, and warmth (triggers associated with mammalian hosts) serve as releasers, causing the adult flea to emerge from the cocoon. Upon emergence, if the flea does not locate a host immediately, it can survive for approximately 7 to 10 days (or longer under high-humidity and low-temperature conditions).

FLEA SUPPRESSION

Because fleas must have blood from a mammalian host to survive, treating host animals is the most efficient and effective means of suppressing fleas. There are several on-animal products that are effective for flea control. Many contain pyrethrins, which are safe, effective products but kill only fleas on the animal at the time of treatment and do not provide residual control. Other over-the-counter compounds include spot-on permethrin products, which are limited to canine use because they can be lethal to cats.

Veterinarians can recommend products that provide several weeks of control with a single application. Products prescribed by veterinarians for on-animal flea control are applied in a small volume (a few milliliters) on the back of the animal's neck. The material distributes over the body surface in skin oils. In addition to spot-on formulations, some products are available as sprays. These adulticides kill fleas on the animal within a few hours, then provide residual flea suppression for several weeks.

To forestall flea infestations, pets can be started on flea developmental inhibitors early in the season. Products containing insect development inhibitors can be applied topically, given orally (once monthly as a pill for dogs or a liquid added to a cat's food), or given as a 6-month injectable formulation for cats. Female fleas that feed on blood of treated animals subsequently are unable to reproduce.

Once pets have been treated, it will take a while for fleas in the environment to die off. Meanwhile, as they emerge, fleas will hop onto the animal; the host will continue to "harvest" fleas from the surrounding environment until they have been killed and no more are emerging. Insect growth regulators can be used to break the flea life cycle. Although these compounds do not kill adult fleas, they do prevent eggs and larvae from completing their development, ensuring that any fleas brought into the area will not establish a sustaining population.

Sanitation is an important flea suppression tactic; by eliminating larval developmental sites and destroying immature stages before they develop to the pestiferous adult stage, pets and people can be protected from fleas. Areas frequented by pets accumulate flea eggs and larval food, so these microhabitats should be vacuumed and treated to prevent flea infestations. These might include areas under furniture, animal bedding and sleeping quarters, and utility rooms or other areas where the pet spends time.

See Also the Following Articles

Bubonic Plague • Medical Entomology • Siphonaptera • Veterinary Entomology

Further Reading

Dryden, M. W. (1997). Fleas. *In* "Mallis Handbook of Pest Control," 8th ed., Chap. 16, pp. 747–770. Franzak & Foster, Cleveland.

Dryden, M. W., and Rust, M. K. (1994). The cat flea: Biology, ecology and control. *Vet. Parasitol.* **52,** 1–19.

Hinkle, N. C., Rust, M. K., and Reierson, D. A. (1997). Biorational approaches to flea (Siphonaptera: Pulicidae) suppression: Present and future. *J. Agric. Entomol.* **14,** 309–321.

Rust, M. K., and Dryden, M. W. (1997). The biology, ecology and management of the cat flea. *Annu. Rev. Entomol.* **42,** 451–473.

Taylor, M. A. (2001). Recent developments in ectoparasiticides. *Vet. J.* **161,** 253–268.

Cave Insects

Francis G. Howarth
Hawaii Biological Survey, B. P. Bishop Museum

Caves and associated subterranean voids harbor extraordinary ecosystems inhabited by equally remarkable animals. Insects and arachnids dominate terrestrial habitats, whereas crustaceans

dominate aquatic systems. This article describes the subterranean biome, highlighting terrestrial systems and the insects that are obligately adapted to live permanently in underground voids.

DISCOVERY AND CHARACTERIZATION OF CAVE ARTHROPODS

Why an animal would abandon the lighted world and lose such adaptive characters as eyes, pigment, and dispersal ability to live permanently in perpetually damp, dark, barren caves has long fascinated both biologists and laymen. In fact, it is these pale, blind obligate cave species that one usually envisages under the rubric of cave animal, and it is this group that is featured in this article. However, numerous other animals live all or part of their life cycles in caves or are regular visitors. Although some cave insects were known in ancient times, the first scientific writings on this topic began in northern Italy in the mid-16th century with the discovery of blind aquatic crustaceans in cave streams. The science of cave biology (biospeleology) was founded in the mid- to late 19th century with studies of limestone caves in southern Europe by Schinner and continued into the early 20th century by Racovitza and colleagues. They devised the currently used classification scheme for cavernicoles, based on the degree of association with caves. Also, in the mid-1800s, obligate cave animals were discovered in Mammoth Cave, Kentucky, and in a few other North American limestone caves, but the study of North American cave faunas generally lagged behind Europe for the next century. The loss of eyes and other apparently adaptive characters led to a revival, circa the turn of the 20th century, of Lamarckian (i.e., acquired rather than inherited characteristics) theories to explain their evolution. During this period several major expeditions went to tropical regions to search for obligate cave faunas, but for a variety of reasons none were found or recognized. The apparent absence of tropical troglobites (obligate cave species) and the relictual nature of temperate cave animals led to the development and general acceptance of the theory that these animals evolved only after populations were stranded in caves by changing climates that extinguished their surface relatives.

Troglobitic Adaptations

The most conspicuous aspect displayed by obligate cave arthropods is the reduction of structures normally considered adaptive (e.g., eyes, pigment, wings, and cuticle thickness). Compare the closely related surface and cave insects shown in Figs. 1A and 1B. Cave species also often lack a circadian rhythm and have relatively low metabolic and reproductive rates. A few characters are often enhanced, including modified structures such as increased hairiness, enlarged sensory organs, longer appendages, and specialized tarsi. These morphological, physiological, and behavioral changes allow the animals to maintain water balance, breathe unusual gas mixtures, disperse, reproduce, and locate food and other resources in their environment. The remarkable convergent evolution of troglomorphy (adaptations to caves) among unrelated cave species in different regions of the world indicates that selective pressures must be similar in all such environments.

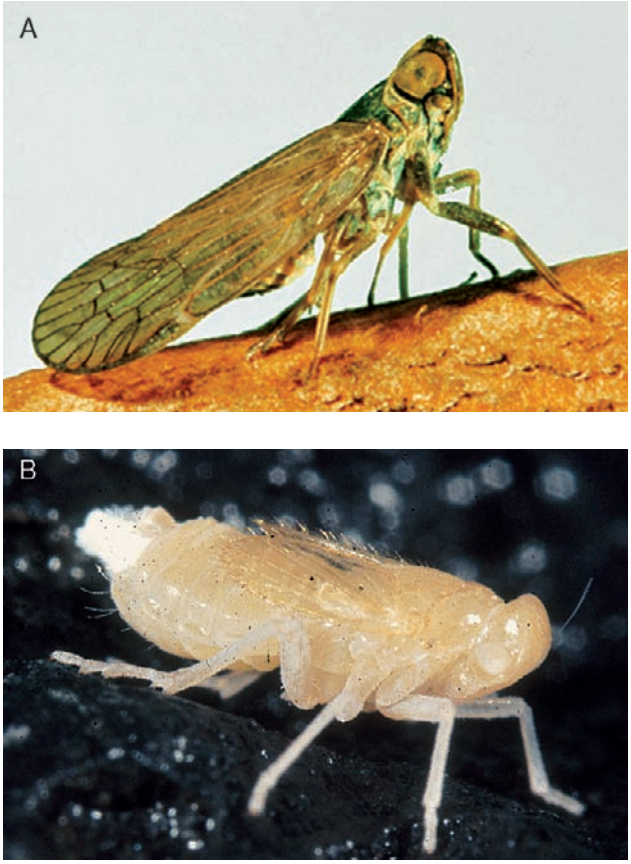

FIGURE 1 Cave and surface cixiid planthoppers. (A) Rain forest *Oliarus* species from Maui Island; note large eyes, dark color, and functional wings. (B) Adult female cave-adapted *Oliarus polyphemus* from Hawaii Island; note absence of eyes, enlarged antennae, and reduced wings and pigment. (Photographs by W. P. Mull, used with permission.)

Taxonomic Overview of Troglobites

TERRESTRIAL CAVE ARTHROPODS Insects, arachnids, and millipedes are the dominant terrestrial groups living in caves. Not all orders are represented, however. Among the Hexapoda, the orders Collembola, Orthoptera, Hemiptera, Coleoptera, and Diptera predominate. The springtails (Collembola) are represented by many troglophilic (facultative cave residents) and troglobitic species and are important scavengers in many caves. Most cavernicolous orthopterans are troglophilic or trogloxenic (roosting in caves), with the cave crickets (Rhiphidophoridae) being the best known. As more tropical caves are studied, many new species of troglobitic true crickets (Gryllidae) are being described. Among

Hemiptera, both suborders occur in caves: true bugs (Heteroptera) and planthoppers and allies (Auchenorrhyncha or Homoptera of some classification systems). Thread-legged bugs (Reduviidae) are common troglophiles in warmer caves, and a few troglobitic forms are known from the tropics. Most species are cryptic, and many new cave species await discovery. Several other hemipteran families contain troglophilic species. Planthoppers, especially Cixiidae (Fig. 1B), are common in tropical caves. Ongoing surveys for cixiids indicate that each isolated cave system may harbor one or more cave-adapted species, and the group may be among the most speciose families in caves, rivaling even the carabid ground beetles in temperate caves.

Beetles (Coleoptera), especially the families Carabidae, Leiodidae, and Staphylinidae, are especially well represented in the temperate caves. For example, the endemic North American ground beetle genus *Pseudanophthalmus* contains at least 250 species, which, with one exception, are found only in caves. Flies (Diptera) are dominant troglophiles in both tropical and temperate caves, but only a few blind, flightless troglobitic species are known.

Troglobitic species are also found in the orders Diplura, Thysanura, Blattodea, Dermaptera, Grylloblattodea, Psocoptera, and Lepidoptera. Troglobitic bristletails (Diplura) occur mainly in temperate caves. Cockroaches (Blattodea) are well represented in tropical caves, and many new species await description. Only a few cave-adapted earwigs are known, and most are from oceanic islands. Grylloblattids are restricted to glaciated mountains in northwestern North America and eastern Asia. They characteristically inhabit caves and crevices; however, most species also venture outside to feed in damp surface habitats. Many moths habitually roost in caves, and some are troglophilic scavengers or root feeders. A few are blind and flightless troglobites.

The arachnids are second only to the insects in numbers of terrestrial cave species. The spiders are common denizens of caves, with numerous troglobitic forms known from temperate and tropical caves. In many tropical caves, spiders instead of ground beetles are the top predators. Pseudoscorpions are also well represented in temperate and tropical caves, and over 300 cave-adapted species representing most families are known. Harvestmen (Opilionida) are more restricted in distribution, but most of the 26 families contain troglobitic species. Some surface species roost in caves in huge numbers. Mites (Acari) are often abundant and diverse in caves, especially species associated with guano. Most terrestrial cave species are troglophilic, but a few families, such as the Rhagidiidae, contain many troglobites. Cavernicolous species are also known among the palpigrades, schizomids, amblypygids, scorpions, and ricinuleids.

Myriapods are also well represented in caves. The millipedes are the third major group of cavernicolous arthropods, especially in temperate caves, where they are often the dominant scavengers in the ecosystem. The orders containing the most cave species are Julida, with numerous troglobites in Europe and North America; Chordeumatida and Polydesmida, with troglobites in Europe, North America and Japan; and Callipodida, with troglobites in Europe and the Near East. Four other orders (Polyxenida, Glomerida, Spirobolida, and Spirostrepida) each have a few cave-adapted species. Cave millipedes from the tropics are still poorly known, and many new species undoubtedly await discovery. Many ground-inhabiting centipedes regularly enter caves. Whether they can live and reproduce underground is unknown for most species, but a few are troglophilic or troglobitic. The rock centipedes (Lithobiomorpha) are widespread and include several troglobitic species. A few troglobitic giant centipedes (Scolopendromorpha) are known from the tropics. An undescribed 8-cm-long Scutigerimorpha from North Queensland, Australia, is one of the largest terrestrial troglobites known.

Two groups of terrestrial Crustacea are found in caves. Isopods in the suborder Oniscidea have adapted to caves many times, especially in the Mediterranean region and in the tropics. Fourteen of the 34 recognized families contain cave species. In contrast, only a few terrestrial amphipods (Talitridae) are found in caves, and most are from islands.

AQUATIC CAVE ARTHROPODS Aquatic subterranean habitats include underground lakes and streams, perched pools of water, water films, and water-filled phreatic aquifers. These aquatic habitats support diverse faunas of troglobitic (or stygobitic) arthropods. By far the dominant group is the crustaceans, with about 2700 cave-adapted species known worldwide. Water mites (Acari) are also well represented, especially in smaller interstitial habitats. Few insects have invaded subterranean aquatic habitats. The most successful group is the dytiscid diving beetles, several species of which are known from aquifers in Africa, Europe, North America, and Japan. Two troglobitic water bugs are known: a blind water scorpion (Nepidae) from a cave in Romania and a terrestrial water treader (Mesoveliidae) from Hawaii.

Zoogeography of Cave Arthropods

Until recently, obligate cave species were thought to occur mainly in temperate limestone caves, and the cave faunas of temperate Europe and North America are well characterized. Diverse cave faunas are also known from Japan, Tasmania, and New Zealand. However, in the past few decades discoveries of significant cave faunas in tropical caves, lava tubes, and even fractured rock layers have revolutionized our understanding of cave biology. These findings suggest that troglobites have evolved wherever suitable subterranean voids are available for sufficient time. They are now known from most regions that have been appropriately investigated. Thus rather than being exceptional, cave adaptation must be a general and predictable process among animals adapting to exploit underground resources.

In hindsight, the early expeditions to the tropics missed troglobites for three main reasons. (1) The environment of caves: Troglobites are restricted to deeper, constantly moist

passages. Because cave temperatures are usually near the mean annual surface temperature (MAST) over the cave and, in the tropics, the surface temperature rises above and falls below MAST almost every day, most tropical caves are subjected to drying winds created by the sinking cold nighttime air. (2) Accessibility: The higher solution rate of limestone in the tropics creates large open cave systems, exacerbating the effects of the daily drying winds and making the deeper moist cave passages, where the troglobites are found, beyond the limits of safe exploration using the equipment available at that time. In addition, the caves found and explored were often bird and bat roosts, and the biologists could fill their containers with new species without going deeper. (3) Systematics: Ironically, many troglobites were collected, but the species belonged to groups unrelated to the animals found in temperate caves, and in fact unrelated to anything the temperate-based taxonomists had seen, so their status in the cave went unrecognized. As in all fields in biology, evolutionary biology is only as good as the systematics research upon which it is based.

Each cave region is inhabited by representatives of the surface fauna currently or historically living over the caves. Only a few surface taxa within each region successfully invaded caves. In general, the surface ancestors possessed characters that facilitated their shift into underground environments; i.e., they were already adapted to live in dark, moist rocky habitats and utilized food that was relatively common in caves. The chief ancestral habitats for terrestrial cave species include rocky margins of rivers, lakes, and seashores; leaf litter and moss in wet forests; and moist rocky terrains. Each cave system harbors relatively few species of troglobites; even the most diverse known fauna—that in the Postojna–Planina System, Slovenia—totals only 84 species. In North America, Mammoth Cave supports the most species (41). Among lava tubes, Bayliss Cave (North Queensland, Australia) contains the highest number (25). Because of the restricted distribution of each species, cave habitats are often likened to islands. Despite the few species found in each cave, the overall number of troglobites is quite large since subterranean habitats are much more extensive and widespread than is often assumed. Karst landscapes cover about 15% of the earth's surface, and cavernous lava and fractured rock habitats have not been mapped but may cover another 5% or more. Submarine caves have barely been investigated, but the diverse fauna derived from marine ancestors found in anchialine systems along seacoasts indicates that caves and cave-like habitats below the seafloor may harbor diverse ecosystems at least in shallow coastal areas.

SUBTERRANEAN BIOME

Caves and Voids

Caves are subterranean voids large enough for humans to enter, but intermediate-sized voids (i.e., mesocaverns) smaller than caves but larger than capillary spaces are also important for terrestrial cave insects. Terrestrial animals rarely exploit capillary-sized spaces underground, but water-filled pore spaces (i.e., interstitial habitats) are often inhabited by numerous tiny species of stygobites. Caves and voids can form in three ways: solution, erosion, and volcanism. The largest and best known caves are dissolved in limestone, calcium carbonate. Limestone is structurally strong yet readily dissolves in weak acid, such as the small amounts of carbonic acid normally found in groundwater. The process is slow, but over millennia large interconnected systems of caves and voids can form in limestone exposed to weathering. Caves created by solution can also form in other soluble rocks, such as gypsum (hydrated calcium sulfate) and dolomite (magnesium calcium carbonate), but the caves formed are usually less stable than those in limestone. Erosional caves form during landslides and tectonic events, as well as by groundwater removing loose material from under a cap rock. Erosional caves are usually ephemeral but, in some areas, they are re-created continuously and so remain available for colonization. Tectonic caves are common on volcanoes, but lava tubes are more familiar cave features. Lava tubes form by the roofing over of lava channels during an eruption. Because the roof insulates the flow, lava tubes become efficient transporters of lava away from the vent, and long complex caves can be built over time by long-lived eruptions. Mesocavernous habitats are more extensive than caves and can be found in rock strata not suitable for supporting cave-size passages. Mesocaverns also occur in fractured rock strata and in cobbles deposited by rivers.

Environment of Caves

The terrestrial cave environment is strongly zonal (Fig. 2). Three zones are obvious: (1) the entrance zone where the surface and underground environments meet; (2) the twilight zone between the limits of vascular plants and total darkness; and (3) the dark zone. From biological and environmental perspectives, the dark zone can be subdivided into three zones: (a) the transition zone where short-term

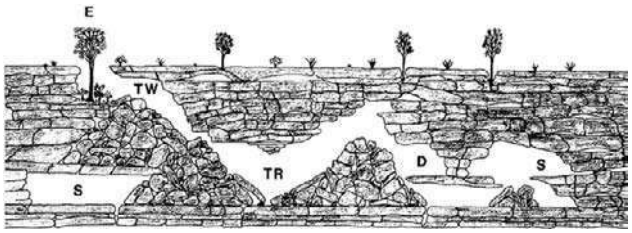

FIGURE 2 Profile view of a representative cave showing the five environmental zones. Not shown to scale; length and depth are compressed. Key: D, deep zone; E, entrance zone; S, stagnant air zone; TR, transition zone; and TW, twilight zone. [Illustration by N. C. Howarth. Reproduced, with permission, from E. C. Dudley (ed.), 1991, "The Unity of Evolutionary Biology," Dioscorides Press, an imprint of Timber Press, Portland, OR.]

climatic events on the surface are still felt; (b) the deep cave zone where the atmosphere remains saturated with water vapor; and (c) the stagnant air zone where decomposition gases, especially carbon dioxide, can accumulate. The boundary between each zone is often dynamic and is determined by size, shape, orientation, and location of entrances in relation to the surface environment and size and shape of the cave passages, as well as to the climate on the surface and availability of water. Because air exchange is reduced in smaller spaces, the environment within most mesocaverns probably remains in the stagnant air zone. Each zone often harbors a different community of organisms, with the obligate cave species found only in the inner two zones. The deep cave and stagnant air zones contain a harsh environment for most surface-dwelling organisms. It is a perpetually dark, wet, three-dimensional maze without many of the cues used by surface species and with often abnormally high concentrations of carbon dioxide. In many caves in temperate regions, the transition zone is evident only in winter when the outside temperature is below cave temperature.

Energy Sources and Nutrient Cycling in Caves

Unlike capillary spaces typical of soils, which act as filters capturing water and nutrients near the surface, caves and mesocaverns act as conduits for water and nutrients. In cavernous regions, a significant amount of organic material sinks or is carried into deeper underground voids where it is inaccessible to most species adapted to surface habitats. The principal mechanisms that transport material underground are sinking streams, percolating rainwater, trogloxenes, animals blundering into caves, and deeply penetrating plant roots. A few cave communities are known to rely on food energy created underground without the aid of sunlight by chemoautotrophic microbes. Sinking streams are more important in transporting food into limestone caves than in lava and other caves, because streams are important in creating and maintaining solution caves. Plants growing on barren rocky substrates such as lava and limestone often must send their roots deep into crevices and caves to obtain water and nutrients. Because higher temperatures result in higher rates of water loss from leaves and higher rates of leaching of tropical soils, and because there is a continuous growing season without a spring recharge of water, plant roots must penetrate deeper underground (sometimes in excess of 100 m) and are, therefore, generally more important in tropical caves than in temperate caves.

Most troglobites are detritivores or scavengers feeding on decaying organic matter and the associated microbes. Living tree roots provide food directly for several obligate cave insects. A relatively large percentage of troglobites are predators, attesting to the role of lost surface animals in bringing in food. It is these available food resources that enable the evolution of troglobites, which are highly specialized to exploit resources within medium-sized subterranean voids. They colonize or temporarily exploit cave-sized passages only where the physical environment is suitable. Most caves appear barren and therefore often are believed to be food-poor environments. However, food can be locally abundant, and exploiting such a patchy resource in a harsh, maze-like environment is probably more critical than paucity per se.

In addition to troglobites many other organisms enter caves. Many arthropods seek out caves for estivation or hibernation sites during periods of harsh weather. Some, such as agrotine moths and cave crickets, use caves for daytime retreats and sometimes oviposition sites and emerge at night to forage in the neighboring forest. Troglophilic arthropods enter to feed on guano and other organic material deposited or brought in by roosting bats, birds, crickets, and other trogloxenes. Parasites and other associates of trogloxenes also live in caves, and some of these, such as nycteribiid and streblid flies on bats, show some troglomorphies. Many leaf-litter and soil arthropods living in caves feed on accumulations of organic material left by sinking streams. These resources are usually more abundant near entrances and in the transition zone. Only a portion of the surface-inhabiting species in each region can cope with the environment and exploit these food resources. Some troglophiles apparently leave caves only to disperse to new sites, but most show no morphological adaptations to living in caves.

CONSERVATION OF CAVE LIFE

The fantastic adaptations displayed by obligate cave animals have long intrigued biologists. Their often narrow environmental tolerances, coupled with their island-like habitats, have reinforced the view that these animals are fragile, lead an endangered existence, and are in need of conservation. However, development of conservation programs is hampered by a severe lack of data about the species present and their status. Discoveries in the past few decades of cave ecosystems in a variety of cavernous rocks in diverse regions have revolutionized our understanding of cave life. We now believe that cave colonization and adaptation are general phenomena and occur wherever there are suitable underground voids available for evolutionary time. Most cave species remain undiscovered; in fact, the cave faunas of large areas containing caves, especially in the tropics, remain unsurveyed and unknown. Unfortunately, many cave systems are being destroyed before their faunas become known. The major anthropogenic threats to cave faunas include (1) mining of the surrounding rock, (2) changes in land use over subterranean habitats such as deforestation and urbanization, (3) alteration of groundwater flow patterns, (4) waste disposal and pollution, (5) invasion by nonindigenous species, (6) disruption of food inputs, and (7) direct human disturbance during visitation. Biological surveys are urgently needed. Also, recent systematic studies reveal that cave arthropod faunas are far more diverse than previously thought, indicating that priority should be focused on recognizing and protecting each distinct population rather than protecting a single population of each conventional species.

Conservation efforts must mitigate threats affecting the system, as well as recognize emerging threats. Generally, species extinctions result from novel perturbations, e.g., new stresses with which a species has had little experience during its evolution. Ecological studies are needed that improve our understanding of the functioning ecosystem, as well as understanding of natural successional processes. However, experimental ecological studies in caves are problematic because in few other habitats are humans so dramatically intruders as in caves. Not only do researchers affect the environment of the passages they study, but also they cannot sample the medium-sized voids where the major activity usually occurs. Caves are a fragile window through which we can see and study the fauna living within cavernous rock. Protected areas must include a sustainable portion of the ecosystem as well as suitable source areas for food and water resources. This usually represents an area larger than the footprint of the known cave.

RESEARCH OPPORTUNITIES

The bizarre adaptations displayed by troglobites make them excellent animals for evolutionary research. Recent advances in phylogenetic methods and molecular techniques provide important new tools for deciphering relationships among cave animals and their surface relatives. The discovery that close surface relatives are still extant for many tropical and island troglobites allows more appropriate comparative studies between species pairs adapted to wildly different environments. These studies should provide more critical understanding of how certain adaptations correlate with environmental parameters, as well as a better understanding of evolution in general. Some of these studies are in progress, for example, the work of Culver and colleagues on *Gammarus minus* in springs and caves in the eastern United States. Individual species of troglobites frequently have restricted distributions even within a given area of caves. Usually such a limited distribution indicates the existence of a barrier to subterranean dispersal, but not always. Critical morphological and behavioral studies, corroborated by modern molecular techniques, are showing that some troglobites thought to be widespread actually are composed of several more or less reproductively isolated populations. It has been assumed that cave adaptation was a dead end and that each of these populations evolved separately from the same or closely related surface ancestors that independently invaded caves. However, recent research by Hoch and colleagues on Hawaiian cixiid planthoppers suggests that some troglobites can disperse to new caves through underground voids and diverge into new species.

Caves are island-like habitats that support distinct ecosystems composed of communities of highly specialized organisms. Because the environment is discrete, rigorous, and easily defined, it provides an ideal system in which to conduct ecological studies. The number of species is usually manageable. The physical environment is rigidly constrained by the geological and environmental setting, and the environmental parameters can be determined with great precision because the habitat is surrounded and moderated by thick layers of rock. However, it is a rigorous, high-stress environment and difficult for humans to access and envision because it is so foreign to human experience. Also, one cannot enter or sample the mesocaverns where perhaps most cave animals live. These disadvantages can be overcome by comparing passages differing in the parameter of interest or by designing experiments that manipulate the parameter being studied in the natural environment. Biospeleology is still in the discovery phase. Although our understanding of cave biology has progressed substantially, results of future studies on evolution and ecology will be exciting and add significantly to our fascination with caves.

See Also the Following Articles

Aquatic Habitats • Conservation

Further Reading

Barr, T. C. (1968). Cave ecology and the evolution of troglobites. *Evol. Biol.* **2,** 35–102.

Camacho, A. I. (ed.) (1992). "The Natural History of Biospeleology." Museo Nacional de Ciencias Naturales, Madrid.

Chapman, P. (1993). "Caves and Cave Life." Harper Collins, London.

Culver, D. C. (1982). "Cave Life: Evolution and Ecology." Harvard University Press, Cambridge, MA.

Culver, D. C. (ed.) (1985). Special issue, regressive evolution. *Nat. Speleol. Soc. Bull.* **47(2),** 70–162.

Culver, D. C., Kane, T. C., and Fong, D. W. (1995). "Adaptation and Natural Selection in Caves. The Evolution of *Gammarus minus*." Harvard University Press, Cambridge, MA.

Hoch, H., and Howarth, F. G. (1993). Evolutionary dynamics of behavioral divergence among populations of the Hawaiian cave-dwelling planthopper *Oliarus polyphemus* (Homoptera: Fulgoroidea). *Pac. Sci.* **47,** 303–318.

Howarth, F. G. (1983). Ecology of cave arthropods. *Annu. Rev. Entomol.* **28,** 365–389.

Humphries, W. F. (ed.) (1993). The biogeography of Cape Range, Western Australia. *Rec. West. Aust. Mus.* Suppl. 45.

Juberthie, C., and Decu, V. (1996). "Encyclopaedia Biospeologica," Vol. I. Soc. Biospeologie, Moulis, France.

Juberthie, C., and Decu, V. (1996). "Encyclopaedia Biospeologica," Vol. II. Soc. Biospeologie, Moulis, France.

Vandel, A. (1965). "Biospeleology. The Biology of Cavernicolous Animals." Pergamon, Oxford. [Translated by B. E. Freeman]

Wilkens, H., Culver, D. C., and Humphries, W. F. (eds.) (2000). "Subterranean Ecosystems." Elsevier, Amsterdam.

Cell Culture

Dwight E. Lynn

U. S. Department of Agriculture

Cell culture is the technique in which cells are removed from an organism and placed in a fluid medium. Under proper conditions, the cells can live and even grow. The

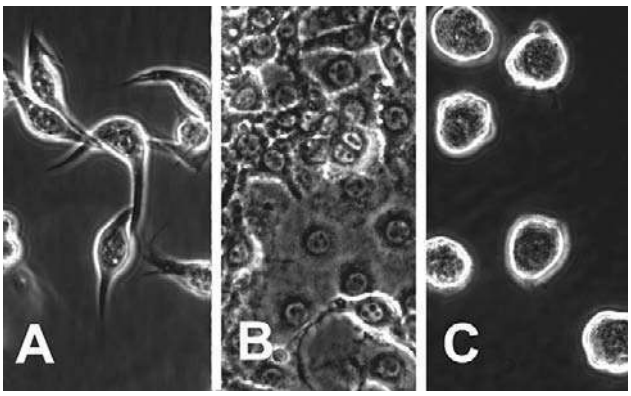

FIGURE 1 Typical appearances of insect cells in culture by phase contrast microscopy: (A) spindle shaped (fibroblast-like), (B) epithelial shaped, and (C) round.

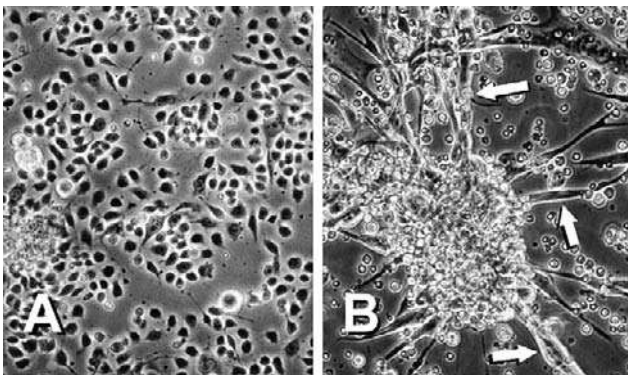

FIGURE 2 Cell shape change in response to treatment with the insect molting hormone: (A) Untreated cells. (B) Cells treated for 2 weeks with 20-hydroxyecdysone. Arrows point to cells that were contracting in the culture.

growth can be characterized by cell division (mitosis) or by other processes, such as differentiation, during which the cells can change into specific types that are capable of functions analogous to tissues or organs in the whole organism. The practice of cell culture (and its close cousins, tissue culture and organ culture) originated in a Yale University laboratory in 1907, when Ross Harrison removed nerves of a frog and maintained them in a simple salt solution for several days. Within a very few years a visiting scientist in Harrison's laboratory, Richard Goldschmidt, reported on the first cell cultures from an insect. For the next half-century, insect cell culture was used periodically in a variety of experiments, such as studying the pathogenesis of viruses, but the field received a great boost when the Australian Thomas D. C. Grace succeeded in obtaining four cell lines from the emperor gum moth, *Antheraea eucalypti.* These lines were capable of continuous growth, requiring periodic subculturing.

In the years since Grace's report, numerous other continuous insect cell lines have been developed—over 500 lines from more than 100 different insect species. Under microscopic examination, cells take on one of several distinct morphologies, including spindle shaped, epithelial, and round to oval (Fig. 1). Cell cultures are frequently used in research and biotechnology.

USES IN PHYSIOLOGY AND DEVELOPMENTAL BIOLOGY

Harrison's earliest work was designed to examine the physiology of a nerve outside the living organism, and similar uses have been made of insect cells and tissue. By removing the tissue or cells from the insect, it is possible to delineate how individual compounds affect them. Some of the most useful work has been with ecdysone, the insect molting hormone. In 1972 Anne-Marie Courgeon showed that exposing a cell line from *Drosophila melanogaster* to β-ecdysone (a particular form of ecdysone now known as 20-hydroxyedysone) caused rounded cells to change to aggregates of highly elongated cells. Lynn and Hung found that a cell line from a small wasp can undergo a similar morphogenesis with the added feature that the elongated cells are highly contractile, like muscle cells (Fig. 2).

USES IN PATHOLOGY

Certain disease-causing organisms, such as viruses, rickettsia, and certain protozoans are obligate pathogens. Cell cultures can be extremely useful with such pathogens because this is often the only way to grow them outside a whole animal or plant and make them much easier to study. Some of the early work with insect cell culture was initiated with a group of viruses known as nucleopolyhedrovirus. These viruses cause diseases in a large number of pest insects and thus are potential biological control agents. Cell cultures grown in large volumes can be used to produce these viruses for biologically based pesticides.

Some insects are also notorious for their ability to transmit diseases (such as malaria and yellow fever) to higher animals and plants. Cell cultures from mosquitoes and other insects can also be used to study these pathogens.

USES IN MEDICINE AND PHARMACOLOGY

Decades ago, researchers showed that a gene in baculoviruses could be replaced through genetic engineering with genes for other proteins. When insect cells are infected with these modified viruses, the cultures can produce a large quantity of the protein. This technique—the baculovirus expression vector system—has facilitated a new use for insect cell cultures for the production of vaccines, growth factors, and other materials useful in medicine. Over 1800 articles have been published on the use of insect cells to produce various recombinant proteins.

See Also the Following Articles
Biotechnology and Insects • Genetic Engineering

Further Reading

Freshney, R. I. (2000). "Culture of Animal Cells: A Manual of Basic Technique." 4th ed. Wiley-Liss, New York.

King, L. A., and Possee, R. D. (1992). "The Baculovirus Expression System: A Laboratory Guide." Chapman & Hall, London.

Lynn, D. E. (1996). Development and characterization of insect cell lines. *Cytotechnology* **20,** 3–11.

Lynn, D. E., and Hung, A. C. F. (1991). Development of continuous cell lines from the egg parasitoids *Trichogramma confusum* and *T. exiguum. Arch. Insect Biochem. Physiol.* **18,** 99–104.

Maramorosch, K., and McIntosh, A. H. (eds.) (1994). "Insect Cell Biotechnology." CRC Press, Boca Raton, FL.

Maramorosch, K., and Mitsuhashi, J. (eds.) (1997). "Invertebrate Cell Culture: Novel Directions and Biotechnology Applications." Science Publishers, Enfield, NH.

Shuler, M. L., Wood, H. A., Granados, R. R., and Hammer, D. A. (eds.) (1995). "Baculovirus Expression Systems and Biopesticides." Wiley-Liss, New York.

Smith, G. E., Fraser, M. J., and Summers, M. D. (1983). Molecular engineering of the *Autographa californica* nuclear polyhedrosis virus genome: Deletion mutations within the polyhedrin gene. *J. Virol.* **46,** 584–593.

Chemical Defense

Murray S. Blum

University of Georgia

Biologists have become keenly aware that insects possess a remarkable ability to biosynthesize a large variety of compounds for use as agents of chemical defense against their omnipresent enemies. Many of these compounds are unique products (e.g., cantharidin, or Spanish fly, produced by blister beetles) with diverse modes of toxicity against a variety of vertebrate and invertebrate predators. These defensive secretions often originate from unlikely sources that appear to optimize the effectiveness of the chemical defensive systems. Ultimately, for countless species of insects, chemical defense and survival are synonymous.

ECLECTIC ORIGINS, FUNCTIONS, AND RESERVOIRS OF DEFENSIVE COMPOUNDS

It would be no exaggeration to state that the tremendous abundance of insects constitutes the primary food source for diverse vertebrate and invertebrate predators. For insects in a variety of orders, blunting the attacks of their omnipresent predators is identified either with the production of defensive compounds in exocrine glands or with the acquisition of these compounds from external sources. These deterrent allomones sometimes represent novel natural products that have a very limited distribution in the Insecta. In short, exocrine compounds, characteristic of species in orders or genera, have evolved to function as versatile agents of chemical defense.

It has been generally assumed that *de novo* biosynthesis characterizes the origins of insect defensive compounds. However, recent investigations suggest that novel insect defensive allomones, including the complex amide pederin from staphylinid beetles (*Paederus* spp.) and unique steroids from dytiscid beetles, are biosynthesized by endosymbiotes. These results raise the question of whether other novel insect allomones, including cantharidin and steroids in chrysomelids and lampyrids, may have microbial origins.

Often, however, the deterrent allomones constitute ingested allelochemicals such as cardenolides (milkweeds) and toxic pyrrolizidine alkaloids (asters, heliotrope). Furthermore, some of these plant natural products have been metabolized after ingestion into products that are suitable for sequestration and use as deterrents, as for ingested steroids from milkweeds by the monarch butterfly, *Danaus plexippus.* These compounds are also transferred to eggs to function as effective predator deterrents. In addition, these allelochemicals may be added to the secretions of exocrine glands, further increasing the deterrent properties of these exudates. The dependence on ingested plant natural products of some insect species is further emphasized by the utilization of "stolen" defensive exudates that essentially represent mixtures of pure plant allelochemicals that have been appropriated, unchanged, from the host plants.

In some species, ingested allelochemicals are sexually transmitted by the male as a copulatory "bonus" for the female. For example, the sperm-rich spermatophore of ithomiine butterflies is accompanied by pyrrolizidine alkaloids that provide protection for the female and her eggs. Importantly, this very adaptive system is functional because the spermatozoa are resistant to the well-known toxic effect of these alkaloids.

Some allelochemicals also possess great selective value for insects as antibiotic agents. Alkaloids such as α-tomatine, a constituent of tomatoes, reduce the infectivity of bacteria and fungi for lepidopterous larvae. Other compounds reduce the activity of viruses and in some cases are highly toxic to insect parasitoids.

Insects have adapted for defensive functions a variety of glands not identified as defensive organs. For example, salivary glands have been converted into defensive structures that deliver deterrent compounds biosynthesized in these glands. Even respiratory structures have assumed the role of deterrent organs as further testimony to the insect emphasis on defensive adaptations. For a variety of insect species, chemical defense is clearly identified with survival.

The defensive value of insect allomones has been further enhanced by the ability of these arthropods to adapt a variety of these natural products to subserve a surprising variety of multiple functions. This phenomenon, semiochemical parsimony, has been particularly emphasized by insect species such as fire ants, whose alkaloidal venoms possess a dazzling variety of pharmacological activities. The same may be said of cantharidin, the potent vesicant from blister beetles (Spanish fly).

Things are seldom what they seem. The sting-associated glands of bees and wasps are obvious candidates for the production of compounds with considerable deterrent activities. These glands have evolved as biosynthetic centers clearly dedicated to the biogenesis of pharmacologically active compounds that can be delivered by the sting in an unambiguous act of defense. On the other hand, some glands clearly identified with nondefensive functions have been adapted by a variety of insect species to function as defensive organs with varied functions. Furthermore, the deterrent efficiency of these secretions may be considerably enhanced by adding repellent plant natural products to the exudate. And insects have not neglected adapting enteric products to discourage their omnipresent predators. If all else fails, many insects eject blood, sometimes fortified with toxic allomones, at their adversaries with startling results. It is no exaggeration to state that for these species, bleeding has often provided an extraordinary means of deterring a variety of aggressive predators.

VARIETY OF SALIVARY DEFENSIVE FUNCTIONS

Salivary Venoms

The spitting cobra, *Naja nigricollis,* has an insect parallel, both in terms of the general chemistry of the saliva and the ability to accurately "fire" the venom at a moving target. For example, *Platymeris rhadamanthus* is a black and orange assassin bug (Reduviidae) that is very conspicuous because of its aposematic (warning) coloration. This insect can eject its saliva for a distance up to 30 cm, and if this enzyme-rich solution (proteases, hyaluronidase, phospholipase) strikes the nose or eye membranes of a vertebrate, intense pain, edema, and considerable vasodilation may follow. The saliva of *P. rhadamanthus* is admirably suited to deter vertebrate predators, including birds and reptiles. This salivary venom has clearly been evolved for predation on invertebrates, and rather than a specific site of action, it is reported to attack many organs simultaneously. Its speed of paralytic action is very pronounced: an American cockroach *(Periplaneta americana)* can be totally immobilized in about 4 s.

Entspannungsschwimmen (Chemically Induced Aquatic Propulsion)

The proteinaceous saliva of the hemipteran *Velia capraii* has been adapted to promote escape from potential predators in aquatic environments. This aquatic true bug will discharge its saliva onto the water surface, a reaction that results in lowering the surface tension of the water behind the bug. Under these circumstances, *V. capraii* is rapidly propelled across the water surface, putting considerable distance between itself and the source of the disturbance. Discharge of saliva posteriorly from the rostrum may project the bug 10 to 25 cm on the contracting water surface on which it is riding.

Allomonal Pheromones

Male bumble bees (*Bombus* spp.) scent-mark territorial sites with cephalic products that are very odoriferous. The secretions, which originate in the cephalic lobes of the salivary glands, are dominated by terpenes, some of which are well-known defensive compounds. This appears to be an excellent example of semiochemical parsimony, with the males utilizing the compounds both as territorial pheromones and as defensive allomones.

Salivary "Glues"

Termite workers in both primitive and highly evolved genera secrete defensive exudates that are rapidly converted to rubberlike or resinous products that can rapidly entangle small predators such as ants. This conversion frequently reflects the polymerization of salivary proteins that have reacted with *p*-benzoquinone, a highly reactive salivary defensive product. Similar systems for generating entangling salivas have been detected in a diversity of termite genera, including *Mastotermes, Microtermes, Hypotermes,* and *Odontotermes.*

Termites in other genera discharge cephalic exudates that are fortified with toxic terpenes. Species of *Nasutitermes* and *Tenuirostritermes* secrete mixtures of compounds that rapidly form a resin that entangles ants and other small predators. The presence of monoterpene hydrocarbons is probably responsible for killing ants and, in addition, may function as an alarm pheromone for recruiting termite soldiers.

NONSALIVARY ENTANGLING SECRETIONS

The posterior abdominal tergites and cerci of cockroaches in a variety of genera are covered with a viscous secretion that can act as an entangling glue for small predators. Species in genera as diverse as *Blatta* and *Pseudoderopeltis* produce proteinaceous secretions on the abdominal tergites that would be readily encountered by predators pursuing these cockroaches. After seizing the cockroaches, predatory centipedes, beetles, and ants rapidly release their prey while cleaning their mouthparts. The fleeing cockroaches generally have more than ample time to effect their escape.

Aphid species in many genera also utilize an entangling secretion as a primary means of defense. In this case the exudate is discharged in response to a confrontation, often hardening to a waxy plaque on an adversary within 30 s. This defensive behavior, which appears to be widespread in the Aphididae, uses tubular secretory organs, the cornicles, on the fifth and sixth abdominal tergites. The secretions, which are dominated by triglycerides, have been characterized in a range of genera, including *Aphis, Myzus, Acyrthosiphon,* and *Therecaphis.* The cornicular secretions are clearly more effective against generalized predators (e.g., ants) than they are against specialized predators (coccinellids, nabids). The secretions also contains alarm pheromones, *E*-β-farnesene

and germacrene A, which release dispersive behavior that may cause aphids to drop off plants.

A variety of glands have been evolved by ants as sources of viscous defensive secretions. Many species in the subfamily Dolichoderinae discharge a pygidial (anal) gland secretion that is dominated by cyclopentanoid monoterpenes such as iridodial, compounds that rapidly polymerize on exposure to air. The viscous polymer effectively entangles small predators such as ants. Myrmicine species in the genus *Pheidole* also use the pygidial glands as a source of an entangling glue and in addition, an alarm pheromone. In contrast, a myrmicine species in the genus *Crematogaster* secretes a potent viscous deterrent from the hypertrophied metapleural glands. On the other hand, minor workers of a *Camponotus* sp. (Formicinae) produce a secretory "glue" in the capacious mandibular glands that extend through the entire body. Mechanical disturbance of the workers results in contraction of the gaster and eventual altruistic rupture, liberating the mandibular gland contents, which are very sticky and readily immobilize attacking ants.

DEFENSIVE FROTHS FROM DIVERSE GLANDS

A surprising diversity of defensive secretions has been converted to froths that may literally bathe small adversaries with compounds that seem to adversely stimulate the olfactory and gustatory receptors of their predators. The independent evolution of deterrent froths by moths, grasshoppers, and ants demonstrates that this form of defensive discharge can be highly efficacious in adverserial contexts.

Species in moth genera in the families Arctiidae (aposematic tiger moths), Hypsidae, and Zygaenidae secrete froths, the production of which is often accompanied by a hissing sound and a pungent odor. The aposematism of these moths is enhanced by secretions discharged from brightly colored areas on or near the prothorax. These secretions do not seem to contain plant natural products but rather, toxic *de novo* synthesized compounds such as pharmacologically active choline esters. Some arctiid froths contain blood, but its importance is unknown.

Frothing is highly adaptive in the ant genus *Crematogaster.* Workers in this very successful myrmicine genus do not possess a hypodermic penetrating sting, but rather, a spatulate sting that is enlarged at the tip. Venom accumulates at the tip and can be smeared onto small adversaries such as ants as if with a paintbrush. This mode of administration of venom is obviously identified with a topical toxicant that can penetrate the insect cuticle much as an insecticide does. There is no indication that tracheal air is added to the venom to generate the discharged froth.

Two grasshopper species produce froths that are derived from a mixture of tracheal air and glandular secretion. Both species are eminently aposematic, and this warning coloration is enhanced by a powerful odor emanating from the froths of the pyrgomorphid *Poekilocerus bufonius* and the acridid *Romalea guttata.*

P. bufonius, a specialized milkweed feeder, is brilliantly colored, exhibiting a dark bluish gray background with contrasting yellow spots and orange hind wings. From a bilobed gland opening between the first two abdominal tergites, disturbed grasshoppers discharge a viscous secretion that is converted to a froth when it mixes with air while passing over the second abdominal spiracle. The froth enhances the aposematism of *P. bufonius* by appearing to be rainbow tinted in contrast to the dark background. This grasshopper is well protected from predation because its exudate contains *de novo* synthesized toxins and sequestered plant natural products that are strongly emetic.

In contrast to *P. bufonius, R. guttata* is a generalist feeder but as with *P. bufonius,* its defensive exudate is discharged as a froth that contains plant natural products as well as compounds synthesized by the grasshopper. If *R. guttata* temporarily specializes on a plant species rich in allelochemicals (*Allium* spp.), its defensive froth can be highly repellent. The aposematism of this acridid rivals that of the pyrgomorphid, and the warning coloration of the former is considerably enhanced by a loud hissing that accompanies the very odoriferous secretory froth.

EXTERNALIZING ALLOMONES BY REFLEX BLEEDING

Many insect species, particularly beetles, externalize their distinctive defensive compounds in a blood carrier rather than discharging them as components in an exocrine secretion.

Cantharidin, the terpenoid anhydride synthesized by adult beetles in the families Meloidae and Oedemeridae, is externalized in blood discharged reflexively from the femorotibial joints. The repellent properties of cantharidin were established more than 100 years ago, and the ability of amphibians to feed on these beetles with impunity has been long known, as well. Cantharidin possesses a wide spectrum of activities, including inducing priapism in the human male, and it has been reported to cause remission of epidermal cancer in mammals. Although its role as a repellent and lesion producer certainly documents its efficacy as a predator deterrent, its potent antifungal activity may be of particularly great adaptiveness in protecting developing meloid embryos from entomopathogenic fungi present in their moist environment.

Autohemorrhage, from the femorotibial joints, is widespread in many species of ladybird beetles (Coccinellidae), most of which are aposematic. The blood is generally fortified with novel alkaloids that are outstanding repellents and emetics (i.e., inducers of vomiting) as well.

Adult fireflies (*Photinus* spp.) produce novel steroids (lucibufagins) that are effective repellents and inducers of emesis in invertebrates and vertebrates. Reflex bleeding from specialized weak spots in the cuticle along the elytra and antennal sockets externalizes these steroids.

Sometimes, rapidly coagulating blood, free of allomones, is used defensively.

BLOOD AS PART OF A GLANDULAR SECRETION

Often the secretions of defensive glands are fortified with blood. The mechanism by which blood penetrates the defensive gland preparatory to being discharged remains to be determined, and indeed, the exact function of the blood itself is not known.

Arctiid moths (e.g., *Arctia caja*) discharge odoriferous froths from prothoracic glands, and these exudates contain pharmacologically active choline esters that are accompanied by blood. A similar system characterizes the pyrgomorphid grasshopper *P. bufonius.* Nymphs of this species possess abdominal defensive glands that sequester two of the six compounds ingested from their milkweed hosts and, in addition, synthesize high concentrations of histamine. These compounds are accompanied by blood as a normal component of the secretion.

NONGLANDULAR DISCHARGES OF PLANT ORIGIN

Certain insects have evolved storage reservoirs for plant natural products that can be discharged in response to traumatic stimuli. This evolutionary development reflects the insect's appropriation of plant allelochemicals (defensive compounds) for subsequent utilization as defensive allomones. In essence, the insects have sequestered the plant's defenses and stored them in reservoirs, where they are available as defensive agents. This defensive system does not require the evolution of any biosynthetic pathways for the storage of compounds in nonglandular reservoirs.

Adults of hemipterous species in the family Lygaeidae possess dorsolateral (reservoirs) and abdominal spaces that contain a fluid very similar to that of the proteins in the blood. This fluid sequesters steroids (cardenolides) present in the milkweeds on which these species feed. The cardenolides are about 100-fold more concentrated in the dorsolateral fluid than they are the blood, and they thus constitute a formidable deterrent system.

Sequestration of plant natural products in nonglandular reservoirs also characterizes larvae of the European sawfly, *Neodiprion sertifer.* Feeding on pine (*Pinus* spp.), these larvae sequester both mono- and sesquiterpenes in capacious diverticular pouches of the foregut. Young larvae, feeding only on pine needles, sequester only three terpenes, whereas older larvae also ingest resin acids. These acids also serve to entangle would-be predators, thus providing a dual protective function.

PLANT NATURAL PRODUCTS IN EXOCRINE SECRETIONS

Herbivorous insects may incorporate plant natural products into exocrine and nonexocrine defensive secretions. By selectively adding proven plant repellent compounds to their own deterrent secretions, insects can increase the effectiveness of their chemical deterrents. These plant-derived compounds are generally unrelated to the constituents in the defensive exudates of their herbivores. In all likelihood, these plant additives may augment the repellency of the deterrents by reacting with olfactory chemoreceptors different from those targeted by the insect-derived repellents.

The large milkweed bug, *Oncopeltus fasciatus,* in common with many species of true bugs, uses the secretion of the metathoracic scent gland as an effective defensive exudate. Nymphs of this species generate defensive secretions with midorsal glandular fluid. The repellent secretions also contain cardenolides derived from the milkweed host plants of this species. These toxic and emetic steroids undoubtedly augment the deterrent effectiveness of the *de novo* synthesized compounds in the glandular exudates.

Similarly, *R. guttata* sequesters in the metathoracic defensive glands plant allelochemicals that can considerably augment the deterrent effectiveness of the secretion. Unlike *O. fasciatus, R. guttata* is a generalist that feeds on and sequesters a potpourri of plant natural products. As a consequence, the compositions of the glandular exudates can be variable, sometimes resulting in secretions that are considerably more repellent than those derived from insects that had fed on a limited number of host plant species.

REGURGITATION AND DEFECATION OF ALLELOCHEMICALS

Enteric defense may be widespread in insects as a means of using the proven repellencies of a variety of plant natural products. In a sense, the intestine is functioning as a defensive organ once repellent plant products have been ingested, and it is likely that the presence of pharmacologically active plant compounds in the intestine renders the insect distasteful or emetic. Therefore, transfer of gut contents to the outside by either regurgitation or defecation could actually constitute the externalization of the internal enteric defenses.

When tactilely stimulated, acridid grasshoppers readily regurgitate, and this discharge, fortified with plant natural products, is very repellent to ants. Similarly, larvae of the moth *Eloria noyesi* regurgitate when molested. The enteric discharge, which contains cocaine extracted from the larval food plant, is very repellent to ants.

Defecation can also serve to externalize deterrent plant natural products. The large milkweed bug defecates readily when subject to traumatic stimuli, the discharge being fortified with emetic and distasteful cardenolides (steroids) derived from the milkweed host plants. The anal discharge, containing concentrated cardenolides, is very repellent to ants.

See Also the Following Articles

Aposematic Coloration • Autohemorrhage • Defensive Behavior • Monarchs • Venom

Further Reading

Blum, M. S. (1981). "Chemical Defenses of Arthropods." Academic Press, New York.

Blum, M. S. (1996). Semiochemical parsimony in the Arthropoda. *Annu. Rev. Entomol.* **41,** 353–374.
Dettner, K. (1987). Chemosystematics and evolution of beetle chemical defenses. *Annu. Rev. Entomol.* **32,** 17–48.
Duffey, S. S. (1977). Arthropod allomones: Chemical effronteries and antagonists. *Proc. XV Int. Cong. Entomol.* Washington, DC, pp. 323–394.
Duffey, S. S. (1980). Sequestration of plant natural products by insects. *Annu. Rev. Entomol.* **25,** 447–477.
Edwards, J. S. (1961). The action and composition of the saliva of an assassin bug *Platymeris rhadamanthus* Gaerst. (Hemiptera: Reduviidae). *J. Expt. Biol.* **38,** 61–77.
Eisner, T. (1970). Chemical defense against predation in arthropods. In "Chemical Ecology" (E. Sondheimer and J. B. Simeone, eds.), pp. 157–217. Academic Press, New York.
Hartmann, T. (1995). Pyrrolizidine alkaloids between plants and insects: A new chapter of an old story. *Chemoecology* **5,** 139–146.
Kellner, R. L. L. (1999). What is the basis of pederin polymorphism in *Paederus riparius* rove beetles? The endosymbiotic hypothesis. *Entomol. Exp. Appl.* **93,** 41–49.
Pasteels, J. M., and Gregoire, J.-C. (1983). The chemical ecology of defense in arthropods. *Annu. Rev. Entomol.* **28,** 263–289.

Chemoreception

B. K. Mitchell
University of Alberta

Insects are acutely aware of many aspects of their environment, as anyone knows who has tried to catch a fly perched on a slice of pizza. In the chemical realm, and depending on the chemicals and insects involved, insects are often outstandingly sensitive. The most famous and best-studied aspects of chemoreception in insects are mate recognition and finding. Like many, if not most animals, insects produce chemicals called pheromones that allow individuals of one sex in a species to recognize and find individuals of the opposite sex. Usually the female produces a mixture of chemicals to which the male responds. Other important, life-or-death decisions largely based on chemicals include choice of site for egg laying, decisions about what to eat and what to avoid, and communications about immediate danger. How insects taste and smell is therefore of great interest and, given that many insects are serious agricultural pests and vectors of disease, research in this area is both fundamental and practical.

As with most physiological systems, model animals are vitally important for scientists who explore the specific workings of what is always a complex series of interactions. For studies of insect chemoreception, adult moths and caterpillars, flies, cockroaches, and leaf beetles have provided some of the best models. Large moths such as the silkworm, *Bombyx mori,* and the tobacco hornworm, *Manduca sexta,* have been essential in studies of pheromones, whereas flies such as the black blowfly, *Phormia regina,* caterpillars such as the cabbage butterfly, *Pieris brassicae,* and *M. sexta,* and beetles such as the Colorado potato beetle, *Leptinotarsa decemlineata,* have helped unravel the role of chemoreception in food and oviposition-related behavior.

For an insect to sense and respond appropriately to the presence of a chemical, or more often a mixture of chemicals, requires a large number of cuticular, cellular, and molecular processes. Because insects are covered in cuticle, it is appropriate to begin there. The cells involved include the sensory cells themselves and closely associated accessory cells whereas the molecules include a wide array of extracellular, intracellular, and membrane-bound proteins. The processes involved in tasting and smelling include sampling the environment, transport of stimulus molecules to receptors, reception, transduction, coding, and transmission to the higher brain centers. This article looks at both gustation (taste) and olfaction (smell), though in many areas detailed knowledge is more complete for smell than it is for taste.

ROLE OF CUTICLE IN TASTE AND SMELL

Insects, like all arthropods, are covered with a chitin–protein complex called cuticle, which in turn is covered with wax to prevent desiccation. For the creature to taste or smell anything, there must be a pathway from the outside to the sensory cells inside. On various parts of the insect body, but particularly on the antennae, mouthparts, legs, and ovipositor (egg-laying structure) insects possess a variety of cuticular elaborations in which are housed chemically sensitive cells. These cuticular structures take the form of hairs (trichoids), pegs, pegs in pits, flat surfaces, and several other shapes. Common to them all is a modified cuticular region that will provide one or more pores through which chemicals can gain entrance. For water conservation, and to keep the important sensory cells functional, these pores cannot allow direct contact of the sensory cell membrane with air. All these pores are small (in the submicrometer range), and there is always a water–protein pathway from the pore to the cell membrane. The cuticular structures plus the associated cells collectively are referred to as sensilla.

Figure 1 represents a reconstruction of a typical mouthpart gustatory sensillum in a caterpillar. All caterpillars so far investigated have this type of sensillum, and it is always important in the food selection processes. The reconstruction is based on careful observations of hundreds of images taken with the electron microscope. The cellular details shown in the drawing cannot be seen with the light microscope. Most of the parts of this drawing below the cuticle could be mistaken for those in an olfactory sensillum. This is because chemosensory cells in insect sensilla are modified cilia and the accessory cells are also basically the same in both types. This involvement of cilia is not surprising, because most of the sensory cells of animals, including light, touch, and hearing, as well as chemical sensors, are modified cilia. Only the sensory cells are modified cilia. The accessory cells are more ordinary, although still specialized, epidermal cells, and they

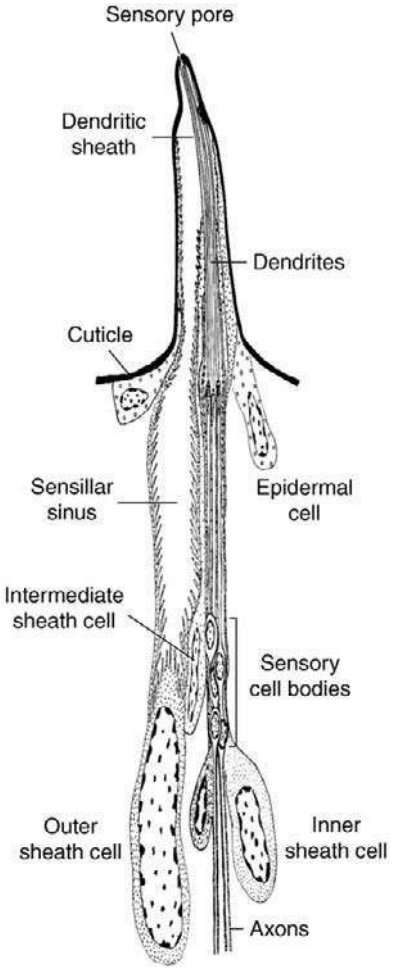

FIGURE 1 Reconstruction of a taste sensillum of the type typically found on the mouthparts of caterpillars. Associated with the maxilla there are four such sensilla, each with four gustatory cells, and it is clear that caterpillars rely heavily on the information provided by the cells to make food choices. The cuticular modification, accessory cells, and sensory cells are all necessary for the sensillum to function properly. In addition to providing the sense of taste, these sensilla are also sensitive to touch. [From Shields, V. D. C. (1994). *Can. J. Zool.* **72,** 2016–2031, as modified by Mitchell, B. K., *et al.* (1999). *Microsc. Res. Technol.* **47,** 401–415.]

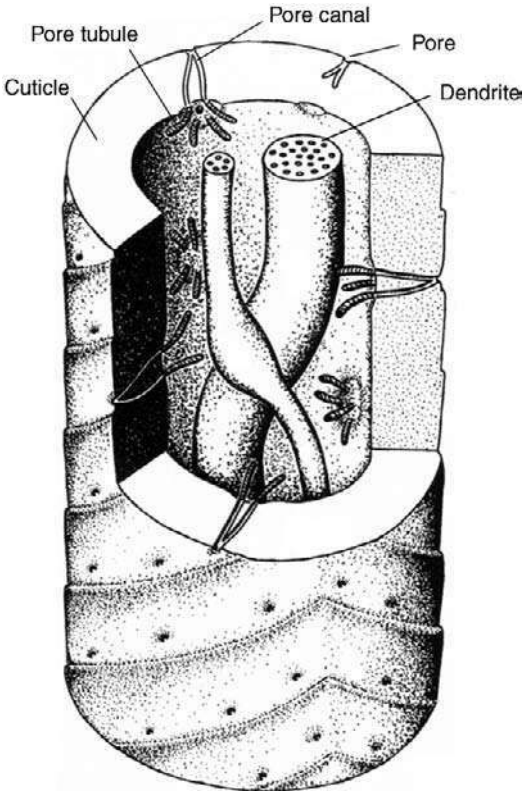

FIGURE 2 Schematic view of a section of a pheromone sensillum in a moth. The features are those revealed in an electron microscopic examination. Olfactory sensilla may have as few as two sensory dendrites, as here, or many more. The arrangement shown is typical of many moth pheromone sensilla. [Relabeled from Keil, T. A. (1999). *In* "Insect Olfaction" (B. S. Hansson, ed.), Fig, 17a, p. 39. © Springer-Verlag GmbH & Co. KG, Berlin.]

have two very different functions. During the development of a sensillum (i.e., between molts) these cells are involved in secreting all the cuticular elements of the sensillum, including the base, the shaft, the cuticular pore or pores, and the dendritic sheath surrounding the dendrites (above the cillary rootlets) of the sensory cells. Once the dendritic sheath is in place, the dendrites are physically separated from the rest of the sensillum lumen, though chemicals can pass through. The dendritic sheath is much longer in taste sensilla, as depicted in Fig. 1, running all the way to the single pore in the tip. The dendritic sheath in olfactory sensilla stops nearer the base of the sensillum, and the dendrites are free in the lumen. In both types, the dendritic sheath provides mechanical stabilization for the sensory cells. When development is complete, the accessory cells provide the particular chemical ionic mix that surrounds the dendrites (note the microvilli in the outer sheath cell). The fluid surrounding the dendrites is very different from the general body fluid (hemolymph), and its high cation concentration is critical in allowing the cell to signal its contact with an appropriate chemical stimulus. This signal is in the form of a potential change across the dendritic cell membrane that is eventually turned into normal action potentials near the sensory cell body.

The structural features discussed so far are shared by olfactory and gustatory sensilla. The major differences between the two types have to do with the way chemicals get into the system and the underlying cuticular modifications. Chemicals enter gustatory sensilla via the single pore in the tip. This pore contains a sugar–protein complex (mucopolysaccharide) that protects the dendrites from desiccation and probably limits the types of chemicals that can pass (though this latter point is in need of further study). Once past this barrier, the chemical enters the solution around the dendrites and potentially can interact with the cell. Olfactory sensilla typically have many pores, and they are different in origin from those in gustatory sensilla. Figure 2 illustrates a section of a typical olfactory sensillum from the pheromone system of a moth. To understand the nature of the numerous pores on this

structure requires knowledge of insect cuticle in general. The surface of insect cuticle is in constant communication with the inside of the animal for the purpose of wax renewal. This communication is provided by numerous pore canals, microscopic and tortuous passages through the cuticle. These canals are filled with a water–wax mix. In olfactory sensilla, the pore canals are taken over for the function of providing access of stimuli to the sensory dendrites. On the inside end of some pore canals are structures called pore tubules; these delicate structures can be seen only in electron micrographs of carefully prepared tissue. It was once thought that pore tubules provided a hydrophobic route for odor molecules to pass from the outside waxy surface of the sensillum to the surface of the dendrite (which is surrounded by water and salts). Discovery of additional molecular components of this system replaced this long-standing and attractive hypothesis.

THE ODOR PATH

Substances animals taste are usually much more water soluble than those that they smell, and the sensory dendrites of both gustatory and olfactory sensilla are in an aqueous medium. Thus, the problem of getting the stimulus to the receptor has received much more attention in olfactory research. In insects, odor molecules first contact the cuticular surface, and because it is waxy, they easily dissolve. From here they move in two dimensions, and some find their way into the opening of a pore canal. Since the pore canal contains wax, passage through it is probably easy, and passage in pore tubules may be similar. Eventually, however, before it arrives at the receptor surface of a dendrite, the hydrophobic odor molecule will encounter water. Recent work, particularly with the antennae of large moths, has uncovered at least two types of protein in the extracellular spaces of sensilla. One type specifically binds chemicals that are part of the moths' pheromone mixture, and are therefore called pheromone binding proteins (PBP). The other type binds less specifically a variety of nonpheromone molecules (e.g., food odors) and are called general odorant binding proteins (GOBP).

The odorant binding proteins (OBP) act as shuttles and carry odor molecules through the aqueous medium to the surface of the dendrite. In the membrane of the sensory cell are receptors for various odors, depending on the specificity of the cell. Cells that respond to only a single pheromone would be expected to have only one type of receptor molecule. More typically, a cell that is sensitive to food odors has a variety of related receptors covering various stimuli. In either case, the odorant binding protein, now carrying the odor molecule, comes in contact with a receptor. What happens next is now under investigation, and there are two competing hypotheses. The OBP may simply deliver the stimulus, which itself then interacts with the receptor protein; or, the stimulus–OBP complex may be the actual stimulus. That is, the receptor site may be configured as to recognize only the combined stimulus and OBP; either alone will not fit. The latter hypothesis may also explain how these systems can turn on and off so quickly: namely, because moths can follow a discontinuous (patchy) odor trail, making minute adjustments in flight pattern on a millisecond scale. This precise behavior is corroborated by electrophysiological measurements showing that the sensory cells can follow an on-and-off pattern of odor stimulation, also in the millisecond range. It is possible that the OBP–stimulus complex, when first formed, is the effective stimulus for the receptor. During the interaction with the receptor, however, the OBP–stimulus complex changes slightly, becomes inactive, and immediately leaves the receptor. Later, it is broken down by other proteins (enzymes) in the sensillum lumen. Figure 3 summarizes this complex series of events and emphasizes the second hypothesis.

CHEMICAL-TO-ELECTRICAL TRANSDUCTION

In almost all studies of animal sensory systems, the stimulus being sensed is in a different energetic form than the chemoelectrical transmission used by the nervous system. Thus, in eyes, light (photon) energy needs to be transduced into chemoelectrical energy via photo pigments. Similarly, with a chemical stimulus–receptor complex, once binding between stimulus and receptor has occurred, the event must be communicated to other parts of the sensory cell to ensure that the end result is a message, composed of action potentials, transmitted to the brain. Understanding of chemical transduction in insects is far enough along to permit the statement that the

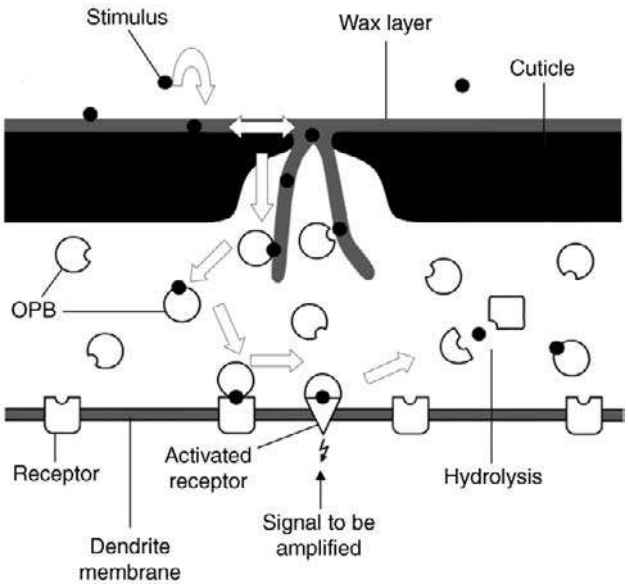

FIGURE 3 Schematic summary of the movement (arrows) of an odor molecule (solid circles) from the surface of a sensillum to the dendritic membrane. Specialized proteins (various shapes) act sequentially as carriers, receptors, and hydrolytic agents to make precise detection of the odorant possible. See text for details. [Relabeled, from Stengl *et al.* (1999). *In* "Insect Olfaction" (B. S. Hansson, ed.), Fig. 1, p. 66. © Springer-Verlag GmbH & Co. KG, Berlin.]

basic elements are probably very much like the arrangement in the vertebrates. There will be differences in detail, but these will continue to be the subjects of active research for some time. Basically, most chemotransduction requires (1) a more or less specific receptor molecule (thus the stimulus–receptor complex can be formed), (2) an amplification step (involving a series of membrane-bound and intracellular molecules) that turns a few stimulus–receptor events into a significant, momentary elevation of some chemical (often calcium) inside the cell, (3) at least one ion channel that senses the rise in calcium and opens, allowing depolarization, and (4) a braking (deactivation) system, composed of more molecular interactions, so the system can be precisely controlled.

Parts of a complete transduction system are beginning to emerge from electrophysiological (patch-clamp) studies of cultured olfactory cells, pharmocological experiments on these cells and on whole-sensillum studies of fly taste sensilla, and from genetic work with *Drosophila* fruit flies. The fruit fly work used specific searches of the now complete *D. melanogaster* genetic database to find some likely candidates for sugar receptor proteins. Carlson has used this information to make specific fluorescent probes, and some of these probes bound only with cells in gustatory sensilla. Combinations of genetic analysis, molecular biology, electrophysiology, and pharmacology will be needed to define all the necessary components.

CHEMOSENSORY CODING AT THE PERIPHERY

In the real world, animals encounter thousands of chemicals. Most of these are meaningless, in the sense that no behavioral response is required, whereas some are critical. A sensory system thus serves two opposing functions. First, the effective sensory system must act as a filter, allowing the animal to ignore most potential stimuli so that it can concentrate on the important ones. Second, the same system must be sensitive, sometimes exquisitely sensitive, to biologically relevant stimuli and must continuously transmit a "summary" report to the brain or central nervous system. The receptor proteins and associated transduction molecules provide the specificity (only some things are adequate stimuli) and sensitivity (the effectiveness of the amplification step in transduction). The nature of the summary neural message is the problem addressed in studies of peripheral sensory coding. In insect chemosensory coding, the problem can be as simple as a few highly specific receptor proteins recognizing a three- or four-component blend of pheromone molecules all housed on a pair of cells found in each of many thousand antennal sensilla. At the other extreme, a leaf beetle may be faced with a food choice of two closely related plants, each with many chemicals to which its tens of gustatory cells are capable of responding. When one is comparing these two scenarios, it is not the number of sensory cells that constitutes the relative scale of the coding problem, but the number of chemical compounds that can be sensed by these cells, and the combinations of compounds that are possible.

In the pheromone example, there are two cell types (each sensillum has one of each type). They respond differently to, for example, four pheromone molecules and not much else. Also, one or two of the pheromone molecules may be completely nonstimulatory to one of the two cells. In addition, only two of the four compounds in the blend may be sufficient to stimulate a full array of behaviors necessary for the male to find the female. The coding problem, though overly simplified to make the point, could thus be reduced to the following: cell A responds only to compound A, and cell B responds only to compound B. Both cells continuously signal to the antennal lobe the levels of compounds A and B detected in the air. If cell A is firing at twice the rate of cell B and both cells are firing at some rate, then the moth flies upwind. Thus the code is a simple comparison, and the large number of cells involved is a kind of amplifier, reflecting the overwhelming importance of the pheromone system to the animal. The two cells, A and B in this example, can be thought of as labeled lines, each sending unique information about the concentration of compound A or B. The central nervous system uses a simple hardwired rule to compare this paired input, and, accordingly, behavior is or is not released.

The beetle, potentially, has a more difficult coding problem. Many experiments have shown that gustatory cells of plant-feeding insects are affected by numerous single plant compounds. Ubiquitous compounds such as water, salts, amino acids, and sugars are sensed by some cells on the mouthparts of all such insects. Less widely distributed chemicals such as alkaloids, terpenes, glucosinolates, and other so-called secondary plant compounds, are stimuli for cells that are variously scattered throughout the class Insecta. To exemplify this coding problem, consider a Colorado potato beetle facing the choice of a potato leaf (host plant) or a tomato leaf (marginal host) (Fig. 4A). The gustatory cells in the beetle's mouthpart sensilla (on the galea), are all sensitive to different compounds. Both direct stimulation by some molecules and inhibition of one molecule by another are known, as are some injury effects in the presence of when too much glycoalkaloid (compounds in potatoes and tomatoes). Not surprisingly, the summary report such a four-cell system sends to the brain comprises two kinds of message, one for potato and one for tomato (Fig. 4B). The complex array of stimuli represented by potato actually stimulate a single cell—the others may well be inhibited. The tomato leaf juice, on the other hand, causes several cells to fire in an inconsistent pattern. The first is another example of a labeled-line type of code; while the second is an across-fiber pattern. In the latter type of code, the brain is receiving information from several physiologically distinct cells, and it is the pattern that is important. It is thought that the across-fiber code pattern prevails in many situations involving complex chemical mixtures. Progress in this area is impeded by the inherent variability of the types of recording possible in the across-fiber pattern (see, e.g., Fig. 4B).

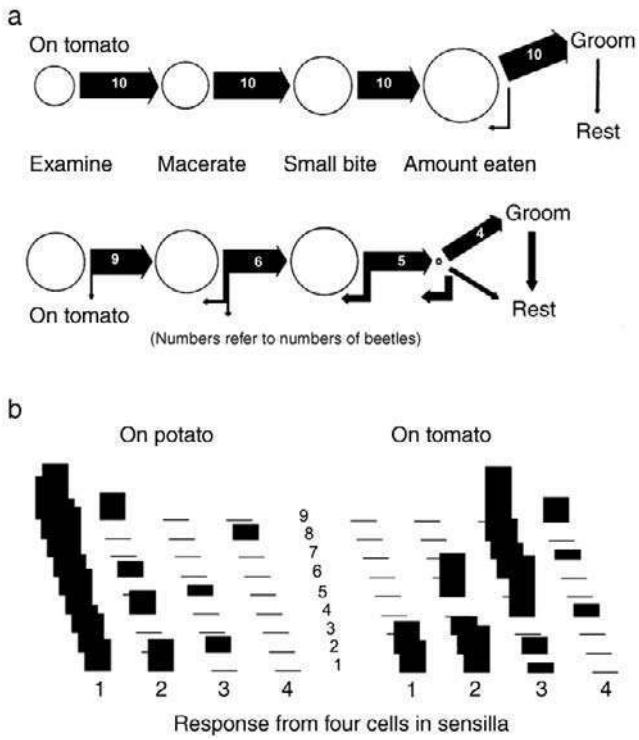

FIGURE 4 (a) Summary of the behaviors exhibited by newly emerged Colorado potato beetle adults when provided with either potato (host plant) or tomato (nonhost plant); numbers of beetles indicated inside heavy arrows. Beetles first examine the leaf, then they squeeze it between their mandibles (macerate) before taking a small bite, which they taste for only a short time. If the plant is acceptable, they very quickly move to sustained feeding. If the plant is less acceptable, few beetles will feed. The decision to not feed is made after considerable time has been spent in examining, macerating, taking small bites, and sometimes repeating one or more of these steps. [Modified from Harrison, G. D., (1987). Host–plant discrimination and evolution of feeding preferences in the Colorado potato beetle, *Leptinotarsa decemlineata. Physiol. Entomol.* **12,** 407–415.] (b) Taste sensilla are important in making the kinds of decisions shown in (a). If potato leaf juice is the stimulus, four cells in nine sensilla on the mouthparts respond by sending a clear, almost labeled-line (cell 1), message to the central nervous system. When tomato leaf juice is the stimulus, a mixed message is provided from the four cells housed in each of the nine sensilla, and this message varies considerably across the available sensilla. The result is a type of across-fiber pattern that signals "do not eat." [Modified from Haley Sperling, J. L., and Mitchell, B. K. (1991). A comparative study of host recognition and the sense of taste in *Leptinotarsa. J. Exp. Biol.* **157,** 439–459. © Company of Biologists LTD.]

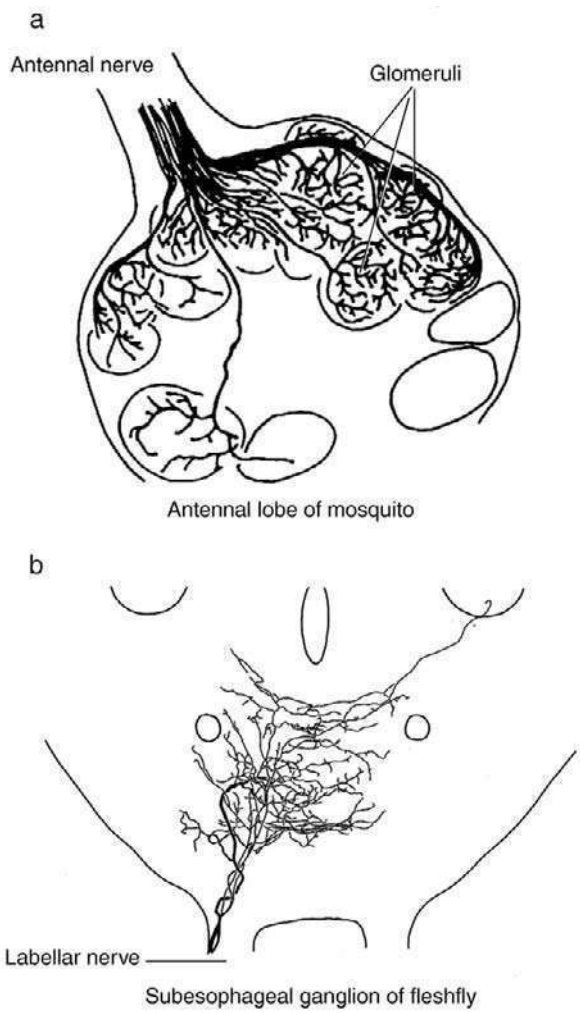

FIGURE 5 In insects, both olfactory and gustatory cells send axons (afferents) directly to the central nervous system. The first synapse (information relay point) is in a particular part of the central nervous system for each sensory modality. (a) Olfactory afferents go to the antennal lobe, where the input is organized in a manner resembling a bunch of grapes—glomerular organization. [From Anton, S., and Homberg, U. (1999). *In* "Insect Olfaction" (B. S. Hansson, ed.), Fig. 6a, p. 110. © Springer-Verlag GmbH & Co. KG, Berlin.] (b) Gustatory afferents from mouthpart sensilla go to the subesophageal ganglion, where they project into a discrete space that is not organized into glomeruli. For both (a) and (b), subsequent processing is done by first-level and higher interneurons. [From Edgecomb, R. S., and Murdock, L. L. (1992). Central projections of axons from the taste hairs on the labellum and tarsi of the blowfly, *Phormia regina* Melgeri. *J. Comp. Neurol.* **315,** 431–444. Reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.]

CENTRAL PROCESSING OF CHEMOSENSORY INPUT

Over the past 20 years, studies of insect olfactory systems have produced a rich literature on the topic of central processing, particularly for pheromonal systems. Work on gustatory systems is far less advanced. The section on insect pheromones provides more information on olfactory processing. This section simply contrasts the gross morphology of the two systems. Both olfactory and gustatory sensory cells are primary neurons; that is, they connect the periphery (sensillum) directly with the central nervous system. Olfactory cells, on the antennae as well as on the palpi, send their axons directly to the antennal lobe, which is a part of the insect brain. Gustatory cells, for the most part, send their axons to the ganglion for the segment in which the sensory cell occurs. Figure 5A shows a typical innervation pattern for antennal and mouthpart olfactory cells in a mosquito, and Fig. 5B shows innervation from the gustatory cells in the mouthparts of a blowfly. A striking difference in the organization of the two parts of the central nervous systems receiving these imputs is repeated across many animal phyla. Olfactory systems are characterized by a glomerular arrangement (like a bunch of grapes) of the neural centers

(neuropile) that receive olfactory afferents (input), but gustatory systems have no such patterned arrangement. The distribution of olfactory inputs into glomeruli suggests a strong association of structure with function, and this is most clearly seen in the macroglomeruli, which receive only pheromonal afferents in male moths. There is undoubtedly an association of structure with function in the way gustatory inputs are arranged, but the lack of a glomerular substructure makes any such system far less obvious. The two ways of organizing chemosensory input, throughout animals, may also point to important differences in coding and or evolution.

See Also the Following Articles

Antennae • Brain • Feeding Behavior • Mechanoreception • Mouthparts • Pheromones

Further Reading

Clyne, P. J., Warr, C. G., and Carlson, J. R. (2000). Candidate taste receptors in *Drosophila. Science* **287,** 1830–1834.

Hansson, B. S. (ed.) (1999). "Insect Olfaction." Springer-Verlag, Berlin, Heidelberg. (See especially Chaps. 1, 2, 3 and 4.)

Hildebrand, J. G., and Shephers, G. M. (1997). Mechanisms of olfactory discrimination: Converging evidence from common principles across phyla. *Annu. Rev. Neurosci.* **20,** 595–631.

Mitchell, B. K. (1994). The chemosensory basis of host–plant recognition in Chrysomelidae. *In* "Novel Aspects of the Biology of the Chrysomelidae" (P. H. Jolivet, M. L. Cox, and E. Petitpierre, eds.), pp. 141–151. Kluwer, Dordrecht, The Netherlands.

Mitchell, B. K., Itagaki, H., and Rivet, M.-P. (1999). Peripheral and central structures involved in insect gustation. *Microsc. Res. Technique* **47,** 401–415.

Pollack, G. S., and Balakrishnan, R. (1997). Taste sensilla of flies: Function, central neuronal projections, and development. *Microsc. Res. Technique* **39,** 532–546.

Schoonhoven, L. M., Jermy, T., and van Loon, J. J. A. (1998). "Insect–Plant Biology." Chapman Hall, London. (See especially Chaps. 5 and 6.)

Steinbrecht, R. A. (1997). Pore structures in insect olfactory sensilla: A review of data and concepts. *Int. J. Insect Morphol. Embryol.* **26,** 229–245.

Zacharuk, R. Y., and Shields, V. D. (1991). Sensilla of immature insects. *Annu. Rev. Entomol.* **36,** 331–354.

Chiggers and Other Disease-Causing Mites

Larry G. Arlian
Wright State University

The Acari (mites and ticks) represent a large array of organisms that exhibit very diverse lifestyles. This article deals with the acarines that are of importance to human health, a group that includes human parasites, natural parasites of other mammals and birds that in particular situations may bite humans, and acarines whose fecal matter, body secretions, and disintegrating bodies are sources of potent allergens.

The parasitic Acari of vertebrates are physiologically dependent on their host and must obtain nourishment from tissue fluids, blood, and cytoplasm from the host to survive, complete the life cycle, and reproduce. Thus, these are obligate parasites. Some species are temporary parasites (e.g., ticks), which visit and feed on the host intermittently. In contrast, other species of parasitic Acari (e.g., scabies and follicle mites) are permanently associated with the host and perish if they become separated from the host. For some species, only one life stage in the life cycle is a parasite (e.g., chiggers), whereas for other species each life stage must feed from a vertebrate host to complete the life cycle (e.g., scabies mites and ticks).

There is usually an intimate interrelationship between acarine parasites and their hosts. Specific host factors, such as carbon dioxide, body odor, and temperature, allow the parasite to locate a host. For example, scabies mites are attracted to the host by body odor and temperature. Permanent parasites may be directed to specific areas of the host body by factors in the skin. The host–parasite interactions for most parasitic acarines have not been well studied and thus are not well understood. This article discusses mites that bite humans, live in the skin of humans, or produce substances that induce immune and/or inflammatory reactions. Because acarine parasites can induce inflammatory and adaptive immune responses, an understanding of the relationship between these two responses is important if one is to understand the symptoms associated with bites from parasitic mites or reactions to body parts, secretions, and fecal matter.

INFLAMMATORY AND IMMUNE RESPONSES

When feeding from the host skin surface, acarine parasites inject or secrete into the host an array of immunogenic and pharmacokinetic molecules. Likewise, acarine parasites that live in the skin, hair follicles, sebaceous glands, and respiratory tree and lungs release immunogenic molecules both while living and after death, from their disintegrating bodies. Substances injected or released may induce an inflammatory (i.e., innate) and/or immune (i.e., adaptive) response by the host. Pharmacokinetic molecules can modulate specific aspects of the host immune or inflammatory responses.

Innate Immune Response

After a person has been bitten by a parasitic acarine, a red (or erythematous) swollen (i.e., edema), and irritated (i.e., painful) lesion may develop at the bite site. These symptoms may be the result of a localized innate inflammatory reaction and not an adaptive immune reaction. In an inflammatory reaction, components of the saliva and body secretions of mites that feed from the skin surface or in tissue (e.g., follicle or scabies mites) cause cells of the skin (epidermis and dermis) such as keratinocytes, fibroblasts, and antigen-presenting cells (Langerhans, macrophages, natural killer cells) to release an array of chemical mediators (cytokines, kinins, and others). These

substances cause arterioles to dilate, which results in increased blood flow to the tissue. Increased blood flow to the skin where a mite has bitten or is located imparts a red appearance. In addition, the tight junctions between endothelial cells of the capillary wall become less tight, which allows fluid from the blood to leak from the capillary lumen into the surrounding tissue, causing it to swell. These cytokines also cause local endothelial cells in the capillaries and white blood cells that pass by in the capillaries to express or increase expression of adhesion molecules (i.e., the receptors) in their surface membranes. White blood cells in the blood vessels stop and adhere to the endothelial cells of the capillary. These cells (cellular infiltrate) then migrate out of the capillary space between endothelial cells to the source of the molecules that induced the reaction. The infiltrating cells may include neutrophils, eosinophils, macrophages, and lymphocytes. The molecules from damaged or stimulated cells and secreted cytokines from the infiltrating cells stimulate pain receptors in the vicinity, causing an irritating sensation. This type of a host response is referred to as innate immunity, and it is not altered with repeated exposure to a particular mite or tick. The time and intensity of the response reaction is the same each time the individual is challenged.

Adaptive Immune Response

In contrast, the molecules introduced into the body by acarine parasites may induce an adaptive immune response that is highly specific for a particular epitope (sequential or structural) on an immunogenic molecule (antigen) from the parasite. An epitope is the part of the antigen that receptors on B and T lymphocytes recognize. The adaptive immune response is stronger and quicker with successive exposures and involves T and B lymphocytes and memory cells of each type. It may be accompanied by an inflammatory reaction too that can be delayed. With the help of type 2 T-helper cells (Th2) B cells become plasma cells that produce antibody directed at the offending molecules from the mite. Activated Th1-type helper cells activate cytotoxic T cells (Tc) that perform functions that kill the parasite directly or damage it. Helper T cells release specific cytokines such as Interleukin 2 (IL-2), interferon γ (IFN-γ), and other interleukins (IL-4, IL-6, IL-10, and IL-13), which act as signals to activate Tc and B cells.

PARASITIC MITES

Family Trombiculidae

Chiggers are the parasitic larval stage of prostigmatid mites that belong to the family Trombiculidae (Fig. 1). Chiggers are also known as harvest bugs in Europe and scrub-itch mites in Asia and Australia. Trombiculid mites are prevalent in moist, warm temperate climates and in tropical climates worldwide. These mites live in moist soil covered with vegetation such as grassy and weedy areas. More than 3000 species of chiggers are known, but only about 15 species frequently bite humans and cause a cutaneous reaction.

FIGURE 1 *Tromicula alfreddugesi,* the mite that causes chiggers. (Photograph © David Scharf.)

Unlike many mites, male and female chiggers do not copulate directly. Instead, males deposit a stalked spermatophore (sperm packet) on the substrate. Females insert it into their genital pores to fertilize the eggs, which are then deposited on moist soils. Larva emerge from the eggs and complete development into an active hexapodal (six-legged) larva (chigger). The larva is parasitic and must feed from a mammal, bird, or reptile host before development can progress to the nymphal stages and the adult. The active nymphal stages and adults are predators and prey on small arthropods (insects and mites) or their eggs. The larval stage (chigger) generally feeds on rodents, mice, birds, and reptiles, and some species bite humans.

Chiggers can cause dermatitis and transmit the agent *Rickettsia tsutsugamushi,* which causes scrub typhus in humans. Scrub typhus is characterized by an ulcer at the site of the bite, high fever, and headache. Scrub typhus is present in tropical climates such as parts of India, Pakistan, Southeast Asia, Philippines, Indonesia, Korea, Japan, China, some Pacific Islands, and coastal Queensland, Australia. The principal vectors are species of the chigger genus *Leptotrombidium.* The reservoir hosts for this disease are rodents (mainly rats). In nature, the pathogen is transferred from rodent to rodent by many chiggers species. Humans become infected when they venture into an enzootic area and are bitten by infected larva. The larval stage feeds only once and acquires the pathogen from infected moles, mice, rats, and other small rodents. Therefore, chiggers can only acquire the rickettsia or, if they were already infected, transmit it, but not both. The rickettsia acquired by the larva is carried (trans-stadially) throughout the developmental stages to the adult. Rickettsia

acquired by the larva multiply in the subsequent developmental life stages and infect the ovaries of the adult, from which they are passed to the egg (transovarially) and then to the larva of the next generation. The rickettsia in transovarially infected larva infect the salivary gland and are transmitted to humans when the larva feed.

The chigger feeds from the surface of the skin much like a tick. Its piercing mouth-parts (chelicerae) are inserted through the epidermis into the dermis. Saliva is introduced into the host during feeding. In humans, these salivary components induce both an innate inflammatory reaction and an adaptive immune response. These reactions are characterized by the production of circulating antibody and by cellular infiltration into the feeding lesion. Repeated exposures result in a more rapid and intense adaptive immune response. It is unclear whether chiggers induce an innate inflammatory response independent of the immune response. Clinically, however, the bite manifests as a reddish (erythematous), swollen (edema), and epidermally-thickened papular and irritating lesion. Histologically, the feeding lesion appears as a cylinder of tightly packed cells surrounding a strawlike channel that extends from the dermis to the skin surface where the chigger is located. The chigger sucks fluids from the surface of the channel until it is engorged, and then it drops off the host. Chiggers do not feed on blood; rather, they feed on extracellular fluid from the dermis.

Family Demodicidae

The prostigmatid mites of the family Demodicidae are small (approx. 100 µm in length) and have an elongated, wormlike body. The podosoma bears retractible, short, stumpy, telescoping legs. The opistosoma is transversly striated and elongate. Two species, *Demodex folliculonum* and *D. brevis,* parasitize humans and are commonly called follicle mites. Both species are most often obtained from the face, particularly along the nose, forehead, scalp, and eyelids. *D. folliculonum* lives in the hair follicle alongside the hair shaft and is positioned with its capitulum (mouthparts) down in the follicle. *D. brevis* resides in the sebaceous gland off the follicle. The entire life cycle is completed in the follicle and sebaceous gland. Generally, these mites cause little pathology in humans who practice good facial hygiene and are not immunocompromised. However, they may be associated with acne, blackheads, and acne rosacea.

Families Laelaptidae, Dermanyssidae, and Macronyssidae

The Mesostigmata contains many species of mites that are parasitic on reptiles, birds, and mammals. Included are hematophagous (blood-feeding) species in the families Laelaptidae, Dermanyssidae, and Macronyssidae. Among these are *Dermanyssus gallinae* (chicken mite), *Ornithonyssus bacoti* (tropical rat mite), *O. bursa* (tropical fowl mite), *O. sylviarum* (northern fowl mite), *Echinolaelaps echidninus* (spiny rat mite), *Liponyssus sanguineus, Haemogamasus pontiger,* and *Eulaelaps stabularis.* These species are attracted to warm objects and usually live on their host or in the nest of their host. Some of these species will attack humans if their normal hosts are not available. This situation may result after roosts and nests of birds (e.g., pigeons, sparrows, starlings) and nests of rodents (mice, rats, squirrels) in homes (attics, behind shutters, etc.) are destroyed. In the absence of a natural host, the mites invade homes and attack humans. Also, species that infest poultry *(O. sylviarium, O. bursa, O. gallinae)* can be a problem for workers who handle infected chickens and turkeys. Bites of these mesostigmatid mites can cause an irritating inflammatory reaction. There may also be an allergic reaction in some individuals, but this remains to be confirmed. *Siponyssoides sanguineus* parasitizes house mice and rats and can transmit *Rickettsia abari,* which causes rickettsial pox in humans. Western equine encephalitis and St. Louis encephalitis viruses have been isolated from *D. gallinae,* but there are no documented cases of transmission of these viruses to humans.

Species in the families Rhinonyssidae, Entonyssidae, and Halarachnidae live in the nasal cavity and lungs of birds and some mammals (e.g., dogs, monkeys, seals, baboons). Human infections by these mites have not been reported.

Family Sarcoptidae

The astigmatid mites, (e.g., *Sarcoptes scabiei*) are permanent obligate parasites that live in the stratum corneum of the skin of at least 17 families of mammals. These mites cause a disease known as scabies. Scabies is a common contagious disease of humans. There is little morphological difference between the strains of *S. scabiei* that parasitize different host mammals, and at this time, the strains from different host species are not considered to be different species by most experts. However, the strains from different host species are host specific and generally cannot permanently infest an unnatural host. For example, the strain from dogs causes only temporary self-limiting infestations in humans, cats, pigs, cattle, goats, and mice, yet scabies naturally occurs on these host species. The host factors and physiological differences between mite strains that do not allow one strain to establish an infestation on strange hosts are not known.

Scabies mites are small. The male and female are 213 to 285 µm and 300 to 504 µm in length, respectively. The life cycle, consisting of egg, larvae, protonymph, tritonymph, and adult males and females, is completed in about 10 to 13 days on the host. All active stages are oval, with a characteristic tortoise like body with stout dorsal setae, cuticular spines, and cuticular striations.

When separated from the host at room temperature, scabies mites must infest a new host within 24 to 36 h to survive. Under cool (4 or 10°C) and humid conditions, females of the strain that infests humans (var. *hominus*) remain infective for at least 4 days. Therefore, fomites (i.e., clothing, bedding, and furniture that harbor dislodged mites) can be important sources of infection for humans. Body odor

and temperature attract these mites to a host. Once on the host skin, females begin to burrow into the skin within minutes, and they can be completely submerged within the stratum corneum within a half-hour. Males, nymphs, and larval stages penetrate more quickly than females.

Scabies is common in nursing homes, day-care centers, and among the general population in the United States. It often mimics other skin diseases and is difficult to diagnose. Scabies is prevalent in some populations in Africa, Central America, South America, Egypt, India, and Australia. Human scabies infestations are manifested in the vicinity of the burrowing mite by itching, red, papular and vesicular lesions. These symptoms generally develop in 6 to 8 weeks after a primary (first) infestation, but they are evident within a few days of a subsequent infestation. Lesions most commonly occur on the interdigital, elbow, and chest (breast area) skin. However, other areas that may be infested are the penis, buttocks, knees, soles and insteps of the feet, wrists, waistline, and axillae.

Scabies mites induce both cell-mediated (Th1) and circulating antibody (Th2) immune responses and an associated inflammatory reaction. The cell-mediated/inflammatory response is characterized by a mixed cellular infiltrate in the skin lesion that consists of plasma cells, lymphocytes, mast cells, neutrophils, Langerhans cells, and eosinophils.

An infestation with scabies induces some immune resistance to subsequent infestations. The balance between the Th1 and Th2 responses appears to be a key aspect in protective immunity. Hosts that develop protective immunity exhibit up-regulated Th1 and weaker Th2 responses. In contrast, hosts that do not develop protective immunity exhibit strongly up-regulated Th2 response (circulating antibody) but a weaker Th1-cell-mediated response. Infected hosts produce serum antibodies to at least 12 antigens from sarcoptic mites. Some of these antigens are cross-reactive with antigens from the related house dust mites *Dermatophagoides farinae, D. pteronyssinus,* and *Euroglyphus maynei.* In some humans, antigens from *S. scabiei* can also induce an IgE-mediated allergic reaction and circulating IgE-type antibody.

Family Pyemotidae

Pyemotid mites are prostigmatids that have an elongate cigar-shaped body with the first two pair of legs widely spaced from the posterior two pair of legs. They have stylettiform (needlelike) chelicerae and are usually parasitic on the larvae of insects. Unlike other mites, pyemotid female mites retain internally the eggs from which the immatures hatch and pass through all developmental stages. As a result, the female's opisthosoma (region behind the last pair of legs) becomes enormously swollen before the offspring are born.

Pyemotes tritici (straw itch mite) and *P. ventricosus* (grain itch mite) are parasitic on the larvae of grain moths, boring and stored grain beetle larvae, and other insects. Humans may contact these species when working with grain and hay. Also, hordes of these insects may emerge from the flowers of cattails brought into a home to make a floral arrangement. These mites will attack humans and cause red, itchy inflammatory dermatitis.

Families Tetranychidae and Eriophyidae

Many species of prostigmatid mites such as those in the families Eriophyidae and Tetranychidae parasitize plants and can become an economic problem on food crops (e.g., fruit trees; vegetable and grain crops) and yard/garden and green houseplants. Humans come into contact with these mites when working in fields, orchards, greenhouses, gardens, and yards, when handling infested food crops/produce, or by living near an area in which food crops are grown. The importance to human health of most of these pest species has yet to be determined. However, it is clearly documented that a few species are the source of allergens that induce allergic reactions in predisposed individuals. Farmers working in apple orchards and children living around citrus orchards have become sensitized and/or had allergic reactions to *Tetranychus urticae* (two-spotted spider mite) and *Panonychus ulmi* (European red mite) and *P. citrilis* (citrus red mite).

Family Phytoseiidae

Humans come into contact with predaceous mites that are used for biological control of pest species such as the tetranychids just mentioned. The predaceous mite *Phytoseilus persimilis,* which feeds on spider mites, can cause allergic reactions.

Family Hemisarcoptidae

Hemisarcoptes cooremani is an astigmatid mite that is a predator of scale insects that parasitize woody plants. The body of this mite is the source of at least two allergenic proteins. Close contact with these mites can result in production of serum IgE and allergic symptoms. Therefore, gardeners and nursery workers may become sensitized to this mite and have allergic reactions.

NONPARASITIC MITES

Family Pyroglyphidae

The family Pyroglyphidae contains mainly species of astigmatid mites that live in the nests of birds and mammals, where they feed on the epidermal detritus (skin, feathers) left by the host. Three species, *Dermatophagoides farinae, D. pteronyssinus,* and *Euroglyphus maynei,* are commonly found in homes of humans. In homes, these mites are most prevalent in high-use areas, where shed skin scales collect and serve as their food. Therefore, the greatest densities are found in carpets around sofas and easy chairs, in fabric-covered overstuffed furniture, and in mattresses. However, they may also be found in bedding, on pillows, on clothing, on

automobile and train seats, and sometimes in schools and in the workplace. Each species is the source of multiple potent allergens that sensitize and trigger allergic reactions in predisposed people. These allergens cause perennial rhinitis, asthma, and atopic dermatitis.

Ambient relative humidity is the key factor that determines the prevalence and geographical distribution of these mites. This is because water vapor in humid air is the main source of water for their survival. They survive and thrive well at relative humidities above 50% but desiccate and die at relative humidities below this. Therefore, dust mites and the allergies they cause are a significant problem only for people who live in humid, tropical, and temperate geographical areas. *D. farinae* and/or *D. pteronyssinus* are prevalent in homes in the United States, Europe, South America, and Asia. Most homes are coinhabited by multiple species. However, the most prevalent species varies both between homes in a geographical area and between geographical areas. For example, in the United States, both *D. farinae* and *D. pteronyssinus* are prevalent in homes. However, in South America, *D. pteronyssinus* is prevalent in homes, whereas *D. farinae* is not.

In temperate climates, population densities of *D. farinae* and *D. pteronyssinus* exhibit pronounced seasonal fluctuations that parallel the seasonal fluctuations in indoor relative humidity. High densities occur during the humid summer and low densities during winter.

The life stages of the dust mites are egg, larva, protonymph, tritonymph, and adult male and female. Length of the life cycle is temperature dependent when relative humidity is above 60%. At 23°C the life cycle takes 34 and 36 days to complete for *D. farinae* and *D. pteronyssinus,* respectively. Females produce 2 or 3 eggs daily during the reproductive period at 23°C. *D. pteronyssinus* takes 23 and 15 days to complete development at 16 and 35°C, respectively. *D. farinae* does not develop well at 16 and 35°C.

A desiccant-resistant quiescent protonymphal stage can develop that allows survival during long periods (months) under dry (low relative humidity) conditions. When relative humidity conditions become optimal, the quiescence is broken and development continues.

Allergens from these mites are associated with fecal material, body secretions, and body anatomy. Fourteen different groups of mite allergens have been characterized. The frequency of reactivity to most of these allergens is above 40% among patients sensitive to dust mites. Sensitivity to allergens varies both within and between individuals. Allergens from one species may be species specific, or they may cross-react with allergens from another mite species. Most patients with sensitivities are allergic to multiple allergens of a species and to multiple mite species.

Families Acaridae, Glycyphagidae, Carpoglyphidae, Echimyopididae, and Chortoglyphidae

Many species of the astigmatid families Acaridae, Glycyphagidae, Carpoglyphidae, Echimyopididae, and Chortoglyphidae are medically important because they are the sources of potent allergens. Many species of these mites are often referred to as "storage mites" because they occur in stored hay, grain, and straw, in processed foods made from grain (flour, baking mixes), and in dust in grain and hay at storage, transfer, and livestock feeding facilities. Humans may be exposed to storage mites, and their allergens, occupationally and in the home. Inhalation or contact on the skin with allergens from storage mites can induce allergic reactions. These mites and their allergens can also occur in bread, pancakes, cakes, pizza, pasta, and bread made from ingredients contaminated with mites. Humans have had anaphylactic reactions after eating these mite-contaminated foods.

Species known to be the sources of allergens include *Blomia tropicalis* (Echimyopididae); *Acarus siro, Tyrophagus putrescentiae, T. longior,* and *Aleuroglyphus ovatus* (Acaridae); *Lepidoglyphus destructor* and *Glycyphagus domesticus* (Glycyphagidae); *Carpoglyphus* spp. (Carpoglyphidae); *Chortoglyphus arcuatus* (Chortoglyphidae); and *Suidasia medanensis* (Suidasiidae). *T. putrescentiae* is the source of 14 allergens, with the number recognized as allergens by individuals ranging from 5 to 11. *B. tropicalis,* which is common in house dust in tropical climates and may be more prevalent than pyroglyphid mites, has been reported in small numbers in some homes in the southern subtropical United States. Several allergens from *B. tropicalis* have been characterized and/or produced by recombinant technology. There is little cross-reactivity between storage mites and house dust mites. However, many patients are sensitive to both storage mites and the pyroglyphid house dust mites.

See Also the Following Articles

Medical Entomology • Mites • Ticks • Veterinary Entomology

Further Reading

Arlian, L. G. (1989). Biology, host relations, and epidemiology of *Sarcoptes scabiei. Annu. Rev. Entomol.* **34,** 139–161.

Arlian, L. G. (1992). Water balance and humidity requirements of house dust mites. *Exp. Appl. Acarol.* **6,** 15–35.

Arlian, L. G. (1996). Immunology of scabies. *In* "The Immunology of Host–Ectoparasitic Arthropod Relationships." (S. Wikel, ed.), pp. 232–258. CAB International, Wallingford, U.K.

Arlian, L. G. (2002). Arthropod allergens and human health. *Annu. Rev. Entomol.* **47,** 395–433.

Arlian, L. G, Bernstein, D., Bernstein, I. L., *et al.* (1992). Prevalence of dust mites in homes of people with asthma living in eight different geographic areas of the United States. *J. Allergy Clin. Immunol.* **90,** 292–300.

Arlian, L. G., Neal, J. S., Morgan, M. S., *et al.* (2001). Reducing relative humidity is a practical way to control dust mites and their allergens in homes in temperate climates. *J. Allergy Clin. Immunol.* **107,** 99–104.

Johansson, E., Johansson, S. G. O., and van Hage-Hamsten, M. (1994). Allergenic characterization of *Acarus siro* and *Tyrophagus putrescentiae* and their cross reactivity with *Lepidoglyphus destructor* and *Dermatophagoides pteronyssinus. Clin. Exp. Allergy* **24,** 743–751.

Kim, Y. K., Lee, M. H., Jee, Y. K., *et al.* (1999). Spider mite allergy in apple-cultivating farmers: European red mite (*Panonychus ulmi*) and the two-spotted spider mite (*Tetranychus urticae*) may be important allergens in the development of work-related asthma and rhinitis symptoms. *J. Allergy Clin. Immunol.* **104,** 1285–1292.

Lee, M. H., Cho, S. H., Park, H. S., *et al.* (2000). Citrus red mite (*Panonychus citrilis*) is a common sensitizing allergen among children living around citrus orchards. *Ann. Allergy Asthma Immunol.* **85,** 200–204.

Radovsky, F. J. (1994). The evolution of parasitism and the distribution of some dermanyssoid mites (Mesostigmata) on vertebrate hosts. *In* "Mites, Ecological and Evolutionary Analysis of Life-History Patterns." (M. A. Houck, ed.), pp. 186–217. Chapman & Hall, New York.

Thomas, W. R., and Smith, W. (1999). Towards defining the full spectrum of house dust mite allergens. *Clin. Exp. Allergy* **29,** 1583–1587.

Wikel, S. K., Ramachandra, R. N., and Bergman, D. K. (1996). Arthropod modulation of host immune responses. *In* "The Immunology of Host-Ectoparasitic Arthropod Relationships." (S. Wikel, ed.), pp. 107–130. CAB International, Wallingford, U.K.

Wrenn, W. (1996). Immune responses to manger mites and chiggers. *In* "The Immunology of Host-Ectoparasitic Arthropod Relationships." (S. Wikel, ed.), pp. 259–289. CAB International, Wallingford, U.K.

Chitin

Ephraim Cohen
The Hebrew University of Jerusalem

Chitin is a globally abundant biopolymer, second only to cellulose and possibly lignin in terms of biomass. Owing to extensive hydrolytic activity mainly by soil and marine chitinolytic microorganisms, chitin is not accumulated in the biosphere, because it is similar to cellulose and unlike lignin. Chitin, which is absent from plants and vertebrates, is present to a small or large extent in most invertebrates, notably in cuticles of arthropods, in primary septum and scar buds of yeast, and in cell walls of most filamentous fungi. Chemically detectable chitin has been verified in 25-million-year-old insect fossils. Chitin is almost invariably covalently or noncovalently associated with other structural molecules in contact with the external environment; examples include carbohydrate polymers in fungi and the cuticular proteins that comprise up to 50% by weight of arthropod cuticles. The chitoprotein supramolecular matrix occurs in peritrophic membranes of insects and in the arthropod exoskeleton, where the rigid chitin microfibrils contribute greatly to its mechanical strength.

STRUCTURE

Chitin is a large water-insoluble, linear aminocarbohydrate homopolymer composed of β_{1-4}-linked *N*-acetyl-D-glucosamine units with a three-dimentional α-helix configuration (Fig. 1). Intramolecular hydrogen bondings stabilize the α-helical configuration of the macromolecule. In nature, chitin polymers coalesce extracellularly by intermolecular hydrogen bonds to form crystalline microfibrils that may appear in various polymorphs (α, β, and γ). The most abundant one in insects is the antiparallel arrangement of the α-chitin polymorph.

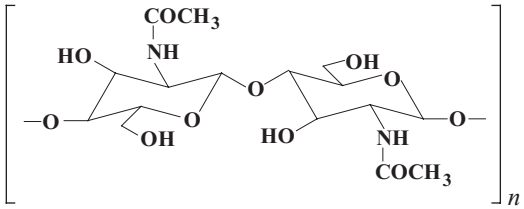

FIGURE 1 The dimer unit of chitin polymer.

CHITIN SYNTHESIS

Chitin synthesis occurs throughout the insect's life cycle and is under hormonal control of ecdysteroids. Bursts of synthetic activity that are associated with the buildup of the new cuticles occur in particular at the last phase of embryonic development, and as larvae or pupae molt. Chitin synthesis is the end result of a cascade of interconnected biochemical and biophysical events that link the mobilization of substrate molecules, polymerization by the enzyme chitin synthase, and translocation of the nascent amino polymer across the plasma membrane (Fig. 2). Individual chitin chains coalesce outside the plasma membrane, forming fibril crystallites by intramolecular hydrogen bonds. The UDP-*N*-acetyl-D-glucosamine substrate is the end point of a series of biochemical transformations that include successive steps of phosphorylation, amination, and acetylation of starting precursors such as trehalose or glucose. Chitin synthase is a relatively large membrane-bound enzyme with multiple transmembrane segments. The active site of the enzyme faces the cytoplasm, and the catalysis involves linking together dimer amino sugar substrates. The question of how chitin polymers are translocated across the cell membrane remains unresolved. Hydrophobic transmembrane segments of chitin synthase are implicated in this process.

The complete chitin synthase cDNA and deduced amino acid sequences of the insects *Drosophila melanogaster* and *Lucila cuprina,* and the nematode *Caenorhabditis elegans* have

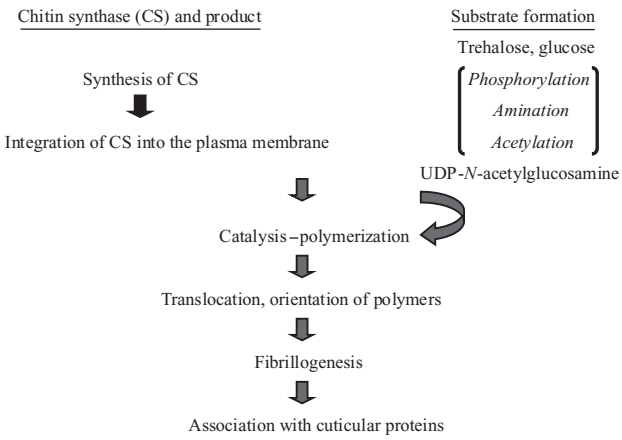

FIGURE 2 Polymer formation and deposition.

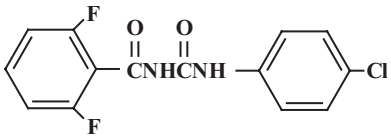

FIGURE 3 Chemical structure of the acylurea compound diflubenzuron (Dimilin).

been described. The deduced amino acid sequence revealed a protein (about 180 kDa) with a large number (15–18) of potential transmembrane segments that may be involved in the translocation of chitin polymers.

CHITIN DEGRADATION

Degradation of chitin is physiologically crucial for normal growth and development of insects. Chitin is degraded by the joint action of chitinase, which yields oligomeric fragments, and exochitinase, or β-*N*-acetylglucosaminidase, which hydrolyzes terminal polymers or dimers. These hydrolytic enzymes are widespread in plants, vertebrates, invertebrates, and microorganisms. During the complex molting process in arthropods, the chitin in the cuticular region (the endocuticle), which is close to the epidermal cells, is degraded. Since chitin microfibrils are tightly associated with various cuticular proteins, proteolytic activity accompanies and facilitates chitin hydrolysis. Hydrolysis of chitin does not occur in the exocuticle, where sclerotization of the cuticular protein takes place. Formation and secretion of chitinases by epidermal cells, processes that are under hormonal control, are vital for the molting process. The mono- and disaccharide degradation products are absorbed by the epithelial cells and may be recycled to serve for biosynthesis of the new chitin.

INHIBITION OF CHITIN SYNTHESIS AND DEGRADATION

Because chitin is present in invertebrates (abundantly in arthropods) and absent from vertebrates and plants, it is a logical target for selective pest control. Acylurea compounds, discovered serendipitously by Dutch scientists in 1972, inhibit chitin synthesis, resulting in deformed and weak cuticles that cause molting failure and death by desiccation. Acylureas do not inhibit the catalytic step of polymerization, and their exact biochemical lesion is unresolved. It appears that the mode of action is associated with the process of chitin translocation from site of catalysis across cell membranes to the region of deposition and fibrillogenesis. The first commercial product reaching the market was diflubenzuron (Dimilin) (Fig. 3), which was followed by a large number of structurally similar bioactive molecules. The acylurea compounds, which act as insect growth regulators, are widely used in integrated pest management (IPM) programs.

See Also the Following Articles

Cuticle • Exoskeleton • Integument • Molting

Further Reading

Cohen, E. (1987). Chitin biochemistry: Synthesis and inhibition. *Annu. Rev. Entomol.* **32,** 71–93.

Cohen, E. (1993). Chitin synthesis and degradation as targets for pesticide action. *Arch. Insect Biochem. Physiol.* **22,** 245–261.

Muzzarelli, R. A. A. (1977). "Chitin." Pergamon Press, Oxford, U.K.

Stankiewicz, B. A., Briggs, D. E. G., Evershed, R. P., Flanerry, M. B., and Wuttke, M. (1997). Preservation of chitin in 25-million-year-old fossils. *Science* **276,** 1541–1543.

Tellam, R. L., Vuocolo, T., Johnson, S. E., Jarmey, J., and Pearson, R. D. (2000). Characterization of insect chitin synthase: cDNA sequence, gene organization and expression. *Eur. J. Biochem.* **267,** 6025–6042.

Verloop, A., and Ferrell, C. D. (1977). Benzoylphenyl ureas—A new group of larvicides interfering with chitin deposition. *In* "Pesticide Chemistry in the 20th Century" (J. R. Plimmer, ed.), pp. 237–270. ACS Symposium Series 37, American Chemical Society, Washington DC.

Chromosomes

Graham C. Webb
The University of Adelaide, Australia

Chromosomes in insects display almost the whole range of variation seen in the chromosomes of higher plants and animals. In these groups the deoxyribonucleic acid (DNA), which contains the genetic code determining development and inheritance, is contained in a nucleus in each cell. At interphase, the DNA is organized into the complex linear structures that are chromosomes, which can be seen in a conveniently condensed state when the cell is dividing.

The study of insect chromosomes is less intensive now than formerly for three possible reasons: (1) the thoroughness of the early investigators, (2) the commercialization of science, which has pushed the study of chromosomes (cytogenetics) in animals toward more lucrative mammalian, and particularly human, fields, and (3) the replacement of cytogenetic with molecular methods. The third point was predicted by Michael White in the conclusion to his famous 1973 textbook, *Animal Cytology and Evolution.* Through his work, almost entirely on insects, White is widely regarded as the founder of the study of evolutionary cytogenetics in animals and one of its foremost authorities; his book remains a most comprehensive authority on most aspects of insect chromosomes.

In 1978 White was firmly convinced that evolution is essentially a cytogenetic process, and he did much to demonstrate this at the level of speciation. At a higher evolutionary level, the integrated chromosomal characteristics of the various insect orders seem to support this view. However more recently authors such as King have de-emphasized the importance of chromosomal changes in species evolution.

SOURCES AND PREPARATION OF CHROMOSOMES FROM INSECTS

Mitotic chromosomes undergoing the familiar stages of prophase, metaphase, anaphase, and telophase can be prepared from any insect somatic tissues with dividing cells. Embryos are the best sources of mitotic divisions, but they are also seen in the midgut ceca of adults and juveniles and in the follicle cells covering very early ova in females.

Insect cytogeneticists now usually use colchicine or other mitostatic agents to arrest the chromosomes at metaphase of mitosis by inhibiting the formation of the spindle fibers required for the cells to progress to anaphase. Squashing, under a coverslip, spreads the chromosomes, and for squash preparations the cells are usually prestained. Insect cytogeneticists now often use air-drying to spread the chromosomes, since this process has the advantage of making the chromosomes immediately available for modern banding and molecular cytogenetic methods.

Male meiosis is very commonly used to analyze the chromosomes of insects and to analyze sex-determining mechanisms. The structure of the insect testis is very favorable to chromosomal studies because each lobe has a single apical cell that divides by a number *(s)* of spermatogonial divisions (Fig. 1A) to yield 2^s primary spermatocytes, which then undergo synchronous first and second meiotic divisions to yield $2^s + 1$ secondary spermatocytes and $2^s + 2$ sperm.

First meiotic prophase in insects involves the usual stages (Fig. 1). Replication of the DNA is followed by the prophase stages of leptotene (strand forming), zygotene (chromosome pairing to form bivalents), pachytene (crossing over to yield recombinants), diplotene (repulsion of the homologues), diakinesis (completion of repulsion), and premetaphase (bivalents fully condensed).

Metaphase I is followed by first anaphase, which can be a very informative stage and, in contrast to mammals, is readily available in insects. Second meiotic division is also readily observed in insects (Fig. 1) and can be useful for confirming events in earlier stages.

Meiotic chromosomes in insect females are difficult to prepare and are usually studied only in special cases, such as parthenogenesis.

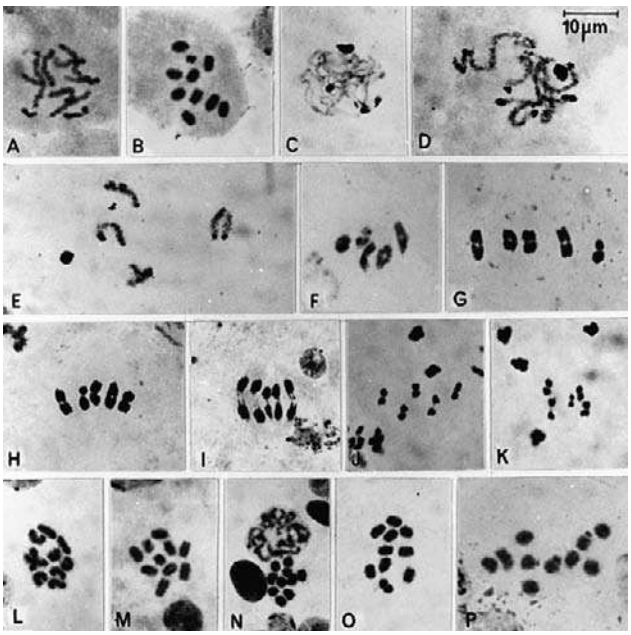

FIGURE 1 Mitotic and meiotic holocentric chromosomes in an earwig, *Labidura truncata.* Orcein-stained squash preparations, B, L, M-P colchicine-treated. (A) Spermatogonial division in prophase with the Y at bottom left and the X to the right, both more condensed than the autosomes. (B) Spermatogonial metaphase with the small Y chromosome obvious. (C) Leptotene, with the sex chromosomes at the top very condensed and the heterochromatic ends of some autosomes also condensed. Two nucleoli are visible, one at 11 o'clock and the other at 5 o'clock. (D) Zygotene–pachytene with the heterochromatic ends of the autosomes more obvious. (E) Diplotene displaying the four autosomal bivalents and the condensed sex chromosomes separately. (F) Diakinesis, one autosomal bivalent showing a chiasmata that is quite interstitial. (G, H) First metaphases with the larger X seem to be paired with the smaller Y. First anaphase with the neocentromere actively moving the chromosomes apart. (J, K) Second metaphases; J shows the X dyad, K shows the smaller Y dyad. (L–P) Female mitotic chromosomes, late and early prophase in L and N, respectively; M–P show metaphases, with O and P showing secondary constrictions. The primary constrictions of fixed centromeres do not show, and uninterrupted chromatids, characteristic of holocentric chromosomes, are particularly obvious in M. [From Giles, E. T., and Webb, G. C. (1973). The systematics and karyotype of *Labidma Truncata* Kirby, 1903 (Dermoptera: Labiduridae). *J. Aust. Entomol. Soc.* **11,** Plate 1, with permission.]

TYPES OF CHROMOSOME IN INSECTS

Autosomal chromosomes are usually represented as two haploid sets, one from each parent, in the diploid tissues of insects. With the addition of the sex chromosomes from each parent, the haploid set is known as *"n"* and the diploid set as "2*n*." Major exceptions to diploidy in both sexes occur, such as in almost all species in the the orders Hymenoptera (ants, bees, and wasps), Thysanoptera (thrips), and some species of Heteroptera and Coleoptera, where the females are diploid, the males being normally haploid (i.e., derived from unfertilized eggs, arrhenotoky). Arrhenotoky determines the sex of about 20% of all animal species. This mechanism has allowed one species of Australian ant, *Myrmecia croslandi,* to achieve the lowest possible chromosome number, $n = 1$, in the parthenogenically derived male.

Chromosomal imprinting has not been demonstrated in insects, so gametes from both sexes are not necessarily required. Indeed, accidental development of unfertilized eggs (thelytoky) can form a parthenogenetic insect if sufficient double-haploid cells arise in critical tissues in the $n/2n$ mosaic.

Sex chromosomes are usually involved in sex determination in insects, but by a variety of genetic mechanisms. The male is usually the heterogametic sex in insects, the exceptions being the orders Lepidoptera (butterflies and moths) and Trichoptera, in which the females are heterogametic. The mammalian system of having genes determining the sex and

other male functions on the Y chromosome almost certainly does not occur in insects. In the earwigs (Dermaptera), male determination by the presence of a Y chromosome seemed to be the rule, until XO/XX mechanisms were found in two species.

As in other animals, the insect heterogametic male has half the number of X chromosomes as the female; most commonly the sexes are XO male and XX female, but multiple X-chromosome systems frequently occur. Fusions of autosomes to the X chromosome can cause the formation of XY/XX and further fusions to form $X_1X_2Y/X_1X_1X_2X_2$ systems. X_1X_2Y males are almost the rule in the mantids (Mantodea). In the most of the Hymenoptera, sex is determined by the diploid females being heterozygous, and the haploid males hemizygous, for multiple alleles at a single genetic locus on one chromosome; that might still be regarded as an X chromosome.

The karyotype is the set of chromosomes, both autosomes and sex chromosomes, in an organism. The karyotype found in 90% of the large family of short-horned grasshoppers, Acrididae, is usually given as $2n♂ = 23$ [22 + X (or XO)], the female karyotype, 23 (22 + XX), being usually inferred from the male. Some authors carefully confirm diploidy and subdifferentiate the autosomes and the sex chromosomes, for example, for the earwigs *Chaetospania brunneri* $2n$ ♂ = 31 (13AA + $X_1X_2X_3X_4Y$) and *Nala lividipes,* with one pair of autosomes being exceptionally long, $2n♂ = 34$ (A^LA^L + 15AA + XY). Since in insects, the karyotypic nomenclature is variable and somewhat confusing, it would seem preferable to adopt the simple karyotypic nomenclature used for mammals [e.g., human: $2n♂ = 46$,XY].

Monocentric chromosomes are the norm in most insect orders (Fig. 2), with the single centromere characterized by a primary constriction, a structure seen in many other animals and in plants. After replication of the DNA and other chromosomal constituents during interphase, the chromosomes at metaphase show two identical chromatids. Following the discovery that each chromatid must be terminated by a telomere, geneticists concluded that it is highly probable that a chromosome must always have two arms, one on each side of the centromere. If these arms are of appreciable length, the chromosomes are called metacentric (arms of about equal length), or submetacentric (arms of unequal length). If one of the arms is very short, perhaps invisible under normal microscopy, the chromosome is said to be acrocentric. It is now widely accepted that a chromosome cannot normally be telocentric (terminated by a centromere). The sequence of nucleotides repeated many times to make up the DNA of the telomeres of most insects is TTAGG, but it is not universal.

Holocentric chromosomes occur in the insect orders Heteroptera, Dermaptera (Fig. 1), Mallophaga, Anoplura, and Lepidoptera. The centromeres are elongated across much of the length of the chromosomes, although not usually extending to the telomeres. During mitotic anaphase, the spindle fibers pull equally on most of the length of the chromosome so that only the distal ends can be seen to be

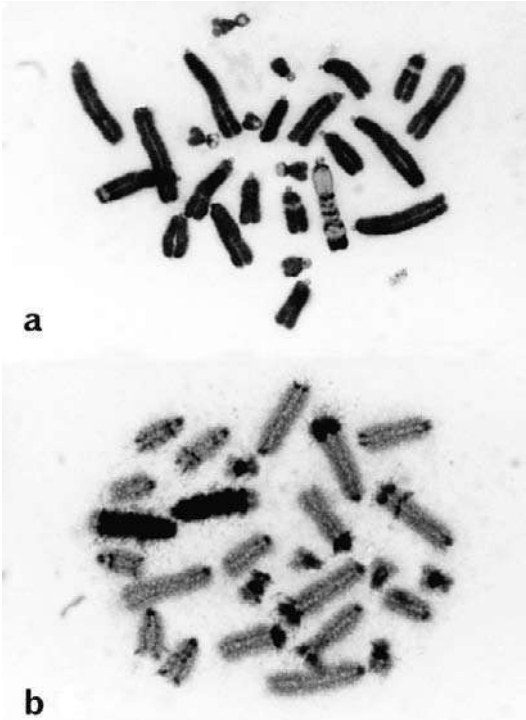

FIGURE 2 Monocentric chromosomes of the locust *Chortoicetes terminifera,* mostly acrocentric with some of the smaller ones submetacentric. (a) With one B chromosome, which is distinctively G-banded by a trypsin treatment that has produced comparatively minor effects in the A chromosomes. (b) with two B chromosomes showing positive C-banding for most of their length. The A chromosomes mostly have small centromeric C bands, but they show variable interstitial and distal C-banded segments.

trailing. Operation of an elongate centromere during first meiosis would break the crossovers, or chiasmata, which have formed between the the paired chromosomes. Apparently to preserve this chromosomal bivalent, the holocentric chromosomes develop neocentric activity at one telomeric end only. The neocentromeres behave like those of monocentric chromosomes, and they persist through second metaphase of meiosis (Fig. 1). Broken holocentric chromosomes seem to be able to retain attachment to the spindle fibers: in earwigs, each piece of a broken chromosome forms a bivalent with a neocentromere during meiosis. Breakage of holocentric chromosomes probably also explains the wide range of chromosome numbers seen in butterflies, from $2n = 14$ to $2n = 446$, and the extreme of $2n = 4$ to $2n = 192$ found by Cook in a single genus of scale insect, *Apiomorpha.* Breakage probably also accounts for the common finding of multiple X chromosomes in insects with holocentric chromosomes. The addition of telomeres to the broken ends of holocentric chromosomes might be a function of the complex enzyme telomerase.

Polytene chromosomes are large chromosomes formed by the repeated replication, without intervening division, of chromatids that remain uncondensed as in interphase (Fig. 3). Polytene chromosomes often contain thousands of chromatid strands, and the homologous chromosomes are usually closely somatically paired, so that inversions in them are accommo-

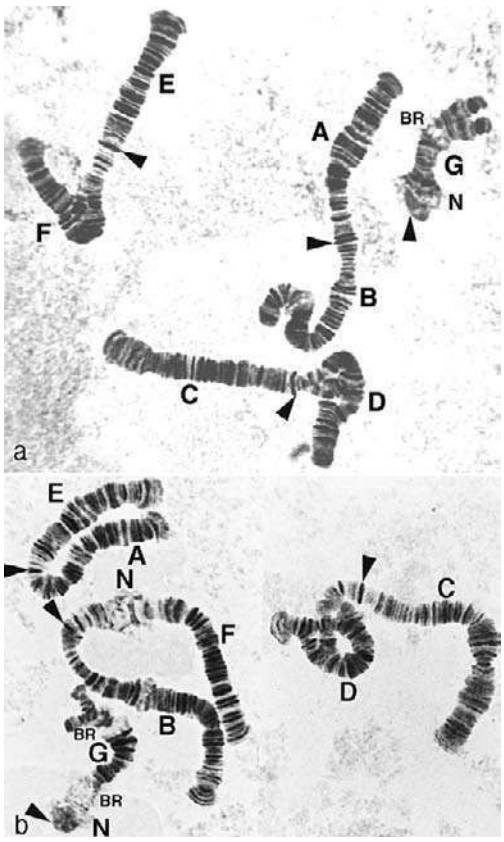

FIGURE 3 Polytene chromosomes in the salivary glands of the larvae of two species of chironomid midge. Orcein-stained squash preparations. (a) From the North American species *Chironomus decorus,* species b; (b) From the Australian species *C. oppositus*. For both species, labels A–F indicate arms of metacentric chromosomes, with arrowheads indicating the centromeres. The acrocentric chromosome G shows some breakdown of somatic pairing at the distal end in both species. Chromosomes AB and EF in *C. decorus* b have undergone whole-arm exchanges to form AE and BF chromosomes in *C. oppositus.* N and BR indicate nucleoli and Balbiani rings, respectively. Loop pairing, resulting from heterozygosity for paracentric inversions, can be seen in arms D and F in *C. decorus* b and in arm D in *C. oppositus* (Images kindly supplied by Dr. Jon Martin, University of Melbourne.)

dated by the formation of loops. Transcription of ribonucleic acid (RNA) from the DNA is accomplished at expanded regions called Balbiani rings, (BR in Fig. 3B), and the attachments of the polytene chromosomes to the nucleoli (N in Fig. 3B) by the nucleolar organizing regions are obvious. Polytene chromosomes have been most famously studied in the salivary glands and other glandular tissues in insects of the order Diptera, particularly in the fruit fly, *Drosophila melanogaster.* They display a large number of bands without any special staining, and the detail revealed is most useful for the localization of DNA sequences of various types, including single gene probes.

Supernumary or B chromosomes occur occasionally in insects of most orders. B chromosomes, when present, are in addition to the always present A chromosomes. Certainly the most variable and spectacular B chromosomes ever seen were found in the Australian plague locust, *Chortoicetes terminifera* (Fig. 2). These B chromosomes display over 20 different banding patterns after treatment with trypsin, and this treatment allowed the harmless identification of carriers of B chromosomes using interphase cells in the hemolymph, thus facilitating breeding experiments. These breeding experiments showed that single B chromosomes in males of *C. terminifera* were distributed into the sperm with a 50% frequency, but in females single B chromosomes were driven into the egg with a frequency of 80%. This meiotic drive in females should have ensured that every individual in the population carried a B chromosome. Since, however, they were found in only 10% of individuals, the B chromosomes must have been lowering the fitness of carriers. The situation supported a "parasitic" mechanism for the maintenance of B chromosomes in the population.

SUBCHROMOSOMAL ORGANIZATION IN INSECTS

Euchromatin and heterochromatin can be distinguished in insects in various ways. Euchromatin contains the active genes, and heterochromatin, contains mainly repetitious, transcriptionally inactive DNA. Heterochromatic segments of the chromosomes can be observed in meiosis because of their high degree of condensation during first prophase (Fig. 1). Heterochromatin may also be detected by hybridization *in situ* of repetitous DNA sequences, such as satellite DNA, to the chromosomes. The DNA of heterochromatin is replicated later in the S phase of the cell cycle than the DNA of the euchromatin. Examples of DNA replication that is both late to start and late to finish has been seen in the B chromosomes of *C. terminifera,* and in the sex chromosomes of the common earwig, *Forficula auricularia.* The C-banding technique (described shortly) can also be used to stain heterochromatic segments.

In most cases the heterochromatin of insects is constitutive (i.e., in a permanent state), but in some insects with peculiar life cycles, such as the Cecidomyidae, individual chromosomes or sets of them may be made facultatively heterochromatic before being eliminated from the soma or the germ line of one of the sexes.

Chromosome banding in insects is largely limited to C-banding (Fig. 2b), originally named because the repetitious DNA proximal to the centromeres was stained. The position of the C bands is often procentric in insects, but these bands can be procentric, terminal, or interstitially distributed on the chromosome arms in different races of the same species, as in the grasshopper *Caledia captiva.*

The results of treating insect chromosomes with trypsin or other reagents that induce the narrow G bands, which are distributed all over the chromosomes of vertebrates, are disappointing in insects. The dark G bands correspond to the chromomeres, regions of the chromosomes that are contracted during meiotic prophase, and since insect chromosomes display chromomeres during meiosis, it is surprising that they do not show typical G bands. The bands

revealed by trypsin treatment of the B chromosomes of *C. terminifera* seem to be exceptional, and they are a reflection of the C-banding patterns of these chromosomes (Fig. 2).

POLYPLOIDY

Polyploidy occurs when the zygote, or first cell, has more than two sets of haploid chromosomes. In insects, polyploidy is mainly restricted to parthenogenetic species and is largely limited to $3n$ and $4n$. Chromosomal sex determination is regarded as a major barrier to the formation of polyploids among bisexual species of insects because duplicated sex chromosomes, such as XXYY, would lead to uniformly XY sperm and therefore no possibility of sex determination.

Endopolyploidy is the occurrence of a multiplicity of the ploidy in the zygote in the somatic tissues of an organism. The term "endomitosis" is used if the chromosomes appear during cycles of endoreduplication but with no formation of a mitotic spindle and no cell division. Endopolyploidy, including endomitosis, is commonly seen in the somatic tissues of most insects. In a special case of endoreduplication, involving only one round of replication under a variety of treatments, the chromosomes that subsequently appear may remain closely associated at the centromeres, forming diplochromosomes, which were first seen in a locust. The formation of polytene chromosomes (Fig. 3) is also a special form of endoreduplication. Endopolyploid cells are very common in the tissues of all insects, and the phenomenon seems to reflect a tendency for insects to increase the bulk of certain tissues by increasing cellular size rather than cell number.

CHROMOSOMAL REARRANGEMENTS

Rearrangements occur in the chromosomes of insects when they occasionally break and rejoin in an irregular fashion. If any chromosomal rearrangement is maintained heterozygously in a population at a frequency greater than can be explained by recurrent chromosomal mutation, it is said to be polymorphic. There are a number of chromosomal rearrangements.

Paracentric Inversions

Paracentric inversions result when two breaks in one chromosome arm rejoin after the excised piece has inverted. These rearrangements are commonly recorded in polytene chromosomes, where the presence of them is shown by the formation of a loop allowing the homologues to be closely paired (Fig. 3). The presence of a chiasmata at meiosis within paracentrically inverted segments results in a dicentric chromosome and an acentric fragment, which cannnot be regularly transmitted. Paracentric inversions survive for long periods in many dipteran species because there is no chiasma formation in males and because the products of female meiosis are organized to ensure that a nonrecombinant for any paracentric inversion is deposited in the egg nucleus, with recombinants being placed in the unused polar bodies.

Because they have the capacity to lock up long combinations of syntenic genes, it has been assumed that inversion polymorphisms can be adaptive. For paracentric inversions, many studies with dipterans have been undertaken to link paracentric inversion polymorphism to aspects of the environment in which the particular insect exists.

Pericentric Inversions

Pericentric inversions result from breaks in each arm of a chromosome that rejoin after the excised piece containing the centromere has inverted. Pericentric inversion polymorphism was perhaps most famously studied in the morabine grasshopper, *Keyacris scurra.* White and coworkers used this rearrangement to develop adaptive topographies (defined by Sewell Wright) for various populations of *K. scurra,* that were on a saddle between adaptive peaks. The duplications and deletions that are the consequences of recombination within mutually pericentrically inverted segments seem to be largely avoided in insects bearing them at polymorphic frequencies. This is because the chromosomes are able to pair during meiosis without the inverted regions undergoing synapsis: so-called "torsion pairing."

Translocations

Translocations result from breaks in two chromosomes that allow exchange of pieces between the chromosomes. For breaks that are interstitial on the chromosome arms, a reciprocal translocation results, and in heterozygotes the translocated chromosomes synapse together at meiosis to form a quadrivalent (or a multivalent, if exchanges are more frequent). Multiple translocation heterozygosity has been observed, with resulting ring multivalents, in the cockroaches (Blattodea).

Centric Fusions

Centric fusions, or Robertsonian translocations, are special cases of translocation in which two breaks are very close to the centromeres of acrocentric chromosomes, causing the formation of a large metacentric or submetacentric chromosome and a very small remnant, which is lost. Centric fusions commonly distinguish chromosomal races or species in insects, but they are not seen to be maintained polymorphic in populations as frequently as they are in mammals. Centric fusions between sex chromosomes and autosomes results in the formation of neo-XY and X_1 X_2 Y systems in insects.

Dissociation

Dissociation, the reverse of fusion, involves the formation of two acrocentrics from a metacentric chromosome. Dissociation is rare because a donor centromere, a short arm, and a telomere are required; however this rearrangement was shown to occur in the dissociation that formed the two chromosomal races of the morabine grasshopper *K. scurra.*

Whole-Arm Interchanges

Whole-arm interchanges occur when chromosomal breaks and rejoinings near the centromeres of metacentric chromosomes result in the exchange of whole chromosome arms (Fig. 3). It has been noted that such exchanges distinguish races and species more frequently than reciprocal translocations, perhaps because the former maintain a sequence of coadapted genes in the arms concerned.

Complex Rearrangement

Complex rearrangements such as insertions, involving three or more breaks, have been noted in insect chromosomes, particularly after damage induced by radiation. Such work, particularly by H. Müller in *D. melanogaster,* led to each arm of the chromosome being defined as oriented from the centromere to the telomere.

See Also the Following Articles

Genetic Engineering • Parthenogenesis • Sex Determination

Further Reading

Cook, L. G. (2000). Extraordinary and extensive karyotypic variation: A 48-fold range in chromosome number in the gall-inducing scale insect *Apiomorpha* (Hemiptera: Coccoidea: Eriococcidae). *Genome* **43,** 169–190.

Crosland, M. W. J., and Crozier, R. H. (1986). *Myrmecia pilosula,* an ant with one pair of chromosomes. *Science* **231,** 1278.

Giles, E. T., and Webb, G. C. (1972). The systematics and karyotype of *Labidura truncata* Kirby, 1903 (Dermaptera: Labiduridae). *J. Aust. Entomol. Soc.* **11,** 253–256.

Gregg, P. C., Webb, G. C., and Adena, M. A. (1984). The dynamics of B chromosomes in populations of the Australian plague-locust, *Chortoicetes terminifera* (Walker). *Can. J. Genet. Cytol.* **26,** 194–208.

Hoy, M. (1994). "Insect Molecular Genetics." Academic Press, San Diego.

John, B. (ed.). (1974–1990). "Animal Cytogenetics," Vol. 3, "Insecta." Bornträger, Berlin.

Jones, R. N., and Rees, H. (1982). "B Chromosomes." Academic Press, London.

King, M. (1993). "Species Evolution: The Role of Chromosome Change." Cambridge University Press, Cambridge, U.K.

Kipling, D. (1995). "The Telomere." Oxford University Press, Oxford, U.K.

Shaw, D. D., Webb, G. C., and Wilkinson, P. (1976). Population cytogenetics of the genus *Caledia* (Orthoptera: Acridinae). II: Variation in the pattern of C-banding. *Chromosoma* **56,** 169–190.

White, M. J. D. (1973). "Animal Cytology and Evolution," 3rd ed. Cambridge University Press, London.

White, M. J. D. (1978). "Modes of Speciation." Freeman, San Francisco.

Chrysalis

Frederick W. Stehr
Michigan State University

A chrysalis (plural chrysalids) is the pupa of a butterfly, usually belonging to the family Papilionidae, Pieridae, or Nymphalidae. It is commonly found suspended or hanging from a leaf, twig, or branch, or even a windowsill, arbor, or other suitable structure. Not all species in these families form chrysalids. For example, the parnassians in the Papilionidae and the wood nymphs (Satyrinae) in the Nymphalidae pupate in a minimal cocoon in grass, leaves, or litter. The pupae of the families Lycaenidae and Riodinidae are also not suspended and usually are concealed in leaves or litter.

FIGURE 1 Pupa of the butterfly *Papilio cresphontes,* showing the strand of silk that holds some butterfly pupae in an upright position. (Photograph from the Teaching Collection, Department of Entomology, Michigan State University.)

In forming the chrysalis, the prepupal caterpillar has to perform the seemingly impossible maneuver of spinning the silk pad to attach its cremaster (caudal pupal hooks) while maintaining its grip; it then must molt the larval skin as it attaches to the silk pad. Members of the Pieridae and Papilionidae (except Parnassinae) secure the chrysalis in an upright position with a band of silk around the middle (Fig. 1).

Chrysalids are usually angular, with projections, tubercles, spines, and sometimes gold or silver flecks. They are often cryptically colored so that they blend into the surrounding materials but some, like the monarch chrysalis, are smooth with gold flecks. In emerging from the chrysalis the adult splits the chrysalis ventrally and dorsally at the anterior end, crawls out, and suspends itself from the pupal skin while its wings expand.

See Also the Following Article

Pupa and Puparium

Further Reading

Stehr, F. W. (ed.) (1987). "Immature Insects," Vol. 1. Kendall/Hunt, Dubuque, IA.

Stehr, F. W. (ed.) (1992). "Immature Insects," Vol. 2. Kendall/Hunt, Dubuque, IA.

Cicadas

M. S. Moulds

Australian Museum, Sydney

Cicadas form a small part of the order Homoptera, a diverse group of insects whose mouthparts comprise a jointed rostrum for piercing and sucking up liquid food. They make up the superfamily Cicadoidea, distinguished by having three ocelli, an antennal flagellum usually of five segments, and a complete tenorium (internal development of the head for attachment of muscles); nymphs burrow and develop underground. The family arrangement for cicadas remains in a state of flux but is now generally accepted as being two families, the majority falling within the Cicadidae, and just two extant species plus some fossil species in the Tettigarctidae. There are almost 2000 named species, with perhaps as many again awaiting description.

Cicadas are mostly tropical or subtropical insects, but many also inhabit temperate regions. Some are minor pests of sugarcane, rice, coffee, and fruit trees, either reducing the vigor of the plants by nymphal feeding or weakening branches by oviposition, which in turn may cause the branches to break under crop load.

STRUCTURE AND FUNCTION

Cicadas typically possess a broad head delimited by a pair of large compound eyes, a large pro- and mesothorax housing mostly wing and leg muscles, a small metathorax, an abdomen that in the male is highly modified to accommodate the organs of sound production and reception, and two pairs of membranous wings that are usually held tentlike over the body at rest.

The head is dominated by a large, noselike postclypeus that houses muscles for sucking sap through the rostrum; the three jewellike ocelli detect the direction of light sources and, if asymmetrically covered, cause erratic flight.

The foreleg femora are characteristically enlarged and swollen. On the nymph these are even more enlarged (Fig. 1), serving the nymph for subterranean tunneling.

The abdomen carries the organs of reproduction and of hearing and, in males, also sound production.

SOUND PRODUCTION AND RECEPTION

Cicadas are best known for their ability to produce loud sound. No other insect has developed such an effective and specialized means of doing so. The calls are mating songs produced only by the males. Each species has its own distinctive song and attracts only females of its own kind (Fig. 2).

The organs of sound production are the tymbals, a pair of ribbed cuticular membranes located on either side of the first abdominal tergite (Fig. 3). In many species the tymbals are partly or entirely concealed by tymbal covers, platelike anterior projections of the second abdominal tergite. Contraction of internal tymbal muscles causes the tymbals to buckle inward, and relaxation of these muscles allows the tymbals to pop back to their original position. The sound produced is amplified by the substantially hollow abdomen, which acts as a resonator.

Many species sing during the heat of the day, but some restrict their calling to semidarkness at dusk. Often the species that sing at dusk are cryptic in coloration and gain further protection from predatory birds by confining their activity to dusk. The loud noise produced by some communal day-singing species actually repels birds, probably because the noise is painful to the birds' ears and interferes with their normal communication. The American periodical cicadas

FIGURE 1 Mature nymph of *Cyclochila australasiae,* lateral view.

FIGURE 2 A mating pair of northern cherrynose, *Macrotristria sylvara* (family Cicadidae). This large and colorful species is found in tropical northeastern Australia. (Photograph by Max Moulds.)

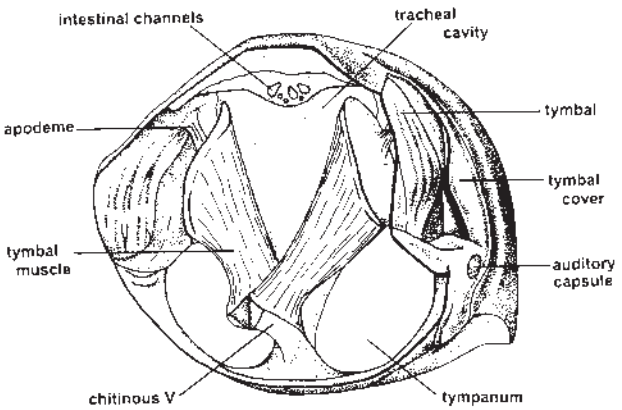

FIGURE 3 Transverse section of male abdomen of *Tamasa tristigma* at the first abdominal segment with the thorax removed. Exposed are the large tymbal muscles anchored basally to a chitinous V and attached dorsally via an apodeme to the sound-producing tymbals. Sound received by the tympana is transferred to the auditory capsules.

have mass emergences, and although their song is not sufficiently loud to repel birds, the number of individuals is so large that predatory birds soon lose their appetite for them.

Both sexes have organs for hearing. Sound is received by a pair of large, mirrorlike membranes, the tympana, which are often concealed below the opercula (Fig. 3). The tympana are connected to an auditory organ by a short slender apodeme. A singing male creases the tympana to avoid being deafened by his own song.

LIFE CYCLE

Eggs are laid in branches of trees and shrubs or in the stems of grasses (the nymphal food plants) in small slits cut into the surface by the female's spearlike ovipositor. The number of eggs laid in each slit varies between both species and individuals. Usually it is about 10 to 16, although the number laid per slit by a single female can range from 3 or fewer to more than 20.

A female makes many egg slits and often distributes her eggs at more than one site. A batch of eggs can number 300 or more. Some species, such as many *Cicadetta,* select only living tissue for oviposition, whereas others choose only dead or dying tissue. Many days, often in excess of 100, may pass before the nymphs hatch.

On hatching, the young nymphs are encased in a thin transparent skin that encloses the appendages separately but restricts their function. These pronymphs quickly wriggle their way along the egg slit to its entrance. A spine at the apex of the abdomen probably assists this exit and also in casting off the pronymphal skin. The young nymphs fall to the ground, whereupon they immediately seek shelter in the soil and later search for a root from which to feed by sucking sap.

Cicadas spend most of their life underground, slowly growing to maturity through five instars (Fig. 1). The length of life cycle is known only for a small number of species. Some grass-feeding species mature within a year. The American periodical cicadas, *Magicicada* species, have a life cycle spanning 13 or 17 years, the longest known for any insect. Periodical cicadas are consistently regular in their life cycle length, but most other cicadas change by a year or two, and even individuals from a single egg batch can mature at different rates.

For most species, emergence from the final nymphal skin occurs during the first few hours after dark; the laborious process can last an hour or more. The adult life usually lasts 2 to 4 weeks, but some grass-dwelling species possibly live only 3 to 4 days. Some of the larger tree-inhabiting species probably live 8 or more weeks.

FAMILY CICADIDAE

The family Cicadidae includes all but two extant species. Two subfamilies are widely recognized, the Cicadinae, which have tymbal covers present, and the Tibicinae, with tymbal covers absent.

FAMILY TETTIGARCTIDAE

The family Tettigarctidae includes the other two extant species, *Tettigarcta crinita* and *T. tomentosa,* both found only in Australia. This family also includes 13 genera known from Cenozoic fossils.

The Tettigarctidae differ from other cicadas in several features. Most notable is the presence of tymbals in both sexes, but instead of producing airborne songs, they create low-level vibrations of the substrate below the adult. These substrate vibrations are detected by sensory empodia between the claws on all legs; the tympana used for hearing in other cicadas are lacking.

See Also the Following Articles

Auchenorrhyncha • Hearing • Sternorrhyncha • Vibrational Communication

Further Reading

Boer, A. J., de (1995). Islands and cicadas adrift in the west-Pacific. Biogeographic patterns related to plate tectonics. *Tijdschr. Entomol.* **138,** 169–244.

Boer, A. J., de, and Duffels, J. P. (1996). Historical biogeography of the cicadas of Wallacea, New Guinea and the West Pacific: A geotectonic explanation. *Palaeogeogr. Palaeoclimato. Palaeoecol.* **124,** 153–177.

Claridge, M. F. (1985). Acoustic signals in the Homoptera: Behaviour, taxonomy, and evolution. *Annu. Rev. Entomol.* **30,** 297–317.

Kato, M. (1956). The Biology of the Cicadas [Bulletin of the Cicadidae Museum]. Iwasaki Shoten, Jinbocho Kanda, Tokyo. [In Japanese; headings, subheadings, captions and index in English. Facsimile reprint, 1981, Scientist Inc., Japan.]

Duffels, J. P. (1988). The cicadas of the Fiji, Samoa and Tonga Islands, their taxonomy and biogeography (Homoptera, Cicadoidea). *Entomonograph* **10,** 1–108.

Duffels, J. P., and van der Laan, P. A. (1985). "Catalogue of the Cicadoidea (Homoptera, Auchenorhyncha) 1956–1980," Series Entomologica, Vol. 34. Junk, Dordrecht, Netherlands.

Metcalf, Z. P. (1963). "General Catalogue of the Homoptera," Fasc. 8, Cicadoidea. Part 1, Cicadidae. Part 2, Tibicinidae. [Species index by Virginia Wade, 1964.] University of North Carolina State College, Raleigh.
Moulds, M. S. (1990). "Australian Cicadas." New South Wales University Press, Kensington, Australia.
Moulton, J. C. (1923). Cicadas of Malaysia. *J. Fed. Malay States Mus.* **11,** 69–182, pls. 1–5.
Williams, K. S., and Simon, C. (1995). The ecology, behaviour, and evolution of periodical cicadas. *Annu. Rev. Entomol.* **40,** 269–295.

Circadian Rhythms

Terry L. Page
Vanderbilt University

Circadian rhythms are daily oscillations in physiology, metabolism, or behavior that persist (or free run) in organisms that have been isolated from periodic fluctuations in the environment. These rhythms are under the control of innate regulatory systems that are based on internal oscillators (or pacemakers) whose periods approximate those of the naturally recurring 24-h environmental cycles. The oscillators are subject to control by a limited number of these environmental cycles that synchronize or entrain the period to exactly 24 h and establish specific phase relationships between the rhythms and the external world (Fig. 1). Light cycles are virtually universally effective in the entrainment of circadian rhythms, and in insects, daily cycles of temperature are also effective.

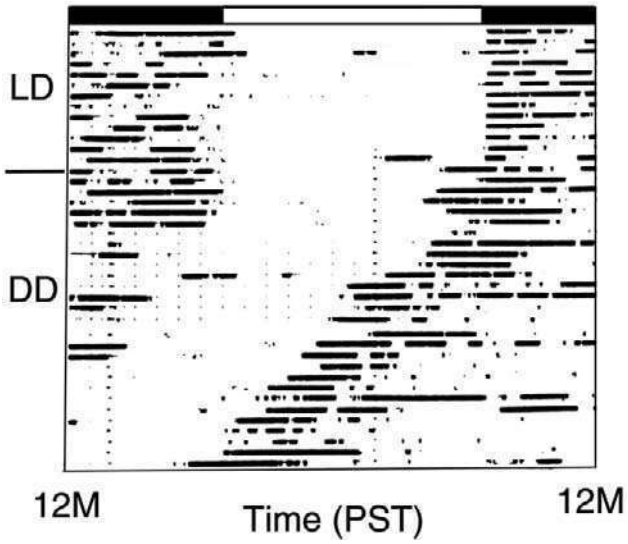

FIGURE 1 Event recording of the wheel-running activity of a cockroach, *L. maderae.* Data for successive days are placed one below the other in chronological order. The bar at the top of the record indicates the light cycle to which the animal was exposed during the first 14 days of the recording. Then animal was then placed in constant darkness (DD) and its endogenously generated, free-running circadian rhythm was expressed for the remainder of the record with a period of about 23.5 h.

FUNCTION OF THE CIRCADIAN SYSTEM

In insects, the circadian system is responsible for imposing daily rhythmicity on a variety of processes, including locomotor activity, stridulation, oviposition, hatching, pupation and pupal eclosion, pheromone release, retinal sensitivity to light, and daily cuticle growth. This list is by no means exhaustive. It is generally accepted that the functional importance of this control is to restrict processes that are best undertaken at a particular phase of the environmental cycle to a particular time of day. It has also been suggested that a secondary role of the circadian system is to provide for internal temporal organization, coordinating the timing of various processes within the individual.

In addition to its role in generating daily rhythms, the circadian clock has been shown to be involved in photoperiodic time measurement for seasonal regulation of reproduction, development, and diapause in many insects. In honey bees (*Apis mellifera*), it is also involved in time measurement necessary for time-compensated sun orientation and in *Zeitgedachtnis,* which is the ability to return at the appropriate time to a food source that is available only at particular times of day. Thus the circadian system functions as a biological clock, capable of providing the individual with information on the time of day and with the ability to measure lapse of time.

PHYSIOLOGICAL BASIS OF CIRCADIAN ORGANIZATION

The study of the anatomical and physiological organization of circadian systems of insects has a long and productive history. The heuristic model generally used in these studies is illustrated in Fig. 2. There are four essential elements: (1) a pacemaker or oscillator that generates the primary timing signal, (2) photoreceptors for entrainment, and two coupling pathways, (3) one that mediates the flow of entrainment information from the photoreceptor to the pacemaker and (4) another that couples the pacemaker to the effector mechanisms that it controls. The model identifies several basic questions. Can the anatomical location of the circadian clock be identified? What are the pathways and mechanisms by which inputs to the pacemaking system regulate its phase and period? Finally, what are the neural and endocrine signals

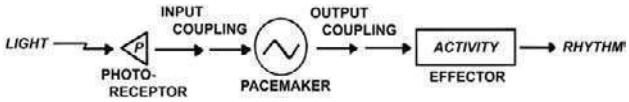

FIGURE 2 Functionally defined model of the circadian system. An entrainment pathway that consists of a photoreceptor and coupling mechanism (input) synchronizes a self-sustaining oscillator (pacemaker) to the external light/dark cycle. The output of the pacemaker regulates the timing of various processes (e.g., activity) via coupling to the effector mechanisms.

by which the pacemaking system regulates the various processes under its control?

Circadian Oscillations Are Generated by Discrete, Localized Populations of Cells

Studies on pacemaker localization in insects have largely focused on behavioral rhythms (locomotor activity or eclosion) and their control by the nervous system. Compelling evidence that the brain is the site of generation of circadian timing signals for rhythms in behavior has been obtained in several species, with much of the early work involving studies on the locomotor activity rhythm of the cockroach.

In 1968 it was first discovered that surgical removal of both optic lobes or disconnecting them from the rest of the brain by section of the optic tracts abolished the activity rhythm of the Madeira cockroach, *Leucophaea maderae.* Results of lesion studies on other cockroach species, several species of crickets, and beetles have also suggested that the optic lobes might contain the pacemaker. Compelling evidence arose from the observation that it is possible to transplant optic lobes between cockroaches whose activity rhythms had quite different free-running periods. Animals that received transplanted optic lobes recovered rhythmicity in a few weeks with regeneration of the optic tracts, and the preoperative period of the donor and the postoperative period of the host were strongly correlated. Thus, the transplantation of the optic lobes not only restored the rhythm of locomotor activity but also, critically, imposed the period of the donor animal's rhythm on the activity of the host. Other studies involving small electrolytic lesions indicated that the cells responsible for generating the circadian signal have their somata and/or processes in the proximal half of the optic lobe, likely in a group of cells located ventrally near the medulla.

In contrast to cockroaches, crickets, and beetles, in a variety of other insects the optic lobes do not appear to be required for rhythmicity and the pacemaker appears instead to reside in the cerebral lobes (midbrain). In a classic series of experiments by James Truman and colleagues it was shown that the circadian pacemaker that controls the timing of eclosion in two silkmoth species, *Hyalophora cecropia* and *Antheraea pernyi,* is located in the cerebral lobes of the brain. The time of day at which eclosion occurs is different for the two species. When the insects are maintained in a photoperiod of 17:7 (L:D) hours, *H. cecropia* emerges shortly after lights-on while *A. pernyi* emerges just before lights-off. Removal of the brain did not prevent eclosion, but did disrupt its timing. However, if the brain was reimplanted in the abdomen, normal rhythmicity was restored under both entrained and free-running conditions. When brains were transplanted between species, individuals exhibited normal species-specific eclosion behavior, but the phase of the rhythm was characteristic of the donor and not the host. The demonstration that the transplanted brains restored rhythmicity and determined the phase of the rhythm left little doubt that the circadian pacemaker that regulates the timing of the eclosion rhythm is located in the brains of these moths. The fact that the pacemaker was located in the cerebral lobes and not the optic lobes was demonstrated by subdividing the brain prior to transplantation. It was found that the optic lobes were unnecessary and that transplantation of the cerebral lobes alone was sufficient to restore rhythmicity.

Similarly, in a variety of dipterans, including the fruit fly, the house fly, the blow fly, and the mosquito, regions of the nervous system controlling locomotor activity rhythms have been dissected with both surgical and genetic lesions, and in each instance the pacemaking oscillation appears to be generated in the cerebral lobes. In the fruit fly, *Drosophila melanogaster,* extensive behavioral and genetic evidence demonstrates a crucial role for the *period (per)* gene in the circadian pacemaker controlling locomotor activity and eclosion rhythms (see later). The *per* gene is widely, and in some cell types rhythmically, expressed in the fly, including the head, thorax, and abdomen; thus, its spatial expression pattern in wild-type flies provided no definitive localization of the central pacemaker. However, the expression pattern has been altered by numerous genetic and molecular manipulations and it has been possible to determine the identity of the pacemaker cells in *Drosophila* by correlating *per* expression in specific cell types with the presence or absence of behavioral rhythmicity. The results suggest that only a few neurons between the lateral protocerebrum and the medulla of the optic lobes, the lateral neurons, are necessary for the generation of a circadian rhythm in locomotor activity.

The potential for further cellular identification of pacemaker neurons in insects was provided by an observation that in cockroaches and crickets optic lobe neurons that fulfilled the predicted anatomical criteria to be pacemaker cells were labeled by an antibody to crustacean pigment-dispersing hormone (PDH). When anti-PDH was applied to *Drosophila* brains, it labeled a ventral subset (LNv) of the *per*-expressing lateral neurons that were identified as pacemaker neurons in genetic studies. Taken together, the results indicated that the PDH-immunoreactive neurons are strong candidates for pacemaker neurons in insects and raise the possiblity that the insect version of crustacean PDH (called pigment-dispersing factor or PDF) may be an important temporal signaling molecule.

Interestingly, the numbers and projection patterns of PDF neurons in cockroaches and crickets are strikingly similar to those of the LNv of *Drosophila,* suggesting that they are functionally homologous. The most salient difference in the morphology of these neurons is in the locations of their somata: between the lobula and the medulla of the optic lobes in cockroaches and crickets, as opposed to between the medulla of the optic lobes and the lateral margin of the cerebral lobes in fruit flies. This difference may be sufficient to account for the fact that the lesion and transplant studies suggested different anatomical organizations for pacemaker structures in the central nervous system in different insects.

Circadian Pacemakers Are Also Found in Tissues outside the Nervous System in Insects

The localization of circadian pacemakers that regulate behavioral rhythms to the brain raised the question of whether other rhythms are controlled by the same clock. In crickets, beetles, and cockroaches studies indicate that the pacemaker regulating the daily rhythm in retinal sensitivity to light as measured by electroretinogram (ERG) amplitude is located in the optic lobe and suggest that the same pacemaker controls both the ERG amplitude and the activity rhythms. However, in other cases rhythms have been found to be regulated by pacemakers outside the nervous system. These include rhythmic secretion of cuticular layers in newly molted cockroaches, the release of sperm from the testis into the seminal ducts in gypsy moths, and the timing of ecdysteroid release from the prothoracic gland of the cynthia moth, *Samia cynthia.* In each of these examples the rhythms were shown to persist *in vitro* in the absence of neural pacemaking structures.

These results indicate that the distribution of circadian pacemaking centers may be widespread in insects. In support of this view, one recent study by Plautz and co-workers with *Drosophila,* in which the *per* promoter was coupled to the coding sequence for luciferase, indicated that rhythmic promoter activity could be detected in a variety of tissues, including the wing, leg, proboscis, and antennae maintained in isolation in tissue culture.

The fact that the circadian system in the individual may be composed of several widely distributed oscillators raises the question of whether there is communication between component oscillators. In general the answer is uncertain. Work on cockroaches has shown that the bilaterally distributed oscillators in the two optic lobes are connected to one another (mutually coupled) and suggested that the coupling was relatively strong. In contrast, both in the beetle, *Blaps gigas,* and in crickets the data indicated that coupling between optic lobe pacemakers is either absent or weak. Coupling relationships among other oscillators have not yet been systematically explored.

Photoreceptors for Entrainment

Extraretinal photoreceptors are typically involved in entrainment of behavioral rhythms. The classical example is the silkworm, in which it was shown that the photoreceptor for entrainment of the eclosion rhythm resides in the brain. Brains were removed from silkworm pupae and were either replaced in the head region or transplanted to the abdomen. The pupae were then placed in holes in a partition that separated two chambers in which the light/dark cycles were out of phase. Whether the pupae entrained to the light cycle to which the anterior end of the pupae was exposed or entrained to the light cycle at the posterior end corresponded to the location of the brain.

Additional evidence for extraretinally mediated entrainment of pacemakers that are located in the nervous system has been obtained in a variety of other insects, including other lepidopterans, dipterans, and orthopterans. In those instances in which there is evidence on the location of the photoreceptor, the brain appears to be the most likely site. However, more precise identification of the cells involved in the phototransduction has not been accomplished.

Even though the compound eyes may not be necessary for entrainment, they may nevertheless participate. In *Drosophila,* for example, genetic lesions to the eyes or the phototransduction pathway can alter the entrainment pattern. Further, there are at least two insects, the cockroach and the cricket, in which the compound eyes appear to be the exclusive photoreceptors for entrainment because sectioning the optic nerves between the eyes and the optic lobe or painting over the compound eyes eliminated entrainment of the locomotor activity rhythm to light cycles.

As noted above, there are several instances in which there is convincing evidence for circadian pacemakers outside the nervous system. In the case of the moth testis, since the rhythm measured *in vitro* responds to light, some cells in the testis–seminal duct complex must be photosensitive. Similarly, in the saturnid moth *S. cynthia,* the photoreceptor for entrainment of the pacemaker in the prothoracic glands appears to be in the gland itself.

Signals to Communicate Timing Information

Another important issue is how circadian oscillators impose periodicity on the various physiological and behavioral processes they control. *A priori,* several alternative mechanisms are plausible. Timing information within the individual could be represented by the level of a circulating hormone, impulse frequency in specific neural circuits, changes in general levels of neural excitability through neuromodulation, or, as the weight of the available evidence suggests, some combination of these mechanisms.

There are a large number of studies that suggest that secretion of a variety of insect hormones, including ecdysone, prothoracicotropic hormone, and eclosion hormone, is under the control of the circadian system during development. The experiments involving the transplantation of the silkworm brain, described above, provide the clearest demonstration of a hormonal link in the control of behavior by the circadian system. The signal for the eclosion behavior is the eclosion hormone that is produced in neurosecretory cells located in a region near the midline of the brain, the pars intercerebralis, and released via the neurohemal organs, the corpora cardiaca. The release of the hormone triggers release of two other peptide hormones, pre-ecdysis-triggering hormone (PETH) and ecdysis triggering hormone (ETH). PETH and ETH act on the central nervous system to initiate a stereotyped sequence of behavior that ultimately results in the emergence of the adult moth from the pupal case.

The role of humoral factors in the regulation of adult behaviors in insects (e.g., locomotor activity) is less clear. In cockroaches and crickets, the timing signal that originates in the optic lobe is transmitted to the brain via the optic tracts, and transmission from the brain to the activity centers in the thorax requires that the connectives of the ventral nerve cord be intact. Nerve impulse activity is rhythmic in both the optic tracts and the cervical connective.

In summary, the mechanism by which circadian phase information is transmitted to behavioral effectors in insects is generally not well understood. The emerging picture is that temporal regulation of behavior involves a modulation of excitability in the central nervous system. Axonal connections between the brain and the lower elements of the central nervous system are clearly required for the maintenance of some behavioral rhythms (e.g., cockroach locomotor activity), whereas others appear to rely heavily on hormonal mechanisms (moth eclosion). An important step in understanding how temporal information is transmitted will be the identification of the signal molecules involved.

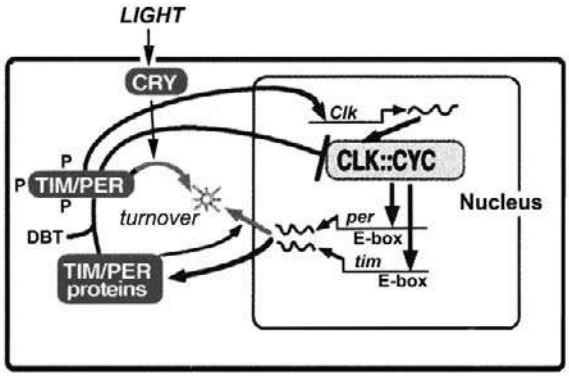

FIGURE 3 Molecular model of the circadian pacemaker of *Drosophila* showing the proposed negative feedback loop of the oscillation. CLK and CYC heterodimers bind to E boxes of nuclear DNA promoting transcription of *per* and *tim* genes. TIM and PER proteins heterodimerize and are phosphorylated by DBT. The heterodimer enters the nucleus and inhibits the positive regulation by the CLK/CYC heterodimer. Light enters the system through CRY and promotes turnover of TIM and PER (modified from Dunlap, 1999).

MOLECULAR BASIS OF CIRCADIAN ORGANIZATION

There has been remarkable progress in the past 15 years in identifying the molecular basis of circadian clocks in a variety of organisms. In animals, much of this progress has resulted from pioneering work with the fruit fly. In 1971 the first clock gene, the *period* gene, was discovered in a mutagenesis screen in *D. melanogaster.* A decade later the gene was cloned, paving the way for studies of the gene's regulation. This work led to the discovery of several other genes in *D. melanogaster* that appear to be part of the clock mechanism, including those involved in entrainment.

Molecular Basis of the Clock

Four genes, the transcriptional regulators *period (per)*, *timeless* (*tim*), *cycle* (*cyc*), and *clock* (*clk*), have been shown to be critical components for generating the basic circadian oscillation. Of the four, three, *per, tim,* and *clk,* are rhythmically expressed and circadian oscillations in both mRNA and protein levels are well documented. The fourth gene, *cyc,* is expressed at relatively constant levels throughout the day. Both CLK and CYC proteins are transcription factors that utilize basic helix-loop-helix domains to bind to E boxes, and both contain protein–protein interaction domains (PAS domains) that likely mediate the association of the two proteins with each other, thus forming heterodimers. The fundamental mechanism for generating the oscillation involves a transcription/ translation negative feedback loop. The basic loop is illustrated in Fig. 3. A heterodimer composed of CLK and CYC binds to promoters of *per* and *tim,* leading to an increase in transcription of these two genes that continues throughout the day. Levels of mRNA for the two genes peak in the early night. Protein products of these two genes increase as well, but peak levels of protein are delayed by several hours, peaking after the middle of the subjective night. PER and TIM themselves form a heterodimer, interacting through PAS domains. The heterodimer moves to the nucleus and functions as the negative element in the feedback loop, acting on the positive regulators CLK and CYC to suppress their activation of the *per/tim* promoters. This leads to a decline in the *per* and *tim* mRNA levels that continues throughout the night. The degradation of PER and TIM allows the cycle to start over.

The time delay between mRNA synthesis and the accumulation of PER and TIM is likely to be a critical element in the generation of the oscillation. PER is unstable in the absence of TIM. The dimerization stabilizes PER and promotes nuclear entry. In addition, both PER and TIM are phosphorylated, probably through the action of a homolog of casein kinase identified as *double-time* (*dbt*). This phosphorylation appears to be involved in regulation of PER turnover.

Mechanism of Entrainment

In *Drosophila,* light acts to cause a rapid decrease in the levels of TIM, and because TIM stabilizes PER, PER levels also decline. In the late day and early night when levels of these proteins are increasing, their destruction delays the progress of the oscillation, whereas in the late night and early day PER and TIM levels are decreasing, and hastening their demise advances the oscillation. Interestingly, genetic ablation of the eyes or mutations in the visual phototransduction pathway, although reducing sensitivity of the circadian clock to light, do not block its entrainment. The altered sensitivity to light observed with mutations that affect the visual system

indicates that an opsin-based photoreceptor can contribute to entrainment of the circadian rhythm of locomotor activity, but the persistence of entrainment in these mutants implicates an extraretinal photoreceptor. Action spectra for entrainment have suggested a flavin-based photoreceptor. Cryptochrome (CRY), is a member of a family of flavoproteins, which includes photolyases and plant blue-light receptors. A mutant allele of the *cry* gene disrupts normal light responses of the locomotor activity rhythm, whereas flies overexpressing CRY are hypersensitive to light pulses. Further, in the periphery, CRY is required for light-dependent TIM degradation. These results suggest that CRY is a central element in the phototransduction pathway for entrainment.

The extent to which the molecular mechanisms detailed for *Drosophila* are applicable to other insects is not yet clear. However, there has been considerable progress in identifying homologous proteins in mammals, and although there are differences in detail, the basic framework of the oscillator seems to have persisted through the evolutionary process, giving confidence that the story will be broadly applicable to other insects as well.

Output of the Molecular Clock

The general supposition is that the clock ultimately regulates rhythms through the regulation of gene expression. This view is supported by the observation that there are several clock-controlled genes (CCGs) in *Drosophila.* However, at this point there are no examples in which a CCG has been linked directly to an overtly expressed physiological or behavioral rhythm, and this is an area of research that is likely to receive increased attention as researchers work to further elucidate the molecular details of the circadian system.

See Also the Following Articles

Brain and Optic Lobes • Drosophila melanogaster

Further Reading

Dunlap, J. C. (1999). Molecular bases for circadian clocks. *Cell* **96,** 271–290.

Giebultowicz, J. M. (2000). Molecular mechanism and cellular distribution of insect circadian clocks. *Annu. Rev. Entomol.* **45,** 769–793.

Hall, J. C. (1995). Tripping along the trail to the molecular mechanisms of biological clocks. *Trends Neurosci.* **18,** 230–240.

Page, T. L. (1985). Clocks and circadian rhythms in insects. *In* "Sensory Physiology" (G. Kerkut and L. Gilbert, eds.), Vol. VI of "Comprehensive Insect Biochemistry, Physiology, and Pharmacology," pp. 577–652. Pergamon Press, Oxford.

Page, T. L. (1990). Circadian organization in the cockroach. *In* "Cockroaches as Models for Neurobiology: Applications in Biomedical Research" (I. Huber, ed.), pp. 225–246. CRC Press, Boca Raton, FL.

Plautz, J. D., Kaneko, M., Hall, J. C., and Kay, S. A. (1997). Independent photoreceptive circadian clocks throughout *Drosophila. Science* **278,** 1632–1635.

Saunders, D. (1982). "Insect Clocks," 2nd ed. Pergamon Press, Oxford.

Zitnan, D., Ross, L. S., Zitnanova, I., Hermesman, J. L., Gill, S. S., and Adams, M. E. (1999). Steroid induction of a peptide hormone gene leads to orchestration of a defined behavioral sequence. *Neuron* **23,** 523–535.

Circulatory System

Thomas A. Miller

University of California, Riverside

Insects have an open circulatory system. This means that the internal organs and tissues are bathed in hemolymph, which is propelled actively to all internal surfaces by specialized pumps, pressure pulses, and body movements and is directed by vessels, tubes, and diaphragms. Without such constant bathing, tissues would die. The internal organs and tissues depend on the circulatory system for the delivery of nutrients, both to carry away excretion products and as the chemical communication pathway by which hormone messengers coordinate development and other processes. So vital is this function that it can be equated to the umbilical cord of a human fetus during development (Fig. 1).

Gas exchange in insects occurs via the tracheal system, which supplies all internal organs with tracheole tubules from spiracular openings in the body wall of terrestrial insects or from gill structures in aquatic insects. However, the hemolymph has the capacity to dissolve carbon dioxide gas in the form of bicarbonate ions. A few insects live in low oxygen environments and have a type of hemoglobin that binds oxygen at very low partial pressures, but for the most part oxygen is supplied and carbon dioxide is removed by ventilation through the tracheal system via the same system.

Besides the functions already mentioned, the circulatory system provides a medium in which battles are fought between the insect host and a myriad of invading disease microorganisms, including viruses, bacteria, fungi, and insect parasites. Principal participants in these interactions are the blood cells or hemocytes.

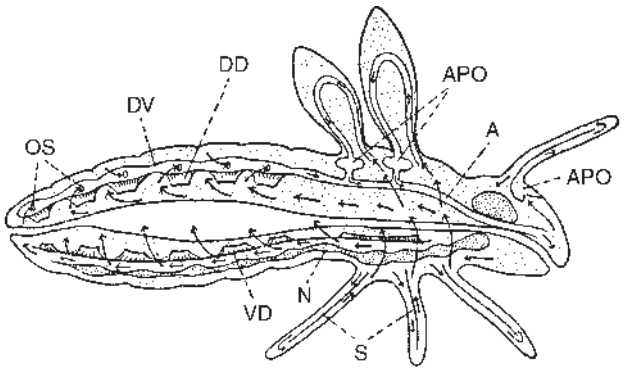

FIGURE 1 Delivery of the hemolymph to all tissues is so vital that a number of structures have evolved to ensure complete circulation including accessory pulsatile organs (APO), aorta (A), dorsal vessel (DV), dorsal diaphragm (DD), ostial valves (OS), and septa (S). Perfusion of the abdominal ventral nerve cord (N) is associated with a ventral diaphragm (VD). [Modified from Wigglesworth, V. B. (1972). "Principles of Insect Physiology." 7th ed. Chapman & Hall, New York, London, with kind permission of Kluwer Academic Publishers.]

While maintaining the body tissues, the circulatory system is the medium in which homeostasis is ensured, including the regulation of pH and inorganic ions, as well as the maintenance of proper levels of amino acids, proteins, nucleic acids, carbohydrates, and lipids. Any change in the hemolymph immediately affects all organs bathed. The time for complete mixing of the hemolymph depends on the size of the insect, but it can be up to 5 min in a resting adult cockroach weighing about a gram. Any substance injected into a healthy insect will eventually appear at the extreme ends of all appendages in a few minutes, emphasizing the efficiency of the delivery mechanisms, which can be marvels of microhydraulic engineering.

DORSAL VESSEL

The principal organ of hemolymph propulsion is the dorsal vessel, or at least it is the most visible organ associated with hemolymph movement. By tradition, the abdominal portion of the dorsal vessel is called the "heart" and the thoracic and cephalic (head) portion the "aorta." Both terms are borrowed from better-known vertebrate structures and give an inaccurate impression of the roles of those structures.

The dorsal vessel, especially in the abdomen of insects, is suspended in the pericardial sinus, which is delimited by the dorsal cuticle and the dorsal diaphragm (when present). Contractions of the dorsal vessel operate against the pull of lateral connective tissues, which are responsible for dilating or opening the vessel (in diastole) following each contractile stroke (called systole or systolic contraction). In most larvae of holometabolous insects (insects that possess a pupal stage) and in most primitive insects, the dorsal vessel is a simple tube running from the rear of the abdomen to the head, where it becomes closely associated with the top of the foregut and then travels under and opens just underneath or in front of the brain. This arrangement ensures a constant supply of nutrients and removal of waste products to and from the brain mass. In addition, the dorsal vessel is often intimately associated with the retrocerebral nervous system (including the hypocerebral ganglion, corpora cardiaca, and corpora allata complex) just behind the brain, which may deliver neurohormones and possibly other hormones into the aorta by way of specialized release sites.

The dorsal vessel has openings called ostia along the sides and ventral surfaces of each segment of the abdominal heart. The most common ostia allow the flow of hemolymph into the dorsal vessel and contain valves to prevent backflow. These are called incurrent ostia. Some insects have openings without valves through which the hemolymph moves constantly; these are called excurrent ostia and are common, for example, on the ventral side of the grasshopper heart, which also has a full complement of paired incurrent ostia associated with each segment in the abdomen.

Occasionally, insects have structures that branch out from the dorsal vessel. In the American cockroach (*Periplaneta americana*) and some other orthopteroid insects (e.g., crickets and mantids) there are paired segmental vessels diverging from the heart laterally. In the cockroach, these vessels are simple sacs of connective tissue and have no inherent musculature, thus providing a simple channel to the lateral aspects of the pericardial sinus in the middle segments of the abdomen. These specialized vessels ensure lateral perfusion of the pericardial sinus in moderate to large insects. Lateral tubes and vessels are not known in small insects.

The dorsal vessel is composed of muscle cells (collectively called the myocardium) that lie sometimes as opposed pairs and sometimes as spiral bands to form the cylinder of the dorsal vessel. The myocardium in all insects is spontaneously active, usually beginning in the embryonic stages. This type of heart is termed myogenic because the electrical activity underlying contractions arises in the myocardium itself. This is in contrast to a neurogenic heart present in, for example, crustaceans such as crabs and lobsters, in which a barrage of nervous impulses drives the heartbeat from a discrete cardiac ganglion center.

In the pupae (and sometimes in resting adults) of holometabolous insects, the heartbeat exhibits reversal during which peristaltic contractile waves first push hemolymph from back to front (anterograde peristalsis) then at other times exclusively from front to back (retrograde peristalsis). Because heartbeat reversal is characteristic of even highly mobile mosquito pupae, reversal of hemolymph flow is thought to be an adaptation to an insect body that is rigid (the front end of a mosquito pupa is a rigid structure in which internal tissue and organs are undergoing drastic changes in shape to form adult structures, including the wings).

Nervous stimulation or mechanical disturbance causes the anterograde pulsations to revert to retrograde peristalsis. Because substances that block nerve impulses can cause the anterograde peristalsis to disappear, nervous signals (possibly inhibitory signals) are assumed to be responsible for alternating between the two peristalsis conditions, with retrograde being the basic condition.

Until recently, little was known about the innervation and control of heartbeat activity. Although the basal heartbeat rate of most insects is around 60 beats min^{-1} at room temperature and at rest (American cockroach and the locust, *Locusta migratoria*), the heartbeat of adult house flies (*Musca domestica*) is extremely unusual in that it fluctuates seemingly at random from over 300 beats min^{-1} to zero regardless of activity of the insect, flying or at rest.

The central nervous system of the adult house fly is composed of the brain and a thoracic ganglion mass. No ganglia are present in the abdomen. Because of this unusual anatomy, the dorsal vessel in the abdomen can be separated from all innervation from the central nervous system simply by cutting between the thorax and abdomen. After this operation, the heartbeat of the fly becomes quite regular at around 60 beats min^{-1}. This indicates that the house fly heart is innervated by both inhibitory and excitatory motor neurons from the central nervous system. Recently Ruthann

Nichols demonstrated inhibition caused by one or more neuropeptides in the fruit fly, *Drosophila melanogaster.*

VENTRAL DIAPHRAGM

The ventral diaphragm plays a prominent role in perfusing the ventral nerve cord of insects (Fig. 1). Nearly 40 years ago Glenn Richards surveyed the ventral diaphragms in insects and found that insects with a well-defined ventral nerve cord in the abdomen also had a well-developed ventral diaphragm. In contrast, insects with the ventral nerve cord condensed into a complex ganglion structure in the thorax invariably lacked a ventral diaphragm. This correlation suggests that the role of the ventral diaphragm is inexorably tied to perfusion of the ventral nerve cord in the abdomen.

The thorax of most insects is so packed with muscles involved in locomotion that other tissues are greatly reduced. Thus, the foregut is a simple tube passing through a small opening in the middle of the thorax and a well-defined ventral diaphragm (if present) is reduced. When present, the ventral diaphragm loosely defines a perineural sinus below and the perivisceral sinus above containing the gut.

In some insects, the ventral diaphragm is a strong muscular structure with a great deal of contractile activity. The activity of the ventral diaphragm is dictated by innervation from the central nervous system. In some large flying insects, the ventral diaphragm assists in hemolymph flow during thermoregulation by facilitating the removal of warm hemolymph from the hot thoracic muscles to the abdomen for cooling.

The intimate association between the ventral diaphragm in insects and perfusion of the ventral nerve cord is strengthened by considering the structure in the American and Madeira (*Leucophaea maderae*) cockroaches that takes the place of a proper diaphragm. In these two insects, four strips of muscles are attached at the back of the thorax and inserted on the ninth sternite. This structure has been called the hyperneural muscle because it does not form a true diaphragm above the ventral nerve cord and therefore is given a distinctive name. The hyperneural muscle is attached near the back of each of the abdominal ganglia, and the muscles contract slowly but not in a rhythmic order.

The hyperneural muscles are electrically inexcitable, which means that they do not contract myogenically, as the myocardium does, but instead are neurally driven by motor neurons located in the ventral ganglia. Thus each of the ventral nerve cords in these two cockroach species (*P. americana* and *L. maderae*) has its own muscle supply that pulls the ganglia back and forth along the midline of the abdomen upon demand. This entire structure is designed to increase the mixing and contact between the ganglia and the hemolymph.

DORSAL DIAPHRAGM

A cross section of the abdomen of insects reveals a pericardial sinus near the dorsal cuticle. The dorsal diaphragm can be a thin sheet of muscular tissue, or it can be fenestrated (Fig. 1). In most cases, there are muscles present in the diaphragm, which are called alary muscles because when vitally stained they give the appearance of "wings" projecting laterally from each abdominal segment of the dorsal vessel. The presence of paired alary muscles and paired ostia in each segment of the dorsal vessel in the abdomen reinforces the concept of "chambers" of the dorsal vessel in each abdominal segment.

Although mistakenly sometimes thought to play a key role in heartbeat, the alary muscles are more properly called muscles of the dorsal diaphragm. Whereas the myocardium is specialized to contract rapidly and constantly, the ultrastructure of the alary muscles is compatible with muscles that contract infrequently and slowly, having long sarcomeres and few mitochondria to provide only moderate amounts of energy.

In some insects, such as the tsetse fly and some moths, the alary muscles of the vestigial dorsal diaphragm extend from lateral cuticular attachment to join the dorsal vessel in the abdomen, turn, and travel along the dorsal vessel for some distance. Where this occurs, it is more difficult to determine the role of such alary muscles in the heartbeat.

ACCESSORY PULSATILE ORGANS

Because the circulation of hemolymph is vital to all insect tissues, several intricate structures ensure circulation of hemolymph through the appendages. Collectively, these are termed the accessory pulsatile organs (APOs), but modifications to ensure circulation in the appendages also include diaphragms and directed channels. When present, APOs occur at the bases of wings, antennae, legs, and cercal appendages at the back of the abdomen.

Early studies of the neuromusculature of the locust leg revealed a proximal bundle of muscles in the extensor tibia (jumping muscle) of the hind leg that exhibited rhythmic contraction. Amputating the hind leg of the grasshopper or locust very near the connection with the thorax, and attaching the femur to a convenient substrate with the tibia pointing straight up, demonstrates this rhythmic activity. After a delay of several minutes, the tibia will move back and forth spontaneously.

A small patch of muscles (called a "leg heart") near the coxal–trochanter–femur joint generates rhythmic pulsations thought to assist in the movement of hemolymph in the large femur of the jumping leg. To ensure hemolymph supply to the entire leg, there is a delivery route out and a collecting route back. Movement of hemolymph in an open circulatory system may be assisted by gross movement of internal organs, such as contractile activity of the Malpighian tubules and of the mid- and hindgut.

EXTRACARDIAC PULSATIONS

First described in 1971, extracardiac pulsations of insects are the simultaneous contractions of intersegmental muscles,

usually of the abdomen of insects, that cause a sharp increase in the pressure in the insect body. The amount of movement accompanying each pulse is too small to be seen, but it can be readily measured as a slight shortening or telescoping of the abdomen as measured from its tip. The extracardiac pulses should not be confused with larger overt movements of the abdomen, especially in bees and bumble bees, that accompany ventilation during times or high activity or exertion such as flight.

Either the extracardiac pulsations occur in coordination with openings of certain of the spiracles, and therefore can play a role in ventilation, or they occur when all the spiracles are tightly closed, hence affecting hemolymph movement. The extracardiac pulsations become suppressed only in quiescent stages of insect development, such as during diapause, but they can be evoked immediately upon disturbance or stimulation.

The extracardiac pulsations are driven by a part of the nervous system for which Karel Slama coined the name "coelopulse nervous system." The pressures induced by extracardiac pulsations are 100 to 500 times greater than pressures caused by contractions of the dorsal vessel and are transmitted by the hemolymph throughout the entire body of the insect, influencing hemolymph movement at some distance from the dorsal vessel and APO structures.

TIDAL FLOW OF HEMOLYMPH

To keep body weight to a minimum, large flying insects decrease the amount of water in the hemolymph. The remaining amount of hemolymph is first delivered into the thorax and directed into the wing veins. To make room for the hemolymph in the wing veins, an accompanying tracheal tube collapses as the hemolymph is pumped into the vein. Thus, the space in the wing veins is first taken up by hemolymph at the expense of air in the tracheal tube and then, as the hemolymph is pumped out, the air reenters the wing vein. Lutz Wasserthal called this periodic exchange of air and hemolymph a "tidal flow" of hemolymph.

The entire circulatory system of the insect is recruited to operate the tidal flow of hemolymph in large flying insects. Thus both the dorsal vessel and the ventral diaphragm are recruited to direct the hemolymph in the proper direction during each tidal cycle. This implies a very sophisticated control mechanism that must operate from the central nervous system.

THERMOREGULATION

Before the extracardiac pulsations were reported and before the tidal flow of hemolymph had been described in insects, Bernd Heinrich wrote about the use of the hemolymph in thermoregulation of flying insects. The optimum temperature for flight muscle contraction in many insects, such as the tobacco hornworm, *Manduca sexta,* is surprisingly high, up to 45°C. Before this moth can fly, it must warm the thorax to near this temperature, which it accomplishes by means of a series of simultaneous isometric contractions of the antagonistic pairs of flight muscles that appear to the casual observer as "shivering," or vibrations of the wings (Fig. 2).

A "thermometer" in the thoracic ganglia detects the proper temperature. When the thoracic temperature is below

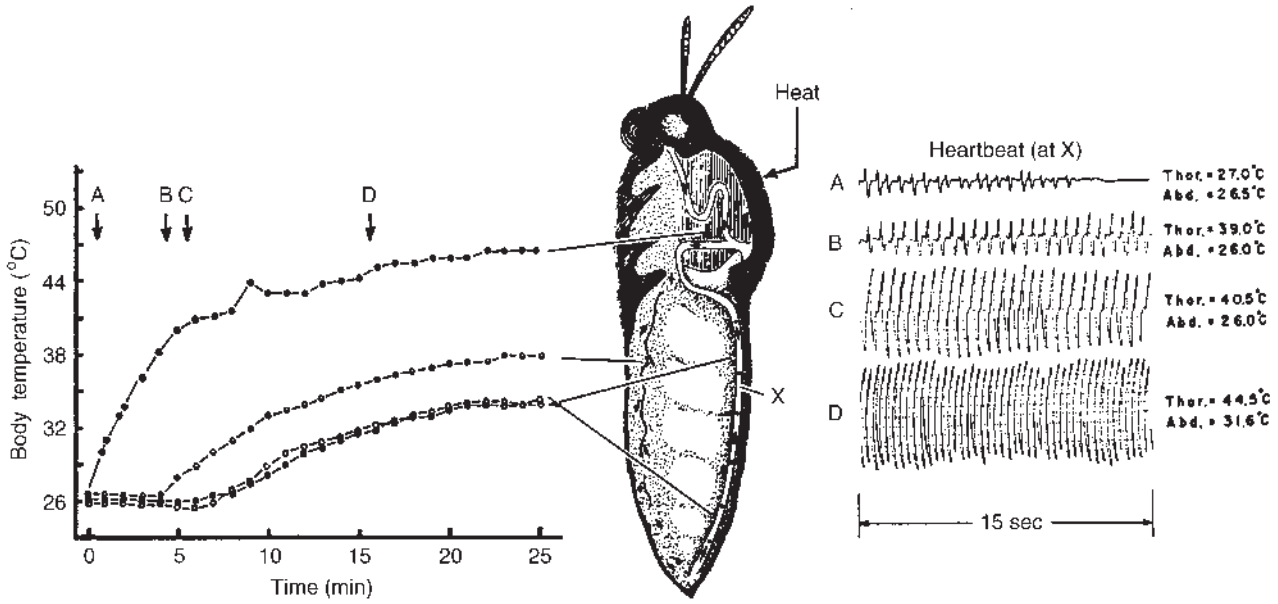

FIGURE 2 Control of thoracic temperature by central nervous control of dorsal vessel contractions during external heating of the thorax (Heat). At the optimum temperature, hemolymph is pumped at maximum frequency and amplitude through the dorsal vessel to conduct heat from the thorax to the abdomen, where is it dissipated. [From Heinrich, B. (1970). Nervous control of the heart during thoracic temperature regulation in a sphinx moth. *Science* **169,** 606–607, Copyright 1970 American Association for the Advancement of Science.]

optimum, the central nervous system signals the dorsal vessel to circulate hemolymph slowly. When the thoracic temperature rises above optimum, the central nervous system brings about maximal amplitude and rate of heartbeat to drive hemolymph through the thoracic muscles. The increased hemolymph flow pulls heat away from the flight muscles in the thorax and eventually delivers hot hemolymph to the abdomen, where the heat is dissipated. Then relatively cool hemolymph is redelivered to the thoracic muscles by the dorsal vessel, completing the thermoregulation cycle.

The warm hemolymph is then delivered to the head and percolates back past the ventral ganglia in the thorax to the abdomen, where the heat is dissipated. The cooler hemolymph is then delivered again to the thorax. The dorsal vessel and the very strong ventral diaphragm in the tobacco hornworm act together to move hemolymph in this analogy to an automobile radiator. When the thorax is too warm, both the amplitude and the frequency of heartbeat contractions are increased, and the rate of delivery of hemolymph increases. When the thorax is too cool, amplitude and frequency of contraction of the dorsal vessel are decreased. The activity of the ventral diaphragm acts in concert with that of the dorsal vessel.

Thermoregulation of the flight muscles of the tobacco hornworm implies a sophisticated nervous control. The overall nervous control can be easily demonstrated by severing the ventral nerve cord between the thorax and abdomen. When this is done, the moth can no longer thermoregulate because the feedback loop of temperature detection by the thoracic ganglia has been destroyed, and control over ventral diaphragm and dorsal vessel contractions has been lost. An extreme form of modified circulatory system to accommodate thermoregulation is shown in Fig. 3.

AUTONOMIC NERVOUS SYSTEM

The tidal flow of hemolymph, the extracardiac pulsations, heartbeat reversal, and thermoregulation all imply a very sophisticated control of circulation by the central nervous system. The central nervous system also plays a role in regulation of the respiratory system. It seems increasingly clear that the activities of circulatory and respiratory systems are coordinated by the central nervous system, perhaps to an extent not fully appreciated, but strongly implied by the tidal flow of hemolymph concept of Lutz Wasserthal.

It would be convenient and satisfying to be able to point out a particular part of the central nervous system and related peripheral nerves in insects that might comprise this regulatory system; however, outside existing evidence that the meso- and/or metathoracic ganglia play a major role in certain of these functions, entomologists know of no such discrete structure or structures, possibly because these interregulatory functions have been undertaken by different parts of the nervous system in different insects. It is known that insects have a number of regulatory mechanisms that can be recruited to achieve such control, from motor and sensory neurons to neurosecretory neurons to neurohormonal organs located all over the insect hemocoel.

See Also the Following Articles

Hemolymph • Immunology • Respiratory System • Thermoregulation

Further Reading

Ai, H., and Kuwasawa, K. (1995). Neural pathways for cardiac reflexes triggered by external mechanical stimuli in larvae of *Bombyx mori. J. Insect Physiol.* **41,** 1119–1131.

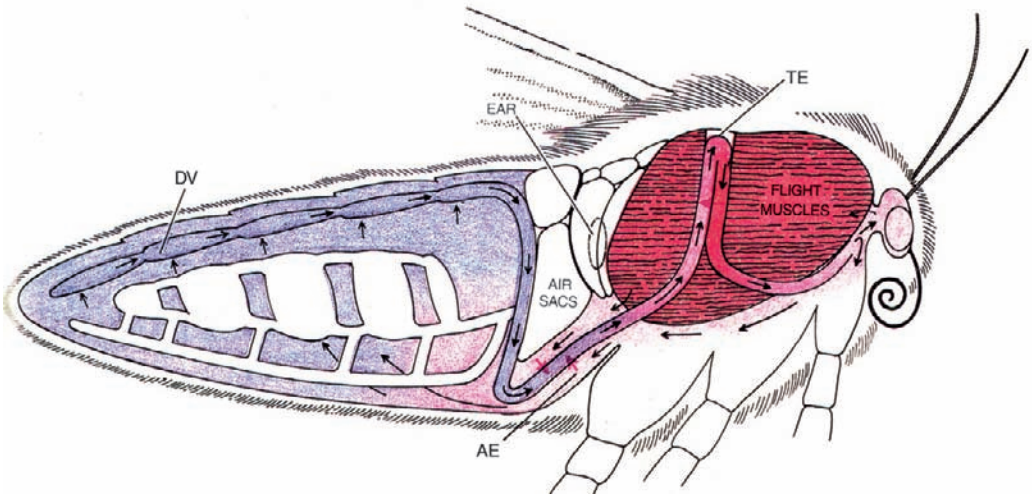

FIGURE 3 Circulation under precise nervous control of the dorsal vessel (DV) keeps the flight muscles (red) warm in the noctuid winter moth with the aid of two strategically placed countercurrent heat exchangers (TE and AE). [Modified from Heinrich, B. (1987). Thermoregulation in winter moths. *Sci. Am.* **256,** 104–111. Illustration by Patricia J. Wynne.]

Heinrich, B. (1971). Temperature regulation in the sphinx moth, *Manduca sexta.* II: Regulation of heat loss by control of blood circulation. *J. Exp. Biol.* **54,** 153–166.
Jones, J. C. (1977). "The Circulatory System of Insects." Thomas, Springfield, IL.
Miller, T. A. (1997). Control of circulation in insects. *Gen. Pharmacol.* **29,** 23–38.
Pass, G. (2000). Accessory pulsatile organs: Evolutionary innovations in insects. *Annu. Rev. Entomol.* **45,** 495–518.
Slama, K. (2000). Extracardiac versus cardiac haemocoelic pulsations in pupae of the mealworm (*Tenebrio molitor* L.). *J. Insect Physiol.* **46,** 977–992.
Wasserthal, L. T. (1996). Interaction of circulation and tracheal ventilation in holometabolous insects. *Adv. Insect Physiol.* **26,** 297–351.

Classification

see *Nomenclature and Classification*

Cockroach

see *Blattodea*

Cocoon

Frederick W. Stehr
Michigan State University

A cocoon is commonly believed to be the silken protective covering within which the caterpillars of many moths and a few butterflies pupate. Other orders of insects also spin silk and form cocoons, including Siphonaptera (fleas), Hymenoptera (ants, bees, and wasps), Neuroptera (lacewings and antlions), and Trichoptera (caddisflies).

Lepidoptera cocoons can be very thick and tough, such as that of the silkworm, *Bombyx mori,* or they can be limited to a relatively few strands of silk that keep the pupa from falling or that hold materials together to form a shelter. Cocoons such as those of the silkworm are composed of a single continuous strand of silk that is unwound in commercial silk production. Other lepidopteran cocoons may also be composed of a single strand, but the strand is usually intertwined in a way that is very difficult to unwind.

There are many kinds of lepidopteran cocoons. Some are formed from substrate materials held together by silk, others are so tough and juglike that they have a special escape lid woven at the end for the emergence of the adult (slug caterpillars, Limacodidae). Some lepidopterans such as the cecropia moth, *Hyalophora cecropia* (Saturniidae), spin an elongate cocoon with a one-way escape hatch at the end. The tent caterpillars, *Malacosoma* (Lasiocampidae), spin a complete cocoon that is infused with a yellow or whitish powder that can be irritating to predators. Emergence is accomplished by secretion of a fluid that softens and dissolves part of the cocoon. Many woollybear caterpillars (Arctiidae) incorporate the caterpillar's spiny setae (which can also be irritating) into the cocoon.

Emergence from a pupa in the soil is usually by the adult crawling upward and often occurs after a rain softens the soil. Other adults emerge by cutting or forcing their way through the wall of the cocoon with sharp structures on the pupal head that are moved by the adult inside the pupal skin. Generally, if the pupa is capable of forward movement, it only partially protrudes from the cocoon, because it is held in place by forward-projecting spines near its rear that anchor it within the cocoon, enabling the adult to pull out of the pupal skin more easily.

Caterpillars that live in bags or cases attach the bag firmly with silk to the host (or some other substrate) and pupate within these structures. Winged male bagworms (Psychidae) emerge from the bag but the females of most species are wingless and remain inside the bag where they are fertilized by the males' insertion of their long abdomens into the open end of the bag. Females usually lay their eggs inside the bag and die.

Hymenopteran cocoons are highly variable in appearance, ranging from the tightly spun cocoons of some parasitic ichneumonid wasps that are suspended on a long silken thread to the more loosely spun cocoons of some braconid wasps such as those found clustered on the outside of caterpillars (Fig. 1). Ants also spin cocoons. When ant nests are opened, many ants rush off carrying larvae and cocoons (the larger, smoother objects that look like short hotdog buns and are thought to be eggs by some). Sawfly larvae spin cocoons that are similar to ant cocoons, and most bee and wasp larvae spin cocoons inside the cells provided by the adults.

Flea (Siphonaptera) larvae spin cocoons covered with debris in areas where the larvae have been feeding. Lacewing larvae

FIGURE 1 Multiple parasitic wasp cocoons (Braconidae) formed by larvae after emerging from the slug caterpillar, *Lithacodes fasciola.* (Photograph by David J. Wagner, University of Connecticut.)

(Neuroptera: Chrysopidae) spin tight, egg-shaped cocoons in some snug spot near where they have been feeding. Caddisfly larvae (Trichoptera) are aquatic; many use their cases made of silk or materials spun together with silk as cocoons by attaching them to the substrate and sealing off the ends.

See Also the Following Articles
Caterpillar • Larva • Pupa and Puparium

Codling Moth

Stephen C. Welter
University of California, Berkeley

Codling moth, *Cydia pomonella* (Fig. 1), is a key insect pest of apples, pears, and walnuts nearly worldwide. Codling moth acquired its name because of its attacks on a greenish, elongated English cooking apple referred to as a "codling" apple. The codling moth was noted to be a pest in Europe as early as 1635, well before the development of large-scale planting of apples or pears; the first report of codling moth in the United States was in 1750.

IMPORTANCE AND DAMAGE

Codling moth spread from its presumed site of origin in Eurasia, along with the cultivation of its hosts, particularly apples; other agricultural hosts include quince, apricot, plum, peach, and nectarine, although codling moth is less of a problem in these crops. Damage results from shallow feeding wounds that cause scarring of the fruit, from direct feeding damage to the fruit pulp or seeds, or from indirect contamination of the fruit by larval feces.

LIFE HISTORY

Codling moth has between one and four generations per year, depending on temperature and other climatic factors.

FIGURE 1 Codling moth *(Cydia pomonella).* (Photograph by Mark Skevington, Whetstone, Leicestershire, U. K.)

Adult codling moth females lay single eggs on the fruit or leaves of their host. Although some larvae feed on the surface of the fruit, most larvae bore directly into the fruit within 24 h, continue to feed briefly under the surface of the skin, and then move through the flesh of the fruit to feed on the seeds. There are five larval instars. Mature larvae exit the fruit and most frequently pupate under the bark. As daylength shortens with the approach of winter, mature fifth instars spin overwintering cocoons under bark, in debris, or wood fruit. The mature larvae spend the winter in a state of arrested development until spring conditions trigger development.

MANAGEMENT

Insecticides

Management of codling moth populations in orchards traditionally has relied on synthetic pesticides. Although newer, more selective pesticides provide effective control of codling moth, older pesticides have been associated with nontarget environmental and human health risks. In addition, the evolution of resistance in codling moth to many different groups of insecticides (the chlorinated hydrocarbons, organophosphates, carbamates, pyrethroids, and newer insect growth regulators) has made the long-term reliance on these compounds more problematic.

Pheromone Mating Disruption

A recent alternative to insecticides relies on the disruption of codling moth mating using sex pheromones. Artificial emitters of the female attractant interfere with the male's ability to find females. The most common dispensers are variations on different reservoir designs, which are tied or placed in orchard tree canopies. Synthetic pheromone from these emitters then permeates the orchard canopies. Although the exact mechanisms explaining this approach are unclear, program efficacy has been demonstrated in almost all growing regions of the world. However, mating disruption is often not efficacious initially in orchards with high pest densities, so that some use of conventional insecticides may be required. Mating disruption has been widely implemented in some areas such as the western United States, where up to 40 to 50% of the pear and apple acreage (e.g., in northern California) uses this technique.

Biological Control

Although management of codling moth based on control by natural enemies has proven elusive, significant reductions in population densities have been made by using both native and introduced natural enemies of codling moth. One of the more thoroughly studied natural enemies in North America, Europe, or the former Soviet Union is the *Trichogramma* egg parasitoid. Large numbers of these minute wasps are periodically released into an orchard to seek out and kill the

eggs of codling moth. The eggs of *Trichogramma* are laid into the eggs of codling moth; the death of the egg occurs as the *Trichogramma* larvae develop. Other parasitoids that attack larval or pupal stages have also been introduced or accidentally released into new regions, including *Pimpla pterelas* and *Ascogaster quadridentata.* However, parasitism levels rarely reach more than 5%, except for some regions in central Asia where levels are as high as 50%. Nonspecialized parasitoids of egg, prepupal, or pupal stages comprise the majority of the natural enemies in North America; more specialized larval parasitoids are found in Europe and Central Asia.

General predators such as birds, predaceous insects, and spiders have been reported as suppressive agents of codling moth; these include woodpeckers, carabid beetles, and mirid bugs.

Although codling moth is susceptible to several diseases, a granulosis virus that can be applied in water, similar to insecticide applications, can cause significant reductions in codling moth densities. However, problems with production, formulation, and the short residual activity of the virus restrict its usage. Some reductions in codling moth populations also have been associated with applications of the bacterium *Bacillus thuringiensis,* but its efficacy is limited.

See Also the Following Articles

Agricultural Entomology • Biological Control of Insect Pests • Integrated Pest Management • Pheromones

Further Reading

Aliniazee, M. T., and Croft, B. A. (1999). Biological control in deciduous fruit crops. *In* "Handbook of Biological Control: Principles and Applications of Biological Control" (T. S. Bellows and T. W. Fisher, eds.), pp. 750–753. Academic Press, San Diego.

Barnes, M. M. (1991). Codling moth occurrence, host race formation, and damage. *In* "Tortricid Pests: Their Biology, Natural Enemies, and Control" (L. P. S. v. d. Geest and H. H. Evenhuis, eds.), Vol. 5, pp. 313–328. Elsevier, Amsterdam.

Calkins, C. O. (1998). Review of the codling moth areawide suppression program in the western United States. *J. Agric. Entomol.* **15,** 327–333.

Cardé, R. T., and Minks, A. K. (1995). Control of moth pests by mating disruption: Successes and constraints. *Annu. Rev. Entomol.* **40,** 559–585.

Cross, J. V., Solomon, M. G., Babandreier, D., Blommers, L., Easterbrook, M. A., Jay, C. N., Jenser, G., Jolly, R. L., Kuhlmann, U., Lilley, R., Olivella, E., Toepfer, S., and Vidal, S. (1999). Biocontrol of pests of apples and pears in northern and central Europe. 2. Parasitoids. *Biocontrol Sci. Technol.* **9,** 277–314.

Coevolution

Douglas J. Futuyma
State University of New York, Stony Brook

The term "coevolution" usually refers to the joint evolution of two or more species or genomes, owing to interactions between them. These interactions include interspecific competition, mutualism, and interactions between "consumers" and "victims" (encompassing predator/prey, herbivore/plant, and parasite/host relationships), as well as other interactions such as mimicry. Although it is often difficult to prove that true coevolution has occurred, it has probably had profound effects on the diversity of organisms and the evolution of their characteristics. Insects have figured prominently in research on coevolution.

CONCEPTS OF COEVOLUTION

Coevolution refers to several processes. One possible form of coevolution is cospeciation, the coordinated branching (speciation) of interacting species (such as host and parasite). To the extent that this has occurred, concordant (or matching) phylogenies of host and parasite clades (or evolutionary lines) would be expected. Cospeciation might be caused by the interaction between species, but it could also result from a joint history of geographic isolation, assuming that divergence and reproductive isolation evolve at similar rates in the two groups. Concordance of the two phylogenies implies a longer history of association, and of opportunity for reciprocal adaptation, than, for example, when parasites or symbionts have frequently switched from one host to another. Host switching can be inferred from certain patterns of discordance between host and symbiont phylogenies. Both cospeciation and host switching have been revealed in herbivorous insects, symbiotic bacteria, and parasites. For example, lice associated with gophers and with certain seabirds appear to have cospeciated to a considerable extent, and endosymbiotic, mutualistic bacteria *(Buchnera)* display almost complete phylogenetic concordance with their aphid hosts, from the family level down through relationships among conspecific populations.

In its most frequent usage, coevolution refers to genetic changes in the characteristics of interacting species resulting from natural selection imposed by each on the other—i.e., reciprocal adaptation of lineages to each other. Such changes are referred to as specific or pairwise coevolution if the evolutionary responses of two species to each other have no impact on their interactions with other species. Diffuse or guild coevolution occurs when the genetic change in at least one species affects its interaction with two or more other species. For example, cucumber genotypes with high levels of the chemical cucurbitacin have enhanced resistance to mites but also enhanced attractiveness to cucumber beetles; this is an instance of a negative genetic correlation in resistance. Early season attack by flea beetles makes sumac plants more susceptible to stem-boring cerambycid beetles, and so resistance to the former would also reduce the impact of the latter.

In one of the seminal papers on coevolution, Ehrlich and Raven postulated in 1964 what has since been named "escape and radiate" coevolution—a process in which evolutionary

changes temporarily reduce or eliminate the ecological interactions between species. Applying this concept to plants and herbivorous insects, Ehrlich and Raven postulated that in response to selection by herbivores, a plant species may evolve new defenses that enable it to escape herbivory and to flourish so well that it gives rise to a clade of descendant species with similar defenses. At some later time, one or more species of herbivores adapt to the defenses and give rise to an adaptive radiation of species that feed on the plant clade. In this scenario, the evolutionary diversification of both herbivores and plants is enhanced by their interactions.

Despite a common misconception, coevolution need not promote stable coexistence of species, and it certainly need not enhance mutual harmony. For example, parasites may evolve to become more virulent or less, depending on their life history. The Darwinian fitness of a genotype of parasite is measured by the average reproductive success of an individual of that genotype. Extracting more resource from a host, thereby reducing its chance of survival, often enhances the parasite's reproductive success, as long as the parasite individual, or its offspring, can escape to new hosts before the current host dies. Evolution of the parasite, by individual selection, may result in such high virulence that the prey or host population is extinguished. Extinction of prey populations does not alter the relative fitnesses of individual parasite genotypes and so does not select for reduced virulence. However, group selection may favor lower virulence or proficiency. If populations of more virulent parasites suffer higher extinction rates than less virulent populations, the species as a whole might evolve lower virulence. Although individual selection is likely to be stronger than group selection in most species, the population structure of some parasites may provide an opportunity for group selection to affect their evolution.

COEVOLUTION OF COMPETING SPECIES

Darwin argued that competition is an important agent of natural selection for adaptation to different habitats or resources by different species. Indeed, a common theme in community ecology is that coexisting species differ in food or other components of their ecological niches and that such differences are ordinarily necessary for species to coexist in the long term.

Quantitative genetic models of the evolution of competitors assume that in each of two or more species, a heritable, continuously varying trait, such as an animal's body size or mouth size, determines the mean and variance of resources (e.g., size of prey) consumed. Because competition for limiting resources decreases an individual's fitness, genotypes of species 1 that use a resource different from that used by species 2 are likely to increase in frequency, so that the mean phenotype (and resource use) shifts away from that of the other species. At evolutionary equilibrium, the species will still overlap in resource use to a greater or lesser extent, depending on the abundance of different resources, but the variance in each (the breadth of resources used) is likely to be lower than in a solitary species. Three or more species may evolve differences from each other in phenotype (e.g., size) and resource use. Such coevolutionary changes should promote coexistence. However, if competition between species is asymmetrical (e.g., if larger individuals reduce the fitness of smaller ones more than the converse), a species may converge toward the other, use its resources, and "chase" it to extinction.

Considerable evidence, mostly from vertebrates, supports this coevolutionary theory. For example, closely related sympatric species of Darwin's finches, woodpeckers, and some other animals each use a narrower variety of food types or microhabitats than do species that occur singly on islands. Evidence for evolutionary response to competition is provided by some instances of character displacement—a greater difference between two species where they occur together than where each occurs alone. Some lakes left by retreating glaciers in northwestern North America are inhabited by a single species of stickleback fish (*Gasterosteus aculeatus* complex), which feeds both near the bottom and in open water. In other lakes, two coexisting species have evolved. Relative to the solitary form, the coexisting species have diverged and specialized in morphology and behavior: one feeds on benthic prey and the other on plankton. Experiments have shown that competition among similar phenotypes reduces growth of juveniles more than among dissimilar phenotypes. In one of the few cases of ecological character displacement reported for insects, sympatric populations of two species of rhinoceros beetles (Scarabaeidae: *Chalcosoma*) overlap less in altitudinal range and differ more in size than allopatric populations. However, it has not been shown that these differences stem from competition for resources.

Coevolution of competitors may explain some patterns in community structure. For example, differences in body size or trophic structures among sympatric pairs of species of bird-eating hawks, carnivorous mammals, and seed-eating Galapagos finches are greater than if the species had been assembled at random. In a remarkable example of coevolutionary consistency, ecologically and morphologically equivalent sets of species of *Anolis* lizards have evolved independently on each of the four islands of the Greater Antilles.

COEVOLUTION OF CONSUMERS AND VICTIMS

We might expect predators and prey, herbivores and plants, and parasites and their hosts to evolve in an "arms race," whereby the victim evolves ever greater resistance, defense, or evasion, and the consumer evolves ever greater proficiency in finding and attacking the victim. However, the coevolutionary dynamics may be more complex than this, because of factors such as costs of adaptation and diffuse coevolution.

Considerable evidence supports the assumption that greater elaboration of a defensive or offensive feature imposes costs resulting from the character's interfering with another function or simply from the energy required for its development.

The population dynamics and the course of character evolution depend on many parameters and are often sensitive to starting conditions. An indefinitely extended arms race or escalation of the two species' characters is unlikely, because the cost of a sufficiently elaborated character eventually exceeds its benefit. Rather, the characters of both the prey and the predator may evolve to an intermediate stable state. Perhaps counterintuitively, species may become less proficient in attack or defense; for instance, a prey species may evolve a lower level of defense if it is so well defended that the predator becomes rare and thus becomes a weaker agent of selection than the energetic cost of defense. In some models, both the population densities and the character means of both species may change indefinitely, either in stable limit cycles or chaotically, and may even result in extinction.

When the consumer feeds on multiple species of victims, or a victim is attacked by multiple consumer species, diffuse coevolution may affect the outcome. For example, if there exists a negative genetic correlation between a host's resistance to different species of parasites, then resistance to each carries a "cost," selection will vary in time and space, depending on the relative abundance of the two parasites, and resistance to each parasite is constrained. Diffuse coevolution can be very difficult to document and might often be sluggish. Because prey species have characteristics (e.g., cryptic coloration, distastefulness, speed of escape) that provide protection against many species of predators, and predators likewise have characteristics that enable them to capture and handle many prey species, changes in the relative abundance of different predators (or prey) may not greatly alter selection. During the "Mesozoic marine revolution," lineages of crustaceans and fishes capable of crushing hard shells evolved, and many groups of molluscs evolved features (e.g., thicker shells, spines) that made them more difficult to crush. Surely these changes reflect diffuse coevolution, but our inability to ascribe changes in any one species to changes in any other one species makes it hard to discern a coevolutionary process.

Predators and Prey

Geographic variation in the identity and strength of interactions among species provides some of the best evidence of coevolution. For example, the shape of the cones of lodgepole pine *(Pinus contorta)* differs among populations, depending on whether its major seed predator, the red squirrel *(Tamiasciurus hudsonicus),* is present or absent. In mountain ranges without red squirrels, red crossbills (*Loxia curvirostra* complex) are abundant seed predators. In these areas, the pine has evolved modifications of the cone that reduce seed extraction by this species of bird, and the shape and size of the crossbill's bill have evolved to enhance seed extraction. Such evidence of coevolution, however, is rare, compared with evidence of unilateral adaptation. For example, Mediterranean populations of the braconid parasitoid *Asobara tabida* have higher "virulence" (capacity to survive host defenses) than northern European populations. Although one of its hosts, *Drosophila melanogaster,* shows a somewhat parallel geographic pattern in defense, the cline in *Asobara* appears to be most parsimoniously explained not by coevolution, but by the fact that *D. melanogaster,* its major host in the south, has stronger defenses than the major northern host, *D. subobscura.*

Parasite–Host Interactions

The evolution of interactions between hosts and parasites (including pathogenic microorganisms) can differ from predator–prey interactions in several respects. Whereas improvement in a predator or prey trait (such as size or fleetness) is likely to enhance fitness regardless of the specific genotype in the opponent species, parasite–host interactions are more likely to be affected by "gene-for-gene" interactions, in which each allele for host resistance is matched by a parasite "virulence allele" that enables the parasite to overcome resistance. Such gene-for-gene relationships have been described for several plant–fungus interactions and for the relationship between the Hessian fly (Cecidomyiidae: *Mayetiola destructor*) and resistant genotypes of wheat. Selection in gene-for-gene systems may be frequency-dependent: as a parasite allele that matches the most common host allele increases in frequency, rare host alleles acquire a selective advantage by conferring resistance against most of the parasites and so increase in frequency and initiate selection for a currently rare virulence allele. The genetic composition of local populations is likely to differ at any one time, because these oscillatory genetic dynamics may be out of phase unless the populations are connected by high gene flow. Geographic variation in genetic composition has been reported for trematodes and snails, trematodes and fish, microsporidians and *Daphnia,* and fungal parasites and plants. In most of these parasite–host pairs, populations of the parasite are best adapted to their local host populations, suggesting that the parasites adapt faster than their hosts.

The fitness of a parasite genotype may be approximately measured by the number of potential hosts it infects, compared with other genotypes. Often, the rate of transmission to new hosts is proportional to the parasite's reproductive rate, which in turn often (though not always) determines the parasite's virulence to the host. For example, the probability that progeny of a virus are transmitted by a mosquito is a function of the density of viral particles in the host's blood. However, the probability of transmission is reduced if the host dies too soon, i.e., if the parasites die before transmission. Such conflicting factors result in an evolutionary equilibrium level of virulence that is determined by several factors, especially the mode of transmission. If transmission is "vertical," i.e., only to the offspring of infected individuals, then parasite

fitness is proportional to the number of surviving host offspring, and selection favors benign, relatively avirulent parasite genotypes. If transmission is "horizontal," i.e., among hosts of the same generation, the equilibrium level of virulence is likely to be higher, because (a) an individual parasite's fitness does not depend on successful reproduction of its individual host and (b) the likelihood is higher that an individual host will be infected by multiple parasite genotypes that compete for transmission to new hosts. As predicted by this theory, among species of nematodes that parasitize fig wasps (Agaonidae), those that are mostly horizontally transmitted cause a greater reduction of their hosts' fitness than those that are vertically transmitted.

Herbivores and Plants

Most of the many thousands of species of herbivorous insects are fairly or highly specialized, feeding on closely related species—sometimes just a single species—of plants. At a proximal level, this specificity is largely the result of behavioral responses to plant features, especially the many "secondary chemicals" that distinguish plant taxa. Insects often react to compounds in nonhost plants as deterrents to oviposition or feeding and to certain compounds in host plants as stimulants. Phytochemicals may not only deter feeding but also reduce insect fitness by acting as toxins or interfering with digestion. It is generally thought that chemical and other differences among plants select for host-specificity in insects, on the supposition that physiological costs impose trade-offs among adaptations to different plant characteristics; however, only a minority of genetic and physiological studies has supported this hypothesis. Other proposed advantages of host specificity include use of specific plants as rendezvous sites for mating, greater efficiency of finding hosts, and predator escape by several means, such as sequestering defensive plant compounds.

Phylogenetic studies show that associations between some insect clades and plant clades are very old, often dating to the

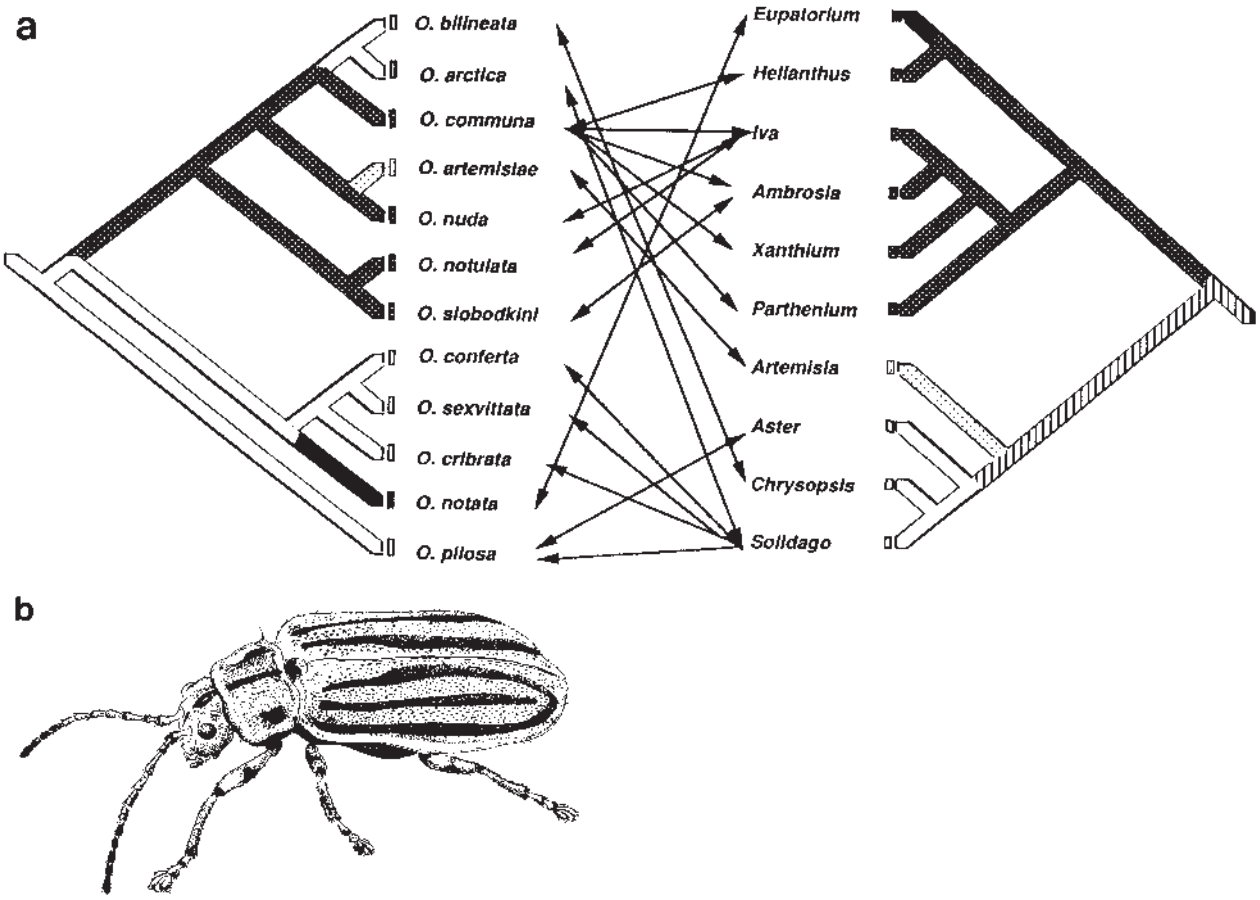

FIGURE 1 (a) Phylogenies, based on DNA sequence data, of leaf beetles in the genus *Ophraella* (left) and their host plants (right). Arrows join each beetle species to its host plant. Different shading patterns represent the four tribes of Asteraceae into which the host plants fall; the shading of branches is a parsimonious inference of the tribes with which ancestral *Ophraella* lineages were associated. Note that most host shifts associated with beetle speciation have been between plants in the same tribe. The incongruence between the phylogeny of the insects and that of their host plants is one of several indications that the beetles and plants did not cospeciate. These plant lineages represent only a few of the tribes of Asteraceae and of the genera within each tribe. [After D. J. Funk *et al.* (1995). *Evolution* **49**, 1008–1017. The Society for the Study of Evolution.] (b) The leaf beetle *O. sexvittata,* which feeds on *Solidago* species, tribe Astereae. (Original illustration by author.)

early Tertiary and in some cases to the Cretaceous or even Jurassic. Nevertheless, only a few instances of cospeciation and phylogenetic concordance have been described. In most cases, much of the speciation within an insect clade has occurred after the host plants diversified, but new species have shifted to plant species closely related to the ancestral host (Fig. 1). That these host shifts have been facilitated by chemical similarity of related plants is supported by instances in which phylogenetic relationships among insect species (e.g., *Blepharida* flea beetles, melitaeine butterflies) more closely match the hosts' chemical similarities than phylogenetic relationships. Patterns of genetic variation in the ability of host-specific *Ophraella* leaf beetles to feed and develop on nonhost plants, all within the Asteraceae, indicated greater genetic potential to adapt to those plants that were most closely related to the insect's normal host.

Although physiological, morphological, behavioral, and phenological adaptations of insects to host plants are many and obvious, demonstrating that plant characters have evolved because of selection for their defensive functions has been more difficult. Many chemical and morphological features of plants have the effect of reducing attack or damage by some or many insect species, but some authors have argued that they actually evolved for physiological reasons or as defenses against mammalian herbivores rather than insects. However, both phytochemicals (e.g., furanocoumarins) and morphological features (e.g., trichomes) have been shown to determine fitness differences among genotypes due to their effect on insect herbivores, and the distribution of many plant compounds among tissues conforms to what we should expect if they were adaptively deployed defenses. Still, there have been few demonstrations of adaptive geographic variation in plant defenses in relation to the abundance or identity of particular herbivorous insects. In one of the few examples of probable coevolution at the population level, populations of wild parsnip *(Pastinaca sativa)* have diverged in their profile of toxic furanocoumarins, and parsnip webworms *(Depressaria pastinacella)* are adapted to their local host population.

Ehrlich and Raven's escape-and-radiate model of coevolution between plants and herbivorous insects has found some support. Most lineages of plants that have independently evolved latex or resin canals (potent deterrents to most insects) are richer in species than their canalless sister groups, supporting the hypothesis that new plant defenses enhance the rate of diversification. Likewise, herbivorous clades of insects are generally more diverse than their nonherbivorous sister groups. Clades of phytophagous beetles that are thought to be primitively associated with gymnosperms have fewer species than sister taxa that have shifted to angiosperms, perhaps because the latter are so very diverse. The diversity of several moth taxa that feed on Apiaceae with presumably "advanced" chemical defenses is greater than that of those that feed on Apiaceae with "primitive" defenses, paralleling the differences in plant diversity, but phylogenetic analysis is needed to confirm that the diversification rate has been enhanced by novel plant defenses and insect counteradaptations.

EVOLUTION OF MUTUALISM

In mutualistic interactions between species, each species uses the other as a resource. That is, each exploits the other, and the degree of exploitation may determine whether an interaction is mutualistic or parasitic. Mutualisms include interactions both between free-living organisms, such as plants and pollinating animals, and between symbionts, one of which spends most of the life cycle on or in the other. Microbes are partners in many symbiotic mutualisms. Mutualists often have adaptations for encouraging the interaction or even nurturing the associate, such as foliar nectaries in plants, which attract ants that defend the plants against herbivores, or the root nodules of legumes, which house and nourish nitrogen-fixing rhizobial bacteria. In some intimate symbioses, the symbiont functions as an organ or organelle, as in the case of host-specific bacteria that reside within special cells in aphids and supply essential amino acids to their host.

For each mutualist, the interaction has both a benefit and a cost. Legumes, for example, obtain nitrogen from rhizobia, but expend energy and materials on the symbionts. Excessive growth of the rhizobia would reduce the plant's growth to the point of diminishing its fitness. Likewise, excessive proliferation of mitochondria or plastids, which originated as symbiotic bacteria, would reduce the fitness of the eukaryotic cell or organism that carries them. Thus, selection will always favor protective mechanisms to prevent overexploitation by an organism's mutualist. Whether selection on a mutualist favors restraint depends on how much an individual's fitness depends on the fitness of its individual host. When a mutualist can readily move from one host to another, as pollinating insects can from plant to plant, it does not suffer from the reproductive failure of any one host, and selfishness or overexploitation may be favored. For example, many pollinating insects "cheat." The larvae of yucca moths *(Tegeticula)* feed on developing yucca seeds in flowers that their mothers actively pollinated. However, several species of *Tegeticula* have independently lost the pollinating behavior, having evolved the habit of ovipositing in flowers that other species have already pollinated. Moreover, the pollinating species lay only a few eggs in each flower, so that the few larvae do not consume all the developing seeds. This reproductive restraint has evolved in response to a defensive tactic of the plant, which aborts developing fruits that contain more than a few eggs. However, the "cheater" species of *Tegeticula* circumvent the plant's defense by laying eggs after the developmental window for fruit abortion, and they lay so many eggs that the larvae consume most or all of the seeds. Deception and cheating has also evolved in some plants, such as orchids that provide no reward to the naive

bees that visit them; other orchids mimic the female sex pheromone of an insect species, the males of which effect pollination by "copulating" with the flower.

Vertical transmission of a symbiont favors restraint and reciprocal benefit, just as it favors lower virulence in parasites, because the fitness of the individual symbiont is then proportional to its host's reproductive success. This principle can explain why internal symbionts such as aphids' bacteria or corals' zooxanthellae (or eukaryotes' mitochondria) divide at rates commensurate with their host's growth. It is conceivable that hosts may evolve mechanisms to prevent horizontal transmission (mixing) of symbionts and thus maintain conditions under which "selfishness" would be disadvantageous to the symbiont. By extension, such principles explain the conditions for the evolution of coordination versus conflict among different genes in a single genome, i.e., the evolution and maintenance of integrated organisms.

CONSEQUENCES OF COEVOLUTION

Coevolution has undoubtedly had major effects on the history and diversity of life. Many of the adaptive differences among organisms—the many thousands of toxic defensive compounds in different plants, insects, and fungi, the many forms of flowers, the diverse growth forms of plants, the sometimes astonishingly specialized diets of animals—have issued from interactions among species. The numbers of species, too, may have been augmented by coevolution, as Ehrlich and Raven proposed. Coevolution among competitors can also augment the species diversity in communities, producing suites of specialized species that finely partition resources among them. In theory, such coevolution may result in ecosystem-level effects such as higher productivity and resource consumption, but the evidence on this subject is very sparse.

See Also the Following Articles

Insectivorous Plants • Parasitoids • Plant–Insect Interactions • Predation • Symbionts

Further Reading

Abrams, P. A. (2000). The evolution of predator–prey interactions: Theory and evidence. *Annu. Rev. Ecol. Syst.* **31,** 79–105.

Benkman, C. W., Holiman, W. C., and Smith, J. W. (2001). The influence of a competitor on the geographic mosaic of coevolution between crossbills and lodgepole pine. *Evolution* **55,** 282–294.

Berenbaum, M., and Zangerl, A. (1998). Chemical phenotype matching between a plant and its insect herbivore. *Proc. Nat. Acad. Sci. USA* **95,** 13473–13478.

Ehrlich, P. R., and Raven, P. A. (1964). Butterflies and plants: A study in coevolution. *Evolution* **18,** 586–608.

Farrell, B. D. (1998). "Inordinate fondness" explained: Why are there so many beetles? *Science* **281,** 555–559.

Frank, S. A. (1996). Models of parasite virulence. *Q. Rev. Biol.* **71,** 37–78.

Fritz, R. S., and Simms, E. L. (eds.) (1992). "Plant Resistance to Herbivores and Pathogens: Ecology, Evolution, and Genetics." University of Chicago Press, Chicago.

Futuyma, D. J. (1998). "Evolutionary Biology," 3rd ed., pp. 539–560. Sinauer, Sunderland, MA.

Futuyma, D. J., and Mitter, C. (1996). Insect–plant interactions: The evolution of component communities. *Philos. Trans. R. Soc. London B* **351,** 1361–1366.

Kraaijeveld, A. R., and Godfray, H. C. J. (1999). Geographic patterns in the evolution of resistance and virulence in *Drosophila* and its parasitoids. *Am. Naturalist* **153,** S61–S74.

Lively, C. M. (1999). Migration, virulence, and the geographic mosaic of adaptation by parasites. *Am. Naturalist* **153,** S34–S47.

Schluter, D. (2000). "The Ecology of Adaptive Radiation." Oxford University Press, New York.

Thompson, J. N. (1994). "The Coevolutionary Process." University of Chicago Press, Chicago.

Thompson, J. N. (1999). Specific hypotheses on the geographic mosaic of coevolution. *Am. Naturalist* **153,** S1–S14.

Wahlberg, N. (2001). The phylogenetics and biochemistry of host–plant specialization in melitaeine butterflies. *Evolution* **55,** 522–537.

Cold/Heat Protection

David L. Denlinger
Ohio State University

As poikilotherms, insects are largely at the mercy of environmental temperatures. There are a few exceptions: some moths, bees, and flies do have the capacity to elevate their body temperature by vigorously contracting their flight muscles to generate heat, and a few species of cicadas, grasshoppers, and other desert species exploit evaporative cooling to lower their body temperature on hot days. The majority of species, however, have a limited capacity to alter their body temperature. Insects survive, perform, and reproduce across a broad temperature range, but they do so with varying levels of success at different temperatures. A thermal performance curve (Fig. 1) can be generated for nearly any quantitative trait. The curve delimits the body temperature at which a certain activity can be performed (tolerance zone). The low extreme is the critical thermal minimum, and the upper extreme is the critical thermal maximum.

Construction of such a curve will demonstrate that any activity has a temperature at which performance is optimal (optimum body temperature). Characteristically, the drop in performance outside the optimum body temperature is more precipitous at the high end of the temperature scale than at the low end. When given a choice, insects will readily select temperatures at which performance is maximized. But, the extremes of the daily temperature cycle and the dominant temperatures that prevail during major portions of the year pose significant obstacles for insect growth, development, and performance. This review describes the nature of the injury inflicted by high and low temperatures and discusses the protective mechanisms used by insects to counter these forms of injury.

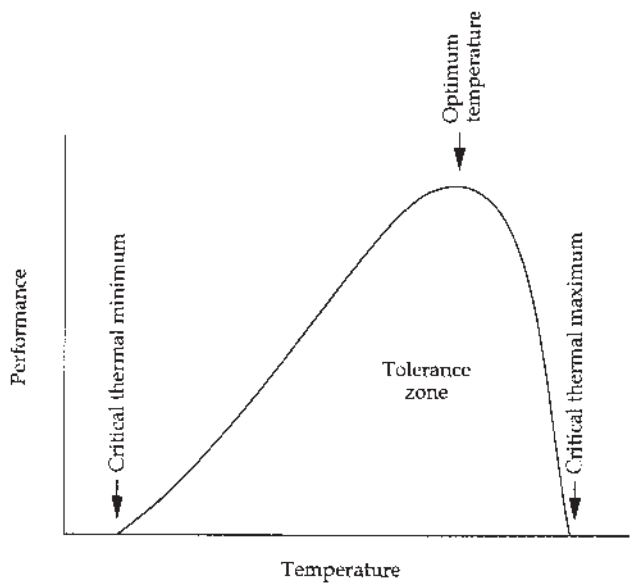

FIGURE 1 Hypothetical performance curve delineating the tolerance zone, critical thermal minimum temperature, optimum temperature, and critical thermal maximum temperature at which any quantitative trait can be performed. Note that the decline in performance above the optimal temperature is usually more precipitous at the high end of the temperature scale than at the low end.

PROTECTION AGAINST HIGH-TEMPERATURE INJURY

Heat Injury

Lethality at high temperature is a function of both temperature and time; the higher the temperature, the shorter the exposure time needed to kill the insect (Fig. 2). But, injury can manifest itself in more subtle forms at less extreme temperatures. For example, temperatures that prevent reproduction are lower than the temperatures that cause immediate mortality. At still less severe temperatures, adults are reproductively functional but emergence may be delayed or occur at the "wrong" time of day.

Heat shock can also produce developmental abnormalities known as phenocopies (developmental abnormalities resembling mutations but caused by environmental conditions), a phenomenon especially well known for the fruit fly, *Drosophila melanogaster.* Flies heat shocked during embryogenesis or metamorphosis yield interesting phenocopies with aberrant adult bristle shapes, colors, and wing formations. Which defect is observed is dependent upon the age of the fly at the time of exposure. The sensitive period for the production of each phenocopy is brief, usually less than 2 h. The various phenocopies are generated by disruption of a heat-sensitive developmental process that is specific to a particular developmental window. For example, heat shock can shut down phenol oxidase, the enzyme needed for melanin production. If heat shock is thus administered during the interval when this enzyme is needed to generate the black color normally associated with bristles, the blond-bristle phenocopy will be produced instead.

At the cellular level, a number of abnormalities are elicited in response to heat stress. These include declines in hemolymph pH; disruption of the normal pattern of protein synthesis; loss of conformational integrity of RNA, DNA, and protein; and deformation of the cellular membrane. Many cell processes are thus vulnerable to injury. Which cell process is the primary site of thermal wounding is still not clear, but two models have been proposed. One model suggests that the plasma membrane is the primary site of thermal wounding. In this model, disruption of the plasma membrane sets in motion a cascade of events involving inactivation of membrane proteins and subsequent leakage of K^+ out of the cell and movement of Ca^{2+} and Na^+ into the

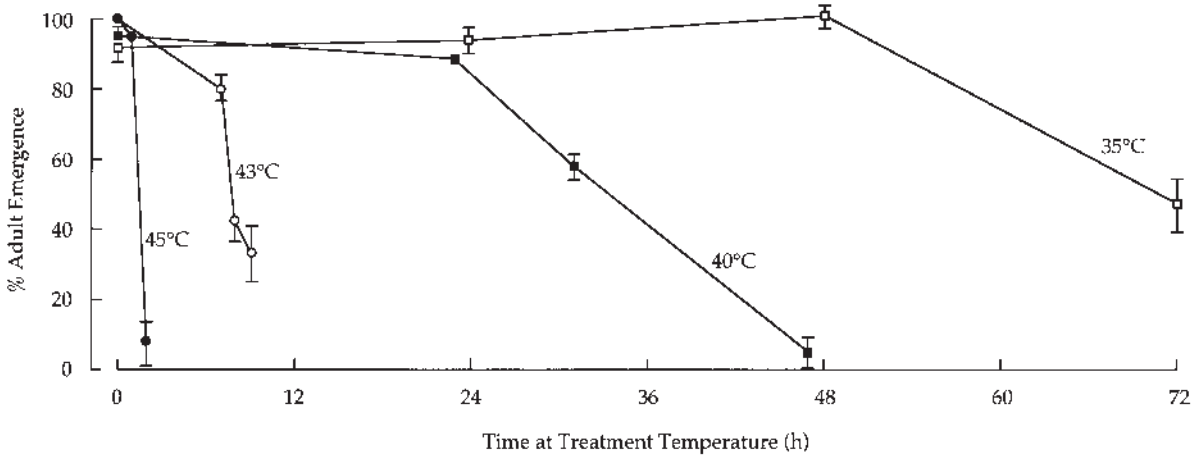

FIGURE 2 Mortality is a function of both temperature and duration of exposure, as demonstrated by the survival curves for *S. crassipalpis.* The flies were exposed to four different temperatures for various durations several days before adult emergence and survival was based on success of adult emergence. [Reproduced, with permission, from Yocum and Denlinger, (1994). Copyright Blackwell Science.]

cell. This loss of the cell's bioelectrical properties leads to a breakdown in cell metabolism, loss of homeostasis, and finally death. An alternative model also focuses on the plasma membrane but suggests that the subsequent protein denaturation is the critical cause of death. Denatured protein adheres to the chromatin and restricts enzymatic access to the DNA. The cell eventually dies as a consequence of an increase in DNA damage. But, it is also evident that an enzyme will lose its metabolic function at a fairly low level of heat stress, long before denaturation is complete. Thus, it is difficult to point to any single factor as the cause of death because high temperature adversely affects many aspects of the cell or organism's physiology simultaneously.

Thermotolerance

Thermotolerance (tolerance of high temperature) can be increased several ways. (1) Genetic adaptation: Differences in thermotolerance can be detected in diverse geographic populations, as well as in laboratory lines that have been selected for heat-shock survival. (2) Long-term acclimation: Rearing individuals for long durations at high temperatures can result in a striking increase in thermotolerance. (3) Rapid heat hardening: A brief exposure to an intermediately high temperature provides protection from injury at a more severe temperature.

Rapid heat hardening is the best studied response. Heat shock is the thermal injury caused by a sudden increase in temperature. This form of injury can be reduced dramatically if an organism is first exposed to an intermediate temperature (rapid heat hardening). For example, the flesh fly, *Sarcophaga crassipalpis,* can tolerate only a brief time at 45°C if it is transferred there directly from 25°C, but survival at 45°C is greatly extended (Fig. 3) if the flies are first exposed to 40°C for 2 h. The thermotolerance that protects against heat-shock injury is acquired quickly, within minutes, reaches a maximum within a few hours, and then decays rather slowly over several days.

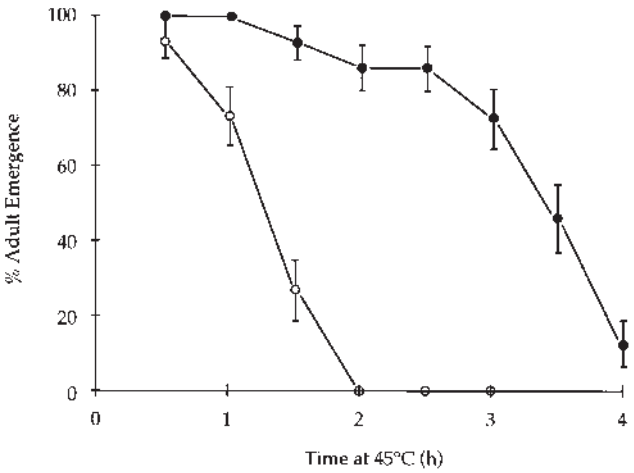

FIGURE 3 Thermotolerance can be demonstrated by the higher survival rates noted in flies that were first exposed to a moderately high temperature. In this example, based on *S. crassipalpis,* flies reared at 25°C and then transferred directly a few days before adult emergence to 45°C (open circles) survived poorly as indicated by success of adult emergence. By contrast, flies that were first exposed to 40°C (solid circles) survived exposure to 45°C much better. [Reproduced, with permission, from Chen *et al.* (1990), copyright Springer-Verlag.]

Heat-shock proteins are the best known contributors to thermotolerance. In response to heat stress, the normal pattern of protein synthesis is suppressed, and concurrently several new proteins, the heat-shock proteins, are synthesized. These proteins are classified according to their molecular mass and in *D. melanogaster* include a high-molecular-mass protein (82 kDa), members of the 70-kDa family, and small heat-shock proteins with molecular masses of 22, 23, 26, and 27 kDa. The most highly expressed heat-shock proteins, members of the heat-shock protein 70 (Hsp70) family, are highly conserved. The gene that encodes Hsp70 is over 50% identical in bacteria and *D. melanogaster.* In response to heat stress Hsp70 levels in the cell may increase more than 1000-fold.

Though heat shock was the first stress known to elicit synthesis of these proteins, it is now evident that many other forms of stress (e.g., heavy metals, alcohols, metabolic poisons, aberrant proteins, cold shock, desiccation) can elicit synthesis of these same proteins. It is thus clear that these proteins are involved in diverse stress responses.

For years the linkage between heat-shock proteins and thermotolerance was based strictly on correlation between the presence of the proteins and the expression of thermotolerance, but more recently the linkage has been strengthened with new experimental evidence. Cultured *D. melanogaster* cells and whole flies transformed with extra copies of the Hsp70 gene acquire thermotolerance more rapidly than normal cells or flies, while cells transformed with Hsp70 antisense genes acquire thermotolerance more slowly.

How do the heat-shock proteins contribute to thermotolerance? Members of the Hsp70 family function as molecular "chaperones" that facilitate the process of protein folding and assembly. Hsp70 can reduce high-temperature damage by interacting with susceptible proteins to prevent their interactions with other reactive surfaces, thus helping to maintain the integrity of proteins present in the cell.

Although heat-shock proteins have received the most attention in studies of thermotolerance, other molecules, including sugars such as trehalose and polyols such as glycerol and sorbitol, are also suspected of contributing to the protective mechanism.

Thermosensitivity

While it is widely appreciated that previous exposure to an elevated temperature can generate tolerance to high temperature (thermotolerance, as discussed earlier), it is less well appreciated that some high temperatures can decrease an

insect's ability to survive a subsequent high-temperature exposure. It is this loss of tolerance that is referred to as thermosensitivity. For example, *S. crassipalpis* appears to readily survive a 1-h exposure to 45°C, but if the fly is subjected to a second high-temperature pulse 1 day later, the effect will be lethal, even if the second pulse is considerably less severe, e.g., 35°C. Such observations suggest that some form of injury caused by the first challenge made the flies considerably more vulnerable to the second heat pulse. Without the second challenge, the initial injury can apparently be repaired, but the problem arises if the insect is challenged a second time before it has fully recovered. The temperatures that produce thermosensitivity are generally above the temperatures that generate thermotolerance.

An intriguing practical implication of thermosensitivity is that the pattern of administering a thermal stress has important consequences for an insect's survival. Two relatively modest pulses of high temperature may be just as effective in causing death as a single pulse of a higher temperature. From an economic perspective, this type of wounding may require less energy input than needed to administer a single pulse of a higher temperature.

PROTECTION AGAINST LOW-TEMPERATURE INJURY

Insects are frequently classified as being either freeze tolerant or freeze susceptible. Freeze tolerance implies that the insect can actually survive ice formation within its body. Relatively few insects have this capacity, but it is well documented in some insects such as the goldenrod gall fly, *Eurosta solidaginis.* By contrast, most insects are freeze susceptible, which means that they cannot tolerate internal ice formation. This, however, does not mean that all freeze-susceptible species can survive temperatures approaching the point at which their body will freeze. Many such species are fatally injured at temperatures well above their freezing point.

Within a single species, huge differences in cold tolerance may be evident in different stages of the life cycle. For example, in *S. crassipalpis,* the adult is the stage most susceptible to cold injury, while the pupa is least susceptible. If the pupa is in diapause, the overwintering state of dormancy, the pupa is even more cold tolerant. Characteristically, diapausing stages are highly tolerant of low temperature and are capable of withstanding far lower temperatures, and for much longer, than nondiapausing stages. In the flesh fly, diapausing pupae can tolerate temperatures of –20°C (a few degrees above their supercooling point) for many months, while nondiapausing pupae will be killed with an exposure of just a few hours to –10°C.

Supercooling and Ice Nucleation

Understanding the nature of supercooling and ice nucleation is critical for understanding the strategies used by insects to survive at subzero temperatures. One might assume that an insect will freeze when its body temperature reaches 0°C, but this does not occur. Instead, the body water supercools, a process that is enhanced in many cases by the production of cryoprotectants that dramatically reduce the freezing point and thus enable the insect to remain unfrozen at temperatures down to –20°C or lower. The temperature at which the body liquid turns to ice is called the supercooling point or the temperature of crystallization. This point is easily detected by monitoring body temperature and noting the appearance of an exotherm, the burst of heat given off by crystallization as the body water freezes. The insect body contains a number of agents that can affect the supercooling point. Cryoprotectants are capable of lowering the supercooling point, whereas ice-nucleating agents elicit the opposite response. Ice-nucleating agents act as catalysts to promote ice nucleation at higher temperatures than would occur in their absence. Formation of ice at rather high temperatures is especially common in freeze-tolerant species. In such cases it is advantageous to initiate ice formation at a rather high subzero temperature, a feature that enhances survival by slowing down the processes of ice formation. By contrast, freeze-susceptible insects exploit the use of cryoprotectants to suppress the supercooling point and thus avoid freezing.

Cold Injury

In addition to lethality, cold injury can be manifested in failure of reproduction and the appearance of developmental abnormalities. A cold shock can induce an extra molt in some species such as the greater wax moth, *Galleria mellonella.* Phenocopy defects, like those noted in *D. melanogaster* at high temperature, can also be elicited by low temperature: The incidence of aristapedia (in which antennae are transformed into legs) increases at low temperature. Sex ratios can be distorted, in some species favoring females and in others, males.

Many freeze-susceptible species are killed at temperatures well above their freezing points. The mechanism involved in this form of nonfreezing injury is poorly understood but may result from a decline in the rate of enzyme function at low temperatures or to irreversible changes in tertiary structure of critical proteins. Nonfreezing injury resulting from low temperature is frequently associated with damage to the plasma membrane. At some point chilling induces fluid to gel phase transitions in cell membranes that result in major alterations in membrane permeability and reduction in activity of membrane-bound enzymes.

Among freeze-tolerant species it is commonly assumed that survival of freezing requires that ice formation be restricted to extracellular spaces. This, however, is not always the case. Intracellular freezing does occur in some tissues such as the fat body cells of the goldenrod gall fly. Ice formation normally is initiated outside of the cell. Only water is added

to the ice lattice, thus the remaining body fluids become more concentrated. This, in turn, causes osmotic removal of water from the cells. Although mechanical injury due to ice formation can be a deleterious effect, it is likely that the primary initial stress results from cell dehydration and the accumulation of excess amounts of solutes in the body fluid. The high concentrations of solutes, particularly electrolytes, can cause protein denaturation and extreme shifts in pH that result in irreversible membrane damage.

Certain systems are more vulnerable to injury than others. The neuromuscular system appears to be particularly vulnerable. As temperatures decline insects gradually lose their ability to fly, and at still lower temperatures they lose their ability to walk. Chill coma, the point at which the insect loses its ability to walk, coincides with the temperature at which the muscles and nerves lose their electrical excitability.

The reproductive system is also quite vulnerable to cold injury. Insects may appear normal but fail to reproduce following a cold shock. Both the number of eggs produced and the fertility of the eggs may be lowered by cold injury.

Cold Hardiness

The injury caused by low temperature can frequently be mitigated by prior exposure to less severe low temperatures. Like the acquisition of thermotolerance at high temperatures, cold hardening enables an insect to survive at low temperatures that would otherwise prove lethal. Cold hardening can be either a long-term process attained after weeks or months at low temperature or a very rapid process (rapid cold hardening) invoked within minutes or hours after exposure to low temperature.

The traditional view of cold hardening depicts a slow process that gradually increases the insect's tolerance to low temperature. As seasonal temperatures drop in the autumn, many insects become progressively more cold hardy. Thus, a field-collected insect from the north temperate region evaluated in January is likely to be more cold tolerant than one collected in September. In contrast, rapid cold hardening is a very fast process that allows an insect to respond to daily changes in temperature. For example, *S. crassipalpis,* when reared at 25°C cannot survive an immediate transfer to –10°C, but if the fly is first placed at 0°C for as short a time as 10 min, it can readily survive a subsequent 2-h exposure to –10°C. The capacity for rapid cold hardening appears to be common among insects and presumably functions in enabling them to track daily and other forms of rapid temperature change.

Several diverse physiological mechanisms contribute to cold hardiness. For freeze-susceptible insects, one of the most important mechanisms involves the elimination of ice nucleators. The presence of ice nucleators limits the insect's ability to supercool; thus getting rid of potential nucleators is a critical feature of cold hardiness. Food particles present in the gut are among the most powerful ice nucleators; thus it is perhaps no surprise that many insects purge their gut prior to overwintering.

Another common cold-hardening mechanism used by freeze-susceptible insects is the synthesis and accumulation of high concentrations of low-molecular-mass polyols (glycerol, sorbitol, mannitol) and sugars (trehalose). Like a classic antifreeze, the polyols and sugars reduce the supercooling point and thus allow the insect to avoid freezing at temperatures well below 0°C. Hemolymph concentrations of polyols sometimes reach multimolar levels.

Thermal hysteresis refers to a difference between the freezing and the melting point of the body fluid. At equilibrium one would expect these two points to be nearly identical, but this relationship can be altered by thermal hysteresis proteins, also known as antifreeze proteins. Thermal hysteresis proteins depress the freezing point while leaving the melting point unchanged. This lowering of the freezing point can thus expand an insect's low-temperature tolerance. Such proteins were first discovered in cold-water, marine fish but were found more recently in several species of beetles.

Ice nucleator proteins function in a manner opposite to that of thermal hysteresis proteins. Rather than inhibiting freezing, these proteins promote freezing. Ice nucleator proteins facilitate the organization of water molecules into embryo crystals, which, in turn, seed the supercooled solution, causing freezing at relatively high temperatures. As discussed above, this is advantageous for freeze-tolerant species.

Synthesis of heat-shock proteins is a well-documented response to high temperature, but some of the same proteins are also synthesized in response to low-temperature shocks. As with heat-shock, the most prominent heat shock protein elicited by cold shock is a member of the 70-kDa family of heat-shock proteins. These stress proteins are most evident following the cold shock, thus suggesting they may play a role in the recovery process.

Insects thus have at their disposal an array of mechanisms to counter the adverse effects of low temperature. Cold hardening can entail a complex suite of responses and should not be regarded as a process driven by a single biochemical event, but species differences are likely to dictate that one particular process may be more important in one species than in another.

PRACTICAL IMPLICATIONS

Exploiting temperature for pest management is an attractive alternative to the use of pesticides. The manipulations can be safely administered and no harmful residues remain. Heat and cold treatments are emerging as the treatment of choice for quarantine treatment of fresh fruits and vegetables. Temperature treatment is especially popular in this industry because the major fumigant, methyl bromide, is being removed from the market due to its role as an ozone depleter. Soils and planting beds are being treated with heat, and both high and low temperatures are being used to treat houses and

other structures. Stored grain can be effectively protected from insects with temperature treatments, and even some field crops can be protected with novel applications of heat applied directly to the plant.

Cold storage is used extensively to increase the "shelf life" of parasitic wasps and other biological control agents, as well as the hosts on which they are reared. The cryopreservation of embryos of *D. melanogaster* and other insects is a goal sought by numerous researchers. This ability could facilitate the long-term maintenance of valuable genetic stocks and reduce the care and expense required to continuously propagate insects used for research.

Insects have a wealth of behavioral and physiological responses to counter the effects of high- and low-temperature stress, and if temperatures are to be exploited for use in integrated pest management systems, these mechanisms must be either overridden or disabled. For example, the generation of thermotolerance can be prevented by applying heat stress in a nonoxygenated environment. Combination treatments that simultaneously apply both heat and anoxia or thermosensitization (application of two temporally separated treatments at moderately high temperatures) are especially attractive because they can cause mortality with less energy input. The low temperatures that prevail during winter are frequently just a few degrees above the insect's lower limit of tolerance. Attempts to further reduce the insect's body temperature by destroying the insect's protective winter habitat offer promise. Recent discoveries of ice-nucleating bacteria and fungi that are active on insects suggest new tools for manipulating the supercooling point. The diverse protective responses operating in insects suggest a similar richness of targets that could be rendered vulnerable to heat or cold injury.

See Also the Following Articles

Diapause • Hibernation • Temperature, Effects on Development and Growth • Thermoregulation

Further Reading

Bowler, K., and Fuller, B. J. (eds.) (1987). "Temperature and Animal Cells." Society for Experimental Biology Symposium 41, Cambridge, U.K.

Chen, C.-P., Lee, R. E., Jr., and Denlinger, D. L. (1990). A comparison of the responses of tropical and temperate flies (Diptera: Sarcophagidae) to cold and heat stress: *J. Comp. Physiol. B* **160,** 543–547.

Hallman, G. J., and Denlinger, D. L. (eds.) (1998). "Temperature Sensitivity in Insects and Application in Integrated Pest Management." Westview Press, Boulder, CO.

Heinrich, B. (1993). "The Hot-Blooded Insects." Harvard University Press, Cambridge, MA.

Johnston, I. A., and Bennett, A. F. (eds.) (1996). "Animals and Temperature: Phenotypic and Evolutionary Adaptation." Society for Experimental Biology Symposium 59, Cambridge, U.K.

Lee, R. E., Jr., and Denlinger, D. L. (eds.) (1991). "Insects at Low Temperature." Chapman & Hall, New York.

Somero, G. (1995). Proteins and temperature. *Annu. Rev. Physiol.* **57,** 43–68.

Yocum, G. D., and Denlinger, D. L. (1994). Anoxia blocks thermotolerance and the induction of rapid cold hardening in the flesh fly, *Sarcophaga crassipalpis. Physiol. Entomol.* **19,** 152–158.

Coleoptera

(Beetles, Weevils, Fireflies)

James K. Liebherr
Cornell University

Joseph V. McHugh
University of Georgia

Beetle diversity so characterizes Earth that instead of telling future extraterrestrial colleagues we come from the "blue planet," we might better state that we come from the "beetle planet." Beetles comprise 25% of all described animals and plants, single-handedly making them the primary contributor to earth's biodiversity. The 350,000 described beetle species are members of largest order of life on Earth, Coleoptera.

Familiar beetles are known by various names including fireflies, ladybugs, june bugs, and weevils. The vast number of beetle species is reflected by a bewildering array of anatomical and biological diversity in the order. Coleoptera are represented in nearly all biogeographic regions and nonmarine habitats. Most adult beetles can fly; when not in use, however, the delicate flight wings are usually concealed beneath protective shell-like elytra, permitting beetles to utilize diverse resources and engage in a broad range of activities that otherwise would be restricted to either winged or wingless insects. Most beetles are herbivores, fungivores, or predaceous carnivores in the larval and adult stages. Many are considered to be serious pests of our homes, forests, crops, and stored products, whereas some beneficial species are regularly employed as biological control agents. Countless curious youngsters, including Charles Darwin, Alfred Russel Wallace, and Henry Walter Bates, have started their broader studies of biology through beetle collecting, as beetle species often are consistently found in specific sorts of habitats.

The technical name, Coleoptera, was coined by Aristotle to signify the hardened, shieldlike forewings (*coleo* = shield + *ptera* = wing). Although several other insect orders possess hardened forewings, beetles are considered to be a monophyletic assemblage based on their sum of shared evolutionary derivations that include the following:

1. A holometabolous life cycle, wherein the larval stages are developmentally separated from the adult by the pupal stage.
2. Possession of hardened forewings, called elytra, that abut medially. Flight is powered predominantly by the metathoracic wings, which are folded longitudinally and usually transversely to lie under the elytra when the beetle is walking or at rest. The mesothoracic scutellum is visible as a triangle situated medially between the bases of the two elytral halves.

3. A prothorax that is distinct from, and most often freely articulating with, the following mesothorax. The meso- and metathoracic segments are fused to form the pterothorax.
4. A generally depressed body shape, whereby the legs are situated on the ventral surface of the body. The leg bases, or coxae, are recessed into cavities formed by heavily sclerotized thoracic sclerites.
5. Abdominal sternites that are much more heavily sclerotized than the tergites. These sternites may close tightly against the lateral edges of the elytra, protecting the hind body from the attentions of predators and parasitoids.
6. Antennae usually with 11 or fewer segments.
7. Terminal genitalia that are not visible when in repose; that is, the male aedeagus and the female ovipositor are retracted into the abdominal apex when not in use.

Insects in several other orders may appear superficially similar to beetles. For example, various Hemiptera in the superfamily Pentatomoidea possess an enlarged triangular scutellum and heavily sclerotized forewings. However, these bugs can be distinguished by their beaklike suctorial mouthparts, whereas beetles retain the more generalized mandibulate mouthparts seen throughout orders such as Odonata, Orthoptera, and Hymenoptera. In addition, the forewings of Hemiptera always retain an apical membranous portion, whereas beetle forewings are consistently sclerotized throughout their length. Also, Dermaptera, or earwigs, exhibit quadrate forewings, looking much like the foreshortened elytra of staphylinoid, or rove beetles. Earwigs, however, exhibit a radial wing folding mechanism versus the transverse folding system of beetles, retain the presence of abdominal cerci, represented by large tonglike forceps at their abdominal apex, and do not undergo complete metamorphosis incorporating the pupal stage.

BEETLE DIVERSITY

Although beetles share characters supporting their common evolutionary origin, remarkable variations have evolved on the beetle theme. For example, adult body size ranges from the 0.4-mm-long *Nanosella fungi* ptiliid feather-winged beetles of North America to the 200-mm-long *Titanus giganteus* cerambycid long-horned beetles of South America. A rough estimate based on maximum dimensions for adult length, breadth, and depth puts the disparity in volume at a factor of 2.8×10^7. Life cycles also can vary in extraordinary ways, depending on the larval food resources used for development. The mushroom-inhabiting aleocharine staphylinid *Phanerota fasciata* completes three instars in 3.2 days at room temperature. Even more impressive, *Anisotoma* round fungus beetles of the family Leiodidae can complete larval development on short-lived slime mold fruiting bodies in as little as 2 days, making them arguably the fastest developing beetles yet recorded. Conversely, C. V. Riley, the first entomologist of the U.S. Department of Agriculture, reported that a larva of the dermestid carpet beetle, *Trogoderma inclusum,* survived for 3.5 years in a tight tin box. These larvae feed on the dried proteinaceous matter in animal remains, and even if Riley's larva had started with a tin full of insect specimens, the feat of solitary confinement is remarkable. *Trogoderma* larvae can even molt to a smaller size under starvation conditions, then regain size by progressively molting when food returns. Stan Beck found that mature larvae molted retrogressively eight times during a year of starvation, dropping from an initial weight of 9.24 mg to a final, svelte 1.38 mg (an 85% weight loss!).

Dramatic variation in reproductive capacity is also observed across the Coleoptera. An abundant plant pest such as the chrysomelid northern corn rootworm, *Diabrotica barberi,* can colonize cornfields and build populations quickly, since each female lays on average nine clutches of eggs, spaced 6 days apart, totaling 274 eggs over the reproductive period. At the opposite extreme we once again find the diminutive, feather-winged Ptiliidae. In eight species of *Bambara* ptiliids from Sri Lanka, the males produce spermatozoons that range in length from 220 to 600 μm; the largest size being more than two-thirds the length of the adult male producing them. After mating, these giant sperm pack the female spermatheca, with up to 28 spermatozoons recorded filling this structure. The length of the female spermathecae of various *Bambara* species is consistent within species and varies in proportion to the length of the complementary male sperm, whereas the diameter of the spermathecal duct varies in proportion to the diameter of the sperm. The female also invests heavily in her progeny, maturing one relatively giant egg in her abdomen at a time. The highly complementary male spermatozoons and female spermathecae ensure reproductive isolation because of biomechanical incompatibilities associated with any attempted interspecific matings.

Beetles are among the earliest diversifying groups of the Holometabola. Together with the orders Megaloptera, Raphidioptera, and Neuroptera, they are classified in the superorder Neuropterodea. The order Coleoptera is divisible into four major lineages, which are recognized as the suborders Archostemata, Adephaga, Myxophaga, and Polyphaga (Table I). Present-day diversity among the four coleopteran suborders is highly skewed toward the Polyphaga. Taking the numbers of beetle species estimated for Australia, John Lawrence and Everard Britton calculated that Archostemata (9 species) make up 0.03% of the Australian beetle fauna, Adephaga, with 2730 species comprise 9.6%, Myxophaga, with 2 species (0.007%), and, with 25,600 species, Polyphoga, dominates at 90.4% of the fauna. Extrapolating these figures to the estimated world total of 350,000 described beetle species suggests that Polyphaga would account for more than 300,000 species.

Consensus concerning the phylogenetic relationships among all four suborders has yet to be achieved. Recent summaries of morphological data and separate efforts using molecular sequence data reach different conclusions based on the character types and sets of taxa included. Recent studies

TABLE I Classification of Beetle Suborders, Series, Superfamilies, and Families of the Order Coleoptera

Suborder Archostemata
- Cupedoidea
  - 1. Ommatidae
  - 2. Cupedidae
  - 3. Micromalthidae

Suborder Adephaga
- Caraboidea
  - 4. Gyrinidae
  - 5. Haliplidae
  - 6. Hygrobiidae
  - 7. Amphizoidae
  - 8. Dytiscidae
  - 9. Noteridae
  - 10. Trachypachidae
  - 11. Carabidae (incl. Rhysodini, Cicindelini)

Suborder Myxophaga
- 12. Torridincolidae
- 13. Cyathoceridae
- 14. Hydroscaphidae
- 15. Microsporidae

Suborder Polyphaga
- Staphyliniformia
  - Hydrophiloidea
    - 16. Hydrophilidae
    - 17. Sphaeritidae
    - 18. Synteliidae
    - 19. Histeridae
  - Staphylinoidea
    - 20. Hydraenidae
    - 21. Ptiliidae
    - 22. Agyrtidae
    - 23. Leiodidae
    - 24. Scydmaenidae
    - 25. Silphidae
    - 26. Staphylinidae
- Sciritiformia
  - Scirtoidea
    - 27. Scirtidae
    - 28. Eucinetidae
    - 29. Clambidae
- Scarabaeiformia
  - Scarabaeoidea
    - 30. Lucanidae
    - 31. Passalidae
    - 32. Trogidae
    - 33. Glaresidae
    - 34. Pleocomidae
    - 35. Diphyllostomatidae
    - 36. Geotrupidae
    - 37. Ochodaeidae
    - 38. Ceratocanthidae
    - 39. Hybosoridae
    - 40. Glaphyridae
    - 41. Scarabaeidae
- Elateriformia
  - Dascilloidea
    - 42. Dascillidae
    - 43. Rhipiceridae
  - Buprestoidea
    - 44. Buprestidae
  - Byrrhoidea
    - 45. Byrrhidae
    - 46. Dryopidae
    - 47. Lutrochidae
    - 48. Elmidae
    - 49. Limnichidae
    - 50. Heteroceridae
    - 51. Psephenidae
    - 52. Callirhipidae
    - 53. Eulichadidae
    - 54. Ptilodactylidae
    - 55. Chelonariidae
    - 56. Cneoglossidae
  - Elateroidea
    - 57. Artematopidae
    - 58. Rhinorhipidae
    - 59. Brachypsectridae
    - 60. Cerophytidae
    - 61. Eucnemidae
    - 62. Throscidae
    - 63. Elateridae
    - 64. Plastoceridae
    - 65. Drilidae
    - 66. Omalisidae
    - 67. Lycidae
    - 68. Telegeusidae
    - 69. Phengodidae
    - 70. Lampyridae
    - 71. Omethidae
    - 72. Cantharidae
- Bostrichiformia
  - Derodontoidea
    - 73. Derodontidae
  - Bostrichoidea
    - 74. Jacobsoniidae
    - 75. Nosodendridae
    - 76. Dermestidae
    - 77. Endecatomidae
    - 78. Bostrichidae
    - 79. Anobiidae
- Cucujiformia
  - Lymexyloidea
    - 80. Lymexylidae
  - Cleroidea
    - 81. Phloiophilidae
    - 82. Trogossitidae
    - 83. Chaetosomatidae
    - 84. Cleridae
    - 85. Acanthocnemidae
    - 86. Phycosecidae
    - 87. Melyridae
  - Cucujoidea
    - 88. Protocucujidae
    - 89. Sphindidae
    - 90. Nitidulidae
    - 91. Monotomidae
    - 92. Boganiidae
    - 93. Helotidae
    - 94. Phloeostichidae
    - 95. Silvanidae
    - 96. Passandridae
    - 97. Cucujidae
    - 98. Laemophloeidae
    - 99. Propalticidae

(continues)

TABLE I (Continued)

100. Phalacridae
101. Hobartiidae
102. Cavognathidae
103. Cryptophagidae
104. Lamingtoniidae
105. Languriidae
106. Erotylidae
107. Biphyllidae
108. Byturidae
109. Bothrideridae
110. Cerylonidae
111. Discolomidae
112. Endomychidae
113. Alexiidae
114. Coccinellidae
115. Corylophidae
116. Latridiidae
Tenebrionoidea
117. Mycetophagidae
118. Archaeocrypticidae
119. Pterogeniidae
120. Ciidae
121. Tetratomidae
122. Melandryidae
123. Mordellidae
124. Rhipiphoridae
125. Colydiidae
126. Monommatidae
127. Zopheridae
128. Perimylopidae
129. Chalcodryidae
130. Trachelostenidae
131. Tenebrionidae
132. Prostomidae
133. Synchroidae
134. Oedemeridae
135. Stenotrachelidae
136. Meloidae
137. Mycteridae
138. Boridae
139. Trictenotomidae
140. Pythidae
141. Pyrochroidae
142. Salpingidae
143. Anthicidae
144. Aderidae
145. Scraptiidae
Chrysomeloidea
146. Cerambycidae
147. Chrysomelidae
Curculionoidea
148. Nemonychidae
149. Anthribidae
150. Urodontidae
151. Oxycorynidae
152. Aglycyderidae
153. Belidae
154. Attelabidae
155. Caridae
156. Ithyceridae
157. Brentidae
158. Curculionidae

Source: Modified from Lawrence, J. F., and Britton, E. B. (1994). "Australian Beetles." Melbourne University Press, Melbourne, Australia.

agree that the Archostemata are the sister group to the other three suborders. The position of Myxophaga remains ambiguous, though Beutel and Haas's comprehensive morphological analysis places them as the sister group to Polyphaga.

The burgeoning discoveries of beetle diversity throughout the course of modern scientific endeavor has begged the question, "Why?" The noted geneticist J. B. S. Haldane, in a lecture on the biological aspects of space exploration, stated that "the Creator, if he exists, has a special preference for beetles, and so we might be more likely to meet them than any other type of animal on a planet that would support life." No single answer provides the definitive biological explanation for the present-day preponderance of beetle diversity. A number of answers are consistent with the pattern of diversity, with some better supported by the comparative totals of species in the different suborders and the major families.

First, the origin of Coleoptera, relatively early in the Triassic compared with other holometabolous orders, provided ample time for diversification. Having been in existence throughout the breakup of Pangaea, which started in the Jurassic, distinct beetle biotas have evolved in place on the various continental fragments of that supercontinent.

Second, beetle diversification has been explained as the result of a successful body plan incorporating protective elytra and a flexibly articulating prothorax. Although beetles are generally not regarded as fast or agile fliers, representatives of various beetle families have routinely colonized the most remote island systems in the world. In many families, the outward appearance and function of the walking beetle has been maintained, while the metathoracic flight wings have been reduced to nonfunctional straps or vestigial flaps. This brachypterous condition eliminates the possibility of winged dispersal by individuals and is associated with increased speciation and endemism, most often in ecologically stable, geographically isolated montane, desert, or island habitats.

Third, as representatives of the Holometabola, the larval and adult beetle life stages have been morphologically decoupled via the intervening pupal stage. Larvae may exhibit morphological specializations not observed in the adult stages, and may live in particular microhabitats not primarily occupied by the adults.

Fourth, the early diversification of beetles in the Jurassic placed many lineages in prime position to exploit ecological opportunities associated with the Cretaceous diversification of flowering plants. Many of the largest families of Polyphaga (e.g., Buprestidae, Scarabaeidae, Chrysomelidae, Cerambycidae, and Curculionidae) include lineages that are intimately associated with angiosperms. These host plant associations are based on the use of various portions of the particular species or sets of species of flowering plants as larval or adult food. In addition, many other beetle groups use fungi as a food source, and fidelity to fungi of particular types is not atypical. The ability to specialize along with their larval and adult hosts has clearly been associated with extensive speciation across the Coleoptera.

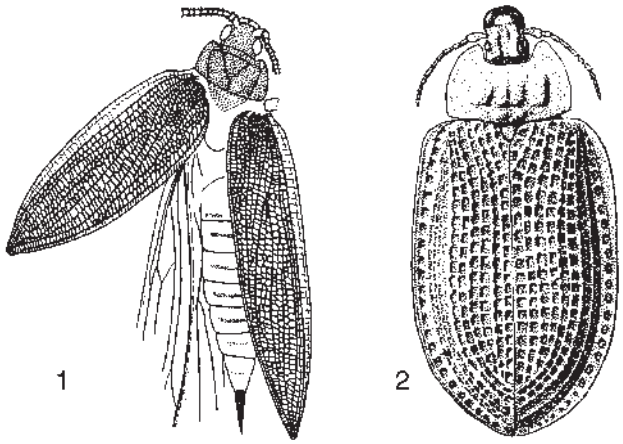

FIGURES 1–2 Fossil beetles. (1) *Moravocoleus permianus* (Tshekardocoleidae: Protocoleoptera, Permian). (© Czech Geological Survey.) (2) *Notocupes picturatus* (Cupedidae: Coleoptera, Triassic).

EVOLUTIONARY HISTORY

The earliest beetlelike insects are known from Lower Permian (280 mya) fossil deposits in Moravia, Czech Republic, and the Ural Mountains of Russia. These insects, classified in the family Tshekardocoleidae, order Protocoleoptera, resemble present-day species of the archostematan families Ommatidae and Cupedidae. They differ from true beetles in having 13-segmented antennae, elytra with more well-developed venation and more irregular longitudinal ribbing, and an abdomen and ovipositor extending beyond the apex of the elytra (Fig. 1).

We now date the origin of true Coleoptera as Triassic, about 240 mya. These fossils exhibit the coleopteran 11-segmented antennae, have more regular longitudinal ribbing on the elytra, and possess internal genitalia (Fig. 2). The earliest fossil beetle faunas have been described from Queensland in Australia, South Africa, and central Asia. The four lineages now recognized as suborders appear to have been extant at this time. The Archostemata were represented by species assignable to Ommatidae and Cupedidae, plus others belonging to families not lasting past the Mesozoic. The Adephaga included species sharing enlarged hind coxal plates such as are seen in present-day Haliplidae, plus other ground beetle-like species of Trachypachidae. Myxophagan ancestors included a variety of genera in the extinct families Catiniidae and Schizophoridae. The currently dominant suborder Polyphaga was represented in these faunas by members of the Elateroidea and Curculionoidea. These earliest beetles inhabited a world made up of early forked-leaved pteridosperms, lycopods, cycads, gingkos, and early conifers. The large animals of these communities included therapsid reptiles and dinosaurs; however, neither birds nor true mammals had yet evolved.

During the Jurassic period (210–145 mya), known family-level beetle diversity increased dramatically. Among the Adephaga, first appearances are documented for the whirlygig beetle family Gyrinidae, the ground beetle family Carabidae, and the predaceous diving beetle family

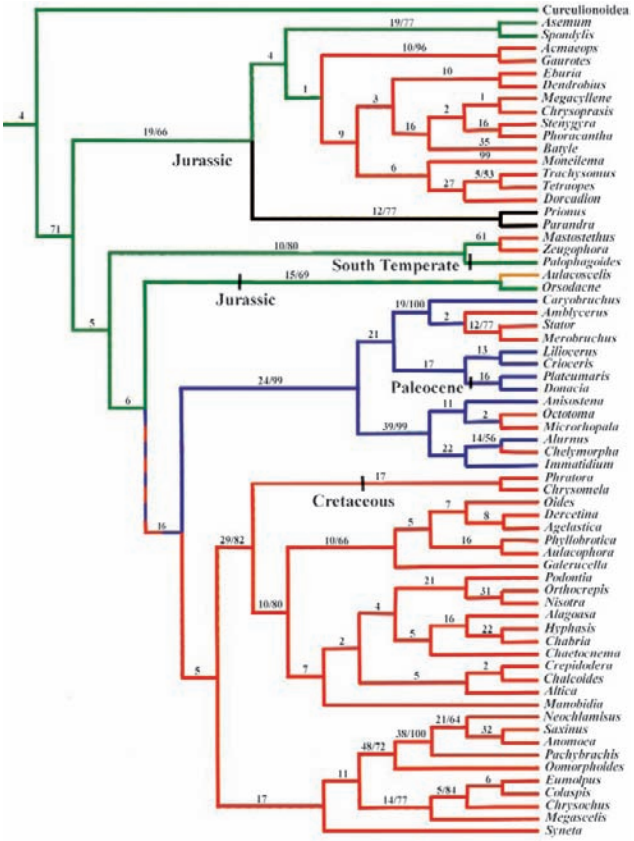

FIGURE 3 Strict-consensus estimate of the phylogeny of Chrysomeloidea and outgroups, with host groups mapped onto the cladogram. Numbers of synapomorphies/bootstrap values exceeding 50% shown along branches. Colors indicate major host group attributable to common ancestor of each group (green, Coniferae; mustard, Cycadales; red, dicotyledonous angiosperms; blue, monocotyledonous angiosperms; black, do not feed on living plants). Approximate ages of Mesozoic and early Tertiary fossils only are indicated where known, since almost all subfamily groups are present in the mid-Tertiary fossil record. [Redrawn with permission from Farrell, B. D. (1998). "Inordinate fondness" explained: Why are there so many beetles? *Science* **281,** 555–559. © 1998 American Association for the Advancement of Science.]

Dytiscidae. In all three families, the predaceous habit would be considered to be the ancestral condition. Among Polyphaga, the major families Staphylinidae, Scarabaeidae, Tenebrionidae, and Chrysomelidae are first documented. Other earliest occurrences include members of the scavenging water beetles (Hydrophilidae), carrion beetles (Silphidae), ovoid bark-gnawing beetles (Trogossitidae), tumbling flower beetles (Mordellidae), sap beetles (Nitidulidae), and false blister beetles (Oedemeridae). Of these, Scarabaeidae, Chrysomelidae, Oedemeridae, Mordellidae, plus the Triassic-aged Curculionoidea are strictly phytophagous or saprophagous. Members of the large, diverse present-day assemblage of Chrysomelidae use a broad diversity of plant hosts, ranging from cycads to conifers to angiosperms. Based on a phylogenetic hypothesis derived from extant species, the basal chrysomelid lineages are associated with primitive conifers (*Araucaria* spp.) and cycads (Fig. 3). The

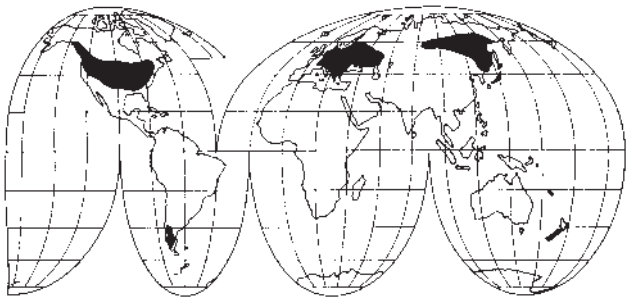

FIGURE 4 World distribution of Derodontidae. Areas supporting species include North America, Europe, Siberia, Japan, the Valdivian forest of Chile, and the South Island of New Zealand. [From Crowson, R. W. (1981). "The Biology of Coleoptera," p. 349, Fig. 2. Academic Press, London.]

Curculionoidea, the sister group to chrysomeloids, also exhibits this ancestral association with conifers and cycads. Third, the larvae of present-day Oedemeridae are borers in conifers. Thus it appears that at least several lineages of phytophagous Coleoptera were in place before the evolutionary advent of the angiosperms.

The Cretaceous witnessed initiation of the most recent round of southern landmass fragmentation, via the opening of the southern Atlantic Ocean and the isolation of New Zealand. South America and Antarctica plus Australia became progressively isolated from Africa, although they maintained contact with one another. Beetle families responded to this pattern of vicariance, with relictual distributions of several extant taxa supporting their origin during this time (Fig. 4). Continuing vicariance of the southern portions of Gondwana continued into early Tertiary, with progressive isolation of Australia, and finally the separation of Antarctica and South America at the start of the Oligocene (38 mya). This last event permitted formation of the circum-Antarctic current, helping plunge the world into a latitudinally zonated climate similar to that of today.

Preservation of beetles in amber has provided unparalleled levels of information about extinct taxa. The deposits of Baltic amber dated at 35 to 50 mya, and Dominican amber dated 15 to 40 mya, open windows onto the transition from the tropical world of the Eocene to the climatically zonated world of today. Most often, amber fossils (Fig. 5) indicate historically broader distributions for taxa presently known from only one continent (Fig. 6). This range contraction, continuing from the Eocene until the present day, suggests one explanation for the current latitudinal pattern of biodiversity. Many of the tropically adapted groups of organisms, of which beetles count significantly, have been progressively excluded from higher latitudes through the advent of cool to cold higher latitude climes, followed by the dramatic climatic perturbations associated with Pleistocene glaciation. G. Russell Coope goes so far as to argue that Pleistocene glaciation has put a halt to speciation of beetles in the temperate zones most influenced by the glaciation. His argument is based a simple fact: as he and his students studied subfossil beetle bits interred in wetland peats throughout various portions of Europe and North America, they found that all species taken from deposits younger than Pliocene could be identified as currently extant. These findings contrast starkly with those from tropical island systems, where speciation may have occurred in far younger areas. In Hawaii, for example, cave-adapted carabid beetles with reduced eyes and elongate legs have evolved from fully eyed, short-legged, epigean ancestors on the younger volcanoes of East Maui and Hawaii Island, which respectively broke the ocean surface no earlier than 750,000 and 430,000 years ago. Numerous Hawaiian beetle radiations in the Carabidae, Anobiidae, Nitidulidae, Cerambycidae, and Curculionidae demonstrate the many rapid and extensive bouts of speciation that occur in newly evolving tropical island communities.

ADULT SPECIALIZATION

It is impossible to argue for or against the proposition that possession of elytra has helped beetles' evolutionary success because possession of elytra is a defining character of "beetle-

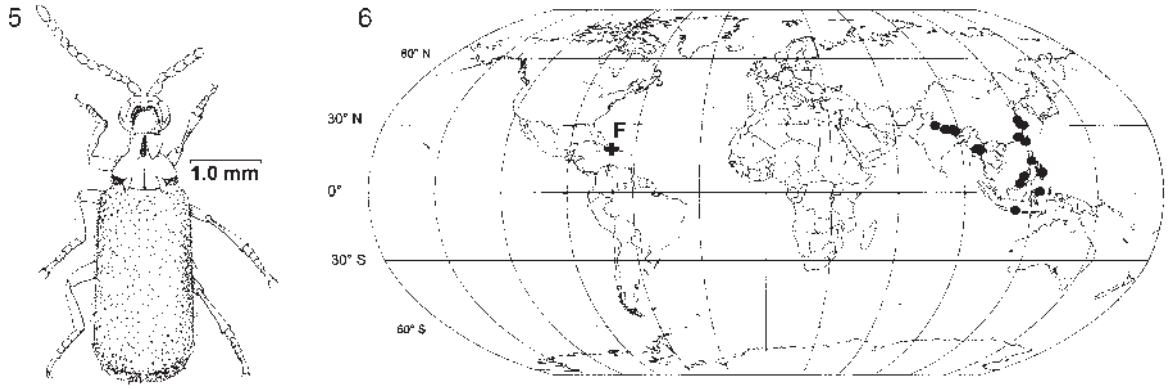

FIGURES 5–6 *Protopaussus pristinus* (Carabidae), described from Dominican amber. (5) Reconstruction of adult, dorsal view. (6) Distribution of *Protopaussus:* **F**, *P. pristinus* fossil; ● localities of extant *Protopaussus.* [From Nagel, P. (1997) New fossil paussids from Dominican amber with notes on the phylogenetic systematics of the paussine complex. *Syst. Entomol.* **22,** 345–362. © Blackwell Science Ltd.]

ness," and beetle families vary so dramatically in their diversity. For example, one of the earliest evolving beetle groups, the Cupedidae, is currently represented by only 26 species worldwide. Possession of elytra thus is only one step among many leading to successful diversification of Coleoptera. Nonetheless, functional study of the beetle body plan illustrates many instances in which "beetleness" has predisposed lineages to enter and proliferate in particular habitats.

Most generally, the organization of beetle bodies that has permitted entry into confining, laminar microhabitats involves (1) thick hard cuticle on the head and prothorax, (2) a prothorax flexibly articulating with the pterothorax, and (3) a pterothorax topped by elytra that cover folded flight wings and soft, expansible abdominal tergites. Carabid beetles utilize wedge-pushing locomotion to move through leaf litter and under loose tree bark. In these beetles, a rounded projection on the base of the hind femur impinges on the metatrochanter, which articulates only in a horizontal plane with the immobile hind coxa. Pulling the hind leg forward pushes the apex of the femur away from the body, thereby elevating the carabid's dorsum (the wedge). This upward motion is then followed by a thrust of the hind legs, forcing the beetle body forward (the push). Using this mechanism, rhysodine carabid beetles can move through dying or even living wood without leaving a trace; the wood simply closes up behind them! Their goal in this unlikely activity is foraging on the amoeboid plasmodia of slime molds (Myxomycetes).

For beetles to both fly and move through confining spaces, their wings must be stowed under the elytra while walking or wedge pushing, yet quickly unfurled for flight. All wing folding is controlled through muscles attached to the wing base; as long as tension is applied so that the radial and cubital veins are pulled apart, the wing surface remains flattened. However, relaxation of this tension brings natural folds into play, so that, with the wing apex folding in upon itself (Fig. 7), the medius comes to lie above the radius posterior (Fig. 8).

Numerous variations on wing folding have evolved depending on elytral configuration. In the archostematan Cupedidae, the wing apex rolls up longitudinally (Fig. 9). Wing folding can proceed even given the evolutionary reduction of wing venation observed in tiny beetles such as Microsporidae or Ptiliidae (Figs. 10–11). In addition to the folding characteristics of the wings, setose binding patches occurring on the wing surface, inner elytral surface, and abdominal terga are used to manage the wing folding, thus ensuring safe stowage of the wing membranes (Fig. 12).

The generalized thickening of cuticle characteristic of archostematans, adephagans, and many polyphagans results in an adult insect that is highly constrained in internal volume. Abdominal sutures between ventrites allow these segments to move against one another so that well-fed or gravid individuals will exhibit an abdomen extending beyond the elytral apex. Nonetheless, longitudinal abdominal extension can be minimized and external structural integrity maintained if the body is allowed to expand in another direction. Beetles accomplish this increase in body volume through dorsoventral expansion of the abdomen.

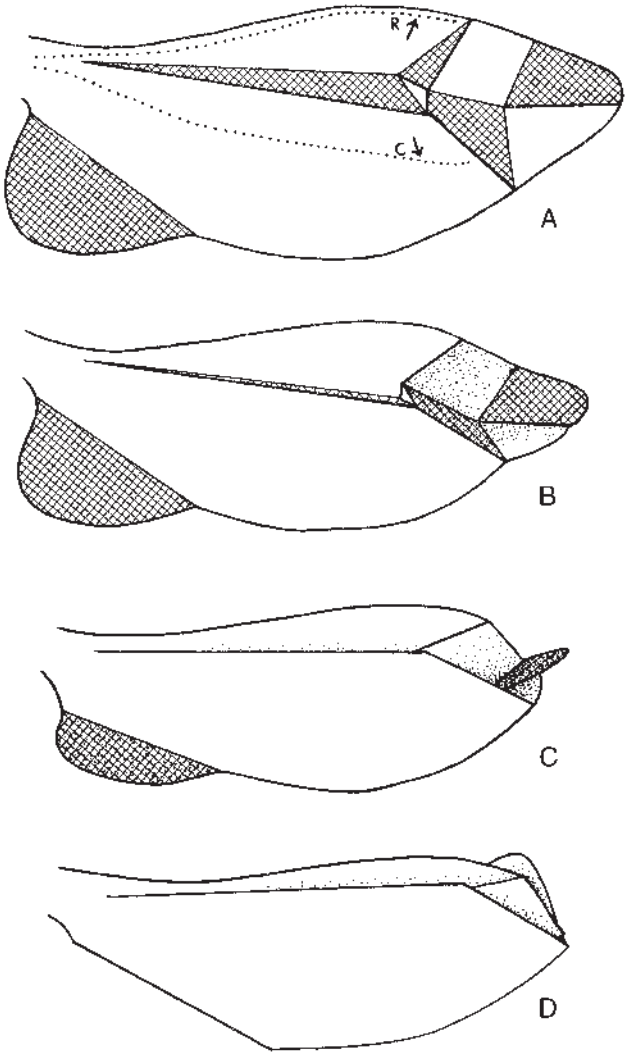

FIGURE 7 Paper model of right wing of *Cantharis* sp. arranged to demonstrate wing folding. Cross-hatched areas face ventrally in fully folded wing. In the extended wing (A) the principal veins—radius (*R*) and cubitus (*C*)—are apart by muscular action from the wing base. When this action ceases, the wing apex automatically folds (B, C) until wing is fully folded (D). [From Hammond, P. M. (1979). Wing-folding mechanisms of beetles, with special reference to investigations of adephagan phylogeny. *In* "Carabid Beetles: Their Evolution, Natural History, and Classification" (T. L. Erwin, G. E. Ball, D. R. Whitehead, and A. L. Halpern, eds.), p. 122, Fig. 1. Junk, the Hague. With kind permission of Kluwer Academic Publishers.]

In a newly eclosing beetle, the lateral reaches of the abdominal tergites lie both between and below the lateral portions of the abdominal sternites. The tergites and sternites are joined by extensive membranes, within which the spiracles are situated. These membranes may stretch, and the tergites may move dorsally relative to the stationary lateral margins of the sternites, dramatically increasing the volume of the abdomen. This volumetric expansion is accomplished without any compromise to the external armor represented

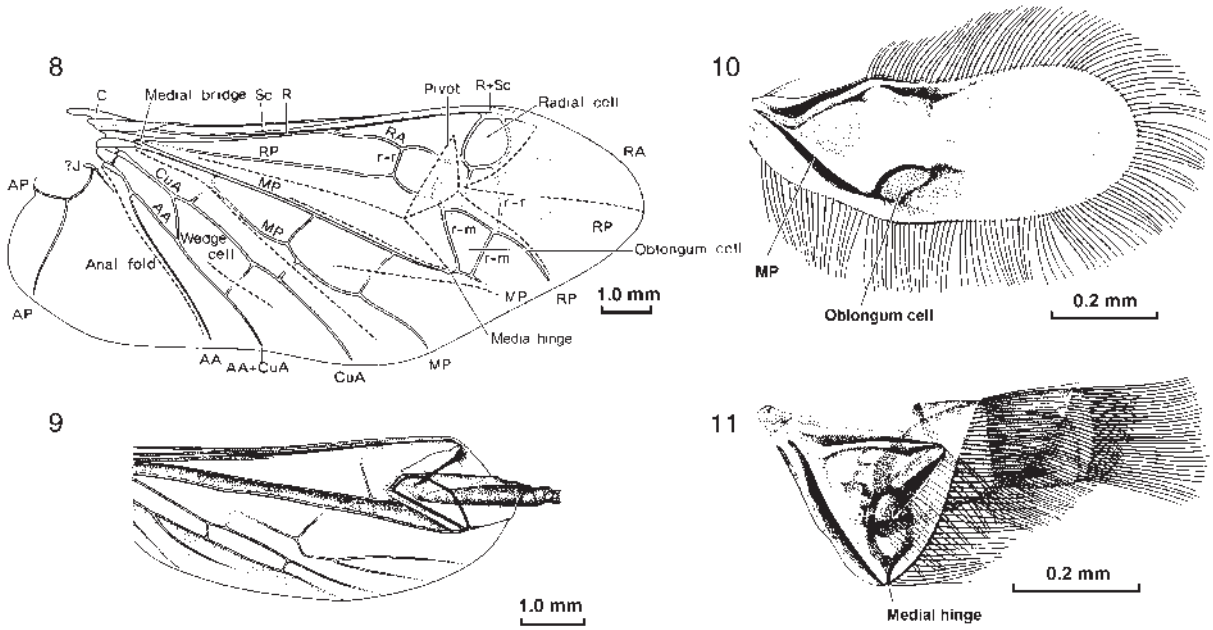

FIGURES 8–11 Hind wings of Coleoptera. (8) *Omma stanleyi* (Ommatidae): AA, anal anterior; AP, anal posterior; C, costa; CuA, cubitus anterior; J, jugal; MP, medial posterior; R, radius; RA, radius anterior; r-m, radial-median crossvein; RP, radius posterior; r-r, radial crossvein; Sc, subcosta. (9) *Adinolepis mathesoni* (Cupedidae). (10) Open hind wing of *Microsporus vensensis* (Microsporidae). (11) Folded wing of *M. ovensensis* (Images provided by copyright holder, CSIRO Entomology, Canberra, ACT, Australia.)

by the cuticle. The soft, flexible abdominal tergites are protected by the elytra, except when the beetle is flying. At this time the soft membranes and flexible tergites are vulnerable to attack by predators or parasites.

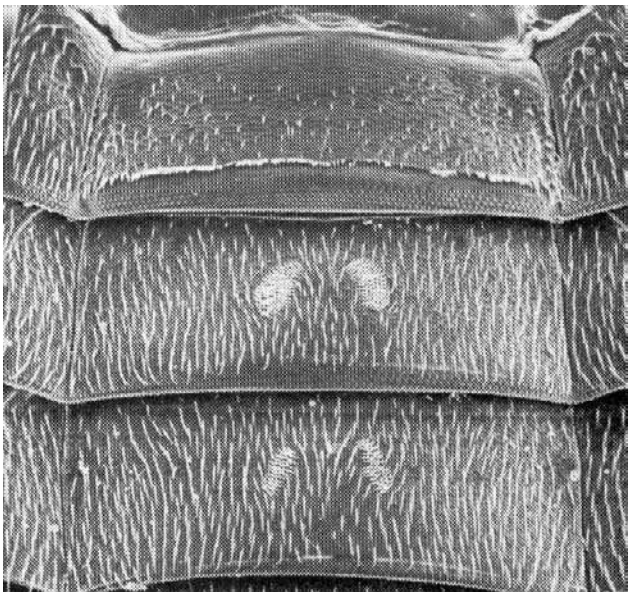

FIGURE 12 Wing-folding spicule patches on abdominal terga, *Xylodromus concinnus* (Staphylinidae) (see Fig. 7). [From Hammond, P. M. (1979). Wing-folding mechanisms of beetles, with special reference to investigations of adephagan physiology. *In* "Carabid Beetles: Their Evolution, Natural History, and Classification" (T. L. Erwin, G. E. Ball, D. R. Whitehead, and A. L. Halpern, eds.), p. 122, Fig. 1. Junk, the Hague. With kind permission of Kluwer Academic Publishers.]

In the floricolous, day-flying Buprestidae and scarab beetles of the subfamily Cetoniinae, flight is undertaken without significant separation or lifting of the elytra, with the metathoracic wings extended under the lateral elytral margins. In the buprestids, this posture allows the aposematic coloration of the elytra to be visible both in flight and at rest. In other polyphagans and the Archostemata, the elytra are held at an angle during flight, beating synchronously with the flight wings, and thereby providing some degree of aerodynamic lift.

Given the need to exchange oxygen and carbon dioxide at a liquid interface on the surfaces of the tracheolar cells, respiration represents the major activity through which an insect can lose water. This source of water loss is of particular importance for an animal of small body volume. The beetle respiratory system opens via large metathoracic spiracles and up to eight pairs of abdominal spiracles, all of which open onto the subelytral cavity. Thus, in addition to controlling gas exchange via the spiracular openings, a beetle can modulate respiration by the position of the abdominal venter relative to the elytra. Reduction of the elytra to the quadrate condition seen in Staphylinidae has resulted in secondary exposure of the abdominal spiracles.

Beetles have invaded freshwater aquatic habitats several times during their evolutionary history. In all instances, adult aquatic beetles retain the spiracular respiratory system of their terrestrial relatives, requiring that they regularly have access to atmospheric gases. The subelytral space provides the means to hold an air bubble while the beetle is active underwater. This bubble can be replenished by periodic surfacing of the beetle, during which the tip of the abdomen breaks the water surface, permitting exchange of gases.

Because of the makeup of our atmosphere (79% nitrogen, 21% oxygen), the subelytral air bubble serves as a compressible gill, permitting extended underwater sojourns. As the beetle uses oxygen, more oxygen diffuses into the bubble from the surrounding water. The carbon dioxide produced through the beetle's respiration, being highly soluble in water, quickly leaves the bubble. Because nitrogen dissolves slowly into the water, there is a gradual reduction in bubble size. The beetle can use up to eight times as much oxygen than was in the original bubble before being required to surface to replenish its air supply. Swimming beetles using these simple subelytral compressible gills include various Adephaga (e.g., Haliplidae and Dytiscidae).

In other families of the aquatic realm, oxygen is supplied to the subelytral bubble by a plastron composed of microfuge hairs or other columnar evaginations of the cuticle that are close together along their outer surface, excluding water by its surface-filming qualities. Oxygen diffuses into the plastron without any change in plastron gas volume, allowing the beetle to remain indefinitely below the water's surface. Nonetheless, plastron respiration can work only in highly oxygenated water, so beetles with plastrons are usually found in moving waters. Plastron breathers also are less active than the adephagan compressible gill breathers, because the plastron cannot provide the high levels of oxygen required for intense activity. This type of structure has evolved repeatedly in the order, being found in the Hydrophilidae, Dryopidae, Elmidae, and some Curculionidae.

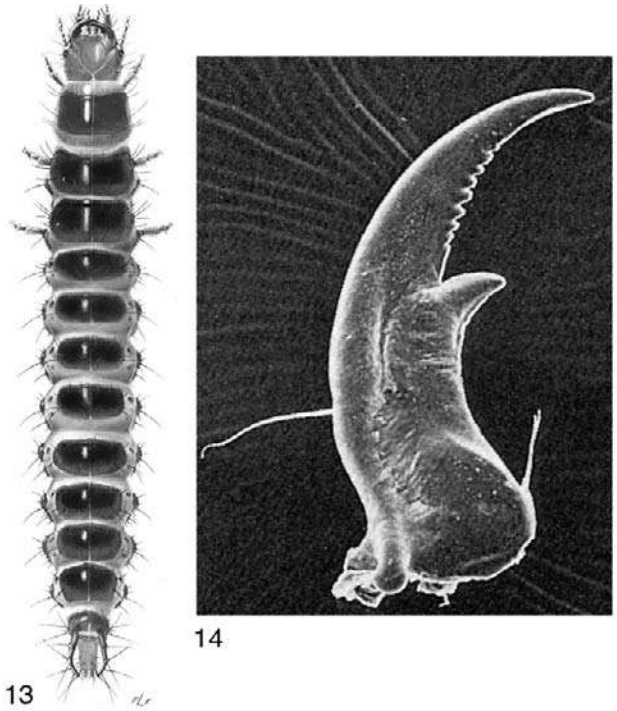

FIGURES 13–14 (13) *Abaris bigenera* mature larva (Carabidae), dorsal view (larval length, 8.3 mm). (Image © F. L. Fawcett.) (14) Right larval mandible, ventral view, *Platynus* sp. (Carabidae). Note absence of basal mola, and presence of large retinacular tooth and serrate incisor. [From Lawrence, J. F. (1991). Order Coleoptera. *In* "Immature Insects," Vol. 2 (F. W. Stehr, ed.), Fig. 34.19. Kendall/Hunt, Dubuque, IA.]

LARVAL SPECIALIZATION

Among the four suborders of Coleoptera, life histories of the predaceous Adephaga most closely resemble those of the beetles' phylogenetic sister group, the neuropteran orders. Adephagan larvae are generally campodeiform, that is, elongate and slightly dorsoventrally flattened, with long thoracic legs and a posteriorly tapered, dorsally sclerotized abdomen (Fig. 13). They typically have anteriorly directed mouthparts that often include elongate, sickle-shaped mandibles with a reduced mola (Fig. 14). The legs are six-segmented (coxa, trochanter, femur, tibia, tarsus, claws) as in the Megaloptera, Raphidioptera, and Neuroptera. The ninth abdominal tergite usually bears a pair of dorsolateral appendages (urogomphi) that may be short and unsegmented, or longer and variously segmented. These are secondarily evolved structures of the Coleoptera, and not homologous with the cerci of, for example, the orthopteroid orders. Adephagan larvae usually develop through three instars before pupation.

The larvae of Archostemata deviate from this generalized configuration by representing the syndrome that has evolved repeatedly in taxa characterized by the larval wood-boring habit. In these groups, the larvae are lightly sclerotized, more or less tubular, with shortened or reduced legs, and various ampullae on the thoracic and abdominal segments (Fig. 15). The archostematan family Micromalthidae exhibits probably the most bizarre set of larval forms and associated life cycle seen in Insecta. The campodeiform first instar is an active triungulin. It molts to become a legless, feeding cerambycoid larva, which in turn may undergo four types of molt. It may pupate directly to become an adult diploid female. Alternatively, it may develop into one of three kinds of larviform reproductive: a thelytokous pedogenetic female that parthenogenetically produces viviparously a number of diploid triungulins; an arrhenotokous pedogenetic female that lays a single egg, from which hatches a stump-legged curculionoid larva that in turn devours the mother, pupates, and then emerges as an adult haploid male; and an amphitokous pedogenetic female, which may produce either form. The hormonal controls of this system are not known, although production of the various larval types seems to be affected by environmental conditions.

The demographic consequences of this life cycle include the ability to quickly multiply and to use available rotting wood in the production of numerous dispersive adults. The triungulin larvae (Fig. 16) can expand the infestation to adjacent portions of the rotten log or timber. The cerambycoid larvae (Fig. 17), more typical of other archostematan larvae, can efficiently feed in confined galleries in rotting wood. The pedogenetic form (Fig. 18) can itself produce many more triungulins, enhancing the rate of increase of the population. The adults (Fig. 19) are produced in massive numbers, with these winged colonists establishing new colonies. Natural

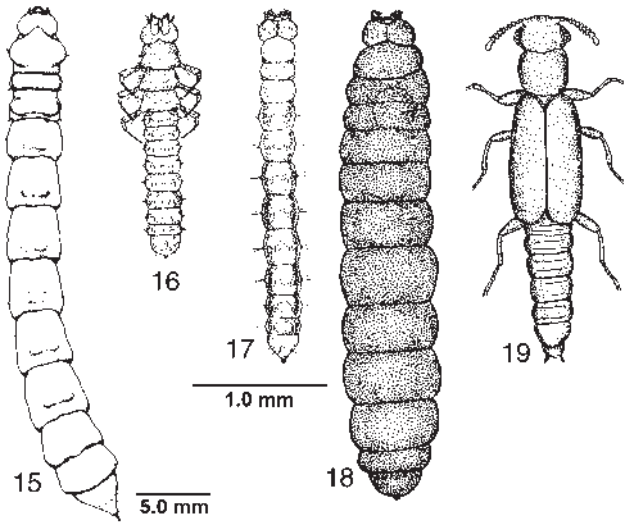

FIGURES 15–19 (15) *Tenomerga concolor* mature larva (Cupedidae), dorsal view. [From Lawrence, J. F. (1991). Order Coleoptera. *In* "Immature Insects," Vol. 2 (F. W. Stehr, ed.), Fig. 34.67a. Kendall/Hunt, Dubuque, IA. Figures 16–19, *Micromalthus debilis* (Micromalthidae), dorsal view. (16) Triungulin first instar larva. (17) Cerambycoid larva. (18) Pedogenetic larva. (19) Adult female. (Drawings, Figs. 16–19, courtesy of the copyright holder, the Royal Entomological Society, London.)

infestations have been reported in large *Quercus* (oak) or *Castanea* (chestnut) logs across the beetles' native range in northeastern North America. Other human-associated infestations have been reported from timbers deep in a South African diamond mine, and in thick oak paneling used to line the vaults of the Federal Reserve Bank in New York City.

The small suborder Myxophaga is characterized by adults and larvae of extremely small size, with both larvae and adults living interstitially in riparian areas, where they feed on algae. As opposed to the Archostemata and Adephaga, the larval legs are five-segmented, with the tarsus and claws fused into a single segment, the tarsungulus. The abdomen may or may not bear urogomphi on the ninth tergite. Like many other beetle species that feed on small particulate matter (pollen, spores, conidia, etc.), the larval mandibles bear a basal mola. Because they are aquatic in all stages, the adults bear a plastron, and the larvae may breathe by means of a plastron that covers the spiracles or via vesicular gills (i.e., a balloonlike expansion of the spiracular peritreme with an apical opening).

It is in the order Polyphaga that divergence of larval and adult lifestyles becomes evolutionarily significant. Among basal polyphagans in the superfamilies Staphylinoidea and Hydrophiloidea, larval anatomy remains generally of the campodeiform type, although mouthparts may be specialized for feeding on fungal food through development of broadly papillate molar regions on the mandible (Figs. 20–21). As in the Myxophaga, the larval leg has five segments. Aquatic forms may bear lateral gills on the thorax or abdomen (Fig. 22). Urogomphi of various configurations also may be present.

The larvae of the superfamilies Dascilloidea (Fig. 23), Byrrhoidea, and Bostrichoidea exhibit a dorsally convex body

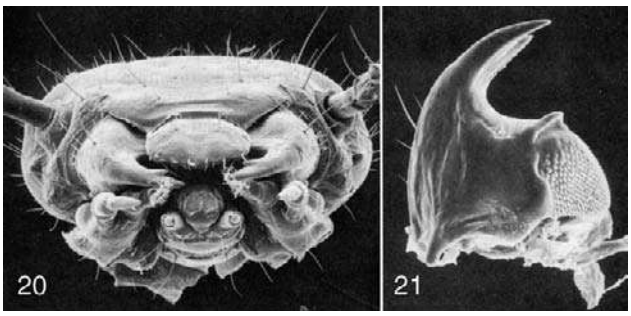

FIGURES 20–21 *Anisotoma errans* larva (Leiodidae). (20) Head capsule, anterior view. (21) Right mandible, ventral view. Note large asperate mola at base. [From Newton, A. F., Jr. (1991). Leiodidae, pp. 327–329. *In* "Immature Insects," Vol. 2 (F. W. Stehr, ed.), Figs. 34.152a and 34.154. Kendall/Hunt, Dubuque, IA.

configuration that has evolved into the much more exaggerated C-shaped grub characteristic of the Scarabaeoidea (Fig. 24). Scarab grubs can develop in a variety of microhabitats. Primitive scarabaeoids such as stag beetles (Lucanidae) and bess beetles (Passalidae) develop as saprophagous larvae in rotting wood. Larvae of the Geotrupidae and scarab subfamilies Scarabaeinae and Onthophaginae develop in mammalian herbivore dung where they also feed on fungi. Flowering plant roots are fed on by larvae of species in the more highly derived scarab subfamilies Melolonthinae, Rutelinae, and Dynastinae. Many species in these subfamilies are of economic concern, because they feed on commodities such as corn, small grains, vegetable crops, grasses, turf, fruits, and nursery stock. The C-shaped larval configuration results in an increased abdominal capacity relative to the head and thoracic forebody. This increased capacity is directly connected to the scarab larva's penchant for feeding on large amounts of food in order to pupate at a large size. Scarabs are well represented among the largest beetles, with the impressive *Goliathus* beetles of Africa and Asia attaining the greatest body mass of any beetle known.

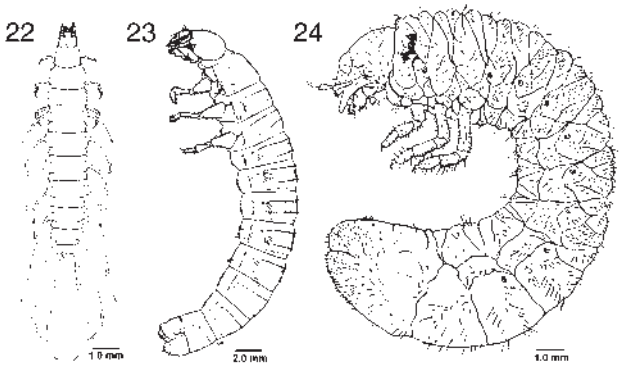

FIGURES 22–24 Beetle larvae. (22) *Berosus metalliceps* (Hydrophilidae), dorsal view. [From Spangler, J. P. (1991). Hydrophilidae, pp. 355–358. *In* "Immature Insects," Vol. 2 (F. W. Stehr, ed.), Fig. 34.296. Kendall/Hunt, Dubuque, IA.] (23) *Dascillus davidsoni* (Dascillidae), lateral view. [From Lawrence, J. F. (1991). Order Coleoptera. *In* "Immature Insects," Vol. 2 (F. W. Stehr, ed.), Fig. 34.323a. Kendall/Hunt, Dubuque, IA.] (24) *Popillia japonica* (Scarabaeidae), lateral view. (© New York Entomological Society.)

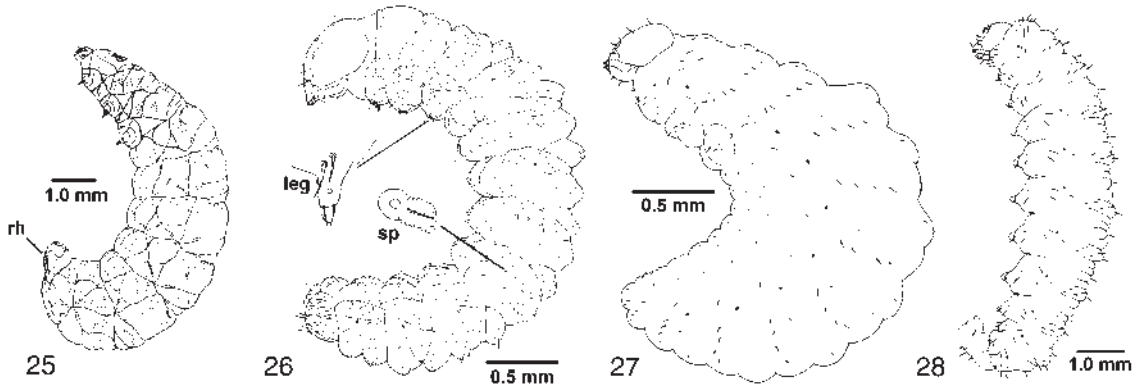

FIGURES 25–28 Larvae of phytophagous Chrysomeloidea and Curculionoidea. (25) *Donacia* sp. (Chrysomelidae), lateral view, sharp respiratory horns (rh) insert into underwater stems of water lily, providing air to the spiracles located at base of horns. (© New York Entomological Society.) (26) *Neocimberis pilosus* (Nemonychidae), lateral view (sp, spiracle). [From Anderson, W. H. (1947). *Ann. Entomol. Soc. Am.* **40,** 489–517. © Entomological Society of America.] (27) *Apion griseum* (Apionidae), lateral view. (28) *Hypera nigrirostris* (Curculionidae), lateral view.

A C-shaped larva has evolved independently in another phytophagous group with concealed larval stages; the Curculionidae. The curculionid sister group, the Chrysomeloidea, is primitively characterized by larval stages superficially similar to those of Eucinetoidea and Dascilloidea, that is, larvae of moderately convex dorsal habitus (Fig. 25). Primitive weevils retain evidence of thoracic legs (Fig. 26); however, all evidence of thoracic appendages has been evolutionarily erased in higher weevils (Fig. 27). As phytophagous weevils have specialized, taxa have moved from being internal feeders to foraging on the external surfaces of their hosts. External feeders such as the lesser clover-leaf weevil gain a foothold on their host plant through ventral abdominal ampullae (Fig. 28), analogous to the prolegs of Hymenoptera and Lepidoptera. A parallel transition from hidden feeders to exposed foliage feeders has also evolved in the weevil sister group, the Chrysomeloidea.

The Cerambycidae comprise one basal division of the chrysomeloids, with all their larvae internal feeders. The Palophaginae represent the earliest divergent lineage of Chrysomelidae, based both on late Jurassic fossils (> 145 mya), and phylogenetic analysis of living species. Larvae of this subfamily attack the male strobili of *Araucaria* (Coniferales: Araucariaceae).

The subfamily Aulacoscelinae represents another early-diverging chrysomelid lineage. Larvae of this group are internal feeders on cycads. From this syndrome of hidden feeding, leaf beetle larvae have evolved to live on open plant tissues of many of the world's angiosperms. Where plants have evolved the ability to incorporate secondary chemical compounds in their tissues, herbivorous chrysomelids have evolved to use these chemicals to recognize food and stimulate oviposition. They have also evolved the ability to sequester these broadly toxic chemicals into their tissues to gain protection from predators. Today it is commonplace to observe brightly colored larvae and adults of protected leaf beetles congregated on exposed plant surfaces, serving as a communal warning to predators regarding their unpalatability.

The wood-boring larval body plan of the Archostemata is well represented in the Polyphaga, having independently evolved in the Buprestidae (Fig. 29), Eucnemidae, and Cerambycidae (Figs. 30–31). Larvae in all these families can bore through freshly dead or dying wood by using their well-developed, anteriorly directed mandibles. Laterally expanded thoracic segments or abdominal ampullae serve to anchor these larvae in their tunnels, facilitating purchase by the

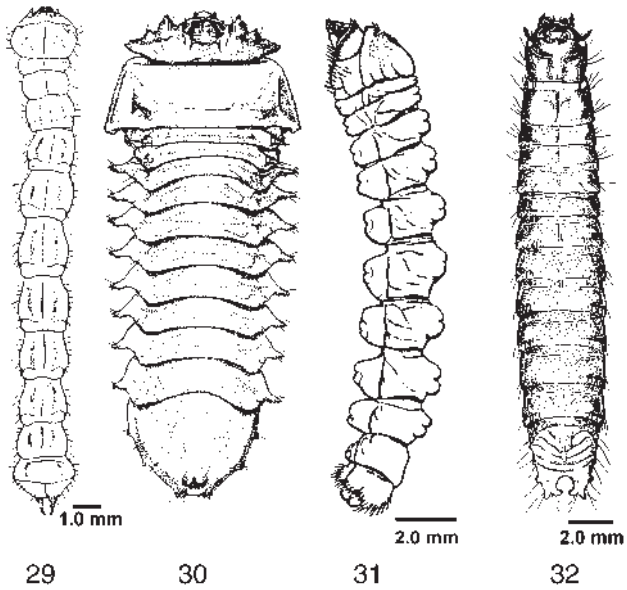

FIGURES 29–32 Larvae of wood-boring beetles. (29) *Agilus anxius* (Buprestidae), dorsal view. (30) Unidentified lepturine larva (Cerambycidae), ventral view, scale unknown. [Figs. 29, 30 from Böving, A. G., and Craighead, F. C. (1930). An illustrated synopsis of the principal forms of the order Coleoptera. *Entomol. Am.* **11**, 1–125. © New York Entomological Society.] (31) *Platyzorilespe variegata* (Cerambycidae), lateral view. [From Gardner, J. C. M. (1944). On some coleopterous larvae from India. *Ind. J. Entomol.* **6,** 111–116. © Entomological Society of India.] (32) *Hemicrepidius memnonius* (Elateridae), dorsal view. [From Dietrich, H. (1945). Cornell University Agricultural Experiment Station Memoir 269, plate IV.2. © Cornell University.]

mandibles on the wood surface. Leg reduction has proceeded during diversification of cerambycid borers, with larvae of more basally divergent subfamilies such as the Prioninae and Lepturinae having shortened thoracic legs (Fig. 30), whereas larvae of the highly derived subfamily Lamiinae (Fig. 31) are legless.

Where wood-boring beetles have gone, similarly shaped predatory beetles have followed. These tubular larvae in the Elateroidea and Cleroidea may be highly sclerotized, and they bear well-sclerotized head capsules and/or urogomphal plates (Fig. 32) that armor them appropriately for their habitats (e.g., under bark, within wood-boring beetle galleries). Elaterid larvae have diverse feeding habits, with many groups being phytophagous or saprophagous. However, all forms, regardless of food habit, imbibe their food as an extraorally predigested liquid.

Other elateroid larvae, such as fireflies (Lampyridae) and soldier beetles (Cantharidae), lack the heavy armor of the concealed gallery feeders, and prey on other arthropods among leaf and ground litter. These larvae use grooved mandibles to suck up the liquefied contents of their prey. In Elateridae, and independently in Phengodidae and Lampyridae, larval and adult stages have evolved the ability to produce light using organs composed of modified cuticular cells. The significance of larval luminescence has been variously explained. For example, night active *Pyrearinus* larvae in the elaterid subfamily Pyrophorinae use light organs to attract flying insect prey to Brazilian termite mounds where they make their home. Phengodid larvae of the genus *Phrixothrix* possess medial photic organs on the head that use red light to illuminate potential prey. But they also possess lateral abdominal light organs that emit green light. These abdominal light organs are homologous with those of Lampyridae and most likely serve to advertise that the larvae are chemically protected. Increasingly complicated light communication systems have evolved in the adult stages of various phengodid and lampyrid taxa.

Cucujoidea and Tenebrionoidea are diverse superfamilies whose larval forms blend imperceptibly into each other morphologically and biologically. It is in these groups that saprophagous and mycophagous feeding habits are associated with extensive larval diversification. Primitive larvae of both superfamilies are similar and typical of Polyphaga in many ways (e.g., five-segmented legs, urogomphi, moderate degree of sclerotization, etc.) Evolutionary trends in one often are mirrored in the other. Cucujoid and tenebrionoid larvae are usually small to moderate in size, and somewhat dorsoventrally compressed. Many are cryptozoic, occurring in leaf litter, under bark, in fungus, or in rotting wood, where they feed on fungi or on fungus-altered plant matter. Groups specialized for feeding on spores, conidia, loose hyphae, or other small particles exhibit various specializations correlated with microphagy. Most notably, these include a well-formed mandibular mola and prostheca (Fig. 33).

Extreme dorsoventral compression of the body has occurred repeatedly in response to the selective pressures of occupying subcortical and interstitial leaf litter habitats. Sometimes (Fig. 34), the body form is simply flattened, with a reorientation of the head to a protracted, prognathous condition and a migration of the leg articulations to more lateral positions. Flattening, however, may be accompanied by an additional transition to an onisciform (or pie-plate-shaped) body through extensive development of tergal flanges, resulting in a broadly oval body outline in some Cerylonidae, Corylophidae (Fig. 35), Discolomidae, and Nilioninae (Tenebrionidae). Larvae specialized for life under bark, in fungi, or in rotting wood typically have short, stout,

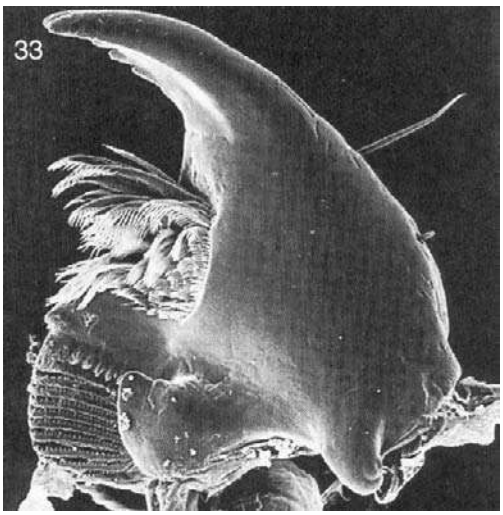

FIGURE 33 Left larval mandible, ventral view, of *Anchorius lineatus* (Biphyllidae), showing basal mola (lower left) and prostheca with comb hairs; mandible width, 0.16 mm. [From Lawrence, J. F. (1989). Mycophagy in the Coleoptera: Feeding strategies and morphological adaptations. *In* "Insect–Fungus Interactions" (N. Wilding, N. M. Collins, P. M. Hammond, and J. F. Webber, eds.), p. 6, Fig. 6. Academic Press, London.]

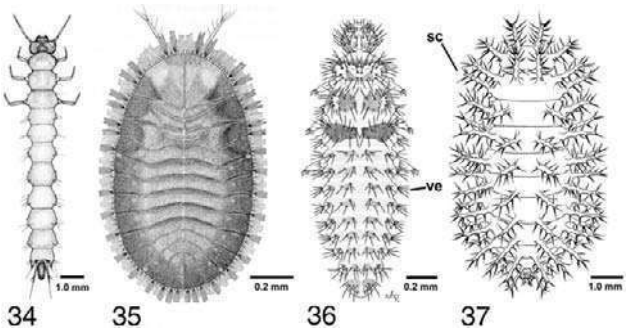

FIGURES 34–35 Flattened beetle larvae, dorsal view. (34) *Dendrophagus americanus* (Cucujidae). [From Lawrence, J. F. (1991). Order Coleoptera. *In* "Immature Insects," Vol. 2 (F. W. Stehr, ed.), Fig. 34.527. Kendall/Hunt, Dubuque, IA.] (35) Corylophidae, genus unknown [From Lawrence, J. F. (1991). In "Immature Insects," Vol. 2 (F. W. Stehr, ed.), Fig. 34.628a. Kendall/Hunt, Dubuque, IA.] Figures 36–37 Larvae of Coccinellidae, dorsal view. (36) Predaceous *Stethoris histrio.* (37) Phytophagous *Epilachna varivestis:* sc, scolus; ve, verruca. [From Le Sage, L. (1991). Coccinellidae, pp. 485–494. *In* "Immature Insects," Vol. 2 (F. W. Stehr, ed.), Fig. 34.570. Kendall/Hunt, Dubuque, IA.]

unarticulated, and unsegmented urogomphi. The apex is typically recurved to point dorsally. This form of urogomphi is thought to help larvae move about in cramped habitats.

Several tenebrionoid and cucujoid groups experienced parallel transitions to a parasitic lifestyle, including Meloidae, Rhipiphoridae, and some Cucujidae and Bothrideridae. The Rhipiphoridae provide a glimpse at parasitism involving both externally and internally feeding stages. In Rhipiphorinae, the triungulin first instar locates and attaches itself to an adult of a suitable hymenopteran host. After being carried back to the host's nest, the triungulin detaches itself and searches for a host larva. Once the host has been located, the larva burrows inside, where feeding continues (endophagy) until the parasitoid becomes greatly swollen. As the host larva reaches maturity, the parasitoid exits from its thorax, switching to feed externally (ectophagy), eventually killing it.

In Rhipidiinae, the reverse sequence of internal and external feeding occurs: the triungulin locates a cockroach as a potential host, inserts its head and thorax into a membranous region on its venter, and begins to feed while most of its body remains outside the host. Later, the larva transforms into a less mobile, legless form and moves entirely inside the host, where it begins to grow rapidly. Near the end of its development, the larva molts to a form with legs and emerges from the host to pupate.

Cucujoid and tenebrionoid taxa that are adapted for external feeding typically have a more eruciform (caterpillar-like) shape resulting from elongation of the legs, reorientation of the head to a more hypognathous position, and dorsoventral inflation to a more cylindrical shape. These external feeders also tend to exhibit defensive modifications. Aposematic coloration is common in these groups. Tergal and pleural armature, which is absent or modest in most cucujoids and tenebrionoids, becomes exaggerated in some predators (e.g., *stethoris,* Coccinellidae, Fig. 36), surface feeding herbivores (e.g., the coccinellid genus *Epilachna,* Fig. 37), and fungus feeders (e.g., the erotylid genus *Aegithus*), to form various structures such as setose, tuberculate verrucae, and complexly branched scoli.

Within Cucujoidea and Tenebrionoidea there is a recurring evolutionary transition from mycophagy/saprophagy to a lifestyle of true phytophagy as a borer in healthy herbaceous stems or wood. This entire sequence can be observed within individual families (e.g., Melandryidae), where there is a range of larval feeding that extends from boring in fungus sporophores to boring in fungus-infested wood and finally to boring in sound wood. Cucujoid and tenebrionoid wood borers tend to have fleshy bodies with conspicuous sclerotized plates usually restricted to the anterior end of the body. The head capsule tends to be prognathous and often bears a median endocarina, an internal keel on the dorsum associated with the development of especially powerful mandibular muscles.

Predatory larval forms have arisen repeatedly within Cucujoidea and Tenebrionoidea, most notably in the Coccinellidae. Accompanying this trophic transition is a suite of morphological changes to produce a campodeiform body (Fig. 36). The head typically has a more prognathous orientation. The mandibles are more prominent and lack a mola.

The beetle pupa is adecticous and usually exarate (i.e., the mandibles are fixed in position, and the head and thoracic appendages are free). Several groups have independently evolved the obtect condition; among them staphylinine Staphylinidae, Ptiliidae, and Coccinellidae, and the hispine Chrysomelidae. If the pupa rests concealed in a pupal chamber, it lies on its dorsum elevated from the substrate by numerous thoracic and abdominal setae. Pupae may be enclosed in a cocoon made of silk (aleocharine Staphylinidae, Tenebrionidae, Curculionidae), fecal material (Passalidae and some Scarabaeidae), or the larval fecal case (cryptocephaline Chrysomelidae).

Exposed pupae, as in Coccinellidae, Chrysomelidae, and Erotylinae (Erotylidae) may remain attached to their host plant or fungus via the sloughed-off last larval cuticle, which encircles the anal portion of the pupa. Such exposed pupae may be protected by defensive secretions remaining in the shed larval skin. Beetle pupae retain the ability to move the abdomen by using the flexible abdominal intersegmental membranes. Sclerotized processes on opposing margins of the abdominal segments, called gin traps, have been suggested as defensive devices used to pinch and drive off mites and other predators.

ECOLOGICAL SPECIALIZATION

One means of estimating the ground plan feeding habits of Coleoptera uses observations of extant taxa, interpreted in context of phylogenetic hypotheses for the various lineages making up the order. By this method, we would deduce that the most primitive beetles were either saprophagous wood borers as larvae, such as extant Archostemata, or that they were campodeiform predators, such as the Adephaga and the sister group to Coleoptera, the Neuroptera+Raphidioptera+Neuroptera. Examining the fossil record of Coleoptera as well as suggestive damage to fossil plants of the Triassic and Jurassic formations containing the earliest beetle fossils provides a second means of making such an estimation. By this method, we find that archostematans and primitive weevils predate fossils of all other types, suggesting that the earliest feeding habits were either saprophagous or herbivorous. Of course, fossil evidence of predation is not likely to be preserved, nor interpretable as such if it were. These two viewpoints, phylogenetic and paleontological, represent the diversity of opinion about how the first beetles lived their lives.

The two viewpoints can be reconciled if we view fossil data drawn from the various periods in light of phylogenetic estimates based on a diversity of taxa and characters. To do this, we must assume that the lifestyles of recent taxa represent those of their related fossil relatives. By this reasoning, it is very apparent that herbivorous taxa have

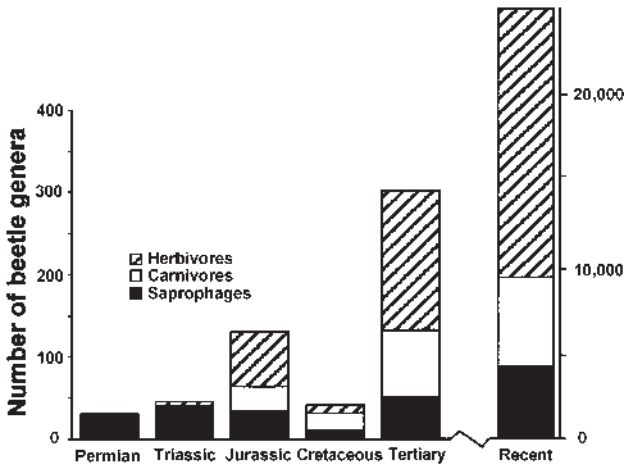

FIGURE 38 The number of beetle genera of each of three trophic levels from Permian to recent epoch. Permian genera represent Protocoleoptera. [Redrawn with permission from Farrell, B. D. (1998). "Inordinate fondness" explained: Why are there so many beetles? *Science* **281,** 555–559. © 1998 American Association for the Advancement of Science.]

constituted a majority of the major life-forms, measured by their recognition as genera, for coleopteran faunas from the Jurassic and Tertiary to recent times (Fig. 38). Even before the advent of the flowering plants, more than half the variety of beetle life-forms had evolved to focus their feeding attentions on plants. We can investigate the impact of the origin and diversification of angiosperms on beetle diversity by looking at beetle sister taxa in which one group is restricted to gymnospermous plants, whereas its sister is found on angiosperms. Brian Farrell examined lineages within the Chrysomeloidea and Curculionoidea. He found angiosperm-feeding taxa to be far more rich in species today than their gymnosperm-feeding sister groups. Clearly, angiosperm feeding has enhanced the species-level diversification of beetles living on them.

In addition to internal feeding on stem tissue, and feeding on saprophagous growth in decaying cambium, the angiosperms offer floral resources unavailable from gymnosperms. Adult beetles of many families characterized by phytophagous, saprophagous, or scavenging larvae may be found feeding in or on flowers, associated exudates, or pollen. Melolonthine and cetoniine scarab beetles, whose larvae are subterranean root feeders or rotten wood feeders, respectively, often feed on flowers. Dermestid beetle larvae scavenge dead animal matter, then move to flowers to feed on pollen after they have eclosed as adults. Once a female dermestid has fed, she becomes negatively phototactic and searches for cavities containing animal remains, where she will oviposit. Other families well represented among the pollen-feeding adults include Buprestidae, Lycidae, Nitidulidae, Mordellidae, Rhipiphoridae, Meloidae, Anthicidae, and Cerambycidae. Meloids and rhipiphorids not only feed at flowers but oviposit there, with their hatching triungulin larvae waiting in the flower to climb on passing bees and wasps, which they parasitize. Feeding on hard pollen grains is facilitated by possession of mandibles bearing a well-developed mola. Such mandibles are also associated with fungal feeding, and families such as the Nitidulidae, Tenebrionidae, and Oedemeridae contain species representing both adult feeding habits; individual oedemerid species have been reported to feed on both fungi and pollen.

Coleopteran relationships with fungi are widespread throughout the order and diverse in form. Approximately 25 extant families of beetles are primarily mycophagous. Greatly unappreciated, however, are the less obvious trophic relationships between fungi and many beetles that are ostensibly saprophagous or phytophagous. Many beetles eat plant tissue only after it has been partially broken down by fungi. Some harbor endosymbiotic fungi that allow digestion of plant tissue or provide essential nutrients. Others are thought to ingest and acquire fungal enzymes that are essential for their existence as herbivores. John Lawrence estimated that as many as half of all beetle families either are truly mycophagous or feed on plant matter that has been altered by fungal enzymes.

Ancient Greeks believed fungi were merely homes of insects. A rich insectan fauna often dwells in larger fungi, and much of it comprises mycophagous and predaceous beetles. Through evolutionary time few fungus taxa have escaped the interest of beetles. Mycophagous families seem to be especially concentrated in the polyphagan superfamilies Cucujoidea, Tenebrionoidea, and Staphylinoidea. However, fungivory arose repeatedly in various other lineages within the order as well.

Fungi are tremendously diverse physically, chemically, behaviorally, and ecologically. Mushrooms, woody conks, puffballs, truffles, yeasts, smuts, rusts, and molds present separate special challenges as food sources. In addition, a single fungus often represents a composite of resources. For example, a single polypore shelf on a log may provide a delicate layer of spore-bearing tissue on the underside, a hard, woody context, and an area where its hyphae penetrate decaying wood. Some mycophagous beetles have a broad range of acceptable hosts; however, many are more selective, feeding only on some portions of fruiting structures from a few species at a particular stage of development or decay. Host specificity tends to be narrower for immature stages. Specialization of beetles has occurred in response to the various resources and challenges that fungi present.

Woody polypore shelves offer large, persistent sources of food for mycovores. There are many different strategies for the use of the soft spore-bearing tissue of wood polypore fungi. Species of *Ellipticus* (Erotylidae) have robust mandibles capable of gouging off chunks of hymenium and its supporting tissues (Fig. 39). Larvae of *Holopsis* (Corylophidae) have found another method of tapping this resource. They use a slender, snoutlike elongation of the head to graze on the inner surface of individual spore tubes (Fig. 40). The Nannosellinae (Ptiliidae) exhibit another evolutionary solution, namely,

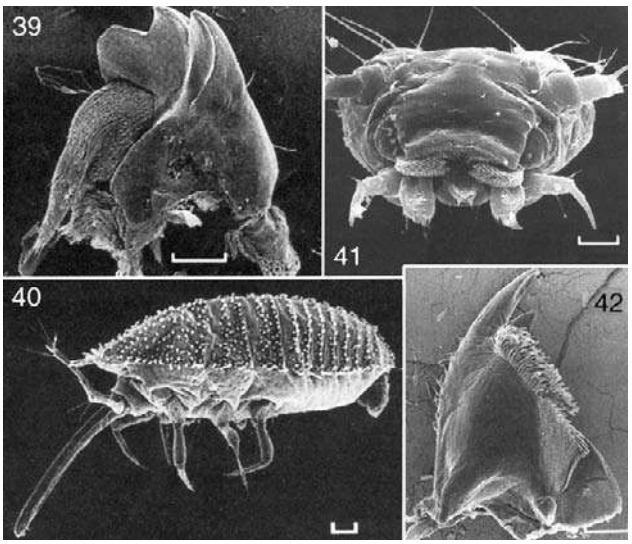

FIGURES 39–42 Feeding structures of fungus-feeding beetle larvae. (39) Larval mandible, ventral view, of *Ellipticus* sp. (Erotylidae); multidentate apex and setose lobe near base (lower left) are used to bite large chunks off fungal substrate, which are then swallowed whole (scale, 100 μm). [From Lawrence, J. F. (1989). Mycophagy in the Coleoptera: Feeding strategies and morphological adaptations. *In* "Insect–Fungus Interactions" (N. Wilding, N. M. Collins, P. M. Hammond, and J. F. Webber, eds.), p. 16, Fig. 21. Academic Press, London.] (40) *Holopsis* sp. (Corylophidae), lateral view, with long feeding rostrum bearing apical mandibles, allowing feeding inside pore tubes of sporocarp fungi (scale, 100 μm). [From Lawrence, J. F. (1989). Mycophagy in the Coleoptera: Feeding strategies and morphological adaptations. *In* "Insect–Fungus Interactions" (N. Wilding, N. M. Collins, P. M. Hammond, and J. F. Webber, eds.), p. 16, Fig. 17. Academic Press, London.] (41) Larval head, anterior view, of *Dasycerus* sp. (Staphylinidae), showing brushy mandibular apices used to remove spores or hyphae from the substrate to the mouth (scale, 50 μ). [From Lawrence, J. F. (1989). Mycophagy in the Coleoptera: Feeding strategies and morphological adaptations. *In* "Insect–Fungus Interactions" (N. Wilding, N. M. Collins, P. M. Hammond, and J. F. Webber, eds.), p. 11, Fig. 12. Academic Press, London.] (42) Larval right mandible, ventral view, of *Nosodendron unicolor* (Nosodendridae), showing food press near base (lower right) that concentrates particulate food while ejecting liquid (scale, 100 μm). [From Lawrence, J. F. (1989). Mycophagy in the Coleoptera: Feeding strategies and morphological adaptations. *In* "Insect–Fungus Interactions" (N. Wilding, N. M. Collins, P. M. Hammond, and J. F. Webber, eds.), p. 6, Fig. 6. Academic Press, London.]

miniaturization: fully grown adults, only 0.4 mm in length, crawl inside individual spore tubes to feed directly on the soft spore-bearing tissue.

Specialists on fleshy mushrooms [e.g., *Oxyporus* (Staphylinidae),] face different challenges. Unable to fly around to look for new mushrooms, larvae must complete their feeding on their ephemeral host before it decays. Many of the fungus beetles that specialize on soft mushrooms exhibit greatly accelerated development. Their mandibles are more bladelike and are capable of slicing through large chunks of soft fungal tissue.

Beetles preferring small, scattered fungal spores or conidia as food often have a suite of features related to their microphagous habits. The mouthparts tend to be brushy and capable of sweeping tiny particles from the substrate into their mouth. These modifications often involve the maxillae, but in larval *Dasycerus* (Staphylinidae) the mandibular apices are modified for this function as well (Fig. 41). The mandibular mola is also commonly modified in spore feeders. Spores are ground between opposing molar grinding surfaces on each mandible, with an action much like that of a millstone grinding wheat into flour. Nosodendridae, which feed partially on yeasts that occur in sap fluxes, use their brushy mouthparts to filter the fungal cells from the fluid (Fig. 42).

Less exploitative symbiotic relationships with fungi also are widespread and diverse within Coleoptera. The best-studied examples of mutualism with fungi are the relationships occurring in the bark and ambrosia beetles (Platypodinae and Scolytinae of the Curculionidae). Perhaps the most familiar case is that of Dutch elm disease. At the corners of this "ecological triangle" are the bark beetles (*Scolytus* spp.), the fungus (*Ceratostomella ulmi*), and the host elm trees (*Ulmus* spp.). Adult beetles nibble on tree twigs and thereby inoculate them with fungal spores. Following germination of the spores, the fungus attacks the tree and ultimately kills it. The beetles prefer to oviposit on recently killed *Ulmus* trees, many of which were recent victims of *C. ulmi*. Upon hatching, their larvae bore about, feeding on fungus-infested wood. The final link in the cycle is completed when newly emerging adults pick up fungal spores as they move around the gallery before flying off to dine on some living elm twigs.

A broad range of variants stems from the basic pattern observed in Dutch elm disease. In some cases the link between the fungus and the beetles weakens to the point of being merely incidental. In Lymexylidae and at least some Platypodinae, the relationship is a tighter, obligatory one in which the beetles farm a fungus to feed their brood. The wood of the host tree is important to the beetle only as a substrate for the fungal garden. In these evolutionarily linked relationships, the beetles often have specialized pockets called mycangia on their body to aid in the transportation of spores or conidia to new substrates (Figs. 43–44). Mycangia sometimes have associated glands that help to keep the fungal tissue viable until it is needed to start a new garden. There also is a tendency for these ambrosia fungi to be less invasive and destructive to the tree, instead staying near the galleries in which they are cultivated. Neither the fungi nor the beetles in these closer relationships can exist independently.

Another solution to digestion of plant matter is seen in some Cerambycidae and Anobiidae. Instead of using fungi to externally convert plant matter to digestible food, they rely on endosymbiotic yeasts and bacteria to accomplish the feat internally. Although yeasts and bacteria are common inhabitants of the gut in many insects, the relationship between some yeasts and beetles is one of obligatory symbiosis. Endosymbiotic yeasts may be harbored in the lumen of the gut, in diverticula (Fig. 45), or in specialized cells in the cytoplasm called mycetocytes. Clusters of

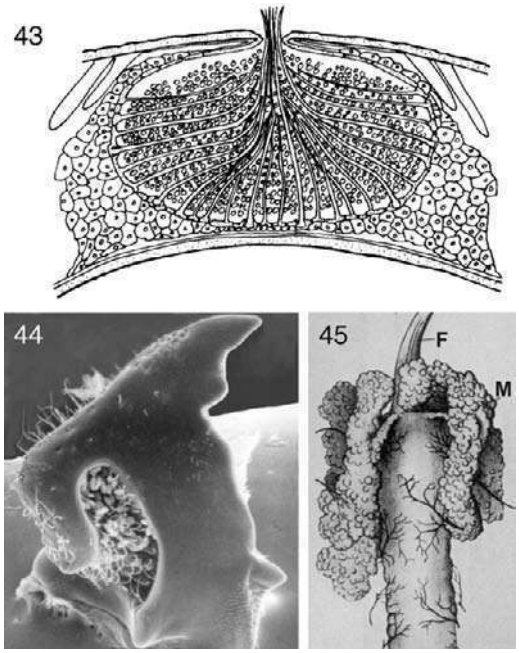

FIGURES 43–45 Beetle mycangia and mycetome. (43) *Scolytoplatypus* sp. (Curculionidae), transverse section of front part of adult pronotum, showing mycangial cavity filled with spores. [From Crowson, R. A. (1981). "The Biology of Coleoptera," p. 562, Fig. 286. Academic Press, London.] (44) *Eurysphindus hirtus* (Sphindidae), left adult mandible, dorsal view, showing spores of myxomycete inside dorsal cavity that is presumed to serve as a mycangium. [From McHugh, J. V. (1993). A revision of *Eurysphindus* LeConte and a review of sphindid classification and phylogeny. *Syst. Entomol.* **18,** 57–92. © Blackwell Science Ltd.] (45) Foregut (F) and anterior portion of midgut of *Lixus* sp. larva (Curculionidae), showing mycetomes (M). [From Crowson, R. A. (1981). "The Biology of Coleoptera," p. 562, Fig. 286. Academic Press, London.]

mycetocytes can form small organs called mycetomes. Yeasts may permit the breakdown of cellulose and provide various nutrients to their host. In the drugstore beetle, *Stegobium paniceum*, endosymbiotic yeasts are credited with providing riboflavin, niacin, pyridoxine, pantothenic acid, folic acid, and biotin.

To perpetuate endosymbiotic relationships, the gut of offspring must be charged with endosymbionts early in development. Yeasts are passed from adult beetles to larvae in various ways. The egg chorion may be inoculated with yeast so that the young are charged upon chewing out of egg and ingesting the chorion. In some Cucujidae, Silvanidae, Lyctidae, and Curculionidae, yeasts migrate into the egg within the female before the chorion is secreted. A third method of yeast transmission results following migration into the testes of the father. The yeast and sperm then enter the egg through the micropyle.

Formerly classified as fungi and studied by mycologists, the Myxomycetes are now recognized as protozoan animals. Despite their phylogenetic position, Myxomycetes are similar to fungi in some respects, and as a result beetle–myxomycete interactions share parallels with beetle–fungus interactions. In the plasmodial stage, Myxomycetes flow around their

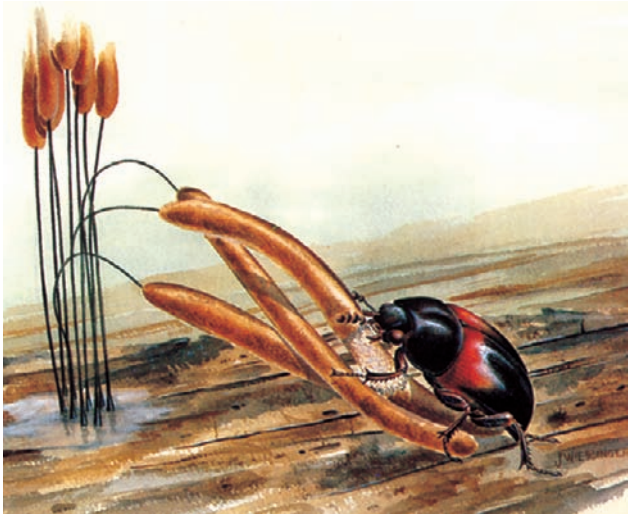

FIGURE 46 *Anisotoma basalis* (Leiodidae) feeding on a *Stemonitis* myxomycete fruiting body. [From McHugh, J. V., and Wheeler, Q. D. (1989). *Cornell Plantations Q.* **44(3),** cover figure. © Cornell Plantations.]

environment, consuming bacteria. Rhysodine Carabidae and Cerylonidae feed, at least facultatively, on the plasmodial stage. When these colonial protozoans well up as plasmodia to form a sporocarp, they take on many funguslike features. This stage has attracted specialist beetles from no fewer than seven families: Leiodidae (Fig. 46), Staphylinidae, Clambidae, Eucinetidae, Cerylonidae, Sphindidae, and Lathridiidae. Pits in the mandibles of Sphindidae, an entirely myxomycophagous family (Fig. 44), and the venter of slime-mold-feeding latridiid species have been found to house myxomycetan spores.

Whereas other holometabolous insect orders such as the Hymenoptera and Diptera include parasitoid lineages of great diversity, the Coleoptera have not diversified to any great extent via parasitism on animal hosts. In addition to meloid and rhipiphorid hymenopteran parasites, parasitism of single host individuals has been infrequently observed. Aleocharine Staphylinidae parasitize the pupae of higher flies (order Diptera, suborder Cyclorrhapha). Within the Carabidae, the bombardier beetles, or Brachinini, parasitize the pupae of Gyrinidae, and species of the genus *Lebia* parasitize chrysomelid leaf beetles. *Lebia* beetles imitate various alticine flea beetle species with which they co-occur. The quick-jumping alticines are protected from predatory birds by their ability to disappear via a jump, suggesting that the *Lebia* have evolved a similar appearance through mimetic evolution. Coccinellid predatory larvae approach the specialization seen in some parasitoids, as some of the smaller species require only one to several homopterous prey individuals to complete larval development. Nonetheless, these species can switch prey species depending on the density of various hosts.

Platypsyllus castoris beetles of the family Leiodidae are specialists on beavers, with both the flattened, highly modified adults (Fig. 47) and the larval stages living in the animals' fur. Related leiodids in the subfamily Leptininae live

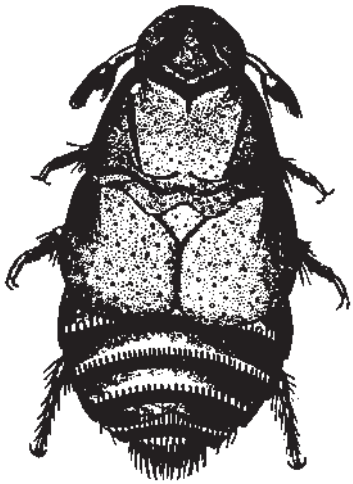

FIGURE 47 *Platypsyllus castoris* (Leiodidae), parasitic on beaver (*Castor* spp.). [From Crowson, R. A. (1981). "The Biology of Coleoptera," p. 549, Fig. 280. Academic Press, London.]

on the bodies of rodents, though they exhibit much less extreme body forms, and a lower level of host specificity, than the beaver beetles. The highly specific host relationship of *Platypsyllus* probably evolved from a more general predaceous habit. Such nest inquilines are found in a variety of lineages within the Staphylinidae, with adults and larvae variously preying on flea larvae or other nest-associated scavengers.

INTRASPECIFIC INTERACTIONS

The newly eclosed adult beetle faces the various tasks of dispersing from the pupal habitat, finding a mate, mating, finding a suitable larval habitat, ovipositing, and possibly guarding or facilitating the development of its young, all the while avoiding natural enemies. To beetles, flight may be a rare event. Many species undergo only a nuptial flight from the larval habitat to a new habitat, where mating and oviposition occur. Others may move from a breeding habitat to a drier microhabitat for overwintering, and thence back to the breeding habitat the next spring, making three flight periods in their lifetime. Others, such as the floricolous cerambycid long-horned beetles, buprestid jewel beetles, herbivorous Chrysomelidae, and Homoptera-feeding Coccinellidae may fly more or less continuously during their adult life span as they move from plant to plant. Beetle flight always requires the unfolding of the flight wings. Typically, beetles will climb some sort of prominence, use their antennae as "windsocks" and orient their body so that their initial liftoff is against the wind, and then open their elytra and unfold their flight wings prior to takeoff. During the nuptial flight, beetles are likely to be reproductively incompetent. In some scarab beetles, vast amounts of air are swallowed prior to flight, resulting in a distended gut unsuitable for feeding.

Mate finding may be facilitated by aggregation on host plants. The crushed leaves of host- and non-host-plant species are attractive to both male and female scarab beetles. Adult emergence occurs as synchronous mass flights. Once near or on the host plants, female pheromones attract male scarabs. Ruteline scarabs utilize sex pheromones derived via fatty acid biosynthesis, whereas scarabs in the not distantly related Melolonthinae utilize amino acid derivatives and terpenoid compounds. The compounds of different classes are released from glands on different parts of the body: ruteline pheromones from epithelial cells lining inner surfaces of the abdominal apical segments, for example, or melolonthine pheromones from eversible glands on the abdominal apex. Pheromones used in the other beetle superfamilies span these scarab pheromone classes (e.g., terpenoids in the Curculionidae, fatty-acid-derived aldehydes and acetates in the Elateridae, esters in the Dermestidae).

Beetles use the other sensory modalities in mate finding, outdoing diversity observed in any other insect order. The anobiid deathwatch beetle acquired its ominous name through the predisposition of its males to bang their head capsules on host wood, telegraphically inquiring whether a receptive female is in the vicinity. This rapping was thought to foretell an imminent death. A males initially taps an average of five times, and if a female responds with a single tap, he moves a short distance and taps once. If he determines that the second female tap is fainter than the first, he turns at various angles to attempt to approach the female. Males receiving no returning female tap to their five-tap overture move greater straight-line distances between tapping bouts, searching greater expanses of wooden habitat for a responsive female (Fig. 48).

The use of light for mate finding has been evolutionarily refined in the Elateroidea, with the Lampyridae using flashing signals produced in abdominal light organs to engage in complex male–female dialogue before mating. These light organs are modified fat body cells with transparent outer surfaces, backed with highly reflective uric acid crystals. The light is highly efficiently produced via the oxidation of luciferin by the enzyme luciferase in the presence of adenosine triphosphate (ATP) and oxygen, producing oxyluciferin, carbon dioxide, and light. Male flashes are composed of species-specific series of flashes of varying duration, composition, and in some instances intensity. Males of different species fly in different patterns and at different heights, while females respond with a simpler flash that encodes species identity by the response delay to the male flashing sequence, by the flash duration, and in several species by a multiple-flash sequence. This sexual communication has been co-opted as a predation mechanism in *Photuris* fireflies. Males and females of these adult-feeding lampyrids use a typical male–female light dialogue to mate, whereupon the female's nervous system is affected so that she sends species-specific mating responses coded for sympatric, smaller *Photinus* species. *Photinus* males who venture too near the faux-*Photinus* female responses sent by the *Photuris* females are eaten.

Males and females may undertake various types of precopulatory behavior before mating. These may involve the sensing of species-specific alkene aphrodisiacs related to cuticular

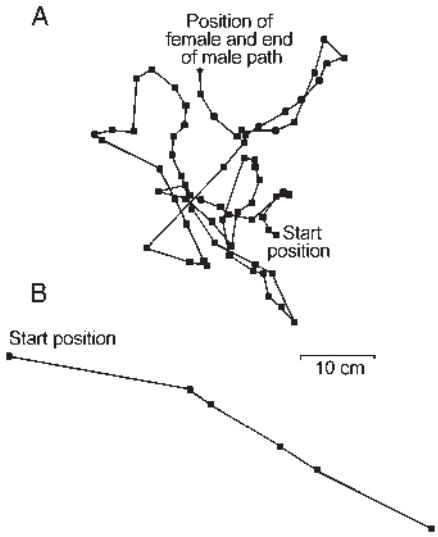

FIGURE 48 Paths taken by male deathwatch beetle, *Xestobium rufovillosum* (squares indicate positions at which male stopped to tap). (A) Female responded to male head tapping and was successfully found by male. (B) Male tapped in absence of any responding female. [Redrawn from Goulson, D., Birch, M. C., and Wyatt, T. D. (1994). Mate location in the deathwatch beetle, *Xestobium rufovillosum* DeGeer (Anobiidae): Orientation to substrate vibrations. *Anim. Behav.* **47,** 899–907.]

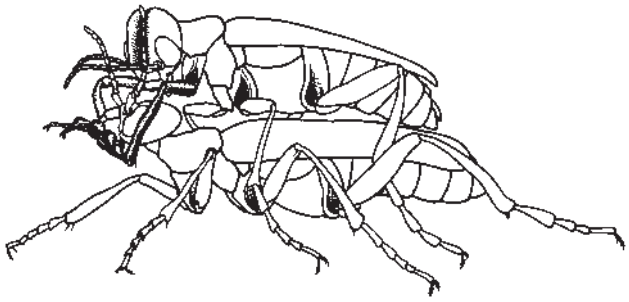

FIGURE 49 *Eupompha fissiceps* (Meloidae), male antennating female while rubbing his tarsi under her head; male antennae bring female's antennae alternately into his cephalic sulcus. [From Pinto, J. D. (1977). *Ann. Entomol. Soc. Am.* **70,** 937–952. © Entomological Society of America.]

hydrocarbons, as in aleocharine Staphylinidae. Males or females may stridulate as part of their behavioral repertoire. In Meloidae, a male will climb onto the dorsum of the female and antennate her head, palps, or antennae. In *Eupompha* meloids, the males draw the antennae of the females along a longitudinal sulcus on the male vertex (Fig. 49). Genitalic insertion by the male is successfully attempted only after antennation of the female. The passage of a nuptial gift of the highly toxic compound cantharidin has been incorporated into mating behavior in the pyrochroid fire beetles. In *Neopyrochroa flabellata*, the female samples an exudate from a transverse sulcus on the male vertex. If the exudate contains the terpenoid cantharidin (better known as the mammalian "aphrodisiac" Spanish fly), the male successfully mates, whereupon he transfers to the female, along with his sperm, about half the cantharidin stored in the accessory glands of his reproductive tract. The female translocates this cantharidin from her spermatheca to the developing eggs, which are thus chemically protected from predation by cantharidin-sensitive predators. Although meloids are known to produce cantharidin, transfer of this compound during meloid mating has not been documented. Conversely, although pyrochroids utilize this chemical in their mating behavior, they do not seem to be able to synthesize it, and the natural source of cantharidin that facilitates their behavior remains to be discovered.

Precopulatory behavior may involve more than a male and a female, especially in species in which male-specific structures have evolved in elaborate fashion. The enlarged male mandibles of stag beetles, Lucanidae, and prominent horns on the heads and pronotum of scarab beetles, are used by males to joust for advantageous mating sites with females. Dynastine and other scarabs seek out branches of shrubs and low trees upon which to mate. Males competitively maneuver for the top position on the branch, which is favored by females entering the fray for mating.

Copulation occurs with male dorsal to the female, the male grasping the female with the fore- and midlegs, and sometimes the mandibles, as in the tiger beetles or cicindeline Carabidae. The male aedeagus is inserted into the female gonopore. An aedeagal internal sac may be everted to place the male's gonopore near the entrance to the female spermatheca, and a spermatophore may be passed that encloses the sperm. Most beetles exhibit a monotrysian female reproductive tract: that is, the eggs pass out of the same structures used for copulation. However, in the dytiscid water beetles, a ditrysian configuration has evolved whereby copulation and oviposition occur via parallel, though connected, passages in the female.

The necessity for mating and copulation has been obviated in various groups of Curculionidae, Chrysomelidae, and Carabidae through thelytokous parthenogenesis. Species may be composed entirely of parthenogenetic populations, or such populations may be restricted to peripheral portions of the range. Parthenogenesis may also be associated with polyploidy, especially in weevils.

Although most mated female beetles oviposit into appropriate microhabitats where the larvae will develop, some families are characterized by eggs being laid in masses (e.g., Coccinellidae). Some tortoise shell chrysomelid females, Cassidinae, will lay eggs in a mass and then hover over the mass through hatching and the early days of the larvae. Female pterostichine carabids of the genera *Abax* and *Molops* similarly guard their eggs, although only until hatching. Females of several staphylinid species of *Oxyporus*, voracious mycovores with large sicklelike mandibles, have been reported to oviposit several eggs within a cavity in a soft mushroom, then stay with the larvae as they quickly develop to pupation over 3 to 6 days.

Ovoviviparity, or the holding of eggs until larvae hatch, has evolved several times across the Coleoptera. Typically it occurs in beetles occupying marginal environments dangerous to egg

development. Chrysomelid females of montane or subarctic species hold developing eggs in the reproductive tract while basking on sun-drenched leaves to hasten egg development before larviposition. In *Pseudomorpha hubbardi* carabid beetles, females hold developing eggs until the larvae can be deposited, whereupon the larvae complete development as inquilines in an ant nest.

Male and female cooperative brood rearing has evolved repeatedly in various groups of Coleoptera. The long-known, and oft-revered dung-rolling Scarabaeidae provision nest burrows with rolled dung balls, upon which the eggs are laid. Females undertake this activity alone in some species, whereas the sexes work together in others. Burrows may be dug before dung balls are cut from mammalian dung pats, requiring navigation from the dung pat to a predetermined burrow location, or the burrow may be dug after the dung ball has been constructed.

In the Australian *Cephalodesmius armiger*, males and females pair up, with males actively foraging for decomposing leaves, flowers, fruit, and seeds, which are brought back to the female, staying in the nest. The female works the plant materials into a compressed ball, to which she adds her fecal material. The microbiological action of fungi from her feces causes fermentation in this external rumen after larval brood balls have been made from the mass. As the larvae develop, they feed on the brood ball from the inside out. When the thickness of the walls of the brood ball drops to about 2 mm, the increased volume of larval stridulations sensed by the female stimulates her to add more decaying material to the brood ball. Four to ten brood balls are made per nesting pair. When the larvae finish their development, the female seals the brood ball with a combination of larval and female feces, the larva having ejected its fecal material through cracks in the brood ball before pupation. Both parents die before adult emergence of their young. The new adult beetles feed on the walls of the brood ball, inoculating their gut with the fungi used by the mothers to produce fermentation in the external rumen.

Like the nest-building scarab, beetles of the silphid carrion beetle genus *Nicrophorus* raise their young on a concealed, highly desirable resource, a decaying carcass. Adult *Nicrophorus* actively fly long distances searching for a carcass. If a male discovers one, he emits a pheromone that attracts a female, with mating occurring on the carcass. Male and female then cooperatively bury the carcass by digging underneath it, and maneuver the corpse into a ball. Their activities isolate the corpse from competing silphids, and insulate it from microorganisms in the soil. After repeated mating, the female lays eggs in the surrounding soil. Upon hatching, the larvae crawl to the carcass, attracted by olfactory cues and adult stridulation. The adults precondition part of the carcass for larval feeding by chewing on it. They first feed the young larvae by regurgitating predigested carrion. Older larvae feed on their own, developing in 1 to 3 weeks, during which the female stays on the carcass. Upon maturation, the larvae crawl into the adjoining soil to pupate, and the female leaves to search for a new carcass. Variations on this scenario include more than one pair of adults supported by a larger carcass, mated females raising their larvae alone when they discover a carcass without a resident male, and larger *Nicrophorus* species usurping a carcass through killing the original colonizing adults and their larvae.

True sociality, wherein more than one generation of adult lives together, and reproduction is restricted to a portion of the individuals, is likely to be rare in Coleoptera. This behavior has been reported only twice, and conclusive studies to completely document interactions among adults and larvae have not been fully documented for either family, represented by the wood-inhabiting bess beetles (Passalidae) and the ambrosia beetle *Austroplatypus incompertus* (Curculionidae: Scolytinae). There is no doubt that adults and larvae live together, and that fungi are passed from generation to generation. For this arrangement to qualify as eusociality, the existence of individuals that assist reproductives but do not themselves reproduce, at least during a portion of their life, must be documented.

INTERSPECIFIC INTERACTIONS

Beetles exhibit defensive behavior that is mostly rooted in the attributes of their cuticle. Many beetles living an exposed portion of their life cycle on vegetation will use the "drop-off" reflex if disturbed (i.e., simply close the legs and tumble off the leaf or branch and fall to the ground, where their often cryptic coloration helps protect them from visually oriented predators). The drop-off reflex can be combined with thanatosis, in which the beetle lies still with legs appressed to the body. Alternatively, the legs may be held at irregular positions by muscular tetanus (catalepsy), or the individual may roll up into a ball with the antennae, mouthparts, and legs hidden from view. More brightly colored species do not use the drop-off reflex. Chrysomelid flea beetles have enlarged hind femora containing strong tibial extensor muscles; a cuticular femoral spring releases the stored energy, catapulting them into the air.

Defensive chemical secretions that protect beetle adults from predators have evolved numerous times. Toluquinone is a defensive constituent common to several major terrestrial families (Carabidae, Staphylinidae, and Tenebrionidae), suggesting that this was one of the earliest defensive secretion types to have evolved. Since quinones are used in the tanning process of new cuticle, they would have been evolutionarily available in large quantities in well-sclerotized ancestral lineages of these families. Their tanning nature is not restricted to insect cuticle, as attested by the darkly stained fingertips of anyone who picks up an oozing *Eleodes* tenebrionid beetle.

Perhaps the most famous defensive chemical reaction in beetles is observed in the crepitating bombardier beetles of the carabid tribe Brachinini. These beetles, like other carabids, possess pygidial defensive glands that empty from the lateral edges of the intersegmental membranes between the seventh

and eighth abdominal segments, Brachinine bombardier beetles plus carabid beetles of several other tribes (Metriini and Paussini) eject a combination of hydroquinones plus hydrogen peroxide held in one chamber of the gland, and catalases plus hydrogen peroxidase held in a second chamber. These chemicals combined result in an explosive ejection of hot (100°C) secretion, with liberation of the oxygen of H_2O_2, thus reducing hydroquinone to quinone, with the released O_2 propelling the spray (Fig. 50).

In addition to quinone compounds, beetles have evolved to use a variety of other defensive chemicals. The more recently evolved carabid beetle groups spray formic acid, a chemical also utilized as a defensive agent by their omnipresent antagonists, the ants, or Formicidae. Brightly colored or starkly patterned beetles are candidates for chemical protection via defensive gland secretions. The buprestid jewel beetles are often colored in black and yellow stripes to appear like the Hymenoptera with which they cohabit in various flowers. Jewel beetles are highly protected by bitter chemicals named buprestins. Not only are these chemicals distasteful to mammals (viz., organic chemists!), but ants reject sugar solutions laced with buprestins. Jewel beetles form mimetic complexes with lycid beetles, themselves protected by defensive secretions composed of various substituted parazines, reportedly among the most powerful odorous substances known (Fig. 51).

Various other beetle families regularly contribute members to lycid-based mimicry rings, including Cerambycidae, Meloidae, and Oedemeridae. Given that the meloids and oedemerids can synthesize cantharidin, it is likely that most beetles in such rings are distasteful, making Müllerian mimicry the dominant basis for such common color patterns (Fig. 51). Other mimicry rings center on the dangerously toxic *Paederus* staphylinid beetles (Fig. 52), the cuticle of which exudes pederin. When such a beetle is scraped or crushed, contact with the pederin released results in human whiplash dermatitis (Fig. 53).

INTERACTIONS WITH HUMANS

Throughout history, humans have had diverse interactions with and perceptions of beetles. Coccinellid beetles were once perceived to have a close association with the Virgin Mary, hence their common name "ladybugs." Ancient Egyptians recognized dung beetles (Scarabaeidae) as a symbol of Ra, the sun god, because of parallels between the beetles' behavior and cosmic activities credited to the deity. Much as the scarabs rolled dung balls across the desert, Ra was thought to guide the sun across the sky each day. The symbolism of sacred scarabs has continued until today, as scarab images are still incorporated into jewelry, signifying good luck to the buyer or wearer.

The mystery and aesthetic beauty of beetles has been captured in paintings, sculptures, dances, poems, songs, and other art forms. Beetles have been used by many cultures for decoration. The brilliant metallic elytra of Buprestidae serve as natural sequins on textiles, and as biological gems in jewelry. In some cultures, beetle horns are included in jewelry because they are thought to increase sexual potency.

Live stag beetles (Lucanidae) are prized as pets in Japan, where a considerable amount of study has been given to their care in captivity. In Thailand the practice of "fighting" male

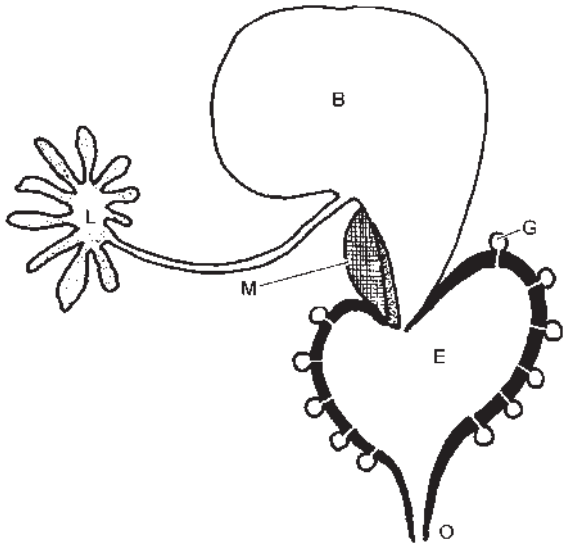

FIGURE 50 Cross section of pygidial defense gland of *Brachinus* bombardier beetle adult (Carabidae): L, secretory lobes; B, collecting vesicle; M, sphincter muscle; E explosion chamber; G, ectodermal glands that secrete catalase; O, outlet. Vesicle B contains mixture of hydroquinone and hydrogen peroxide, exploded by catalase, when it passes into E. [From Crowson, R. A. (1981). "The Biology of Coleoptera," p. 502, Fig. 265. Academic Press, London.]

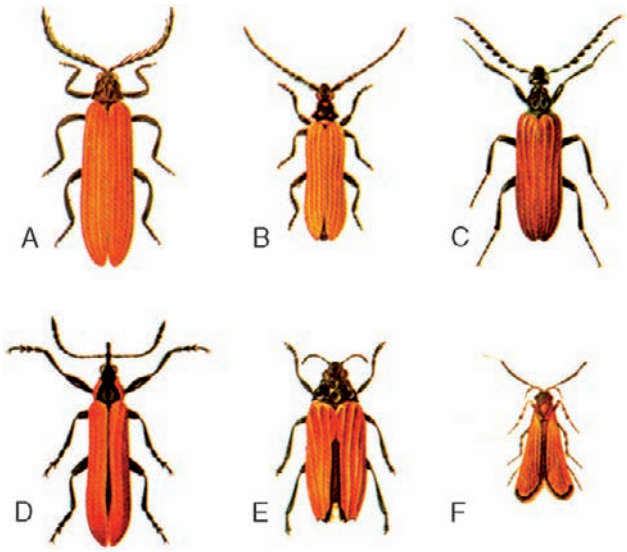

FIGURE 51 Five Australian beetles and a moth forming part of a mimicry ring: (A) *Metriorrhynchus rhipidius* (Lycidae), (B) *Eroschema poweri* (Cerambycidae), (C) *Tmesidera rufipennis* (Meloidae), (D) *Rhinotia haemoptera* (Belidae), (E) *Stigmodera nasuta* (Buprestidae), (F) *Snellenia lineata* (Lepidoptera: Oecophoridae). (Images provided by copyright holder, CSIRO Entomology, Canberra, ACT, Australia.)

52 53

FIGURES 52–53 (52) *Paederus cruenticollis* (Staphylinidae) exhibiting warning coloration observed in many species of this genus. (Image provided by copyright holder, CSIRO Entomology, Canberra, ACT, Australia.) (53) Dermatitis linearis on human forearm at 66 h after an adult *Paederus* beetle had been crushed on volunteer's skin. (Photograph courtesy J. Howard Frank.)

Hercules beetles (Scarabaeidae) is a traditional source of entertainment. With a referee controlling the action, two males are introduced into an arena. When a female is placed nearby, her mating pheromones trigger the combatants to engage each other. The match ends and a victor is declared when one male becomes exhausted or backs down from the advances of his opponent. In Central America local craftsmen blur the distinction between "pet" and "jewelry" by gluing rhinestones, glass beads, and a small chain to the dorsal surface of zopherid beetles. When the tiny chain is pinned to clothing, the tethered beetle becomes living jewelry.

Entomophagy, the eating of insects, is common in many parts of the world, and beetles often make up part of the menu. Larvae of palm weevils (Curculionidae) are considered to be a delicacy on the islands of the South Pacific. Similarly the fleshy, sausagelike larvae of various long-horned beetles (Cerambycidae) and scarabs are relished by people around the world. Mealworms, the larvae of some tenebrionid beetles, are easily reared and have become standard fare for culinary demonstrations of entomophagy.

Beetles attract the most attention when they become economic pests of agriculture, horticulture, and forestry. Two families, the snout beetles (Curculionidae) and the leaf beetles (Chrysomelidae), include many serious pest species. In the middle to late 1800s, the Colorado potato beetle, *Leptinotarsa decimlineata* (Chrysomelidae) abruptly expanded its range across North America and then colonized Europe and neighboring regions. Great efforts were made to thwart the invader each time it appeared, but ultimately the beetles succeeded. Throughout the 20th century an epic battle was waged against the notorious boll weevil, *Anthonomus grandis grandis* (Curculionidae), in the Cotton Belt of the southern United States, where it inflicted great financial losses. A sustained and coordinated effort to control this pest succeeded in eradicating the boll weevil from portions of several states by the turn of the millennium.

Predaceous ladybugs are often used in biological control to suppress populations of homopterous crop pests (i.e., aphids and scales). In the first successful biological control introduction, an Australian ladybug, *Rodalia cardinalis*, suppressed the cottony cushion scale (Hemiptera) on citrus crops in California. Phytophagous beetles have been employed to control weeds. In the 1960s the cattle-rearing industry in Australia faced a dilemma: because cows are not native to the continent, no natural bovine dung entomofauna was available to use their feces. Therefore cow patties persisted for months, during which time they served as breeding grounds for pestiferous horn flies. After careful study, Australian entomologists introduced South African *Onthophagus* dung beetles (Scarabaeidae). The measure was successful, and the problem quickly abated.

FIGURE 54 The jewel scarab, *Chrysina cusuquensis,* known only from a restricted fragment of forest in northern Guatemala. (Photograph courtesy of David Hawks.)

Perhaps the least appreciated human–beetle interactions are those in which human population pressure inflicts a negative impact on beetle populations. Coleopteran diversity is largely attributable to their specialization for particular geographic locales, microhabitats, and food. As human populations grow and people alter the Earth for their needs, destruction of spatially restricted resources is an inevitable result, leading to extinction of species associated with those resources. Ironically, a characteristic that helped Coleoptera to attain the astounding degree of diversity that it exhibits today also predisposes many beetle species to anthropogenic extinction (Fig. 54).

See Also the Following Articles

Boll Weevil • Cultural Entomology • Dung Beetles • Hymenoptera • Japanese Beetle • June Beetles • Ladybugs

Further Reading

Arnett, R. H., Jr., and Thomas, M. C. (eds.). (2001). "American Beetles." Vol. 1. CRC Press, Boca Raton, FL.

Arnett, R. H., Jr., Thomas, M. C. Skelley, P. E., and Frank, J. H. (2002). "American Beetles." Vol. 2. CRC Press, Boca Raton, FL.

Beutel, R. G., and Haas, F. (2000). Phylogenetic relationships of the suborders of Coleoptera (Insecta). *Cladistics* **16,** 103–141.

Branham, M. A., and Wenzel, J. W. (2001). The evolution of bioluminescence in cantharoids (Coleoptera: Elateroidea). *Fla. Entomol.* **84,** 565–586.

Carpenter, F. M. (1992). Arthropoda 4. *In* "Treatise on Invertebrate Paleontology," Part R (R. L. Kaesler, ed.). Geological Society of America, Boulder, CO, and University of Kansas, Lawrence.

Choe, J. C., and Crespi, B. J. (eds.). (1997). "The Evolution of Social Behavior in Insects and Arthropods." Cambridge University Press, Cambridge, U.K.

Crowson, R. A. (1981). "The Biology of Coleoptera." Academic Press, London.

Elias, S. A. (1994). "Quaternary Insects and Their Environments." Smithsonian Institution Press, Washington, DC.

Evans, A. V., and Bellamy, C. L. (1996). "An Inordinate Fondness for Beetles." Holt, New York.

Farrell, B. D. (1998). "Inordinate fondness" explained: Why are there so many beetles? *Science* **281,** 555–559.

Lawrence, J. F. (1989). Mycophagy in the Coleoptera: Feeding strategies and morphological adaptations. *In* "Insect–Fungus Interactions" (N. Wilding, N. M. Collins, P. M. Hammond, and J. F. Webber, eds.), pp. 1–23. Academic Press, London.

Lawrence, J. F. (1991). Order Coleoptera. *In* "Immature Insects." Vol. 2 (F. W. Stehr, ed.), pp. 144–658. Kendall/Hunt, Dubuque, IA.

Lawrence, J. F., and Britton, E. B. (1994). "Australian Beetles." Melbourne University Press, Melbourne, Australia.

Lawrence, J. F., Hastings, A. M., Dallwitz, M. J., Paine, T. A., and Zurcher, E. J. (2000). Beetles of the World (CD-ROM, Windows version). CSIRO Publishing, Victoria, Australia.

Lawrence, J. F., Hastings, A. M., Dallwitz, M. J., Paine T. A., and Zurcher, E. J. (2000). Beetle Larvae of the World (CD-ROM, Windows version). CSIRO Publishing, Victoria, Australia.

McCormick, J. P., and Carrel, J. E. (1997). Cantharidin biosynthesis and function in meloid beetles. *In* "Pheromone Biochemistry" (G. D. Prestwich and G. J. Blomquist, eds.), pp. 307–350. Academic Press, London.

Collection and Preservation

Charles V. Covell, Jr.
University of Louisville

Insect collecting often begins in youth, when one discovers the love of making specimens for school, scouts, 4-H clubs, and other projects or as a fascinating pastime in its own right. The great diversity and numbers of insects, plus their rapid life cycles, usually mean insect populations can afford to give up some of their numbers and not be adversely affected by most collecting activities.

As one becomes engaged in various facets of insect biology as a researcher, the collection of specimens is important for taxonomic research, ecological studies, bioassessment and biomonitoring, and physiological and genetic studies. Because each labeled specimen is a historical record of that species' occurrence in time and place, proper methods of collecting, preparing, labeling, and storing are vital.

The general habitats, collecting equipment needs, and methods of collection and storage for the major insect orders and order groupings are presented in Table I. Below, a description of each type of equipment and its use are given. For more extensive illustrations and descriptions consult the books listed under Further Reading.

BASIC EQUIPMENT FOR COLLECTING INSECTS

1. Aerial net—A net bag made of translucent netting so one can see what's inside; it can be used as a beating net if needed. The net is used to grab insects off plants or to cover them on the ground. Since insects tend to crawl or fly upward, hold the net so they move toward its closed end once they are inside.

2. Beating (or sweeping) net—A heavy cloth bag, perhaps with small netted area at the bottom; it is used to sweep "like a broom" through vegetation many times. To use, strongly wave the net to concentrate insects in bottom of the net before placing net with insects into a killing jar until movement ceases. Then pick out what is desired and allow the rest to revive and go free.

3. Aquatic net—A heavy-duty metal hoop that can be D-shaped or round supports the netting. The former type is best for stream bottoms. The mesh and heavy cloth skirt have to be strong enough to take a beating. To use, hold the net against the bottom of the stream riffle and disturb the substrate upstream to allow insects to flow into net, or "work" the net among plants or debris to catch pond insects.

4. Malaise trap—A tent-like structure made of netting and designed to direct insects that encounter it to climb upward and follow the seams to a collecting container into which they fall. Container can be designed for live capture or killing in alcohol or by means of a dry poison such as cyanide powder.

5. Lights and light traps—A battery-powered light bulb such as a 15-W fluorescent "black light" or self-ballasted mercury vapor lamp can be hung from a tree limb or other support about one-half meter in front of a white sheet strung between two trees in the forest. The collector then picks the desired insects off the sheet. Various trap designs are available from supply houses, in which lights attract the insects that hit one of four vanes (or baffles) surrounding the bulb and above a funnel, into which the insects fall when they hit a vane. Ethyl acetate in tins with "wicks" of cloth provide a killing agent; crumpled paper also can be used in the bucket below for live capture.

6. Pitfall traps—Tin cans, jars, or pails can be placed in holes dug in the ground and filled with earth to the outside rims. One may bait with dead animal matter or other attractants. Ethylene glycol (antifreeze) is often used as a

TABLE I Collection and Preservation of Insect Specimens for Insect Orders

Taxon	Habitat	Equipment to use	Collection method	Preparation
Protura, Diplura, and Collembola	Leaf litter, rotten logs and stumps, birds' nests, other detritus	Berlese funnel, aspirator, wet brush	Place in funnel for several days, jar of alcohol beneath, light above	70% EtOH, mount on microslides
Thysanura and Microcoryphia	Buildings (silverfish), leaf litter, logs, seashores	Forceps, Berlese funnel	Same as above	70% EtOH
Ephemeroptera	Naiads: streams, rivers, lakes Adults: fields and forests	Dip nets, grab samplers Aerial nets, light traps	Kick samples, pick off stones Pick off plants or from light sheet	70% EtOH
Odonata	Naiads: streams, lakes, ponds	Dip nets	Dredge or kick sample with net	70% EtOH, place in envelope, wings folded over back, and card with collecting data; spread for display
	Adults: fields, near streams and ponds	Aerial nets	Sweep fast from behind with net	
Plecoptera	Naiads: streams Adults: along streams, at lights	Aquatic nets Light trap, aerial and sweep nets, light trapping	Kick-netting in riffles, pick off stones, sweep shore vegetation	70% EtOH
Orthoptera and other orthopteroids	Fields, forests, gardens, and other terrestrial habitats	Sweep nets, light traps, aerial nets, hand capture	Sweep and aerial netting, light trap sampling	Mount on insect pins, support body until dry
Hemiptera, Homoptera, and other hemipteroids	All terrestrial habitats	Sweep nets, beating sheet, examine plants, light traps	Sweep and aerial netting, light trap sampling	Pin large bugs, small ones on card points or store in 70% EtOH, scales on microslides
Phthiraptera	Avian and mammalian hosts	Forceps, aspirator	Scrape fur and feathers	70% EtOH, mount on microslides
Thysanoptera	Plant axils, flower parts, and other plant parts	Aspirator	Examine plants and aspirate	70% EtOH, mount on microslides
Neuroptera and Megaloptera	Larvae aquatic (mostly streams) or on plants	Aquatic nets, sweep nets and light traps	Kick sampling in riffles, sweep vegetation, examine trap samples	70% EtOH or pin
Coleoptera	All habitats	Aquatic, aerial, and sweep nets: light, malaise and pitfall traps	Bait pitfall traps with rotting animal flesh, other methods as above	Pin or mount on card points
Mecoptera	Woodland glades, understories	Sweep and aerial nets, light trap	Follow and net individuals seen, use light trap (Meropeidae)	Pin or place in 70% EtOH
Lepidoptera	All habitats, esp. fields and woods	Aerial net, sweep net, bait, malaise and light traps	Net resting butterflies, bait traps with rotting animal flesh and excrement or fermenting fruit, sweep or examine plants for larvae	Relax, and then spread on spreading boards, use 70% EtOH or special fluids for larvae
Trichoptera	Running water, esp. streams for larvae	Aquatic, sweeping, and aerial nets	Kick samples for some larvae, others must be picked in cases off rocks in stream	Store all stages in 70% EtOH
	Adults may be near or far from breeding sites	Malaise and light traps	Adults come to lights or can be swept from streamside vegetation	
Diptera	All habitats; larvae most common in aquatic or moist habitats in water and land or animal hosts	All kinds of nets, dippers, light traps, malaise traps	Examine plant and animal hosts, capture in net, traps	Pin, place on card points, or store in 70% EtOH
Siphonaptera	Bodies and nests of birds and mammals	Aspirator or moistened brush, sweep net	Comb animal, break up nest over white background, sweep grassy areas around infested buildings	Place in 70% EtOH, mount on microslides
Hymenoptera	All terrestrial habitats	Nets, all trap types	Collect from flowers, sweep, extract from light, malaise, pitfall, and other traps	Mount on pins; on card points or in 70% EtOH if small

killing agent. Walls of boards can also be erected narrowing to the opening of the pitfall to direct arthropods to the pit.

7. Beating sheet—A square of bed sheet or similar white cloth placed under a bush or tree to catch insects when they are knocked off after the plants are struck with a large stick, such as an axe handle. Insects are then collected by aspirator or forceps.

8. Aspirator—A tube plugged with a rubber cork in which are inserted two tubes: one bent and used to point at tiny insects; the other connected to a rubber tube for inhaling quickly to suck the insect into the tube. The latter one has a tiny screen attached to the inside end to prevent insects from getting into one's mouth.

9. Berlese funnel—A commercial funnel of any size is needed, equipped with a screen inserted just above the narrow spout to prevent material from falling out. Leaf litter, birds' nests, and other organic matter are put into the funnel, which is mounted on a rack or ring stand. A light bulb is placed over the top to dry out the organic material, driving arthropods downward as they seek moisture. Insects then fall through the screen and into a jar of 70% alcohol placed under the spout. The Berlese funnel is left in place until the organic matter is completely dried out.

10. Relaxing box—A tight container (plastic refrigerator boxes are excellent) is chosen in a size needed. Cut or fold paper toweling to line the bottom of the box at least 1 cm deep. Moisten the paper thoroughly with water, but leave no water standing. Add a small amount of an anti-mold chemical such as paradichlorobenzene or carbolic acid (phenol). Place a piece of stiff cardboard above the wet paper as a platform for the specimens. Freshly killed insects, or dried ones you wish to pin or spread, can be softened in the box. If left in the "relaxer" too long, however, they may mold or turn mushy and disintegrate.

11. Killing jar—A glass or plastic jar of desired size can be made into a killing jar by putting about a 1-cm layer of plaster of Paris in the bottom, or use just a pad of absorbent material such as cellucotton, cotton, or soft tissue. A fluid killing agent such as ethyl acetate or fingernail polish remover containing acetone is added to be absorbed by the plaster or other material. Be sure not to have any fluid on the walls of the jar, or specimens will be spoiled. If you use cotton or other absorbent material, cut a cardboard disk to separate the insects from the pad of killing agent.

METHODS

Insects are prepared for study and storage in three basic ways: pinning, fluid storage, and mounting on microslides. Adult insects or the immature forms of hard-bodied insects such as those with incomplete metamorphosis are pinned through the thorax of the body, unless too tiny, and then they are mounted on card points (see later). Insect pins, available from supply houses, are long and very sharp. They range from tiny headless "minuten nadeln" for mounting specimens on tiny blocks of foam, which in turn are put on regular insect pins, to pins that are numbered to match the general size of the insect. Size 000 is the smallest made and bends very easily. Most small insects that can be pinned are at least 5 mm in length, with a thorax big enough to hold the pin. Most medium and large insects are pinned on sizes 1 to 3. Sizes 4 to 7 are sometimes available for large specimens.

Preserving Insect Specimens in Fluid

Insects that are too small, or the bodies of which are too brittle or soft, should not be pinned. They should be stored in glass vials in 70% ethyl alcohol (EtOH). Other special fluids, especially those that preserve colors, can be learned from the works under Further Reading. Actually collecting in alcohol can be done using traps of any type (light, malaise, pitfall, and some bait traps). The larger insects can be dried out later and pinned. However, collecting in fluid is NOT recommended for collecting Lepidoptera (butterflies and moths) or Culicidae (mosquitoes) because they have patterns formed of colored scales and those may be ruined by the fluid.

Vials used are often of the "patent lip" type with neoprene stoppers. The author prefers to use 4-dram vials with size 0 stoppers and store them in plastic racks and cardboard boxes with partitions available from supply houses. A better alternative is the screw-cap vial, which should be equipped with "polyseal" plastic sealing inserts. One of the biggest problems with liquid-stored specimens is the drying out of the fluid. I believe the latter storage to be superior because the alcohol does not discolor with years nor does the cap change shape (stoppers swell or stick to the glass).

Many tiny insects such as lice, fleas, and thrips can be stored in EtOH until such time as they can be made into permanent microslide mounts with Euparol, Canada balsam, or some other mounting medium.

Pinning Insect Specimens

1. Be sure the insects to be pinned are soft enough so that they will not crumble when you handle them and attempt to pierce them with the pin. These can be just-caught, or they can be softened, if dry, in a relaxing box (see earlier).

2. Select the pin and pierce the high point of the thorax with the point. Push the pin straight through the thorax. Check straightness by observing from front and side to see if the pin is perfectly perpendicular to both the transverse and the longitudinal axis of the insect.

3. When the insect, such as a grasshopper, has a middorsal ridge in the thorax, pin just to the right of the ridge.

4. For beetles, insert the pin in the right elytron (front wing) close to the midline. Do not pin beetles through the prothorax.

5. Push the pin on through when you are satisfied with the position. One-third to one-fourth of the pin should be showing above the insect's thorax.

6. If the abdomen or legs are drooping, push the pinned insect into a block of foam plastic or a cardboard box to

support these parts until they are dry. Then remove the insect and label it.

7. Most museum specimens do not have legs and antennae adjusted to a life-like position when they are pinned. However, for display purposes or personal satisfaction one may move these body parts into desired positions on the foam or cardboard support and fix them temporarily with pins over or against them.

Placing Insects on Card Points

A card point is a small wedge of high quality (100% cotton content) cardstock, punched from the sheet with a special punch obtainable from a supply house. There are several different shapes, but the author prefers the ones with the wide end rounded.

Card pointing is used for tiny insects that are hardbodied enough not to lose shape when dried. Size usually ranges from 1 to 5 mm or slightly larger in length. The author normally selects from large samples of dried specimens collected in sweep samples or light traps.

1. Punch out a number of card points. Place them on top of a firm foam plastic or cardboard surface.
2. Push the point of an insect pin into the wide end a short way from the very end, and push the card point up the pin by inserting the pin with the card point into the top hole of a 3-step pinning block (wooden block with three fine holes of different depths to provide uniform heights of labels on pins) and pushing the point up until it stops. It should be about 1/3 the distance from the top of the pin.
3. Use forceps to turn the very tip of the card point downward at a right angle to form a vertical surface.
4. Put a tiny dab of glue on the vertical surface you have made with the forceps. When doing a number of specimens, put a small drop of glue on a piece of card or paper to use (although it will tend to harden on the surface after a minute or two).
5. Position the insect so that the right side of the thorax is accessible, and touch the glue-covered surface of the card point to the right side of the thorax. (The insect should appear to be "holding onto the card point with its right hand"). Use forceps to position the insect firmly against the glued surface and have it positioned so that its orientation to the ground is as it would be in life.
6. Fill out your insect label with locality, date, and collector's name. Trim it to be as small as possible (avoid large, oversized "barn door" labels). Labels should be printed on 100% cotton light card stock in permanent black (India) ink or can be done on a postscript laser printer.
7. Position the label on the pin and push it up the pin at the middle hole of the 3-step pinning block. As you read the label, the card point and insect should be projecting to the left of the pin shaft. Make sure both card point and label are not tilted or crooked.
8. Place the specimen in a temporary holding unit tray until it can be identified and put in the collection. Identification labels should be affixed below the collecting data label and in a position so that both labels can be read from the same angle. The lowest step on the 3-step pinning block is normally used for the identification label.

INSTRUCTIONS FOR SPREADING BUTTERFLIES AND MOTHS

1. Have all needed items ready: well-softened specimens (stored in relaxing box or freezer after collecting), spreading boards of proper sizes, straight strips of tracing or waxed paper or other material, setting needles or picks, insect pins for specimens, glass-headed pins for holding paper in place (insect or dressmaker's pins okay). See Fig. 1.
2. Fix paper strips along the side boards of spreading board, slightly back from the notch to allow you to work the wings into place. Use two or three pins at the top of board to hold the paper even down its length.
3. Push a proper sized insect pin straight down through the thorax of specimen, so it is not tilted in any direction. Push the pin far enough that the top of the thorax is one-third to one-fourth the distance down from the pin head (Fig. 1A).
4. Push the pin down into the soft material in the notch of the spreading board so that it is not tilted in any direction. Also, push it far enough that the wings, when out straight to the side of the insect, rest flat on the side boards of the spreading board. Be sure you do not place the insect too close to the top of the board (leave room to pull wings into proper position).
5. Push an insect pin down along the left rear of the thorax, behind the base of the left hind wing, to keep the body from swinging left as you position the wings.
6. Place paper over the wings. Hold the left-hand paper strip in the thumb and forefinger of your left hand while you now begin to position the wings.
7. Insert a sharp insect pin or setting pick behind the costa vein close to the base of the left forewing. Swing that wing upward until the inner (anal) margin is at a right angle to the plane of the body (Fig. 1B). Be sure not to let the hind wing pop out from below the forewing. Insert a glass headed pin into the paper above the costa near the base and inner margin near the anal angle to hold the wing secure.
8. Pull the left hind wing forward by inserting a setting pin or pick behind the radial vein near the wing base, and swinging it forward. Leave a small triangular space between the outer margin of the hind wing and the inner margin of the forewing. Fasten paper over the left hind wing by putting a pin below it near the wing base.
9. Repeat procedures 6 to 8 on the right side, and be sure you have produced symmetrical results (Fig. 1C).
10. Position the antennae with pins to look as shown in Fig. 1. The abdomen may need to be supported with crossed pins beneath it or held down straight with crossed pins above it.

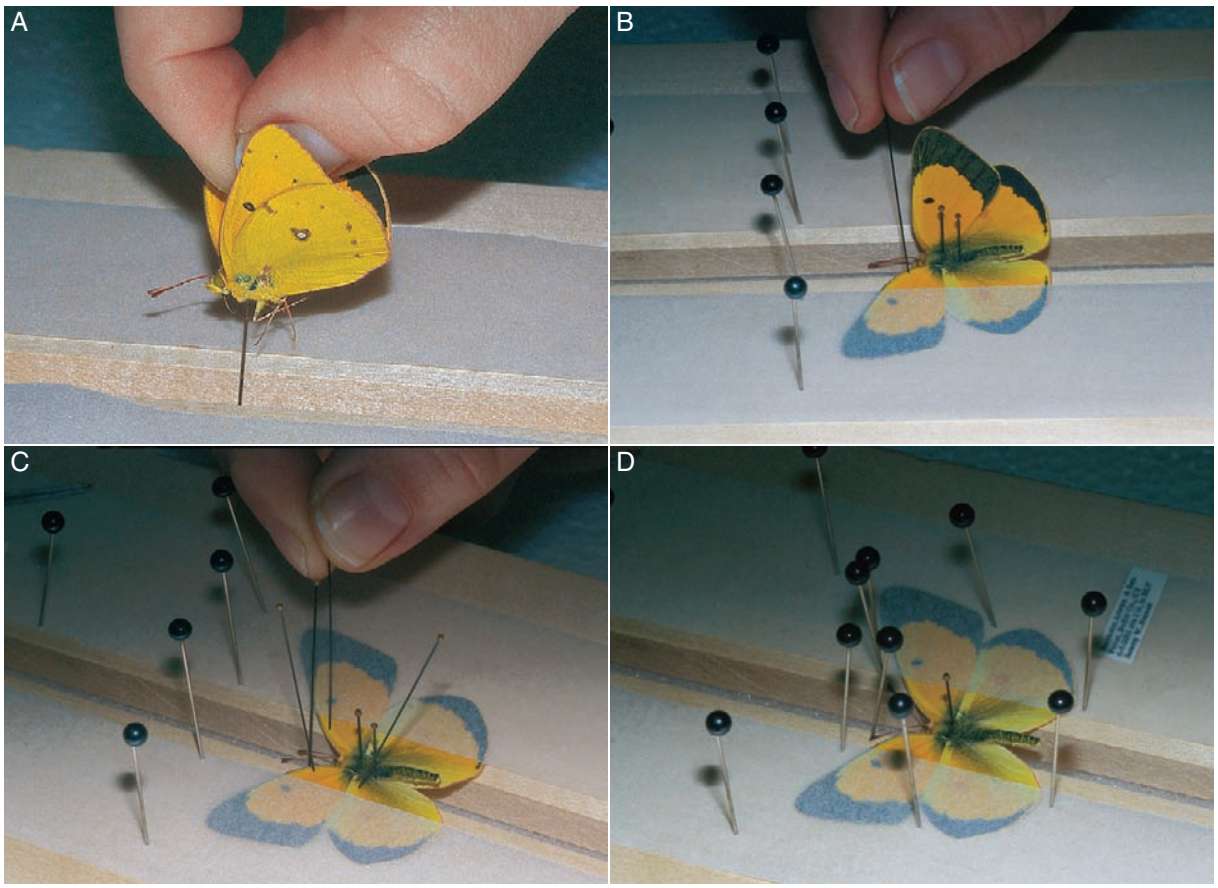

FIGURE 1 (A) Orange sulfur butterfly *(Colias eurytheme)* with insect pin inserted at proper height, ready to place in groove of spreading board. (B) Insect pin inserted behind thick costa margin and pulled forward so that inner margin of forewing is at right angle to groove. (C) After left hindwing is pulled forward and secured, right forewing and hindwing positioned to match left. (D) Glass-headed pins in proper position to hold tracing paper tight for at least one week, until the insect dries and can be removed; label ready to add.

11. Write data (where, when, and by whom collected) on the paper strip holding down the wing or make a label and tuck it under the paper strip until the specimen is taken off the board (Fig. 1D).

12. Add other specimens below, as close together as you can, if you have many specimens to spread.

13. Make a notation of the date of spreading on the paper strip to remind you of how long the specimens have been on the boards.

14. Store the board in a pest-free, dry place such as a steel or wood cabinet. Fumigation of the storage enclosure is recommended.

15. Allow specimens to dry for at least a week, longer if possible. If the abdomen is completely dry and stiff, the specimen should be ready to remove.

STORAGE OF SPECIMENS

Specimens that you would normally pin or spread after pinning can be placed in envelopes. This is known as "papering." Glassine stamp envelopes are excellent, but any kind will do. To make triangular paper envelopes, cut rectangles of paper, one side about a half-inch longer than the other. Fold into a triangle and then fold down the remaining "flaps" after putting the insects inside. Be sure butterflies and moths have wings folded over their backs for best results. They can be softened in the relaxing box at any later time. Don't forget to put collection data on the envelope.

Storage of pinned and papered specimens must be in tight containers so that museum pests such as Dermestidae (carpet beetles) and booklice (Psocoptera) cannot get to them. These can also be repelled by fumigants such as napthalene (moth flakes or moth balls), PDB (paradichlorobenzen), or dichlorvos-impregnated "strips" cut into blocks. However, the trend is away from museum fumigants because of possible health problems from exposure to them. The better method is freezing. Whole boxes can be left in a freezer for a few days on an annual basis to kill any pests that may have entered.

Drawers and boxes housing pinned specimens must have tight-fitting lids with inner flanges higher than the outer

walls of the unit. Thus, a tight seal can be achieved, which usually keeps pests out. Equipment dealers offer high quality "Schmitt" boxes and standard cabinet drawers of different dimensions (Cornell, U.S. National Museum, and California Academy types are most common), as well as cabinets to house them. Homemade boxes and cigar boxes will do in a pinch; just add a foam plastic lining. However, one cannot expect such boxes to be pest-proof without fumigation.

Vials with alcohol-preserved specimens and microscope slides can be stored in special boxes or cabinets also available from dealers or built yourself.

See Also the Following Articles

Museums and Display Collections • Photography of Insects • Population Ecology

Further Reading

Anonymous (1992). "How to Make an Insect Collection." BioQuip Products, Inc., Gardena, CA 90248.

Borrer, D. J., Triplehorn, C. A., and Johnson, N. F. (1992). "Introduction to the Study of Insects," 6th ed. Harcourt Brace, New York.

Borror, D. J., and White, R. E. (1970). "A Field Guide to the Insects." Houghton–Mifflin, Boston.

Covell, C. V., Jr. (1984). "A Field Guide to the Moths of Eastern North America." Houghton–Mifflin, Boston.

Martin, J. E. H. (1977). "The Insects and Arachnids of Canada," Part 2, "Collecting, Preparing and Preserving Insects, Mites and Spiders." Biosystematics Research Institute, Ottawa, Ontario. [Publication No. 1643]

Merritt, R. W., and Cummins, K. W. (1996). "An Introduction to the Aquatic Insects of North America," 3rd ed. Kendall/Hunt, Dubuque, IA.

Opler, P. A. (1998). "A Field Guide to the Butterflies of Eastern North America." Houghton–Mifflin, Boston.

White, R. E. (1983). "A Field Guide to the Beetles of North America." Houghton–Mifflin, Boston.

Winter, W. D. (2000). "Basic Techniques for Observing and Studying Moths and Butterflies." Lepidopterists' Society, Natural History Museum, Los Angeles, CA 90007-4057.

Collembola

(Springtails, Snow Fleas)

Kenneth A. Christiansen
Grinnell College, Grinnell, Iowa

Peter Bellinger
California State University, Northridge

Collembola or springtails comprise one of the most widespread and abundant groups of terrestrial arthropods. They are found everywhere, to the utmost reaches of multicellular animals in the Antarctic and Arctic and in all habitats except the open oceans and deep areas of large lakes. These all-wingless hexapods range in adult size from 0.4 to over 10 mm. Their small size generally results in their being overlooked, but they display an enormous range of body forms (Fig. 1), habitats, and habits. While most feed on fungi, bacteria, and decaying vegetation, some are carnivores, others are herbivores, and a number are fluid feeders. There are many commensal but no parasitic forms. They are most common in soils and leaf litter, but many species live in vegetation, littoral and neustonic habitats, caves, and ice fields or glaciers. Collembola have been classified with the insects but are now generally considered to belong to an order closely related to the Diplura and Protura. There are approximately 9000 described species belonging to about 27 families (Table I).

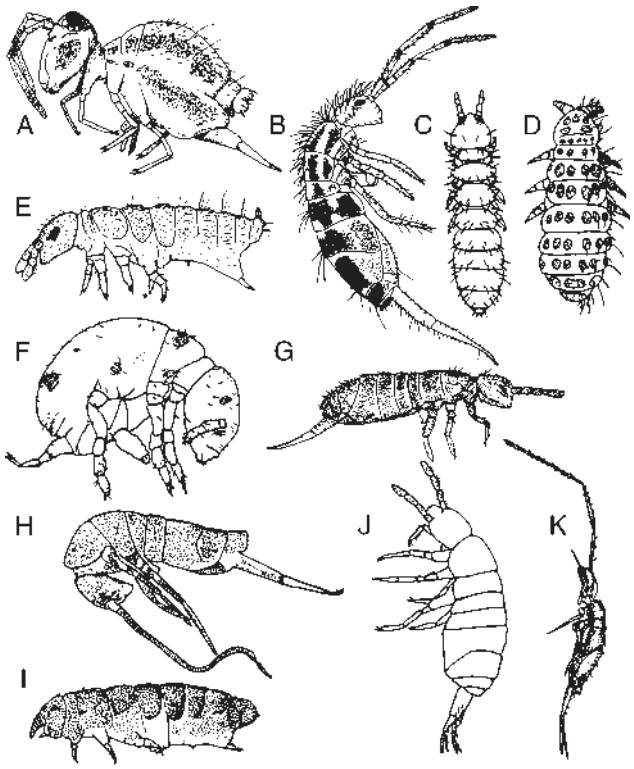

FIGURE 1 Variety of Collembola forms (not to scale). (A) Sminthuridae. (B) Entomobryidae. (C) Onychiuridae. (D) Neanuridae. (E) Hypogastruridae. (F) Neelidae. (G) Isotomidae. (H) Tomoceridae. (I) Odontellidae. (J) Oncopoduridae. (K) Paronellidae.

ANATOMY

All Collembola are primitively wingless hexapods. All have three thoracic segments and six or fewer abdominal segments, including a telson consisting of a dorsal and two ventral valves surrounding the anus.

There are typically four antennal segments, each with musculature (this distinguishes them from true insects, with three, and Diplura, with many antennal segments). Collembola vary enormously in form and somewhat in internal anatomy,

TABLE I Families and Numbers of Species of Collembola

Family	Number of species
Suborder Arthropleona	
Hypogastruidae	800
Odontellidae	150
Brachystomellidae	130
Neanuridae	1500
Onychiuridae	800
Poduridae	1
Isotogastruridae	5
Isotomidae	1500
Coenaletidae	2
Actaletidae	9
Entomobryidae	1800
Microfalculidae	1
Paronellidae	450
Cyphoderidae	185
Oncopoduridae	50
Tomoceridae	200
Mackenziellidae	1
Suborder Symphypleona	
Sminthurididae	170
Katiannidae	350
Sturmiidae	1
Spinothecidae	8
Dicyrtomidae	220
Bourletiellidae	300
Sminthuridae	270
Suborder Neelipleona	
Neelidae	30

but all lack Malpighian tubules and most have paired labial nephridia that empty into the ventral groove at the base of the labium. One universal and unique feature is the ventral tube or collophore (Fig. 2)—a distally weakly paired projection from the first abdominal segment with membranous, sometimes eversible, distal margins. Probable functions include imbibition, excretion, respiration, and adhesion to smooth surface. Collembolan mouthparts are said to be entognathous, being concealed by the head capsule, and typically adapted for chewing. The mandible usually has apical teeth and a molar

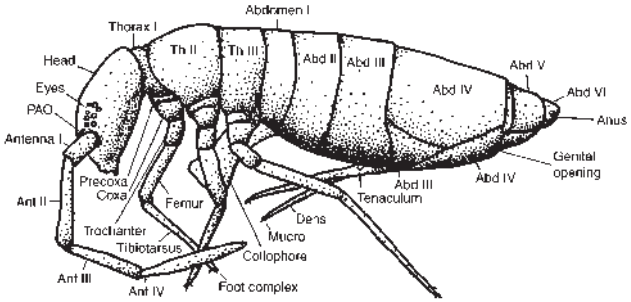

FIGURE 2 Typical Collembola anatomy.

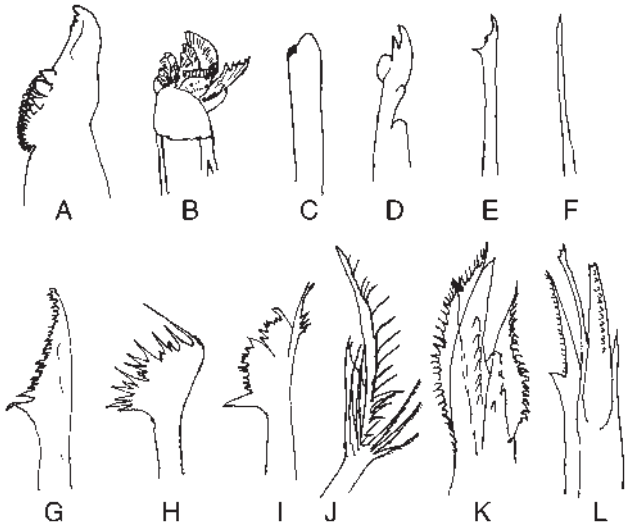

FIGURE 3 Collembolan mouthparts. (A) typical mandible and (B) maxilla; (C) reduced mandible and (D) maxilla of Cyphoderidae; (E) piercing and sucking mandible and (F) maxilla of *Neanura.* (G–I) Various mandibles of Neanuridae and (J–L) various maxillae of Neanuridae.

plate, and the maxilla varies greatly and bears a number of complex lamellae.

In some Neanuridae and a few other groups, the mouthparts are simplified and the mandible may be lost (in connection with adaptation for specialized, including liquid, diets). In other Neanuridae the mandibles and maxillae show an inexplicable complexity (Fig. 3) and diversity of form equal to that seen in any other order of insects. The mouth opening is connected to the anterior surface of the ventral tube by a ventral groove through which fluids may flow.

Collembola are equipped maximally with 8 + 8 ommatidia but often have a supplementary light sensory organ between the antennae on the dorsum of the head. A few Collembola possess rudimentary trachea; however, respiration is normally through their thin cuticle and the membranous surface of the ventral tube. The reproductive system consists of paired ovaries or testes opening on the venter on the fifth abdominal segment. Collembolan legs consist of one or two apparent subcoxal segments, a coxa, femur, trochanter, fused tibiotarsus, and distal, normally four bladed, unguis. An opposable smaller lamellate unguiculus is usually present.

Most Collembola have a forked ventral jumping apparatus or furcula on the fourth abdominal segment, consisting of a single basal manubrium and paired distal dentes and mucrones. It is held in place by the latch-like tenaculum on the third abdominal segment. When the tenaculum releases, the furcula catapults the animal, as much as 10 cm. All Collembola are covered with setae but their number, size, and structure vary greatly from group to group. The cuticle of Collembola is extremely varied and often has elaborate surface structures.

FOSSIL HISTORY

The first fossil Collembola occur in the 400 million-years-old Rhynie chert deposits of the Devonian, although there are secondary fossil hints of earlier Collembola occurrence. These fossils display very modern collembolan features, including typical entognathous, chewing mandibles; ventral tube; and, probably, a furcula. The single described species—*Rhyniella praecursor*—has been placed in a variety of families, including recently Isotomidae; however, all family placement must be considered very tentative and it is likely that one or two additional species are in this deposit. A single specimen of a very probable member of the family Entomobryidae was found in Permian shale of South Africa but extensive collembolan fossils are limited to amber of the Cretaceous, Oligocene, Miocene, and Pliocene. Collembola represent only a small fraction of the hexapods found in amber, and they are absent from many amber deposits; however, there are over 70 specimens from late Cretaceous Canadian amber, over 160 from the Baltic Eocene amber, about 130 from Miocene amber of Chiapas and the Dominican Republic, and 16 from Pliocene Japanese amber. The Cretaceous material has no identifiable specimens of extant genera and most specimens can be placed in one of eight extinct genera. All the remaining amber specimens can be placed in extant genera and in a few cases in extant species. Since the Eocene, generic extinction appears to have been absent, a unique feature among hexapods well represented in Eocene deposits.

VARIETY OF BODY FORM

Although the generally considered primitive Collembola (Fig. 2) display most of the features described above, most genera differ from this. All families have some forms with reduced numbers of eyes, and Neanuridae, Hypogastruridae, and Isotomidae (Figs. 1E and 1G) often have reduced or no furcula. The Neanuridae (Fig. 1D) often have large spines on the body as well as spectacularly complex mouthparts. Indeed these are so complex and varied (Fig. 3) that species can be identified by their mouthparts alone. The Onychiuridae (Fig. 1C) all lack eyes and almost all lack pigment and a furcula. They are characterized by the presence of pseudocelli through which defensive toxic and/or repulsive fluids are secreted. These along with the Hypogastruridae, Poduridae, and Neanuridae have well developed, seta bearing, first thoracic segments; the remaining families all have greatly reduced, nonsetaceous, first thoracic segments (Fig. 2), and some families have fusion of abdominal segments. The Neelidae and Sminthuridae have the first four abdominal segments fused and more or less fused with thoracic segments. Some Entomobryidae and Sminthuridae (as well as most Tomoceridae) have antennal subsegmentation, giving the appearance of more than four antennal segments. The largest species are found in the Neanuridae, Entomobryidae, and Tomoceridae, often reaching 5 mm and occasionally over 10 mm in length, but the Neelidae and Mackenziellidae rarely reach 1 mm.

HABITATS AND HABITS

Most Collembola in temperate and arctic zones live in the soil or ground litter, but there are several groups, most notably the Sminthuridae, that largely inhabit vegetation. In tropical regions Collembola are abundant in trees and epiphytic plants. In rain forests, they are rare in soils but abundant in trees. Collembola are abundant in many caves and are frequent in marine littoral zones. They are also common in the interstitial sand regions of marine beaches and the surface of standing fresh water. In all these examples there are many species specialized for these habitats. Collembola have recently been discovered at depths up to 20 m in both fresh and salt water, but nothing is known of the habits of such forms. Many species are found in bird and mammal nests, and microcavernicole habitats are frequently exploited but such forms show no particular specializations, being also found either in litter or in soil habitats. Ant and termite nests are frequently occupied, and one family, the Cyphoderidae, consists largely of species limited to and highly adapted for life in these habitats. Some of the most striking examples of presumed commensalism occur in the genus *Axelsonia* (Isotomidae), of which one species lives in the gill chamber of land crabs, and in the family Coenaletidae, of which all species are confined to the shell of terrestrial hermit crabs.

The forms living in the different habitats often display a suite of morphological characteristics correlated to their habitat. Thus, forms that have reduced furcula, reduced or no eyes, weak pigment, and reduced pointed tenent hairs are characteristically found in soil. Forms with no eyes or pigment; well-developed furcula; elongate, slender untoothed ungues; and reduced, pointed tenent hairs (troglomorphic) are almost always cave dwelling. Almost all species with well-

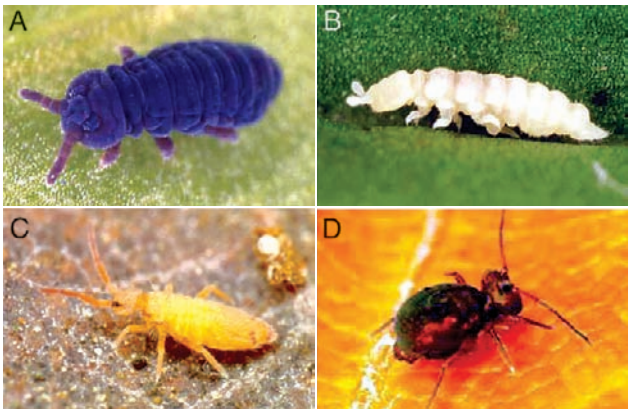

FIGURE 4 Collembolans discovered in various habitats in Reading, UK. (A) *Podura aquatica* (Poduridae), from the surface of a garden pond. (B) *Kalaphorura burmeisteri* (Onychiuridae), from soil. (C) *Dicyrtoma fusca* (Dicyrtomidae), from leaf litter. (D) *Entomobrya nicoleti* (Entomobryidae), under surface debris. (Photographs by Steve Hopkin.)

marked color patterns and well-developed furcula are either litter or vegetation dwelling (Fig. 4).

REPRODUCTION AND DEVELOPMENT

Fertilization is internal; however, exchange of sperm occurs in a variety of fashions. Sexual receptivity is associated with adult molting and in some species pheromones to facilitate aggregation of sexes. The sperm is produced in a packet, often with a stalk holding it above the substrate. In some groups (most Onychiuridae) these packets are produced randomly and fertilization occurs by accidental contact of the female with the packet of sperm. In a number of species the packets are produced only in the presence of females, but the most elaborate procedures are seen in *Podura aquatica* and the Sminthuridae. Here, often, there are elaborate courtship and maneuvering associated with fertilization. This is often accompanied by modifications in male anatomy, which ensure the appropriate species response and/or positioning for sperm packet uptake. Most of these species are brightly colored and patterned, which may also be associated with species recognition. In these forms, sexual dimorphism is the rule and often extreme. This is also true of many marine littoral species, but in these, the method of sperm transfer is still unknown and the function of the dimorphic structures (usually male) is unclear.

In some members of the family Isotomidae secondary sexual characters alternate with molts, being expressed in stages in which the animals are sexually receptive and not expressed in stages in which they are not receptive. In most Collembola there is little or no sexual dimorphism and sexes can be separated only by the difference in their genital openings. Both males and females occur in most species but parthenogenesis is common, especially in some genera of the Tullbergiidae.

Development is direct, with the young generally very similar to the adults except for the absence of sexually associated features and body ratios and some aspects of the setae clothing. The main exception to this generalization is in the Tomoceridae, whose juveniles have been assigned to genera different from those of the adults. Collembola continue to molt after reaching sexual maturity and some species can molt very large numbers of times (the record is 52). They stop reproducing at some point and later molts result in reduced rather than increased size. Although some Collembola have been known to live more than 5 years in captivity, their life span in the wild is undoubtedly much shorter.

UNUSUAL FEATURES

One remarkable feature of some members of the family Onychiuridae is that some male-only specialized setae on the venter of the abdomen achieve full development only several molts after sexual maturity. Their function is unknown.

Many species of Collembola, almost entirely of the families Isotomidae and Hypogastruridae, go through a period of reduced activity, wherein they develop a unique morphology, often associated with the development of heavy abdominal spines and wrinkled surface and reduced mouthparts and digestive systems. When this is associated with particular ecological conditions (most commonly drying or elevated temperature), it is termed ecomorphosis: feeding ceases and the structural changes are usually striking. The cessation of the causal conditions results in a quick molt and return to normal anatomy and activity. When these conditions are part of a regular cycle the process is called cyclomorphosis.

A number of Collembola are also capable of anhydrobiosis, that is, they can become completely dry without dying. In some (but not all) instances these animals forms small ball-like capsules around themselves before entering this state. If wetted, the animal resumes normal activity in an hour or two. Recent studies with sand dune Collembola suggest that this capacity may be more widespread than currently established. Another unusual feature of Collembola is the ability of some species to live very long periods without food. This characteristic appears to be best developed in some cave forms, and in several instances animals reproduced after not being fed for 30 weeks. The longest survival was a specimen of *Onychiurus,* which lived over a year without food and was then accidentally killed.

ECOLOGY AND ROLES IN ECOSYSTEM

Because Collembola are found in all habitats, from the coldest to the hottest supporting multicellular life, and from treetops to the deepest soil layers supporting multicellular animals, it is clear that their responses to various abiotic conditions must vary enormously. Humidity is usually the most important factor in determining Collembola distribution. High humidity is seldom a problem for Collembola but desiccation is often serious. Collembola resist desiccation by moving into microenvironments of high humidity (under stones or into deeper soil layers) and/or limiting activity to nights and by morphological adaptations (such as cuticular thickening, ornamentation, and scales). Some species, as already discussed, change form radically and cease feeding, while others go into anhydrobiosis. Many species lay eggs that are much more resistant to drying and they survive desiccation in this stage, often accompanying this with short postembryonic life cycles.

Collembolans have vastly different temperature tolerances and preferences, ranging from a species of *Sminthurides* found in volcanic vents with temperatures as high as 48°C to an Antarctic species shown to survive temperatures below –30°C. Survival (and activity) in low temperatures has been studied extensively. Some Collembola are primarily inhabitants of glaciers and ice fields and others are dominant members of the arthropod faunas of high latitudes. Winter-active Collembola in temperate climates often build up large numbers under snow and on suitable warm days pour out

onto the snow in vast numbers as snow fleas. Extreme cold tolerance always involves supercooling with the accumulation of cryoprotective substances.

Oxygen requirements of Collembola also vary enormously. The greatest tolerances discovered are in the Antarctic *Cryptopygus antarcticus,* which has a 30% survival rate after 30 days in pure nitrogen atmosphere. In many Collembola, respiration when submerged is via air films surrounding the animals as a result of their hydrophobic cuticle, but this apparently not necessary in all forms. In many forms the eggs are more resistant to immersion than in other stages.

Collembola, even in uniform soils, are never randomly distributed, but show strong clumping because of pheromones or local food abundance or simply as a result of limited dispersion after founding events and subsequent population growth.

Competition between Collembola species in cultures has in at least a few instances shown that there is no evidence for competitive exclusion, even under long-term clearly competitive conditions. In addition it has been shown that interactions between two species can be either positive or negative depending upon the nature of the interaction (airborne allomones, substrate-transmitted allomones, or direct contact).

While most soil- and litter-inhabiting Collembola feed primarily on decaying vegetation and fungi (and appear to be general feeders), experimental studies have shown that, given a choice, they may be very selective as to both the decay state and nature of the vegetation and the species of fungi. A number of Collembola are occasionally or primarily (and in a few species exclusively) carnivores, different species feeding on a variety of organisms, ranging from rotifers to other Collembola. Probably the most commonly eaten prey is nematodes. Vegetation-inhabiting Collembola eat primarily unicellular algae, pollen, and soft parts of vegetation and fungal spores. Many Collembola are coprophagic, feeding largely on arthropod feces. Some littoral species appear to feed largely on diatoms or unicellular algae, and forms with piercing–sucking mouthparts feed largely on fungal hyphae juices. Thus their primary role in the environment is that of reducer; however, another major role is that of prey. The ability to jump is the major defense mechanism of Collembola; however, many Poduromorpha, particularly those with the furcula short or absent, have body fluids that are repellent to predators, and they may release these by reflex bleeding when attacked. Most carnivorous soil organisms feed on Collembola, and many beetles, ants, and wasps are specialized for feeding on them.

HUMAN INTERACTIONS

Collembola rarely interact overtly with humans. There are few agricultural pests and, except for the introduced Lucerne flea *(Sminthurus viridis)* in Australia, which is a pest in pastures and horticultural crops, these are of little economic importance. There are no parasitic Collembola and they are not known to transmit any diseases. Mass emergences occur and may cause a temporary problem with household infestation but they are generally short lived there. The true household Collembola are unobtrusive and generally overlooked. Collembola play an important role in the development and maintenance of healthy soils, but this is not generally appreciated. Here they are usually abundant and may reach densities up to a trillion per square meter.

See Also the Following Articles

Arthropoda • Diplura • Protura

Further Reading

Christiansen, K., and Bellinger, P. (1998). "The Collembola of North America North of the Rio Grande: A Taxonomic Analysis," 2nd ed. Grinnell College, Grinnell, IA.

Fjellberg, A. (1998). The Collembola of Fennoscandia and Denmark. Part I. Poduromorpha. *Fauna Entomol. Scand.* **35,** 183.

Hopkin, S. P. (1997). "Biology of the Springtails (Insecta, Collembola)." Oxford University Press, London.

Lubbock, J. (1873). "Monograph of the Collembola and Thysanura." Ray Soc., London.

Maynard, E. A. (1951). "A Monograph of the Collembola or Springtail Insects of New York State." Comstock, Ithaca, NY.

Colonies

Sean O'Donnell
University of Washington, Seattle

Some species of insects spend much or all of their life living in organized social groups called colonies. Insect colonies have long fascinated biologists because they resemble superorganisms. Although insect societies are composed of distinct individuals, they possess group organization and coherence. Colonies exhibit emergent developmental properties, which are characteristics that cannot be explained or predicted by examining the behavior of their component parts. Insect colonies can serve as useful models of biological processes that occur in other complex living systems. One powerful analogy has been to compare the initiation, growth, and reproduction of an insect colony to the process of development of multicellular organisms.

Like individual plants and animals, insect colonies are initiated by propagules that are produced by parents (mother colonies); they then grow, reproduce, and often decline in old age. However, a wide array of developmental patterns have evolved in insect societies. Some of this variation can be explained by abiotic factors, such as the climate that prevails in the geographic range of a given species. Seasonality of temperature, daylength, and rainfall appear to have far-reaching effects on colony development. Climatic variables are not the whole story, however, since a diversity of colony

cycles can be found among closely related species that live in the same area. Pressure from natural enemies, such as predators and parasites, as well as pressure from social competitors, has shaped the evolution of colony development.

MODES OF FOUNDATION: INDEPENDENT, SWARMING, AND BUDDING

Parent Colony Investment Decisions

SIZE VERSUS NUMBER OF PROPAGULES Insect colonies vary widely in the amount of investment they make in each of their offspring colonies. At the low end are independently founded colonies, wherein single inseminated females (such as eusocial thrips and aphids and some Hymenoptera) initiate new colonies alone. In these species, the colony passes through a solitary phase. Examples of independent founders include sweat bees (Halictidae), bumble bees *(Bombus),* several genera of paper wasps [most Vespinae (hornets and yellowjackets), *Parapolybia,* some *Ropalidia, Mischocyttarus,* and *Polistes*], and many ants (Formicidae). In some species, the lone foundress may be later joined by one or more conspecific cofoundresses. In other species, cofoundresses are not tolerated. In many termites, the smallest possible social group founds the new colony: a single reproductive male–female pair. At the other extreme, the relatively large colonies of some species issue discrete colony-founding swarms. Swarms are made up of reproductives and workers that migrate to a new nest site as a coordinated unit. Swarms often include a sizeable portion of the worker force, and they represent a large investment. Swarm-founding lineages include honey bees *(Apis),* swarm-founding wasps (tribe Epiponini), and army ants *(Eciton).*

There is an inherent trade-off between the size of the offspring colony propagule and the number of propagules that a given parent colony can produce. Large propagules are logically restricted to species with large colony sizes, but not all large-colony species reproduce by swarming or budding. *Vespula* paper wasps, higher termites (Termitidae), and leafcutter ants *(Atta)* achieve mature colony sizes of thousands or millions of adults, yet reproduce by issuing solitary dispersing reproductives. Production of new colonies by swarms has evolved independently in bees (honey and stingless bees), paper wasps (Neotropical *Epiponini,* some *Ropalidia, Provespa,* and *Polybioides*), and ants (*Eciton* army ants). Some species of ants produce new colonies by budding, wherein portions of the colony that occupy discrete nests gradually reduce interchange of members and eventually become independent.

SURVIVAL OF PROPAGULES: PREDATORS AND ENVIRONMENTAL EFFECTS One important set of selective pressures that may explain variation in propagule size is negative biotic interactions. These can take the form of predation, attack by other natural enemies such as parasites, and conflict with conspecific competitors. Larger incipient colonies result from swarming and budding. These larger groups possess a defensive worker force and are more likely to resist destruction or consumption by enemies.

Abiotic challenges may also select for larger numbers of participants during incipient colony formation. Larger social groups may be better able to resist desiccation and temperature fluctuations, especially when they nest in enclosed spaces. Interesting in this regard are ant colonies that exhibit seasonal polydomy. Polydomy occurs when a single colony occupies several distinct nest cavities or structures. *Leptothorax* ants nest in small cavities in the leaf litter, such as hollow twigs. The colonies of some *Leptothorax* species divide themselves among several nests in summer when milder weather prevails, later coalescing into a single nest cavity as winter approaches.

Independent Foundation and Options for Social Cooperation

TO JOIN OR NOT TO JOIN In some species of independent-founding eusocial Hymenoptera, reproductives have the option of joining an already-initiated nest as a cofoundress, rather than starting one of their own. The degree of division of reproductive rights among the cofoundresses can be analyzed as a type of social contract. Often, the cooperating females are closely related. Differences in social status and reproductive capacity may be influenced by the degree of genetic relatedness among the cofoundresses. Dominant females can attempt to monopolize reproduction, or they can share a portion of reproduction as an incentive to stay and help on the part of subordinates. Kin selection theory predicts that the incipient society should be more equitable if the social partners are less closely related, since a greater incentive to help is required of nonrelatives. Cooperative colony founding may also represent a form of bet hedging and may be favored irrespective of genetic relatedness. If lone nest founders have little chance of succeeding, then cooperating can be favored by all individuals, even in the face of complete reproductive division of labor. In some cases, such as bull-horn *Acacia*-inhabiting *Pseudomyrmex* ants, female reproductives of different species may occupy a young plant, even though only one colony will eventually emerge to monopolize the tree.

USURPATION AND SOCIAL PARASITISM Another option for reproductives of some species is to steal or usurp a young colony from a conspecific or from another species. Social parasitism occurs when an invading reproductive uses the workers of a nest she did not construct to rear her reproductive offspring. A range of degrees of integration of social parasites into their host colonies can be observed in a diversity of insect lineages. Good examples occur in yellowjacket wasps (Vespinae), European *Polistes* paper wasps, bumble bees and their *Psythris* parasites, and ants. In the

simplest cases, queens attack conspecific colonies and kill the resident reproductive, taking over the worker force. Simple heterospecific parasitism is similar to conspecific takeovers, in that the invading queen kills the resident queen. Often, females of socially parasitic species exhibit adaptations to improve their chances of winning queen vs queen combat, such as enlarged heads and mandibles. Parasitic species are often incapable of producing workers of their own, so the colony switches to producing new parasite reproductives after a takeover. In some species of ants, the socially parasitic queens are better integrated into the host society (e.g., *Teleutomyrmex* invading *Tetramorium* colonies). The parasitic queens coexist with the host queen and allow her to continue to produce a worker force, while the parasites produce reproductive offspring.

Social Groups as Founding Units

DIVISION OF LABOR When new colonies are founded by swarms or by buds, a worker force is always present. One potential advantage to this strategy is the increased efficiency of the colony resulting from division of tasks among the group members. An important form of division of labor, which swarm-founders generally exploit, is the removal of the reproductives from the need to perform such risky and expensive tasks as food collection and nest defense. Division of labor is often weaker in independently founded colonies and is absent by definition for solitary foundresses.

DEFENSE A group of workers can protect incipient colonies from natural enemies. New nests that are left unattended when solitary foundresses leave to forage are often attacked by parasites and predators. Survival of colony propagules increases dramatically with group size, particularly in areas where negative biotic pressures are most intense. Several studies of independent-founding paper wasps (*Polistes* and *Mischocyttarus* spp.) have shown that young colonies with cofoundresses fare dramatically better than singly founded nests.

THE NEED FOR COMMUNICATION A special challenge facing swarm-founding species, and perhaps to a lesser extent budding species, is the need to coordinate movement from the parent nest to the offspring nest site. Special communicative mechanisms are used, such as the dance language in honey bees (*Apis* spp.), and trail pheromones in stingless bees and epiponine wasps. The need to evolve communicative mechanisms may constrain the evolution of swarming as a mode of colony foundation.

COLONY GROWTH

Social insects provide interesting and accessible models for testing life history theory because workers are roughly equivalent to the soma or body of a metazoan organism, while the reproductives can be treated as the germ or reproductive line. This analogy becomes weaker when the workers have some opportunity for direct reproduction. Nonetheless, insect colonies often develop in ways that suggest a trade-off between growth (i.e., worker production) and reproduction [i.e., production of gynes (new queens) and males].

One important decision that colonies make is the size at which to reproduce. This varies widely among even closely related species. For example, average size for mature colonies varies over at least five orders of magnitude among eusocial paper wasps (Vespidae).

Another important concern is the timing of reproduction. In some species, colonies exhibit a big-bang pattern of reproduction. The worker population of the nest increases as the growing season progresses, often exponentially. At some critical point in development the colony ceases to produce workers, switching entirely to the production of gynes and/or males. Colony decline or senescence follows reproduction. Temperate bumble bees and yellowjackets often approximate a big-bang approach to reproduction, and their life cycles resemble those of annual plants. Other species produce workers and reproductives simultaneously. In the extreme case, some males and gynes may emerge among the earliest offspring from the nest. For example, some Neotropical paper wasps (*Mischocyttarus* spp.) exhibit a great deal of overlap of worker and reproductive production. Swarm-founding species frequently undergo several bouts of reproduction, issuing reproductive swarms sequentially over a long period without undergoing parent colony decline.

For eusocial Hymenoptera, production of male offspring is potentially costly to the colony, and selection for labor efficiency may act to delay male production. Male Hymenoptera rarely work for their colonies and are often thought to represent a drain on colony resources. This cost does not accrue to termites, both sexes of which participate fully as workers.

REPRODUCTION

Timing and Synchrony

In seasonal habitats, the proper conditions for nest foundation can be constrained to a narrow window of time. This can select for a high degree of synchrony among colonies in a population in the timing of release of reproductives. In some species, reproductive offspring that depart from their natal nest must mate and either overwinter or initiate a new nest or perish. This pattern is apparently common to many ants and termites. In other species, reproductive females (honey bees, some bumble bees) and males (other bumble bees, some tropical *Mischocyttarus* wasps) can leave to find mates, but then return to the natal nest.

Sex Ratios and Sex Allocation

Beyond the germ line vs soma distinction, investment in the different sexes is an important consideration facing many

insect societies. This is especially important for eusocial Hymenoptera, the males of which perform little or no labor for their colonies to offset their production and maintenance. Sex allocation theory, which attempts to predict the optimal investment an organism should make in the proportion of male compared to female offspring, has been applied to ant colonies. There is some evidence to suggest that, as predicted, colonies alter their relative amount of investment in males and gynes, depending on such environmental conditions as food availability.

COLONY SENESCENCE

Seasonal Effects versus Programmed Senescence

Colonies of many temperate eusocial insects are annual and appear to exhibit a programmed decline and senescence. Colony decline appears to be related to queen longevity and queen condition. For example, late-season colony breakdown appears to follow a decline in the queen's ability to suppress worker reproduction or the queen's death, in temperate bumble bees, *Polistes* paper wasps, and yellowjackets. The queen is not the whole story, however. Queens lost early in the season can be replaced or supplanted by reproducing workers without colony decline. The larger worker forces that are present late in colony development may be harder for the queens or their replacements to suppress. However, closely related species in less seasonal habitats do not exhibit time- or stage-determined colony decline. The plasticity of colony development exhibited in subtropical and tropical habitats by temperate invaders, such as German yellowjackets *(Vespula germanica),* may provide valuable insights into the factors that cause colony decline. German yellowjacket colonies in invaded sites (e.g., Hawaii and New Zealand) can be polygynous, accepting new queens into established nests, and are often perennial. These colonies can grow to much larger sizes than occur in temperate habitats, and the invading populations have become serious pests.

Can Colonies Be Immortal?

When abiotic forces do not terminate colonies, their longevity can be determined by the longevity of the reproductives. Queens and nests of *Atta* leafcutter ants may survive a decade or more in the wild. If colonies can replace dying queens, there is no inherent limit on colony longevity. The polygyne (multiple queen) of the imported fire ants *Solenopsis invicta* in the United States is an example of a species that accepts new, young queens into active nests. In this case, colonies may not senesce, and the observed upper limit on colony longevity will be set by the background rate of colony mortality. In other words, the chance of colony termination may be independent of colony age. Particularly interesting in this regard are some unicolonial invasive ant species, such as the Argentine ant, *Linepithema humile.* In habitats outside their native South American range, such as the western United States, these ants fail to show internest aggression. Colony boundaries are fluid, and workers, brood, and reproductives are freely exchanged among nests. The entire population, which at present extends over a range greater than 1000 km in length, functions as a single colony. Colony longevity therefore equals the time to population extinction, and these may prove to be the longest lived insect colonies.

See Also the Following Articles

Ants • Apis Species • Division of Labor • Isoptera • Recruitment Communication • Sociality

Further Reading

Hölldobler, B., and Wilson, E. O. (1990). "The Ants." Harvard University Press, Cambridge, MA.

Jeanne, R. L. (1991). The swarm-founding Polistinae. *In* "The Social Biology of Wasps" (K. G. Ross and R. W. Matthews, eds.), pp. 191–231. Cornell University Press, Ithaca, NY.

Jeanne, R. L., and Davidson, D. W. (1984). Population regulation in social insects. *In* "Ecological Entomology" (C. B. Huffaker and R. L. Raab, eds.), pp. 559–590. Wiley, New York.

Oster, G. W., and Wilson, E. O. (1978). "Caste and Ecology in the Social Insects." Princeton University Press, Princeton, NJ.

Tschinkel, W. R. (1991). Insect sociometry: A field in search of data. *Insect Soc.* **38,** 77–82.

Colorado Potato Beetle

George G. Kennedy
North Carolina State University

The Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae), is the most devastating, defoliating, insect pest of potato *(Solanum tuberosum).* Uncontrolled, it is capable of causing complete crop failure. The potato beetle is important because of the damage it causes to potato and some related crops, as well as its extraordinary ability to evolve resistance to insecticides used in its control.

GEOGRAPHICAL SPREAD AND HOST RANGE

The Colorado potato beetle is native to Mexico. It was first recorded in the United States in 1811, feeding on a native plant, buffalo bur *(Solanum rostratum)* near the Iowa/Nebraska border. It was first reported as a pest on potato in Nebraska in 1859. The expansion of its host range to include potato allowed the beetle to spread rapidly eastward, moving among farm and garden plantings of potato. By 1874, it had expanded its geographic range to the East Coast of the United States. The potato beetle now occurs in North America throughout Mexico, the United States, and Canada, except

FIGURE 1 Adult female Colorado potato beetle depositing eggs on foliage of potato. Note evidence of feeding by the beetle at the leaflet tip. (Image © 2001–2003 www.arttoday.com.)

California, Nevada, and the coastal area of the Pacific Northwest, between 15 and 55° N latitude. It was accidentally introduced into France in 1922 and subsequently spread throughout Europe (except Great Britain) and the former Soviet Union; it now occurs in China, Greece, Turkey, and northern Iran.

The host range of the Colorado potato beetle is largely restricted to plants in the genus *Solanum* but includes some plants, such as tomato *(Lycopersicon),* in related genera. Although geographically isolated potato beetle populations vary in their ability to utilize particular plant species as hosts, potato is the preferred host for most populations.

LIFE HISTORY AND CROP INJURY

Colorado potato beetles overwinter as adults in the soil within potato fields or in field margins. There are typically one to three generations per year, depending on latitude and the availability of host plants. Adult Colorado potato beetles are oval and approximately 9.5 mm in length and 6.4 mm in width. They are yellow-orange with 10 narrow, black, longitudinal stripes on their elytra (Fig. 1). Adults typically consume 130 to 1200 mm^2 of foliage per day and are highly fecund, depositing up to 3000 yellow eggs in clusters of 10 to 50 eggs on the lower surface of host leaves over a period of several weeks. All larvae within an egg mass hatch simultaneously, typically within 4 to 14 days, depending on temperature. There are four instars, and larvae have a distinctive "hunchbacked" appearance, a black head capsule, and two rows of black spots on each side of the body. Instars 1 and 2 are brick red, whereas instars 3 and 4 are pink to salmon. The larvae are voracious feeders, with fourth instars consuming as much as 500 mm^2 of potato foliage per day. Larval development requires as little as 8 days or as long as 28 days at average temperatures of 29 and 14°C, respectively. Mature fourth instars burrow into the soil where they pupate. The pupal stage typically lasts 8 to 18 days, depending on temperature.

The Colorado potato beetle is primarily a pest of potatoes, but in some locations is also a pest of tomato *(L. esculentum)* and eggplant *(Solanum melongena).* Damage results from defoliation by adult and larval feeding. In potato, yield reductions are related to both the amount of defoliation and the stage of plant growth during which it occurs. Yield reductions in tomato and eggplant result from feeding injury to the fruits, as well as from defoliation.

MANAGEMENT

Although a number of cultural measures, including crop rotation, isolation from previous potato crops, planting of nonpreferred and early maturing potato varieties, and use of trap crops, were recommended measures for potato beetle control, hand removal of adults, eggs, and larvae from infested plants was the primary means of control prior to the introduction of the insecticide Paris Green (copper acetoarsenite) in the late 1800s. Arsenic-based insecticides remained the primary means of control until DDT replaced them in the late 1940s. Resistance of the Colorado potato beetle to DDT was first reported in New York in the early 1950s. Resistance to other chlorinated hydrocarbon insecticides soon followed throughout much of the potato-growing region of the eastern United States. A series of insecticides was used to control the beetle during the succeeding decades and the potato beetle developed resistance to each. By the early 1980s, insecticide resistance had reached a crisis level. In many locations, potato beetle populations could not be controlled using insecticides. This stimulated a burst of research activity, which resulted in the development of more holistic pest management approaches. These involved foliar applications of the bacterial pathogen *Bacillus thuringiensis tenebrionis,* crop rotation, naturally occurring biological control, scouting and the use of economic thresholds, and the use of narrow-spectrum insecticides. By the late 1990s, several new, highly effective, narrow-spectrum insecticides had become available to control resistant potato beetle populations. Currently, potato beetle management relies on these new insecticides but heavily emphasizes their use within a pest management context, which is designed to minimize selection for insecticide resistance and negative environmental impacts.

During the mid-1990s, transgenic potato varieties were commercialized that expressed a protein from *B. thuringiensis*

tenebrionis, which is highly toxic to the Colorado potato beetle. These varieties produce high-quality potatoes and are highly effective in controlling the potato beetle. Nonetheless, they have received only limited use because of their inability to compete with insecticides that controlled other insect pests (aphids and leafhoppers) in addition to the Colorado potato beetle and because of concern that consumers would not buy potato products made from transgenic potatoes. It is not clear at this time whether transgenic potato varieties will play a significant role in the future management of the Colorado potato beetle.

See Also the Following Articles

Agricultural Entomology • Coleoptera • Insecticide and Acaricide Resistance • Integrated Pest Management • Plant–Insect Interactions

Further Reading

Bishop, B. A., and Grafius, E. J. (1996). Insecticide resistance in the Colorado potato beetle. *In* "The Classification, Phylogeny and Genetics" (P. H. A. Jolivet and M. L. Cox, eds.), Vol. 1 of "Chrysomelidae Biology." SPB Academic Pub., Amsterdam.

Casagrande, R. A. (1987). The Colorado potato beetle: 125 years of mismanagement. *Bull. Entomol. Soc. Am.* **33,** 142–150.

Ferro, D. N. (2000). Success and failure of Bt products: Colorado potato beetle—A case history. *In* "Emerging Technologies for Integrated Pest Management: Concepts, Research, and Implementation" (G. G. Kennedy and T. B. Sutton, eds.). APS Press, St. Paul.

Hare, J. D. (1990). Ecology and management of the Colorado potato beetle. *Annu. Rev. Entomol.* **35,** 81–100.

Lashomb, J. H., and Casagrande, R. (eds.) (1981). "Advances in Potato Pest Management." Hutchinson Ross Pub., Stroudsburg, PA.

Zehnder, G. W., Powelson, M. L., Jansson, R. K., and Raman, K. V. (eds.) (1994). "Advances in Potato Pest Biology and Management." APS Press, St. Paul.

Coloration

Helen Ghiradella
State University of New York, Albany

Coloration is, as the word implies, the tapestry of hues with which an organism arrays the surfaces that it presents to the world. The signals thus produced may aid in species identification, camouflage, warning, and temperature regulation; all in all, they serve as a mute "language" with which an individual organism may communicate its place in the community within which it lives.

Insects are master chemists whose virtuosity is particularly evident in the design of the cuticle, the nonliving material that makes up the exoskeleton and serves as the boundary between the living animal and the outside world. Cuticle, a composite of chitin fibrils and various proteins and lipids, can be tailored for strength, rigidity, flexibility, permeability, or elasticity, as needs dictate. It is also a technical and artistic medium with which insects, who are also master physicists and optical engineers, manipulate light to attire themselves with brilliant color on their bodies and wings. This article briefly reviews the bases of this ability. It begins, however, with an overview of the physics of color production, particularly with respect to structural colors, because only with this background can the reader really appreciate what a biological system, in its handling of light and color, can do.

TYPES OF COLOR

"Light" by definition involves wavelengths within the visible part of the electromagnetic spectrum. For humans it consists of wavelengths ranging from approximately 400 nm (violet) to approximately 725 nm (red). Many organisms, including insects, extend this range into the near ultraviolet (300–400 nm). "White" light for a particular organism consists of all wavelengths visible to that organism. Colored light has an incomplete spectrum in which only some wavelengths are represented.

Matter interacts with white light in various ways to produce color. One way is by selective absorption of particular wavelengths by a chemical, or pigment. The absorbed wavelengths (which are determined by the pigment's molecular structure) are essentially subtracted from the total spectrum, whereas the rest are reflected or transmitted to produce the visible color. Because pigments subtract colors, as additional pigments are added to a mix, additional wavelengths are absorbed and lost to view, changing the perceived color. When all wavelengths of the visible spectrum are absorbed, we call the sensation "black." (This is a somewhat simplified view: visual physiologists and psychophysicists would point out that additional processing by the visual system tempers what humans actually "see.") Pigmentary colors may be found in the cuticle or, if that be transparent, in the underlying tissues and even in the gut contents.

A second basis for color is structural, caused by the interaction of white light with minute and precise arrays on or in the material. The effects depend on the architecture, rather than the chemical makeup of the material. Light may be reflected, refracted, or scattered, but it is not absorbed, and so structural colors are "additive": if two are combined, both sets of wavelengths are represented in the final effect. If all wavelengths of the visible spectrum are reflected, we call the sensation "white." (Technically, white, even if caused by a pigment, is always a structural color, because it is the absence of any absorption of light.) Because the underlying architecture must generally be precise and stable, most structural colors are typically produced by stiff, nonliving materials, and of these insect cuticle is literally a brilliant example.

In biological systems, pigmentary colors are more common in the "warm" range—red, orange, and yellow—although green and blue pigments do exist. Biological

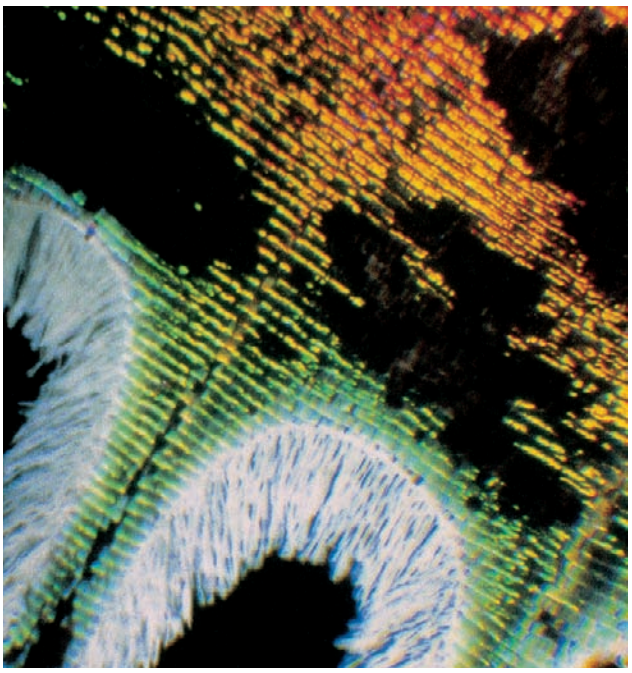

FIGURE 1 *Uranus riphaeus,* portion of hind wing, showing the typical lepidopteran investiture of shingle-like scales (on the surface) and bristles (at the edges). The scales in the black areas are colored by a pigment, probably melanin, whereas the iridescent scales and the white bristles are structurally colored.

structural colors, in contrast, are more likely to be "cool"—green, blue, violet, and ultraviolet. Figure 1 shows part of a butterfly wing: the dark colors are pigmentary, whereas the iridescent colors and the whites are structural. Many insects display both types, which are sometimes used together to produce yet additional effects. For example, a structural blue may be added to a pigmentary red to make a luminous violet, or a structural color may be "deepened" or intensified by a "backing" pigment that absorbs stray light leaking in from the "wrong" direction.

This article considers both pigmentary and structural colors. The following is a review of insect pigments, abstracted from the reviews of Chapman, Fox, and Nijhout.

INSECT PIGMENTS

Insects can make most of their pigments (some apparently from waste products that were historically simply stored or excreted), whereas others must come from their diets. Several general classes of pigments are recognized. These differ in the color ranges they generate and in the precursors used to produce them. As they share the same underlying mechanism of color production (selective absorption of some wavelengths of light), they can be reviewed with a simple list.

Melanins are black, brown, tan, or reddish brown pigments whose production and deployment involve a complex system of gene products and biochemical pathways. They are often present as granules in the exocuticle, although in lepidopteran scales they may be diffusely distributed, and they are responsible for most of the dark patterning in the body and wings. Eumelanin, the black form, commonly requires dopamine and tyrosine as precursors, while the chemistry of phaeomelanin, the brown, tan, or reddish brown form, is less well understood and may require the incorporation of additional kinds of molecules into the compound.

Pterins are white, yellow, or red pigments derived from a purine, guanosine triphosphate. Some function as cofactors of enzymes important in growth and differentiation; they may help control these processes. They are also cofactors in ommochrome (see later) production and often occur with these latter pigments, for example in the screening-pigment cells in the ommatidia of the eyes.

Ommochromes are red, yellow, or brown pigments derived from tryptophan, which they may serve to use up if it is in excess supply during times of high protein turnover (e.g., in metamorphosis). They usually occur in granules coupled with proteins and, as mentioned above, are present as screening pigments in the eyes as well as in the colors on the body. In insects displaying Tyndall blue (see later), they may serve as background pigments to absorb extraneous light.

Tetrapyrroles are pigments commonly classified into two groups. The first, the ring-shaped porphyrins, may add and incorporate iron to become hemes, which in turn may link to proteins to become (1) cytochromes, proteins important in cellular respiration in all higher organisms, or (2) hemoglobin, the protein that vertebrates and other organisms use to facilitate oxygen transport to their cells. Of necessity, all insects make cytochromes. Some that live in habitats of very low oxygen tension may make hemoglobin as well.

The other class of tetrapyrroles, the bilins, may in themselves be green or may link with proteins to make blue chromoproteins. These may in turn link with carotenoid pigments (see later) to make many insect greens.

Papiliochromes are yellow and red/brown pigments found only in butterflies of the family Papilionidae.

Quinone pigments are pigments of uncertain origin found in the Homoptera. Anthraquinones are found in members of the family Coccidae, in which they give red and sometimes yellow coloration; these include cochineal dye of historical importance. Aphins are characteristic of aphids, to whom they impart a purple or black coloration.

Carotenoids are yellow, orange, red, and, if bound to the appropriate protein, blue pigments that are made from dietary carotenes and their oxidized derivatives, xanthophylls. In combination with blue pigments (often bilins) they may produce an insect green, insectoverdin. They are also sources of retinal, a component of the photopigment of the eye.

Flavonoids are plant-derived pigments that produce cream or yellow colors, particularly in the Lepidoptera. Like the carotenoids, they cannot be synthesized but must come from the diet.

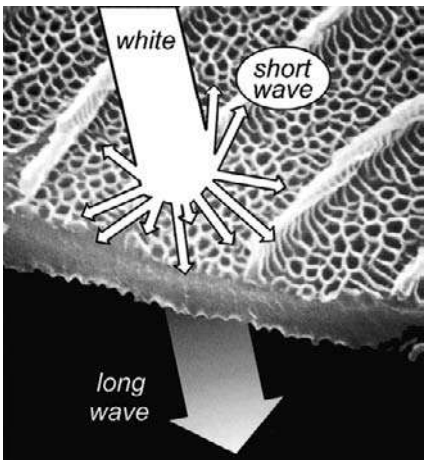

FIGURE 2 Scattering of light to produce Tyndall blue. When full-spectrum (white) light encounters structures or particles of the right dimensions, the shorter wavelengths are preferentially scattered in all directions, including toward the eye of the observer, who sees a blue color. The longer wave light passes through unscattered (and therefore bypasses the observer).

STRUCTURAL COLORS

There are many mechanisms by which structural colors can be produced. All depend directly or indirectly on the fact that a particular piece of material scatters or refracts different wavelengths of light to different degrees. This property of the material can be expressed in terms of its index of refraction, *n,* a measure of the degree to which a given wavelength of light entering the material is "retarded" or slowed down. For insect cuticle, *n* typically ranges from 1.5 for long-wave (red) light to 1.6 for short-wave (UV) light, although in special cases *n* less than 1.4 has been reported (for comparison, *n* for air is by definition 1). Structural colors described so far in biological systems fall into two general classes, scattering and interference.

Scattering

Scattering of light occurs when white light encounters a distributed cloud or array of molecules, particles, or other structures (Fig. 2). At least some of the component wavelengths of the beam will be reflected in random directions, including toward the observer. If the scattering agents are relatively large (700 nm or more), all visible wavelengths are scattered, and the resulting color is a matte white (the color of whole milk is an example of such scattering). If the particles are smaller (in the 400 nm range), the short wavelengths are scattered to a much greater degree than the long ones, which tend to pass on through the system and not reach the eye of the observer. The resulting color, Tyndall blue, is commonly seen in blue eyes and bluejay feathers; in insects it occurs in blue dragonflies and in some blue butterflies. Often, the blue structure is underlaid by a layer of ommochrome pigments, which, as mentioned above, deepen and intensify the color by absorbing stray light. Lacking such pigment backing, the blue is a dilute "powder" blue.

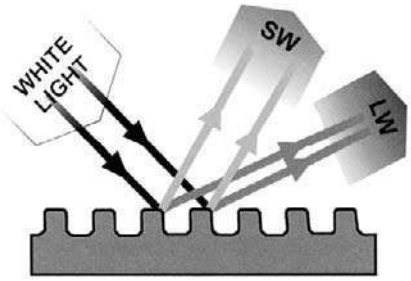

FIGURE 3 Diffraction (in this example, from a grating). Light hitting an edge or discontinuity gets bent or refracted to different degrees, depending on its wavelength. When it is then reflected, which of the component wavelengths are reinforced varies with the position of the observer, so that from one angle shorter wave light (SW) predominates, whereas from another, longer wave (LW) light predominates.

Interference

The general category of interference includes those situations in which the rays of a beam of white light are temporarily separated and then brought back together in such a manner that some have traveled a longer path than others. Depending on geometry, when the rays recombine, certain wavelengths are in phase and reinforced ("constructive interference"), whereas others are out of phase and cancel each other ("destructive interference"). The results are the brilliant, shimmering colors we call "iridescent." There are many ways of producing iridescence; this article considers only those of known importance in insects.

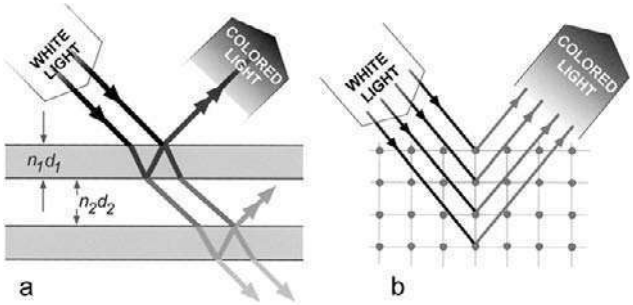

FIGURE 4 Two forms of interference from layers. (a) Thin film. A thin film can be described in terms of its optical thickness, its index of refraction, *n,* times its actual thickness, *d.* When white light encounters such a film, part of the light reflects from the top surface and part from the bottom. When these two beams recombine, those wavelengths four times the optical thickness of the film are constructively reinforced and the others not. If many films are stacked, light not reflected by the first film may be so by the others; if the films are alternated with others of equal optical thickness but of a different refractive index (so that $n_1d_1 = n_2d_2$), the stack reflects essentially all light of the reinforced wavelength. (b) Lattice. A lattice of points, spheres, or other structures reflects light in a manner analogous to that of a crystal. Each plane reflects part of a beam and transmits the rest (transmitted light not diagramed here). If the planes are evenly spaced, they reflect light the wavelength of which is twice the spacing, i.e., they will form a half-wave reflector. As in the case of thin films, with enough reflective planes, essentially all the light of the reinforced wavelength will be reflected.

DIFFRACTION Diffraction occurs when light strikes the edge of a slit, groove, or ridge. Different wavelengths bend around the edge to different degrees and the spectrum fans out into its components. If many such grooves or ridges occur in a regularly spaced array (for example, a "diffraction grating" such as that in Fig. 3) light of different wavelengths is reinforced at different angles so that the colors change with the position of the viewer (e.g., consider iridescent bumper stickers and other shimmering plastic labels). Many insect cuticles have fine gratings etched into them; these and the ridge and crossrib structures (see later) of some lepidopteran scales and bristles produce diffraction colors.

THIN-FILM INTERFERENCE Thin-film interference involves, as the name implies, the interaction of light with ultrathin films of a material (e.g., iridescence from soap bubbles and oil slicks). Light reflecting from the top surface of such a film interacts with that reflecting from the bottom surface (Fig. 4a) and depending on the optical thickness of the film (its index of refraction, n, times its actual thickness, d), some wavelengths are reinforced and others not. Because the wavelengths of the reinforced light are four times the optical thickness of the film (i.e., a film of 100 nm optical thickness results in reflected light of wavelength 400 nm), such films are commonly called "quarter-wave interference reflectors" or "quarter-wave films." Because a slanted beam of light has to penetrate a greater thickness of film, thereby changing the effective optical geometry, thin-film colors shift toward the shorter wavelengths when the films are tilted with respect to the light source (e.g., the familiar blue of the morpho butterflies becomes more violet).

Of course any film thin enough to act as a quarter-wave reflector can catch and reflect only a portion of the incident light; the rest passes through. The presence of other films below the first increases the likelihood that light will be reflected, and in fact the most efficient of these reflector systems are stacks of thin films of the material in question, separated by other films with a different refractive index or by air (n = 1), so that the light is reflected from layers of alternating high and low n. If all the films are equivalent in nd, their optical thickness, the emerging colors are relatively pure, whereas varied spacing produces a less intense but broader range of reflection. As in all these systems, there may be behind the "mirror" a layer of pigment that intensifies the color by eliminating stray light that would otherwise interfere with the efficiency of the interference and thereby dilute the color.

LATTICES Many iridescent colors are produced, not by thin films per se but by thin-film analogues, systems that achieve similar effects without actual discrete films. One such mechanism is the "Bragg" or space lattice (Fig. 4b), a highly regular array of spheres or other units. Light entering such a lattice is reflected from the various layers, and the beams interfere in a manner analogous to that in thin-film stacks. In this example, the wavelength reinforced is twice that of the spacing between the layers of the lattice, which therefore acts as a half-wave reflector. The familiar brilliance of the mineral opal is an example of this type of interference, caused in this instance by a lattice of tiny silica spheres. These lattices are very common in the biological world; those described so far in insects are "reverse" lattices, consisting of spheres of air in a matrix of cuticle.

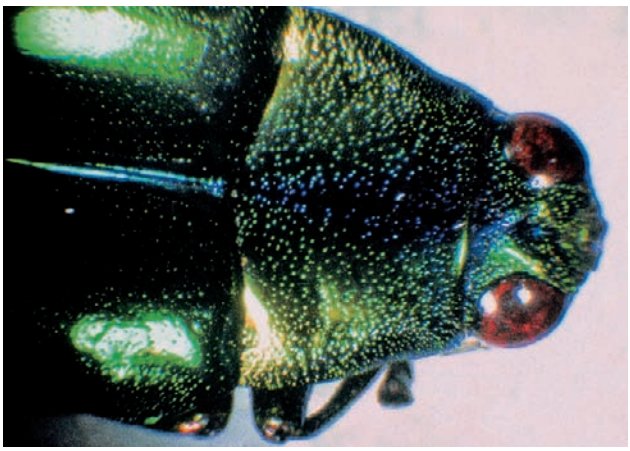

FIGURE 5 This beetle shows the metallic coloration typical of many beetles and flies. The colors have at least two possible origins: they may be caused by a thin-film stack in the exocuticle (or sometimes the endocuticle) (Fig. 6d) or they may be the result of a helicoidal arrangement of chitin fibrils in the exocuticle (Fig. 6e). The latter effect is analogous to that produced by certain types of liquid crystal in common technological use. The red and black coloration in the eyes, on the other hand, is almost certainly pigmentary.

HELICOIDS The metallically colored cuticles of many beetles and flies (Fig. 5) either are thin film (Fig. 6d) or owe their iridescence to yet another mechanism, one analogous to that shown by the familiar and brightly colored liquid crystal displays in our electronic world. Cuticle is of course a composite of chitin fibrils in a complex matrix that is laid down sequentially in what can be considered a series of layers. If the fibrils in a particular layer are lined up in the same direction, the layer exhibits form birefringence, i.e., different indices of refraction parallel to and normal to the fibrils. In many cuticles, the layers precess, that is, each is laid down slightly rotated relative to the previous one (Fig. 6e). In essence, the structure can be considered a helicoid, and like all helical structures, it repeats itself with a certain spacing (called a "pitch"). As the layers precess, so does the difference in refractive index, so that viewed from a given direction a helicoidal array displays what are essentially layers of alternate high and low n, reminiscent of those in thin films. (Unlike thin films, helicoids also circularly polarize light, which insects may be able to see and which may therefore carry additional information to them.) If the spacing is regular and the pitch is appropriate, the helicoid behaves like a half-wave interference reflector, i.e., it reflects light of wavelengths twice the pitch. In the typical metallic cuticles,

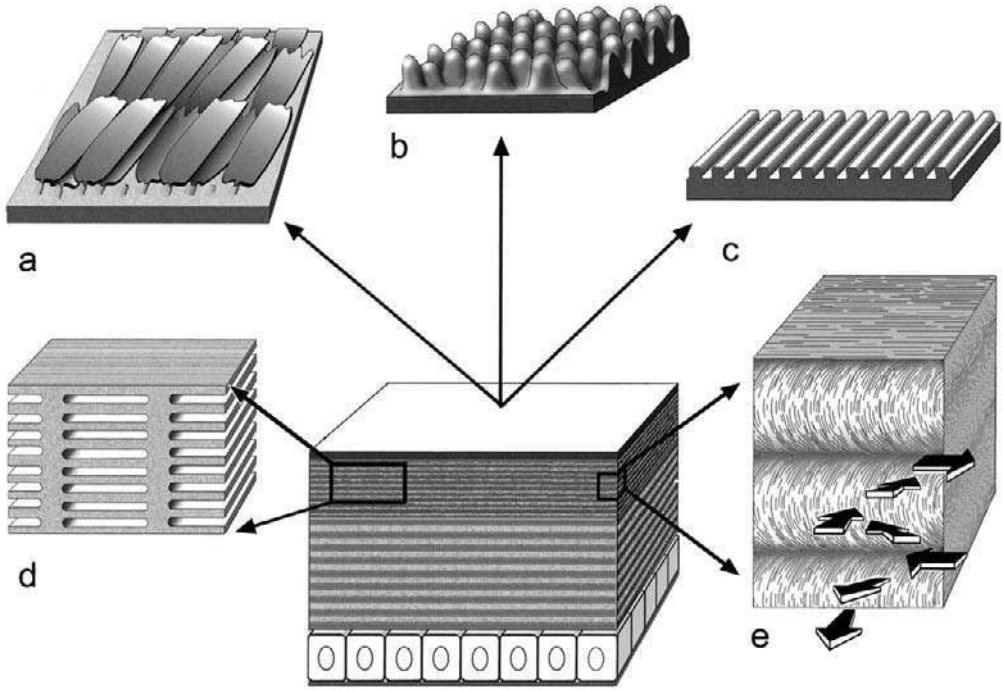

FIGURE 6 How to make an interference color. A block of hard insect cuticle (bottom center) typically consists of a relatively thin epicuticle (here represented as a featureless covering layer) and an inner procuticle, which in turn consists of a distal exocuticle and an inner endocuticle (this diagram also shows the attendant epithelial cells). The layering of the procuticle is common in most (but not all) cuticles and is the visible manifestation of the helicoidal architecture of the chitin fibrils. Such a block of cuticle may be modified in any of several ways to manipulate light: (a) The surface investiture (the scales and/or bristles) may be modified to produce scattering or iridescent colors (see Fig. 7). (b, c) The cuticle surface may be sculpted into fine protuberances that serve as an antiglare coating (see Figs. 8 and 9) or into fine parallel grooves that act as diffraction gratings. (d) Part of the procuticle may be elaborated into a quarter-wave thin-film reflector stack. (e) The chitin fibrils of the exocuticle may be arranged in a "helicoidal" array, analogous to that in a liquid crystal and producing color by a similar mechanism. (The apparent parabolic bending of the fibers is an optical illusion.)

the helicoids of the exocuticle are so tuned, and because the helicoidal arrangements of their fibrils resemble that of the molecules in one iridescent class of liquid crystals, they are often referred to as "liquid crystal analogs." Some insects intensify the effect by doping the cuticle with uric acid, which increases its birefringence.

BASES OF STRUCTURAL COLORS IN INSECTS

As mentioned above, because scattering colors (whites and Tyndall blues) can be produced by granules or droplets as well as hard structures, these may come from the epidermis and internal tissues, as well as from the integument. Interference colors, which require stable structures to produce them, are limited to the cuticle and its investiture. Figure 6 shows diagrammatically a patch of cuticle with its two basic layers, the thin outer epicuticle and the inner procuticle. The procuticle commonly shows the helicoidal arrangement described above, which results in a banded or layered appearance in section. In hard or stiffened cuticle, the procuticle is commonly further subdivided into a cross-linked, more tightly woven distal exocuticle and a basal, more loosely structured endocuticle.

The cuticular surface and the exocuticle are most likely to be modified to produce structural colors, although in some insects the endocuticle may be as well. Several possibilities exist (Fig. 6). For example, the surface may be invested with layers of scales and/or bristles (Fig. 6a), which carry the color, especially in the Lepidoptera (see later). Alternatively, it can be sculpted into a series of nipple-like protuberances (Fig. 6b—more about this later) or into the fine grooves that characterize diffraction gratings (Fig. 6c). In the exocuticle (and sometimes the endocuticle) metallic colors can be produced by stacks of thin films of alternating refractive indices (n = 1.58 alternating with n = 1.38 has been measured in one of these systems) (Fig. 6d) or by appropriately tuned helical rotation of the chitin fibrils (Fig. 6e). (As yet another example of insect command of light, in many corneas, the helicoidal architecture of the cuticle is tailored not to produce structural colors but to control refractive index, so that incoming light is appropriately focused as it enters the ommatidia.)

Scales and bristles are particularly impressive in the variety and complexity of their architecture (Fig. 7). They commonly exist in two and sometimes three layers on the body or wing surface (Fig. 7a), and each layer may be modified in shape and color. A typical scale consists of a flattened sleeve of cuticle whose lower surface (that toward the wing) is relatively featureless, whereas the upper surface is elaborated into a reticular network of longitudinal ridges joined at intervals by

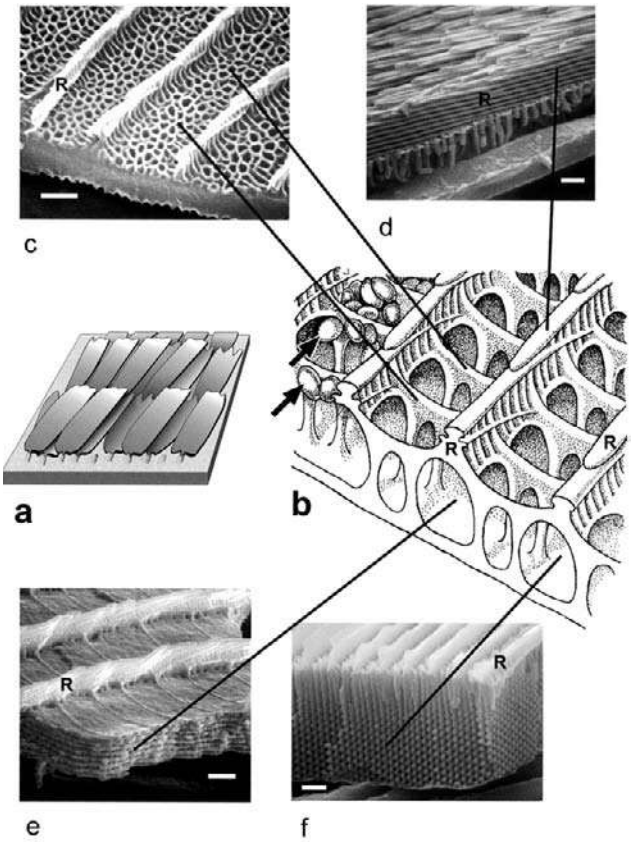

FIGURE 7 Closer look at the investiture (scales and bristles), which in the Lepidoptera typically carries the color. Scales and bristles are complex cuticular structures each elaborated by a single cell, and they are often both pigmentarily and structurally colored. (a) As in Fig. 6a, a patch of cuticle surface showing several overlapping scales and one empty socket. (b) Diagrammatic view of a small fragment of a more or less typical unspecialized scale. The scale may be thought of as a flattened sac, the two surfaces of which are joined by fine pillars. (A bristle would be cylindrical, rather than flattened, but it is essentially the same type of structure.) The upper surface is a rectangular grid made up of longitudinal ridges (R) joined at regular intervals by transverse crossribs from which, in some species, hang pigment granules (arrows) (in other insects, pigment is incorporated into the cuticle itself). Ridges and crossribs together frame a series of windows opening into the interior of the scale. Virtually any part of this basic scale may be elaborated into a reflective structure. In the following examples, scales have been fractured to show their interior structures; lines indicate which basic scale structures have been elaborated to produce each structural color. [Modified from Ghiradella, H. (1998). Hairs, bristles and scales. *In* "Insecta." (M. Locke, ed.), vol. 11A of "Microscopic Anatomy of Invertebrates" (F. W. Harrison, ed.) pp. 257–287. Wiley, New York. Copyright 1998 John Wiley & Sons. Reprinted by permission of John Wiley & Sons.] (c) *Papilio zalmoxis*, fragment of upper scale surface. The ridges are low and unornamented, but the crossribs have "filled in" the windows with a network of "alveolae" that scatter light to produce a Tyndall blue color (compare Fig. 2). Bar, 1 μm. (d) *Morpho menelaus,* fragment of deep blue iridescent scale, fractured longitudinally to show a side view of a ridge (R), together with the pillars that join it to the bottom layer of the scale. The ridge (and those behind it) has been elaborated into stacks of slanting thin films that reflect the characteristic blue of this butterfly. Bar, 1 μm. (e) *Urania riphaeus,* fractured green iridescent scale (see Fig. 1). The ridges and crossribs are not particularly elaborate, but the interior of the scale is filled with a stack of thin films that produce the color. Bar, 1 μm. (f) *Teinopalpus* sp., fractured green iridescent scale. The scale interior is filled with a space lattice that produces the color. Bar, 1 μm.

FIGURE 8 Effects of an antiglare coating. Even though the wing of this clearwing moth is somewhat wrinkled and parts of it would therefore be expected to reflect light, its matte surface (Fig. 9) allows the text to be read through it with minimal loss.

transverse crossribs (Fig. 7b). Fine flutings or microribs line the sides of the ridges and sometimes run out across the crossribs. Slender pillars join top and bottom surfaces. Pigments in some groups (typically the Pieridae) may exist in discrete granules, whereas in other insects they are laid into the scale cuticle itself.

Virtually any part of this basic scale can be elaborated to produce a structural color. The spacing of the ridges and/or crossribs may be appropriate to produce diffraction colors. The crossribs and microribs may extend to fill in the windows with a network of "alveolae" that reflects Tyndall blue (Fig. 7c). The scale ridges may bear stacks of thin films (examples known so far reflect green, blue, or ultraviolet) (Fig. 7d). The interior of the scale may be filled with stacks of thin films tuned to produce green or blue (Fig. 7e), or it may contain a space lattice that reflects iridescent green (Fig. 7f). And, as mentioned earlier, these structural colors may be combined with pigments to give yet additional colors and effects.

More detailed study of some of these systems is revealing yet more complicated and sophisticated optical effects. For example, in blue Morpho butterflies, the deep blue iridescent scales (whose color comes from thin-film iridescence on the ridges—Fig. 7d) are overlaid by a layer of "glass scales," which, though otherwise transparent, do have iridescent ridges. The apparent function of the glass scales is to broaden the effective angle of reflection (see Vukusic *et al.,* 1999). The iridescent scales of *Papilio palinurus* have stacks of internal thin films, but rather than being flat, the stacks are puckered into shallow cup-shaped depressions whose bottoms reflect yellow light, whereas the sides reflect blue, giving the human observer the sensation of green (see Vukusic *et al.,* 2000)—it is not known why these animals have developed this mechanism to produce green scales when other iridescent greens are produced by more conventional thin films or by lattices. There are other intriguing scale and bristle types whose optics are now being studied, and from these insect systems new and sophisticated insights into the effective control of light can be expected.

ANTIGLARE COATINGS

Insect handling of light does not stop with the production of colors. Figure 8 shows the clarity with which light may be transmitted through the wing of a clearwing moth. Although

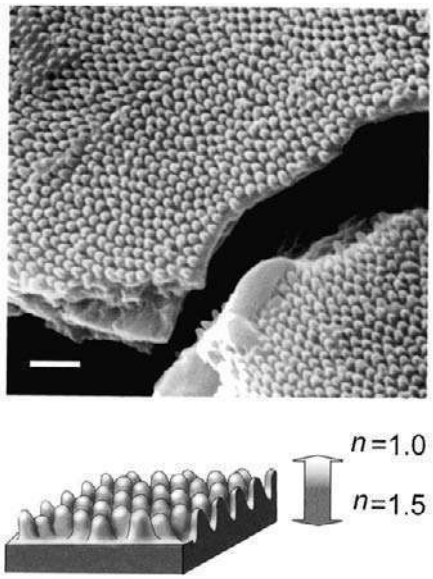

FIGURE 9 (Top) *Podesesia syringae,* patch of wing fractured to show its internal structure as well as the fine protuberances or nipples that form the antiglare coating. A few of those on the wing reverse show through the break at bottom center. Bar, 1 μm. (Bottom) Basis of the antiglare effect. The tapered shape of the protuberances produces a gradual change in refractive index from that of air (n = 1) to that of cuticle (n = 1.5 in this example), so that at the interface there is neither refraction nor reflection to disturb the passage of light.

the cuticle is somewhat rippled and one would expect some of its surfaces to show glare, they do not. Figure 9 shows why: the wing surface is covered by fine arrays of protuberances that are commonly found on cuticles that are engineered not to maximize the reflection of light but to minimize it (besides the wings of these clearwing moths, such arrays have been reported on the eye corneas of nocturnal moths). The arrays provide a gradual transition in refractive index from that of air (n = 1) to that of cuticle (typically n = 1.5 to 1.6) so that there is no sharp interface to refract or reflect light as it passes from one phase to the other (the basis for antiglare coatings on eyeglasses).

In summary, the complexity associated with insect colors extends, for pigments, to the sophisticated biochemistry with which insects make (and often recycle) the compounds that characterize their chemical colors and, for structural colors, to the production and control, often by single cells, of the precise cuticular architecture reviewed here. Other effects abound; for example, many insects are capable of physiological color change, by reversibly hydrating or dehydrating their cuticles to change the optical thickness of the layers or by moving pigment about.

PERSPECTIVE

An easy question is why such arrays of color? Insects share the same challenges as humans and so they use color and patterning for species and mate recognition, camouflage, startling potential predators, and mimicry. Energy is also almost certainly a factor: dark colors absorb more heat, and butterflies, for example, may use pigments and possibly interference mechanisms to increase the absorption of infrared. It can only be speculated as to why structural colors predominate at the short end of the visible spectrum. As a biological material, cuticle is assumed to have a limited range of refractive indices and if so, only shorter wavelengths may be refracted and scattered effectively enough to produce the needed effects. It could also be that short-wave structural colors are metabolically "cheaper" (i.e., require less energy to produce) or easier to make than short-wave pigments, which do seem relatively rare in biological systems. Further study may enlighten both biologists and engineers.

How are these structures and color patterns made? On one time scale the question is developmental: how can an animal transform its genetic information into the complicated structures observed? This is the general question of pattern formation, the nested series of instructions that must be carried out by a developing organism on many levels at once. A developing butterfly must specify, for example, the general shape of its wing, the precise venation pattern, the distribution of scale and/or bristle types on both sides and on all edges of the wing, the distribution of pigment(s), and finally whether scales are to be structurally colored and if so, what type of structure they are to have. Nijhout has presented a compact and authoritative review of pattern formation in butterfly wing systems; many other researchers are currently studying the molecular and genetic mechanisms underlying pigment formation and deployment. Common themes are emerging, but much still remains to be done, especially on the role of physical forces that almost certainly work along with the biochemical ones to bring forth the final form.

The formation of the microarchitecture underlying structural color systems is less well understood. Ghiradella in 1998 reviewed what was known about development of structural colors in scale systems, and Neville in 1993 presented a comprehensive review of the formation of helicoidal and other fibrous composite systems. However, despite their value as potential models for human research and development, particularly of optical systems, very little is known about these systems. There surely are lessons to be learned here. For example, Bragg lattices are of interest to engineers seeking more efficient transmission of information along optical fibers, and scale optics is becoming of interest to the photonics research community, which is seeking to develop structures and materials that can control light for purposes of communication, paints, surface coatings, electronic displays, etc. Again, the insect systems have a lot to teach us, especially since their structures are made at room temperature and without toxic solvents.

On the longer time scale, how did these systems evolve? In some examples there are grounds for speculation. As mentioned above, many pigments may have originally been metabolic by-products that, because of actual or potential

ability to absorb some wavelengths of light, were somehow co-opted for purposes of display. The helicoidal arrangement of chitin fibrils in cuticle is part of a larger structural adaptation of cuticle as a building material. As in all skeletons, fibril orientation in cuticle is tailored to local challenges. Helicoidal arrangements, with their multidirectional fibril orientation, are well equipped to to provide toughness and strength in the face of multidirectional stresses and are common in areas exposed to such stresses. Having evolved such a helicoidal arrangement to confer a particular type of strength, the animals needed only to make the pitches of the helices regular and to tune them to have fine iridescent reflectors at the same time.

The evolution of the thin-film, diffraction, and other systems is at present a very open question. They appear to be of great antiquity. Parker reported diffraction and antiglare structures in Burgess shale fossils and suggested that the emergence at the beginning of the Cambrian period of image-forming eyes (to quote Parker, "…the lights were effectively turned on…") may have produced extreme selection pressure for potential prey animals to develop rigid armor (with its inherent potential for forming structural colors) and at the same time a need for, and an opportunity to develop, camouflage, recognition patterns, and all the other common uses and expressions of biological color. While there is no question about the utility of these color mechanisms, it is still hard to imagine how so much can have been accomplished, even with millions of years of research and development in a competitive and presumably highly selective world.

To this point coloration has been considered in terms of passive and static displays on the surfaces of insects. But in living insects, the color-producing structures are situated on a moving body with moving appendages, and so the displays are modulated over time. The resulting signals are four dimensional, which adds to them a richness of information that we cannot begin to appreciate, especially because the true capabilities of the insect eye (which is much "faster" than that of the human) in its processing of either color or movement are not known.

The subject of insect mastery of light must also include bioluminescence. Lantern types and flash patterns come in a variety of forms; superficially, the mechanisms by which they are produced seem to differ radically from those already discussed. But here too, the insect displays mastery of architecture—in the design of the lantern cells themselves—and of chemistry to create light signals that can be controlled in space and time, but at those times of day when sunlight is not available to power the display. In doing so, the insects have truly made "the lights come on," replacing in their signaling the warmth of sunlight with their own cold light. As researchers continue to learn about these systems, they are exploring worlds within worlds of complexity and can only gain in appreciation of the enormous capabilities of biological systems in their communication with their environments … and with each other.

See Also the Following Articles

Aposematic Coloration • Bioluminescence • Cuticle • Eyes and Vision • Industrial Melanism • Integument

Further Reading

Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. (2001). "From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design." Blackwell Sci., Malden, MA.

Chapman, R. F. (1998). "The Insects: Structure and Function." Cambridge University Press, Cambridge, U.K.

Fox, D. L. (1976). "Animal Biochromes and Structural Colors." University of California Press, Berkeley.

Ghiradella, H. (1998). Hairs, bristles and scales. *In* "Insecta" (M. Locke, ed.), Vol. 11A of "Microscopic Anatomy of Invertebrates," (F. W. Harrison, ed.), pp. 257–287. Wiley, New York.

Kunzig, R. (2001). Trapping light. *Discover* **22,** 72–79.

Mason, C. W. (1926). Structural colors in insects. I. *J. Phys. Chem.* **30,** 383–395.

Mason, C. W. (1927). Structural colors in insects. II. *J. Phys. Chem.* **31,** 321–354.

Mason, C. W. (1927). Structural colors in insects. III. *J. Phys. Chem.* **31,** 1856–1872.

Neville, A. C. (1975). "Biology of the Arthropod Cuticle." Springer-Verlag, New York.

Neville, A. C. (1993). "Biology of Fibrous Composites." Cambridge University Press, Cambridge, U.K.

Nijhout, H. F. (1991). "The Development and Evolution of Butterfly Wing Patterns." Random House (Smithsonian Inst. Press), Washington, DC.

Parker, A. (1999). Light-reflection strategies. *Am. Sci.* **87,** 248–255.

Parker, A. (1999). The Cambrian light switch. *Biologist* **46,** 26–30.

Vukusic, P., Sambles, J. R., Lawrence, C. R., and Wootton, R. J. (1999). Quantified interference and diffraction in single *Morpho* butterfly scales. *Proc. R. Soc. London B* **266,** 1403–1411.

Vukusic, P., Sambles, J. R., and Lawrence, C. R. (2000). Colour mixing in wing scales of a butterfly. *Nature* **404,** 457.

Commercial Products from Insects

Michael E. Irwin and Gail E. Kampmeier

University of Illinois and Illinois Natural History Survey

When people contemplate how insects are marketed as consumer products, images of novelties, gimmick foods, cuddly toys, odd adornments, and cartoon images are invoked, calling forth a range of emotions from repugnance to warmth. But that is just the tip of the iceberg. Insects can be very big business. They and their products are sold for crop pollination, pharmaceuticals, health and agricultural protection, and human, pet, and livestock nutrition, as implements for conducting research, and for a host of other uses. This article focuses on commercialization of insects and their products.

FIGURE 1 Bees hired out for pollination of apple orchard. (Photograph by Eugene Killion.)

MARKETING LIVING INSECTS

Crop Pollination

Flowering plants are fertilized by several groups of insects. By far the most common pollinators are bees, and the honey bee, *Apis mellifera,* plays the dominant role in pollinating large tracts of agriculture. The domestication of the honey bee for pollinating crops had its beginnings at least 4000 years ago. Since that time, beekeeping has flourished and is now a thriving industry. In the United States alone, $15 billion worth of crops (fruits, vegetables, flowers) are pollinated by domesticated honey bees each year. Commercial apiaries lease their beehives to growers who need their crops pollinated. The keepers manage the hives, moving the bees from field to field to ensure crop pollination (Fig. 1). Although worker bees are not sold as such, their labor is. Moreover, the commercial interdependency of the honey bee industry is complex. Keepers buy high-quality queen bees from specialized suppliers, who, along with the keepers, purchase bee-tending equipment from other specialized suppliers, and the entire industry is dependent on information contained in specialized books, journals, and magazines.

A number of crops are more efficiently pollinated by bees of other kinds. Leaf-cutting bees, or mason bees, are a good example. These "solitary" bees, unlike honey bees, do not live in colonies. Solitary bees produce no honey or wax but are relatively docile and not likely to sting. One species of leaf-cutting bee, *Osmia cornifrons,* is widely used in Japan for apple pollination. It was imported to the eastern and midwestern United States for the same purpose. Another leaf-cutting bee, *O. lignaria,* a native to parts of the United States, is also widely used for orchard pollination. Pollinating a hectare of apples requires on average either 750 female hornfaced bees *(O. cornifrons),* 600 female blue orchard bees *(O. lignaria),* or 50,000 honey bee workers. Other mason bees, bred and sold to alfalfa growers in the western portions of the United States, ensure the production of high quality alfalfa seed. Not only are mason bees better pollinators of a number of crops, they are also immune to the devastating effects of tracheal and varroa mites, which can decimate honey bee colonies.

Another crop pollinator is the bumble bee, which is less affected by extreme weather than the honey bee and is better adapted to perform under confined greenhouse conditions. By vibrating as they extract nectar and pollen, bumble bees efficiently pollinate flowers and encourage high fruit set under greenhouse conditions. Bumble bees are bred, reared, and packaged for sale to growers for pollinating vegetable crops (particularly tomatoes) grown under greenhouse and plastic tunnel conditions. Entire industries are founded on the production and sale of bumble bees, especially in the Mediterranean region, from Spain to Israel.

Agricultural and Human Protection

One who has never witnessed the devastation of a crop by insect pests would be alarmed by the rapidity with which it can occur. One of the best ways to counter the buildup and devastation caused by insect pests is to unleash on them their own natural enemies. A vibrant industry is built on supplying the natural enemies or "beneficials" needed to manage pests and pest outbreaks, both for protecting agriculture and for preserving human health. These beneficials can take the form of insect pathogens, insects that prey on the pests (predators), insects that parasitize them (parasitoids), or insects that destroy weeds. This industry is increasingly in demand as growers, horticulturists, home gardeners, and vector control organizations alike turn from chemically oriented pest suppression measures to the principles of integrated pest management (IPM) and practices more attuned to organic farming. Many companies are in the business of rearing and supplying beneficial organisms, not just for agriculture and health in their broadest senses, but also for parklands, green corridors, and home gardens. This challenging industry must take into account knowledge of the systematics of the pests and the beneficial organisms that attack them, methods to efficiently and inexpensively mass-produce the desired beneficials, ways to maintain genetically viable and aggressive beneficial organisms, procedures to efficiently transport beneficials to targeted release sites, and knowledge to ensure that the habitats of the release sites are conducive to optimal utilization by the beneficials for controlling the pest. This last point is especially critical because the beneficials can simply move from the release site and take up residence elsewhere, providing a neighbor, instead of the grower who purchased them, with pest suppression.

Although the mass rearing and marketing of beneficial insects is an expanding business, the mass irradiation and release of sterile males is of a considerably larger scale. The Mediterranean fruit fly, New World screwworm, tsetse fly, and boll weevil have been successfully controlled through inundative releases of sterile males. The technology of irradiation is so notably complex, the scale of releases so

great, and the costs of mass irradiation so high that these services are almost always provided by a governmental agency. The International Atomic Energy Agency (IAEA), the Food and Agriculture Organization of the United Nations (FAO), and the U.S. Department of Agriculture (USDA) have been instrumental in pioneering sterile male irradiation and release. The equipment needed to mass-rear, irradiate, and release sterile males is costly. Industry, its major supplier, is exploiting the need for this technology by developing and marketing specialized products.

Live Insects and Human Therapy

The thought of using live insects to treat human ailments would make most pale, but the results can sometimes outperform drugs and surgery typical of more traditional Western medicines. Honey bees, fly maggots, ants, and *Plasmodium*-carrying mosquitoes have all been used in human therapy.

The venom of honey bees is used to ameliorate inflammatory and autoimmune conditions such as multiple sclerosis, arthritis, rheumatism, chronic pain, neurological diseases, asthma, and dermatological conditions. The venom can be administered by humans or injected via the sting of a bee. Venom therapy is widely used in China, Korea, Bulgaria, Romania, Russia, Brazil, and the United States. Much of the research with venom therapy in the United States focuses on treatment of multiple sclerosis and chronic pain. Of the more than 40 components identified in bee venom, 18 are considered to be active. One of these, melittin, is among the most powerful anti-inflammatory substances known. Although not a conventional form of treatment in the United States, anecdotal evidence of its efficacy is accumulating, and companies market bees and bee products for therapeutic purposes. This form of therapy should be undertaken only with qualified supervision and, since some people go into anaphylactic shock when stung by bees, should be administered with adequate precaution.

Maggot debridement therapy uses maggots of *Phaenicia sericata* to cleanse wounds of necrotic tissue without attacking healthy underlying tissues. Maggots have been used to treat abscesses, burns, cellulitis, gangrene, ulcers, osteomyelitis, and mastoiditis. Their use has lessened the need for amputations and has been especially useful where diabetes is a complicating factor. Therapy involving the cleansing effect of these maggots dates to the 16th century. Despite the pioneering work in the early part of the 20th century, the practice of debridement therapy fell to disuse with the advent of antibiotics and new surgical techniques in the mid-1940s. The increase in resistance to antibiotics in the late 1980s elicited a resurgence of interest in debridement therapy. The mechanisms underlying success of this treatment remain poorly understood to this day.

Live ants, particularly Amazonian army ants and carpenter ants of Africa, India, and the Mediterranean region, have been used to close wounds and surgical incisions. The sharp mandibles of the soldiers lock when their jaws are closed, irrevocably fastened in place even if the bodies are severed from their heads. Although these live suturing instruments are unlikely to grace surgical theaters in modern hospitals, they have long been used by native peoples.

Another use of live insects for therapy has fallen to disuse now that better alternatives are available. Before other treatments were available, it was known that the progress of syphilis could be halted when the body temperature was raised above 40°C. Because the effects of syphilis were so devastating, mosquitoes bearing a relatively mild strain of malaria were used to infect such patients. The *Plasmodium* pathogen caused high fevers that exterminated the syphilis pathogen. Although the patients were then infected with malaria, the cure was deemed worth the consequences.

Living Insects on Parade

There is little doubt that insects fascinate. Perhaps that is why they are so often featured in zoos and living museum displays, sold as pets, bred and released to celebrate special events, filmed and videotaped for movie and television productions, and ubiquitously adopted for live entertainment and education.

Why anyone would purchase live immature insects and rear them to adulthood may baffle some, but marketing immatures for that purpose is a thriving industry. Hobbyists in the United Kingdom and around the world order exotic butterfly chrysalids for the sheer joy of observing the spectacularly adorned adults emerge. Ant farms, butterfly houses with living chrysalids, a wide variety of butterfly immatures, exotic live tropical stick insects, praying mantid and cockroach oothecae, pea aphids, *Drosophila,* mealworms, and lady beetles are sold directly by suppliers or by auction. Rearing these insects is both fun and educational. For insect collectors, it offers a way to obtain exotic species.

Insect zoos, petting zoos, live museum displays, and insects in botanical gardens provide a sometimes exotic backdrop for educating the curious. Staged cricket and beetle fights are popular pastimes in Japan and elsewhere in Southeast Asia. Entomology departments at universities and museum and entomological societies in a number of countries have sponsored insect expositions (insect expos) that draw school groups and families from great distances to view these fascinating creatures firsthand. Cockroach races are regular features at insect expos. The human flea, *Pulex irritans,* was the center of attraction in American flea circuses, where their antics would attract Depression-era audiences to see a show at more than the cost of a double-feature movie.

Butterflies are a charismatic group of insects that are recognized and appreciated by almost everyone. The butterflies' spectacular, often iridescent beauty has caught the eye of naturalists and collectors alike. In Victorian times, Lord Rothschild employed more than 400 explorers to seek

out and collect butterflies for what became the largest personal butterfly collection in the world. Although rearing, buying, and trading butterflies has been a popular pastime in Europe since the days of Queen Victoria, one of the first butterfly houses was inaugurated only relatively recently, in 1977, to attract tourists to Britain's island of Guernsey, whose poor weather left little to recommend it. After Guernsey's commercial tomato industry failed and the plastic growing houses were abandoned, someone thought to plant tropical gardens in those plastic houses and populate them with exotic butterflies. The idea was a success and was copied elsewhere well into the 1980s. In 1977, exotic butterfly suppliers were unknown, but the industry has since spread to Thailand, Malaysia, the Philippines, El Salvador, Costa Rica, Taiwan, Kenya, Madagascar, and the United States.

Why not release butterflies at your wedding ceremony or next celebration? Environmentally correct, butterflies can be impressive when they take flight. Forget the rice, birdseed, and confetti. Painted lady *(Vanessa cardui)* and monarch *(Danaus plexippus)* butterflies are bred and sold for such releases. One particularly poignant use of butterfly releases took place at a fund-raising event in Costa Rica, where each member of the legislature released a butterfly and simultaneously called out the name of a street child to whom the release was dedicated, lifting the children's hopes and aspirations skyward.

Who can resist the calming effects of sounds emanating from nature? Many commercial stores play bird, frog, and other nature sounds as a means of enticing customers to come in and shop. Compact discs (CDs) are recorded and sold for commercial use and for households wishing to bring that calming quality into the home. Among the insect sounds recorded and marketed are those of cicadas, grasshoppers, tree crickets, mole crickets, ground crickets, and katydids singing; June beetles flying; honey bees, bumble bees, yellowjackets, and midges swarming; and medleys of insects communicating or otherwise sounding off in nature. These recordings are sometimes played at insect expos to help bring a sense of reality to those who come to imbibe the amazing presence of the insect world.

The entertainment industry takes advantage of fascinating, educational, scary, and exciting properties of insects by featuring them in movies and on television. Insects are topics of education and wonder on various television series. The cost of producing these documentaries, largely filmed in the wild, is covered by sponsorship. Like early films with insect subjects, children's movies rarely film living insects; instead they use graphical characterizations and cartoon images. Fictional films made for more mature audiences, however, usually present the frightening or horrifying aspects of insects, and for this purpose, people are hired as "insect wranglers" to supply and manage live insects on the set. Such management demands a basic understanding of insect behavior, including knowing how to influence the insects to "act" in the way desired by the film director. Discovering that dead insects were easier to manipulate than live ones, Wladislaw Starewicz wired dead specimens and manipulated them frame by frame to simulate desired actions in his early short, *The Fight of the Stag Beetles.* In the 1978 film, *The Swarm,* killer bee invasions were depicted by filming actual bees.

Waging War with Insects

Insect-borne diseases have taken the lives of countless soldiers throughout the ages. Millions have fallen to malaria, yellow fever, dengue, and a host of other diseases transmitted by mosquitoes. The purposeful waging of war with living insects dates to at least the 14th century when the Tartar army catapulted bodies of bubonic plague victims into Kaffa. Although knowledge that fleas spread this dread disease would not come until much later, the tactic nonetheless served its purpose.

Using insects to destroy agricultural crops seems to have emerged as a weapon of war only in modern times. Harlequin bugs, *Murgantia histrionica,* were introduced into the South, presumably in an effort to destroy the crops of the Confederacy during the American Civil War. Insects were used in both world wars as purposeful weapons. During World War II, the Japanese undertook the first large-scale use of insects as weapons of war by mass-producing an astonishing 500 million fleas bearing plague bacilli per year! In 1950, during the Cold War, the United States was accused of dropping Colorado potato beetles over East Germany. The Korean War brought to the Far East theater some 14 additional insects purportedly propagated in the United States as agricultural and medical warfare agents. The Vietnam War introduced additional entomological agents of war, especially as vectors of anticrop agents like plant viruses (e.g., beet curly top and Fiji disease), and fungi, (e.g., fire blight, cornwilt).

It was not until 1972 that insects were explicitly banned as weapons of mass destruction by the Biological Weapons Convention. Even though the mass production of these biological weapons was carried out exclusively by governmental agencies acting in secret, the trickle-down effects on local economies of producing entomological "weapons" must have been notable.

Entomological warfare does not stop with wars, where humans square off against each other. In 1990, another relatively large-scale war was waged, this time on the illicit drug trade. In fact, the U.S. government allocated $6.5 million to investigate, breed, and air-drop lepidopterous caterpillars to devour fields planted to coca in tropical Peru.

Insect Identification Services

There are so many insects in this world that it is difficult to identify them. Only by having authoritative determinations can many of the various insect-oriented industries succeed. Because of this demand, identification services have sprung

up around the world. Some are geared toward the identification of agriculturally or medically related insects, but many focus on identifying insects in the context of biodiversity, especially of benthic invertebrates.

FARMING INSECTS FOR THEIR PRODUCTS AND BY-PRODUCTS

Not only living insects are marketed. Dead insects and products derived from them can also be of high commercial value. In fact, insect products and by-products probably account for the lion's share of insect commercialization.

Implements for Research

Insects provide critical basic tools for studying a great many aspects of biology. Because *Drosophila melanogaster,* a common fruit fly, is small, has a short life cycle, and is inexpensive and easy to rear, it is an extremely valuable organism for biological research, particularly in the fields of genetics and developmental biology. *Drosophila* has been used extensively and intensively as a model organism for research for almost a century, primarily to uncover the relatedness of genes to proteins and to study and map the underlying mechanisms of genetic inheritance and gene expression. More recently, the field of developmental biology, especially embryology, has relied on *Drosophila* in explorations of how a complex organism arises from a relatively simple fertilized egg. The genome of *Drosophila*, recently sequenced, maps the gene structure of that seminal organismal model. Gene products such as *Drosophila* polypeptides and transcripts, and investigative tools such as the *Drosophila* Activity Monitor for circadian rhythm research, provide highly marketable products for the scientific supply industry. Moreover, specific, even mutant strains of *Drosophila* may be purchased, as well as supplies for rearing and maintaining cultures, and specialized equipment for conducting experiments.

Insect products are also marketed for other research functions. For instance, they are used for genetic and molecular markers. The enzyme luciferase, derived from fireflies, is an excellent marker for assaying gene expression. These markers are produced and sold commercially. Indeed, specialized equipment for detecting the expressed bioluminescence is also marketed. Cell lines derived from insects is another powerful research tool. For example, protein-based human and veterinary vaccines and therapeutic proteins are produced by using baculovirus expression vector systems in insect cell lines. Human and animal protein products derived from insect cell lines are marketed for a number of purposes, including drug screening and clinical trials.

Food Products

Insects are an extremely rich source of high-quality proteins, fats, essential vitamins, and minerals. It is therefore not surprising that dead insects and products derived from them are marketed for their nutritional value. These products can take the form of human food, pet food, and livestock feed.

HONEY One can hardly think of insects as a source of human food without envisioning honey, diligently produced by worker bees. Honey was used as a sweetener in ancient Egypt and continues to be popular today, both in cooking and for sweetening foods to be consumed immediately. Entire industries are built around honey bees both as crop pollinators and as master producers of honey. The latter industry ends with the sale of honey products on the supermarket shelf, but the intermediaries are varied and include, beyond those involved in rearing honey bees, equipment for extracting honey from combs, devices for straining and clarifying honey, and beekeeping books and magazines that keep the honey producer up to date on the latest developments in the industry.

HUMAN FOOD One often thinks of insects as human food in a novelty context, like being dared to eat fried mealworms, crickets, or chocolate-covered ants at the county fair. But insects have been a serious source of human nutrition for a very long time. This association substantially waned as urbanization and "westernization" spread, but in the less developed corners of the globe it continues unabated. Accordingly, about 500 species in some 260 genera and 70 families of insects are used for human food somewhere in the world, especially in central and southern Africa, Asia, Australia, and Latin America. Even in the West, insect foods need not be a novelty. Where they are consumed, insects provide 5.10% of the annual animal protein of indigenous peoples. Some Native American peoples consumed saturniid moth larvae as a main part of their diets. Currently, more than 100 species of insects are sold as human food at local markets in rural Mexico, where they constitute a regular part of the local diets. In Thailand, the specialized sex pheromone gland from giant water bugs provides a flavoring to shrimp paste. Thus, marketing insect-derived foodstuffs in selected regions of the globe contributes to local economies, but repugnancy of insect foods in western cultures continues to thwart economic opportunity for mass-producing and marketing these products in the West.

PET FOOD Birds, lizards, fish, caiman, crocodiles, turtles, and a host of other insectivorous pets survive and breed much better if supplied with protein and nutrients that are available from live or dead insects. Rearing and selling these insects to the public is a thriving business (Fig. 2) Madagascar hissing roaches are sold as reptile food, whereas crickets are marketed for consumption by a variety of pets.

LIVESTOCK FEED Beyond pet food, insects can provide a highly nutritious food source for domestic animals and livestock. Although low in such amino acids as methionine-cysteine, arginine, and tryptophan, when supplemented by

FIGURE 2 Superworms *(Zophobus morio)* are popular large mealworms originating from South America available in pet stores for reptile and amphibian food. (Photograph by Gail Kampmeier.)

these, insect protein forms an excellent feed. Under clinical trials, white rats (the universal experimental animal for testing new medical and pharmaceutical findings) fed Mormon cricket meal demonstrated the great potential of insects as a major source of protein for rats. China recognizes the potential nutritive value of insects as feed for fish, poultry, pig, and farm-grown mink. There, experiments have demonstrated that insect-derived diets are cost-effective alternatives to more conventional fish meal diets. House fly larvae and pupae, silkworm pupae, and mealworm larvae are the major source of these insect-based diets. Fly larvae fed on poultry manure have been experimentally incorporated back into poultry feed. When this system is in place, it will take the concept of recycling to a whole new level.

Secretions and Dyes

A number of insects have the ability to secrete substances such as waxes and resins through specialized glands. Dyes too can be extracted from insect tissues. Many of these products are of high commercial value.

SERICULTURE Among fine fabrics made of natural products such as wool, cotton, linen, and leather, silk is almost always the most highly prized. Silk cloth is woven from a secretion of the silkworm, *Bombyx mori.* In the Orient, sericulture, a 4700-year-old industry, has built up around this insect and its precious secretion. The silk is a continuous-filament fiber consisting of fibroin protein, secreted from two larval salivary glands in the insect's head, and a gum called sericin, which cements the two filaments together. Silkworm larvae secrete this substance to weave cocoons within which they pupate. To obtain the fibroin protein filaments, cocoons are softened in hot water to remove the sericin. Single filaments are drawn from cocoons in water bowls and combined to form yarn, which is drawn under tension and wound onto reels, dried, packed according to quality, and sold as raw silk. It was once believed that silklike synthetic fibers would replace silk, thus decimating the silk industry, but that has not occurred. In fact, world silk production nearly doubled over the last 30 years. Together, China and Japan manufacture more than half of the world production. Other countries, like Nepal, are intensifying their silk production. The sericulture industry is complex, and many suppliers commercially produce and sell products to culture silkworms, obtain the raw silk, refine the silk, weave it, produce clothing from it, and sell the products on the market. Wild sericulture also exists: that is, fibers from cocoons other than the silkworm are used, often by native peoples, in a similar manner. This industry is less relevant to the modern world of commerce, but it fuels local industry and provides clothing and other needs of native peoples, especially in India.

SPIDER SILK Spider silk, like that of silkworms, is composed of fibroin. However, unlike silkworms, which secrete silk from salivary glands in the head, spiders secrete silk from glands at the tip of the abdomen. Depending of the type of silk that is to be made, the spider mixes the fluid from up to six different glands and regulates the speed and volume of release. Spider silk is an extraordinarily strong and elastic material. On a weight basis, it is stronger than steel; a pencil-thick strand of silk is strong enough to snare a Boeing 747 airplane in midair. DuPont advertises that the company's researchers are studying biopolymer structures of the spiderwebs. They have used recombinant DNA technology to produce analogues of spider silk in yeast and bacteria and are planning to promote this synthesized material for all manner of construction purposes.

ROYAL JELLY Royal jelly, a substance secreted by the salivary glands of worker honey bees, stimulates the growth and development of queen honey bees. It is one of the most difficult of all foods to harvest, commanding astronomical prices because of its scarcity and high demand, fueled by belief in its healing properties. What royal jelly can do for humans is controversial, but it purportedly reinvigorates the body and extends the life span. Pantothenic acid, a major ingredient, is useful in treating some bone and joint disorders. Rheumatoid arthritis symptoms may subside with the injection of this acid. When pantothenic acid is combined with royal jelly, even better results are reported. This product is sold by many health food companies.

BEESWAX Glands on the underside of young worker honey bee abdomens secrete small wax platelets, which are masticated and molded inside the hive into a comb of hexagonal cells that are then filled with honey. Additional wax is used to cap the cells for honey storage. Of all the

primary products of the honey bee, wax has been, and remains, the most versatile and most widely used material. For centuries, beeswax has been regarded as the best material for making candles. An excellent wax for polishing woods and floors, it is also an ingredient in general-purpose varnishes. It has uses in packaging, processing, and preserving foods, and as a separation agent in the confectionery industry and in cigarette filters. Textiles and papers are waterproofed with products containing beeswax. Emulsions containing beeswax clean and soften leather goods.

Batik, an Asian method of coloring cloth, is based on the principle that wax (traditionally beeswax) protects areas that are not to be stained when the cloth is immersed in the dye solution. This protection feature is used for waterproofing and as an anticorrosion rust inhibitor to prevent dissolution of the metal in steel drums used to store and ship honey. Materials for embedding or electrically insulating circuits of high and ultrahigh frequency include beeswax. Beeswax is used as a binder when lubricant characteristics are desired or if mixtures are to be ingested. It is an ingredient in slow-release pellets of pyrethrum pesticides. Glass can be etched with hydrofluoric acid when areas that are not to be etched have been protected with beeswax. Various inks, pens, markers, and even carbon paper often contain small amounts of beeswax. Ancient jewelers and artisans formed delicate objects from wax and cast them later in precious metals. Colors of 2000-year-old wall paintings, as well as wrappings of Egyptian mummies, contain beeswax. Beeswax has long found use in medicines and body lotions. As a coating for pills, beeswax facilitates ingestion. Other products in which beeswax is a traditional ingredient are grafting wax, crayons, sealing wax, protective car polishes, and thread for sewing sails and shoes.

RESINS Shellac has been in use for 3200 years and is made from an insect native to India and Myanmar, the lac scale, *Laccifer lacca.* Lac females infest branches of fig trees and cover their bodies with a resinous secretion that hardens into a shield. Between 17,000 and 90,000 insects are needed to produce a pound of lac. The resins are ground to free the lac granules, which are then crushed and boiled in water. The floating lac is skimmed off, dried, and placed in burlap bags, which are stretched over a fire. As it is heated, the bags are twisted and the melted lac drips out. Before hardening, the lac is stretched like toffee. After hardening, the lac is broken into pieces and sold. Lac is the basic ingredient of a vast list of products besides shellac, including stiffening agents in the toes and soles of shoes and felt hats, shoe polishes, artificial fruits, lithographic ink, glazes in confections, phonographic records, playing card finishes, and hair dyes.

INKS Iron gall ink is arguably the most important ink in the history of the Western civilization. It is made of vitriol, gum, water, and, most notably, tannin extracted from Aleppo galls. Oaks produce Aleppo galls in response to a chemical substance secreted by larvae of the cynipid wasp, *Cynips gallae-tinctoriae.* The gall provides both food and protection for the larva. Tannin content of the gall is highest before the wasp exits. Iron gall ink is still sold and used for many purposes. Because iron gall ink is indelible, it was the ink of choice for documentation from the late Middle Ages to the middle of the twentieth century. It was very popular with artists as a drawing ink, used with quill, reed pen, or brush. It is now used by the U.S. Treasury in the ink for printing money. The range of objects that contain iron gall ink is enormous. It was used for most manuscripts, music scores, drawings, letters, maps, and official documents such as wills, bookkeeping records, logs, and real estate transactions.

DYES Historically, adult female Mediterranean scales (*Kermes iticies* and *K. vermilio*), Oriental lac insects *(Kerria lacca),* central European scales *(Porphyrophora polonica),* and New World cochineal scales *(Dactylopius coccus)* were used in the preparation of red dye by a number of indigenous populations. Today, cochineal dye is the most important. It is obtained from an extract of the bodies of scale females found feeding on a cactus native to Mexico and Central America. The insects' bodies contain the pigment called carminic acid, which is effective in repelling potential predators such as ants. This substance is obtained by subjecting a mass of the crushed insects to steam or dry heat. Because 70,000 scale bodies are needed to produce a pound of cochineal, the dye is extremely expensive. Once commonly used as a scarlet-red mordant dye for wool and as a food coloring, cochineal has been largely replaced by synthetic products. It continues to be used as a coloring agent in cosmetics and beverages. Furthermore, the art of cochineal dying is practiced by natives in southern Mexico. The cochineal scale is still widely cultivated as a source of commercial dye in the Canary Islands and in parts of Central and South America. It is sold and chiefly used now as a biological stain.

Pharmacology

Even 3600 years ago, insects, their parts, and toxins derived from them were used to alleviate a number of human ills. Some of the remedies were less than effective (e.g., notably hirsute flies and bees used to treat baldness). Other insect-derived remedies were more credible because they have at their core a chemical property that today confirms their efficacy. For example, the hemolymph of cicadas has a high sodium ion concentration and was recommended in preparations to treat bladder and kidney dysfunction. Hemolymph is known to possess antibacterial properties and has thus been recommended in prescriptions to treat bacterial infections and sepsis. Traditional Chinese medicine includes a wealth of insects and other arthropods in its pharmacopoeia. Dried cockroaches, blister beetles, maggots, silkworm larvae, cicada exuviae, cicada nymphs and adults, and recipes using mole crickets, mantid oothecae, and silkworm frass can be purchased at traditional Chinese drugstores.

Aside from bee venom therapy described earlier, products from honey bees have long been used to promote health and as a food source (Fig. 3). Honey, royal jelly, bee pollen, and propolis are all sold to treat a variety of ailments from anorexia to insomnia to cardiovascular diseases, and to promote wound healing. More information can be obtained from the American Apitherapy Society.

Blister beetles are the major source of cantharidin, the active ingredient of "Spanish fly." This chemical has been used to topically treat warts and can be ingested for its aphrodisiac properties. Acute renal failure and death can arise from overdosing on cantharidin. These findings have prompted the removal of cantharidin from use in the United States, but Chinese researchers have discovered that beetles (e.g., *Mylabris phalerata* and *M. cichorii*), long used in traditional medicines, contain antitumor properties. Researchers are attempting to balance the potential cancer-fighting properties with undesirable side effects by testing less toxic analogues of cantharidin.

Adornments and Displays

Certain insects lend themselves or their products to the making of spectacular jewelry. Beetles are probably the most notable because of their durable, often iridescent, hardened forewings, called elytra, and interesting body shapes. They can be made into brooches or encased in plastic for key chains and paperweights; many tropical species are reared specifically for this purpose. Beetle elytra have also been woven into textiles. Insect galls and morpho butterfly and dragonfly wings have been incorporated into jewelry designs. Caddisfly larvae glue together tiny stones, grains of sand, and bits of litter to form cases that protect them in their aquatic environment. Furnished specific materials such as gold nuggets, shells, or semiprecious stones, they will incorporate these materials into their protective cases, which can then be harvested and made into earrings, necklaces, tie tacks, and pins. Insects trapped in fossil amber also are sold for jewelry and displays. Although butterflies and beetles are commonly encountered in displays, a wide variety of insects are sold for those purposes, as well whether as decoration or for educational uses.

FIGURE 3 Honey, skin care products, and cough lozenges all make use of products from honey bees. (Photograph by Ann Coddington Rast.)

Party Favors and Pranks

For the prankster, live Madagascar hissing cockroaches are sold as party favors and "stocking stuffers" for the holidays. Honey bees embedded in plastic cubes shaped like ice can be purchased to be placed in a guest's drink. Mexican jumping beans, which are bean seeds containing larvae of a small moth *Carpocapsa saltitans,* have been popular as novelties for decades.

MARKETING INSECTS IN THE ENVIRONMENT

A more nebulous category of insect commercialization surrounds the marketing of insects in the wild. Bioprospecting, ecotourism, and conservation enhancement are modes through which insects are marketed in an environmental context. These modes frequently interact to serve the broader intent of environmental protection.

Biodiversity Prospecting

Biodiversity prospecting involves the exploration, extraction, and screening of commercially valuable genetic and biochemically active compounds of plants, arthropods, and microorganisms for pharmaceutical development and agricultural and industrial use. That some 200 pharmaceutical corporations and biotechnology companies are now stalking the wilds in search of biological riches is convincing evidence of the economic potential ascribed to bioprospecting. The vast array of insect compounds that are being discovered, reexamined, and put to new uses in disease treatments lags behind that of the botanicals currently being exploited. In combination with the tendency of many insects to sequester or change plant compounds they have ingested, there is an enormous untapped source of potential insect or insect derived compounds for medicine in the biodiversity of this planet.

With advances in molecular biology and the availability of more sophisticated diagnostic screening tools, it is increasingly cost-effective for commercial organizations in search of new pharmaceuticals to seek out natural products. This trend has resulted in a soaring market for candidate biological specimens, a market that currently tops $40 million per year for the pharmaceutical industry alone. Because of the difficulty many governments have encountered in maintaining sovereignty and control over their resources, there has been a surge of interest in legislation governing access to resources and in ensuring that host countries benefit from the commercial products fashioned from their native species.

Ecotourism

Tourism is the leading economic sector in several tropical countries. It is dependent on the lure of a warm climate,

relatively low prices, and perceptions of relaxation, excitement, and even educational appeal. Ecotourism takes advantage of the attractiveness of adventure by offering the enticement and wonder of nature in an exotic setting. Insects, too, provide tourist attractions, and perhaps the best example involves the monarch butterfly, a popular insect in North America. A tropical species, it extends its range northward well into Canada during the growing season but cannot overwinter there. Individuals retreat southward for thousands of kilometers each autumn to take up residence in climes more amenable to their survival. These butterflies are attractive to ecotourism enterprises precisely because of this pattern of movement accompanying the remarkable biology of the insects. Almost anyone can view these beautiful butterflies flitting around meadows and parklands during the summer months. But as autumn approaches, they begin remarkable journeys southward and westward towards one of two destinations, depending on where they grew up. Those east of the Rocky Mountains migrate to the Monarch Butterfly Biosphere Reserve at the high-altitude oyamel fir forests of Michoacán, in central Mexico, where they overwinter in extraordinary aggregations of millions of individuals. Those born west of the Continental Divide migrate southwestward and take up residency in the monterey pines, cypress, and introduced eucalyptus trees of Natural Bridges State Park and Monarch Grove Sanctuary in Pacific Grove, on the Monterey Peninsula of California, where they too overwinter in large aggregations. These two localities are ecotourist destinations. Entire tourist industries surrounding each locality are based on this amazing insect and its habitat. Accommodation, guided tours, and, in the case of Pacific Grove, considerable emphasis on fine dining, are featured. Organizations like Friends of the Monarch in Pacific Grove promote this ecotourism.

Conservation Pursuits

Conservation efforts fold together the concepts of ecotourism and bioprospecting in an effort to protect the landscape and the biota it contains. One intent of ecotourism is to sustain the environments that attract the tourists, permitting the business to remain viable. The indigenous Ejido community of central Mexico, for example, depended on income from logging in the buffer zone of the Sierra Chincua sanctuary, the largest and most pristine monarch butterfly overwintering area in the world. Through a leasing contract, the community agreed to cease logging sanctuary forests in exchange for compensation of lost income from ecotourism profits. When agreements are made with the care of the earth as a goal, bioprospecting can also be an instrument for conservation.

Although not big business, conservation efforts can involve the production and sale of insects. Indigenous populations that use natural areas will maintain them if profitable industries, based on gathering and selling renewable resources of the system, can be developed. Jewelry made from beetle elytra and sold at local tourist markets is an example. Insects are sometimes bred and released into the wild to enhance the preservation of the species. A butterfly breeding industry has emerged in many corners of the world where pupae are sold to collectors and accumulated for release into habitats where the species is, for one reason or another, becoming rare. In Papua New Guinea, participants in a butterfly farming project sell live and preserved butterflies to collectors around the world. They earn between $2500 and $5000 per year, 50 to 100 times the average per-capita income of $50. Residents who gain from this industry have a stake in protecting the local environment where wild butterfly stocks originate. Conservation groups encourage the sale of reared butterflies because that reduces the pressure on threatened and endangered species in the wild. Furthermore, by releasing a portion of the reared specimens back into the wild, the industry encourages ecotourism, which, in turn, brings added wealth to the community. A butterfly ranching project in Barra del Colorado in northeastern Costa Rica, is an example. It provides sustainable income for its participants and assigns a portion of the stock bred from wild and captive butterflies for release back into the wild.

See Also the Following Articles

Bee Products • Food, Insects as • Honey • Medicine, Insects in • Silk Production

Further Reading

Akre, R. D., Hansen, L. D., and Zack, R. S. (1991). Insect jewelry. *Am. Entomol.* **37(2),** 91–95.

Beekeeping/Apiculture/Imkerei/Apicultura. (2002). http://www.beekeeping.org/ Last updated January 10, 2002. Accessed March 12, 2002.

Cherry, R. H. (1987). History of sericulture. *Bull. Entomol. Soc. Am.* **33(2),** 83–84.

Crane, E. (1983). "The Archaeology of Beekeeping." Cornell University Press, Ithaca, NY.

DeFoliart, G. R. (1992). Insects—An overlooked food resource. (J. Adams, ed.), *In* "Insect Potpourri: Adventures in Entomology." pp. 44–48. Sandhill Crane Press, Gainesville, FL.

Family of Nature Websites. (2000). The Butterfly Website: Public butterfly gardens and zoos. http://butterflywebsite.com/gardens/index.cfm/ Last updated June 14, 2000. Accessed March 12, 2002.

Genetics Society. (1997). FlyBase: A database of the *Drosophila* genome. http://flybase.bio.indiana.edu/ Accessed March 12, 2002.

Krell, R. (1996). Value-added products from beekeeping. FAO Agricultural Service. Bulltin 124. http://www.fao.org/docrep/w0076e/w0076e00.htm/ Accessed March 12, 2002.

Lockwood, J. A. (1987). Entomological warfare: History of the use of insects as weapons of war. *Bull. Entomol. Soc. Am.* **33,** 76–82.

Manning, G. (2000). A quick and simple introduction to *Drosophila melanogaster.* http://ceolas.org/fly/intro.html/ Last updated Oct. 23, 2000. Accessed March 12, 2002.

National Honey Board. (2001). http://www.nhb.org/ Accessed March 12, 2002.

Reid, W. V., Laird, S. A., Meyer, C. A., Gámez, R., Sittenfeld, A., Janzen, D. H., Gollin, M. A., and Juma, C. (1993). Biodiversity prospecting: Using genetic resources for sustainable development. World Resources Institute. Washington, DC.

Strickler, K. (2000). Solitary bees: An addition to honeybees. http://www.pollinatorparadise.com/Solitary_Bees/SOLITARY.HTM/ Last updated 21 Feb. 2001. Accessed 12 March 2002.

Tauber, M. J., Tauber, C. A., Daane, K. M., and Hagen, K. S. (2000). Commercialization of predators: Recent lessons from green lacewings (Neuroptera: Chrysopidae: *Chrysoperla*). *Am. Entomol.* **46(1),** 26–38.
Thompson, J. C. (1996). "Manuscript Inks." Caber Press, Portland, OR.

Conservation

Tim R. New
La Trobe University, Melbourne, Australia

Insect conservation includes two main contexts. Insects may be conservation "targets," whereby particular species become the focus of concern because of their perceived decline in abundance or distribution, or insects may be conservation "tools," in which they are incorporated into broader aspects of conservation concern through their sensitivity to environmental changes and used as "signals" to monitor or herald changes to natural environments. This role is facilitated by their high richness and diversity in most terrestrial and freshwater environments. Both contexts reflect concern over human intervention with the natural world and the desire to sustain both components (i.e., species and equivalent entities) and processes in natural ecosystems. The major roles of insects in sustaining ecosystem services and processes acknowledge their immense richness and biomass and are reflected in E. O. Wilson's famous characterization of invertebrates as "the little things that run the world."

Nevertheless, with few exceptions, ideas of conserving insects are difficult for many people to accept. In contrast to higher vertebrates and many vascular plants, which people accept readily as objects worthy of conservation, insects have a poor image and are more commonly viewed as objects for suppression or elimination. They are regarded broadly as pests or nuisances or by some disparaging epithet such as "bird food" (however important that categorization may be in sustaining community integrity).

Insect conservation has a long history, mainly through focus on the more popular groups, such as predominantly butterflies, dragonflies, and some showy beetles. These insects are accepted widely as "worthy," simply because people like them and regard them as harmless. It is also revealing to see the commonly polarized perceptions of "a butterfly" and "a moth" despite these being artificial segregates of the same insect order. Concerns arose over decline of particular species from the mid-19th century onward. Initial concerns, and the foundations of modern insect conservation practice, were in western Europe and North America but have expanded to encompass many parts of the world. Conservation in practice includes application of biological knowledge to manage or sustain species and other higher ecological levels, which reflects the total biodiversity and linkages that occur within the complex, imposed framework of regulation and socioeconomic needs that provides for ever-increasing human populations. "Biodiversity" encompasses both taxonomic and genetic diversity, with conservation aiming, broadly, to prevent its loss—either by the extinction of threatened or rare entities or by preventing other entities from decline to that state. As major components of biodiversity, in terms of species richness, ubiquity, and ecological variety, insects are an important and increasingly appreciated component of global conservation need.

PROBLEMS WITH INSECT CONSERVATION

Traditionally, most insects have been largely disregarded in conservation, on the premise that they may be secure under measures taken to conserve more charismatic taxa such as warm-blooded vertebrates. The latter are supposed widely to act as "umbrellas" for most or all coexisting species, but this idea is now recognized as oversimplistic, because many invertebrates are ecologically specialized and need detailed management to sustain them in the face of environmental change. However, without past emphasis on vertebrates, many habitats and sites recognized as of considerable importance for insects would surely have been lost. One attraction of basing conservation on groups such as birds or mammals is simply that they are relatively well known: their diversity is limited and tangible, most of the species are named, and many are recognizable without having to capture and kill the animals for detailed examination; their biology and habitat needs are reasonably well understood, and their distributional ranges and patterns defined; even the numbers and population sizes of many species can be evaluated reliably. Parallels with large showy butterflies and dragonflies have led to these being referred to as "birdwatchers' bugs," but they contrast dramatically with most other insect groups. Uncertainties over levels of species richness, that most species are still unnamed or even uncollected, and fragmentary or nonexistent ecological and distributional knowledge provide severe impediments to defining the patterns of diversity and distributions that may constitute the template for conservation evaluation. For many insect habitats in most parts of the world, we have little idea of insect species richness and identity. Many insect species are known solely from long-dead museum specimens and may never be seen alive. Of the world's 12 "megadiverse countries" (collectively estimated to harbor more than 70% of earth's animal and higher plant species), only for Australia can reasonably informed approximations of the extent, distribution, and ecological features of the insect fauna be deduced sufficiently to make conservation recommendations above the universal need to safeguard natural habitats. For the far more species-rich tropical countries, the paucity of resident entomologists and differing priorities render such data very approximate and their accumulation a low priority. Costa Rica, recently subject to an internationally sponsored biodiversity inventory through its national Biodiversity Institute, is an important exception. It is salutary to reflect

that a decline in individual insect species in well-known (predominantly temperate-region) faunas can arouse substantial conservation interest and action, whereas tropical habitats supporting far more insect species than the total fauna of any European country disappear rapidly.

There is little reason to doubt that numerous insect species have become extinct as a direct result of human activities during the past few decades, although most have not been documented, and that the process continues. Insects are a major component of what has sometimes been referred to as "the sixth great extinction," considered likely to result in the loss of a substantial proportion of the world's species within a few decades. Efforts in insect conservation are an important avenue to increasing the understanding of human impact on natural ecosystems and of the subtle steps needed to safeguard them in the face of accelerating losses. However, the complexity of the issues involved demands a clear perspective and allocation of priorities, so that limited funding and expertise can be deployed for the greatest benefit.

Developing such perspective has involved: (a) increasing fundamental documentation of patterns of insect species richness on a variety of geographical scales, perhaps streamlining the process by concentrating on selected focal taxa because of the immense difficulty of enumerating all insect groups; (b) selecting the most deserving taxa for conservation targets, based on urgency of need to prevent extinction; (c) defining and alleviating threats to taxa and to their host environments; (d) public and administrative education to communicate the importance of insects in the natural world, and hence the need for their conservation; and (e) evaluating the contributions of insects in broader conservation activities. These parameters recognize that, despite ethical problems with any such selection, the diversity of insect species is such as to necessitate some form of "triage" in selecting the most deserving species for management and recovery action. One consequence has been a tendency to increase the scale of conservation concern; whereas single species are the most popular conservation targets, because they are defined tangible entities to which people can relate easily, their value as "flagships" or "umbrellas" for their habitats and other community members is of increasing importance in seeking wider benefits. Most fundamentally, support will never be sufficient to treat all deserving insect species individually as conservation targets needing expensive long-term recovery actions, and so any constructive shortcuts must be explored.

RANKING TAXA FOR CONSERVATION PRIORITY

The World Conservation Union (IUCN) has initiated schemes whereby species can be signaled as of conservation concern through being included on a global *Red List of Threatened Animals* and progressively allocated to a category of threat severity based on quantitative estimates of risk of extinction. The year 2000 *Red List* includes 747 insect species, including representatives of 15 orders, but is dominated by Lepidoptera (284), Hymenoptera (152), and Odonata (137 species); daunting though this number may seem, it is no more than the detected tip of the iceberg of needy insect taxa. Many of the species included have not been evaluated critically in relation to their close relatives, for example; some are listed as the result of the zeal of individual nominators; and many insect groups have no such champions to promote their welfare.

A number of regional red data books dealing with insects have established more local priorities, as have a greater number of Action Plans and similar documents arising from country- or state-based conservation legislations. In common with other taxa, the "listing" of an insect on a schedule of protected taxa often confers legal obligation to define and pursue the necessary conservation measures needed to ensure its well-being. It is important to recognize that simply being "rare" does not necessarily indicate conservation need. Many insects are known from single localities or otherwise very small areas. "Rarity" has connotations of one or more of small numbers, limited distribution, and ecological specialization, with the rarest species being ecological specialists occupying very small areas and occurring in very low numbers. However, rarity can be a stable condition. Conservation concern arises more properly from threats caused by human intervention increasing the level of rarity, such that a risk of extinction is imposed on a formerly stable balance or a trajectory of decline is accelerated. Conditions of rarity may predispose the species to stochastic effects and increase its vulnerability if the external threat spectrum increases.

Unlike many conservation assessments for vertebrates, quantitative population data on insects are rarely available, and even large numerical fluctuations between successive generations may be entirely normal. Detection of numerical decline is thereby difficult, and the quantitative thresholds for allocating a species to the IUCN categories of "critically endangered," "endangered," and "vulnerable" generally cannot be met. For most insects for which *any* biological information is available, which is a small minority, even the basic pattern of population structure is generally unclear. Many butterfly species previously assumed to have closed populations, for example, are now known to manifest a metapopulation structure, wherein discrete demographic units (nominally subpopulations) occur disjunctly in patches of habitat across a wider area, and the whole population is maintained through rolling series of extirpations and recolonizations of the suitable habitat patches in the wider environment. Thus, even loss of whole apparent populations may be entirely normal, and the practical conservation dilemma is to distinguish these from declines and loss caused by imposition of external threats.

THREATENING PROCESSES

The following are the major threats cited in decline and loss of insect species and assemblages.

Habitat Loss and Change

Many insects depend on very intricate and specialized ecological conditions for their survival, so that critical habitat and resource parameters can be very subtle. Whereas destruction of a forest, for example, is an obvious form of habitat loss, relatively small changes in vegetation composition or microclimate may lead to decline of ecologically specialized insects. Many lycaenid butterflies, for example, depend on a tripartite association whereby their caterpillars have obligate mutualistic relationships with particular species of ants, as well as specific larval food plants, so that both of these are critical resources, in addition to the need for nectar sources for the adult butterflies. On a broader scale, many insects are limited to or associated with particular vegetation types, so that any process that diminishes forests, grasslands, heathlands, alpine meadows, mangroves, and many other habitats may harm them. The area needed for many insect populations to be self-sustaining is not large—colonies of many butterfly species can thrive on areas of less than a hectare—so that the widespread pattern of habitat fragmentation through agricultural and urban conversion so damaging to many other taxa may not necessarily be harmful for insects. But, by the same token, even limited habitat destruction or change might exterminate the entire population or species.

Despite wide supposition to the contrary, many insects do not disperse readily or far. Some butterflies are reluctant to traverse even narrow bands of open ground between sheltered or shaded habitats, so even apparently unobtrusive habitat fragmentation (such as by construction of an access road) may have severe demographic and genetic consequences through promoting isolation.

Habitat loss is the paramount threat concern in insect conservation and is potentially universal. Many insect conservation programs stress the need for habitat security and management as the most important single conservation measure. The latter aspect is critical; simply that an insect is represented in a high-quality reserve such as a National Park does not in itself guarantee its well-being, because conditions may continue to change through succession or management for other priorities. For example, in Britain, several butterflies declined following changed grassland management involving removal of grazing by domestic animals or rabbits. This led to the decline of attendant ant species because of changes in ground microclimates from denser overlying vegetation; particular grazing regimes are an integral part of habitat management for such taxa. Simply "locking up" a habitat in a reserve may be a vital first step in ensuring security, but is not an end point in conservation practice.

Invasive Species

Replacement of native flora by exotic plants has characterized much human endeavor. In Australia, native grasslands in the south east are regarded as among the country's most endangered ecosystems, having been reduced to around 1% of their original extent. There is still "plenty of grassland," but most of it is composed of exotic grass species introduced to improve pasture quality for domestic stock. Many insects (including wingless morabine grasshoppers and some Lepidoptera) that depend on native grass species are now of considerable conservation concern, as representing putative remnant populations confined to small patches of their original much wider range. Introduced plants, be they agricultural or forestry crops, weeds or ornamentals, provide opportunity for exotic herbivores to establish and thrive—often as insect pests demanding control in order to protect commercial interests. Classical biological control of introduced pest weeds and arthropods has led to numerous introductions of insect consumers, be they herbivores, predators, or parasitoids. The practice has aroused concern among insect conservationists, because of the propensity of some such taxa to invade natural environments and attack native species, rather than being restricted to the (predominantly) agroecosystem environments where their impacts are needed. Protocols for screening for safety of biological control agents continue to improve, but some recent pest management practices need careful appraisal. Neoclassical biological control (whereby exotic natural enemies are introduced to combat native pest species) is a highly controversial practice, for example, as witness the recent debate over the possible side effects of exotic wasp parasitoids against innocuous native grasshoppers coexisting with the few destructive rangeland species in North America. In such situations, lack of host specificity is a prerequisite, as the agent is to attack "new species"; the agents are thereby seen as predisposed to become invasive and attack a wider host spectrum. For classical biological control, much concern has arisen from isolated island environments such as Hawaii, where there is strong suggestion that extinctions of sensitive native insects have resulted from nonspecific agents invading natural environments. A tachinid fly, *Compsilura concinnata,* introduced to combat gypsy moth (an introduced major forest pest in North America), is known to attack a wide variety of native Lepidoptera, with recent concerns for its effects on some giant silkmoths (Saturniidae).

Invasive social Hymenoptera, particularly ants such as the Argentine ant, *Linepithima humile,* and bigheaded ant, *Pheidole megacephala,* and vespoid wasps (such as *Vespula* in New Zealand), are known to outcompete native species and to disrupt the structure of natural communities in many parts of the world. As with other invasive taxa, many of these insects are extremely difficult to eradicate once they become established, and continuing spread is a major conservation concern.

Exploitation

Exploitation of insects as a threat revolves largely around issues of "overcollecting," a highly controversial and emotive theme in insect conservation. Collector demands for rare butterflies and beetles, in particular, have led to the listing of a number of species on schedules of the Convention on

International Trade in Endangered Species (CITES) as either prohibited in trade (a few species such as Queen Alexandra's birdwing butterfly, *Ornithoptera alexandrae*) or for which numbers in trade must be monitored. Much protective legislation for insects prohibits or restricts take of specimens. Individual rare insects can command sums of many thousands of dollars on the black market or more openly in dealers' catalogs. Trade in insects is the predominant aspect of exploitation and has three main components (as nominated by Collins and Morris for swallowtail butterflies): the low-value high-volume trade (mainly in common species, for the souvenir trade and general collector supply), the high-value low-volume trade (of very rare species for collectors), and the live trade (mainly of long-lived showy species for display in butterfly houses). The second of these is the major conservation concern, with potential for illicit measures to circumvent protective measures for exceedingly scarce taxa and which has led to development of butterfly farming (or butterfly ranching) activities to help satisfy demand for high-quality reared specimens. This approach was pioneered in Papua New Guinea and has major conservation benefits in helping to reduce human pressures on primary forest habitats. Rearing butterflies for sale through a centralized government-supported agency has provided income sufficient to curtail needs for continued agricultural development in places and has allowed people to recognize forests as resources on which their sustainable incomes depend.

However, and despite widespread assertion to the contrary, there is very little evidence that overcollecting is a common threat to insect species or populations. It is almost always subsidiary to changes to habitats. Very small, isolated populations of highly desirable species may indeed be "tipped over the brink" by imposition of any additional pressures and mortality, but measures to prohibit take must be seen as a responsible action. They should be balanced against the possible loss of information to be gained from hobbyists, who have contributed most of the information available on collectable insects such as butterflies.

Pesticides and Pollution

Pesticides are a special category of environmental pollutants, in that they are chemicals designed specifically to kill insects, rather than simply the by-product of industrial and other manufacturing processes. Pesticides can have nontarget effects, with the practice of greatest conservation concern being aerial spraying of insects in noncrop environments, either accidentally or purposely. Wiest's sphinx moth, *Euproserpinus weisti,* was almost exterminated when its last known site in Colorado was sprayed with malathion, for example, and aerial use of fenitrothion against massing plague locusts in Australia before they reach cropping areas remains controversial and a stimulus to develop alternative management strategies.

Other forms of chemical contamination, of both land and freshwater environments, have been documented as harmful to insect assemblages, and the more widespread acid rains in the northern hemisphere have undoubtedly threatened insects in forest environments. Many local pollutant effects could be cited as occasional threats. Concerns have been expressed widely over effects of insecticides on dispersive pollinators such as honey bees that can forage 5 km or more from their hives.

INSECTS AS CONSERVATION TARGETS

The most familiar forms of insect conservation are associated with campaigns to conserve individual species, often "crisis-management exercises" stimulated by perception of decline or of current or impending threats from anthropogenic changes to the environment. Many are initially local exercises, but the species may also be of national or (as regional endemics) global concern. Detection and quantification of conservation need are followed, ideally, by well-designed and effectively coordinated management or recovery plans, with adequate monitoring to determine their effectiveness. Sound biological knowledge of the species underlies any such program, and in many programs an initial research phase to elucidate key ecological features must precede optimal management. However, for this to occur, the species may need interim regulatory protection, such as a moratorium on future despoiling of its habitat.

With few exceptions, decline of insects has been difficult to quantify, because of lack of knowledge of population dynamics and absence of historical data on numbers and distribution. Many declines have been inferred from loss of habitat, on the basis of persistence in small remnant habitat patches and presumed losses elsewhere. For some well-known faunas, particularly for butterflies in western Europe, disappearances have been documented more effectively through a century or more of collector intelligence and accumulation of labeled specimens and information. The most complete example is for the British butterflies, a fauna of fewer than 60 resident species for which data are sufficient to plot reliable high-resolution maps of species incidence and change over much of the past century. Such "atlases" have been produced on a 10 by 10-km^2 scale and have progressively spawned similar compilations for other insect groups and countries, together with numerous recording schemes to define current situations. In North America, the annual Fourth of July Butterfly Count developed through the Xerces Society is an important initiative helping to define the template for insect conservation needs. However, for most of the tropics, which is the most species-rich part of the world for insects, such schemes are impracticable because of the lack of sufficiently informed resident entomologists/hobbyists and the complexity of the faunas, as well as vastly different local priorities.

Most concern for species, then, arises from perceived or anticipated declines to taxa considered rare or threatened in some way. For practical conservation the need is to define the severity of the threat(s), integrating this with knowledge of the species' ecology, and to determine and pursue the measures

needed to alleviate the threat and either (a) prevent decline and/or loss or (b) enable recovery from decline to occur. Development of an Action Plan or Recovery Plan is often accompanied by listing the species on some form of legal or regulatory protection schedule, a step that commonly accords formal obligation to investigate and pursue conservation needs and confers priority for limited support over nonlisted species. However, listing has sometimes been seen as an end in itself, being considered as practical conservation rather than as a facilitating mechanism for practical conservation measures. Many legislations provide for eventual delisting of species as secure, and this can come about in two ways:

1. The more intensive examination of the species that results from listing it reveals that it is more secure than supposed previously, so that continued categorization as of conservation concern is not warranted.
2. Recovery (or broader management) measures are successful and render the species, and its habitat, secure.

Either outcome is positive, but the process emphasizes the need for periodic review of all listed species to determine changes in security and the effectiveness of conservation measures. The practical steps needed, as for a variety of other taxa, are varied but may include increasing habitat security, intensifying site management (for example, by enriching it with food plants and eliminating competitive weeds), increasing insect numbers and distribution through *ex situ* measures such as captive breeding and release, and translocations to sites within the historical range from strong donor colonies. Any such program should be monitored fully and coordinated and managed effectively by a species recovery team whose membership includes informed entomologists. Because of the novelty of insect species management in this way, many such teams still tend to focus on expertise derived from vertebrates rather than reflect invertebrate expertise strongly.

In addition to species focus, assemblages or communities of insects are sometimes adopted as conservation targets, leading to a larger scale of consideration.

INSECTS AS CONSERVATION TOOLS

The high numbers of insect species and higher taxonomic and ecological categories (guilds) comprise a significant proportion of easily sampled biodiversity, with many easily categorized forms, in terrestrial and freshwater environments. They have attracted considerable attention in attempts to document communities and to measure the impacts of changes, whereby aspects of diversity, species composition, and ecological integrity can be evaluated by using insects as highly informative surrogates or indicators in various ways. A burgeoning literature on these topics reflects movements to conserve entities above single species and emphasizes the growing awareness of the key roles of insects in ecosystems and as "signals" of environmental health. Not all insects are amenable to use in this collective way; they are simply too poorly known. The desirable features of insect groups used as indicators include their high diversity and abundance, being widespread within the target ecosystems; being taxonomically tractable and recognizable (not always to species, and genera or families can be used instead of species in some groups, but it is highly desirable that adequate handbooks and identification keys suitable for use by nonspecialists are available); being easily sampled quantitatively or semiquantitatively by simple methods; showing demonstrable changes in response to particular sets of disturbances or otherwise being ecologically responsive; and being sufficiently understood biologically that normal fluctuations in abundance, incidence, and distribution are not confounded with disturbance effects. The best documented insect groups have naturally attracted most attention, and it is sometimes difficult to distinguish the really useful insect groups from those that simply have strong advocates but less proven worth. One constructive approach is to focus on several different taxonomic groups simultaneously and so to incorporate additional ecological breadth.

Another useful approach has been to determine the incidence of "functional groups" among ecologically diverse taxa (such as ants) in which different genera, tribes, or subfamilies coexist but have different trophic habits; respond to different physical, vegetational, or climatic cues; and interact in various ways. Local faunas can thus be characterized in functional terms and changes in balance of the different guilds used to evaluate environmental changes in, often, subtle ways. In freshwater environments, groups such as chironomid flies are diverse and have likewise been used to signal wider effects of pollution or temperature changes. In such contexts, insect indicators are the equivalent of the "miner's canary," with the strong implication that their responses may be sufficiently subtle to indicate environmental changes before the effects are reflected in other changes in biota. On a broader scale, the dependence of some insects on particular microclimates may provide a basis to monitor effects of longer term climate change. Thus, in Britain many insects are on the northernmost fringe of a broader European distribution and are confined to south-facing slopes with high insolation; with change in climate, their distribution may well also change.

Insects also have values as flagship or umbrella taxa, much as with some vertebrates in the past. Selected popular insects can capture public sympathy and are of vital importance in spreading advocacy for insect conservation and the broader values of invertebrates. These need not be indicators, but species adopted as local or broader emblems for conservation have been one of the main imperatives in development of insect conservation through bodies such as the Xerces Society (United States) and the former Joint Committee for Conservation of British Insects (United Kingdom, most recently known as Invertebrate Link). Many leading entomological societies now have sections for members

interested in conservation and conservation committees to help serve these wider interests. Even more broadly, characterization of community condition in terms of "representativeness" or "typicalness" or the principle of selecting nature reserves on their values as centers of diversity, evolution, or endemism can all benefit from incorporating insects in such evaluations rather than relying on low-diversity, sometimes atypical, vertebrate assessments alone.

See Also the Following Articles

Biodiversity • Biogeographical Patterns • Greenhouse Gases, Global Warming, and Insects • Endangered Insects • Insect Zoos

Further Reading

Collins, N. M., and Thomas, J. A. (eds.) (1991). "Conservation of Insects and Their Habitats." Academic Press, London.

Collins, N. M., and Morris, M. G. (1985). "Threatened Swallowtail Butterflies of the World." IUCN, Gland/Cambridge.

Gaston, K. J., New, T. R., and Samways, M. J. (eds.) (1994). "Perspectives on Insect Conservation." Intercept, Andover, MA.

Journal of Insect Conservation (various dates). Kluwer, Dordrecht.

New, T. R. (1995). "Introduction to Invertebrate Conservation Biology." Oxford University Press, Oxford.

New, T. R. (1997). "Butterfly Conservation," 2nd ed. Oxford University Press, Melbourne.

New, T. R. (1998). "Invertebrate Surveys for Conservation." Oxford University Press, Oxford.

New, T. R., Pyle, R. M., Thomas, J. A., Thomas, C. D., and Hammond, P. C. (1995). Butterfly conservation management. *Annu. Rev. Entomol.* **40,** 57–83.

Pyle, R., Bentzien, M., and Opler, P. (1981). Insect conservation. *Annu. Rev. Entomol.* **26,** 233–258.

Samways, M. J. (1994). "Insect Conservation Biology." Chapman & Hall, London.

Wells, S. M., Pyle, R. M., and Collins, N. M. (1983). "The IUCN Invertebrate Red Data Book." IUCN, Gland/Cambridge.

Crickets

Richard D. Alexander
University of Michigan

Daniel Otte
Philadelphia Academy of Natural Sciences

Crickets are insects in the order Orthoptera that comprise the ensiferan family Gryllidae. Some authors regard them as the superfamily Grylloidae with four families: Myrmecophilidae, Gryllotalpidae, Mogoplistidae, and Gryllidae. The group dates from the Triassic Period and today includes 3726 known living species and 43 extinct ones, 22 extant subfamilies and 7 extinct ones, 528 extant genera and 27 extinct ones. Most extant subfamilies are distributed worldwide.

DIAGNOSTIC FEATURES

The Orthoptera also include katydids, long-horned and meadow grasshoppers, short-horned grasshoppers and locusts, pigmy locusts, and wetas. Orthoptera are related to stick insects (order Phasmatodea), cockroaches (order Blattodea), and mantids (order Mantodea), all of which lack jumping hind legs. Phasmatodea have three tarsal segments, Blattodea and Mantodea five tarsal segments. Crickets are further classified in the suborder Ensifera, the members of which share jumping hind legs, two pairs of wings (rarely one) or none, either three or four tarsal segments, and thread-like antennae that are longer than the body except in subterranean forms.

Crickets all have long thread-like antennae, two slender tactual abdominal cerci, three tarsal segments, and some bulbous sensory setae basally on the insides of the cerci. No other insects share all these features; the last is closest to a single defining trait, shared by only certain Stenopelmatidae (Jerusalem crickets with four tarsal segments).

Some Ensifera have been called crickets mainly because they resemble Gryllidae, lack established common names, or otherwise have obscure family connections. All of these, however, have four tarsal segments and are probably related more closely to one another, and to other noncrickets, than to crickets. Examples are bush crickets (a term used in Europe for certain Tettigoniidae), sand and stone crickets (Schizodactylidae), and camel crickets, cave crickets, wetas, and their relatives (Tettigoniidae, Gryllacrididae, Rhaphidophoridae, Anostostomatidae, Stenopelmatidae, and Cooloolidae). Some caeliferan insects related to grasshoppers are called pigmy mole crickets and have two tarsal segments.

VARIATION

The smallest crickets are tiny, wingless forms comprising the subfamily Myrmecophilinae (ca. 1 mm); they apparently live and reproduce only in ant nests. The largest are the short-tailed crickets (Brachytrupinae) called bull crickets (ca. 5 cm); they excavate burrows a meter or more deep. Different cricket groups vary from having slender, fragile, whitish or greenish bodies with virtually transparent forewings (tree

FIGURE 1 Adult male *Gryllus pennsylvanicus.* (Photograph courtesy of David H. Funk.)

FIGURE 2 Adult male *Oecanthus latipennis,* with the forewings in singing position. (Photograph courtesy of David H. Funk.)

crickets: Oecanthinae, Fig. 1) to heavy-bodied, aggressive brown and black defenders of burrows and territories (field crickets: Gryllinae, Fig. 2, short-tailed crickets: Bachytrupinae; mole crickets: Gryllotalpinae). James Thurber said of one grylline, the sturdily built European burrowing field cricket *(Gryllus campestris),* that it "has the aspect of a wrecked Buick."

HABITATS

Crickets live in virtually all terrestrial habitats from treetops to a meter or more beneath the ground. Members of multiple subfamilies live in or near treetops and in bushes, grasses, and other herbaceous plants (Oecanthinae, Mogoplistinae, Eneopterinae, Podoscirtinae, Trigonidiinae) (Fig 3); on the soil surface (Nemobiinae, Gryllinae); in caves (Phalangopsinae, Pentacentrinae); and in shallow or deep burrows (Gryllotalpinae, Brachytrupinae). Some excavate burrows in logs or standing trees (Pteroplistinae). Some beach-dwelling species of Trigonidiinae run and jump readily on water.

FIGURE 3 Adult male *Orocharis saltator.* (Photograph courtesy of David H. Funk.)

Females of different groups lay eggs in stems or twigs, in wood, under bark, in the ground, or in burrows. Apparently all females in the widely distributed burrowing subfamilies Brachytrupinae (short-tailed crickets, 223 species) and Gryllotalpinae (mole crickets, 76 species) are parental toward their eggs and also toward their juveniles.

GEOGRAPHIC DISTRIBUTION

Crickets occur almost everywhere on the earth between the regions of taiga vegetation that start at approximately 55° north and south latitude. Excellent dispersers, they are the only orthopterans that readily colonize far-flung Pacific islands. Some colonizers fly, others tend to lay their eggs in wood, which is carried by water far and wide. Flightless soil ovipositors are poor colonizers, except via human transport. The largest numbers of cricket species audible from a single location occur in tropical regions. In Malaysia, along about 1 km of road northeast of Kuala Lumpur, the songs of 88 species can be heard, including almost every extreme of structure in cricket songs across the world. In the richest habitats of tropical Queensland (Australia), midwestern United States (southern Ohio), and some other tropical locations, up to about 25 species can be heard in single locales. Mute species, more difficult to locate, must be added to these numbers to obtain the number of cricket species in each region.

WINGS AND FLIGHT

The forewings of crickets, when present, are typically stiff and leathery; the hind wings are membranous and fold fan-like under the forewing when not being used. The hind wings can be miniature nonflying organs (microptery), longer than the forewings (macroptery), or absent. Some macropterous individuals shed their hind wings. The hind wings may also be pulled off and eaten by their bearer or by a female being courted by a macropterous male. Some macropterous crickets, such as the subtropical and tropical American species, *Gryllus assimilis,* take off, fly, and land so adeptly as to be wasp-like; others, such as mole crickets, fly in almost comically ponderous and slow manners, some with their abdomens hanging almost vertically.

SONGS AND COMMUNICATION

In most cricket species the males chirp or trill, producing clear, rhythmic, musical sounds distinctive to their family. An upturned scraper on one forewing is rubbed along a row of fewer than 10 to more than 1300 teeth, on the underside of the other forewing; tooth number, often species-specific, correlates with pulse rate and length. The dominant frequency in the sound depends upon the tooth-strike rate,

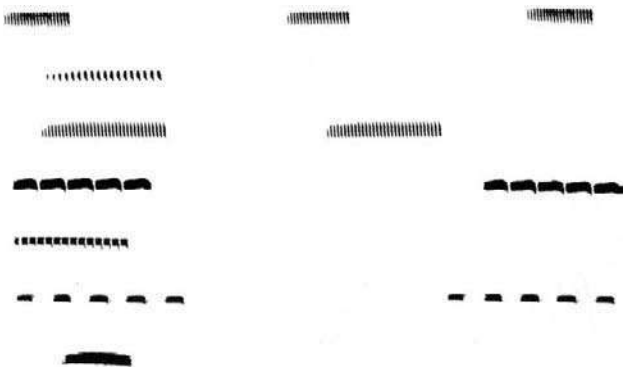

FIGURE 4 Drawings from audiospectrographs of the songs of 7 of the 17 known species of Western Australian desert crickets in the genus Eurygryllodes. Top to bottom the species are warrilla (a), warrilla (b), warrami, wirangis, yoothapina, buntinus, and diminutus. *E. warrilla* (a) and (b) have not yet been treated as different species because too little is known about them, and the available specimens have not been distinguished morphologically (from Otte and Alexander, 1983, p. 81).

ca. 1.5 to 10 kHz (Figs. 4 and 5). Sounds are pulses caused by the individual closing strokes of the wings and separated by the silent opening strokes. Pulse rates vary from one every 3 s (a Hawaiian trigonidiine) to more than 200 per second (a Malaysian gryllotalpine). The communicative significance of the songs lies in the rates and patterning of the pulses; pitch is important because the tympana on the forelegs are for each species tuned to the sounds of conspecific individuals. The calling song, best known, attracts females and challenges nearby males. The courtship or mating song induces the female to move into the copulatory position, and the fighting chirps (most obvious in surface-dwelling and subterranean species) challenge other males at close range.

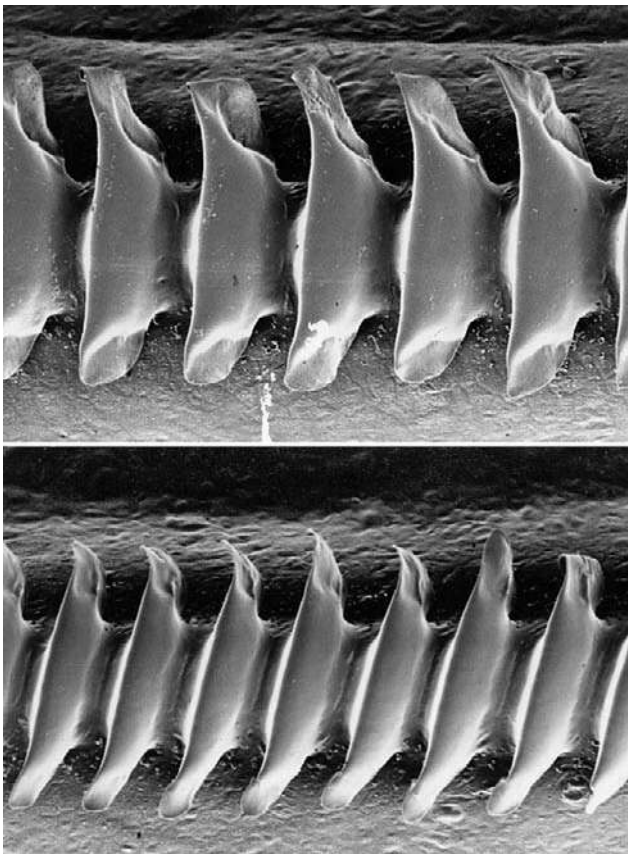

FIGURE 5 Teeth on the stridulatory files of field crickets (genus *Gryllus*) from the Galapagos Islands.

Cricket sounds vary with temperature. The North American snowy tree cricket *(Oecanthus fultoni),* sometimes called the thermometer cricket, chirps at a steady countable rate, such that degrees Fahrenheit is represented by the number of chirps in 15 s plus 40. Individuals of this nocturnal species also synchronize their chirps, so that a dense colony produces an intense, monotonous beat. Ralph Waldo Emerson said that if moonlight could be heard it would sound like this intense, beautiful cricket song.

Tree- and bush-living crickets tend to sing only at night, those in grasses and on the ground both day and night; presumably this situation has come about because of a history of trade-offs between obtaining mates and avoiding predators, mainly birds. Some crickets are also parasitized by flies that locate males by song. In Europe and Australia certain sounds of nightjars (Caprimulgidae) are so similar to the songs of mole crickets in the same locations that it seems likely (though unproved) that these insectivorous birds attract mole crickets, both sexes of which respond to songs in flight.

Crickets were among the first musicians on earth. The calling song of each species today—as with all so-called "singing" insects—is invariably distinct from the songs of all other species that breed in the same times and places. Species that mature and mate in different regions, or at different times in the same region, sometimes have songs so similar as to be unlikely to be distinguished by the crickets.

Cricket acoustical communication evidently evolved but once, yet has been lost many times. In Australia the stridulum (file) has been lost at least 27 times, in Africa 17 times. The auditory tympana are often retained after acoustical communication has ceased, but only in flying species, suggesting that the tympana are used to avoid bat predation (Fig. 6).

Varying shapes and locations of the auditory tympana on the fore tibiae of crickets, and variations in the structure of the stridulatory device on the forewings of males, cause the morphological devices of crickets, as well as their songs, to be important to students of phylogeny and classification. Because of their species distinctiveness, cricket songs are unusually fine tools for locating previously unknown species and for the rapid study of geographic and ecological distribution, biogeography and phylogeny, species density variations, population size, seasonal and geographic overlap of species, character displacement, and the nature of the life cycle and overwintering stage from the seasonal distribution of adults.

Biologists have analyzed cricket songs intensively for almost 70 years because they are audible, recognizable, and amenable to sophisticated acoustical, physiological, and

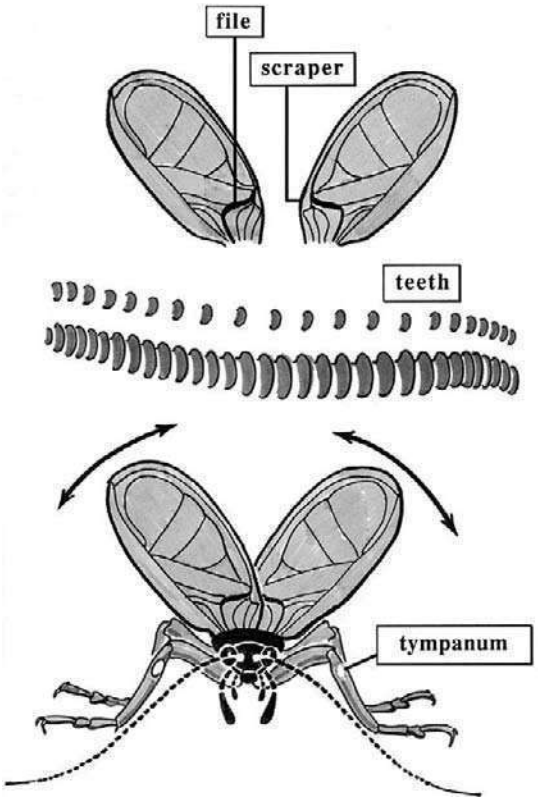

FIGURE 6 The sound-producing and auditory apparatus of a tree cricket. The stridulatory file is located on the underside of the forewing near its base. Each time the forewings close during song the teeth are rubbed against a scraper located on the underside of the other forewing. The forewings are thus caused to vibrate, more or less as a whole, as fast as the teeth are struck, producing separated pulses of a whistle-like tone. The hearing organ, or tympanum, is located on the upper portion of the front tibiae. It is tuned to the frequency of its own species' wing vibrations, causing it to transmit the rate and pattern of pulse production of all songs at that frequency to the central nervous system. The central nervous system is able to distinguish its own species' songs from those of other species with songs of similar frequencies that sing in the same times and places. In some crickets the tympanum is exposed through openings on both the inner and the outer faces of the tibia; in others there is only one opening, most often on the outer surface of the tibia, but in some (typically burrowing) species on the inner surface only.

behavioral analysis and because hybridization of different species provides insight into the genetics of song differences. As a result of the use of song to locate new species, during the past 50 years the number of known cricket species has more than doubled. With respect to genetic background, nerve and muscle physiology, and behavioral functions, the cricket acoustical system is almost certainly the most thoroughly understood of all animal communicative systems.

EGG-LAYING

Most female crickets inject their eggs into the soil or into plant stems through long, slender ovipositors. The oviposition slashes of tree crickets often seriously damage berry canes and small twigs. Females of the two subterranean subfamilies do not inject their eggs into the soil and have lost the external ovipositor; a few Brachytrupinae retain short ovipositors and inject eggs shallowly into burrow walls. Most subterranean forms lay only a few eggs in one place in an underground chamber. In those studied, such as the North American species *Anurogryllus arboreus,* plant materials are also stored in the burrow, and a special defecation chamber is excavated; juveniles eat stored plant materials, small unfertilized eggs produced by the mother after the young hatch, and eventually the body of the dead mother. A female may dig her own burrow or ferociously take over and defend the burrow of a courting male before allowing him to mate with her at the entrance.

LIFE HISTORIES

In northern (and probably southern) latitudes most crickets overwinter as eggs and mature in late summer. A few burrowers overwinter as partly grown juveniles and mature in early summer. There are 6 to 12 nymphal molts, and the adults usually live 6 to 8 weeks. In latitudes with significant winters, life cycles vary from one generation every 2 years in a mole cricket to two generations each year. Diapause occurs in the overwintering stage. Nondiapausing crickets such as the house cricket *(Acheta domesticus)* have a generation time of a few weeks, varying with rearing temperature. Diapause also occurs during droughts in some tropical countries. Eggs or adults live through droughts, with rain causing nymphs to hatch and adults to oviposit.

FOOD

Many crickets are omnivorous. Some seem to feed almost entirely on vegetable matter, yet sometimes consume carrion and even ferociously kill and eat other insects. Several species frequent human dwellings and refuse heaps, most notably *A. domesticus* and the decorated cricket, *Gryllodes sigillatus*. Subterranean species feed mostly on roots and can be injurious when abundant in crops, gardens, lawns, golf courses, and newly reseeded forests.

HUMAN USE

House and field crickets are reared and sold in large numbers as fish bait and food for laboratory animals in many parts of the world. In the Orient, male crickets are caged for their songs, and staged battles between aggressive males in surface and subterranean species have been a favorite sport for over 1000 years. Males can be primed for serious fights—even to the death—by tickling them with brushes resembling antennae and by providing them with lairs made of small boxes that enhance their motivation to defend the site. Burrowing crickets can be extracted from their burrows by eliciting aggressive reactions to tiny pebbles rolled into the burrow followed by blades of grass used so as to imitate cricket antennae. Allowing a male to mate sets him into guarding the female for further mating, thus also priming him to fight ferociously.

MATING SYSTEM

The long-range female-attracting songs and long tactual cerci of crickets are components of a unique mating system, some aspects of which evidently trace to the earliest instances of copulation in the insect line and help explain changes leading to the current major groups of insects. Thus, none of the primitively wingless modern insects copulate, while all winged and secondarily wingless insects do, the majority with the male mounting the female and in some way holding or forcing her. In primitively wingless insects, however, a sac or bulb containing the sperm (a spermatophore) is transferred indirectly to the female without direct copulation. Like crickets, some of these particular primitively wingless insects possess prominent tactual cerci (e.g., Thysanura), used to guide the female during spermatophore transfer, as also in cockroaches and mayflies. In all insect groups of ancient origin that have prominent tactual cerci, transfer of the spermatophore is a luring act in which the female either mounts (winged and secondarily wingless forms) or stands beside the male (primitively wingless forms). In some crickets, such as the field cricket genus *Gryllus,* the copulatory act appears unique among all animals in being entirely luring, with no evidence of controlling force by the male at any stage. The female is attracted initially by the long-range calling song and then by the male's close-range courtship song and probably the fluttering touches of his antennae (Fig. 7). As in nearly all crickets, most close relatives of crickets, and most cockroaches and mayflies (the last aerially), the female mounts (or flies above) the male in the copulatory act. Apparently in correlation with the male field cricket having minimal ability to clasp the female's genital parts, the spermatophore is transferred quickly, in 15 to 90 s. The spermatophore is osmotically self-emptying, so that sperm injection occurs largely after the female dismounts from the male. In forms related to crickets, such as Tettigoniidae and Caelifera, in which males have evolved terminal claspers on the abdomen, the tactual cerci have disappeared and copulation is much lengthier. In Caelifera the mating act has evolved such that the male mounts the female, though still reaching beneath her to attach the genitalia; here, unlike Tettigoniidae, the antennae have also become much shorter. Apparently luring copulatory acts in insects have repeatedly evolved into acts involving significant force, but the reverse does not seem to have happened. Groups of features related to the history of insect mating acts have significance for interpreting changes in diagnostic features of major groups of insects, including cerci, antennae, genitalia, wing structure, long-range communication, and modes of pair formation.

FIGURE 7 Adult female (left) and male (right) *Phyllopalpus pulchellus*, the latter with forewings in singing position. (Photograph courtesy of David H. Funk.)

Distinctive morphological and behavioral features of crickets, especially those related to their methods of pair formation and mating behavior, make them a pivotal group in understanding insect evolution and phylogeny.

See Also the Following Articles

Cultural Entomology • Folk Beliefs and Superstitions • Hearing • Orthoptera

Further Reading

Alexander, R. D. (1961). Aggressiveness, territoriality, and sexual behaviour in field crickets (Orthoptera: Gryllidae). *Behaviour* **17,** 130–223.

Alexander, R. D. (1966) The evolution of cricket chirps. *Nat. History Mag.* **75,** 26–31.

Alexander, R. D. (1968). Life cycle origins, speciation, and related phenomena in crickets. *Q. Rev. Biol.* **43,** 1–42.

Alexander, R. D. (1969). Arthropods. *In* "Animal Communication" (R. Sebeok, ed.), 1st ed., pp. 167–215. Indiana University Press, Bloomington.

Alexander, R. D., and Otte, D. (1967). The evolution of genitalia and mating behavior in crickets (Gryllidae) and other Orthoptera. University of Michigan Museum of Zoology Miscellaneous Publication No. **133,** pp. 1–62.

Huber, F., Moore, T. E., and Loher, W. (eds.) (1989). "Cricket Behaviour and Neurobiology." Cornell University Press, Ithaca, NY.

Otte, D. (1992). Evolution of cricket songs. *J. Orthoptera Res.* **1,** 25–49.

Otte, D. (1994). The crickets of Hawaii: Origins, systematics, and evolution. *In* "Publications on Orthopteran Diversity." The Orthopterists' Society, Academy of Natural Sciences, Philadelphia.

Otte, D. (1994). Orthoptera species file 1: Crickets (Grylloidea). *In* "Publications on Orthopteran Diversity." The Orthopterists' Society, Academy of Natural Sciences, Philadelphia.

Otte, D., and Alexander, R. D. (1983). The Australian crickets (Orthoptera: Gryllidae). Academy of Natural Sciences of Philadelphia Monograph No. **22,** pp. 1–447.

Otte, D. and Naskrecki, P. (1997). Orthoptera species file online at http://viceroy.eeb.uconn.edu./Orthoptera.

Crypsis

Paul M. Brakefield

Leiden University

Organisms with bright and conspicuous color patterns tend to attract the most attention both scientifically and aesthetically. However, the majority of insects and other animals rely on camouflage or crypsis for survival from predators that hunt them by sight. Furthermore, crypsis may extend to include the other senses, namely, smell, touch, and sound. Indeed, any stimulus or signal that can alert a potential predator could be expected to become part of a

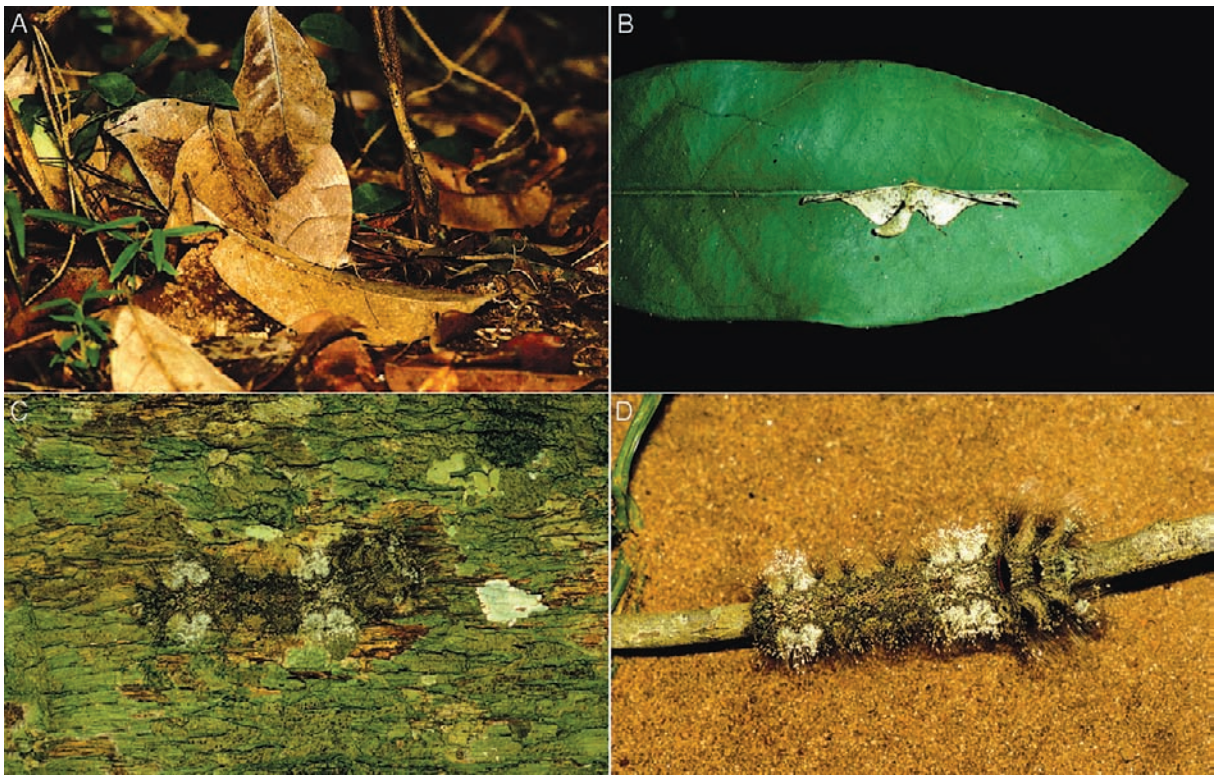

FIGURE 1 Crypsis illustrated for different insects. (A) An individual of the dry season form of the evening brown, *Melanitis leda,* resting among dead leaves on the forest floor in the Shimba Hills, Kenya. The insect is at the center with head pointing to the right; forewing length is ca. 4.5 cm. (B) A small moth that resembles a dead patch on a large leaf in a forest in Costa Rica (wing span is ca. 3 cm). (C) The caterpillar of a moth of the family Lasiocampidae resting on a tree trunk in the Shimba Hills, Kenya; it is ca. 6 cm in length and is positioned horizontally, head to the right, in the center of the photo (image has been rotated 90 degrees). (D) The same larva when actively moving in the same direction along a twig. See text for further details. (Photographs by the author.)

coordinated suite of cryptic traits. A form of crypsis is also shown by some predators that disguise themselves by assuming the same color and patterns as the background on which they hunt. H. B. Cott in 1940 wrote perhaps the best known book on animal color patterns, but many of the great entomologists of the 19th century had already considered insect camouflage. It is not usual to consider insect crypsis as a subject of applied biology but there are certainly many parallels with military expertise in either the hiding of or the searching for personnel and armaments in a landscape.

COLOR MATCHING AND CRYPSIS

An insect that is perfectly camouflaged is perhaps one of the most striking exhibitions of the power of evolution by natural selection to mold and adapt organisms to fit their environment and to maximize survival and reproductive success. Wonderful examples of camouflage are presented by many species of insects, including some butterflies in tropical forests (Fig. 1A), which rest on carpets of dead brown leaves. The apparent perfection of crypsis is emphasized in many such insects by a similarity of, and matching of, the color pattern of the wings, body, and appendages to the background on which they normally rest. The color pattern of these different body parts and structures must involve different genetic and developmental pathways, and yet evolution has led to a corresponding perfection of matching, albeit using entirely different mechanisms of pattern formation. Such an example of an underlying complexity of patterning is given by some caterpillars of the family Lasiocampidae that rest on the bark of trees and survive by resemblance to the background color pattern of the bark, including epiphytic lichens and algae (Fig. 1C). Such larvae are encircled by long hairs that are flattened around their margin when at rest. This breaks up their shape, smoothing their outline. These hairs are also patterned in a very specific way and one that is fully coordinated with the body cuticle, including the short bristles of the dorsal areas of the body segments. These elements are exposed, and the whole insect becomes highly conspicuous as soon as a larva is forced to move along a twig of fine diameter (Fig. 1D).

Furthermore, color matching in cryspis is almost always only one component of the strategy for survival; both habitat choice and, frequently, the adoption of very specific patterns of behavior and activity are required for effective crypsis. One such example is shown by some species of moths that attain crypsis by appearing to be a dead patch of tissue within a large leaf on which they rest (Fig. 1B). They achieve this not only through the generally brown color of their wings and some details of patterning, which may resemble small

patches of fungal-attacked leaf tissue, but also through a precise positioning on the leaf. For example, the moth in the photograph of Fig. 1B has rolled up the leading edge of its forewing, wrapped its abdomen along the trailing edge of one hind wing, hidden its appendages, and positioned itself alongside the midrib of the leaf.

Despite the potential fascination of understanding crypsis, it is only relatively recently that scientists have begun to analyze what is meant precisely when it is stated that an organism is well camouflaged. John Endler in 1978 stated that "a color pattern is cryptic if it resembles a random sample of the background perceived by predators at the time and age, and in the microhabitat where prey is most vulnerable to visually hunting predators." There are several crucial components in this definition. First, a color pattern is cryptic only with respect to the specific environment in which the organism is potentially encountered by the predator or the guild of predators to whom the pattern is an adaptive response. What is a cryptic pattern on the resting background of that environment may be conspicuous and ineffective on any other background. Second, the effectiveness of a particular pattern is considered with respect to the normal time and lighting conditions under which crypsis is functional. Third, to be cryptic the color pattern of a prey organism must essentially reflect a random sample of the background on which it rests.

INDUSTRIAL MELANISM AND CRYPSIS

Perhaps the first analysis of crypsis and the evolution of a color pattern from the perspective of changes in camouflage involved industrial melanism in the salt-and-pepper moth, *Biston betularia.* Industrial melanism refers to an association of high frequencies of dark, melanic forms or phenotypes of a species with high levels of air pollution. The fundamental components of this classic example of the evolution of an adaptive trait also apply to numerous other species of moth and other insects that have evolved melanism as a response to environments influenced by air pollution. These components are: (1) the environment was changed by air pollution in such a way that the camouflage of the "typical" or wild type of color pattern was impaired, (2) a mutant phenotype occurred in this new environment that had a functional design or color pattern that improved survival from birds hunting the moths at rest, and (3) the dominant allele at the gene that specified this favored mutant phenotype then increased in frequency under the influence of natural selection, leading to the species exhibiting industrial melanism.

In the salt-and-pepper moth, we know from museum collections that prior to the middle of the 19th century in northern England the moths had pale-colored wings with a speckling of dark dots (the typical form). Also, up until that time in the early industrial revolution the bark of trees was predominantly pale and covered in epiphytic lichens and algae. The salt-and-pepper moth rests on bark, and females lay their eggs under foliose lichens or in cracks in the bark. The moths are active at night and rely on background matching and crypsis for survival from birds during daylight hours. Survival enables males to mate at night and females to lay their eggs over a number of nights. The gaseous (e.g., sulfur dioxide) and particulate (soot) air pollution produced by industry both killed the epiphytic communities on the trees and blackened the resting surfaces of the moths. The typical, pale-colored moths became more conspicuous. The fully black, melanic form known as *carbonaria* was not collected until 1848, near Manchester. It may have occurred shortly before through a mutation (producing a new allele of the gene), or perhaps it had already existed for some time in that region as a rare allele. Whatever its precise origin, the *carbonaria* form rose rapidly in frequency and spread extensively through the industrial regions of Great Britain over the following decades; the adult moth as well as newly emerged larvae can move long distances. Clear geographical associations were established between the amount of air pollution and the frequency of the fully melanic *carbonaria* and also of several intermediate melanic forms known as *insularia.*

Up until the mid-20th century this remained a verbal, albeit persuasive, reasoning for the evolution of melanism as an adaptive response to a changed environment. It was only then that some classic early experiments in evolutionary biology began to add scientific rigor to this explanation. Several researchers performed a series of experiments that showed beyond doubt that, whereas the survival of the pale typical form was higher in rural, unpolluted regions of Great Britain than that of the *carbonaria* form, this relationship is reversed in the polluted industrial environments. Although there have been discussions about the precise details of some of these types of experiments, the fundamental finding of a switch in survival and relative fitnesses (reproductive success) of the pale and dark phenotypes across the extreme environments, principally the result of corresponding changes in crypsis, has been corroborated. Other differences in fitness among the phenotypes that are not directly related to the visual differences in color pattern may also be involved in determining the precise dynamics of the evolution.

There has, however, more recently been an additional finding that proves beyond any doubt the role of evolution by natural selection. Great Britain and other countries in northern Europe have over the past few decades reduced levels of air pollution from soot and gases such as sulfur dioxide. This has in turn led to declines in the frequencies of the melanic forms and the coining of the phrase "evolution in reverse." As the resting environment returns, at least in a qualitative sense, back toward the original, unpolluted state, the relative fitnesses are also reversed, leading to present-day declines in melanism. Although it has not been precisely quantified, the conclusion must be that in previously polluted regions, while the fully black melanic *(carbonaria)* has again become conspicuous and vulnerable to birds, the paler typicals have become well camouflaged on the changed background.

ANALYSIS OF CRYPSIS

This example of the salt-and-pepper moth illustrates that crypsis still needed to be scientifically measured and fully quantified. In 1984 Endler began to use early techniques of image analysis to mathematically describe how well matched in terms of color patterning were moths in a North American woodland community with respect to different potential resting environments. If crypsis is "optimal" the patterning of the insect will represent a random assemblage of the pattern elements of the background. Endler also pointed out that there will be matching with respect to different components of the color patterns of both insect and resting background, namely, size, color, shape, and brightness. In some backgrounds, such as pine needles or bark with striations, the component of orientation should also be added. Failure to match with respect to any one of these components will lead to mismatching and ineffective crypsis. Because the color vision of many predators, including birds and insects, extends into the ultraviolet part of the spectrum, when color matching in crypsis is considered it often has to include the UV. Researchers have recently begun to use computer-generated patterns, image analysis, and "visual predators" to explore more fully the potential effects of interactions among predators and their prey that lead to the evolution of cryptic color patterns.

Cryptic color patterns may also include an element of banding, which is disruptive and can serve to break up the outline of the prey. Usually, such an element also has to blend into the resting background in terms of the prey representing a random assemblage of its pattern. However, this restriction is perhaps relaxed when crypsis is used only to protect a prey from a distance, such as in the brightly colored, banded moth caterpillers, including the cinnabar, *Tyria jacobaeae,* and the strikingly striped forewings of some arctiid moths, for example, *Callimorpha quadripunctaria.*

CRYPSIS AND NATURAL SELECTION

Although testing of these ideas, at least in the context of animal color patterns and their camouflage, has not been completed, Endler has also performed experiments with guppies that dramatically illustrate the power of natural selection to lead to the evolution of effective crypsis. Male guppies can be very colorful with a patterning of bright spots and patches on their lateral flanks and fins. Laboratory experiments in which females can choose whether to mate with males of different patterns show that there is female preference for the more brightly colored males. In the wild in Trinidad, there is a correlation between the degree of color patterning on males in a population and the presence of predatory fish and invertebrates ranging from weak to strong mortality factors on guppies. Male fish are colorful and brightly patterned when either no predators or only weak predators are present, whereas they are drab and unpatterned brown fish when strong predators such as certain cichlids are present. A series of experimental pools with natural backgrounds in a greenhouse was established to examine the efficacy of natural selection on crypsis in this system. Endler showed that guppy populations with the weak predators showed no divergence over subsequent generations in their average color pattern; in contrast, in those pools to which strong predators were added the guppies showed a marked and progressive decline in the brightness and spottiness of the males. This result was highly consistent with selection favoring a more effective crypsis through a lower conspicuousness and improved background matching of the prey populations. In the absence of such strong predators, the balance of sexual selection through female choice and of natural selection by visually hunting predators favors colorful males because they survive to maturity and then achieve a higher mating success than their less colorful competitors.

Such a balance of selection on animal color patterns is probably the norm in natural populations. Thus, in animal communication, a color pattern is usually a compromise between being conspicuous to conspecifics and being poorly visible to predators (or prey). Indeed, one of the potential disadvantages of adopting crypsis as the primary means of survival is that it almost inevitably ties the organism down to a sedentary style of life at least during the hours of daylight. In contrast, when organisms are distasteful and adopt a conspicuous, aposematic lifestyle or when they evolve Batesian mimicry to resemble such warningly colored species, there is no such disadvantage associated with daytime activity.

INTERACTION OF CRYPSIS AND OTHER DEFENSES

In many insects, an organism may not rely only on crypsis for survival. There may be some secondary means of defense once crypsis has failed and the prey has been detected by a potential predator. Insects that are cryptic at a distance but conspicuous when seen close up (including the banded larvae and arctiid moths mentioned above) are often chemically protected. This type of multiple defense is also illustrated by the moth caterpillar in Fig. 1C. If the caterpillar is disturbed and begins to move it can expose a series of glands in the dorsal cuticle of several segments toward the front of the body. These are visible as a pair of partial bands in Fig. 1D, the largest immediately to the right of the largest white-colored region. These produce a pungent odor and probably provide a potential chemical defense against birds and other predators.

The effectiveness of crypsis will also show complex interactions with the visual processing abilities of the specialist predator or the guild of predators. Some insects that rely on camouflage for survival often exhibit extreme individual variation. One example is the tropical evening brown, *Melanitis leda.* This large brown butterfly is common

throughout the Old World tropics. In wet–dry seasonal environments, the species shows classical seasonal polyphenism (i.e., distinct color patterns that result from phenotypic plasticity), with a wet season form having conspicuous marginal eyespots and a cryptic dry season form without such eyespots. The latter form relies on survival through crypsis on a resting background of dead brown leaves (Fig. 1A). In large numbers of the dry season form it is difficult to find two individuals with exactly the same color pattern. Dramatic variation across individuals is produced by high genetic variation in several different pattern elements across the wing (such as the contrast and brightness of particular patches and bands and the background wing color in different regions). This variation can be interpreted as an evolutionary response involving "apostatic selection" to make it more difficult for browsing predators in the leaf litter to form a specific "search image" for a particular form of dead leaf pattern corresponding to the color pattern of the prey. Although like many of the detailed ideas about the significance of crypsis and particular animal color patterns, this hypothesis remains to be tested rigorously, it does once again illustrate the fascination of crypsis.

See Also the Following Articles

Aposematic Coloration • Defensive Behavior • Eyes and Vision • Industrial Melanism • Mimicry

Further Reading

Bond, A. B., and Kamil, A. C. (2002). Visual predators select for crypticity and polymorphism in virtual prey. *Nature* **415,** 609–613.

Bradbury, J. W., and Vehrencamp, S. L. (1998). "Principles of Animal Communication." Sinauer, Sunderland, MA.

Cott, H. B. (1940). "Adaptive Coloration in Animals." Methuen, London.

Endler, J. A. (1978). A predator's view of animal colour patterns. *Evol. Biol.* **11,** 319–364.

Endler, J. A. (1984). Progressive background matching in moths, and a quantitative measure of crypsis. *Biol. J. Linn. Soc.* **22,** 187–231.

Kettlewell, H. B. D. (1973). "The Evolution of Melanism." Clarendon Press, Oxford.

Majerus, M. E. N. (1998). "Melanism: Evolution in Action." Oxford University Press, Oxford.

Sargent, T. D. (1976). "Legion of Night: The Underwing Moths." University of Massachusetts Press, Amherst.

Thery, M., and Casas, J. (2002). Predator and prey views of spider camouflage. *Nature* **415,** 133.

Cultural Entomology

James N. Hogue

California State University, Northridge

Since the dawn of humanity, the organisms that share our world have captured our imagination and influenced our thoughts, dreams, and fears. This influence is particularly true of insects, which impact nearly every facet of human activity. In addition to serving as objects of scientific inquiry, competitors for resources, carriers of disease, and food, insects have made a marked impact on the cultural aspects of human societies. Cultural entomology is the study of the role of insects in those human affairs that are practiced for the nourishment of the mind and soul, such as language and literature, music, folklore, religion, art, and recreation. These activities that pervade primitive and modern human societies are concerned primarily with life's meaning rather than its function.

Despite their extra appendages and different strategies for making a living, insects look and behave enough like humans to serve as models for friends, enemies, teachers, and entertainers. This status permits insects to act as objects on which to impart human qualities and as the source of qualities that can be incorporated into the framework of human ideology and social structure. It is not surprising then to find insects playing a host of roles in the oral and written traditions throughout human history, ranging from folk tales to the holy writings of the world's most prominent religions.

FOLKLORE, MYTHOLOGY, AND RELIGION

The derivation of stories and myths is a universal tendency of all human societies. Both myths and folk tales differ enormously in their morphology and their social function. They are used to mediate perceived contradictions in phenomena observed in the natural world, they serve as vehicles of wish fulfillment, they may embody a lesson, or they may serve to preserve a piece of a culture's history. Myth and folklore also differ from one another in their origin and purpose, but application of these distinctions is difficult to discuss here. Originally, mythology meant no more than telling stories, such as traditional tales passed from generation to generation. Later, some of these tales acquired new meaning and status and evolved more symbolic or religious functions. All tales, whether classified as folklore or myth, are not generated in isolation, but derive their inspiration, elements, and messages from the environment, including the host of other species that surround us. These tales are often used to derive commonsense explanations of natural phenomena observed in the environment. Conversely, such observations may also serve as the basis for the superstitious beliefs and tales surrounding aspects of human existence such as healing practices and other utilitarian activities such as agriculture.

Entomological mythology commonly employs transformations of beings between the insect and the human form (and combinations thereof), the acquisition of souls by insects, and ultimately the deification of insect forms. Insects are also used symbolically throughout the world's religions in a variety of roles.

Insects figure prominently in the creation myths of many cultures. The widespread recognition of insects in this role probably stems from an innate recognition of insects as ancient members of the living world that must have been

present at its creation or soon thereafter. Beetles, for example, play central roles in the creation myths of two native American tribes. According to the Cherokee of the southeast, the world was originally covered by water. The first land was brought forth by the water beetle that dived under the water and brought mud to the surface.

The behavior of beetles in the genus *Eleodes* (Tenebrionidae), which raise their abdomens in the air by standing on their heads when disturbed, is explained by the role this beetle played in the creation of the universe according to the Cochiti of the American Southwest. The beetle was given the responsibility of transporting a bag of stars that would later be carefully named and placed into the sky. But the beetle's carelessness resulted in most of the stars being spilled into what is now the Milky Way. He was punished with blindness and today expresses his embarrassment at his mistake by hiding his head when approached. Any hope of this beetle regaining its sight was apparently completely lost at a later date, as this seems to be the same beetle that lost his eyes in a bet with a spider; this is how the spider got its extra sets of eyes.

Insects appear throughout Mayan codices and Aztec reliefs. The use of insects in this manner indicates an appreciation of their existence and their inclusion in cultural events. In addition to scorpions and some unknown bugs and worms, references to seven different insects are found in the Mayan book of the dawn of life, the Popul Vuh. These include lice, leafcutter ants, mosquitoes, fireflies, bees, yellowjackets, and another type of wasp. Yellowjackets were used as weapons by the Quiché against the enemy tribes during an attack on the Quiché citadel at Hacauitz. Fireflies were used by the brothers Hunahpu and Xbalanque, who later became the sun and the full moon, respectively. They placed these insects in the tips of cigars as false lights to deceive the Xibalban sentries of the underworld that watched over them during their night in the Dark House.

Observations of metamorphosis led people throughout history and from various parts of the world to equate pupation with death of the earthbound larva and the emergence of the often beautiful, winged adult with resurrection. The adult insect is additionally equated with the soul in many circumstances. The equation of souls or spirits of the afterlife with imaginal insects may be why angels are traditionally depicted bearing wings. Insect analogies in descriptions of death, resurrection, and the journey to the afterlife continue to be used today. For example, a Doris Stickney book uses a story of growth and emergence of dragonflies to explain a Christian concept of death to young children.

Insects have also been incorporated into the astrological and cosmological traditions of various societies. Aquatic insects were used as water symbols associated with the coming of rain by Chumash astrologers of southern California, who believed that rain was a gift from the sun. The guardians of the four cardinal points in Warao (Orinoco delta of Venezuela) cosmology are insects: arboreal termites, two kinds of stingless bees, and a paper wasp. There is even a constellation of the southern fly, *Musca australis.*

In Aztec culture, Xochiquetzal, represented by the swallowtail butterfly, *Papilio multicaudatus,* was the goddess of beauty, love, and flowers; patron of domestic labor and the courtesans; and the symbol of the soul and the dead. The mother deity and goddess of human sacrifice, war, and travelers, Itzpapalotl (the saturniid moth *Rothschildia*), was also the personification of the earth and moon. Images of these and other deified insects are found in many Aztec and Mayan reliefs.

The most famous deified insect is the scarab of ancient Egypt. The scarab is a symbol of the sun god Khepera (Fig. 1) and also equated with the creator god Atum. One representation of the scarab was as the agent responsible for moving the sun through the sky, in the manner that these beetles move balls of dung across the ground. Another prominent representation of the scarab was that of the soul emerging from the body, and it was commonly associated with mummies. Just as the actions of the beetles and balls of earth and dung give rise to new beetles, the buried human dead will rise again. Scarab figures are nearly always found on Egyptian mummy sarcophagi, and amulets and pendants bearing the scarab likeness were worn as jewelry by royalty

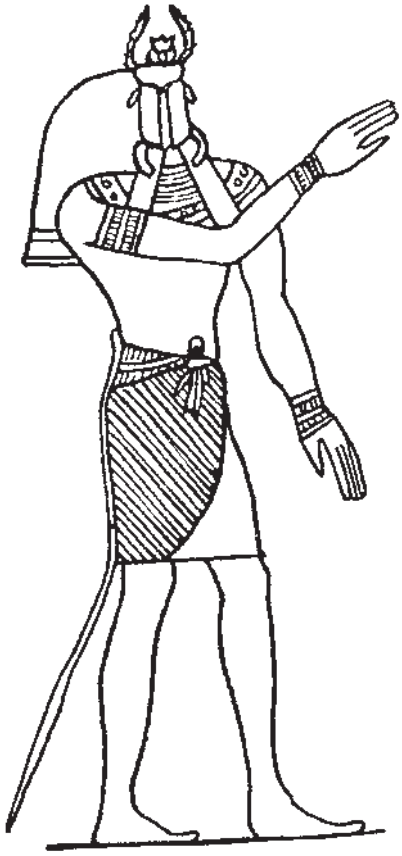

FIGURE 1 Depiction of the Egyptian scarab god, Khepera. [Illustration modified from Bodenheimer (1928).]

and included in funeral caches as symbols of new life. Another testament of the association of these beetles with life comes from Saint Ambrose, the Archbishop of Milan, who wrote of Jesus as "the good Scarabaeus, who rolled up before him the hitherto unshapen mud of our bodies."

Recent English translations of the Bible, based solely on the original texts, have shed new light on biblical references to insects, particularly with respect to the identity of the insects themselves. Of the 98 references to insects in the Revised English Version, most focus on negative aspects of their activities and as vehicles for God's wrath. Three of the 10 plagues (maggot infestation, swarming flies, locusts) visited upon Egypt prior to the Exodus were mediated by insects. Other references deal with more utilitarian or beneficial aspects of insect life. Insects are included as part of the instructions of what kinds of animals are permitted as food (Leviticus 11:22), some writings are merely observations of insects and their habits (Exodus 16:20), and other references use them as examples of virtuous characteristics (Proverbs 6:6–8 and 30:25–27). Sometimes insects are used metaphorically, as in Psalms 118:12: "They surround me like bees at the honey, they attack me, as fire attacks brushwood, but in the Lord's name I will drive them away." One or more kinds of scarab beetle may have even served as the inspiration for the prophetic visions of Ezekial.

SYMBOLISM AND REVERENCE

Throughout human existence, many insects have been admired for their ingenuity, beauty, fantastic shapes, and behaviors. In some instances, the use of insects as totemic figures that may symbolize ancestry or kinship of humans with these organisms leads to a deep sense of adoration and reverence. In other cases, the resultant admiration has developed into a reverence for their inspirational and historical nature and a medium for symbolizing a variety of aspects of human life. In these situations, characterizations of organisms, in both illustration and sculpture, act as vehicles to convey human feelings rather than as objective expressions of entomological facts. Insect symbolism is best developed in the most advanced ancient civilizations of Egypt, Greece, and especially Central America, where the people were surrounded by a multitude of insects.

Of all the insect groups, the flies (Diptera) most frequently play negative roles in human symbolism. Flies typically represent evil, pestilence, torment, disease, and all things dirty. This association is likely a result of the fact that those flies most familiar to people have a close association with filth. Beelzebub, the Lord of the Flies, is a fallen angel who presides as a leader of demons and an agent of destruction and putrefaction. In the ancient lore of Persia, the devil Ahriman created an evil counterpart for every element of good put on Earth by the creator. Many insects, particularly flies, were thus formed and they continue to be associated with evil and filth. Some flies were considered so bad that they became symbols of qualities revered by humans. The Order of the Golden Fly was a military decoration of the New Kingdom of Egypt (1550–1069 B.C.) awarded for valor. Derived from encounters with the stable fly, *Stomoxys calcitrans,* soldiers observed these flies to fiercely bite and return to bite again, even in the face of persecution.

Because of the perceived similarities between human and insect societies, social insects figure prominently in the symbolic representation of insects. Social insects such as ants, termites, and some bees represent desirable qualities such as unity, cooperation, and industriousness. For example, ants represent the benefits of teamwork and cooperation for the good of all. Many symbolic depictions feature the ancient activities of honey hunting and beekeeping. In Europe, bees and hives also are widely used in various signage and as heraldic emblems, perhaps extolling various qualities of bees upon their bearer. A fine example of the latter is found on the coat of arms of Pope Urban VIII, Maffeo Barberini, who consecrated the present church in St. Peter's Basilica in 1626. The three Barberini bees adorn various ornamentations at the church and many papal objects located in the Vatican museum, including the building itself. In the United States, honey bees are used to symbolize virtuous qualities. The designation of Utah as the "Beehive State" originates from the adoption of the beehive as a motif by the Mormon leaders in 1849 and may be based on impressions of the bees as hard-working, industrious creatures.

Some insect groups have such wide representation in the symbolism of past and present human societies that it is impossible to make general statements about their meaning. Butterflies and moths, for example, are very common elements in symbolism of societies worldwide. Within the limited scope of Western art, Ronald Gagliardi describes the use of butterflies and moths in 74 different symbolic contexts. These insects adorn the artwork of many societies, not only because of their beauty but also because they are widely used to symbolize spirits. Butterflies are often equated with the souls of the dead or sometimes of souls passing through Purgatory (Irish folklore) and are thus often used to represent life after death. The Greek goddess Psyche, who represented the soul, is typically depicted bearing butterfly wings. Moths are depicted as a symbol of the soul's quest for truth, and just as the moth is attracted to a light, so the soul is drawn to divine truth.

Butterfly images are common adornments of pottery, featherwork, and the deeply religious characters hewn in stone in ancient Mexico. The Hopi of the American Southwest have a ritual called the "Butterfly Dance" and have kachina figures that anthropomorphize the butterfly spirit. The Blackfoot Indians of North America believe that dreams are brought to sleeping people by the butterfly.

Symbolic depictions of insects also serve to bestow honor on the insects themselves. Insects have been featured on a few coins and on several thousand postage stamps worldwide (Fig. 2). One fifth century Roman coin bears a honey bee,

FIGURE 2 Postage stamps featuring a selection of moths. (Stamps from the author's collection.)

the emblem of the city of Ephesus. Their "Great Mother" was also known as the Queen Bee and her priestesses were called "melissae," from the Greek for honey bee, in analogous reference to worker bees and their servitude to the queen. There are currently 39 U.S. states that have designated an official state insect, chosen typically to represent something beautiful or inspirational from the state or merely an insect familiar to many. The honey bee is the insect of choice for 16 states and has been nominated, along with the monarch butterfly, as a candidate for the national insect of the United States.

Some insects, particularly those that symbolize aggression, have found their way onto the playing field in the form of sports team mascots. Teams often choose insects, such as wasps, that symbolize aggression (e.g., the Charlotte Hornets of the National Basketball Association and the Yellowjackets of Georgia Institute of Technology).

Other forms of insect celebration involve periodical events of recognition or appreciation for the actions or beauty of local insects. Cities and towns celebrate the beneficial industry and products of some insects, such as honey bees, or the pestiferous activities of those insects that affect the local economy. Annual festivals are held to celebrate honey bees in Illinois and Georgia, woollybear caterpillars in Kentucky and North Carolina, and monarch butterflies in California. In other places, celebrations recognizing the local impact of pestiferous insects are held, including a fire ant festival in Texas, a phylloxera festival in Spain, and a permanent tribute to a weevil in Alabama. The monument to the boll weevil in Enterprise, Alabama, is a large statue of a woman holding a larger than life weevil high over her head. It was dedicated in 1919 to honor the pest for the roll it played in the history of the town. The farmers were forced to switch from planting cotton to a diversity of other crops, particularly peanuts, and the town prospered as a result.

A very successful type of organized celebration of insects that has become common in recent years is the insect fair. These events serve to congregate people with a common interest in insects where they can participate in and enjoy a variety of insect-based fun, contests, food, and dialogue. Insect fairs also provide opportunities to see and purchase nearly anything of entomological interest.

ART, LANGUAGE, AND LITERATURE

Art draws its inspiration from the environment of the artist. It is therefore not surprising that insects have pervaded all forms and aspects of art. Images of insects are found as adornment on all types of objects from textiles and pottery to weapons and jewelry and even the tattoos on human bodies. Insects are also found, either as the primary subjects or merely as curious elements of lesser status, throughout all types of paintings.

Like the illustration of other animals, insect illustration began as a form of decoration. The earliest clearly identifiable drawing of an insect apparently dates to around 20,000 years ago. It is of a cave-dwelling rhaphidophorid cave cricket, inscribed on a piece of bison bone by Cro-Magnon people in southern France. From this humble beginning, depictions of insects have adorned everything from the walls of caves and temples of ancient societies to the paintings and textiles of modern artisans.

Insects, particularly butterflies, were used for decorative purposes in the painted illuminations of medieval manuscripts. By the 15th century, insects had become as important as birds in this respect. The rich iconographic use of insects at this time, often associated with folklore composed of a mixture of misinformation and factual representations, formed the basis on which the first scientific naturalists started their work in the 16th century. One insect painter, who was primarily an artist rather than a biologist, was Maria Sibylla Meriam. She reconciled the old aesthetic realism of medieval origin with the new tradition of practical engravings of the elaborately illustrated natural history treatises of the day and helped form the foundations of modern scientific investigations and writings on insect subjects.

In addition to paintings and textiles bearing artistic depictions of insects, their bodies, parts, and products often serve as the media for art. The metallic, brightly colored elytra of some buprestid beetles have been used as decorative cover on sculptures and textiles and as accessories in jewelry. Similarly, pieces of the colorful wings of butterflies are used in various parts of the world in collages to create artistic images. Beeswax was used to fashion figures and was the wax used to make the positive images in the "lost wax" technique for casting metal figures that originated in the third millennium B.C. in the Middle East.

One art form in which insects have been widely used as models is jewelry. Jewelry resembling insects has been used as aesthetic adornment around the world, throughout history and currently. The insects most commonly used as models for jewelry are beetles, flies, bees, butterflies, and dragonflies.

Some of these, such as flies and bees, had symbolic significance in ancient societies. Others, such as dragonflies and butterflies, are more likely used because of their beauty. A particularly interesting form of insect-based adornment is living jewelry. In Mexico, small jewels, glass beads, and metallic ornaments are set or glued to the elytra and pronotum of living ironclad beetles (Zopheridae) that are then attached to a fine chain pinned to the blouse and allowed to act as a living brooch. Some brilliantly metallic buprestids are used in a similar manner in parts of tropical Asia, and living fireflies and luminescent elaterids are used as decorations in hair or attached to clothing.

Many 15th and 16th century paintings include the motif of common-looking flies perched on various subjects, including people. The depiction of flies in this manner was done in mischievous jest or to invoke shock, perhaps to symbolize the worthiness of even the smallest objects of creation in association with the images of humans, as an expression of artistic privilege, or to indicate that the person in a portrait had died. Sometimes flies were included simply as imitation of such *musca depicta* done by previous painters.

In addition to their roles in mythology and folklore, insects and their symbolic representations have been adapted into the language and philosophy of various cultures. Symbols are used to suggest some idea or quality other than itself. One example is insect symbols in ideographic or phonetic symbols in written language. Examples are found in Assyro-Babylonian cuneiform and the ideographic writing of the Chinese and Japanese. The Greek word for mosquito, "Konops," is the source of the word for canopy, such as that made of mosquito netting. The medieval word "mead" refers to an alcoholic drink made from fermented honey and water that was used as an elixir. This word is the basis for the word "medicine" in recognition of its purported healing properties, and the word "madness" is in reference to the state of some people under the influence of mead. Insects even form the basis for geographical place names. Chapultepec, the hill of the grasshoppers, is where the castle of Aztec Emperor Montezuma stood in what is now part of Mexico City. Urubamba, which means the plain of the insect, is the sacred valley of the Incas near Cuzco in Peru. Japan was once known as Akitsushima, meaning dragonfly island.

Insects have also lent their names and attributes to a variety of descriptors of people and their personalities. People may be described as "busy as a bee," "nit picky," or "antsy." They may act "merry as a cricket" or feel as though they have "butterflies in their stomach." Connotations associated with particular insects may be used to convey similar traits in people. In many parts of the world, the reference of someone as a cockroach signifies an utter contempt for the individual and implies that their life is without value. The Spanish word for butterfly, "mariposa," is street slang for male homosexuals in Mexico. Lastly, insects enter language as metaphor. For example the self-ascribed desirable qualities of boxer Muhammed Ali are that he can "float like a butterfly and sting like a bee."

Aside from purely scientific works, insects have been represented in word and verse in a variety of contexts. In literature, insects are found as subjects of humor, as examples of aesthetically interesting natural subjects of wonder and appreciation, and as characters in fairy tales and in science fiction, mystery, and fantasy novels. Insects sometimes even serve as the storytellers themselves.

The essence of insects in literary humor typically involves the superimposition of insects into aspects of human behavior. The depiction of insects engaged in human activities is a common avenue of insect humor. This is particularly true of the role of insects in comic strips and cartoons, such as in the Far Side cartoons by Gary Larson. In other works, factual entomological information is cleverly presented in a humorous format. Such essays serve to popularize insects and their study, to educate, and, of most relevance here, to entertain.

Insects with endearing qualities, such as beautiful appearance or song, are used in fanciful stories and celebrated in poetry and verse. In Roald Dahl's *James and the Giant Peach,* a group of larger than life insects join a young boy as companions in a surreal adventure inside a monstrous fruit. A cricket and its song play a central role in Charles Dicken's fairy tale of home, *The Cricket on the Hearth.* The people of the house are gladdened and cherish the pleasant voice of the cricket as they listen to its fireside music. The melodious tune made for a happy home and served as an inspiration for those that heard it. Selections of insect poetry are typically written to convey particular feelings or to celebrate insects themselves. A contemporary example comes from the late D. K. McE. Kevan, the author of many humorous entomocentric verses, who wrote *An Embiopteran Epitaph* (reprinted from the *Bulletin of the Entomological Society of Canada* 6(1), 29, 1974).

We embiid web-spinners,
When seeking out our dinners,
Run back and forth in tunnels made of silk;
But, when we get the urge,
We occasion'ly emerge
From beneath a log, or places of that ilk.

We like our climates warm;
We're of dimorphic form;
We're soft and have a tendency to shrink.
"One does not often see 'em!"
Says the man in the museum,
But we're really not so rare as people think!

On the other hand, insects with undesirable qualities or strange traits are typically the subject of horror and mystery stories. Hundreds of science fiction and fantasy stories that use insects in a variety of prominent roles have been published. Franz Kafka's short story *The Metamorphosis* is about a young man who awakes one morning to find out he has turned into a giant insect. In Edgar Alan Poe's tale of the

hunt for a pirate's treasure, *The Gold Bug,* an insect is used to find the buried loot. As per the directions on a coded map, the gold beetle, tied to the end of a string and passed through the left eye of a skull nailed high in a tall tree, indicates the spot of a landmark from which the location of the treasure can be deciphered.

The role of insects in science fiction is particularly well established in film, where various insects appear as horrific creatures. Some of these insect fear films, e.g., *The Hellstrom Chronicle* (1971), merely embellish factual information in order to prey on the entomophobic tendencies of the general populace and the potential fleetingness of the future of humans on Earth in the face of the insect hordes. Others use fantastic representations of insects with supranormal characteristics, typically the result of science and technology gone awry, to instill fear and malevolence toward the insect characters, and as a lesson of what can happen when humans arrogantly fool with nature. Ants are common subjects in these roles and appear as giant mutants invading southern California in *Them* (1954) and a housing development in Florida in *Empire of the Ants* (1977). In *The Naked Jungle* (1954) and *Phase IV* (1974), the ant attackers are of normal size, but possess supernatural intelligence and aggression. Because they are widely despised by humans, cockroaches and flies are predisposed to be good villains in these films. In *Bug* (1975), hordes of carnivorous, self-combustible cockroaches wreak havoc on the population; and in the classic insect horror film *The Fly* (1958), the bodies of scientist and insect become inextricably combined with horrific consequences.

Not all fiction films starring insects depict them in a negative light. Insects sometimes fill the role of funny or entertaining characters. For example, in *Joe's Apartment* (1996), the singing and dancing cockroaches are crudely humorous roommates. The literary or cinematic use of insects in humor or as subjects of entertainment invariably leads to the creation of bugfolk. Bugfolk are humanized insects and other related arthropods that dress or talk like humans or are little people with wings, antennae, or other insect features. Bugfolk appear in nearly every literary and art form and are favorite characters for young audiences because of their teaching and entertainment abilities.

Certainly the most familiar bugfolk to Americans and many others worldwide is Jiminy Cricket, of Walt Disney's 1940 animated film *Pinocchio.* Like many of his kind, but unlike his true insect model, Jiminy Cricket bears only four limbs and acts and appears very human. Although morphological correctness is commonly practiced in more recently derived motion picture bugfolk, four-leggedness continues to be seen particularly when a friendly character relationship is desired. Six leggedness, e.g., the evil "Hopper" and his gang of grasshoppers in Disney's animated feature *A Bug's Life,* is used perhaps to provide a farther-from-human image and invoke disdain. Many other bugfolk are featured in comics, as children's toys, and as subjects in literature and art.

FIGURE 3 Trio of bugfolk extolling the virtues of sociality. [Illustration from *Episodes of Insect Life,* by Acheta Domestica, M.E.S. (1851).]

The use of bugfolk in literature and film enables people to see and learn something about themselves through these characters, in perhaps a different light than would be achieved through a strictly human relationship. For example, insect humor often involves a comparison of human behavior and what an insect might be supposed to do in comparable situations. In this manner, insects are found dressed as humans engaged in human activities, such as attending a festive party or dance, or as subjects in amusing or thought-provoking situations (Fig. 3). Particularly creative illustrations of anthropomorphized insects enjoying themselves are found in Grandeville's *Scènes de la vie privée et publique des animaux* and in Aldridge's *Butterfly Ball and the Grasshopper's Feast.*

MUSIC, ENTERTAINMENT, AND CEREMONY

The songs, sounds, and other qualities of insects have inspired many musicians and songwriters. The sounds produced by various insects serve as songs for direct enjoyment or as the inspiration for man-made music. Singing insects have a rich social history in Asia where celebrations and festivals are routinely held. People in both Japan and China have long kept singing insects, chiefly crickets and katydids, in small cages, like birds, for the enjoyment of their songs. The inspiring influence of insect sound for human musicians is exemplified in Nicolas Rimsky-Korsakov's famous musical composition the *Flight of the Bumblebee* and in that of another piece by Korsakov's pupil Anatol Liadov, the *Dance of the Mosquito*. As subject matter in song, insects such as cockroaches are common in blues and folk songs such as the famous Mexican folk song "La Cucaracha," about the troubles of a cockroach down on his luck.

Although insect collection and observation is generally done as an educational activity, many people find great enjoyment in capturing insects for specimens, to keep as pets, and to use in a variety of entertaining tasks. This is particularly true of children living in rural areas of Japan where insects have achieved a lofty cultural status.

These activities support an entire industry devoted to providing the equipment used to capture, observe, and keep insects in captivity. Some insects, particularly large dynastine scarabs and lucanids, are even mass reared and sold in vending machines.

Insects serve as the models for games or may be active, albeit unwilling, participants in a variety of six-legged sporting events. In the children's game "Cootie," the object is to be the first player to assemble a complete insect from a set of body parts such as antennae, proboscis, and six legs. In many parts of the world children fly insects instead of kites. Large insects, such as big beetles and dragonflies, are tethered to strings and allowed to fly for the amusement of people. In places where they occur naturally, large male dynastine scarabs or lucanids are collected and made to fight each other for sport.

Bouts and games involving insects are a source of enjoyment as well as an opportunity for gambling, such as with cricket fighting in China and Thailand and water bug roulette. In the latter contest, water beetles or water bugs are released into the center of a circular container filled with water. The inside perimeter is bounded by a continuous series of marked slots into which the insect can enter. Entrance of a particular insect into a slot is analogous to the landing of the ball on a particular number on a roulette wheel and the appropriate prize is awarded. In addition to being pitted against each other in battle, insects are commonly matched in foot races. For example, caterpillar races are held in Banner Elk, North Carolina, during the Woolly-Bear Festival, and cockroach racing is popular in many parts of the world, particularly in China and India.

Other forms of insect-based entertainment for humans include flea circuses and entertaining displays of both living and dead insects. Flea circuses use tiny performing fleas that are "trained" to perform a variety of circus acts for the amusement of the audience. Living insects are displayed in venues such as butterfly houses, where they can be viewed and enjoyed flying about their enclosures by an appreciative public. Dead insects have been similarly displayed as objects of aesthetic pleasure, sometimes with added adornments such as miniature clothing. Dead fleas are dressed in tiny costumes and displayed in folk art exhibits in Mexico. In Plano, Texas, the Cockroach Hall of Fame Museum features dead roaches dressed as famous people engaged in various activities.

As is true for other organisms that are held in high regard and for those that serve some utilitarian function such as food, some insects fill symbolic roles in human ceremonies. For example, although insects are regularly eaten in many parts of the world for sustenance, the consumption of insects was sometimes reserved for ceremonies or other special occasions (Fig. 4). The Kaua of Brazil perform a dance known as the "Dance of the Dung Beetles" that is used to drive away demons. The dancers attempt to transfer powers to themselves from the spirit world by taking on the image of the beetles. They do this by imitating the actions of beetles rolling a ball of dung. Other insects play a more active role and are used for a particular ceremonial or ritualistic purpose. Because of their powerful stings, giant hunting ants are used by indigenous peoples in Amazonia in male initiation and virility rites ceremonies. Large numbers of ants are tied to a woven mat and the mat with the now enraged ants is applied to the initiate's bare skin. Those who endure the excruciating pain without complaint, and live, are deemed worthy.

FIGURE 4 Grasshoppers being carried to a feast to celebrate the Assyrian defeat of the Elamites, from a relief of Ashurbanipal at Ninevah. [Illustration modified from Bodenheimer (1928).]

The Indians of central and southern California also made ceremonial use of ants. Male youths of the Kitanemuk, Tübatulabal, and Kawaiisu were taken by their elders for three days of fasting, after which they were given numerous live "red" or "yellow" ants to eat. The ants were consumed in order to gain power and induce a trance-like state during which spiritual insight would be gained.

HISTORICAL EVENTS MEDIATED BY INSECTS

Finally, insects have made their mark on human cultures by influencing events that shape history, such as wars, or by

changing the way societies can or cannot accomplish things. The Panama Canal was built and ultimately controlled by the United States in part because the earlier effort by France was thwarted by mosquito-borne yellow fever. As vectors of African sleeping sickness, *Glossina* spp. (Diptera) have made huge pieces of land in Africa uninhabitable by humans. Bubonic plague, spread by its flea vector, helped cause drastic changes in the social and economic structure in Europe during the 14th and 15th centuries. The populations, and thus the sites, of more than one ancient eastern Mediterranean city moved because of the actions of insects, particularly flies.

In many military campaigns, the number of casualties attributed to insects has exceeded that caused by actual fighting. The activities of insects, primarily by transmitting disease to troops in battle, have determined the outcome of entire wars. Napoleon's invading army lost hundreds of thousands of men and was decimated by the louse-borne disease typhus during their eastward march across Europe in 1812 and 1813.

Insects have also served as important determinants in the fates of human societies and economies throughout human history. The survival of the Israelites during their extended journey through the Sinai Desert was apparently made possible by insects. The manna that they gathered, ate, and survived upon was most likely the excretions of scale insects. If not for the arrival and help of divinely inspired seagulls, a plague of mormon crickets in 1848 may have ruined the crops and doomed the Mormons soon after their arrival in their new home in Utah. The silk trade was central to the economy of the Chinese Empire as was cochineal to the Aztecs of central Mexico. This is also true on a smaller scale for producers of honey and shellac, and for the thriving modern-day trade in insects sold for scientific, educational, and hobbyist uses.

The action of insects even helped to revolutionize the production of one product that has greatly shaped the whole of human civilization over the past 2000 years. Since the "invention" of paper was first proclaimed to the Chinese emperor Ho Ti in 105 A.D. by Ts'ai Lun, a variety of plant fibers were used in the production of paper. As writing flourished, supplies of raw materials for making paper became in short supply. Such was the situation in 16th century Europe, where paper was made from cotton and linen. It was here that observations of paper wasps inspired the French naturalist and physicist Rene Antoine Ferchault de Réaumer to suggest the use of wood as a papermaking fiber in 1719. These wasps, which chew wood and mix the fibers with saliva to make their nests, served as the inspiration for the use of the plentiful fiber on which modern papermaking is based.

CONCLUSION

The sources of published information dealing with the roles of insects in human culture have until recently been found in a diffuse body of literature. Such information is often hidden in historical documents, anthropological works, and ethnoentomological notations in travel logs and journals. In the past 20 years, a wider aesthetic and cultural appreciation for insects has been realized. The celebration of insects and their attributes as they relate to the development of human societies is generally accepted as a worthwhile endeavor. This interest has spawned a number of review articles and books that summarize, synthesize, and sometimes popularize much of the previously diffuse literature and serve as a starting point for those interested in this fascinating subject. Some periodicals, namely, *The American Entomologist,* regularly publish cultural entomological articles, and the recently introduced periodical, *Cultural Entomology Digest* (http://www.bugbios.com/ced/), is devoted entirely to this topic.

Along with the modernization of the world, the perceived relevance of insects to human life is slowly eroded. As this happens, the various roles of insects in human cultural affairs may change or be lost. However, many people continue to carry mythological modes of thought, expression, and communication into this supposedly scientific age and others still find pleasure in observing and contemplating their six-legged companions on Earth. Therefore, the importance of insects as subjects of entertainment and aesthetic pleasure should continue to enter into the thoughts of future people and mold aspects of human culture. As some relationships between human and insect are lost, others are formed. Because of the dominant place in the function of the world's ecosystems and their influence on human existence, insects have played and will continue to play a prominent role in our perception of life and pursuit of aesthetically pleasing activities and for the enlightenment of human societies.

See Also the Following Articles

Entomological Societies • Folk Beliefs and Superstitions • History of Entomology • Movies, Insects in • Stamps, Insects and

Further Reading

Akre, R. D., Hansen, L. D., and Zack, R. S. (1991). Insect jewelry. *Am. Entomol.* **37,** 90–95.

Berenbaum, M. R. (1995). "Bugs in the System." Addison Wesley, Reading, MA.

Berenbaum, M. R. (2000). "Buzzwords." Henry Press, Washington, DC.

Bodenheimer, F. S. (1928). "Materialien zur Geschichteder Entomologie bis Linné," Vol. I. Junk, Berlin.

Cherry, R. H. (1993). Insects in the mythology of native Americans. *Am. Entomol.* **39,** 16–21.

Clausen, L. W. (1954). "Insect Fact and Folklore." MacMillan Co., New York.

Cloudsley-Thompson, J. L. (1976). "Insects and History." St. Martin's Press, New York.

Dicke, M. (2000). Insects in western art. *Am. Entomol.* **46,** 228–236.

Gagliardi, R. A. (1976). "The Butterfly and Moth as Symbols in Western Art." Southern Connecticut State College, New Haven. [Masters thesis]

Hamel, D. R. (1991). "Atlas of Insects on Stamps of the World." Tico Press, Falls Church, VA.

Hogue, C. L. (1980). Commentaries in cultural entomology. 1. Definition of cultural entomology. *Entomol. News* **91,** 33–36.

Hogue, C. L. (1985). Amazonian insect myths. *Terra* **23,** 10–15.

Hogue, C. L. (1987). Cultural entomology. *Annu. Rev. Entomol.* **32,** 181–199.
Kritzky, G., and Cherry, R. (2000). "Insect Mythology." Writers Club Press, San Jose, CA.
Laurent, E. L. (2000). Children, 'insects' and play in Japan. *In* "Companion Animals and Us" (A. L. Podberscek, E. S. Paul, and J. A. Serpell, eds.), pp. 61–89. Cambridge University Press, Cambridge, U.K.
Pearson, G. A. (1996). Insect tattoos on humans: A "demographic" study. *Am. Entomol.* **42,** 99–105.
Stickney, D. (1997). "Water Bugs and Dragonflies." Pilgrim Press, Cleveland.
Tedlock, D. (1985). "Popul Vuh." Simon & Schuster, New York.

Cuticle

Svend O. Andersen
Copenhagen University

The cuticle is an extracellular layer that covers the complete external surface of insects, as well as the surfaces of their foreguts and hindguts, and acts both as a skeleton for muscle attachment and as a protective barrier between the animal and its environment. The cuticle is an integral part of a complex dynamic tissue, the integument, which also includes the cuticle-producing epidermal cells, and various glands and sense organs.

GENERAL PROPERTIES OF CUTICLES

The cuticular layer varies in thickness from a few micrometers to a few millimeters, depending upon the insect species, developmental stage, and body region, but cuticles typically are between 100 and 300 μm thick. Cuticles are highly diverse in their mechanical properties. They can be divided into two groups: stiff and hard cuticles, and soft and pliant cuticles. Intermediate degrees of stiffness also exist, and some types of cuticle have special properties, such as rubberlike elasticity or extreme extensibility. Cuticles differ in color and in surface sculpturing, but electron microscopy shows that all types of cuticle are built according to a common plan. The details in structure and properties of the various cuticular regions are such that for each species they are optimal for the functioning of the living insect in its natural surroundings.

EPICUTICLE

The outermost layer of a cuticle is called epicuticle; it forms a continuous layer covering the complete cuticular surface. Seldom more than 2 μm thick, it is responsible for the waterproofing properties of the cuticle. Electron microscopy shows that the epicuticle can be subdivided into several layers, of which the inner epicuticle, also called the dense layer, is the thickest. It is covered by the thin, outer epicuticle, sometimes called the cuticulin layer, which is assumed to be responsible for the mechanical stiffness of the epicuticle. The inner and outer epicuticle are composed of polymerized lipids and protein, and they contain no chitin. These two layers remain poorly characterized because they are difficult to purify, dissolve, and degrade.

The outer epicuticle is covered by a waterproofing wax layer, containing complex mixtures of extractable lipids, secreted during the molting process from integumental oenocytes and epidermal cells. This layer is again covered by a protective cement layer, secreted immediately after ecdysis from glands in the integument.

The extractable lipids in the wax layer have been characterized for several insect species. They appear to be species-specific mixtures of a wide range of lipids, including normal and branched, saturated and unsaturated hydrocarbons, fatty acids, alcohols, esters, sterols, and aldehydes. Differences in lipid composition have been used to discern closely related insect species. The epicuticular lipid composition also can vary between instars and sex of the same species, and these lipids often play an essential role in recognition and communication between insects.

PROCUTICLE

The region of the cuticle, located between epicuticle and the epidermal cell layer, is called procuticle; it constitutes the main part of the total cuticle. Histologically, the sclerotized regions (sclerites) are often subdivided into layers with different staining properties: (1) the outermost layer, the exocuticle, may be dark colored because of sclerotization, but is refractory to staining; (2) the innermost, uncolored layer, the endocuticle, stains blue; and (3) in between one often observes a layer of mesocuticle, staining red with Mallory triple stain. The flexible cuticle (arthrodial membranes), which connects the sclerites, stains blue with Mallory throughout most of its thickness. Exocuticle may correspond to the part of the procuticle deposited before ecdysis, stabilized by sclerotization. Mesocuticle plus endocuticle often correspond to the post-ecdysially deposited procuticle, and if these layers are sclerotized at all, it is only slightly.

The procuticle consists mainly of chitin and proteins; water is an essential component, and other materials, such as lipids, phenolic compunds, salts, pigments, and uric acid may be present. Chitin (poly 1,4-β-*N*-acetylglucosamine) is a polysaccharide, present as long and nearly straight microfibrils, usually about 2.8 nm in diameter and of indeterminate length. The filaments tend to run parallel to the cuticular surface, but columns of chitin filaments running perpendicular to the surface have been described for some types of cuticle (lepidopteran larval cuticle). The function of such chitinous columns remains uncertain.

The chitin microfibrils are organized in various patterns, and the organization seems to have importance for the mechanical properties of the cuticle. The most commonly observed patterns are the heliocoidal pattern, where the

microfibril direction changes by a small, constant angle between neighboring layers; the preferred, unidirectional orientation, where the fibrils run in the same direction in all layers, and the pseudo-orthogonal orientation, where unidirectional layers of chitin microfibrils alternate with layers running at nearly right angles to each other. In certain cuticles the pattern of chitin microfibrils depends on a daily rhythm: in locust tibiae, heliocoidal cuticle is deposited during the night and unidirectional cuticle is deposited during the day, making it possible to determine the number of days since ecdysis.

The chitin microfibrils are embedded in a protein matrix; the protein content tends to equal the chitin content in flexible cuticles and is usually three to four times higher than the chitin content in hard cuticles. The number of different proteins present in a given type of cuticle can vary from about 10 to 100. Different types of protein are present in flexible and hard cuticles; the proteins are species specific, and some of them are also specific for certain cuticular regions. A characteristic amino acid sequence region, common to a large number of cuticular proteins, is supposed to have a function in the linking of proteins to the chitin microfibrils. The proteins are often extractable immediately after deposition. In many cuticular regions, however, they are later rendered inextractable by sclerotization, whereby low molecular weight phenolic compounds are covalently incorporated into the cuticular matrix, cross-linking the proteins, and making the cuticle harder and stiffer, and more difficult to digest with enzymes.

Sclerotization may occur soon after a molt when the insect has expanded its new cuticle to a larger size, but the regions that are not enlarged may have been sclerotized in the pharate stage, which is the stage that is present before emergence from the exuvium, or old cuticle. The elastic protein, resilin, present in rubberlike cuticular regions, is cross-linked as soon as it is deposited extracellularly. The cross-linking process is different from that in sclerotized cuticle because no low molecular weight compounds are involved, but tyrosine residues in the protein chains are oxidatively coupled to each other, forming di- and trityrosine residues.

SUBCUTICLE

A narrow, histochemically distinct layer, called subcuticle, is situated between the procuticle and the epidermal cells. It stains positively for muco- and glycoproteins. It has been suggested that it serves to bind cuticle and epidermis together and that this layer is the deposition zone, where new cuticular material is assembled and added to the already existing cuticle.

See Also the Following Articles

Chitin • Exoskeleton • Integument • Molting

Further Reading

Andersen, S. O. (1985). Sclerotization and tanning of the cuticle. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut, and L. I. Gilbert, eds.), Vol. 3, Chap. 2. Pergamon Press, Oxford, U.K.

Blomquist, G. J., and Dillwith, J. W. (1985). Cuticular lipids. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut, and L. I. Gilbert, eds.), Vol. 3, Chap. 4. Pergamon Press, Oxford, U.K.

Hepburn, H. R. (ed.). (1976). "The Insect Integument." Elsevier, Amsterdam.

Hepburn, H. R. (1985). Structure of the integument. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut, and L. I. Gilbert, eds.), Vol. 3, Chap. 1. Pergamon Press, Oxford, U.K.

Kramer, K. J., Dziadik-Turner, C., and Koga, D. (1985). Chitin metabolism in insects. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut, and L. I. Gilbert, eds.), Vol. 3, Chap. 3. Pergamon Press, Oxford, U.K.

Neville, A. C. (1975). "Biology of the Arthropod Cuticle." Springer-Verlag, Berlin.

Wigglesworth, V. B. (1972). "The Principles of Insect Physiology." 7th ed. Chapman & Hall, London.

D

Daddy-Long-Legs
(Opiliones)

Gonzalo Giribet
Harvard University

The Opiliones, commonly known as daddy-long-legs, harvestmen, shepherd spiders, or harvest spiders (among many other names), are a very interesting group of arachnids that are well known by farmerss. Opiliones constitute the third most speciose arachnid order (after Acari and Araneae), comprising approximately 1500 genera and 5000 species in 45 families. They are the only nonmite or tick arachnids that ingest vegetable matter, but generally they prey on insects, other arachnids, snails, and worms and have the ability to ingest particulate food; this is unlike most arachnids, which ingest only liquefied substances.

Opiliones are divided into four suborders: Cyphophthalmi, Eupnoi, Dyspnoi, and Laniatores.

DESCRIPTION

Daddy-long-legs present all the typical characteristics of arachnids, with the body divided into two regions, cephalothorax and abdomen, although these two regions are not clearly differentiated, giving daddy-long-legs the aspect of "waistless spiders." The cephalothorax generally has a pair of median eyes on top of an ocular tubercle. The eyes are simple, i.e., not compound as in insects and crustaceans. The cyphophthalmids lack eyes entirely or have a pair of lateral eyes. The cephalothorax also bears a pair of chemical-secreting organs, known as repugnatorial glands.

The cephalothorax has one pair of chelicerae for manipulating the food particles, one pair of palps of either tactile or prehensile function, and four pairs of walking legs, enormously long in some Eupnoi and Laniatores species, surpassing 15 cm in some species. The palps of most Laniatores are relatively large and have two rows of spines acting as a grasping organ. The second pair of walking legs is sometimes modified and acquires a tactile function.

Another distinctive characteristic of the Opiliones is that the females have a long ovipositor with sensory organs on the tip that are used to check the soil quality where they will lay the eggs. Except for some mites, similar organs are not known for any other arachnids. The males have a penis or copulatory organ, which may be muscular or alternatively operated by hydraulic pressure. Copulatory organs are also unique among the arachnids, again with the exception of certain mites. Fertilization is thus internal and direct (unlike in spiders, which use the palps for the indirect internal fertilization).

Figure 1 is an example of a typical daddy-long-leg, *Odiellus troguloides,* from the western Mediterranean.

FIGURE 1 *O. troguloides,* one of the most typical daddy-long-legs from the western Mediterranean region, with an elongated body reaching almost 1 cm in length. This species has considerable sexual dimorphism, females being much larger and more globose than males. Juveniles of this species hatch in the spring, quickly reach maturity, and die in the fall.

LIFE HISTORY AND BIOLOGY

Opiliones are oviparous and deposit between one (in cyphophthalmids) and several hundred (in phalangiids) eggs. Life cycles and longevity are variable. Many species live 1 year, with embryonic development occurring during the winter, with hatching in the spring, and reaching maturity in the fall, after five to seven molting periods. This is the typical seasonal life history of most Northern Hemisphere phalangiids. Others have an overlap of adults and juveniles throughout their life cycles during the favorable seasons, dying in the winter. Finally, cyphophthalmids and most laniatorids live several years, with cases recorded up to 5 years.

Sexual dimorphism is evident in some species. All cyphophthalmid males have a spur on the tarsus of the fourth walking leg. This structure, named an adenostyle, possibly secretes a pheromone. The families Pettalidae and Sironidae in the Cyphophthalmi have male anal glands, and the pettalids may have extreme modifications of the male anal regions.

Opiliones are generally small to medium in size (body measuring less than 1 mm to almost 2.5 cm in the European species *Trogulus torosus*), inhabit all types of moist to wet habitats, and occur on all the continents. The Laniatores include the large (up to more than 2 cm) and the most colorful Opiliones, and their distribution reaches a peak of diversity in tropical regions and in the Southern Hemisphere. The Eupnoi and Dyspnoi are more widely distributed, but especially abundant in the Northern Hemisphere. Finally, the Cyphophthalmi are distributed more uniformly worldwide, but are the smallest (down to 1 mm) and most obscure of the Opiliones.

No Opiliones are harmful to humans, and they do not contain any type of venom or other substance. Some Opiliones are reported as highly poisonous although not having the capacity of biting humans. This myth seems to be a confusion with the highly neurotoxic venom of some spiders. These are differentiated from Opiliones by the presence of a waist that separates the prosoma from the opisthosoma, among many other characters. In fact, Opiliones are supposed to be beneficial, and they are good indicators of undisturbed environments.

See Also the Following Articles

Arthropoda and Related Groups • Spiders

Further Reading

Edgar, A. L. (1990). Opiliones (Phalangida). *In* "Soil Biology Guide" (D. L. Dindal, ed.), pp. 529–581. Wiley, New York.

Giribet, G., Edgecombe, G. D., Wheeler, W. C., and Babbitt, C. (2002). Phylogeny and systematic position of Opiliones: A combined analysis of chelicerate relationships using morphological and molecular data. *Cladistics* **18,** 5–70.

Hillyard, P. D., and Sankey, J. H. P. (1989). "Harvestmen." Brill, Leiden.

Shear, W. A. (1982). Opiliones. *In* "Synopsis and Classification of Living Organisms" (S. P. Parker, ed.), pp. 104–110. McGraw–Hill, New York.

Damselfly

see *Odonata*

Dance Language

P. Kirk Visscher
University of California, Riverside

The "dance language" of honey bees refers to patterned, repetitive movements performed by bees that serve to communicate to their nestmates the location of food sources or nest sites.

RECRUITMENT

If a saucer of honey is placed outdoors, many hours or days may go by before a bee finds it and feeds on it. Soon after this first visit, however, large numbers of bees will arrive. Interest in honey bees goes back to prehistory, because their colonies provided human ancestors' most concentrated source of sugar. At least as far back as Aristotle, people have inferred that the bees that first discover a food source must recruit their nestmates to share in the collection of the food, thus accounting for the rapid buildup once a discovery has been made. The same kind of buildup occurs at flowers, bees' natural source of their sugary food.

Recruitment to food is one of the most important adaptations of nearly all social insects, and there are many forms of recruitment among them. Being able to recruit nestmates to food sources allows colonies of insects to realize one of the advantages of living in groups: the ability to harvest food that would not be as readily available to an individual foraging alone. Such edible items might include prey bigger than an individual could subdue, food resources that are rich but so widely scattered that an individual would not be likely to discover their source and sources that are ephemeral and thus more effectively harvested by means of group foraging during the short time the source is available. Cooperative foraging also is important to social animals in overcoming one of the inherent disadvantages of group living: since members of groups generally will compete with other members of the group for local food resources, without some compensating foraging advantage, solitary individuals would have better access to food than those in groups.

SIGNIFICANCE OF THE DANCE LANGUAGE

The best known of the mechanisms of recruitment in social insects is the honey bee (*Apis* spp.) dance language, in terms

of both its fame outside the realm of specialists and the depth in which it has been studied. The dance language is famous for a number of reasons. It is frequently cited as the premiere example of symbolic communication among nonhuman animals, and it is one of the first and best examples of such communication aside from human language. The discovery that mere insects could perform such a complex behavior led to a reassessment of the behavioral complexity possible among these animals with relatively small nervous systems, which had formerly been regarded as simple automatons governed by instinct and reflex. Finally, the dance language has provided a tool for studying the perceptual world and behavioral response of bees that has illuminated our understanding of their vision, olfaction, memory, orientation, learning, and social organization, and has provided a model for understanding these areas about insects in general.

DISCOVERY OF THE DANCE LANGUAGE

Observers of bees had repeatedly noted that sometimes a bee in a colony will perform repeated circular movements, closely followed by other bees, but it was Karl von Frisch who firmly established the connection between these movements and recruitment, and, in the course of a long career, discovered many aspects of communication by the dance language.

Von Frisch began his studies of the dance language in 1919, with the simple yet powerful approach of marking bees with paint as they fed at a flower he had enriched with a drop of sugar syrup (and in later experiments with a simple scented syrup feeder). He then watched their behavior when they returned to a glass-walled observation beehive. He observed his marked bees doing circular "round dances," which were followed attentively by other bees in the hive. He then observed that bees, presumably those that had followed the dances, would investigate nearby flowers of the same type as those at which the marked bee had fed but did not investigate flowers of other types as much. Von Frisch inferred that the dance stimulated recruits to look for food, and that odor in the nectar, and on the body of the dancing bee, communicated to the recruits the scent to seek. He also described a "waggle" form of the dance in which a dancing bee rapidly waggles her abdomen laterally while moving in a particular direction on the comb, then turns back more or less to the starting point, repeats the waggle on the same course, turns back the other way, and so on, describing a squat figure-eight with the waggle in the middle. The artificially small scale of his early work, in a small, walled, Munich garden, caused von Frisch to mistakenly conclude that the two kinds of dance he saw indicated different types of food. The waggle dancers often had pollen on their legs, whereas the bees he provided with nectar did not, so he concluded that waggle dances indicated pollen and the round dances nectar.

This error persisted for 25 years, but von Frisch himself discovered the full story when, during World War II, he was forced to take his studies away from the war-torn city to rural Brunnwinkel, Austria. There, in 1944 and 1945, working under conditions that more accurately reflected the natural scale of bees' foraging, he found that when bees fed at long distances from the hive they performed the waggle dances for nectar, as well. At the same time, he also made the startling discovery that the bees were communicating the direction and distance to the food source, as well as its odor.

COMMUNICATION OF DISTANCE AND DIRECTION IN THE DANCE

The waggle dance of honey bees can be thought of as a miniaturized reenactment of the flight from the hive to the food source (Fig. 1). As the flight distance to the food becomes longer, the duration of the waggle portion of the dance also becomes longer. The angle that a bee flies during the flight to the food, relative to the sun azimuth (the horizontal component of the direction toward the sun), is mirrored in the angle on the comb at which the waggle portion of the dance is performed. If the food is to be found directly toward the sun, a bee will dance straight upward. If the food is directly away from the sun, the bee will dance straight downward. If food is at 35° to the right of the sun, then the dance is performed with the waggle run at 35° to the right of vertical, and so forth. Bees make a transition from round dances for food sources near the nest to waggle dances at greater distance, with the transitional distance varying somewhat between different subspecies of *A. mellifera.*

While the bee is waggling her abdomen, she also produces bursts of buzzing sound from her wings, which are perceived by dance-following bees with the Johnston's organ at the base of the antennae. Recent work by Wolfgang Kirchner has shown that even the round dance contains directional

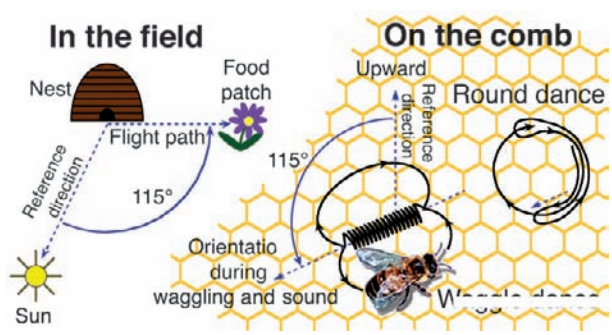

FIGURE 1 How direction to the food patch is encoded in the honey bee dance language. As a bee flies to flowers in the field (left), she learns the direction to the food patch relative to a reference direction of the sun azimuth (here the food is 115° to the left of the sun). When she dances on the vertical combs of the dark hive (right), she uses the direction upward as a reference and performs the waggling portion of the dance at the same angle, relative to this upward reference, to indicate that the food is to be found relative to the sun direction reference in the field (here, 115° to the left of upward). Dancing bees produce buzzing sounds during the waggle portion of the dance. In the round dance (far right), the dancing bee changes direction more randomly and does not waggle, but does buzz when moving in the direction that would indicate the direction to the food.

information, because these sounds are produced at the time in the round dance at which the circling bee is facing in the direction on the comb in which waggle runs would be performed for more distant food sources in the same direction. However, recruit bees seem to search the vicinity of the nest equally in all directions in response to round dances. This scatter in search area, however, is not really greater than the area searched at greater distances, though because of its proximity to the nest it includes all angles, whereas more distant searches are mostly within a restricted range of angles. Thus it is uncertain whether recruit bees can perceive the direction information in round dances.

MEASUREMENT OF DISTANCE AND DIRECTION

The ability to bees to communicate distance and direction to a food source requires that the recruiting bee and the recruits be able to measure these parameters. The study of how bees do this provides an example of how the dance language gives a readout of the perception of the bees. This in turn makes possible sophisticated analyses of the mechanisms by which bees acquire the information, analyses that are vastly more difficult to perform with insects that do not report their findings in a format entomologists have learned to decode.

Von Frisch found that wind, height differences between the feeder and hive, or adding additional weights or airfoils to bees changed the tempo of their dances. This finding indicated that something about these conditions had changed the bees' perception of distance to the food source. One aspect that was changed was the time of flight to the source, but the changes in dance tempo did not correlate well with the changes in flight time, and so this was rejected as the way the bees measured distance. Instead, it was concluded that the bees were measuring energy use, because all these conditions would affect energy use. This was consistent with observations that, on the flight to the food source, either a headwind or flying uphill would increase perceived distance, whereas either a tailwind or flying downhill would decrease it.

However, more recent work by Harald Esch and others suggests that it is not energy that is measured, but the movement of landscape objects across the visual field, or optic flow. Humans experience the apparent motion of landmarks as faster when riding in a car than when flying in an airplane. Similarly, when a bee flies close to the ground, she experiences rapid optic flow, whereas at greater altitudes the optic flow is less. In von Frisch's experiments, the changing conditions also affected the height off the ground of the bees' flight, so that energy use and optic flow were confounded. In experiments in which bees are trained to feeders at different distances from the ground, the distance that a bee perceives, as indicated by the tempo of her dances, is shorter for higher feeders, even though more energy is needed to fly to them and the length of the flight path is greater. The progress of entomologists' understanding of the mechanism by which bees measure distance provides an excellent example of how the conclusions from an experiment may reject incorrect hypotheses, but may also accept incorrect ones, if the predictions of the latter are the same as another alternative hypothesis not considered in the design of the experiment.

Martin Lindauer described the way in which bees measure the angle of their body with respect to gravity, using groups of sensory hairs in the joints between head and thorax and thorax and abdomen. When Lindauer severed the nerves to these hairs, bees were no longer able to do oriented dances on a vertical comb. When flying in the field, bees use their compound eyes to measure their angle of flight relative to the sun, searching out the patterns of polarized light in the blue sky itself, even if the sun is not visible. The polarized light is produced by a phenomenon called Rayleigh scattering; the angles of polarization occur in a pattern that is consistent relative to the position of the sun, and this pattern moves across the sky as the earth moves relative to the sun. Rüdiger Wehner and S. Rössel discovered that the bees use a "celestial compass" to interpret the polarization patterns, which consists of the layout of ommatidia in the dorsal portion of the bees' compound eyes. Each ommatidium is selectively sensitive to a particular angle of polarization of light, and each ommatidium also gathers light exclusively from a particular region of the visual field of the bee. The layout of the ommatidia is such that when a bee is facing directly away from the sun, each ommatidium is looking at the region of the sky that contains the angle of polarized light to which it is most sensitive. Thus, as the bee rotates in flight, the summed response from these specialized ommatidia will reach a peak when the bee is aligned with the sun azimuth and fall away as she turns off it. Although the way in which a bee uses this system to hold a fixed course at a particular angle relative to the sun is not known yet, this compass provides a beautiful example of how a solution to a tremendously complex analysis can be built into the design of the sensory system, so that only relatively simple neural processing is needed to execute the behavior.

USE OF THE DANCE LANGUAGE

Honey bees are known to use the dance language to recruit nestmates in several contexts. In the context of foraging, bees dance to indicate the location of sources of nectar, pollen, water, and propolis (a resinous material collected from plants and used to seal cracks and waterproof the nest cavity). As far as is known, the dances for these different materials are the same, but this area has not been systematically investigated.

When a swarm of bees leaves its natal colony to build a new nest elsewhere, scout bees report the location of cavities they have found by means of the dance, and other bees inspect the advertised sites and may dance in turn. Over the course of hours or days the swarm as a whole makes a choice among the alternative sites discovered by different scouts and arrives at a unanimous decision on a single site. The swarm then takes off and flies to the new nest site. Only a small minority of the bees in the swarm has ever visited the chosen

cavity. Therefore, although the information transferred by the dance could be important in guiding other bees to the site, there are probably other mechanisms, perhaps visual or olfactory, involved as well. The question of how swarms find their way, and the question of just how the dance language is used in the course of the swarm coming to a collective decision on a single nestsite, are still being investigated.

The sharing of information about food sources makes it possible for a honey bee colony to serve as an information center, pooling the reconnaissance of its many foragers, surveying a vast area around the nest, and focusing the bulk of its foraging force on the best sources discovered. In the 1980s a study by Kirk Visscher and Tom Seeley decoded the dances of a colony living in a deciduous forest in New York State to show the dynamics of colony food patch use that result from these interactions. Research by Seeley has shown that integration of foraging information via the dance language is quite flexible, and Seeley has worked out many of the mechanisms by which a honey bee colony responds rapidly to changes in the relative quality of food sources and colony need for food.

THE DANCE LANGUAGE CONTROVERSY

In the late 1960s Adrian Wenner, Patrick Wells, and Dennis Johnson challenged von Frisch's interpretation of the bee dances. While they did not question that the dances contain correlations of distance and direction, they pointed out that many experiments claimed by von Frisch to show that bees actually used this vector information in their searches could also be interpreted as the bees simply orienting with respect to odors. These ambiguous results were recorded when the recruiters' feeder was placed in the center of an array of scented bait stations and recruits were observed to come more frequently to stations near the center. This behavior, von Frisch's critics argued, would be predicted regardless of whether bees were using distance and direction (and odor) information or just odor information. Johnson and Wenner performed experiments at relatively short distances and with strong odors, and the results followed the expectations of recruits relying strongly on odor produced by bees feeding at the bait stations, but not the expectations of the location information in the dance.

Not all of von Frisch's experimental results were readily reinterpreted in terms of the odor-only hypothesis. For example, when a hive is turned on its side, bees are unable to use gravity as a reference for their dances and so do disoriented dances, and von Frisch showed that recruits were less well oriented under these conditions, although odor cues would not have been affected. Several lines of subsequent work have indicated that the search distribution of recruits can indeed be influenced by distance and direction information from the dance alone. The challenge in such studies is that normally odor information and dance vector information is highly correlated, so definitive experiments required means of unlinking them.

In the 1970s James Gould unlinked the location (and odor) of the food source on which dancers had foraged from the directional information in their dances. To achieve this, he shined a bright light from the side as bees danced. In this situation, recruiters or recruits will normally perform or interpret dances using the position of the light as the "sun" angle reference, rather than the direction upward. However, if a bee's ocelli are painted over with opaque paint, the bee becomes less sensitive to light, and so this shift in reference does not occur. By having recruiters with painted ocelli (and a reference of up) dancing, followed by recruits with unpainted ocelli (and reading the dances relative to a reference of the light, at some other angle), Gould was able to show that recruits could interpret a direction from the dance that was independent of the direction to the food source. The recruits then searched principally in the direction predicted by the modified dance information, rather than the true direction of the feeder, as would have been predicted by the odor-only hypothesis.

In the early 1990s Axel Michelsen, Martin Lindauer, and Wolfgang Kirchner constructed a computer-controlled robot bee that mimicked the behavior of a dancing bee. Recruits followed this robot bee and searched for food preferentially in the directions indicated by the dance angles programmed for the robot. Changes in the length of the robot bee's dances also changed the distribution of distances at which recruits were captured. The robot bee recruited rather imprecisely, with even more scatter than the rather large scatter of recruits from real bee dances. However, the demonstration that changing nothing but the computer programming was enough to cause significant shifts in the search distribution of recruits in the predicted manner was conclusive evidence that recruits were decoding distance and direction information from the dances.

CURRENT QUESTIONS

Although it is now quite clear that bees do decode the dances, odor does play a strong role in recruitment to food sources. It is appropriate to think of the dance as giving recruits a general idea of the direction and distance to the food source. Recruits then search in this area for sources matching the odors they have learned from the food carried by the dancing bee. Depending on the distribution of available food sources, the distance and direction information might be crucial in organizing a colony's food collection, or relatively unimportant. However, the relative importance of these two mechanisms in different habitats is just beginning to be investigated.

The angular scatter in the dance itself decreases with increasing distance indicated, as von Frisch reported. This change in scatter may be the result of changing duration of the waggle runs of the dance, but it also may be an adaptation to recruit bees to patches of more or less constant size at varying distances. This idea is supported by Seeley and Burmann's finding that the dances of scouts for nest sites,

which are always single points rather than patches, have less scatter than those of nectar foragers. However, these results differ from those reported by Will Towne on the same issue, and this remains a question of current research.

The evolutionary origin of the honey bee dance remains incompletely discerned. All species of *Apis* perform recruitment dances, though there are interspecific variations in a number of the aspects discussed earlier. The stingless bees (Meliponini), the bumble bees (Bombini), and the orchid bees (Euglossini) are the closest relatives of *Apis,* but the phylogeny of these different taxa within the Apidae remains controversial. Stingless bees are highly social and have a variety of mechanisms of recruitment that may provide possible antecedents to the dance language, but a determinination of how the current form of the dance language might have arisen from these components must await both a greater understanding of recruitment mechanisms within the stingless bees and a more firmly established phylogeny within the family Apidae.

See Also the Following Articles

Apis Species • Feeding Behavior • Orientation • Recruitment Communication

Further Reading

Dyer, F. C. (2002). The biology of the dance language. *Annu Rev. Entomol.* **47,** 917–949.

Frisch, K. von. (1967). "The Dance Language and Orientation of Bees." Harvard University Press, Cambridge, MA.

Moffett, M. W. (1990). Dance of the electronic bee. *Natl. Geogr.* **177(1),** 135–140.

Seeley, T. D. (1995). "The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies." Harvard University Press, Cambridge, MA.

DDT

Fumio Matsumura
University of California, Davis

DDT (dichloro-diphenyl-trichloroethane) is an old insecticide that has been banned from use in most countries of the world since the 1970s. However, DDT, its metabolites, and some of its derivatives, which are mostly produced as impurities in technical insecticide preparations, still contaminate the environment. DDT residues continue to cause deleterious biological effects, most notably, environmental endocrine disruptions.

From the viewpoint of environmental toxicology and chemistry, DDT is by far the best-studied chemical. Many models of bioaccumulation, atmospheric transport, transfer mechanisms within soil compartments, and from soil to air, and soil to water are based on data generated from studies of DDT residues in the environment.

CHEMICAL CHARACTERISTICS

DDT is one of several typical chlorinated hydrocarbon insecticides discovered in the early 1940s and known for their persistent insecticidal activities, their lipophilic attributes, and their stable chemical properties. The insecticidal properties of DDT itself were discovered in 1939 by Paul H. Müller of Switzerland, who later received the Nobel Prize for his work. Since DDT was the first organic synthetic insecticide that possessed advantages such as low mammalian toxicity, wide spectrum, long-lasting properties, and low cost in comparison to arsenicals and other inorganic insecticides, most entomologists embraced its use to such an extent that more than 100 million pounds of DDT was being produced annually by the mid-1950s.

The insecticidal active ingredient of DDT preparations is *p,p′*-DDT (Fig. 1A). Its 1-dechlorination product, *p,p′*-DDD (Fig. 1B) retains reasonable levels of toxicity for some insects, but its dehydrochlorination product, *p,p′*-DDE (Fig. 1C), shows no insecticidal property, although *p,p′*-DDE could still have a toxic effect in other organisms. Other components often found in insecticidal DDT preparations are *o,p′*-DDT, *p,p′*-DDD, and *o,p′*-DDD. All these can be found as environmental residues.

DDE (dichloro-diphenyl-ethylene) one of the residues derived from DDT most frequently found in the environment, is produced mainly by metabolic activities in biological systems and is particularly prevalent in insects and in some mammalian species. Although both *p,p′*-DDE and *o,p*-DDE are found in the environment, the former is more abundant and more frequently encountered. In assessing residue levels of all DDT-derived compounds today, scientists express the entire spectrum of DDT-related (DDT-R) compounds or DDT-derived compounds as total DDT residue, or DDTs.

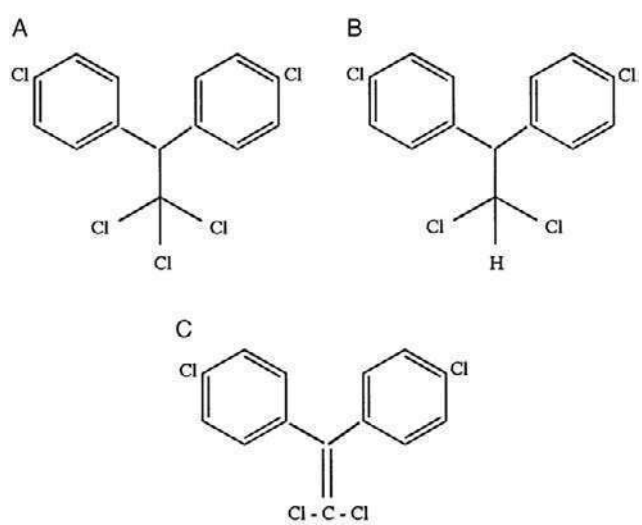

FIGURE 1 (A) 1,1,1-Trichlor-2,2-bis (*p*-chlorophenyl) ethane (*p, p′*-DDT). (B) *p, p′*-DDD. (C) *p, p′*-DDE.

EFFECTS ON INSECTS

The main action mechanism by which *p,p′*-DDT causes the death of insects is the destabilization of the sodium channel, the main vehicle that propagates excitation signals on the surface of neurons, so that affected neurons become easily excitable. Insects poisoned by DDT show typical hyperexcitation symptoms that lead to exhaustion and death. This phenomenon may be better understood as an electrophysiological manifestation in which neurons affected by DDT show a typical excitation pattern called "repetitive discharges." Such a neuron that has been excited by a stimulus remains in an excited state and continues to discharge for several minutes.

The most well-known use of insecticidal DDT is probably for mosquito control in malaria eradication programs. The most frequently used technique was that of "wall painting" of the interior of buildings with DDT in areas where malaria was prevalent. Because mosquitoes transmit malaria directly from human to human (i.e., without going through other hosts), this method effectively cuts off the link to continued transmission. The two key properties of DDT responsible for its effectiveness are the extreme susceptibility of mosquitoes to DDT and the long-lasting nature of DDT, particularly in indoors and dry environments.

DDT was also well known for its role in the control of cotton insect pests that posed a serious problem to cotton growers in the southern United States. The most commonly used formulation was a mixture of DDT and toxaphene. DDT was also used to control many other pests including the bark beetle vectoring Dutch elm disease, locusts, and forest pests (e.g., spruce budworm); these wider uses resulted in environmental loading of DDT-R.

ENVIRONMENTAL EFFECTS

Although *p,p′*-DDT is really the only component of DDT-R potent enough to be an insecticidal ingredient (as far as environmental effects are concerned), all the DDT-related compounds are presumed to be potentially toxic. Perhaps the best example of the extreme toxicity of DDE is its effects on bird reproduction. Because DDT is slowly converted into DDE in the environment over many years, environmental samples of DDT-R today are actually mostly DDE.

Another important compound is *o,p′*-DDT, which is known to mimic the actions of estrogen in several vertebrate biological systems. The action of *o,p′*-DDT can be attributed to its ability to bind to the estrogen receptor as an agonist, like estrogen itself, and to activate estrogen signals in the organism. Interestingly, *p,p′*-DDE acts as an antagonist to the androgen receptor in males, thereby blocking male sex hormone signaling in many vertebrate species.

Of all the effects of DDT-related compounds on wildlife, the biological damage cited most frequently is that of eggshell thinning. This phenomenon was originally reported by Ratcliffe in 1967 and verified by Anderson and Hickey in 1976 in North America. In addition to DDT, both DDE and polychlorinated biphenyls also have deleterious effects on eggshell production. Eggs affected by these chemicals crack easily and contribute to the decline of vulnerable bird species.

Eggshell thinning is not the only harmful effect for which DDT-R has been implicated. DDT-R has also been shown to contribute to the increased mortality as well as myriad reproductive problems among a broad range of wildlife including birds, fish, and other aquatic organisms. Behavioral changes are also caused by exposure to DDT-R.

A current view among scientists is to interpret many of these effects as "endocrine disruptions" caused by the hydrocarbon pollutants, with DDT-R being one of the prominent study materials. Certainly, DDT-R, particularly *o,p′*-DDT, acts in an estrogen-like manner, whereas *p,p′*-DDE acts as an anti-androgen. Deleterious effects of such endocrine disruptions by DDT-R in birds are well documented. Because disruptions of endocrine actions, including those of some vitamins, are expected to cause serious effects on reproduction, development, and nutritional balance of animals, this topic is likely attract increased attention in the scientific community.

Despite the difficulty of conducting and evaluating environmental effects studies, evidence for the harmful biological effects of DDT on wildlife and ecosystems has been overwhelming. Clearly, the decision to ban the use of DDT was sound.

See Also the Following Articles

Insecticides • Integrated Pest Management • Pollution

Further Reading

Bradley, D. J. (1998). The particular and general issue of specificity and verticality in the history of malaria control. *Parasitologia (Rome)* **40,** 5–10.

Fry, D. M. (1995). Reproductive effects in birds exposed to pesticides and industrial chemicals. *Environ. Health Perspect.* **103** (suppl. 7)**,** 165–171.

Matsumura, F. (1985). "Toxicology of Insecticides," 2nd ed, pp. 51–55. Plenum Press, New York.

Metcalf, R. M. (1955). "Organic Insecticides," pp. 127–180. Interscience, New York.

Peakall, D. (1970). Pesticides and the reproduction of birds. *Sci. Am.* **222,** 72–78.

Peterle, T. J. (1991). "Wildlife Toxicology," pp. 157–172. Van Nostrand Reinhold, New York.

Ratcliffe, D. A. (1967). Decrease in eggshell weight in certain birds of prey. *Nature* **215,** 208–210.

Defensive Behavior

Justin O. Schmidt
Southwestern Biological Institute, Tucson

Defensive behaviors are the responses of organisms to perceived threats by potential predators. The responses can

be active and obvious to an outside observer, including the predator; they can be subtle and difficult to observe; or they can be completely inapparent. Obvious responses might include escape flight or changing to a menacing posture, a subtle response might be the "freezing" of a slowly moving insect, and an inapparent behavior might be the warming of flight muscles by a large moth or beetle in anticipation of flight from a detected predator. The goal of this article is to describe the major types of defensive behaviors of insects and to illustrate how, when, and why these defenses are of survival benefit.

All animals must eat. Food choices for animals are limited to materials derived from other life forms, with flesh from animals being among the richest food sources in energy and nutrients. This sets the evolutionary stage for fierce competition among organisms to eat others, yet not to be eaten themselves. Good defenses and defensive behaviors tip the balance from mere survival of a population (or its extinction) to success and domination of a niche.

Survival and reproduction are the key elements of life. For both elements, defense is a paramount feature; without defense, survival, and therefore reproduction, is unlikely. Insects must defend against microorganisms, parasites, and predators and use different strategies against each. The defenses against these attackers differ. The ultimate defense against microorganisms, including bacteria, viruses, protozoans, and fungi, is the immune system. Parasites pose a different challenge. These multicellular organisms live in or on the insect body, sapping vital nutrients and reserves, sometimes damaging essential tissues or organs and causing death. Defenses against parasites are primarily behavioral and life history strategies, with backup from the immune system after parasite attack. Parasitoids are a curious group of attackers that share properties of both parasites and predators. They, like parasites, live in or on the body of the host insect and feed on its blood and tissues. Like parasites, they also do not immediately kill the insect. Parasitoids resemble predators, in fact some consider them predators, because their mode of delivery usually involves direct physical attack on the prey. Parasitoids differ from parasites because instead of directly killing the host, the parasitoid lays one or more eggs or larvae on or in the host and then the parasitoid larva(e) consumes and kills the host. Insect defenses against parasitoids are a combination of defenses used against predators and against parasites. Attacks by adult parasitoids are met with behavioral and morphological defenses similar to those used against predators. Deposited eggs and larvae are resisted by encapsulation by the immune system and other physiological defenses. Predators, in contrast to microorganisms and parasites, directly attack and kill or paralyze their insect prey. They also possess more complex nervous systems than parasites and use this added ability to combine enhanced sensory awareness, decision-making, and learning to challenge the limits of the insect prey to detect, respond to, and defend against the predator. Consequently insects have evolved a myriad of defenses of dazzling form and complexity against their predatory enemies.

FACTORS AFFECTING PREDATORS AND PREY

The predator–prey equation is never constant. Age and size of respective predator and prey, hunger levels, population sizes, presence of alternative prey, and behavioral factors are ever changing. Size of a potential predator relative to prey size is an obvious factor affecting defensive capabilities of an insect. For example, an ant's mandibles might be an effective defense against a small jumping spider, but likely are ineffective against an anteater. Hunger is an important, often overlooked, factor in the defensive equation. Investigators sometimes starve a potential predator for a period of time to ensure that it is hungry when presented with a potential prey. This often yields false impressions, because a starved predator is much more likely to try to attack almost anything that might be edible than would a well-fed predator. An analogy from human experience is instructive. Humans faced with starvation from war or other disasters have eaten rats or cockroaches in an effort to survive, but these same people would not consider such items when not starving. The effect of predator hunger can greatly affect the success of insect defenses. Hunger level can also affect prey insects by inducing them to forage for food during more dangerous times and for longer periods.

Population levels also influence the success of insect defenses. Cryptic (concealed) caterpillars (Figs. 1 and 2) are at more risk of failure of their defensive concealment when high populations of paper wasps *(Polistes)* are present than at times of low wasp numbers. The opposite situation, high populations of prey, can turn the defensive tables in favor of the prey. The synchronous emergence of periodic cicadas and mayflies not only serves reproductive benefit but also saturates the predators in the environment, reducing the risk to each individual cicada or mayfly. Presence or absence of

FIGURE 1 Looper caterpillar (Geometridae) with fleshy body projections whose shape and appearance closely match the vegetation of its host plant, *Polygonella* sp., providing excellent camouflage and protection from visually searching predators. (Author photograph, Florida, U.S.A.)

FIGURE 2 Second instar of *Eumorpha typhon* cryptically matching background grape leaf as it rests. (Author photograph, Arizona, U.S.A.)

alternative prey affects the success of various insect defenses. Some species of caterpillars exhibit several different color patterns. These color differences form the basis for "apostatic selection," which confers protective benefit on the rare color morph (form). The rarer morph is safer because birds adopt search images or searching behaviors oriented toward discovery of the common-color morph and often miss rare-color morphs that do not fit the image. Predator behavioral factors can determine the effectiveness of the defensive behavior of prey. Prey speed and flight ability often provide excellent protection from predators that actively search for prey. Flies and bees rarely fall prey to roving spiders, but often are captured by ambush sit-and-wait crab spiders. In the examples of vinegaroons (whiptail scorpions, *Mastigoproctus giganteus*) and tarantula spiders, which are classical ambush predators, fast and powerful prey such as sulpugids (wind scorpions) and centipedes are surprised by the ambush predator and fall prey. The element of surprise is crucial. Without surprise, the powerful jaws and quickness of alert sulpugids and centipedes make them formidable prey that would be difficult, if not impossible, to overcome by predators that have, at best, only equal equipment.

OVERRIDING PRINCIPLES

Predator–prey relationships are not chaotic, but are based upon time-tested principles with constant refining and generation of new approaches. Principles for prey include: (1) it is better to avoid attack than to defend against an attack; (2) the higher the cost or penalty suffered by a predator, the greater protection gained by the prey; and (3) avoid sharing time and space with as many potential predators as possible. Combinations of these principles form the basis for most successful defenses of insects. However, the story is complex: insects usually face not just one, or even a few, species of predators, but rather a whole suite of potential predators. And the biggest, fiercest predator usually is not the one that poses the greatest risk. To be successful, insect populations must maximize their defensive success against the summation of all predator attacks encountered. This is the combination of the probability of attack by each potential type of predator and the probability of success in countering the attack. Herein lies the experimentalist's dilemma: how can we know all of the predators that have impacted the evolution of an insect's defenses, the frequency of the attacks by each predator type, the success of the attacks by each predator type, and whether "phantom" predators, predators that might have posed serious threats in the past or become threats only at very infrequent intervals, exist. In addition, to access accurately the role of predators, the effects of predators during both bottleneck and outbreak periods of the insect's population must be evaluated. A clever experimentalist is an outstanding observer of natural history and the biology of the insect in question and successfully controls as many variables as possible to resemble nature.

EVOLUTIONARY STRATEGIES AND DEFENSIVE BEHAVIORS

Crypsis

Crypsis, or avoiding detection by blending into the background, is one of the most common and successful defenses. Classical examples of crypsis include stick insects in the order Phasmida, leaf-mimicking moths, and ambush bugs (Phymatidae) that resemble the flowers in which they hide. In the first two examples, crypsis functions to avoid detection by visually hunting predators such as birds and monkeys. The crypsis of ambush bugs serves dual purposes of concealment from potential predators and from their prey, flower-visiting insects. Variations of the cryptic theme can take many forms. Many caterpillars, moths, and other insects are patterned and colored like the vegetation they eat or the twigs, bark, or other substrate upon which they rest (Figs. 1 and 2). In this way they blend toward invisibility in the eyes of all but the best predators. Other cryptic specialists resemble dead objects ranging from bird droppings, for some swallowtail butterfly caterpillars (*Papilio* spp.), to stones, for some grasshoppers *(Eremocharis insignis)* and toad bugs (Gelastocoridae). Cryptic insects match behavior to lifestyle. To maintain their concealment cryptic insects tend to move little during the day, and when they do move it is slow and deliberate to avoid notice. Exceptions are cryptic predators that must move swiftly during the act of prey capture. Cryptic insects tend to select resting backgrounds, lighting conditions, and positions to match their own appearance. How they recognize and choose matching backgrounds is unclear.

Aposematism

The opposite of crypsis is aposematism, or a warning signal to predators. Aposematic insects usually appear and behave

FIGURE 3 Aposematic venomous caterpillar, *Automeris metzli* (Saturniidae), whose painful urticating hairs can cause a long-lasting burning rash. (Michael F. Wilson photograph, Chihuahua, Mexico; used with permission.)

very different from cryptic insects. These warning-colored insects sport bright pattern and color combinations of reds, yellows, oranges, whites, and blacks that are notably conspicuous within the background setting and advertise the insect's presence (Fig. 3). Insects with warning coloration, like cryptic insects, tend to move slowly in their environment. Similarity ends there; the slow, deliberate movement of aposematically colored insects serves not to avoid detection, but to provide time for potential predators to detect and recognize the insect before acting. Warning coloration boldly signals that this insect is toxic, is bad tasting, or can sting or cause injury.

Aposematic warnings have evolved for every sensory system used by predators. Acoustic stridulations, hisses, and other sounds advertise to potential predators that the emitting insect is dangerous. Toxic tiger moths (Arctiidae) send loud return sounds to approaching insectivorous bats. These sounds warn the bat that the moth is unpalatable and potentially harmful, and may also interfere with the echolocation capabilities of bats. Often aposematic sounds share similarities: they have broad frequency ranges, have low pattern complexity, and resemble "white noise." Such signals are readily detected by a wide range of predators, differ from most nonwarning sounds produced by insects or other animals, and are poorly structured for precise intraspecific communication, e.g., courtship sounds. These are exactly the properties a good warning signal needs: they are readily detected, conspicuous, and generalizable and they indicate no form of communication other than warning. Aposematic sounds are produced by a wide range of insects, including velvet ants (Mutillidae), bees and flies, assassin bugs (Reduviidae), moths, many groups of beetles, cockroaches, and grasshoppers.

Aposematic odors and tastes serving as warnings need not be toxic, only readily apparent. Examples of warning odors include the strong, unpleasant fragrances of pyrazines produced by butterflies and a variety of other insects, ketones produced by velvet ants, and uncharacterized compounds produced in mandibular glands of tarantula hawks (Pompilidae, *Pepsis*). Warning tastes are probably widespread among noxious insects, but few have been investigated. Predators such as lizards, birds, fish, and spiders will often "taste" an insect before actually killing and eating it. If the sampled insect has a compound with an aposematic taste, it is often released unharmed. Tiger moths externalize some of their pyrrolizidines, which, when tasted by orb weaver spiders, cause the spider to cut its web to release the unharmed moth.

Mimicry

Aposematic organisms having effective defensive systems become models for biological copycats. Some of these mimicking organisms are cheaters who do not actually possess noxious or dangerous properties like those of the models they resemble. They "trick" predators into perceiving them as noxious and unpalatable. This kind of mimicry is called Batesian mimicry, in honor of the famous naturalist Henry Bates, who first recognized the phenomenon. The pipevine swallowtail *(Battus philenor)* is a toxic butterfly that serves as the model for several Batesian mimics, including the dark morph of the palatable female (but not male) tiger swallowtail *(Papilio glaucus).* In regions where model pipevine swallowtails are abundant, most tiger swallowtail females are mimics; but in areas where pipevine swallowtail models are rare, many female tiger swallowtails have patterns that are not mimics. Other mimicking species are "truthful" because they themselves are noxious or toxic and resemble other noxious or toxic insects. Such "Müllerian" mimicry abounds among butterflies, especially within the passion-flower-feeding butterflies in the genus *Heliconius* in South America or the queen butterflies (*Danaus* and relatives) of Africa. Between the extremes of Batesian and Müllerian mimicry exists a continuous gradation of mimicking species that possess varying degrees of noxiousness.

Allomones

Allomones are chemical defenses. Many hundreds of allomonal chemicals have been identified from insects, most being distasteful, damaging, or toxic to other animals or synergizing the activity of other active chemicals in the secretion. Beyond their common properties for defense, allomones share little chemical similarity and contain compounds of almost all imaginable types, including organic acids, alcohols, ketones, aldehydes, esters and lactones, hydrocarbons, terpenes, phenolics, quinones and hydroquinones, amines, alkaloids, sulfurous compounds, steroids, polysaccharides, peptides, and proteins. Allomones are the defensive arsenal's "backup artillery" used to blunt direct attack by predators. The effectiveness of allomones is embodied in the common names of some insects and their relatives: stink bugs, blister

beetles, bombardier beetles, and vinegaroons. The aldehydic secretions of stink bugs spread over the body surface, providing a repellent odor barrier that doubles as a repugnant contact liquid when touched by a predator. Bombardier beetles (*Brachinus* spp.) and vinegaroons (*Mastigoproctus* spp.) spray corrosive quinones/hydroquinones and concentrated acetic acid solutions, respectively, at approaching predators, thereby actively extending protection to a distance and reducing the risks of physical attack by the predator. Insect venoms are effective specialized allomones typically consisting of water-soluble proteins and other components that are injected into the body of an assailant. The effectiveness of stings as defenses against large predators is evident from the names "killer bees," "fire ants," and "cow killers." Other names, including hornet and wasp, have become terms synonymous with pain, power, and fear.

DEFENSIVE PLOYS

For virtually any sensory system and behavior used by predators to detect prey, prey insects have evolved counterstrategies or defensive ploys. For predators relying mainly on vision, insects possess physical properties and behaviors either to avoid being seen or to maximize being seen. For predators that rely primarily on sound, prey have counterbehaviors to minimize sound generation, to evade sound-emitting predators, or to counter with effective sounds of their own. For predators using olfaction as their primary searching sense, prey have evolved systems to reduce their own odor, to mask it, to mimic the odors of unsuitable prey, or to blunt sensory orientation with allomones and aposematic odors. As a rule, insects supplement general defenses and behavioral strategies with multiple suites of defenses directed toward specific sensory systems.

Concealment and Hiding

Hiding from predators is a nearly universal tactic of insects. Even well-defended insects such as stinging wasps conceal their nests within dense vegetation, among roots, or in holes. Aposematic insects such as *Dasymutilla occidentalis* tend to rest in concealed places during periods of inactivity and run and hide under leaves or among vegetation when an approaching potential predator is sensed. Toxic butterflies often rest with wings folded and in among vegetation that hides them from view. Many other insects are masters of concealment and are so cryptically colored and patterned that finding them in a photograph has become an educational and entertaining challenge for children and adults alike. Some insects are concealed only during particular times. Caterpillars are commonly concealed on bark or in the ground while at rest, but are more apparent while feeding on leaves.

Concealment and hiding take many forms. Less obvious than daytime concealment against visual predators, but an equally frequent and effective defense, is use of time for concealment. Looper caterpillars (Geometridae) and others that are cryptically colored and concealed feed on leaves during daylight. At night when they are not cryptic or concealed from spiders, beetles, ants, and predators that search for prey mechanically and by olfaction and vibration, some conceal themselves by terminating feeding and hanging below the vegetation on long silken threads. An extreme of concealment is used by mayflies, which "conceal" their adulthood by reducing their adult life to only about a day, just enough time to mate, lay eggs, and die.

Escape

If an insect is capable of flying, jumping, running, crawling, or dropping to safety, escape is often the first response to detection by a predator. Nevertheless, not all insects can, or do, attempt to escape when approached by predators. Many larvae live in confined spaces and can move little or slowly, and none can fly. These individuals must rely primarily on other means of protection such as concealment, crypsis, or chemical defense. Other species rely first on their aposematism and/or noxious nature for protection and only attempt escape secondarily.

Fighting Back

When concealment and escape fail, most insects resist and fight back by biting with mandibles, kicking and struggling, stinging, and releasing allomones, with varying success. Leg spurs and spines are used effectively by some large moths, including sphinx moths in the genera *Manduca* and *Eumorpha* and cockroaches *(Archimandrita marmorata),* to defeat the grasp of even large potential predators. These sharp spines not only can painfully pierce skin but also can anchor strong kicks to free the slippery insect from grasp. Some male wasps possess either sharp genitalia or separate "pseudostings" that are jabbed into grasping predators. Jabbed predators might mistake pseudostings for actual stings of female wasps and release the male.

Pain

Pain is the early warning system to indicate that bodily damage is occurring, has occurred, or is about to occur. Bodily damage is a serious threat and risk to an organism's ability to survive, feed, and reproduce. When given a choice between a meal with accompanying pain (plus perceived bodily damage) and the loss of a meal, predators often opt for the latter. The venomous stings of wasps, bees, and ants are legendary for their abilities to cause pain and deter predation. Spiny caterpillars and an assortment of biting bugs and beetles, including assassin bugs (Reduviidae), giant water bugs (Belostomatidae), water scorpions (Nepidae), and predaceous diving beetles (Dytiscidae), also produce painful venoms. Allomones can be effective by causing immediate

pain. Examples are formic acid, sprayed by ants in the subfamily Formicinae, carabid beetles, and notodontid caterpillars, and quinones, released by tenebrionid and carabid beetles. Although bites and kicks might induce pain, their overall effectiveness relative to venoms and allomones suffers from lesser ability to produce pain and from predator familiarity with them and their expected effects.

Warnings

Stereotyped warnings are used to threaten and intimidate predators. Paper wasps (*Polistes* spp.) on their nest face large adversaries with raised wings, waving front legs, abdomens curved toward the predator, and wings flipped, fluttered, or buzzed. These threats inform the predator that it is spotted and an attack will ensue if the advance continues. Hissing cockroaches *(Grophadorhina portentosa)* threaten by hissing, which resembles the defensive hiss of a snake. Many flies and harmless bees and wasps buzz loudly when grabbed. These aposematic buzzes sound similar to those of painfully stinging honey bees and wasps and often serve as effective warnings.

Surprise and Startle

Insects that perceive an approaching predator can use the elements of surprise and startle to escape. Surprise combined with rapid escape flight are often a sufficient defense. If a predator is adept at pursuit, the execution of surprise and rapid flight, followed by instantaneous concealment upon landing, becomes a powerful defense. Startle is the combination of the elements of surprise and fright. Examples of startle are dull cryptic moths, which, when detected by a bird or monkey, flash their hidden hind wings, revealing bright colors (Fig. 4) or large frightening eyespots mimicking those of an owl or large predator (Fig. 5). Desert clacker grasshoppers, *Arphia pseudoneitana,* are grand masters of surprise and startle. When approached by a large animal, these inconspicuous desert grassland insects suddenly jump into the air, fly away amidst a confusion of bright red wing flashes and loud clacking noise, and disappear into distant vegetation as suddenly as they appeared. In addition to startle, this behavioral display provides the predator with a search image for red color, which can cause the predator to overlook the dull grasshopper.

Confusion

Individuals in schools of fish, flocks of birds, and herds of running African ungulates present difficult targets for predators. Use of confusion of predators via mass motion is little studied in insects but likely is an important defensive behavior in some situations. The constant movement and hopping of masses of migratory locusts and the seemingly erratic circling flights of flies disturbed from a fresh cow patty are likely examples of the use of confusion as a defense.

Aggregations

When many individuals aggregate in a group, each member receives protection through the presence of the others. Not only does a group present fewer locations with prey but also the individuals within a group gain protection through reduction in the chance of being the chosen prey by a discovering predator. Aggregation as defense is particularly effective if the individuals are toxic or are defended as are ladybird beetles (Coccinellidae), monarch butterflies, milkweed bugs (Lygaeidae), or social wasps on a nest. In these insects, a potential predator need sample only one or a few individuals to learn the unsuitability of the whole.

Association

Protection can be achieved by living near or associating with a defended or noxious species. The tropical paper wasp, *Mischocyttarus immarginatus,* prefers to make its small nest with few individuals near the much larger nest of stinging *Polybia occidentalis,* a common social wasp. The arrangement seems to provide protection for the *Mischocyttarus* from vertebrate predators. No potential benefit to the *Polybia* has been demonstrated. Honey bees and some species of ants are known to nest in portions of termite mounds. The exact

FIGURE 4 Underwing moth (Noctuidae) with cryptic leaf-mimicking front wings (note that the mimicry even includes "mold" spots) that normally cover the bright hind wings, which are exposed to startle predators. (Author photograph, Borneo.)

FIGURE 5 Warning eyespots resembling the eyes of an owl or mammal are displayed as a threat by the leaf-mimicking silk moth, *Automeris cecrops,* when disturbed. (Michael F. Wilson photograph, Arizona, U.S.A; used with permission.)

nature of these associations is unclear and the ants generally attack termites if given the chance. The benefits to the bees and ants are more obvious; they not only share the moderated temperature and humidity environments produced by the termites, but they also gain protection and reduced risk of discovery by being in the termite mound.

Sociality

Unlike a solitary individual that must detect and defend against predators alone, individuals of social species enjoy benefits of group defense. A group of many coordinated individuals can more readily detect predators than solitary individuals, can then recruit others via alarm pheromones or vibrational signals to the common defense, and can launch effective attacks *en masse.* Group attacks are particularly effective when individuals possess painful stings or bites and when the attackers are nonreproductive workers who can sacrifice themselves in battle with little reproductive loss to themselves or the colony as a whole. Predators confronted by a "cloud" of attackers cannot devote attention to defending against each attacker, and reduced predator vigilance enhances attacker chances of scoring an effective sting or bite. Sociality and defense are strongly synergistic and, when combined, go a long way toward explaining the success of social insects.

Overall, insects as a class have taken defensive behaviors to levels unsurpassed in number, complexity, and creative diversity within animal life.

See Also the Following Articles

Aposematic Coloration • Crypsis • Ladybugs • Mimicry • Venom • Wasps

Further Reading

Blum, M. S. (1981). "Chemical Defenses of Arthropods." Academic Press, New York.

Cott, H. B. (1940). "Adaptive Coloration in Animals." Methuen, London.

Edmunds, M. (1974). "Defence in Animals." Longman, Essex, U.K.

Evans, D. L., and Schmidt, J. O. (eds.) (1990). "Insect Defenses: Adaptive Mechanisms and Strategies of Prey and Predators." State University of New York Press, Albany.

Schmidt, J. O. (1982). Biochemistry of insect venoms. *Annu. Rev. Entomol.* **27,** 339–368.

Wickler, W. (1968). "Mimicry in Plants and Animals." McGraw–Hill, New York.

Dengue

Thomas W. Scott
University of California, Davis

Dengue is a human disease caused by a virus that is transmitted from one person to another by the bite of infected mosquitoes. Worldwide, dengue virus infections cause more human morbidity and mortality than any other arthropod-borne virus. It is estimated that 2.5 to 3.0 billion people are at risk of infection each year, and millions have been infected during recent epidemics. Dengue occurs throughout the tropics, where incidence rates have steadily increased since the 1950s. The most severely affected areas are urban centers of Southeast Asia, where dengue hemorrhagic fever and dengue shock syndrome (DHF/DSS) are among the leading causes of pediatric hospitalization. During the past 20 years dengue has emerged as a major international public health threat; during that time, changes in dengue epidemiology were most pronounced in the Americas. Epidemics in Cuba and Venezuela during the early 1980s have elevated concern that the progression of dengue outbreaks in the Western Hemisphere is following a pattern similar to that observed in Asia over the past 50 years, putting people in the New World tropics at increased risk for severe, life-threatening disease.

HISTORY

Dengue viruses are believed to have originated in tropical forested habitats, moved from there to rural environments, and finally invaded urban centers. The word dengue most likely originated from Swahili, and following a series of modifications in pronunciation and spelling the word evolved to its present form. The earliest recorded epidemics of a dengue-like illness were in China during the Chin dynasty (265–420 AD). During the 18th and 19th centuries sporadic epidemics were reported in Asia and the Americas. Following World War II, the pattern changed from one of periodic outbreaks to one of continuous transmission of multiple virus serotypes in Asian cities. It was from that situation that DHF/DSS surfaced in 1954 in the Philippines.

DISEASE

Dengue fever (DF), DHF, and DSS are caused by four closely related but antigenically distinct single-stranded RNA viruses (DEN-1, 2, 3, and 4) in the genus *Flavivirus*, family Flaviridae. All four serotypes cause a range of human disease, including asymptomatic infections; undifferentiated fever; classic DF, an acute febrile illness with headache, body aches, and rash; DHF; and DSS. Sequential infections with different serotypes are possible because infection with one serotype provides lifelong protection from a homologous infection, but is only briefly cross-protective against heterologous serotypes. Even though most infections, especially in children under 15 years of age, are asymptomatic, it is estimated that annually there are between 50 and 100 million DF cases and between 250,000 and 500,000 DHF/DSS cases worldwide. If untreated, the case fatality rate for DHF/DSS can approach 20%; however, with supportive therapy (fluid and electrolyte management and oxygen) fewer than 1% of severely ill patients die.

The etiology of serious illness is not completely understood. Risk of DHF/DSS is highest in places where two or more viral serotypes are simultaneously transmitted. People with preexisting dengue antibodies, which can be obtained actively from a previous infection or passively by infants from their mothers, are 100-fold more likely to experience severe disease following infection with a heterologous virus serotype than are people without preexisting anti-dengue antibodies. The mechanism most commonly attributed to severe dengue is immune enhancement. Preexisting nonneutralizing antibodies complex with the infecting heterologous serotype virus, which enhances phagocytosis (entry) and replication of virus in mononuclear cells, which then leads to increased vascular permeability and hemorrhaging. An alternative explanation for severe disease is that different serotypes or strains of virus vary in virulence. Severe disease and death have been reported following primary dengue infections. Recent studies in the Americas suggest that the two hypotheses are not mutually exclusive. Immune enhancement may require secondary infection by genotypes of virus from Southeast Asia.

VIRUS TRANSMISSION

Although other mosquitoes in the subgenus *Stegomyia,* including *Aedes albopictus* and *Ae. polynesiensis,* have been implicated as vectors in jungles and rural habitats, and on islands in the South Pacific, *Ae. aegypti* is the most important dengue vector. Although there is some evidence of sylvatic transmission cycles that include nonhuman primates and vertical transmission from an infected female mosquito to her progeny, the majority of dengue virus transmission is between mosquitoes and human hosts. Horizontal virus transmission begins when a mosquito imbibes viremic human blood. Virus enters and replicates in midgut epithelial cells, disseminates to the hemocoel, and infects secondary target tissues, including the salivary glands. Following replication in salivary gland acinar cells, virus is released into the salivary matrix and can be transmitted the next time the infective mosquito probes its mouthparts into a human host in an attempt to locate blood. Extrinsic incubation in the mosquito requires 10 or more days, depending on the ambient temperature. Once infective, *Ae. aegypti* can transmit virus each time it probes its mouthparts into a host or imbibes a blood meal. Incubation in the human host typically ranges from 4 to 7 days, after which the person is viremic for ~5 days. Fever subsides in concert with the inability to detect virus in the blood.

MOSQUITO VECTOR

Ae. aegypti is uniquely adapted to a close association with humans and efficient transmission of dengue virus. Immature forms develop primarily in man-made containers. Highly anthropophilic adult females rest inside houses where they feed on human blood. Unlike most other mosquito species, which engage in a feeding duality of plant carbohydrates for synthesis of nutrient reserves and blood for egg development, female *Ae. aegypti* forego sugar meals and feed almost exclusively and frequently on human blood. Relatively low concentrations of the amino acid isoleucine in human blood are believed to be responsible for the ability of *Ae. aegytpi* to use only blood to meet their energy needs and to complete vitellogenesis. Females fed only human blood have higher measures of fitness (survival and reproduction), and thus a selective advantage, over those fed sugar and blood. Because females seldom disperse beyond 100 meters and consequently food, mates, and oviposition substrates are readily available within the human habitations where they reside, rapid synthesis of glycogen from sugar substrates for extended flight is not necessary. To meet their energy and reproductive needs, females must imbibe more than one blood meal in each gontorophic cycle, something that increases contact with human hosts and opportunities to contract or transmit a viral infection. Because of their unusual propensity to make frequent and preferential contact with humans, *Ae. aegypti* is an exceedingly efficient vector of dengue virus even though compared to other mosquito species they are not especially susceptible to virus infection. Relatively low *Ae. aegypti* population densities have been associated with virus transmission. It is expected that entomological thresholds for dengue virus are quite low.

CONTROL

Dengue control is dependent on control of *Ae. aegypti* because there is no licensed vaccine for dengue, which will need to be tetravalent because of the phenomena of immune enhancement, and no clinical cure has been found. Despite a history of detailed study of *Ae. aegypti,* vector control programs for dengue control are often nonexistent or ineffective. Outdoor applications of aerosol insecticides to kill adults have in most instances not been effective because the majority of females rest indoors where they avoid contact with the insecticide. Over any considerable period of time, control of immatures using chemicals, biological control, or community-based source reduction has been effective only in authoritative systems in which negative consequences are associated with noncompliance. Disease control based on genetic manipulation of mosquito vectors, rendering them incapable of transmitting virus, is currently being investigated but this method will require extensive evaluation before it can be deployed. Of great concern is the observation that explosive dengue epidemics occur even when *Ae. aegypti* population densities are low. This apparent paradox illustrates that aspects of *Ae. aegypti* biology other than population density, such as their blood-feeding behavior, duration of extrinsic incubation, and female survival, play an important role in defining virus transmission dynamics. Control of dengue constitutes a formidable challenge for public health officials.

See Also the Following Articles

Medical Entomology • Mosquitoes • Zoonoses

Further Reading

Gubler, D. J., and Kuno, G. (1997). "Dengue and Dengue Hemorrhagic Fever," p. 462. CAB Int., Wallingford, U.K.

Halstead, S. B., Rojanasuphot, S., and Sangkawibha, N. (1983). Original antigenic sin in dengue. *Am. J. Trop. Med. Hyg.* **32,** 154–156.

Harrington, L. C., Edman, J. D., and Scott, T. W. (2001). Why do female *Aedes aegypti* (Diptera: Culicidae) feed preferentially and frequently on human blood? *J. Med. Entomol.* **38,** 411–422.

Kuno, G. (1995). Review of the factors modulating dengue transmission. *Epidemiol. Rev.* **17,** 321–335.

Monath, T. P. (1994). Dengue: The risk to developed and developing countries. *Proc. Natl. Acad. Sci. USA.* **91,** 2395–2400.

Rigau-Perez, J., Clark, G. G., Gubler, D. J., Reiter, P., Sanders, E. J., and Vordam, A. V., (1998). Dengue and dengue haemorrhagic fever. *Lancet* **352,** 971–977.

Scott, T. W., Amerasinghe, P. H., Morrison, A. C., Lorenz, L. H., Clark, G. G., Strickman, D., Kittayapong, P., and Edman, J. D., (2000). Longitudinal studies of *Aedes aegypti* (L.) (Diperta: Culicidae) in Thailand and Puerto Rico: Blood feeding frequency. *J. Med. Entomol.* **37,** 89–101.

Watts, D. M., Porter, K. R., Putvatana, P., Vasquez, B., Calampa, C., Hayes, C. G., and Halstead, S. B. (1999). Failure of secondary infection with American genotype dengue 2 to cause dengue haemorrhagic fever. *Lancet* **354,** 1431–1434.

Dermaptera
(Earwigs)

Susan M. Rankin and James O. Palmer
Allegheny College

The Dermaptera (earwigs) comprise a small, relatively old, hemimetabolous order of insects characterized in their external anatomy by paired cerci (forceps) at the posterior end, and (in winged forms) short tegmina incompletely covering hind wings that are also unique structurally (Fig. 1). Behaviorally, earwigs are thigmotactic, nocturnal, and subsocial, in a system whereby the female parent broods, grooms, and defends eggs and young nymphs (Fig. 2). Internal anatomy is typical of orthopteroids, except that the corpora allata have undergone fusion to a single median structure, and the paired ovaries are primitively polytrophic (i.e., each follicle contains an oocyte and a single nurse cell).

Earwigs are members of the orthopteroid assemblage and have a strong sister-group relationship with the Dictyoptera; they also may be closely related to the Grylloblattodea. Four suborders are generally recognized, and of the three extant ones, the Hemimerina and Arixinina are small groups of viviparous ectoparasites of vertebrates; most species of earwigs are oviparous members of the third group, the Forficulina. Typically, classification schemes have relied on features of the male genitalia. Earwigs are not of medical importance; they do not crawl in people's ears (occasional anecdotal accounts notwithstanding), and they do not bite, although some may pinch with their forcepslike cerci. Some earwigs may be pests of gardens or households; alternatively, some species are important biocontrol agents, feeding on agricultural pests, such as aphids, armyworms, mites, and scale insects.

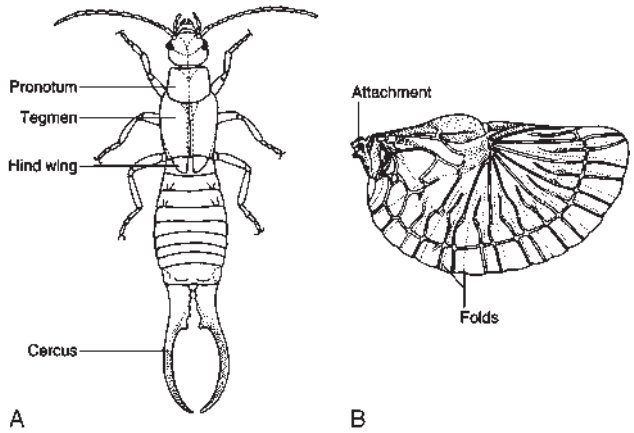

FIGURE 1 European earwig *(Forficula auricularia):* (A) adult male and (B) his right hind wing. [Reprinted with the permission of Scribner, a Division of Simon & Schuster, from "College Entomology" by E. O. Essig (Macmillan, New York, 1942).]

GENERAL CHARACTERISTICS

In 1773 DeGeer coined the term Dermaptera (but used the name for all orthopteroids). Kirby in 1815 introduced the name in its current sense, as a small insect order of about 2000 described species. The oldest known examples of dermapterans are Jurassic fossils dating from about 208 mya. These elongate, slender, hemimetabolous (incompletely metamorphic) insects have chewing mouthparts, three-segmented tarsi (in extant groups), (usually) compound eyes, and no ocelli. The presence of abdominal cerci makes them easy to distinguish from beetles. The cerci are typically forcepslike (though they are filiform in at least some parasitic forms) and are sexually dimorphic (Fig. 1). The forceps are used for a variety of purposes, including prey capture,

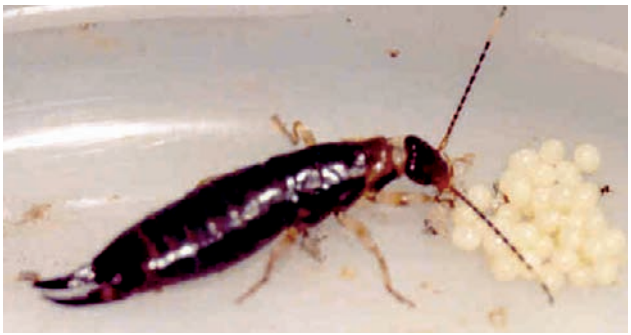

FIGURE 2 Female ringlegged earwig *(Euborellia annulipes),* brooding over her clutch of eggs.

defense, fighting, and as aids in copulation and in folding of hind wings. Earwigs are diploid (i.e., they have a double set of chromosomes); males are heterogametic (i.e., they produce gametes with different sex chromosomes, e.g., X and Y).

Some earwig species are wingless as adults, but most have short front wings (tegmina) that do not cover the abdomen. The derivation of the name "dermaptera" (*derma,* skin, *ptera,* wing) refers to the thickened or "skinlike" front wings. The hind wings are unlike those of any other group of insects: they are semicircular and membranous, with radially arranged veins. They fold fanlike beneath the front pair when the insects are at rest (Fig. 1). The derivation of the common name (earwig) may be a corruption of "earwing," in reference to the hind wing resemblance to a human ear. Alternatively, it could be a reference to the ancient Anglo-Saxon legend that these insects crawl in ears of sleeping humans. Additionally, the forceps of some species look like instruments once used for piercing women's ears for earrings.

Earwigs typically display parental care of offspring (although there are almost no observations of maternal care in the viviparous, ectoparasitic forms). Eggs are typically deposited in soil (or protected whorls of monocotyledonous plants); the females "roost" on the eggs until the young hatch and then they care for them (Fig. 2). The period of maternal care appears to be a time of nonfeeding of the brooding female; physiologically, her levels of both juvenile hormone and ecdysteroids (see later) are likely low during the period of egg care.

INTERNAL ANATOMY

The major elements of the neuroendocrine system are the brain, the subesophageal ganglion, three thoracic ganglia, and six abdominal ganglia (with thick, paired connectives between the ventral ganglia). Paired neurohemal corpora cardiaca are connected to the brain and frontal ganglion by strong nervous connections; the closely associated single median corpus allatum produces and releases juvenile hormone III and is in close proximity to the neurohemal dorsal aorta. The digestive system contains the typical regions of fore-, mid-, and hindgut (though gastic caecae are lacking); the midgut–hindgut junction is characterized by the presence of numerous long, slender (excretory) Malpighian tubules.

The female reproductive system consists of paired ovaries, lateral oviducts, a median oviduct, spermatheca (for sperm storage), and genital chamber. Earwigs are unusual in that the female genital opening (gonopore) is just behind the seventh abdominal segment. The ovaries are primitively polytrophic; in some species the long ovarioles branch off the lateral oviduct, while in others, short ovarioles appear in series around the oviduct. The viviparous species are pseudoplacental, with egg maturation and embryonic development taking place in the greatly enlarged vitellarium. The male reproductive system is complex, with paired testes, paired vasa deferentia, paired or single vesicula seminalis (for sperm storage), and a paired or single common ejaculatory duct ending in the sclerotized virga (penislike structure). Anatomy of the male reproductive system has been used extensively in classification schemes of earwigs, as discussed shortly.

RELATIONSHIPS TO OTHER INSECTS

Earwigs are members of the informal orthopteroid assemblage and share a sister-group relationship with the dictyopterans (cockroaches). The Grylloblattodea (rock crawlers) may be linked to the Dictyoptera/Dermaptera. Alternatively, earwigs may be closer to the Grylloblattodea than to any other orthopteroid order.

PHYLOGENY AND DISTRIBUTION OF DERMAPTERA

Early in the 20th century, Burr established suborders of the Dermaptera recognized by most contemporary systematists. The four suborders (three of them recent) are as follows:

1. Archidermaptera, represented by 10 fossil specimens from the Jurassic; they are characterized by unsegmented cerci and tarsi having four or five segments.
2. Forficulina, the suborder containing most earwigs (i.e., 1800 described species, in 180 genera); cerci are unsegmented (except in a few primitive larvae) and forcepslike.
3. Hemimerina, composed of 10 species in one genus; they have filiform (segmented) cerci, and are wingless, blind, viviparous (pseudoplacental) ectoparasites of African rats.
4. Arixenina, composed of five species in two genera; like the Hemimerina, they are viviparous (pseudoplacental), wingless, blind, and ectoparasitic of vertebrates. The Arixenina live on bats in Malayan–Philippine region; Popham considered this group to be a sister group of the Labiidae.

Popham based his phylogeny of families on the structure of the male genitalia (Fig. 3). The Hemimerina consists only of

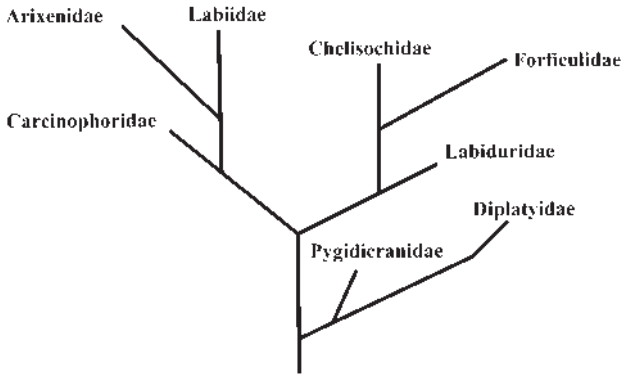

FIGURE 3 Phylogeny of the families of the Dermaptera. [Amended figure reprinted with the permission of Cambridge University Press from Popham, E. (1965). The functional morphology of reproductive organs of the common earwig *(Forficula auricularia)* and other Dermaptera with reference to the natural classification of the order. *J. Zool.* **146**, 1–43.]

the family Hemimerinidae. The Pygidicranidae are frequently regarded as the most primitive family of earwigs, in as much as the males have two functional penis lobes. The Carcinophoridae also show less specialized, ancestral traits (and are typically treated as wingless, though some are fully winged). The Arixinidae and Labiidae are treated as sister groups (though this is not generally accepted). The Chelisochidae and Forficulidae are the most specialized families (i.e., display the most derived traits). The Forficulidae are considered to be "higher earwigs" (more recently derived) because the males have a single functional penis lobe. This family is the best-represented family in North America.

The current geographical distribution of most families of earwigs was largely determined by continental drift, with two main centers of radiation before the Triassic opening of the Pacific ocean being the equatorial region of the eastern Pacific (Pygidicranidae, Carcinophoridae, Arixinidae, and Labiidae) and the Afro-Indian circumtropical center (Labiduridae, Chelosochidae, and Forficulidae).

The distribution has also been largely affected by climatic conditions: earwigs were "discouraged" from spreading northward from the tropics by mountain ranges of southern Europe and Asia; only more specialized members have become established in the Palearctic region (and none have been reported in the polar regions). Although most earwigs have wings, they seldom fly. The cosmopolitan distribution of some species can be attributed to their habit of hiding in crevices, especially in timber or other material that is transported by commerce.

NATURAL HISTORIES AND BIOTIC ASSOCIATIONS

Development

In oviparous species such as European earwig, *Forficula auricularia,* and ringlegged earwig, *Euborellia annulipes,* clutches of ovoid, creamy white eggs are laid in protected burrows; eggs can be up to 2 mm in length in larger species. Nymphs generally resemble adults but can be distiguished from them by lighter color, shorter antennae, a male-type 10-segmented abdomen (rather than the 8-segmented abdomen of the adult female), and typically female-type forceps. Sexes are not easily distinguished externally in nymphs.

Habits

Most earwigs are thigmotactic and nocturnal, inhabiting crevices of various types, bark, fallen logs, and debris. Cavernicolous (cave-dwelling) blind species have been reported in the Hawaiian islands and in South Africa. Food typically consists of a wide array of living and dead plant and animal matter. Some earwigs have scent glands opening onto the dorsal side of the third and fourth abdominal segments, and from these they can squirt a foul-smelling yellowish-brown fluid some 10 or so cm, presumably for protection.

Reproductive Strategies

Among the Forficulina, reproductive strategies range from oviparous iteroparity (many clutches, the likely ancestral condition) to semelparity (one clutch, a trait derived in colder climates). In contrast, the Hemimeridae are viviparous ectoparasites, residing in the fur of central African rats of the genera *Cricetomys* and *Beamys* and dining on the skin and body secretions of the host, while the viviparous Arixenina are ectoparasites of bats.

Natural Enemies

Some beetles, toads, snakes, birds, and bats have been reported as predators of earwigs. *F. auricularia* can be parasitized by gregarines (sporozoans, which may be harmless) in the gut, as well as nematodes, cestodes, mites, and tachinid flies (a potential biocontrol agent). A fungal parasite has also been associated with several earwig species.

MEDICAL AND ECONOMIC IMPORTANCE

Earwigs are harmless to humans: they carry no known pathogens of humans, and their mouthparts are incapable of biting humans (although some species can pinch).

Some genera (e.g., *Forficula, Labidura, Euborellia*) are repeatedly reported as a pest of homes, gardens, and orchards. Their thigmotactic nature, coupled with (known and suspected) aggregation pheromones, can lead to high densities of earwigs in and around homes. In gardens, earwigs may attack seedlings and soft fruit. Management in backyard gardens can be accomplished by persistent trapping in bamboo tubes, rolled-up newspaper, or low-sided cans filled with vegetable oil. Removing refuge sites, such as ivy and piles of leaves, is also helpful.

Some species are also of importance to commercial agriculture, being pests of ginger, maize, and of honey bee colonies. However, earwigs are also regarded as valuable biocontrol agents for crop pests, consuming armyworms, aphids (of various types), mites, scale insects, sugarcane rootstock borers, and tropical corn borers. Several dermapteran species are found in commercial egg houses and have potential as biocontrol agents for fly eggs and larvae.

See Also the Following Articles

Cave Insects • Juvenile Hormone • Orthoptera

Further Reading

Burr, M. (1911). "Genera Insectorum" (P. Wytsman, ed.), pp. 1–112. L. Desmet-Verteneuil, Bruxelles.

Crumb, S. E., Eide, P. M., and Bonn, A. E. (1941). The European earwig. U.S. Department of Agriculture Technical Bulletin 766.

DeGeer, C. (1773). "Mémoires pour servir à l'histoire des insectes," Vol. 3. Hesselberg, Stockholm. Lucas, W. J. (transl.) (1920). "A Monograph of the British Orthoptera." The Ray Society, London.

Giles, E. T. (1963). The comparative external morphology and affinities of the Dermaptera. *Trans R. Entomol. Soc. Lond.* **115**, 95–164.

Popham, E. J. (1963). The geographical distribution of the Dermaptera. *Entomologist* June, 131–144.
Popham, E. J. (1965). The functional morphology of the reproductive organs of the common earwig (*Forficula auricularia*) and other Dermaptera with reference to the natural classification of the order. *J. Zool.* **146**, 1–43.

Development, Hormonal Control of

Michael E. Adams
University of California, Riverside

The majority of insects undergo embryonic development within an egg and then advance through a series of immature larval stages culminating in metamorphosis to the adult, reproductive form. It has been suggested by Lynn Riddiford and James Truman that the stunning evolutionary success of insects is attributable to complete metamorphosis, whereby the immature larva, essentially a gut covered with cuticle, is exquisitely adapted for resource exploitation, rapid growth, and avoidance of competition with its conspecific adult, reproductive stage. Metamorphosis is a magnificent transformation of one body form into a completely different one under the control of hormones. Upon reaching the requisite body size, precisely timed hormonal signals are released, committing the animal to a postembryonic rebirth. Carroll Williams summarized this basic developmental process with the following anecdote:

> *The earth-bound stages built enormous digestive tracts and hauled them around on caterpillar treads. Later in the life-history these assets could be liquidated and reinvested in the construction of an entirely new organism—a flying machine devoted to sex.*

Conversion of the energy accumulated by the larval form into the adult during metamorphosis is a fascinating process of organismal remodeling. It is accomplished via programmed cell death of larval-specific cells, reprogramming of others, and postembryonic birth of new cells from imaginal disc tissues upon receipt of precisely timed and coordinated hormonal signals.

INSECT BODY PLANS AND DEVELOPMENTAL PROGRAMS

Three distinct patterns of growth are observed in insects, distinguished by body form and type of metamorphosis: ametabolous, hemimetabolous, and holometabolous. Insect orders exhibiting ametabolous development belong to the Apterygota, or wingless insects. Included in this group are the primitive orders Protura, Collembola, and Diplura, whose earliest stages are miniature adults in form, except for the absence of external genitalia. They grow continuously and lack metamorphosis. Attainment of reproductive competence occurs at an indefinite time in development, and they continue to molt even as adults. Little is known about the hormonal control of development in the Ametabola.

The majority of insects are winged and have either partial or complete metamorphosis. The Hemimetabola, such as grasshoppers and crickets, emerge from the egg formed as small, immature versions of the adult and are called nymphs. They lack wings and functional reproductive organs. After a series of molts, the number usually constant from one generation to another, nymphs pass directly to the winged, reproductive adult stage in a single step. This mode of development is referred to as incomplete metamorphosis. The more advanced insect orders, including moths, beetles, flies, and wasps, develop as vermiform (wormlike) larvae during the immature stages. The complete metamorphosis of these groups is a two-step process in which a sessile, nonfeeding pupal stage is intermediate between larva and adult. The pupal stage allows for a complete change in body form from larva to winged, hexapod adult. This total transformation of body form during complete metamorphosis requires a high degree of postembryonic cellular programming under the control of hormones. Three types of cellular processes are dictated, mostly by hormonal signaling involving ecdysteroids and juvenile hormones. First, many larval-specific structures such as body wall muscles and neurons must be eliminated through a programmed cell death known as apoptosis. Second, some cells persist to the adult stage, but are extensively remodeled to serve adult functions. Finally, new cells derived from imaginal discs are born.

The diversity of insect groups places limits on generalizations about hormonal control of development. The effects of hormones in one group may not be the same for other groups, because of different patterns of growth and cellular specification. In the Lepidoptera, for example, larval epidermal cells change their cuticle secretory program at metamorphosis and switch to production of pupal and subsequently adult cuticle. Exogenous juvenile hormone (JH) application at this time maintains the larval secretory program, resulting in supernumerary (extra) instars. Similarly, properly timed application of JH in the early pupal stage of moths causes a second pupal stage. However, the fruit fly *Drosophila* and other higher flies are largely resistant to such effects of JH. This may relate to differences in the developmental program of fly epidermal cells. The entire epidermis in the head and thorax is programmed for secretion of larval cuticle only and dies at metamorphosis. It is replaced by imaginal disc tissue, which remains undifferentiated throughout larval life in the presence of JH. Abdominal epidermal cells also die after pupation and are replaced by abdominal histoblasts. Thus the development program of epidermis in higher flies is entirely distinct from the Lepidoptera, and its lack of response to JH in the early stages may be a consequence of this very different design.

These differences in responses to hormones complicate interpretation of many findings, especially because it has become fashionable to test overall hypotheses using the moth, *Manduca,* for physiology and endocrinology experiments, and *Drosophila* for genetic manipulations. Generalizing about common mechanisms for these two evolutionarily distant groups should be done cautiously.

INSECT DEVELOPMENTAL HORMONES

Ecdysteroids

Ecdysteroids are relatively polar steroid hormones released by the prothoracic glands in immature stages and the gonads in adults. They are the chief regulators of gene expression in insect development and reproduction. The first structure to be elucidated was α-ecdysone (αE), the immediate precursor to 20-hydroxyecdysone (20HE), accepted as the main protagonist in ecdysteroid actions. While 20HE indeed is associated with the majority of ecdysteroid actions, there is evidence to support αE as a signaling molecule in certain instances. Other suspected ecdysteroids are 20,26HE and makisterones.

Juvenile Hormones

JHs, sesquiterpenoid derivatives from the sterol synthesis pathway, are released by the corpora allata. Six types are known in insects, with JH III being the predominant form. The chief action of JHs is to modulate ecdysteroid-mediated gene expression. No receptors for JH have been clearly identified, but it seems likely that the hormone interacts with intracellular receptors or proteins that modulate ecdysteroid signaling. The signature of JH action is to promote expression of the immature phenotype.

Prothoracicotropic Hormone

Prothoracicotropic hormone (PTTH) is a large peptide hormone released by brain neurosecretory cells from terminals in the corpora cardiaca or corpora allata. PTTH stimulates the prothoracic gland to synthesize and release αE or, in some instances, 3-dehydroecdysone, which are both then converted to 20HE in the hemolymph or by target tissues. PTTH is a homodimer consisting of two 12-kDa subunits joined by a disulfide bond. Release of PTTH is regulated by sensory inputs to the brain, which convey information about body size and nutritional state.

Bursicon

Bursicon is a 30- to 40-kDa peptide that accelerates sclerotization of cuticle. It is released from neurosecretory cells in the brain and ventral nerve cord after each ecdysis. Since insects are particularly vulnerable to predation during and after ecdysis, rapid hardening of the cuticle maximizes survival.

Eclosion Hormone

Eclosion hormone (EH) is a 62-amino-acid peptide hormone that mediates circadian-mediated eclosion to the adult stage and each larval ecdysis. Eclosion hormone is released into the blood from ventral median neurosecretory cells of the brain, causing release of ecdysis-triggering hormone (ETH) from Inka cells of the epitracheal endocrine system. It is also implicated in the elevation of cyclic GMP in a subset of neurons in the central nervous system that initiate ecdysis behavior.

Ecdysis-Triggering Hormones

ETHs are peptide hormones synthesized and released from Inka cells at the end of the molt. ETHs act directly on the central nervous system (CNS) to cause a behavioral sequence that leads to shedding of the cuticle. ETH release is caused by eclosion hormone, and the action of ETH on the CNS leads to release of eclosion hormone from VM neurons. It has been proposed that ETH and EH engage in a positive-feedback signaling pathway that results in depletion of ETH from Inka cells, which is necessary for the transition from preecdysis to ecdysis behaviors.

MODES OF HORMONE ACTION

Hormones are chemical messengers that travel throughout the body to effect responses in specific tissues. Targeted cells have receptors that, upon binding the hormone, transduce the signal into a cellular response. Insect hormones regulate development by activation of either intracellular receptors or receptors at the cell membrane.

Ecdysteroids and Juvenile Hormones Activate Intracellular Receptors

Ecdysteroids and juvenile hormones are relatively lipophilic signaling molecules able to easily traverse the cell membrane. Upon entry, they bind to intracellular proteins called nuclear receptors or nuclear transcription factors, which reside either in the cytoplasm or in the nucleus. Regardless of their initial location, hormone binding triggers passage to the nucleus, where the receptor forms a complex with other proteins and then binds directly to DNA, inducing or repressing gene expression.

Several factors govern diverse, stage-specific responses of target cells to ecdysteroids. First, ecdysteroid receptors occur as multiple subtypes, including EcR-A, EcR-B1, and EcR-B2. The response to ecdysteroids is governed by the subtype expressed by target cells, as well as which subtype of USP (USP-1 or USP-2), the EcR partner, is expressed. The transient availability of receptors in target cells leads to sensitive periods at specific stages of development. For example, both EcR-A and EcR-B1 are present throughout larval life, but the ratio of the two favors EcR-B1. In contrast, levels of EcR-A

increase during metamorphosis. Second, the affinity of EcR subtypes can vary, as can the kinetics of a given response to an ecdysteroid peak. These patterns of receptor and dimer partner expression appear to mediate different cellular responses to ecdysteroids at different developmental times.

Peptide Hormones Activate Cell-Surface Receptors

Most peptide hormones bind to a class of integral membrane proteins on the plasma membrane of the target cell. The majority of these are G-protein-coupled receptors that trigger intracellular second messenger cascades.

EMBRYOGENESIS

Distinct patterns of embryonic development are observed in hemimetabolous and holometabolous insects. In both types, the embryo produces multiple cuticular layers, and the appearance of these coincides with pulses of ecdysteroids. Juvenile hormone levels are generally low during early embryogenesis, but climb later to program nymphal or larval cuticle formation upon appearance of an ecdysteroid peak.

In the hemimetabolous grasshopper *Locusta,* four peaks of ecdysteroids are observed, corresponding to production of serosal cuticle and three embryonic cuticle layers (Fig. 1A). JH levels are elevated immediately after oviposition, because of maternal contribution to the yolk, but rapidly decrease to low levels. At ~20% of embryonic development and before prothoracic glands are developed, the first ecdysteroid peak consisting exclusively of αE occurs in the presence of relatively low JH levels. This first peak comes from the release of maternal ecdysteroids stored as polar conjugates, and shortly thereafter the serosal cuticle is secreted. A second αE peak occurs at ~30% development, leading to formation of the first embryonic cuticle. A third ecdysteroid peak occurs just after differentiation of the prothoracic glands, and this coincides with the first appearance of 20HE and 20,26HE. Nevertheless, levels of αE together with 20,26HE predominate at this time, leading to the secretion of the second embryonic cuticle, which Truman and Riddiford refer to as pronymphal cuticle. It is the first layer of cuticle tough enough to require shedding via ecdysis behavior. The coincidence of the early, αE peaks with cuticle secretion suggests that αE is not only a 20HE precursor, but also a biologically active hormone at certain times of development. At 70% development, the first embryonic cuticle is shed and followed quickly by a large peak of ecdysteroids. This is the first exposure of the embryo to substantial 20HE levels in the presence of JH, leading to synthesis of the first instar nymphal cuticle. This peak of ecdysteroids contains large amounts of αE, 20HE, and 20,26HE.

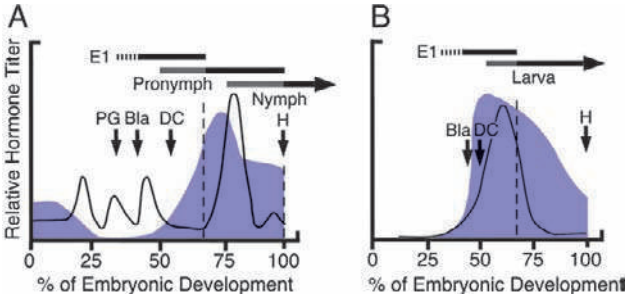

FIGURE 1 Hormone levels during embryonic molts in a hemimetabolous insect, the grasshopper *Locusta migratoria* (A), and in a holometabolous insect, the moth, *Manduca sexta* (B). During *Locusta* embryogenesis, four peaks of ecdysteroid are observed, each corresponding to secretion of a layer of cuticle. The first peak of predominantly αE at 20% development initiates secretion of the serosal cuticle (not shown). A second peak prior to prothoracic gland differentiation (PG) causes secretion of the first embryonic cuticle (E1). Just after blastokinesis (Bla), a third ecdysteroid peak leads to secretion of the pronymph cuticle. Shading of the pronymph time line at the top of the plot indicates the pharate stage, which ends with ecdysis of the E1 cuticle (vertical dotted line). The fourth ecdysteroid peak, occurring for the first time in the presence of JH, contains approximately equal amounts of αE, 20HE, and 20,26HE. This causes secretion of the first-stage nymphal cuticle. At hatching the nymph sheds the pronymph cuticle upon escaping from the substrate. Secretion of the first embryonic cuticle in *Manduca* occurs in the absence of an ecdysteroid peak. The first larval cuticle is secreted in response to elevated ecdysteroids in the presence of JH just after dorsal closure (DC). Note that JH levels in *Manduca* rise earlier in development than in *Locusta* and that ecdysteroid signaling always occurs in the presence of JH. (Adapted from Truman and Riddiford, 1999, 2002.)

At hatching, the grasshopper emerges from the egg under the ground, still surrounded by the pronymphal cuticle. Despite its hexapod body plan, the animal exhibits a classic vermiform (wormlike) locomotory pattern as it escapes the egg pod and maneuvers through the substrate to the surface. In a matter of seconds to minutes, the pronymphal cuticle is shed, and the animal stretches its legs and switches abruptly to hexapod behavior. Truman and Riddiford have called attention to many similarities between the hemimetabolous pronymph and the holometabolous larva, suggesting that the latter has resulted from a hormonal shift in embryogenesis, resulting in an extended postembryonic phase of pronymph development. The ancestral pronymph undergoes an extended, multistage developmental sequence as a larva in the Holometabola.

The importance of a JH-free period during early embryogenesis of hemimetabolous insects (grasshopper—*Schistocerca,* cricket—*Acheta*) has been demonstrated by treatment of eggs with JH analogs. This results in inhibition of blastokinesis, reduction in the number of embryonic cuticle layers produced, premature appearance of nymphal cuticle and mouthparts, and reduced body size.

Embryogenesis in holometabolous Lepidoptera is somewhat simpler, with the secretion of only three cuticles, one serosal and two embryonic cuticles. Levels of both ecdysteroids and JH are undetectable early in embryogenesis, but rise earlier compared with the Hemimetabola, or at about 30% development (Fig. 1B), preceding the ecdysteroid peak. Therefore, unlike hemimetabolous embryogenesis, the first exposure to ecdysteroids occurs in the presence of JH, leading to production of the first larval cuticle. An

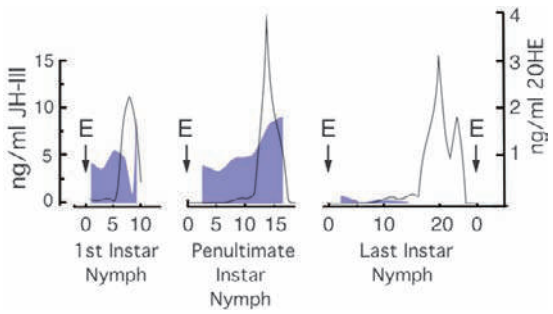

FIGURE 2 Hormonal regulation of development in postembryonic stages of the cockroach *Nauphoeta cinerea,* representing hemimetabolous development. During the immature stages of *Nauphoeta,* JH levels are elevated each time ecdysteroid levels rise to initiate a molt, resulting in secretion of nymphal cuticle. Following ecdysis to the last instar nymph, JH levels drop precipitously. Adult commitment is signaled by a biphasic ecdysteroid peak in the absence of JH. Incomplete metamorphosis occurs during the period between ecdysteroid elevation and ecdysis to the adult stage. (Adapted from Lanzrein *et al.*, 1985.)

embryonic ecdysis occurs at 70% development, and first instar larval hatching does not involve cuticle shedding. The importance of a JH-free period observed for hemimetabolous insects is not the case for embryogenesis in the Holometabola. Embryos are largely insensitive to exogenous JH treatment, suggesting perhaps the absence of receptors for these hormones until later in embryogenesis.

LARVAL DEVELOPMENT

The Intermolt

Because the exoskeleton places limits on growth, insect development occurs in stages, each ending with molting and cuticle shedding, or ecdysis. During the intermolt, which follows ecdysis, JH levels are maintained around 1 to 10 ng/ml in the blood (Fig. 2). It is presumed that these JH levels promote a high metabolic rate, active feeding behavior, synthesis of larval cuticle proteins, and continuous proliferation (but not differentiation) of imaginal discs. Growth during the immature stages is possible because the immature integument is predominantly unsclerotized procuticle, which is quite flexible compared with hard, sclerotized adult cuticle. Several mechanisms allow for larval cuticle expansion. The epidermal cells add new protein to the cuticle throughout the intermolt, increasing the surface area by intussusception. In addition, new cuticle in *Manduca* is deposited in vertical columns that are gradually reoriented during the feeding stage to allow for expansion. The increase in size during the larval stage can be quite impressive, as in fifth-instar *Manduca,* which increases its body weight from ~1 g on the first day of development to 15 g by the end of the instar, and its cuticular surface area by approximately fivefold just prior to pupation. Blood-feeding insects such as *Rhodnius* are known to release serotonin after a blood meal, which acts as a plasticizing agent, facilitating the enormous expansion of the body wall after a blood meal.

The Molt

At some point during each immature stage, growth results in a decision by the brain to initiate the molt. In *Rhodnius,* the simplest case known, stretch-receptor input from the abdomen to the brain causes release of PTTH, which induces synthesis and secretion of ecdysteroids from the prothoracic glands. In most insects, the decision to release PTTH is more complicated and less well understood, but it has to do with body weight, nutritional state, and time spent at that stage. The immediate effects of ecdysteroid elevation include cessation of feeding and apolysis, the detachment of the old cuticle from underlying epidermal cells. Apolysis of larvae results in head capsule slip, which occurs because the new head capsule is larger than the old one. This is the most visible sign that the molt has been initiated. If elevation of ecdysteroids occurs in the presence of JH, epidermal cells maintain the secretory program for immature phenotype, and larval cuticle is secreted (Fig. 2). Through the action of molting fluid, most components of the old cuticle are broken down and recycled into the new layer.

During the period of new cuticle synthesis, ecdysteroids also orchestrate gene expression crucial to the synthesis and action of peptide hormones that control ecdysis behaviors. Ecdysis is a complex process in which the old cuticle is shed not only from the surface of the animal, but also from the lining of the foregut, the hindgut, and the inner walls of the tracheal system. Success in this process depends on completion of new cuticular synthesis, attachment of the musculature to this new cuticle, and digestion of the old cuticle. In addition, the animal prepares for a sequence of Houdini-like escape behaviors necessary to shed the old cuticle. These consist of preecdysis and ecdysis behaviors. The ability to perform these behaviors depends on orchestration of a peptide signaling cascade involving the central nervous system and the epitracheal endocrine system.

For the ecdysis signaling cascade to be functional at the appropriate time, ecdysteroids orchestrate gene expression in four ways. Genes are activated in epitracheal glands to increase production of ecdysis triggering hormones (ETHs). Release of ETHs initiates ecdysis behaviors through direct action on the CNS. Although the CNS is not sensitive to ETHs during the feeding stage, acquisition of sensitivity occurs upon elevation of ecdysteroids, specifically around the time of apolysis. Third, the nervous system becomes competent to release EH, a peptide hormone that targets Inka cells to cause release of ETH. Finally, elevated ecdysteroids exert a negative influence on the secretory competence of Inka cells. As long as ecdysteroids remain high, Inka cells are unable to secrete ETHs in response to EH exposure. This latter effect of ecdysteroids, to block release of ETHs from Inka cells,

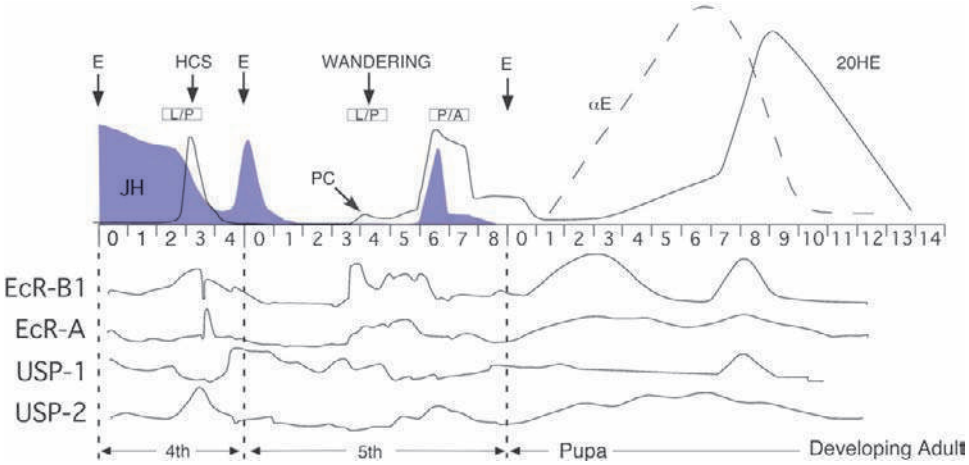

FIGURE 3 Hormonal regulation of development in postembryonic stages of the moth, *M. sexta,* exemplifying the Holometabola. Elevated JH levels occur throughout larval development in *Manduca,* in which metamorphosis is encoded by two ecdysteroid peaks. During the fifth instar, JH levels drop and an ecdysteroid pupal commitment peak (PC) signals a change in the way epidermal cells respond to the secretory program from larval to pupal phenotype. This peak also triggers cessation of feeding and wandering behavior. The next time epidermal cells are exposed to ecdysteroid and JH on day 6 leads to secretion of pupal cuticle. During the pupal stage, αE and 20HE rise in the complete absence of JH, signaling adult commitment. Sensitive periods of epidermal cell commitment are shown as L/P (larval vs pupal specification) and P/A (pupal vs adult specification of imaginal disc tissue). Fluctuating levels of ecdysteroid receptor subtypes (EcR- B1, EcR-A) and USP subtypes (USP-1, USP-2) are shown below. (Adapted from Riddiford and Truman, 2001, and Baker *et al.,* 1987.)

appears designed to ensure that ecdysis does not occur prematurely. The mechanism of block involves a crucial step in the EH-induced secretory mechanism of Inka cells. Declining ecdysteroid levels at the end of the molt provide the necessary signal permitting expression of one or more genes needed for secretory competence.

It is believed that initiation of ecdysis behaviors occurs as a result of an ongoing conversation between EH neurons and Inka cells. When Inka cells become sensitive to EH, ETH release initiates preecdysis behavior, which is thought to loosen the remaining connections between the new and the old cuticle. The transition from preecdysis to ecdysis occurs upon depletion of ETH from Inka cells. It is thought that the action of ETH activates a downstream cascade of peptide signaling within the CNS to regulate each unit of the behavioral sequence. Included in this cascade is a neuropeptide called crustacean cardioactive peptide (CCAP), named for its initial discovery and biological activity. In the context of insect ecdysis, CCAP appears to be an immediate chemical signal within the CNS for activation of peristaltic ecdysis behavior. Upon escaping the old cuticle, the animal is surrounded by a new soft cuticle and is therefore extremely vulnerable to injury. Release of the neuropeptide bursicon from neurosecretory cells of the CNS accelerates sclerotization of the cuticle.

In summary, ETH, EH, CCAP, and bursicon regulate ecdysis at all stages. This includes embryonic ecdysis in *Manduca,* and adult eclosion.

METAMORPHOSIS

The transition from immature to adult is signaled by the elevation of ecdysteroid levels in the absence of JH. This is a one-step process in hemimetabolous insects. During the last nymphal instar of the cockroach *Nauphoeta,* JH levels fall from 5 to 10 ng/ml to less than 1 ng/ml prior to the next ecdysteroid peak (Fig. 2). Appearance of ecdysteroids at this low JH level signals a commitment to an adult gene expression pattern. Some examples of cellular responses to this adult commitment peak include mitosis in wingpad tissue, development of flight muscles, competence of gonadal accessory glands to differentiate, formation of external genitalia, and reorganization of the nervous system to accommodate these new adult structures. Also included are more subtle alterations, such as the relative proportions of body parts and addition of secondary sexual characteristics such as acoustic organs for communication.

In the Holometabola, complete metamorphosis requires an intervening pupal stage for remodeling of the larva into an adult. During the last larval instar of *Manduca,* ecdysteroids rise on two occasions, first in the absence of JH and later in its presence (Fig. 3). The first ecdysteroid pulse is a small one during days 3 to 4, referred to as the pupal commitment peak. This is the first time in the life history of the animal that a peak of 20HE occurs in the absence of JH. This triggers cessation of feeding and a wandering behavior aimed at locating a suitable site for pupation. The pupal commitment peak prepares the genome for its response to the next ecdysteroid peak. Although the second pulse of ecdysteroids occurs in the presence of JH, the commitment peak has changed the response of epidermal cells from a larval to a pupal secretory program. Similarly, imaginal discs respond to this peak by differentiating into adult tissues, something not observed in the previous larval stages.

In the Lepidoptera, the new hormonal milieu that triggers metamorphosis produces striking changes in the CNS and musculature. These tissues must be drastically altered during

construction of the adult body form. Simultaneously, undifferentiated cells in imaginal discs proliferate and differentiate. For these tissues, the pupal commitment peak sets the stage for three types of cellular responses to the subsequent ecdysteroid peak: programmed cell death (apoptosis), cellular remodeling, or differentiation of imaginal discs. For example, some motoneurons that innervate larval-specific structures such as prolegs die shortly after pupal ecdysis. Others persist because of their involvement in the motor patterns involved in adult eclosion and die soon thereafter. Most larval neurons survive, but are remodeled to play roles in adult behavior.

The precise mechanisms governing cellular responses to the pupal commitment peak remain obscure, but the identification of EcR and USP subtypes has allowed monitoring of their expression during metamorphosis, for example, the response of epidermal cells of *Manduca* to the ecdysteroid peak during days 2 to 3 of the fourth instar by up-regulation EcR-B1, no change in EcR-A, and down-regulation of USP-1 (Fig. 3). However, during the fifth instar, the pupal commitment peak is correlated with sharply increased EcR-B1 *and* EcR-A expression, an altered USP-1 response. These altered patterns of expression apparently encode a change in downstream gene expression, leading to a pupal phenotype as well as imaginal disc differentiation during this stage.

During the pupal stage, ecdysteroid levels rise in the complete absence of JH (Fig. 3). This signals commitment to the adult phenotype and accelerated development of imaginal discs. It is remarkable that αE begins to rise on day 1 of the pupal stage, well before elevation of 20HE, and this is correlated with increases in both EcR-B1 and EcR-A expression. This suggests that αE may have a hormonal role itself in programming the adult stage. Elevation of 20HE occurs in two phases, one beginning on day 3 and a second, steeper rise on day 7. The slow gradual rise coincides with the adult commitment phase, whereas the steeper rise beginning on day 7 is associated with differentiation of new tissues. This latter phase coincides with a rapid rise of EcR-B1 and USP-1 expression.

The hormonal signaling mechanisms governing metamorphosis are complex and include a diversity of hormones, receptors, and varying temporal patterns of hormone release and receptor expression. Ecdysteroid signaling in the presence or absence of JH can set the stage for changes in the programming of target tissues, such as epidermal cells that secrete cuticle. Depending on the responses of EcR and USP subtypes, qualitatively different cellular programs are initiated.

See Also the Following Articles

Ecdysteroids • Embryogenesis • Imaginal Discs • Juvenile Hormone • Metamorphosis • Mating Behaviors • Temperature, Effects on Development and Growth

Further Reading

Baker, F. C., Tsai, L. W., Reuter, C. C., and Schooley, D. A. (1987). *In-vivo* fluctuation of JH, JH acid and ecdysteroid titer and JH esterase activity during development of fifth stadium *Manduca sexta. Insect Biochem.* **17,** 989–996.

Ewer, J., Gammie, S. C., and Truman, J. W. (1997). Control of insect ecdysis by a positive-feedback endocrine system: Roles of eclosion hormone and ecdysis triggering hormone. *J. Exp. Biol.* **200,** 869–881.

Gammie, S. C., and Truman, J. W. (1997). Neuropeptide hierarchies and the activation of sequential motor behaviors in the hawkmoth, *Manduca sexta. J. Neurosci.* **17,** 4389–4397.

Gilbert, L. I., Tata, J. R., and Atkinson, B. G. (1996). "Metamorphosis: Postembryonic Reprogramming of Gene Expression in Amphibian and Insect Cells." Academic Press, San Diego.

Lanzrein, B., Gentinetta, V., Abegglen, H., Baker, F. C., Miller, C. A., and Schooley, D. A. (1985). Titers of ecdysone, 20-hydroxyecdysone and juvenile hormone III throughout the life cycle of a hemimetabolous insect, the ovoviviparous cockroach *Nauphoeta cinerea*. *Experientia (Basel)* **41,** 913–917.

Lageaux, M., Hetru, C., Goltzene, F., Kappler, C., and Hoffmann, J. A. (1979). Ecdysone titer and metabolism in relation to cuticulogenesis in embryos of *Locusta migratoria. J. Insect Physiol.* **25,** 709–723.

Nijhout, H. F. (1994). "Insect Hormones." Princeton University Press, Princeton, NJ.

Riddiford, L. M., Cherbas, P., and Truman, J. W. (2001). Ecdysone receptors and their biological actions. *Vitam. Horm.* **60,** 1–73.

Temin, G., Zander, M., and Roussel, J.-P. (1986). Physico-chemical (GC–MS) measurements of juvenile hormone III titres during embryogenesis of *Locusta migratoriasta. Int. J. Invertebr. Rep. Dev.* **9,** 105–112.

Truman, J. W., and Riddiford, L. M. (1999). The origins of insect metamorphosis. *Nature* **401,** 447–452.

Truman, J. W., and Riddiford, L. M. (2002). Endocrine insights into the evolution of metamorphosis in insects. *Annu. Rev. Entomol.* **47,** 467–500.

Zitnan, D., Kingan, T. G., Hermesman, J., and Adams, M. E. (1996). Identification of ecdysis-triggering hormone from an epitracheal endocrine system. *Science* **271,** 88–91.

Zitnan, D., Ross, L. S., Zitnanova, I., Hermesman, J. L., Gill, S. S., and Adams, M. E. (1999). Steroid induction of a peptide hormone gene leads to orchestration of a defined behavioral sequence. *Neuron* **23,** 523–535.

Diapause

David L. Denlinger
Ohio State University

Diapause is a form of developmental arrest in insects that is much like hibernation in higher animals. It enables insects and related arthropods to circumvent adverse seasons. Winter is most commonly avoided in temperate zones, but diapause is also used to avoid hot, dry summers and periods of food shortage in the tropics. Unlike quiescence, which represents a halt in development elicited immediately at any stage by an adverse condition, diapause is a developmental response that is expressed only during a specific developmental stage, which depends on the species of insect. For example, the commercial silkworm *(Bombyx mori)* always diapauses as an early embryo, the European corn borer *(Ostrinia nubilalis)* as a fifth instar, the cecropia moth *(Hyalophora cecropia)* as a pupa, and the Colorado potato beetle *(Leptinotarsa decemlineata)* as an adult. A few species are capable of entering diapause several times,

but this usually occurs only in species living at high latitudes for which several years may be required for the completion of development. If the diapause occurs in response to environmental cues it is referred to as "facultative diapause," but if it occurs during each generation regardless of the environmental cues it receives, it is considered an "obligatory diapause." Facultative diapause is by far the more common, but several important species such as the gypsy moth *(Lymantria dispar)* have an obligatory diapause.

Embryonic diapauses are common in many of the Lepidoptera, in many Hemiptera, and in some Diptera such as mosquitoes. The arrest can occur at any stage of embryonic development, from shortly after fertilization (e.g., commercial silkworm) until after the first instar has already been fully formed (e.g., gypsy moth). Larval diapauses, especially common in the Lepidoptera, are most frequent in the final instar but they sometimes occur in earlier instars as well, e.g., the second instar of the spruce budworm *(Choristoneura fumiferana).* Pupal diapause is well known for the Lepidoptera and Diptera. Usually the arrest of pupal development occurs in the true pupal stage, but there are a few examples of diapause occurring in pharate adults (after completion of adult differentiation but before adult eclosion). Adult diapause is common in the Coleoptera, Hemiptera and Homoptera, Hymenoptera, Orthoptera, and Neuroptera, as well as some Diptera and Lepidoptera. Adult diapause, sometime referred to as a reproductive diapause, represents a halt in reproduction. Ovaries of females remain small, and the oocytes within the ovarioles contain little or no yolk. In males of some species, the testes remain small during diapause, but in others the testes are well developed and contain sperm. Male accessory glands, the organs that produce spermatophores and factors responsible for sperm activation, usually remain small and inactive during diapause. Mating behavior is strongly suppressed during diapause. In wasps, mating takes place in the autumn; males die soon thereafter and only the females overwinter in diapause. In many other insects, both sexes overwinter and mating takes place in the spring, after diapause has been terminated. Some species, such as lacewings and weevils, mate both before and after diapause.

In preparation for diapause, the insect usually sequesters additional energy reserves and moves to a site that is somewhat protected from the full onslaught of the inclement environmental conditions. Such sites may be underground, beneath debris on the soil, within galls and other plant tissues, or inside cocoons or other structures constructed by the insect. A migratory flight may be a preparatory step for diapause. This may include a short flight to a fence row or a local wooded area, but in the extreme it may be a long-distance flight, as made by the monarch butterfly *(Danaus plexippus)* when it leaves its summer habitat in Canada and the northern regions of the United States and flies to the highlands of Mexico or California to spend the winter in an adult diapause.

Upon entering diapause, development (or reproduction if it is an adult diapause) is halted and metabolic activity is suppressed. Usually, feeding ceases during diapause; thus, the insect is forced to survive on the energy reserves it has garnered prior to the onset of diapause. It is not unusual for an insect destined for diapause to sequester twice as much lipid reserves as its counterpart that is not programmed to enter diapause. The economic utilization of these reserves is enhanced by the suppression of metabolism, and for poikilotherms such as insects, the low temperatures prevailing during winter further serve to conserve energy reserves. Another challenge faced by diapausing insects is the lack of access to free water. Although some insects may drink during diapause, certain stages such as embryos and pupae do not have this option. This lack of water poses special constraints for an organism as small as an insect. Their large surface-to-volume ratios make insects particularly vulnerable to water loss across the surface of their integument. Two features appear to be common adaptations for maintaining water balance during the long months of diapause. The cuticles of many diapausing insects are coated with extra thick layers of wax that are effective in retarding water loss. In addition, a number of diapausing insects are capable of absorbing atmospheric water vapor directly through their cuticle using a mechanism that is not yet clearly understood.

Color changes are sometimes noted for diapausing individuals. For example, diapausing larvae of the southwestern corn borer, *Diatraea grandiosella,* are white, whereas their nondiapausing counterparts are brown. Reproductively active adults of a lacewing, *Chrysopa carnea,* are green but turn brown when they enter diapause in the autumn. In the spring, when the lacewings become reproductively active, they again turn green. Such changes presumably serve to camouflage the insect and help it blend with the dominant colors of the seasonal environment.

Flight muscles in many beetles and bugs degenerate when the adults enter diapause. Flight muscles are particularly expensive to maintain, thus their degeneration presumably saves energy that would otherwise be expended for maintenance of this tissue.

Several species that diapause as adults, especially beetles, bugs, and butterflies, are found in aggregations. For species that are distasteful, aggregations are likely to provide protection from predators. Such aggregations, however, may also provide another important function by providing a more stable microenvironment. In diapausing aggregations of a tropical fungus beetle, *Stenotarsus rotundus,* the beetle's metabolic rate is inversely related to group size and relative humidity. By forming an aggregation the beetles create a stable, high humidity in their environment, a feature that serves to reduce metabolic rate.

Being in diapause does not, by itself, ensure winter survival. The small size of insects implies that they quickly assume a body temperature close to that of the environment, and their body water is thus vulnerable to freezing. Diapausing insects that live in temperate and polar regions have a host of behavioral, physiological, and biochemical adaptations that

enable them to survive at low temperature. A few insects such as the goldenrod gall fly, *Eurosta solidaginis,* are freeze tolerant, which implies that they can actually survive body freezing. But, the majority of insects cannot tolerate body freezing. Such freeze-intolerant or freeze-avoiding insects prevent body freezing by several mechanisms. For example, selection of a thermally buffered microhabitat is a first line of defense. Ice nucleators such as food particles or microbes are usually eliminated from the digestive tract to reduce sites for ice formation. Glycerol, sorbitol, or other polyols serve as classic antifreezes that are synthesized and released into the body to suppress the supercooling point. Several proteins, including thermal hysteresis proteins, ice nucleator proteins, and heat-shock proteins, also contribute to cold hardiness. In some insects, such as flesh flies *(Sarcophaga),* cold hardiness is directly linked to diapause, indicating that the same genetic program that dictates diapause also results in cold hardiness. In other insects, for example the European corn borer, the two programs are regulated independently: the European corn borer enters diapause without initially being cold hardy, but it becomes cold hardy later in the season in response to prevailing low temperatures.

Diapause thus represents a syndrome of developmental, physiological, biochemical, and behavioral attributes that together serve to enhance survival during seasons of environmental adversity.

ENVIRONMENTAL REGULATION

Obligatory diapause is not elicited by environmental cues. It simply occurs in each generation when the insect reaches a certain developmental stage. In the example of the gypsy moth, diapause occurs when the embryo has completed its development and the first instar is nearly ready to hatch. With the exception of a few aberrant individuals, the gypsy moth always halts development at this time, regardless of the environmental cues they receive. In this example, environmental conditions, mainly temperature, determine when diapause should be terminated but play no role in programming the insect to enter diapause.

This is in contrast to the majority of insects, those with a facultative diapause, which use environmental cues to decide whether to enter diapause. If a certain environmental cue is received during a sensitive period the insect will enter diapause, but if this cue is not received or not received at the correct time, development will proceed without interruption. This design feature enables an insect to track seasonal changes and regulate its development accordingly. Many insects can produce multiple generations each year, and insects with a facultative diapause frequently produce spring and summer generations without diapause and then produce a generation in late summer or autumn that enters an overwintering diapause. The environmental cue used most widely to signal diapause induction is photoperiod, but temperature, food quality, and other factors may contribute to the decision.

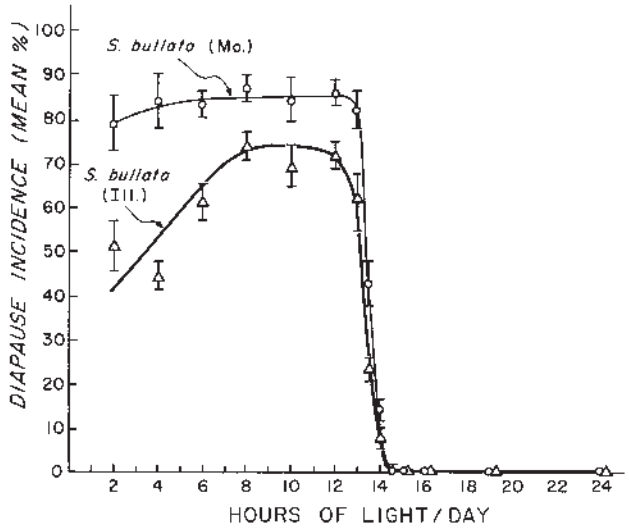

FIGURE 1 Photoperiodic response curves for pupal diapause induction in two populations of the flesh fly *S. bullata* from Illinois and Missouri. Fly cultures were maintained at 25°C under the range of daylengths indicated, and the incidence of pupal diapause was recorded. The critical daylength in this case is 13.5 h of light/day. (Reproduced, with permission, from Denlinger, 1972.)

Photoperiod

Seasonal change in daylength has all the design features that are desirable in a reliable indicator that can be used for predicting upcoming periods of inclemency. It is mathematically accurate and can be used to effectively foretell the advent of winter or other seasons that are to be avoided. The developmental period that is sensitive to photoperiod usually occurs far in advance of the actual diapause stage. Thus, diapause is not usually an immediate reaction to photoperiod but occurs in response to signals received at an earlier stage. Such early programming offers the insect a period to prepare for diapause by sequestering food reserves and making other preparatory adjustments prior to the actual onset of the developmental arrest.

For many of the insects that overwinter in diapause in the temperate regions, short daylengths dictate the expression of diapause. In the example shown in Fig. 1, flesh flies reared at long daylengths, those longer than 13.5 h, develop without interruption, but at daylengths shorter that 13.5 h, the majority enter diapause as pupae. The daylength marking the transition, 13.5 h in this example, is referred to as the critical daylength. The shape of the photoperiod response curve shown in Fig. 1 is common for temperate species that overwinter in diapause, but the curves may have different forms. For some species, especially those that undergo a summer estivation and reproduce in the autumn, long rather than short daylengths may be used to program diapause. Other species may respond to only a narrow range of daylengths for diapause induction, whereas daylengths both shorter and longer avert diapause.

Near the equator, seasonal changes in daylength are progressively less pronounced; yet insects living as close as 5°

north or south of the equator are still capable of using photoperiodic cues to regulate diapause. Diapause still exists in insects living in equatorial regions, but cues derived from temperature, rainfall, and food quality take precedence over photoperiod.

The photoperiodic response controlling diapause varies among geographic populations. Populations living at lower latitudes characteristically respond to shorter critical daylengths. An increase in latitude of 5° results in an increase in critical daylength of approximately 30 min. This pattern of variation is closely related to the latitudinal temperature gradient and is well documented in species of *Drosophila* that inhabit the Japanese archipelago. The species that occur in the subtropical zone exhibit only a weak diapause or no diapause at all. As one moves northward in the archipeligo, the diapause response becomes more pronounced and the flies use longer critical daylengths for diapause induction.

The period sensitive to photoperiod usually does not encompass the entire prediapause period, but instead a shorter interval, usually well in advance of the actual diapause stage. For example, the pupal diapause in the flesh fly *S. crassipalpis* is programmed during a photosensitive stage that includes the final 2 days of embryonic development and the first 2 days of larval life. In the tobacco hornworm, *Manduca sexta,* a species that also has a pupal diapause the photosensitive stage is much longer; it begins during embryonic development and continues through the feeding phase of the fifth instar. In the silkworm, *B. mori,* embryonic diapause is programmed during the mother's period of embryonic development. This timing of the photosensitive stage thus facilitates the channeling of development toward diapause at an early stage and allows sufficient time for the preparative phase of diapause.

The duration of diapause, often called diapause depth, also may depend on photoperiod. For example, in the lacewing, *C. carnea,* diapause depth is controlled by photoperiod in such a way that the adult diapause is deeper when it is induced earlier in the autumn, thus preventing an untimely termination of diapause before the onset of winter. And, in the tobacco hornworm, *M. sexta,* the duration of pupal diapause is a function of the number of short days the embryo and larva have received. Exposure to a few short days, such as would occur in mid- to late summer, results in a long diapause, while exposure exclusively to short days, an event that could occur only in early autumn, results in a short diapause. Such qualitative responses to photoperiod allow the insect to fine tune its development to fit the changing season.

Photoperiodic information is perceived through a receptor in the brain, integrated and stored in the brain, and then translated into the endocrine events that control the induction and maintenance of diapause. The location of the photoreceptor responsible for measurement of daylength has been studied in relatively few insects, but in most of them the compound eyes and ocelli are not the conduit for this information. Surgical destruction of these visual centers or coating the eyes with an opaque paint usually does not interfere with the photoreception involved in the programming of diapause. The photoperiodic signal appears to impinge directly on the brain, but the exact location of these extraretinal photoreceptors has not been elucidated. As in many other plants and animals, the photoperiodic response in insects is primarily a blue-light response. Cryptochromes, proteins involved in photoperiodic responses in a diverse array of organisms, are present in insects and are likely to be implicated in this response. Several important clock genes have been identified in insects, but thus far their involvement in photoperiodism has not been well established.

The role for photoperiod in the environmental regulation of diapause is mainly in the inductive phase of diapause. There are a few species that use daylength as a direct environmental cue for diapause termination. More commonly, photoperiod may influence the rate of diapause development, which in turn does impact the duration of diapause, but frequently diapause development proceeds at a rate determined by temperature rather than photoperiod.

Temperature

Temperature provides another important seasonal cue for diapause induction, but the daily fluctuations in temperature mean that it is less reliable than photoperiod in this regard. Frequently, a short-day response is enhanced by low temperature. For example, the maximum diapause response observed for flesh flies shown in Fig. 1 is approximately 80%. But, this was for flies reared at 25°C, and if the temperature is lowered to 18°C, the diapause incidence is elevated to nearly 100%. In these flies the critical photoperiod is not influenced by temperature, but in some insects the critical photoperiod may shift as well.

Near the equator, where seasonal changes in daylength are too subtle to be used as environmental cues, temperature may replace photoperiod as the primary environmental regulator of diapause, as it does for flesh flies living in East Africa: daylength has no influence on the expression of diapause, but instead low daytime temperatures experienced in July and August are used to program the flies for pupal diapause.

A period of chilling may be essential for diapause termination. Diapausing insects often cannot resume development or reproduction immediately upon transfer to favorable conditions but require a period of chilling. Although some insects do not absolutely require chilling before initiating development many will terminate diapause more quickly if they have first been chilled for a few months.

HORMONAL REGULATION

The juvenile hormones (JHs) and ecdysteroids, two of the major families of insect hormones that direct insect development, metamorphosis, and reproduction, are intimately involved in regulating diapause. The JHs, which are isoprenoids secreted by the corpus allatum (CA), maintain the

juvenile characters during the premetamorphic molts, while the steroid hormones from the prothoracic gland (PG), ecdysone and related compounds, dictate the decision to molt. In turn, the CA and PG are regulated by both neural and humoral factors from the brain. Brain neuropeptides governing the CA can exert either a stimulatory (allatotropins) or an inhibitory action (allatoinhibins) on the CA. The dominant regulator of the PG is the brain neuropeptide prothoracicotropic hormone (PTTH). These hormones, together with diapause hormone, a unique neuropeptide that regulates the embryonic diapause of the commercial silkworm, are the key hormonal regulators of insect diapause. In certain situations, the presence of one or more of these hormones promotes diapause, while in others it is the absence of a certain hormone that causes diapause.

Embryonic Diapause

The best understood hormonal mechanism regulating embryonic diapause is based on the silkworm. In this species diapause intercedes early during embryogenesis, just before segmentation. The developmental fate of the embryo is determined by the presence or absence of diapause hormone (DH), a neuropeptide secreted by the mother's subesophageal ganglion. In the presence of DH, the ovariole produces eggs that enter diapause, and when the hormone is not present the eggs develop without the interruption of diapause. Whether the mother releases DH is dependent upon the photoperiod she was exposed to as an embryo. Thus, the mother's photoperiodic history dictates whether she will release the DH needed to influence the diapause fate of her progeny.

The structure of DH has been defined, as well as the sequence of the cDNA that encodes the peptide. DH appears to exert its effect on diapause by influencing carbohydrate metabolism. In the presence of DH, the developing oocytes incorporate glycogen stores, which in turn are converted to sorbitol. Sorbitol was originally thought to function simply as a cryoprotectant, but recent work suggests that sorbitol may actually be involved in shutting down development in the embryo. The addition of sorbitol to an embryo that is programmed to develop without diapause elicits a developmental arrest; in contrast, the removal of sorbitol from diapause-programmed embryos enables the embryos to develop without diapause.

No other diapauses appear to rely on DH for diapause regulation. It appears to be a hormonal regulator unique to the silkworm. In the gypsy moth diapause occurs at the end of embryogenesis, just before hatching of the first instar. The diapause of this species appears to be regulated by maintenance of a high ecdysteroid titer. As long as the ecdysteroid titer remains high, the pharate first instar remains locked in diapause. Only when the ecdysteroid titer drops in the spring is the gypsy moth free to terminate its diapause and hatch. Yet another mechanism seems to operate in the giant silkmoth, *Antheraea yamamai.* In this insect, an unidentified repressive factor from the mesothorax inhibits the action of a maturation factor from the abdomen. The fact that all three species that have been examined display different endocrine control mechanisms suggests a wealth of mechanisms operating in the regulation of these early stage diapauses.

Larval Diapause

Larval diapause frequently intercedes at the end of larval life, just before the onset of pupation and metamorphosis, but it is not at all uncommon in earlier instars as well. Common to most examples of larval diapause is a shutdown in the brain–prothoracic gland axis. In the absence of ecdysteroids from the PG, the larva fails to initiate the next molt. The failure of the brain to release ecdysteroids can usually be directly attributed to the brain's failure to release PTTH. In a number of species, JH may also play a role. For example, in the southwestern corn borer, *D. grandiosella,* the JH titer remains elevated throughout diapause, and the diapause can be terminated only when the JH titer drops. In some other species such as the European corn borer, *O. nubilalis,* the JH titer is high in early diapause but then declines and remains low throughout the remainder of diapause. No role for JH is apparent in several other insects: the larval diapause of both the parasitic wasp, *Nasonia vitripennis,* and the blow fly, *Calliphora vicina,* can be explained strictly as an ecdysteroid deficiency.

Pupal Diapause

Pupal diapause is the consequence of a shutdown in the brain–prothoracic gland axis. Thus, in the absence of ecdysteroids from the PG the progression of adult differentiation is halted. At the termination of diapause ecdysteroids are again released, triggering adult development. In *H. cecropia* a period of chilling is required before the brain can stimulate the PG to release ecdysteroids. Pupal diapauses can usually be quickly terminated with an injection of 20-hydroxyecdysone. Usually the absence of ecdysteroids can be attributed directly to a failure of the brain to release the neuropeptide PTTH needed to stimulate the PG to synthesize ecdysteroids, but in some insects, e.g., *Heliothis zea,* PTTH is released shortly after pupation, but pupa fail to develop until the PG has been chilled adequately.

Unlike larval diapause there is no evidence suggesting that JH regulates pupal diapause induction or termination, yet JH is indeed present during pupal diapause in some species. In flesh flies, cycles of JH activity apparently drive infradian cycles (4-day periodicity at 25°C) of metabolic activity that persist throughout diapause.

Adult Diapause

A shutdown in JH synthesis is a key feature in the regulation of adult diapause. The corpora allata, the endocrine glands that synthesize and release JH, are characteristically small during

diapause. Application of exogenous JH or implantation of active corpora allata into a diapausing individual usually prompts the termination of diapause. Conversely, the surgical extirpation of the corpora allata from a nondiapausing adult causes the adult to enter a diapause-like state. Measurement of the JH titer also supports the idea that adult diapause is the consequence of a shutdown of the corpora allata: the titer of JH typically drops as the insect enters diapause and increases again when diapause is terminated.

It is the brain that regulates the corpora allata, and both nervous and humoral pathways are involved in its regulation. In the Colorado potato beetle, the brain exerts its control over the corpora allata by a humoral mechanism, but in the linden bug, *Pyrrhocoris apterus,* nervous control is also involved. Ecdysteroids may also be involved in some species. The ecdysteroid titer is nearly twice as high in Colorado potato beetles destined for diapause than in those that are not destined to enter diapause, and an injection of ecdysteroids can terminate adult diapause in *Drosophila melanogaster.*

MOLECULAR MECHANISMS

The environmental cues that regulate diapause have been well defined, and there is also a fairly good understanding of the downstream hormonal signals that serve to coordinate diapause. But, the molecular underpinning of diapause remains poorly understood. Is diapause simply a shutdown in gene expression or does it represent the expression of a unique set of genes? An examination of the synthesis of brain proteins in flesh flies suggests that far fewer proteins are synthesized in the brain during diapause but, in addition, the brains of diapausing flies synthesize a set of proteins not observed in brains of nondiapausing flies. This suggests that diapause represents both a shutdown in gene expression and the expression of a unique set of genes.

One of the most conspicuous groups of genes that are diapause up-regulated is that of the heat-shock proteins. Both heat-shock protein 70 (Hsp70) and one of the small heat-shock proteins (Hsp23) are up-regulated in flesh flies during diapause. The Hsps are up-regulated upon entry into diapause, remain elevated throughout diapause, and then drop sharply at diapause termination. But, not all heat-shock proteins are up-regulated during diapause. Hsp90, by contrast, is actually down-regulated. The Hsps may offer protection from environmental stresses during diapause and possibly contribute to the cell cycle arrest that characterizes diapause.

Genes that are diapause down-regulated are potentially of equal interest. Among the genes in this category is the gene that encodes proliferating cell nuclear antigen, a cell cycle regulator. The down-regulation of this gene during diapause may be important in bringing about the cell cycle arrest. As more genes are examined, it is evident that certain genes are expressed throughout diapause, others are turned off during diapause, while still others are expressed only during early or late diapause or may be expressed intermittently during diapause.

It is still too early to know if common sets of genes are expressed during diapause in different species and different life stages, but preliminary data suggest that the expression patterns of at least some of the genes, those that encode Hsp70, may be shared across species and life stages.

See Also the Following Articles

Aestivation • Bombyx mori • Cold/Heat Protection • Dormancy • Juvenile Hormone

Further Reading

Danks, H. V. (1987). "Insect Dormancy: An Ecological Perspective." Biological Survey of Canada, Ottawa.

Denlinger, D. L. (1972). Induction and termination of pupal diapause in *Sarcophaga* flesh flies. *Biol. Bull.* **142,** 11–24.

Denlinger, D. L. (1985). Hormonal control of diapause. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 8, pp. 353–412. Pergamon Press, Oxford.

Denlinger, D. L. (2002). Regulation of diapause. *Annu. Rev. Entomol.* **47**, 93–122.

Denlinger, D. L., Giebultowicz, J. M., and Saunders, D. S. (eds.) (2001). "Insect Timing: Circadian Rhythmicity to Seasonality." Elsevier, Amsterdam.

Lee, R. E., Jr., and Denlinger, D. L. (eds.) (1991). "Insects at Low Temperature." Chapman & Hall, New York.

Saunders, D. S. (2002). "Insect Clocks," 3rd edition. Elsevier, Amsterdam.

Tauber, M. J., Tauber, C. A., and Masaki, S. (1986). "Seasonal Adaptations of Insects." Oxford University Press, Oxford.

Digestion

Walter R. Terra
University of São Paulo, Brazil

Digestion is the process by which food molecules are broken down into smaller molecules that are able to be absorbed by the gut tissue. Most food molecules requiring digestion are polymers such as proteins and starch, and are sequentially digested through three phases (Fig. 1). Primary digestion is the dispersion and reduction in molecular size of the polymers and results in oligomers. During intermediate digestion, these undergo a further reduction in molecular size to dimers, which in final digestion form monomers. Digestion usually occurs under the action of digestive enzymes from the midgut, with minor or no participation of salivary enzymes. In most insects, midgut pH is either mildly acidic or neutral. Lepidopteran and trichopteran larvae, scarabaeid beetles, and nematoceran flies have alkaline midguts, whereas cyclorrhaphous flies have a very acidic section in the middle of the midgut. The midgut is, as a rule, an oxidizing site, although in some wool-digesting insects it is a reducing site, a condition necessary to break disulfide bonds in keratin, thus facilitating enzymatic hydrolysis.

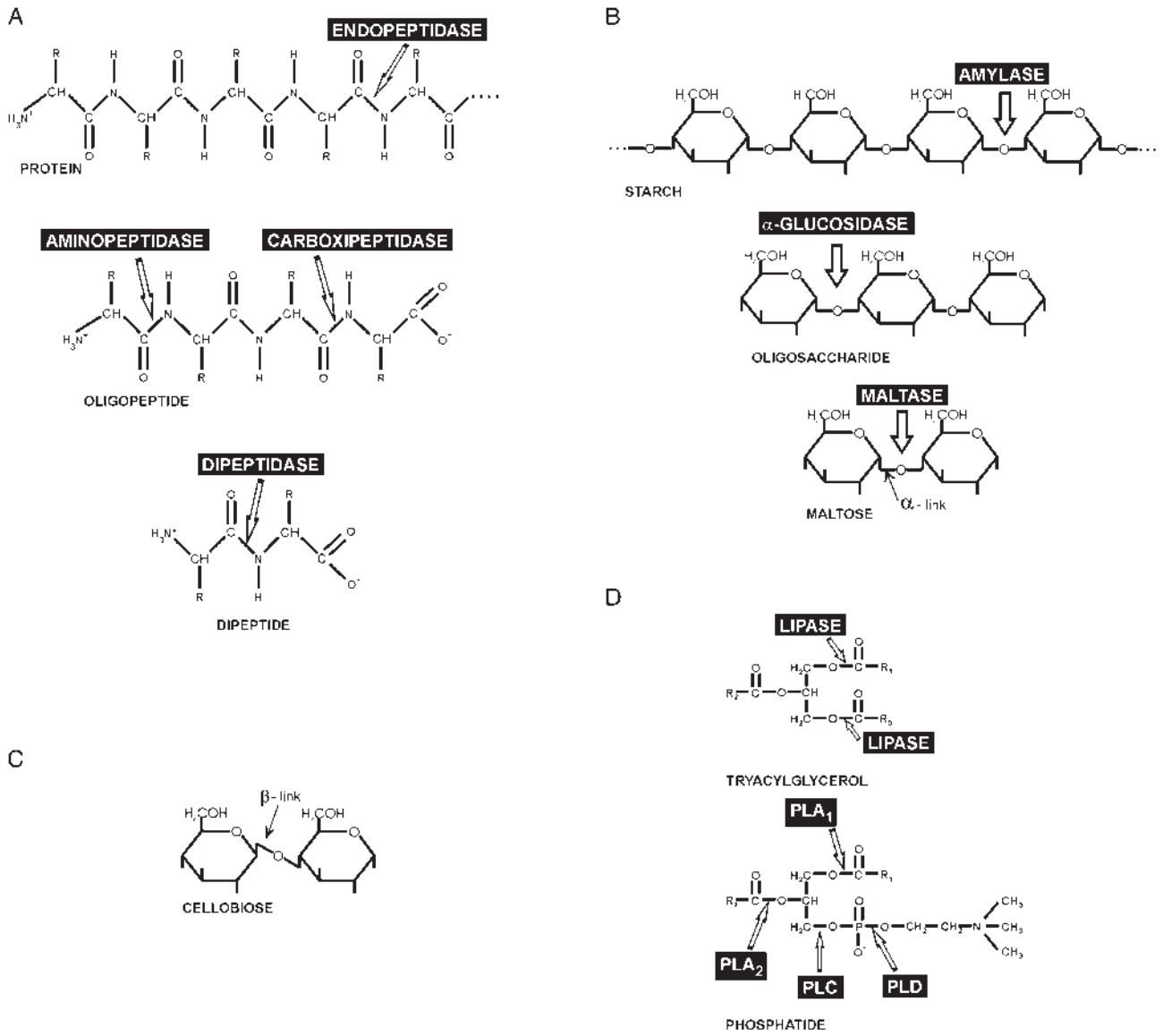

FIGURE 1 Digestion of important nutrient classes. Arrows point to bonds cleaved by enzymes. (A) Protein digestion; R, different amino acid moieties. (B) Starch digestion. (C) β-linked glucoside. (D) Lipid digestion; PL, phospholipase; R, fatty acyl moieties.

DIGESTION OF PROTEINS

Initial digestion of proteins is carried out by proteinases (endopeptidases), which are enzymes able to cleave the internal peptide bonds of proteins (Fig. 1A). Different endopeptidases are necessary to do this because the amino acid residues vary along the peptide chain (R is a variable group in Fig. 1A). Proteinases may differ in specificity toward the reactant protein (substrate) and are grouped according to their reaction mechanism into the subclasses: serine, cysteine, and aspartic proteinases. Trypsin, chymotrypsin, and elastase are serine proteinases that are widely distributed in insects and have molecular masses in the range 20 to 35 kDa and alkaline pH optima. Trypsin preferentially hydrolyzes (its primary specificity) peptide bonds in the carboxyl end of amino acids with basic R groups (Arg, Lys); chymotrypsin is preferential toward large hydrophobic R groups (e.g., Phe, Tyr) and elastase, toward small hydrophobic R groups (e.g., Ala). The activity of the enzymes also depends on the amino acid residues neighboring the bond to be cleaved. This may explain the differences in susceptibilities of insects to strains of *Bacillus thuringiensis,* because the deleterious effects depend on the previous proteolysis of the bacterial endotoxin. Related to this is the growing evidence that insects fed on trypsin inhibitor-containing food express new trypsin molecules insensitive to the inhibitors. These inhibitors are proteins and their binding to the enzyme has molecular requirements similar to those of the substrate.

Cysteine and aspartic proteinases are the only midgut proteinases in hemipterans and they occur in addition to serine proteinases in cucujiformia beetles. Their occurrence in Hemiptera is interpreted as a consequence of the loss of the usual digestive serine proteinases associated with the adaptation of hemipteran ancestors to a diet lacking proteins (plant sap), followed by the use of lysosome-like enzymes in adapting to a new predatory habit. The presence of cysteine and aspartic proteinases in cucujiformia beetles is likely an ancestral adaptation to circumvent proteolytic inhibition caused by trypsin inhibitors in ingested seeds. Cysteine and aspartic proteinases have pH optima of 5.5 to 6.0 and 3.2 to 3.5 and molecular masses of 20 to 40 kDa and 60 to 80 kDa, respectively. Because of their pH optima, aspartic proteinases are not very active in the mildly acidic midguts of Hemiptera and cucujiformia beetles, but are very important in the middle midguts (pH 3.5) of cyclorrhaphous flies.

Intermediate digestion of proteins is accomplished by exopeptidases, enzymes that remove amino acids from the N-terminal (aminopeptidases) or C-terminal (carboxypeptidases) ends of oligopeptides (fragments of proteins) (Fig. 1A). Insect aminopeptidases have molecular masses in the range 90 to 130 kDa, have pH optima of 7.2 to 9.0, have no marked specificity toward the N-terminal amino acid, and are usually associated with the microvillar membranes of midgut cells. Therefore, the action of aminopeptidase is restricted to the surface of midgut cells. Because aminopeptidases are frequently active on dipeptides, they are also involved in protein-terminal digestion together with dipeptidases. Aminopeptidases may account for as much as 55% of the midgut microvillar proteins in larvae of the yellow mealworm, *Tenebrio molitor.* Probably because of this, in many insects aminopeptidases are the preferred targets of *B. thuringiensis* endotoxins. These toxins, after binding to aminopeptidase (or other receptors), form channels through which cell contents leak, leading to insect death. The most important insect carboxypeptidases have alkaline pH optima, have molecular masses in the range 20 to 50 kDa, and require a divalent metal for activity. They are classified as carboxypeptidase A or B depending on their activity upon neutral/acid or basic C-terminal amino acids, respectively.

DIGESTION OF CARBOHYDRATES

Initial and intermediate digestion of starch (or glycogen) is accomplished by α-amylase. This enzyme cleaves internal bonds of the polysaccharide until it is reduced to small oligosaccharides or disaccharides (Fig. 1B). The amylases are not very active in digesting intact starch granules, making mastication prior to ingestion important. Insect amylases depend on calcium ions for activity or stability, they are activated by chloride ions (amylases in Lepidoptera are exceptions), their molecular masses are found in the range 48 to 68 kDa, and their pH optima vary widely (4.8–9.8) depending on the insect taxon. As described for trypsin, insects feeding on amylase inhibitor-containing food express new amylase molecules insensitive to the inhibitors.

The final digestion of starch chains occurs under α-glucosidases, enzymes that sequentially remove glucosyl residues from the nonreducing ends of short oligomaltosaccharides. If the saccharide is a disaccharide, it is named maltose (Fig. 1B). Because of that, α-glucosidase is also called maltase. As a rule, sucrose (glucose α1,β2-fructose) is hydrolyzed by α-glucosidase. If an enzyme is able to hydrolyze sucrose, but not maltose, it is likely a β-fructosidase, an enzyme attacking sucrose by the fructosyl residue. Sucrose is found in large amounts in nectar and phloem sap and in lesser amounts in some fruits and leaves.

The important insect hemolymph and fungal sugar trehalose (glucose α1,α1-glucose) is hydrolyzed only by the specific enzyme trehalase. This digestive enzyme occurs in luminal contents or immobilized at the surface of midgut cells and also as an enzyme at the midgut basal cell membrane, making available glucose from hemolymph trehalose.

Although cellulose is abundant in plants, most plant-feeding insects, such as caterpillars and grasshoppers, do not use it. Cellulose is a nonramified chain of glucose units linked by β-1,4 bonds (Fig. 1C) arranged in a crystalline structure that is difficult to disrupt. Thus, cellulose digestion is unlikely to be advantageous to an insect that can meet its dietary requirements using more easily digested food constituents. The cellulase activity found in some plant feeders facilitates the access of digestive enzymes to the plant cells ingested by the insects. True cellulose digestion is restricted to insects that have, as a rule, nutritionally poor diets, as exemplified by termites, woodroaches, and cerambycid and scarabaeid beetles. There is growing evidence that insects secrete enzymes able to hydrolyze crystalline cellulose, challenging the long-standing belief that microbial symbionts are necessary for cellulose digestion. The end products of cellulase action are glucose and cellobiose (Fig. 1C); the latter is hydrolyzed by a β-glucosidase.

Hemicellulose is a mixture of polysaccharides associated with cellulose in plant cell walls. They are β-1,4- and/or β-1,3-linked glycan chains made up mainly of glucose (glucans), xylose (xylans), and other monosaccharides. The polysaccharides are hydrolyzed by a variety of enzymes from which xylanases, laminarinases, and lichenases are the best known. The end products of the actions of these enzymes are monosaccharides and β-linked oligosaccharides. The final digestion of those chains occurs under the actions of β-glycosidases that sequentially remove glycosyl (usually glucosyl, galactosyl, or xylosyl) residues from the nonreducing end of the β-linked oligosaccharides. As these may be cellobiose, β-glycosidase is frequently also named cellobiase. Thus, β-glycosidases complete the digestion of cellulose and hemicelluloses.

A special β-glycosidase (aryl β-glycosidase) acts on glycolipids and *in vivo* probably removes a galactose from

monogalactosyldiacylglycerol that together with digalactosyldiacylglycerol is a major lipid of photosynthetic tissues. Digalactosyldiacylglycerol is converted into monogalactosyldiacylglycerol by the action of an α-galactosidase. The aryl β-glycosidase also acts on plant glycosides that are noxious after hydrolysis. Insects circumvent these problems by detoxifying the products of hydrolysis or by repressing the synthesis and secretion of this enzyme while maintaining constant the synthesis and secretion of the other β-glycosidases.

DIGESTION OF LIPIDS AND PHOSPHATES

Oils and fats are triacylglycerols and are hydrolyzed by a triacylglycerol lipase that preferentially removes the outer ester links of the substrate (Fig. 1D) and acts only on the water–lipid interface. This interface is increased by surfactants that, in contrast to the bile salts of vertebrates, are mainly lysophosphatides. The resulting 2-monoacylglycerol may be absorbed or further hydrolyzed before absorption.

Membrane lipids include glycolipids, such as galactosyldiacylglycerol and phosphatides. After the removal of galactose residues from mono- and digalactosyldiacylglycerol, which leaves diacylglycerol, it is hydrolyzed as described for triacylglycerols. Phospholipase A removes one fatty acid from the phosphatide, resulting in a lysophosphatide (Fig. 1D) that forms micellar aggregates, causing the solubilization of cell membranes. Lysophosphatide seems to be absorbed intact by insects.

Nonspecific phosphatases remove phosphate moieties from phosphorylated compounds to make their absorption easier. Phosphatases are active in an alkaline or acid medium.

See Also the Following Articles

Metabolism • Nutrition • Phytophagous Insects • Salivary Glands

Further Reading

Cristofoletti, P. T., and Terra, W. R. (1999). Specificity, anchoring and subsites in the active center of a microvillar aminopeptidase purified from *Tenebrio molitor* (Coleoptera) midgut cells. *Insect Biochem. Mol. Biol.* **29,** 807–819.

Kerkut, G. A., and Gilbert, L. I. (eds.) (1985). "Comprehensive Insect Physiology, Biochemistry and Pharmacology." Pergamon Press, Oxford. [See especially Vol. 4, Chaps. 5 and 7]

Lehane, M. J., and Billingsley, P. F. (eds.) (1996). "Biology of the Insect Midgut." Chapman & Hall, London. [See especially Chaps. 3, 6, 7, 11, and 14]

Silva, C. P., Terra, W. R., de Sa, M. F. G., Samuels, R. I., Isejima, E. M., Bifano, T. D., and Almeida, J. S. (2001). Induction of digestive alpha-amylases in larvae of *Zabrotes subfasciatus* (Coleoptera: Bruchidae) in response to ingestion of common bean alpha-amylase inhibitor 1. *J. Insect Physiol.* **47,** 1283–1290.

Terra, W. R., and Ferreira, C. (1994). Insect digestive enzymes: Properties, compartmentalization and function. *Comp. Biochem. Physiol.* **109B,** 1–62.

Vonk, H. J., and Western, J. R. H. (1984) "Comparative Biochemistry and Physiology of Enzymatic Digestion." Academic Press, London.

Watanabe, H., Noda, H., Tokuda, G., and Lo, N. (1998). A cellulase gene of termite origin. *Nature* **394,** 330–331.

Digestive System

Walter R. Terra and Clélia Ferreira

University of São Paulo, Brazil

The digestive system consists of the alimentary canal (gut) and salivary glands, and is responsible for all steps in food processing: digestion, absorption, and feces delivery and elimination. These steps occur along the gut. The anterior (foregut) and posterior (hindgut) parts of the gut have cells covered by a cuticle whereas, in the midgut, cells are separated from the food by a filmlike anatomical structure referred to as the peritrophic membrane. Salivary glands are associated with the foregut and may be important in food intake but usually not in digestion. Remarkable adaptations are found in taxa with very specialized diets, such as cicadas (plant sap), dung beetles (feces), and termites (wood), and in insects with short life spans, as exemplified by flies and moths. Digestion is carried out by insect digestive enzymes, apparently without participation of symbiotic microorganisms.

GUT MORPHOLOGY AND FUNCTION

Figure 1 is a generalized diagram of the insect gut. The foregut begins at the mouth, includes the cibarium (preoral cavity formed by mouthparts), the pharynx, the esophagus, and the crop (a dilated portion, as in Fig. 2A, or a diverticulum, like Fig. 2K). The crop is a storage organ in many insects and also serves as a site for digestion in others. The foregut is lined by a cuticle that is nonpermeable to hydrophilic molecules and in some insects is reduced to a straight tube (Fig. 2F). The proventriculus is a triturating (grinding into fine particles) organ in some insects, and in most it provides a valve controlling the entry of food into the midgut, which is the main site of digestion and absorption of nutrients.

The midgut includes a simple tube (ventriculus) from which blind sacs (gastric or midgut ceca) may branch, usually from its anterior end (Fig. 2A). Midgut ceca may also occur along the midgut in rings (Fig. 2F) or not (Fig. 2H) or in the

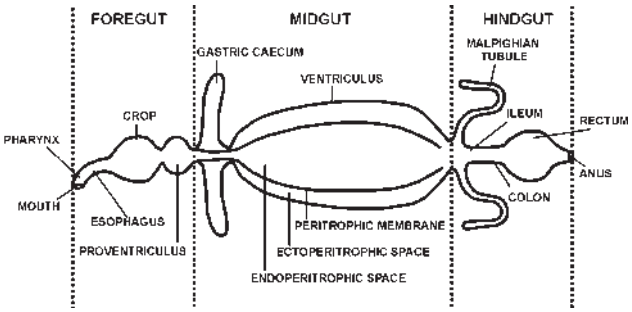

FIGURE 1 Generalized diagram of the insect gut.

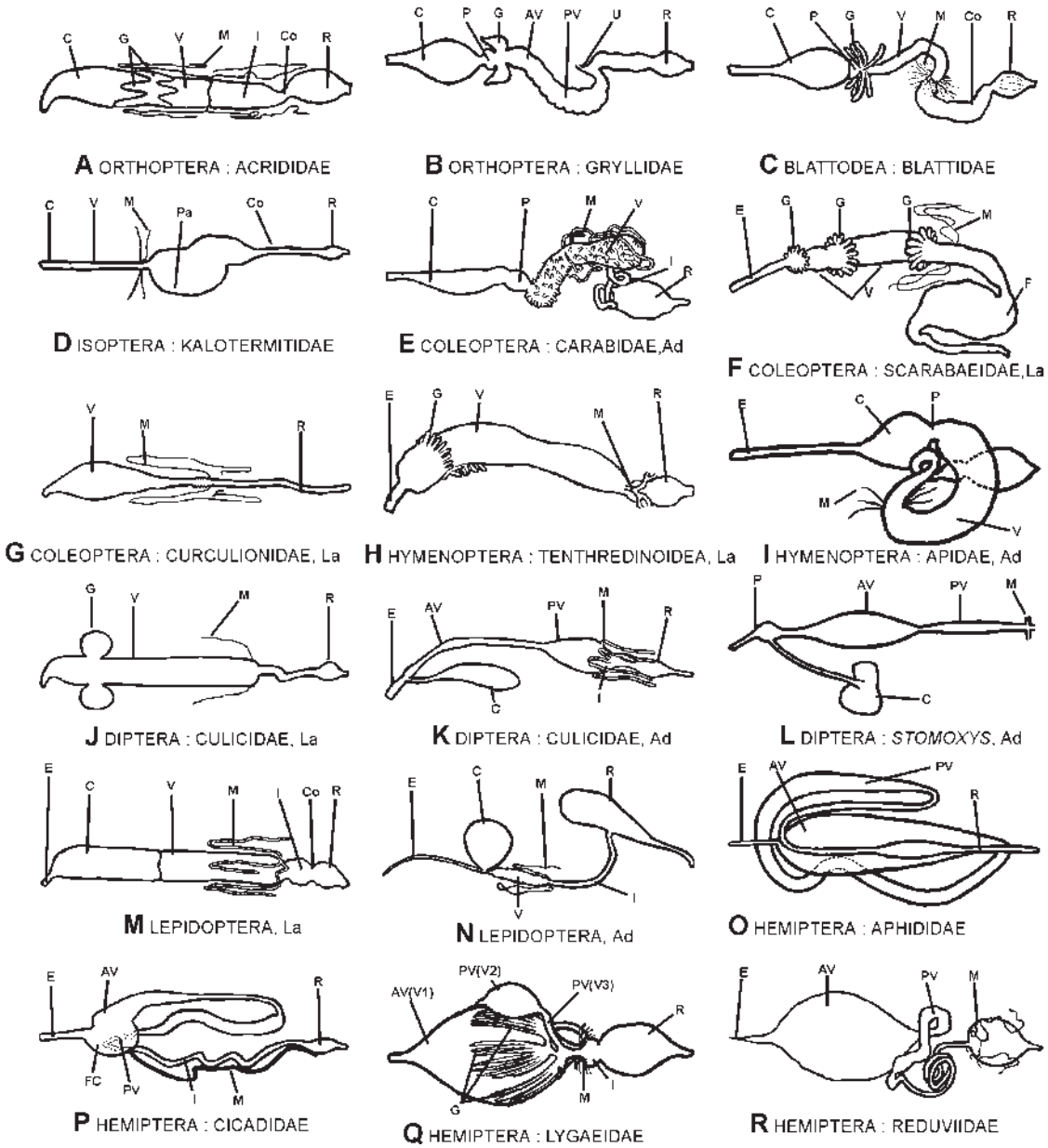

FIGURE 2 Major insect gut types: Ad, adult; AV, anterior ventriculus (midgut); C, crop; Co, colon; E, esophagus; F, fermentation chamber; FC, filter chamber; G, midgut (gastric) ceca; I, ileum; La, larva; M, Malpighian tubules; P, proventriculus; Pa, paunch; PV, posterior ventriculus (midgut); R, rectum; V, ventriculus. Not drawn to scale. [Based partly on Terra, W. R. (1988). Physiology and biochemistry of insect digestion: An evolutionary perspective. *Brazilian J. Med. Biol. Res.* **21,** 675–734.]

posterior midgut (Fig. 2Q). In most insects, the midgut is lined with a filmlike anatomical structure (peritrophic membrane) that separates the luminal contents into two compartments: the endoperitrophic space (inside the membrane) and the ectoperitrophic space (outside the membrane). Some insects have a stomach, which is an enlargement of the midgut to store food (Fig. 2R). In the region of the sphincter (pylorus) separating the midgut from the hindgut, Malpighian tubules branch off the gut. Malpighian tubules are excretory organs that individually empty in the gut and may be joined to form a ureter (Fig. 2B); in some species, however, they are absent (Fig. 2O).

The hindgut includes the ileum, colon, and rectum (which is involved in the absorption of water and ions) and terminates with the anus. The hindgut is lined by a cuticle (usually impermeable); although in some insects it is reduced to a straight tube (Fig. 2G), in others it is modified in a fermentation chamber (Fig. 2F) or paunch (Fig. 2D), with both structures storing ingested food and harboring microorganisms that have a controversial role in assisting cellulose digestion.

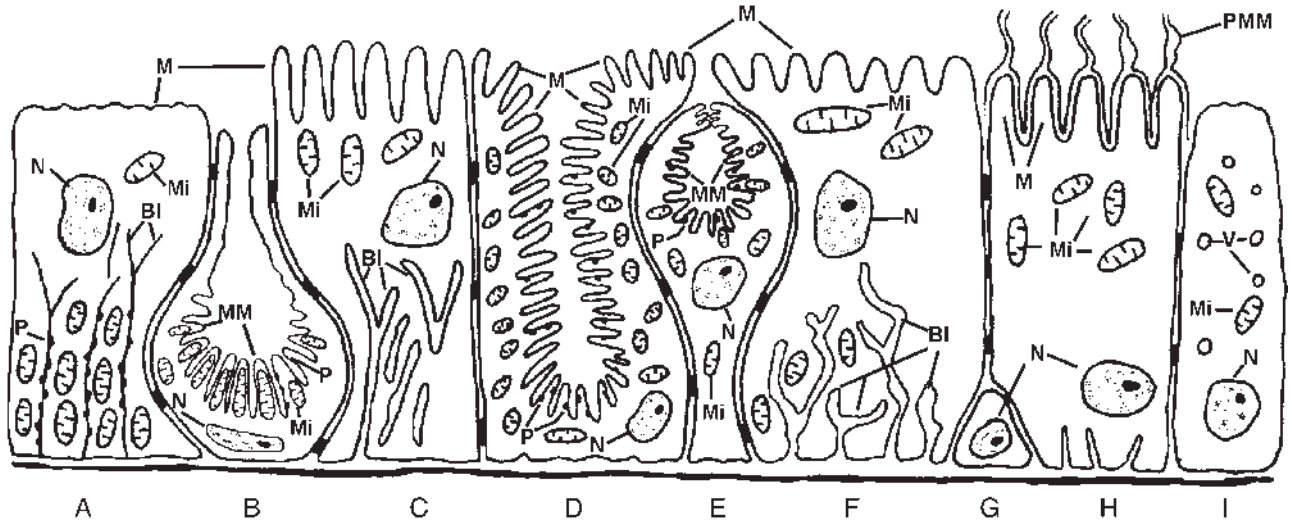

FIGURE 3 Diagrammatic representation of typical insect midgut cells: (A) columnar cell with plasma membrane infoldings arranged in long and narrow channels, usually occurring in fluid-absorbing tissues; (B) lepidopteran long-necked goblet cell; (C) columnar cell with highly-developed basal plasma membrane infoldings displaying few openings into the underlying space, usually occurring in fluid-absorbing tissue; (D) cyclorrhaphan dipteran oxyntic (cuprophilic) cell; (E) lepidopteran stalked goblet cell; (F) columnar cell with highly developed plasma membrane infoldings with numerous openings into the underlying space, frequently present in fluid-secreting tissue; (G) regenerative cell; (H) hemipteran midgut cell; (I) endocrine cell. Note particles (portasomes) studding the cytoplasmic side of the apical membranes in B, D, and E and of the basal plasma membranes in A. Abbreviations: Bl, basal plasma membrane infoldings; M, microvilli; Mi, mitochondria; MM, modified microvilli; N, nucleus; P, portasomes; PMM, perimicrovillar membranes; V, vesicles.

The gut epithelium is always simple and rests on a basal lamina that is surrounded by conspicuous circular and a few longitudinal muscles, the organization of which varies according to species. Wavelike contractions of the circular muscles cause peristalsis, propelling the food bolus along the gut. The gut is oxygenated by the tracheal system, and whereas the foregut and hindgut are well innervated, the same is not true for the midgut. The gut is also connected to the body wall through the extrinsic visceral muscles. These act as dilators of the gut, mainly at the foregut, where they form a pump highly developed in fluid feeders (cibarium pump), exemplified by sap (Hemiptera) and blood (Hemiptera and Diptera) feeders. However it is also present in chewing insects (pharyngeal pump), which are thus enabled to drink water and to pump air into the gut during the molts. The gut sensory system includes the chemoreceptors in the cibarium and stretch receptors associated with muscles of the foregut and hindgut.

Salivary glands are labial or mandibular glands opening in the cibarium. They are usually absent in Coleoptera. The saliva lubricates the mouthparts, may contain an array of compounds associated with blood intake, or may be used as a fixative of the stylets of sap-sucking bugs. Saliva usually contains only amylase and maltase or no enzymes at all, although in a few hemipteran predators it may have the whole complement of proteolytic enzymes.

The epithelium of the midgut is composed of a major type of cell usually named columnar cell, although it may have other forms (Fig. 3A, C, F); it also contains regenerative cells (Fig. 3G) that are often collected together in nests at the base of the epithelium, cells (Fig. 3I) whose purpose is not understood but are generally believed to have an endocrine function, and also specialized cells (goblet cells, Fig. 3B, E; oxyntic cells, Fig. 3D; hemipteran midgut cell, Fig. 3H).

The peritrophic membrane is made up of a matrix of proteins (peritrophins) and chitin to which other components (e.g., enzymes, food molecules) may associate. This anatomical structure is sometimes called the peritrophic matrix, but this term is better avoided because it does not convey the idea of a film and suggests that it is the fundamental substance of some structure. The argument that "membrane" means a lipid bilayer does not hold here because the peritrophic membrane is an anatomical structure, not a cell part. Peritrophins have domains similar to mucins (gastrointestinal mucus proteins) and other domains able to bind chitin. This suggests that the peritrophic membrane may have derived from an ancestral mucus. According to this hypothesis, the peritrophins evolved from mucins by acquiring chitin-binding domains. The parallel evolution of chitin secretion by midgut cells led to the formation of the chitin–protein network characteristic of the peritrophic membrane. The details of peritrophic membrane formation are not known, although there is evidence that peritrophins are released by exocytosis (Fig. 4A) in Diptera or by microaprocrine secretion (Fig. 4D) in Lepidoptera and somehow interlocked with chitin fibers that are synthesized at the luminal surface of midgut cells.

The formation of the peritrophic membrane may occur in part of the midgut or in the entire organ (type I), or only at the entrance of the midgut (cardia) (type II). The two types

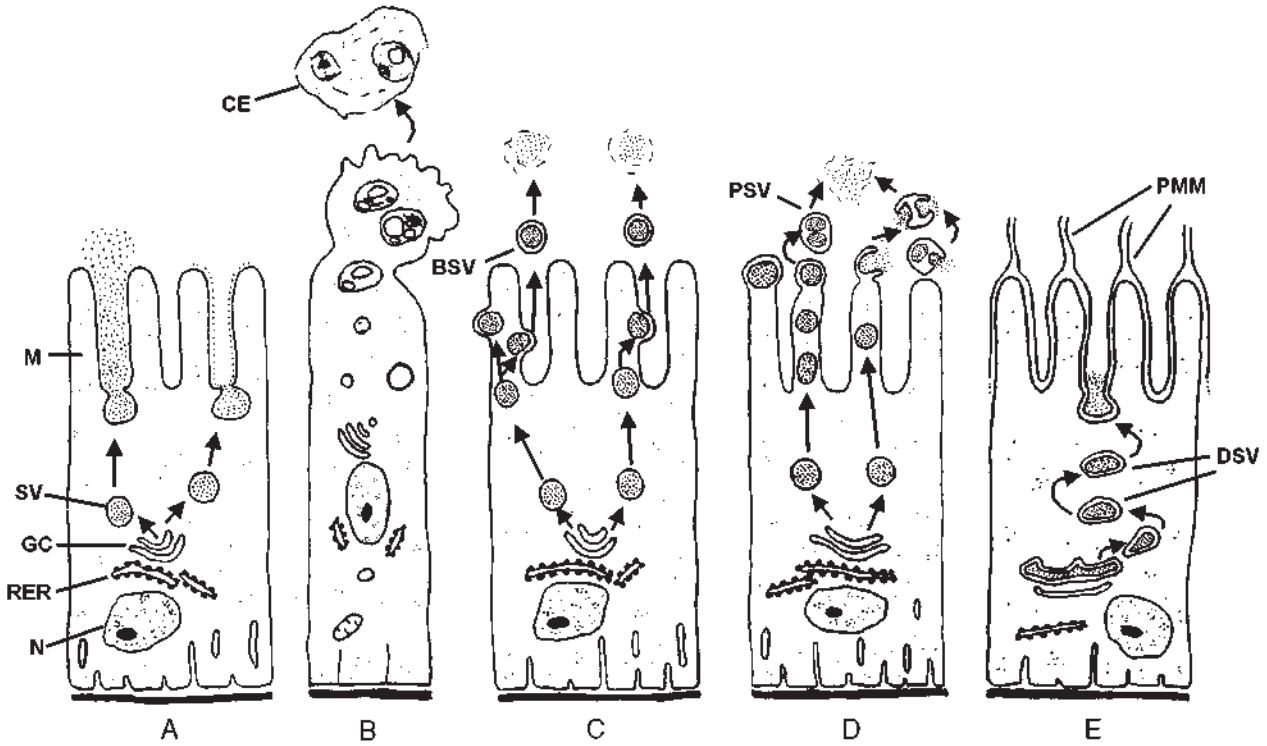

FIGURE 4 Models for secretory processes of insect digestive enzymes; (A) exocytic secretion, (B) apocrine secretion, (C) microapocrine secretion with budding vesicles, (D) microapocrine secretion with pinched-off vesicles, and (E) modified exocytic secretion in hemipteran midgut cell. Abbreviations: BSV, budding secretory vesicle; CE, cellular extrusion; DSV, double-membrane secretory vesicle; GC, Golgi complex; M, microvilli; N, nucleus, PMM, perimicrovillar membrane; PSV, pinched-off secretory vesicle; RER, rough endoplasmic reticulum; SV, secretory vesicle.

of membrane differ in their constituent peritrophins and in their supramolecular organization. Type I peritrophic membrane occurs in most insects, whereas type II is restricted to larval and adult (except hematophagous) mosquitoes and flies (Diptera) and a few adult Lepidoptera. Although a peritrophic membrane is found in most insects, it does not occur in Hemiptera and Thysanoptera, which have perimicrovillar membranes in their cells (Fig. 3H). The other insects that do not seem to have a peritrophic membrane are adult Lepidoptera, Phthiraptera, Psocoptera, Zoraptera, Strepsiptera, Raphidioptera, Megaloptera, and Siphonaptera as well as bruchid beetles and some adult ants (Hymenoptera). Most of the pores of the peritrophic membrane are in the range of 7 to 9 nm, although some may be as large as 36 nm. Thus, the peritrophic membrane hinders the free movement of molecules, dividing the midgut lumen into two compartments (Fig. 1) with different molecules. The functions of this structure include those of the ancestral mucus (protection against food abrasion and microorganism invasion) and several roles associated with the compartmentalization of the midgut. These roles result in improvements in digestive efficiency and assist in decreasing digestive enzyme excretion, and in restricting the production of the final products of digestion close to their transporters, thus facilitating absorption.

DIGESTIVE PHYSIOLOGY

Overview

The study of digestive physiology involves the spatial organization of digestive events in the insect gut. Digestive enzymes that participate in primary digestion (cleavage of polymers like protein and starch), secondary digestion (action on oligomers exemplified by polypeptides and dextrans), and final digestion (hydrolysis of dimers as dipeptides and disaccharides) are assayed in different gut compartments. Samples of the ectoperitrophic space contents (Fig. 1) are collected by puncturing the midgut ceca with a capillary or by washing the luminal face of midgut tissue. Midgut tissue enzymes are intracellular, glycocalyx-associated or microvillar membrane-bound. In addition to the distribution of digestive enzymes, the spatial organization of digestion depends on midgut fluxes. Gut fluid fluxes are inferred with the use of dyes. Secretory regions transport injected dye into the gut lumen, whereas absorbing regions accumulate orally fed dyes. Upon studying the spatial organization of the digestive events in insects of different taxa and diets, it was realized that the insects may be grouped relative to their digestive physiology, assuming they have common ancestors. Those putative ancestors correspond to basic gut plans from which groups of insects may have evolved by adapting to different diets.

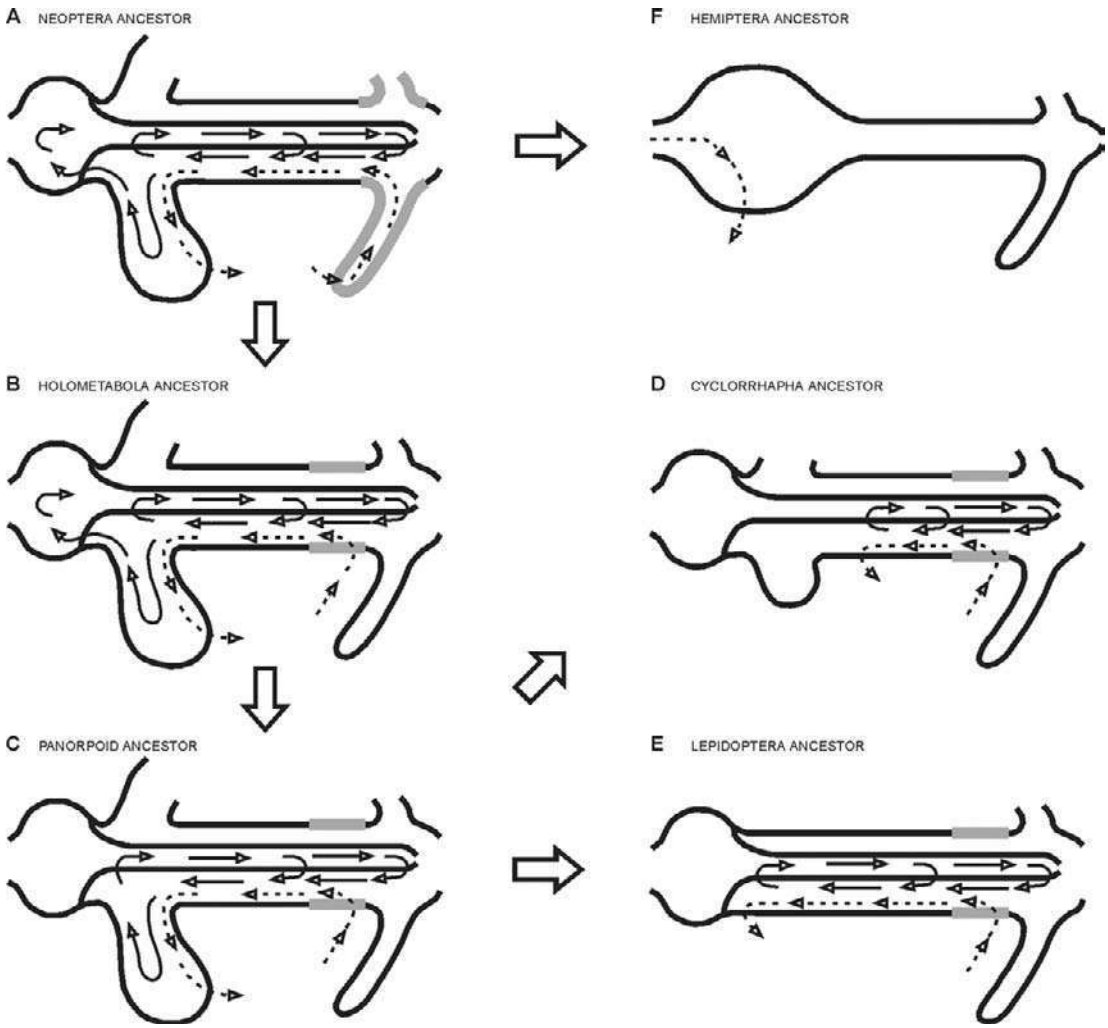

FIGURE 5 Diagrammatic representation of water fluxes (dashed arrows) and of the circulation of digestive enzymes (solid arrows) in putative insect ancestors that correspond to the major basic gut plans. In Neoptera ancestors (A), midgut digestive enzymes pass into the crop. Countercurrent fluxes depend on the secretion of fluid by the Malpighian tubules and its absorption by the ceca. Enzymes involved in initial, intermediate, and final digestion circulate freely among gut compartments. Holometabola ancestors (B) are similar except that secretion of fluid occurs in posterior ventriculus. Panorpoid (Lepidoptera and Diptera assemblage) ancestors (C) display countercurrent fluxes like Holometabola ancestors, midgut enzymes are not found in crop, and only the enzymes involved in initial digestion pass through the peritrophic membrane. Enzymes involved in intermediate digestion are restricted to the ectoperitrophic space and those responsible for terminal digestion are immobilized at the surface of midgut cells. Cyclorrhapha ancestors (D) have a reduction in ceca, absorption of fluid in middle midgut, and anterior midgut playing a storage role. Lepidoptera ancestors (E) are similar to panorpoid ancestors, except that anterior midgut replaced the ceca in fluid absorption. Hemiptera ancestors (F) lost crop, ceca, and fluid-secreting regions. Fluid is absorbed in anterior midgut.

Neopteran insects evolved along three lines: the Polyneoptera (which include Blattodea, Isoptera, and Orthoptera), the Paraneoptera (which include Hemiptera), and the Holometabola (which include Coleoptera, Hymenoptera, Diptera, and Lepidoptera). Polyneoptera and Paraneoptera evolved as external feeders occupying the ground surface, on vegetation, or in litter, and developed distinct feeding habits. Some of these habits are very specialized (e.g., feeding wood and sucking plant sap), implying adaptative changes of the digestive system. Major trends in the evolution of Holometabola were the divergence in food habits between larvae and adults and the exploitation of new food sources, exemplified by endoparasitism and by boring or mining living or dead wood, foliage, fruits, or seeds. This biological variation was accompanied by modifications in the digestive system. Among the panorpoid Holometabola (an assemblage that includes Diptera and Lepidoptera), new selective pressures resulted from the occupation of more exposed or ephemeral ecological niches. Following this trend, those pressures led to shortening life spans, so that the insects may have more generations per year, thus ensuring species survival even if large mortality occurs at each generation. Associated with this trend, the digestive system evolved to become more efficient to support faster life cycles.

The basic plan of digestive physiology for most winged insects (Neoptera ancestors) is summarized in Fig. 5A. In

these ancestors, the major part of digestion is carried out in the crop by digestive enzymes propelled by antiperistalsis forward from the midgut. Saliva plays a minor role or no role at all in digestion. After a while, following ingestion, the crop contracts, transferring digestive enzymes and partly digested food into the ventriculus. The anterior ventriculus is acid and has high carbohydrase activity, whereas the posterior ventriculus is alkaline and has high proteinase activity. This differentiation along the midgut may be an adaptation to the instability of ancestral carbohydrases in the presence of proteinases. The food bolus moves backward in the midgut of the insect by peristalsis. As soon as the polymeric food molecules have been digested to become small enough to pass through the peritrophic membrane, they diffuse with the digestive enzymes into the ectoperitrophic space (Fig. 1). The enzymes and nutrients are then displaced toward the ceca with a countercurrent flux caused by secretion of fluid at the Malpighian tubules and its absorption back by cells (similar to Figs. 3A, C) at the ceca (Fig. 5A), where final digestion is completed and nutrient absorption occurs. When the insect starts a new meal, the ceca contents are moved into the crop. As a consequence of the countercurrent flux, digestive enzymes occur as a decreasing gradient in the midgut, and lower amounts are excreted.

The Neoptera basic plan is the source of that of the Polyneoptera orders and evolved to the basic plans of Paraneoptera and Holometabola. Lack of data limits the proposition of a basic plan to a single Paraneoptera order, Hemiptera. Symbiont microorganisms may occur in large numbers in insect gut. For example, the bacteria *Nocardia rhodnii* may represent up to 5% of the *Rhodnius prolixus* midgut dry weight. The symbionts are believed to provide nutrient factors (such as B vitamins or fermentation end products) to the host. Microorganism symbionts have rarely been associated with digestion, and the few that are known are implicated with cellulose digestion only.

Polyneoptera

BLATTODEA AND MANTODEA Cockroaches, which are among the first neopteran insects to appear in the fossil record, are extremely generalized in most morphological features. They are usually omnivorous. In spite of the lack of detailed data on midgut fluxes and enzyme distribution, it is thought that digestion in cockroaches occurs as described for the Neoptera ancestor (Fig. 5A), except that part of the final digestion of proteins occurs on the surface of midgut cells. Another difference observed is the enlargement of hindgut structures (Fig. 2C), noted mainly in wood-feeding cockroaches. These hindgut structures harbor bacteria producing acetate and butyrate from ingested wood or other cellulose-containing materials. Acetate and butyrate are absorbed by the hindgut of all cockroaches, but this activity is more remarkable with wood roaches. Cellulose digestion may be accomplished by bacteria, but there is evidence that wood roaches have their own cellulases. Mantids have a capacious crop, and a short midgut and hindgut. It is probable that the major part of digestion takes place in their crops.

ISOPTERA Termites are derived from and are more adapted than wood roaches in dealing with refractory materials such as wood and humus. Associated with this specialization, they lost the crop and midgut ceca and enlarged their hindgut structures (Fig. 2D). Termites digest cellulose with their own cellulase, and the products pass from the midgut into the hindgut, where they are converted into acetate and butyrate by hindgut bacteria as in wood roaches. Symbiotic bacteria are also responsible for nitrogen fixation in hindgut, resulting in bacterial protein. This is incorporated into the termite body mass after being expelled in feces by one individual and being ingested and digested by another. This explains the ability of termites to develop successfully in diets very poor in protein.

ORTHOPTERA Grasshoppers feed mainly on grasses, and their digestive physiology clearly evolved from the neopteran ancestor. Carbohydrate digestion occurs mainly in the crop, under the action of midgut enzymes, whereas protein digestion and final carbohydrate digestion take place at the anterior midgut ceca. The abundant saliva (devoid of significant enzymes) produced by grasshoppers saturate the absorbing sites in the midgut ceca, thus hindering the countercurrent flux of fluid. This probably avoids excessive accumulation of noxious wastes in the ceca, and makes possible the high relative food consumption observed among locusts in their migratory phases. Starving grasshoppers present midgut countercurrent fluxes. Cellulase found in some grasshoppers is believed to facilitate the access of digestive enzymes to the plant cells ingested by the insects by degrading the cellulose framework of cell walls. Crickets are omnivorous or predatory insects with most starch and protein digestion occurring in their capacious crop (Fig. 2B).

Paraneoptera

HEMIPTERA The characteristics of the Paraneoptera ancestors cannot be inferred because midgut function data are available only for Hemiptera. The Hemiptera comprise insects of several suborders (e.g., cicadas, leafhoppers, aphids, and fulgorids) that feed almost exclusively on plant sap, and insects of the taxon Heteroptera (e.g., assassin bugs, plant bugs, stinkbugs, and lygaeid bugs) that are adapted to different diets. The ancestor of the entire order is supposed to be a sapsucker similar to present-day cicadas and fulgorids.

The hemipteran ancestor (Fig. 5F) differs remarkably from the neopteran ancestor, as a consequence of adaptations to feeding on plant sap. These differences consist of the lack of crop and anterior midgut ceca, loss of the enzymes involved in

initial and intermediate digestion and loss of the peritrophic membrane associated with the lack of luminal digestion, and, finally, the presence of hemipteran midgut cells (Fig. 3H), which have their microvilli ensheathed by an outer (perimicrovillar) membrane. The perimicrovillar membrane maintains a constant distance from the microvillar membrane, extends toward the luminal compartment with a dead end, and limits a closed compartment, the perimicrovillar space (Fig. 3H). Ongoing research suggests that aphids have modified perimicrovillar membranes.

Sap-sucking Hemiptera may suck phloem or xylem sap. Phloem sap is rich in sucrose (0.15–0.73 M) and relatively poor in free amino acids (15–65 mM) and minerals, whereas xylem fluid is poor in amino acids (3–10 mM) and contains monosaccharides (about 1.5 mM), organic acids, potassium ions (about 6 mM), and other minerals. Thus, except for dimer (sucrose) hydrolysis, no food digestion is necessary in sapsuckers. The major problem facing a sap-sucking insect is to absorb nutrients, such as essential amino acids, that are present in very low concentrations in sap. Amino acids may be absorbed according to a hypothesized mechanism that depends on perimicrovillar membranes. In phloem feeders such as aphids, this process may have an assimilation efficiency of 55% for amino acids and only 5% for sugars, whereas in xylem feeders such as leafhoppers, about 99% of dietary amino acids and carbohydrates are absorbed.

Organic compounds in xylem sap need to be concentrated before they can be absorbed by the perimicrovillar system. This occurs in the filter chamber (Fig. 2P) of Cicadoidea and Cercopoidea, which concentrates xylem sap 10-fold, or in the filter chamber of Cicadelloidea (phloem feeders), which is able to concentrate dilute phloem about 2.5-fold. The filter chamber consists of a thin-walled, dilated anterior midgut in close contact with the posterior midgut and the proximal ends of the Malpighian tubules. This arrangement enables water to pass directly from the anterior midgut to the Malpighian tubules, concentrating food in midgut.

The evolution of Heteroptera was associated with regaining the ability to digest polymers. Because the appropriate digestive enzymes were lost, these insects instead used enzymes derived from lysosomes. Lysosomes are cell organelles involved in intracellular digestion carried out by special proteinases referred to as cathepsins. Compartmentalization of digestion was maintained by the perimicrovillar membranes as a substitute for the lacking peritrophic membrane. Digestion in the two major Heteroptera taxa—Cimicomorpha, exemplified by the blood feeder *R. prolixus,* and Pentatomorpha, exemplified by the seed sucker *Dysdercus peruvianus*—is similar. The dilated anterior midgut stores food and absorbs water and, at least in *D. peruvianus,* also absorbs glucose. Digestion of proteins and absorption of amino acids occur in the posterior ventriculus. Most protein digestion occurs in lumen with the aid of a cysteine proteinase and ends in the perimicrovillar space under the action of aminopeptidases and dipeptidases. Many Heteroptera feed on parenchymal tissues of plants. In some of these insects, excess water passes from the expanded anterior midgut to the closely associated midgut ceca, which protrude from the posterior midgut (Fig. 2Q). These ceca may also contain symbiont bacteria.

Holometabola

The basic gut plan of the Holometabola (Fig. 5B) is similar to that of Neoptera except that fluid secretion occurs in the posterior ventriculus by cells similar to Fig. 3F, instead of by the Malpighian tubules. Because the posterior midgut fluid, unlike Malpighian tubular fluid, does not contain wastes, the accumulation of wastes in ceca is decreased. There is an evolutionary trend leading to the loss of anterior midgut ceca in holometabolous insects and an increase in the use of anterior ventricular cells for water absorption. Ceca loss probably further decreases the accumulation of noxious substances in the midgut, which would be more serious in insects that have high relative food consumption rates, such as is common among Holometabola. Digestive systems may change remarkably between larvae and adults of holometabolous insects. Despite these changes, adult digestive systems probably evolved in parallel to larval systems because, except for minor differences, the compartmentalization of digestion in larvae and adults seems to be similar.

The basic plan of Coleoptera and Hymenoptera did not evolve dramatically from the Holometabola ancestor, whereas the basic plan of Diptera and Lepidoptera (panorpoid ancestor, Fig. 5C) presents important differences. Thus, panorpoid ancestors have countercurrent fluxes like Holometabola ancestors but differ from these in the lack of crop digestion, in midgut differentiation in luminal pH, and in which compartment is responsible for each phase of digestion. In Holometabola ancestors, all phases of digestion occur in the endoperitrophic space (Fig. 1), whereas in panorpoid ancestors only initial digestion occurs in that region. In the latter ancestors, intermediate digestion is carried out by free enzymes in the ectoperitrophic space and final digestion occurs at the midgut cell surface by immobilized enzymes. The free digestive enzymes do not pass through the peritrophic membrane because they are larger than the peritrophic membrane pores. Immobilized enzymes may be either soluble enzymes entrapped in the cell glycocalyx or membrane-bound enzymes, which are those embedded in the lipid bilayer forming the microvillar membranes (intrinsic proteins). As a consequence of the compartmentalization of digestive events in panorpoid insects, there is an increase in the efficiency of digestion of polymeric food by allowing the removal of the oligomeric molecules from the endoperitrophic space, which in turn is powered by the recycling mechanism associated with the midgut fluxes. Because oligomers may be substrates or inhibitors for some polymer hydrolases, their presence should decrease the rate of polymer degradation. A fast

polymer degradation ensures that polymers are not excreted, hence increases their digestibility. Another consequence of compartmentalization is an increase in the efficiency of oligomeric food hydrolysis by allowing the transference of oligomeric molecules to the ectoperitrophic space and by restricting oligomer hydrolases to this compartment. In these conditions, oligomer hydrolysis occurs in the absence of probable partial inhibition (because of nonproductive binding) by polymer food and presumed nonspecific binding by nondispersed undigested food. This process leads to the production of food monomers only in the neighborhood of the midgut cell surface, causing an increase in the concentration of the final products of digestion close to their transporters, thus facilitating absorption.

COLEOPTERA Larvae and adults of Coleoptera usually display the same feeding habit; that is, both are plant feeders (although adults may feed on the aerial parts, whereas the larvae may feed on the roots of the same plant) or both are predatory. Coleoptera ancestors are like Holometabola ancestors except for the anterior midgut ceca, which were lost and replaced in function by the anterior midgut. Nevertheless, there are evolutionary trends leading to a great reduction or loss of the crop and, similar to panorpoid orders, occurrence of final digestion at the surface of midgut cells. Thus, in predatory Carabidae most of the digestive phases occur in the crop by means of midgut enzymes, whereas in predatory larvae of Elateridae initial digestion occurs extraorally by the action of enzymes regurgitated onto their prey. The preliquefied material is then ingested by the larvae, and its digestion is finished at the surface of midgut cells. The entire digestive process occurs in the larval endoperitrophic space of Dermestidae. In Tenebrionidae, the final digestion of proteins takes place at midgut cell surface; in Curculionidae and Cerambycidae, the final digestion of all nutrients is carried out at midgut cell surface. It has been proposed that Cerambycidae larvae acquire the capacity to digest cellulose by ingesting fungal cellulases while feeding on fungus-infested wood. In contrast, Coccinellidae adults use their own cellulase to digest cellulose. The distribution of enzymes in gut regions of adult Tenebrionidae is similar to that of their larvae. This suggests that the overall pattern of digestion in larvae and adults of Coleoptera is similar even though (in contrast to adults) beetle larvae usually lack a crop. Insects of the series Cucujiformia (which includes Tenebrionidae, Chrysomelidae, Bruchidae, and Curculionidae) have cysteine proteinases in addition to (or in place of) serine proteinases as digestive enzymes, suggesting that the ancestors of the whole taxon were insects adapted to feed on seeds rich in serine proteinase inhibitors.

Scarabaeidae and several related families are relatively isolated in the series Elateriformia and evolved considerably from the Coleoptera ancestor. Scarabid larvae, exemplified by dung beetles, usually feed on cellulose materials undergoing degradation by a fungus-rich flora. Digestion occurs in the midgut, which has three rows of ceca (Fig. 2F), with a ventral groove between the middle and posterior row. The alkalinity of gut contents increase to almost pH 12 along the midgut ventral groove. This high pH probably enhances cellulose digestion, which occurs mainly in the hindgut fermentation chamber (Fig. 2F), likely through the action of bacterial cell-bound enzymes. The final product of cellulose degradation is mainly acetic acid, which is absorbed through the hindgut wall. Whether scarabid larvae ingest feces to obtain nitrogen compounds, as described above for termites, is a matter of controversy.

HYMENOPTERA Hymenoptera comprise several primitive suborders (including sawflies and horntails) and Apocrita. Apocrita are divided into Parasitica, which are parasites of other insects, and Aculeata, in which the piercing ovipositor of Parasitica evolved into a stinging organ. The first Apocrita were probably close to the ichneumon flies, whose larvae develop on the surface or inside the body of the host insect. Probably because of that, the larvae of Apocrita present a midgut that is closed at its rear end, and remains unconnected with the hindgut until the time of pupation. Hymenoptera ancestors differ from the Holometabola ancestor in the lack of anterior midgut ceca, which are replaced by the anterior midgut in the function of fluid absorption, and in the absence of midgut enzymes in the crop. Wood wasp larvae of the genus *Sirex* are believed to be able to digest and assimilate wood constituents by acquiring cellulase, xylanase, and possibly other enzymes from fungi present in wood on which they feed. In larval bees, most digestion occurs in the endoperitrophic space. Countercurrent fluxes seem to occur, but the midgut luminal pH gradient hypothetically present in the Hymenoptera ancestor was lost.

Adult bees ingest nectar and pollen. Sucrose from nectar is hydrolyzed in the crop (Fig. 2I) by the action of a sucrase from the hypopharyngeal glands. After ingestion, pollen grains extrude their protoplasm into the ventriculus, where digestion occurs. Worker ants feed on nectar, honeydew, plant sap, or partly digested food regurgitated by their larvae. Thus, they seem to display only intermediate and (or) final digestion.

DIPTERA The Diptera evolved along two major lines: an assemblage of suborders corresponding to the mosquitoes, including the basal Diptera, and the suborder Brachycera, which includes the most evolved flies (Cyclorrhapha). The Diptera ancestor is similar to the panorpoid ancestor (Fig. 5C) in having the enzymes involved in intermediate digestion free in the ectoperitrophic fluid (mainly in the large ceca), whereas the enzymes of terminal digestion are membrane bound at the midgut cell microvilli. Although these characteristics are observed in most nonbrachyceran larvae, the more evolved of these larvae may show reduction in size of midgut ceca (e.g., Culicidae, Fig. 2k). Nonhematophagous adults store liquid food (nectar or decay products) in their

crops. Digestion occurs in their midgut as in larvae. Nectar ingested by mosquitoes (males and females) is stored in the crop, and digested and absorbed at the anterior midgut. Blood, which is sucked only by females, passes to the posterior midgut, where it is digested and absorbed.

The Cyclorrhapha ancestor (Fig. 5D) evolved dramatically from the panorpoid ancestor (Fig. 5C), apparently as a result of adaptations to a diet consisting mainly of bacteria. Digestive events in Cyclorrhapha larvae are exemplified by larvae of the house fly *Musca domestica.* These larvae ingest food rich in bacteria. In the anterior midgut there is a decrease in the starch content of the food bolus, facilitating bacteria death. The bolus now passes into the middle midgut where bacteria are killed by the combined action of low pH, a special lysozyme, and an aspartic proteinase. Finally, the material released by bacteria is digested in the posterior midgut, as is observed in the whole midgut of insects of other taxa. Countercurrent fluxes occur in the posterior midgut powered by secretion of fluid in the distal part of the posterior midgut and its absorption back into the middle midgut. The middle midgut has specialized cells for buffering the luminal contents in the acidic zone (Fig. 3D), in addition to those functioning in fluid absorption (Fig. 3A). Except for a few bloodsuckers, Cyclorrhaphan adults feed mainly on liquids associated with decaying material (rich in bacteria) in a way similar to house fly adults. That is, they salivate (or regurgitate their crop contents) onto their food. After the dispersed material has been ingested, starch digestion is accomplished primarily in the crop by the action of salivary amylase. Digestion is followed in the midgut, essentially as described for larvae. The stable fly, *Stomoxys calcitrans,* stores and concentrates the blood meal in the anterior midgut and gradually passes it to the posterior midgut, where digestion takes place, resembling what occurs in larvae. These adults lack the characteristic cyclorrhaphan middle midgut and the associated low luminal pH. Stable flies occasionally take nectar.

LEPIDOPTERA Lepidopteran ancestors (Fig. 5E) differ from panorpoid ancestors because they lack midgut ceca, have all their digestive enzymes (except those of initial digestion) immobilized at the midgut cell surface, and present long-necked goblet cells (Fig. 3B) and stalked goblet cells (Fig. 3E) in the anterior and posterior larval midgut regions, respectively. Goblet cells excrete K^+ ions, which are absorbed from leaves ingested by larvae. Goblet cells also seem to assist anterior columnar cells in water absorption and posterior columnar cells in water secretion. Although most lepidopteran larvae have a common pattern of digestion, species that feed on unique diets generally display some adaptations. *Tineola bisselliella* (Tineidae) larvae feed on wool and display a highly reducing midgut for cleaving the disulfide bonds in keratin to facilitate proteolytic hydrolysis of this otherwise insoluble protein. Wax moths *(Galleria mellonella)* infest beehives and digest and absorb wax. The participation of symbiotic bacteria in this process is controversial. Another adaptation has apparently occurred in lepidopteran adults that feed solely on nectar. Digestion of nectar requires only the action of an α-glucosidase (or a β-fructosidase) to hydrolyze sucrose, the major component present. Nectar-feeding lepidopteran adults have amylase in salivary glands and several glycosidases and peptidases in the midgut. The occurrence of the whole complement of digestive enzymes in nectar-feeding moths may explain, at least on enzymological grounds, the adaptation of some adult Lepidoptera to new feeding habits such as blood and pollen.

ABSORPTION OF WATER AND NUTRIENTS

Overview and Absorption of Lipids

Absorption is the passage of molecules and ions from the gut lumen into the gut cells, thus traversing the cuticle (if present) and the cell plasma membrane. Absorption depends on the permeability of those barriers and on the concentration ratio of a compound in gut lumen and inside gut cells. The permeability of cuticles is variable, whereas that of the plasma membrane is greater for water (a fact not well understood) and for hydrophobic compounds, reflecting the ease with which they solubilize in the lipid bilayers characteristic of cell membranes. Thus, absorption of hydrophilic compounds requires special devices (transporters) to help the molecules find their way through the cell membrane. These transporters are transmembrane proteins that bind the molecule to be transported in a membrane face and, after suitable conformational changes, deliver the molecule from the other face. A uniporter is a transporter that carries a single solute, whereas symporters and antiporters are transporters that carry two solutes into the same and opposite directions, respectively. To transport molecules against a concentration gradient, the process must be energized by coupling with ATP hydrolysis or with cotransport of another molecule down its concentration gradient. Transporter-mediated absorption may be inhibited by molecules resembling those of the transported solute, and its velocity attains a maximum (transporter becomes saturated) at a high solute concentration. This behavior is not observed in the case of simple diffusion, exemplified by lipid absorption.

The study of gut absorption in insects is difficult because of the small size of these animals, which frequently hinders the use of methods developed for studying vertebrates. Absorption sites in insect guts are identified by feeding groups of insects with known dye solutions and then dissecting insects at different periods of time. If the insect is large enough, absorption studies can go further, using gut sections mounted as a sac and measuring the rates at which compounds traverse it under different conditions.

Tracer studies showed that lipid is absorbed more heavily in the anterior than in the posterior midgut of insects, thus following the tissue distribution of fatty acid binding

proteins. These proteins are thought to facilitate fatty acid uptake by cells, by decreasing their diffusion back from cells to the gut lumen and by targeting them to specific metabolic pathways. The fatty acids acetate and butyrate are absorbed in significant amounts by the hindgut of insects utilizing cellulose.

Water

Water absorption in the midgut occurs associated with midgut fluid fluxes, but in large amounts it is characteristic of insects feeding on dilute diets, of blood feeders, and also of insects in which salivation is important (e.g., grasshoppers, seed-sucker bugs). Water absorption in the hindgut is part of the water conservation mechanism that is important in all terrestrial insects. Frequently, special cell aggregates (rectal pads) are involved in this process. Water uptake is thought to depend on salt being pumped into spaces enclosed by the basolateral infoldings (Fig. 3A, C) of the absorbing cell. This creates an osmotic pressure that moves water into these restricted spaces. The resulting hydrostatic pressure drives water into the hemolymph, with salts being absorbed from the water on its way out. In insects feeding on dry diets in dry habitats (e.g., *T. molitor, D. maculatus*), hindgut water absorption is improved by the cryptonephridial system, which consists of an association of Malpighian tubules and rectal pads. Less sophisticated forms of this system are found in other insects.

Ions, Amino Acids, and Sugars

In insects, as is usual for all animals, most nutrient absorption occurs in the midgut through symporters, with ions being cotransported down the concentration gradient. Favorable ion gradients are maintained by ion pumps. The most ubiquitous of these pumps is the ATP-driven Na^+, K^+-antiporter (Na^+-K^+-ATPase) localized in the midgut cell basal membrane. Another important ion pump is the H^+, K^+-ATPase found in the goblet cell (Fig. 3B, E) microvillar membranes of lepidopteran larvae. Those pumps maintain cell Na^+ and cell K^+ low in insects with Na^+-rich and K^+-rich diets, respectively. The midgut of lepidopteran larvae contains K^+-dependent symporters for amino acids, consistent with the K^+-rich plant diets of these larvae. Such symporters are relatively insensitive to amino acid sizes and shapes, but have narrow specificities towards charge. Thus, there are usually symporters for neutral, acid, and basic amino acids. Insects with high-Na^+ diets seem to have midgut amino acid symporters dependent on Na^+, as shown in cockroaches. The absorption of glucose was shown in several insects to depend on a difference of concentration between midgut lumen and cells. This transport increases as the concentration of luminal glucose increases (no saturation) and is not inhibited by molecules similar to glucose, at least in the range of concentrations tested. This finding led to the speculation that, in insects, glucose is absorbed by simple diffusion. Nevertheless, this is probably false, since a hydrophylic molecule is not expected to pass through membranes without the help of a transporter.

The primary urine produced in Malpighian tubules contains salts and amino acids, and passes into the hindgut together with food remains. Salts are absorbed in the hindgut by means of special pumps, like those for chloride and calcium. Amino acids, at least in locusts, are absorbed in the hindgut through a Na^+-dependent amino acid symporter.

MIDGUT SECRETORY MECHANISMS

Insects are continuous (e.g., Lepidoptera and Diptera larvae) or discontinuous (e.g., predators and hematophagous insects) feeders. Synthesis and secretion of digestive enzymes in continuous feeders seem to be constitutive; that is, these functions occur continuously, whereas in discontinuous feeders they are regulated. It is widely believed (without clear evidence) that putative endocrine cells (Fig. 3I) occurring in the midgut could, like similar cells in vertebrates, play a role in regulating midgut events. The presence of food in the midgut is necessary to stimulate synthesis and secretion of digestive enzyme. This was clearly shown in mosquitoes.

Mosquitoes express constitutively small amounts of a trypsin called early trypsin. After a blood meal, early trypsins generate free amino acids and small peptides from blood proteins. These compounds are the initial signals that induce the synthesis and secretion of large amounts of late trypsins, which complete protein digestion.

Like all animal proteins, digestive enzymes are synthesized in the rough endoplasmic reticulum, processed in the Golgi complex, and packed into secretory vesicles (Fig. 4). There are several mechanisms by which the contents of the secretory vesicles are freed in the midgut lumen. During exocytic secretion, secretory vesicles fuse with the midgut cell apical membrane, emptying their contents without any loss of cytoplasm (Fig. 4A). In contrast, apocrine secretion involves the loss of at least 10% of the apical cytoplasm following the release of secretory vesicles (Fig. 4B). These have previously undergone fusions originating larger vesicles that after release eventually free their contents by solubilization (Fig. 4B). When the loss of cytoplasm is very small, the secretory mechanism is called microapocrine. Microaprocrine secretion consists of releasing budding double-membrane vesicles (Fig. 4C) or, at least in insect midguts, pinched-off vesicles that may contain a single or several secretory vesicles (Fig. 4D). In both apocrine and microapocrine secretion, the secretory vesicle contents are released by membrane fusion and/or by membrane solubilization due to high pH contents or to the presence of detergents.

Secretion by hemipteran midgut cells displays special features because the cells have perimicrovillar membranes, in addition to microvillar ones (Fig. 3H): double-membrane vesicles bud from modified (double-membrane) Golgi

structures (Fig. 4E). The double-membrane vesicles move to the cell apex, their outer membranes fuse with the microvillar membrane, and their inner membranes fuse with the perimicrovillar membranes, emptying their contents (Fig. 4E). Because apocrine and microapocrine mechanisms waste membrane and cytoplasm material, these mechanisms are preferred only when they present advantages over the exocytic mechanism. This occurs when a burst of digestive enzymes is needed, as in hematophagous flies after a blood meal, and when secretion occurs in a midgut region responsible for water absorption, a common situation in the anterior midgut of most insects. An exocytic mechanism in a water-absorptive region is not efficient, because the movement of fluid toward the cells would prevent uniform diffusion of the material secreted. Fluid movement has little effect on apocrine and microapocrine secretion because the enzymes are released from budded or pinched-off secretory vesicles far from cells. Since posterior midgut cells usually secrete fluid, no problem arises in the dispersion of material released by exocytosis by these cells. Microapocrine mechanisms seem to be an improvement relative to apocrine mechanisms, because they waste less material. This is consistent with the observation that apocrine mechanisms were found in less evolved grasshoppers and beetles, whereas microaprocine mechanisms were described in the more evolved moths.

See Also the Following Articles

Blood Sucking • Excretion • Feeding Behavior • Symbionts

Further Reading

Chapman, R. F. (1998). "The Insects: Structure and Function." 4th ed. Cambridge University Press, Cambridge, U.K. (See especially Chaps. 2–4).

Cristofoletti, P. T., Ribeiro, A. F., and Terra, W. R. (2001). Apocrine secretion of amylase and exocytosis of trypsin along the midgut of *Tenebrio molitor* larvae. *J. Insect Physiol.* **47,** 143–155.

Daly, H. V., Doyen, J. T., and Purcell III, A. H. (1998). "Introduction to Insect Biology and Diversity." 2nd ed. Oxford University Press, Oxford, U.K. (See especially Chap. 15.)

Dow, J. A. T. (1986). Insect midgut function. *Adv. Insect Physiol.* **19,** 187–328.

Kerkut, G. A., and Gilbert, L. I. (eds.). (1985). "Comprehensive Insect Physiology, Biochemistry and Pharmacology," 13 vols. Pergamon Press, Oxford, U.K. (See especially Vol. 4, Chaps. 4–6.)

Lehane, M. J., and Billingsley, P. F. (1996). "Biology of the Insect Midgut." Chapman & Hall, London.

Silva, C. P., Ribeiro, A. F., Gulbenkian, S., Terra, W. R. (1995). Organization, origin and function of the outer microvillar (perimicrovillar) membranes of *Dysdercus peruvianus* (Hemiptera) midgut cells. *J. Insect Physiol.* **41,** 1093–1103.

Terra, W. R. (1990). Evolution of digestive systems of insects. *Annu. Rev. Entomol.* **35,** 181–200.

Terra, W. R., and Ferreira, C. (1994). Insect digestive enzymes: Properties, compartmentalization and function. *Comp. Biochem. Physiol.* **109B,** 1–62.

Terra, W. R. (2001). The origin and functions of the insect peritrophic membrane and peritrophic gel. *Arch. Insect Biochem. Physiol.* **47,** 47–61.

Vonk, H. J., and Western, J. R. H. (1984). "Comparative Biochemistry and Physiology of Enzymatic Digestion." Academic Press, London.

Wolfersberger, M. G. (2000). Amino acid transport in insects. *Annu. Rev. Entomol.* **45,** 111–120.

Diplura

Robert T. Allen

Author and Consultant, Paris, Arkansas

The Diplura are a group of primitive Arthropods usually included in the class Insecta. Typically they live in the soil, under the bark of decaying trees, under rocks, and under decaying leaf debris. The majority of the species are small, less than 5 mm, but some members of the superfamily Japygoidea, *Atlasjapyx atlas,* may be 60 mm in length. The Diplura are worldwide in distribution, with about 1000 described species assigned to eight families. The name Diplura is derived from the presence of paired caudal appendages.

CLASSIFICATION

All Diplura possess the following defining characters: (1) elongate body, (2) 10th abdominal segment with a pair of caudal cerci or one segmented forcep-like pincers, (3) absence of eyes, (4) entognathous mouthparts, and (5) two pair of spiracles on the thorax (Fig. 1).

There is no doubt that these organisms are primitive arthropods standing near the base of the evolutionary lineage that led to the class Insecta. Whether they should be included in the Insecta or merit a class unto themselves is debatable. Kristensen assigned them to their own class (class and order Diplura) in 1991. This suggestion, although it may be correct, has not been followed in general textbooks of entomology. The classification within the order Diplura has developed gradually over the past 100 years as our knowledge of the group increased. Table I lists the major taxa and their general distribution.

The 1000 or so species that have thus far been described may represent only 50% or less of the actual world fauna. Distribution records of the known taxa are poor, with many species known from a single locality. A great deal of work is left to be accomplished before an accurate idea of the diversity and distribution of the Diplura is known.

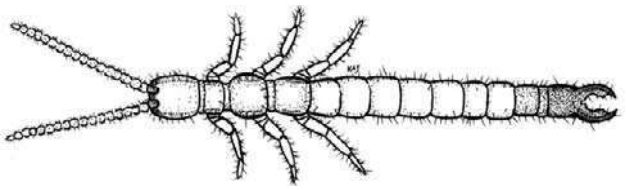

FIGURE 1 Dipluran (*Parajapyx* sp.). (Illustration by K. A. Justus.)

TABLE I Classification of the Diplura

Class or order	Diplura
Suborder	Rhabdura
Superfamily	Projapygoidea
Family	Anajapygidae (1 genus *Anajapyx,* 4+ spp.; CA, MX)
Family	Projapygidae (2 genera *Projapyx,* 7 spp.; *Symphylurinus,* 19+ spp.; West AF, Brazil, MX, AU, CH, IN)
Superfamily	Campodoidea
Family	Procampodeidae (1 genus, *Procampodea,* 2 spp.; CA)
Family	Campodeidae
Subfamily	Campodeinae (30+ genera, 160–200+ spp.; cosmopolitan)
Subfamily	Lepidocampinae (1 genus; tropical, cosmopolitan)
Subfamily	Hemicampinae (2 genera, 4 spp.; US, MX, SA)
Subfamily	Plusiocampinae (5 genera, 40+ spp.; cosmopolitan)
Subfamily	Syncampinae (1 genus, 1 sp.; China)
Family	Octostigmatidae (1 genus, 1 sp.; South Pacific)
Suborder	Dicellurata
Superfamily	Japygoidea
Family	Parajapygidae (cosmopolitan)
Family	Dinjapygidae (Peru, Bolivia)
Family	Japygidae
Subfamily	Heterojapyginae (AU, NZ, Mad., Pamit, Tibet)
Subfamily	Japyginae (cosmopolitan)
Subfamily	Evalljapyginae (NA, CA)
Subfamily	Provalljapyginae (NA, Brazil)

Note. CA, Central America; MX, Mexico; AF, Africa; AU, Australia; CH, China; IN, India; SA, South America; NZ, New Zealand; Mad., Madagascar; NA, North America; US, United States.

COLLECTING AND SPECIMEN PREPARATION

Because of their small size and obscure living habitats Diplura are not well represented in most collections. They are, however, easily collected in most ecosystems. The collector needs small vials of ethyl alcohol, a size 00 or 000 camel hair brush, a small tool to turn rocks or pry away bark, and a keen eye. When a rock is turned over, the specimens may be on the underside of the rock or on the soil. They usually begin to move rapidly once exposed and it is necessary to act quickly. The brush is wetted in the vial of alcohol and then touched to the specimen. Usually the specimen will adhere to the brush and can then be transferred to the alcohol vial. Some of the larger Japygidae may be collected using the fingers or a pair of forceps. Once specimens are collected they must be mounted on microscope slides for study and identification.

There is no single source that allows ready identification of Diplura. The sources listed under Further Reading will assist in keying specimens to families and genera and provide a guide to the numerous papers necessary for species identification.

BIOLOGY

Relatively few studies have been done on the Diplura, and thus we know very little about their habits. However, the studies that have been published have recorded the basic biological characteristics of the group. The males deposit sperm bundles in the soil and females pick up these sperm bundles and become fertilized. Eggs may be deposited randomly and in clusters. Some japygids suspend eggs at the end of a filimentous stalk. The prelarvae hatch in 7 to 16 days depending on the species. The prelarva does not feed and moves very little. The prelarva molts in about 2 days. The newly molted immature is fully mobile and feeds readily on whatever food source is available. After the second molt the immature form possesses the major setae and other anatomical characters used for identification. During the fourth or fifth molt the individual becomes sexually mature as evidenced by the appearance of the sex organs along the posterior margin of sternum VIII. Diplura continue to molt throughout their lives, adding clothing setae on the various sclerites and regenerating damaged body appendages.

Both major groups of Diplura appear to be omnivores. Many species are predators as well as scavengers. Foods that have been recorded include other Diplura, mites, Collembola, Symphyla, Isopoda, fly and beetle larvae, small arthropods of any class, enchytraeid worms, fungal spores, and mycelia. Some species have been observed feeding on the roots of living plants, including peanuts, sugarcane, and melons.

See Also the Following Articles

Arthropoda and Related Groups • Protura

Further Reading

Allen, R. T. (1995). Key to the species of Campodea (Campodea) from eastern North America and description of a new species from Virginia (Diplura: Campodeidae). *Ann. Entomol. Soc. Am.* **88,** 255–262.

Ferguson, L. M. (1990). Insecta: Diplura. *In* "Soil Biology Guide" (D. L. Dindall, ed.), pp. 951–963. Wiley, New York.

Kristensen, N. P. (1991). Phylogeny of extant hexapods. *In* "The Insects of Australia" (CSIRO, ed.), 2nd ed., Vol. I, pp. 125–140. Melbourne University Press, Carlton.

Paclt, J. (1957). Diplura. *In* "Genera Insectorum" (P. Wytsman, ed.), pp. 1–123. Crainhem, Belgium.

Diptera

(Flies, Mosquitoes, Midges, Gnats)

Richard W. Merritt
Michigan State University

Gregory W. Courtney
Iowa State University

Joe B. Keiper
The Cleveland Museum of Natural History

The Diptera, commonly called true flies or two-winged flies, are a group of familiar insects that includes mosquitoes, black flies, midges, fruit flies, and house flies. The Diptera are among the most diverse insect orders, with approximately 124,000 described species. These insects are diverse not only in species richness but also in their structural variety, ecological habits, and economic importance. The group is ubiquitous and cosmopolitan, having successfully colonized nearly every habitat and all continents, including Antarctica. Although brachyptery (wings reduced) or aptery (wings absent) are known in some Diptera (e.g., some Mycetophilidae, Tipulidae, Phoridae, and Hippoboscidae), adults usually are winged and active fliers. Depending on the group, adults can be nonfeeding or feeding, with the latter including diets of blood, nectar, and other liquefied organic materials.

Larval Diptera are legless and found in a variety of terrestrial and aquatic habitats. Most larvae are free-living and crawl or swim actively in water (e.g., Simuliidae, Culicidae, Chironomidae, Ptychopteridae, Blephariceridae), sediments (e.g., Tipulidae, Psychodidae, Ceratopongonidae, Tabanidae), wood (e.g., Tipulidae, Mycetophlidae), fruit (e.g., Drosophilidae, Tephritidae), or decaying organic material (e.g., Muscidae, Ephydridae, Sphaeroceridae, Sarcophagidae). Other larvae inhabit the tissues of living organisms (e.g., Oestridae, Tachinidae).

As expected for a ubiquitous group with diverse habits and habitats, the Diptera are of considerable economic importance. Pestiferous groups can have significant impacts in agriculture (e.g., Agromyzidae, Tephritidae), forestry (e.g., Cecidomyiidae), animal health (e.g., Oestridae), and human health (e.g., Culicidae, Simuliidae, Psychodidae). Other groups can be a general nuisance if present in high numbers (e.g., Muscidae, Ceratopogonidae) or because of allergic reactions to detached body hairs (e.g., Chironomidae). Despite these negative impacts, flies can play a valuable role as scavengers (e.g., Mycetophilidae, Muscidae, Calliphoridae), parasitoids and predators of other insects (e.g., Tachinidae, Empididae, Asilidae), pollinators (e.g., Syrphidae, Stratiomyiidae, Bombyliidae), food for vertebrates (e.g., Chironomidae, Tipulidae), bioindicators of water quality (e.g., Chironomidae, Blephariceridae), and tools for scientific research (e.g., Drosophilidae).

MORPHOLOGY

Because of the structural variety in Diptera, especially among larvae, it is difficult to generalize about morphology. Despite this variety, flies share a number of features. Except for certain forms (e.g., cave-dwelling species), adult flies usually possess large compound eyes. In some species, eyes meet or almost meet dorsally (holoptic); in other groups, eyes are widely separated (dichoptic). Further modifications include eyes that are divided into distinct dorsal and ventral components, a feature found in many Simuliidae, Blephariceridae, and other groups. These modifications are among many that might be related to swarming behavior. The regions of a fly head include the vertex, a dorsomedial area above and posterior to the eyes; the frons, an area extending from the vertex to the antennal insertions; and the face, which extends from the antennal insertions to the clypeus, a region intimately associated with the mouthparts. All of these areas can bear a variety of setae, the number and position of which often are useful in identification.

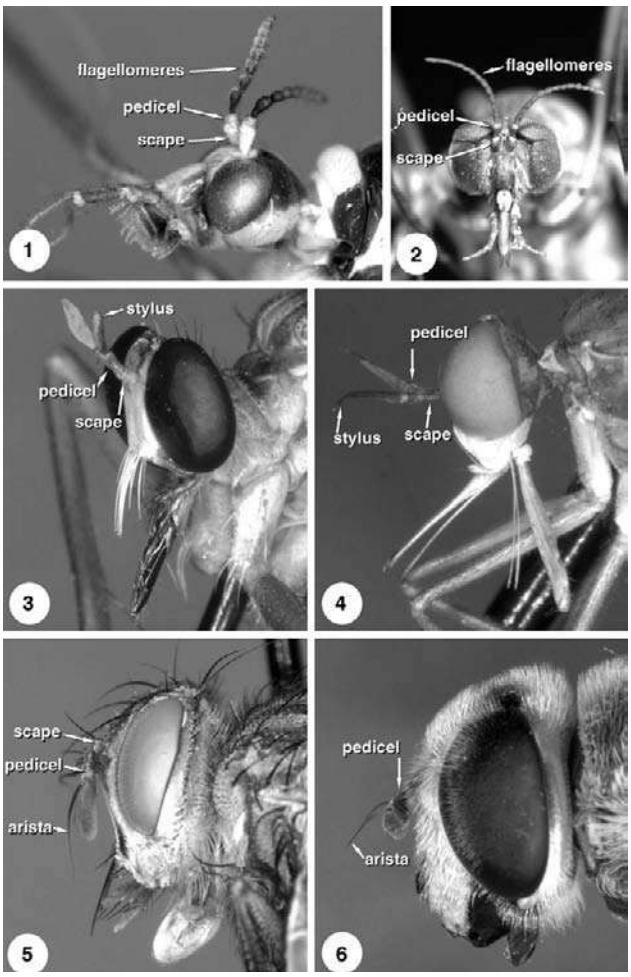

FIGURES 1–6 Adult head of (1) Tipulidae, (2) Blephariceridae, (3) Asilidae, (4) Empididae, (5) Tachinidae, (6) Syrphidae. (Photographs by G. Courtney.)

Nearly all flies have well-developed antennae, with the flagellum being the most varied component. In nematocerous families, the antennae are usually composed of many segments and are filiform, plumose, or pectinate (Figs. 1–2), whereas brachycerous flies typically have the first flagellomere enlarged and the remaining flagellomeres stylate or aristate (Figs. 3–6). The mouthparts of adult flies also vary between groups, ranging from vestigial forms (e.g., Deuterophlebidae, Oestridae) to those that are well developed. The latter include two general types: (1) piercing and sucking, as seen in simuliids, culicids, and asilids, and (2) lapping and sucking, as seen in tipulids and most brachycerous groups. Typically, the proboscis comprises the unpaired labrum–epipharynx, labium, and hypopharynx and the paired

mandibles and maxillae. In most groups, the base of each maxilla bears a distinct palpus and the apex of the labium is modified into a labellum, which consists of membranous lobes derived from the labial palpi.

Perhaps the most distinct feature of the adult fly is the single pair of wings (hence, the ordinal name, Diptera, meaning "two wings"). A related characteristic is the highly modified thorax, with a reduced prothorax and metathorax, and a greatly enlarged mesothorax. The latter includes several prominent dorsal and lateral sclerites and, internally, houses much of the wing musculature. Wing venation varies greatly throughout the Diptera and can be extremely important for identification. The metathoracic wings are modified into distinct club-shaped halteres, which are thought to play an important role as balancing organs. Interestingly, halteres are distinct in some groups that are otherwise wingless (e.g., Hippoboscidae). The legs of an adult fly are typical of most insects, each with a coxa, trochanter, femur, tibia, and, in nearly all groups, a tarsus comprising five tarsomeres. Beyond this basic arrangement, there is considerable diversity of leg structure in Diptera, with this diversity often providing useful taxonomic information.

The adult abdomen also shows considerable variety. In basic structure, the abdomen consists of 11 segments, the last 2 or 3 of which are highly modified for reproduction. Most abdominal segments consist of a dorsal and ventral sclerite, connected laterally by a pleural membrane of varying width. There is a general trend toward a shortening of the abdomen in Diptera (cf. Tipulidae and Muscidae). The terminalia of Diptera are complex, highly variable, and of considerable use in taxonomic and phylogenetic studies. Details of terminalic structure are beyond the scope of this article; however, the structural variety of Diptera terminalia and the controversy about interpreting their homologies can be found in some of the general references listed at the end.

The dipteran pupa also varies considerably in form. Some fly pupae look like a cross between the worm-like larva and the adult, whereas others are relatively featureless and seed-like in appearance. The former are typical of the Nematocera and are described as obtect, or having the appendages fused to the body (Figs. 7–10). For instance, a crane fly (Tipulidae) pupa has identifiable head, thoracic, and abdominal segments, but the antennal sheaths, legs, and wing pads adhere to the pupal body (Fig. 9). Nematocerous pupae are frequently leathery to the touch. The exterior of the nematoceran pupa may be adorned with spines, gill-like respiratory devices, or locomotory paddles (Figs. 7–10). The Brachycera and Cyclorrhapha form the pupal stage in a different, more concealed manner. Families of the so-called higher Diptera form pupae that are described as coarctate, which literally means "compacted" or "contracted" (Figs. 11–15). These taxa (e.g., Syrphidae, Drosophilidae, Muscidae) form a puparium that is composed of the hardened skin of the last larval instar (Fig. 14). This relatively tough, desiccation-resistant structure houses and protects the pupa; the adult

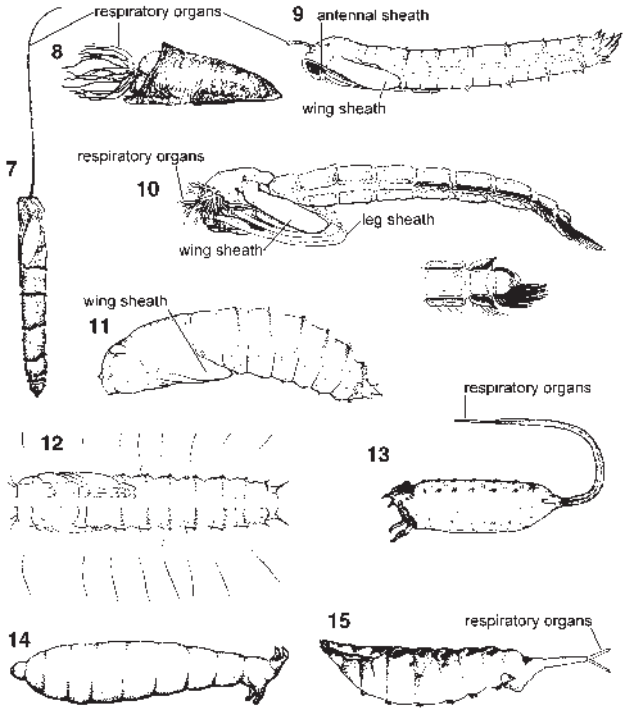

FIGURES 7–15 Pupa of (7) Ptychopteridae, (8) Simuliidae, (9) Tipulidae, (10) Chironomidae, showing anal division below, (11) Tabanidae, (12) Empididae, (13) Syrphidae, (14) Muscidae, (15) Ephydridae. (All illustrations modified, with permission, from Merritt and Cummins, 1996.)

also forms within the puparium. The enclosed adult must break through the puparial skin and does so by extruding a balloon-like structure from the frons called the ptilinum. The ptilinum is used to break the cephalic cap, a lid-like structure positioned anteriorly on the puparium, thus liberating the teneral (or newly emerged) adult. Very few external features are noticeable on the puparium, although careful examination will reveal the spiracles through which atmospheric air is obtained by the pupa.

Diptera larvae can be distinguished from the larvae of most other insects by the lack of jointed thoracic legs. In other features, larval dipterans show tremendous structural variety. This variation is exemplified by cranial structure. Larvae of most nematocerous flies are eucephalic, i.e., characterized by a complete, fully exposed, and heavily sclerotized head capsule (Figs. 17–19 and 24). Larval tipulids are special among nematocerous flies, as the head capsule often is fully retracted into the thorax (Fig. 16) and the posterior cranial margin may possess small to extensive longitudinal incisions (Fig. 23). In contrast to the condition in nematoceran larvae, the cranial sclerites of brachyceran larvae are greatly reduced or absent. The hemicephalic head capsule of many orthorrhaphous Brachycera consists of slender arms and rods that are partly retracted into the thorax (Figs. 25–26). The culmination of cranial reduction is in the acephalic head of larval Cyclorrhapha, in which the external portions of the head are membranous, and much of the head is retracted into the

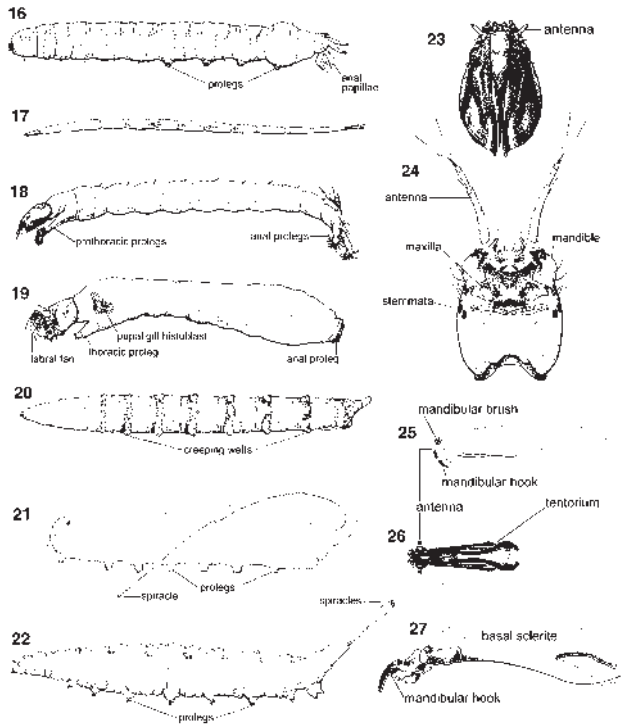

FIGURES 16–27 Larva of (16) Tipulidae, (17) Ceratopogonidae, (18) Chironomidae, (19) Simuliidae, (20) Tabanidae, (21) Syrphidae, (22) Ephydridae. Larval head capsule of (23) Tipulidae, (24) Chironomidae. Cranial sclerites and mouth parts of (25) Tabanidae, (26) Dolichopodidae. (27) Cephalopharyngeal skeleton of Sciomyzidae. (All illustrations modified, with permission, from Merritt and Cummins, 1996.)

thorax (Fig. 27). The internal portion, or cephalopharyngeal skeleton, is thought to comprise the remnants of internal cranial sclerites (tentorium) and various mouthparts. Although referred to as "acephalic," the primary difference between the head of a cyclorrhaphan larva and that of a nematoceran larva is that most of the constituent segments are withdrawn into the thorax and thus externally hidden (Fig. 22). Cranial modifications are accompanied by general changes in the shape and rotation of mandibles and other mouthparts. The mandible of larval nematocerans typically consists of a stout, toothed structure that moves in a horizontal or oblique plane and operates as a biting and chewing organ. The brachyceran larval mandible usually is more claw-like, has fewer teeth along the inner surface, moves in a vertical plane, and operates as a piercing or slashing organ.

In most Diptera larvae, the thorax and abdomen are soft, flexible, and only occasionally provided with sclerotized plates. The thorax usually consists of three distinct segments and the abdomen usually eight or nine segments (Figs. 17–19). Body form varies almost as much as does cranial diversity and ecological habits. In many nematoceran groups (e.g., most Chironomidae, Tipulidae, and Simuliidae), the body is subcylindrical (Figs. 16, 18, and 19). Other groups are predominantly fusiform (e.g., Cecidomyiidae) or elongated and serpentine (e.g., Ceratopogonidae) (Fig. 17). The latter body form is common in groups inhabiting soil and interstitial aquatic habitats. The larvae of some groups (e.g., Culicidae) are unusual in that the thoracic segments are indistinctly differentiated and form a single large segment that is wider than the rest of the body (Fig. 48). The typical body shape of a cyclorrhaphan larva is that of a maggot (i.e., pointed at the anterior end, with the thoracic segments approaching the maximum body diameter). The variation in body form is particularly impressive in families whose larvae feed on a variety of substrates (e.g., Syrphidae). Cyclorrhaphan larvae can be dorsoventrally flattened, a feature often associated with the presence of segmental or branched body protuberances. The syrphid genus *Microdon* has one of the most unusual larvae, being ventrally flattened, dorsally dome-shaped, and sluglike in overall appearance. Larvae with parasitoid and parasitic life styles (e.g., Pipunculidae, Oestridae) are often extremely stout or pear-shaped, their body form being closely adapted to that of the host.

Despite the absence of jointed thoracic legs, locomotion is highly diverse in fly larvae, reflecting the group's diversity in habitat and habits. Locomotory appendages operate through a combination of turgor pressure and muscle action and include creeping welts, prolegs, and other specialized structures (e.g., suctorial discs). Creeping welts are transverse, swollen areas (ridges) that bear one to several modified setae or spines; creeping welts are characteristic of several groups, including many crane flies, dance flies, and deer and horse flies (Fig. 20). Among orthorrhaphous groups, ventral creeping welts are common in the larvae of Rhagionidae and Empididae. Cyclorrhaphan larvae typically use creeping welts as anchoring devices, with welts usually comprising bands of small spines on abdominal segments. The distribution and morphology of creeping welts vary considerably between families, species, instars, and segments. Prolegs usually are paired, round, elongate, fleshy, retractile processes that bear apical spines or crochets; prolegs come in a diversity of shapes, sizes, and positions and are typical of Chironomidae, Deuterophlebidae, Simuliidae, Rhagionidae, and various members of other groups (Figs. 16, 18, 19, 21, and 22). Other specialized structures used for locomotion or attachment include friction pads and suctorial discs. Several genera of Psychodidae possess friction pads, which are areas of modified cuticle on the ventral surface of the thorax or abdomen. Functionally similar structures may occur in certain Ephydridae, particularly in groups inhabiting waterfalls and thin films of flowing water. Suctorial disks are true suction devices on the ventral body surface of larval net-winged midges and are an obvious adaptation to life in torrential streams.

Larval Diptera show a variety of respiratory adaptations, many a reflection of life in fluid or semifluid habitats. The basic respiratory system comprises an internal system of tracheae and the external spiracles. Respiration may be directly from the atmosphere, from plant tissues, or from oxygenated fluids. The presence of hemoglobin in the blood of some midges can assist the absorption of oxygen. Many aquatic

larvae, particularly those from well-oxygenated streams, are apneustic (lack spiracles) and absorb oxygen directly through the skin. Some families (e.g., Psychodidae) possess spiracles on the prothorax and last abdominal segment, whereas others (e.g., Culicidae and most cyclorrhaphans) have spiracles on only the last segment. In several groups (e.g., many Ephydridae and Syrphidae), the spiracles are at the end of a retractile respiratory siphon (Figs. 21 and 22).

PHYLOGENY AND CLASSIFICATION

Traditionally the Diptera have been divided into two or three suborders: Nematocera ("lower" Diptera) and Brachycera ("higher" Diptera), with the latter sometimes divided further into the Orthorrhapha and Cyclorrhapha. Although there is general agreement that the Diptera, Brachycera, Cyclorrhapha, and a few other subordinate groups are monophyletic, there is comparably general agreement that the Nematocera is a paraphyletic or grade-level grouping. No synapomorphies (shared, derived characters) unite the Nematocera, and the Brachycera are thought to have originated from some subgroup within the Nematocera. Despite this, it is useful to mention some of the primitive features shared by most nematocerans. The name itself ("Nematocera") refers to the fact that adults of these flies typically have long, multisegmented antennae. Furthermore, adult nematocerans generally are slender, delicate, long-legged flies (e.g., Tipulidae and Culicidae); however, the group also includes some rather stout-bodied flies (e.g., Simuliidae and Ceratopogonidae). Larval nematocerans typically have a well-developed, sclerotized head capsule, and their mandibles usually rotate at a horizontal or oblique angle. Brachycera are characterized by the short, three-segmented antennae, the last segment of which is usually either stylate or aristate. Brachyceran larvae usually have a hemicephalic or acephalic head capsule, consisting mostly of slender, sclerotized rods that are partly or largely retracted into the thorax. Within the Brachycera, there are additional differences between orthorrhaphous and cyclorrhaphous groups. The former group, which includes Rhagionidae, Tabanidae, Stratiomyiidae, and a few other families, is similar to nematocerous Diptera in that it is considered a paraphyletic group. Finally, within the Cyclorrhapha are two major subgroups, the presumed paraphyletic Aschiza (includes Phoridae and Syrphidae) and the monophyletic Schizophora (includes the majority of Brachycera, such as Tephritidae, Drosophilidae, Ephydridae, Agromyzidae, Muscidae, and Tachinidae).

ECOLOGY

Life History

As a holometabolous insect, or one that undergoes complete metamorphosis, the dipteran life cycle includes a series of distinct stages or instars. A typical life cycle consists of a brief egg stage (usually a few days or weeks, but sometimes much longer), three or four instars (typically three in Brachycera, four in nematocerous flies, and more in simuliids, tabanids, and a few others), a pupal stage of varying length, and an adult stage that lasts from less than 2 h (Deuterophlebidae) to several weeks or even months (some female Culicidae). The eggs of aquatic flies are usually laid singly, in small clusters, or in loose or compact masses in or near the water and attached to rocks or vegetation. In Deuterophlebiidae and certain members of some other groups, the female crawls beneath the water to select oviposition sites, a behavior that ensures eggs are placed in a suitable larval habitat. The latter also is typical of many terrestrial flies, such as calliphorids, which will lay their eggs near the body openings (eyes, nose, mouth, anus) of carcasses. Some tephritid fruit flies use a rigid ovipositor to pierce plant tissue. Oviposition in parasitic flies can be complex and may involve placement of eggs in or on the host or in areas frequented by the host. Some parasitoids (e.g., some tachinid flies) produce eggs that are ingested by a feeding host, then larvae hatch inside the host and penetrate the gut wall. Furthermore, some parasitic groups will oviposit on a blood-feeding arthropod (e.g., tick or another fly), with the heat of the next host stimulating hatching.

All instars occur in the same habitat in most taxa. Exceptions include flies that demonstrate hypermetamorphosis, which is characterized by an active, slender first instar (planidium) and grublike, endoparasitic later instars. Acroceridae, Nemestrinidae, and Bombyliidae are among the better known groups with hypermetamorphic representatives. In general, the duration of the first larval stage is shortest, whereas that of the last instar is much greater, often several weeks or even months.

Habitat

The diversity of Diptera habitats is partly a reflection of the different ecological roles of larvae and adults, with larvae generally adapted for feeding and growth and adults for reproduction and dispersal. Whereas fly larvae occur in both terrestrial and aquatic habitats, virtually all adults are terrestrial and capable of flight. Wingless and, therefore, flightless groups include certain tipulids, marine chironomids, and phorids, as well as ectoparasitic adults of Hippoboscidae and Nycterobiidae. Adult flies are arguably one of the most aerial of organisms. Swarms of flies, which usually consist primarily of males, are a common sight in many areas. These aggregations, often for the purpose of enhancing male visibility to prospective female mates, may be seen along roadsides, over certain trees or bushes, above sunlit pools along streams, at the summits of hills, in sunny gaps of forest canopies, or at any number of other swarm markers. Swarming is probably a primitive feature of Diptera, which might explain the prevalence of this behavior in nematocerous groups. These Diptera and other flies share a number of structural features that might be adapted for

swarming, including enlarged compound eyes and wings with well-developed anal lobes. These features and others are thought to assist flies in both maneuvering in flight and perceiving conspecific individuals in swarms. Swarming and related behaviors are especially developed in Bibionidae and Empididae. Males of the latter group are known for their predaceous habits and the elaborate behaviors and "nuptial gifts" for prospective female mates. Other groups (e.g., Bombyliidae, Syrphidae) are among the most agile flying insects, being particularly adept at hovering.

Diptera larvae have colonized a variety of terrestrial and aquatic habitats, including water (e.g., Simuliidae, Culicidae, Chironomidae), soil and damp sediments (e.g., Tipulidae, Ceratopogonidae, Tabanidae), rotting wood (e.g., Tipulidae, Mycetophilidae,), fruit (e.g., Tephritidae), decaying organic material (e.g., Muscidae, Sarcophagidae), and the tissues of living organisms (e.g., Sciomyzidae, Oestridae, Tachinidae). Despite this diversity of habitats, most larvae are in a broad sense aquatic. Even "terrestrial" groups from decomposing vegetation, carcasses, leaf litter, rotting wood, or soil often live in a rather aqueous environment. This requirement for a damp environment partly reflects that the larval cuticle is usually thin, soft, and susceptible to drying. Truly aquatic larvae occur in coastal marine, saline, and estuarine waters, shallow and deep lakes, ponds, cold and hot springs, plant cavities (phytotelmata), artificial containers, slow to torrential streams, groundwater zones, and even natural seeps of crude petroleum! Aquatic habits are most prevalent in larvae of nematocerous flies, including all or most Culicidae, Simuliidae, and Chironomidae. Among brachycerous flies, aquatic habits are most common in ephydrids, sciomyzids, and tabanids. In some groups, such as muscoid flies, only a few species are aquatic.

Trophic Relationships

Their trophic diversity and numerical abundance make the Diptera an important component in many ecosystems, both as primary consumers and as a food resource for other organisms. Trophic diversity is reflected in the wide range of larval feeding habits, which encompass nearly every category. In some groups (e.g., asilids, most empidids), larvae and adults belong to the same trophic category; in other groups (e.g., simuliids, tachinids) these life stages usually adopt different feeding strategies; in still others, feeding can be restricted to only the larvae or adults (e.g., chironomids, hippoboscids, and nycteribiids). The latter comprise primarily the so-called Pupipara, in which the females are hematophagous and do not lay eggs and instead give birth to fully formed larvae (i.e.,viviparous development). In addition to the above-mentioned variety of feeding habits, some groups may feed on multiple food resources during the same life stage (e.g., larvae that can be both saprophagous and predaceous and adults that are both nectarivorous and hematophagous). Larval sciomyzids may feed on dead or living mollusks, and some ephydrid larvae may consume algal, bacterial, or detrital resources during the same instar.

Saprophagous habits are among the most prevalent in Diptera, especially in brachycerous groups. Many fly larvae feed on decaying organic material or organic detritus, in which the resident bacteria and other microorganisms are the primary source of nutrition. Among the more common sources of these materials are animal carcasses, which are frequently colonized by callphorids, muscids, phorids, sphaerocerids, and others. The sequence of colonization is often quite predictable, which contributes to the use of Diptera in forensic studies. Decaying fruit and vegetable material also is colonized by many groups, including especially otitid, sphaerocerid, and muscid flies. Decomposing plant fragments can be an important food resource in aquatic habitats, where it is consumed by the larvae of tipulids, ephydrids, otitids, and other groups. These groups and others (e.g., Psychodidae, Syrphidae, Stratiomyiidae) also contain many species that feed on decaying, fine organic matter and associated microorganisms. Most Culicidae and Simuliidae consume fine particulate organic matter of varying size and quality, but use modified mouth-brushes or labral fans to extract particles from water. In most other saprophagous groups, including aquatic species, a sieve-like pharyngeal filter is used to concentrate microorganisms and other organic particles, whereas those feeding on carrion have well-developed mouthhooks for shredding and macerating raw meat.

Phytophagous groups, which consume live plants (including algae and fungi), are well represented by the larvae of bibionids, cecidomyiids, mycetophilids, tipulids, phorids, tephritids, and agromyzids. Many of these flies can be serious agricultural pests. Aquatic habitats contain numerous flies that consume the thin films of algae and organic matter that occur on rocks and other substrata. Among the more obvious of these aquatic grazers are blepharicerids and certain species of psychodids, simuliids, and ephydrids.

Most predaceous Diptera attack other invertebrates as their primary food. Many families (e.g., Chironomidae, Culicidae, Tipulidae, and Ephydridae) contain a few predaceous species, whereas other groups (e.g., Ceratopogonidae and nearly all noncyclorrhaphan Brachycera) feed primarily or exclusively on invertebrates. Vertebrate prey (frogs and salamanders) can be part of the diet of larval Tabanidae. Whereas predaceous larvae typically kill multiple hosts, parasitic and parasitoid larvae generally attack only one host. Parasitoids typically will kill that host, often after a long association with it. Twenty-two families of Diptera include parasitoid members, with tachinid flies perhaps the best known of these. Dipterans are parasitoids of other invertebrates, mostly other arthropods. Because other insects (some pests) often are attacked, parasitoids often are useful for biological control. The Diptera also includes several true parasites, which attack but do not kill the host, such as oestrids and various other groups that often exhibit distinct and complex migrations in vertebrate hosts.

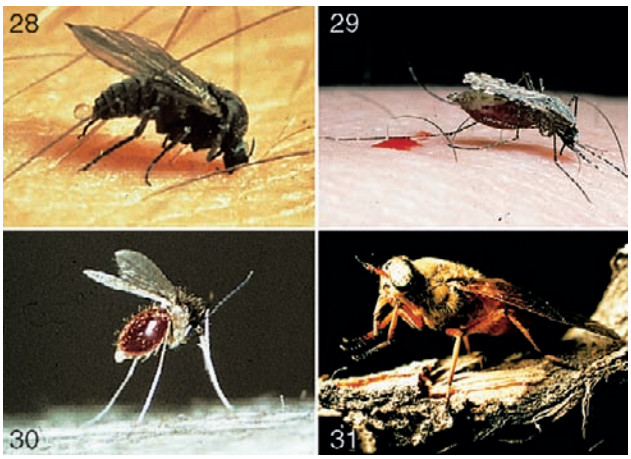

FIGURES 28–31 (28) Female black fly adult (Simuliidae) taking a blood meal. (29) Female mosquito adult (Culicidae: *Anopheles*) taking a blood meal. (Photographs by R. W. Merritt.) (30) Female sand fly adult (Psychodidae) taking a blood meal. (Photograph by B. Chaniotis.) (31) Female horse fly adult (Tabanidae). (Photograph by R. W. Merritt.)

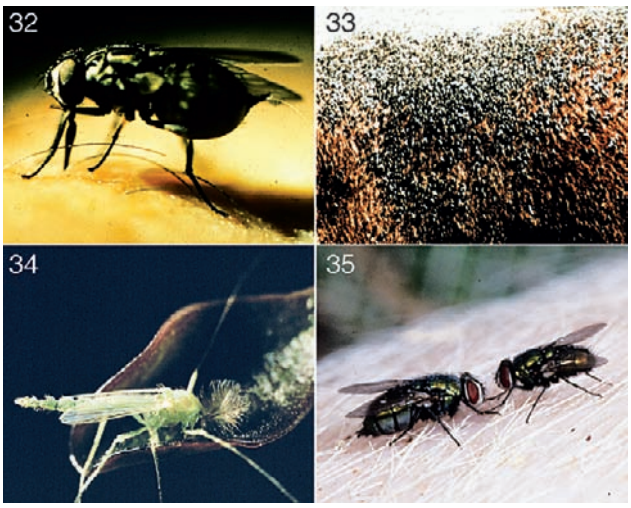

FIGURES 32–35 (32) Female stable fly adult (Muscidae: *Stomoxys calcitrans*) taking a blood meal. (Photograph by E. Hansens.) (33) Horn flies (Muscidae: *Haematobia irritans*) resting and feeding on the back of a bull. (Photograph by R. W. Merritt.) (34) Adult male midge (Chironomidae). (Photograph by R. F. Harwood.) (35) Adult blow flies (Calliphoridae: *Phaenicia sericata*) on a pig. (Photograph by M. J. Higgins.)

ECONOMIC IMPORTANCE

Injurious Families

Several families of Diptera are of major economic importance and involved in the transmission of more disease pathogens to humans and other animals than any other group of arthropods. Biting flies cause annoyance that impacts tourism, recreation, land development, and industrial and agricultural production, whereas their effects on livestock can cause reduced milk, egg, and meat production.

The adults have mouthparts that have very effective piercing stylets, enabling these flies to "bite" and suck blood. Some major families with this characteristic include members of Simuliidae (Fig. 28), Culicidae (Fig. 29), Psychodidae (Fig. 30), Ceratopogonidae, Tabanidae (Fig. 31), and the blood-sucking Muscidae (Figs. 32 and 33). The bites from these groups can often cause severe allergic reactions, resulting in intense itching, rashes, and local swelling or, in some instances, hospitalization as a consequence of toxemia or anaphylactic shock.

Some of the major human and other animal diseases resulting from the transmission of causative organisms by Diptera include human onchocerciasis (river blindness) by Simuliidae; leishmaniasis (sand fly fever) by phlebotomine sand flies belonging to the family Psychodidae; several protozoan and viral diseases of domestic and wild animals, poultry, and waterfowl by Simuliidae and Ceratopogonidae; malaria, yellow fever, filariasis, dengue, dog heartworm, the encephalitides, and related viral diseases by Culicidae; and tularemia and animal trypanosomiases by Tabanidae. Several other species belong to the blood-sucking muscoid flies and include the tsetse fly of Africa, responsible for transmitting the pathogen causing human sleeping sickness, and the stable fly (Muscidae) (Fig. 32), whose vicious bites can annoy humans in recreational areas, and bother domestic animals such as horses, cattle, and sheep. The horn fly (Muscidae) (Fig. 33) is a well-established biting cattle pest throughout many tropical and temperate areas of the northern hemisphere, whereas its close muscoid relative, the buffalo fly, is particularly important to cattle and dairy industries of Australia.

In addition to the biting habits and disease agent transmission of the above groups, flies can cause annoyance and interference with human comfort. Members of the genus *Hippelates* in the family Chloropidae are referred to as "eye gnats" because they frequently are attracted to the eyes of the victim, feed on secretions, and may assist in the entrance for pathogenic organisms. A muscoid fly having similar habits, known as the "face fly," has been associated with the transmission of "pink eye" to cattle. Several other species of muscoid flies (e.g., house fly, bush fly, latrine fly) generally breed in excrement and at times can be economically important pests of humans and/or domestic animals. Two families of Diptera that can cause annoyance and constitute a nuisance by their sheer numbers emerging from ponds and lakes are the Chironomidae (nonbiting midges) (Fig. 34) and the Chaoboridae (chaoborid gnats). These are commonly mistaken for mosquitoes (Culicidae), but do not bite. When one encounters swarms of these midges or gnats, it is difficult to keep them out of one's eyes or avoid inhaling them.

The dipteran families Calliphoridae (blow flies) (Fig. 35) and Sarcophagidae (flesh flies) (Fig. 36) are the major producers of myiasis, i.e., the infestation of organs and tissues of humans or other animals by fly maggots. The larvae of these groups feed on necrotic tissue and may accidently be ingested or invade wounds of humans and domestic animals, causing severe discomfort and subsequent secondary infections. The primary and secondary screwworm flies (Calliphoridae)

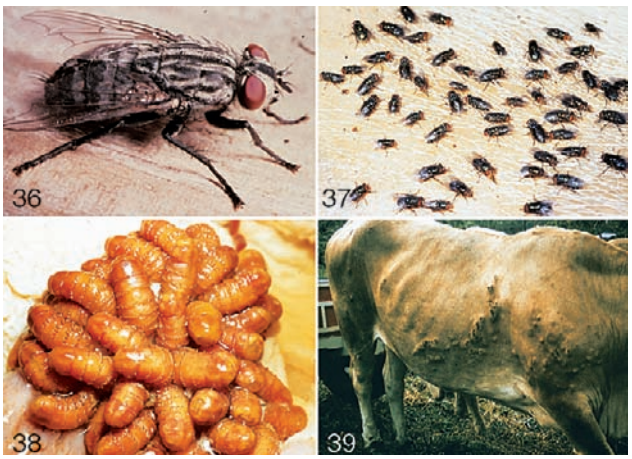

FIGURES 36–39 (36) Adult flesh fly (Sarcophagidae). (Photograph by R. W. Merritt.) (37) Secondary screwworms (Calliphoridae: *Cochlimyia macelleria*) on a pig. (Photograph by M. J. Higgins.) (38) Horse bot fly larvae (Oestridae: *Gasterophilus intestinalis*) attached to the stomach of a horse. (Photograph by R. W. Merritt.) (39) Human bot fly larvae (Oestridae: *Dermatobia hominus*) under the hide of an ox in Costa Rica. (Photograph by L. Green.)

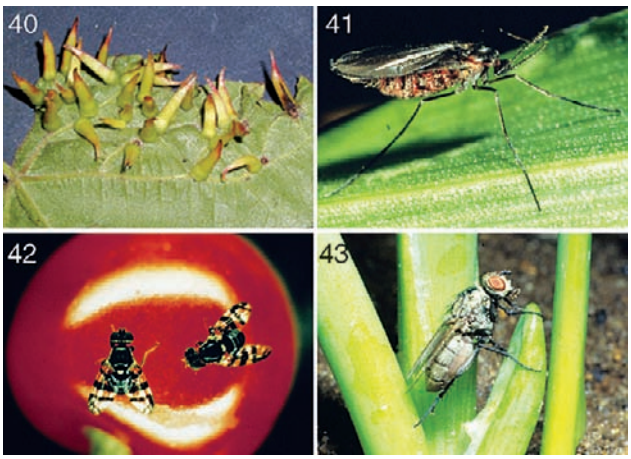

FIGURES 40–43 (40) Cecidomyiid gall on grape leaves. (Photograph by R. Isaacs.) (41) Hessian fly (Cecidomyiidae: *Mayetiola destructor*). (42) Cherry fruit fly adults (Tephritidae: *Rhagoletis cingulata*) on cherry. (Photographs by Department of Entomology, Michigan State University.) (43) Onion maggot adult (Anthomyiidae: *Delia antiqua*). (Photograph by J. Spencer.)

(Fig. 37) are attracted to the wounds and sores of animals, and the former was one of the most serious pests of livestock in the United States until it was eradicated through the sterile male release program. In recent times, the identification and aging of the larvae of some species of blow (Calliphoridae) and flesh flies (Sarcophagidae) have proved useful in establishing the time of death in forensic investigations.

One other family, the Oestridae (cattle, sheep, horse, human, and rodent bot flies), is involved in enteric myiasis of animals and sometimes humans. Damage caused by horse bots (*Gasterophilus* spp.) (Fig. 38) varies from violent reactions by horses due to the flies ovipositing, to irritation by larvae when burrowing into the oral tissue and susequent interference with digestion. The larvae of cattle grubs (*Hypoderma* spp.) migrate through the host's body and eventually reach the upper back where they cut a small opening in the hide and remain there for some time. Economic losses in cattle result from reduction in milk production, weight loss, and damage to hides. Another species of bot fly, the human bot or torsalo *(Dermatobia hominis),* is common in parts of Mexico and Central and South America. It parasitizes a wide range of hosts, including humans, but is a more serious pest of cattle and oxen in these areas (Fig. 39).

Several families of Diptera are economically important to agriculture. The Cecidomyiidae or gall gnats "sting" the plant and make it grow a "gall home" for them (Fig. 40), within which they find not only shelter but also adequate and abundant food. Examples are the goldenrod ball gall and the pine cone gall. Some very destructive species in this family, such as the Hessian fly (Fig. 41), chrysanthemum gall midge, and wheat, pear, and cloverseed midge, feed on cultivated crops and do not always form galls. The Tephritidae, or fruit flies, contain some species whose larvae bore into the stems of plants; some produce galls, others are leaf miners, and most important of all are those that bore into the flesh of fruits and vegetables. The latter include some of the most important of all economic insects, specifically the apple maggot, cherry fruit flies (Fig. 42), walnut husk fly, and Mexican, Mediterranean, oriental, olive, and melon fruit flies. The Anthomyiidae, or root maggot flies, have larvae that feed on decaying vegetable matter from which a number have adopted the habit of attacking the roots of vegetables. These include the cabbage maggot, onion maggot (Fig. 43), seed corn maggot, and spinach leafminer. Larvae of the family Agromyzidae are known as leafminers and feed between the leaf surfaces, leaving light-colored, narrow, winding mines or large blotches that decrease photosynthesis and make produce unsalable. The leaves are weakened and the mines promote disease and decay.

Beneficial Families

The Diptera contain several families that can be considered beneficial to humans and their environment. First, and most important, is the role of all Diptera in food chains in nature. Groups such as Culicidae, Chironomidae, and Simuliidae occur in large numbers as larvae and adults and provide a major prey base for many other invertebrates as well as vertebrates such as fish, birds, bats, and amphibians. In turn, several families contain predators and parasitoids as larvae and adults, including the Asilidae, Empididae, Dolichopodidae, Syrphidae, and Tachinidae. Many families are important decomposers and recyclers of decaying organic matter of different types. Examples include the Psychodidae, Tipulidae, Stratiomyiidae, Mycetophilidae, Sciaridae, Sepsidae, Coleopidae, Muscidae, Calliphoridae, Sarcophagidae, Phoridae, Syrphidae, and Sphaeroceridae. Some Diptera are

important pollinators of flowers and include some species of Syrphidae, Bombyliidae, and even adult male Culicidae who visit flowers to imbibe nectar.

Some families of aquatic Diptera have been important in water quality and bioassessment studies to classify the degree of pollution in a water body. For example, larvae belonging to the midge genus *Chironomus* in the family Chironomidae have been referred to as blood worms because of the hemoglobin in their blood. These and another group known as the "rat-tailed maggots" (Syrphidae: *Eristalis*) are often used as indicators of polluted water or water low in oxygen. The presence of Simulidae in a stream generally indicates clean well-aerated water. The Culicidae and Chironomidae have members that are associated with both polluted and clean water habitats. Finally, some Diptera have been the subject of study for scientists throughout the world. For example, chironomid midges are used in acute and chronic laboratory toxicity studies to compare toxicants and the factors affecting toxicity and to ultimately predict the environmental effects of the toxicant. The small fruit fly, *Drosophila* (Drosophilidae) (Fig. 44), has been the organism of choice in most genetic studies for years and has contributed significantly to studies ranging from neurobiology to evolutionary theory. Overall, the Diptera represent an order containing a variety of species that are economically very beneficial and equally injurious to humans.

BIOLOGY OF SELECTED FAMILIES

Suborder Nematocera

TIPULIDAE Crane flies (Fig. 45) are a diverse group of 14,000 species that inhabit a variety of freshwater and terrestrial habitats. Larvae are significant shredders *(Tipula, Pedicia)* of leaves that enter streams and are predators *(Hexatoma, Dicranota)* in aquatic habitats. The moist transition zone between aquatic and terrestrial areas supports a distinct assemblage of species (e.g., *Erioptera, Ormosia*). Terrestrial habitats are home to species that feed on coniferous *(Limonia)* and deciduous *(Epiphragma)* rotting logs or decaying organic material *(Tipula)* and that may even be pestiferous consumers of sod *(Tipula)*. A few species can tolerate high salinity and inhabit the rocky intertidal zones of marine habitats. The adults generally do not feed, although they are frequently mistaken for "giant mosquitoes." A few taxa possess a long proboscis that presumably allows nectar feeding. Large and gangly, adult crane flies are easily taken by vertebrate predators such as birds.

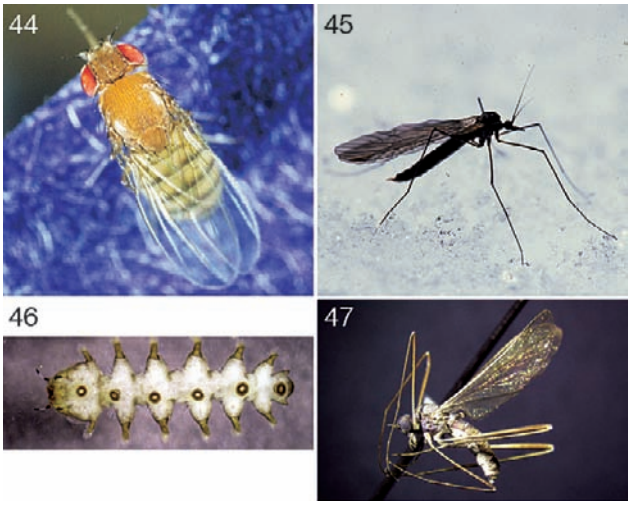

FIGURE 44–47 (44) Adult female small fruit fly (Drosophilidae: *Drosophila*). (Photograph by R. D. Akre.) (45) Adult crane fly (Tipulidae). (Photograph by Department of Entomology, Michigan State University.) (46) Ventral view of larva of Blephariceridae showing suctorial discs. (47) Adult net-winged midge (Blephariceridae). (Photographs by G. W. Courtney.)

PSYCHODIDAE Sand flies (Fig. 30), drain flies, and moth flies are typical representatives of this family and contain 2500 species. Adult sand flies *(Phlebotomus)* are tropical hematophagous (blood-feeding) flies that can transmit leishmaniasis, a disease caused by parasitic protozoa spread by sand-fly bites. However, most psychodids do not bite and are harmless to humans and livestock. Drain flies (*Psychoda* and *Telmatoscopus*) and moth flies *(Psychoda)* resemble tiny moths (about 2–4 mm in length) with hairy, pointed wings. The former have larvae that develop on the rich organic material that builds up in domestic pipes and drains and can be abundant in households and public restrooms. Moth flies have aquatic to semiaquatic larvae that breathe atmospheric oxygen by maintaining contact with the atmosphere using hydrofuge hairs on their posterior spiracles. Eutrophic lakes, marshes, and wastewater treatment plants may produce large numbers of adults. As detritivores, the larvae of moth flies probably are significant nutrient recyclers in lentic ecosystems.

BLEPHARICERIDAE The net-winged midges (300 species) have peculiar larvae (Fig. 46) that use ventral suckers (suctorial disks) to maintain their positions on rocky substrates in torrential streams. A hydraulic, piston-like apparatus gives the larvae the ability to generate suction that allows their suckers to work—even waterfall habitats are occupied by blephariceridae larvae. The mouthparts are positioned ventrally on the head capsule and are specialized for scraping thin algal films off of rocks within fast-flowing environments. Diatoms and other unicellular algae are most often consumed, but fungi and bacteria may also be included in the larval diet. Pupae are also firmly attached to rocks within the flow with permanent suction pads. The adult (Fig. 47) will emerge and maintain a brief grip on the attached pupal skin as the exoskeleton hardens prior to flight. Little deviation from these habits has been documented within the Blephariceridae. Adults are known as net-winged midges because of the finely divided venation of the wings.

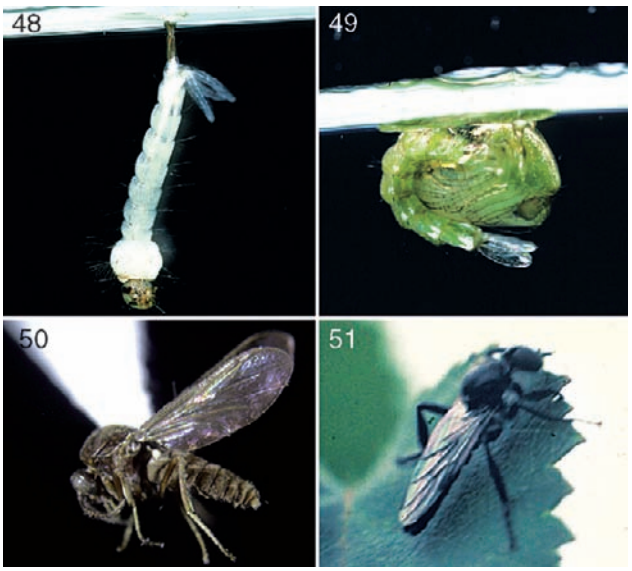

FIGURES 48–51 (48) Larva of mosquito (Culicidae: *Aedes aegypti*). (49) Pupa of mosquito (Culicidae: *Anopheles quadrimaculatus*). (Photographs by R. W. Merritt.) (50) Adult Ceratopogonidae. (Photograph by G. W. Courtney.) (51) Adult march fly (Bibionidae). (Photograph by Department of Entomology, Michigan State University.)

CULICIDAE Mosquitoes (3000 species) (Fig. 29) are well-recognized for their roles in disease agent transmission and as pests to humans, livestock, birds, and a variety of other vertebrate hosts. However, adults may emerge in high numbers and provide ample food for avian, bat, and certain predatory invertebrate populations. Mosquitoes exhibit the ability to colonize new aquatic habitats quickly and can survive in confined container habitats. In terms of mosquito control, the ecological importance of the larvae, pupae, and adults is rarely considered. The larvae are mostly filter feeders, but some scrape organic material and algae from solid substrates in standing water habitats. The clearance rate of particles from standing water is impressive and may alter the characteristics, such as turbidity, of the water the insects inhabit. Larval populations are a major component of the neuston, or water-surface inhabitants, and maintain contact with the atmosphere with their spiracles. Larvae (Fig. 48) are known as "wrigglers" because of their frantic swimming action that allows them to dive when threatened; lessening of light intensity by a mere shadow will initiate the wriggling action in some *Culex,* making them difficult to collect. Some taxa, such as *Culex, Culiseta,* and *Aedes,* have their spiracles positioned apically on respiratory siphons; others, such as *Anopheles,* lack this breathing-tube apparatus. *Mansonia* and a few other genera possess siphons that are specialized for piercing the roots of wetland plants such as cattails to obtain oxygen and therefore do not need to come to the water surface to breath. The mosquito pupa (Fig. 49) is free-swimming with respiratory trumpets that allow individuals to obtain atmospheric oxygen; pupae are known as "tumblers" because of their tumbling action that propels them below the surface when disturbed. Emergence occurs quickly at the water surface as the pupal skin breaks to liberate the adult. Although females feed on sugar sources and may take a blood meal for the purpose of egg production, males feed only on nectar and lack bloodsucking proclivities.

CERATOPOGONIDAE This family is known as biting midges, punkies, and no-see-ums and contains 5500 species. The adults (Fig. 50) are minute bloodsuckers that swarm around mammalian hosts, including humans (e.g., *Culicoides, Lasiohelea*). Certain taxa also feed on other invertebrates as ectoparasites, including crane flies, dragonflies, and mantids. The tiny black to gray adults frequently have darkly patterned wings and relatively long antennae. Larvae are encountered in a variety of standing water habitats, including saturated mud and sand, tree holes *(Dasyhelea)*, rain pools, marshes, lakes, and even hot spring algal mats *(Bezzia)*. The genus *Leptoconops* can be pestiferous and biting adults are encountered at ocean-side beaches. The larval feeding habits of biting midges consist mostly of scavenging and predatory behavior.

SIMULIIDAE Although the general public is often aware of the pest nature of mosquitoes, knowledge of blood feeding by black flies (1500 species) is often restricted to anglers and those who recreate within or near aquatic systems. Like mosquitoes, the larvae play an important role as filter feeders; however, black flies are restricted to flowing water systems. Larval simuliids spin a patch of silken webbing on the surface of riffle rocks and maintain a hold on the webbing with hooks positioned on the posterior abdominal segment. The mouthparts are modified in many species (e.g., *Simulium, Prosimulium*) and resemble head fans that allow the larvae to capture organic particles, including materials as small as bacteria. Rocky substrates below dam spillways where organic-rich water flows may support tens of thousands of larvae per square meter. Other species are more mobile and scrape or collect food materials from benthic substrates (*Gymnopais* and *Twinnia*). The larvae are apneustic (i.e., lack spiracles) and therefore require moving water for cutaneous respiration. Pupae are firmly attached to areas of rocks exposed to current where thoracic pupal respiratory organs (gill-like structures) dangle in a downstream direction, supplementing spiracular respiration (Fig. 8). Most of the pupa is enclosed within a sheath-like cocoon. As in mosquitoes, males have weak mouthparts and may feed on nectar, whereas the females of most species have short probosci and cutting mouthparts for obtaining a blood meal from their vertebrate hosts. The adults are small and grayish black, lack distinct patterns on the wings, and have short antennae (Fig. 28). Some species can swarm in large numbers and are capable of causing shock in domestic animals (e.g., cattle) due to blood loss.

BIBIONIDAE March flies (700 species) are named for their early spring appearance in temperate habitats. The stout, dark-colored adults (Fig. 51) feed on flowers; in contrast, the worm-like larvae are general detritivores and can be found in organic soils and compost heaps in abundance. The common genus *Bibio* overwinters as larvae prior to forming pupae after being exposed to cold temperatures. One predaceous species, *Plecia nearctica,* was introduced to the southeastern United States to control mosquitoes. Although its impact on mosquitoes is questionable, its impact on human residents is very real. The adults appear for brief periods (about 2 weeks) in such large numbers as to smear automobile windshields and clog radiators. The smashed bodies may even damage a car's paint if not washed off quickly. The adults are known as "love bugs" because males and females are frequently seen flying *in copula.*

SCIARIDAE These are known as dark-winged fungus gnats (1000 species) because of their small, gnat-like size and smoky grayish-black wings. *Sciara* is the genus most frequently encountered by people, as the pale, slender larvae develop in a variety of materials, including potting soil used in greenhouses and household planters. In nature, larvae consume the fungus-rich detritus formed under the bark of rotting trees, as well as within the logs themselves. Organic-rich compost heaps and mushrooms are also inhabited by sciarid immatures. Although a few taxa are pests (e.g., *Pnyxia* attacks mushrooms), most species of this common family are harmless.

CECIDOMYIIDAE Gall midges and gall gnats are minute flies that are abundant, species-rich (4500 species), and cosmopolitan. More than 1000 species occur in North America alone, and many undescribed species await taxonomic attention. Most species form distinctive galls within which the maggots develop (Fig. 40). Indeed, it is frequently easier to determine what species is attacking a plant based on gall morphology rather than adult or larval morphology. Many people are quite familiar with the plants that are affected, such as the damage from maple leaf spot *(Rhabdophaga)*, or the attack of the Hessian fly *(Mayetiola destructor)*, which can be a serious pest of wheat. However, the Cecidomyiidae as a family shows impressive breadth in the plant species it attacks. A few species (e.g., *Miastor*) exhibit paedogenesis, whereby the larvae reproduce. The "mother larva" produces a number of larvae within her body, which eventually consume the mother and then escape.

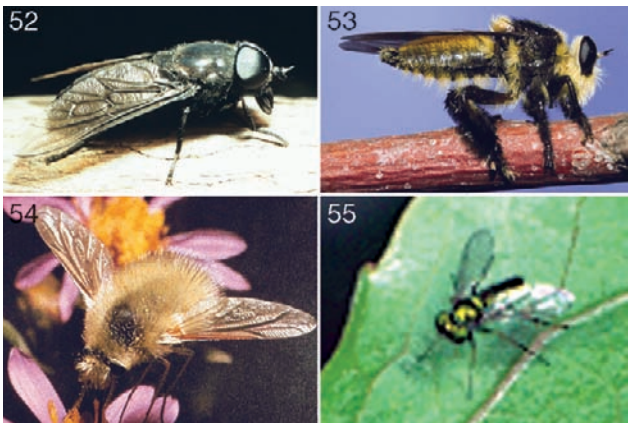

FIGURES 52–55 (52) Adult deer fly (Tabanidae). (Photograph by R. W. Merritt.) (53) Adult robber fly (Asilidae). (Photograph by R. W. Sites.) (54) Adult bee fly (Bombyliidae). (Photograph by R. D. Akre.) (55) Adult long-legged fly (Dolichopodidae). (Photograph by Department of Entomology, Michigan State University.)

Suborder Brachycera

TABANIDAE The horse flies (e.g., *Tabanus*, *Hybomitra*) (Fig. 52) and deer flies *(Chrysops, Silvius)* (Fig. 52) contain 3000 species and are a familiar insect group to people who frequent rural outdoor areas. The adults are rapid fliers; one species was estimated to fly over 150 km per hour! Eggs are normally laid in masses, frequently on vegetation overhanging water or saturated soils. The cryptic larvae are restricted to aquatic and semiaquatic habitats where most species are predators of other invertebrates. The life cycle generally takes about 1 year to complete, whereas some of the larger horse flies require up to 3 years. Although most species inhabit stagnant habitats, some are found at the margins of streams. Females are blood-feeders and may inflict a painful bite. Rather than puncturing a host's skin and sucking blood like mosquitoes, tabanids create a laceration on the host's skin and quickly lap up the pooling blood before retreating. The attack on livestock can be so severe as to reduce milk yields in dairy cattle. Like many other families of biting flies, females use visual cues to locate hosts and also sense plumes of carbon dioxide produced during vertebrate respiration. Horse flies tend to be large (about 10–25 mm) with nearly colorless or smoky wings, whereas deer flies are smaller (around 8 mm) and have yellow or black bodies that support darkly patterned wings. Human disease transmission by Tabanidae (i.e., tularemia, anthrax) is possible, but not significant in North America. However, transmission can be significant in other areas of the world (e.g., Africa).

RHAGIONIDAE The snipe flies (500 species) superficially resemble some deer flies, but have a more slender body. Most common in woodlands, snipe flies are often dull yellow to brown (e.g., *Rhagio*), but the gold-backed snipe fly *(Chrysopilus ornatus)* of eastern North America has brilliant gold hairs adorning the thorax and abdomen. Most adults are nectar feeders, whereas a few taxa are predators of flying insects. Larval rhagionids tend to be predators of small invertebrates within masses of rotting wood, organic-rich soil, or compost. One genus *(Symphoromyia)* of western North America has blood-feeding adults that will bite humans in woodland areas.

MYDIDAE The largest adult Diptera are the mydas flies, with 400 species. Some tropical species are over 50 mm. The adults are dark and have red to yellowish coloration on some abdominal segments. Little biological information is available on this family, although the larvae are predators found in decaying logs in woodlands. Pupae occur a few centimeters below the soil, and are adorned with heavy spikes for digging to the surface just prior to adult emergence. The adults are also thought to be predators that specialize in capturing other flying insects, but a fair number of species have vestigial mouthparts. The females of the latter may simply live on the accumulated fatty tissue in the abdomen.

ASILIDAE The robber flies (5500 species) occur in a vast number of terrestrial habitats; most adult activity occurs in areas that are sunny or at least partially sun lit. Adults (Fig. 53) may reach approximately 30 mm in length (e.g., *Proctacanthus*), whereas others are less than 10 mm in length (e.g., *Holocephala*). There is great morphological variation in this family among adults, but all species share in common a conspicuously sunken vertex. Adults are predators that are able to take larger prey such as dragonflies, but the selected prey size varies among species. The type of prey, whether stationary, crawling, or flying, is also species-specific among robber flies. The mouthparts contain a stout proboscis that the adult uses to exsanguinate prey species. Some species (e.g., *Laphria*) mimic bumble bees, which reduces predatory attempts on the adult. Most larvae live in soil or rotting wood where they hunt other insect larvae and nymphs; however, some species are ectoparasitic on Diptera, Coleoptera, Hymenoptera, and Orthoptera immatures. Very few life history studies have been done on the Asilidae.

BOMBYLIIDAE Bee flies (5000 species) are stout, hairy-bodied flies and, as the name implies, adults are frequently mistaken for hymenopterans because of their bee-like appearance (Fig. 54). Furthermore, the adult behavior often involves hovering at flowers, beelike, and extending a long proboscis to obtain nectar while in flight! Some taxa have bold patterns on the wings (e.g., *Anthrax, Exoprosopa*) or have the anterior margin of the wing darkened (e.g., *Bombylius*). Although a widespread family, most species occur in arid areas. The biology of the immature stages remains unknown for most species, but it appears that all species for which the larval feeding habits are known are parasitic on Diptera, Lepidoptera, Hymenoptera, Coleoptera, and Neuroptera larvae or pupae. Many larvae have relatively large, tong-shaped mandibles, presumably suited to aid the parasitic life history. A few bombyliid species consume grasshopper eggs.

DOLICHOPODIDAE The long-legged flies (5000 species) are small to minute flies that are often brilliant green, blue, or copper colored (Fig. 55). Males have genitalia that are nearly as long as the other abdominal segments combined (e.g., *Dolichopus*). Adults participate in complex courtship rituals, and males of some species have legs adorned with flattened hair-like scales used as flags to communicate with females during courtship. The family is impressively diverse in its habitat use, as adults can be abundant in freshwater marshes and lake edges, stream margins, woodlands, open fields, and coastal marine areas. Adults appear to be exclusively predaceous. Larvae have been taken from water, damp soil, grass stems, under bark, and other places. Most taxa are predaceous, but a few (e.g., *Thrypticus*) are phytophagous. One genus *(Medetera)* has predaceous larvae that feed on bark beetles.

Suborder Cyclorrhapha

PHORIDAE This family, also known as humpbacked flies and scuttle flies (3000 species), is another group of flies that exploits a wide range of habitats and exhibits diverse feeding habits. The humpbacked appearance and reduced venation make the adults easy to identify. Many species are consumers of decaying organic matter and can infest household garbage cans on occasion; the females are strongly attracted to the odor of decay. Other species are more unusual, specializing on the consumption of slug eggs *(Megaselia)* or parasitic on spiders, millipedes, and at least nine insect orders. Some species are currently targeted as potential biocontrol agents of fire ants, a serious pest in the southern United States. One species is known as the coffin fly *(Conicera tibialis)* because it was reported to maintain many generations on a single human body in the confines of a buried casket.

SYRPHIDAE Like the bee flies, flower fly adults (6000 species) resemble Hymenoptera and can mimic bees, bumble bees, hornets, and others (Fig. 56). Syrphids have the ability to hover (thus, they are also known as hover flies), and

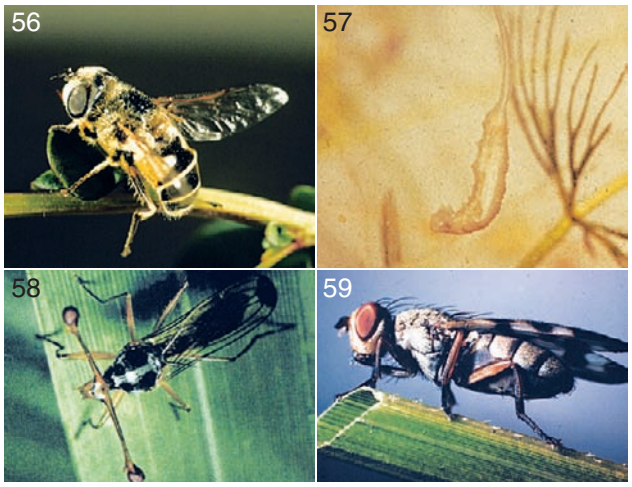

FIGURES 56–59 (56) Adult flower fly (Syrphidae: *Eristalis* sp.). (Photograph by R. F. Harwood.) (57) Rat-tailed maggot (Syrphidae). (Photograph by R. W. Merritt.) (58) Adult stalk-eyed fly (Diopsidae). (Photograph by R. D. Akre.) (59) Adult picture-winged fly (Otitidae: *Melieria similies*). (Photograph by J. A. Novak.)

individuals are frequently found flying near flower heads where they obtain nectar. Adults are common near wetlands and lakes (e.g., *Eristalis, Allograpta*) but are also abundant in terrestrial areas (e.g., *Merodon, Syrphus*) where appropriate flowering vegetation grows. In aquatic habitats the larvae, or rat-tailed maggots, as they are called, are collector–gatherers and may use retractable siphons for respiration (Fig. 57). Emergent vegetation, or vegetation at the aquatic–terrestrial interface, may be infested with aphids and other homopterous herbivores that certain flower fly larvae devour upon discovery. Some aphids enter plant stems compromised by boring larvae of other orders (e.g., Lepidoptera) to feed on decaying plant juices; it is not uncommon to find flower fly larvae that have also entered the damaged areas of plants to obtain prey. Predatory habits are also seen in terrestrial habitats, where larvae of certain species will inhabit dung, rotting logs, and decaying vegetation, as well as the exterior of plants. Overall, larval habitats are diverse, including dung, rotting cactus, peat, and hymenopteran nests. Some species have also been implicated in intestinal myasis.

DIOPSIDAE The stalk-eyed flies (Fig. 58) are one of the most aptly named and morphologically unusual dipteran families (150 species). Each eye and its antenna are positioned at the end of individual stalks that protrude laterally from the head; the distance from eye to eye may be approximately equal to the entire body length! North America's one species *(Sphaerocephala brevicornis)* has very short eye stalks and breeds in decaying organic matter. The adults of this species exhibit no particular courtship displays, whereas highly adorned males of Afrotropical species *(Diopsis)* battle with one another using their stalks as levers during aggressive "wrestling matches." The larvae of *Diopsis* are herbivorous, and some species develop within the stems of rice plants.

OTITIDAE These flies are also known as picture-winged flies because their boldly patterned wings are used in courtship and species recognition (Fig. 59). Adults are commonly found walking along vegetation flashing their wings. Otitids are abundant in both aquatic and terrestrial habitats (800 species). In marshes and vegetated lake margins, picture-winged flies are herbivores *(Eumetopiella),* secondary invaders of damaged plants *(Chaetopsis),* and general scavengers *(Seioptera).* Herbivory also occurs in terrestrial species *(Tetanops, Tritoxa),* but scavenging of decaying organic material appears to be more common *(Delphinia, Euxestra, Notogramma).* Some species also attack fungi *(Pseudotephritis).*

PYRGOTIDAE Although some species of certain fly families (e.g., Tipulidae) are attracted to collecting lights at night, pyrgotid flies are unusual in that they are exclusively nocturnal. These flies (200 species) are relatively large and usually have strongly patterned wings (Fig. 60). Adults (e.g., *Pyrgota, Sphecomyiella)* seek scarab beetles, most notably June beetles, and apparently attack flying beetles by laying a single egg on the dorsum of the thorax or abdomen that is exposed when a beetle's elytra and wings are spread. The larva hatches from the egg and burrows into the body, acting as a parasitoid. The feeding larva eventually kills the host and consumes the remaining tissue. Larvae pupate within the hollowed host, and the adult exits the beetle exoskeleton to continue the life cycle.

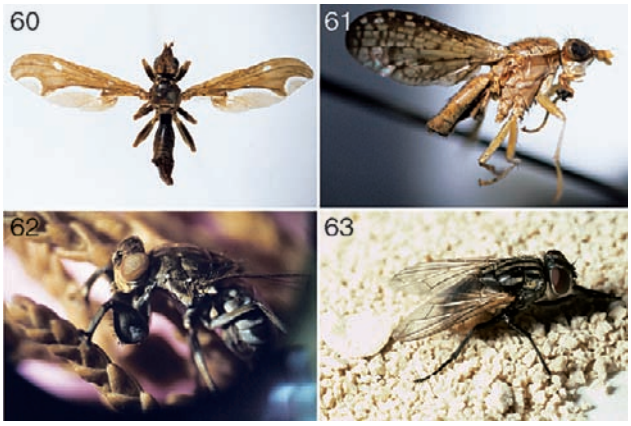

FIGURES 60–63 (60) Adult of Pyrgotidae. (61) Adult marsh fly (Scromyzidae: *Limnia*). (62) Adult shore fly (Ephydridae: *Ochthera mantis*). (Photographs by The Cleveland Museum of Natural History.) (63) House fly adult (Muscidae: *Musca domestica*). (Photograph by R. W. Merritt.)

TEPHRITIDAE These true fruit flies (4000 species) are essentially entirely terrestrial in their habitat selection, although the host plants exploited by the family sometimes grow at the margins of lakes and marshes. Adults (Fig. 42) oviposit on the flower heads of the plant family Compositae or on fleshy fruits. Like the Otitidae, the wings of most adults are distinctly patterned, and adults flash their wings during courtship; this behavior has earned the Tephritidae a second common name, "peacock flies." Species tend to be fairly specific in their host plant preferences or at least attack a narrow spectrum of plant taxa. Fruit fly species are also specific in the area of a plant that they infest. Some species are frugivorous *(Ceratitis, Rhagoletis),* seed-head predators *(Euaresta, Trupanea, Tephritis),* gallmakers *(Eurosta),* or leafminers *(Euleia).* Frugivorous larvae damage the host fruit, causing it to rot quickly; seed predators of select young, developing seeds. Galls may be formed on a variety of plant areas, including stems, leaves, and flower heads.

DRYOMYZIDAE These are relatively uncommon flies (300 species), with the biology of only 2 of the 8 North American species known. *Dryomyza anilis* is a scavenger and breeds in decaying mammalian carcasses; it can be reared on raw ground beef. A contrasting life history is found in *Oedoparena glauca,* which preys on barnacles in the intertidal zone of western North American shorelines. This character makes *O. glauca* one of the truly marine insects, as it is tied intimately to an ocean-inhabiting invertebrate. Adults lay their eggs into the barnacle's operculum when dropping tide levels expose them. Larvae consume the soft tissue, and

mature larvae frequently move to new barnacles to continue feeding. Pupariation occurs within the final host.

SEPSIDAE The black scavenger flies are fairly abundant in both aquatic *(Enicomira, Themira)* and terrestrial *(Sepsis)* environments (250 species), where the larvae are scavengers of decaying organic matter. Dung of a variety of mammalian animals, carcasses, rotting snails, and washed-up seaweed have been exploited. The adults of many species are easily recognized by their rounded heads and the presence of a black dot at the apex of each wing.

SCIOMYZIDAE Called marsh flies and snail-killing flies, neither name encompasses all the habits of this well-studied family (500 species). Some species (e.g., *Dictya, Limnia)* are found in marshes, but some species of certain genera (e.g., *Sciomyza, Pherbellia)* are fully terrestrial. Many species are larval parasitoids or predators of snails, and some attack slugs (e.g., certain *Tetanocera, Euthycera*) or fingernail clams *(Renocera);* the larvae of one genus *(Antichaeta)* prey on the eggs of aquatic and semiaquatic snails. The adults range in color from yellowish brown to brownish black, have antennae that may be long or short, and vary in size from a few millimeters to nearly 1 cm (Fig. 61). However, the trophic niche of exploiting freshwater or terrestrial Mollusca (i.e., snails, slugs, and clams) ties all Sciomyzidae together evolutionarily. Ovipositional habits vary from certain species that lay their eggs directly on the host (e.g., *Sciomyza*) to species that lay eggs on plants, thus requiring larvae to search for their hosts (e.g., *Tetanocera*). Only one species *(Sepedonella nana)* from Africa seems to deviate from the trophic tie to mollusks, as laboratory-reared larvae have fed and survived on aquatic oligochaetes in the laboratory.

CHAMAEMYIIDAE Aphid flies are predators of aphids, mealy bugs, and other homopterous herbivores (250 species). Adults lay their eggs on plant surfaces, in galls, or in the egg sacs of scale insects, and the maggots can reduce homopteran populations *(Leucopis).* One report from Mexico showed that the adults of an aphid fly fed on the secretions of vertebrate animals and that the larvae may have developed in bird nests *(Paraleucopis).* Some taxa are found within emergent and shoreline vegetation of aquatic habitats, whereas others are encountered in woodland or open fields.

PIOPHILIDAE These flies (70 species) are most commonly represented by the cheese-skipper fly *(Piophila casei),* a cosmopolitan consumer of proteinaceous materials. Larvae frequently infest cheese and exhibit the rather peculiar escape strategy of grabbing the posterior body segment with their mouth hooks to form a U shape and then releasing their grip, which causes the larvae to propel, or skip, away from their original location. These behaviors give them the name cheese-skipper, even though the larvae are also known to consume the drying tissues of aging mammalian carcasses and dung. While most species appear to be scavengers of decaying materials and carcasses *(Protopiophila)* or mushrooms *(Amphipogon),* the larvae of some species are parasitic on avians *(Neottiophilum).*

SPHAEROCERIDAE The small dung flies represent a speciose family (2500 species) that have predominantly scavenging larval feeding habits. One genus *(Leptocera)* consistently appears to be scavengers; however, the spatial niches inhabited by the larvae are highly diverse. Larvae have been found in decaying vegetable matter, sewage, dung, dung beetle broods, stranded masses of seaweed, fungi, slime molds, carrion, and the organic matter accumulated within cups of bromeliads. Muddy, organic-rich margins of aquatic habitats, such as marshes and ponds, will support virtual clouds of adults. Some species are quite habitat-specific, such as those that inhabit bogs. The adults of common species have long, stiff bristles dorsally and are black to gray in color, and the arista is several times longer than the other segments of the antenna.

EPHYDRIDAE These flies are also known as shore flies, and most taxa are associated with aquatic habitats. This is one of the most genera-rich families of Diptera (1300 species) and one of the most diverse in feeding habits. Larvae are consumers of decaying organic matter (e.g., *Discocerina*), secondary stem borers of damaged plants (e.g., *Typopsilopa*), primary herbivores (e.g., *Hydrellia*), generalist feeders of algae (e.g., *Scatella*), specialist consumers of algae and cyanophytes (e.g., *Hyadina*), diatom specialists (e.g., *Parydra*), predators (e.g., *Ochthera*), and consumers of spider eggs *(Trimerina).* Virtually all aquatic habitats, from flowing water to stagnant environments, temporary to permanent, fresh water to hypersaline, and cold water to hot springs, are occupied by ephydrids. Terrestrial environments are less likely to support ephydrid populations, but these flies are found in moist woodlands and even in sod from suburban areas. Shore flies can be abundant in human-made habitats, including constructed wetlands and sewage treatment plants. It is difficult to make generalizations about the overall morphology of this family, except that the adults tend to be small; the smallest adults *(Lemnaphila),* only a couple of millimeters across, mine the thalli of duckweed plants. Adult body color ranges from silvery gray to jet black, and the wings are completely colorless to highly patterned with various shades of gray and brown (Fig. 62). The species, habitat, and feeding diversity have led the dipterist Harold Oldroyd to state that the shore flies are currently "in the flower of their evolution."

DROSOPHILIDAE Pomace flies, vinegar flies, and small fruit flies are another highly species-rich family (3000 species). The latter name has led to some confusion as the Tephritidae are also known as fruit flies. The adults (Fig. 44) are small (generally only a few millimeters in length), but can disperse about 10 km in 1 day. Some drosophilids are frugivorous, but a vast array of food sources are used. For example, many feed

on fungi (e.g., *Amiota, Mycodrosophila, Stegana, Scaptomyza*), living flowers (e.g., *Apenthecia, Styloptera*) or are predaceous on other invertebrates (e.g., *Rhinoleucophenga, Cacoxenus, Acletoxenus*). Indeed, the genus *Drosophila,* mostly known for the experimental studies of *D. melanogaster,* in the wild exhibits a vast trophic ecology and includes species that develop in rotting vegetation, rotting fruit, tree sap, fungi, living flowers, and plant stems and that prey on other invertebrates. Two *Drosophila* species are commensal with crabs: larvae live attached to the crab exoskeleton and consume semiliquid excretions from the crab or develop in the crab's branchial chamber and consume its microflora.

CHLOROPIDAE Flies is this family are called chloropid flies or frit flies (2000 species). Larvae are generally scavengers of decaying organic matter, secondary invaders of damaged plants, or primary herbivores. Adults are common and abundant among vegetation in terrestrial and aquatic situations. Species have been reared from dozens of different substrates around the world. Grasslands commonly support populations of certain genera (e.g., *Meromyza, Parectecephala*), and fungi are the sole food of others (e.g., *Fiebrigella, Apotropina*). However, a great diversity occurs in vegetated areas of aquatic habitats (e.g., *Chlorops, Epichlorops, Eribolus, Diplotoxa*) where most taxa are detritivores in decaying masses of vegetation, secondary stem borers, and primary herbivores of aquatic or semiaquatic plants. A few taxa are predatory on Homoptera (e.g., *Thaumatomyia*). Other, less common, larval food sources include dung *(Cadrema),* decaying wood, and bird nest debris *(Gaurax),* and one Australian genus *(Batrachomyia)* is subcutaneously parasitic on frogs and toads. The adults tend to have rounded flagellomeres and range from dull colored to bright yellow or green, and many species have a distinctly shiny triangle positioned at the vertex of the head.

AGROMYZIDAE These are known as the leaf-miner flies because of their highly herbivorous nature. Like the Chloropidae, agromyzids are well represented in both aquatic and terrestrial environments (2000 species). Herbaceous and woody plants are both attacked, but larvae tend to feed on a single host plant, or a narrow spectrum as host plants as leaf miners, stem borers, or seed head predators. Wetland taxa can form large populations in which both monocot and dicot flora are used as host plants (e.g., *Agromyza, Cerodontha, Liriomyza, Phytomyza*). One species *(Melanagromyza diantheraea)* is a specialist stem borer of water willow, a flowering plant found at the edge of streams. The females of this fly lay eggs on the exterior of the plant, and upon hatching the larvae burrow into the stem to initiate feeding. Leaf mines are frequently seen as dead or brown areas on a leaf surface, and mine morphology is sometimes distinctive enough to determine which agromyzid species is responsible for plant damage.

ANTHOMYZIDAE These common flies are an example about which little is known of their biology (50 species). One genus *(Anthomyza)* has small yellowish adults that may feed on the culms of wetland sedges, but it is unclear if they are herbivorous or act as secondary stem borers after plants have been attacked by other herbivorous insects.

MUSCIDAE This large family (4000 species) includes anthropophilic species such as the house fly *(Musca domestica)* (Fig. 63) and the stable fly *(Stomoxys calcitrans)* (Fig. 32). The house fly is well known for its "filthy habits," and the stable fly bites both humans and livestock. The reproductive rate of the house fly is noteworthy, as one female can eventually give rise to 2 billion other female flies after several summer generations are produced (assuming all flies live, which is never the case). A short life cyle (12–14 days required for development from egg to adult in summer temperatures) is at least partially responsible for the success of this species and is necessary for developing in such ephemeral, human-made habitats such as dung heaps, garbage cans, and mammalian road kill. However, most muscid species are not directly associated with human populations. The larval feeding habits found among the Muscidae include herbivory *(Atherigona, Dichaetomyia),* scavenging *(Graphomyia),* and predatory behaviors *(Coenosia, Lispe, Spilogona).* A few taxa cause myiasis in birds *(Muscina)* or are avian blood feeders *(Philornis).* Some Muscidae form a cocoon prior to pupariation (formation of puparia), which is uncommon among Diptera. Adult muscid flies may be predaceous on other insects, but most are generalized scavengers or feed on pollen.

OESTRIDAE These are commonly known as bot or warble flies (40 species). The larvae of all species are endoparasites. Species that attack livestock burrow into the host skin to feed on living tissue and either form their pupae under the skin, forming warbles *(Hypoderma),* or drop off the host and pupariate in soil *(Oestris).* Four species of the horse bot fly *(Gasterophilus)* infest the alimentary tract of horses (Fig. 38), donkeys, and mules. One genus *(Cuterebra)* (Fig. 64) infests lagomorphs and rodents and is among the biggest bot flies (about 2.5 cm). The human bot fly *(Dermatobia)* lays eggs on mosquitoes and other biting flies. When a larva hatches, it hangs onto the bloodsucker's leg until it lands on a human to take obtain a blood meal. The maggot then drops onto the host and burrows into the skin. Human bot flies are restricted to the Neotropical areas of the world and use a variety of mammalian hosts in addition to humans. North American vacationers and

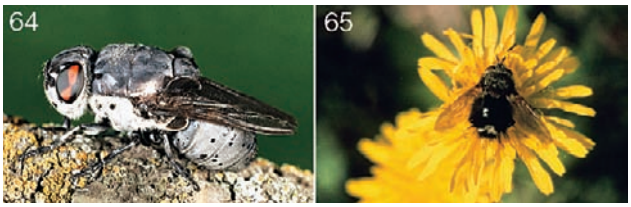

FIGURE 64–65 (64) Adult rodent bot fly (Oestridae: *Cuterebra jellisoni*). (Photograph by C. Baird and R. D. Akre.) (65) Adult of Tachinidae. (Photograph by E. A. Elsner.)

workers visiting the fly's home range frequently return home with a painful welt, under which lays a feeding maggot that respires through a small hole in the person's skin. The experience is painful, and most infected travelers have the larva removed surgically prior to pupariation or adult emergence.

NYCTERIBIIDAE Bat flies (250 species) are specialized ectoparasites of bats. The spider-like adults lack wings, which probably reduces the host's chances of removing the fly. Females bear mature living young (i.e., they are pupiparous), which is uncommon among Diptera. Larvae receive nutrients produced by glands within the female abdomen. The female bat flies deposit the larva on the walls of bat roosts, and females of some species will sit on top of the larva and briefly press it to the wall to ensure good adhesion. Pupariation occurs quickly without larval feeding. The adult emerges, and then seeks a host to continue the bloodsucking habit. Therefore, larval feeding does not occur outside of the adult female fly.

CALLIPHORIDAE The blow flies are thought to have been given their name from Homer's classic book, *The Iliad,* in which he wrote about the "blows of flies" infesting the wounds of injured and dead soldiers. Most of the 1000 species of this cosmopolitan family are attracted to rotting flesh and can sense the chemical scent of decay within minutes of death. In nature, the adults accelerate the decomposition of all types of vertebrate carcasses, and most blow flies (e.g., *Calliphora, Cochliomyia, Lucillia, Phaenicia*) are specialist scavengers. Because of the ability to find dead bodies rapidly, forensic scientists use the stage of larval development (i.e., age of a larva) found on corpses of people who died from suspicious causes as a way to determine the time between death and corpse discovery. It is well documented that by using this method, the time of death often can be estimated with a fair amount of accuracy. Other taxa of blow fly, however, exhibit other feeding habits, such as parasitism of land snails *(Helicobosca),* earthworms *(Pollenia),* and amphibians *(Bufolucilia).* The adults are also known as blue bottle and green bottle flies because some taxa have metallic brightly colored bodies (Figs. 35 and 37).

SARCOPHAGIDAE Flesh flies have been given a name that often contrasts their biology. Only a few of the 2500 species invade or consume carrion *(Sarcophaga)* or living tissue *(Wohlfahrtia).* Dung (e.g., *Ravinia*) is more commonly used. Many taxa are parasitic on other invertebrates (snails, earthworms, insects, and others), whereas some specialize in consuming the decaying bodies of insects found in the bottoms of pitcher plants of wetland habitats. Females are viviparous, young hatch within the female's abdomen, and she deposits them as first instars on the desired substrate. This may give flesh flies an advantage over potential competitors for food because mortality of eggs by predation or parasitism is avoided, and larvae can feed immediately rather than waiting to hatch for some days prior to feeding. The adults of most genera are easily recognizable by the gray thorax possessing longitudinal black stripes (Fig. 36).

RHINOPHORIDAE This is an unusual fly family in that nearly all of the 100 species for which biological details are known are specialist endoparasitoids of terrestrial isopods (also known as sow bugs, pill bugs, and potato bugs). This family represents the only dipterans that attack isopods. The larval life is tenuous, because larvae hatch from eggs laid in moist soil and must wait for a passing isopod. Perhaps the proleg-like apparatus present on the first instar is an adaptation to securing itself to a host. Both species of rhinophorids found in North America were probably introduced from Europe.

TACHINIDAE These flies (9500 species) (Fig. 65) are important parasites of a variety of other insects and are used in biological control programs against pestiferous Lepidoptera. Eggs are deposited on hosts or in areas where hosts are common. Some species retain their eggs so that they will hatch almost immediately after being laid; this strategy prevents the loss of the egg if the host happens to molt shortly after oviposition. Insertion of the egg through the epidermis of the host has evolved in a few species *(Phorocera).* An alternative strategy used by some taxa is to lay many eggs on partially consumed plants; when a potential host returns to continue feeding, the eggs are consumed along with the plant material. A few eggs survive maceration by the mandibles, and the larvae hatch within the host's foregut. Other genera broadcast their eggs, and the larvae burrow selectively into soil or rotting wood where they actively seek a host insect. Most often, tachinid flies attack only one species or a narrow spectrum of hosts; however, a small number *(Compsilura)* have been reared from some 200 different animal host species.

See Also the Following Articles

Drosophila melanogaster • *House Fly* • *Mosquitoes* • *Tsetse Fly*

Further Reading

Brown, B. V. (2001). Flies, gnats, and mosquitoes. *In* "Encylopedia of Biodiversity" (S. A. Levin, ed.), pp. 815–826. Academic Press, London.

Cole, F. R. (with collaboration of E. I. Schlinger) (1969). "The Flies of Western North America." University of California Press, Berkeley.

Courtney, G. W., Teskey, H. J., Merritt, R. W., and Foote, B. A. (1996). Aquatic Diptera, Part One, Larvae of aquatic Diptera. *In* "An Introduction to the Aquatic Insects of North America." (R. W. Merritt and K. W. Cummins, eds.), pp. 484–514. Kendall–Hunt, Dubuque, IA.

Dethier, V. G. (1963). "To Know a Fly." McGraw–Hill, Columbus, OH.

Feener, D. H., Jr., and Brown, B. V. (1997). Diptera as parasitoids. *Annu. Rev. Entomol.* **42,** 73–97.

Hennig, W. (1973). Diptera (Zweiflügler). *In* "Arthropoda," Hälfte 2, "Insecta," No. 2, "Spezielles, 31," (Helmecke *et al.,* eds.), Vol. IV of "Handbuch der Zoologie: Eine Naturgeschichte der Stämme des Tierreiches." Berlin/New York.

James, M. T. (1948). "The Flies That Cause Myiasis in Man." U.S. Department of Agriculture, Washington, DC. [Publication No. 631]

McAlpine, J. F., Peterson, B. V., Shewell, G. E., Teskey, H. J., Vockeroth, J. R., and Wood, D. M. (coordinators) (1981). "Manual of Nearctic Diptera," Vol. 1. Research Branch, Agricultural Canada. [Monograph 27]

McAlpine, J. F., Peterson, B. V., Shewell, G. E., Teskey, H. J., Vockeroth, J. R., and Wood, D. M. (coordinators) (1987). "Manual of Nearctic Diptera," Vol. 2. Research Branch, Agricultural Canada. [Monograph 28]
McAlpine, J. F., and Wood, D. M. (coordinators). "Manual of Nearctic Diptera," Vol. 3. Research Branch, Agricultural Canada. [Monograph 32]
Merritt, R. W., and Cummins, K. W. (1996). "An Introduction to the Aquatic Insects of North America." Kendall–Hunt, Dubuque, IA.
Merritt, R. W., Webb, D. W., and Schlinger, E. I. (1996). Aquatic Diptera, Part Two, Pupae and adults of aquatic Diptera. *In* "An Introduction to the Aquatic Insects of North America" (R. W. Merritt, and K. W. Cummins, eds.), pp. 515–548. Kendall–Hunt, Dubuque, IA.
Oldroyd, H. (1964). "The Natural History of Flies." Norton, New York.
Oosterbroek, P., and Courtney, G. W. (1995). Phylogeny of the nematocerous families of Diptera (Insecta). *Zool. J. Linn. Soc.* **115,** 267–311.
Papp, L., and Darvas, B. (eds.) (2000). "Manual of Palaearctic Diptera," Vol. 1. Science Herald, Budapest.
Yeates, D. K., and Wiegmann, B. M. (1999). Congruence and controversy: Toward a higher-level phylogeny of Diptera. *Annu. Rev. Entomol.* **44,** 397–428.

Diversity

see *Biodiversity*

Division of Labor in Insect Societies

Gene E. Robinson
University of Illinois, Urbana-Champaign

Division of labor is fundamental to the organization of the insect societies and is thought to be one of the principal factors in their ecological success. Different activities are performed simultaneously by specialized individuals in social insect colonies, which is more efficient than if tasks are performed sequentially by unspecialized individuals.

Division of labor is one of the defining characteristics of the most extreme form of sociality in the animal kingdom, "eusociality." Eusociality is defined by three traits: (1) cooperative care of young by members of the same colony, (2) an overlap of at least two generations of adults in the same colony, and (3) division of labor for reproduction, with (more or less) sterile individuals working on behalf of fecund colony members. It is now recognized by many biologists that eusociality extends to taxa beyond the ants, bees, wasps (Hymenoptera), and termites (Isoptera). This article focuses on the societies of the classic social insects, particularly the Hymenoptera, because they have the most elaborate and well-studied systems of division of labor.

DIVISION OF LABOR FOR REPRODUCTION

Females dominate the functioning of insect societies, even in termite societies, in which males play more diverse roles than in hymenopteran societies. There are two types of females in an insect society, queens and workers. Queens specialize in reproduction and may lay up to several thousand worker eggs per day. Workers are either completely or partially sterile, engage in little if any personal reproduction, and perform all tasks related to colony growth and maintenance. Worker sterility occurs because the ovaries do not develop or because critical steps in oogenesis do not occur. Worker sterility occurs either during preadult stages or during adulthood.

In many species of social insects, queens and workers are distinguished by striking morphological differences. A queen can have huge ovaries and a sperm storage organ that maintains viable sperm for years. The most striking morphological differences between queens and workers occur as a result of caste determination, which occurs during preadult stages. Caste determination has an endocrine basis. Research on the honey bee, *Apis mellifera,* and the bumble bee, *Bombus terrestris,* has shown that a high hemolymph titer of juvenile hormone (JH) during a critical period of larval development induces queen development. JH and presumably other hormones trigger a variety of processes that ultimately result in the production of either a worker or a queen. For example, caste-specific apoptosis (cell death) occurs in the ovaries of worker-destined honey bees and is associated with low titers of JH and ecdysteroid. Molecular analyses of endocrine-mediated caste determination have just begun. Some of the first findings involve caste-specific differences in the expression of genes that are associated with metabolism and protein synthesis, reflecting the fact that developing queens are metabolically more active than developing workers.

Little is known about how extrinsic factors act on endocrine-mediated developmental processes to influence caste determination. There is a strong circumstantial link between diet and JH in honey bee larvae, but how nutritional information acts to elevate JH levels is still largely unknown. In other species, extrinsic factors that influence caste determination include temperature and social factors such as behavioral interactions and pheromones released by adult colony members. These might affect the larvae directly or might influence the treatment accorded them by a colony's workers.

In societies in which queens and workers have strong morphological differences, the major mechanisms for queen domination of reproduction appear to be primer pheromones produced by queens. However, only one queen primer pheromone has been well characterized, that being the mandibular pheromone of the queen honey bee. Workers exposed to queen pheromones show little or no ovary development or egg-laying behavior. In other species of social insects, the physical differences between queens and workers can be very slight. Division of labor for reproduction in these

"primitively eusocial" species is achieved by a dominance hierarchy that is established and maintained by direct behavioral mechanisms, including pushing, biting, and physical prevention of egg laying. Behavioral domination is an ongoing process because some workers are physiologically capable of producing offspring and do, under some circumstances.

Queen behavior and pheromones affect adult worker neuroendocrine systems to reduce reproductive potential. JH has been implicated in the regulation of division of labor for reproduction in some, but not all, species studied to date, especially *B. terrestris;* the paper wasp, *Polistes gallicus;* and the fire ant, *Solenopsis invicta.* This is consistent with the function of JH as a hormone promoting reproductive development. JH does not appear to play this traditional role in adult *A. mellifera.* Ecdysteroids and biogenic amines also are suspected of being involved in the regulation of division of labor for reproduction among adult queens and workers, but a clear picture has not yet emerged.

DIVISION OF LABOR AMONG WORKERS

In most insect societies there also is a division of labor among the workers for tasks related to colony growth and maintenance. The evolution of a highly structured worker force is generally seen as an evolutionary consequence of the developmental divergence between queens and workers. Once workers were limited to serving largely as helpers, their characteristics could be shaped further by natural selection acting at the level of the colony to increase colony fitness. This perspective is consistent with the observation that the most intricate systems of division of labor among workers are found in species with the strongest division of labor for reproduction.

Age-related division of labor is the most common form of worker organization. Workers typically work inside the nest when they are young and shift to defending the nest and foraging outside when they are older. In the more elaborate forms of age-related division of labor, such as in honey bee colonies, workers perform a sequence of jobs in the nest before they mature into foragers. Physiological changes accompany this behavioral development to increase the efficiency with which particular tasks are performed. Among these are changes in metabolism, diet, and glandular secretions.

A less common but more extreme form of division of labor among workers is based on differences in worker morphology. This is seen in a minority of ant species and nearly all termites. Morphological differences among workers result from processes similar to worker–queen caste determination and morphologically distinct worker castes are recognized. For example, small ant workers (minors) typically labor in the nest, whereas bigger individuals (majors) defend and forage. Sometimes this form of division of labor also involves dramatic morphological adaptations in some worker castes, such as soldiers with huge and powerful mandibles and the ability to release a variety of potent defensive compounds.

A third form of division of labor among workers involves individual variability independent of age or morphology that results in an even finer grained social system. There are differences in the rate at which individual workers grow; some show precocious behavioral development, while others mature more slowly. There also are differences between individuals in the degree of task specialization. For example, foragers may specialize in the collection of a particular resource, such as some honey bees that collect only nectar or only pollen. It also has been found that some workers simply work harder than others.

The prevailing behavioral explanation for these three forms of division of labor among workers involves the application of the stimulus–response concept. Workers are thought to differ in behavior because of differences in exposure to, perception of, or response thresholds to stimuli that evoke the performance of a specific task. These differences can result from differences in worker genotype, age, experience, or morphological caste. There is some behavioral evidence for differences among workers in stimulus perception and response thresholds; challenges for the future are to more precisely define the nature of the stimuli and extend these analyses to the neural levels.

Some endocrine and neural mechanisms regulating age-related division of labor have been discovered, primarily in honey bees. Changes in hemolymph titers of JH act to influence the rate and timing of behavioral development, but JH is not required for a worker to mature into a forager. Evidence for a similar role for JH has been found in the advanced eusocial tropical wasp *Polybia occidentalis.* JH also affects the activity of exocrine glands that produce brood food and alarm pheromones in honey bees, apparently acting to ensure that physiological changes are coordinated with behavioral development. As JH receptors have not yet been identified in any insect, it is not known whether JH exerts its effects on division of labor directly in the brain, on other target tissues, or at a variety of sites. Octopamine acts as a neuromodulator in honey bees. Higher levels of octopamine, particularly in the antennal lobes of the brain, increase the likelihood of foraging. Changes in brain structure also occur as a worker bee matures into a forager, particularly in the antennal lobes and mushroom bodies, but the functional significance of these changes is unknown. As with caste determination, molecular analyses of behavioral development have only recently been initiated. Differences in the expression of several genes have been detected in the brains of younger and older honey bee workers. The orchestration of the neural and behavioral plasticity that underlies age-related division of labor is undoubtedly based on changes in the expression of many genes in the brain and other tissues as well.

Mechanisms underlying morphologically based systems of worker division of labor also have been studied. Morphological differences among adult workers have their origin in

pathways of development that diverge during the larval stage. Information on worker caste differentiation, drawn largely from studies of *Pheidole* ants, suggests mechanisms similar to those involved in queen–worker caste determination. Both larval nutrition and JH have been implicated in the differentiation of *Pheidole* minors and soldiers.

Genetics is one factor influencing the third form of division of labor among workers, individual variability among workers that is independent of age or morphology. Genotypic variation within colonies arises as a consequence of multiple mating by queens or multiple queens in a colony. This genotypic variation is strongly associated with behavioral differences between individuals within a colony. Genotypic variation in honey bee colonies is known to influence how specialized a worker becomes on a particular task or the age at which it shifts from nest work to foraging. For example, quantitative trait loci have been found that are associated with variation in the tendency of honey bees to collect either nectar or pollen. These findings can lead to the identification of differences in specific genes that contribute to individual differences in behavioral specialization. Genotypic effects on division of labor also have been documented in several ant and wasp species.

PLASTICITY IN COLONY DIVISION OF LABOR

Colony division of labor, though highly structured, also shows great plasticity. Colonies respond to changing needs by adjusting the ratios of individual workers engaged in different tasks. This is a consequence of the flexibility of the individual workers. For example, there is plasticity in age-related division of labor, with workers able to respond to changes in colony age demography with accelerated, retarded, or reversed behavioral development. Workers also can shift to emphasizing a different task that is part of their age-specific repertoire or they can simply work harder. Morphologically specialized workers can be induced to shift their behavior; majors, normally specialized in foraging or defense, can care for the brood in the absence of minor workers. This plasticity in division of labor contributes to the reproductive success of a colony by enabling it to continue to grow, develop, and ultimately produce a new generation of reproductive males and females during changing colony conditions.

Plasticity in division of labor in advanced eusocial species is achieved by a variety of mechanisms of behavioral integration. These mechanisms enable workers to respond to fragmentary information with actions that are appropriate to the state of the whole colony. This makes sense because it is unlikely that any individual workers have the cognitive abilities to monitor the state of their whole colony and then perform the tasks that are needed most or direct others to do so.

Mechanisms of worker behavioral integration often involve social interactions. For example, in many species, including *Polybia* wasps and honey bees, nest workers routinely relieve the foragers of their newly acquired loads, whether nest material or food. Foragers that are unloaded immediately upon their return to the nest are likely to continue foraging for the same resource, apparently because the quick unloading signals to them that they have brought something of high value back to the colony. In contrast, foragers that experience a significant time delay before being unloaded respond by changing their behavior, perhaps shifting to the collection of another resource. The nutritional status of a fire ant colony strongly influences the behavior of its foragers, with the relevant information transferred during social feeding. In colonies of the desert-dwelling red harvester ant, *Pogonomyrmex barbatus,* workers obtain information on the needs of the colony by changes in their encounter patterns with members of various task specialist groups. For example, red harvester ant foragers are more likely to leave the nest to forage when they encounter greater numbers of successful returning foragers.

Social inhibition is a potent mechanism of integration in insect colonies. In colonies of honey bees, social inhibition acts to keep the division of labor synchronized with changes in colony age demography. Older workers inhibit the rate of maturation of younger workers. Some young workers in a colony deficient in older workers, for example, exposed to lower levels of social inhibition, respond by becoming precocious foragers. The specific honey bee worker factor that causes this inhibition has not yet been identified, but other sources of social inhibition have, emanating from the queen and the brood. The regulation of the size of the soldier force in *Pheidole* colonies also is based on a process of social inhibition. In this case, the presence of adult soldiers inhibits the production of new soldiers. Involvement of a pheromone is suspected, but no specific soldier inhibition pheromone has been identified yet.

The integration of activity in primitively social insect societies appears to be more centralized than in advanced eusocial societies. Primitively eusocial colonies often contain only a few dozen individuals, making centralized control more feasible. Queens act as central pacemakers and modulate worker activity via behavioral interactions in sweat bees and polistine wasps. Queens do not appear to be able to get workers to shift to different tasks, but they do cause them to work harder at the tasks they are already doing.

We are far from understanding how the behavior of individual workers is integrated into a well-functioning colony. Studies of behavioral integration are aided by various kinds of theoretical models. In some models, an insect colony is likened to a developing organism, i.e., the "superorganism" metaphor. In other models, an insect colony is analyzed with perspectives from neural network theory, with individual workers serving as analogs of individual neurons. Still other models view an insect colony as a self-organizing entity and use complex systems theory to develop ideas on colony function.

See Also the Following Articles

Ants • Caste • Colonies • Hymenoptera • Isoptera • Recruitment Communication • Sociality

Further Reading

Beshers, S. N., Robinson, G. E., and Mittenthal, J. (1999). The response threshold concept and division of labor. *In* "Information Processing in Social Insects" (C. Detrain *et al.,* eds.), pp. 115–141. Birkhauser, Basel.

Bloch, G., Wheeler, D. L., and Robinson, G. E. (2002). Endocrine influences on the organization of insect societies. *In* "Hormones, Brain and Behavior" (D. Pfaff *et al.,* eds.), pp. 195–237. Academic Press, San Diego.

Detrain, C., Deneubourg, J. L., and Pasteels, J. M. (eds.) "Information Processing in Social Insects." Birkhauser, Basel.

Page, R. E., and Robinson, G. E. (1991). The genetics of division of labour in honey bee colonies. *Adv. Insect Physiol.* **23,** 117–171.

Robinson, G. E. (1992). The regulation of division of labor in insect societies. *Annu. Rev. Entomol.* **37,** 637–665.

Robinson, G. E. (1998). From society to genes with the honey bee. *Am. Sci.* **86,** 456–462.

Wilson, E. O. (1971). "The Insect Societies." Harvard University Press, Cambridge, MA.

Dobsonfly

see *Megaloptera*

Dog Heartworm

Edward G. Platzer
University of California, Riverside

Dog heartworm, *Dirofilaria immitis,* is an important filarial nematode infection of dogs and canids primarily, but can occur in other mammals and occasionally humans. It usually occupies the pulmonary arteries and the right ventricle of the heart. Dog heartworm is transmitted by mosquitoes.

BIOLOGY

D. immitis is a filarial nematode in the superfamily Filarioidea (order Spirurida, class Secernentea). The males are 12 to 20 cm in length and 0.7 to 0.9 mm in diameter, with a spirally coiled posterior end. The females are 25 to 31 cm in length and 1.0 to 1.3 mm in diameter. *D. immitis* was first found in the United States, but it occurs globally, with a tendency for increased prevalence in humid warm regions conducive to abundant mosquito populations. In the United States, the prevalence can be as great as 45% in dogs within 150 miles of the Gulf of Mexico coast, Atlantic coast north to New Jersey, and Mississippi River and its tributaries. The prevalence elsewhere in the United States is generally less than 5%. Although dogs are the primary host, *D. immitis* has been found in coyotes, wolves, dingoes, foxes, sea lions, harbor seals, wolverines, ferrets, and cats. The number of nematodes per dog is variable, ranging from single nematodes to as many as 250. Cats are less tolerant, with maximum parasite loads of 1 to 3 nematodes. *D. immitis* can persist for 5 to 7 years in dogs and 2 to 3 years in cats.

Mature female nematodes reproduce ovoviparously and the microfilaria are released from the uterus of the worm via the vaginal opening into the host blood. Microfilaria are 220 to 330 μm in length and 5 to 7 μm in width. Under laboratory conditions, over 60 species of mosquitoes are competent hosts for *D. immitis.* When mosquitoes take a blood meal, microfilaria are ingested and reside in the mosquito midgut briefly and then migrate to the Malpighian tubules, where they enter the cells and shorten. After a short developmental period, the juvenile nematodes leave the cells day 6 to 7 postinfection and enter the lumen of the tubules where the first molt (day 10) takes place, with the formation of the second-state juveniles (J2). After further growth and differentiation, the J2 molts (days 10–14 postinfection) to the infective stage (J3), which reaches a length of 1.3 mm. The J3 migrate through the hemocoel to the proboscis sheath of the mosquito. Development of the juvenile stages ceases if ambient temperatures decline below 15°C, which constitutes a constraint on the distribution of *D. immitis.* When the infected mosquito takes a blood meal, the J3 escape from the proboscis sheath, dropping onto the host in a droplet of hemolymph; they enter the host through the wound made by the piercing mouthparts of the mosquito. The J3 enter the subcutaneous tissue where they undergo the third molt to the fourth-stage juveniles, which reside in subcutaneous tissues or muscle of the abdomen or thorax for about 60 days, at which time the last molt to the adult stage occurs. The nematodes are now 12 to 15 mm long and enter the pulmonary arteries and attain lengths of 3.2 to 11 cm by 85 to 120 days postinfection. Fertilized females can be found 120 days postinfection and microfilaria enter the blood 6 to 9 months postinfection. Microfilaria can survive in the blood for $2^1/_2$ years. Host treatment with tetracycline leads to the loss of the endosymbionts and a concomitant reduction in survival and reproduction of *D. immitis.*

HOST PATHOLOGY

Usually no signs of infection are present in dogs until 8 to 9 months postinfection. Symptoms are generally related to the intensity of infection. The presence of 25 worms may be tolerated by the dog with no signs of disease. Symptoms such as reduced exercise tolerance and coughing increase significantly when the nematode burden doubles and serious symptoms, such as dsypnea (difficulty in breathing), hepatomegaly (enlargement of the liver), syncope (temporary loss of consciousness), and ascites (fluid accumulation in the abdomen) appear. At this level, death may occur. Pathogenesis is related primarily to inflammation of pulmonary arteries and lungs induced by the adult stage of *D. immitis.* This chronic inflammatory process in combination with the physical obstruction by the nematodes of blood flow leads to pul-

monary hypertension and heart failure. Diagnosis depends on an accurate history, recognition of symptoms, and diagnostic procedures such as microfilarial detection, serology, clinical laboratory tests, radiology, ultrasonography, and angiography. Microfilaria are detected through concentration of the microfilaria from a small quantity of blood by the Knott's test or filtration. In some dog heartworm infections (occult heartworm infection), microfilaria are not detectable and diagnosis relies on serology and other diagnostic tools.

MANAGEMENT

Treatment is effective. The goal is elimination of the primary disease agent, the adults. Two arsenical drugs are currently approved for use in dogs, melarsomine hydrochloride (Immiticide; Merial) and thiacetarsamide sodium (Caparsolate; Merial). Treatment requires concomitant restriction of exercise and use of anti-inflammatory support to reduce the possibility of pulmonary thromboembolism from the dead heartworms. Microfilaria are eliminated by secondary treatment with ivermectin or milbemycin. Prevention of heartworm infection is safer and more economical than treatment and is accomplished readily by routine administration of diethylcarbamazine (daily in the diet) or one of the macrolide anthelminthics (monthly treatment; ivermectin, milbemycin, moxidectin, selamectin) during the transmission season.

See Also the Following Articles

Blood Sucking • Mosquitoes • River Blindness • Veterinary Entomology • Wolbachia • Zoonoses, Arthropod-Borne

Further Reading

American Heartworm Association (2001). http://www.heartwormsociety.org/.

Anderson, R. C. (2000). "Nematode Parasites of Vertebrates: Their Development and Transmission," 2nd ed. CAB Int., Wallingford, UK.

Bowman, D. D. (1999). "Georgis' Parasitology for Veterinarians," 7th ed. Saunders, Philadelphia.

Taylor, M. J., and Hoerauf, A. (1999). *Wolbachia* bacteria of filarial nematodes. *Parasitol. Today* **14,** 437–442.

Dormancy

Richard E. Lee, Jr.
Miami University, Oxford, Ohio

Dormancy is an inactive state associated with metabolic depression and arrested development that promotes the survival of insects during periods of harsh environmental conditions, including high or low temperatures or moisture conditions and reduced food quality or availability. Diapause may occur in any life stage (egg, larva, pupa, or adult); however, for a given species it generally occurs in only one stage.

Dormancy may be manifested in a variety of forms that differ widely in their intensity and duration. Consequently, it has proved difficult to classify these forms into discrete categories, and general agreement on terminology is lacking. Nonetheless, dormant states range from a moderate depression of short duration (quiescence) to a profound and extended period of metabolic suppression and developmental arrest (diapause).

Quiescence commonly refers to short periods of dormancy that are directly induced by adverse environmental conditions, principally low or high temperature. It also has the advantage of being quickly reversible upon the return of favorable conditions; this rapid response may be especially important in extreme environments, such as alpine regions or deserts, where access to food and favorable conditions are intermittent and unpredictable.

In contrast, diapause is not directly induced, but is triggered by genetically programmed responses to environmental cues that occur in advance of adverse conditions. Anticipatory induction allows time for substantial physiological changes prior to the arrival of adverse conditions. These changes may include accumulation of lipid and glycogen reserves, deposition of cuticular lipids that enhance desiccation resistance, suppression of gametogenesis, decreased metabolic rate, increased tolerance of anoxia, and low temperature.

Although moisture conditions, temperature, and host-plant quality may serve as cues for the induction of diapause, photoperiod is the factor that has been identified most commonly in this regard and is the one that has been the subject of the most investigation. Furthermore, these environmental conditions may interact to promote or inhibit the induction of diapause; an unusually cool autumn may induce diapause sooner than would be expected based solely on photoperiod. The cue for diapause induction need not be received by the life stage that enters diapause. For example, in many species adults that experience short daylengths produce diapause eggs, while those exposed to longer photoperiods do not.

See Also the Following Articles

Aestivation • Cold/Heat Protection • Diapause

Further Reading

Danks, H. V. (1987). "Insect Dormancy: An Ecological Perspective." Biological Survey of Canada (Terrestrial Arthropods), Ottawa.

Denlinger, D. L. (1986). Dormancy in tropical insects. *Annu. Rev. Entomol.* **31,** 239–264.

Leather, S. E., Walters, K. F. A., and Bale, J. S. (1993). "The Ecology of Insect Overwintering." Cambridge University Press, New York.

Lee, R. E., and Denlinger, D. L. (eds.) (1991). "Insects at Low Temperature." Chapman & Hall, New York. [See especially Chaps. 8, 9, 10, and 14]

Tauber, M. J., Tauber, C. A., and Masaki, S. (1986). "Seasonal Adaptations of Insects." Oxford University Press, New York.

Dragonfly

see *Odonata*

Drosophila melanogaster

Patrick M. O'Grady
American Museum of Natural History

When biologists refer to *"Drosophila"* they usually mean *Drosophila melanogaster.* This small, inconspicuous species has become one of the premiere model systems in modern biology. Research on *D. melanogaster* over the past century has led to better understanding of virtually every discipline of biology, especially genetics and developmental and evolutionary biology. This work has applications not merely to the biology of flies and other insects, but also into the causes of a variety of human diseases. The most powerful aspect of *Drosophila* as a model system is the ease with which its genome can be manipulated through a variety of genetic techniques, including germline transformation with transposons. The genomics revolution promises to expand the utility of *D. melanogaster* and make this species not only a pivotal tool in understanding the evolution and working of the genome but also an important model for bioinformatics and genome annotation.

HISTORY OF *DROSOPHILA* RESEARCH

D. melanogaster was first described by Meigen in 1830. Subsequent taxonomists described this species under at least five different names from 1830 to 1941. The profusion of names was likely due to the quick spread of this species throughout the world as a result of the fruit trade.

Drosophila research began in the early 1900s when a number of scientists, most notably T. H. Morgan, began to use *D. melanogaster* as a model organism for studies of genetics. W. E. Castle was the first to bring this species into the lab and develop many of the culture techniques still used today. It was Morgan's group at Columbia University, however, that fully took advantage of this species as a research model. Morgan, up to that time, had been experimenting with marine invertebrates in an effort to understand a number of developmental processes. He was looking for a small, rapidly developing species that produced large numbers of progeny and was both easy and inexpensive to maintain and manipulate in the laboratory. Early in these studies, it became clear that *D. melanogaster* was just such a model system. In 1912 Morgan's group had isolated roughly two dozen mutants. Morgan and his colleagues began to use these mutants to provide experimental evidence for the chromosome theory of inheritance, and they devised methods for gene mapping that are still used today.

Drosophila was an important model organism throughout the 20th century. Ed Lewis began working on homeotic mutants in the 1950s. His work focused on the bithorax gene complex. Most Diptera have only a single set of wings on the mesothoracic segment, but these mutant flies had two pairs of wings, one each on the meso- and metathoracic segments (Fig. 1). This set of genes has since proved to be the major control switch for body axis development and is conserved in many organisms, including humans. The Nüsslein–Volhard and Wieschaus screens of the early 1980s further advanced the use of *D. melanogaster* as a model system to study the development of more complex organisms. The future Nobel laureates elegantly showed the genetic control of development, mapping many of the genes involved in forming the major body axes in nearly all metazoans.

FIGURE 1 The ultrabithorax mutant. [From Lawrence P. A. (1992). "The Making of a Fly: The Genetics of Animal Design." Blackwell Scientific Publications, Oxford, U.K., with permission of the publisher.]

D. melanogaster continues to be an important model system in biological research, and the *Drosophila* Genome Project has completed the entire genome sequence of *Drosophila melanogaster*. This work, described by Adams and colleagues in 2000, has provided researchers with an immense amount of data that can be used to understand the mechanisms of development and the evolution of the genome. As of late 2002, a reference search of Flybase (http://flybase.bio.indiana.edu/) recovers about 20,000 papers with the query terms *"Drosophila melanogaster."* Furthermore, GenBank (http://www.ncbi.nlm.nih.gov/) currently contains over 322,000 nucleotide entries for this species. Several stock centers around the world are dedicated to maintaining live cultures of *D. melanogaster* and its relatives for research. For example, the Bloomington Stock Center (http://flystocks.bio.indiana.edu/) currently has about 8700 different lines, mostly mutants of *D. melanogaster,* and the Tucson Stock Center (http://stockcenter.arl.arizona.edu/) maintains about 1300 cultures from nearly 300 species in the family Drosophilidae.

ECOLOGY AND LIFE CYCLE

Drosophila melanogaster originated in tropical west Africa and has spread around the world, primarily through its commensal associations with humans. This species is a generalist and breeds in a variety of rotting fruits in its

natural environment. It was first recorded on the east coast of North America in the 1870s following the end of the American Civil War and the expansion of the fruit trade.

Like all other members of the family Drosophilidae, *D. melanogaster* is holometabolous and undergoes a complete metamorphosis. Development times vary, depending on temperature. Typical *Drosophila* laboratories maintain flies between 18 and 25°C. Stocks or infrequently used strains are usually kept at lower temperatures to slow development and reduce the amount of stock changing required. Complete development takes about 3 weeks at 18°C. At 25°C, embryonic development is completed roughly 1 day after the egg is laid. The fly then goes through three larval stages prior to pupation. Larvae are motile and work their way through the food media feeding on yeast and bacteria. After 4 days, the larvae enter a stationary pupal stage. Pupation takes approximately 4 days, after which time adults emerge from the pupal case. After they eclose, females require about 2 to 3 days to develop mature eggs. Therefore, at 25°C about 10 or 11 days is required to complete a cycle from egg to egg. At higher temperatures (29–30°C), pupal lethality and female sterility begin to have an effect on culture viability.

After the adult ecloses, it takes between 6 and 12 h for both males and females to begin mating. Genetic crosses require known paternity. Females are collected prior to reaching sexual maturity and isolated from males, so controlled crosses can be made. Mean adult life span is 40 to 50 days, although some individuals may live up to 80 days. A single female can lay as many as 75 eggs in a day, for a total of perhaps 500 eggs in a 10-day period.

MORPHOLOGY AND PHYLOGENY

The family Drosophilidae is divided into a number of genera, subgenera, species groups, and species subgroups; this system gives each species a "taxonomic address" that loosely defines relationships within the family. For example, *D. melanogaster* is placed in the genus *Drosophila,* subgenus *Sophophora,* and *melanogaster* species group and subgroup (Fig. 2). *D. melanogaster* is a typical drosophilid and possesses a number of the characteristics, such as red eyes and plumose arista, that delineate this family. Along with the other taxa in the *melanogaster* and *obscura* species groups, *D. melanogaster* bears a single sex comb on its first tarsal segment. These are 7 to 12 thickened setae (hairs), which are closely set in a row, or comb. The number and position of the setae diagnose *D. melanogaster* from all but the most closely related species.

Within the Afrotropical *melanogaster* species subgroup, *D. melanogaster* is most closely related to the triad of species containing *D. simulans, D. sechellia,* and *D. mauritiana,* the common ancestor of which is thought to have diverged between 2 and 3 mya. *D. simulans,* a closely related species that is also cosmopolitan, can be differentiated only by examining the characters of the male genitalia, namely the number of prensisetae and the shape of the epandrial lobes.

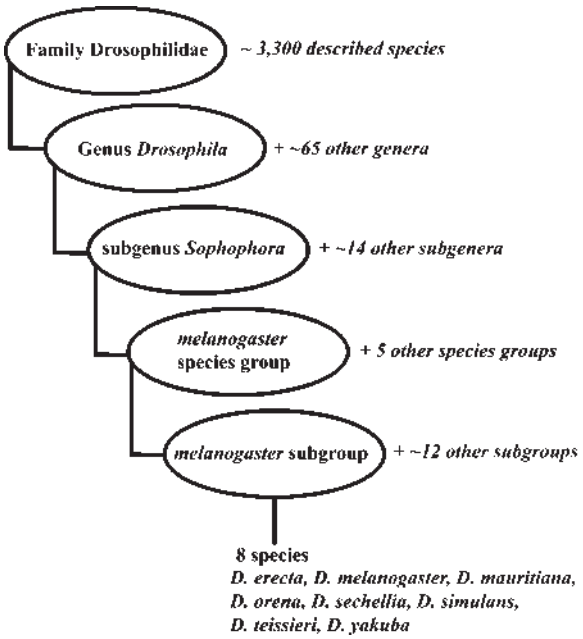

FIGURE 2 Placement of *D. melanogaster* within the family Drosophilidae. [Modified after Powell, J. R. (1997). "Progress and Prospects in Evolutionary Biology. The *Drosophila* Model." Oxford University Press, New York.]

DROSOPHILA MELANOGASTER AS A GENETIC MODEL

Over the past 100 years, geneticists have built a large "toolbox" of specialized methods that allow them to manipulate the genome of *D. melanogaster* with more deftness than is possible with any other organism. These methods have largely taken advantage of some of *Drosophila*'s inherent characteristics, such as the lack of recombination in males. Some widely used techniques include polytene chromosome visualization and *in situ* hybridization, using balancers and other cytological aberrations for genetic crosses, and germline transformation using *P* elements and other transposons to examine gene expression and to tag genes for cloning.

The chromosomes found in the larval salivary glands are highly duplicated, allowing a characteristic banding pattern to be visualized with a compound light microscope (Fig. 3). Polytene chromosomes, which allow researchers to observe and study large-scale genetic rearrangements such as inversions, duplications, translocations, and deletions, have been used by geneticists to answer a variety of questions. Early work focused on understanding chromosome mechanics and using deletions to map the location of specific genes. Molecular geneticists have used the polytene chromosome in conjunction with *in situ* hybridization to more specifically localize the chromosomal site of specific cloned genes or gene fragments. For example, small fragments of DNA can be amplified by using the polymerase chain reaction (PCR), incorporating radioactive or bioluminescent probes as labels and with hybridization to the polytene chromosomes. In addition, evolutionary and population geneticists have used inversion patterns to reconstruct the history of species and populations.

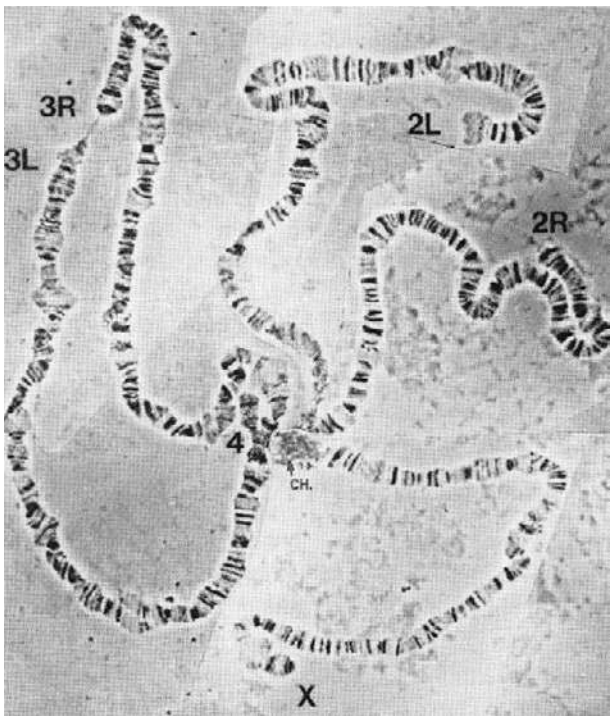

FIGURE 3 Polytene chromosome of *D. melanogaster:* X, X chromosome; 2R, right arm of second chromosome; 2L, left arm of second chromosome; 3R, right arm of third chromosome; 3L, left arm of third chromosome; 4, fourth chromosome; CH, chromocenter. [From Krimbas C., and Powell, J. R. (eds.) (1992). "*Drosophila* Inversion Polymorphism," Fig. 2B, p. 344. CRC Press, Boca Raton, FL, with permission.]

Balancers are multiply inverted chromosomes that repress recombination and are useful for making controlled genetic crosses as well as keeping homozygous lethal mutant genes in culture. In addition to being marked with a visible phenotype, such as curly wings, balancer chromosomes are often homozygous lethal, making crosses and the establishment of multiple mutant stocks much simpler.

Transposable elements (TEs) are native components of the genomes of nearly all organisms. TEs typically encode a protein, called transposase (some also move via a method that is mediated by reverse transcriptase), which can catalyze the movement of the element throughout the genome. Transposons have been used to mutagenize and clone genes, as well as to study spatial and temporal patterns of gene expression. The *P* transposable element was isolated after several researchers noticed an aberrant syndrome of hybrid sterility when certain geographic strains were crossed. This sterility was caused by the introduction of *P* elements into a genetic background lacking these transposons. This transposon has become the most versatile and widely used tool in modern *Drosophila* genetics.

Since their discovery, *P* elements have been heavily modified and are now used extensively to manipulate *Drosophila* germline DNA. The transposase coding regions have been removed and replaced with a wild-type marker gene, resulting in an inactive transposon with a dominant marker. It is possible to introduce such a *P* element into the germline of mutant embryos by injecting the embryos with cloned *P* element DNA and a buffer containing the active transposase. The offspring carrying the *P* element construct will have a wild-type phenotype because the marker will rescue the mutant phenotype of the recipient strain. Such transformed lines can be used in mutagenesis screens by crossing the inactive *P* element construct line to a stock engineered to contain active transposase. Because *P* elements insert at random into the genome, they are very effective mutagenic agents and can insert into a gene, thereby disrupting its function. Once a phenotype has been observed, the transposase can be "crossed out," leaving a stable *P* element insertion into a gene of interest. The mutagenized gene can be easily cloned by means of a variety of techniques (e.g., inverse PCR) because the sequence of the *P* element is known and a "transposon tag" is present in the gene of interest. Other powerful techniques that exploit transposons are enhancer trapping and the flipase recombination system.

THE GENUS *DROSOPHILA* AS A MODEL SYSTEM

In addition to referring to the single species *D. melanogaster, "Drosophila"* can also refer to the entire genus *Drosophila,* a spectacular radiation of roughly 1500 described species. This genus can be found throughout the world in every conceivable habitat, from tropical rain forests to subarctic regions. Generally, these species are saprophytic, feeding and ovipositing in rotting plant and, sometimes, animal material. Members of this genus have been used as a model system for understanding evolutionary biology. A number of *Drosophila* groups, such as the *obscura, repleta,* and *virilis* species groups, have become prominent model systems in evolutionary biology. Such studies include chromosome and molecular evolution, the mechanisms of species formation, phylogeny, ecology, and behavior.

See Also the Following Articles

Chromosomes • Diptera • Genetic Engineering • Research Tools, Insects as

Further Reading

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000). The genome sequence of *Drosophila melanogaster. Science* **287,** 2185–2195.

Ashburner, M. (1989). "*Drosophila:* A Laboratory Handbook." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Ashburner, M., Carson, H. L., and Thompson, J. N., Jr. (eds.). (1981—1986). "The Genetics and Biology of *Drosophila.*" Academic Press, New York. (See especially Wheeler, M. R., pp. 1–84, and Lemeunier *et al.,* pp. 147–256.)

Barker, J. F. S., Starmer, W. T., and MacIntyre, R. J. (eds.). (1990). "Ecological and Evolutionary Genetics of *Drosophila.*" Plenum Press, New York.

Kohler, R. E. (1994). "Lords of the Fly: *Drosophila* Genetics and the Experimental Life." University of Chicago Press, Chicago.

Krimbas, C., and Powell, J. R. (eds.). (1992). "*Drosophila* Inversion Polymorphism." CRC Press, Boca Raton, FL.

Lawrence P. A. (1992). "The Making of a Fly: The Genetics of Animal Design." Blackwell Scientific, Oxford, U.K.

Patterson, J. T., and Stone, W. S. (1952). "Evolution in the Genus *Drosophila.*" Macmillan, New York.
Powell, J. R. (1997). "Progress and Prospects in Evolutionary Biology: The *Drosophila* Model." Oxford University Press, New York.
Sullivan, W., Ashburner, M, and Hawley, R. S. (eds.). (2001). "*Drosophila* Protocols." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. (See especially, Pardue, M.-L., pp. 119–129.)
Throckmorton, L. H. (1975). The phylogeny, ecology, and geography of *Drosophila. In* "Handbook of Genetics," Vol. 3, "Invertebrates of Genetic Interest." (R. C. King, ed.), pp. 421–469. Plenum Press, New York.

Dung Beetles

James Ridsdill-Smith
Commonwealth Scientific and Industrial Research Organisation, Australia

Dung beetles are specialized to feed and breed on an ephemeral and discrete food resource, namely, the piles of dung produced by herbivorous warm-blooded animals. Adults are strong flyers and can search for some distance to find fresh dung. Most species make tunnels in the soil and remove dung from the pat, which is packed into the tunnel to form a brood mass in which a single egg is laid. Ecosystems contain many coexisting dung beetle species, particularly in tropical grasslands. Intraspecific and interspecific competition for dung is high, and beetles show diverse behavior to reduce its effects. Fecundity of dung beetles is extremely low, the eggs are relatively large, and adult investment in nesting behavior is high, as is illustrated in male/female adult reproductive behavior and in brood care by female beetles.

Dung beetles have a place in history, in which the ball-rolling species, *Scarabeus sacer* (Fig. 1), was sacred to the early Egyptians (Fig. 2). The ability of large ball-rolling beetles to create a perfect sphere, which is then rolled along the soil surface, was taken to be representing Khepri, a great scarab beetle, rolling the globe of the rising sun. The scarabeus, symbol of the sun, is often depicted hovering with outstretched wings.The new beetle emerges from the inactive pupa, representing rebirth or reincarnation. Scarab amulets often were placed over the heart of the dead to simulate rebirth or worn widely by the living to bring good luck.

BIOLOGY

Dung as a Resource

When scarabaeine dung beetles are abundant, numbers can be observed swarming upwind in the odor plume from fresh dung. Volatile compounds produced from the fresh dung attract adult dung beetles, and most arrive within the first few hours after the dung is deposited. Up to 16,000 beetles have been recorded at a single elephant dropping, to which 4000 beetles were attracted in 15 min. Beetles leave when most of the dung is buried or when feeding activity has removed most of the moisture from the dung (referred to as shredding). Adult beetles feed on the liquid "soup" in the dung. The incisor lobe of the adults is flattened and fringed for handling soft food, and the particulate components of the food are filtered out before being ingested. The galae and laciniae of the adults have special brushes for collecting food. Dung produced by herbivorous animals is highly variable in size and consistency and may range from mounds weighing over 1 kg from an elephant to pellets of about 1 g from a rabbit. The water content of dung is high at deposition (90% water), but dung dries out quickly, the rate depending on both temperature and the size of the dung pat. Typically, dung is used as a resource by dung beetles for 1 to 4 weeks, although small sheep pellets may dry out in 3 h in summer and are relatively little used by dung beetles after this time.

FIGURE 1 Adult *S. sacer.* Unworn tibiae indicate a newly emerged beetle. (Photograph courtesy of CSIRO Entomology.)

FIGURE 2 Egyptian sacred scarabs. Top center is a winged scarab pectoral in blue faience with holes for attachment. Bottom left is a heart scarab in blue faience. Bottom right is an inscribed heart scarab in felspar. (Photograph courtesy of Ashmolean Museum, University of Oxford.)

Dung Burial for Brood Masses

The behavior for which the dung beetle is best known is the removal of dung from the pat and burial in the ground as provisioning for their offspring. Three groups are distinguished based on their behavior in creating a brood mass, the compacted dung in which a single egg is laid. In teleocoprids, a sphere is made from pieces of dung at the dung pat or from pellets of dung (Fig. 3A). Beetles roll the ball away from the dung pat, usually with the hind legs. It is buried in the soil, and a single egg is laid in a small cavity. In paracoprids, the beetle digs a tunnel in the soil under a dung pat, cuts off pieces of dung using its front legs, head, and body, and carries them down the tunnel where they are packed into the end to form a compacted brood mass. As each brood mass is completed, a single egg is laid. Soil is then placed over the brood mass and another brood mass is made. Branching tunnels may be made containing many brood masses with eggs (Fig. 3B). The size and shape of the brood mass, and the depth of the brood mass in the soil, are characteristic for each species. These will be affected by soil moisture and soil hardness. Endocoprid species construct brood balls in cavities within the dung pat.

Life Cycle

When a female beetle is ready to lay her eggs, she constructs a chamber in the top of the brood mass and lays an egg on a small pedestal, which prevents it from coming into contact with the surrounding dung (Fig. 3C). The larva hatches after a week or two and feeds on the dung of its brood mass (Fig. 3D). Larvae have biting mouthparts, unlike the adults, and can use the fiber content of the dung. They typically complete three instars over about 12 weeks and then undergo a pupal stage (Fig. 3E) 1 to 4 weeks before turning into an adult (Fig. 3F). The adults emerge from the brood shells, dig their way to the soil surface, and then fly off and find fresh dung on which to feed. Depending on the biology of the individual species, there may be periods of diapause or quiescence by mature larvae, pupae, or adults during development. Such adaptations are usually related to enhancing survival over, for example, a dry summer or a cold winter and can delay the completion of the life cycle by several months. Dung beetle species are univoltine, completing one generation a year, or multivoltine, completing several.

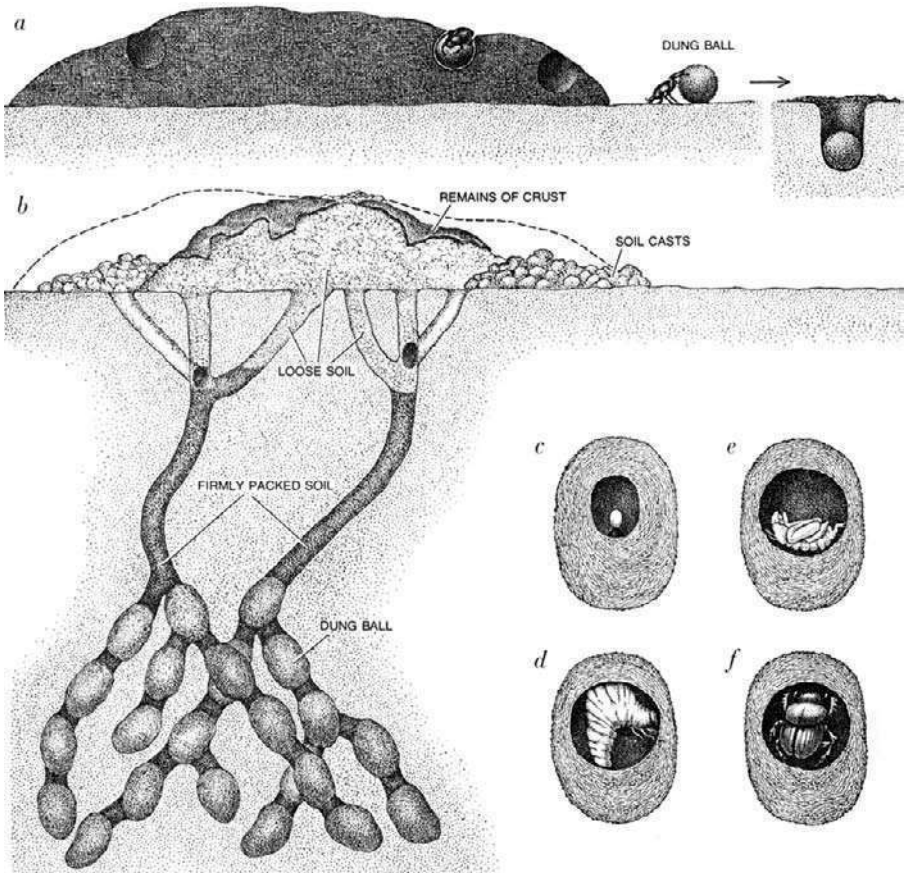

FIGURE 3 Dung beetle reproduction in cattle dung. (a) Dung pat with a teleocoprid species removing a ball of dung and burying it. (b) Dung pat with a paracoprid species producing brood masses in tunnels beneath the pat. Brood mass containing: (c) egg, (d) larva, (e) pupa, and (f) young adult. (Illustration by Tom Prentis from Waterhouse, 1974; reproduced with permission.)

Dung Quality

Beetle egg laying is very sensitive to seasonal changes in dung quality. This quality is influenced by several factors, including rainfall, which affects the growth of plants on which grazing animals are feeding and hence the quality of dung they produce, and the plant species grazed upon. Under favorable laboratory conditions, the rate of egg production of *Onthophagus binodis* on dung collected from cattle grazing on dry summer annual pasture in winter rainfall regions of Australia is 7% of that on dung from cattle grazing on green spring annual pasture. *Onitis alexis,* a larger species, is somewhat less affected by the same seasonal changes in dung quality. The rate of egg production of *Euoniticellus intermedius* on dung collected from cattle grazing on dry winter pasture in summer rainfall regions of Australia is 30% of that on dung from green summer pastures.

Egg Laying

The adult female reproductive system of Scarabaeinae has only a single ovary, consisting of a single ovariole. Newly emerged beetles have no differentiated oocytes in their single ovary. The eggs develop sequentially during a period of maturation feeding. The terminal oocyte is the only one ready to be laid at any one time. If conditions are unsuitable for oviposition, the oocyte is extruded from the ovariole into the hemocoele and nutrients are resorbed.

Fecundity of scarabaeine dung beetles is very low, but they produce relatively large eggs. The length of an average egg is about 33% of the adult female body, and the volume of the egg is about 2.5% of that of the female. Most species probably produce as few as 20 eggs/female/year in the field, because weather conditions or dung quality are rarely ideal for adult reproduction, and some species produce 5 or fewer. Competition for dung at times beetles are ovipositing is high. Adult investment in nesting behavior is high to enhance the success of the offspring that are produced. This is illustrated here in terms of adult male/female reproductive behavior and in female brood care.

ADULT REPRODUCTIVE BEHAVIOR

Recognition of beetles of the same species as mates is important, as beetles are frequently present in large numbers in fresh dung. Male beetles court females by tapping them with their head and forelegs prior to successful copulation. The males of many species produce pheromones that are probably involved in close-range species recognition and in sexual attraction. Pheromones are released via pumping movements from forelegs, or from abdominal sternites, depending on the species.

Male-to-male intraspecific competition occurs in tunnels in the soil for the possession of a female making a brood mass where she is to lay her egg. Of the two males, the larger is usually successful in pushing the smaller beetle away from the tunnel and the female. This behavior is widely reported in species in both the larger genera such as *Scarabaeus, Kheper,* and *Typhoeus,* and the smaller genera such as *Onthophagus.* The size of male beetles and of horns arising from their head, thorax, and clypeus can be very variable. In many of the smaller species, beetles exhibit dimorphic male morphology. Both large horned major and small hornless minor morphs (forms) coexist in the field. Females produce more brood masses, and the brood masses are larger, in the presence of horned males than in the presence of hornless males. Offspring size is determined by the size of the brood mass used to provision the larva. The horned males assist the females in dung provisioning, providing each egg with more dung. They also guard the tunnels where the females are producing brood masses from other males. The hornless males show alternative mating strategies, by sneak mating with unguarded females, but they do not assist the females in providing dung for brood masses after mating.

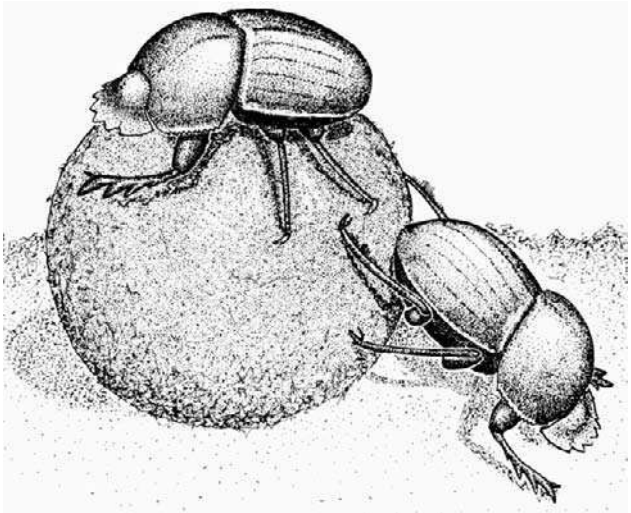

FIGURE 4 A pair of *K. nigroaeneus* rolling a brood ball. The female sits on the ball while the male rolls it backward (beetles approx 2 cm in length). (Reproduced, with permission, from Edwards and Aschenborn, 1988.)

BROOD CARE AND SUBSOCIAL BEHAVIOR

Investment in nesting behavior that enhances the survival of offspring is evident in brood care by adult female beetles. The female remains underground with the brood ball, providing care of her offspring during its development. In *Copris,* a male and a female make a chamber in the soil and carry dung down to construct a large dung cake of about 100 g. The male then leaves and the burrow is sealed. The female cuts up the cake into about four brood balls and lays one egg in each. The mother then remains with the brood balls during the development of her larvae, caring for them, repairing damage to the balls by the larvae, and removing fungal growth from the outside. Such behavior doubles the survival rate of the offspring. The mother emerges from the soil after her offspring have emerged from their brood balls.

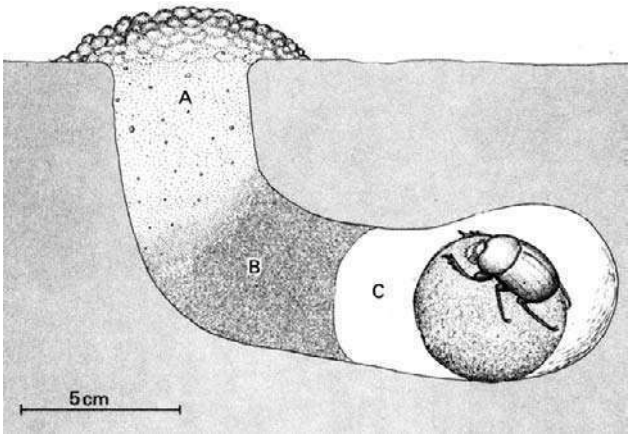

FIGURE 5 Female *K. nigroaeneus* in a brood chamber with a single brood ball. Male has left and packed the chamber entrance closed. (A) Loose soil, (B) hard packed soil, and (C) chamber. (Reproduced, with permission, from Edwards and Aschenborn, 1988.)

An extreme case of low fecundity occurs with *Kheper nigroaeneus,* a large ball-rolling species, which produces only a single offspring on each nesting occasion. A pair of beetles forms a single large dung ball over 4 cm in diameter. The male rolls the ball away from the dung pat while the female clings to the top of it (Fig. 4). The ball is buried via a diagonal tunnel to a chamber about 14 cm deep. The male then leaves the nest while the female beetle stays in the chamber and lays a single egg. She remains underground with the ball for the 12 weeks taken for egg and larval development (Fig. 5). As with *Copris* spp., the presence of the female greatly enhances offspring survival. After the new adults have emerged, the mother leaves the nest.

ECOSYSTEM-LEVEL PATTERNS

Species Richness and Competition

Worldwide there are about 4000 species of scarabaeine dung beetles. Local species richness is generally related to the species richness of large mammalian grazing animals, although cattle dung does support a rich dung beetle fauna around the world. The most competitive assemblages of dung beetles occur in tropical grasslands, where up to 120 species can be present in local areas; here, competition for the dung resource is high.

Intraspecific competition for dung occurs in a range of beetle species, expressed as a reduction in the number of eggs laid/female at high densities. There is also evidence for interspecific competition in which the presence of beetles of one species reduces egg production of a second species. In these instances the competition is frequently asymmetric, and the larger species has a greater effect on the smaller species than the other way round, particularly at high beetle densities. The competitive advantage of larger beetles is associated with preemptive dung burial, whereby they bury a greater proportion of dung in the first day.

Interspecific competition is avoided where there are differences in patterns of dung use and reproductive strategies. Species can be allocated into functional groups on the way they use dung and where it is buried under the pats. Competition is reduced also by the occurrence of aggregated spatial distributions between the discrete dung pats as a result of timing of optimal flight activity and of dung or soil preferences. Of the species present in a region, only some will have a preference for the niche represented by each dung pat, including the time and place at which it is dropped.

Australian Dung Beetle Program

The arrival of farmers from Europe in the 19th century led to a major change in the Australian landscape, resulting from planting pastures and introducing domesticated grazing animals such as sheep and cattle. Native beetles, active only for restricted periods of the year, and occupying mainly heath and other undisturbed habitats, were unable to use effectively the large quantities of cattle dung in the newly created pasture habitats. Dung fauna in these pastures thus consisted of a high abundance of dung-breeding pest flies, but few dung beetles or predatory beetles. George Bornemissza of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) suggested introducing to Australia exotic scarabaeine dung beetles adapted to open pasture habitats and active mainly at the time of year native beetles were not, to correct this imbalance in pastures, to improve nutrient cycling, and to control nuisance flies breeding in the dung. A total of 46 species of scarabaeine dung beetles were introduced into Australia between 1967 and 1995, of which 26 species are established. The project has greatly increased the rate of dung recycling in Australia and reduced the population of at least one important dung-breeding nuisance pest, the Australian bushfly, *Musca vetustissima.*

See also the Following Articles

Coleoptera • Cultural Entomology • Mating Behaviors • Parental Care

Further Reading

Bornemissza, G. F. (1976). The Australian dung beetle project—1965–1975. Australian Meat Research Committee Review No. 30, 1–30.

Doube, B. M. (1990). A functional classification for analysis of the structure of dung beetle assemblages. *Ecol. Entomol.* **15,** 371–383

Edwards, P. B., and Aschenborn, H. H. (1988). Male reproductive behaviour of the African ball-rolling dung beetle, *Kheper nigroaeneus* (Coleoptera: Scarabaeidae). *Coleopterists Bull.* **42,** 17–27.

Halffter, G., and Matthews, E. G. (1999). "The Natural History of Dung Beetles of the Subfamily Scarabaeinae." Reprint Medical Books, Palermo.

Hanski, I., and Cambefort, Y. (eds.) (1991). "Dung Beetle Ecology." Princeton University Press, Princeton, NJ.

Ridsdill-Smith, T. J. (1991). Competition in dung insects. *In* "Insect Reproductive Behaviour" (W. J. Bailey and T. J. Ridsdill-Smith, eds.). Chapman & Hall, London.

Waterhouse, D. F. (1974). The biological control of dung. *Sci. Am.* **230,** 100–109.

E

Earwig

see *Dermaptera*

Ecdysis

see *Molting*

Ecdysteroids

Michael E. Adams
University of California, Riverside

The ecdysteroids are steroid hormones that, in combination with juvenile hormones, program gene expression appropriate for each stage of insect development. Beginning at embryogenesis, ecdysteroid levels rise transiently during each stage to initiate molting. Depending on the level of juvenile hormone, elevated ecdysteroids mobilize nuclear receptors of several types that bind directly to DNA to either promote or suppress gene expression. During the adult stage of some insects, ecdysteroids play important roles in reproduction, principally in the development of gametes.

DISCOVERY

Early in the 20th century, scientists began to demonstrate roles for circulating hormones in the development and maturation of insects. Stefan Kopec, working in Poland, observed that removal of the brain in gypsy moths (*Lymantria dispar*), caused developmental arrest. Amazingly, reimplantation of the brain allowed development to resume. In other experiments, Kopec used an experimental tool called a ligature to demonstrate the timing of hormone release from the brain. A ligature is applied with a loop of string that is pulled tight, like a tourniquet, interrupting blood flow between front and back blood compartments. If a ligature is tied around a last instar before the "critical period" of 5 to 7 days, the front part of the animal will develop pupal cuticle, whereas development in the back part will be arrested (Fig. 1). However if the ligature is applied after the critical period, both sides will develop. These observations implied that something was released from the front end of the animal during the interval between days 5 and 7 of development. Consistent with Kopec's earlier work, this signal was later shown to come from the brain. These ingenious experiments illustrated the surprising fact that the insect brain, in addition to its well-known function in electrical signaling, is a secretory organ, controlling developmental processes through release of a hormone into the bloodstream.

The experiments of Kopec were published in two classic papers in the early 1920s, but little happened for about a

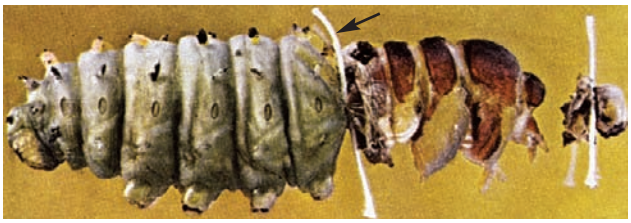

FIGURE 1 Ligature tied prior to the critical period (arrow) leads to arrested development in the posterior part of the animal, where green larval cuticle is retained. The thorax synthesizes a new, dark pupal cuticle in response to ecdysteroid release by the prothoracic glands. [Modified from Farb, P. (1962). "The Insects." Time, Incorporated, New York.]

decade. The notion that hormones program insect development was not readily accepted because thinking was dominated by mid-19th century studies of reproductive hormones in birds. These experiments showed that implantation of the avian male testes causes masculinizing effects, whereas removal of the testes produced loss of male characteristics. Such experiments performed in insects had no apparent effect, and as a consequence it became accepted that insects did not engage in hormonal signaling.

Nevertheless, Vincent Wigglesworth, S. Fukuda, and later Carroll Williams extended Kopec's work, by providing evidence for a second signal located in the thorax of the insect that is released in response to the brain hormone. It became evident that implantation of the brain, as Kopec had done, worked only if the brain was placed in the thoracic area. When the brain was implanted into an isolated insect abdomen, developmental arrest persisted. The source of this second factor was the prothoracic gland. Williams showed that, if both the brain and the prothoracic glands were implanted into an isolated abdomen, development resumed. The prothoracic glands alone could accomplish this, provided they had prior exposure to the brain. It therefore became evident that the brain provides a hormonal signal that induces release from the prothoracic glands of a "molting hormone" that is critical to promoting growth.

CHEMICAL NATURE OF ECDYSTEROIDS

Attempts to isolate the molting hormone, or "ecdysone" as it came to be known, began in the 1940s and continued for the next 10 years, until the efforts of two German chemists, Butenandt and Karlson, yielded microgram quantities of pure ecdysone. Because of its water solubility, ecdysone at first was not recognized as a steroid. Soon X-ray studies of ecdysone crystals revealed it as a unique steroid with five hydroxyl substituents, accounting for its ability to dissolve in water (Fig. 2).

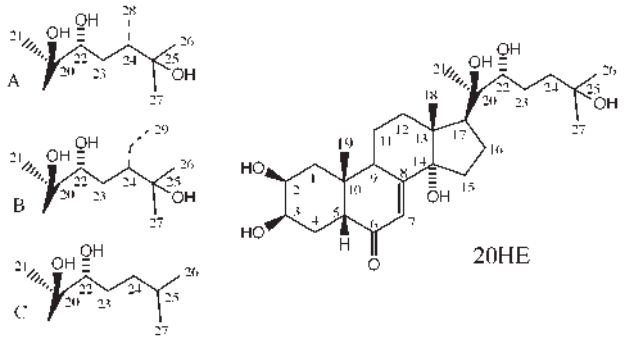

FIGURE 2 The structure of 20-hydroxyecdysone (20HE) and related ecdysteroids. The hydroxyl substituent of 20HE at position 20 confers biological activity to ecdysone. Many insects obtain phytoecdysones from plants and convert them to biologically active forms. These molecules differ from 20HE only in the number of carbons in the alkyl side chain, shown for (A) makisterone A (28 carbons), (B) makisterone C (29 carbons), and (C) ponasterone A, which lacks the hydroxyl group at position 25.

The first ecdysone structure soon was recognized as a precursor to the biologically active material. Upon release from the prothoracic glands, ecdysone, or its 3-dehydro form in some insects, is converted to the active form upon arriving at target tissues. The tissues capable of responding to the hormone produce the enzymes needed to attach a single hydroxyl group to the carbon at position 20, making it 20-hydroxyecdysone (20HE). This substance had already been identified in crustaceans as crustecdysone. Identical molecules function as the biologically active hormone in insects and in crustaceans.

Not long after the identification of ecdysone, phytoecdysones were discovered in plants. The first, ponasterone A, was discovered by Koji Nakanishi, and an identical molecule was later found in some crustaceans. Ponasterone A, which differs from ecdysone in lacking a single hydroxyl at position 25, eventually proved useful in radiolabeled form for the characterization of ecdysteroid receptors. Hundreds of phytoecdysones have been identified, including 20HE itself. Reasons for the presence of phytoecdysones in plants are unclear, but they may serve defensive roles by disrupting the growth of herbivorous insects.

As more insects were examined, additional configurations of the basic ecdysone structure were found. This group of molecules now is collectively known as ecdysteroids. Insects require cholesterol in the diet to synthesize ecdysteroids. Often phytosterols such as campesterol and β-stigmosterol are converted to produce hormonally active makisterones A and C, respectively (Fig. 2).

SOURCES AND FUNCTIONS FOR ECDYSTEROIDS

During the immature stages of development, the chief source of ecdysteroids is the prothoracic gland, a diffuse organ located in the thorax. In higher insects such as flies (Diptera) and bees (Hymenoptera), the prothoracic gland has become part of a composite structure called the ring gland. The precursor ecdysone or 3-dehydroecdysone is synthesized and immediately released into the blood. Then 20-hydroxymonoxgenase is converted to 20HE in target tissues such as epidermal cells, salivary glands, fat body, nervous system, gut, and imaginal discs.

During each stage of development, feeding and growth are followed by a sudden elevation of ecdysteroids, which induces animals to stop feeding and to engage in a new round of gene expression appropriate for the next stage. The epidermal cells begin to secrete a new layer of cuticle and to take back most of the old cuticle, recycling the chitin and protein recovered into the new layer. If the next stage is to be larval, ecdysteroids circulate with high levels of juvenile hormones, and a new set of larval characters is expressed. However, when larval development is complete, metamorphosis is signaled by short ecdysteroid pulses in the absence of juvenile hormones. Completely new structures are created as the insect undergoes the process of changing from a caterpillar into a reproductive, winged adult.

The actions of ecdysteroids can be observed almost immediately after their release as "puffs" on the large polytene chromosomes in the salivary glands of some flies. Even before ecdysteroids were chemically identified, Clever and Karlson noticed puffs within hours of injecting the natural hormone and suggested that 20HE acted to regulate gene expression. This was an excellent insight, for it is now known that ecdysteroids have direct actions to either activate or suppress the expression of many genes via nuclear receptors and that the puffs correspond to gene loci at which transcription takes place.

A model explaining relationships between temporally distinct puffs was proposed by Michael Ashburner in the early 1970s. He found that just after ecdysteroid release, a set of early puffs could be observed, followed by early-late and late puffs. With the identification of ecdysteroid receptors, his model now views early and early-late puffs as evidence of transcriptional activity induced by ecdysteroid receptors (EcRs) and other nuclear receptors. The gene products resulting from these events then give rise to late puffs and also repress further transcriptional activity at early puffs. In this way, a complex but coordinated series of gene expression events occurs, initiated by EcRs.

Having been present throughout larval development, the prothoracic glands degenerate during metamorphosis. Nevertheless, ecdysteroids persist in the adult stage, where they play important reproductive functions. The source of ecdysteroids in adults remained a mystery for many years, although anecdotal accounts implicated the mobile oenocytes as a possibility. In the 1970s, Henry Hagedorn provided a breakthrough, showing that mosquito ovaries produce large quantities of 20HE and that the hormone is required for vitellogenesis, or yolk deposition in developing oocytes. The precise source of ecdysteroids is the follicle cell layer surrounding the oocyte. Since Hagedorn's initial finding, ecdysteroids have been identified also in the testes, where they are involved in sperm maturation.

MOLECULAR BASIS OF ECDYSTEROID SIGNALING

Ecdysteroids belong to a large class of steroid chemical signaling molecules. Because of their lipophilic character, they pass through the cell membrane easily. Whether they affect the cell upon entry depends on the presence or absence of specialized proteins belonging to a large class of soluble, diffusible nuclear receptors. Nuclear receptors get their name from the part of the cell in which they conduct their business, which is the regulation of gene expression. Early in the 1990s, David Hogness and colleagues discovered in fruit flies a class of nuclear receptors they called EcRs. Upon their activation by ecdysone binding, EcRs bind directly to DNA at "ecdysone response elements" (EcRs) to turn genes on or off. Further work showed that to bind with high affinity to DNA, the EcR first finds a partner protein to form a doublet, or "dimer" complex. It is this protein dimer that, together with coactivator proteins, binds to EcREs, resulting in regulation of gene expression.

It is well known that EcRs affect cells in many different ways, causing some to differentiate into muscles and some into glands, and others to form particular kinds of cuticle and cuticular structures appropriate for a larva, a pupa, or an adult. The process by which ecdysteroid receptors encode this diversity of effects is very complicated, and many questions are under current investigation. But it is known already that several different types of ecdysteroid receptor occur in insects, including EcRA, EcrB1, and EcRB2. These receptor "subtypes" occur at different stages of development and can be specific to particular tissues. Thus at least some of the stage-specific effects seem to depend on this diversity of receptors. Another point is that the partners with which they bind to DNA probably vary substantially, providing a further level of combinatorial diversity.

ECDYSONE-BASED INSECT CONTROL

One way of managing pest insect populations is to target unique aspects of their physiology. Because molting is a particularly unique aspect of insect biology, scientists have attempted to learn more about hormonal control systems to be able to design "magic bullets" targeting only insects. The first complete synthesis of ecdysone was accomplished by John Siddall in the late 1960s, but development of such a complicated molecule has not proven to be commercially viable. As insecticides go, the ecdysone structure itself is rather complex and would be expensive to produce on a large scale; moreover, it is too unstable in the environment to be useful in field applications.

Nevertheless, chemicals with unexpected biological activities are produced every year by the chemical industry, and it has become routine to test these compounds, using all available biological assays. This has led to unexpected successes on many occasions, including the serendipitous discovery of the first ecdysone agonists, or "ecdysanoids," by Rohm & Haas in the 1980s. Keith Wing and colleagues at Rohm & Haas found that a series of bisacylhydrazines had astounding ecdysone-like activity, even though their structures were not recognizable as steroidlike. Application of these compounds caused premature insect molting, differentiation of cells, and death of insects due to improper

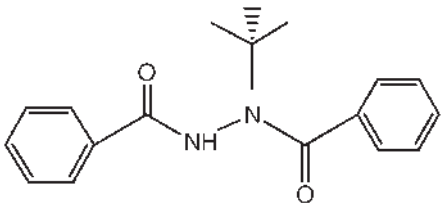

FIGURE 3 The structure of RH5849, one of the first ecdysone agonists developed at Rohm & Haas.

programming of development. The ecdysanoids have been commercially developed for agricultural pest control and represent an encouraging example of how targeting basic insect physiological processes can lead to safer, more environmentally sound insect control agents.

See Also the Following Articles

Development, Hormonal Control of • Juvenile Hormone • Metamorphosis • Molting

Further Reading

Ashburner, M., Chihara, C., Meltzer, P., and Richards, G. (1973). Temporal control of puffing activity in polytene chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* **38,** 655–662.

Clever, U., and Karlson, P. (1960). Induktion von Puff-Veranderungen in den Speicheldrusenchromosomen von *Chironomus tentans* durch Ecdyson. *Exp. Cell. Res.* **20,** 623–626.

Farb, P. (1962). "The Insects." New York, Time Incorporated.

Hagedorn, H. H., O'Connor, J. D., Fuchs, M. S., Sage, B., Schlaeger, D. A. and Bohm, M. K. (1975). The ovary as a source of alpha-ecdysone in an adult mosquito. *Proc. Natl. Acad. Sci. U.S.A.* **72(8),** 3255–3259.

Koelle, M. R., Talbot, W. S., Segraves, W. A., Bender, M. T., Cherbas, P., and Hogness, D. S. (1991). The *Drosophila* EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell* **67(1),** 59–77.

Nakanishi, K. (1992). Past and present studies with ponasterones, the first insect molting hormones from plants. *Steroids* **57(12),** 649–657.

Riddiford, L. M., Cherbas, P., and Truman, J. W. (2000). Ecdysone receptors and their biological actions. *Vitam. Horm.* **60,** 1–73.

Wigglesworth, V. B. (1983). Historical perspectives. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 7, pp. 1–24. Pergamon Press, Oxford, U.K.

Wing, K. D., Slawecki, R. A., and Carlson, G. R. (1988). RH 5849, a nonsteroidal ecdysone agonist: Effects on larval Lepidoptera. *Science* **241,** 470–472.

Eggs

Diana E. Wheeler

University of Arizona

Most insects use fertilized, nutrient-rich eggs to reproduce. In some insects, however, eggs can develop into embryos without fertilization or egg nutrients. And in a few insects, embryos develop directly inside the female insect's body. Nutrient-rich eggs are a resource to parasitoids and predators, as well as to embryos, and so insects use a variety of methods to protect their eggs.

TYPICAL INSECT EGGS

Typical insect eggs contain nutrients to support embryogenesis and produce newly emerged first instars. Most eggs contain large amounts of lipid, for use as building material and energy, and yolk proteins, for the amino acids needed to build a larval insect body. Eggs also contain a cytoplasmic "starter kit" for development that includes cellular machinery such as ribosomes. In species with symbiotic bacteria or protists, eggs are inoculated with a small population of the mutualistic microbes.

Insects typically have internal fertilization, and fertilized eggs contain one set of chromosomes from each parent. Eggs are laid in protected places in environments where young are likely to find food. For example, many butterflies lay their eggs on larval food plants, mosquitoes lay their eggs in water in which larval food grows, and parasitoids lay eggs in, on, or near a host insect. Because mature eggs are usually covered by a thin shell, there must be a way for sperm to penetrate the shell before it is laid, and a way to accommodate water balance and respiratory needs afterward. Sperm enter through an opening called the micropyle. Water and air can pass through specialized regions of the eggshell and embryonic membranes. Finally, some insect groups, remarkably, produce offspring without sperm, egg nutrients, or both.

INSECTS FROM EGGS WITHOUT FERTILIZATION

The sex of hymenopteran insects normally is determined by the number of sets of chromosomes. Unfertilized, haploid eggs have only their mother's set and develop as males. Fertilized, diploid eggs have chromosome sets from both their parents and develop as females. Mated females have control over when sperm is released from the spermatheca to fertilize eggs. Therefore, they can adjust their offsprings' sex ratio in response to a variety of cues. Hymenoptera are particularly susceptible to manipulation of sex determination by parasitic microbes. *Wolbachia,* for example, can alter sex determination so that haploid, and therefore unfertilized, eggs develop as females.

Aphids have complex life cycles that often include female forms that reproduce parthenogenetically, producing female clones of themselves. In such aphids, diploid oocytes form in the germarium and begin development without fertilization.

INSECTS FROM EGGS LACKING NUTRIENTS

Stored nutrients are a major feature in typical insect eggs, but diverse insects have reduced amounts of yolk or lack it entirely. Obviously, if nutrients for embryonic development are not provided in eggs, they must come from another source. Two major alternate sources are the mother and other insects, which can serve as hosts for both embryonic and larval development.

Females that provide nutrients to their embryos, in addition to or instead of egg materials, are termed viviparous. The parthenogenetic aphids described earlier are viviparous: they produce first instars rather than eggs. Young aphid embryos shed their covering of follicle cells and break the strands of tissue that connect them to the germarium. Then, they absorb the necessary nutrients directly from the mother's body.

The viviparous Pacific beetle cockroach, *Diploptera punctata,* produces eggs with insufficient yolk to support complete embryonic development. During pregnancy, females make a supplementary protein, termed roach milk, that is taken up by embryos. Females in a few groups, such as the order Strepsiptera, are neotenous. Neotenic insects do not go through a full metamorphosis, and they reproduce during the larval stage. Strepsipteran females lack oviducts, so that eggs produced in the ovaries are released into the blood. The eggs can contain some, but insufficient amounts of yolk. To fertilize these eggs, sperm must move from the genital canal into the blood. When embryogenesis is complete, larvae use the genital canals to leave the mother's body.

Some parasitoids also develop inside other insects, but here the insect is the parasitized host. Female parasitoids that oviposit directly inside other insect eggs, larvae, or adults are likely to produce small eggs with little or no nutrients. The parasitoid embryos, lacking their own supply of nutrients, then use the host's body to supply materials for their own development.

ECOLOGY OF EGGS

Eggs, the first life stage of insects, can be important ecologically. For example, eggs are the diapausing stage in many insects, with embryogenesis stopping at a species-specific point. Eggs of silkworms (*Bombyx mori*) have an obligatory diapause that coincides with winter under natural conditions. Embryonic diapause has been studied extensively in silkworms because delayed development can be a nuisance from an industrial perspective. Gypsy moth eggs also diapause during winter, but they arrest development at a later stage. Embryos complete embryogenesis and overwinter as unhatched larvae.

Daylength is the most common cue for inducing diapause, but moisture, temperature, and food quality can also be important. The environmental cues that cause eggs to stop developing can be detected by females and then passed on to signal the eggs. Alternatively, the cues can be detected directly by the eggs and embryos. Later, eggs must break diapause in response to another environmental combination of daylength, temperature, and moisture.

Eggs are rich sources of nutrients and therefore pose a great "temptation" to parasitoids, parasites, and predators. Insects protect their eggs in a variety of ways. For example, eggshells can be thick and protective, or cryptic (difficult to detect). Eggs can be laid in protected places. Primarily females, but sometimes males, can contribute chemical repellents or toxins to eggs to deter attacks. A variety of insects stay with their egg masses and actively protect them.

See Also the Following Articles

Diapause • Ovarioles • Parthenogenesis • Spermatheca • Symbionts • Vitellogenesis

Further Reading

Chapman, R. F. (1998). "The Insects: Structure and Function," pp. 298–312. Cambridge University Press, Cambridge, U.K.

Hinton, H. E. (1981). "Biology of Insect Eggs," Vols. 1 and 2. Pergamon Press, Oxford, U.K.

Egg Coverings

Diana E. Wheeler
University of Arizona

The protective coverings females provide their eggs reflect the full range of environments exploited by insects. Egg coverings fall into two major categories: those produced by follicle cells and those produced by accessory glands of the reproductive tract. Follicle cells secrete the chorion, or insect eggshell. The design of the chorion is important in fertilization, egg respiration, and water balance. Coverings produced by accessory glands provide additional protection from the elements, predators, and parasites.

THE EGGSHELL

Layers of the Eggshell

As the oocyte develops, follicle cells secrete structural proteins, along with other substances, as layers. The making of insect eggshell has been best documented in the fruit fly *Drosophila melanogaster.* The vitelline envelope, sometimes considered to be the first layer of the chorion, is the first and innermost layer produced. Next, a layer of wax is secreted, giving the egg greater resistance to desiccation. Then, several more chorionic layers are produced, commonly with sheetlike inner and outer layers separated by a pillared region enclosing air spaces. In insects such as Odonata and Ephemeroptera, which lack accessory glands for producing additional secretions, follicle cells may also produce gelatinous or adhesive coatings. Figure 1 shows the chorionic structures of *Antherea polyphemus,* the polyphemus moth.

Functions of the Eggshell

The eggshell is a layer of armor protecting the egg and developing embryo from the elements, predators, and parasites. Eggshell shape, texture, and color can also provide protection through camouflage and warning coloration.

The protection must be breached, however, to allow for vital functions. First, sperm must be able to enter and fertilize mature eggs. Micropyles, which are openings allowing sperm access to the egg interior, often have distinct architecture, such as the conical protrusion in *D. melanogaster.* Some insect eggs have multiple micropyles.

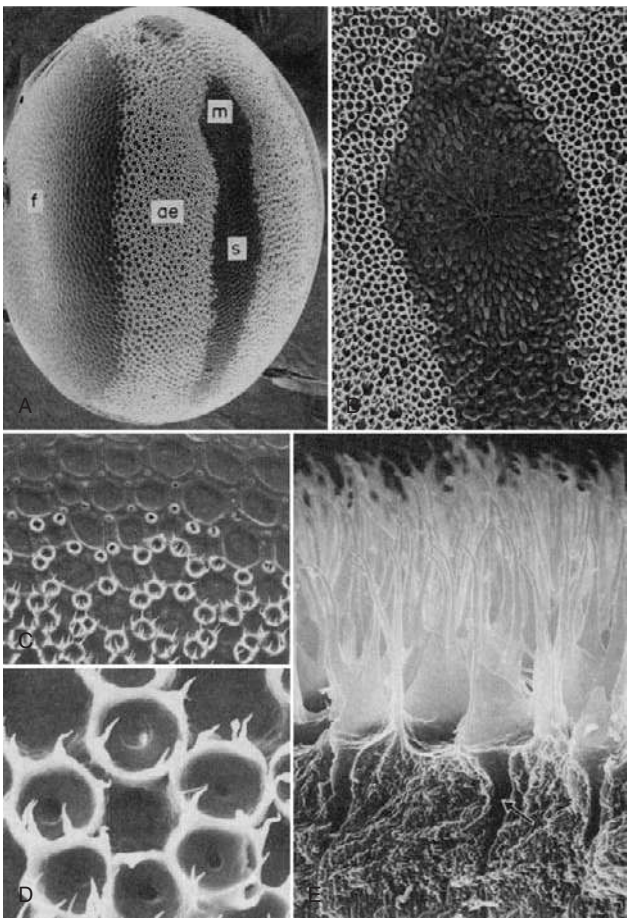

FIGURE 1 Chorionic structures of the polyphemus moth *Antherea polyphemus.* (A) Regional differences in chorion structures. Note micropyle (m) and aeropyle (ae) regions. (B) Micropyle region magnified. (C–E) Aeropyle region. [From Margaratis, L. H. (1985). Structure and physiology of eggshell. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," Vol. 1 (Kerkut, G. A., and Gilbert, L. I., eds.), Figs. 35 and 72A–C. Pergamon Press, Oxford, U.K.]

Second, the developing embryo is very active metabolically and must respire. Various architectural features of the chorion allow the embryo to exchange gases in both air and water. The interior chorion layer, with its pillared structure, harbors a thin layer of air that can connect directly to the atmosphere by openings termed aeropyles. In moist environments, when air trapped by the inner chorion cannot access atmospheric air directly, the trapped air layer functions as a plastron. Plastrons allow diffusion of gases between water and the air space based on differential partial pressures. The proportion of the egg surface that functions as plastron or aeropyle reflects moisture levels generally found in a particular habitat. To enhance respiratory capacity, eggshells may have extensions, called respiratory horns, that increase the surface area for gas exchange.

Third, some insect eggs are able to absorb environmental water to replace that lost by evaporation. Hydropyles, which are regions specialized for water uptake, may include chorionic layers in addition to the oocyte membranes interior to the vitelline envelope.

Finally, the embryo itself must be able to emerge from the eggshell when development is complete. Lines of weakness built into the eggshell, based on the interruption of some of the chorionic layers, create a "door," or operculum, through which the first-stage larva can emerge.

OTHER EGG COVERINGS

The colleterial glands (based on *colle* from Greek for glue) are accessory glands of the female's reproductive system that produce egg coatings. Most simply, glue is produced to attach eggs to a substrate. In addition, some coatings deter predators or parasites chemically.

Insects in the orthopteroid orders Blattodea, Mantodea, and Orthoptera, as well as some beetle taxa, secrete an egg case or pod surrounding the eggs, to give additional protection from desiccation and predation. As with eggshells, these egg casings must be constructed to allow respiration and hatching. In cockroach eggs, the air space surrounding each embryo opens into a ventilated air duct in the keel. In grasshoppers, the colleterial gland secretions are churned into a froth in which the eggs are suspended. The entire oviposition hole is filled with the frothy material, which then hardens to form a plug. Mantids produce a similar substance from their colleterial glands, which they mold into an egg case that is attached to a flat surface or suitable vegetation.

Beetles are also known to produce egg cases, with the most complex occurring in the cassidine Chrysomelidae. Less complex ootheca have been reported in a variety of other beetle groups as well. Colleterial glands in hydrophilid beetles produce silk that is used to form a cocoon for the egg mass.

Various insect taxa use additional material to enhance protection for eggs. In some beetles, fecal material and/or secretions from anal glands apparently provide chemical defense. Some Lepidoptera use urticating (or irritating) hairs from the larval skin to protect eggs. As adult females emerge from the cocoons, they pick up discarded larval setae (hairs) with their anal tufts and later deposit them on eggs. Nonurticating adult scales, most commonly from the anal tuft, are also used by some species to create an effective physical barrier.

See Also the Following Articles

Accessory Glands • Ovarioles • Reproduction, Female

Further Reading

Hinton, H. E. (1981). "Biology of Insect Eggs," Vols. 1 and 2. Pergamon Press, Oxford, U.K.

Margaratis, L. H. (1985). Structure and physiology of eggshell. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," Vol. 1 (Kerkut, G. A., and Gilbert, L. I., eds.), pp. 153–230. Pergamon Press, Oxford, U.K.

Embiidina

(Embioptera, Webspinners)

Edward S. Ross
California Academy of Sciences

Embiidina (Embioptera, "webspinners," "foot spinners," "embiids") are warm-climate-adapted insects. Only about 300 species are named, but the order perhaps comprises about 2000 species. Embiids are perhaps most closely related to stoneflies and stick insects (Phasmida) but, probably since Carboniferous or Permian times, have followed their own specialized evolutionary line. The most peculiar feature of all embiids, regardless of developmental stage or sex, is an ability to spin hundreds of strands of silk with each stroke of the greatly enlarged, gland-packed basal segment of the foretarsi.

The silk is formed into narrow galleries serving as protective runways in or on the food supply—weathered bark, lichens, moss, or leaf litter. In arid regions the galleries extend deep into soil and there serve as refuges from heat and desiccation. The primordial habitat is tropical forest, where predation-reducing galleries of most species radiate on edible surfaces of tree trunks. When disturbed, an embiid quickly darts backward into the depths of the labyrinth or into a crevice beneath such cover.

HABITS AND SPECIALIZATION

Except for short, hazardous dispersal of adults, embiids almost never leave the shelter of their self-created microenvironment, and most of the order's anatomical and behavioral characteristics foster very smooth, rapid, reverse movement in narrow galleries. Such specializations include the following:

1. A linear, short-legged, supple body with the head projected forward (Figs. 1 and 2).

FIGURE 1 Typical adult female: *"Aposthonia"* n. sp. (family Oligotomidae) of Thailand; body length, 18.00 mm. Females of all species are apterous, and neotenic.

FIGURE 2 Typical adult male, wings in repose: *Antipaluria caribbeana* (family Clothodidae) of northern Venezuela; body length, 20.00 mm. Males of species inhabiting arid regions often are apterous, or subapterous.

2. Rapid reverse movement aided by great enlargement of depressor muscle of the hind tibiae.
3. Highly sensitive cerci serving as tactile guides during backward movement.
4. Complete apterism, and thereby elimination, of projecting structures in all females due to endocrinal arresting of development of adult anatomy (neoteny or pedomorphosis) at an early nymphal stage. Males of many species, especially in arid regions, also are apterous or subapterous.
5. Flexibility and forward folding of wings of adult males (Fig. 3)—an advantage in reverse movement as a means of reducing the barb effect against gallery walls and thus a slowing of reverse movement and thereby increased predation.
6. A compensating ability temporarily to stiffen wings for flight by increasing the blood pressure in the full length of the anterior radius (RA), the cubitus, and the anal vein (Fig. 4). This unique wing specialization must have early evolved in both sexes but has been supplanted in females by complete apterism through neoteny.

FIGURE 3 Forward wing-flip during defensive, reverse movement of a male: *"Aposthonia"* n. sp. (family Oligotomidae) of Thailand. In repose, wings are flexible; when extended for flight they are temporarily stiffened by blood pressure, particularly in the full length of the anterior radius vein.

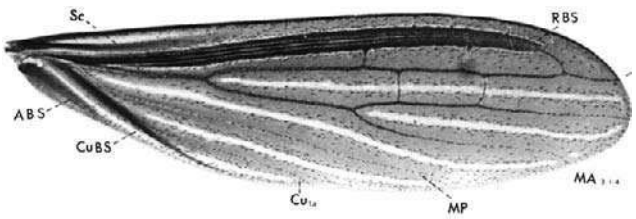

FIGURE 4 Typical embiid forewing: *Pararhagadochir trinitatis* (family Embiidae), Venezuela; wing length 10.0 mm. Most important are the blood sinus veins, especially the anterior radius (RBS, radial blood sinus); less important are the cubital blood sinus (CuBS) and the anal blood sinus (ABS). Hyaline stripes between veins also characterize embiid wings.

In spite of the great antiquity of the order, one that must predate the fragmentation of Pangaea, embiids constitute a single adaptive type of organism, one with a single "ground plan." As in earthworms, no great diversity of body form has occurred because of the physical uniformity of the galleries. As a result, it is difficult to sight-recognize various higher taxa. However, ages of evolutionary diversification are reflected in complex external genitalia of adult males, their head structures, and other characters useful in classification. Females and nymphs are difficult to identify without associated adult males. Adult females offer some anatomical characters in their paragenital sternites and hind tarsi; but size and coloration are most useful for species recognition of females.

RANGE

Embiidina are endemic to all continental landmasses presently in tropical latitudes. The principal evolutionary centers, in order of importance are Africa, the Americas, tropical Asia, and Australia. There also are natural extensions into adjacent temperate regions, such as southern United States and Europe. Several species, particularly of the Asian family Oligotomidae, have widely spread in both ancient and modern commerce and are the most frequently collected species of the order (males are attracted to lights).

Recent extensive, worldwide collecting by Ross has greatly increased the number of known taxa, but most await descriptions now in progress.

The most generalized species (family Clothodidae) occur in tropical South America. The large, diverse family Embiidae comprises a number of subfamilies found in South America, Africa and adjacent Palearctic regions, and Asia as far east as Myanmar. Southeastern Asia and Australia have the peculiar families Embonychidae and Notoligotomidae, and others soon to be described in the literature.

Anisembiidae are confined to the New World tropics and adjacent warm regions. The peculiar family Australembiidae is restricted to the eastern portion of Australia. The Oligotomidae occur almost entirely in tropical Asia and Australia. The large family Teratembiidae is mostly Neotropical and Afrotropical; only a few of its species occur in tropical Asia.

See Also the Following Articles

Phasmida • Silk Production

Further Reading

Ross, E. S. (1970). Biosystematics of the Embioptera. *Annu. Rev. Entomol.* **15,** 157–171.

Ross, E. S. (2000). EMBIA. Contributions to the biosystematics of the insect order Embiidina. 1: Origin, relationships and integumental anatomy of the insect order Embiidina. *Occas. Pap. Calif. Acad. Sci.* **149,** 1–53.

Ross, E. S. (2000). EMBIA. Contributions to the biosystematics of the insect order Embiidina. 2: A review of the biology of the Embiidina. *Occas. Pap. Calif. Acad. Sci.* **149,** 1–36.

Embryogenesis

Lisa Nagy
University of Arizona

Miodrag Grbić
University of Western Ontario, Canada

Embryogenesis is the process by which a larva or a juvenile is built from a single egg. The fertilized egg divides to produce hundreds of cells that grow, move, and differentiate into all the organs and tissues required to form a larva or juvenile. Embryogenesis is extremely diverse in different insect species. For each generalization that can be made about some aspect of insect development, numerous variations and multiple exceptions exist. In some species a single egg gives rise to several thousand larvae, in others, embryos devour their mothers prior to hatching. The most extreme variations are found among insects that parasitize other insects. This article presents a generalized view of some of the more regular features of insect development.

EGG MEMBRANES

Most insects lay eggs in terrestrial environments. For the most part, an insect egg forms a self-reliant developmental system that is generally impervious to the external environment, although sensitive to temperature, which serves as an important cue for many developmental events. Insect eggs are typically quite large, both in absolute dimensions and relative to maternal body size, and well-provisioned with yolk. Eggs vary from about 0.02 to 20 mm in length. To prevent desiccation, they are covered by some of the most resistant and impenetrable egg coverings found in the animal kingdom. Egg contents are protected by a vitelline membrane and covered by an external hard shell, the chorion. The chorion, vitelline membrane, and egg membrane itself surround the internal contents of the fertilized egg: the zygote nucleus and two types of macromolecule. Nutritive material such as yolk

proteins, lipids, and carbohydrates are used for nourishment and growth of the embryo. Patterning molecules (such as specialized proteins and mRNAs) direct major events in embryogenesis including establishment of embryonic polarity, segmentation, and gastrulation.

EGG CLEAVAGE

Development in nearly all animals involves a period in which the egg is subdivided into increasingly smaller cells. Compared with other animals, insect eggs undergo an unusual type of cleavage. In most animals, cleavage involves subdivision of both cytoplasm and nuclear material, to form individual cells called blastomeres. In contrast, the early cleavages of most insects involve only nuclear subdivisions (karyokinesis) and are not accompanied by cleavage of the cytoplasm (cytokinesis). This type of cleavage is called syncytial cleavage and results in the formation of a common compartment (syncytium), where up to several thousand nuclei reside (Fig. 1).

It is unclear how syncytial cleavage evolved in insects. The sister group to insects, the entognathans, has both total egg cleavage (Collembola) and syncytial cleavage (Diplura). Basal arthropods such as chelicerates, which include horseshoe crabs, spiders, and scorpions, have both syncytial and total cleavages. Even Onychophora, which are believed to be a sister group to Arthropoda, exhibit both syncytial cleavage (in oviparous species with yolky eggs) and total cleavage (in placental, viviparous species). However, it is reasonable to believe that syncytial cleavage in arthropods and insects evolved from total cleavage in the ancestor to the arthropods.

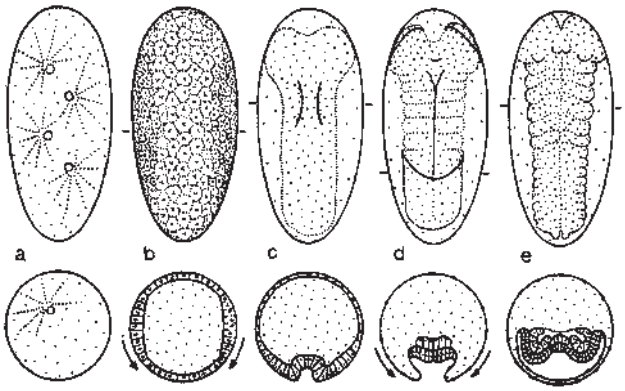

FIGURE 1 Diagram of the basic pattern of early insect embryogenesis: ventral views of eggs, anterior poles at top, are shown above cross sections at the levels indicated by bars in top row. (a) Syncytial cleavage. (b) Formation of the cellular blastoderm: arrows show that the lateral cells are coalescing toward the ventral surface to form the germ anlage. (c) Gastrulation. The prospective mesoderm begins invagination along the midline of the germ anlage. (d) Germ band after gastrulation, with segment borders (dotted) and amniotic folds forming: arrows indicate the movement of the serosal cells to enclose and cover the developing germ band. (e) Advanced germ band stage, with appendage buds, and transient coelomic sacs formed by the mesoderm. [Adapted from Sander, K., *et al.* (1985).]

BLASTODERM FORMATION

In most insects, the early syncytial cleavages proceed rapidly to form a syncytium with up 6000 nuclei. Syncytial nuclei are surrounded by islands of cytoplasm that separate the nuclei from one another. In general, these early cleavages are synchronized and, as the nuclei divide, they separate, such that there is regular spacing between them. Nuclei and associated cytoplasm are referred to as energids. Upon reaching a critical density, the energids migrate to the periphery of the egg. The arrival of the energids at the surface is sometimes visually apparent as bumps along the surface. At the periphery, the energids continue to undergo several rounds of division. In most insects, some of the nuclei remain in the yolk mass. These nuclei subsequently cellularize and become vitellophages, which serve to break down yolk to be used for embryo nutrition. Following the arrival of energids at the egg periphery, the egg membrane invaginates from the egg surface to surround each of the individual syncytial nuclei, marking the end of the syncytial stage of development. The single sheet of cells thus formed at the periphery of the egg is the cellular blastoderm (Fig. 1).

FORMATION OF GERM CELLS

In some species, distinctive granular inclusions can be found in the posterior cytoplasm of the egg. The cells that inherit these granules become the germ cells and eventually migrate into the ovaries or testes to become sperm and eggs. When the germ cells are ablated, the germline is missing and the individual is sterile, as noted in 1911 by Hegner. The nuclear energids (in some species, such as *Drosophila*) arrive at the posterior before reaching any other egg region, and the cells that will include this specialized cytoplasm cellularize earlier than any of the other cells. Germ cells rarely grow or divide during embryogenesis. The early segregation of these cells is thought to protect them from potential errors incurred during division and differentiation that might damage the genetic material necessary to build the next generation. In other species (e.g., most Lepidoptera), segregation of the germline occurs in the middle of the blastoderm; in other species no apparent germ cells can be detected at the blastoderm stage.

SEROSAL FORMATION

Only blastoderm cells destined to form the embryo coalesce to form the germ anlage, which later develops into the germ band. The cells that do not contribute to the germ anlage form an extraembryonic membrane called the serosa. In most species, the boundary between the future serosa and the future embryo ruptures, and the serosal cells migrate over and envelope the embryonic primordium and yolk cells (Fig. 1). However, there is variation in how the serosa is formed. In extreme cases like dipterans, the serosa cells do not migrate over the germ anlage but remain as a cluster of cells on one

side of the egg. In addition to the serosa, a second protective membrane, the amnion, forms later from the cells immediately adjacent to the germ anlage. These cells proliferate, flatten, and elongate. As they extend over the germ band, they resemble a sleeping bag being pulled up from the posterior and down from the presumptive head lobes of the germ band. Ultimately, the amnion cells meet in the middle of the embryo and form a single cell layer that lies between the embryo and the now separate serosa. In some derived, holometabolous species (e.g., *Drosophila*), this membrane has become vestigial, and the cells never migrate over the germ band.

GERM ANLAGE FORMATION

The size of the germ anlage varies relative to the length of the egg. In nearly all species, the nuclei arrive at the periphery to form a blastoderm that encompasses the whole surface of the egg. In metamorphic species, such as fruit flies and honey bees, the germ anlage forms from nearly the entire blastoderm surface. However, in direct developing species (such as the grasshopper and cricket), after the formation of a uniform synctyial blastoderm, nuclei migrate and aggregate near the posterior pole, where the germ anlage forms. The germ anlage thus forms from a relatively small proportion of the blastoderm. In the former case, called long-germ-type embryos, the complete body pattern (head, gnathal, thoracic, and abdominal segments) is patterned at the blastoderm stage and all segments appear nearly simultaneously in development. In contrast, in short-germ-type embryos, the head lobes, the most anterior trunk segments, and the posterior terminus are patterned first. Additional segments are added progressively, through proliferative growth. Some insects develop with germ types intermediate between these two extremes. The pleisiomorphic condition for insects is believed to be an intermediate-sized germ anlage. Short/intermediate germ embryogenesis is predominant in direct-developing hemimetabolous insects; more derived, metamorphic insects exhibit long-germ development. However, this division is not clear-cut. In some insect families, closely related species can exhibit both short- and long-germ types of development.

GASTRULATION

Formation of the cellular blastoderm is followed by gastrulation, the process of cellular invagination that results in the formation of a layered embryo comprising two germ layers. Cells that remain at the blastoderm periphery will form the ectoderm, and cells that invaginate below the ectoderm will form the mesoderm. The presumptive mesoderm in most species consists of a strip of cells along the ventral midline (Fig. 1). Gastrulation can happen in any number of ways: by the mesodermal cells invaginating simultaneously, as in *Drosophila,* or by cells invaginating sequentially, beginning at the anterior, while the gastrulation furrow progresses toward the posterior. Most often the presumptive mesoderm lies along the ventral midline, but in the apterygote thysanuran *Thermobia domestica,* cells migrate inward from every part of the germ band. Regardless of the mechanism of gastrulation, the end result is a bilayered embryo, with mesodermal precursors underlying the ectoderm.

SEGMENTATION

Segmentation refers to the process by which repeated units of similar groups of cells, the metameres, are created. Segmentation proceeds nearly simultaneously with gastrulation. Current understanding of the process of segmentation comes from the genetic dissection of development in *Drosophila* by Ed Lewis, Christiane Nüsslein-Volhard, and Eric Wieschaus. These researchers used a large-scale mutant screen to uncover developmental defects in *Drosophila.* They found a complex genetic regulatory cascade that specifies the insect body plan (Fig. 2), which has since been shown to have many commonalties with molecular patterning in vertebrate embryos. The three were awarded the Nobel Prize for their efforts. However, even before the genetic dissection of segmentation, the elegant work of Klaus Sander had defined the basic mechanisms of embryo patterning that helped in the interpretation of the new genetic data. By analyzing the outcomes of embryonic manipulations of leafhopper embryos *(Eucelis),* Sander concluded that two morphogenetic gradients specify the pattern elements along the anteroposterior axis of

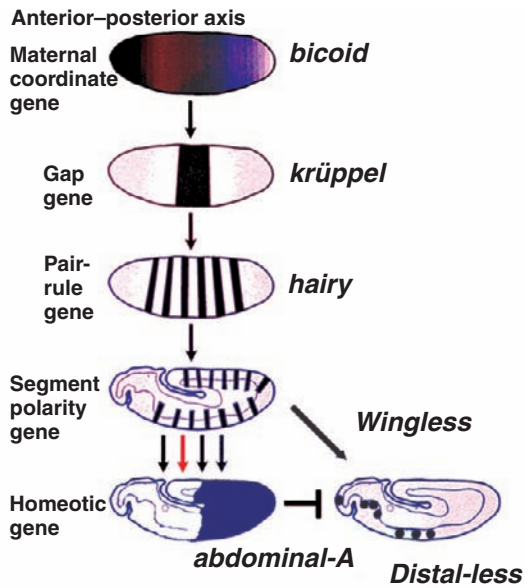

FIGURE 2 Simplified diagram of the segmentation gene cascade in *Drosophila melanogaster* and its relation to limb development. Diagrammatic representation of some of regulatory interactions between genes in the *Drosophila* segmentation cascade: a maternal coordinate gene, a representative gap gene, a pair-rule gene, the segment polarity gene, a homeotic gene, and the limb-patterning gene *Distal-less.* [Modified from Nagy, L. (1998). *Am. Zool.* **38**(6).]

the germ band: one gradient with a high point in the posterior pole and another with a high point at the anterior pole (Fig. 3). In 1976 Sander reviewed these experiments, and a large body of other experimental work on insect embryos, setting the stage for modern understanding of the molecular basis of insect development.

MOLECULAR CIRCUITS THAT REGULATE SEGMENTATION

In the cascade of gene activity that generates the segmental pattern of the embryo (Fig. 2), segmentation proceeds by a progressive refinement of positional information that will eventually specify groups of cells that form metameric units. Refinement is initiated with maternally provided proteins that form gradients from the anterior to the posterior of the egg and early cleavage stages. These gradients of maternal proteins provide the coordinates that position the front and back of the embryo; hence the genes that encode these proteins are called the maternal coordinate genes. The function of these maternal gradients is to activate the gap genes, which are so named because when their function is lacking, the segmental pattern of the embryo has large gaps in it (e.g., the three thoracic segments may be missing, or the first few abdominal segments.

The proteins encoded by the gap genes in turn activate the pair-rule genes. The pair-rules genes are expressed in every other segment and represent the first apparent metameric pattern. It was somewhat surprising that the first metameres produced by the pair-rule genes during embryogenesis do not correspond to the adult segments, but rather consist of a unit with double-segment periodicity. When pair-rule genes are absent, the larva has only half the normal number of segments. The pair-rule proteins then activate the segment polarity genes, a set of genes expressed in a segmentally reiterated manner. Finally, the homeotic genes are activated in a region-specific manner. The homeotic genes are a well-studied group of genes that are responsible for conferring segment character. They provide information on whether an individual segment will be a specific mouthpart, thoracic, or abdominal segment.

Much of what has been learned about the molecular process of segmentation is from *Drosophila;* how much is representative of a general process for all insects is not yet known. The segment polarity and homeotic genes, as well as their presumed functions, seem to be conserved in all insects examined so far; however, the activity of the maternal coordinate, gap, and pair-rule genes is more variable (Fig. 4). Because of the variation in the formation of germ anlage, it is not surprising that the earliest stages of the segmentation gene cascade established in *Drosophila* do not function in more ancestral insects. Exactly how short-germ-type embryos establish their segmental pattern remains to be discovered.

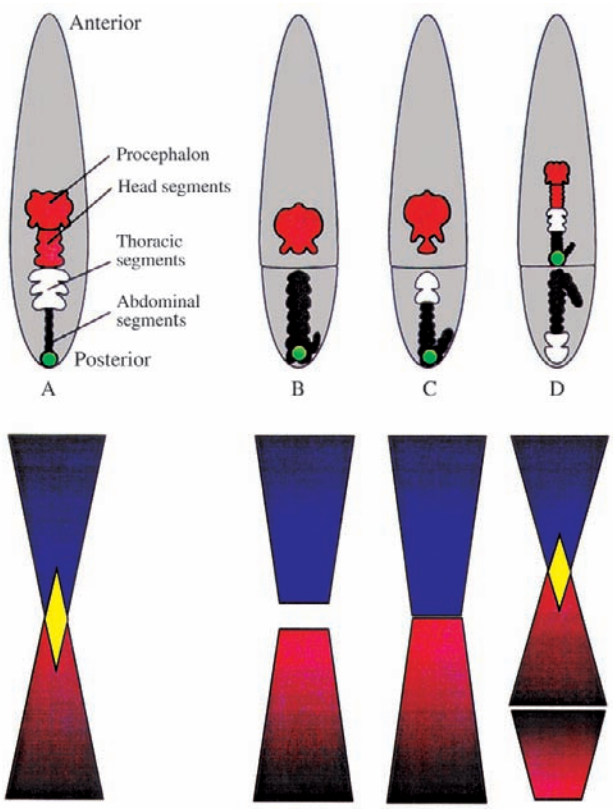

FIGURE 3 Double gradients. (A) Leafhopper embryo consists of head (red), thoracic (white), and abdominal segments (black). Posterior pole of the egg is marked by bacterial symbiont (green). (B) After early egg ligation, anterior and posterior fragments form fewer segments than in the normal embryo. (C) Late ligations result in more segments formed in anterior and posterior fragments. (D) Finally, when posterior material has been displaced anteriorly and the egg ligated just below the symbiont marker a mirror-image duplication was formed. Schematic of corresponding anterior (blue) and posterior (red) gradients and their overlap (yellow) below images illustrates possible anterior and posterior gradients and their interactions in each experimental intervention.

THE GERM BAND AND DORSAL CLOSURE

The germ band is a two-layered structure, comprising both ectoderm and mesoderm, that represents the outline of the final body plan along both axes. As the embryo grows, the germ band transforms from this essentially two-dimensional, two-layered sheet into a three-dimensional larva. From anterior to posterior, all the segments are represented. Individual segments first become visible near the anterior end, where the ectoderm differentiates into the brain and compound eyes. Protrusions develop anterior to the mouth opening that will eventually grow to form the labrum (front lip of mouthparts) and the antennae. The next segment, the intercalary segment, develops a transient limb bud, which is later retracted. This bud may be a remnant of a second pair of antennae found in this position along the anterior–posterior axis in crustaceans. Each of the first three segments behind the mouth form paired appendages that become the mouthparts: mandibles, maxillae, and labium. The next three segments

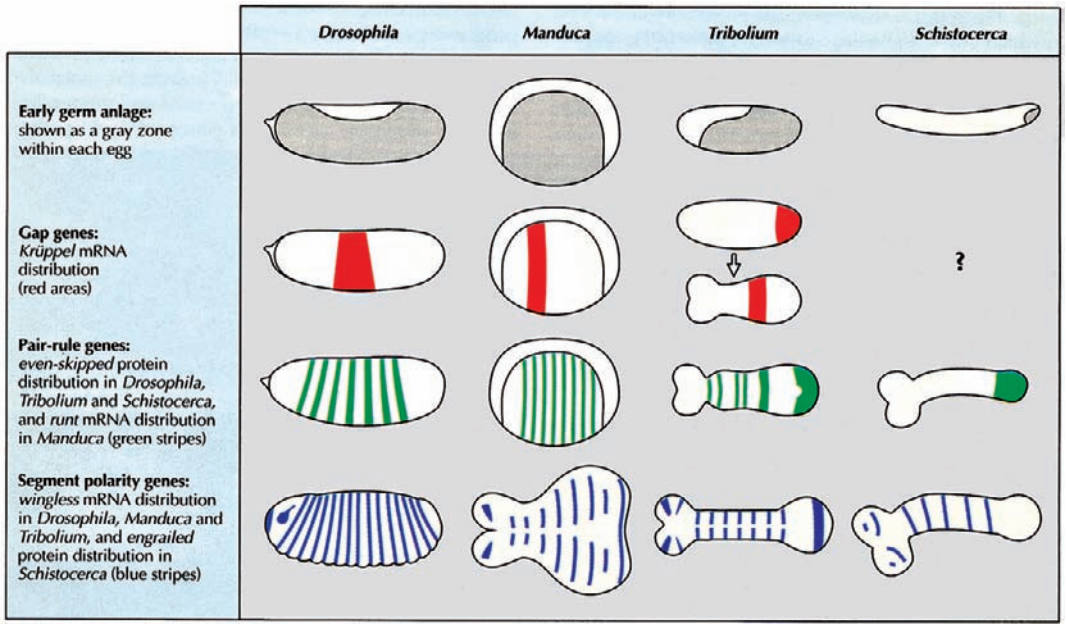

FIGURE 4 Comparative germ band and expression patterns of segmentation genes. [Modified from Nagy, L. (1994). *Curr. Biol.* **4**, 811–814.]

develop into the thorax and form appendages that become walking legs. During the remainder of embryogenesis, as organs develop and differentiate, the flanks of the germ band, both ectoderm and mesoderm, grow laterally and extend around the yolk. The two edges of the germ band meet and fuse along the dorsal midline, such that the mesodermal and ectodermal layers now enclose the yolk.

ORGANOGENESIS

When the germ band is fully segmented and gastrulation is complete, the remainder of embryogenesis involves the differentiation of the ectoderm and mesoderm into the organ systems of the larva or juvenile. The ectoderm gives rise to the bulk of the larval or adult form. Most obviously the ectoderm forms the "skin" of the larvae, marked by numerous bristles and hairs. In addition, the nervous system develops from the ventral ectoderm, and the tracheal system develops from invaginations of the lateral ectoderm. Ocelli, salivary glands, a prothoracic gland, corpora allata, molting glands, oenocytes, and silk glands also develop as ectodermal invaginations. Finally, two additional invaginations of the ectoderm occur:

1. The stomadeum occurs in a central position near the anterior of the germ band, and once invaginated, these cells proliferate in a posterior direction to form the foregut.
2. The proctodeal invagination occurs in the terminal segment, and these cells grow anteriorly to form the hindgut.

Malpighian tubules, the insect excretory organ, develop from outpocketings of the proctodeum. The invaginated mesoderm initially forms a pair of transient coelomic sacs in each segment (Fig. 1E). From these, the dorsal vessel, or heart, the internal reproductive organs, muscles, fat body, subesophageal gland, and hemocytes will form. The midgut arises from a third germ layer, the endoderm, that develops at the edge of the fore- and hindgut invaginations and eventually fuses with them to complete the gut. During the remainder of development both the mesodermal and ectodermal organ primordia all undergo differentiation into tissue-specific cell types and cell rearrangements required to form the final organ structures.

APPENDAGE DEVELOPMENT

In direct-developing hemimetabolous insects, leg and wings develop as direct outpocketings from the lateral embryonic ectoderm. Leg buds appear early, just after the completion of gastrulation, whereas wing buds appear later in development, after the lateral ectoderm has grown dorsally. In many metamorphic insects, rather than outpocketing, a cluster of cells that will form the adult leg and wing invaginates below the ectoderm. These cells become the leg and wing imaginal discs and do not undergo any further differentiation until later larval stages.

The molecular basis of positioning the limb primordia within the embryo is also well established in *Drosophila* and seems to be similar in many respects throughout both hemi- and holometabolous insects. The same information required to pattern the body axis (Fig. 2) is used to pattern the limb primordia. Every segment has the capacity to form a limb, and limbs appear at a discrete boundary formed at the intersection of the segment polarity genes, and the graded signals that are used to pattern the dorsal ventral axis of the embryo. Limb primordia are marked by the expression of the

Distal-less gene. The absence of *Distal-less* gene function results in the loss of distal limb structures up to the proximal limb segment, which is the coxa. The limbless abdomen characteristic of insects is created by a subsequent repression of *Distal-less* gene activity by several of the homeotic genes.

HATCHING

Upon completion of organ formation and histogenesis, the embryo begins to stretch and contract its newly formed muscles, and gas is secreted into the trachea. Embryogenesis is over when the maternally supplied yolk has been consumed. Hatching is achieved by any number of means, but typically, hatching is a mechanical process, in which the larva either chews its way out of the chorion, grows by imbibing air until the chorion cracks, or uses a special egg burster. There is sometimes an enzymatic digestion of the eggshell, but complicated hydrostatic mechanisms are used, as well. The hatchling emerges as a first instar (larva or nymph).

See Also the Following Articles

Eggs • Imaginal Discs • Segmentation • Vitellogenesis

Further Reading

Cohen, S., and Juergens, G. (1989). Proximal–distal pattern formation in *Drosophila:* Graded requirement for *Distal-less* gene activity during limb development. *Roux's Arch. Dev. Biol.* **198,** 157–169.

Counce, S. J., and Waddington, C. H. (1972). "Developmental Systems: Insects," Vol. 2. Academic Press, London.

Hegner, R. W. (1911). Experiments with Chrysomelid beetles. III: The effects of killing parts of the eggs of *Leptinotarsa decemlineata. Biol. Bull.* **20**, 237–251 (1911).

Kume, M., and Dan, K. (1968). "Invertebrate Embryology." Garland Publishing, New York.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila. Nature* **287,** 795–801.

Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila. Nature* **276,** 565–570.

Sander, K. (1976). Specification of the basic body pattern in insect embryogenesis. *Adv. Insect Physiol.* **12,** 125–238.

Sander, K., Gutzeit, H. O., and Jaeckle, H. (1985). Insect embryogenesis: Morphology, physiology, genetical, and molecular aspects. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology." (G. A. Kerkut and L. I. Gilbert, eds.). Pergamon Press, Oxford, U.K.

Endangered Insects

Scott Hoffman Black and Mace Vaughan

The Xerces Society

The Xerces blue butterfly, Antioch katydid, Tobias' caddisfly, Roberts's alloperlan stonefly, Colorado burrowing mayfly, and Rocky Mountain grasshopper all were driven extinct by humans, and all foreshadow the fate of the world's endangered insects. With almost 1 million described species, insects eclipse all other forms of animal life on Earth, not only in sheer numbers, diversity, and biomass, but also in their importance to functioning ecosystems. However, human-induced changes to the natural environment endanger vast numbers of these organisms, threatening them and the vital services they provide with extinction.

INSECT DIVERSITY AND IMPORTANCE

As biologist J. B. S. Haldane noted more than 60 years ago, "The creator must have an inordinate fondness for beetles." The more than 300,000 species of beetle to which Haldane referred are representative of the great diversity of insects. Measured by the number of formally described species, insects are by far the most diverse group of organisms on Earth. More than 950,000 species of insects have been described, comprising 72% of the total identified animal species on Earth.

Even more remarkable are the estimates of how many insects we have not cataloged. Most insect species that have been classified and named to date are from temperate zones, but tropical habitats harbor far more. Smithsonian Institution entomologist Terry Erwin has suggested that as many as 30 million insect species may exist based on extrapolations from the number of beetles found in particular tropical tree species. The most conservative estimates suggest that 5 to 8 million insect species have not been discovered. This number contrasts sharply with the 5,000 to 10,000 species of vertebrates that may await discovery and description around the world.

The sheer number and mass of insects reflect their enormous ecological impact. The world's ecosystems depend upon insects for pollination, decomposition, soil aeration, and nutrient and energy cycling. As Harvard biologist E. O. Wilson wrote, "So important are insects and other land dwelling arthropods, that if all were to disappear, humanity probably could not last more than a few months."

INSECT ENDANGERMENT

A report by the World Commission on Environment and Development noted, "there is a growing consensus that species are disappearing at rates never before witnessed on the planet" but that "we have no accurate figures on current rates of extinctions, as most of the species vanishing are the least documented, such as insects in tropical forests." Scientists and conservationists agree that insect species are going extinct. But how many have been lost and how many more are at risk remains unclear.

Extinct Insects

The International Union for Conservation of Nature and Natural Resources (IUCN) lists 72 insects as extinct worldwide. In the United States, the Natural Heritage Program lists 160 insect species either as presumed extinct or

as missing and possibly extinct. Many scientists believe that these numbers drastically underestimate actual insect extinction and that many hundreds, or perhaps thousands, of species have gone extinct unnoticed in North America and Europe in the past 2 centuries. The loss in tropical areas has probably been much greater.

For example, the Antioch katydid, *Neduba extincta,* from California was described in 1977 from preserved specimens collected 40 years earlier. Searches of its sand dune habitat, now largely destroyed, have proved fruitless. The Tobias' caddisfly, *Hydropsyche tobiasi,* was described in 1977 from specimens collected on the Rhine River in the 1920s. None have been seen since.

In some instances, insects that at one time were very common have disappeared. During the mid-1800s, immense swarms of the Rocky Mountain grasshopper, *Melanoplus spretus,* periodically migrated from the northern Rocky Mountains and destroyed crops throughout the western and central portions of the United States and Canada. However, in the late 1880s this species began a precipitous decline. Some believe that a natural population crash combined with habitat destruction and introduced species led the Rocky Mountain grasshopper to extinction. If a widespread species can vanish because of human activity, the fate of many endemic tropical species must hang in the balance as their only habitat is destroyed.

Endangered Insects

Based on available information we can deduce that a very large number of insects are endangered. The majority of animals on the planet are insects and, if the factors that endanger other animals also affect insects, the number of endangered insects must be very large.

According to the 2000 IUCN *Red List of Threatened Species,* 163 insects are listed as critically endangered or endangered worldwide. In 1987, West Germany classified 34% of its 10,290 insect and other invertebrate species as threatened or endangered and, in Austria, this figure was 22% of 9694 invertebrate species. More recent figures from 2000 for Great Britain show that 10.8% of its 14,634 described insect species are rare, vulnerable, or endangered. In the United States, both the U.S. Fish and Wildlife Service (USFWS) and the Natural Heritage Program track endangered species, including insects. The USFWS lists 44 insects as either endangered or threatened, whereas the Natural Heritage Program lists 165 insects as either critically imperiled or imperiled.

Are these figures on endangered insect species realistic? Because we lack an enormous amount of information on the taxonomy, life history, and distribution of insects and because endangered species documentation is biased in favor of vertebrates, we certainly are underestimating the number of at-risk insect species. To illustrate, only 7 and 4% of the endangered animal species listed by the IUCN and USFWS, respectively, are insects, yet insects make up more than 72% of global animal diversity. Of all the vertebrates described in the United States, 17.9% are listed as threatened or endangered. If we assume that insects and vertebrates face similar destructive forces at similar levels of intensity, then one should expect to find on the order of 29,000 at-risk insects in the United States alone. Although this assumption oversimplifies the situation, it shows that the 44 insects listed as endangered and threatened by USFWS are a significant underestimate. The Natural Heritage Program may be closer to the mark for select groups of insects for which we have more information. It estimates that 43% of stoneflies, 19% of tiger beetles and butterflies, and 17% of dragonflies and damselflies are critically imperiled or imperiled in the United States. In addition, according to the IUCN *Red Book of Swallowtails,* 10% of swallowtail butterflies are considered threatened. Swallowtails are the only group of insects to have been assessed worldwide.

IMPORTANCE OF ENDANGERED INSECTS

A rare and endangered species of insect is unlikely to determine the fate of a large ecological system, but as a group they may have a large effect. Ecosystem functions, such as the recycling of nutrients, often are done by specialists like the American burying beetle rather than generalists. There are innumerable specialized insects that feed on particular kinds of wood, dung, or carrion. For instance, the plates that cover the shells of tortoises are made of keratin, a protein few scavengers can digest. However, in Florida there is a moth, *Ceratophaga vicinella,* whose caterpillar appears to have specialized on a diet of dead gopher tortoise shells.

Endangered species also can play a linchpin role in small, specialized systems, such as caves, oceanic islands, or some pollinator–plant relationships. For example, many plant species rely on one or a few pollinators. Decreased abundance or loss of any of these pollinators can have dramatic consequences, especially if a plant depends on a single, obligate pollinator.

Some endangered species might provide useful products, such as new defenses against diseases and tools for studying various ecosystem or organismal processes, as well as direct material benefits. For instance, the conservation of several species of butterflies is helped by the market value of aesthetically pleasing specimens or of live specimens for butterfly houses that charge admission.

In addition to these material reasons for conserving endangered insects, we also have the responsibility of caring for the rich biological heritage we leave to future generations. At this time, we cannot begin to grasp the full value of biodiversity and, thus, it is in our best interest to be conservative.

CAUSES OF ENDANGERMENT

Insects become endangered because of the same destructive forces faced by many other animals. According to the IUCN, the leading causes of animal endangerment are habitat

destruction, displacement by introduced species, alteration of habitat by chemical pollutants (such as pesticides), hybridization with other species, and overharvesting. Many at-risk insects are threatened by more than one of these causes. For example, according to the Natural Heritage Program there are six tiger beetles and 33 butterflies that are imperiled or federally listed under the U.S. Endangered Species Act. The major threat to all six tiger beetles is habitat degradation and loss. Two of these beetles also are threatened by overcollecting. For the 33 butterflies, 97% are threatened by habitat loss, 36% by alien species, 24% by pollution, and 30% by overcollecting.

Insects as a group are not at risk because many species are generalists or widely distributed. A significant proportion of the total diversity of insects, however, is composed of species that are highly specialized or are restricted to one or a few small patches of habitat. The giant flightless darkling beetle, *Polposipus herculeanus,* for instance, lives only on dead trees on tiny Frigate Island in the Seychelles. The stonefly *Capnia lacustra* exists only in Lake Tahoe and is also the only stonefly in the world known to be fully aquatic in the adult stage. Another unusual stonefly, *Cosumnoperla hypocrema,* is known from only one intermittent spring in the Cosumnes River Basin in California.

Habitat Destruction

Agriculture, commercial development, outdoor recreation (including off-road vehicles), pollution, and water development rank as the most frequent causes of habitat degradation affecting federally listed endangered and threatened insect species in the United States (Fig. 1). Commercial and residential developments often are situated on sites that have naturally high diversity, such as along rivers or near bays and estuaries. Urban development in the southeastern United States and California has had particularly strong impacts on native insects because of the high rates of insect endemism where these cities were built. The best known case is that of San Francisco, California, which now almost entirely covers what was once one of the major coastal dune ecosystems in western North America. Three dune butterflies, which were endemic to this region, are now extinct: *Cercyonis sthenele sthenele, Glaucopsyche xerces* (Fig. 2), and *Plebeius icarioides pheres.* Three other butterflies, *Speyeria callippe callippe, Callophrys mossi bayensis,* and *Plebeius icarioides missionensis,* are now limited to the San Bruno Mountains just south of San Francisco, the last remnant of the San Francisco hills ecosystem.

Conversion of natural habitats for agriculture, particularly for planted food and fiber crops (e.g., cotton), is one of the most extensive land uses and, according to Robert Pyle (a noted lepidopterist and author), has resulted in the greatest loss of native insect populations. The most serious losses of endemic insects to agricultural conversion have taken place in the tropics, but because of the lack of knowledge of insects in these regions, it is impossible to know the extent of this destruction.

Dams and other water development are implicated in the decline of 21% of federally listed insect species. Impoundments destroy habitat for native aquatic organisms, such as stoneflies, as well as some terrestrial insects. For example, the damming of the Columbia River in Oregon and Washington resulted in the destruction of much of the sand bar habitat of the tiger beetle, *Cicindela columbica.*

Although we have no numbers, insects most likely are lost to large-scale timber management. Studies have shown that there is higher invertebrate diversity, as well as endemism, in late successional forests than in younger stands, and less than 10% of U.S. native forests remain intact. Widespread use of off-road vehicles also threatens some species. For example, vehicles have crushed the larval burrows of the tiger beetle, *Cicindela dorsalis,* along beaches to such an extent that this

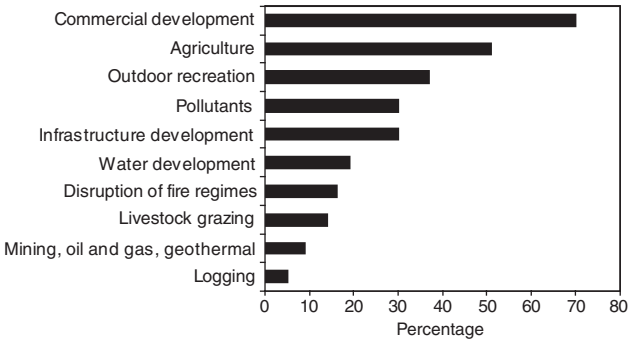

FIGURE 1 Percentages of the 43 U.S. federally listed threatened and endangered insect species affected by different causes of habitat destruction or degradation, as of December 2001. It is important to note that the habitats of most listed species are being degraded by more than one cause. (Bar graph format modified from B. A. Stein *et al.,* 2000. Data modified from D. S. Wilcove *et al.,* 1998, Quantifying threats to imperiled species in the United States. *Bioscience* **48,** 607–615.)

FIGURE 2 The Xerces blue butterfly *(G. xerces),* was one of the first butterflies in North America known to become extinct as a result of human interference. It was driven to extinction as San Francisco expanded over the butterfly's habitat. (Photograph courtesy of C. B. Barr and the Essig Museum of Entomology, University of California, Berkeley.)

once widespread, abundant species has been eliminated throughout most of its range. Wetland draining also has taken its toll. The draining of fens in England caused the extirpation of the butterfly *Lycaena dispar* in 1851 and possibly other insects as well. Capping of springs led to the loss of the fritillary butterfly, *Speyeria nokomis coerulescens,* in the U.S. portion of its range.

The biggest unknown is, of course, the loss of tropical rainforest. Tropical rain forests may hold the majority of terrestrial insect diversity and are being converted to agriculture and other uses at an alarming rate. As rainforests around the world are clear-cut, insects are bound to go with them.

Alien Species

The introduction of various exotic organisms (whether intentional or not) has affected native insects, both directly and indirectly. For example, introduced plants may out compete native plants and, thus, lead to the loss of insect host plants or habitat. Introduced plant diseases also can wreak havoc on insect populations. A classic example involves the American chestnut. Mature examples of the tree disappeared throughout its range following the accidental introduction of chestnut blight. At least five microlepidopterans, including the chestnut borer, *Synanthedon castaneae,* are believed to have gone extinct because of the loss of their host plant. Some aquatic insect species are restricted to small mountain lakes in the United States and have been impacted by introductions of nonnative fish. On the Island of Oahu, a species of *Megalagrion* damselfly is uniformly absent in stream reaches where nonnative mosquitofish in the family Poeciliidae have been introduced.

Intentional introductions of insects also many harm native insects. Over the past 50 years, nonnative insects often have been released to control nonnative pest insects. Although the damage to nontarget, native insects from these biological controls is rarely documented, some evidence is surfacing that it may be significant. For example, a parasitoid fly, *Compsilura concinnata,* that was released repeatedly in North America from 1906 to 1986 as a biological control against several pests, including the introduced gypsy moth, is implicated in the declines of four species of giant silk moths (Lepidoptera: Saturniidae) in New England. Another study in Hawaii found that 83% of parasitoids reared from native moths were former biological control agents.

Overcollecting

Although overcollecting has not been shown to harm healthy populations of insects, it may be an important threat to insect species with very small populations and is included in the list of threats to many of the federally protected insect species in the United States. The Endangered Species Act expressly forbids the collection of endangered or threatened species, and most insect conservationists feel that collecting from small populations should be done only for well-designed, hypothesis-driven, scientific studies. It is not too much to ask that scientists rise to this standard when studying populations that are at risk.

Other Potential Threats

Pesticides and other pollutants are implicated in the decline of many native bees and some aquatic insects, although the degree of impact is not conclusive. Lights along streets and highways also have been implicated in losses of nocturnal insects, particularly large moths. Finally, even though we cannot specify the exact effects of climate change at this time, it could lead to endangerment of endemic insects with specific, narrow habitat requirements. A changing climate may be especially detrimental to species that cannot disperse, like the Uncompahgre fritillary butterfly *(Boloria improba acrocnema),* which is restricted to high mountain slopes in southern Colorado.

PROTECTING AT-RISK INSECT SPECIES

Conservationists have concluded that the current, widespread destruction of the earth's biodiversity must be matched by a conservation response an order of magnitude greater than currently exists.

Protecting Habitat

Ultimately, to protect any species one must protect its habitat. Some insects need only small areas to thrive, and even backyard gardens may help some pollinator insects. Large swaths of land set aside as reserves, wilderness, national parks, and conservation easements ultimately may benefit insects and other invertebrates. Recent evidence, however, shows that some reserves, with management plans tailored to vertebrates, do little to protect insects such as butterflies.

One important caveat for setting aside land for insects is that species often have subtle habitat requirements and can be lost even from reserves because of apparently minor habitat changes. For example, larvae of the large blue butterfly *(Maculinea arion)* are obligate parasites of red ant colonies *(Myrmica sabuleti).* In 1979, this butterfly went extinct in England because habitat was not managed for these red ants. The large blue subsequently has been reintroduced successfully to appropriately managed sites in England using a subspecies from Sweden.

Federal Laws and Legislative Efforts

Federal legislation is vital to the protection of endangered insects. In the United States, the formal listing of species as threatened or endangered under federal or state endangered species legislation has been an extremely effective habitat protection tool because (1) these species are protected by law

and (2) money is allocated for recovery efforts. In addition to this protection, a listing as "sensitive" or "indicator species" under U.S. Forest Service National Forest Management Act regulations, or even a formal listing from nongovernmental organizations such as IUCN and the Natural Heritage Program, raises visibility and an awareness of these species. This increased attention may lead to the stricter legal protection of a federal listing under the U.S. Endangered Species Act.

Other countries also have legislative efforts to protect insects and other invertebrates. In 1986, the Committee of Ministers of the Council of Europe adopted a charter favoring the protection of invertebrates. This charter has raised awareness to the plight of endangered invertebrates and, in some cases, led to habitat protection. For most developing countries in the world, protective legislation for insects is either lacking or only sporadically applied. One exception is Papua New Guinea, where there is legislation, as well as a management program, that protects the rarest birdwing butterflies, allows only citizens to sell native insects, and protects some insect habitat.

Research

Before we can work to protect insects and other invertebrates we need to know, at least, what species are present, if populations are stable or declining, and the habitat needs of these populations. In the long run, more emphasis needs to be placed on invertebrate survey, systematics, taxonomy, and population ecology so that these species can be identified and cataloged and their life histories understood. Research needs to go hand in hand with conservation, for a catalog of extinct species is of little use.

Insects as Commodities

Conservation-based ranching of butterflies and other charismatic insects, like scarabs, can protect and conserve critical habitat for threatened species where the appropriate tropical forests remain intact and where live insect export is legal. The tropical forests of Central and Latin America, the Philippines, Madagascar, Kenya, Malaysian Borneo, Jamaica, and Indonesian Irian Jaya meet these criteria. These ranches not only offer protection to these charismatic insects and their habitat, but also serve as a sustainable means of economic development.

We differentiate between butterfly farming and ranching. According to the Convention on International Trade in Endangered Species (CITES) "farming" operations are essentially closed systems, no longer dependent upon regular infusions of wild stock to produce successive generations in captivity. Ranching operations, on the other hand, are open-ended and depend upon a recurrent infusion of wild stock (such as by harvesting early instar larvae in the wild and then growing them out in controlled environments). Using the CITES terminology, butterfly ranching is preferable to farming because the viability of ranching efforts depends upon the continued availability of wild habitat from which to take the needed stock. This assumes, of course, that any harvest from the wild is sufficiently controlled so as not to be excessive.

Education

To conserve insects successfully, the general public, scientists, land managers, and conservationists need to understand the extraordinary value that these organisms provide. It is unlikely that very many people will develop an affinity for these animals, but it is plausible that a more compelling depiction of the contributions insects make to human welfare and survival will improve the public's attitude toward these organisms. An ambitious public education program would enhance recognition of the positive values of invertebrates and, indeed, all biological diversity.

THE TIME IS NOW

The number of endangered insects is large and growing. The rate of destruction and degradation of natural habitats currently is so great that there are not nearly enough biologists to even catalog, much less study, the species that are suddenly on the edge of extinction. In Indonesia, approximately 1.3 million hectares of tropical forest were cut in 2001. In Argentina, 7964 metric tons of insecticides were used in 1998. In the United States, imported red fire ants have infested over 260 million acres in the southeast. These examples of threats to endangered insects continue to mount across the world. The time is now for agencies, scientists, conservationists, and land managers to promote the conservation of imperiled insects.

See Also the Following Articles

Biodiversity • Conservation • Greenhouse Gases, Global Warming, and Insects • Pollution, Insect Response to

Further Reading

Bean, J. M. (1993). Invertebrates and the Endangered Species Act. *Wings,* Summer.

Buchmann, S. L., and Nabhan, G. P. (1996). "The Forgotten Pollinators." Island Press, Washington, DC.

Collins, N. M., and Thomas, J. A. (eds.) (1991). "The Conservation of Insects and Their Habitat." Academic Press, London.

Deyrup, M. (2001). Endangered terrestrial invertebrates. *In* "Encyclopedia of Biodiversity" (S. A. Levin, ed.), Vol. 2. Academic Press, New York.

Deyrup, M., and Eisner, T. (2001). Interviews at the edge of a cliff. *Wings,* Fall.

Hoffman Black, S. C., Shepherd M., and Mackey Allen, M. (2001). Endangered invertebrates: The case for greater attention to invertebrate conservation. *Endangered Species Update* **18,** 41–49.

Kellert, S. R. (1993). Values and perceptions of invertebrates. *Conservat. Biol.* **7,** 845–855.

Opler, P. (1995). Conservation and management of butterfly diversity in North America. *In* "Ecology and Conservation of Butterflies" (A. S. Pullin, ed.). Chapman & Hall, London.
Pyle, R. M., Bentzien, M., and Opler, P. (1981). Insect conservation. *Annu. Rev. Entomol.* **26,** 233–258.
Samways, M. J. (1994). "Insect Conservation Biology." Chapman & Hall, London.
Stein, B. A., Kutner, L. S., and Adams, J. S. (eds.) (2000). "Precious Heritage: The Status of Biodiversity in the United States." Oxford University Press, Oxford.
Wilson, E. O. (1992). "The Diversity of Life." Norton, New York.

Endopterygota

Endopterygota is a division of the class Insecta in the phylum Arthropoda. The orders of insects in this division have wing rudiments that develop internally, and external wings are not evident until pupal metamorphosis occurs. There are three superorders in this division: the Neuropteroidea (which includes the orders Coleoptera, Megaloptera, Neuroptera, Raphidioptera, and Strepsiptera), the Mecopteroidea (Diptera, Lepidoptera, Mecoptera, Siphonaptera, and Trichoptera), and the Hymenopteroidea (Hymenoptera). There is no morphological resemblance between larvae and adults, and habitats and biology of adults differ greatly from larvae.

Entomological Societies

Alan I. Kaplan
East Bay Regional Park District, Berkeley, California

Entomological societies, as a category of learned societies, grew out of small, localized groups whose members lived less than a day's carriage ride from each other. In an age when specimens could not be entrusted to an irregular (or nonexistent) postal service, visiting the "cabinet" (collection) of a fellow member to see an actual specimen was necessary. Today, international societies have hundreds to thousands of members; activities range from having only a journal subscription in common to annual meetings with thousands of participants. From a largely amateur base in the 19th century, entomological societies have grown increasingly professionalized, a pattern similar to the societies covering ornithology and botany, which also had their beginnings as organized sciences with broad, nonprofessional participation.

In response to increased professionalization of entomology in the 20th century, a large number of regional and international specialized societies have arisen, to serve both professional and amateur entomologists having particular systematic or disciplinary interests. From a time when all entomologists were amateurs (mid-19th century), through a period of increased professionalization resulting in marginalization of amateurs (early to late 20th century), amateur entomologists continue their contributions to the field today.

Herbert Osborn, in his *Brief History of Entomology* published in 1952, wrote, "the origin of entomological societies is to me still a mystery." We now know why: the first entomological society in the world was founded in London, sometime between 1720 and 1742. The exact date is uncertain because the collection, books, and regalia (and presumably the minutes) of this group, the (first) Aurelian Society, were destroyed in the Great Cornhill Fire of March 25, 1748. Its meeting place, Swan Tavern on Exchange Street, was burnt to the ground; the members, then in session, barely escaped with their lives.

GENERAL FEATURES OF SOCIETIES

Entomological societies share some common features: Membership requires payment of dues to maintain the organization, there is often a "pro forma" election to membership held at a meeting of the society, and prospective members are rarely refused. For example, the only person ever turned down for membership in the New York Entomological Society was "the author of a new version of the theory of spontaneous generation!" Honorary membership (usually limited to a small number) is offered to accomplished and distinguished entomologists in the home country of the society or from other countries. Distinguished Regular members may be elevated to Fellow status. There is often a category of nonlocal membership, usually referred to as "corresponding." Regular meetings are held, at least annually, often more frequently, with guest speakers and the opportunities for members to provide a greater number of shorter presentations. There are constitutions and by-laws, with officers who preside over business meetings. Field trips ("field days") to collect insects were a major feature of 19th and early 20th century society meetings, and annual meeting circulars and programs will suggest collecting opportunities near to meeting sites. Some societies have a tradition of insect protection: As early as 1896, the Royal Entomological Society (London) (RES) had a committee to look into protecting British insects from extinction. In 1988, the RES became the first entomological society to join the International Union of Conservation of Nature. But as far back as the second International Congress of Entomology, held at Oxford in 1912, N. C. Rothschild spoke on steps taken to protect insects in Great Britain. The British Entomology and Natural History Society formed several Conservation Working Groups in 1994 to bring the expertise of its members to bear on matters relating to conservation of the invertebrate fauna of the United Kingdom and to express the field naturalists' views of which species deserve special attention.

The publications of these societies—as proceedings, journals, memoirs, annals, bulletins, and newsletters—have been the main vehicle for dissemination of scientific information and more personal information about the work and lives of entomologists since the founding of the early societies. For example, many societies begin a publishing program the same year or soon after their founding (the French began *Annales* in 1832, the year the Société Entomologique de France was founded; the Royal Entomological Society (London) began its *Transactions* in 1834, a year after its founding). Publications have served as a medium of exchange with other societies in order to build up another feature, that of society library. The American Entomological Society library, with over 15,000 volumes, has been incorporated since 1947 into the Academy of Natural Sciences (Philadelphia, PA) library; the Pacific Coast Entomological Society does not maintain a separate library, but journals received in exchange for its publication, the *Pan-Pacific Entomologist,* and books received for review therein are deposited in the library of its host institution, the California Academy of Sciences.

Several societies (for example, the Amateur Entomologists' Society, Orthopterists' Society, Entomological Society of America) have produced a series of handbooks and guides for identifying insects and have a regular publishing program outside of the usual journal- and memoir-type series. The Brooklyn Entomological Society (BES) took as a goal the publication of alphabetical lists of scientific terms used in technical descriptions in entomology. The first entomological vocabulary published in North America was 800 terms and definitions, in Vol. 6 of the *Bulletin* of the BES in 1886. This was followed in 1906 by the BES-sponsored *Glossary, an Explanation of Terms Used in Entomology* by J. B. Smith (4000 entries). The BES published J. A. Torre-Bueno's *A Glossary of Entomology* in 1937 (10,000 terms, 12,000 definitions), and a supplement to it in 1960 by G. S. Tulloch added 500 new terms and revised 160 others. The Entomological Society of America publishes the only society-sponsored list of insect common names; in other countries this is usually a function of the department or ministry of agriculture.

Insect collections of societies and their members have become important components of the holdings of large institutions: For example, the Academy of Natural Sciences in Philadelphia has the collection of the American Entomological Society; the New York Entomological Society collection has been incorporated into those of the American Museum of Natural History. Members of the Pacific Coast Entomological Society often deposit type specimens of species described in the Society's journal (*Pan-Pacific Entomologist*) with the collection of the California Academy of Sciences. Society collections have at times been controversial: The (third) Aurelian Society in England, founded in 1801, dissolved 5 years later because of the odious requirement that members donate their best specimens to a central society collection. Disagreements over the deposition and loan of specimens of the Entomological Society of Philadelphia collection led to its expulsion in 1862 of its first president, the eminent coleopterist John L. LeConte.

Another feature of some societies is a youth program: Membership is offered to young people at a discounted rate, special publications are aimed at them, occasional exhibits are developed to tour schools or be displayed at annual meetings, and field days featuring insect collecting trips are planned. The New York Entomological Society formed a Junior Division in 1958. The Entomological Society of America has had a Youth Membership category since 1989. The Young Entomologists' Society (U.S.A.) traces its origin to the Teen International Entomology Group, founded in 1965 by a teenager as a worldwide correspondence club to exchange letters and specimens with like-minded teens around the world.

These youth programs may take time to develop. For example, the Royal Entomological Society (London) Youth Development Scheme of 1990 had hopes of local and regional participation by its Fellows, which did not materialize, and the program failed a year later. But from it came the Bug Club, now a national organization in Great Britain.

ORIGINS

The earliest scientific societies were founded in Europe in the mid-16th and early 17th centuries, but the first entomological societies came about in England in the mid-18th century for the purpose of sharing knowledge of the Lepidoptera. These were the Society of Aurelians (also called the [first] Aurelian Society, formed sometime between 1720 and 1742) and its successor, the (second) Aurelian Society, formed in 1762. "Aurelia" is a classical name for the chrysalis of a butterfly; an aurelian is a butterfly collector. The (first) Aurelian Society was finished by the Great Cornhill Fire of 1748; the second ceased in 1767 because of personality clashes among members; a third Aurelian Society, founded in 1801, disappeared by 1806.

The oldest entomological society still in existence—the Entomological Club of London, founded in 1826—has had only eight members at a time since its inception and meets one evening each month to dine at members' homes or other places. It also hosts the annual Verrall Supper for entomologists, a tradition since 1887. The oldest existing national entomological societies are Société Entomologique de France (1832), the Royal Entomological Society (London) (1833), and the Nederlandsche Entomologische Vereeniging (1845). See Table I for a list of societies that are 100 years old or older and are still in existence as of 2001.

NORTH AMERICAN SOCIETIES

The first North American entomological society was the (first) Entomological Society of Pennsylvania which, from 1842 to 1853, took as its only mission the taxonomic description of American insect species. It did not try to spread knowledge of insects to the general public, nor to encourage the study of economic (applied) entomology, nor to inves-

TABLE I Entomological Societies Founded in 1901 or Earlier, Still in Existence in 2001

Entomological Club (London), 1826
Société Entomologique de France, 1832
Royal Entomological Society (London), 1833
Nederlandsche Entomologische Vereeniging, 1845
Société Royale d'Entomologie de Belgique, 1855
American Entomological Society, 1859 (as Entomological Society of Philadelphia)
Russian Entomological Society, 1859
Entomological Society of Canada, 1863
Entomologisk Förening (Copenhagen), 1868
Società Entomologia Italiana, 1868
British Entomology and Natural History Society, 1872 (as South London Entomological Society)
Cambridge Entomological Club, 1874
Lancashire and Cheshire Entomological Society, 1877
Entomologiska Foreningen I Stockholm, 1879
Internationaler Entomologischer Verein, 1884
Entomological Society of Washington (Washington DC), 1884
Entomological Society of America, 1889 (as Association of Official Economic Entomologists)
New York Entomological Society, 1892
Union des Entomologistes Belges, 1896
Entomologischer Verein "Apollo" e.V., 1897
Association de Coleopteristes de la Region Parisienne, 1901
Pacific Coast Entomological Society, 1901 (as California Entomological Club)

tigate insect natural history. Although there were between 40 and 70 agricultural societies in the United States at this time (chiefly with the aim of county fair exhibitions), the first U.S. society's members feared that an association with this applied sphere of entomology would weaken the efforts of American entomologists to gain the respect of European entomologists. The Entomological Society of Pennsylvania had as its main project the publication of a catalog (a list of species with nomenclatural data, such as author names, dates of publication, synonyms, and taxonomic references) of American Coleoptera. It was defunct by 1853, when the Smithsonian Institution published the founding member and only president F. E. Melsheimer's *Catalogue of the Described Coleoptera of the United States.*

The oldest North American entomological society still in existence is the American Entomological Society, founded in 1859 as the Entomological Society of Philadelphia and renamed in 1867. For several years (1865–1867), this society published the first journal to be devoted to economic entomology, *The Practical Entomologist.* Its *Transactions* have been published since 1867, and it has published *Entomological News* since 1890.

The entomological societies of Canada began with the (first) Entomological Society of Canada, formed in Toronto in 1863 (an organizing meeting had been held the year before). Its journal, *The Canadian Entomologist,* has been published continuously since 1868. When support was obtained for this publication from the Council of Agriculture and Arts Association of Ontario, the society's name was changed to Entomological Society of Ontario in 1871 to reflect this support, but it still served as something of a national organization. Autonomous regional societies grew up across Canada (for example, in British Columbia in 1902, in Nova Scotia in 1914, and in Manitoba in 1945). A (second) Entomological Society of Canada emerged in 1950 to link the provincial societies into a truly national organization.

The current Entomological Society of America (ESA) was formed in 1953 from the union of the American Association of Economic Entomologists (AAEE) and the (first) Entomological Society of America. The AAEE was formed in 1889 (originally for state and federal entomologists, its first title was Association of Official Economic Entomologists; "Official" was soon dropped and "American" added in 1909). It grew out of the Entomological Club, a subsection of the Natural History Section of the American Association for the Advancement of Science (AAAS), which itself originated at an AAAS meeting in Hartford, Connecticut, in 1874, the year when entomologist John L. LeConte was AAAS president.

Dissatisfaction with the applied emphasis of AAEE and need for a societal home for academic and noneconomic entomologists led to the formation of the (first) Entomological Society of America in 1906, with Professor J. H. Comstock of Cornell University as its president. In 1908, both societies began publishing journals that are still printed today: the *Annals of the Entomological Society of America* and the *Journal of Economic Entomology.* In addition, the merged ESA produces other journals, including (since 1972) *Environmental Entomology* and, since 1986, has owned and published the *Journal of Medical Entomology* (originally published by the Bishop Museum of Hawaii).

Throughout the period 1906 to 1953, the two U.S. national entomological societies often held joint annual meetings, so members of both (and there was considerable overlap in membership) could participate in each meeting. By the time of the 1953 merger, membership in the AAEE was triple that of the ESA, reflecting the expansion of the applied entomology field in the age of modern insecticides. The strong regional (branch) divisions of the reorganized ESA closely follow the premerger AAEE structure, as do most of the subject sections within the organization (for example, sections for Regulatory and Extension Entomology and Crop Protection Entomology). Today's ESA sponsors a unique program of board certification for professional entomologists. In the past, the ESA had been the institutional home of the American Registry of Professional Entomologists (about 15% of ESA members in 1989 were registered with ARPE). A code of ethics and education and experience requirements, plus testing and continuing education, are elements in the process of "professionalization" that have emerged in only a few scientific fields (medicine being the best example), but these are hallmarks of a technology-oriented profession (civil, structural, and geologic engineering, for example). Entomology's unique status as a technology *and* a science emerged in the late 20th century, resulting in some confusion over status and prestige, which a process of professional certification clarified for some ESA members.

SOCIETAL GROWTH

In 1956, Curtis Sabrosky published a near-exhaustive list of entomological societies that had existed and/or were still in existence. By 1956, at least 70 entomological societies had begun, blossomed (or not), and then faded away. That year, there were 96 active regional or national entomological societies (not counting those devoted to applied aspects such as apiculture or pest control). Only 10 specialty societies, devoted to a taxonomic group or some other special, non-applied aspect, were listed: 3 devoted to Coleoptera (in Austria, the United States, and Japan), 4 for Lepidoptera (3 in Japan and 1 in the United States), and 1 international society for the study of social insects. The post-1956 period has seen a huge development of specialty societies. Today, there are at least 92 specialty societies (including 1 for conservation of invertebrates in general and 1 devoted to young entomologists). There has been increased worldwide interest in Lepidoptera; at least 22 new societies formed in the period from 1970 to 1999. Although there are specialty societies devoted to at least nine orders of insects, the majority are devoted to just three: Coleoptera (12), Lepidoptera (39), and Odonata (12). In addition, regular international symposia are held on other, smaller orders (for example, Trichoptera and Ephemeroptera). Just as improved optics (binoculars, spotting scopes, cameras) and better field guides stimulated the mid-20th century interest in bird-watching, similar improvements in entomological materials (for example, a recent series of books on identifying butterflies and dragonflies with binoculars) have made Lepidoptera and, more recently, Odonata popular subjects of observation and interest to larger numbers of people, who in turn have formed many new entomological societies devoted to their insect passions.

ROLE OF AMATEUR ENTOMOLOGISTS

Professionalization and an improved image as scientists has been an issue for entomologists since the 19th century. From their origin in amateur lepidopterist clubs and local societies of collectors interested in the taxonomic position of their specimens and little else, entomological societies grew into associations of applied scientists who recognized the contribution of a client base of agriculturalists but did not make a place for them or for hobbyist entomologists (amateurs) in their national organizations. Once applied entomologists in the United States organized into the AAEE, they neither encouraged nor discouraged nonprofessional participation, but instead created two classes of membership. One was a professional category, which required educational qualifications and vocational activity; the other was an "associate" category, which was a second-class membership for amateurs and others with inadequate qualifications.

Yet, amateur entomologists, who had been the founders of the field and its earliest supporters as collectors and benefactors, continued to make contributions. Early societies at times depended on wealthy amateurs for support. The best example is the support given to the American Entomological Society (AES) by one of its founders, Thomas B. Wilson, an executive of the Pennsylvania Railroad. He paid for the AES's building and was patron of its library and collection. Wilson provided a sinecure for E. T. Cresson, Sr. (one of North America's greatest hymenopterists), as his private secretary, which in reality supported Cresson as curator of the AES collection and its corresponding secretary for many years. When Wilson died, Cresson worked for an insurance company for the next 40 years (1869–1910). "The Wilson Fund" was still supporting AES publications in 1984, almost 120 years after their provider's death.

The first national Canadian entomological society was formed in 1863 by a 25-year-old divinity student, Charles J. S. Bethune, and a 28-year-old pharmacist, William Saunders. They founded its journal, *The Canadian Entomologist,* in 1868 and were the sole contributors to its first two numbers. Each went on to distinguished careers in Canadian entomology (Bethune as Professor of Entomology and Zoology at Ontario Agricultural College, Saunders as the first Director of Experimental Farms [agricultural experiment stations] for the Dominion of Canada) but they both had begun the Entomological Society of Canada as amateurs.

Amateur entomologists were always welcome to publish in the *Journal of the New York Entomological Society.* Annie T. Slosson was a NYES founder (in 1892) and the largest financial supporter of its journal—her donated specimens of Lepidoptera raised the most money at the Society's auctions. She was a well-known collector and contributed many journal articles, though she did not publish new taxonomic names for the species she gathered; she preferred to send them to specialists to describe.

The Cambridge Entomological Club was saved from extinction by an infusion of amateur members. The Club was founded at Harvard in 1874 with 12 members. It reached 48 members by the time of incorporation in 1877, but had declined in 1902 to just 7 (of whom only 3 or 4 attended meetings at any one time). The Club combined with a local amateur society, The Harris Club, with its 38 members, in 1903, and active amateur members have been an important component of the Cambridge Entomological Club ever since. The participation by amateurs was encouraged in these early days by ant specialist, Professor William M. Wheeler, who supported amateur naturalists and said, "We have all known amateurs who could make an enthusiastic naturalist out of an indifferent lad in the course of an afternoon's rambling, and, alas, professors who could destroy a dozen budding naturalists in the course of an hour's lecture."

In the first volume (1908) of *Annals of the Entomological Society of America,* the Canadian entomologist H. H. Lyman, himself an amateur, urged the society "to secure the support and cooperation of the great body of amateur entomologists." This has not been accomplished. Almost 80 years later, in 1986, a

survey of the ESA's 9111 members found only 31 (0.5% of the 5505 respondents) who described themselves as amateurs. But amateur entomologists have found welcome and a home in regional and local societies devoted to taxonomic specialties (Coleoptera and Lepidoptera predominately, but Odonata have become popular) or geographically restricted. Since 1939, the Amateur Entomologists' Society (United Kingdom) has been a flagship of the great amateur enterprise, publishing its bulletin and a large number of identification guides and handbooks.

A survey of adult amateur entomologists in 1987 by Janice Matthews found that they often suffer from being stereotyped by professional entomologists as less qualified or educated and get a cool (or even hostile) reception from professionals. Amateurs actually produce the great bulk of educational outreach on entomological topics (for example, programs for school children, other amateur naturalists, and the public at large). Adult amateur entomologists' professional lives align very closely with those of professional entomologists: Amateur entomologists are doing science and math in their daily work; they are in education; they are in service occupations (by comparison, pest control work is also a service occupation). The science background of amateur entomologists can be as strong as that of professionals, but is often in a related field. None of the respondents to Matthews' study reported that a professional entomologist influenced their childhood interest in insects; the failure of the Youth Development Scheme of the Royal Entomological Society (London) in 1990 was attributed to just this kind of lack of interest on the part of its members toward young entomologists.

That amateur entomologists have made, and continue to make, great contributions to entomology is unquestioned. From the great coleopterist P. F. M. A. DeJean (Napoleon's general and aide-de-camp at Waterloo), to 19th century lepidopterist William H. Edwards (a lawyer and coal company president), to civil engineer Richard H. Stretch (who first warned of the economic dangers of cottony cushion scale in California in 1872), through a long list of physician–entomologists (for example, H. Bernard Kettlewell, who was a general practitioner while pursuing his studies of melanism in Lepidoptera), to the great student of leaf-mining flies, Kenneth Spencer (an electronics executive, he published 74 papers before retiring in 1969 and then published 45 more papers in the next 20 years), to the Parisian taxi driver Pierre Morvan (honored with the Rolex Enterprise Award in 1987 for his biogeographic study of Asian ground beetles, he is a self-taught entomologist and author of over 50 scientific publications), entomology advances through the efforts of its many amateur practitioners.

See Also the Following Article

History of Entomology

Further Reading

Allen, D. E. (1976). "The Naturalist in Britain: A Social History." Princeton University Press, Princeton, N.J.

Connor, J. T. H. (1982). Of butterfly nets and beetle bottles: The Entomological Society of Canada, 1863–1960. *HSTC Bull.* **6,** 151–171.

Mallis, A. (1971). "American Entomologists." Rutgers University Press, New Brunswick, NJ.

Matthews, J. R. (1988). Adult amateur experiences in entomology: Breaking the stereotypes. *Bull. Entomol. Soc. Am.* **34(4),** 157–161.

Osborn, H. (1952). "Brief History of Entomology." Spahr and Glenn, Columbus, OH.

Sabrosky, C. W. (1956). Entomological societies. *Bull. Entomol. Soc. Am.* **2(1),** 1–22.

Salmon, M. A. (2000). "The Aurelian Legacy: British Butterflies and Their Collectors." University of California Press, Berkeley.

Scientific Reference Resources (2001). http://www.sciref.org.

Smith, E. H. (1989). The Entomological Society of America: The first hundred years, 1889–1989. *Bull. Entomol. Soc. Am.* **35(3),** 10–32.

Sorensen, W. C. (1995). "Brethren of the Net: American Entomology, 1840–1880." University of Alabama Press, Tuscaloosa.

Ephemeroptera

(Mayflies)

John E. Brittain

Natural History Museums and Botanical Garden, University of Oslo

Michel Sartori

Museum of Zoology, Lausanne

Mayflies (order Ephemeroptera) date from Carboniferous and Permian times and represent the oldest order of the existing winged insects. They are unique among the insects in having two winged adult stages, the subimago and imago (Fig. 1). Adult mayflies do not feed; instead, they rely on reserves built up during their nymphal life. As adults they generally live from 1 to 2 h to a few days, and mayflies spend most of their life in the aquatic environment, either as eggs or nymphs. The nymphal life span in mayflies varies from 3 to 4 weeks to more than 2 years. The length of egg development varies from ovoviviparity (i.e., the release of live offspring) to a period of up to 10 to 11 months in some arctic/alpine species.

Because of their winged adult stage and a propensity for drift (i.e., downstream movements) as nymphs, mayflies are often among the first macroinvertebrates to colonize virgin habitats. However, over longer distances their dispersal capacity is limited, owing to their fragility and short adult life. Mayflies are found in almost all types of freshwater habitat throughout the world, although in the Arctic and in mountain areas above the tree line there are few species. Mayfly faunas on oceanic islands and isolated mountain areas have few species, and they are usually restricted to the Baetidae and/or Caenidae. Their greatest diversity is in lotic habitats in temperate and tropical regions, where they are an important link in the food chain, from primary production by algae and plants to secondary consumers such as fish. Mayflies are used extensively as indicators of pollution and environmental change.

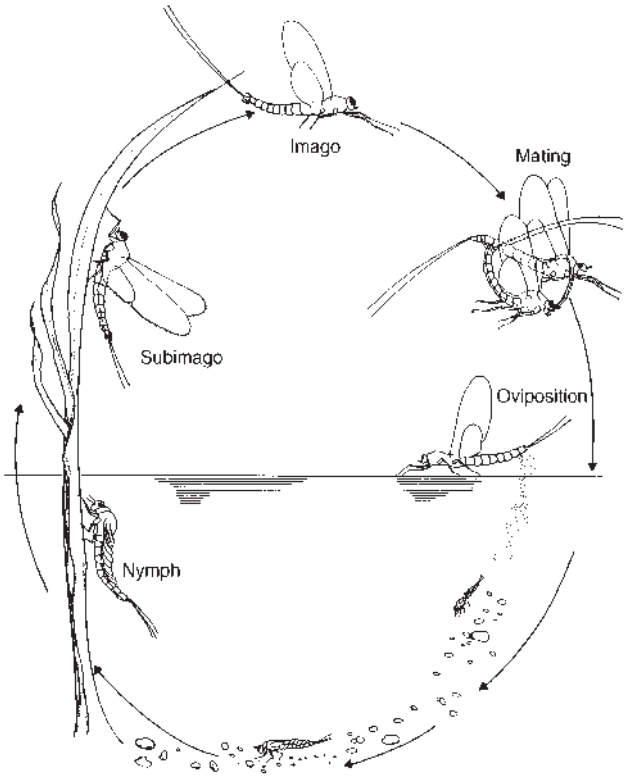

FIGURE 1 Mayfly life cycle showing the alternation between the aquatic and terrestrial environments. Mayflies are unique in having two winged stages, the subimago and imago. The adult life is very short and most of the time is spent in the aquatic environment.

ORIGINS AND EVOLUTION

Ephemeroptera are among the oldest known winged insects still extant. Carboniferous fossils have been ascribed to mayfly precursors or even mayflies. Permian data confirm that the order was already present at the end of the Paleozoic. Ephemeroptera reached their highest diversity during the Mesozoic, mainly in the Jurassic and Cretaceous. All these species belong to extinct families. The Tertiary fauna, as documented by for instance Baltic amber, is undeniably modern, with both the extinct and living genera of modern families.

The relationship of Ephemeroptera with other modern winged insects is still a subject of debate. Together with the Odonata, mayflies were traditionally placed in the Paleoptera, which was considered the sister group of all other extant primarily winged orders. More recently, it was suggested that Ephemeroptera per se are the sister group of Odonata + Neoptera. This idea is based on a number of features unique to mayflies, such as the presence of a subimaginal stage, the nonfunctionality of the adult mouthparts, and the presence of only one axillary plate in the wing articulation. This hypothesis is also supported by anatomical data: female mayflies exhibit telotrophic meroistic ovaries instead of panoistic ones as found in Odonata.

CLASSIFICATION AND PHYLOGENY

The Ephemeroptera are numerically a small order of insects, with about 3000 described species within more than 375 genera and 37 families (Table I). About 350 species occur in Europe, and 670 in North America. During the 1990s, partly as a result of the discovery of new taxa, especially in tropical areas, where the mayfly fauna is still poorly known, 10 new families, 75 genera, and more than 500 species were added. The expansion of the order is also the result of several phylogenetic analyses that led to a narrower concept of supraspecific taxa. As a consequence of these important changes, there is no real consensus about the higher classification of Ephemeroptera (superfamilies, suborders, or infraorders). Based on the structure of the nymphal wing pads, mayflies were traditionally divided into two suborders, Pannota (with fused wing pads) and Schistonota (with free wing pads). That the latter suborder is paraphyletic is now well documented, but there is no agreement about the composition and even the names of these higher taxa.

ADULTS

The adult mayfly has two main functions, mating and oviposition, which produce a general uniformity in structure. The prominent turbinate eyes of males, especially well-developed in the Baetidae and some Leptophlebiidae, provide both high acuity and good sensitivity. This enables them to detect and capture single females in a swarm at low light intensities.

The forelegs of most mayflies also show sexual differences; those of the male are unusually long for grasping and holding the female during mating. In the Polymitarcyidae, the middle and hind legs of the male and all the legs of the female are reduced, and in *Dolania* (Behningiidae) all the legs of both sexes are reduced. In *Dolania* and several members of the Polymitarcyidae and Palingeniidae, the females remain in the subimaginal stage. The reason for two winged stages has provoked much discussion. It has been suggested that this primitive trait is maintained because there has not been the selective pressure on the short-lived stages to produce just a single molt. Another explanation is that two molts are necessary to complete the elongation of the caudal filaments and forelegs of the adults. Most mayflies have two pairs of wings, but in the Caenidae, Tricorythidae, Baetidae, and some Leptophlebiidae, the hind wings are reduced or even absent.

Fecundity

Spermatogenesis and oogenesis are generally completed in the final nymphal instar, and the eggs and sperm are physiologically mature in the subimago. Most species produce 500 to 3000 eggs, but values range from less than 100 in *Dolania* to 12,000 in *Palingenia,* and the fecundity values recorded for the females of the larger species of mayfly are higher than in most other insect groups except the social

TABLE I Overview of the Mayfly Families and the Approximate Number of Genera and Species

Family	Genera	Species	Biogeography
Acanthametropodidae	2	5	Asia and North America
Ameletidae	2	45	Asia, Europe, and North America
Ameletopsidae	4	10	Australia, New Zealand, and South America
Ametropodidae	1	5	Asia, Europe, and North America
Arthropleidae	1	5	Europe and North America
Baetidae	95	700	Worldwide
Baetiscidae	1	12	North America
Behningiidae	3	5	Asia, Europe, and North America
Caenidae	13	100	Worldwide
Coloburiscidae	3	5	Australia, New Zealand, and South America
Coryphoridae	1	1	South America
Dipteromimidae	1	1	Japan
Ephemerellidae [a]	16	300	Worldwide except Australia and New Zealand
Ephemeridae [b]	8	30	Worldwide except Australia
Ephemerythidae	1	5	Africa
Euthyplociidae	5	15	Asia, Madagascar, and South America
Heptageniidae	28	500	Africa, Asia, Europe, and North America
Isonychiidae	1	20	Asia, Europe, North and South America
Leptohyphidae	7	120	North and South America
Leptophlebiidae	120	900	Worldwide
Metretopodidae	2	10	Asia, Europe, and North America
Neoephemeridae	3	10	Asia, Europe, and North America
Nesameletidae	3	5	Australia, New Zealand, and South America
Oligoneuriidae	11	45	Worldwide except Australia and New Zealand
Oniscigastridae	3	10	Australia, New Zealand, and South America
Palingeniidae	7	30	Asia, Europe, and Madagascar
Polymitarcyidae [c]	7	65	Worldwide except Australia and New Zealand
Potamanthidae	3	25	Africa, Asia, Europe, and North America
Prosopistomatidae	1	15	Africa, Asia, Australia, Europe, and Madagascar
Pseudironidae	1	1	North America
Rallidentidae	1	1	New Zealand
Siphlaenigmatidae	1	1	New Zealand
Siphlonuridae	4	30	Asia, Europe, and North America
Teloganellidae	1	1	Asia
Teloganodidae	7	15	Africa, Asia, and Madagascar
Tricorythidae [d]	6	30	Africa, Asia, and Madagascar
Vietnamellidae [e]	2	5	Asia and Australia
Total	**376**	**3083**	

[a]Including Melanamerellinae. [b]Including Pentageniinae and Ichthybotinae. [c]Including Exeuthyplociinae. [d]Including Dicercomyzinae and Machadorythinae. [e]Austremerellidae.

Compiled with the assistance of Jean-Luc Gattolliat (Lausanne) and Jan Peters and Michael D. Hubbard (Tallahassee).

Hymenoptera. In species with a long emergence period or with a bivoltine life cycle (having two summer emergence periods), early emerging females are larger and therefore more fecund than those emerging later.

Mating and Swarming

Swarming in adults is a male activity, apart from the Caenidae and Tricorythidae, where both males and females may participate. The females fly into these swarms, and mating occurs almost immediately and usually in flight. Swarming may take place over the water itself, over the shore area, or even away from the water. Most swarms are positioned according to terrain markers such as areas of vegetation, the shoreline, and trees. The time of swarming varies considerably, although dusk is the most common time of day in temperate regions. Parthenogenesis has been reported in about 50 mayfly species, although it is not obligatory as a rule.

Oviposition

The majority of mayflies, including most Ephemeridae, Heptageniidae, and Leptophlebiidae, oviposit by descending to the water and releasing a few eggs at a time by dipping their abdomen into the water. Species of *Ephemerella, Siphlonurus,* and *Centroptilum,* however, release all their eggs in a single batch that separates immediately on contact with water. In *Habroleptoides* and some Heptageniidae the female rests on a stone above the water, and dips her abdomen into the water to lay the eggs. This is taken a stage further in several species of *Baetis* in which the female actually goes underwater and lays her eggs on suitable substrate, often under stones.

EGGS

Mayfly eggs have a variety of attachment structures that enable them to adhere to submerged objects or to the substrate. Differences in egg morphology have enabled the construction of identification keys, purely on the basis of eggs. This has provided a useful complement, not only to studies of phylogeny, but also to taxonomy, since identification of female adults by means of external characters is often difficult.

Development

Most nymphs hatch at temperatures in the range of 3 to 21°C. However, in the North American *Hexagenia rigida,* the nymphs hatch successfully between 12 and 32°C and even at 36°C if incubation is started at lower temperatures. In *Tricorythodes minutus,* nymphs hatch between 7.5 and 23°C, but mortality is least at 23°C. Hatching success is variable, ranging from over 90% in several *Baetis* and *Hexagenia species* to less than 50% in the Heptageniidae studied. Excluding the few ovoviviparous species, the total length of the egg development period varies from a week in *H. rigida* to almost a year in *Parameletus columbiae.* Temperature is the major factor determining the length of the period of egg development in mayflies. There is no indication that photoperiod influences egg development time. Ovoviviparity is rare in the mayflies and is restricted to the Baetidae. In North America, a number of species in the genus *Callibaetis* are ovoviviparous.

NYMPHS

In contrast to the adults, mayfly nymphs show considerable diversity in habit and appearance. Differences do not always follow taxonomic lines, and convergent and parallel evolution seems to be common (Fig. 2).

Growth and Development

Mayflies have a large number of postembryonic molts. Estimates of the number of instars vary between 10 and 50; most are in the range 15 to 25. The number of instars for a particular species does not seem to be constant, but probably varies within certain limits. Environmental conditions, such as food quality and temperature, may affect instar number. Because of its simplicity, by far the most common measure of development and growth in mayflies has been body length, although head width and other body dimensions also have been used. However, growth of the various body parts is not always isometric. Many authors have also used body weight, and the length–weight relationship is usually well expressed by a power function.

Nymphal growth rates are influenced by several environmental factors, although the major growth regulator is mean temperature, the scale of diurnal fluctuations, or total degrees-days. Other factors, such as food and current velocity, may exert a modifying influence on growth rates. No true diapausing nymphal stage has been reported in the Ephemeroptera, although growth rates often are very low during the winter.

Respiration

The gills of mayflies are very diverse in form, ranging from a single plate in *Ameletus* to fibrillar tufts in *Hexagenia.* Respiratory tufts are sometimes developed on other parts of the body besides the abdomen, such as those at the base of the coxa in *Isonychia* and *Dactylobaetis.* In several families the second abdominal gill has developed into an operculate (lidlike) gill cover for the remaining gills, and in certain Heptageniidae the gills are markedly expanded so that they together form an adhesion disc. In many of the Siphlonuridae, the gills are used as swimming paddles, which has been put forward as their original function. In respiring, the gills may function either as respiratory organs or as ventilatory organs for other respiratory exchange surfaces.

High rates of oxygen consumption are often reported in association with emergence and gonad maturation. High water temperatures at that time may mean that low oxygen concentrations can be critical. Many burrowing Ephemeridae and pond-dwelling Baetidae are able to survive moderately low oxygen concentrations, especially for short periods. However, so far only one species, the European *Cloeon dipterum,* has been shown to survive long-term anoxia.

Population Movements

During the final stages of nymphal life there is a movement to and a concentration in the shallower areas of lakes and rivers. In running waters, springtime mass movements of mayfly nymphs along the banks of the main river and into slower flowing tributary streams or into areas flooded by spring snowmelt have been observed. In running water, mayfly nymphs may move down into the substratum in response to spates or as part of a daily rhythm. Generally, however, mayflies do not extend far down into the substratum (i.e., the hyporheic zone).

Mayflies, especially Baetidae, are a major component of invertebrate drift in running waters. Their drift shows a strong diel periodicity, with a peak during the hours of darkness. Drift rates are not constant for a particular species, and the larger size classes are usually more in evidence. Other factors that have been shown to influence mayfly drift include changes in current velocity and discharge, increased sediment loading, temperature changes, oxygen conditions, density, food availability, and predators.

EMERGENCE

Emergence, the transition from the aquatic nymph to the terrestrial subimago, is a critical period for mayflies. Their movement up to the water surface makes them especially vulnerable to aquatic and aerial predators. Shedding of the nymphal skin usually occurs at the water surface on some object, such as a stone or macrophyte stem, or in midwater.

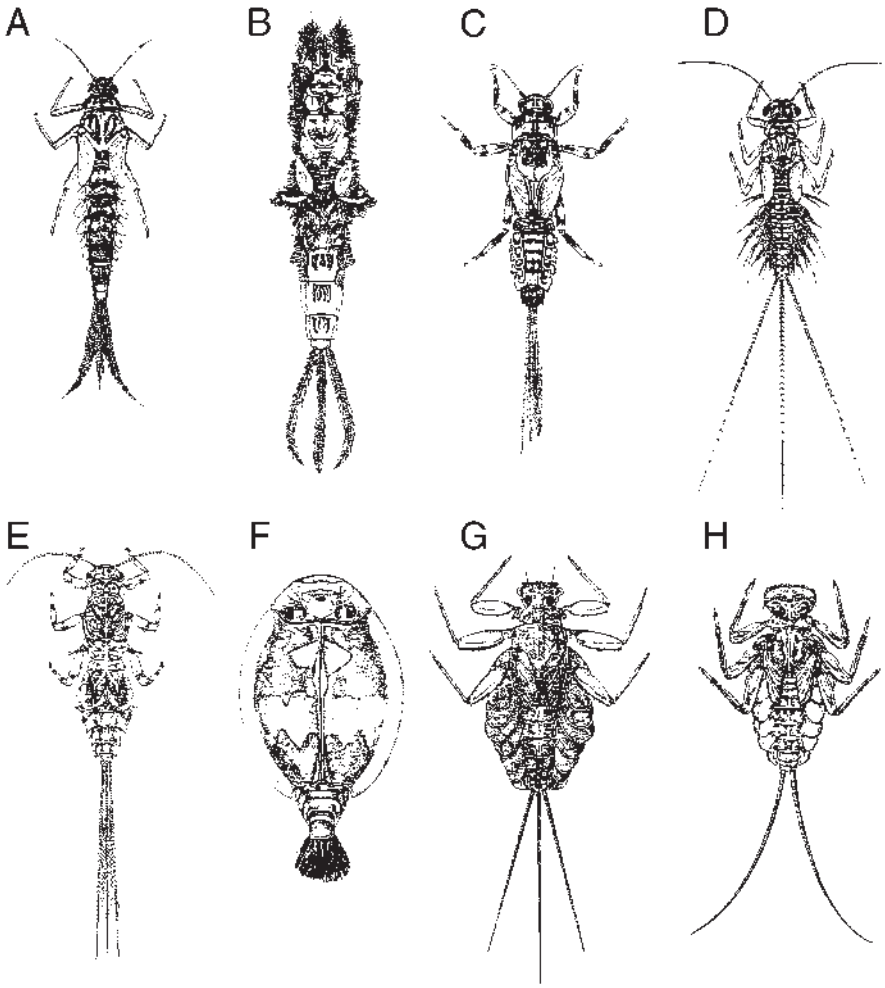

FIGURE 2 Mayfly nymphs: (A) *Baetis subalpinus* (family Baetidae), (B) *Ephemera danica* (family Ephemeridae), (C) *Ephemerella mucronata* (family Ephemerellidae) (D) *Leptophlebia vespertina* (family Leptophlebiidae), (E) *Caenis robusta* (family Caenidae) (F) *Prosopistoma boreus* (family Prospistomatidae), (G) *Lepeorus thierryi* (family Leptophlebiidae), and (H) *Epeorus alpicola* (family Heptageniidae). Illustrations show some of the large range in morphology, often related to habitat and food habits and not necessarily to family relationships. For example, *L. thierryi* and *E. alpicola* are morphologically similar and adapted to fast-running waters but belong to different families.

The latter location is more typical of the burrowing species that inhabit deeper waters and of a number of river species. Genera such as *Siphlonurus, Isonychia,* and *Baetisca* crawl completely out of the water before they molt.

Diel Patterns

In temperate regions, the crepuscular emergence of mayflies is well known. However, dusk is not the only time of day that mayflies emerge, although most species exhibit clear diel patterns of emergence that are, with few exceptions, characteristic for a given species, genus, or even a whole family. For example, the emergence of the short-lived Caenidae invariably takes place either at dawn or dusk and seems to be controlled by light intensity. Several baetid and leptophlebiid genera emerge around midday. In temperate areas, the higher daytime air temperatures are less restrictive for flight activity, although the adults are probably more susceptible to predation.

In the tropics and warm temperate regions, night air temperatures are less restrictive, and to escape from daytime predators it seems that most longer-lived forms emerge during the first two hours of darkness. The shorter-lived genera, such as *Caenis,* are subject to fewer restraints on their emergence, and there are few constant differences between tropical and temperate species.

The daily emergence of males and females is usually synchronous, especially in the short-lived forms, although there may be an excess of males at the start of the day's emergence. In species in which the females oviposit as subimagos, the males, which molt to imago, emerge well before the females.

Seasonal Patterns

Mayflies have distinct and finite emergence periods, especially in temperate and arctic areas. In the tropics, emergence is often nonseasonal, although some species have clear

emergence patterns. The lunar rhythm of emergence from a number of lakes of the African species *Povilla adusta,* is well known. The burrowing mayflies of the Ephemeridae, Polymitarcyidae, and Oligoneuriidae are noted for their sporadic mass emergence. The mass emergence of *Hexagenia* from the Mississippi River has been well documented. There are latitudinal and altitudinal gradients in the timing of emergence. For example, in both North American and European *Leptophlebia,* emergence occurs progressively later as one moves northward. In a similar way, the onset of emergence is delayed with increasing altitude. In habitats with several mayfly species, peak emergence of the major species may be separated in time, especially in congeneric species.

It has been suggested that emergence falls into two main categories: synchronized and dispersed, and represents two approaches for reducing adult mortality. Synchronous emergence attempts to saturate a potential predator, and dispersed emergence seeks to lower the possibility of predator–prey encounters. However, emergence pattern can vary with abundance and locality, and from year to year within the same species.

Water temperature thresholds, often in conjunction with rising temperatures, are important for both seasonal and daily emergence of many mayflies. Photoperiod has also been suggested as a potential factor regulating seasonal emergence in mayflies; few concrete data are available, however, and successful emergence occurred when nymphs were reared in complete darkness. Other abiotic factors may also affect daily emergence totals.

LIFE CYCLES

There is an extensive literature on mayfly life cycles, although mostly from temperate areas in Europe and North America. However, care should be taken in the interpretation of mayfly life cycles, especially when only field observations are available. Particular care is necessary in interpreting the length of time for egg development from field data.

Several authors have classified mayfly life cycles; most have used a combination of voltinism, duration of egg development, and nymphal growth rates as criteria. Multivoltine species usually have two or three generations in temperate regions, often a slowly growing winter generation and one or two rapidly growing summer generations. Limited data from the tropics, where many species are nonseasonal, indicate that some species go through about four and possibly up to six generations during the course of a year.

In temperate areas, the univoltine life cycle is the most widespread type. Several authors have distinguished two main types of univoltine cycle: when overwintering occurs during the nymphal stage after a relatively short egg developmental period, and when hatching occurs in the spring after a long period of egg development. Semivoltinism, with generation times up to 3 years, is relatively uncommon in mayflies.

Mayfly life cycles show a distinct trend from the tropics to the Arctic. In the tropics, nonseasonal multivoltine cycles predominate, with seasonality becoming more distinct in mountainous and continental areas. As one approaches the Arctic, univoltine cycles dominate.

Many mayflies exhibit flexibility in life cycle, whereas some mayflies (e.g., the widespread species *Leptophlebia cupida*) have a univoltine winter cycle over a wide range of latitudes and climates. However, a number of common and widespread species display a considerable degree of life cycle flexibility throughout their distributional range. This is perhaps best exemplified by many Baetidae, which may switch from multivoltine to univoltine depending on climate. The North American *Hexagenia* show a similar flexibility.

ABIOTIC AND BIOTIC RELATIONSHIPS

Nutrition

The majority of mayfly nymphs are herbivores, feeding on detritus and periphyton (algal communities on stones and plants). This explains their relative uniformity in mouthparts. The modifications that are present are a result of different food-gathering mechanisms rather than differences in diet. The herbivorous mayflies fall into two main categories, collectors and scrapers. Among the collectors, several genera are filter feeders, with setae on the mouthparts or forelegs acting as filters. Oligoneuriidae, Leptophlebiidae, Siphlonuridae, and the Heptageniidae have several genera that are probably filter feeders. By using their gills to produce a current of water through their burrows, several of the Ephemeridae and Polymitarcyidae may, at least for part of their food supply, be regarded as filter feeders. To supplement their diet, *Povilla* nymphs, especially the larger ones, leave their burrows at night and graze on periphyton. Most mayflies, however, are fineparticle detritivores. These include many Siphlonuridae, Baetidae, Leptophlebiidae, Metretopodidae, Ephemerellidae, Caenidae, and Baetiscidae, as well as some Heptageniidae. Members of the other major feeding group within the mayflies, scrapers, feed on the periphyton present on mineral and organic surfaces. These include representatives of several mayfly families, notably the Baetidae, Heptageniidae, Leptophlebiidae, and Caenidae. Shredders are probably also represented among mayflies.

True omnivory is of limited occurrence in the mayflies and is restricted to some species in genera such as *Isonychia, Siphlonurus, Stenonema,* and *Ephemera.* The predatory habit is also relatively uncommon in the mayflies. In North America, *Dolania, Analetris,* and the heptageniid, *Pseudiron, Spinadis,* and *Anepeorus* feed largely on chironomids. The baetid genera *Centroptiloides* and *Raptobaetopus* have carnivorous nymphs. Within the Prosopistomatidae there are also carnivorous species. Several species, such as *Siphlonurus occidentalis* and *Stenonema fuscum,* may change from a predominantly detrital diet in the early instars to one containing a significant proportion or even a dominance of animal material in the mature nymphs.

The time for food to pass through the gut is often short, and in *Baetis, Cloeon,* and *Tricorythodes* it has been shown to

be only about 30 mins. *Hexagenia* nymphs feed continuously during the day and night, and at most temperatures they ingest over 100% of their dry body weight per day. In contrast, values for the surface-dwelling collector *Stenonema* are much lower and vary between 2 and 22% of dry body weight per day. The carnivorous *Dolania,* feeding more intermittently but on a higher energy diet, has consumption indices similar to those of *Stenonema.* Studies have shown little or no cellulase activity in mayflies, whereas the proteolytic activity of trypsin- and pepsinlike enzymes is very high.

Predation

Mayfly nymphs are eaten by a wide range of aquatic invertebrate predators, including stoneflies, caddisflies, alderflies, dragonflies, water beetles, leeches, triclads, and crayfish. Mayflies are also important food organisms for fish. Birds and winged insects, such as Odonata, also prey on mayfly adults. Birds may take both the aquatic nymphs and the aerial adults. Several other animal groups, including spiders, amphibians, marsupials, and insectivorous mammals such as bats and shrews, have been reported to take mayflies. Many parasites also utilize these food chain links.

Symbiosis, Phoresy, and Parasitism

There is a wide range of organisms that live on or in mayflies. They include the normal spectrum of protozoan, nematode, and trematode parasites, and phoretic and commensal relationships with other organisms occur, as well. Chironomids in the genus *Symbiocladius* are ectoparasites and may cause sterility, although ectoparasites in the genus *Epoicocladius* do not seem to be detrimental to their host. In fact the cleaning effect, especially of the gills, may facilitate oxygen uptake in the mayfly. Mayflies can also be commensal, and two baetid genera, *Symbiocloeon* from Thailand and *Mutelocloeon* from West Africa, live between the gills of freshwater mussels.

DISTRIBUTION AND ABUNDANCE

Because of their fragility and short adult life, mayflies are generally rather limited in their dispersal powers. Together with their ancient origin and the strict association of larvae with freshwaters habitats, Ephemeroptera represent an interesting group for biogeographical analyses. The Siphlonuridae and allied families, typically cool-adapted mayflies, are mainly distributed in the temperate Northern Hemisphere, except for the Oniscigastridae, Nesameletidae, Rallidentidae, and Ameletopsidae, which are confined to New Zealand, Australia, and southern South America. We can hypothesize that this lineage was already present on the Pangaea, and radiated later on in Laurasia (Northern Hemisphere continent). Gondwanian representatives (Southern Hemisphere continent) expanded over the transantarctic land bridge and were confined to cool habitats.

The weak dispersal power of mayflies also results in a high percentage of endemism. Many species colonizing cool running waters in the European Alps are found nowhere else, but have related species in the Pyrenees or the Carpathians. On some islands, such as Madagascar and New Caledonia, endemism in mayflies reaches 100%. In contrast, many species that are effective dispersers may have very wide distributions.

Worldwide, two families, the Leptophlebiidae and the Baetidae, are especially important both in terms of abundance and diversity, representing half of the known species. In contrast, the Siphlaenigmatidae (New Zealand) and Dipteromimidae (Japan) encompass only one species apiece.

The distribution and abundance of mayflies has received considerable attention. Within the basic zoogeographical limitations, abiotic factors, notably temperature, substratum, water quality, and, in running water, current speed, seem to be the most important. Other factors, such as ice, floods, drought, food, and competition, may also influence abundance and distribution. Generally, the number of mayfly species decreases with increasing altitude.

Many lotic mayflies are either dorsoventrally flattened or streamlined as an adaptation to life in swift current. The physical substratum also traps different amounts of detritus and silt, and this is a major factor influencing microdistribution. The richest mayfly community is often found in association with aquatic vegetation, which, as well as providing shelter, functions as a detrital trap and as a substratum for periphyton. For burrowing mayflies, the presence of the correct substratum is obviously a major determinant of both macro- and microdistribution. In lakes, the highest mayfly diversity occurs in the shallow littoral areas. At deeper levels, the mayfly fauna, although often reaching high densities, is usually poor in species. Mayflies are generally absent from the profundal (the deep waters where light does not penetrate) of lakes. Many mayflies can tolerate a wide range of salinities, and a few species within the Baetidae, Caenidae, and Leptophlebiidae occur in brackish water.

Mayflies constitute a major part of the macroinvertebrate biomass and production in freshwater habitats. Seasonal variation in density, biomass, and annual production are strongly influenced by life cycle parameters, indicating the importance of correct life cycle information in production studies. Most mayfly production values, expressed in terms of dry weight per square meter per year, are in the range of 0.1 to 10.0 g.

HUMAN INFLUENCE

Humans increasingly affect the distribution and abundance of mayflies and, by virtue of their widespread occurrence and importance in aquatic food webs and particularly in fish production, mayflies have been widely used as indicators of water quality. Mayflies often occur in habitats of a particular trophic status, and increased eutrophication due to human activities can lead to the reduction or even extinction of certain species. *Baetis* species are often among the most tolerant of mayflies to pollution. In North America, the use of mayflies

as indicators of water quality has not escaped attention. The mass emergence of burrowing mayflies from Lake Erie and the Mississippi River has provided a useful barometer of water quality. Organic and nutrient enrichment of Lake Erie in the 1940s and 1950s led to an increase in the intensity and frequency of mass emergence of *Hexagenia* until 1953, when prolonged periods of oxygen depletion in the hypolimnion (the lower layer of cold water in lakes that stratify) caused the population to crash to virtual extinction. However, improvement of water quality has now led to a resurgence of emerging swarms. Mayflies, particularly *Hexagenia,* have been used in numerous bioassays for various pollutants. Pesticides also affect nontarget organisms such as mayflies, and Canadian studies in connection with blackfly control have demonstrated catastrophic drift and reduced biomass in mayfly populations over long distances in rivers treated with methoxychlor. Although most mayflies are adversely affected by petroleum products, a few species may show small increases owing to the extensive algal growth that often occurs on oiled substrates.

Acidification of fresh waters is a major threat to mayfly communities. Many mayflies are affected adversely by low pH, and emergence is a particularly critical period. The genus *Baetis* seems to be particularly sensitive and is often replaced by less sensitive *Leptophlebia* and *Siphlonurus.*

River and lake regulation (e.g., by impoundment in reservoirs) for water supply and power can have profound effects on the mayfly community, especially when there is a hypolimnion drain. For example, an increase in winter temperatures and a fall in summer temperatures may remove obligatory life cycle thresholds, leading to extinction. Fecundity may also be influenced by changes in water temperature. In reservoirs themselves, lentic (still water) conditions and increased water level fluctuations usually produce a reduced mayfly fauna, although there may be an increase in the abundance of burrowing and silt-dwelling species. The flooding of new areas can also create new habitats for mayflies, and in many of the large African reservoirs the mayfly *Povilla adusta* has developed large populations, which burrow into the submerged trees and play an important role in tree breakdown. It has recently been demonstrated that ovipositing mayflies are deceived by asphalt roads because the strongly polarized light reflected from the surface mimics a water surface, thus representing a threat to successful reproduction. Climate change scenarios involve changes in water temperatures, which in turn will affect many of the facets of mayfly biology and lead to changes in mayfly communities.

See Also the Following Articles

Aquatic Habitats • Pollution • Respiratory System

Further Reading

Alba-Tercedor, J., and Sanchez-Ortega, A. (eds.). (1991). "Overview and Strategies of Ephemeroptera and Plecoptera." Sandhill Crane Press, Gainesville, FL.

Brittain, J. E. (1982). Biology of mayflies. *Annu. Rev. Entomol.* **27,** 119–147.

Campbell, I. C. (ed.). (1990). "Mayflies and Stoneflies: Life Histories and Biology." Kluwer, Dordrecht, the Netherlands.

Ciborowski, J. H., and Corkum, L. D. (eds.). (1995). "Current Directions in Research on Ephemeroptera." Canadian Scholars' Press, Toronto.

Florida Agricultral & Mechanical University, http://www.famu.edu/mayfly/

Hubbard, M. D. (1990). "Mayflies of the World—A Catalogue of the Family and Genus Group Taxa (Insecta: Ephemeroptera)." Sandhill Crane Press, Gainesville FL.

Landolt, P., and Sartori, M. (eds.). (1997). "Ephemeroptera & Plecoptera: Biology–Ecology–Systematics." MTL Fribourg, Switzerland.

McCafferty, P. (1981). "Aquatic Entomology." Science Books International, Boston.

Purdue University, Department of Entomology. http://www.entm.purdue.edu/Entomology/research/mayfly/mayfly.html

Estivation

see *Aestivation*

Evolution

see *Phylogeny*

Excretion

Timothy J. Bradley
University of California, Irvine

Excretion is the elimination from the body of excess ions, water, and metabolic wastes. As in all organisms, excretion in insects serves to promote the appropriate regulation (homeostasis) of the intracellular environment as the key to organismal well-being and survival. Regulation of the intracellular environment depends in turn on appropriate regulation of the extracellular fluids. In insects, these extracellular fluids are contained in the hemolymph and in the interstitial fluids surrounding the cells. Excretion therefore refers to the processes by which wastes, as well as excess ions and water, are eliminated from the hemolymph. Wastes produced in the cells and transported into the hemolymph are concentrated and excreted by the excretory organs.

STRUCTURE OF THE EXCRETORY ORGANS

The processes of excretion in insects are carried out largely by the organs of the insect gut. These include the midgut, the

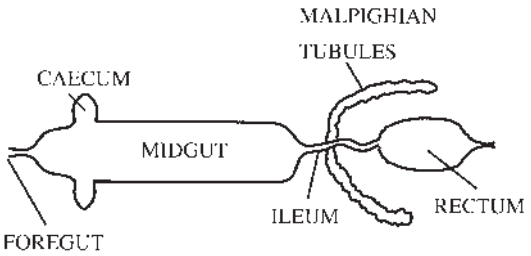

FIGURE 1 A diagrammatic representation of the insect gut.

Malpighian tubules, and the structures in the hindgut, namely the ileum and rectum (Fig. 1).

Midgut

The midgut is a tubular epithelium. Upon ingestion, food and fluids move through the esophagus and pass directly into the midgut. Because the cells in the midgut epithelium are derived from embryonic endoderm, the midgut is not lined with cuticle. Most of the cells in the midgut are involved in the secretion of digestive fluids and the absorption of nutrients from the midgut lumen. These secretory and absorptive cells have apical microvilli that greatly increase the surface area available for inward and outward transport. In insects that feed periodically (such as adult mosquitoes), the microvilli shorten during nonfeeding periods and lengthen following ingestion. Many insects have additional cell types termed goblet cells that are thought to be involved in the secretion of fluids that modify the acidity and alkalinity (pH) of the luminal fluid. These goblet cells have been intensively investigated in Lepidoptera, where they serve to produce a markedly alkaline pH.

The basal surface of the midgut cells possesses a network of longitudinal and circular muscles that upon contraction can produce peristaltic waves. These contractions serve to move the food along the gut and stir the midgut contents during digestion. Many insects possess globular outpocketings in the anterior region of the midgut, termed ceca. The cells types in the ceca are generally differentiated from those in the midgut proper.

Malpighian Tubules

The Malpighian tubules are the site of urine formation in all insects except the Collembola, Thysanura, and aphids. The Malpighian tubules are tubular epithelia that are diverticulae (outpocketed extensions) of the gut itself. The tubules open into the gut near the midgut–hindgut junction, and the lumina of these two tubular epithelia are continuous. The contents of the tubules flow into the gut lumen; the ends of the tubules distal to the gut are closed. Fluid is produced in the Malpighian tubules by secretion; and because the tubules are closed at the distal end, hydrostatic pressure builds up and fluid flows through the tubules into the gut.

The number of Malpighian tubules is quite variable depending on the insect species. Bloodsucking Hemiptera (e.g., *Rhodnius prolixus*) and higher Diptera (e.g., *Drosophila melanogaster*) have as few as four tubules, whereas the desert locust *(Schistocerca gregaria)* has hundreds. Attached to the Malpighian tubules of many insects are longitudinal muscles. When these muscles contract, the tubules are waved about in the hemolymph, presumably for the purpose of stirring the fluid adjacent to the tubules and promoting fluid and solute transport. These tubules may also serve the more general function of promoting hemolymph circulation throughout the abdomen.

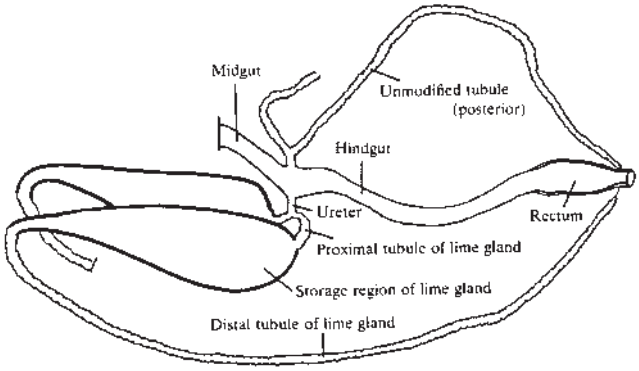

FIGURE 2 A schematic diagram of the Malpighian tubules of *E. hians.* The Malpighian tubules are differentiated on the left and right side. The hindgut is composed, from anterior to posterior, of the ileum, colon, and rectum. [From Herbst, D. B., and Bradley, T. J. (1989). A Malpighian tubule lime gland in an insect inhabiting alkaline salt lakes. *J. Exp. Biol.* **145,** 63–78.]

The Malpighian tubules of all species examined to date contain more than one cell type. In some cases, a single epithelial region contains two or more cell types (regions with heterologous cell types) reflecting, presumably, separate physiological roles for each cell type. In other species, the tubules are divided into distinct regions, each consisting of a single cell type (regions with homologous cell types). In these insects, each tubule region has a distinct function in transport. Finally, in many insects, the tubules show regional specialization as well as multiple cell types within a region. It is presumed that each cell type in these tubules performs a distinct function.

As an example of cell type heterogeneity, consider the Malpighian tubules of the larvae of the brine fly, *Ephydra hians* (Fig. 2). The tubules in this insect are differentiated on each side of the body, as well as along their length. On one side of the body is a pair of tubules called the lime gland tubules. The distal ends of these are secretory and contain two regions that can be differentiated on the basis of cell color: one white, one yellow. More proximal to the gut are expanded regions of the tubules that serve to store concentric concretions in the tubule lumen. Finally, two of these cells combine in a common ureter that empties into the gut. On the opposite side, the tubules have only the yellow and white regions of the tubules, with no storage section. This example illustrates the variety of cell types that can exist in a single tubule. The details of transport function in these and other highly complex tubules have not been fully elucidated.

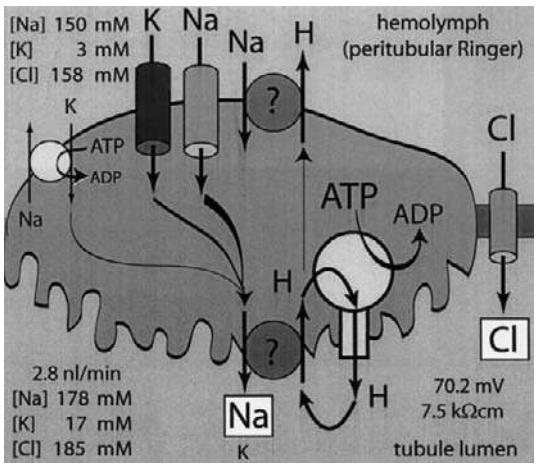

FIGURE 3 Model of the transport processes occurring in the Malpighian tubules of adult mosquitoes, based on the work of Klaus Beyenbach using the species *A. aegypti*. Upon stimulation with mosquito natriuretic peptide (MNP), rates of transepithelial fluid secretion increase from an unstimulated rate of 0.4 nl min^{-1} to 2.8 nl min^{-1}. In parallel, Na^+ concentrations in the secreted fluid rise, and K^+ concentrations fall. Electrophysiological studies reveal that MNP, working via cyclic AMP, induces an increase in basolateral membrane Na^+ conductance, presumably through the actions of Na^+ channels in the basolateral membrane of principal cells. The hyperpolarization of the transepithelial voltage and the decrease in transepithelial resistance are consistent with the activation of Na^+ channels in the basolateral membrane of principal cells. (Figure and legend provided by Klaus Beyenbach.)

In most Malpighian tubules, formation of the primary urine occurs in a cell type often referred to as the primary cell (Fig. 3). This cell type has extensive apical microvilli, often containing a central core of microfilaments. Frequently, these microvilli contain fingerlike extensions of the mitochondria and even of the endoplasmic reticulum. The basal surface of the cells exhibits deep infolds, often again closely associated with mitochondria by means of structures termed scalariform junctions. The intercellular spaces are occupied apically by septate or continuous junctions. More basally, the intercellular space contains gap junctions or undifferentiated basolateral membranes indistinguishable from the basal membrane infolds.

Numerous other cell types occur in the Malpighian tubules. It is presumed that each histologically distinguishable cell type performs a unique function within the tubules. In addition, distinct functions have been found in some cell types in the absence of histological or ultrastructural differentiation. A common cell type in Malpighian tubules is the stellate or secondary cell. This cell type possesses smaller microvilli than the primary cells, and these microvilli contain no mitochondria. As described later in the section on function, the secondary cells may be involved in modification of the primary urine produced by the primary cells.

Ileum

Posterior to the midgut, most insects possess a segment of gut referred to as the ileum. Because this region is part of the hindgut, it is covered on the apical surface by cuticle. The cells often show deep apical and basal membrane infoldings reflecting the role of these cells in fluid and solute transport. The basal surface of the cells is covered by layers of longitudinal and circular muscle that serve to generate peristaltic movements of the gut. These muscular contractions move the gut contents through the gut and may also serve an important role in reducing unstirred layers adjacent to both the apical and basal membranes of the epithelium.

Historically, the Malpighian tubules and rectum have been assumed to carry out most of the fluid transport in the posterior regions of the gut. The ileal epithelium is smaller in diameter, with less highly developed apical and basal membrane infolds, and a lower mitochondrial density than is observed in rectum. Nonetheless, when the transport properties of the ileum are investigated, this region of the gut is always found to be carrying out important transport functions.

Rectum

All insects possess an enlarged chamber called the rectum near the posterior end of the gut. The structure of the rectum can vary substantially from species to species. The rectal lumen is covered by a thick cuticle. Posterior to the rectum, insects possess an anal canal through which the feces and urine are eliminated. The strong bands of muscle surrounding the rectum contract during defecation, expelling the feces and urine into the external environment through the anus.

The recta of terrestrial insects are large and very active organs. Regions within the rectum are highly differentiated and contain cells with deep membrane folds. If (as in many orthopterans and lepidopterans) these cells are contained in thickened ridges that extend into the rectal lumen, they are referred to as rectal pads. If (as in many adult Diptera) the cells extend into the lumen as fingerlike structures, often on a thin stalk, they are referred to as rectal papillae. The rectal pads and the papillae are the major sites of fluid resorption and urine concentration. The cells in the rectal pads have a complicated array of intercellular junctions associated with the active sites of ion secretion and resorption. The evolution of a rectal structure permitting the formation of a urine hyperosmotic to the hemolymph was, along with the evolution of a waxy cuticle, a major adaptive event permitting insects to invade drier terrestrial habitats.

Both rectal pads and rectal papillae are covered with thick cuticle to protect the underlying epithelial cells from abrasion by the fecal material. Between the rectal pads, the rectum of terrestrial insects possesses a thin, cuticle-lined epithelium that, upon unfolding and stretching, permits rectal swelling during feces and fluid accumulation. In *E. hians,* the rectum is merely a thin, distensible epithelium that expands upon filling with feces prior to defecation. In this insect, the task of modifying the urine is carried out by the colon, which lies just anterior to the rectum. The presence of a rectum without transport capabilities is very unusual. In most insects the rectum is the major organ responsible for osmotic regulation of the urine.

Excretion outside the Gut

Storage excretion of concentric concretions containing calcium salts of urate and carbonate occurs in the fat body cells of most insects. Some insects contain nephrocytes, cells in the head that also store wastes and toxic elements. Finally, in collembolans and thysanurans, cephalic nephridial glands, sometimes termed labial glands, are responsible for excretory function, including the excretion of nitrogenous waste.

FUNCTION OF THE EXCRETORY ORGANS

Midgut

TRANSPORT The primary function of the midgut in insects is the digestion and uptake of ingested nutritive materials. There are two processes that occur in the midgut and contribute to excretion: pH regulation and storage excretion. The midguts of many insects secrete fluids that contribute to extreme alkalinity or acidity in the gut. In Lepidoptera, for example, the midgut epithelium consists of goblet and columnar epithelial cells. The goblet cells are responsible for transporting ions and creating a highly alkaline solution in the midgut lumen. The columnar cells contribute to digestion of the food and uptake of ions.

The apical membrane of the goblet cell contains hydrogen pumps (H^+-ATPases) that use adenosine triphosphate (ATP) as their source of energy. These pumps transport hydrogen ions into the cuplike apical crypt of the goblet cell. The resulting electrical and pH gradient is used in an exchange process (antiporter) to exchange potassium ions for the hydrogen ion. Potassium ions associated with hydroxyl ions remain in the lumen following the hydrogen/potassium exchange process. These ions diffuse from the goblet cavity into the midgut lumen, with the result that the potassium hydroxide (KOH) causes the lumen to be very alkaline. The high potassium concentration in the lumen is, in turn, used as an energy source for the uptake of amino acids from the digested foodstuffs. This uptake occurs by means of transporters (symporters) in the apical cell membranes of the columnar cells that use the potassium gradient to cotransport potassium ions and amino acids from the lumen into the cell interior.

Various functions have been proposed for the highly alkaline pH in the midgut of Lepidoptera. Clearly, such an extreme pH would serve to kill pathogens and to saponify lipids in the ingested food material. It has also been proposed that this pH serves to reduce the solubility and activity of tannic acids in the food of Lepidoptera, protecting herbivorous larvae from the high concentrations of these toxic compounds found in the leaves of many plants.

In other insects, for example, mosquito larvae, low-pH (acid) conditions are observed in the midgut. The precise mechanisms by which this pH is produced and maintained remain to be elucidated, but there is evidence in mosquitoes that the midgut epithelium is also energized by the H^+ -ATPase.

STORAGE EXCRETION IN THE MIDGUT The columnar midgut cells of insects often contain concentric mineralized concretions. The major cation in these concretions has been shown to be calcium. The anion can be quite variable and has been found to be largely urate or carbonate, depending on the cell type in which the concretions are located. In the midgut, the concretions have also been shown to play a role in the storage excretion of toxic cations such as copper, zinc, iron, and selenium. Each of these elements is toxic in high concentrations but plays a critical role in metabolism in trace amounts. The concretions may therefore play both a protective and a homeostatic role in insects by regulating the free concentration of these ions and metals in the body. They reduce toxic concentrations and serve as a reservoir for these elements, which can be used for physiological purposes when required. Storage excretion in the form of concentric concretions also occurs in the Malpighian tubules.

Malpighian Tubules

TRANSPORT The production of urine in insects occurs by the active transport of ions across the epithelium from hemolymph to tubule lumen. This process generates an osmotic concentration that drives the movement of water across the epithelium as well. Generally, Malpighian tubules have a high permeability coefficient for water (a low osmotic resistance), and as a result water moves rapidly across the epithelium in response to relatively low osmotic concentration gradients. These concentration differences across the epithelium are indeed so low that they have been difficult to measure. Recent experimental results have led to a general consensus, however, that an osmotic gradient of a few milliosmoles is sufficient to account for the observed rates of water movement across Malpighian tubules.

Let us consider first the active transport of ions across the epithelium, and then the passive movement of water that follows. In insects, potassium is the predominant cation transported across the epithelium of the Malpighian tubules. Insects evolved as a distinct clade on land, feeding on plants and detritus. It may be that their dependence on potassium as the major cation used for fluid transport evolved at this time as well. Certainly, animals of marine origin, such as the vertebrates and crustaceans, rely on sodium as the principal cation for driving fluid movements.

In the cell interior, the Malpighian tubule cells have a negative electrical charge relative to the hemolymph. This electrical potential facilitates the entry of potassium into the cells. Thermodynamically speaking, therefore, the most energetically costly transmembrane movement for potassium in the epithelium occurs as this cation crosses the apical membrane. The process by which this occurs has been very difficult to unravel, but in recent years it has been demonstrated that a very active hydrogen ATPase, related to vacuolar H^+-ATPase found widely in eukaryotic cells, is located on the apical membrane of Malpighian tubule cells. This transporter

moves hydrogen ions from the cell interior into the tubule lumen, thereby setting up a large electrochemical gradient for hydrogen ions. Depending on the circumstances, this electrochemical gradient can be predominantly expressed as a pH gradient or as an electrical gradient. In either circumstance, this electrochemical gradient is thought to serve as an energy reservoir that can be subsequently used for a variety of coupled transport processes.

For example, if the energy contained in the hydrogen ion gradient is used in an antiporter (a transporter that couples ion movement in one direction to ion movement in the opposite direction) that exchanges hydrogen for potassium at the apical membrane, the result of hydrogen transport with subsequent hydrogen exchange for potassium is a net active transport of potassium. Potassium is moved against its electrochemical gradient by the antiporter, using the energy contained in the transmembrane hydrogen ion gradient. In some Malpighian tubules, similar mechanisms may exist for the coupled transport of other cations, (e.g., sodium, calcium, or magnesium). In bloodsucking insects such as adult mosquitoes and the reduviid *R. prolixus,* the plasma portion of the blood meal also provides the insect with a substantial intake of sodium. Any amount of sodium that exceeds the physiological needs of the insect must be excreted. In these insects, a substantial active transport of sodium occurs accords the Malpighian tubules. This process is thought to be driven across the apical membrane by a hydrogen–sodium exchange mechanism.

Cation transport across the epithelium also requires a process for entry of the ions into the cytoplasm from the hemolymph. In most insects, the basal membrane contains an enzyme (Na^+-K^+ ATPase) that uses the energy in ATP to transport sodium actively out of the cell and potassium actively in. As a result, the cytoplasm is greatly enriched in potassium. The passive outward diffusion of this ion through barium-sensitive channels produces an electrical potential across the basal membrane, the inside of which is negatively charged. This potential can be used for a variety of transport functions; one that seems to be almost universally present in Malpighian tubules is the bumetanide-sensitive $Na^+/K^+/2Cl^-$, cotransporter. This transporter uses the energy contained in the sodium gradient to move one sodium, one potassium, and two chloride ions simultaneously from the hemolymph into the cytoplasm. This process serves to provide chloride to the cell interior, as well as sodium in the cell types to which sodium is moved transepithelially.

The movement of anions across the epithelium involves one or more of three distinct transport pathways, depending on the characteristics of the tubules. In the first, chloride is thought to move across the same cells as those in which the cations are transported. As mentioned earlier, the interior of the cells is electrically negative relative to the cell exterior. The movement of chloride into the cell is therefore a thermodynamically active process. As already discussed, it may be driven by the sodium electrochemical gradient in a cotransport process. In other insect species in which the Malpighian tubule cells in a single region of the tubule are differentiated into two or more cell types, chloride ions have been shown to enter the lumen via a cell type distinct from that involved in active cation transport. For example, in *D. melanogaster,* the fluid-transporting segment of the tubules contains both primary and stellate cells. The former cells are the sites of active potassium transport, the latter the site of chloride flux from the hemolymph to the tubule lumen. Although this movement of chloride into the tubule lumen is thermodynamically downhill, the precise mechanism of chloride transport is presently unknown.

A third process has been described in the Malpighian tubules of adult mosquitoes. Although these insects also possess primary and stellate cells in the Malpighian tubules, it has been proposed that chloride moves into the lumen of the tubules via the intercellular junctions. In fact, this process has been shown to be under hormonal control. The movements of chloride into the Malpighian tubules may therefore be quite variable depending on cell types. The movements of anions are much less well characterized at this time, with regard to the molecules that drive the process, than are the movements of cations. A model of ion transport at both the apical and basal membrane of the Malpighian tubules of adult mosquitoes is shown in Fig. 3.

Regardless of the mechanisms by which cations and anions enter the Malpighian tubules, it is clear that the types of ions transported can vary greatly with the species of insect and will depend on an individual's physiological needs and demands. Thus since blood-sucking insects ingest a large amount of sodium compared with other insects, the Malpighian tubules of bloodsuckers contain specific mechanisms designed to reduce the large sodium load. Species of mosquitoes whose larvae can survive in salt water must ingest the medium and eliminate the ions as means of obtaining water. Those species that have been investigated can excrete magnesium and sulfate via the Malpighian tubules. The larvae of brine flies (ephydrids) generate crystals in the lumina of the tubules that are rich in calcium and carbonate. Both these ions must be transported across the epithelium, although the combination forms insoluble crystals that reduce the activity of these ions in the lumen.

The Malpighian tubules of insects are also the site of excretion of the waste products of energy and nitrogen metabolism. Acid by-products of energy metabolism have been shown to be actively transported into the lumen from the hemolymph. The precise molecular mechanisms of the process remain unclear, but the process is of paramount importance for the insects in maintaining acid/base balance and energy homeostasis. The by-products of nitrogen metabolism are also excreted by the Malpighian tubules.

In aquatic insects, ammonia may be excreted, but in most insects and certainly in terrestrial forms, urea and uric acid predominate. Both these compounds are actively removed from the hemolymph by the Malpighian tubules. The transport of uric acid has been investigated in *R. prolixus,* in which the blood meal provides a very protein-rich meal

requiring intense capacity for the elimination of nitrogenous waste. In *Rhodnius,* the primary urine is produced in the most upstream portions of the Malpighian tubules, the upper tubule. This urine is modified in the downstream section (the lower tubule) through the resorption of potassium and chloride. This process serves to return potassium to the hemolymph, and to remove waste from the hemolymph, through the retrieval of an isosmotic fluid. Thus hemolymph volume is retained and the sodium in the urine is concentrated. In addition, uric acid is transported in the lower tubule from hemolymph to urine. Potassium urate is fairly insoluble, particularly at neutral to acid pH. As a result, crystals of uric acid form in the urine. This process further removes osmotically active compounds from the urine, allowing the additional movement of water from the urine to the hemolymph by osmosis.

CONCENTRIC CONCRETIONS IN THE MALPIGHIAN TUBULES Concentric concretions occur in the midgut, where they are thought to contribute to excretion by storage in an insoluble form of salts containing calcium, magnesium, manganese, copper, cadmium, and zinc. Identical concretions are found intracellularly in the Malpighian tubules. These concretions are thought to perform an identical function, namely storage of ions in an insoluble form either for subsequent use or as a means of removing the ions from the body. In the Malpighian tubules, however, these concretions also appear in the tubule lumen, a location from which they can move into the gut and be eliminated with the excrement. It has been suggested by many authors that the intracellular concretions in the cells of the Malpighian tubules can be transported by exocytosis into the lumen of the tubules. Although there are occasionally physiological conditions in which the concretions disappear from the cells and appear in the lumen, it has not been unambiguously demonstrated that the crystals move from one location to the other intact. Instead, it is likely that the crystals are dissolved within the cells of the Malpighian tubules, that the soluble ions are transported into the lumen, and that the crystals are formed anew in the tubule lumen. Crystals are formed in some tubule segments (e.g., in the lower tubule of *R. prolixus*) where no crystals exist in the cells.

The crystals in the midgut, fat body, and Malpighian tubules are concentric and perfectly round. This is in marked contrast to the natural structure of the crystals formed by the same salts in solution. Uric acid crystals, for example, have sharp corners and sometimes take a needlelike form. It is thought that the concentric concretions avoid acicularity through the activity of organic compounds that are known to be a substantial component of the concretions. The compounds are thought to nucleate and direct crystal formation, leading to the formation of round concretions. This spherical shape is less damaging to the cells of the tubules and can be excreted from the tubules and gut with little tissue damage. Ultrastructurally identical concretions are observed in the urine of birds, which is rich in uric acid. It has been proposed that organic compounds are excreted into the tubule lumen, where they nucleate and guide the formation of the concretions. The ions contained in the concretions can vary greatly, ranging from potassium urate in some tissues to calcium carbonate in others. Even though such crystals should be quite distinct in shape, the concretions produced by the insects all have the same distinct concentric, spherical shape. This set of properties argues that the organic compounds have a profound effect on crystal form and formation.

Ileum

The principal function of the ileum is to act as a tubular epithelium that serves to transport to the rectum the undigested remains of the food from the midgut and fluid from the Malpighian tubules. This transport occurs by peristaltic movements of the circular and longitudinal muscles surrounding the ileum. The ileum also engages in important transepithelial transport functions. This has been investigated in considerable detail in the locust *Schistocerca gregaria.* In this species, potassium and chloride ions are transported from the lumen of the ileum into the hemolymph. This transport is iso-osmotic. It therefore does not contribute directly to osmotic regulation but serves instead to reduce the volume of the urine and to retain valuable ions and water in the hemolymph. This transport is under hormonal control, presumably to allow the insect to modulate the return of water to the hemolymph depending on whether osmotic condition of the animal dictates a diuretic or an antidiuretic response.

In some insects, an additional segment of the hindgut exists, which is termed the colon. Although this segment is hard to distinguish with the unaided eye, it is functionally and histologically distinct from the ileum. In larvae of *E. hians,* for example, an ileal segment occurs near the midgut, while a colonic segment of the hindgut lies between the ileum and the rectum. It has been shown that active ion transport occurs in the colon. The colon has a relatively low osmotic permeability, thus allowing the secretion in this segment of a fluid that is strongly hyperosmotic to the hemolymph. Production of hyperosmotic excreta is crucial for the osmotic regulation in this species because the insects live in the waters of a saline lake, the osmotic concentration of which is six times more concentrated than the hemolymph. The ions transported in the colon include sodium, chloride, and sulfate. Sulfate ions are large in comparison to other transported ions; therefore the transport of sulfate through an epithelium capable of maintaining a substantial osmotic gradient is unusual. The larvae of the blowfly, *Sarcophaga bullata,* have also been shown to engage in active transport in the colon. In this species, the colon is a major site for the excretion of nitrogenous waste in the form of ammonium ion. Because these larvae feed in rotting flesh, the active transport of ammonium is a critical adaptive feature in the physiology of the species.

Rectum

In most insects, the rectum is the most active ion-transporting organ on a per-gram basis. All fluids and solids deriving from the midgut and Malpighian tubules pass through the ileum and enter the rectum before being excreted. The rectum is therefore the last location in the gut in which the ionic and osmotic concentration of the excreta can be modified to meet the regulatory needs of the insect.

In terrestrial animals, the requirements for osmotic homeostasis vacillate between the production of a dilute excreta (diuresis) and the production of a concentrated excreta (antidiuresis). Control of the rectum is therefore a critical element in the maintenance of osmotic homeostasis in the hemolymph. None of the other elements of the excretory system discussed thus far are capable of producing a fluid differing in osmotic concentration from that of the hemolymph.

The role of the rectum in terrestrial insects has been most intensively studied in *S. gregaria.* In this insect, the cells in the rectal pads serve to transport a hypo-osmotic fluid from the lumen into the hemolymph. This serves to produce excreta with a very high osmotic concentration and, in the process, conserve water in the hemolymph.

The process by which the locust transports a hypo-osmotic fluid is complicated and, unlike a functionally analogous process in the kidney of mammals, it requires cells of only a single type, the cuticle-covered rectal pad cells. In the parts of the rectum differentiated into rectal pads, the cells underlying the cuticle have deep apical infolds associated with numerous mitochondria. The rectal epithelium in the regions of the rectal pads is thick, meaning that the rectal pad cells comprise a tall, columnar epithelium. The intercellular junctions in these cells are highly convoluted and contain open spaces or intercellular swellings in the clefts between the cells.

The process of fluid resorption from the lumen begins with the active transport of ions across the apical membrane (i.e., from the lumen to the intracellular compartment). Once in the cytoplasm, the ions are transported across the intercellular membrane into the enlarged spaces in the intercellular clefts. The compounds transported are principally potassium and chloride, although other compounds including acetate and proline are actively transported out of the lumen as well. These transported compounds produce a fluid with high osmotic concentration. It is thought that water is drawn from the lumen into the intercellular clefts, probably through the apical septate junctions. As a result, fluid accumulates in the intercellular clefts and in the open spaces in the intercellular regions. From here, the fluid flows extracellularly between the cells in a basal direction toward the hemolymph. It is thought that as this fluid flows, transporters within the lateral cell membranes remove ions. If these membranes have a low osmotic permeability, ions can move across with little water following. As a result, ions are removed faster than water can follow, resulting in a fluid that is hypo-osmotic not only to the lumen but also to the hemolymph.

Under conditions in which the insect is well hydrated (e.g., after eating lush vegetation), the rectum removes ions from the rectal lumen but little water follows, presumably because either the site or the rate of transport in the more lateral and basal membranes has been modified. This produces a dilute urine, the excretion of which serves the osmotic needs of the insect.

Aquatic insects are similarly dependent on the rectum for the final modification of the urine prior to excretion. In freshwater insects, the fluid derived from the midgut and Malpighian tubules is iso-osmotic to the hemolymph. Excretion of this fluid would lead to rapid loss of ions and the death of the animal. The rectum serves to transport ions from this primary urine back into the hemolymph.

Transport of potassium and chloride has been documented for number of freshwater insects. These transport mechanisms are relatively easy to demonstrate because the fluid entering the rectum from the Malpighian tubules is enriched in these two ions, and the excreted urine leaving the rectum much depleted.

Rectal function has also been investigated in aquatic insects residing in hyperosmotic media, for example, in saline-tolerant dipteran larvae inhabiting coastal and desert saline waters. In species of *Aedes* inhabiting these waters, the rectum is differentiated into two segments. The anterior rectal segment is identical in function to the rectum of freshwater species and serves to remove ions from the urine under conditions in which the larvae find themselves in hypo-osmotic media (i.e., fresh water). When the larvae hatch in saltwater, or when the medium becomes concentrated because of evaporation, the posterior rectal segment becomes active. This segment has a single cell type, which is characterized by deep apical and basal infolds associated with numerous mitochondria. The cells actively transport ions from the hemolymph into the rectal lumen. Because the epithelium has a low osmotic permeability, ions are transported faster than water can follow. As a result, a concentrated urine is produced by secretion in this segment, which has been called the salt gland.

The ions transported in the posterior rectal segment vary with the environment in which the larvae occur. In seawater, sodium, magnesium, and chloride predominate. In bicarbonate-rich waters, a concentrated fluid is secreted, and the urine is rich in sodium and bicarbonate. The precise molecular mechanisms of ion transport in the rectum, as well as their neuronal or hormonal control, are poorly known for aquatic insects.

See Also the Following Articles

Digestion • Fat Body • Hemolymph • Water and Ion Balance

Further Reading

Beyenbach, K. W. (1995). Mechanisms and regulation of epithelial transport across Malpighian tubules. *J. Insect Physiol.* **41,** 197–207.

Bradley, T. J. (1985). The excretory system: Structure and physiology. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), pp. 421–465. Pergamon Press, Oxford, U.K.

Bradley, T. J. (1998). Malpighian tubules. *In* "Microscopic Anatomy of the Invertebrates," Vol. XI, "Insecta" (M. Locke and F. W. Harrison, eds). pp. 809–829. Liss, New York.
Chapman, R. F. (1998). "The Insects: Structure and Function." Cambridge University Press, Cambridge, U.K.
Herbst, D. B., and Bradley, T. J. (1989). A Malpighian tubule lime gland in an insect inhabiting alkaline salt lakes. *J. Exp. Biol.* **145,** 63–78.
Phillips, J. E., and Audsley, N. (1995). Neuropeptide control of ion and fluid transport across locust hindgut. *Am. Zool.* **35,** 503–514.
Wigglesworth, V. B. (1965). "The Principles of Insect Physiology," 6th ed. Methuen., London.

Exopterygota

Exopterygota is a division of the class Insecta in the phylum Arthropoda. The orders of insects in this division have wings that develop externally during the maturation of the larva (which is variously referred to as a larva, nymph, or naiad). There are two superorders in this division: the Orthopteroidea (which includes the orders Blattodea, Dermaptera, Embiidina, Grylloblattodea, Isoptera, Mantodea, Mantophasmatodea, Orthoptera, Phasmatodea, and Plecoptera) and the Hemipteroidea (Hemiptera, Phthiraptera, Psocoptera, Thysanoptera, and Zoraptera). Except for the developed wings and genitalia, there is a strong morphological resemblance between larvae and adults (although habitats and biology may differ greatly).

Exoskeleton

Svend O. Andersen
Copenhagen University

The exoskeleton is noncellular material that is located on top of the epidermal cell layer and constitutes the outermost part of the integument. The local properties and appearance of the exoskeleton are highly variable, and nearly all visible features of an insect result from the exoskeleton. The exoskeleton serves as a barrier between the interior of the insect and the environment, preventing desiccation and the penetration of microorganisms. Muscles governing the insect's movements are attached to the exoskeleton.

Although the exoskeleton is a continuous structure, its mechanical properties differ from region to region. Sometimes the transition between regions is gradual, but often it is quite abrupt; pliant and elastic regions can thus border on hard and heavily sclerotized regions. Most exoskeletal regions of soft-bodied larvae, such as larvae of moths and flies, are soft and pliant, and only restricted regions of their exoskeletons are hard and stiff, such as legs, head capsule, and mandibles. Most of the body surface of adult, winged insects is covered by a stiff exocuticle, which can be somewhat flexible and bendable but also serves as a hard protective armor. The exoskeleton covering the dorsal abdomen of many beetle species is thin and easily flexed, whereas the ventral abdominal exoskeleton of the same animals is hard and resistant. The mechanical properties of all exoskeletal regions are precisely adapted to be optimal for the lifestyle of the insect.

FORMATION OF THE EXOSKELETON

The exoskeleton is produced and modified by the epidermal cell layer, and each cell in the epidermis must have the necessary information for producing and depositing the right amount of the right cuticular components at the right time; some of them will later have to modify the secreted products to give a mature material. The timing of the various events is often hormonally controlled, but the quantitative information on how much to produce must be inherent in individual epidermal cells.

A new exocuticle is produced at each molt. A thin, lipid-rich epicuticle is initially secreted from the epidermal cells and deposited beneath the old cuticle, followed by secretion of a thicker procuticle, consisting of chitin and proteins. To allow growth, the total surface area of the new cuticle is larger than that of the old one, and expansion and stretching of the new cuticle take place during and after emergence from the old cuticle (exuvium). Some exoskeletal regions, such as the head capsule, mouthparts, and spines, may be sclerotized before ecdysis; this will aid emergence from the old cuticle. These regions cannot be further expanded but will keep their pre-ecdysial size and shape. Other exoskeletal regions are soft and pliant at ecdysis and are sclerotized soon after emergence when cuticular expansion is complete; as soon as the sclerotization process has started, these regions are irreversibly locked in their new shape.

Sclerotization not only makes the exoskeleton harder and stiffer, it also makes the proteins inextractable and more resistant to enzymatic digestion. Before sclerotization, the exoskeletal proteins are bound to each other and to chitin by various noncovalent links, such as electrostatic interactions, hydrogen bonds, and hydrophobic interactions. Such links can be weakened by changes in pH and ionic strength, making the cuticle more pliant, because displacements of the cuticular components will be easier. During the sclerotization process the proteins are linked firmly to each other, polymerized sclerotizing material fills the voids between proteins and chitin molecules, the cuticle is dehydrated, and deformations of the material will be more difficult.

The sclerotization precursors, *N*-acetyldopamine (NADA) and *N*-β-alanyldopamine (NBAD), are synthesized from tyrosine in the epidermal cells. The tyrosine molecules are transformed by decarboxylation and hydroxylation to

dopamine, which is acylated to NADA and NBAD. These precursors are secreted from the epidermal cells into the cuticular matrix, where they encounter enzymes (phenoloxidases), which oxidize them to the corresponding orthoquinones. Oxidases of different types (tyrosinases, laccases, peroxidases) have been reported and characterized from cuticle. The quinones produced are highly reactive; they will react spontaneously with histidine and lysine residues in the matrix proteins, resulting in cross-links between neighboring proteins, and they will also react with each other, resulting in complex phenolic polymer mixtures. Depending on the precise reaction conditions, the exoskeleton may remain colorless, or a lighter or darker brown coloration may appear during sclerotization.

The water content of the exoskeleton decreases during incorporation of the sclerotizing precursors into the matrix, probably from a decrease in the number of positively charged amino acid residues in the cuticular proteins, which makes the matrix proteins less hydrophilic. Exclusion of water from the intracuticular voids from accumulation of polymerized material also presumably contributes to dehydration of the exoskeletal material. Often only the exocuticular layer of the sclerites is sclerotized, but in some insects the sclerotization process continues for extended periods after ecdysis, resulting in sclerotization of parts of the endocuticle, although to a lesser extent than the exocuticle.

Both the loss in cuticular water content and the formation of cross-links between proteins contribute to a stabilization of the exoskeletal material. The amounts of sclerotizing material incorporated into the various exoskeletal regions varies from less than 1% to more than 10% of cuticular dry weight. These differences are assumed to be responsible for most of the variation in hardness and stiffness of the various exoskeletal regions. Exocuticle tends to be harder and more difficult to deform than endocuticle, presumably because of more extensive sclerotization. The endocuticular layer will tend to be compressed when a piece of exoskeleton is bent, whereas the stiffer exocuticle will be little deformed, although it will be in tension.

MUSCLE ATTACHMENTS

The muscles that act on the exoskeleton are connected to the basal surface of the epidermal cells by means of desmosomes. The muscular forces are transferred through the cells by a rich array of microtubules, running in parallel from the basal to the apical surface of the cells, where they attach to tonofilaments stretching into the cuticular material. The muscles are often attached to infoldings of the exoskeleton, the apodemes, which can stretch deep into the body of the insect, allowing larger muscles to act on the same skeletal region.

ELASTIC EXOSKELETONS

Some small exoskeletal regions are characterized by a rubberlike elasticity; they can undergo considerable deformation when exposed to mechanical stresses and return to their original shape when unstressed. The amount of energy used for deformation is almost completely recovered during relaxation. Its elasticity is the result of the matrix protein resilin. Resilin-containing ligaments are used for energy storage when a fast release of mechanical energy is needed: for example, in the flight system of insects and in the jumping systems of fleas and click beetles. Most resilin-containing ligaments contain chitin microfibrils, making them inextensible, but readily flexible, but there are some ligaments that consist of nearly pure resilin and are devoid of chitin. Such ligaments can be reversibly stretched to three to four times their unstrained length before breaking. The protein chains in resilin are cross-linked by a mechanism different from that used for the solid cuticle; the chains are linked together by covalent bonds formed between side chains of tyrosine residues during the secretion of soluble resilin from the epidermal cells. The elastic properties of the cross-linked material are due to the flexibility and random coiling of the chain segments between cross-links.

PLASTICIZATION

Sometimes the mechanical properties of the exoskeleton can be changed rapidly and reversibly. In bloodsucking bugs (e.g., nymphs of *Rhodnius prolixus*), the abdominal cuticle is stiff and inextensible before a blood meal. When a meal is initiated, the abdominal cuticle is plasticized, enabling the animal to gorge itself with a volume of blood 10 to 12 times larger than the total volume of the animal before the meal. To do this, stretch receptors send nerve impulses via the central nervous system to axons terminating in the abdominal epidermis. A neurohormone is released from these nerve endings, and the epidermal cells respond by effecting a slight decrease in intracuticular pH. The water content of the abdominal cuticle increases simultaneously, probably owing to the pH change, and the interactions between cuticular proteins decrease, resulting in increased plasticity of the cuticular material.

To facilitate emergence from the old cuticle during ecdysis, the stretchability of the new, pharate cuticle may be temporarily increased to make it easier for the animal to escape from the rather stiff exuvium and facilitate expansion of the new cuticle after emergence. In the tobacco horn worm *Manduca sexta,* and probably in many other insects, the plasticization of the pharate adult cuticle is triggered by release of eclosion hormone into the hemolymph. As in *Rhodnius* nymphal abdominal cuticle, the plasticization of *Manduca* pharate cuticle at emergence is probably due to an intracuticular pH decrease in combination with increased hydration.

Newly emerged blowflies, which must dig free of the soil before they can expand to their proper size, have a relatively stiff cuticle until they have reached the surface and can begin to swallow air for expansion. For a brief period, their cuticle is plasticized, from release of the neurohormone bursicon. This hormone also plays a role in initiating sclerotization and

deposition of endocuticle in the blowflies and probably in other insects.

VISCOELASTICITY

Most types of cuticle are more or less viscoelastic; when exposed to a deforming force for extended periods, they will suffer a slight, time-dependent elongation, and recovery after release of the force may not be complete. A special type of highly stretchable, viscoelastic cuticle is found in the abdominal intersegmental membranes of sexually mature female locusts. This stretchability allows elongation of the abdomen necessary for depositing eggs in the soil at a sufficient depth. The membranes in both male and female locusts are soft and pliable, but not very stretchable, as long as the animals are sexually immature. When sexual maturation is initiated in the females by resumed production of juvenile hormone, the organization of the chitin microfibrils in the intersegmental membranes changes from a helicoidal arrangement to one that is perpendicular to the long axis of the animal; at the same time, special hydrophilic proteins are deposited in the membranes. The fully mature intersegmental membranes stretch when loaded, but recover only partly when the load is released. When reloaded with the same load as before, they elongate significantly more than during the first load, and by repeated application of even small loading forces the females can elongate the membranes to about 10 to 15 times their relaxed length, corresponding to a threefold elongation of the total abdomen. Such stretching enables the female locust to deposit eggs in the soil to a depth of 10 to 12 cm.

METAL REINFORCEMENT

The mandibles of plant-eating insects are often extremely hard and abrasion resistant because of incorporation of metals, such as zinc and manganese, in the cuticular matrix of the cutting edge of the mandibles. Up to 5% zinc has been registered in some mandibles.

PROTECTIVE BARRIER

The exoskeleton serves also as a water-impermeable barrier, protecting the insect against desiccation. The main part of the barrier is located in the wax-covered epicuticle.

An important function for the exoskeleton is to act as a barrier preventing microorganisms from access to interior of the animal. Soft, pliant cuticles are more easily damaged and penetrated by microorganisms than the sclerotized regions, but they contain a defense system of inactive precursors of phenoloxidases. When the cuticle is damaged, these precursors are activated by limited proteolysis to active phenoloxidases, which will oxidize tyrosine and other phenols to highly reactive quinones. The reaction products are toxic for microorganisms, and they will close minor wounds in the cuticular surface.

COLORATION

Often the result of various pigments present in granules in the epidermal cells, the colors of insects can also be due to colored material in the cuticle, diffraction or interference of light caused by special cuticular structures, or the Tyndall effect.

A brown coloration in the cuticle develops often during sclerotization of the exocuticle, especially when NBAD is used as precursor for the sclerotization agents, whereas uncolored and transparent cuticles results when NADA is the sole sclerotization precursor. The intensity of the color varies from very light brown over tan to a very dark brown, which can be difficult to discern from the genuinely black cuticles that contain melanins. Melanins are formed when free tyrosine or dopamine is oxidized to orthoquinones, which readily polymerize to complex, black, intractable materials. Melanins either are diffusely distributed in the cuticle or occur in discrete, membrane-bounded granules.

Structural colors of the cuticle from interference of light can be caused by regularly spaced layers in the cuticle in, for example, the cornea of the compound eyes in many flies. Light reflected from the individual layers will interfere to give colors varying with the angle of reflection. Structural colors may also be produced by diffraction of light by regularly spaced microscopic structures on the cuticular surface. The brilliant colors of many beetle species are due to such surface diffraction.

Light scattered by sufficiently small particles ($< 0.7\ \mu m$ in diameter) looks blue because of the Tyndall effect, as in the blue colors of many dragonflies. The light-scattering particles may be located in the epidermal cells underlying a transparent cuticle, or the light may be scattered by a very fine bloom of wax filaments deposited on the cuticular surface after emergence.

SENSE ORGANS

Several exoskeletal structures are involved in sense perception. Various types of mechanoreceptor are involved in registering the exact position of, and deformation in, the various exoskeletal regions and body parts, movements of surrounding objects, currents of air or water, vibrations in the substrate, and sound oscillations. Chemoreceptors are involved in registering and discerning the presence of various chemical substances; these receptors can be contact chemoreceptors (taste) or olfactory chemoreceptors (smell). Many of the sense organs take the form of setae (bristles, hairs, etc.), which are sensilla consisting of an elongated cuticular structure in connection with the sensory cell(s). A trichogen cell in the epidermis produces a more or less elongated structure, which can be variously shaped, often as a flexible hair, a rigid spine, or an arched dome. The hairs are usually connected to the surrounding cuticle by a joint, flexible membrane, and the sensory cell responds to deformations of the cutaneous membrane. The campaniform sensilla are rigidly connected

to the surrounding cuticle, and they respond to tensions in the dome shaped cuticle.

The cuticle covering the elongated sensilla of olfactory chemoreceptors contains numerous narrow pores, allowing access for the airborne stimulatory molecules into the interior of the sensilla, where they come in contact with and stimulate the dendritic membrane of the sensory cell. The contact chemoreceptors are constructed according to the same principle, but they often contain a single larger pore through which molecules can get access to the sensory cell.

A characteristic feature of the visual system in insects is that both the compound eyes and the single eyes (ocelli) are covered by a transparent cuticle, the lens or cornea, through which light reaches the light-sensitive cells. Both the corneal cuticles and the cuticles used for construction of the other sense organs are constructed according to the common cuticular plan.

See Also the Following Articles

Chemoreception • Coloration • Cuticle • Mechanoreception • Molting

Further Reading

Andersen, S. O. (1985). Sclerotization and tanning of the cuticle. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut, and L. I. Gilbert, eds.), Vol. 3, Chap. 2. Pergamon Press, Oxford, U.K.

Bereiter-Hahn, J., Matoltsy, A. G., and Richards, K. S. (eds.) (1984). "Biology of the Integument," Vol. 1, "Invertebrates." Springer-Verlag, Berlin. (See especially Chaps. 27–35.)

Blomquist, G. J., and Dillwith, J. W. (1985). Cuticular lipids. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," (G. A. Kerkut, and L. I. Gilbert, eds.), Vol. 3, Chap. 4. Pergamon Press, Oxford, U.K.

Hepburn, H. R. (ed.) (1976). "The Insect Integument." Elsevier, Amsterdam.

Hepburn, H. R. (1985). Structure of the integument. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," (G. A. Kerkut, and L. I. Gilbert, eds.), Vol. 3, Chap. 1. Pergamon Press, Oxford, U.K.

Kramer, K. J., Dziadik-Turner, C., and Koga, D. (1985). Chitin metabolism in insects. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," (G. A. Kerkut, and L. I. Gilbert, eds.), Vol. 3, Chap. 3. Pergamon Press, Oxford, U.K.

Neville, A. C. (1975). "Biology of the Arthropod Cuticle." Springer-Verlag, Berlin.

Wigglesworth, V. B. (1972). "The Principles of Insect Physiology." 7th ed. Chapman & Hall, London.

Extension Entomology

Nancy C. Hinkle and Beverly Sparks
University of Georgia

Linda J. Mason
Purdue University

Karen M. Vail
University of Tennessee, Knoxville

Land-grant institutions have teaching, research, and outreach (service) as their missions. Cooperative extension is the university's face to the state's citizenry, just as teaching faculty are the university's face to students and research faculty are the component visible to their academic peers around the world. In linking the university to the public, extension entomologists translate research results into practical applications and convey them to end users, while simultaneously apprising university researchers of real-world needs.

HISTORY OF COOPERATIVE EXTENSION

The Smith–Lever Act created the Cooperative Extension Service in 1914. However, several key legislative acts preceded Smith–Lever and these acts were critical in leading to the formation of the Cooperative Extension Service. The Morrill Acts of 1862 and 1890 (also known as the Land-Grant Acts) authorized that each state be granted 30,000 acres (12,141 ha) of public land for each senator and representative of the states in Congress at that time. Revenue generated from these lands was to be used for endowment, support, and maintenance of at least one college to teach fields of study related to agriculture and mechanical arts "to promote the liberal and practical education of the industrial classes in the several pursuits and professions in life."

The second Morrill Act provided funding to establish the 1890 land-grant institutions. Under the conditions of legal racial separation in the South during the late 1800s, black students were not permitted to attend the original land-grant institutions. Passage of the second Morrill Act expanded the 1862 system of land-grant universities to include historically black institutions.

The Hatch Act is often likened to a sturdy bridge between the Morrill Acts and the Smith–Lever Act. Signed on March 2, 1887, the Hatch Act gave this nation its network of agricultural experiment stations. The Hatch Act states that experiment stations should "conduct original and other research, investigations and experiments bearing directly on and contributing to the establishment and maintenance of a permanent and effective agricultural industry." These experiment stations were charged with conducting research for effective and efficient production of food and fiber. Research findings from systems across the country revised farming methods to fit America's diverse geography, making farmers more productive.

The federal–state research partnerships funded through the Hatch Act supported research that addressed "hunger and poverty and the drudgery of subsistence agriculture production." From its inception, research stations created by the Hatch Act were designed to meet the needs of agriculture in the areas in which the experiment stations were located, but the research generated often has far-reaching applications. In fact, research supported by Hatch Act funding benefits every person in the United States and much of the world.

The Smith–Lever Act of 1914 created the Cooperative Extension Service. Senator Hoke Smith (Georgia) and Representative Frank Lever (South Carolina) introduced this act "to aid in diffusing among the people of the United States useful and practical information on subjects relating to agriculture and home economics, and to encourage application of the same." This legislation created a partnership between the U.S. Department of Agriculture, the land-grant universities, and the 1890 institutions that was charged to provide outreach education to the citizens of each state. In practical terms this legislation created the ability for representatives of land-grant universities and 1890 institutions to work with farm families on their farms to introduce research-based advances in agriculture, home economics, and other fields.

Today, this educational system includes professionals in each of America's land-grant universities (in the 50 U.S. states, Puerto Rico, the Virgin Islands, Guam, Northern Marianas, American Samoa, Micronesia, and the District of Columbia) and in 16 1890 historically black, land-grant universities plus Tuskegee University.

The Cooperative Extension Service is a partnership between the U.S. Department of Agriculture, the land-grant institutions, and the 1890 institutions. Legislation in various states has also enabled local governments in the nation's counties to become a fourth legal partner in this educational endeavor. Organization of the Cooperative Extension Service at national, international, state, regional, and county levels is discussed below.

ORGANIZATION AT THE NATIONAL LEVEL

At the national level, the Cooperative Extension Service is an integral part of the Cooperative State Research, Education, and Extension Service (CSREES). The CSREES is a national research and education network that links education programs of the U.S. Department of Agriculture with land-grant institutions, with 1890 institutions, with agricultural experiment stations, with Cooperative Extension Services, with schools of forestry, and with colleges of agriculture, colleges of veterinary medicine, and colleges of human sciences. CSREES, in cooperation with all these partners, develops and supports research and extension programs in the food and agricultural sciences and related environmental and human sciences. Examples of some program areas in which CSREES and its partners are currently working include improving agricultural productivity; protecting animal and plant health; promoting human nutrition and health; strengthening children, youth, and families; and revitalizing rural American communities.

CSREES serves as a critical connection between research and extension. CSREES works with extension educators on identifying and communicating agricultural, environmental, and community problems (Table I). These problems are then relayed to researchers at the land-grant institutions and agricultural experiment stations. Working together, these partners initiate and stimulate new research that provides solutions to real-world problems.

TABLE I Extension Educational Programs within CSREES

1. Provide model education programs on food safety; sustainable agriculture; water quality; children, youth, and families; health; environmental stewardship; and community economic development in all 50 states, all U.S. territories, and the District of Columbia.
2. Represent over 9600 local extension agents working in 3150 counties.
3. Engage 5.6 million youth in 4-H programs for personal development and community service.
4. Involve 3 million trained volunteers who work with outreach education programs nationwide.
5. Provide farm safety education programs in all 50 states and Puerto Rico.
6. Provide pesticide applicator programs that train over half a million people each year in safe and environmentally sound pesticide use.
7. Participate in international education programs taught by over 200 extension professionals in 17 countries.

INTERNATIONAL EXTENSION ORGANIZATION

Although some industrialized countries have attempted to reduce costs by delegating extension responsibilities to the private sector, with varying degrees of success, most developing countries have modeled their extension systems on the U.S. paradigm. Frequently, extension outreach in third-world nations is funded by such agencies as the U.S. Agency for International Development, the World Bank, and the United Nations' Food and Agriculture Organization. These technology transfer programs typically are most effective when closely linked with university research programs, permitting rapid transmittal and adoption of research results. Alternatively, the outreach may be handled by such governmental entities as the ministry of agriculture.

STATE, REGIONAL, AND COUNTY ORGANIZATION

Organizational structure of the Cooperative Extension Services varies greatly in size from state to state. In general, leadership of Cooperative Extension Services within each state is the responsibility of the dean and/or director of the agricultural college of the land-grant university and/or 1890 institution within each state. These directors provide leadership to an administrative staff that often includes associate and/or assistant deans of extension, directors of county operations, department heads and/or extension program leaders within various scientific disciplines, and directors of units that support programming. In states with numerous counties, the organizational structure often

includes regional administrators that serve under the Cooperative Extension director.

The extension director, along with administrators within each scientific discipline, also oversees a faculty of extension specialists. These specialists serve as educational resources to county agents and their clientele in various subject matter/disciplines. Extension specialists are most often administratively based within their academic department and may be located on the main campus of the land-grant university, at experiment stations or, occasionally, within county extension offices.

The organization of an extension office at the county level also varies greatly from state to state and county to county. In areas of the United States that have very low populations there are county offices with only one county extension agent, responsible for the administration and delivery of programming in all subject areas. In more populated areas, county extension offices often house several agents with program areas divided among agents.

Cooperative Extension at the County Level

The interface between extension entomologists on the state and county staff tends to follow a similar model in the majority of states. At the state level the positions are usually tied to a university academic unit and filled with a Ph.D.-level entomologist. These persons would either be full-time extension or have a partial extension appointment combined with other duties, including teaching and/or research. At the county level, job responsibilities and qualifications may vary; however, some common models are evident. County-level extension entomologists usually are termed agents, advisors, or educators, terms used synonymously within this article. Additionally, entomology positions typically fall under either Agriculture and Natural Resources (ANR) or 4-H and Youth programmatic areas.

County agents initially were itinerant teachers hired for their practical farm and home experiences. Today extension educators are highly trained, often specialized, professionals. Generally extension educator positions require a master's degree or a bachelor's degree with significant related experience. At least one degree in a discipline related to the specialty area is usually required. Specialty areas may include entomology but could be any related field such as botany, plant pathology, agronomy, horticulture, general agriculture, soil science, or animal science. Forty-one percent of educators have one-half or more of their job assignments in agriculture. County agents with agricultural backgrounds are expanding their roles to serve urban/suburban clientele as programs such as Master Gardeners become more successful.

County-based extension entomologists, whether ANR or 4-H and youth based, need to be highly skilled, technically based professionals with excellent people, writing, and presentation skills; multitasking abilities; and willingness to work flexible hours.

Extension Specialists

Although extension agents are located in the counties and are expected to have broad expertise, extension specialists typically are housed on university campuses and specialize in discipline areas. The position of Cooperative Extension Specialist is one of statewide leadership toward university colleagues, agricultural industries, consumers, youth, policymakers, and governmental and other agencies. The specialist keeps campus and county colleagues and clientele apprised of emerging issues and research findings and directions, works with them to develop applications of research knowledge to specific problems, and provides educational leadership and technical information support for county staff/clientele.

A Cooperative Extension Specialist is a primary liaison with university research units, providing leadership, facilitating teamwork, developing collaborative relationships with colleagues, and ensuring appropriate external input into research and educational program planning by the Agricultural Experiment Station (AES) and Cooperative Extension. Ideally, the AES–extension relationship is a seamless continuum, with extension identifying timely research opportunities to AES colleagues and conveying research results to clientele. The specialist also defines and considers needs of relevant clientele groups in planning, development, and execution of applied research and education programs.

EXTENSION TEACHING Specialists provide leadership for nonformal education of end users, intermediate users, and the public. In addition to directing planning and coordination of statewide extension education and information transfer programs related to areas of responsibility, specialists facilitate coordination of work group activities with appropriate internal and external organizations. Specialists serve as scientific and technical resources on work groups, providing disciplinary input and perspective.

Specialists' education efforts are directed toward four main clientele groups—county agents, producer/professional groups, public/private agencies, and the general public. They educate and serve as teaching resources in areas of responsibility for extension county/area personnel via individual consultations, conferences, and workshops. In addition to formal teaching at training sessions, specialists provide one-on-one consultation in person, electronically, and by telephone.

Specialists prepare and evaluate educational materials, such as publications, newsletters, slide sets, videotapes, computer software, and other learning aids, to extend subject matter information to county staff and the public sector. Because county agents are the main public interface, specialists focus on "training the trainers," developing county skills to serve clientele. In addition, specialists assist agents in customizing materials for their clientele and disseminate industry-appropriate articles through relevant channels.

Although term-length, resident classroom instruction is not the norm for full-time extension specialists, they may

participate in teaching programs (via lectures and seminars) of relevant campus-based courses. Doing so permits specialists to serve as models for students developing careers in extension while fostering interactions with undergraduate and graduate students, providing these groups a vision of the third function of a university. In addition, specialists train graduate students, serve on advisory committees, and participate in other graduate education activities.

APPLIED RESEARCH AND OTHER CREATIVE WORK Like their AES counterparts, specialists are expected to plan, conduct, and publish results of applied research/creative activity directed toward resolution of important issues or problems, independently or, more commonly, in collaboration with other research and extension personnel (including county agents). In addition, specialists provide leadership for planning and coordination of applied research activities related to areas of responsibility with departmental and other researchers, encouraging interdisciplinary collaboration and work-group participation.

Research and creative activity include synthesis and interpretation of extant knowledge, an integral aspect of the Smith–Lever mission. Extension fulfills its role by assisting in formulating policy and establishing regulatory standards and mechanisms, providing science-based information upon which policy decisions are made, and serving as the university's liaison with nongovernmental organizations and historically underserved groups.

PROFESSIONAL ACTIVITY Specialists participate in appropriate professional societies and educational organizations and serve on state, regional, national, and international committees; review panels; and editorial boards. Enhanced professional stature accrues to the reputations of specialists' home institutions in addition to reflecting positively on CSREES.

UNIVERSITY AND PUBLIC SERVICE As good university citizens, specialists participate in activities of committees within the department, college, campus, and other university entities. Serving as liaisons, specialists respond to regulatory and state and federal agencies, external groups, industry organizations, and the media on issues related to areas of expertise, as well as representing the university to producer groups and other organizations.

The value of Cooperative Extension is its ability to design, develop, and deliver educational programs that meet the unique needs of people as they adjust to change. The Smith–Lever Act specifies that the main function of Cooperative Extension is synthesis of existing knowledge, ancillary to creation of new knowledge. The complementarity of AES and Cooperative Extension is demonstrated not only in that extension takes AES's discoveries to the people but also in extension's conveying the needs of the citizenry to AES researchers, ensuring that these issues are addressed.

See Also the Following Articles

Agricultural Entomology • Regulatory Entomology

Further Reading

Bartholomew, H. M., and Smith, K. L. (1990). Stresses of multicounty agent positions. *J. Extension* **28(4),** 1–6.

Bushaw, D. W. (1996). The scholarship of extension. *J. Extension* **34(4),** 5–8.

Cooper, A. W., and Graham, D. L. (2001). Competencies needed to be successful county agents and county supervisors. *J. Extension* **39(1),** 1–11.

Gray, M. E., and Steffey, K. L. (1998). Status of extension entomology programs: A national assessment. *Am. Entomol.* **44,** 9–13.

Jones, M. P. (1944). Extension entomology activities in wartime. *J. Econ. Entomol.* **37(3),** 354–356.

Jones, M. P. (1950). Extension entomology. *J. Econ. Entomol.* **43(5),** 736–739.

Lincoln, C., and Blair, B. D. (1977). Extension entomology: A critique. *Annu. Rev. Entomol.* **22,** 139–155.

Palm, C. E. (1954). The growing responsibility of entomology to human welfare. *J. Econ. Entomol.* **47(1),** 1–6.

Patton, M. Q. (1986). To educate a people. *J. Extension* **24,** 21–22.

Steffey, K. L., and Gray, M. E. (1992). Extension–research synergism: Enhancing the continuum from discovery to delivery. *Am. Entomol.* **38,** 204–205.

Eyes and Vision

Michael F. Land

University of Sussex, Brighton

Insect eyes are of two basic types: compound (or multifaceted) and simple (or single chambered). In adults, the principal organs of sight are nearly always compound eyes, although simple eyes—often quite good ones—are frequently present in immatures. Despite the major differences in their form and construction, compound and simple eyes perform essentially the same job of splitting up the incoming light according to its direction of origin (Fig. 1). Compound eyes are of two distinct and optically different kinds: apposition eyes, in which each receptor cluster has its own lens, and superposition eyes, in which the image at any point on the retina is the product of many lenses.

APPOSITION EYES

History of Insect Optics

The facets of compound eyes of insects are too small to be resolved with the naked eye, and it required the invention of the microscope in the 17th century before they could be properly depicted. The process of working out how compound eyes functioned took more than 2 centuries from Robert Hooke's first drawing of "The Grey Drone Fly" (probably a male horse fly) in his *Micrographia* of 1665 to

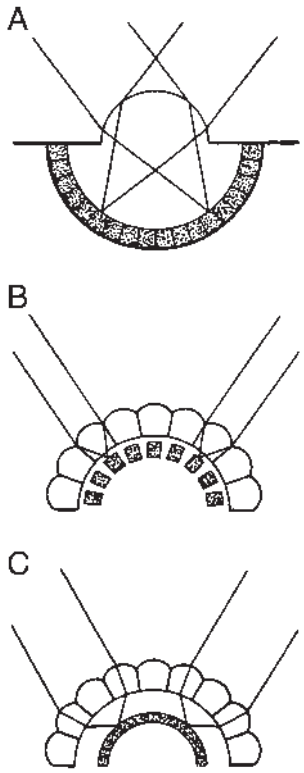

FIGURE 1 The three types of eye found in insects. (A) Simple, or single-chambered, (B) apposition compound, (C) superposition compound. The receptors are shown stippled. (Reproduced, with permission, from Land and Nilsson, 2002.)

the essentially modern account by Sigmund Exner in 1891. The first person to look through the optical array of an insect eye was Antoni van Leeuwenhoek, and his observations caused a controversy that was not fully resolved until the 1960s. The following quotation comes from a letter from Leeuwenhoek to the Royal Society of London, which was published in 1695.

Last summer I looked at an insect's cornea through my microscope. The cornea was mounted at some larger distance from the objective as it was usually done when observing small objects. Then I moved the burning flame of a candle up and down at such a distance from the cornea that the candle shed its light through it. What I observed by looking into the microscope were the inverted images of the burning flame: not one image, but some hundred images. As small as they were, I could see them all moving.

Evidently, each facet of the eye (at least in apposition eyes) does produce an inverted image, even though the geometry of the eye as a whole dictates that the overall image is erect (Fig. 1). What, then, does the insect see? Do the receptors (typically eight) beneath each lens resolve the inverted images, or do they just indicate the average intensity across the field of view of the ommatidium? (An ommatidium is the "unit" of a compound eye, consisting of the lens, receptors, and associated structures. See Fig. 2A).

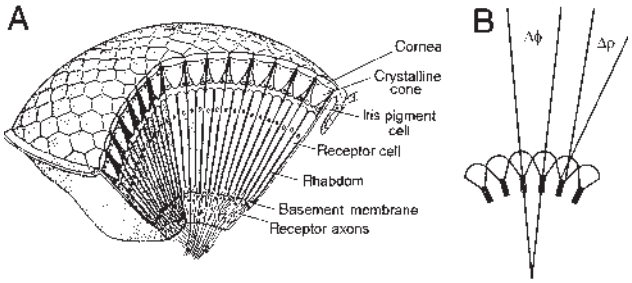

FIGURE 2 (A) Basic structure of an apposition eye, showing its construction from ommatidial elements. (B) Definitions of the interommatidial angle, Δφ, and rhabdom acceptance angle, Δρ. (Reproduced, with permission, from Land and Nilsson, 2002.)

Remarkably, the answer depends on the animal. By the 1870s histological studies had shown that in most apposition eyes the eight receptor cells in each ommatidium contribute to a single radial structure, known as a rhabdom (Greek for rod; Figs. 2 and 3). Much later, in the 1950s, this material was found to be made up of photoreceptive membrane covering large numbers of long narrow microvilli, but even by the time that Exner wrote his monograph in 1891 it was

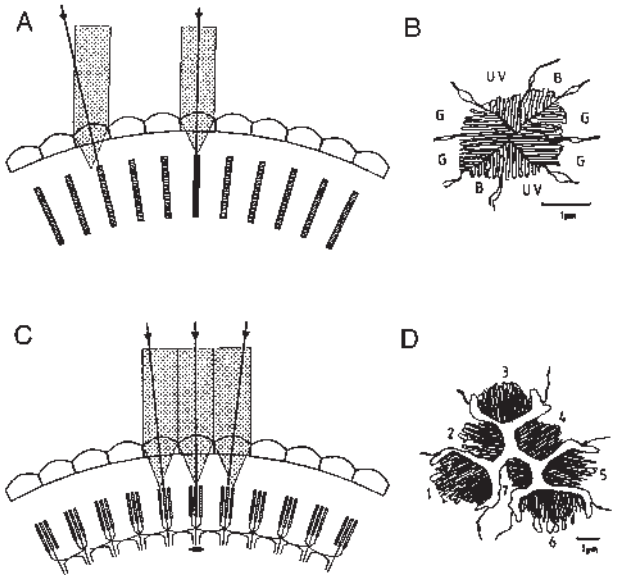

FIGURE 3 Optical comparison of an apposition eye (A,B) and a neural superposition eye (C,D). In an apposition eye each rhabdom (hatched) views light from a slightly different direction (arrows), and the rhabdoms (B), although made up from eight receptors, have a fused structure that acts as a single light guide. UV, B, and G indicate the receptor elements that respond to ultraviolet, blue, and green in an ommatidium from the eye of a worker bee. In neural superposition eyes, light from a single direction is imaged onto different rhabdomeres in adjacent ommatidia (C). The axons from all receptors imaging the same point collect together in the first synaptic layer (the lamina, Fig. 5) so that here the image has the same structure as in an ordinary apposition eye. The section (D) shows the arrangement of the separated rhabdomeres in an ommatidium from a fly. The six outer rhabdomeres (1–6) all send axons to different adjacent laminar "cartridges" (as in C). The central pair (7 overlying 8) bypass the lamina and go straight to the next ganglion, the medulla. (Reproduced, with permission, from Land and Nilsson, 2002.)

clear that the rhabdom was the structure sensitive to light. Optically, each ommatidium works as follows. The inverted image that Leeuwenhoek saw is focused onto the distal tip of the rhabdom. Having a slightly higher refractive index than its surroundings, the rhabdom behaves as a light guide, so that the light that enters its distal tip travels down the structure, trapped by total internal reflection. Any spatial information in the image that enters the rhabdom tip is lost, scrambled by the multiple reflections within the light guide, so that the rhabdom itself acts as a photocell that averages all the light that enters it. Its field of view is defined, in geometric terms, by the angle that the tip subtends at the nodal point of the corneal lens ($\Delta\rho$; Fig. 2B), and in a typical apposition eye this acceptance angle is approximately the same as the angle between the ommatidial axes (the interommatidial angle, $\Delta\varphi$ Fig. 2B). Thus the field of view of one rhabdom abuts (or "apposes," hence the name) the field of its neighbor, producing an overall erect image made up of a mosaic of adjacent fields of view.

Although the eight receptors that contribute to the rhabdom share the same visual field, it does not mean that they supply the same information. The labels UV, B, and G on the cross section of a bee rhabdom in Fig. 3B indicate the regions of the spectrum that the cells respond to best. Most insects have trichromatic color vision, just as humans do, although their visible spectrum is shifted toward shorter wavelengths compared with ours (Fig. 4). Some butterflies and dragonflies have four-color vision.

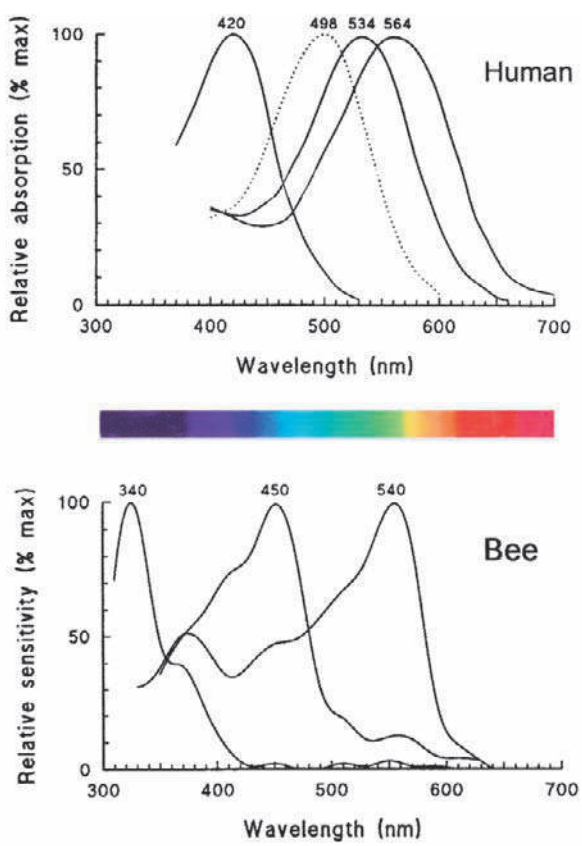

FIGURE 4 The spectral sensitivity curves for the three human cone mechanisms (and rods, dotted) and the corresponding three curves for a bee. The spectrum shows the colors as they appear to human eyes. (Reproduced, with permission, from Land and Nilsson, 2002.)

The second feature of the bee rhabdom (Fig. 3B) is that the microvilli making up the structure are arranged in orthogonal sets. It has been known since the work of Karl von Frisch in the 1940s that bees can navigate using the pattern of polarized light in the sky. This capacity arises from the way the photoreceptor molecules are arranged on the microvilli. A geometric consequence of the cylindrical shape of the microvilli is that there will be twice as many light-sensitive chromophore groups of the rhodopsin molecules aligned parallel to the long axis of each microvillus than at right angles to it. This, in turn, means that the receptors respond best to light polarized parallel to this axis. In fact bees use a special dorsal region of the eye (the POL area) to analyze sky polarization; in the rest of the eye the receptors are twisted to abolish polarization sensitivity, so that it does not interfere with color vision. Polarization vision is also used by some insects, such as the water bug *Notonecta,* to detect water surfaces, which polarize light strongly.

The description of apposition optics given above holds for most diurnal insects (e.g., bees, grasshoppers, and dragonflies), but it does not apply to the true (two-winged) flies, the Diptera. Since 1879, when Grenacher observed that the receptors in fly ommatidia have separate photoreceptive structures (rhabdomeres) that do not contribute to a common rhabdom, there had been suspicions that flies might actually be resolving the Leeuwenhoek images. In the focal plane of the lens of a fly ommatidium, the distal tips of the rhabdomeres are separated from each other and form a characteristic pattern (Fig. 3D) that resolves the image into seven parts (there are eight receptors, but the central pair lie one above the other). This raises the obvious question: how are these seven-pixel inverted images welded together to form the overall erect image, if indeed that is what occurs? Kuno Kirschfeld finally solved this conundrum in 1967. It turns out that the angle between the fields of view of adjacent rhabdomeres within an ommatidium (about 1.5° in a blow fly) is identical to the angle between neighboring ommatidial axes. Furthermore, the fields of each of the six peripheral rhabdomeres in one fly ommatidium are aligned, in the space around the fly, with the field of the central rhabdomere of one of the neighboring ommatidia (Fig 3C). Thus, each point in space is viewed by seven rhabdomeres in seven adjacent ommatidia. What does this complicated and seemingly redundant arrangement achieve? To answer this it is necessary to know what happens to the signals from the seven receptors that view the same point, and that turns out to be the most astonishing part of the story. Beneath each ommatidium, the emerging receptor axon bundle undergoes a 180° twist before the individual neurons disperse to nearby regions of the first optic ganglion (the lamina) that correspond to the adjacent ommatidia. The net result of this impressive

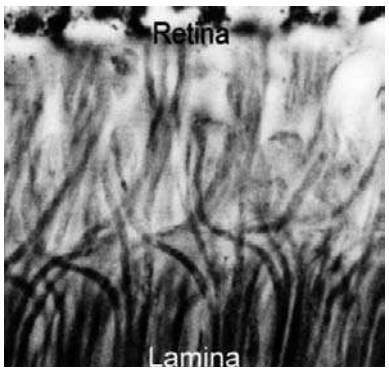

FIGURE 5 The interchange of axons that occurs between retina and lamina of a blow fly *(Calliphora),* which makes possible the neural superposition mechanism of Fig. 3C.

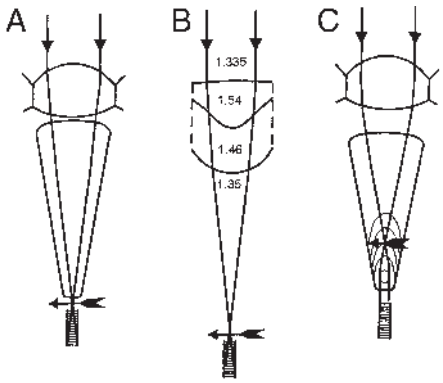

FIGURE 6 Four mechanisms of image formation in apposition eyes. (A) Corneal lens (bee, fly). (B) Multisurface lens (water bugs). (C) Lens/lens-cylinder afocal combination (butterflies). Details in text. (Reproduced, with permission, from Land and Nilsson, 2002.)

feat of neural knitting (Fig. 5) is that all the axons that "look at" the same point in space finish up making connections with the same cells in the lamina. Thus, as far as the lamina is concerned, the image is exactly the same as it would be in a conventional apposition eye, except that the signal, in terms of photon captures, is seven times stronger. One advantage of the extra signal is that it provides flies with a short period at dawn and dusk when they can see well, but when the eyesight of their predators and competitors is less sensitive and so less effective at detecting small objects.

Kirschfeld called this arrangement "neural superposition," because, as in optical superposition (see later), the contributions of a number of ommatidia are superimposed in the final image. One might ask: could the signal not have been made stronger simply by increasing the diameter of the rhabdom in a conventional apposition eye? Indeed it could, but that would mean increasing the rhabdom acceptance angle ($\Delta\rho$; Fig. 2B) at the same time, which in turn would mean a loss of resolution for the eye as a whole. The beauty of the fly solution, and undoubtedly the reason why it evolved, is that it involves no increase in acceptance angle, provided the rhabdomeres are properly aligned. There are strong hints that something like neural superposition occurs in other insect groups (some beetles, earwigs, water bugs, and crane flies) but it is only in the advanced flies that the perfect nearest-neighbors arrangement is known to be achieved.

Imaging Mechanisms

The structures that form the images in the ommatidia of apposition eyes are quite varied (Fig. 6). In terrestrial insects, as in terrestrial vertebrates, the simplest way to produce an image is to make the cornea curved (Fig. 6A). Ordinary spherical-surface optics then apply, and an image is formed about four radii of curvature behind the front face. In aquatic insects such as the water bug *Notonecta,* the external surface of the cornea has little power because of the reduction in refractive index difference (Fig. 6B). It is augmented by two other surfaces, the rear of the lens and an unusually curved interface in the center of the lens whose function may be to correct one of the defects of spherical surfaces—spherical aberration.

The eyes of butterflies, which resemble ordinary apposition eyes in nearly all respects, have an optical system that is subtly different from the arrangement in Fig. 6A. Instead of forming an image at the rhabdom tip, as in the eye of a bee or locust, the image lies within the crystalline cone. The proximal part of the cone contains a very powerful lens cylinder that makes the focused light parallel again, so that it reaches the rhabdom as a beam that just fits the rhabdom (Figs. 6C, and 17). This arrangement, known as afocal apposition because there is no external focus, has much in common with the superposition optical system of moths, to which butterflies are closely related, and will be considered later.

Resolution

For any eye, the resolution of the image seen by the brain is determined by the fineness with which the ommatidial mosaic samples the environment, represented by the interommatidial angle, $\Delta\varphi$ (Fig. 2B), and by the quality of the image received by each rhabdom, represented by the rhabdom acceptance angle $\Delta\rho$ (Fig. 2B). (Although the eight receptors that contribute to each rhabdom usually have different spectral and polarization responses, they all share a common field of view.) In asymmetric eyes (which most are) $\Delta\varphi$ may be different along different axes of the facet array, but for present purposes $\Delta\varphi$ is taken to be the average of the angle measured along each of the three axes of the array. In the central region of a bee eye, $\Delta\varphi$ is about 1.7°. An extensive table of values can be found in a recent review by Land in 1997.

One would expect that apposition eyes would show a rough match between the interommatidial angle and the acceptance angle ($\Delta\rho$) of a single rhabdom, the argument being that no individual rhabdom can resolve detail finer than $\Delta\rho$, so there is no point spacing the directions of view of ommatidia closer than this angle. The acceptance angle $\Delta\rho$

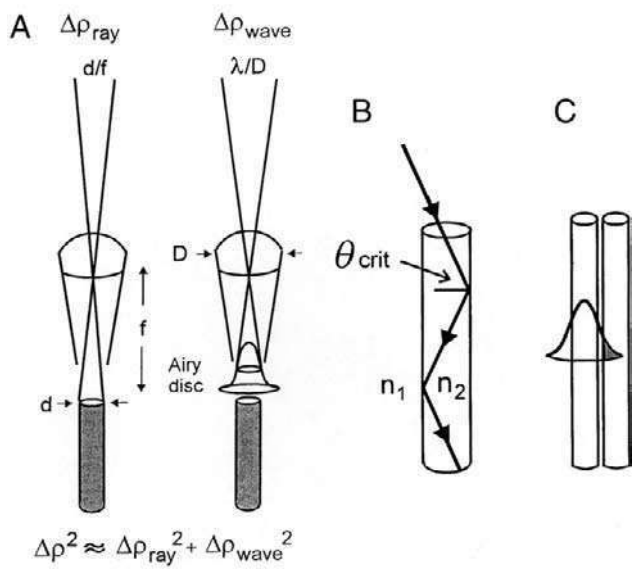

FIGURE 7 (A) The acceptance angle (Δρ) of an ommatidium results from a combination of the Airy diffraction pattern (point-spread function), given by λ/*D* (right), and the geometrical angular width of the rhabdom *(d/f)* at the nodal point of the lens (left). (B) Light is trapped in a rhabdom by total internal reflection, which occurs when the angle the light makes with a normal to the wall is greater than the critical angle, given by $\sin \theta_{crit} = n_1/n_2$, the ratio of the refractive indices outside and inside the rhabdom. A typical rhabdom can trap a cone of light about 22° wide. (C) In narrow light-guiding structures some of the light is actually outside the fiber, and can potentially be caught by adjacent fibers and so spoil resolution.

is actually a combination of the contributions of ray and wave optics (Fig. 7A). Geometrically, $\Delta\rho_{ray}$ is the angle subtended by the rhabdom tip at the nodal point of the facet lens, i.e., the rhabdom diameter divided by the focal length (*d/f* radians). Typical values (for a bee) are 2 μm for *d* and 60 μm for *f*, which makes $\Delta\rho_{ray}$ 0.033 radians, or 1.9°. In wave optics, the limit to image quality is set by diffraction, specifically by the angle subtended by the Airy disk (the diffraction image of a point source), and this is given by λ/*D* radians. If the wavelength (λ) is 0.5 μm and the facet diameter (*D*) is 25 μm, then $\Delta\rho_{wave}$ is 0.02 radians, or 1.1°. To obtain the final value for Δρ, $\Delta\rho_{ray}$ and $\Delta\rho_{wave}$ have to be combined, and unfortunately the proper way of doing this (convolution, taking the wave-guide properties of the rhabdom into account) is very complicated. A simple approximation is given by $\Delta\rho^2 = \Delta\rho_{ray}^2 + \Delta\rho_{wave}^2$. This is adequate for most purposes but tends to overestimate Δρ slightly. Using this approximation, Δρ for the bee data is 2.2°, somewhat larger than Δφ. Typically in light-adapted diurnal insects the ratio of Δρ to Δφ is about 1:1.

The neural superposition eyes of dipterans have an additional constraint, namely that the separation of the tips of the rhabdomeres must match the interommatidial angle. In a house fly, Δφ is about 2°, and with an ommatidial focal length of 70 μm, this means that the tip separation must be 2.4 μm, which does not leave a great deal of room (Fig. 3D). Because narrow light guides, such as rhabdomeres, tend to be "leaky," with a substantial fraction of the light energy outside the guide itself (Fig. 7C), there needs to be an adequate gap between one rhabdomere and the next to prevent cross talk. In flies there is a 1-μm gap between adjacent rhabdomeres (Fig. 3D), which means that the rhabdomeres themselves must be very narrow. They have a distal tip diameter that is also about 1 μm, making them among the narrowest photoreceptors in any animal. In most other respects, however, neural superposition eyes are optically similar to other apposition eyes.

Diffraction and Eye Size

In a short and remarkable article titled "Insect sight and the defining power of compound eyes," published in 1894, Henry Mallock, an optical instrument maker, described insect vision in these terms: "The best of the eyes…would give a picture about as good as if executed in rather coarse wool-work and viewed at a distance of a foot."

Why is insect vision so poor? The problem, as Mallock recognized for the first time, is diffraction. Compound eyes have very small lenses compared with the lenses of single-chambered eyes, and because the size of the diffraction blur circle (the Airy disk) is inversely proportional to aperture diameter, the blur circles are large and the resolution correspondingly poor (Fig. 7A). A 25-μm diameter facet of a bee produces an Airy disc that is just over 1° wide in angular terms. One degree is about the size of a thumbnail at arm's length, so one can imagine a bee's world made up of pixels of about that size. In terms of the acuity of our own eyes (Δφ about 0.01°), this is not very good at all.

Mallock's article goes on to discuss what a compound eye with human resolution would look like, and he came to the astonishing conclusion that it would need to be more than 20 m in diameter, or bigger than a house. The reason for this is clear: the human eye achieves high resolution by having a daylight pupil diameter of 2 mm, 80 times the diameter of a bee lens. For a bee to have the same resolution, diffraction requires that all its lenses would need to have this diameter, and to exploit all the detail in the scene they would need to be spaced at 0.5 arcmin angular intervals, the same as the receptors in our fovea. In a spherical eye, the interommatidial angle (Δφ) is the angle subtended by one lens diameter at the center of the eye (*D/r* radians, where *r* is the eye radius), which gives $r = D/\Delta\varphi$. With Δφ = 0.5 arcmin of arc (0.000145 radians; 1 radian = 57.3° and 1° = 60 arcmin), and *D* = 2 mm, the radius of curvature will be 13.8 m and the diameter twice this. (Kirschfeld has pointed out that this calculation is a little unfair because resolution in the human eye falls off dramatically away from the fovea, to a tenth of its maximum value at 20° from the fovea, and even less farther out. Taking this into account the "human" compound eye can be shrunk in size considerably, to an irreducible 1 m diameter, which still looks very clumsy). Dragonflies seem to approach the limit of what it is possible with an apposition

eye. Their eyes are 8 mm or more in diameter, have up to 30,000 facets each, and resolve about 0.25° in their most acute region. This is still poor compared with what is achievable by any camera-type eye of the same diameter.

The outcome of this discussion is that it is very hard for an apposition eye to improve its resolution; it simply gets too big. Space is thus at a premium; a little extra resolution here must be bought by a bit less there, and for this reason the different visual priorities of arthropods with different lifestyles show up in the distribution of interommatidial angles, and often facet sizes, across the eye.

Sensitivity

The sensitivity of an eye is the ratio of the amount of light received by a single photoreceptor to the amount emitted by the surface that eye is imaging. It can be used to work out the numbers of photons that individual receptors receive, and this determines the way in which the eye will perform under dim light conditions. Sensitivity can be calculated from the formula $S = 0.62\ D^2\ \Delta\rho^2$, where D is the lens diameter and $\Delta\rho$ the rhabdom acceptance angle (Figs. 2B and 7A) (we ignore the effect of receptor length here). Although D is roughly 100 times greater in a human eye than in a bee ommatidium, $\Delta\rho$ is about 100 times smaller (approximately 0.015° compared with 1.5°), so that the value of S is very similar in the bee and the human. Thus, the range of illumination conditions over which an insect with an apposition eye can operate is similar to that of a mammal using its cone system. Mammals can also see at much lower intensities, by pooling the responses of rods over quite large retinal areas (effectively increasing $\Delta\rho$). It is unlikely that pooling occurs to any great extent in insect eyes.

When discussing sensitivity, "adaptation" can have two meanings. Different eyes may be adapted in the evolutionary sense to work permanently under conditions of high or low illumination, e.g., night or day, deep sea or surface. Alternatively, the same eye can be said to be light- or dark-adapted via reversible and temporary changes in its optical anatomy. In both cases, the above equation is the key to interpreting changes and differences.

Light and Dark Adaptation

Temporary light and dark adaptation mechanisms take a number of forms in apposition eyes. Some are illustrated in Fig. 8 and include the following: (A) an iris mechanism just above the distal tip of the rhabdom that restricts the effective value of $\Delta\rho$. In the case of crane flies (Tipulidae), which have an arrangement of six outer and two central rhabdomeres, the iris cuts off the outer six in the light, leaving only the central pair. (B) A "longitudinal pupil" consisting of large numbers of very small pigment granules that move into the region immediately around the rhabdom in the light and withdraw in the dark is a second form. The main effect of this is to absorb the wave-guided light that travels just outside the rhabdom. This is replaced with light within the rhabdom, and this is absorbed in turn, so that light is progressively "bled" out of the rhabdom. This mechanism is particularly important in higher Diptera (house flies, etc.) and in butterflies, and it can work in a matter of seconds. (C) The rhabdom dimensions may themselves change, usually over a period of hours. This mechanism may involve the resynthesis of photoreceptive membrane in the dark and its sequestration in the light. In addition to these changes there are electrical and enzymatic changes in the receptors themselves that alter the gain of transduction and increase response time in the dark.

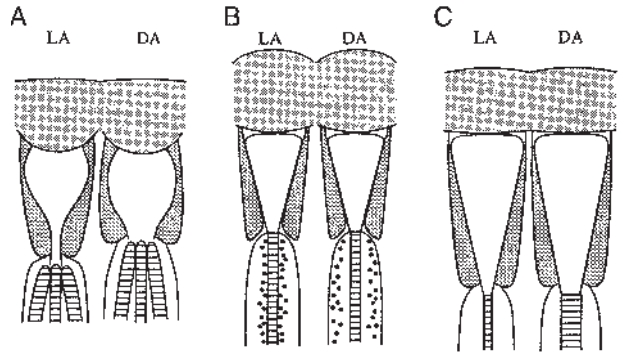

FIGURE 8 Three mechanisms of dark adaptation in apposition eyes of insects (see text). (Reproduced, with permission, from Land and Nilsson, 2002.)

Ecological Variations in Apposition Design

As we have seen, the optical design of apposition eyes means that there is no spare room on the head surface, and what there is needs to be used as efficiently as possible. A survey of the apposition eyes of insects and crustaceans leads to the conclusion that there are three main patterns of acuity distribution that one can identify fairly easily. These are identified in Fig. 9, which illustrates the ecological reasons for these patterns (Figs. 9A–9C) and examples of the distributions themselves (Figs. 9D–9F). Figure 9D shows the pattern related to the motion across the eye encountered in forward locomotion, especially flight. Figure 9E has an "acute zone" associated with predation or sex, these zones sometimes developing into separate components of a double eye. In Fig. 9F the narrow horizontal strip of high resolution is associated with environments such as water surfaces and sand flats, where almost all important activity takes place around the horizon.

THE FORWARD FLIGHT PATTERN When an animal is moving through the world, the objects in the world appear to move backward across the eye. Objects to the sides move faster than those in front, and there is a point in the direction of the animal's travel (the "focus of expansion") where there is no image motion. Objects farther away move more slowly

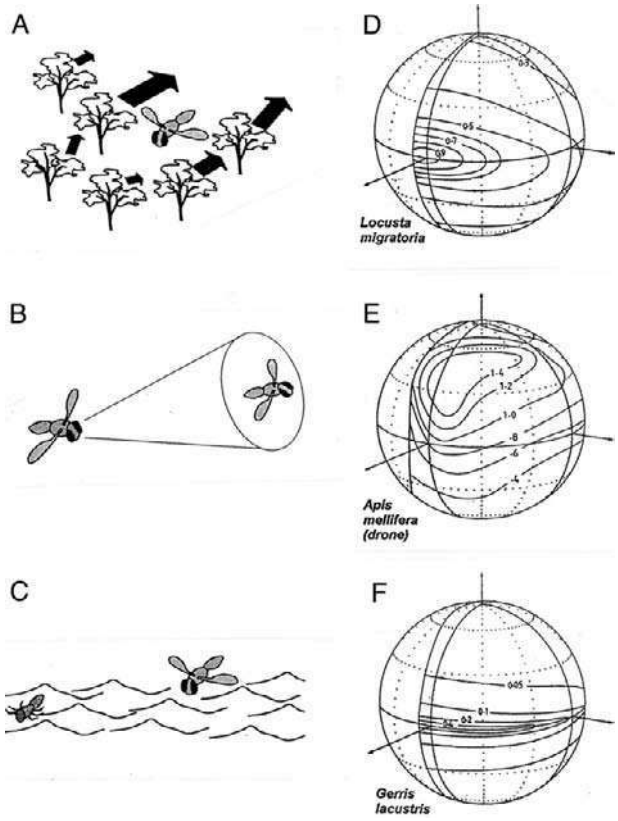

FIGURE 9 (A–C) Three situations that lead to asymmetries in the distribution of resolution in apposition compound eyes. (A) Flight through vegetation. (B) Chasing mates or prey. (C) Flight close to flat surfaces. (D–F) Plots of the density of ommatidial axes around the eyes of three insects, corresponding to the three situations in A–C. (D) Locust (forward flight pattern). (E) Drone bee (chasing females). (F) Water strider (hunting on water surface). Contours show the numbers of ommatidial axes per square degree of space around the animal. (Reproduced, with permission, from Land and Nilsson, 2002.)

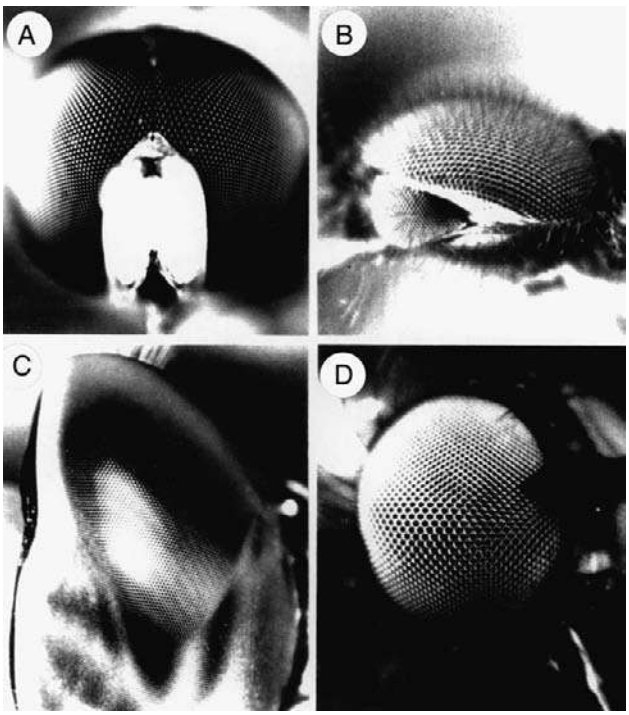

FIGURE 10 Eyes in which facet size reflects local resolution. Paradoxically, large facets produce high resolution. (A) *Syritta* (syrphid male), (B) *Dilophus* (bibionid male), (C) *Aeschna* (dragonfly), (D) *Hilara* (empid fly). See text. (Reproduced, with permission, from Land and Nilsson, 2002.)

than near objects. Clearly, near objects to the side are likely to move so fast across the retina as to cause blurring, and if this is the case it would be economical to use fewer receptors there, as high resolution is not usable. For a bee or butterfly flying half a meter from foliage, the blur streak can be estimated to be about 2.3° long. It follows that there is little point in having lateral-pointing receptors closer together than 2 or 3°, however good the resolution at the front of the eye may be. This seems to be borne out in practice. In the butterfly *Heteronympha merope,* for example, the horizontal interommatidial angle decreases from 1.4° in front to 2.6° at the side. Bees, butterflies, and acridid grasshoppers are flying insects, and their eyes all show decreasing horizontal interommatidial angles from front to rear, consistent with these ideas. Nonflying insects, e.g., many tettigonid grasshoppers, have more or less spherical eyes, without this gradient. In all the flying groups there is another, separate gradient of vertical interommatidial angles; they are smallest around the eye's equator and increase toward both dorsal and ventral poles. This results in a band around the equator with enhanced vertical acuity. The most likely reason for this vertical gradient is that the region around the eye's equator contains the highest density of information important to the animal, especially if it is an insect that feeds on flowers.

The combined effects of these two gradients on the overall density of ommatidial axes are shown for a locust in Fig. 9D, in which the contours represent the number of ommatidial axes per square degree on the sphere surrounding the animal. Worker bees and female blow flies *(Calliphora)* show a similar pattern, although in male flies and drone honey bees, this pattern is distorted to give a more pronounced acute zone concerned with mate capture (also Fig. 9E).

ACUTE ZONES CONCERNED WITH PREY CAPTURE AND MATING Many insects have a forward- or upward-pointing region of high acuity, related either to the capture of other insect prey or to the pursuit in flight of females by males (Fig. 9E). When both sexes have the specialization (mantids, dragonflies, robber flies), predation is the reason, but more commonly it is only the male that has the acute zone (simuliid black flies, hover flies, mayflies, drone bees), indicating a role in sexual pursuit. The acute zones vary considerably. In male house flies and blow flies, they may involve little more than a local increase in the acuity of the "forward flight" acute zone common to both sexes (see earlier). However, in other insects the acute zone may be in a

separate eye, as is the case with the dorsal eyes of male bibionid flies (Fig.10B). In these more extreme double eyes, the upward-pointing part is often specialized for detecting other small animals against the sky.

Good examples of forward-directed acute zones are found in the praying mantids, predators in which both sexes ambush prey. The eyes have large, binocularly overlapping acute zones that are used to center potential prey before it is struck with the spiked forelegs. Mantids provide the only known example in insects in which prey distance is determined by binocular triangulation. The interommatidial angle ($\Delta\varphi$) in *Tenodera australasiae* varies from 0.6° in the acute zone center to 2.5° laterally. Facet diameters decrease from 50 μm in the acute zone to 35 μm peripherally, but this is less of a decrease than would be expected from diffraction considerations alone.

In many male dipterans an acute zone associated with sexual pursuit is typically situated 20 to 30° above the flight direction. In *Calliphora* flies it is characterized by a low value for $\Delta\varphi$ of 1.07° compared with 1.28° in the female. In house flies and probably in other flies there are also anatomical differences at the receptor level that suggest that this region (it has been called the "love spot") is specifically adapted for improved sensitivity. This is no doubt caused by the very fast response times required for high-speed chasing. Male flies also have a number of "male-specific" interneurons in the optic ganglia, which are undoubtedly involved in the organization of pursuit behavior.

In the small hover fly *Syritta pipiens* the sex difference is particularly striking. In the male's acute zone, $\Delta\varphi$ is about 0.6°, nearly three times smaller than elsewhere in the eye or anywhere in the female eye (Fig. 10A). Drone bees have a similar anterodorsal acute zone, where the density of ommatidial axes is three to four times greater than anywhere in the female eye (Fig. 9E). They use this region when they chase the queen and can be induced to chase a dummy queen on a string subtending only 0.32°, much smaller than the ommatidial acceptance angle of 1.2°. This implies that the trigger for pursuit is a brief decrease of about 6% in the intensity received by single rhabdoms.

Most of the animals just discussed have to detect their prey or mates against a background of foliage, a far from easy task. However, many insects have simplified the problem by using the sky as a background, against which any non-luminous object becomes a dark spot. Thus, one finds not only upward-pointing acute zones but also double eyes with one component directed skyward (Figs. 10B and 10C). For example, dragonflies hunt other insects on the wing and have acute zones with a variety of configurations. Many in fact have two acute zones, one forward pointing, and presumably concerned with forward flight as discussed above, and another directed dorsally and used to detect prey. The migratory, fast-flying aeschnids have the largest eyes and most impressive acute zones. Exactly 28,672 ommatidia have been counted in one eye of *Anax junius,* which has the smallest interommatidial angles of any insect (0.24° in the dorsal acute zone) and facets of corresponding size (62 μm). The dorsal acute zone takes the form of a narrow band of high resolution extending across the upper eye along a great circle, 50 to 60° up from the forward direction. The axis density (five per square degree) is twice that in the forward acute zone and five times higher than in a male blow fly. The dorsal acute zone is easily visible as a wedge of enlarged facets (Fig. 10C). Presumably the great high-acuity stripe in *Anax* is used to trawl through the air, picking out insects against the sky much as the scan line on a radar set picks up aircraft.

Simuliid flies have divided eyes and use the upper part to detect potential mates against the sky. They can do this at a distance of 0.5 m, when a female subtends an angle of only 0.2°. As in drone bees, this is a small fraction of an acceptance angle. The eyes of male bibionid flies are similarly divided (Fig. 10B), with larger facets and smaller interommatidial angles in the dorsal eye (1.6° compared with 3.7°, in *Bibio marci*). The upper eyes are used exclusively for the detection of females; movement of stripes around the lower eye evokes a strong optomotor turning response (the almost universal visual behavior used by insects to prevent involuntary rotation) but the dorsal eye is quite unresponsive to this kind of stimulus.

HORIZONTAL ACUTE ZONES As we have seen, many flying insects have a zone of increased vertical acuity around the horizon, no doubt reflecting the visual importance of this part of the surroundings. The visual field of the locust in Fig. 9A shows this clearly. There are environments where this region is even more important.

Insects that fly over water have a similarly narrow equatorial field of interest. Empid flies hunt close to the surfaces of ponds, again looking for stranded insects, and they have a horizontal acute zone that can be recognized by a linear region of enlarged facets around the eye (Fig.10D). In *Rhamphomyia tephraea,* vertical interommatidial angles are only 0.5° in this 15°-high region, rising to 2° above and below it.

Water surfaces themselves provide a similarly constrained field of view, and water striders *(Gerris)* that hunt prey stranded in the surface film have a narrow acute band imaging this region, as shown in Fig. 9F. This has a height of only about 10°, centered on the horizon, and within this the vertical interommatidial angle in the frontal region is only 0.55°, which is close to the diffraction limit and impressive in an eye with only 920 ommatidia.

SUPERPOSITION EYES

The Nature of Superposition Imagery

From the outside, apposition and superposition eyes are almost indistinguishable. Both are convex structures with

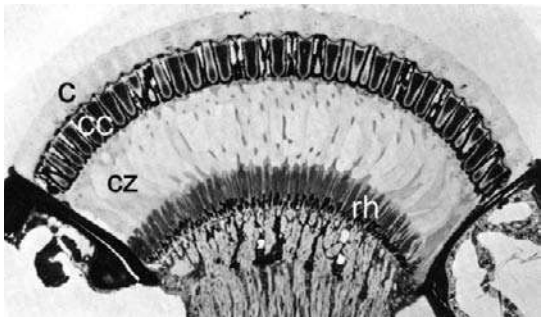

FIGURE 11 Section through the superposition eye of a dung beetle (*Onitis westermanni*). c, cornea; cc, crystalline cones; cz, clear zone; rh, rhabdoms. (Photograph by Dr. S. Caveney. Reproduced, with permission, from Land and Nilsson, 2002.)

facets of similar dimensions and are clearly variants of the same general design. But there the resemblance ends. Internally, there are several crucial anatomical differences: the retina is a single sheet, not broken up into discrete ommatidial units as in apposition eyes, and it lies deep in the eye, typically about halfway between the center of curvature and the cornea. Between the retina and the optical structures beneath the cornea there is a zone with very little in it, the clear zone, across which rays are focused—the equivalent of the vitreous space in a camera-type eye (Fig. 11). The optical devices themselves are complex—in insects they are nearly always tiny refracting telescopes—although to a cursory examination most do not look very different from the lens structures of apposition eyes.

The real surprise is optical. All superposition eyes produce a single deep-lying erect image in the vicinity of the retina. This distinguishes them not only from apposition eyes, which have multiple inverted images, but also from camera-type eyes in which the image is inverted. Clearly, we are dealing here with something quite out of the ordinary. Around the turn of the 20th century there were a number of successful attempts to photograph these images. A recent attempt by the author to re-create this photographic feat, in a firefly eye, is shown in Fig. 12 (right), in which the single erect image should be contrasted with the multiple inverted images of an eye of the apposition type (Fig. 12, left). It turns out that it is important to use a beetle (such as a firefly) for this. Other insects, in particular moths, have superposition eyes but there the optical structures that create the image are not joined to the cornea, and they are swept away when the eye is cleaned to make a lens for photography. In beetles, however, the optical elements are continuous with the cornea and so survive the removal of the eye's internal structures.

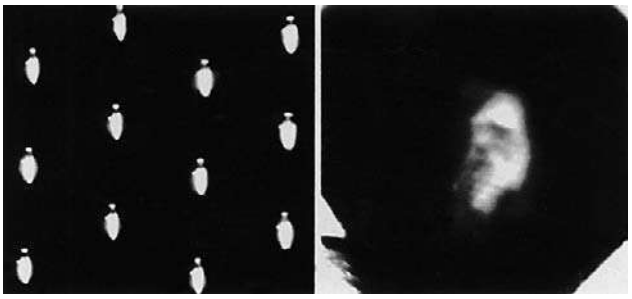

FIGURE 12 Left: Apposition-type inverted images photographed behind the cleaned cornea of a robber fly (Asilidae). Right: Photograph of an influential 19th century naturalist, taken through the superposition optics of the cleaned cornea of a firefly (*Photuris* sp.) (Reproduced, with permission, from Land and Nilsson, 2002.)

The credit for the discovery and elucidation of this remarkable piece of optics is due to Sigmund Exner, who worked on the problem throughout the 1880s and published his complete findings in 1891. Exner showed that the only way an erect image could be formed was for the optical elements to behave in a rather strange way, as shown in Fig. 13A. Basically what each has to do is not form an image from a parallel beam as in a conventional lens, but redirect light back across the element's axis, to form another parallel beam on the same side of the axis (Fig. 13B). Exner realized that although a single lens would not do the job, a two-lens telescope would, and he went on to demonstrate (as well as he could

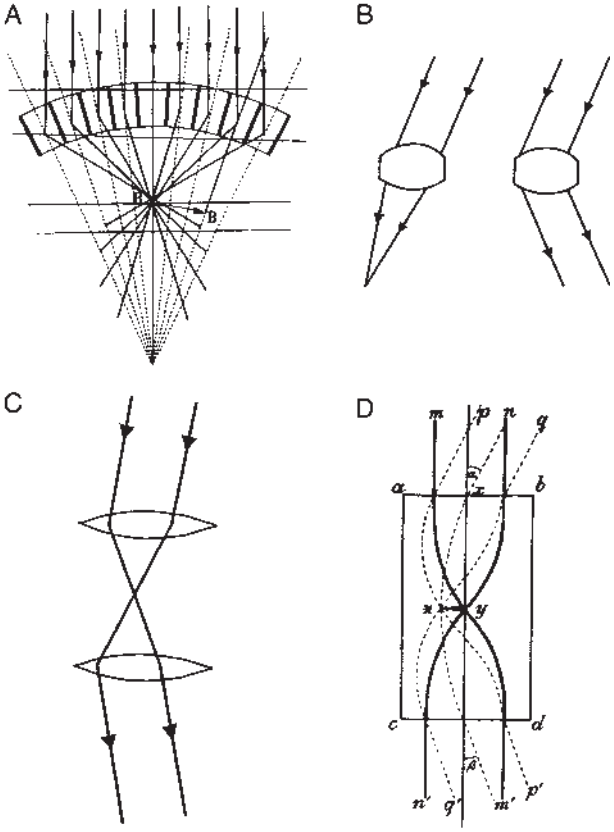

FIGURE 13 (A) Sigmund Exner's diagram of ray paths in a superposition eye. Note that the rays are bent in a "dog-leg" path by the optical elements. (B) An ordinary lens (left) will not produce the ray bending at the right, as required in (A). (C) A two-lens telescope is needed to redirect a light beam back to the same side of the axis, as in (A). (D) Exner proposed a lens cylinder equivalent to the telescope, in which rays are bent within the structure by a parabolic gradient of refractive index, highest in the center. (Reproduced, with permission, from Land and Nilsson, 2002.)

with the technology of the time) that such structures were indeed present in the superposition eyes of insects.

Telescopes and Lens Cylinders

In a lens-based superposition eye, the optical elements need to act as simple inverting telescopes that redirect the entering beam of light back across the axis, as shown in Fig. 13C. The most straightforward way to do this is to have two lenses separated by the sum of their individual focal lengths, with an image plane between them (Fig. 13C). Exner realized that, given plausible refractive indices and the curvatures of the structures revealed by histology, there was not enough ray-bending power in each element of a beetle eye to make this possible. He came up with the idea that structures must have an internal refractive index gradient similar to that in the *Limulus* horseshoe crab eye. The result would be that most of the ray bending would occur within the tissue, rather than at its external surfaces. The pure form of this structure, a flat-ended cylinder with a radial parabolic refractive index gradient, Exner called a lens cylinder. He showed that, depending on its length, it could act as a single lens or as a pair of lenses making up an inverting telescope of the kind required for superposition optics (Fig. 13D). Although Exner did not have the means in his time of establishing whether beetles and moths had optical elements with the required refractive index gradient, numerous studies since the advent of interference microscopy have shown that his brilliant conjecture was correct.

Resolution and Sensitivity

The geometry of a superposition eye is shown in Fig. 14. The peculiarities of this type of image formation mean that the nodal point of the eye (the point through which rays pass undeviated) is at the center of curvature, and the focal length is the distance out from the center to the image. The interrhabdom angle ($\Delta\varphi$) is s/f, where s is the rhabdom separation, just as in a camera-type eye. As in apposition eyes,

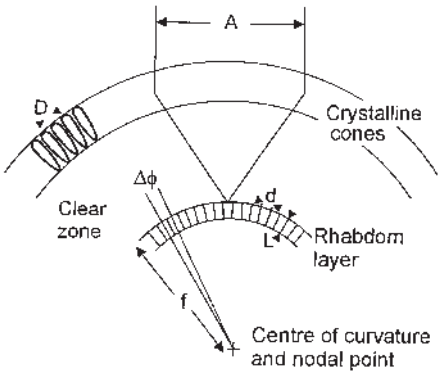

FIGURE 14 Optical definitions that apply to superposition eyes. D, facet diameter; A, superposition aperture; $\Delta\varphi$, interreceptor angle (compare Fig. 2B); d, rhabdom diameter; L, rhabdom length; f, focal length. (Reproduced, with permission, from Land and Nilsson, 2002.)

the rhabdom acceptance angle is a combination of the geometrical subtense of a rhabdom *(d/f)* and the width of the blur circle provided by the optics (Fig. 7A).

In the past, there has been a belief that superposition eyes suffer from poor resolution, mainly because of the difficulty of conceiving how the large numbers of ray bundles contributing to a single point on the image could be directed there with sufficient accuracy. However, this reputation seems not to be justified, except perhaps in extreme cases. A careful study by Peter McIntyre and Stan Caveney on the eyes of dung beetles that fly at different times of the day and night found that in the day-flying *Onitis belial* about 50 optical elements (the effective superposition aperture) contributed to the image at any one point, and in the nocturnal *O. aygulus* the number was close to 300. *O. belial* had a calculated rhabdom acceptance angle ($\Delta\rho$) of 2.2°, which is comparable with values from many apposition eyes, and in *O. aygulus* $\Delta\rho$ was somewhat larger, 3.0°, which is still quite impressive for an eye with such a huge aperture. These modeling studies have since been confirmed by electrophysiological recordings from single receptors. In the Australian day-flying moth *Phalanoides tristifica,* the image quality has been measured directly with an ophthalmoscopic method that uses the eye's own optics to view the retina and images on it. The result was that $\Delta\rho$, the acceptance angle of a rhabdom when viewing a point in space, was 1.58°, of which optical blur contributed only 1.28°. This is itself only slightly larger than the half-width of the Airy diffraction image from a single facet. Thus, a superposition eye in which 140 elements contribute to a point image has optics that are almost as good as is physically possible. (Although the superposition pupil is many times wider than an individual facet, it does not behave for diffraction purposes as a single large lens, and the Airy disk diameter depends on the diameter of single facets, just as in apposition eyes.)

Size for size, superposition eyes are more sensitive than apposition eyes, which is why they are most commonly encountered in animals such as moths and fireflies that are active at night. For an apposition eye and a superposition eye of the same size and the same resolution, the sensitivity of the superposition eye (with an aperture 10 facets wide) is 100 times that of the apposition eye, meaning that it will work just as well at light levels 100 times lower.

Eye Glow and the Superposition Pupil

Most moths have a reflecting layer (tapetum) behind the rhabdoms. Its function is the same as the tapetum in the eye of a cat: to double the light path through the photoreceptors and so to improve their photon catch. In some diurnal moths, a reflector also surrounds each rhabdom, optically isolating it from its neighbors. In dark-adapted eyes, the tapetum causes the eye to glow when viewed from the same direction as the illuminating beam (Fig. 15). In some diurnal moths, such as the sphingid *Macroglossum,* the glow is always

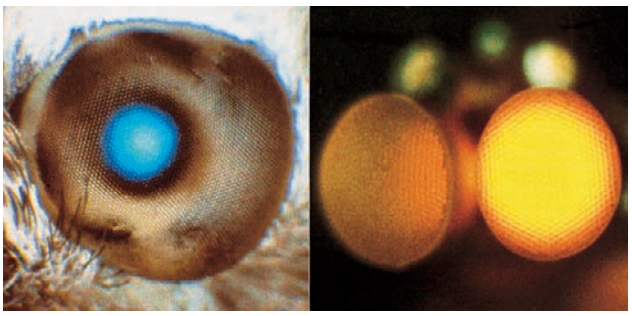

FIGURE 15 Left: Blue light reflected from the tapertum of the day-flying hummingbird hawk moth *(Macroglossum).* The bright area corresponds to the superposition pupil. Right: Superposition dorsal eyes of a male mayfly *(Centroptilum).* The yellow color is not from a tapetum, but results from the scattering of long wavelengths by screening pigment. (Photographs by Dr. D.-E. Nilsson. Reproduced, with permission, from Land and Nilsson, 2002.)

visible. The mechanism is similar to that in a cat's eye. The optical system forms a point image of the light source on the tapetum, or close to it, and this point acts as an emitter of light which, on passing through the optics again, emerges as a roughly parallel beam.

If the optics are good, that is to say they really do bring a parallel beam to a point in the image, then the patch of glow seen at the surface of the eye will have the same diameter as the beam that entered the eye. This is the superposition pupil (i.e., the amount of eye surface from which rays contribute to each point on the image (Fig. 15). Eye glow can also provide a useful test of image quality. If the glow can be seen only over a narrow angle (a few degrees) from the direction of the illuminating beam, then the retinal image must itself be very small. On the other hand, if the glow can be seen over a wide angle, this indicates either that there is a large blur circle on the retina or that the tapetum is situated a long way from the focus.

Light and Dark Adaptation

The high sensitivity of most superposition eyes means that they must protect their visual pigment in daylight and so need adaptation mechanisms that can reduce image brightness by several orders of magnitude. The main mechanism of light adaptation in superposition eyes consists of pigment movements that result in the progressive interception of rays from the outer zones of the superposition pupil (Fig. 16). This reduction may ultimately result in light from only a single facet reaching a single point in the image, which is essentially the apposition condition.

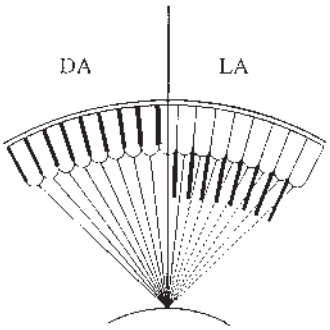

FIGURE 16 Mechanism of dark and light adaptation (DA, LA) in superposition eyes. Screening pigment migrates inward, cutting off the outer rays in the image-forming bundle. (Reproduced, with permission, from Land and Nilsson, 2002.)

The eye glow (Fig. 15) provides a means of monitoring the process of light and dark adaptation. As oblique rays across the clear zone are cut off during light adaptation (Fig. 16), the brilliance of the glow and the size of the patch are reduced, often disappearing completely. In the dark, these slowly return. In insects with refracting superposition eyes, the main pigment movement is a longitudinal inward migration of granules in both the primary and the secondary pigment cells. In the dark, the granules are bunched up between the crystalline cones, and with the onset of light they extend inward, over a matter of minutes, to occupy much of the clear zone.

Interestingly, the trigger for pigment migration in some moths is not provided by photoreception in the rhabdoms themselves. In the crepuscular sphingid moth *Deilephila,* a region immediately beneath each crystalline cone initiates pigment migration, when illuminated with ultraviolet light, and the much deeper lying rhabdoms are not involved. However, in the owl fly *Ascalaphus,* a day-flying neuropteran with double superposition eyes, the pigment movements can be triggered from both the region below the cones and the rhabdoms themselves.

Single and Double Eyes

In superposition eyes, major departures from spherical symmetry are rare because the geometry of the eye is constrained by the shared optics (the hummingbird hawk moth *Macroglossum* is an exception in this respect, with a visibly asymmetric eye, but excellent resolution everywhere). One way around this problem is the use of double eyes, in which each part is essentially separate from the other and has its own radius of curvature. Although common among crustacean groups such as mysids and euphausiids, double superposition eyes are uncommon among insects. As mentioned earlier, owl flies *(Ascalaphus)* have double superposition eyes. Male mayflies have a pair of dorsal superposition eyes, which they use for sighting females against the sky, in a way similar to that of bibionid flies (Fig. 10B). However, the lower eyes, present in both sexes and responsible for other visual activities, are of the apposition type. The field of view of the dorsal eye is small, and it is adjusted to the environmental circumstances of the species; those species swarming in woods with small gaps in the canopy have the narrowest fields.

Afocal Apposition: The Eyes of Butterflies

Butterflies and moths are classified together in the Lepidoptera and are undoubtedly very closely related. Most

butterflies [skippers (Hesperidae) are the exception] have eyes that behave in most respects as apposition eyes. They have long narrow rhabdoms abutting the bases of the crystalline cones, no clear zone, and complex pseudopupils. Many moths, on the other hand, have refracting superposition eyes with wide, deep-lying rhabdoms, clear zones, and eye glow. Transitions between the eye types must have occurred a number of times within the moths, as well as between moths and butterflies. A very similar picture emerges in the beetles, most of which have apposition eyes, but a substantial number of nocturnal and crepuscular groups, including the dung beetles and the fireflies, have superposition optics.

It is not very easy to see how it is possible to get from one type of eye to the other, without going through an intermediate that does not work. Apposition eyes use simple lenses and superposition eyes two-lens telescopes (or the equivalent lens cylinder devices), and there does not seem to be much room for compromise. In the case of butterflies we do know the answer: in 1984 Dan-Eric Nilsson and his colleagues discovered that their apposition eyes actually have an extreme form of superposition optics in the ommatidia, in which the proximal lens in each telescopic pair has become not weaker, as one might have guessed, but extremely powerful (see Fig. 6C).

The way this works is shown in Fig. 17A. As in a normal superposition eye, a combination of the curved cornea and a weak lens cylinder in the distal region of each crystalline cone results in the formation of an image within the crystalline cone, about 10 μm in front of its proximal tip. This focused light then encounters a lens with an extraordinarily short focal length, about 5 μm. The discovery of this lens involved taking thin frozen sections from the tiny region at the base of the crystalline cone and examining their image-forming properties. The last 10 μm of the cone produced excellent images. The effect of this second lens is to bring the light focused by the first (distal) lens back into a parallel beam, just as in a superposition eye. The essential difference is that, whereas in a superposition eye the magnification of the telescopic pair of lenses rarely exceeds –2, here it is much greater. The large difference in the focal length of the distal and proximal lenses gives an overall magnification of –6.4 in the nymphalid butterfly *Heteronympha merope.*

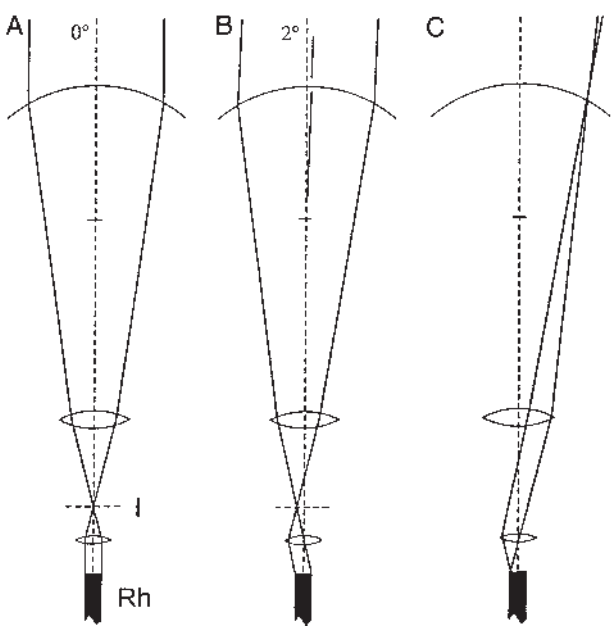

FIGURE 17 "Afocal" apposition in butterfly eyes. (A and B) Although each ommatidium acts independently, like an apposition eye, the optical elements function as telescopes with an internal image, as in superposition eyes (Fig. 13). The wide beam of light reaching the cornea is reduced to "fit" the rhabdom (see text). (C) A consequence of this arrangement is that the rhabdom tip is imaged onto the cornea. I, image plane; Rh, rhabdom. (Reproduced, with permission, from Land and Nilsson, 2002.)

This high magnification has two important consequences, illustrated in Fig. 17B. The first is that the beam that emerges from the proximal tip makes an angle with the axis that is 6.4 times greater than the beam that entered the facet from outside. A ray making an angle of 1° with the facet axis emerges at 6.4°, and similarly a beam 3° wide at the cornea emerges into the rhabdom as a 19.2°-wide beam. The significance of this is that a rhabdom with a refractive index of 1.36 will just contain (by total internal reflection; Fig. 7B) a beam 22° wide, which in turn means that the acceptance angle of the ommatidium will be limited to just over 3°: light making higher angles with the rhabdom wall will escape and be absorbed by the surrounding pigment. Thus, in this kind of eye, the ommatidial acceptance angle is limited principally by the refractive index of the rhabdom, not (as in a conventional apposition eye) by its diameter (Fig. 7A). The second effect of the magnification is to reduce the diameter of the beam leaving the base of the crystalline cone by a factor of about 9 (angular magnification × refractive index), compared with that entering the facet. The entering beam is limited by the facet diameter, typically about 20 μm. The beam leaving the crystalline cone and entering the rhabdom is squashed down to a diameter of 2.1 μm, which is indeed close to the diameter of a butterfly rhabdom. Thus rhabdom diameter and facet diameter are related and between them determine the effective aperture of the ommatidium and hence its sensitivity. Bright-light butterflies tend to have smaller facets (20 μm) and narrow rhabdoms (1.5–2 μm), whereas the crepuscular Australian butterfly, *Melanitis leda,* has 35-μm facets and 5-μm rhabdoms. A further consequence of this optical system is that the rhabdom tip is imaged onto the cornea (Fig. 17C), which means that one can sometimes see magnified versions in the cornea of the wave-guide mode phenomena that occur in the rhabdom.

What we have seen is that butterfly eyes behave as apposition eyes, because light entering a single facet is received by a single rhabdom. They are called afocal because light is not focused on the rhabdom tip as in most apposition eyes, but enters the rhabdom as a parallel beam. In their fundamental optical design, however, these ommatidia remain of the superposition type, constructed from two-lens telescopes. This makes it easy to understand how different

lepidopteran groups managed to switch readily from the diurnal (apposition) version of the afocal eye to the nocturnal (superposition) version. To become nocturnal, the powers of the distal and proximal lenses must become more equal, the receptor layer moves to a deeper location, and gradually more and more facets contribute to the image. There are no blind intermediaries.

SIMPLE CORNEAL EYES IN INSECTS

Insect simple eyes, or ocelli, fall into two main groups: the larval eyes of holometabolous insects and the dorsal ocelli present in most winged adult insects. In both, the curved air/tissue cornea interface is the main refracting surface, although as in vertebrate eyes, a lens of some kind often augments the optical power of the system and aids in the formation of the image.

Larval Ocelli

In insects with a distinct larval stage, the ocelli are the only eyes the larvae possess. They vary greatly in size and complexity. The larvae of flies have no more than a small group of light-sensitive cells on either side of the head. Lepidopteran caterpillars, however, have ocelli with lenses and a structure resembling that of a single ommatidium from a compound eye. In each ocellus in the *Isia,* seven receptors contribute to a two-tiered rhabdom containing the photopigment (Fig. 18A). There seems little possibility of spatial resolution within each ocellus, but as it appears that the fields of view of the 12 ocelli do not overlap, they are capable of providing a 12-"pixel" sampling mosaic of the surroundings. These ocelli do, however, resolve color; three spectral types of receptor have been found in butterfly larval ocelli.

The ant lion *Euroleon* (Neuroptera) also has six ocelli on each side of the head, borne on a small turret (Fig. 18B). Unlike caterpillars, however, each has an extended retina of 40 to 50 receptors, giving interreceptor angles ($\Delta\varphi$) of 5 to 10°. Although this resolution is not impressive, it is presumably enough to allow the animals to detect their prey, e.g., moving ants, at a distance of about 1 cm. Sawflies (Hymenoptera) have larvae with a single pair of ocelli, each with an in-focus retina covering a hemisphere (Fig. 18C). The rhabdoms in *Perga* are made up of the contributions from eight receptors (much as in an ordinary compound eye) and are spaced 20 μm apart, giving an interreceptor angle of 4 to 6°. These larvae are vegetarian, and it seems that the main function of the ocelli is to direct the larvae to their host plants. However, *Perga* larvae will also track moving objects with their head and defend themselves by spitting regurgitated sap.

The most impressive of all larval ocelli are found in tiger beetles *(Cicindela).* These have a lifestyle similar to that of ant lions, ambushing insect prey as they pass their burrows (Figs. 18D–18F). There are again six ocelli on each side of the head, but two are much larger than the others. The largest has a diameter of 0.2 mm and a retina containing 6350 receptors. The interreceptor angle is about 1.8°, comparable with or better than the resolution of the compound eyes of most adult insects. This raises the interesting question as to why the insects did not retain eyes like this into adult life.

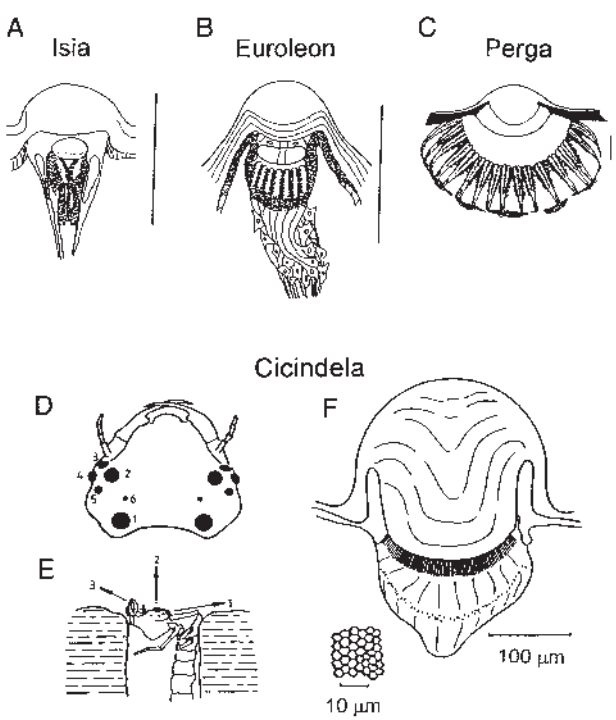

FIGURE 18 Simple eyes (ocelli) of increasing complexity in larval insects. (A) Lepidopteran, (B) Neuropteran, (C) Hymenopteran. Scale bars, 0.1 mm. (D–F) Large simple eyes of tiger beetle larvae *(Cicindela).* They are used to spot prey (usually ants), which they ambush from the burrow. (D) Head with six pairs of eyes. (E) Larva in ambush position. (F) Largest ocellus showing corneal lens and retina. Inset: Tangential section of retina. (Reproduced, with permission, from Land and Nilsson, 2002.)

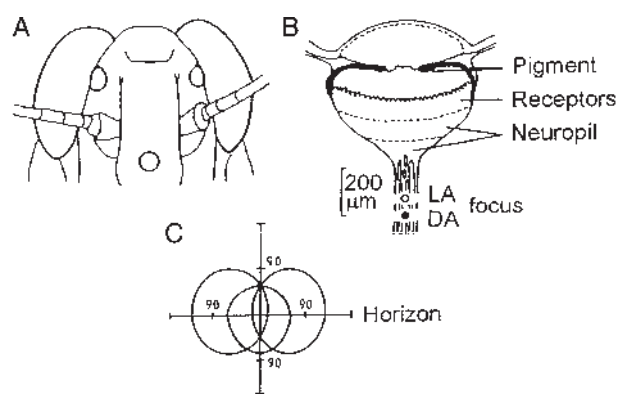

FIGURE 19 Dorsal ocelli of adult locust. (A) Positions of frontal and lateral ocelli on head. (B) Section of an ocellus, showing the different layers and the positions of the focus in light- and dark-adapted states. The focus is a long distance behind the receptor layers. (C) Fields of view of the three ocelli straddling the horizon. (Reproduced, with permission, from Land and Nilsson, 2002.)

Dorsal Ocelli of Adults

Adult insects that fly typically have three simple eyes on the top of their heads. These dorsal ocelli resemble larval ocelli in possessing a lens and (like sawfly larvae) an extended retina (Fig. 19), but they are not embryologically related to the larval eyes. Some dorsal ocelli have tapeta, and some a mobile iris. They each have a wide field of view of 150° or more and may have as many as 10,000 receptors. So far all this suggests that these are "good" eyes, like those of hunting spiders. However, they are profoundly out of focus, with the retina much too close to the lens. For example, in the blow fly *Calliphora* the receptors extend from 40 to 100 μm behind the lens, but the focus is at 120 μm.

What then are they for? Recent studies mainly support the idea that the ocelli are horizon detectors, involved in enabling an insect to make fast corrections for pitch and roll. The defocus then makes sense; high spatial frequency clutter such as leaves and branches will be removed, allowing the receptors to respond to changes in the overall distribution of light in the sky. The idea that these ocelli contribute to flight equilibrium is supported by the fact that the receptors converge massively onto a relatively few second-order neurons that project directly into the optomotor system.

See Also the Following Articles

Brain and Optic Lobes • Ocelli and Stemmata

Further Reading

Exner, S. (1989). "The Physiology of the Compound Eyes of Insects and Crustaceans." Springer-Verlag, Berlin. [Translated from the 1891 German ed. by R. C. Hardie.]

Land, M. F. (1997). Visual acuity in insects. *Annu. Rev. Entomol.* **42,** 147–177.

Land, M. F., and Nilsson, D.-E. (2002). "Animal Eyes." Oxford University Press, London.

Nilsson, D.-E. (1989). Optics and evolution of the compound eye. *In* "Facets of Vision" (D. G. Stavenga and R. C. Hardie, eds.), pp. 30–73. Springer-Verlag, Berlin.

Wehner, R. (1981). Spatial vision in arthropods. *In* "Handbook of Sensory Physiology VII/6C" (H. Autrum, ed.), pp. 287–616. Springer-Verlag, Berlin.

F

Fat Body

Ephraim Cohen
The Hebrew University of Jerusalem

The insect fat body is a mesodermal tissue composed of a meshwork of loose lobes suspended in the hemocoel and bathed in the insect hemolymph. The tissue is composed primarily of vacuolated rounded or polyhedral cells called adipocytes or trophocytes, which commonly harbor stored inclusions of proteins, lipids, and glycogen. In certain insect species, mycetocytes (cells containing symbiontic microorganisms) and urocytes (cells containing nitrogenous waste product in the form of uric acid) are present. The fat body is also associated with connective tissue and various blood cell types. Being a major biosynthetic and storage organ in insects, the insect fat body is equivalent to the vertebrate liver. It is the prime location of intermediary metabolism and detoxification processes, as well as storage and excretion of glycogen, lipids, and proteins. Storage of reserves is characteristic of the larval fat body cells. Such reserves are subsequently used for metamorphosis in holometabolous insects and for flight and reproduction in adults.

STRUCTURAL ORGANIZATION

Although the insect fat body is widely distributed throughout the hemocoel, two major regions can be distinguished. Near the integument and musculature is the peripheral (subcuticular) fat body, which largely functions for storage. The second layer, the perivisceral (gut) fat body, which surrounds the alimentary canal, is more metabolically active than the previous layer. The fat body tissue surrounds other insect organs such as brain and nervous tissues, gonads, and muscles. It is noteworthy that the fat body is intimately associated with nearly all vital tissues and organs in the insect body, including the tracheal system, the musculature, the Malpighian tubules, and the hemolymph. This spatial organization is well adapted to the physiology and the open circulatory diffusion system of insects, thereby facilitating absorption and release of metabolites and nutrients.

FAT BODY CELLS

Adipocytes (trophocytes) are the predominant cell type associated with metabolic and storage functions. In young cells, a few inclusions can be detected and the nuclei are round. As the cells mature and accumulate nutritional reserves, they become vacuolated and the nuclei are compressed. The colors of adipocytes, which depend on the insect species and change with maturation, range from white, yellow, tan, and brown to blue.

Urocytes are special cells common in cockroaches, which sequester uric acid (the main end product of nitrogen metabolism in terrestrial insects) for excretion and storage. They are degenerate cells, which unlike adipocytes, lack organelles such as mitochondria, ribosomes, or the endoplasmic reticulum.

Mycetocytes are cells that harbor symbiontic microorganisms and may serve for nutritional purposes. Mycetocytes are in proximity to urocytes, a spatial organization that implies some sort of physiological–biochemical interaction.

The adipocytes are arranged in two or three layers in the periphery of the fat body lobe, and the more metabolically active cells face the circulatory system. The mycetocytes are located in the center of the lobe surrounded by urocytes.

Other cell types associated with the fat body, including various blood cells, can be found adhered to fat body cells. Oenocytes, which are large ectodermally derived cells, have also been observed to be attached to adipocytes. Their exact physiological role is unresolved.

METABOLIC FUNCTIONS

The fat body participates in myriad metabolic activities and functions. Absorption from hemolymph and buildup of

intracellular storage nutrients in the form of lipid droplets, carbohydrate (glycogen) deposits, and protein granules during the immature stages are aimed at accumulating reserves for later stages, and primarily to serve adult activities. Fat body cells, having homeostatic functions related to metabolism, respond to nutritional and hormonal cues that regulate and modulate blood sugars, lipids, and proteins at larval and mature stages.

As in vertebrates, the oxidative metabolism is mediated via the tricarboxylic acid (TCA) cycle and the electron transport enzyme systems. The fat body contains enzymes mediating the gluconeogenesis process as well as enzymatic systems with a detoxification role to manage harmful endogenous metabolites and toxic xenobiotic compounds. Detoxifying enzyme systems include microsomal mixed function oxidases, in which the cytochrome P450 is predominant, various hydrolytic enzymes (esterases, phosphoesterases), and conjugating systems.

The cells synthesize the various blood proteins (lipoproteins, glycolipoproteins), which include juvenile hormone (JH) carrier proteins (protecting JH from degradation), diglyceride carrier proteins, diapause proteins, and, particularly at the adult stage, production of vitellogenins (yolk proteins) that are absorbed by the maturing oocytes. Fat body cells also synthesize JH esterase, which regulates levels of JH in the insect blood, and enzymes involved in purine metabolism. Generally, proteins released into the insect blood during larval development are sequestered by the adipocytes, forming large intracellular granules until their use during metamorphosis. Triglycerides, which are the major form of stored lipids, are mobilized when needed and released into the hemolymph in the form of diglycerides accompanied by the production of specific carrier proteins. Trehalose, produced by the fat body, constitutes the major disaccharide in the insect blood. Glycogen, which is the principal form of stored carbohydrates, is mainly present in the peripheral fat body adipocytes. Glycogen is synthesized (by glycogen synthase) and hydrolyzed (by glycogen phosphorylase), by these enzymes active in the fat body cells. The hydrolytic products are mobilized at molting and metamorphosis to serve as precursors required for chitin synthesis and formation of the new cuticle.

ENDOCRINE CONTROL OF FAT BODY METABOLISM

Neuroendocrine secretions from brain and ganglia, ecdysteroids (molting hormones), JHs, and the myriad corpora cardiaca neurosecretions affect the metabolic state of the adipocytes. These endocrine secretions are strongly influenced by stimuli from internal and external environments, and they function to coordinate and integrate crucial metabolic activities involved in molting, growth, metamorphosis, and reproduction. The fat body is a target tissue for endocrine regulation as is illustrated shortly. Stored glycogen and proteins are mobilized during the molting process to form the newly synthesized cuticular chitin–protein complex. The blood level of trehalose is regulated by a corpora cardiaca neurohormone. The adipokinetic hormone from the corpora cardiaca stimulates the adipocyes to release diglycerides and the accompanied lipoprotein carrier, and enhances lipid oxidation to fuel flight in favor of carbohydrate oxidation. Synthesis and release of vitellogenins by the female fat body cells usually are under the control of JH, although in certain insect species also the molting hormone is involved.

FAT BODY DURING DEVELOPMENT AND METAMORPHOSIS

During the period of metamorphosis the fat body tissue undergoes extensive morphological, histological, biochemical, and organizational changes. These processes are triggered by the molting hormone on the background presence of extremely low levels of the JH. Such alterations have been thoroughly studied in dipterans and lepidopterans. Two major strategies for transforming the larval fat body into an adult tissue exist: (1) the histolytic pathway, in which the larval fat body adipocytes in dipteran species are completely histolyzed and the adult new tissue is formed from undifferentiated stem cells, and (2) the remodeling pathway, in which adipocytes in the larval stage of lepidopteran insects dissociate at metamorphosis into individual cells before being reassociated into the adult new tissue. In certain holometabolous insect species, a combination of the two processes takes place.

Dynamic exchanges of nutrients between fat body cells and the hemolymph compartments are evident throughout the life cycle of holometabolous insects (Fig. 1). Buildup of reserves and their partial use at the molting periods are characteristic of the larval stages. During the prepupal period, mass quantities of reserve material are accumulated in the fat body cells. Lysis of fat body cells in higher dipteran species at metamorphosis results in the discharge of stored reserves into the hemolymph. However, as the new adult fat body cells are

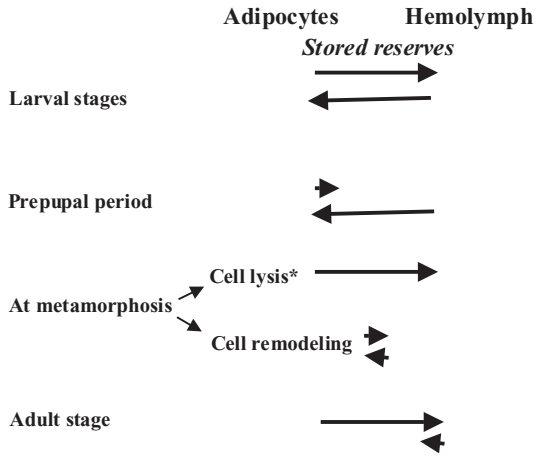

FIGURE 1 Exchange of stored reserves between fat body cells and hemolymph during the life cycle of holometabolous insects. Asterisk indicates that later, as stem cells are differentiated into adult fat body cells, a buildup of reserves occurs.

formed, nutrients are reabsorbed. In contrast, in lepidopteran species in which cell remodeling occurs, a status quo prevails in that fat body cells in adults are depleted of reserve materials, which are used for locomotion and reproduction.

See Also the Following Articles

Excretion • Juvenile Hormone • Metabolism • Vitellogenesis

Further Reading

Dean, R. L., Locke, M., and Collins, J. (1985). Structure of the fat body. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," Vol. 3 (G. A. Kerkut and L. I. Gilbert, eds.), pp. 155–210. Pergamon Press, Oxford, U.K.

Haunerland, N. H., and Shirk, P. D. (1995). Regional and functional differentiation in the insect fat body. *Annu. Rev. Entomol.* **40,** 121–145.

Keeley, L. L. (1985). Physiology and biochemistry of the fat body. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" Vol. 3 (G. A. Kerkut and L. I, Gilbert, eds.), pp. 211–248. Pergamon Press, Oxford, U.K.

Law, J. H., and Wells, M. A. (1989). Insects as biochemical models. *J. Biol. Chem.* **264,** 16355–16638.

Feeding Behavior

R. F. Chapman

University of Arizona, Tucson

The most extensive studies of insect feeding behavior and its regulation have focused on two insects with completely different feeding habits: the adult black blow fly *(Phormia regina),* which is a fluid feeder, and the final nymphal stage of the migratory locust *(Locusta migratoria),* a grass-feeding insect. Although their feeding habits are different, there are many common features in their feeding behavior patterns and the mechanisms that control their feeding behavior. It is likely, though unproven, that similar principles are involved in other insects, and this article focuses on these two insects, making reference to other species and other feeding habits where data are available. V. G. Dethier initiated and guided the earlier work on the blow fly, summarized later by Stoffolano. The studies on the locust were initiated by R. F. Chapman and developed by E. A. Bernays and S. J. Simpson.

WHY INSECTS HAVE DISCRETE MEALS

Although casual observation may give the impression that insects feed nonstop, many insects eat in discrete meals separated by periods of nonfeeding (Figs. 1A and 1B). This is most extreme in species that feed on vertebrate blood: larval *Rhodnius prolixus* take a single meal in each developmental stage, and adult female mosquitoes usually have a single blood meal associated with each vitellogenic cycle. Nectar-feeding insects and phytophagous insects also feed in discrete meals, but the degree to which this is true in other insects has not been investigated. For predatory insects, a single prey item is commonly not sufficient to produce satiation, and it is likely that a "meal" would involve several prey, just as a "meal" for a nectar-feeding insect involves foraging from a number of flowers because no single flower contains sufficient nectar to produce satiation.

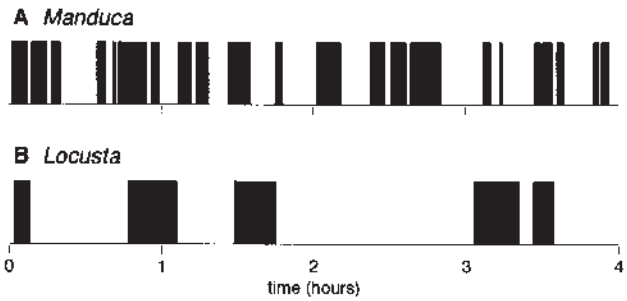

FIGURE 1 Patterns of meals taken by a caterpillar and a locust. Periods of feeding are shown by black bars. Notice that short pauses sometimes occur within meals, but the intervals between meals are markedly longer. (A) A final stage larva of *M. sexta* feeding on tobacco leaves. [Reproduced, with permission, from Reynolds, S. E., Yeomans, M. R., and Timmins, W. A. (1986). The feeding behaviour of caterpillars on tobacco and artificial diet. *Physiol. Entomol.* **11,** 39–51.] (B) A final stage nymph of *L. migratoria* feeding on wheat. [Reproduced, with permission, from Blaney, W. M., Chapman, R. F., and Wilson, A. (1973). The pattern of feeding of *Locusta migratoria. Acrida* **2,** 119–137.]

The underlying causes of this behavior are probably both physiological and ecological. Energy is expended in acquiring food and initiating digestion so that, when food is first ingested, there is a net loss of energy. Subsequently, as food is digested and absorbed, there is a gain in resources and energy, but as the process continues the rate of gain declines and the net gain plateaus as digestion and absorption are completed (Fig. 2). Consequently, there is an optimal period for which an insect should retain food in its gut before replacing it with

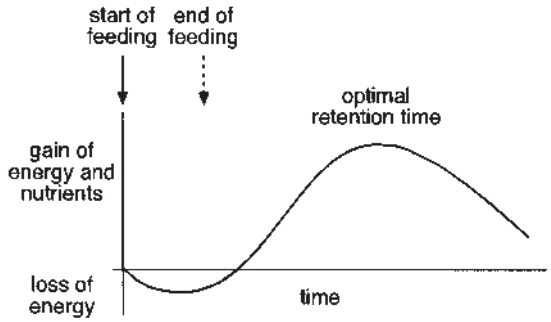

FIGURE 2 Nutrient and energy returns associated with eating a discrete meal. At first, the insect expends energy in obtaining its food. As the food is digested, nutrients are absorbed increasingly rapidly but as the nutrients are removed from the food the nutrient return decreases. To optimize the rate of nutrient return, most insects have discrete meals, with intervals between meals that approximate to peak rates of return. [Reproduced, with permission, from Sibly, R. M., and Calow, P. (1986). "Physiological Ecology of Animals." Blackwell Science, Oxford.]

newer, undigested food. Thus, it is advantageous for the insect to eat a discrete amount of food (a meal) and process it before taking more food.

The risk of predation has almost certainly also had a major role in shaping feeding behavior. Predation risks are much higher during feeding presumably because of the movements made by the insect that can be detected visually or mechanically by potential predators. For example, genista caterpillars, *Uresiphita reversalis,* feed for only about 3% of the day, yet 80% of predation by anthocorid bugs occurs during this period. Similarly, although tobacco hornworm caterpillars *(Manduca sexta)* on tobacco plants in a greenhouse fed for only about 7% of the time, 20% of predation occurred during this period.

Grasshoppers typically move away from a feeding site following a meal, sometimes backing down into the mass of vegetation and remaining unmoving and hidden until the time to feed again approaches. The caterpillars of *U. reversalis* move into silken shelters between meals. Most blood-sucking insects leave the host as soon as they are replete, usually moving to shaded places where they are inconspicuous, and it is probably true that most insects move away from the immediate area of feeding where food-related cues might reveal their presence to predators.

As a consequence of feeding in discrete, relatively short meals, the time spent feeding by most insects is only a small proportion of the available time; for most of the remainder they remain inactive and presumably minimize predation risks. Blood-sucking insects, which commonly ingest more than their own weight of food in a single meal, feed for less than 1% of the time; nectar-feeding butterflies and flies (feeding on unlimited supplies of nectar in the laboratory) feed for up to 14% of the time and this is true also for grasshoppers, both in the laboratory and in the field. All these insects have part of the gut modified for temporary food storage. Final-stage caterpillars of the tobacco hornworm spend about 35% of the time feeding in the field. In grasshoppers, the reduction in activity after feeding is controlled, at least partly, by a hormone released from the corpora cardiaca at the end of a meal. Hormonal release is induced by distension of the crop at this time.

Phloem-feeding homopterans appear to differ from most other insects. Planthoppers and aphids do not have discrete meals and ingest food more or less continuously. The phloem provides a continuous supply of sugars and free amino acids, requiring little or no digestion, so the availability of nutrients for absorption remains virtually unchanged over time. Under these circumstances the physiological necessity of eating discrete meals is eliminated. In these insects, the act of feeding is not associated with obvious body movements because once the feeding tube is plugged into a phloem sieve tube, the insect remains in one place for hours. This probably applies to xylem-feeding insects, which also need to process very large amounts of fluid because of the low concentrations of nutrients in xylem. Filter-feeding aquatic insects, such as some mosquito larvae, also probably feed continuously.

THE START OF FEEDING

As the time from the previous meal (the intermeal interval) gets longer, the likelihood that the insect will respond to food stimuli increases. A locust starts to move again and so the likelihood of encountering food is increased. Other factors, not related to the food, may also further increase the probability of feeding. In a locust, a sudden increase in light intensity or the act of defecation may have such effects. Conversely, an encounter with a highly unpalatable food source may delay the start of feeding and careless movements by an observer may have a similar effect. Simpson demonstrated that, in the migratory locust, there was in addition a tendency for meals to begin with some pattern of regularity which, in his observations, had a period varying from 12 to 16.5 min in different individuals. This does not mean that feeding or some other activity occurred every 15 min, but when it did so it was usually at some multiple of 15 min from a set time, which he determined to be lights-on in his experiments (Fig. 3). There is now evidence for similar rhythms in the caterpillars of an arctiid moth, *Grammia geneura,* and the sphingid *M. sexta.* The evidence for the latter is based on field observations, and the rhythm had a period of 3 to 4 min. The discovery of these rhythms was dependent on detailed, long-term observations on individual insects. Such sets of observa-

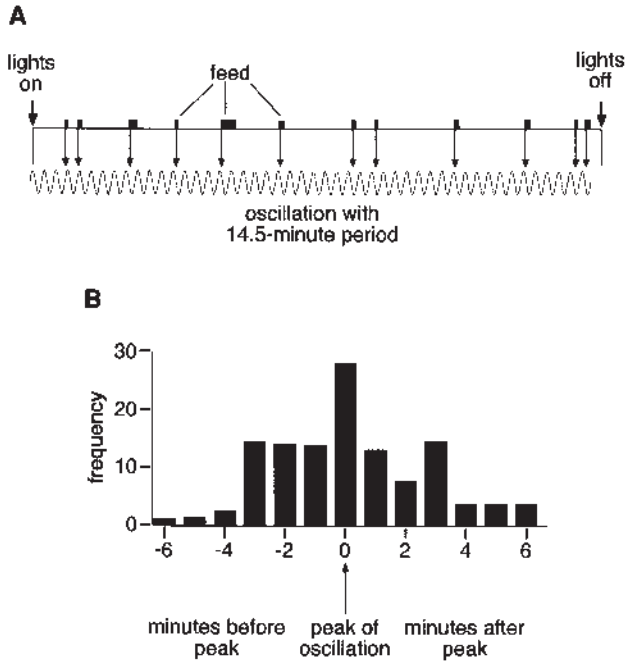

FIGURE 3 Oscillation underlying the feeding behavior of the migratory locust, *L. migratoria.* (A) Feeding record of an individual during a 12-h light phase. Notice that each meal begins close to the peak of a 14.5-min oscillation. (B) Times at which feeding started relative to the peak of the oscillation for eight insects on one 12-h day. [Reproduced, with permission, from Simpson, S. J. (1981). An oscillation underlying feeding and a number of other behaviours in fifth-instar *Locusta migratoria* nymphs. *Physiol. Entomol.* **6,** 315–324.]

tions are rare, and the extent to which similar rhythms occur in other insects is not known because observations are lacking.

The effects of these varying factors on feeding can be accounted for by an, as yet hypothetical, excitatory state in the central nervous system first postulated for the blow fly and subsequently elaborated for the migratory locust (Fig. 4). Only when the central excitatory state exceeds a certain threshold can feeding occur, but feeding is not an automatic consequence of reaching the threshold; it is a probabilistic event. At the end of a meal, the central excitatory state is assumed to be depressed below threshold. As time since the previous meal increases, so does the level of the central excitatory state so that it approaches and ultimately exceeds threshold. Rhythmic changes in the central excitatory state are presumed to account for the basic rhythmicity of feeding, and other events, such as defecation, may temporarily elevate it, while others (disturbance) may depress it.

THE SIZE OF A MEAL

The size of a meal, assuming the food supply to be unlimited, depends on the net phagostimulatory effects of the food and the nutrient requirements of the insect. Net phagostimulatory effect is the balance between nutrient components of the food that stimulate taste receptors leading to feeding and any other factors, such as the presence of toxic compounds or undue hardness, that tend to inhibit feeding. For many insects, sugars are major phagostimulants and higher concentrations in the food result in larger meals. Amino acids may also influence the phagostimulatory input. Plant secondary compounds, such as alkaloids, are often feeding deterrents even for insects, such as the blow fly, whose food does not normally contain them. Inorganic salts at higher concentrations are also deterrent although at low concentrations they may stimulate feeding. Phagostimulatory compounds do not just switch on feeding that then continues until the insect is replete; their continued input is necessary for feeding to continue and the behavior of some insects appears to reflect this. Chemosensory receptors usually adapt within a few seconds if continually stimulated, but the palpation behavior of grasshoppers and caterpillars appears to permit a continual flow of information by bringing the receptors into contact with the food for frequent very brief periods. The sensilla on the palp tips of a locust make about 10 contacts per second and each contact may be only 10 to 20 ms. As a result they remain largely unadapted.

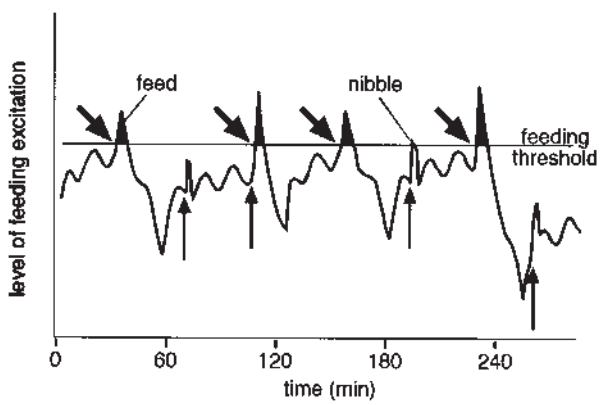

FIGURE 4 Model of the control of feeding in a locust eating wheat. Similar principles are believed to apply to other insects. The irregular line shows the level of feeding excitation (the central excitatory state). When this exceeds a threshold, the insect feeds. Notice that after a meal, the excitatory state declines sharply. Subsequently it rises slowly and the level oscillates with a period of about 15 min. Defecation (upwardly pointing arrows) causes a sudden rise in excitation. If this causes excitation to exceed the threshold, the insect may feed. Biting the food (oblique arrows) releases juices from the food and phagostimulants cause a sharp rise in the central excitatory state. (Reproduced, with permission, from Simpson, 1995.)

Although continual stimulation is important to maintain feeding, meal size seems to be determined by the level of phagostimulation when the insect first bites into its food and releases the internal fluids containing a mixture of stimulating chemicals. This was demonstrated by an experiment in which the mouthpart chemoreceptors of locusts were stimulated with a highly phagostimulatory solution that they were not allowed to ingest. These insects subsequently ate larger meals than others stimulated with water alone, despite the fact that during the meal the receptors of both sets of insects were equally stimulated. Events before feeding started determined how much was eaten. Comparable experiments have shown that distasteful compounds can reduce meal size. Such experiments are interpreted as reflecting changes in the central excitatory state. A high concentration of phagostimulant is believed to elevate the central excitatory state well above threshold and feeding continues, provided some level of input is maintained, until the excitatory state declines to threshold. Thus, the higher the initial level, the longer it takes the excitatory state to reach threshold and the larger is the meal. It is supposed that a high level of deterrent compounds would inhibit feeding by depressing the level of the central excitatory state below threshold.

The elevated level of the central excitatory state is also believed to account for the "dances" of flies and the palpation behavior of locusts and grasshoppers following loss of contact with food early in a meal. When a fly loses contact with a drop of sugar it moves in an irregular path with frequent turns as if "searching" for the food. The more concentrated the solution, the more frequent the turns (Fig. 5A). If a locust loses contact with its food it palpates vigorously and such behavior lasts longer if loss of contact occurs earlier in a meal. Toward the end of a meal, loss of contact with the food results in only a limited period of palpation (Fig. 5B). The so-called searching behavior of other insects, such as that described for coccinellid larvae when feeding on aphids and temporarily losing contact with the prey, probably has a similar basis.

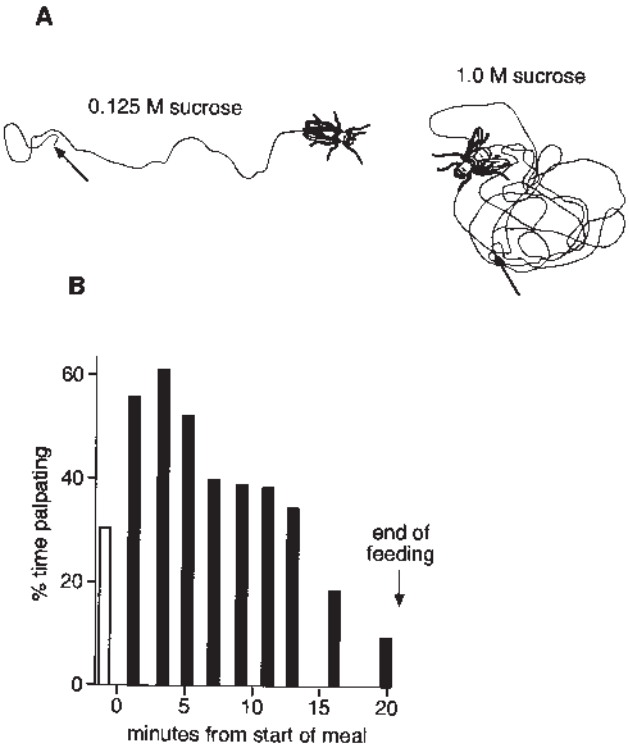

FIGURE 5 Searching for food after loss of contact during a meal. (A) Blow flies dance when they lose contact with a drop of sugar. After experiencing more concentrated sugar solutions the high rate of turning is much more sustained. [Reprinted, with permission, from Dethier, V. G. (1957). Communication by insects: Physiology of dancing. *Science* **125,** 331–336. Copyright 1957 American Association for the Advancement of Science.] (B) The migratory locust palpates when it loses contact with a blade of grass. Each bar represents the percentage of time palpating after losing contact with the food at a different stage of the meal. Soon after the start of a meal it palpates for most of the time, but toward the end of a meal it is less persistent. The open bar represents the time palpating just before starting to feed. [Reprinted, with permission, from Bernays, E. A., and Chapman, R. F. (1974). The regulation of food intake by acridids. *In* "Experimental Analysis of Insect Behaviour" (L. Barton Browne, ed.). Springer–Verlag, Berlin. Copyright Springer–Verlag.]

THE END OF A MEAL

The end of a meal is ultimately determined by the degree of distension of the part of the gut in which the food is temporarily stored. In grass-feeding grasshoppers and nectar-feeding flies, this temporary store is the crop (part of the foregut). In *R. prolixus,* feeding on vertebrate blood, the food is stored in the anterior midgut. In each case, the degree of distension is monitored by some form of stretch receptor. In locusts, these receptors are multipolar cells on the wall of the foregut, and receptors on the most anterior part of the foregut, which is the last part to fill; these receptors are responsible for inhibiting further feeding. *R. prolixus* has chordotonal organs in the body wall. The input from these stretch receptors has an inhibitory effect on feeding, presumably by leading to a decline in the level of the central excitatory state to below threshold. If food quality and the nutritional and feeding status of the insect are constant, stretch receptor input determines that an insect ingests a similar amount of food at each meal.

CHANGES IN FEEDING BEHAVIOR

The pattern of feeding changes with the age of the insect, its previous experience, and its nutritional needs. Phytophagous insects in general tend to eat greater amounts in the middle of a developmental stage and more in the light than in the dark. The average meal size taken by the final-stage nymph of the migratory locust, for example, increases from about 50 mg on the day of molting to almost 100 mg 4 days later, whereas the average interval between meals declined from 82 to 71 min. At night, the insects take fewer meals even though the temperature may be constant.

Some phytophagous insects become less selective if they experience a long period without food and this has given rise to some confusion in the literature. For experimental purposes, it is often convenient to use insects that feed readily when presented with food. This is achieved by depriving them of food, often for 24-h periods. However, because such insects are less selective than insects with continual access to food, grasshoppers, for example, were generally considered to be unselective in their choice of foods. More critical observations, however, show that this is not accurate. With increasing periods of food deprivation, several grasshoppers have been shown to accept a wider range of food plants. It is probable that this acceptance of previously unacceptable plants reflects a need for water rather than for other specific nutrients, although this hypothesis has not been thoroughly investigated. It is, however, clear that a well-hydrated locust actively moves away from wet filter paper, whereas a dehydrated one attempts to eat it. Similarly, dehydrated flies drink water, whereas hydrated ones do not.

The tendency of grasshoppers and caterpillars, and probably other insects, that are deprived to sample food that would otherwise be rejected can play a major part in the subsequent acceptance of food. This becomes possible because taste receptors that initially signaled rejection because of some distasteful component of the food become habituated and are no longer stimulated by the distasteful compound. At the same time, detoxifying enzymes are probably mobilized within the insect, providing it with the capacity to minimize any harmful effects that the compound might have.

The nutritional requirements of insects vary through life and this is reflected by changes in their feeding behavior. During larval development, the amount of food consumed is usually maximal in the middle of each developmental stage, falling to zero for a period before each molt. Changes also occur in adults in relation to somatic development and, in females of many species, in relation to egg development. This variation is illustrated for adult red locusts *(Nomadacris septemfasciata)* in Fig. 6. When the insect first becomes an adult the cuticle is soft and the flight muscles are poorly developed. During this teneral period, both sexes feed actively

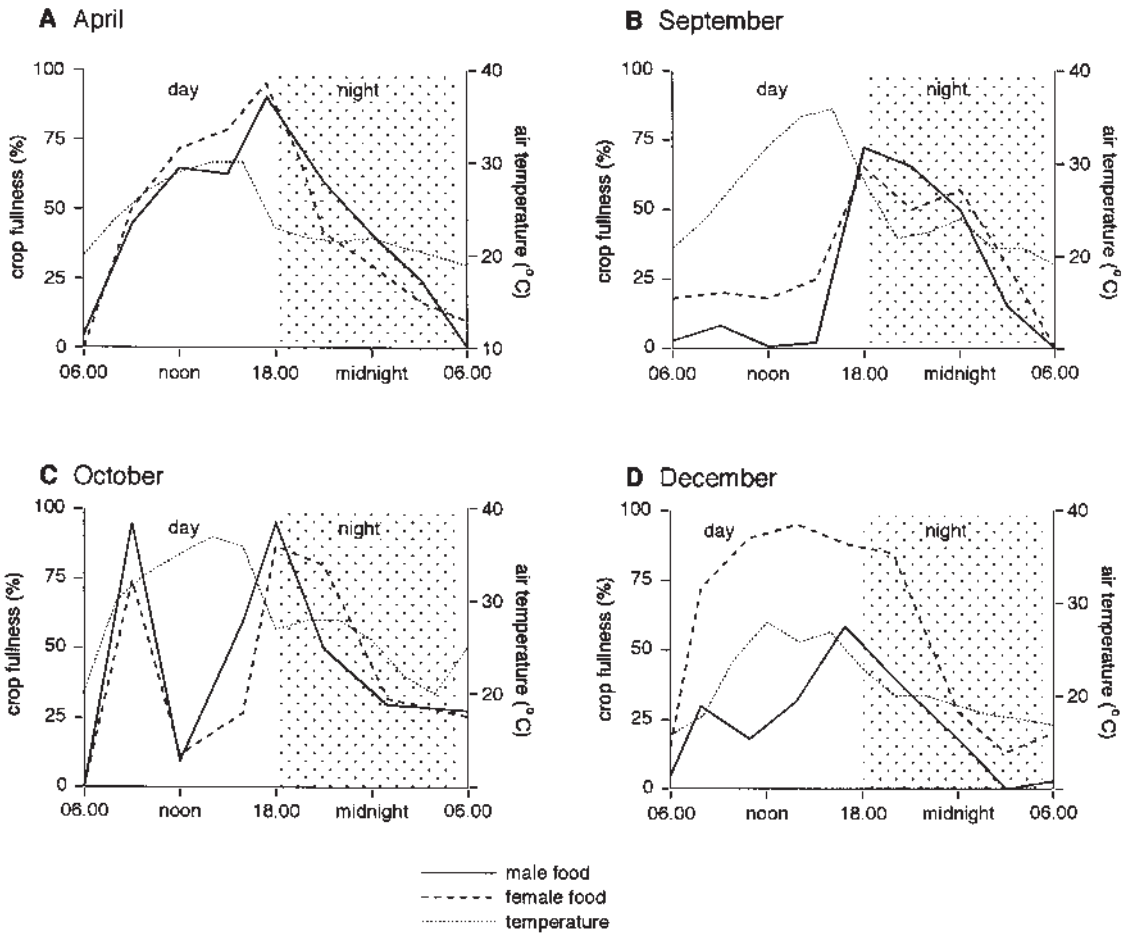

FIGURE 6 Variation in feeding behavior during adult life. The red locust, *N. septemfasciata,* in the field. (A) April. Soon after becoming adult both sexes feed for most of the day. This is a period of somatic growth when flight muscles and cuticle become fully developed. (B) September. Despite moderately high temperatures during the day, very little feeding occurs until late afternoon. The insects are in reproductive diapause. (C) October. Little feeding occurs in the middle of the day, perhaps because of the high temperature. The insects are beginning to become sexually mature. (D) December. Females eat much more than males during the period of egg development. All these samples were taken from the same generation and population of insects, which live for about 9 months as adults. Each graph shows the amount of food in the foreguts of a sample of insects taken at each time point over a 24-h period; 100% would indicate that all the locusts were full, 0% that they were all empty. When the temperature is 30°C or above, the foregut becomes more than half empty within an hour, so that crop fullness above 50% during the day indicates recent feeding. At 25°C and below, the food takes several hours to leave the foregut so that night time values largely reflect feeding before dark. [Reproduced, with permission, from Chapman, R. F. (1957). Observations on the feeding of adults of the red locust. *Br. J. Anim. Behav.* **5,** 60–75.]

during the day (Fig. 6A). Subsequently, the insects enter reproductive diapause and feeding is reduced to a single meal each day (Fig. 6B). During the reproductive period, females eat much more than males (Fig. 6D).

Among some adult flies and grasshoppers, there is good evidence that mature females change their feeding behavior to acquire protein for the synthesis of vitellogenin. This is most obvious in blood-sucking flies, such as mosquitoes and tabanids, females of which use nectar as a flight fuel, but vertebrate blood as their primary protein source. Males of these same species feed only on nectar. This is also true of blow flies. Mature female grasshoppers, given the opportunity, tend to select food with a higher protein level than do males or immature females. Thus, they tend to eat the seed heads of developing grain rather than foliage.

Under laboratory conditions, when fed on artificial diets, locusts and caterpillars are able to select from the foods with different amounts of proteins and carbohydrates to maintain an appropriate balance of the two classes of compound. Locusts can make the adjustment from one meal to the next, with an interval of less than an hour between meals. The extent to which insects can fine tune their nutritional balance when feeding on natural food with much smaller deficiencies of protein or carbohydrate has yet to be demonstrated.

FEEDING BEHAVIOR UNDER NATURAL CONDITIONS

In the field, feeding behavior is determined to a large extent by environmental factors, although relatively few extensive studies have been carried out. Temperature has a major effect on feeding behavior, as it does on other insect activities, with

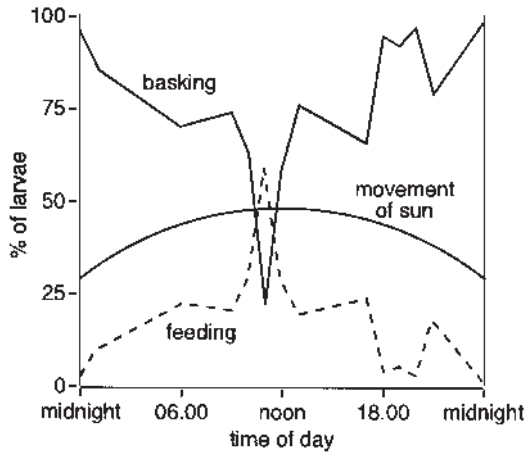

FIGURE 7 Feeding is limited by temperature. Most caterpillars of *G. groenlandica,* living within the Arctic Circle, feed during a 2-h window when the sun is at its zenith. For most of the time, the insects bask to raise their body temperatures, enabling them to feed efficiently; during feeding their temperature falls rapidly. [Reproduced, with permission, from Kukal, O., Heinrich, B., and Duman, J. G. (1988). Behavioural thermoregulation in the freeze-tolerant Arctic caterpillar, *Gynaephora groenlandica. J. Exp. Biol.* **138,** 181–193. Copyright Company of Biologists.]

little feeding occurring at low or at very high temperatures (Figs. 6A and 6C). The effects of temperature are most obvious in insects living under extreme conditions of low or high temperature. For *Gynaephora groenlandica* caterpillars living within the Arctic Circle, feeding is possible only when the insect has raised its body temperature by basking. As a result, most feeding occurs in a relatively narrow window of time around noon each day, when the sun is highest in the sky (Fig. 7).

Darkness also tends to reduce feeding. For many visually foraging insects, finding food at night is impossible, although night-flying moths obtain nectar only during darkness and some blood-sucking insects, such as mosquitoes, feed most actively at night or in the crepuscular periods. These insects locate their host primarily by odor, although night-blooming flowers often also present conspicuous targets because of their size and whiteness.

Biotic factors may also have a profound effect. For example, a caterpillar of *M. sexta* that has defended itself from the attack of a tachinid fly does not feed for some time after it has successfully repelled the attacker.

See Also the Following Articles

Blood Sucking • Mouthparts • Salivary Glands

Further Reading

Bernays, E. A. (1997). Feeding by lepidopteran larvae is dangerous. *Physiol. Entomol.* **22,** 121–123.

Bernays, E. A., and Simpson, S. J. (1995). Control of food intake. *Adv. Insect Physiol.* **16,** 59–118.

Chapman, R. F., and de Boer, G. (eds.) (1995). "Regulatory Mechanisms of Insect Feeding." Chapman & Hall, New York.

Dethier, V. G. (1976). "The Hungry Fly." Harvard University Press, Cambridge, MA.

Simpson, S. J. (1995). Regulation of a meal: Chewing insects. *In* "Regulatory Mechanisms of Insect Feeding." (R. F. Chapman and G. de Boer, eds.), pp. 137–156. Chapman & Hall, New York.

Simpson, S. J., Raubenheimer, D., and Chambers, P. G. (1995). The mechanisms of nutritional homeostasis. *In* "Regulatory Mechanisms of Insect Feeding" (R. F. Chapman and G. de Boer, eds.), pp. 251–278. Chapman & Hall, New York.

Stoffolano, J. G. (1995). Regulation of a carbohydrate meal in the adult Diptera, Lepidoptera, and Hymenoptera. *In* "Regulatory Mechanisms of Insect Feeding" (R. F. Chapman and G. de Boer, eds.), pp. 210–247. Chapman & Hall, New York.

Fire Ants

Les Greenberg
University of California, Riverside

Most residents of the southeastern United States are very familiar with fire ants. These reddish brown ants are well known for their aggressiveness and stings that produce a burning sensation. The term "fire ant" actually applies to a group of New World ant species in the genus *Solenopsis.* Many people refer to them as "red ants," although this term is also used to refer to the larger red harvester ants found in desert climates. In Spanish, the fire ant is sometimes called *hormiga colorada,* and in Portuguese it is *formiga de fogo.* In North America there are four native fire ant species, two introduced species, and two hybrid forms. The two imported species in the United States are the red imported fire ant, *Solenopsis invicta,* and the black imported fire ant, *S. richteri.* The former has spread throughout the southeastern part of the country, whereas the latter is restricted to northeastern Mississippi, northern Alabama, and northwestern Georgia. Seventeen fire ant species are currently described from South America.

ORIGIN AND SPREAD

There seems little doubt that the most important fire ant pest, *S. invicta,* traveled from South America to Mobile, Alabama, in ship ballast between 1930 and 1940. It spread in all directions from there, limited only by cold winters or desert drought conditions. *S. richteri* may have arrived earlier (perhaps in 1919), only to be largely displaced by *S. invicta.* The latter is currently established in 11 states (Florida, Georgia, Alabama, Mississippi, Louisiana, Texas, North Carolina, South Carolina, Arkansas, Tennessee, Oklahoma) and Puerto Rico. A new infestation was discovered in California in 1998, and eradication efforts are under way.

It was predicted that imported fire ants could not survive a winter when the minimum temperature was below –12°C. However, infestations now occurring in southern Oklahoma and Tennessee have led to a revised estimate of –18°C. Another limiting factor is rainfall. The deserts of west Texas

have proven a barrier to the fire ant's progression westward. It is likely that annual rainfall of less than 25 cm precludes *S. invicta* from becoming established unless there is also irrigation.

IDENTIFICATION AND BEHAVIOR

All fire ants have two segments in their narrow waist and antennae with 10 segments. The workers range in size from small to large (for *S. invicta,* about 2–5 mm in length; queens are about 7 mm long). To be able to sting, fire ants must first gain leverage with their mandibles by biting; they then curve around the abdomen to insert the stinger. The fire ant injects venom consisting mainly of piperidine alkaloids that produce a burning sensation. Shortly thereafter, a red spot is usually visible. The burning sensation is short-lived, followed by itching. In most people, a white pustule will develop at the site within a few hours. These pustules are sometimes called "sterile pustules" because they are not produced by infectious bacteria. The pustules can last from days to weeks and can become infected if they are scratched. The venom also contains a small amount of protein (about 1%) that can cause anaphylactic shock in susceptible individuals. Fire ants can sting repeatedly; therefore, stinging ants should be brushed off rapidly.

Mounds and Foraging Behavior

When undisturbed, the typically dome-shaped mounds of *S. invicta* can reach heights of 30 cm or more above the ground. These mounds allow the workers to respond to local conditions by moving up and down with their brood and queens according to temperature and humidity. Exit holes are usually not apparent on the mounds themselves, but foraging trails extend outward from the mound just below the surface. During floods, fire ants move to the upper parts of the mound. If the water gets any higher, the ants grasp each other to form floating rafts that carry the brood and queens downstream. During droughts, fire ants can extend their tunnels down 6 m or more in search of moisture.

Fire ant workers can feed only on liquids: they have filters in their digestive tract that prevent the ingestion of solids. Only the fourth instar can digest solids directly, and it is the only path for processing of solid food particles in the colony. Workers deposit insect parts and other solids on the larva's "food basket." After feeding on these solid foods, the larva secretes liquids that are licked up by workers and distributed around the colony.

Mating Behavior

Like most other ants, fire ants have mating flights. In the southeastern United States, flights are most frequent in the spring following rain and subsequent sunshine. After the rain, workers fill the queens with food to prepare them for the flight. Dissection of queens at this time shows a large drop of yellow oil in their crops. In midafternoon the workers open large exit holes in the mounds to allow quick exit of the males and females. Workers become very agitated and start chasing the reproductives, which then climb vertical objects nearby from which they fly. Mating occurs in the sky. The males drop to the ground and die shortly thereafter. The queens also land, quickly shed their wings, and search for a place to dig a tunnel. The queen will close the tunnel and start to lay eggs, producing her first workers in about one month. Queens typically live 6 or more years. Because they mate only once, they must store live sperm for the rest of their lives. For this purpose, they have a transparent sac in their abdomens called the spermatheca that is filled with over a million sperm after mating. When the queen lays an egg, she can open a valve on the spermatheca, allowing the escape of sperm to fertilize her eggs. These diploid eggs give rise to females, either workers or new queens. If she does not release sperm, the egg she lays is haploid and becomes a male, as is typical for all haplodiploid social Hymenoptera.

Number of Queens

There are two forms of *S. invicta* in the United States. Originally this species was thought to be monogyne, having one queen per colony. Polygyne, or multiple-queen fire ant colonies, were first described from Mississippi in the early 1970s. In these colonies there can be dozens of fertilized queens. The queens are not aggressive toward each other and are frequently together in one part of the nest. One or more of these queens may be dominant, laying more eggs and receiving more food than the others. Polygyne queens in a colony are not closely related, suggesting that they are adopted from outside sources. They are also smaller on the average than monogyne queens and lay fewer eggs. However, the total number of eggs laid by all the polygyne queens in a colony exceeds that produced by a single monogyne queen. On the average, polygyne colonies also have smaller workers: there is a negative correlation between the number of queens and the average worker size. Furthermore, the polygyne form is not aggressive toward conspecifics, whereas the monogyne form will fight with nearby conspecifics. The polygyne form can bud off new colonies of queens and workers and is thereby able to quickly populate an area with fire ants. The polygyne form is now predominant in Texas and has also been found in Florida, Georgia, and even South America.

IMPACT ON PEOPLE AND THE ENVIRONMENT

Imported fire ants arrived in the United States without their native parasites and predators. In addition, few native ant species are able to compete with them. When fire ants encounter other ant species they vibrate their gasters (abdomens) and protruded stingers, spraying their venom and chasing away most other ants. They can displace native

FIGURE 1 A child's arm showing many pustules after numerous fire ant stings. (Photograph courtesy of the U.S. Department of Agriculture.)

ant species and quickly become the dominant ant species and significant pests. Fire ants will attack newly hatched birds both on the ground and in trees. Fawns and calves can be stung in their eyes and blinded if they are dropped on fire ant mounds. Soft plant tissue, such as okra, can be destroyed. Ants in mounds at the bases of trees can eventually girdle and kill trees. The sick and elderly, pets and children in backyards, campers, and picnickers, all can become victims of fire ant stings (Fig. 1). One oddity about fire ants is their evident attraction to electrical fields. They frequently enter electrical boxes such as outside air conditioners, traffic boxes, and lights, where they chew wires and short out the circuits.

POSSIBLE REMEDIES

There are many insecticides that control fire ants. Baits are advantageous because it is not necessary to find the mounds; ants carry the bait back to their nests. Most fire ant baits consist of corncob grits coated with soybean oil as an attractant. Typically, a toxicant or insect growth regulator is dissolved in the oil. However, none of these solutions are permanent. Efforts are now under way to bring into the United States some of the fire ant parasites and predators from South America. Decapitating flies (Phoridae) are one promising predator now being released in Florida and elsewhere. These flies lay their eggs on fire ants. The grub that hatches invades the ant's head, where it consumes its brain. Ultimately the ant's head falls off and a new fly emerges. Although these flies parasitize only a small percentage of ants, they do interfere with the ant's foraging behavior and may make fire ants less competitive with other ants. Other potential biocontrol agents include protozoal parasites, soil fungus, and even parasitic ants from South America. Ultimately a broad approach using chemicals and biological agents will best manage this invasive species.

See Also the Following Articles

Ants • Introduced Insects • Regulatory Entomology • Venom

Further Reading

Greenberg, L., Fletcher, D. J. C., and Vinson, S. B. (1985) Differences in worker size and mound distribution in monogynous and polygynous colonies of the fire ant *Solenopsis invicta* Buren. *J. Kans. Entomol. Soc.* **58,** 9–18.

Porter, S. D. (2000). Host specificity and risk assessment of releasing the decapitating fly *Pseudacteon curvatus* as a classical biocontrol agent for imported fire ants. *Biol. Control* **19,** 35–47.

Trager, J. C. (1991). A revision of the fire ants, *Solenopsis geminata* group (Hymenoptera: Formicidae: Myrmicinae). *J. N.Y. Entomol. Soc.* **99,** 141–198.

Tschinkel, W. R. (1998). The reproductive biology of fire ant societies. *Bioscience* **48,** 593–605.

Vander Meer, R. K. (1996). Potential role of pheromones in fire ant control. *In* "Pest Management in the Subtropics: Integrated Pest Management: A Florida Perspective" (D. Rosen, F. D. Bennett, and J. L. Capinera, eds.), pp. 223–232. Intercept Ltd., Andover, U.K.

Vargo, E. L., and Hulsey, C. D. (2000). Multiple glandular origins of queen pheromones in the fire ant *Solenopsis invicta. J. Insect Physiol.* **46,** 1151–1159.

Vinson, S. B. (1997). Invasion of the red imported fire ant (Hymenoptera: Formicidae): Spread, biology, and impact. *Am. Entomol.* **43,** 23–39.

Flea

see *Siphonaptera*

Flight

Michael Dickinson
California Institute of Technology

Robert Dudley
University of California, Berkeley

From their first appearance in the late Paleozoic, winged insects have emerged as critical components of nearly all terrestrial ecosystems. Many important behavioral features of insects, including evasion of predators, dispersal, and reproductive strategies, rely in some way on flight behavior. Wings themselves, as cuticular structures, have no intrinsic musculature and are moved instead via thoracic deformations and by muscles that insert directly at the wing base. The back and forth motion of the wing through the wingbeat, as well as the rotation about its longitudinal axis at the beginning

and end of each stroke, creates unsteady aerodynamic forces of continuously changing direction and magnitude. Flight control relies on multiple sensory modalities to maintain stable trajectories and to maneuver via bilaterally asymmetric motions of the wings and body. Flight is energetically costly, and the delivery of oxygen to flight musculature is limited by diffusion in the fine branches of the tracheal system. Heat simultaneously produced by contraction of inherently inefficient muscles may be co-opted in the regulation of body temperature during flight to further enhance performance. Forces of both natural and sexual selection have contributed synergistically to the evolution of insect flight performance and maneuverability. Contemporary insect diversity largely comprises extensive radiations of miniaturized species. Flight biomechanics of these small insects is complicated by the viscous nature of airflows and depends in part on the high wingbeat frequencies enabled by a specialized muscle type termed asynchronous muscle. Asynchronous flight muscle has evolved independently more than eight times among the winged insects and enables muscles to generate increased mechanical power by trading sarcoplasmic reticulum for more contractile fibrils and mitochondria. The majority of insect species-level diversity appears to derive indirectly from this flight-related innovation that facilitates miniaturization.

EVOLUTION OF FLIGHT

Although many features have contributed to the radiation of insects in terrestrial ecosystems, the evolution of actively powered flight is almost certainly the key innovation responsible for their remarkable success. The relative abundance of extant winged (ptergygote) insects to wingless (apterygote) insects (a ratio of at least 500,000:1 in species richness) manifests the potent advantages of flight. Although the selective advantages of flight are obvious, the means by which ancestral hexapods evolved wings and associated flight behavior are not. Because flight is such a specialized form of behavior and is associated with morphological and physiological traits that represent extreme forms of the basic arthropod body plan, reconstructing the series of functional intermediates between flightless ancestors and flying insects continues to pose a challenging problem. The evolution of flight involves two distinct, but overlapping, questions. First, what is the morphological structure from which wings arose? And second, what suite of selective forces drove the evolution of wings as aerodynamic structures?

Morphological Origin of Wings

The morphological origin of the wing in pterosaurs, birds, and bats is unequivocal; in all these animals it arose from a modification of the forelimb. Insect wings are novel structures, at least in the sense that they are not homologous with the legs. Biologists have long debated which structure served as the anatomical precursor of insect wings. Of the various theories that have been proposed, the dominant view until recently was that wings arose from rigid lateral extensions of the notum. Such a scenario seems at first plausible, given the structural plasticity of the thoracic exoskeleton in extant insects. When considering function of flight morphology as a whole, however, the most complicated feature of the wing is not the flat distal blade that serves as the aerodynamic surface, but rather the complicated hinge with its associated muscle attachments that enables the wing to flap and rotate during the stroke. Over the past 15 years an alternative hypothesis, that the wings evolved from basal branches of the leg, has emerged from work in a number of disciplines. This theory owes much to the work of the paleontologist Jarmila Kukalova-Peck, who challenged the widely held view that insects, as distinct from other arthropods, possess unbranched limbs. According to her alternative view, the ancestors of winged insects possessed biramous appendages and used a developmental-genetic program for limb development that they shared with crustaceans and other arthropods. The structure that gave rise to the wing may have been a dorsal branch, or exite, of a precoxal segment of the leg called the epicoxa. Whereas the epicoxa has been lost or incorporated into the pleurum of the thorax, its exite has been retained as a wing. However, rather than classifying the wings of extant insects as direct morphological homologues of epicoxal podites in ancestral apterygotes, it may be more precise to view both as arising from homologous morphogenetic programs. The leg podite theory of wing origin solves an enigmatic step in the evolution of functional wings, the formation of the wing hinge and its complex arrangement of muscles. As a leg branch, the protowing would have been endowed with joints and muscles long before it ever took on an aerodynamic role. Further, because legs are replete with various mechanosensory structures, the protowing would have inherited the campaniform sensilla, stretch receptors, or chordotonal organs that may have mediated the reflexes and motor patterns that presumably served as the foundation of flight control circuitry. Although several lines of evidence support a leg podite origin for insect wings, this intriguing issue is far from resolved and the consensus may change with additional fossil evidence and further comparative studies of arthropod development.

Functional Origin of Flight Behavior

Hindered by the inherent difficulty of extracting behavior and physiology from fossil evidence, the functional origin of flight remains enigmatic. The fact that wings arose from small structures poses the same problem that Darwin first recognized for all organs of great complexity—it is difficult to reconstruct a series of functional intermediates between a tiny leg podite and an aerodynamic surface capable of sustaining active flight. The aerodynamic performance of a wing increases with length and surface area. Thus, a small wing is incapable of generating enough force to sustain active flight. Without a selective

pressure driving the wing to larger sizes, how did the structure initially attain the size required to support active flight? It is unlikely that any single selective pressure was responsible for the hypertrophy of the wings. For example, if the direct ancestors of pterygotes possessed an aquatic nymph stage, protowings might have served as gill covers or respiratory paddles. Given the high density and buoyancy offered by an aquatic medium, it is even possible that wings may have functioned as hydrodynamic structures for underwater propulsion. However, no matter what role they may have played in the aquatic stage of life history, the use of wings in air would necessitate a substantial increase in size.

Hypotheses attempting to explain the early selective engine for true aerial flight segregate into two basic types. One set of hypotheses suggests that early selective pressures for an increase in wing size had nothing to do with aerodynamics per se, but rather with some other size-dependent selective force. For example, the use of wings as reflectors and conduits in basking butterflies has led to the proposal that wings first served a thermoregulatory role. Other possibilities include the use of wings in sexual displays or copulatory offerings by males. The second set of hypotheses asserts that protowings functioned aerodynamically before they were large enough to support active flight. For example, small wings might serve to increase glide angle or offer added stability during controlled descents. The utility of small protowings in gliding behavior might have been enhanced by their serial repetition, and fossil evidence indicates that protowings were present on the prothorax and abdominal segments in some groups of early insects. Vegetation and surface topography would have served as the most convenient launching points for gliding or parachuting insects. Another possibility is that protowings may have prolonged jumps, thereby serving as an important anti-predator behavior in response to the coevolutionary radiation of terrestrial predators at the time. Recently, James Marden suggested that protowings may have served as aerodynamic structures used to either sail or flutter insects across the surfaces of streams and ponds. This intriguing hypothesis is based on the behavior of extant stoneflies that skim across streams in this manner when the temperature is too low for their flight muscles to generate sufficient mechanical power to sustain flight. The atmospheric composition at the time, in which both oxygen levels and air density were elevated by today's standards, might also have aided the transition to active flight (see later). Whatever selective pressures led to the evolution of flight, analyses of insect phylogeny strongly suggest that flight evolved only once within the clade. However, no behavior that has been proposed as a model for ancestral pterygotes, such as sun basking or surface skimming, maps into the current phylogeny in a way that is entirely consistent with it being an ancestral trait. With no definitive means of excluding any of the proposed scenarios, the functional origins of insect flight are likely to remain alluring, controversial, and unresolved for years to come.

AERODYNAMICS

Conventional Aerodynamics

The scientific study of insect flight is haunted by the widely told story of an engineer who proved that a bumblebee could not fly. Although the flight of insects is indeed more complicated than that of airplanes, the underlying physics is nevertheless fully explicable within the rubric of modern fluid mechanics. To understand how insects fly by flapping their wings, it is useful to first consider the means by which fixed-wing aircraft create aerodynamic forces. The design of conventional airplanes is based on the steady-state principle that the flow of air around the wings and the resulting forces generated by that flow are stable over time. As the wing of a plane moves through the air, it meets the oncoming flow at a small inclination, termed the angle of attack. As the flow of air approaches the leading edge of the wing, it divides into two streams on the undersurface of the wing. Because of the viscous behavior of air (a general property of all "fluids," including liquids and gases), the two streams meet again smoothly at the sharp trailing edge. For the flow to separate under the wing, but meet again at the trailing edge, the upper stream must travel faster than the lower because it covers a greater distance. By Bernoulli's principle, this higher velocity generates lower pressure, which sucks the wing upward producing lift.

Although the explanation of flight based on Bernoulli's principle is sufficient for simple situations, engineers and physicists often use a mathematical transformation to quantify the velocity difference above and below the wing and analyze more complex situations. Subtracting the background flow caused by the speed of the airplane from the local flow near the wing uncovers a net circular movement of air around the wing called vorticity. Cohesive filaments or loops of vorticity are called vortices, a term that also applies to more familiar flow structures such as tornadoes, whirlpools, and smoke rings. Although the net circular flow of air around a wing is a mathematical abstraction, wings are, in effect, vortex generators. At a low angle of attack a wing creates a bound vortex, so named because the center of vorticity is located within the wing. The Kutta–Joukowski theorem, perhaps the most essential equation in aerodynamics, states that the lift generated by each section of a wing is proportional to the strength of the vorticity it creates, a quantity termed circulation. The simplest way of increasing the amount of circulation, and thus the lift, is to increase the angle of attack. At angles of attack above about 10°, however, the flow over the top surface separates as it rounds the leading edge, resulting in a catastrophic loss of lift known as stall. For a wing operating according to conventional aerodynamics, the stall angle places an upper limit on the amount of stable circulation, and thus lift, that a wing can continuously generate. Early analyses of insect flight aerodynamics applied conventional steady-state theory unto the complex motion of flapping wings. This approach, termed quasi-steady theory, is equivalent to "freezing" the wing at one position within the stroke cycle

and then testing it at that particular velocity and angle of attack in a wind tunnel under steady flow. If conventional theory were sufficient, then a series of such measurements repeated for each point in the stroke cycle should sum up to the animal's body weight. In most cases such simple quasi-steady approaches cannot account for the forces required to sustain flight, indicating that unsteady aerodynamic mechanisms play an important role in insect flight.

Scaling Parameters

Before discussing such mechanisms in detail, it is useful to introduce two important parameters that help organize the great diversity of flight patterns in insects. The first term, the Reynolds number, quantifies how changes in body size, wingbeat frequency, and atmospheric conditions affect aerodynamic mechanisms. The wing or body of an insect encounters two forces as it moves through the air, a shear force caused by fluid viscosity and an inertial force from the fluid momentum. The dimensionless Reynolds number is simply the ratio between these two forces and, for insects, is equal to the product of wing velocity, wing length, and air density divided by air viscosity. Reynolds numbers vary among insects from about 10 for the tiniest to 10,000 for the largest insect. At high Reynolds numbers, the inertial behavior of the air dominates and wings generate pressure forces acting perpendicular to their surface. At a Reynolds number less than 1, a viscous shear force dominates, acting parallel and opposite to the direction of motion. Recent measurements of force production by flapping wings indicate that aerodynamic performance is remarkably constant across a range of Reynolds numbers spanning from about 100 to 5000—encompassing the operating range of most insects. Nevertheless, miniaturization is a common theme in insect evolution, and many species are so small that viscous forces, if not dominant, are large enough to greatly influence force production. The functional peculiarities of lower Reynolds numbers are manifest in the unique wing morphology of the smallest insects, including the brush-like wings of thrips and the whip-like wings of some miniaturized beetles. Although the kinematics used by these tiny insects is as yet unknown, it is possible that they flap their "wings" in such a way as to generate an excess of viscous drag during the downstroke, akin to the power strokes of aquatic plankton. Reynolds numbers are also used to construct large mechanical models of flapping insects for the purpose of directly measuring aerodynamic forces and visualizing flow. This technique, termed dynamic scaling, is based on the principle that the fluid-based forces acting on two geometrically similar but different-sized objects are the same as long as the Reynolds numbers are identical.

Another important dimensionless parameter, called the advance ratio, is useful in coarsely assessing whether conventional steady-state aerodynamics is sufficient to explain force production. The advance ratio is simply the animal's airspeed divided by the flapping velocity of its wings. At one extreme, an infinitely high advance ratio indicates that an animal is gliding, and all the air flowing past the wings derives from the motion of the body as a whole, which is a condition amenable to conventional steady-state aerodynamics. Even if the wings flap up and down, steady-state approximations may be valid as long as the forward speed is substantially greater than the velocity of the wings. The situation is much more complicated for hovering or near-hovering conditions, in which the insect is essentially stationary and most of the airflow encountered by the wings is generated by their back-and-forth motion. Under these conditions, the flow of air around the wing changes substantially throughout the stroke, and the analysis of aerodynamic forces is more complex. Low advance ratio flight is typical of many insects, particularly those of small body size, and is characterized by a motion in which the wings flap back and forth in a roughly horizontal plane. During the two strokes (somewhat inappropriately named the downstroke and upstroke), the wings translate through the air at high angles of attack creating elevated vorticity. At the end of each stroke, the wings rapidly flip over such that the dorsal surface of the wing faces upward during the downstroke, and the ventral surfaces faces upward during the upstroke. As it flips, the wing sheds the vorticity it created in the previous stroke, thereby adding to a complex vortex wake that forms underneath the stroke plane akin to the downwash beneath a hovering helicopter. Thus, at the start of each stroke the wing travels not through still air, but through its own wake. These three peculiarities of wing motion during flapping flight, (1) the high angle of attack during translation, (2) the rapid rotation between strokes, and (3) the influence of the wake on subsequent flow of air around the wings, all profoundly influence the manner by which insects create and modify aerodynamic forces.

Aerodynamic Mechanisms

The total force created throughout a stroke by a flapping wing may be conveniently separated into four main components: translational force, rotational force, wake capture, and inertial force (Fig. 1). Inertial force results from the acceleration of the wing back and forth during each stroke. Although the mass of the wings is small, the acceleration is great and the resulting inertial forces are substantial. Peak values during stroke reversal may be many times greater than the aerodynamic forces. However, because the flapping motion is largely sinusoidal, wing inertia averages close to zero over each stroke and thus does not contribute to the average forces acting on the body. Another component of inertial force derives from the acceleration of the air displaced by the wing as it accelerates, termed virtual mass. Although the precise volume of air disturbed by an accelerating or rotating wing is difficult to calculate, conservative estimates indicate that added mass inertia is relatively small compared with the wing mass inertia and other aerodynamic components. Thus, although wing and virtual mass inertia may complicate the

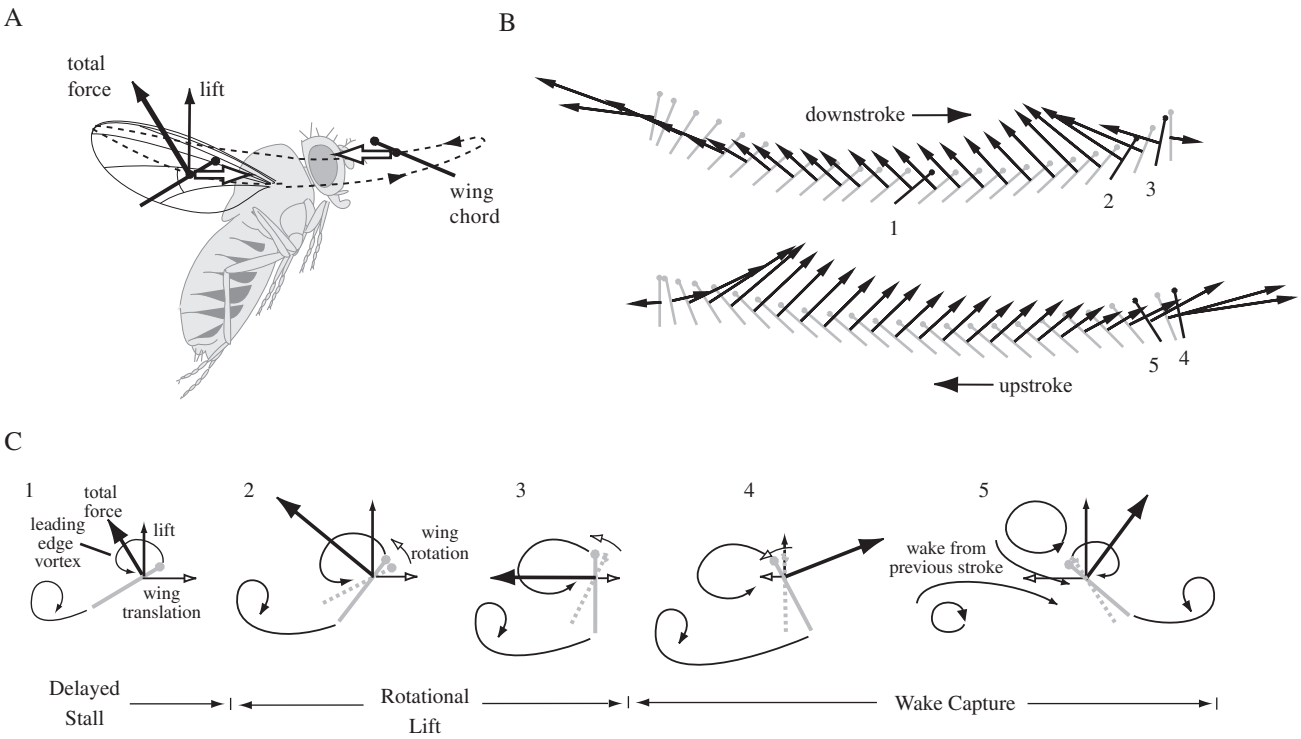

FIGURE 1 Summary of aerodynamic mechanisms used by insects. (A) Under hovering or near-hovering conditions, insects flap their wings back and forth at a high angle of attack during each stroke. The wing path is shown by the dotted line; white arrows indicate wing motion. Between strokes, the wing rapidly rotates so that the dorsal surface faces up during the downstroke, while the ventral portion faces up during the upstroke. The total aerodynamic force (thick arrow) acts perpendicular to the surface of the wing and may be decomposed into orthogonal lift and drag components (thin arrows). (B) Diagram of wing motion indicating magnitude and orientation of the total aerodynamic force vector (black arrows) generated throughout the stroke. Gray lines indicate instantaneous position of the wing at temporally equidistant points during each stroke. Small circles indicate the leading edge of the wing. Time moves left to right during downstroke and right to left during upstroke. (C) Drawings indicate air flow around wing and resulting forces at points within the wing stroke (as indicated in B). Delayed stall (1) results from the formation of a leading edge vortex on the wing. Rotational lift (2 and 3) results from the rapid rotation of the wing at the end of the stroke. Wake capture (4 and 5) results from the collision of the wing with the wake shed during the previous stroke.

precise time course of force production, they are typically ignored in biomechanical analyses of insect flight.

The salient feature that distinguishes the translational forces of insect wings from those generated by airplanes, helicopters, and most birds is that the wings flap back and forth at relatively high angles of attack (30–45°). At such high angles, the stream of air separates from the leading edge of the wing, forming a characteristic flow structure called a leading edge vortex. The lift resulting from the leading edge vortex is much greater than that produced by the bound circulation generated at lower angles of attack. This transient increase in lift at the start of motion at high angles of attack, termed delayed stall, was first recognized by aeronautics engineers in England in the early 1930s, but is too brief to be of use to most aircraft. After only a few moments, the vortex structure grows too large and is shed into the wake, resulting in a precipitous drop in lift. Insects, however, can make use of the initial stages of stall because their wings strokes are so brief. At the end of each stroke the wing sheds the vortex, rotates, and develops a new leading edge vortex swirling in the opposite direction. In addition, complex three-dimensional aspects of the flow, such as a base-to-tip helical flow of air through the center of the vortex, may in some cases remove energy from the structure, enabling it to remain attached to the wing throughout the stroke.

Rotating objects are themselves sources of circulation and concomitant lift production, which is why a tennis ball hit with backspin rises as it moves through the air. The physics of circular balls and flat wings is somewhat different because the wing has sharp leading and trailing edges, but in both examples the act of rotation serves as a source of circulation, creating a faster flow of air over the top surface resulting in an elevated pressure force. The direction of this rotational force is such that the wing generates positive lift if it flips over before stroke reversal and negative lift if the wing flip is delayed until after stroke reversal, which are kinematic conditions roughly analogous to backspin and topspin on a tennis ball. Thus, unlike the translational component of the total aerodynamic force, the rotational component is strongly dependent on the precise timing of rotation between strokes. For this reason, rotational forces may be particularly important for steering and flight control.

After reversing direction, the wing does not move through undisturbed air, but rather collides with the wake generated

during the previous stroke. Because the leading edge vortex moves downward after it is shed, its influence on the flow around the wing is maximal at the start of wing translation, but then rapidly diminishes. Nevertheless, the instantaneous air velocity experienced by the wing at the start of each stroke can be substantially greater than that caused by its own flapping speed. Under certain conditions, this increased flow can result in additional force by a mechanism called wake capture. Because a vortex wake represents the energy lost to the fluid by a moving object, wake capture is an aerodynamic mechanism that enables an insect to recover some of the energy otherwise lost to the air. As with rotational forces, wake capture may play a particularly important role in flight control and maneuverability. By changing the timing and speed of wing rotation, insects can manipulate the magnitude and direction of forces during stroke reversal, thereby manipulating force moments around the body's center of mass.

The wake generated by the wings influences aerodynamic forces in other ways. Vortices shed from the wings drive a column of air downward from the plane of wing motion, which is a change in fluid momentum that is equivalent to the average upward force on the wings. This downwash alters the flow around the wings, but reduces the effective aerodynamic angle of attack and thus attenuates the production of translational forces. In addition, flow interactions may occur among the wings on the same insect. For example, in some insects the close apposition and subsequent rotation of the wings at the beginning of the downstroke, termed the "clap" and "fling," augment force production at the start of the stroke by enhancing the development of the leading edge vortex. In four-winged insects such as dragonflies, the wake of the forewing might under certain conditions increase the forces created by the hind wing.

Although certain general aerodynamic principles apply to all insects, the precise details of flight aerodynamics likely vary in concert with the extreme morphological and behavioral diversity found among the species. The force-generating mechanisms described above, as well as additional mechanisms yet to be discovered, are best viewed as a palette from which the flight behavior of any given species is constructed. The long-term goal for the study of insect flight aerodynamics is not only to uncover the mechanism by which any particular species stays in the air but also to show how it manipulates various mechanisms to maneuver and accomplish the aerial behaviors that are necessary for its survival and reproduction. Recent work in elucidating specific aerodynamic mechanisms must be viewed as only a starting point toward a more comprehensive understanding of flight mechanics and behavior.

NEURAL CONTROL

Sensory Systems

The extreme morphological adaptations associated with flight behavior in insects are paralleled by equally impressive specializations within the nervous system. Perhaps most extreme among these alterations relative to the basic neural organization of wingless insects is the hypertrophy of the compound eyes and associated visual ganglia. Large eyes capable of rapid response and broad adaptation to ambient light level are characteristic of diurnal insects such as butterflies, dragonflies, bees, wasps, and true flies. The visual system provides essential sensory feedback for flight control in most diurnal species and is used for a variety of tasks, including velocity and altitude control, obstacle avoidance, landing responses, target recognition, and spatial memory. Features of the anatomy and physiology of the visual system of individual species correlate well with flight behavior and habitat. The elevated translational and rotational speeds characteristic of flight, particularly compared with those of walking and running, place a premium on rapid response time of the visual system. The enhanced visual processing speeds of insects is exemplified by the flicker fusion rate of house flies, which at roughly 300 Hz is the highest found among all animals.

In addition to the eyes, several other sensory modalities on the head provide critical feedback during flight. Although incapable of extracting detailed spatial information, output from the three ocelli helps to stabilize pitch and roll. Because the associated neural computations are relatively simple, the ocellar system can detect and process changes in body orientation more rapidly than can the visual system. Hair cells on the head and mechanoreceptors at the base of the antennae are capable of measuring the magnitude and direction of airflow during flight. In conjunction with visual measurements of ground speed, the input from these wind-sensitive cells is crucial for calculating ambient wind direction, an important capability for flying upwind or tracking odor sources, which are detected in part by chemosensory sensilla on the antennae.

Although sensory structures on the head provide relatively slow tonic cues used for modulating wing motion or body posture over many wingbeats, sensory input from mechanosensory cells on the thorax provides fast phasic input that can alter wing movements on a cycle-by-cycle basis. These mechanosensory structures include the tegula, an organ below the wing that is stimulated during the downstroke, and stretch receptors embedded in the wing hinge that fire during the upstroke. Wing veins contain arrays of tiny campaniform sensilla that encode deformations of the wing surface throughout the stroke. In flies, these arrays are greatly elaborated at the base of tiny drumstick-shaped hind wings called halteres, which function as equilibrium organs. Associated sensory fields detect the Coriolis forces that deflect the beating haltere when the animal's body rotates during flight. Remarkably, a similar specialization is found among stresipterans, but in these insects it is the forewing that has been transformed into an equilibrium organ, whereas the hind wings retain the aerodynamic function. Although the precise role of the thoracic mechanosensory organs varies from species to species, their general function is to tune the output

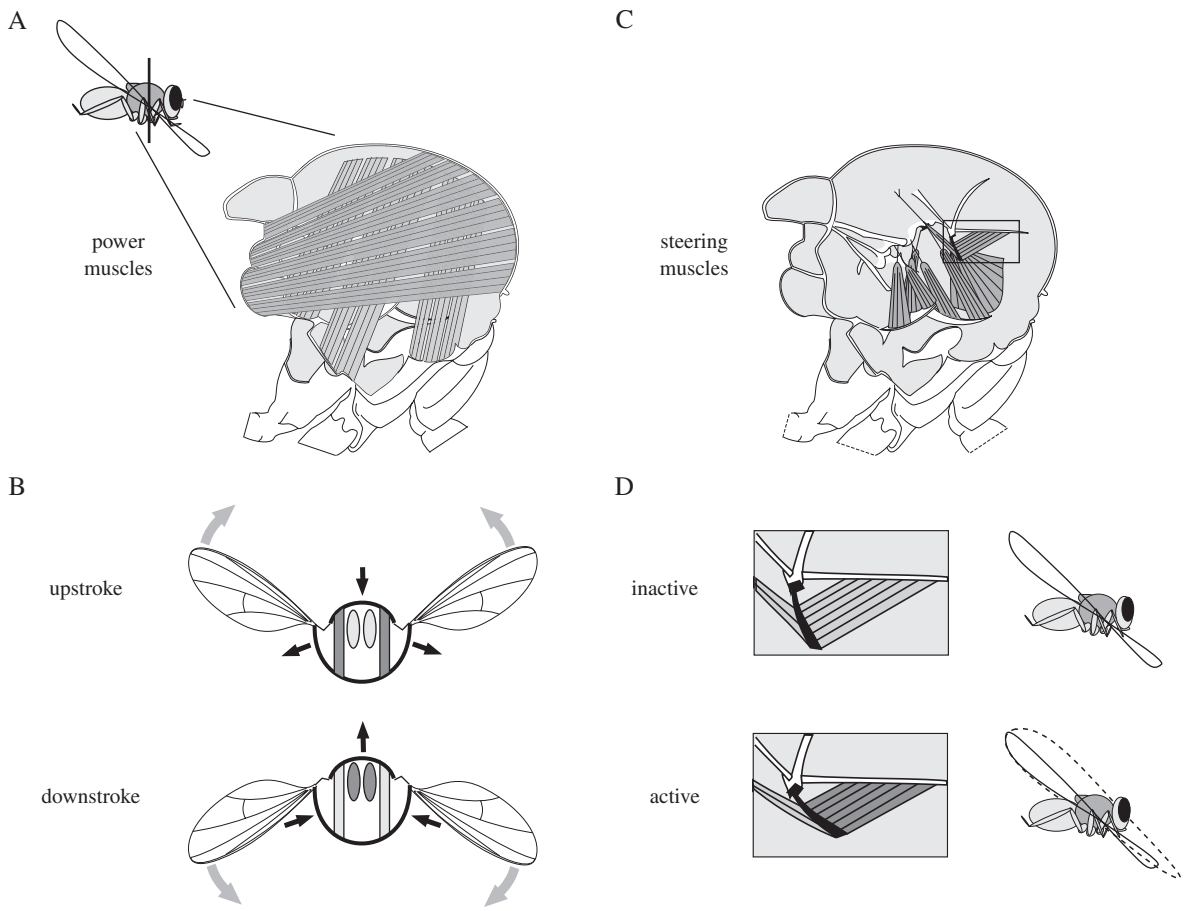

FIGURE 2 In insects using an asynchronous flight motor, the wing muscles are segregated into two anatomically, physiologically, and functionally distinct groups. (A) The large indirect power muscles, which fill the thorax, are arranged in two antagonistic groups. (B) A cross section through the thorax (as indicated by line in A) showing the action of the power muscles. The laterally placed dorsoventral muscles drive the upstroke, whereas the more medial dorsolongitudinal muscles drive the downstroke. The contraction of each muscle set stretches the antagonist group thereby activating the next phase of oscillation. The motion of the thorax is indicated by black arrows, motion of wings is shown by gray arrows. (C) The arrangement of direct steering muscles. (D) Illustration of how activity of a steering muscle changes wing motion (enlargement of rectangular region in C); as muscle becomes active (dark gray), wing trajectory changes.

of interneurons and motor neurons that pattern the activity of the flight muscles. The phasic, phase-locked nature of this feedback is important because the mechanical properties of steering muscles are extremely sensitive to the precise time at which they are activated within the wingbeat cycle.

Muscles

As with sensory systems, motor systems of insects exhibit many specializations related to flight behavior. Unlike the wings of birds, bats, and pterosaurs, insect wings contain no intrinsic muscles. The wing is attached to the thorax by a complicated hinge structure that amplifies the tiny strains of the flight musculature into the large sweeping motions of the wing. The hinge is composed of a connected set of hard sclerotized elements (the wing sclerites or pteralia) embedded within a matrix of more compliant cuticle. Flight muscles may be segregated into two morphological groups according to how they transmit force to the wing. Direct flight muscles insert upon apodemes connected directly to the wing sclerites. In contrast, indirect flight muscles insert within the thorax some distance from the base of the wing. Odonates are distinct in possessing only direct flight muscles, whereas most insects possess some combination of direct and indirect muscles. In many of the most species-rich orders, including the Coleoptera, Hymenoptera, and Diptera, direct and indirect muscles differ physiologically and serve distinct functions (Fig. 2).

Large indirect "power" muscles provide the mechanical energy to drive the gross up-and-down motion of the wings, whereas a set of small direct "steering" muscles controls the fine changes in wing kinematics during flight. Each contraction in a steering muscle is activated one for one, by action potentials in presynaptic motor neurons, but contractions in the power muscles are asynchronous with motor input. By a molecular mechanism not yet fully understood, rapid stretch activates the crossbridges in asynchronous muscles, causing them to shorten after a brief delay. The low-frequency drive of motor neurons is sufficient to elevate calcium concentration within the sarcoplasm of asynchronous muscle to a level that maintains crossbridges in a stretch-activated state. During

flight, contractions within sets of antagonist downstroke and upstroke muscles provide the requisite mechanical stretch to activate each other. Stretch activation frees muscles from the requirement of an extensive sarcoplasmic reticulum (SR), which is necessary in synchronous muscle for the release and subsequent uptake of calcium during twitches.

Asynchronous muscles are capable of generating elevated levels of mechanical power because their internal volume is filled almost exclusively with contractile fibrils and mitochondria. The advantage of stretch activation is especially strong at high frequencies for which typical twitch muscles would require an enormous surface area of SR, severely compromising their ability to generate power. Thus, asynchronous fliers can attain much higher wingbeat frequencies, and thus smaller body size, than can insects using synchronous flight muscles. The mechanical efficiency of asynchronous muscles should also be high because the normal costs associated with cyclic release and uptake of calcium through the SR are not incurred. Because their contraction is only partially controlled by the nervous system, indirect asynchronous muscles are ill-suited to mediating rapid changes in wing motion. The nervous system exerts its control of flight behavior primarily through the action of the direct synchronous steering muscles.

Pattern-Generating Circuits

The motor neurons that innervate insect flight muscles are driven by complex rhythm-generating circuits within the nervous system. Seminal studies by Don Wilson on locust flight led to the discovery of central pattern generators (CPGs), circuits consisting of interneurons and motor neurons capable of generating rhythmic patterns in the complete absence of phasic sensory feedback. Cells within CPGs excite and inhibit the motor neurons of upstroke and downstroke muscles so that they fire antiphasically during the stroke cycle. Even stretch-activated muscles are driven by CPGs, although the firing rate is roughly 10 times lower than wingbeat frequency. Although there is no doubt that insect nervous systems contain CPGs, research pioneered by Kier Pearson and colleagues has demonstrated that sensory feedback from thoracic mechanosensory structures plays an essential role in patterning motor output during flight. For example, electrical stimulation of wing stretch receptor cells can reset the timing of the flight rhythm in locusts—thus fulfilling a strict criterion that is used to test whether a neuron is a member of a CPG. The circuitry underlying flight behavior is best described as a distributed pattern-generating network, consisting of both central and peripheral neurons.

ENERGETICS

Fuel and Oxygen Delivery

Metabolic rates during flight exceed resting values by a factor of 50 to 200, and the thoracic muscles of flying insects exhibit the highest mass-specific rates of oxygen consumption known for any locomotor muscle. Mitochondrial densities within flight muscle fibers are correspondingly high, ranging in some insects to values as high as 45% of the total muscle volume. Energy during flight is derived almost entirely from the oxidation of chemical fuels; anaerobic pathways are absent from flight muscles. Metabolic fuels diffuse from the hemolymph surrounding muscle fibers to the point of oxidation within mitochondria, whereas bulk movement of hemolymph within the body cavity transports fuels from the abdominal fat body to the thoracic musculature. The type and composition of the fuel used in flight (i.e., lipids, carbohydrates, or amino acids) vary with phylogenetic association and may even change with time during a single flight duration in some species. Oxygen influx and carbon dioxide efflux during flight occur primarily via diffusion within tracheal pathways, but may be augmented by convective motion. For example, contraction of flight muscles and the associated deformations of the thorax can compress and expand internal air sacs and even first- and second-order tracheal branches. Although most higher order branches within the tracheolar network are unlikely to experience convective pumping, muscular contraction may augment diffusion by deforming tracheoles that invaginate muscle fibers.

One important issue relating to flight energetics concerns the limits of insect body size. In dragonflies, studies of tracheal geometry suggest an upper limit to thoracic radius of about 0.5 cm if diffusion alone supplies oxygen during flight. The thoraces of many extant insects are well above this limit, however, and the relative contribution of convection to oxygen supply has yet to be determined for any insect. The existence of some flight-related constraint on maximum body size is supported by the observation that many large insects (e.g., the giant stick insects of Southeast Asia) are secondarily flightless. In a modern species of dragonfly, flight metabolic rates vary in direct proportion to ambient oxygen concentration, a result that is consistent with diffusion-limited oxygen transport. The existence of widespread gigantism in late Paleozoic insects (and among other arthropods) during periods of elevated atmospheric oxygen concentration provides further evidence for diffusive limits to flight metabolism, and thus body size, of flying insects.

Energy Requirements for Flight

Although selection has presumably acted to minimize mechanical power expenditure, most of the energy consumed during flight is lost as heat in the flight musculature. Estimates for the mechanical efficiency of insect flight muscle range from only 4 to 30%, depending on taxon and assumptions as to the amount of elastic energy storage within the thorax. Thus, a comparatively small fraction of the fuel an insect consumes is available as mechanical power to drive the wings. This mechanical energy must support three requirements: parasite, inertial, and aerodynamic power. Inertial power is the power

required to accelerate the wings back and forth during the stroke. Unless inertial power is substantially greater than aerodynamic power, even moderate elastic storage within the thorax renders inertial costs small. Parasite power is the work required to overcome the drag on the animal's body as it moves through the air. Thus, parasite power is negligible at low advance ratios, but increases with the cube of flight speed. The aerodynamic power is the rate of the work the wings perform on the air, which may be further subdivided into induced power, the cost of generating lift, and profile power, the cost of overcoming drag on the wings. Because the lift-to-drag ratio for most wing kinematic patterns capable of generating sufficient lift is quite low, profile power requirements may substantially exceed the induced power, especially in smaller insects. Also, recent measurements of drag on dynamically scaled model insect wings indicate that values of profile power may be two to three times higher than previously thought. Underestimates of aerodynamic power resulting from unrealistically low values for wing drag may explain the low estimates of mechanical efficiency for asynchronous flight muscle.

The variation in power requirements with forward airspeed is of ecological and evolutionary interest because of its implications for optimal foraging and dispersal strategies. Both direct measurements and aerodynamic modeling of bumble bees in forward flight suggest that mechanical power requirements are approximately constant over an airspeed range of 0 to 4.5 m/s. In contrast, calculations for various lepidopteran and odonatan species show substantial increases in mechanical power expenditure, with forward airspeed due to the rise in parasite power. In situations in which parasite power is large relative to aerodynamic power, the choice of airspeed during flight has significant energetic implications. One study with dragonflies suggests, in fact, that maximum flight speeds are determined predominantly by the dramatic increase in body drag and associated power requirements at extreme airspeeds.

Temperature Effects

As with many features of flight muscle physiology, power production is strongly temperature-dependent, an effect that has several implications for overall flight performance. Measurements on isolated muscles show that mechanical power output typically increases with temperature and is maximal near muscle temperatures characteristic of the free-flying insect. However, the temperature dependence of power output differs greatly among taxa, and although some insects can instantly take off from the surface of glaciers, others must warm their thoraces to 40°C before their muscle generates sufficient power to sustain flight. In insects for which the flight muscles require elevated temperatures to attain adequate performance, the heat generated during flight that results from low muscle contractile efficiency is available as a source with which to regulate thoracic temperature. In small insects, most metabolic heat generated during flight is lost via convective cooling, and body temperature is close if not equal to ambient air temperature. In larger insects, however, metabolic heat gain is high relative to convective loss and body temperatures are correspondingly elevated. Many large insects regulate internal heat distribution via control of hemolymph circulation between the thorax and the abdomen, using the latter to radiate excess heat. The dramatic amounts of heat produced by muscular contraction are illustrated by the capacity of bumble bees and of some moths to maintain thoracic temperatures exceeding 30°C when ambient air temperature is only 2 to 3°C. Evolution of such thermoregulatory capacity in many insects is consistent with strong historical selection on muscle performance to meet the exacting energetic demands of flight. Further evidence supporting the link between thermoregulation and flight is the phenomenon known as preflight warm-up. In larger insect taxa, pronounced contraction of the thoracic muscles and low-amplitude wing vibrations precede flight. These actions elevate thoracic temperature to values at which the muscles yield sufficient mechanical power for takeoff. Ontogenetic variation in the temperature dependence of muscle power output can also be substantial. In some dragonflies, for example, thermal sensitivity of force production by flight muscle is correlated with changes in the expression of myosin isoforms through development. This finding suggests that physiological features of flight performance are matched to particular environmental conditions and selective demands.

ECOLOGY AND DIVERSITY

Wing Arrangement

The origin of wings was followed by an explosive diversification of insect orders. Many Carboniferous insects possessed wings of approximately equivalent size, shape, and aerodynamic function that were probably limited to low-amplitude flapping. Equivalently sized fore- and hind wings persist to this day in at least seven orders. However, major differences in the sizes of meso- and metathoracic wings are evident in both contemporary fauna and fossils from the Paleozoic. With the exception of the Coleoptera and Strepsiptera, enlarged hind wings are for the most part confined to extant exopterygote orders. Many endopterygote orders (Hymenoptera, Lepidoptera, Diptera), by contrast, reduce the aerodynamic role of the hind wings. In many insects in which the hind wings provide aerodynamic force, the forewings have been modified for supplemental function. Far from isolated events, the evolutionary transformation of the forewing into either a tegmen or an elytron has occurred at least three times at the ordinal level. Elytra of the Coleoptera have much reduced aerodynamic roles relative to the hind wings and provide for greater mechanical resistance to crushing in conjunction with increased sclerotization of the body as a whole. A similar functional role may be hypothesized for tegminized forewings (e.g., Blattodea and Orthoptera) and for the hemelytra of Hemiptera. Insect wings may also serve a variety of behavioral

functions unrelated to flight, including sound production and visual communication. None of these functions are mutually exclusive, although the role of aerodynamic force production remains paramount for at least one wing pair.

Flightlessness

The behavioral and ecological advantages of flight notwithstanding, flightlessness has evolved independently many times in insects. Approximately 5% of the extant insect fauna may be classified as flightless, if all forms of variable wing expression and of reduced flight musculature are included. One common feature of the otherwise diverse manifestations of flightlessness is a reduced need for locomotor mobility. Selection for maintaining flight may be weak if this capability is not required for dispersal, reproductive behavior, or predator avoidance. Even in flying species, the costly development of wings and associated musculature may not occur under all ecological conditions.

Flight Diversity and Body Size

Changes in body size represent major trends in the evolution of winged insects. Although direct paleontological evidence is not available, body lengths of the first flying insects were probably in the range of 2 to 4 cm. Substantial increases in body length appear to have occurred by the mid-Carboniferous, and gigantism relative to today's forms was typical of many late Paleozoic insects as well as of other arthropods. The most parsimonious explanation for Paleozoic gigantism is a contemporaneous increase in atmospheric oxygen concentrations, possibly to values as high as 35% relative to today's 21%. Such high oxygen concentrations, together with higher diffusion constants due to an increase in total atmospheric pressure, would have relaxed diffusional constraints on flight metabolism and thus would have permitted the evolution of giant flying forms. Increased atmospheric density would also have yielded increased augmented lift production during flight, both effects possibly being advantageous during the initial periods of wing evolution. Furthermore, geophysical evidence suggests a decline in atmospheric oxygen concentration through the mid- to end-Permian. As would be consistent with asphyxiation on a geological time scale, all giant terrestrial arthropod taxa of the late Paleozoic went extinct by the end of the era.

In sharp contrast to the late Paleozoic giants, the contemporary insect fauna is characterized by a diversity of miniaturized forms. For example, mean adult beetle body length lies between 4 and 5 mm. Much of the wealth of dipteran and hymenopteran diversity is similarly associated with small body sizes, particularly among the parasitoid and hyperparasitoid taxa. Wingbeat frequencies vary inversely proportional to body size, and today's small insects typically fly with wingbeat frequencies in excess of 100 Hz, rates achievable only with the use of asynchronous muscle. Thus,

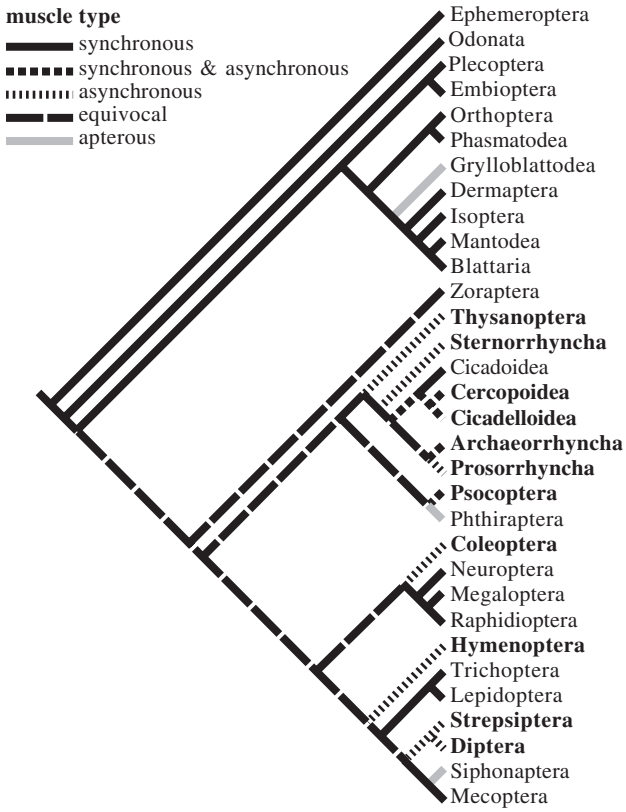

FIGURE 3 Phylogenetic distribution of asynchronous flight muscle. The paraphyletic assemblage Homoptera is here represented at lower taxonomic levels of suborders and superfamilies. Equivocal branch designations indicate either an unknown (e.g., Zoraptera) or an unresolved character state.

the acquisition of asynchronous flight muscle has played a major role in morphological diversification among different insect orders. Asynchronous flight muscle is phylogenetically derived relative to synchronous precursors and has evolved repeatedly among pterygote lineages (Fig. 3). Because flight at small body sizes mandates elevated wingbeat frequencies, this repeated evolutionary acquisition of asynchronous muscle may have facilitated taxonomic radiations of small insects. For example, systematic comparison of sister insect lineages that differ in muscle type statistically demonstrates a decrease in mean body size and an increase in species number if asynchronous flight muscle is present. Three-quarters of all described insect species and three of the four largest orders (i.e., Coleoptera, Diptera, and Hymenoptera) are characterized by asynchronous muscle. Because higher wingbeat frequencies yield increased aerodynamic force, asynchronous muscle may also permit a reduced wing area relative to body mass. This effect may help to explain how one wing pair in many insect groups evolved nonaerodynamic roles.

Flight Behavior and Ecology

Flight plays a central role in the life history patterns of most pterygote insects. A partial list of important insect behaviors

mediated by flight includes pollination, phytophagy, hematophagy, escape from predators, mate acquisition, and migration. Forces of both natural and sexual selection have demanded ever-increasing flight performance from insects through evolutionary time, whereas different selective agents are often mutually reinforcing. For example, intra- and intersexual selection often acts synergistically on maneuverability, as does escape from predation attempts by bats, birds, and other insects. Coevolutionary defensive responses among insects, including increased maneuverability and erratic flight styles, parallel the diverse radiations of insectivorous vertebrates worldwide. The morphological and behavioral mimicry among certain chemically defended insects provides wonderful testimony to the strength of such natural selection.

Another major coevolutionary theme in the terrestrial biosphere concerns relationships between flying insects and plants. Phytophagy and pollination by insects are particularly influenced by three-dimensional aerial mobility, the capacity for which dramatically increases access to nutritional resources and suitable oviposition sites. The antiquity of such interactions is well demonstrated by fossil evidence for feeding on plants in the Upper Carboniferous, whereas high rates of herbivory imposed by insects characterize most present-day floras. The evolutionary presence of flying insects has similarly influenced the reproductive biology of many plants. Contemporary angiosperms are pollinated primarily by a broad diversity of insect taxa, most of which are miniaturized forms that can hover at flowers either before or during pollination. Small body size facilitates both incidental and intentional dispersal by wind, and as a consequence tiny insects can act as long-distance pollen vectors.

Continuous aerial entrainment by winds interacts with the large individual numbers of insects worldwide to result in a transient but substantial population of insects moving at heights up to 10 km from the earth's surface. Remarkably, insects from continental faunas have been captured in the mid-Pacific far from any land mass or island. The ability to decouple the flight trajectory from ambient winds depends on the relative magnitude of insect airspeeds, which but rarely exceed typical wind speeds. Thus, directed movement is likely only a few meters from the ground or within canopies of vegetation. Dispersal, on the other hand, is readily attained simply by flying upward into moving air masses. Even migratory flights of larger, more powerful insects (such as locusts and butterflies) are influenced by the directionality of prevailing winds.

See Also the Following Articles

Anatomy • Migration • Muscle System • Odonata • Swimming • Walking and Jumping • Wings

Further Reading

Brodsky, A. K. (1994). "The Evolution of Insect Flight." Oxford University Press, Oxford.

Dalton, S. (1975). "Borne on the Wind: The Extraordinary World of Insects in Flight." Reader's Digest Press, New York.

Dickinson, M. H. (2001). Solving the mystery of insect flight. *Sci. Am.* **284,** 34–41.

Dickinson, M., Lehmann, F.-O., and Sane, S. (2001). Wing rotation and the aerodynamic basis of insect flight. *Science* **284,** 1881–2044.

Dudley, R. (2000). "The Biomechanics of Insect Flight: Form, Function, Evolution." Princeton University Press, Princeton, NJ.

Ellington, C. P. (1999). The novel aerodynamics of insect flight: Applications to micro-air vehicles. *J. Exp. Biol.* **202,** 3439–3448.

Ellington, C. P., Van den Berg, C., Willmot, A. P., and Thomas, A. L. R. (1996). Leading edge vortices in insect flight. *Nature* **384,** 626–630.

Harrison, J. F., and Roberts, S. P. (2000). Flight respiration and energetics. *Annu. Rev. Physiol.* **62,** 179–205.

Heinrich, B. (1993). "The Hot-Blooded Insects: Strategies and Mechanisms of Thermoregulation." Harvard University Press, Cambridge, MA.

Josephson, R., Malamud, J. G., and Stokes, D. R. (2000). Asynchronous muscle: A primer. *J. Exp. Biol.* **203,** 2713–2722.

Marden, J. H. (2000). Variability in the size, composition, and function of insect flight muscles. *Annu. Rev. Physiol.* **62,** 157–178.

Nachtigall, W. (1974). "Insects in Flight: A Glimpse behind the Scenes in Biophysical Research." McGraw–Hill, New York.

Wootton, R. J. (1992). Functional morphology of insect wings. *Annu. Rev. Entomol.* **37,** 113–140.

Young, D., and Simmons, P. (1999). "Nerve Cells and Animal Behavior," 2nd edition. Cambridge University Press, Cambridge, UK.

Fly

see *Diptera*

Folk Beliefs and Superstitions

James N. Hogue
California State University, Northridge

An integral part of any society's cultural heritage is the collection of stories and traditions passed from generation to generation through the ages. Folklore serves to define a people's identity by mirroring its beliefs, concerns, and fantasies. The passing on of traditional tales thus creates a lasting, tangible bond between the living and their ancestors and provides meaning to a people's existence in the present.

As with other aspects of human culture, the characters, lessons, and motivation of human folklore are greatly influenced by a people's surroundings. It is therefore not surprising that ubiquitous insects are common elements in the variety of traditional stories told by people from both past and present societies. Among the variety of loosely categorized stories, traditions, and beliefs typically passed orally through time in human societies are folk tales and superstitious beliefs.

Although such a medium for passing on information often results in inconsistent transmission across time, it provides an avenue for the creative embellishment of stories that reflects the ideals of the teller and the contemporary state of the particular culture. Other aspects of a people's culture passed on in this way, such as religion or mythology, can be thought of as fundamentally different and thus treated separately.

FOLKLORE AND MYTHOLOGY

Although one certainly grades into the other and it is impossible to generalize across every situation, there are some marked differences between folk tales and stories classified as mythology. Myths are typically more infused with expressions of the unconscious and have more symbolic or religious significance than folk tales. Entire societies are grounded in myths, not folklore. Myths also accompany rituals and ceremonies much more so than folk tales. For example, myths are a common medium for reenactments of the past, such as the creation of the world or other significant events in a culture's history. This is true sometimes for stories treated as folklore, but they are much less engrained with ritual or any extended meaning much beyond the particular focus of the tale. In general, folk tales tend to be more for entertainment, whereas myths tend to be more for spiritual instruction.

Folk tales and superstitions serve a variety of purposes, some of which reveal possible reasons for their origin. Some beliefs and tales deal with societal problems and incompatibilities between culture and nature. Some are used as vehicles for wish fulfillment or as a means to speculate on explanations for phenomena observed in nature. They also may serve as mere tales for the enjoyment and entertainment of both the teller and the listener. Other tales serve to instill moral lessons or provide instructions for living one's life in a particular way. Such tales, because they are inherently interesting and entertaining, are particularly suited to providing historical or moral instruction to young minds that might otherwise be unreceptive to such teachings.

TALES THAT EXPLAIN OBSERVED PHENOMENA

Many folk tales dealing with insects are based on fanciful explanations of natural phenomena. The ancient tale of the bugonia apparently originated from such confusion. Bugonia comes from a Greek word that means ox progeny, and is based on the notion that a swarm of honey bees could be spontaneously generated from the rotting carcass of an ox (Fig. 1). This was not merely a description of something perceived to occur in nature, but was a means whereby people could generate many new individuals of these beneficial insects. For this to be successful, precise instructions had to be followed regarding the proper methods and timing of the slaughter and preparation of the ox carcass. This European tale is also found in Chinese and Japanese folklore, and similar beliefs existed for the generation of other bee-like insects from the carcasses of other animals, namely wasps from horses and hornets from mules.

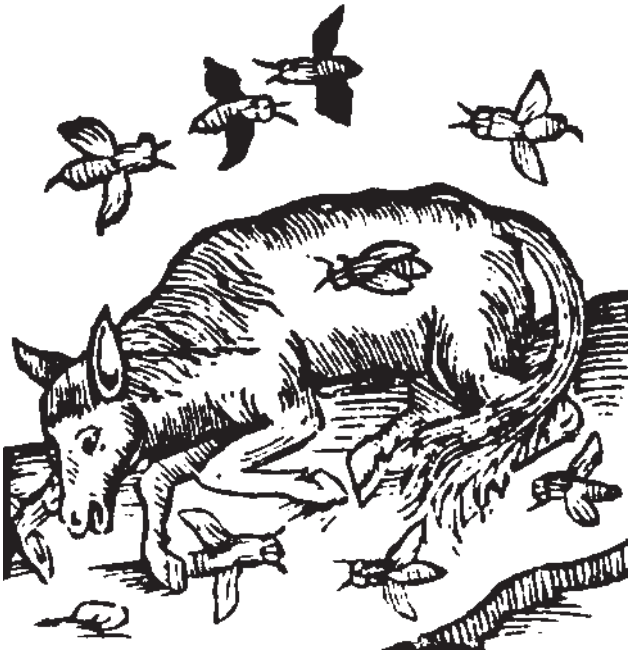

FIGURE 1 A 16th century depiction of spontaneous generation of honey bees from a dead ox. [Illustration modified from Bodenheimer (1928).]

The bugonia tale originated in ancient Egypt, in a place and time at which the ox and the bee were revered as gods. A Biblical reference to this phenomenon (Joshua 14:8) attests to the antiquity of this belief. This tale of ancient times persisted well into the 1600s, when more careful observations of insect biology led to other explanations. In 1883, the eminent dipterist C. R. Osten Sacken proposed an explanation for the origin of the bugonia story that led to it being discredited. The supposed bees and wasps occurring in the carcasses of dead animals were in fact the drone fly, *Eristalis tenax* (Syrphidae). This fly, a Batesian mimic of honey bees, breeds in putrefying organic matter and could easily be mistaken for a bee by the untrained eye.

The presence of particular forest clearings in western Amazonia is ascribed to the activities of forest gnomes known as *chulla chaqui.* These mischievous creatures live near the clearings and eat the fruits of the only shrub that grows in such places, *Duroia* spp. (Rubiaceae). Like many other figures in zoological folklore, these gnomes can take on the appearance of other forest creatures. They take particular delight in transforming into a brilliant blue morpho butterfly, whereby they attract the attention of human visitors and lure them into the forest, only to disappear and leave the disoriented humans lost. In reality, these clearings are formed and maintained by ants that live in a symbiotic relationship with *Duroia.* The ants clear potential competitor seedlings from areas around their myrmecophytic host plants in exchange for a place to live. This folk explanation is similar to that for small clearings in temperate forests or fields that serve as

FIGURE 2 A typical insect-winged fairy. [Illustration by Ellen Edmonson from "Honey Bees and Fairy Dust," by Mary Geisler Phillips (1926).]

places where woodland fairies commonly gather. Fairies, those furtive, entomologically inspired imaginary beings of diminutive human form that typically bear insect-like wings (Fig. 2), are common figures in European folklore. In fields, the clearings known as fairy rings, supposedly caused by dancing fairies, are in reality caused by underground fungi that make their appearance in the form of an ever-expanding ring of mushrooms that encircles a bare patch among the surrounding vegetation.

Another example of transformation surrounds the explanation for the name of a famous insect used by people as an object of adornment. In Mexico, the jewel-adorned ironclad beetles that serve as living brooches are known as *Ma'Kech* after a legend about a Mayan Prince of Yucatan who is said to have escaped his lover's guards by transforming, with the help of the Moon Goddess, into this beetle. His lover was so impressed by his resolve that she uttered *"Ma'Kech."* This phrase not only means "you are a man" but also means "does not eat," and refers to this insect's and the Prince's ability to go without food for long periods of time.

The phenomenon of crypsis (imitating the background in form, color, pattern, or behavior by an organism to avoid detection) is explained by some indigenous peoples using yet another example of transformation. It is said that leaves can transform into insects such as katydids and mantids. This is a reasonable explanation given the striking leaf-like appearance of these insects. The transformation of plants into insects was also implicated by some early European naturalists in their explanation of the issuance of insects from galls, nuts, and fruit. This fanciful theory supposed that the various insects brought forth from these sources were generated by the "vegetative and sensitive soul" of the plant. The origin of another organism associated with insects is similarly explained. The elongate fruiting bodies of certain fungi (*Cordyceps* spp.) that commonly attack insects in the American tropics are thought to be the first stage in the development of particular jungle vines that are used for binding poles.

Explanations for insect behavior often take the form of folk tales. Several stories and rhymes tell the tale of the origin of the katydid's song. These short tales typically center on a girl or young woman named Katy who is accused of committing some bad deed such as deceiving or killing another person. The shameful act is immortalized by the singing insects in the trees that continue to debate whether "Katy did" or "Katy didn't." Some insects spend the greater part of their lives boring through and feeding upon living or dead wood. According to a story from the Tahltan of British Columbia, these insects were tricked into searching for their food in this manner by another insect. Long ago a beetle larva and a mosquito lived together. Every day, the larva watched his friend the mosquito come home engorged with blood. Upon being asked by the beetle larva where he was able to regularly find food, the mosquito, not wanting to give up his secret, replied that he sucked his meal out of trees. The next day the larva began boring into wood looking for food, an activity that continues to this day.

In addition to biological phenomena, stories about the origins of some geophysical entities similarly incorporate insects. The origin of fire has been attributed to the actions of fireflies that were responsible for starting the mythological first campfire. According to the Yagua Indians of the upper Amazon, the origin of the river is a result of the misguided actions of insects. Before the existence of the river, the water used by people came from the "tree of water" that, when cut, would release some of this precious liquid. In an effort to liberate more water, wood-boring insects were deceptively used by some children in a plan that damaged the tree such that it released all its waters at once. This resulted in the formation in the mighty Amazon River.

BENEFICIARY TALES AND ENTOMOPHOBIC LEGENDS

Another force behind some folklore is a means of obtaining some diffuse or ancillary benefit for the originator or propagator of the tale. The tale of the Machaca among some inhabitants of Amazonia is a good example. The purportedly deadly consequences of the bite from the Machaca, which in

actuality is the harmless but menacing looking fulgorid *Fulgora laternaria,* can be thwarted by having sexual relations within 24 h. These insects instill fear and should be avoided, but should the unfortunate happen, a cure is available. Such "sex antidotes" are fairly widespread among folk cures. The potential benefits to those disseminating such tales are obvious.

Other superstitious beliefs benefit particular insects by protecting them from undue harm from people. The Cornish believed that fairies were the souls of ancient heathen people that were too good for Hell but too bad for Heaven. These beings had gradually shrunk from their natural size to that of ants. It was therefore unlucky to kill ants. Similar tales of bad luck when people willingly or inadvertently step on or otherwise harm particular insects are found throughout the world. This is particularly true for insects perceived as beautiful or beneficial to human endeavors such as butterflies and ladybird beetles.

Some insect folklore stems from a general dislike of insects by people and serves to pass this feeling on to others and propagate fright and ill will toward insects. In some stories, insects may be stigmatized with imagined, dangerous qualities. This is most common for insects that have a frightening appearance and gives reason for them to be despised and avoided. Dragonflies and damselflies, for example, are the bearers of nearly 100 English folk names related to their appearance or supposed behaviors. One of their names is "the devil's darning needle," referring to their ability to sew closed the mouth, nostrils, and eyelids of someone unfortunate enough to be the focus of their displeasure. Other examples focus on fanciful abilities of certain pestiferous species to invade nearly any aspect of human life. One fictitious tale describes the plight of an unlucky woman who kept her hair pinned up for such a long time that it became infested with cockroaches.

A little known legend surrounds the comings and goings of body lice, *Pediculus humanus humanus,* an ectoparasite long associated with humans. There was a belief during the 16th century that during trans-Atlantic voyages, lice on the heads and bodies of mariners would miraculously disappear from the westward traveler at a line of longitude roughly 100 leagues west of the Azores. Furthermore, these parasites would return to the eastbound sailors at the same meridian. The basis of this sailor's tale is unclear, but it may be loosely related to the effects that the increase in ambient temperature and the associated shedding of clothing had on the number of observed lice as ships approached more tropical climes.

INSECTS AS OMENS AND SOOTHSAYERS

Insects that are most commonly featured in human folklore are those that most closely associate with humans or impact human affairs. It is not surprising then that insects such as cockroaches, mosquitoes, and bees are some of the most common subjects in stories and superstitions in which an insect's presence or activity is related to significant events in people's lives.

Because humans have practiced honey hunting and beekeeping for thousands of years, it is not surprising that there is much folklore surrounding these activities. The discovery and collection of honey is reason for merriment and joy in many hunter–gatherer societies and much significance has been attributed to the presence of bees and their role as makers of honey. The activities of foraging honey bees are used to predict the weather. When bees forage far from the hive, good weather is expected, but when they forage nearby, poor weather is sure to come. In ancient Rome, swarms of bees foretold impending misfortune. The significance of the timing of bee swarms is exemplified by the following rhyme:

A swarm of bees in May, is worth a load of hay.
A swarm of bees in June, is worth a silver spoon.
A swarm of bees in July, is not worth a fly.

This saying is relevant to the beekeeper whose summer swarms of bees heading off into the distance mean lost assets.

In addition to bees, the presence and behavior of other insects are used to predict the weather. The most widely known insect-mediated weather forecaster is the larvae of some tiger moths (Arctiidae), known as woollybear caterpillars. These caterpillars, in particular those of the banded woollybear, *Pyrrharctia isabella,* are thickly covered with erect black hairs and have a band of reddish brown hairs encircling the middle of their body. The width of the central band supposedly predicts the weather conditions of the coming winter. Narrow bands indicate a long, cold winter, whereas wide bands indicate a short, relatively warm winter. Other insects associated with weather forecasts are butterflies, flies, wasps, and ants. The Zuni of the American Southwest say to expect rain when the white butterfly flies from the southwest. American folklore tells us that when the gnats swarm, rain and warmer weather are believed to be coming, and when hornets build nests near the ground a harsh winter is expected. Rain is expected when ants withdraw into their nests or if someone steps on an ant. The European stag beetle, *Lucanus cervus,* is supposed to be able to attract thunderbolts. This association is perhaps explained because these beetles were commonly found in old oak trees that were often struck by lightning. Because of this belief, these beetles were sacred to Thor, the Germanic god of thunder.

The association of particular insects with common events in distant parts of the world sometimes depends on the characteristics of a particular taxon. Praying mantids are considered pious prophets or soothsayers in various parts of the world. There is also a considerable body of folklore associated with ladybird beetles. Named after the Virgin Mary (Our Lady), these beetles are widely equated with good luck and are often associated with the ability to portend happy events. These beetles are reputed to have been sacred to Freyja, the ancient Norse goddess of love. To harm one of these insects would certainly bring bad luck. That the most common European species of ladybird has seven spots is the basis for one explanation why this beetle is venerated in this

part of the world. The number 7 has long been considered a mystical, powerful, and "perfect" number. In southeastern France, a young girl can predict the year when she will marry by placing a ladybird beetle on her finger and counting the years aloud until the beetle flies away. In other instances, an insect's significance depends on characteristics or behaviors shared between quite different taxa.

In general, the insects found in the folklore of a particular place are drawn from the local fauna. Consequently, significant events common to people worldwide are associated with different species of insect. In British folklore, the presence of deathwatch beetles (Anobiidae) is correlated with the demise of someone in the household. These beetles that live in wood, such as that framing an old English house, send telegraphic messages to each other by tapping their heads on the tunnel walls. This tapping sound is audible to people when all else is quiet, such as in a silent room during a bedside deathwatch. In parts of the Neotropics, the activities of termites fulfill this role as a harbinger of death in a similar manner. Other insects associated with impending death include the appearance of lice in one's dreams, cockroaches flying in one's room, and the sighting of a death's head hawk moth *(Acherontia atropos).* The scales on the dorsum of the thorax of this moth form the readily recognizable image of a whitish human skull against a dark background. The association of this moth with death in the minds of humans was inevitable.

Often the appearance of a given insect conveys a different meaning in different places or at different times. For example, in some parts of the world, a cricket in the house means good luck, but in other places the presence of this insect means ill fortune. According to one superstition in Brazil, careless contact with fireflies can cause blindness, but in the hands of a curandeiro (folk healer or medicine man), fireflies can be used to cure blindness.

INSTRUCTIONAL TALES

In addition to being entertaining, some folk tales serve as a useful means of instruction. Many tales are told to convey a moral message or pass on useful information in an interesting, amusing, and hence more easily remembered format. An example is that of Aesop's fable of the ant and the grasshopper. While the ant concerned himself all summer with gathering provisions for the upcoming winter, the grasshopper spent his time in leisure and song. The grasshopper even derided the ant for spending so much of his time at work instead of play. When winter came, the grasshopper was not prepared and suffered the consequences of his folly. The ant on the other hand, lived comfortably through the winter on the stores he gathered all summer. The activities of these insects in this story are used to show the importance of preparation for future times of necessity. In addition to ants, the behavior of other social insects such as termites, honey bees, and wasps is commonly used to exemplify the benefits of cooperation, diligence, and hard work.

FOLK MEDICINE

Folk remedies for the treatment of the innumerable ailments that befall humans and their animals are found worldwide. Although less important than herbal remedies, insects play a role in the folklore of healing and drug use. One of the most well-known insect-derived folk medicines is cantharidin. This powerful vesicant is derived from dried blister beetles, particularly *Lytta vesicatoria.* Although cantharidin can be extremely toxic to humans, as recently as the early 1900s cantharidin was used to treat a variety of ailments such as asthma, epilepsy, warts, sterility, and bedwetting. In Europe, where the drug as well as the beetle is known as "Spanish fly," powdered cantharidin was taken orally for its purported qualities as an aphrodisiac. Cossinus, a close friend of the Roman Emperor Nero, reportedly died when an Egyptian doctor gave him "cantharis" to drink for treatment of a skin disease.

Many other insects and insect-derived products have been, and sometimes continue to be, used to improve health and treat disease. One product of insects that is widely used today in the context of what might be called folk medicine is bee pollen. The consumption of bee pollen is said to improve general health and increase stamina. Tonics and teas derived from nearly every insect order, from bedbugs to beetles and cicadas, have found their way into the human apothecary. In China, exuviae left behind by newly emerged adults are used to prepare a tonic to treat eye disease and ailments of the lungs and liver and to soothe crying children. Another particularly interesting use of insect-derived pharmaceuticals in China has recently received much publicity. A tonic made from the fruiting body of the entomophagous fungus *Cordyceps sinensis* is considered a general-health and stress-relieving tonic. The fungus is collected in the wild from the dead caterpillar hosts of the hepialid moth *Hepilus fabricius.* In addition to the variety of ailments purportedly treated with this tonic, caterpillar fungus is also used to improve stamina and endurance. The tremendous performances of Chinese female distance runners in the early 1990s were attributed in part to the use of this caterpillar fungus tonic as part of their training regimen.

Although generally based on some empirical observation some time in the distant past, the validity of insect-based folk medicines should not be assumed, even on the grounds of widespread and long-term use. This is particularly true of aphrodisiacs. The symbolic, religious, and ceremonial associations common to the historical use of many drugs tend to obscure evidence on actual potency. On the other hand, the medicinal use of insects in folk remedies should not be dismissed outright as untrue. Each insect species possesses a unique biochemistry that has the potential to perform any number of medicinal tasks. Some insect-based folk remedies, such as the use of bee venom to treat arthritis and rheumatism, may eventually find a place in modern medicine or may at least serve as the basis for the derivation of modern treatments.

In addition to folk remedies that use insects to cure ailments, another body of insect folklore deals with ways to

rid ourselves of pestiferous insects. Pliny the Elder wrote that one sure way to rid one's fields of pests, particularly plagues of cantharid beetles, is to have a menstruating woman walk through the field. This treatment was said to cause the "caterpillars, worms, beetles, and other vermin to fall to the ground." One widespread remedy for an infestation of cockroaches is to seal a few roaches and three coins in an envelope and leave it outside. Whoever picks up the envelope would not only be a little richer, but would also be the new owner of your roaches. A simpler remedy was to slip some roaches to some unsuspecting acquaintance to take home with them with the assurance that your roaches would soon follow. Problem ants can be dealt with in much the same way. By rolling several of the bothersome ants in a leaf and leaving it at a neighbor's house, you could be sure that the ants in your house would soon depart to take up residence with the neighbor. Similarly, some folklore deals with how other animals rid themselves of pestiferous insects. Scottish foxes infested with fleas were said to hold a lock of wool in their mouth and then slowly submerse themselves in water until only the nose and the wool were above water. In trying to escape the water, the fleas end up on the fox's nose and the wool. To finish the job, the fox puts its nose under water and releases the wool along with its passengers of fleas.

CONCLUSION

The acculturation of indigenous peoples worldwide, the disconnection of people from the natural world, and the spread of scientifically based knowledge facilitated by modern means of communication all work in concert to diminish the generation and proliferation of folklore in modern societies. With the ongoing loss of folk traditions in the modern world, so goes the place of insects in traditional folk beliefs and superstitions. Some continue to live on in various forms among contemporary societies, especially among indigenous people in places where the impact of the modern world has yet to take hold. The tidbits of factual observations and the incorrect information that came from antiquity through folklore formed the basis on which the first truly scientific naturalists started their work in the 16th century and led to modern scientific investigation. What was once a way of dealing with problems of unknown nature in our world is now being supplanted by stories of mere curiosity, which although largely untrue, provide interesting insight into the origin and development of human societies, cultures, and religions and are aesthetically pleasing to study.

See Also the Following Articles

Beekeeping • Cultural Entomology • Food, Insects as • Ladybugs • Medicine, Insects in • Teaching Resources

Further Reading

Bodenheimer, F. S. (1928). "Materialien zur Geschichteder Entomologie bis Linné," Vol. I. Junk, Berlin.
Clausen, L. W. (1954). "Insect Fact and Folklore." Macmillan Co., New York.
Hogue, C. L. (1981). Commentaries in cultural entomology. 2. The myth of the louse line. *Entomol. News* **92,** 53–55.
Hogue, C. L. (1985). Amazonian insect myths. *Terra* **23,** 10–15.
Hogue, C. L. (1987). Cultural entomology. *Annu. Rev. Entomol.* **32,** 181–199.
Kirk, G. S. (1970). "Myth: Its Meaning and Functions in Ancient and Other Cultures." Cambridge University Press, London.
Kritzky, G., and Cherry, R. (2000). "Insect Mythology." Writers Club Press, San Jose, CA.

Food, Insects as

Gene R. DeFoliart
University of Wisconsin, Madison

Insects in certain taxonomic groups have played an important role in the history of human nutrition. Although their use as food has long been taboo in almost all Western cultures, their traditional use in tropical and subtropical countries continues to be widespread and to provide significant benefits—nutritional, economic, and ecological—especially for rural communities. The potential benefits of continued or wider use are obvious enough that there seems to be a lessening of the negative attitude in the West.

The type of metamorphosis undergone by an edible species determines which life stage(s) is likely to be consumed. In the insect orders with simple or incomplete metamorphosis (i.e., the Hemimetabola), the life stages usually eaten are the nymphs and/or adults. These orders include the Orthoptera (grasshoppers, locusts, katydids, crickets), Isoptera (termites), Heteroptera (true bugs), and Homoptera (cicadas). Legs, wings, head, and any other hard parts are usually removed before cooking. Orders having complex or complete metamorphosis (i.e., the Hemimetabola) include the Lepidoptera (moths, butterflies), Coleoptera (beetles, weevils), and Hymenoptera (bees, ants, wasps). The life stage usually eaten is the larva, but sometimes it is the pupa or, rarely, the adult.

The insects used as food are, for the most part, clean-living in their choice of food and habitat. Most feed on leaves or other parts of plants. Some of the coleopterous and lepidopterous larvae are wood borers in either dead or living trees and bushes; some, such as cicada nymphs, feed on plant roots. Some hemipterans and coleopterans are aquatic, and some of these and other edible insects are predaceous. Some hymenopterans such as wasps provision their nests with insect prey upon which the young feed. Some edible species have other aesthetic qualities. Some African termites are architects, erecting earthen cathedral-like termitaria that may rise to heights of 3 or 4 m or more. Cicadas and crickets are songsters.

To collect wild insects for use as food, one should be knowledgeable about which local species are edible, particularly in Western cultures in which insects are not among

traditional foods that are widely recognized. Some insects secrete toxins or sequester toxic chemicals from food plants or serve as a source of injectant, ingestant, contactant, or inhalant allergens. Bright colors, especially red, or showy behavior such as slow, deliberate flight may suggest that an insect contains toxins, or is unpalatable, and should be avoided.

There are many environmental and ecological ramifications relevant to the use of insects as food. Because of the large number of insect species and the consequently wide variety of plants used as hosts, in general, insects are potentially capable of converting a much wider range of vegetation and waste substances into animal biomass than are the animals currently considered acceptable as food by Western cultures. Many plants that either are not used efficiently or are not used at all in food production serve as hosts for edible insects. In Mexico, it has been suggested that some plants that are widespread and characteristic of arid regions, but of limited food value, such as mesquite, madrono, and some cacti, could be used for cultivation of their associated insects, the weevil *Metamasius spinolae* and the larva of the skipper butterfly, *Aegiale hesperiarus.* The protein and fat content of these insects is many times higher than that of their plant hosts. In general, insects also are higher in their food conversion efficiency than are other food animals when both are fed diets of high quality (see the house cricket, *Acheta domesticus*).

ORTHOPTERA (GRASSHOPPERS, LOCUSTS, KATYDIDS, CRICKETS)

Family Acrididae (Shorthorned Grasshoppers)

Grasshoppers and locusts are included among the foods of almost every culture having any history of using insects as food. In southern Africa, before there were crops to protect, the arrival of a locust swarm, some of which were dense enough to block out the sun, was hailed with rejoicing as a time of harvest. Villagers collected them in the evenings after the swarms had alighted and were benumbed by the cool of the night. The locusts were roasted or boiled or, when plentiful, dried and crushed in mortars to make a much appreciated flour. Sometimes the flour or porridge was mixed with honey to make a sort of cake. Early reports noted that indigenous populations with access either to these vast locust swarms or to winged termites soon grew "visibly fatter and in better condition than before." Grasshoppers were also an important food of Indian tribes in western North America. Various methods of harvest were used, but, most frequently, the grasshoppers were encircled by a number of people and driven into a pit previously dug or onto a bed of coals. Thus, slightly roasted, they could either be eaten or dried and kept for winter food.

In more modern times, within the past 20 years, grasshopper harvest has at one time or another replaced insecticide spraying in parts of Mexico, Thailand, and the Philippines. *Sphenarium* is the grasshopper genus of greatest commercial food importance in Mexico. The rice grasshopper, *Oxya velox,* was formerly widely eaten in Japan and Korea. Following reduced use of pesticides on rice in both countries, it is again increasing in numbers. Known as *inago* in Japan it is now found in supermarkets as a luxury item; known as *metdugi* in Korea, it is considered a health food.

Family Gryllidae (Crickets)

Several species of crickets are important as food. In Southeast Asia, *Brachytrupes portentosus* lives in tunnels that are about 30 cm deep, usually one cricket per hole, and comes out only at night. They feed on young plants and are an agricultural pest. They are collected by digging, by filling the holes with water, or as they fly around lights at night. After the wings are removed they are eviscerated, then fried, grilled, or put into curry as a substitute for meat. They are sold by villagers in the markets. In the market at Chiang Mai in Thailand, the shopkeeper takes the crickets live from a plastic bag and spits them longitudinally from head to abdomen on a bamboo stick, three or four crickets per stick. They are then fried in oil in front of shoppers.

Another species of *Brachytrupes,* the sand cricket *(B. membranaceus),* occurs widely in eastern Africa. Like its cousin in Asia, its presence is indicated by a small heap of soil pushed out from its burrow. It is usually collected by the women and children, and as many as 100 can be collected in a day. It has been said of the sand cricket, "When well prepared it is considered a delicacy, for it turns an ordinary meal into a dinner." In Zimbabwe and likely elsewhere, *B. membranaceus* is one of the species that has increased in numbers in recent years because it is particularly suited to the new kinds of agroecosystems. It is now a significant pest in sand-soil fields, and it is sold in urban markets.

The cricket most readily available to Western insect gourmets is the cosmopolitan house cricket, *A. domesticus* (Fig. 1), which is widely reared commercially as food for pets and other small animals. Studies in the United States led to estimates that this cricket, when kept at temperatures of

FIGURE 1 Mass-reared edible house crickets, *A. domesticus.*

30°C or higher and fed diets equal in quality to those used in bringing conventional livestock to market condition, shows a food conversion efficiency about twice as high as those of broiler chicks and pigs, four times higher than sheep, and nearly six times higher than steers when losses caused by dressing percentage and carcass trim are taken into account. In addition, female crickets have much higher fecundity than beef animals; each cricket lays 1200 to 1500 eggs over a period of 3 to 4 weeks. In beef production, by contrast, four animals exist in the breeding herd for each market animal produced, thus giving crickets a true food conversion efficiency close to 20 times better than that of beef.

ISOPTERA (TERMITES)

Termites are a highly regarded food throughout sub-Saharan Africa. They are eaten raw, fried, or roasted and are found widely in village markets. The fungus-growing termites of the genus *Macrotermes* (family Termitidae) are the most widely used as food.

The large winged adults (sexual forms) are collected as they emerge from the nests on their mating flights at the beginning of the rainy season. They are strongly attracted to light and this behavior is utilized in harvesting them. The high termitaria of some species of *Macrotermes* are a spectacular feature of the African landscape; they are even considered private property in some areas. In southern Congo (Kinshaza), the termitaria average three to five per hectare and may cover 4.3 to 7.8% of the *miombo* woodland. The flora of the mounds is characteristic and quite different from that of the surrounding *miombo,* thus increasing habitat diversity. This diversity is in danger of being lost, however, because in suburban regions and towns, the characteristic flora and fauna are being destroyed and the mounds converted to other uses.

Winged termites are a rich source of fat; analysis of *M. subhyalinus* in Angola yielded an energy value of 613 kcal/100 g and *M. falciger* in Zimbabwe a value of 761 kcal/100 g on an ash-free basis. Protein and fat content of the latter were shown to be 41.8 and 44.3%, respectively, on a dry weight basis. In addition to the winged adults, soldier termites are also eaten.

The high regard in which winged termites are held in Zambia has been documented by Silow in 1983 and appears typical for other countries in Africa: "The Mbunda, Nikangala, Lucazi, Luvale, Cokwe, and Yauma generally agree that the meat of *Macrotermes* spp. [winged sexuals] is better than meat of animals, birds, [or] fish. Perhaps one or another of the edible caterpillars is comparable with them, but most of my informants are of the opinion that [*Macrotermes*] or honey is the best existing food." Silow notes that there are a few people who simply do not like termites and that some missionaries have condemned termite eating as a heathen custom. But he further states that Bemba, Namwanga, Nyanja, and Nsenga whom he has met unanimously declare that *Macrotermes* winged adults "are more delicious than anything else, or at least among the most delicious dishes."

HETEROPTERA (TRUE BUGS)

Family Belostomatidae (Giant Water Bugs)

A giant water bug, *Lethocerus indicus,* is widely eaten in Southeast Asia and is especially popular in Thailand, where it is known as *ma-lang-da-na.* The bugs are 5 to 8 cm in length and fierce predators on frogs, large insects, and small fish. They are caught using nets or at lights, to which they are attracted. There are many methods of preparation, including roasting, frying, steaming, and grilling. After cooking they may be pounded and used for flavoring sauces and curries. The males secrete a fragrant liquid from two abdominal glands and are made into a much-prized sauce to accompany meat and fish. In the markets, males sell for three or four times the price for females. Artificial water bug flavoring is now produced, but people still prefer to eat the real bugs.

Imported bugs (known as *mangda*) from Thailand and extracts of the bug (known as "mangdana essence") can be found in Southeast Asian food shops in California. They are popular with Thai and Laotian customers who use them to make a bug paste called *nam prik mangda* prepared by mashing a whole bug with salt, sugar, garlic, shallots, fish sauce, lime juice, and hot Thai capsicum peppers. The *nam prik mangda* is commonly used as a vegetable dip and as a topping for cooked rice. The extracts known as mangdana essence can be used as a substitute for a whole bug in the preparation of *nam prik mangda,* but they are considered inferior in taste to that prepared from a whole bug.

Families Corixidae (Water Boatmen) and Notonectidae (Backswimmers)

The famous Mexican "caviar," also known as *ahuahutle,* is composed of the eggs of several species in these families. These insects formerly bred in tremendous numbers in the alkaline lakes of central Mexico and were the basis of aquatic farming for centuries. Lake water pollution has now reduced their numbers. The eggs are harvested by what amounts to setting oviposition trap lines. Bundles of shore grass are tied together and weighted with a stone and then distributed by canoe. They are left in place for about 3 weeks during which the adult bugs swim up and lay their eggs on the submerged grass. The bundles are then collected, brought to shore, and dried in the sun. When dry, they are shaken and the eggs fall off. The "caviar" is a true delicacy that appears on the menus of the finest restaurants in Mexico.

HOMOPTERA (CICADAS AND OTHERS)

When there is an emergence of one of the species of periodical cicadas (family Cicadidae), many Americans, for whatever reasons, seem to regard them as legitimate fun food. During a recent (1990) emergence in Chicago and northern Illinois, for example, the *Chicago Sun-Times* carried several articles, the second of which began: "Millions of tasty,

entrees-if-you-dare will be available for the gathering during the next month in northern Illinois, and some Chicagoans will want to know how cicada fanciers prepare them." Several recipes were provided. Articles described cicada biology and how to prevent damage caused by egg laying on very young plants and urged Chicagoans to forego the use of insecticidal sprays. There were many radio reports, a cicada hotline, and even *Time* magazine published a recipe.

There are six species of periodical cicadas *(Magicicada)* in North America, three with a 13-year cycle and three with a 17-year cycle. The nymph remains in the soil, feeding on the roots of various plants until ready for the final molt. It then digs itself out of the ground, climbs the nearest tree or shrub, and attaches itself firmly. The adult lives for a month or longer. The so-called dog-day cicadas, such as those of the genus *Tibicen,* have shorter life cycles, but even they require at least 4 years. Cicadas are eaten in many countries, but probably most widely in the countries of southeastern Asia.

LEPIDOPTERA (BUTTERFLIES, MOTHS)

Lepidopterans reach their maximum food importance in Africa where, in many countries, more than 20 species are consumed. In the southern part of Congo (Kinshasa), for example, caterpillars of at least 35 species are consumed.

Family Saturniidae (Giant Silkworms)

In 1980, Malaisse and Parent analyzed 23 species (17 of which were Saturniidae) eaten by humans, using samples that were prepared in a manner identical to that preceding their culinary preparation and then dehydrated. Crude protein content averaged 63.5%, kcal per 100 g dry weight averaged 457, and most species proved to be an excellent source of iron, 100 g averaging (in 21 species analyzed for iron) 335% of the daily requirement. In Angola, the saturniid caterpillar *Usta terpsichore* was found to be a rich source of zinc, iron, thiamine, and riboflavin.

Probably the best known of the edible caterpillars is *Gonimbrasia belina,* the so-called "mopanie worm," which is widely eaten in southern Africa. The South African Bureau of Standards has estimated annual sales of mopanie through agricultural cooperative markets at about 40,000 bags, each containing 40 kg of traditionally prepared, dried caterpillars. This total represents only those entering reported channels of commerce and does not include those privately collected and consumed or sold. The caterpillars, up to 10 cm in length, grip the host plant tightly and cannot be shaken off; they must be picked by hand. A good picker in an average infestation can collect 18 kg per hour. In areas where they are abundant and bulk-dried, they are first eviscerated and then roasted for 15 min before being spread out to dry. About 2 days are required for the product to become dry enough for storage.

The mopanie caterpillar is an important food item and is actively traded not only in South Africa but also in Botswana and Zimbabwe and is exported by the ton to Zambia. From extensive studies in South Africa, Dreyer and Wehmeyer concluded in 1982 that "the consumption of mopanie caterpillars can to a substantial degree supplement the predominantly cereal diet with many of the protective nutrients." The amino acid composition of the dried caterpillars is relatively complete, with high proportions of lysine and tryptophan (which are limiting in maize protein) and of methionine (limiting in legume seed proteins). There is increasing concern in South Africa that the mopanie caterpillar might be collected to the point of extinction.

In Malawi, *G. belina* and another saturniid, *Gynanisa maia,* still occur abundantly in Kasungu National Park; the larvae are in season from mid-October to December, a time of year when food stocks of families living near the Park are running low. The caterpillars are nonexistent outside the Park because of the absence of host trees, which have been displaced by extensive agriculture. A study has shown that opening Kasungu National Park to controlled sustainable use, such as caterpillar harvest, by local people can reduce the problems of poaching in parks and other protected areas. Similarly, of ecological benefit, observations in Zambia have shown that there are very few late bushfires in areas where the caterpillars of *Gy. maia* are found. Fires late in the dry season cause considerable damage by killing trees, reducing regrowth, and increasing erosion. The caterpillars are a highly prized food, and in areas where they are abundant they provide the incentive for people to burn early, thereby protecting the caterpillars and enhancing woodland regeneration. There are other examples in Africa where protection of caterpillars as a food resource enhances biodiversity.

Family Bombycidae (Silkworm Moths)

A by-product of the silk industry, pupae of *Bombyx mori* remain after the silk is reeled from the cocoons. These pupae are widely sold, often canned, in markets in Eastern Asia. In China, the pupae, along with waste materials from the reeling factories and from the silkworm rearing, are also used as fish food in pond-fish culture. Canned pupae are exported, especially from Korea, and can be found in Asian food shops in the United States.

Family Cossidae (Carpenterworms, Leopard Moths)

Many insects were important foods for the Aborigines of Australia and among the most prized were the *witchety* or *witjuti* grubs, several species of root-boring cossid larvae belonging to the genus *Xyleutes.* Tindale conducted in 1953 the definitive study on these insects and stated, "Aborigines with access to *witjuti* grubs usually are healthy and properly nourished. … Women and children spend much time digging for them and a healthy baby seems often to have one dangling from its mouth in much the same way that one of our children would be satisfied with a baby comforter." Over

a period of several months spent observing nomadic Pitjandjaras in the Mann and Musgrave Ranges, it was noted that part of nearly every day's diet consisted of these larvae. Tindale states elsewhere that the taste of *witchety* grub, "when lightly cooked in hot ashes, would delight a gourmet."

Recently in Australia there has been an explosion of interest in native, or "bush tucker," foods, including *witchety* grubs and other insects such as *bardi* grubs (Cerambycidae) and honey ants. Bush food is increasingly found in restaurants frequented by tourists, and book stores are well-stocked with books on bush tucker. *Witchety* grubs are on the menu of the posh restaurant Rountrees on Sydney's North Shore; the chef says of them, "They have a nice, nutty flavor when roasted."

Family Megathymidae (Giant Skippers)

The larva of the giant skipper butterfly, *Aegiale hesperiaris,* known as *gusano blanco de maguey,* or the white agave or maguey worm, is in demand by people of all social classes in Mexico. Whereas campesinos with access to maguey plants can collect their own larvae to eat or to sell, restaurants in the larger towns and cities charge as much as U.S. $25 per plate. The *gusanos* are served fried or roasted in butter, chili, or garlic sauce. They are also exported as gourmet food. Two other edible insects are associated with the maguey. The pink worm of the maguey, *Xyleutes redtenbachi* (family Cossidae), also called the red agave worm or *gusano rojo de maguey,* is the larva used in bottles of tequila. They are sold in the markets and are also used to season sauces or they may be roasted or fried with salt and eaten in tacos. Along the maguey's roots are often colonies of ants, which serve as a source of the prized *escamoles,* or so-called "ant eggs," which actually are ant pupae.

Family Pyralidae (Wax Moths, Grass Moths)

Taylor and Carter wrote in 1976 as follows: "Larvae of the greater wax moth *(Galleria mellonella)* are tasty and, fortunately, easily reared, hardy and odorless. If only they were commercially available, we would probably have centered most of our recipes around them. They are our favorite insect. They are thin-skinned, tender, and succulent. They would appear to lend themselves to commercial exploitation as snack items." The authors note that the larvae, when dropped into hot vegetable oil, immediately swell, elongate and burst, looking then not like an insect, but like popcorn, and having the flavor of potato chips, corn puffs, or the like. These larvae, known as wax moths, are now available from various dealers in North America.

COLEOPTERA (BEETLES, WEEVILS)

Family Curculionidae (Snout Beetles, Weevils)

The larvae of palm weevils, several species of *Rhynchophorus,* also called palm worms, are widely eaten and greatly esteemed. A modern cookbook on Cameroon cuisine includes a recipe describing "coconut larvae" as "a favorite dish offered only to good friends." The major species are *Rhynchophorus palmarum* in the Western Hemisphere, *R. phoenicis* in Africa, and *R. ferrugineus* and *R. bilineatus* in southeastern Asia, Indonesia, and the western Pacific. All of these species have long been semicultivated or "farmed" by indigenous peoples and are excellent examples of how harvests of edible insects from natural populations can be increased by intentional creation of additional breeding sites. Cultivation consists basically of cutting down palms and leaving the logs in the forest with the expectation that larvae will be ready to harvest from the decaying pith 1 to 3 months later. The flavor of the sago grub *(R. ferrugineus papuanus)* in Papua New Guinea has been described as "tender and sweet with a slightly nutty flavor." The insect not only sells regularly in local food markets and is bought by foreigners as well as Papua New Guineans, it also is the focus of annual "grub festivals."

Palm weevils are also destructive pests of palms and, in the Western Hemisphere, are vectors of the nematode *Bursaphelenchus cocophilus,* the causal agent of red-ring disease. Although insecticides have been used in attempts to control the weevils, emphasis is on cultural methods. With the palm worms considered such a delicacy, it has been suggested it might be possible to combine increased production with more efficient recycling of dead and diseased palms and as part of reduced-pesticide integrated pest management (IPM) programs and disease control on coconut and other palms.

Family Scarabaeidae (June Beetles, Dung Beetles, Rhinoceros Beetles)

Of the several edible groups within this family, the most interesting is probably the subfamily Dynastinae or giant rhinoceros beetles, particularly the genus *Oryctes.* Three species, including two that breed mainly in dead standing palms, are eaten in Africa, whereas *Oryctes rhinoceros* is a major pest of palm in Asia and the western Pacific. Main hosts of the adult beetles are coconut, oil, and date palms, whereas the larvae live in a variety of dead but not yet decomposed plant material, including dead standing coconut palms, stumps and logs on the ground, and other types of decaying wood, as well as compost, dung heaps, rotting straw, rotting coconut husks, coffee and cacao pulp waste, refuse from sugar cane factories, ricemills, and sawmills, and other wastes from agricultural processing. Control of rhinoceros beetles is based on sanitation and cultural practices similar to those recommended for *Rhynchophorus* weevils, suggesting that *Oryctes* might also be incorporated into palm IPM programs, recycling an endless variety of tropical wastes into animal protein and fat.

Family Cerambycidae (Longhorned Beetles)

In this family, it is the larvae, primarily, that are used as food. They are wood borers in both living and dead trees and in

logs and stumps. They have long life cycles, a year or more, so would not be good candidates for mass-rearing under controlled conditions. A major genus, with edible species, is *Batocera* in Asia.

Family Tenebrionidae (Darkling Beetles)

Tenebrionids have a bad reputation as pests of meal, flour, and other stored and packaged cereal foods, but, despite this, the yellow mealworm, *Tenebrio molitor,* has been reared by zoos, aquaria, and commercial dealers as food for birds, fish, and a variety of small animals since at least the 18th century. Their easy availability makes them one of the insects most commonly recommended for inclusion in recipes in the West. There is a problem of quinone contamination in some tenebrionid-infested food products, but this appears to be much less a problem in *T. molitor* than in species of the genus *Tribolium.*

HYMENOPTERA (ANTS, BEES, WASPS)

Family Apidae (Honey Bees)

Honey is prized by many indigenous cultures, and bee pupae/mature larvae, sometimes called "grubs" or "brood," are often as highly prized as the honey. In southeast Asia, three species of wild bees, *Apis dorsata, A. florea,* and *A. indica,* are important sources of honey, wax, and brood. *A. dorsata* is the largest species and its nests, in the higher branches of large trees, may be up to 2 m in diameter. Its honey is also the most expensive, but honey from *A. florea* is most commonly found in the markets. People often eat the grubs uncooked, but they are also fried or put into soup. In Latin America, the grubs of *A. mellifera* and of species in several genera of stingless bees (subfamily Meliponinae) are used as food, and some of the bees, in Brazil and Mexico, for example, are semidomesticated. Bees, including stingless species, are also important in Africa. In some places, such as the Congo (Kinshasa), honey and brood are still harvested by cutting down the tree although the practice has been much criticized. Apiculture in the United States is based on the introduced honey bee, *A. mellifera,* and it has been suggested that, because of its good public image, this species might be a valuable tool in helping to reshape attitudes toward insects as food in the United States.

Family Formicidae (Ants)

Many kinds of ants serve as food in different parts of the world and they are generally considered delicacies. In Colombia, for example, toasted leafcutter ants (genus *Atta*) are said to constitute the highest attainment of Colombian cookery. A campesino, by collecting and selling *Atta* ants, can earn during the 3-month season the equivalent of a year's wages for the average rural worker. The genus is restricted to the Western Hemisphere. Only the alates are eaten, the large females being especially prized. They are collected as they swarm from the nest by the thousands on their mating flights during the early part of the rainy season. Two species, *A. cephalotes* and *A. sexdens,* are the most widely consumed, being relished across northern South America, with the former extending up into Mexico.

Fungus gardens grown on chewed leaf fragments are tended in the underground chambers of the large nests of the leafcutter ants. The fungus converts cellulose into carbohydrates that can be metabolized by the ants, thus allowing them to tap the virtually inexhaustible supply of cellulose in their forest environment. Forest trees are able to survive the grazing pressure of the ants, but the ants are serious pests of many cultivated trees and other crops when nests are located at the edges of forests adjacent to cultivated areas.

Escamoles are eaten by all social classes in Mexico and the ants have been described as the most enjoyable and expensive edible insect in the markets. Although called "ant eggs," *escamoles* are mainly mature larvae/pupae of two species of the genus *Liometopum, L. apiculatum* and *L. occidentale* var. *luctuosum.* Digging out the underground nest where the *escamoles* are found is very labor intensive. After harvest of ants from the nest (two or three times per year between February and June), the nest is covered with nopal, dried grass, and fresh weeds to maintain an environment suitable for survival and regrowth of the colony. People who collect *escamoles,* known as *escamoleros,* sometimes make more money during the harvest season than other rural people make during the entire year. The *Liometopum* ant is considered such a special treat in Mexico that it is the subject of songs, dances, and festivities.

Honey ants are a source of sweet treats in Mexico (species of *Myrmecocystus*) and Australia (several species in the genera *Camponotus* and *Melophorus*). Specialized worker ants, called repletes, store the honey in the abdomen, which may become the size of a small marble. They are eaten by grasping the head of the ant and sucking the honey from the abdomen. In Australia, *Camponotus inflatus* develops the largest repletes and they are considered a great luxury by the indigenous population. The repletes are found in galleries in the underground nests, where they are immobile and must be fed by the workers. Some aborigines expend much time and effort digging for the repletes but they only partially dig up the nests so as not to destroy the colonies and thus to preserve this valuable resource.

See Also the Following Articles

Commercial Products from Insects • Cultural Entomology

Further Reading

Bukkens, S. G. F. (1997). The nutritional value of edible insects. *Ecol. Food Nutr.* **36,** 287–319.

de Conconi, J. R. E. (1982). "Los Insectos como Fuente de Proteinas en el Futuro." Editorial Limusa, Mexico City.

DeFoliart, G. R. (1995). Edible insects as minilivestock. *Biodiversity Conserv.* **4,** 306–321.

DeFoliart, G. R. (1997). An overview of the role of edible insects in preserving biodiversity. *Ecol. Food Nutr.* **36,** 109–132.
DeFoliart, G. R. (1999). Insects as food: Why the Western attitude is important. *Annu. Rev. Entomol.* **44,** 21–50.
Dreyer, J. J., and Wehmeyer, A. S. (1982). On the nutritive value of mopanie worms. *S. Afr. J. Sci.* **78,** 33–35.
Finke, M. D., DeFoliart, G. R., and Benevenga, N. J. (1989). Use of a four-parameter logistic model to evaluate the quality of the protein from three insect species when fed to rats. *J. Nutr.* **119,** 864–871.
Malaisse, F., and Parent, G. (1980). Les chenilles comestibles du Shaba meridional (Zaire). *Nat. Belges* **61(1),** 2–24.
Menzel, P., and D'Aluisio, F. (1998). "Man Eating Bugs: The Art and Science of Eating Insects." Ten Speed Press, Berkeley, CA.
Oliveira, J. F. S., Passos de Carvalho, S. J., Bruno de Sousa, R. F. X., and Magdalena Sinao, M. (1976). The nutritional value of four species of insects consumed in Angola. *Ecol. Food Nutr.* **5,** 91–97.
Paoletti, M. G., and Bukkens, S. G. F. (eds.) (1997). Minilivestock. *Ecol. Food Nutr.* **36,** 95–346. [Special Issue]
Ramos-Elorduy, J. (1998). "Creepy Crawly Cuisine: The Gourmet Guide to Edible Insects." Park Street Press, Rochester, VT.
Silow, C. A. (1983). Notes on Ngangala and Nkoya ethnozoology: Ants and termites. *Ethnol. Stud.* **36,** 1–177.
Taylor, R. L., and Carter, B. J. (1976). "Entertaining with Insects, or: The Original Guide to Insect Cookery." Woodbridge Press, Santa Barbara, CA.
Tindale, N. B. (1953). On some Australian Cossidae including the moth of the witjuti (witchety) grub. *Trans. R. Soc. S. Aust.* **76,** 56–65.

Forensic Entomology

M. Lee Goff
Chaminade University of Honolulu

In its broadest sense, forensic entomology includes any situation in which insects or their actions become evidence within the legal system. Medicocriminal entomology involves insects as evidence in a criminal case, most frequently homicide, and this is the area that has been most closely associated with the term "forensic entomology" by the general public and, in fact, most entomologists. The use of insects and other arthropods as evidence in criminal investigations dates from 12th century China. Other records appear sporadically in both the forensic and the entomological literature from various parts of the world until a resurgence of interest in the field in the mid-1980s. Prior to this period, the primary application of entomological evidence was to determine the postmortem interval of decomposed bodies. Although this remains the primary application of forensic entomology in criminal investigations, it is now recognized that insects and other arthropods can provide insights into movement of a corpse following death, assessment of wounds (antemortem versus postmortem), characteristics of a crime scene, and abuse and neglect of children and the elderly, as well as serving as alternate specimens for toxicological analyses and sources of human DNA.

INSECTS AS INDICATORS OF THE POSTMORTEM INTERVAL

Decomposing remains provide a temporary microhabitat offering a progressively changing food source to a variety of organisms, ranging from bacteria and fungi to vertebrate scavengers. The arthropods constitute a major element of this fauna, with the insects as the predominant taxa in terrestrial environments worldwide, in terms of both numbers of individuals and species diversity. In North Carolina, for example, 522 species in three phyla were recovered from decomposing pig carcasses, and of these, 84% were insects. In the Hawaiian Islands, 133 different kinds of arthropods were collected from pig and cat carcasses and of these, 83% were insects. There have been numerous decomposition studies conducted worldwide using different animal models, ranging from lizards and toads to elephants. There has been considerable variation in the numbers of different taxa recovered. These differences may be related to both geographic variation and differences in the animal models used.

RELATIONSHIPS OF INSECTS TO THE REMAINS

The use of insects to estimate the postmortem interval requires an understanding of the insect's life cycle, the relationship of the insect to the remains, and the relationship of the remains to the habitat in which they are discovered. Insects pass through a number of distinct stages during their life cycle. Using a blow fly in the family Calliphoridae as an example, the female fly arrives at the body and deposits eggs in body openings associated with the head, anus, and genitals, or in wounds. After hatching, larvae or maggots feed on the decomposing tissues. There are three larval stages, with a molt in between each stage. Once the maggot is fully developed, it ceases to feed and moves away from the remains before pupariation. The puparium is an inactive stage during which the larval tissues are reorganized to produce the adult fly.

The insects encountered on a corpse in any given habitat consist of species unique to that particular habitat and those having a wider distribution. The unique components may be restricted to a particular geographic area or a particular habitat type within a given geographic area. Those taxa having wider distributions are frequently encountered in several different habitat types and are typically highly mobile species. Many of those taxa closely tied to carrion show this wider pattern of distribution. In estimating the postmortem interval, taxa from both components may, under given circumstances, provide essential information on the history of the corpse.

Of those insects having a direct relationship to the corpse, there are four basic relationships, as described below.

Necrophagous Species

Those taxa feeding on the corpse compose this group. This includes many of the Diptera [Calliphoridae (blow flies) and

Sarcophagidae (flesh flies)] and Coleoptera (Silphidae and Dermestidae). These species may be the most significant taxa for use in postmortem interval estimates during the earlier stages of decomposition, defined here as days 1 to 14.

Parasites and Predators of Necrophagous Species

This is the second most significant group of carrion-frequenting taxa and includes Coleoptera (Silphidae, Staphylinidae, and Histeridae), Diptera (Calliphoridae and Stratiomyidae), and hymenopteran parasitoids of larvae and puparia of Diptera. In some instances, dipteran larvae that are necrophages during the early portions of their development turn into predators.

Omnivorous Species

Ants, wasps, and some beetles, which feed on both the corpse and associated arthropods, compose this group. Large populations of these may severely retard the rate of carcass removal by depleting populations of necrophagous species.

Adventive Species

This category includes those taxa that use the corpse as an extension of their own natural habitat, as in the case of the Collembola, spiders, and centipedes. Acari in the families Acaridae, Lardoglyphidae, and Winterschmidtiidae that feed on molds and fungi growing on the corpse may be included in this category. Of less certain association are the various Gamasida and Actinedida, including the Macrochelidae, Parasitidae, Parholaspidae, Cheyletidae, and Raphignathidae, that feed on other acarine groups and nematodes.

DECOMPOSITION

Although there have been many decomposition studies conducted in different parts of the world and under different environmental conditions, most studies have been conducted in temperate areas and fewer in tropical and subtropical habitats. Common to the majority of these studies has been an attempt to divide the decompositional process into a series of discrete stages. Decomposition is, in nature, a continuous process and so discrete combinations of physical parameters and arthropod assemblages do not occur. There is a value to these stages, however, in providing reference points when faced with the problem of explaining the events associated with decomposition to a jury.

Regardless of locality, there are certain common patterns. The faunas involved tend to be regional, except for some widely distributed species of Diptera and Coleoptera, but the families involved are somewhat stable. The division of decomposition into five stages can be applied to most studies.

Fresh Stage

This stage begins at the moment of death and ends when bloating is first evident. The first insects to arrive at the corpse are flies in the families Calliphoridae and Sarcophagidae. Adult females investigate the corpse, frequently feed, and then, depending on the species of fly, deposit either eggs or larvae around the natural body openings associated with the head (eyes, nose, mouth, and ears) and anogenital regions. Wounds are secondary sites of attraction to tropical species but also may be of major significance in temperate environs.

Bloated Stage

Putrefaction, the principal component of decomposition, begins. Gases produced by the metabolic activities of anaerobic bacteria first cause a slight inflation of the abdomen and, later, the corpse appears balloon-like and fully inflated. Internal temperatures rise during this stage as the result of bacterial decay and metabolic activities of feeding dipteran larvae. Calliphoridae are strongly attracted to the corpse during this stage. As the corpse inflates, fluids are forced from natural body openings and seep into the soil. These fluids combined with the by-products (ammonia, etc.) produced by the metabolic activities of the dipteran larvae cause the soil beneath the corpse to become alkaline and the normal soil fauna departs.

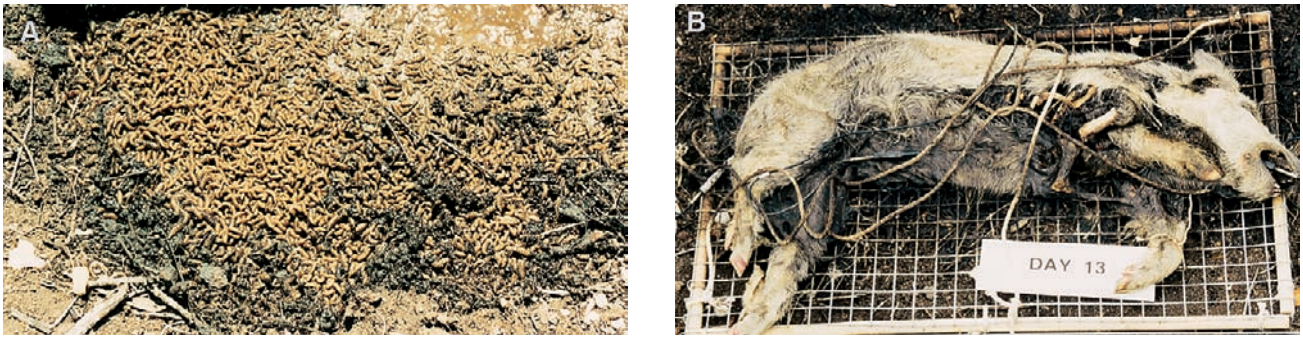

FIGURE 1 Pig carcass during decay stage of decomposition. (A) Day 8: active maggot mass consists primarily of third instar *Chrysomya rufifacies.* (B) Day 13 (end of decay stage): maggots have completed development and migrated away from carcass for pupariation.

Decay Stage

This is the only stage in the decomposition process which has a distinct starting point. The decay stage begins when the skin is broken, allowing gases to escape and the corpse to deflate. Dipteran larvae form large feeding masses and are the predominant taxa present (Fig. 1A). Although some predatory forms, such as beetles, wasps, and ants, are present during the bloated stage, both necrophagous and predatory taxa are observed in large numbers during the later portions of the decay stage. By the end of this stage, most Calliphoridae and Sarcophagidae have completed their development and departed the corpse for pupariation (Fig. 1B). Dipteran larvae will have removed most of the soft tissue from the corpse by the end of the decay stage.

Postdecay Stage

As the remains are reduced to skin, cartilage, and bone, Diptera cease to be predominant. In xerophytic and mesophytic habitats, various Coleoptera predominate throughout this stage, and the diversity of these taxa increases. Associated with this increase is an increase in the numbers of parasites and predators of beetles. In wet habitats (swamps, rain forests, etc.), however, other taxa, primarily Diptera, and their predator/parasite complexes predominate.

Skeletal Stage

This stage is reached when only bones and hair remain. No obvious carrion-frequenting taxa generally are present and there is a gradual return of the normal soil fauna to the area under the corpse. An examination of the soil during the early portions of this stage will reveal various acarine groups that may be of use in estimating the postmortem interval. There is no definitive end to this stage and changes in the soil fauna may be detectable months or even years following the death.

BIOLOGY OF DIPTERA

One of the major problems facing forensic entomologists is the accurate identification of the larvae collected from the remains. Too frequently, the entomologist must work with dead specimens collected by crime scene investigators and submitted in marginal states of preservation. Even when the local faunas are well known, identification of these specimens is difficult, especially for early instars. Work by Erzinclioglu in England and Liu and Greenberg in the United States has provided identification keys to larvae of forensically important taxa. Given the wide distributions of many of the sarcosaprophagous taxa, these keys have wider application than their regional nature implies.

Seasonal variation in the populations of Calliphoridae have been documented in North America and Europe. There are also successional patterns in Calliphoridae. In northern Mississippi, for example, *Phaenicia caeruleiviridis* was the first species to arrive during spring, whereas *Cynomyopsis cadaverina* was the first during the fall and winter months.

Considerable emphasis has been placed on how temperature influences the duration of the various stadia for different species of flies. Accumulated degree hours (ADH) or accumulated degree days (ADD) can be used to estimate the postmortem interval. Ambient temperature data from weather stations in the vicinity of the corpses are used as an indicator of the temperatures at which larvae developed. These temperatures, however, may not reflect the temperatures at which the larvae actually develop. Based on a Hawaiian study, internal temperatures associated with the maggot masses can be as much as 22°C above ambient. Similar observations have been made for human corpses in the former Soviet Union.

In instances of heavy maggot infestations, it is obvious that there is little, if any, direct relationship between ambient air temperature and the temperatures at which the maggots are developing. Although heat generated by maggot masses influences the rate of larval development, this heat generation may not occur immediately, but requires a period of several days to develop. Such a delay may be because of a lack of an organized maggot mass during the early instars. For corpses found during cool weather when colonizing fly populations are low, ADH or ADD calculations generally are more accurate than in higher temperatures with high fly populations. In addition to the rate of development, temperatures may also serve to limit the species that can use the corpse for development. Only some species of Calliphoridae can tolerate high temperatures inside a maggot mass during development. Thus, maggot-generated heat can influence the rate of maggot development, the nature of the corpse arthropod community, the character of the subcorpse community, and the validity of ADH or ADD calculation-based postmortem interval estimates.

FACTORS DELAYING INVASION OF THE CORPSE

The initial invasion of the corpse by insects and other arthropods starts the clock that is ultimately interpreted to give the estimate of the time since death. The basic assumption underlying this is that the invasion occurs soon after death. In decomposition studies, fly activity begins as soon as 10 min after death, but there are factors that may delay this invasion. Invasion can be delayed by wrapping of the corpse or submersion in water. Adverse climatic factors such as cloud cover, temperature, and rainfall may inhibit adult fly activity. Darkness has long been believed to inhibit calliphorid activity, but nocturnal oviposition has been observed in Calliphoridae commonly associated with decomposing human remains in North America and Hawaii. In these instances, temperatures during oviposition were above 20°C. Although nocturnal activity and oviposition may be expected for some species in tropical habitats, it appears to

be the exception rather than the rule in temperate regions. Clearly, care must be taken in forensic interpretations dealing with nocturnal activity of Calliphoridae.

APPLICATIONS TO DEATH INVESTIGATIONS

In the preceding sections, the biological basis for using insects as evidence in estimating postmortem interval (time since death), patterns of decomposition, and several of the problems have been presented. An understanding of insect biology and ecology is necessary as a template for determining postmortem interval, but it does not always provide an accurate estimate. In processing insect data, a forensic entomologist must bear in mind that an estimate based on insect activity only determines the period of insect activity. This time frame may not correspond to the entire period of time since death. Each case is unique and must be analyzed on its own merits with careful attention to all data available.

As noted earlier during discussions of decomposition, there is great geographic and seasonal variation in insect populations associated with a corpse. Even in areas where the species populations are similar, there may be seasonal changes in developmental patterns which may complicate the analyses.

In addition to gathering and processing of entomological data, it is necessary to develop protocols for cooperation between the forensic entomologist and various law enforcement agencies, medical examiners, coroners, and the courts. As these individuals better understand the functioning of the forensic entomologist, the end results of the investigations will improve.

COLLECTION PROTOCOLS

Several workers have independently formulated their own protocols for collecting and processing entomological evidence. Among those published, Lord and Burger provide the best general procedures.

ESTIMATION OF THE POSTMORTEM INTERVAL

In the actual process of estimating the postmortem interval from entomological data, each case will be unique. Regardless, the process tends to follow the same general sequence of events:

1. The stage or physical state of decomposition based on physical parameters for the corpse should be determined, and any indications of disturbance or dismemberment of the corpse that may have occurred following death are noted. If the community collected under the corpse does not conform to the observed stage during later analysis, the possibility of movement of the corpse following death must be considered.
2. Specimens collected from the corpse and crime scene must be identified as completely as possible. Immature stages must frequently be reared to the adult stage for final species-level identifications. Representative samples of immatures (species and stages) must be preserved properly to provide a record of what was present on the corpse at the time of examination.
3. For outdoor scenes, appropriate climatic data (temperatures, rainfall, cloud cover, etc.) from weather stations and on-scene observations must be obtained. Aspects of the scene that may serve to influence the effects of these climatic factors on arthropod invasion of the corpse should be considered (cover by vegetation, shading by trees, slope of the ground, burial of the corpse). For indoor scenes, the temperatures (automatic heating, thermostat settings, air conditioner) possible for the time period in question should be noted; positioning of the corpse relative to windows and doors may be significant in terms of both heat and solar radiation.
4. From the autopsy report or one's own observations, the sites of infestations by arthropods should be noted.
5. The postmortem interval is estimated. In the earlier stages of decomposition, this estimate may be based on the developmental cycles of dipteran larvae. In simplest form, the time required to reach the most mature stage of development of the earliest arriving species on the corpse under prevailing conditions would correspond to the minimum postmortem interval. Further consideration must be given to factors that could delay the onset of insect activity (climatic factors, wrapping of the corpse, seasonal variation). When these factors are considered, the final postmortem interval estimate may be greater than the estimated faunal ages. During this comparison, both presence and absence of taxa and developmental stage must be considered. For this reason, it is essential that collections from the corpse and the surrounding area be as complete as possible. In general, the parameters for the postmortem interval estimate will become wider as the time since death increases. During the earlier stages of decomposition, the estimate may be expressed conveniently in terms of hours, whereas later it may be in days, months, or even seasons of the year.

MOVEMENT OF THE BODY FOLLOWING DEATH

Although insects and other arthropods are among the most widely distributed organisms on earth, they are often quite specific in their distributions. As noted earlier, the fauna of the decomposing corpse is composed of species having a wide distribution and species specific to the particular area in which the body is discovered. Presence of species not typically associated with the habitat in which the body is discovered is an indication that the victim died in one location and the body was exposed to insect activity for a period of time. Following this, the body was transported to another location where a second colonization by insects took place.

ASSESSMENT OF TRAUMA

The initial sites of fly egg-laying activity on a decomposing body will normally be the natural body openings associated with the head, anus, and genitals. For many species, the pre-

sence of blood associated with wounds is also attractive and egg-laying occurs at those sites. A wound that occurs prior to death (antemortem), while the heart is still beating, produces blood, which is attractive to the flies. If the wound occurs following death (postmortem), similar quantities of blood are not associated with the wound and it will be less attractive to flies for oviposition. If a body is encountered during the earlier stages of decomposition with significant infestations of maggots in areas other than the natural body openings, the possibility of antemortem wounds must be considered. In like manner, if there are wounds present on the body, but infestations are primarily restricted to the natural body openings, these wounds may well be postmortem artifacts.

ABUSE AND NEGLECT

Although most of the applications of entomological evidence involve the dead, in some instances entomological evidence may involve living victims. In these instances, the maggots are encountered feeding on live tissues. This is a condition known as myiasis and is an obligate condition in the life cycles of many flies, primarily in the families Calliphoridae and Sarcophagidae. In other species of flies, myiasis may be a facultative situation. Stages of development recovered from wounds or sores on children, the elderly, or those otherwise unable to care for themselves can be used to document the period of abuse or neglect.

An additional potential problem is in situations in which myiasis occurs prior to death. If the sites of infestations are not noted, the period of development of the maggots on the living individual may be added to the postmortem development, thus increasing the estimated time since death. Any departure from the normal pattern of arthropod invasion of a corpse should be a cause for care in interpretation of entomological evidence.

ENTOMOTOXICOLOGY

Over the past 2 decades, drug-related deaths have increased in the United States and other countries. In many instances, these deaths are not immediately reported and the remains may be undiscovered for several days. Because of decompositional processes, estimations of the time of death or postmortem interval are based on analyses of insects and other arthropods infesting the remains. The data most frequently employed are those associated with insect development rates and successional patterns. Recently, the accuracy of these estimations has been questioned in deaths involving narcotic intoxication. Relatively few studies are currently available detailing the effects of drugs, such as cocaine and heroin, in decomposing tissues on the rates and patterns of development of carrion-feeding arthropods. Additionally, there are few data dealing with effects of other tissue contaminants, such as toxins and environmental pollutants, in decomposing tissues on rates and/or developmental patterns of arthropods using such tissues as food. Interest also has focused on the potential use of arthropods as alternate specimens for toxicological analyses, in situations in which more normal specimens of blood, tissue, or body fluids are not available. These two areas now comprise entomotoxicology.

Detection of Drugs and Toxins

It is not unusual for remains to be discovered in a highly decomposed or skeletal stage, when there is insufficient tissue for toxicological analyses. There frequently are, however, arthropods or their cast larval or puparial skins still associated with the remains. Various toxic and controlled substances can be detected by analyses of these arthropods and their residues. Generally, arthropod materials have been homogenized and then processed in the same manner as other tissues or fluids of toxicological interest. Analytic procedures include radioimmunoassay, gas chromatography, thin layer chromatography, and high-performance liquid chromatography–mass spectrometry.

Effects of Drugs on Development of Insects

Although many of the studies mentioned documented the potential for use of maggots and puparia as alternate specimens for toxicological analyses, few were concerned with the potential effects of these drugs on the development of the insects ingesting them. In providing an estimate of the postmortem interval, particularly within the first 2 to 4 weeks of decomposition, it has been assumed that the insects will develop at predictable rates for given environmental conditions. That this might not always be true was established by studies on the effects of cocaine on development of the sarcophagid *Boettcherisca peregrina.* In this example, maggots were reared on tissues from rabbits that had received known dosages of cocaine, corresponding to 0.5, 1.0, and 2.0 times the median lethal dosage by weight. Two patterns of development were noted. Control and sublethal-dosage colonies developed at approximately the same rate, as indicated by total body length. In contrast, the colonies fed on tissues from the lethal and twice-lethal dosages developed more rapidly. This difference continued until maximum size was attained and the postfeeding portion of the third instar was reached. Due to the increased rate of development during the feeding stages, pupariation occurred first in the lethal and twice-lethal colonies, but the actual duration of the puparial period was the same for all colonies and there were no detectable differences in puparial mortality.

The potential significance of these alterations in the rates of larval and puparial development is illustrated in trying to establish the postmortem interval of a Caucasian woman, approximately 20 years of age, discovered in a pine woods area northeast of Spokane, Washington. The body was in the early bloated stage of decomposition and had extensive populations of maggots on the face and upper torso. Maggots

were submitted to the entomologist after being refrigerated for 5 days and reared to the adult stage. Two species were identified from the adults: *C. cadaverina* and *Phaenicia sericata.* Typically, *P. sericata* oviposits within 24 h following death, whereas *C. cadaverina* oviposits 1 to 2 days following death. Three size classes of maggots were present on the corpse. The first consisted of maggots measuring 6 to 9 mm in length that were consistent of a period of development of approximately 7 days. The second consisted of smaller maggots, consistent with continued oviposition by adult flies. The third consisted of a single maggot measuring 17.7 mm in length and indicative of a developmental period, under prevailing conditions at the scene, of approximately 3 weeks. Given the other data associated with the case, this period did not seem possible. The possibility that this maggot had migrated from another nearby source was eliminated, as no carrion could be located nearby and the probability of only a single maggot migrating was low. The alternate explanation was that the maggot's growth rate had been accelerated in some manner. It was learned that the victim had a history of cocaine abuse and that she had snorted cocaine shortly before her death. This maggot had most probably developed in a particular pocket in the nasal region containing a significant amount of cocaine.

SOURCES OF HUMAN DNA

Recent advances in technology have provided means to identify individual hosts from analyses of hematophagous arthropod blood meals. Using polymerase chain reaction and amplified fragment length polymorphism DNA characterization procedures, two human DNA genetic markers, D1S80 and HUMT01, have been typed from human DNA material derived from excreta from the human crab louse, *Pthirus pubis,* fed on human volunteers. Although these results are preliminary, they demonstrate the potential for this technology to provide individual characterizations for cases of rape or homicide in which hematophagous arthropods are encountered by investigators at the scene. More recently, human mitochondrial DNA (mtDNA) has been isolated, amplified, and sequenced from crab lice fed on human volunteers. This study demonstrates the potential for mtDNA analyses to characterize individual hosts even from desiccated and frozen arthropod blood meals, and it is yet another example of the stability of mtDNA.

CONCLUSIONS

Applications of entomological evidence have increased significantly since the 1980s. Although the major application is estimation of the postmortem interval, entomological evidence also can be applied to toxicology, DNA analyses, and aspects of crime scene assessment. Although there have been many advances, there is still a need for much basic research. Questions remain with regard to the life cycles of many of the necrophagous arthropods, particularly Coleoptera. One problem in application of entomological evidence is the diversity of the insect fauna which changes from one geographic area to the next. Databases are still needed for many parts of the world where the sarcosaprophagous fauna is poorly known. Even in areas relatively well investigated, there may be significant microgeographic variation in the insect fauna. Additional investigations are needed in the areas of toxicological analyses and applications of DNA technology. A current underlying problem is the relative lack of trained forensic entomologists. Most individuals now working in the field are employed in academic institutions where only a small part of their research effort is assigned to forensic concerns.

See Also the Following Article

Medical Entomology

Further Reading

Catts, E. P., and Goff, M. L. (1992). Forensic entomology in criminal investigations. *Annu. Rev. Entomol.* **37,** 253–272.

Early, M., and Goff, M. L. (1986). Arthropod succession patterns in exposed carrion on the island of O'ahu, Hawaiian Islands, USA. *J. Med. Entomol.* **23,** 520–531.

Goff, M. L. (1993). Estimation of postmortem interval using arthropod development and successional patterns. *Forens. Sci. Rev.* **5,** 81–94.

Goff, M. L. (2000). "A Fly for the Prosecution." Harvard University Press, Cambridge, MA.

Goff, M. L., Charbonneau, S., and Sullivan, W. (1991). Presence of fecal material in diapers as a potential source of error in estimations of postmortem interval using arthropod development rates. *J. Forens. Sci.* **36,** 1603–1606.

Goff, M. L., Omori, A. I., and Goodbrod, J. R. (1989). Effect of cocaine in tissues on the rate of development of *Boettcherisca peregrina* (Diptera: Sarcophagidae). *J. Med. Entomol.* **26,** 91–93.

Greenberg, B. (1991). Flies as forensic indicators. *J. Med. Entomol.* **28,** 565–577.

Liu, D., and Greenberg, B. (1989). Immature stages of some flies of forensic importance. *Ann. Entomol. Soc. Am.* **82,** 80–93.

Lord, W. D., and Burger, J. F. (1983). Collection and preservation of forensically important entomological materials. *J. Forens. Sci.* **28,** 936–944.

Payne, J. A. (1965). A summer carrion study of the baby pig *Sus scrofa* Linnaeus. *Ecology* **46,** 592–602.

Schoenly, K., Goff, M. L., and Early, M. (1992). A BASIC algorithm for calculating the postmortem interval from arthropod successional data. *J. Forens. Sci.* **37,** 808–823.

Smith, K. V. G. (1986). "A Manual of Forensic Entomology." British Museum (Natural History), London.

Forest Habitats

David L. Wood
University of California, Berkeley

Andrew J. Storer
Michigan Technological University, Houghton

Forest habitats exhibit extraordinary diversity. Within them are complex assemblages of species, and myriad interactions

occur among these species. Forest habitats comprise not only forested landscapes but also associated lakes, streams, and meadows. All orders of insects can be found in forested habitats. Even insects that occur in salt water can be found in some forest habitats, such as in tropical mangrove forests. All feeding habits are thus represented, that is, scavengers (especially in soils and beneath the bark of dead trees) and phytophagous, entomophagous, and parasitic species. In California, over 5000 species of insect inhabit oak forests alone. In this article we give greatest emphasis to those orders that feed on trees and the predators and parasitoids that consume these phytophagous species. Insects found primarily in soils, lakes, and streams; those that feed on other vegetation found in forests; and those that parasitize other animals are considered elsewhere in this encyclopedia.

THE DIVERSITY OF FOREST HABITATS

Forest habitat diversity is a reflection of geographic location of the forest, as well as the influences of humans on the forest. The broadest definition of a forest habitat is that of a habitat in which trees are a significant component. Natural forests also vary in complexity based on species composition resulting from a combination of environmental factors. A number of classification systems for natural forests have been proposed, but perhaps the most useful is that developed by the Food and Agriculture Organization of the United Nations in the 1970s. This system divides the natural forests of the world into five broad categories:

1. The cool coniferous forests are a circumpolar belt of boreal forests across northern latitudes. They occur between tundra to the north and temperate mixed forests to the south. Conifers are the dominant species in these forests, but they also include a few species of broadleaf trees.
2. The temperate mixed forests are found south of the cool coniferous forests and also are found in parts of the southern hemisphere. Pines and deciduous and evergreen broadleaf trees tend to dominate these forests.
3. Tropical moist evergreen forests are rainforests characterized by high annual precipitation (>2000 mm) evenly distributed during the year. Amazonia, western equatorial Africa, and the Indo-Malayan region are the three main regions for this forest type.
4. Tropical moist deciduous forests are found in tropical regions with 1000–2000 mm annual rainfall and a dry season for 1 or more months. These forests include monsoon forests in Asia where a dry period of 2 to 6 months is followed by heavy rains.
5. Dry forests are found in both temperate and tropical zones where annual precipitation is less than 1000 mm and are low and simply structured wooded areas.

These forest types may be further subdivided according to dominant species types and climatic differences among other factors. Additional forest types can also be recognized, including those heavily influenced by humans, such as urban forests and plantations.

The species of host plants naturally influence the species of phytophagous insects present in a forest. Additionally, human influences affect the species and density of insects in a forest. Forest management in natural forests may involve fire suppression that may favor insects that can exploit shade-tolerant tree species. For example, the fir engraver, *Scolytus ventralis,* has increased in density in mixed conifer forests of the Sierra Nevada in California because fire suppression has led to a dense understory of white fir, *Abies concolar*. Regeneration of shade-intolerant pines is suppressed and this in turn leads to an overly dense forest dominated by white fir, *Abies concolar*. Periodic logging or thinning may result in a build up of insect populations in cut material or associated waste. For example, neglect of forest hygiene following thinning operations in European pine forests may result in an increase in populations of the European pine shoot beetle, *Tomicus piniperda,* which subsequently may damage the shoots of standing trees during its adult feeding stage. Clear-cutting and replanting or natural regeneration of forests favors insects that feed in young trees. For example, the pine weevil, *Hylobius abietis,* is perhaps the most significant forest insect pest in Europe. Adults feed on the bark at the base of conifer seedlings, resulting in seedling mortality. In many locations, any efforts to establish a plantation or naturally regenerated forest must have a management protocol to reduce the impact of this weevil.

In human-made forests, the species composition may be very simple, as in monocultures, or more complex, depending on the management goals of the plantation. Age and size classes of trees are frequently very limited. These plantations may comprise native or exotic tree species and may be on previously forested land or on previously unforested land. Perhaps the simplest forest habitat is a plantation monoculture of even-aged trees (such as in Monterey pine, *Pinus radiata*, plantations in New Zealand or loblolly pine, *P. taeda*, plantations in areas of the southeastern United States). This habitat is nevertheless very diverse as a result of the array of different ecological niches present on the dominant tree species. Usually, plantation monocultures involve areas of forest that are even-aged and that therefore encounter pressure from phytophagous insects that are adapted to, or more pestiferous during, certain stages of forest growth. For example, the gouty pitch midge, *Cecidomyia piniinopsis,* is a pest of young ponderosa pines, *Pinus ponderosa*, in the Sierra Nevada but is less significant on mature trees. Plantation forestry may involve tree species that are not native to the location where a forest is planted. Any insects that occur in these forests, but which were not present prior to the existence of that forest, are therefore introduced. Often these insect populations are not under natural regulatory pressures that limit their numbers. For example, as spruce forests have been planted in areas of Europe, using spruce as exotic species, the European spruce beetle, *Dendroctonus micans,* has moved westward from Eurasia during the 20th century. Outbreaks of this bark beetle have

been most severe at the edge of its expanding range before natural enemies spread or are introduced into the areas now exploited by *D. micans.* In the creation of new forest habitats, the interactions among the species in the forest are unpredictable, because an essentially novel habitat has been created.

DIVERSITY OF INSECTS IN FORESTS

Among all plants, trees present the most diverse habitats for insects to occupy. Insects feed on all parts of the tree, i.e., vegetative structures such as leaves, stems, and roots and reproductive structures such as flowers, fruits, and seeds. Some insects are specialized to feed on phloem and/or xylem tissues, dead sapwood, and heartwood. Insects that feed on these structures and tissues vary in size from 1–2 mm (scale insects) to 6 cm (longhorned beetles). Life cycles (from egg to adult) can be completed in a few days or weeks (aphids) or be prolonged for 50 years (metallic wood borers).

Although all orders of insects are found in forest habitats, only a small number feed on the trees or are the predators and parasitoids of these taxa. Species from the following orders are generally referred to as "forest insects": Hemiptera, Isoptera, Orthoptera, and Thysanoptera (in the Exopterygota, which undergo incomplete metamorphosis) and Coleoptera, Diptera, Hymenoptera, Lepidoptera, and Raphidioptera (in the Endopterygota, which undergo complete metamorphosis).

Habitat diversity is greatly favored by the large size of trees, both in mass and in height. Thus, both the abiotic and the biotic environment can vary considerably from the roots to the upper canopy. For example, a bark beetle, *Hylastes nigrinus,* spends its life cycle feeding beneath the bark on the roots of Douglas-fir, *Pseudotsuga menziesii,* growing in California, whereas a cone midge, *Contarina oregonensis,* feeds in a gall in the seed coat within a cone. Old-growth Douglas-firs with a height over 100 m and a diameter of 3 m are not uncommon. Over 240 insect species from the above-mentioned orders are listed as feeding on Douglas-firs. However, only a few of these would be found at the same time on a large, living tree. Similar diversity of insects feeding on trees can be found in broadleaved deciduous forests and broadleaved evergreen forests. Deciduous oaks in England are fed upon by over 280 insect species, and some evergreen oaks in coastal California provide food for about 300 insect species.

This complexity is magnified further when the guilds of insects in the canopy are considered. The phytophagous insects are represented by chewing, mining, gall-making, and sap-sucking species. Chewing insects are found in the Coleoptera, Lepidoptera, Hymenoptera, and Orthoptera; mining species in the Hymenoptera, Lepidoptera, and Diptera; gall-making species in the Diptera and Hymenoptera; and sap-sucking species in the Hemiptera (Homoptera and Heteroptera) and Thysanoptera. Predators are found in the Coleoptera, Diptera, Hymenoptera, and Raphidioptera and parasites in the Diptera and Hymenoptera. Insects that feed on epiphytes such as lichens, algae, and mosses are found in the Psocoptera (bark lice), Collembola (springtails), Dermaptera (earwigs), and Plecoptera (stoneflies). Mosquitos (Diptera) breed in water contained in tree holes, which are decayed cavities in the wood.

Many detritivores are found in the canopy as well as on other parts of the tree. They feed on many food sources such as protozoa, bacteria, fungi, nematodes, and small particles of plant and animal tissues. These species are found for example in the Psocoptera, Collembola, and Blattoidea.

Finally, many insects that do not feed on the trunk or foliage use them for a resting or hiding place, mating, pupation and/or hibernation, or estivation. These temporary residents have been classified as "tourists." Many of these insects feed on the surrounding vegetation. Their predators and parasitoids would be included as tourists also. For example, ants and spiders, which are very important predators of tree-inhabiting insects, use tree trunks as roadways from the forest floor to the canopy in search of prey. Another example, a sawfly, *Strongylogaster distans,* that feeds on bracken fern in the Sierra Nevada of California, pupates in the bark crevices of nearby ponderosa pines.

In summary, about half of the insect orders are directly or indirectly associated with trees. As with humans, insects use trees for food, shelter, support, and travel.

ROLE OF INSECTS IN FOREST SUCCESSION

Although trees can live for thousands of years (e.g., giant sequoia, *Sequoiadendron gigantea*), natural forests are dynamic plant communities that can change very slowly over thousands of years or very quickly over a few days to a few years. Fires, volcanic eruptions, strong winds, and snow and ice often have dramatic effects on forest succession. Insects can also have this effect during outbreaks (referred to as epidemics or gradations). For example, bark beetles, together with their associated blue-stain fungi, are known to kill millions of trees over thousands of hectares. The forest shown in Fig. 1 may change from an old

FIGURE 1 Mountain pine beetle, *Dendroctonus ponderosae* (Coleoptera: Scolytidae), killed lodgepole pine, *Pinus contorta* var. *latifolia,* in 1982 in Glacier National Park, Montana. [Photograph by Mark D. McGregor, www.forestryimages.org.]

growth (ca. 80 years), even-aged, predominantly lodgepole pine forest to a younger forest of lodgepole pine if a fire occurs subsequent to the mountain pine beetle infestation. Lodgepole pine cones often remain closed for many years and thus can store a large seed crop on the tree. Following a fire, which opens the cones, an almost pure forest of lodgepole pine is established over very large areas. Thus, the stage is set for another mountain pine beetle outbreak in 80 to 100 years.

Insects and fungi are the most important biotic agents that affect forest succession. In general, these organisms influence the number and growth rate of trees through space and time. Insects that feed on seeds influence reproduction of trees, shoot-feeding and defoliating insects influence growth rates of trees, and insects that infest the main stem of trees accelerate the mortality of trees in a forest. All of these effects occur through interactions with other biotic and abiotic agents as part of the complex forest cycle that involves gaps, building, and mature and degenerate phases. Insects therefore impact forest succession by influencing the forest cycle. The changes in these parameters affect the distribution and abundance of trees at a given moment in time and can be projected into the future using mathematical modeling techniques.

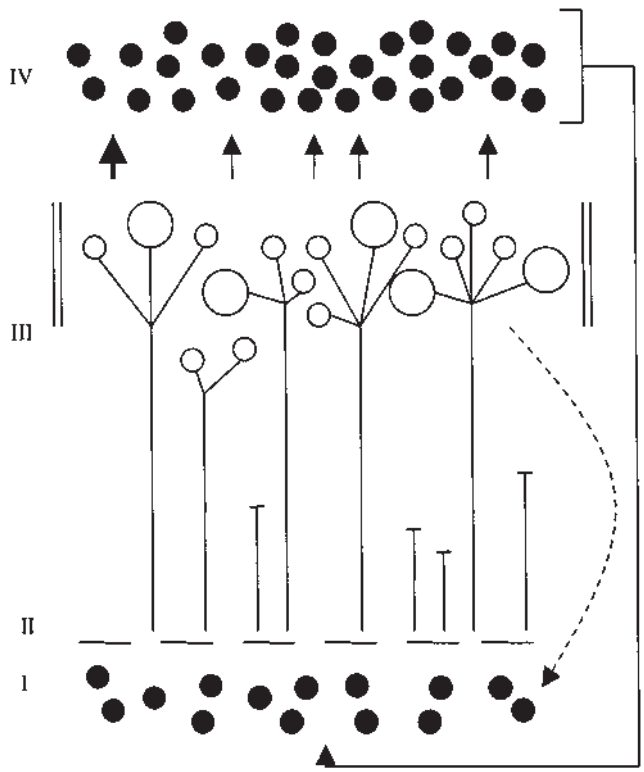

FIGURE 2 Diagrammatic representation of the plant life cycle, illustrating a tree as a series of modular units (the shoots). There are four basic components in the life cycle: (I) the seed bank, (II) the recruitment and establishment of individuals of the population from the seed bank, (III) the growth of individuals, and, (IV) reproduction and dispersal. (Reproduced from Coulson and Witter, "Forest Entomology: Ecology and Management. Copyright 1984. This material is used by permission from John Wiley & Sons, Inc.)

Insects affect the growth rate and mortality of trees by feeding on various parts of the tree. Thus they affect the life cycle of trees (Fig. 2) by influencing the size of the seed bank, the amount of recruitment, the growth of individuals in height and volume, and the reproduction and dispersal of the tree species. Insects consume seeds within fruiting structures or on the ground, thus reducing the size of the seed bank. They consume young seedlings, thus decreasing the recruitment of new trees into the population. They kill tips and shoots, suck plant fluids from the phloem and xylem, and consume the foliage, thus reducing the photosynthetic capacity of the tree and, as a result, reduce the growth of individuals. Insects also kill trees by reducing their growth rate so they cannot compete with other individuals of the same or different species for light, space, water, and nutrients. Trees are killed quickly, often in a few weeks, following severe defoliation by moths or sawflies and by bark beetles that introduce pathogenic fungi into the phloem and xylem. Direct and indirect interactions occur between insect species and between insects and forest pathogens. Root disease pathogens may indirectly affect bark beetle populations by weakening trees and thus predisposing them to attack by bark beetles. Direct interactions with pathogens may occur when, for example, insects transmit the pathogen, as with the pitch canker pathogen, *Fusarium circinatum,* in California and the Dutch elm disease pathogen, *Ophiostoma novo-ulmi,* in Europe and North America.

The examples drawn upon in this article emphasize those insect taxa that affect the survival of trees and in turn influence successional patterns of forests. This treatment thus includes phytophagous species and their natural enemies. Some authors refer to this interaction between insects and plants as a predaceous or parasitic relationship. The predator kills the host but the parasite does not kill the host directly. Insects may also form commensalistic and mutualistic associations with trees. In a commensalistic interaction, the tree is not affected, for example, when it is used as a resting place, whereas the tree benefits in a mutualistic interaction, such as with a pollinator. Our objective then is to emphasize the negative interactions between insects and their host trees, in which only the insect benefits and the tree is debilitated or killed. The tree does, however, benefit in the long term at the population level, at which succession allows nutrient turnover and reproduction allows adaptation of the tree species to environmental changes.

FEEDING GROUPS

Insects can be assigned to feeding groups based on the part of the tree they attack and the method of feeding they use. In forest habitats, these feeding groups are:

A. Insects that feed on cones and seeds
B. Insects that feed on shoots and tips
C. Insects that feed on foliage
D. Insects that feed on the trunk and large branches
E. Insects that feed on roots

In the following discussion, species have been selected from each feeding group to demonstrate the great diversity of insect adaptations to their tree hosts. An example is given from each of the major orders of insects associated with each feeding group. The examples are taken mostly from the Old (Palearctic) and New (Nearctic) World fauna that have received the greatest scientific attention to date. We select examples based on insect orders, knowledge base, and a diversity of biological attributes, all within each feeding group. This approach should give the greatest insights into the mechanisms of insect interactions with their host trees. Thus, the reciprocal responses of the host and the insect are considered at the level of the individual tree and how this affects forest successional patterns.

Insects That Feed on Cones and Seeds

Important species of cone and seed insects are found in the Coleoptera, Diptera, Heteroptera, Hymenoptera, and Lepidoptera. They damage the reproductive structures of trees, resulting in reduced seed production. These insects may feed on the seed itself or on structures associated with the seed. The number of insects feeding in conifer cones is very large when the small size of the feeding substrate is considered. For example, in western Europe, cone- and seed-feeding insects number some 59 species from 30 genera, 13 families, and 4 orders.

***Conophthorus* spp. (Cone Beetles); Coleoptera: Scolytidae (Bark and Ambrosia Beetles)** Species in this genus colonize only pine species in North America. Morphological divergence is greatly reduced and therefore, species separation is difficult. Thus, many species were described based on their host associations. Recent studies of pheromone components and DNA are helping to establish the status of these sibling (closely related and difficult to distinguish) species.

Except for the time spent between emergence from the brood cone and colonization of a new cone, the entire life cycle is spent within the cone. A single female tunnels into a new cone in the spring through the cone scales near the supporting stem (as in the "hard" pines, e.g., ponderosa pine) or into the supporting stem of "soft pines" (e.g., eastern white pine, *Pinus strobus*). She excavates a gallery along the cone axis and deposits eggs in the gallery wall. The gallery is packed with frass, which is composed of uneaten fragments of cone tissues and excrement. The male mates with the female inside the cone prior to gallery elongation. One female can infest more than one cone. In general, infested cones from the soft pines fall prematurely from the tree, while infested hard pine cones remain attached to the branch until they drop in the fall. Beetle attacks cause the death of the cone before and after the cone reaches full size. New brood adults can overwinter in the cone or emerge and mine twigs on living trees (*C. ponderosae* on sugar pine, *Pinus lambertiana*) or mine first-year cones (*C. radiatae* on Monterey pine) and spend the winter as nonreproducing adults.

Other insect species enter the galleries of the cone beetle. The dry twig and cone beetle, *Ernobius punctulatus* (Anobiidae), feeds in the drying cone tissues that follow cone beetle feeding activity. A parasitic wasp (Bethylidae: *Cephalonomia* species) enters the cone beetle tunnel to lay eggs. The wasp larvae then feed on the cone beetle larvae.

***Contarina oregonensis* (Douglas-Fir Cone Midge); Diptera: Cecidomyiidae (Gall Midges)** This midge infests cones of Douglas-fir in California, Oregon, Washington, and British Columbia. Eggs are laid in young female cones as they open for pollination in the spring. Larval feeding tunnels stimulate gall formation on the seed coat, which destroys the seed. Several galls may form on the seed coat. In the fall, the mature larvae leave the cone and fall into the litter beneath the tree where they spend the winter. There is one generation each year; however, some individuals remain in a resting stage in the litter for more than 1 year. Arrested development, termed diapause, is common in insects that infest cones and seeds. This mechanism permits the insects to survive through periods when few cones are produced. Many other species of insects can inhabit these cones with this midge, including moths, beetles, and wasps.

***Megastigmus spermatrophus* (Douglas-Fir Seed Chalcid); Hymenoptera: Torymidae (Torymids)** Most species in this family are parasites of gall-forming insects, which are usually other wasps or flies. However, in the Douglas-fir seed chalcid, the female inserts her long ovipositor through the cone scales and into the seed where she deposits an egg. The larva develops entirely within the seed coat. The seed coat continues to develop normally. In the spring of the following year the adult cuts a round hole in the mature seed coat and emerges to begin a new generation. Adult emergence may be delayed 2 to 3 years. Cones are harvested in the fall to extract seed for future plantings. Infested seeds cannot be distinguished from uninfested seeds by external appearance. This species has been introduced into Europe and has become a serious pest in plantations of Douglas-fir. Parasites reared from infested cones in western North America have been introduced into France to help reduce populations of this seed chalcid. *M. speculatris* was also introduced from North America and now infests cones of Siberian fir, *Abies sibirica* and Nordmanns fir, *A. nordmanniana* in Europe.

***Leptoglossus occidentalis* (Western Conifer Seed Bug); Heteroptera: Coreidae (Coreid Bugs)** This insect is named the "leaf-footed bug" because the tibiae of the hind legs are broad and flat while the tibiae of the first two pairs of legs are tubular in shape. This insect feeds on the seed of many conifer species, including Douglas-fir, incense cedar, *Libocedrus decurrens*, and ponderosa pine. This bug sucks plant juices, which results in damage to the seed. The eggs are barrel-shaped and are deposited in rows on the needles. The adults overwinter in protected areas including inside buildings. The broad host range of this species favors survival during years with low cone production in one or more host species.

***Cydia strobilella* (Spruce Seed Moth); Lepidoptera: Tortricidae (Tortrix Moths)** The female spruce seed moth oviposits between the cone scales of spruce flowers. After hatching, the larvae feed on the central stalk of the cone before moving to the developing seeds. Each larva consumes several seeds to complete its development. Fourth instars return to the cone stalk to overwinter and may remain in extended diapause for 2 or more winters. Higher than average summer temperatures result in extensive cone production in the following year and also stimulate the larvae of spruce seed moth to break diapause. However, even in good years, not all larvae within a cone break diapause. In some European seed orchards, over 50% of cones may be damaged by this species.

Insects That Feed on Shoots and Tips

Insects feeding in or on shoots and tips are found in the orders Coleoptera, Diptera, Hemiptera, and Lepidoptera. These insects may feed externally or internally on these structures. In addition, they may induce the tree to produce a gall as a result of their feeding activities. This gall may have nutritional as well as protective benefits to the insect. Some insects that feed externally gain protection by, for example, producing a spittle mass, as in spittlebugs, or wax, as in woolly aphids. Feeding may be on foliage in addition to the shoot and/or tip.

***T. piniperda* (Pine Shoot Beetle); Coleoptera: Scolytidae (Bark and Ambrosia Beetles)** In spring, adults of the pine shoot beetle emerge from overwintering sites in the "duff," the needle and bark litter on the ground. They also overwinter in short galleries below the litter layer. They attack trees weakened by root disease (e.g., annosus root disease caused by *Heterobasidion annosum*) and broken and cut trees, but are also capable of killing apparently healthy trees. There can be two or more early emergence periods that result in "sister" broods. This bark beetle infests several pine species, but Scots pine, *Pinus sylvestris,* is its principal host in Europe. The female excavates an egg gallery in the phloem, which is parallel with the grain of the wood. The male mates with the female in the egg gallery. She cuts niches in the gallery wall and deposits one egg in each niche. Larvae then tunnel laterally from the egg gallery. The larval gallery gradually becomes wider as the larvae develop. Pupation occurs in the early summer and emergence occurs in midsummer. These adults fly to living trees and tunnel into the axils of young shoots, which causes the shoots to die and break off. The feeding that occurs in the shoots is necessary for maturation of the gonads. The beetles emerge from the dying shoots and attack host material as described above. This habit has been termed "maturation feeding." Death of the terminal shoot causes flattened canopies. The pine shoot beetle was discovered in Ohio in 1992 and has since spread to adjoining states and Ontario, Canada. It infests eastern white pine and Scots pine Christmas tree plantations.

***C. piniinopsis* (Gouty Pitch Midge); Diptera: Cecidomyiidae (Gall Midges)** The gouty pitch midge feeds on a large number of pine species in North America. Only the hard pines are infested. It is usually found feeding on young, open-grown pines in both natural stands and plantations. Damaged shoots are first observed in early summer, when new shoots droop and turn yellow and then red. Repeated infestations in northern California slow tree growth and eventually kill the tree. Heaviest infestations occur on ponderosa pines exhibiting sticky twigs while lightest infestations occur on dry, powdery stems, suggesting that some genotypes are resistant to the feeding-induced damage caused by this midge. The bright red larvae overwinter in small resinous pits beneath the bark. In spring larvae migrate to the needles where they pupate. There is one generation per year. Dead shoots resemble frost-damaged tips.

***Adelges cooleyi* (Cooley Spruce Gall Adelgid); Homoptera: Adelgidae (Adelgids)** The Cooley spruce gall adelgid is distributed from coast to coast in North America where it infests many species of spruce, including Sitka, *Picea sitchensis*, Engelmann, *P. engelmannii*, white, *P. glauca*, blue, *P. pungens*, and Brewer, *P. brewerana*. In the west, Douglas-fir is the alternate host. It was introduced into Europe along with Douglas-fir. In Great Britain, it is named the Douglas-fir woolly aphid. Infestations are also found on Sitka spruce, which has also been introduced into Europe as a plantation species. This aphid has a complex life cycle. When both hosts are present, there can be six life stages in addition to eggs and crawlers. The life cycle, which includes all life stages, is completed in 2 years. Immature females overwinter under bark scales near the tips of spruce twigs. In the spring, they develop into mature females, termed stem-mothers. They each lay up to several hundred eggs under a white cottony, waxy mass. The newly hatched nymphs feed at the base of needles. Light green to purple elongate galls are formed that enclose the nymphs. A chamber at the base of the needle may contain 3 to 30 wingless adelgids, which are covered with a white wax. A few winged adelgids may also be produced. The galls vary between 12 and 75 mm in length. After the nymphs leave the gall, it turns brown and hardens, often remaining on the tree for many years. Infestations kill branch tips, which stunts and deforms the trees. Nymphs transform to adults on needles and fly to Douglas-fir and lay eggs on newly developing needles, shoots, and cones. Infested needles are twisted and chlorotic and can drop from trees in large numbers. No galls are formed on Douglas-fir. Winged adults produced on this host fly back to spruce to oviposit. Where each host occurs independent of the other, continuous generations are produced on one host. In areas of western North America where spruce is rare or absent, two parthenogenetic generations are produced on the needles of Douglas-fir in one season. No galls are produced on spruce where Douglas-fir is absent.

***Rhyacionia buoliana* (European Pine Shoot Moth); Lepidoptera: Olethreutidae (Olethreutid Moths)** This moth occurs throughout Europe where it feeds on almost all

pine species. It was first discovered in America in New York State in 1914 and later was found in British Columbia, Washington, and Oregon. It has not spread into other western North American states. This moth is a serious pest of pine plantations in Europe and North and South America. Recently, it has spread rapidly throughout the extensive Monterey pine plantations in Chile. Lodgepole pine planted in France are often heavily damaged, as are red pine, *Pinus resinosa,* plantations in the states and provinces that border the Great Lakes in North America. In summer, the female lays yellowish, disk-shaped eggs in rows on needles as well as under and on bud scales and on small twigs. The larvae hatch from the eggs and feed on the needles where they are attached to the twig. The larvae then spin a silken web between the needle sheath and the twig. The web is usually coated with resin. The larvae tunnel into the bud, which produces a resinous crust. These larvae overwinter under this crust or within the bud. In the spring, they mine another bud and then feed on the base of needles. They pupate within mined shoots, but the pupal skin protrudes from the shoot so that adult emergence is not impeded by the shoot or resin. The mined shoots are killed and buds formed later produce new shoots that grow around the dead shoot. Repeated infestations produce a bush-shaped tree. A crook is formed on the main stem when the terminal bud is killed.

Insects That Feed on Foliage

The foliage is the predominant photosynthetic part of the plant. Loss of foliage results in loss of photosynthetic area with resulting reduction in the production of carbohydrate. Defoliation may be tolerated in some species of trees, but less so in others. Successive cycles of defoliation, however, are rarely tolerated and often result in tree death, usually in association with other insect groups. Insects that feed on the foliage of trees are predominantly found in the orders Coleoptera, Diptera, Hemiptera, Lepidoptera, and Orthoptera. Most of these feed externally on the foliage, though some mine into the foliage for part or all of their feeding stage. Various larvae in the Coleoptera, Diptera, and Lepidoptera mine the leaves of broadleafed plants, including trees. This activity protects the insect from desiccation and from predators. Many Hymenoptera induce galls to form on leaves as a result of oviposition into the leaf. These galls can take on many shapes and all protect the developing larva from desiccation and predation. Additional protection from abiotic and biotic effects may be provided by webbing to hold needles together, leaf rolling, or producing a silken nest in which insects congregate for protection from predators, as in the case of tent caterpillars.

***Pyrrhalta luteola* (Elm Leaf Beetle); Coleoptera: Chrysomelidae (Leaf Beetles)** The elm leaf beetle is a native European chrysomelid beetle. It infests all species of elm. This beetle often causes severe growth loss, which weakens trees and, in turn, makes them susceptible to invasion by bark beetles and the Dutch elm disease fungus, which they carry into the tree.

The elm leaf beetle overwinters as an adult in sheltered, dry habitats, such as inside houses and barns and under loose bark on trees. They can become a nuisance in fall and spring as they enter and leave hibernation sites inside residences. They can be active in houses during the winter. Adults emerge from overwintering quarters and fly to elms at the time leaves are beginning to emerge from buds in spring. They lay eggs (up to 800 per female) in groups or irregular rows, along the major leaf veins on the underside of leaves and on nearby twigs. Larvae emerge from eggs in late spring and feed on the underside of leaves, causing skeletonization, where small veins are visible between the major lateral veins that emanate from the midrib. The later instars consume some of the area between these lateral veins. The leaves soon become desiccated and turn brown. At high densities, the entire tree turns brown and appears to be dead. However, in late summer some trees refoliate with smaller leaves. Repeated infestations can kill trees. Two or more generations are produced each season. This insect was introduced into the United States in the late 1890s and has since spread to the west coast. It feeds on all species of elm; however, American elm, *Ulmus americana*, and the introduced Siberian, *U. pumila*, and European elms are severely damaged by this beetle. Introduction of parasites into the United States has met with mixed control results.

***Elatobium abietinum* (Green Spruce Aphid); Homoptera: Aphididae (Aphids)** The green spruce aphid feeds on native and introduced spruce species in Europe and frequently occurs in forest nurseries. Introduced North American species such as Sitka spruce are most susceptible, whereas most European species, with the notable exception of Norway spruce, *Picea abies*, are less suitable as hosts. In parts of Europe a sexual as well as an asexual stage is found, and overwintering occurs as eggs. Elsewhere, such as in Great Britain, the sexual stage is absent, and parthenogenetically reproducing females remain on the needles during the winter. In North America, where this aphid is presumed to be an introduction, only the parthenogenetic form is known. Most feeding occurs when amino acid levels in the leaves are high in the winter and early spring. In late spring/early summer, longer day length coincides with a drop in amino acid levels, and winged individuals (alatae) are produced, which disperse to new hosts. Most outbreaks coincide with mild winters. The nymphs and adults feed on the lower side of older leaves where they use piercing mouthparts to access the phloem sap. They may completely defoliate trees and attack current-year needles later in the year.

***Pristophora erichsonii* (Larch Sawfly); Hymenoptera: Tenthredinidae (Sawflies)** The larch sawfly is a native European species. It was first recorded in North America in 1880, in Alberta, Canada, in 1930, in Oregon in 1964, and in Alaska in 1965. Some strains are now believed to be native to North America. The prepupal stage overwinters in a

cocoon in the ground or duff under defoliated trees. Adults emerge in spring and as late as late summer depending on location. They lay eggs in rows under the bark of newly elongating shoots. This oviposition behavior causes the shoots to droop. The common name of this family is derived from the saw-like ovipositor that females use to cut slits into the bark where they lay eggs. This species reproduces parthenogenetically. Larvae emerge from eggs and feed on the needle margins and then move to the older needle clusters. They feed gregariously and eat most of the needles on one shoot before moving to another. Between summer and fall, mature larvae drop to the ground and spin a paper-like cocoon in the forest litter. Here, they enter a diapause where they spend the winter. However, some individuals overwinter 2 or 3 years.

Successive defoliation results in tree mortality, especially eastern, *Larix laricina,* and western larch, *L. occidentalis,* in North America, and European larch *L. decidua.* It has been most destructive in the states around the Great Lakes. Where black spruce, *Picea mariana,* and eastern larch co-occur, black spruce is favored in succession because the larch sawfly kills both larch seedlings and mature trees. In the Alps, when aphid densities are high, larch sawfly densities are low, due in part to predation by ants that are tending aphids seeking their honeydew. Biological control efforts in the western United States have been successful in lowering the overall average densities of the larch sawfly. This program is considered one of the most successful biological control programs aimed at a widely distributed forest insect. During studies of the parasitoids of this sawfly, encysted larvae were discovered in certain populations. A cellular layer surrounding the larva is produced by the hemolymph (phagocytosis), a rare phenomenon in insects.

***Lymantria dispar* (Gypsy Moth); Lepidoptera: Lymantriidae (Tussock Moths)** The gypsy moth is an important defoliator of hardwood trees in Europe and North America. It was introduced into Massachusetts in 1869. Since then it has slowly spread southward to North Carolina and westward into Wisconsin. Small populations have been eradicated repeatedly in California, Oregon, Washington, and British Columbia. In North America, massive defoliation has occurred over thousands of hectares. This species feeds on over 50 tree species, including mostly hardwoods, such as apple, beech, basswood, *Tilia americana*, elm, hornbeam, oak, poplar, and willow. During outbreaks, larch and pine are consumed. In southern Europe and northern Africa, outbreaks occur at irregular intervals. In North America, considerable resources are directed toward control and slowing the spread to uninfested areas. The Asian form of this species has been found in the western United States and Canada, but apparently has not become established.

In spring, larvae emerge from overwintering egg masses, at about the time oak leaves emerge from buds. The young caterpillars move to the tops of trees where they feed on the bases of young leaves or they chew small holes in the leaf surface. They also drop from trees on silken threads and are windblown. Older larvae feed mostly on the leaf margins. The larger veins and midribs of leaves are usually not consumed. At high densities (tens of thousands per hectare), the larvae literally "eat themselves out of house and home." They move to new areas in search of food or pupate, often giving rise to small adults. The larvae are very colorful, exhibiting six longitudinal rows of tubercles. On the thorax and first two abdominal segments the tubercles are blue and on the rest of the abdomen they are red. Larvae crawl down the tree and aggregate in sheltered places. At night they crawl up the tree to feed. In June, larvae pupate on many substrates, including rocks, limbs and trunks of trees, picnic tables, automobiles, and forest debris. In North America, females are flightless, and after emergence from the pupae they crawl a short distance and emit a sex pheromone that attracts male moths. After mating, the females lay eggs in a mass and cover them with larval hairs and a frothy substance. These egg masses may be transported on vehicles or on lawn furniture and rocks that are moved inside vehicles. This means of dispersal gave rise to the common name, "gypsy" moth. The eggs overwinter in a diapause state. The Asian gypsy moth is of major concern because the females are excellent fliers. It has frequently been intercepted on the west coast of the United States, and flying females enhance the dispersal of this species if it becomes established. Many natural enemies of this moth, including parasitoids, predators, and protozoa, have been collected in Europe and introduced in North America. These biological control studies are among the earliest attempts to use indigenous natural enemies to lower densities of an introduced pest.

***Coleotechnites milleri* (Lodgepole Needleminer); Lepidoptera: Gelechiidae (Gelechiid Moths)** The lodgepole needleminer has a unique life history in the Sierra Nevada of California. The adults are synchronized to emerge in midsummer in odd-numbered years. Adults are not known from even-numbered years. However, another population of this species (or an undescribed species) that occurs in a nearby mountain range in western Nevada also has a 2-year life cycle, but adults emerge in even-numbered years. Another undescribed species feeds on lodgepole pine in central Oregon, but this species has a 1-year life cycle.

After mating, females lay eggs in mined needles or on branches near the needle fascicles. First instars enter near the tip of a single needle and remain in the mine through the winter. During the even-numbered season, fourth instars mine several needles and overwinter in a single, mined needle. In spring of the odd-numbered years, fifth instars feed on many needles and finally pupate in a mined needle.

The most severe outbreaks occur in mature lodgepole pine stands and occur over large areas. In the high-elevation forests (4000 m) near Yosemite National Park, many outbreaks have been recorded, some lasting 16 to 18 years, before returning to low densities. There does not appear to be a regular pattern to these epidemics. Severe defoliation

weakens trees, which predisposes them to infestation by the mountain pine beetle, *Dendroctonus ponderosae.* Thus these two insect species convert these extensive mature to overmature lodgepole pine forests to a young forest of lodgepole pine. The Sierra Nevada form of lodgepole pine, *Pinus contorta* var. *murrayana,* produces largely nonserotinous cones that are opened by sunlight. Furthermore, lodgepole pine is moderately shade tolerant and can thus regenerate in the understory of larger trees. When needle miner populations increase to epidemic levels, young trees are released to grow and seeds germinate to produce new seedlings.

The decline of outbreaks has been attributed to rain, and often snow, during the mating period, abundance of parasites and predators, and reduction in amount of host foliage for colonization. The biology of host-specific and generalist parasites has been investigated in an attempt to determine the basis for the 2-year synchronized life cycle.

***Diapheromera femorata* (Walkingstick); Phasmatodea: Heteronemiidae (Walkingsticks)** Walkingsticks are very unusual insects because, as their name signifies, they are long, slender, and oval and thus resemble a stick or defoliated branch. Most species are wingless. Their body color varies from gray to green to brown to red, which is similar to the leaves and branches on which they are feeding. Through this camouflage they may escape predation from birds. In late summer and fall they deposit their hard, seed-like eggs directly on the ground or by dropping them from trees where they are feeding. At high population levels, the large number of eggs falling on the forest floor literally sounds like rain. Most young hatch during the following spring throughout the southern United States, whereas in the northern portion of its range, the young hatch in the second spring. Occasionally, severe outbreaks occur in the northern United States that last several years, causing extensive tree mortality. Two defoliations can occur during the same season. They feed on a variety of hardwood species, including cherry, basswood, birch, aspen, elm, hickory, oak, locusts, and dogwood.

Insects That Feed on the Trunk and Large Branches

Insects that feed on the trunk and larger branches of trees are found in the Coleoptera, Hemiptera, Hymenoptera, Isoptera, and Lepidoptera. Insects that tunnel into the trunk and large branches of trees may exploit two resources: they may derive most of their nutrition from the phloem, or they may tunnel into the wood and derive their nutrition from the xylem. Insects that do not tunnel into the host, such as scale insects, are sucking insects and use nutrients in the xylem or phloem depending on the species. Many insects that infest the wood of trees, such as termites, serve as primary decomposers of the woody material.

***D. ponderosae* (Mountain Pine Beetle); Coleoptera: Scolytidae (Bark and Ambrosia Beetles)** The mountain pine beetle is one of the most important forest insects in North America. It is widely distributed in many pine species throughout western North America, from southeastern Alaska to northern Baja California, and eastward through the Yukon territory in Canada and the Rocky Mountains of both the United States and Canada to the Black Hills of South Dakota. The eight-toothed spruce bark beetle, *Ips typographus,* shows a similarly wide distribution in Europe and Asia but it has a more limited host range than *D. ponderosae.* During outbreaks, this bark beetle kills millions of trees over large areas. It infests some of the most widely distributed and important timber-producing species, including ponderosa pine, western white pine, *P. monticola,* and sugar pine. Extensive mortality of ponderosa pine at the turn of the 19th century in the Black Hills of South Dakota attracted public attention that led to the establishment of the Federal Forest Insect Research Program in the United States. The effects of mountain pine beetle are greatest in climax lodgepole pine forests where lodgepole pine is self-perpetuating or in even-aged stands where shade-tolerant species are not abundant enough to replace lodgepole pine. These even-aged forests are usually created by a stand-replacement fire. Mountain pine beetle infestations kill the largest trees and leave behind small-diameter, low-vigor, and mistletoe-infested trees. Young seedlings that have survived in the understory because of their shade tolerance are then released. New seedlings appear in new openings where trees have died. The resultant forest following mountain pine beetle outbreak is uneven aged and multistoried.

The life histories of tree-killing *Dendroctonus* (meaning "tree killers") species are generally similar. The female emerges from overwintering sites beneath the bark of trees killed in the previous year. She penetrates the outer bark and begins feeding in the phloem. She releases a pheromone component, *trans*-verbenol, and a host monoterpene hydrocarbon, myrcene (in lodgepole pine) or α-pinene (in western white pine). The male joins the female gallery and releases *exo*-brevicomin, another pheromone component. This mixture of compounds is highly attractive to other mountain pine beetles and a massive aggregation of beetles occurs on the tree. The attracted beetles introduce a pathogenic bluestain fungus, *Ophiostoma clavigerum,* which, together with tunneling females, causes the death of the tree. The growth of this bluestain fungus in the sapwood interrupts water conduction (a vascular wilt) to the crown. As the tree begins to die, females excavate egg galleries that are positioned vertically in the trunk, 30 to 90 cm in length. The egg gallery etches the sapwood superficially and may be quite sinuous, as with the western pine beetle, *D. brevicomis.* Eggs are laid in niches cut in the phloem on both sides of the gallery. Larvae hatch from the eggs and they excavate lateral galleries that increase in size as the larvae grow. A pupal cell is excavated at the end of the gallery, and the larvae transform to pupae and then to adults in these cells. Both larvae and adults are known to overwinter. Depending on the latitude and altitude one generation may take 2 years to complete or there may be as many as two generations and a partial third in 1 year.

***Cryptococcus fagisuga* (Beech Scale); Hemiptera: Eriococcidae (Eriococcids)** The beech scale is a native European species that was introduced into Halifax, Nova Scotia, about 1890. Since then it has slowly spread eastward to Toronto, Ontario, and Ohio and south to West Virginia and western Virginia. Infestations are found in North Carolina, Tennessee, and Michigan. The beech scale in combination with the fungus *Nectria coccinea* var. *faginata* causes beech bark disease, which has killed a large number of native and ornamental beeches in Europe and the northeastern United States. European beech, *Fagus sylvatica,* and American beech, *F. grandifolia,* are severely impacted by this scale and its associated fungus, and it also infests other native as well as ornamental beeches from the Orient and Europe.

In late spring and early summer, pale-yellow females deposit up to 50 yellow-colored eggs in groups of 5 to 8 that are coated with a white-colored wax-like substance. Only wingless females are known for this species. The newly hatched nymphs or "crawlers" search out a location on the bark to settle and insert their tubular mouthparts. During this crawler stage they can be dispersed by wind, often over 100 meters. This sessile stage overwinters. After egg-laying the female dies. There is one generation produced each year. At high densities, the bark on the trunk and lower portion of the branches is completely white with scales. Under these conditions, the bark is killed and forms pits and ultimately ruptures, allowing entry of *N. coccinea* var. *faginata.* This fungus, along with the native species, *N. galligena* and *N. ochroleuca,* kills the cambium and sapwood in these areas, which results in interruption of water and food transport. The trunk often turns red with the fruit bodies produced by this fungus. Tree mortality often ensues after a few years. Extended periods of drought increase the rate of mortality caused by the interaction of this scale and fungus with their host. Cold temperatures (–38°C) often kill the scales above the snow pack. In Europe, the ambrosia beetle, *Trypodendron domesticum,* attacks beeches weakened by this scale and its associated fungus. Several decay-causing fungi are also associated with dying trees.

***Sirex noctilio* (Steely-Blue Wood Wasp); Hymenoptera: Siricidae (Horntails/Wood Wasps)** *S. noctilio* is indigenous to Europe where it infests pines weakened by fire, insects, and diseases. In its native habitat it seldom causes tree mortality. However, this species was introduced into New Zealand in the late 19th century, where it became a serious agent of mortality in planted forests of Monterey pine. It has since been introduced into Monterey pine plantations in Australia and South Africa. Females have a long ovipositor that extends straight back from the anus and is often mistaken for a stinger. The female drills a hole about 12 mm deep through the bark and into the sapwood. As many as four side holes are drilled from the primary entrance hole. One hole is filled with mucus from the female's accessory glands, as well as spores of the symbiotic fungus, *Amylosterium areolatus,* that reside in a specialized gland (mycangium) at the base of the ovipositor. Eggs are deposited in the other side tunnels. Adults tend to aggregate on weakened trees for oviposition. The mucus secretion has toxic properties that cause interruption of water conduction in the sapwood (i.e., a vascular wilt). Needles are subsequently killed and they fall from the tree. Young larvae feed on the fungus growing in the egg chamber. Older larvae tunnel deeply into the sapwood and their finely divided frass (mostly woodchips) is tightly packed into their galleries. They pupate near the surface and excavate a round hole in the bark through which they emerge.

FIGURE 3 Female horntail, *Urocerus gigas* (Hymenoptera: Siricidae), ovipositing in the end grain of a European larch log. [Photograph by Andrew J. Storer.]

Horntails occur in both hardwood and softwood species and they are found in forests throughout the northern hemisphere (Fig. 3). The life cycle can be completed in 1 year, but in some cases development is delayed for several years.

A major biological control program for *Sirex noctilio* was undertaken and many species of hymenopteran parasitoids were collected from Europe and North America and introduced into New Zealand and Australia. Two species of nematodes, *Deladenus siricidicola* and *D. wilsoni,* were introduced into Australia where they have caused a significant reduction in tree mortality. These nematodes infect the reproductive system and prevent ovarian development.

***Reticulitermes flavipes* (Eastern Subterranean Termite); Isoptera: Rhinotermitidae** Subterranean termites are social insects that live in colonies in the soil. They feed on a variety of cellulosic materials as their principal energy source. Termites and fungi are the most important organisms that can digest plant cell walls and thus reduce cellulose from a complex polysaccharide to simple sugars. Symbiotic fungi and protozoa live in the hindguts of termites and possess the enzymes needed to digest cellulose. Thus termites play an important role in recycling nutrients in a forest. Subterranean termites coexist in soils with many other microorganisms, especially fungi. Some of these fungi have been shown to produce chemicals that are the same as the trail-following pheromones produced by the termites, as well as chemicals that act as feeding stimulants and deterrents to termites.

The eastern subterranean termite is distributed throughout the eastern and mid-western part of the United States where it feeds on most species of wood found in

forested habitats. Related species of subterranean termites are found in western North America and in Europe and Asia. Subterranean termite colonies contain seven castes: larvae or immatures, workers, soldiers, nymphs, winged (alate) primary reproductives, wingless (dealate) primary reproductives, and supplementary reproductives. Workers feed all the castes. Supplementary reproductives can occur in colonies with primary reproductives. They take over the oviposition role if the primary "queen" reproductive dies. Soldiers have large hardened (sclerotized) heads and mandibles, which they use to defend the colony from invaders, especially ants.

With the onset of warm rains in spring or fall (western United States), large numbers of winged females and males (termed swarmers) emerge from earthen tubes extending out of the ground or from the surface of logs, stumps, and wooden structures. These adults are weak fliers and soon drop to the ground where they lose their wings. Copulating adults paired end-to-end can be seen running about. The mated female seeks a damp piece of wood and excavates a chamber where she lays her first eggs. The young hatch into a worker caste and when large enough, they begin foraging for food to feed the queen and other castes as they are produced.

Subterranean termites are a critical component of forest habitats because of their important role in recycling wood. However, they are also the most destructive pests of human habitations made from wood. They construct earthen tubes from the ground into wooden structures and thus can consume wood in dry habitats long distances from the colony in the ground. They often escape notice because they travel in hollowed out timbers and in tubes constructed in wall cavities. Their galleries parallel the grain of the timbers and are excavated first in the early wood (springwood) and later in the late wood (summerwood). This damage to structures can be extensive before it is discovered. A large pest control industry has developed to protect wooden structures from termites and other wood-destroying insects and fungi.

***Synanthedon sequoiae* (Sequoia Pitch Moth); Lepidoptera: Sesiidae (Clearwing Moths)** The sequoia pitch moth is attracted to the resin that flows from wounded pine trees. Larvae hatch from eggs deposited individually on the bark of limbs and the trunk and tunnel into the phloem–cambial layer beneath the outer bark. As they feed on the phloem, they excavate a chamber that fills with resin and frass. This material is deposited on the outer bark surface over the chamber and forms a white resinous mass that is interspersed with reddish brown fecal material. As this resin mass ages, the surface turns reddish brown. Resin streaming is occasionally observed from these masses. Larvae pupate beneath the resinous mass, but with the pupal integument extending through the mass. This enables the moth to emerge from the mass without contacting the resin. These resinous masses often exhibit two or three pupal cases, indicating that these areas are often recolonized by later generations of this species. Oviposition occurs in summer. The life cycle is completed in 1 or 2 years.

Infestations of this species rarely cause tree mortality directly. However, heavy infestations may result in attraction of tree-killing bark beetles. This moth infests many native species of pines as well as European species such as Scots pine, Austrian pine, *Pinus nigra,* maritime pine, *P. pinaster,* and Italian stone pine, *P. pinea.* Pines growing in urban environments can be heavily infested around wounds caused by pruning and by vehicles. The sequoia pitch moth is distributed throughout western North America, including California, Oregon, Washington, Idaho, Montana, and British Columbia. Coast redwood, *Sequoia sempervirens,* is not a known host for this moth. Similar pitch masses are produced by a related moth, *S. pini,* which infests many pine and spruce species in eastern North America and in the midwestern states.

Insects That Feed on Roots

Insects that feed on the roots of trees may be sap feeders or may enter the host tissue and exploit the phloem, cambium and/or xylem. Representatives are found in the Coleoptera, Hemiptera, and Hymenoptera. The underground portion of the tree represents a protected environment for root-feeding insects.

***H. abietis* (Pine Weevil); Coleoptera: Curculionidae (Weevils)** The pine weevil is native to the forests of northern Europe and Asia where it colonizes roots of weakened and recently dead conifers. Larvae emerge from eggs laid in the bark and tunnel through the phloem and score the sapwood surface. They pupate in cells excavated in the outer sapwood. Its abundance has been increased dramatically by thinning and clear-cutting of large forested areas and replanting with pines. Under these conditions, the weevils breed in the stumps and emerge in the next season to feed on seedlings planted after logging. This species has the reputation of being one of only a few species of which one individual can kill one or more healthy trees, although they are small trees. Grasshoppers (Acrididae) and scarabs (Scarabaeidae) also have this capability. Adults feed on the tender young bark of most coniferous species. They are present from early spring through early fall. Control of this species is necessary in new conifer plantations. One generation is produced in 1 to 3 years, depending on latitude. The pales weevil, *H. pales,* has a comparable biology in the eastern United States. These insects are considered to be among the most destructive pests of conifer plantations in both Europe and North America.

***Pachypappa termulae* (Spruce Root Aphid); Homoptera: Pemphigidae (Pemphigids)** The spruce root aphid is cream colored and feeds in colonies on the fine roots of standing trees. These colonies are covered with waxy wool. Spruce trees are most commonly infested, though other conifers may also be affected. As with all aphids, this is a sap-feeding species that has sucking mouthparts. Aboveground symptoms are usually absent except in nursery stock, where an overall decline of plant health may be observed. Often

these insects are noticed only during transplantation of nursery stock. This species can be found year round in parts of Europe and North America and is thought to be associated with dry conditions.

***Camponotus modoc* (Carpenter ant); Hymenoptera: Formicidae (Ants)** *C. modoc* excavates galleries in fire-scarred and rotted standing and fallen trees of many coniferous species growing in western North America, including pines, true firs, Douglas-fir, western redcedar, *Thuja plicata*, and giant sequoia. In a study of uprooting and breakage of over mature giant sequoia in the Sierra Nevada of California, carpenter ants and decay fungi were often associated with tree failure. Most of the roots of recently uprooted trees showed evidence of advanced decay. Carpenter ant galleries were observed in the few functional roots that were present. These roots showed evidence of early-to-moderate stage decay and carpenter ants were observed tunneling in these areas. This excavation activity is probably associated with the establishment of subsidiary colonies. These ants are found tending aphids, e.g., *Cinara* spp., in colonies on understory white fir and sugar pine trees. *C. modoc* was also observed excavating cavities in the bark of the root collar of young white firs. These cavities were later colonized by *Cinara* spp. that were tended by these ants. These shade-tolerant conifers have greatly increased in abundance because of fire exclusion in the old-growth giant sequoia groves. Thus carpenter ant abundance may have also increased as a result of increased aphid colonies on these understory trees.

Many insect families are found almost exclusively in forest habitats and have not been mentioned above because of space limitations. These include, for example, the wood-boring beetles in the families Buprestidae, Cerambycidae, Platypodidae, Micromalthidae, Anobiidae, Bostrichidae, and Lyctidae. In the Lepidoptera, the family Cossidae is almost entirely found infesting wood. Similarly, a very large number of species of moths, wasps, aphids, and leafhoppers feed on the foliage of trees.

FOREST PEST MANAGEMENT

Insect populations may reach levels in both natural and human-made forests where they are considered pests. High populations of cone and seed insects in a conifer seed orchard, mortality of seedlings resulting from feeding by weevils, defoliation of mature trees, and bark beetle outbreaks are all examples for which a management activity aimed at reducing the density of the insect population may be appropriate. These management activities all have a financial cost associated with them, which needs to be weighed against the anticipated benefit from the activity. With a few notable exceptions, the use of traditional insecticides to control forest insects is rarely practical, because of cost, or desirable because of impacts on nontarget organisms. Exceptions may be high-value urban trees and trees in seed orchards. In these situations, the area being targeted with the pesticide is limited. Biological pesticides are, however, very useful in limiting the impacts of defoliating Lepidoptera. Extracts from the bacterium *Bacillus thuringiensis* (Bt) are used in aerial applications where defoliators threaten to kill large tracts of trees. Extensive areas in North America are treated with Bt to reduce the damage from the gypsy moth and to slow its spread. Insect outbreaks may be a result of fire suppression that creates an overmature forest that is susceptible to these insects. Long-term management strategies should therefore address this issue rather than use short-term solutions such as insecticides.

In populations of native and introduced forest insects, especially defoliators, it has been possible to identify viruses that are specific to their host insect. Rearing and release of these viruses offer promise for pest control in some situations, especially for defoliating Lepidoptera and Hymenoptera. The nuclear polyhedrosis virus (NPV) of the European pine sawfly, *Neodiprion sertifer,* has been applied to thousands of hectares of forests in Europe over the past 35 years. An NPV that infects gypsy moth has also been used in efforts to slow the spread of this species in North America. The accidental introduction of an NPV in the 1930s, along with a parasite of the European spruce sawfly, *Diprion hercyniae,* into eastern Canada has reduced this species to very low densities.

Biological control has achieved considerable success in managing forest insect populations. It is especially useful when addressing problems caused by introduced insects. For example, as the European spruce beetle spread westward across Europe during the 20th century, outbreaks were most severe on the edge of its range, in part because of the lack of natural enemies in newly invaded areas. Mass releases of the predatory beetle *Rhizophagus grandis* had some success in limiting the impact of this species in many European countries. Similarly, in urban forests where many tree species are introduced, management of exotic pests is often achieved by introduction of predatory or parasitic insects. In California, eucalyptus are widely planted in urban landscapes, and in the latter part of the 20th century, many new pests were introduced. Prompt identification and monitoring of pest populations coupled with introduction of appropriate biological control agents have successfully reduced the adverse impacts of many of these introduced insects, especially the eucalyptus longhorned beetle, *Phorocantha semipunctata.*

Much scientific research has focused on the pheromone systems of forest insects, especially defoliators and bark beetles. Pheromones have been incorporated into management plans aimed at reducing the impacts of the target pest species. For example, mountain pine beetle pheromones are used to bait living lodgepole pines that are to be logged following infestation. This treatment attracts beetles into plots that will be harvested, and the beetles are removed with the logs. The treatment is aimed at reducing mortality in the unharvested areas. It prevents trees from dying at more

scattered locations in these unharvested areas where most trees would not be harvested because of cost and accessibility. Other techniques for reducing the impacts of target pest species can be achieved in a number of ways. Pheromones can be used to monitor insect populations as a guide for when to apply pesticides or embark on an alternative management strategy. They are also used to detect invasive pest species in new areas. Throughout the western states of North America, pheromone traps are placed to detect gypsy moth so that early detection can be met with rapid response eradication programs. Similar use of pheromone traps and traps baited with other behaviorally active compounds occur at ports to detect the arrival of exotic organisms.

Attempts have been made to use pheromones to mass trap insects and thereby reduce their density to nonpestiferous levels. Because insect populations naturally decline at some point following an outbreak due to a range of factors, it is often difficult to determine whether such mass trapping efforts were in part responsible for a return of the insect population to endemic levels. In Scandinavia, mass trapping of the eight-toothed spruce bark beetle occurred during 1979 and 1980, and billions of insects were caught. The subsequent drop in population size of this insect may have been in part a result of this intensive mass trapping effort. Mass trapping may be useful locally where a small-scale disturbance may have resulted in elevated activity of a particular insect. This situation may occur when management activities have produced a large amount of host material. Pheromone traps may be deployed in this situation to reduce populations. Trap logs that have become infested with bark beetles are removed from the forest in an attempt to lower population levels. Pheromones may be used to bait the logs or trees to enhance the effectiveness of this technique.

Pheromones have also been used in larger doses to confuse or disrupt their natural function. Male moths following pheromone plumes to locate females for mating are not able to find the female if there is a large concentration of the pheromone present, such that the plume is effectively hidden. Western pine shoot borer, *Eucosma sonomana,* pheromone release of 10 to 20 g per hectare has been shown to reduce damage by over 65% in ponderosa pine plantations in the western United States.

Many forest management practices are available for use to reduce the impacts of insect populations. Removal of susceptible trees, thinning of the stand to increase tree vigor, prescribed burning to eliminate susceptible host material, and enhancement of natural enemy populations are all tools available for the management of forest insect populations. The thinning of lodgepole pine stands reduces tree mortality from mountain pine beetle. Prescribed fire following harvesting operations in ponderosa pine forests reduces populations of the California five-spined ips, *Ips paraconfusus,* that breeds in logging debris. This practice lowers the probability of these populations emerging and killing living trees in the area.

THE FUTURE

Insects perform essential roles in forests as part of these complex ecosystems. However, when transplanted from one forest habitat to another, or when the natural forest habitat is disturbed, their role in the ecology of the forest may change. For example, the eastern five-spined ips, *I. grandicollis,* generally infests stressed or diseased trees in its native habitat in the southeastern United States and is not considered a primary agent of tree mortality. Its introduction into Australia in the mid-20th century has resulted in significant tree mortality. This beetle infests many pine species throughout the eastern United States but is now infesting a Californian pine, Monterey pine, in Australia. Efforts have been made to find biological control agents in North America that are effective in Australia. Increased global transport of forest products undoubtedly will result in introductions of both pestiferous and potentially pestiferous insects into different parts of the world. Ongoing efforts to reduce this occurrence are essential if we are to maintain forest habitats that meet environmental and economic goals. Furthermore, research that addresses environmentally sound methods for the integrated management of forest pests is essential if humans are to minimize their indirect impacts on forest habitats resulting from the introduction of new insect species. Some insect outbreaks are important natural processes in forests, whereas others are the product of human activities. Recognizing the causes of outbreaks, detecting signs of imminent outbreaks, and having appropriate management options available help to ensure the viability of forest habitats and the environmental and economic benefits they provide.

See Also the Following Articles

Biological Control • Gypsy Moth • Phytophagous Insects • Soil Habitats

Further Reading

Bevan, D. (1987). "Forest Insects: A Guide to Insects Feeding on Trees in Britain." Forestry Commission, London. [Handbook 1]

Coulson, R. N., and Witter, J. A. (1984). "Forest Entomology, Ecology and Management." Wiley–Interscience, New York.

Dajaz, R. (2000). "Insects and Forests." Lavoisier, Paris.

Drooz, A. T. (ed.) (1985). "Insects of Eastern Forests." U.S. Department of Agriculture, Forest Service, Miscellaneous Publication No. 1426.

Furniss, R. L., and Carolin, V. M. (1977). "Western Forest Insects." U.S. Department of Agriculture, Forest Service, Miscellaneous Publication No. 1339.

Johnson, W. T., and Lyon, H. H. (1988). "Insects That Feed on Trees and Shrubs," 2nd edition. Cornell University Press, Ithaca, NY.

McCullough, D. G., Werner, R. A., and Neumann, D. (1998). Fire and insects in northern and boreal forest ecosystems of North America. *Annu. Rev. Entomol.* **43,** 107–127.

Paine, T. D., Raffa, K. F., and Harrington, T. C. (1997) Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annu. Rev. Entomol.* **42,** 179–206.

Speight, M. R., and Wainhouse, D. (1989). "Ecology and Management of Forest Insects." Clarendon Press, Oxford.

Fossil Record

David Grimaldi
American Museum of Natural History, New York

Insects are the most evolutionarily successful group of organisms in the 4-billion-year history of life on earth, with perhaps 5 million species alive today and untold millions of extinct species. Although fossils of insects are not as abundant as has been found for some other types of organisms, the insect fossil record extends back for 400 million years, making them among the oldest terrestrial animals known, and the fossils contribute unique insight into the evolutionary history of insects. Particularly significant periods in the evolution of insects are the Paleozoic, Triassic, and Cretaceous. Key features that gave rise to their spectacular success, notably flight and complete metamorphosis, originated at least 300 and 250 million years ago (mya), respectively.

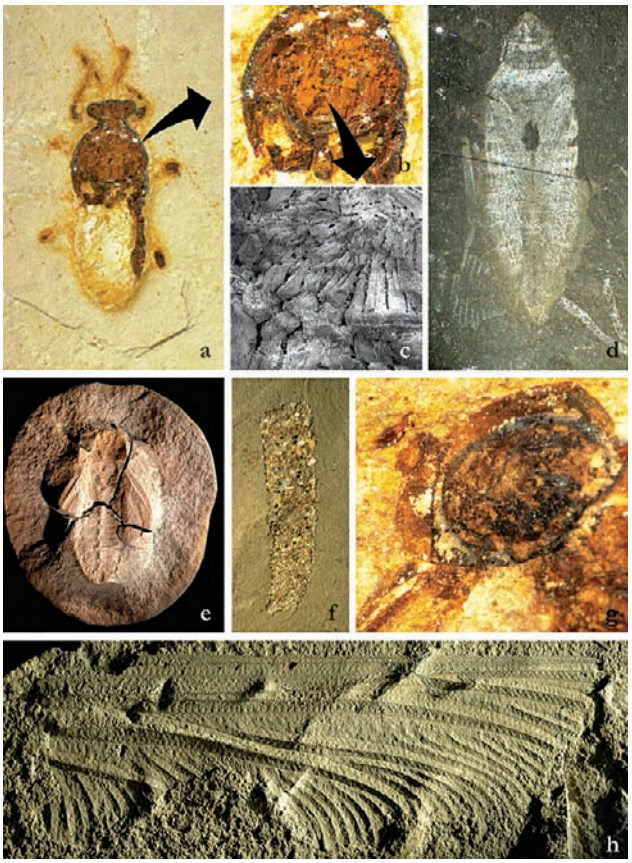

FIGURE 1 Various kinds of fossil insects, modes of fossilization, and degree of preservation. Different scales. (a–c) Iron hydroxide concretion of a heteropteran in Cretaceous limestone from Brazil, showing preservation of thoracic muscles [b, light micrograph; c, scanning electron micrograph (SEM)]. (d) Silvery carbon film of a belostomatid on fine-grained, Triassic shale (Virginia). (e) Nymph of †*Herdina* (in this article, † signifies an extinct group) in ironstone concretion from the Carboniferous of Mazon Creek, Illinois. (f) Trichopteran case of sand pebbles in volcanic shale from Florissant, Colorado (late Eocene/early Oligocene). (g) Part of the head of tabanid fly from Florissant, showing the eye facets. (h) Portion of the wing of †*Typus* ("Protodonata"), from the Permian of Elmo, Kansas, with the wing fluting preserved.

FOSSILS

Fossils are generally, but not necessarily, extinct species whose remains have been preserved for thousands to millions of years. The remains are most commonly mineralized replacements of original tissues (Figs. 1 and 2); in rare situations portions of the original organism are preserved (Figs. 2b–2e, 2g, and 5d–5f). Remains of apparently existing (or extant) species that are thousands to several million years old are sometimes called *subfossils.* Earth's fossils are dominated by organisms from marine continental shelves, where deep sediments efficiently preserved durable calcified parts such as shells and skeletons. Terrestrial life is less well known in the fossil record and largely is preserved as bones, leaves, and pollen in freshwater sediments. Well-preserved insects, by comparison, are rare; their occurrence depends on conditions under which

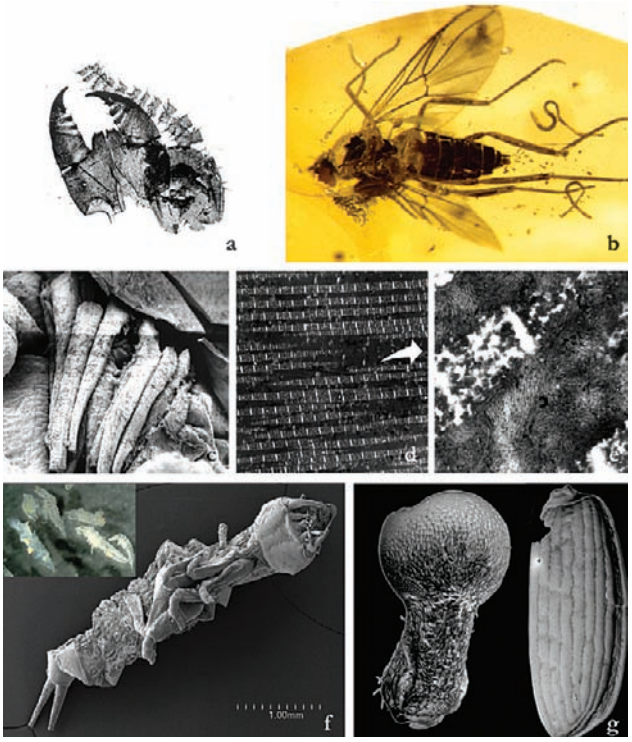

FIGURE 2 Exceptional preservation of fossil terrestrial arthropods. (a) Centipede from Upper Devonian (New York). (Courtesy of William Shear, Hampton–Sydney College.) (b) *Mycetobia* woodgnat in Miocene amber (Dominican Republic), with parasitic nematodes bursting from abdomen. (c–e) Flight muscles of meliponine bee in Dominican amber, showing ultrastructural preservation of myofibrils (d) and even the fingerprint-like mitochondria (e). (f) Silicified replicas of early instar dytiscoid beetle, from Miocene of California (inset, photomicrograph; SEM is larger). (g) Cuticular remains of extant beetle species from the Wisconsin stage (ca. 10,000–80,000 years ago) of Alaska (left, weevil head; right, carabid elytron). (Courtesy of Scott Elias, University of Colorado.)

the sediments were fine grained, anoxic (i.e., lacking oxygen), and deposited rapidly but without significant disturbance. Even amber is usually deposited in lacustrine or swampy sediments; it would otherwise completely decompose from oxidation and other processes. Insects preserved in such sediments were aquatic or semiaquatic and died *in situ* (autochthonous), or their bodies were transported via winds or water from surrounding habitats (allochthonous).

SIGNIFICANCE OF FOSSILS

There are at least five reasons why fossils are uniquely significant for understanding the evolutionary history of organisms:

1. Fossils provide the only direct record of extinct lineages, such as giant dragonfly-like forms from the Carboniferous and Permian (e.g., †Meganeuridae) (in this article, † signifies an extinct group).

2. Fossils reveal patterns and timing of extinctions and radiations. The mass extinction at the end of the Permian, for example, was the most cataclysmic event in the history of life and may have caused the extirpation of the †Paleodictyopteroidea (known almost exclusively from the Carboniferous and Permian). The extinctions at the Cretaceous/Tertiary boundary that extinguished the remaining dinosaurs, ammonites, and other groups, appear to have had little impact on families of insects. Although insects have been affected by some mass extinction events, major lineages of insects appear particularly resistant to extinction.

3. Fossils provide the only direct information on the ages of lineages. Because there is never assurance that a fossil is the earliest, original occurrence of a taxon, the age of the earliest fossil is the minimum age of a taxon. This information, in conjunction with the phylogenetic positions of all fossils in a group, can be used to estimate actual ages and significant gaps in the fossil record. Such information is further useful for calibrating and then estimating rates of change, such as rates of genetic change among living species and dates of divergence.

4. Fossils may assist in reconstructing the phylogeny of a group. While fossils are widely acknowledged to possess combinations of characters unique from those of living species, their significance in reconstructing phylogenetic relationships is controversial. Nonetheless, fossils also provide the only direct evidence for the evolutionary sequence of character change. For example, in the fossil record of the Blattaria (cockroaches and their primitive roach-like relatives), tegminous forewings and a large, discoid pronotum appear well before the loss of an external ovipositor—a sequence not revealed by the study of living species alone.

5. Fossils can provide evidence that a taxon is old enough for its distribution to have been affected, for example, by Cretaceous continental drift. Indeed, many families of insects extend to the Cretaceous or even earlier, whereas some large groups are apparently too young to have been affected by continental drift, such as the ditrysoan Lepidoptera and the schizophoran Diptera. Often, too, a fossil is found outside the present-day range of its group, indicating formerly widespead distribution. A famous example is the occurrence of Glossinidae (tsetse flies) in the Cenozoic of North America and Europe.

PRESERVATION OF FOSSIL INSECTS

The small size and external cuticle of insects are largely responsible for the many modes of fossilization, which are much more varied than for vertebrates and plants. Insect fossils are most commonly encountered as impressions or compressions in sediments (Figs. 1d, 1g, 1h, 5a and 5b), generally as disarticulated cuticle and particularly as wings because these are especially resistant to decay. Because wing venation has many systematically significant features, isolated wings often can be identified at least to family level. Generally, remains in sediments are highly compressed, but can still reveal microscopic structures such as flagellomeres, tarsomeres, microtrichia, wing scales, and even color patterns. Some are preserved as concretions, which are three-dimensional permineralized replicas of the original animal (Figs. 1a–c, 1e and 2f).

The finest preservation of insects is in amber (Figs. 2b–2e and 5d–5f). These formed when the resin was originally viscous and sticky, and small organisms became mired and then engulfed by the flows; they were embalmed so thoroughly as to preserve parasites, soft internal organs and tissues, and even organelles of cells (Figs. 2b–2e). The putatively most ancient DNA in the geological record is reported from insects preserved in amber, but authenticity of the DNA is disputed by those who unsuccessfully attempted to replicate these results.

Exceptional preservation is also seen in some insects preserved free in sediments. Terrestrial arthropods in several Devonian deposits of eastern North America are preserved as original cuticle, with even microscopic sensilla and setae preserved (Fig. 2a). In several Miocene deposits from California, insects are preserved in nodules as perfect three-dimensional silicified replicas (Fig. 2f). Similar relief and microscopic fidelity are found in carbonized remains in Cretaceous clays, rendered by ancient forest fires that charcoalified small organisms buried in leaf litter. Traces of insects have also been preserved as tracks, burrows, nests, galleries, feeding damage, and larval cases (e.g., Fig. 1f). Lepidoptera, for example, are very rarely preserved in rocks, probably because they are so soft-bodied, but larval mines characteristic of various microlepidopterans occur in some fossil leaves.

The various modes of fossilization each have their biases. Entrapment in amber is biased against larger insects that could extract themselves from the resin and against insects that live in open, nonforested habitats. Also, the earliest insects preserved in amber are from only the Lower Cretaceous, some 275 million years after the earliest known hexapods appeared.

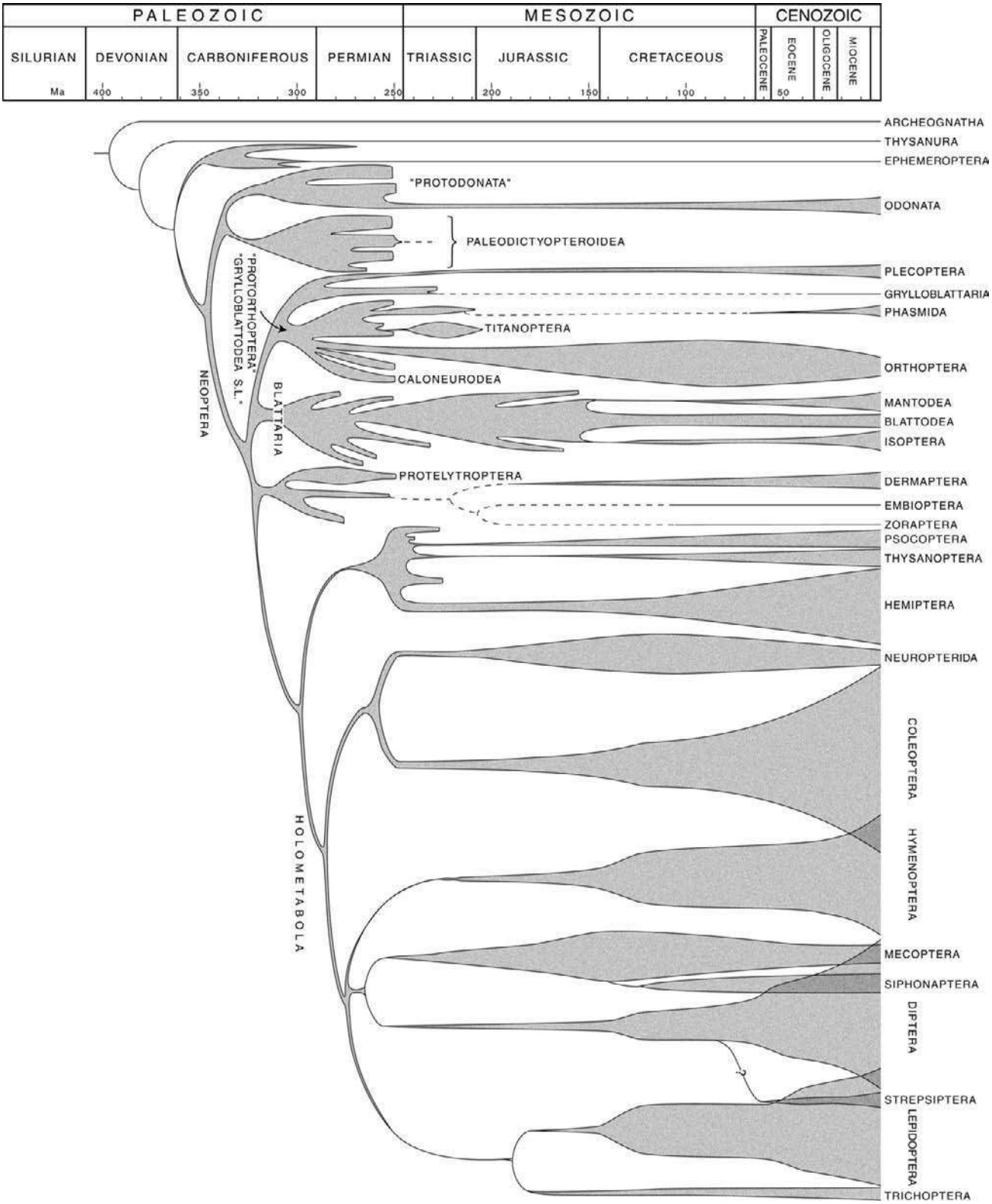

FIGURE 3 Living and extinct orders of insects, their possible relationships, and chronology. Width of lineages is a rough approximation of diversity. Some groups with a meager or nonexistent fossil record (i.e., Phthiraptera, related to Psocoptera) are not included.

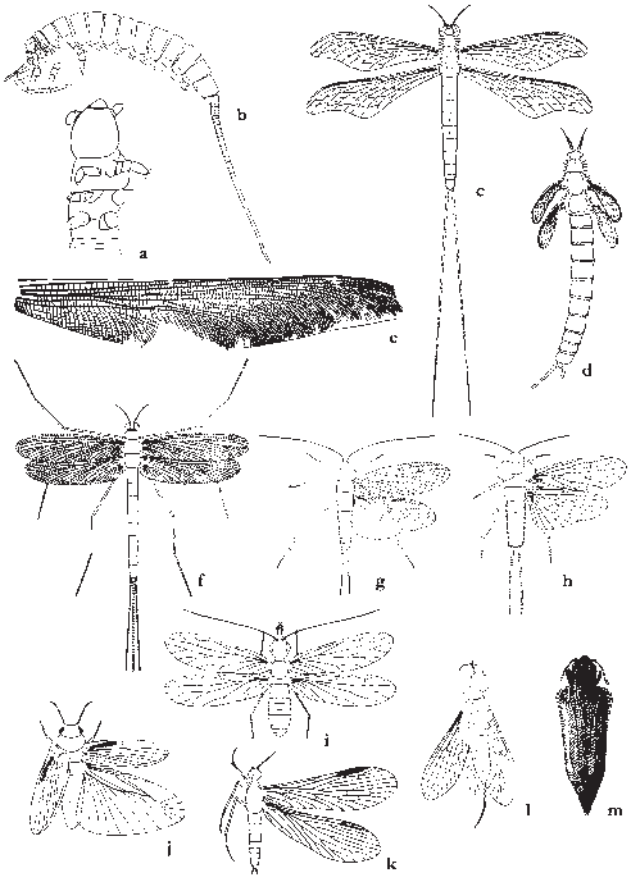

FIGURE 4 Representative Paleozoic hexapods. (a) Devonian. (c and d) Carboniferous. All others Permian. (a) †*Rhyniella* (Collembola). (b) †*Dasyleptis* (Archaeognatha). (c, d) †*Mischoptera* adult (c) and nymph (d) (†Megasecoptera). (e) †*Meganeuropsis* ("Protodonata"). (f) †*Protereisma* (near Ephemeroptera). (g) †*Liomopterum* (Paraplecoptera). (h) †*Lemmatophora* (near Plecoptera). (i) †*Dichentomum* (near Psocoptera). (j) †*Protelytron* (†Protelytroptera). (k) †*Permopanorpa* (Mecoptera). (l) †*Sojanoraphidia* (Raphidioptera). (m) †*Sylvacoleus* (Coleoptera). Not to the same scale. (Reproduced, with permission of the publisher, the Geological Society of America, Boulder, Colorado, from Carpenter, 1992.)

Preservation in rock is biased against smaller insects, and microscopic features are usually not visible against the grain of the matrix. Collectively, though, the fossil record of insects is actually much better than most paleontologists realize.

SUMMARY OF THE INSECT FOSSIL RECORD

Figure 3 summarizes the chronology, approximate diversity, and possible relationships of major groups of living and extinct insects.

Paleozoic (570–245 Mya)

The earliest remains of terrestrial animals are arthropods from the Silurian, including primitive scorpions, millipedes, and †Trigonotarbida (extinct, primitive arachnids). Two major hypotheses on the origins of the hexapods (including insects) are that they are most closely related to either the myriapods (centipedes and/or millipedes) (all comprising the Tracheata, or Atelocerata) or the Crustacea. Crustacea may actually have the oldest fossil record of all animals (formerly held by the trilobites) because some Precambrian fossils have recently been reinterpreted as crustaceans. If hexapods are closely related to crustaceans, it is most likely to be a group within Crustacea, and the earliest evidence of this stem group will probably be found in the Silurian.

THE DEVONIAN (408–362 MYA) The Devonian was a period when the continents were largely inundated and joined into the large supercontinent Pangaea. Lycopods, lycopsids, and horsetails (Equisitoidea) were common terrestrial plants; primitive groups of fishes radiated, and amphibians first appeared. The earliest record of hexapods is from the Rhynie chert of Scotland, ca. 400 mya, and is the collembolan †*Rhyniella praecursor* (Fig. 4a), although a pair of mandibles in this chert is suggestive of dicondylic insect mandibles. The Rhynie chert also contains mites, trigonotarbids, and a primitive spider.

Approximately 10 million years younger is an undescribed archaeognathan (=Microcoryphia) from the Gaspé Peninsula in eastern Canada. Apterygotes, each with a single, long caudal filament ("Monura" = Archaeognatha), occur in the Carboniferous and Devonian (Fig. 4b). Unidentified insect remains from 378 mya are known from Gilboa, New York, along with centipedes (Chilopoda) (Fig. 2a), true spiders (Araneae), trigonotarbids, oribatid mites, and pseudoscorpions. The Gaspé and Gilboa remains are original cuticle.

THE CARBONIFEROUS (362–290 MYA) The Carboniferous period is famous for the wet, warm climates and extensive swamps of mosses, ferns, seed ferns, horsetails, and calamites. Remains of insects are scattered throughout Carboniferous coal deposits (particularly blattarian wings); two particularly important deposits are Mazon Creek, Illinois, and Commentry, France. The earliest pterygotes appear in the Carboniferous, including the Blattaria, †Caloneurodea, primitive stem-group ephemeropterans (Fig. 4f), Orthoptera, †Paleodictyopteroidea (Figs. 4c and 4d), †"Protodonata" (Fig. 4e), and †"Protorthoptera"; the latter two are paraphyletic assemblages of primitive pterygotes.

Hypotheses on the evolution of insect wings include their use originally as gills or gill covers, or for mating displays, but early outgrowths of the insect pleuron most plausibly served in gliding. Feeding damage on plants is also recorded first in the Carboniferous, in the form of punctures and deep holes probably made by the long, beaked mouthparts of paleodictyopteroid insects. Thus, insects have been evolving in close association with plants for at least 350 million years, which is longer than any other group of terrestrial animals. Arborescent plants appear in the Upper Devonian, and as Carboniferous insects increasingly dwelled in them to feed, gliding probably became so adaptive for escape and dispersal that flapping wings and powered flight evolved rather suddenly.

Putative Holometabola are recorded from the Carboniferous. One is a larva from Mazon Creek, *Srokalarva berthei,* many features of which are inconsistent with extant holometabolan larvae, including segmented abdominal legs, ocelli, and possible compound eyes. Legs and body segments of *Srokalarva* are undifferentiated, as in myriapods. Some tree fern galls (ca. 300 mya) are attributed to the Holometabola on the basis of size of frass pellets in the galls. Some Paleozoic arthropods were considerably larger than living relatives, and Carboniferous gall-making mites are also known, and so it is possible that large mites caused these ancient galls. The earliest definitive Holometabola occur in the Permian.

Blattaria (Blattoptera) This group consists of the Dictyoptera (mantises, termites, and ovipositorless roaches) plus the Paleozoic and Mesozoic "roachoids" that possessed an ovipositor. Very early, Carboniferous blattarians possessed a large, discoid pronotum and coriaceous forewings with a distinctive CuP vein. From the Carboniferous to the Cretaceous there was a gradual diminishment in the length of the ovipositor. The first true, ovipositorless roaches as well as probable oothecae appeared in the Triassic. Thus, the common view that modern roaches are exceptionally ancient is inaccurate.

†Caloneurodea and †Miomoptera Both orders are known only from the Upper Carboniferous to Permian and, with Orthoptera and Blattaria, were among the earliest known Neoptera. These insects had homonomous wings with small anal lobes. In Caloneurodea the wings had many crossveins that formed numerous square cells. Wing structure and unsegmented cerci (the latter known from only few genera) suggest relationships with the orthopteroids. The wings of Miomoptera were shorter and broader and had very reduced venation. A close relationship of Miomoptera is plausibly with the Psocoptera, but probably with the Hemiptera. With a wing length of 3 to 5 mm, these are the smallest Paleozoic pterygotes known.

Orthoptera This is an ancient Recent order extending from the Carboniferous, from which time even the distinctive synapomorphy (advanced defining feature) of saltatorial hind legs is preserved (†Oedischiidae). By the Triassic, stridulatory organs on wings evolved, and some extant families had first appeared (Haglidae, Gryllidae). †Elcanidae and related families (Permian to Cretaceous) are distinctive for the reduced forewing venation with numerous parallel M and Cu veins; they were probably the most diverse and abundant orthopterans in the Late Paleozoic and Early Mesozoic. Many extant families do not appear until the Cretaceous (e.g., Eumastacidae, Gryllotalpidae, Tridactylidae, Tetrigidae) or even the Cenozoic (Acrididae, Tettigoniidae, Gryllacrididae).

†Paleodictyopteroidea This group includes the orders †Diaphanopterodea, †Megasecoptera, and †Paleodictyoptera, which had very long cerci, an ovipositor, and wings with little or no anal lobe (all primitive features). It is the only definitively monophyletic group from the Paleozoic, defined by a long, rigid beak with five stylets, as well as (primitively) a well-developed pair of possible maxillary palps. Nymphs were terrestrial, with large, free wing pads bearing rudimentary venation (Fig. 4d) and mouthparts similar to those of the adults. The group is among the most primitive pterygotes, and their distinctive parapronotal lobes (often with a rudimentary venation) probably are part of the ground plan of pterygote insects. The Paleodictyoptera are probably paraphyletic with respect to the other two orders, because they possess complete wing venation with an archedictyon. Diaphanopterodea folded wings over the abdomen at rest, an ability convergent with Neoptera. Megasecoptera had substantially reduced venation, often with a graded series of crossveins (Fig. 4c). Some had striking color patterns on the wings (e.g., Spilapteridae), and some were impressively large (e.g., *Homioptera,* approximately 40-cm wing span).

†"Protodonata" This group is a paraphyletic assemblage of primitive pterygotes similar to true Odonata but lacking the derived features of modern, true Odonata, such as the nodus, pterostigma, and arculus. Some (i.e., Permian Meganeuridae) were the largest insects ever, with a 70-cm wing span (Fig. 4e). They existed from the Upper Carboniferous to the Permian and were clearly the dominant aerial predators for some 100 million years until flying reptiles appeared in the Triassic. Fossils with the venation of true Odonata did not appear until the Triassic.

†"Protorthoptera" (=†"Paraplecoptera") (Fig. 4g) This is another paraphyletic assemblage of primitive pterygotes, with affinities not necessarily suggested by their name. Some refer to this group as the Grylloblattida *(sensu lato),* but there are no derived features that indicate a relationship of these extinct insects with this relict, extant order. Some had forewings that were coriaceous (i.e., sclerotized and leathery), but usually with an archedictyon. Unlike Orthoptera, most did not have hind wings with expanded anal lobes, and none are known to have folded their forewings roof-like over the abdomen (instead, folded flat and over each other), none had saltatorial hind legs nor a tarsomere number reduced to 3. Some had raptorial forelegs; †Geraridae had large pronotal spines.

PERMIAN (290–245 MYA) The Permian was a relatively short but very important period, when extensive mountain ranges were formed, such as the Appalachians; this caused the interior climates of continents to become cooler and drier. Extensive glaciation of the Southern Hemisphere also occurred. Voltziales and glossopterid plants radiated. Insect orders from the Carboniferous extended into the Permian, and many even extensively diversified. Numerous new orders appeared in the Permian: the †Protelytroptera (Fig. 4j) and primitive relatives of the Plecoptera ("Paraplecoptera") (Fig. 4h), Psocoptera (Fig. 4i), Mecoptera (Fig. 4k), Raphidioptera (Fig. 4l), Neuroptera, and Coleoptera (Fig. 4m), the last four orders being the earliest definitive records of the Holometabola. The most significant and diverse Permian

insect deposit is from Elmo, Kansas (ca. 260 mya); others are from New South Wales, Australia (ca. 240 mya) and central Eurasia (ca. 250 mya).

The Permian is biologically most notable for the mass extinction that marks the brief interval between the end of this period and the beginning of the Mesozoic. It is estimated that as much as 95% of all Permian life forms became extinct by the early Triassic, although this is based largely on the marine fossil record. It is difficult to estimate the impact of the Permo-Triassic extinction on insects because many of the Paleozoic orders are paraphyletic, and it is likely that particular lineages within each survived into the Mesozoic (Fig. 3). One definitive monophyletic group, the Paleodictyopterodea, may have become extinct at the end of the Permian, but a possible survivor of this group *(Thuringopteryx)* occurred in the Triassic. Otherwise, major groups of insects show few effects of the most cataclysmic extinction known.

Hemiptera The earliest Permian records are the auchenorrhynchan-like wings of †Dunstaniidae, †Palaeontinidae, and †Prosbolidae, which also extend into the Mesozoic. True auchenorrhynchans from the Triassic are Cercopoidea, Cicadoidea, Cicadellidae, Cixiiidae, and Membracoidea. The earliest Sternorrhyncha are †Archescytinidae, †Pincombeidae, †Boreoscytidae, and †Protopsylidiidae (all from the Permian into the Mesozoic). The first diverse records of true aphids and coccoids occur in Cretaceous ambers; psylloids are older. Heteroptera are slightly younger than homopterans. Presumed Permian heteropterans are †*Actinoscytina* (†Progonocimidae) and †*Paraknightia* (†Paraknightiidae). Venation of the former is barely different from Auchenorrhyncha; the latter had expanded parapronotal lobes, a large ovipositor, and forewings with unusual venation (perhaps an aberrant roach). The earliest true Heteroptera are various predatory aquatic bugs (Nepomorpha) from the Triassic of Virginia (Fig. 1d). By the Jurassic, phytophagous pentatomorphs and cimicomorphs appeared, and modern families were widespread in the Cretaceous.

Neuropterida This group includes the orders Raphidioptera and Neuroptera. Systematic position of the one Permian family of putative raphidiopteran (†Sojanoraphidiidae) (Fig. 4l) has been considered doubtful, although it had a long ovipositor distinctive to this order and a series of short costal crossveins (distinctive to Neuropterida); the wing venation, however, is quite primitive. Jurassic and Cretaceous raphidiopterans were diverse, belonging to the †Alloraphidiidae, †Baissopteridae, and †Mesoraphidiidae. Raphidioptera is relict today, with three genera in two families having a disjunct distribution in the Northern Hemisphere. The only Southern Hemisphere raphidiopterans are from the Lower Cretaceous of Brazil, indicating that the group was formerly more widespread.

Definitive Neuroptera also appear in the Permian (†Archeosmylidae, †Palaemerobiidae, †Permithonidae, and †Sialidopsidae). Triassic diversity is very poorly known (e.g., Psychopsidae), and by the Jurassic there appeared forms that are primitive relatives or members of the Chrysopidae, Coniopterygidae, Nymphidae, Osmylidae, and Polystoechotidae. An impressive Mesozoic family was the †Kalligrammatidae, with broad, patterned wings and a long, rigid proboscis. In the Cretaceous are the earliest records of the Ascalaphidae, Berothidae, Mantispidae, and Myrmeleontidae. Berothidae were particularly abundant and diverse in Cretaceous ambers. An unusual, dipterous mantispid (†*Mantispidiptera*) is known in Cretaceous amber from New Jersey.

Coleoptera The earliest Permian forms (i.e., †Tshekardocoleidae) had long, pointed, coriaceous forewings with definitive venation and arrays of small cells and pits (Fig. 4m). Beetles similar to modern cupedoids were diverse in the Upper Permian and the Mesozoic. Cupedoidea today are a small, relict group. Fossil forms were probably all wood borers as larvae and adults, like modern relatives. By the Triassic, Adephaga appeared (Carabidae, Trachypachidae) as did some basal Polyphaga (definitive Staphylinidae). By the Jurassic the first Hydradephaga (†Coptoclavidae, Dytiscidae, Gyrinidae) appeared, as well as other living families (Elateridae, Hydrophilidae, Silphidae). The earliest records of many living families are from the Cretaceous and the lower Tertiary (especially Baltic amber). Interpretation of compression fossil beetles is greatly compromised by the lack of wing venation characters and restriction to the use of simple features such as elytral structure.

Plecoptera Permian †Palaeperlidae, †Perlopseidae, Eustheniidae, and Taeniopterygidae had venation consistent with this order. Several extant families are recorded from the Jurassic and Cretaceous. †Lemmatophoridae and †Liomopteridae (Figs. 4g and 4h) had expanded parapronotal lobes (sometimes with distinctive venation like the wings) and have been placed in "Protorthoptera." These families are probably related to the Plecoptera because they had nymphs with abdominal tracheal gills and so were probably aquatic.

†Protelytroptera Found only in the Permian, this group is characterized by narrow, elytrous/tegminous forewings and with hind wings with an expansive anal lobe and radiating venation (the latter one reason they have been allied with Dermaptera) (Fig. 4j). The forewings of some species have very few veins; others have intricate reticulation. The †Umenocoleidae from the Cretaceous, formerly placed in this order, are actually blattarians.

Psocoptera The earliest putative members of this order are the Permian †Psocidiidae, but these have cerci and five tarsomeres (vs lack of cerci and generally three tarsomeres in living species) (Fig. 4i). †Psocidiidae do possess the areola postica (short, terminal branch of vein CuA), although this feature also occurs in the Zoraptera and Embioptera. Psocoptera are poorly known in the Triassic, but are diverse and abundant in Cretaceous ambers, which include many living families (e.g., Prionoglariidae, Lepidopscocidae). †Lophioneuridae (Permian–Cretaceous) have traditionally been placed in this order, but the reduced wing venation (especially in Permian †*Zoropsocus*) and (where preserved) the

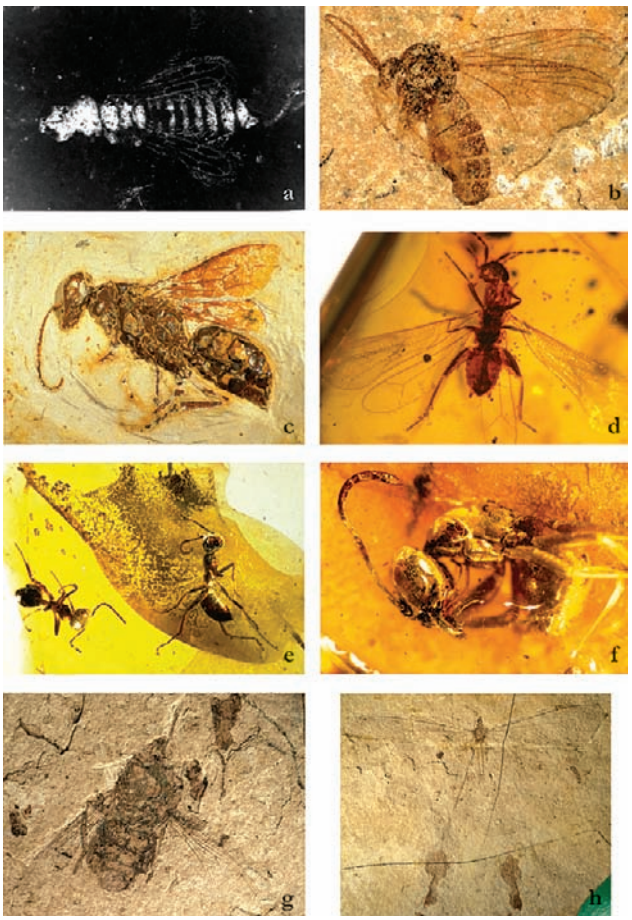

FIGURE 5 Representative Mesozoic (a–f) and Cenozoic (g, h) insects (not to the same scale). (a) Primitive, oldest known thysanopteran (Triassic, Virginia). (b) †Protorhyphidae (Diptera) (Jurassic, Kazakhstan). (c) Sphecidae (Cretaceous, Brazil). (d) Oldest known zorapteran (Cretaceous amber, Burma). (e and f) Oldest definitive ants [Formicidae; *Sphecomyrma* (e), *Kyromyrma* (f)] (Cretaceous amber, New Jersey). (g) Large, extinct tsetse fly (Diptera: Glossinidae) (U. Eocene/L.Oligocene, Colorado). (h) Spoon-winged lacewing (Neuroptera: Nemopteridae) (Colorado).

narrowed, projected mouthparts indicate relationship to the thrips (Thysanoptera). †Lophioneuridae lack extensive vein reduction, the marginal fringe, and the flagellum and tarsal structure distinctive to true thrips.

Mesozoic

The Mesozoic was an era when terrestrial ecosystems became modern, with the rise of cycads and ginkgos and the diversification of conifers in the Triassic and Jurassic. In the Lower Cretaceous there occurred what was probably the most profound evolutionary event: the explosive radiation of the angiosperms.

TRIASSIC (245–208 MYA) This was a period when arid and semiarid savannas developed and when the first mammals, dinosaurs, and pterosaurs also appeared. Most deposits with insects are from the Upper Triassic (especially Carnian, ca. 230 mya), including South Africa; Virginia; Queensland, Australia; Tadjikistan; and France. The first evidence of a diverse freshwater insect fauna appears during the Triassic, as did the oldest living families: Cercopidae, Cicadellidae, Cixiidae, and Membracidae (Auchenorrhyncha); Belostomatidae (Fig. 1d); Naucoridae (Heteroptera); Carabidae; Staphylinidae; Trachypachidae (Coleoptera); Anisopodidae, Chironomidae, and Tipulidae s.l. (Diptera); and Xyelidae (Hymenoptera). The first true Odonata, Heteroptera, Thysanoptera (Fig. 5a), Diptera, and Hymenoptera appeared during this time. The Triassic Hymenoptera consisted entirely of primitive symphytans; the Diptera consisted of diverse nematocerans. †Titanoptera were restricted to the Triassic and the size of some species fit their name (e.g., *Gigatitan,* ca. 33-cm wing span). Others were no larger than typical orthopterans, to which they are probably most closely related. They may have been Early Mesozoic analogues of the predatory mantises because their forelegs, when preserved, were spiny and apparently raptorial.

JURASSIC (208–145 MYA) One of the more significant events during this period was the origin of birds in the Upper Jurassic. Birds and (much later) bats are the only other skilled fliers and are perhaps the most important predators of modern insects. Numerous Jurassic deposits of insects occur in Europe and Asia: Grimmen and Solnhofen, Germany (the latter famous for the earliest birds, *Archaeopteryx*); Dorset, England; Issyk-Kul, Kirghizstan; and, the most diverse and productive site of all, Karatau, Kazakhstan. Jurassic deposits of insects are barely known in North America and in the Southern Hemisphere. The earliest aculeate (stinging) Hymenoptera appeared in the Upper Jurassic and were of the family †Bethylonymidae. Diptera in the Jurassic were diverse nematocerans (i.e., Fig. 5b), and the oldest definitive brachycerans evolved, dominated by Rhagionidae.

CRETACEOUS (145–65 MYA) In many respects, the insect fauna of the earlier part of the Cretaceous has more similarity to the Jurassic than to the later part of the Cretaceous. No doubt this is the result of the radiations of angiosperms 120–100 mya. Today, insects pollinate some 85% of the angiosperms, and so pollinating Cretaceous insects must have helped spawn the diversification of angiosperms; also, the diversification of phytophagous insects (and probably their parasitoids) was promoted by angiosperms. Radiations of some very speciose insect groups began during the Cretaceous, including the Scarabaeoidea, Cerambycidae, Chrysomelidae, and Curculionoidea (Coleoptera) and the Lepidoptera (the largest lineage of phytophagous organisms). The Yixian Formation of China (Lower Cretaceous, ca. 130 mya) has yielded important vertebrate fossils, as well as early angiosperms and nemestrinid flies with long proboscides; these are the earliest records of specialized insect pollination. Early sphecoid wasps (Fig. 5c), empidid and other flies, and

some beetles were probably very important, generalized pollinators of early angiosperms. Other important Cretaceous deposits are from Koonwarra, Victoria, Australia; Orapa, Botswana; Ceará, Brazil; Purbeck, Dorset, and Weald, England; Baissa, Russia; and Llérida, Spain. Isoptera appeared first in the Cretaceous, represented entirely by the primitive families Hodotermitidae, Termopsidae, Mastotermitidae, and Kalotermitidae. True, eusocial termites and ants existed for at least 50 million years before they became abundant in the Cenozoic. Eusocial wasps and bees did not first appear until the Upper Cretaceous.

Cretaceous continental drift fragmented Gondwanaland and Laurasia into the continents seen today. One result was fragmentation of ancestral ranges, the areas most famous for this being disjunct temperate regions of the southernmost regions of Africa, South America, Australia, and New Zealand (the "Austral Region"). This region harbors many primitive, relict insect groups.

Conifers were still diverse in the Cretaceous, which produced virtually all of the amber found during the Cretaceous. The oldest amber with insect inclusions is from the Lower Cretaceous of Lebanon, Japan, and England. Highly fossiliferous, somewhat younger amber deposits occur in northern Burma, western Canada, New Jersey, northern Spain, and the Taimyr Peninsula in northern Siberia. These have been exceptionally important in the study of smaller insects, having revealed the earliest bee (Apoidea), ants (Formicidae) (Figs. 5e and 5f), Zoraptera (Fig. 5d), and Stepsipteran, as well as the oldest definitive Embioptera and Lepidoptera and many families of insects. All Cretaceous Lepidoptera belong to phylogenetically basal families. Cretaceous ambers have also revealed a great diversity of primitive parasitoid Hymenoptera, sternorrhynchan hemipterans (especially Coccoidea), and empidoid flies. Cyclorrhaphan Diptera were rare and primitive in the Cretaceous.

Many of the oldest records of blood-sucking insects are from the Cretaceous: Ceratopogonidae, Culicidae, phlebotomine Psychodidae, Tabanidae, and a possible stem group to the Siphonaptera († *Tarwinia,* from Koonwarra). Unlike modern fleas, † *Tarwinia* did not have jumping hind legs and had long antennae, and so it is at best a very primitive relative of fleas. Apterous mecopteroids from the Cretaceous of Baissa and Transbaikalia, Russia, are also believed to have been vertebrate ectoparasites, perhaps of pterosaurs or feathered dinosaurs. Simuliidae appeared first in the Upper Jurassic but were still scarce and primitive in the Cretaceous. Vertebrate ectoparasitism probably first appeared in the Jurassic, but were fully developed in the Cretaceous.

Cenozoic (65 Mya–Recent)

The sudden radiation of orders of modern mammals occurred in the Paleocene (65–56 mya). This period is very poorly known from the insect fossil record, the most diverse deposit being the Fur Formation from Denmark, which preserved giant ants, noctuid moths, and others. The first butterflies (Papilionoidea) are known from the upper Paleocene.

The Eocene (56–35 mya) is far better known for insects, with extensive compression deposits from British Columbia, Canada (Horsefly, Merritt, Princeton); Green River, Wyoming; Gurnet Bay and Bembridge Marls, Isle of Wight, England; and Florissant, Colorado. Florissant (e.g., Figs. 5g and 5h) was an ancient lake inundated with volcanic ash and is probably the most diverse compression fossil insect deposit from the Cenozoic. The greatest diversity of fossil insects is preserved in Baltic amber, huge deposits of which occur throughout northern Europe (from Eocene to Oligocene). The Baltic amber has preserved the first diverse faunas of schizophoran Diptera, ditrysian Lepidoptera, advanced termites, bees, and ants, all of which are relatively young, very speciose groups whose greatest radiations occurred in the Cenozoic. At least 30 species of bees have been found in Baltic amber; paradoxically, the great majority of them are corbiculate bees belonging to extinct genera. The diversity of advanced eusocial bees was much greater than today. The oldest definitive muscoid flies and fleas occur in Baltic amber. Other important Eocene amber deposits are from Fushun, China, and Arkansas.

Deposits from the Oligocene (35–23 mya) and the Miocene (23–5 mya) indicate that the insect fauna was essentially modern, with virtually all species (extinct) belonging to modern genera. Particularly rich Oligocene deposits are from the Ruby River Basin of Montana (compression) and Chiapas, Mexico (amber). Rich Miocene deposits occur in Oeningen, Switzerland (compression), and the Dominican Republic (amber). The age of the Dominican amber has been cited as Eocene but is definitively Miocene. The Dominican amber forest was very similar to contemporary neotropical forests, but there were some groups now extinct from the Caribbean (meliponine and euglossine bees, certain genera of ants, and others) or even from the Western Hemisphere (*Mastotermes* termites and some genera of ants, acrocerid and phorid flies, and others). Dramatic extirpation of ranges has also been found in other Cenozoic deposits, such as tsetse (now entirely African) (Fig. 5g) and nemopterid lacewings (Fig. 5h, now Southern Hemisphere) from Florissant and many currently austral taxa preserved in Baltic and Cretaceous ambers of the Northern Hemisphere. Although some specimens of insects preserved in Cenozoic ambers are very similar to modern species, it is unclear (especially genetically) if these are conspecific. The average duration of insect species is difficult to estimate, but the upper limit probably extends to 10 mya.

Study of Pliocene (5–1.5 mya), Pleistocene (1.5–0.01 mya), and Holocene (10,000 years ago–Recent) deposits has been extremely useful along with that of fossil pollen in reconstructing paleoclimates and ecological succession. These remains are preserved in existing or ancient lakebeds, bogs, and tarpits. Paleoclimatic use of these remains depends on the availability of indicator taxa or readily identified extant species whose distributions are well documented. Particularly persistent and abundant in lakebeds, and

therefore commonly used, are fragments of beetles (Fig. 2g) and the larval head capsules of chironomid midges.

See Also the Following Articles

Amber • Coevolution • Endangered Insects • Nomenclature and Classification • Phylogeny • Wings

Further Reading

Carpenter, F. M. (1992). "Superclass Hexapoda," Vols. 3 and 4 of "Treatise on Invertebrate Paleontology," Part R, "Arthropoda 4." University Press of Kansas, Lawrence, and Geological Society of America, Boulder, CO.

Elias, S. (1994). "Quaternary Insects and Their Environments." Random House (Smithsonian Inst. Press), Washington, DC.

Evenhuis, N. L. (1994). "Catalogue of the Fossil Flies of the World (Insecta: Diptera)." Backhuys, Leiden.

Grimaldi, D. A. (1996). "Amber: Window to the Past." Abrams/Am. Mus. Nat. History, New York.

Grimaldi, D. A., and Cumming, J. (1999). Brachyceran Diptera in Cretaceous ambers and Mesozoic diversification of the Eremoneura. *Bull. Am. Mus. Nat. Hist.* **239.**

Hennig, W. (1981). "Insect Phylogeny." Wiley, New York.

Kukalova-Peck, J. (1991). Fossil history and the evolution of hexapod structures. *In* "The Insects of Australia," Vol. 1. Cornell University Press, Ithaca, NY.

Labandeira, C. C. (1998). Early history of arthropod and vascular plant associations. *Annu. Rev. Earth Planet. Sci.* **26,** 329–377.

Larsson, S. G. (1978). "Baltic Amber: A Paleobiological Study." Scandinavian Sci. Press, Copenhagen.

Rasnitsyn, A. P., and Quicke, D. L. J. (eds.) (2002). "History of Insects." Kluwer Acad, Publ., Dordrecht.

Rohdendorf, B. B. (1962). "Fundamentals of Paleontology," Vol. 9. "Arthropoda, Tracheata, Chelicerata." 1991 English translation of original Russian published by Amerind Pub., New Delhi.

Freshwater Habitats

see *Aquatic Habitats*

Fruit Fly

see *Drosophila melanogaster*

G

Gallmaking and Insects

Katherine N. Schick and Donald L. Dahlsten
University of California, Berkeley

Plant galls, which are abnormal growths of plant tissue that often resemble plant organs, can be induced by a wide variety of different insect species. The gallmaking insect stimulates the host plant through a complex chemical interaction so that the resulting gall is much more than a simple response to wounding or feeding damage. The precise gall form and position of the gall on the host plant is consistent and characteristic for each species of gallmaking insect.

Although cecidology, the formal study of plant galls, was initiated in 1679 with Malpighi's study of gallmaking insects, humans have admired and utilized galls for thousands of years. For example, Gallic acid (3,4,5-trihydroxybenzoic acid) was first derived from an oak gall induced by the cynipid wasp *Andricus gallaetinctoriae.* These galls have been commercially traded from source trees in the Middle East and the gallic acid derived from them has been used historically as a dye as well as an antiseptic astringent skin medication. Derivatives are used as photographic developers and the ink base made from these galls has been used to make permanent inks for such purposes as printing the U.S. dollar bill.

In 16th century England, the cynipid gall, *Biorhiza pallida,* was used for personal ornamentation. On May 29, the English parliament's official "Oak Apple Day," sprigs of oak leaves and gilded galls were worn to commemorate restoration of the English monarchy.

CECIDIA: PLANT GALLS

Cecidia, or plant galls, are abnormal growths of plant tissue under the influence of a parasitic organism. Within the growing cecidium, plant cells proliferate (hyperplasy) and enlarge (hyptertrophy) into a characteristic structure specific to that particular gallmaking organism. The organism inducing cecidogenesis (gall formation) receives nourishment and shelter while the host plant seldom benefits. Plant galls are induced by a variety of organisms including bacteria, fungi, nematodes, and arthropods.

Insect gallmakers span seven orders (see Table I) within which two entire insect families (Cecidomyiidae, the gall midges, and Cynipidae, the gall wasps) are found only within

TABLE I Insect Gallmakers

Insect order	Gallmaking insects: galls
Thysanoptera (thrips)	Thrips: roll and fold galls on leaves and buds, mostly tropical.
Heteroptera (true bugs)	Tingidae (lace bugs): galls on flowers.
Homoptera (aphids, hoppers, and scales)	Adelgidae, Aphididae, Asterolecaniidae, Cercopidae, Cicadellidae, Coccidae, Diaspididae, Eriococcidae, Eriosomatidae, Kermidae, Phylloxeridae, and Psyllidae (leafhoppers, aphids, and scales): pouch and roll galls on leaf, stem, and root; mostly of woody plants.
Coleoptera (beetles)	Curculionidae (weevils): galls on Brassicaceae.
Lepidoptera (moths)	Aegeriidae, Coleophoridae, Cosmopterygidae, Gelechiidae, Heliozelidae, Lycaenidae, Orneodidae, Pterophoridae, and Tortricidae (mostly small moths): mostly fusiform galls on stems or petioles.
Diptera (flies)	Agromyzidae, Anthomyzidae, Cecidomyiidae, Chloropidae, Platypezidae, and Tephritidae (gall midges, fruit flies, and leafminers): a variety of galls on woody and herbaceous dicots and monocots.
Hymenoptera (wasps)	Agaonidae, Cynipidae, Eurytomidae, Pteromalidae, and Tenthredinidae (sawflies, fig wasps, seed chalcids, and gall wasps): galls on all plant parts of mostly woody plants and a few herbaceous plants.

cecidia. While individual insect species make a characteristic gall on only one part of a single plant species, the thousands of gallmaking insects induce cecidia on nearly all plant parts of a wide variety of plant species worldwide.

CECIDOGENESIS: MECHANISMS OF GALL INDUCTION

The process of cecidogenesis (gall induction) involves increased levels of plant growth regulators (auxins, cytokinins, gibberellins, abscisic acid, etc.), the stimuli for which differ among taxa of gallmaking insect. For example, galls induced by tenthredinid sawflys (Hymenoptera) form in response to chemicals produced in the female accessory gland and placed on the plant at the time of oviposition. However, the chemical stimulus for galls induced by cynipid gall wasps (Hymenoptera) is released with larval feeding, and gall formation ceases if the larva dies. The exact mechanism by which insects induce gall structures characteristic to that insect species and markedly different from those of other gallmaking insects is still poorly understood.

The plant tissue stimulated to form a gall is always unspecialized parenchyma. As these plant cells undergo hyperplasy and hypertrophy, some cells may specialize to form the characteristic structures associated with that gall. However, some of the gall cells always remain unspecialized. These parenchyma cells sequester macronutrients (such as amino acids) and micronutrient minerals (such as calcium, iron, and magnesium) so that the galls act as physiological sinks in the host plant.

ECOLOGY OF GALLMAKING

Biology of Gallmakers

Formation of the gall has an adaptive advantage to the gallmaking insect, for nutritive gall tissues feed the growing larva and the gall structure hides it from natural enemies. These insect benefits of gallmaking are produced at a cost of photosynthate and energy to the host plant. The majority of insect gallmakers are plant parasites with a notable exception: fig wasps in the family Agaonidae (Hymenoptera) form mutualistic associations with their host plant (in the genus *Ficus*) as pollinators. Agaonid wasps from male flowers are introduced into fig orchards in the ancient agricultural process of caprification to allow pollination of the female flowers necessary to produce the fruit.

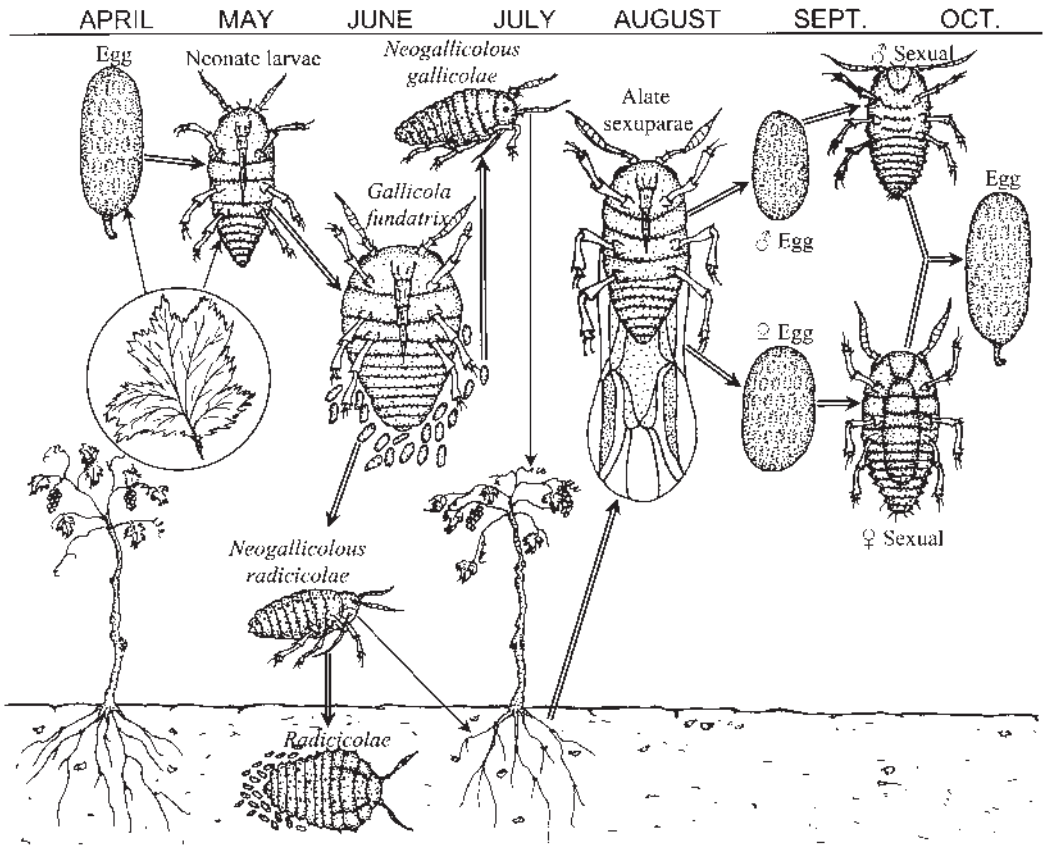

FIGURE 1 Heterogeny in grape phylloxera. In October males and females of the bisexual generation emerge, mate, and lay eggs on leaves. Their offspring form blister galls on grape leaves. In April the all-female leaf gallers lay eggs on leaves or drop to roots to lay eggs. In June all-female root gallers lay eggs on roots. Their offspring form nodular galls on grape roots. In late August a winged generation of females crawls out of the soil to lay their eggs (of the bisexual generation) under bark.

A number of gallmaking insect species exhibit heterogeny, alternating generations that include both sexes with generations including only females. Alternate generations frequently make very different galls on different parts of a plant, as does the homopteran grape phylloxera, *Daktulosphaira vitifoliae,* of which one generation induces galls on the leaves and stems while another generation induces galls on the roots of grapes. (Fig. 1). Some alternating generations of gallmaking insects form galls on different host plants, as is found in the cynipid gall wasp, *Andricus quercuscalicis,* with the unisexual generation forming galls on acorns of *Quercus robur* (English oak) and the bisexual generation forming galls on staminate flowers of *Q. cerris* (Turkey oak).

Other Insects in the Gall Community

Although a plant gall would appear to offer a place for an insect to escape predators and parasites, it is also a sedentary structure where natural enemies can predictably locate insect larvae. Predators of hidden insect larvae, such as woodpeckers, regularly prey upon galls. Many parasitoids, insects which receive their nutrients from a single host insect, eventually killing the host, have adapted to the predictability of plant galls.

In addition to gallmakers and parasitoids, the gall community also contains inquilines, which are insects which live in the gall and consume the plant gall tissue. Some inquilines deliberately kill the gallmaker, probably to prevent lignification of the gall tissues. However, many galls have both gallmaker and inquiline emerging as adults unless one or both have succumbed to parasitoids. The presence of inquilines, and the chemical stimulants they secrete, can sometimes alter the shape of the final gall, especially among cynipid gall wasps.

ECONOMIC IMPACT OF GALLS

The majority of plant galls harm the host plant only by diverting plant resources and thus have little economic impact. The economic impacts of gallmaking insects include the benefits of fig pollination as well as some negative economic effects in the form of crop losses.

Three examples of gallmaking insects with negative economic impacts are the Hessian fly (Diptera: Cecidomyiidae), the grape phylloxera (Homoptera: Phylloxeridae), and the oriental chestnut gall wasp (Hymenoptera: Cynipidae). The Hessian fly, *Mayetiola destructor,* is well known for the damage it causes to wheat crops. This fly does not cause galls on wheat, but rather exhibits a kind of facultative cecidogenesis by inducing galls only when the eggs are laid on barley, where the galls are much more damaging to the host plant than the necroses it forms on wheat. Grape phylloxera, *D. vitifoliae,* nearly destroyed the wine industries of Europe in the late 19th century. Although the leaves of European vines are resistant and are not infested with the leaf- and stem-galling forms, generations of wingless females gall the roots, eventually killing the vines. The only successful treatment has been to graft grapes onto certain resistant root strains. The economic impact of the oriental chestnut gall wasp, *Dryocosmus kuriphilus,* results from the loss of nut crop as galls form on the buds and flowers of chestnut trees (*Castanea* spp.), although the trees remain healthy.

See Also the Following Articles

Agricultural Entomology • Commercial Products from Insects • Host Seeking, for Plants • Phytotoxemia

Further Reading

Ananthakrishnan, T. N. (1984). "The Biology of Gall Insects." Oxford & IBH, New Delhi.
Csóka, G. (1997). "Plant Galls." Forest Research Inst., Budapest.
Csóka, G., Mattson, W. J., Stone, G., and Price, P. (1998). "The Biology of Gall-Inducing Arthropods." Forest Service, USDA, St. Paul.
Darlington, A. (1975). "The Pocket Encyclopaedia of Plant Galls in Colour." Blandford Press, Poole.
Meyer, J. (1987). "Plant Galls and Gall Inducers." Gebrüder, Borntraeger, Stuttgart.
Redfern, M., and Askew, R. R. (1992). "Naturalists' Handbooks." 17. "Plant Galls." Richmond Pub., Slough, England.
Shorthouse, J. D., and Rohfritsch, O. (1992). "Biology of Insect-Induced Galls." Oxford University Press, New York.
Williams, M. A. J. (1994). "Plant Galls: Organisms, Interactions, Populations," Special Vol. 49 of the Systematics Association. Clarendon Press, Oxford.

Genetically Modified Plants

D. A. Andow
University of Minnesota, St. Paul

Although there is no one absolute definition of a genetically modified (GM) plant, to some it is any plant that has had its genes deliberately altered by humans, by whatever means. This definition includes all plants produced by conventional plant breeding. Even though specific genes cannot be altered deliberately using conventional plant breeding, these conventional processes modify many genes simultaneously in statistically predictable ways. Hence conventionally produced plants can be considered GM plants, broadly speaking.

To most others, a GM plant is more narrowly defined as a plant that has been produced using transgenic methods. These plants are also called transgenic or genetically engineered plants. Transgenic methods are molecular methods that enable the transfer of a gene or potentially a group of genes from an individual of one species to an individual or individuals of a different species. Currently, there are two common methods by which purified genes are introduced into plant cells: one uses the Ti plasmid of *Agrobacterium tumefaciens* to transfer the gene as a part of the plasmid; the other uses a metal particle or fiber or an electric pulse to

pierce the cell wall and carry the gene into the nucleus (also called gene gun or electroporation). Transgenic methods enable humans to alter specific genes deliberately. The term transgenic is sometimes restricted only to genetic transfers across the species boundary, but usually includes molecular gene transfers within species as well.

The European Union uses GM plants in the narrow sense in discussing the regulation of biotechnology. Within the Cartagena Protocol on Biosafety under the Convention on Biodiversity, the term living modified organism is defined as a GM plant (narrow sense) that is intended to be grown, which excludes grain shipments and most other trade from consideration. The U.S. Office of Science and Technology Policy, which authorized the Coordinated Framework for the regulation of biotechnology in the United States uses GM plants in the narrow sense, except that it is broadened slightly to include a couple of methods that would normally be considered conventional methods. All of these organizations have chosen a definition of GM plants to exclude conventional plant breeding, in part because they do not want to regulate conventional plant breeding.

Although it has become less common, several others have used GM plants (broad sense) to blur the distinction between transgenic plants and conventionally produced plants. Because commercialized conventionally produced plants and their food products have generally been assumed to be safe for the environment and human consumption, blurring the distinction has often been a device to suggest that there are few legitimate concerns about transgenic plants. More recently, this argument has been reversed. Some conventionally produced plants are in fact potentially harmful to the environment or human health, and it may become useful to evaluate the potential dangers associated with some of these plants.

WHY THERE ARE SO MANY *BACILLUS THURINGIENSIS* PLANTS

The first commercial transgenic crops were planted in China during the early 1990s. These were primarily virus-resistant tobacco and tomato. In the United States, the first commercialized crop was Calgene's FLVR SAVR tomato in 1994. This product was not a commercial success in part because it did not pack well for shipping. Initially, a variety of transgenic crops were planted (Table Ia); by 1999, however, four crops dominated: soybean, corn, cotton, and canola. The primary traits of these GM plants are herbicide tolerance and insect resistance (Table Ib). In 1999, herbicide-tolerant soybeans, Bt *(Bacillus thuringiensis)* corn, herbicide-tolerant corn, Bt cotton, herbicide-tolerant cotton, and herbicide-tolerant canola accounted for over 99% of the commercial transgenic crops grown worldwide. All of the insecticidal-transgenic crops currently available are based on *cry* toxin genes from *B. thuringiensis,* and a few now under development are based on other toxin-coding genes from *B. thuringiensis.* The *cry* genes code for crystalline proteins that are toxic to some insects.

Bt genes have been incorporated into broccoli, cabbage, canola, cotton, corn, eggplant, poplar, potato, soybean, tobacco, and tomato, and the commercially available crops during 2001 in the United States are Bt corn and Bt cotton. Since their introduction during 1995, the cropping area of all of these transgenic crops has grown substantially (Table II). By 1999, Bt corn was grown on 9.6 million ha. Bt cotton lagged behind substantially in total area because about five times more corn than cotton is grown in the United States. Clearly, Bt corn in the United States is one of the dominant transgenic crops in the world today. Interestingly, Bt corn area has decreased during 2000, probably in response to market uncertainty. In contrast,

TABLE I Area of Transgenic Crops in the World from 1996 to 2000 (in Millions of Hectares)

	1996	1997	1998	1999	2000
(a) Crop					
Soybean	0.5	5.1	14.5	21.6	25.8
Maize	0.3	3.2	8.3	11.1	10.3
Cotton	0.8	1.4	2.5	3.7	5.3
Canola	0.1	1.3	2.5	3.4	2.8
Tobacco	1.0	1.7	+	+	+
Tomato	0.1	0.1	+	+	+
Potato	+	+	+	+	+
(b) Trait					
Herbicide tolerance	0.7	6.9	20.1	31.0	35.8
Insect resistance	1.0	4.7	8.0	11.8	11.5
Virus resistance	1.1	1.8	+	+	+
Quality traits	+	+	+	+	+
(c) Total	2.8	12.8	27.8	39.9	44.2

Note. +, <100,000 ha were grown. The first commercial crops were planted in China during the early 1990s. The first commercial production in the United States was tomatoes during 1994. Several crops were first commercialized during 1995, including Bt corn. (a) Area by crop. Several minor crops are not listed. (b) Area by transgenic trait. Values do not always sum to the worldwide total because some crops have more than one transgenic trait. (c) Total area worldwide.

TABLE II Area of Bt Corn, Bt Cotton, and Bt Potato, 1995–1997 and 2000 for the United States Only; 1998–1999 for the World, Although Nearly All Was Planted in the United States

	1995	1996	1997	1998	1999	2000	2001
Bt maize	+	0.3	2.8	6.7	9.6	6.3	5.3
Bt cotton	+	0.7	0.8	na	2.2	2.2	2.3
Bt potato	+	3650	10,000	20,000	23,000	na	+

Note. Corn and cotton data are in million hectares. Potato data are in hectares. na, data not available; +, <100,000 ha for maize and cotton and <1000 ha for potato.

Bt cotton area increased, perhaps tied to increased demand for herbicide-tolerant cotton.

Several factors probably account for the predominance of Bt genes in transgenic crops. First and foremost, transgenic technology is relatively new, and consequently the products are those that are technically feasible, are readily accomplished, and have a clear path to commercialization. The *cry* genes present in Bt have been technically easier to use than genes from plant or animal sources. The *cry* gene structure is simple, and Cry toxins require no posttranscriptional or posttranslational processing to be functional. Initially, *cry* genes did not express very well in plants, in part because bacterial DNA is A-T rich, whereas plant DNA is G-C rich. By directed mutagenesis, the "wobble" codon and other triplet redundancies were converted from A or T to G or C, and the resulting *cry* genes were expressed more consistently in plants. The commercialization path for Bt crops was believed to be relatively clear. Several Cry toxins have been used in commercial formulations of insecticides and these have been considered much safer for the environment and much less toxic to humans than nearly all other synthetic chemical insecticides. More recently, however, the differences between the Cry toxins in the insecticides and the Cry toxins in the transgenic plants have received increasing scientific and regulatory attention. Another reason for the predominance of *cry* genes is that as each particular Bt crop variety was approved for commercial use by U.S. regulatory agencies, the regulatory requirements became increasingly clear, making it increasingly easy to prepare for regulatory evaluation. Thus, the transformation technology, the regulatory environment, and the toxicological and insecticidal characteristics of Cry toxins have contributed to the present abundance of Bt crops.

THE POTENTIAL OF THIS TECHNOLOGY

There are many different kinds of transgenic crops under development, including some that are expected to be commercialized in the next few years, others that may reach commercial status on a midterm horizon sometime during the next decade, and finally others that are mere ideas that require significant research breakthroughs before they can be realized. Most of these aim to address one of four broad needs: improved agricultural characteristics, improved postharvest processing, improved food quality and other novel products for human use, and improved mitigation of environmental pollution.

Among the transgenic traits near to commercial release are new Bt genes that provide protection against additional types of insect pests. One of these traits is a gene that protects corn against rootworm damage. Some traits are based on Cry toxins, but others are based on novel Bt toxins (a binary toxin requiring two components that are toxic to insects). Other agriculturally useful traits include tolerance to abiotic stresses, such as drought, salt, and cold. These traits, however, probably require significant research breakthroughs before they can become realities.

Transgenic technology is also being applied to several commercially important tree species, including poplar, eucalyptus, aspen, sweet gum, white spruce, walnut, and apple. The first traits being genetically engineered into trees are herbicide tolerance and insect resistance, which may be useful for establishing and maintaining young trees and protecting valuable fruits. Several traits are under development to better adapt trees to postharvest processing. For example, the lignin content of certain tree species is being engineered to improve pulping, the process by which wood fibers are separated to make paper. Reduced lignin may improve the efficiency of paper production and reduce pollution from the paper-production process.

Transgenic technologies are being applied to alter macronutrients, starch, protein, oil, and micronutrients in several food crops, such as maize and soybean. For example, soy protein is deficient in the essential amino acids that contain sulfur. Increased production of these essential amino acids has been approached through genetic engineering by altering the activity of the enzymes that synthesize them, by overproducing a protein that contains them, or by blocking production of major proteins that lack them, which thereby increases the percentage of these amino acids in the remaining proteins. There are 17 minerals and 13 vitamins required at minimum levels to prevent nutritional disorders, and all of these have attracted biotechnology research. For example, increased levels of tocopherol, the lipid-soluble antioxidant vitamin E, has been engineered in the model plant *Arabidopsis thaliana* by overexpressing the gene responsible for the last step in vitamin E synthesis. Using a similar approach, current efforts are creating transgenic soybean and canola plants with enhanced levels of vitamin E.

In the future, transgenic plants may be grown to mitigate pollution. For example, it has been proposed that transgenic plants could contribute to removing or detoxifying heavy

metal pollutants in contaminated soils ("phytoremediation"). A particular problem in some locations is mercury. Plants have already been created that can accumulate mercury, and genetic engineering may be able to enhance this ability so that growing such plants could become a potential solution for cleaning up mercury pollution at despoiled sites.

Few of these future potential applications are directly related to insects or insect control. However, each of them is likely to create significant needs for entomological investigations. Traits such as abiotic stress tolerance, altered lignin content, altered macronutrient or micronutrient content, and metal accumulation are all likely to have profound effects on plant physiology and growth. Because insect herbivores and their natural enemies are very sensitive to changes in plant physiology, growth rates, and morphological structure, these new transgenic crops are likely to create new kinds of pest control problems that will require entomological solutions. Indeed, it may become necessary to devise variety-specific pest management systems for some of these novel plants.

THE LIMITS OF THIS TECHNOLOGY

The rate at which new transgenic traits can be expected to appear in the near future depends largely on the number of genes encoding them. Traits controlled by single genes, or traits that can be altered or eliminated by the loss of expression through silencing of a single gene or group of related genes, have been the first developed and commercialized. Genetically complex traits probably will require additional years of research to understand them, let alone to express and regulate them in a genetically engineered crop species. Nevertheless, many complex traits, including those controlling adaptation to abiotic stresses such as drought and salinity, flowering and reproduction, and hybrid vigor, are being investigated, and it is possible that some of these traits could appear in transgenic crop plants during the next decade.

The social acceptability of products from transgenic plants has affected and may continue to affect their adoption. Social acceptability has many components, including environmental and human health risks, food choices, the ownership of agricultural inputs and production process, the future structure of agriculture, and so on. This issue is extremely complex and volatile, and it may take several more years before it has stabilized sufficiently so that it will be possible to anticipate how various societies around the world might accept or reject transgenic plants.

Although these and other factors are involved currently in limiting the application of transgenic plants, the central trade-off that may limit it ultimately may involve a classic gene–environment trade-off in crop production. A broadly adapted trait may not be able to be used optimally without a corresponding management system. For example, the short-statured rice varieties of the Green Revolution required an intensive management system oriented around high fertilizer and pesticide use to attain their high yields across vast areas of Asia. These varieties, whether transgenic or conventional, are limited by the applicability of the attendant management system, which creates the environment in which they can flourish. Conversely, a plant with a trait adapted to specific environments associated with a single field or valley would not be broadly applicable to vast growing regions because the environment would not be equally suitable throughout. The scale of use of such locally adapted varieties depends on a system of seed production and distribution that reliably delivers the appropriate seed at the proper time. The system of self-propagated land races admirably meets these needs, but it remains to be seen whether a commercial seed distribution system is able to deliver varieties that are so locally adapted. Finally, as noted above, transgenic plant varieties are being developed for extremely specialized uses on areas of only hundreds of hectares. The scale of use of these specialty varieties will probably depend on how many different such uses are commercially successful as well as the intensivity of management needed to produce them. In conclusion, if technical factors and social acceptance do not limit the adoption of transgenic crops, the costliness of management associated with the specialty transgenic crops, the applicability of management of the broadly adapted transgenic crops, and the delivery of seed of locally adapted transgenic crops likely will limit their applicability, much like conventional varieties are limited today.

POTENTIAL ADVERSE EFFECTS

The risks associated with transgenic plants stem from, but are not directly caused by, the nature of the transformation process. First, transgenic methods enable traits to be expressed that have never before been expressed in a plant. This widened range of traits creates potential risks that should be evaluated. Second, the present transgenic methods cannot incorporate foreign DNA into precise locations in the plant genome. Because expression of genes can depend on where the gene occurs in the genome, and because the incorporation can be complex, the expression of the transgene cannot be predicted completely. The scope of potential traits and the uncertainty associated with trait expression create the circumstances requiring the evaluation of risks to human health or the environment.

Because transgenes code for proteins, human health risks associated with these proteins and products produced by these proteins are possible. These potential risks include creation of novel toxicants, possible shifts in the nutritional content of food, and the possible creation of novel allergens. Most of the scientific attention has focused on allergens, because they are difficult to assess and there has been an increase in the incidence of food allergies. Novel proteins and their products can be altered after synthesis by alterations in amino acid sequence and by reactions with other chemicals, such as glycosylation. Assessing each of these possibilities will be challenging.

Environmental risks stem from several types of potential effects: (1) effects associated with the movement of the

transgene itself and its subsequent expression in a different organism or species, (2) effects associated directly or indirectly with the transgenic plant as a whole, (3) nontarget effects associated with the transgene product outside of the plant, (4) resistance evolution in the targeted pest populations, and (5) indirect effects on human health that are mediated by the environment. The European Union (EU) recognizes affects on genetic diversity as a separate category of environmental effect in the modified 90/220 directive. The United States government has not recognized this as an environmental effect because it believes that it is the effects of altered genetic diversity, such as increased extinction rate, a compromised genetic resource, inbreeding depression, or increased vulnerability to environmental stresses, that are the actual environmental hazards. The EU recognizes this category as a precautionary measure, because the effects of movement of the transgenes are uncertain and are at present incompletely characterized. By recognizing the more easily measured, intermediate effects on genetic diversity as a potential effect, the EU risk analysis will address all of the effects caused by movement of transgenes without having to assess them specifically.

Hazards Associated with Movement of the Genes

Horizontal transfer is the nonsexual transfer of genetic material from one organism into the genome of another. Although there are no cases of transgenes moving horizontally from plants to any other organism at rates higher than normal, new discoveries could change the assessment and significance of this risk. Pollen dispersal provides an opportunity for the sexual transfer of transgenes to relatives of the crop, including other varieties of that crop, related crops, and wild relatives. Potential effects include the evolution of increased weediness (i.e., more vigorous agricultural weeds, more invasive plants) or increased risk of extinction of native species by hybridization.

Hazards Associated with the Whole Plant

The transgenic plant itself may become an environmental hazard if the traits it receives improve its fitness and ecological performance. Although many crop plants may pose little hazard, insofar as they are unable to survive without human assistance, most crops have weedy and/or wild populations in some part of their global distribution. In these areas, transgenes that improve fitness could increase weediness of the crop. In addition, because transformation includes forage grasses, poplars, alfalfa, sunflowers, wild rice, and many horticultural species, the risk of invasiveness may increase.

Nontarget Hazards

Nontarget organisms are any species that are not the direct target of the transgenic crop, and consequently, the list of potential nontarget species is very long. These organisms can be grouped conveniently into five categories: (1) beneficial species, including natural enemies of pests (lacewings, ladybird beetles, parasitic wasps, and microbial parasites), and pollinators (bees, flies, beetles, butterflies and moths, birds, and bats); (2) nontarget pests; (3) soil organisms, which usually are difficult to study and identify to species; (4) species of conservation concern, including endangered species and popular, charismatic species (monarch butterfly); and (5) biodiversity, which is the entire group of species in an area.

Hazards of Resistance Evolution

Resistance evolution can occur in pests that are targeted for control by or associated with the transgenic crop. If the pest becomes resistant, then alternative, more environmentally damaging controls may be used. Insects, weeds, and microbial pathogens all have the potential to overcome most control tactics used against them. Insect resistance to Bt crops is considered inevitable, and efforts are being made to manage resistance evolution to these transgenic crops.

Indirect Hazards

Transgenic crops can have indirect environmental impacts, especially when scaled up for commercial production. Many of these effects are associated with changes in production practices or cropping systems. For example, transgenic maize resistant to corn rootworms may lead to an expansion of continuous corn (corn planted after corn) and its attendant environmental risks, such as soil erosion. In addition, it is possible that crops transformed to produce pharmaceutical or other industrial compounds might mate with plants grown for human consumption with the unanticipated result of novel chemicals in the human food supply.

See Also the Following Articles

Biotechnology • Genetic Engineering • Insecticides • Pathogens of Insects

Further Reading

Andow, D. A. (2001). Resisting resistance to *Bt* corn. *In* "Genetically Engineered Organisms: Assessing Environmental and Human Health Effects" (D. K. Letourneau and B. E. Burrows, eds.), pp. 99–124. CRC Press, Boca Raton, FL.

DellaPenna, D. (1999). Nutritional genomics: Manipulating plant micronutrients to improve human health. *Science* **285,** 375–379.

Ellstrand, N. C., and Elam, D. R. (1993). Population genetic consequences of small population size: Implications for plant conservation. *Annu. Rev. Ecol. Sys.* **24,** 217–242.

Georghiou, G. P. (1986). The magnitude of the resistance problem. *In* "Pesticide Resistance: Strategies and Tactics for Management," pp. 14–39. National Academy Press, Washington, DC.

Green, M. B., LeBaron, H. M., and Moberg, W. K. (eds.) (1990). "Managing Resistance to Agrochemicals: From Fundamental Research to Practical Strategies." American Chemical Society, Washington, DC.

Hansen, L. C., and Obrycki, J. J. (2000). Field deposition of Bt transgenic corn pollen: Lethal effects on the monarch butterfly. *Oecologia* **125,** 241–248.

Hilbeck, A., Moar, W. J., Pusztai-Carey, M., Filippini, A., and Bigler, F. (1998). Toxicity of *Bacillus thuringiensis* Cry1Ab toxin to the predator *Chrysoperla carnea* (Neuroptera: Chrysopidae). *Environ. Entomol.* **27,** 1255–1263.

Losey, J. E., Rayor, L. S., and Carter, M. E. (1999). Transgenic pollen harms monarch larvae. *Nature* **399,** 214.

Munkvold, G. P., Hellmich, R. L., and Rice, L. G. (1999). Comparison of fumonisin concentrations in kernels of transgenic Bt maize hybrids and nontransgenic hybrids. *Plant Dis.* **93,** 130–138.

National Research Council (2002). "Environmental effects of transgenic plants: The scope and adequacy of regulation." National Academy Press, Washington, DC.

Saxena, D., Flores, S., and Stotzky, G. (1999). Insecticidal toxin from *Bacillus thuringiensis* in root exudates of transgenic corn. *Nature* **402,** 480.

Shintani, D., and DellaPenna, D. (1998). Elevating the vitamin E content of plants through metabolic engineering. *Science* **282,** 2098–2100.

Snow, A. A., and Moran-Palma, P. (1997). Commercialization of transgenic plants: Potential ecological risks. *BioScience* **47,** 86–96.

Genetic Engineering

Peter W. Atkinson
University of California, Riverside

David A. O'Brochta
University of Maryland Biotechnology Institute

Tremendous progress has been made in the development of genetic engineering technologies in insects. This article emphasizes studies with the vinegar fly, *Drosophila melanogaster,* because, as a result of its successful use as a genetic model for eukaryotic genetic systems, developments in genetic engineering with this species establish a benchmark for what can, or could, be done in other insect species. The article also discusses the more recent use of transposable-element-based genetic transformation procedures in nondrosophilid insects and concludes that many of the tools required for genetic manipulations of nondrosophilid insects are now available.

The term "genetic engineering" is typically taken to refer to the direct manipulation of genes. It has become synonymous with a more general term, "DNA technology," which has come to encompass all contemporary molecular-based techniques. However, many insect geneticists were using "DNA technology" before the development of recombinant DNA technology in the 1970s and 1980s, Genetic control approaches applied to such insect pests as the Mediterranean fruit fly *(Ceratitis capitata)* (medfly), the mosquito *(Culex tarsalis),* and the Australian sheep blowfly *(Lucilia cuprina)* used genetics to develop new strains that could be used in insect control and/or eradication programs. The tools of these pioneers were not DNA modification and restriction enzymes, thermocyclers, or DNA sequencers, but rather radiation sources, microscopes, and the knowledge that mutations and chromosomal rearrangements could be created and selected for. These tools have now been surpassed, but one aim remains the same: the generation and application of new genetic strains of insects that can be used to control pest insect species. The development of sophisticated genetic tools, in conjunction with the rapid progress being made in genomics, will provide insect scientists with the ability to characterize and manipulate, in hitherto unimaginable ways, insect genes.

GENETIC ENGINEERING IN *DROSOPHILA MELANOGASTER*

Genetic Technologies Are More Advanced in *Drosophila* Than in Other Insect Species

One cannot discuss the genetic manipulation of insects without describing the molecular genetic tools that are available in *D. melanogaster.* Traditionally, a gulf has existed between entomologists who view the harmless vinegar fly as being distant to the problems of insect control and *Drosophila* geneticists who utilize the many biological attributes of *Drosophila* to understand the basis of gene action. This gulf will close as comparative genomics reveals similarities and differences in the conservation of many genes and molecular pathways between *Drosophila* and other insect species. The power of this comparative approach to modern biology will offer insect scientists and traditional entomologists exciting opportunities to bring the power of genetics and molecular biology to the control of insects. The development and application of these tools is what insect scientists seek to achieve in pestiferous and beneficial insects.

Genetic engineering in *D. melanogaster* is an extremely mature technology. It is founded on several independent phenomena:

1. The presence of a transposable element, called the *P* element, which is an efficient genetic transformation vector. This vector has been available and exploited since the early 1980s.
2. The ability to create and maintain genetic mutants by traditional techniques such as chemical- or radiation-induced mutagenesis or by transposon insertion mutagenesis, and the construction and availability of balancer chromosomes to maintain many of these mutants.
3. The presence of strains that lack the *P* element, thus providing recipient strains suitable for *P* element transformation.
4. The completion of the *Drosophila* genome project and the public availability of the data generated.

These planks of achievement are a consequence of the intense and sustained research that has been invested into *Drosophila* over the course of the last 90 years. The picture in all other insect species is, by comparison, sparse. For example, transposable elements capable of transforming nondrosophilid species have been available only since 1996.

Also, traditional mutagenesis approaches have been used to generate mutations in a handful of insect species. Many of these have been lost because of problems arising from the rearing of these species (it should be noted that what attracted T. H. Morgan to *Drosophila* was the ease with which it could be reared and mated in the laboratory) and, often, because it had been necessary to depend on a handful of dedicated workers to maintain these strains. (In *Drosophila,* by contrast, there are central repositories for strain maintenance as well as hundreds, if not thousands, of researchers who maintain even the most problematic genetic stocks.) Except for medfly, balancer chromosomes have not been constructed in nondrososphild insects.

Two other factors are important. The interactions, if any, of the transposable elements so far known to transform non-drosophilids with components of the insect genome remain unknown, as do the molecular mechanisms by which these elements move both within and between insect genomes. Second, to date, no insect species other than *Drosophila* has had its entire genome sequenced. Some mosquito genomes are the target for current and future genomic projects.

Transformation Technologies in *D. melanogaster*

***P* ELEMENT TRANSFORMATION** Population geneticists in the 1970s had observed that when males from certain strains of flies recently established from wild populations were mated to females from long-established laboratory populations, a number of abnormal traits were observed in the progeny. These traits included high rates of mutation, sterility, and recombination in males, traits that are not usually seen in this species. Collectively the traits that arose only when specific hybrid insects were created were thought to be manifestations of a single syndrome that became known as "hybrid dysgenesis." Because the factors responsible for this syndrome were transmitted by males, they were referred to as paternal or *"P"* factors. Those working on *P*-factor-mediated hybrid dysgenesis quickly realized that there were multiple factors that mapped to many different locations within the genome. Some of the mutations that were induced during hybrid dysgenesis were very unstable and were themselves capable of mutating further to result in more extreme phenotypes or to revert to wild type. This instability as well as other genetic observations suggested that the *P* factors were mobile genetic elements or transposons. As long ago as 1989, Engels gave a comprehensive description of the *P* element, *P* factors, and their use in genetic transformation.

Concurrent with the efforts of population geneticists to understand the phenomenon of hybrid dysgenesis were efforts of molecular geneticists to clone genes from *D. melanogaster.* The eye color gene known as *white* was one of the first genes to be cloned from this species, largely owing to the great amount of genetic analysis that had been done on this locus. Having the *white* gene cloned provided a unique opportunity to isolate *P* factors. Because *P* factors were responsible for causing mutations, a genetic screen was performed to seek mutations induced by hybrid dysgenesis of the *white* gene. The reasoning behind this experiment was that once a *P*-factor-induced mutation of the *white* gene was obtained, it should, in theory, be readily cloned by conventional genomic DNA library screening using the wild-type allele of the *white* gene as a probe. By comparing the mutant allele with the wild-type allele, the nature of the *P* factor might be deduced. As expected, mutations of the *white* gene induced by hybrid dysgenesis contained insertions, and the insertion sequences had all the characteristics of a transposable element. In fact, *P* factors were transposable elements and became known as *P* elements. Complete *P* elements were about 3 kb in length and contained four open reading frames encoding for a protein essential for *P* element movement. The terminal sequences of *P* elements consisted of inverted repeat sequences of 31 bp. In structure, the *P* elements generally resembled other transposable elements that had been isolated from bacteria and were referred to as short, inverted repeat–type transposable elements.

The physical isolation of an active transposable element from *D. melanogaster* provided researchers with a unique opportunity to integrate foreign DNA into the chromosomes of this species. Efforts to integrate exogenous DNA into the chromosomes of insects can be traced back to the late 1960s. While there was an interest in genetically transforming insects and a few reports of minor successes, there was no reliable method for creating transgenic *D. melanogaster.* The *P* element solved that problem. It was reasoned that if the terminal, noncoding sequences of the element, which serve as signal sequences directing the cutting and pasting of the element, were attached to any piece of DNA, that piece of DNA would acquire the mobility properties of a *P* element. Furthermore, if this altered transposable element could be introduced into a cell that was going to form gametes and the element jumped (i.e., transposed) onto one of the chromosomes of this cell, then the gametes arising from this cell would be transgenic and would give rise to transgenic progeny. This reasoning proved to be precisely correct. Under the appropriate conditions, genes to which the terminal noncoding sequences of a *P* element have been attached can readily integrate into the chromosomes of presumptive germ cells and lead to the efficient creation of transgenic insects (Fig. 1).

This relatively simple technology helped fuel a revolution in the study of this model organism. Today, this transposable element forms the basis for a suite of technologies that allow researchers to identify and analyze genes in a variety of ways. The *P* element gene transformation system has also served as a paradigm for the development of similar technologies applicable to other species of insects.

GENE TAGGING WITH TRANSPOSABLE ELEMENTS The key to the identification and isolation of *P* elements was the availability of the cloned *white* gene. The *white* locus was used as a trap; once the *P* element had been identified and cloned, it could be used as a way of identifying and isolating

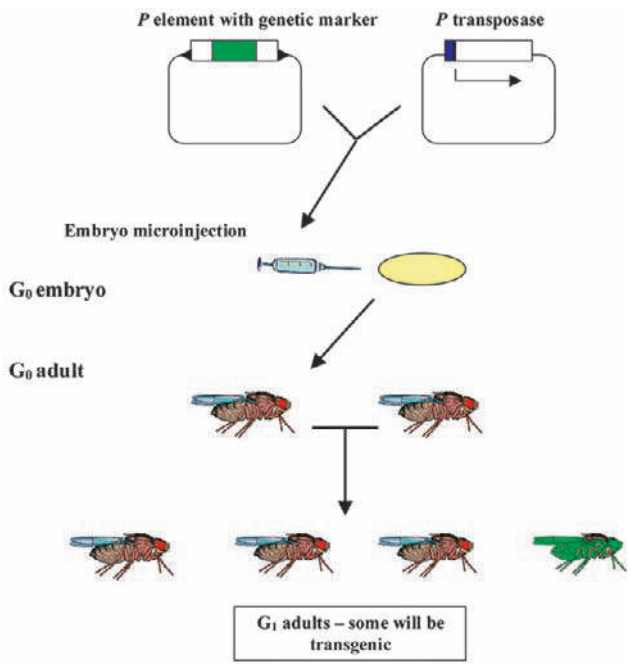

FIGURE 1 *P* element transformation of *D. melanogaster.* Two plasmids, one containing a *P* element into which has been cloned a genetic marker (green) and a helper plasmid containing the *P* element transposase (white) placed under the control of an inducible promoter (blue) are coinjected into embryos. The *P* element inverted terminal repeats are shown as black arrowheads. G_0 adults arising from injected embryos are not transgenic but some will contain a percentage of gametes containing the *P* transposable element. These adults are outcrossed and G_1 progeny are examined for the presence of transgenic individuals (green fly).

genes. As already discussed, one of the prominent features of *P* element movement (as revealed by the phenomenon of hybrid dysgenesis) was the creation of mutations. These mutations are caused by the insertion of the *P* element into an essential region of a gene, thereby altering its level or pattern of expression.

Mutations and their associated phenotypes define genetic loci. The existence of a mutant insect with an altered eye color defines a locus that plays some role in eye pigmentation. Although the existence of a mutant reveals the presence of a gene and its location, it does not provide researchers with a means of readily isolating the DNA containing the gene. If, however, the mutation is caused by the insertion of a sequence, such as a *P* element, and we know the sequence of the insertion sequence, we can use this information to isolate the DNA of the gene that was mutated. By making a genomic DNA library from the mutant insect, one can use conventional DNA hybridization techniques to identify sequences in the library that contain the *P* element. Because the mutation was caused by the insertion of the *P* element into a gene, the DNA adjacent to the *P* element is likely to be the gene responsible for the mutant phenotype. This methodology of transposon tagging is very powerful and has been used not only in *D. melanogaster* but in a number of other organisms as well. The requirements for an effective transposon-tagging system that ensures unambiguous gene identification are an active transposon that has little integration site specificity and insect strains that contain few or only one transposon-tagging transposable element. Roberts has described the use of the *P* element for gene tagging and enhancer trapping.

ENHANCER TRAPPING Transposable element-based mutagenesis or transposon tagging is a powerful technology with one limitation: it can identify only genes that have a recognizable mutant phenotype following element integration. Many of the genes that one mutates either do not result in a visible phenotype or cause the death of the organism. Such genes will never be recovered from a screen based on transposon tagging.

A complementary methodology that does not rely on mutagenesis for gene identification is called enhancer trapping. Enhancers are gene expression regulatory elements, and they function to fine-tune the control of gene expression, temporally and spatially. They are quite distinct from gene promoters in that enhancers are not sites of RNA polymerase binding but are instead sites for protein binding that influence when and how often RNA polymerase will associate with a promoter. A remarkable and useful feature of enhancers is their ability to act over long distances by mechanisms that are not entirely clear. That is, an enhancer may be located hundreds or even thousands of bases away from its target promoter. If a new promoter is inserted near the enhancer, it too will become regulated by that enhancer. This phenomenon provides a clever, nonmutagenic method for gene identification based on patterns of gene expression called enhancer trapping.

Like transposon tagging, enhancer trapping relies on the movement of a transposable element. The element in this case has been engineered to contain a gene whose expression is readily detected. Today the green fluorescent protein from the jellyfish is a common choice. The reporter gene has been engineered to contain a minimal basal promoter, meaning that it contains an RNA polymerase binding site but no associated enhancers. Consequently, this enhancerless gene construct does not result in reporter gene expression unless the transposon in which it is contained integrates near an active enhancer. The presence of enhancers can be detected by moving the transposon around the genome and looking for expression of the reporter gene. By identifying enhancers with particular properties, one then has indirectly identified the genes controlled by these enhancers. Often, the genes regulated by enhancers identified by using this method are located in the proximity of the enhancer. The significant difference between this method of gene identification and transposon tagging is that enhancer detection does not require mutating the enhancer or its associated gene. Consequently, genes that may not have been detected by a transposon tagging screen might be detected using an enhancer trap (Fig. 2).

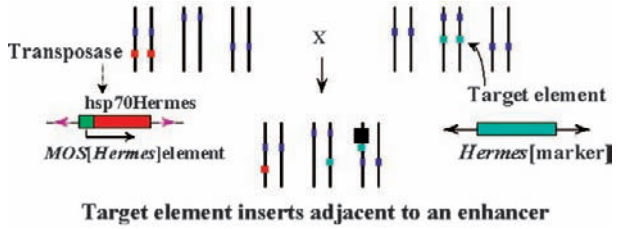

FIGURE 2 Example of enhancer trapping in insects: three pairs of chromosome with their centromeres (purple). One strain contains a *MOS* element into which the *Hermes* transposase has been cloned (orange). The *MOS* inverted terminal repeats are shown as pink arrowheads. A second strain contains a *Hermes* element containing a genetic marker (blue) placed under the control of a weak promoter. The two strains are crossed, whereupon the *Hermes* transposase causes the *Hermes* elements to move to new regions of the insect genome. Should a *Hermes* element insert near an enhancer element (black box), the genetic marker in the *Hermes* element would show the same tissue- and stage-specific expression of the gene controlled by the enhancer. The gene and the enhancer can then be cloned by standard gene tagging techniques.

Transposon tagging and enhancer trapping are rather intense genetic methods for gene identification. Such methods require the ability to efficiently perform genetic crosses, to recognize mutants or desirable reporter gene expression patterns, and then to maintain large numbers of distinct genetic lines of insects. Although *Drosophila* is readily amenable to such manipulations, other insects may be less so. Nevertheless these methods will be of great value to those entomologists working on a variety of insect species.

HOMOLOGOUS RECOMBINATION Transposon tagging and enhancer trapping are methods of identifying genes based on a phenotype: a mutant phenotype in transposon tagging, an expression phenotype in enhancer trapping. The availability of essentially the entire DNA sequence of the genome of *D. melanogaster* has permitted the identification of genes based entirely on DNA sequence patterns. Often the role of these genes is completely unknown because flies with mutations in these genes have not been identified. Without the ability to examine the phenotypes of flies with mutant alleles of the gene, gene function must be deduced entirely by other means, such as patterns of expression or analysis of the protein gene product. Today, however, it is possible for researchers who know the DNA sequence of a specific gene to create *D. melanogaster* with mutations in that gene. This method of targeted mutagenesis relies on the process of homologous recombination.

Homologous recombination, the process of gene exchange that typically occurs during meiosis, depends on the association of DNA sequences that are identical or nearly identical. Breaks in one of the strands of a DNA duplex can result in this strand becoming associated with its homologue on another chromosome, leading to gene exchange. It is now possible to exchange a gene located on a chromosome of a fly with a nearly identical gene created in the laboratory. This somewhat involved process relies on the use of a site-specific recombinase and a site-specific endonuclease, but it is potentially a method that will be generally applicable to any insect. Rong and Golic have described this technology in *D. melanogaster.*

The strategy behind using homologous recombination takes advantage of the high recombinogenicity of linear molecules of DNA. Such molecules will preferentially recombine with sequences homologous to the sequence at the end of the linear molecule. Gene targeting by homologous recombination in *D. melanogaster* is based on a clever method for generating the highly recombinogenic targeting molecule *in vivo.* The process begins by creating a transgenic insect using, for example, a *P* element gene vector that contains the targeting sequences flanked by site-specific recombination sites such as the FRT sites of the FLP recombinase system. When FLP recombinase is expressed (from a previously integrated transgene) in the insect, the FRT sites will recombine causing the targeting gene to be excised from the integrated gene vector. This recombination event results in the creation of extrachromosomal circular molecules in the nuclei of the insect. These extrachromosomal circles are then linearized by expressing a site-specific endonuclease (from a previously integrated transgene) that recognizes a DNA sequence that has been placed in the targeting gene in such a way that digestion results in the target gene sequences being located at the ends of the linearized circle. This highly recombinogenic molecule will then recombine with the chromosomal homologue, resulting in gene disruption.

Homologous gene replacement has been achieved for two *Drosophila* genes, the *yellow* gene and the *pugilist* gene, and most likely will be applicable to a large number of *D. melanogaster* genes. In particular it should enable gene function to be assigned to the thousands of new genes identified in the *Drosophila* genome project through replacing the wild-type forms with nonfunctional mutations that have been created *in vitro.*

A prerequisite for targeted gene replacement is a set of transgenic insects that can express the appropriate restriction enzyme and the FLP recombinase. This is readily achieved in *D. melanogaster* and now can also be accomplished, in principle, in other insect species because transposable elements exist that can be used to genetically transform them. The FLP recombinase system has been shown to function correctly in the yellow fever mosquito, *Aedes aegypti,* and most likely will function in all insects into which it is placed. Similarly, the ability of a yeast restriction enzyme to function in *Drosophila* suggests that it should also function correctly in a range of insect species into which it is placed.

GENETIC ENGINEERING IN NONDROSOPHILID INSECTS

Genetic Transformation of Nondrosophilid Insects

The *P* element paradigm is successful in nondrosophilid insects. Despite many attempts, the *P* element was found to be unusable as a gene vector in nondrosophilid insect species.

The reason for the narrow host range of *P* is unknown; however, it has been proposed that *P* is dependent for its mobility, in part, on the presence of host-encoded factors. These are thought to be absent, or at least sufficiently diverged, to prevent the mobility of *P* in these species. The *P* element is, however, not required for insect transformation because of the discovery and performance of four transposable elements, each from a separate family of transposable elements. Each of these is endowed with a broad host range, and each can transform *D. melanogaster* as well as a number of nondrosophilid species. They are briefly described below.

What is conserved between drosophilid and nondrosophilid transformation has been described as the *P* element paradigm. This refers to the mode of transformation. The *P* element and the four elements described shortly are class II transposable elements. They all transposase by a "DNA-only" type of mechanism—no production of an RNA intermediate is needed. These elements have an overall structure that is shared between them. They are short (< 4 kb), have inverted terminal repeated sequences, and encode a transposase enzyme that catalyzes the movement of the transposable element from one genomic location to the next. The same methodology is used to introduce these transposable elements regardless of species. Typically two plasmids are coinjected into preblastoderm embryos. One plasmid contains the transposable element, into which has been placed a genetic marker and an effector gene—a gene meant to alter the phenotype of the insect in a desired way. The placement of the marker gene and the effector gene interrupts and inactivates the transposase gene within the element, necessitating the use of a second plasmid containing the corresponding transposase, which is typically placed under the control of an inducible promoter such as the hsp70 promoter of *D. melanogaster.* This transposase mediates the transposition of the transposable element from the donor plasmid to the genome of the developing germline cells. As for *D. melanogaster* transformation, the individual arising from the injected embryo is not transformed; rather, it contains genetically transformed gametes. Individuals are mated, and transgenic insects are screened for in the next generation.

TRANSPOSABLE ELEMENTS USED FOR NONDROSOPHILID INSECT TRANSFORMATION Four transposable elements can be used to genetically transform nondrosophilid insects: *piggyBac, Hermes, Mariner,* and *Minos.*

piggyBac The 2.5-kb *piggyBac* element has 13-bp inverted terminal repeats and 4-bp direct repeats located proximally to these. It contains a 2.1-kb open reading frame that encodes a transposase enzyme. *piggyBac* was discovered through its ability to transpose from the chromosomes of the Cabbage looper *Trichoplusia ni* into the genome of a baculovirus that had infected this TN368 cell line. Transposition of *piggyBac* into the baculovirus genome led to a mutation that resulted in few polyhedra being generated, in turn causing a clear change in cell morphology. *piggyBac* inserts only at TTAA sites and generates duplications of this sequence at the target site. Excision of *piggyBac* is precise—unlike other class II insect transposable elements, no deletions or additions of DNA remain at the empty excision site. *piggyBac* has found wide use as a gene vector in insects and has been used to genetically transform the flies *C. capitata, Bactrocera dorsalis, Anastrepha suspensa, Musca domestica, L. cuprina,* and *D. melanogaster*; the mosquitoes *Anopheles albimanus, An. stephensi, An. gambiae,* and *Ae. aegypti*; the moths *Bombyx mori* and *Pectinphora gossypiella*; and the beetle *Tribolium castaneum.* Little is known about the distribution of *piggyBac* throughout insects, although highly similar elements have recently been found in three strains of *B. dorsalis.* Over the 1.5 kb of nucleic acid sequence examined, these *B. dorsalis* elements are 95 to 98% identical to the element originally isolated from *T. ni* cells. Two of these *B. dorsalis piggyBac*-like sequences contain small deletions that interrupt the open reading frame, whereas the third has an intact open reading frame over the region examined. Conceptual translation of this region yields a sequence identity of 92% compared with the corresponding region of the *T. ni piggyBac* tranposase. The basis of the distribution of *piggyBac*-like elements combined with the possible effect that incumbent *piggyBac*-like sequences may have on introduced elements in transgenic lines is a fertile field for investigation.

Hermes *Hermes* elements are members of the *hAT* family of transposable elements that are widely dispersed in animals and plants. Some members of this family, such as the *Ac* element of maize and the *Tam3* element of snapdragon, have a broad host range, and this attribute is shared with the *Hermes* element. *Hermes* was isolated from the house fly, *M. domestica,* and was first recognized by its ability to cross-mobilize the related *hobo* element when this was introduced into house fly embryos by microinjection. The 2.7-kb *Hermes* elements contain 17-bp inverted terminal repeats and a 1.8-kb open reading frame that encodes a transposase of 70 kDa. *Hermes* elements exhibit a preference for inserting at 5′-GTnnnnAC 3′ sites and create 8-bp duplications of these sites upon insertion. They have been used to genetically transform *D. melanogaster, C. capitata, Stomoxys calcitrans, Ae. aegypti, Culex quinquefasciatus,* and *T. casteneum.* Plasmid-based transposition assays have shown that *Hermes* can transposase in several other insect species as well. *Hermes* transposes by a cut-and-paste mode of transposition in higher Diptera but seems to integrate by another, transposase-dependent mode in mosquito germlines. The molecular basis of this remains unknown. *Hermes* elements can interact with the related *hobo* element (and vice versa) when both are present in the genome of *D. melanogaster.*

Mariner *Mariner* elements are widespread among arthropods. They are approximately 1.3 kb with inverted terminal repeats typically around 30 bp long. *Mariner* elements can be present in an extremely high copy number in

some species; however, it seems likely that only a handful (if any) of these may contain a single open reading frame that encodes an active transposase of approximately 33 kDa. Based on DNA sequence comparisons, five different subfamilies of *Mariner* elements exist in arthropods. The distribution of members of these subfamilies is inconsistent with the established evolutionary histories of their host species and it is now accepted that *Mariner* elements have been horizontally transferred throughout evolutionary time. At present only one naturally occurring, active *Mariner* element has been discovered. This is the *MOS* element from *Drosophila mauritiana* and has been used to genetically transform *D. melanogaster, Ae. aegypti,* and *M. domestica.* Indeed *MOS* displays a broad host range and has been used to genetically transform *Leishmania,* chickens, and zebrafish. The mobility characteristics of *MOS* are preserved in these species; it transposes by a cut-and-paste mechanism and inserts at, and duplicates, TA nucleotides. A second active element, *Himar,* was constructed based on a consensus of *Mariner* sequences obtained from the horn fly, *Haemotobia irritans. Himar* is active in *Escherichia coli* but so far is inactive in insects.

Minos The *Minos* element is a member of the *Tc*1 family of transposable elements. The *Tc*1 family of elements is related in sequence and mobility properties to the *Mariner* family of elements, and both are grouped into a single superfamily of elements. *Minos* elements are approximately 1.8 kb and possess long, 254-bp inverted terminal repeats. *Minos* contains two long open reading frames that are interrupted by an intron. Conceptual translation of the *Minos* transposase gene reveals a greater than 40% identity with the *Tc*1 transposase of *Caenorhabditis elegans. Minos* has been used to genetically transform *C. capitata, D. melanogaster*, and *An. stephensi.*

Transposable Elements in New Hosts

These four transposable elements just discussed provide the means by which genes can be introduced into pest insect species. Although these elements represent four different transposable element families, the transformation frequencies achieved, with some exceptions, are in the range of 1 to 10%. It seems likely that all will enjoy use as gene vectors in a range of insect species, and all may well be subject to interactions with endogenous transposable elements or other host factors present in these species. This is an important point that is not encountered by geneticists working on *Drosophila.* The recipient strains used for *P* element transformation are devoid of *P* elements (and any other related elements) and are deliberately chosen for this reason. This is not possible in other insect species in which the composition of the target genome with respect to transposable elements is unknown.

Whether interactions with endogenous transposable elements and/or host factors occur at levels that detrimentally affect transgenic stability is an issue that must be addressed. Central to this is development of a complete understanding of how these transposable elements are regulated both in their original host species and in species into which they have been introduced.

Genetic Markers

The development of universal genetic marker genes, together with the identification of promoters to drive their expression in heterologous species, has played a major role in the extension of genetic engineering into nondrosophilid insects. Natural and modified forms of the green fluorescent protein (GFP) gene of the jellyfish, *Aequeria victoria,* have enabled transgenic insects in several species to be easily identified from nontransgenic siblings at most stages of development. These include *D. melanogaster, C. capitata, B. dorsalis, Ae. aegypti, An. stephensi, Cx. quinquefasciatus, P. gossypiella, T. casteneum,* and *S. calcitrans.* In these species, the GFP gene has been placed under the control of a promoter that enables either organelle-specific or tissue-specific expression of the marker gene to occur. Examples of the former are the actin5C and polyubiquitin promoters of *D. melanogaster.* Examples of the latter are the Pax6 and actin88 promoters. The actin88 promoter is from *D. melanogaster* and is specifically expressed in the indirect flight muscles of the pharate adult and adults. Pax6 is a member of the Pax family of transcription factors and is specifically involved in the development of the eye and central nervous system.

The enhanced GFP (EGFP), cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and Ds Red forms of the fluorescent protein genes can also function as genetic markers in insects.

OTHER APPROACHES TO GENETIC ENGINEERING IN INSECTS

FLP/FRT Recombinase in Nondrosophilid Insects

The FLP/FRT recombinase system of the yeast *Saccharomyces cerevisiae* can also function correctly in at least one nondrosophilid species. Plasmid-based excision and integration assays showed that the FLP recombinase enzymes could recognize and recombine FRT sites in the soma of developing *Ae. aegypti* embryos. Excision at the FRT sites was high—60% of plasmids examined had undergone an excision event mediated by FLP recombinase. Integration, as measured by the formation of heterodimeric plasmids arising from the recombination between two plasmids each containing an FRT site, occurred at a low, but statistically significant, frequency. The ability of the FLP/FRT recombinase system to function correctly in *Drosophila* and *Aedes* suggests that it should function across a range of insect species. It cannot, however, be used to directly genetically transform an insect species because to achieve this, FRT sites must first be introduced into the target genome by some

other means, such as transposable elements. When combined with transposable element technology, this yeast recombination system should allow investigators to undertake precise manipulations of both introduced and host DNA. This ability will be of particular importance if DNA sequences necessary for the movement of transposable elements need to inactivated, (e.g., for regulatory reasons) following initial integration of the element into the target genome.

RNA-Mediated Interference (RNAi) in Insects

RNA-mediated interference (RNAi) refers to the targeted disruption of gene expression arising from the introduction of double-stranded RNA (dsRNA) into the cell. This disruption is targeted only to RNA molecules homologous to the invading dsRNA. It was initially characterized in plants and in the nematode *C. elegans* but is now thought to be a general phenomenon of eukaryotic cells that enables them to overcome invasions of RNA viruses. The mechanism by which RNAi works is unknown. It does not work through a simple titration of nascent or messenger RNA as would occur for an antisense RNA-based mechanism because the RNAi response can be elicited by far fewer dsRNA molecules per cell than, target RNA molecules. A catalytic mechanism in which the presence of dsRNA induces the destruction of homologous cellular RNAs has been recently proposed. RNAi technology has been harnessed to allow the targeted inactivation of specific genes and will prove to be a valuable component of genomics projects in those species in which nucleic acids can be introduced into cells. In its original experimental design, the effects of RNAi were not inherited. RNAi technology has recently been combined with *P* transposable element technology in *D. melanogaster* to produce heritable RNAi-mediated gene inactivation. Thus it is possible to examine the function of genes expressed in later stages of development of this insect and also the generation of genetically stable mutant lines in which production of the dsRNA can be induced or terminated based on the promoter used to drive expression of the extended hairpin loop RNA. RNAi technology should be extendable into other insect species in which transformation systems exist, and its exploitation in insects such as mosquitoes will enable the effects of the selective inactivation of specific genes to be quickly determined. This will represent a significant advance over traditional methods of creating and isolating mutants in these insect species that have not traditionally been amenable to genetic analyses.

EXAMPLES OF INSECT GENETIC ENGINEERING FOR INSECT POPULATION CONTROL

Transgenic technology in nondrosophilid insects has already been applied to examine promoter function and gene expression in transgenic lines of *Ae. aegypti* and *C. capitata.* In addition, recent work performed in *D. melanogaster* illustrates how transgenic approaches may be applied to pest insect control in the foreseeable future. This approach involves using transgenic technology to develop new genetic sexing strains. Although these experiments have been performed in *D. melanogaster,* the concepts are applicable to any insect species in which transgenic technology has been developed, and the ability to generate and test novel genetic strains in pest insect species should result from such additional experiments.

Both systems were centered on exploiting the tetracycline-controlled transactivator (rTA) gene, which is inactivated in the presence of tetracycline. As a dietary component, tetracycline can readily be administered to *Drosophila* larvae in measured doses. Both systems consist of two components, which are combined in a single strain when transgenic lines containing each component are crossed. The rTA gene was placed under the control of the enhancer from the yolk protein 1 *(yp1)* gene of *D. melanogaster.* This enhancer results in fat-body- and female-specific expression of the *yp1* gene. The second component of their system was a proapoptosis gene *(head involution defective—hid),* the expression of which leads to apoptosis and the death of the organism. The *hid* gene was placed under the control of the tetracycline operator (tetO), which contains the binding site for the rTA protein. Thus, in females the yp1-rTA gene is induced and, in the absence of tetracycline in the diet, the rTA protein binds to the tetO sequence and so induces the expression of the *hid* gene. All transgenic females that were raised in the absence of tetracycline and possessed both components of this lethal genetic system died. When tetracycline was added to the diet, the rTA protein was inactivated, and there was no female lethality. Males containing both components were unaffected on either diet.

These experiments clearly demonstrate that transgenic technology can be used to construct efficient genetic sexing strains in at least one species of insect—*D. melanogaster.* The genes, promoters, and enhancers chosen to do so are predicted to be of generic use in insects. The tetracycline-controlled transactivator system is from bacteria and, given that it functions correctly in *Drosophila,* will most likely be applicable to all insects in which tetracycline, or its analogues, can be delivered in measured doses. Female-specific enhancers would be expected to exist in nondrosophilids, should the *D. melanogaster* enhancers not function correctly in these species. Similarly, should conditional lethal alleles of *Drosophila* genes not function in other species, it should be possible to generate analogous mutants either by established procedures or by employing an RNAi-based approach. The extension of these strategies into nondrosophilid insects requires, in the end, genetic transformation procedures and, as already discussed, several of these now exist for nondrosophilid insect species.

CONCLUDING REMARKS

For many years the absence of genetic transformation techniques for nondrosophilid insect species was seen as bottleneck for the full extension into these important pest

species of strategies based on molecular genetics. The development of successful transposable-element-based transformation technologies enables the potential of these strategies to be tested at last. Insect geneticists have at their disposal gene vectors, universal genetic markers, promoters that can be utilized in heterologous insect species, and many target genes to test and manipulate. In addition, as outlined here, there is reason to be confident that generic techniques such as gene tagging, enhancer trapping, homologous recombination, FRT/FLP recombination, and RNAi-based gene silencing can now also be applied to insects other than *D. melanogaster.* Reports of sex-specific lethal genetic systems working in *Drosophila* have been published, and there is every expectation that similar systems will soon be established and tested in pest insects. All these technologies are precise—targeting only the genes that investigators seek to change—and the effects on a laboratory population can be predicted and are unambiguous. How successfully these technologies can be extended into pest insects, both in the laboratory and in the field, will be a matter of some interest in the years ahead.

See Also the Following Articles

Drosophila melanogaster • *Sterile Insect Technique*

Further Reading

Atkinson, P. W., and James, A. A. (2002). Germ-line transformants spreading out to many insect species. *In* "Advances in Genetics" (J. C. Hall, ed.), Vol. 47, pp. 49–86.

Atkinson, P. W., Pinkerton, A. C., and O'Brochta, D. A. (2001). Genetic transformation systems in insects. *Ann. Rev. Entomol.* **46,** 317–346.

Berghammer, A. J., Klingler, M., and Wimmer, E. A. (1999). A universal marker for transgenic insects. *Nature* **402,** 370–371.

Collins, F. H., and James, A. A. (1996). Genetic modification of mosquitoes. *Sci. Med.* **3,** 52–61.

Engels, W. R. (1989). *P* elements in *Drosophila. In* "Mobile DNA" (D. Berg and M. Howe, eds.), pp. 437–484. American Society for Microbiology, Washington, DC.

Handler, A. M., and James, A. A. (2000). "Insect Transgenesis—Methods and Applications." CRC Press, Boca Raton, FL.

Heinrich, J. C., and Scott, M. J. (2000). A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release strain. *Proc. Natl. Acad. Sci. USA* **97,** 8229–8232.

Horn, C., and Wimmer, E. A. (2000). A versatile vector set for animal transgenesis. *Dev. Genes Evol.* **201,** 630–637.

Kennerdell, J. R., and Carthew, R. W. (2000). Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nat. Biotechnol.* **18,** 896–898.

Montgomery, M. K., Xu, S., and Fire A. (1998). RNA as target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA* **95,** 15502–15507.

Morris, A. C., Schaub, T. L., and James, A. A. (1991). FLP-mediated recombination in the vector mosquito, *Aedes aegypti. Nucleic Acids Res.* **19,** 5895–5900.

O'Brochta, D. A., and Atkinson, P. W. (1998). Building the better bug. *Sci. Am.* **279,** 90–95.

Roberts, D. B. (1998). "*Drosophila*—A Practical Approach." 2nd ed. IRL Press at Oxford University Press, Oxford.

Rong, Y. S., and Golic, K. G. (2000). Gene targeting by homologous recombination in *Drosophila. Science* **288,** 2013–2018.

Rong, Y. S., and Golic, K. G. (2001). A targeted gene knockout in *Drosophila. Genetics* **157,** 1307–1312.

Genetic Variation

George K. Roderick

University of California, Berkeley

The genome, the entire collection of an organism's genetic material, provides the blueprint containing information that dictates all biological forms and functions. Without change, this blueprint would be passed identically to future generations, preserving past genetic structures that have proven to be successful, but also constraining potential future adaptation to new situations. Only when genetic variability is present can processes such as selection, genetic drift, and migration act to change the frequencies of genetic variants and in so doing allow evolution that may lead to adaptation and ultimately even speciation. Thus, understanding genetic variation—its origins, maintenance, and pattern of change—is critical to understanding the diversity of life.

GENETIC VARIABILITY: TYPES AND ORIGINS

How much genetic variation exists in natural populations? Before easy access to DNA sequences themselves, genotypic variants or polymorphisms were examined at the level of chromosome banding, particularly in the salivary glands of *Drosophila,* and through genetically based variation in enzymes as revealed through allozyme electrophoresis. Studies of enzyme variation in the mid-1960s by Lewontin and Hubby working with *Drosophila pseudoobscura* showed that an unexpectedly high number of loci were polymorphic (two or more alleles were found in 30% of all loci examined) and that over all loci nearly every individual was genetically unique. This work prompted the question of what was responsible for all this genetic variation, and in particular, did natural selection maintain this polymorphism or was the variation selectively neutral, being influenced only by processes of random genetic drift? That many loci are in fact polymorphic has been confirmed more recently with information directly from DNA sequences of both protein-coding and noncoding regions of DNA.

How is this variation then expressed in the observable phenotype? The link between the genotype and the phenotype is often relatively straightforward. Many phenotypic traits are determined by only one or a few genetic loci. However, other traits are influenced not by one or two loci, but by many loci, each with a relatively small effect; here, the link between genotype and phenotype is described by quantitative genetics. Usually such traits have measurable genetic and environmental components and frequently an interaction between the two. For such traits, the amount of variation that is "genetic" is described as the heritability, usually denoted h^2, which can be estimated through breeding studies or by examining relatives of known genetic relatedness. Crosses of inbred lines that

differ in traits of interest can be used to determine the inheritance, approximate number, and relative importance of loci responsible for variation in those traits. In such analyses, loci responsible for variation in quantitative traits are termed quantitative trait loci.

The origins of genetic variability rest in the processes of mutation, which are typically classified by the type of change caused by the mutational event. Mutations can arise through substitution (one nucleotide is replaced by another), recombination (crossing over and gene conversion), deletion (one or more nucleotides are removed), insertion (one or more nucleotides are added), or inversion (180° rotation of a double-stranded DNA segment). In protein-coding regions, some nucleotide substitutions do not change the amino acid for which they code and such substitutions are termed synonymous or silent. Substitutions that change the amino acid are termed nonsynonymous. Rates of mutation vary widely and can be influenced by the internal genetic environment as well as by the external environment. Aspects of the genetic environment that have been shown to affect rates of mutation include the functional role of the region (whether coding or not) and, for a given base, its position within a gene (e.g., stems or loops of ribosomal DNA) or within a codon (e.g., third positions change much more frequently than first or second positions due to redundancy in the genetic code). Rates of mutation are also affected by genome size and type, i.e., whether organellar or nuclear. For example, in insects, rates of mutation for protein-coding regions are typically higher for haploid mitochondrial DNA than for diploid nuclear DNA, a fact usually attributed to a lack of an efficient mechanism for DNA repair in insect mitochondrial DNA. Recent findings have shown that parts of the genomes of organisms may originate from other sources. For example, recent research on Orthoptera and Diptera has revealed nuclear DNA inserts of what were previously mitochondrial genes, and transposable elements, highly mobile pieces of DNA, are likely widespread in insect nuclear genomes. Of course, rates of mutation are also affected by external environmental variables, such as temperature and radiation.

MAINTENANCE OF GENETIC VARIABILITY

At equilibrium, the gene frequencies in a population will not change from one generation to the next. This important concept has been formalized in the Hardy–Weinberg Principle (H–W), which forms the foundation for the general understanding of population genetics of sexually reproducing organisms and, by extension, the understanding of the genetic theory of evolution. H–W is best thought of as a population genetic "null" model. The basic idea is that a single generation of random mating results in genotype frequencies that are directly predictable by the frequency of alleles in the population, no matter what history gave rise to the current mixture. This concept results in the familiar equation for a system of two alleles at one locus,

$$p^2 + 2pq + q^2 = 1, \quad (1)$$

where p and q are the frequencies of the two alleles in the population and p^2, q^2, and $2pq$ are the frequencies of the three possible genotypes (two homozygotes and heterozygote, respectively). A number of assumptions are explicit in H–W, including random mating, infinite population size (i.e., no random genetic drift), no gene flow or migration, no mutation, and individuals all having equal probabilities of survival and reproduction (i.e., no natural selection). It is the study of deviations from these assumptions that makes the principle so useful, particularly for examination of nonrandom mating, genetic drift (chance events), gene flow and migration, and selection.

ADAPTATION

Genetic variation is fundamental to Darwin's theory of evolution through natural selection, although when the idea was initially developed the mechanisms of inheritance were not yet known. Selection acts to favor some phenotypes over others, resulting in differences in relative fitness. The extent to which these phenotypes have a genetic basis determines whether those phenotypes that survive will pass on their attributes to their offspring. This process of genetic change through natural selection is termed adaptation. It can most readily be studied in recognizably polymorphic species, i.e., those in which genetic variability can be monitored over space and time. Accordingly, perhaps the best known study of natural selection in the wild is that of industrial melanism in the peppered moth, *Biston betularia,* in Britain. The typical form *(typica)* of this moth is light colored with black, pepper-like spots and is well camouflaged on lichen-covered birch trees. Following industrialization of certain areas of England, a darker form *(carbonaria)* started to appear in great abundance. This prompted the question, were the typical forms more subject to bird predation where the lichen on trees was soot covered and darker, leaving the darker forms in greater numbers? This issue was addressed by Kettlewell, in the 1950s, who manipulated the relative frequency of morphs and monitored their success. His findings were consistent with the hypothesis of predation—*typica* are eaten from dark backgrounds much more often than from light backgrounds, and vice versa for *carbonaria.* Further, a more recent reanalysis has demonstrated that, associated with the postindustrial reduction in pollution, there has been a shift to a lower frequency of the darker *carbonaria* forms. Still, a number of issues remain unresolved, demonstrating the difficulty involved in studying such systems in nature. For example, the *carbonaria* form was been found to persist even in areas relatively unaffected by pollution and two melanic forms have been shown exist in nature.

Adaptation has been studied using a number of other genetically determined polymorphisms in nature as indicators of variability. These studies include not only the

now familiar color polymorphisms, such as are found in butterfly wing patterns, ladybird beetles, walking sticks, and happy face spiders, but also behavioral polymorphisms such as caste structure in social insects. Because of the relative ease by which polymorphisms can be measured phenotypically, and the often relatively simple genetic basis that underlies the variation, studies of polymorphisms have contributed much to the general understanding of natural selection in the field, particularly in demonstrating that selection can be a very powerful force and that balancing selection can maintain genetic variability.

SPECIATION

How much genetic divergence is necessary for two populations to maintain themselves as separate species? In part, the answer to this question depends upon how the species themselves were formed. In a study of allozyme polymorphism in allopatric populations (those found in different localities) within the *Drosophila willistoni* group in Venezuela, Ayala and colleagues showed that isolated populations of the same species were very similar genetically, with a mean genetic identity (the proportion of loci identical in two samples) of $I = 0.970$. However, the genetic identity was lower in populations that were practically reproductively isolated (termed "semispecies"; $I \approx 0.8$) and even lower in well-recognized, reproductively isolated sibling species ($I = 0.517$) and closely related, but nonsibling, species ($I = 0.352$).

Species formed in sympatry (i.e., in the same locality) might potentially be more similar genetically than those formed in allopatry, particularly if only a small number of loci are important for initial divergence. Sympatric speciation is most commonly associated with true fruit flies (Tephritidae) in the genus *Rhagoletis.* Bush in the 1960s suggested that speciation in these flies may have occurred following shifts in host use within the same habitat. He offered as an example of the process the host shift of the apple maggot, *Rhagoletis pomonella,* from its native hawthorn host to introduced apples in New York State in the 1860s. There was considerable resistance initially to the concept of a nonallopatric mode of speciation. However, more recent work by Bush and colleagues has provided a convincing case for sympatric speciation in *Rhagoletis,* with differences between species being maintained through genetic control of emergence times on the different hosts. There also exists at least partial premating reproductive isolation associated with host-plant fidelity. Berlocher in 1976 measured the mean genetic identity between *R. pomonella* and two other different species of *Rhagoletis* that were likely formed in sympatry through host shifts and found the species to be nearly identical genetically ($I = 0.980$ and $I = 0.989$) and significantly more similar than were *Rhagoletis* species formed in allopatry and the *Drosophila* species noted above.

The conclusion from these studies is that the average genetic divergence itself is not critical for speciation, nor for the maintenance of reproductively isolated species, but rather that the effects of a small set of nonrandom loci may be important in species formation. Recent advances by Feder, facilitated through better understanding of the *Rhagoletis* genome, and by others studying *Heliconius* butterflies and pea aphids suggest that changes in only relatively few functional loci can lead to rapid speciation.

THE IMPORTANCE OF GENETIC VARIABILITY

How important is genetic diversity itself to the persistence of insect populations? This question is difficult to answer. Clearly, the presence of resistant alleles in pest populations has led to the development of resistance to numerous insecticides or to the development of virulence on resistant plant varieties. For example, it is not uncommon to find alleles for resistance in mass screening that takes place through field applications of insecticides or in the field use of transgenic plants expressing insecticidal proteins. By contrast, it is more difficult to detect such alleles in laboratory studies in which only a relatively small handful of individuals are sampled. Yet, for many species, genetic diversity seems to matter little. Many parasitoids, for example, are notoriously lacking in genetic polymorphisms, perhaps as a result of years of brother–sister mating. Invasive species such as the Mediterranean fruit fly, *Ceratitis capitata,* have reduced genetic variability, likely as a result of successive population bottlenecks, yet survive well while expanding their range. Indeed, the lack of variability may explain the success of some invasive species. Invasive populations of the Argentine ant, *Linepithema humile,* for example, have lost the genetic ability to distinguish one colony from another, thus escaping from population control imposed by intercolony conflict.

In these examples, high levels of genetic variability do not seem to be important for survival. However, it should be noted that the pests and invasive species that are the largest economic problems are only a very small sample of the diversity of insect species and that these species have experienced selection over many generations to be successful as pests or invasives. During this period such species may have overcome potential genetic obstacles associated with low levels of genetic variability. In natural populations of insects, however, the importance of genetic diversity still remains poorly understood.

APPLICATIONS

The genetic variability held within individuals and populations can provide critical insights into the structure and dynamics of populations that would otherwise be difficult or impossible to study. For example, variation in allele frequencies among populations can be used to assess the genetic structure of populations, summarized by F_{ST} (calculated as the variance in allele frequencies standardized by the mean) or related measures. From these measures, with appropriate assumptions,

it is often possible to infer additional information on the biological dynamics in those populations, including attributes such as effective population size (N_e), degree of inbreeding, and rates of gene flow or migration. Genetic variability also provides the opportunity for tracing the history of populations, species, and their ancestors through methods that recognize the genealogical nature of genetic material.

A second application that uses information on genetic variation is in the area of insect pest management. For example, an understanding of the genetic basis of resistance to chemical insecticides (whether administered externally or through genetic modification of plants) has been critical in the development of strategies to delay the evolution of resistance in herbivorous insects. Genetic variability is also important in the ability of insects imported for biological control to establish themselves, as well as in their potential to attack nontarget hosts.

Finally, genetic information has been of great value in the area of conservation biology and biodiversity, as for example in efforts to determine which insect species or populations are most worthy of protection. Specific applications include genetic estimation of population sizes and spread and assessment of genetic or phylogenetic uniqueness for assignment of conservation priorities. The full value of such measures based on genetic variability has yet to be realized, and future developments will almost certainly provide further insights into both past histories and future trajectories.

See Also the Following Articles

Chromosomes • Conservation • Industrial Melanism • Insecticide and Acaricide Resistance • Ladybugs

Further Reading

Berlocher, S. H., and Feder, J. L. (2002). Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annu. Rev. Entomol.* **29,** 403–433.

Falconer, D. S., and MacKay, T. F. C. (1996). "Introduction to Quantitative Genetics," 3rd ed. Longman, Harlow.

Ford, E. B. (1971). "Ecological Genetics," 3rd ed. Chapman & Hall, London.

Futuyma, D. J. (1997). "Evolutionary Biology," 3rd ed. Sinauer, Sunderland, MA.

Graur, D., and Li, W.-H. (1999). "Fundamentals of Molecular Evolution," 2nd ed. Sinauer, Sunderland, MA.

Hartl, D. L., and Clark, A. G. (1997). "Principles of Population Genetics," 3rd ed. Sinauer, Sunderland, MA.

Howard, D. J., and Berlocher, S. H. (eds.) (1998). "Endless Forms: Species and Speciation." Oxford University Press, Oxford.

Kettlewell, H. B. D. (1973). "The Evolution of Melanism." Clarendon Press, Oxford.

Price, P. W. (1996). "Biological Evolution." Saunders College, Fort Worth.

Grasshopper

see *Orthoptera*

Grassland Habitats

Teja Tscharntke, Ingolf Steffan-Dewenter, Andreas Kruess, and Carsten Thies

University of Göttingen

Grasslands are plant communities that are based on grasses and herbs, and in which shrubs are rare and trees are absent. Perennial grasses represent the dominant species of grasslands, and make up the largest portion of their biomass, but not necessarily of their species richness. Grasses are often followed by legumes in abundance and herbs from many other plant families. Grassland is the natural vegetation in areas of low or strongly seasonal rainfall (250–1000 mm), but naturally occurring mammalian herbivory (e.g., by elephants) may also effectively suppress establishment of trees. Grasslands naturally encompass a wide range of habitat and vegetation types and span a large latitudinal gradient, from tropical grassland (savannas) to temperate grassland (the prairie in North America and the steppe in Eurasia) to the arctic tundra, totaling about 25% of the earth's land surface. Herbivory in temperate grasslands is dominated by insects, whereas large ungulate herbivores dominate in tropical grasslands. The temperate meadows and pastures are seminatural grasslands growing in essentially deforested areas with a forest climate, and their succession to forests is inhibited by mowing, burning, and human-controlled grazing.

INSECT COMMUNITIES ON GRASSLANDS

Grasslands are habitats for many insects and may harbor an extraordinarily species-rich community. One temperate old-field grassland may be habitat for more than 1500 insect species, whereas cereal fields, which are monocultures of annual grasses, may contain 900 species. Compared to forests, the structural complexity of the grassland vegetation is obviously simpler, so that the insect diversity is reduced. Similarly, the litter layer of forests is larger and more heterogeneous, with a correspondingly richer decomposer community. Further features of grassland-specific insect communities include the dominance of species adapted to feed on grasses.

Plant and insect communities of grasslands greatly differ depending on climate, soil type, and management practices. Some marked differences are apparent between the plant–insect communities of temperate and tropical habitats. Plant species richness, which determines much of the insect diversity, may be only 10 to 15 species in intensively managed and highly fertilized grasslands, but 50 to 70 in extensively managed and low-input temperate grasslands. In contrast, tropical grasslands may contain over 200 plant species. Chalk-rich temperate grasslands with abundant earthworm populations tend to support the highest faunal biomass (often >100 g fresh mass

per square meter), whereas arid and semiarid steppe and desert soils, dominated by microfauna such as protozoans and nematodes, may have a biomass of only 1 g/m^2. Tropical grasslands and tundra tend to be somewhere in-between. Termites and ants are dominant groups in tropical and subtropical grasslands, some surface-dwelling predatory arachnids such as scorpions and solifugids are restricted to warm, arid soils, and cold tolerance limits the range of many species in arctic and antarctic conditions. The ways in which insect communities of grasslands are influenced will be the subject of the remainder of this article.

Which insect species attack grasses, and what are the typical plants of grasslands? Ectophages, which feed externally on leaf tissue by chewing, scraping, or sucking, are distinct from endophytic feeders, which include leafminers, gallers, and borers. Grass foliage-chewing insects belong primarily to the Orthoptera, Lepidoptera, Coleoptera (mainly Chrysomelidae and Curculionidae), Hymenoptera (Tenthredinidae), and Phasmida. Of the specialized grass chewers in Great Britain <2% are Coleoptera, 6% Lepidoptera, 6% Hymenoptera, and 41% Orthoptera (the grasshoppers). Specialization on grasses appears to be particularly important in grasshoppers, and their abundance in grasslands is high. Sap-feeders on grasses are Homoptera (Auchenorrhyncha, Sternorrhyncha, Pseudococcidae), Heteroptera (mainly Miridae), and Thysanoptera. The endophagous, mostly stem-boring herbivores belong primarily to the the Diptera (mainly Cecidomyiidae, Chloropidae, Agromyzidae), Hymenoptera (Cephidae, Eurytomidae), Lepidoptera (mainly Pyralidae, Noctuidae), Coleoptera (mainly Cerambycidae, Mordellidae, Chrysomelidae), and mites (Acari).

The number of endophagous insect species associated with grass species can be predicted by (1) the annual–perennial dichotomy (annuals, in contrast to perennials, support almost no endophagous insects); (2) the mean shoot length; and (3) the abundance of the host plant. Annuals have impoverished communities and both shoot length and abundance are positively correlated with insect diversity. Plant height is usually a surrogate for the complexity of the plant architecture and, generally, a well-known predictor of plant–insect ratios. Furthermore, the more widely distributed and abundant a plant is, the more insects it should encounter in its evolutionary history. Annuals, which typically dominate in early-successional habitats, are characterized by a faster relative growth rate than perennials and, therefore, a short exposure time, so that they are a spatiotemporally unpredictable resource for insects.

Grasses (which are monocots) are hosts of many specialized endophagous insects and a multitude of ectophagous insects, a pattern that shows no principal difference from that of dicots. Host-plant preferences of oligophagous grass feeders often differ between grass species. Even more, variability among the many commercially available strains of the perennial ryegrass *Lolium perenne* to frit fly (the stem-boring chloropid fly *Oscinella frit*) attack is greater than the variability between many pasture species. Attack of many species, such as frit fly, are negatively correlated with silica content, which presumably influences the females' choice of oviposition site and larval performance. Further, wild biotypes are often better resources than grasses grown from commercially available seeds and support richer insect communities, which may be of importance for sowings with a nature-conservation background.

GRASSES AS FOOD RESOURCE

Grasses make up the largest portion of the grassland biomass; consequently, the insect communities of grasslands are determined more by the monocotyledonous Poaceae than by the dicotyledonous herb families. Grasses differ from the Dicotyledonae in that their architecture is simple, and the intercalary meristems, which substitute for growth from terminal buds, are protected by hard leaf sheaths. Most grasses lack the variety of secondary compounds that deter herbivory in most dicotyledons. For example, cyanogens and toxic terpenoids are rare, and alkaloids are present in <0.2% of grass species but in 20% of all vascular plants. Grass-feeding insects such as the oligophagous grasshoppers select their pooid-grass host plants in that they simply reject plant tissues enriched with secondary compounds (deterrents), while no phagostimulants characterizing grasses as a group have been found. Grasses are not toxic, but this does not mean that they are little protected from herbivory; just the contrary is true (see below).

Endophytic fungi have been considered acquired chemical defenses in grasses, and the main mechanism is the production of mycotoxins, notably alkaloids. The presence of these seed-borne *Neotyphodium* endophyte fungi may cause dramatic toxicosis to grazing livestock, best known from *L. perenne* and *Festuca arundinacea.* In addition to deterring vertebrate herbivory, these endophytes are also well known for increasing resistance to insect pests, microorganisms, and drought. Endophytes may also alter attack of natural enemies in that they enhance larval development time of the herbivore (the slow growth–high mortality hypothesis) or directly affect immature enemies, e.g., parasitoids feeding on the toxic tissues of their hosts.

Within and among grass species, a considerable chemical and morphological variability may be found. Nutrient availability of grass shoots is greatly determined by the shoots' age. Fresh internodes have high concentrations of the major nutrients (water, protein, minerals) and reduced concentrations of plant-resistance factors (raw fiber, silicate). High levels of plant nitrogen are generally associated with a high assimilation efficiency and density of phytophagous insects.

HERBIVORE–PLANT INTERACTIONS

Long-term experiments with chemical control to eliminate insect herbivores indicated an average annual yield loss of

15%, and nematode control increased biomass by 12 to 28%. The biomass losses appeared to be mainly the result of frit fly and other stem-boring Diptera, which kill the central grass shoots, root-feeding wireworms (*Agriotes* spp., Elateridae), root-feeding scarabeid grubs, the range caterpillar *Hemileuca oliviae,* armyworms (*Spodoptera* spp.), grass worms (*Crambus* spp.), the Mormon cricket (*Anabrus simplex*), and leatherjackets (*Tipula* spp., in wetter soils). Planthoppers (Auchenorrhyncha), grass bugs (genera *Labops, Irbisia, Leptopterna*) and grasshoppers (Acrididae), and plant-feeding nematodes may also be important pests. In Sweden, the grass-feeding antler moth *Cerapteryx graminis* may reach densities of 100 to 1500 individuals per square meter; their corresponding effects on grass biomass consequently enhance herb populations. In the years following *C. graminis* outbreaks, shifts from herb dominance to renewed grass dominance show effects of competitive release and the return to competitive exclusion.

In temperate grasslands, the below-ground standing crop of insects is 2 to 10 times greater than the aboveground insect mass, although the effects of below-ground insects remain largely unseen, unless scarabeid beetle larvae or nematodes cause heavy decreases in shoot growth or even kill grass over large areas. In a latitudinal gradient across North American grasslands, root-to-shoot ratios vary from 2:1 to 13:1, with high values in cooler climates; tropical grasslands have even lower ratios (0.2:1 to 2.6:1). As can be expected from these data, the soil fauna is less abundant in tropical savannas and forests compared to temperate ecosystems. Earthworms usually dominate the soil biomass, but in the tropics, termites and ants are particularly important. These below-ground species can be a key in nutrient dynamics determining plant growth and aboveground plant–insect interactions.

Grasses are well adapted to herbivory and, in general, tolerate grazing better than herb species; therefore enhanced grazing pressure increases the fraction of grasses in pastures. The high resistance, tolerance, and compensatory ability of grasses are the result of (1) the generally high silicate content, lignification of vascular bundles, and additional sclerenchyma in mature leaves that make foliage hard to chew and digest; (2) the rapid induction of dormant buds that develop into lateral shoots following defoliation or destruction of apical meristems, which is based on the below-ground nutrient reserves; (3) the location of meristematic zones that are in many instances near the ground and not at the top of the plant, where they would be better accessible to grazers; and (4) the compensatory photosynthesis and growth stimulation by bovine saliva, which may also play a role. However, the concept of a herbivore-optimization curve or even grass-grazer mutualisms overestimates the compensatory abilities of grasses and grasslands.

Grazing causes much sprouting from dormant buds and converts tall canopies into shorter and denser grazing lawns. Heavily grazed pooid populations are smaller and have higher silicate concentrations, exhibiting ecotypic variation as a result of different grazing histories. The mitigation of predation by the highly silicified grasses is presumably not confined to mammals, because the mandibles of many grass-chewing insects are adapted to biting and grinding and are analogous to the teeth of grazing mammals. Although the evolution of siliceous grass leaves appears to be driven by many stress factors (including drought and fungal attack), both mammal and insect herbivory may have been important factors.

CONSERVATION OF SPECIES-RICH GRASSLANDS

Insect diversity in grassland ecosystems can be best predicted by floral diversity or related characteristics of vegetation structure, especially biomass and structural heterogeneity of the plant community. Species richness of butterflies, wild bees, phytophagous beetles, true bugs, etc., was found to be positively related to the species richness of plants. However, age of the habitat as well as fragment size is known to disproportionally enhance the number of species in higher trophic levels. The fraction of specialized predators and parasitoids increases greatly with area and age of grasslands, although the plant species richness may respond little.

Intermediate levels of vegetation disturbance, caused by ants, rodents, foxes, rabbits, sheep, and other mammals, significantly increase species richness of vegetation with consequent effects on the insect community. For example, gaps reduce the likelihood of competitive exclusion in a plant community when space is monopolized by a few dominant species. The openings are rapidly exploited by seedlings. Rotational management also may enhance grassland heterogeneity, creating a mosaic of old and young, tall and short, early and late successional patches.

Mineral or organic fertilization of meadows or rangeland increases biomass and may also enhance palatability of the nitrogen-rich foliage, resulting in higher insect densities. But the main result of continued grassland fertilization is a steady reduction in plant species richness with a corresponding loss of insect species.

Grasslands established by sowing are colonized in the beginning by relatively few insects. As these grasslands age, communities become more species rich in both plants and insects, and the biotic interactions, such as between predators and their prey or parasitoids and their hosts, increase. Ants and subterranean insects are absent on newly created fields because establishment of nests and populations needs time, and their highest densities occur in mature grasslands. Percentage of macroptery (i.e., those with full wings) in dimorphic insects such as grass-feeding planthoppers is high in early successional habitats, whereas brachypterous (short-winged) species dominate in persistent habitats.

The destruction and fragmentation of habitats has become one of the major threats to biodiversity. Not all insect species are equally affected by habitat fragmentation: species of higher trophic levels, rare species, species with specific habitat requirements, species with greatly fluctuating populations, and species

with poor dispersal abilities are expected to be more prone to extinction. For example, butterfly communities on calcareous grasslands show positive species–area relationships, and the most specialized and endangered butterflies profit most from large grassland fragments. For a few butterflies, morphological characters associated with flight ability have been shown to change with isolation of limestone habitat fragments. This indicates that habitat fragmentation in simple, human-dominated landscapes may also have evolutionary consequences for the life-history traits within populations. As a result of changes in community structure, interspecific interactions such as plant–pollinator interactions may be disrupted. Grasses are wind-pollinated, but most herbs, such as the many legume species that typically play a major role in nutrient-poor or extensively managed grasslands, depend on insect pollination. Populations of pollinating bees need nectar and pollen resources as well as suitable nesting sites. Both may be limiting in small grassland fragments, and so very small plant patches usually receive fewer pollinator visits. Plant ecologists have found clear evidence that pollination efficiency, gene flow by pollen dispersal, and seed set are reduced in small calcareous grasslands. Habitat fragmentation is known to also affect specialized populations of higher trophic levels, for example, in a plant–herbivore–parasitoid food chain. Communities of monophagous butterflies show a steeper increase with the area of grasslands than communities of plants. Theoretical models and empirical evidence show that specialized parasitoids (and predators) suffer even more, so that food chain length tends to be shortened and herbivores tend to become released from possible control of their natural enemies.

Habitat quality of species-rich grasslands such as the calcareous grasslands mainly depends on the opposing forces of management (see below) and succession. Speed of succession may be related to fragment size because late-successional shrubs and trees often invade from the edge. Abandoned grasslands will often increase in species richness of both plants and insects, but they will certainly decrease on late-successional grasslands (for example after 10–20 years of abandonment), when shrubs and trees become dominant. Many specialized butterflies mainly occur on regularly mown or grazed calcareous grasslands as they appear to rely on warm microclimates and host plants associated with only sparse vegetation (Fig. 1). The rare British butterfly *Hesperia comma* prefers small plants of the grass *Festuca ovina* surrounded by sunny bare ground and nectar resources as oviposition sites. Death of rabbits from the myxomatosis virus appeared to enhance population declines of this butterfly, because the reduced rabbit populations caused less grazing. Ground-nesting species such as solitary bees are also more abundant on regularly mown or grazed grasslands, because the sparse vegetation and open soil provide nesting sites and thereby greatly enhance populations. In contrast, aboveground-nesting solitary bees are enhanced by dense, high, and woody vegetation that offers the necessary plant material for nest construction. Altogether, species will profit from early-, mid-, or late-successional stages depending on their life-history traits, and highest overall diversity should be conserved with a mosaic of different successional stages.

FIGURE 1 Calcareous grasslands belong to the most species-rich habitat types in Central Europe and depend on annual cutting or grazing (near Göttingen, Germany, photograph by Jochen Krauss).

MANAGEMENT OF GRASSLAND

Cutting, grazing, and burning are typical methods of grassland management. As management alters plant growth and vegetation structure profoundly, the community of associated insects will also change. The insects' responses greatly differ between functional and taxonomic groups, and consequently it is often difficult to decide which management strategy is best in the conservation of overall diversity. Closely cut (or grazed) grasslands typically have an impoverished insect fauna. This is partly the result of pronounced vertical stratification of species using different parts of the sward canopy during the growing season (Fig. 2) and has been shown for planthoppers and leafhoppers (Auchenorrhyncha) as well as phytophagous beetles (Coleoptera). In particular, cutting affects flower visitors, pollen feeders, and grass-seed feeders among the gall midges (Cecidomyiidae) and other groups (Miridae, Chloropidae, Thripidae). In general, vegetation height and, therefore, the structural complexity of grasses decrease with intensity of grazing or mowing, and the complexity of plant architecture is a good predictor of insect species richness. The positive correlation between aboveground plant biomass and insect species richness is well established, whereas root feeders and other soil invertebrates are often more abundant at intermediate levels of grazing or mowing (of temperate grassland) than on unaffected patches. When grassland has been left unmanaged for a few years, the hemipterous and coleopterous fauna quickly recovers. With increasing number of mowings per year, which may be best observed in the sometimes extremely often mown urban turf-grass areas, species richness of both plants and insects (e.g., planthoppers

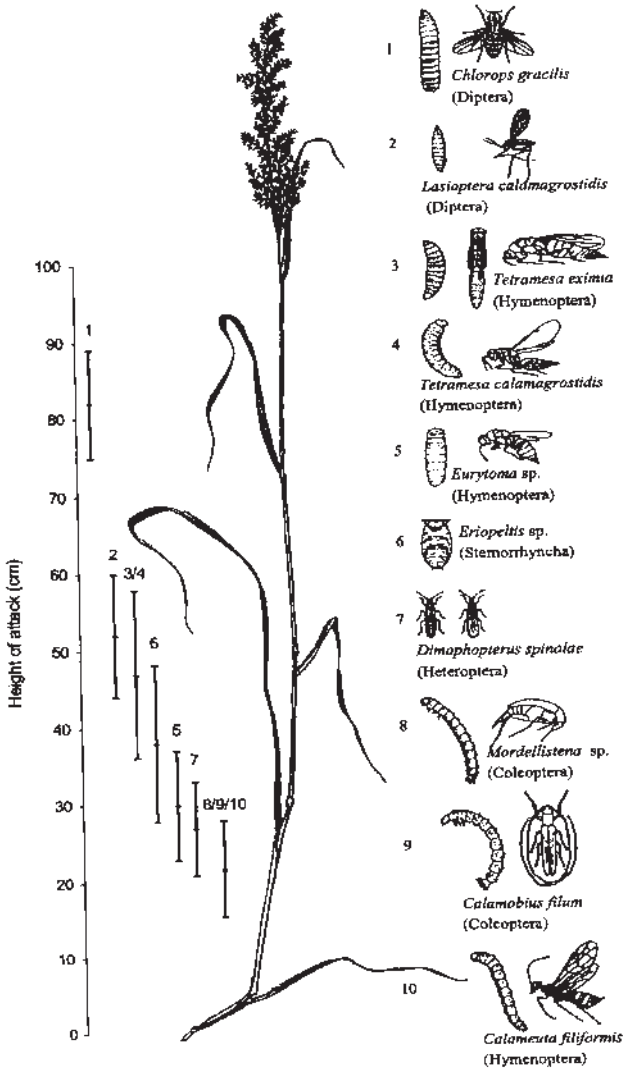

FIGURE 2 The stem-boring insects feeding on pure stands of the grass *Calamagrostis epigeios.* Height of attack (arithmetic means and 95% confidence limits) is given for each of the 10 species. (Reproduced, with permission, from M. Dubbert, T. Tscharntke, and S. Vidal, 1998, *Ecol. Entomol.* **23,** 271–280.)

and true bugs) becomes extremely poor. Moderate cutting or grazing may promote grasshopper populations, possibly through tillering rejuvenation or through changes in the proportions of nutritious grasses. Because the quality of grass shoots as a food resource declines with age, the induction of tillers and side shoots by cutting make nutritious food available later in the season (for example, for the populations of many grasshoppers and enhanced infestations of stem-boring Diptera). In conclusion, the effects of cutting or grazing on insect communities can be divided into short-term effects (simplification of plant architecture, regrowth of young and nutrient-rich plants) and long-term effects caused by changes in the structure of plant communities.

Grazing adds to the effects of cutting in that grazers feed selectively on the more palatable plants, compress or loosen the ground by trampling, and fertilize grassland patches by urination and the deposition of dung. Accordingly, the changes in the plant community following grazing affect insect community structure in a complex way and make the habitat more heterogeneous than a homogeneous cutting regime. Further, grazing is a gradual form of vegetation removal, except at high stocking densities, and thereby differs from the large-scale disturbance of cutting (or burning). Cattle feed on taller vegetation than sheep and may open up tall vegetation. The cattle's trampling effects are usually high compared to those of sheep and enhance vegetation heterogeneity with disturbed and bare areas that improve habitat quality of many invertebrates. High cattle densities, however, lead to short and uniform swards and create problems due to vegetation damage, especially on wetter soils and on slopes. Most of the nutrients removed by grazing are returned through the deposition of urine and dung. Cow dung harbors a unique and speciose insect community. The breakdown of ungulate dung in temperate environments is enhanced by fly maggots, such as *Scatophaga* sp. and dung-burying beetles such as the scarabeid *Geotrupes* sp. Deposition of bovine dung poses no problems where bovines have an evolutionarily associated fauna that exploits the fecal resources. However, in Australia, native detritivores could not process cow dung because cows were brought over by the first English colonists only at the end of the 18th century. The loss of pasture under dung has imposed a huge economic problem to agriculture in Australia, and only the decision in 1963 to establish African dung beetles there led to a solution.

Burning grassland is less common in Europe than in America or Australia, where burning is a widespread natural phenomenon. Burning, like cutting and grazing, tends to produce a greater floristic uniformity, and it is considered to be very detrimental to grassland invertebrates. Controlled burning has been suggested as an alternative to chemical or biological control of pest arthropods. Direct effects are diverse, depend on the intensity of the burn, and include the escape of many flying insects as well as few changes in many soil insects. Indirect effects are the xeric conditions after burning and the mineral-rich regrowth after burning, which for many animals is a superior resource quality.

See Also the Following Articles

Biodiversity • Plant–Insect Interactions • Soil Habitats

Further Reading

Curry, J. P. (1994). "Grassland Invertebrates." Chapman & Hall, London.

Siemann, E. (1998). Experimental tests of effects of plant productivity and diversity on grassland arthropod diversity. *Ecology* **79,** 2057–2070.

Stanton, N. L. (1988). The underground in grasslands. *Annu. Rev. Ecol. Syst.* **19,** 573–589.

Steffan-Dewenter, I., and Tscharntke, T. (2002). Insect communities and biotic interactions on fragmented calcareous grasslands—A minireview. *Biol. Conserv.,* **104,** 275–284.

Tscharntke, T., and Greiler, H.-J. (1995). Insect communities, grasses, and grasslands. *Annu. Rev. of Entomol.* **40,** 535–558.

Watts, J. G., Huddleston, E. W., and Owens, J. C. (1982). Rangeland entomology. *Annu. Rev. Entomol.* **27,** 283–311.

Greenhouse Gases, Global Warming, and Insects

Peter Stiling
University of South Florida

Greenhouse gases, the gases involved in determining the Earth's average temperature and climate, are accumulating at a rapid rate within the atmosphere. Such gases include carbon dioxide, methane, nitrous oxide, ozone, and chlorofluorocarbons. By far the most important of these is carbon dioxide, CO_2, whose contribution to the total greenhouse gas warming effect is at least 50%. For this reason, nearly all research pertaining to insects and greenhouse gases has focused on the response of insects to elevated levels of CO_2. Global atmospheric carbon dioxide levels are increasing at an astonishing rate, mainly because of the burning of fossil fuels. The atmospheric concentration of CO_2 has increased from a preindustrial level of about 270 ppm to a current level of about 365 ppm, an increase of nearly 100 ppm or 35%. According to some reports, the atmospheric concentration of CO_2 will likely stabilize at four times the preindustrial levels. Most studies indicate that CO_2 levels will at least double from preindustrial levels over the next five to ten decades. This increase represents one of the most large-scale and wide-reaching perturbations to the environment.

Many of the changes in insect populations likely to result from elevated CO_2 will be brought about by changes in plant chemistry. The chemical changes in plants result from increases in plant carbon, decreases in nitrogen, and increases in levels of defensive compounds such as phenolics. In addition, global warming, which will result from elevated levels of greenhouse gases, may increase the reproductive capabilities of some insects and change their distributional ranges. This could change the abundance of some pest species and disease vectors. Some of the myriad effects of elevated CO_2 on insects are summarized in Fig. 1.

STUDYING THE EFFECTS OF GREENHOUSE GASES ON INSECTS

Most published studies on the effect of CO_2 on insects tell of experiments in which plants and insects are confined to CO_2 levels of 700 to 710 ppm, or about double the current level. Such experiments are typically conducted in the laboratory, where well-watered potted plants are grown in nutrient-rich soil and maintained under elevated CO_2 for several months. Insects are introduced onto these experimental plants, and their feeding rates and performance are measured.

Studying the effects of elevated CO_2 on altered temperature and rainfall patterns, and the effects on insects of these modifications of the environment is much more problematic. It is not easy to warm whole communities in the field, except through the use of greenhouses—which tend to change many other features such as precipitation patterns. Thus the effect of temperature is often studied on laboratory populations. In addition, mathematical models are used to determine the likely range alterations of plants and insects in the face of increased global temperatures and changes in precipitation patterns.

CHANGES IN PLANT CHEMISTRY

Plants commonly respond to elevated CO_2 by increasing their rates of photosynthesis. Higher rates of photosynthesis usually result in higher accumulations of carbon-rich carbohydrates. Furthermore, the increased atmospheric CO_2 levels mean that stomatal conductance is reduced because plants can get sufficient atmospheric CO_2 into their leaves even when their stomates are closed more often. A reduction in stomatal conductance results in greater efficiency of water use by plants, because less water is lost through transpiration. Both these factors have important effects on plant chemistry.

First, increased carbon uptake by plants results in higher plant growth rates, with leaf area index, woody biomass, and below-ground biomass sometimes increased by as much as 25 to 50%. Despite the increase in plant growth, there is usually no increase in the availability of soil nutrients, particularly nitrogen, and these nutrients must be spread further among the available plant biomass. The usual result is a decrease in total plant nitrogen because nitrogen is diluted over the entire plant. Herbivore growth is most often limited by nitrogen rather than by carbon, so that plants grown in atmospheres of elevated CO_2 become poorer quality forage. Plant–water content usually affects digestibility, so that the poorer quality diet is partly offset by an increased ease of digestion.

The second major change in plants grown under conditions of elevated CO_2, namely, is a change in the ratio of carbon to nitrogen (C:N), as described in the preceding section. This has major implications for the concentration of defensive compounds in the leaves, the so-called secondary chemicals. Carbon-based secondary chemicals often increase and deter insect feeding. The overall effect of increased CO_2 on insect herbivores is to decrease plant palatability because of decreases in nitrogen levels and increases in secondary chemicals.

For secondary chemicals, the increases seem to be greatest for soluble phenolic compounds, especially condensed tannins, which are found in a variety of trees, especially oaks. These compounds are known to negatively affect many herbivorous insect species. Yet for other defensive compounds, such as linear furanocoumarins, found in celery, and monoterpenes and sesquiterpenes, found in peppermint, little increase has been noted.

CHANGES TO INSECT HERBIVORE FEEDING

There have been over 40 studies of the performance of insect herbivores of various types under conditions of elevated CO_2.

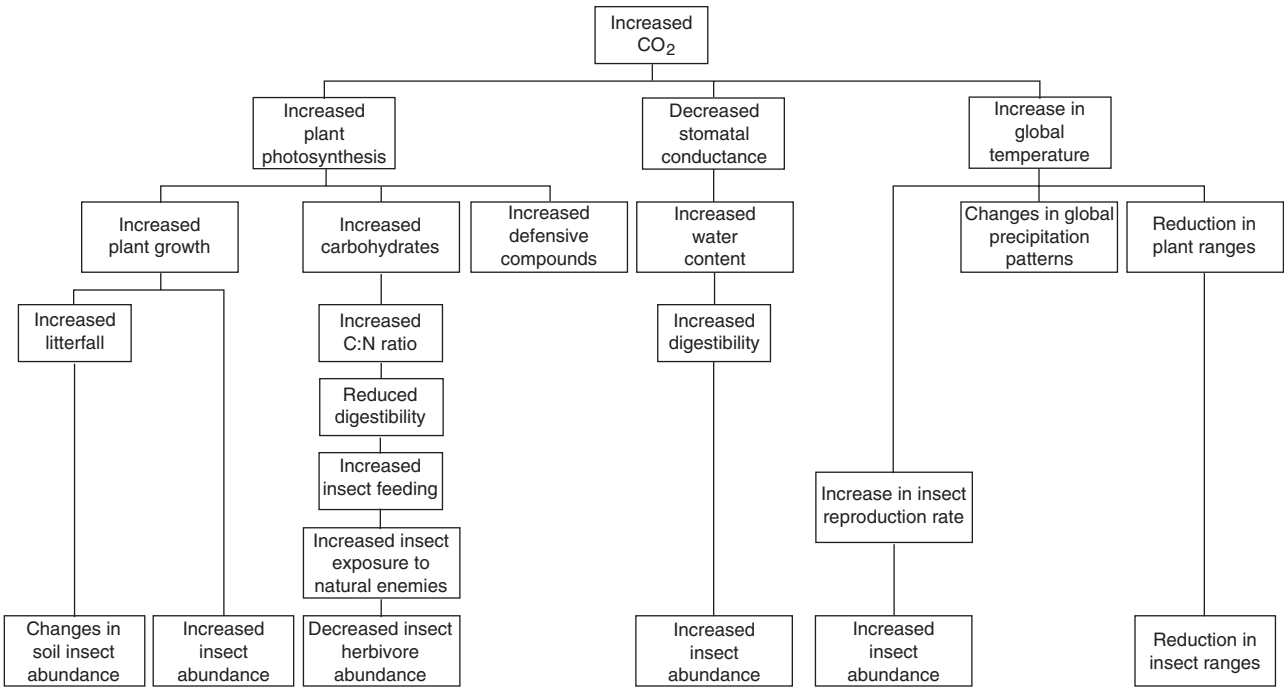

FIGURE 1 Schematic representation of the effects of elevated CO_2 on insects.

The majority of these, over 80%, have been conducted with leaf-chewing insects, especially lepidopteran caterpillars. The most commonly reported change is that food consumption increases as the insects struggle to obtain sufficient nitrogen in their diet. The efficiency of food conversion to insect biomass (conversion efficiency) decreases, probably because of the increased concentration of secondary chemicals, such as tannins, which bind digestive enzymes and render them less effective. Thus, it takes insects much longer to develop, and their final weight is often reduced. Early instars seem to be more susceptible than late instars. Of course such changes in diet could, in theory, be partly offset by the increase in digestibility due to the increased water content. However, at least in the studies done so far, the net outcome of elevated CO_2 on herbivorous insect digestibility has been negative.

The responses just outlined may vary somewhat according to the feeding guild of insects involved. Thus, chewing insects, which often digest the whole leaf and encounter both reduced nitrogen levels and increased defensive compounds, are particularly susceptible to changes in nitrogen and phenolics. Insects that feed in a different way may be less susceptible. Phloem and xylem feeders in particular may be less affected by CO_2 because they feed on plant sap, which is low in defensive compounds. Seed feeders also may be less affected by increased CO_2 because these plants try to maintain high levels of nitrogen in their reproductive parts. In cotton, for example, the C:N ratio of cotton balls is unaffected by elevated CO_2 and lepidopterans feeding there are unaffected. The concern is that pest insects could be stimulated to feed on these reproductive parts when the quality of the remainder of the plant decreases, which in turn would increase the pest status of some insects.

Of course CO_2 has also the direct effect of increasing temperature via the greenhouse effect, which may stimulate feeding activity because of increased metabolic rate in higher temperatures. Studies on the green peach aphid, *Myzus persicae,* a pest of many crops, suggest that elevated temperature increases aphid population growth rate and thus the likelihood that aphids will become more important pests in the future. In this case, both elevated CO_2 and elevated temperature increased aphid densities in experiments. Since, however, very few experiments have examined both CO_2 concentration and temperature in factorial experiments, the generality of the aphid results is unknown. It is also possible that the effects of elevated CO_2 and elevated temperature could cancel each other out for other insect species, especially leaf chewers.

EFFECTS OF CHANGED PLANT CHEMISTRY ON INSECT DENSITIES AND MORTALITIES

There has been relatively little research into how CO_2-mediated changes in plant chemistry affect insect densities and mortalities. This is because most plant–insect work has been done in laboratory conditions, where insects are fed foliage grown in elevated or ambient CO_2 and insect weight gains, losses, and digestibility coefficients are measured. To predict the effects of elevated CO_2 on insect densities, a

population must be established on CO_2-treated foliage. However, in the few cases where insects have been reared from first instars through to pupae and adults, a significant decrease has been found in resultant population sizes in over 30% of the cases. This is usually because nutritionally inadequate foliage kills the immature insects. In some species (e.g., leafminers), we can get a good estimate of host-plant-induced mortality. Here, larvae that die from nutritional inadequacy are entombed with the leaf and can be counted, permitting an accurate assessment of deaths induced by the host plant.

To study the effects of elevated CO_2 on the interactions of insect herbivores with their natural enemies, such as predators and parasites, whole communities containing insect herbivores and their predators, parasites, and diseases are exposed to elevated CO_2. Such community-wide exposure has proved to be very difficult to achieve in the laboratory. Only where whole communities of plants and insects are exposed to elevated CO_2 in the field is it possible to fully address the effects of CO_2 on natural enemies. Experiments like this are very costly to do because of the huge quantities of CO_2 needed to arrive at a large enough increase in CO_2 under field conditions. However, it is widely thought that the net result of increased plant consumption and slower growth by herbivorous insects in elevated CO_2 is likely to result in increased exposure to natural enemies. For example, consumption of additional foliage increases the probability of ingestion of viruses or pathogenic bacteria, such as *Bacillus thuringiensis,* which can cause death. Once again, leafmining insects are a valuable study organism with which to examine the effects of elevated CO_2 on attack rate by natural enemies. This is because the leafmines themselves leave a permanent record of the fate of the insect inside. Parasite larvae can often be found within a mine, or the emerging adult parasitoids leave characteristic small shotgun-like holes in the upper mine surface. A recent study by the author, at Kennedy Space Center, was able to examine attack rates of leafminers by parasitic Hymenoptera in field chambers under conditions of ambient and elevated CO_2. The open-topped chambers contained the full complement of herbivores and their natural enemies on naturally occurring oak vegetation. Leafminer density was reduced inside the chambers, and leaf nitrogen content was reduced. The leafminers died more frequently inside the mines in elevated CO_2 and the mine area was bigger, indicating that larval leafminers had to eat more. Attack rate by natural enemies, particularly parasitoids, was significantly increased inside the chambers in which CO_2 was elevated. Perhaps the leafminers had created bigger, more obvious mines. Alternatively, their developmental time might have been slower in elevated CO_2, exposing them to natural enemies for a longer time, or they might have been physiologically less well able to resist attack.

In other systems, aphids known to produce alarm pheromones show a reduced capacity to do so under elevated CO_2. Once again, the result is an increased susceptibility to natural enemy attack. Finally, increased global temperatures are also likely to increase parasite and predator abundance as a result of increased population growth rates. This in turn could also lead to higher insect herbivore mortalities.

EFFECTS OF TEMPERATURE CHANGES ON INSECT DISTRIBUTION PATTERNS

Greenhouse gases are likely to change insect distribution patterns both directly, via increases in temperature and rainfall, and indirectly, via changes in the distribution of host plants. Recent research on a sample of 35 nonmigratory European butterflies showed that 63% had ranges that shifted to the north by 35 to 240 km during the 20th century, while only 3% shifted to the south. Thus for many insects, global warming has already changed range boundaries. The data appear to be robust because for most of these species, northward shifts have been shown in more than one country. Furthermore, the data appear to be robust across families, with many members of the Lycaenidae, Nymphalinae, Satyrinae, and Hesperiidae showing such range shifts. The northward shifts of the butterflies are of the same magnitude as the shift in climatic isotherms, which have moved about 120 km north as Europe has warmed by about 0.8°C.

Changes in rainfall, likely to have at least as big an impact as rising temperatures, have not been much studied. Global rainfall patterns clearly will change as a result of changes in global temperature, with many coastal areas becoming wetter and many interior continental areas becoming drier. This set of changes will affect the distribution of host plants and the insects that live on them. In addition, rainfall changes can directly affect the hatching of immatures from eggs laid in the soil, including eggs of many species of locust. Increased soil moisture increases the likelihood of locust outbreaks because it increases hatching and stimulates growth of host plants on which the locusts feed. The threat of locust plagues in new areas of the globe is therefore very real.

Margaret Davis, a paleobotanist from the University of Minnesota, showed that in the event of a CO_2 doubling, beech trees, presently distributed throughout the eastern United States and southeastern Canada, would die back in all areas except northern Maine, northern New Brunswick, and southern Quebec. Of course favorable new locations would develop in central Quebec, but the trees would take a long time to colonize such areas. Presumably the animals that feed on beech trees, including insect herbivores, would suffer a severe range contraction too, though this has not yet been studied.

EFFECT OF GREENHOUSE GASES ON SOIL INSECTS

It is doubtful that soil-inhabiting insects will respond directly to increased levels of CO_2 because of existing high

concentrations in the soil. However, there are many likely indirect effects of CO_2 on soil insects. Light interception by a larger canopy may lower soil temperature and moisture. The most important change, though, is likely to be increased litterfall. Soil organic matter is likely to accumulate, rendering grasslands and forests net sinks of carbon under conditions of elevated CO_2. However, before senescence, most leaf nitrogen is reabsorbed by the plant, so that whereas living leaves in elevated CO_2 generally have a lower nitrogen content than leaves in ambient CO_2 conditions, litter quality remains unchanged. However, the increased volume of litter is likely to increase the number of litter-decomposing insects there.

In addition, increased root production may benefit root-feeding insects. Total numbers of Collembola per kilogram of soil have been shown to be significantly higher in experimental laboratory-based mesocosms where CO_2 levels were 60% above ambient. Species composition of Collembola also changed. Part of these increases in Collembola may be due to changes in abundance of mycorrhizal and nonmycorrhizal fungi on which they feed.

AQUATIC COMMUNITIES

Aquatic insect communities are unlikely to be directly influenced by increased CO_2 as much as terrestrial systems are. However, the increased litterfall associated with forest productivity is likely to increase allochthonous (i.e., leaf fall from riparian zones) litter input into forest streams and lakes. Such increased litter input is likely to increase stream insect populations. Litter quality itself, because it does not generally differ between ambient and elevated CO_2 treatments, is unlikely to affect aquatic decomposer communities. This has been verified by adding litter from ambient and elevated CO_2 to laboratory microcosms (simulated treeholes) and examining effects on eastern treehole mosquitoes, *Aedes triseriatus.* No differences in mosquito development time or survival were found. However, the elevated water temperatures and precipitation may increase the abundance of disease vectors such as mosquitoes. On the other hand, some cold water species may be reduced in abundance.

ELEVATED CO_2 AND DISEASE VECTORS

One of the main concerns voiced about global warming is that the delicate balance between diseases, their vectors, and humans might be upset as tropical climates that are so hospitable to spawning and spreading diseases move poleward. The spread of infectious diseases is controlled by the range of their vectors—mosquitoes and other insects. Increases in temperatures mean increases in the activity and ranges of these vectors.

Data on recent trends support this observation. An increase of one degree Celsius in the average temperature in Rwanda in 1987 was accompanied by a 337% rise in the incidence of malaria that year as mosquitoes moved into mountainous areas they had not previously inhabited. Also, *Aedes aegypti,* a mosquito that carries dengue and yellow fever, has extended its range high into the mountain areas of such diverse areas as Colombia, India, and Kenya. Although global warming is expected to deliver its most deadly punch in the tropical areas of the world, where over 500-million people are affected (and 2.7 million die), the United States is not immune. A computer model by a Dutch public health team proposed that an average global temperature increase of 3°C in the next century could result in 50 to 80 million new cases of malaria each year. In the United States, public health facilities are likely to keep new incidences of disease in humans to a minimum, because of vaccinations. But disease outbreak in wildlife, which is not vaccinated, could be more severe.

See Also the Following Articles

Aquatic Habitats • Growth, Individual • Malaria • Pollution • Temperature, Effects on Development and Growth

Further Reading

Bezemer, T. M. and Jones, T. H. (1998). Plant–insect herbivore interactions in elevated and atmospheric CO_2: Quantitative analysis and guild effects. *Olkos* **82,** 212–222.

Coviella, C. E., and Trumble, J. T. (1999). Effects of elevated atmospheric carbon dioxide on insect–plant interactions. *Conserv. Biol.* **13,** 700–712.

Drake, B., Gonzalez-Meler, M., and Long, S. P. (1997). More efficient plants: A consequence of rising atmospheric CO_2. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48,** 607–637.

Houghton, J. T., Meira Filko, L. G., Callander, B. A., Harris, M., Kattenburg, A., and Maskell, K. (1995). "Climate Change 1995. Science of Climate Change." Cambridge University Press, New York.

Keeling, C. D., and Whorf, T. P. (2000). Atmospheric CO_2 records from sites in the SIO via sampling network. *In* "Trends: A Compendium of Data on Global Change." Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN.

Parmessan, C., Ryrholm, N., Stefanesu, C., Hill, J. K., Thomas, C. P., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tammaru, T., Tennett, W. J., Thomas, J. A., and Warren, M. (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399,** 579–583.

Stiling, P. (2002). "Ecology: Theories and Applications." 4th ed. Prentice Hall, Upper Saddle River, NJ.

Stiling, P., Rossi, A. M., Hungate, B., Dijkstra, P. D., Hinkle, C. R., Knott, W. M., and Drake, B. (1999). Decreased leaf-miner abundance in elevated CO_2: Reduced leaf quality and increased parasitoid attack. *Ecol. Appl.* **9,** 240–244.

Growth, Individual

Martin B. Berg
Loyola University Chicago

Richard W. Merritt
Michigan State University

The growth of individual insects proceeds in a progressive manner throughout the immature period of development,

although the rate of growth can vary depending on a variety of factors such as molting frequency, temperature, and nutrition. Growth can be measured as an increase in biomass or body size, although biomass can be more variable than body size owing to differences in food and water intake. Growth usually is distinguished from development in that the latter refers to the various morphological and physiological changes that occur throughout the life span of an insect as it progresses toward maturation.

HOW INSECTS GROW

The limited ability of the rigid exoskeleton of insects to expand imposes a considerable constraint on individual growth resulting in the necessity to shed (i.e., to molt) the old exoskeleton, a process termed ecdysis, for growth to continue. Some insects (e.g., collembolans, diplurans, and thysanurans) exhibit indeterminate growth and continue to molt even after reaching the adult stage, although little if any increase in biomass occurs. In contrast, the majority of insects exhibit determinate growth in which both growth and molting cease upon reaching the reproductively mature last instar (i.e., the adult).

The pattern of individual growth differs depending on whether growth is measured as an increase in biomass or as an increase in body size. When measured as an increase in biomass, individual growth occurs between molts and is more or less continuous in most insects, although decreases in biomass often occur at the time of molting. When measured as an increase in body size, however, individual growth is largely dependent on the amount of sclerotization of the insect or of a particular body part. Membranous body parts, and those insects that are not highly sclerotized, can continuously increase in size between molts as folds in the cuticle expand, whereas insects or body parts that are more highly sclerotized increase in size immediately following each molt and exhibit a more discontinuous pattern.

Different body parts of insects may exhibit either isometric or allometric growth compared with the body as a whole. Isometric growth occurs when body parts grow at the same rate as the body as a whole, i.e., body length. Allometric growth occurs when body parts grow at rates different from that of the body and can be expressed as a power function of the form $x = ky^a$, where x is the dimension of the whole, y is the dimension of the part, a is the growth coefficient, and k is a constant. Thus, a straight line results from a log–log plot of body part size vs body length (Fig. 1).

INCREASES IN BIOMASS

Biomass increase of insects during development can be appreciable, with immature final instars weighing as much as 1000 or even 10,000 times greater than first instars. Changes in biomass can be expressed as either an absolute (weight/time) or a relative (weight/weight/time) increase in biomass. Absolute growth usually is greatest in later instars and for the tobacco hornworm, *Manduca sexta,* larval weight increases as much as 90% (10-fold) in the final two instars. Similarly, 90% of the growth of larval *Paratendipes albimanus* (Diptera: Chironomidae) occurs in the final 10% of the life cycle. In contrast, relative growth rates normally decrease in later stages of development as the insect increases in size.

INCREASES IN BODY SIZE AND DYAR'S LAW

In 1890, H. G. Dyar noted that the head capsule widths of lepidopteran larvae followed a geometric progression in growth. During development of an immature insect, increases in body size occur in discrete steps, with highly sclerotized body parts exhibiting predictable and regular increases by a relatively constant factor, subsequently known as Dyar's Law (or Rule). Although initially based on observations of lepidopteran larvae, Dyar's Law has been applied to immature insects in general and refers to the geometric progression in the size of sclerotized structures that is constant throughout development. Dyar's Law has been widely used in entomological studies to discern instars of immature insects and also has been used to predict the size of instars missing from samples. The ability to distinguish instars is crucial to accurately describe insect life histories and growth patterns and is widely used in secondary production studies.

Because membranous portions of the less sclerotized cuticle (e.g., intersegmental membranes of the abdomen) allow the body to grow more or less continuously, overall body size is not considered a good indicator of instar. In addition, the increase in body size at each molt varies with different species and the growth of various body parts of many insects may differ from the growth rate of the body as a whole (i.e., allometric growth). In contrast, the rigid exoskeleton of

FIGURE 1 Allometric relationships between body part length and body length for *P. albimanus* (Diptera: Chironomidae) (Data from Ward and Cummins, 1978.)

immature insects prevents expansion and results in discontinuous growth of highly sclerotized cuticular parts (e.g., head capsule, legs). Thus, these structures increase in size incrementally in a stepwise manner following a molt. It is these distinct increases in the size of sclerotized structures that allows for the distinguishing of different instars with little overlap occurring between size classes.

The publication of Dyar's observations was in response to two previous papers that presented contradictory data concerning the number of molts in other species. Dyar studied the number of molts in 39 individuals of 28 species of caterpillars and chose the head capsule as the structure to measure for ease of measurement and because it was not subjected to growth during each stadium as was body length. The taxa chosen by Dyar ranged from 4 to 10 instars. Dyar calculated head-width ratios of successive instars and found that the progression was often nearly constant for a given species (mean = 1.5; range 1.3–1.7), what one would expect from a geometric progression. Dyar then calculated expected head capsule widths of each instar by multiplying the width of the final instar by this ratio and then back-calculating to the first instar. To test the applicability of the ratio, Dyar compared calculated head widths to those observed from reared specimens. Using this approach, it was possible to detect whether some instars had been missed or mismeasured. The most common method for detecting these problems is to plot the logarithm of the head capsule width measurement (or a measurement of another highly sclerotized structure) against the appropriate instar (Fig. 2). Conformity to Dyar's Law results in a straight line the slope of which is constant for a given species. According to Dyar's Law, deviations from a straight line indicate potentially missed instars or errors in measurement.

Although Dyar's Law has been widely used in entomological studies, the progression in the size of sclerotized body parts is not always constant and can be influenced by abiotic and biotic factors such as temperature and food. In addition, apparent contradictions to Dyar's Law occur when two requisite conditions are not met: (1) the number of instars is constant and (2) head capsule growth occurs only at ecdysis. Despite these constraints, approximately 80% of the entomological studies published from 1980 to 2000 that have examined the validity of Dyar's observations provided support for his law.

There has been some disagreement as to whether Dyar should be credited with the findings of geometric progression in growth. In a paper published 4 years prior to Dyar's article, Brooks reported that total larval length of a species of crustacean stomatopod (Stomatopoda: Squillidae) sequentially increased in size by a factor of 1.25 at each molt. Brooks also noted, as did Dyar, that this relationship could be used to determine whether larval stages were missing from the series. Thus, Dyar's Law may occasionally also be referred to as Brooks' Law (or Rule) in the literature. It is likely that entomologists were unaware of Brooks' observations because they were documented in a specialized publication on stomatopods.

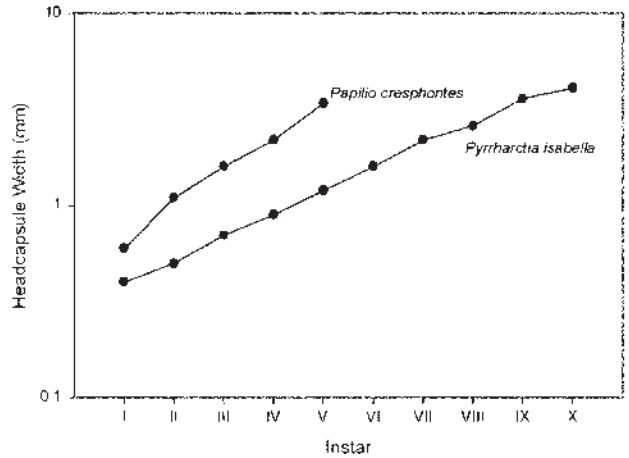

FIGURE 2 Conformity of head capsule width to Dyar's Law for giant swallowtail (*Pa. cresphontes,* Lepidoptera: Papilionidae) and banded woollybear (*Py. isabella,* Lepidoptera: Arctiidae) (Data from Dyar, 1890.)

EFFECTS OF TEMPERATURE ON INDIVIDUAL GROWTH

Because insects are ectotherms, temperature can have a profound impact on individual growth. In general, insect growth is correlated with environmental temperature; however, the strength of this relationship may be species- and habitat-specific. For example, many terrestrial insects exhibit little to no growth at low temperatures (0–4°C) because of either reduced feeding rates or overall low metabolic activity. Aquatic insects, in contrast, particularly those whose evolutionary ancestral habitat was in cold streams or lakes, can exhibit high growth rates during winter, assuming the presence of an adequate food supply. Although feeding rates of these insects may be low at reduced temperatures, basal metabolic needs also are low and result in the ability of the insect to direct more of the energy derived from ingestion to growth. The interactions between food quantity and quality and temperature are complex. In some streams, water temperature can influence the growth of microbial populations attached to detrital particles that are ingested by filter-feeding aquatic insects, and in turn, this can enhance their growth rate. This type of indirect control on aquatic insect growth and survivorship is difficult to separate from the direct effects of temperature on insect metabolism.

Elevated environmental temperatures also can result in either high or low larval growth rates because of the influence of temperature on larval ingestion, digestion, and development time. In general, larval ingestion and digestion are positively correlated with temperature and should result in increased growth rates at higher temperatures. The effect of temperature on growth rate, however, is confounded by an inverse relationship between ingestion rates and assimilation rates. In contrast, high temperatures can accelerate developmental time, resulting in the inability of larvae to maximize absolute growth rates.

EFFECTS OF NUTRITION ON INDIVIDUAL GROWTH

Individual growth of immature insects is strongly influenced by food availability, feeding selectivity, and food quality. Insect growth is often directly related to food availability, in that larval growth rates are highest in the presence of an abundant food supply. Because availability of food resources can vary temporally, growth responses also can be expected to vary throughout the year. For example, several species of stream chironomids (Diptera: Chironomidae) that feed primarily, but not exclusively, on attached algae, exhibit periods of maximal larval growth that coincide with times of the year when instream algal production is highest.

Feeding selectivity and food quality, however, also influence the relationship between growth and food availability. Even in the presence of an apparently abundant food supply, larval growth rates may be reduced if that food resource is not preferred or if it is of low quality or lacking essential nutrients. For example, the leaf-shredding crane fly, *Tipula abdominalis,* shows a strong preference for hickory, maple, and American chestnut leaves and a low preference for American beech, white oak, and red oak leaves. Larval growth rates are highest on the more preferred leaf types and lowest on the less preferred leaves. The higher growth rates on more preferred leaves are not due to higher food conversion efficiencies, but rather result from increased consumption rates due to a more palatable food source.

Although food availability may not be a limiting factor for growth in some insects, food quality may impose a substantial constraint to larval growth. Differences in food quality also can affect larval growth rates and the ability to complete development and reproduce. Absolute and relative growth rates of the caddisfly *Clistoronia magnifica* (Trichoptera: Limnephilidae) reared on diets differing in quality were significantly higher on a diet rich in triglycerides (conditioned alder leaves plus whole wheat grains) compared with diets of conditioned alder leaves, conditioned alder leaves with a fatty acid mixture, conditioned alder leaves plus hyphomycete fungi, or hyphomycete fungi alone. In addition, only larvae fed the high-triglyceride diet successfully completed development and reproduced. Some insects, however, even in the absence of higher quality food, can maintain a relatively uniform growth rate throughout development by increasing ingestion rates of lower quality food.

See Also the Following Articles

Body Size • Development, Hormonal Control of • Exoskeleton • Feeding Behavior • Nutrition • Temperature, Effects on Development and Growth

Further Reading

Berg, M. B., and Hellenthal, R. A. (1992). Life histories and growth of lotic chironomids (Diptera: Chironomidae). *Ann. Entomol. Soc. Am.* **85,** 578–589.

Chapman, R. F. (1998). "The Insects: Structure and Function." Cambridge University Press, Cambridge, U.K.

Crosby, T. K. (1972). Dyar's rule predated by Brook's rule. *N. Z. Entomol.* **5,** 175–176.

Dyar, H. G. (1890). The number of molts of lepidopterous larvae. *Psyche* **5,** 420–422.

Epstein, M. E., and Henson, P. M. (1992). Digging for Dyar. *Am. Entomol.* **34,** 148–169.

Gillott, C. (1980). "Entomology." Plenum Press, New York.

Richards, O. W., and Davies, R. G. (1979). "Structure, Physiology, and Development," Vol. I of "Imm's General Textbook of Entomology." Chapman & Hall, London.

Sweeney, B. W. (1984). Factors influencing life-history patterns of aquatic insects. *In* "The Ecology of Aquatic Insects" (V. H. Resh and D. M. Rosenberg, eds.), pp. 56–100. Praeger, New York.

Ward, G. M., and Cummins, K. W. (1978). Life history and growth pattern of *Paratendipes albimanus* in a Michigan headwater stream. *Ann. Entomol. Soc. Am.* **71,** 272–284.

Wigglesworth, V. B. (1965). "The Principles of Insect Physiology." Methuen, London.

Grylloblattodea

(Rock Crawlers, Ice Crawlers)

D. C. F. Rentz
California Academy of Sciences

The Grylloblattodea, or ice crawlers, are a small group of soft-bodied, apterous, terrestrial and termitelike insects confined to the Northern Hemisphere. They occur under rocks in forest leaf litter or above the treeline in the high mountains. In parts of the United States, Canada, and China they are found under rocks in soil at or below freezing temperature. In Japan and Korea, they are active in midsummer in the deep leaf litter of mixed conifer and deciduous forests, where daily ambient temperatures approach 30°C. In Japan and Korea some have been found in caves; in the western United States, they are known from subterranean lava tubes. Grylloblattids are slender, depressed insects covered with fine hairs and having reduced eyes. Adults range from 2 to 3.5 cm in length. Ice crawlers appear to be primarily nocturnal. They are considered to be some of the most primitive of orthopteroid insects and have been thought to be related to cockroaches (Blattodea) and stick insects (Phasmida).

CLASSIFICATION AND FEATURES OF THE ORDER

The Grylloblattodea comprise a single family, Grylloblattidae, which includes only four genera. *Grylloblatta* occurs in North America and Canada, where at least 11 species are known. *Grylloblattina* is known from a few species from Siberia. *Grylloblattella* has species known from Korea, Japan, and Siberia. *Galloisiana* contains at least 10 species from Japan. The group, however, is undoubtedly larger, since new species are discovered every few years.

FIGURE 1 Copulating pair of grylloblattids, *G. nipponensis:* the female is the lower individual (note ovipositor); the male has the eversible sac of the left phallomere exposed. [From Nagashima *et al. In* Ando, H. (ed.). (1982). "Biology of the Notoptera," p. 48, Kashiyo-Insatu, Nagano, Japan.]

Grylloblattids are cryptic, ground-dwelling insects that prefer wet habitats and cool temperatures. They shun light and occur under stones or in dense leaf litter. Species of *Grylloblatta* that live under rocks under or near snowbanks emerge after dark and feed as scavengers or predators on dead or dying insects that have been blown onto the snow from lower elevations. During winter, the species probably occupy the airspace between the ground and overlying snowpack, where they remain active at temperatures of 0°C. Massive fat bodies build up prior to winter; during winter, the insects may feed on decaying plant material. Korean species live under debris on the floor of caves at only 200 m altitude and apparently never venture forth from the cave habitat. The most widespread species in Japan, *Galloisiana nipponensis,* is found at elevations ranging from 300 to 3000 m, where the insects live under stones and in the leaf litter of thick, mixed coniferous and hardwood forests. These species consume both insects and plant material day and night.

The resemblance of grylloblattids to Dermaptera has been shown to be superficial and is associated only with the fact that both have a projecting head. The head, however, is typical of orthopteroid insects. The antennae are elongate and thin. The mouthparts are structured like those of a predator. Ocelli are absent, and the eye comprises fewer ommatidia in young instars than in adults. The legs are simple, slender, and not suited for jumping. The abdomen comprises 11 segments, with the cerci long and flexible and the male genitalia asymmetrical. The ovipositor comprises three pairs of slender, tapering, partly free valves.

Although grylloblattids are normally considered to be cool-adapted insects, they cannot withstand temperatures much below 0°C. At –5.5 to 8.0°C they stress. Contrary to the popular belief that they can withstand very low temperatures, they can be killed by ice formation within the body as a result of their low levels of glycerol, sorbitol, or erythrol.

Copulation has been observed in a few species. In *G. nipponensis,* the female is chased and seized by the male. The resulting copulation can last from 30 min to 4 h. Males always assume a position on the right side of the female as a response to the male's asymmetrical genitalia. Oviposition occurs 10 to 50 days after copulation. Females lay eggs with the elongate ovipositor in wood or under stones and decaying plant material. None have been found in moss. Oviposition for each egg takes about 3 min, and females lay 5 or 6 eggs per day to a total of about 30 eggs. A captive female laid 145 eggs in her lifetime. The large, black eggs, develop over periods of from 5 months to 3 years.

See Also the Following Articles

Blattodea • Orthoptera • Phasmida

Further Reading

Ando, H. (ed.). (1982). "Biology of the Notoptera." Kashiyo-Insatsu, Nagano, Japan.

Storozhenko, S. (1989). A review of the family Gryllop(b)lattidae (Insecta). *Articulata* **13,** 167–181.

Gypsy Moth

Joseph S. Elkinton
University of Massachusetts

The gypsy moth, *Lymantria dispar,* is one of the world's most damaging defoliators of hardwood forest trees. It is native to Europe and Asia. It was introduced from Europe to North America near Boston, Massachusetts, in 1869 and has been spreading slowly south and west ever since. A large body of research has focused on the biology, management, and population dynamics of this species.

GEOGRAPHIC RANGE AND SPREAD

Gypsy moth occurs throughout much of the Northern Hemisphere. Its native range stretches from Japan, China, and Siberia across Russia to western Europe and as far south as the Atlas Mountains of North Africa. In North America, gypsy moth has spread over much of the eastern United States and Canada. Currently, the leading edge of the infestation stretches from North Carolina to Wisconsin and adjacent regions of Ontario. Because female gypsy moths from Europe have wings but do not fly, the rate of spread of this insect has been extremely slow. The spread occurs when newly hatched larvae spin down on silken threads and are blown in the wind. Most of this dispersal is less than 50 m, although some larvae are carried by wind currents for greater distances. Natural spread has been augmented by inadvertent human transport of egg masses laid on vehicles or other backyard objects, often to locations well outside the region infested by gypsy moth in the northeastern United States. Such new infestations have been eradicated at many sites

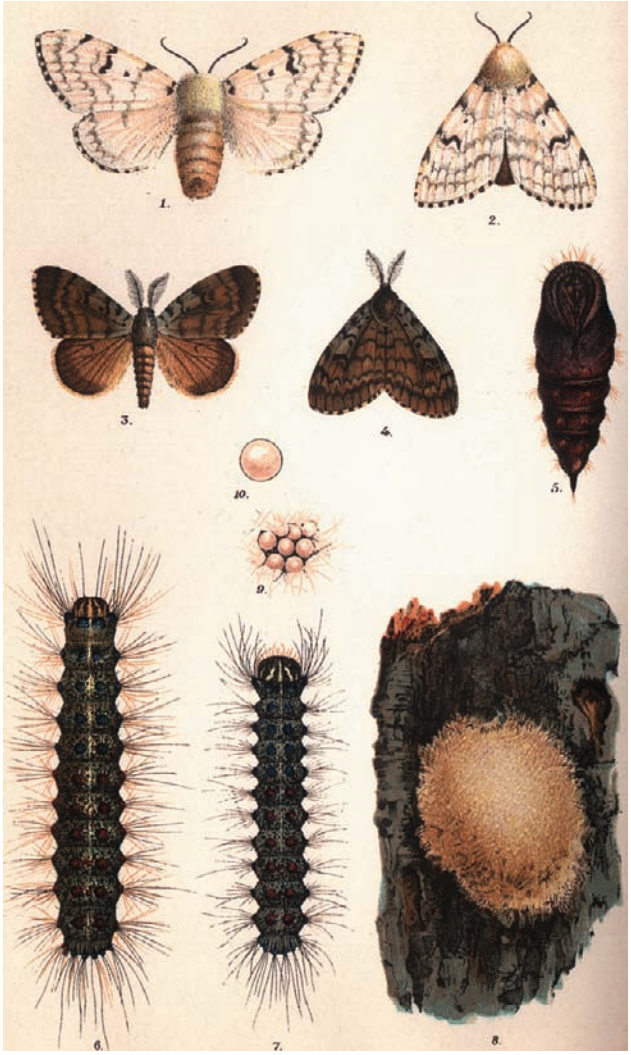

FIGURE 1 Life stages of gypsy moth. (1,2) Adult female; (3,4) adult male, (5) pupae; (6,7) larvae; (8) egg mass; (9,10) individual eggs. (Reproduced from J. Bridgham, 1896, 43rd Annual Report, Massachusetts State Board of Agriculture.)

throughout North America. In recent years, attention has shifted to new introductions of gypsy moth from Asia, particularly the Russian Far East. Adult female gypsy moths from these regions are able to fly. Consequently, the rate of spread would be much faster, if they became established. These introductions have occurred at several locations in North America, mainly in the Pacific Northwest, but thus far, none have become established.

LIFE HISTORY AND HOST TREES

The gypsy moth female lays a single egg mass (Fig. 1), usually on the stems of trees, and she covers the eggs with her body hairs. The egg mass typically contains from 100 to 600 eggs. The eggs are laid in midsummer, but overwinter in this stage. Larvae developing within the egg enter diapause and hatch the following spring at the time of host-tree budbreak. Emerging larvae climb to the tops of trees, where many of them spin down on silken threads and are borne away by the wind. If larvae land on an acceptable host tree, they begin feeding. They develop through five instars for males or, frequently, six for females throughout May and June. Beginning in the fourth instar, larvae seek resting locations during daylight hours, either in the forest litter at the base of trees or under bark flaps on tree stems. They usually pupate in these same locations. The adults emerge after about 12 days in the pupal stage. Soon after eclosion, the female releases a sex pheromone from a gland on the tip of her abdomen. Males locate females by flying upwind when they detect the pheromone. After mating, the female lays her egg mass, often just a few centimeters from where she eclosed, and dies soon thereafter. There is one generation per year.

Gypsy moth larvae feed on a wide range of tree species. Favored tree species include oaks (*Quercus* spp.), aspen (*Populus* spp.), and, in Japan, Japanese larch, *Larix leptolepis.* Gypsy moth outbreaks occur in forests that are dominated by these species. Gypsy moth will feed on many other tree species, such as maple (*Acer* spp.) and many conifers, but significant damage to these trees usually occurs only in gypsy moth outbreaks, when more favored hosts have already been defoliated. If defoliation is complete, most deciduous hardwood trees will put out a new set of leaves. Most trees will survive one defoliation, but if outbreaks persist for several years in a row, a significant proportion of the trees may die. Trees that survive defoliation suffer growth loss in subsequent years.

DISEASES OF GYPSY MOTH

As with most insects, gypsy moths are host to a suite of natural enemies and these play a pivotal role in the dynamics of the gypsy moth populations. There are two major diseases: a nuclear polyhedrosis virus and a fungal pathogen. The virus, *LdMNPV,* causes epizootics that are largely responsible for the collapse of gypsy moth outbreaks. Similar viruses terminate the outbreaks of many defoliating Lepidoptera. High mortality from these viral diseases occurs only in dense populations, because transmission of the virus takes place when larvae feed on leaves contaminated by cadavers of larvae that have previously died from the disease. Encounters with cadavers are only likely in dense populations. Transmission of *LdMNPV* from one generation of gypsy moths to the next occurs primarily by way of external contamination of the egg mass; larvae become infected as they emerge from the mass in the spring. It is not entirely clear how the virus persists at low density, but it does survive in the forest litter for several decades.

The fungal pathogen *Entomophaga maimaiga* was, until recently, known only in the Far East, especially Japan. In 1989, a dramatic epizootic of *E. maimaiga* occurred throughout the northeastern United States from Pennsylvania to Maine. In

subsequent years, the fungus spread across the mid-Atlantic states and was introduced intentionally by researchers to Virginia and Michigan. It is now established throughout the region infested by gypsy moth in North America. *E. maimaiga* produces two kinds of spores: conidia and resting spores. The conidia are released from cadavers and are carried by wind currents to uninfected larvae, which they infect by penetrating the cuticle; these conidia are responsible for the rapid spread of *E. maimaiga* in North America. Late instars produce resting spores that overwinter in the forest litter, where they persist for up to 10 years before germinating to infect new gypsy moths. A. Hajek and colleagues analyzed the DNA of *E. maimaiga* and showed that the pathogen in North America is identical to *E. maimaiga* in Japan. How *E. maimaiga* was introduced into North America is unknown. Since 1989, it has continued to cause high levels of mortality in gypsy moth populations, particularly in years with high rainfall in May and June. A key difference from *LdMNPV* is that *E. maimaiga* causes substantial mortality in low-density as well as in high-density populations of gypsy moth. This means that *E. maimaiga* can prevent outbreaks from occurring, whereas *LdMNPV* can only cause the collapse of outbreak populations.

PARASITOIDS

As with most insects, various parasitoid species attack the different life stages of gypsy moth. In North America, efforts to introduce parasitoids of gypsy moth from Europe and Asia began around 1905, and 10 species have been established. The egg parasitoid *Ooencyrtus kuvanae* (Encyrtidae) from Japan is frequently observed on gypsy moth egg masses in late summer and may cause as much as 30% mortality of the eggs. Larval parasitoids include *Cotesia melanoscela* (Braconidae) and the tachinids *Blepharipa pratensis, Compsilura concinnata,* and *Parasetigena silvestris.* The most common pupal parasitoid is *Brachymeria intermedia* (Chalcididae). The impact of these parasitoids on gypsy moth populations remains equivocal. Total mortality caused by parasitoids in North America is typically below 50% and is not consistently density dependent, so that their ability to regulate gypsy moth populations is in doubt. In Europe, on the other hand, parasitism of gypsy moth is often much higher than that observed in North America. European gypsy moths are attacked by several parasitoid species that were never established successfully in North America. It seems likely that parasitoids are responsible for preventing gypsy moth outbreaks, which are rare in western Europe, but more common in central and southern Europe.

PREDATORS

Compared with parasitoids, an even larger community of vertebrate and invertebrate predators feeds on gypsy moth. Very little is known about the impact of most predators, because predation is extremely difficult to measure. Many bird species worldwide feed on gypsy moth, but it is generally believed that most birds dislike the hairy cuticle of gypsy moths and avoid them.

Research groups led by H. Bess in the 1940s and R. Campbell in the 1970s concluded that predation by small mammals, particularly the white-footed mouse, *Peromyscus leucopus,* has a major impact on low-density gypsy moth populations. The mice feed on late instars and pupae, particularly on the forest floor. Both research groups demonstrated an increase in gypsy moth survival in forest plots from which small mammals had been removed or excluded. More recently, J. Elkinton and colleagues showed that predation on gypsy moth pupae was strongly correlated with density of mice and that gypsy moth densities increased when mouse densities declined. The density of mice, in turn, was correlated with the abundance of acorns, which, in oak-dominated forests, are their principal overwintering food. Indeed, there are many studies that link forest-dwelling mice to abundance of acorn crops. Poor acorn crops, which occur on a regional scale and are caused by a variety of weather events, could thus be the ultimate trigger of gypsy moth outbreaks. C. Jones and colleagues provided further experimental proof of these ideas. They removed mice from experimental plots and observed an increase in gypsy moth and they augmented food in other plots and observed an increase in mice.

There are a number of invertebrate predators of gypsy moth. One of these is the introduced ground beetle, *Calosoma sycophanta* (Carabidae). R. Weseloh has shown that this insect becomes quite abundant in outbreak populations of gypsy moth and may cause substantial mortality. Several researchers have documented predation by ants as a significant source of mortality in low-density populations, but it is usually much less than predation by mice.

POPULATION DYNAMICS

In the 1970s R. Campbell developed the first comprehensive theory of gypsy moth dynamics, wherein populations alternate between low-density and high-density phases, each maintained by different factors and sources of mortality. At low density, predators, particularly mice, maintain gypsy moth populations indefinitely at a low-density equilibrium. The concept of an equilibrium implies that predation is density dependent, which means that it increases as gypsy moth density increases until total mortality balances fecundity of gypsy moth and the population density stops growing. The equilibrium is an unstable one, however, because the density-dependent response of most natural enemies is constrained by a variety of factors. For example, most predators or parasitoids have their own natural enemies. These constraints produce a threshold density above which gypsy moth population growth outpaces the mortality caused by natural enemies, and as a result densities of gypsy moth increase rapidly into an outbreak phase. At the much higher

outbreak densities, a different set of natural enemies becomes predominant. These natural enemies, coupled with competition among gypsy moths for available foliage, limit further increases in gypsy moth density. The outbreak population either persists for several generations at high density or collapses back to the low-density phase. The model is particularly appropriate when the principal mortality factors maintaining the low-density equilibrium are generalist predators, such as mice. Unlike specialist natural enemies, whose densities often track those of their hosts, the densities of mice are not determined by gypsy moths. Instead their densities are determined by their overwintering food supply, mainly acorns. Whether this conceptual model is an accurate description of the gypsy moth system remains to be demonstrated.

Gypsy moth outbreaks are spatially synchronized (Fig. 2). Outbreaks tend to occur simultaneously over a large region. A cause of this pattern may be acorn crops, which are heavy or meager over a region in response to regional weather patterns. Indeed, as indicated by P. Moran, populations of many species are synchronized by a variety of weather-related influences. Weather may synchronize gypsy moth populations by exerting a common influence across populations on many of the factors that affect gypsy moth growth and survival. Many gypsy moth researchers believe that there is a 10-year cycle of abundance of gypsy moth embedded in the erratic temporal pattern evident in Fig. 2. Analyses of these data by D. Williams and A. Liebhold provide some support for this view. If these cycles exist, the factors that cause them remain unknown. Analyses by P. Turchin of decade-long records of

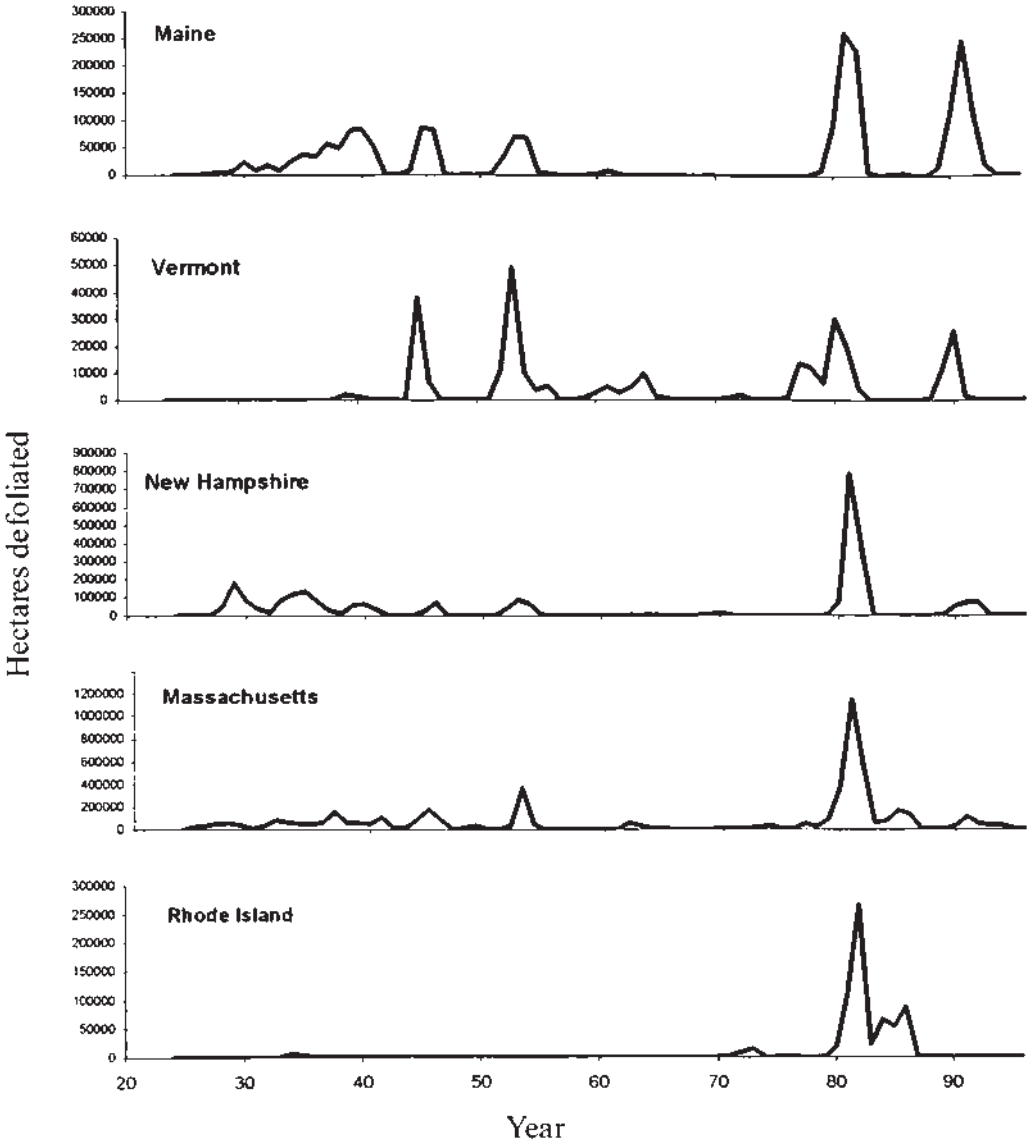

FIGURE 2 Historical record of area defoliated by gypsy moth in each of five states in northeastern United States from 1924 to 1996. (Reproduced from Liebhold *et al.,* 2000, *Pop. Ecol.* **42,** 257–266, © Springer-Verlag, with permission.)

gypsy moth egg-mass density in Yugoslavia provide more convincing support for regular cycles in that region, evidently caused by fluctuations in parasitism.

Adding to the complexity of gypsy moth dynamics in North America is the recent appearance of the fungal pathogen *E. maimaiga.* In the northeastern United States, epizootics of this fungus have occurred nearly every year since 1989, except in years that were extremely dry. We still have much to learn about this agent, but it causes substantial mortality in both low- and high-density populations and appears to have prevented several incipient outbreaks. It appears that incidence of *E. maimaiga,* which is largely determined by rainfall in May and June, is now a prime determinant of whether outbreaks occur.

MANAGEMENT OF GYPSY MOTH

Management of gypsy moth in North America and elsewhere has evolved over time as different tools became available and as public attitudes toward pesticide use have changed. In the 1960s, large areas were sprayed by air with DDT. DDT was banned in the late 1960s and was supplanted by other chemical pesticides, such as carbaryl. In the late 1980s, *Bacillus thuringiensis* (Bt), a bacterial pesticide, became a viable alternative to chemical pesticides and became the material of choice in many regions. The advantage of Bt is that it is more selective than most chemical pesticides; it affects only larval Lepidoptera that feed on Bt-contaminated foliage. In addition, government agencies in the generally infested region in northeastern United States concluded that large-scale application of pesticides against gypsy moth was neither ecologically acceptable nor worth the considerable expense. Most forest trees survive gypsy moth outbreaks, and the outbreak populations soon collapse on their own. Control activities in these regions generally aim at foliage protection on high-value trees, rather than suppression of gypsy moth populations.

See Also the Following Articles

Biological Control • Forest Habitats • Pathogens of Insects • Population Ecology

Further Reading

Campbell, R. W., and Sloane, R. J. (1977). Forest stand responses to defoliation by gypsy moth. *Forest Science Monogr.* No. 19.

Doane, C. C., and McManus, M. L. (eds.) (1981). The gypsy moth: Research toward integrated pest management. USDA Forest Service Tech. Bull. 1584.

Elkinton, J. S., and Liebhold, A. M. (1990). Population dynamics of gypsy moth in North America. *Annu. Rev. Entomol.* **35,** 571–596.

Hajek, A. E. (1999). Pathology and epizootiology of *Entomophaga maimaiga* infections in forest Lepidoptera. *Microbiol. Mol. Biol. Rev.* **63,** 814–835.

Liebhold, A. M., and McManus, M. L. (1999). The evolving use of insecticides in gypsy moth management. *J. Forestry* **97,** 20–23.

Liebhold, A. M., Elkinton, J. S., Williams, D., and Muzika, R.-M. (2000). What causes outbreaks of gypsy moth in North America? *Populat. Ecol.* **42,** 257–266.

Wallner, W. E. (ed.) (1989). Lymantriidae: Comparisons of features of new and old world tussock moths. USDA Forest Service Gen. Tech. Rep. NE-123.

Williams, D. W., Fuester, R. W., Metterhouse, W. W., Balaam, R. J., Bullock, R. H., Chianese, R. J., and Reardon, R. C. (1992). Incidence and ecological relationships of parasitism in larval populations of *Lymantria dispar* (Lepidoptera: Lymantriidae). *Biol. Control* **2,** 35–43.

H

Hearing

Jayne Yack
Carleton University

Ron Hoy
Cornell University

Among all terrestrial animals, only vertebrates and insects are richly endowed with a sense of hearing. By "hearing," we usually mean the ability to detect minute, time-varying changes in air pressure that we familiarly experience as "sound." Under this restricted definition, we can say that audition has evolved in at least seven orders of insects, including all of the major orders except the Hymenoptera (wasps, ants, bees). However, if we were to include under "hearing" the ability to detect sound waves in water and solids, or the displacement of molecules in a sound's near field, then the number of "auditive" insects would expand enormously to include not only the Hymenoptera, but even small orders such as Plecoptera (stone flies) and Isoptera (termites). Initially, we focus on the form and function of tympanal ears, which are organs that are sensitive to sound signals that are propagated through the air or water as fluctuations in pressure and which come to mind when we (humans) use the term "hearing with ears." Following this, we provide some examples of nontympanal hearing organs.

Using sound, vertebrates and insects are often capable of sensing, identifying, and locating their predators, prey, conspecific rivals, and mates by hearing their intentional or unintentional acoustic signals. As might be expected, natural selection has shaped the form and function of hearing organs ("ears") in insects over evolutionary time. In this respect, the ears of insects show much greater diversity than those of vertebrates, for reasons that will be apparent in our discussion. However, it must be emphasized that despite the scope of morphological diversity among insect ears, there is a morphological "bauplan" (structural design) that underlies their great range in behavioral and physiological function.

MORPHOLOGICAL REQUIREMENTS

There is tremendous morphological diversity of insect ears (Fig. 1). The multitude of different ear designs and locations reflects the unique physical and behavioral challenges faced by each insect. Yet despite their many differences, most ears follow a similar morphological plan. Each typically consists of three identifiable substructures: a tympanal membrane, a tracheal air chamber, and a chordotonal sensory organ.

Tympanal Membrane

The tympanal membrane (eardrum) is a thinned region of exoskeleton, typically supported by a chitinous ring and stretched across an enlarged air-filled cavity. Sound impinges upon the membrane, setting it and its associated nerve cells into motion. The thickness of the membrane can vary from 1 to over 100 μm. The ultra-sound-sensitive ears of many nocturnal Lepidoptera (butterflies and moths), for example, are so thin that they are transparent. Such fragile membranes are typically protected within body cavities or by external flaps of cuticle. In contrast, the thicker, opaque tympanal membranes of some diurnal butterflies or grasshoppers are conspicuously positioned on the outer surface of the body.

Tracheal Air Chamber

The internal face of the eardrum backs onto an enlarged air-filled chamber, which forms part of the tracheal respiratory system. In some ears, the air chambers are connected directly to other sound input sources (spiracles or contralateral ears) or resonating chambers via the tracheal system. In a few rare cases [e.g., green lacewings (Chrysopidae) and some water bugs (Corixidae)] the tympanic chambers are largely fluid filled.

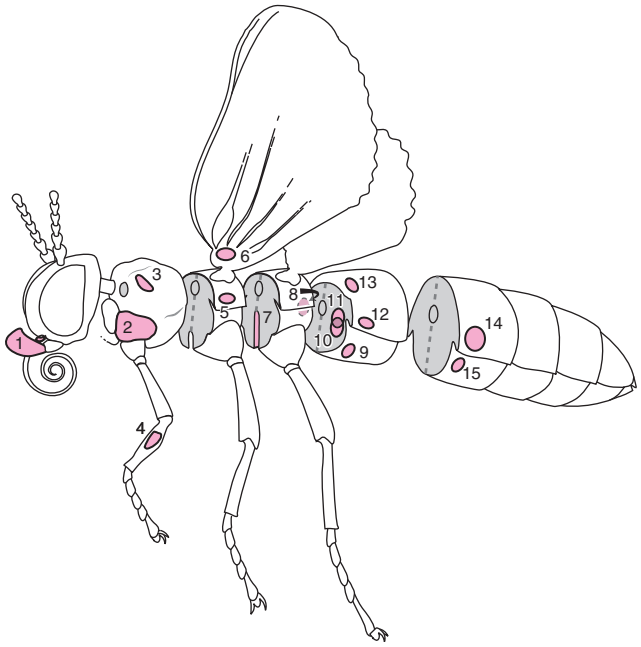

FIGURE 1 A schematic drawing of a generalized insect showing 15 body locations where tympanal ears have been identified. Each number represents a position on the body where an ear has evolved independently in one or more taxa, although all species within a taxonomic division do not necessarily possess ears. (1) Lepidoptera: Sphingidae (Choerocampini, Acherontini). Location: palp–pilifer region. (2) Diptera: Sarcophagidae, Tachinidae. Location: ventral inflation of prosternum, between coxa. (3) Coleoptera: Scarabidae, Dynastinae. Location: dorsolateral region of prosternum. (4) Orthoptera: Ensifera: Gryllidae, Tettigoniidae. Location: tibia of foreleg. (5) Heteroptera: Hydrocorisae (water boatmen). Location: lateral mesothorax, ventral to wing base. (6) Lepidoptera: Papilionoidea, Hedyloidea; Neuroptera: Chrysopidae. Location: base of ventral forewing. (7) Dictyoptera: Mantodea. Location: within a deep groove between the metathoracic legs. (8) Lepidoptera: Noctuoidea. Location: within a cavity on the posterior metathorax. (9) Lepidoptera: Pyraloidea. Location: within a cavity on ventral surface of first abdominal segment. (10) Lepidoptera: Geometridae. Location: within a cavity on anterior side of first abdominal segment. (11) Lepidoptera: Drepanidae. Location: internalized tympanal membrane located between two air-filled chambers on first abdominal segment. (12) Orthoptera: Acrididae. Location: lateral surface of first abdominal segment. (13) Coleoptera: Cicindelidae. Location: dorsal surface of first abdominal segment, beneath the elytra. (14) Homoptera: Cicadidae. Location: within cavity on lateral second abdominal segment. (15) Lepidoptera: Uraniidae. Location: within cavity at the anterior (females) or posterior (males) end of the second abdominal segment. (Illustration by M. Nelson.)

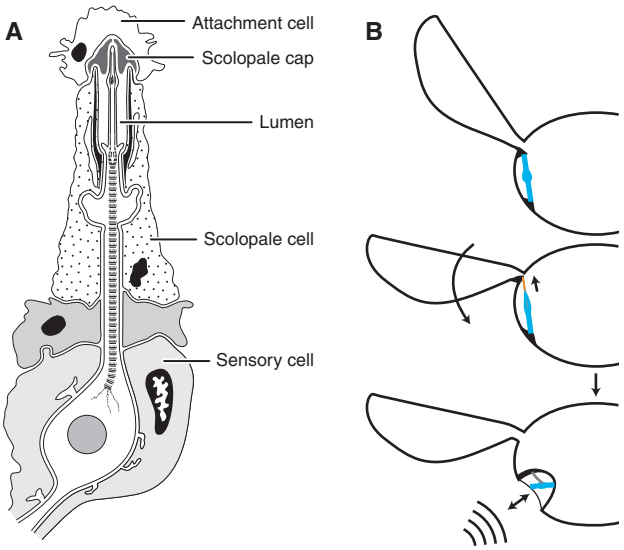

FIGURE 2 Insect tympanal sensory receptors. (A) A typical tympanal scolopidial organ, consisting of three cell types. The dendrite of a bipolar sensory neuron projects into a fluid-filled space (lumen) formed by the walls of a enveloping scolopale cell. The distal tip of the dendrite inserts into the scolopale cap, an extracellular secretion of the scolopale cell. The attachment cell connects the sensory neuron and scolopale cell to the tympanal membrane, either directly or indirectly via a tracheal air sac. A chordotonal organ may have from one to several thousand scolopidia. (B) A schematic diagram depicting the hypothetical transition from a wing-hinge proprioceptive chordotonal organ to a tympanal hearing chordotonal organ. The top two images show a chordotonal organ functioning as a proprioceptor monitoring wing movements. At the bottom the chordotonal organ has been mechanically isolated within a rigid tympanal cavity and attaches to a thinned region of cuticle (the tympanic membrane) that detects sounds. (A was modified, with permission, from E. G. Gray, 1960, The fine structure of the insect ear, *Philos. Trans. R. Soc. B* **243,** 75–94. Illustrations by M. Nelson.)

Chordotonal Organ

Associated with the inner surface of the tympanal membrane is one to several chordotonal organs. Chordotonal organs are specialized mechanoreceptors unique to insects and crustaceans, but not unique to ears. Each chordotonal organ comprises one or more individual sensory units called scolopidia, and each scolopidium consists of three cells arranged in a linear array: a sensory cell, a scolopale cell, and an attachment cell (Fig. 2A). The total number of scolopidia in an ear ranges from one in some moths (Notodontidae) to almost 2000 in the bladder grasshopper *(Bullacris membracioides).* A chordotonal organ may attach directly to the inner surface of the tympanic membrane (e.g., Fig. 4E) or to tracheal air sacs indirectly associated with the tympanum. The axons of the sensory neurons collectively make up the auditory nerve that forms the neural link between the mechanosensory stimulation of the eardrum and its "perception" by the central nervous system. Vibrations of the eardrum and/or air chamber cause bioelectric currents to flow in the sensory cell, initiating action potentials in the auditory nerves and signaling neurons in the auditory pathways of the nervous system.

MECHANISMS

Sound waves are fluctuations in pressure traveling through a medium away from a source of mechanical disturbance. Tympanal ears are designed to detect these minute pressure changes traveling through air or water. Depending on the structural design of the ears, they may convey information about the location, frequency, and intensity of an acoustic stimulus.

There are two distinct types of insect tympanal ears: pressure receivers and pressure difference receivers (or pressure gradient receiver). In a pressure receiver, the tympanal membrane forms one side of an otherwise enclosed air chamber. The

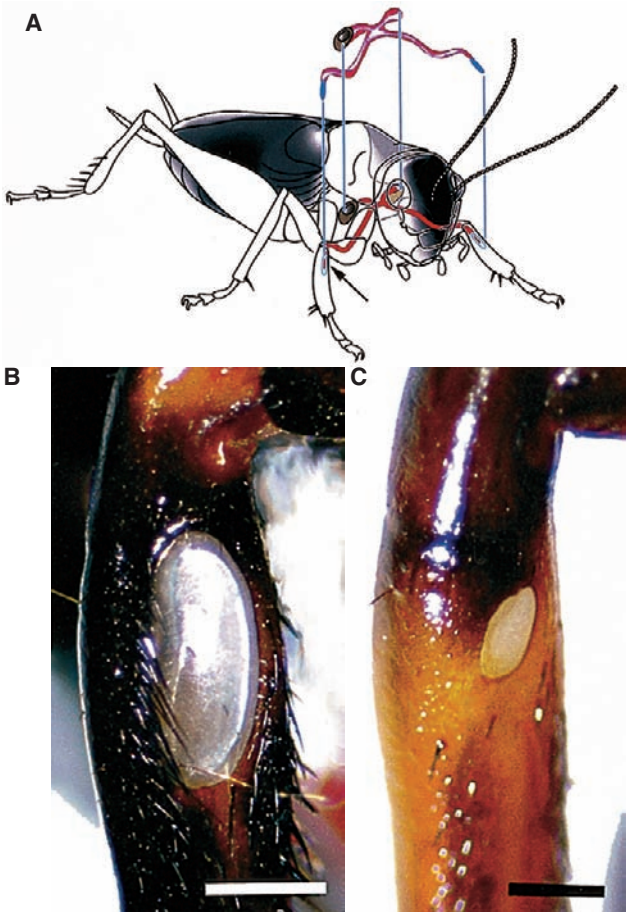

FIGURE 3 Schematic illustrations of structures associated with hearing in crickets (Ensifera). (A) The locations of the tibial ears and the H-shaped tracheal system connecting the contralateral tympanal membranes and spiracles are shown. (Illustration by M. Nelson.) (B and C) Light micrographs of the larger posterior and smaller anterior tympanal membranes located on the tibia of the forelegs in *Gryllus bimaculatus.* (Courtesy of A. C. Mason.)

membrane vibrates in response to sound waves arriving only to the external surface of the tympanal membrane. In insects whose ears are sensitive to high-frequency sounds (like many bat-detecting moth ears), the sound shadow of the body is sufficient to cause a difference in sound amplitudes arriving to each ear, which provides useful directional cues to the animal. Many insects using lower frequency sounds are faced with the difficulty of being able to localize sounds that arrive almost simultaneously to both ears. They overcome this problem using pressure-difference receivers in which the sound impinges on both the outer and the inner surfaces of the membrane, and the membrane vibrates in response to the difference in pressure between the two sides (e.g., Fig. 3).

A third and very unusual way to localize sounds has been "invented" by certain parasitic tachinid and sarcophagid flies. These flies localize their hosts (field crickets, katydids, and cicadas) by hearing and then homing in on the mating calls. The ears possessed by these flies are conventional chordotonal organs, containing 70 or more scolopidia. However, the fly's eardrums are mechanically connected, which confers an acute sense of directionality such that they can detect and localize their calling hosts/prey from a distance of tens of meters, even from high in the air. Such ears appear to be unique to these parasitoid flies.

The frequencies detectable by insect tympanal ears range from a few kilohertz (e.g., the water boatman *Corixa*) to over 100 kHz (e.g., some tettigoniid species). This broad bandwidth of frequency sensitivity is used by some insects for mate calling and in others it enables them to detect their predators (e.g., echolocating bats). Some insects have the ability to discriminate between different sound frequencies, which may be achieved at the level of the tympanal membrane or the chordotonal organ. To date, most insect ears studied appear to be tone deaf (i.e., different tones are indistinguishable).

LOCATION

Ears can be located just about anywhere on insects (Fig. 1). There is one striking difference between insect ears and vertebrate ears: those of the latter, because of developmental constraints, are always on opposite sides of the head and always on the head, whereas insect ears have been found on virtually every part of the body. Of course, in any given species of insect, ears are always found on the same part of the body, but from group to group, ears may be found on the head, legs, wings, thorax, or abdomen.

Although the ears of most insects are clearly recognized by the presence of a conspicuous tympanal membrane either on the body's external surface or within an ear cavity, the ears of other insects are morphologically cryptic. For example, some hawk moths (superfamily Sphingoidea) possess hearing organs in their mouthparts, and by inflating their palps while they are feeding on flowers at dusk, they create functional ears that alert them to the echolocation calls of bats. The "cyclopean" ear of the praying mantis is located within a deep groove between the hind legs, and the tympanal membranes are not morphologically distinct from surrounding body parts, even to the trained eye. Even more cryptic are the ears of the Madagascar hissing cockroach, *Gromphadorhina portentosa,* in which there is no obvious eardrum or tympanal membrane overlying the internal chordotonal organs contained within the tibiae of its forelegs.

Although most insects possess a single pair of hearing organs, there are at least two reports of insects with multiple ears. One lineage of praying mantids has two sets of ears: one located between the mesothoracic legs and tuned to ultrasonic frequencies (25 to 40 kHz) and another between the metathoracic legs and tuned to lower frequencies (2–4 kHz). The multieared bladder grasshopper *(B. membracioides)* of South Africa has six pairs of serially repeated abdominal ears that function in detecting mating calls at distances of ~2 km.

DEVELOPMENTAL ORIGINS

In vertebrates, ears are always located on the head, behind the eyes, and above the jaw, but in insects they are virtually

anywhere on the body. Why are they not confined to one place like those of vertebrates? Evolutionary and developmental origins of the vertebrate ear, drawing from embryological anlagen (organs or structures in their earliest state) such as the gill arches, neural crest, and the otic capsule, have constrained the ears to their canonical position in the head. Insect ears require but two structural modifications, cuticular and spiracular, combined with innervation by a chordotonal organ. In the insect body, chordotonal organs do not function only as hearing organs, but are actually widely distributed throughout the body, where they act as detectors of self-induced body movements (proprioceptors) or substrate vibrations. Insect bodies are made up of a series of hard cuticular plates joined by flexible membranes, and chordotonal organs are frequently suspended between moving joints. We now know through developmental studies and comparative anatomy that it is not so difficult to "construct" an ear by making a few peripheral modifications to an existing proprioceptor and its surrounding cuticular and tracheal structures. By simply thinning the cuticle, enlarging the surrounding tracheal air sacs to allow membrane vibration, and mechanically isolating the sensory organ from body movements, a proprioceptor can be converted into a sound-pressure receiver (Fig. 2B). For example, the bat-detecting ears on the thorax of noctuoid moths are thought to have evolved from proprioceptors monitoring wing movements. Given the jointed, segmental body plan of any insect, the ubiquitous branching of the respiratory tracheae lining the inner face of the cuticle, and the widespread occurrence of chordotonal organs that span different segments in the body and appendages, the precursors of an insect ear can be found virtually at every joint in the body and appendages. There seem to be few developmental constraints in positioning an ear, should an adaptive need arise.

Given so many possibilities for developing an ear, how does it come about that a particular insect possesses an ear in one place, say its forelegs, and another species in another, say its abdomen? One can imagine a host of anatomical, biophysical, ecological, and evolutionary constraints and advantages that might play important roles in the selection process. Such factors as the distances between the ears, the degree of protection offered by surrounding structures, the availability of tracheal air sacs, or the preexisting connections to the central nervous system may all be important. For example, if the function of hearing in a flying nocturnal insect like a moth is to detect and avoid echolocating bats, then a proprioceptor with preestablished neural connections to wing flight musculature would be a better ear candidate than, say, a leg proprioceptor. Or, possibly, the reason the thorax and abdomen are so "busy" with ears (Fig. 1) could be because these locations offer maximum interaural distance and a high degree of protection.

EVOLUTION

Tympanal ears have evolved independently at least 20 times within the class Insecta, and this number surely underestimates the number of ears that do exist. Still, hearing appears to be the exception, not the rule, for insects, with only 7 of the 25 recognized extant neopteran orders having tympanate species. The Orthoptera (crickets, grasshoppers, and katydids) and the Lepidoptera (moths and butterflies) boast the largest number of eared species. In two of the most speciose orders, the Diptera (flies and mosquitos) and the Coleoptera (beetles), tympanal ears are rare, and surprisingly, the Hymenoptera (wasps, ants, and bees) are completely atympanate as far as we know. Following is a brief introduction to the major taxonomic groups for which tympanal ears have been identified to date.

Orthoptera

Orthopteran ears come in two different forms and are divided nicely by the taxonomy of the order. The Acridoidea (locusts, grasshoppers) have ears on either side of the first abdominal segment. The tympanal membranes are nearly circular, opaque, and clearly visible upon inspection with the naked eye. Tracheal sacs connect both ears, allowing the locust to determine the direction of a sound source. The acridid ear is one of the few insect ears known to have the capability of pitch discrimination. About 60 to 80 scolopidia form four separate groups that attach to different regions of the eardrum, which, in turn, resonate to different sound frequencies. The ears of extant acrididis function primarily in conspecific communication, but comparative evidence suggests that the primitive function was for predator detection.

In the suborder Ensifera [crickets (Gryllidae), katydids (Tettigoniidae)] the ears occur just below the "knee" region, on the tibia of the forelegs. Each leg has two eardrums—one on each side of the leg (Fig. 3). The tympanal membranes are connected to other sound input sources (the spiracles, contralateral ear) via a system of tracheal tubes and air chambers, which play important roles in directional hearing. The ensiferan auditory chordotonal organ (crista acoustica) has typically between 60 and 80 sense cells arranged in a linear array down the leg, connecting indirectly to the tympanal membrane by tracheal air sacs. Like the acridid ear, the ensiferan ear is capable of pitch discrimination. Presumably, the function of hearing in primitive Ensifera was conspecific communication, which remains the primary function in extant species. Some species are sensitive to ultrasound and use their ears to detect bats in addition to communicating with conspecifics. Fossil records demonstrate that the ancestors of modern Ensifera, which predate the appearance of bats by at least 100 million years, had ears on their legs and therefore, ultrasonic hearing for defense against bats seems to have evolved secondarily.

Lepidoptera

Tympanal ears have evolved more times within the Lepidoptera than any other insect order. To date, at least seven ears of independent origins have been described. Ears are located in the mouthparts (Sphingoidea), at the base of the forewings (butterflies: Hedyloidea, Nymphalidae), in the

thorax (Noctuoidea), or on anterior abdominal segments (Tineoidea, Pyraloidea, Drepanoidea, Geometroidea, Uraniidae). Moth ears are among the simplest of all insect ears in that they have very few auditory sensory cells (one, two, or four). K. D. Roeder first demonstrated that the primary function of hearing in moths is to detect the ultrasonic echolocation cries of insectivorous bats. Despite their simplicity, moth ears are capable of determining the distance and direction of an approaching bat. Flying moths evade bats by either turning away from a distant bat or engaging in evasive and erratic flight maneuvers to avoid a sudden and unexpected attack. W. E. Conner has shown that some species, including many day-flying tiger moths (Arctiidae), use their hearing in social interactions. In most Arctiidae, ultrasound production functions to either jam a bat's echolocation calls or warn the bat of a distasteful meal (aposomatic display), but in those species that use sounds for social interactions, both hearing and sound production have secondarily taken on different roles.

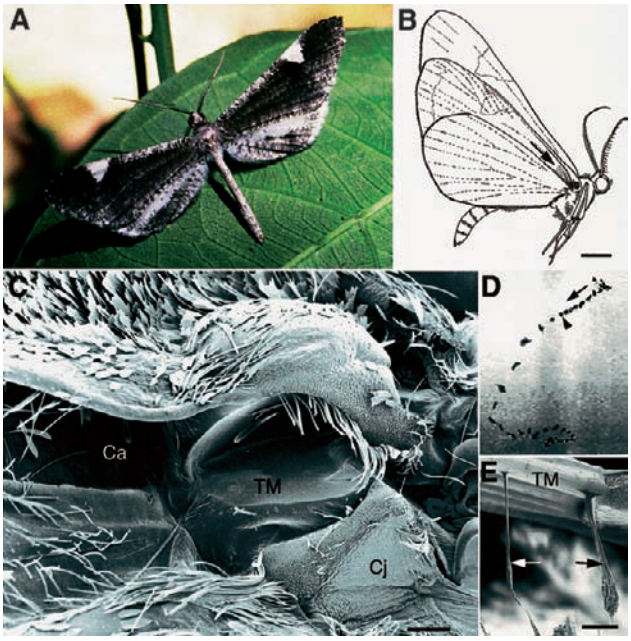

FIGURE 4 Nocturnal butterflies of the superfamily Hedyloidea possess ultrasound-sensitive ears on their wings that mediate evasive flight maneuvers to avoid bats. (A) A male *Macrosoma heliconiaria* (Hedylidae). (B) Lateral view of *M. heliconiaria,* showing the general location of the right ear. An arrow points down the canal toward the tympanic cavity where the tympanic membrane resides. Scale bar, 3 mm. (C) Scanning electron micrograph of the right tympanal ear. The hind wing and a dense fringe of scales have been removed to reveal the tympanic cavity. Ca, canal; Cj, conjuctiva. Scale bar, 110 µm. (D) Consecutive video images (30 frames/s) of a free-flying *M. heliconiaria* responding to a short (~250 ms), high-frequency (25 kHz), high intensity (>100 dB) sound. The direction of flight is marked with an arrow and the stimulus onset with an arrowhead. (E) Scanning electron micrograph of the two most proximal chordotonal organs (arrows), viewed from inside the tympanic chamber. The largest organ attaches to the proximal border of the tympanal frame (black arrow). TM, tympanic membrane. Scale bar, 50 µm. (A and E courtesy of J. Yack. B–D were modified, with permission, from J. E. Yack and J. H. Fullard, 2000, Ultrasonic hearing in nocturnal butterflies. *Nature* **403,** 265–266.)

Hearing in butterflies has only begun to be explored. The hedylid butterflies of the neotropics, unusual because of their nocturnal habits, have ultra-sound-sensitive ears on their wings that function as bat detectors (Fig. 4). Some diurnal butterflies of the family Nymphalidae possess ear-like structures on their wings, and there is evidence that the "cracker" butterfly *(Hamadryas feronia)* uses its ears for conspecific communication. No doubt there will be further examples of butterfly hearing in the near future, since tympanal-like structures have been described anatomically in many species.

Homoptera and Heteroptera

Most people are familiar with the loud "buzzing" sounds of cicadas (Cicadidae) during hot weather; these are typically males calling to females. Cicada ears are located within cavities on the ventral side of the second abdominal segment. They are among the largest of all insect ears, with over 1000 scolopidia in each ear. It has been suggested that the large number of sense cells enhances the ear's sensitivity for long distance communication, but cicadas do not appear to have more sensitive hearing organs than other insects with far fewer scolopidia.

Water bugs (subfamily Hydrocorisae) are reported to have ears on various body parts, including the mesothorax, metathorax, or first abdominal segment. The best known is the ear of the water boatman, *Corixa* (Corixidae), which occurs on the lateral mesothorax between the wing and the leg. The ears, like those of moths, are simple, with only two auditory cells, both tuned to low-frequency sounds (1–2 kHz), within the range of conspecific calls. The insect carries a bubble of water with it to allow the membrane to vibrate under water. Unlike for most other insect ears, the tympanal membrane is backed by fluid, not air. Hearing in corixids appears to function primarily for mate attraction.

Diptera

Although it had been well documented that certain parasitic flies were attracted to the songs of crickets, katydids, and cicadas, until recently, it was not known how these flies were eavesdropping on their hosts. Two families of parasitoid flies (Sarcophagidae and Tachinidae) have independently evolved a pair of peculiar ears on their prosternum, just below the head, in the "neck" region (Fig. 5). The gravid females use their ears to locate singing insect hosts upon which to lay their predaceous larvae. The ears of the two parasitoid groups are described earlier in this article. The design features of these dipteran ears show remarkable convergence in anatomy and function, despite the fact that tachinids and sarcophagids are only distantly related. This suggests that evolutionary and developmental constraints are at work here in ways that we do not yet understand.

Mantodea

Until recently, praying mantids were thought to be deaf; now we know that 65% of all mantid species can hear. The ears

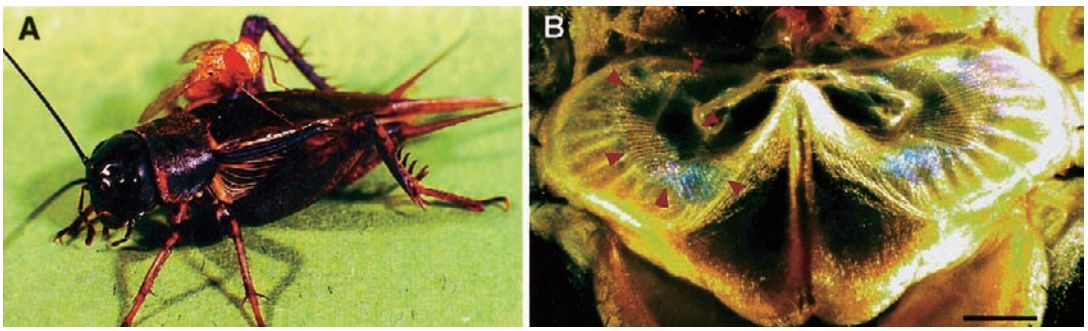

FIGURE 5 The prothoracic tympanal ear of a parasitoid tachinid fly, *Ormia ochracea* (Tachinidae). (A) A female fly resting on its host (*Gryllus integer*). (B) A light micrograph of the prosternal ear in a female. The outline of the tympanal membrane is indicated with arrows. (Photographs provided courtesy of D. Robert, R. Hoy, and G. Haldimann.)

occur in the most bizarre location: the two tympanal membranes face one another inside a narrow groove between the metathoracic legs. The mantid ear is functionally an "auditory cyclops" because the close proximity of the ear drums (less than 150 μm) provides no directional information to the animal. The ears function as bat detectors and are most sensitive between 25 and 50 kHz.

Coleoptera

Tympanal ears have been described in two beetle families to date. Several species of the genus *Cicindela* (Cicindelidae) have ears on the dorsal surface of the first abdominal segment, beneath the wings. Some scarabs (two tribes in the subfamily Dynastinae) have ears located just beneath the neck membranes (pronotal shield). The ears of both families are tuned to ultrasonic frequencies, and there is strong evidence that they function as bat detectors. It is a little bit surprising that there are not more examples of hearing in beetles. Given the large number of species, the wide diversity of niches worldwide, and numerous reports of sound production, we expect that more examples will be uncovered.

Neuroptera

Green lacewings (Chrysopidae) have an ear near the base of each forewing in a location similar to that of the ears of some butterflies. The ear consists of a swelling of the radial vein, with a region of very thin cuticle on the ventral side that functions as a tympanal membrane. Like the corixid ear, the tympanal chamber is predominantly fluid filled. The ears respond to sounds between 40 and 60 kHz and are sufficiently sensitive to detect echolocating bats at close distances.

NONTYMPANAL HEARING ORGANS

Until now we have focused on tympanal ears, which are sensitive to traveling waves of changing pressure in air and water, known as the acoustic far field. In the broadest sense of the word, however, hearing encompasses the detection of near-field sounds, as well as vibrations traveling through solid substrates. By and large, the near field can be thought of as a short distance, a few body lengths, from the sound source. Substrate vibrational signals have also been described as "seismic communication." In this larger sense, then, we could argue that most insects can hear. We highlight a few recent developments in the study of these alternative, but no doubt widespread, forms of hearing.

Detecting Near-Field Sounds

When a sound is produced, air particles are being pushed back and forth near the source of disturbance. These particle movements, or near-field sounds, are generally of low frequency (typically below 1 kHz) and, unlike pressure waves, do not travel far from the sound source, in many instances, just a few body lengths in distance. A small, light, and ponderable object occurring within the near-field sound will move in response to the vibrating air molecules. In insects, loosely attached setae or antennae are commonly used for detecting particle velocity. Depending on the structure of the receptor organ, and its position relative to the sound source, near-field receptors can offer information about the direction and the intensity of a sound source.

Some caterpillars can detect the near-field sounds produced by the beating wings of a flying wasp up to distances of 70 cm. Specialized hairs on the dorsal thorax of the caterpillar are displaced in the sound's near field, eliciting an evasive response, such as freezing or dropping from a leaf. Similar particle-displacement-sensitive setae on the cerci of crickets and cockroaches function in predator avoidance and possibly for close-range conspecific communication.

The antennae of many insects also function as near-field sound detectors. In many Diptera (mosquitos, chironomids, and fruit flies), for example, the males are attracted to the near-field "buzzing sounds" of females (Fig. 6). The long flagella of the antennae in male mosquitoes resonate to the tune of flying females, and these antennal movements in turn stimulate many thousands of scolopidia in the Johnston's organ, a chordotonal organ located at the base of the

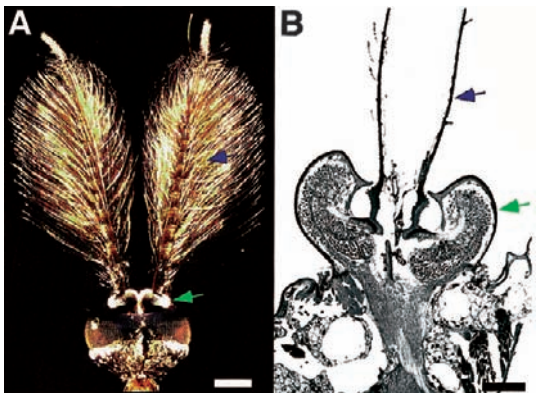

FIGURE 6 The near-field sound hearing organs of a mosquito. (A) Micrograph of the head of a male *Toxorhynchites brevipalpis,* showing the plumous flagella of the antennae (blue arrows), which resonate in response to the near-field sounds of flying females. (B) A cross section through the antennal base, showing the location of the Johnston's organ (green arrows), where the auditory scolopidia are located. (A is courtesy of D. Huber. B was adapted, with permission, from M. C. Göpfert and D. Robert, 2001, Active auditory mechanics in mosquitoes, *Proc. R. Soc. London B* **268,** 333–339.)

antennae. The Johnston's organ may also be involved in the detection of near-field sounds produced by honey bee waggle dances. It is also the sensory organ by which the well-known *Drosophila melanogaster* detects the patterned wingbeats that make up the mating songs of these species.

Detecting Substrate-Borne Vibrations

The detection of vibrations traveling through solid substrates may be one of the most ubiquitous but least appreciated forms of acoustic communication in insects. With recent technological advances in the detection of substrate-borne signals (e.g., laser vibrometry and piezoelectric sensors), we are learning that a large number of insects can detect vibrations produced by both intentional and unintentional senders. In fact, most insect orders probably include species capable of detecting vibrations. At present, however, there are few complete studies on this subject. We provide two recent examples out of many possibilities.

Membracid treehoppers (Homoptera: Membracidae) transmit alarm calls through tree stems to communicate with conspecifics. Nymphs living in colonies of up to 100 individuals on the stem of a host plant produce coordinated waves of vibrations when threatened by a predator. The mother treehopper detects these alarm calls and rushes to defend her offspring, by kicking her hind legs and fanning her wings at the intruder (Fig. 7A). For many species of caterpillars that live in social or crowded conditions, vibrational signaling may be a principal means of communication. The common North American masked-birch caterpillar (*Drepana arcuata*) engages in acoustic "battles" with invading conspecifics. The nest owner drums and scrapes its mandibles and scrapes modified "oars" against the leaf in ritualized acoustic displays. Acoustic "duels" between residents and intruders can last

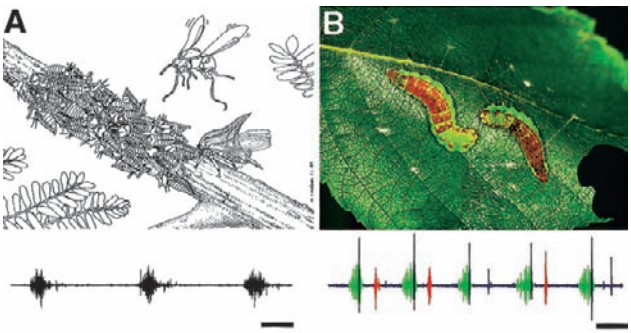

FIGURE 7 Acoustic communication through substrate-borne vibrations. (A) A group of treehopper nymphs (*Umbonia crassicornis*) use coordinated substrate-borne vibrations to warn their mother of an approaching predatory wasp. The mother detects the signals and rushes to the defense of her offspring. The waveform represents three group signals from an aggregation of nymphs. Scale bar, 560 ms. (Reproduced, with permission, from R. B. Cocroft, 2002, Antipredator defense as a limited resource: Unequal predation risk in broods of an insect with maternal care, *Behav. Ecol.* **13(1),** 125–133. Waveform courtesy of R. B. Cocroft.) (B) Two masked birch caterpillars (*Drepana arcuata*) engaged in an acoustic "duel" over a silken nest on a birch leaf. The waveform depicts the three signal types (green, anal scrapes; blue, mandible drumming; orange, mandible scraping) used by the caterpillars. (Reproduced, with permission, from J. E. Yack *et al.,* 2001, Caterpillar talk: Acoustically mediated territoriality in larval Lepidoptera, *Proc. Natl. Acad. Sci. USA* **98(20),** 11371–11375.)

from a few minutes to a few hours (Fig. 7B). At present, we know little about how insects detect substrate vibrations. The subgenual organ (a chordotonal just "below the knee" in many insects) functions as a vibration receptor in some groups (like some crickets and termites), but for most insects, the receptor organs are yet to be identified. Clearly further research is required before we gain a full appreciation of this important form of communication in insects.

FUNCTION

Because of their physical properties, acoustic signals are highly adaptive for certain kinds of behavioral interactions: sound waves can travel at any time of the day or night, through thick vegetation or muddy water; they convey information instantaneously and can be transmitted over long distances; and sounds are easy to localize, do not leave lingering traces, and can transmit large amounts of information per unit time. For the majority of insects, acoustic communication functions primarily in reproductive behavior and predator avoidance, but may also be used for detecting prey or host species (parasitic flies, wasps; predatory water striders, ant lions) or calling to conspecifics to form aggregations (sawfly larvae) or warn of danger (termites, treehoppers).

For humans, the most conspicuous sounds commonly heard from insects are the loud chirps and trills of field crickets, the long raspy choruses of katydids by night, and the intense, shrill-like buzzes and rattles of cicadas by day. These are the mating calls emitted by males in order to attract conspecific females. Sounds used in reproductive interactions

function in species recognition, courting, pair maintenance, female mate choice, and male–male competition. The hearing organs of crickets, katydids, grasshoppers, mosquitos, and cicadas are used primarily for these purposes and are sharply tuned to the calls of conspecifics. The features of these mating calls have surely been shaped by sexual selection.

Many insects have ears for the sole function of detecting predators. Many nocturnally active insects (most Lepidoptera, some mantids, beetles, and lacewings) have ears tuned to the ultrasonic vocalizations of insectivorous bats that use biosonar to detect and home in on their prey. Unlike ears specifically designed for conspecific communication, the ears of predator detectors are usually more broadly tuned and more simple in their design, sometimes having only a few auditory cells per ear.

CONCLUSION

Since antiquity we have known that many insects produce sounds, but only during the past 150 years have scientists realized that some insects can hear. Detailed descriptions of ear anatomy, and the behaviors associated with hearing, began in the early 1800s, providing the basis for current developments in the field of insect bioacoustics. Over the past 40 years there have been significant advances in the field: many new ears have been discovered, and previous claims to tympanal hearing (based on morphological studies) have been validated. With the development of new instruments for detecting acoustic signals outside the realm of human perception (e.g., ultrasound, solid-substrate-borne vibrations), and for recording neurophysiological responses to sound, we are now beginning to better appreciate the immense diversity of insect sound receptor organs.

There is still much to learn about insect hearing. We know little, for example, about the chain of physical and bioelectrical events leading to sound reception at the level of the auditory cells or how acoustic sensory responses are integrated at the level of the central nervous system to promote adaptive behaviors. New tympanal ears will no doubt turn up in the years ahead, but perhaps most significantly, future explorations into substrate-vibrational and near-field sound communication are sure to yield exciting insights into how insects communicate acoustically.

See Also the Following Articles

Mating Behaviors • Mechanoreception • Orientation • Vibrational Communication

Further Reading

Bailey, W. J. (1991). "Acoustic Behavior of Insects." Chapman & Hall, London/New York.

Ewing, A. W. (1989). "Arthropod Bioacoustics." Cornell University Press, Ithaca, NY.

Fullard, J. H., and Yack, J. E. (1993). The evolutionary biology of insect hearing. *Trends Ecol. Evol.* **8(7),** 248–252.

Hoy, R. R., and Robert, D. (1996). Tympanal hearing in insects. *Annu. Rev. Entomol.* **41,** 433–450.

Hoy, R. R., Popper, A. N., and Fay, R. R. (eds.) (1998). "Comparative hearing: Insects. Springer Handbook of Auditory Research." Springer-Verlag, New York.

Michelsen, A. (1979). Insect ears as mechanical systems. *Am. Sci.* **67,** 696–706.

Michelsen, A., and Larsen, O. N. (1985). Hearing and sound. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.). Pergamon, Oxford.

Robert, D., and Hoy, R. R. (1998). The evolutionary innovation of tympanal hearing in Diptera. *In* "Comparative Hearing: Insects" (R. R. Hoy, A. N. Popper, and R. R. Fay, eds.), Chap. 6. Springer-Verlag, New York.

Roeder, K. D. (1967). "Nerve Cells and Insect Behavior." Harvard University Press, Cambridge, MA.

Scoble, M. J. (1995). Hearing, sound and scent. *In* "The Lepidoptera. Form, Function and Diversity." Oxford University Press, Oxford.

Yack, J. E., and Fullard, J. H. (1993). What is an insect ear? *Ann. Entomol. Soc. Am.* **86(6),** 677–682.

Yager, D. D. (1999). Structure, development, and evolution of insect auditory systems. *Microsc. Res. Tech.* **47,** 380–400.

Hemiptera

see *Auchenorrhyncha; Prosorrhyncha; Sternorrhyncha*

Hemolymph

Michael R. Kanost
Kansas State University

Hemolymph is the circulating fluid or "blood" of insects. It moves through the open circulatory system, directly bathing the organs and tissues. Insect hemolymph differs substantially from vertebrate blood, with the absence of erythrocytes and a high concentration of free amino acids being two of the common distinguishing features. The main component of hemolymph is water, which functions as a solvent for a variety of molecules. Water in hemolymph makes up 20 to 50% of the total water in insect bodies, with larval stages generally having a larger relative hemolymph volume than adults. Hemolymph serves as a water storage pool for use by tissues during desiccation and as a storage depot for other types of chemicals. It also contains circulating cells called hemocytes. Hemolymph can function as a hydraulic fluid, for example, in the expansion of a newly molted butterfly's wings. Hemolymph serves important roles in the immune system and in transport of hormones, nutrients, and metabolites.

INORGANIC COMPONENTS

The composition of inorganic ions in hemolymph varies widely among different insect groups. The pH of the

hemolymph of most insects is in the range of 6.4 to 6.8. Apterygotes contain high levels of sodium and chloride, similar to mammalian blood. In hemolymph of exopterygotes, sodium and chloride are also high but magnesium makes up a large portion of the total inorganic cations. In endopterygotes, particularly Lepidoptera, Coleoptera, and Hymenoptera, concentrations of sodium and chloride tend to be much lower and are replaced with high levels of potassium, magnesium, and organic anions. This difference has been attributed to the coevolution of these insect groups with flowering plants and the consequent dietary importance of leaves (which contain high concentrations of magnesium and potassium). However, the concentration of inorganic ions is not a function of only the diet, because insects are able to regulate the ion composition of hemolymph to some degree.

LOW-MOLECULAR-WEIGHT ORGANIC COMPONENTS

Citric acid and other organic acids and organic phosphates (such as glycerol 1-phosphate and sorbitol 6-phosphate) account for much of the anion content in hemolymph from many insect species. The most abundant carbohydrate in hemolymph of most insects is the disaccharide trehalose. Transport of trehalose as an energy source for tissues is an important function of the hemolymph. Trehalose levels are hormonally regulated and can be increased through synthesis from glucose phosphate derived from glycogen stored in the fat body. Glucose may also be present in hemolymph, although generally at a lower concentration than trehalose. In some insects, diapause or exposure to low temperatures can stimulate synthesis of glycerol and sorbitol (from glycogen stored in fat body). The resulting high concentration of these compounds in hemolymph depresses the freezing point and protects the insects from damage that would occur if ice crystals were to form in hemolymph.

Hydrophobic lipoidal compounds present in hemolymph are carried by specific transport proteins. Diacylglycerol is the major transported form of lipid in most insects, but triacylglycerol, fatty acids, phospholipids, and cholesterol are also present. Pigments such as β-carotene, riboflavin, and biliverdin, which give hemolymph of many insects a characteristic yellow or green color, are also carried by specific proteins.

Free amino acids are present at high concentration (up to 200 mM) in hemolymph and make a major contribution to hemolymph osmolarity. All 20 of the amino acids found in proteins exist as free amino acids in hemolymph. Although the relative concentrations of the amino acids vary in different species, glutamine and proline are typically abundant. Proline is known to serve as an energy source for flight muscles in some species. Hemolymph may also contain some amino acids that are not found in proteins, such as β-alanine and taurine. Tyrosine, which is metabolized for use in cuticle sclerotization, often occurs in hemolymph as a conjugate with glucose, phosphate, or β-alanine, which increases its solubility. The phosphate and glucose substituents are removed from tyrosine by specific enzymes when tyrosine is needed for sclerotization. Catecholamines derived from tyrosine, which are used in cuticle sclerotization and pigmentation, are also present in hemolymph as conjugated forms.

PLASMA PROTEINS

Proteins are a major component of the hemolymph plasma. Typical protein concentrations in plasma range from 10 to 100 mg/ml. In most species, the concentration of proteins in plasma increases during each instar and decreases at each molt. The fat body is responsible for the synthesis of the majority of plasma proteins, but there is also a contribution of some specific proteins from epidermis and hemocytes. Plasma from each species contains a few very abundant proteins and more than a hundred other proteins at much lower concentrations. Although the identities and functions of the major proteins are understood, many of the minor hemolymph proteins have not yet been thoroughly investigated.

Storage Proteins

The most abundant proteins in larval hemolymph belong to a class known as storage proteins or hexamerins (because they are assembled from six ~80-kDa polypeptide subunits). The storage proteins are synthesized by the fat body and reach extremely high concentrations in the last instar. At the end of this stage, most of the storage proteins are taken back into the fat body, through interaction with specific receptors, and stored in protein granules. During metamorphosis the storage proteins are broken down into free amino acids, which are used for synthesis of other proteins required in the adult stage. In some exopterygotes, hexamerins are again synthesized by the adult, although their function at this developmental stage is unclear. The hexamerins can be classified according to their amino acid compositions. Those rich in the aromatic amino acids (phenylalanine, tyrosine, and tryptophan) are called arylphorins, whereas another group of hexamerins are known as methionine-rich storage proteins. In addition, some other proteins that function as storage proteins but are not similar in sequence to the hexamerins have been identified in lepidopterans.

Transport Proteins

Several hemolymph proteins function to transport small molecules that have low solubility in water. Insect plasma contains two proteins that specifically bind iron; ferritin appears to sequester dietary iron, whereas transferrin acts as a shuttle to transport iron between tissues.

The most abundant transport protein in hemolymph is lipophorin, which transports lipids between tissues. Like lipoproteins in mammalian plasma, lipophorin is composed of proteins that complex with lipids in such a way that the lipids are protected from contact with the surrounding water.

Lipophorin docks with specific receptors on the surface of tissues to either accept or unload diacylgycerol. Lipophorin contains two polypeptide subunits, apolipophorin-I and apolipophorin-II, which are produced by proteolytic cleavage of a larger protein precursor. In insects that use lipids as a fuel for flight muscles, diacylglycerol is released from the fat body into the hemolymph under control of a peptide hormone known as adipokinetic hormone. As lipophorin accepts large amounts of diacylglycerol, its volume increases and its density decreases as it is converted from high-density lipophorin to low-density lipophorin. Low-density lipophorin contains a third type of protein subunit, apolipophorin-III, which binds to the surface to stabilize the expanding lipid–water interface.

Juvenile hormone (JH), a sesquiterpenoid lipid, has low solubility in water and is transported through hemolymph bound to a specific carrier protein. JH binding proteins of ~30 kDa have been well characterized from plasma of lepidopterans, whereas in other insect orders lipophorin or a specific hexamerin takes on the role of JH transport. In addition to keeping JH in solution, these proteins also protect the hormone from degradative enzymes that help to regulate JH concentration in plasma. JH binding proteins may also aid in delivery of the hormone to target tissues.

Egg Yolk Proteins

In adult female insects, certain proteins synthesized by the fat body and secreted into the hemolymph are delivered to the ovary, where they are taken up by developing oocytes. The most abundant of these is called vitellogenin. Once vitellogenin becomes a part of the egg yolk, it is called vitellin. Vitellogenins are typically large, phosphorylated lipoglycoproteins that are expressed specifically in adult females. Lipophorin is also taken up from hemolymph into eggs and provides additional lipids for use by the developing embryo. Vitellogenin and lipophorin are related in their amino acid sequences, indicating that they have a common ancestral gene. Vitellogenin, lipophorin, and a few other plasma proteins are taken up into oocytes by receptor-mediated endocytosis.

Proteins and Peptides Involved in Immune Responses

A group of plasma proteins functions in defense against microbial infection. Hemolymph of many insects contains lysozyme, an enzyme that degrades bacterial cell walls. In addition, low-molecular-weight antimicrobial peptides are synthesized in response to bacterial or fungal infection. Many of these peptides act by disrupting the integrity of bacterial cell membranes. Phenoloxidase, an enzyme present in plasma of some species and stored in hemocytes of others, is synthesized as an inactive precursor, prophenoloxidase. In response to infection or injury, prophenoloxidase is activated and catalyzes the production of quinones that polymerize to form the pigment melanin, which helps to trap and kill invading organisms. The tendency of hemolymph to darken has been known for more than 100 years, but this melanization has only recently become understood at a molecular level. Plasma contains proteins that bind to carbohydrates on the surface of microorganisms. This causes activation of a cascade of proteases that results in the proteolytic activation of prophenoloxidase. To regulate this immune response, plasma contains several types of proteins that function as protease inhibitors.

HEMOCYTES

The circulating cells in hemolymph are called hemocytes. Insects lack erythrocytes, and hemocytes cannot be directly equated with vertebrate leukocytes. Some fraction of hemocytes remains sessile and attached to the surfaces of tissues, and in some species (mosquitoes, for example) such cells may account for a majority of the hemocytes. Several different morphological types of hemocytes can be identified in each insect species. Some commonly observed hemocyte types are illustrated in Fig. 1. Prohemocytes are small, round cells that may be precursors from which some other cell types develop. Granular hemocytes contain conspicuous cytoplasmic granules that can be discharged as part of a defensive response

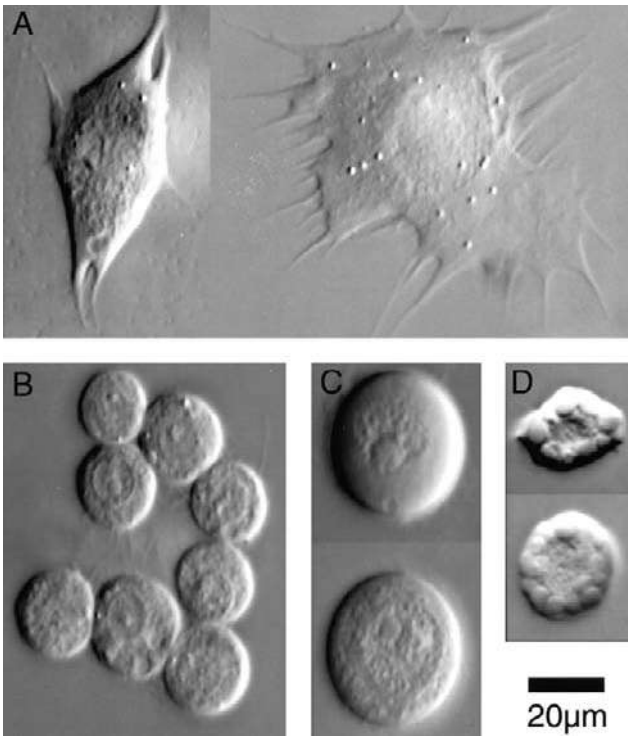

FIGURE 1 Examples of hemocyte types from a lepidopteran, *Manduca sexta.* (A) Plasmatocytes. The plasmatocyte shown on the left has just begun to spread, whereas the one on the right has spread extensively. (B) Granulocytes. (C) Oenocytoids. (D) Spherulocytes.

to invading parasites. Plasmatocytes usually contain few granules and are characterized by their ability to change from round or spindle-shaped cells in suspension to extensively flattened, ameboid cells after attaching to a substrate. Spherule cells contain very large cytoplasmic granules, which may contain mucopolysaccharides. Oenocytoids are large cells that synthesize prophenoloxidase.

Plasmatocytes and granular hemocytes are usually the two most abundant hemocyte types, although their proportions can vary between species and within a species at different developmental stages. These two hemocyte types participate in immune responses, including: (1) phagocytosis of small organisms such as bacteria; (2) nodule formation, in which multiple hemocytes aggregate to trap microorganisms; and (3) encapsulation, in which hemocytes attach to the surface of a larger parasite and form a multilayered hemocyte capsule, in which the parasite is killed. Nodules and capsules often become melanized through the action of phenoloxidase. Hemocytes, especially plasmatocytes, also aggregate in a type of coagulation response, sealing wounds to prevent hemolymph loss. Another function of hemocytes is in synthesis of the extracellular matrix that covers tissues exposed to the hemolymph. Granular hemocytes appear to be the primary cell type involved in this aspect of hemocyte function.

See Also the Following Articles

Circulatory System • Fat Body • Immunology • Vitellogenesis • Water and Ion Balance, Hormonal Control of

Further Reading

Gillespie, J. P., Kanost, M. R., and Trenczek, T. (1997). Biological mediators of insect immunity. *Annu. Rev. Entomol.* **42,** 611–643.

Gupta, A. P. (1985). Cellular elements in the hemolymph. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 3. Pergamon, New York.

Haunerland, N. H. (1996). Insect storage proteins: Gene families and receptors. *Insect Biochem. Mol. Biol.* **26,** 755–765.

Kanost, M. R., Kawooya, J. K., Law, J. H., Ryan, R. O., Van Heusden, M. C., and Ziegler, R. (1990). Insect haemolymph proteins. *Adv. Insect Physiol.* **22,** 299–396.

Lackie, A. M. (1988). Haemocyte behaviour. *Adv. Insect Physiol.* **21,** 85–178.

Mullins, D. E. (1985). Chemistry and physiology of the hemolymph. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 3. Pergamon, New York.

Ryan, R. O., and van der Horst, D. J. (2000). Lipid transport biochemistry and its role in energy production. *Annu. Rev. Entomol.* **45,** 233–260.

Telfer, W. H., and Kunkel, J. G. (1991). The function and evolution of insect storage hexamers. *Annu. Rev. Entomol.* **36,** 205–228.

Wyatt, G. R. (1961). The biochemistry of insect hemolymph. *Annu. Rev. Entomol.* **6,** 75–102.

Heteroptera

see *Prosorrhyncha*

Hibernation

Richard E. Lee, Jr.
Miami University, Oxford, Ohio

Hibernation refers to the state in which animals pass the winter. In discussion of insects overwintering is often used as a synonym for hibernation; usually hibernation is associated with entry into a dormant state. Estivation is a term used for animals that become dormant in the summer. Some insects enter an extended period of dormancy, referred to as estivohibernation, which begins in summer and continues through the winter. This term also is used commonly in connection with those mammals that lower their body temperature slightly (e.g., carnivorean lethargy in bears and skunks) or extensively (e.g., hibernation *sensu stricto* in ground squirrels) during the winter.

It is critical for insects to synchronize their periods of feeding, growth, and reproductive activities with those times of the year when food is available and environmental conditions are suitable. Hibernation generally includes entry into diapause, a dormant state that promotes survival by depressing metabolism and energy utilization when host plants and other food sources are unavailable. Typically, this also includes reduced morphogenesis in immature stages; hibernating adults typically hibernate before reproducing.

Specific behavioral changes are often associated with movement to overwintering sites, termed hibernacula. One of the most extreme examples is the monarch butterfly *(Danaus plexippus),* which may migrate more than 5000 km from southern Canada and New England to mountain sites in central Mexico. Other insects migrate locally as they move to hibernacula within the soil where they may burrow to avoid exposure to winter cold. Still others seek sites beneath rocks, logs, bark, and leaf litter.

In temperate regions, insects typically hibernate in a specific life stage. The egg, larva, and pupa are more common stages for overwintering than is the adult. However, alpine and polar insects or others living in extreme environments, in which growing seasons are short or unpredictable, and having life cycles that may be extended over several years may hibernate multiple times in one or more life stages.

In temperate regions, most hibernating insects enhance their resistance to environmental extremes, particularly cold and desiccation. During autumn, many species markedly enhance their tolerance to low temperature, termed cold-hardening. Most insects are freezing intolerant and are unable to survive freezing within their body fluids. These species typically enhance their capacity to supercool (i.e., remain unfrozen at temperatures below the melting point of their hemolymph) by synthesizing glycerol, sorbitol, trehalose, or other cryoprotective compounds, often at high concentrations of 1 M or more. Production of antifreeze proteins, avoidance

of inoculative freezing by external ice, and ridding the body of ice nucleators that may catalyze ice formation are other mechanisms used to avoid lethal freezing.

In contrast, a few insects are freeze tolerant and can survive the freezing of 70% or more of their body water. Cryoprotectants are also commonly synthesized by freeze-tolerant species, as are ice-nucleating proteins that induce ice formation at high subzero temperatures. Overwintering insects also may acquire exceptionally high levels of desiccation resistance, comparable to those of desert species.

Behavioral and physiological changes associated with hibernation, diapause, and cold-hardening are commonly triggered by environmental cues, including photoperiod, temperature, moisture conditions, and changes in host plant quality. These cues ensure that adaptive responses occur before severe winter conditions arrive.

See Also the Following Articles

Aestivation • Cold/Heat Protection • Diapause • Dormancy • Monarchs

Further Reading

Danks, H. V. (1987). "Insect Dormancy: An Ecological Perspective." Biological Survey of Canada (Terrestrial Arthropods), Ottawa.

Denlinger, D. L., Giebultowicz, J. M., and Saunders, D. S. (eds.) (2001). "Insect Timing: Circadian Rhythmicity to Seasonality." Elsevier, Amsterdam.

Hallman, G. J., and Denlinger, D. L. (eds.) (1998). "Temperature Sensitivity in Insects and Application in Integrated Pest Management." Westview, Boulder, CO.

Leather, S. E., Walters, K. F. A., and Bale, J. S. (1993). "The Ecology of Insect Overwintering." Cambridge University Press, New York.

Lee, R. E., and Denlinger, D. L. (eds.) (1991). "Insects at Low Temperature." Chapman & Hall, New York.

Tauber, M. J., Tauber, C. A., and Masaki, S. (1986). "Seasonal Adaptations of Insects." Oxford University Press, New York.

History of Entomology

Edward H. Smith
Cornell University (Emeritus)
Ashville, North Carolina

Janet R. Smith
Asheville, North Carolina

This brief history traces the interactions of humans and insects dating from the adoption of agriculture and its inherent ecological disruptions. Humankind's early preoccupation with survival focused on insects as relentless pests, competitors for food and fiber, threats to health and comfort. The high hopes following World War II for relief from the bondage of insects through the use of chemical insecticides such as DDT proved unrealistic. The reassessment that followed led to a concept based on ecological principles which is referred to as integrated pest management (IPM). In this system, multiple control technologies are used, with the additive effect being to hold insect injury at acceptable levels while avoiding excessive environmental insult. The age-old struggle continues; entomologists are now armed with the lessons of the past; advances in insecticidal chemistry, bio logical control, and cultural methods; and visionary new technologies based on genetic modification of plants and animals. Simultaneously, the rise of the environmental movement and ecological awareness has placed insects in a new context, highlighting their essential role in biodiversity on which the viability of the Earth depends. The vision for the 21st century calls for compatibility between insect control and conservation; both are prerequisites to human well-being. Stewardship of the Earth is the greatest challenge ahead and one that places awesome responsibility on the shoulders of entomologists.

IN THE BEGINNING

The history of life on earth reaches back some 4 billion years. From this beginning the long evolutionary trail unwound. Along the way, 99% of the forms that appeared met with extinction.

The great exterminations that have occurred since the appearance of insects in the Devonian period, 400 mya, revealed insects' remarkable survival qualities. Insects witnessed the last of the trilobites that preceded them by 175 million years. By the time the dinosaurs appeared in the Triassic period, 210 mya, the major orders of insects existing today were already well established. Dinosaurs became extinct 66 mya, leaving a niche occupied in time by mammals. The mammals, in turn, provided a niche for insects, offering furry cover and warm meals. The disappearance of the dinosaurs coincided with a great radiation of insects based on insects' symbiosis with flowering plants. For the past 150 million years, the flowering plants and insects have honed their intricate coevolution, which accounts for their immense biodiversity on which human habitability of the earth depends.

Insects have withstood trial by ice and fire, meteorite strikes, volcanic eruptions, global dust veils, acid rain, and continental upheavals. This evolutionary experience is encoded in their DNA and attests to the advantage of their small size, external skeleton, flight, metamorphosis, and specialized systems of reproduction. These are significant credentials in insects' rivalry with *Homo sapiens,* a species that draws on an evolutionary history of a scant 7 million years.

COEXISTENCE, HUMANS AND INSECTS

Class Insecta has plagued and fascinated humans for all of their history. The most striking features of the Insecta are diversity and numerical superiority. Of the 5 to 30 million

species estimated to compose the global flora and fauna, approximately 1.7 million have been named, more than half of them being insects. It is estimated that insects make up 75% of the known animal kingdom.

Because insects occupy almost every conceivable terrestrial niche, they interact with humans in countless ways that accord them status as "pests." This same diversity bestows on insects essential roles in the functioning of the biosphere as a sustainable biological system. Considering the countless interactions between humans and insects, it is not surprising that insects have become fixed in the fabric of human culture. They have become important components of our art, language, literature, music, philosophy, and religion. In addition, insects are remarkable sources of knowledge, ideal models for the study of biological processes including genetics, physiology, and molecular biology.

Professional entomologists find challenge in our universities where they engage in teaching, in conducting research to advance knowledge, and in extension, applying knowledge to the solution of applied entomological problems. In addition to professional entomologists, amateur naturalists are drawn to the study of insects because of their form, color, and behavior. As the expanding human population continues its modification of the natural habitat, the interface between humans and insects will become more problematic.

ENTOMOLOGICAL ROOTS

A mere 10,000 years ago, primitive hunter–gatherers made a great leap forward; they entered into partnership with plants, and agriculture was born. Thereafter, humans would seek to alter ecosystems to their own advantage. They would intervene to favor some plants and animals over others, thereby altering the evolutionary process that had shaped the biological world for the preceding 4 billion years.

Insects in their coevolution with plants and animals had been a powerful force in shaping the biosphere. They posed a primary threat to humans' alteration of ecosystems to provide food, fiber, shelter, and comfort for themselves and their domestic animals. The struggle that followed moved through stages of ignorance, myth, religion, and then enlightenment through science and technology.

Our brief historical sweep will jump the seemingly long sleep of ancient civilizations and go to the Greek civilization in the time of Aristotle (384–322 B.C.). The orderly study of biology began with his speculations. He relied on his own observations, defined the field, posed the questions, and accumulated evidence to answer them. Aristotle's vision of rationality lay dormant for centuries until the Renaissance. In the meantime, Judaism and Christianity imposed a new concept, one focused on God and creation as depicted in the book of Genesis. Accordingly, God created the world, directing Man to "be fruitful and multiply…" and "with dominion over every living thing that moves upon the earth" (Genesis 1:28). Man was not part of nature. Nature was subservient to Man.

The Scientific Revolution of the 16th and 17th centuries marked the beginning of modern science and included mathematics, mechanics, and astronomy but had little impact on biology. While the revolution rejected superstition, magic, and the dogma of medieval theologians, it did not reject the ideological bias of the Judeo–Christian religion. The hand of God was still directing the course of the natural world.

Not until the 17th and 18th centuries was entomology advanced as a field of study within zoology. Anton van Leeuwenhoek (1632–1723) of Holland used the microscope to extend the power of the human eye. He was obsessed with the study of detail, including the morphology and specialized organs of insects. His revelations established insects as proper subjects for scientific study. Francesco Redi (1626–1697) of Italy demonstrated in 1668 that insects arose, not from spontaneous generation, but from eggs laid by fertilized females. Jan Swammerdam (1637–1680) of Holland did superb anatomical work on insects, including the honey bee.

The excitement of these discoveries was further enhanced by the flow of exotic plants and animals brought back from voyages to other continents. Charles Darwin's voyage of *HMS Beagle* in 1831–1836 followed this tradition. The wealth of material acquired made students aware of the need to classify the organisms collected and to assemble specimens in orderly collections. Other investigators focused on the activity of insects in the field and their role as pollinators and as agricultural pests.

A prerequisite to advancing the study of insects was the development of a classification system that would bring order out of chaos. The Swedish naturalist Carl Linnaeus (1707–1778) met this need. Although trained in medicine, he studied botany extensively and turned to the classification of plants, animals, and minerals. His *Systema Naturae* (10th ed., 1758) is still regarded as the foundation stone of zoological nomenclature. He greatly simplified insect classification by using insect wings (hence the suffix *-ptere,* meaning wing, for most order names) as the basis for classification. The other great feature of his system was the designation of genus and species each by a single word, thus providing a binomial system to replace the unwieldy descriptive names employed earlier. Linnaeus's "artificial" system of insect systematics based only on wings was in time modified by adding other characters to construct a "natural" system.

Another great naturalist, René Antoine Ferchault de Réaumur of France (1683–1757), infused a new perspective into the emerging study of insects. He deplored the confusion that existed regarding metamorphosis, distribution, and "industries" of insects. He championed the study of insects out of sheer curiosity, claiming that useful discoveries would be made in the process. His six volumes of *Memoires pour Servir a l'Histoire des Insectes* (1734–1742) with their exacting attention to morphology and function, complete with accurate drawings, established a new standard of excellence.

The work of Linnaeus and Réaumur provided the templates for orderly classification and elucidation of fundamental and applied aspects of entomology. Their works

were extended and refined by French naturalists Pierre André Latreille (1762–1833), Georges Cuvier (1769–1832), and Jean Lamarck (1744–1829). By the 19th century, entomology was firmly established in European zoological science. The taxonomic treaties established in this process were to provide the guides to the classification of American insects. These sources were augmented by two sources in Great Britain: Gilbert White's (1720–1793) *The Natural History and Antiquities of Selborne* (1789) and William Kirby (1759–1850) and William Spencer's (1783–1860) *Introduction to Entomology* (1816–1826). The writings of these field naturalists on the biological characteristics of insects made insects, at one point, the most popular component of natural history in Victorian England. In addition, they contributed to the development of the biological species concept. This concept, essential to the understanding of biological communities, recognizes a species as a reproductively (genetically) isolated group of inbreeding populations.

The next step in the unfolding of the biological sciences was a giant one: the publication, in 1859, of Charles Darwin's (1809–1882) theory, *On the Origin of Species.* This event placed conceptual biology in a new light. In a single stroke, Darwin's work challenged the natural theology that had dominated biological thought for 3 centuries. Natural theology had been elaborated in John Ray's (1627–1705) *The Wisdom of God Manifested in the Works of the Creation* published in 1691. The concept provided a truce between science and religion. It contended that God created the world and the evidence of His omnipotence was to be found in the study of His creatures.

There was no middle ground between Darwin and Ray. Darwin provided a new way of viewing biology. The living world had evolved; it could be explained on the basis of descent with change. It was noteworthy to entomologists that much of Darwin's supporting evidence was derived from his study of insects dating from his days at Cambridge University (1828), where he was an avid insect collector.

Intense debate followed the publication of Darwin's theory. Nowhere was the debate more intense than at Harvard University, where Asa Gray (1810–1888), a botanist and staunch Darwinian, challenged Louis Agassiz (1807–1873), the foremost naturalist of the world and unrelenting defender of the creationist view.

An unlikely pair of outspoken individuals led the pro-Darwinian entomologists of North America. Benjamin D. Walsh (1801–1869), who had collected beetles with Darwin at Cambridge University, emerged from obscurity on the Illinois frontier to assert his long dormant entomological interest and declare his support for Darwinism. He was joined by the youthful Charles V. Riley (1843–1895), a fellow Englishman and self-taught entomologist who was then writing on entomology for the *Prairie Farmer,* the leading farm journal of the Midwest (Fig. 1). Their early collaboration on the issue of Darwinism bode well for the future. In time, evolution was accepted by biologists as a fundamental basis of the discipline, although segments of the public still oppose it as a challenge to the Creation story as reported in the Old Testament book of Genesis. The distinguished geneticist, Theodosius Dobzhansky (1900–1975), summarized the views in the major fields of biology in his essay entitled, "Nothing in Biology Makes Sense Except in the Light of Evolution" (1973).

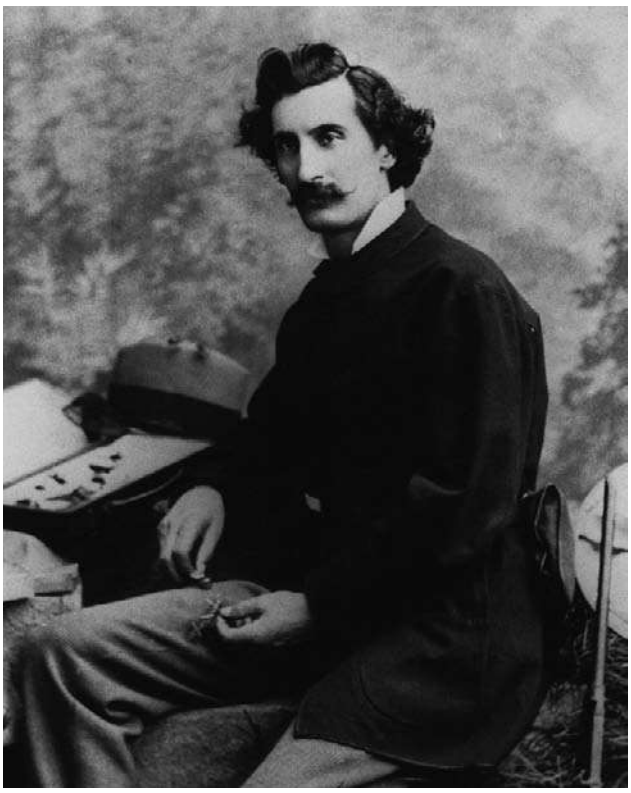

FIGURE 1 Charles V. Riley's nine annual reports as the state entomologist of Missouri (1868–1876) figured prominently in the knowledge base on which the philosophy of insect control in North America was based.

ENTOMOLOGY IN THE NEW WORLD

Nurturing Environment and Supporting Institutions

EMPIRE AND INSECTS Early entomological developments in the United States occurred in the climate of Thomas Jefferson's America. Jefferson took office as the nation's third president in March 1801 and proceeded to sell his vision to his countrymen, then numbering just over 5 million. The United States had only recently gained its independence from England. Louisiana, the land stretching from the Mississippi to the Pacific Ocean, although claimed by Spain, was available and being considered by Russia, France, and England. Of these three, France, energized by Napoleon, was feared the most. Despite the tenuousness of the situation, Jefferson clung to his vision of this vast land mass, stretching from sea to sea, being united under a stable government of the United States. Napoleon's agreement to sell Louisiana came as a

surprise; its purchase, although not applauded by some leading politicians, was a masterful stroke. It doubled the land mass of the United States and resolved the feuding over control of the Mississippi River. But what had been purchased? To answer that question, Jefferson dispatched an expedition, led by Captain Meriweather Lewis and Lieutenant William Clark (1804–1805). Commissioned to explore the new acquisition, their report, on their return, removed some of the mystique of the Pacific Northwest but provided only tantalizing glimpses of the area's natural history. Unfortunately, the expedition included no trained naturalists.

The follow-up to Lewis and Clark's report came in 1819–1820. Major Stephen H. Long, under authorization of President Monroe, led an expedition of "Gentlemen of Science" to study this vast unexplored territory, its natural history, and the American Indians of the Rocky Mountain region. Entomology was well represented by Thomas Say, whose affiliation with the Academy of Natural Sciences of Philadelphia had won him the reputation as "perhaps the most brilliant zoologist in the country." Say epitomized the confidence and vision of the young nation's leaders in natural history; he was tall, handsome, and resplendent in the uniform of the Long Expedition, with its high-collared, gold-buttoned, green jacket and black military trousers.

Although Say's primary interest was insects, he covered the entire field of botany and zoology and conducted studies on the American Indians. His entomological studies provided the foundation for his *American Entomology,* published in three volumes (1824, 1825, and 1828). These were the first books on North American insects. They were beautifully illustrated by the artistic talents of his wife, Lucy, and Titian Peale of the distinguished family of Philadelphia artists. They provided a stimulus for American entomologists and signaled their emancipation from the European centers to which the study of American insects had previously been consigned, named by Europeans, and retained in their collections. With the spirit of the Revolution still vibrant, consigning American insects to European collections offended national pride.

The Long Expedition imparted an American quality to the study of the nation's natural history. The Gentlemen of Science who manned the expedition were trained in the nation's centers of learning. They were not closet naturalists; they were adventurers. Their spirit was described by Say himself: "If our utmost exertions can perform only a part of a projected task, they may, at the same time, claim the praise due to the adventurous pioneer, for removing the difficulties in favor of our successors" *(American Entomology).*

THE CULTURAL CENTERS Natural history found strong support in the cultural centers of Philadelphia, Boston, and New York. Philadelphia led the way, with the American Philosophical Society and the influence of the distinguished Benjamin Franklin (1706–1790). The Academy of Natural Sciences, founded in 1812, nurtured the founding of the Entomological Society of Philadelphia in 1859, which in turn launched the *Practical Entomologist* in 1865.

Boston looked to Harvard College and the Massachusetts Society for Promoting Agriculture. William D. Peck (1763–1822) was appointed Professor of Natural History at Harvard in 1805 and offered the first lectures in entomology in North America. Thaddeus William Harris (1795–1853), a physician turned Harvard librarian, found time to become entomological author, teacher, collector, and correspondent. His report on *Insects Injurious to Vegetation* (1841) summarized the knowledge of insect control in Europe and North America, earning him the title, "Father of Economic Entomology."

New York asserted its interest by appointing Amos Eaton and John E. LeConte to the Lyceum of Natural History of New York. Other distinguished leaders included John Abbot (1751–1840), Thomas Say (1787–1834), and Frederick V. Melsheimer (1749–1814). A striking feature of these men and their institutions was their support of both classical and applied entomology. The individuals were well trained by the standards of the day, often completing training in medicine or theology, because there was no specific training in entomology.

Systematics was the primary entomological interest, followed by aid to agriculture, which was beset with countless insect pests. These leaders experienced the frustration of gaining access to the European literature, founding periodicals for the publications of their own findings, and establishing reference collections.

In the 1840s, American entomologists turned to the task of establishing the institutions that would sever their European dependence. The institutional framework took shape rapidly, led by the American Association for the Advancement of Science (AAAS), founded in 1847. It marked a transition from amateur to professional status, provided a national scientific forum, and nurtured the founding of professional societies. Within approximately 2 decades, 1859 to 1881, five additional societies were established in North America: the Entomological Society of Philadelphia (1859), the Entomological Society of Canada (1862), the Cambridge Entomological Society (1874), the Brooklyn Entomological Society (1872), and the Entomological Club of AAAS (1872). The institutional framework was now in place to expand the scientific and technical dimensions of entomology.

STATE AND FEDERAL ACTION IN THE UNITED STATES

Agriculture held the key to moving the nation from an agrarian to an industrial society. The farmer was viewed as the noblest and most independent man in society. Unlike in Europe, the availability of fertile soil was seemingly unlimited. While great physical obstacles lay in the path of progress, the greatest was the limits of the human intellect. With these elements in the national outlook, it followed that state and federal action would augment the private efforts in support of agriculture and entomology. In 1854, two landmark appointments were made: Townend Glover was appointed to

the Federal Patent Office for work in the newly established Bureau of Entomology, and New York State, responding to pressure from the New York State Agricultural Society, appointed Asa Fitch as its first state entomologist. Illinois and Missouri followed suit in 1868 with the appointments of Benjamin D. Walsh and Charles V. Riley, respectively.

These appointments represented historic landmarks, because state and federal funds were appropriated in support of agriculture with entomology in the vanguard. These men were able individuals whose evangelical zeal and sound professional grounding were attuned to national goals. Their publications, with Charles V. Riley's nine Missouri reports forming the core, laid the foundation for applied entomology in North America.

National goals for agriculture led to enabling federal legislation in three steps. First, the Morrill Land Grant Act of 1862 provided grants to each state, the proceeds from which funded a college, "to teach such branches of learning as are related to agriculture and the mechanic arts. . . ." A research dimension was added in 1887 by the provisions of the Hatch Act, a state experiment station being added to each college and coordinated by a central office in the Department of Agriculture in Washington, DC. The events cited in the foregoing, occurring within approximately 3 decades, provided an impetus for applied entomology that was unprecedented in the world.

Cooperative Extension, the outreach arm of the Land Grant University, which had been active from the start, was formally recognized and funded by the Smith-Lever Act of 1914. This institutional framework with its catalytic feedback from teaching, research, and extension has been recognized as one of the greatest educational innovations of all time.

With economic entomology rapidly expanding under the stimulus of the experiment stations, Charles V. Riley, now Chief of the Bureau of Entomology, perceived the need for a national organization to advance the goals of economic entomology. His organizational abilities and partnership with his Canadian counterpart, James Fletcher, led to the establishment of the American Association of Economic Entomologists in 1889. At Riley's insistence, the association focused on economic entomology, leaving unmet the needs of the broader dimensions of biology, taxonomy, morphology, and faunistic studies of insects. In 1906, the Entomological Society of America was organized to meet these needs, with John Henry Comstock of Cornell University serving as president. With the various forces that shaped these professional and governmental institutions in mind, we can examine how the institutions responded to the challenge posed by insect pests.

Insect Pests

That the world is not awash in insects, despite their remarkable potential for reproduction, attests to the "balance of nature." But nature's balance, while avoiding extremes, does not preclude insect activity that is annoying to humans. Insects that take humans' crops or blood, and invade their dwellings, are termed "pests." The term has no biological significance; it only expresses a human perception.

Let us examine a few insect outbreaks that occurred in agricultural, medical, and veterinary entomology in the late 1800s. They were to test the mettle of the institutions crafted to address such problems. They taught us much about insects, ourselves, and our vast land mass with its unique biomes and punctuated with its geographical features: the Rockies in the west, the Appalachian range in the east, the Great Plains, the Great Lakes, and the Mississippi. It was this abundance of land that appealed to the early settlers from land-poor Europe. It was from the vastness of the land with its rich flora and fauna coupled with the democratic spirit of its people that the American dream was fashioned. However, the dream's social fabric was not matched by its concept of stewardship of the land. New England's forest primeval had to be breached to make way for agriculture. Conquering nature was viewed as a prelude to progress, and the American Indian and the bison fared poorly under this credo.

AGRICULTURAL ENTOMOLOGY The Colorado potato beetle, *Leptinotarsa decemlineata,* existed in the foothills of the Rockies on the buffalo bur, *Solanum rostratum.* As pioneer settlers pushed westward with their crops, the beetle colonized the cultivated potato, *Solanum tuberosum,* and began its eastern migration along the "potato trail." It was observed as a potato pest in Nebraska in 1859, reached the Atlantic coast by 1874, and traveled thence to Europe in 1876, where it remains an important pest.

The early search for control measures established the arsenical, Paris green, an industrial pigment, as an effective poison. It soon became the standard treatment and was the first widely used poison to kill by ingestion.

The early marketing of insecticides invited fraud through adulteration and false claims. It was not until 1910 that federal legislation was passed requiring labeling to reveal efficacy and ingredients in the two most widely used insecticides, Paris green and lead arsenate.

The boll weevil, *Anthonomous grandis grandis,* crossed the Rio Grande to Texas in 1894 and began its eastward trek, occupying the entire 1,500,000-km^2 cotton belt by 1925. Efforts to impede its progress by establishing no-cotton barriers failed for lack of community compliance. Countless control measures were tried but insecticides eventually won as the first line of defense. Calcium arsenate was adopted for control in about 1920, and its use soon reached 20,000 tons per year. This marked a new scale of area-wide pesticide treatment with its attendant environmental and human safety problems.

The social and economic impact of the weevil was incalculable. The prosperity of the south evolved around a single crop, cotton. With its loss, the economic infrastructure collapsed, and panic ensued. Black laborers left, mortgages were foreclosed, and banks failed. The potential for economic disaster in the wake of insect outbreak was seared in the

memory of the people of the cotton-producing states. Only the Civil War had greater impact on the economic and social life of the southern states than did the boll weevil.

The Rocky Mountain grasshopper, *Melanoplus spretus,* appeared in an epidemic eastward migration, borne on wind currents from the foothills of the Rockies to the Mississippi valley in 1874–1876. Presumably, this migration was in response to the agricultural disruption of the ecosystem that had been dominated by the American bison. The ravages of these hordes of airborne insects created a crisis for the affected states, whose governors appealed to Washington, DC, for federal intervention. In response, the U.S. Entomological Commission was created with the colorful Charles V. Riley as its chairman. This was not a staid Washington bureaucracy; it was a mobile force that reached out to the crisis whenever it arose. Riley scoffed at Sundays devoted to prayer for divine intervention to restrain the pest. Rather, he urged the people to adopt control measures based on intricate knowledge of the life history of the pest. His insights led to bold predictions of the pest's demise from natural causes. With some sound observations and a modicum of luck, his predictions held. The Commission, despite its denials, was credited with solving the problem, bringing new credibility to entomologists and federal aid to the states.

The gypsy moth, *Lymantria dispar,* was introduced from France, not by accident but by design, by Leopold Trouvelot (1827–1895), a Harvard astronomer and amateur entomologist, who was interested in silk-producing moths. Larvae, emerging from egg masses he had imported, escaped from his Medford, Massachusetts, residence in 1869. After a period of 20 years, the moth reappeared in an epidemic outbreak, it having been mistakenly overlooked as a native species. With this head start, the scorched earth practice of cutting and burning infested trees, augmented by arsenical sprays, failed. The effort did stimulate advances in the technology of spray machines. Today, this introduced pest has spread westward and southward, occupying a swatch from the Great Lakes to the Carolinas.

Biological control was enthusiastically touted following the spectacular success achieved by Charles V. Riley's innovative introduction of the Vedalia, *Rodolia cardinalis* (Coleoptera), into California to destroy the cottony cushion scale, *Icerya purchasi.* The pest had been introduced from Australia about 1868 and soon threatened destruction of the state's citrus industry. Two years after the introduction of the predator, the pest was miraculously under control.

The whole array of control measures, cultural, mechanical, chemical, plant resistance, and biological, was employed in seeking to cope with these problems. The entomologists were influenced by expectations and perspectives of their farmer clientele. The farmer's time frame was established by harvest date and sale of the crop; his risk tolerance was low. Insecticides provided immediate and predictable results and they became the backbone of control programs.

Meanwhile, several factors intensified the pressure for insect control, including monoculture, susceptible crops, exacting market standards, and introduced pests; all required greater intervention and modification of the agroecosystem.

MEDICAL AND VETERINARY ENTOMOLOGY The foundation for modern medical and veterinary entomology was laid by Louis Pasteur, a French microbiologist, who formulated the theory of microbial causation of disease based on his work with the silkworm *Bombyx mori* in 1887. Without benefit of the germ theory, Josiah Nott, a Mobile, Alabama, physician, proposed (1848) that the causative agents of malaria and yellow fever were transmitted by mosquitoes. In 1881, Carlos Finlay, a Cuban physician, postulated that mosquitoes transmitted the yellow fever agent, setting the stage for Major Walter Reed and associates to verify his claim. In 1897, Ronald Ross demonstrated the occurrence of the malaria parasite in mosquitoes that fed on a human patient whose blood contained the parasite, thus leading to the elucidation of the epidemiology of malaria.

In 1889, Theobold Smith discovered the causative agent of Texas cattle fever and, working with F. L. Kilbowen, showed in 1893 that the cattle tick, *Boophilus annulatus,* was the vector. Their work paved the way for tick prevention and development of the cattle industry in the southern United States.

The experiences cited above in control of insects of agricultural, medical, and veterinary importance were unprecedented in the American experience and left no region untouched. They revealed the social, political, biological, economic, and environmental dimensions of insect problems. A nation leaning so heavily on agriculture was sensitive to the impact of these problems on the nation's well-being.

The fundamental principles gleaned from these experiences were to shape the philosophy of insect control for the future. They included the following: (1) taxonomic knowledge of the vast insect fauna is a prerequisite for detection and development of control programs; (2) advances in international commerce breach the ancient oceanic barriers to the dispersal of insects and increase the likelihood of introducing exotic species; (3) introduced species, uninhibited by their natural controls, often become major pests in their new habitat; (4) the economic well-being of vast regions of the nation is vulnerable to insect attack; (5) intervention at the federal level is required for insect problems beyond the scope of individual states; (6) alterations of ecosystems trigger changes in patterns of insect behavior; (7) the use of insecticides requires federal regulations to protect the user, the public, and the environment; and (8) an informed public will underwrite sound programs of insect control.

PARTNERSHIP IN PEST CONTROL, 1880 TO WORLD WAR I

The institutional framework for colleges of agriculture was well established by 1880. The first department of entomology was founded at Cornell University in 1874 under the leadership of John Henry Comstock; others followed shortly.

The primary objective of the department was to train students who wished to become farmers, to identify and classify the insect fauna, to study life cycles, to devise control measures, and to train farmers in their use.

As insecticides became a more important component of production technology, an alliance of increasing importance developed among the agricultural constituency, the agricultural colleges, and the chemical industry. As this partnership developed, agribusiness expanded to provide the goods and services required for agricultural production and grew even more mechanized, technical, and capital intensive. The three partners shared a common objective, pooling their resources to increase the efficiency of agricultural production that accrued ultimately to the benefit of the consumer.

The arrangement involved the agricultural constituency, lending political support to the agricultural colleges in return for their services. The colleges then aided the chemical industry by testing their products and giving their stamp of approval, which enhanced their marketability. A grateful chemical industry provided grants to the entomology departments, which were always short of operational funds. The deans at the agricultural colleges had the difficult task of being broker between the college faculty, with its leaning toward basic research, and the farm constituency seeking low-risk pest control programs. The arrangement was an American innovation that seemed to please everyone. Furthermore, the chemical industry was greatly stimulated by the economic and political activities of World War I. Food and fiber production was given high priority and new discoveries advanced the pesticide industry.

The period from 1880 to 1940 witnessed the maturing of the Agricultural Experiment Stations as a national, highly coordinated network. The extension entomologists became the connecting link between the agricultural college and the agricultural producer. Because insecticides had become the first line of defense, the growing chemical industry added strength to this already solid partnership with the colleges and farmers in the aftermath of World War II. Furthermore, this was the threshold of an era of discovery of new molecules that would affect biological processes of plants and animals.

Along the way some ominous straws in the wind signaled trouble ahead. In 1913, The San Jose scale, *Quadraspidiotus perniciosus,* developed resistance to lime sulfur but the phenomenon was not recognized as an expression of Darwinian selection. In 1928, the codling moth, *Cydia pomonella,* was shown to be resistant to lead arsenate. The apple industry was dealt a severe blow in the mid-1930s when British markets rejected fruit from the United States because of high arsenical residues. Simultaneously, fruit trees were showing loss of vigor because of insecticidal toxicity to foliage and accumulation of residues in the soil.

Although these disquieting revelations were not widely publicized, they were of concern to entomologists, as they and the pesticide industry were marshaled to meet the greater demands for food and fiber required for World War II. Even then, the sustainability of insecticidal control was clearly in doubt.

ENTOMOLOGY, POST-WORLD WAR II

Technology's Triumph

The explosion of the atom bomb over Nagasaki in 1945 brought a dramatic end to World War II and in so doing, highlighted the role of science and technology in the victory. Following this, the age-old ritual of "beating swords into plowshares" turned scientific and technical advances to peaceful ends.

No field emerged with more exciting prospects than did the field of entomology. DDT, with its wartime secrecy removed, was hailed as the answer to insect control. Its employment in arresting an epidemic of typhus in Naples in 1943–1944 dramatically neutralized the lethal companion of armed conflict, vector-borne disease.

Overnight the entomological community documented DDT's remarkable effectiveness in controlling insect pests of agricultural, medical, and veterinary importance. The race was on, and an old alliance assumed new vigor. The Land Grant Universities joined with industry and agriculture to exploit the new possibilities of chemical pest control.

Although industrial grants to the Agricultural Experiment Stations to fund trials of mutual interest dated from the early 1930s, they assumed a greater role in Experiment Station research as the partnership geared up for a new era in the synthesis of pesticides. The chlorinated hydrocarbons, with DDT their prototype, yielded related compounds followed by the development of organophosphates, methyl carbamates, and pyretheroids, all neuroactive chemicals. By the 1950s, post-WWII insecticides had become the mainstay of insect control, with the prewar calls for biological and cultural controls in eclipse.

Professional Societies

Professional societies perform important functions. They establish the ethical and intellectual standards of the discipline, provide liaisons with the scientific community and the public, and provide continuity and the written record. In North America, three major professional societies have served the entomologists of Canada, Mexico, and the United States: the Entomological Society of Canada, La Sociedad Mexicana de Entomologia, and the Entomological Society of America (ESA). Today, the membership in these three societies stands at about 8000. The new challenge to entomological societies will be to hold to their traditional goals while embracing the environmental and biodiversity crusades, advancing IPM, and increasing interdisciplinary collaboration. Because the ESA is the largest of the three societies and enjoys substantial joint membership from the other two, we will focus on its recent history.

As the scope of applied and basic entomology grew under the stimulus of post-World War II goals, the two primary entomological societies in the United States, the American Association of Economic Entomologists (AAEE) and the ESA, recognized two common needs. They needed professional

management to better process publications and provide other services for their members. In addition, professional entomologists needed to address the big policy issues posed by the growing importance of the biological sciences in human affairs. Big science was emerging as a national goal, and the landscape was changing physically. For example, the stately elms were vanishing, casualties to an insect-vectored disease. The gypsy moth and the imported fire ant were on the move. The Green Revolution had captured the imagination of the great humanitarian organizations. Therefore, entomology was being summoned to new levels of leadership.

A merger of the two organizations seemed to offer advantages in both management and unity. This view was not unanimous, particularly among ESA members, who were outnumbered three to one by AAEE. After protracted debate and two ballots, the two societies merged in 1953 as the Entomological Society of America.

It soon became evident that although the new ESA gained administrative efficiency, it was not unified in its philosophy and the more numerous and vocal members oriented to applied entomology predominated. Stress arising from this dichotomy within the membership has characterized the ESA over the nearly 5 decades since the merger.

FIGURE 2 Rachel Carson's *Silent Spring* (1962) focused public attention on the pesticide issue. Her crusade catalyzed the environmental movement. (Harbrace photograph.)

Enter Rachel Carson

Early warnings of the danger of insecticide mania were sounded within the entomological community, but these warnings were largely ignored. It was the publication of Rachel Carson's *Silent Spring* in 1962 that triggered the avalanche of public concern (Fig. 2). She lamented that "so primitive a science has armed itself with the most modern and terrible weapons, and that in turning them against the insect, it has turned them against the earth."

Overnight, her exhortation changed the public's perception of entomologists. Their traditional obscurity was swept away. They were in the public eye, viewed as allies with the corporate giants, poisoners of robins and the earth, all under the pious veil of aiding the consumer by aiding the farmer.

Response to Rachel Carson's charges came largely from industry, whose strategy was to discount the witness. This proved ineffective as the accumulating evidence reinforced her concerns. Practices that endangered birds, especially the national symbol, the American eagle, were certain to stir emotions.

The public debate on the pesticide issue did not reach the agenda of the Entomological Society of America. The Society formed by merger in 1953 had not become a forum for debate on issues concerning the field. Instead, plenary sessions at the Society's annual meetings were largely ceremonial with substantive debate occurring in the subject matter sections representing the various areas of specialization among their members. In short, there was no philosophical common denominator that united the Society.

Carson's *Silent Spring* became a corner stone of the environmental movement. Her thesis focused on the "web of life," which placed humanity's relationship to all forms of life in an ecological context. The concern for the environment and the natural world triggered by *Silent Spring* melded with other concerns for humans—women's rights, the war in Vietnam, and Native American rights—to give rise to the broadly based environmental movement embracing the rights of humans and nature, animate and inanimate. This great philosophical debate extending over the past 4 decades proceeded without the active involvement of the entomological community.

One of the immediate effects of *Silent Spring* was to make pesticide policy a matter of public debate. While focusing on DDT, the issue became broader and embraced the central tenet of the environmental movement: that human intervention had become the dominant environmental influence on the planet. The fact that, despite the restricted use of DDT, its residues could be found in Antarctica implicated entomologists in the global insult.

In 1967, a group of concerned individuals formed the Environmental Defense Fund, its object being to use litigation in defense of citizens' right to a clean environment. In protracted public hearings, entomologists were called to testify that their practices were not infringing on citizens' rights to a clean environment. It was an uncomfortable defensive position in which these dedicated "defenders of agriculture" were placed. In 1972, a decade after *Silent Spring,* the Environmental Protection Agency banned DDT. Its meteoric rise and fall, from discovery to banning, had spanned only 3 decades.

Economic entomologists in general viewed *Silent Spring* as an attack on their professional competence and integrity. Since the late 19th century, they had cultivated a self-image as dedicated public servants, bringing science "to the distressed husbandman" whose labors were closely aligned with the national interest. This explains in part their emotional response and sense of hurt that has lingered among entomologists of the DDT era.

Ecology's Promise

Economic entomologists of the 1960s faced two daunting challenges, the loss of public confidence in the aftermath of *Silent Spring* and the failure of their programs of insect control. These were powerful incentives for reassessment.

In the early 1950s, well before *Silent Spring,* the concerns regarding the insecticidal treadmill led a group of entomologists at the University of California at Berkeley and at Riverside to reassess control practices. Drawing on the biological control heritage pioneered by Harry Scott Smith (1883–1957) they sought to "integrate" features of biological and chemical control. This concept with further refinement led to the adoption by the late 1960s of IPM. In practice, IPM seeks to integrate multiple control measures into a cohesive package, the additive impact of which would hold pests within acceptable levels with minimum adverse environmental impact.

The abbreviation "IPM" was soon adopted worldwide to identify a holistic approach to pest control. Its enthusiastic reception reflected the optimism accorded a new paradigm, one that placed pest control on an ecological foundation.

The most impressive feature of the movement has been its evolving nature. The underlying theory and the fundamental question that has plagued population ecologists, "What factors determine the number and distribution of animals?" remain under debate. Views on the role of pesticides in IPM are likewise evolving. Progress has been made in tailoring pest-specific insecticides with reduced environmental disruption. Although impressive gains have been made in specific programs, the IPM era has not resulted in a major decline in the total quantities of pesticides used.

The euphoria induced by the IPM concept has run its course, and its promise after 3 decades is a subject of lively debate. One of the problems affecting acceptance and support of IPM is the difficulty of assessing its effectiveness. It is a complex system with many obstacles and a restricted database for evaluating a variety of constraints: technical, financial, educational, organizational, and social. Whatever the outcome, there is no turning back. The human intellect has been unable to construct a more promising strategy for keeping humans' age-old competitors in check. For millions of people threatened with disease and hunger, IPM constitutes their safety net for tomorrow.

In seeking to understand the strategy employed by applied entomologists, it is helpful to note historical perspective. Applied entomologists were late to embrace ecology despite the entreaties of their distinguished president of the Entomological Society of America in 1912, Stephen A. Forbes. He insisted that "the economic entomologist is an ecologist pure and simple whether he considers himself so or not." In retrospect, it appears that entomologists opted for the certainty of insecticides favored by their farmer clientele over the uncertain promise of ecology. Their adherence to the conventional wisdom of insecticidal control in the DDT era tarnished their image as environmentalists.

In medical and veterinary entomology, the post-World War II experience with the miracle insecticides paralleled the experience with agricultural pests. First, there was euphoria following the miraculous effectiveness of the insecticides. So promising were the prospects that in 1955 the World Health Organization (WHO) proposed global eradication of malaria. However, the development of resistant strains of vectors and parasites as well as economic and political factors doomed the eradication program. In 1976, the WHO abandoned eradication in favor of more modest programs of control. Research languished under the demoralizing effect of this decision. With antimalarial drugs and insecticides losing their effectiveness, the battle against malaria was being lost. Alarmed at these developments, the WHO, in 1993, called for a renewed global effort. The initially slow response has gained support, with unprecedented funding available in 2000 for new initiatives.

An ambitious objective is the development of a vaccine against malaria. Although the scientific obstacles are enormous, researchers are now predicting a successful vaccine by about 2010. The most ambitious and futuristic of all approaches to combating malaria is creating a strain of *Anopheles gambiae* mosquito unable to transmit the parasite. To displace the native vectors involves three steps: find genes that interrupt the parasite's life cycle, develop techniques to transfer those genes into the mosquito, and finally, develop ways of replacing existing mosquito populations with the genetically engineered model. One additional hurdle remains. With such a mosquito in hand, there may be strong resistance to releasing such transgenic forms into nature. In the meantime, the disease continues to cast its shadow over the malaria-endemic areas of the world, which are home to 40% of the world's population.

Advancing the Science

The past 5 decades have witnessed remarkable advances in both applied and basic entomology. The collapse of chemical control of insects forced reassessment, which gave rise to IPM. While the pesticide issue dominated public interest, basic science was forging ahead.

The primary stimulus was the discovery in 1953 that the compound deoxyribonucleic acid (DNA) encodes genetic information that provides the blueprint for synthesis and cellular differentiation; this discovery elucidated the great mystery of life, the cell's ability to self-replicate. Many aspects of biology were catalyzed by the discovery. It dramatically reaffirmed Darwin's hypothesis of common descent and

revealed evolutionary pathways. The studies of molecular systematics that followed have resulted in accumulation of much DNA sequence data from most insect groups. These data complement and enhance the morphological and ecological data of classical systematics, thereby making substantial contributions to evolutionary biology.

The DNA breakthrough also paved the way for biotechnology, the introduction of genes from various species into plant and animal species. With biotechnology, plants can be engineered to produce their own pesticides. Corn can be altered to contain a pesticide produced by the bacterium, *Bacillus thuringiensis,* with such corn being designated "Bt" corn.

This technology can place the cornucopia of biodiversity in the service of agriculture and medicine. But this novel technology comes with complex ethical and scientific issues in environmental stewardship and human health. The scientific community as well as the general public is in strong disagreement over the introduction of genetically modified organisms (GMOs) into the ecosystem. Caution should prevail until some basic questions are answered. For instance, what impact will the thousands of acres of Bt corn have on the complex of beneficial and injurious insects on the modified plants?

GMOs conferring drought tolerance to food crops of developing countries would be welcomed additions as the food supply grows more tenuous with rising populations and civil disruption. In weighing these options, it needs to be recognized that an infrastructure to monitor such crops is not in place at present. The role of GMOs in IPM is likewise uncertain in developed countries, and resistance to GMOs is particularly acute in European countries.

Industry has been quick to recognize biotechnology's commercial potential, and substantial segments of the seed market have been given over to GMOs. Questions of a scientific nature are joined by social and economic questions. For instance, should corporate interest determine and control the genetic profile of the three crops, corn, rice, and wheat, that provide sustenance for most of the peoples of the world?

The economic issues surrounding GMOs seem to overshadow the more basic environmental issues they pose. For instance, the biodiversity program seeks to conserve natural forms, whereas biotechnology seeks to replace natural forms with modified ones exempted from evolutionary testing. Over time, how will this practice affect the gene pool, the timeless and priceless biological resource?

Although elucidation of the structure and function of DNA is clearly the most important discovery of the 20th century, other discoveries have greatly advanced our understanding of insects. This progress is due in large measure to technical advances in fields such as insect olfaction, acoustics, flight, and communication (e.g., by pheromones).

Such advances have in turn altered the way scientists communicate in person and in professional literature. They became more informal and democratic. The excitement was often centered in youth, in graduate students, with women strongly represented.

The excitement of discovery and exuberant professional exchange produced masses of data leading to new specialized journals. The worldwide computer network has catalyzed the processing and exchange of data among colleagues on a global scale. The predominant use of English in scientific journals has reduced language barriers. Thus, in both applied and basic entomology, the latter half of the 20th century has represented a new order, new methodologies, new discoveries, and new organizational arrangements. The paths of progress in the multifaceted phases of entomology are well documented in the *Annual Review of Entomology,* published since 1956.

Historical Perspective

The preceding 2 centuries of entomological enterprise in North America have been directed primarily to two activities: (1) protecting humans' food, fiber, and health and (2) basic research to advance knowledge of insects. These were and continue to be appropriate objectives in the national interest.

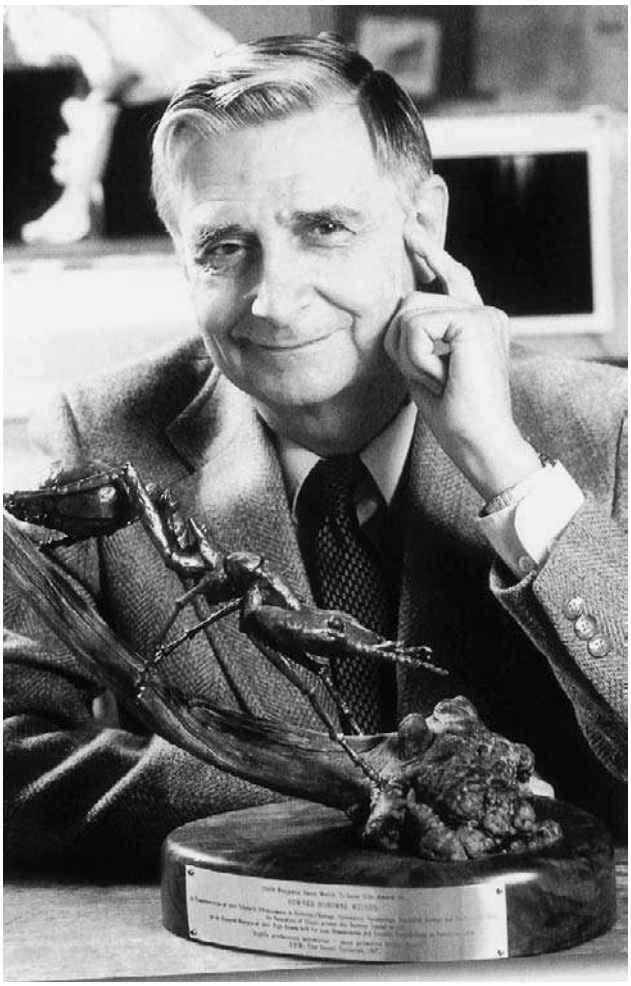

FIGURE 3 Edward O. Wilson's prolific writings advanced the science of entomology, established the field of sociobiology, and led the 21st century movement to preserve biodiversity.

In the past half century the environmental movement and the emerging science of ecology have highlighted two salient points: (1) insects play a vital role in the sustainability of the global biosphere and (2) the biodiversity essential to sustainability is threatened by human intervention. The factors of habitat destruction, pollution, and introduction of exotic species are believed to account for extinction rates much higher than before the coming of humans. In addition, the ecological impact of global warming looms on the horizon. The movement to preserve biodiversity has been led by E. O. Wilson (Fig. 3) following publication of his *The Diversity of Life* (1992). The concept has been generally accepted and is now part of the American culture.

The extinction dilemma poses new challenges to the field of entomology and calls for modification of the image entomologists hold of themselves and of the institutions established in the past to deal with entomological matters. The new order calls for entomological statesmanship that looks beyond entomology's traditional agricultural constituency to the global environmental issues. Thus, the age-old challenge of insect control will be joined with the challenge of insect conservation.

See Also the Following Articles

Agricultural Entomology • Biological Control • Entomological Societies • Extension Entomology • Integrated Pest Management • Regulatory Entomology

Further Reading

"Annual Review of Entomology" (1956–2003). Vols. 1–48. Annual Reviews, Palo Alto, CA.

Berenbaum, M. R. (1995). "Bugs in the System, Insects and Their Impact on Human Affairs." Addison–Wesley, Reading, MA.

Dunlap, T. R. (1981). "DDT: Science, Citizen and Public Policy." Princeton University Press, Princeton, NJ.

Howard, L. O. (1930). "A History of Applied Entomology," Vol. 84. Smithsonian Inst., Washington, DC.

Mallis, A. (1971). "American Entomologists." Rutgers University Press, New Brunswick, NJ.

Mayr, E. (1982). "The Growth of Biological Thought. Diversity, Evolution, and Inheritance." Harvard University Press, Cambridge, MA.

Paladino, P. (1996). "Entomology, Ecology and Agriculture: The Making of Scientific Careers in North America, 1885–1985." Harwood Academic, Amsterdam.

Perkins, J. A. (1966). "The University in Transition," p. 43. Princeton University Press, Princeton, NJ.

Perkins, J. H. (1982). "Insects, Experts, and the Insecticide Crisis." Plenum Press, New York.

Smith, R. F., Mitter, T. E., and Smith, C. H. (eds.) (1973). "History of Entomology." Annual Reviews, Palo Alto, CA.

Sorensen, W. C. (1995). "Brethren of the Net: American Entomology, 1840–1880." University of Alabama Press, Tuscaloosa.

Stein, B. A., Kutner, L. S., and Adams, J. S. (eds.) (2000). "Precious Heritage: The Status of Biodiversity in the United States." Oxford University Press, Oxford.

Whorton, J. (1974). "Before Silent Spring: Pesticides and Public Health in Pre-DDT America." Princeton University Press. Princeton, NJ.

Wilson, E. O. (1992). "The Diversity of Life." Harvard University Press, Cambridge, MA.

Wilson, E. O. (1994). "Biophilia." Harvard University Press, Cambridge, MA.

Worster, D. (1988). "Nature's Economy: A History of Ecological Ideas." Cambridge University Press, New York.

Homeostasis, Behavioral

P. Kirk Visscher
University of California, Riverside

Behavioral homeostasis refers to mechanisms of behavior that allow an insect or group of insects to maintain conditions within a certain range of values. These conditions may be the temperature of the body or the environment, internal water balance or environmental humidity, nutritional state or food stores, the balance between different activities of the individual or of the group, or the number and composition of individuals in a group. Behavioral mechanisms of homeostasis are important to individual insects, whether solitary individuals or part of a group, and include such nearly universal behaviors as feeding and drinking, as well as behavior concerned with thermoregulation and habitat choice. This article, though, is mostly concerned with homeostasis in groups of insects, such as the colonies of bees, wasps, ants, and termites. Individual behavioral homeostasis in physiological regulation, thermoregulation, and habitat choice are covered elsewhere in this encyclopedia.

ENVIRONMENTAL REGULATION BY GROUPS OF INSECTS

Insects are relatively small animals, with high surface-to-volume ratios. Because of this, they readily lose body heat or water to the environment (or gain heat if the ambient temperature is high). However, a few species of insects form large groups that are able to exert some control over these processes. The most striking examples of this come from the social insects (the wasps, ants, bees, and termites), but some other insects also form groups that enhance homeostasis (Fig. 1).

The control of groups of insects over heat exchange may take two forms. First, they may form a cluster that effectively makes them more similar collectively to larger organisms. If the surface-to-volume ratio is of a cluster of insects rather than an individual, it has a smaller value, and heat exchange is slower. Second, most social insects construct nests, and the architecture of these nests can result in the interior environment being substantially different from the ambient environment outside the nest.

Honey Bees

Honey bees exhibit both of the above strategies. Honey bees (*Apis* spp.) arose in the tropics, but *A. mellifera* and *A. cerana*

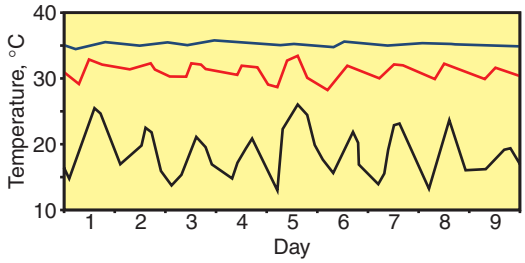

FIGURE 1 Honey bees (*Apis mellifera,* blue line) and yellowjacket wasps (*Vespula vulgaris,* red line) both maintain their nests at temperatures that fluctuate less than outside air temperatures (black line). At cool outside temperatures, as here, the nests are kept warmer than ambient. Note that the honey bee colony, with tens of thousands of workers, achieves more precise temperature homeostasis than the wasp colony, with only hundreds of workers. (Data from H. Kemper and E. Döhring, 1967, *Die Sozialen Faltenwespen Mitteleuropas,* Parey, Berlin.)

have colonized much of the temperate zone as well. These honey bees are unique among temperate insects in maintaining a high temperature in their nests throughout the winter, even when environmental temperatures are dramatically lower. For example, an *A. mellifera* colony can maintain a temperature in the center of its winter cluster inside the nest of 35°C, even when the temperature outside the nest is –40°C. The bees accomplish this by clustering together tightly so that the bees themselves, as well as the nest structure (often a hollow tree or wooden beehive) serve as insulation. The bees consume honey as metabolic fuel and contract their large flight muscles to create heat. Bees on the outside do get chilled, but they trade places with bees in the warm interior from time to time. Even when a bee colony is not in a nest, as when they are moving as a swarm to a new homesite, they maintain warm temperatures inside the cluster of thousands of bees.

The environment is not always cold, so temperature homeostasis for a bee colony sometimes involves cooling the nest. Honey bees fan their wings to move outside air through a colony to remove excess metabolic heat (and carbon dioxide). When this does not cool the colony enough, the bees begin to collect water and evaporate it within the nest to provide cooling. Also, when the nest becomes too warm, many bees leave the cavity and cluster outside the nest, reducing the heat input from their metabolism.

Termites

Many species of termites, like honey bees, live in large groups. Indeed, the largest colonies of social insects occur among the termites, some species of which may have several million individuals in a nest. Unlike honey bees, termite workers do not have wings, and so they cannot move air by fanning. Instead, some species of termites rely on the structure of the nest to regulate temperature and humidity. *Macrotermes subhyalinus* colonies, for example, construct tall "chimneys" on their nests. These chimneys are thought to increase airflow in two ways. As the metabolic heat of the termite colony (and the fungus gardens that they cultivate in the nest) warms the air in the chimneys, it rises and is replaced by cooler air from passages near the ground. Also, when wind blows across the open tops of the chimneys, the Bernoulli effect causes lower air pressure at the chimney top and draws air upward.

The climate-control nest structure of another species of *Macrotermes, M. bellicosus,* was described by Martin Lüscher. These termites build nests with a closed-circuit air circulation system. Air warmed by metabolic heat rises in central galleries of the nest but then enters channels on the outer ribs of the nest. Here, it loses heat to the outside through the nest material, and the denser, cooled air settles to chambers at the base of the nest, from which it is drawn to replace the rising air in the central nest, over and over again. As the air passes through the thin outer channels, carbon dioxide diffuses out and oxygen diffuses inward. This system allows gas exchange and cooling, while limiting water loss.

In Australia, *Amitermes meridionalis* nests are constructed as flat towers, always oriented with their long axis north and south. The result is that they are warmed by sun as it rises in the east early in the morning and strikes their broad side, but they receive relatively little sunshine at midday when the sun is in the north and strikes the nest edge on. These termites are known to sense the earth's magnetic field and use it to coordinate the nest-building activity of the colony's many workers to achieve this striking geographic orientation of the nest.

Tent Caterpillars

Although less organized in their social behavior than most social insects, tent caterpillars use some of the same thermal strategies to get a jump on the warm season. The larvae of tent caterpillars cluster together and form tents from silk that they produce. A group of caterpillars clusters together inside the tent during the night, where both the tent and the presence of many clustered insects reduce heat losses. The higher temperatures that the caterpillars experience allow them to develop more quickly than they would if they were isolated and exposed to the low temperatures that are common, especially at night, in their environment. Tent caterpillar behavior is adapted to keeping with the group. They find their way back to the tent by trails of odors and silk that are laid down as the caterpillars move from the tent to the foliage on which they feed during the day.

COMMUNICATION AND GROUP ACTIVITIES

An individual organism must allocate its time and resources between food collection, reproduction, habitat selection, and other activities. In the social insects, one sees similar behavioral adaptations. There is added complexity, though, because in social insects they occur at both the level of the individual and the level of the group. Group-level adaptations include

the regulation of numbers of individuals in the colony, the allocation of reproduction between workers and sexual forms, the division of labor among individuals (e.g., caste), and the social organization of food collection (e.g., recruitment). All of these homeostatic activities by colonies of insects require mechanisms of communication to coordinate the activities of multiple individuals. It is for this reason that the social insects provide so many of the examples of communication among insects, because in nonsocial insects, communication is largely restricted to behavior associated with mating or defense. As in other insect groups, much of the communication in social insects is carried on chemically, by means of pheromones.

Homeostasis is fundamental to the survival of organisms, because the processes of life occur in a well-regulated manner only within a certain range of conditions. The same could be said about the processes conferring advantages of group living on those insects that live in groups. If a colony is too large, or fails to coordinate its activities in foraging, reproduction, or defense, it may perish. It is the function of behavioral mechanisms of homestasis to regulate both the group environment and the properties of the group itself in a manner that preserves its efficient functioning.

See Also the Following Articles

Dance Language • Magnetic Sense • Nest Building • Recruitment Communication • Thermoregulation

Further Reading

Abe, T., Bignell, D. E., and Higashi, M. (2000). "Termites: Evolution, Sociality, Symbioses, Ecology." Kluwer Academic, Dordrecht.

Heinrich, B. (1993). "The Hot-Blooded Insects: Strategies and Mechanisms of Thermoregulation." Harvard University Press, Cambridge, MA.

Seeley, T. D. (1994). "The Wisdom of the Hive: Social Physiology of Honey Bee Colonies." Harvard University Press, Cambridge, MA.

von Frisch, K., and von Frisch, O. (1974). "Animal Architecture." Harcourt Brace Jovanovich, New York.

Wilson, E. O. (1971). "The Insect Societies." Harvard University Press, Cambridge, MA.

Homoptera

see *Auchenorrhyncha; Sternorrhyncha*

Honey

Eva Crane
International Bee Research Association

Honey is a sweet substance produced by social bees and some other social insects. They collect nectar or honeydew from living plants and transform it into honey, which they store in their combs. Most nectar collected by bees contains from 15 to 50% sugars; these are mainly sucrose in some plants, glucose and fructose in others, and all three in the rest. There are also many minor constituents.

THE ORIGIN OF HONEY

The Bees That Produce Honey and How They Do So

Certain social insects produce and store honey as a nonperishable food for use in dearth periods. The insects include all honey bees (*Apis* spp.) and stingless bees (*Melipona* and *Trigona* spp.) and also certain species of social wasps in South America (*Nectarina*) and honey ants, e.g., *Melophorus inflatus* in Australia. Honey-producing species whose colonies die out at the end of the active season, which are most social wasps and bumble bees (*Bombus* spp.), store comparatively little honey, and it is not economically important.

Production of honey by the honey bee, *A. mellifera,* has been studied most. Foraging workers collect nectar from plants and, when they return to their colony, "house bees" (young workers) take it from them and deposit it in cells of the comb. Bees evaporate water from it by manipulations that increase its surface area, while other bees fan to maintain a current of warm air through the hive. During this process (and even during its transport to the hive), secretions from the bees' hypopharyngeal glands are added to it. These contain the enzyme invertase, which inverts sucrose into fructose and glucose. At hive temperatures the solubility of glucose in a solution of fructose is unusually high, and the final honey has a very high sugar content, around or even above 80%. The relative amounts of the two sugars depend on the nectar sources. Fructose is more soluble in water than glucose, and high-fructose honeys (e.g., from *Robinia pseudoacacia*) rarely if ever granulate (crystallize), whereas high-glucose honeys (e.g., from dandelion, *Taraxacum officinale*) do so very quickly. Gentle warming of granulated honey redissolves the crystals.

Plant Sources of Honey

Most nectar is produced by flowers, although a few plants have extrafloral nectaries, including cotton (*Gossypium barbadense* and *G. hirsutum*) and rubber (*Hevea brasiliensis*). A number of world nectar plants have been classified according to the weight of honey that may be produced from a hectare of the plant in bloom, and the following are among those reported to be in the highest class (over 500 kg honey/ha or pounds/acre): *Epilobium angustifolium* (fireweed, rosebay willowherb), *Melilotus alba* (white melilot), *Phacelia tanacetifolia* (phacelia), *R. pseudoacacia* (false acacia, black locust), *Thymus vulgaris* (common thyme), and *Trifolium pratense* (red clover).

Where honeydew is available, bees collect it as well as nectar; it is sap from the host plants of certain plant-sucking insects in the order Hemiptera (Stenorhynchota) that excrete

part of the sap they ingest. Honeydew honey, which contains certain sugars not in floral honey, lacks any floral fragrance. It is much favored in some regions where it is produced—such as the Black Forest in Germany—but elsewhere the more delicate flavor of honey from nectar is much preferred.

COMPOSITION AND PROPERTIES OF HONEY

The composition of honey varies according to its plant origin and the weather conditions when the honey was produced. An analysis of 490 U.S. honeys gave an average of 69.5% fructose + glucose, 8.8% other sugars, 17.2% water, and small amounts of free acids, lactones, ash, and nitrogen.

The color of fresh honey varies with plant source, and honeys become darker during storage, especially at high temperatures. The color of liquid honey is important in marketing, and a number of countries have established systems for color grading.

Crystallization of honey (often referred to as granulation) is of great importance. It is a reversible process that changes liquid (run) honey into solid (set) honey, and it consists of the spontaneous crystallization of glucose (dextrose) monohydrate from a supersaturated solution. In Europe, liquid honey has been preferred, and honey from *R. pseudoacacia*—which rarely granulates—is favored; on the other hand, in Canada granulated honey is the norm, possibly because a common source is alfalfa *(Medicago sativa),* and its honey granulates rapidly.

The aroma and flavor of a sample of honey depend on its plant source, and beekeepers learn to recognize the plant origins of the honeys their bees produce. Honey processed for sale on the mass market is usually blended to maintain a constant product.

HONEY PROCESSING

In a beekeeper's hives the bees store honey in the combs of an upper honey box that is removed when it is full. Bees may be cleared from combs in the honey box by various methods: brushing and shaking bees off combs, using a bee-escape board through which bees can leave the honey box but not return, using a bee repellent, or blowing the bees out of the boxes with a stream of air.

Most honey is separated from the wax comb and processed for sale in containers, without any comb or wax. Processing the honey is likely to consist of the following stages: (1) clearing bees from the combs to be harvested, which are then taken to the honey house; (2) warming the combs to 32 to 35°C; (3) uncapping the combs and dealing with the cappings; (4) extracting the honey from the combs in a centrifuge; (5) clarifying the honey by passing it through a strainer and/or baffle tank; (6) flash heating and pressure filtering (in large processing plants in some countries); (7) if desired, initiating controlled granulation, on a large or small scale.

Honey is hygroscopic, and it should not be exposed to air with a relative humidity above 60%, or it may absorb water. (Some operators reduce the water content of honey slightly during stage 2 of processing or between stages 5 and 7.)

The processing of honey for sale either liquid or granulated is obviated if honey combs themselves are sold. Traditionally, "sections" were miniature wooden frames fitted with a very thin wax comb foundation on which the bees built cells, filled them with honey and—the beekeeper hoped—completely sealed them; the weight of honey in each section sold was usually 0.5 kg or 1 lb. However, perfectly sealed sections are difficult to produce, and in the 1900s several easier ways were devised to prepare honey in the comb for sale.

One alternative is cut-comb honey. To produce it the beekeeper inserts large frames fitted with extra-thin unwired wax foundation in the hive, harvesting them when full of honey (unlike sections, they need not be entirely capped). Each frame is placed on a flat surface and the comb cut out of its frame with a heated knife. Fully capped areas of it are cut into portions for sale, and honey is allowed to drain from the cut edges. Each piece is packaged in a heat-sealed box or a sheet of transparent plastic.

An easier alternative is to sell a jar containing a piece of honey comb and filled up with liquid honey; this is referred to as chunk honey.

HONEY AS A PRODUCT

Present World Production and Consumption

According to figures available, 1.1 million tonnes were produced in 1999. Honey yields per hive are usually highest in countries with an extensive belt between latitudes 23 and 30° (N or S), including China, Argentina, Mexico, and Australia.

Honey as Food

Honey from bees' nests was probably eaten by some mammals, including bears, before humans did so, and chimpanzees have been observed using tools to get access to honey in bees' nests in a tree. In Africa, India, and Spain, rock art from Mesolithic times and later shows human honey hunters harvesting from nests in trees or rocks.

Within the historical period, the use of honey is recorded from around 3000 B.C. onward. In India it was used by the famous surgeon Susruta around 1400 B.C. and much praised in the Vedas, sacred Hindu books collected together about 1500 B.C. In Rome, Columella judged honeys by their plant source, that from thyme being the best. Honey from a few plants, including *Rhododendron,* is toxic; in 399 B.C. when Xenophon's army retreated from Persia across Pontus in Asia Minor, the soldiers ate honey near the Black Sea coast that probably came from *R. ponticum.* It made them very ill, but they recovered by the third day. Records of baking with honey survive from 1200 B.C. onward in ancient Egypt.

Honey in Medicine

Honey has been regarded as a health-giving substance since ancient times, and Pythagoras (ca. 530 B.C.) was said to have attributed his long life to his constant use of it. Honey is a common ingredient of cough mixtures and lozenges and is often recommended as a symptomatic treatment for dyspepsia and peptic ulcers; the organism *Helibacter pylori,* which is a common cause of peptic ulceration, is inhibited by honey. Some sufferers from hay fever may be helped by eating honey that contains pollen.

A beneficial effect on wound healing has been known since early times, and a mechanism for this was established in the 1960s. The hypopharyngeal glands of *A. mellifera* workers secrete the enzyme glucose oxidase; this enters the honey, and in the presence of water a small amount of hydrogen peroxide is produced, which is bactericidal. Also, honey is hygroscopic, so it extracts exudates from infected lesions. For these reasons, honey is currently used in a number of hospitals, especially on wounds that are difficult to dress.

Honey in Alcoholic Drinks

From ancient times onward, drinks have been made by fermenting fruit to make wine, or cereals to make ale or beer. In many regions where bees were kept in hives, an important use of honey was its fermentation to produce an alcoholic drink, often referred to as mead. Where vines were grown, wine had a higher social status than honey-based drinks and tended to displace them, but honey-based drinks remained important north of the warmer vine-growing areas. In tropical Africa "honey beer" was made by fermenting honey for a short period.

See Also the Following Articles

Beekeeping • Commercial Products from Insects • Medicine, Insects in

Further Reading

Crane, E. (1975). "Honey: A Comprehensive Survey." Heinemann, London.
Crane, E. (1999). "The World History of Beekeeping and Honey Hunting." Duckworth, London.
Crane, E., Walker, P., and Day, R. (1984). "Directory of Important World Honey Sources." International Bee Research Association, London.
Graham, J. G. (ed.) (1992). "The Hive and the Honey Bee," revised edition, Chap. 21. Dadant, Hamilton, IL.
Killion, E. E. (1981). "Honey in the Comb," revised edition. Dadant, Hamilton, IL.
Riches, H. R. C. (2000). "Medical Aspects of Beekeeping." HR Books, Northwood, U.K.

Honey Bee

see *Apis Species*

Honeydew

Thomas E. Mittler
University of California, Berkeley
Angela E. Douglas
University of York, United Kingdom

Honeydew is a sugar-rich liquid released from the anus of some phloem sap-feeding insects of the order Homoptera. It consists principally of the residue of ingested phloem sap after digestion and assimilation in the insect gut, but it also contains waste products of insect metabolism eliminated via the gut. Honeydew deposited onto plant or other surfaces is an important source of energy-rich food for other animals, including some flies, parasitoids, ants, and microorganisms, and is used as a foraging cue by insect predators and parasitoids of some homopterans. Ants also collect the honeydew directly as it is being released from the producing insect, a behavior known as ant attendance.

THE NATURE OF HONEYDEW

Honeydew has historically been a source of wonderment. For example, the honeydew produced by the coccid *Trabutina mannipara* on tamarisk trees may have been the "manna from heaven" on which the Israelites fed during their escape from Egypt; honeydew has also been described as "the milk of Paradise" by Samuel Taylor Coleridge. However, the biological nature of honeydew is more mundane. The phloem sap of plants contains very high concentrations of sugars, usually the disaccharide sucrose or oligosaccharides of the raffinose family. Phloem-feeding insects ingest very large amounts of sugars relative to other essential nutrients, and up to 90% of the ingested sugar may be egested via the anus, and this sugar-rich material is honeydew.

Honeydew is produced by phloem-feeding insects, not by xylem feeders. Most phloem-feeding insects are members of the homopteran suborder Sternorrhyncha, which includes aphids, whitefly, mealybugs, and psyllids, or the homopteran suborder Auchenorrhyncha, which includes planthoppers and leafhoppers. Honeydew production is not dependent on either gut anastamoses (filter chambers) or Malpighian tubules absent in many and all aphids, respectively. It is released exclusively from the anus. The cornicles of aphids (capable of discharging defensive secretions and pheromones) are not involved.

PRODUCTION

The amount of honeydew produced by insects can be substantial. For example, first instars of the willow aphid *Tuberolachnus salignus* release more honeydew than their own

body weight on an hourly basis. However, the rate of honeydew production by other aphids is generally considerably lower.

The fluid that comprises honeydew accumulates in the rectum and is then ejected as a single droplet via the anus at fairly regular intervals, usually once every 15 to 40 min. Both the volume of each droplet and the frequency of production are influenced by many factors, including aphid age, size, species, and host plant.

Honeydew can be hazardous for the insect producing it because droplets may smother the insect and promote microbial growth on or near the insect. Various mechanisms reduce these risks in insects both on exposed plant surfaces and in confined spaces, including galls. On exposed surfaces, aphids generally project the honeydew droplet from the anus to a distance of up to several centimeters by either a kicking action of one of the hind legs or by using the cauda (a small appendage dorsal to the anus) to catapult the droplet ventrally. The legs and cauda bear hydrophobic cuticular hairs that prevent the sticky honeydew from adhering to the insect surface. Honeydew production by some species has been reported to be interrupted in windy conditions, presumably to avoid being smothered by the honeydew.

Insects in galls or other confined spaces (e.g., subterranean forms) are not wetted by their honeydew because they are coated with hydrophobic wax secreted from cuticular glands. In addition, each honeydew droplet produced by these insects tends not to be projected away from the insect body but remains poised at the anus until it is coated by wax. In some insects, notably certain nymphal psyllids, the honeydew and waxes coalesce to form a gelatinous or crystalline substance (called lerp), which acts as a solid protective covering for the growing insects.

Honeydew production may serve purposes additional to the elimination of waste compounds, with implications for the pattern of honeydew production. This is illustrated by two phenomena: "honeydew-panting" and ant-tending.

The former behavior is displayed by certain aphids, notably *T. salignus,* at elevated temperatures. The aphids raise their abdomen almost at right angles to the plant surface with their mid- and hind legs extended, and small honeydew droplets are alternately protruded and retracted from the anus. This behavior may cool the aphids as a result of evaporational water loss from the droplets.

The ejection of honeydew from some insects is modified by the attendance of ants or other insects. When solicited by an ant, the insect releases the honeydew droplet slowly and holds it at the anus while the ant imbibes, and, if the droplet is not removed by an ant, the insect may repeatedly extrude and retract the droplet, as if advertising the availability of honeydew. However, certain obligately ant-tended species apparently excrete honeydew droplets only in response to solicitations by their attending ants. The cauda and rectal musculature of some ant-tended aphids is much reduced, presumably reflecting their dependence on ants to remove their honeydew.

COMPOSITION

From a physiological perspective, honeydew is dominated by first, egesta, the components of ingested phloem sap that have not been assimilated by the insect (some phloem sap compounds may be enzymatically altered by gut enzymes prior to egestion), and second, excreta, waste products of insect metabolism that are eliminated via the gut after transfer from the body tissues to the gut lumen.

In quantitative terms, honeydew is dominated by sugars. In the best studied group, the aphids, the principal sugars in honeydew are typically different from the sugar ingested by the insect and usually of higher molecular weight. This reflects the vital osmoregulatory function of the gut in phloem-feeding insects. The osmotic pressure of phloem sap is generally considerably higher than the osmotic pressure of the insect body fluids, creating a tendency for the insect to lose tissue water to the gut.

Amino acids are the principal nitrogenous compound in phloem sap. Not all of the ingested amino acids are absorbed across the insect gut and assimilated (estimates of assimilation efficiency vary from 60 to 99%), and amino acids are routinely recovered from insect honeydew.

Honeydew also contains nitrogenous excretory compounds but generally at very low concentrations. This is because the symbiotic microorganisms in homopterans act as an internal sink for waste nitrogen compounds. For example, uric acid, the principal nitrogen waste compound of the planthopper *Niloparvata lugens,* is not voided in the honeydew, but retained within the insect body and metabolized by the insect's symbiotic yeasts. Similarly, ammonia, the dominant waste nitrogen compound of aphids, is in low concentration in their honeydew because their symbiotic bacteria *Buchnera* consume much of the ammonia synthesized by these insects.

Honeydew may contain microorganisms and viruses derived either from the ingested phloem sap (and passed directly through the gut) or from the resident insect microbiota. For example, aphids feeding on plants infected with the luteovirus barley yellow dwarf virus will ingest viral particles from the phloem sap; those particles that are not transported into the insect hemocoel are expelled in the aphid honeydew. Plant viruses multiplying in the insects may also pass into the gut and occur in honeydew.

HONEYDEW AS FOOD

Many insects in several orders, including Diptera, Hymenoptera, Lepidoptera, Coleoptera, and Neuroptera, feed on honeydew that has fallen onto plant or other surfaces. Among these insects are herbivores (e.g., tephritid flies, butterflies, and moths) and many entomophagous taxa, such as chrysopids, coccinellids, syrphids, tachinid flies, and hymenopteran parasitoids. A number of nectivorous birds in Mexico and Australia forage on honeydew and lerp; lerp is also

consumed by flying foxes in Australia. Other small mammals and reptiles also feed on honeydew, and dipterous vectors of human diseases (e.g., mosquitoes and phlebotomine sand flies) may rely on honeydew for an energy source.

Because it is usually freely accessible on leaf surfaces, it can readily be imbibed by insects that lack the specialized mouthparts needed to exploit floral nectar. However, several features of honeydew reduce its availability and suitability as food. First, there is the tendency of honeydew sugars to crystallize. Second, the performance of various predators and parasitoids is generally lower on honeydew than on nectar. Third, plant-derived secondary compounds in certain honeydews are toxic to other insects. It has been suggested that insects have selection pressure to produce honeydew of little nutritional value; hence, potential competitors and natural enemies may be one factor shaping honeydew composition.

When animals consume honeydew as it is voided they are described as "tending." By far the most widespread group of tenders are the ants, including most species of the subfamilies Myrmicinae, Dolichoderinae, and Formicinae. Other insects reported to tend homopterans include polybiine wasps (e.g., *Brachygastra* and *Parachartergus* spp. associated with membracids and planthoppers, respectively) and silvanid beetles (e.g., *Coccidotrophus* spp. with the mealybug *Pseudococcus breviceps*). It is believed widely that only insects tend honeydew-producing homopterans, but it has been demonstrated recently that several Madagascan gekkoes stimulate planthoppers of the family Flatidae to release honeydew droplets on which they feed.

In the interactions involving ants, both the ants and their tended homopterans generally benefit from the association, which is therefore described as mutualistic. Access to honeydew has been shown to enhance the rate of increase of ant colonies, but the magnitude of the nutritional benefit varies widely with ant species and environmental circumstance. Predominantly predaceous species feed on honeydew only very occasionally; some ants switch between tending and preying on homopterans, depending on the nutritional quality of the honeydew (as shaped by plant physiology) and the nutritional needs of the ant colony, and honeydew accounts for more than half of the diet of many temperate wood ants of the genus *Formica* and is the dominant, even sole, food of certain subterranean ants, such as *Acropyga* spp., and of *Solenopsis* (fire ants).

An indication that honeydew is an important food source for many tending ants is that the ants protect the tended homopterans from predators such as lacewings, syrphids, and coccinellids. In addition, certain ant species enhance their supply and quality of honeydew by transporting their tended homopterans to suitable parts of the host plant where the phloem nitrogen content is high and concentration of toxic plant chemicals is low (e.g., *Lasius* and aphids of genus *Stomaphis*), and some members of the genus *Acropyga* that tend coccids bear live coccids in their mandibles during the nuptial flight. The homopteran partner benefits from the protection from natural enemies and ant-mediated removal of honeydew, as frequently indicated by elevated rates of population increase in field conditions.

Honeydew has been used as a source of food by people. Encrustations of honeydew produced by scale insects have been eaten since biblical times in the Middle East. Certain groups of Australian Aborigines and American Indians also used lerp from psyllids and honeydew from scale insects as a source of sugar. In Central Europe, large amounts of honeydew are consumed indirectly, because the honeydew of aphids on conifers is the principal, and sometimes sole, source of food for some honey bees. The honey produced from this source, often referred to as Wald Honig (forest honey), is considered of inferior in quality to floral honey but is, nevertheless, consumed extensively.

Honeydew may also be an important source of carbon and nitrogen for microorganisms. For 2 decades, this topic has been influenced by an as yet experimentally unsupported hypothesis that the use of nutrients in insect honeydew by soil microorganisms would mobilize soil nutrients and enhance nitrogen fixation and thus promote plant nutrition.

In agricultural contexts, the growth of molds on deposited honeydew can depress plant photosynthesis and crop yield and contaminate fruits, vegetables, and flowers, making them unmarketable. For example, sooty molds arising from untreated infestations of greenhouse whitefly can halve the yield of glasshouse tomato crops, and cotton growers in the United States have suffered financially as a result of the "gray cotton" caused by sooty mold growing on cotton lint contaminated with whitefly and aphid honeydew. Other detrimental effects of honeydew are sticky sidewalks, glazed windshields, and gummed up harvesting machines.

HONEYDEW AS A KAIROMONE

The smell or taste of honeydew on the plant surface is used as cues by various predators and parasitoids of homopterans to locate their hemipteran prey or, for reproductive females, as a stimulus for oviposition. This has led to the use of "artificial honeydew" sugar sprays onto crops to increase the numbers and effectiveness of natural enemies.

See Also the Following Articles

Aphids • Auchenorrhyncha • Food, Insects as • Sternorryncha

Further Reading

Budenberg, W. J. (1990). Honeydew as a contact kairomone for aphid parasitoids. *Entomol. Exp. Appl.* **55,** 139–148.

Douglas, A. E. (1998). Nutritional interactions in insect–microbial symbioses: Aphids and their symbiotic bacteria *Buchnera. Annu. Rev. Entomol.* **43,** 17–37.

Kiss, A. (1981). Melizitose, aphids and ants. *Oikos* **37,** 382.

Mittler, T. E. (1958). The excretion of honeydew by *Tuberolachnus salignus* (Gmelin). *Proc. R. Entomol. Soc. (London) A* **33,** 49–55.

Mittler, T. E. (1958). Studies on the feeding and nutrition of *Tuberolachnus salignus* (Gmelin). II. The nitrogen and sugar composition of ingested phloem sap and excreted honeydew. *J. Exp. Biol.* **35,** 74–84.

Molyneux, R. J., Campbell, B. C., and Dreyer, D. L. (1990). Honeydew analysis for detecting phloem transport of plant natural products. *J. Chem. Ecol.* **16,** 1899–1909.

Owen, D. F., and Wiegert, R. G. (1976). Do consumers maximize plant fitness? *Oikos* **27,** 488–492.

Stadler, B., Michalzik, B., and Muller, T. (1998). Linking aphid ecology with nutrient fluxes in a coniferous forest. *Ecology* **79,** 1514–1525.

Wackers, F. L. (2000). Do oligosaccharides reduce the suitability of honeydew for predators and parasitoids? A further facet to the function of insect-synthesized honeydew sugars. *Oikos* **90,** 197–201.

Way, M. J. (1963). Mutualism between ants and honeydew-producing Homoptera. *Annu. Rev. Entomol.* **8,** 307–344.

Hornet

see *Wasps*

Host Seeking, by Parasitoids

Ronald M. Weseloh
Connecticut Agricultural Experiment Station

Parasitoids are holometabolous insects that are free living as adults; their larvae are parasites within the bodies of other insects, which they invariably kill as they develop. Most parasitoids are small-to-large wasplike insects in the hymenopteran superfamilies Ichneumonoidea, Chacidoidea, Serphoidea, and Cynipoidea, or are flies in the dipteran family Tachinidae. Adult females, which are well suited for this task, almost always carry out host seeking in parasitoids. Most have wings and are active fliers, making it possible for them to explore large areas, relative to their body size. Most also have well-developed legs that facilitate the exploration of complicated surfaces. Typically, they possess tactile and chemosensory receptors on the antennae, feet, mouthparts, or ovipositor, and they have good visual acuity. Their ability to find specific host species may be important because many larval parasitoids exist inside other insects that often are capable of mounting immune responses unless the parasitoid is well adapted. Since, in addition, many parasitoids are small, it is easy for them to occupy restricted niches. Thus, the behaviors involved in host seeking in parasitoids are diverse and well developed. In fact, host seeking can be conveniently broken down into the overlapping, hierarchical categories of host habitat finding, host finding, and host acceptance.

HOST HABITAT FINDING

A female parasitoid may find herself far from potential hosts. This could occur if the host stage she emerges from is different from the one attacked. Also, many parasitoid females have a preoviposition period before eggs are ready to be laid. During this interval of a few days to several weeks, the parasitoid may leave the vicinity of the host to mate and obtain nourishment. For example, the ichneumonid wasp *Pimpla ruficollis* is a parasitoid of the European pine shoot moth, *Rhyaciona buoliana.* Yet for the first few weeks of her adult life, she is repelled by the odor of pine, and thus avoids the forest where the host is located. As a consequence, the initial stage in host seeking in many parasitoids is to search for locations where the host is likely to occur. Parasitoids often respond to general stimuli such as light, humidity, or vegetation form, leading them to meadows, forests, swamps, ponds, soil, or different vegetation strata. These behaviors considerably narrow the areas that must be actually searched for hosts. Many parasitoids are also attracted to volatile chemicals from plants. For example, after initially being repelled by pine, *P. ruficollis* females that are ready to oviposit are attracted by pine odors, and thus are drawn back to the forest. The ichneumonid *Itoplectis conquisitor* is attracted to the odor of Scots pine but not red pine, and does not attack lepidopterous hosts on the latter. In olfactometer tests, the aphid parasitoid *Diaeretiella rapae* is attracted to collard leaves. Wind tunnel experiments have shown that when such a parasitoid perceives a plant volatile, she reacts by walking or flying upwind (aenemotaxis), thus often leading to the plants where her preferred host feeds. In fact, in some cases the plant attraction is so important that the parasitoid host range encompasses the often diverse herbivores that feed on that plant rather than hosts that are taxonomically closely related. Plants are not the only habitat characteristics that can produce such attractive volatiles. Parasitoids of carrion-feeding flies are attracted to fresh or decaying meat, and parasitoids of *Drosophila* fruit flies respond to odors from yeast in decaying fruits where their hosts are likely to be present.

HOST FINDING

Unless random search is important, parasitoids usually find their hosts as a result of cues derived directly or indirectly from the host itself, often after they have entered the host habitat. Often host cues are perceived at close range. For example, the braconid *Cardiochiles nigriceps,* increases its searching when it contacts secretions produced by the mandibular glands of its host, *Heliothis* spp. caterpillars, as the latter feed on plants. The braconid *Microplitis croceipes* searches areas contaminated by chemicals contained in the feces of *Heliothis* caterpillars, whereas the braconid *Cotesia melanoscela* intensely searches leaf areas where host gypsy moth *(Lymantria dispar)* caterpillars have deposited silk strands. Once these materials have been perceived, stereotyped searching behaviors occur that typically consist of intense examination of the area with the antennae or tarsi. Parasitoids often also decrease their walking speed (orthokinesis) and/or increase turning rates (klinokinesis).

These behaviors, which serve to keep the parasitoid in the area having the host products, often lead to host discovery.

Other parasitoids are attracted from longer distances directly to hosts. Some parasitoid females respond to pheromones produced by their host. Parasitoids of the European elm bark beetle, *Scolytus multistriatus,* are attracted to "multilure," the aggregation pheromone of adult beetles. *Aphytis* spp. (Hymenoptera: Aphelinidae) are attracted to the sex pheromone produced by their host, California red scale *(Aonidiella aurantii).* A number of true bugs produce sex pheromones that are attractive to a variety of fly (Tachinidae) and to hymenopteran parasitoids. Some parasitoids are drawn from a distance to chemicals produced by plants in response to damage caused by host herbivores. For example, the braconid larval parasitoid *Cotesia marginiventris* responds to volatile terpenoids released from corn seedlings as a result of eating damage caused by host *Spodoptera* caterpillars. These chemicals may be components of the induced resistance that plants have developed against pathogens and herbivores. Even leaves not directly damaged by a herbivore may produce such materials. The induced plant chemicals may, along with materials directly produced by herbivores, serve to attract parasitoids. For example, parasitoids of bark beetles are attracted to a combination of plant chemicals produced by trees as a reaction to the mass attack of the beetles as well as to the aggregation pheromone produced by these beetles. Simultaneous responses of parasitoids to long-range host and plant cues illustrate that the division between the categories of habitat and host finding is often arbitrary.

HOST ACCEPTANCE

Although perhaps not technically part of host-seeking behavior, whether a host can be recognized as such after it has been contacted by a parasitoid is very important for the parasitoid. If a female cannot recognize a host as suitable for her progeny, habitat and host-finding activities would be wasted. Parasitoids have evolved behaviors that enable them to accurately choose suitable hosts. Many detect chemicals in the host cuticle or egg chorion that enable them to differentiate one potential host from another. These they usually detect with their antennae (Hymenoptera) or front tarsi (tachinid flies). Some parasitoids are also able to distinguish between hosts after insertion of the ovipositor by use of sense organs on the egg-laying organ itself. Acceptance of hosts via other sensory modalities, such as touch, sound, or sight, have also been documented. For instance, a *Trichogramma* wasp female examines a host egg with her antennae to determine its size. The ichneumonid parasitoid *Campoletis sonorensis* is influenced by host caterpillar shape. A cylindrical shape that approximated the shape of the *Heliothis virescens* host was more effective in stimulating oviposition than round or flat shapes. The egg–larval parasitoid *Chelonus texanus* accepts host lepidopteran eggs that have a rough or sculptured surface rather than a smooth one. Hairs from the body of gypsy moth larvae are enough to cause examination behavior in the parasitoid *C. melanoscela.* Movements perceived by sight or through vibrations of the substrate are important cues for a number of different parasitoids.

HOST DISCRIMINATION

A behavior related to host seeking involves the discrimination between hosts that have already been parasitized and those that have not. Some parasitoid females, after they have successfully parasitized a host, deposit a chemical marker on the surface that serves to tell other parasitoids of the same or other species that the host is already parasitized. Such marking pheromones have most often been found in parasitoids that attack sedentary hosts, such as eggs or pupae. Often the oldest parasitoid in a host is the one that survives; consequently, host marking saves other parasitoids from wasting time and eggs on a host in which their offspring are likely to perish. Also parasitoids whose ovipositor has been inserted into a host are often able to distinguish parasitized from unparasitized hosts. Whether the cues perceived are the result of marking materials specifically injected by the first parasitoid or of chemical changes in the host resulting from parasitoid development is generally not known.

VARIATIONS IN HOST-SEEKING BEHAVIOR

Although the host-seeking process in parasitoids may be very efficient, such that eggs are deposited only in host species suitable for their development, other strategies are used. Some parasitoids lay eggs in an area likely to be inhabited by their host. The larvae hatching from these eggs then must find their own way to the host. Members of the hymenopteran family Eucharidae are parasitic on ant larvae. Adult females lay eggs on or in plants. Each hatching larva is a planidium and so is free living and waits until it can attach to a passing adult ant, whereby it is taken into the nest and transfers to ant larvae. Many immature blister beetles (family Meloidae) are also parasitoids. Adult females lay eggs in the soil or sometimes on plants, and the emerging larvae are called triungulins. These active larvae find their own way to the eggs of locusts or nests of solitary bees, where they devour the eggs and/or provisions of the hosts. (Strictly speaking, meloids should probably be called "egg predators"; but their impact is much like that of true parasitoids.) Some tachinid flies lay large numbers of very small eggs on foliage that potential hosts may eat. In some cases, the females are attracted to damaged leaves, and this increases the chances of success for their larvae, but suitable hosts may never ingest many eggs. Nevertheless, these parasitoids may be as host specific as those that actively search for hosts.

FUNCTIONAL TERMINOLOGY FOR BEHAVIORAL CHEMICALS

A terminology has been developed for chemicals that function as signals between organisms, paralleling the activity of the

chemical cues involved in host-seeking behavior of parasitoids. All chemical attractants, arrestants, and so on that are important as modifiers of behavior between different organisms are grouped under the general term "semiochemicals." Pheromones are semiochemicals that serve to communicate between organisms of the same species; sex pheromones are an obvious example. Allelochemicals have effects between species and are further divided into those depending on whether the producing or receiving organism is helped or hurt by the signal. If the species producing the material is helped and the receiving one is harmed, the chemical is called an allomone. Examples of allomones include repellents that a stinkbug may produce to ward off predators such as ants or birds. A substance that harms the producing species but helps the receiving one is called a kairomone. The chemicals produced by insect hosts that serve as cues to parasitoids are kairomones because the parasitoid exploits them to the host's detriment. Usage is important here. A sex pheromone may attract a male to a female moth; but if a parasitoid cues on this chemical, the substance also functions as a kairomone. There are some materials that benefit both sender and receiver species. These are called synomones, and in the present context, the plant volatiles that attract parasitoids to host plants are synomones because they make it easier for the parasitoid to find herbivores damaging the plant.

IDENTITY OF HOST-SEEKING CHEMICALS

Some progress has been made in identifying chemicals important in host-seeking behavior of parasitoids. Volatiles involved in host habitat finding include ethanol and especially acetaldehyde produced in rotting peaches that attract the braconid, *Biosteres longicaudatus,* a parasitoid of tephritid fruit flies. Allyl isothiocyanate produced by crucifers is attractive to the braconid parasitoid of aphids *D. rapae.* The straight-chain hydrocarbons docosane, tricosane, tetracosane, and pentacosane from the scales of adult *Helicoverpa zea* moths are cues used by *Trichogramma evanescens* to locate host eggs. In the frass of *H. zea,* 13-methylhentriacontane is an examination-stimulating cue for the braconid parasitoid *M. croceipes,* as is heptanoic acid in the frass of the potato tuberworm *(Phthorimaea operculella)* for the braconid parasitoid, *Orgilus lepidus.* Members of a series of methyl branched hen-, do-, and tritriacontaines from the mandibular glands of *Heliothis virescens* serve to intensify searching of the braconid, *C. nigriceps* on areas of leaves damaged by host feeding. Long-chain hydrocarbons (heptacosane, nonacosane, and several dimethyl compounds) in the cuticle of gypsy moth pupae are important in the host acceptance behavior of the chalcid, *Brachymeria intermedia.* The ichnuemonid parasitoid *I. conquisitor* oviposits into a wax-covered cylinder of water mixed with the amino acids arginine, isoleucine, methionine, lysine, leucine, and serine, as well as magnesium chloride. The tachinid fly *Cyzenis albincans* lays very small, microtype eggs on oak foliage that caterpillars of the winter moth (*Operophtera brumata*) eat. The parasitoid is stimulated to lay eggs in the presence of sugars exuded by damaged oak leaves, thus increasing the chance that its host will be nearby.

LEARNING AND HOST SEEKING

Insect behavior is sometimes perceived as a rigid, instinctive, inherited phenomenon not subject to change. However, there is ample evidence that many insects vary their behavior depending on circumstances, and that often learning is involved. This is also true for parasitoids. Many female parasitoids respond to host stimuli more strongly after they have parasitized a host. The heightened response, which may take the form of faster host finding and/or more intensive searching, may be considered to be a form of reward conditioning in that the female responds more avidly to host stimuli once she has been "rewarded" by being able to oviposit. Indeed, the response to host stimuli may wane if the parasitoid is prevented from oviposition. This occurs in the ichneumonid *Campolitis sonorensis* if it is not allowed to oviposit after contacting host frass or damaged plant material. Also, when the eucoilid parasitoid *Leptopilina heterotoma* is not able to oviposit, it becomes unresponsive to host cues (*Drosophila* larvae), but it searches more avidly if placed in a novel environment. Another ability that some parasitoids demonstrate is associative learning. In this type of learning, the parasitoid becomes able to associate a nonhost stimulus with the presence of hosts. For example, the ichneumonid parasitoid *I. conquisitor* can learn to distinguish between different shapes, sizes, and colors of artificial tubes holding host lepidopterous pupae, depending on which ones it has been allowed to oviposit in. Another ichneumonid, *Venturia canescens,* which attacks lepidopterous larvae in cereals, can learn to associate the presence of hosts with the odor of a nonhost chemical such as geraniol. Also, the braconid *Bracon mellitor* learned to associate with its host an antibiotic incorporated into the artificial diet of that host, the boll weevil. Thus, the sensory modalities of vision, olfaction, and contact chemoreception may be involved in the process of associative learning.

The advantage of such flexibility is probably greatest for parasitoids that are not strictly host specific. Suitably malleable behavior would help these parasitoids take advantage of changes in host and habitat composition. Learning in parasitoid searching behavior has recently generated much research interest, so it is likely that many more examples will be forthcoming.

TRITROPHIC INTERACTIONS

The involvement of plant volatile chemicals in the host-seeking behavior of parasitoids has an ecological and an evolutionary aspect. By facilitating the parasitization of herbivores feeding on a plant these synomones aid both the plant and the parasitoid. As such, plant and parasitoid would be expected to coevolve, resulting in some finely developed systems of signal and response. Examples of tritrophic

interactions in which such coevolution may have occurred include situations in which a parasitoid is attracted to volatile chemicals produced by a plant only when that plant has been damaged by herbivores. This makes the signal more meaningful to the parasitoid than a signal produced by all plants at all times. The plant also presumably benefits by not wasting resources to produce a signal that is not needed. However, another possible explanation is that substances produced as a result of injury are part of an induced resistance response of the plant to herbivory; thus the primary purpose of the material would be to decrease foliage palatability or otherwise directly harm the herbivore. A parasitoid might evolve to use these materials, but still have little or no impact on the evolution of the plant responses. Indeed, most work on coevolution in insects and plants has emphasized the plant–herbivore interactions, yet there is little solid information about tritrophic interactions. However, coevolution between parasitoid and plants is still a theoretical possibility, and researchers are beginning to study this interaction.

IMPLICATIONS FOR BIOLOGICAL CONTROL

In biological control, high searching capacity is considered to be a very desirable trait of natural enemies. Host-seeking behavior influences searching capacity greatly. The more readily a parasitoid can find a host, the better it will realize its full reproductive potential. Also, for a parasitoid to control a host at low densities, it is necessary to find that host under conditions of scarcity. By using specific cues, especially volatile chemicals produced directly or indirectly by hosts, many parasitoids are able to find these hosts very well. Research on host-seeking behavior has progressed far enough to permit the development of some general concepts that should aid biological control workers as they evaluate the effectiveness of parasitoids. For example, chemicals are very prominent as host habitat-finding and host-finding cues. Chemicals involved in host habitat finding are usually perceived from relatively long distances, and they orient parasitoids to travel upwind. Chemicals involved in host finding may be perceived from long distances, but often these are plant materials induced by feeding damage by herbivore hosts. Many chemical cues emanating directly from the host are short range or can be perceived only upon contact. Thus, plants strongly influence initial stages of the host-seeking process. There has been much concern about whether exotic natural enemies imported for biological control can have the detrimental effect of attacking nonpest species, especially endangered species. Because host-seeking behavior effectively determines the host ranges of many parasitoids, the general concepts developed from studies of host-seeking behavior should aid in efforts to delineate host ranges of parasitoids. For example, candidate parasitoids are now often screened for their ability to attack nontarget hosts. These screening tests usually occur in a laboratory, often using insect hosts that are removed from their usual plant hosts. Thus, only host-seeking behaviors associated with cues directly derived from hosts are assessed. This ignores the often strong winnowing effect that attractions to plants exert on potential host ranges, and the laboratory-derived host range may be substantially wider than the natural range. Also important is learning, because the degree of flexibility in host seeking could quantify the likelihood of host switching in parasitoids.

The above-mentioned considerations illustrate the most important implications that host-seeking behavior in parasitoids has on practical biological control, but there is another dimension. As already mentioned, *Trichogramma* egg parasitoids intensively search areas in which scales from female moth hosts have been deposited. The main attractive material in these scales is tricosane. When tricosane is artificially deposited on foliage containing eggs of the host moth, the resulting parasitism by *Trichogramma* is higher than in areas not having tricosane. Thus, the direct use of such kairomones could improve pest control by manipulating the behavior of natural enemies. Although such schemes have so far not been economically viable, similar manipulations may prove to be workable in situations not yet tested. Continued research may well lead to some useful control methods.

See Also the Following Articles

Coevolution • Hypermetamorphosis • Learning • Oviposition Behavior • Parasitoids

Further Reading

Nordlund, D. A., Jones, R. L., and Lewis., W. J. (eds.). (1981). "Semiochemicals: Their Role in Pest Control." Wiley, New York.

Vet, L. E. M., Lewis, W. J., and Cardé, R. T. (1995). Parasitoid foraging and learning. *In* "Chemical Ecology of Insects" (R. T. Cardé and W. J. Bell, eds.), Vol. 2. Chapman & Hall, New York.

Vinson, S. B. (1984). Parasitoid–host relationships. *In* "Chemical Ecology of Insects" (W. J. Bell and R. T. Cardé, eds.). Sinauer, Sunderland, MA.

Vinson, S. B. (1998). The general host selection behavior of parasitoid Hymenoptera and a comparison of initial strategies utilized by larvaphagous and oophagous species. *Biol. Control* **11,** 79–96.

Vinson, S. B. (1999). Parasitoid manipulation as a plant defense strategy. *Ann. Entomol. Soc. Am.* **92,** 812–828.

Host Seeking, for Plants

Elizabeth A. Bernays
University of Arizona

Plant-feeding insects may find their hosts by seeking appropriate habitats, by increases in activity that maximize the chances of encountering a plant, by completely random activity in combination with strong arrestant properties of the host, or by attraction to a plant from a distance by smell or vision or both. Often generalized plant odors are attractive, but commonly host-specific odors can be distinguished by

specialist insects, and recently it has become known that most such insects are highly sensitive to one or a few host odors that are particularly attractive. Host color and shape can be important in plants with characteristic visual features and in insects that are day flying, although visually mediated responses are usually relatively unspecific.

FINDING HOSTS INDIRECTLY

Habitat location appears to be the first step for a number of species, although it is difficult to prove in practice. For example, grass-feeding grasshoppers are attracted to open habitats where grasses are generally abundant, but there are likely to be other reasons for this behavior. The pierid butterfly *Euchloe belemia,* however, is attracted to patches of thorn plants where, typically, its small host plants grow most densely.

Some small insects that find their hosts within fairly short distances may simply rely on increases in activity and turning behavior when they detect the appropriate odor, so that they are more likely to encounter their hosts. Among chrysomelid beetles, some engage in random movements that show little change with level of host odor; crucifer flea beetles in the genus *Phyllotreta,* for example, move randomly within and between host patches. It has been shown mathematically that random activity is important in overall efficiency of search strategies because, under many circumstances and especially when host signals are weak and plant suitability variable, exploratory search enhances the likelihood that an individual will contact the better plants.

USING ODORS TO FIND HOSTS

There are many examples of insects being attracted to the odors of their host plants, both by flying and by walking or crawling. Generalists such as the moths *Trichoplusia ni* and *Heliothis virescens* and the desert locust *Schistocerca gregaria* fly or walk upwind in wind tunnels toward general green plant odors, and there are examples among all orders of specialist herbivores being attracted to chemicals arising specifically from their host plants (Table I).

Because of air turbulence, concentration gradients that an insect might follow do not generally exist, except within centimeters of the plant. Instead, there are pockets of odor-carrying air that are carried downwind in a rough plume, and an insect encounters and perceives an irregular series of these pockets. An insect usually responds in two stages. First, there is "arousal," preparing the insect to respond to some further stimulus. Then orientation occurs, either on the substrate or in the air. Usually the orientation response is a response to the wind, with the insect turning upwind, and this is termed an odor-induced upwind (positive) anemotaxis. Among flying moths, this has been demonstrated clearly in wind tunnels, and the same kind of response may be seen when males fly upwind toward the source of female pheromone.

TABLE I Example of Host Plant Volatiles Attracting Specific Phytophagous Insects

Insect	Chemical(s) or host odors
Cavariella aegopodii (carrot aphid)	Carvone (one of the host volatiles)
Brevicoryne brassicae (cabbage aphid)	Isothiocyanates (host volatiles)
Aphis gossypii (cotton aphid)	Host plant odor
Leptinotarsa decemlineata (Colorado potato beetle)	Host plant odor
Ceutorhynchus assimilis (Cabbage seedpod weevil)	Isothiocyanates (host volatiles)
Psila rosae (carrot fly)	Mixture of five host volatiles
Delia antiqua (onion maggot)	Disulfides (host volatiles)
Acrolepiopsis assectella (leek moth)	Thiosufinates (host volatiles)
Plutella xylostella (diamondback moth)	Host plant odor
Manduca sexta (tobacco hornworm)	Host plant odor
Heliothis subflexa (groundcherry moth)	Host plant odor

Once airborne, the insect needs to monitor its ground speed, so that it can increase its airspeed if the wind is strong. To do this it uses visual information (i.e., image movement across the eyes from front to back). If the wind is too strong and the insect is unable to keep the images flowing, it turns and flies downwind or lands. The use of visual images by a flying or swimming insect to maintain orientation to a current flow is called an optomotor reaction. It enables the insect to maintain an orientation at any angle to the wind, not just directly up- or downwind. If the insect is unable to see the pattern of objects on the ground, it cannot orient. As well as generally flying upwind in response to a particular odor, many moths and beetles follow zigzag flight paths. This behavior, which evidently is programmed in the insect central nervous system, has the possible function of increasing the chances of encountering a pocket of odor.

Walking insects also show odor-induced anemotaxis. This has been demonstrated in locust nymphs, certain beetles, and aphids, for example, where individuals walk upwind in response to host odors.

In a number of smaller insects such as phytophagous flies, the odor-induced anemotaxis is slightly different; this is well studied are the onion maggot, *Delia antiqua,* and the cabbage maggot, *D. radicum.* In these species, after perception of the host odor, an individual fly turns into the wind and makes short flights. After landing, and again detecting the odor, it reorients into the wind and takes off. This tactic is particularly effective for host finding in vegetation, where the path to the food plant may be rather devious and the odor plume very broken.

A different response to host odor after the initial arousal is to move toward or land on a relevant visual target. This odor-induced visual orientation is believed to occur, for example, in the cabbage seed weevil, which uses odor-conditioned anemotaxis from a distance and then odor-conditioned landing responses on yellow targets close to the source. A number of insect species may be readily trapped by

means of a yellow water trap combined with a host odor source, and it is probably generally true that landing responses induced by the host odor are responsible.

Insects living in soil use odors alone to find hosts. Since, the air moves little in soil, steep gradients of volatile chemicals can be achieved and maintained. Carbon dioxide is commonly used by such larvae, but for specialists, host-specific compounds are also used. Root-feeding larvae, such as that of the carrot fly, *Psila rosae,* and the corn rootworm, respond by moving directly up a concentration gradient. Larvae of the carrot fly respond to a mixture of five compounds found in carrot odor.

For insects that fly or walk, the distances from which olfactory cues elicit responses vary from less than a meter as in the Colorado potato beetle, *Leptinotarsa decemlineata,* to about 30 m in some bark beetles and 100 m in some flies such as the onion maggot. Those that crawl in soil respond from just a few centimeters.

USING VISION TO FIND HOSTS

Visual attraction can result from responding to the color or form of the host plant. Because these vary so greatly within a species, and because there is relatively little specificity of shape among plant species, visual responses often occur only in the presence of an appropriate olfactory signal.

In a few examples, visual responses to host features have been demonstrated without the presence of odors. Walking insects of several species are attracted to narrow vertical targets in a plain arena, but the precise significance of this attraction is unknown. Perhaps it is a response to potential vegetation or shelter. Several species of butterflies, however, have been shown to land preferentially on leaves of particular shapes, with further discrimination occurring only after landing. Shape may interact with color as in the apple maggot, *Rhagoletis pomonella.* Host odors play a role here, but when colored rectangles are offered, the only color to attract flies is yellow, perhaps representing vegetation. If colored spheres are presented, the red and black shapes attract flies, perhaps representing the host fruit.

With respect to color, both wavelength and intensity are important. *D. radicum* lands preferentially on leaves with a leaf reflectance pattern characteristic of its host, whereas the western flower thrips, *Frankliniella occidentalis,* land most on yellows and whites, and more at highest intensities of reflected light. Patterns can also matter. For example, females of *Heliconius* butterflies lay their eggs on *Passiflora* leaves but tend not to oviposit on leaves that already have eggs on them. This is known to be a visual response to the yellow eggs, because if the eggs are painted green to match the leaf, butterflies do not discriminate against them.

A response to color is often coupled with a chemical cue. *Pieris rapae* require the presence of glucosinolates to oviposit but still responds to these chemicals only if they are on blue, yellow, green, or white substrates. Females reject red or black substrates.

Visual cues are usually important only at close range, though occasionally they attract specific herbivores from 10 m or so. This is true for the apple maggot, which has a very clear signal in the bright red fruits of its host substrate.

LEARNING IN HOST SEEKING

Although the studies are few, it is clear that many insects take advantage of experience in their foraging activities and thus improve efficiency of host finding. For example, butterflies learn to land on leaf shapes that resemble their hosts' leaf shapes, making many fewer mistakes with experience, and they learn many visual cues, especially color, when these are coupled with nectar rewards. Grasshoppers have been shown to learn that certain colored backgrounds are associated with the presence of high-quality food, and the time taken to find the food inside colored boxes in laboratory training experiments with *Melanoplus sanguinipes* was reduced from about 40 min for naive individuals to less than 10 min after a single experience.

Less is known about olfactory learning, but the work so far suggests that it may be more important than visual learning. Grasshoppers in experiments have been trained with different food odors associated with high-protein, low-carbohydrate diets and low-protein, high-carbohydrate diets. They were then fed untreated diets of one or the other type of imbalance until they were relatively deprived of one or the other major nutrient. Then, given a choice, grasshoppers tended to select against the odor that had originally been paired with the unbalanced food. Thus, if they were overfed protein and underfed carbohydrate, they were more likely to avoid the odor that had originally been paired with high-protein food and instead be attracted to the odor that had originally been paired with high-carbohydrate food.

Food aversion learning has been demonstrated in grasshoppers and caterpillars, whereby individuals having a deleterious postingestive experience after eating a certain food thereafter reject it or eat little of it. However, the role of odor and the importance of the associated cues in behaviors prior to contact have not yet been investigated.

ECOLOGICAL INTERACTIONS

The abiotic environment and the presence of other organisms influence host-seeking behavior in nature. Among abiotic factors, temperature constraints and needs are probably the most important. For example, thermoregulating grasshoppers choose sites off the ground for cooling, and warm sunny substrates for basking. This can dictate the plants that are immediately available for feeding upon, so selection of thermoregulatory sites influences food selection. For example, the black lubber grasshopper, *Taeniopoda eques,* is highly polyphagous; when temperatures become very high in its desert environment in the middle of the day, however, it roosts as high off the ground as possible on mesquite or

acacia bushes, and thus, any feeding is on these plants. During the cooler mornings and evenings it feeds only on plants at ground level. Many temperate butterflies seek out sunny or warm patches, and thus plants in those patches. For example, the meadow brown butterfly, *Pararge aegeria,* oviposits on various grasses but the actual choice depends on the temperature of the leaves, which in turn is influenced by whether the leaves are in sun or shade.

Wind is important for most insects. Wind speed and constancy influence odor plumes used by orienting insects. The wind speed also limits flight, with larger, stronger flying species remaining airborne at higher speeds. Very small insects are often carried by wind, and depending on the terrain, are deposited preferentially in certain places, such as the lee side of trees and hedges.

The presence of certain nonhost plants and the relative abundance or clumpiness of the host plant can alter the detailed behaviors involved in host seeking. For example, butterflies ovipositing in a habitat where two or more host plant species occur commonly tend to choose the species they laid eggs on previously, so that they land more often on the common host. In other insects, the host being selected for oviposition is dependent on factors such as the need for additional resources. In one example, the celery fly, *Phylophylla heraclei,* requires trees near to the celery host because this is where mating occurs and the adult food of aphid honeydew is available.

Insects that show odor-induced anemotaxis to their host plants presented alone in a wind tunnel in the laboratory do not always show the same behavior in field situations. For example, the Colorado potato beetle is attracted, at least from short distances, to its preferred host, potato. However, if nonhosts are also present, the response may be reduced or absent, and the host odor is said to be masked. Such interactions reduce the distance over which some host odors can be detected by phytophagous insects, and the phenomenon may be one of the mechanisms involved reduction of pest numbers in certain crop mixtures.

In addition, some insects are influenced by olfactory or visual evidence of prior occupation of a plant, competitors of the same or different species, and of the presence of natural enemies.

PHYSIOLOGY OF THE HERBIVORE

Host-seeking behavior is restricted to times when the ovipositing or feeding insect is in a suitable physiological state. For example, insects about to molt do not feed and are generally not responsive to host odors; in adult females, a load of eggs ready for laying alters motivation so that searching for a host takes priority over other behaviors. Similarly, an insect that has been deprived of food seeks hosts more readily than one that is replete. In nymphs of the desert locust, for example, positive anemotactic responses to the odor of grass in a wind tunnel were not seen in well-fed individuals but were dramatic in nymphs that had been deprived of food for 4 h.

In the bean aphid, *Aphis fabae,* winged individuals that fly distances from one host to another are attracted, when they take off, to the short wavelengths of the blue sky. After flying certain distances, they are preferentially attracted to the longer wavelengths of yellow, so that they then tend to land on plants in the vicinity. A number of aphid species bias their landings toward the yellower greens that often are associated with plants in an appropriate physiological state rather than toward plants of a particular species.

See Also the Following Articles

Eyes and Vision • Learning • Migration • Orientation • Phytophagous Insects

Further Reading

Barton Browne, L. (1993). Physiologically induced changes in resource oriented behavior. *Annu. Rev. Entomol.* **38,** 1–25.

Bell, W. J. (1991). "Searching Behaviour: The Behavioural Ecology of Finding Resources." Chapman & Hall, London.

Bell, W. J., Kipp, L. R., and Collins, R. D. (1995). The role of chemo-orientation in search behavior. *In* "Chemical Ecology of Insects" (R. Cardé and W. J. Bell, eds.), Vol. 2, pp. 105–152. Chapman & Hall, New York.

Bernays, E. A., and Chapman, R. F. (1994). "Host-Plant Selection by Phytophagous Insects." Chapman & Hall, New York.

Jones, R. E. (1991). Host location and oviposition on plants. *In* "Reproductive Behaviour of Insects" (W. J. Bailey and J. Ridsill-Smith, eds), pp. 108–138. Chapman & Hall, London.

Morris, W. F., and Kareiva P. M. (1991). How insect herbivores find suitable host plants: The interplay between random and nonrandom movement. *In* "Insect–Plant Interactions" (E. A. Bernays, ed.), Vol. III, pp. 175–208. CRC Press, Boca Raton, FL.

Murlis, J., Elkinton, J. S., and Cardé, R. (1992). Odor plumes and how insects use them. *Annu. Rev. Entomol.* **37,** 505–532.

Papaj, D. R., and Prokopy, R. J. (1989). Ecological and evolutionary aspects of learning in phytophagous insects. *Annu. Rev. Entomol.* **34,** 315–350.

Prokopy, R. J. (1986). Visual and olfactory stimulus interaction in resource finding by insects. *In* "Mechanisms in Insect Olfaction" (T. L. Payne, M. C. Birch, and C. E. J. Kennedy, eds.), pp. 81–90. Clarendon Press, Oxford, U.K.

Schoonhoven, L. M., Jermy, T., and van Loon, J. J. A. (1998). "Insect–Plant Biology: From Physiology to Evolution." Chapman & Hall, London.

Visser, J. H. (1988). Host plant finding by insects: Orientation, sensory input and search patterns. *J. Insect Physiol.* **34,** 259–268.

House Fly
(*Musca domestica*)

Gregory A. Dahlem
Northern Kentucky University

The house fly, *Musca domestica* (Fig. 1), is one of the best known and most widely distributed insects known to humans. It is a classic example of a synanthropic animal, one that lives in association with humans and their domesticated

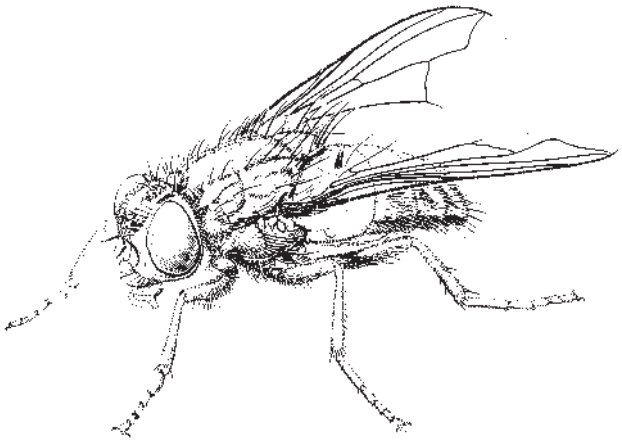

FIGURE 1 *M. domestica.* [After Huckett, H. C., and Vockeroth, J. R. (1987). Muscidae. *In* "Manual of Nearctic Diptera" (J. F. McAlpine, B. V. Peterson, G. E. Shewell, H. J. Teskey, J. R. Vockeroth, and D. M. Wood, eds.), Vol. 2. Biosystematic Research Institute Research Monograph **28.** For the Department of Agriculture and Agri-Food, Government of Canada. © Minister of Public Works and Government Services, Canada, 1987. Reproduced with the permission of the Minister of Public Works and Government Services, 2001.]

animals. House flies occur and thrive wherever humans are found but are very rare in natural or wild areas throughout the world. "Insects will survive long after humans disappear" is a common expression, but it is not true of the common house fly. House flies would likely not be able to survive in the absence of humans because their relationships are so closely linked.

The available literature on the house fly is vast. A computer search of only one database (BIOSIS) using *"Musca domestica"* as the key words, yielded well over 1000 references for the period between 1990 and 2000. A complete synopsis of the known information on this one species is not possible in this short article, but the following information should give the reader a better appreciation for this common species and the readings listed at the end will enable the curious to pursue this topic further. House flies are included in the group known as the calyptrate Diptera, which includes the Muscidae, Anthomyiidae, Calliphoridae, Sarcophagidae, Tachinidae, and several smaller families.

LIFE CYCLE AND BEHAVIOR

Mating Behavior

Courtship and copulatory behaviors are the most important and the most complex behaviors exhibited by the house fly. Visual, chemical, tactile, and auditory cues are all used, to various degrees, in courtship and copulation. The elimination of a male's production of, or a female's reception of, any one stimulus may not greatly affect mating success. However, if combinations of stimuli are simultaneously eliminated, mating can be significantly affected. In general, males mate as often as they can, whereas females mate just once.

The courtship behaviors of *M. domestica,* and many other calyptrate Diptera, are initiated when the male first sights a prospective femalelike object. Males are not very discriminatory in their initial choice of partners and strike other males, other species of flies, and small inanimate objects moving through their visual field.

The discovery of cuticular hydrocarbons that serve as sex pheromones in the house fly triggered a burst of investigations into the role of such pheromones in the mating behavior of the calyptrate Diptera. No evidence has been found that these pheromones are olfactory stimulants; rather, they appear exclusively to be contact excitants. In house flies, chemosensilla involved with contact chemoreception (or "taste") are located on both the mouthparts and the tarsi. Thus, house flies can "taste" with their feet. When a male touches a female with his tarsi, as he grasps her upon initial contact, he can use the female-produced sex pheromone to determine whether a potential mate is of the correct species, sex, and even mating status. This pheromone is a very important stimulus for the male, and a male repeatedly attempts to copulate with an object that is of appropriate size and "tastes right."

Behaviors that involve the touching or bodily movement of males and females beginning after the initial contact may be elicited by tactile cues. Tactile cues may be given by either sex during courtship, but it appears that the male's role is much more complex than that of the female. High-speed photography has shown that the house fly's courtship is extremely brief and complex. The highly ritualized sequence of movements that the male performs immediately after contact with the female seems to be very important to the female in her choice of potential mating partners. Males often strike females in midflight and perform the courtship ritual during their plunge to the ground. If the courtship is performed to the female's satisfaction, she allows the male to mate with her. If not, she can dislodge the male and stop mating from occuring by performing one of several different rejection maneuvers.

Mating pairs of *M. domestica* are normally quiescent during copulation. However, if the pair is disturbed they move, and the female flies, short distances carrying the male on her back.

Development

The number of eggs that mature in a fly's ovaries at one time is about 120. The female requires both sugar and protein meals for egg production. After copulation, egg laying takes place in 4 to 8 days. The female requires nearly a day to deposit the eggs, which may be deposited in a single mass or distributed in a number of locations. Each female is capable of developing several batches of eggs during her lifetime. Animal manure is the preferred ovipositional substrate, although a variety of decaying organic material can be used if fecal material is not available.

Developmental time is highly temperature dependent. Hatching usually takes place within one day after oviposition. Larval development occurs rapidly, with the larva (maggot)

normally passing through all three instars in 5 to 9 days. After full larval development, the third instar turns into a dark, cylindrical puparium, which is composed of the sclerotized skin of the last (third) larval stage. The process of pupation normally lasts about 5-days.

Most adults live for 2 to 3 weeks at normal summertime temperatures in the temperate regions. During the summer in temperate regions of the world, the entire life cycle (from egg to egg-laying adult) can be accomplished in 10 days to 2 weeks.

Flight and Dispersal

House flies are comparatively slow fliers, with a normal flight speed of about 2 m s^{-1} (or 7.2 km/h^{-1}). They have an innate tendency to disperse from their rearing site, even when conditions are favorable. Capture–release studies with marked house flies indicate that 85 to 95% of the flies stay within a 3-km radius of their release point after 4 days, although a few individuals may travel over as much as 20 km.

House flies are one of very few species of fly that purposely enter human structures, such as houses and barns. This propensity for entering dark openings has implications on the dispersal of this species, because house flies readily enter cargo or passenger areas of trucks, trains, ships, and airplanes. By this means, gene flow between geographically distant populations is easily, although accidentally, promoted.

Ability to Land and Walk on Ceilings and Vertical Surfaces

A common question about house flies is, "How do the flies land on ceilings, and how do they walk up smooth vertical surfaces, such as glass windows?" In landing on ceilings, a house fly normally performs a "half-roll" and reaches its legs out to the ceiling. Contact of the tarsi with the ceiling inhibits flight and the fly comes to rest, generally facing the direction that it was flying. A close look at the structures found on the tips of the tarsi help to explain the fly's ability to cling and walk on ceilings or smooth glass windows. The apical tarsal segment bears a pair of curved claws that are used to cling to rough surfaces. At the base of each claw is a padlike structure, called the pulvillus, which bears a large number of glandular setae. These setae are coated with secretions that make them sticky, allowing the fly to walk on vertical, or even inverted, smooth surfaces.

THE HOUSE FLY AS A VECTOR OF HUMAN AND ANIMAL DISEASE

House flies, particularly in large numbers, are a nuisance to humans when they enter houses, land and feed on human food, and spot windows with their feces. Of greater importance to humans, however, is their ability to spread human and veterinary disease agents. House flies have been associated with over 100 pathogens that can cause disease in humans and animals. Unlike the pathogens responsible for many other insect-borne diseases, the pathogens spread by the house fly do not usually multiply within the fly, nor do they require association with the fly for part of their life cycle. Instead, the usual association between house flies and pathogenic organisms is one of physical transmission of pathogens the flies pick up on their bodies at one feeding site (e.g., a garbage can or manure pile) and transfer to human and/or animal food when they land and feed. House flies have been associated with the transfer of a variety of viral and bacterial diseases, such as typhoid fever, cholera, dysentery, and infantile diarrhea, as well as a variety of parasitic worms.

See Also the Following Articles

Chemoreception • Medical Entomology • Urban Habitats

Further Reading

Colwell, A. E., and Shorey, H. H. (1977). Female-produced stimuli influencing courtship of male house flies *(Musca domestica). Ann. Entomol. Soc. Am.* **70,** 303–308.

Crosskey, R. W., and Lane, R. P. (1995). House flies, blowflies and other allies (calyptrate Diptera). *In* "Medical Insects and Arachnids" (R. P. Lane and R. W. Crosskey, eds.). Chapman & Hall, London.

Greenberg, B. (1971). "Flies and Disease," Vol. I of "Ecology, Classification and Biotic Associations." Princeton University Press. Princeton, NJ.

Greenberg, B. (1973). "Flies and Disease." Vol. II of "Biology and Disease Transmission." Princeton University Press. Princeton, NJ.

Huckett, H. C., and Vockeroth, J. R. (1987). Muscidae. *In* "Manual of Nearctic Diptera" (J. F. McAlpine, B.V. Peterson, G. E. Shewell, H. J. Teskey, J. F. Vockeroth, and D. M. Wood, eds.), Vol. 2, pp. 1115–1131. Biosystematic Research Institute Research Monograph **28,** Ottawa, Ontario, Canada.

Tobin, E. N., and Stoffolano, J. G., Jr. (1973). The courtship of *Musca* species found in North America. 1. The house fly, *Musca domestica. Ann. Entomol. Soc. Am.* **66,** 1249–1257.

West, L. S. (1951). "The Housefly. Its Natural History, Medical Importance, and Control." Comstock, Ithaca, NY.

Hymenoptera

(Ants, Bees, Wasps)

Donald L. J. Quicke

Imperial College, University of London

The Hymenoptera are a major order of holometabolous insects. That is, they undergo complete metamorphosis with distinct egg, larval, pupal, and adult stages. They are one of the five megadiverse insect orders along with Coleoptera, Diptera, Lepidoptera, and Heteroptera, and perhaps even the most species rich of any insect order—certainly this is true at temperate latitudes. They are generally cosmopolitan and, except for some specialized groups, they are most speciose in the tropics.

Although after working with Hymenoptera for a while it becomes easy to recognize members of this order, there are almost no conspicuous defining characters, because the majority of hymenopteran attributes are plesiomorphic; that is, they are shared with the common ancestors of various other orders. It is not surprising therefore that although almost everyone on earth, save perhaps those living at extreme northern latitudes, is familiar with ants, bees, and social wasps and has vernacular names for these particular taxa, there is not a single vernacular name in any language that refers to them in toto. Hymenoptera are also diverse in terms of their life histories: they include phytophagous, parasitoid, and predatory taxa, both solitary and highly social species, and they range in size from the rather large and intimidating spider-hunting pompilid wasps that can reach 12 cm wingspan down to the tiniest parasitic wasps that are approximately 0.1 mm in length (males of the wingless mymarid chalcidoid, *Dicopomorpha echmepterygis*). It is hard to overstate their ecological importance because they are collectively involved in so many types of interaction, and it is likely that many are effectively keystone species in their own habitats.

The Hymenoptera get their name from the Greek words *humen* and *pteron,* meaning membrane and wing, respectively, and this gives the first clue to identifying them. Excluding the numerous exceptions of apterous and brachypterous species that are widely distributed through the order, hymenopterans possess two pairs of membranous wings that are devoid of scales. The forewings are larger than the hind wings, and the two are interlocked during flight by a row of special hooks called hamules (or hamuli) that are on the anterior margin of the hind wing; these hamuli engage (or interlock) with a fold on the posterior edge of the forewings. This system makes the Hymenoptera functionally dipterous (two-winged) during flight, since the wing surfaces on either side of the body acts as a single aerofoil. Hamules are unique to this order of insects.

GENERAL BIOLOGY

Since Hymenoptera is a very large order, it is not surprising that a considerable number of biologies and life history strategies are exhibited by its various taxa. Broadly speaking, the basal lineages are phytophagous as larvae, feeding both ecto- and endophytically on a large range of herbs, shrubs, and trees; few tropical pergid sawflies, even feed on slime molds! The great majority of the remaining species are either parasitoids of other insects or predators of insects (e.g., the yellow-jackets or social wasps, which are members of the Vespidae) or spiders. However, among the higher taxa, there have also been several reversals to phytophagy, especially through the formation of galls on plants (cecidogenesis). The bees (Apidae) and one other, small tropical group, the Masarinae within the Vespidae, have evolved to make use of pollen and nectar as a larval food source.

Development

Like other holometabolous groups, hymenopterans primitively pass through generally five instars, though the number of instars is typically smaller in endoparasitic taxa, and in one such genus there seems to be just a single instar. The final instars of hymenopterans are rather morphologically conservative, with most sawflies having rather caterpillar-like larvae with well-developed true legs and variously developed prolegs on several of the abdominal segments. Adoption of an endophytic way of life by cephoid sawflies and wood wasps was accompanied by a reduction in the prolegs and more generally by reduction of sensory structures. Final instar apocritan wasp larvae are all quite similar and are termed hymenopteriform. They are superficially rather maggotlike in that they lack legs and other processes and often have a rather reduced head. However, many endoparasitoids have highly bizarre, first instars characteristic of their particular families, and for which a variety of specific terms have been coined.

The pupal stage of hymenopterans is exarate; that is, the antennae, legs, and wings are free from the body (in contrast to the Lepidoptera, e.g., in which these components are fused with the body). The pupae tend to be rather delicate and are easily damaged. All sawflies and most members of the Ichneumonoidea + Aculeata clade produce a silken cocoon to protect the pupa. Most of the other parasitic taxa do not, however, probably because they pupate within the host remains or, if they pupate externally, do so in a location where the pupa is likely to be protected by the surroundings, such as within a leaf mine, gall, or wood boring.

Key Features in Hymenoptera Evolution

Given the huge size of the order, it is interesting to consider what features have enabled hymenopterans to be so successful in terms of both individuals and total number of species. Most attention has focused on a small number of features such as selection of oviposition site, modification of that site, the use of venoms, and the evolution of the thin wasp waist, all of which are discussed in this article. In addition, the unusual form of sex determination mechanism, haplodiploidy, may have been particularly important in the evolution of sociality. It is likely, however that few of these traits have operated in isolation, and it is the interactions of these and other factors that have been important. Thus, for example, evolution of sociality may have been facilitated by the sex determination mechanism but also requires the abilities to remember where the nest is, to recognize nestmates, and to be able to defend the nest.

Several studies have emphasized that the success of the Hymenoptera has probably been a consequence of the general tendency of these insects to provide their offspring with particularly nutritious food sources, and when necessary (and that has been often) to modify poorer foods to better ones. Although this may be most familiar in terms of the provisioning of larvae in the nests by the social wasps and bees, such

TABLE I Most Recent Classification of the Sawflies and Wood Wasps

Superfamily	Family	Described extant species	Notes
Xyeloidea	Xyelidae	50	Most ancient family, with Holarctic distribution.
Pamphilioidea	Pamphiliidae	250	Sometimes called webspinning sawflies after the habit of early instars. Some are pest species.
	Megalodontesidae	40	Rare group with little known about biology; some feed on Apiaceae and on Rutaceae.
Tenthredinoidea	Argidae	800	A common group.
	Blasticotommidae	~10	Usually uncommon, larvae live in a ball of foam of ferns.
	Cimbicidae	130	Occasionally common, often rather large and beelike sawflies.
	Pergidae	500	Principally southern group, especially in Australia and South America.
	Diprionidae	90	Pine sawflies.
	Tenthredinidae	4000	Very common and speciose in temperate areas, uncommon but moderately diverse in the tropics. Most are exophytic with caterpillar-like larvae. Some are gall formers
Cephoidea	Cephidae	80	Stem sawflies, elongate, associated with grasses and rosaceous shrubs.
Anaxyeloidea	Anaxyelidae	1	A single species from western United States associated with fire-damaged trees.
Siricoidea	Siricidae	95	Horntail wood wasps.
Xiphydroidea	Xiphydriidae	100	Horntail wood wasps.
Orussoidea	Orussidae	75	Parasitic sawflies.

behaviors and physiological adaptations are to be seen all through the order and are manifested in many different ways.

First, there is egg placement and the larval food resource. The morphology of the ovipositor has been crucial in this respect. The hymenopteran ovipositor is used not only for laying eggs, it is also used to pass venom and/or other secretions to the place of oviposition. In the parasitoid taxa, these venoms either cause paralysis of the host or are important in overcoming the host's immune response against the parasitoid. The ovipositor is typically well supplied with sensilla, and the insects receive and interpret the resulting sensory information and use it in deciding whether they have located a site or host suitable for egg laying. This organ has been especially well studied in parasitoid taxa, and such observations have been used to test many evolutionary concepts.

In the majority of the aculeates (stinging wasps, bees, and ants) the egg-laying role has been lost, but the same structures are still present and are used for envenomation of prey or enemies. The venoms of most of these act on the nervous systems or nerve–muscle junctions of their prey insects, permanently paralyzing them. In this sense, the venoms are rendering their larval food sources manipulable and safe by preventing the prey insect from wriggling or moving to damage the wasp's developing young.

"Venoms" were important even before the evolution of parasitoidism. For example, at least some and possibly most wood wasps inject chemicals into their host trees along with their eggs and symbiotic fungi fragments, and these toxins probably either kill the living cambium cells or in some other way help the symbiotic fungi to overcome the trees' defenses so that the wood wasp larvae can feed on the developing nutritious fungal hyphae. As often happens, these conclusions are based on relatively few data, and observations of other species are very much needed.

Evolution of the thin wasp waist, which defines a large group of families called the Apocrita, was another absolutely key feature in that it greatly increased the mobility of the posterior abdomen relative to the thorax. This in turn allowed greater control of the ovipositor and greater variety in its use; later, it allowed the sting, which is in fact just a derived ovipositor, to be much more effective as a weapon of defense and offense. It is interesting that vertebrates can learn that a bee or wasp can deliver a sting and that part of the recognition of this ability involves the very conspicuous abdominal movements of the insect as it probes for a vulnerable spot with its sting. Because male Hymenoptera never possess stings (because the males do not have an ovipositor-derived apparatus!), they are harmless in this respect. Often, however, males very effectively mimic female wasp stinging movements such that people, and probably many experienced predators, do not take the risk and quickly release them—this behavior has been termed *aide-mémoire* mimicry.

The wasp waist, contrary to many people's initial expectations, is actually not located between the thorax and abdomen, but is a constriction between the first and second abdominal segments (Fig. 1). In the ants, posterior abdominal mobility is increased even more by second and sometimes third constrictions between the second and third, and third and fourth, abdominal segments, which give rise to the distinct node or nodes between the middle and posteromost body regions. There is a very good reason for the wasp waist to be positioned after the first abdominal segment. Higher hymenopterans are typically strong fliers, and their longitudinal flight muscles are consequently large. Because these muscles are attached internally on the anterior of thorax (actually the mesonotum) and posteriorly on a large internalized chitinous phragma that slants posteriorly, if there were a constriction immediately behind the last (third)

TABLE II Generally Accepted Classification of the Apocrita[a]

Superfamily	Family	Described extant species	Notes
Stephanoidea	Stephanidae	200	No common name. Idiobiont parasitoids of wood-boring beetle larvae.
Megalyroidea	Megalyridae	50	No common name. Usually rare, presumed idiobiont parasitoids of wood-boring beetle and aculeate wasp larvae.
Trigonaloidea	Trigonalidae	100	No common name. Complex hyperparasitic (rarely primary parasitic) life cycle.
Evanioidea	Evaniidae	500	Ensign wasps. Often common in the tropics, endoparasitic or egg-predatory in cockroach oothecae.
	Gasteruptiidae	420	Moderately common, cosmopolitan, kleptoparasites of solitary bees. Their larvae kill the host bee's egg and consume the latter's pollen food store.
	Aulacidae	200	Koinobiont endoparasitoid of wood wasps and of wood-boring Coleoptera larvae.
Ceraphronoidea	Ceraphronidae	350	Small to very small, very common wasps with a wide range of parasitic biologies.
	Megaspilidae	450	Small to very small, common wasps with a wide range of parasitic biologies.
Platygastroidea	Platygastridae	1,100	Extremely common, usually small, cosmopolitan, mainly koinobiont endoparasitoids of Diptera larvae, but other biologies and hosts known.
	Scelionidae	3,000	Extremely common, usually small, consmopolitan, idiobiont egg parasitoids of many groups of insects and spiders.
Proctotrupoidea	Austroniidae	3	Extremely rare, Australian, biology unknown.
	Diapriidae	2,300	Very common, idiobiont and koinobiont endoparasitoids, mainly of Diptera larvae/pupae.
	Heloridae	7	Usually uncommon, koinobiont endoparasitoids of Neuroptera (Chrysopidae) larvae.
	Maamingidae	2	Most recently described family, known only from New Zealand, biology unknown.
	Monomachiidae	20	Parasitoids of Diptera (Stratiomyidae) in Australia and South America.
	Pelecinidae	3	Moderately common, large, entirely New World, koinobiont endoparasitoids of subterranean Coleoptera larvae.
	Peradeniidae	2	Very uncommon, Australian, biology unknown.
	Proctotrupidae	310	Common, mainly northern, koinobiont endoparasitoids, mainly of Coleoptera larvae, mostly in soil or litter layer.
	Renyxidae	2	Holarctic, extremely rare, biology unknown.
	Roproniidae	18	Usually uncommon, Holarctic and Oriental, parasitoids of sawflies.
	Vanhorniidae	5	Generally uncommon, New World parasitoids of eucnemid beetle larvae.
Mymarommatoidea	Mymarommatidae	14	Very uncommon, minute, biology unknown but guessed to be egg parasitoids.
Chalcidoidea	~20 families	19,000	Chalcids. Approximately 20 families are recognized, most with diverse biologies.
Cynipoidea	Ibaliidae	50	Egg-larval, koinobiont endoparasitoids of wood wasps.
	Liopteridae	50	Endoparasitoids (probably koinobiont) of wood-boring Coleoptera larvae in the tropics.
	Cynipidae	1,000	True gall wasps and also inquilines in other cynipid galls.
	Figitidae	1,500	Very common endoparasitoids of Diptera, of Neuroptera, and of hymenopterous parasitoids of aphids.
	Austrocynipidae	1	Extremely rare, parasitic on Lepidoptera larvae in *Araucaria* cones in Australia.
Ichneumonoidea	Ichneumonidae	22,000	Very common, biologically diverse though not including any egg parasitoids.
	Braconidae	20,000	Very common, biologically diverse though not including any egg parasitoids.
Chrysidoidea	Bethylidae	2,000	Common and widespread, small ectoparasitoids of small beetle and moth larvae in semicryptic locations; some show parental care.
	Chrysididae	3,000	Common, cosmopolitan. Chrysidines are mainly larval parasitoids of solitary vespid wasps and bees; cleptines attack sawfly prepupae; others are idiobiont egg parasitoids of stick insects.
	Dryinidae	950	Common, koinobiont parasitoids of larger Auchenorrhyncha (e.g., Cicadelloidea), some developing partially externally.
	Embolemidae	16	Uncommon, cosmopolitan; one species parasitic on Heteroptera nymphs.
	Plumariidae	20	Very rare tropical wasps, biology unknown.
	Sclerogibbidae	10	Very rare, koinobiont ectoparasitoids of webspinners (Embioptera).
	Scolebythidae	3	Very rare, tropical ectoparasitoids (probably idiobiont) of wood-boring beetle larvae.
Vespoidea	Bradynobaenodae	200	Very rare, cosmopolitan, may be koinobiont ectoparasitoids of sun-spiders (Solipugida). Great sexual dimorphism.
	Formicidae	10,000	Ants. Extremely common and cosmopolitan. Most are eusocial but also includes social parasites of other ants and slave makers.
	Mutilidae	5,000	Velvet ants. Not true ants, commonest in arid tropics, these are idiobiont ectoparasitoids of aculate larvae and pupae in their cells. Great sexual dimorphism.

(continues)

TABLE II *Continued*

Superfamily	Family	Described extant species	Notes
Vespoidea (cont.)	Pompilidae	4,000	Spider wasps. Sometimes large. Some are ectoparasitoids of spiders *in situ;* most relocate spider prey to a new site.
	Rhopalosommatidae	35	Moderately common, ectoparasitoids of crickets (Gryllidae).
	Sapygidae	80	Kleptoparasitoids of solitary bees. Usually uncommon but occasionally a pest.
	Scoliidae	300	Often large, primarily tropical or warm temperate, idiobiont ectoparasitoids of subterranean beetle larvae.
	Sierolomorphidae	10	Very rare, Americas and Oriental region, biology unknown.
	Tiphiidae	1,500	Often common, mainly tropical, idiobiont ectoparasitoids mainly of subterranean beetle larvae. Great sexual dimorphism.
	Vespidae	4,000	Very common, includes the familiar social wasps or yellowjackets, mason or potter wasps. Females progressively provision their larvae with chewed insect/spider tissue, or in Masarinae, pollen.
Apoidea	Apidae	30,000	Bees, from solitary to highly social. Progressively provision larvae with pollen.
	Sphecidae	8,000	Almost entirely solitary, nest-building predators of insects and spiders.

[a]The Chrysidoidea, Vespoidea, and Apoidea comprise the aculeate Hymenoptera. The true number of species in the Ceraphronoidea, Platygastroidea, Diapriidae, Chalcidoidea, and Ichneumonoidea, in particular, are likely to greatly exceed the figures given here.

thoracic segment, the size of the flight muscles would be greatly restricted. By having the first abdominal segment fused to the thorax, larger flight muscles can be accommodated. Thus, the middle body part of an apocritan hymenopteran is comprised of the pro-, meso-, and metathorax, plus the first abdominal segment, the latter being termed the propodeum. Of course, this nomenclature has often led to confusion among those who lack detailed familiarity with wasp physiology. Nowadays, to avoid ambiguity, it is becoming increasingly common to refer to the middle body region as the mesosoma and the part behind it as the metasoma. Further nomenclatural confusion can arise when, as in the ants and some parasitic wasps, the first metasomal segment (i.e., the second abdominal segment) is greatly reduced. The most conspicuous part of the metasoma is then referred to as the gaster.

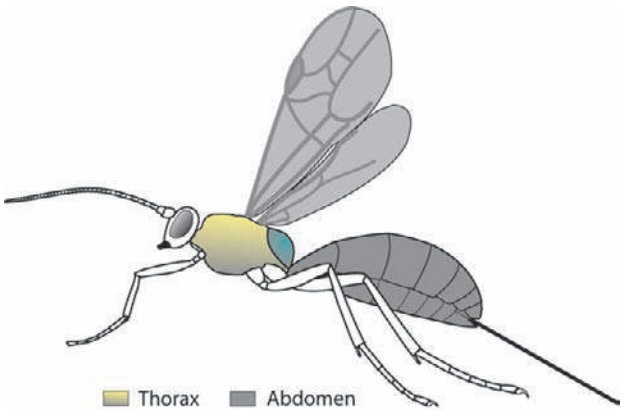

FIGURE 1 Stylized illustration of an apocritan hymenopteran showing the wasp waist between the first and second abdominal segments. The first abdominal segment, called the propodeum, is broadly attached to the thorax (colored blue), and the combined structure is referred to as the mesosoma. The abdomen (darker shading) posterior to the mesosoma is often called the metasoma.

Venoms

Most if not all Hymenoptera, even the sawflies (which are phytophagous), have a "venom" gland associated with the ovipositor or sting. In fact, most have at least two distinct glands, the venom gland proper, also referred to as the acid gland on account of its typical histological staining properties, and a second gland, referred to as the alkaline gland or Dufour's gland in the Aculeata (see later). Details of the function of the secretions of these glands are known for only a few species; for the sawflies, such knowledge is almost completely lacking. It may reasonably be assumed that the initial function was production of lubricants that assisted passage of the egg down the ovipositor; but natural selection would have acted quickly to favor organisms that showed an ability to benefit from modification of the substrate.

In the parasitic wasps, the venom gland products are thought to be primarily associated with two roles: to help overcome the host's immune defense against the parasitoid's egg or larva, and sometimes, especially among the ectoparasitic idiobionts, to paralyze the host. In some taxa, perhaps in many, there may also be secretory products from parts of the female reproductive tract itself, and these may play important roles in overcoming host immunity. Of great interest among these are the "polydnavirus soups" produced by the calyx gland in a few groups of Ichneumonoidea; this gland is a modified part of the lateral oviduct.

The venom glands in the aculeate wasps are the source of the well-known, pain-inducing toxins that many social and some solitary Hymenoptera use to such good effect in self-defense. One solitary aculeate, a mutillid wasp (or velvet ant), is commonly called the camel-killer because its venom is reputedly strong enough to have that effect, and there are anecdotal reports of soldiers who have been incapacitated by the pain caused by encounters with this substance. Typically, however, the venoms are rather less fearsome. These pain-

causing venoms are very specialized and contain a variety of neurotoxins, including, in some taxa, small ringlike peptides that insert themselves in cell membranes and cause depolarization of nerve cells, and consequently pain. It is easy to envisage how these peptides could have evolved from toxins that were originally selected to cause paralysis of the arthropod prey of their ancestors, although probably most are so highly modified that any initial similarity has been lost. A few of the larger parasitic wasps have also developed pain-causing venoms for defense, but their stings are quite mild and the effects short-lived compared with those of many aculeates.

The main function of Dufour's gland seems to be the production of substances that are involved in intraspecific communication. In the parasitic wasps, the gland probably serves primarily as a source of marking pheromones that indicate where an egg has been laid, likely to minimize self- and intraspecific superparasitism. Among the social aculeates, the functions of this gland have been greatly elaborated, and it is the source of many other pheromones that are involved in colony organization.

The Ovipositor: A Key Organ

Since Hymenoptera in general are known to take great care in the placement of their eggs, it is not surprising that the ovipositor is an important organ, and one that has shown many specializations for particular modes of life.

The hymenopteran ovipositor is derived from abdominal appendages and comprises three independently movable parts, called valves, that together form the egg canal. The dorsal valve is a fused structure, but the ventral ones are separate. There is no intrinsic ovipositor musculature; rather, the movements of the valves depend on muscles within the abdomen that pull on the internal apodemes of the three valves. Nevertheless, several parasitic taxa have evolved mechanisms that enable them to steer their ovipositors and thus increase their chances of successfully attacking a mobile host that might otherwise be able to wriggle away from its attacker. Although the penetration of the substrate by wasp ovipositors is usually referred to as "drilling," it is important to realize that there is no circular motion: penetration is achieved by the to-and-fro motion of the three valves relative to one another. In the simplest mode of operation, one valve has a projection or nodus that interlocks with the substrate, and this acts as a support for the others to be pushed forward.

The sawflies get their name from the laterally compressed, strongly serrated, ovipositors with which they insert their eggs under plant cuticle. These ovipositors are unsuited, however, for penetration of wood, and the wood wasps' ovipositors are longer and rounder in cross section, with serrations used for rasping wood fibers, located just at the tip (see Fig. 2). Most of the parasitic Hymenoptera have a similar ovipositor except that in many of those with exposed hosts it is much shorter and has reduced serrations because there is no substrate to "drill" through. Some ovipositors are very long (up to 12 times longer than the wasp's body), and various mechanisms and behaviors have evolved to enable the wasp to manipulate them.

FIGURE 2 A large ichneumonid wasp, *Megarhyssa* sp., using its ovipositor to "drill" through a tree trunk to reach its host, a siricid wood wasp larva. The ovipositor is very thin and pointing between the fore legs; the large black structures are the protective ovipositor sheaths. (Photograph by Nathan Schiff.)

The Road to Parasitoidism

The evolution of parasitoidism has long been of interest, and several possible scenarios have been discussed at various times. As our understanding of the phylogenetic relationships of the basal parasitic wasps has firmed up over recent years, it now seems most probable that the transition to a parasitoid way of life occurred first among some ancestral wood wasp, because the closest extant sister group of the parasitic Hymenoptera is almost certainly the Orussidae; which collectively are derived from wood wasp ancestors. The most widely discussed and generally accepted proposal for the evolution of parasitism in the Hymenoptera envisages an ancestral wood wasp gaining an advantage by producing a larva that could encounter, kill, and eat another wood-boring insect—possibly the larva of another wood wasp or of a beetle—because such a food item would have a greater nutritional value than the plant diet. An advantage therefore would have been gained if the female ancestral parasitoid were to seek out for oviposition sites branches where such food bonuses occurred. This could occur only if the prey item, such as a batch of eggs or perhaps a minimally mobile prepupa or pupa, did not pose a danger to the ancestral parasitoid. The next step would be from a facultative utilization of vulnerable prey insects to supplement a plant-based diet with an obligate one, eventually eliminating the need to consume plant material. So by evolution of the wood wasps, prey species become hosts. At first, as in extant orussids (see later) the larva might have done the final prey location, but the protoparasitoid in this scenario always evolves to use its ovipositor to injure or

kill the prey, perhaps by stabbing. Subsequently, venoms evolved greater sophistication until they were able to induce permanent host paralysis, and even to help keep the host fresher for longer by preventing it from becoming infected with fungi and bacteria.

Gall Formers

As with several other insect orders, some groups of Hymenoptera have evolved ways of making plants produce especially safe and nutritious places (i.e., plant galls) in which to shelter their young. Because hymenopterans produce venoms that contain a wide range of pharmacologically active compounds that affect host insect physiology, it is not very surprising that some of these might affect plants—and the typical plant response to damage is cell proliferation, forming a callus. Thus, venom usage may have preadapted parasitic wasps for the evolution of gall forming (cecidogenesis). We cannot know whether this has always been true, and at least in the extant Cynipidae, the gall wasp family, the phytotoxins that stimulate gall production are produced by the wasp's larva and are not components of the female venom. Gall-forming sawflies, however, do seem to have undergone this preadaptation. True gall wasps (cynipids) have colonized and diversified upon only a few plant groups, notably Rosaceae and Fagaceae, although the other gall-forming Hymenoptera collectively attack a huge range of plants.

One family of chalcidoid wasps, which are also gall formers in a sense, are of particular interest. These are the pollinating fig wasps belonging to the Agaonidae. These insects develop within the ovaries of fig flowers, consuming as larvae the galled tissue. Many people eat figs, but few know that originally at least, the cultivated fig and all its relatives in the genus *Ficus* (Moraceae), relied absolutely on the pollinating activities of these specialized agaonid wasps. Figs are actually not fruit in the strict sense but are syconia—that is, a hollow flasklike structure containing many flowers. The female fig wasps collect pollen from the flowers within the fig in which they have developed, and upon entering another developing fig, actively pollinate the flowers. Figs and their pollinator wasps have evolved in a tight association, and each of the 400 or so species of fig (all members of the genus *Ficus*) has one or, rarely, a couple of pollinating wasp species associated with it that pollinate no other fig species. In these wasps, the males are especially highly modified and do not leave the interior of the fig as adults. Instead they compete with one another for mates and frequently kill competitors with their large mandibles. Males are also responsible for chewing an exit hole through the wall of the fig that enables the female pollinator fig wasps to escape and go in search of new figs.

The Road to Sociality

Perhaps the greatest claim to fame of Hymenoptera is that the order includes several highly social groups of insects, various bees, yellowjacket (vespid) wasps, and ants. The only other insect order with such a large number of highly social species are the termites (Isoptera). Within the Hymenoptera, sociality has evolved on quite a few separate occasions, and much consideration has been given to the reasons for this circumstance. Probably a variety of factors have contributed.

Provision of a good food source for the larvae first involved selection of a suitable host plant species, but with the evolution of parasitoidism, the degree of selectivity increased. Further selectivity is apparent with the very narrow host ranges of many of the parasitic taxa and, also, their ability to assess the suitability of individual hosts of the right species. This may be viewed as a progression in the degree of individual attention provided to each individual offspring. In several lineages, all in the Aculeata, additional behaviors have evolved to make hosts, in a sense, more suitable—at first these changes consisted of moving a host to a slightly preferable location before oviposition. This is seen in several members of the Bethylidae. Some bethylids also show a degree of parental care in that the female remains with her single brood through their development to guard them against predators and to guard the host against other parasitoids including conspecific females. In the bethylid wasps that is about the limit of brood care, but in a number of groups, notably among the pompilids and sphecids, the female wasp prepares a hideaway in which to cache the insect or spider that will provide her offspring with food. Probably at first, once a host had been identified, the female would locate, dig, or modify a burrow; a further evolutionary step was likely the postponement of burrow construction until the search for a host had begun. This stage required the behavioral sophistication of being able to remember the location of the burrow and to relocate it once a prey had been found. This step was crucial for the evolution of sociality because nest members must be able to locate their nests after foraging expeditions. Another major development allowed by this evolutionary advance was the use of hosts, now usually called prey, that are smaller than what would be necessary for the development of the wasp, because it was now possible to bring back multiple individual hosts for each of the wasp's larvae. This is the stage exhibited by many sphecid wasps. In these, the paralyzed prey are first accumulated until there are sufficient, then an egg is deposited on the cache, and this set is sealed into a cell, after which the female starts collecting more prey for her next egg. Bees and vespid wasps have independently dispensed with sealing their larvae in individual closed cells with all the food that they need; instead, they provide food continuously upon demand—preprepared food, that is, rather than whole prey individuals.

Reproduction and Sex Determination

As far as is known, all hymenopterans have a haplodiploid sex determination system, which means that haploid individuals (having only one set of chromosomes and resulting from

unfertilized eggs) are males, whereas diploid individuals (having two copies of each chromosome and resulting from fertilized eggs) are females. This sex determination system is found in a few other groups of organisms, notably in thrips and some rotifers. Development of unfertilized eggs into males is a form of parthenogenesis and is termed arrhenotoky.

There is certainly more than one sex determination mechanism even within this haplodiploid system, and this has important consequences in matters such as biological control. The best understood (or surmised) mechanism, called complementary sex determination, is characterized by the occasional occurrence of diploid (but infertile) males and by the tendency of the proportion of these to increase with inbreeding. Culturing species with this sex determination mechanism is difficult because small population sizes result in the gradual loss of sex alleles and so an increase in frequency of diploid males and loss of colony vigor. In the rearing of insects for biological control programs, the appearance of an abnormally high number of diploid males can be a very serious setback, because when colonies become inbred, they go extinct.

COMPLEMENTARY SEX DETERMINATION (CSD) Although the exact molecular details are unresolved, there is a reasonable hypothesis that CSD may involve polymeric proteins, the most simple of which are dimers. A heterodimers (i.e., an association of two different protein chains) has a different shape from a homodimer (two identical protein chains), and this shape difference determines the sex of the offspring. Because haploid individuals can make only one form of the protein (they only have one gene locus), they must make the homodimeric form, and this means that they will develop as males. It seems that in natural populations there are typically quite a few sex alleles (roughly between 6 and 50), with the result that the proportion of fertilized eggs that contain two copies of the same allele is rather small. Thus the proportion of eggs that either fail to develop or produce infertile diploid males would be expected to be small, as well.

NONCOMPLEMENTARY SEX DETERMINATION Not all Hymenoptera can possess CSD, as is evidenced by the routine inbreeding that occurs, for example, in some parasitic wasps. In these species, a female lays on or in a single host an often large brood of eggs consisting mostly of daughters (from fertilized eggs) and a single haploid male (or at least a very low number of sons), which fertilizes all his sisters when they emerge. This goes on for many generations and undoubtedly must lead to increased homozygosity, but these wasps show no progressive change in sex ratio or fecundity as CSD would necessarily cause. However, what sex determination system is involved in these insects is not known, and while a gene dosage mechanism is widely postulated and seems highly likely, it is not proven.

RELATEDNESS AND MATING A particular consequence of haplodiploidy that has been invoked as a major reason for the multiple independent evolutions of sociality within the Hymenoptera is that it leads to a change in the normal degrees of relatedness between a mother and her sons and daughters and between siblings, and this difference is most pronounced if the female has mated only once. The reasons are as follows. A male offspring gets all his genes from his mother (because he is haploid and so has no male parent), but he gets only half of the mother's chromosomes; in the Hymenoptera, therefore, a son is 50% related to the mother just as in mammals. However, if the female has mated only once, her daughters, which come from fertilized eggs, contain half her set of chromosomes plus 100% of those from the male parent (because he is haploid and all his sperm are identical). Overall, therefore, each daughter is 75% related to each other daughter, whereas daughters are related to their mother only 50%. The argument regarding sociality is that because sisters in Hymenoptera are more closely related to each other than they are to any potential offspring they could have, their fitness will be better enhanced by helping their mother to produce more sisters than by reproducing themselves.

The vast majority of species of social ants, bees, and wasps only mate once. Thus the foregoing arguments may generally hold. However, there are exceptions, and honey bee queens typically mate about a dozen times. Thus for these insects the disparity in relatedness between offspring and sisters is much closer to the 50:50 ratio of normal diploid taxa.

PHYLOGENY, CLASSIFICATION, AND WHAT PARTICULAR GROUPS DO

The Hymenoptera are member of the monophyletic group of insects known familiarly as the Holometabola, and it is among these therefore that their relationships must be sought. The search for such relationships, however, has thus far proved to be rather difficult, and despite a growing body of molecular data, a consensus about the exact relationships of the order has yet to emerge. Weak evidence has been put forward to suggest a relationship with the Mecoptera, but the characters involved are liable to homoplasy. One possibility is that the Hymenoptera form a sister group to the whole of the rest of the Holometabola (or at least, of the extant holometabolan orders).

Classification

For a long time the Hymenoptera have been broadly divided into three groups: the sawflies or "Symphyta," the aculeates, or stinging and social wasps, bees, and ants and their close allies, and the remainder, which are typically referred to as the parasitic wasps or "Parasitica," even though many of them are not parasitic at all. It has long been realized that this is an unnatural arrangement, and slowly an attitude more consistent with our understanding of hymenopteran phylogeny has been filtering into the literature. Thus, although it is still useful to be able to refer to sawflies as a group, the use of a formal classificatory term for them to the

exclusion of all other Hymenoptera (i.e., Symphyta) is rather unsatisfactory. This is because it has long been known that the "Symphyta" are a basal grade that leads to the wasp-waisted Hymenoptera; "Symphyta" are therefore not a monophyletic group (i.e., all the descendants of its single basal species are not included); rather, they are a paraphyletic one, and so from a cladistic standpoint, should be recognized as such. The situation, though, is complicated because most languages, lack a vernacular term for all the sawflies exclusive of the wasp-waisted taxa. Thus, use of informal terms, typically indicated as such by the use of quotation marks, is seen more and more. Terms like "Symphyta" indicate that the name is being used as a handle of convenience and should not be taken as a reference to a monophyletic group.

The wasp-waisted Hymenoptera have also been traditionally classified into two groups, the "Aculeata" and the "Parasitica," the former group being largely (though not entirely) distinguished from the latter by the loss of use of the ovipositor for egg laying, being used instead for stinging prey and/or potential enemies. As with the "Symphyta," the "Parasitica" are no longer recognized as a formal group because there is a reasonable consensus that the aculeates are derived from within them, so rendering them paraphyletic. The Aculeata are strongly supported as monophyletic, at least on morphological grounds.

Sawflies and Wood Wasps

The basal representatives of the order from within which the parasitic and social families of wasps have evolved are the sawflies and wood wasps. There is little doubt that the most ancient extant family of the Hymenoptera is the Xyelidae, represented today by two subfamilies, one whose larvae feed on gymnosperms (either pollen of male cones or on buds or young shoots), and the other having rather caterpillar-like larvae that feed exposed on the leaves of elms and walnut species. Xyelid sawflies possess distinctive antennae (Fig. 3) with the first section of the flagellum very enlarged (possibly as a result of the fusion of multiple segments). Fossil xyelids date from the middle or late Triassic, some 200 mya, and are the earliest known fossil Hymenoptera.

Three other superfamilies of typical sawflies are recognized: the Pamphiloidea, which are relatively uncommon (though they include a few pest species), the Cephoidea or stem sawflies, which are rather slender and feed endophytically in grasses and a few woody plant stems, and the very large and speciose Tenthredinoidea. All these are predominantly Holarctic in distribution, although it is likely that at least the tenthredinoids have moderate species diversity in the tropics, even though they are typically quite uncommon there. Some tenthredinoids form galls (e.g., the genera *Euura* and *Pontania*), and it appears that secretions from the female sawfly contain the cecidogenic compounds. A few are interesting because they show a degree of parental care (Fig. 4), and in some Australian pergids, the caterpillar-like larvae form resting aggregations during the day but disperse over the food plant at night to feed on the leaves, apparently communicating by means of vibrations.

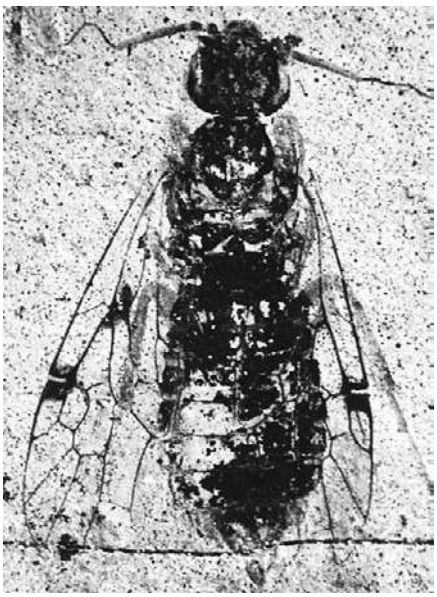

FIGURE 3 *Angaridyela vitimica,* a typical xyelid sawfly from the Lower Cretaceous of Baissa in Siberia (body 10.5 mm long). Note the enlarged third antennal segment. (Photograph by Alexandr P. Rasnitsyn, Paleontological Institute, Russian Academy of Sciences.)

The wood wasps were recognized as a group of three superfamilies, the Siricoidea, Xiphydroidea, and Anaxyeloidea (by Vilhelmsen in 2001), although the greater part of the literature on wasps refers to them simply as Siricoidea. This division is intended to reflect better the phylogeny of the group (Fig. 5), which forms a grade rather than a monophyletic group. Their endoxylous larvae are typically associated with fungus-infected wood, and the adults are responsible for transporting these fungi to their host trees in a special mycangial pouch that is associated with the reproductive system and lies near the base of the ovipositor. In addition to

FIGURE 4 A female pergid sawfly guarding her brood of first instars. (Photograph by Nathan Schiff.)

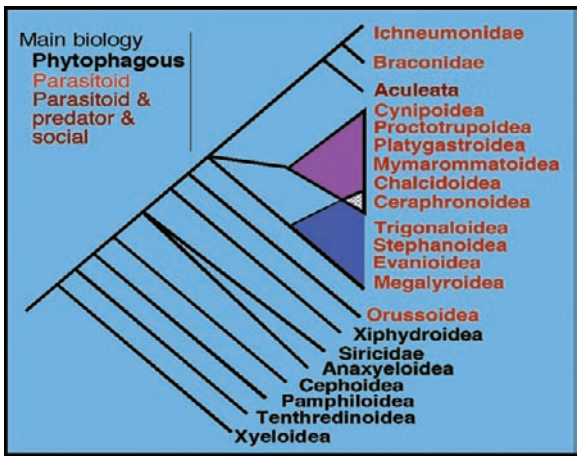

FIGURE 5 Working scheme of Hymenoptera higher level phylogeny. The relationships among the basal, principally phytophagous sawflies and wood wasps are becoming well established, but only a few broad groupings are widely accepted among the Apocrita.

these fungi, at least in some siricid species, components of the secretions injected into the tree at the time of oviposition are actually phytotoxic and may be important in helping their fungi overcome the tree's defenses. Some siricids are important pests of conifer plantations.

There is a very large body of evidence that the Orussidae form the sister group of all the wasp-waisted or apocritan Hymenoptera and, therefore, its biology is of such interest that the family warrants some separate discussion, small though it is. There have been few detailed biological studies on the group, but it is clear that some, probably all, are parasitic on wood-boring insect larvae, although in at least one species the final instar continues to feed on and in its host when it is fairly well decayed. The orussids cannot be considered to represent the ancestral condition of the parasitic wasps as a whole because they show many very derived features not found in any other Hymenoptera. It is tempting, however, to consider at least that their biology is closely similar to and perhaps not modified much from the earliest parasitic wasps. Recent studies have shown that they use vibrational sounding (i.e., echolocation through a solid substrate) to detect their host boring. The females' antennal apices are massive and solid and are used to tap the substrate, and their foretibiae contain massive subgenual organs that are used to detect the vibrations that are transmitted through the wood. This form of host location has evolved on several independent occasions within the order, but the Orussidae are particularly interesting because the egg may be laid into a boring some distance from the host (in the studied species a beetle pupa), and it is the first instar orussid that seeks out the host.

Apocrita or Wasp-Waisted Hymenoptera

In terms of numbers of species, the Apocrita is dominated by parasitoids, but the very great biological developments leading to the evolution of sociality in bees, yellowjacket wasps, and ants has had the effect of polarizing study and discussion, and, in the past, also classification. It has long been appreciated that the aculeate Hymenoptera, which include the social taxa, have evolved from within the Apocrita (in fact, they seem to be most closely related to one particular superfamily, the Ichneumonoidea).

THE "PARASITICA" The "Parasitica" is a paraphyletic group (with respect to the aculeates) that comprises some 11 or 12 superfamilies, the exact number being a little unstable because rapid advances in phylogenetic studies are tending to suggest that some previously recognized taxa are paraphyletic or polyphyletic. Consequently, the number is likely to increase by a few when new evidence leads to robust and better resolved phylogenetic trees. A rather conservative phylogenetic hypothesis is shown in Fig. 5, but even this is problematic. Areas of particular uncertainty are the monophyly of the Proctotrupoidea (even after the Platygastroidea, once included therein, have been removed to a separate group), and of the evaniomorph superfamilies (viz., Megalyroidea, Evanioidea, Trigonaloidea, Ceraphronoidea, and Stephanoidea). Even within this grouping it is not yet totally certain that the three families constituting the Evanioidea form a monophyletic group; they have very different biologies, and no unique synapomorphies have been found to unite the component families.

The vast majority of known and undescribed species are parasitoids of other insects. The term "parasitoid" is used almost universally nowadays to distinguish the interactions of these insects from those of parasites, although purely for reasons of euphony, we still often refer to them as parasitic wasps rather than parasitoid wasps. Whereas true parasites live off the living bodies of their hosts, they seldom kill them; indeed, it would usually be maladaptive for them to do so because the longer the host lives, generally the greater will be the reproductive opportunity of the parasite. Parasitoids, however, always kill their host and treat it as a single meal that will provide all the food necessary for their own development. Once the parasitoid wasp has eaten all of the host that it needs, the host is no longer of use, and so there is no need to leave it to recover. In this respect, parasitoids are rather akin to predators but, unlike predators, they require only a single host (i.e., prey) individual to provide all their needs. (Predators, on the other hand, eat multiple, often very many, prey during their life span.) Even so, a few groups of "parasitic" wasps actually behave more like predators. These include species that attack egg masses of, for example, spiders, and species whose larvae will eat not one egg but many or all in the spider's batch.

IDIOBIONTS AND KOINOBIONTS Almost all parasitoids can be classified into one of two classes defined by whether their hosts continue developing after parasitization (the koinobiont strategy) or whether further host develop-

ment is curtailed at that time (the idiobiont strategy). These strategies explain many other aspects of the wasp's biology, such as longevity, egg development, egg size, and fecundity, and host range, as discussed in greater detail elsewhere in this volume and in works by Godfray and Quicke. There has been general agreement that the first parasitic wasps were probably idiobionts. However, it is far from proven that this means that within any given parasitic wasp family or superfamily, the idiobiont taxa are necessarily ancestral. Indeed, the most parsimonious explanation for the distribution of parasitoid life history strategies on several of the most widely cited phylogenies suggests that endoparasitoids have evolved into ectoparasitoids on many occasions within the Chalcidoidea, though probably not within the Ichneumonoidea.

Aculeate Hymenoptera

The very familiar, often social, ants (Formicidae), bees (Apidae or Apiinae depending on classification system), and yellowjacket wasps (Vespidae) have rather dominated the traditional classifications of Hymenoptera. These taxa are united with a number of other, rather less familiar ones to form a monophyletic group called the Aculeata. Most members of this clade have derived biologies, though some families are still functionally parasitoids. In these, however, the ovipositor proper is not used for reaching hosts, and only in a few (possibly derived) cases is it actually used for passing the egg. Even among the essentially parasitic species there is a strong tendency toward physically manipulating hosts. Many spider wasps (Pompilidae), for example, drag paralyzed hosts to a hiding place before ovipositing on them, sometimes biting off the spiders' legs to facilitate handling and carrying. The least-derived biology is to use a preexisting cavity as a hiding place, and subsequent evolution has led to many levels of modification or *de novo* nest construction.

SOCIAL KLEPTOPARASITES AND SLAVE MAKING The great resources afforded by social insect nests have attracted the evolution of numerous thieves, some of which are other hymenopterans, often closely related to the species being robbed. The most interesting of these interactions are exhibited by social kleptoparasites when a female of a nonsocial species replaces the queen of a social one and makes use of the workers in the host colony to rear her own brood, rather than relying on more of their own siblings. This ability obviously depends on sophisticated mimicry, not just visual, but more particularly chemical and tactile, or otherwise the usurped workers would detect the intruder. Social kleptoparasitic taxa have evolved independently on numerous occasions and are known among bees (Apidae), vespid wasps, and ants. "Slave making," another form of social parasitism, is known only among the ants, but within this family it has been evolved by several different lineages. The workers of one species are abducted by the slave-making species and forced to forage for the latter.

RELATIONSHIPS TO MAN

Beneficial Species

The number of beneficial hymenopterans greatly outweighs the number of harmful ones, though some taxa fall into both categories depending on circumstance. Even the vespid wasps (yellowjackets), which are well-known nuisances at picnics and barbecues, and occasionally have serious medical consequences, are responsible for eating a very large number of other insects (their larvae are fed almost entirely on chewed-up insect muscle), and in agricultural settings undoubtedly devour many pest insects. Humans have made use of the voracious insect-eating capacities of some ants for many years. For example, Chinese citrus growers have traditionally transferred tree ants, *Oecophylla* spp., into their orange groves to consume potential pests, and central European foresters have had considerable success in controlling pest outbreaks by transporting into forests the nests of wood ants, *Formica rufa.*

Biological control programs have made a great deal of use of parasitic Hymenoptera to control host pest populations, sometimes with spectacular results, and when such controls work, the cost–benefit ratio is very favorable. It is becoming increasingly possible to use commercially produced parasitic wasps to control pests in private gardens and greenhouses as well as on large commercial enterprises, and this has obvious desirable features such as reducing the need to apply pesticides.

Bees and Honey

Probably best known of the beneficial affects of Hymenoptera is the production of honey, which people have valued as a sweetener since prehistory. Originally various honey-producing wild bee nests would have been harvested, and indeed are still harvested by indigenous peoples on several continents. However, one species, the European hive bee or honey bee, *Apis mellifera,* was found to be both productive and manageable, and it was effectively domesticated by getting it to nest in artificial hives several thousand years ago. The first records of beekeeping come from ancient Egyptian wall paintings, some 2500 years B.C. Bees also produce wax from which their larval cells are formed, and this is also widely used for a variety of purposes, from candles to cosmetics and from pharmaceuticals to polishes.

Although most people think of *A. mellifera* when they talk of bees, there are in fact more than 25,000 bee species in the world, the majority of which are solitary rather than social. Collectively, bees are of immense economic importance because of their major role in plant pollination, which includes the pollination of a large number of crop species, and this service is far more valuable than the production of honey and beeswax. In addition to the economic importance of bees, most ecosystems rely on them for pollination, largely because bees have evolved, diversified, and specialized together with the angiosperm plants they pollinate, and it cannot be doubted that loss of many bee species would have

a long-term dramatic impact on the floral composition of many habitats. Interest has been increasing of late in non-*Apis* bees because of the realization that they may be excellent pollinators of many crops, sometimes better than honey bees, and that changes in habitat use have led to a decline in many of these useful species. Additionally, *Varoa,* a parasitic mite that has spread from its natural eastern honey bee host, *A. cerana,* onto the common honey bee, has resulted in serious losses of the latter and thus in pollination rates.

All bees feed their larvae on pollen that is harvested and processed by the adult females (worker bees among the highly eusocial taxa). Typically the pollen is mixed with nectar, although plant oils are sometimes used. The honey made, for example, by *Apis* is not a food for the bee larvae at all, but rather a high-energy food source that is produced and stored by the social bees for times when their normal source of energy food, nectar, is in short supply or simply unavailable—as in temperate winters.

Pests and Medical Importance

Considering the great abundance and species diversity of the Hymenoptera, it could easily be argued that the order includes very few pests. There is general awareness that many bees and most social wasps are capable of delivering a quite painful sting, but for most people these stings are relatively minor, if unpleasant, quickly forgotten occurrences. In the United States approximately 50 people die each year as a result of hymenopteran stings. To put this statistic into perspective, about 25 people are killed each year by lightning in the United States.

The major cause of death from hymenopteran sting is respiratory block, sometimes as a result of a sting in the inside the throat as a consequence of inhaling a wasp or bee, but also often because of more systemic reactions. Approximately 2% of the population are hypersensitive to Hymenoptera stings, but although such sensitization puts the victims at a greater risk, only about one death per year in the United States is attributable to a hyperallergenic response.

Some sawflies and gall wasps are economically important pests of crop and ornamental plants, but many of these have potential roles in the biological control of weeds as well. Notable pest sawflies include both external foliage feeders such as the pine sawflies (Diprionidae), which can be major defoliators of coniferous forests, and concealed feeders such as the wheat stem sawfly, *Cephus cinctus* (Cephidae) and the siricid wood wasp, *Sirex noctilio,* with the latter having caused considerable harm to pine *(Pinus radiata)* plantations in Australia and New Zealand.

Probably the largest number of pest taxa are to be found among the ants. There are a number of introduced taxa in many parts of the world that cause considerable damage to crops, are harmful to livestock, and have nuisance value to humans. Three of many possible examples will serve to illustrate the range of harmful interactions that can occur. The tiny pharaoh ant, *Monomorium pharaonis,* a tropical species widely introduced into buildings in the Northern Hemisphere, often gets into hospitals, where it is a potential vector of bacteria—it has the habit of getting into almost anything in its foraging, including under bandages. The Argentine ant, *Linepithema humile,* is a highly invasive species that fights and outcompetes other ants in introduced regions, and its huge colonies cause damage in agricultural situations because they protect aphids and other honeydew-forming pest insects. Third, the red imported fire ant, *Solenopsis invicta,* which has spread alarmingly in the southern United States in recent years, is very aggressive, especially when its mounds are disturbed. Fire ants will sting people, pets, and livestock many times, with resulting pain, blisters, and even systemic reactions, and small pets are sometimes killed by them. Interestingly, many of these pests are far less harmful in their native regions.

IDENTIFICATION

As might be expected with such a large insect order, identification of hymenopteran species, but also genera and even families, is not without its difficulties. Even in parts of the world where the insect fauna is quite well known, such as Europe and North America, there will be many groups for which there are no satisfactory identification keys. In less well known parts of the world, almost any reasonably sized sample will contain numerous undescribed species and within some families even genera. The presence of so many unclassified hymenopterans reflects a combination of innate taxonomic difficulties such as the insects' small size, the large number of similar species, and often a great deal of superficial convergence. In addition, because the order, with few exceptions, has not attracted a great deal of amateur attention, relatively little work has been done on it. Recent years have, however, seen vast improvements to the situation. Several well-illustrated keys to all families have been published, as well as several major works on the more popular aculeates. Particularly useful works are by Gauld and Bolton and Goulet and Huber.

See Also the Following Articles

Ants • Apis Species • Division of Labor • Sex Determination • Sociality • Venom • Wasps

Further Reading

Askew, R. R. (1971). "Parasitic Insects." Heinemann Educational, London.

Askew, R. R. (1984). The biology of gall wasps. *In* "Biology of Gall Insects" (T. N. Ananthakrishnan, ed.), pp. 223–271. Edward Arnold, London.

Cook, J. M. (1993). Sex determination in the Hymenoptera—A review of models and evidence. *Heredity* **71,** 421–435.

Gauld, I. D., and Bolton, B. (1988). "The Hymenoptera." Oxford University Press, Oxford, U.K.

Godfray, H. C. J. (1993) "Parasitoids: Behavioral and Evolutionary Ecology." Princeton University Press, Princeton, NJ.

Goulet, H., and Huber, J. T. (eds.). (1993). "Hymenoptera of the World: An Identification Guide to Families." Agriculture Canada.

Greathead, D. J. (1986). Parasitoids in classical biological control. *In* "Insect Parasitoids." (J. Waage, and D. Greathead, eds.). Academic Press, London.
Hölldobler, B., and Wilson, E. O. (1990). "The Ants." Belknap Press of Harvard University Press, Cambridge, MA.
O'Toole, C., and Raw, A. (1991). "Bees of the World." Blandford, London.
Quicke, D. L. J. (1997). "Parasitic Wasps." Chapman & Hall, London.
Rasnitsyn, A. P., and Quicke, D. L. J. (eds.) (2002). "The History of Insects." Kluwer, Dordrecht, the Netherlands.
Ronquist, F., Rasnitsyn, A. P., Roy, A., Eriksson, K., and Lindgren, M. (1999). Phylogeny of the Hymenoptera: A cladistic reanalysis of Rasnitsyn's (1988) data. *Zool. Scripta* **28,** 13–50.
Thompson, G. H. (1960). "The Alder Woodwasp and Its Insect Enemies." Film rereleased as video by World Educational Films, http://www.worldeducationalfilms.com.
Vilhelmsen, L. (2001). Phylogeny and classification of the extant basal lineages of the Hymenoptera (Insecta). *Zool. J. Linnaean Soc. (Lond.)* **131,** 393–442.
Wilson, E. O. (1971) "The Insect Societies." Belknap Press of Harvard University Press, Cambridge, MA.

Hypermetamorphosis

John D. Pinto
University of California, Riverside

Hypermetamorphosis is a form of complete insect metamorphosis or holometaboly in which at least one of the instars in the life cycle differs considerably from the others. The term heteromorphosis, preferred by some entomologists, carries a degree of ambiguity in that it also refers to the relatively minor differences characterizing consecutive instars in virtually all insects, as well as to the phenomenon of organ replacement following mutilation. Hypermetamorphosis is most common in parasitoids, where it usually is the first instar that deviates structurally and behaviorally from the others. In some groups one or more of the subsequent instars also are distinctive. In the same way that holometabolous development allows a division of function between the larva and adult insect, hypermetamorphosis can be viewed as an exaggerated form of holometaboly characterized by additional division of function within the larval stage.

GENERAL CHARACTERISTICS AND TERMINOLOGY

Two broad categories of hypermetamorphosis can be recognized in insects. In the most widespread form, there is a decoupling of oviposition site and the larval food; in the other, the oviposition and larval feeding sites are identical. For convenience, these can be referred to as type I and type II hypermetamorphosis, respectively.

Type I adult females do not oviposit directly at the larval feeding site; instead, the first instars must find the food source. Such larvae are active, slender, and well sclerotized (Fig. 1a); they are further characterized by their ability to exist for a considerable time without nourishment and without becoming desiccated. Depending on the group, the first instar may attain the feeding site by direct searching, or indirectly by phoresy, in which it is carried to its food usually by the host itself. Once on the larval food, the first instar begins feeding, and subsequently molts into a grublike and less mobile larva (Fig. 1b). Phoretic larvae are commonly equipped with elongate caudal cerci and a terminal suction process (or pygopod), which allows them to stand erect, thus facilitating contact with passing hosts. Although comparative studies are few, hypermetamorphic taxa seem to be more fecund than nonhypermetamorphic relatives. This apparent difference is attributed to the reduced likelihood of larvae finding suitable hosts. Two general characteristics of adult

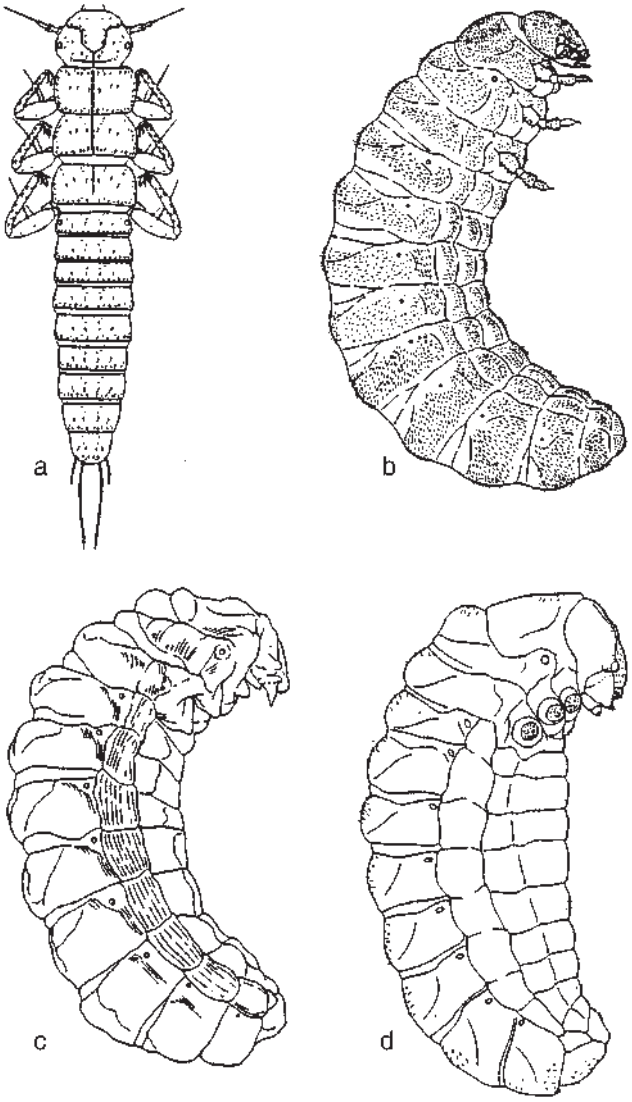

FIGURE 1 Larvae of *Meloe dianellus* (Coleoptera: Meloidae) illustrating the four types of meloid larvae: (a) first instar (planidium with legs), (b) first grub (fifth instar), (c) coarctate (sixth instar), and (d) second grub (seventh instar). Natural lengths: a, 2 mm; b–d, 7–10 mm. [From Pinto, J., and Selander, R. (1970). Illinois Biological Monographs, No. 42.]

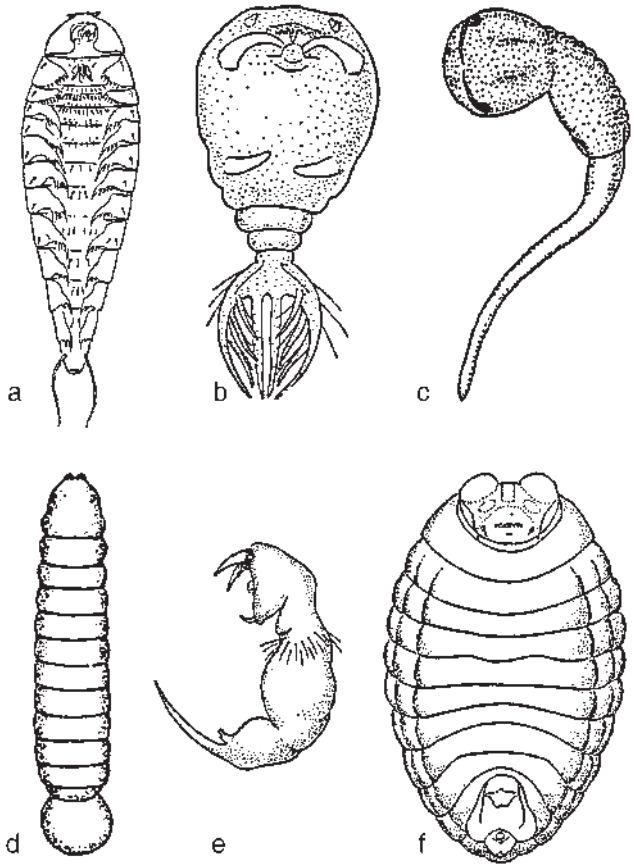

FIGURE 2 Larvae of hypermetamorphic Hymenoptera. (a) first instar *Perilampus hyalinus* (planidium without legs) [from Smith, H. (1912) *U. S. Bur. Ent. Tech. Ser.,* Bull. 19]; (b) first instar *Platygaster* sp. (cyclopiform) [from Kulagin, N. (1898), *Zeitschr. Wiss. Zool.* **63,** 195–235]; (c) first instar *Aridelus* sp. (mandibulate and caudate) [from Kirkpatrick, T. (1937), *Trans. Roy. Ent. Soc. London* **86,** 247–343]; (d) second instar *Apanteles* sp. (vesiculate) [from Allen, W. (1958) *Hilgardia* **27,** 1–42]; (e) first instar *Hadronotus ajax* (teleaform); and (f) third instar, *H. ajax* [from Schell, S. (1943) *Ann. Ent. Soc. Amer.* **36,** 625–635].

females with this form of development are the absence of a well-developed ovipositor and the laying of eggs in masses rather than one at a time. Type I hypermetamorphosis is found in the Strepsiptera (all groups), Neuroptera (Mantispidae), Coleoptera (several families), Diptera (Nemestrinidae, Acroceridae, most Bombyliidae, some Tachinidae), Hymenoptera (Perilampidae, Eucharitidae, some Ichneumonidae), and Lepidoptera (Epipyropidae).

Although there is some inconsistency in usage, the active first instar in type I hypermetamorphic taxa has generally been referred to as a *triungulin* or triungulinid if it has legs (Neuroptera, Coleoptera, Strepsiptera, Lepidoptera; Fig. 1a), or as a *planidium* if it is legless (Diptera, Hymenoptera; Fig. 2a). However, triungulin is an inappropriate term for most of the groups it is applied to. Usage stems from the fact that the first instar of *Meloe* (Coleoptera: Meloidae), the first hypermetamorphic group studied, has trident-shaped claws. Such claw structure does not occur in other hypermetamorphic families and is uncommon in the Meloidae itself. Thus, planidium, signifying "little wanderer," although usually restricted to the larvae of Hymenoptera and Diptera, is perhaps more appropriately applied to all type I first instars. This usage is followed here and was by adopted R. E. Snodgrass in his 1954 review of insect metamorphosis.

In type II hypermetamorphosis the oviposition and larval feeding sites do not differ. In these insects the first instar (sometimes the second as well) has a distinctive morphology (Fig. 2b–e) compared with the relatively simple and often amorphous subsequent instars (Fig. 2f), but these modifictions are associated with functions other than host finding. The adaptive significance attributed to the often bizarre larvae of type II hypermetamorphic groups includes locomotion, protection and combat, ingestion, and respiration. This type of development occurs in several families of parasitic Hymenoptera and in the dipteran family Cryptochetidae *(Cryptochetum).* Correlated with the greater possibilities of larval function, the morphology of type II first instars is considerably more variable than that found in type I taxa.

Reviews by Clausen and Hagen identified 12 structural forms of first instars in the parasitic Hymenoptera with type II hypermetamorphosis. The generalized type in nonhypermetamorphic species (and in hypermetamorphic species as well after the first instar) is referred to as hymenopteriform. It is ovoid or fusiform and relatively featureless (Fig. 2f). Examples of deviations in hypermetabolic groups include the following. The caudate type has a taillike prolongation of the terminal segment and is found in many Ichneumonoidea and Chalcidoidea (Fig. 2c). It is believed to be an adaptation for locomotion within the host and/or food absorption. The mandibulate and cyclopiform types have enlarged falcate mandibles and are found in several groups (e.g., Ichneumonoidea, Diapriidae, Platygasteridae) (Fig. 2b,c). They are associated with feeding as well as combat in cases of multiple or superparasitism. The vesiculate larva (several Ichneumonoidea) is similar to the hymenopteriform type, except its hindgut is evaginated to form an external vesicle to aid in respiration within the host (Fig. 2d). Several of these larval types are characteristic of taxonomic groups. For example, the cyclopiform larva (Fig. 2b) characterizes the Platygasteridae, and the teleaform larva (Fig. 2e) is typical of Scelionidae.

Hypermetamorphic development not falling into types I or II occurs in the Hydroptilidae in the order Trichoptera (caddisflies), a group with aquatic larvae. There are modifications in shape, setation, sclerotization, and leg development in the fifth instar not found in the four preceding instars. The first four instars of hydroptilids are free living, whereas the fifth lives within a case constructed of varying materials.

The term hypermetamorphosis generally is applied only to the Holometabola. However, the validity of the practice of excluding hemimetabolous groups belonging to the Sterrnorhyncha (Homoptera), namely, the Aleyrodidae (whiteflies) and Coccoidea (scale insects), is questionable.

Although these groups are phytophagous they also have instars with considerable differences in structure and function. The first instar, or crawler, is responsible for locating the feeding site and for dispersal. The following instars are sessile and modified for a sedentary existence.

Type I hypermetamorphosis with its planidium larva is widespread in insects and is an excellent example of convergence in evolution. It is convenient to summarize the occurrence of this form of development by surveying the groups where it is known to occur.

SURVEY OF INSECT TAXA DISPLAYING TYPE I HYPERMETAMORPHOSIS

Neuroptera

Only the Mantispidae have true hypermetamorphosis. Most of the species feed on spider eggs associated with a single egg sac. The first instar either enters a previously constructed egg sac or attaches onto a female spider and enters the sac as she constructs it. Other species feed on the larvae of various aculeate Hymenoptera. Several of these are phoretic and reach the food source by attaching to the adult bee or wasp.

Lepidoptera

The only lepidopteran family with hypermetamorphosis is the Epipyropidae. This group is parasitic on various Homoptera, a unique association for the generally phytophagous Lepidoptera. The active first instar seeks out a host and the grublike instars that follow occur on the body of a single homopteran.

Coleoptera

Hypermetamorphosis occurs in several families of beetles. Most are parasitoids of other insects. The larvae of Bothrideridae, and a few Carabidae and Staphylinidae, attack the larvae or pupae of other coleopterans or dipterans. The genus *Sandalus* (Rhipiceridae) feeds on cicada nymphs. All are hypermetamorphic to some degree, with active and relatively long-legged planidia followed by short-legged, grublike feeding instars.

Hypermetamorphosis perhaps is best known in the Meloidae and Rhipiphoridae. Most meloids parasitize grasshopper eggs or, more commonly, the larvae and provisions of soil- or wood-nesting bees. The planidia of several groups are phoretic on adults of their hymenopteran host. There are four distinctive larval types in the typical meloid life cycle: planidium, first grub, coarctate, and second grub (Fig. 1). The planidium (Fig. 1a) encounters the food source. The first grub (Fig. 1b), consisting of four instars, is the primary feeding stage. The coarctate (Fig. 1c), a quiescent instar, is adapted for overwintering or aestivation. The second grub (Fig. 1d) is a nonfeeding instar that precedes pupation. Each has a distinct phenotype, the most unusual being the immobile coarctate with its thick, highly sclerotized cuticle, aborted appendages, closed mouth and anus, and vestigial musculature.

The Rhipiphoridae include parasitoids of immature Hymenoptera (Rhipiphorinae) and Blattodea (Rhipidiinae). The planidia of Rhipiphorinae are phoretic on adults of their host. After being carried to the host nesting cells, they burrow into the body of the host larva, eventually molting into a grub that emerges to feed externally. The planidia of Rhipidiinae attach directly to their cockroach host and eventually molt into a legless and amorphous larva, which enters the host to feed. The last instar regains poorly developed legs, exits the host, and moves away for pupation.

Among nonparasitoid groups of Coleoptera, hypermetamorphosis is known to occur in one genus of Eucnemidae *(Rhacopus)* and in the Micromalthidae. The larvae of both groups feed in decaying wood and have a planidial-like first instar. The single species of Micromalthidae, *Micromalthus debilis,* has perhaps the most complicated life cycle known in insects, with several distinct morphological and reproductive types of larvae. Hypermetamorphosis has also been reported in the Megalopodidae.

Strepsiptera

The Strepsiptera parasitize Thysanura, Orthoptera, Hemiptera, Diptera, and Hymenoptera. The free-living planidium encounters the host. If parasitic on a hemimetabolous host, the larva may immediately penetrate its body and develop. Those parasitizing Hymenoptera are phoretic and are carried to the nesting site, where the immature stages are attacked. The planidium molts into an endoparasitic legless grublike larva, which may have several instars. This secondary larva lacks mouthparts and feeds by diffusion through the cuticle. Male Strepsiptera are free living; females are neotenic and most remain on their host.

Diptera

Hypermetamorphosis is known in the Acroceridae, Nemestrinidae, most Bombyliidae, and some Tachinidae. All have legless, well-sclerotized, and active planida that search for the host, followed by soft-bodied maggotlike larvae. The Acroceridae are internal parasitoids of spiders. The Nemestrinidae are endoparasites of grasshoppers and beetles. The Bombyliidae are parasitoids of immature Lepidoptera, Hymenoptera, Coleoptera, Neuroptera, and Diptera; some prey on grasshopper eggs. They are either endo- or ectoparasitic.

Hymenoptera

A legless planidial larva (Fig. 2a) is known in three families of Hymenoptera: Perilampidae, Eucharitidae, and Ichneumonidae *(Euceros).* It is followed by a soft-bodied, relatively immobile

larva. The Eucharitidae are endoparasitic on mature larvae or pupae of ants and are phoretic. Eggs are laid on vegetation. The planidia of most groups attach to worker ants and are carried to the nest, where they transfer to larvae for feeding. The Perilampidae is a closely related family, similar bionomically to the Eucharitidae but known to parasitize several orders of insects. In the Ichneumonidae, the planidium of *E. frigidus* attaches to the integument of sawfly larvae and eventually transfers to feed on the larva of other ichneumonid species that are primary parasites of the sawfly.

See Also the Following Articles

Host Seeking, by Parasitoids • Larva • Metamorphosis • Phoresy

Further Reading

Clausen, C. P. (1940). "Entomophagous Insects." McGraw-Hill, New York.
Clausen, C. P. (1976). Phoresy among entomophagous insects. *Annu. Rev. Entomol.* **21,** 343–368.
Crowson, R. A. (1981). "The Biology of Coleoptera." Academic Press, London.
Hagen, K. S. (1964). Developmental stages of parasites. *In* "Biological Control of Insect Pests and Weeds" (P. H. DeBach and E. Schlinger, eds.), pp. 168–246. Chapman & Hall, London.
Snodgrass, R. E. (1954). "Insect Metamorphosis." Smithsonian Miscellaneous Collections, 122 (9). Smithsonian Institution, Washington, DC.
Stehr, F. W. (ed.) (1991). "Immature Insects," Vol. 2 (see Chaps. 33–35, 37). Kendall Hunt, Dubuque, IA.

Hyperparasitism

Daniel J. Sullivan
Fordham University

Hyperparasitism is a highly evolved behavior in the Hymenoptera and in a few species of Diptera and Coleoptera, in which an adult hyperparasitoid (or secondary parasitoid) oviposits on or in a primary parasitoid host that has attacked another (usually herbivorous) insect species. The larval offspring of the hyperparasitoid cause the death of the primary parasitoid. Ecologists emphasize this interaction as a food-web "community." This article focuses on the hymenopteran microwasps in which hyperparasitism occurs.

There are a variety of behaviors by hyperparasitoids depending on the species of secondary and primary microwasp parasitoids which in turn are influenced by the species of phytophagous host, often an insect pest. In addition, there is an economic interest in hyperparasitism because if primary parasitoids are considered to be beneficial insects when used in biological control programs, it would seem that hyperparasitoids that attack primary parasitoids would be detrimental. However, hyperparasitoids may play a positive role by preventing extreme oscillations of the primary parasitoids that might reduce the numbers of the phytophagous host enough to cause the local elimination of both the insect pest and the beneficial primary parasitoid.

EVOLUTION

Hyperparasitism has evolved in only three insect orders: in Hymenoptera (in 17 families) and in a few species of Diptera and Coleoptera. Its evolution was preceded by that of primary parasitism that evolved in the Hymenoptera during the Jurassic, about 135 mya. In the primary parasitoids, ectophagous feeding probably evolved before endophagous, with the parasitoid egg deposited near or on the host rather than in it. Hence, ectophagous parasitoids usually attacked concealed hosts, often within galleries in wood or plant galls. The use of venom by primary parasitoids apparently developed very early and produced physiological changes in the host. Although the venom of the more ancestral ectophagous parasitoids resulted in idiobiosis (permanent paralysis or death), the venom of the specialized endophagous species tended toward koinobiosis (temporary or nonlethal paralysis).

Facultative hyperparasitism probably evolved from primary ectophagous parasitoids because few special adaptations are needed to oviposit and feed externally on a primary parasitoid as well as on the primary's phytophagous host. Obligate hyperparasitism has a wider taxonomic distribution and may have evolved via facultative hyperparasitism as an opportunistic behavior to specialize only in attacking readily available primary parasitoid hosts—especially if they share similar physiological and/or ecological attributes. Hence, it is not surprising that hyperparasitoid species can be either ecto- or endophagous, whereas some are idiobionts and others are koinobionts.

The host spectrum of hyperparasitoids is broader at the species level than that of primary parasitoids, but hyperparasitism is usually restricted to immature stages of hymenopteran hosts (larvae and/or pupae) that are natural enemies mainly of phytophagous insects in the Hemiptera (mainly suborder Sternorrhyncha), Lepidoptera, and the hymenopteran suborder Symphyta. Hyperparasitoids rarely attack the egg and adult stages of primary parasitoids. Also interesting is that some families of Hymenoptera that are well known for their species of primary parasitoids (Braconidae, Trichogrammatidae, Aphidiidae, Mymaridae, and almost the entire superfamily Proctotrupoidea) do not seem to have evolved any hyperparasitoids. Similarly, in the order Diptera, hyperparasitoids are absent in some important parasitic groups such as the family Tachinidae.

MULTITROPHIC ECOLOGY

There are two complementary ways that ecologists look at the interacting food-web community involving hyperparasitoids. One aspect is the "bottom-up" effect beginning with the first trophic level that shows both inter- and intraspecific plant

variation influencing the ecology and behavior of the second trophic level of phytophagous (herbivorous) insects. This in turn is one of the fundamental determinants of the third trophic level of entomophagous (carnivorous) insects such as the primary parasitoid microwasps. Finally, insect hyperparasitism is the highly evolved fourth trophic level, wherein a secondary parasitoid microwasp oviposits on or in the primary parasitoid and kills it. The other complementary aspect is the "top-down" view by which the hyperparasitoid at the fourth trophic level exerts selective pressure on the primary parasitoid at the third trophic level. The next interaction is with the phytophagous insect at the second trophic level that in turn is feeding on the plant at the first trophic level. This food-web community can have economic importance if the plant is an agricultural food crop or even a forest used for commercial lumbering or as a park.

APHID COMPLEX

Over many decades, studies of hyperparasitism have been conducted on the primary parasitoid microwasps in the Hymenoptera that attack the Hemiptera in the suborder Sternorrhyncha, and in particular the superfamily Aphidoidea, with special emphasis on the family Aphididae. The aphid–primary parasitoid–hyperparasitoid food web has been used as a model system in community ecology partly because of the economic importance of aphids as worldwide pests on a variety of agricultural crops and forests, but also because of the relative ease of rearing aphids, their primary parasitoids, and hyperparasitoids in the laboratory and/or greenhouse for precise behavioral and ecological studies. Primary aphid parasitoids are found in only two families of Hymenoptera: all genera of the Aphidiidae (Braconidae: Aphidiinae), and the Aphelinidae (*Aphelinus* and related genera). These primary parasitoids of aphids are in turn attacked by many genera of hyperparasitoids in three hymenopteran superfamilies such as Chalcidoidea (Pteromalidae: *Asaphes;* Encyrtidae: *Syrphophagus;* Eulophidae: *Tetrastichus*), Cynipoidea (Alloxystidae: *Alloxysta*), and Ceraphronoidea (Megaspilidae: *Dendrocerus*).

Aphid hyperparasitoids can be divided into two categories based on adult ovipositional and larval feeding behaviors:

1. The female wasp of endophagous species such as *Alloxysta* (*Charips*) *victrix* (Fig. 1A) deposits her egg inside the primary parasitoid larva while it is still feeding on and developing inside the live aphid, but before the aphid has become mummified (a mummy is the hardened exoskeleton of the dead aphid that remains attached to the leaf). Being a koinobiont hyperparasitoid, the larva usually does not hatch until after the mummy has been formed by the primary parasitoid larva. Then the hyperparasitic larva feeds internally on the primary parasitoid larval host.

2. The female wasp of ectophagous species such as *Asaphes lucens* (Fig. 1B) and *Dendrocerus (Lygocerus) carpenteri* (Fig. 1C) deposits her egg on the surface of the primary parasitoid larva after the aphid has been killed and mummified. To do this, the female must first drill a hole in the mummy in which to deposit her egg. Then the hyperparasitic larva feeds externally on the primary parasitoid larva while both are still inside the mummy. The venom of species varies such that *Asaphes* is an idiobiont, whereas *Dendrocerus* is a koinobiont. An unusual species is *Syrphophagus (Aphidencyrtus) aphidivorus,* whose females display a "dual ovipositional" behavior by attacking primary parasitoid larvae within both living aphids (Fig. 1D), and also inside dead aphid mummies (Fig. 1E). Either way, *S. aphidivorus* develops as an endophagous koinobiont hyperparasitoid.

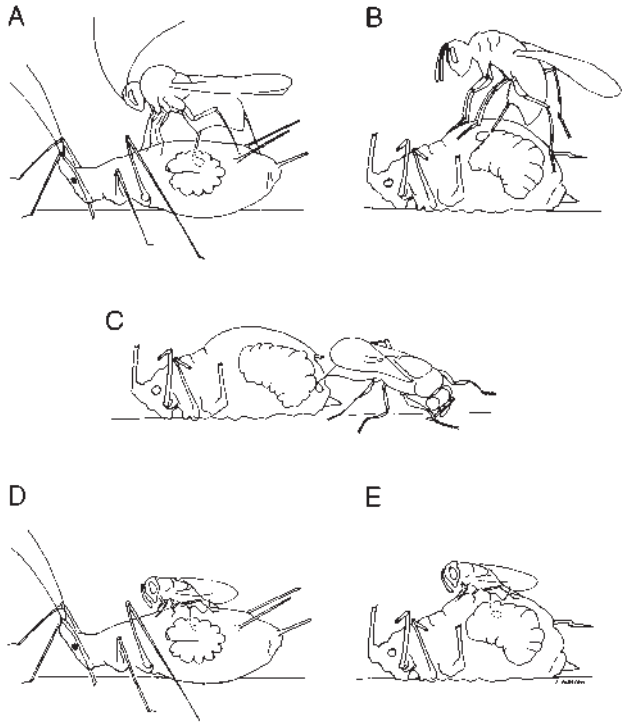

FIGURE 1 Female ovipositional behavior of four genera of aphid hyperparasitoids. (A) Endophagous koinobiont *Alloxysta victrix* jumps on a live parasitized aphid and deposits her egg internally inside the primary parasitoid microwasp larva while the aphid is still alive, but before mummy formation. (B) Ectophagous idiobiont *Asaphes lucens* stands on top of a dead aphid mummy, drills a hole, and deposits her egg externally on the surface of the primary parasitoid larva developing inside the mummy. (C) Ectophagous koinobiont *D. carpenteri* stands on the leaf, backs into the dead aphid mummy, drills a hole, and deposits her egg externally on the surface of the primary parasitoid larva developing inside the mummy. (D) and (E) "Dual ovipositional" behavior of endophagous koinobiont *S. aphidivorus.* (D) *Syrphophagus* stands on top of a live parasitized aphid and deposits her egg internally inside the primary parasitoid larva while the aphid is still alive, but before mummy formation. (E) *Syrphophagus* stands on top of a dead aphid mummy, drills a hole, and deposits her egg internally inside the primary parasitoid larva developing inside the mummy. [Reprinted from Sullivan, D. J. (1988). Aphid hyperparasites. *In* "Aphids, Their Biology, Natural Enemies and Control" (A. K. Minks, and P. Harrewijn, eds.), Vol. 2B, 192, with permission from Elsevier Science.]

However, in all species of aphid hyperparasitoids, further development is similar to that of primary parasitoids, such that pupation is also inside the dead aphid mummy. Then the single adult hyperparasitoid emerges by cutting a hole in the dorsum of the mummy. After pulling itself out, the new adult male or female is ready for mating. From the time of the female's attack on the primary parasitoid larva/pupa, the period of hyperparasitoid development from egg deposition to adult emergence varies with different hyperparasitoid species from as short as 16 days to 25 or more.

NONAPHID COMPLEX

Hyperparasitism exists not only with aphids, but also with almost all insect taxa. There is a complex of primary and secondary parasitoids associated with other members of the suborder Sternorrhyncha (scale insects, whiteflies, mealybugs, psyllids), as well as in various orders such as the Lepidoptera (gypsy moth, leafminers, budworms, stem borers), Diptera (gall makers, leafminers), Hymenoptera (cynipid gall makers, leafcutter bees, sawflies), and Coleoptera (lady beetles, weevils).

BIOLOGICAL CONTROL

Biological control uses predators, parasitoids, and pathogens to reduce an insect pest's population to an acceptable level. Because primary parasitoids are considered beneficial, hyperparasitoids may interfere with such programs. It has been the policy in biological control to exclude "exotic" (nonindigenous) *obligate* hyperparasitoids by enforcing quarantine procedures during importation. It is less easy to decide what to do with "exotic" *facultative* hyperparasitoids when no exclusively primary parasitoids are available, and so perhaps admitting these could be beneficial. Such dilemmas are evaluated with caution, and the solution depends on the seriousness of the insect pest problem. On the other hand, "indigenous" hyperparasitoids (whether obligate or facultative) already exist in the ecosystem and may or may not interfere with the beneficial "exotic" primary parasitoid.

CONCLUSIONS

Hyperparasitism intrigues entomologists because of its multidisciplinary relationship to evolution, ecology, behavior, biological control, taxonomy, and mathematical models. More field studies are needed to determine whether hyperparasitoids are always detrimental to biological control programs. Perhaps, instead, they could have a beneficial influence by regulating the extreme/detrimental population oscillations of the beneficial primary parasitoids.

See Also the Following Articles

Biological Control • Parasitoids

Further Reading

Godfray, H. C. J. (1994). "Parasitoids: Behavioral and Evolutionary Ecology." Princeton University Press, Princeton, NJ.

Gordh, G. (1981). The phenomenon of insect hyperparasitism and its taxonomic occurrence in the Insecta. *In* "The Role of Hyperparasitism in Biological Control: A Symposium" (D. Rosen, ed.), Publication 4103, pp. 10–18. University of California Press, Berkeley.

Holt, R. D., and Hochberg, M. E. (1998). The coexistence of competing parasites. II: Hyperparasitism and food chain dynamics. *J. Theor. Biol.* **193,** 485–495.

Mackauer, M., and Völkl, W. (1993). Regulation of aphid populations by aphidiid wasps: Does parasitoid foraging behaviour or hyperparasitism limit impact? *Oecologia* **94,** 339–350.

Price, P. W., Bouton, C. E., Gross, P., McPheron, B. A., Thompson, J. N., and Weis, A. E. (1980). Interactions among three trophic levels: Influence of plants on interactions between herbivores and natural enemies. *Annu. Rev. Ecol. Syst.* **11,** 41–65.

Stiling, P., and Rossi, A. M. (1994). The window of parasitoid vulnerability to hyperparasitism: Template for parasitoid complex structure. *In* "Parasitoid Community Ecology" (B. Hawkins, and W. Sheehan, eds.), pp. 228–244. Oxford University Press, Oxford, U.K.

Sullivan, D. J. (1987). Insect hyperparasitism. *Annu. Rev. Entomol.* **32,** 49–70.

Sullivan, D. J. (1988). Aphid hyperparasites. *In* "Aphids, Their Biology, Natural Enemies and Control" (A. K. Minks, and P. Harrewijn, eds.), Vol. 2B, pp. 189–203. Elsevier, Amsterdam.

Sullivan, D. J., and Völkl, W. (1999). Hyperparasitism: Multitrophic ecology and behavior. *Annu. Rev. Entomol.* **44,** 291–315.

Whitfield, J. B. (1998). Phylogeny and evolution of the host–parasitoid relationship in the Hymenoptera. *Annu. Rev. Entomol.* **43,** 129–151.

I

Imaginal Discs

Seth S. Blair
University of Wisconsin-Madison

The term "imaginal disc" is used to describe structures found in the larvae of the Holometabola. Holometabolous insects can be defined as those in which the final instar metamorphoses into a radically different adult during a quiescent pupal stage; they are thought to be a monophyletic group, distinct from the Hemimetabola. During the metamorphosis of Holometabola, the epidermis must form novel structures that were lacking in the larva and, in some insects larval tissues that were lost must be replaced. The cells that give rise to the new epidermal tissues of the adult (imago) are often referred to as histoblasts. When histoblasts are organized into morphologically distinct clusters, these structures are commonly referred to as imaginal discs or imaginal buds. This article discusses this definition and briefly reviews some of the experimental studies examining the biology and, especially, the development of imaginal discs.

WHAT IS AN IMAGINAL DISC?

The terms "imaginal disc" and "histoblast" are more pragmatic than precisely defined, and their usage varies from author to author because imaginal discs vary considerably between taxa in number, morphology, and development. For example, some authors have defined imaginal discs and histoblast cells based on their undifferentiated state and early appearance in development, since in some taxa imaginal discs are formed in the embryo. These early-developing discs may sometimes secrete a thin cuticle-like substance, but they do not contribute appreciably to larval life and thus can be considered to be specialized, relatively undifferentiated structures set aside for adult development. However, in other taxa the imaginal discs cannot be detected until the final instar and are apparently derived from cuticle-secreting, differentiated larval cells. Several authors have argued that late-developing discs reflect the ancestral condition within the Holometabola.

The morphology of imaginal discs also varies. The typical imaginal disc is a pocket or sac of cells that has invaginated from the larval epidermis and is destined to form part or all of an adult appendage, compound eye, or genitalia. One portion of the sac has thickened to form the disc epithelium, whereas the rest forms a thinner peripodial membrane; the space within the sac is termed the peripodial cavity (Fig. 1). This sac evaginates ("everts") during metamorphosis and contributes to the adult cuticle (Fig. 1). However, the positions within the disc of the disc epithelial cells and peripodial cells vary, as does the disc's degree of invagination, and some authors have defined several categories of discs or quasidiscs. For example, some discs have invaginated a large distance from the larval epidermis and remain connected to it by only a thin peripodial stalk. At the other extreme, as in the wings of some Coleoptera, the regions of disc epidermis that form adult structures never invaginate from the larval epidermis. Some of these discs can still be recognized as thickenings of the larval epidermis. However, in other examples, such as the leg of the lepidopteran *Pieris brassicae,* the histoblasts cannot be easily recognized, and the novel portions of a single adult appendage develop from several zones of dividing cells within the larval appendage.

Thus, sometimes it is only its eventual contribution to a novel adult structure that makes a cell a histoblast, distinct from neighboring cells in the larval epidermis. Moreover, the organization of histoblasts into a morphologically identifiable imaginal disc is simply one of several ways these cells can be arranged. And though the term "imaginal disc" is used only for holometabolous insects, it is only the invaginated morphology of most discs that distinguishes them from structures such as the external wing pads of hemimetabolous

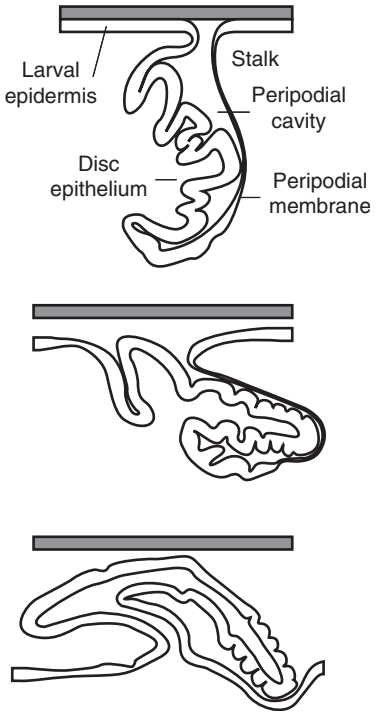

FIGURE 1 Eversion of *D. melanogaster* leg disc, shown in cross section. [After Blair, S. S. (1999). *Drosophila* imaginal disc development: Patterning the adult fly. *In* "Development-Genetics, Epigenetics, and Environmental Regulation" (V. E. A. Russo, D. Cove, L. Edgar, R. Jaenisch, and F. Salamini, eds.), p. 348. © Springer-Verlag GmbH & Co. KG, Heidelberg.]

pterogotes, which also are morphologically distinct and form adult structures. Thus, the evolution of invaginated wing imaginal discs from the evaginated wing pad may have required only a morphological change, rather than the evolution of a radically different cell type or organ. In fact, a similar evolution has apparently occurred in some Hemimetabola, such as thrips, which have invaginated wing precursors instead of external wing pads.

EXPERIMENTAL STUDIES ON IMAGINAL DISCS

The biology of imaginal discs has been described in a number of species but has been experimentally analyzed in only a few. As with other metamorphosing organs, there have been studies on the endocrine control of imaginal disc eversion, differentiation, and cuticle secretion. Other types of experiment, such as ablation or transplantation of portions of imaginal discs, have been used to study aspects of developmental patterning, such as the development of pigmentation patterns in the wings of butterflies and moths. And in a sense, all studies of the evolution of adult Holometabola reflect on the histoblast cells and discs from which portions of those adults are formed.

However, the most extensively studied imaginal discs are those of the fruit fly, *Drosophila melanogaster.* This species, which has been the subject of intense genetic studies for a century, has become a model for the study of a variety of biological problems. Gaining an understanding of the imaginal discs in this species has been particularly critical because the entire adult epidermis is derived from imaginal discs or disclike structures. Thus, any mutation that alters the external form of the adult *D. melanogaster* does so by altering the development of these imaginal structures. Thus a large literature has appeared on the genetics, development, and cell biology of these structures and, via comparisons with other drosophilids, a number of evolutionary studies have been published, as well.

The imaginal discs of *D. melanogaster* are at one extreme in the spectrum of imaginal tissues, and it is not clear to what extent the mechanisms underlying this apparently derived state are shared by other Holometabola. As in other cyclorrhaphan Diptera, the metamorphosis of *D. melanogaster* is unusual in several respects. First, the cuticle of the last instar is not shed during the formation of the pupa (pupariation), as occurs in most Holometabola. Rather, the larval cuticle is retained and converted into an outer covering, the pupal case. Second, in *D. melanogaster* the nondisc cells of the larval epidermis are polyploid (containing more than the diploid number of chromosomes), and these die during the early stages of metamorphosis. In many other Holometabola, large portions of the larval epidermis and appendages are retained in the adult. Finally, in *D. melanogaster* the imaginal disc primordia are formed during embryonic development, rather than during the last instar as they are in some other Holometabola.

In *D. melanogaster* embryos each imaginal disc primordium contains 10 to 40 cells, which divide during the three instars to form as many as 50,000 cells by late third instar. The arrangement of imaginal discs in the late third instar is shown in Fig. 2. The adult head is derived from a pair of fused eye–antennal discs, as well as pairs of proboscis (labial) and labral (cibarial or clypeolabral) discs. The adult thorax is derived from the three pairs of ventral leg discs and, dorsally, pairs of prothoracic (humeral), wing, and haltere discs. Each of these contributes to the body wall, and also forms the appendage for which it is named. Each segment of the adult abdomen is formed from four pairs of small "histoblast nests" and the genitals from the genital disc.

At the late third instar each imaginal disc consists of a simple epithelial sac, invaginated from the larval epithelium (Fig. 1). One surface of each sac forms the thickened and folded disc epithelium, from which most of the adult structures are derived. The other surface of the sac is the thinner peripodial membrane, and each disc remains connected to the larval epithelium by a long, narrow stalk. During the first few hours of pupariation each disc everts through the stalk, expands, and eventually sutures together with adjacent discs. The peripodial membrane is lost during this process, and the polyploid cells of the larval epidermis are histolysed and replaced. The portions of each disc epithelium that are fated to form the appendages unfold and lengthen: the prospective legs and antennae form long tubes, whereas the prospective

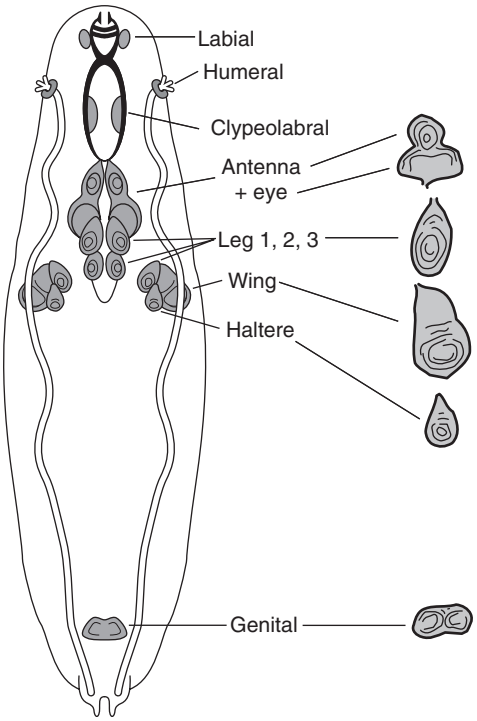

FIGURE 2 Imaginal discs in late third instar *D. melanogaster.* [From Blair, S. S. (1999). *Drosophila* imaginal disc development: Patterning the adult fly. *In* "Development-Genetics, Epigenetics, and Environmental Regulation" (V. E. A. Russo, D. Cove, L. Edgar, R. Jaenisch, and F. Salamini, eds.), p. 348. © Springer-Verlag GmbH & Co. KG, Heidelberg.]

dorsal and ventral surfaces of the wing lengthen and flatten together. The disc cells then secrete the pupal cuticle, which is separated from the epithelial surface shortly thereafter. After a delay during which there is further morphogenesis and differentiation, the adult cuticle is secreted.

Many aspects of the biology of imaginal discs of *D. melanogaster* have been examined over the years and continue to be the subjects of intense research. These include the hormonal control of disc development and cuticle secretion, the molecular genetics of cell division and cell growth, and various problems in cell biology. Perhaps the most striking advances, however, have been made in developmental patterning by examining for the most part the large wing, leg, and eye–antennal discs. Space constraints prevent the discussion of many fascinating and important research areas, such as the patterning of the compound eye, the formation of sensory organs, and the planar polarity of the disc epithelium. Instead, the following discussion briefly reviews some early developmental events that help establish the identities and axes of the discs.

PATTERNING IMAGINAL DISCS IN *D. MELANOGASTER*

The identity of each disc primordium is specified during embryonic development. This specification relies on the disc-

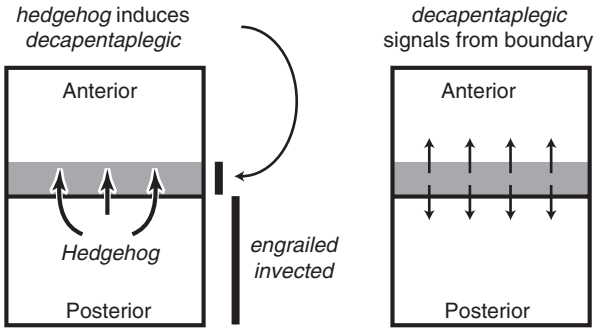

FIGURE 3 Signaling and anterior boundary cells in wing disc of *D. melanogaster.*

specific expression of a small number of genes, encoding largely transcription factors or their cofactors. The loss or misexpression of these genes causes dramatic homeotic transformations of disc identities. The localized expression of these transcription factors is to a large extent inherited from or signaled by the embryonic epidermis, which was subdivided into anterior–posterior and dorsal–ventral domains prior to the formation of the disc primordia.

How are different tissues formed within each disc? Even at late third instar most imaginal disc cells appear morphologically similar; only a few (mostly sensory) elements have begun to differentiate terminally. However, the premetamorphic imaginal disc is in no sense a *tabula rasa,* since it has already been subdivided into number of regions. In fact, most disc primordia are subdivided into anterior and posterior lineage "compartments" from the earliest stages of their development. This subdivision is controlled by the posterior-specific expression of the *engrailed* and *invected* transcription factors in each disc, which act as a binary switch, controlling the choice between posterior and anterior identities and preventing cells from crossing between compartments. The wing and haltere discs are further subdivided into prospective dorsal and ventral lineage compartments, in this case by the dorsal-specific expression of the *apterous* transcription factor.

Why have lineage compartments? Cells in different lineage compartments have different signaling capabilities, and this has important developmental consequences. Posterior cells secrete the signaling molecule Hedgehog, but only anterior cells are capable of responding to that signal (Fig. 3). Because Hedgehog diffuses only a short distance into the anterior compartment, it induces the formation of a specialized group of cells just anterior to the compartment boundary. The formation of this group of boundary cells is critical for appendage development because boundary cells in turn secrete several important signals (Fig. 3), including Decapentaplegic (a member of the BMP family of morphogens, generated in the wing, dorsal leg, and dorsal antenna), and Wingless (a member of the Wnt family of morphogens, generated in ventral leg and ventral antenna). Cells outside the boundary region judge their approximate position in the disc by the

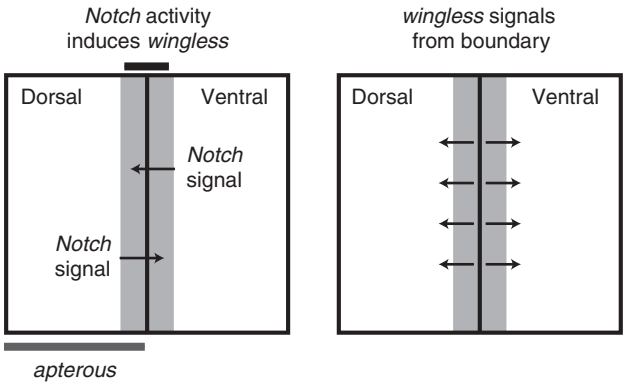

FIGURE 4 Signaling and dorsal–ventral boundary cells in the wing disc of *D. melanogaster.*

levels of boundary signal they detect. Thus, these boundaries define important axes used to pattern the entire disc.

The subdivision of the wing into dorsal and ventral compartments by the dorsal expression of *apterous* plays a very similar role. This subdivision establishes reciprocal signaling between dorsal and ventral cells, this time via the *Notch* pathway, which results in the specification of cells on either side of the dorsal–ventral boundary (Fig. 4). These cells secrete Wingless, which helps pattern the wing blade along the proximodistal axis.

Do the other Holometabola share these axis-defining patterning mechanisms? There have been to date only a few descriptions of gene expression in the imaginal discs of other taxa, and no experimental tests of the type that verified these patterning mechanisms in *Drosophila.* However, it seems likely that at least some of the fundamental features are shared. All insect appendages so far examined (both larval and adult) are subdivided into apparent anterior and posterior compartments by the posterior expression of *engrailed*-like transcription factors, including the wing discs of the lepidopteran *Precis coenia.* Similarly, the wing blade of *P. coenia* is subdivided into dorsal and ventral domains by the dorsal expression of an *apterous*-like molecule. As in *D. melanogaster,* a *wingless*-like molecule is expressed along the dorsal–ventral compartment boundary in *P. coenia.*

See Also the Following Articles

Development, Hormonal Control of • *Drosophila melanogaster* • *Embryogenesis*

Further Reading

Blair, S. S. (1999). *Drosophila* imaginal disc development: Patterning the adult fly. *In* "Development-Genetics, Epigenetics and Environmental Regulation" (V. E. A. Russo, D. Cove, L. Edgar, R. Jaenisch, and F. Salamini, eds.), pp. 347–370. Springer-Verlag, Heidelberg.

Carroll, S. B., Gates, J., Keys, D. N., Paddock, S. W., Panganiban, G. E. F., Selegue, J. E., and Williams, J. A. (1994). Pattern formation and eyespot determination in butterfly wings. *Science* **265,** 109–114.

Fristrom, D. K., and Fristrom, J. W. (1993). The metamorphic development of the adult epidermis. *In* "The Development of *Drosophila melanogaster*" (M. Bate and A. Martinez Arias, eds.), pp. 843–897. Cold Spring Harbor Laboratory Press, Plainview, NY.

Kim, C.-W. (1959). The differentiation centre inducing development from larval to adult leg in *Pieris brassicae* (Lepidoptera). *J. Embryol. Exp. Morphol.* **7,** 572–582.

Švácha, P. (1992). What are and what are not imaginal discs: Reevaluation of some basic concepts (Insecta, Holometabola). *Dev. Biol.* **154,** 101–117.

Tower, W. L. (1903). The origin and development of the wings of Coleoptera. *Zool. Jahrb. Anat. Ontog. Tiere* **17,** 517–572.

Truman, J. W., and Riddiford, L. M. (1999). The origins of insect metamorphosis. *Nature* **401,** 447–452.

Immunology

Nancy E. Beckage
University of California, Riverside

Insects lack immunoglobulins but nevertheless mount a variety of effective immune responses to parasites and pathogens. In permissive or susceptible hosts, either no response is mobilized or the responses induced fail to counter the invader, so that the host is successfully infected. In resistant (also called refractory) hosts, the parasite or pathogen is thwarted, and infection is prevented. Regardless of the outcome, the result is a dynamic interplay of host and parasite genes, with the products of host resistance/susceptibility genes being counterbalanced by the virulence/avirulence characteristics of the invader (Fig. 1).

In insects, both cellular and humoral immune responses figure prominently in host defense, with many parasitic and pathogenic infections resulting in the deployment of defenses of both types. Insect hemocytes mobilize the cellular defenses, which include phagocytosis, nodulation, and encapsulation. In the main classes of humoral defenses, the following substances are produced: antibacterial, antifungal, and (presumably) antiviral molecules, the melanizing enzyme phenoloxidase, and agglutinins or clotting factors.

Although much recent progress has been made in deciphering the cellular and humoral aspects of insect immune responses, much less is known about the recognition mechanisms responsible for the initial discimination of "nonself" material at the host–invader interface that sets the stage for the immune response. This is especially true for the recognition of multicellular parasites, which seem to use a combination of active and passive strategies to avoid being detected as foreign.

CELLULAR IMMUNE RESPONSES

Insect Hemocytes

Initially, hemocytes were classified on the basis of morphological criteria alone, resulting in the publication of numerous

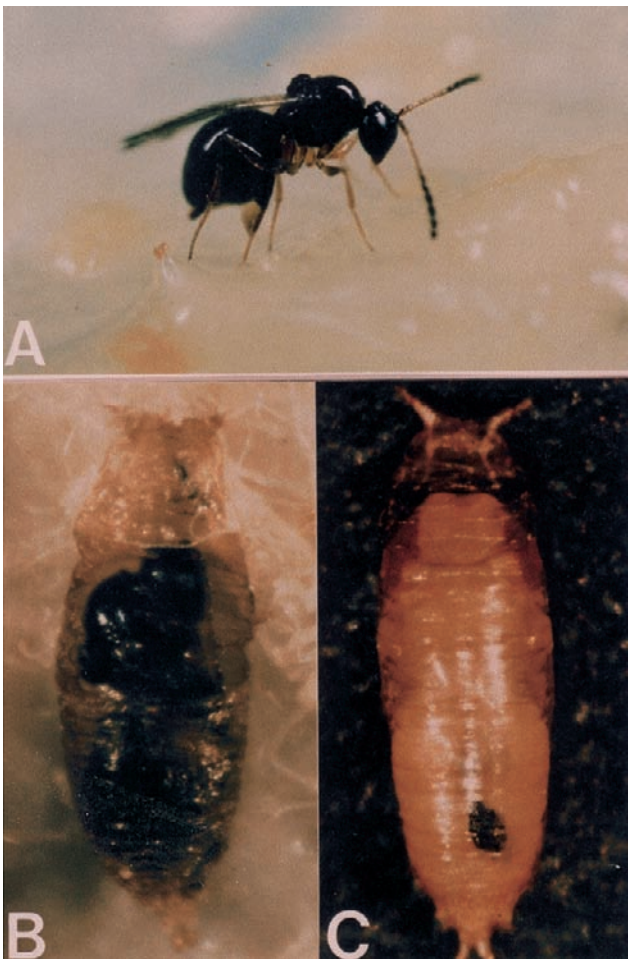

FIGURE 1 (A) The parasitoid wasp, *L. boulardi,* preparing to oviposit in *D. melanogaster* host larvae. If the host strain is of the susceptible genotype, and the wasp is of the virulent strain of this parasitoid species, the wasp will develop normally in the host (B) and successfully emerge while the host dies. However, if the host expresses resistance genes, the parasitoid will be encapsulated and melanized (C), and the host will survive to adulthood. (Photographs courtesy of Dr. Yves Carton, CNRS, Gif-sur-Yvette, France.)

conflicting classification schemes, even for the same species. The different hemocyte types display marked inter-, and even intra-, species variability in appearance and behavior, further complicating classification. Additionally, hemocytes that are examined before they have attached to a substrate are difficult to identify in the unspread state, in contrast to spreading cells (e.g., plasmatocytes). Today, tools such as monoclonal antibodies, which bind cell-type-specific epitopes, are used in combination with other biochemical markers to facilitate identification of the different hemocyte classes. In addition, techniques such as density gradient centrifugation have been employed to purify homogeneous populations of the individual hemocyte morphotypes, facilitating *in vitro* studies of their biochemistry and behavior.

In the Lepidoptera, the two most abundant hemocyte classes are plasmatocytes and granulocytes, which are the primary phagocytic and encapsulative cells. In the higher Diptera including *Drosophila melanogaster,* the multifunctional lamellocytes play these roles, and crystal cells carry the melanizing enzyme phenoloxidase. In contrast, in lepidopterans including the tobacco hornworm, *Manduca sexta,* the bulk of this enzyme is carried in the plasma. Taxonomic variations are seen in the total hemocyte count, with the Lepidoptera having abundant hemocytes in contrast to dipterans (e.g., mosquitoes), which have many fewer cells per microliter of hemolymph.

The fat body is also an immunoresponsive organ and synthesizes a variety of antimicrobial peptides, enzymes (e.g., lysozyme), and other immunoreactive molecules. Indeed, the fat body represents the primary source of hemolymph-borne macromolecules and is the most metabolically active tissue in the insect. The gut, which is constantly assaulted with pathogens ingested by the insect, also produces a battery of antibacterial and antifungal agents to counter these invaders. During molting, when the newly synthesized cuticle is most fragile and the cuticular linings of the fore- and hindguts are shed, bacteria are released into the lumen of the gut, resulting in enhanced production of antimicrobial peptides by the gut when the animal is most vulnerable to infection.

Hemocyte-Mediated Immune Responses

Phagocytosis is the process by which pathogens such as bacteria and small particles (< 1 μm in diameter) are engulfed by host hemocytes, culminating in death of the invader. The membrane of the cell invaginates, and the pathogen is engulfed in a membrane-bound vesicle into which lytic enzymes are released, causing the pathogen's demise. This process appears to be mediated by prostaglandins (eicosanoids) produced by the hemocytes. The phagocytic cells include plasmatocytes and granulocytes, although other cell types may also participate in this response to a lesser degree. Large numbers of bacteria-laden hemocytes may clump together to form nodules, which attach to the host's internal tissues and are removed from circulation. Frequently these nodules are melanized and deposited on lobes of fat body, the Malpighian tubules, or gut tissues.

To counter eukaryotic invaders (i.e., parasites) that are too large to be phagocytosed by a single cell, multiple classes of hemocytes cooperate in the mobilization of the multiphasic encapsulation reaction. Usually the capsule is formed from several hundred to several thousand cells that form a dense capsule enclosing the parasite. In the initial phases of encapsulation of parasitoid eggs/larvae, or encapsulation of abiotic (nonliving) implants such as Sephadex beads, granulocytes are the first cells to make contact with the target. These cells, which contain large numbers of refractile granules in their cytoplasm, attach to the target surface and then release their contents in a degranulation reaction, forming a sticky matrix on the surface of the target. The degranulation event triggers the attachment of multiple layers of plasmatocytes, which are fibroblast-like cells that spread, flatten out, and encase the parasite in a multilayered sheath of cells. Sometimes multiple

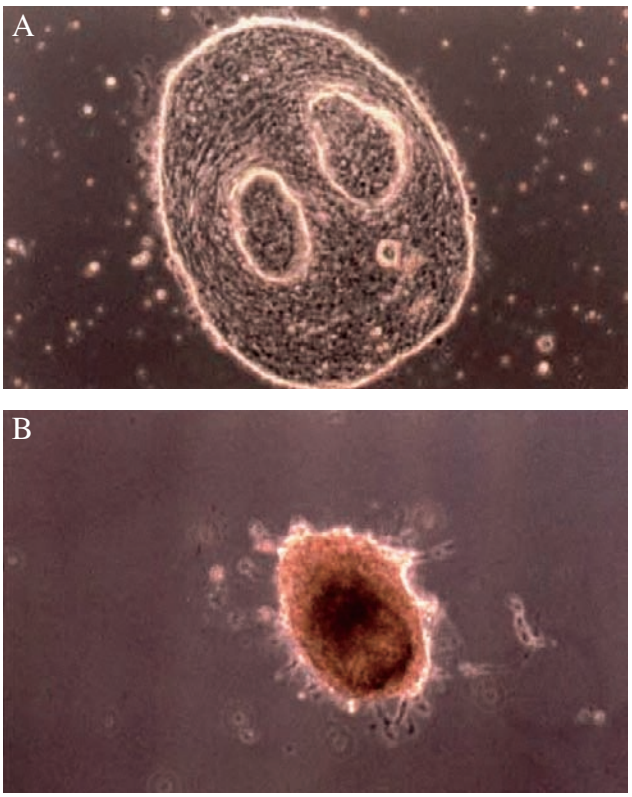

FIGURE 2 Light micrograph series showing fully encapsulated *C. congregata* eggs recovered from the nonpermissive host *P. occidentalis* 72 h postparasitization of the host larva. Multiple eggs are sometimes engulfed in a single capsule (A), which frequently shows signs of partial or complete melanization (B). [Reprinted from Harwood, S. H., *et al.* (1998). Production of early expressed parasitism-specific proteins in alternate sphingid hosts of the braconid wasp *Cotesia congregata. J. Inv. Pathol.* **71,** 271–279, with permission from Academic Press.]

targets are enclosed in a single capsule, as occurs with eggs of the parasitoid *Cotesia congregata,* in the nonpermissive host *Pachyshpinx occidentalis* (Fig. 2A). The encapsulation reaction is terminated by the adherence of additional granular cells to form a thin envelope around the completed capsule. Like nodules, fully formed capsules frequently adhere to the host's internal tissues (including fat bodies, Malpighian tubules, gut, or salivary glands) and are thereby removed from circulation. In rare instances, the encapsulated parasite may actually be extruded from the hemocoel and pass through the epidermis, to be shed with the host's exuvial cuticle in a molt during "cuticular encystment," which occurs when a parasitoid develops in a nonpermissive host.

During the encapsulation process or upon its completion, the innermost layer of cells deposits melanin or its toxic quionone precursors over the surface of the invader, regardless of whether the foreign material is an abiotic or biotic target (Figs. 2B, 3A,B). However, the occurrence of melanization is variable, and sometimes capsules persist for the remaining life span of the host, showing no signs of melanization. Some species of hosts with encapsulated parasitoids undergo metamorphosis and live to the adult stage (e.g., *D. melanogaster* with encapsulated *Leptipolina boulardi* parasitoid egg; Fig. 1), whereas other hosts die prematurely, frequently showing symptoms of endocrine disruption. When the host dies prematurely, the parasitoid fails to survive. When the parasitoid *C. congregata* is encapsulated in host larvae of the tobacco hornworm, which is normally permissive for this parastoid, the "hosts" with encapsulated parasitoids molt to supernumerary instars and then to nonviable larval–pupal intermediates. These developmental symptoms suggest that the host's juvenile hormone titer is elevated to a level high enough to interfere with normal pupation. In these instances, endocrine disruption seems to be mediated by the factors (i.e., polydisperse DNA viruses: polydnaviruses) injected by the wasp into the host during parasitization and thus even in the absence of developing wasps, host development is arrested prematurely.

During parasitization, endoparasitoids belonging to the hymenopteran families Braconidae and Ichneumonidae inject into their lepidopteran hosts polydnaviruses that subsequently play a critical role in suppressing the host immune response. These viruses, which are integrated in the genomic DNA of the wasp and undergo replication only in the female's ovary, rapidly enter host hemocytes following parasitization, and viral genes are expressed. Depending on the host–parasitoid combination, the host's hemocytes either alter their behavior and fail to spread (thereby inhibiting the encapsulation response) or, alternatively, undergo fragmentation and programmed cell death. In *M. sexta,* larvae parasitized by *C. congregata,* massive numbers of dead and dying hemocytes undergo clumping and then are removed from circulation soon after parasitization, resulting in a dramatic drop in the host's total hemocyte count. By 8 days postparasitization, new cells have differentiated from prohemocytes, and the host regains its ability to encapsulate Sephadex beads. However, the living parasitoid larvae remain unencapsulated (Fig. 3A), suggesting that they escape being detected as foreign by another mechanism. However, if second-instar parasitoids are dissected from a host, killed, then implanted into a surrogate "host" caterpillar, the parasitoids are avidly encapsulated and melanized (Fig. 3B), suggesting that something unique about the living parasitoid surface suppresses a host immune response. These simple observations suggest that the living parasitoid larvae either escape being detected as foreign by mechanisms that may involve host antigen mimicry (or masking) or by the presence of specific as yet unidentified surface molecules that prevent their recognition as "nonself" by hemocytes.

The final biochemical events that culminate in death of the parasite remain unclear for the most part. Although melanin and its precursors are toxic, much recent evidence points to the potential role of other toxic molecules such as reactive intermediates of nitrogen (nitric oxide) or oxygen (superoxide), released by cells localized in the innermost layers of the capsule, in causing lethality. Although it was formerly presumed that death was induced by asphyxiation of the parasite or parasitoid inside the capsule, biochemically

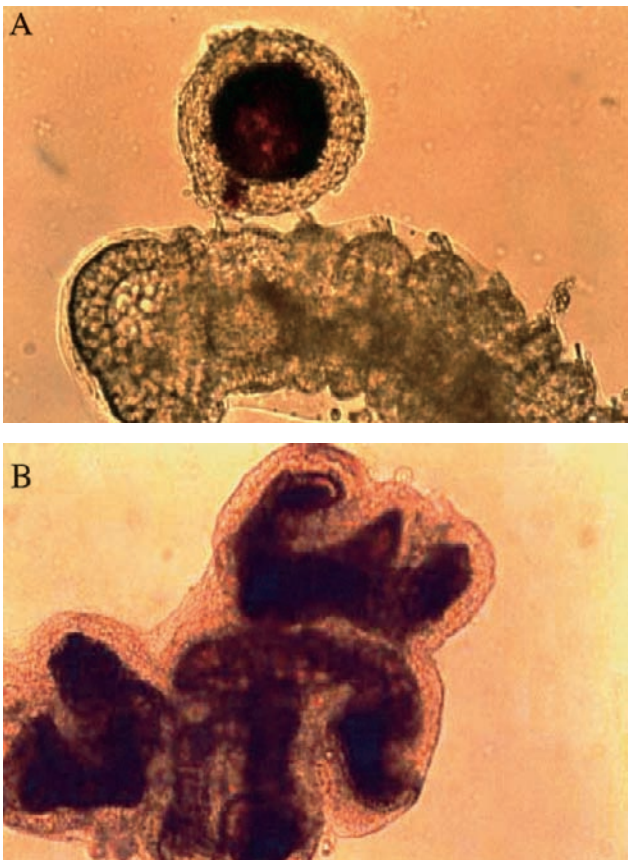

FIGURE 3 (A) Evidence that living *C. congregata* parasitoids are not encapsulated larvae of the host, *M. sexta,* which are permissive for this parasitoid, even though the host may mobilize hemocytes to encapsulate abiotic targets such as Sephadex beads. The bead shows hemocytic encapsulation and associated melanization, whereas the parasitoid larva has avoided this response, suggesting that the living parasitoids actively evade the host's immune response by antigen mimicry, secretion of immunosuppressive molecules, or other mechanisms. (B) Encapsulation of first-instar *C. congregata* that had been dissected from a fifth-instar host tobacco hornworm, killed by immersion in ethanol, then implanted into the hemocoel of a "surrogate host." Thick, agglutinated, and melanized capsules surround the larvae 24 h after implantation. [Photographs from Lavine, M., and Beckage, N. (1995). Polydnaviruses: Potent mediators of host insect immune dysfunction. *Parasitol. Today* **11,** 368–378, with permission from Elsevier Science.]

mediated parasite-killing strategies now seem to be important in causing the ultimate death of the invader. Phagocytosis is followed by the generation of these cytotoxic intermediates in the phagolysozome, indicating that this pathway is also important to the function of phagocytic cells.

The antiviral defenses of insects are just now beginning to be deciphered, although recent observations made in parasitized insects have yielded some insights. Larvae of *M. sexta* that are parasitized by *C. congregata* are dramatically more susceptible to the *Autographa californica* nucleopolyhedrovirus than nonparasitized larvae of the same age, which are normally semipermissive hosts of this pathogenic baculovirus. Parasitized larvae die faster and at higher rates than nonparasitized larvae infected with the same dose of the occluded form of the baculovirus. This enhanced susceptibility appears because of the inactivation of the cellular immune response of the host induced by the wasp's polydnavirus. Thus, there is likely to be a cellular immune response to the baculovirus, which is suppressed in parasitized larvae. In nonparasitized larvae, cellular plaques composed of hemocytes clump on cells localized in the host's tracheal epithelium that harbor the virus. These plaques do not form in parasitized larvae, and the virus is rapidly disseminated throughout the body cavity via the tracheolar epithelium, uninhibited by the hemocytic response. Although the polydnavirus genes that render the host more susceptible to the baculovirus have yet to be isolated, they offer promise for formulation of baculovirus biopesticides with enhanced potency and a broader host range for control of lepiopteran insect pests.

HUMORAL IMMUNE RESPONSES

Plasma-Borne Factors

In addition to hemocyte-mediated immune reactions, insects possess a variety of potent plasma-borne defense molecules that are toxic to parasites and pathogens. Usually these are synthesized by the fat body or the hemocytes and secreted into the plasma, where they act either on the invader directly or via the hemocytes in altering their behavior to enhance the immune response. A battery of antibacterial proteins are produced by many insects, including defensins, drosocin, cecropins, attacins, and diptericins, depending on the species, and, in addition, the ubiquitous lysozyme family of antibacterial proteins. These proteins differ in their specificity for gram-positive versus gram-negative bacteria, with some acting on both types of bacteria with varying degrees of potency. Often the proteins disrupt the bacterial cell membrane function by inducing pore formation, causing lysis of the cell. Antifungal molecules are also produced by insects, providing a first line of defense against fungal invaders that often infect the insect via its cuticle, which is penetrated by the fungus. Although insects are not known to produce antiviral interferon-like molecules, the mobilization of biochemical defenses against viruses seems to be a likely component contributing to the evolution of viral resistance in insect populations treated with viral biopesticides. Although antiviral resistance has been characterized at the insect population level, the cellular and molecular mechanisms contributing to resistance remain relatively ill-defined.

Fortuitously, several antimicrobial peptides such as defensins have also been shown to have antiparasite activity, killing malaria parasites and filarial nematodes in insects that are injected with these molecules. Hence, molecular geneticists are now exploiting defensin genes in the production of transgenic mosquitoes that show up-regulation of defensin gene expression under regulation of tissue-specific promoters either in the gut (malaria) or flight muscle (filaria) where the parasites develop.

In *Drosophila,* mosquitoes, and other insects, the activation of transmembrane *toll* receptors mediates the physiological response to microbial ligands or septic injury, leading to activation of killing mechanisms such as production of nitric oxide, resulting in death of the invader. The *toll* receptors are conserved across a wide range of animal phyla including Mammalia, indicating that this pathway, which insects share with a variety of species, is likely of ancient evolutionary origin.

The phenoloxidase pathway is activated by the synthesis of DOPA from tyrosine via the action of the monophenoloxidase enzyme (also called tyrosinase). Then DOPA is converted to DOPA quinone by diphenoloxidase, and thence to melanin via a series of toxic intermediates. Phenoloxidase activity may be associated with hemocytes, as occurs in mosquitoes, or as in lepidopterans (e.g., *M. sexta*), it may be secreted into the plasma. The first step is the activation of prophenoloxidase by a serine protease, which cleaves a peptide from the proenzyme, generating the active phenoloxidase molecule.

In many species of parasitized lepidopterans, including *M. sexta* larvae parasitized by *C. congregata,* levels of hemolymph phenoloxidase activity have been found to be suppressed following parasitization, which benefits the parasitoid by inhibiting this immunoreaction. This effect seems to be expression of polydnavirus genes that inhibit translation of the phenoloxidase mRNA, thereby suppressing levels of this enzyme in the blood.

In refractory strains of mosquitoes, melanization of the malaria ookinete occurs in the midgut wall, apparently without the intervention of phenoloxidase derived from hemocytes. Disease transmission stops because melanized parasites die trapped in the gut without ever moving to the hemocoel and salivary gland. Hence, there is widespread interest in using phenoloxidase genes to bioengineer refractory transgenic mosquitoes to halt malaria transmission. One approach has been to link this gene to the vitellogenin promoter, which is activated when the mosquito takes a blood meal in preparation for production of eggs.

Cross-Talk between the Cellular and Humoral Immune Response Networks

Blood cells in vertebrates produce many cytokines, which act at close range on other immunocompetent cells. In insects, the characterization of cytokines is less well documented, but recent evidence indicates that factors such as plasmatocyte-spreading peptide (which was first isolated in the moth *Pseudoplusia includens*) are produced by plasmatocytes and act to stimulate spreading of the hemocyte over the surface of the parasite. Other cytokines, which likely play a role in the cell-to-cell communication events that accompany encapsulation, have yet to be characterized. In parasitized lepidopterans, these cytokines may include plasmatocyte and granulocyte depletion factors. Without these cell types, parasite encapsulation cannot occur, and thus the number of viable circulating cells available to mount the encapsulation response is drastically reduced.

UNRESOLVED QUESTIONS

The role of chemical mediators in the cell-to-cell communication pathways required for successful immune responses as complicated as encapsulation have only just begun to be explored. Also, the initial signaling events that discriminate "nonself" at the most basic level remain remarkably obscure. This is particularly true for the recognition of eukaryotic parasites, which seem to use a complex mix of strategies such as host immunosuppression, antigen masking, and antigen mimicry to avoid being detected as foreign. Progress in identifying resistance and susceptibility genes in insect hosts, and the complementary virulence/avirulence genes in parasites has begun only recently, although efforts are intensifying to clarify these genetic components, particularly in *Drosophila.* Although insects lack "memory cells" in their immune system, some long-lived insects such as cockroaches exhibit an enhanced response to a secondary challenge, raising the possibility that insects can show a sensitization response, albeit via as yet unknown mechanisms. Although the antibacterial defenses of insects have been well characterized in the past two decades, even at the molecular level, the defenses mobilized by insects against invasion by viral pathogens are only now being delineated, despite intensive efforts to clarify mechanisms of antiviral resistence. In most species, the roles of hormones such as juvenile hormone and ecdysteroid in regulating immunity have yet to be deciphered. The prospect of developing transgenic insects expressing refractoriness-related traits offers promise of the ability to control disease transmission by insect vectors. Likewise, the prospects of genetic manipulation of agricultural pests to enhance their susceptibility to biopesticides, and to increase the virulence of the pathogen, offer great promise to agriculture. Clearly, many challenging questions need to be addressed to decipher the tactics of parasite/pathogen offense and host defense.

See Also the Following Articles

Genetic Engineering • Hyperparasitism • Pathogens of Insects

Further Reading

Beckage, N. E. (1997). The parasitic wasp's secret weapon. *Sci. Am.* **277,** 50–55.

Beckage, N. E. (1998). Modulation of immune responses to parasitoids by polydnaviruses. *Parasitology* **116,** S57–S64.

Beckage, N. E., Thompson, S. N., and Federici, B. A. (eds.). (1993). "Parasites and Pathogens of Insects": Vol. 1, "Parasites," Vol. 2, "Pathogens." Academic Press, San Diego.

Carton, Y., and Nappi, A. J. (2001). Immunogenetic aspects of the cellular immune response of *Drosophila* against parasitoids. *Immunogenetics* **52,** 157–164.

Dimopoulos, G., Muller, H. M., Levashina, E. A., and Kafatos, F. C. (2001). Innate immune defense against malaria infection in the mosquito. *Curr. Opin. Immunol.* **13,** 79–88.

Gillespie, J. P., Kanost, M. R., and Trenczek, T. (1997). Biological mediators of insect immunity. *Annu. Rev. Entomol.* **42,** 611–643.

Hoffmann, J. A., and Reichhart, J. M. (1997). *Drosophila* immunity. *Trends Cell Biol.* **7,** 309–316.

Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., and Ezekowiltz, R. A. (1999). Phylogenetic perspectives in innate immunity. *Science* **284,** 1313–1318.

Gupta, A. P. (ed.). (1991). "Immunology of Insects and Other Arthropods." CRC Press, Boca Raton, FL.
Imler, J. L., and Hoffmann, J. A. (2001). *toll* Receptors in innate immunity. *Trends Cell Biol.* **11,** 304–311.
Paskewitz, S. M., and Gorman, M. J. (1999). Mosquito immunity and malaria parasites. *Am. Entomol.* **45,** 80–94.
Pathak, J. P. N. (ed.). (1993). "Insect Immunity." Kluwer Academic Publishers, Dordrecht, Netherlands.
Schmidt, O., Theopold, U., and Strand, M. (2001). Innate immunity and its evasion and suppression by hymenopteran parastioids. *Bioessays* **2,** 344–351.
Vass, E. and Nappi, A. J. (2001). Fruit fly immunity. *BioScience* **51,** 529–535.
Wiesner, A., Dunphy, G. B., Marmaras, V. J., Morishima, I., Sugumaran, M., and Yamakawa, M. (eds.). (1998). "Techniques in Insect Immunology." SOS Publications, Fair Haven, NJ.

Industrial Melanism

Michael E. N. Majerus
University of Cambridge

Industrial melanism may be defined as a proportional increase of dark, or melanin, pigments in individuals of a population, caused by changes in the environment resulting from industrial pollution. Both increases in the frequencies of distinct melanic forms and the general darkening of some or all forms within a population may be involved.

The increase in dark forms of some species of moth in industrial regions of western Europe, and latterly elsewhere, has provided some of the best known, most easily understood, and most often quoted examples of evolution in action. Increases in pollution following the industrial revolution led to changes in the environment. In particular, sulfur dioxide denuded trees and other substrates of lichens, while particulate air pollution blackened the resulting surfaces. In response to these changes, many species of moth and some other invertebrates that rely on camouflage for defense against some predators have changed their coloration, becoming darker, in line with the darkening of the substrates that they rest upon by day. These changes have occurred largely in the past 150 years and are cited as examples that illustrate the central mechanism of Charles Darwin's theory of evolution: natural selection.

TYPES OF INDUSTRIAL MELANISM

Three categories of industrial melanism have been recognized:

A. Full industrial melanic polymorphism involves distinct melanic forms that have arisen since the industrial revolution and have increased as a consequence of the effects of industrialization on the environment.

B. Partial industrial melanic polymorphism involves polymorphic species that had melanic forms prior to the industrial revolution. These forms have increased in frequency following and as a consequence of the effects of industrialization.

C. Polygenic industrial melanism involves species in which the average ground color of some or all members of a population has darkened gradually as a consequence of the effects of industrialization.

It should be noted that melanism is a common phenomenon throughout the animal kingdom, with many factors unrelated to industrialization or pollution influencing the success of melanic forms in some species.

FULL INDUSTRIAL MELANIC POLYMORPHISM

The Peppered Moth

The peppered moth, *Biston betularia,* has dominated the literature on industrial melanism. In Britain, the ancestral form of this species (form *typica*) is white, liberally speckled with dark brown or black scales (Fig. 1). In 1848, a predominantly black form of *B. betularia,* form *carbonaria* (Fig. 2) was recorded in Manchester, England. Within 50 years, 98% of Mancunian peppered moths were black. From this original location, *carbonaria* spread to many other parts of Britain. The renowned Victorian lepidopterist J. W. Tutt was the first to suggest that camouflage and bird predation could be involved in the spread of *carbonaria.* In 1896, arguing that the typical form was camouflaged well on surfaces covered by foliose lichens, he noted that the nature of many natural surfaces had changed as a consequence of pollutants resulting from heavy industry. In particular, the combined effects of sulfur dioxide, which killed foliose lichens, and soot fallout, which blackened the denuded surfaces, had led to darker and more uniform substrates. He stated that on these surfaces the *carbonaria* form would be better camouflaged than *typica* and so gain protection from bird predation. Natural selection, through the medium of differential bird predation, augmented by "hereditary tendency," had led to an increase in the frequency of the black form.

FIGURE 1 The typical form of the peppered moth, *B. betularia.*

FIGURE 2 The *carbonaria* form of the peppered moth.

FIGURE 4 A peppered moth in its natural resting position, beneath a lateral tree branch.

Tutt's hypothesis was largely rejected, both at the time and for a considerable period thereafter, because most entomologists and ornithologists concurred in the view that birds are not major predators of cryptic, day-resting moths. A variety of other explanations of the increase in melanic forms of some moths were thus put forward during the first half of the 20th century (pollutants acting as mutagenic agents, Lamarckian evolution, heterozygote advantage).

One important advance during this period was the calculation by Haldane in 1924 that *carbonaria* would have to have been one and a half times as fit as *typica* to account for the rapidity of the rise in *carbonaria* frequency in Manchester. This fitness difference was much higher than most evolutionary biologists of the time thought feasible.

Not until the 1950s was Tutt's bird predation explanation of the rise of *carbonaria* in polluted regions tested by scientific experimentation. Dr. Bernard Kettlewell, using direct observation of the predation of live moths released onto tree trunks, and mark–release–recapture techniques, in two populations, one in a polluted and the other in a nonpolluted oak woodland, obtained strong evidence to support Tutt's differential predation hypothesis. Both experiments showed that the *typica* form of the moth had lower fitness than *carbonaria* in the polluted woodland, but a higher fitness in the nonpolluted wood. It was the fact that Kettlewell obtained reciprocal results in the two environments that made his conclusions so convincing. Kettlewell also mapped the frequency of *carbonaria* against sulfur dioxide and soot fallout, finding a significant correlation between the frequency of *carbonaria* and both pollutants, that with sulfur dioxide being strongest. This correlation between high melanic frequencies and high levels of pollutants has been reinforced by the finding that *carbonaria* frequencies have declined following decreases in pollution levels as a result of anti-pollution and smoke control legislation.

FIGURE 3 The peppered moth has a third intermediate form, *insularia.*

The elements of the basic story of the peppered moth that are usually related are therefore:

1. The peppered moth has two distinct forms.
2. These forms are genetically controlled.
3. Peppered moths rest by day on tree trunks.
4. Birds find peppered moths on tree trunks and eat them.
5. The likelihood of a moth being found by a bird depends on its degree of camouflage.
6. Nonmelanic peppered moths are better camouflaged than melanics on lichen-covered tree trunks in rural areas. Melanic peppered moths are better camouflaged than nonmelanics in industrial areas where tree trunks have been denuded of lichens and blackened by soot fallout.
7. The frequencies of melanic and nonmelanic moths in a particular area depend on the level of bird predation of each form and the rate of migration of moths into the area from adjacent districts in which the form frequencies are different.

Since Kettlewell's research, other studies on the peppered moth, which have included work on the intermediate form, *insularia* (Fig. 3), and the natural resting behavior of the moth (Fig. 4), have refined some of the details of the case. However, Kettlewell's basic qualitative deductions remain valid.

TABLE I The Decline in the Frequency of the *carbonaria* Form of the Peppered Moth Since Anti-pollution Legislation at Two Sites in England

Year	Caldy Common, West Kirby, northwest England	West Cambridge, England
1960/1961	94.2	94.8
1965	90.2	—
1970	90.8	75.0
1975	86.6	64.7
1980/1981	76.9	45.9
1985	53.5	39.5
1990	33.1	22.2
1995	17.6	19.2
1998/2000	11.5	15.1

Note. Similar differences have been recorded in the United States (see text and Grant *et al.*, 1995, 1996).

Industrial Melanism in Reverse

In the 1950s, anti-pollution legislation was introduced in industrial countries on both sides of the Atlantic. This legislation led to declines in both sulfur dioxide and particulate soot emissions. Subsequently, the frequencies of the melanic forms have declined considerably in industrial regions in Britain (Table I). Current rates of decline are broadly in line with theoretical predictions using computer simulations. If the decline in *carbonaria* continues at its present rate, this form will be reduced in Britain to the status of a rare mutation by 2020. A similar decline in the frequency of the melanic form (f. *swettaria*) of the American subspecies of the peppered moth, *B. betularia cognataria,* has occurred in some parts of the United States.

Data on the declines of *carbonaria* in Britain and *swettaria* in America are important for three reasons. First, they show that evolution is not a one-way process. Evolutionary changes can be reversed if the selective factors that lead to them are reversed. Second, the data sets from different populations in Europe and America are, in effect, replicate natural experiments. The consistency in the patterns of increase and decrease in the frequencies of melanic forms correlated to pollution levels adds weight to the selective explanation of the evolutionary changes observed. Third, the accord between predicted decreases in melanic frequency and the observed frequency currently being obtained argues that the factors incorporated into the models are broadly correct.

Other Examples of Full Industrial Melanic Polymorphism

The case of the peppered moth is not unique. A small number of other examples of full industrial melanic polymorphism are known. The melanic forms in most of these cases are controlled by dominant alleles of single genes.

FIGURE 5 Nonmelanic and melanic forms of the brindled beauty, *L. hirtaria.*

An exception is that of the brindled beauty, *Lycia hirtaria,* in which the melanic form *nigra* (Fig. 5) is controlled by a recessive allele. That most recent melanic forms are genetically dominant is not surprising because a dominant mutation will be fully expressed as soon as it arises and will quickly be favored by selection if advantageous. Recessive mutations would not be exposed to selection until they occurred in homozygotes, in which their effects would be expressed phenotypically.

In some species showing full industrial melanic polymorphism, such as the lobster moth, *Stauropus fagi* (Fig. 6), melanism developed at roughly the same time as in the peppered moth. In others, industrial melanism has developed much more recently, as in the cases of the sprawler, *Brachionycha sphinx,* and the early grey, *Xylocampa areola,* in which industrial melanism developed only in the second half of the 20th century. The reason that industrial melanism did not evolve earlier in these species is probably serendipitous: a melanic mutation simply did not occur previously in an appropriate population.

The different timings of the initial occurrence of industrial melanics of different species help emphasize that natural selection cannot cause change unless phenotypic variation

FIGURE 6 Nonmelanic and melanic forms of the lobster moth, *S. fagi.*

exists. This is manifest in the oak beauty moth, *Biston strataria,* the closest British relative of the peppered moth. The oak beauty has a melanic form, *melanaria,* which is a common industrial melanic in Holland, but has never been recorded, except as a rare mutation, in Britain. In terms of its ecology, behavior, and distribution, the oak beauty is similar to the peppered moth. However, the *melanaria* mutation seems never to have arisen in Britain in favorable circumstances nor has this form reached Britain from continental Europe as a migrant. Melanism in the oak beauty in Europe can be contrasted with that of another moth, the figure of eighty, *Tethea ocularis.* The melanic form, *fusca,* of this species was known in Belgium and Holland in the early part of the 20th century, but was absent from Britain. This form arrived in southern England, by migration, in the mid-1940s. Following its arrival, f. *fusca* spread to many industrial parts of Britain and increased in frequency rapidly, although its frequency is now declining again in response to reductions in pollution.

The current declines in melanism seen in the peppered moth, the figure of eighty, and several other species, following anti-pollution legislation suggest that future studies of industrial melanism may have to shift to countries in which industrialization is still increasing and anti-pollution measures are as yet limited.

PARTIAL INDUSTRIAL MELANIC POLYMORPHISM

Melanic forms of many species of moth are independent of industrialization. The factors that can favor melanism are numerous and varied. These have been discussed in detail by Kettlewell and Majerus. Their relevance to industrial melanism is that in some moths, the presence of melanic forms prior to, and independent of, industrialization provided a repository of melanic variants that were favored as pollution levels increased.

Indeed, it is likely that the majority of moths that exhibit melanic polymorphism, with melanic frequency correlated to pollution levels, had melanic forms occurring at relatively low equilibrium frequencies prior to the industrial revolution. Changes in the environment resulting from increased pollution favored these dark forms and their frequencies increased. The willow beauty, *Peribatodes rhomboidaria,* illustrates the idea well. In Britain, this species has long had a nonindustrial melanic form, *perfumaria.* The *perfumaria* form greatly increased in frequency in industrial regions in the late 19th century. In the 20th century, *perfumaria,* which still occurs at low frequency in some rural areas, particularly in Scotland, was replaced in industrial areas, but not elsewhere, by an even darker form, f. *rebeli.* Here then, f. *perfumaria* should be regarded as a partial industrial melanic, while f. *rebeli* is a full industrial melanic.

Many probable instances of partial industrial melanic polymorphism could be cited, but rather few of the species that fall into this class of melanism have been investigated in any depth. Exceptions are the pale brindled beauty, *Phigalia pilosaria;* the mottled beauty, *Alcis repandata;* and the green brindled crescent, *Allophyes oxyacanthae.* All are trunk-resters, the increase in melanics in industrial regions being attributed to increased crypsis.

Some of these species show morph-specific habitat preferences. Morph-specific habitat preferences in Lepidoptera showing melanic polymorphism were first suggested to explain abrupt differences in melanic frequencies of the mottled beauty and the tawny-barred angle, *Semiothisa liturata,* either side of sharp habitat boundaries. Such differences have subsequently been recorded in 14 species, in all cases melanics having higher frequencies in woodland with dense canopies than in adjacent more open habitats.

Many species that now have industrial melanics first evolved melanism in specific ecological circumstances prior to industrialization. It is known, for example, that a number of species now show melanic polymorphism in unpolluted ancient coniferous forests, such as Rannoch Black Wood in Scotland. Similar habitats were more widespread in the past and are likely to have supported melanic forms. These melanic forms would have been at a selective disadvantage if they moved from areas with the specific ecological circumstances to which they were adapted. Consequently, the melanics evolved behaviors that restricted them to such habitats. Recent changes in forestry and land usage and increases in pollution have provided new habitats (e.g., conifer plantations, polluted woodlands) with ecological conditions that favor melanics. The melanics have consequently spread and risen in frequency, producing examples of partial industrial melanic polymorphism in which morph-specific habitat preferences are retained to some extent.

POLYGENIC INDUSTRIAL MELANISM

Of all categories of melanism, polygenic industrial melanism has been the least considered and is the most difficult to address. Examination of specimens collected over the past century and a half suggests that many species have experienced a gradual darkening of the colors and loss of patterning in industrial regions, irrespective of morph. Although some of this change may be attributed to the gradual fading that occurs in museum specimens with time, it is difficult to ascribe all of the differences to this phenomenon. Comparison of series of specimens of six species, from rural and industrial regions, collected between 1880 and 1914 with those collected between 1992 and 1996 showed that the ground color had darkened more in industrial regions than in the rural areas.

This gradual darkening is probably the result of selection acting on polygenic variation. Small variations in the color patterns of many species are known to be controlled by many genes, each having a small effect. The selective predation of lighter and thus less cryptic forms in regions affected by particulate air pollution will result in those alleles which produce darker morphs increasing in frequency. It is difficult to see

how this hypothesis can be tested. However, if it is correct, the recent decrease in pollution should lead to a reversal of this trend, with ground colors lightening and patterns becoming more clearly defined again. Novel, digital methods of measuring the spectral reflectance of surfaces and storing data should allow measurement without reliance on museum specimens or photographs, both of which may fade with time.

MELANISM AND THE STUDY OF EVOLUTION

The significance of industrial melanism in the Lepidoptera to evolutionary biology has been considerable. It has provided one of the best observed examples of evolutionary change caused by natural selection and has shown that Darwinian selection can be a strong force. In the peppered moth, differential bird predation, together with migration, has been primarily responsible for the rise and fall of the melanic form *carbonaria.*

Although the story of the peppered moth is undoubtedly more complex than usually related, data accumulated during the past 40 years have done nothing to undermine Tutt's initial hypothesis of the role of differential bird predation or Kettlewell's experimental demonstrations of this role.

Within the Lepidoptera, the factors responsible for melanism and the forms of melanism that result are very variable. Because a great variety of factors may promote melanism, it may be misleading to extrapolate from one population to another, let alone from one species to another. Even within one class of melanism, the relative influence of different aspects of a species' biology will vary between species. Each species that has evolved melanic forms will have done so in the presence of a variety of different intrinsic and extrinsic circumstances. The differences in the factors affecting melanism in even the few well-studied cases suggest that there is still enormous scope for original research into this phenomenon. However, in species in which melanism is strongly correlated with pollution levels, such as the peppered moth, we are rapidly running out of time to pursue research into this phenomenon as melanics decline.

See Also the Following Articles

Coloration • Crypsis • Genetic Variation • Lepidoptera • Pollution • Thermoregulation

Further Reading

Cook, L. M. (2000). Changing view on melanic moths. *Biol. J. Linn. Soc.* **69,** 431–441.

Grant, B. S., Owen, D. F., and Clarke, C. A. (1995). Decline of melanic moths. *Nature* **373,** 565.

Grant, B. S., Owen, D. F., and Clarke, C. A. (1996). Parallel rise and fall of melanic peppered moths in America and Britain. *J. Hered.* **87,** 351–357.

Haldane, J. B. S. (1924). A mathematical theory of natural and artificial selection. *Trans. Cambridge Philos. Soc.* **23,** 19–41.

Howlett, R. J., and Majerus, M. E. N. (1987). The understanding of industrial melanism in the peppered moth (*Biston betularia*) (Lepidoptera: Geometridae). *Biol. J. Linn. Soc.* **30,** 31–44.

Kettlewell, H. B. D. (1955). Selection experiments on industrial melanism in the Lepidoptera. *Heredity* **9,** 323–342.

Kettlewell, H. B. D. (1956). Further selection experiments on industrial melanism in the Lepidoptera. *Heredity* **10,** 287–301.

Kettlewell, H. B. D. (1973). "The Evolution of Melanism." Clarendon Press, Oxford.

Majerus, M. E. N. (1998). "Melanism: Evolution in Action." Oxford University Press, Oxford.

Majerus, M. E. N., Brunton, C. F. A., and Stalker, J. (2000). A bird's eye view of the peppered moth. *J. Evol. Biol.* **13,** 155–159.

Owen, D. F. (1961). Industrial melanism in North American moths. *Am. Nat.* **95,** 227–233.

Tutt, J. W. (1896). "British Moths." George Routledge, London.

Insecta, Overview

Vincent H. Resh
University of California, Berkeley

Ring T. Cardé
University of California, Riverside

The species-rich superclass Hexapoda includes all insects and their near relatives that share the characteristic arrangement of having, as adults, three major body regions and six legs. The number of described insect species has increased greatly from the time of the early catalogers of life. Those 18th-century pioneers in biodiversity, such as Carl Linné, would not have conceived that there would be nearly a million species described by the 21st century. Most estimates today suggest that this number represents only 10 to 30% of the actual number of insect species thought to exist. The richness of living things is essentially the result of insect richness; animal biodiversity is therefore, in reality, mainly insect biodiversity.

Within the class Insecta, major forms of insects are grouped in orders. Ordinal-level groups represent divergent lineages that are nearly always recognizable by a set of distinctive characteristics. Almost always, an adult insect can be readily determined to order at a glance. The number of recognized orders has fluctuated slightly as entomologists' understanding of the included taxa and methods for classifying have developed. Classification schemes are both organizational systems and true scientific hypotheses. In this way they are dynamic, changing as new information becomes available. There are several important ways in which a classification may evolve. One is the subjective change in taxonomic rank. For example, in the 1950s all Ephemeroptera (the order containing mayflies) in North America were assigned to three families, and today they are in 21. Mostly, this is the result of raising subfamilies to family status.

If substantial evidence is found that a group previously recognized as an order is paraphyletic (i.e., does not contain all descendants of that group), then new monophyletic arrangements will be proposed. A good example is the order

Hemiptera. The taxa included in the order now were traditionally divided into two groups (often given ordinal status): Heteroptera (true bugs) and Homoptera. Recent analyses suggest a more complicated pattern of relationships. Three groups within the order Hemiptera are treated separately as suborders in this encyclopedia: Auchenorrhyncha (cicadas, spittlebugs, leafhoppers, treehoppers); Prosorrhyncha (Heteroptera and Coleorrhyncha); and Sternorrhyncha (aphids, psyllids, scale insects, whiteflies).

Perhaps the most exciting way that classifications may change is by the discovery of something genuinely novel. However, in insects, finding a truly new order (i.e., a group of taxa that have a combination of characteristics unique at that level) is an astounding event. At the time that this encyclopedia was ready for printing, a new order (Mantophasmatodea) was discovered in southern Africa. It is the first order of living insects to be described in over 80 years! This discovery reinforces the point that there is much left to learn about our earth's biodiversity.

In many classification systems, orders are grouped into superorders, but what comprises a superorder is far from fixed. For example, zorapterans are viewed as being in the superorder Orthopteroidea by some and in the superorder Hemipteroidea by others. Although a close link between Trichoptera (caddisflies) and Lepidoptera (moths and butterflies) is supported by a wealth of concordant evidence

TABLE I The Orders of Insects and Other Members of the Arthropod Superclass Hexapoda

Hierarchical category	Taxon	Families	Species
Class	Parainsecta		
Order	Protura	8	600
	Collembola (springtails)	27	9,000
Class	Entognatha		
Order	Diplura	8	1,000
Class	Insecta		
Subclass	Apterygota		
Order	Archaeognatha (bristletails)	2	500
	Zygentoma (thysanurans, silverfish)	5	400
Subclass	Pterygota		
Infraclass	Paleoptera		
Order	Ephemeroptera (mayflies)	37	3,000
	Odonata (dragonflies, damselflies)	31	5,500
Infraclass	Neoptera		
Division	Endopterygota		
Order	Blattodea (cockroaches)	5	4,000
	Mantodea (mantids)	8	1,800
	Isoptera (termites, white ants)	7	2,500
	Grylloblattodea (rock crawlers)	1	25
	Dermaptera (earwigs)	7	2,000
	Plecoptera (stoneflies)	16	2,000
	Embiidina (webspinners)	8	300
	Orthoptera (grasshoppers, katydids)	23	20,000
	Phasmida (walkingsticks)	2	3,000
	Mantophasmatodea	?	?
	Zoraptera	1	32
Superorder	Hemipteroidea		
Order	Psocoptera (booklice, barklice)	17	4,400
	Phthiraptera (biting lice, sucking lice)	24	4,900
	Hemiptera (true bugs)	104	55,000
	Thysanoptera (thrips)	9	5,000
Division	Endopterygota		
Order	Megaloptera (alderflies, dobsonflies)	2	300
	Raphidioptera (snakeflies)	2	260
	Neuroptera (lacewings, ant lions)	17	6,000
	Coleoptera (beetles)	135	350,000
	Strepsiptera	8	550
	Mecoptera (scorpion flies)	7	550
	Diptera (flies)	117	125,000
	Siphonaptera (fleas)	15	2,600
	Lepidoptera (moths, butterflies)	120	160,000
	Trichoptera (caddisflies)	45	11,000
	Hymenoptera (ants, bees, wasps)	73	150,000

and consequently is undisputed, relationships among many endopterygote (e.g., Coleoptera and Hymenoptera) orders are unclear because different data sets present conflicting evidence. Some orders [e.g., Collembola (springtails) and Protura] are considered to be noninsects (placed in the class Parainsecta), but evidence clearly places them in the superclass Hexapoda with insects.

The classification presented here is a snapshot of the current hypotheses of insect relationships. Because the field of systematics that underlies this classification scheme is ever evolving, future arrangements will undoubtedly shift. The extant (or existing) orders of insects, their common names, and estimates of their worldwide species and family richness are listed in Table I.

See Also the Following Articles
Arthropoda and Related Groups • Phylogeny of Insects

Insecticides

Fumio Matsumura
University of California, Davis

Insecticides is the term coined to describe chemicals used to control pest insects and related invertebrate pest species. Insects are by far the most important species against which these chemicals are targeted. Other major groups of pest organisms include mites, ticks, and nematodes. Acaricides (for the control of mites and ticks) and nematocides (for the control of nematodes) are chemicals specifically used to control these pests, but they are still considered subgroups of the broadly defined "insecticides" group.

Not all insecticides are designed to kill pest insects, despite the use of the suffix "-cides" which gives the connotation of biocidal agents. Insecticides have been defined to include any chemical that can be used to reduce damage caused by insects. Thus, nonlethal chemicals such as pheromones, repellents, hormone mimics, growth regulators, feeding inhibitors, anorectic agents (which cause loss of appetite), behavioral disrupters, food attractants (used in traps and as bait), and anesthetics, as well as those causing physical problems such as surfactants, sticky substances, desiccants, and barriers (such as oil film on the surface of water for mosquito larval control) are considered to be insecticides.

BRIEF HISTORY

Insecticides used prior to the 1940s were mostly inorganic compounds such as arsenicals. After World War II, DDT and other chlorinated pesticides came on the market. There is no question about the spectacular insect-controlling effects achieved on many crops, and populations of some pests that affect both public and veterinary health were greatly diminished. The shortcomings of these compounds, particularly their lack of selectivity and harmful environmental effects, were eventually realized, however, leading to the termination of their use by the late 1970s. Meanwhile, organophosphorus and carbamate insecticides gained in popularity and have established themselves as two of the major classes of insecticides. Many of them offer at least some degree of selectivity (malathion is particularly outstanding in this regard) and are less persistent in the environment. In more recent years, functional synthetic analogues of naturally occurring toxic chemicals were developed. Pyrethroids, for example, are essentially synthetic mimics of naturally occurring pyrethrins found in the flowers of species of chrysanthemum. The synthetic neonicotinoids mimic naturally occurring nicotine from tobacco plants. Useful microbial products were also developed in the 1980s and 1990s; examples are *Bacillus thuringiensis* (Bt) toxins, avermectins, and spinosyns. Modern insecticides used today are generally very selective, mostly affecting only the targeted pest insect. They are potent, requiring only small quantities to achieve their effects, and they are much less persistent in the environment.

CLASSIFICATION OF INSECTICIDES

Synthetic organic insecticides may be divided into several major classes: (1) chlorinated hydrocarbons, (2) organophosphorus compounds (often referred to as organophosphates), (3) carbamates, (4) pyrethroids, (5) nicotinoids, (6) fumigants, (7) GABA receptor antagonists, (8) chitin synthesis inhibitors (benzoylureas), (9) mitochondrial poisons, and (10) insect hormone mimics. These classifications are based on either group-specific chemical characteristics (classes 1–6) or their action mechanisms (classes 7–10).

Other insecticides belonging to minor classes (i.e., fewer compounds per class or less frequent use) are (11) botanically derived naturally occurring insecticides (other than pyrethroids and nicotinoids), (12) microbially produced insecticides, (13) synergists, (14) semiochemicals such as attractants, including pheromones, (15) insect repellents or feeding deterrents, and (16) behavior-modifying agents for use on insects.

USE PATTERNS

Insecticides as a class of pesticides constitute about one-quarter of total pesticides (approximately a billion pounds per year) used in the United States. By far the largest volume of pesticides used is herbicides (620 million pounds), followed by insecticides (247 million pounds) and fungicides (131 million pounds) (all 1993). Approximately 75% of all pesticides used in 1993 were for the control of agricultural pests. Other uses are for pests found in the home (including gardens), industry, commerce, and public and veterinary

health. The top 17 insecticides (used in 1993) were (1) chlorpyrifos, (2) terbufos, (3) methyl parathion, (4) carbofuran, (5) carbaryl, (6) phorate, (7) cryolite, (8) aldicarb, (9) propargite, (10) acephate, (11) malathion, (12) fenofos, (13) methomyl, (14) dimethoate, (15) azinphos-methyl, (16) ethyl parathion, and (17) profenfos. Most of these are organophosphates (1–3, 6, 10–12, 14–17) or carbamates (4, 5, 8, 13), but propargite is a sulfite ascaricide and cryolite (sodium fluoroaluminate, Na_3AlF_6) is a naturally occurring inorganic fluoride compound. Of these, the use of methyl parathion (3) and ethyl parathion (16) has been phased out. Among organochlorine insecticides, most of which have been eliminated, the only ones remaining are endosulfan (19th) and dicofol (22nd). The most popular pyrethroid is permethrin (25th, approximately 1,000,000 pounds) followed by cypermethrin (225,000 pounds) and fenvalerate (66,000 pounds). Pyrethroids are used in much lower quantities than organophosphates and carbamates mainly because the former compounds are much more powerful than the latter, and therefore only small amounts of pyrethroids per hectare are needed to control insect pests.

MECHANISMS OF ACTION OF INSECTICIDES

The great majority of insecticides used today are nerve poisons. This is because insects have highly developed nervous systems and, furthermore, many of their sensory receptors are exposed to the atmosphere outside the insect body. The insect nervous system relies on several key functions that have been exploited as the targets of insecticides: the sodium channel, acetylcholinesterase, the γ-aminobutyric acid (GABA) receptor, and the acetylcholine receptor.

The sodium channel, which is the insecticidal target of DDT, pyrethroids, pyrethrins, and other minor classes of insecticides, lines the outer surface of the neurons and functions as the voltage-dependent sodium ion pore (i.e., the pore opens or closes depending on the change in voltage). Upon the arrival of stimuli, this pore allows the selective entry of sodium ions into the neuron for a brief moment and then abruptly shuts down the flow (this phenomenon is called "inactivation"). Thereafter, the sodium channel goes through an internal rearrangement to recover its original state. Such an action causes a brief local equalization of sodium ions between the outside and the inside of the neuron (depolarization), and this change is sensed as a local signal for excitation by the affected neuron. These insecticides delay the shutdown process and furthermore delay the recovery process, resulting in a prolongation of the period of excitation. Insects thus affected continue in a state of hyperexcitation, leading to exhaustion and, at high doses of the insecticide, death.

The next important insecticidal target is acetylcholinesterase, which is attacked by organophosphorus and carbamate insecticides. This enzyme, by inducing hydrolysis, inactivates the interneuron nerve transmitter acetylcholine. This excitatory transmitter is released upon the arrival of a signal from the distal end of one neuron, travels across the intercellular gap, arrives at the frontal end of the second neuron, and reacts with its specific acetylcholine receptor on the surface that sends the signal of excitation to the second neuron. It is important to stress here that such a successful signal transmission must be followed with an abrupt termination of the action of the transmitter; this allows for the second neuron to recover quickly enough and thereby stay ready for the next message, maintaining the normal function of the message-transmitting neuron. This termination action is mainly carried out by acetylcholinesterase, which eliminates acetylcholine from the vicinity of the acetylcholine receptor of the second neuron. All organophosphorus and carbamate insecticides, or their active metabolites, show potent inhibitory actions on acetylcholinesterase of insects as well as other animals. The insects affected by these chemicals show overt signs of excitation, exhaustion and, at sufficient doses, death.

The acetylcholine receptor also can be deactivated to cause the same type of hyperexcitation. Indeed, nicotinoids (which include naturally occurring nicotine analogues and their modern derivatives, sometimes called "neonicotinoids"), such as imidacloprid, are known to directly activate the acetylcholine receptor, just like acetylcholine. Nicotine's excitatory action is well known. Neonicotinoid derivatives readily penetrate the insect's body and nerve sheath, arriving at critical sites of neurons, and persisting there long enough to exert a powerful excitant effect.

The GABA receptor, in contrast, acts as the receiver for the inhibitory transmitter, GABA. That is, unlike acetylcholine, it is not an excitatory transmitter. The signal generated by this GABA–GABA receptor interaction is converted to the opening of chloride channels, which upon the arrival of the signal permit Cl^- ions to come into the signal receiving cells (either neurons or muscle cells), to make them nonresponsive to excitation stimuli. Those insecticides—chlorinated hydrocarbon insecticides, cyclodienes (such as γ-HCH, dieldrin, endosulfan, toxaphene), and more modern insecticides (such as fipronil)—render the chloride channel inoperative so that chloride ions cannot come into the cells. Cells thus affected fail to receive the inhibitory signal of GABA and therefore cannot counterbalance any excitatory forces. One group of insecticides, avermectin analogues, keep the chloride channel stuck in the open position, an action opposite from that of the excitation-inducing insecticides. These compounds induce long-lasting inhibition of excitation in insects. Insects thus affected by avermectin analogues show diminished activities, nonresponse to stimuli, and slow death through paralysis.

Certainly there are other mechanisms by which normal functions of insects may be affected. The main ones are as follows:

1. Mitochondrial poisons, such as rotenone, which causes respiratory failure.

2. Inhibitors of cuticle formation, via the action of dimilin, including the rest of the diflubenzyron derivatives, which cause difficulty with molting and maintaining protective shields.

3. Insect hormone mimics such as juvenile hormone analogues that keep affected insects as immature forms (this method is effective against insects that cause damage only as adults, e.g., mosquitoes). Another group is ecdysone analogues, which affect insect development, including molting.

4. *B. thuringiensis* toxins, which mainly affect the potassium channel in insect digestive systems.

5. Formamidine analogues, such as chlordimeform, which mimics octopamine, a naturally occurring transmitter/hormone, by acting on its receptor. Octopamine is used by insects and mites to control their behavior (among many of its actions), and therefore chlordimeform analogues are known to modify many behavioral patterns of insects and mites, and thereby protect crops from those pests.

INSECTICIDE RESISTANCE

In 1958 A. W. A. Brown's landmark publication, *Insecticide Resistance in Arthropods,* established the principle that insects as well as other related invertebrates are capable of developing resistance to insecticides through natural selection. The probability of the development of resistance largely depends on (1) the frequency of the resistance-conferring gene in the given population, (2) the level of selection pressure, (3) the degree to which resistant gene density is diluted by susceptible genes through influx of individuals from untreated areas, and (4) the stability of the resistance gene in the given population. In some cases, once established, resistance genes may persist in the same locality for many years. A good example may be the pyrethroid resistance of the moth *Helicoverpa armigera,* in Australia.

How insects develop resistance to insecticides is a topic that has fascinated many entomologists. Basically, there are two major ways through which insect pests acquire resistance: increased detoxification capabilities and alteration of the insecticide target sites (target sensitivity). The first type of resistance occurs more frequently than the second type, as well as all others. Detoxification of toxic insecticidal chemicals is carried out by specialized enzymes designed to handle all chemicals toxic to insects, not just insecticides. Insects, particularly those feeding on plants that produce naturally large amounts of toxic chemicals, have well-developed detoxification enzymes. There are three major types of detoxification enzymes: (1) broad-spectrum oxidases such as mixed function oxidases catalyzed by cytochrome P450, (2) hydrolases that break up esters, ethers, and epoxides, and (3) conjugation systems such as glutathione *S*-transferase, which are mediated to cover up the reactive part of the toxic chemical and further facilitate its removal. Every type of detoxification enzyme has been documented to play a role in the development of some form of resistance against various classes of insecticides.

In determining which type of detoxification enzymes will become the key player in the development of resistance, the most important factor for consideration is the chemical properties of the insecticide. For instance, carbamates and pyrethrins are readily detoxified by mixed function oxidases; therefore, if resistance is reported against these insecticides, one must first look for increased activities of mixed function oxidases in the resistant insects. If higher activity levels are found, the resistance spectrum (i.e., cross-resistance of carbamate-resistant insects to other types of insecticide) is usually wide because mixed function oxidases are capable of detoxifying chemicals of many different types. In contrast, organophosphorus and pyrethroid insecticides are mainly degraded by hydrolases. Thus, the involvement of an increased hydrolytic enzyme activity may be suspected when insects develop resistance against these chemicals.

A good example of this is malathion resistance. Malathion molecules contain two extra carboxylic acid ethyl ester parts. Malathion-resistant insects always show increased carboxylesterase activity. Esterases of these types are not broad-spectrum enzymes, and therefore malathion resistance is usually specific (i.e., usually the insects resistant to malathion are not resistant to other insecticides). Insecticides with labile halogens, epoxides, methoxy unsaturation, and some aliphatic unsaturation may be degraded through these glutathione-mediated detoxification systems, and hence their elevated presence could be suspected to cause resistance. This scheme is, however, merely a rough guess about the possible mechanism of development of metabolic resistance. Indeed, unexpected and unique resistance mechanisms have been reported to occur in some combinations of insecticides and insects (e.g., DDT resistance in *Drosophila*). The recommended method of identification of the metabolic cause is to co-treat insects with the insecticide and specific inhibitors for each type of metabolic detoxification system, such as piperonyl butoxide for mixed function oxidase and DEF for esterases.

In studies of mechanisms for target insensitivity resistance, mutations occurring in the sodium channel, the GABA receptor, and acetylcholinesterase have been found in insects resistant to DDT/pyrethroids, cyclodiene insecticides, and organophosphorus and carbamate insecticides, respectively. Those resistances are characterized by their specificity (low degrees of cross-resistance) and the general stability of resistance among insect populations in given localities.

REGULATIONS OF INSECTICIDE USES

Insecticides, like all other types of pesticide, are highly regulated by governments in all countries. In the United States, the main law governing the use of insecticides is the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), which mandates registration with the U.S. Environmental Protection Agency of all insecticides used in the country. The initial data requirements for successful registration (so-called Tier 1) depend on the extensiveness of

the projected use, the levels of acute toxicity of the agent contemplated, its effectiveness as an insect control agent (called "efficacy"), the intended modes of usage, and the availability of background knowledge, among other requirements. Occasionally, experimental use permits are given after this Tier 1 examination/process (e.g., for insect pheromones, which are already known to be almost nontoxic and are to be used only for a specific pest in small areas). Usually, however, registrants are required to go through a much more extensive and rigorous process of registration, data procurement, and evaluation. For example, extensive tests are required for acute, chronic (such as carcinogenicity tests), genetic, pathogenic, reproductive, hormonal, and immune toxicities along with the environmental behavior of chemicals and limited wildlife toxicities. Such registration processes, which must be completed before a new chemical pesticide can be sold in the United States, typically require 7 to 10 years and roughly $100 million.

Despite the thoroughness of the registration processes, occasionally problems come to the attention of the scientific community or the regulatory agencies. Sometimes, for example, old pesticides are registered despite the availability of extensive records of their actual use. This is partly the result of the relative ease of the registration process in the past and partly from the absence of the main registrant, who is not economically motivated to reregister the compound because the patents for those chemicals (and thereby the exclusive marketing right) have expired. The second type of problem is due to the failure of the regulatory agency/scientific community to address the special vulnerability of certain groups of human populations or ecosystems. Examples include the lack of toxicological data on infants and embryos, women, and the aged, and science's incomplete understanding of the hormonal effects of pesticides on humans and wildlife. The third type of problem is caused mostly by unforeseen scientific or technological developments, or unfortunate circumstances that are difficult to predict. The question of the safety of genetically modified crops and the assessment of strategies to study the recently discovered skin-hypersensitizing action of some pesticides serve as examples.

A recent trend is to look at this issue from the consumer's side. A good example is the enactment of the Food Quality Protection Act (FQPA), which addresses the presence of pesticide residues and other toxic chemicals in food and drinking water. A key part of this regulation is the consideration of children's health. Here, an extra safety factor of 10× is demanded to accommodate the postulated extra vulnerability of embryos, infants, and developing children. This requirement is enforced unless registrants can provide actual safety data to demonstrate that the susceptibility of these groups to the hazardous effects of the compound is equal to or less than that of adults.

In the end, the toxicological methods of evaluation, including overall risk assessment approaches, address the majority of health concerns. Future improvements are needed, however, to deal with unresolved environmental and human health risks.

See Also the Following Articles

DDT • Integrated Pest Management • Regulatory Entomology

Further Reading

Brown, A. W. A. (1951). "Insect Control by Chemicals," pp. 117–207. Wiley, New York.

Crouse, G. D., and Sparks, T. C. (1998). Naturally derived materials as products and leads for insect control: The spinosyns. *Rev. Toxicol.* **2,** 133–146.

Elliott, M., Farnham, A. W., Janes, N. F., Needham, P. H., and Pulman, D. A. (1974). Synthetic insecticide with a new order of activity. *Nature* **248,** 710–720.

Forrester, N., Cahill, M., Bird, L. J., and Layland, J. K. (1993). Management of pyrethroid and endosulfan resistance in *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Australia. *Bull. Entomol. Res. Suppl. Ser.* **O(1),** I–VIII, 1–132.

Kasuda, Y. (1999). Development of and future prospects for pyrethroid chemistry. *Pestic. Sci.* **55,** 775–782.

Marrone, P. G. (1999). Microbial pesticides and natural products as alternative. *Outlook Agric.* **28(3),** 149–154.

Plunkett, L. M. (1999). Do current FIFRA testing guidelines protect infants and children? Lead as a case study. *Regul. Toxicol. Pharmacol.* **29(1),** 80–87.

Yamamoto, I., Yabuta, G., Tomizawa, M., Saito, T., Miyamoto, T., and Kagabu, S. (1995). Molecular mechanism for selectivity toxicity of nicotinoids and neonicotinoids. *J. Pestic. Sci.* **20,** 33–40.

Insecticide and Acaricide Resistance

Gregor J. Devine and Ian Denholm

Rothamsted Research, U.K.

Insecticide resistance is an example of a dynamic evolutionary process in which chance mutations conferring protection against insecticides are selected for in treated populations. This article reviews the origins and mechanisms of resistance, the factors that influence its severity, and the current options for combating its detrimental impact on agricultural productivity and human health.

INTRODUCTION

The genetic variation inherent in all populations is the consequence of random mutations within individuals, their recombination through meiosis, and the dispersal of genes between populations (gene flow). This variation is then shaped by the chance events of genetic drift and by the deterministic process of natural selection. The latter phenomenon eliminates alleles (gene variants) that reduce the fitness of an organism and preserve those that are neutral or that increase fitness. In eukaryotes, the phenotypic changes (adaptations)

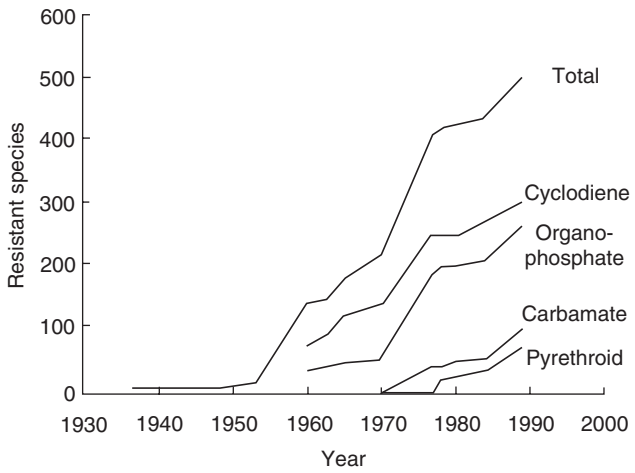

FIGURE 1 Increase in the number of arthropod species reported to resist insecticides over time, in total, and in response to the four most widely used classes of insecticide [Adapted with permission from Georghiou, G. P. (1990). Overview of insecticide resistance. *In* "Managing Resistance to Agrochemicals" (M. D. Green, H. M. Le Baron, and W. K. Moberg, eds.), pp. 18–14. ACS Symposium Series 421. Copyright (1990) American Chemical Society, Washington, DC.]

that result from this process are seldom visible over a human lifetime. The development of pesticide resistance by arthropods, however, is a spectacular exception to the rule.

Since the 1940s, synthetic insecticides have been used on an increasing scale to control the insects and mites that cause immense crop losses and pose major threats to public and animal health. However, because many of the target species have evolved resistance, some of these chemical control programs are failing. At the current time, more than 500 arthropod species have evolved resistance to at least one pesticide, and a few populations of some of those species are now resistant to all, or almost all, of the available products (Fig. 1).

The evolution of insecticide resistance has undoubtedly contributed to overall increases in the application of chemicals to crops. About 500,000 metric tons of insecticide is now applied each year in the United States alone, with obvious implications for both human health and the environment. Yet resistant insects continue to affect our agricultural productivity and our ability to combat vectors of disease. As a result, insecticide resistance imposes a huge economic burden on much of the world (in the United States alone, annual losses in crop and forest productivity have been estimated at $1.4 billion). Moreover, it is proving impossible to combat resistance by embarking on a chemical arms race. The development of a new insecticide takes 8 to 10 years at a cost of $20 to 40 million, and the rate of discovery of new insecticidal molecules, unaffected by current resistance mechanisms, seems to be on the wane.

Within just a few years of the registration of some of these new molecules, resistant insect populations have evolved.

DIAGNOSIS OF RESISTANCE

Although a large number of laboratory bioassay methods have been developed for detecting and characterizing resistance, most of these are limited to defining phenotypes and provide little or no information on the underlying genes or mechanisms. Thus, although bioassays remain the indispensable mainstay of most large-scale resistance monitoring programs, much attention is being paid to developing more incisive techniques that not only offer greater precision and turnover rates, but also diagnose the type of mechanism(s) present and, whenever possible, the genotypes of resistant insects. A variety of approaches are being adopted for this purpose, including electrophoretic or immunological detection of resistance-causing enzymes, kinetic and end-point assays for quantifying the activity of enzymes or their inhibition by insecticides, and DNA-based diagnostics for mutant resistance alleles.

The sensitivity of these techniques is exemplified by work on the green peach aphid, *Myzus persicae.* In northern Europe, this insect possesses three coexisting resistance mechanisms: an overproduced carboxylesterase conferring resistance to organophosphates, an altered acetylcholinesterase conferring resistance to certain carbamates, and target-site resistance (i.e., knockdown resistance, kdr) to pyrethroids. These mechanisms collectively confer strong resistance in this species to virtually all available aphicides. Fortunately, it is now possible to diagnose all three mechanisms in individual aphids by using an immunoassay for the overproduced esterase, a kinetic microplate assay for the mutant AChE, and a molecular diagnostic for the *kdr* allele. The combined use of these techniques against field populations provides up-to-date information on the incidence of the mechanisms and serves to inform growers of potential control problems and in the development of optimal strategies for the management of *M. persicae.*

EXTENT OF RESISTANCE

In some insects, resistance extends only to a few closely related compounds in a single chemical class. It may be very weak or restricted to a small part of the insects' geographical range. At the other extreme, some widespread pests, such as anopheline mosquitoes (e.g., *Anopheles gambiae*), the diamondback moth *(Plutella xylostella),* the Colorado potato beetle *(Leptinotarsa decemlineata),* and the sweet potato whitefly *(Bemisia tabaci)* now resist most or all of the insecticides available for their control. The most extensively used insecticide classes—organochlorines, organophosphates, carbamates, and pyrethroids—have generally been the most seriously compromised by resistance, and many principles relating to the origin and evolution of resistance can be demonstrated solely by reference to these fast-acting neurotoxins. In recent years, however, there has also been a worrying increase in resistance to more novel insecticides. These include compounds attacking the developmental pathways of arthropods (e.g., benzoylphenylureas), their respiratory

processes [e.g., mitochondrial electron transport inhibiting (METI) acaricides], their digestive systems [e.g., *Bacillus thuringiensis* (Bt) endotoxins], and pathways associated with the regulation of their nervous processes (e.g., neonicotinoids).

ORIGINS AND BREADTH OF RESISTANCE

Insecticides are not considered to be mutagenic at their field application rates and are, therefore, not the causative agents of insecticide resistance. Rather they act to select favorable mutations inherent in the population to which they are applied. Some attempts to estimate the rates at which resistant mutations occur have been made. The treatment of blow flies *(Lucilia cuprina)* with a chemical mutagen resulted in the production of dieldrin-resistant target-site mutations in less than one per million individuals. Other studies, however, have found the incidence of resistant mutations to be worryingly high. A recessive allele conferring resistance to Bt toxins in unselected populations of the tobacco budworm, *Heliothis virescens,* was estimated to be present in about one in every thousand individuals in some areas of North America. Sixteen in every hundred insects were found to carry a Bt-resistant allele in unselected populations of the pink bollworm, *Pectinophora gossypiella,* in Arizonan cotton fields. Despite this, Bt cotton remains effective in the control of these species, suggesting that such estimates need to be interpreted carefully. Less empirical measures of mutation rates are extremely variable (10^{-3} to 10^{-16}), but they will undoubtedly be dependent on the resistance mechanism involved.

Resistant mutations seldom confer protection to just a single toxin. Most commonly, they exhibit differing levels of resistance to a range of related and unrelated insecticides. In its strictest sense, the term *cross-resistance* refers to the ability of a single mechanism to confer resistance to several insecticides simultaneously. A more complex situation is that of *multiple resistance,* reflecting the coexistence of two or more resistance mechanisms, each with its own specific cross-resistance characteristics. Disentangling cross-resistance from multiple resistance, even at the phenotypic level, is one of the most challenging aspects of resistance research.

Cross-resistance patterns are inherently difficult to predict in advance, because mechanisms based on both increased detoxification and altered target sites can differ substantially in their specificity. The most commonly encountered patterns of cross-resistance tend to be limited to compounds in the same chemical class (equivalent to the term "side-resistance" as used by parasitologists). However, even these patterns can be very idiosyncratic. For example, organophosphate resistance based on increased detoxification or target-site alteration can be broad ranging across this group or highly specific to a few chemicals with particular structural similarities. The breadth of target-site resistance to pyrethroids in houseflies is also dependent on the resistance allele present. The *kdr* allele itself affects almost all compounds in this class to a similar extent (~ 10-fold resistance), whereas resistance due to the more potent *super-kdr* allele is highly dependent on the alcohol moiety of pyrethroid molecules, ranging from about 10-fold to virtual immunity. Cross-resistance between insecticide classes is even harder to anticipate, especially for broad-spectrum detoxification systems whose specificity depends not on insecticides having the same mode of action, but on the occurrence of common structural features that bind with detoxifying enzymes.

Empirical approaches for distinguishing between cross-resistance and multiple resistance include repeated back-crossing of resistant populations to fully susceptible ones, to establish whether resistance to two chemicals cosegregates consistently, and reciprocal selection experiments, whereby populations selected for resistance to one chemical are examined for a correlated change in response to another. If available, biochemical or molecular diagnostics for specific resistance genes can assist considerably with tracking the outcome of genetic crosses or with assigning cross-resistance patterns to particular mechanisms.

MECHANISMS OF RESISTANCE AND THEIR HOMOLOGY

Depending on the mechanism involved, resistance has been shown to arise through structural alterations of genes encoding target-site proteins or detoxifying enzymes, or through processes affecting gene expression (e.g., amplification or altered transcription). Examples of the former include the following.

- Enhanced metabolism of insecticides by cytochrome P_{450} monoxygenases can potentially confer resistance to most chemical classes. Much of the evidence for this mechanism is indirect, based on the ability of monoxygenase inhibitors to reduce the magnitude of resistance when used in combination with insecticides in bioassays.
- Enhanced activity of glutathione *S*-transferases (GSTs) is considered to be potentially important in resistance to some classes of insecticide, including organophosphates. Like monoxygenases, GSTs, exist in numerous molecular forms with distinct properties, making correlations of enzyme activity with resistance very challenging and often ambiguous.
- Enhanced hydrolysis or sequestration by esterases (e.g., carboxylesterases) capable of binding to and cleaving carboxylester and phosphotriester bonds undoubtedly plays an important role in resistance to organophosphates and pyrethroids. Biochemically, this is the best-characterized detoxification mechanism. Sometimes (e.g., for mosquitoes, blowflies, and *M. persicae*) the esterases have been identified and sequenced at the molecular level. Resistance caused by increased esterase activity can arise through a qualitative change in an enzyme, improving its hydrolytic capacity, or (as in mosquitoes and aphids) a quantitative change in the titer of a particular enzyme that already exists in susceptible insects.

The following examples appear to show that although some adaptations to the environment are unpredictable (e.g.,

the modifications of the forelimbs for flight are very different in birds, bats, and pterodactyls), the opportunities for insects to modify or reduce binding of insecticides, hence to develop target-site-based resistance mechanisms, are very limited indeed. It is conceivable that most of the mutations that confer such resistance do not allow the organism to retain normal functioning of the nervous system.

- Pyrethroids act primarily by binding to and blocking the voltage-gated sodium channel of nerve membranes. Knockdown resistance, or insensitivity of this target site, is now unequivocally attributed to structural modifications in a sodium channel protein. The same amino acid substitution (leucine 1014 to phenylalanine) in a sodium channel protein confers a "basal" kdr phenotype in a range of species including house flies, cockroaches, the green peach aphid, the diamondback moth, and a mosquito *(A. gambiae).* This phenotype may subsequently be enhanced (to "super-kdr" resistance) by further mutations that also recur between species.
- GABA receptors are targets for several insecticide classes including cyclodienes (a subclass of the organochlorines), avermectins, and fipronils. The primary mechanism of resistance to cyclodienes and fipronils involves modification of a particular GABA receptor subunit, resulting in substantial target-site insensitivity to these insecticides. The target-site mechanism of cyclodiene resistance has been attributed to the same amino acid substitution (alanine 302 to serine) in the GABA receptors of several species of diverse taxonomic origin including *Drosophila,* several beetles, a mosquito *(Aedes aegypti),* a whitefly *(B. tabaci),* and a cockroach *(Blatella germanica).* When susceptible individuals of the sheep blowfly *(L. cuprina)* were exposed to the mutagen ethyl methanesulfonate (EMS), and their progeny screened for resistance to dieldrin (a cyclodiene), surviving insects exhibited an alanine-to-serine amino acid substitution in the GABA receptor identical to that found in nature.
- Organophosphates and carbamates exert their toxicity by inhibiting the enzyme acetylcholinesterase (AChE), thereby impairing the transmission of nerve impulses across cholinergic synapses. Mutant forms of AChE showing reduced inhibition by these insecticides have been demonstrated in several insect and mite species. Biochemical and molecular analyses of insecticide-insensitive AChE have shown that pests may possess several different mutant forms of this enzyme with contrasting insensitivity profiles, thereby conferring distinct patterns of resistance to these two insecticide classes.

Some of these resistance mechanisms are illustrated schematically in Fig. 2.

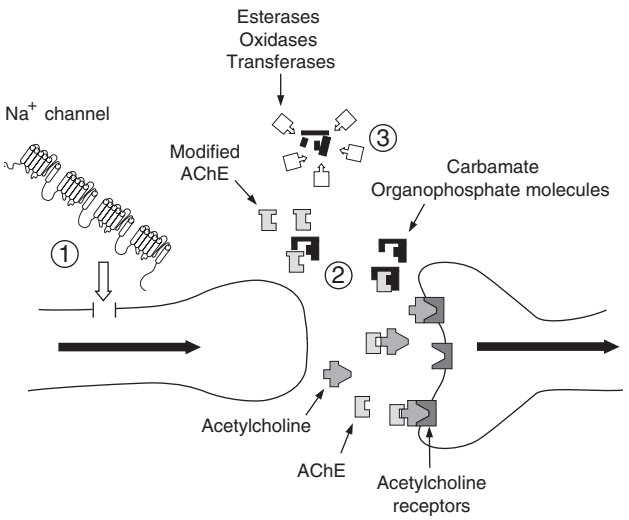

FIGURE 2 Schematic diagram of a nerve synapse showing examples of insecticide resistance mechanisms: (1) changes in the structure of the sodium channel confer kdr or super-kdr target-site resistance to pyrethroids; (2) modified AChE is no longer bound by organophosphates and remains available to break down acetylcholine molecules after neurotransmission across the synapse; (3) detoxifying enzymes degrade or sequester insecticides before they reach their targets in the nervous system.

SPREAD OF RESISTANCE GENES

The recurrence of specific resistance mutations within and between taxa begs another question: Have such mutations arisen repeatedly within the same species, or have they appeared on a limited number of occasions and subsequently spread through migration and/or human agency? Although there is molecular evidence for some resistance genes having several independent origins in the same species (e.g., for target-site resistance to cyclodienes in the red flour beetle, *Tribolium castaneum*), other examples suggest that some mechanisms have arisen only once.

Organophosphate resistance in the mosquito *Culex pipiens* is primarily conferred by allozymes at two closely linked loci (esterases A and B), coding for insecticide-detoxifying carboxylesterases. Overproduced allozymes (resulting from amplification of A or B genes) tend to recur in geographically disjunct areas. This situation could be explained by recurrent mutations generating each amplification event *de novo,* or by a nonrecurrent mutation that has spread within and between populations. Restriction mapping of DNA around the esterase genes suggest the latter, with large-scale gene flow attributable to passive migration of mosquitoes on ships and/or airplanes. A new resistance allele in southern France is known to have originated in the vicinity of the international airport and seaport at Marseilles.

Resistance to organophosphates in the aphid *M. persicae,* is also attributable to the amplification of a gene encoding an insecticide-detoxifying carboxylesterase. Despite the often widespread dispersion of these amplified genes in the aphid genome, restriction analyses have indicated that all copies are in the same immediate genetic background. This suggests that amplification occurred only once, whereupon the amplified DNA was moved intact around the genome through chromosomal rearrangements.

FACTORS AFFECTING THE EVOLUTION OF RESISTANCE

As an evolutionary trait, insecticide resistance is unusual in that we can identify the main selection pressure with ease, but the rate at which resistance develops is governed by numerous biotic and abiotic factors. These include the genetics and ecology of the pests and their resistance mechanisms, and the operational factors that relate to the chemical itself and to its application. To manage resistance effectively, an assessment of genetic, ecological, and operational risk is required. Although this can be done empirically on a species-by-species basis, one of the great challenges of the future is to understand why some species seem to have a greater tendency to become resistant than others.

Genetic Influences

To predict how quickly resistance will become established, it is necessary to understand how resistant alleles affect the survival of phenotypes in the field. For example, the dominance of resistance genes exerts a major influence on selection rates. In laboratory bioassays evaluating the relative survival of susceptible homozygotes (SS), heterozygotes (RS), and resistance homozygotes (RR) over several insecticide concentrations, RS individuals usually respond in an intermediate manner. In the field, however, dominance is dependent on the concentration of insecticide applied and its uniformity over space and time. Even when the initial concentration is sufficient to kill RS individuals (rendering resistance effectively recessive), upon weathering or decay of residues, this genotype may later show increased survival, with resistance becoming functionally dominant in expression. When resistance genes are still rare, hence mainly present in heterozygous condition, this sequence can have a profound effect in accelerating the selection of resistance genes to economically damaging frequencies.

The diverse mating systems of insects also influence the rate at which resistance evolves. Although most research has focused on outcrossing diploid species (typified by members of the Lepidoptera, Coleoptera, and Diptera), systems based on haplodiploidy and parthenogenesis also occur among key agricultural pests. In haplodiploid systems, males are usually produced uniparentally from unfertilized, haploid eggs, and females are produced biparentally from fertilized, diploid eggs. The primary consequence of this arrangement (exemplified by whiteflies, spider mites, and phytophagous thrips) is that resistance genes are exposed to selection from the outset in the hemizygous males, irrespective of intrinsic dominance or recessiveness. Whether a resistance gene is dominant, semidominant, or recessive, resistance can develop at a similar rate.

Most species of aphid undergo periods of parthenogenesis (in which eggs develop and give rise to live offspring in the absence of a paternal genetic contribution) promoting the selection of clones with the highest levels of resistance and/or the most damaging combination of resistance mechanisms. In fully anholocyclic (asexual) populations, such as those of *M. persicae* in northern Europe, the influence of parthenogenesis has led to strong and persistent associations between resistance mechanisms within clonal lineages.

Ecological Influences

Fecundity and generation times have a huge bearing on the evolution of resistance in a population. The greater the number of individuals, and the faster they reproduce and attain maturity, the higher the likelihood that a favorable mutation will occur, and be maintained in the population. Faster growth and higher population numbers will also have an effect on the size of a pest population, and therefore the need for insecticide treatment.

The dispersal capabilities of pests can also act as primary determinants of resistance development. Movement of pests between untreated and treated parts of their range may delay the evolution of resistance because of the diluting effect of susceptible immigrants. Conversely, large-scale movement can also accelerate the spread of resistance by transferring resistance alleles between localities. A good example relates to the two major bollworm species (Lepidoptera: Noctuidae) attacking cotton in Australia. Only the cotton bollworm *Helicoverpa armigera,* has developed strong resistance. *H. punctigera,* despite being an equally important cotton pest, has remained susceptible to all insecticide classes. The most likely explanation is that *H. punctigera* occurs in greater abundance on a larger range of unsprayed hosts than *H. armigera,* thereby maintaining a large pool of unselected, susceptible individuals, which dilute resistant mutations arising on treated crops.

Operational Influences

Operational factors are at human discretion and can be manipulated to influence selection rates. Factors exerting a major influence in this respect include the rate, method, and frequency of applications, their biological persistence, and whether insecticides are used singly or as mixtures of active ingredients.

Equating operational factors with selection is often difficult, since without detailed knowledge of the mechanisms present it is impossible to test many of the assumptions on which genetic models of resistance are based. If resistance alleles are present, the only entirely nonselecting insecticide doses will be ones sufficiently high to overpower all individuals, regardless of their genetic composition, or ones so low that they kill no insects at all. The latter is obviously a trivial option. Prospects of achieving the former depend critically on the potency and dominance of resistance genes present. A pragmatic solution to this dilemma is to set

application doses as far above the tolerance range of homozygous, susceptible individuals as economic and environmental constraints permit, in the hope that any heterozygotes that do arise will be effectively controlled. However, this approach will obviously be ineffective if resistance turns out to be more common than suspected (resulting in the presence of homozygous resistant individuals) or if resistance alleles exhibit an unexpectedly high degree of dominance (and heterozygotes are therefore phenotypically resistant). Unless a high proportion of insects escape exposure altogether, the consequence could then be very rapid and effective selection for homozygous resistant populations.

In practice, concerns about optimizing dose rates to avoid resistance are secondary to those related to the application process itself. Delivery systems and/or habitats promoting uneven or inadequate coverage will generally be more prone to select for resistance, because, under these circumstances, pests are likely to encounter suboptimal doses of toxins that will permit survival of heterozygous individuals.

The timing of insecticide applications relative to the life cycle of a pest can also be an important determinant of resistance. A good example of this is found in the selection of pyrethroid resistance in *H. armigera* in Australia. On cotton foliage freshly treated with the recommended field dose, pyrethroids killed larvae up to 3 to 4 days old irrespective of whether they were resistant by laboratory criteria. Since the sensitivity to pyrethroids of larvae of all genotypes was found to decline with increasing larval size, the greatest discrimination between susceptible and resistant phenotypes occurred when larvae achieved a threshold age. Targeting of insecticides against newly hatched larvae, as is generally advocated for bollworm control, not only increases the likelihood of contacting larvae at the most exposed stage in their development but also offers the greatest prospect of retarding resistance by overpowering its expression. It may also have the effect of reducing genetic variation and therefore the potential number of resistant mutations. Indeed, it is also possible to impose genetic "bottlenecks" by applying pesticides when populations are already low (e.g, when they are overwintering). Although such a tactic might be beneficial where populations are fully susceptible, if resistant mutations are already present, it might act to increase their frequency.

In theory, the application of two or more unrelated chemicals as insecticide mixtures offers substantial benefits for delaying the selection of resistance. The underlying principle is one of "redundant killing," whereby any individuals already resistant to one insecticide are killed by simultaneous exposure to another, and vice versa. However, achieving this objective requires that each type of resistance be rare and that both ingredients persist throughout the effective life of an application. Otherwise, one compound will exert greater selection pressure than the other, and the advantage of applying a mixture will be lost.

Fitness of Resistant Individuals

In the absence of insecticidal selection pressure, resistance genes can impose fitness costs on their carriers. Sometimes these costs are quite subtle and difficult to determine. In *M. persicae,* resistant individuals are less inclined to move from senescing to younger leaves and are therefore more vulnerable to isolation and starvation after leaf abscission. These costs appear to contribute to a decline in the frequency of resistant insects between cropping seasons.

COMBATING INSECTICIDE RESISTANCE

Insecticide resistance management (IRM) aims to intervene in the evolutionary process and either overcome resistance or prevent its appearance in the first place. There are several practical, economic, and political constraints on the choice of possible IRM tactics and the precision with which they can be applied:

- The properties of any resistance genes present are often unknown, and knowledge of pest ecology may still be rudimentary.
- It is often necessary to contend with a whole pest complex rather then just a single pest species.
- There may be a very limited number of insecticides available for use in management strategies.
- For highly mobile pests, at least, countermeasures may need to be standardized and synchronized over large areas, sometimes whole countries.
- Resistance is a dynamic phenomenon; that is, any mechanisms already known to exist may change over time.
- To promote compliance with management strategies, the tactics adopted should be as unambiguous, rational, and simple as possible.

A strategy first implemented on Australian cotton in 1983 against *H. armigera* illustrated many features of large-scale attempts at resistance management. Introduced in response to unexpected, but still localized, outbreaks of pyrethroid resistance in *H. armigera,* the strategy was based primarily on the concept of insecticide rotation. The threat of pyrethroid resistance was countered by restricting these chemicals to a maximum of three sprays within a prescribed time period coincident with peak bollworm damage. To diversify the selection pressures being applied, farmers were required to use alternative insecticide classes at other stages of the cropping season.

Initially, this strategy had the desired effect of preventing a systematic increase in the frequency of pyrethroid-resistant phenotypes. Additional recommendations, including the targeting of insecticides against newly hatched larvae (the most vulnerable life stage) and the plowing in of cotton stubble to destroy resistant pupae overwintering in the soil, undoubtedly contributed to this success. Unfortunately, the restrictions placed on pyrethroid use were insufficient to combat resistance in the long-term, and it has been necessary to

revise the strategy to place greater emphasis on the strategic use of nonpyrethroids against this pest.

Another strategy incorporating a wide range of chemical and nonchemical countermeasures was introduced on Israeli cotton in 1987. The primary objective was conservation of the effectiveness of insecticides against *B. tabaci.* Under recommendations coordinated by the Israeli Cotton Board, important new whitefly insecticides are restricted to a single application per season within an alternation strategy optimized to contend with the entire cotton pest complex and to exploit biological control agents to the greatest extent possible. One major achievement of this strategy has been a dramatic reduction in the number of insecticide applications against the whole range of cotton pests, but especially against *B. tabaci.* Sprays against whiteflies now average fewer than two per growing season compared with over 14 per season in 1986. Most importantly, the strategy has generated an ideal environment for releasing additional new insecticides onto cotton and for managing them effectively from the outset.

An integral part of delaying or preventing the evolution of resistance is the preservation of the innate "susceptibility" of a pest species. This is arguably as valuable a genetic resource as those of the rice, wheat or apple "gene banks" that are so carefully tended in institutes around the globe. The most effective way to conserve susceptibility, based both on evolutionary models and on empirical evidence, is to ensure the presence of pesticide-free "refugia" in which susceptible genotypes may survive and reproduce. The inclusion of refugia as essential components of IRM strategies is a recent phenomenon, signaling that pest management is no longer simply about eradication, but is now at least partially focused on conservation.

TRANSGENIC PLANTS

A recent development in crop protection has been the release of crop plants genetically engineered to express genes for insecticidal toxins derived from the microbe *B. thuringiensis.* In 2001 the total area worldwide planted to Bt plants was estimated to exceed 12 million ha. Existing toxin genes in Bt cotton and corn are active specifically against certain key lepidopteran pests (especially bollworms and corn borers); another engineered into potatoes provides protection against the Colorado potato beetle.

Aside from their commercial prospects, insect-tolerant transgenic crops offer numerous potential benefits to agriculture. By affording constitutive expression of toxins in plant tissues throughout a growing season, the incorporation of Bt genes into crops could reduce dramatically the use of conventional broad-spectrum insecticides against insect pests, as well as remove the dependence of pest control on extrinsic factors such as climate and on the efficiency of traditional application methods. However, this high and persistent level of expression also introduces a considerable risk of pests adapting rapidly to resist genetically engineered toxins. To date, there are no substantiated reports of resistance selected directly by exposure to commercial transgenic crops, but resistance to conventional Bt sprays (selected in either the laboratory or the field) has been reported in more than a dozen insect species. Research into the causes and inheritance of such resistance is providing valuable insights into the threats facing Bt plants and the efficacy of possible countermeasures.

Tactics proposed for sustaining the effectiveness of Bt plants have many parallels with those considered for managing resistance to conventional insecticides. However, they are more limited in scope because of the long persistence and constitutive expression of engineered toxins, and because of the limited diversity of transgenes currently available. Indeed, for existing "single-gene" plants, the only prudent and readily implementable tactic is to ensure that substantial numbers of pests survive in nontransgenic refugia. These can be incorporated into the crop itself, or or they may comprise alternative host plants. The success of this strategy is dependent on some key assumptions: (1) that resistant mutations are recessive or at least only partially dominant, so that their heterozygous forms can be controlled by the toxins expressed; (2) that refugia will produce enough susceptible insects to ensure that insects carrying resistant alleles do not meet and mate; and (3) that resistant alleles will carry a fitness cost, rendering insects less fit when the selection pressure is removed (e.g, outside the growing season when the insect is dependent on other crops).

In the longer term, there are potentially more durable options for resistance management: stacking (or pyramiding) of two or more genes in the same cultivar, or possibly rotations of cultivars expressing different single toxins. Whatever measures are adopted, it is essential that plants expressing transgenes be exploited as components of multitactic strategies rather than as a panacea for resistance problems with conventional insecticides.

RESISTANCE IN NONPEST SPECIES

The ability of insect predators and parasitoids to develop pesticide resistance would be of enormous benefit to pest management strategies that are chemically dependent. Although pyrethroid and organophosphate resistance has been documented in predatory mites (e.g., *Typhlodromus pyri* in orchards and *Amblyseius womersleyi* in tea fields) and hymenopterous parasitoids (e.g., *Aphytis holoxanthus* in orchards and *Anisopteromalus calandrae* in grain stores), reports of insecticide-resistant beneficial species from the field are far rarer than they are for pest species. Reporting bias aside, the most likely reasons for this are the difficulty in host location when both natural enemy and host are under selection pressure and, in comparison with herbivorous species, the possibility that the enzyme systems of predators and parasites are less well adapted to detoxify xenobiotics.

Resistance may, therefore, be more likely to develop if the hosts or prey are themselves resistant, thereby making their location easier. For example, a parasitic wasp *(A. calandrae)* of a stored-grain beetle *(Sitophilus oryzae)* is resistant to insecticides, and it is thought that this adaptation has been encouraged because the host organism is sheltered from insecticides by the grain kernels it inhabits.

Many attempts have been made to select resistance in beneficial species in the laboratory, but limitations on the size of the populations (and hence their genetic variability) that can be maintained under these conditions means that resistance tends to arise through the development of polygenic traits. Once released into natural populations, these are more likely to fragment and dissipate than rarer, but generally more robust, single mutations.

In general, when resistance does occur in nonpest species, its mechanisms are similar to those exhibited by pest species. Organophosphate resistance in strains of *A. calandrae* has been linked to the presence of carboxylesterase-like enzymes similar to those conferring organophosphate resistance to the *M. persicae.* The expression level of the carboxylesterase-like enzyme in this wasp is approximately 30-fold higher in the resistant strain relative to that in the susceptible strain, and the mechanism seems to have its basis in a single nucleotide replacement. Organophosphate resistance in strains of the warehouse pirate bug *(Xylocoris flavipes)* has also been linked to the presence of a carboxylesterase. Resistance to this chemical group in the lacewing, *Chrysopa scelestes,* has been attributed to increased AChE activity.

CONCLUDING REMARKS

Research on the topic of insecticide resistance has provided invaluable insights into the origin and nature of adaptations, and these are proving to have broad significance for understanding genetic responses to change in the environment. In many respects the continuing battle against resistance is as good an example of coevolution as any and is a clear illustration of how such processes generate biological diversity. In this instance, however, the diversity being created is undesirable from a human standpoint and, because of the threat posed to susceptible genotypes, probably temporary.

It is important to note that the pest management problems posed by the evolution of resistance are not unique to control strategies that use conventional insecticides. The utilization of host plant resistance is a case in point. Resistance to insects in crop plants is selected by screening for genes that provide resistance in the laboratory or in field plots, then crossing those genes into crop strains with other desirable characteristics. At least six major genes for resistance to the Hessian fly *(Mayetiola destructor)* have been successively bred into wheat over the past two decades. In each instance, the introductions of new resistant mutations in the plant were rendered useless by the evolution of corresponding protective adaptations in the fly. Another example of such coevolution comes from the use of semiochemical tools for pest control. In many parts of Asia, a synthetic pheromone is used to disrupt mating in the tea tortrix moth *(Adoxophyes honmai),* the larvae of which can cause severe damage in tea plantations. Researchers in Japan have recently reported the evolution of a new biotype of this species that exhibits reduced sensitivity to the pheromone. Such events make it clear that regardless of whether the major strategies for pest management continue to use conventional chemicals, the "arms race" between insect evolution and human ingenuity will continue to present major challenges.

See Also the Following Articles

Agricultural Entomology • Biotechnology and Insects • Genetically Modified Plants • Genetic Variation

Further Reading

Denholm, I., and Rowland, M. W. (1992). Tactics for managing pesticide resistance in arthropods—Theory and practice. *Annu. Rev. Entomol.* **37,** 91–112.

Devonshire, A. L., Field, L. M., Foster, S. P., Moores, G. D., Williamson, M. S., and Blackman, R. L. (1998). The evolution of insecticide resistance in the peach-potato aphid, *Myzus persicae. Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci.* **353,** 1677–1684.

ffrench-Constant, R. H., Pittendrigh, B., Vaughan, A., and Anthony, N. (1998). Why are there so few resistance-associated mutations in insecticide target genes? *Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci.* **353,** 1685–1693.

Forrester, N. W., Cahill, M., Bird, L. J., and Layland, J. K. (1993). Management of pyrethroid and endosulfan resistance in *Helicoverpa armigera* (Lepidoptera, Noctuidae) in Australia. *Bull. Entomol. Res.* suppl. 1.

Horowitz, A. R., Forer, G., and Ishaaya, I. (1994). Managing resistance in *Bemisia tabaci* in Israel with emphasis on cotton. *Pestic. Sci.* **42,** 113–122.

Martinez Torres, D., Devonshire, A. L., and Williamson, M. S. (1997). Molecular studies of knockdown resistance to pyrethroids: Cloning of domain II sodium channel gene sequences from insects. *Pestic. Sci.* **51,** 265–270.

McKenzie, J. A. (1996). "Ecological and Evolutionary Aspects of Insecticide Resistance." R. G. Landes, Austin, TX.

Pimental, D., Acquay, H., Biltonen, M., Rice, P., Silva, M., Nelson, J., Lipner, V., Giordano, S,. Horowitz, A., and Damore, M. (1992) Environmental and economic costs of pesticide use. *Bioscience* **42,** 750–760.

Raymond, M., Chevillon, C., Guillemaud, T., Lenormand, T., and Pasteur, N. (1998). An overview of the evolution of overproduced esterases in the mosquito *Culex pipiens. Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci.* **353,** 1707–1711.

Roush, R. T. (1989). Designing resistance management programs—how can you choose? *Pestic. Sci.* **26,** 423–441.

Roush, R. T. (1997). Bt-transgenic crops: Just another pretty insecticide or a chance for a new start in resistance management? *Pestic. Sci.* **51,** 328–334.

Soderlund, D. M., and Bloomquist, J. R. (1990). Molecular mechanisms of insecticide resistance. *In* "Pesticide Resistance in Arthropods" (R. T. Roush and B. E. Tabashnik, eds.), pp. 58–96. Chapman & Hall, London.

Tabashnik, B. E., Liu, Y. B., Malvar, T., Heckel, D. G., Masson, L., and Ferre, J. (1998). Insect resistance to *Bacillus thuringiensis:* Uniform or diverse? *Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci.* **353,** 1751–1756.

Insectivorous Plants

Lewis J. Feldman
University of California, Berkeley

The term "insectivorous" was used by Charles Darwin to characterize a group of plants that seemed to trap and feed on insects. Since Darwin's time, observations have revealed that these plants capture and interact with a greater variety of animals, which can include spiders, lizards, sow bugs, tadpoles, and frogs, and judging from some reports, even mammals such as rats and rabbits. Hence, because of this varied diet, many workers now prefer to describe such plants as carnivorous, rather than solely insectivorous. Yet the interactions between carnivorous plants and animals go beyond the presence of certain creatures as items on a plant's menu. While it is true that the most spectacular and usually the most obvious activities of carnivorous plants seem to be in the often elaborate mechanisms for capture and digestion of prey, many other (often more subtle) associations, occur between these plants and animals. Researchers are just at the beginnings of learning about these other fascinating interactions.

THE CARNIVOROUS HABIT

Plant carnivory is a rarity, occurring in only about 550 out of approximately 250,000 plant species. The carnivorous habit is not obligate, and carnivorous plants can grow without an insect meal, depending instead on photosynthesis and minerals supplied from the soil. In general, carnivorous plants grow in sunny areas, and in mineral-deficient, sometimes sandy soils. Often these soils have standing or gently moving water, with any dissolved minerals from the soil being easily carried away by the flowing water. The carnivorous plant habitat is typically low in nitrogen and phosphorus and, some reports suggest, in potassium as well. In this sort of habitat, plants that have alternative strategies for obtaining essential minerals are at a competitive advantage. The capture of insects and other animals thus provides carnivorous plants with a supplemental source of essential nutrients.

The carnivorous habit depends on an ability to trap prey. In the vast majority of carnivorous plants, the trap represents a modification of the entire leaf or of structures borne on the leaf. Given this rather straightforward requirement of a trap, it should be easy enough to characterize a plant as carnivorous, or not. However, the picture is not so simple: many plants can trap insects yet are not considered to be carnivorous. What truly distinguishes a plant as carnivorous is not only a trapping ability, but also a mechanism to digest prey and to absorb the prey's nutrients.

Digestion implies an ability of the plants to break down the trapped prey into its component chemicals, to be able to absorb them as nutrients. It is the specifics of this digestion that have caused some controversy. Some workers consider a plant to be carnivorous only if it has an inherent ability to digest prey—that is, if the plant itself produces enzymes to break down the insect. Other plants, sometimes considered to be semicarnivorous, are able to trap prey but depend on the assistance of other organisms, usually, but not always, microbes, to digest the captured insects. However, for this short article, a plant is considered to be carnivorous if it traps and has a means, of its own making or not, for digesting prey.

The trap of a carnivorous plant is a true marvel, designed to attract, capture, digest, and then absorb nutrients from the prey. Traps can be grouped by whether they are "passive," with no or slowly moving "parts," often relying on gravity to aid in capture of the insect, or "active," exhibiting some sort of usually rapid movement. Perhaps the most familiar examples of passive traps are the sundews *(Drosera)* and the pitcher plants (*Sarracenia* and *Darlingtonia* in temperate climes, and *Nepenthes* in the tropics).

PASSIVE TRAPS

The sundews capture their prey by producing from stalked glands an adhesive, or glue (the drop of "dew"), which captures and holds fast the insect. As the prey struggles, it is covered with the sticky mucilage, and as a consequence, suffocates. The stalked glands then bend in toward the prey; in some species, the entire leaf enfolds the prey. A second type of gland on the leaf secretes digestive enzymes and acids, initiating the breakdown and subsequent absorption of nutrients. Darwin was so enthralled with the sundews that about two-thirds on his book *Insectivorous Plants* is devoted to this group. He notes his surprise at "finding how large a number of insects were caught by the leaves of the common sundew," and speculates that "as this plant is extremely common in some districts, the number of insects thus annually slaughtered must be prodigious."

A fascinating variation on the carnivorous plant passive trap theme is shown by plants that comprise the genus *Roridula.* These plants, considered by some workers not to be truly carnivorous, are native to South Africa and may be near extinction. Individuals in this group have leaves covered with stalked mucilage-secreting glands, which as in the sundews, capture and hold fast insects. This is where the carnivorous story would end, were it not for another player, an assassin bug (Heteroptera: Reduviidae). Large numbers of these capsid bugs may inhabit *Roridula* and are able to traverse the leaves without themselves being ensnared by the glue. When other insects are captured by the plant, the assassin bugs move to the trapped prey, suck out their liquid contents, and, some time afterward, secrete a nutritious substance that is absorbed by the leaf and nourishes the plant.

The second major group of plants having passive traps are the pitcher plants. In this group the leaf becomes variously modified, often into a tube, and develops at the base of the

FIGURE 1 Habit view of the California pitcher plant *(Darlingtonia californica)* growing in a bog in northern California. The leaf is divided into a hood region and tube region. Digestion occurs in a well of water at the base of the tube.

tube a well that must fill with water for the pitcher to function as a trap. The temperate species of pitcher plants (*Darlingtonia* in the western United States and *Sarracenia* in eastern North America) (Figs. 1 and 2) are usually terrestrial. In these plants the leaf lures flying insects by producing nectar that sometimes covers colorful appendages. Crawling insects follow nectar trails running along the outside of the leaf. The nectar trails lead to the mouth of the tube, where the surface is smooth and slippery and from which the insect can easily lose its foothold, thus falling into the watery well. Escape from the well is almost impossible, since the inside wall of the tube-leaf is lined with downward-pointing hairs. One would think that flying insects could fly out if they started to fall. To counter this possibility, pitcher plants such as *Darlingtonia* have developed a hooded leaf, transparent and sealed at its top. When an insect tries to leave the leaf, it flies toward light coming through the transparent upper portion of the hood. Since, however, the exit is sealed, eventually the insect becomes so exhausted that it falls into the well. There is some suggestion that pitcher plants may produce a "drug" to confuse the flying insect, and fluids in the well may contain substances that stun and quiet the fallen insect. In addition to nectar serving as an attractant, the possible development of ultraviolet signaling, as employed by flowers to attract pollinators, may also serve to lure insects to the trap. Many temperate pitcher plants secrete hydrolytic enzymes into the liquid in the well, thereby digesting the insect, whereas other pitcher plants (e.g., *Darlingtonia*) produce none of their own digestive enzymes but instead rely on bacteria to decay the insect. In either model, the digestive enzymes can be quite powerful, with only the hardest parts of insects, such as legs or shells, remaining undigested.

FIGURE 2 Eastern pitcher plant *(Sarracenia purpurea)* from Ohio. In this species a hood is absent.

The tropical pitcher plants belong to the genus *Nepenthes,* so named by Linnaeus after the drug "nepenthe," which Helen of Troy was said to have dispensed in drink to soldiers to "relieve their sorrow and grief." In giving this name, Linnaeus noted, "What botanist would not be filled with admiration if, after a long journey, he should find this wonderful plant. In his astonishment, past ills would be forgotten when beholding the admirable work of the creator!" In *Nepenthes,* the pitcher develops at the end of a leaflike petiole. Indeed, the complexity and variety of pitchers in *Nepenthes* strains one's credulity, for it is hard to believe that what one is looking at is a leaf. Like their temperate cousins, *Nepenthes* spp. produce nectar to lure prey, which subsequently become intoxicated, lose their foothold, and fall into the trap. *Nepenthes* spp. generally produce climbing stems, thus elevating the pitchers, and perhaps thereby making them more accessible to potential prey.

Species of *Nepenthes,* and likely all carnivorous plants, do not seem to be designed to trap one particular species. An inventory of the traps shows that their diets are ever-changing and can be quite varied. For example, from 10 *Nepenthes* pitchers over a season, Erber found arthropods of 150 identified species, belonging principally to the orders Diptera, Hymenoptera, and Collembola, and the families Fromicidae, Aphididae, and Acarina. Similar tallies in *Sarracinea* reveal victims of 115 families belonging to 14 orders of insects, including several species of Mollusca. This strategy no doubt ensures some prey, if, for example, a particular insect species is not present one year, or becomes

extinct, and also allows the plant to trap a variety of insects over a very long season. Additionally, such variety may be important in supplying a diversity of nutrients.

ACTIVE TRAPS

Plants with rapid movement usually first come to mind when carnivorous plants are mentioned. These so-called active traps imprison their prey by a quick movement of all or part of the leaf. Into this group are placed several genera; one well known, the Venus flytrap *(Dionaea)* (Fig. 3) and the others less familiar (e.g., the bladderworts, *Utricularia*) (Fig. 4). It is believed that the rapid movement comes about when the prey makes contact with a triggering mechanism, resulting in the generation of a small electric current and the activation or closure of the trap. As with the passive traps, the lure for the insect is usually some sweet nectar. In the case of the Venus flytrap leaf, in which the two halves of the blade are joined along one side, as in an open book, nectar is produced on the inner surface of the leaf (Fig. 3). As an insect wanders along this inner surface collecting nectar, it may contact the trigger hairs, of which there are about two to four on the inner surface of each half-leaf. For the trap to close, either an individual trigger hair "must be touched twice, or two different trigger hairs must be touched sequentially, within a time period that is neither too short (< 0.75 s) nor too long (> 20 s)." When the hair or hairs are touched within the right time interval, the trap literally snaps shut, though at first, not completely. Initially, small openings remain between the two halves, presumably to allow smaller insects to escape from the trap. When unsuitable prey gain their release, the trap reopens and awaits the main course. But if the insect is unable to escape through the small openings and continues to struggle, the trap closes more fully. Subsequently, enzymes are secreted by special gland cells, and the insect is digested and its nutrients absorbed by the leaf.

Less well known as active trappers but possessing traps more complex than the Venus flytraps are the bladderworts, which grow in wet or periodically wet areas. Bladderworts, the largest genus of carnivorous plants, grow worldwide, on every continent. They develop diminutive, often microscopic traps that cover the leaves (Fig. 4). The size of the trap determines what creatures will enter: paramecia, rotifers, water fleas, worms, and mosquito larvae, for example. As in the Venus flytrap, contact with trigger hairs initiates the trapping mechanism, which involves the opening of a "door" leading to a chamber maintained under a vacuum, a sucking in of the prey, and a resealing of the trap; all this occurs within 10 to 15 thousandths of a second! With the secretion of enzymes, the prey is digested, usually within hours. There is some speculation that the trap can also lure prey.

FIGURE 3 Habit view of the Venus flytrap (*Dionaea* sp.). In this genus the blade is divided into two halves, which are attached along one side. On the inner surfaces of the blade a lure is produced, and here too are located the trigger hairs.

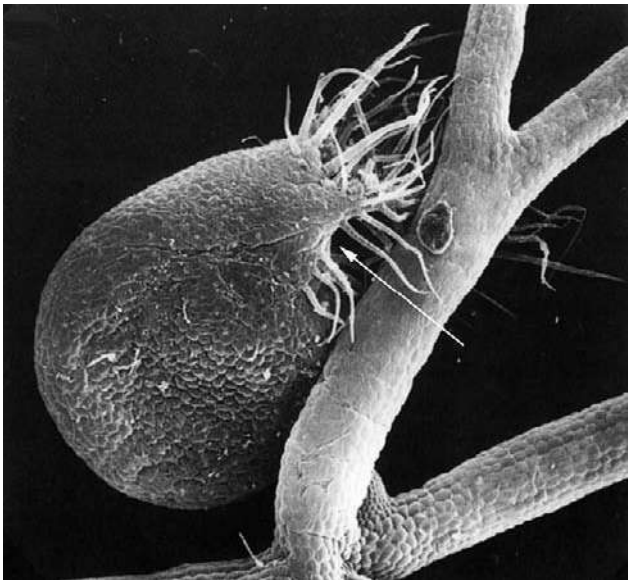

FIGURE 4 Scanning electron micrograph view of the trap of *Utricularia neglecta.* The large hairs ("antennae") may act as guides luring the prey to the trap mouth (arrow). [After Juniper, B. E., Robins R. J., and Joel, D. M. (1989). "The Carnivorous Plants." Academic Press, London. Reprinted with permission.]

OTHER INTERACTIONS BETWEEN CARNIVOROUS PLANTS AND INSECTS

Thus far, plant–insect interactions have been presented in the context of insects serving as prey. But the associations can be much more varied and complex, as seen in the instance of the assassin bug and *Roridula.* Entomologists are now only beginning to appreciate the many other ways in which carnivorous plants and insects/animals interact. Some of the more fascinating examples are found in the pitcher plants, where other animals turn the traps to their own advantage. Spiders often can be found prowling about the mouth of pitchers, then lowering themselves on silken strands to retrieve prey from the pitcher well. In *Nepenthes,* spiders also use the pitcher for protection from predators. If a predator is detected, the spider will lower itself on a silk thread to the pool and, if necessary, will even hide under the water until danger passes.

Other insects spend part or most of their lives in the wells of pitchers. For example, the pitcherplant mosquito *(Wyeomyia smithii)* lays its eggs on the moist inner surface of the leaf, or more often, in the pool of liquid. The larvae hatch and feed on detritus from trapped insects, bacteria, and protozoans. As winter approaches, the larvae go into a dormant state and overwinter in the pitcher, exiting the pitcher in the spring as adult mosquitoes. In climates where water freezes, the larvae spend winter frozen in the ice of the pitcher. Interestingly, the liquid that digests the trapped insects seems to have no detrimental effects on the larvae. Fish fly larvae also live in pitchers and, like the mosquito larvae, are not injured by the pitcher's digestive enzymes because, it is speculated, their bodies produce a protective substance.

One of the most fascinating examples of an association between plant and insect, benefiting the insect, is seen in species of *Exyra* moths (Noctuidae) that exploit the pitcher leaf to shelter their young. The cycle begins with the female moth entering an open leaf and laying its eggs on the inner wall of the pitcher leaf. When the larvae hatch, they move about on silken strands, feeding on the inner wall. As they grow, hence becoming more visible to predators, the larvae move to the top of the pitcher, severing vascular strands carrying water to the upper regions of the pitcher leaf, which causes the top of the pitcher leaf to dry, collapse, and fold over the opening. The developing larvae are now shielded from predators. Just before a larva prepares to pupate, it chews a hole in the wall of the leaf. Through this hole, the moth exits the leaf.

Yet another example of an insect exploiting its association with carnivorous plants is the solitary sarracenia wasp *(Chlorion harrisi).* This insect uses the pitcher as an incubator for its eggs. In preparation for the laying of eggs, the wasp packs into the bottom of the pitcher tube a layer of grass, which is then overlayered with freshly killed grasshoppers or crickets. This process may be repeated several times, resulting in alternating layers of insects and plant materials. Eggs are then laid among the dead insects and the whole construction covered by another layer of grass. The eggs can now develop protected, and when the young hatch, they have a food supply.

The last example represents what may come closest to a commensal, or symbiotic, association between plant and insect. Certain species of *Nepenthes* (e.g., *N. bicalarate*) have enlarged, hollow petioles in which ants take up residence. In return for this "home" (domatia), the ants, it is suggested, protect the plant from predators.

See Also the Following Article

Plant–Insect Interactions

Further Reading

D'amato, P. (1998). "The Savage Garden." Ten Speed Press, Berkeley, CA.
Darwin, C. (1875). "Insectivorous Plants." Appleton, New York.
Juniper, B. E., Robins, R. J. and Joel, D. M. (1989). "The Carnivorous Plants." Academic Press, London.
Lerner, C. (1983). "Pitcher Plants: The Elegant Insect Traps." Morrow, New York.
Lloyd, F. E. (1942). "The Carnivorous Plants." Chronica Botanica, Waltham, MA.

Insectivorous Vertebrates

Insects are a part of the diet of a vast array of animals. Indeed, there are almost no animal groups that do not include some representatives that consume insects, incidentally or intentionally, actively or accidentally. Even obligate herbivores consume insects, although never in any large number. Several groups of animals are insectivorous as juveniles and shift dietary focus as adults, whereas other groups become insectivorous only as adults.

DIVERSITY OF INSECTIVORY

A simple list of the insectivorous terrestrial animals would be lengthy and include many freshwater fishes, most frogs and salamanders, lizards and snakes, and birds and mammals. Because marine ecosystems do not include many insects, the incidence of insectivory among marine invertebrates and vertebrates is much less common.

In freshwater fishes, insects are ubiquitous and widely consumed. In streams, for example, 70 to 90% of the macroinvertebrates are insects, comprising as much as 99% of the numbers of individual organisms and 99% of the biomass. Fishes take advantage of these abundant resources and consume insects from all of the 13 orders of insects with aquatic life stages.

Terrestrial salamanders and almost all frogs, which are often filter feeders or herbivores as larvae, become

predominantly insectivorous as adults. Frogs and salamanders that remain aquatic, or become so secondarily, possess a wide array of dietary choices that can include other aquatic life, including (but not limited to) insect adults or larvae.

Lizards, more so than snakes, include insects in the diet. Consumption of insects also occurs in other reptiles such as turtles and juvenile crocodilians. Insect prey selected by lizards is somewhat size dependent; smaller lizards consume more small insects, whereas larger lizards can also consume larger insects. Some lizards that are insectivores as juveniles become more herbivorous as adults. In snakes, smaller insectivores become more carnivorous as they get larger, focusing on other vertebrate prey, especially mammals, frogs, and other snakes. Among some groups of lizards and snakes, specialization for insectivory is a familiar pattern; in these instances, ants and termites are most frequently consumed.

There are many birds that consume insects as a dominant part of the diet. Some of these insectivorous lineages include pipits and wagtails (Motacillidae), bulbuls and allies (Campephagidae, Pycnonotidae, Chloropseidae), waxwings and allies (Ptilogonatidae, Bombycillidae, Dulidae), dippers (Cinclidae), warblers and gnatcatchers (Sylviidae, Parulidae), flycatchers (Muscicapidae), and titmice, nuthatches, and treecreepers (Paridae, Sittidae, Certhiidae). Many birds capture insect prey in flight, whereas other birds forage in shrubs and trees or on the ground. Some birds specialize by obtaining insect food grooming large mammals or following behind large mammalian herbivores and foraging on the insects disturbed by large mammal movements.

Mammals include many insectivorous groups, some generalists and others obligate specialists. Most of those that specialize in eating insects eat either ants or termites. Generalized insectivores will eat insects along with other arthropods such as centipedes, millipedes, spiders, and scorpions. Marsupials, bats, primates, rodents, carnivores, and other groups of mammals include insectivorous lineages. Specialized feeding on particular insects, especially ants and termites, occurs among a few frogs, many lizards, and some snakes and has occurred in several different lineages of mammals (see Table I).

ANATOMICAL SPECIALIZATIONS FOR INSECTIVORY

Although eating insects is a dominant part of many diets, anatomical and behavioral specializations for insectivory are not as widespread. Many lizards, for example, feed on whatever suitable prey item might be available. Nevertheless, there are several anatomical specializations that seem to assist in the capture of insect prey. Perhaps the most obvious and remarkable trait is a highly projectile tongue. This sort of tongue evolved many times, in many different lineages, and in many different ways. Most frogs have a projectile tongue whose intrinsic muscles attach to the lingual edge of the symphysis of the jaw. The tongue is flipped out of the mouth, in the same way that a catapult works, so that the back of the tongue after it is extruded contacts the prey first. Salamanders have evolved several different types of projectile tongues. One group of lungless salamanders, the plethodontids, uses the hyobranchial skeleton and associated muscles, once used to ventilate lungs, in a tongue protrusion mechanism that is quite spectacular. Contraction of these muscles results in protrusion of the tongue as well as large parts of hyobranchial skeleton, resulting in an extruded tongue that can reach 80% of body length.

Chameleon lizards have an unmatched ability to accurately aim, project, and hit arboreal insect prey. The muscles that chameleons use to accomplish this ballistic feat contract faster than any other vertebrate muscle. Furthermore, the tongue

TABLE I Examples of Frog, Lizard, and Mammal Lineages in Which Termites and/or Ants Are a Significant Part of the Diet

Frogs	Mammals
Dendrobatidae—poison frogs	Canidae—dogs and foxes
Microhylidae—narrowmouth frogs and toads	Cercopithecidae—Old World monkeys
Pelobatidae—spadefoot toads	Cricetidae—New World mice, hamsters, etc.
Rhinophrynidae—Mesoamerican burrowing toads	Dasypodidae—armadillos
Lizards	Didelphidae—opposums
Agamidae—angelheads, calotes, dragon lizards, and allies	Herpestidae—mongooses
Amphisbaenia—wormlizards	Lorisidae—lorises
Gekkonidae—geckos and pygopods	Manidae—pangolins
Iguaninae	Muridae—Old World rats
Lacertidae—wall lizards, rock lizards, and allies	Myrmecobiidae—marsupial anteaters
Phrynosomatinae	Myrmecophagidae—anteaters
Scincidae—skinks	Orycteropidae—aardvarks
Teiidae—whiptail lizards, tegus, and allies	Sciuridae—squirrels
Tropidurinae	Tachyglossidae—echidnas
Snakes	Talpidae—moles
Anomalepididae—early blindsnakes	Thylacomidae
Leptotyphlopidae—threadsnakes and wormsnakes	Vespertilionidae—bats
Typhlopidae—blindsnakes	

Note. Not all the taxa that are a part of these lineages eat only ants and/or termites. There are many other lineages not listed that eat insects.

can be accurately projected up to 200% of body length. The tongue has a large tip covered with viscous mucous. A muscle in the large fleshy tip contracts just after the tongue tip strikes a prey item, creating a slight vacuum that assists in prey capture.

Mammals eat insects and the most distant ancestors of mammals may have been insectivores. Generalized features derived from the primitive amniote condition are associated with mammalian feeding, including a longer secondary pallate, heterodont dentition, higher metabolic rates and more active foraging behaviors. Several of these features are reversed in obligate ant- and termite-eating mammals. For example, anteaters, pangolins, and the echidna, numbat, and aardvark possess highly simplified teeth few in number or lack teeth entirely. Some ant and termite specialists have lower metabolic rates, but it is not clear if these rates are retained from a primitive mammalian ancestor or if they are a response to prey that may present chemical challenges to typical mammalian digestive systems. A long, sticky, and protrusible tongue is a common feature among ant- and termite-feeding mammals. Details of tongue anatomy confirm that many of these lineages evolved these specializations independently.

See Also the Following Articles

Food, Insects as • Predation

Further Reading

Cushing, C. E., and Allan, J. D. (2001). "Streams: Their Ecology and Life." Academic Press, San Diego.

Gerking, S. D. (1994). "Feeding Ecology of Fish." Academic Press, San Diego.

Schwenk, K. (ed.) (2000). "Feeding: Form, Function, and Evolution in Tetrapod Vertebrates." Academic Press, San Diego.

Thorp, J. H., and Covich, A. P. (eds.) (2001). "Ecology and Classification of North American Freshwater Invertebrates," 2nd ed. Academic Press, San Diego.

Zug, G. R., Vitt, L. J., and Caldwell, J. P. (2001). "Herpetology: An Introductory Biology of Amphibians and Reptiles," 2nd ed. Academic Press, San Diego.

Insect Zoos

Leslie Saul-Gershenz

Center for Ecosystem Survival, San Francisco, California

Insect zoos, exhibits that display live insects and arthropods to the general public, have gained recognition for their educational value and broad appeal. They have demonstrated that insects are interesting to millions of people, not just a select few. Insects represent the majority of terrestrial species on earth and display a dazzling diversity of lifestyles, behaviors, and adaptations. The great naturalists Pliny, Fabre, Wallace, Darwin, Belt, and Wheeler possessed a boundless desire to observe living insects in the natural world. The observation of living insects in the field and lab still excites the imagination of naturalists today, and the general public has been increasingly infected with this enthusiasm through exposure to the lives of insects through superb nature films and photography. Insect zoos provide opportunities for positive firsthand observations of live insects to millions of people. These exhibits are intended to inspire a curiosity to learn more, and to balance the myths and misrepresentations promoted by the cinema and other commercial media. Insectariums have proven to be enormously popular with the public wherever they appear, whether modest or grand, whether showcasing native or exotic species. Today there are over 100 insect zoo exhibits throughout the world.

INSECT ZOOS DEFINED

The term "insect zoo" has been applied to facilities of many different types. Defined broadly, an insect zoo or insectarium is an exhibit facility dedicated to the display of live insects housed in a separate room, building, or distinct exhibit hall and maintained primarily for public visitation. Insect zoos typically are permanent, year-round facilities that house live insects and related groups of arthropods (arachnids, centipedes, millipedes, and crustaceans) and occasionally representatives of other invertebrate groups. Insect zoos have been built in zoological parks, natural history museums, botanical gardens, county parks, horticultural centers, amusement parks, nature reserves, and universities, and on privately owned land. Interest in their development has increased (excluding the period between the first and second World Wars) with growing public interest in biodiversity and the documented success of insect exhibits (Table I).

Most facilities contain a series of terrariums, where species are displayed in naturalistic mini-environments. A major insectarium is a comprehensive coverage of the class Insecta, representing many different orders such as Coleoptera, Hymenoptera, Orthoptera, and Mantodea. Observation honey bee hives, ant and termite colonies, walkingsticks, katydids, lubber grasshoppers, and assassin bugs are a few examples of typical displays. These facilities can be distinguished from collections of a few species of arthropod housed in a reptile house or aquarium or included in an exhibit that focuses on the interpretation of a particular ecosystem. Many tropical rain forest exhibits today include a few displays of insects, often a leafcutter ant exhibit and a few other invertebrates as nominal representatives of the vast diversity of invertebrates. However, overall these exhibits emphasize the vertebrate fauna of rain forests and present a relatively minor treatment of the subject of invertebrates. Major insectariums typically have between 30 to 100 live displays, exhibiting up to 100 species of arthropods.

A butterfly house or lepidoptery is a type of live insect facility that primarily displays members of the order

TABLE I List of Selected Insect Exhibits Built Worldwide from 1797 through 2001

Exhibit	Date of establishment
Europe	
Small early acclimatization collections	
Jardin des Plantes, Muséum d'Histoire Naturelle et Ménagerie, Paris, France	1797
Jardin d'Acclimatisation, Paris, France	1860
Permanent exhibits	
Insect House, London Zoo, London, England (renovation 1913, Web of Life, 1999)	1881
Artis Zoo, Amsterdam, Netherlands	1898
Frankfurt Zoo, Frankfurt, Germany (renovation 1957)	1904
Zoologischer Garten Köln, Cologne, Germany (new Insektarium 1971)	1905–1929
Budapest Zoo, Budapest, Hungary (renovations: Vivarium, 1970; Butterflies, 2000)	1907
Zoo-Aquarium Berlin, Berlin, Germany (renovated 1978–1983)	1913
Zoologischer Garten, Leipzig, Germany	1913
Gruga Park, Essen, Germany	1956
Sherbourne Butterfly House, Dorset, England	1960
Löbbecke Museum and Aquazoo, Düsseldorf, Germany (renovation in 1987)	1970
Guernsey Butterfly Farm, England	1977
Insektarium, Zoological–Botanic Garden Wilhelma, Stuttgart, Germany	1980
London Butterfly House, Syon Park, England	1981
Stratford-upon-Avon Butterfly Farm, Stratford-upon-Avon, England	1985
Papiliorama-Nocturama, Marin Center, Neuchâtel, Switzerland	1988
Dortmund Zoo, Dortmund, Germany	1991
Bug World, Bristol Zoo, Bristol, England (renovated in 1996)	1992
Butterflies Center, Girona, Spain	1995
Mariposario del Drago, Tenerife, Spain	1997
Idea Schmetterlings-Paradies Neuenmark, Germany	1998
Krefeld Zoo, Krefeld, Germany	1998
Micropolis, Saint-Léon-en Lévezou France	2000
The Butterfly Arc, Montegrotto Terme, Italy (seasonal, 1988)	2001
Projeckledare Aquademin, Sweden "Science Centre," Göteborg, Sweden	2001
Asia	
Takarazuka Insectarium, Takarazuka Zoological and Botanical Garden, Japan	1954–1967
Toshima-en Insectarium Toshima-en Amusement Park, Tokyo, Japan	1957
Insectarium, Tama Zoo, Japan (renovations in 1966, 1975, 1988)	1961
Penang Butterfly House, Penang, Malaysia	1986
Fragile Forest, Singapore Zoo, Singapore	1998
North America	
United States	
Early short-lived or seasonal exhibits	
Bronx Zoo, New York Zoological Society, New York	1910, 1940, 1945
Goddard State Park, Providence, Rhode Island	1934–1937
Chicago (Brookfield) Zoo, Brookfield, Illinois	1938, 1947–1950
Flushing Meadow Zoo, New York	1969–1970
Insect Zoo, Smithsonian National Museum of Natural History, Washington, DC	1971
Permanent year-round exhibits	
Arizona–Sonoran Desert Museum, Tucson, Arizona	1952
Otto Orkin Insect Zoo, Smithsonian Museum of Natural History, Washington, DC	1976
World of the Insect, Cincinnati Zoo and Botanic Garden, Cincinnati, Ohio	1978
Insect Zoo, San Francisco Zoological Gardens, San Francisco, California	1979
Invertebrate Exhibit, National Zoological Park, Washington DC	1987
Butterfly World, Coconut Creek, Florida	1988
Cecil B. Day Butterfly Center, Callaway Gardens, Pine Mountain, Georgia	1988
Ralph K. Parsons Insect Zoo, Natural History Museum of Los Angeles County, Los Angeles, California	1992
Cypress Gardens, Winter Haven, Florida	1992
Terminix Insect City, Fort Worth Zoo, Fort Worth, Texas	1992
Butterfly Encounter, San Diego Wild Animal Park, Escondido, California	1993
Moody Gardens, Galveston, Texas	1993
Cockerell Butterfly House, Houston Natural Science Museum, Houston, Texas	1994
Butterfly Pavilion and Insect Center, Westminster, Colorado	1995
Detroit Zoo, Royal Oak, Michigan	1995
Bug World, Woodland Park Zoo, Seattle, Washington	1996
Sophia M. Sachs Butterfly House, Chesterfield, Missouri	1998

(*continues*)

TABLE I *(Continued)*

Exhibit	Date of establishment
Chicago Academy of Sciences, Chicago, Illinois	1999
Tropical Butterfly House and Insect Village, Pacific Science Center, Seattle, Washington	1999
Mosanto Insectarium, St. Louis Zoo, St. Louis, Missouri	2000
Puelicher Butterfly Wing Milwaukee Public Museum, Milwaukee, Wisconsin	2000
Magic Wings, South Deerfield, Massachusetts	2000
Butterfly House & Insectarium, North Carolina Museum of Life & Science, Durham, North Carolina	2000
Canada	
L'Insectarium de Montréal, Montréal, Quebec	1990
The Niagara Parks Butterfly Conservatory, Niagara Falls, Ontario	1996
Newfoundland Insectarium, Deer Lake, Newfoundland	1997
Victoria Bug Zoo, Victoria, British Columbia	1997
F. Jean MacLeod Butterfly Gallery, Science North, Sudbury, Ontario	2000
L'Arche des Papillons c.m. Inc., Quebec	2000
Australia	
Butterfly House, Royal Melbourne Zoological Park, Parkville, Victoria	1985
Australian Butterfly Sanctuary, Kuranda, Queensland	1987
Insectarium of Victoria, Woodend	1993
New Zealand	
Butterfly and Orchid Garden, Thames	1999
Latin America	
The Butterfly Farm, La Guacima de Alajuela, Costa Rica	1990
The Butterfly Garden at Shipstern Nature Reserve, Belize	1990
Spirogyra Butterfly Garden, San Jose, Costa Rica	1992
La Selva Butterfly Farm—Primary Forest, Ecuador	1992
Green Hills Butterfly Ranch and Botanical Collection, Cayo District, Belize	1997
Tropical Wings Nature Center, Cayo District, Belize	1998
Africa	
Insectarium de la Réunion, Réunion	1992
Butterfly World, Klapmuts, South Africa	1996
French West Indies	
La Ferme des Papillons, St. Martin, French West Indies	1994

Lepidoptera, typically in a large walk-through immersion exhibit enclosed in glass or screening, such as the Cockerell Butterfly House in Houston, Texas. Some facilities are combinations of both insect zoos and butterfly houses. Year-round walk-through butterfly houses vary in size. The Sophia M. Sachs Butterfly House in Chesterfield, Missouri, covers 511 square meters, whereas the Penang Butterfly House in Penang, Malaysia, is much larger (1474 m^2). Seasonal exhibits are generally smaller. A few facilities, such as the insectarium at the Tama Zoo in Japan, display grasshoppers and aquatic species in walk-through immersion exhibits more typical of butterfly houses.

PURPOSE AND VALUE OF INSECT ZOOS

Live insects are one of the best teaching tools for children and adults alike. Despite an enormous range and intensity of educational, research, and conservation activities, most insectariums and butterfly displays offer some level of educational programming that promotes the appreciation and understanding of insect life. Aside from the actual live exhibits and graphics, these programs take the form of informal presentations, hands-on opportunities to touch live animals, formal classes and lectures from kindergarten to university level, teacher training, field trips, outreach programs, printed educational materials, multimedia materials, Web-based resources, and special events such as insect fairs and film festivals. In addition to these more lofty goals, insects have proven to be enormously popular with the public and thus are used to generate increased visitation and revenue for various types of nonprofit organizations and commercial enterprises.

Basic and applied research is carried out by some facilities, resulting in presentations at conferences and publications in conference proceedings and scientific journals. Several insectariums participate in captive breeding programs for threatened or endangered species such as the American burying beetle, *Nicrophorus americanus,* the Italian ground beetle, *Chrysocarabus olympiae,* the giant wetas, *Deinacrida* spp., and the Polynesian tree snails, *Partula* spp. For example, London Zoo currently manages six invertebrate conservation programs, involving 38 species. *In situ* conservation programs are also supported by a number of insectariums and their parent organizations.

Some of the early insectariums began with the immediate goal of raising food for insectivorous zoo animals such as the Tama Zoo's Insectarium, which opened in its first form in 1961. It was felt that decreases in wild populations of grasshoppers due to widespread use of insecticides required the development

of stable food sources. An earlier insectarium at Tokyo's Toshima-en Amusement Park was developed with the idea of using its live insect residents for making nature films.

DIVERSITY OF EXHIBIT TECHNIQUES

Facilities that house live insects are diverse in construction as well as in the type of sponsoring institution. To face the challenges of exhibiting small, short-lived, seasonally limited, diapausing animals with radically different life stages, major facilities rely heavily on the in-house maintenance of breeding colonies so that specimens are available year-round. Founder stock is collected from the wild or obtained through exchange or purchase from other insectaries to establish and maintain genetically healthy colonies. In contrast, many butterfly houses rely on independently run breeding facilities, often in the species's country of origin. Pupae are generally received weekly from various butterfly ranches or farms. Some butterfly houses also have supplemental in-house breeding colonies of selected species of butterflies. Many for-profit butterfly houses maintain breeding populations both for their own display and to sell to other butterfly houses.

Exhibit techniques have not changed much during the 120 years since the opening of the first insect exhibit at the London Zoo in 1881 (Fig. 1). Even in 1881, informative labels and preserved specimens of insects accompanied the live insect displays. The major change has been the inclusion of interactive techniques and graphics to enhance the interpretive experience of the display. Today, insectariums are really zoo–museum hybrids. Interactive computer modules, microscopes, audio tracks, video loops, models, robotics, and cultural artifacts are now used to enrich and enliven the educational messages. Special displays on topics related to cultural entomology have been included in some facilities such as the Insect Zoo at the San Francisco Zoo, which has produced special exhibits on ancient cricket cages of China, insects as human food, and aquatic insects in terms of fly fishing and fly tying in North America. These techniques combine to address different visitor learning styles, ages, and interests. Some facilities have educational outdoor garden displays to focus attention on native insects and plant interactions, and a few facilities adjoin or are sister organizations to native wildlife reserves.

FIGURE 1 The London Zoo opened the first major insect exhibit in the world in 1881. This contemporary photograph was made by collector William Hornaday. [From "Zoological Gardens Illustrated" (photograph album), Vol. 1, "London Zoological Gardens." © Wildlife Conservation Society, headquartered at the Bronx Zoo.]

Most often, insect exhibit facilities are associated with larger institutions. Zoological gardens such as the London Zoo, Cologne Zoo, Berlin Zoo, Tama Zoo, Cincinnati Zoo, San Francisco Zoo, and St. Louis Zoo all contain major insectariums. General histories of zoos and aquariums, while focused on vertebrates, can be found in the encyclopedic memory and archives of Marvin Jones (San Diego Zoo) and in edited volumes by Kisling, and Hoage and Deiss. Natural history museums such as the Smithsonian National Museum of Natural History, the Natural History Museum of Los Angeles County, and the Houston Museum of Natural Science have insect zoos or butterfly houses; the Cecil B. Day Butterfly Center is located within Callaway Gardens in Pine Mountain, Georgia; universities such as Kansas State University, Michigan State University, the University of Joensuu in Finland, and the University of Alberta in Canada all have seasonal butterfly gardens or insect zoos. The Insectarium de Montréal (Canada), the Insectarium of Victoria (Australia), the Butterfly Pavilion and Insect Center (Colorado), and the Sophia M. Sachs Butterfly House (Missouri) are examples of independent stand-alone facilities. Insectariums vary from tax-supported municipal institutions to nonprofit organizations to for-profit corporations. The for-profit businesses such as the Penang Butterfly House in Malaysia, Stratford-upon-Avon Butterfly Farm in England, Butterfly World in Florida, and the Australian Butterfly Sanctuary in Australia serve as both public exhibits and commercial suppliers. While many are permanent, year-round facilities, the popularity of live insect exhibits (and perhaps the short life span and easy transportability of insects) has allowed for the explosion of temporary or seasonal exhibits, particularly butterfly displays.

INSECTARIUMS AROUND THE WORLD

During the 19th century, expanding empires, increased trade, and improved transportation and communication stimulated interest in exotic wildlife. European powers sent expeditions to bring back specimens for potential domestication and commercial use. Illustrated publications on biological subjects appeared, and books recounting the adventures of naturalist–explorers allowed the public the vicarious thrill of discovery. With the emergence of modern systematics, the number of described genera rose. This was a fertile time of new discoveries and new theories. Darwin,

Wallace, and others pondered the origin of species. From 1828 through 1914, each year an average of two zoological institutions opened throughout the world; the total was 168, with 86 in Europe alone. Themes of metamorphosis and evolution permeated art and literature. The Industrial Revolution transformed commercial production and altered attitudes toward nature as reflected in the diverse philosophies encompassed by the Art Nouveau movement (~1870–1914). Menageries toured, zoological gardens sprang up, and exotic nature became a source of fashionable urban pleasure and scientific study.

Europe

FRANCE Loisel's 1912 history of zoos documents the early development of zoos in general and includes some interesting details about a few insect exhibits. In France, the Jardin des Plantes was created in 1793 adjacent to the Muséum d'Histoire Naturelle and the Ménagerie d'Observation Zoologique. A chair of Insects and Worms was created, and its first occupant was Jean-Baptiste Lamarck. The Jardin's mission was exploration to find species of plants and animals for utility or ornament, with instruction of the public as a minor goal. In 1797 the visitors could view silkworms and stroll through the grounds past honey bees, which were housed in a large glass hexagonal structure. In the middle of the 19th century a disease killed almost all the silkworms in France. In 1860 the Jardin Zoologique d'Acclimatisation was created on the western fringe of Paris to focus on the study of animals of economic utility. New species of silkworm were imported from China and India and kept at the Jardin d'Acclimatisation. Two of these species were acclimatized, and a fertile hybrid was produced. French scientist Louis Pasteur, in 1870, rescued the silk industry by discovering that the then epidemic "pebrine disease" of silkworms, now known to be borne by *Nosema bombycis,* could be prevented through microscopic examination of adult moths and isolation of uninfected stock. These advances set the trend for a more scientific approach to silk production.

In France today, a handful of butterfly houses have appeared as independent facilities. In 2000, a new theme park called Micropolis opened in the town Saint-Léon-en Lévezou, which was home to the famous French entomologist Jean-Henri Fabre.

ENGLAND AND THE NETHERLANDS In 1828 the Zoological Society of London (London Zoo) opened, and visiting it became a popular recreational pastime. Commercial butterfly farms appeared as early as 1865 in Colchester, England. In 1881, a mere 53 years after its debut, the London Zoo opened the first major public insect house. In what had been a refreshment room, a series of terrariums on tables along the wall and in the center of the room displayed live specimens of silkworms, aquatic insects, and other invertebrates. Display tanks were well labeled, and mounted specimens enhanced the exhibit.

Seventeen years after the opening of London's insect zoo, in 1898, the Artis Zoo in Amsterdam opened the Insectarium founded by schoolteacher Rudolph A. Polak, who feared that the urban population was becoming alienated from nature. Polak was part of the "Biologisch Reveil" (reveil means wake-up call) movement—to remind people of their connection to the natural world. His goal was to use insects to bring children into close contact with nature, and this tradition is still carried on today at Artis Zoo. Amazingly, Polak continued to work as a teacher and ran the Insectarium as a hobby.

In the early 1900s, other insect-viewing opportunities existed as well. In 1900 a businessman operated a butterfly farm in Kent, where he bred various species to sell to collectors, museums, and universities in England and America. In 1908 Londoners could pay 6 pence (equivalent to perhaps $10 today) to visit the storefront menagerie display of an enterprising optician, filled with display cases of live bees and ants. The London Zoo, where the formal display of live insects really began, celebrated the new millennium by creating the Web of Life facility in 1999, boasting 65 live displays, 156 invertebrate species, a room-sized enclosure for desert locusts, *Schistocerca gregaria* (including a half-buried jeep for atmosphere), and a giant anteater.

Long involvement with the silk industry and an active amateur naturalist community seemed to make England fertile ground for the birth of modern-day butterfly houses. Businessman David Lowe, who created the Guernsey Butterfly Farm in 1977, went on to create seven other facilities in England by 1984. The London Butterfly House in Syon Park opened in 1981, and Stratford-upon-Avon Butterfly Farm, which followed in 1985, includes an extensive insect zoo exhibit adjoining the free-flight butterfly display. In 1986 there were approximately 40 butterfly houses in England. They number closer to 20 today.

GERMANY AND EASTERN EUROPE The Frankfurt Zoo opened its first insect house in 1904 as a seasonal exhibit in the summer (in the winter it was used for storks). In 1957 Frankfurt Zoo built a new large insectarium on top floor above the renovated aquarium. The Zoologischer Garten Köln (Cologne Zoo) opened an insect house in 1905 and maintained it through 1929. A new insectarium was built in 1971, with a butterfly room at its entrance on the first floor of the new aquarium-terrarium, with approximately 60 species maintained in this facility. The Budapest Zoo in Hungary opened its first insect exhibit in 1907 and renovated the vivarium in 1970, exhibiting 68 species. In 2000, a butterfly exhibit was added to the zoo. In 1913 the Berlin Zoo opened a large insect exhibit on the third floor of the new aquarium, as did the Zoologischer Garten Leipzig. The Berlin Zoo's building, destroyed during World War II, was rebuilt after 1945. The exhibit was renovated again in 1978 through 1983

and maintains over 35 species. An insect section was opened in Gruga Park in Essen in 1956, and in 1970 the Löbbecke Museum in Düsseldorf built a large insect exhibit associated with its aquarium. In 1987 the new Löbbecke Museum and Aquazoo opened an insectarium as the centerpiece of the whole new building. In 1980 the Zoological–Botanic Garden Wilhelma in Stuttgart built an insectarium exhibiting 35 species. The Dortmund Zoo and the Krefeld Zoo built butterfly exhibits in 1991 and 1998, respectively. The Noorder Zoo in the Netherlands, the Zoologicka Zaharada Praha in Czechoslovakia, and the Tiergarten Schönbrunn in Austria also have insect zoo exhibits.

Asia

Although no records have been found to document the early exhibition of live insects in China, the Chinese have long had a complex appreciation of the insect world. The development of silkworm culture and silk production, the development of bee culture, the early and extensive use of insects in traditional medicine, and the use of crickets as pets in the Táng dynasty (618–906) suggests an advanced appreciation of the utility and aesthetics of insect life.

Yajima describes the development of insectariums in Japan. The Insectarium at the Takarazuka Zoological and Botanical Garden opened in 1954 and was expanded in 1967. The Toshima-en Insectarium, at the Toshima-en Amusement Park in Tokyo, opened in the late 1950s. In 1961, Yajima went on to design the first incarnation of the Insectarium at the Tama Zoo. In 1966 Tama Zoo's Insectarium opened a walk-through grasshopper greenhouse filled with 6000 grasshoppers in addition to a walk-through butterfly house. A firefly building was added in 1975. In 1988 a new insectarium was built at Tama Zoo, which mixed traditional terrarium-type exhibits with walk-through exhibits without guardrails of Orthoptera, fireflies, butterflies, beetles, ants, and aquatic insects. Insect Ecological Land had a construction cost of over $5 million and increased attendance to the zoo by 20%. At its opening, 94 species were maintained by a staff of 12. As of 1995, there were 30 live insect exhibits in Japan alone.

A few other insectariums and butterfly houses can be found throughout Asia. The Penang Butterfly House in Malaysia was founded in 1986. Attached to the butterfly display is an exhibit of Malaysian insects and arthropods. The Fragile Forest exhibit, which opened at the Singapore Zoo in 1998, displays butterflies and other insects.

North America

Insectariums developed more slowly in the United States. Though the early entomologist Thomas Say contributed to bringing respect to the science of entomology in the early part of the 19th century, the zoological gardens in the United States remained focused on the display of vertebrates and the natural history museums focused on nonliving exhibits. Interest in representing invertebrate biodiversity either taxonomically or zoogeographically in these institutions lagged behind that shown by their European counterparts. Enduring insectariums did not take hold in America until the late 1970s.

However, a number of early successful experiments at both zoos and museums proved extremely popular with the public. The New York Zoological Society Bulletin chronicles several early attempts. Raymond L. Ditmars, Curator of Reptiles at the Bronx Zoo and world-renowned herpetologist, actually began his career in science as an entomologist working as an assistant curator in the Entomology Department at the American Museum of Natural History. Throughout his tenure, he maintained a keen interest in insects. In 1910, under his direction, a live arthropod exhibit consisting of 56 cages and containing silk moths in various life stages, lubber grasshoppers, Hercules beetles, and tarantulas was set up as an experiment. A portion of the moths that emerged from the silk moth collections were mounted and sold as souvenirs by the Bureau of Information in the Lion House. The exhibit was so popular with visitors that plans were made to make it a permanent feature at the zoo. The permanent facility was never built, but Ditmars persevered and in 1940, a Department of Insects was created with Ditmars as curator. Ditmars died in 1942, however, without having brought the plan to fruition. Brayton Eddy, an entomologist for the state of Rhode Island, was hired as the new curator of insects in 1945. Eddy's experience included the development of a seasonal live insect zoo at Goddard State Park in Providence, housed in the first floor of the stately mansion on the property from 1934 to 1937. Then Eddy died unexpectedly in 1950, and again a planned permanent insect exhibit never got started. The new exhibit was to have featured local insects and tropical imports sent by Dr. William Beebe from his research station in Venezuela.

During the 1930s and 1940s, the Chicago Zoo (later renamed the Brookfield Zoo) was planning and experimenting with its own invertebrate exhibit. Construction on the Insect House (also called the Invertebrate House or the Special Exhibit and Demonstration Building) was completed between 1934 and 1938. Grace Olive Riley, acting as curator of reptiles and invertebrates, was succeeded by Robert Snediger as reptile curator. Snediger organized the "Animals Without Backbones" exhibit, which ran from 1947 into the early 1950s. Bees, cockroaches, aquatic insects, spiders, scorpions, centipedes, leeches, and other invertebrates were displayed, along with amoebas. Though planned as a permanent feature by Brookfield Zoo's early designers, the Insect Building was later converted into the zoo's library. The exhibits developed in the late 19th and early 20th centuries were not very different from the exhibits of today. Glass terrariums with screened lids displayed on tables containing local and exotic insect life were the state of the art in 1881, 1910, the 1940s, 1970s and 1990s.

Worth mentioning is the arthropod exhibit that was included as a permanent feature in the Arizona–Sonoran

Desert Museum from its opening date in 1952. Although not large, either in species number or in size of tanks, this collection was innovative in that it included invertebrates in the interpretation of an ecosystem. Then 20 years passed during which the early invertebrate exhibits seem to have vanished from the consciousness of zoo administrators.

Another period of experimentation with short-lived insect exhibitions emerged. An insect zoo was created at the Flushing Meadows Zoo from 1969 to 1970, and another was assembled at the Smithsonian National Museum of Natural History in 1971. The immense popularity of the exhibit at the Smithsonian resulted in the opening of the museum's permanent insect zoo in 1976, in time to greet the International Congress of Entomology, which held its meeting that year in Washington, D.C. A natural history museum had now stepped permanently into the world of live insect display, an endeavor formerly restricted to zoological gardens or commercial rearing facilities staffed by persons used to dealing with the challenges and demands of live animals. From the start, the Smithsonian staff interacted with visitors daily with informal hands-on demonstrations, and the species inventory included both local and exotic specimens collected by Smithsonian entomologists in the field.

Cincinnati Zoo followed the Smithsonian in 1978 and opened a new building containing 68 live displays and housing 70 to 100 species. This facility, the World of the Insect, set a new standard for insectariums in the United States. It also set a new standard for monetary commitment to insect exhibits, with a price tag of one million dollars, including the cost of the new stand-alone building, custom-made terrariums, colorful graphics, and educational interactive exhibits borrowed from museum methodology.

The Insect Zoo at the San Francisco Zoo opened in 1979 as a temporary summer exhibit in an unused auditorium building in the children's zoo section. Zoo attendance instantly increased by 50%, and the collection became a permanent facility, the second largest in the United States for two decades, containing 35 displays and maintaining 70 species.

In the United States, these three permanent insectariums created in the 1970s (at the Smithsonian, the Cincinnati Zoo, and the San Francisco Zoo) were really "arthropod zoos," with an occasional mollusk, leech, or annelid thrown in. These exhibits became models for many subsequent projects that followed throughout the country. Emphasis was placed on the development of year-round breeding colonies and use of native and exotic species.

The only other invertebrate exhibit to open in the following decade excluding the butterfly houses, was the Invertebrate Exhibit, which opened in 1987 on the lower floor of the Reptile House of the National Zoo, showing tropical insects alongside marine invertebrates such as cephalopods. A pollinarium exhibit was added in 1996.

The only nonbutterfly insect exhibit to emerge in the 1990s was the Ralph K. Parsons Insect Zoo at the Natural History Museum of Los Angeles County. Modest at its opening in 1992, it has grown to 40 live displays.

New walk-through butterfly exhibits dominated the late 1980s and early 1990s modeled on free-flight greenhouse exhibits in England. In 1988 the $3 million Cecil B. Day Butterfly Center at Callaway Gardens opened in Georgia. Two for-profit Florida endeavors, Butterfly World and the butterfly house at Cypress Gardens, opened in 1988 and 1992, respectively. Back in the zoo world, the San Diego Wild Animal Park's Butterfly Encounter debuted in 1993. In 1994, the Houston Museum of Natural Science opened the Cockerell Butterfly Center, a three-story immersion butterfly glasshouse attached to the museum, attracting 700,000 visitors during its first year of operation. A live arthropod exhibit room was later added to this facility. In 1995, the Butterfly Pavilion and Insect Center, became the first stand-alone nonprofit insect facility in the United States.

In the 1990s at least 20 seasonal butterfly houses emerged, including several supported by universities such as Michigan State University and Kansas State University under the auspices of their respective Departments of Entomology. The insectarium at the St. Louis Zoo opened in 2000. Costing $4 million and maintaining 80 to 100 species in addition to a butterfly display, it became the most significant facility to open in over a decade.

Parallel developments have occurred in Canada, but the most notable was the development of the Insectarium de Montréal founded by Georges Brossard in 1990. This $8 million exhibit mixed classic European museum design with multimedia interactive displays and challenged the live animal exhibit community to set a higher standard for the construction of new insectariums to educate the public about insects, which represent 80 to 95% of the animal species on terrestrial earth.

Australia

In Australia, the Melbourne Zoo was the first to get into the invertebrate business by opening a butterfly house in 1985. The Insectarium of Victoria and the Victorian Institute of Invertebrate Sciences opened at Heathcote in 1993, moving in 1998 to a permanent home in Woodend. On display is *Astacopsis gouldi,* or the giant yabby or crayfish, one of the largest freshwater invertebrates on earth.

Latin America and Africa

Butterfly exhibit houses in Latin America are most commonly derived from commercial butterfly ranching and farming operations. The Butterfly Farm in La Guacima de Alajuela, Costa Rica, founded in 1990, and Spirogyra Butterfly Garden in San Jose, Costa Rica, established in 1992, typify these facilities. The Green Hills Butterfly Ranch in Belmopan, Belize, is a display as well as commercial supplier, as is the La Selva Butterfly Farm in Ecuador. Butterfly World in Klapmuts, South Africa, opened in 1996.

BIODIVERSITY CONSERVATION AND INSECT EXHIBITS

Insectariums can play a strong role in conservation education. They can also play a role in the conservation of biodiversity in nature by providing economic opportunities for local populations that are alternatives to destructive resource extraction such as logging, mining, or conversion of forest land to agricultural enterprises such as cocoa, coffee, or oil palm plantations. It is difficult if not impossible to conserve natural resources where alternative economic opportunities are limited or absent. In response, as general awareness and concern for loss of tropical forest ecosystems and biodiversity emerged, the government of Papua New Guinea deemed insects a national resource and candidate for economic development. This policy resulted in the establishment of the Insect Farming and Trading Agency (IFTA) in 1978, to create income-producing opportunities for villages through nondestructive extraction of forest resources while at the same time creating incentive for preserving forest habitat. In 1983 a report published by an advisory committee of the National Research Council, in cooperation with the IFTA, promoted the idea of butterfly ranching and farming to supply research scientists, butterfly collectors, and other commercial uses. The report estimated that the current trade was between $10 and 20 million annually. Gram for gram, butterflies became more valuable than cattle. In 1981 it was estimated that an industrious butterfly farmer could earn from $100 to $3000 a year versus the mean per capita of $50. The IFTA sells $400,000 worth of stock annually and provides income for 1500 villagers in Papua New Guinea.

The worldwide explosion of public walk-through butterfly houses in the 1980s and 1990s created a new market for butterfly ranching and farming projects. Nairobi University scientists, working with the East African Natural History Society and the National Museums of Kenya, began the Kipepeo Project, (*kipepeo* is Swahili for butterfly) in 1993 to protect the Arabuko–Sokoke Forest in Kenya. A United Nations grant was awarded to develop sustainable utilization of butterfly biodiversity for the benefit of surrounding rural communities. In addition, a butterfly house was established as an ecotourism attraction to diversify the coastal tourism industry and to promote conservation of the Arabuko–Sokoke Forest. Several butterfly ranching projects have been created in Costa Rica to promote the conservation of remaining remnants of forest habitat in that country. These projects rely on the purchase of pupae by live butterfly exhibits. Organizations such as the U.S. Agency for International Development, The Nature Conservancy, Conservation International, and the World Wildlife Fund have now participated in the development of similar projects in Central America, and in Irian Jaya and Sulawesi in Indonesia.

Papiliorama-Nocturama Tropical Gardens in Neuchâtel, Switzerland, an exhibit and nonprofit organization that opened in 1988, invests income into a sister foundation, the International Tropical Conservation Foundation, which runs the 8.9 ha Shipstern Nature Reserve in Belize. Shipstern also has its own public butterfly house on the reserve grounds.

Approaches to conservation have shifted in the past decade from focusing on single-species protection to ecosystem preservation. In 1987, for example, the Center for Ecosystem Survival, a nonprofit consortium serving zoos, aquariums, museums, and botanical gardens, was created in San Francisco to raise funds for biodiversity conservation through habitat protection, conserving invertebrates and plants as well as vertebrates and other forms of life. In 2001, a total of 112 informal science institutions participated in this program and raised over $2 million for habitat purchase and protection, fueled in part by increased public awareness of the magnitude and importance of insect biodiversity.

See Also the Following Articles

Conservation • Museums and Display Collections • Photography • Rearing of Insects • Teaching Resources

Further Reading

Chicago Zoological Society. (1947). "Guide Book, Chicago Zoological Park." Brookfield, IL.

Hoage, R., and Deiss, W. A. (eds.), (1996). "New Worlds, New Animals." Johns Hopkins University Press, Baltimore.

Kisling, V. N. (ed.). (2001). "Zoo and Aquarium History—Ancient Animal Collections to Zoological Gardens." CRC Press, Boca Raton, FL.

Loisel, G. (1912). "Epoque contemporaine," Vol III in "Histoire des ménageries de l'antiquité à nos jours." Henri Laurens/Octave Doin et fils, Paris.

New York Zoological Society. (1911). "Zoological Park Notes," Zoological Society Bulletin 47, pp. 788–789.

New York Zoological Society. (1940). "Notes from the Park and Aquarium," Zoological Society Bulletin 43(6), pp. 198–203.

Parsons, M. (1992). Butterfly farming and conservation in the Indo-Australian region. *Trop. Lepid.* **3,** 1–28.

Saul-Gershenz, L. S., Arnold, R. J., and Scriber, J. M. (1995). Design of captive environments for endangered invertebrates. *In* "Conservation of Endangered Species in Captivity: An Interdisciplinary Approach," State University of New York Press (E. F. Gibbons, B. S. Durant, and J. Demarest, eds.). Albany.

Yajima, M. (1996). The development of insectariums and their future. *Int. Zoo News* **43,** 484–491.

Integrated Pest Management

Ronald Prokopy
University of Massachusetts

Marcos Kogan
Oregon State University

Across millennia, humans have used a variety of approaches in attempts to maintain pest insects at tolerable levels. The character of these approaches has evolved over time. Since

the late 1960s, an approach termed integrated pest management (IPM) has ascended to dominance internationally. In essence, IPM is a decision-based process involving coordinated use of multiple tactics for optimizing the control of all classes of pests (arthropods, microbial pathogens, vertebrates, weeds) in an ecologically and economically sound manner.

HOW INSECT PESTS ORIGINATE

The manner by which an insect has become a pest influences the strategy used for managing it within an IPM framework. Some species have reached the status of pest because human thresholds for tolerating them have decreased as economic well-being has increased. Other species have become pests as a consequence of the availability of abundant resources for their survival and reproduction; resource concentration in time and space is conducive to rapid buildup of insect populations. Still other species have become pests in regions of their origin when they expanded their host range from comparatively unimportant native hosts to economically important introduced hosts. Finally, some species have become pests almost instantly upon being transported by humans to new locales having favorable resources but devoid of effective natural enemies.

CLASSIFICATION OF INSECT PESTS

For purposes of devising an appropriate IPM strategy, insect pests may be classified as key, secondary, or induced. Key pests are those whose populations, if unmanaged, repeatedly exceed tolerable levels; these are the principal focus of IPM endeavors. Secondary pests are those whose populations occasionally reach intolerable levels; their potential threat is recognized by the very act of devising IPM strategies. Induced pests are those whose populations rarely exceed tolerable levels under natural conditions, but if they become resistant to pesticides or other single-tactic control measures that harm their natural enemies, they can reach outbreak proportions. Usually induced pests return to nonpest status under a true IPM approach.

EVOLUTION OF CONCEPT OF IPM

Through trial and error across centuries, humans gradually came to use tactics such as cultural control, host resistance, and biological control in efforts to protect themselves, livestock, crops, and forests against pests. By the latter half of the 19th century, some pest control strategies that blended these tactics could rightfully be considered to be precursors of modern IPM strategies. During the 20th century, efforts to maintain pests at tolerable levels became more formalized as they became more intensive; these efforts can be considered as having progressed along the following four pathways.

"Pest control" was the terminology used during the first half of the 20th century to describe the set of actions taken to avoid, attenuate, or delay the impact of pests. Early in the century, inorganic and botanical insecticides gained increasing prominence as a control tactic against pest insects. By mid-century, use of organosynthetic insecticides supplanted virtually all other tactics in becoming the dominant approach to insect pest control, especially in developed countries.

"Integrated control" surfaced about 1950 in California as a concept of pest control aimed at combining and integrating biological and chemical control. The emergence and eventual popularity of this concept had roots in problems associated with overuse of organosynthetic insecticides, particularly a surge of key, secondary, and induced insecticide-resistant pests and environmental harm caused by insecticides. The publication of *Silent Spring* by Rachael Carson in 1962 was especially important in spurring widespread interest in the principles of integrated control.

"Pest management" represents a shortened version of "protective population management," a concept coined in 1964 by Australian ecologists to emphasize direct human interference in maintaining pests at tolerable levels, as opposed to reliance on unmanaged natural abiotic and biotic factors acting on pest populations. The concept of pest management embraced a broader range of pest control tactics than did the concept of integrated control. Until the late 1960s, both these concepts flourished simultaneously in accenting the need to move beyond use of insecticides as a sole pest control tactic.

"Integrated pest management" originated in 1968 as a contraction of "integrated pest population management," an expression used first in 1967 by R. F. Smith and R. van den Bosch. Soon afterward, the abbreviation "IPM" came into use worldwide for signifying a desirable and holistic approach to controlling pests. Numerous definitions have been put forward for IPM, but one that captures the broad and essential elements of many others is the following: IPM is a decision support system for the selection and use of pest control tactics, singly or harmoniously coordinated into a management strategy, based on cost–benefit analyses that take into account the interests of and impacts on producers, society, and the environment. No single definition, however, is likely to encompass all facets of IPM for all time. IPM has been and likely will continue to be an evolving concept. For example, a compendium of IPM definitions, available on the World Wide Web (http://www.ippc.orst.edu/IPMdefinitions/define.html?), listed 67 definitions in mid 2002.

ECOLOGICAL FOUNDATION OF IPM

Ecology is the study of relationships among organisms and their environment. Consideration of these relationships usually begins with focus on individuals or populations of a single species and subsequently broadens to include communities of organisms and, eventually, ecosystems.

Conceptually, the foundation of IPM is ecological. Ideal IPM programs are those that fully embrace ecosystem structure and processes in time and space. In reality,

ecological complexity increases dramatically with each step from population to community to ecosystem. Such complexity challenges the realization of ideal IPM programs.

One approach that has been taken toward recognition of the ecological foundation of IPM, especially in developing countries, accentuates the pest-suppressing properties inherent in natural ecosystems as primary building blocks for the construct of human-designed ecosystems. After system construction based on ecological principles, the intent is to minimize human intervention to the greatest extent possible while still maintaining pest populations within tolerable levels.

Another approach, common in developed countries, takes as its starting point an existing ecosystem constructed by humans and aims at reducing negative impact in a succession of steps used in the management of pests. Such an approach usually commences with a focus on the ecology of a single-species population and may expand to consideration of structures and processes associated with communities and ecosystems.

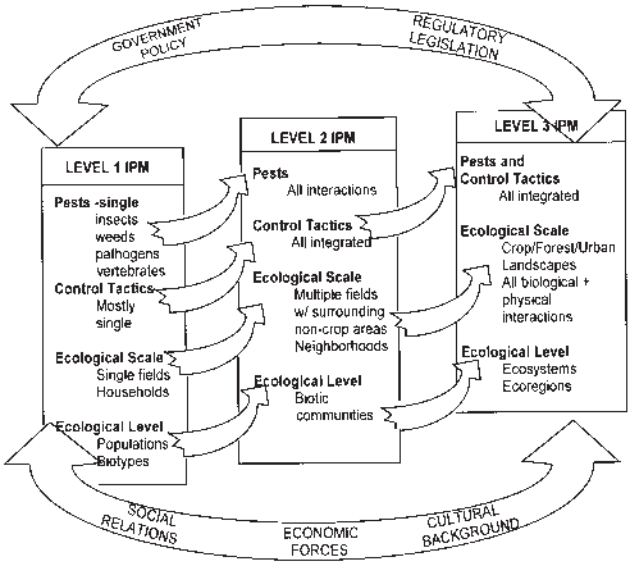

FIGURE 1 Levels of IPM integration: main targets, ecological scales, and levels of ecological complexity.

ADVANCING LEVELS OF INTEGRATION IN IPM IMPLEMENTATION

A hierarchical structure results from the ecological foundation of IPM that lends itself to viewing IPM as progressing in scope through a series of ecologically rooted steps. These steps are characterized by ascending levels of complexity and spatial scale: from focus on a single-species population in a restricted locale to focus on a community of pest and other organisms in a larger area to focus on a whole ecosystem. Coincident with this ecocentric hierarchy is another stepwise hierarchy conceived of as a vehicle for measuring progress in achieving the goals of IPM. This hierarchy comprises a succession of levels from single-tactic (almost invariably based on pesticide use) to multitactic measures of pest and habitat management. The steps further involve ascending from focus on a single pest species in a single class of pests (e.g., insects) to multiple pest species across all classes of pests (insects, microbial pathogens, vertebrates, and weeds). These distinctive hierarchies can be blended in the form of a continuum of advancing levels of integration in IPM implementation. Three more or less distinctive levels along the continuum are described here (Fig. 1). Degree of success in integration at each level is shaped not only by ecological processes but also by government policy, regulatory legislation, social relations, economic forces, and cultural background that may enhance or constrain progress.

FIRST-LEVEL IPM

In the most basic and also the most widely practiced form of IPM, emphasis is on monitoring development and/or abundance of a single pest species at a single locale (e.g., a household, cow barn, greenhouse, cropped field, or woodlot) and using thresholds for deciding whether to take action. Application of a pesticide is by far the most common form of action taken under first-level IPM. Integration occurs when abundance of natural enemies of the pest in question also is considered in the decision-making process and when selection among candidate pesticides involves explicit attention to minimizing harm to these and other beneficial organisms. This form of IPM has been characterized by some as "integrated pesticide management."

Monitoring Pest Development

Because the developmental rate of an arthropod is regulated largely by temperature, the monitoring of developmental rate for pest management purposes usually takes the form of measuring accumulation of heat units above a threshold temperature at which development begins. At temperatures above these fostering the maximal developmental rate, development may decrease. Such decrease has not been investigated for most pest arthropods and has not yet played a significant role in making pest management decisions.

The simplest and most prevalent approach to measuring accumulation of heat units above developmental threshold temperature involves use of degree-days (DD). For a specific date, the number of accumulated DD equals the average temperature of that date minus the developmental threshold temperature of the arthropod. Several procedures have been devised to estimate average daily temperature. The most common one, albeit somewhat crude, consists simply of averaging the maximum and the minimum ambient temperature of the day. To illustrate the DD approach, if the high and low temperature for a given day were 30 and 20°C, respectively, with a developmental threshold temperature of 10°C, then 15 DD would have accumulated on that day.

Pest development as monitored by DD accumulation may benefit decision making under first-level IPM in several ways, particularly for optimal timing of management activities. For example, ability to predict when a majority of pest adults is about to emerge from pupae is useful for optimal timing of deployment of traps for monitoring adults. Knowledge about when oviposition is likely to begin and peak can facilitate optimal timing of pesticide application against newly hatched larvae, which often is the stage most vulnerable to pesticide treatment. Sometimes this determination is made in conjunction with date of first capture of adults by traps, known as a "biofix" point for initiation of DD accumulation. The ability to forecast when a majority of larvae or nymphs is at a particular growth stage can aid in optimal timing of sampling their abundance and the abundance of their natural enemies.

Monitoring Pest Abundance

Ideally, an IPM practitioner would have available a precise count of the number of individuals of an insect pest species present in an area of concern; realistically, obtaining information on absolute densities of pests is prohibitively costly. Therefore, most practitioners rely on imprecise estimates of pest population density obtained by using one or more population sampling techniques. The intent is to capture a more or less consistent, if unknown, proportion of the pest population. Choice of appropriate sampling technique varies considerably according to pest species and developmental stage.

For sampling comparatively mobile individuals such as adults, traps using odor and/or visual stimuli are common tools. Odor stimuli usually consist of synthetic equivalents of either attractive sex odors (sex or aggregating pheromones) or attractive food or host odors. Visual stimuli normally rely on synthetic mimics of visually attractive sites where feeding, mating, or egg laying occurs.

For sampling less mobile individuals such as larvae, common techniques include visual searching of the target area accompanied by direct counts of detected pests, use of a sweep net (especially effective for sampling individuals on foliage of nonwoody plants), and use of a loose or framed cloth placed beneath vegetation that is shaken or tapped to dislodge pests. Sampling immobile individuals such as eggs or pupae usually is done by visual inspection.

To obtain an acceptably accurate and cost-effective estimate of the size of a pest population by means of one of these techniques, careful attention must be given to the program under which sampling is conducted. Effective sampling programs take into account the daily activity pattern of the target species as well as its characteristic spatial distribution (uniform, random, or clumped). Historically, most programs have incorporated sampling at several or numerous sites in a target area to acquire sufficient representation of the size of a pest population; then researchers have counted the sampled individuals of the target pest. New programs developed for some pests simplify these procedures. Sequential sampling is an approach that optimizes the number of sampling sites needed for classifying a pest population as below or above a density requiring action. Binomial sampling is an approach that classifies an individual species as either present or absent at a sampling site, thereby precluding the need to count all members of that species taken in a sample. Both these simplifying approaches require substantial species-specific background information for their development and use.

The emerging technologies of global positioning systems (GPS) and geographical information systems (GIS) offer unsurpassed capability of aiding in the mapping of site-specific variation in characteristics of areas under consideration for sampling. A GPS uses triangulation of signals from a constellation of satellites to identify the precise location (within a meter) of an area on the earth's surface. A GIS is a computer program for the mapping and spatial analysis of georeferenced information. GIS capabilities include assemblage, storage, manipulation, retrieval, and graphic display of information about attributes of precise locations identified through GPS. Such information can be exceptionally useful in forming associations between characteristics of a specific locale (e.g., terrain, soil, extent of vegetative growth, microclimate) and density of a population (Fig. 2). For pests, sampling can be directed toward specific sites in which densities are suspected to be highest.

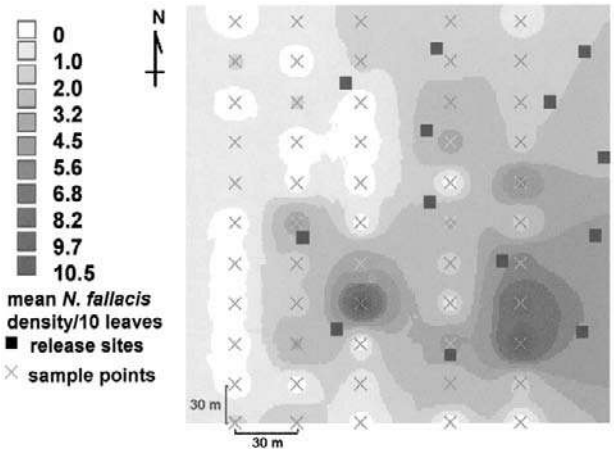

FIGURE 2 Dispersal of *Neoseiulus fallacis* for biological control of spider mites in a strawberry field, 8 to 15 weeks following release of 100 adult females at each of 15 sites: squares, release sites; crosses, sample points. This distribution is due to ambulatory foliar movement and aerial dispersal (dominant winds from south and southwest). Data represented using GIS (GRASS v. 4.1). [From Coop, L. B., and Croft, B. A. (1995). *Neoseiulus fallacis:* Dispersal and biological conrol of *Tetranychus urticae* following minimal inoculation into a strawberry field. *Exp. Appl. Acarol.* **19,** 31–43. Reproduced by permission from the authors and Chapman & Hall (now Kluwer Academic Publishers).]

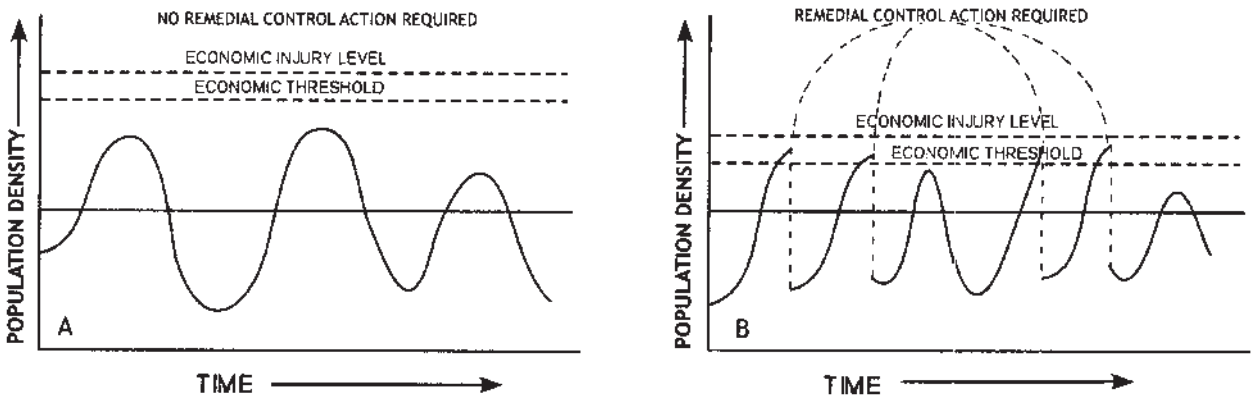

FIGURE 3 Graphs depicting theoretical population fluctuations of two insect pest species.

Deciding Whether to Take Action

Several approaches have been developed for deciding whether an insect pest population has or has not reached a level requiring intervention, such as an insecticide application. For agricultural purposes, the approach used most often centers on the concept of "economic injury level" (EIL), formalized in 1959 by V. H. Stern, R. F. Smith, R. van den Bosch, and K. S. Hagen and defined by them as the "lowest pest population density that will cause economic damage." These entomologists also proposed a related concept, which they termed the "economic threshold" (ET), defined as the "pest density at which control measures should be applied to prevent an increasing pest population from reaching the economic injury level" (Fig. 3).

The decision-making concepts of EIL and ET have been fundamental to the development and implementation of first-level IPM, particularly for insect management in agriculture (but less pronouncedly for disease, vertebrate, and weed management). They have been especially useful when insect pest populations are expected to increase over time within a crop, can be sampled reliably, can be related in a predictable way to reduction in crop yield or quality, and can be controlled readily by taking immediate action (e.g., application of insecticide) to prevent further damage. They are less valuable when human comfort or aesthetics, rather than economic damage, is paramount. Even for agriculture, the concepts of EIL and ET cannot be applied rigidly because of inherent unpredictability of such factors as future weather (which can markedly affect rate of pest population growth and degree of crop susceptibility to a pest) and future value of the crop in the marketplace. Also, a type of action that may require considerable time before reducing pest density, such as application of a biocontrol technique, is likely to be less appropriate than an insecticide application within an EIL/ET framework.

A refinement of the concept of EIL, put forward by L. P. Pedigo and L. G. Higley, introduces the element of environmental quality into the decision-making process. Negative effects of insecticides on natural enemies of pests and on other organisms in the environment are treated as costs in addition to monetary costs associated with insecticide application. Quantifying environmental costs has proven to be challenging and subject to much debate, but progress under this refined concept of EIL nonetheless has been made.

In 1984 W. L. Sterling advanced the concept of "inaction level," which is the density of natural enemies sufficient to maintain a pest below the EIL. M. P. Hoffmann and collaborators developed a sampling program for eggs of tomato fruitworm, *Helicoverpa zea,* that permits ready identification and quantification of parasitized eggs in addition to healthy eggs. If the level of egg parasitism is determined to be too low to prevent larval numbers from exceeding the EIL, the lowest adequate rate of a "soft" pesticide (one having least impact on parasitoids and other nontarget organisms) is recommended. This approach is an example of application of the inaction level concept and represents a high degree of first-level IPM implementation.

Area-wide IPM is an expansion of first-level IPM that may represent a significant transitional step toward second and third levels of IPM. Under area-wide IPM, the key pest of a crop is targeted for management by means of the most effective noninsecticidal approach. For example, the codling moth, *Cydia pomonella,* is managed in apple and pear orchards by tactics such as the pheromone mating disruption technique or the sterile insect release method, that impair normal reproduction. Such tactics are implemented over areas large enough to preclude recolonization by fertile females from adjacent areas. By reducing the impact of broad-spectrum insecticides, natural enemies are preserved and are usually capable of regulating most secondary pests in the crop. As area-wide IPM programs expand to incorporate multiple pest interactions, they become natural springboards to higher level integration in IPM systems.

SECOND-LEVEL IPM

Second level IPM is intermediate between basic and advanced. It is receiving increased research attention, but inherent complexities have greatly limited its effectiveness.

Emphasis is on management of key pests of all classes and their associated natural enemies comprising a community (e.g., a village of dwellings, an entire farm, a wooded area surrounding a village). Emphasis also is on substituting a variety of comparatively environmentally benign management tactics (e.g., cultural management, host resistance, biological control, behavioral control), to the greatest extent possible, for the therapeutic practice of pesticide application. Decision makers must determine how best to integrate these tactics to achieve long-term suppression of pests within a cost–benefit framework.

Cultural management is purposeful manipulation of the environment to reduce pest abundance. It is most effective when directed at the most vulnerable life stage of a pest. Four forms of habitat or environmental manipulation aimed at controlling pests have been practiced for centuries in agriculture: crop rotation, timing of planting or harvest to minimize pest damage, sanitation or elimination of noneconomic resources available for pest reproduction, and polyculture, or the interplanting of different crops to diffuse resource concentration. More recent practices include the planting or encouragement of selected types of noneconomic vegetation in the vicinity of crops or as cover crops to serve as harborage for natural enemies of pest insects. Analogues of these practices have been developed for managing insect pests in nonagricultural situations.

Host resistance is any inherited characteristic of a host that lessens effects of an attacking pest. For centuries, humans may have unknowingly or intentionally selected for cultivars of plants or breeds of animals that are best able to withstand pests. Modern breeding programs, however, often have placed more emphasis on increasing yields than on protection against pests. Entomologically, resistant traits are preadaptive characteristics of a host that reduce its detectability, acceptance, or nutritional value, or enhance its toxicity to a pest insect. Molecular genetics techniques that facilitate introduction of specific pest resistance genes into cultivars or breeds possessing desirable commercial or aesthetic traits are beginning to replace traditional resistance breeding approaches.

In biological control, parasitoids, predators, or pathogens are deployed as natural enemies in the reduction of pest populations. Of the myriad insect species that could become pests, most do not because they are suppressed effectively by naturally occurring populations of biological control agents. Natural levels of biocontrol, however, often are insufficient for IPM purposes. Biological control then takes the form of importing absent natural enemies from other locales (termed importation or classical biocontrol), augmenting existing natural enemies by rearing and then releasing substantial numbers into the target community (termed augmentation), or tailoring management tactics to reduce negative effects on existing natural enemy populations (termed conservation). The latter is the most widely practiced form of biocontrol.

Behavioral control is manipulation of the behavior of pest individuals to prevent them from causing harm or unpleasantness. Because of the expense and technological challenges associated with its use, behavioral control usually is directed only at key pests. Behavioral control may involve use of natural or synthetic chemical or physical stimuli to lure pests to sites where they are killed, or use of such stimuli to disrupt the ability of pests to find or use a potential resource. An ideal form of behavioral control might involve joint use of disruptive and attractive stimuli to achieve maximum effect, but this form is not yet widespread in practice.

Vineyards in parts of Europe and North America represent one of the few areas in which pest management is practiced effectively under the second-level IPM concept. Besides using essential elements of first-level IPM for insect pests, certain practitioners of vineyard IPM in these locations blend host plant resistance with cultural, biological, and behavioral controls for suppression of key pest insects and also use a suite of cultural controls for managing key disease and weed pests. This approach has resulted in marked reduction in pesticide use and greater stability of relationships among organisms comprising vineyard communities.

THIRD-LEVEL IPM

Although many IPM practitioners aspire to implement third-level IPM, the most advanced form remains largely in an embryonic state of development. Emphasis is on using principles and practices of second-level IPM in harmony with all other elements that affect long-term productivity or well-being of an ecosystem. Such elements include sound horticultural or husbandry practices (for agriculture), sound forest management (for silviculture), and sound community health practices (for villages or subunits of cities). Third-level IPM features attention to environmental and societal costs and benefits in the making of pest management decisions. The focal ecosystem may be an entity no larger than a community, as considered under second-level IPM, or it could be an entity as extensive as an ecological region. For crops, third-level IPM is roughly synonymous with the concept of integrated crop management. It is not, however, synonymous with organic agriculture, which disallows some materials and practices acceptable under third-level IPM.

Spearheaded by P. E. Kenmore, an approach has been developed for growing rice in developing parts of Asia that reflects many of the tenets of third-level IPM. This approach acccentuates societal contribution to the IPM decision-making process. It involves weekly gatherings of small groups of rice farmers, accompanied by experienced pest management personnel, who jointly conduct agroecosystem observations, engage in data analysis, and consider local ecosystem structure, environmental health, and a range of immediate and long-term tactics before making pest management decisions. This process has resulted in dramatic increases in awareness by entire communities of an advanced form of IPM and often dramatic decreases in use of pesticides on rice. It stands in contrast to modes of decision making

and levels of popular awareness characteristic of less advanced forms of IPM implementation in many developed countries, where it is commonplace for farmers to make IPM decisions either acting alone or at most interacting with a private consultant, government extension representative, or employee of a pesticide distributor.

IPM AND SUSTAINABLE DEVELOPMENT

In a broad sense, sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs. Sustainable development is rooted in the concept of ecosystem integrity and permeates all facets of human endeavor, whether economic, social, or cultural.

The scope of concerns and practices of third-level IPM as described here corresponds closely to that of sustainable development. Each emphasizes preservation of processes associated with natural ecosystems, long-term well-being of humans as members of communities, economic viability, and deployment of exogenous resources only after careful consideration. For agriculture, concepts underlying third-level IPM can be equated with concepts underlying sustainable agriculture. For both, one can expect concepts to evolve further over time.

See Also the Following Articles

Agricultural Entomology • Biological Control • Extension Entomology • Insecticides • Physical Control • Population Ecology • Sterile Insect Technique

Further Reading

Altieri, M. A., and Nicholls, C. I. (1999). Biodiversity, ecosystem function, and insect pest management in agricultural systems. *In* "Biodiversity in Agroecosystems" (W. W. Collins and C. O. Qualset, eds.), pp. 69–83. CRC Press, Boca Raton, FL.

Benbrook, C. M. (1996). "Pest Management at the Crossroads." Consumers Union, Yonkers, NY.

Coop, L. B., and Croft, B. A. (1995). *Neoseiulus fallacis:* Dispersal and biological control of *Tetranychus urticae* following minimal inoculations into a strawberry field. *Exp. Appl. Acarol.* **19,** 31–43.

Dent, D. (1995). "Integrated Pest Management." Chapman & Hall, New York.

Higley, L. G., and Pedigo, L. P. (1993). Economic injury level concepts and their use in sustaining environmental quality. *Agric. Ecosys. Environ.* **46,** 233–243.

Kenmore, P. E. (1996). Integrated rice pest management. *In* "Biotechnology and Integrated Pest Management" (G. J. Persley, ed.), pp. 76–97. CAB International, Wallingford, Oxon, U.K.

Kennedy, G. G., and Sutton, T. B. (2000). "Emerging Technologies for Integrated Pest Management." APS Press, St. Paul, MN.

Kogan, M. (1998). Integrated pest management: historical perspectives and contemporary developments. *Annu. Rev. Entomol.* **43,** 243–270.

Lewis, W. J., van Lenteren, J. C., Phatak, S. C., and Tumlinson, J. H. (1997). A total system approach to sustainable pest management. *Proc. Natl. Acad. Sci. USA* **94,** 12243–12248.

Metcalf, R. L., and Luckman, W. H. (1994). "Introduction to Insect Pest Management." Wiley, New York.

Pedigo, L. P. (1996). "Entomology and Pest Management." Prentice-Hall, Upper Saddle River, NJ.

Prokopy, R. J. (1993). Stepwise progress toward IPM and sustainable agriculture. *IPM Pract.* **15(3),** 1–4.

Sterling, W. (1984). Action and inaction levels in pest management. Texas Agricultural Experimental Station, Bulletin 1480. Texas A&M University, College Station.

Tan, K. H. (2000). "Area-Wide Control of Fruit Flies and Other Insect Pests." Pulau Pinang, Malaysia.

Zalom, F. G. (2000). Moving along the IPM continuum. *Proc. Am. Soc. Enol. Vitic.* 50th Meeting, Seattle, WA 2000, pp. 356–359.

Integument

Svend O. Andersen
Copenhagen University

The integument is the external layer of tissue that covers the outer surface of insects and the surfaces of the foregut and hindgut. It is composed of the epidermis, which is a continuous single-layered epithelium, and an underlying thin basal lamina plus the extracellular cuticle that lies on top of the epidermis.

BASAL LAMINA

The basal lamina separates the epidermal cells from the hemolymph in the body cavity; it varies in thickness from 0.15 to 0.5 μm and is composed of structural proteins including collagens, glycoproteins, and glycosaminoglycans. It is negatively charged and can act as a filter between the hemolymph and the epidermal cells, regulating which molecules gain access to the cells.

EPIDERMAL CELLS

The epidermal cells are attached to the basal lamina by hemidesmosomes, which anchor the cell membrane to collagen fibers in the basal lamina. Near their base, the cells are attached to each other by desmosomes; near their apical end, they are attached to each other by a narrow, impermeable zone (the adhering zonule), effectively separating the cuticular compartment from the lateral space between cells. Below the zonule are bands of septate desmosomes, which may be adhesive, and gap junctions through which the cells can communicate chemically with each other by interchange of low molecular weight compounds.

The cuticular materials (chitin and proteins) are secreted from the apical surface of the epidermal cells into the subcuticular space, or deposition zone, where they are assembled into an intact cuticle. The apical surface is folded into shorter or longer microvilli, depending on the stage of the molting cycle and the secretory activity of the cells.

THE MOLTING PROCESS

The cuticle is a rather inextensible structure, and to grow, insects need to shed their old cuticle at intervals after having produced a new one with a larger surface area. The whole process, from breaking the connections between the epidermal cells and the cuticle (apolysis) to emerging from the remnants of the old cuticle (ecdysis), is called molting. A molt is initiated by a brief increase in concentration of the molting hormone, ecdysterone. Whether the new cuticle will become a larval, a pupal, or an adult cuticle depends on the concentration of juvenile hormone during the ecdysterone surge.

The first microscopically visible sign of molting is apolysis, which is the formation of a narrow space between the apical membrane of the epidermal cells and the inner surface of the cuticle. The cell membrane is folded into rather low microvilli, which carry small, dense plaques at the tip. Patches of a new epicuticle form at the top of the plaques, and gradually these patches merge to form a thin continuous layer, called the outer epicuticle or the cuticulin layer, that effectively separates the old cuticle from the space in which new cuticle is deposited.

The epidermal cells undergo mitotic divisions at the onset of molting, resulting in an increase of cell number and total epidermal surface. To allow the animal to increase in size at molting, the epicuticle deposited above the apical cell surface has a larger area than the former epicuticle, and the epidermis together with the epicuticle is folded to be accommodated in the space available inside the old cuticle.

The epidermal cells secrete a mixture of hydrolytic enzymes (the molting fluid) into the space below the old cuticle. The enzymes (proteases, peptidases, chitinase, and glucosidases) degrade the endocuticular chitin and proteins in the old cuticle to free *N*-acetylglucosamine and amino acids to be resorbed by the insect and reused for building the new cuticle.

While the old cuticle is being digested, the new inner epicuticle is deposited beneath the outer epicuticle and the new procuticle starts forming. Chitin is synthesized by an enzyme complex (chitin–synthetase) located at the tip of the microvilli, and chitin microfibrils grow from here into the subcuticular space, the deposition zone. The cuticular proteins are synthesized intracellularly and transported via secretory vesicles from the Golgi complex to the apical plasma membrane, where by exocytosis they are secreted into the subcuticular lumen. The chitin and protein molecules are in some way organized into a macromolecular complex in the deposition zone between cells and cuticle, possibly by a process of self-assembly.

The procuticle grows steadily in thickness until ecdysis, when the partially digested old cuticle ruptures and the insect emerges. Free of the old cuticle (exuvium), the insect expands the new cuticle to a predetermined size, often dependent on the area of the epicuticle. Cuticular deposition of chitin and proteins is resumed and continues for several days after ecdysis, and extended regions of the new cuticle are hardened (sclerotized) by oxidative incorporation of phenolic compounds into the cuticular matrix. In many insects, sclerotization and deposition of endocuticle after ecdysis are governed by the neurohormone bursicon.

Secretion of material from the epidermal cells occurs not only at the apical surface. Some of the proteins synthesized by the cells are exported to the hemocoel via the basolateral membrane system, and others, such as arylphorins, are synthesized in the fat body and secreted into the hemolymph to be taken up by the epidermal cells and incorporated into the cuticle.

PORE CANALS

Most cuticles contain pore canals, minute ducts that traverse the cuticle from the apical surface of the epidermal cells to or close to the cuticular surface. When viewed with an electron microscope they may seem empty, but they often contain one or more cuticular filaments composed of wax and lipids. Cytoplasmic processes may extend into the ends of the pore canals during cuticle deposition. The pore canals are generally assumed to be a transport route for lipids and possibly also for sclerotizing agents and proteins to the epicuticle and outer exocuticle.

SENSE ORGANS

The integument of insects contains a large number of sensory cell types, involved in transferring information from the environment to the insect. The sensory cells in the epidermis are often connected to specific cuticular structures, forming sense organs of various types, such as contact chemoreceptors (taste), olfactory chemoreceptors (smell), and mechanoreceptors, which register any small distortion of the cuticle caused either by the movements of the animal or by influences from the environment.

OENOCYTES

Oenocytes, a special cell type, often are present between the basal region of the epidermal cells and the basal lamina, or they may adhere to the hemolymphal surface of the basal lamina. Electron microscope studies show that they have a highly developed smooth endoplasmic reticulum, characteristic of cells engaged in hydrocarbon synthesis. Oenocytes synthesize hydrocarbon waxes, which are transferred to the wax layer of the epicuticle, presumably via the epidermal cells and the procuticular pore canals. It has been suggested that the oenocytes also provide lipids to the outer and inner epicuticle as well as to the sclerotized regions of the exocuticle.

INTEGUMENTAL GLANDS

Epidermal glands are present in many types of integument; they often consist of a single cell surrounding a cavity, which

functions as a product reservoir and is connected to the cuticular surface by a small duct. Glands of this type are assumed to be responsible for forming and maintaining the cement layer, which after ecdysis is spread as a protective layer on top of the epicuticular wax layer. Other integumental glands of this type produce various chemical defense secretions, which may be forcibly ejected when the insect is disturbed.

Integumental glands also may consist of epidermal cells without a cavity and duct, but with direct contact to the inner surface of the cuticle, through which the secretion passes. Often the secretions of such glands function as pheromones, playing a part in the communication between individuals.

COLORATION

The colors of most insects are the result of pigments located in cuticle and epidermis, or they may be physical colors caused by surface structures in the cuticle. Colored material in hemolymph or internal organs can also contribute to the insect's color if the integument is transparent. Ommochromes, pteridines, carotenoids, bile pigments, melanins, and urates are the most widespread and important of the epidermal pigments. The colored light reflected from such epidermal pigments passes the overlying cuticle before it reaches the eye of the observer, and the observed color is influenced by the amount of colored material present in the cuticle.

See Also the Following Articles

Chemical Defense • Coloration • Cuticle • Exoskeleton • Molting

Further Reading

Bereiter-Hahn, J., Matoltsy, A. G., and Richards, K. S. (eds.). (1984). "Biology of the Integument," Vol. 1 of "Invertebrates." Springer-Verlag, Berlin. (See especially Chaps. 27–35.)

Binnington, K., and Retnakaran, A. (eds.). (1991). "Physiology of the Insect Epidermis." CSIRO Publications, Melbourne, Australia.

Hepburn, H. R. (ed.). (1976). "The Insect Integument." Elsevier, Amsterdam.

Kerkut, G. A., and Gilbert, L. I. (eds.). (1985). "Comprehensive Insect Physiology, Biochemistry and Pharmacology," 13 vols. Pergamon Press, Oxford, U.K. (See especially Vol. 3, Chaps. 1–4.)

Ohnishi, E., and Ishizaki, H. (eds.). (1990). "Molting and Metamorphosis." Japan Scientific Societies Press Tokyo, Japan.

Introduced Insects

Daniel Simberloff
University of Tennessee

An introduced species is a species that did not achieve its current taxonomic status in some location by natural evolutionary processes. Such species are often said to be nonindigenous or adventive to the location of introduction. Among nonindigenous species, those carried to the location by humans, either deliberately or inadvertently, are called "introduced species," whereas those that arrived of their own volition are termed "immigrant." However, all nonindigenous species are often colloquially termed "introduced." "Exotic" and "alien" are sometimes used for all nonindigenous species.

Introduced insects generate major ecological and economic impacts by a variety of means. Many species are introduced in various regions, and they span most insect orders. Their means of introduction are numerous, but most arrive with human help. Only a minority become problematic, and the reasons why a newly arrived species survives or fails to establish a population, and, if established, has or does not have a major effect, are often mysterious. Some problematic introduced species can be eradicated, and several management procedures can adequately control undesirable species when eradication attempts are unsuccessful.

NUMBERS

Introduced species often comprise a substantial fraction of a regional entomofauna. For example, in Great Britain there are about 21,000 native insects and about 200 established, introduced ones (comprising ~1% of the total). For the contiguous United States, approximately 84,000 native species have been identified, as have 1862 established, introduced species (~2% of the total). Florida has relatively more introduced species: 11,509 native and 993 introduced (8%). For oceanic islands, the introduced proportion can be much greater. Thus, for the Hawaiian Islands, there are about 5400 native insects and 2600 introduced ones (32% introduced), whereas the mid-Atlantic island Tristan da Cunha has 84 native insects and 32 introduced ones (28% introduced). The greater proportions of introduced species on islands probably result more from the mathematics of smaller numbers of native species and relatively larger numbers of attempted introductions than from some inherent invasibility of island communities or stronger "biotic resistance" of continental ones.

SYSTEMATICS

Established introduced insects are not a systematically random group. An order may be over- or underrepresented, reflecting primarily the reasons for its introduction and/or the means of its arrival. This pattern is exemplified in Fig. 1, which depicts the proportional fractions, by order, of insects of the world compared with introduced insects of the contiguous United States. Thus, far more species of homopterans, hemipterans, and thysanopterans are introduced than would be expected based on their numbers in the world, because these orders are associated with introduced agricultural and ornamental plants. Hymenopterans are overrepresented

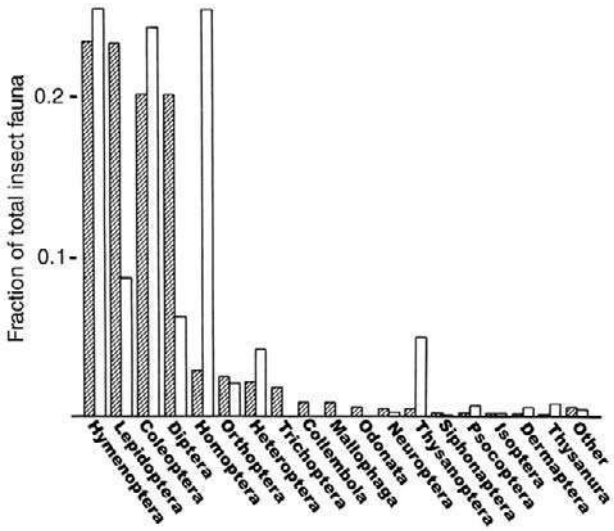

FIGURE 1 Fractional representation of different insect orders of the world (cross-hatched) and established introduced insects of the United States (clear). [From Simberloff, D. (1986). Introduced insects: a biogeographic and systematic perspective. *In* "Ecology of Biological Invasions of North America and Hawaii (H. A. Mooney and J. A. Drake, eds.), Fig. 1.1, p. 6. © Springer-Verlag GmbH & Co. K G, Heidelberg. Data from Arnett, R. H. (1983). Status of the taxonomy of the insects of America north of Mexico: a preliminary report prepared for the subcommittee for the insect fauna of North America project. (Privately printed) and Sailer, R. I. (1983). History of insect introductions. *In* "Exotic Plant Pests and North American Agriculture" (C. Graham and C. Wilson, eds.), pp. 15–38. Academic Press, New York.]

perhaps partly because of their wide use in biological control and partly because their haplodiploidy allows a single female, which can produce both male and female offspring, to found a population. Haplodiploidy may also lower inbreeding depression in founding populations. Haplodiploidy is also found among homopterans and thysanopterans, both overrepresented groups. At the family level, 26 families are overrepresented, and other patterns become clear. In the contiguous United States, coccinellid beetles and anthocorid bugs are both overrepresented; species in both groups were introduced as biological control predators. Tephritid flies were introduced deliberately for control of knapweeds and inadvertently on cultivated plants. Ten overrepresented families are in the Hymenoptera, and seven of these were introduced for biological control. Two thrips families in this list are associated with cultivated plants, as are the adelgids (pine and spruce aphids). Oestrid flies are overrepresented, and these are internal parasites of mammals, including livestock; they probably arrived with their hosts. Dermestid and anobiid beetles are overrepresented, and these stored product pests probably arrived with their food sources.

WHY, WHEN, AND HOW INSECTS ARE INTRODUCED

Some introduced insects arrive in new locations on their own and are true immigrants. Of course, any simple range extension could bring a new introduced species in this sense, but as a rule this status is restricted to species arriving at a distant location by a discontinuous dispersal, rather than gradually diffusing from a neighboring site. Thus, the monarch butterfly *(Danaus plexippus)* in Australia might not be considered to be introduced. It arrived around 1870 and established a population, having spread through the Pacific during the 19th century largely unaided by humans. The monarch migrates long distances in its native North America, is an occasional straggler in Europe, and was recorded successively in Hawaii, the Caroline Islands, Tonga, and New Zealand before it reached Australia.

Some insects are introduced deliberately by humans. A few insects arrive as pets or pet food; recent pet price lists include mantids, walkingsticks, spider wasps, velvet ants, and dung and blister beetles. In Florida in 1989, giant Madagascan hissing cockroaches *(Gromphadorhina portentosa)* became highly popular as pets; at least some were released to the wild, where they survive. Similarly, caterpillars of a Chilean moth, *Chilecomadia morrei,* are sold as reptile food in the United States and Europe. The European honey bee *(Apis mellifera)* has been widely introduced for both the production of honey and crop pollination. The Asian silkworm *(Bombyx mori)* has been widely introduced along with its host plant, mulberry, for silk production. Even introductions that fail to establish a commercial industry can nonetheless establish a population. The Asian ailanthus moth *(Samia walkeri)* was brought to the United States in an attempt to found a silk industry on the ailanthus tree. Though the industry foundered, the moth remains. The gypsy moth *(Lymantria dispar)* was brought to North America to establish a silk industry; its predictable escape established one of North America's major pests.

Most deliberate insect introductions are for biological control. Although weeds and insect pests of agriculture are the usual targets, there are others. Thus, *Paratrechina fulva,* a Brazilian ant, was introduced to Colombia to control poisonous snakes, whereas over 45 dung beetle species were introduced to Australia to break down droppings of introduced livestock (~33% of established populations). Insects introduced to attack insects are either predators (primarily coccinellid beetles, but including other beetles, hemipterans, and neuropterans) or parasitoids (mostly hymenopterans, but including some dipterans). Insects introduced to control plant pests include primarily flies, beetles, and moths, although others such as bugs and thrips have been used. The number of insect species introduced for biological control purposes is substantial. For example, of approximately 2600 introduced insect species established in the Hawaiian islands, roughly 400 were introduced for biological control. About twice as many species introduced there for this purpose perished.

Far more insects are introduced inadvertently than deliberately by humans. Pathways are myriad. Soil ballast was an early predominant mode of entry to North America—many of the first introduced insects were soil beetles from

southwestern England. This pathway is less common now, but insects are still carried in rootballs around cultivated plants and in soil on heavy equipment. Phytophages, particularly homopterans, dominated introductions to North America in the 19th century with the advent of fast steamships and a proliferation of imported nursery stock, and imported plants are still a major means of introducing insects worldwide. The grape phylloxera *(Daktulosphaira vitifoliae)* that devastated French vineyards in the 19th century arrived on saplings or cuttings of American vines. Insects can also be carried in water. The yellow fever mosquito *(Aedes aegypti)* probably arrived in colonial America in drinking water casks, while the Asian tiger mosquito *(A. albopicta)* reached North America in the 1980s in scrap tires from Japan. This is probably the route taken by *A. japonicus,* which arrived in the United States in 1998 and transmits West Nile virus. These latter two mosquitoes have recently been detected in used tires in New Zealand. Wooden packing material brought the Asian longhorned beetle *(Anoplophora glabripennis)* to the United States, whereas the European elm bark beetle *(Scolytus multistriatus)* that transmits Dutch elm disease arrived in North America on unpeeled veneer logs of European elm. The growth of international tourism can enhance the rate of insect introduction; in 1992 an Australian tourist returned from South America with a wound containing maggots of the New World screwworm *Cochliomyia hominivorax.*

ESTABLISHMENT AND SPREAD

The majority of introduced insects, like most introduced species, do not survive, although only biological control introductions generate substantial data on failed introductions. For parasitoid species introduced to control insect pests, only about 30% establish populations, whereas for all insects introduced for plant control, the comparable figure is about 60%. Because biocontrol candidates are chosen and often tested for survival in the target environment, one might expect failure rates for inadvertently introduced introductions to be even higher. For most taxa, invasion biologists believe that 5 to 20% of introduced species establish populations, although many of these may remain for years or in perpetuity near the point of introduction, and a large fraction are restricted to anthropogenous habitats such as human habitations or agricultural fields.

An arriving propagule must be large enough to survive the initial threat of demographic stochasticity—that is, random elimination of so many individuals during the first few generations that the population fails. Data from the biological control literature show that probability of establishment increases with propagule size and number of attempts, but many very small propagules have established large, widespread populations. For example, a single fertilized female of the cochineal insect *Dactylopius opuntiae* from Sri Lanka initiated a large, ongoing population on Mauritius. In Puerto Rico, two females of a unisexual race of the encyrtid wasp *Hambletonia pseudococcina* were used to rear 7000 individuals, which were released, established, and quickly spread. For a parthenogenetic species, at least the difficulty of finding a mate is obviated, but demographic stochasticity has other components.

Even assuming an adequate propagule size, the environment, both physical and biotic, must be suitable for a species to survive and spread. Predators, parasites, competitors, and pathogens can eliminate an introduced species or restrict its ambit. For example, the Asian aphelinid wasp parasitoid *Aphytis fisheri,* introduced to California to control California red scale *(Aonidiella aurantii),* failed to establish because of competition from previously introduced *A. melinus* and *A. lingnanensis.* Conversely, the absence of natural enemies from its native range is often posited as the reason for the success of some invaders, such as the cynipid *Andricus quercuscalicis,* introduced into Great Britain. By contrast, the presence of another species, such as a food plant or a symbiont, might be necessary for an invader to survive. The monarch butterfly would not have survived in Australia but for the prior introduction of its host milkweeds. The physical environment is probably an even more frequent reason for introduced species to perish. A temperate climate does not augur well for a newly arrived tropical insect, but even subtler physical differences can be crucial to a species's survival. For instance, synanthropic species are unlikely to survive if they arrive in pristine natural habitats.

IMPACTS

Although quantifying the impact of a new species is an unsolved challenge, it is safe to say that most introduced species do not generate major impacts. However, some are enormously damaging, whereas others are highly beneficial. The variety of impacts is staggering.

Many introduced insects prey on natives. This activity can be useful, as in biocontrol introductions such as that of the Australian vedalia *(Rodolia cardinalis)* to attack the cottony cushion scale *(Icerya purchasi).* Predation can also be extremely damaging; on Christmas Island, the introduced yellow crazy ant *Anoplolepis gracilipes* has locally devastated populations of the dominant red crab *Gecarcoidea natalis.* Because the crab controls seedling recruitment and litter breakdown, the entire community is affected. Parasites can also be beneficial or harmful. The wasp *A. melinus* has effectively controlled California red scale in parts of California. Alternatively, sheep blowfly *(Lucilia cuprina),* introduced to Australia from Africa, caused staggering losses. Herbivory can similarly be beneficial or detrimental. The South American flea beetle, *Agasicles hygrophila,* effectively controls alligatorweed in Florida. However, phytophagous insect crop pests impose staggering costs. The alfalfa weevil *(Hypera postica)* caused $500 million in losses in the United States in 1990 alone.

Resource competition is subtler than predation, parasitism, and herbivory, but many introduced insects outcompete

natives. The European sevenspotted lady beetle *(Coccinella septempunctata),* introduced to the United States for control of the Russian wheat aphid *(Diuraphis noxia),* has locally outcompeted several native lady beetles. European honey bees outcompete the native bee *Osmia pumila* for pollen in New York State. Introduced insects can even outcompete vertebrates. The introduced wasps *Vespula germanica* and *V. vulgaris* in New Zealand outcompete an endemic parrot for honeydew produced by a scale insect *(Ultracoelostoma assimile)* and have locally lowered parrot populations.

Introduced insects can transmit or be reservoirs of diseases of humans, domestic animals, cultivated plants, and wild animals and plants. The spread of yellow fever and dengue as the vector mosquito *A. aegypti* dispersed throughout the tropics, and of malaria to Brazil with the introduction of its vector *Anopheles gambiae,* are notable examples of human disease organisms transmitted by introduced insects. Cat fleas *(Ctenocephalides felis)* and dog fleas *(C. canis),* introduced to Australia with their hosts, are intermediate hosts for the dog tapeworm *(Dipylidium caninum).* The southern house mosquito, *Culex quinquefasciatus,* was accidentally introduced to the Hawaiian islands in 1826. Subsequently it transmitted the disease organism causing avian malaria, introduced with resistant Eurasian songbirds, to susceptible native birds and helped to exclude them from low elevations. The mosquito *A. japonicus* transmits West Nile virus to both birds and humans in the northeastern United States. Animal disease vectors may be useful biocontrol introductions. For example, the rabbit flea *(Spilopsyllus cuniculi),* the main vector of the organism causing myxomatosis in Europe, has been introduced (so far unsuccessfully) in Australia to attempt to boost disease transmission. In the lab, it also transmits calicivirus.

Among plants, Dutch elm disease was dispersed to and through North America with the European elm bark beetle. Beechbark disease spread throughout northeastern North America after the causal fungus was introduced from Europe around 1890 with its vector, the beech scale *(Cryptococcus fagisuga).* The wine industry in California is threatened by the recent introduction of the glassy-winged sharpshooter *(Homalodisca coagulata)* from the southeastern United States. The sharpshooter spreads an incurable bacterial disease of grape vines, a malady long present but rarely a problem until this vector arrived.

Introduced species often exacerbate one another's impacts, a process termed "invasional meltdown." Sometimes this interaction occurs when coevolved mutualists invade a region separately. In south Florida, over 60 species of ornamental figs were not invasive because their obligatory pollinating wasps were absent. Since the 1970s, three such wasps (*Parapristina* spp.) have arrived, and three formerly innocuous fig species have begun spreading in natural areas. However, invasional meltdown need not involve coevolved species. In California citrus orchards, the Argentine ant *(Linepithema humile)* tends and protects the Asian California red scale, thereby exacerbating its impact. Similarly, in Hawaii, the African bigheaded ant *(Pheidole megacephala)* protects the tropical American gray pineapple mealybug *(Dysmicoccus neobrevipes)* from coccinellids introduced for biological control.

Relative to the numbers of species introduced, insects rarely cause enormous ecological (as opposed to economic) damage. Introduced species whose impacts ripple through entire communities usually do so by changing the habitat dramatically, and such change agents are mostly plants (which become structural dominants or modify fire regimes) or pathogens, which attack dominant plants. Occasionally mammals can generate an enormous ecosystem impact by trampling or grazing. A recent list of the world's 100 worst introduced species included 15 insects, but at most one would qualify as having a huge ecosystem-wide impact: the yellow crazy (or long-legged) ant, which removes the keystone red crab species on Christmas Island. Of the 15 insects, five are ants. In addition to the yellow crazy ant, the Argentine ant, the bigheaded ant, the little fire ant *(Wasmannia auropunctata),* and the red imported fire ant *(Solenopsis invicta)* all affect other ants greatly, and sometimes other insects, but to date none has had the dramatic impact of certain plants and mammals. Some species among the 15 transmit organisms that cause human diseases (*Aedes albopicta* and *Anopheles quadrimaculatus*) and others are agricultural pests, such as the sweetpotato whitefly *(Bemisia tabaci)* and several of the ants. The Formosan termite *(Coptotermes formosanus shiraki)* has caused enormous damage to housing in New Orleans. With respect to ecosystem-wide damage to natural areas, however, the only members of the list that might qualify, aside from the yellow crazy ant, are the gypsy moth in North America, by virtue of its devastating impact on dominant trees, and the Argentine ant, because it has greatly lowered densities of native seed-carrying ants in the fynbos of South Africa. Other insects may have ecosystemic impacts by removing dominant plants. The beetle transmitting the Dutch elm disease organism has already been noted. The Asian balsam woolly adelgid *(Adelges piceae)* has eliminated the dominant Fraser fir throughout the high southern Appalachians, whereas the hemlock woolly adelgid *(A. tsugae)* has locally killed large fractions of hemlocks in much of eastern North America.

Impacts of an introduced species can occur after a substantial lag period during which the species can seem to be innocuous. For example, the beetle *Chrysolina quadrigemina,* introduced to Australia in 1939 to control St. John's wort, seemed to die out but resurfaced and spread in 1942. Such lags are mysterious; they are often attributed to favorable changes in the environment or to evolution of the invader, but evidence for these phenomena is generally lacking. Some introduced insects achieve great numbers and appear to have a major impact, but the population suddenly crashes, again for reasons poorly understood. The European browntail moth *(Euproctis chrysorrhoea)* followed this trajectory in New England and eastern Canada.

EVOLUTION

The conditions under which a species is introduced to a new region (isolation from parent population, small propagule size [usually], and different physical and biotic environment) should be conducive to rapid evolution. There has been little study of this phenomenon, but some striking examples have emerged. *Drosophila subobscura,* introduced to the Americas from the Old World around 1980, spread widely and by 2000 had evolved a cline of increasing total wing length with latitude in North America phenotypically similar to that in its native range. The ichneumonid *Bathyplectes curculionis,* introduced to the western United States for biological control of introduced alfalfa weevils (*Hypera* spp.), evolved in less than 10 years to become less susceptible to the encapsulation reaction of its host.

A phenomenon widely reported among introduced vertebrates and plants, particularly in North America and Eurasia, is hybridization with native species, sometimes to the point of a sort of genetic extinction of the latter. Although hybridization is known to have played an important role in insect evolution, hybridization between native and recently introduced insects is rarely if ever reported. This absence of data may reflect biological differences or simply less genetic study of insects. There are instances of introduced populations hybridizing with one another, most notably the Italian and African strains of the honey bee. The red imported fire ant hybridizes extensively with the previously introduced black imported fire ant *(Solenopsis richteri)* in Tennessee.

ERADICATION AND MANAGEMENT

Eradication is one possible response to an introduced species, particularly one that has not dispersed widely. Many introduced insects have been eradicated, some from substantial areas. Perhaps the most impressive is the chemical eradication of the African malaria mosquito *A. gambiae* from 31,000 km^2 of northeastern Brazil in 1939–1940. The Mediterranean fruit fly *(Ceratitis capitata)* was eradicated over 18 months from a 20-county region of Florida by a strict quarantine, destruction of produce and plants, trapping, and insecticide sprays. More recent attempts to eradicate *C. capitata* in both Florida and California may not have been as successful. Although victory has been declared repeatedly, reappearances are frequent and may constitute either new invasions or simply recovery by uneradicated remnant populations.

The development of the sterile-male technique in the United States against the New World screwworm gave tremendous impetus to the eradication approach. Release of massive numbers of sterile males so reduced the probability of fruitful mating by females that this species disappeared totally from the island of Curaçao in 1954–1955, and this method greatly aided eradication of this fly from the southeastern United States in 1958–1959. The melon fly *(Bactrocera cucurbitae)* was eliminated from Rota Island by this method. The male annihilation method, in which males are attracted and destroyed, has also succeeded in eradicating introduced fruitfly populations from islands, including the Oriental fruit fly *(Dacus dorsalis)* from Rota and Guam and the melon fly from Nauru. Male annihilation followed by release of sterile males eradicated the melon fly from the entire Ryukyu Archipelago. The white-spotted tussock moth *(Orgyia thyellina)* was eradicated from greater Auckland, New Zealand, by pheromone lures plus spraying of *Bacillus thuringiensis.*

There have also been disastrous failures of expensive eradication campaigns, such as the $200 million attempt to eradicate chemically the red imported fire ant from the southeastern United States, an effort that imposed greater mortality on native insects than on the invader. Quick detection, rapid response, sufficient resources to finish the project, and adequate regulatory power to enforce cooperation have proven most conducive to successful eradication.

If eradication fails or is not attempted, chemical and biological control are the two methods most commonly attempted to manage introduced insects. There are successes and failures for both methods. The nontarget and human health impacts of early generation insecticides such as DDT are legendary. Though more recent chemicals minimize or eliminate this problem, chemical control frequently is problematic for two main, related reasons. First, insects evolve resistance to chemicals; second, expense can be far too great, especially when ever larger amounts must be used because of resistance. For large natural areas, expense of continued chemical applications can be particularly prohibitive.

Biological control is attractive because the expense of development and deployment is lower, and because, although a host may evolve resistance, the biological control agent itself can evolve countermeasures (as witness *B. curculionis,* discussed earlier). However, the success rate of biological control is rather low. For instance, for parasitoids introduced for insect control, only 10% have been effective. Furthermore, biocontrol agents can affect nontargets, as the sevenspotted lady beetle has done, and these nontarget impacts can be generated by established biocontrol agents that are not even effective against their targets (about three times as many biocontrol parasitoids establish populations as actually control the target pest). The tachinid fly *Compsilura concinnata* has failed to control gypsy moths in North America, but it is believed to be responsible for the decline of several large native moths.

See Also the Following Articles

Conservation • Fire Ants • Gypsy Moth • Island Biogeography and Evolution • Neotropical African Bees • Regulatory Entomology

Further Reading

New, T. R. (1994). "Exotic Insects in Australia." Gleneagles, Adelaide.

Sailer, R. I. (1983). History of insect introductions. *In* "Exotic Pest Plants and North American Agriculture" (C. Graham and C. Wilson, eds.), pp. 15–38. Academic Press, New York.

Simberloff, D. (1986). Introduced insects: A biogeographic and systematic perspective. *In* "Ecology of Invasions of North America and Hawaii" (H. A. Mooney and J. A. Drake, eds.). pp. 3–26. Springer-Verlag, New York.
Simberloff, D. (1989). Which insect introductions succeed and which fail? *In* "Biological Invasions. A Global Perspective" (J. A. Drake, H. A. Mooney, F. diCastri, R. H. Groves, F. J. Kruger, M. Rejmánek, and M. Williamson, eds.), pp. 61–75. Wiley, Chichester.
Williams, D. F. (ed.). (1994). "Exotic Ants. Biology, Impact, and Control of Introduced Species." Westview, Boulder, CO.
Williamson, M. (1996). "Biological Invasions." Chapman & Hall, London.

Island Biogeography and Evolution

George K. Roderick and Rosemary G. Gillespie
University of California, Berkeley

Studies of insects have played a major role in the general understanding of the biota of islands, touching on all areas of biogeography, ecology, evolution, and conservation. The notable writings of Darwin and Wallace were influenced heavily by the biological diversity that each witnessed on islands and by the processes inferred to underlie that diversity. Rather more recently, studies exploiting the discrete nature of islands have given rise to pervasive organizing theories of community ecology, in particular MacArthur and Wilson's equilibrium theory of island biogeography (ETIB). With the advent of accessible molecular genetic tools, research on islands has allowed unique insights into the processes that generate biotic diversity, especially the mechanisms of speciation. Unfortunately, islands are also prime targets for biological invasions, mediated largely by anthropogenic disturbance. The severity of such impacts on island biotas may result from their evolution in isolation, but it is certainly compounded by their characteristically small population sizes. Yet, for many islands extinction among arthropods is largely unknown, although this circumstance may be attributable more to lack of knowledge than to any innate security that arthropods might possess. Indeed, it is likely that many island arthropods will go extinct before they have been collected and described.

THEORY OF ISLAND BIOGEOGRAPHY

Larger islands contain more species. This idea was formalized by MacArthur and Wilson in the 1960s with the development of the ETIB. This theory relates species and area by the formula $S = cA^z$, where S is species number, A is area, c is a constant measuring overall species richness, and z measures the extent to which increases in area have diminishing returns in terms of species number. Values of z tend to vary between 0.18 and 0.35; that is, doubling the species number requires increasing the area by a factor lying between 7 and 100. The premise of the theory is that the rate of immigration decreases with increasing distance from the source, whereas the rate of extinction decreases with increasing island size. The balance of these processes results in an equilibrium number of species on any one island. As the number of resident species on an island increases, the chance of an unrepresented species arriving on that island decreases and the likelihood of extinction of any one resident species increases. The predictions of the model are as follows: (1) the number of species on an island should change little once the equilibrium has been reached; (2) there should be continual turnover of species, with some becoming extinct and others immigrating; (3) small islands should support fewer species than large islands; and (4) species richness should decline with remoteness of the island, since islands farther from the source will have lower rates of immigration.

Rigorous tests of the ETIB have been surprisingly few, and they have supported some aspects of the theory but not others. For example, Simberloff used insecticides to defaunate mangrove islands and found that species of insects and spiders accumulated to an equilibrium number. However, contrary to expectation, turnover of species was not randomly distributed among species—species of particular types were likely to colonize or go extinct. Species numbers have been found to be affected unpredictably by both area and isolation; yet other work has shown that an equilibrium does not exist, or that parameters other than area per se may dictate species richness. Such factors include habitat diversity, climatic conditions, island age, and even the status of knowledge concerning the presence of resident species. However, the theory has proven to be remarkably useful and, although it was developed for islands, it has had relevance for the study of ecological communities of many kinds.

ADAPTIVE RADIATION

The ETIB assumes that islands are within the geographic distance into which a species is likely to disperse, thus maintaining genetic populations between source and island populations. On islands that are beyond the range within which populations can maintain genetic contact with source populations, one might predict (based on the theory) that few species should be present. But this tends not to happen. Isolated islands that are formed initially without life are often found to have large numbers of closely related species. When single colonists, isolated genetically from their source population, give rise to a series of species that have diversified ecologically, the phenomenon is termed adaptive radiation. Usually it occurs beyond the so-called radiation zone, or normal range of dispersal of a given organism. Species that form through adaptive radiation are typically neoendemics, formed *in situ* and found nowhere else. Among arthropods, the Hawaiian Islands hold the record in having the largest number of neoendemics, an extraor-

dinary 98% of the fauna. Founder effects, behavioral isolation, ecological isolation, and host-associated isolation have all been implicated in the process of adaptive radiation. For insects, particularly noted examples include *Drosophila* flies, which are well known for their diversity of mating behaviors, as well as lineages of crickets that have diversified in song repertoire, sap-feeding planthoppers that have proliferated by switching between plant hosts, and beetles that have formed new species on different substrates. Diversification may follow a predictable pattern, at least in some groups; for example, among *Tetragnatha* spiders, similar ecological sets of species have evolved over and over again on each of the different Hawaiian Islands.

Compared with their hypothetical colonizing ancestors, species on remote oceanic islands are often characterized by a reduction in dispersal ability. Indeed, they are often found to have very narrow ranges of dispersal. Moreover, the individual species that colonize remote islands are a small sample of the continental source. They are therefore said to be "disharmonic" and not representative of the biotic diversity on continents, a phenomenon accentuated by the frequent proliferation of successful colonists.

NEOENDEMIC AND PALEOENDEMIC ISLANDS

Neoendemics typically form on isolated islands that have been created *de novo* and have abundant empty ecological space into which those few colonists can diversify. Besides Hawaii, other volcanic archipelagoes, including the Marquesas, Societies, and Galapagos in the Pacific and the Canaries in the Atlantic, have provided ideal conditions for the formation of neoendemics. However, species can also form on fragment islands, formed as a mass of land has broken away from a larger continental region. Examples of such islands include some of the Caribbean islands, and the islands of New Zealand and Madagascar. As these islands, formed upon losing connection with a continental source, become more isolated, gene flow between island and continental populations may become insufficient to overcome genetic divergence. Unlike volcanic islands that form in isolation, starting without any species and accumulating species through time, fragment islands are usually ecologically saturated at the time of separation and tend to lose species through ecological time, a process termed relaxation. Over evolutionary time, the species on these islands may change through relictualization, with the formation of paleoendemics, usually without adaptive radiation.

HABITAT ISLANDS

Many of the ideas originally developed for islands in the sea have been extended also to so-called habitat islands of a particular habitat type in a matrix of unlike terrain. Most such islands are fragments of habitats that were historically connected, such as remnant trees and forest patches. For habitat islands, as for islands in the sea, ecological and evolutionary processes are governed largely by isolation, time, and the nature of the matrix relative to the dispersal abilities of the organisms in question. Habitat islands, because they are discrete and easy to manipulate compared with islands in the sea, have been exploited in the development of many ecological principles, including those related to metapopulation dynamics and the physical design of nature reserves.

CONSERVATION

The biota of islands is often unique—for example, the islands of the Pacific have been designated a biodiversity hot spot. Assessing this diversity, particularly for arthropods, is problematic. The major impediment is a lack of taxonomic understanding of arthropods on many islands, particularly those that are more remote. New species are being collected at a remarkable rate in areas such as French Polynesia, Madagascar, and even the relatively well-studied Canary Islands, New Zealand, Hawaii, and the Galapagos, yet the training of arthropod systematists has lagged behind.

Anthropogenic disturbance has also had its impact not only in present times, but historically, as witnessed through the colonization of the Pacific by Polynesians several thousands of years ago. A number of characteristics of arthropods populations on islands, including high local endemism, limited dispersal abilities, and small population sizes, make them particularly vulnerable to both demographic accidents and environmental change. In addition, islands have been impacted heavily by invasive species, many of which are also arthropods. The impact has been both direct, such as through the extirpation of species by invasive predatory ants, and indirect, such as through diseases of vertebrates having mosquitoes as vectors.

Although islands have long served as extraordinary laboratories for studying processes associated with the generation of diversity, they are now contributing to understanding of processes leading to the loss of diversity. For example, studies of invasive species on islands have shown the importance of environmental factors as well as species-specific attributes that facilitate biological invasions and its negative effects. New tools are urgently needed: rapid biodiversity assessment techniques that bypass traditional taxonomic identification will be important in recognizing areas of high conservation priority, as will genetic or ecological approaches that can distinguish native species from those introduced in more recent history.

See Also the Following Articles

Biodiversity • Biogeographical Patterns • Cave Insects • Introduced Insects

Further Reading

Darwin, C. (1859). "The Origin of Species by Means of Natural Selection." John Murray, London.

Gillespie, R. G., and Roderick, G. K. (2002). Arthropods on islands: Colonization, speciation, and conservation. *Annu. Rev. Entomol.* **47,** 595–632.

Howarth, F. G., and Mull, W. P. (1992). "Hawaiian Insects and Their Kin." University of Hawaii Press, Honolulu.

MacArthur, R. H., and Wilson, E. O. (1967). "The Theory of Island Biogeography." Princeton University Press, Princeton, NJ.
Wagner, W. L., and Funk, V. A. (eds.). (1995). "Hawaiian Biogeography, Evolution on a Hot Spot Archipelago." Smithsonian Institution Press, Washington, DC.
Wallace, A. R. (1902). "Island Life." 3rd ed., Macmillan, London.
Whittaker, R. J. (1998). "Island Biogeography: Ecology, Evolution, and Conservation." Oxford University Press, Oxford, U.K.
Williamson, M. H. (1981). "Island Populations." Oxford University Press, Oxford, U.K.

Isoptera
(Termites)

Vernard R. Lewis
University of California, Berkeley

The ordinal name Isoptera refers to the two pairs of straight and very similar wings that termites have as reproductive adults. The common name, of Latin origin, translates as "woodworm." Termites are small and white to tan or sometimes black. They are sometimes called "white ants" and can be confused with true ants (Hymenoptera). However, a closer look reveals two easily observed, distinguishing features: termites have straight antennae and a broad waist between the thorax and the abdomen, whereas ants have elbowed antennae and a narrow waist. For reproductive forms, termites have four equally sized wings, whereas ants have two pairs of dissimilarly sized wings.

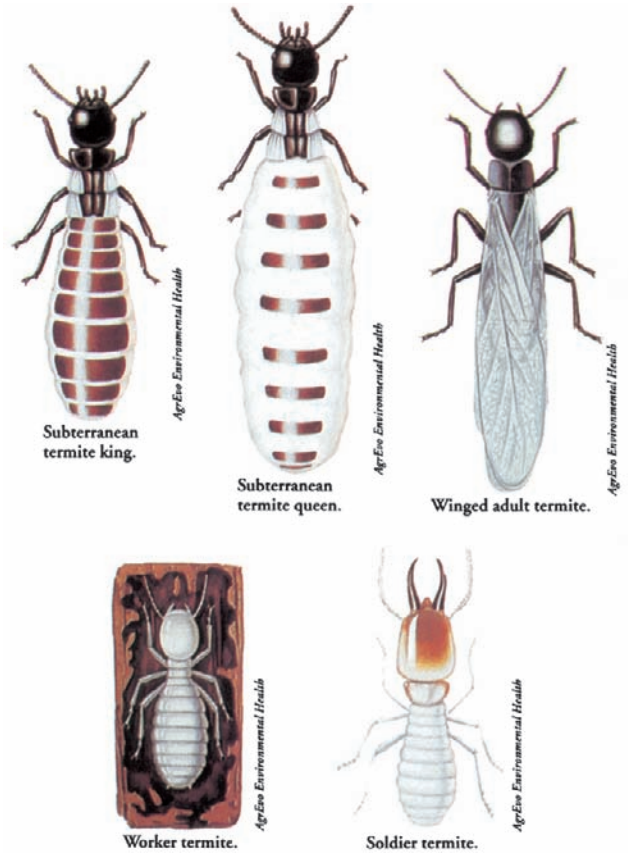

FIGURE 1 Castes for Isoptera. A lower termite group, *Reticultermes,* is represented. A large queen is depicted in the center. A king is to the left of the queen. A worker and soldier are below. (Adapted, with permission from Aventis Environmental Science, from *The Mallis Handbook of Pest Control,* 1997.)

IMPORTANT FAMILIES OF TERMITES

The earliest known fossil termites date to the Cretaceous, about 130 mya. There are >2600 species of termites worldwide. Undoubtedly, more will be recognized with improved methods of discerning cryptic species and after intensive collecting of tropical and remote regions. Termites are most closely related to cockroaches and mantids. The greatest continental termite diversity is in Africa, where there are over 1000 species. Polar continents have none, and North America with 50 species and Europe with 10 species are intermediate in termite diversity.

Development is incomplete metamorphosis containing castes that include nymph, worker, pseudergate, soldier, and several types of reproductives (Figs. 1, 2, and 3). Nymphs hatch from eggs and molt at least three times before becoming functional workers. Workers are wingless, do not lay eggs, and, except for the family Hodotermitidae, are blind. Worker and pseudergate castes are the most numerous in a colony and conduct all major foraging and nest-building activities. Soldiers defend colonies with fearsome mandibles and/or chemical squirts from a nasus, a frontal projection from their heads. Soldiers, including nasutes, cannot feed themselves. Reproductives consist of a royal pair, the original colony founders, but supplementary and replacement reproductives (neotenics) can be generated from workers, nymphs, or other immatures dependent on pheromonal cues from the queen and environmental factors.

Termite families traditionally were categorized as lower or higher. However, this categorization may change soon as newer classification systems are adopted. Lower termites (families Mastotermitidae, Kalotermitidae, Termopsidae, Hodotermitidae, Rhinotermitidae, and Serritermitidae) have symbiotic intestinal protozoa and bacteria. Higher termites (Termitidae) have intestinal bacteria.

Termite identification at the family and genus level is determined using reproductive adults or soldiers or, in some groups, workers. All living termites can be divided into seven families as follows.

Mastotermitidae

This family contains the most primitive living termite, *Mastotermes darwiniensis* (Fig. 4), now limited to Northern

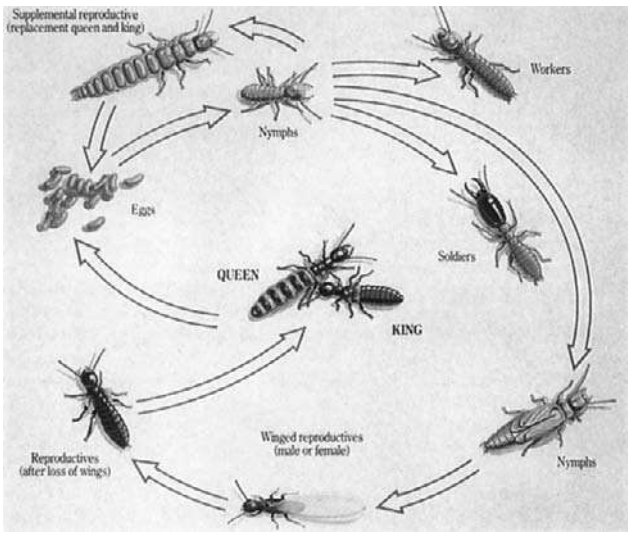

FIGURE 2 Life cycle of the termite. Lower termite family depicted. (Adapted, with permission from FMC Corp., from *The Mallis Handbook of Pest Control,* 1997.)

Australia. In appearance, these termites are light brown, robust, and about 8–10 mm in length. This family is recognized by the presence of an anal lobe in the hind wing of the reproductive adults and five-segmented tarsi. The hind wings are very similar to those of some cockroaches, considered a sister group to termites. Like cockroaches, reproductive females also lay egg cases containing up to 24 eggs arranged in two regular rows. Although egg masses contain few eggs, *Mastotermes* has many neotenic reproductives (no primary queen has ever been found in the field), and colonies can reach a population size of millions. Soldiers have powerful mandibles and excrete a toxic brown substance from their buccal cavity that repels intruders.

FIGURE 3 Castes of the Formosan subterranean termite, *Coptotermes formosanus* (Isoptera: Rhinotermitidae). In the center is a queen with large physogastric abdomen containing eggs. A king with a physogastric abdomen lies next to the queen. Soldiers have brown tear-shaped heads with sickle mandibles. A worker is also shown. (Photograph courtesy of Dr. Minoru Tamashiro, University of Hawaii.)

FIGURE 4 *Mastotermes darwiniensis,* the most primitive termite from Darwin, Australia. Tertiary-era fossils contain species from this family. Reproductive adults are in the center. Soldiers have large heads and mandibles. Smaller termites are workers. (Photograph courtesy of Dr. Barbara Thorne, University of Maryland.)

Kalotermitidae

Members of this family are commonly called "dry-wood termites" for their habit of nesting in wood above the soil level, although exceptions occur. Some dry-wood termites have subterranean habits, whereas others prefer rotten and damp wood. Dry-wood termites are brownish and are considered medium-sized termites, 10–13 mm in length. This family is recognizable by the presence of ocelli and two-segmented cerci in the alate form. There are more than 400 species worldwide. Dry-wood termites are common on most continents. Colonies are moderate in size and contain several thousand individuals, most of which function as workers. The queen lays about 1 dozen or so eggs per day.

Termopsidae

The damp-wood termites nest in wet and rotting wood, especially fallen logs and stumps in forests. Damp-wood termites were formerly grouped within harvester termites (Hodotermitidae), but now are considered a separate family. Damp-wood termites are among the largest termites, some reaching almost 25 mm in length. Most individuals retain marked developmental plasticity. There are about 20 species and they are limited to forests in the Americas, Eurasia, Africa, and Australia. Egg production per queen is relatively low (≤30 per day) and colony size is moderate, up to approximately 10,000.

Hodotermitidae

Members of this grass-harvesting family of 15 species are similar in appearance to damp-wood termites and are quite large (>15 mm in length). Reproductive adults lack ocelli and their cerci have three to eight segments. Modern species are savanna grass feeders and occur in Africa, the Middle East,

and India. Queen egg production and colony size are similar to those of damp-wood termites.

Rhinotermitidae

Commonly called subterranean termites, this family typically requires its nest to contact the soil. However, exceptions occur (genera *Coptotermes* and *Recticulitermes*). Most species in this family are recognizable by their reproductive adults and a flat pronotum behind the head of soldiers. For some species, a fontanelle gland is present on the head of soldiers that produces a defensive fluid. Workers and soldiers are small (<5 mm) and are very pale white. More than 300 species are recognized. They occur on most continents except in polar and near-polar regions and are serious pests of structures. Queens of some species can produce more than 100 eggs per day, and colonies can number from the tens of thousands into the millions. Some mound-builders and aerial-nesters are found in this family. Aerial-nesting species still maintain contact to the soil for water by runways constructed from soil and saliva.

Serritermitidae

This family is very similar in appearance and closely related to subterranean termites (Rhinotermitidae). It also requires its nest to contact the soil. A single species occurs in South America. Soldiers have serrated teeth along the entire inner margin of the mandibles.

Termitidae

This family contains builders of the great mounds (up to 8 m high) that occur in the tropics, mainly in Asia, Africa, Australia, and South America. There are a few species in North America and none occur in Europe. More than 1800 species have been described, many from Africa. Termitids are distinguished by two prominent teeth on the left mandible of reproductive adults and a saddle-shaped pronotum. The Termitidae have a true worker caste. Workers are very small (<5 mm) and pale or dark in color. Many species have nasute soldiers. Members of this family are some of the most prolific producers of eggs in the animal kingdom. A queen can produce more than 10 million eggs in a single year.

TERMITE BIOLOGY AND ECOLOGY

Termites live in colonies that are social and can be long-lived. Colonies are composed of castes that conduct all tasks for survival (Figs. 1, 2, and 3). Some termite queens are larger than the length of a human thumb and can lay more than a thousand eggs per day. The king is also long-lived and mates intermittently to provide sperm to the queen. Some of the longest living insects are termites: some termite mounds and their queens are thought to be more than 70 years old and Aborigine folklore claims some mounds in Australia are over 200 years old. There are no methods to age a queen.

Termites are herbivores, fungivores (i.e., plant or fungus feeders), and humivores (soil feeders). They feed on cellulose, directly from plants, dead or alive, or indirectly from fungus arising from decaying plant material within mounds. Plants are made of cellulose, a polysaccharide that is composed of glucose units. The traditional view is that termites rely on intestinal gut microorganisms for cellulose digestion. However, there is growing evidence that termites also use their own enzymes for cellulose digestion.

Before mating and starting new colonies, new kings and queens, called alates or swarmers, depart the colony and fly (Fig. 2). They mate after landing on the ground. Swarming behavior varies considerably among termite families and species, but occurs most frequently during the rainy season. However, dry-wood termites can swarm during hot days, or sometimes evenings, of summer. A mated king and queen lose their wings and find a suitable nesting site near or in wood where they construct a small chamber that they enter and seal. The queen soon begins laying eggs, and both the king and the queen feed the young predigested food until they are capable of feeding themselves. Once workers and nymphs are produced, the king and queen are fed by the workers and cease feeding on wood. The exchange of food among colony members is called trophallaxis. Social insects exchange food in two ways, stomodaeal and proctodaeal trophallaxis. Termites use the latter method for food and symbiont exchange, mouth to anus. Symbionts are protozoa and bacteria that occur in the hindgut of termites. These microbes help digest cellulose, the major food source for termites.

The reproductive adults have functional eyes, needed for flight and initial finding of nest sites. The blind workers and soldiers live deep in nests, soil, or mounds and do not require or need vision. They already are in contact with or close to their food source.

Termites can also communicate through chemical, acoustical, and tactile signals. Two termite trail pheromones, *(Z,Z,E)*-(3,6,8)-dodecatrienol and *(E)*-6-cembrene, have been identified. These messages are produced in a sternal gland on the underside of the termite's abdomen. However, other chemical signals, such as those used for alarm and colony recognition, are produced from other glands located throughout their body. Many termite behaviors (e.g., trail following, alarm, and sexual communication) are mediated by pheromones. Soldiers also produce chemicals that are important for colony defense. Colony recognition and colony spacing are thought to be regulated by cuticular hydrocarbons. These waxy compounds are produced over the exterior cuticle of termites and spread throughout the colony. Termites can also communicate danger by "head-banging" of soldiers, in which they tap their heads in galleries to alert their nestmates.

Termites play a major role in recycling wood and plant material, but their tunneling effort also ensures that soils are

porous, aerated, and enriched in minerals and nutrients, all of which improve plant growth. For example, termite activity in the desert areas of west and north Africa helps to reclaim soils damaged by overgrazing. Termites are an important food source for many other animals, including reptiles, birds, and mammals. Termite mounds and trees hollowed out by termites provide shelter and breeding sites for birds, mammals, and other insects.

Termites also contribute to atmospheric gases. The most abundant gases produced are carbon dioxide and methane. Both are greenhouse gases, but they are not produced in sufficient quantities to have negative effects on the atmosphere.

TERMITES AS PESTS

Some termites are destructive feeders and consume homes, other wooden structures, and agricultural crops. In some regions of the world, tunneling by termites damages dams, which then results in flooding. Worldwide, several billion dollars is spent annually for the control and repair of damage caused by termites. In the United States alone, over $1 billion is spent annually for termite control and damage repairs. Globally, subterranean termites (Rhinotermitidae: genera *Reticulitermes, Coptotermes, Heterotermes,* and *Psammotermes*) are the most responsible for the control and damage costs. Dry-wood termites (Kalotermitidae: genera *Incisitermes* and *Cryptotermes*) have lesser importance as structural pests and are more prevalent in coastal, arid, or semiarid regions. Termites as agricultural pests are confined primarily to the Asia, Africa, South America, and Australia. The major pest species belong to the genera *Microtermes, Macrotermes,* and *Odontotermes* (Termitidae) in Africa and Indo-Malaysia. *Mastotermes* (Mastotermitidae) is an important pest in Australia, whereas *Cornitermes* and *Procornitermes* (Termitidae) are important pests in South America. Damage varies from superficial to killing the plant. Healthy plants can tolerate some termite damage with reduced yields. In general, exotic plants and stressed plants are most prone to termite attack.

TERMITE CONTROL/MANAGEMENT

Before termites in structures can be treated, the extent of the infestation must be assessed. Visual searching and probing of wood are the dominant means of inspection. However, the efficacy of visual searches is questionable, because structures have inaccessible areas. Several nonvisual detection methods are used, including electronic stethoscopes, dogs, methane gas detectors, and microwave and acoustic emission devices, but each of these technologies has some limitations. For subterranean termites, wood-baited monitoring stations can identify the presence and delimit the extent of colonies. Some species of subterranean termite have colonies as large as several million individuals, and these forage over an area of more than 10,000 m^2. Other termite species have much smaller colonies and forage within areas of only a few square meters. There is considerable debate about the methods and accuracy in reporting termite numbers and foraging behavior.

Termite control is most regulated in North America, Europe, and Australia. However, in many countries controlling termites is achieved by the hand removal of queens and nests, flooding nests, or drenching them with used motor oil. Soil drenches with liquid termiticides injected into the soil beneath structures to protect foundations and structural wood is the dominant control tactic for subterranean termites for several continents. Chlorinated hydrocarbon insecticides, such as chlordane, have been used extensively for subterranean termite control because of their long persistence, >30 years in the soil. Because of persistence and suspicions of health-related problems, chlordane has been removed from many markets. Chloronicotinyls and phenyl pyrazoles are new compounds marketed for termite control. The use of toxic baits (e.g., containing chitin and metabolic inhibitors) and physical barriers (sand and stainless steel mesh) for controlling subterranean termites are also gaining acceptance. Techniques to prevent infestations of subterranean termites include using wood pressure-treated with oil and water-soluble chemicals.

Surveys of pest control firms in the United States reveal that poor building practices, particularly wood in contact with soil and cracks in concrete foundations, lead to many of the subterranean termite infestations. Experimental efforts have been made to control soil-dwelling termites using biological control agents, such as argentine ants and nematodes. However, these methods have not yet been proven effective.

Dry-wood termite colonies are usually above soil level in structures, small, and difficult to detect. Treatments include whole-structure applications of fumigants (such as sulfuryl fluoride and methyl bromide) and heat. Chemicals, heat, freezing, microwaves, and electricity are used for localized or spot treatments of dry-wood termites.

See Also the Following Articles

Blattodea • Caste • Sociality • Urban Habitats

Further Reading

Abe, T., Gignell, D. E., and Higashi, M. (eds.) (2000). "Termites: Evolution, Sociality, Symbioses, Ecology." Kluwer Academic, Dordrecht.

Forschler, B. T. (1998). Subterranean termite biology in relation to prevention and removal of structural infestation. *In* "NPCA Research Report on Termites." NPCA, Dann Loring, VA.

Kofoid, C. A. (ed.) (1934). "Termites and Termite Control." University of California Press, Berkeley.

Krishna, K., and Weesner, F. M. (1969). "Biology of Termites," Vols. 1 and 2. Academic Press, New York.

Lewis, V. R., and Haverty, M. I. (1996). Evaluation of six techniques for control of the western drywood termite (Isoptera: Kalotermitidae) in structures. *J. Econ. Entomol.* **89,** 922–934.

Pearce, M. J. (1997). "Termites: Biology and Pest Management." CAB Int., Oxon, UK.

Potter, M. F. (1997). Termites. *In* "Mallis Handbook of Pest Control" (A. Mallis, ed.), 8th ed., pp. 232–333. Franzak & Foster, Cleveland.

Su, N.-Y., and Scheffrahn, R. H. (1990). Economically important termites in the United States and their control. *Sociobiology* **17,** 77–94.

Su, N.-Y., and Scheffrahn, S. H. (2000). Termites as pests of buildings. *In* "Termites: Evolution, Sociality, Symbioses, Ecology (T. Abe, D. E. Bignell, and M. Higashi, eds.), pp. 437–453. Kluwer Academic, Boston.

Thorne, B. L. (1998). Biology of subterranean termites of the genus *Reticulitermes. In* "NPCA Research Report on Subterranean Termites." NPCA, Dunn Loring, VA.

Thorne, B. L., Russek-Cohen, E., Forschler, B. T., Breisch, N. L., and Traniello, J. F. A. (1996). Evaluation of mark–release–recapture methods for estimating forager population size of subterranean termite (Isoptera: Rhinotermitidae) colonies. *Environ. Entomol.* **25,** 938–951.

Watanabe, H., Noda, H., Tokuda, G., and Lo, N. (1998). A cellulase gene of termite origin. *Nature* **394,** 330–331.

Wood, T. G., and Pearce, M. J. (1991). Termites in Africa: The environmental impact of control measures and damage to crops, trees, rangeland and rural buildings. *Sociobiology* **19,** 221–234.

Japanese Beetle

David W. Held and Daniel A. Potter
University of Kentucky

The Japanese beetle, *Popillia japonica,* is among the most polyphagous of plant-feeding insects. The adults skeletonize the foliage, or feed on the flowers or fruits, of nearly 300 species of wild or cultivated plants. Favored hosts include many woody and herbaceous landscape plants, garden plants, fruits, and field crops. The larvae, or grubs, develop in the soil where they feed on roots of turf and pasture grasses, vegetables, nursery seedlings, and field crops. Hundreds of millions of dollars are spent annually for controlling the adults and grubs, and in state and federal regulatory efforts aimed at limiting the beetle's rate of spread in the United States and elsewhere.

The Japanese beetle was first discovered in the United States in 1916, near Riverton, New Jersey. How it was inadvertently transported from its native Japan is not known; however, the grubs may have arrived in soil around the roots of nursery plants. The species is not a major pest in Japan, where suitable grassland habitat is limited and natural enemies keep this beetle in check. The eastern United States, however, provided a favorable climate, with abundant moist turf as habitat for the eggs and larvae, numerous adult food plants and, at that time, no host-specific natural enemies. Populations increased and spread rapidly. By 2000, the beetle was established in all states east of the Mississippi River except for Florida, and in parts of Wisconsin, Minnesota, Iowa, and Nebraska. It also has spread north into southern Ontario and Quebec, Canada.

DESCRIPTION

Japanese beetles belong to the family Scarabaeidae, subfamily Rutelinae. Adults are broadly oval, 8 to 11 mm in length, metallic green, with coppery brown elytra that do not quite cover the end of the abdomen (Fig. 1A). The abdomen bears five patches of white hairs on either side, and another pair near its tip. Females, which tend to be slightly larger than males, have an elongate, spatula-shaped spur on the foretibia, used for digging. This spur is shorter and pointed in males.

Larvae are typical scarabaeiform grubs: C-shaped, grayish to cream colored, with three pairs of jointed legs, a distinct yellow-brown head capsule, and chewing mouthparts (Fig. 1B). Neonate grubs are about 1.5 mm in length, whereas the length of full-sized third instars is about 32 mm. The underside of the last abdominal segment, just anterior to the anal slit, bears two short rows of hairs forming a tiny, truncated V. This pattern distinguishes Japanese beetle larvae from the larvae of other common scarabs. The end of the abdomen appears dark because of ingested soil and food.

LIFE CYCLE

Japanese beetles have a one-year life cycle in most parts of their range. Adults occur from June to August. Upon emergence

FIGURE 1 (A) Japanese beetles with characteristic feeding damage. (B) Japanese beetle grub.

from the soil, virgin females emit a volatile sex pheromone that attracts clusters of males. Subsequent matings occur on food plants. The beetles typically feed from the upper surface of leaves, chewing out the tissue between the veins and leaving a lacelike skeleton. Adults also feed on petals of flowers such as roses, and on developing fruits or berries. Food plants growing in sunny locations are preferred. Usually the beetles begin to feed on foliage near the top of a plant, regardless of its height. They often aggregate on particular shoots or plants. This phenomenon results from both sexes being attracted to blends of aromatic volatile compounds released from beetle-damaged leaves. Despite the beetles' broad host range, some plant species are rarely or never fed upon. Closely related cultivars within species may also differ in susceptibility. Resistance probably results from presence of feeding deterrents (e.g., certain phenolics) or other secondary plant compounds. Some plants (e.g., geranium, *Pelargonium hortorum*) are palatable to the beetles but cause paralysis or other toxic effects.

After feeding, gravid females fly to moist turf, pasture, or agricultural fields, where they burrow down to lay small clutches of eggs in the upper 8 cm of soil. Females alternate between feeding and egg laying; each female may emerge from the soil, fly to host plants, feed, mate again, and return to the soil 15 or more times, laying 40 to 60 eggs in her lifetime. The pearly white eggs, oval when first laid, swell with soil moisture to a diameter of about 1.5 mm. Larvae hatch in 2 to 3 weeks, usually by early to mid-August. Larvae feed just below the soil surface, consuming plant roots and organic matter. When grubs are numerous, the root system of turf grasses may be completely severed, such that the turf wilts and dies, and can be pulled from the soil like a loose carpet. Most grubs are third instars by September. About the time of first frost, the grubs move deeper (about 15–30 cm) to overwinter. In early spring, as soil temperatures warm to about 10°C, the grubs move back to the upper 2.5 to 5.0 cm of soil and resume feeding for about 4 to 6 weeks, after which they again go deeper and form an earthen cell in which to pupate. The first adults begin emerging a few weeks later.

PREDATORS AND PARASITIZERS

Vertebrate predators such as skunks, raccoons, birds, and moles may dig in infested areas to feed on the grubs. Indigenous predatory insects, including ants and ground beetles, feed on the eggs and young grubs. Birds, fish, and other insectivores eat the adults. From 1920 to 1933, entomologists searched for, and imported, numerous natural enemies from Asia and released them for biological control in areas infested with Japanese beetles. Only a few of these became established. The most widely distributed are two species of tiphiid wasps, *Tiphia vernalis* and *T. popilliavora,* whose larvae are ectoparasitoids of the grubs, and *Istocheta aldrichi,* a tachinid fly that parasitizes the adults. The grubs are susceptible to parasitic nematodes, as well as several lethal microbial pathogens, especially the milky disease bacterium, *Paenibacillus popilliae.*

See Also the Following Articles

Introduced Insects • Regulatory Entomology • Soil Habitats

Further Reading

Fleming, W. E. (1968). Biological control of the Japanese beetle. U.S. Department of Agriculture Technical Bulletin 1383.

Fleming, W. E. (1972). Biology of the Japanese beetle. U.S. Department of Agriculture Technical Bulletin 1449.

Loughrin, J. H., Potter, D. A., Hamilton-Kemp, T. R., and Byers, M. E. (1996). Role of feeding-induced plant odors in aggregative behavior of the Japanese beetle. *Environ. Entomol.* **25,** 1188–1191.

Potter, D. A., and Held, D. W. (1999). Absence of food-aversion learning by a polyphagous scarab, *Popillia japonica,* following intoxication by geranium, *Pelargonium hortorum. Entomol. Exp. Appl.* **91,** 83–88.

Potter, D. A., and Held, D. W. (2002). Biology and management of the Japanese beetle. *Annu. Rev. Entomol.* **47,** 175–205.

Vittum, P. J., Villani, M. G., and Tashiro, H. (1999). "Turfgrass Insects of the United States and Canada." 2nd ed. Cornell University Press, Ithaca, NY.

June Beetles

Daniel A. Potter and David W. Held
University of Kentucky

June beetles, sometimes called May beetles or june bugs, are heavy-bodied, brownish, plant-feeding scarab beetles (Fig. 1A). Almost all species are nocturnal in their habits. The adults are voracious feeders on leaves of many deciduous trees, shrubs, and some herbaceous plants. Their larvae, called white grubs, develop in the soil, where they feed on plant roots and can be pests of turf and pasture grasses, young nursery stock, corn, small grains, potatoes, strawberries, and other agricultural crops.

June beetles belong to the genus *Phyllophaga* (formerly *Lactosterna*) in the family Scarabaeidae, subfamily Melolonthinae. They occur in both the New and Old Worlds. In North America north of Mexico about 200 species are known, with many found in the north-central and eastern United States. They also have been reported from South and Central America, the West Indies, eastern and southern Asia, and the islands of the Pacific and Indian Oceans.

DESCRIPTION

June beetles average from 12 to over 25 mm in length, with a cylindrical or oblong body shape, dense hair on the metasternum, and lamellate antennae that end in a three-segmented club that is longer in males than in females. Each tarsal claw bears a small tooth near the middle. Coloration ranges from tan to mahogany to dark chocolate brown. The elytra of some species are hairy, whereas in others they are nearly smooth.

Larvae are typical scarab grubs: cream colored, C-shaped when feeding or at rest, with a brown head capsule, with

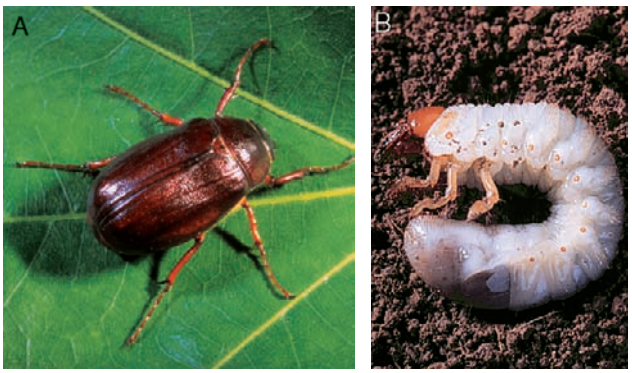

FIGURE 1 (A) A typical June beetle, *Phyllophaga* sp. (B) June beetle larva, or white grub.

chewing mouthparts and three pairs of jointed legs (Fig. 1B). The hind part of the abdomen usually appears dark because of ingested food and soil. The ventral surface of the last abdominal segment bears two parallel rows of short spines in a pattern that resembles a zipper. There are three instars. Full-sized grubs of most species are 25 to 38 mm in length.

LIFE CYCLE

Most June beetles have 2- or 3-year life cycles, although a few species have cycles lasting 1 or 4 years. Adults typically are active from April to June. The beetles emerge after sundown and fly to the tops of trees to feed and mate, returning to the soil before dawn. They are clumsy fliers and often are attracted to outdoor lights. Mated females fly to turf, pasture, or agricultural fields and burrow down 5 to 15 cm to lay eggs in the soil. Each female lays 20 to 50 eggs in her lifetime. Eggs are pearly white, about 2.5 mm long, and elliptical at first, becoming more spherical as the embryo develops. Hatching occurs in about 3 weeks, and the young grubs begin feeding on fine roots and organic matter.

Larvae of species with 2-year cycles typically overwinter as second instars. They resume feeding in early spring, molting again in April or May. Third instars attain their full size by summer's end. Pupation occurs underground, in an earthen cell. Most species transform to adults by late autumn, but the beetles remain underground until the following spring. Grubs of species with 3-year cycles feed throughout the first two summers, hibernating twice and pupating midway through the third summer. Their adults are usually fully formed by autumn but do not emerge from the soil until the following spring. Because of overlapping generations and presence of more than one species, several sizes of June beetle grubs may be found together at a given site.

ELIMINATING JUNE BEETLE GRUBS

June beetle grubs are susceptible to various microbial pathogens, including specific strains of the milky disease bacterium *Paenibacillus popilliae.* Wasps and flies of several kinds parasitize the larvae or beetles, and predatory insects (e.g., ants, carabids) feed on the eggs and young larvae. The grubs also attract vertebrate predators, including insectivorous birds, skunks, raccoons, moles, and armadillos. In the past, farmers were advised to clean June beetles out of heavily infested fields by pasturing the land with hogs, which would root out and eat the grubs. Today this objective is more typically accomplished through crop rotation, or with soil insecticides.

See Also the Following Articles

Coleoptera • Hibernation • Soil Habitats

Further Reading

King, A. B. S. (1984). Biology and identification of white grubs of economic importance in Central America. *Trop. Pest Manag.* **30,** 36–50.

Luginbill, P., and Painter, H. R. (1953). May beetles of the United States and Canada. U.S. Department of Agriculture Technical Bulletin 1060.

Pike, K. S., Rivers, R. L., Oseto, C. Y., and Mayo, Z. B. (1976). A world bibliography of the genus *Phyllophaga.* University of Nebraska Miscellaneous Publication 31.

Vittum, P. J., Villani, M. G., and Tashiro, H. (1999). "Turfgrass Insects of the United States and Canada." 2nd ed. Cornell University Press, Ithaca, NY.

Ritcher, P. O. (1966). "White Grubs and Their Allies." Oregon State University Press, Corvallis.

Juvenile Hormones

Michael E. Adams
University of California, Riverside

Juvenile hormones (JHs), acting in concert with ecdysteroids, orchestrate the expression of larval-specific genes each time the insect molts to a new stage. These morphogenetic effects include determination of an immature body form and internal organs, hardness and color of the cuticle, and accompanying physiology and behaviors. At the conclusion of larval development, juvenile hormone levels drop at critical times, allowing ecdysteroids to program expression of pupal and adult characteristics. Juvenile hormones return in the adult stage, in which they have gonadotropic functions in connection with reproduction. In addition to their morphogenetic and gonadotropic actions, juvenile hormones are involved in dormancy and various types of polyphenisms, including caste determination in social insects.

DISCOVERY

The discovery of juvenile hormones began in the 1930s, with a series of ingenious experiments conducted by Sir Vincent Wigglesworth aimed at elucidating the hormonal control of molting. Trained as a medical doctor, Wigglesworth dedicated his life to basic studies of insect physiology, believing that knowledge gained would hold the keys to controlling insect

vector-borne disease and agricultural pests. As a model experimental insect, Wigglesworth chose the Chagas' disease vector *Rhodnius prolixus,* otherwise known as the "kissing bug" because of its habit of sucking blood from the lips of sleeping humans. The choice of *Rhodnius* was inspired, because its development is closely timed to its blood meals. This allowed Wigglesworth to precisely determine the physiological stage of the insects to coincide with his experimental manipulations. He found that 3 days after a blood meal, hormones are released into the *Rhodnius* system, stimulating the molt to the next stage. By performing a number of surgical procedures on the bug he demonstrated the source and timing of hormone release. Part of the advantage of working with insects as experimental animals is that they can survive for long periods without such seemingly vital organs as the brain, a fact that Wigglesworth took advantage of. He found that decapitation of animals prior to a 3-day critical period led to an arrest in development, even though the animal would remain alive for many months. If the brain was reimplanted, development resumed. He also found that the blood of a normally developing animal could reactivate development in the headless animal. This was achieved via a technique called parabiosis, in which the developmentally arrested animal was joined to the normal one by means of a tube, which allowed blood from the two animals to mix. With these experiments, Wigglesworth demonstrated that hormones released from the brain trigger molting. This discovery actually had been made more than a decade earlier by Stefan Kopec, working with gypsy moth, but Wigglesworth's experiments revealed a new type of hormone, one that influenced the form taken by the animal after each molt.

Wigglesworth fundamentally changed the thinking about insect development, specifically the distinction between regulation of growth and regulation of form by separate hormones. *Rhodnius* passes through five nymphal stages before molting to the adult form. It is easy to tell the adults from the nymphs, because of differences in pattern and color of the cuticle, as well as the fact that only adults have wings. Wigglesworth found that parabiosis of a fifth (last)-stage nymph with a young nymph prevented the former's metamorphosis to the adult stage. Instead, the animal molted to a sixth-stage nymph, an extra immature stage that never occurs normally. A chemical in the blood of the young insect promoted continued expression of larval characters, and this factor came to be known as the "juvenile hormone." The source of the juvenile hormone was traced to a pair of small glands behind the brain called the corpora allata (Fig. 1). Surgical removal of the gland did not interfere with molting, but drastically altered the form taken after the molt, causing animals to become precocious adults. Reimplantation of the glands led to the return of larval characters.

While the corpora allata proved to be the sole source of juvenile hormone, only very small amounts were available from the gland for chemical studies. The short supply of JH greatly constrained experimentation, slowing the process of discovery considerably. A breakthrough came with the discovery of large amounts of JH in abdomens of adult male silk moths by Carroll Williams. Ether extracts produced a dark orange material he called the golden oil. Such an abundance of juvenile hormone in a male adult at first was surprising, but already Wigglesworth had noted the essential gonadotropic role of JH associated with reproduction, that is, stimulation of egg and sperm development. The reason for enormous quantities of JH in male adult abdomens is likely caused by its inclusion in spermatophores, which contain sperm together with nutritive and hormonal stores and are provided to the female during mating for fertilization and nutrition of developing eggs. Williams is credited with stimulating the modern era of JH research, by making available enough of the natural hormone to conduct biological experiments on its modes of action in many types of insects. This work also provided quantities of starting material sufficient for the eventual isolation and chemical identification of the hormone, which occurred in the late 1960s and early 1970s.

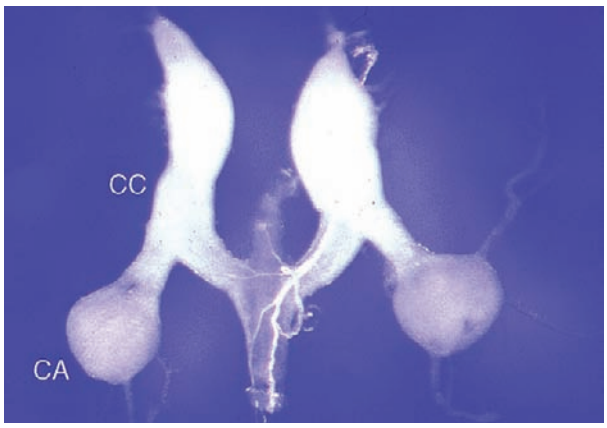

FIGURE 1 Photomicrograph of the corpora allata (CA), paired, spherical glands that are the sole source of the juvenile hormones in insects. Also shown are the elongated, white corpora cardiaca (CC). The CC and CA are positioned behind the brain, where they release hormones into the blood. Structures shown were dissected from the cockroach, *Periplaneta americana.* (Photograph courtesy of Dr. S. J. Kramer.)

CHEMISTRY OF JUVENILE HORMONES

The juvenile hormones are lipophilic sesquiterpenoid derivatives of farnesoic acid. Their chemical nature came into focus in the early 1960s, beginning with observations that farnesoid components of beetle excreta had JH-like activity in bioassays. Although far less potent than the native hormone, these substances, including farnesol and its oxidized form farnesal, were suspected to be JH precursors. Subsequent synthesis of methyl farnesoate with an epoxide at position 10–11 by William Bowers in 1965 gave a highly potent compound. It is indeed ironic that this compound was discovered 8 years later to be the most ubiquitous of the natural juvenile hormones, JH-III (Fig. 2).

Another interesting prologue to the discovery of the natural juvenile hormones was the discovery of a curious

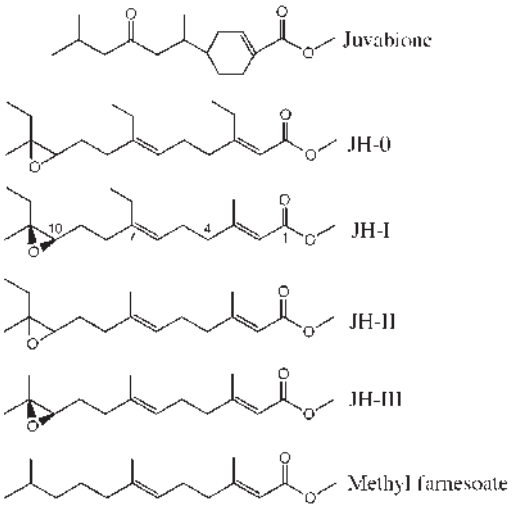

FIGURE 2 Structures of natural juvenile hormones and related compounds. Juvabione is the paper factor from North American balsam fir discovered by Williams, Slama, and Bowers. JH-I was the first natural JH discovered by Röller, and JH-II and JH-III soon followed. JH-III is present in most groups of insects, including the Lepidoptera. JH-0 is found in moth eggs, but its biological function is unclear. JH-0, JH-I, and JH-II are generally confined to the Lepidoptera. Methyl farnesoate, a JH precursor, has been isolated from some insects and crustaceans, in which it may serve as the active JH.

"paper factor" by Karel Slama and Carroll Williams in the mid-1960s. Slama had carried insects from his laboratory in Czechoslovakia to the laboratory of Williams at Harvard University for a series of joint experiments. Some weeks after arriving the insects began to develop extra stages and many died. Nothing of the sort was noticed in Czechoslovakia. It was eventually determined that the paper towels used to line the containers holding the insects contained compounds with JH-like biological activity. The substances were absorbed through the insect cuticle upon contact with the paper towels. The paper factor turned out to be a mixture of terpenoids in the wood pulp from which the paper towels were manufactured. These compounds were found only in American and Canadian balsam fir and not in European trees. One of these substances was chemically identified by Bowers and colleagues as "juvabione" (Fig. 2), which had a structure reminiscent of farnesol.

These studies provided new information on two critical issues of the time. First was the question of juvenile hormone structure. Observations that farnesol and methyl farnesoate, along with juvabione, possessed JH-like biological activity made it likely that terpenoid chemistry was involved. This guided further attempts to chemically define the natural material(s). The second issue was whether juvenile hormones or analogs such as juvabione could be used to produce, as Carroll Williams suggested, a new class of "third-generation" pesticides. Juvabione constituted a relatively stable terpenoid with potent insecticidal activity and hence stimulated further interest in this novel concept for new insecticides that would be both highly insect-selective and safe for warm-blooded animals.

With this as background, Röller and colleagues isolated sufficient quantities of juvenile hormone from silk moth abdomens for identification of the first natural juvenile hormone in 1967. The carbon skeleton was identified as a 15-carbon sesquiterpene substituted at positions 7 and 11 with ethyl groups. Further key structural features were the presence of a methyl ester and an epoxide at carbons 10–11. This landmark achievement was followed rapidly by publication of a second JH, also from moths. These molecules were named JH-I and JH-II, respectively, differing only at carbon 7, which was ethylated in JH-I and methylated in JH-II (Fig. 2). These two juvenile hormones are largely restricted to the Lepidoptera. Within a few years, a third hormone was identified by a completely different approach. In this instance, corpora allata from the moth *Manduca sexta* were removed and placed in organ culture containing a radiolabeled methyl donor, ^{14}C-labeled methionine. The corpora allata incorporated the ^{14}C-methyl group into the ester moiety of juvenile hormone. This isotope-labeled synthetic product was isolated and identified as JH-III. It is the most cosmopolitan of juvenile hormones, occurring in most insect groups, including the Lepidoptera. It has methyl groups at positions 7 and 11 (Fig. 2). Three additional JH structures have been identified: JH-0 and 4-methyl JH from moth eggs and JH-bisepoxide from the fruit fly *Drosophila.*

The juvenile hormones are derived from acetyl CoA and/or propionyl CoA via mevalonic acid and homomevalonic acid in the sterol biosynthetic pathway. The final steps of JH-III biosynthesis go by way of farnesol → farnesoic acid → methyl farnesoate, to which an epoxide is formed at carbons 10–11. Owing to their low aqueous solubility, the juvenile hormones are transported through the blood via binding proteins upon their release from the corpora allata. These binding proteins also protect JH from degradative enzymes.

In some insects, the corpora allata synthesize a precursor of the biologically active form of JH, which is converted to the active form in target tissues. For example, silk moth adults produce JH acid in the corpora allata and convert it to JH-I in the accessory glands of the abdomen. It is also known that the ovaries of certain species of mosquito can synthesize JH from precursors under culture conditions. Whether this occurs under natural conditions *in vivo* has not been demonstrated.

The levels of JH in the blood are regulated through a combination of synthesis and degradation. Synthesis by the corpora allata is promoted by neuropeptides called allatropins. So far, only one allatotropin has been identified from the tobacco hawkmoth *M. sexta.* Surprisingly, the peptide is active only in the adult stage. More compelling evidence has been provided for the existence of allatostatins. These are neuropeptides synthesized in brain neurons that project to the corpora allata. Their release from nerve endings in the gland inhibits the synthesis of juvenile hormone.

The removal of juvenile hormones already in the blood, a necessary condition for metamorphosis, occurs through two enzymatic degradation pathways. One is through cleavage of

the ester bond by JH esterase, the other through epoxide destruction by epoxide hydrolases.

BIOLOGICAL ACTIONS

Morphogenetic Effects

The presence of juvenile hormones in the blood promotes expression of juvenile characters, chief among these being an immature body form or morphology. For insects such as grasshoppers, which undergo incomplete metamorphosis, the effects are not so visible outwardly. The early stages look like miniature adults except for the absence of wings, but they also lack functional reproductive organs. However, for those insects such as flies, moths, and bees, which undergo complete metamorphosis, the effects are extreme. The immature stages are wormlike with no wings or legs.

Determination of holometabolous immature or larval body plan by juvenile hormone represents its morphogenetic action. The decision to develop larval characters during development is made near the end of each larval stage, when ecdysteroid levels increase to initiate the molt. Elevation of ecdysteroids causes a cessation of feeding and new round of gene expression appropriate for the next stage of development. If juvenile hormones are present at this time, genes for larval characters are expressed, whereas genes appropriate for pupal or adult characters are repressed. A primary larval character is the type of cuticle secreted by the epidermal cells. Larval cuticle is lighter and more flexible than pupal or adult cuticle, which are characteristics resulting from expression of larval cuticle protein genes that predominate under the influence of juvenile hormones. The flexibility of larval cuticle has to do with the absence of cross-linking between proteins and between proteins and chitin, the latter constituting the polysaccharide component of the cuticle. In contrast, pupal and adult cuticles are hard and dark, indicating a high degree of sclerotization and melanization.

Specification of pupal or adult features by JH, or lack thereof, is associated with transient, hormone-sensitive periods during development. It is important to note that the actions of JH depend not only on its presence in the blood, but also on the ability of cellular targets to respond. This latter condition presumably reflects the presence of suitable receptors required to mediate the action of the hormone. It has been observed in many studies that JH responsiveness occurs after priming, which could be associated with expression of receptor genes or other molecules necessary to complete the signaling pathway. For example, in moths, specification of a wandering period in preparation for pupation occurs upon the appearance of ecdysteroid peaks in the complete absence of JH. These "commitment" peaks of ecdysteroid prime the system to respond later in the same stage to elevated JH levels, which specify pupal features.

In many insects, considerable development of the gonads takes place during the larval stages. This is especially true for insects such as the silkworm *Bombyx mori,* which does not feed during the adult stage. Within hours of emergence, these animals mate and lay eggs. The ability to mate and produce viable eggs so soon after emergence means that gonadal development is well along during the larval and pupal stages. Juvenile hormones promote the development of gonads and gametes during the immature stages, but must disappear in order for final developmental steps to be completed. This drop in JH levels just prior to the pupal stage therefore serves both morphogenetic and gonadotropic functions.

Effects of JH in the Adult Stage

The decrease in JH levels just prior to metamorphosis is only a temporary condition. The corpora allata are retained in the adult stage, and JH eventually reappears to regulate adult reproductive functions. JH promotes sperm and egg development and hence is said to have "gonadotropic" functions. In the female, JH directly promotes the synthesis of lipo- and glycoproteins in the fat body and their uptake into the developing oocyte. This process, called vitellogenesis, is essentially yolk deposition. In many insects, JH levels rise and fall in a cyclic fashion as discrete batches of oocytes go through the vitellogenic process.

In other instances, as in some mosquitos, JH exposure leads to "competence" of the fat body to synthesize vitellogenic proteins upon later exposure to ecdysteroids. Likewise, JH exposure is required to induce competence of the ovaries to respond later to peptide hormones from the nervous system, thus stimulating uptake of vitellogenic proteins. In these instances, the gonadotropic actions of JH appear to be priming steps in preparation for ecdysteroid action.

Gonadotropic functions of JH in the male have to do with growth of the sperm. Sperm growth requires JH in many insects. However, maturation from spermatocytes to motile spermatids requires a drop in JH. As observed for oocyte development, JH exerts both positive and negative influences in sperm development.

Polyphenism and Caste Determination

Many insects have the remarkable ability to develop into alternate forms as they become adults. These alternate forms together with accompanying physiology and behavior, referred to as polyphenism, do not reflect differences in the genetic makeup of individuals. Rather, they result from a particular pattern of gene expression under hormonal control. Most polyphenisms are controlled by juvenile hormones acting at certain sensitive periods during immature development.

Some of the most common instances are caste polyphenisms observed in social insects such as bees, ants, and termites. In these insect societies, larvae can develop into workers, soldiers, or queens, depending on the diet they are raised on and the hormonal levels that result. If bee larvae are reared in a special cell in the hive and consistently fed a nutritious "royal jelly" beginning during the third instar, they develop into queens.

Treatment with JH will mimic this effect. If this feeding is delayed, larvae develop into workers instead. In certain ants, development of queens is regulated during embryonic development by JH levels. During postembryonic development, larvae fed a high-protein diet produce large amounts of JH, bringing blood levels to a threshold necessary for specification of soldier phenotype. If larvae are fed a diet lower in protein, JH levels are correspondingly lower, and development to worker is specified. The number of soldiers in the colony also is determined by a soldier-inhibiting pheromone, which elevates the JH threshold for soldier specification. Alternative body forms and behaviors in insect colonies provide for cooperative functions between members of the society to serve the greater whole.

Many types of phase polyphenisms occur in nonsocial insects. For example, locusts occur either in solitary or in migratory phases, depending on population density. Differences in both behavior and physiology are characteristic of these phases. Solitary locusts are sedentary, pale green, yellow, or brown, and have short wings and large ovaries. Crowding causes the switch to the gregarious phase, in which individuals are brightly colored, have longer wings and smaller ovaries, and are easily induced to engage in long flights. Both JH and peptide neurosecretory hormones from the brain are involved in the determination of these two phases.

Aphids exhibit at least two different types of phase polyphenism as a response to seasonal conditions: food quality and crowding. In one type, adults switch between winged or apterous (no wings) forms. The other type has to do with the mode of reproduction, either sexual or parthenogenetic. During the longer days of spring and summer, apterous, parthenogenetic females predominate, and juvenile hormone is involved in specification of these forms. As winter approaches, winged forms are produced, allowing for dispersal. Later, in autumn, males and females mate and lay eggs, which overwinter and hatch in the spring. In this context, body forms and accompanying dispersal or migratory behaviors maximize survival as the season changes.

In summary, JH and other neurosecretory hormones are important determinants of polyphenisms, which result in different body forms, reproductive physiologies, and behaviors in the adult stage. It is emphasized that such variability of form and function is not the result of genetic differences between individuals, which would be classified as polymorphisms. Rather, insects have the enormous potential to change form in response to environmental conditions through hormonal control mechanisms. Depending on the needs of a social colony, or changes of season and in food availability, the complex endocrinology of insects enables them to assume various alter egos to enhance success and survival.

Behavioral Effects of JH

The presence or absence of juvenile hormones has profound effects on behavior, some of which have been mentioned above. Throughout the stages of immature development, JHs program gene expression in the nervous system for the expression of behaviors appropriate for juvenile life, including, for example, locomotion, host or prey seeking, feeding, and silk spinning. From the point of view of behavior, the larva is an animal completely different from its later adult form.

In moths, the disappearance of JH at the end of the last instar allows ecdysteroids to program new behaviors appropriate for metamorphosis. Insects stop feeding, void their guts, and engage in wandering behaviors to locate a suitable pupation site. This accomplished, a series of behaviors leads to silk spinning for cocoon construction.

Upon becoming adults, female mosquitos initiate the search for a blood meal and become sexually receptive to males only after release of JH into the blood. In milkweed bugs, JH levels are influenced by daylength, temperature, and food quality. Under short daylengths, JH levels drop, and insects engage in migratory behavior immediately after molting to the adult stage. However, long days and warm conditions lead to high JH levels, whereby flight is inhibited and reproduction ensues.

Grasshopper females that have had corpora allata removed rebuff male sexual advances until JH is reintroduced by injection. In crickets, the male sings a species-specific calling song to attract the female for mating. The female responsiveness to this song is enhanced by elevated JH levels. These examples serve to illustrate the dramatic effects that JH has on the behavior of insects, effects that are specific and appropriate for each particular life stage.

Dormancy–Diapause

Insects are able to enter prolonged states of dormancy referred to as diapause, allowing them to resist freezing and low food supplies during the winter. Diapause can occur at any stage (egg, larva, pupa, or adult) and is triggered by decreasing daylength, low temperatures, decreased food or food quality, or a combination of these factors. The insect response to these environmental factors is mediated by a variety of hormones, depending on the stage and species.

Adult diapause is largely synonymous with reproductive diapause. Beetles, butterflies, and flies enter a reproductive diapause when the brain inhibits synthesis of JHs by the corpora allata. The lack of JHs leads to both physiological and behavioral changes, including cessation of vitellogenesis, loss of flight muscle, increasing stores of lipid in the fat body, burrowing, and construction of hibernacula (overwintering chambers). Implantation of corpora allata or injection of JHs reverses reproductive diapause.

JH involvement in larval diapause also has been documented. The southwestern corn borer *Diatraea grandiosella* enters diapause during the last instar when JH levels are depressed but are still high enough to inhibit development to the pupal stage. The animal spins a hibernaculum, exhibits a light pigmentation, and actually undergoes several "stationary" molts. Diapause in this stage lasts as long as JH levels remain elevated.

MOLECULAR BASIS OF ACTION

It is presumed that juvenile hormones exert their effects through receptor activation. However, it is a curious and surprising fact that, almost 70 years after its discovery by Wigglesworth, no definitive receptors have been identified. Even though no clear example of a JH receptor has been defined, a number of tantalizing possibilities have been suggested. One of these involves a nuclear receptor called ultraspiracle, or USP, a protein that is well known to regulate gene expression by forming a dimer complex with ecdysteroid receptors. The complex then binds to regulatory sequences on genes to turn them on or off. Grace Jones and Alan Sharp have demonstrated that JHs bind specifically to USP, although the affinity for this binding is lower than is generally expected for hormone–receptor interactions. It is proposed that JH binding to USP may influence how it interacts with ecdysteroid receptors (EcR) to regulate gene expression. The influence of JH and USP on EcR actions would seem to be a very plausible scenario for joint actions of JH and ecdysteroids, but further work is needed before USP is confirmed as a JH receptor.

Although early accounts of juvenile hormones focused on their uniqueness with respect to insect biology, the elusiveness of JH receptors has prompted a closer look at possible similarities between signaling mechanisms common to insects and mammals. Indeed, the chemical structure of JHs resembles those of retinoids and farnesoids, both of which function in mammalian nuclear signaling by activating retinoic acid receptors, retinoid X receptors, and the farnesoid X receptor. JH and farnesoids are capable of activating some of these receptors, and some retinoids are known to have JH-like activity. It also has been observed that vertebrate thyroid hormones mimic some of the actions of JHs. Efforts are under way to identify receptors homologous to their mammalian counterparts as possible JH receptors.

Recently, workers taking a genetic approach to the problem identified a strain of fruit flies resistant to methoprene, an insecticidal juvenile hormone analog (see next section for details). The resistant flies have a defect in a gene that encodes MET, a protein related to the vertebrate aryl hydrocarbon receptor, which upon binding a diverse range of hydrocarbons activates a battery of genes involved in their metabolism. If the MET protein has similar properties, this might help explain why many synthetic chemicals such as fenoxycarb and pyriproxyfen have very potent JH-like effects, but bear little obvious structural similarity to natural JHs.

The failure after so many years to define a receptor for JH may indicate that, for this particular hormone, signaling does not conform to conventional modes of action. Perhaps JH binds to certain proteins, which then act as coeffectors or adaptor proteins to amplify or modify transduction of signals initiated by other hormones at conventional receptors. The obvious example is modification of ecdysteroid receptor action. It turns out that MET also is related to steroid receptor coactivators, which could bind to EcR and/or USP to modify their effects on gene expression. Although little is known about the specific actions of MET at the present time, the MET resistance gene may hold the key to understanding the elusive molecular action of JHs.

JUVENILE HORMONES AND INSECT CONTROL

The discovery of juvenile hormones in the late 1960s by Röller and others stimulated a period of great excitement regarding the concept of third-generation pesticides foreseen by Carroll Williams in the early 1960s. It was known that juvenile hormones and related substances such as juvabione could disrupt insect development with lethal effects. Likewise, surgical removal of the corpora allata led to precocious metamorphosis, also with lethal effects. It therefore seemed that insect hormones or their analogs could be synthesized and used to accomplish a form of "birth control" for insects.

This idea occurred to Carl Djerassi and Alejandro Zafferoni, two former colleagues at the Syntex Corporation, who were involved in synthesis of the first human contraceptives that led to the birth control pill. They formed a new company called Zoecon, a name chosen to denote "animal control" through the use of hormones and related chemical analogs. Their principal objective was to develop insect hormones for use as birth control agents specific for this group of animals. Chemists at Zoecon soon produced analogs of JH called "juvenoids" that were much more stable and could penetrate the cuticle. One of the first of these analogs to be granted a registration from the Environmental Protection Agency was methoprene (Fig. 3), a compound with outstanding biological activity against mosquitos, fleas, and biting flies. By mimicking JH, methoprene prevents treated insects from completing metamorphosis, and insects die during the pupal stage. Other juvenoids such as hydroprene (Fig. 3) are more effective against insects with incomplete metamorphosis, such as cockroaches. Treated cockroaches actually reach the

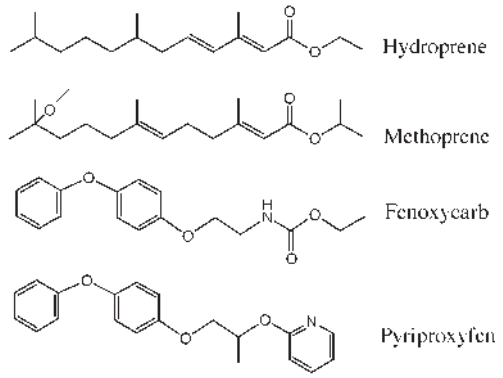

FIGURE 3 Structures of synthetic juvenile hormone analogs, commonly called juvenoids. Methoprene has been useful in the control of mosquitos, fleas, and biting flies, while hydroprene was developed for cockroach control in dwellings. Fenoxycarb and pyriproxyfen, heterocyclic compounds with little resemblence to JHs, nevertheless have potent JH-like biological activity against a wide range of insects.

adult stage, but the presence of juvenoid during the transition to the adult results in only a partial adult phenotype in which many adult features are abnormal. For instance the gonads are not fully developed, leading to sterility, and crinkled wings are an obvious morphological defect.

Subsequent efforts by several agrochemical companies have generated a variety of juvenoids, many with potent JH-like biological activity but with aryl rings substituted for isoprenoid units and without obvious similarities to natural JHs. These include fenoxycarb and pyriproxyfen (Fig. 3). Juvenoids have proved to be commercially successful for insects that are pests in the adult stage. However, because they do not control insects in the immature stages, they have not proved useful for large-scale agricultural pest control. For this purpose, juvenile hormone antagonists are needed for induction of precocious metamorphosis. So far, that goal remains as elusive as the search for the JH receptor.

See Also the Following Articles

Development, Hormonal Control of • *Dormancy* • *Ecdysteroids* • *Hibernation* • *Mating Behaviors* • *Migration* • *Molting* • *Vitellogenesis*

Further Reading

Ashok, M., Turner, C., and Wilson, T. G. (1998). Insect juvenile hormone resistance gene homology with the bHLH-PAS family of transcriptional regulators. *Proc. Natl. Acad. Sci. USA* **95,** 2761–2766.

Feyereisen, R. (1998). Juvenile hormone resistance: [No PASaran! *Proc. Natl. Acad. Sci. USA* **95,** 2725–2726.

Gilbert, L. I., Granger, N. A. and Roe, R. M. (2000). The juvenile hormones: Historical facts and speculations on future research directions. *Insect. Biochem. Mol. Biol.* **30,** 617–644.

Henrick, C. (1995). Juvenoids. *In* "Agrochemicals from Natural Products" (C. R. A. Godfrey, ed.), pp. 147–213. Dekker, New York.

Jones, G., and Sharp, P. A. (1997). Ultraspiracle: An invertebrate nuclear receptor for juvenile hormones. *Proc. Natl. Acad. Sci. USA* **94,** 13499–13503.

Nijhout, H. F. (1994). "Insect Hormones." Princeton University Press, Princeton, NJ.

Riddiford, L. M. (1994). Cellular and molecular actions of juvenile hormone. I. General considerations and premetamorphic actions. *Adv. Insect Physiol.* **24,** 213–274.

Truman, J. W., and Riddiford, L. M. (2002) Endocrine insights into the evolution of metamorphosis in insects. *Annu. Rev. Entomol.* **47,** 467–500.

Wigglesworth, V. B. (1983). Historical perspectives. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 7, pp. 1–24. Pergamon Press, New York.

Williams, C. M. (1967). Third-generation pesticides. *Sci. Am.* **217,** 13.

Wyatt, G. R., and Davey, K. G. (1996) Cellular and molecular actions of juvenile hormone. II. Roles of juvenile hormones in adult insects. *Adv. Insect Physiol.* **26,** 1–155.

Katydid

see *Orthoptera*

Killer Bees

see *Neotropical African Bees*

L

Lacewing

see *Neuroptera*

Ladybugs

Michael E. N. Majerus
University of Cambridge

Ladybugs are one of the most familiar groups of insects. These beetles have received attention in both pure and applied areas of biological research. In some senses they are typical insects, having regular life cycles comprising egg, larval, pupal, and adult stages. However, close scrutiny of the behavior and habits of ladybugs has revealed a variety of fascinating evolutionary and ecological features, including color pattern polymorphism, extreme promiscuity, cannibalism, sexually transmitted diseases, and biased sex ratios, some of which seem to be contrary to theoretical expectation. Here the basic biology of ladybugs and some of these conundrums are considered.

As a group, the ladybugs are the most popular of beetles. The bright colors of many species and their reputation of being beneficial, because many species eat plant pests, are at the root of this popularity. In many parts of the world ladybugs are named after religious figures and are revered, often being considered harbingers of good fortune. Indeed, the common English-language name for this family of beetles derives from the Virgin Mary. Ladybirds are "Our Lady's birds."

DESCRIPTION

Ladybugs are beetles of the family Coccinellidae. This family consists of about 5200 known species of small to medium-sized, oval, oblong oval, or hemispherical beetles. The dorsal surface is convex and the ventral surface is flat. The forewings, or elytra, are strong and are often brightly colored, sporting two or more strongly contrasting colors in a bold pattern. Not all species are red with black spots. Almost every color of the rainbow is found as the predominant color of some species of ladybug. These ground colors are usually allied to a second color, which differs starkly from the first, particularly with respect to tone. Thus, ladybugs may be red and black, or yellow and black, or black and white, or dark blue and orange, and so on. Sometimes the spots are replaced with stripes or a checkered pattern. The elytra cover the membranous flight wings, which are folded away when the beetle is not in flight.

THE LADYBUG LIFE CYCLE

Ladybugs through the Year

The life cycle of ladybugs has four stages: egg, larva, pupa, and adult. The length and timing of the different stages varies greatly with geographic region. Mating usually occurs when food is available, and eggs are laid in the vicinity of larval food. In contrast to many other insects, the two feeding stages (larvae and adults) usually have the same diet. In regions with winter and summer seasons, reproduction usually occurs in late spring and early summer. In some climates, reproduction can continue throughout the summer, with several generations being produced. However, in places with hot summers, some ladybugs have a dormant period (or

aestivation) in the hottest months, sometimes having a second period of reproduction in the fall. The winter is generally unfavorable for ladybug reproduction, and ladybugs usually pass the winter as dormant adults. In wet/dry seasonal climates, particularly in the tropics, many ladybugs are dormant through the dry season, beginning to reproduce at the start of the wet season when food becomes more readily available.

The rate of development of ladybugs, like that of other insects, depends largely on ambient temperature. In a species such as *Adalia bipunctata* in a temperate climate, the egg stage lasts from about 4 to 8 days; larvae feed for about 3 weeks. When they stop feeding, they form a humped prepupa and shed the final larval skin about 24 h later to produce a pupa that is attached to the substrate at its posterior. The pupal stage lasts 7 to 10 days. When the adult emerges, the elytra are pale yellow and unpatterned. Hemolymph is pumped into the elytra and flight wings to expand them, and the color patterns develop over the next day or two. Adult ladybugs live for up to a year.

The eggs of most species of ladybugs are bright yellow and are laid upright in batches (Fig. 1) in the vicinity of food. Newly hatched larvae habitually eat any remaining eggs in their clutch and then disperse to find food. For many species, this food is in the form of small sapsucking insects such as aphids or coccids. However, some species feed on fungi, while others are true vegetarians, eating the foliage of plants. The larvae (Fig. 2) are usually elongate, and the ratio of leg length to body length is variable, being correlated to diet. The pupa (Fig. 3) is usually formed on the host plant. Both larvae and pupae may be brightly colored and patterned.

FIGURE 2 Larva of *A. bipunctata.*

Generalist and Specialist Ladybugs

Broadly, different species of ladybugs can be split into generalists and specialists on the basis of their dietary array and the range of habitats that they live in. Most of the commonest species feed on a variety of aphid species and move from one host plant to another as aphid colonies wax and wane. However, some species have a specialized diet and so are confined to specific habitats where their food occurs. This is true of some of the aphid feeders, as well as for many of the species that feed on other diets, such as coccids, mildews, the leaves of plants or the pollen, and nectar of flowers, as their principal food. Many of these species have evolved precise adaptations to their diets and habitats.

LADYBUG COLOR PATTERNS

Warning Colors and Chemical Defense

The bright, eye-catching color patterns of most ladybugs are their first line of defense against many predators. The bold markings of one bright color set on a background of another contrasting color provide a memorable image that warns potential predators that ladybugs have hidden defenses, being foul smelling and evil tasting.

The chemical defenses of ladybugs involve a range of chemicals: alkaloids, histamines, cardiac glucosides, quinolenes, and pyrazines, some of which are synthesized by the beetles while others are sequestered from food.

Anyone who has picked up a ladybug a little roughly will have noticed that they often secrete a yellow fluid. This

FIGURE 1 Egg clutch of *Harmonia axyridis* laid among aphids.

FIGURE 3 Pupa of *H. axyridis.*

FIGURE 4 *Anatis ocellata* reflex bleeding.

FIGURE 5 Melanic and nonmelanic forms of *A. bipunctata.*

behavior, called reflex-bleeding, is part of their defense. The fluid is filtered hemolymph and is exuded through pores in the leg joint (Fig. 4), whence it runs along grooves to form small droplets at the edge of the pronotum and elytra. This "reflex blood" contains a cocktail of volatile chemicals that have a strong and acrid scent to deter naïve predators.

Many species of ladybug share the same basic color combinations, red with black spots or yellow with black spots being the most common. From an evolutionary perspective, the similarities between many species are beneficial to all. The reasoning is simply that the more chemically defended species share the same color pattern, the smaller the number of individuals of each species likely to be harmed by naïve predators as the latter learn to associate a particular color pattern combination with unpalatability. This type of resemblance, involving a complex of species that resemble one another and are all unpalatable, is known as Müllerian mimicry.

Most of the generalist ladybugs have fairly simple patterns of just two strikingly different colors. However, some of the habitat specialists have more complex coloration. For example, some of the reed-bed specialists, such as *Anisosticta novemdecimpunctata,* have the ability to change color during their adult life. Through the fall and winter these ladybugs are beige with black spots and are well camouflaged between the old browning reed leaves where they overwinter. In spring, when the ladybugs move to new green reeds to feed and reproduce, their elytra become flushed with red pigment, thus giving the ladybugs a warning pattern.

Polymorphism

Many ladybug species have variable color patterns, and, in many cases, distinct color forms occur together as genetic polymorphisms. Thus, for example, in many parts of Asia, several different forms of *Harmonia axyridis,* can be found. The different forms are controlled genetically, with the inheritance of most depending on differences in just one gene. The existence of these genetic polymorphisms is surprising because the theory of warning coloration leads to the expectation that all members of the species will look the same. Considerable research time has been expended on this evolutionary conundrum, particularly in *A. bipunctata,* which has some forms that are mainly red with black spots and other forms that are black with red spots (Fig. 5). The factors implicated in the evolution and maintenance of the forms of this species include different levels of activity (black surfaces warm up more rapidly than red ones), sexual selection by female choice (some females have a genetic preference to mate with black males), and different levels of unpalatability to different predators. A fully convincing explanation for these polymorphisms has yet to be found.

REPRODUCTIVE BIOLOGY

Promiscuity

The reproductive biology of ladybugs raises several evolutionary problems. Both male and female ladybugs are highly promiscuous. Theoretically, females that produce large and energetically expensive germ cells should mate only often enough to ensure high fertilization rates. This theoretical limitation reflects the energetic and temporal costs of copulation and the possibility of contracting sexually transmitted diseases. Yet females of *A. bipunctata* mate about 10 times as often as they need to to fertilize all their eggs. Such promiscuity may represent a bet-hedging strategy, or it may afford a means of providing conditions for sperm competition. A hedging strategy addresses the unpredictability of the environments that will face a female's progeny: by mating with a wide variety of different males, she is assured of producing genetically diverse offspring, at least some of which may have genes appropriate to the unknown future habitat. With respect to sperm competition, female ladybugs store the sperm they receive from males in a storage organ called a spermatheca. Thus, after mating with many males, the spermatheca will contain contributions from a number of different males, and these sperm will have to compete for the opportunity to fertilize the eggs.

Female Mate Choice

The study of the reproductive biology of ladybugs has an important place in evolutionary biology, for one of Darwin's mechanisms of evolution was first demonstrated on a ladybug. In addition to natural selection, Darwin argued that some characteristics of some organisms were the result of sexual selection through either male competition or female choice of mates. That females may have a genetically controlled preference to mate with males of a particular genetic type was first demonstrated in *A. bipunctata,* over a hundred years after the theory was first proposed. In brief, it was shown that some females carry a single gene that is expressed as a preference to mate with melanic rather than nonmelanic males, irrespective of their own color. Subsequently, mating preferences were shown to exist in other species of ladybug.

Apparent Waste of Sperm

Male ladybugs also present some interesting problems. In a single copulation, for example, a male *A. bipunctata* can transfer to a female up to three sperm packages, or spermatophores. The spermatheca of a female can store about 18,000 sperm. An average spermatophore contains about 14,000 sperm. Therefore, a male that transfers three spermatophores passes more than twice the number of sperm a female can contain. This apparent waste is difficult to comprehend. Possibly by transferring an excess of sperm, the male is indulging in a coarse type of sperm competition, in which sperm in the female's spermatheca from previous matings are flushed out.

Consequences of Promiscuity: Sexually Transmitted Diseases

Not only do both sexes of many species mate many times, but the duration of each copulation is considerable, lasting several hours in many species. This promiscuity has had one obvious consequence: some species of ladybug are infected by sexually transmitted diseases. Such diseases are generally rare in invertebrates, yet both sexually transmitted mites and fungi infect ladybugs. The mite *Coccipolipus hippodamiae,* which appears to specialize on ladybugs, lives under the elytra, with its mouthparts embedded into the elytra, from which it sucks hemolymph. Mite larvae emerging from eggs produced by the adult females travel to the posterior end of their host before moving onto a new host when the ladybug copulates. A sexually transmitted fungus (in the *Laboulbeniales*) also occurs.

CANNIBALISM

Ladybugs indulge in cannibalism. Both adults and larvae will resort to eating conspecifics and sometimes other species, particularly when other food is scarce. The most vulnerable individuals are those that either are immobile (eggs, ecdyzing larvae, prepupae, pupae) or have a soft exoskeleton (recently ecdyzed larvae, newly formed pupae, newly emerged adults). Aphidophagous species tend to be more prone to cannibalism than those with other diets, largely because of the ephemeral nature of their prey and because they are more prone to large fluctuations in population density.

MALE-KILLING BACTERIA AND LADYBUG SEX RATIOS

The population sex ratio of the majority of sexually reproducing organisms is close to 1:1; selection will normally promote the production of the rarer sex, so that the stable strategy is for sex ratio equality. Female-biased sex ratios were first recorded in the ladybug *A. bipunctata* from Russia in the 1940s. Some females were found to produce only female offspring. The trait was inherited maternally. Subsequent research has shown that male embryos die while in the egg as a result of the action of bacteria such as *Wolbachia.* These male-killing bacteria live in the cytoplasm of cells and are transmitted from infected mothers to their eggs. Although the bacteria in male eggs die when they kill their host, they benefit clonally identical copies of themselves in their host's siblings, which consume the dead male eggs. The additional resources gained by these neonate female larvae increase their fitness and hence that of the bacteria that they carry.

PEST CONTROL

The benefits of allowing ladybugs to eat plant pests have long been recognized. Their importance in controlling aphids on hops in England was noted as early as 1815. For over a hundred years, many attempts have been made to use ladybugs as biological control agents of plant pests such as aphids and coccids. The first reported attempt, and still one of the most successful, was the introduction into California of an Australian ladybug, *Rodalia cardinalis,* to control the cottony cushion scale, *Icerya purchasi,* in 1888/1889. This project, costing just $1500, saw an almost immediate return because the orange crop in California increased threefold in 1890. It was the startling economic success of this project that began the biological control "explosion" that occurred through the first half of the 20th century, until the development of cheap and effective synthetic insecticides.

Not all attempts to use ladybugs in biological control have been as successful as that involving *R. cardinalis,* and in general, the successes reported have involved ladybugs that have been used to control scale insects (Coccidae) and mealybugs (*Pseudococcus* spp.). Ladybugs introduced to control aphids on a large scale have been less efficient, largely because aphid populations increase much more rapidly than do ladybug populations. This means that once aphid populations on a crop have reached sufficient density to attract ladybugs in numbers, the aphid population is already causing damage.

Despite this shortcoming, ladybugs are widely used on a smaller scale to reduce aphid populations. One species, the convergent lady beetle, *Hippodamia convergens,* is of particular note. This species aggregates in vast numbers in high mountain valleys, to pass the winter. Huge numbers are collected annually from these aggregations. The ladybugs are then packaged and stored under precisely controlled, cold conditions until the spring, when they are sold through garden centers or by mail order to ecologically minded gardeners and organic farmers.

See Also the Following Articles

Aposematic Coloration • Biological Control • Mating Behaviors • Mimicry • Predation • Sexual Selection

Further Reading

Gordon, R. D. (1985). The Coccinellidae (Coleoptera) of America north of Mexico. *J. N. Y. Entomol. Soc.* **93,** 1–912.

Hemptinne, J.-L. and Dixon, A. F. G. (1991). Why have ladybirds generally been so ineffective in biological control? *In* "Behaviour and Impact of Aphidophaga" (L. Polgar, R. J. Chambers, A. F. G. Dixon, and I. Hodek, eds.), pp. 149–157. SPB Academic Publishing, The Hague.

Hodek, I., and Honek, A. (1996). "Ecology of Coccinellidae." Kluwer, Dordrecht, Netherlands.

Iablokoff-Khnzorian, S. M. (1982). "Les Coccinelles Coléopteres-Coccinellidae Tribu Coccinellini des regions Paléarctique et Orientale." Société Nouvelle des Editions Boubée, Paris.

Majerus, M. E. N. (1994). "Ladybirds," Vol. 81 in New Naturalist Series. HarperCollins, London.

Majerus, M. E. N., and Hurst, G. D. D. (1997). Ladybirds as a model system for the study of male-killing symbionts. *Entomophaga,* **42,** 13–20.

Majerus, M. E. N., and Kearns, P. W. E. (1989). "Ladybirds," Vol. 10 in "Naturalists' Handbooks." Richmond, Slough, U.K.

Larva

Frederick W. Stehr
Michigan State University

The term "larva" is currently used for all immatures that are not eggs, pupae, or adults. When larva is used in this comprehensive sense, the subcategories include exopterygote larva (for Hemimetabola, which have the wingpads developing externally) and endopterygote larva (for Holometabola, which have the wings developing internally as histoblasts in the larva, becoming external wingpads in the pupal stage).

Larvae occur in a great diversity of sizes, shapes, and colors. Colorful ones almost always live in exposed habitats where their colors and shapes offer cryptic concealment or where their bright colors and spines warn potential predators that they are not to be eaten. Larvae that live in concealed habitats are nearly always combinations of white, gray, black, or brown.

Instar has been conventionally defined as the stage the larva is in between molts. Stadium is defined as the interval of time between molts. Others contend that the instar is properly defined as the stage the larva is in between apolysis (separation of the old cuticle) and the molt to the next stage, and a pharate (next-stage) larva would be present before the next molt.

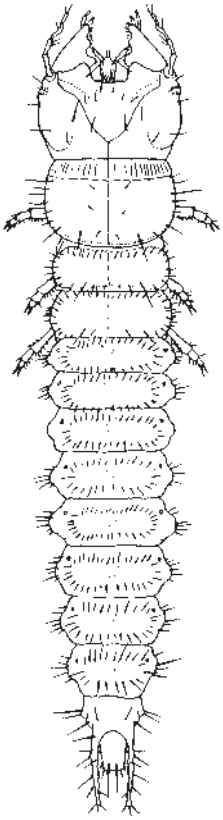

FIGURE 1 A campodeiform ground beetle larva, *Harpalus* (Carabidae). (Reproduced from A. Peterson, 1951, *Larvae of Insects,* Vol. 2, with permission of Jon A. Peterson.)

There are some general terms used for types of holometabolous larvae that have broad usage. Campodeiform larvae (Fig. 1) are somewhat flattened and have an elongate body, thoracic legs that are well developed, a head that is directed forward, no abdominal prolegs, and antennae and cerci that are usually conspicuous. This larval type is common in the Coleoptera (beetles), Megaloptera (dobsonflies and fishflies), Neuroptera (lacewings and antlions), and Raphidioptera (snakeflies).

Elateriform larvae (Fig. 2A) are somewhat similar to campodeiform larvae, but their body is more elongate, subcylindrical, and more heavily sclerotized. This type is common in the Elateridae (click beetles) and other Coleoptera.

Scarabaeiform larvae (Fig. 3) have a C-shaped, whitish body, a dark head, and well-developed thoracic legs. White grubs (Coleoptera: Scarabaeidae) are the best example.

Eruciform larvae are caterpillar-like and have a cylindrical body and well-developed thoracic legs, and prolegs are present. This type is common in the Lepidoptera (butterflies and moths), Mecoptera (scorpionflies), and Hymenoptera (sawflies only).

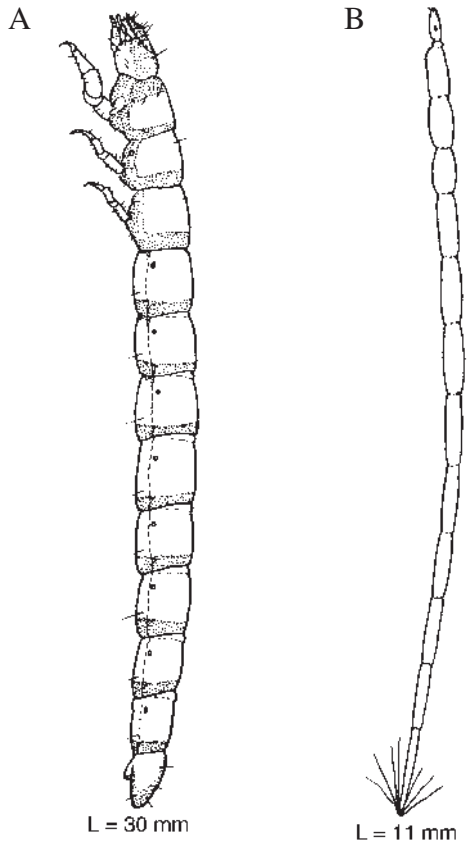

FIGURE 2 (A) A heavily sclerotized elateriform larva (lateral). (Reproduced from A. Peterson, 1951, *Larvae of Insects,* Vol. 2, with permission of Jon A. Peterson.) (B) An elongate, legless, vermiform "wormlike" larva. (Reproduced from A. Peterson, 1951, *Larvae of Insects,* Vol. 2, with permission of Jon A. Peterson.)

FIGURE 4 A legless, peg-shaped maggot of the higher flies. (Reproduced from A. Peterson, 1951, *Larvae of Insects,* Vol. 2, with permission of Jon A. Peterson.)

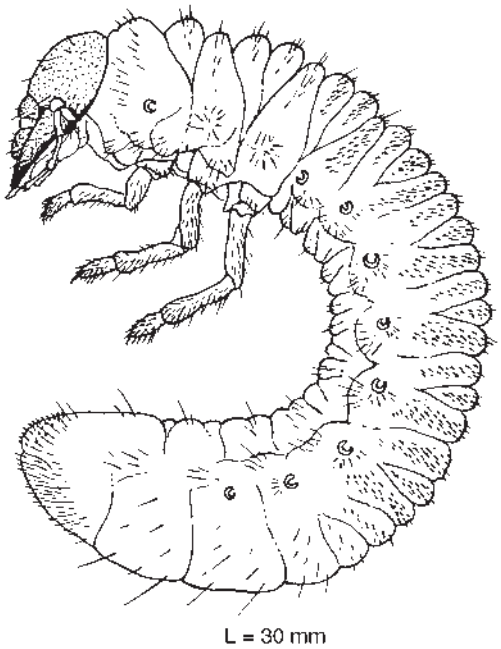

FIGURE 3 A C-shaped scarabaeiform larva, *Phyllophaga* (lateral). (Reproduced from A. Peterson, 1951, *Larvae of Insects,* Vol. 2, with permission of Jon A. Peterson.)

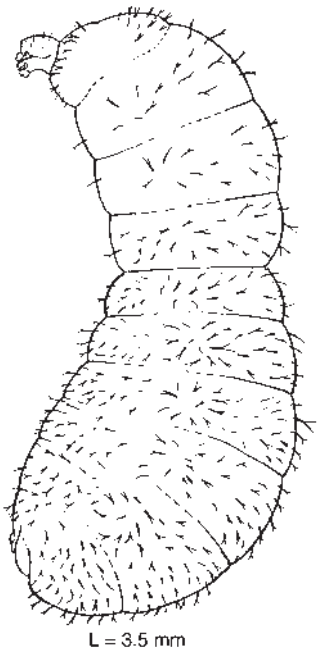

FIGURE 5 A baglike ant larva. It and other similar larvae are commonly called "grubs." (Reproduced from A. Peterson, 1951 *Larvae of Insects,* Vol. 2, with permission of Jon A. Peterson.)

Vermiform larvae (Fig. 2B) are "wormlike." This is an ill-defined term, but it is generally applied to an elongate, legless larva with or without a conspicuous head. Maggots (Fig. 4) are the larvae of higher Diptera. Their shape is peg-like and tapering toward the anterior end. They are legless, have a greatly reduced head (no head capsule), and have conspicuous mouthhook(s). The posterior end bears a pair of conspicuous spiracles.

Grub (Fig. 5) is an imprecise term that is often applied to "comma-shaped" larvae with or without legs or having greatly reduced legs. This term is commonly applied to weevil larvae and other Coleoptera larvae and to many larvae of the higher Hymenoptera (ants, bees, and wasps) that often have reduced or inconspicuous heads and appendages.

See Also the Following Articles

Caterpillars • Cocoon • Eggs • Metamorphosis • Pupa and Puparium

Further Reading

Stehr, F. W. (ed.) (1987) "Immature Insects," Vol. 1. Kendall/Hunt, Dubuque, IA.

Stehr, F. W. (ed.) (1992). "Immature Insects," Vol. 2. Kendall/Hunt, Dubuque, IA.

Learning

Daniel R. Papaj
University of Arizona

A butterfly learns to search for the shape of its preferred host plant's leaves and, contemporaneously, for the color of preferred nectar sources. A parasitoid wasp learns color, pattern, and odor components of its insect host's microhabitat. A grasshopper avoids feeding on a plant associated with a recent digestive malaise. An emerging adult paper wasp imprints on odor cues in its nest, using the odors to distinguish nestmates from nonnestmates. A male damselfy learns to recognize andromorphs (male-mimicking females) as females. A fly improves its depth perception with experience. A bee memorizes a sequence of visual landmarks between its nest and a patch of flowers, as well as the distance between landmarks. All of these are examples of learning, a phenomenon that is ubiquitous throughout the animal kingdom and, as these examples illustrate, well represented within insects. In fact, learning has been documented in all major insect orders. While best studied in the context of foraging for food or oviposition sites, evidence of learning has also been obtained in relation to water consumption, mate finding and choice, territoriality, predator avoidance, dispersal, migration, kin recognition, and thermoregulation.

LEARNING CHARACTERIZED

Characteristics

Learning eludes an easy, satisfying definition, but the following characteristics constitute a useful guide. Learning involves an enduring change in behavior with experience, the change usually progressing gradually with continued experience to some asymptote. Learned behavior is often modified by novel experiences, and effects of experience eventually wane if not reinforced.

Associative vs Nonassociative Learning

Learning can be categorized as nonassociative or associative. Nonassociative learning includes habituation and sensitization. Habituation involves the waning of a response to a stimulus upon repeated presentation of that stimulus. Alternately, repeated presentation of a stimulus sometimes enhances a response to that stimulus and often to related stimuli, a process termed sensitization. Associative learning involves pairing a stimulus with another stimulus, or with a motor pattern, such that the response to the first stimulus is altered as a consequence of the pairing. Associative learning is typically evaluated in two kinds of paradigms: classical (Pavlovian) conditioning and instrumental conditioning.

Classical and Instrumental Conditioning

In classical conditioning, an unconditioned stimulus (US) that elicits an unconditioned response is paired in time and space with a novel stimulus, the conditioned stimulus (CS). As a consequence of the pairing, the CS subsequently elicits a conditioned response. Both appetitive and aversive forms of classical conditioning have been documented in insects. Most of what we know about classical conditioning in insects has involved classical conditioning of the proboscis extension reflex (PER), principally in honey bees.

A case for associative learning is strengthened by evidence of discrimination learning. Discrimination learning (sometimes called differential conditioning) controls for effects of sensitization to a CS by training to two CSs, one which is reinforced with a reward (CS+) and one which is not (CS–). If learning is associative, response to the CS+ only is heightened, relative to controls. Discrimination learning is well documented in bees, hymenopterous parasitoids, moths, butterflies, cockroaches, and fruit flies. A case for associative learning is similarly supported if learning is restricted to forward pairing. In forward pairing, the CS is presented shortly before the US, whereas, in backward pairing, the CS is presented shortly after the US. Insects, like vertebrates, show strong learning in forward-pairing regimes but little or no learning in backward-pairing or random-pairing regimes.

In instrumental conditioning (roughly equivalent to operant conditioning), presentation of a reinforcing stimulus is contingent upon the insect's own motor actions. For example,

an entirely novel motor pattern can be generated through a process of trial and error, as when a bee or butterfly learns how to extract nectar from a flower. A standard operant conditioning paradigm in the laboratory requires a tethered orthopteran to move its leg in response to an electric shock, heat, or access to food. A headless roach learns such a task, demonstrating that conditioning can occur at the level of ganglia.

Learning a given task in nature probably involves a combination of stimulus–stimulus and stimulus–response associations. When an insect pollinator learns nectar-extraction routines for different flowers, for example, it simultaneously learns identifying features of each type of flower, allowing the appropriate motor routine to be expressed on the appropriate flower.

Miscellaneous Types

Various forms of associative learning beyond the basic types have special meaning to students of learning. Food aversion learning, strongly implicated in work on vespid wasps, grasshoppers, mantids, and caterpillars, involves avoidance of food stimuli associated with a digestive malaise. Food aversion learning is noteworthy because an aversion can form even when a long period of time (hours) passes between ingestion of a food and the resulting illness.

Spatial learning is an important component of insect navigation. Commonly traversed routes are learned during homing by ants, bees, and wasps, and traplining is learned by bees and butterflies. Honey bees may additionally possess a topographically organized landscape memory that allows them to navigate along a novel route. Spatial learning is useful in contexts other than movement of the whole organism; for instance, bees learn to discriminate textures with their antennae and use such learning to evaluate the microtexture of flower petals.

One form of learning of significance in vertebrates which has not been documented to date in insects is observational learning, in which a subject imitates the motor actions of a demonstrator. Nevertheless, social interactions do influence what insects learn. Honey bees and bumble bees, for example, evaluate floral scents borne by returning foragers and forage selectively for those scents.

LEARNING PROCESSES

Many associative learning processes that have been described for vertebrates have also been shown in insects. The following list of selected processes is derived mainly from work on honey bees, unless otherwise noted.

Generalization refers to an animal's tendency to respond to stimuli that were not reinforced but that are related to a reinforced stimulus (A+) along some perceptual dimension. Moths and honey bees have been shown to generalize odors according to similarities in functional groups and carbon-chain length.

Blocking occurs when an animal that first learns to respond to a stimulus (A+), and is then reinforced on A and a novel stimulus, B, presented together ([AB]+), subsequently fails to show a heightened response to B alone, relative to controls. Learning of stimulus B has been blocked by coupling with the previously learned stimulus A. Blocking illustrates that temporal pairing between a CS and a US is not sufficient for associative learning to take place; rather, a new CS must convey new information in order to be learned. Whereas blocking is a robust phenomenon in vertebrates, studies of blocking in bees and fruit flies have yielded mixed results. Where blocking has been demonstrated, it seems to be restricted to intramodal stimuli (e.g., odor blocking in honey bees).

Overshadowing occurs when an animal reinforced on a compound of stimuli A and B ([AB]+) shows little response to B alone, relative to when reinforced on B alone (B+). As with blocking, overshadowing illustrates that temporal pairing between a CS and a US is not sufficient for associative learning to take place.

Sensory preconditioning occurs when an insect presented simultaneously with two stimuli in the absence of reinforcement ([AB]–), then reinforced on one stimulus (A+), subsequently shows a heightened response not only to A but also to B. During exposure to [AB], the insect learns that A and B belong together. Observed in *Drosophila* fruit flies and honey bees, sensory preconditioning illustrates that a stimulus does not have to be paired directly with a US in order for an association between the stimulus and the US to form.

Second-order conditioning refers to the capacity for a stimulus, once conditioned, to serve as a US in the conditioning of another stimulus. Second-order conditioning may play a major role in learning complex mixtures of stimuli, such as odor blends.

Patterning is evaluated by reinforcing two stimuli in turn (A+, then B+) and then explicitly not reinforcing a compound of those stimuli ([AB]–). Under this protocol, PER odor conditioning in bees shows "negative patterning," responses being greater to individual odors than to the compound. This result can be explained only if the insect treats the compound [AB] as a unit and relates it to the absence of reinforcement. Such learning is referred to as configural learning.

Rule extraction has been demonstrated with the use of delayed matching-to-sample tasks in which honey bees are required to respond to a stimulus that matches a sample stimulus recently experienced. Bees not only solve the task but also transfer the matching to stimuli not previously reinforced. For example, bees trained to match a color can subsequently match patterns of lines and, remarkably, bees trained to match an odor can subsequently match colors too. Such results have been interpreted to mean that insects can form a concept of "sameness." When trained in a delayed non-matching-to-sample task in which they must choose the stimulus that does not match the sample, bees again perform well and make

similar transfers, showing a grasp of a "difference" relation. Bees also learn to extract bilateral symmetry from a series of rewarded patterns and subsequently transfer that extraction to evaluation of novel patterns.

MEMORY

Associative memory in insects, as in vertebrates and other animals, is time-dependent and phasic. Recent work on fruit flies and honey bees suggests as many as five memory phases: (1) an early and (2) a late form of short-term memory [eSTM and lSTM], (3) a midterm memory [MTM], and two forms of long-term memory (in honey bees, characterized as (4) an early form [eLTM] and (5) a late form [lLTM]; in *Drosophila,* characterized as (4) an anesthetic-resistant form and (5) a parallel, susceptible form). STM forms immediately upon association, is short-lived (seconds to minutes), and is relatively easily erased by conflicting information or treatment by cooling or shock. eSTM is characterized by a relatively nonspecific appetitive arousal and is highly suseptible to interference by new, conflicting information or by cooling. lSTM is more stable, is more specific, and takes longer to form than eSTM. The transition from STM to MTM after a single learning trial requires several minutes. MTM is more resistant to interference than STM, requiring hours to decay.

LTM takes longer still to form than either STM or MTM, involves longer lasting changes (hours to weeks), and is relatively resistant to interference. In bees, formation of LTM requires multiple learning trials. LTM is highly context-specific; landmarks learned by bees around their feeder, for example, may be entirely ignored when presented at a novel location. eLTM and lLTM have been distinguished in terms of the effects of inhibiting protein synthesis: synthesis inhibition after 24 h degrades memory, whereas inhibition after 3 days does not. Effects of inhibition depend on the time between learning trials, with closely spaced trials (termed "massed trials") resulting in memory that is independent of protein synthesis.

The underlying processes involved in memory formation are beginning to be revealed. In honey bees, a "value" neuron, the VUM_{mx1} of the subesophageal ganglion, which fires in response to sucrose stimulation, is proposed to be part of the US pathway. In PER odor conditioning studies, artificial depolarization of the VUM_{mx1} neuron just following presentation of an odor generates a conditioned response to the odor. The VUM_{mx1} neuron, which uses octopamine as a neurotransmitter, converges on two brain neuropils, the antennal lobe and the mushroom bodies. Consistent with these observations, olfactory memories can be established by odor-coupled injection of octopamine into either the antennal lobe or the mushroom bodies. The pattern of octopamine effects suggests that antennal lobe processes may relate more to eSTM, whereas mushroom body processes may relate more to lSTM and LTM.

To what degree these findings pertain only to honey bees or only to odor learning is uncertain. Analysis of *Drosophila* mutants suggests that the mushroom bodies are important for odor learning but dispensable for visual or tactile learning. Studies of locusts have indicated effects of feeding experience on diet choice that resemble discrimination learning, but are based on an entirely novel mechanism. This taste-feedback mechanism involves adjustments in the level of sensitivity to nutrients in the hemolymph.

FUNCTION OF LEARNING

In a sense, the function of associative learning is obvious. Animals learn by association to orient toward stimuli predicting positively rewarding resources (such as sugar, pollen, food plant, hosts) and away from stimuli predicting negatively rewarding events (shock, heat, toxins, predators). Likewise, habituation is a means for reducing energy-wasteful, time-consuming responses to meaningless stimuli. In either case, however, learning is needed only if the appropriate responses cannot be predicted without benefit of experience, else an insect could respond (or not respond) innately. Even in an unpredictable environment, whether learning yields higher fitness than innate behavior depends on the relative costs of learning. A robust assessment of costs and benefits of learning has proved elusive, perhaps in part because individual fitness in nature is especially difficult to measure in *Drosophila* and honey bees, the systems in which learning processes and mechanisms have been best studied.

ADAPTATION, CONSTRAINT, AND LEARNING

Limits to Learning and Memory

Of interest to behavioral ecologists is the degree to which learned behavior reflects adaptation by natural selection versus constraints on selection. Generalization (see above), for example, may seem at first to reflect a constraint on learning, but conceivably represents an adaptive mechanism of imprecision. A pollinator, for example, that responded only to the precise odor blend emitted by the first rewarding flower encountered might never visit another flower, owing to among-flower variation in the blend.

A classic case study of limits on learning and memory in nature that interested Darwin himself concerns the tendency for bees, butterflies, and other pollinators to show greater fidelity to one or a few floral species than expected based on the profitabilities of those species. According to one point of view, this so-called floral constancy is dictated by limits on the acquisition, retention, and/or retrieval of stored information about the floral resource.

That foraging success in insects is limited in terms of acquisition and retention seems unlikely at the level of LTM. LTM in insects, as mentioned above, is extraordinarily durable and the amount of information that can be maintained in LTM, as currently understood, is extremely impressive. Butterflies can learn visual cues in two foraging

modes (nectar collection and oviposition) simultaneously, showing meaningful responses in each instance in just a single trial. Bees can be trained to distinguish multiple rewarded stimuli from multiple unrewarded ones and to link features of eight or more different flower species to the time of day at which nectar is available. These features include flower color, odor, pattern, and microtexture. In addition, a bee learns the location, profitability, and visual landscape associated with a rewarding patch of flowers, as well as the route between hive and patch and, in conjunction with the sun compass used to navigate, even the pattern of movement of the sun through the sky.

Retention at the level of LTM is similarly impressive. Bees have been shown to retain LTM without reinforcement for several weeks, a period of time comparable to average worker life expectancy. In *Tribolium* beetles and *Drosophila,* there is evidence that memory formed in the larval stage persists through metamorphosis.

If pollinators are limited at all in memory, it may be at the level of STM. As noted above, STM is particularly vulnerable to conflicting information; this fact may make it difficult for a bee once fixed on a flower type to switch to a novel one. Alternatively, the key to floral constancy may lie in the retrieval of stored information, specifically a constraint on the minimum time required to activate information stored in LTM and a limited capacity to activate multiple memories at once (together, limits on what for vertebrates has been referred to as working memory).

Learning and Memory as Products of Adaptation

An alternative, albeit not mutually exclusive, view holds that natural selection generates an adaptive balance between activation and suppression of memory, tuning that balance finely to the specific ecological requirements of a given species. For example, floral constancy might conceivably permit workers in a colony to partition floral resources efficiently, in which case the properties of learning and memory that contribute to constancy would be viewed as adaptive. It has even been proposed that memory dynamics in bees are tightly matched to foraging activity rhythms as well as the spatial patterning of the floral resource.

Abundant propositions as to adaptive specialization in learning have been made, especially from a comparative standpoint: "Insects of a given species should be prepared to learn particularly well those stimuli relevant to that species' needs." "Social insects should learn better than solitary ones (owing to the demands of a complex and unpredictable social environment)." "Generalist insects should learn better than specialists." For none of these propositions is there compelling evidence, nor will there be until better descriptions are made of learning in an ecological context, learning protocols are brought closer in rigor to those employed in comparative psychology, and more insect species are evaluated.

For now, the primary comparison to be made is a comparison between learning in insects and in vertebrates. Here, the pattern is one of shared features. Despite significant phylogenetic distance between insects and vertebrates, and despite substantial differences in their underlying physiology, there is a remarkable congruence in the diversity and form of learning processes in these taxa (see above). The similarities may reflect shared ancestry, evolutionary convergence, or both. A finding of evolutionary convergence would imply that certain universal, yet to be clearly defined functional principles govern the evolution of learning and memory processes.

See Also the Following Articles

Brain • Dance Language • Neuropeptides

Further Reading

Abramson, C. I., Yuan, A. I., and Goff, T. (1990). "Invertebrate Learning: A Source Book." Am. Psychol. Assoc., Washington, DC.

Bitterman, M. E. (1996). Comparative analysis of learning in honeybees. *Anim. Learn. Behav.* **24,** 123–141.

Chittka, L., Thomson, J. D., and Waser, N. M. (1999). Flower constancy, insect psychology, and plant evolution. *Naturwissenschaften* **86,** 361–367.

Dukas, R. (ed.) (1998). "Cognitive Ecology." University of Chicago Press, Chicago.

Menzel, R. (2001). Searching for the memory trace in a mini-brain, the honeybee. *Learn. Memory* **8,** 53–62.

Papaj, D. R., and Lewis, A. C. (eds.) (1993). "Insect Learning: An Ecological and Evolutionary Perspective." Chapman & Hall, New York.

Shettleworth, S. (1998). "Cognition, Evolution and Behavior." Oxford University Press, Oxford.

Smith, B. H. (1996). The role of attention in learning about odorants. *Biol. Bull.* **191,** 76–83.

Tully, T. (1997). Regulation of gene expression and its role in long-term memory and synaptic plasticity. *Proc. Natl. Acad. Sci.* USA **94,** 4239–4241.

Legs

Peter H. Adler

Clemson University

One of the most generally known and oft-repeated facts about insects is that they possess three pairs of legs, one pair each on the prothorax, mesothorax, and metathorax. Indeed, this condition is in the fundamental ground plan of insects and is amply represented in the fossil record. The condition inspired Latreille's taxon Hexapoda (Greek *hexa*, six, and *poda,* foot). Exceptions to the hexapodous condition are found in the apodous, or legless, insects that have secondarily lost their legs, typically as a result of selection for an obligatory parasitic or sedentary existence.

The six-legged condition is derived from an ancestral arrangement in which legs occurred on the majority of body

segments. Over evolutionary time, the serially uniform legs became modified in the insectan lineage into the characteristic mouthparts, thoracic legs, and various abdominal appendages, such as cerci and genitalia, while typically becoming lost on other abdominal segments. Further evolution of the basic six-legged condition in the insectan lineage has resulted in an enormous diversity of structure and function. This structural and functional diversity of legs, along with the acquisition of wings without the loss of legs, which is a condition unique to insects, undoubtedly has been a key factor in the numerical success of insects and their representation in nearly every habitat on the planet. The exquisite diversity in leg structure plays an important role in the taxonomy and classification of insects.

STRUCTURE

In the classic textbook interpretation, the insectan leg has six well-sclerotized segments, arranged proximal to distal: the coxa, trochanter, femur, tibia, tarsus, and pretarsus. A more fundamental and complete segmentation scheme, which facilitates the recognition of leg and leg-derived homologies among all arthropods, involves 10 or 11 segments. These segments include the epicoxa (debatably present as the wing articulation and a fused portion of the tergum), subcoxa (absorbed into the pleuron), coxa, trochanter, prefemur (typically fused with the trochanter), femur, patella (fused with the tibia), tibia, basitarsus, eutarsus (often subdivided), and pretarsus. A more modern interpretation of the free leg of extant insects, therefore, depicts it as consisting of seven to eight distinct segments, which are the classical six plus a basitarsus and a prefemur, in some insects.

Each segment in the insectan leg, unless secondarily lost or fused, is independently movable by muscles inserted on its base. Thus, subdivisions of the eutarsus, marked by flexible cuticle but without corresponding internal muscles, are not true segments; these subdivisions are referred to as tarsomeres. The areas of flexion between segments are joints, and the well-sclerotized contact points in the joints are the condyles. The various joints contribute to the mechanical efficiency of the leg. The articulation between the coxa and the body, for example, allows the leg to move forward and rearward, whereas that between the coxa and the trochanter allows the leg to be lifted at the end of the backstroke and depressed at the beginning of the backstroke.

Leg joints are of two types. Monocondylic joints have a single point of articulation, somewhat like a ball-and-socket joint, and usually are situated dorsally. They allow considerable freedom of movement and are characteristic of the legs of larval insects. Dicondylic joints consist of an anterior and a posterior condyle, or a dorsal and ventral condyle in the case of the trochanterofemoral joint. They typically limit movement to that of a hinge. Adult legs usually have dicondylic joints, although the tibiotarsal joint is often monocondylic.

The coxa (plural coxae) is typically short and rather stout, although it varies in shape among taxa. It is set in a coxal cavity and articulates with the thorax at the coxal process of the pleural sulcus (groove). Quite often, it also articulates with the thoracic trochantin and sternum, somewhat restricting its movement. To withstand the forces of movement, the coxa is strengthened by a ringlike basicostal sulcus that sets off a basal sclerite, the basicoxite. Internally, the basicostal sulcus is expressed as a ridge, the basicosta, that provides for muscle attachment. Posterior to the point of articulation, the basicoxite is called the meron and in insects such as adult Neuroptera and Lepidoptera, it can be quite large. In higher Diptera, the meron is detached from the coxa and forms a plate in the mesothoracic pleuron. In some insects, an additional external groove, the coxal sulcus, divides the coxa lengthwise.

The trochanter is small and freely movable in a vertical direction on the coxa, but it is often rather fixed to the base of the femur. In the larvae and adults of numerous fossil insects and a few extant taxa, such as Odonata, two trochanteral segments are present, the distal one being the prefemur.

The femur (plural femora) is usually the largest and strongest segment of the leg. Its size is related to the mass of the tibial extensor muscles within it, varying from a small, thick segment in larval insects to the enormous segment in the hind leg of jumping Orthoptera. The femur often is equipped with spines and other cuticular modifications, especially in predatory insects.

The tibia (plural tibiae) typically is long and slender in adult insects. Proximally, it is bent slightly toward the femur, allowing the shaft of the tibia to be flexed close against the femur for more locomotory power in insects such as grasshoppers. It often bears spines for grooming or for engaging the substrate to aid in locomotion. Many insects also have apical or subapical movable spurs on their tibiae.

The tarsus (plural tarsi) is a simple, undivided segment in holometabolous larvae and basal hexapods such as Protura and some Collembola. In collembolans, the tarsus and tibia are fused into a single tibiotarsus. In most insects, a separate segment, the basitarsus, is present and the eutarsus is subdivided into two to four sections or tarsomeres. The ventral surface of the basitarsus and eutarsus often bears pads called tarsal pulvilli that aid movement on smooth surfaces and are especially well developed in some Orthoptera. The basitarsus and eutarsus generally are well endowed with sensory hairs and chemoreceptors. The ventral surface often has a secretory epithelium that produces a wax, possibly for waterproofing, inhibiting the uptake of undesirable water-soluble compounds, or preventing entrapment in surface films.

The pretarsus, also called the acropod or posttarsus, arises from the distal end of the eutarsus. In the Protura, Collembola, and larvae of many holometabolous insects, the pretarsus is a simple, clawlike segment. Typically, however, the pretarsus consists of a membranous base, a pair of hollow claws (ungues), and various sclerites and lobes. A sclerotized unguitractor plate articulates with the eutarsus into which the plate is partly

invaginated. The muscles that flex the claws are inserted on a process of this plate. The claws articulate with the unguifer, a median process at the distal end of the eutarsus. A saclike, hollow lobe, the arolium, arises between the claws. In adult Diptera, other than crane flies, an arolium is absent. Instead, a padlike lobe called the pulvillus (plural pulvilli) arises from an auxiliary plate (auxilia) beneath the base of each claw, while an unpaired, lobelike or bristlelike process, the empodium, stems medially from the unguitractor plate. The pretarsal pads and lobes are covered with adhesive setae (tenent hairs) that allow the insect to climb and hold onto smooth surfaces.

Variation in structure, and hence function, can be found among the three pairs of legs within an individual, as well as between larvae and adults, between males and females, and among taxa. The thoracic legs of many larval insects are serially uniform or, sometimes, lacking. The legs of adults often vary in structure among the three pairs, although even adults of some insects (e.g., some female Coccidae, Psychidae, and Strepsiptera) are devoid of legs. The variation among the three pairs of legs in adults often is associated with acquisition of food, courtship, and mating. Developmental variation occurs in holometabolous insects, which have simple, rather generalized legs in the larvae and more specialized legs in the adults. Among the hemimetabolous groups, many Hemiptera–Heteroptera gain a tarsomere in the final molt.

Sexual dimorphism in leg structure is particularly prevalent. The reduced forelegs of nymphalid butterflies have short tarsomeres in females but lack all segments beyond the tibia in males. The forelegs of male Ephemeroptera are typically elongated to grasp the female. The slender, elongate legs of crane flies are even longer in males than in females for species in which a guarding male stands over the female during oviposition; in some crane flies the distalmost tarsomere is prehensile in males and used for holding females. The hind femora of males of many Coreidae are enlarged for intrasexual fighting. The pulvilli beneath the claws of some flies, such as Tachinidae, are considerably larger in males than in females.

FUNCTION

The majority of insectan legs are either elongate, slender, and designed for walking and climbing or cursorial, i.e., adapted for running, as in the cockroach (Fig. 1A). During walking, the legs form alternating triangles of support, with the fore and hind legs of one side and the middle leg of the opposite side contacting the substrate as the other three legs move forward. Various modifications allow the legs to be used in other forms of locomotion. Enlarged hind legs of many Orthoptera, fleas, and other insects are saltatorial, meaning they are designed for jumping. The jump of insects such as fleas is aided by a rubber-like protein called resilin in the cuticle that stores and subsequently releases energy for the jump. Powerful, spadelike forelegs of mole crickets, scarab beetles, burrowing mayflies, and other insects are fossorial, or adapted for digging and rapid burrowing (Fig. 1B). Flattened, fringed legs of aquatic insects such as dytiscid and gyrinid beetles and notonectid backswimmers serve as oars for paddling or swimming (natatorial legs), while long legs with hydrophobic tarsal hairs and anteapical claws, as seen in water striders (Gerridae), are for skating on the surface of water. The legs of some insects, although well developed, have lost their associated locomotory function. The spiny legs of Odonata, for example, are designed for perching or seizing and holding prey captured while the dragonfly is in flight; the legs are ineffectual for walking.

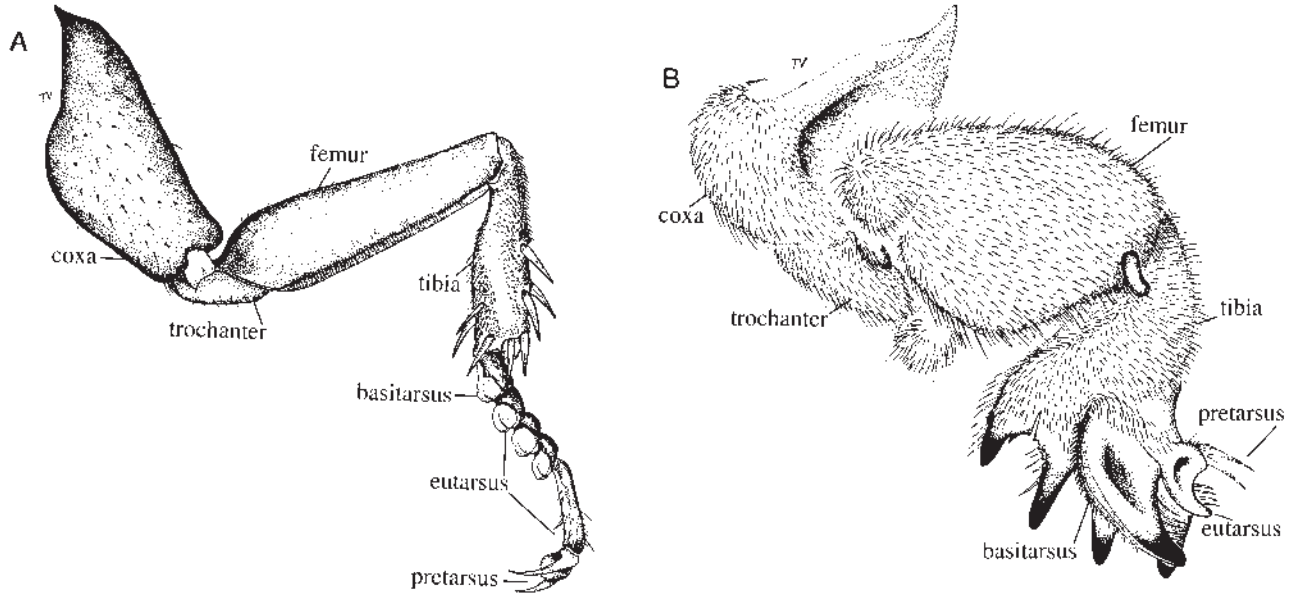

FIGURE 1 (A) Cursorial foreleg of the Madeira cockroach *(Leucophaea maderae).* Illustration by T. S. Vshivkova. (B) Fossorial foreleg of the northern mole cricket *(Neocurtilla hexadactyla).* Illustration by T. S. Vshivkova.

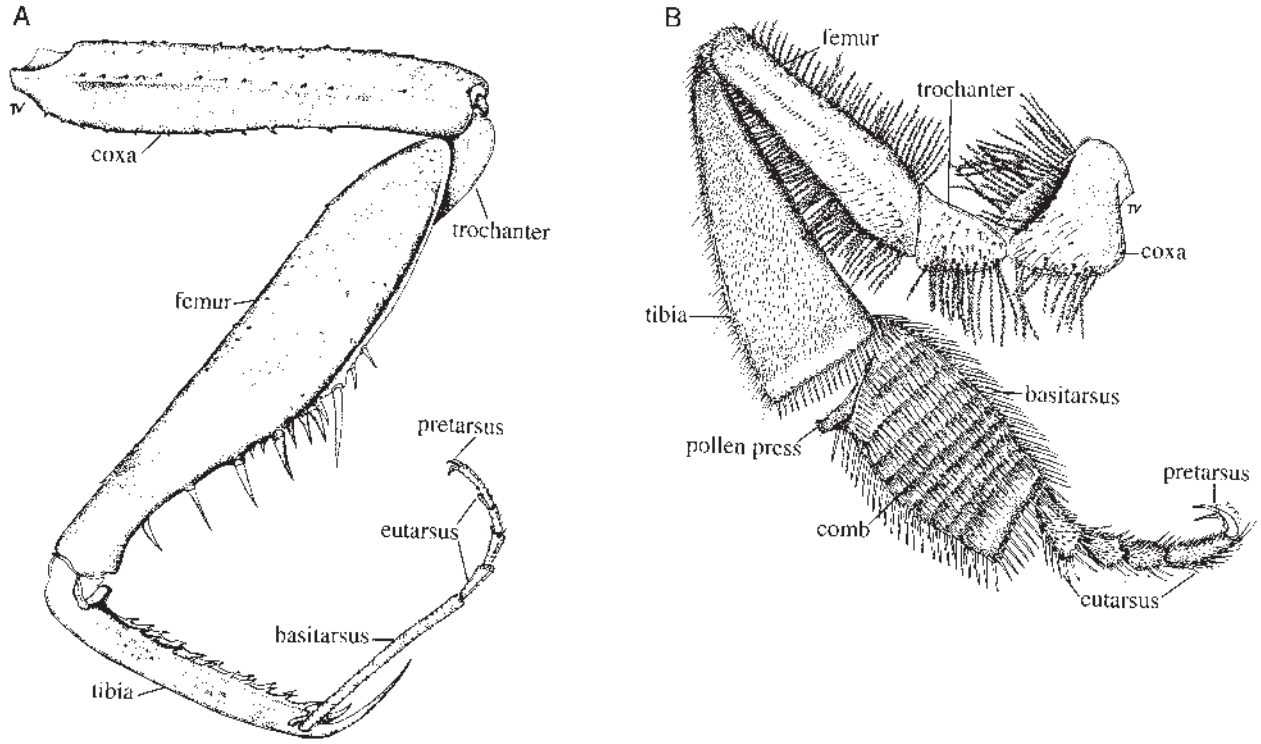

FIGURE 2 (A) Raptorial foreleg of the Carolina mantid *(Stagmomantis carolina).* Illustration by T. S. Vshivkova. (B) Inner surface of the hind leg of the honey bee *(Apis mellifera).* Illustration by T. S. Vshivkova.

Although typically regarded as agents of locomotion, the legs have assumed a wide range of additional, or altogether different, functions. Often, function can be inferred from structure; for example, the thickened, spinose legs of many insects signify a predacious mode of life. Raptorial legs (Fig. 2A), i.e., those designed to seize prey, have arisen independently in many insectan lineages. Either of the three sets of thoracic legs can be raptorial, but the trait is probably most often expressed in the forelegs (e.g., in Mantodea and Reduviidae) and less frequently in the middle legs (e.g., in some Empididae) and hind legs (e.g., in some Mecoptera).

Legs also play an important defensive role, not only in permitting escape by running, jumping, burrowing, and swimming, but also in ways such as kicking and slashing. The spines on the legs of many insects, when used in defense, effectively deter predators and competitors and can inflict considerable damage. Insects such as stink bugs and treehoppers deliver powerful kicks at parasitoids and predators that attempt to attack their young. Autotomy, or the loss of legs at predetermined points of weakness, often at the level of the trochanter, occurs in insects such as crane flies, leaving a predator with only a leg in its clutches as the insect escapes. Legs, whether lost through autotomy or accident, often can be regenerated to various degrees if one or more molts follow the amputation.

All legs are equipped with an extensive arrangement of sensory structures that allow the insect to feel, hear, and taste, providing the insect with its initial assessment of the environment. Chemoreceptors, which are especially prevalent on the basitarsus and eutarsus, provide sensory input on environmental substances and can be used to determine the acceptability of food, ovipositional substrates, and perhaps mates. Mechanoreceptors, most often in the form of hair organs, but also campaniform, chordotonal, and plate organs, provide sensory information on position, movement, and vibrations borne by air and substrate.

In many insects, the legs are used in sound production. Familiar examples include the shorthorned grasshoppers, which have a stridulatory mechanism on the hind femur, involving a series of pegs—the scraper—that is rubbed across a ridged wing vein. Some larval hydropsychid caddisflies have a scraper on the prothoracic femur that is rubbed against a file on the venter of the head. Legs also can be used to produce sound for intraspecific communication by drumming them against a substrate, as in some Orthoptera.

To maintain hygiene, insects spend considerable time preening and grooming their body and appendages. Grooming typically is effected by various leg structures, which can be in the form of cuticular combs (ctenidia), setal brushes, grooves, and notches. The cleaning setae on the foretibia of certain heteropterans are mirror images of the arrangement of antennal setae. The hind leg of honey bees is specially modified to groom pollen from the plumose hairs of the body (Fig. 2B). Combs on the inner surface of the hind basitarsus remove the pollen from the body hairs and pass it to the pollen press between the tibia and the basitarsus. Closure of the press forces the pollen into the pollen basket (corbiculum) on the outer

surface of the tibia where the pollen bolus is held in place by rows of hairs. Once in the hive, the honey bee removes the pollen, with the aid of an apical spur on the middle tibia.

Legs often are used to hold onto objects, and they bear the relevant modifications, including enlarged segments to house increased musculature, various spines and setae, and adhesive organs. The grasping function is seen, for example, in the pincerlike, spiny raptorial forelegs of many predacious insects. It also is expressed dramatically in certain sucking lice in which the claw folds against a thumblike, spinose process of the enlarged tibia. Flies that feed on the blood of birds typically have a thumblike lobe at the base of each of their talonlike claws that helps them grasp feather barbules. Grasping is common during mating, and especially the males of many insects have legs designed to secure and hold their mates. Adhesion to objects such as mates and prey can be achieved with suction discs on the legs. Male dytiscid beetles have a flattened, disklike arrangement on each foreleg that is formed of the basitarsus and the succeeding two tarsomeres; all three structures bear minute suction cups ventrally that can be applied to the elytra of the female.

The colors and patterns of legs vary from subtle to stark, although their function is often poorly understood. Long-legged insects such as phantom crane flies (Ptychopteridae) and some mosquitoes often have banded legs that might render the insect less conspicuous through disruptive coloration. Other configurations of pattern and color play a role in camouflage, mimicry, and courtship. In flies such as some Syrphidae and Micropezidae, the forelegs resemble the antennae of aculeate Hymenoptera, reinforcing the remarkable overall resemblance of fly to wasp.

Other functions ascribed to the legs are often highly specialized. In the Embiidina, the basitarus of each foreleg houses multiple silk glands, and each gland is connected to a seta with an apical pore through which the silk is extruded. The inflated basitarsus of each leg in phantom crane flies contains a tracheal sac, perhaps aiding buoyancy during the driftlike flight. Some flies have specialized areas on their legs, particularly on the tibia, that possibly produce pheromones. Insects such as Chironomidae seem to use the legs much as a second set of antennae. In Protura, which lack antennae, the forelegs probably have assumed an antennal (i.e., sensory) function. Various ornamentations on insectan legs can serve a courtship or intrasexual combative role, as in some coreid bugs.

See Also the Following Articles

Anatomy • Segmentation • Swimming • Walking and Jumping

Further Reading

Adler, P. H., and Adler, C. R. L. (1991). Mating behavior and the evolutionary significance of mate guarding in three species of crane flies (Diptera: Tipulidae). *J. Insect Behav.* **4,** 619–632.

Chapman, R. F. (1998). "The Insects: Structure and Function," 4th ed. Cambridge University Press, Cambridge, UK.

Hlavac, T. F. (1975). Grooming systems of insects: Structure, mechanics. *Ann. Entomol. Soc. Am.* **68,** 823–826.

Kukalova-Peck, J. (1992). The "Uniramia" do not exist: The ground plan of the Pterygota as revealed by Permian Diaphanopterodea from Russia (Insecta: Paleodictyopteroidea). *Can. J. Zool.* **70,** 236–255.

McAlpine, J. F. (1981). Morphology and terminology—Adults. *In* "Manual of Nearctic Diptera" (J. F. McAlpine, B. V. Peterson, G. E. Shewell, H. J. Teskey, J. R. Vockeroth, and D. M. Wood, eds.), Vol. 1, Monogr. 27, pp. 9–63. Research Branch, Agriculture Canada, Ottawa.

Mitchell, P. L. (1980). Combat and territorial defense of *Acanthocephala femorata* (Hemiptera: Coredíae). *Ann. Entomol. Soc. Am.* **73,** 404–408.

Snodgrass, R. E. (1935). "Principles of Insect Morphology." McGraw–Hill, New York.

Lepidoptera
(Moths, Butterflies)

Jerry A. Powell
University of California, Berkeley

Moths and butterflies make up the order Lepidoptera, and they are among the most familiar and easily recognized insects. The Lepidoptera is defined as a monophyletic lineage by a suite of more than 20 derived features, the most obvious of which are the scales and proboscis. The scales are modified, flattened hairs that cover the body and wings, shingle-like, and are the source of the extraordinary variety of color patterns typical of these insects. In all but the most primitive forms, feeding by adults is accomplished by pumping in liquid via a tubular proboscis (haustellum), which usually is elongate and coiled under the head. The sister group of Lepidoptera, the Trichoptera (caddisflies), lack this development of mouthparts and the covering of scales and possess caudal cerci on the abdomen, which are not present in Lepidoptera.

Like other holometabolous insects, lepidopterans pass through egg, larval, pupal, and adult stages. Mating and egg deposition are carried out by the adult moths and butterflies. Within the eggs, embryos develop to fully formed larvae. The larvae, commonly called caterpillars, feed and grow, which is accomplished by a series of stages (instars). At maturity they transform to pupae, usually within silken cocoons spun by the larvae, although many species pupate without a cocoon. Metamorphosis to the adult occurs during the pupal stage, and the fully developed adult breaks the pupal shell to emerge. Adults of most species feed, but they do not grow. Diapause, an arrested state of development, may occur in any of these stages, prolonging life and enabling the insect to bypass seasons that are unsuitable for growth and reproduction.

The Lepidoptera is one of the two or three largest orders of insects, with an estimated 160,000 named species. Based on specimens in collections and extrapolating from recent studies of Central American moths, we believe that fewer than one-half of the known species have been named by

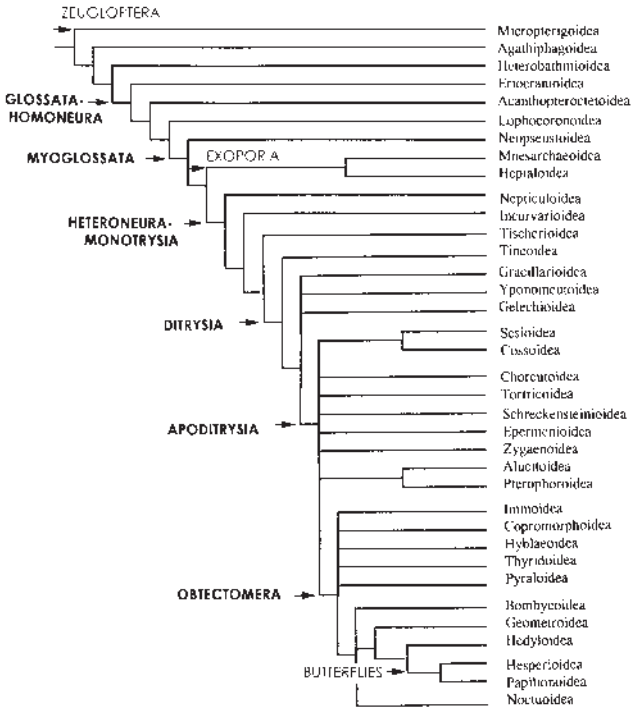

FIGURE 1 Hypothesis of phylogenetic relationships of extant lepidopteran superfamilies. Successively more derived clades representing major morphological changes are indicated in boldface to the left (modified from Kristensen and Skalski, 1999).

taxonomists; even in North America, an estimated one-third of the fauna is undescribed. Thus, a realistic projection of the total world Lepidoptera species number is not possible, but certainly it exceeds 350,000 and may be much larger. Much of this diversity can be attributed to the radiation of species in association with flowering plants. Lepidoptera represent the single most diverse lineage of organisms to have evolved primarily dependent upon angiosperm plants, and their numbers exceed those of the other major plant-feeding insects, Heteroptera, Homoptera, and Coleoptera (Chrysomeloidea and Curculionoidea). Figure 1 depicts the hypothesized evolutionary lineages and lists currently recognized superfamilies of Lepidoptera.

MORPHOLOGY

Adult

The body framework (Fig 2) consists of a hardened (sclerotized) exoskeleton made up of a head capsule with appendages; three fused thoracic segments, each with legs, and two pairs of wings, on the middle (mesothoracic) and third (metathoracic) segments; and an abdomen, which has 10 segments, is less sclerotized than the thorax, and is movable by intersegmental membranes. Complex genital structures of external origin arise from abdominal segments A8–10, and often there are accessory structures (pouches, glands, hair brushes) associated with sound reception, courtship, or other functions.

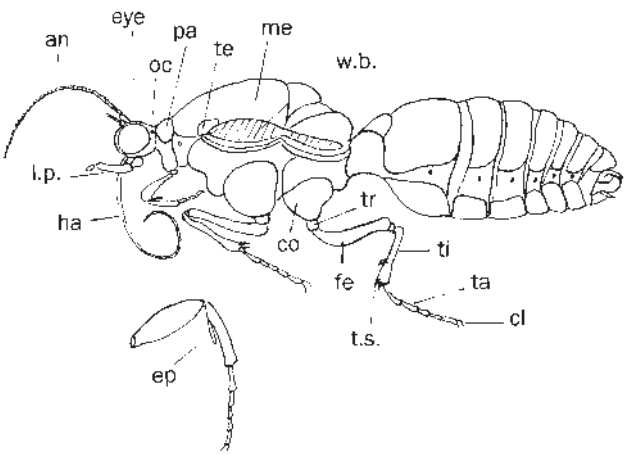

FIGURE 2 Schematic representation of the exoskeletal anatomy of a ditrysian moth, with prothoracic leg enlarged below. Head: an, antenna; eye, compound eye; oc, ocellus; l.p., labial palpus; ha, haustellum (proboscis); Thorax: pa, patagium; te, tegula; me, mesoscutum; w.b., wing base; co, coxa; tr, trochanter; fe, femur; ti, tibia; t.s., tibial spurs; ta, tarsomeres; cl, tarsal claws; ep, epiphysis. Abdomen: tergites and sternites 1–7 and spiracles shown.

HEAD Structures include paired simple eyes (ocelli) and scaleless, raised spots (chaetosema), which are unique to Lepidoptera, although one or both are lost in many taxa (Figs. 2–4). There is enormous variation in the form of the antennae, often between the sexes of a species, being filiform or with the flagellar segments variously enlarged or branched. Antennae of butterflies are enlarged distally, forming apical clubs, while those of moths are not, although some moths have distally enlarged antennae that are tapered or hooked to the tip. The mouthparts of the most primitive moth families retain functional mandibles as in their mecopteroid ancestors, but in the majority of moths the mandibles are lost, and the maxillary galeae are elongate and joined to form a tubular

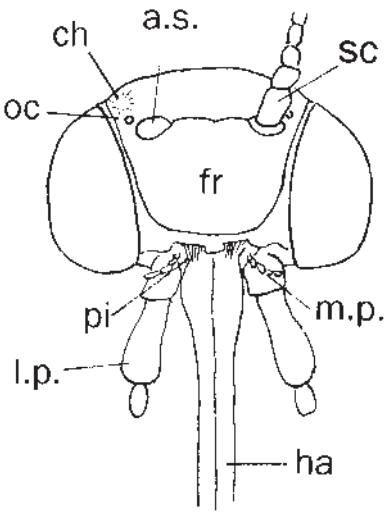

FIGURE 3 Descaled lepidopteran head, frontal aspect. ch, chaetosema; oc, ocellus; a.s., antennal socket; sc, scape; fr, frons; pi, pilifer; m.p., maxillary palpus; l.p., labial palpus; ha, haustellum, consisting of fused galeae.

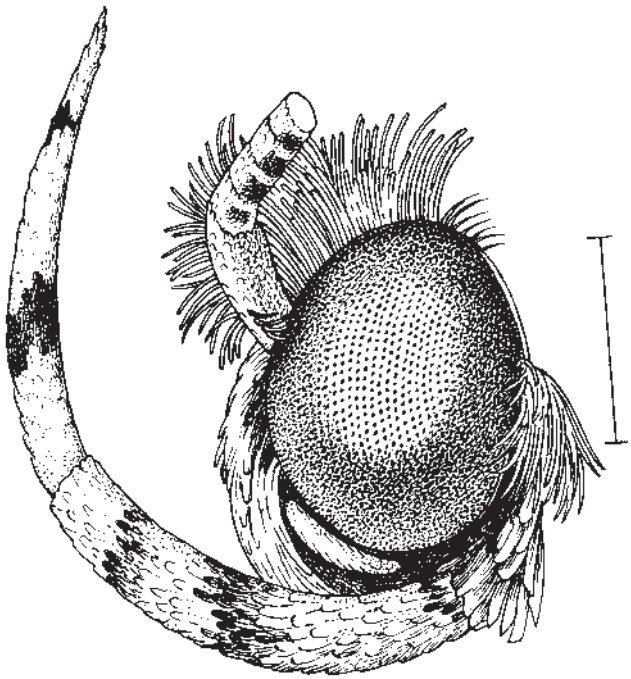

FIGURE 4 Head of ethmiid moth, showing the strongly upcurved labial palpus that is characteristic of most Gelechioidea. Scale bar = 1.0 mm.

proboscis (haustellum) with musculature that enables it to be coiled under the head when not being used to suck nectar from flowers or other fluids into the digestive tract by a pumping action. The maxillary palpi consist of one to five segments and in primitive moths are conspicuous, often folded. The labial palpi are more prominent in most Lepidoptera and vary in curvature and length, but they are not folded.

THORAX The pro-, meso-, and metathorax are fused, each consisting of a series of nonmovable sclerites (Fig. 2). In primitive groups the meso- and metathorax and their wings are similar in size, but in derived families the mesothorax is larger and has more powerful musculature, and the forewing has more rigid vein structure, especially on the leading edge. In the largest superfamily, Noctuoidea, the metathorax is modified posteriorly into a pair of tympanal organs. The tibia of the foreleg has an articulated epiphysis on the inner surface, a uniquely derived feature in Lepidoptera, usually with a comb of stout setae, that is used to clean the antennae and proboscis by drawing them through the gap between the comb and the tibia. The wings are tiny and soft at eclosion from the pupa, then rapidly expand by circulation of blood pumped into the flaccid veins, causing them to extend, stretching the wing membranes to full size, after which they rapidly harden, with the membranes pressed closely together, and the system of tubular veins provides structure. Homologies of the six vein systems are discernible across all families of Lepidoptera, and the configuration of veins has been used extensively in classification. In the most primitive moths the fore wing (FW) and

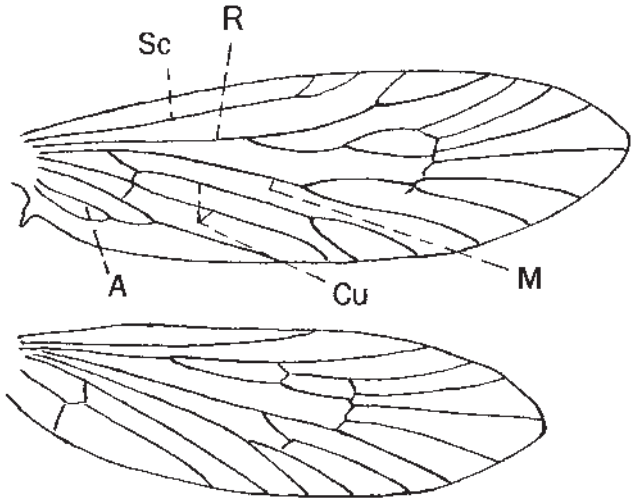

FIGURE 5 Wing venation of a homoneurous moth (Eriocraniidae). Vein systems: Sc, subcostal; R, radial; M, medial; Cu, cubital; A, anal.

hind wing (HW) are similar in shape and wing venation (homoneurous) (Fig. 5), while the more derived groups have lost parts of the vein systems and have fewer remaining in the HW than in the FW (heteroneurous) (Fig. 6). There are various wing-coupling mechanisms by which the FW and HW are linked to facilitate flight. Primitive homoneurous moths have an enlarged lobe at the base of the FW (jugum) that folds under the HW when the insect is at rest but extends over the HW in flight, which does not couple the wings efficiently. Most moths have the HW frenulum that hooks under the FW retinaculum, the development of which varies among taxa and between the sexes of many species.

In a few groups (e.g., Psychidae, Lymantriidae) females of many species are flightless, having very reduced wings (brachypterous), or are apterous and may not even shed the

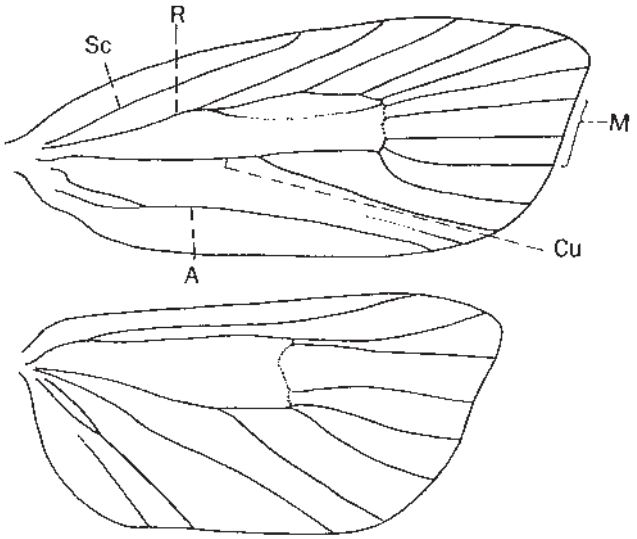

FIGURE 6 Wing venation of a heteroneurous moth (Tortricidae). Abbreviations as in Fig. 5.

pupal skin. Brachyptery has evolved many times independently, such as in high montane and winter-active species of various families in Europe, North America, and Australia. Both sexes are flightless in species of several families on remote southern oceanic islands and in one species of Scythrididae that occurs only on windswept coastal sand dunes in California.

ABDOMEN The abdomen has segments A7–10 or A8–10 modified to form external parts of the genitalia; the sternum of A1 in homoneurous families is small and is lost in other Lepidoptera. Articulation of the thorax and abdomen in derived families is accomplished by musculature attached to sclerotized struts (apodemes) that project from abdominal sternite 2. There are paired tympanal organs at the base of the abdomen in Pyraloidea and Geometroidea. Various male glandular organs associated with courtship occur on the abdomen in several families. Usually these are developed as expandable hair brushes or tufts, or as thin-walled, eversible sacs (coremata), from the intersegmental membrane at the base of the genitalia or on other segments.

The genitalia of Lepidoptera are highly complex and provide the basis for taxonomic species discrimination in most families and often generic or family-defining characteristics. In the male (Fig. 7) the valvae, which are thought to provide clasping stability during mating, usually are large, more or less covering the other structures in respose, and usually are densely setate on the inner surface, scaled exteriorly, and the most visible part of the genitalia externally. The phallus, which is separately articulated and passes through the diaphragma, is sclerotized and contains the membranous vesica, the intromittent organ. The vesica often is armed with cornuti, which sometimes are deciduous and deposited in the female. Sperm are produced in paired testes and pass through a duct leading to the vesica and are deposited in a spermatophore produced by the male accessory glands during mating. The precise functions of most of the external, sclerotized parts of the genitalia are unknown, and they vary independently in form, being uniform in some taxa, variable in others, and thus of differing taxonomic value from one taxon to another.

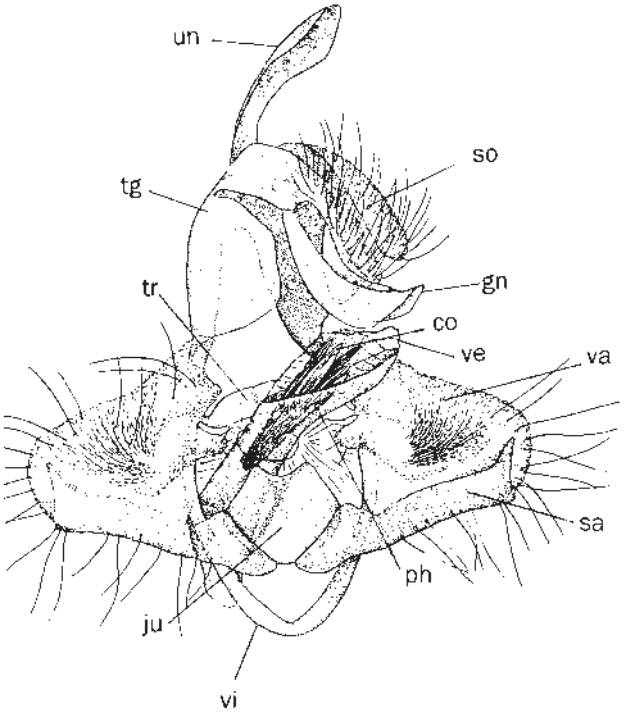

FIGURE 7 Male genitalia of a ditrysian moth (Tortricidae), venterolateral aspect with valvae reflexed. un, uncus; tg, tegumen; so, socii; gn, gnathos; tr, transtilla; ju, juxta; va, valva; sa, sacculus; vi, vinculum; ph, phallus (aedeagus); ve, vesica; co, cornuti.

In the female there are three fundamental types of genitalia. Primitive moths possess a single genital aperture near the posterior end of the abdomen, through which both copulation and oviposition occur (monotrysian). Other Lepidoptera have separate apertures for copulation and oviposition; Hepialidae and related families are exoporian (i.e., the spermatozoa are conveyed from the gonopore, or ostium bursae, to the ovipore via an external groove). All remaining families are ditrysian (i.e., having internal ducts that carry the sperm from the copulatory tract to oviduct) (Fig. 8). This feature defines the Ditrysia, comprising most of the superfamilies and more than 98% of the species. The papillae anales typically are soft and covered with sensory setae but in many taxa are modified for various kinds of oviposition, such as piercing. Both the ductus and the corpus bursae are variously modified in different taxa, the corpus often with one or more thorn-like sclerotized signa that may aid in retaining the spermatophore. Sperm are transported from the corpus bursae through the ductus seminalis to the bulla seminalis and ultimately to the oviduct. The musculature that controls the ovipositor and papillae anales, often involving extension and telescoping the abdomen, as well as the copulatory aperture, is inserted on the posterior and anterior apophyses.

INTERNAL ANATOMY Lepidoptera possess the same fundamental internal systems for breathing, blood circulation, digestion, excretion, central nerves, and endocrine functions as do other holometabolous insects (see relevant articles).

Egg

With few exceptions, female Lepidoptera produce eggs that are deposited externally after fertilization in the oviduct (Figs. 9 and 10). Moth and butterfly eggs vary enormously in size, shape, surface sculpture, and arrangement during oviposition. Within lineages such as families, larger species produce larger eggs, but depending upon the family, the sizes and numbers differ greatly. For example, females of hepialids, including some of the largest moths in the world, produce vast numbers of tiny eggs (20,000–30,000 or more by a single female) that are broadcast in the habitat. Conversely some small moths and butterflies produce few, relatively large eggs.

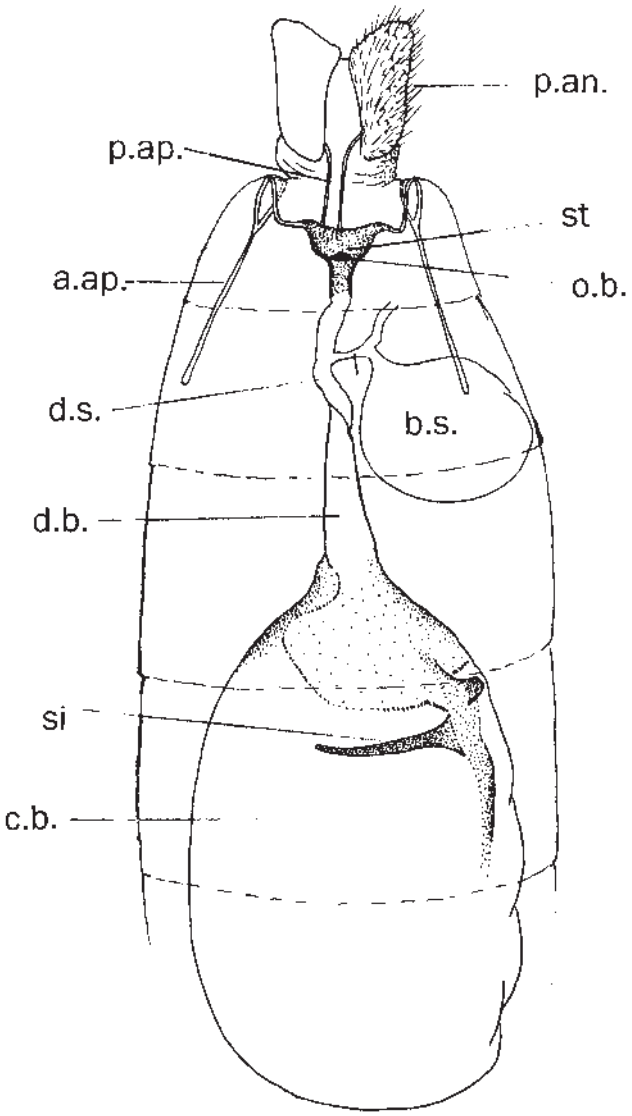

FIGURE 8 Female genitalia of a ditrysian moth (Tortricidae), ventral aspect; broken lines represent segments of abdominal pelt. p.an., papilla anale; p.ap., posterior apophysis; a.ap, anterior apophysis; st, sterigma; o.b., ostium bursae; d.b., ductus bursae; c.b., corpus bursae; si, signum; d.s., ductus seminalis; b.s., bulla seminalis.

The shell (chorion) is soft during development and quickly hardens after oviposition, assuming a regular form consistent for the species and often characteristic for genera or families. The chorion may be smooth or strengthened by raised longitudinal ribs or transverse ridges or both. At one end there is a tiny pore (micropyle), through which the sperm enters, surrounded by a rosette of radiating lines or ridges. Two types of egg form are defined, those laid horizontally, with the micropyle at one end, which are usually more or less flat, and those that are upright, with the micropyle at the top. Flat eggs are prevalent in the more ancestral lineages, microlepidoptera, while most derived groups, larger moths and butterflies, have upright eggs with more rigid and ornamented chorion. Eggs of either type are laid singly or in groups; flat eggs are sometimes

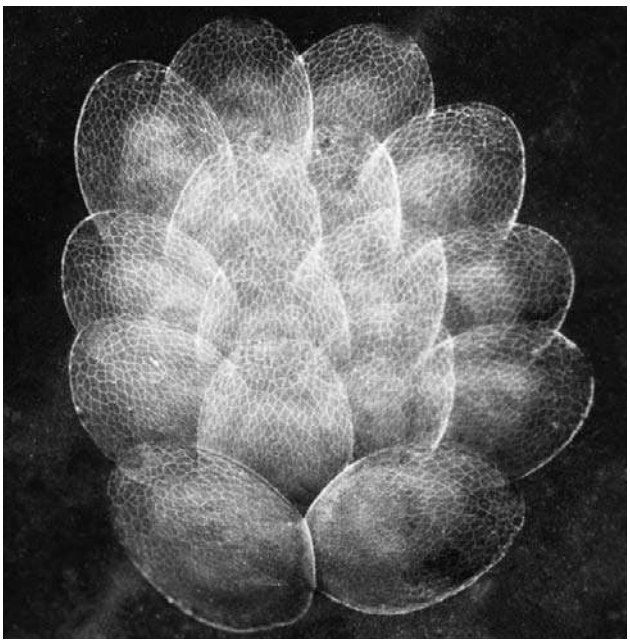

FIGURE 9 Shells of the flat type ditrysian moth eggs (*Amorbia,* Tortricidae), which in this instance are deposited overlapping, in regularly arranged imbricate masses (photograph by A. Blaker).

deposited shingle-like, with the micropylar ends protruding partway over the preceding row (Fig. 9), while upright eggs are arranged side by side, like rows of miniature barrels (Fig. 10). Usually the eggs are glued to the substrate by a secretion of the female accessory (colleterial) glands, applied within the oviduct, sometimes forming a thick, paint-like covering to egg masses. Eggs may be covered with debris collected by the female or hairs or scales from her abdomen or wings or may be surrounded by fences of upright scales, but lepidopteran eggs are not tended or guarded by the adults.

Embryonic development is related to temperature, proceeding more rapidly under warmer conditions, but the rate is physiologically and hormonally controlled in many instances. It requires 7 to 14 days in most Lepidoptera but may be

FIGURE 10 Eggs of the upright type of a ditrysian moth (Arctiidae) (photograph by R. Coville).

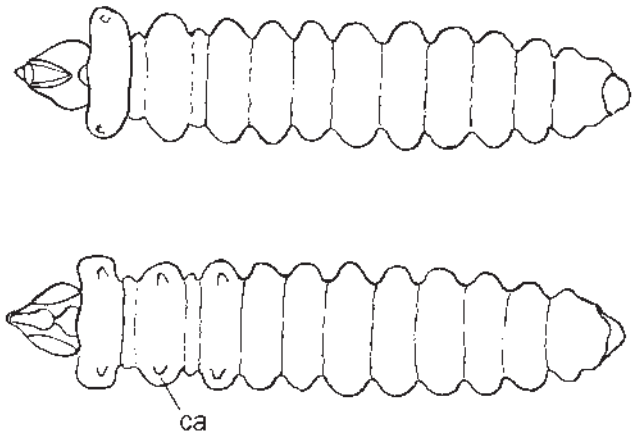

FIGURE 11 Flattened body form of a leaf mining larva (Tischeriidae), dorsal aspect above, ventral below. ca, ambulatory calli that represent vestigial remnants of the thoracic legs.

delayed for many weeks or months in species that overwinter in the egg stage.

Larva

The head (Figs. 12 and 14) is sclerotized, usually rounded (flattened in leaf-mining species, Fig. 11), with large lateral lobes, each bearing an ellipse of usually six simple eyes (stemmata) ventrolaterally and systematically arranged primary setae and are joined by a median suture, which is flanked by two narrow adfrontal sclerites. The mouthparts may be directed downward (hypognathous) or forward (prognathous). The labium is weak but carries a spinneret behind the mouthparts ventrally, which distributes the silk produced by modified salivary glands. The thorax has spiracles on the meso- and metathoracic segments, except in some aquatic pyraloids that have external gills. The abdomen usually has spiracles on segments 1 to 8, restricted to segments 1 to 3 or absent in some aquatic pyraloids. There are paired, ventral, fleshy, and nonsegmented leglike organs on all segments in the most primitive moths, while on others they are restricted to segments 3 to 6 (ventral prolegs) and 10 (anal prolegs), equipped with circles or bands of tiny hooks (crotchets) that aid in grasping and walking. The prolegs are fewer in Geometridae

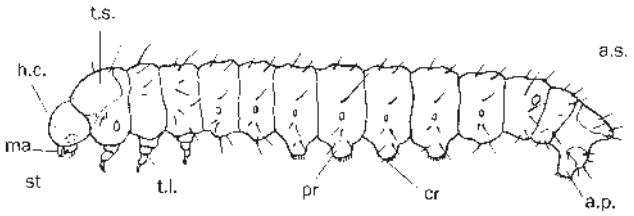

FIGURE 12 Typical form of a ditrysian caterpillar (Cossidae), lateral aspect. h.c., head capsule; ma, mandible; st, spinneret; t.s., thoracic shield; t.l., thoracic leg; sp, spiracle; pr, abdominal proleg; a.s., anal shield; a.pr., anal proleg; cr, crotchets.

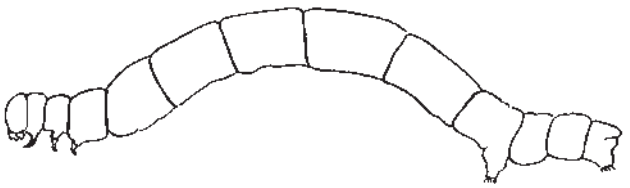

FIGURE 13 Body form of Geometridae larva (inchworm), lateral aspect, lacking prolegs on abdominal segments 1–5.

(Fig. 13) and some other groups and are lost in some borers (e.g., Prodoxidae), leaf miners (e.g., Eriocraniidae, Nepticulidae), and sand-dwelling larvae (a few Noctuidae). In some groups, A10 has a musculated anal fork used to flip frass away from the larval shelter.

There are sensory setae on the head and body integument, and the homology of their primary arrangements (chaetotaxy) (Fig. 15) can be compared in all but the few most primitive families. Their patterns have been valuable to understanding evolutionary trends and to identification of larvae, although the primary arrangement is lost or replaced by numerous secondary setae in many taxa, at least in later instars. The adfrontal sutures, arrangement of stemmata, and crotchet-bearing abdominal prolegs distinguish Lepidoptera from other insect larvae.

Pupa

The head, thorax, and abdomen of the pupa resemble those of the adult and can be recognized externally (Fig. 16). The mandibles of the most primitive families are functional and

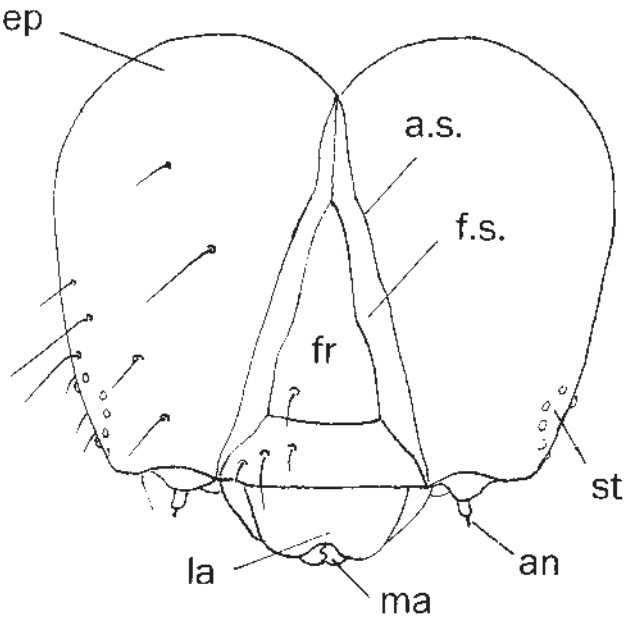

FIGURE 14 Schematic representation of the head capsule of a larval ditrysian moth, frontal aspect. ep, epicranial lobe; st, stemmata; a.s., adfrontal suture; f.s., frontal suture; fr, frons; la, labrum; ma, mandible; an, antenna.

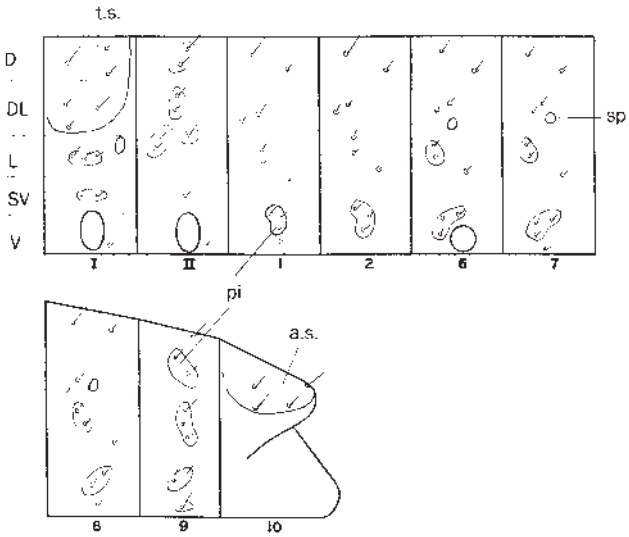

FIGURE 15 Chaetotaxy (setal map) of a larval ditrysian moth (Tortricidae); each rectangle represents one body segment from mid dorsum (upper border) to mid venter (lower border). I, II, pro- and mesothoracic segments; 1, 2, etc., abdominal segments. Setal groups: D, dorsal; DL, dorsolateral; L, lateral; SV, subventral; V, ventral; t.s., thoracic shield; a.s., anal shield; sp, spiracle; pi, pinacula, which are raised and often pigmented.

used to cut open the cocoon preceding eclosion of the adult. In other moths the head is sometimes provided with a beak or other armature that assists in the eclosion process. The appendages of the head and thorax are each encased in cuticle and in most Lepidoptera are fused to the venter of the body, with the wing cases wrapped around, adjacent to the antennae and mouthparts. Abdominal segments 7 to 10 are fused. In the more ancestral families some of the other

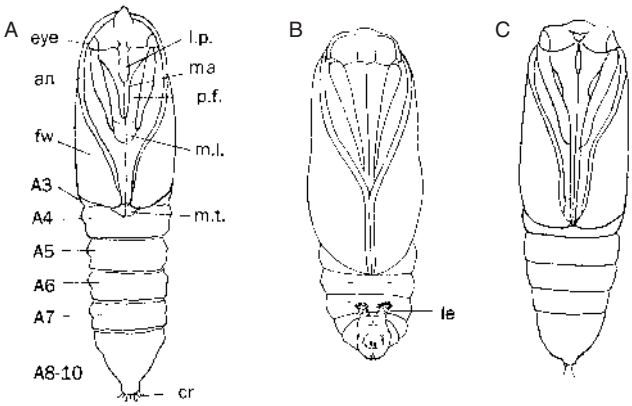

FIGURE 16 Pupae of ditrysian moths, ventral aspect. (A) Tortricidae, with abdominal segments 4–7 movable, enabling pupal movement forward at emergence. (B) Ethmiidae, with pupal movement restricted to flexible segments 5–6, and the pupa remains in place at emergence, a characteristic of Gelechioidea. (C) Noctuidae (Obtectomera) with all segments immobile. 1.p., labial palpus; ma, maxilla including galeae (haustellum); p.f., prothoracic femur; m.l., mesothoracic leg; m.t., metathoracic tarsus; an, antenna; fw, forewing; A3–10, abdominal segments 3–10; cr, cremaster; le, leglike extensions of the 9th abdominal segment bearing hooked setae that anchor the pupa in lieu of a cremaster (A, C redrawn from Mosher, 1916).

segments are movable (Fig. 16A), usually provided with backwardly directed spines or spurs, and the pupa wriggles forward to protrude from the cocoon or burrow just before moth eclosion. Gelechioidea and derived moths (Obtectomera, Fig. 1) and butterflies are obtect, with fused abdominal segments (Figs. 16B and 16C). They remain in place, and adult eclosion occurs along a silken track or other means prepared by the larva or directly from the pupa, in butterflies and some moth groups that do not spin cocoons. Many species have a cremaster, hooked setae at the tip of the abdomen that anchor the pupa inside the cocoon or at the terminus of a silk emergence track, enabling pressure from the emerging adult to break the pupal shell. Others lack the cremaster but are held within a tight cocoon, in an earthen cell, or by a silk girdle. The integument is soft, smooth, and green or whitish when first formed but soon hardens and turns brown in most Lepidoptera. Those that pupate exposed, including butterflies, Pterophoridae, and some Gelechioidea, are mottled green or brownish and often have prominent spines or ridges that aid in camouflage.

BIOLOGY

Success of Lepidoptera populations is dependent upon several factors in the climatic and biotic environment, interrelated with the insects' behavior. First, larval foods, and for most species adult nourishment, must be available. Climatic conditions suitable for mating and oviposition, larval feeding, and pupation are necessary. Females must find appropriate places for deposition of eggs. Larvae must sense proper foods, eat, molt, grow, and pupate. Pupae need to avoid desiccation and other factors that might prevent successful adult eclosion. Finally, egg, larval, and pupal parasites and predators have to combine to take all but two of the offspring of each female (whose eggs may number 200–600 or more) that survive physical dangers, but on average they cannot exceed that, in order to maintain stable population levels.

Adult Behavior

Males usually begin emergence and peak in numbers a few days ahead of females. Both are sexually mature upon eclosion, and males of nearly all moths are attracted by chemical signals (pheromones) emitted by "calling" females. Hence, in most Lepidoptera mating takes place soon after female eclosion, and she has mature eggs ready to be fertilized and deposited within the first 24 h. Mate-seeking involves primarily visual cues in most butterflies, although there may be short-range pheromones produced by one or both sexes that mitigate courtship. Males, and females too in most species, mate more than once. It is assumed that sperm precedence prevails, wherein the most recent male's sperm is effective.

Adults of both sexes of most Lepidoptera feed and in confinement die quickly if water is not available. Feeding on honey-enriched fluids extends the life of some moths and

increases fecundity. Most macromoths and butterflies feed at flowers, imbibing nectar, whereas most micromoths do not and apparently gain nourishment from extrafloral nectaries, sap flows, and honeydew secreted by aphids or other Homoptera. Exceptions occur in diurnal microlepidoptera (e.g., Adelidae, Sesiidae, Heliodinidae, Scythrididae, Plutellidae, and Tortricidae, but not nocturnal species of the latter three families), which visit flowers, often other than the larval hosts. The mouthparts are nonfunctional in a few families (e.g., Lasiocampidae, Lymantriidae) and in specialized species such as winter-active Geometridae and Ethmiidae, and females possess mature eggs upon eclosion.

Host-plant selection is made primarily by the female, which seeks by chemical and tactile cues the proper substrate or habitat for oviposition. This choice is made by instinct, inherited genetically, and the newly hatched larvae also require specific stimuli, detected by chemoreceptors on the antennae and mouthparts; in host-specific species, they starve if the proper plant is not available, ignoring plants or synthetic diets that are quite acceptable and sufficient for nourishment of generalist species.

Most butterflies and moths live only a few days, until mating and egg laying are accomplished, but some are active for several weeks, or they may overwinter as adults and become active on warm days. Some adult microlepidoptera enter a prereproductive state lasting through summer and winter, followed by mating and oviposition in early spring.

Larval Development

The newly formed larva, or caterpillar, first bites its way out of the eggshell, leaving a crescentic slit or ragged hole at the micropylar end. Some species then eat the reminder of the eggshell. All growth takes place during the larval stages, so caterpillars consume enough nutrients to carry through cocoon formation, pupation, and metamorphosis to the adult. It must be sufficient for the moth or butterfly to move to its first feeding or, in species with nonfeeding adults, enough to provide for complete egg development of the next generation. To accommodate growth, the larva molts its skin (cuticle) several times, through successively larger stages (instars). Most Lepidoptera undergo five or six instars, but many larvae that feed on detritus or dry plant material undergo indeterminate numbers of instars.

Silk is produced by paired labial glands. It is composed of two proteins secreted in a viscous fluid in two strands, which consolidate as they leave the spinneret and contact the air. Its functions are many: first instars of many species are dispersed by air currents on silk strands; many or most species lay down a silk line as they move, enabling them to cling to substrates; silk is used by most external-feeding micromoths to form shelters in foliage or other food sources, and some construct portable cases from which they feed; others line tunnels with silk in fruits, stems, roots, or soil from which they forage to feed. Finally, silk is used in cocoon formation

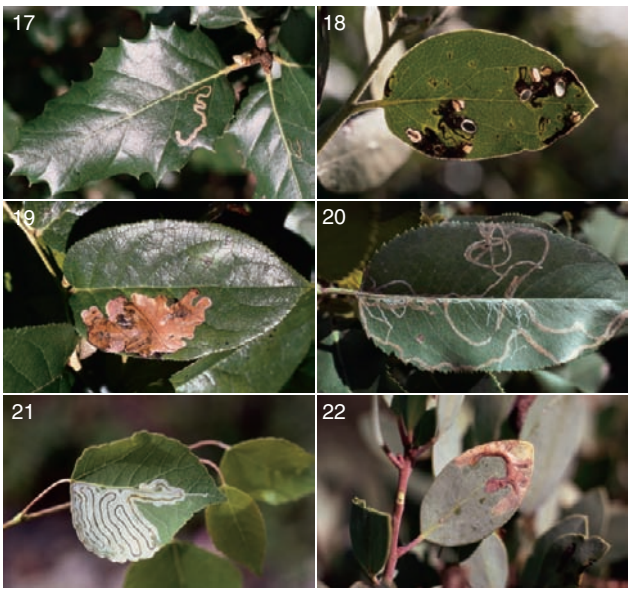

FIGURES 17–22 Leaf mines. (17) *Stigmella variella* (Nepticulidae) on *Quercus agrifolia;* (18) mature larvae of *Coptodisca arbutiella* (Heliozelidae, Incurvarioidea) and their abandoned mines, on *Arctostaphylos;* (19) *Cameraria gaultheriella* (Gracillariidae) on *Gaultheria shallon;* (20) *Marmara arbutiella* (Gracillariidae) on *Arbutus menziesii;* (21) *Phyllocnistis populiella* (Phyllocnistidae, Gracillarioidea) on *Populus tremuloides;* (22) *Epinotia nigralbana* (Tortricidae) on *Arctostaphylos* (photographs by J. Powell, all California, except Fig. 21, Alaska).

preceding pupation, within the larval shelter or gallery or separately, sometimes as a characteristically shaped structure.

Larval habits vary widely and often are quite specific for a family, genus, or species. These include leaf mining, in which a larva spends it entire life within a leaf, and the depth and form of the mines are consistent such that the moth family or genus often is recognizable from the mine (Figs. 17–22). Other types of internal feeding include stem mining; boring in seeds, stems, and roots (Figs. 25 and 26); or feeding in galls developed by plants, stimulated by the larvae (Figs. 27 and 28). Many external-feeding caterpillars avoid adverse conditions by seeking shelter in leaf litter at the base of the plant or in tunnels during the day and emerge at night to feed, when temperatures are cooler, humidity is higher, and diurnal predators are not active. Many macromoth and butterfly larvae remain exposed, motionless, protected by cryptic coloration, body form, and behavior (Figs. 29–32), or even camouflaged by a coat of flower bits or debris that collect on hooked body setae. Larvae of a few genera live gregariously in silken tents that shield them from climatic extremes (Fig. 33). Many others are protected from vertebrate predators by toxic chemicals they sequester, and advertize their presence by bright colors (aposematic) (Fig. 34).

The duration of larval development varies greatly with the feeding and life cycle types, even within families and genera. The time required to reach maturity also is dependent upon temperature within species, such as between seasonal generations. Most Lepidoptera grow slowly in early instars,

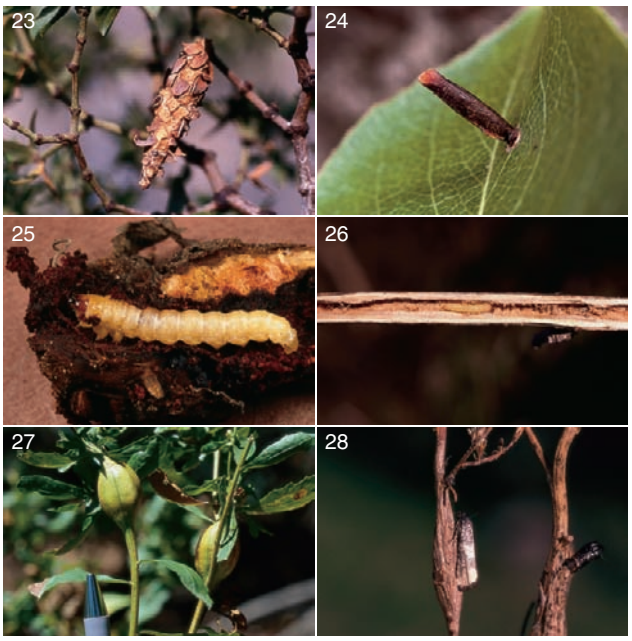

FIGURES 23–28 Case-bearers, borers, and gall inducers. (23) *Thyridopteryx meadii* (Psychidae, Tineoidea), case on *Larrea tridentata;* (24) *Coleophora* species (Coleophoridae, Gelechioidea) on *Malus;* (25) larva of *Synanthedon sequoiae* (Sesiidae, Sesioidea) under bark of a conifer; (26) larva of *Grapholita edwardsiana* (Tortricidae) in stem of *Lupinus arboreus;* (27) stem galls induced by *Gnorimoschema baccharisella* (Gelechiidae) on *Baccharis pilularis;* (28) stem galls caused by *Epiblema rudei* (Tortricidae), with newly emerged moth and its pupal shell on *Gutierrezia* (photographs by J. Powell, except 24, 25 by R. Coville, all California).

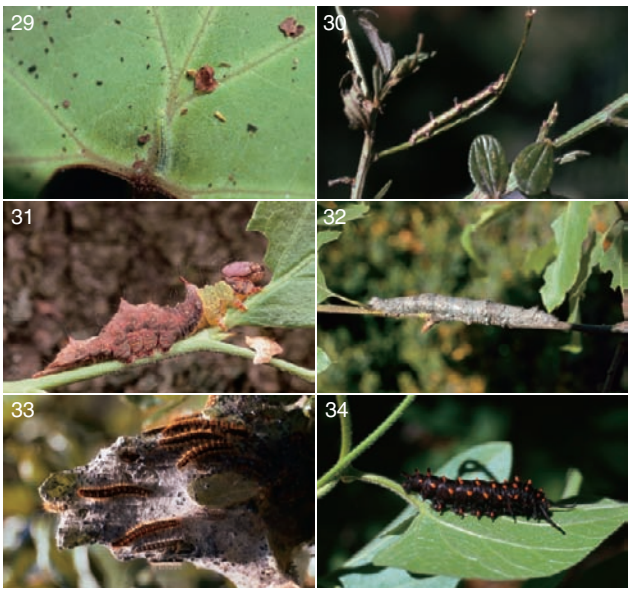

FIGURES 29–34 Cryptic and aposematic caterpillars. (29) *Oidaematophorus* species (Pterophroidae) on *Petasites palmatus;* (30) stick-like larva of *Sicya macularia* (Geometridae) on *Ceanothus thyrsiflorus;* (31) *Schizura unicornis* (Notodontidae, Noctuoidea) on unidentified tree; (32) *Catocala* species (Noctuidae) on *Quercus kelloggii;* (33) tent caterpillars, *Malacosoma californicum* (Lasiocampidae, Bombycoidea), on *Quercus agrifolia;* (34) *Battus philenor* (Papilionidae) on *Aristolochia californica* (photographs by J. Powell, except 31 by R. Coville, 32 by D. Wagner, all California except 31, British Columbia).

increasing in size much more rapidly in later instars, particularly the last. Growth after eclosion from the egg to maturity usually takes 30 to 50 days, but sometimes is more rapid, as few as 18 or 19 days. Larval life can extend much longer, particularly in species that enter quiescent phases at lower temperatures, intermittently feeding when warmer, or in detritus-feeders, which can simply wait long periods when food is not suitable. Such species may live 100 to 140 days before pupation, and those that enter obligate diapause, usually as first or last instar, typically spend 9 or 10 months as inactive larvae in addition to their feeding and growth period.

Larval Foods

The nutritional requirements of many caterpillars are generally similar. Synthetic diets that contain the same basic elements, casein, sucrose, salt, cellulose, wheat germ, amino acids, and vitamins, incorporated in an agar base, are successfully used for rearing many kinds of Lepidoptera. However, sometimes species that are specific to particular plants do not accept a synthetic diet. Hence, nutritional value alone may not be sufficient to elicit feeding, and natural plant chemicals act either as cues for feeding or as deterrents, often the same chemical in both roles with different larval species.

The majority of Lepidoptera caterpillars are phytophagous, consuming living plants, almost exclusively flowering plants, and primarily angiosperms. All parts of plants are eaten, each kind of caterpillar specializing on its particular niche, leaves, flowers, fruit, stems, or roots. Some species feed internally (endophagous) as leafminers and seed or root borers, others externally (exophagous), either concealed in shelters constructed with silk or exposed. Larvae of the most primitive family, Micropterigidae, consume liverworts and mosses or are general feeders on green plants, fern sporangia, or fungal spores in moist habitats. Some other groups of moths do not feed on flowering plants (e.g., Tineidae), but specialize on wood-rot fungi (Polyporaceae) or are detritivores on the ground, under bark of dead tree limbs, or in abandoned insect and spider nests or feed on animal products in mammal burrows, bird nests, or scats, and a few can digest wool. Many species feed on fallen leaves, notably Oecophoridae and Tortricidae on *Eucalyptus* (Myrtaceae) in Australia, and several groups of Noctuidae in wet forest habitats. Some Lepidoptera specialize on lichens (lithosiine Arctiidae, some Psychidae and Xylorictidae), mosses (some Crambidae), or ferns (unrelated species, mainly on oceanic islands). A few Lepidoptera are predaceous on scale insects or other Homoptera or in ant nests. A Hawaiian geometrid moth *(Eupithecia)* is predaceous on adult flies, which it catches by seizing the fly with elongate prolegs. Other members of the worldwide genus *Eupithecia* are plant feeders.

Virtually every kind of flowering plant is eaten by one or more species of caterpillar. Food preferences vary enormously among families; they are summarized in the accounts of the

major families that follow. Nearly all internal feeders, such as leafminers, stem and root borers, and gall inducers, and most other microlepidoptera are specialists on one or a few related plants, whereas perhaps half or more of external-feeding macromoth species are generalists within habitats, such as ground-dwelling cutworms feeding on low-growing herbaceous plants or shrub- and tree-feeding species. Most butterfly species are specialists.

Pupal Development

The duration of pupation during which metamorphosis to the adult occurs varies with temperature, usually requiring about 10 to 12 days, but many species require several weeks or hibernate as pupae, often for 10 months or more.

Pupal movement is an important adaptation in primitive moths and basal Ditrysia. The pupa moves forward just preceding adult eclosion and either anchors by the cremaster to silk or wedges in the emergence aperture, which is prepared by the larva to be slightly narrower than the pupal abdomen. This movement is aided by rows of dorsal, backwardly projecting spines. Gelechioidea and the Obtectomera (Figs. 1, 16B, 16C) have independently derived fusion of abdominal segments that restricts movement, enabling turning within the cocoon but not forward movement, and the adult emerges directly from the pupation site. Pupae respond to tactile stimuli, including potential predators and probing by a parasitoid wasp ovipositor, by turning or wriggling. Some moth pupae have special structures on the abdomen that produce clicking or rattling sounds when the wriggling abdomen strikes the walls of the pupal cells or parchment-like cocoon, or sounds are produced by rubbing fine pegs or rasp-like surfaces on adjacent segments. Such sounds may aid in pupal defense.

Life Cycle

Most Lepidoptera in temperate climates undergo a single annual generation (univoltine), although many have two discrete seasonal broods (bivoltine), and some produce continuous generations as long as favorable temperature conditions prevail (multivoltine). Diapause, a state of arrested development regulated by hormones, controls the life-cycle pattern and enables populations to survive during unfavorable times (winter, dry season, etc.) when necessary resources are not available. Diapause may be the single most important adaptation leading to species radiation of Lepidoptera in northern climates and high mountains, in the world's deserts and tropical dry season habitats, and in other places where insects could not grow and reproduce continuously. In Lepidoptera, diapause occurs primarily in eggs, in first or last instars, in pupae, or as a reproductive delay in adults, depending on the species. In Mediterranean climates, larval feeding typically occurs in spring when foliation peaks, and diapause lasts through the dry season in summer and hibernation in winter. Some species aestivate in diapause as prepupal larvae or pupae, fly in autumn, and then hibernate as adults or eggs. Multivoltine species enter diapause at the end of the growing season, often triggered by decreasing day length, or the larvae simply wait in a quiescent state, feeding slowly on warm days through winter, and metamorphose, and adults eclose with warmer temperature in spring.

Most tropical Lepidoptera are too poorly documented to estimate the proportion of multivoltine to other life-cycle patterns. Some species migrate from wet regions to dry forest habitats at the beginning of the rainy season to take advantage of the newly available resources, but others undergo diapause through the dry season.

Many Lepidoptera are capable of maintaining the diapause to a second or later season if appropriate climatic conditions do not occur. This happens as a regular phenomenon in species adapted to seed feeding on plants with biennial crops such as conifers or sporadically in species that depend upon resources that are limited to a specific season but are erratic in abundance, such as flowering and fruiting by desert plants. Numerous prepupal larvae of yucca moths (Prodoxidae) have metamorphosed synchronously after 8 to 30 years in diapause under experimental conditions.

SIGNIFICANCE IN NATURAL AND HUMAN COMMUNITIES

The major role of Lepidoptera in natural communities is primary consumer of plants. Moths and butterflies make up the largest single evolutionary lineage adapted to depend upon living plants, in terms of species numbers and, in many communities, in biomass as well. Females of most species produce 200 to 600 eggs within a few days, vastly more in some species (1000–30,000), releasing a potentially enormous load of caterpillars onto particular plant species or plant groups such as herbs or woody shrubs and trees. Therefore, an important food resource is available for specialized parasitoid wasps and flies, general invertebrate predators such as spiders, mites, ants, and social wasps, and vertebrate predators, especially birds. There have been estimates of 80,000 caterpillars of several species feeding on a single oak tree and many times that number during outbreaks of single species that defoliate forest trees. Thus caterpillars comprise a major component of biological communities, affecting foraging by birds, buildup of yellowjacket colonies, and insect disease epidemics. A secondary role as decomposers also is filled by Lepidoptera. Tineidae, several groups of Gelechioidea (particularly Oecophoridae in Australia), and some Noctuidae and other moths are detritivores and assist in reducing fallen leaves and fruit, fungi, and animal products (hair, feathers, predator scats) to humus. Finally, a few species are secondary consumers, predaceous on scale insects or other Homoptera in natural communities.

Lepidoptera larvae damage plants grown for human use (food, lumber, cotton, garden ornamentals) and our stored products (grain, flour, nuts, woolen clothes and carpets). Most agricultural damage occurs because monoculture crops

are grown in places distant from the natural enemies of the pest species, which themselves usually have been introduced by human activities to a new region. Wide-scale insecticide suppression of pest species has further increased problems because local parasites and invertebrate predators are eliminated, and the pest species become resistant to the insecticides by selection for survivors of repeated treatments. Similarly, pests of stored food and wool products have been transported worldwide by human activities. Lepidoptera probably are the most important insect group as plant defoliators (e.g., spruce budworm, the economically most important insect in Canada; larch budworm in Europe) and they cause huge losses by damage to fruits (e.g., codling moth, the "worm" in apples), corn (corn earworm, European corn borer), potatoes (potato tuberworm), cotton (pink bollworm), and many other crops and garden plants. They are a major problem in stored meal, grain, and nuts (Angoumois grain moth, Indian meal moth, Mediterranean flour moth) and woolen products (casemaking clothes moth, webbing clothes moth, tapestry moth, and others). Still others infest bee nests, eating the combs (greater and lesser wax moths).

Conversely, some moths are believed to play significant roles in pollination in natural communities, especially Sphingidae and Noctuidae, and they may aid in crop pollination in some instances. Several Lepidoptera have been purposefully introduced to act as biological control agents against noxious plants. Notable examples include a pyralid, the cactus moth, from Argentina used to successfully suppress millions of acres of introduced prickly pear cactus in Australia; an arctiid, the cinnabar moth from Europe, on tansy ragwort in the Pacific states of North America; and several Mexican species against lantana in Hawaii.

FOSSIL RECORD AND EVOLUTION

A widely accepted phylogenetic hypothesis of relationships among lepidopteran evolutionary lineages, based on morphological characteristics in living forms, primarily of the adults, is shown in Fig. 1. The problem in such analysis is that we do not know what kinds of species might have preceded and interceded with the primitive extant lineages, each of which is now represented by one or a few relict genera that have divergent larval features not shared with other Lepidoptera. Moreover, the fossil record is of little use in revealing clues to "missing links," and the preservation usually fails to provide information on critical characteristics, particularly those of the larvae and pupae.

Fossil Record

There are fossils of Triassic age assigned to Trichoptera (caddisflies), the presumed sister group of Lepidoptera, and so branching of the two lineages could have occurred in the early Mesozoic (Fig. 35). The earliest fossil recognized as lepidopteran is a small scaled wing from the Lower Jurassic of

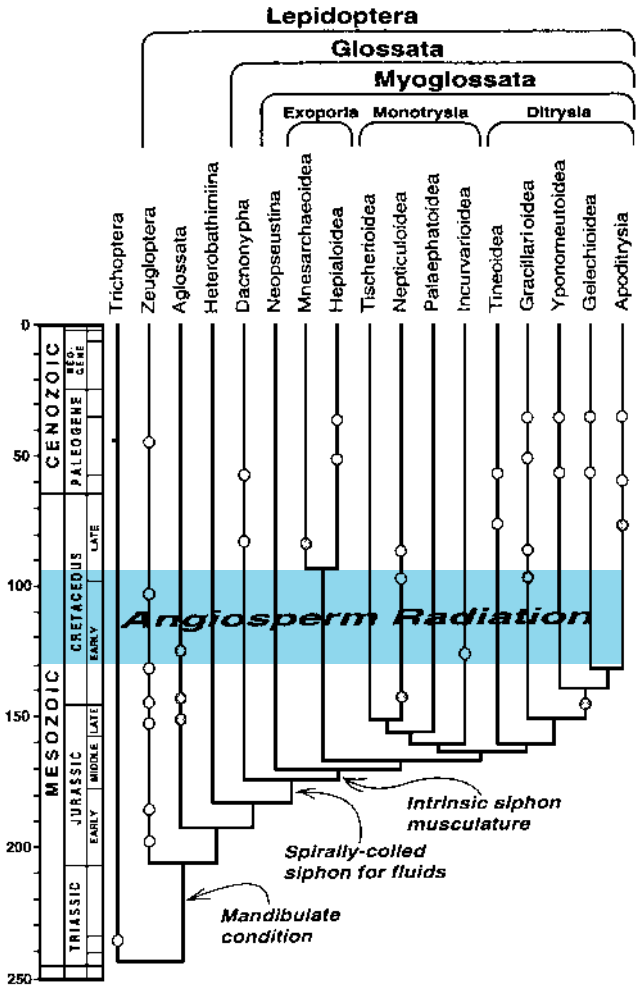

FIGURE 35 Phylogenetic hypothesis of major lepidopteran lineages superimposed on the geologic time scale, with fossil occurrences indicated. Open dots, reliable identifications; shaded dots, questionable assignments. Angiosperm radiation spans 130 to 95 mya from the earliest recognized occurrence of pollen to the time when angiosperms became the dominant vegetation (modified from Labandeira *et al.,* 1994).

Dorset, England. It was placed in a separate family, Archaeolepidae, suggested as a sister group to the Micropterigidae, but without characters known that might establish its relationships. Four genera were described from Upper Jurassic tuffites from Russia. Among these, two were assigned to Micropterigidae and two to Glossata and Ditrysia, but only one of them, *Protolepis,* possesses visible mouthpart structures. They were interpreted as a siphon formed of maxillary galeae, which would imply existence of Glossata, 20 to 30 mya, prior to the radiation of angiosperm plants during the early Cretaceous. That interpretation has been questioned, the structures possibly being maxillary palpi, and therefore the fossil may represent an extinct lineage of Aglossata. By the early Cretaceous there are well preserved Micropterigidae and an incurvariid (Heteroneura) in amber, and by the late Cretaceous several kinds of leaf mines representing modern families and host plant

associations, both heteroneuran (Nepticulidae) and ditrysian (Phyllocnistidae, Gracillariidae), as well as a ditrysian larval head capsule of a free-living form such as Tineidae. That is, the fundamental clades of Lepidoptera are all represented before the beginning of the Tertiary. Hence, although Lepidoptera is the most recently evolved major insect order, its radiation was relatively rapid, paralleling that of the angiosperms, the major lineages having evolved between ca. 140 and 90 mya.

Morphological Evolution

Major changes in morphological adaptation in adult feeding, oviposition mode, wing structure, and larval locomotion are indicated by Figs. 1 and 35. The relict moths of ancient lineages (Micropterigidae, Agathiphagidae, Heterobathmiidae) share features of ancestral mecopteroids, functional mandibles in adults and pupae, similar fore- and hind wings with complete venation, and a single female genital aperture. However, larvae of their extant species differ greatly from one another, each adapted for a particular life-style. Micropterigid larvae are free-living ground dwellers in moist environments, with well-developed thoracic legs, no crotchet-bearing abdominal prolegs, and fluid-filled chambers in the cuticle. Agathiphagids are legless borers in primitive gymnosperm seeds with reduced head sclerotization and sutures and few stemmata. Heterobathmiids are flattened leafminers of southern beech, having a prognathous head with prominent adfrontal ridges, as well as seven stemmata laterally and thoracic legs with large, subdivided trochanters (unique in Lepidoptera), but no abdominal prolegs.

Adult Glossata (Eriocraniidae and all subsequent lineages) lack functional mandibles and feed by a proboscis formed of the maxillary galeae. Basal glossatan lineages have a piercing ovipositor and retain functional mandibles in the pupa, used to cut the cocoon at eclosion. The larvae have a spinneret. Several derived features occur beginning with the Exoporia (Mnesarchaeidae and Hepialidae): The ovipore and gonopore are separate, connected by an external groove for sperm transfer; the larvae have differentiated prolegs on abdominal segments 3 to 6 and 10, with circles of crotchets; and silk is used for various activities, not just cocoon formation, the ancestral condition in Lepidoptera. Functional pupal mandibles are lost and there is no piercing ovipositor. Differentiated size, shape, and venation between fore- and hind wings appear in the Heteroneura. The thoracic legs, crotchet-bearing larval prolegs, and silk webbing are lost by larvae of Nepticuloidea, which are severely modified for leaf mining. An independently derived piercing ovipositor occurs in Incurvarioidea, some of which have secondarily legless larvae.

The last fundamental change, leading to the Ditrysia, is the internal system for storage and transfer of sperm from the gonopore to oviduct. Evidently this had evolved by the mid-Cretaceous, when larval mines of Gracillarioidea appear in the fossil record. The most successful lineages, in terms of extant diversity, Pyraloidea, Geometroidea, and Noctuoidea, which are defined by independently derived tympanal organs, presumably originated coincident with radiation of the bats during the late Paleocene and early Eocene. The earliest butterfly fossils also date from late Paleocene–Eocene times.

Ecological Scenario

Questions remain concerning the origins of angiosperm feeding in basal lepidopteran lineages that led to major radiations of Lepidoptera. The ground-dwelling larvae of Micropterigidae are generalists, either detritivores or fungivores in leaf litter or feeding on low-growing green plants in moist habitats, including bryophytes and soft angiosperm leaves. Similar habits occur in Exoporia (Mnesarchaeidae and Hepialidae, except that many hepialids feed on roots or burrow into stems of woody angiosperms) and in basal Ditrysia (Tineidae, except that none feeds on green plants). By contrast, extant larvae of the other lower Lepidoptera are endophagous feeders that specialize on particular flowering plants (larvae of Lophocoronidae and Neopseustidae are unknown, but their ovipositor types indicate that at least early instars are internal feeders). We assume ground-dwelling, generalist habits are similar to those of mecopteroid ancestors of the Trichoptera–Lepidoptera clade, but we do not know if that mode of life persisted in basal members of all lineages through to the Ditrysia. If so, adaptation to endophagy and to specialist angiosperm feeding might have occurred at least four times, in heterobathmiids, in an eriocraniid + acanthopteroctetid + lophocoronid + neopseustid lineage, in nepticuloids, and, probably independently, in incurvarioids, when a piercing ovipositor reappears, and finally in a palaephatid + tischeriid lineage. If an unknown angiosperm-feeding lineage was the common ancestor, at least two reversals to ground-dwelling, external-feeding, generalist caterpillars characterized by multiple morphological reversals must be postulated for exoporians and again for Tineidae. In either scenario, there were independent origins of a piercing ovipositor (at least twice) and endophagous larval feeding accompanied by numerous derived morphological specializations in larvae (several times). Repeated shifts to angiosperm feeding (Fig. 36) may have been facultative, as it is in extant micropterigids, and multiple adaptations to endophagy imply parallel evolutionary trends, a more parsimonious scenario than multiple reversals to an ancestral morphological and behavioral ground plan.

CLASSIFICATION

Historically the Lepidoptera have been classified in four or five suborders, all but one of which are primitive moths that retain ancestral characteristics as relict, morphologically dissimilar groups. All the more derived moths and butterflies, more than 98% of the described species, comprise one evolutionary lineage, or clade, the Ditrysia. In recent decades, much progress has been made in detailed analyses of the relationships of the

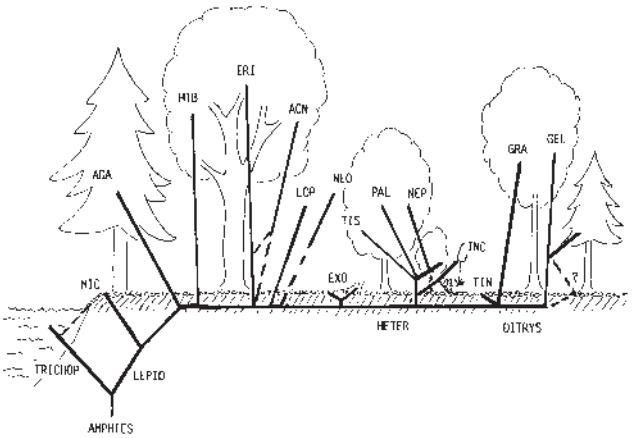

FIGURE 36 Cartoon representing a theoretical scenario of the origins of angiosperm feeding that led to the radiation of Lepidopteran during the Cretaceous. Ground-dwelling mecopteroid-like ancestor gave rise to the Trichoptera-Lepidoptera split, then successively to the ancestor of extant Micropterigidae (MIC) and several specialized, radically differing, angiosperm-feeding lineages. The ancestral ground dwelling caterpillar form is presumed to have been retained in Exoporia (EXO, Mnesarchaeoidea, Hepialoidea) and basal Ditrysia (Tineoidea). AGA, Agathiphagoidea; HTB, Heterobathmioidea; ERI, Eriocranioidea; ACN, Acanthopteroctetoidea; LOP, Lophocoronoidea; NEO, Neopseustoidea; HETER, Heteroneura; TIS, Tischerioidea; PAL, Palaephatoidea; NEP, Nepticuloidea; INC, Incurvarioidea; DITRYS, Ditrysia; TIN, Tineoidea; GRA, Gracillarioidea; GEL, Gelechioidea.

primitive groups, aided by discoveries of new taxa and previously unknown larvae and pupae. Phylogenetic analyses have shown the primitive lineages to be paraphyletic with respect to the rest of the Lepidoptera (Fig. 1), and consequently, the use of suborders and other ranks between order and superfamily has been abandoned by lepidopterists. On the other hand, we continue to recognize the obligate categories (family, genus, species) for purposes of names and communication across related lineages. Historically, the family has been the common denominator level for communication among entomologists, including for Lepidoptera, but in recent decades there has been a proliferation of both family and superfamily divisions such that the superfamily has become a commonly used and understood rank for lepidopterists. Recent authors have treated more than 120 families of Lepidoptera, and there is considerable discrepancy between analyses within some of the 45 to 48 superfamilies. Morphological and biological traits of the larger, worldwide superfamilies and families are summarized in the text that follows.

Primitive Lineages

ZEUGLOPTERA—MICROPTERIGOIDEA Micropterigidae are the most primitive lepidopterans, living fossils. There are micropterigids recognizable as modern genera preserved in amber dating back to dinosaur times in the early Cretaceous (125 mya). Adults (Fig. 37) are small (FW length 3–6 mm), often colorful, with metallic sheens of bronze or purple and yellow forewing markings, usually active in the daytime. They are characterized by numerous ancestral traits not shared by other moths, most notably retention of functional mandibles, which are used to feed on pollen of various trees in Europe, more primitive plants, sedges, Winteraceae, and fern spores in New Caledonia and Madagascar. A more complete, Mecoptera-like wing venation led to proposal of this group as a separate order, the Zeugloptera, but overall evidence indicates the combined Zeugloptera + other Lepidoptera as a sister group to the caddisflies (Trichoptera). The larvae are wholly unlike caterpillars of other Lepidoptera; they are plump, somewhat hexagonal in cross section, with long antennae and short thoracic legs, and they lack the abdominal prolegs with crotchets typical of most Lepidoptera. The larvae live in moist leaf litter among mosses or in rotting wood, habitats with high moisture conditions; the cuticle has specializations unique among arthropods, with exo- and endocuticle separated by a fluid-filled space leading via pores to chambers in the exocuticle, overlaid by sticky pellicle to which particles of debris adhere. The pattern of primary setae on the body is unlike that of other moth larvae. Larvae of some species feed on liverworts, but most micropterigids are generalists, feeding on detritus, fungal hyphae, or angiosperm leaves. About 120 species are known worldwide, in a disjunct, relictual distribution pattern. More than half the named species are in the genus *Micropteryx* in the Palaearctic region, while only 2 are known in North America *(Epimartyria);* there is a greater diversity of genera in the Orient and southwest Pacific, particularly New Zealand, eastern Australia, and New Caledonia, which has about 50 species.

There are two other families of Aglossata: the Agathiphagidae (two species), caddisfly-like moths whose larvae are legless borers in the seed of primitive gymnopserms *(Agathis)* in Australia and south Pacific Islands, and Heterobathmiidae (nine species), which are similar moths to micropterigids but their larvae are leafminers in southern beech *(Nothofagus)* in Chile and Argentina.

GLOSSATA—HOMONEURA The majority of Lepidoptera comprise the Glossata, the monophyly of which is well supported by a suite of derived characters. The most obvious traits that distinguish glosssatans are the adult mouthparts: the mandibles are nonfunctional and maxillary galeae elongated, forming a proboscis that is coiled in repose, accompanied by reduction of the head capsule and its cuticular thickening associated with mandibular musculature. The basal lineages retain ancestral features of the wings: similarly shaped fore- and hind wings with relatively complete venation (homoneurous) and the jugal lobe at the base of the forewing. Females in these families have a flattened, sclerotized abdominal apex with serrate edges, forming a "saw," which is everted to cut into host-plant leaves to deposit the eggs.

Eriocranioidea Eriocraniidae form a Holarctic counterpart to the South American Heterobathmiidae, resembling them superficially as adults and larvae, mining primarily in birch and oak (Fagales) in early spring. Adults are small moths (FW length 4–6.5 mm) with relatively narrow wings covered by

FIGURES 37–64 Adults and larvae of microlepidoptera. Micropterigoidea: (37) *Epimartyria pardella* (Micropterigidae) (California). Incurvarioidea: (38) *Coptodisca arbutiella* (Heliozelidae) ovipositing into leaf of *Arbutus menziesii* (California); (39) *Adela septentrionella* (Adelidae) ovipositing into buds of *Holodiscus discolor* (California); (40) *Greya reticulata* (Prodoxidae), ovipositing into bud of *Sanicula* (California); (41) *Tegeticula maculata* (Prodoxidae) ovipositing into ovary of *Yucca whipplei* (California). Tineoidea: (42) *Tinea pellionella* (Tineidae) (Texas); (43) Larval cases of *Tinea pellionella* on wool fabric (Texas). Gracillarioidea: (44) *Caloptilia reticulata* (Gracillariidae) (California). Yponomeutoidea: (45) *Atteva punctella* (Yponomeutidae) nectaring (Illinois); (46) *Ypsolopha maculatella* (Plutellidae) nectaring at flower of Asteraceae, whereas the larval host is *Ephedra* (California). Gelechioidea: (47) *Antaeotricha* species (Stenomatidae) (Illinois); (48) *Ethmia arctostaphylella* (Ethmiidae), bird dropping-like resting posture on *Eriodictyon,* the larval host (California); (49) Larva of *Ethmia delliella* (Ethmiidae), which feeds on *Cordia* (Costa Rica); (50) *Arotrura longissima* (Scythrididae) nectaring at flowers of *Senecio,* whereas the larval host is *Lycium* (California); (51) *Esperia sulphurella* (Oecophoridae, Oecophorinae) (California); (52) *Callimima lophoptera* (Oecophorinae) (Australia); (53) *Coleophora* species (Coleophoridae) (California); (54) *Holcocera* species (Blastobasidae) (California); (55) *Telphusa latifasciella* (Gelechiidae) (Illinois). Choreutoidea: (56) *Tebenna gemmalis* (Choreutidae) nectaring at flowers of *Achillea,* whereas the larval host is *Wyethia* (California). Sesioidea: (57) *Synanthedon sequoiae* (Sesiidae) (California); (58) *Castnia* species (Castniidae) (French Guiana). Cossoidea: (59) *Acossus* species (Cossidae) (California). Tortricoidea: (60) *Argyrotaenia citrana* (Tortricidae, Tortricinae) (California); (61) *Synnoma lynosyrana* (Tortricidae, Tortricinae), flightless female in calling posture on *Chrysothamnus,* the larval host plant (California); (62) *Pseudatteria leopardana* (Tortricidae, Chlidanotinae), a diurnal and presumed distasteful species (Costa Rica). Alucitoidea: (63) *Alucita* species (Alucitidae) (Colorado). Pterophoroidea: (64) Platyptilinae species (Pterophoridae) (Costa Rica). (Photographs by: I. Common, 52; C. Covell, 58; R. Coville, 45, 47, 51, 53, 54, 55, 56, 57, 60, 64; H. Daly, 50; J. Hafernik, 42, 43, 59; P. Opler, 62, 63; J. Powell, 38, 39, 40, 41, 44, 46, 48, 49, 61; D. Wagner, 37).

iridescent, simple scales and hairs, often golden with purplish markings. Most are diurnal and fly in early spring just as the host trees are beginning to leaf out. The larvae are legless miners in newly expanded leaves, forming "baggy" full-depth mines. They mature quickly and enter the soil for pupation, and the mines dry and deteriorate after the leaf hardens. The pupae are mandibulate, and the emerging pharate adult uses the mandibles to cut through the cocoon and reach the soil surface the following spring. Larval foods are birch (Betulaceae), oak (Fagaceae), and other Fagales, or Rosaceae (1 species). There are about 20 species assigned to five genera, with about half the species in Europe and Asia and half in North America.

A related family, Acanthopteroctetidae, with two species in the western United States, formerly was included in the Eriocranioidea, has been given superfamily status based on its more derived type of scales and first thoracic spiracle. The larvae of one species are miners in Rhamnaceae.

EXOPORIA Within the homoneurous Glossata, two superfamilies comprise the Exoporia, the Mnesarchaeoidea (15 species), a relict group of small, eriocraniid-like moths in New Zealand, and the worldwide Hepialoidea. Monophyly of the two is established by the unique configuration of the female genital system, which is shared by and interpreted as homologous in these otherwise quite dissimilar moths. The copulatory orifice is separate from the ovipore, but there is no internal connection between the two. Sperm is transferred via a groove in the body wall below the ovipore.

Hepialoidea The Hepialoidea is the most successful of the Homoneura and more primitive lineages in terms of extant diversity. The superfamily is characterized by having reduced mouthparts, with the proboscis absent or short and evidently nonfunctional. Hepialidae are large moths, even enormous in some genera, well represented on all nonpolar continents. Four other hepialoid families, Anomosetidae, Neotheoridae, Palaeosetidae, and Prototheoridae, are Southern Hemisphere relicts represented by one to a few species and are smaller moths.

Adults of the Hepialidae are large to very large (FW 10–120 mm), including some of the largest Lepidoptera in the world, *Trichophassus* in South America and *Zelotypia* in Australia, with a 10-in. wing span, and often beautifully colored in greens and pinks. The females carry enormous numbers of eggs—one female of *Trictena* in Australia laid 29,000 eggs and had another 15,000 in her ovaries when dissected—and therefore are bulky, heavy-bodied creatures, surpassing in weight the largest sphingid and saturniid moths. Hepialid males form groups, or leks, that fly together at dusk, as a ritual of courtship behavior; especially suggestive of the common name "ghost moths" is one European species that has white forewings and forms ghost-like clouds. The larvae are elongate, cylindrical, with fully developed thoracic legs and abdominal prolegs that bear rings of crotchets. They have primary setae distributed in patterns that are homologous with those of ditrysian larvae, and they lack secondary setae. Hepialid larvae are concealed feeders, living in silken galleries in leaf litter and grasslands, in tunnels in roots or trunks, feeding indiscriminately on pteridophyte, gymnosperm, or angiosperm plants. Early instars of some species feed on decaying wood and fungi associated with it, and then bore into tree trunks in later instars. There are about 550 named species in 50+ genera worldwide, best developed in Australia and Africa.

HETERONEURA—MONOTRYSIA All the remaining Lepidoptera have different fore- and hindwing shapes and venation, with reduced radial system in the hind wing, and the hind wing usually is smaller. They possess a frenulum–retinaculum wing-coupling mechanism, and they have lost the first abdominal sternite. The five most basal superfamilies of the Heteroneura retain the ancestral monotrysian female reproductive system, but they share no derived characteristics that would unite them as monophyletic. The three most diverse of these are Nepticuloidea, Incurvarioidea, and Tischerioidea.

Nepticuloidea These are tiny moths whose larvae are leaf and stem miners. Specialization on diverse flowering plants has led this group to become the most speciose of the primitive Lepidoptera.

Nepticulidae: Adult nepticulids include the smallest Lepidoptera (FW length 1.5 to 4.5 mm), characterized by having the basal antennal segment (scape) usually greatly enlarged, forming a cap over the upper half of the relatively large eye. The head is rough-scaled, and the mouthparts are primitive, with long, folded maxillary palpi, and rudimentary proboscis with galeae not joined, used to lap up moisture and honeydew secreted by aphids. The FW is relatively broad with long scale fringes. The larvae are legless, obligate leafminers, typically forming a serpentine track beginning just below the egg cemented to the leaf surface, gradually enlarging to a full-depth tube or irregular blotch (Fig. 17). At maturity the larva cuts a crescentic slit in the upper cuticle and drops to the ground to form a tough silken cocoon. Larval foods usually are mature leaves of woody plants, although a few larvae mine stems or cause petiole galls. Individual species are specialists, using more than 40 families of angiosperms, primarily Fagaceae and Rosaceae in the Holarctic; some species groups are specialists on one plant family, such as Anacardiaceae, Polygonaceae, or Fabaceae. There are nearly 800 described species, placed in 11 genera, occurring in all nonpolar regions. No accurate estimate is available, but the total named includes fewer than 10% of the species in tropical regions and probably less than half the North American species.

A related family, Opostegidae (100+ species), are slightly larger moths (FW 1.8 to 8.3 mm), with enormous eye caps, completely obscuring the eyes from a frontal view. The FW is relatively broad, white with sparse black markings, and the apices often are strongly bent upward. The larvae are leafminers in Rutaceae and cambium miners in stems and fruit of Betulaceae, Ranunculaceae, Polygonaceae, Saxifragaceae in the Holarctic, and Fagaceae in Chile.

Tischerioidea Tischeriidae adults are very small (FW 2.7 to 5 mm), with lanceolate wings of white, gray, or yellow. They

are nocturnal with large eyes and when at rest they perch with the head appressed to the substrate and tail end lifted at a 45° angle. The larvae are slightly flattened leafminers with thoracic legs reduced to two vestigial segments or ambulatory calli, abdominal prolegs rudimentary with crotchets (Fig. 11). The linear or blotch mines are characterized by a heavily silk-lined nest within which the larva retreats when not feeding; the mines have been recorded from nine angiosperm families, most commonly Fagaceae, Rosaceae, and Asteraceae, more diverse on the last than is true of other lepidopterous miners. This group is primarily Holarctic, with about 80 described species in one genus, a few of which are in the Neotropical, Ethiopian, and Indo-Malayan regions, none in Australia and Oceania.

Incurvarioidea These are tiny to small moths having diverse biologies, but females all have a piercing ovipositor specialized for inserting the eggs into plant tissue, often the ovules or young seed. There are six families, Cecidosidae (seven species, gall inducers in South America and Africa), Crinopterigidae (one Mediterranean species, a larval case bearer with habits similar to some coleophorids), and four that are diverse and widespread: Heliozelidae, Incurvariidae, Adelidae, and Prodoxidae.

Heliozelidae: Species of this family occur worldwide but because of their minute size and diurnal habits they are rarely seen and many more species are known from the characteristic abandoned larval mines than there are named. Adults (Fig. 38) are tiny to small (FW 1.7 to 7.0 mm), typically with iridescent, metallic-appearing scaling. The eyes are small, characteristic of diurnal microlepidoptera. The larvae are flattened, usually legless leaf miners, having a thorax with paired ventral and dorsal movable calli; abdominal prolegs are absent. Early instars form a short, serpentine mine, and then enlarge it to a full-depth blotch (Fig. 18). The last instar constructs a portable case by cutting lenticular disks from the upper and lower epidermis and joining them with silk, giving rise to the common name "shield-bearers"; the abandoned mines, with their distinctive "shot holes" are highly characteristic of heliozelids (Fig. 18). The larva crawls off and descends by a silken thread to attach to a lower leaf or bark, where pupation occurs in the portable case. Heliozelids are host specific, using at least 17 families of usually woody angiosperms, with preponderance in Myrtaceae in Australia and Cornaceae and Vitaceae in the Holarctic, the only Lepidoptera to specialize on the latter. There are more than 100 described species in about 12 genera, distributed in all major faunal realms except New Zealand but poorly known in tropical regions.

Incurvariidae: Adults are small (FW 3.5 to 9 mm), with rough head scaling, relatively small eyes, and proboscis short, half the palpi length; maxillary palpi are elongate and folded. They are somber moths with dark, monochromatic wings, sometimes iridescent brown, bronze, or bluish. The larvae are moderately flattened, with well-developed thoracic legs and reduced abdominal prolegs. Early instars form blotch mines; later the larvae cut through the upper and lower epidermis to remove oval sections, which they sew together to form a portable case. Larvae of a few genera remain in the mines throughout feeding, and then cut out a case in which they pupate. Oviposition is host specific; the ancestral, southern-continent genera use Myrtaceae or Proteaceae, while Holarctic incurvariids use about 10 unrelated angiosperm families. There are about 100 described species in 11 genera, mostly Australian and Palaearctic; they are poorly represented in Africa and the Western Hemisphere.

Adelidae (Fig. 39) are best known for their enormously long antennae, often 2.5 or 3 times the forewing length. Usually they are much longer in the male, which in many species possesses greatly enlarged eyes, but the eyes are small in some species, irrespective of antennal length. Holarctic and Neotropical species (Adelinae) are small (FW 4.5–9 mm), diurnal moths, often brightly colored, iridescent green, blue, or purplish, with white antennae, while the primarily African Nematopogoninae are crepuscular or nocturnal and dull colored. Both sexes have a well-developed proboscis and seek nectar from various flowers other than the larval food plant. Males of the large-eyed species form small, dancing groups, reacting to one another during mate seeking. Females insert the eggs into the base of the ovaries of unopened flowers. First instars of the few species studied in detail feed in the developing ovules; after molting they drop to the ground and construct flat, portable cases from silk and debris and feed on fallen leaves or the lower leaves of the host plant, which often are short-lived annuals. Pupation the following spring occurs in the figure 8-shaped case, with the long antennae free and coiled several times around the abdomen. Oviposition is restricted to one or a few closely related plants, which include members of at least 18 angiosperm families. Biologies of the Nematopogoninae of the Southern Hemisphere are poorly known. There are more than 300 species in five genera, occurring in all faunal regions except New Zealand.

Prodoxidae are famous for the close symbiotic relationship between species of *Tegeticula* and yucca plants (Agavaceae). Females possess enormous "tentacles," appendages of the maxillary palpi, which are unique among all insects, used to gather pollen that is purposefully transferred to the stigmas while visiting other flowers for oviposition, thus ensuring cross-pollination. Other kinds of insects are not attracted to yucca flowers to collect pollen. Females are believed to leave a pheromone signal at the oviposition sites that deters later visiting females so that only a few larvae feed in any given seed pod and many unaffected seeds are produced. Adults (Figs. 40 and 41) are small (FW 4 to 16 mm), generally dull colored, white or gray, although a few *Greya* and *Prodoxus* species have patterned or iridescent bronze-colored forewings. The maxillary and labial palpi are relatively prominent but usually shorter than the proboscis. The Agavaceae-feeding prodoxines apparently do not seek nectar, although individuals of *Tegeticula maculata* have been recorded living up to 9 days in the field. Early instars of the more ancestral genera (*Lampronia, Greya*) feed in young ovules of the host plant and then leave to spin overwintering shelters on the ground.

In early spring they feed in flower beds or foliage shoots of the newly foliating host plants. These caterpillars and those of the pollen-carrying genera are stout, highly mobile, with well-developed thoracic legs, lacking abdominal prolegs, while those of the bogus-yucca moths *(Prodoxus)* are completely legless and apparently blind, living their entire life within the gallery and pupating there. Prepupal larvae of yucca moths are capable of maintaining the diapause for several years if optimal winter conditions are not experienced, up to 30 years followed by successful, seasonally synchronized development, in experimental trials. Holarctic Lamproniinae and species of the basal prodoxine genus *Greya* specialize on Rosaceae, Ericaceae, or Saxifragaceae, while the more derived prodoxines are Agavaceae specialists. Prodoxidae are predominately Holarctic, with Lamproniinae mainly Palaearctic and Prodoxinae largely Nearctic, with a few species ranging into southern Mexico. About 75 species in 10 genera are known.

Ditrysia

The Ditrysia includes 98% or more of the described species, most of the superfamilies and families, almost all of the external plant-feeding caterpillars, and most of the special adaptations for prey avoidance. All members possess reproductive systems based on separate female copulatory and oviposition orifices with internal ducts for transfer of the sperm.

TINEOIDEA The tineoids are generally recognized as the most ancestral living group of the Ditrysia. Most tineoids have erect, roughened head scaling and elongate, five-segmented maxillary palpi that are folded, usually longer than the labial palpi, which have lateral bristles, while the haustellum has short, unconnected galeae, used to lap up surface moisture from detritus or fungi. Females of most species possess elongate apophyses of segments A9 and A10 that anchor musculature, enabling the ovipositor to be telescoped outward to inject the eggs into crevices or other niches in the habitat. Five families are regarded as comprising the superfamily, two of which are worldwide and more species rich, Tineidae and Psychidae. The others are smaller families of restricted distribution, Eriocottidae (70+ species) in the Mediterranean region and southern Africa to Australia and Taiwan, Acrolophidae (280 species) in the Neotropical and Nearctic regions, and Arrhenophanidae (30 species), Neotropical.

Tineidae (Fig. 42) are slender, small to moderately large moths (FW length 2.5–25 mm), usually shining brown, tan, or whitish, with FW patterns of black on pale or yellow on dark. Tineids are most easily recognized by the rough head vestiture and the short (or absent) proboscis. They lack bipectination of the male antennae, characteristic of other tineoid families. The larvae are slender with integument usually lacking color pattern, often living within silken tubes or portable cases (Fig. 43). All instars have well-developed thoracic legs and abdominal prolegs with a single circle of crotchets. Larval foods—Tineids do not feed on flowering plants; they mostly are generalist detritivores or fungivores, and members of some subfamilies tend to be specialists on animal products such as fur or feathers (e.g., Tineinae). Some are capable of digesting wool, including several cosmopolitan species that feed on woolen clothes and other manmade products. Others are primarily fungus-feeders (e.g., Scardiinae, Nemapogoninae), especially in sporophores of wood-rot fungi (Polyporaceae) or wood permeated by the hyphae, sometimes quite specialized in host preference. Fungivory or detrivory presumably was the ground plan for the family and therefore for the Ditrysia. Larvipary, wherein eggs mature and first instars emerge within an enlarged oviduct in the female, is known in numerous Andean and Indo-Australian Tineinae. Many fewer eggs are produced than by tineids with conventional reproductive systems. There are more than 3000 described species worldwide, probably less than half the number known in collections, especially in tropical regions. These are assigned to more than 300 genera in 15 subfamilies.

Psychidae: The common name "bagworms" applies to psychids because the larvae live in portable cases constructed from silk, plastered with debris or symmetrically arranged pieces of host plants (Fig. 23). Adults are small and slender to rather large and heavy bodied (FW length 4–28 mm). Males are fully winged, while females of some species may be fully winged, short-winged, wingless, or even larviform and never leave the larval case. Some species are female only (parthenogenetic) or bisexual only in some populations. The head vestiture is roughened, with long, slender scales directed forward, and the antennae often are strongly bipectinate in males, particularly in species with flightless females, but are filiform in both sexes of species having winged females. Nearly all psychids are gray or brown without color patterns. Psychid larvae are stout compared to tineids, with the head and thorax larger and more heavily sclerotized than the posteriorly tapered abdomen and variously pigmented. The thoracic legs are well developed and are used to pull the cases along on the host plant, while the abdominal prolegs are reduced. Bagworms feed on lichens, grasses, conifer foliage, or leaves of angiosperm trees and shrubs, sometimes as specialists but often as generalists. At maturity the larva attaches the case to a substrate and then inverts itself and pupates in the case with the head toward the distal (older) end, whence the moth emerges. There are nearly 1000 described species from all faunal regions, about 85% in the Old World. Psychids are generally better studied than most microlepidopterans, owing to their fascinating behavior, biologies, and genetic complexity associated with the larviform females and parthenogenesis in five unrelated genera.

GRACILLARIOIDEA This is the major clade of Lepidoptera adapted for larval mining in leaves (Figs. 19–21). Gracillarioids primarily mine woody trees, shrubs, and vines of angiosperms and conifers. The larvae are obligate leaf-, stem-, or fruit-miners in early instars; in many genera larvae leave the mines to feed exposed or in webs externally. The adults lack the tineoid lateral bristles of the

labial palpi, have a smoothly scaled frons, and usually have a well-developed, elongate, coiled proboscis.

Gracillariidae: Adults (Fig. 44) are nocturnal, often are brightly colored, with the FW patterned in metallic orange, bronze, purple, or yellow. They are small, slender moths (FW 2–10 mm) with a head with smooth scaling directed forward over the front and tufts of erect scales on the crown in Lithocolletinae; the antennae are 0.8 times to much longer than the FW and filiform. The HW is lanceolate with scale fringe broader than the wing. The larvae characteristically are hypermetamorphic with more than one form in successive instars; early instars are modified for mining, flattened with legs reduced or lacking, transforming in the third or later instar to cylindrical caterpillars, with chewing mouthparts and fully developed legs. They use silk to buckle the mine into a tent-like shelter or feed externally, often folding a leaf into a tightly closed shelter in which they graze. In some genera a variously modified, nonfeeding instar spins the cocoon. Pupation occurs outside the mine in most genera. Nearly all gracillariids are specialists on one or a few closely related plants, typically woody angiosperms, including more than 80 plant families. More than 2000 species have been described from all major faunal regions, assigned to about 75 genera, and certainly a much greater number remain to be defined, especially in tropical forests.

Phyllocnistidae: Adults are tiny, slender moths (FW 2–3 mm) with long antennae, often with shining white or silvery FW, delicately banded with gray and rust distally; HW are lanceolate with a much broader fringe. The larvae are flat with legs reduced to stubs, mouthparts highly modified for sap feeding; they create extremely long, meandering or regularly zig-zag, subcutaneous mines (Fig. 21), often in new, still-soft leaves, causing them to curl conspicuously. More than 20 angiosperm families have been recorded as larval hosts, but many others in Central America are used, judging from the ubiquitous mines, probably all made by undescribed species. Phyllocnistid mines are described from mid-Cretaceous (97 mya) Magnoliidae, the earliest known ditrysian leaf mining. Fewer than 100 species are described, a small fragment of the fauna; mines are found on more kinds of plants at one lowland forest locality in Costa Rica than there are named New World species.

Bucculatricidae: Adults are tiny to small moths (FW 2.5–7 mm), most easily recognized by their elongate frons and large, erect tuft of scales on the vertex. The appendages are short, antennae 0.6 to 0.9 times the FW length. The wings are lanceolate, FW often with tufts of upraised scales. Larvae are hypermetamorphic; they are legless leaf miners in the first two instars and later have well-developed legs, feeding externally as exposed grazers. Bucculatricid species are host-plant specialists, with more than 20 angiosperm plant families recorded, Asteraceae and Fagaceae dominant in the Holarctic; many species use Cupressaceae. There are about 250 described species, distributed on all continents except New Zealand, most numerous in the Holarctic.

YPONOMEUTOIDEA This superfamily includes a heterogeneous conglomeration of dissimilar microlepidopterans that are grouped by default, i.e., the nonapoditrysian Ditrysia that have nonmotile pupae and lack the scaled proboscis typical of Gelechioidea.

Yponomeutidae: Adults are slender moths with elongate FW, ranging from tiny (FW 3.2–6.8 mm), metallic golden, purple, or gray and white nocturnal *Argyresthia* to larger (FW 9–15 mm) and brightly colored, diurnal moths in *Atteva* (Fig. 45) and the white ermine moths in *Yponomeuta.* Larvae of typical yponomeutines live communally in extensive webs on trees and shrubs, sometimes causing economic damage to fruit trees; those of some *Zelleria* damage growing tips of conifers. Larvae of *Argyresthia* are miners in angiosperm buds or conifer foliage. Most species are specific to one or a few plants; Argyresthiinae feed on at least 13 gymnosperm and dicot families, with more than 40% of recorded species on conifers, including 25% on Cupressaceae, a greater degree of adaptation to conifers than by other Lepidoptera. There are about 600 described species, occurring in all biotic regions, with both the phylogeny and the taxonomy in tropical and south temperate faunas yet to be resolved.

Plutellidae and Ypsolophidae traditionally were treated as one family. The adults (Fig. 46) are small moths (FW 6–13 mm) with distinctive labial palpi, with the second segment broadly scaled and the third slender, smooth scaled and upcurved from the second preapically. FW are narrow to lanceolate, with a flared terminal fringe. Large pleural lobes enclose the male genitalia in ypsolophids but are small and narrow in plutellids. These are typically nocturnal moths with yellow, brown, or gray FW, often with linear markings. The diamondback moth *(Plutella xylostella),* a ubiquitous pest of cabbage, cauliflower, and other plants of the mustard family, is the best known plutellid. The larvae are slender, tapered toward both ends, often with elongate abdominal prolegs, and pale green with unpatterned integument, and they live in slight webs as external feeders. Pupation occurs in large-meshed, open cocoons (*Plutella* group) or dense, envelope-like cocoons (*Ypsolopha* group). Most *Plutella* feed on Brassicaceae, the only moth lineage adapted to do so, while members of this family group as a whole use 50+ families of angiosperms, rarely monocots, and a few gymnosperms, including Ephedraceae and Cupressaceae. Nearly 300 described species are assigned to one or the other of these families, but the systematics relationships and descriptive inventory are incompletely known, especially in Southern Hemisphere faunas.

Glyphipterigidae are small (FW usually 3.2–10 mm), diurnal moths with a smooth-scaled head and porrect or decumbent, slightly upcurved labial palpi, often metallic gray; the FW have metallic markings and parallel, white chevron marks from costa and dorsal margin. The last abdominal tergum is greatly enlarged, forming a hood over the genitalia. The larvae are borers in seed, flowering stems, terminal buds, or leaves, primarily in sedges and rushes (Cyperaceae, Juncaceae), less commonly in grasses, and rarely in dicots,

including Crassulaceae. Nearly 400 species are described in 20+ genera. The family is cosmopolitan and well represented in temperate, Palaearctic, Nearctic, and Australian regions, including New Zealand.

Heliodinidae: Adults are tiny, diurnal moths (FW mostly 3.2–5 mm; rarely to 8 mm), resplendent in shining metallic body and wing scaling, often red or orange with raised silver or lead-colored spots. Adults of many species hold the hind legs aloft when perched, which has been regarded as characteristic of heliodinids, but not all of them do so. Larvae are unpigmented, grub-like, and host specific as leafminers or stem or seed borers, or a few feed externally in flowers and fruit. This is the only lepidopteran family to specialize on Caryophyllales (90% of the known hosts), especially Nyctaginaceae. About 75 species are described worldwide, but the majority occur in the southwestern United States and Mexico.

GELECHIOIDEA This is the largest superfamily of micromoths by far, and because vast numbers of species remain undescribed, Gelechioidea may surpass Noctuoidea as the most diverse group of Lepidoptera. For example, even in North America only about one-third of the species known in collections are named, and a total of 4400 species is projected, 1000 more than Noctuoidea. In tropical regions the inventory is imponderably incomplete—more than 1100 species of Gelechioidea have been counted at one rainforest reserve in Costa Rica, comprising 20% of all Lepidoptera believed to occur there, the majority unnamed. In Australia, gelechioids are estimated to make up about 40% of the Lepidoptera species.

Gelechioids all have overlapping scales on the dorsal surface of the haustellum, up to half its length, and most have a smooth-scaled head, four-segmented maxillary palpi, and upward curved labial palpi with the third segment long and acute (Figs. 4 and 55). The pupa is nonmotile (obtect), remaining in the cocoon until emergence of the moth. There is general agreement on the phylogenetic unity of this superfamily, but there have been wide differences of opinion on the number and relationships of the included families in recent analyses. About 25 groups have been treated as families, of which only the larger, worldwide ones are mentioned here.

Elachistid Assemblage This group of taxa is defined by having modified abdominal articulation in the pupa, lateral condyles, on A5–A6 and A6–A7, that prevent lateral movement (Fig. 16B), although those of Elachistidae s. str. are polymorphic and questionably homologous. The first three groups have been treated as families or subfamilies of Oecophoridae or Elachistidae. They are broad-winged moths with strongly curved labial palpi, often exceeding the top of the head (e.g., Fig. 4).

Stenomatidae: Adults (Fig. 47) are small to moderate sized (FW 5–25 mm), with rectangular to nearly oval FW. The valvae of male genitalia have setae with prominent, multilobed apices. The male antennae usually have long cilia. Larvae of only a small proportion of described species are known; they are relatively stout caterpillars, often with heavily pigmented integument. They are external feeders in concealed shelters on diverse angiosperms (16+ families), predominately Myrtaceae in the Southern Hemisphere and Fagales in the Nearctic. More than 30 genera and 1200 species are named, 90% from the Neotropical region, where many are not yet described. This group and the Ethmiidae tend to be mutually exclusive on a broad geographical scale. Stenomatids are species rich in the southeastern United States and lowland wet forests of central America and northern South America, while ethmiids are speciose in arid parts of western North America and thorn forest regions of Mexico, Central America, and the Antilles.

Ethmiidae: Adults (Fig. 48) are small moths (FW mostly 4–16 mm, rarely 24 mm), having elongate, narrow FW, often dark with a sinuate pale band along the dorsal edge that renders a bird-dropping appearance when the moth rests, or white with black spotting, superficially resembling *Yponomeuta,* and some tropical species are colorful. Most are nocturnal, but some high montane species are diurnal, as are a group of species in the southwestern Nearctic that fly in early spring, adapted to use annual plants. The family is characterized by a strongly recurved basal part of the phallus, secondary SV setae of the larva, and two separately derived pupal-anchoring mechanisms, either development of "anal legs," ventral, setiferous, forward-directed extensions of the ninth segment (Fig. 16B) or grasping of the exuvial head capsule or cocoon silk between abdominal segments 6 and 7. Diapause occurs in the pupal stage, and development can be delayed several years, an adaptation to unpredictable, arid habitats. About 80% of the species for which the often colorful larvae (Fig. 49) are known feed on Boraginales (Boraginaceae, Ehretiaceae, Hydrophyllaceae). About 300 species are named worldwide, with the greatest richness in areas of seasonal drought, especially microphyllous thorn forests of the northern Neotropical Region. This group is better studied than most micromoths, excepting the African fauna, and most species in collections are described.

Depressariidae: Adults are small (FW 7–16 mm) and resemble the two preceding groups, with FW usually rectangular and labial palpi slender and strongly curved upward, lacking the multilobed setae of the valvae, the strongly recurved phallobase, and special pupal anchoring mechanisms that characterize stenomatids and ethmiids. The larvae mostly are leaf tiers but some bore into stems or seed, using at least 17 families of dicots, with strong specialization on Apiaceae and Asteraceae in the Holarctic. More than 600 species are described, assigned to 80+ genera, occurring in all major faunal realms and best represented in north temperate and tropical regions.

Elachistidae: Adults are tiny (FW 2.5–6.5 mm) with narrow wings; the HW fringe is much wider, and FW are usually white or black with white markings. The labial palpi are slender, strongly curved upward. The larvae are flattened, with head prognathous and recessed into the first body

segment and legs short, adapted for mining. *Elachista* typically mine monocots (Poaceae, Cyperaceae, Juncaceae), while other species are miners in a few dicot families. There are about 250 described species worldwide, but they are mainly Holarctic.

Xyloryctid Assemblage The remaining families of Gelechioidea lack modified lateral articulation on the abdominal segments of the pupa but otherwise are not defined as a lineage by a derived feature.

Xyloryctidae: Adults are small to moderately large (FW 10–33 mm), nocturnal moths, often brightly colored, shining white or yellow patterned with black or brown, and relatively heavy bodied, having spiniform setae on the posterior part of abdominal terga 2 to 6. The labial palpi usually are long, strongly curved, and slender. Larvae are robust caterpillars with a bordered submental plate and secondary SV setae on abdominal segments 3 to 6. They form silken tubes or shelters in lichens, on bark, or among foliage. Some species feed in bark or tunnel in bark or stems and drag leaves back to the gallery at night. Larval foods are lichens, living angiosperm plants (20+ families), about half the known species on Myrtaceae and Proteaceae, or dead eucalypt leaves. More than 500 described species are assigned to 60+ genera, in Africa, the Indo-Australian region, and Polynesia, with the greatest numbers in Australia.

Scythrididae are tiny to small (FW 3–12 mm) stiletto-shaped moths with narrow wings that wrap around the body, rendering a tapered appearance from thorax to wing tips (Fig. 50). Most are somber colored, gray or brown, with darker or white FW markings, the nocturnal species tending to be white or pale gray with larger eyes, while diurnal scythrids are mostly dark brown with small eyes and visit flowers for nectar. The male genitalia display an astonishing array of forms, from a relatively unmodified gelechioid plan in most nocturnal species to an extremely reduced and modified by fusion, often asymmetrical form that defies interpretation of homologies of the structures, mostly in diurnal forms. A remarkable species on coastal sand dunes in California, *Areniscythris brachypteris,* is flightless in both sexes, has greatly enlarged hind legs that enable it to leap 20 times or more its body length, and buries itself at night. The larvae are slender, with a small head, tapered toward both ends, usually without integumental markings but with sclerotized rings around setae SD1 on the abdomen. The head capsule has a submental pit in most genera, at least in early instars, which with the setal rings indicates relationship to Blastobasidae. Most species live in frail webs and feed on growing tips of herbaceous plants; some are leafminers in early instars. At least 20 families of angiosperms are hosts, mostly dicots, with Cistaceae and Asteraceae prevalent in the Palaearctic; larvae of a few species eat lichens, mosses, grasses, or cacti. There are 700+ described species in 26 genera, with vast numbers awaiting study and naming (e.g., 90% of the North American fauna). Worldwide but most numerous in arid and seasonal drought areas such as the southwestern Nearctic, these moths are rare or lacking in wet tropical habitats.

Oecophoridae (s. str.): As now defined this is a worldwide assemblage of dissimilar moths, with a tremendous radiation of forms in Australia. Adults (Figs. 51 and 52) are small (FW 4–23 mm) with narrow to broad wings, mostly dull colored in Holarctic genera but wildly variable and colorful in Australia, where the FW are patterned in yellow, rose, rust, and browns. The abdominal terga are usually without spiniform setal bands. The larvae are cylindrical, with head often darkly sclerotized, sometimes with reduced numbers of stemmata and integument usually not pigmented; the thoracic legs and abdominal prolegs are well developed. Most feed on dead plant material, leaf litter, and other vegetative refuse, and the rich fauna in Australia depends mainly on *Eucalyptus* (Myrtaceae), with about 60% feeding on fallen leaves and 25% on living foliage. There are more than 3000 described species in 500+ genera worldwide; this is the dominant group in Australia, with 2200+ named species in 340 genera and a projected 35–40% of the Lepidoptera fauna. Several species are cosmopolitan household moths whose larvae feed in stored meal, potted plant humus, etc.

Coleophoridae are typically tiny to small (FW 3.5–13 mm), very slender moths with lanceolate wings (Fig. 53); the HW fringe is much wider than the wing. Usually these moths have rather long, nearly straight labial palpi that project forward, often slightly drooping. Paired patches of spiniform setae on the abdominal terga define their relationship with the Momphinae, which are tiny, stout moths with thick, diverging, upward-turned palpi. Mostly dull colored, yellowish, white, gray, or brownish, the FW often have linear, pale, or dark streaks. The larvae are slender, with very reduced abdominal prolegs in Coleophorinae, which are leafminers in the first instar and then live in a portable case constructed of silk covered with sand grains or pieces of plant material (Fig. 24); they feed by mining outward from the affixed case, forming characteristic, round mines with a central hole, as they move from spot to spot. Larvae of Momphinae are more stout, are grub-like, and feed within growing tips, stems, or galls they cause. Coleophorinae feed on more than 30 plant families, including conifers (rare); monocots, especially Juncaceae; and diverse dicots. Momphinae specialize on Onagraceae (70% of host records), the only Lepidopteran group to do so. More than 1100 species are described, with an estimated 500 unnamed species known in the Nearctic; they exist worldwide but are mainly Holarctic and are absent from Neotropical rainforests.

Blastobasidae: Adults (Fig. 54) are small (FW 4–15 mm), nocturnal moths with narrow wings and a short abdomen that bears a conspicuous, transverse row of stout, rust-colored or black, spiniform setae on each segment dorsally. The labial palpi are usually short, strongly curved upward, and appressed to the head. Blastobasids are uniformly dull colored, usually gray, having FW with whitish or black steaks, sometimes tan or yellowish. The larvae are slender, cylindrical, often with heavily pigmented integument, a labium with a submental pit, and SD1 setae usually with sclerotized rings. Larval foods—

The larvae are mainly scavengers, living in a wide variety of situations such as abandoned nests of insects, galleries of stem and root borers, and detritus associated with aphid and scale insect colonies, occasionally eating living insects, and a few species feed on living plant material. Owing to their consistently drab appearance and uniform genitalia structures, this family has been neglected in systematics studies. Worldwide, but more diverse in the Nearctic and Neotropical regions, there are 500+ described species and possibly 5 to 10 times that many awaiting study.

Cosmopterigidae: This is a diverse group not defined by any uniquely derived characteristic. The adults are tiny to small (FW 3–13 mm), slender with lanceolate or somewhat broad wings, lacking a gnathos, and having a strongly hooked aedeagus. The larvae are morphologically most similar to Gelechiidae. Habits are diverse, most of these insects are internal feeders, leafminers or bud, stem, bark, or root borers, sometimes causing gall formation by the host plant. Hence they tend to be stout with short legs, without secondary setae and little integumental pigmentation. More than 25 families of angiosperms are hosts. Typical cosmopterigines are leafminers, *Cosmopterix* often in monocots; others are seed feeders, and many are scavengers, in and under old bark, dead stems, etc., and in ferns and palms, especially on oceanic islands. Larvae of *Euclemensia* are predaceous on scale insects. Worldwide, there are 1650 described species in 100+ genera. The genus *Hyposmocoma* in Hawaii is the most famous example, with an estimated 450 mostly unnamed species that occupy diverse larval niches, including living plants, in dead wood or stems, on lichens, some feeding from a portable case, and in freshwater and littoral habitats, analogous to the Galapagos finches, but with a vastly more species-rich insular radiation.

Gelechiidae: One of the major families of micromoths, especially in temperate latitudes, the adults are most easily recognized by the hind wing shape, with the terminal margin indented below the acute apex. Adults (Fig. 55) are tiny to small (FW 3–12 mm, a few tropical species to 18 mm). The great majority are nocturnal, somber colored, brown, gray, or black, but some are colorfully patterned. The larvae usually form concealed shelters in new growing tips of trees and shrubs, but many are leafminers at least in early instars, or stem and root borers, and a few live in plant galls they cause (Fig. 27). Some feed in seeds or dead plant materials, while a broad diversity of living gymnosperms and angiosperms (80+ families) are used. A few feed on ferns or mosses, especially on oceanic islands. There are more than 4500 named species placed in 500+ genera and unknown numbers of undescribed species (e.g., estimated 60% of the North American species). These moths are most diverse in temperate zone areas, including deserts and other seasonally arid habitats. Several are important agricultural insects, including the pink bollworm *(Pectinophora gossypiella),* a threat to cotton growers worldwide; Angoumois grain moth *(Sitotroga ceralella),* which feeds in stored grains; potato tuber moth *(Phthorimaea operculella);* and many conifer needle miners.

APODITRYSIA All the more derived Lepidoptera are grouped in the Apoditrysia by possession of shortened apodemes on the second abdominal sternum that have enlarged bases, contrasted with the more ancestral state, continuations of longitudinal costae (venulae) of the sternal plate. Several superfamilies comprising the non-obtectomeran Apoditrysia retain the ancestral movable and spined abdominal segments in the pupa, which moves forward to protrude from the shelter, facilitating the moth's eclosion. The more derived superfamilies of this lineage, the Obtectomera (the pyraloids and macromoths), have nonmotile pupae.

Choreutoidea The family Choreutidae is a small group of phenotypically similar moths that have a scaled proboscis, an independent development from the Gelechiidae and Pyraloidea, and deposit upright eggs. The adults (Fig. 56) are small (FW 2–9 mm), diurnal, with broad wings, which they twitch in a characteristic fashion as they strut jerkily about on host plant leaves. The male antennae usually have long ventral setae. Choreutids are mostly dark colored, with black or brown wings marked by metallic gray, white, or silver-white; some tropical species are orange, with harlequin patterns. The larvae are slender with elongate abdominal prolegs, living externally in slight webs, from which they graze on leaf surfaces. Larval foods include diverse dicot angiosperms (17+ families), concentrating on Moraceae in tropical regions, Fabaceae, Urticaceae, and Asteraceae in the Holarctic. About 400 species are described worldwide, with many undescribed tropical species.

Sesioidea This superfamily consists of three families, Brachodidae, Sesiidae, and Castniidae, the adults of which are markedly dissimilar in appearance and behavior; their proposed relationship is based on subtle features: the eye is more strongly pigmented anteriorly, they have large patagia, and the larvae have an unusual crotchet arrangement, two transverse, uniordinal rows.

Sesiidae: These moths are wasp-like (Fig. 57), with FW basally narrow and relatively short HW, usually lacking scales except along the veins and distal margin. The wings are tightly coupled, with the posterior margin of the FW bent down, engaging with the upcurved costal margin of the HW, and both have rows of stiff scales that interlock, in addition to the normal frenulum and retinaculum. Sesiids are small to moderately large and heavy bodied (FW length 5–28 mm), diurnal or crepuscular, and almost all species resemble wasps or bees, often startlingly so. This involves not only clear wings and a colorful, banded abdomen, but the legs are modified with tufts, even to the extent of having yellow-tipped scales resembling pollen carried by bees. *Alcathoe* are black with bright orange wings, and the males have a long, slender, scaled process from the tip of the abdomen, which in flight resembles the trailing leg posture of tarantula hawks (Pompilidae, *Pepsis*). Sesiids often visit flowers with quick, darting flights. The larvae (Fig. 25) are borers in stems, bark, and roots; they are stout, with heavily sclerotized head and mandibles and unpigmented integument. Larval foods

include 40+ families of flowering plants, including conifers but not monocots. There are more than 1100 described species in 120 genera, speciose in both temperate and tropical regions. Sesiidae may be the most completely known of any microlepidoptera. Many species are agricultural pests, borers in stems of berry and squash vines, fruit trees, and conifer trunks.

Castniidae are primarily tropical, large (FW 24–190 mm), strong-flying, diurnal moths that have broad, often colorful wings, and the antennae are swollen distally, so they resemble butterflies (Fig. 58). The antennal tip is abruptly narrowed and bears a tuft of long hairs, features shared with Sesiidae. Some species are crepuscular and occasionally come to lights. Larvae are stout, cylindrical, with short legs and a head retractable into the thorax. They feed in plant stems or form tunnels in soil and feed on subterranean plant parts. The larval life requires 4.5 months to 2 years. All confirmed feeding records are for monocots, and several species are pests of sugarcane, banana, or oil palm. There are about 180 species, placed in 33 genera. The distribution suggests a Gondwanan origin, with the subfamily Taschininae in Southeast Asia, the sister group of Castniinae in Central and South America and Australia. Castniids occur in tropical, subtropical, and warm temperate regions and are lacking from most of the Holarctic, southern South America, Tasmania, and New Zealand.

Cossoidea Cossidae: Goat moths are small to large, hawk-moth-like, robust moths (Fig. 59) (FW 4–70 mm, rarely to 125 mm), having short, usually bipectinate antennae in the male; the proboscis is not functional, reduced to a small triangular lobe; labial palpi are three-segmented, short, and upturned. The wing venation is more primitive than in related superfamilies, with the median vein complete and branched within the cell. Cossids typically are nocturnal, drab, mostly gray with black striae; a few have brown or orange patches. The eggs usually are laid in groups in crevices or under bark and may be produced in vast numbers, 18,000 counted for one species in Australia. First instars disperse and bore into branches or trunks of living shrubs or trees, sometimes living gregariously; later instars are stout, cylindrical, with heavily sclerotized head and mandibles and unpigmented integument. Larvae require 1 to 4 years to mature. Recorded hosts include at least 17 families of angiosperms, including 1 monocot, with woody legumes (Fabaceae) accounting for 25% of the records. More than 20% of the species are polyphagous. Cossids occur worldwide, with greatest numbers in tropical regions (40% Neotropical), including more than 670 described species placed in 83 genera.

Tortricoidea Tortricidae: This is a large and relatively homogeneous family, with three subfamilies, Chlidanotinae, Tortricinae, and Olethreutinae, each of which, along with several subordinate taxa, at times have been treated as families. Females possess modified papillae anales, which have been rotated 90% from the ancestral lateral position to form flat, expanded pads facing ventrally, usually with the outline of shoe soles. Adults (Figs. 60–62) are small to moderately large (FW mostly 3–25 mm, to 28 mm in the Asian Ceracini), generally with rectangular FW and broader, plicate HW. The antennae are about 0.6 times the FW length, filiform, with sensory setae in males typically short, but long in some groups; the labial palpi are porrect or bent upward but not curved as in Gelechioidea. Most species are nocturnal, with the FW cryptically colored in gray, brown, rust, or tan, but some species have colorful markings. A few are spectacularly polymorphic, notably species of *Acleris* (Tortricini)—2 species in England have more than 100 named color forms. Because many tortricids are economically important as agricultural and forest pests, there is a vast literature on the biology, ecology, host-plant selection, oviposition behavior, pheromone chemistry, etc.—for example, more than 6000 references on the spruce budworms (*Choristoneura fumiferana* species complex) in North America. In all but the most ancestral tribes the eggs are flat, scale-like, and deposited singly in Chlidanotinae and Olethreutinae and the more ancestral tribes of Tortricinae, but derived Tortricinae deposit small to large, symmetrically shingled masses (100–150 eggs) (Fig. 9). Females of the Neotropical tribe Atteriini have thick mats of specialized (corethrogyne) scales of two types on sterna A6 and A7, which they deposit on and as upright fences around the egg masses. Two Australian Archipini and Epitymbiini fence the egg mass with scales from costal tufts on the HW. The larvae are cylindrical without secondary setae, with setal pattern and crotchet arrangements similar to those of Cossidae, usually with little or no integumental pigmentation other than the setal pinacula. Larvae of most Chlidanotinae and Olethreutinae feed as borers in stems (Fig. 26), roots, buds, or seeds, and most are specialists in host-plant selection. Larvae of a few species are miners in leaves (Fig. 22) or conifer needles or cause plant galls (Fig. 28). By contrast, nearly all Tortricinae are external feeders, often polyphagous, that form leaf rolls or other shelters in foliage, but species of Cochylini bore into buds and stems. Pupation usually occurs in the larval shelter or gallery, although some drop to the ground to pupate, especially those that diapause over winter as pupae or prepupal larvae in the Holarctic. External feeders of all three subfamilies possess an "anal fork," used to flip frass away from the larval shelter. An enormous array of plants serve as hosts. A few small tribes (few genera) are specialists, e.g., Bactrini on monocots. About 8500 described species are placed in 720+ genera, and incalculable numbers are unnamed in tropical regions—e.g., 70 to 80% of recently studied species of Neotropical tortricines have been previously unnamed. Rich faunas occur in all biogeographic regions. In addition to the spruce budworms, important economic Tortricinae include the light brown apple moth *(Epiphyas postvittana)* in Australia and fruittree leafroller *(Archips argyrospilus)* in North America, while Olethreutinae include the codling moth *(Cydia pomonella),* pea moth *(C. nigricana),* larch and spruce budworms *(Zieraphera),* several seed and cone borers *(Cydia),* and pine tip borers *(Rhyacionia)* in the Holarctic.

Zygaenoidea This is a conglomeration of families of very different appearing moths and larvae. All share two features,

larval head retractile into thorax and second abdominal spiracle of the pupa covered by wings. The mouthparts are vestigial in all the families treated here except Zygaenidae. The larvae are stout with the ventral surface slug-like, having short prolegs or suckers, often resembling caterpillars of lycaenid butterflies. Feeding habits vary, including predators of Homoptera or ants. Defensive secretions containing cyanoglucides are produced by some zygaenoids. Some families have been intensively studied because of their interesting larval habits, reproductive behavior, or chemical ecology. Taxonomists define 12 families, most of them small, with fewer than 50 species and limited to one geographical region.

Epipyropidae: Adults are small (FW 4–10 mm), FW triangular, HW round, blackish or gray, rarely with white or orange HW. The antennae are bipectinate in both sexes, more broadly in the male. The moths are crepuscular or nocturnal but rarely come to lights. Females produce large numbers of eggs, up to 3000 in one African species, that are deposited on foliage of a plant frequented by the planthopper hosts. The larvae are hypermetamorphic, with first instar triungulin-like, slender, tapering posteriorly with long thoracic legs and ventral ambulatory setae. The first instar seeks the host, attaches by means of silk, and transforms to a grub-like larva, which may feed on secretions produced by the homopteran, and the host remains active. The larva is covered by wax secreted by glands in its integument. Pupation occurs away from the host in a dense cocoon impregnated with wax. Epipyropids feed on body fluids and secretions of fulgoroid leafhoppers of several families. There are about 40 described species assigned to nine genera, occurring in pantropical and warm temperate regions.

The Australian family Cyclotornidae (12 species) are similar to the Epipyropidae with an even more bizarre life cycle. Females deposit large numbers of eggs (1400 counted for one female) on vegetation infested with cicadellid leaf hoppers or Psyllidae (Homoptera). The active first instars follow *Iridomyrmex* ant trails leading to the homopterans. After feeding, the moth larva leaves and molts into a brightly colored, flat, broadened, scale-like larva, which curls its abdomen upward to expose the anus where a secretion is produced that is much sought after by the ants. The cyclotornid is seized by an ant and carried back to the nest, where it feeds on ant larvae and pupae.

Limacodidae: Adults (Fig. 65) are small to medium-sized (FW 6–35 mm), mostly nocturnal moths with a stout body, relatively short, broad wings, and densely scaled head and body. The antennae are bipectinate in male, often to about half the length of the body. Mostly brightly colored, in yellow, tan, browns, these moths assume a characteristic resting posture, with the body held at an acute angle to the substrate. Larvae are hypermetamorphic, with first instar, which often is nonfeeding, oval, flat, and bearing rows of large spines that are not homologous with basic setal patterns of other Lepidoptera. Later instars (Fig. 66) are lycaenid-like in form, lack abdominal prolegs, and often have ventral suckers on segments A1 to A7 that adhere to the foliage, aided by a fluid secreted over the cuticle, and they move slug-like by peristaltic waves passing along the sole-like venter. Different species exhibit a great variety of body form, smooth, sometimes with gelatinous warts, with protuberances, spines, and hairs, some of which are stinging (nettle caterpillars), or densely hairy. Later instars often are green, but in stinging species they are brightly colored. The fecal pellets are characteristically cup-shaped. The cocoon, which usually is hard, incorporating calcium oxalate that is secreted by the prepupal larva, has a circular, dehiscent lid. A variety of dicot trees and shrubs serve as larval hosts, with many species polyphagous, and some feed on monocots, including coconut, banana, rice, sugarcane, and other economically important plants. There are more than 1000 described species, occurring in all zoogeographical regions, with greatest numbers in the tropics.

The sister family to the Limacodidae, the Dalceridae (40 species), is limited to the Neotropical region. Adults (Fig. 67) are small moths (FW 6.5–24 mm), white, yellow, or orange, with broad, rounded wings and hairy bodies, similar to epipyropids and limacodids in having vestigial proboscis and small labial palpi; the antennae are bipectinate, more broadly in males. The larva (Fig. 68) has a brush-like spinneret, dorsum and sides with gelatinous humps secreted from glandular setae, venter with translucent cuticle, thoracic legs short, prolegs not developed, and crotchets absent in early instars, appearing in the final two.

Megalopygidae: Adults (Fig. 69) are also similar to Limacodidae but generally larger (FW 11–28 mm) and relatively heavy bodied, with FW markings often brightly colored and elaborately patterned and scales hair-like, cleft, or tripartite. The larvae (Fig. 70) are superficially similar to some limacodids, but with three rows of spined protuberances (scoli) bearing variable development of typically urticating setae, sometimes beneath tufts of long, silky hairs like a fur coat (puss caterpillars), with the unusual complement of seven pairs of prolegs, those of A2 and A7 lacking crotchets. Pupation occurs in a tough, tapered cocoon blended into a branch. There are about 260 species, primarily Neotropical, with several genera ranging into the United States. The small subfamilies Somabrachyinae, in the Mediterranean region and Africa, and the Neotropical Aidinae are sometimes separated as families.

Zygaenidae comprise a worldwide group of dissimilar moths exhibiting a large number of remarkable specializations. Nearly all are diurnal and many are slow-flying and brightly colored (aposematic) (Fig. 71) and are avoided by birds and other predators because both adults and larvae biosynthesize cyanoglucides. They are able to release hydrocyanic acid by enzymatic breakdown of the cyanoglucides. This has enabled many to take part in complex mimicry relationships. Moreover, adults and larvae are unusually resistant to cyanide, and naive collectors are startled to see a moth survive for half an hour in a potent vial that kills other moths in seconds. Adults are small to large (FW 5–50+ mm), with head and labial palpi smooth scaled, and they commonly visit flowers for nectar.

FIGURES 65–92 Adults and larvae of zygaenoid, pyraloid, and macro moths. Zygaenoidea: (65) *Parasa indeterminata* (Limacodidae) (New Jersey); (66) Larva of *Isa textula* (Limacodidae) (Maryland); (67) *Dalcerides ingenita* (Dalceridae) (Arizona); (68) larva of *D. ingenita* (Dalceridae) (Arizona); (69) *Trosia revocans* (Megalopgyidae) (Amazonas); (70) larva of *Monoleuca semifascia* (Megalopygidae) (New Jersey); (71) *Zygaena ephialtes* (Zygaenidae), pair in copulo (France). Pyraloidea: (72) *Petrophila confusalis* (Crambidae, Nymphulinae) (California); (73) Pyraustinae species (Crambidae) (Costa Rica); *Crambus* species (Crambidae) nectaring at composite flower, whereas the larval host is a grass (Arizona). Geometroidea: (75) *Urania fulgens* (Uraniidae) (Ecuador); (76) *Dichorda illustraria* (Geometridae) (California); (77) *Neoterpes edwardsata* (Geometridae) (California); (78) flightless female of *Tescalsia giulianiata* (Geometridae), a winter moth (California). Bombycoidea: (79) *Phyllodesma* species (Lasiocampidae) (California); (80) *Bombyx mori* (Bombycidae), pair in copulo of the commercial silk moth; (81) *Eacles* species (Saturniidae, Citheroniinae) (Costa Rica); (82) *Hemileuca eglanterina* (Saturniidae, Hemileucinae), a diurnal species, female ovipositing (California); (83) *Argema maenas* (Saturniidae, Saturniinae) (Malaysia); (84) *Smerinthus cerisyi* (Sphingidae) in predator avoidance posture (Utah); (85) *Hemaris senta* (Sphingidae), a diurnal bumble bee mimic (California). Noctuoidea: (86) *Clostera apicalis* (Notodontidae) (California); (87) *Phryganidea californica* (Notodontidae, Dioptinae); (88) larva of *D. californica,* the California oak moth; (89) *Lymantria dispar* (Lymantriidae), mating pair of the notorious gypsy moth (Russia); (90) *Orgyia vetusta* (Lymantriidae), wingless female tussock moth (California); (91) larvae of *O. vetusta;* (92) *Horama panthelon* (Arctiidae) (Texas). (Photographs by: E. Buckner, 71; R. Cardé, 89; C. Covell, 75, 83, R. Coville, 72, 73, 76, 77, 79, 80, 81, 86, 87, 88, 90; J. Hafernik, 74; C. Hanson, 67; L. Penland, 65, 68; J. Powell, 78, 84, 91; D. Rubinoff, 82, 85, 92; J. Ruffin, 66, 70; K. Sandved, 69).

The antennae are often thick and either clubbed in both sexes or bipectinate in males and narrowly bipectinate or filiform in females. Glands located between the eyes and the base of the proboscis produce a whitish or yellow liquid or foam when the moth is disturbed. They are often metallic colored, particularly vivid in European *Zygaena,* in bright red and metallic blues. The eggs are deposited in rows, clusters, or overlapping patches and sometimes covered with scales from a special abdominal hair tuft, which in some Australian genera are urticating. The larvae are stout and broad, with the head usually retractile under the extended prothorax; the body is roughened and covered with dense secondary setae. Larval foods include numerous plant families, although the species are mostly host specific. European zygaenines specialize on Celastraceae, cyanogenic Fabaceae, and noncyanogenic Apiaceae; some Australian Procridinae rely on Dilleniaceae, Myrtaceae, or Vitaceae, and the last is used by some Nearctic species. About 1000 species are described, and the family is relatively well studied owing to the colorful forms, diurnal habits, and mimicry associated with the chemical ecology.

Alucitoidea Alucitidae: The many-plume moths are so called because the wings are deeply cleft, so each wing has six fringed, plume-like segments (Fig. 63). The family was classified with the true plume moths (Pterophoridae) from which the alucitids differ by having discrete bands of spines on some or all of abdominal terga 2 to 7 and a relatively unspecialized pupa formed in a cocoon. Adults are small (FW 3–13 mm), slender moths that are unmistakable by their wing structure, but in two tropical genera with the largest members of the family, the wings are divided only a short distance. The antennae are filiform, proboscis is well developed, and labial palpi are usually fairly long, porrect or upcurved. Most are gray or brown, delicately banded with tan or white. They are nocturnal and collapse the wing plumes when at rest, holding them out from the body so as to resemble narrow-winged pyralids, but when active they strut about with the wings fully expanded, like miniature peacocks. The larvae are borers in flower buds, shoots, or in galls; the body is stout with short legs with short setae on inconspicuous bases. Pupation occurs in the larval shelter or in leaf litter in a cocoon. Larval hosts include at least eight dicot families, especially Caprifoliaceae in the Holarctic and Bignoniaceae and Rubiaceae in Australia and tropical regions; one species is a coffee pest in Africa. There are about 130 species described and likely there are many more in tropical regions.

Pterophoroidea Pterophoridae: The plume moths are recognizable by their deeply cleft wings in all but the most ancestral genera. They lack proboscis scaling and abdominal tympana, the hind tibia is 2 or more times the length of the femur, and abdominal terga 2 and 3 are elongated. Adults (Fig. 64) are small (FW 4–18 mm), with long and slender bodies, legs, and wings; FW is cleft for about 0.25 to 0.33 its length and the HW deeply twice-cleft in most species. When at rest, the plumes are overlaid and rolled under the leading edge of the FW, resembling sticks held out from the body. They usually are nocturnal and dull colored, tan, brown, or gray with paler and darker markings; although a few are colorful members of tropical mimicry complexes. The larvae (Fig. 29) typically are elongate and cylindrical with long prolegs. Most are external feeders on foliage and usually have dense setae that may be forked, clubbed, or glandular, secreting a sticky fluid. Some species are borers in stems and have short setae, even a strongly sclerotized anal shield with two stout horn-like processes resembling the urogomphi of coleopteran wood borers. The pupae are strongly spined or setose and either are formed in the galleries or are affixed to host-plant stems or debris by the anal cremaster to a silk pad, fully exposed, and they can bend and curl over by their movable abdomen. More than 20 dicot families are recorded as larval hosts, principally Asteraceae, usually herbs, but not monocots. The larvae of *Buckleria* in Europe are remarkable for feeding on sundews (Droseraceae). About 1000 species are described worldwide, with many unnamed tropical species.

APODITRYSIA: OBTECTOMERA

Pyraloidea This is one of the largest superfamilies of Lepidoptera, with more than 17,500 species described and probably at least as many more from tropical regions awaiting study. The fundamental features that define the pyraloids are a basally scaled proboscis, well-developed maxillary palpi, and tympana consisting of paired chambers on the venter of abdominal segment 2. In recent decades specialists have agreed that differences in the morphology of the tympana and other adult and larval characters warrant treating the former Pyralidae as two families, the Crambidae and Pyralidae. While it is difficult to recognize crambids and pyralids as distinct groups on the basis of superficial appearance owing to the enormous variability within each, subfamilies of each family are distinctive. Pyraloids occur worldwide, other than Antarctica, and range from high alpine to low desert and tropical habitats but are most prevalent at low and middle elevations in the tropics. They are highly successful at dispersal and colonization and are especially well represented on oceanic islands.

Pyralidae: These are small to relatively large moths (FW 5–75 mm, mostly under 30 mm) that have the tympanal organs almost completely closed, with their conjunctiva and tympanum in the same plane; they have vein R5 of the FW stalked or fused with R3+R4 and lateral "arms" at the base of the uncus in the male genitalia; the larvae almost always have a sclerotized ring around the base of seta SD1 on abdominal segment 8 and often around SD1 of the metathorax. The larvae usually are stout and cylindrical, with relatively short legs and setae; the body typically is unpigmented, although some species are well patterned, even brightly colored. Almost all are concealed feeders, most often borers in seed, fruit, or stems or live in tunnels they construct in the soil beneath plants. Many others construct shelters among tied leaves, often of quite tough silk. A variety of flowering plants are hosts, as well as wood-rot fungi (Xylariaceae), dry vegetable

matter including seeds, and the papery structure of social Hymenoptera nests (*Galleria, Achroia, Aphomia,* Galleriinae). Many are household and granary pests that have been transported worldwide by human activities (*Corcyra,* Galleriinae; *Pyralis,* Pyralinae; *Plodia, Ephestia, Anagasta, Ectomyelois,* Phycitinae). A few species are predaceous on scale insects (*Laetilia,* Phycitinae) or live in ant nests (some Chrysauginae); three genera of Chrysauginae feed in the dung of sloths, and the adults live in their fur. Several species of Phycitinae have been used for biological control of cactus. The majority of species feed on flowering plants, including conifers (*Dioryctria,* Phycitinae); and monocots, including pests of coconut and other palms (*Tirathaba,* Galleriinae) and corn (Epipaschiinae), and a wide range of forest trees, ornamental shrubs, and crops are damaged, especially in tropical regions. There are more than 6000 described species, about two-thirds of which are Phycitinae, and the tropical faunas are not thoroughly studied.

Crambidae: Crambids are more diverse and variable in morphology and biology than pyralids, but all share "open" tympanal organs (i.e., with a wide anterior aperture, and the conjectiva and tympanum meet at a distinct angle); vein R5 of the FW usually is not stalked with R3+4, male genitalia are without basal uncus "arms," and A8 of the larva lacks sclerotized rings around seta SD1. About 14 subfamilies are defined, and usually members of each can be recognized on superficial bases. Adults (Figs. 72–74) vary from small and slender to large and stout bodied (FW 3.5–47 mm), with FW narrow and HW plicate and folded under it at rest (e.g., Crambinae, Fig. 74), to broadly triangular with HW similarly colored and held flat (e.g., Pyraustinae, Fig. 73). The labial palpus is prominent, obliquely ascending or porrect (typically very long in Crambinae, snout moths), maxillary palpi are typically small, often with broadened scaling that connects a profile of frons with the labial palpus. Proboscis is usually well developed. Legs are long, and those of males often have structural modifications and/or androconial scale tufts. Larval food habits and morphology vary among subfamilies: Crambinae (lawn moths) live either as ground dwellers feeding primarily on grasses or as stem borers in various monocots; Schoenobiinae are borers in marsh grasses; Cybalomiinae and Evergestiinae specialize on Brassicaceae and Capparidaceae; Midliinae are borers in Araceae; Musotiminae larvae feed on ferns, a rare niche for Lepidoptera; many Odontiinae are leafminers or flower, bud, and stem borers, on diverse dicots; and Pyraustinae, which make up 65% of the known species of Crambidae, are mostly webworms but some borers, in an enormous variety of monocots and dicots, pest species of many crops, including corn, bananas, palms, pasture grasses, cucurbits, tomatoes and other solanaceous fruits, coffee and other tropical trees, garden mints, and conifers. Scopariinae are specialists on mosses and lycopods, tunneling in the roots and stems, and on ferns, or on seed-bearing vascular plants. Nymphulinae larvae are aquatic, living either in ponds on vascular plants, often in cases, or in rapid streams, usually under webs on rocks and feeding on algae. Larvae either breathe through open spiracles, living in air-filled cases or stems, or absorb dissolved oxygen through tracheal gills. Pupae are formed in cases in chambers within the plants or in gas-permeable cocoons and breathe through the spiracles. In the Palaearctic *Acentria* (Schoenobiinae), females are wingless, enter the water, and are parthenogenetic until a bisexual generation of winged adults late in the season. More than 11,500 species are described, nearly 90% of which are members of Nymphulinae, Crambinae, or Pyraustinae.

Geometroidea This group includes five families, Drepanidae, Epicopeiidae, Sematuridae, Uraniidae, and Geometridae, although they have been separated as two or three superfamilies by some authors, based primarily on the structure of the larval mandibles and tympanal organs. Geometroids typically are broad-winged, with slender bodies, small to large moths (FW 5–78 mm); they have abdominal tympana of structures different from those of pyraloids and lack scaling of the proboscis. Larvae of Uraniidae have well-developed abdominal prolegs, while most Drepanidae and Geometridae have some of the prolegs vestigial or absent.

Drepanidae: Adults have internal abdominal tympana unlike any other Lepidoptera, associated with the dorsal–ventral sclerites that connect tergum 1 with sternum 2, opening dorsally. Adults are medium sized to large (FW 8–31 mm), broad winged and geometrid-like in typical Drepaninae, often with the FW apex produced or curved; Thyatirinae are stout bodied and noctuoid-like. The antennae usually are short, lamellate or bipectinate to the tip for most of the length, sometimes filiform. The larvae have few secondary setae or rarely numerous but very short setae, sometimes with an eversible vesicle just above the prothoracic coxa in Drepaninae, often with notodontid-like protuberances in Thyatirinae; anal prolegs are usually vestigial but those of A3 to A6 are well developed, and the anal shield is conspicuously elongated. At least 20 families of diverse dicots and one monocot (Zingiberaceae) are hosts; some Holarctic species are generalists. More than 650 species are described in 120+ genera, mostly in the Holarctic and Oriental regions; Drepaninae are absent in the Neotropics. A few species are pests of coffee.

Uraniidae: This family is defined by the sexual dimorphism of the tympanal organs, which are on the lateral, posterior part of tergum A2 in males and on the lateral part of sternum A2 in females. Adults (Fig. 75) are small to large (FW 7–78 mm), broad winged, usually with a relatively slender body; wings are often resplendent in brilliant, iridescent colors; HW veins are sometimes produced into one or several swallowtail butterfly-like tails. Most are nocturnal, but 3 tropical genera are day flying, including *Urania,* some species of which are famous for their massive migratory flights, involving thousands of the spectacularly colored moths. Epipleminae are smaller, nocturnal, and dull colored; they rest with the FW extended and rolled, HW appressed to the body. The antennae are filiform,

lamellate, or pectinate, sometimes thickened preapically. The larvae are more or less bare, with few secondary setae, occasionally with spatulate setae, and prolegs are well developed. Larval foods are recorded in about a dozen dicot families, including specializations on Oleaceae, Asclepiadaceae, and Euphorbiaceae (all known larvae of Uraniinae feed on euphorbs), unusual for Lepidoptera and likely sources of distasteful qualities, and many species appear to be aposematic. Worldwide, these moths are primarily pantropical; around 700 described species are assigned to 90 genera.

Sematuridae are tropical, similar superficially to Uraniinae, often brightly colored with HW tails, but adults lack abdominal tympana, and the antennae usually are distally thickened, like those of skipper butterflies.

Geometridae: This is one of the three most speciose families of Lepidoptera. Adults (Figs. 76–78) are small to large (FW 5–55 mm), typically with broad wings and slender body and abdomen with basal tympanal organs in deep, ventrolateral cavities. The great majority of geometrids are nocturnal and rest by day with the wings outspread, with cryptic resemblance to tree bark, lichens, green or fallen, brown leaves, often ornate with lines simulating leaf veins or spots resembling necrotic or eaten areas of foliage. Some genera hold the wings upright, butterfly-style, and their undersides are cryptically colored. Some geometrids are day flying, including early spring species in the Holarctic or mimetic species in tropical regions. A few species that are winter moths or occur at high elevations have flightless females (Fig. 78), and these have the mouthparts and tympana reduced or vestigial. The larvae usually have prolegs of segments A3 to A5 reduced or absent (Fig. 13), so the caterpillar walks by advancing in measured, looping steps, from which both the family name and the common name (inchworms) are derived. They are bare but with a variety of colors, protuberances, ornate body forms, and behavior; most are exposed feeders that depend upon cryptic resemblance to flowers, leaves, twigs, etc., to avoid predators (Fig. 30). Most species are general feeders on trees or shrubs. Owing to their worldwide species richness, an enormous variety of gymnosperms and angiosperms are eaten. At least 21,000 species and 1500 genera have been described, and no doubt many remain unnamed, especially in tropical regions. Several species are defoliators of hardwoods or conifers (e.g., spring and fall cankerworms, *Paleacrita vernata* and *Alsophila pometaria,* and the hemlock looper, *Lambdina fiscellaria,* in North America), but geometrids are not a major pest group.

Bombycoidea These are macromoths that have no thoracic or abdominal tympana. The group is distinguished by deep clefts between the prescutum and the mesoscutum of the mesothorax. There are 12 families, including 4 worldwide groups summarized here that are sometimes treated as superfamilies: Lasiocampidae, Bombycidae, Saturniidae, and Sphingidae. Other bombycoids are Mimallonidae, a mostly Neotropical family with 200 species; Anthelidae, a small Indo-Australian group; Eupterotidae, worldwide with 300 species; Endromidae, Mirinidae (Palaearctic), and Carthaeidae (Australian), each with 1 or 2 species; and 2 small Eurasian and African families related to Sphingidae, Lemoniidae and Brahmaeidae.

Lasiocampidae: Adults (Fig. 79) are small to large (FW 9–80 mm) with moderately broad wings and stout bodies. The eyes usually have fine hairs between the facets; mouthparts are absent or vestigial; antennae are bipectinate to the tip in both sexes; labial palpi are porrect, the first segment with a scaleless patch bearing sensory setae, unique in Lepidoptera. The wings have interlocking tiny setae but no retinaculum and frenulum. Lasiocampids are nocturnal, but males of a few species are primarily diurnal, mostly somber-colored moths, browns and tan. The female often is much larger than the male, with a stout abdomen having a fully developed complement of eggs upon emergence, sometimes deposited in a single large mass. The larvae do not have the fore coxae fused as in other bombycoids. The body is covered with dense, often long, unbranched secondary setae and frequently is brightly colored. Larvae of some genera are gregarious (e.g., Holarctic *Malacosoma,* tent caterpillars, Fig. 33), and some have urticating hairs that produce a skin rash in humans. Pupation occurs in a dense, parchment-like cocoon. Larval foods are diverse, mainly trees and woody shrubs, Betulaceae, Fagaceae, and Salicaceae in the Holarctic; low-growing Asteraceae, Brassicaceae, and grasses in Africa; and predominately Myrtaceae in Australia but also mistletoes (Santalaceae). Some species are polyphagous. There are about 1500 named species, in 150 genera, occurring worldwide, mostly in the tropics; they are absent from New Zealand.

Bombycidae: Including Apatelodidae of recent authors, this family is defined by a few subtle skeletal features and by short pupal galeae that fail to reach the foreleg apices. Adults are small to medium sized (FW 9–28 mm), nocturnal with broad wings, which usually are held outward from the body when at rest. The mouthparts are vestigial. The antennae are bipectinate nearly to the tip in both sexes, and the thorax and legs are clothed in long, hairlike scales. Larvae are similar to lasiocampids, but the secondary setae are either short and minute or long in Apatelodinae. The fore coxae are separate (Apatelodinae) or fused. Segment A8 has a middorsal scolus except in Apatelodinae. Larval foods are largely Bignoniacease, Symplocacaceae, Moraceae, and Theaceae—some species are reported to defoliate commercial tea. There are about 350 species referred to 40 genera, worldwide except in Europe, best represented in the Oriental and Neotropical regions. The silk moth (*Bombyx mori,* Fig. 80) is the most well known species; it has been domesticated for centuries and is not known in the wild, its adults no longer capable of flight.

Saturniidae: The emperor or giant silk moths include many of the world's most spectacular moths. They are medium sized to very large (FW 14–130+ mm), broad winged with highly variable color patterns (Figs. 81–83), often with eyespots with concentric rings (hence the family name), which presumably act as a defense mechanism against predators of the sluggish moths, many of which must warm themselves by pumping the

abdomen before they can fly. Most genera are nocturnal, some mating only during early morning hours, while species of a few genera are diurnal (e.g., *Hemileuca* in North America, buck moths, Fig. 82). The proboscis and maxillary palpi are rudimentary or absent. The male antennae are bi- or quadripectinate, those of the female filiform to quadripectinate. The male FW usually is produced apically, while the hind margin of the HW sometimes has short to very long tails. The larvae are stout with a prominent, smooth head and body protuberances that often are branched; secondary setae are numerous but small, mostly on the ventral half of body and prolegs. The scoli of some genera bear poisonous spines that cause nettle-like stings in humans, and females of some species coat the eggs with urticating hairs that cause dermatitis. Abdominal segment 8 usually has a middorsal horn or scolus. Pupation occurs in a strong, sometimes double-walled silken cocoon that may be covered with plant fragments or in the soil without a cocoon. Many saturniids are polyphagous, and an enormous array of plant hosts are recorded—90 genera in 48 plant families recorded for one species of *Attacus*—but some are specialized feeders. There are about 1500 described species in 165 genera, occurring worldwide except at the highest latitudes, most abundant in moderate to high-elevation habitats, richest in the Neotropical region, particularly the South American Andes.

Sphingidae: Sphinx or hawk moths are among the largest, most easily recognized, and best known Lepidoptera. Adults (Figs. 84 and 85) are medium sized to very large (FW 16–90 mm), having a stout body with the abdomen typically tapering posteriorly. The FW is narrow and HW relatively short, its hind margin produced, angulate at the tip of veins 1A + 2A, emarginate beyond. The antennae are distinctive, usually lamellate ventrally or bi- to quadripectinate, tapering toward the apex, which is upturned or hooked; males with two rows of long sensillae that meet dorsally; shorter and filiform in females. The proboscis usually is well developed, sometimes much longer than the body, and used to imbibe nectar while hovering in flight, hummingbird-style, and in some habitats sphingids have significant roles in pollination. Most are nocturnal, extremely strong fliers, among the fastest insects, and several are well-known long-distance migrants. Some genera are diurnal, a few resembling bumble bees, with mostly transparent wings (e.g., *Hemaris,* Fig. 85); such species have fully scaled wings upon eclosion, but after drying, the scales are shed, all but along the margin and veins. Wing coupling is usually by frenulum–retinaculum, a long bristle in males, multiple setae in females. The larvae have a prominent, triangular or globose head; the body is covered densely with minute secondary setae and usually no other setae or protuberances except a middorsal horn or button on A8. The lateral markings are distinctive, each abdominal segment with an oblique stripe ascending posteriorly, those of A7 reaching the base of horn on A8. Color is often polymorphic; they usually feed completely exposed and rely upon cryptic coloration for protection. A distinctive characteristic is their resting pose, with the thorax raised and head turned down, resembling the pose of an Egyptian sphinx. Pupation usually occurs without a cocoon, in soil or ground litter, but rarely in a silken cocoon. Sphingids feed on a very broad range of gymnosperms and angiosperms, often specialists on plants with chemical defenses that repel most insects, including Apocynaceae, Cleaceae, Solanaceae, Rubiaceae, and Violaceae. There are more than 1200 described species in about 200 genera, distributed worldwide, best represented in the tropics. It is probably better inventoried and cataloged than any other moth family.

Noctuoidea This is the largest superfamily by far, with more than 7200 genera proposed for nearly 60,000 species, about 40% of all described Lepidoptera. As such, there is a tremendous variety of form, size, color, morphology, and behavior in both larvae and adults. Four major, cosmopolitan families are recognized, Notodontidae, Lymantriidae, Arctiidae, and Noctuidae, from which several regional lineages have been split as families by some authors. Monophyly of Noctuoidea is unequivocally based on complex metathoracic tympanal organs and associated abdominal structures. The tympana are assumed to have evolved in response to bat echolocation, and many species have been observed to engage in bat avoidance behavior upon receipt of their sounds. Tympana may also receive mating signals, especially in Arctiidae.

Notodontidae: With Thaumetopoeidae and Dioptidae included as subfamilies, this is a large and diverse family. Adults (Figs. 86 and 87) are medium sized to large (FW 16–50+ mm), typically with relatively long FW and stout body that extends 2 or more times the width of the HW. The head often has scale tufts or crests; antennae are usually bipectinate to the tip in the male, filiform or sometimes bipectinate in the female. Proboscis is usually well developed and coiled; labial palpi are often quite short. The abdomen is densely covered with long, slender scales and sometimes dorsal scale tufts at the base. The tips of the tibial spurs are serrated. These are mostly dull-colored, tan, brown, or gray moths, but many tropical dioptines are diurnal and brightly colored, involved in mimicry complexes. The larval body is stout, nearly bare, sometimes with long secondary setae, often possessing one or more protuberances, a modified body form (Fig. 31), a median knob or horn on A9, or anal prolegs modified into slender, single or double caudal processes (stenopods). All larvae except Dioptinae have two MD setae above the spiracle on abdominal segments, whereas other noctuoids have only one. Late instars have a smooth mandibular cutting edge, derived from the serrate ancestral state in other noctuoids. When disturbed, some species emit formic acid or ketones from a cervical gland (adenosoma). Many have various cryptic colors correlated with modified body forms, but others are brightly colored and aposematic, especially Dioptinae (Fig. 88). Larval foods include a wide diversity of dicot angiosperms, mainly woody shrubs and trees, and a few feed on grasses. Many specialize on plants containing toxic substances, including Anacardiaceae, Apocynaceae, Aristolochiaceae, Fabaceae, Passifloraceae, and Violaceae. There are more than 2800 described species,

distributed worldwide except in New Zealand and the Pacific Islands and particularly rich in the Neotropics. A few are occassional defoliators of orchard or forest trees.

Lymantriidae: The tussock moths are so called because larvae of many have thick tufts of erect secondary setae on the dorsum, like those of a toothbrush (Fig. 91). Adults are small to moderately large (FW 7–45 mm), nocturnal, usually dull-colored moths, mostly brown or yellow; they are usually broad winged and densely hairy; the FW is triagnulate, HW rounded, hidden under the FW at rest, with the wings appressed to the substrate and the densely hairy forelegs extended in front of the head. Females of some genera are flightless, their wings vestigial (Fig. 90). Antennae of males and usually females are bipectinate to tip. The proboscis is vestigial or absent. Females have a large, abdominal tuft of deciduous scales used to cover the egg masses. Males of many genera possess a tymbal organ, a pair of finely corrugated pockets on A3. The larval integument often is brightly colored, with hair tufts from verrucae, but the body form is not strongly modified as in notodontids. Lymantriid larvae all possess a mid-dorsal, eversible gland, often yellow or red, on segment A6 and usually another on A7. The larval hairs of *Euproctis* in Australia and some other genera are hollow, barbed, and urticating and cause a severe skin rash in humans. The body hairs are woven into the cocoons and often retrieved by the emerging female and redeposited on the egg masses and then used by first instars, which feed in protected aggregations. Many lymantriids are polyphagous, frequently on arborescent shrubs or trees, and a broad array of plants is eaten. Appreciable generic radiations feed on flowers and fruit of low herbs and grasses, and many Asian and African species feed on algae, fungi, and detritus. There are more than 2500 species placed in 360 genera, distributed in all geographic regions, reaching their greatest development in the Old World tropics. Many species in several genera are forest defoliators in Europe, North America, and Indo-Australian tropics, the most notorious being the Palaearctic gypsy moth (*Lymantria dispar,* Fig. 89), which was introduced into North America in the late 1800s.

Arctiidae: The tiger moth family includes Ctenuchidae and Pericopidae, formerly treated as families, which now are interpreted as artificial groupings. There are three subfamilies, Lithosiinae, Syntominae, and Arctiinae, all characterized by a pair of dorsal, eversible, single or branched pheromone glands from the terminal abdominal segment in the females. In addition, many arctiines and lithosiines and some syntomines have metathoracic tymbal organs in both sexes. Sound is produced by contraction of muscles that deform ridges on the tymbals rapidly to produce bursts of ultrasonic clicks, stimulated by tactile cues or in response to hunting signals of bats. Many species of all three subfamilies have prothoracic glands from which a liquid is extruded containing acetylcholines and histamines and probably pyrozines, the odor of which is believed to signal distasteful or toxic properties to predators. Adults (Figs. 92 and 93) are small to moderately large (FW 5–50 mm), usually brightly colored moths with a myriad of patterns, predominately orange, red, and black, often aposematic, advertising their toxic qualities and involved in mimicry complexes, or they resemble wasps, with mostly scaleless wings. Females of some species have a large tuft of abdominal scales used to mix with or cover egg masses. Most arctiids are nocturnal and come to lights, even many tropical species with elaborate Hymenoptera resemblances. Some genera are strictly diurnal, such as North American *Ctenucha, Gnophaela,* and *Lycomorpha* that accompany aggregations of lycid beetles they mimic. Larvae of many arctiids have dense secondary setae over the body (woolly bears, Fig. 94), but setae are sparse in some arctiids, especially Lithosiinae. The body form is typically cylindrical with a full complement of prolegs; most are rapid crawlers and can move great distances in search of food or pupation sites. Lithosiinae possess an enlarged basal molar area of the mandibles, used to macerate algae and lichens, which are the principal foods (rarely liverworts and mosses). Some Syntominae also feed on algae and lichens, many are scavengers or fungivores, and some feed on flowers, especially Asteraceae, or grasses. Arctiinae are polyphagous plant feeders or specialize on one of a variety of angiosperms, some genera on latex-producing plants (Apocynaceae, Euphorbiaceae, Moraceae) or those with toxins (Asteraceae, Boraginaceae). More than 6000 species are described in 750+ genera, occurring worldwide, particularly rich in tropical regions. Tropical lithosiines are poorly studied, and there are numerous undescribed species.

Noctuidae: This is the largest family of Lepidoptera, with more than 35,000 species grouped into about 30 subfamilies, many of which have been regarded as families, including Pantheinae and Nolinae, by recent authors. The composition and classification of most of these subfamilies is debatable, with critical larval characteristics unknown for most genera. Members of this vast lineage do not share a distinguishing derived feature; they posses the wing venation and tympanal form of advanced noctuoids, retain a well-developed proboscis, and lack the eversible abdominal gland present in larval Lymantriidae and the eversible pheromone glands of arctiid females. Adults (Figs. 95–97) are small to very large (FW 4–140 mm), including *Thysania agrippina* in Central America, with the world's largest insect wing expanse, exceeding 10 in. The FW is triangular to narrow and the HW broad, usually folding fan-like under the FW when at rest. The body is relatively heavy, the thorax having powerful musculature in most noctuids, and many of these moths migrate long distances. The head has long scales, sometimes forming erect tufts or a conical projection in front; the proboscis is long and coiled, its tip armed with thorns for piercing fruit in some genera. Wing coupling is accomplished by frenulum–retinaculum, usually a single bristle in males, two or three finer bristles in females. The abdomen often has eversible coremata or tufts or pouches of specialized scales. Most noctuids are nocturnal and somber colored, with the FW cryptic against bark or leaf litter when the moths are at rest during the daytime. Some have brightly colored HW that are flashed when the insect is disturbed, presumably

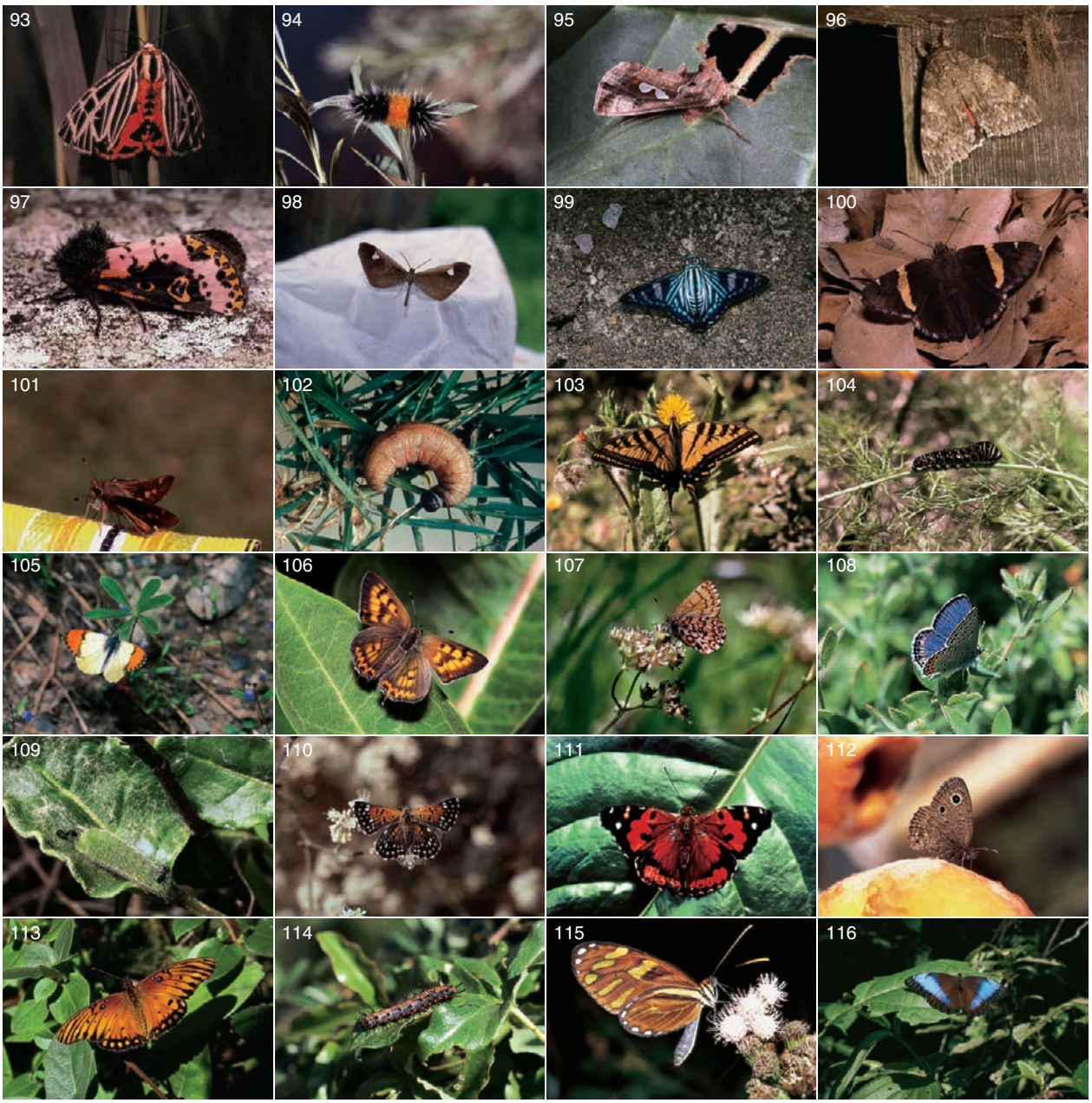

FIGURES 93–116 Adults and larvae of Noctuoidea, Hedyloidea, and butterflies. Noctuoidea: (93) *Apantesis (Grammia) virgo* (Arctiidae) (E. U.S.); (94) larva of *Lophocampa maculata* (Arctiidae) on *Salix* (California); (95) *Megalographa biloba* (Noctuidae) (California); (96) *Catocala* species (Noctuidae) (California); (97) *Xanthopastis timais* (Noctuidae) (Florida). Hedyloidea: (98) *Macrosoma* species (Hedylidae) (Ecuador). Hesperioidea: (99) *Phocides* species (Hesperiidae, Pyrrhopyginae) (Ecuador); (100) *Autochton cellus* (Hesperiidae, Pyrginae) (Texas); (101) *Poanes melane* (Hesperiidae, Hesperiinae) (California); (102) larva of Hesperiidae (California). Papilionoidea: (103) *Papilio rutulus* (Papilionidae) (California); (104) larva of *Papilio polyxenes* (Papilionidae) (Costa Rica); (105) *Anthocaris stella* (Pieridae) (California); (106) *Lycaena (Tharsalea) arota* (Lycaenidae, Lycaeninae) (California); (107) *Callophrys (Incisalia) eryphon* (Lycaenidae, Theclinae) (California); (108) *Plebeius acmon* (Lycaenidae, Polyommatinae) (California); (109) larva of *Plebeius acmon* on *Eriogonum,* tended by ants (California); (110) *Apodemia mormo* (Lycaenidae, Riodininae) nectaring at *Eriogonum,* the larval host (California); (111) *Vanessa tameamea* (Nymphalidae, Nymphalinae) (Hawaii); (112) *Cercyonis oetus* (Nymphalidae, Satyrinae) feeding on ripe fruit (Nevada); (113) *Agraulis vanillae* (Nymphalidae, Heliconiinae) (California), (114) larva of *A. vanillae* on *Passiflora* (California); (115) Ithomiini species (Nymphalidae, Heliconiinae) (Costa Rica); (116) *Morpho peleides* (Nymphalidae, Morphinae) (Costa Rica). (Photographs by: C. Covell, 98, 99; R. Coville, 95, 96, 97; J. Hafernik, 100, 105, 106, 109; W. Hartgreaves, 111; W. Middlekauff, 102; P. Opler, 115; J. Powell, 94, 101, 103, 104, 107, 108, 110, 112, 113, 114, 116; unknown, 93.)

having a startle effect on would-be predators (e.g., *Catocala,* underwings, Fig. 96). Other Catocalinae hold the wings out flat at rest, with the HW cryptically colored, matching the FW. Some noctuids are diurnal and brightly colored, like flowers they visit, while many high-latitude and montane species are diurnal and dark colored. The larvae typically are

cylindrical robust caterpillars, bare with only primary setae (cutworms). Those of some Catocalinae have lateral fringes that appress to the substrate, eliminating shadows (Fig. 32). Acronictinae, Pantheinae, and Nolinae have various secondary setae, sometimes in rows or as tufts on verrucae similar to the arctiids. The prolegs are reduced in some subfamilies, with those of segments A3 to A5 nonfunctional (semiloopers), especially in sedentary species that feed on herbs. Noctuids feed on all kinds of plants, probably nearly every gymnosperm and angiosperm family. Many are polyphagous, foraging on low-growing plants at night, while others are specialists on one or a few plants, including those with toxic chemicals or latex, such as Anacardiaceae, Apocynaceae, Asclepiadaceae, Euphorbiaceae, Moraceae, Urticaceae, and Vitaceae, as well as grasses and sedges, Liliaceae and Amaryllidaceae. Several groups feed on fallen leaves or in plant detritus (e.g., Hypenodinae) or on algae, lichens, and fungal-ridden plant matter. Several species are predaceous on scale insects or feed on detritus in spiderwebs or mammal nests. Many are important economically worldwide, especially polyphagous cutworms and armyworms (e.g., *Agrotis, Autographa, Trichoplusia, Heliothis, Pseudaletia*), eating soybean, sugarcane, cereal, legume, rice, and other field crops. There are more than 35,000 named species in 4200+ genera, worldwide, occurring from high elevations above timberline to low deserts and especially numerous in tropical regions.

Hedyloidea Hedylids are peculiar moths that superficially resemble geometrids, with which they were classified until recently proposed as the ancestral butterfly lineage. The hypothesis is based on 10 characteristics: mesothoracic aorta configuration, six features of the adult skeletal structure, an upright egg, larval anal comb, and pupal girdle. None, however, is unique or universal for all members of either Hedyloidea or Papilionoidea to affirm their monophyly. Probably these resemblances evolved independently, as they are not present consistently in ancestral members of the respective butterfly lineages. Adults (Fig. 98) are nocturnal, medium sized (FW 16–32 mm), with broad, semitranslucent wings, weakly scaled in patterns of gray or brownish and white. Wings coupled by a retinaculum and frenulum with a single bristle in the male, weak bristles in the female, typical of most ditrysian moths. The resting posture is characteristic, with the thorax titled so the HW nearly touch the substrate, and the slender abdomen is raised above them. The head is small, eyes are large, proboscis is well developed, labial palpi are ascending. The antennae are usually filiform, bipectinate in a few species, lacking the apical club of butterflies. There are small, tympana-like structures at the base of the FW, similar to those of some Nymphalidae, and there are no abdominal tympana that would link hedylids with Geometridae. The forelegs of males are reduced and not used for walking, like nymphalid butterflies. The egg is pierid-like, upright, spindle-shaped, and ribbed. The larval head is bizarre with elongate, trifid, barbed horns, similar to some nymphalids; the body is smooth and slender, and the last abdominal segment is bifid, somewhat like some notodontid moths and satyrine butterflies. The pupa is exposed, anchored by a silken girdle, similar to Pieridae and some Papilionidae. Recorded larval foods are Euphorbiaceae, Malpighiaceae, Malvaceae, and Sterculiaceae. There are about 40 species, all in the genus *Macrosoma,* restricted to the Central and South American tropics, Cuba, and Trinidad.

Hesperioidea Hesperiidae includes six subfamilies, and the giant skippers, formerly accorded family status (Megathymidae), have been relegated to a subset of the subfamily Hesperiinae by recent authors. Adults (Figs. 99–101) are small to moderately large (FW 8–35 mm), stout bodied with powerful thoracic musculature; they are called skippers because of their quick, darting flights. The third axillary sclerite at the FW base is unusually wide and forms an irregular Y-shaped structure for muscle attachment, a defining character for Hesperiidae. The HW has an area of very small, specialized scales at the base of R + Sc, also not found in other Lepidoptera. Skippers typically perch with the wings outspread (Pyrginae, Fig. 100) or hold the HW out horizontally and the FW upright, slightly cocked open (Hesperiinae, Fig. 101), or they close both wings above the body like other butterflies, especially when taking nectar. The head is broader than the thorax, with the antennae widely separated and enlarged distally into a club with its tip (apiculus) attenuated and curved. The flagellum is strongly bent at the club in Pyrrhopyginae. The proboscis is well developed, and nearly all species feed at flowers, bird droppings, or other nutrient sources. Most skippers are rather drab, predominantly tan, brown, gray, or black, but many tropical species are colorful. The larval head is prominent (Fig. 102), frequently with protruding lobes, separated from the body by a constricted "neck" in all but the *Megathymus* group, a unique condition in Lepidoptera. In megathymids the head is narrower than the thorax, and the pupa moves in the larval tunnel, protruding at adult eclosion, a unique reversal among Obtectomera. The terminal body segment has an anal comb, analogous to that of some moths, used to flip frass from the larval shelter, sometimes remarkable distances. Trapetzinae in Australia feed on monocots, mainly Xanthorraceae, Poaceae, and Cyperaceae, as do Hesperiinae, mostly on grasses. Coeliadinae of the Old World tropics, Pyrrhopyginae in the New World, and Pyrginae specialize on dicots, with more than 50 plant families recorded, and a few feed on monocots. About 3500 species are described in 500+ genera, and many species complexes in tropical regions are not thoroughly studied. Hesperiidae are distributed worldwide except in New Zealand, with greatest richness in the Neotropical region.

Papilionoidea To most people butterflies are among the most conspicuous and recognizable insects. Their diurnal behavior, aesthetic beauty, and limited species numbers render them favorite subjects for beginning naturalists and amateur collectors and teaching insect metamorphosis to primary school children. Moreover, butterflies have been of special significance to biologists, in studies of geographical

distribution patterns, chemical defenses and mimicry, migration, genetics and population biology, and host-plant relationships. In recent decades they have become poster children in conservation efforts. Butterflies are vastly better studied than most moths, but they are negligible in insect biodiversity, making up fewer than 0.1% of all insect species and 9% of described Lepidoptera (likely less than 4%, were moths equally well cataloged). Several adult skeletal features define members of this superfamily as monophyletic, and the antennae have apical clubs, without an apiculus like that of Hesperiidae and not subapically broadened as in sphingid, castniid, and sesiid moths. Systematists recognize four families of butterflies: Papilionidae, Pieridae, Lycaenidae (including riodinids), and Nymphalidae (including libytheids, satyrids, and danaids). There are an estimated 14,000 species, a total that is greatly inflated relative to moths and other insects, because they are more thoroughly studied, with a propensity by the industrious specialists to accord species status to geographically disjunct populations that differ in color patterns and size but not morphologically or by molecular analysis.

Papilionidae: The swallowtail family is one of the most easily recognizable of Lepidoptera, one whose phylogeny is well supported, based on several wing venation and skeletal features and the eversible gland of the larval prothorax (osmeterium). Adults (Fig. 103) are medium sized to very large (FW 14–105 mm), including the largest butterflies (birdwings of the Indo-Australian region); wings are broad, with the FW triangulate and the HW rounded, and often with one or more veins extended into "tails." Most are brightly colored, often aposematic, warning of their distasteful properties, and many swallowtails are models in mimicry complexes; some are polymorphic, with several forms each with different corresponding mimics, often nymphalids. Swallowtails are strong fliers and have been recorded dispersing several miles from a point of origin. Some tropical species engage in mass migrations. The larvae (Figs. 34 and 104) are plump, often ornate with filaments or protuberances. They appear bare but usually have numerous tiny secondary setae. They live exposed on foliage, inactive by day, and depend upon cryptic coloration for protection, often resembling bird droppings in early instars and then graduating to foliage or flower colors, as they grow, or they are aposematic, brightly colored (Fig. 34). The osmeterium is horn-like, forked, usually bright pink or orange, and everted to emit a foul aroma intended to ward off predators. The caterpillars feed on a wide variety of dicot angiosperms, including several groups on Aristolochiaceae and others on Magnoliaceae, Apiaceae, Rutaceae, Lauraceae, and other plants not used by most Lepidoptera. There are about 600 species in 26 genera, with virtually all the world's species described, distributed worldwide, with Parnassiinae in high latitudes and elevations of the Holarctic and Papilioninae mostly subtropical and tropical and with their greatest richness in the Old World tropics.

Pieridae: The whites and sulphurs make up a well-established, monophyletic family, based on several characters: the presence of pterin pigments in the wing scales; the foretarsi with inner claw subequal in length to the outer, whereas the inner is much shorter in other butterflies; and wing venation and thoracic skeletal features. Adults (Fig. 105) are small to medium sized (FW 11–48 mm), broad winged, mostly white, yellow, or orange, with some tropical species brightly colored, containing flavone pigments, mimicking other butterflies (e.g., South American Dismorphiinae). Most pierids display sexual dimorphism in color patterns, sometimes to the extreme, and many have marked seasonal variation. Remarkable mass migrations by some tropical pierids occur, often moving from seasonally dry to wet habitats. Larvae are slender caterpillars, relatively uniform in structure, without protuberances, and covered with short secondary setae, and each segment is divided superficially into six annulets. They are mainly green, including the head, or spotted with yellow and blue in species that feed in flowers. Some species possess an anal comb. Tropical Dismorphiinae feed on legumes, as do some Coliadinae, while most Pierinae specialize on Brassicaceae, Capparidaceae, Loranthaceae, or Santalaceae. The American genus *Neophasia* feeds on pines. Several species have achieved important pest status, particularly the cabbage white *(Pieris rapae),* which was introduced from Europe into North America in the 1880s, feeding on cabbage and other crucifer crops, and species of *Colias,* feeding on alfalfa. More than 1000 species in 75 genera have been described, including probably nearly all the world's species. This group is cosmopolitan except in New Zealand and the Pacific Islands, with greatest development in the tropics. Species range to the extreme limits of Lepidoptera habitats, *Colias* to 83° N latitude and *Baltia* to 5000 m (16,350 ft) in the Himalayas and several genera to similar elevations in the Andes.

Lycaenidae: The coppers (Fig. 106), hairstreaks (Fig. 107), blues (Fig. 108), and metalmarks (Fig. 110) together form a diverse butterfly family, with a remarkable array of larval biologies. Inclusion of the metalmarks (Riodininae) is debatable because they have several uniquely derived traits and because they have foreleg morphology and function that resemble those of Nymphalidae. However, exclusion of the riodinids leaves the remainder of the Lycaenidae an incomplete lineage (paraphyletic). Adults (Figs. 106–108, 110) are mostly small (FW 6–25 mm; Neotropical *Eumaeus* and African *Liphyra* reach 35 mm) and the upper surface of their wings is usually brightly colored, entirely or patterned, in blue, orange, or red, often brilliantly metallic, especially in the males, while the undersides, which are exposed when the butterfly is inactive and the wings are held together above the body, tend to be more cryptic, gray, brown, or green. The wings are relatively broad, the FW usually triangular and the HW rounded; most hairstreaks and a few blues have one or more slender filaments arising from the hind margin, often preceded by a colorful eyespot on the underside. During perching the wings are moved alternately, giving an impression of antennal movements, a behavior thought to deflect predator attack to the HW rather than to the head and thorax. Metalmarks exhibit

a bewildering array of wing forms and color patterns, especially in tropical species, resembling diverse kinds of butterflies and moths. Lycaenid antennal bases are adjacent to and usually indenting the eyes. In Riodininae the antennae usually are long, more than half the FW length, and the forelegs are atrophied in males. The antennae are shorter and the male forelegs functional in other lycaenids. Lycaenid larvae are peculiar caterpillars, shaped like a sowbug, with the body segments broadened laterally and the small head retractable and hidden under the thorax; they are usually covered with short secondary setae, giving a velvety appearance (Fig. 109). Species that live in association with ants are bare, and Riodininae usually have long secondary setae. All lack eversible prothoracic glands characteristic of other butterfly caterpillars. Many Lycaenidae have evolved glands on the last abdominal segment that produce a sweet, honeydew-like fluid that is much sought after by ants, which display various behaviors. Some tend and "milk" the larvae on their food plants (thereby presumably warding off parasites and invertebrate predators) (Fig. 109), others transport the young caterpillars to their nests, where they are fed by the ants or eat the ant brood. A wide variety of flowering plants serve as hosts, including a few conifers and monocots. Most species are specialists, but some are polyphagous. Larvae of the African Poretiinae feed on algae and lichens. Those of Miletinae (African and 1 species in North America, *Feneseca*) feed exclusively on Homoptera or their secretions or in ant nests. The association with ants has developed in many Riodininae and unrelated genera of other lycaenids. Feeding on legumes has led to minor pest status for a few species, including the bean lycaenid *(Strymon melinus)* in North America, a polyphagous species also called the cotton square-borer. There are more than 6000 described species in 640+ genera, and many tropical taxa are not thoroughly studied. Lycaenids occur worldwide, with endemic species even in New Zealand and the Pacific Islands, but the majority occur in the Neotropics and Africa.

Nymphalidae: This is a large and diverse family that includes the typical Nymphalinae (brush-footed butterflies, admirals, checkerspots, Fig. 111), Libytheinae (snout butterflies), Satyrinae (wood nymphs, ringlets, Fig. 112), Heliconiinae (long wings, fritillaries, Fig. 113), Morphinae (morpho and owl butterflies, Fig. 116), and Danainae (milkweed and glasswing butterflies, Fig. 115). All possess three longitudinal ridges (carinae) on the ventral surface of the antennae that are unique in Lepidoptera, and the forelegs of males are reduced or modified (less so in Libytheinae), usually lacking claws and nonfunctional for walking. Adults are small to very large (FW usually 10–50 mm, ranging to 75 mm in tropical *Morpho* and *Caligo*), mostly broad winged except in Heliconinae, and usually brightly colored, often with orange, black, and white dominating, but mostly brown and tan in Satyrinae. Many tropical nymphalids are involved in mimicry complexes, either as models (Danainae, Heliconiinae) or as mimics of them or other distasteful butterflies and moths and/or they benefit in both roles. Glasswing butterflies (Ithomiini) live primarily in deep shade of tropical forests and have sparsely scaled areas or transparent wings, with subtle color patterns, while owl butterflies fly at dusk. They, morphos, and satyrines have conspicuous eye-like spots near the margins of the wing undersides, presumably confusing would-be predators or diverting their attacks from the body. The larvae are cylindrical caterpillars with full complement of abdominal prolegs (Fig. 114), but there are diverse modifications, e.g., densely spinose or with dorsal projections (verrucae) that are spinose (Nymphalinae), smooth with filaments (Danainae), smooth with bifid caudal segment (Satyrinae), pubescent with hair tufts and usually bifid caudally (Morphinae). The pupa hangs head downward, attached by a cremaster, without a silken girdle. The larvae feed on a diverse array of flowering plants, with considerable specialization within subfamilies: Morphinae and Satyrinae almost exclusively on monocots, including Arecaceae, Bromeliadaceae, Helioconiacae, and Musaceae (a few species are pests on bananas) in the tropics, mostly Poaceae and Cyperaceae in the Holarctic, with 2 genera on Selaginellaceae; other nymphalids eat mostly dicot angiosperms, often specializing on plants with toxic chemicals (e.g., Heliconiinae on Flacourtiaceae, Passifloraceae, Urticaceae, Violaceae) or latex-producing plants (Danainae on Apocynaceae, Asclepiadaceae, Moraceae). About 6500 described species are placed in 630+ genera, occurring worldwide, ranging from Arctic–Alpine *Boloria* in the Holarctic to extremely rich tropical faunas in most subfamilies, several of which are not represented in New Zealand.

See Also the Following Articles

Caterpillars • Collection and Preservation • Scales and Setae • Trichoptera • Wings

Further Reading

Common, I. F. B. (1990). "Moths of Australia." Melbourne University Press and Brill, Leiden/New York.

Hering, E. M. (1951). "Biology of the Leaf Miners." Junk, The Hague.

Kristensen, N. P. (1984). Studies on the morphology and systematics of primitive Lepidoptera (Insecta). *Steenstrupia* **10,** 141–191.

Kristensen, N. P. (ed.) (1999). Lepidoptera, moths and butterflies, Vol. 1, Evolution, systematics, and biogeography. *In* "Arthropoda: Insecta" (M. Fischer, ed.), Part 35 of Vol. IV of "Handbook of Zoology." W. deGruyter, Berlin/New York.

Kristensen, N. P., and Skalski, A. W. (1999). Phylogeny and paleontology. *In* "Arthropoda: Insects." (N. P. Kristensen, ed.), Part 35 of Vol. IV of "Handbook of Zoology," Ch. 2, pp. 7–25. W. deGuyter, Berlin/New York.

Labandeira, C. C., Dilcher, D. L., Davis, D. R., and Wagner, D. L. (1994). Ninety-seven million years of angiosperm-insect association: Paleobiological insights into the meaning of coevolution. *Proc. Natl. Acad. Sci. USA* **91,** 12278–12282.

Medvedev, G. S. (ed.) (1987). "Lepidoptera," Part 1. Vol. IV of "Keys to the Insects of the European Part of the USSR." Oxonia Press, New Delhi.

Medvedev, G. S. (ed.) (1990). "Lepidoptera," Part 2. Vol. IV of "Keys to the Insects of the European Part of the USSR." Brill, Leiden/New York.

Mosher, E. (1916). A classification of the Lepidoptera based on characters of the Pupa. *Bull. Illinois State Lab. Nat. Hist.* **12,** 14–159.

Sbordoni, V., and Forestiero, S. (1984). "Butterflies of the World." Times Books, Random House, New York.

Scoble, M. J. (1992). "The Lepidoptera: Form, Function, and Diversity." Oxford University Press, Oxford.

Snodgrass, R. E. (1961). The caterpillar and the butterfly. *Smithsonian Misc. Collect.* **143,** 1–51.
Stehr, F. W. (ed.) (1987). Lepidoptera. *In* "Immature Insects," Chap. 26, pp. 288–596. Kendall/Hunt, Dubuque, IA.
Zimmerman, E. C. (1978). "Microlepidoptera," Parts 1 and 2. Vol. 9 of "Insects of Hawaii." University of Hawaii Press, Honolulu.

Lice

see *Phthiraptera*

Lice, Human

(*Pediculus* and *Pthirus*)

Terri L. Meinking
University of Miami School of Medicine

Lice are wingless, bloodsucking insects that belong to the order Phthiraptera. Although there are currently 4000 species of lice recognized, only 560 species suck blood and feed on mammals. Lice are very host specific; therefore human lice cannot be transmitted to or from other mammals. There are only three species of lice that infest humans: *Pediculus humanus humanus,* the body or "clothing" louse; *Pediculus h. capitis,* the head louse (Fig. 1) and *Pthirus* (or *Phthirus*) *pubis,* the crab or "pubic" louse (Fig. 3).

FIGURE 1 Head louse nit, adult female (left) and male (right). Body lice and head lice are identical in appearance except for size. The body louse is about 25% larger than the head louse. The female has an invaginated V-shape (center). (Photography by Bruce Hard.)

BIOLOGY

Lice, like all other insects, have six legs. They are hemimetabolous in development, meaning that they do not go through a complete metamorphosis like mosquitoes or fleas. There are three nymphal or "instar" stages that all look like a miniature adult. Within 7 to 12 days after eggs (nits) are laid, nymphs hatch, which will molt three times before becoming adults.

Within a few hours of hatching, the nymph must find a human blood meal or it will die of starvation and dehydration. At all stages, lice have a very tough, leathery cuticle capable of considerable expansion after feeding, usually taking in blood meals up to one-third of their own body weight every few hours. Regular feedings occur every 4 to 6 h with head and crab lice, although body lice can survive for days without a blood meal. Several feedings occur between the shed of each chitinous exoskeleton or cuticle. The first, second, and third nymphal stages last 3 to 4 days each. It is not until the final molt that the sex can be determined. Females are usually 20% larger than males of the same species, as well as longer, wider, and rounder, with the posterior portion of the female terminating in an invaginated V shape (Fig. 1). Gender identification is more difficult with *Pthirus* than with *Pediculus.*

Within 2 days of molting the female will feed several times, copulate, and begin laying an average of 3 to 6 eggs per day, with body lice laying more and crab lice laying fewer eggs. The female louse attaches the nits to hairs or fibers (in the case of body lice) by secreting a glue for which there is no solvent (Fig. 2). The life span of a louse from hatching through adult is 30 to 42 days.

HABITAT/EPIDEMIOLOGY

Pediculus h. humanus

The body louse lives in clothing or bedding and travels to the host only to feed. The nits are laid on the clothing fibers, especially in the seams and collars. This is truly an infestation of individuals who are unable to wash themselves or their clothing.

FIGURE 2 Head lice eggs are 0.8 mm in length. Note the glue that attaches the nit to the hair shaft. (Photography by Bruce Hard.)

Infestations with the body louse occur more commonly in individuals crowded together because of war, natural disaster, refugee status, or, in cooler climates, homelessness. They give all lice a "bad name" because people associate all lice with poor hygiene; however, this is the only human louse associated with lack of cleanliness.

Pediculus h. capitis

The head louse likes a clean healthy head and is found commonly on children 3 to 11 years of age. The nits are laid on the hair, usually close to the scalp for warmth. Head lice are primarily transmitted by contact, which is more frequent in younger children; however, older siblings, parents, childcare workers, and teachers may also be infested. Sharing of brushes, combs, hats, helmets, and other headgear and hair accessories also can result in transmission.

Pthirus pubis

The crab louse (Fig. 3) prefers the pubic and perianal areas, but contrary to older literature, can also be found on the beard, mustache, scalp, axillae, eyebrows and eyelashes, or any hairy part of the body. Unlike head lice, which tend to be somewhat particular, crab lice are found in all levels of society, and are generally transmitted by sexual contact. HIV-positive and immunocompromised individuals are more difficult to treat. Adult crab lice can remain alive off the host for at least 36 h; therefore, transmission by infested towels and bedding is more common than originally thought.

DISEASE TRANSMISSION

Pediculus h. humanus

Disease transmission primarily occurs as a result of contact with infested lice fecal pellets. The body louse is capable of transmitting the following diseases: epidemic typhus (caused by *Rickettsia prowazekii*), murine typhus (caused by *R. typhii,* trench fever (caused by *Bartonella quintana*), and relapsing fever (caused by *Borrelia recurrentis*).

FIGURE 3 Crab louse. *Pthirus pubis* is broader than *Pediculus* spp., and resembles a crab.

Pediculus h. capitis and *Pthirus pubis*

Head lice and crab lice are capable of transmitting group A *Streptococcus pyogenes* and *Staphylococcus aureus.* Neither head nor crab lice have been studied as transmitters of blood-borne diseases, but this is an area of current interest.

TREATMENT

Pediculus h. humanus

Clothing and bedding should be disinfested by washing in hot water and then drying in a hot cycle (65°C, 149°F). Permethrin or malathion dusting powders are effective in treatment for mass eradication of body lice infestations. For individual cases, people should be treated with prescription permethrin 5% topical cream or oral ivermectin.

Pediculus h. capitis

Over-the-counter lice treatments that contain 1% permethrin or natural pyrethrin products can be used, but resistance has been reported in the United States and other countries. Prescription malathion lotion and oral ivermectin are the most effective treatments. Lindane products should not be used because of toxicity and resistance.

Pthirus pubis

Although head lice products may be used on crab lice, the most effective treatment is 5% permethrin cream or oral ivermectin. All hairy areas of the body should be treated, including the scalp.

Phthiriasis palpebrarum

Since topical treatments should not be used around the eyes, crab lice of the eyelashes should be treated with petroleum jelly, which suffocates lice and eggs and lubricates the lashes, making nit removal easier.

See Also the Following Articles

Blood Sucking • *Medical Entomology* • *Phthiraptera*

Further Reading

Alexander, J. O. (1984). "Arthropods and Human Skin." Springer-Verlag, Berlin.

Andrews, M. (1976). "The Life That Lives on Man." Taplinger, New York.

Meinking, T. L., and Taplin, D. (1995). Infestations. *In* "Pediatric Dermatology," 2nd ed. (L. A. Schachner and R. C. Hansen, eds.), pp. 1347–1392. Churchill Livingstone, New York.

Meinking, T. L. (1999). Infestations. *Curr. Problems Dermatol.* **11(3)**, 75–118.
Meinking, T. L., and Taplin, D. (1995). Infestations: Pediculosis. *In* "Sexually Transmitted Diseases: Advances in Diagnosis and Treatment" (P. Elsner and A. Eichmann, eds.), pp. 157–163. Karger, Basel.
Meinking, T. L., Burkhart, C. G., and Burkhart, C. N. (1999). Ectoparasitic diseases in dermatology: Reassessment of scabies and pediculosis. *Adv. Dermatol.* **15,** 67–108.

Locusts

R. F. Chapman
University of Arizona

Locusts are medium-sized to large grasshopper-like insects that form swarms of hundreds of millions of individuals with the potential to migrate long distances in the tropics and subtropics. Locusts differ from grasshoppers in their responses to crowding. Locusts behave as solitary insects immediately after hatching or when maintained in isolation, but if they are forcibly crowded for as little as 6 h they subsequently tend to group together, or exhibit gregarious behavior. In contrast, if grasshoppers are kept in a crowd they usually remain as solitary insects and show no tendency to come together. A few grasshopper species, however, do have some tendency to gregarize when forcibly grouped, although this does not occur naturally. Thus, there is no absolute distinction between a "grasshopper" and a "locust."

Locusts do not comprise a single taxonomic group. Rather, they occur in three subfamilies of Acrididae, the Cyrtacanthacridinae, Oedipodinae, and Gomphocerinae. Even within a genus, some species exhibit the swarming habits of locusts, while others lack the habit and never swarm. This is most obvious in the genus *Schistocerca,* where a majority of the American species are nonswarming, and *Dociostaurus,* with many nonswarming species in Asia. *Locusta migratoria,* which extends from Australia and eastern Asia to Europe and West Africa, has a number of subspecies, which differ in their propensity to swarm. The species usually regarded as locusts and their distributions are given in Table I.

In addition to the behavioral change, locusts exhibit a marked color change when crowded. In isolation they are often green or exhibit a more or less uniform color matching that of the background; crowded locusts, however, exhibit a striking black and yellow or orange coloration in the nymphal stages. These changes, however, are not peculiar to locusts; similar changes are shown by some grasshoppers. The proportions of different parts of the body also differ between locusts reared in isolation and those reared in crowds. The most striking difference occurs in the migratory locust, where insects reared in isolation have a strongly crested pronotum (upper surface of the first thoracic segment), but in crowded insects the upper surface of the pronotum is saddle shaped.

PHASE THEORY OF LOCUSTS

Locusts do not swarm continuously; periods of swarming may last for several years but are separated by times when no swarms are reported (i.e., recession periods). The mystery of their apparent disappearance during recession periods was solved for *L. migratoria* by B. P. Uvarov in a paper published in 1921. He proposed the phase theory of locusts, suggesting that during recession periods the insects exist in a form that differs phenotypically and behaviorally from swarming locusts. He referred to these two forms as the "solitary" (later called "solitarious") and "gregarious" phases. *Locusta* in the solitarious phase had, until that time, been placed in a different genus, *Pachytylus.* Subsequent work showed that similar phases occur in the other locust species. Solitarious locusts are typically cryptically colored, relatively inactive as nymphs (but not necessarily as adults; see later), and live in isolation. Gregarious locusts are conspicuous, with contrasting colors; they form bands (as larvae) or swarms (as adults), and are usually highly mobile. The change from one form to another does not occur in a regular manner, but is dependent on environmental conditions. For example, a period of grouping may increase the tendency of nymphs to group (i.e., to become gregarious), but if they then become isolated again they will tend to lose these characteristics. Consequently, there is no regularity in the timing of outbreaks when swarms occur.

PHASE CHANGE IN THE FIELD

During recessions between outbreaks, solitarious locusts may be very widely distributed, but sometimes these insects are very uncommon. The transformation from solitarious to gregarious involves several discrete phenomena that were first recognized by J. S. Kennedy. Outbreaks are initiated by conditions that favor successful breeding, leading to an increase in the population size. Then the population becomes concentrated in particular areas as other parts become uninhabitable. This leads to aggregation in which the previously isolated individuals are forced into intimate contact with one another. Finally, in the process of gregarization, the behavior and physiology of the insects is changed, and they now tend to aggregate spontaneously. Concentration, aggregation, and gregarization are generally dependent on a drying out of the habitat following good breeding conditions.

Population increases may occur in any part of the distribution area of the insects, but swarming of most species seems to originate only in what are called outbreak areas. The peculiarity of these areas is that only here do the conditions for population increase, concentration, and aggregation coexist. The migratory locust has only a single outbreak area in Africa, even though the species is widespread and often common in many parts of the continent. This outbreak area is the delta of the middle River Niger in Mali. The area is unique because, in addition to local rain, it receives

TABLE I The Species of Locusts

Subfamily/Species	Common name	Distribution
Cyrtacanthacridinae		
Anacridium melanorhodon	Sahelian tree locust	Sahel, eastern Africa, southwestern Arabian peninsula
Anacridium wernerellum	Sudanese tree locust	Sahel, eastern Africa
Nomadacris septemfasciata	Red locust	Southern Africa
Nomadacris succincta	Bombay locust	Southwest Asia
Schistocerca cancellata	South American locust	South America
Schistocerca gregaria	Desert locust	Northern Africa, Arabia, Indian subcontinent
Schistocerca piceifrons	Central American locust	Central America
Oedipodinae		
Chortoicetes terminifera	Australian plague locust	Australia
Locusta migratoria	Migratory locust	Southern Europe, Africa south of the Sahara, Malagasy Republic, southern Russia, China, Japan, Philippines, Australia
Locustana pardalina	Brown locust	Republic of South Africa, Mozambique
Gomphocerinae		
Dociostaurus maroccanus	Moroccan locust	Mediterranean countries, Middle Eastern countries east to Kyrgyzstan

considerable moisture from precipitation on the mountains of Senegal, the source of the Niger. From the mountains, the river runs inland to a low-lying area in Mali, where it branches to form a delta before flowing southwest through Nigeria to the Atlantic Ocean. The combined effects of rain and river flooding in the delta region produce an extended growing period for vegetation and so enable the locusts to have as many as four generations within a year, whereas elsewhere in Africa the species usually has only two. As a result, huge population increases can occur. However, as the floodplains dry out, suitable areas of vegetation become increasingly restricted, and the insects are first concentrated and then aggregated into smaller areas, where gregarization occurs. Biogeographical analysis of the occurrence of swarms during the last great outbreak of the migratory locust, which lasted from 1930 until 1940, shows clearly that the plague originated from the single outbreak area and spread progressively over Africa south of the Sahara.

The red locust, in southern Africa, unlike the migratory locust, has only a single annual generation. It has several outbreak areas in Tanzania and Zambia that, like the middle Niger, are floodplains. The red locust outbreak areas, however, have either no or very limited outflow of water. As a result, the water that accumulates and sometimes forms a lake has become salty over time. Nymphal development coincides with the rainy season, when extensive flooding produces lush, tall grasslands in which the locusts feed. As the vegetation dies, its distribution becomes more limited, perhaps as a consequence of increasingly saline conditions in the slightly lower parts of the floodplain and as the area of vegetation becomes more restricted, the locusts become concentrated, with the potential to give rise to swarms.

The desert locust differs from these species in that swarms arise in different places depending on the success of breeding and vegetation changes; there is no evidence of any single outbreak area from which the plagues of the 20th century originated.

MIGRATION

Locust swarms fly during the day and, if they are flying close to the ground, often tend to stream in one direction. This is still true at any one position within a higher flying swarm; in the swarm as a whole, however, the orientation of these streams is random. This would rapidly cause the swarm to disperse except that upon reaching the edge of the swarm, individuals turn back into it. It is not known what stimuli produce this behavior, but vision, sound, or even smell may be involved. Because the locusts within the swarm are, effectively, randomly oriented, the swarm itself has no directional movement and is carried downwind. The rate of displacement of swarms flying close to the ground is less than the airspeed because the insects tend to land at intervals, taking off again as the rest of the swarm passes. In high-flying swarms, however, this is not possible. The locusts may be carried on thermals as high as 3000 m above the ground, and then the swarms are displaced downwind at about the speed of the wind. If the winds are light and variable, swarm displacement is negligible. With sustained winds, however, displacements over hundreds or even thousands of kilometers can occur. This behavior is one of the factors enabling the desert locust to survive in some of the most arid regions on earth, the Sahara and Arabian deserts. Downwind displacement takes the insects to areas of wind convergence, where rain is most likely to occur, if it occurs at all, so that the chances of the insects breeding and producing viable offspring are greatly increased. Because wind patterns are not completely reliable, however, this strategy is not always effective. As a result, swarms of desert locusts in West Africa are sometimes carried out into the Atlantic or north to western Europe. The most spectacular recorded flight occurred in October 1988, when huge swarms were carried right across the Atlantic, with large numbers reaching the Caribbean and the northern coasts of South America, a distance of about 6000 km from the insects' source in West Africa.

The bulk of the work leading to our current understanding of swarm behavior was carried out by R. C. Rainey and Z. Waloff, working on the desert locust in the 1950s and 1960s. The assumption is that swarms of other locusts behave in the same way, although the other species do not generally form such massive swarms and are much less well studied.

It is also known that adults of the solitarious phases of at least some locust species migrate, but they do so at night. Evidence for night migration by solitarious individuals exists for *Anacridium* spp., *L. migratoria migratorioides* (the African subspecies), *Locustana pardalina,* and *S. gregaria. Nomadacris septemfasciata,* on the other hand, appears to be sedentary in the solitarious phase. In this respect, locusts are similar to tropical grasshoppers, many of which are sedentary, whereas a few are known to make extensive night migrations. These solitary migrants, unlike day-flying swarms, deliberately climb to relatively high altitudes (200–500 m above the ground) and then may maintain flight for some hours, although probably a majority of flights are relatively short. In the case of *L. m. migratorioides,* regular flights occur within the floodplains of the outbreak area and from them to the surrounding semiarid country, where breeding may occur. Return migrations to the floodplains also occur, and this strategy moves populations with the Inter-Tropical convergence, along which rain is likely. These flights are sometimes downwind, but there is also some evidence from radar observations that the insects can maintain a particular heading despite shifts in wind direction. These seasonal movements make an important contribution to the survival of the insects. Night flights by the Australian plague locust, *Chorthoicetes terminifera,* are also well documented, but these generally are of shorter range.

CONTROL OF PHASE

The physiology of phase change is not yet fully understood. It has been known for some time that grouping can be induced in isolated nymphs by touching individuals with fine wires dangling from a rotating circle, indicating that it is primarily physical contact with other locusts that initiates gregarization. More recent work has shown that touching the hind femora is more effective than touching other parts of the body. Presumably, the effects are registered by mechanoreceptors on the femora, leading to a change in the nervous system that alters the insect's behavior toward gregariousness. It is very likely, though not yet proved, that this sequence involves neuromodulators. A peptide hormone that enters the hemolymph via the corpora cardiaca induces the dark coloration of gregarious nymphs.

Pheromones play a part in the maintenance of gregarization. A number of experiments indicate that gregarious locusts of both sexes produce a gregarization pheromone. In adult desert locusts, benzaldehyde, veratrole, guaiacol, phenol, and phenylacetonitrile are its major components. This pheromone enhances the tendency to group as well as having some effect on color change. Solitarious locusts do not produce the full suite of compounds in comparable concentrations. It has been shown that the chemicals are produced from plant material ingested by the locusts and that bacteria are responsible for their production. Locusts reared on axenic (microbe-free) diets do not produce the pheromone. Mature males in the gregarious phase of both the desert and migratory locusts produce from epidermal glands a pheromone that accelerates maturation of insects of either sex. The major component of this pheromone in the desert locust is phenylacetonitrile. Its effect under natural conditions is, presumably, to tend to synchronize oviposition by the individuals in a swarm, which increases the likelihood that the first-stage nymphs, when they hatch, will be present in large numbers and so will be likely to interact with each other and gregarize. A chemical produced in the accessory glands of gregarious females of the desert locust promotes gregarious behavior and coloration in the nymphs hatching from the eggs; solitarious females do not produce the chemical. The chemical is contained in the frothy material that forms a plug above the egg mass and that is interpolated in spaces between the eggs. There is thus a marked intergenerational effect of phase with gregarious females producing offspring that already have some characteristics of gregarious individuals.

EVOLUTION OF SWARMING BEHAVIOR

It was once thought that the contrasting coloration of gregarious nymphs was likely to have a function in promoting gregarious behavior, but experimental evidence does not support this. Recent studies with the desert locust show that when the locusts feed on plants containing deterrent chemicals, such as the alkaloid hyoscyamine, predaceous lizards rapidly learn to avoid individuals with gregarious coloration but do not avoid solitariously colored nymphs even when they have eaten the same food. Other plants in the desert areas that are the habitat of *S. gregaria* also contain potentially noxious compounds, and it may be that the gregarious coloration results from selection for warning coloration. This, in turn, may have led to gregarious behavior, since aposematic insects commonly group together. Whether similar arguments can be applied to other locust species is not known.

The tendency to migrate is clearly an adaptation to living in arid habitats, enabling the insects to colonize new areas before the initial food supply is totally depleted. This is most clearly seen in the desert locust. Because some grasshoppers in these same habitats exhibit annual migrations, and some solitary locusts are also known to migrate, it must be supposed that swarm migrations arose from these individual movements. This, however, involved a switch from nighttime migration within the insects' boundary layer, where flight can be directed by the insect, to daytime flight that is often outside the boundary layer and displacement is largely determined by the wind.

See Also the Following Articles

Migration • Orthoptera

Further Reading

McCaffery, A. R., Simpson, S. J., Islam, M. S., and Roessingh, P. (1998). A gregarizing factor present in the egg pod foam of the desert locust *Schistocerca gregaria*. *J. Exp. Biol.* **201,** 347–363.

Pener, M. P., and Yerushalmi, Y. (1998). The physiology of locust phase polymorphism: An update. *J. Insect Physiol.* **44,** 365–377.

Sword, G. A., and Simpson, S. J. (1999). Is there an intraspecific role for density-dependent colour change in the desert locust? *Anim. Behav.* **59,** 861–870.

Sword, G. A., Simpson, S. J., el Hadi, O. T. M., and Wilps, H. (2000). Density-dependent aposematism in the desert locust. *Proc. R. Soc. Lond. A* **267,** 63–68.

Tawfik, A. I., Tanaka, S., de Loof, A., Schoofs, L., Baggerman, G., Waelkens, E., Derua, R., Milner, Y., Yerushalmi, Y, and Pener, M. P. (1999). Identification of the gregarization-associated dark-pigmentotropin in locusts through an albino mutant. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 7083–7087.

Uvarov, B. P. (1966). "Grasshoppers and Locusts," Vol. 1. Cambridge University Press, Cambridge, U.K.

Uvarov, B. P. (1977). "Grasshoppers and Locusts," Vol. 2. Cambridge University Press, Cambridge, U.K.

M

Magnetic Sense

John Klotz
University of California, Riverside

Rudolf Jander
University of Kansas, Lawrence

Experimental evidence for magnetic field sensitivity has been reported in insects belonging to various orders, including Isoptera (termites), Diptera (flies), Coleoptera (beetles), Hymenoptera (ants and bees), and Lepidoptera (moths and butterflies). There is evidence that a few insect species obtain directional information from geomagnetic fields for compass orientation. Two alternative properties of the local geomagnetic vector could serve this purpose. Like a number of birds, animals either make use of the direction in which the dip angle points ("inclination compass") regardless of the field's polarity or sense the local declination and polarity ("polarity compass"). Which of these alternatives pertains to insects has been investigated in only one species, the yellow mealworm *(Tenebrio molitor)* (Coleoptera), which makes use of the polarity compass. The sensory system that mediates magnetoreception in insects has not been identified definitively, though one favored hypothesis is based on the detection of magnetic fields using particles of magnetite.

Magnetic compass orientation can be useful for insects in the context of home range (topographic) orientation and during long-distance migration, especially in the absence of visual compass cues. Both honey bees building combs in darkness and blind termites building oriented mounds appear to use magnetoreception for aligning their structures. On the other hand, it is difficult to imagine how an insect could make adaptive use of sensing the absolute strength of the local geomagnetic field.

MAGNETIC COMPASS ORIENTATION

Multiple directional orientation or compass orientation to artificially induced magnetic fields has been shown in several species. The insects always responded to changes in the magnetic field's declination, which implies sensing of magnetic polarity. In contrast, a geomagnetic inclination compass, as used by some migrating birds, has not been demonstrated for any insect.

Home Range Orientation in Social Insects

The magnetic sense of insects and its adaptive importance have been most thoroughly investigated in social insects such as ants, bees, and termites that require highly developed orientation skills to find and then communicate to their nestmates the location of resources within their home ranges. Much is known about how ants and bees use visual cues such as the sun, polarized light, the moon, and landmarks for spatial orientation. For most navigating insects, the primacy of visual cues must be taken into account in experiments designed to investigate magnetic field orientation. Indeed, an insect's competence in magnetic field orientation may be hidden if more salient cues such as light are present. However termite workers and soldiers, which have poor vision at best, rely more on nonvisual cues.

THE GEOMAGNETIC FIELD AS A BACKUP CUE FOR ANTS

Experiments with naturally foraging weaver ants, *Oecophylla smargdina,* revealed that their sense of direction is stronger and more accurate under clear than under overcast skies. In addition, when ants were tested for orientation indoors after displacement from their outdoor foraging trail, those exposed to overcast conditions maintained the correct trail heading but others exposed to clear skies did not. The difference in response indicated the ants' use of a nonvisual cue that is overridden by celestial cues if they are present. Support for

this hypothesis came with further experiments showing that ants trailing in dim, diffuse light reversed their heading when exposed to an artificially induced magnetic field with polarity opposite that of the geomagnetic field. Wood ants, *Formica rufa,* have also been shown to use magnetic field orientation when directional light cues are unavailable. These experiments with two species of ants suggest a hierarchically organized orientation system designed so that the primary light compass is more efficient than the magnetic compass, which serves as a backup when directional light cues are absent.

MAGNETIC DIRECTION AS A REFERENCE FOR LANDMARK LEARNING IN HONEY BEES In flight, the honey bee, *Apis mellifera,* also uses a magnetic compass in home range orientation. Foraging bees approaching the vicinity of their "target" learn the precise location of resources with respect to surrounding landmarks so they can return to the same place in the future. The most popular hypothesis assumes fast "snapshot"-like recall of near-target constellations of landmarks. The returning bee finds the target location by matching the current perception of landmarks with the "snapshot memories" of them. While learning the spatial relations of landmarks, bees face in a preferred compass direction, using directional light and the geomagnetic field. Honey bees trained in an artificial field with polarity reversed to the geomagnetic field face landmarks in the opposite direction. Hence, their magnetic compass may provide directional information as a frame of reference for the memorized landmarks.

A MAGNETIC CUE FOR HOMING TERMITES All termites are social insects that have evolved a different set of adaptations for home range orientation. Termites are specialized for foraging underground and in enclosed spaces. The eyesight of workers and soldiers either has regressed or has been lost completely. All foraging termites depend heavily on pheromone trails for finding their way back home. However, as in ants, such trails do not provide any cue that helps to discriminate between the outward and homeward direction. The geomagnetic field could provide such a cue. This has indeed been demonstrated in the blind African grass-harvesting *Trinervitermes geminatus* (Termitidae: Nasutitermitinae) which, unlike the majority of termite species, is an open-air forager. Homing orientation in returning workers is substantially disturbed by distortions of the geomagnetic field due to weak bar magnets. Whether geomagnetic field orientation is widespread among termites is still an open question.

Migration in Moths and Butterflies

Long-distance compass migration has evolved in relatively few species of insects as an adaptation for dispersal and for coping with seasonal climatic changes. Examples are found among dragonflies (Anisoptera), true bugs (Heteroptera: e.g., the large milkweed bug, *Oncopeltus fasciatus*), and moths and butterflies (Lepidoptera). The implied geographic orientation mechanism could, plausibly, make use of magnetic compass orientation, especially during nocturnal migration and migration under dense overcast. Some evidence supports this possibility.

Two nocturnal cross-country migrants, the large yellow underwing moth, *Noctua pronuba,* and the heart-and-dart moth, *Agrotis exclamationis,* reversed their direction of orientation in four-armed bioassay arenas when they were exposed to reversals of an artificial magnetic field.

The most spectacular example of geographic orientation in insects is the massive annual fall migration of the monarch butterfly, *Danaus plexippus.* Eastern North American populations of monarchs migrate over 3000 km to winter in the mountains of Mexico. Experimental evidence substantiates their use of a sun compass for geographic orientation, and some experiments suggest their use of a magnetic compass as well.

For years entomologists have speculated about magnetic compass orientation in migrating monarch butterflies. The first supportive evidence came in field experiments: migratory butterflies were exposed to a brief pulse of an induced magnetic field 15,000 times the intensity of the geomagnetic field, whereupon the treated butterflies were released and tracked to determine their direction of flight. Two control groups of butterflies were also tested. One of the control groups received the same treatment as the experimental group except for the magnetic pulse. The other group received no treatment and was composed of naturally occurring butterflies migrating through the test area. Both control groups of butterflies kept their normal migratory flight direction to the southwest, but directional headings of the magnetically treated group were randomly distributed, indicating disorientation. Because these experiments were conducted on clear days, however, the sun was also available as a cue. Thus conflicting information from the butterflies' sun and magnetic compasses may have caused the insects' disorientation.

OTHER EFFECTS OF MAGNETIC FIELDS ON THE ORIENTATION BEHAVIOR OF INSECTS

Some of the earliest and most detailed studies of magnetic field sensitivity were also conducted with social insects. Once again the honey bee was the focus of intense research, but this time the investigators studied its communication behavior. To recruit and guide nestmates to a newly discovered resource, a honey bee performs a dance indicating to her followers the direction and distance of the resource from the hive. The dance is usually performed in darkness on a comb's vertical surface. The flight direction to the resource in reference to the sun is transposed by the bee to the direction of her dance with respect to gravity on the comb. If the resource is in the direction of the sun, the dance is directed upward; if away from the sun, downward; and if in other locations, at various angles to the vertical. Small systematic errors in the directional component of this dance are correlated with daily fluctuations in intensity of the geomagnetic field. These errors disappear when the bees

dance in an artificial magnetic field that compensates for the earth's field.

When honey bees are forced to dance on a horizontal surface in the dark, their dances become aligned with the cardinal and intercardinal axes of the geomagnetic field. This response intensifies when the magnetic field is artificially enhanced and disappears when the field is canceled.

Evidence also suggests that honey bees use magnetic fields in nest construction. Bees that are transferred to a new hive construct combs that are oriented in approximately the same magnetic direction as those in their old hive. In one study, bees built abnormal combs when they were exposed to magnets during construction.

Among the most spectacular and magnificent termite mounds are those of *Amitermes meridionalis* (Termitidae: Amitermitinae) in tropical Australia near the town of Darwin. These massive tombstonelike black structures reach up to 4 m in height, and their long horizontal axes align near perfectly north–south. Similar but less perfectly oriented and shaped mounds are constructed by *A. laurensis* on the Cape York Peninsula of Australia. It is more than tempting to refer to these mound builders as "magnetic termites."

Indeed some good evidence supports this label. If a strong, permanent magnet is buried underground where a new colony starts to build, the resulting structure is misshapen and lacks clear orientation.

In addition to nest alignment, numerous studies have identified insects that align the body axis to magnetic fields. Resting termites, flies, and honey bees adopt positions aligned with the cardinal axes of a magnetic field.

Finally, orientation transfer sometimes occurs from light orientation to magnetic compass orientation. When a yellow mealworm moves away from a light source, it remembers its current magnetic compass bearing. If the directional light is turned off, the course direction is maintained with the help of the remembered magnetic compass bearing.

POSSIBLE SENSORY MECHANISMS

Several hypotheses have been proposed to explain how animals sense magnetic fields. There is circumstantial but no definitive evidence in insects for two such sensory mechanisms. One type of mechanism could be based on the magnetic sensitivity of some chemical or photochemical reactions. If such reactions are linked to light reception in the eye, then changing the wavelength of ambient visible light could alter the directional orientation to the geomagnetic field. Such effects have been obtained in male *Drosophila melanogaster* (Diptera) as well as in some birds. A second mechanism could be based on the interaction between the geomagnetic field and intracellular, submicroscopic magnetite particles that have been found in some insects, including ants, honey bees and monarch butterflies.

See Also the Following Articles

Dance Language • Learning • Monarchs • Orientation

Further Reading

Jander, R. (1977). Orientation ecology. *In* "Encyclopedia of Ethology" (B. Grzimek, ed.), pp. 145–163. Van Nostrand Reinhold, New York.

Jander, R., and Jander, U. (1998). The light and magnetic compass of the weaver ant, *Oecophylla smaragdina,* (Hymenoptera: Formicidae). *Ethology* **104,** 743-758.

Perez, S. M., Taylor, O., and Jander, R. (1999). The effect of a strong magnetic field on monarch butterfly *(Danaus plexippus)* migratory behavior. *Naturwissen schaften* **86,** 140–143.

Walker, M. M. (1997). Magnetic orientation and the magnetic sense in arthropods. *In* "Orientation and Communication in Arthropods" (M. Lehrer, ed.), pp. 187–213. Birkhäuser Verlag, Basel.

Weaver, J. C., Vaughan, T. E., and Astumian, R. D. (2000). Biological sensing of small field differences by magnetically sensitive chemical reactions. *Nature* **405,** 707–709.

Wiltschko, R., and Wiltschko, W. (1995). "Magnetic Orientation in Animals." Springer-Verlag, New York.

Malaria

William K. Reisen
University of California, Davis

Malaria is a pyrogenic (fever-producing) disease caused by infection with one of four species of parasitic protozoa in the genus *Plasmodium* and is the most important arthropod-transmitted pathogen in the world today, in terms of numbers of cases, deaths, and economic burden. Acquired from the bite of an infective *Anopheles* mosquito or from infected blood products, malarial parasites continue to suppress development in Africa and parts of Asia and are emerging as a critical health issue in tropical Central and South America. Expanding and rapid global commerce and travel provide an effective conduit for malaria parasites to be reintroduced into currently malaria-free areas.

HISTORY AND DISCOVERY

There is little doubt that there has been a long evolutionary association between humans and malaria. The ascent of the human species and its dispersal from the African center of origin into Europe and Asia most likely were accompanied by host-specific and coevolved species of plasmodia. *Vivax* malaria possibly accompanied early Asian voyagers to the New World across the Pacific Ocean, whereas *falciparum* malaria probably was introduced into the New World from Africa with the post-Columbus slave trade.

Malarial disease has impacted human health throughout recorded history. References to seasonal intermittent fevers abound in the earliest Assyrian, Chinese, and Indian religious and medical writings; however, it was not until the 5th century B.C. that Hippocrates related the distribution of cases to specific seasons and residence near marshes. Malaria has altered the course of human history by afflicting political, scientific, and

religious leaders as well as decimating invading armies. The Romans associated marshes with intermittent fevers and attempted to reduce their occurrence through swamp drainage. The term malaria was derived from the Italian *mal' aria* (bad air), drawing from the association between foul-smelling marsh gases and the occurrence of this disease. In the 1600s, powders from the bark of the Peruvian "quina-quina" tree (now known as quinine) were discovered in South America and shown to be therapeutic against certain seasonal fevers. Shortages of these natural medicinal powders and the resulting impact of malaria on military campaigns during World War I stimulated research to develop antimalarial drugs and resulted in the formulation of atebrin in 1930 and chloroquine in 1934. Although marshes, mosquitoes, poor living conditions, and malaria were associated throughout history, it was not until 1880 that Laveran first observed parasites in the blood of fever patients and 1897 that Ross found malarial parasites in an *Anopheles* mosquito that previously had fed on a malaria patient. The following year, Ross worked out the complex life cycle of the malaria parasite using a *Culex* mosquito–sparrow malaria model. Shortly afterward Grassi and colleagues elucidated the life cycle of the human parasite and with Manson demonstrated that protection from mosquito bites provided protection from infection.

The now-confirmed relationship between malarial infection and mosquitoes led to expanded control efforts by chemically treating or reducing surface water where larval mosquitoes occurred. In 1936, the insecticidal properties of DDT were discovered by Muller and Weisman. DDT spraying was used in the successful eradication of introduced *Anopheles gambiae* mosquitoes from Brazil in 1939–1940 and Egypt in 1942–1945. These successes and the eradication of malaria from the United States by spraying the inside walls of houses with DDT set the stage for the 14th World Health Assembly to adopt a global malaria eradication strategy that was implemented by the World Health Organization from 1957 through 1969. However, failure to sustain the effort and funding after initial success, disregard for the magnitude of the malaria problem in Africa, and the onset of insecticide resistance in several key vector species resulted in a global collapse of this effort that was followed by a general resurgence of malaria throughout tropical regions of the world. In 1992, the World Health Organization again selected malaria as the target disease for a global initiative to improve human health and in 1998 launched the new "Roll Back Malaria" campaign to reduce malaria by 50% by the year 2010. Only time will determine if the human host finally will rid itself of its malaria burden.

PARASITES AND LIFE CYCLE

Four species of human *Plasmodium* may be identified, in part, by clinical symptoms such as the pattern of fever and chills (Table I), morphology and staining characteristics of the parasite within red blood cells, antigenic properties determined by serology, or genetic sequence. The *Plasmodium* life cycle is complex (Fig. 1). The female *Anopheles* becomes infected when gametocytes are ingested during blood feeding. Sexual union of gametocytes occurs in the mosquito midgut, after which the resulting ookinete penetrates the midgut wall and forms an oocyst. After asexual reproduction, the oocyst ruptures and the motile sporozoites make their way to the salivary glands. Humans become infected during blood feeding by the infective mosquito host when sporozoites are expectorated with mosquito saliva into the wound created by the mosquito bite. After entry into the circulatory system of the human host, sporozoites rapidly enter the liver where asexual reproduction occurs. Liberation from the liver may occur rapidly or be delayed, depending upon the species and strain of parasite (Table I). Once in the bloodstream, parasites rapidly enter red blood cells where they multiply asexually. The synchronous liberation of parasites from the red blood cells results in

TABLE I Characteristics of Human Infection with Four Species of *Plasmodium*

Characteristic	*Plasmodium* species			
	vivax	*ovale*	*malariae*	*falciparum*
Incubation period (days)	13 (12–17)[a]	17 (16–18)	28 (18–40)	12 (9–14)
Exoerythrocytic cycle[b]	Present	Present	??	Absent
Merozoites/tissue schizont	>10,000	15,000	2,000	40,000
Erythrocytic cycle (h)[c]	48	49–50	72	ca. 48
Parasitemia (avg. per ml)	20,000	9,000	6,000	20,000–500,000
Attack severity	Mild–severe	Mild	Mild	Severe
Paroxysm duration (h)	8–12	8–12	8–10	16–36 or longer
Relapses	++	++	+++	None
Period of recurrence	Long	Long	Very long	Short
Duration of untreated infection (years)	1.5–3	1.5–3	3–50	1–2

Note. Modified from Bruce-Chwatt (1980).
[a]Strain dependent, may be up to 9 months.
[b]Continued production of merozoites within the liver.
[c]Time between red blood cell infection and rupture indicated by the pattern of paroxysms.

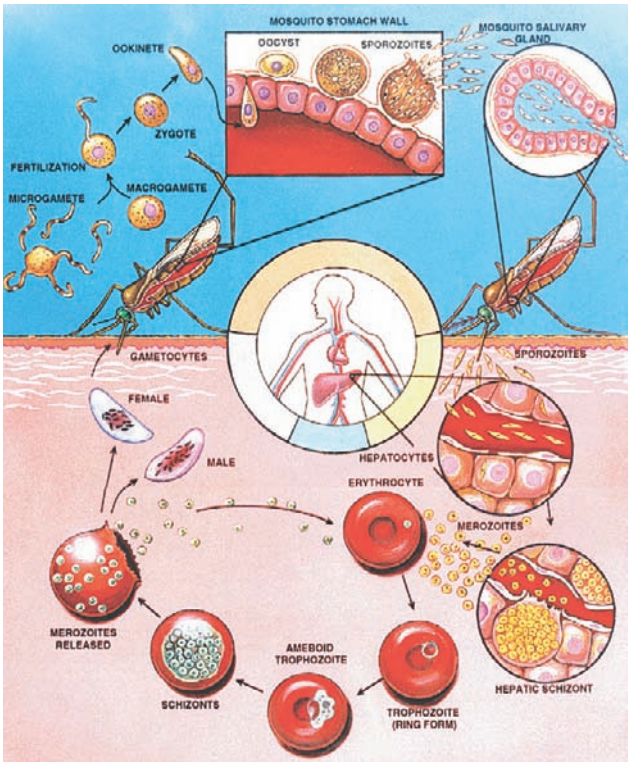

FIGURE 1 Generalized life cycle of the four human-infecting *Plasmodium* species. (From the U.S. Centers for Disease Control and Prevention.)

rhythmic paroxysms characteristic of malarial disease. As the infection progresses, gametocytes are formed in the peripheral bloodstream from where they are ingested by blood-feeding mosquitoes, thereby completing the life cycle.

MOSQUITO VECTORS

Only female mosquitoes in the genus *Anopheles* serve as definitive hosts for the four species of human malarial parasites. Of these, species in the subgenus *Cellia* account for most of the current global transmission and include members of the *A. gambiae* complex *(gambiae, arabiensis)* and *A. funestus* in subSaharan Africa and the *A. culicifacies* complex, *A. fluviatilis* complex, *A. stephensi,* and *A. minimus* in Asia. Historically, the *A. maculipennis* complex was important in the Mediterranean and Europe, whereas species in other subgenera such as *A. darlingi* and *A. albitaris* have been responsible for the resurgence of malaria in South America.

EPIDEMIOLOGY AND DISEASE

Malaria remains a critical health problem of global proportions, causing an estimated 500 million clinical cases and 2.7 million deaths annually. It is a health problem of crisis proportions and a severe economic burden in 90 countries inhabited by 2.4 billion people (roughly 40% of the world population). The temporal concordance between crop growing and malaria transmission seasons frequently results in a serious loss of agricultural productivity. The distribution of malaria in time and space and the efficiency of transmission are limited by temperature requirements for the development of the *Plasmodium* parasites within their poikilothermic *Anopheles* hosts and the abundance, bionomics, and behavior of the different *Anopheles* vectors. *P. vivax* can develop at temperatures as low as 14.5°C and is found at colder latitudes and higher elevations than *P. falciparum,* which requires temperatures above 16°C (Fig. 2). In addition to ambient temperature, transmission efficiency depends almost entirely on *Anopheles* bionomics expressed as vectorial capacity; species that are long lived, rapidly develop parasites, and feed frequently on humans are the most efficient transmitters of malaria parasites.

The incubation period between infection and clinical illness varies among malarial species and strains, being shortest for *P. falciparum* and as long as 9 months for some northern strains of *P. vivax* (Table I). Illness is characterized by the malarial paroxysm and, if untreated, increases in severity as the number of parasites multiplies logarithmically. Typical

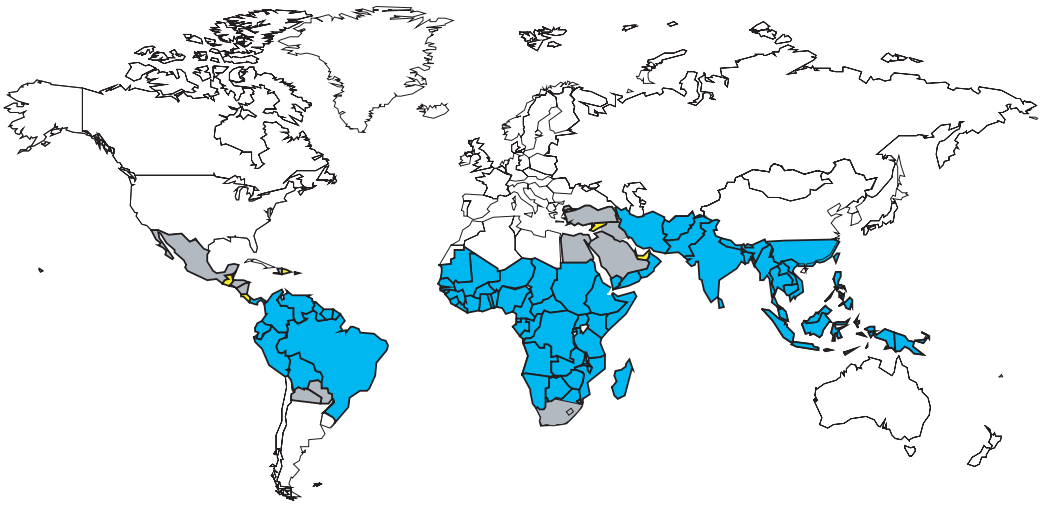

FIGURE 2 Global distribution of malaria. Extended *vixax* area shown in gray. (From the U.S. Centers for Disease Control and Prevention.)

complications include anemia and splenomegaly. In *P. falciparum* infection, changes in the structure of infected red blood cells creates congestion and blockages within the circulatory system, causing coma (brain hemorrhages), jaundice, and "blackwater fever" with the passing of black urine (liver failure, nephritis), and severe dysentery (dehydration, renal failure). Infection during pregnancy frequently leads to abortion, stillbirth, and neonatal mortality. Some liver stages of *P. vivax* and *P. ovale* remain dormant and, if untreated, may relapse for years after the initial infection.

Infection imparts transient immunity that is maintained in endemic areas by almost constant reinfection. In hyperendemic areas, morbidity and mortality are highest among the nonimmune, including travelers, infants, and pregnant women. Adults in these areas tolerate chronic infections and present a constant source of gametocytes for mosquito infection. Malaria mortality has selected for resistant and semiresistant phenotypes from the genome of affected human populations, leading to the evolution and persistence of traits such as sickle cell anemia and Duffy blood group antigen that alter the structure and surface of red blood cells, making them resistant to parasite infection.

TREATMENT AND CONTROL

Treatment has emphasized the use of chemical derivatives of the quinoline ring, originally found in quinine and present in chloroquine and primaquine. Primaquine has the important feature of destroying the liver stages of *vivax* and *ovale,* thereby eliminating relapses. Resistance has led to the development of alternative drugs, including proguanil, mefloquin, pyrimethamine, and sulfonamide; however, in some areas of Southeast Asia treatment of patients infected with resistant strains must revert to quinine with tetracycline. Extracts from plants of the genus *Artemisia* represent a new class of drugs from Asia that are undergoing clinical trials for the treatment of drug-resistant malaria.

Public health control efforts targeting eradication combined active-case detection and treatment with adult mosquito abatement. Active-case detection emphasized complete village-level surveys, the presumptive treatment of fever cases with chloroquine, and verification of malaria infection by slide examination. Residual house spraying with DDT and later malathion targeted indoor resting *Anopheles* females in an attempt to interrupt the transmission cycle. This combined approach resulted in remarkable successes in areas such as Sri Lanka, Pakistan, and India, where the primary vector, *A. culicifacies,* rests almost exclusively within houses and cattle sheds. However, interest and funding to sustain successful programs waned and eventually collapsed. Recently eradication has changed to control, is limited to passive case detection and treatment, and has been incorporated into general village-level health programs.

In addition to research to improve and expand the number of drugs for patient therapy, three control approaches currently are being investigated: (1) personal protection by sleeping under pyrethroid impregnated bed nets (the simplest and least expensive technology that is currently available), (2) vaccination (protection that probably will be of short duration and is expected within 7–10 years), and (3) genetic manipulation of vector competence in *Anopheles* (species-specific, costly, and untried). Success of malaria intervention in developing countries most likely will continue to be hindered by inadequate delivery systems, political unrest, and the low socioeconomic level of most rural populations. In the modern era, successes in malaria control typically have accompanied advances in education, economic well-being, and medical delivery systems.

See Also the Following Articles

DDT • Medical Entomology • Mosquitoes

Further Reading

Bruce-Chwatt, L. J. (1980). "Essential Malariology." Heinemann, London.

Garnham, P. C. C. (1966). "Malaria Parasites and Other Haemosporidia." Blackwell Sci., Oxford.

Macdonald, G. (1957). "The Epidemiology and Control of Malaria." Oxford University Press, London.

Malaria Foundation International. (1998). Worldwide impact of malaria. http://malaria.org/bginfo.html.

Oaks, S. C., Mitchell, V. S., Pearson, G. W., and Carpenter, S. C. J. (eds.) (1991). "Malaria: Obstacles and Opportunities. Report of the Committee for the Study of Malaria Prevention and Control: Status Review and Alternative Strategies." Division of International Health, Institute of Medicine, National Academy Press, Washington, DC.

Russell, P. F., West, L. S., Manwell, R. D., and Macdonald, G. (1963). "Practical Malariology." Oxford University Press, London.

World Health Organization. (1998). Malaria. Fact Sheet No. 94. www.who.int/inf-fs/en/fact094.html.

Mallophaga

see *Phthiraptera*

Mantodea

(Praying Mantids)

Lawrence E. Hurd
Washington and Lee University

The praying mantis (from the Greek for "prophet"), or mantid (from the name of one family in this group), has been mentioned in literature at least since the Egyptian *Book of the Dead.* However, surprisingly little formal scientific research has been done on mantids compared with many other orthopteroid insects, e.g., grasshoppers. No doubt at least some of the reason for this lies in the greater relative economic importance of

insects that compete with humans for food. In any case, to date there has been but a single book summarizing research with mantids, *The Praying Mantids.* There are more than 1800 species worldwide, most of which inhabit the tropics. Among the 20 or so species that occur in the continental United States, the best known, most abundant mantids are 3 species that were introduced from Europe and Asia over the past century.

CLASSIFICATION OF MANTIDS

There is some disagreement as to the proper classification of mantids within the class Insecta. Although they clearly are related to cockroaches, grasshoppers, crickets, stick insects, and termites, most modern systematists recognize that mantids have sufficiently distinctive morphological characteristics to warrant taxonomic separation from these groups. The prevailing view is that praying mantids comprise several families (including Mantidae), either within the order Dictyoptera, suborder Mantodea, or else within a separate order, the Mantida. The most obvious morphological features that characterize this group are a highly mobile head, elongated prothorax (most anterior midbody segment), and especially the raptorial front legs attached to the prothorax. The forelegs are folded when the animal is at rest, giving it an attitude of being in prayer (hence, the common name for the group). The combination of these features is unique among insects.

Much of the uncertainty over classification of praying mantids is because a coherent phylogeny of this group is lacking. The fossil record for this group is both scanty, and recent: fossil mantids date no earlier than the Cenozoic. Therefore, scientists have had to infer phylogenetic relationships from a variety of features such as body shape, presence of auditory organs, and genetics of sex determination. These features have led to different conclusions as to whether, for example, modern mantids have monophyletic or polyphyletic origins. The recent discovery of the new order, Mantophasmatodea, hints at a close relationship between mantids and stick insects (Phasmida). However, based upon morphological and molecular evidence to date, the closest relatives of mantids appear to be Blattodea (cockroaches) and Isoptera (termites), both of which have much earlier origins in the fossil record.

Most biologists agree that there are many more species of insects on earth than have been identified so far, and this is almost certainly true with mantids. There are currently 8 families and 28 subfamilies in this suborder. The most important family (about 80% of all named species) is Mantidae, consisting of 21 subfamilies and 263 genera. This family contains some of the most abundant and widely distributed species on earth including the Chinese *(Tenodera aridifolia sinensis)* and European *(Mantis religiosa)* mantids that occur widely in the temperate zones of Europe, Asia, and North America.

Three families in Mantodea are well represented in the Old World in addition to Mantidae: Amorphoscelididae (two subfamilies found widely distributed in Africa and Australia), Empusidae (eight genera in Africa and Asia), and Eremiaphilidae (two genera of ground-dwelling desert species in Africa and Asia). Another tropical family, Hymenopodidae, contains three subfamilies that include some relatively rare and spectacular flower mimics such as *Hymenopus coronatus* (Asia) and *Pseudocreobotra ocellata* (Africa).

The remaining three families of Mantodea are less diverse tropical groups: Mantoididae, (with a single neotropical genus, *Mantoida*), Chaeteessidae (with only one neotropical genus, *Chaeteessa*), and the most primitive family in the order, Metallyticidae (with a single Malaysian genus, *Metallyticus,* named for their characteristic metallic coloration).

NEUROPHYSIOLOGY AND BEHAVIOR

Perhaps the majority of scientific studies of praying mantids, particularly during the past few decades, have involved the interaction of neurophysiology and behavior. These interactions include the role of binocular vision in estimating distance, hearing of ultrasound and its possible use for avoidance of bat predation, and behaviors associated with defense, sex, and prey capture. Mantids are models of behavioral complexity beyond the imaginations of earlier researchers. They are capable of integrating much detailed information from their environment and have exhibited an astonishingly sophisticated array of responses to stimuli.

Sexual behavior and cannibalism in mantids are particularly noteworthy, partly because they have received much anecdotal mention in the literature. Sexual behavior varies among species, but in general females attract males through a combination of airborne pheromones and visual cues. In many species females either cannot or do not fly, and so males find females by flying upwind along the pheromone plume. This places males at greater risk of predation than females because they are more apparent to birds while in flight. When a male finds a female he incurs even greater jeopardy from his intended mate, depending on her hunger level. Sometimes males are simply captured and eaten before they have a chance to mate, but unless the female is very hungry, he usually is able to mount her. However, a hungry female may decapitate and partially consume a male during copulation without interrupting the transfer of sperm.

The noted French naturalist of the late 19th and early 20th centuries, J.-H. Fabre, described in lurid detail the cannibalistic mating habits of female mantids in his laboratory. This behavior was once interpreted as "adaptive suicide" by the male, to invest both his sperm and his nutrients in the next generation. This requires the assumption that a cannibalized male can be sure he is the father of a female's brood and not simply a food item for a female that has already been fertilized by a previous suitor. The simpler, modern explanation for this behavior is that a female attracts males both for sperm and for nutrition at the end of the growing season when alternate prey are scarce and she has to gain significant mass to produce viable eggs. Males have no choice in the matter, because they cannot discern which pheromone-emitting females have been mated (Fig. 1).

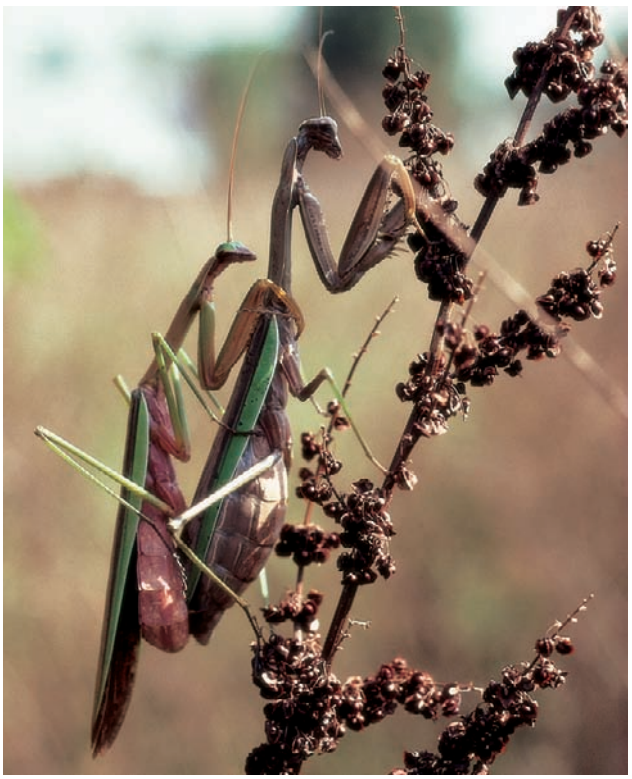

FIGURE 1 A mating pair of Chinese mantids, *Tenodera aridifolia senensis.* The female's abdomen is already swollen with eggs. This is the most widespread and abundant species in the eastern United States and may have the widest global distribution of any mantid species.

ECOLOGY

Studies of the feeding in praying mantids link behavior and ecology. Praying mantids are bitrophic, feeding both on herbivorous arthropods and on other carnivores (e.g., spiders), including cannibalizing each other. The fact that all of these processes may be occurring simultaneously in the same ecosystem can complicate definition of the ecological role of these predators in ecosystem structure and dynamics.

Experimental studies show that bitrophic mantids have both direct (prey reduction) and indirect (prey enhancement) effects, because competition with, or predation on, other predators may reduce predation on some prey species. Mantids tend to eat many arthropods that are beneficial to plants, including pollinators such as bees and butterflies and predators such as wasps and spiders. Whether it is a good idea to add these predators to one's garden as an agent of biological pest control is not always clear. Much more evidence is required before generalizations can be made with confidence, but experiments have demonstrated that under natural conditions mantids can instigate a trophic cascade (top-down effect) whereby plant productivity is enhanced when mantids feed on herbivorous insects.

See Also the Following Articles

Hearing • Orthoptera • Predation

Further Reading

Helfer, J. (1963). "How to Know the Grasshoppers, Cockroaches and Their Allies." Brown, Dubuque, IA.

Hurd, L. E., and Eisenberg, R. M. (1990). Arthropod community responses to manipulation of a bitrophic predator guild. *Ecology* **76,** 2107–2114.

Prete, F. R., Wells, H., Wells, P. H., and L. E. Hurd (eds.) (1999). "The Praying Mantids." Johns Hopkins Press, Baltimore.

Mantophasmatodea

Klaus-Dieter Klass
Museum für Tierkunde, Dresden

Mantophasmatodea are hemimetabolous, wingless pterygote insects, 11 to 25 mm in length, found in Africa. Their body structure is fairly generalized, but a dorsal process on the tarsi, an unusual course of the subgenal sulcus on the head, and a medioventral projection on the male subgenital plate are unique features for this order. Discovered in 2001, the Mantophasmatodea are the most recently described order of insects. Knowledge about them is expected to increase rapidly.

SYSTEMATICS AND DISTRIBUTION

The description of the order and its first two species was based on the first extant specimens that were recognized: two museum specimens described as *Mantophasma zephyra* (collected in 1909 in Namibia; Fig. 1) and *M. subsolana* (collected in 1950 in Tanzania). Members of *Mantophasma* have fairly small eyes and ventral rows of spines on the fore- and midleg femora and tibiae. Two other extant species, yet undescribed, also occur in Namibia. One of them (the "gladiator") has larger eyes than *Mantophasma* and stout spines on the thoracic terga. Specimens recorded from the Western and Northern Cape Provinces of South Africa represent several additional undescribed species; some were collected 100 years ago. *Raptophasma* from Baltic Eocene amber (about 45 mya), with two known species, resembles *Mantophasma,* but has large eyes and stouter, spineless legs. These Tertiary fossils show that the order once also occurred in northern Europe, at a time when its climate was warm and humid.

FIGURE 1 *M. zephyra,* female. (Reprinted from Klass *et al.* 2002, *Science* **296,** 1456–1459. Copyright 2002 American Association for the Advancement of Science.)

ANATOMY

The hypognathous head lacks ocelli and bears orthopteroid mouthparts and long, multisegmented antennae. Wings are lacking. The pleuron of the prothorax is fully exposed. The legs have elongate coxae. In the five-segmented tarsi (Fig. 2A) the three basal tarsomeres are fused. The dorsal membrane beyond the third tarsomere (tm3) bears a characteristic triangular process (dpt in Fig. 2A), and the arolium (arl) of the pretarsus is conspicuously large.

In the abdomen, both tergum I and coxosternum I are free from the metathorax. The small spiracles lie in the pleural membrane and have a muscular closing device. In the male (Fig. 2B, C), coxosternum IX forms a subgenital lobe with a median spatulate process (spp in Fig. 2B) but without styli. The phallomeres (male genitalia) are reduced to membranous lobes around the gonopore. Behind them a transverse, asymmetrically produced sclerite articulates upon tergum X, which is similar to the vomer in Phasmatodea. The female (Fig. 2D) has a short subgenital lobe formed by parts of coxosternum VIII. The genital opening lies above it on segment VIII. The ovipositor comprises clawlike gonoplacs (third valves), blunt gonapophyses VIII (first valves), and gonapophyses IX (second valves) fused with the gonoplacs and interlocked with the gonapophyses VIII. The one-segmented cerci are long claspers in the male but short in the female (Fig. 2B–D).

The foregut has a large proventriculus (gizzard) armed with weak sclerites that terminate posteriorly in three successive whorls of lobes. Midgut ceca are a pair of short and wide lateral pouches. The abdomen has a ventral diaphragm. In the nerve cord, abdominal ganglion VII is free from the terminal ganglion including neuromeres VIII and the following. The egg lacks a defined operculum but has a circumferential ridge; the chorion displays a hexagonal pattern of grooves that are traversed by delicate bars.

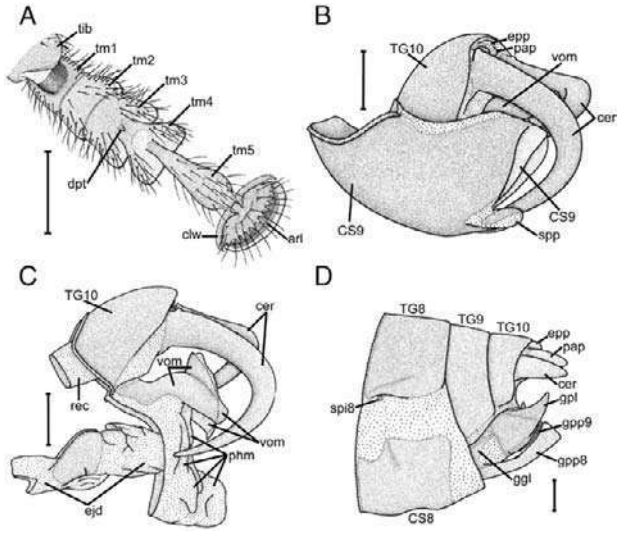

FIGURE 2 (A) *M. zephyra,* dorsal view of tarsus of female. Lateral views of male *M. subsolana* postabdomen with (B) and without (C) coxosternum IX, genitalia exposed. (D) Lateral view of female *M. zephyra* postabdomen. Scale: 0.5 mm. Abbreviations: arl, arolium; cer, cerci; clw, claw; CS8,9, abdominal coxosterna VIII and IX (subgenital plates); dpt, process beyond third tarsomere; ejd, ejaculatory duct; epp, epiproct; ggl, gonangulum; gpl, gonoplac; gpp8,9, gonapophyses VIII and IX; pap, paraproct; phm, phallomeres; rec, rectum; spi8, spiracle VIII; spp, spatulate process of male subgenital lobe; TG8,9,10, abdominal terga VIII, IX, and X; tib, tibia; tm1–5, tarsomeres 1–5; vom, vomerlike element. (Part A reprinted from the Lehrbuch der Speziellen Zoologie, Vol. 1, Part 5: Insecta. Copyright Spektrum Akademischer Verlag, Heidelberg, Berlin. Parts B–D reprinted with permission from Klass *et al.* 2002, *Science* **296,** 1456–1459. Copyright 2002 American Association for the Advancement of Science.)

PHYLOGENETIC RELATIONSHIPS

Mantophasmatodeans superficially resemble insects of the other "orthopteroid" orders. However, they lack the apomorphies (i.e., derived characters) of these, such as prognathous head, prothoracic repellent glands, and elongated female subgenital plate of Phasmatodea; the perforated tentorium, female subgenital lobe from coxosternum VII, and fused abdominal ganglia VII and VIIIff of Dictyoptera (mantises, cockroaches, and termites); the prognathous head, membranous sac on abdominal segment I, and loss of muscled closing devices of abdominal spiracles found in Grylloblattodea; and the pronotum overfolding the prothoracic pleura, the thickened hind femora, and the anterior intervalvula in the ovipositor found in Orthoptera.

Mantophasmatodea are assigned to the Pterygota by their lack of a noncuticular trunk endoskeleton and of the lateral parts of abdominal tergum XI, and to the Pterygota–Neoptera by the valvelike gonoplacs lacking styli and the presence of apodemes on the abdominal spiracles. Otherwise, their phylogenetic position is unclear. Phasmida and Grylloblattodea are the most promising candidates as the sister group of Mantophasmatodea.

HABITATS AND LIFE HISTORY FEATURES

Mantophasmatodea prey on other insects, which they catch by means of their strong and usually spinose fore- and midlegs. They are found in relatively dry and stony habitats, where occasional heavy rain leads to formation of temporary pools of water. The structure of the egg chorion suggests that the egg can overcome temporary flooding by plastron respiration.

See Also the Following Articles

Grylloblattodea • Insecta, Overview • Mantodea • Phasmida

Further Reading

Klass, K.-D., Zompro, O., Kristensen, N. P., and Adis, J. (2002). Mantophasmatodea: A new insect order with extant members in the Afrotropics. *Science* **296,** 1456–1459.

Zompro, O. (2001). The Phasmatodea and *Raptophasma* n. gen., Orthoptera incertae sedis, in Baltic amber (Insecta: Orthoptera). *Mitt. Geol. Palaeontol. Inst. Univ. Hamburg* **85,** 229–261.

Zompro, O., Adis, J. and Weitschat, W. (2002). A review of the order Mantophasmatodea. *Zool. Anzeiger.* **241.**

Marine Insects

Lanna Cheng
Scripps Institution of Oceanography

Although insects are undoubtedly the most common animals on land, very few species appear to live in the sea. However, they are actually rather well represented in diverse coastal marine or saline habitats. A marine insect is any insect that spends at least part of its life cycle in the marine environment, which includes any habitat from the upper intertidal to the open ocean. Among the 30 or so recognized insect orders, marine members occur in more than one-half of them (Table I). The most important species are found in Collembola, Heteroptera, Homoptera, Coleoptera, and Diptera. There are also many species of Mallophaga (biting lice) and Anoplura (sucking lice) whose hosts live in or on the sea (mammals or seabirds). A Web page on marine insects, www.unk.edu/marineinsects, is available.

HABITATS

Marine habitats can be divided either by salinity or by their position relative to the tidal level. Three types of saline habitats are generally accepted, based on their salt content (in parts per thousand): brackish (0.5–32), sea (34–37), and inland saline (0.5–250). Marine biologists, on the other hand, have traditionally divided coastal habitats into various zones according to their coverage by seawater or exposure to the sun. Three major zones are recognized: supralittoral (covered only during highest spring tides), littoral or intertidal (covered regularly between high and low tides), and sublittoral (never exposed even during the lowest low tides). An additional important habitat for marine insects is the pelagic zone, which comprises the open ocean far from the shore.

The majority of marine insects occur in the intertidal zones, which can be further categorized by the types of vegetation associated with them, e.g., seagrasses and rushes *(Spartina, Juncus),* seaweeds (green, blue-green, brown, or red), mangroves *(Rhizophora, Avicennia, Bruguiera, Sonneratia),* or other higher plants *(Xylocarpa, Acanthus).* The salinity of water in the various intertidal habitats tends to be variable or brackish. Larvae of several marine chironomids (Diptera) live among submerged vegetation in the sublittoral zone, which may include various green plants *(Enhalus, Halophila, Halodule)* and algae *(Halimeda, Corallina).*

The occurrences of various insect orders in different marine habitats are given in Table I. Five habitat categories are used in this table: pelagic, coastal, intertidal, mangrove, and saltmarsh. Brackish water habitats are commonly associated with mangroves in the tropics but with saltmarshes in temperate regions.

TABLE I Occurrences of Insects in Marine Environments by Taxonomic Grouping and Habitats

		Habitat				
Taxonomic group	**Common name**	**P**	**C**	**I**	**M**	**S**
Subclass Apterygota						
Protura	proturans	–	–	–	–	–
Collembola	springtails	–	+	+	+	+
Diplura	diplurans	–	–	–	–	–
Microcoryphia	jumping bristletails	–	–	+	+	+
Thysanura	bristletails	–	–	+	+	+
Subclass Pterygota						
Ephemeroptera	mayflies	–	–	–	–	+
Odonata	dragonflies and damselflies	–	–	+	+	+
Blattodea	cockroaches	–	–	–	–	–
Isoptera	termites	–	–	–	+	+
Mantodea	mantids	–	–	–	–	–
Grylloblattodea	icebugs	–	–	–	–	–
Phasmatodea	stick insects (walkingsticks)	–	–	–	+	–
Orthoptera	grasshoppers, crickets, etc.	–	–	+	+	+
Dermaptera	earwigs	–	–	+	–	–
Embiidina	webspinners	–	–	–	–	–
Zoraptera	zorapterans	–	–	–	–	–
Plecoptera	stoneflies	–	–	–	–	–
Psocoptera	psocids, booklice	–	–	–	–	+
Mallophaga*	chewing lice	+	+	+	+	+
Anoplura*	sucking lice	+	+	+	+	+
Thysanoptera	thrips	–	–	–	–	+
Heteroptera	true bugs	+	+	+	+	+
Homoptera	cicadas, aphids, etc.	–	–	+	+	+
Megaloptera	alderflies, etc.	–	–	–	–	–
Raphidioptera	snakeflies	–	–	–	–	–
Neuroptera	lacewings, etc.	–	–	–	–	+
Coleoptera	beetles	–	–	+	+	+
Strepsiptera*	strepsipterans	–	–	–	–	+
Mecoptera	scorpionflies	–	–	–	–	–
Trichoptera	caddisflies	–	–	+	+	+
Lepidoptera	butterflies and moths	–	–	–	+	+
Diptera	flies	–	+	+	+	+
Siphonaptera	fleas	–	–	–	–	–
Hymenoptera	bees, wasps, and ants	–	–	+	+	+

Note. P, pelagic; C, coastal; I, intertidal; M, mangrove; S, saltmarsh; +, present; –, absent or no data; *, only habitats of hosts marine. From Cheng and Frank (1993).

TAXONOMIC GROUPS

Apterygota

Among the five known orders, marine members are found only in Collembola, Microcoryphia, and Thysanura. There are few marine thysanurans. The genus *Petrobius* has several marine species living on rocky shores in Europe, whereas species of *Neomachilis* can be found living under rock or in crevices in the upper intertidal in California and probably elsewhere in North America. Some species in at least six families of Collembola (Onychiuridae, Hypogastruridae, Neanuridae, Isotomidae, Entomobryidae, and Acraletidae) live in various intertidal habitats. The best studied and most widely distributed is *Anurida maritima* (Neanuridae), commonly

found in rocky upper intertidal zones or saltmarshes. It lives in crevices among rocks and comes out to feed at low tide when the habitat is exposed. It is able to withstand submergence under seawater during high tide for periods up to 4 h by surrounding itself with an air bubble that acts as a compressible gas gill. Its orientation rhythm is endogenous and synchronized with the tides. These animals are usually negatively phototactic (i.e., going away from light), but between 2 and 7 h after low tide a large proportion of the population becomes positively phototactic (i.e., going toward light). Brightness of the horizon appears to be the main cue for the animals to move toward higher ground, where they seek shelter after foraging at low tide. When population densities become too high for any given crevice, these collembolans emerge and allow themselves to be dispersed at high tide by currents. They may then become stranded on the beach. Not infrequently huge aggregations consisting of millions of collembolans can be seen on beaches in various parts of the world.

Heteroptera

This is one of the most important orders in the marine environment, with nine families represented. Four families are semiaquatic and live at the water surface, five are shore dwellers, and only one, the Corixidae, is truly aquatic.

FAMILY GERRIDAE This is by far the most conspicuous and diverse family in the marine environment, with three subfamilies (Halobatinae, Trepobatinae, and Rhagadotarsinae), five genera (*Asclepios, Halobates, Stenobates, Rheumatometroides,* and *Rheumatobates*), and over 60 species. They can be found in various habitats ranging from near-shore mangrove streams, intertidal reefs, coastal lagoons, bays and estuaries, to the open ocean. The majority of the species are found in the Indo-Pacific region. The best studied genus in terms of taxonomy, distribution, ecology, phylogeny, and evolution is *Halobates* (Fig. 1). This genus is almost exclusively marine and contains 45 described species, including 5 that live a wholly pelagic life on the ocean surface. Although no living *Halobates* are known from the Mediterranean, a fossil species was described from an Eocene deposit (45 mya) in Verona, Italy.

FAMILY VELIIDAE This is also represented in the marine environment by three subfamilies (Rhagoveliinae, Microveliinae, and Haloveliinae), five genera, and more than 50 species. The genera *Trochopus* (5 species) and *Husseyella* (3 species) are confined to coastal bays, mangroves, and estuaries of the Americas, whereas *Xenobates, Halovelia,* and *Haloveloides* are know only from the Indo-Pacific. The latter three genera consist of small bugs, with adults measuring not more than 3 mm in body length. *Xenobates* spp., common among mangrove plants, are often overlooked, but there are at least 16 species. *Halovelia* is the largest genus of the Veliidae, with more than 35 species. These tiny bugs live in crevices among intertidal rocks and corals on tropical seashores and emerge to feed only at low tide. Although extant *Halovelia* species are confined to the Indo-Pacific, a fossil species was recently described from an Oligo-Miocene Dominican amber (15–45 mya).

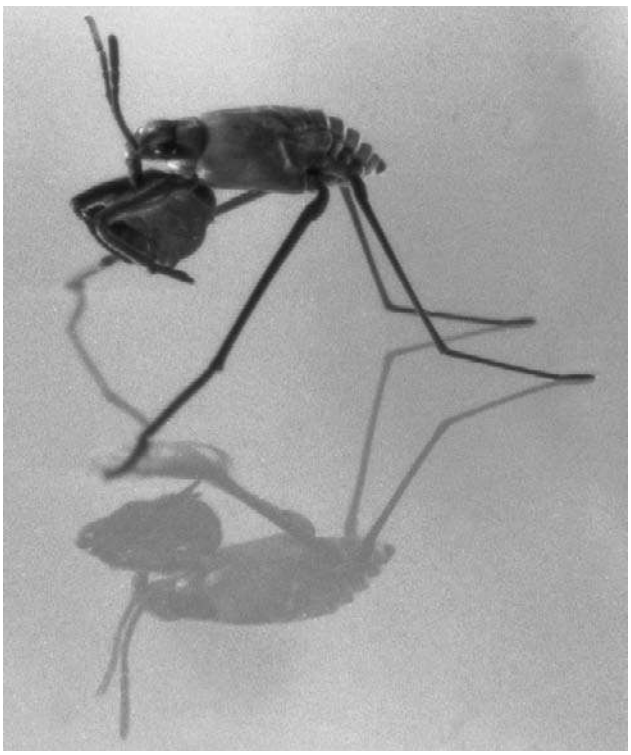

FIGURE 1 *H. sericeus* female (Heteroptera: Gerridae) feeding on *Drosophila* at the sea surface.

FAMILY MESOVELIIDAE Only two genera are represented in the marine environment. While most species of *Mesovelia* are found in freshwater habitats, at least one is halophilous and regularly found in tidal water. However, both known species of *Speovelia* are marine.

FAMILY HERMATOBATIDAE This is a monotypic, exclusively marine family with eight recognized species, all except one being found in the Indo-Pacific. These unique bugs are associated exclusively with intertidal rocks or coral rubble, where they hide in crevices during high tide and come out to feed at low tide. However, some must remain active at the sea surface because individuals have been caught in net tows offshore or with light during the night at high tide.

FAMILY CORIXIDAE This is the only truly aquatic heteropteran family to be found in saline environments. Although about 60 species belonging to 12 genera have been reported from saline waters, most of them normally occur in freshwater habitats but are able to tolerate saline conditions. Many species are common in inland saline lakes, but only a few are found there predominantly. *Trichocorixa verticalis,* the most widely distributed species, is commonly found in saltmarsh pools, but

specimens have occasionally been captured in plankton tows near shore. It is most tolerant of salinity changes and has been found to live and even to breed in ponds with a salinity approaching saturation (>300 ppt). Unlike most marine gerrids, which are wingless and flightless, adult corixids are winged and often disperse by flight.

SALDIDAE AND OTHER SHORE BUGS Five families of shore bugs (Saldidae, Gelastocoridae, Ochteridae, Omaniidae, and Aphelocheiridae) have representatives in marine environments. The most important family is Saldidae, which has at least 15 genera with more than 50 marine species worldwide. Most live in salt marshes. Winged forms occur in most shore bugs, and some species are rather strong fliers.

Homoptera

The Homoptera are rather poorly represented in the marine environment. In the Aphididae the best studied are *Pemphigus* spp. which feed on the roots of the saltmarsh *Aster* in Britain and probably elsewhere in Europe. Several species of herbivorous homopterans (Delphacidae, Issidae, Cicadellidae, and Cicadidae) feed and breed in seagrass beds in tidal saltmarshes. *Prokelisia marginata* (Delphacidae) is confined to beds of the seagrass *Spartina alterniflora* along the Atlantic coast of North America.

Coleoptera

Representatives of more than 20 families of beetles are found in marine environments, but none are truly aquatic. Most of the species occur in the intertidal zones among sand, rocks, algae, or wrack. Some are found in brackish waters, saltmarshes, or sand dunes. The most important marine families are Staphylinidae, Carabidae, Curculionidae, and Tenebrionidae.

FAMILY STAPHYLINIDAE This is the most important family, with more than 300 marine species. The genus *Cafius* is exclusively marine, with about 50 species. They are generally found on beaches associated with wrack (piles of stranded seaweeds), where they feed (as predators) and breed. The genus *Bledius,* with well over 400 species, has about 10% occupying marine habitats. Unlike most other staphylinids, they are not predatory. Females lay eggs in burrows and guard them from fungal attack and predation (a presocial behavior). The narrow openings of burrows of *B. spectabilis* have been shown to prevent flooding by tides. Adults and larvae of some other *Bledius* species dwell in inland salt flats. *Bryothinusa,* a genus of at least 24 species of small beetles, is exclusively marine.

OTHER FAMILIES Tiger beetles (Carabidae, subfamily Cicindelinae) include about 2300 species that occupy mainly terrestrial habitats. Adults of a few *Cicindela* species are common on sandy seashores and invade the intertidal areas during low tide. These predatory beetles run very fast and are difficult to catch. Females lay eggs in burrows where the larvae, after hatching, may remain for 2 or more years in the larval stage. Females of certain *Cicendela* species have been found to choose soils with specific salinity or shade conditions for laying eggs. Some genera of Curculionidae are known exclusively from driftwood or stranded seaweed on beaches, where they breed and the larvae develop. Beetles in several other families (Chrysomelidae, Cerambycidae, Curculionidae) are pests of tropical mangrove trees, feeding on flower buds, leaves, or bark and in some instances causing considerable damage.

Trichoptera

Caddisflies are predominantly freshwater insects, but several families breed in brackish water. The Chathamidae are exclusively marine and are known only from New Zealand and Australia. The four species belong to two genera, *Chathamia* and *Philanisus.* The adults are winged and can be found flying among intertidal rock pools. Eggs of the most widely distributed species, *P. plebeius,* are sometimes laid in the coelomic cavity of intertidal starfish or among coralline algal turf. The larvae use bits of coralline algae to construct their tubes (Fig. 2). They feed on various intertidal algae that may remain submerged at high tide.

Diptera

In addition to saltmarsh mosquitoes (Culicidae), biting midges (Ceratopogonidae), horse flies, and deer flies (Tabanidae), some of which are of great medical and/or economic importance, many other dipterans are found in various saline habitats. Almost all adult flies are winged, but the larvae of many species are truly aquatic and may remain submerged throughout their entire larval lives. The most commonly encountered nonbiting beach insects are probably seaweed flies belonging to at least five families (Coelopidae, Dryomyzidae, Muscidae, Borboridae, and Anthomyiidae). They are all associated with wrack or cast seaweeds where the adults feed and breed, and their life cycles tend to be synchronized with the tidal rhythm. Members of the

FIGURE 2 *P. plebeius* larvae in tubes constructed with coralline alga (Trichoptera: Chathamidae).

FIGURE 3 *Po. cottoni* male (Diptera: Chironomidae).

predatory family Canaceidae are exclusively marine and occur in the intertidal zone. Some of the most common invertebrates of inland salt lakes are brine flies in the family Ephydridae. Millions of these flies can sometimes be seen clustering along the lake shores, where they provide abundant food for shore birds.

Nonbiting flies in the families Chironomidae, Dolichopodidae, and Tipulidae are often associated with intertidal algal turf or submerged marine plants, where the larvae live and feed. The best studied is undoubtedly the chironomid *Clunio marinus,* whose life cycle is controlled by lunar as well as circadian rhythms. The most curious chironomid is probably *Pontomyia* (Fig. 3), an exclusively marine genus with four species. *Pontomyia* has an extremely short adult life span (30 min to 3 h), dying shortly after mating or egg laying. The eggs, embedded in a gelatinous matrix, are laid in a coil that sinks to the sea bottom. The larvae feed on submerged marine algae, and the pupae float to the sea surface shortly before emergence. The timing of emergence may be controlled by light, lunar, and tidal cycles.

Other Orders

Caterpillars of many species of moths feed on the leaves of mangrove trees, but their biology is poorly known. Other insect orders are represented in the marine environments by only a few species, e.g., the earwig *Anisolabis littorea* (Dermaptera) in New Zealand and the larvae of two sisyrids (Neuroptera) that live in brackish-water sponges. In addition, certain species of grasshopper (Orthoptera), ant (Hymenoptera), and even termite (Isoptera) can be common in some saltmarsh or mangrove areas.

See Also the Following Articles

Aquatic Habitats

Further Reading

Cheng, L. (ed.) (1976). "Marine Insects." North-Holland, Amsterdam.

Cheng, L. (1989). Factors limiting the distribution of *Halobates* species. *In* "Reproduction Genetics and Distributions of Marine Organisms" (J. S. Ryland and P. A. Tyler, eds.), pp. 357–362. Olsen & Olsen, Fredenborg, Denmark.

Cheng, L., and Frank, J. H. (1993). Marine insects and their reproduction. *Oceanogr. Mar. Biol. Annu. Rev.* **31,** 479–506.

Christiansen, K., and Bellinger, P. (1988). Marine littoral Collembola of north and central America. *Bull. Mar. Sci.* **42,** 215–245.

Denno, R. F., and Peterson, M. A. (2000). Caught between the devil and the deep blue sea, mobile planthoppers elude natural enemies and deteriorating host plants. *Am. Entomol.* **46,** 95–109.

Evans, W. G. (1980). Insecta, Chilopoda, and Arachnida: Insects and allies. *In* "Intertidal Invertebrates of California" (R. H. Morris, D. P. Abbott, and E. C. Haderlie, eds.), pp. 641–658. Stanford University Press, Stanford, CA.

Foster, W. A., and Benton, T. G. (1992). Sex ratio, local mate competition and mating behaviour in the aphid *Pemphigus-Spyrothecae. Behav. Ecol. Sociobiol.* **30,** 297–230.

Hoback, W. W., Golick, D. A., Svatos, T. M., Spomer, S. M., and Higley, L. G. (2000). Salinity and shade preferences result in ovipositional differences between sympatric tiger beetle species. *Ecol. Entomol.* **25,** 180–87.

Hogarth, P. J. (1999). "The Biology of Mangroves." Oxford University Press, Oxford.

Soong, K., Chen, G. F., and Cao, J. R. (1999). Life history studies of the flightless marine midges *Pontomyia* spp. (Diptera: Chironomidae). *Zool. Stud.* **38,** 466–473.

Zinkler, D., Rüssbeck, R., Biefang, M., and Baumgärtl, H. (1999). Intertidal respiration of *Anurida maritima* (Collembola: Neanuridae). *Eur. J. Entomol.* **96,** 205–209.

Mating Behaviors

Darryl T. Gwynne

University of Toronto

Mating behavior is typically viewed as comprising all events from pair formation through courtship to the final breakup of the mating pair. In most pterygote insects, sperm transfer is achieved through copulation. In contrast, in the few studied apterygotes, including both Insecta (Archaeognatha and Thysanura) and Ellipura (Collembola, Protura, and Diplura), sperm transfer is indirect; the spermatophore is placed on the substrate and is picked up by the female either following a period of courtship or with the pair making no contact at all. This article focuses on events occurring after the male and female have made physical contact; pair formation in insects is covered elsewhere. The main theme here is the function and adaptive significance of mating behaviors.

There is a vast amount of published information on the mating behaviors of insects. These behaviors have traditionally been viewed as relatively invariant within species. However, it is now evident that insect mating can show a great deal of adaptive variation and flexibility. As an introduction to this variation, consider insects that use carrion, a resource that can attract both males and females and thus serves as a location for mating. The complexity and plasticity of mating behavior observed in carrion insects easily rival those of other animals, including vertebrates (even Shakespeare's Romeo felt that "more courtship lives in carrion-flies than Romeo").

One courting carrion fly is attracted to the dry hide and bones of large old carrion sources, the main larval food for the

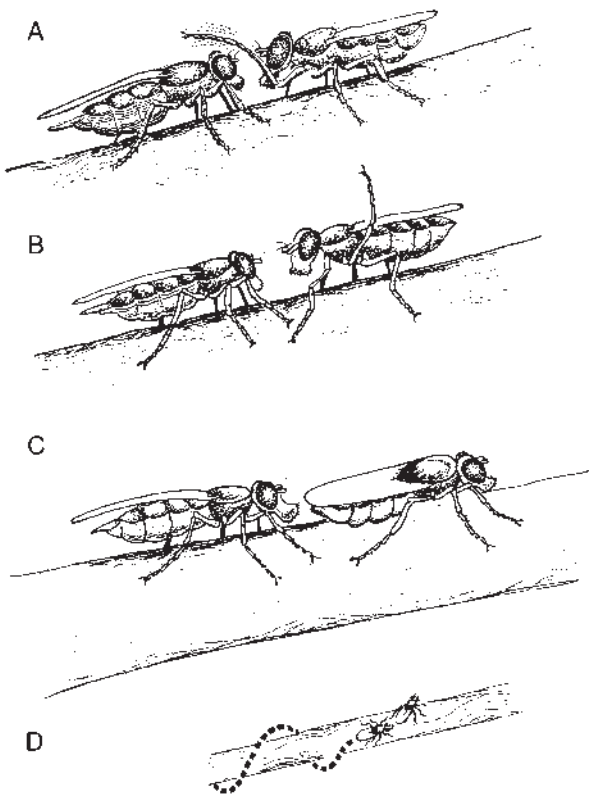

FIGURE 1 A fly similar to piophilids in exhibiting complex male courtship is the otitid, *Physiphora demandata.* The male first taps the female with a foreleg (A), then raises a middle leg (B), and turns and presents his abdomen to the female, who extends her proboscis to touch his abdomen (C). This can be followed by the female backing up in a spiral path, appearing to pull the male backwards (D). (Reproduced, with permission, from Alcock and Pyle (1979) *Z. Tierpsychol.* **49,** 354.)

species. The fly *Prochyliza xanthostoma* (Piophilidae) shows a remarkably complex courtship that can last for over 15 min (see Fig. 1 for another example). A courting male approaches the female while stepping rapidly from side to side and striking his abdomen downward. Males can repeat these vigorous movements and vary in the degree to which they display. If the female stops moving, the male stops courting, orients, and then slowly creeps toward her, occasionally repeating earlier parts of his routine. Individual males vary greatly in the vigor and length of the courtship display and this may represent variation in a signal of male quality used by females to select the best mates (Table I, Nos. 5 and 6).

Courtship by males of another temperate-zone carrion fly also appears to mediate female discrimination among males. Females of the fly *Dryomyza anilis* (Dryomyzidae) lay eggs on small carrion items such as dead fish. For this fly, courtship on the carrion occurs *after* copulation. A single courtship sequence consists of the male's genital claspers tapping vigorously on the female's external genitalia and then lifting and releasing her abdomen. Males vary in the number of genital tapping sequences performed and the number of sequences correlates with greater fertilization success. Bouts of tapping are followed by oviposition during which the male guards his mate from rivals. A male's success in fertilization is apparently achieved by the female biasing the distribution of sperm within her sperm storage organs (see Box 1). A similar influence on the success of courting male red flour beetles (*Tribolium castaneum:* Tenebrionidae) comes from a display in which the male rubs the female's elytra.

In a beetle that buries carrion, considerable variation occurs in reproductive behavior after the sexes have paired up and mated.

TABLE I The Functions and Context of Mating Behaviors

	Before copulating	While inseminating (copulating)	After insemination and copulation
(1) To communicate information about sex (gender), possibly to suppress aggressive (in males) or cannibalistic tendencies (in predatory species).	X		
(2) To synchronize mating behavior, such as when physiological mechanisms synchronize the behavior of the sexes.	X	X	
(3) To perform movements associated with positioning of genitalia, transferring ejaculates, and uncoupling.	X	X	X
(4) To communicate species information to prevent costly interactions (e.g., mate-finding movements or inviable offspring) with the wrong species.	X	X	
(5) To communicate information about direct benefits (for mates or offspring) supplied during or after mating such as:			
(a) fecundity or number of ejaculated sperm (fertility) and the ability to supply nutrients (nuptial meals) and	X	X	
(b) territory quality, or level of parental care.	X	X	X
(6) To communicate information about indirect benefits (i.e., for offspring) such as compatibility of genotypes or genetic quality.	X	X	X
(7) To communicate competitive ability to rivals.	X	X	X
(8) To resolve struggles between the sexes that reflect conflict over whether to mate at all, when to terminate copulation, or whether the partner mates with another individual.	X	X	X

Box 1. A Broader View of Courtship: The Concept of "Cryptic" Sexual Selection

If males behave in order to maximize the number of surviving offspring, then male success is best estimated as fertilization success rather than success in mating many females. Thus, both direct competition between males and female discrimination should not end at copulation. The full development of this insight has coincided with the advent of a number of molecular-genetics methods to assign paternity. In insects the potential for paternity competition is high because females typically mate with more than one male and store their long-lived sperm in specialized organs. Indeed, males have been found to possess adaptations that incapacitate, physically displace, or remove rival ejaculates, examples being copulatory movements in dragonflies, such as *Calopteryx maculata,* in which penis brushes remove virtually all rival sperm from the female sperm storage organs. In fact, male insects can enhance fertilization sucess even after sperm transfer has occurred. For example, males are known to transfer chastity-enforcing chemicals to females or substances that cause the females to increase the rate of laying eggs. Ultimately, however, mechanisms that bias fertilization success are under female control and so are probably best viewed as female discrimination, a phenomenon that can be revealed by experimentally removing "male-control" effects on the paternity of offspring. Female adaptations include fertilization biases caused by moving favored ejaculates. Also, in species with "last ejaculate stored is the first to be used" mechanisms, females mating high-quality males (representing genetic quality in species with no paternal care, Table I, No. 6) can simply increase the rate of egg laying (e.g., *Oecanthus* tree crickets and *Hylobittacus* scorpionflies) or differentially allocate more resources to these eggs, thereby increasing offspring fitness (no examples from insects, but this is known in birds). Finally, the consequences of these "cryptic" sexual selection mechanisms are: (i) a male's courtship that signals his quality, e.g., genital copulatory displays (Table I, Nos. 5 and 6), can occur at any point during mating until the female oviposits (thus making courtship synchronous with mating) and (ii) the Darwinian division between primary (e.g., penes and testes) and secondary (e.g., the peacock's tail) sexual structures is blurred: male structures such as dragonfly penis brushes and large testes that are adapted to deliver large numbers of gametes into the sperm competition lottery are probably sexually selected devices. The vast diversity and complexity of insect genitalia may result from these processes: species with multiple-mating females are known to have more complex male genitalia than species in which females mate only once.

A male and female of the beetle *Nicrophorus defodiens* (Silphidae) cooperate both to defend a mouse-sized carcass from intrusions by other *Nicrophorus* and to bury the carrion in an underground chamber where it becomes food for the pair's offspring. However, the behavior of the sexes is quite different after interment of a rat-sized carcass, one large enough to support more offspring than can be produced by the initial pair. Here, conflict between the pair becomes evident when the male produces a pheromone to attract additional females. The male's signal causes his resident mate to try and thwart this signaling by mounting and biting the male (Table I, No. 8). Conflict stems from a potential sexual difference in success on larger carrion. On this food resource, the male stands to gain substantially from the increased number of larvae he fathers when mating several females, whereas his first mate can expect only decreased fitness owing to increased larval competition.

Sexual conflict and flexibility in mating behavior is also apparent in the postmating interactions of a neotropical rove beetle, *Leistotrophus versicolor* (Staphylinidae). Male and female *L. versicolor* are attracted to carrion (and occasionally dung) not as an oviposition site but as a place to prey on flies. After a pair copulates, a male can be observed to attack and bite his mate, often running after her for up to half a meter. However, this behavior occurs only when there are a number of rival males present. Male aggression appears to serve in driving the female away from the carrion, thus preventing her from mating with other males (Table I, No. 8). An alternative possibility is that aggession toward females is a form of postcopulatory courtship (see Box 1 and Table I, No. 6).

SEXUAL CONFLICT DURING MATING

These episodes of insect mating reveal how Darwinian selection theory can be used to understand variation in behavior. The basic underlying assumption of this theory is that individuals behave in such a way as to yield the greatest number of surviving progeny. This theoretical insight suggests that courtship and copulation should rarely be a cooperative venture. Cooperation in courtship was the prevailing view among biologists at one time, in part because courtship was thought to synchronize mating events (Table I, No. 2). In contrast to this view, much research indicates that the sexes are often in conflict. Sexual conflict is expected to be common because the reproductive interests of male and female are so often at odds (Table I, No. 8). Conversely, cooperation is expected in the few cases in which male and female interests are similar. For example, interactions between a pair of burying beetles are mainly cooperative after they have interred a mouse-sized carcass. In contrast, when a larger carcass is buried, conflict is evident because, unlike the situation with small carcasses, the male has an opportunity to increase his reproductive success by attracting additional mates, whereas any added larvae from such matings probably decrease the initial female's success. The latter situation exemplifies a type of sexual

Box 2. Risk of Predation and Mating Behavior

Insects engaged in mating activities are known to assess the risks of predation and to adaptively change their behavior. For example, the typical song preference shown by female crickets, *Gryllus integer* (Gryllidae), can be overcome if the female can safely approach a less preferred song. In water striders, *Aquarius remigius* (Hemiptera: Gerridae), high predation risk appears to reduce male activity, thus decreasing their tendency to harrass females. This in turn allows large males to achieve high mating success (possibly because females can be more selective or can more easily avoid mating with smaller males). The threat of predation from fish, insect, and spider predators of this species and other gerrids can cause a decrease in mating frequency as well as in the duration of copulation.

FIGURE 2 Sexual conflict behavior in insects can be seen in mating struggles between a male and a female. Here a pair of scorpionflies, *Hylobittacus similis,* are using their hind tarsi in a tug-of-war over a nuptial prey (a blow fly) item captured by the male.

conflict expected in the reproductive activities of animals because of a sexual difference in reproductive strategy: males typically maximize the number of females mated so as to maximize fertilization success, whereas females maximize fecundity and offspring quality. This sexual difference also causes conflict when already-mated females are harrassed by promiscuously mating males. Examples come from the precopulatory struggles often observed in insects. A well-studied case involves water striders (Heteroptera: Gerridae). When a male uses forelegs and genitalia to secure a female for copulation, a vigorous struggle ensues during which the female attempts to dislodge him. Superfluous matings can be costly to females in terms of increased predation risk (Box 2) and energetic cost. To reduce such costs female *Gerris incognitus* have evolved upcurved abdominal spines that appear to function in thwarting male mating attempts. Another possible purpose for precopulatory struggles is that they test male quality (Table I, No. 6; see also the example of seaweed flies considered under Genetic Quality and Mate Choice).

Precopulatory struggles appear to be a result of sexual conflict in species in which males feed their mates, because females pay a cost if the size of their meal is reduced in any way. Thus a newly paired male and female scorpionfly (Mecoptera: Bittacidae) can both be seen to pull on the prey offering (Fig. 2). Conflict comes from males holding back the prey in order to conserve food for copulations with other mates and the female attempting to begin her meal as soon as possible. Conflict in some mate-feeding insects is particularly evident in the struggle between the sexes when a male attempts to force a copulation without providing the beneficial meal to his mate. To overcome female resistance, males of both panorpid scorpionflies and haglid orthopterans have specialized abdominal organs that function in holding onto females during forced matings.

A widespread form of sexual conflict arises from multiple mating by a female, which increases her success while compromising her mate's confidence of paternity. This conflict is evident in rove beetle males that drive their mates away from rivals. Striking examples also occur in male adaptations that not only enforce chastity in the female but also reduce her survival. Examples include toxic chemicals in the seminal fluid of fruit flies, *Drosophila melanogaster* (Drosophilidae), and damaging spines on the penis of lowpea weevils, *Callosobruchus maculatus* (Bruchidae), both of which decrease female life span. In the beetle, females appear to reduce injury to the reproductive tract by vigorously kicking males in order to terminate copulation (Table I, No. 8).

SEXUAL DIFFERENCES IN MATING BEHAVIOR

Advantages that females might obtain from choosing to mate with more than one male include acquiring goods and services—such as nuptial meals—or enhancing offspring quality by remating when a high-quality male is encountered (Box 1). This point highlights the basic sexual difference in mating behavior: typically females are choosy when it comes to the males that father their offspring, whereas males compete and display as a way to obtain multiple matings.

The factors controlling these typical sexual differences in behavior stem from the basic difference in the way males and females maximize reproductive success. Females usually invest more in individual offspring than males by providing materials for egg production and, in some species, caring for progeny. These maternal activities mean that fewer females than males are available for mating, thus causing males to compete for the limiting sex. Therefore sexual selection is greater on males than on females. This theory predicts that in species in which males invest more in offspring than females, sexual selection on the

sexes will be reversed, causing a reversal in the mating roles, i.e., competitive females and choosy males. This prediction has been upheld in experiments with several katydid species (Orthoptera: Tettigoniidae). These species are useful experimental organisms because mating roles are flexible; when food in the environment becomes scarce females compete for mates and males are choosy. Hungry females fight to obtain matings because each copulation comes with a nuptial meal, a nutritious spermatophore (in contrast to sexual selection on males—to increase fertilizations—the sexual selection on meal-seeking females is to increase number of matings). In support of the theory, food scarcity causes an increase in relative investment in individual offspring (eggs) because there is an increase in the material in eggs derived from males—their spermatophore nutrients. The degree of choosiness shown by a sex should also be influenced by variation in the quality of the sex being chosen. Members of a sex are expected to be choosy if variation in the quality of potential mating partners is high.

SIGNALS TO MATES DURING MATING

Material Benefits and Mate Choice

Rejection of a mate (usually of males by females) is only one of the explanations for the failure of a pair to mate successfully. Other causes of a breakup of pairs are certain changes in the physical environment and a threat of harm from predators (see Box 2) or rival males. A number of studies have ruled out these alternatives and have thus shown that certain mating behaviors function in choosing mates, for example, individuals noted to move between signaling or swarming members of the opposite sex before mating with one of them and individuals pulling away from their mates after the mating sequence has begun.

As predicted by theory, female choice of mates is more widespread than male choice. Some of the clearest examples of mate choice come from species in which females obtain material benefits from males (Table I, Nos. 5a and 5b). One case involves scorpionflies (family Bittacidae), of which females attracted to males presenting food gifts of prey will manipulate the offering (Fig. 2) and reject males presenting small prey. Female scorpionflies also discriminate against such males by breaking off copulation prematurely. Only a male that transfers a large prey will complete copulation by supplying both a full complement of sperm and chastity-inducing substances.

Virtually all cases of male choice in insects appear to involve the acquisition of material benefits from large females, specifically the large number or size of eggs possessed by heavyweights. This has been noted in a number of insects, including tettigoniid orthopterans, cerambycid and brentid beetles, and empidid flies. Selection on males to mate with the most fecund females should mean that preferred females would not display their genetic quality (Table I, No. 6) because investment in costly displays probably indicates reduced fecundity. In fact, fecundity selection on females predicts that they will usually not evolve in displays of quality in the first place. In what may be an exception to this prediction, female empidid flies, *Rhamphomyia longicauda,* display inflated abdomens and fringed legs to choosy males while flying in all-female swarms. The size of the inflated female, as perceived by a male entering the swarm, seems to be a poor predictor of her fecundity and may instead advertise her genetic quality.

As noted earlier, male choice is expected when the mating roles are reversed or when there is a high degree of variance in the quality of females. Examples of the former include male Mormon crickets (*Anabrus simplex:* Tettigoniidae) pulling away from mounted lightweight females—apparently after weighing them—and males of the empidid flies *Rhamphomyia sociabilis* and *Empis borealis* choosing large, fecund individuals from within all-female swarms. In contrast to male mate choice in role-reversed systems, male choice that evolves in response to a high variance in female quality typically is often found with a high degree of male–male competition, i.e., sexual selection on males. Indeed, male choice in this situation can be caused by sexual selection among males to mate with the highest quality females. An example of male choice when females vary in quality includes winter moths, *Operophtera brumata* (Lepidoptera, Geometridae), and red flour beetles. Finally, local population variation in the primary sex ratio can affect the likelihood of male choice; in red milkweed beetles, *Tetraopes tetraopthalmus* (Cerambycidae), a scarcity of males is associated with a higher degree of male choice.

Genetic Quality and Mate Choice

In theory, females are expected to show choice to obtain indirect benefits, i.e., benefits that enhance the genetic quality of offspring. Female *Dryomyza* flies appear to do this by biasing fertilization after evaluating male copulatory courtship. But what sorts of indirect benefits do choosing females obtain? In yellow dung flies, *Scathophaga stercoraria* (Scathophagidae), females can favor the stored sperm from males with genotypes likely to enhance offspring growth. For genotypes common in the population this involves mating with a male of a similar genotype when the environment is constant but choosing a male of a different genotype when the environment is unpredictable. In other insects, the cue to a male's genetic quality is consistent between females. For example, following a courtship consisting of wing-flicking and pheromone displays, older females of *Colias* butterflies (Pieridae) show mating preferences for genotypes that fly well and are long-lived. Male and female calopterygid dragonflies also court using wing displays and the size of wing spots. The latter is a sexually dimorphic trait in European *Calopteryx splendens* and is correlated with several aspects of male quality, including the level of immunocompetence, developmental stability, and resistance to gut parasites. In yet another dipteran, the seaweed fly, *Coelopa frigida* (Coelopidae), mating interactions involve prolonged premating struggles in which a mounted male can be dislodged as a result of kicking and shaking by the female. Such struggles favor matings with large males and this female bias enhances

the genetic fitness of her offspring: progeny of large males tend to be heterozygous for a chromosomal inversion that increases offspring viability. However, female choice in this system may be maintained by more than good-genes sexual selection. Females carrying the inversion genotype show a strong preference for large males. Genes for preference thus appear to be linked with genes for the male display trait, suggesting a form of female-choice sexual selection, termed "runaway" or "Fisherian" sexual selection (after the originator of this idea, R. L. Fisher), in which mothers gain by producing "sexy sons," those that are highly attractive to females.

Mating Preferences for the Correct Species

One result of the expected rapid evolutionary change from runaway sexual selection may be speciation through behavioral isolation. Speciation results when there is sufficient between-population divergence in the female preference and the linked male display that a side effect of intraspecific mating preferences is discrimination against males from other populations (see also Box 1). This "effect" hypothesis for species discrimination differs from the hypothesis that certain female mating preferences have evolved to *function* in avoiding costly interactions with other species. An example of the latter involves the fruit flies *Drosophila pseudoobscura* and *D. persimilis,* in which hybrid matings result in decreased reproductive success because sons are sterile. Female *Drosophila* assess the wing-vibration displays of males, and *D. pseudoobscura* females collected from areas where the two species co-occur (sympatry) reject courting *D. persimilis* males more frequently than females collected from areas with no species overlap (allopatry). This result was not the result of differences in courtship by the males with the two types of females. These findings indicate that female discrimination against heterospecific male courtship has been reinforced in areas where maladaptive hybridization is likely to occur.

A high degree of discrimination against courtship by heterospecific individuals in sympatry has also been noted in *Calopteryx* dragonflies in which both sexes display patterned wings in precopulatory courtship. Compared with areas of allopatry, male *C. maculata* in sympatry discriminate more against the wing patterns of *C. aequabilis* females, and mate preferences during courtship appear to have reinforced wing-pattern differences between the species. For example, in a north–south transect in eastern North America, the proportion of pigmented wing area of both sexes is greater in areas of sympatry than in areas of allopatry.

SIGNALS TO RIVALS DURING MATING

In another North American calopterygid, *Hetaerina americana,* variation in wing-spot displays reflects an evolutionary history of competition between males; males with larger wing spots are more successful in defending mating territories than males with smaller spots. As males with experimentally enhanced spots suffer a cost (increased mortality), these signal patterns appear to have evolved as honest indicators of fighting ability. These signal indicators may convey information about male ability to rivals during wing-waving displays directed at females. Other behaviors taking place during mating appear to function in a competitive context (Table I, No. 7). For example, males of the carrion beetle mentioned previously *(L. versicolor)* will occasionally mimic female behavior during interactions with potential mates (Fig. 3). Pseudofemale behavior reflects a remarkable plasticity in the way the male beetles obtain matings. The most profitable way to gain access to females is to fight to defend the carcass that attracts them. However, males can obtain some matings subversively by mimicking female behavior and thus avoiding costly fighting. This form of behavior is conditional on the relative size difference between opponents: a male will engage a smaller male in a fight but will switch to pseudofemale behavior if his rival is larger. There are other mating behaviors that appear to reflect male–male competitive interactions. For example, after attracting females, males of some singing insects switch from song to a more reclusive signal such as substrate vibration, apparently as a way of avoiding courtship behavior that attracts rivals.

HOMOSEXUAL BEHAVIOR AND MATING MISTAKES

A male *L. versicolor* beetle can be duped into courting a small female-mimicking rival (Fig. 3) and there are a few other examples, such as in some butterflies and dragonflies, of homosexual mating mistakes when certain males adaptively resemble females. Homosexual mounting can also occur among insects of which the males have not evolved to mimic females. This male behavior is widespread in animals and appears to be simply an effect of poor sex recognition; strong selection on males to mate frequently causes them to mount any object that resembles a female. Examples of mating mistakes can even include inanimate objects, such as in the case of *Julidomorpha bakewelli,* an Australian buprestid beetle, the males of which attempt to copulate with beer bottles with a coloration and reflection pattern resembling the female's elytra (Fig. 4).

Poor sex recognition appears to be the explanation of why males of another beetle, *Diaprepes abbreviatus,* mount conspecific males. A big difference between this species and others, however, is that females also perform homosexual mountings. In this case, however, mounting appears to be an adaptive reproductive strategy rather than a mating mistake. Laboratory experiments with this species reveal that a mounted pair of females attracts males. In fact, large males attempt to mate more often with paired females than with single large or small females. As both the mounting and the mounted females had similar probabilities of copulating with the attracted male, it appears that the mounted pair mimics a heterosexual pair in order to incite the attraction of large, competitive (i.e., high-quality) males.

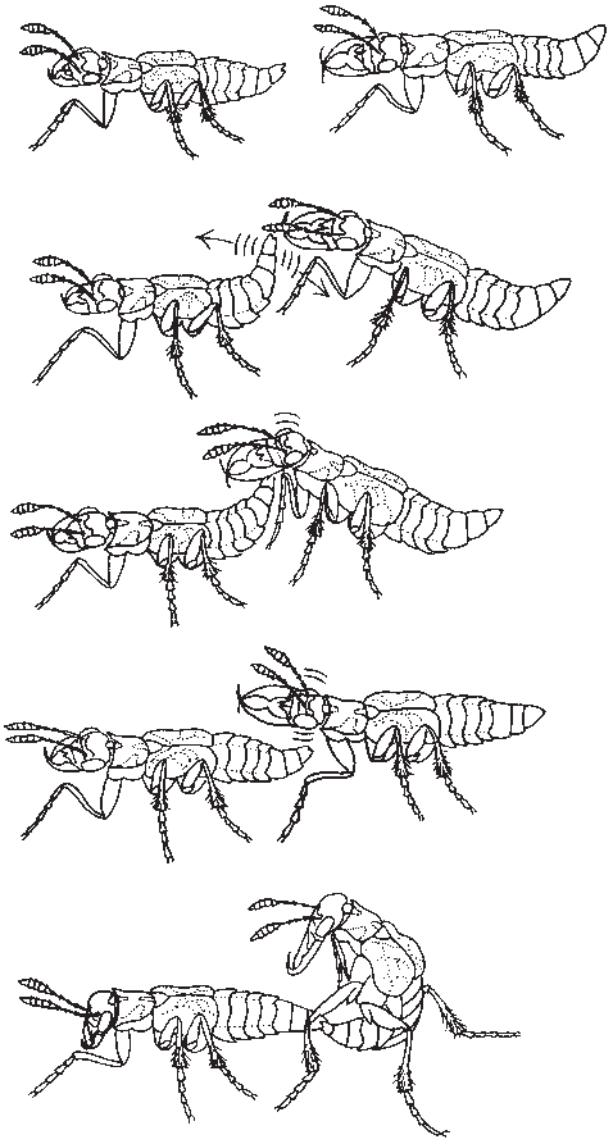

FIGURE 3 A male staphylinid beetle, *L. versicolor,* can avoid being chased by a rival from the carrion source by mimicking female behavior. The mimic male turns and presents his abdomen to the approaching rival, which antennates the abdominal tip and taps it with his head. Copulation (bottom) is the only stage of a heterosexual encounter that is not represented in these homosexual encounters (because the mimic male breaks up the encounter by walking away). (Reproduced, with permission, from Forsyth and Alcock (1990) *Behav. Ecol. Sociobiol.* **26,** 326.)

CONCLUSIONS

Observations of insect mating behaviors reveal a great diversity, some of which is a result of plasticity within species. The examples discussed here show how an understanding of the function of both inter- and intraspecific variation in mating behaviors can be gained by examining the consequences of behavior for the reproductive success of the mating male and female. Functions of insect mating that were proposed before the widespread use of the "selectionist" approach (Table I, Nos. 1 and 2) can be subsumed into this framework. For example, movements involved in delivering sperm are undoubtedly subject to sexual selection. And, any observations of synchronized courtship in a species inevitably lead to the question of how such synchrony enhances the reproductive success of the male and female. There is a wealth of behavioral diversity for future research, including apterygote insects, a virtually unstudied group and one of great interest because they lack copulation.

FIGURE 4 A mating error by a male: a *J. bakewelli* male mounts a beer bottle. Note how the aedeagus (penis) is extended.

See Also the Following Articles

Hearing • Pheromones • Reproduction • Sexual Selection

Further Reading

Alcock, J. A., and Gwynne, D. T. (1991). Evolution of insect mating systems: The impact of individual selectionist thinking. *In* "Reproductive Behaviour in Insects: Individuals and Populations" (W. J. Bailey and J. Ridsdill Smith, eds.). Chapman & Hall, London.

Arnqvist, G. (1998). Comparative evidence for the evolution of genitalia by sexual selection. *Nature* **393,** 784–786.

Cade, W. H. (1985). Insect mating and courtship behaviour. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.). Pergamon Press, Oxford.

Choe, J. C., and Crespi, B. (eds.) (1997). "The Evolution of Mating Systems in Insects and Arachnids." Cambridge University Press, Cambridge, UK.

Eberhard, W. G. (1996). "Female Control: Sexual Selection by Cryptic Female Choice," Princeton University Press, Princeton, NJ.

Eggert, A.-K., and Sakaluk, S. K. (1995). Female-coerced monogamy in burying beetles. *Behav. Ecol. Sociobiol.* **37,** 147–153.

Lima, S. L. (1998). Stress and decision making under the risk of predation: Recent developments from behavioral, reproductive, and ecological perspectives. *Adv. Stud. Behav.* **27,** 215–290.

Lloyd, J. E. (1979). Mating behavior and natural selection. *Fla. Entomol.* **62,** 17–34.

Rowe, L., Arnqvist, G., Sih, A., and Krupa, J. J. (1994). Sexual conflict and the evolutionary ecology of mating patterns—Water striders as a model system. *Trends Ecol. Evol.* **9,** 289–293.

Simmons, L. W. (2001). "Sperm Competition and Its Evolutionary Consequences in Insects." Princeton University Press, Princeton, NJ.

Thornhill, R., and Alcock, J. (1983). "The Evolution of Insect Mating Systems." Harvard University Press, Cambridge, MA.

Mayfly

see *Ephemeroptera*

Mechanoreception

Andrew S. French and Päivi H. Torkkeli
Dalhousie University, Halifax, Nova Scotia

Mechanoreception is the sense that allows insects to detect their external and internal mechanical environments, including physical orientation, acceleration, vibration, sound, and displacement. The integument and internal organs contain a wide variety of mechanoreceptors. Prominent receptors, such as surface hairs that mediate touch or auditory organs, have been studied extensively, but many other physiological functions also depend on mechanosensory signals.

Arthropod mechanoreceptors are divided into two morphological groups: Type I, or cuticular, and Type II, or multipolar. Type I are ciliated receptors, associated with the cuticle, and have their nerve cell bodies in the periphery, close to the sensory endings. They can be subdivided into three major groups (Fig. 1). Hairlike receptors are found on the outer surface in a variety of shapes and sizes, from long, thin hairs to short pegs and scales. A sensory neuron is closely apposed to the base of the hair and its dendrite contains microtubules ending in a structure called the tubular body. It is assumed that movement of the hair compresses the ending, with the tubular body perhaps providing a rigid structure against which the compression can work. Hair receptors can contain additional sensory neurons, such as chemoreceptor neurons in taste hairs. Campaniform (bell-shaped) sensilla are also found on the outer surface, particularly in compact groups near the joints, where they detect stress in the cuticle. Stress moves the bell inward, compressing the dendritic tip containing the tubular body. Chordotonal receptors are generally found farther beneath the integument, although they can be connected to the integument by attachment structures. They serve several functions, including hearing and joint movement detection. They generally lack tubular bodies but have dense scolopales surrounding the dendrites and often have multiple mechanosensory neurons.

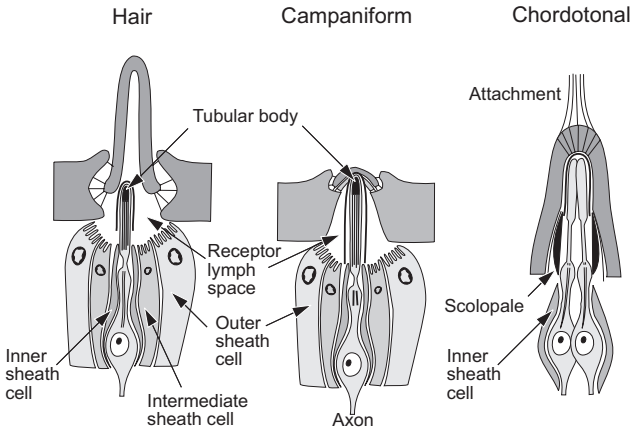

FIGURE 1 The three major groups of insect cuticular mechanoreceptors. The receptor lymph space surrounding the sensory ending is formed by a layer of sheath cells and epithelial cells connected by tight junctions. The numbers of sheath cells and their nomenclatures are both variable (see text).

Type II mechanoreceptors are nonciliated neurons, whose central cell bodies have many fine dendritic endings, each of which is apparently mechanosensitive, but lacks the detailed structures seen in Type I receptors. Type II receptors are found in many internal structures, predominantly associated with mesodermal tissues, including the musculature, where they detect muscle tension.

Studies of mechanoreceptor morphology have used many techniques, including light microscopy, scanning and transmission electron microscopy, and immunohistochemistry. Receptor electrophysiology has been studied by three basic methods: (1) Extracellular recordings observe the receptor currents flowing along the axon. (2) Epithelial recordings measure the current flowing through the relatively low resistance of the thin socket tissue or through a cut hair. (3) Intracellular recordings give direct measurements of membrane potentials and currents.

Mechanosensation is commonly viewed as a three-stage process in which a mechanical event is first coupled to the receptor cell membrane by mechanical structures, then transduced into a receptor current at the cell membrane, and finally encoded into action potentials for transmission of information to the central nervous system.

DEVELOPMENT OF MECHANORECEPTORS

Type I sensory neurons are surrounded by specialized sheath cells of varying numbers and names, although the terms trichogen (hair-forming) and tormogen (sheath-forming) are commonly used for the innermost two layers of sheath cells. Development of these cells has been well characterized in several species, but especially in *Drosophila* external bristles, for which many of the genes involved have been identified. A single sensory organ precursor cell divides to give two different secondary precursors, IIA and IIB. IIA divides to form one trichogen and one tormogen cell. IIB gives rise to the neuron, another sheath cell, and sometimes an additional glial cell. The neuron then forms an axon that grows into the central nervous system. A variety of other noncellular structures, including sheaths, are also found, particularly in dendritic regions. The development of Type II receptors is less well understood.

MECHANICAL COMPONENTS

Extracellular tissues, often with elaborate structures, surround the sensory cells. These structures modify the spatial and temporal sensitivities of the receptors, and they are often designed to interact with the outside environment or other parts of the animal, such as cercal hairs detecting air movements or hair plates detecting joint rotation. External structures usually allow detection of mechanical events at some distance from

the sensory cell but make the displacement at the receptor cell membrane smaller than the original movement. Estimates of this attenuation suggest that threshold movements at the cell membrane leading to sensation are in the range 1 to 5 nm.

TRANSDUCTION AND ENCODING

Mechanically activated ion channels, probably located in the tips of the sensory dendrites, transduce the mechanical stimulus into a receptor current. The channels are permeable to potassium ions and the receptor lymph spaces have high concentrations of potassium. The channels may have time-dependent properties that contribute to receptor behavior, but separating these from the time-dependent properties of any external mechanical structures is difficult. Current flowing through the channels causes a receptor potential that is encoded into action potentials. Mechanotransduction currents are more sensitive to temperature than most other membrane currents, with activation energy values of 12 to 22 kcal/mol, which are similar to those required to break chemical bonds and significantly higher than the energy barriers associated with ionic diffusion or conductance through ionic channels. Some crucial stage in the link between membrane tension and ion channel opening may lead to this high energetic barrier.

The receptor potential is encoded into action potentials using several different sodium and potassium currents. Action potentials propagate into the central nervous system along axons in nerve roots of the segmental ganglia. Afferent axons have a size range of 1 to 20 μm and conduction velocities are typically 1 to 5 m/s. Information is transmitted from mechanoreceptor axons into the central nervous system via cholinergic synapses.

CENTRAL, PERIPHERAL, AND HUMORAL MODULATION

Many mechanoreceptors receive GABAergic inhibitory efferent innervation close to the output synapses of their axon terminals. This presynaptic innervation modulates afferent mechanoreceptor information. Some mechanosensory neurons are also modulated by efferent innervation in the periphery and by circulating chemicals such as biogenic amines. Of these, octopamine has been most thoroughly studied, but without clear conclusions, because octopamine can increase or decrease firing frequency, even in the same neuron. Studies in locusts and cockroaches suggest that octopamine receptors are located on the peripheral regions of mechanoreceptors.

The extent and functions of peripheral modulation remain to be seen. It is the latest in a series of surprises about the complexity of insect mechanotransduction, but probably not the last.

See Also the Following Articles

Orientation • Vibrational Communication

Further Reading

Abdelilah-Seyfried, S., Chan, Y. M., Zeng, C., Justice, N. J., Younger-Shepherd, S., Sharp, L. E., Barbel, S., Meadows, S. A., Jan, L. Y., and Jan, Y. N. (2000). A gain-of-function screen for genes that affect the development of the *Drosophila* adult external sensory organ. *Genetics* **155,** 733–752.

Bräunig, P., and Eder, M. (1998). Locust dorsal unpaired median (DUM) neurones directly innervate and modulate hindleg proprioceptors. *J. Exp. Biol.* **201,** 3333–3338.

Burrows, M. (1996). "The Neurobiology of an Insect Brain." Oxford University Press, Oxford/New York/Tokyo.

Chapman, K. M., and Pankhurst, J. H. (1967). Conduction velocities and their temperature coefficients in sensory nerve fibres of cockroach legs. *J. Exp. Biol.* **46,** 63–84.

Field, L. H., and Matheson, T. (1998). Chordotonal organs of insects. *Adv. Insect Physiol.* **27,** 1–228.

French, A. S. (1988). Transduction mechanisms of mechanosensilla. *Annu. Rev. Entomol.* **33,** 39–58.

French, A. S. (1992). Mechanotransduction. *Annu. Rev. Physiol.* **54,** 135–152.

Höger, U., and French, A. S. (1999). Temperature sensitivity of transduction and action potential conduction in a spider mechanoreceptor. *Pflugers Arch.* **438,** 837–842.

McIver, S. B. (1985). Mechanoreception. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.). Pergamon Press, Oxford.

Ramirez, J. M., and Orchard, I. (1990). Octopaminergic modulation of the forewing stretch receptor in the locust, *Locusta migratoria. J. Exp. Biol.* **149,** 255–279.

Spinola, S. M., and Chapman, K. M. (1975). Proprioceptive indentation of the campaniform sensilla of cockroach legs. *J. Comp. Physiol.* **96,** 257.

Torkkeli, P. H., and French, A. S. (1995). Slowly inactivating outward currents in a cuticular mechanoreceptor neuron of the cockroach *(Periplaneta americana). J. Neurophysiol.* **74,** 1200–1211.

Zacharuk, R. Y. (1985). Antennae and sensilla. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.). Pergamon Press, Oxford.

Zhang, B. G., Torkkeli, P. H., and French, A. S. (1992). Octopamine selectively modifies the slow component of sensory adaptation in an insect mechanoreceptor. *Brain Res.* **591,** 351–355.

Zill, S. N., Ridgel, A. L., Dicaprio, R. A., and Frazier, S. F. (1999). Load signalling by cockroach trochanteral campaniform sensilla. *Brain Res.* **822,** 271–275.

Mecoptera

(Scorpionflies, Hangingflies)

George W. Byers

University of Kansas

Mecoptera (scorpionflies, hangingflies, and others) are holometabolous insects in which the head is characterized by a downward projecting rostrum or beak, at the end of which are chewing mouthparts. They are usually slender bodied and have four long, narrow wings that are membranous and often marked with dark bands, spots, or darkening along the numerous crossveins. There are exceptions, however. A few species are brachypterous or wingless; species in family Boreidae

have hardened, highly modified wings, which are nearly oval and scalelike in females, slender and somewhat curved apically in males. The rostrum is not unusually prolonged in species of Nannochoristidae; and the wings of species of Meropeidae are only about 2.5 times longer than their greatest width.

Although Mecoptera are one of the minor orders of insects, with only about 550 living species so far made known, they are of much interest to entomologists. This is largely because fossil Mecoptera are among the oldest remains of insects with complete metamorphosis (of Permian age) and have been regarded by some as ancestral to the more recently evolved and vastly larger orders Diptera and Lepidoptera. Fossil evidence suggests that the Mecoptera were once one of the larger orders of holometabolous insects. Past diversity, based mainly on wing venation, has led to placement of some 350 fossil species in 87 genera in 34 families. Thus, modern Mecoptera are survivors of millions of years of evolutionary development.

FAMILIES AND GENERA

Following are the nine families of extant Mecoptera, with the number of genera included in each and the approximate geographical distribution: Panorpidae—3 genera, in Europe, Asia, and North America; Panorpodidae—2 genera, in easternmost Asia, and North America; Bittacidae—17 genera, in North and South America, Africa, Asia, Australia, and Europe; Boreidae—3 genera, in North America, Asia, and Europe; Choristidae—3 genera, in Australia; Nannochoristidae—2 genera, in Australia, New Zealand, and southern South America; Apteropanorpidae—1 genus, in Australia (Tasmania); Meropeidae—2 genera, in North America and southwestern Australia; Eomeropidae (formerly Notiothaumidae)—1 genus, in South America.

Most Mecoptera belong to the families Panorpidae (scorpionflies) and Bittacidae (hangingflies). Scorpionflies are so called because abdominal segments 7 and 8 of the male are slender, and segment 9 is abruptly enlarged and often held above the back, recalling the sting of a scorpion (Fig. 1). The female abdomen tapers to a slender tip. Both male and female hangingflies are slender bodied and, having a single, large (raptorial) claw at the end of each tarsus, are unable to stand on the upper surfaces of leaves but suspend themselves from twigs or edges of leaves (Fig. 2).

ZOOGEOGRAPHY

Mecoptera present a variety of zoogeographical patterns, from highly localized (as in *Notiothauma,* which is endemic in central Chile, or *Apteropanorpa,* found only in Tasmania) to virtually cosmopolitan (as the Bittacidae, which occur in the temperate and tropical parts of six continents). Only Bittacidae have been found in Africa. Some very disjunct occurrences suggest relics of ancient, more widespread ranges (as *Merope* in eastern North America and the very similar *Austromerope* in southwestern Australia; or Nannochoristidae in southeastern Australia, New Zealand, and southern South America).

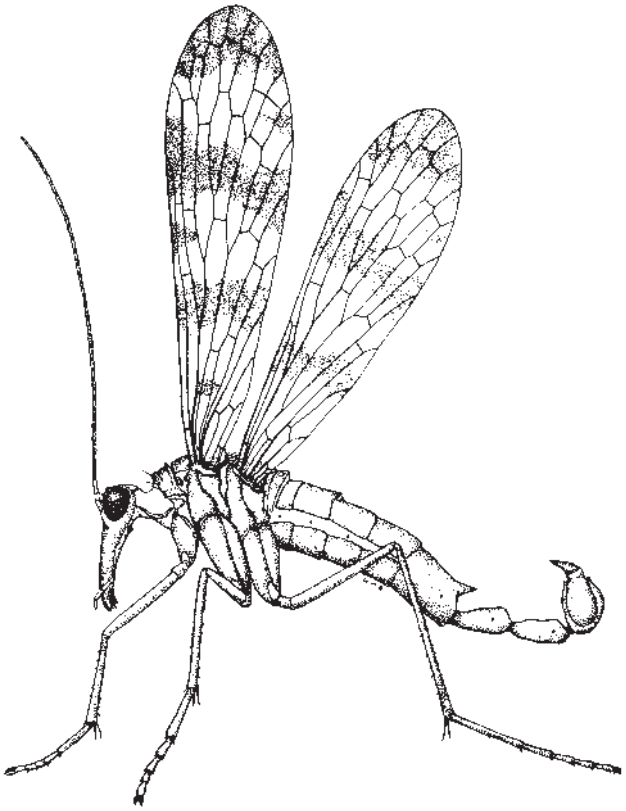

FIGURE 1 A male scorpionfly, showing anatomical characteristics of family Panorpidae. (Illustration courtesy of Holt, Rinehart and Winston, Inc.)

HABITAT

Broad-leaved, herbaceous plants shaded by trees are the usual habitat of panorpids, bittacids, and most species in the smaller families. Some species, however, are most commonly found in vegetation near forest borders, whereas others occur in somewhat similar but more uniformly shaded habitats well within a forest. Elevation clearly affects the distribution of Mecoptera. Boreidae, for example, reach the adult stage in the cold part of the year (and being darkly colored are most often seen on the surface of snow) and at high elevations and high latitudes have a longer seasonal occurrence than boreids at lower elevations. In Mexico, Bittacidae are usually found at elevations below 1500 m, while Panorpidae occur above that level and up to over 3000 m.

DIET

Both adults and larvae of *Panorpa* are scavengers, feeding usually on dead insects, and less often on other dead organisms, including some small vertebrates. Adults occasionally eat pollen and associated parts of flowers; they may even invade the webs of spiders to feed on entangled

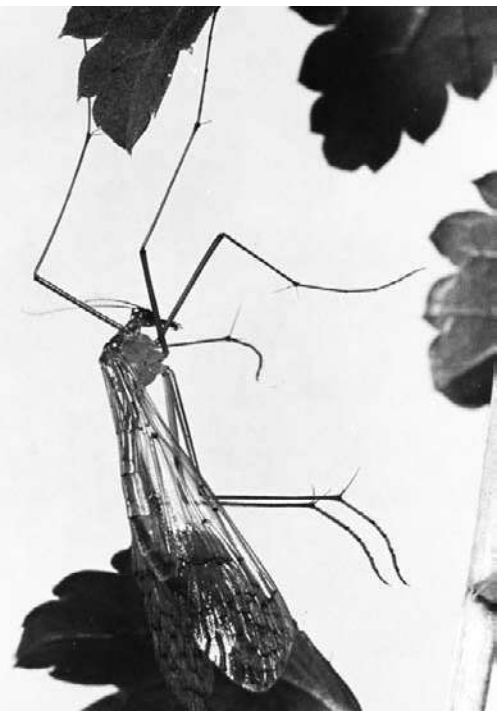

FIGURE 2 A hangingfly, *Bittacus* sp., in characteristic resting posture. (Photograph courtesy of D. W. Webb and Illinois Natural History Survey.)

insects (and are sometimes trapped themselves). Adult Bittacidae (Fig. 2) are predaceous, capturing a variety of insects and occasionally other arthropods by means of the raptorial hind tarsi. Larval bittacids, in contrast, are primarily saprophagous on dead insects but may occasionally feed on plant matter. Both adult and larval Boreidae feed on mosses. Adults of *Brachypanorpa* (Panorpodidae) have been observed scraping the upper epidermis from leaves of herbaceous plants; their larval diet is not known. There is some evidence that *Notiothauma* (Eomeropidae) is also a plant feeder. Larvae of aquatic Nannochoristidae feed on small larvae of chironomid flies, but nothing is known of the adult diet.

BEHAVIOR

Feeding by adults of some Mecoptera is often associated with mating behavior (Fig. 3). Some male bittacids, for example, capture insects as nuptial gifts, then extrude pheromone-dispersing vesicles between abdominal terga 6–7 and 7–8 to advertise their presence to females in the vicinity. Females seem to make an evaluation of the gift prey, leaving the male if his offering is too small or unpalatable. They may then respond to the pheromone of another male. Males of some Panorpidae, when they have found a good source of food, make known their presence by means of pheromonal vesicles within the enlarged genital segment. When a female is attracted closely enough, the male clamps the costal edge of her forewing in a structure formed by a peg on the anterior fourth abdominal tergum and an overlapping portion of the third (the notal

FIGURE 3 Mating pair of Japanese bittacids feeding on nuptial prey (an opilionid) initially offered to the female by the male. (Photograph courtesy of Yasushi Iwasaki.)

organ); mating then ensues as the female feeds. In other species of *Panorpa,* the male may build a small column of brownish saliva, gelatinous as it air-dries, which serves as a nuptial gift in place of a dead insect.

EGGS

Eggs of most Mecoptera are ovoid but approximately equally rounded at the ends. Those of *Bittacus* species are roughly cuboidal with the flattened surfaces shallowly impressed. The chorion of eggs of some species of *Panorpa* is smooth, while in others it is coated with a fine network of polygonal cells. In *Notiothauma,* the chorion is yellowish and granular; in *Austromerope,* the chorion is smooth. Eggs are ordinarily inserted into preexisting cavities in the soil by females of *Panorpa,* or in soil or rotten wood by *Brachypanorpa.* Females of these genera have been observed probing with the extended abdomen for some time before oviposition occurred. Female Bittacidae typically hang from the edge of a leaf and drop their eggs among dead leaves and other plant debris on the ground. Boreids deposit their eggs among the rhizoids of the mosses that will later be a source of food. As the embryonic larva develops and nears the time of hatching, the egg swells noticeably. In *Panorpa,* the increase is as great as 38%; larval eyes and mandibles become visible through the chorion. Eggs of *Bittacus,* roughly cuboidal when laid, become subspherical; those of *Harpobittacus* also become rounded and may double in size.

The egg stage is brief for some species (5–10 days for some American species of *Panorpa;* 14–16 days for panorpids living above 2250 m in Taiwan; 12–15 days in *Chorista*). In contrast, there may be an egg diapause, which together with larval development within the egg may last 216 to 256 days

in one North American *Bittacus,* and up to 290 days in a Japanese *Bittacus.* Larvae from diapausing eggs of *Boreus* laid in early winter hatch the following spring.

LARVAE

Larvae of Panorpidae, Bittacidae, and Choristidae are eruciform; that is, they have somewhat the appearance of caterpillars. The head is well sclerotized, as is the dorsum of the prothorax; the legs are nearly conical, their skin largely membranous; and there are subconical prolegs on abdominal segments 1 to 8. Larval bittacids have paired, elongate, three-branched, fleshy structures on the back of the mesothorax, metathorax, and nine abdominal segments. Panorpids and choristids have setae in most of the corresponding positions. The larvae of Panorpodidae and Boreidae, described as scarabaeiform, lack abdominal prolegs and conspicuous dorsal setae. Larvae of Nannochoristidae are unlike those of the other families, so far as known. They are aquatic, with slender, elongate, almost eel-like bodies and no prolegs. Their mouthparts are directed forward, not downward as in eruciform larvae.

A striking characteristic of many larval Mecoptera is presence of compound eyes, whereas larvae of most holometabolous insects have only one or a few simple eyes at each side. The larvae of Panorpidae and Choristidae have 30 or more ommatidia in each compound eye, while those of the Bittacidae have seven, Boreidae usually only three, and larvae of Panorpodidae lack eyes altogether. Larvae of Nannochoristidae have eyes with indistinctly formed ommatidia. Because their larvae are so different from the larvae of other families in the order (and because of some characteristics of the adults, such as a short rostrum), the Nannochoristidae have sometimes been considered as a group distinct from other Mecoptera.

Duration of larval life varies with the species involved, with temperature and season, availability of food, and length of prepupal diapause. Under favorable conditions, a larva of *Panorpa* may pass through three growth stages and become a fourth instar in about a month. During the final larval stage, feeding and growth continue for several days. But then the larva prepares a cavity in the soil, or other material in its habitat, stops feeding and becomes quiescent as it begins the period of prepupal diapause (inactivity) in this cavity. In species of *Panorpa* that have spring and late-summer generations, prepupal diapause in summer lasts only about 5 weeks, while in overwintering larvae it requires about 6.5 to 7 months. In nearctic species of *Bittacus* it lasts about 7 months. The pupal stage is much shorter: 10 to 21 days in *Panorpa,* 15 to 24 days in *Bittacus,* and 37 to 40 days in *Boreus.*

PUPAE

Described as exarate, the pupa has the legs rather loosely drawn up against its ventral surface and the wings tightly folded within their sheaths, which are not closely adhered to the body. The pupal rostrum is shorter than that of the adult and the mandibles larger. Otherwise, the pupal body has generally the form of the adult developing within.

See Also the Following Articles

Fossil Record • Mating Behaviors

Further Reading

Byers, G. W., and Thornhill, R. (1983). Biology of the Mecoptera. *Annu. Rev. Entomol.* **28,** 203–228.

Kaltenbach, A. (1978). Mecoptera (Schnabelhafte, Schnabelfliegen). *In* "Handbuch der Zoologie." Vol. 4, Part 2, Section 2, pp. 1–111. deGruyter, Berlin.

Setty, L. R. (1940). Biology and morphology of some North American Bittacidae (order Mecoptera). *Am. Midl. Nat.* **23,** 257–353.

Thornhill, R. (1980). Competition and coexistence among *Panorpa* scorpionflies (Mecoptera: Panorpidae). *Ecol. Monogr.* **50,** 179–197.

Thornhill, R. (1980). Mate choice in *Hylobittacus apicalis* (Insecta: Mecoptera) and its relation to some models of female choice. *Evolution* **34,** 519–538.

Medical Entomology

John D. Edman
University of California, Davis

Medical entomology is concerned with the impact of insects and related arthropods on the mental and physical health of humans, domestic animals, and wildlife. It is often subdivided into public health entomology and veterinary entomology. These divisions are tenuous since many of the same arthropods cause similar injuries and diseases in both humans and other animals. The history of medical entomology dates from the end of the 19th century, when arthropods were first shown to transmit important human diseases such as filariasis and malaria. Most arthropod-borne diseases are zoonotic infections that occur naturally in nonhuman hosts. Malaria, dengue fever, and most forms of filariasis are important exceptions. Household pests such as cockroaches and filth flies are sometimes included within medical entomology. When synanthropic flies and cockroaches mechanically contaminate food or other media with infectious organisms, they are clearly of medical importance. Nonetheless, these insects are generally treated in greater depth within the scope of urban entomology.

MEDICAL IMPORTANCE OF ARTHROPODS

Arthropods influence animal health in multiple ways. The most significant impact involves their role as primary vectors and alternate hosts of many devastating infectious disease agents. Parasitic agents transmitted by hematophagous arthropods include filariae, protozoa, bacteria, rickettsiae, and viruses. Arthropods also affect the health of vertebrates directly by triggering altered mental states (delusional parasitosis and

entomophobia/arachnophobia), contact allergies, feeding annoyance and blood loss, envenomization, and myiasis.

Each year these arthropod relationships collectively cause the death of millions of humans and bring illness to hundreds of millions more. Their impact is greatest in poor tropical countries where they are one of the major factors limiting animal production, agricultural productivity, economic development, and well-being. Several vector-borne diseases (e.g., plague, malaria, leishmaniasis, yellow fever, and dengue) apparently were transported from the Old World to the Western Hemisphere via the slave trade or ship crews. Related species of insects in the New World were able to successfully maintain and transmit these introduced parasites. In recent years, several native but previously unrecognized arthropod-borne infections, including tick-borne Lyme disease and ehrlichiosis, have been discovered in the United States and elsewhere. Well-known diseases such as dengue and malaria have been reemerging or expanding in many regions of the world. Other diseases such as West Nile fever recently have been introduced into new regions with devastating effects. A combination of population growth, rapid movement of people and other animals, and environmental disruption has contributed to the growing threat posed by vector-borne diseases. Control of these diseases is complicated by a lack of investment in public health and development of drug-resistant parasites and insecticide-resistant vectors.

ORDERS AND FAMILIES OF MEDICAL CONCERN

Phylum Arthropoda contains several classes of invertebrates that have direct medical importance. In addition to insects, these include arachnids (spiders, mites, ticks, and scorpions), millipedes, and centipedes. Some crustaceans (sowbugs, copepods), molluscs (snails), arachnids (oribatid mites), and insects (mainly beetle larvae) are intermediate hosts for parasitic worms.

Orders and families of blood-feeding species are of major medical significance (Table I). Most but not all of these are involved in the transmission of microparasites. Nonetheless, the simple act of feeding by arthropod ectoparasites can result in blood loss, anemia, stress, discomfort, allergic reactions, and reduced productivity. A few insects are specialized ectoparasites on humans. Bed bugs and kissing bugs are nest parasites and some have permanently invaded human dwellings and feed at

TABLE I Orders and Families with Important Blood-Feeding Insects, Ticks, and Mites

Taxon	Common name	Blood feeding	Types of hosts
Insecta (=Hexopododa)			
Order Anoplura	Sucking lice	♂, ♀, nymph	Mammals
Order Mallophaga[a]	Chewing lice	♂, ♀, nymph	Birds/mammals
Order Heteroptera			
Family Cimicidae	Bed/bird/bat bugs	♂, ♀, nymph	Mammals/birds
Family Triatomidae	Kissing bugs	♂, ♀, nymph	Mammals/birds
Order Siphonaptera	Fleas	♂, ♀	Mammals/birds
Order Diptera			
Family Culicidae	Mosquitoes	♀	All vertebrates
Family Simuliidae	Black flies	♀	Mammals/birds
Family Ceratopogonidae	Biting midges	♀	All vertebrates
Family Psychodidae	Sand flies	♀	Mammals/reptiles
Family Tabanidae	Horse/deer flies	♂, ♀	Mammals
Family Rhagionidae	Snipe flies	♂, ♀	Mammals
Family Muscidae			
Stomoxys	Stable flies	♂, ♀	Mammals
Haematobia	Horn/bush flies	♂, ♀	Mammals
Musca	Cattle flies	♂, ♀	Mammals
Family Glossinidae	Tsetse flies	♂, ♀	Mammals/reptiles
Family Hippoboscidae	Louse flies, keds	♂, ♀	Mammals/birds
Family Nycteribiidae	Bat flies	♂, ♀	Bats
Family Streblidae	Bat flies	♂, ♀	Bats
Arachnida (subclass Acari)			
Family Ixodidae	Hard ticks	♂, ♀, L, N[b]	Birds/mammals/reptiles
Family Argasidae	Soft ticks	♂, ♀, L, N	Birds/mammals
Family Trombiculidae	Chigger mites	Larvae	Birds/mammals
Family Dermanyssidae	Mesostig mites	♂, ♀, L, N	Birds/mammals
Family Macronyssidae	Mesostig mites	♂, ♀, L, N	Birds/mammals
Family Laelapidae	Mesostig mites	♂, ♀, L, N	Birds/mammals
Family Demodicidae[a]	Follicle mites	♂, ♀, L, N	Mammals
Family Psoroptidae[a]	Mange mites	♂, ♀, L, N	Mammals
Family Sarcoptidae[a]	Scab mites	♂, ♀, L, N	Mammals

[a]Ectoparasites that feed mainly on skin/feather tissues rather than on blood.
[b]L, larvae; N, nymph.

night on sleeping people. Humans are also parasitized by three species of sucking lice (i.e., pubic louse, head louse, and body louse), each of which specializes in a different region of the body. Biting flies such as mosquitoes, black flies, deer flies, and biting midges can reach annoyance levels that make outdoor activities nearly impossible, but these hordes normally feed on a wide variety of domestic and wild animals. Only a few biting flies such as the yellow-fever mosquito *(Aedes aegypti),* the tropical house mosquito *(Culex pipiens quinquefasciatus),* and some vector species of *Anopheles, Simulium, Phlebotomus,* and *Lutzomyia* feed preferentially on humans.

Both insects and arachnids can cause harm because of the venoms they contain. Venomous insects are found mainly in the order Hymenoptera within the families Formicidae (ants), Vespidae (yellowjackets, hornets), Mutillidae (velvet ants), and Apidae (honey bees, bumble bees). Some Coleoptera (e.g., Meloidae, Staphlinidae, Chrysomlelidae, Dermestidae) and Lepidoptera (e.g., Noctuidae, Saturniidae, Sphingidae, Nymphalidae) also produce toxic defensive secretions through specialized glands and urticating hairs or have hemolymph that is toxic to vertebrates if the insect is crushed. These secretions are especially toxic when exposed to mucus or lachrymal glands. Spiders in the genera *Loxosceles* (recluses), *Latrodectus* (widows), *Atrax* (Australian funnel-webs), *Harpactirella* (South Africa), *Lycosa* (Central and South America), and *Phoneutria* (Brazil) include some of the most highly venomous species. All scorpions have venomous stings and those within the family Buthidae can be fatal to humans. Desert regions of the Americas, the Mediterranean, and northern Africa are home to most of the highly poisonous scorpions. Centipedes also have venomous bites and millipedes have toxic defensive secretions, but these normally are not severe or life threatening.

Arthropod allergens that cause acute asthma in humans are mostly associated with house dust mites (*Dermatophagoides* spp.) and cockroaches *(Periplaneta, Blattella);* however, the airborne wing scales or fine setae associated with large populations of other insects such as mayflies, caddisflies, and gypsy moths may invoke allergic reactions in sensitized individuals. Three families of dipterous insects (Calliphoridae, Sarcophagidae, Oestridae) include species whose larvae are obligate parasites living within the flesh of vertebrates, a condition referred to as myiasis. The unusual parasitic mites that live within the feathers, nasal passages, and lungs of birds are relatively benign.

DIRECT INJURIES CAUSED BY ARTHROPODS

The various direct effects of arthropods on humans are summarized in Table II. The most common of these is the asthma suffered by millions of people, especially children, who are allergic to the fine airborne particulates generated by insects and mites. These allergens are most often associated with feces or decomposing body parts. The advent of air conditioning and wall-to-wall carpeting seems to have exacerbated this problem especially among the middle and upper classes. Asthma is one of the fastest growing medical problems, particularly among children. Respiratory failure is not an uncommon outcome. Concern over life-threatening multiple stings and allergic reactions (anaphylaxis) to the venom of insects has increased as a result of the introduction and spread of the hybrid of the African honey bee *(Apis mellifera adansonii)* and the imported fire ant *(Solenopsis invicta)* in the Western Hemisphere.

Myiasis is a serious problem in animal production, especially in the neotropics where millions of dollars are lost annually due to these tissue-invading flies. These flies are often found within families that include species that normally feed on the decaying tissues of dead or wounded animals, i.e., flesh flies in the family Sarcophagidae and blow flies in the family Calliphoridae. Some species of Calliphoridae (and related families) sometimes will facultatively invade living tissues while others are so restricted to dead tissues that they are used in maggot therapy to clean deep wounds. All members of the four subfamilies of bot flies (Oestrinae, Gasterophilinae, Hypoderminae, and Cuterebrinae of the family Oestridae) are obligate parasites. The torsalo *(Dermatobia hominis)* is a neotropical dipteran whose eggs are glued to the abdomen of biting flies, and its larvae emerge during blood feeding by the host fly.

ARTHROPOD TRANSMISSION OF MICROPARASITES

Vector-borne diseases can be either biologically or mechanically transmitted. In mechanical transmission, vector mouthparts

TABLE II Direct Effects of Arthropods on Humans and Other Animals

Condition	Health effects	Arthropods involved
Delusional parasitosis	Irrational or destructive acts	Imagined skin parasites
Entomophobia	Stress and mental fatigue	Spiders, wasps
Airborne allergies	Inflamation and respiratory distress	Usually cockroaches and house dust mites
Irritation and blood loss	Allergic skin reactions; pain, itching, inflammation; stress, anemia, and death	Blood-feeding and skin-invading ectoparasites
Envenomization	Arthus and anaphylactic reactions; neurological and cytolytic damage; pain, inflammation, and death	Those with toxic stings, bites, setae, or fluids
Myiasis	Tissue damage, prolonged pain; weight loss, stress, secondary infection, and death	Dipteran maggots

serve as contaminated hypodermic needles since there is no replication or development of the microparasite in the vector; the vector does not serve as an alternate host as in the case of biologically transmitted diseases. For infectious organisms to be mechanically transmitted by arthropods efficiently, they must be abundant in circulating blood or cutaneous tissues and able to survive external exposure. Diseases that are mechanically transmitted by arthropods generally have other transmission mechanisms as well.

Biological transmission takes one of three forms: *propagative* transmission, which involves the replication of eukaryotic parasites and dissemination to the salivary glands prior to transmission; *cyclodevelopmental* transmission, which occurs among filarial parasites, in which development of the parasite to the infective stage is required prior to transmission but there is no increase in the number of parasites; or *cyclopropagative* transmission, which involves both development and multiplication by the parasite as occurs with protozoan parasites such as *Plasmodium, Leishmania,* and *Trypanosoma.*

After the appropriate extrinsic incubation period for replication and/or development of the parasite, the arthropod vector is said to be infective, i.e., able to transmit. Infectious agents are passed to the vertebrate host through a number of different routes: (1) transovarially from the female to her offspring or transtadially from one stage to the next, (2) venereally from infected males to uninfected females, (3) through cofeeding when infected and uninfected vectors group feed, or (4) horizontally through (a) infective saliva injected during feeding, (b) regurgitation of parasites blocking the food canal, (c) defecation of infective feces on the skin, or (d) active escape from the mouthparts and invasion of the skin. In some cases, hosts must assist the transmission process by crushing the infective insect, scratching the contaminated area, and rubbing the eyes. Transmission may be promoted by the physiological or behavioral affect of parasites on the vector. Invasion of host cells or tissues by the parasite can be modulated by host immune responses to the salivary secretions of the vector.

The efficiency of transmission is determined primarily by the "competence" of the arthropod species to support development/replication of the parasite and by the ecology and behavior of the arthropod species. The latter determines the temporal and spatial connection between hosts and potential vectors. Thus, the density, feeding frequency, and host preferences of the vector play critical roles in establishing the *vectorial capacity* (number of infective bites received daily by a single host) of any given species. Environmental conditions also play an important role.

DISEASES TRANSMITTED BY ARTHROPODS

The major diseases transmitted by arthropods are listed in Table III. Protozoan parasites dominate this list in terms of worldwide importance. Malaria is the single most significant vector-borne disease, with an estimated 300 million people infected annually and over 1 million deaths among young children in Africa alone. It is endemic in most tropical and subtropical regions of the world, where it has been resurging since eradication attempts ended some 30 years ago. It is transmitted by *Anopheles* mosquitoes and can be successfully treated with drugs if promptly available. Drug resistance is a growing problem. Tsetse-transmitted African trypanosomiasis remains a serious human disease in parts of tropical Africa but its impact on domestic cattle production is even more severe. Wild bovines are the reservoir of the acute Rhodesian form, and humans and porcines are the reservoir of the chronic Gambian form. American trypanosomiasis is restricted to the mountain regions of tropical America and is often a silent disease that leads to early death. Great efforts are under way to control this zoonotic disease by spraying residual pesticides and by constructing houses with materials that prevent invasion by domesticated kissing bugs. Visceral and cutaneous forms of leishmaniasis affect millions of people and cause significant mortality in Africa, Asia, and South America. The reservoirs for these parasites are dogs, rodents, and a variety of other wild mammals. The cutaneous form also exists in southern Europe and Central America, but disfigurement rather than mortality is usually associated with this form.

Lymphatic filariasis (elephantiasis), transmitted by *Culex quinquefasciatus* and other human-biting mosquitoes, infects many millions of residents throughout the tropics. The long-lived nematode parasites cause debilitation, especially of the lower limbs, but seldom result in death. Onchocerciasis (river blindness), also a filarial infection, is limited to sub-Saharan Africa and some coffee-growing regions in Central and South America. Both of these nematode infections can be prevented by treatment with a new drug, ivermectin. River blindness has been effectively controlled in much of West Africa in recent years through pesticidal elimination of the black fly vectors breeding in streams and the use of anti-helminthics to treat the human population. There are a large number of bacterial and rickettsial infections transmitted by arthropods (mainly ticks, fleas, lice, and mites) annually but none compare to the impact of plague and typhus epidemics in earlier times. Some of these, such as tick-borne Lyme disease and ehrlichiosis, have just been recognized within the past 20 years. Lyme disease is the most prevalent vector-borne disease in the United States. Protective vaccines exist for few bacterial and rickettsial diseases, but all respond to timely antibiotic therapy.

Arthropod-borne viral diseases (arboviruses) are transmitted mainly by mosquitoes and biting midge, but Russian spring– summer encephalitis and several other zoonotic infections (especially of livestock) are transmitted by ticks. Historically, yellow fever was the most important human arboviral disease but today it has been replaced by dengue fever. Dengue viruses infect over a million people annually and can produce a fatal hemorrhagic disease, especially among children. This disease has reinvaded the Western Hemisphere in recent decades and now causes hundreds of thousands of cases annually. Vaccines exist for a

TABLE III Some Important Diseases Transmitted by Arthropod Vectors

Microparasite	Disease	Arthropod vector	Distribution
	Transmitted biologically		
Nematodes			
Dirofilaria	Canine heartworm	Mosquitoes	Worldwide
Brugia, Wucheraria	Lymphatic filariasis	Mosquitoes	Tropics
Onchocerca	Riverblindness	Black flies	Africa, Central and South America
Protozoa			
Leishmania	Visceral and cutaneous leishmaniasis	Sand flies	Tropics and warm temperate areas
Trypanosoma spp.	Sleeping sickness and Nagana of cattle	Tsetse flies	Sub-Saharan Africa
Trypanosoma cruzi	Chagas disease	Kissing bugs	Neotropics
Plasmodium	Malaria	Mosquitoes	Mostly tropical
Theileria	Theileriosis	Hard ticks	Africa, Southern Europe
Babesia	Babesiosis	Hard ticks	Widespread
Bacteria			
Bartonella	Carrions disease	Sand flies	South America
	Trench fever	Body lice	Worldwide
	Cat-scratch fever	Fleas	Widespread
Borrelia	Lyme disease	Hard ticks	North America, Eurasia
	Relapsing fever	Soft ticks, lice	Worldwide
Yersinia	Plague	Fleas	Worldwide
Francisella	Tularemia	Hard ticks	Worldwide
Rickettsia and other obligate intracellular bacteria			
Rickettsia	Epidemic typhus	Body lice	Africa, Americas
	Murine typhus	Fleas	Widespread
	Spotted fevers	Hard ticks	Widespread
Orientia	Scrub typhus	Chigger mites	Southeast Asia
Cowdria	Heartwater	Hard ticks	Sub-Saharan Africa
Anaplasma	Anaplasmosis	Hard ticks	Worldwide
Ehrlichia	Ehrlichiosis	Hard ticks	Widespread
Arthropod-borne viruses (arboviruses)[a]			
Flaviviruses	Yellow fever, dengue, WN, SLE, JE, MVE, ROC, WSL	Mosquitoes	Tropics
	RSSE, OMSK, KFD, LI, POW	Hard ticks	Widespread
Bunyaviruses	CE, LAC, RVF	Mosquitoes	Africa, North America
	ORO	Biting midges	South America
	CCHF, NSD	Hard ticks	Africa, Eurasia
	SFF	Sand flies	South America, Africa, and Eurasia
Togaviruses	EEE, WEE, VEE, RR	Mosquitoes	Widespread
Rhabdoviruses	VSV, BEF	Biting flies	Widespread
Reoviruses	Bluetongue, AHS, EHD	Biting midges	Widespread
	Colorado tick fever	Hard ticks	Western North America
Unnamed	African swine fever	Soft ticks	Africa
	Transmitted mechanically		
Protozoa			
Trypanosoma	Trypanosomiasis	Biting flies	Widespread
Bacteria			
Treponema	Yaws/pinta	Eyes gnats	Tropics
Bacillus	Anthrax	Biting flies	Widespread
Anaplasma	Anaplasmosis	Biting flies	Widespread
Various Anaerobes	Summer mastitis	Head flies	Widespread
Viruses			
Poxvirus	Myxomatosis, fowlpox	Biting flies	Worldwide
Retrovirus	Equine infectious anemia	Tabanids	Worldwide

[a]WN, West Nile; SLE, St. Louis encephalitis; JE, Japanese encephalitis; MVE, Murray Valley encephalitis; WSL, Wesselsbron; RSSE, Russian spring summer encephalitis; OMSK, Omsk hemorrhagic fever; KFD, Kyasanur Forest disease; LI, Louping ill; POW, Powassan; CE, California encephalitis; LAC, La Crosse encephalitis; RVF, Rift Valley fever; ORO, Racio; CCHF, Crimean-Congo hemorrhagic fever; NSD, Nairobi sheep disease; SFF, sand fly fever; EEE, Eastern equine encephalitis; WEE, Western equine encephalitis; VEE, Venezuelan equine encephalitis; RR, Ross River; VSV, vesicular stomatitis virus; BEF, bovine ephemeral fever; AHS, African horse sickness; EHD, epizootic hemorrhagic disease.

few arboviral diseases (yellow fever and Japanese encephalitis) but vector control is often the only preventative measure.

DISEASE AND VECTOR MANAGEMENT

Efforts to develop vaccines for a wide range of vector-borne diseases, including malaria, have been vigorously supported but success has been slow. Antigenic variation in parasites has been a major deterrent. New drugs for treatment of parasitic diseases also are under development and several important new drugs for treatment of helminth and protozoan parasites have been marketed in recent years. Nonetheless, vector control is often the first line of defense against the transmission of these diseases and, during active epidemics, this is the only option outside of public education. Control of vertebrate reservoir animals has occasionally been practiced for diseases that are maintained by rodents. An early example was control of wild bovine reservoirs in parts of Africa to control trypanosomiasis in humans (sleeping sickness) and cattle (Nagana). Vector control programs are generally based on surveillance systems that monitor and report cases of disease or vector population levels. A variety of tools are available to manage vector populations, including chemical pesticides, biological controls, habitat alteration, and personal protection strategies (e.g., screens, bed nets, and repellents). Currently, genetic control strategies are receiving much attention. Optimal vector control programs utilize integrated vector management strategies and strive to maintain vector populations below the threshold densities required for transmission. Targeting the immature stages of vectors that blood feed and transmit disease only as adults normally is more efficient and cost effective. Lack of public health funds is the major limitation on surveillance and vector control programs, especially in developing countries where these diseases have the greatest impact.

See Also the Following Articles

Bubonic Plague • Chiggers and Other Disease-Causing Mites • Lice, Human • Malaria • River Blindness • Ticks • Tsetse Fly • Veterinary Entomology • Yellow Fever • Zoonoses, Arthropod-Borne

Further Reading

Busvine, J. R. (1976). "Insects, Hygiene and History." University of London Press, London.

Eldridge, B. F., and Edman, J. D. (2000). "Medical Entomology: A Textbook on Public Health and Veterinary Problems Caused by Arthropods." Kluwer Academic, Dordrecht/Norwell, MA.

Harwood, R. F., and James, M. T. (1979). "Entomology in Human and Animal Health," 7th ed. Macmillan, New York.

Kettle, D. S. (1994). "Medical and Veterinary Entomology," 2nd ed. Cambridge University Press, Cambridge, U.K.

Lane, R. P., and Crosskey, R. W. (1993). "Medical Insects and Arachnids." Chapman & Hall, London/New York.

Lederberg, J., Shope, R. E., and Oaks, S. C., Jr. (1992). "Emerging Infections." National Academy Press, New York.

Service, M. W. (1978). A brief history of medical entomology. *J. Med. Entomol.* **14,** 603–626.

Service, M. W. (2000). "Medical Entomology for Students," 2nd ed. Kluwer Academic, Dordrecht/Norwell, MA.

Medicine, Insects in

Ronald A. Sherman
University of California, Irvine

Throughout history, humans have used insects and their products therapeutically. Ingested, injected, or topically applied, insects have been used to treat an assortment of respiratory, gastrointestinal, cardiac, neuromuscular, and infectious diseases. To this day, therapeutic insects are prescribed worldwide. Medicinal maggots and honey bee venom therapy are two examples of therapeutic roles played by insects.

INTRODUCTION AND HISTORICAL OVERVIEW

Many insects are ingested for their medicinal value as well as their nutritional value. Roasted, boiled, or powdered cockroaches (Blattodea: Blattidae) have been ingested by people of many cultures to treat respiratory diseases. Various beetles (Coleoptera) are noted to be useful in the treatment of intestinal diseases. The blister beetle, *Lytta vesicatoria* (Spanishfly), is the source of cantharidin, a vesicant that was ingested in Europe as an aphrodisiac. Stinkbugs (Heteroptera: Pentatomidae) in China and termites (Isoptera) in India were used for the same function.

Up until a few decades ago, patients with neurosyphilis often were cured with an inoculation of malaria (*Plasmodium* spp.). The syphilis pathogen *(Treponema pallidum)* was killed by the recurrent fevers and/or other still unknown interactions with the malaria parasites. The malaria was then eradicated with quinine or another species-specific antimalarial agent. Initially, malaria inoculation was achieved by transferring the blood of a malaria-infected neurosyphilitic to a nonparasitized neurosyphilitic. This practice was soon replaced by mosquito-transmitted inoculations.

"Caterpillar fungus" *(dong chong xia cao)* is a Chinese moth larva (Hepialidae: *Hepialus oblifurcus*) infected with an entomopathogenic (insect-killing) fungus, *Cordyceps sinensis* (Clavicipatales: Ascomycotina). Ingestion of the caterpillar fungus reportedly strengthens and rejuvenates the body. The substance achieved international notoriety in 1993, when fungus-drinking Chinese athletes set new world track records. At a cost of approximately $1000 per kilogram, caterpillar fungus is often prepared as a broth; both the broth and the caterpillar are eaten.

Arthropods have played a role in wound care for centuries. The use of large ant or beetle mandibles for holding together wound edges has been documented in many countries. Honey

and spiderwebs both have been used to dress wounds and prevent infection. Maggot therapy—the topical application of blowfly larvae (*Phaenicia, Lucilia,* and *Phormia*) to treat infected wounds—has been practiced around the world for at least 70 years.

MAGGOT THERAPY

The practice of maggot therapy is based on observations that wounds naturally infested with maggots (wound myiasis) often are free of infection and debris. For centuries, European military surgeons described how the maggot-laden wounds of soldiers not promptly removed from the battlefield often appeared clean, once the larvae were wiped away. Soldiers' maggot-infested wounds seemed to heal better than wounds that had not been infested. After his own observations of wound myiasis on the battlefields of World War I, William Baer intentionally placed blowfly larvae into the chronic wounds of his patients at Johns Hopkins and Children's Hospital in Baltimore. Baer first presented his results in 1929; by 1935, thousands of physicians and surgeons had embraced this practice. Many hospitals maintained their own therapeutic fly colonies; other practitioners obtained maggots from pharmaceutical companies.

Maggot therapy all but disappeared during the 1940s. The reasons are purely speculative but probably include the development of antibiotics and the refinement of surgical techniques that came about during World War II. The relatively high cost of maggots ($5 for a bottle of 1000 larvae) may have been another factor. Over the next several decades, therapeutic myiasis was performed only rarely, and only as a last resort, in patients who failed to respond to aggressive surgical and antibiotic treatments. The 1980s brought about the realization that surgery and antibiotics could not cure all wounds. Many infections were now resistant to the once omnipotent antimicrobials. The 1990s saw the reemergence of maggot therapy to treat many of these nonhealing wounds. Today, live fly larvae are once again used in over a thousand centers worldwide for treating chronic wounds.

Maggots effectively treat open wounds by removing dead and infected tissue (debridement), killing bacteria (disinfection), and stimulating the wound to heal. To understand the procedure of maggot therapy, it is necessary to review the natural history of the fly. Many species of blowflies (Calliphoridae) naturally "blow" or lay their eggs on carrion, feces, or the dead (necrotic) tissue of a living host. Upon hatching, the larvae ingest this tissue as it is liquefied by the maggots' digestive secretions. Within 3 to 7 days, the maggots leave what remains of their meal and pupate underground or in some other protected site. One to three weeks later, adult flies emerge. Therapeutic maggots—blowfly larvae that have been disinfected ("sterile" maggots)—are placed on wounds at a density of about 5 to 10 cm^{-2}. Larvae are retained on the wound for about 48 h, in cagelike dressings. After one or more such cycles of treatment, the wound is often free of necrotic tissue and able to accept a skin graft or heal spontaneously.

FIGURE 1 The right foot of a 73-year-old man who was treated for 3 years by orthopedic and podiatric surgery for his chronic foot ulcers. (A) before, (B) during, and (C) 1 year after maggot debridement therapy.

Treatments can be administered in the hospital, clinic, or nursing home, or at home.

Maggot therapy has been used to treat pressure ulcers, venous stasis ulcers, diabetic foot ulcers, burns, traumatic wounds, and nonhealing postsurgical wounds (Fig. 1). Compared with conventional wound therapy, medicinal maggots are credited with more rapid debridement and wound healing. Maggot therapy has reportedly saved numerous limbs from amputation and other surgical procedures. Other advantages of maggot therapy include its simplicity and safety, and (by current standards) the relatively low cost of treatment.

APITHERAPY

Honey bee *(Apis mellifera)* venom, propolis, royal jelly, beeswax, and honey all are used therapeutically. Medicinal use of any of these products can be considered to be "apitherapy," but many authorities use the term specifically to denote the clinical use of honey bee venom. Bee venom contains a multitude of polypeptides, enzymes (phospholipase A_2, hyaluronidase), and biologically active amines (histamine, dopamine, noradrenaline). The mechanisms by which venom exerts its beneficial actions are unknown, but might include an anti-inflammatory effect resulting from alterations seen in pituitary and adrenal gland function, local effects on the nerves and blood vessels, and stimulation of acupuncture-like pathways. A combination of these and other mechanisms may explain the diversity of benefits attributed to venom therapy.

Apitherapists have successfully treated rheumatological disorders (rheumatoid and psoriatic arthritis, gout, fibromyalgia), neurological diseases (multiple sclerosis, chronic pain syndromes), immunological diseases (scleroderma, systemic lupus erythematosis), and other chronic illnesses. Some therapists

administer the treatments in the form of an increasing number of bee stings; other practitioners inject a partially purified extract of the bee venom, in gradually increasing doses. Serious toxic reactions to the venom are uncommon because patients are educated about the signs and symptoms of venom reactions; they are also supervised closely following treatment and are given ready access to medical care after leaving the therapist's office or apiary.

See Also the Following Articles

Bee Products • Beeswax • History of Entomology • Venom

Further Reading

Baer, W. S. (1929). Sacro-iliac joint—arthritis deformans—viable antiseptic in chronic osteomyelitis. *Proc. Int. Assembly Inter-state Postgrad. Med. Assoc. North Am.* **371,** 365–372.

Chernin, E. (1984). The malaria therapy of neurosyphilis. *J. Parasitol.* **70,** 611–617.

Gudger, E. W. (1925). Stitching wounds with the mandibles of ants and beetles; a minor contribution to the history of surgery. *J. Am. Med. Assoc.* **84,** 1861–1865.

Kim, C. M.-H. (1997). Apitherapy (bee venom therapy). *In* "Potentiating Health and the Crisis of the Immune System: Integrative Approaches to the Prevention and Treatment of Modern Diseases" (A. Mizrahi, S Fulder, and N. Sheinman, eds.), pp. 243–270. Plenum Press, New York.

Leclercq, M. (1969). "Entomological Parasitology; The Relations between Entomology and the Medical Sciences" (G. Lapage, transl.). Pergamon Press, Oxford, U.K.

Mizrahi, A., and Lensky, Y. (eds.) (1996). "Bee Products: Properties, Applications, and Apitherapy." Plenum Press, New York.

Mumcuoglu, K. Y., Ingber, A., Gilead, L., Stessman, J., Friedmann, R., Schulman, H., Bichucher, H., Ioffe-Uspensky, I., Miller, J., Galun, R., and Raz, I. (1999). Maggot therapy for the treatment of diabetic foot ulcers. *Diabetes Care* **21,** 2030–2031.

Sherman, R. A. (1998). Maggot therapy in modern medicine. *Infect. Med.* **15,** 651–656.

Sherman, R. A., Hall, M. J. R., and Thomas, S. (2000). Medicinal maggots: An ancient remedy for some contemporary afflictions. *Annu. Rev. Entomol.* **45,** 55–81.

Steinkraus, D. C., and Whitfield, J. B. (1994). Chinese caterpillar fungus and world record runners. *Am. Entomol.* **40,** 235–239.

Taylor, R. L. (1975). "Butterflies in My Stomach." Woodbridge Press, Santa Barbara, CA.

Thomas, S., Jones, M., Shutler, S., and Jones, S. (1996). Using larvae in modern wound management. *J. Wound Care* **5,** 60–69.

Megaloptera
(Alderflies, Dobsonflies)

N. H. Anderson
Oregon State University

The Megaloptera, which include the alderflies, dobsonflies, fishflies, and hellgrammites, are a small order of neuropterous insects with 250 to 300 species. They are often considered to be the most primitive group of insects with complete metamorphosis. In the fossil record, Megaloptera first occur about 250 mya in the late Permian. Their sister group is the Raphidioptera (snakeflies), and both these orders are closely related to the Neuroptera (lacewings and antlions). Among the major orders, they are most closely related to the Coleoptera (beetles). Larvae of all Megaloptera are aquatic predators. Their association with cool, well-oxygenated waters probably accounts for their greater species diversity in temperate regions than in the tropics. Although they are found throughout the world, the distributions are discontinuous, which is characteristic of a relict fauna.

DIAGNOSIS AND CLASSIFICATION

Megaloptera can be distinguished from other orders of insects by the following combination of characters: holometaboly, terrestrial adults and eggs, predaceous aquatic larvae, and exarate terrestrial pupae.

The adults are large, short-lived, and soft-bodied; the head is broad and flattened, with chewing mouthparts, large bulging compound eyes, and long many-segmented antennae; there are two pairs of similar membranous wings, held rooflike over the body, with all major veins present and many cross-veins; the five-segmented tarsi have paired apical

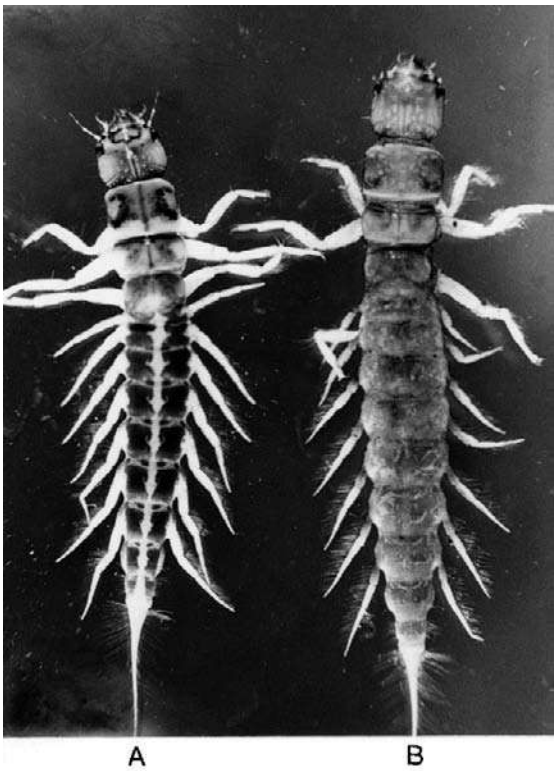

FIGURE 1 Mature larvae of *Sialis:* (A) *S. rotunda* and (B) *S. californica.* [From Azam, K. M., and Anderson, N. H. (1969). Life history and habits of *Sialis rotunda* and *S. californica* in Western Oregon. *Ann. Entomol. Soc. Am.* **62,** 549–558.]

FIGURE 2 *Sialis* adult.

claws; and the abdomen has 10 segments, eight pairs of spiracles, and lacks cerci.

The larvae have a well-sclerotized head, with long toothed mandibles, four- or five-segmented antennae, and six lateral stemmata (eyespots); the thorax is sclerotized dorsally, with a quadrate pronotum, meso- and metathoracic spiracles, and elongate five-segmented legs with two tarsal claws; and the abdomen is soft, with seven or eight pairs of lateral filaments, and spiracles on segments 1 to 8.

The order contains two families, Sialidae and Corydalidae. Sialids, or alderflies, are slow, awkwardly flying insects that range from 10 to 15 mm in length. Their bodies are black, brown, or yellowish orange. The wings are held tentlike over the abdomen so that they bear some resemblance to caddisflies (Trichoptera). The Corydalidae are much larger, 40 to 75 mm in length. Many species have pale, smoky wings mottled with brown, whereas others are nearly black, with white markings. This family is divided into two distinctive subfamilies: the Corydalinae or dobsonflies (larvae are hellgrammites), and the Chauliodinae or fishflies.

FAMILY SIALIDAE

Of the eight genera in the family Sialidae (Figs. 1–3), only the widespread Holarctic genus *Sialis* has received much attention. The life history of *S. lutaria* has been studied extensively in Europe, and the larvae have been used in behavioral and physiological experiments. *Sialis* larvae are found in many habitats, ranging from small springs to large rivers and from ponds to large lakes. They usually occur where the substrate

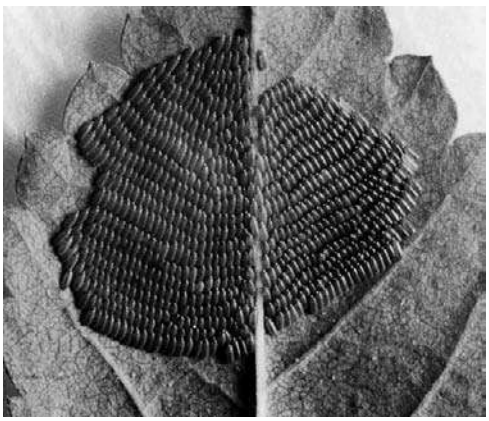

FIGURE 3 *S. californica* egg masses. [From Azam, K. M., and Anderson, N. H. (1969). Life history and habits of *Sialis rotunda* and *S. californica* in Western Oregon. *Ann. Entomol. Soc. Am.* **62,** 549–558.]

is soft and where dead leaves and other detritus have accumulated. Larvae may dig into the substrates to a depth of several centimeters. Sympatric species may be ecologically segregated based on habitat. For example, in Oregon *S. rotunda* occurs in ponds with a muddy bottom, whereas *S. californica* is usually found in pools or glides of streams. However, the two species can be found together in backwater reaches.

The life cycle takes 1 or 2 years depending on physical and biological conditions. The colder and less productive habitats such as trout streams and mountain lakes lengthen the period of larval life, whereas rapid development occurs in warm, productive habitats such as lowland ponds, warm lakes, and muddy rivers.

Larvae feed nonselectively on small invertebrates such as insect larvae, annelid worms, crustaceans, and mollusks. The prey is seized by the elongate mandibles and forelegs and worked into the mouth with the aid of labrum, maxillae, and labium. Sometimes only the softer abdominal parts of prey are eaten. Cannibalism occurs, especially in high-density situations.

During spring or early summer the final instars move to shallow areas near the shore. They then leave the water and prepare to pupate. The transformation also involves a switch in respiration from using aqueous dissolved oxygen via the gills to intake of atmospheric oxygen through the abdominal and thoracic spiracles. Pupation occurs in an unlined chamber dug 1 to 10 cm into soil or litter.

The adults emerge after a pupal instar of about 2 weeks. Adults are most active during midday. In mating, the male crawls beneath the female's abdomen from the rear and raises his abdomen upward and forward to couple the genitalia. A spermatophore is passed to the female within a few minutes, copulation is terminated, and oviposition occurs within a day.

Adult *Sialis* have biting mouthparts, and some observations suggest that they visit flowers to feed.

The eggs are deposited on a variety of substrates projecting over the water. The underside of overhanging leaves is the most common oviposition site. Egg masses contain 300 to 900 eggs. There are two distinctive types of egg mass; in one, the long

axis of the egg is almost parallel to the substrate, and in the other it is almost upright. The incubation period is about 10 days to 2 weeks. Eggs are sometimes parasitized by tiny *Trichogramma* wasps.

In hatching, the larva pushes its head against the chorion below the micropylar projection. The toothed, V-shaped egg burster then ruptures the chorion, initiating a jagged tear through which the larva emerges. When the larva leaves the egg, a postembryonic molt occurs, and the egg burster and embryonic membrane are left attached to the eggshell. Then the appendages and abdominal filaments are expanded and the larva becomes active. It drops to the water surface, where it is quickly wetted to pass through the surface film, and then swims to the bottom. There are about 10 instars before pupation.

CORYDALIDAE

Subfamily Corydalinae

The subfamily Corydalinae includes nine genera and is found mainly in the tropics or subtropics. Only four species of *Corydalus* occur in America north of Mexico, whereas there are three genera and almost 50 species of Corydalinae described from the Neotropics. Corydalinae contain the largest Megaloptera, with some adults having a wingspan greater than 150 mm.

C. cornutus, the dobsonfly, ranges over most of North America east of the Continental Divide, from Canada to Mexico. The larvae, called hellgrammites, are large (to 65 mm), dominant predators in stream riffles that feed opportunistically on invertebrate prey. Final instars leave the stream to pupate in the soil under rocks, or in rotting logs.

C. cornutus is unusual in that the male has elongate, hornlike mandibles that are half the length of the body. The mandibles of the female are also large, but only as long as the head (Fig. 4). Mating occurs shortly after emergence. The male has scent glands between abdominal segments 8 and 9, which evidently produce a sex "stimulant." The male places his enlarged mandibles over the wings of the female for a short time before mating.

Females apparently feed on fruit juices or other liquid food, since the eggs are undeveloped at emergence and require considerable yolk deposition before oviposition. *Corydalus* females are reported to tear apart flowers to feed on nectar. Male dobsonflies do not feed, but they imbibe some water.

FIGURE 4 *Corydalus* (dobsonfly) female. (Photograph by B. M. Drees.)

Females oviposit on objects overhanging the water. Eggs are encased in a white protective material in oval masses of one to five layers and contain over 1000 eggs. A female may deposit two or three egg masses. Adults live for about a week and females die shortly after ovipositing. The incubation period is 2 to 3 weeks.

The life cycle is temperature dependent. In north central Texas, dobsonflies are univoltine. However, northern populations may have a life cycle of 2 to 3 years and are larger as adults.

Subfamily Chauliodinae

There are 16 genera in the subfamily Chauliodinae, and the 18 species account for most of the diversity of megalopterans in America north of Mexico. Five genera and 10 species are known from the Neotropics.

The Chauliodinae have a wide geographical distribution and occur in a wider range of habitats than do sialids or corydalids. Several genera (e.g., *Neohermes, Dysmicohermes, Protochauliodes*) inhabit intermittent streams, whereas others (e.g., *Orohermes*) are found in cold, permanent streams and rivers as well as in the adjacent intermittent tributaries. *Chauliodes* occurs in slow waters, in swamps and ponds, and sometimes in intermittent habitats. These larvae can exploit low-oxygen habitats because the terminal spiracles are on contractile tubes.

The life cycle is also quite variable, with growth and development being temperature and habitat dependent. For example, the life cycle of *Neohermes* varies from 2 to 5 years depending on the duration of flow in temporary streams. Estimates of the number of instars range from 9 to 12. Females may have one more larval molt than males to account for their larger size.

Fishfly larvae are generalist predators, but in intermittent habitats they also may be scavengers, feeding on corpses of individuals stranded by receding waters.

Mating occurs in a tail-to-tail position with no display or special mating behavior (but pheromone attraction may be involved in some genera). The male *Nigronia* backs to the female, the genitalia are clasped, and they remain in copulation for several hours. Spermatophore transfer has not been observed for *Nigronia* or *Corydalus,* although it is reported for *Sialis.*

The final instar of *N. serricornis* leaves the water to pupate in a shallow cell that is dug into the soil. The exarate pupa is light colored but darkens in the last 2 days before emergence. The pupa is active; it turns in the cell, which helps in maintaining cell structure. Pupae do not feed but can use the sizable mandibles and will bite an object that is placed in contact. The pupal period ranges from 2 to 4 weeks depending on temperature.

Like most megalopterans, *Nigronia* adults are awkward and weak in flight. They fly throughout the day, with oviposition occurring in the afternoon. *Nigronia* adults readily take fluids, especially sugar solution and mashed fruit. Adult feeding is needed because some individuals live for 2 weeks.

Reproductive organs in *Nigronia* are mature at emergence, as is the case with *Sialis* but unlike *Corydalus.* A bimodal occurrence in egg deposition demonstrated that after their first oviposition females waited a week to lay a second mass. Egg masses are composed of one to five layers. Maturation of *Nigronia* eggs requires 2 to 3 weeks. Hatching begins with individuals in the uppermost layer and continues to the lower level.

Larvae of fishflies and corydalids often have colonies of epizooics attached to the exoskeleton. These growths, which are readily apparent at low magnification, include stalked protozoans, rotifers, and filamentous algae. Phoretic Chironomidae, for example, *Plecopteracoluthus downesi,* are often found strapped in cases attached to thoracic sternites.

See Also the Following Articles

Aquatic Habitats • Neuroptera • Raphidioptera

Further Reading

Azam, K. M., and Anderson, N. H. (1969). Life history and habits of *Sialis rotunda* and *S. californica* in western Oregon. *Ann. Entomol. Soc. Am.* **62,** 549–558.

Brigham, W. U. (1982). Megaloptera. *In* "Aquatic Insects and Oligochaetes of North and South Carolina" (A. R. Brigham, W. U. Brigham, and A. Gnilka, eds.). Midwest Aquatic Enterprises, Mohamet, IL.

Brown, A. V., and Fitzpatrick, L. C. (1978). Life history and population energetics of the dobsonfly, *Corydalus cornutus. Ecology* **59,** 1091–1108.

Contreras-Ramos, A., and Harris, S. C. (1998). The immature stages of *Platyneuromus* (Corydalidae), with a key to the genera of larval Megaloptera of Mexico. *J. North Am. Benthol. Soc.* **17,** 489–517.

Evans, E. D. (1972). A study of the Megaloptera of the Pacific coastal region of the United States. Ph.D. thesis, Oregon State University.

Evans, E. D., and Neunzig, H. H. (1996). Megaloptera and aquatic Neuroptera. *In* "Aquatic Insects of North America" (R. W. Merritt and K. W. Cummins, eds.). 3rd ed., Chap. 16. Kendall/Hunt, Dubuque, IA.

Kukalova-Peck, J. (1991). Fossil history and the evolution of hexapod structures. *In* "The Insects of Australia" (I. D. Naumann *et al.,* eds.), 2nd ed., Chap. 6. Melbourne University Press, Melbourne.

McCafferty, W. P. (1981). Fishflies, dobsonflies, and alderflies (order Megaloptera). *In* "Aquatic Entomology," Chap. 11. Science Books International, Boston.

Oswald, J. D., and Penny, N. D. (1991). Genus-group names of the Neuroptera, Megaloptera, and Raphidioptera of the world. Occasional Paper No. 147 of the California Academy of Sciences.

Petersen, R. C. (1974). Life history and bionomics of *Nigronia serricornis* (Say) (Megaloptera: Corydalidae). Ph.D. dissertation, Michigan State University.

Theischinger, G. (1991). Megaloptera. *In* "The Insects of Australia" (I. D. Naumann *et al.,* eds.). 2nd ed., Chap. 32. Melbourne University Press, Melbourne.

Metabolism

S. N. Thompson
University of California, Riverside

R. K. Suarez
University of California, Santa Barbara

Metabolism refers to the thousands of chemical reactions that occur in the cell. These reactions link together in defined series to form pathways. Metabolic pathways are interdependent and exquisitely regulated for the efficient extraction of energy from fuels, catabolism, and the synthesis of biological macromolecules, anabolism. Metabolism is a subject area of biochemistry, which also includes the structural chemistry of biological molecules and the chemistry of molecular genetics, the chemical processes involved in the storage and inheritance of biological information.

The biological diversity of insects has enabled focused study on the metabolic bases for many physiological capabilities that are unique to insects or their arthropod and close relatives. The literature on insect biochemistry is extensive. Early studies of insect biochemistry focused on chemical content, individual chemical reactions, metabolic rate, and respiration. Much of this was discussed in the seven editions of Sir V. B. Wigglesworth's *The Principles of Insect Physiology,* which first appeared in 1939. As advances were made other comprehensive treatments appeared, including the 1964 edition of M. Rockstein's *Physiology of Insecta* series and D. Gilmour's *The Biochemistry of Insects,* which first appeared in 1961. *The Biochemistry of Insects,* edited by M. Rockstein, appeared in 1978. More recently, insect metabolism was described in several volumes of the treatise *Comparative Insect Physiology, Biochemistry and Pharmacology,* edited by G. A. Kerkut and L. I. Gilbert (1985).

INTERMEDIARY METABOLISM

Insects share with other invertebrates most of the common pathways of carbohydrate, lipid, and amino acid metabolism. Although much has been presumed based on overt similarities to more extensive studies of mammals and other higher taxonomic groups, many aspects of intermediary metabolism have been examined in a number of insects and different insect tissues. Much of intermediary metabolism, including synthesis and storage of carbohydrate and fat, takes place in the fat body.

The metabolism and utilization of the glucose disaccharide trehalose as the principal hemolymph or blood sugar is unique to insects and some other invertebrates. Unlike glucose, trehalose is a nonreducing sugar, a sugar not readily oxidized by common oxidizing agents. First described from an insect by G. R. Wyatt and his associates in pupae of the polyphemus moth, *Antheraea polyphemus,* trehalose occurs in many insects at variable but high levels. In lepidopteran insects, trehalose levels are commonly as high as 100 mM. The concentration of trehalose in insects often greatly exceeds levels of glucose in the blood of mammals. Blood glucose in humans typically is about 5 mM, a value that would be considered very low for trehalose in hemolymph. With few exceptions, glucose occurs in insect hemolymph at levels less than 5 mM and often at less than 1 mM. Trehalose serves multiple functions, as a storage carbohydrate that serves as a fuel for flight and as a cryoprotectant, protecting insects from damage during overwintering

in cold climes. The hemolymph level of trehalose plays an important role in regulating carbohydrate intake and maintaining nutritional homeostasis. Levels of trehalose in the hemolymph are maintained by a complex interaction of nutrient intake and metabolism.

Trehalose is synthesized in the fat body from two metabolic intermediates of glycolysis, glucose 1-phosphate and glucose 6-phosphate. The reactions synthesizing trehalose are catalyzed by trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase. The sources of glucose for trehalose synthesis include dietary sucrose, glycogen, and gluconeogenesis, dietary sugar being the sole source of glucose under fed conditions. Trehalose formation from glycogen has been described in several insects, including the American cockroach *Periplaneta americana* and tobacco hornworm *Manduca sexta* during starvation. The breakdown of glycogen to glucose is due to activation of the enzyme glycogen phosphorylase, first demonstrated by J. E. Steele and his associates to be under endocrine control by a neurohormone released from the corpora cardiaca in the brain. The induction of a "hypertrehalosemic" hormone RNA transcript in the central nervous system of the cockroach *Blaberus discoidalis* in response to starvation was recently demonstrated. Glucose synthesis, followed by trehalose formation, via gluconeogenesis has been reported only in *M. sexta* and was induced when nymphs were maintained on low-carbohydrate diets. Starvation did not induce gluconeogenesis.

Most insects obtain energy principally from aerobic respiration, but many species have some capacity for anaerobic energy metabolism when exposed to hypoxic or anoxic conditions. This is best known in aquatic insects such as midge larvae, in which the fermentation products may include lactate, ethanol, and acetate. For example, the midge *Chaoborus crystallinis* accumulates succinate, suggesting that this species is capable of anaerobic respiration, possibly involving fumarate reductase for ATP production. Polyol formation during diapause of some insects is another example of anaerobic metabolism.

CHEMISTRY AND METABOLISM OF SCLEROTIZATION

The hardness of cured exocuticle is the result of cross-linking or polymerization between molecules of the protein sclerotin and/or cross-linking between sclerotin and chitin. The chemical composition of sclerotin and the chemistry of the polymerization process were first understood from organochemical analyses of the products resulting from chemical degradation of cuticle. Current knowledge of cuticular structure came about through nondestructive nuclear magnetic resonance (NMR) analysis of intact cuticle. Although sclerotization is principally the result of the reaction of cuticular proteins and chitin with quinones derived from *N*-acetyldopamine, the metabolism of sclerotization is complex and sclerotized cuticles vary greatly in their physical and chemical properties because of metabolic variation.

The metabolism involves principally quinone methide intermediates first described by M. Sugumaran and his collegues. Reactions of quinone methides produce so-called β-sclerotin, with cross-linkages involving the β carbon of the catecholamine side chain, as well as dehydrodopamine intermediates producing linkages to the benzoyl ring. The metabolism of the quinone and dehydrodopamine intermediates is integrated, such that both the side chain and the ring structure may be involved in an individual cross-linked structure. Phenoloxidases, quinone isomerases, and quinone tautomerases are the principal enzymes involved in the metabolism of these various metabolic intermediates.

The structure of the cross-linkage between sclerotin and chitin in the pupal cuticle of *M. sexta* was recently established by K. Kramer, J. Schaefer, and associates through the use of solid-state NMR spectroscopy. As shown in Fig. 1, the structure demonstrates cross-linkage of phenolic and quinone intermediates with the imidizole ring of histidine residues of the protein and the β-hydroxyl group on Carbon 4 of *N*-acetylglucosamine units of chitin.

CHEMISTRY AND METABOLISM OF PIGMENTS

The diversity of insect coloration is in large measure because of an abundance of pigments. Combinations of pigments together with effects of light diffraction, refraction, and interference involving various anatomical structures produce the array of exotic colors familiar to insect observers. Many insects synthesize melanins, ommochromes, porphyrins, pteridines, and/or quinones. Other pigments such as flavonoids and carotenoids, although not synthesized, are often sequestered by insects from plants and contribute to coloration. Two pigment groups are notable: ommochromes, first reported from the eyes of insects, and papilochromes, a unique group

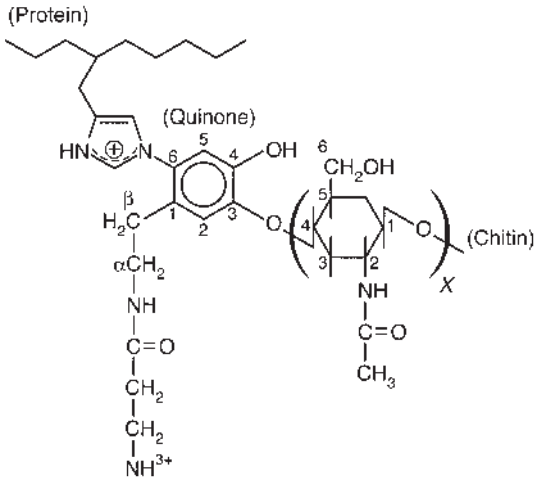

FIGURE 1 Structure of the protein–chitin cross-linkage in the pupal cuticle of *M. sexta.* Protein may be linked through the 1 or the 3 nitrogen of the imidizole ring to the 2, 5, or 6 ring carbon of the quinone derivative, and carbon 4, or other carbons of chitin may be linked to phenoxy carbon 3 or 4 of the quinone. (Adapted from Schaefer *et al.,* 1987.)

that occurs in the bodies and wings of the butterfly family Papilionidae.

Ommochromes are polymers of heterocyclic phenoxines, distributed among a variety of different insect tissues, producing yellow, red, and brown coloration. They are synthesized from tryptophan in a metabolic pathway involving kyneurenine derivatives. In the compound eye, ommochromes form the principal masking pigments that surround and isolate the individual ommatidia and thus, the origin of the name. Several eye-color mutants, described in several insect species, result from the absence of enzymatic function at specific steps in the synthetic pathway. Identification of these steps in *Drosophila* was one of the early comfirmations of A. Garrod's one gene–one enzyme hypothesis. The ommochrome biosynthetic pathway in the coloration of *M. sexta* larvae is hormonally regulated.

Papiliochromes are novel white, yellow, and red pigments whose synthesis intersects the well-known metabolic pathways the melanins and ommochromes. For butterflies of the genus *Papilio,* the precursors are β-alanine, tyrosine, and tryptophan. Papiliochromes accumulate in the wing scales and their distribution varies with the butterfly species. Recent studies on papiliochrome synthesis demonstrated that, as in the case of sclerotization, quinone methides derived from tyrosine are intermediates. The synthesis involves the nonenzymatic condensation of *N*-β-alanyldopamine quinone methide with L-kynurenine to produce a mixture of two diastereoisomers of papilochrome II, a white pigment. Papiliochrome II is a peptide in which the two aromatic rings are linked by a bridge between the aromatic amino group of kynurenine and the catecholamine side chain of norepinephrine derived from the quinone.

ENERGY METABOLISM DURING FLIGHT

Insect flight muscles are obligately aerobic, deriving energy from O_2-dependent substrate oxidation to CO_2 and H_2O. Small insects in flight achieve the highest known mass-specific rates of aerobic metabolism among animals. Of the estimated one-half million insect species capable of flight, the metabolism of only a few has been subjected to detailed examination. Insect species differ in the extent to which carbohydrates (principally trehalose), fats (mainly diacylglycerol), and proline (an amino acid) are used to fuel flight. A scheme summarizing the relevant pathways is shown in Fig. 2.

In some species of locusts and moths, flight is initially fueled by carbohydrate. Prolonged flight follows activation of

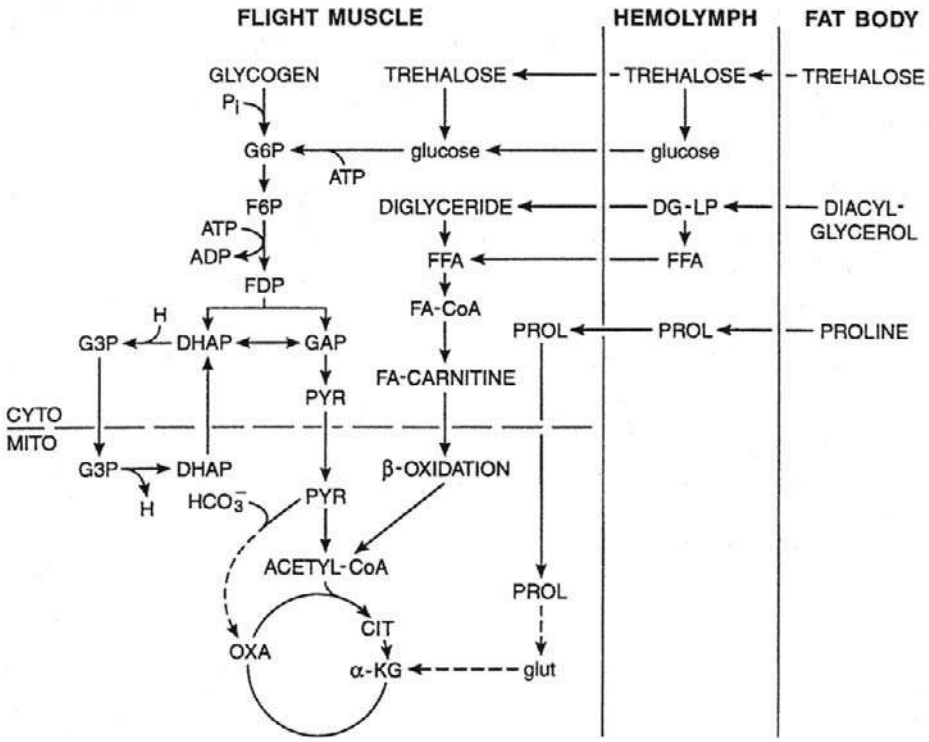

FIGURE 2 Metabolic scheme showing pathways of carbohydrate, fat, and proline oxidation in insect flight muscles. Included are the anaplerotic roles of pyruvate carboxylation and proline oxidation in some species, as well as the β-glycerophosphate shuttle (involving G3P and DHAP) for transferring reducing equivalents from cytoplasm to mitochondria. The contribution of each pathway varies according to species. In some species, substrate use may change with time in flight. Acetyl-CoA, acetyl coenzyme A; ADP, adenosine diphosphate; ATP, adenosine triphosphate; CIT, citrate; CYTO, cytosol; DG-LP, lipoprotein-bound diacylglycerol; DHAP, dehydroxyacetone phosphate; F6P, fructose 6-phosphate; FDP, fructose bisphosphate; FA-carnitine, fatty acid–carnitine; FA-CoA, fatty acid–coenzyme A; FFA, free fatty acids; G3P, glycerol 3-phosphate; G6P glucose 6-phosphate; GAP, glyceraldehyde 3-phosphate; glut, glutamate; HCO_3^-, bicarbonate ion; H, nicotinamide adenine dinucleotide; α-KG, ketoglutarate; MITO, mitochondrion; OXA, oxaloacetate, PROL, proline; PYR, pyruvate. (Adapted from Storey, 1985.)

fatty acid oxidation and inhibition of carbohydrate oxidation, which are processes triggered by adipokinetic hormones. Bees appear unable to make use of fats to fuel flight, but instead oxidize carbohydrate, while using proline as a carbon source for augmenting the concentration of Krebs cycle intermediates. Pyruvate carboxylation serves a similar, anaplerotic role in these species. Carbohydrate and proline are oxidized simultaneously to support steady-state flight by flies and some beetles, but there is considerable interspecific variation as to which fuel is used as the primary carbon source. The tsetse fly may be unique in using proline as the sole fuel for flight.

A huge increase in the rate of ATP utilization in muscle occurs during the transition from rest to flight. In species possessing synchronous flight muscles (e.g., locusts, butterflies, and moths), this is the result of increased activities of actomyosin-ATPase, Ca^{2+}-ATPase, and Na^{+}/K^{+}-ATPase. In insects with asynchronous muscles, actomyosin-ATPase accounts for most of the ATP hydrolyzed, and the energetic cost of excitation–contraction coupling (involving Ca^{2+}-ATPase and Na^{+}/K^{+}-ATPase) is lower in comparison with synchronous muscles. Contraction frequencies in asynchronous muscles are not limited by maximum rates of Ca^{2+} cycling, which may explain why asynchronous muscles, which are unique to insects, are found in 75% of insect species, including flies, beetles, and many species of wasps and bees.

ATPase activities increase upon the initiation of flight, and metabolic control mechanisms activate pathways of substrate catabolism, mitochondrial respiration, and oxidative phosphorylation to maintain ATP concentration and muscle function. Cellular rates of ATP hydrolysis and synthesis in insect flight muscle are so exquisitely matched that ATP concentrations are maintained within narrow limits, despite up to 100-fold increases in the rate of ATP turnover during the transition from rest to flight. The regulatory mechanisms that make possible these large flux changes between rest and flight remain poorly understood, and until recently, how such phenomenal steady-state rates of aerobic metabolism are achieved was largely unknown. Recent work has revealed that these are made possible by high concentrations of catabolic enzymes in the flight muscles, the operation of some of these enzymes at high fractional velocities, high mitochondrial content, large inner membrane surface areas per unit mitochondrial volume, and unusually high rates of electron flow between respiratory chain enzymes. Evolution has produced no locomotory muscles capable of higher rates of aerobic metabolism.

METABOLISM OF COLD HARDINESS

Many insects that exhibit arrested development during periods of cold, or routinely live in cold environments, display a resistance or tolerance to freezing. The phenomenon of supercooling was described for various insects over 100 years ago. Although supercooling temperatures are generally –10 to –15°C, supercooling temperature may approach –50°C. In most cases, low-molecular-weight cryoprotectants, including some amino acids, mannitol, trehalose, sorbitol, and particularly glycerol, accumulate in the hemolymph and act as antifreeze. Cryoprotectants typically reach levels of 20 to 30% or more of fresh body weight as temperature falls and the level of cryoprotectant often is inversely related to the supercooling point. In freeze-tolerant species ice-nucleating proteins accumulate and the supercooling point increases.

Among the first metabolic studies on cryprotectant synthesis were those of H. Chino on glycerol and sorbitol formation in diapausing *Bombyx mori* silkworm eggs. The metabolism is hormonally regulated and is summarized in Fig. 3.

Although most insects synthesize polyols from glycogen in a similar manner, their synthesis in *Bombyx* eggs not only provides cryoprotection, but also maintains redox balance in response to a reduced level of oxidative respiration. In this case, the formation of polyols is an example of anaerobic fermentation.

Regulation of polyol formation is poorly understood, but cold-induced activation of glycogen phosphorylase is involved. Studies with *Eurosta solidaginis* fly larvae demonstrate independent control over glycerol and sorbitol formation, and hormonal cues may be important. In insects that accumulate trehalose for cryoprotection, inhibition of phosphofructokinase has been suggested as the mechanism for shifting metabolism in the direction of sugar formation. A temperature-dependent change in the balance between glucose oxidation by glycolysis and the pentose phosphate pathway may also affect the balance between trehalose, sorbitol, and glycerol.

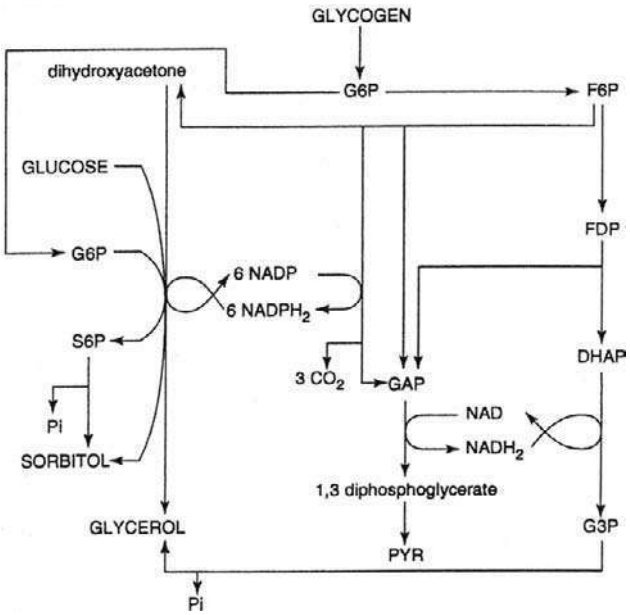

FIGURE 3 Metabolic scheme showing pathways of carbohydrate oxidation and polyol synthesis in *B. mori* silkworm eggs. CO_2, carbon dioxide; NAD, oxidized nicotinamide adenine dinucleotide; $NADH_2$, reduced NAD; Pi, inorganic phosphate; S6P, sorbitol 6-phosphate. Other abbreviations as in Fig. 2. (Adapted from Gilmour, 1965, "The Metabolism of Insects." W. H. Freeman, San Francisco.)

In conclusion, insects exhibit many unique metabolic characteristics. The brevity of this article allows discussion of only a few select examples. Much current molecular work focuses on silk fibroin synthesis and the molecular action of hormones. Relatively few investigations have been conducted on molecular-genetic aspects of metabolism and metabolic regulation. The available techniques in cell and molecular biology, including differential display and microarray analysis, offer marvelous opportunities to examine the "transcriptional physiology" of hormonal and nutritional regulation of intermediary metabolism. The metabolic diversity of the Insecta also awaits more detailed exploration within the context of their ecology and behavior.

See Also the Following Articles

Cuticle • Digestion • Fat Body • Flight • Muscle System • Temperature, Effects on Development and Growth • Thermoregulation

Further Reading

Bursell, E. (1981). The role of proline in energy metabolism. *In* "Energy Metabolism in Insects" (R. G. H. Downer, ed.), pp. 135–154. Plenum, New York.

Candy, D. J. (1985). Intermediary metabolism. *In* "Comprehensive Insect Biochemistry, Physiology and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 10, pp. 1–41. Pergamon Press, New York.

Harshman, L. G., and James, A. A. (1998). Differential gene expression in insects: Transcriptional control. *Annu. Rev. Entomol.* **43,** 671–700.

Hoffman, K. H. (1984). Color and color changes. *In* "Environmental Physiology and Biochemistry of Insects" (K. H. Hoffman, ed.), pp. 206–224. Springer-Verlag, Heidelberg.

Josephson, R. K., and Young, D. (1987). Fiber ultrastructure and contraction kinetics in insect fast muscles. *Am. Zool.* **27,** 991–1000.

Josephson, R. K., Malamud, J. G., and Stokes, D. R. (2000). Asynchronous muscle: A primer. *J. Exp. Biol.* **203,** 2713–2722.

Kayser, H. (1985). Pigments. *In* "Comprehensive Insect Biochemistry, Physiology and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 10, pp. 368–415. Pergamon Press, New York.

Lee, R. E., and Denlinger, D. L. (1991). "Insects at Low Temperature." Chapman & Hall, New York.

Lewis, D. K., Bradfield, J. Y., and Keeley, L. L. (1998). Feeding effects on gene expression of the hypertrehalosemic hormone in the cockroach, *Blaberus discoidalis. J. Insect Physiol.* **44,** 967–972.

Linzen, B. (1974). The tryptophan/ommochrome pathway in insects. *Adv. Insect Physiol.* **10,** 117–246.

Moore, M. V., and Lee, R. E. (1991). Surviving the big chill: Overwintering strategies of aquatic and terrestrial insects. *Am. Entomol.* Summer, 111–118.

Schaefer, J., Kramer, K. J., Garrow, J. R., Jacob, G. S., Stejskal, E. O., Hopkins, T. L., and Speirs, R. D. (1987). Aromatic cross-links in insect cuticle: Detection by solid-state ^{13}C and ^{15}N NMR. *Science* **235,** 1200–1204.

Simpson, S. J., and Raubenheimer, D. (1993). The central role of the haemolymph in the regulation of nutrient intake in insects. *Physiol. Entomol.* **18,** 395–403.

Storey, K. B. (1985). Metabolic biochemistry of insect flight. *In* "Circulation, Respiration and Metabolism" (R. Gilles, ed.), pp. 193–207. Springer-Verlag, Berlin/Heidelberg.

Storey, K. B., and Storey, J. M. (1991). Biochemistry of cryoprotectants. *In* "Insects at Low Temperature" (R. E. Lee and D. L. Denlinger, eds.), pp. 64–93. Chapman & Hall, New York.

Suarez, R. K. (2000). Energy metabolism during insect flight: Biochemical design and physiological performance. *Physiol. Biochem. Zool.* **73,** 765–771.

Sugumaran, M. (1998). Unified mechanism for sclerotization of insect cuticle. *Adv. Insect Physiol.* **27,** 229–334.

Metamorphosis

Frederick W. Stehr
Michigan State University

Metamorphosis means change in form. Most organisms undergo a change in form as they grow from an embryo to an adult. Some changes are radical and the immatures bear little resemblance to the adults; others are more gradual, with the immatures looking very much like the adults.

The term "larva" has very broad usage in invertebrate zoology, being applied to an assortment of forms (often the dispersive state) in virtually all invertebrate phyla. In Arthropoda other than insects, larvae is most often used for first stages, as it is in the mites and ticks, and for the first-stage hexapod larvae of millipedes, with "nymph" being used for second-stage mites and ticks. However, in continental Europe (especially France) "nymphe" refers to a pupa; in English-speaking countries "pupa" is universally used for the stage between the last instar and the adult of insects with complete metamorphosis (Holometabola).

In the insects, larva has been used in different ways, including such diverse forms as the immatures of the most primitive order Protura and of the most advanced order Hymenoptera (sawflies, ants, wasps, and bees). The termites present an interesting problem: some authors use nymph for all juvenile termites, whereas others use larva for those lacking wingpads, and nymph for those having wingpads. This is further complicated because the supplementary reproductives may be wingless or bear wingpads, even though the two forms are functionally equivalent.

The kind of metamorphosis insects undergo is closely related to which of the subclasses a taxon belongs to. In the Apterygota metamorphosis is either anamorphic in the Protura, in which three abdominal segments are added as the individual develops to an adult, or it is ametabolous, in which the number of molts is indefinite and molting may continue throughout life after sexual maturity (Collembola, Diplura, Archeognatha, and Zygotrema).

In the Pterygota there are two fundamental kinds of metamorphosis: the hemimetabola develop through the egg, larva, and adult stages, and the holometabola develop through the egg, larva, pupa, and adult stages. There are also other terms (defined below) that have been used to describe variations in metamorphosis.

LARVAE VS NYMPHS AND NAIADS

Defining a larva is also necessary because its use has been highly variable in the Insecta. In 1918, Comstock proposed

restricting the term larva to juveniles of the holometabolous orders, nymph to the juveniles of his paurometabolous (nonholometabolous) terrestrial orders, and naiad to the juveniles of his hemimetabolous nonholometabolous aquatic orders (Ephemeroptera, Odonata, and Plecoptera). Because these three aquatic orders have a much greater change in form from the last instar to the adult than the terrestrial hemimetabolous orders, there was some basis for Comstock's proposal to call them naiads. However, the Ephemeroptera and Odonata are Paleoptera, which cannot fold their wings, whereas the Plecoptera are Neoptera, which can fold their wings over their back, so they are not closely related.

Currently there is a tendency to use larva for all immature insects that are not eggs, pupae, or adults and the term "immature insect" for all life stages except adults, no matter how many specialized names are applied to the various developmental forms in the different orders.

There is little difficulty in defining an egg or an adult, but naming and defining the instars or stages that may occur between egg and adult can be problematic. Some insects are larviparous, never depositing eggs; some multiply from a single egg by polyembryony, and some are sexually mature as immatures (paedogenesis or neoteny). Nevertheless, all of them undergo a series of molts as they grow. When larva is used in the comprehensive sense, the subcategories "exopterygote larva" (Hemimetabola that have the wingpads developing externally) and "endopterygote larva" (Holometabola that have the wingpads appearing externally in the pupal stage but having developed from internal larval histoblasts) are useful for pterygote immatures. A useful term roughly equivalent to larva in the comprehensive sense is "juvenile," which can be used as a general term for nonadult larvae of all orders.

KINDS OF METAMORPHOSIS

Below are terms that are currently widely used for different types of metamorphosis. Most species are either holometabolous or hemimetabolous, with more than 85% of them holometabolous and most of the rest hemimetabolous.

Anamorphosis

This term means development with fewer body segments at hatching than when mature, which is found in the Protura, in which three abdominal segments are added anterior to the tail as the individual develops to an adult. Because of this some workers believe proturans are not true insects.

Ametabolous

Ametabolous means development with the major change being an increase in size until sexually mature. The number of molts is indefinite, and molting may continue throughout life; it is found in Apterygota (excluding the Protura).

Simple Metamorphosis

This is a broad term covering all types of metamorphosis except holometabolous.

Hemimetabolous (Gradual, Incomplete, Direct, Paurometabolous)

Development through egg, larva, and adult is covered by this term, which includes everything except ametabolous, anamorphosis, and holometabolous. Among the hemimetabolous insects, most species are found in three orders, the Orthoptera, the Heteroptera, and the Homoptera.

Holometabolous (Complete, Indirect)

This means development through egg, larva, pupa, and adult.

All insects do not fit neatly into Hemimetabola or Holometabola. Some Hemimetabola are intermediate in having one or more nonfeeding stages before the adult instar and in having a last instar that forms into a pupa. For example, in the whiteflies (Homoptera: Aleyrodidae) the first instar is active, but subsequent instars are stationary, resembling and feeding like scale insects, and the last instar stops feeding and becomes a pupa, with the wings developing internally. This could be termed holometabolous, but the homopterans as a whole are certainly hemimetabolous. In the Thysanoptera (thrips), there are two feeding instars followed by two or three nonfeeding instars, the propupa and pupa, which may be contained within a cocoon formed by the last feeding instar. This is certainly closer to Holometabola than to Hemimetabola.

See Also the Following Articles

Development, Hormonal Control of • *Hypermetamorphosis* • *Molting*

Further Reading

Balls, M., and Bownes, M. (eds.) (1985). "Metamorphosis." Clarendon Press, Oxford.

Comstock, J. A. (1918). Nymphs, naiads and larvae. *Ann. Entomol. Soc. Am.* **2,** 222–224.

Etkin, W., and Gilbert, L. I. (eds.) (1968). "Metamorphosis." Appleton–Century–Crofts, New York.

Hall, B. K., and Wake, M. H. (eds.) (1999). "The Origin and Evolution of Larval Forms." Academic Press, New York.

Stehr, F. W. (ed.) (1987). "Immature Insects," Vol. 1. Kendall/Hunt, Dubuque, IA.

Stehr, F. W. (ed.) (1992). "Immature Insects," Vol. 2. Kendall/Hunt, Dubuque, IA.

Migration

Hugh Dingle
University of California, Davis

Migration, the major movement behavior of insects, allows them to escape deteriorating habitats, to colonize new areas, or to seek temporary shelter such as overwintering sites. It involves a complex of traits that include development, physiology, morphology, and behavior, and it is a major component of the life histories of many species. These trait complexes or syndromes are adjusted by natural selection in complex ways that increase the fitness and therefore the success of migrants.

MIGRATION AND OTHER MOVEMENTS

The movements characteristic of organisms can be roughly divided into two broad categories, those that are immediately responsive to resources and those that are not (Box 1). Within the category of immediate responses to resources are two further broad types. The first type consists of "station-keeping" responses that serve to keep the organism on its territory or within the home range in which it carries out most of its life functions and spends most of its time. Included within station-keeping movements are resource-sensitive behaviors crucial for survival. Examples are foraging, territorial behavior, and commuting, which is a periodic, often daily, round trip for resources. Foraging may be for any resource, including food, shelter, or mates; and commuting, which can also be considered to be a form of extended foraging, may involve travel over considerable distances. The commuting trails of leafcutter ants, for example, may extend for hundreds of meters both horizontally along the forest floor and vertically into the canopy. Foraging, commuting, and territorial behaviors are all readily responsive to resources: thus a female butterfly stops searching (foraging) upon encountering a host plant on which to oviposit, and a territorial forest drosophilid fly is bounded by the borders of its leaf display ground.

Ranging movements take organisms on exploratory journeys beyond the current home range and serve to locate a new home range or territory. Like station-keeping movements, ranging movement is a facultative response to resources, and like foraging or commuting, it ceases when a new resource (here in the form of previously unoccupied living space) is encountered. Ranging movements may also extend considerable distances but, like station keeping, still belong in the category of activities that are proximately resource sensitive.

Movement that is not immediately responsive to resources constitutes the distinct sort of behavior that is migration. Taking an organism beyond both its current home range and beyond neighboring potential home ranges, migration is physiologically distinct from all other movements. It is so distinct because sensory inputs from resources that would ordinarily cause movement to cease do not stop migration. Thus, a characteristic of migration is that the organism undertaking it is undistracted by and fails to respond to food or mates, otherwise so necessary a part of life functions. Furthermore, migration is usually triggered by environmental cues, such as photoperiod, which forecast habitat change rather than being directly responsive to the change itself (usually a deterioration in the quality or availability of resources). Other characteristics of migration include distinct initiating and terminating behaviors. Many insects climb to the top of a bush or tree branch to take off on migratory flights, behavior they show at no other time. Sensory responses may also change, as in aphids that are sensitive to blue light from the sky during the takeoff phase of migratory flight but become increasingly sensitive to yellow light, the characteristic wavelength of young host plants, as migration proceeds. Thus, migration is not defined by the distance traveled or by whether it is a "round trip." Rather, it is defined in terms of the physiological and behavioral responses to resources; this behavior is true of all organisms, not just insects.

The movement behavior of individuals also has an outcome for the population of which those individuals are a part. This outcome involves displacement for a greater or lesser distance, but at either distance it involves removal from the home range. It can also result in the scattering or dispersal of individuals within the population; thus "dispersal" is a population phenomenon, not an individual movement. Movement can also result in congregation by mutual attraction or aggregation in a habitat. Both tendencies result in a decrease in the mean distance among individuals and contrast with dispersal, which increases mean distances. Note that all the movements just described can contribute to aggregation, congregation, or dispersal, depending on species and ecological circumstance.

Three examples of the sorts of population outcome attributable to migratory behavior occur in locusts, aphids,

Box 1. Types of Insect Movement

Migration is a type of movement displayed by insects, but it differs from all other types because migratory insects (and other migrants as well) are unresponsive to suitable resources. There are two broad categories of movements:

I. Movements that are directed by resources and/or home range.
   A. Station keeping: examples are foraging, commuting (periodic, usually daily, movements), and territorial behavior.
   B. Ranging: movement to explore an area, often for a new home range or territory.

II. Movement not directly responsive to a resource or home range: here migration is undistracted movement with cessation primed (thresholds lowered) by the movement itself. Responses to resources are suppressed or suspended.

and moths of the genus *Heliothis* (and the very similar *Helicoverpa*). Locusts are a group of grasshopper species that under crowded conditions undergo remarkable behavioral and morphological changes known as phase transformation. In the desert locust, *Schistocerca gregaria,* perhaps the most extreme example of the phenomenon, crowded nymphs (hoppers) are strikingly black and yellow, whereas isolated individuals are a pale brown or green. Crowded adults are larger and display differences in body proportions that readily distinguish them from their isolated counterparts. It is behavior, however, that most distinguishes the two forms. Isolated individuals display no mutual attraction, forage independently, and migrate at night. Crowded individuals show a high degree of mutual attraction and form large swarms that can number in the millions. When a swarm is feeding, locusts at the rear are constantly running out of food, overflying the body of the swarm, and landing at the leading edge. The result is a "rolling" movement across country in extended foraging. If a swarm enters an area with no forage, it may rise in unison and be carried for some distance by the wind. If this happens for a long period or repeatedly, the individuals in the swarm may switch to migratory behavior and cover considerable distances to descend again in regions of fresh plant growth. These aggregated swarms are major pests over much of Africa, occasionally reaching adjacent areas of the Middle East and in several notable instances the New World. The arrival of a swarm can mean that "not any green thing" (Exodus 10:15) is left for human consumption. It is the aggregation and migration that make the desert locust such a notorious pest. It would be much less a pest if its characteristic behavior led to dispersal rather than swarming.

Various species of aphid are also capable of spreading far and wide by migration. In Europe, an extensive monitoring network coordinated by English and French entomologists has mapped the seasonal spread of bean aphid, *Aphis fabae.* Large concentrations appear first in central France in early to midsummer. The species then spreads westward and northward over succeeding generations so that by late summer the aphid has reached high densities as far north as Scotland. In North America, monitoring of the corn leaf aphid, *Rhopalosiphum maidis,* has indicated the arrival of large numbers in the cornfields of Illinois. Analysis of weather systems suggests these aphids have come from as far away as Texas and were transported on wind streams. Studies of a number of other insects indicate that the Mississippi Valley is a major spring migration route for wind-transported insects to the agricultural regions of the upper Midwest.

Heliothis moths breed following rainfall in tropical and subtropical arid regions. If productivity is high on the new vegetation on which they feed, large populations of migrating moths are produced and are carried by winds to agricultural areas. In Australia, moths are transported in spring to wheat- and cotton-growing regions in New South Wales from breeding areas in interior regions of New South Wales and Queensland. The location of rainfall in the interior of Australia is unpredictable from year to year, and considerable effort has gone into locating areas in which rain has fallen, determining whether this precipitation is sufficient to produce large moth populations, and forecasting the arrival of migrating moths in conjunction with weather systems so that necessary control measures can be undertaken and unnecessary ones avoided. In the spring in North America, there is similar breeding of *Heliothis* moths in northern Mexico and southern Texas and migration northward on winds.

HISTORICAL BACKGROUND

In the 1930s C. B. Williams collected and summarized the available information on insect migration. The two books that resulted were largely responsible for bringing to the attention of entomologists and other biologists the fact that the phenomenon was a common one. Williams focused on large insects such as butterflies and dragonflies, and he adopted the prevailing notion, derived largely from birds, that only round-trip movement could be called "migratory." The way entomologists now think about insect migration is primarily the result of the work of four Britons: C. G. Johnson, J. S. Kennedy, T. R. E. Southwood, and L. R. Taylor, beginning around 1960. Johnson and Kennedy stressed that insect movements vary with respect to physiology and function, and their ideas revamped notions concerning the behavioral and life history aspects of migration. Southwood showed that the type of habitat determines the likelihood of migration among insects, and Taylor noted the importance of movement to the dynamics of populations in both time and space. Combined, the work of all four made explicit that migration is a distinct behavior with consequences for populations.

The distinct nature of migratory behavior was precisely outlined by Kennedy in his studies of the flight of the summer parthenogenetic females of *A. fabae.* He used a flight chamber that allowed him to analyze the responses of free-flying aphids (Box 2). Key aspects of migratory flight that distinguished it from other types of flight were revealed by the flight chamber experiments. The aphids tested were the winged or alate form produced under crowded conditions. The uncrowded wingless females larviposit (bear live young) as soon as they make contact with a suitable host leaf. In contrast, the winged migrants do not larviposit until they have completed at least some flight. Furthermore, landing responses are primed by migratory flight: the longer the flight, the lower the threshold for landing. Finally, there is reciprocal interaction between flight and settling, since settling responses (i.e., probing a leaf to test its suitability and subsequent moving to the underside of the leaf to larviposit) can prime flight if they fail to be completed by attaching via the mouthparts and producing offspring. This flight after incomplete settling may actually be stronger than that occurring at the beginning of migration.

Box 2. The Kennedy Flight Chamber

J. S. Kennedy used this device in studies of insect migration. His experiments analyzed the performance of free-flying aphids and their landing and foraging responses. The lever arm can be twisted to shake the aphid off the platform, forcing it to fly, and it can be rotated out of the light and presented to the flying aphid again at will. Host plant leaves of different ages and leaves of different species of plant can be presented to permit investigators to observe variation in landing responses. Free flight is maintained by wind from the top of the chamber, and the wind speed is varied with the butterfly valve, whose setting thus indicates the strength of flight as measured by rate of climb which is balanced by the downward windspeed. (Figure reproduced from Dingle, H. (1996). "Migration: The Biology of Life on The Move." Used by permission of Oxford University Press, Inc.).

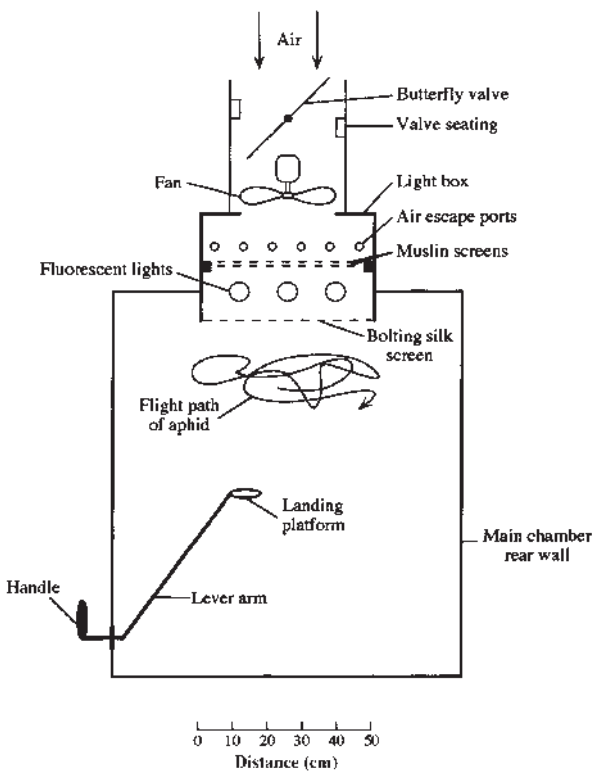

Migration thus is qualitatively different from other movement because station-keeping responses such as landing and probing (foraging) are inhibited by flight, but flight also primes them and promotes their later recurrence. Based on the behavior of migrating aphids, Kennedy provided a complete predictive definition of migration, as follows: migratory behavior is persistent and straightened-out movement effected by the animal's own locomotory exertions or by its active embarkation upon a vehicle; it depends on some temporary inhibition of station-keeping responses, but promotes their eventual disinhibition and recurrence.

By explicitly focusing on the interaction between growth and reproductive behaviors (station keeping) and migratory behavior, Kennedy put migration firmly in the context of life histories. This context was also emphasized by Johnson, who identified the "oogenesis-flight syndrome" as characteristic of insect migration. Johnson noted that in a high proportion of migratory insects, especially in females, flight is limited to individuals with immature reproductive systems. It thus seemed that migration was based on an interaction between flight and the maturation of reproduction. Implicit was the assumption that migration and reproduction were alternative physiological states, with trade-offs in the mobilization of energy and materials. Johnson further postulated that this life history syndrome would be mediated by juvenile hormone, a postulate now amply demonstrated (see later: Migratory Syndromes).

The population dynamical aspects of insect migration were assessed by Southwood and Taylor. Southwood placed migration into an ecological and evolutionary context by summarizing evidence that migration is characteristic of insects living in temporary habitats, such as seasonal pools or early successional fields. This condition is in contrast to that of insects with more permanent habitats such as forests or large lakes; such insects are nonmigratory and often even wingless. This pattern of migration as a response to transitory environments was later formalized by Southwood in 1977 in the ratio H/τ, with H the duration of the habitat and τ the generation time of the insect. The frequency of migration in populations or species increases as the ratio shrinks toward unity, as later nicely demonstrated, for example, in leafhoppers (Homoptera) by Denno and colleagues. Taylor stressed the role of migratory behavior in the spatial dynamics of insect migration. He noted that migrants can disperse or coalesce, depending on whether individuals attract or repell one another and on atmospheric dynamics. This behavior can create a mosaic of insect densities over the landscape. He also initiated the Rothamsted Insect Survey, an array of traps to sample insects in the air, from which data on numbers were taken and analyzed at the Rothamsted Experiment Station near London. This network allowed tracking and forecasting of insect pests such as aphids with major practical implications for insect control.

THE CAUSES OF MIGRATION

Since Southwood's original statement of the relationship, it has become apparent that the impermanence of habitats is indeed the primary selective force driving insect migration. Much of this impermanence is a function of season, and as with other well-known migrants such as many fish or birds, seasonality is a common factor in insect migrations. Most seasonal migrations are on a relatively small scale, with distances traveled only a few hundreds or thousands of meters, but others cover much greater distances. Examples of short-distance migrations

to overwintering diapause sites include the Colorado potato beetle, *Leptinotarsa decemlineata,* several species of common seed-feeding hemipterans, including *Lygaeus kalmii,* a milkweed bug of Europe (in Sweden it often flies to lighthouses to diapause), and several species of lady beetles (Coccinelidae). In some lady beetles migration extends to several kilometers. In California, *Hippodamia convergens,* the convergent lady beetle, overwinters at high altitudes in the Sierra Nevada and migrates to agricultural areas in the Central Valley in March. Beginning in June, offspring of the early spring migrants fly back to intermediate altitudes and form aggregations. There is then a later movement to higher altitudes to overwinter so that the migration to overwintering sites is a two-step process.

Other insect migrants make spectacularly long journeys. The best known of these is that made by the eastern North American population of the monarch butterfly, *Danaus plexippus,* studied extensively by Lincoln Brower and F. A. Urquhart. This butterfly cannot overwinter in the temperate zone, so it must migrate to southern overwintering sites. The short days of autumn cause adult butterflies to enter reproductive diapause, and they undertake a southward journey of 3000 km or more. The majority of the eastern population overwinters en masse in a very few high-altitude protected sites in the Transvolcanic Range of central Mexico, where they arrive in the late autumn. Beginning in February, the aggregations break up, mating occurs, and the butterflies begin to move northward. Identification of the chemical cardenolide "fingerprints" of the milkweeds eaten by monarchs when they are caterpillars and stored in the adults has revealed that the overwintering generation stops and breeds on the spring flush of milkweeds along the coastal plain of the Gulf of Mexico. It is the offspring of these individuals that invade regions farther north beginning in late May and early June. Thus, as with fall California convergent lady beetles, the spring migration is at least a two-step process. A very similar migration pattern occurs in the same region in the large milkweed bug, *Oncopeltus fasciatus,* and it, too, occurs in two stages in the spring.

The migration of western populations of the monarch is more complicated. These populations overwinter in aggregations along the coast of southern California, where winter climatic conditions are similar to the aggregation sites in Mexico. When the aggregations begin to break up, as early as late January, the butterflies move to early sprouting milkweeds in the Coast Ranges and breed there. The next generation moves both to more coastal milkweeds and to milkweeds that grow farther inland as far east as the Rocky Mountains and as far north as the Canadian border, so that as in the eastern populations, the spring migration occurs in two stages over two generations. A very similar pattern occurs in the monarch population introduced into eastern Australia, with overwintering near the coast, a migration inland in the spring, and a return to the coast in the autumn. In more northern parts of the Australian range, there may be year-round breeding in coastal and subcoastal regions.

One way to assess the influence of ephemeral habitats on the evolution of insect migration is to survey across species and populations occurring in different kinds of habitats. In Europe, a number of species and populations of water striders (Hemiptera: Gerridae) occur over an array of habitats, from small, temporary ponds to large lakes and permanent streams to isolated permanent bogs. Species in the more temporary bodies of water have wings and undertake regular migrations to locate their aquatic habitats as they appear and disappear in the landscape. At the opposite extreme in permanent lakes and bogs, there are species that are wingless. Across habitats with varying degrees of permanence are populations and species of water striders with varying proportions of winged and wingless individuals determined primarily environmentally (polyphenisms) where habitat change is predictable, and primarily genetically (polymorphisms), where change is increasingly random with respect to the life cycle.

A second example of the influence of habitat duration on the occurrence of migrants within a fauna occurs in Australian butterflies. Often, latitude predicts the amount of migration that will occur, especially where there is adequate rainfall and seasonal change is largely a function of temperature. In eastern North America, for example, 98% of the variance in the proportion of migratory birds is explained by latitude, with a higher proportion of migrants at northern latitudes. Similarly in eastern Australia, where the climate is warmer overall but still temperate with adequate rainfall, 72% of the variance in the proportion of butterfly migrants is explained by latitude. The situation is quite different in the dry regions of Australia west of the Great Dividing Range. Here latitude accounts for less than 1% of the variance in proportion of butterfly migrants, and climate variables that indicate rainfall patterns, such as soil moisture, which accounts for about 50% of the variance, are much better indicators of migration. The amount of rainfall is not correlated with latitude, and so latitude does not predict migration. In this dry climate it is the availability of erratic rainfall that counts, and only migrants that can take advantage of the ephemeral flushes of vegetation that follow such rainfall. Thus, as with *Heliothis* moths, migration allows some butterflies to exploit a dry and ephemeral habitat.

MIGRATORY SYNDROMES

Accompanying migratory behavior is a syndrome of traits that act in coordination to increase fitness. These traits vary from enzymes to life history characters and, being influenced by subsets of the same genes, are genetically correlated. At the physiological level, insects (like most other migrant organisms) use fat as fuel, primarily for two reasons. First, fat metabolizes to produce about twice as much energy as carbohydrate or protein; and second, it requires no water for storage (in contrast, storage of 1 g of carbohydrate requires 3 g of water). Insects such as the monarch butterfly and the large milkweed bug shift lipids from yolk formation to fat storage under the influence of the shorter photoperiods of autumn and just prior to migration. The flight muscles of migrants are also adapted to the energetic demands

of lengthy flight. Enzymes active in oxidative metabolism, such as citrate synthase, and in fatty acid oxidation, such as β-hydroxyl coenzyme A dehydrogenase, or HOAD, tend to show higher levels of activity in the flight muscle of migrants compared with that shown in nonmigrants. This difference is most apparent where there are wing polymorphisms and migrants have longer wings.

The most important hormone involved with insect migration is juvenile hormone (JH). It has influence not only on the coordination of the various relationships of the oogenesis-flight syndrome but also has direct effects on migratory flight. In many insect migrants such as the monarch butterfly, short photoperiods result in reduced outputs of JH from the corpus allatum. This reduction in JH output in turn leads to a reduction in ovarian and egg development, which is then accompanied by migratory flight. In several species of migrant insects, prolonging of the prereproductive period by reduced JH titers results in the triggering and maintenance of migratory flight. At the same time it has been demonstrated in several migratory species (such as the large milkweed bug, the convergent lady beetle, and the monarch butterfly) that JH directly stimulates migration. Implants of corpora allata, the source of JH, or topical application of JH or some of its chemical analogues, are effective in increasing flight in migrants. In some insects such as the monarch, adipokinetic hormone (AKH—involved in promoting fat metabolism) also stimulates additional flight. The effects of JH and AKH in the monarch are illustrated in Fig. 1.

Because migratory flight occurs when reproduction is delayed by reduced JH concentrations, it is logical to inquire what level of JH determines migration. This question was answered for the large milkweed bug by M. A. Rankin. She selected for delayed onset of flight, which also resulted in delayed reproduction. Rankin measured JH titers in the blood during the prereproductive period and showed that JH titers were low when there was no flight or reproduction; intermediate titers stimulated flight, and high titers stimulated oogenesis. Thus, if titers rose only to intermediate levels, as might occur under short days, for example, flight but not reproduction would be triggered. These JH titers may also be regulated by JH esterase, the enzyme that breaks down JH. In wing-polymorphic crickets, high concentrations of JH result in short-winged individuals. Artificial selection experiments that increased the frequency of long wings also resulted in increased amounts of JH esterase in the blood and so reduced amounts of JH. Selection also demonstrated that it was possible to change both mean and threshold JH esterase activity. The possible role of JH esterase in fully winged migratory insects remains to be studied.

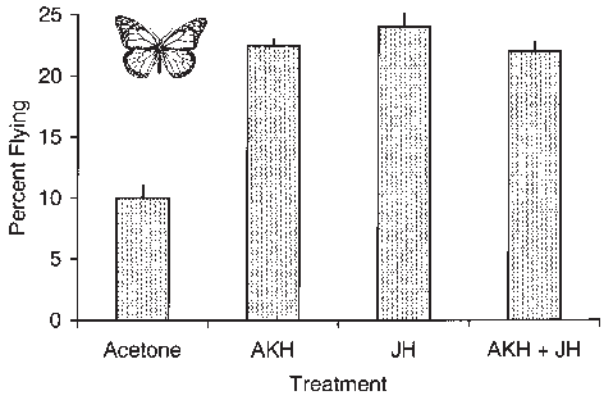

FIGURE 1 Influence of JH and AKH on tethered flight in the monarch butterfly. Flight (%) longer than 30 min is the index of migratory flight. The butterflies received topical application of hormones or the acetone control, and subsequent flight duration was determined. Both hormones, singly or together, increased flight over controls.

An additional behavioral aspect of the migratory flights of many insects is the ability to maintain a more or less constant direction during migration. Mostly, this directionality has been studied in butterflies, although some other large insects such as dragonflies and larger Hymenoptera also seem to maintain a constant direction when migrating. The monarch butterfly in eastern North America flies in a steady southward or southwestward direction in the autumn, flight directions that lead to the overwintering sites in central Mexico. In the spring, the migratory flight is to the north. Compilations of observations of several species of Australian butterflies, including the monarch, reveal that the insects fly south or southwest in the spring and north or northeast in the autumn. The apparently migratory flights of a few species occur in the same direction no matter what the season, a phenomenon that has yet to be explained. In Europe migratory flights of the butterfly *Pieris brassicae* are consistent with both season and geography. Autumn migrants from northern Germany fly south or south by southeast, whereas migrants in the south of France fly to the southwest, which takes them to Spain rather than over the Mediterranean. Further experiments have demonstrated that butterflies that have diapaused, as they do during the winter, fly north when they migrate, but those emerging from nondiapause (summer) pupae fly toward the south. Seasonal winds also frequently carry migrating insects in the "correct" direction. Monarch butterflies in eastern North America frequently soar and are carried southward by northerly winds, and simulations of the migration of the large milkweed bug from the same region indicate that a portion of the population reaches southern overwintering areas regardless of whether they orient.

Where the mechanism of orientation has been studied, the evidence suggests that it is a time-compensated sun compass. To use the sun effectively for orientation, organisms must be able to compensate for its daily passage across the sky by reference to a "biological clock." To demonstrate that an organism is using a time-compensated sun compass, it is necessary to clock-shift it by maintaining it in a daily light cycle that is out of phase with the ambient cycle and to then show that its orientation is displaced by an amount consistent with the clock shift. A displacement of 6 h, for example, should lead to a directional change of 90°; whether

the change is plus or minus depends on the direction of the clock shift. Experiments with southward-moving monarch butterflies suggest that this pattern is indeed followed. In tropical Panama, two species of migrating pierid butterflies, *Aphrissa statira* and *Phoebis argante,* regularly maintain a directional flight across Gatun Lake in the Panama Canal. Clock-shift experiments resulted in changes in the direction of orientation that were consistent with a sun compass, even though there was also a component imposed by wind drift. The sophisticated orientation mechanisms of honey bees, ants, and certain other insects incorporating a sun compass, imply that orientation is probably widespread. The future will undoubtedly reveal the presence of a sun compass in other migrants, as well as the presence of other mechanisms, especially in nocturnal migrants. Radar observations indicate that the passage of many species of nocturnal migrants is specific to winds of a certain direction, but the means by which this preference is enforced are unknown.

In addition to behavioral and physiological characters, migration syndromes often include life history traits such as the age at first reproduction and fecundity, particularly in many wing-polymorphic insects. Typically, in these species the short-winged or wingless forms reproduce earlier and display higher fecundities than their long-winged counterparts. This dichotomy is at least in part because of trade-offs between flight and reproduction. The metabolically active flight muscles that accompany long wings and migration are costly to maintain, requiring considerably more maintenance energy than the thoracic musculature of wingless or short-winged individuals. In contrast, the later reproducing individuals, with lower egg production, are often longer lived.

Migration syndromes that include life history traits are the result of underlying genetic mechanisms, as revealed in artificial experiments using the large milkweed bug. Like all flying insects, this migrant can be induced to fly by removing substrate contact. Bugs that are glued at the prothorax to a tether will fly if contact with the tarsi is removed. An individual in the migratory state can fly on the tether for several hours, and the duration of flight can be used as an index of migration. Artificial selection can be used to increase the proportion of individuals making long (or short) flights, with the duration of flights also affected. Selection was used to both increase and decrease the proportion of bugs undertaking long flights. In addition to flight, wing length and fecundity responded to this selective regime. The bugs of the line with a higher proportion of long flights also had longer wings on average, and the females of this line produced more eggs during the first 5 days of reproductive life. This means that the genes influencing flight also influenced wing length and fecundity, most likely via pleiotropic effects. Longer term selection experiments on wing length, which also resulted in higher fecundities and increased flight as wing length increased, suggested that linkage disequilibruim is unlikely. Parallel selection experiments on a population that did not migrate failed to reveal genetic correlations among wing length, flight, and fecundity, indicating that the genetically based syndrome of correlations among these traits is unique to migratory populations. The selection experiments reveal that natural selection has produced an adaptive migratory syndrome that includes wing length and fecundity. Interestingly, the age at first reproduction is unaffected by selection.

The conclusion from the brief survey of insect migration is that this behavior involves more than simply extended movement to escape to a new habitat. Rather, migration is a trait of considerable complexity, requiring knowledge of behavior, development, ecology, physiology, and genetics to provide a full understanding of its evolution and function.

See Also the Following Articles

Aphids • Juvenile Hormone • Locusts • Magnetic Sense • Monarchs • Orientation

Further Reading

Brower, L. P., and Malcolm, S. B. (1991). Animal migrations: Endangered phenomena. *Am. Zool.* **31,** 232–242.

Dingle, H. (1996). "Migration: The Biology of Life on the Move." Oxford University Press, New York.

Dingle, H. (2001). The evolution of migratory syndromes in insects. *In* "Insect Movement: Mechanisms and Consequences" (I. Woiwood and D. R. Reynolds, eds.), pp. 159–181. CAB International, London.

Johnson, C. G. (1969). "Migration and Dispersal of Insects by Flight." Methuen, London.

Kennedy, J. S. (1985). Migration: Behavioral and ecological. *In* "Migration: Mechanisms and Adaptive Significance" (M. A. Rankin, ed.). *Contr. Mar. Sci.* **27** (suppl.), 5–26.

Oliveira, E. G., Srygley, R. B., and Dudley, R. (1998). Do Neotropical migrant butterflies navigate using a solar compass? *J. Exp. Biol.* **201,** 3317–3331.

Rankin, M. A. (1991). Endocrine effects on migration. *Am. Zool.* **31,** 217–230.

Rankin, M. A., and Burchsted, J. C. A. (1992). The cost of migration in insects. *Annu. Rev. Entomol.* **37,** 533–559.

Zera, A. J., and Denno, R. F. (1997). Physiology and ecology of dispersal polymorphisms in insects. *Annu. Rev. Entomol.* **42,** 207–231.

Mimicry

Mathieu Joron

Leiden University, The Netherlands

Mimicry is the adaptive resemblance in signal between several species in a locality. The most spectacular and intriguing cases are those of accurate resemblance between distantly related animals, such as spiders mimicking ants. Closely related species can also benefit from mutual resemblance, in which case mimicry results from selection against signal divergence.

The vast majority of the hundreds of thousands of insect species are described and identifiable on the basis of morphological characters. This bewildering diversity is, however, ordered because species share characters with their

relatives—and one of the taxonomist's tasks is indeed to recognize, among the shared and divergent characters, a sign of the relatedness of the taxa. Nevertheless, some distantly related species may share a common morphology. Such resemblance may be the result of evolutionary convergence, i.e., parallel lifestyles leading to the selection of similar morphological structures; in this case, resemblance per se is not under selection. On the contrary, when a character is taken as a signal between individuals, one species may benefit from bearing the same signal as one already used by another species; then selection acts directly to favor increased resemblance.

AN INTERACTION BETWEEN THREE PROTAGONISTS

The Discovery of Mimicry and the Development of Evolutionary Hypotheses

Mimicry in insects has been a puzzle for entomologists long before the Darwinian concept of natural selection, but the explanations for mimicry are tightly linked to the development of evolutionary thinking. While he was traveling in the Amazon with Alfred Russel Wallace in 1842, British entomologist Henry Walter Bates noted that distantly related butterfly species bore the same wing color pattern. Moreover, these communities of species changed their shared pattern in concert across localities. Among these species were the very abundant Ithomiinae (called "Danaoid Heliconiidae" then, now a subfamily in the Nymphalidae) and rarer Dismorphiinae (called Leptalidae then, now a subfamily in the Pieridae). Bates, as a pioneer evolutionist (but after Darwin published his *On the Origin of Species*), developed an adaptive explanation for the resemblance. Hypothesizing that ithomiines were inedible to most predators, he proposed that the edible pierids would benefit from being mistaken for their defended counterparts and would thus be selected to resemble them. Edward B. Poulton later named this kind of mimicry after him as Batesian mimicry, when an edible species mimics a distasteful one.

Bates also realized that some apparently inedible ithomiine species in the genus *Napeogenes* seemed to mimic other inedible Ithomiinae. He proposed that, in fact, rare species, whatever their palatability, should benefit from resembling defended common species. It was, however, more difficult to understand the resemblance of abundant and distasteful *Melinaea, Mechanitis* (Ithomiinae), *Lycorea* (Danainae), and some *Heliconius* (Heliconiinae) from Peru and Colombia, so he assumed the resemblance was the result of some inorganic or environmental factors. In 1879, German naturalist Fritz Müller was the first to develop a mathematical demonstration that two unpalatable prey could benefit from mutual resemblance. He understood that, if the community of predators had to kill a certain (fixed) number of prey to learn to avoid them, two indistinguishable distasteful species would together suffer this mortality and both reduce their death rate per unit time. Müller actually showed that this benefit was biased in favor of the rarer species, to a factor equal to the square of the ratio of the species' abundance. Therefore, unequal population sizes translate into even more unequal, although still mutual, benefits: Müllerian mimicry, thus defined, could be beneficial for both species, and perhaps also for the predators, in contrast to parasitic Batesian mimicry.

Mimicry: An Interaction between Senders and Receivers

Mimicry typically involves at least three protagonists, two senders and one receiver, with the receiver judging the resemblance of the signals from the two senders (Fig. 1). Obviously, both the senders and the receiver should be found in the same locality for the mimicry to be possible, although time lags or geographic separation between senders may be plausible if receivers have a long-term associative memory and/or migrate. In a habitat, many senders will converge on the same signal, thereby forming what is called a mimicry complex, or mimicry ring. Signals may involve different sensory modalities, depending on the receiver's sensory ecology: static visual signals (e.g., warning color patterns in butterflies, recognizable body shapes in ants), motion (flight behaviors), acoustic signals (hissing and clicking in many Arctiidae moths), olfactory/chemical signals (pheromones or the so-called "cuticular hydrocarbon profiles" by which social Hymenoptera recognize one another), or tactile signals (used by brood parasites of ants to be allowed to enter their nests). Signaling is indeed often multimodal.

Apparent complications may arise when, for example, one of the senders is also the receiver. For example, a predator may mimic the appearance of its prey when approaching it (aggressive mimicry in some spiders or chemical/tactile mimicry for brood parasites); the prey is thus fooled by the predator via its own conspecific signal. The two senders can also be the

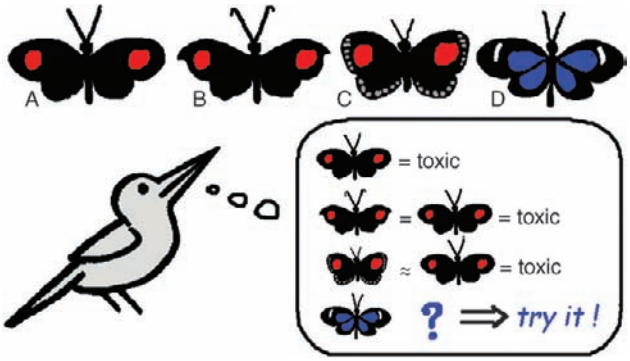

FIGURE 1 Conditioned predators and signaling prey. Predators are known to generalize their knowledge of distasteful prey to other resembling prey. Therefore, once predators recognize one prey as distasteful (prey A), other prey may gain from mimicry, whatever their palatability (prey B and C). If the prey is palatable (prey C), its mimetic gain becomes limited by its abundance in the locality. Finally, a conspicuous prey with a (nonmimetic) pattern new to the predator should suffer higher mortality, making the evolution of diversity in warning color and mimicry a puzzle.

same species. This is sometimes the case in chemical-sequestering phytophagous insects when unpalatability varies among individuals in the same population (e.g., *Danaus gillipus* in Florida), leading to so-called "automimicry" of palatable toward unpalatable individuals in the same species. Similarly, male Hymenoptera do not have the defenses that females have.

However, the present article is not organized around these numerous classificatory distinctions, which are based on subtle differences in the identities of senders and receivers or ecological situations. Instead, it highlights the important evolutionary dynamics that arise from whether receivers are expected to try to discriminate or generalize on the senders' signals or, in other words, from senders sending honest compared to dishonest signals. This should bring into perspective some of the main and still unresolved puzzles in mimicry theory, such as the rise and maintenance of diversity in mimicry signals. Most examples are chosen from the butterfly genera that represent today's best known mimetic organisms, such as *Papilio* and *Heliconius;* indeed, our knowledge of the ecology and genetics of mimicry in these genera is unequaled by any other group of insects.

FREQUENCY-DEPENDENT POPULATION PROCESSES

Batesian Mimicry and Negative Frequency Dependence

THEORY AND CONSEQUENCES In Batesian mimicry, one of the sender species, the mimic, sends a dishonest signal to deceive the receiver—e.g., a predator. It is thought that deception is possible only if the receiver has previously inherited or acquired knowledge about this signal. There is ample evidence that (1) vertebrate predators (birds, lizards) can learn to recognize distasteful prey, (2) that they can be deceived by mimicry, and (3) that mimics gain from the resemblance. The most famous Batesian mimic is probably the viceroy butterfly, *Limenitis archippus,* which mimics the monarch *Danaus plexippus,* although this relationship is now questioned (because viceroys can be unpalatable). Hoverflies (Diptera: Syrphidae), diurnal moths (Sesiidae, Sphingidae), striped beetles (Cerambycidae), or crane flies (Tipulidae) are well-known Batesian mimics of wasps and bees (Fig. 2).

Clearly, the efficiency of the deception is directly linked to the probability that predators have knowledge of the prey. It thus depends on the ratio of models and mimics in the population of prey (Fig. 1). As in host–parasite systems, the fitness of Batesian ("parasitic") mimics therefore depends negatively on their proportion in the prey community. Negative frequency dependence, the selective advantage to rare forms, is thought to be a strong force favoring and maintaining diversity in many ecological situations in nature. In Batesian mimics, any new (or rare) mutant resembling another protected model will be favored, leading to a balanced polymorphism between the two mimetic forms. Negative frequency dependence also predicts that the local number of Batesian species should be dependent on the abundance of the model(s).

FIGURE 2 Batesian mimicry. The day-flying moth *Synanthedon tipuliformis* (Sesiidae) (top) is a Batesian mimic of stinging wasps in Europe. The resemblance is very accurate, and the moth is very rare compared to its wasp models, so that it is not often observed. Similarly, but in a totally different group, the beetle *Clytus arietis* (bottom) mimics wasps and is sometimes seen on blossoms. These two examples illustrate how the same general appearance can be achieved by morphological changes of totally different nature in different groups of insects. (Photographs copyright 1998–2002 Hania Arentsen and Hans Arentsen, reproduced, with permission, from The Garden Safari, http://www.gardensafari.net.)

Many, but by no means all, Batesian mimics are indeed polymorphic. Among the most famous is the African swallowtail *Papilio dardanus,* which may have three co-occurring forms that mimic different species of the Danaine genus *Amauris. Hypolimnas misippus* (Nymphalinae) is another African butterfly that has four forms mimetic of *Danaus chrysippus.* In South America, the swallowtail *Eurytides lisithous* has up to three forms that mimic the co-occurring *Parides* species (Papilionidae), whereas in Southeast Asia the famous *Papilio memnon* also mimics three or more different papilionid models. In the Diptera, the Old World hoverflies *Volucella bombylans* and *Merodon equestris* are examples of polymorphic species mimicking bumble bees.

EVIDENCE FOR NEGATIVE FREQUENCY DEPENDENCE Although experimental demonstration that Batesian polymorphisms stem from negative frequency dependence is still

lacking, there is a lot of evidence for negative frequency dependence itself. A first line of evidence comes from the observation of patterns of abundance of models and mimics in nature. For example, the North American butterfly *Battus philenor* is known to be unpalatable to most birds and is believed to act as model for a number of edible mimics in the "black" mimicry ring. In one of them, *Papilio glaucus,* females are found as a mimetic and a nonmimetic (male-like) form, and the proportion of the mimetic form tends to be higher where its model *B. philenor* is more abundant. Similarly, the resemblance of the mimic *Papilio troilus* to *B. philenor* is higher where the latter is abundant. These give an overall pattern of mimics' occurrence consistent with negative frequency dependence. Moreover, field experiments directly showed a strong selective advantage to mimetic vs nonmimetic *Callosamia promethea* day-flying moths, another Batesian mimic of *B. philenor.*

Experimental approaches give more insight into the mechanisms involved in frequency dependence. In experiments, captive or wild predators can be tested with a variety of artificial or real prey, and the mimic/model proportions can be experimentally changed to explore how it affects the preys' survival. Traditional experiments were carried out in the 1970s with mealworms or pastry baits colored with food dyes, and/or dipped in quinine to make them distasteful, and exposed to garden birds in suburban Britain. Such experiments do suggest that a rare mimic has an advantage over a common one if the "model" is slightly distasteful, which demonstrates frequency-dependent selection. However, if the "models" were made very distasteful, the advantage of being rare decreased and eventually vanished. Laboratory experiments can also be used to search for evidence of frequency dependence, while avoiding potential confounding effects of field experiments. Experiments with captive great tits as predators showed that the mortality of both mimics and models depended on the frequency of the model and that both models and mimics survived better when mimics were fewer.

These experiments tell us that the intensity of frequency-dependent selection in mimics is highly dependent on the palatability of the models. To see its selective advantage decrease, the palatable mimic must become very common, or the model must be not very distasteful. This suggests there is some kind of effective "equivalence" between relative numbers of prey encountered and their relative levels of toxicity.

Positive Frequency Dependence in Müllerian Mimicry

THEORY: THE DISADVANTAGE OF RARE FORMS Warning signals, or aposematism, evolve because prey bearing signals that predators associate better with unprofitability (e.g., harmful prey) survive better. The evolution of warning signals brings some apparent paradoxes that are not treated in that entry. However, there is plenty of evidence that aposematic prey are easily learned and subsequently avoided by vertebrate predators. Both the warning prey and the learning predator benefit from a correct interpretation of the signal. Under such an "honest signaling" framework, rare or new variants within a prey population should not be recognized as distasteful and should suffer higher predation (Fig. 1). This selection against rare forms translates into positive frequency-dependent selection: rare mutants are removed, leading to monomorphism in all populations.

Because predators select only on prey appearance, the selective pressure does not stop at the species boundary: several protected prey species may be selected to use the same warning signal, i.e., become Müllerian mimics. Although the phenomenon is not necessarily symmetrical, two or several defended species should all benefit from sharing a warning signal, which reduces their per capita predation rate. As more and more individuals join in the mimicry ring, the protection given by the signal becomes stronger. Therefore, the direct, and naïve, prediction is that all unpalatable prey of a similar size in a habitat should converge into a mimicry ring.

EVIDENCE FOR THE FREQUENCY-DEPENDENT BENEFITS OF MÜLLERIAN MIMICRY Although comparative and/or biogeographical studies give strong support to the theory, the first convincing experimental evidence came from pastry-bait experiments with garden birds that tend to attack rare distasteful baits more often than common ones. Recently, laboratory experiments also showed strong selection against new rare warningly colored prey items. However, field evidence with free-living prey is crucial for a validation of these results. In one experiment, J. Mallet reciprocally transplanted *Heliconius erato* individuals between populations in which *H. erato* have different wing patterns, thus effectively releasing rare "mutant" and "control" butterflies into the host populations. A strong selective advantage of about 50% was calculated for the commoner form. More recently, to avoid the potential pathology of color patterns being adaptations to local habitat conditions in addition to mimicry, D. D. Kapan used a similar reciprocal release–recapture technique but used polymorphic populations of the butterfly *H. cydno.* In this species, two morphs coexist but participate in two different mimicry rings that differ in relative abundance in different locations in Ecuador. Life expectancy was 12 days for the locally common forms and only 2 days for the locally uncommon forms. These field data give unequivocal evidence for strong selection against rare forms in these Müllerian species.

CONSEQUENCES AND CHALLENGES Strong purifying selection now seems well supported by theoretical, comparative, and experimental evidence. To evolve a new pattern, a toxic prey would have to pass an apparently impassable initial disadvantage, survive a transient polymorphism, and win the aposematic competition with alternative warning signals. It is therefore no surprise that most distasteful Müllerian mimics are indeed monomorphic in local populations (Fig. 3) and

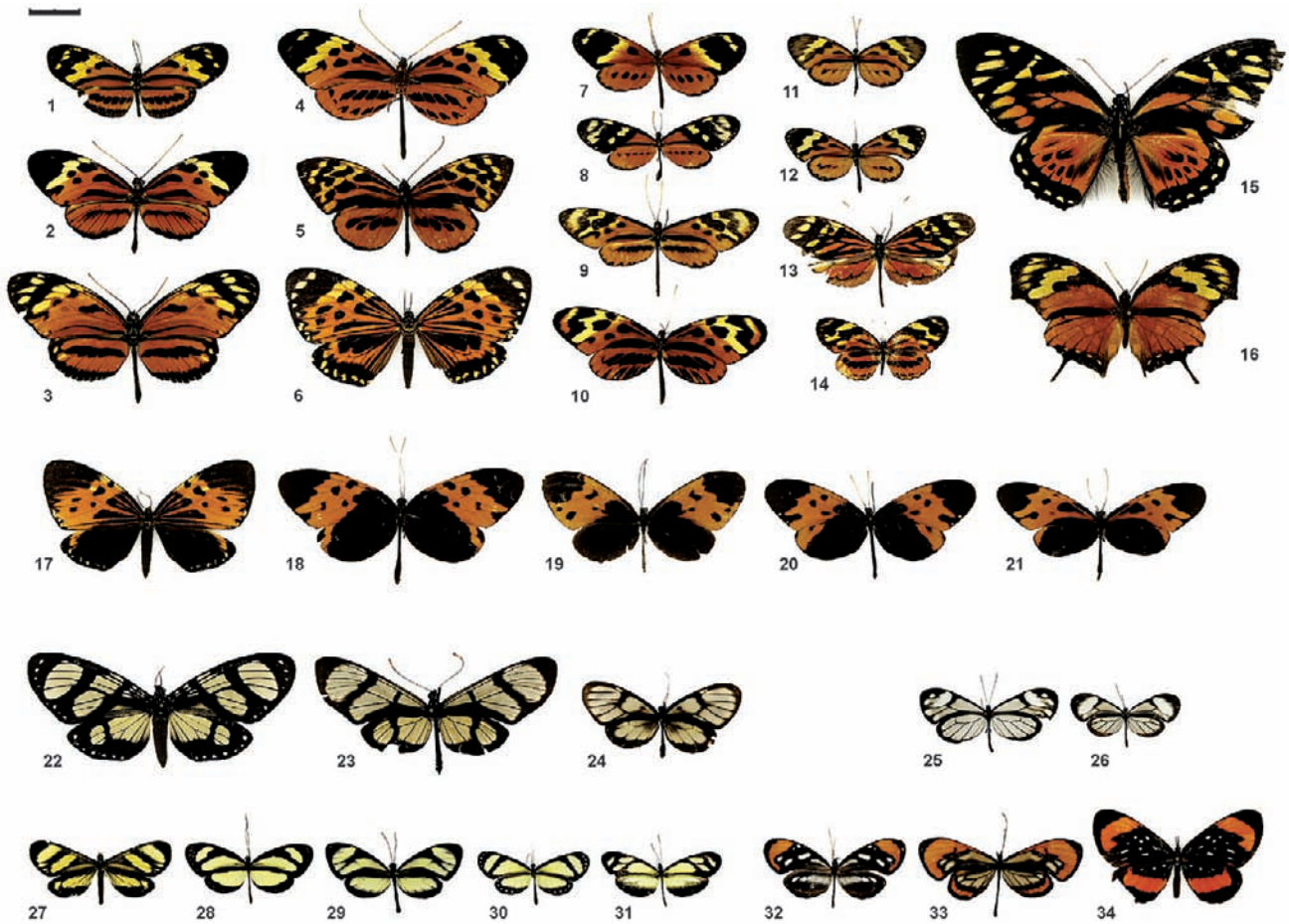

FIGURE 3 Six butterfly mimicry rings from eastern Peru. The mimicry rings (groups of mimetic species) presented here are dominated by butterflies in the Ithomiinae and occur in the forests around the city of Tarapoto. Following G. W. Beccaloni's nomenclature, these mimicry rings are Tiger (1–16), Melanic tiger (17–21), Large transparent (22–24), Small transparent (25 and 26), Small yellow (27–31), and Orange-tip (32–34) mimicry rings. At least 5 other mimicry rings can be recognized involving Heliconiinae and/or Ithomiinae in this area, which brings the total to at least 11 mimicry rings for these two butterfly subfamilies. Many more species, not featured here, belong to these mimicry rings, particularly Ithomiines and especially in the Small transparent group. The Tiger mimicry ring involves a lot of species and the size distribution is almost continuous from small to very big. This may be because as more and more Müllerian mimics join the mimicry ring, predators might generalize more, and the selection for close resemblance could be somewhat relaxed. Note that some day-flying moths (6, 17, 22, 27) participate in these mimicry rings, probably as Müllerian mimics (they reflex-bleed bitter hemolymph when handled). Butterflies 13–16 and 31 are supposed to be Batesian mimics since they belong to palatable groups within their families. See more species belonging to these mimicry rings in Figs. 4 and 5. All butterflies are Nymphalidae: Ithomiinae, except 1–3 (Nymphalidae: Heliconiinae), 14 (Nymphalidae: Melitaeinae), 16 (Nymphalidae: Charaxinae), 15 (Papilionidae), 13 and 31 (Pieridae: Dismorphiinae), 34 (Riodinidae), and 6, 17, and 22 (Arctiidae: Pericopinae). Scientific names: 1, *Eueides isabella;* 2, *Heliconius pardalinus;* 3, *H. hecale;* 4, *Melinaea menophilus;* 5, *Tithorea harmonia;* 6, *Chetone histriona* sp.; 7, *Napeogenes larina;* 8, *Mechanitis lysimnia;* 9, *Mec. polymnia;* 10, *Mec. mazaeus plagifera* ssp.; 11, *Ceratinia tutia;* 12, *Hypothyris cantobrica;* 13, *Dismorphia amphiona;* 14, *Eresia* sp.; 15, *Pterouros zagreus;* 16, *Consul fabius;* 17, *Chetone histriona;* 18, *Mel. marsaeus;* 19, *Hyposcada anchiala;* 20, *Hypot. mansuetus;* 21, *Mec. mazaeus deceptus;* 22, *Notophyson heliconides;* 23, *Methona confusa;* 24, *Godyris zavaleta;* 25, *Greta andromica;* 26, *Pseudoscada florula;* 27, *Notodontid* moth; 28, *Aeria eurimedia;* 29, *Ithomia salapia;* 30, *Scada* sp.; 31, *Moschoneura* sp.; 32, *Hypos. illinissa;* 33, *Hypoleria sarepta;* 34, *Stalachtis euterpe.* Scale bar, 2 cm.

that polymorphisms are usually restricted to narrow hybrid zones between color-pattern races. In *H. erato,* in which two color races abut, frequency-dependent selection maintains a sharp boundary, alternative forms being positively reinforced on either side of a steep cline. Many species join Müllerian mimicry rings, which itself represents interspecific evidence for strong frequency-dependent selection.

However, in contrast with such extremely conservative forces, diversity is present at all levels in mimicry (Fig. 3). At a *macroevolutionary level,* aposematic and mimetic groups typically undergo rapid mimetic radiations into numerous species and races differing in color pattern, like heliconiine butterflies or pyrrhocorid red bugs. At the *community level,* many radically different mimicry rings coexist in the same habitat (e.g., five or six coexisting rings just within the *Heliconius* of Costa Rica, at least seven or eight rings just within the Ithomiinae of the Peruvian Amazon—Fig. 3.). At the *biogeographical level,* many aposematic species show a bewildering diversification in more or less sharply defined mimetic races. Finally, at the *population level,* several chemically defended species show mimetic polymorphism. For instance, the bumble bee *Bombus rufocintus* has two

FIGURE 4 Polymorphic Müllerian mimicry. The Amazonian butterfly *H. numata* (Nymphalidae: Heliconiinae—right column) is a Müllerian mimic in a variety of tiger-pattern mimicry rings. Each population (here around the city of Tarapoto in Eastern Peru) is polymorphic and up to seven forms may coexist, each being an exceptionally accurate mimic of species in the genus *Melinaea* (Nymphalidae: Ithomiinae—left column). Spatial variation in selection pressure is probably what maintains the polymorphism, by a balance between local selection for mimicry of the commonest *Melinaea* species and movement of individuals (gene flow) between neighboring localities selected for different wing patterns. From top to bottom (left column): *Melinaea ludovica ludovica, Mel. satevis cydon, Mel. marsaeus mothone, Mel. marsaeus phasiana, Mel. menophilus* ssp. nov., *Mel. menophilus hicetas,* and *Mel. marsaeus mothone.* (Right column) *H. numata* forms *silvana, elegans, aurora, arcuella, tarapotensis, timaeus,* and *bicoloratus.* Scale bar, 2 cm.

mimetic forms in North America, the burnet moth *Zygaena ephialtes* has two sympatric forms in Italy, the African monarch *D. chrysippus* has four main color forms coexisting in a large areas in East Africa, and the Amazonian *Heliconius numata* shows the most astounding polymorphic mimicry with up to 7 to 10 forms in the Andean foothills. In each of these cases, the different forms closely match the different local mimicry rings (Fig. 4).

This rampant diversity does not question the existence of frequency dependence itself, but the details of how purifying selection may or may not prevent the evolution of diversity. It may also question the validity of the two classical categories of protective mimicry (Batesian and Müllerian) and the existence of a sharp divide between them along the spectrum of prey palatability. Explaining these unexpected cases is therefore central to our understanding of signal evolution in distasteful insects.

The Palatability Spectrum and Predator Psychology

MODELS OF MIMICRY EVOLUTION Case studies and experiments on mimicry are practically difficult, are time consuming, and inform us only on potential processes in particular cases. They are thus not always very informative as to which processes are generally important in the evolution of mimetic diversity. For these reasons, mathematical models simulating mimicry evolution have been widely used. Models of mimicry evolution have been traditionally of two different types: "evolutionary dynamics" models have concentrated on trait evolution in the prey populations, underestimating the effects of the details of predator behavior; "receiver psychology" models have concentrated on the effect of predator cognitive abilities in driving the costs and benefits to mimetic prey, but largely ignored evolutionary processes in the prey populations, particularly frequency or density dependence. The second category of models are those that "traditionally" pose a threat to the validity of the Batesian/Müllerian distinction, and M. P. Speed even coined the new term "quasi-Batesian mimicry" for the strange, though purportedly common, intermediate dynamics that his model highlighted.

The main discrepancies lie in the way predators are thought to respond to prey palatability and density. Speed's models assumed that predators attack a fixed fraction of a prey in a population, irrespective of their total number (linear frequency dependence), and that this fraction depends on the palatability of the species. In a mixture of prey of differing palatability, the resulting fraction killed would be intermediate between the fractions lost in each prey in the absence of mimicry, leading to one prey species benefiting and the other suffering from mimicry. This view, however, leads to the strange prediction that as more mildly unpalatable prey are present, the attacked fraction (per unit time) can increase. In contrast, J. Mallet and the author argued that predators are unlikely to be sensitive to frequency

per se and should instead need only to attack a fixed number of prey before learning, making the "attacked fraction" a decreasing function of the total number of prey bearing the pattern. This should lead to a strongly nonlinear, effectively hyperbolic frequency dependence. The attacked fraction (per unit time) should always decrease when the total number of unpalatable prey increases, whatever their relative unpalatability.

The debate is still very much active, and decisive data are surprisingly scarce. In an experiment with pastry baits and wild passerines, Speed showed that the attack fraction of a mimetic pair was indeed intermediate between that of either "species" alone. Furthermore, birds seemed to learn only to a certain extent; that is, they never completely stopped attacking the unpalatable items. Despite some problems in the experimental design (no predator monitoring during the study, artificially high prey density, prey predictability, zero cost of experimenting for birds), these data remain a puzzle and may hint at more complex learning processes than a pure number-dependent dose response. More decisive evidence came from L. Lindström's study, in which novel toxic prey were introduced into a great tit's foraging arena at varying frequencies (=densities in this setting). Although the total number of attacked toxic prey increased with their initial frequency, the attack fraction decreased. Her data support the validity of nonlinear frequency dependence, although the idea of a strictly fixed number of prey killed could be an oversimplification. Absolute numbers of prey attacked may increase with warning signal density, but proportion will inevitably decrease, which should lead to a traditional Batesian–Müllerian distinction.

THE STRENGTH OF THE SELECTION Müller's number-dependent model also leads to a prediction that has hitherto been largely overlooked. At low densities, selection should act strongly against any transient polymorphism, but at higher densities, selection quickly becomes weak at intermediate form frequencies. This leads to effective neutrality of polymorphism once it is established in abundant species. Kapan's field experiments, in which *H. cydno* were released at varying density, showed precisely this trend. Polymorphism could therefore be nonadaptive but very weakly selected against by predators.

Numerical Mimicry and Density-Dependent Processes

The studies of J. Allen and his collaborators, and others, show that prey selection by predators can be frequency dependent in palatable, cryptic prey, i.e., even in the absence of mimicry of unprofitable prey. This is probably caused, in part, from predators using search images when foraging. For instance, at low densities of a particular kind of (palatable) prey, predators usually prey on the more common form, which corresponds to their search image, imposing a *negative* frequency dependence. Cryptic prey may be globally numerous in a habitat, but because they are camouflaged, their apparent density to predators is bound to be low. This leads to the diversification of cryptic patterns, and perhaps the selection of plastic (partly environmentally induced) color-pattern genetic control, in prey. In contrast, at high density, predators usually prey on the odd phenotypes preferentially, even among perfectly palatable prey, effectively leading to a *positive* frequency-dependent selection on morphology.

Gregarious palatable prey that are at locally high density and that presumably rely on predator satiation to escape predation, might then be selected for mutual resemblance. Such a prey might be called "warningly colored," whereas the appearance itself is not protective. This idea led to the supposition that several prey species that co-occur at unusually high densities, like mud-puddling butterflies or schooling fishes, might evolve "numerical" or "arithmetic" mimicry by simple frequency-dependent predation unrelated to unprofitability. Prey traits like color, shape, and especially locomotor behavior are therefore thought to be under purifying selection in mixed-species aggregations. This attractive idea remains largely untested in insects, although R. B. Srygley proposed the pair of bright orange butterflies *Dryas julia* (Heliconiinae) and *Marpesia petreus* (Nymphalinae) as a potential candidate.

Female-Limited Mimicry

Some of the most spectacular and best studied cases of Batesian polymorphism are found in swallowtails, and in some species only the female is mimetic (see an example in Fig. 5). This peculiar tendency to sex-specific polymorphism seems to be restricted to butterflies (Papilionidae and Pieridae), and virtually no other case of sex-limited mimicry seems to be reported for other insects (except for male-limited mimicry in some moths). Female-limited mimicry was often viewed as a result of negative frequency dependence: if mimicry is restricted to one sex, the effective mimetic population size is only about half that of a nondimorphic species, reducing deleterious effects of parasitism onto the warning signal. But this group-selection argument cannot in itself explain why females tend to become mimetic more often than males and why mechanisms arise that restrict the mimicry to one sex. However, more proximal, individual-selection arguments are not lacking. First, mimicry may be more beneficial to one sex than to the other. For instance, female butterflies have a less agile flight because of egg load and a more "predictable" flight when searching oviposition sites, and they suffer higher rates of attacks by visual predators. Second, male wing patterns can be constrained by sexual selection, via either female choice or male–male interactions: males could not evolve Batesian mimicry without losing mating opportunities. In experiments with North American swallowtails (of which only females mimic *B. philenor*), male *P. glaucus* painted with the mimetic pattern

FIGURE 5 Female-limited mimicry in *Perrhybris pyrrha* (Pieridae), Eastern Peru. The female (top) is a Batesian mimic of the tiger-patterned Ithomiines and Helicomiines (see Fig.3), while the male (bottom) has retained a typical pierid white coloration. Scale bar, 2 cm.

had a lower mating success than normal yellow males; similarly, painted *P. polyxenes* males had a lower success in male–male fights and therefore held lower quality territories around hilltops. In these insects, the wing coloration appears to bear signals directed either to conspecific males or to predators, which creates a potential conflict leading to sex-limited polymorphism. It is interesting to note that *Papilio* and *Eurytides* species that mimic *Parides* (Papilionidae) in South America do not exhibit female-limited mimicry; different modes of sexual selection (e.g., absence of territoriality) may operate in the forest understory habitat. In a different ecological setting, diurnal males of the North American silkmoth *Callosamia promethea* are exposed to visual predators, and mimicry of *B. philenor* is limited to males; female *Callosamia* fly at night and benefit more by crypsis during the day.

MIMICRY AND THE EVOLUTION OF SIGNAL FORM

Resemblance and Homology

Mimicry can arise as soon as the signal is effectively copied, i.e., as soon as superficial resemblance is attained. Therefore, mimics usually bear characters similar to those of their models, but these are often clearly nonhomologous in terms of genes and mechanisms of development. For instance, red spots near the base of the wing in *P. memnon* mimic the spots on the bodies of their models. The translucency and iridescence of distasteful Ithomiinae clearwing butterflies is mimicked by white raylets in dioptine and pyralid day-flying moths and provide the same impression in motion. Similarly, the black-wing patterning of some flies seems to mimic the superposition of wings over the abdomen in their wasp models. Therefore, mimics from distant phylogenetic groups are certainly under very different functional and developmental constraints to create a mimetic impression. Selection will retain the first characters that suddenly increase overall similarity. Which initial step is made will therefore strongly influence which route is selected to achieve mimicry.

The Genetics of Mimicry: Polymorphisms and Supergene Evolution

THE DEBATE The genetical study of the evolution of mimicry was first dominated by a debate between gradualists (Fisher) and mutationists (Goldschmidt). Goldschmidt proposed that "systemic" mutations could affect the whole wing pattern of butterflies in one step and that models and mimics, although not using the same genes, were using at least the same developmental pathways. Because this view could not account for the obvious nonhomologies, like those pointed out above, Fisher and others claimed that mimicry was achieved by slow microevolutionary steps and the gradual accumulation of resemblance alleles.

Decisive steps toward a resolution of the debate came principally from the study, by C. Clarke and P. Sheppard in the 1960s, of Batesian butterfly mimics in which color pattern is easy to define and analyze and gene effects are straightforward to identify. Polymorphic mimics, particularly *Papilio* species, of which different forms could be crossed by breeding experiments (including hand pairing), were particularly useful. It appeared that color pattern is mainly inherited at one or few major loci, affecting the whole pattern. From rare recombinants, it could be shown that these loci were in fact supergenes, that is, arrays of tightly linked small-effect genes. Several additional unlinked "modifier" loci were also shown to increase resemblance via interaction and epistasis with the supergene. Goldschmidt's ideas seemed refuted.

However, although supergenes seem to be a necessary condition for the evolution of polymorphism (otherwise numerous nonmimetic, unfit recombinants would be produced), how they evolve is another issue. Theoretical models suggested that supergenes could not be achieved by simple gradual reduction in recombination. In the absence of spatial variation in selection pressures, tighter linkage cannot evolve by small steps via Fisherian gradual evolution, because good combinations of alleles are immediately broken up by recombination. Instead, gene clusters should preexist the evolution of polymorphism.

THE TWO-STEP HYPOTHESIS These results led to a unifying, now widely accepted two-step mechanism of mimicry evolution: (1) mutations at genes of major effect first

allow a phenotypic leap achieving an approximate resemblance to a particular model. Once these mutations have increased in the population, (2) resemblance can be enhanced through the gradual selection of epistatic modifiers. This two-step mechanism is supported by three lines of evidence. First, empirical evidence from butterflies suggests the existence of a small number of major-effect genes and numerous small-effect modifiers. In fact some of these genes of major effect could even include a series of regulatory upstream elements and transcription factors, now known to be involved in the development of butterfly color patterns. Pigment pathway genes and scale maturation regulators can also have very dramatic effects on the color patterns. Second, population genetics and dynamics models support the prediction that a major phenotypic jump is necessary to cross the deep fitness valleys in a rugged fitness landscape, after which gradual, Fisherian evolution may proceed to enhance resemblance. Finally, experiments show that birds associate cryptic patterns with edibility and generalize those in such a way that only profoundly deviant prey are treated as separate cases by the birds and memorized as warning patterns when appropriate. These experiments also indicate that increased resemblance is still significantly advantageous in imperfect mimics, supporting the second step of the two-step scenario.

LARGESSE OF THE GENOME Another, but not exclusive, route to supergenes for mimicry is called largesse of the genome, put forward by J. R. G. Turner. Under this scenario, it is believed that the modification of a trait can be achieved by so many different genes that some of them will inevitably happen to be linked. Among the many possible combinations of loci, selection could simply sieve out the ones that involve linked genes. This hypothesis is particularly likely for loss-of-function phenotypes that can be achieved by mutating any step in the development, like the loss of tail in the African swallowtail *P. dardanus.* Similarly, that different mimetic species use nonhomologous supergenes can be viewed as indirect evidence for the validity of the largesse of the genome hypothesis in the broad sense.

SUPERGENES IN MÜLLERIAN MIMICS: A PUZZLE Müllerian mimics being usually monomorphic locally, supergenes are not expected to control wing patterns, and multilocus control was hypothesized to be the norm. This basic prediction has, however, constantly been challenged by *Heliconius* color-pattern genetics, which show that a limited set of genes of large effect and supergenes control most of the racial color-pattern variation. In the polymorphic *H. numata,* one single gene seems to control the entire wing pattern, with as many as seven alleles, each allele bringing resemblance to a specific mimetic pattern (Fig. 4). Tight gene clusters are also found, to a lesser extent, in polymorphic *H. cydno,* in *H. melpomene,* and in *H. erato.* The existence of these supergenes seems puzzling. It is possible that butterfly color patterns in general are under the control of relatively few conserved genes, at least in some lineages, such as developmental regulatory genes involved in eyespot formation.

In toxic prey, strong selection against any new form and the impossibility of gradual color-pattern changes have been theoretically and empirically demonstrated. It follows that, like Batesian mimics, Müllerian mimics seem to need an initial phenotypic leap, perhaps involving multimodal signal modifications, to jump either to an already protected pattern or away from predators' generalization of cryptic prey. Therefore, it is perhaps no surprise that most exaggerated signal forms studied are under the control of relatively few genes, following the same two-step scenario as in Batesian mimicry. Moreover, switches from one mimetic pattern to another are likely selected only if the new mutant's mimetic characters are not randomly recombined in its descendants. This imposes another constraint (or "sieve") on the genetic architecture for new mimetic patterns to be selected out of a transiently polymorphic population. It is therefore remarkable to note that although Batesian and Müllerian evolutionary dynamics are radically different, and are even perhaps engaged in an evolutionary arms race, the evolution of their signals might require a similar (though nonhomologous) genetic predisposition.

Myrmecomorphy

Ants represent the most abundant group of organisms in most biota and have powerful multimodal defenses such as acid taste, aggressive biting, painful sting, and social defense. For these reasons, foraging ants are generally little subject to predation and act as ideal models in mimicry rings. Many insects and spiders indeed have an altered morphology and resemble ants, a phenomenon called myrmecomorphy. For instance, several salticid spider genera such *Myrmarachne* or *Synmosyna* are bewilderingly good ant mimics. It is also common to spot ant-like myrid nymphs (Heteroptera) running among leafcutting *Atta* ants or *Ecitomorpha* staphilinid beetles among *Eciton* army ant columns. The adaptive significance of ant-like morphology has been the subject of considerable debate. For instance, several ant-like spiders are believed to mimic ants as a trick to approach and prey on their ant models ("aggressive mimicry"); some ant-like bugs use the same trick to approach and prey on ant-tended aphids. However, most ant-like insects are phytophagous, do not prey on foraging ants, and usually mimic the locally abundant ant species. They are therefore good Batesian mimicry candidates. The interesting aspect of ant mimicry is that, although small birds, lizards, or amphibians may be important predators on ant-sized insects, there are grounds to think that arthropod predators with developed visual skills could be the prime receivers selecting for ant mimicry. For instance, wasps in the Pompilidae are known as important predators of jumping spiders, but ignore ants, thus potentially selecting for ant-like morphology and behavior. Jumping spiders themselves are visual predators hunting insects and also tend to avoid stinging ants as prey. Although

the cognitive abilities of arthropods are not well researched, several studies using mantids, assassin bugs *Sinea* sp., or crab spiders show that they are capable of associative learning and discriminate against ant-like prey. Despite the difference in visual acuity and cognitive abilities between vertebrates and arthropods, it is interesting to note that arthropod predators are likely responsible for visual mimicry that is very accurate to our eyes.

The Importance of Behavior and Motion

Myrmecomorphy highlights a crucial aspect of mimicry: the importance of behavior. Predators integrate many aspects of prey appearance when making a decision of whether to attack, and behavior is an important part of multimodal signals. Ants are characterized by jerky (e.g., *Pseudomyrmex* spp.) or zigzag (e.g., *Crematogaster* spp.) movements that their mimics adopt. Constant waving of antennae seems to be a common feature of ants, which mimics, such as ant-mimicking spiders (Salticidae) or spider-wasp-mimicking leaf-footed bugs (Coreidae), copy by waving their front legs. Because motion considerably enhances visibility, it is hardly surprising that details of the behavior make important identification cues for the predators. For instance, although slow flight in aposematic butterflies may save energy, slowness itself is certainly recognized as such by predators that can select on extremely minute details of flight unnoticeable to the human eye. R. B. Srygley's work on locomotor mimicry has shown that the two butterflies *H. erato* and *H. sapho* differ in the asymmetry of the upward and downward wing strokes, which their respective (Müllerian) mimics *H. melpomene* and *H. cydno* copy accurately in Panama. Batesian mimics usually retain escape behaviors characteristic of their groups: the lazily flying Neotropical butterfly *Consul fabius* (Nymphalidae: Charaxinae) (see Fig. 3) can start rapid escape flight when detected; ant-mimicking salticid spiders are also usually reluctant to jump unless attacked.

The tendency for predators to generalize the characteristics of palatable prey, on which they actually feed, probably selects aposematic signals away from these morphologies, and behavioral signals are no exception. Rapid jerky flight is usually characteristic of a tasty prey, a profit that predators have to weigh against the time and energy costs associated with catching the prey. Unconventional behaviors like the flight of *Heliconius* butterflies or the looping of honey bees make them highly noticeable to predators. This imposes an additional visibility cost on incipient mimetic prey; for the resemblance to be selected, such cost has to be offset by a significant reduction in predation. These considerations suggest that mimetic behavioral change probably evolves in much the same way as morphological characters do, i.e., a two-step process.

Escape Mimicry

Unpalatability is not the only way to be unprofitable to predators. Fast, efficient escape is another way for preys to teach predators that pursuit is useless and will bring no reward: predators unable to consume the desired prey may associate this frustration with the prey appearance and reduce their attacks on this prey altogether. Even if the prey can be seized, predators probably trade off the energy spent and the (often low) nutritional reward. In several experiments birds were shown to be able to decrease their attack rates when the presented prey would quickly disappear ("escape") during their attacks, and conspicuousness of the prey tended to enhance the response. Therefore, evasive prey could advertise their escaping abilities by color patterns, which other prey may mimic. At least three kinds of characters may enhance the difficulty of catching an evasive prey: erratic flight (like that of Pierids), fast and maneuverable flight (like that of charaxine butterflies), or high reactivity (like that of syrphid hoverflies). Typically, these escape specialists are all palatable to predators. Some species of the Neotropical butterfly genera *Adelpha* (subfamily Nymphalinae) and *Doxocopa* (subfamily Charaxinae) show convergent appearance and exhibit extremely quick escape when slightly disturbed, followed by very fast flight. Their resemblance is hypothesized by R. B. Srygley to be a case of escape mimicry. The poor resemblance of some hoverflies to their purported hymenopteran models has also led to the hypothesis that groups of syrphid species could represent an escape mimicry ring on their own.

Poor Mimicry

At least to our eyes, the model's color pattern is not always copied very accurately. Many syrphid flies, for instance, are difficult to assign to particular mimicry rings, although they seem to mimic the general appearance of Hymenoptera. The heterogeneity in mimetic accuracy has led biologists to propose adaptive and nonadaptive hypotheses, none of which seems very strongly supported at present. (1) The null hypothesis is that poor mimics are no mimics: many mimicry associations have been claimed on the general appearance of an insect, whereas careful examination of the geographic covariation of purported models and mimics may reveal evidence against them. In the case of inaccurate mimics, this method is not very powerful because the mimetic association itself is hard to define, so such covariation is difficult if not impossible to judge. (2) Another nonadaptive scenario is that accurate mimicry may not always be possible, either because of functional constraints/trade-offs on the modified organs or because of genetic or developmental constraints on the variation available in populations. Mimicry may then asymptotically reach a maximum level of resemblance, contingent on the route followed in the initial stages of the mimetic change. Again, this is theoretically plausible, but difficult to test. (3) Among the adaptive explanations for inaccurate mimicry is the hypothesis that these species are in the initial stages of their mimetic change and that our instantaneous view of evolution doesn't show us the complete picture. (4) Another adaptive scenario is that predators have biases and perceptions

different from those of humans and are likely to generalize more in some directions than in others, leading to the possibility that mimics that look very inaccurate to us are in fact very good mimics for a predator. Generalization is also dependent on the strength of the harmfulness of the models, perhaps allowing lower levels of accuracy. This may be the case for poor mimicry in some hoverflies. The ultimate adaptationist hypothesis is that inaccuracy itself may be beneficial. It could either (5) allow the mimic to benefit from the protection of several different models, perhaps in a heterogeneous environmental context, or (6)—a related hypothesis—create conflict in the predators' recognition, which may give the mimic more time and chances to escape.

MIMICRY, COMMUNITY ECOLOGY, AND MACROEVOLUTIONARY PATTERNS

Habitat Heterogeneity, Spatial Dynamics, and the Coexistence of Mimicry Rings

The efficiency of a warning pattern depends on the abundance of that pattern in the habitat. Therefore, as new species join a particular mimicry ring, the protection given by the pattern increases, and more species should converge on this best protected pattern. Ultimately, all species should converge on a single mimicry ring. But nature seems to behave in a totally different way. In any one habitat, particularly in tropical environments, aposematic insects of similar size and shape usually cluster into a number of distinct mimicry complexes or mimicry rings.

MULTIPLE MIMICRY RINGS IN THE COMMUNITY One possibility is that different mimicry rings are found in different microhabitats. If predators do not move between microhabitats, or retain microhabitat-specific information, insect species in different microhabitats could converge on different adaptive peaks. Flight height has been invoked as a possible explanation, following the rainforest stratification paradigm, but evidence from butterflies is rather equivocal. However, host-plant stratification and different nocturnal roosting heights in Neotropical butterflies have received empirical support. Forest maturity and succession stage influence the host-plant composition and may allow the maintenance of multiple mimicry rings in a mosaic habitat.

MULTIPLE MIMICRY RINGS WITHIN A SPECIES If some species are patchily distributed because of their microhabitat requirements, each "subpopulation" may be particularly sensitive to genetic drift and allow the local predators to learn and select a different color pattern in different patches. Once locally stabilized, the new pattern may be hard to remove. Indeed, local positive frequency dependence is both very efficient at stabilizing patterns around fitness peaks and slow at removing already established suboptimal patterns. Any slight difference in microhabitat quality or patchiness of the species involved will increase the local apparent abundance of particular patterns to particular predators, further decreasing the power of selection to achieve ultimate convergence.

This "mosaic mimetic environment" theory can help explain some problematic cases of Müllerian polymorphism. For instance, *Laparus doris* is a Heliconiine butterfly (Nymphalidae) that has up to four coexisting forms in some populations, some of which are probably mimetic and others are not. The maintenance of polymorphism in this species could be attributed to its high larval and pupal gregariousness (several hundreds of individuals), which results in a patchy distribution of the adults. When hundreds of butterflies suddenly emerge from one single vine, they make up their own local mimetic environment, and the mimetic environment prior to the mass emergence might be effectively neutral to *L. doris.*

If the species composition and the resulting mimetic environment are spatially variable, polymorphism can evolve in microhabitat generalists, with gene flow across these microhabitats. For example, the Amazonian polymorphic species *H. numata* is selected toward different mimetic patterns in different localities that may represent different microhabitats for their more specialized models in the genus *Melinaea* (subfamily Ithomiinae) (Fig. 4). The balance between local selection and gene flow in a mosaic habitat (and perhaps weak selection against polymorphism as suggested earlier) can therefore maintain a nonadaptive, although widespread, polymorphism in *H. numata*.

Coevolution in Mimicry

EVOLUTIONARY RATES AND THE COEVOLUTIONARY CHASE Despite many potential sieves constraining mimicry, several to many edible species can end up mimicking a particular warning pattern in a parasitic way. In such cases, is it possible that a "Batesian-overload" threshold is reached, beyond which the efficiency of the signal is severely lowered? Batesian mimics are indeed parasites of the honest signals of their models, and so the models should escape their mimics by evolving a new warning pattern. However, this escape would be transient because the new pattern would soon attract new Batesian mimics, resulting in an evolutionary arms race, or coevolutionary chase, between the model and its mimics. Some authors suggested that this chase could be a cause of the mimetic diversity in both models and mimics and that cyclical interactions could arise in some cases. However, first, theory has shown that mimics always evolve faster than their models, because they gain a lot more from mimicry than models lose from being mimicked. Any gradual move of the model should be quickly matched by a similar evolution in the mimic. Second, the models, which are the prime educators of local predators, are under strong purifying selection against any new warning pattern. This strong intraspecific conservative force should in the vast majority of cases be stronger than the deleterious effects of being mimicked and preclude pattern

change in the models. Coevolutionary changes between Batesian mimics and their models should therefore be stopped in their early stages by a stronger selection for the status quo, and both the models and their mimics should be trapped in the same warning pattern. Only by a phenotypic leap toward an already established warning pattern (Müllerian mimicry) or by crossing a fitness valley thanks to local genetic drift could the model ever escape its mimics.

MUTUALISM AND COEVOLUTION IN MÜLLERIAN MIMICRY In contrast with the *unilateral* Batesian evolution in which mimics outrun their models, Müllerian mimicry was traditionally thought to involve *mutual* resemblance of the species involved, as if all had moved toward some halfway phenotype. Of course, Müller himself and others were quick to point out that the mutual benefits were not even, but lopsided, i.e., typically the rarer or the less distasteful species would benefit more than the more common or better defended one (respectively). However mutualistic the relation is, coevolution has often been assumed in Müllerian associations, and the protagonists are usually called "comimics" just because it is difficult to know if one species is driving the association. Coevolution also predicts that geographic divergence and pattern changes should be parallel in both species of comimics, like in the mimetic pair *H. erato* and *H. melpomene* in tropical America, presumably leading to parallel phylogenies. However, DNA sequences from mitochondrial and nuclear genes show distinct phylogenetic topologies in these two species and distinctly nonparallel evolution.

In fact, there are a number of grounds on which to believe that the asymmetrical relationship leads to one-sided signal evolution even in Müllerian mimicry, one species being a mimic and the other a model. First, because of number dependence, mimetic change of a rarer species toward a commoner species will be retained, but the reverse is not true: by mimicry of a less common species, the commoner species would lose the protection of its own ancestral pattern, and a change toward a rarer pattern would be initially disadvantageous. The commoner species is therefore effectively locked in its pattern, and initial changes are only likely in the rarer species. Second, given the selection against nonmimetic intermediates, the mutants in the rarer species will have to be roughly mimetic of their new model to be selected, thus bringing the ultimate shared signal closer to that of the common species. Once this initial step is made by the mimic, there could be gradual "coevolution" to refine the resemblance, but the resulting change in color pattern will inevitably be more pronounced in the mimic, the model remaining more or less unchanged. Because Müllerian pairs are of a mimic–model nature, even with mutual benefits, the prediction for parallel evolution is therefore not likely to be valid. Indeed, in the mimetic pair *H. erato/H. melpomene,* the phylogeography suggests that *H. melpomene* has radiated onto preexisting *H. erato* color-pattern races, thus colonizing all color-pattern niches protected by *H. erato* in South America.

Mimicry, Speciation, and Radiations

Racial boundaries in mimetic butterflies are usually very permeable to genetic exchange, since selection acts primarily on color-pattern genes. However, because clines moving geographically are likely stopped at ecological boundaries, the resulting racial boundaries are likely to rest on ecological gradients. Racial boundaries between mimetic color patterns could therefore be reinforced by adaptation to local ecology on either side of the cline, leading to speciation. Color-pattern diversification could then accelerate speciation by allowing both postmating reproductive isolation, because of a higher mortality of nonmimetic hybrid offspring, and premating isolation if color pattern itself is used as a mating cue by the insect. For these reasons, mimicry has the potential to accelerate speciation. The pattern of mimetic associations in *Heliconius* butterflies seems indeed to indicate that speciation and mimetic switches are usually coincident: sister species usually differ in their mimetic color pattern. Direct evidence of the role of color pattern in mate choice has been gathered for the sister species pairs *H. erato/H. himera* and *H. melpomene/H. cydno.* The first two species are geographically separated across an ecological gradient in the Andes. The second pair is sympatric, although the species also differ in ecological requirements in a patchy distribution. In both pairs, therefore, color-pattern and mimetic switches probably accelerated speciation initiated by ecological adaptation. It is unknown how general this mimicry-based speciation is in mimetic insects but it could be an important consequence of the rampant and apparently easy diversification of mimetic patterns at the intraspecific level. The genetic predisposition of mimetic species to evolve polymorphism—the first stage toward speciation—might explain why mimetic lineages are usually very speciose and undergo rapid radiations, both geographically and phylogenetically.

See Also the Following Articles

Aposematic Coloration • Coevolution • Defensive Behavior • Industrial Melanism • Monarchs • Predation

Further Reading

Bates, H. W. (1862). Contributions to an insect fauna of the Amazon valley: Lepidoptera: Heliconidae. *Trans. Linn. Soc. London* **23,** 495–566.

Bates, H. W. (1863). "A Naturalist on the River Amazon." Murray, London.

Beccaloni, G. W. (1997). Ecology, natural history and behaviour of ithomiine butterflies and their mimics in Ecuador (Lepidoptera: Nymphalidae: Ithomiinae). *Trop. Lepidoptera* **8,** 103–124.

Brower, L. P. (ed.) (1988). "Mimicry and the Evolutionary Process." University of Chicago Press, Chicago.

Edmunds, M. (1974). "Defence in Animals. A Survey of Anti-predator Defences." Longman, New York.

Jiggins, C. D., Naisbit, R. E., Coe, R. L., and Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. *Nature* **411,** 302–305.

Kapan, D. D. (2001). Three-butterfly system provides a field test of Müllerian mimicry. *Nature* **409,** 338–340.

Mallet, J. (1993). Speciation, raciation, and color pattern evolution in *Heliconius* butterflies: Evidence from hybrid zones. *In* "Hybrid Zones and the Evolutionary Process" (R. G. Harrisson, ed.). Oxford University Press, New York.

Mallet, J., and Joron, M. (1999). Evolution of diversity in warning color and mimicry: Polymorphisms, shifting balance and speciation. *Annu. Rev. Ecol. Syst.* **30,** 201–233.
Müller, F. (1879). *Ituna* and *Thyridia:* A remarkable case of mimicry in butterflies. *Trans. Entomol. Soc. London* **1879.**
Pasteur, G. (1982). A classificatory review of mimicry systems. *Annu. Rev. Ecol. Syst.* **13,** 169–199.
Rowe, C. (ed.) (2001). "Warning Signals and Mimicry." Kluwer Academic, Dordrecht. [Special issue of *Evolutionary Ecology,* 1999, **13(7/8)**]
Turner, J. R. G. (1977). Butterfly mimicry: The genetical evolution of an adaptation. *Evol. Biol.* **10,** 163–206.
Wallace, A. R. (1879). The protective colours of animals. *In* "Science for All" (R. Brown, ed.). Cassell, Petter, Galpin, London.
Wickler, W. (1968). "Mimicry in Plants and Animals." McGraw–Hill, New York.

Mites

Barry M. OConnor
University of Michigan

Mites comprise the Acari, which are the largest group within the arthropod class Arachnida, with over 48,000 described species. This number is misleading because it is estimated that only between 5 and 10% of all mite species have been formally described. In contrast with other arachnid groups such as spiders and scorpions, mites are distinctive in both their small size (adult body length ranging from 0.1 to 30 mm) and their ecological diversity. Some mites are predators, like almost all other arachnids, but mites may also feed on plants, fungi, or microorganisms or as parasites on or in the bodies of other animals. Mites are among the oldest known groups of arthropods, with a fossil record beginning in the Devonian period.

BODY STRUCTURE

Unlike insects, with bodies divided into head, thorax, and abdomen, the arachnid body is ancestrally divided into two functional units, the prosoma (the first six body segments) and the opisthosoma (the remaining segments). The body of a mite is further modified in that these original units are fused. A secondary subdivision separates the first two body segments into a structure termed the gnathosoma, specialized for feeding, and the remainder of the body, termed the idiosoma, containing organs of locomotion, digestion, and reproduction. Most mites show no evidence of external body segmentation, other than the serial appendages. The gnathosoma bears the first two pairs of appendages, the chelicerae, which may retain the ancestral chelate, or pincer-like form, or may be highly modified as stylets for piercing and sucking; the pedipalps, which may be almost leg-like, are strongly modified for grasping prey or attaching to a host or highly reduced. The anterior idiosoma typically bears four pairs of walking legs, the first pair of which may be modified as antenna-like, sensory structures. Legs may also be modified for attaching to a host. Occasionally legs of males are modified for grasping a female during mating or for intraspecific combat.

The mite's body cuticle may be entirely soft, divided into a number of hard, sclerotized plates, or almost entirely sclerotized. In a few mites, crystalline, mineral salts also strengthen the cuticle. Such modifications balance the needs for flexibility in movement and protection from predators. The body surface bears setae, typically hair-like sensory organs, arranged in characteristic patterns in different subgroups of mites. Setae are primarily hair-like, but may take on an incredible variety of shapes, from thick spines, to flat plates, to highly branched, feather-like forms. The pedipalps and legs also bear tactile setae as well as chemosensory structures termed solenidia, which are organs of smell and taste, and other specialized sensilla that are sensitive to infrared radiation. Simple eyes, or ocelli, may be present on the anterior idiosoma, and specialized sensory organs, the trichobothria, on the anterior idiosoma or legs may detect vibrations or electric fields.

Like other arthropods, the inside of a mite's body is a hollow cavity, the hemocoel, in which the internal organs are surrounded by fluid, the hemolymph. Hemolymph distributes food materials and waste products and contains hemocytes, which are the cells that serve as the mite's immune system, but it does not contain oxygen-binding proteins as are found in the blood of vertebrates and some other arthropods. The mite's digestive system is divided into the three parts typical of arthropods: foregut, midgut, and hindgut. The midgut may be divided into diverticulae for food storage, particularly in parasitic mites. Some mites lack a connection between the midgut and the hindgut; these mites feed only on fluids and do not defecate. The hindgut in these mites is transformed into an excretory organ for elimination of nitrogenous wastes. Other mites, with entire guts, may have Malpighian tubules, like insects, extending from the junction of the midgut and hindgut as excretory organs. The internal reproductive system typically consists of a single ovary (paired in the Astigmata) in the female and paired testes in the male. Females typically possess a spermatheca for sperm storage after insemination, and both sexes have various accessory glands and ducts to the exterior as part of the system. Tracheal systems for respiration have evolved independently a number of times in the Acari. These open at spiracles, or stigmata, on various parts of the body in different groups. Other mites lack any respiratory system, and gas exchange occurs through the cuticle in these groups.

CLASSIFICATION OF MITES

The classifications of mites used by various authors vary considerably in the number of higher categories recognized and

the hierarchical ranking of the various groups. The simplest system, used by Walter and Proctor (1999), recognizes three orders within Acari: Opilioacariformes, Parasitiformes, and Acariformes. The Opilioacariformes, comprising a single family with about 20 species, is considered the most primitive. These mites are relatively large (2–3 mm) and resemble small opilionids in their general form, having a leathery cuticle that retains traces of external segmentation. These mites resemble the Parasitiformes in having a tracheal system opening laterally on the body, but they have four pairs of stigmatal openings in contrast to the single opening of the Parasitiformes. Opilioacarids resemble some Acariformes in feeding on solid food particles and bearing a pair of rutella, which are sclerotized food-processing structures located near the ventral apex of the gnathosoma.

The order Parasitiformes is a diverse group comprising 76 families divided among three suborders: Gamasida (or Mesostigmata), Ixodida, and Holothyrida. Compared with the Acariformes, this order is morphologically relatively conservative, with most species retaining the same basic body plan. The Holothyrida includes 3 families and around 30 species of heavily sclerotized, predatory or scavenging mites of tropical regions. The Ixodida, or ticks, includes 3 families and around 850 species exclusively parasitic on vertebrate hosts. The vast majority of parasitiform mites are included in the Gamasida, with 70 families. Most gamasid mites retain the ancestral predatory life-style, but the group includes a number of parasites of vertebrates and other arthropods, a few mites which feed on pollen or fungi, and one small group of detritivores capable of feeding on solid food particles.

The order Acariformes is the largest and most diverse group of mites, in terms of both its morphology and its ecological diversity. Hundreds of families are recognized, and over 30,000 species are included. Acariform mites are characterized by the internalization of the basal leg segment, the coxa, leaving the next segment, the trochanter, as the first functional leg segment. Most acariform mites also possess structures termed "genital papillae." While associated with the genital region in the postlarval instars, these structures are actually osmoregulatory organs.

The order Acariformes is conveniently divided into two suborders, Trombidiformes (largely equivalent to the Prostigmata of some authors) and Sarcoptiformes (including the Oribatida and Astigmata of some authors). Most trombidiform mites have tracheal systems opening on or near the gnathosoma. Many have strongly modified chelicerae adapted for piercing animal prey, plant tissue, or the skin of a host animal. Sarcoptiform mites ancestrally feed on solid food and have gnathosomal rutella, like the Opilioacariformes. Tracheal systems opening at the leg bases or anterior dorsal idiosoma have evolved independently several times in this group. Sarcoptiform mites are most diverse in soil habitats, but many have adapted to patchy habitats and have developed commensal or parasitic associations with vertebrates and other arthropods (Fig. 1).

LIFE CYCLES, DEVELOPMENT, AND REPRODUCTION

Most mites exhibit a fixed developmental pattern, passing through the same number of instars regardless of how much food is available. The most complete pattern consists of egg, prelarva, larva, protonymph, deutonymph, tritonymph, and adult. The prelarva and larva are distinguished by having only three pairs of legs; the fourth pair is added at the protonymphal molt. Other immature stages are distinguished from each other by a characteristic pattern of additions of leg and body setae. The prelarva is typically a short-lived stage, either passed completely within the egg or, if it actually hatches, having a highly regressive morphology. The few active prelarvae known do not feed and typically begin the molt to the larval stage within hours after hatching from the egg. This life cycle is found in the Opilioacariformes and ancestrally in the Acariformes. Reductions from this number of instars appear in other groups of mites. Within the Parasitiformes, the prelarva is not observed, and the tritonymph is retained only in some Holothyrida. In one family of Ixodida, the Argasidae, the number of nymphal instars is not fixed. Molts take place after each blood meal in these ticks, but the adult morphology develops only when the mite reaches a minimum body size. In many trombidiform mites, the last nymphal instar is suppressed, and in some extreme cases, all immature stages are suppressed and passed within the body of the female mite. After an extreme form of engorgement termed "physogastry" on fungal food or host-insect hemolymph, these females give birth to fully developed adults. Another developmental pattern observed in the large trombidiform subgroup, the Parasitengona, involves alternation of active and inactive instars. Active stages in this life cycle include the larva, deutonymph, and adult, while the prelarva, protonymph, and tritonymph are morphologically regressive, inactive stages.

Mites exhibit a variety of reproductive strategies and modes of sperm transfer. Ancestrally, mites appear to practice indirect sperm transfer, with males producing and depositing a package of sperm, termed a spermatophore, on the substrate. Females then take the spermatophores into their reproductive tract. This type of reproduction is found in most acariform subgroups, and individuals of the two sexes may or may not be in close contact at the moment of insemination. In known parasitiform groups, males typically use their chelicerae to assist in directly inserting a spermatophore into the female's primary genital opening (as in Ixodida and primitive Gamasida), or the male chelicerae bear an organ termed the spermatodactyl which is used to transfer sperm from the male's genital opening into secondarily developed sperm induction pores near the bases of the female's legs. These paired openings lead to a median spermatheca which is connected directly to the ovary, where fertilization takes place. Direct insemination involving the development of an intromittent organ, the aedeagus in the male, has appeared independently in several groups of acariform mites. Secondary sexual dimorphism typically accompanies

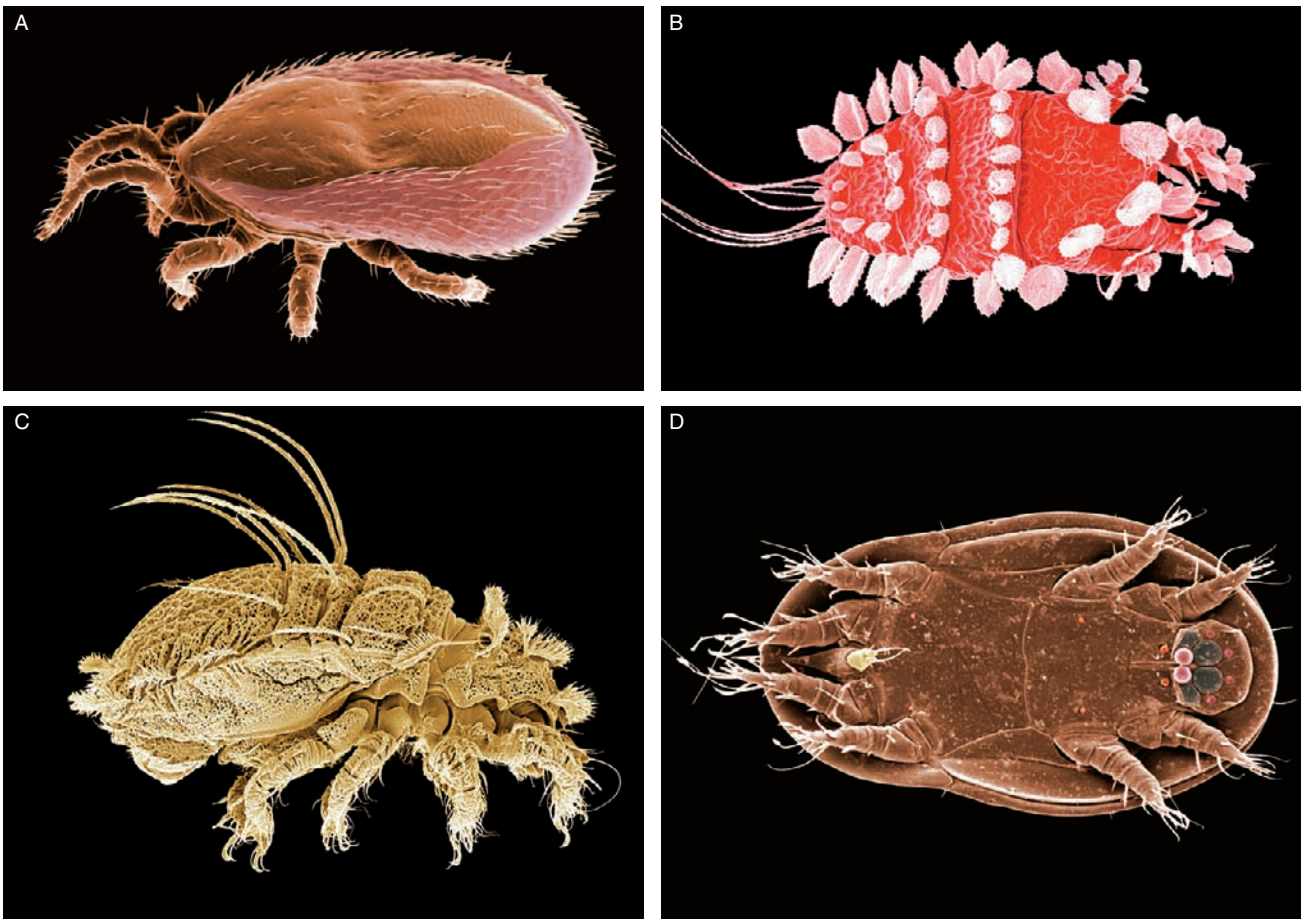

FIGURE 1 (A) Tropical rat mite, *Ornithonyssus bacoti,* a vertebrate parasite. (B) Peacock mite, *Tuckerella* sp., a plant-feeding mite. (C) *Cosmochthonius* sp., a soil mite. (Photographs by D. E. Walter and C. Meacham.) (D) Heteromorphic deutonymphs, phoretic on a predatory mite. (Photograph by D. E. Walter.)

direct mating, with males often having modified appendages for holding the female during mating. Many such males also practice precopulatory guarding of immature females, either merely waiting near a juvenile female about to molt or actively attaching to her. This last behavior is taken to an extreme in the sarcoptiform family Chirodiscidae, species of which live on the hairs of mammals, in which immature females are legless and unable to move. They must be found upon hatching by an adult male who uses an elaborate clasping organ to attach to, and carry about, the juvenile female until her legs appear at adult eclosion.

Sex determination mechanisms and reproductive modes also vary widely throughout the Acari. Some mites are diploid in both sexes, with males having either a Y chromosome or no sex chromosome. Other mites are arrhenotokus, in that females are diploid and males haploid. Such males develop from unfertilized eggs. An unusual reproductive mode, termed parahaploidy, is found in some Gamasida. In these mites, fertilization is necessary for egg development, but in males, the paternal genome is inactivated shortly after the first embryonic cell divisions, and adult males are functionally haploid. Finally, thelytoky, or all-female parthenogenesis, is found in many groups of mites. Such mites reproduce clonally, with diploid eggs developing directly into females without fertilization.

ECOLOGY

Mites exhibit a breadth of ecological interactions unmatched in any other arthropod group. Mites may be found in all geographic provinces, from tropical rainforests to arctic tundra and rocky outcrops in Antarctica and from desert habitats to the deep ocean trenches. They dominate the microarthropod fauna of the soil where they may be found several meters deep or even in groundwater. They occur in all types of aquatic habitats, including freshwater lakes, streams, seepage areas, and even hot springs. Unlike insects, mites are also quite diverse in marine habitats, ranging from the intertidal zone to the deep trenches.

Soil Mites

A single square meter of temperate forest soil may contain upward of 250,000 mites, belonging to a hundred different families. In the litter and upper layers of organic soil, mites play many roles in food webs based on decaying plant materials.

Gamasid and some trombidiform mites are the dominant predators in such systems, feeding on nematodes, small annelids, collembolans, other mites, and the eggs of insects. These predatory mites have developed several strategies for prey detection and capture, from active, foraging species in the gamasid families Laelapidae and Parasitidae, and the trombidiform families Raphignathidae and Cunaxidae, to more sedentary species in the trombidiform families Caeculidae and Cheyletidae, with palps or forelegs modified as traps for unwary prey. Most soil-inhabiting mites, however, are detritivores or fungivores, feeding directly on decaying organic materials or on fungi or microorganisms growing upon them. The greatest diversity of detritivores belongs to the sarcoptiform subgroups collectively known as oribatid mites. These mites are typically slow moving and may take up to 3 years to complete the life cycle. Adults tend to be well sclerotized as a defense against predators, while soft-bodied juveniles may burrow into substrates to avoid predation. Oribatid mites are primarily detritivores, feeding directly on particulate organic material. Others preferentially scrape decaying leaves for their microbial or fungal floras. Despite their numbers, compared with earthworms or other larger soil invertebrates, mites actually process a relatively smaller amount of organic material and are thus of less importance in converting biomass to nutrients again available to plants. However, in terms of the cycling of particular nutrients, notably calcium, mites play an essential role. Mites are also extremely important in the dispersal of bacterial and fungal agents of organic decomposition. Mites feeding on such substrates ingest bacteria or spores that can often pass undigested through the mites' guts. The movement of the mites through the soil, with the associated deposition of fecal pellets containing decomposer propagules, provides a much more efficient dispersal of these organisms than simple physical processes.

Specialized soil types have specialized mite faunas. Dry, sandy, and nutrient-poor soils typically harbor a fauna of primitive acariform mites that show little morphological change from their Devonian fossil ancestors. This entire community may consist of such living fossils, with this type of nutrient-poor soil likely similar to the original terrestrial environment at the time of the first land-colonizing animals. Another highly specialized fauna of mites inhabits the deeper layers of mineral soils. Because there is little organic material that filters down to these layers, many of the mites feed directly on the sparse microbial flora or are predators on nematodes that are able to extract nutrients from the limited resources. Deep soil mites tend to be quite small and soft-bodied and may be elongated to allow for movement through very tight spaces between mineral soil particles. Most are effectively aquatic because the deep soils are often saturated, with the interstitial spaces filled with water.

Mites in Patchy Habitats

Mites living in large, continuous habitats such as the soil and litter layers generally have limited dispersal capability. Being small and lacking wings, mites would seem to be limited to such habitats. However, mites also form major components of the microarthropod communities associated with patchy habitats, which are those separated by distances greater than mites' ability to walk. Such mite communities occur in habitats such as decaying logs, dung and manure, carrion, fungal fruiting bodies, nests of insects and vertebrates, and other concentrations of organic matter such as treeholes, sap flows, and other specialized habitats associated with plants.

Common in patchy habitats are specialized gamasid mites in the families Parasitidae, Macrochelidae, Laelapidae, Digamasellidae, and Uropodidae. These species are typically predators on insect eggs and larvae, particularly those of the Diptera that also frequent patchy habitats, and nematodes. Among the Trombidiformes, species in the subgroup called Heterostigmata are largely associated with patchy habitats, feeding primarily on fungi. The most diverse group of mites in patchy habitats is the sarcoptiform subgroup Astigmata, which appears to have had its origin in such associations.

All of these groups, and some others as well, are able to exploit these habitats, which are generally unavailable to most mites, through a specialized dispersal mode termed phoresy. Phoresy involves one organism utilizing another, larger organism to facilitate its dispersal. In all of the mentioned groups of mites, one life stage is typically specialized for phoretic dispersal on an insect, myriapod, crustacean, or mammal host. Gamasid mites disperse either as inseminated females or as deutonymphs, which is the final juvenile stage in this group. Female laelapid and macrochelid mites typically attach to insect carriers by grasping host setae or other structures with their chelicerae. Parasitid and digamasellid mites disperse as deutonymphs, often in the space under the elytra of beetles, and may roam freely over the insect's body. In the Uropodidae, the deutonymph is often specialized for dispersal and may attach to the host by secreting a sticky substance from posterior ventral glands. This material is drawn into a stalk that hardens in air and connects the mite to its host.

Heterostigmatid mites disperse as adult females, with many species exhibiting a polymorphism in this stage. Nondispersing females have normally developed anterior legs and are not attracted to insect or mammal hosts. Dispersing females, or phoretomorphs, have very enlarged forelegs with a grasping claw that allows attachment onto insect setae or mammalian hair. In the Astigmata, the deutonymph is highly specialized for dispersal. These deutonymphs look nothing like the preceding or following instars, having no mouth or mouthparts, but bearing suckers or claspers at the posterior end of the body for attaching to a host. They are typically heavily sclerotized and able to withstand major fluctuations in environmental conditions. Many astigmatid mites inhabit naturally occurring patches of organic matter such as decaying wood or mushrooms and disperse on any insects that frequent the habitat. Others have developed closer associations with particular insects, notably nest-building bees, wasps, ants, and termites, and depend on these insects not only for dispersal, but also for

creation of the habitat in which the mites live. Still other astigmatid mites have adapted to the nests of vertebrates, with many species inhabiting mammal nests specialized for phoretic dispersal on the mammalian host itself. Species living in birds' nests still disperse on nest-inhabiting insects such as beetles and fleas.

A few of these phoretic associations between mites and insects have become mutualistic, with mites providing either "cleaning services" or "pest control" for their hosts. Some species of Old World carpenter bees may carry several species of mites on their bodies. The astigmatid mites in these communities are kleptoparasites, feeding on the provisions intended for the bee's offspring. A female bee may also carry trombidiform mites in the genus *Cheletophyes* (family Cheyletidae) in specialized pouches, termed acarinaria, on the thorax. These mites are obligate predators of the astigmatid mites. The same female bee may also have a large acarinarium in the anterior part of the abdomen, carrying large (2–3 mm) gamasid mites in the genus *Dinogamasus* (family Laelapidae). These mites have modified chelicerae that scrape the cuticle of the bee larva and remove potentially pathogenic microorganisms and fungal spores as well as cuticular exudates. Some astigmatid mites in the family Histiostomatidae are mutualists in the nests of sweat bees (family Halictidae). Feeding stages of the mites have highly modified chelicerae for filter feeding. These mites wander over the nectar and pollen provisions, straining potentially harmful microorganisms. Deutonymphal mites ride in a rudimentary acarinarium on the propodeum or anterior gaster of the female bees.

Aquatic Mites

A number of different groups of mites have successfully colonized and diversified in aquatic habitats. The most diverse of these, with over 40 families, and 5000 species, is a group within the trombidiform subgroup Parasitengona that is termed Hydracarina or, more simply, water mites. This lineage is characterized by the enlargement or multiplication of the genital papillae (often termed acetabula in this group), acariform organs of osmoregulation that allow these mites to maintain ionic balance in hypoosmotic freshwater environments. The parasitengone life cycle is unusual, with its alternation of active and inactive stages, and in most terrestrial and aquatic species, the larva is parasitic, typically on an adult, flying insect. This parasitic larva not only acquires nutrients by feeding on its host, but also is able to disperse over some distance while on the host. The deutonymph and adult stages are typically active predators on other arthropods or their eggs. Water mites are primarily inhabitants of freshwater habitats including temporary ponds, permanent ponds and lakes, streams and rivers, and interstitial waters. One family of water mites, the Pontarachnidae, has invaded marine, intertidal waters and has lost the genital papillae, while another, the Thermacaridae, is restricted to hot springs and capable of surviving temperatures close to 50°C. Mites in standing waters may crawl about on the substrate or aquatic vegetation, but many species have morphological adaptations for active swimming. These include anterior displacement of the leg bases and long setae, termed swimming hairs, on the legs. These mites actively seek and capture aquatic crustaceans and small insect larvae. Mites inhabiting running waters are typically smaller, with flattened bodies, robust legs, and often sclerotized plates on the body. These mites crawl on and in the substrate, feeding on the eggs of aquatic insects and other microinvertebrates. Some of these mites are specialized predators of the eggs of the same insect species used as hosts by their larvae. Some stream-inhabiting species live deep in the interstitial waters, often having quite elongate bodies for squeezing through the spaces between rock and sand particles. Many water mites are brightly colored, either retaining the red color common among the ancestral, terrestrial Parasitengona or becoming a cryptic blue or green. Some water mites have modified the ancestral parasitengone life cycle by producing fewer, larger eggs. Larvae hatching from these eggs transform to deutonymphs without feeding or dispersing on a host.

Another relatively large group of aquatic mites forms a separate trombidiform lineage, the family Halacaridae. These mites are most diverse in marine habitats, with most species found in intertidal waters. Some halacarids, however, have been collected in abyssal depths up to 7000 m. Feeding ecology of halacarids varies, with some species retaining the ancestral predatory behavior, while others feed on algae or as parasites on crustaceans, echinoderms, or cnidarians. Some halacarids have reinvaded freshwater habitats, presumably via groundwater connections. Such mites are often collected from well water, and a number of species are restricted to freshwater habitats.

Other groups of mites contain aquatic taxa, but none has diversified to the extent seen in the water mites and Halacaridae. Mites in the oribatid family Hydrozetidae are often collected on aquatic vegetation, while those in the family Trimalaconothridae occur in the substrates of ponds and streams. The sarcoptiform group Astigmata includes the family Hyadesiidae, all species of which live in marine, intertidal habitats. These mites are unusual among the Astigmata in living in more or less continuous habitats, and they have lost the dispersing deutonymph from the life cycle. Some species in the family Algophagidae live in brackish waters, and one is known from a fast-flowing river. Other Astigmata live in temporary aquatic habitats, such as water-filled treeholes and other phytotelmata, or water-filled plant cavities, such as pitcher plants, bromeliads, the leaf axils of aroids, and the flower bracts of heliconias and related plants. These species still retain the phoretic deutonymph that disperses on an insect host. Relatively few gamasid mites have become aquatic, but some species in the family Ascidae live in phytotelm habitats or regularly flooded swamp or flood-plain soils. Some of these have a modified cuticle around their respiratory openings that functions as a plastron, holding a bubble of air against the spiracle when the mite is submerged.

Mites on Plants

Unlike all other arachnid groups, several groups of mites have evolved the ability to feed on living plant tissue. Most species belong to one of several lineages of Trombidiformes, each of which has independently evolved this capability, but all share the modification of the chelicerae into piercing stylets. One lineage, the superfamily Tetranychoidea, contains the spider mites and their relatives. Spider mites (family Tetranychidae) are so named because some species utilize silk in constructing webbing on leaves or pads for oviposition and also for dispersal via ballooning much in the manner of some spiders. Silk production is not unique to this group, however, because it is also found in related trombidiform groups not associated with plants. Tetranychoid mites have elongate cheliceral stylets that pierce leaf or root tissue and feed on cell contents or on interstitial fluids. Most species are relatively host specific and do little damage, but some, such as the twospotted spider mite, *Tetranychus urticae,* are polyphagous and are serious pests of agricultural crops, particularly herbaceous annuals, such as beans, and fruit trees. Another tetranychoid group, the false spider mites, or flat mites (family Tenuipalpidae), also includes serious agricultural pests.

A second lineage of plant-feeding mites, the Eriophyoidea, contains extremely tiny species with a highly modified body form. These elongate, worm-like mites have only two pairs of legs at the anterior end of the body, but possess a sucker at the posterior end and move inchworm fashion over plant surfaces. All are obligate plant feeders, using their short stylets to pierce individual cells. Most species are highly host specific, and a single plant species may harbor many species in this group, most of which simply wander over the leaf surfaces. Large populations of such mites may cause loss of color in leaves, leading to the common name rust mites. Another common name, gall mite, refers to the ability of some species to induce characteristic galls on leaves, buds, stems, flowers, or fruits of their host plants. Salivary chemicals mimic certain plant growth hormones and induce the formation of galls in which the mites live. Simple erineum galls form when epidermal cells produce elongate hairlike growths upon which the mites feed. Pouch galls are like erinea but actually form into elongate cavities within which the mites live. Mite-induced proliferation of woody tissue causes "witches' brooms" on trees. Although rusting and gall formation are often unsightly and may affect fruit set in orchard crops, the most important effects of eriophyoid mites on agricultural systems are as vectors of viral pathogens such as wheat streak mosaic virus. On the other hand, other, highly host-specific, eriophyoid mites have been used as virus vectors in the biological control of weeds.

Other plant-feeding mites occur in the families Pentheleidae, including the redlegged earth mite, *Halotydeus destructor,* a serious pest of grasses and herbaceous plants in the Southern Hemisphere, and Tarsonemidae. This last family includes such serious agricultural pests as the broad mite, *Polyphagotarsonemus latus,* which, true to its scientific name, is a polyphagous pest of many agricultural crops.

Other mites living on plants are beneficial as predators of phytophagous mites. Chief among these are species in the gamasid family Phytoseiidae. These mites range from generalist to specialist predators, often attacking economically important species of spider mites, flat mites, and eriophyoids. Several species of Phytoseiidae are commercially marketed for biological control of these pests.

Parasitic Mites

A great many lineages of mites contain parasites of vertebrate and invertebrate animals, some of which are of importance in human and veterinary medicine. Most important of these is the Ixodida, the ticks, but the Gamasida contains a diversity of parasites of reptiles, birds, mammals, insects, myriapods, and crustaceans, most of which belong to the mite superfamily Dermanyssoidea. Among the vertebrate parasites, several different types of parasitism occur. The simplest of these is facultative parasitism, in which typically nest-inhabiting predators may feed opportunistically from a wound on a bird or small mammal host. Other nest-inhabiting mites are obligate parasites, but get on the host only to feed. Notable among these are the northern fowl mite, *Ornithonyssus sylviarum* (family Macronyssidae), and the chicken mite, *Dermanyssus gallinae* (family Dermanyssidae), both of which parasitize a variety of wild birds and domestic poultry and will bite people. Many of these mites have chelicerae modified for piercing and sucking blood or tissue fluid. Finally, some gamasids have become permanent parasites, spending all their time on the host's body. These may have enlarged claws or spurs on the body for holding onto the host. Several different groups of dermanyssoid mites have become endoparasites, living in the respiratory tract of snakes, birds, and some mammals, notably dogs and seals. Some species in the family Rhinonyssidae can cause respiratory distress in cage birds. Other endoparasites inhabit the ear canals of ungulates such as cattle and goats. Some parasitic gamasids act as vectors of bacterial, viral, and protozoan pathogens to their normal hosts, but only one, the dermanyssid *Liponyssoides sanguineus,* acts as a vector for a bacterial pathogen from mice to humans, causing the disease known as rickettsialpox.

Other dermanyssoid mites are parasitic on arthropods, with the most important being the honey bee parasite, *Varroa destructor* (family Laelapidae). This mite is responsible for the worldwide decline in populations of the European honey bee, *Apis mellifera.* The mites feed on hemolymph of bee larvae, causing the adult bee that develops to have aborted wings that prevent the bee from foraging. Buildup of mites in a bee colony causes its destruction over time. This mite became a pest after colonizing *A. mellifera* from its ancestral host, the Asian honey bee, *Apis cerana.* In the normal host, this mite is not pathogenic to the colony because populations do not reach damaging levels.

Several groups of trombidiform mites have become parasitic, the largest of which is the Parasitengona. This group includes the water mites discussed above, but also a number of terrestrial groups. Larvae of most species parasitize insects, in

which they may reduce fecundity or longevity. Larvae of one family, the Trombiculidae, or chiggers, parasitize vertebrate hosts. All groups of terrestrial vertebrates may serve as hosts for this very large group of parasites. Most of the over 5000 described species are known only from their parasitic larval stage. Chiggers feed on tissue fluid and lysed host tissue. On most hosts, they appear not to affect the host negatively, but species biting humans induce an immune reaction that not only causes the death of the chigger, but causes a relatively long-lasting, itchy lesion. Different species of chiggers have achieved "pest" status in various parts of the world. Most are merely irritants to human hosts, but some species in the genus *Leptotrombidium,* ranging from Japan and Korea, west to Pakistan, and south to northern Australia, act as vectors of a serious bacterial pathogen from rats to humans. The disease, termed scrub typhus or tsutsugamushi disease, can be fatal if untreated. A chigger is able to vector the pathogen, the bacterium, *Orientia tsutsugamushi,* despite parasitizing only a single host in its lifetime, because the pathogen remains in the mite's body and enters the eggs of the female mite. Thus, larval chiggers are capable of transmitting the pathogen at hatching.

Another diverse group of trombidiform parasites is included in the superfamily Cheyletoidea. Different families in this group parasitize reptiles, birds, and mammals, with the family Demodicidae containing two species specifically parasitic on humans. *Demodex folliculorum,* an elongate, worm-like mite, inhabits the hair follicles of the face and occasionally other body regions, whereas *Demodex brevis* lives in the sebaceous, or sweat, glands in the skin. Although heavy infestations have been linked to acne rosacea, most people harbor these mites with no discernable effect. Other demodicids can be more pathogenic in their normal hosts, such as *Demodex canis* in dogs and *Demodex bovis* in cattle. The former can cause a mange condition, with hair loss and irritated skin, especially in puppies, while the latter causes large nodules full of mites to form in the skin. Species in other cheyletoid families parasitize birds, living on the skin, in feather follicles, or inside feather quills. One interesting group in the family Cheyletidae lives within feather quills, but feeds as predators on other quill-inhabiting mites.

The trombidiform lineage Heterostigmata includes many parasites of insects. The honey bee tracheal mite, *Acarapis woodi* (family Tarsonemidae), is of considerable economic importance as a pest in the respiratory system of honey bees. Other parasitic Tarsonemidae inhabit the defensive glands of coreid Hemiptera, one of the most unusual habitats known, even among mites!

Among the Sarcoptiformes, the Astigmata includes a great diversity of parasitic species, the hosts of which include mammals, birds, and insects. Certain nest-inhabiting astigmatid mites that ancestrally dispersed via phoretic deutonymphs have modified the nature of the association. Instead of merely attaching to the hair or skin of the host and simply dispersing, deutonymphs in several groups in the superfamily Glycyphagoidea associated with small mammal hosts, and species in the family Hypoderatidae with bird hosts, enter either the hair follicles or the subcutaneous tissue of their host. Despite lacking a mouth and functional gut, these deutonymphs engorge, with some Hypoderatidae increasing their body volume up to 1000-fold. The mode of nutrient acquisition in these parasites is unknown, but some are able to complete the remaining, free-living part of the life cycle in the host's nest without additional food.

Other groups of astigmatid mites have become permanent parasites of birds or mammals, eliminating the deutonymph from the life cycle. Among mammal hosts, these mites are most diverse on marsupials, rodents, insectivores, primates, and bats, with relatively few occurring on carnivores or ungulates. Most are relatively nonpathogenic, feeding primarily on sebaceous materials on the hair shafts. Others, however, can cause problems for their hosts.

Species in the family Psoroptidae live on the host's skin or in the ears and feed by abrading the skin with their chelicerae and imbibing tissue fluids. These mites irritate the skin and cause itching. Several species of psoroptid mites occur on domestic animals, notably the carnivore ear mite, *Otodectes cynotis,* common in cats and dogs, and the scab mites in the genera *Psoroptes* and *Chorioptes* in horses, cattle, sheep, and others. The sheep scab mite, *Psoroptes ovis,* particularly causes economic damage by causing loss of wool.

Probably most important among parasitic astigmatid mites are species in the family Sarcoptidae. Commonly known as mange mites, species in several genera can parasitize humans and domestic animals. Naturally most diverse on marsupials, bats, primates, and rodents, several species have been able to colonize new hosts. *Sarcoptes scabiei* is ancestrally a parasite of humans, causing the skin disease scabies. Like other sarcoptids, these mites burrow into the superficial layers of the skin. In healthy humans, this disease is an itchy annoyance, but in immune-compromised individuals, a serious condition known as crusted scabies can develop in which the patient may harbor millions of mites in large, crusty lesions all over the body. *S. scabiei* has also been able to colonize many domestic animals, notably dogs, pigs, cattle, camels, and others, in which the disease known as sarcoptic mange can be fatal due to the large mite populations and secondary bacterial infections.

Many other astigmatid mites parasitize birds, in which, again, most do not cause harm. Feather mites may be very diverse on an individual bird, with one parrot species known to harbor almost 40 species. Like their fur mite counterparts on mammals, these mites feed on skin oils and do not harm the host. Others, however, may parasitize the feather follicles, skin, or respiratory tract. Skin-inhabiting species in the family Knemidokoptidae can be quite pathogenic in domestic poultry, cage birds, and wild species. Endoparasitic species in the family Cytoditidae live in the air sacs and can cause respiratory distress in poultry.

IMPORTANCE OF MITES

As indicated above, there are a number of instances in which mites are important to humans. Many species are serious

pests of agricultural crops, either through direct damage or indirectly as vectors of plant pathogens. Other species are parasitic on domestic animals and cause losses in meat, egg, and fiber production. Others, such as the human scabies mite, are direct agents of human disease or, as in the case of chiggers and ticks, vectors of pathogens. Other mites may affect humans by infesting stored food products. Many species of Astigmata are known as stored-product mites because they have moved from their ancestral rodent nest habitats into human food stores. Such mites may also cause damage in animal feed by causing allergic reactions in livestock and are also known to cause skin irritation in humans handling infested materials. A related group of astigmatid mites, also ancestrally nest inhabiting, is the family Pyroglyphidae. These mites have colonized human habitations from bird nests and are the primary source for allergens in house dust. Commonly known as "house dust mites," species particularly in the genus *Dermatophagoides* produce many proteins that induce allergic responses in sensitive individuals. House dust allergy may take the form of respiratory distress or skin irritation. Mites typically inhabit beds, chairs, and carpets in houses, and their shed skins and feces provide the bulk of the allergens in house dust extracts.

On the other hand, as indicated above, some mites are beneficial to humans in their role as biological control agents against agricultural pests. Also, the natural role of mites in providing "ecosystem services" in the form of nutrient cycling cannot be overlooked.

See Also the Following Articles

Arthropoda and Related Groups • Chiggers and Other Disease-Causing Mites • Medical Entomology • Neosomy • Plant Diseases and Insects • Predation • Ticks

Further Reading

Evans, G. O. (1992). "Principles of Acarology." CABI, Wallingford, UK.

Krantz, G. W. (1978). "A Manual of Acarology," 2nd ed. Oregon State University Press, Corvallis.

Walter, D. E., and Proctor, H. C. (1999). "Mites: Ecology, Evolution and Behaviour." CABI, Wallingford, UK.

Woolley, T. A. (1988). "Acarology: Mites and Human Welfare." Wiley, New York.

Molting

Lynn M. Riddiford
University of Washington

Molting is the process of producing a new cuticle and the subsequent shedding (or ecdysis) of the old cuticle. This molt is orchestrated by a series of hormones so that it can be triggered by both internal and external cues.

THE MOLTING PROCESS

Cuticle Production

The cuticle is the outer covering of the insect and is its exoskeleton to which the muscles are attached (Fig. 1). The outermost layer is called the epicuticle; under this is the exocuticle followed by the endocuticle. In some systems, the exo- and endocuticle are classed together as the procuticle. In some insects, only the epi- and exocuticle are deposited before ecdysis, with the endocuticle following ecdysis, whereas in others, some endocuticle may be deposited before ecdysis. The epicuticle is composed of only protein, whereas the exo- and endocuticle contain both chitin and protein in varying proportions depending on the type of cuticle, i.e., whether rigid or flexible. Chitin is a polymer of *N*-acetylglucosamine and can be cross-linked to the protein components of the cuticle in a process called sclerotization or hardening, which usually occurs in the exocuticle just after the shedding of the old cuticle and expansion of the new cuticle. After sclerotization the insect is able to move, feed, fly, etc. The rigid parts of the cuticle are then set and cannot be expanded, whereas flexible cuticle may expand either by a simple unfolding of the new epicuticle or in response to hormonal signals. When the epicuticle has completely unfolded, further expansion is impossible and the larva must molt in order to grow further. Molting is also necessary at the end of larval life for metamorphosis.

The epidermis is a single cell layer that produces the cuticle that lies above it (Fig. 1). During the intermolt period, the epidermis actively deposits lamellate endocuticle, especially in those regions where the cuticle is extensible. The chitin and protein are secreted as plaques at the tips of the

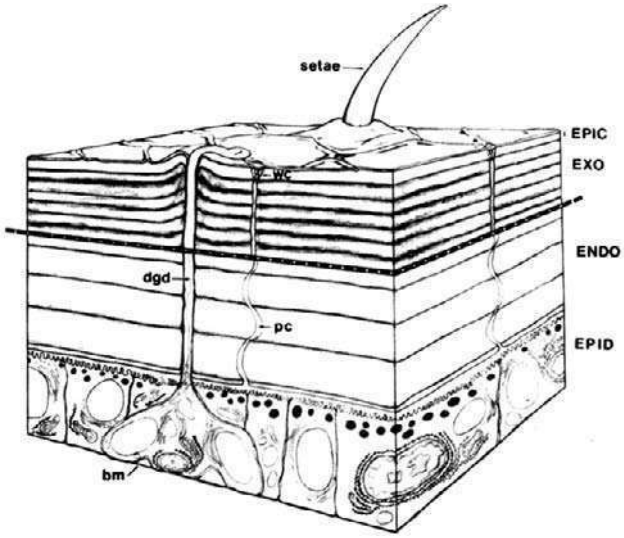

FIGURE 1 Diagram of the relationship of the epidermis (EPID) to the overlying cuticle that it produces. EPIC, epicuticle; EXO, exocuticle; ENDO, endocuticle; bm, basement membrane; dgd, dermal gland (Verson's gland) duct; pc, pore canal; setae, cuticular hair; wc, wax canal. [From Hadley, N. (1982). *J. Exp. Zool.* **222,** 239–248, Copyright © 1982. Reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.]

microvilli at the apical surface of the epidermal cells. Above the plaques in the extracellular space, the cuticle arises by self-assembly of the chitin microfibrils and the secreted proteins. As the larva grows, the epidermal cells underneath the extensible cuticle also grow. During this cell growth, there is DNA synthesis and the epidermal cells may become polyploid (i.e., >2*n*). In the case of a soft-bodied insect such as the larva of the tobacco hornworm, *Manduca sexta,* the epicuticle is deposited in folds to allow for growth during the succeeding intermolt period during which time the epicuticle then unfolds. In this instance, the underlying endocuticle expands via a series of apical expansion points created by the deposition by the epidermal cell of vertical chitin microfilaments during the first day after ecdysis. To accommodate these two processes that must occur at the same time, an unstructured layer is deposited between the epicuticle and the lamellate procuticle.

At the onset of the molt, the epidermal cells detach from the overlying cuticle and go through a burst of RNA synthesis including that of ribosomal RNA. At this time, the cells (including those that are polyploid) may divide. Endocuticle synthesis ceases and is followed by secretion of inactive proteolytic and chitinolytic enzymes, and other proteins, to form the "molting gel" that fills the space between the old cuticle and the apical border of the epidermis. Later activation of these enzymes at the end of the molt leads to digestion of the old endocuticle. Other cellular events preparatory to the deposition of the new cuticle, such as cellular shape changes that prefigure the surface conformation, also occur. Then, cuticulin is deposited first at the tips of the plasma membrane plaques followed by deposition between the plaques to form a complete layer. Under this layer the epicuticle precursors (lipid, protein, and polyphenols) are secreted and self-assemble on the inner face of the cuticulin layer. The whole structure is then stabilized by the action of phenoloxidases that cross-link the polyphenols and the proteins. Subsequently, the apical membrane of the epidermal cells withdraws from the patterned surface and begins to form the procuticle, beginning with the exocuticle.

Digestion and Ecdysis of the Old Cuticle

During most of the formation of the new cuticle, the old cuticle remains intact and the muscles remain attached to this cuticle to allow the insect to move. At the end of the molt shortly before ecdysis, specific proteases are secreted into the molting gel to clip off a part of the inactive chitinases and proteases to render these enzymes active. These enzymes work together to digest both the protein and the chitin in the old endocuticle down to its component amino acids and *N*-acetylglucosamine sugars. This molting fluid then is resorbed into the hemolymph for its components to be recycled for production of the next cuticle or for other uses. Resorption is thought to occur in one of two ways, either back through the new cuticle and the epidermis or through the gut via swallowing and uptake in the hindgut.

Near the end of molting fluid resorption, the insect begins the process of shedding the old cuticle or ecdysis. This shedding occurs in a stereotyped sequence of behaviors. The preecdysis behavior is characterized by a series of coordinated movements that serve to loosen the muscle attachments to the old cuticle. This phase is followed by ecdysis behavior itself, which often is a series of peristaltic waves that travel from posterior to anterior and cause the animal to rupture the old cuticle anteriorly and to escape headfirst. The cuticle opens at ecdysial sutures that are areas of the old cuticle lacking exocuticle so that all but the epicuticle has been digested. In insects with rigid head capsules such as lepidopteran caterpillars, the head capsule has slipped down over the forming mandibles early in the molt to allow the formation of a larger head capsule. At the time of ecdysis, the old head capsule separates from the remainder of the old cuticle and falls off as the new larva walks out of its old cuticle.

At the time of ecdysis, a waterproofing cement layer is deposited on top of the epicuticle by the secretion of dermal glands known as Verson's glands (Fig. 1). This layer is spread over the surface by the movements of the animal under the old cuticle as it sheds its old cuticle. In some cases, a waxy layer is secreted on top of this layer in the first few days after ecdysis for further prevention of desiccation. This secretion occurs through the pore canals that traverse the cuticle from the epidermal cell to the surface of the cuticle (Fig. 1).

Postecdysial Expansion and Sclerotization

After ecdysis the animal fills its tracheae with air and also swallows air in order to expand the new larger cuticle. When it attains its final size, the new cuticle hardens and may also darken (tan) to varying degrees depending on whether the cuticle is to be flexible or rigid. In many insects there is preecdysial tanning and hardening of certain key structures such as the mandibles or the crochets on the abdominal prolegs of caterpillars used for grasping.

Sclerotization is the process of hardening the exocuticle by cross-linking the proteins together and the proteins with chitin to form a stabilized structure suitable for an exoskeleton that anchors the muscles to allow movement. The primary cross-linking agents are *N*-acetyldopamine and *N*-β-alanyldopamine. The latter is found in tan cuticles such as those of many lepidopteran pupae. Both compounds are derived from the amino acid tyrosine through a series of enzymatic steps of which the key enzymes are phenoloxidase for conversion of tyrosine to dopa and dopa decarboxylase for conversion of dopa to dopamine.

HORMONAL CONTROL OF MOLTING

Prothoracicotropic Hormone and Ecdysone

The molt is initiated by the release of prothoracicotropic hormone (PTTH), a neuropeptide, from the brain (Fig. 2). In

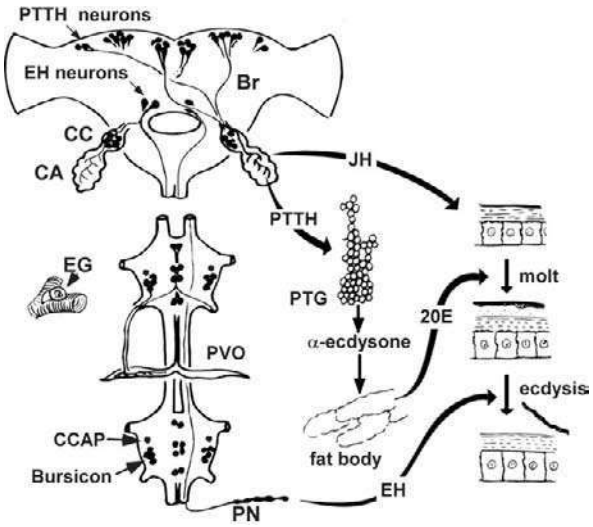

FIGURE 2 Diagram of the insect neuroendocrine system that is involved in molting. Br, brain; CA, corpora allata; CC, corpora cardiaca; CCAP, crustacean cardioactive hormone; EG, epitracheal glands; EH, eclosion hormone; JH, juvenile hormone; PN, proctodeal nerve; PTG, prothoracic gland; PTTH, prothoracicotropic hormone; PVO, perivisceral organ; 20E, 20-hydroxyecdysone.

Lepidoptera there are only two pairs of neurosecretory cells that make and release PTTH (Fig. 2). These cells send their axons out to the corpora cardiaca (or corpora allata in Lepidoptera) where they terminate and store the PTTH in vesicles until either internal or external environmental signals stimulate the cells to fire action potentials that cause release of their vesicular contents. In the blood-sucking bug *Rhodnius prolixus,* Vincent B. Wigglesworth in his classic studies of the insect endocrine system showed that the blood meal is the signal for PTTH release. Distention of the abdomen by the blood is relayed to the brain by stretch receptors. In many larvae such as the large milkweed bug, *Oncopeltus fasciatus,* and *M. sexta,* size determines the time of PTTH release although the sensory pathway utilized has not been determined. In *Manduca,* the time of day at which the larva attains critical size is also important because PTTH can be released only during a particular time period during the night called a "gate." Thus, if the larva attains critical size after the gate closes, PTTH is not released until the gate opens the following night. During the intervening time the larva continues to feed and grow and so will be larger than one that attained critical size during the open gate period and immediately released PTTH. Thus, the brain can integrate the sensory input from both the internal and the external environments and direct the time of molting.

PTTH acts on the prothoracic glands to cause the synthesis and secretion of the steroid hormone α-ecdysone into the hemolymph. α-Ecdysone (E) is converted to 20-hydroxyecdysone (20E) (Fig. 2) by the fat body, the Malpighian tubules, and sometimes other tissues. During the rise of the ecdysteroids for molting, one observes that E appears first in the hemolymph (blood) followed by 20E and later 20,26-dihydroxyecdysone (20,26E). 20E is the main component at the peak of the titer, and 20,26E is the primary ecdysteroid at the end of the molt.

Both E and 20E play a role in molting, with E being important in early cellular changes in the initiation of the molt, such as proliferation, and 20E necessary for the differentiative changes. Experiments with tissues in culture show that low levels of 20E can mimic the early predifferentiative changes caused by E, whereas high levels of E are unable to cause the differentiative events caused by 20E. Although high 20E initiates the deposition of the cuticulin and epicuticle, procuticle consisting of chitin and protein can be deposited only when the 20E levels in the hemolymph have declined. Experiments both with imaginal discs (precursors of adult structures found in larvae of moths and flies) and with *Manduca* abdominal epidermis in culture have shown that in the continuous presence of 20E, neither chitin nor the exo- and endocuticular proteins are deposited, although the epicuticle can be. To form a new cuticle *in vitro,* tissues taken during the intermolt period must first be exposed to 20E for a time commensurate with their exposure during the molt, and then they must be transferred to hormone-free medium. Although 20,26E is present for a long time during the decline of ecdysteroids at the end of the molt, so far no role has been found for it either in the timing of events that occur at the end of the molt, such as the production of new endocuticle or of dopamine, or in any aspect of ecdysis behavior. Therefore, at the present time, 20,26E is still considered to be an inactive metabolite of 20E.

Juvenile Hormone

Juvenile hormone (JH) is a sesquiterpenoid produced by the corpora allata (Fig. 2) and is present throughout nymphal and larval life in all insects. Its primary action is to prevent metamorphosis in response to ecdysone at the time of the molt as first demonstrated by V. B. Wigglesworth with *Rhodnius* and Carroll Williams in the wild silk moth *Hyalophora cecropia.* Consequently, Williams called it the "status quo" hormone.

At the beginning of or during the final nymphal or larval stage, the corpora allata cease production of JH, and JH in the hemolymph declines to undetectable levels. Then when ecdysone next rises, it causes metamorphosis. In most insects metamorphosis of the epidermis consists of switches in developmental programs from nymph to adult or from larva to pupa to adult. This switch is best understood in the epidermis of *Manduca* in which a combination of *in vivo* and *in vitro* experiments by Lynn Riddiford and her colleagues has shown that 20E acts directly on the epidermal cells to cause them to become pupally committed and that JH prevents this switch. The pupally committed epidermis no longer can form a larval cuticle, but can form a pupal cuticle only during a larval molt (as assayed by implantation into a penultimate stage larva). A similar critical ecdysone-induced switch to adult commitment is seen at the onset of the adult molt of the nymph or of the pupa.

In *Drosophila* and other higher flies, there is massive cell death of larval tissues at metamorphosis except for the nervous system and the Malpighian tubules. The imaginal discs that have only proliferated during larval life then take over and form the pupal and adult structures. The larval abdominal epidermis, however, makes the pupal cuticle while the abdominal histoblasts proliferate during the formation of the pupa. These new cells then spread over the abdomen, displacing the larval cells (which die) during the adult molt and producing the adult abdominal cuticle. JH is present during larval life in these flies and presumably suppresses precocious metamorphosis, but cannot prevent the onset of differentiation of the discs in response to ecdysone at the end of the final larval instar. By contrast, JH given at this time can prevent the metamorphosis of the abdominal histoblasts so that a fly is formed with a normal head and thorax but a pupal-like abdomen.

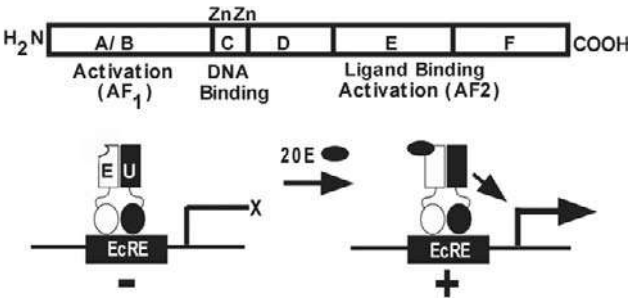

FIGURE 3 Diagram of the modular nature of the nuclear receptor molecule based on the structure of the ecdysone receptor (top) and of the heterodimeric complex of the ecdysone receptor (E) and ultraspiracle (U) on an ecdysone response element (EcRE) in the promoter of an ecdysone-activated gene in the absence (–) and the presence (+) of the active hormone 20-hydroxyecdysone (20E) (bottom). See text for details. AF1 and AF2, activation domains that interact with other proteins to cause activation of transcription when the hormone ligand is bound. (Modified, with permission, from Truman and Riddiford, 2002.)

Molecular Basis of the Action of Ecdysone and Juvenile Hormone in Molting

The molecular basis of the action of ecdysone was first studied using the giant polytene chromosomes of the salivary glands of flies, namely the midge *Chironomus tentans* and the fruit fly *Drosophila melanogaster.* During larval life the gland cells enlarge and the chromosomes replicate but the DNA strands do not separate. Consequently, by the final larval stage the chromosomes are readily visible in nuclear squashes and show bulges at specific locations known as "puffs." These puffs appear and disappear in a dynamic fashion and are sites of messenger RNA (mRNA) transcription from specific genes. When these glands are exposed to ecdysone, a few new puffs appear within 15 to 30 min ("early" puffs) followed by their regression and a second series of puffs in 3 to 4 h ("late" puffs). The early puffs appear in response to ecdysone even when protein synthesis is prevented, whereas the late puffs do not. In the early 1960s these findings led Peter Karlson to suggest that ecdysone acted directly on genes to regulate their activity, a hypothesis that has since been proven true for all steroid hormones. Based on the precise timing of the effects on puffing seen after the addition of 20E to the salivary glands, Michael Ashburner suggested in 1974 that ecdysone acted by directly activating the early genes that produced mRNA to make proteins that in turn activated the late genes and inhibited the early genes. This "Ashburner cascade" was essentially confirmed in the 1990s when the genes involved were isolated and their products identified. The following is the modern version of ecdysone action.

Ecdysone enters the cell and goes to the nucleus where it combines with the ecdysone receptor (EcR), a protein in the nuclear receptor superfamily that has the characteristic structure seen in Fig. 3 (top). The DNA binding domain (C in Fig. 3) consists of 66 amino acids and contains two cysteine–cysteine "zinc fingers" by which the zinc is held coordinately by the four cysteines. The first zinc finger is involved in binding to the DNA, whereas the second is necessary for stabilizing that binding by protein–protein interactions. The ligand-binding domain (E in Fig. 3) has a pocket that binds the ecdysone and on so doing the molecule changes its shape and may interact with coactivator proteins. The N-terminal A/B domain (Fig. 3) is also important for interactions with other proteins that are necessary to activate a gene. Normally the ecdysone receptor forms a heterodimer with ultraspiracle (USP), another member of the nuclear receptor family. This heterodimer binds to the DNA at a particular sequence called the ecdysone-response element in the gene "promoter" (a region usually but not always upstream of the gene that is necessary to turn the gene on and off) (Fig. 3, bottom). In the absence of hormone, the EcR/USP heterodimer keeps the gene suppressed; then when ecdysone appears and binds to EcR, the gene may be activated.

Most of the early genes activated by ecdysone encode transcription factors, proteins that bind to DNA in promoter regions and may activate or inactivate transcription of those genes. In the ecdysone cascade, these transcription factors activate the late genes, which are either general or tissue-specific genes that are involved in the molting process for a particular tissue. They also inactivate the early genes and the intermolt genes such as those for endocuticle synthesis in the epidermis. In addition, these early transcription factors are important in the regulation of the so-called "delayed early" genes that encode transcription factors and are activated by ecdysone but require protein synthesis for their full activation. All of these early and delayed early factors require continuous exposure to ecdysone for their continued transcription, but may be present for differing periods of time depending on the dynamics of their inhibition by the various ecdysone-induced factors. The few early genes that do not encode transcription factors instead encode proteins that are likely critical to the molting process, such as a calcium-binding protein, an ATP-binding cassette membrane

transporter, and a protein in imaginal discs important for the cell shape changes that occur at metamorphosis.

The coordination of events such as endocuticle synthesis and later dopamine production that occur during the decline of the ecdysteroid titer also is dependent on ecdysone-regulated transcription factors. These factors appear only after exposure to 20E followed by its removal. One can show that the timing of this appearance varies among the factors in a sequence similar to that seen in the animal and is presumably dependent on the levels of both 20E and the various inhibitory transcription factors that 20E has induced.

JH has been found to inhibit the appearance of a set of ecdysone-induced transcription factors named the broad complex (BR-C) in the epidermis during larval molts. In both *Manduca* and *Drosophila* the BR-C factors appear in this tissue at the time of pupal commitment and are also present during pupal cuticle synthesis. In *Drosophila* mutants that lack the *broad* gene, larvae develop to the final larval stage but cannot metamorphose because the BR-C factors are active in various tissues at this time. The BR-C factors are members of a family of chromatin-associated transcription factors. In *Drosophila* salivary glands at the onset of metamorphosis, they are associated with both the switching off of a larval-specific gene and the switching on of some of the glue protein genes. Because of the close correlation between the appearance of BR-C and the pupal commitment of *Manduca* epidermis in response to 20E in the absence of JH, the inhibition of BR-C transcription by JH may be one of the key roles of JH in preventing epidermal metamorphosis. The molecular mechanism of JH action in preventing the 20E induction of BR-C or any of the other switching actions of 20E is not yet known.

Hormonal Control of Ecdysis: ETH, EH, and CCAP

The molt culminates in the shedding of the old cuticle during ecdysis, which is followed by the expansion of the new cuticle and then its hardening and often darkening or tanning. A cascade of small peptide hormones that are released after the new cuticle is formed and the ecdysone titer has declined below a threshold level initiates ecdysis. This cascade has been studied in detail by James Truman and co-workers and by Dushan Zitnan. The precise nature of the signal that initiates this cascade is not yet understood, but at a certain time near or at the end of molting fluid resorption, the epitracheal glands just below the spiracles (the openings of the tracheae to the outside air) release a small peptide, ecdysis-triggering hormone (ETH) (Figs. 2 and 4). ETH enters the central nervous system (CNS) and initiates a sequence of behavior called the preecdysis behavior that serves to loosen the muscle attachments from the old cuticle. ETH also acts on a set of neurosecretory cells in the ventromedial region of the brain to cause the release of eclosion hormone (EH) both into the ventral nervous system and into the hemolymph from their endings in the proctodeal nerve along the hindgut (Fig. 4). In the ventral

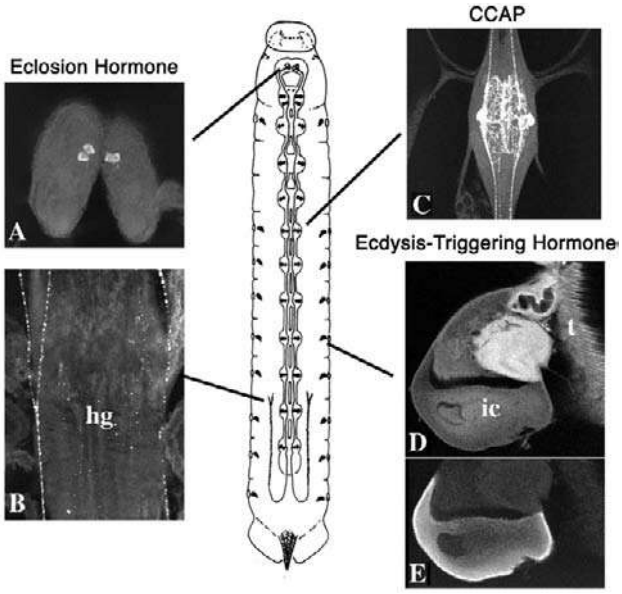

FIGURE 4 Diagram of the neuroendocrine system controlling ecdysis in the tobacco hornworm larva, *M. sexta,* with pictures of the different cells producing the peptides involved. (A) Eclosion hormone (EH) neuron cell bodies in the brain. (B) EH stores in the proctodeal nerve along the hindgut (hg). (C) Neurons in the ventral nerve cord that contain crustacean cardioactive peptide (CCAP). (D) The epitracheal gland showing the Inka cell (ic) that contains ecdysis-triggering hormone (ETH) and the autofluorescent secretory duct complex. t, trachea. (E) ETH in the Inka cell. [Modified, with permission, from Ewer *et al.* (1997). *J. Exp. Biol.* **200,** 869–881, © 1997 by Company of Biologists Ltd.]

nervous system, EH activates a network of neurons that contain crustacean cardioactive peptide (CCAP) (so named because it was first isolated from the crab in which it increases the beating of the heart) (Fig. 4). In response to EH release within the CNS, CCAP is released both into the CNS and into the hemolymph via the neurons' endings in the pervisceral organ (PVO) (Fig. 2). The action of CCAP in the CNS is to trigger the ecdysis behavior that allows the animal to shed its old cuticle. Some of these neurons also apparently contain and release bursicon, the tanning hormone (see below), into the hemolymph at the same time as they release CCAP.

In moths and flies, the sequences of the preecdysis and ecdysis behaviors are found to be stereotyped programs that once initiated by the hormone in question run their course. This behavior can even be seen in the isolated nervous system in the absence of sensory feedback, indicating a preprogrammed network that has only to be triggered by an external signal. The absence of ecdysone is necessary for this system to function since ecdysone suppresses the release of ETH from the epitracheal glands, thereby preventing this cascade of peptides. Ecdysone also decreases the excitability of the EH cells.

EH release also is controlled by photoperiod in many species so that its release, like that of PTTH, can occur only during a certain gate during the day. A good example of this control was shown by the classical studies of Truman on the

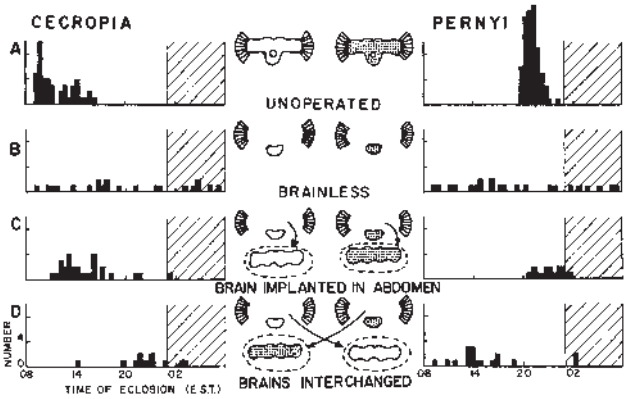

FIGURE 5 Role of the brain in causing timed eclosion of the two wild silk moths, *A. pernyi* (pernyi) and *H. cecropia* (cecropia). See text for details. [Reproduced from Truman (1972), with permission of PUDOC Press, Wageningen, The Netherlands.]

eclosion (adult ecdysis) of two species of wild silk moth, *H. cecropia* and *Antheraea pernyi* (Fig. 5). *H. cecropia* normally emerges in the morning and *A. pernyi* emerges in the late afternoon. When the brain was removed, both species emerged randomly throughout the day and night and showed very uncoordinated ecdysis behavior. When the brains were switched between the species, *H. cecropia* containing implanted *A. pernyi* brains emerged late in the day and *A. pernyi* with *H. cecropia* brains emerged in the morning in a coordinated manner. Thus, EH release from the implanted brain occurs at the time of day dictated by the donor brain, indicating that both detection of the photoperiod signals and the clock that determines the timing of EH release are located in the brain. Also, these studies indicated that EH is necessary for the coordination of ecdysis behavior. Specific destruction of the EH-releasing cells in *Drosophila* brains in the embryo and early larva resulted in about half the animals being unable to complete the larval molts, whereas the remaining flies emerged in an uncoordinated manner, again indicating that EH is necessary for behavioral coordination.

In the hemolymph EH acts back on the epitracheal glands in a positive feedback to cause further release of ETH so that within a few minutes, both the ETH and the EH cells are depleted of their stored peptides. EH also causes the filling of the new tracheae with air and the Verson's glands to release their waterproofing products over the surface of the animal as the insect is shedding its old cuticle.

Hormonal Control of Cuticular Expansion and Hardening: Bursicon, CCAP

After ecdysis the insect expands the new cuticle, then the cuticle hardens. Both CCAP and bursicon released from the PVOs into the hemolymph are involved. CCAP stimulates an increase in heart rate that is associated with the expansion of the new cuticle.

Bursicon is thought to initiate both of the above processes. For example, *Rickets* mutants in *Drosophila* lack the putative bursicon receptor ecdyse but show no wing expansion or other expansion behaviors and do not harden their cuticle. Bursicon is a large protein that is synthesized by various neurosecretory cells within the CNS (Fig. 2) but has not yet been chemically characterized. It is thought to act via cyclic AMP in the epidermal cell to trigger the cross-linking activities of *N*-acetyldopamine and *N*-β-alanyldopamine that cause sclerotization and tanning. Its plasticizing role in cuticular expansion has not been well studied.

See Also the Following Articles

Chitin • Cuticle • Development, Hormonal Control of • Ecdysteroids • Exoskeleton • Hemolymph • Juvenile Hormone • Metamorphosis

Further Reading

Bayer, C., von Kalm, L., and Fristrom, J. W. (1996). Gene regulation in imaginal disc and salivary gland development during *Drosophila* metamorphosis. *In* "Metamorphosis: Postembryonic Reprogramming of Gene Expression in Amphibian and Insect Cells" (L. I. Gilbert, J. R. Tata, and B. G. Atkinson, eds.), pp. 321–361. Academic Press, San Diego.

Ewer, J., Gammie, S. C., and Truman, J. W. (1997). Control of insect ecdysis by a positive-feedback endocrine system: Roles of eclosion hormone and ecdysis triggering hormone. *J. Exp. Biol.* **200,** 869–881.

Gammie, S., and Truman, J. W. (1999). Eclosion hormone provides a link between ecdysis-triggering hormone and crustacean cardioactive peptide in the neuroendocrine cascade that controls ecdysis behavior. *J. Exp. Biol.* **202,** 343–352.

Hadley, N. F. (1982). Cuticle ultrastructure with respect to the lipid waterproofing barrier. *J. Exp. Zool.* **222,** 239–248.

Henrich, V. C., Rybczynski, R., and Gilbert, L. I. (1999). Peptide hormones, steroid hormones, and puffs: Mechanisms and models in insect development. *Vit. Horm.* **55,** 73–125.

Hopkins, T. L., and Kramer, K. J. (1992). Insect cuticle sclerotization. *Annu. Rev. Entomol.* **37,** 273–302.

Locke, M. (1998). Epidermis. *In* "Insecta," Vol. 11A of "Microscopic Anatomy of Invertebrates," pp. 75–138. Wiley–Liss, New York.

Nijhout, H. F. (1994). "Insect Hormones." Princeton University Press, Princeton, NJ.

Reynolds, S. E., and Samuels, R. I. (1996). Physiology and biochemistry of insect moulting fluid. *Adv. Insect Physiol.* **26,** 157–232.

Riddiford, L. M. (1994). Cellular and molecular actions of juvenile hormone. I. General considerations and premetamorphic actions. *Adv. Insect Physiol.* **24,** 213–274.

Riddiford, L. M., Hiruma, K., Lan, Q., and Zhou, B. (1999). Regulation and role of nuclear hormone receptors during larval molting and metamorphosis of Lepidoptera. *Am. Zool.* **39,** 736–746.

Truman, J. W. (1972). Circadian rhythms and physiology with special reference to neuroendocrine processes in insects. *In* "Proceedings: International Symposium on Circadian Rhythmicity," pp. 111–135. PUDOC Press, Wageningen.

Truman, J. W., and Riddiford, L. M. (2002). Insect developmental hormones and their mechanism of action. *In* "Hormones, Brain and Behavior" (D. Pfaff, A. Arnold, A. Etgen, S. Fahrbach, and R. Rubin, eds.), Vol. 2, pp. 841–873. Academic Press, San Diego.

Zhou, B., and Riddiford, L. M. (2001). Hormonal regulation and patterning of the broad complex in the epidermis and wing discs of the tobacco hornworm, *Manduca sexta. Dev. Biol.* **231,** 125–137.

Zitnan, D., Kingan, T. G., Hermesman, J. L., and Adams, M. E. (1996). Identification of ecdysis triggering hormone from an epitracheal endocrine system. *Science* **271,** 88–91.

Monarchs

Lincoln P. Brower
Sweet Briar College

The monarch butterfly (*Danaus plexippus,* Nymphalidae) belongs to the tropical subfamily Danainae, the members of which are called milkweed butterflies because their larval host plants occur mainly in the milkweed family, Asclepiadaceae. With the exception of the monarch, most of the 157 known Danainae species are limited to tropical regions in Malaysia, Africa, South America, and the Greater Antilles. The adults of several species exhibit both short-distance migrations during the dry and wet seasons and social clustering behavior. This suggests that the long-distance migration and overwintering-aggregation behavior of the monarch butterfly in North America was evolutionarily elaborated from an ancient (plesiomorphic) character of the taxon.

During the late Cenozoic, the milkweed genus *Asclepias* underwent an adaptive radiation that produced 108 species in temperate North America, ranging from Mexico to the boreal forests of Canada and from the Atlantic to the Pacific coasts. As the climate alternated between hot and cold periods during the Pleistocene, the North American flora periodically advanced and retreated. The author hypothesizes that the monarch tracked the geographic expansion and retraction of its milkweed hosts and in the process refined its inherited ancestral ability to move between habitats. Because the monarch cannot tolerate temperate zone winters, natural selection would have favored those individuals that moved southward as summer waned. As time passed, the migration syndrome gradually evolved to become increasingly sophisticated, ending in the present round-trip migration, one of the most complex in the animal kingdom (Figs. 1A and 1B).

UNPALATABILITY AND MIMICRY

C. V. Riley, the most famous 19th century entomologist in North America, proposed in 1871 that the monarch was distasteful and advertised its unpalatability with its conspicuous behavior and its spotted pattern of bright orange, black, and white. Time has proven him correct and a remarkable coevolutionary interaction of the monarch with the North American *Asclepias* species resulted in great refinements of the

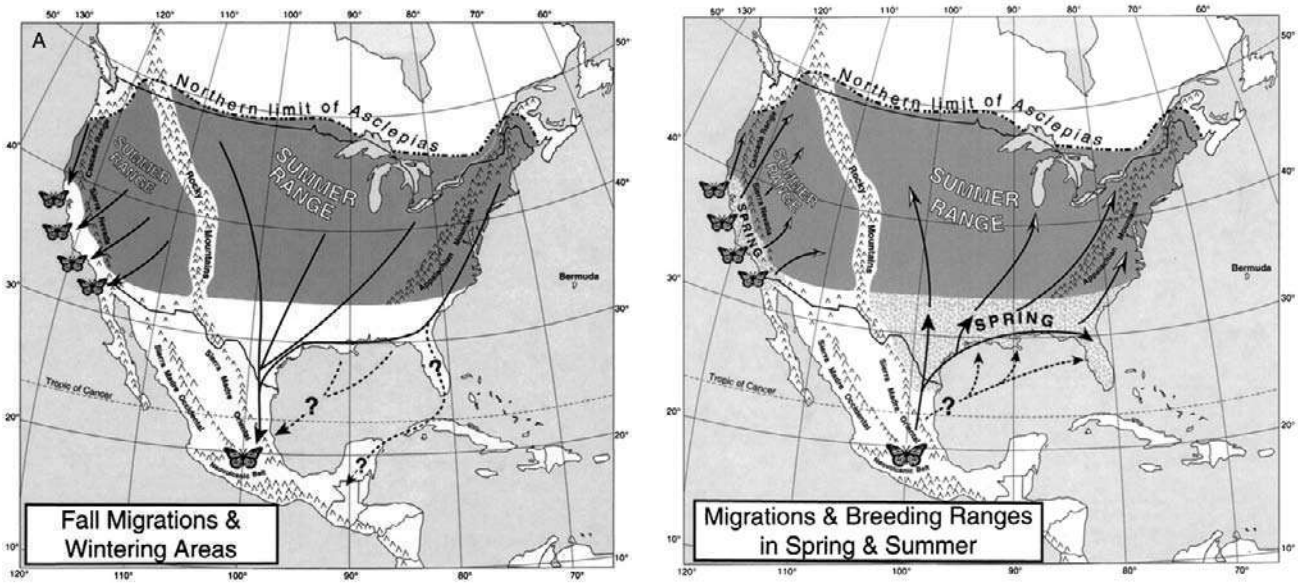

FIGURE 1 (A) Two migratory populations of the monarch butterfly occur in North America. The western population breeds west of the Rocky Mountains during the spring and summer and migrates to numerous overwintering sites, mainly along the California coast. The second, much larger eastern population breeds over several generations east of the Rocky Mountains and in the autumn migrates southward to overwintering sites in the high peaks of the Transverse Neovolcanic Belt, south of the Tropic of Cancer in central Mexico. Migration across the Gulf of Mexico and through Florida and to Guatemala remains hypothetical. (Reproduced, from Brower, 1995, with permission of the Lepidopterists' Society). (B) Spring remigrations of the monarch butterfly in North America. Western monarchs leave the coastal overwintering areas in early spring and reestablish their summer breeding range as shown. Monarchs that overwintered in Mexico remigrate at the end of March to the Gulf Coast states, where they oviposit on southern milkweeds *(Asclepias)* and produce the first new spring generation of adults by the end of April to early May. These butterflies migrate northeasterly across the midwestern states to southern Canada, laying eggs along the way and establishing a large second generation in the western and central Great Lakes region. The midwestern component of the second-generation monarchs is produced in June and they appear to continue the migration eastward over the Appalachians. One or two more summer generations (depending on temperature) follow in the Midwest and east of the Appalachians, with the last generation entering reproductive diapause and migrating southward in the autumn. Spring remigrations over the Gulf of Mexico and through Cuba and Florida remain hypothetical. (Reproduced from Brower, 1995, with permission of the Lepidopterists' Society.)

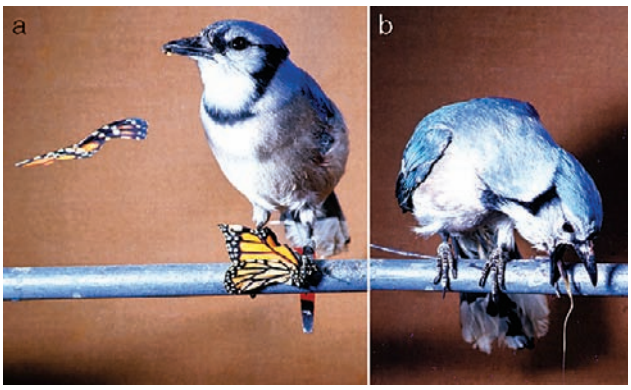

FIGURE 2 (a) A blue jay eats a monarch butterfly that contains the emetic heart poisons that its larva had sequestered from a milkweed plant. (b) About 15 min later, the jay sickens and vomits. One unpleasant experience is sufficient for most jays to avoid all further monarchs on sight. (Photographs by L. P. Brower.)

monarch's chemical defense. Milkweeds synthesize differing arrays and amounts of vertebrate heart poisons, known as cardenolides. These are bitter-tasting chemicals that cause severe vomiting when ingested. Monarch larvae are insensitive to these molecules and, as they feed on the milkweeds, they sequester and store them in their bodies. The poisons are passed on into the chrysalids and then to the adults in sufficient amounts to sicken vertebrate predators, especially birds and mice. Laboratory experiments with blue jays *(Cyanocitta cristata)* have shown that some monarchs are so toxic that once a bird has eaten one, the noxious experience is so intense that the bird not only refuses monarchs in future encounters, but may actually retch at the sight of another (Figs. 2A and 2B).

Riley also proposed that the unrelated viceroy butterfly *(Limenitis archippus)* had evolved through natural selection to mimic the color pattern of the monarch. The viceroy was originally considered palatable and a so-called Batesian mimic, but recent studies suggest that it is also unpalatable. Therefore the monarch and viceroy have most likely converged on a common warning color pattern and are Müllerian mimics. However, the situation is more complex because some milkweeds lack cardenolides. As a result, monarch larvae that feed on nontoxic milkweeds produce palatable butterflies, whereas those that feed on toxic species produce unpalatables. This discovery gave rise to the concept of automimicry, in which palatable members of a species are identical in appearance to, but are protected by their exact resemblance to, the unpalatable ones. Thus in the wild, monarchs exhibit a "palatability spectrum," with the result that mixed populations of monarchs and viceroys may simultaneously exhibit Batesian mimicry, Müllerian mimicry, and automimicry.

THE EASTERN AND WESTERN MIGRATORY POPULATIONS

Monarchs that breed west of the Rocky Mountains have been known since the mid-19th century to migrate during the fall from their breeding areas to numerous overwintering sites along the coast of California, the most famous of which is in Pacific Grove on the Monterey Peninsula. Almost certainly attracted to each other by visual and pheromone signals, the

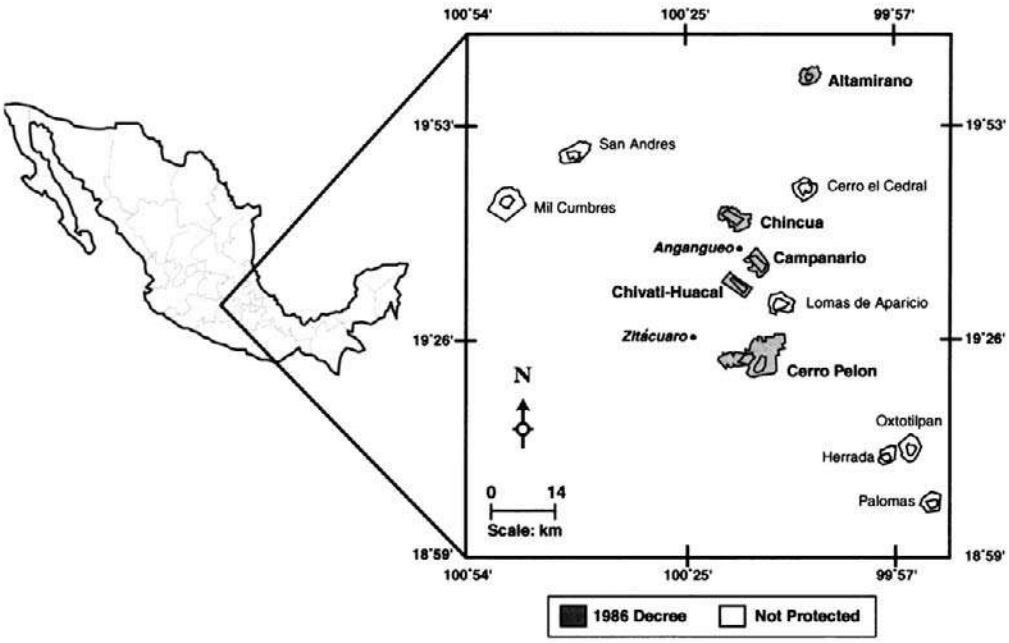

FIGURE 3 Locations of 12 mountain massifs on which monarch butterfly overwintering occurs in the oyamel–pine forest ecosystem of central Mexico, about 120 km west of Mexico City. The total area shown is approximately 11,700 km^2. The five shaded polygons are overwintering areas protected by presidential decree in 1986 (16,100 ha). The unshaded polygons are seven additional overwintering areas that currently are not protected. (From Brower *et al.,* 2002, with permission of Odyssey Press.)

butterflies aggregate from mid-October through February in spectacularly dense clusters on the live branches of Monterey pines, cypress, and eucalyptus. As spring advances, the surviving monarchs mate and migrate inland, where the females seek out the spring flush of milkweeds on which they lay singly up to 400 eggs before they die.

It was long suspected that the much larger eastern population of monarchs—those that breed during the summer in a 2.6 million-km^2 area east of the Rocky Mountains—migrates to Mexico. Fred and Norah Urquhart at the University of Toronto developed a wing-tagging program involving hundreds of collaborating amateurs that finally led one of their associates, Kenneth Brugger, to discover the first overwintering site in Mexico on January 2, 1975. Subsequent searching by Lincoln Brower, William Calvert, and their Mexican colleagues located overwintering sites on 12 separate mountain ranges within a 30 by 60-km area of central Mexico (Fig. 3). Unlike the coastal overwintering sites in California, the overwintering areas in Mexico all occur above 3200 m altitude in a coniferous oyamel fir–pine ecosystem. This is a very limited ecosystem in the Transverse Neovolcanic Belt of mountains that run across Mexico, just south of the Tropic of Cancer.

THE MEXICO OVERWINTERING PHENOMENON

The numbers of butterflies in the Mexican overwintering colonies are astoundingly large (Fig. 4). Early research, based on mark, release, and recapture studies in California, suggested that the butterfly densities in Mexico are 10 times greater and led to an estimate of 10 million monarchs per hectare (2.47 acres) of forest. However, a catastrophic winter storm in January 2002 killed so many butterflies (one sample had more than 50,000 dead butterflies in a single square meter!) that the revised density estimate is now at least 5 times this. In other words, the fir and pine trees are festooned with 50 million monarchs per hectare. So far the maximum combined area occupied by all known colonies is about 20 hectares, a total of at least 1 billion monarchs. The extraordinary beauty and mystery of the densely aggregated cluster of monarchs in these colonies is now well known. They represent one of the greatest biological wonders on this planet.

The extreme unpalatability of monarchs was probably one factor that allowed the species to elaborate the extremely dense winter aggregation behavior. Were they not chemically protected, the aggregations—an enormous potential food supply—would be exploited by vertebrate predators. In fact, field studies have shown that most, but not all species of birds and mice in the overwintering areas in both California and Mexico avoid eating monarchs. However, in Mexico, orioles and grosbeaks have broken through the chemical protection and killed an average of 15,000 monarchs per day in one colony, i.e., more than a million butterflies during the overwintering season. Again, the old adage holds that no protection, no matter how sophisticated, can ever be perfect.

FIGURE 4 Monarch butterflies festooning oyamel trees in the Sierra Chincua on a clear day, 1991. (Photograph by L. P. Brower.)

THE MULTIGENERATIONAL MIGRATION SYNDROME

Because various milkweed species fed upon by monarch larvae synthesize arrays of chemically distinct cardenolides, it is possible to extract the molecules from the butterflies and by thin-layer chromatography to obtain a chemical fingerprint that indicates which species of milkweed each monarch ate when it was a caterpillar. Using this technique, Brower and his colleagues determined that individual monarchs that have survived the winter in Mexico remigrate in the spring to the Gulf coastal states where they lay their eggs and then die. The ensuing new spring generation then continues the migration northward into the Great Lakes region and establishes the first summer generation.

By July, the first summer generation of monarchs disperses east to the Atlantic coast and west to the Rocky Mountains and produces at least one more generation of adults. By mid-August, shortening daylength and colder nights reduce juvenile hormone production in the final summer generation. This prevents gonadal maturation and the butterflies become gregarious and begin their fall migration to Mexico. As shown by Gibo, these migrants are adept at using thermal lift and tail winds. As dusk approaches, the butterflies drift down and aggregate on trees where they spend the night. When the winds blow from the south, the butterflies interrupt their migration and seek out fields of flowers and assiduously drink nectar. They convert the sugar in the nectar into lipid and store it in their abdominal fat bodies. By the time the migrants

reach central Texas, they have increased their lipid content by 500%. This reserve is crucial both to sustain them over their 5 months' overwintering period and to fuel the subsequent 1 month's spring remigration.

UNRESOLVED QUESTIONS

The individual monarchs of the last summer generation are genetically programmed to perform this migration to central Mexico in the fall, to survive the winter, and then to remigrate to the southern United States in the spring. Thus adult monarchs born in the Toronto area in August and returning to central Texas the following April traverse a distance of more than 5000 km. Their ability to find their way through deserts and mountain passes, to compensate for wind drift, and finally to locate the very small areas in Mexico that they have never before encountered remains a mystery. It will probably be solved when satellites can monitor electronic tags placed on individual butterflies.

The migratory orientation of individual monarchs shifts from south in the fall to north in the spring. How they maintain a particular course is poorly understood. While recent evidence casts doubt on the possibility that they may use magnetic orientation, there is strong evidence that sun compass orientation is involved. A complementary hypothesis is that monarch individuals have an internal clock that ticks away in all life stages and shifts the potential angle of the migration direction 1° per day throughout the year. Thus, at the spring equinox, the monarchs head out of their Mexican overwintering areas on a due north course (0°). The new spring generation, about 45 days later, heads northeastward toward the Great Lakes (45°), and the next generation that is produced about 90 days after the spring equinox heads due east (90°). By the fall equinox (September 21), their heading would be 180°, i.e., due south, changing to southwesterly as they migrate southward and finally reach the overwintering areas in November and December (Fig. 5).

Another unresolved question is the degree to which the eastern and western North American migratory populations are geographically isolated from each other. Molecular evidence suggests little differentiation, and it is possible that the western population is derived from and ultimately dependent on monarchs that get displaced westward by occasional strong northwesterly winds during the spring remigration from Mexico. Until the natural interchange between the eastern and the western populations is better understood, it seems prudent to avoid experimental and frivolous commercial transfers between them.

MIGRATION AND OVERWINTERING: ENDANGERED BIOLOGICAL PHENOMENA

The overwintering monarchs in Mexico are highly adapted to the oyamel fir–pine forest ecosystem. The forest provides a microclimatic envelope that protects the butterflies during

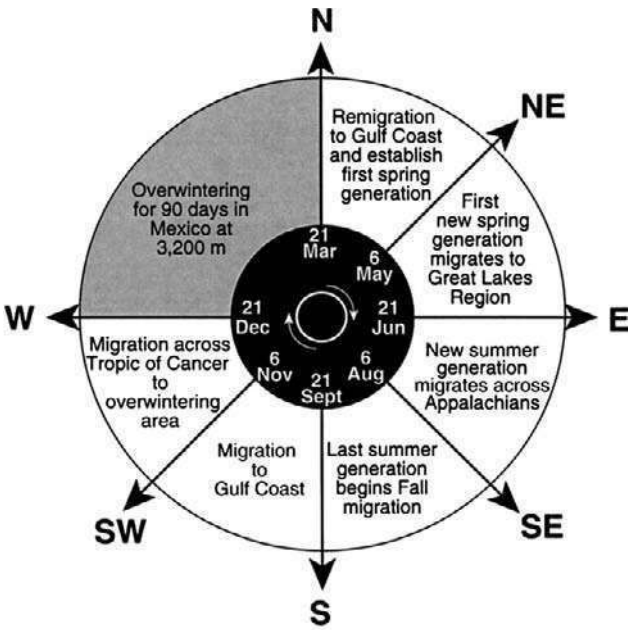

FIGURE 5 The rotational orientation hypothesis holds that all generations of monarch butterflies are migratory and that their orientation shifts clockwise at a rate of 1° per day throughout the year. The number of generations produced in each annual cycle varies from 3 to 5 and is temperature dependent. The spring equinox apparently triggers the northward remigration from Mexico. As time proceeds, the hypothesis holds that the orientation of subsequent generations shifts as shown, with some degree of scattering in each successive geographic displacement. Once the last generation has reached the overwintering sites, their migratory activity is repressed, but their orientation clock is assumed to continue running. By the time the cycle is complete at the new spring equinox, the overwintered butterflies are primed to migrate due north. (Reproduced from Brower, 1996, with permission of the Journal of Experimental Biology.)

the 5 winter months and allows them to remain in reproductive dormancy until the northern milkweed flora resurges in the spring. The high altitude and low latitude selected by the butterflies combine to provide a microclimate beneath the forest canopy that is generally wet enough to prevent the butterflies from desiccating, warm enough to prevent their freezing, and cool enough to preserve their lipid reserves. However, their overwintering can be precarious.

Anderson and Brower determined that adult monarchs can tolerate temperatures to about –8°C if they are dry, but when wetted by rain and exposed to the night sky they lose most of their natural freezing resistance. They concluded that an intact forest serves both as a blanket and as an umbrella for the overwintering monarchs and that removal of even a single large tree exposes and jeopardizes thousands of monarchs during winter storms. Unfortunately this warning was realized in January 2002 when a severe northern cold front penetrated the overwintering region. The ensuing storm soaked and then froze more than a quarter of a billion butterflies in the Chincua and Campanario colonies. The severity of this kill was exacerbated by the fact that the surrounding forests have been thinned and severely fragmented, and despite three presidential decrees

supposedly protecting the overwintering forests, extensive legal and illegal logging is accelerating throughout the region.

Monarchs are also becoming rapidly imperiled in their eastern breeding range by industrialized agriculture that is eliminating milkweeds and nectar sources. An assay based on hydrogen and carbon isotope ratios that vary systematically with respect to the geographic origin of the monarch's *Asclepias syriaca* food plants has determined that the major area of summer breeding in the United States coincides with the corn belt that has replaced the former grassland ecosystem in the midwestern United States. This has ominous implications because of the genetic engineering of crops to be resistant to herbicides. Massive and increasingly sophisticated herbicide spraying is killing the principal milkweed food plant of the monarch *(A. syriaca)* as well as most of the native flora over tens of thousands of hectares. Thus monarchs are losing both their larval food resources and their access to the diversity of flowers that provide critical nectar resources.

Because of these combined pressures on the breeding, migratory, and overwintering habitats, the migration of the monarch butterfly in North America has become an endangered biological phenomenon. The remarkable syndrome manifested by the monarch butterfly is too great a cultural and scientific treasure to allow these rampantly destructive processes to continue. Time is rapidly running out.

See Also the Following Articles

Chemical Defense • Lepidoptera • Migration • Mimicry

Further Reading

Ackery, P. R., and Vane-Wright, R. I. (1984). "Milkweed Butterflies: Their Cladistics and Biology." Cornell University Press, Ithaca, N Y.

Alonso-Mejia, A., Rendon-Salinas, E., Montesinos-Patino, E., and Brower, L. P. (1997). Use of lipid reserves by monarch butterflies overwintering in Mexico: Implications for conservation. *Ecol. Appl.* **7,** 934–947.

Brower, L. P. (1995). Understanding and misunderstanding the migration of the monarch butterfly (Nymphalidae) in North America: 1857–1995. *J. Lepidopterists' Soc.* **49,** 304–385.

Brower, L. P. (1996). Monarch butterfly orientation: Missing pieces of a magnificent puzzle. *J. Exp. Biol.* **199,** 93–103.

Brower, L. P. (1999). Biological necessities for monarch butterfly overwintering in relation to the oyamel forest ecosystem in Mexico. *In* "Paper Presentations: 1997 North American Conference on the Monarch Butterfly (Morelia, Mexico)" (J. Hoth, L. Merino, K. Oberhauser, I. Pisanty, S. Price, and T. Wilkinson, eds.), pp. 11–28. Commission for Environmental Cooperation, Montreal, Canada.

Brower, L. P., and Malcolm, S. B. (1991). Animal migrations: Endangered phenomena. *Am. Zool.* **31,** 265–276.

Brower, L. P., Fink, L. S., Brower, A. V. Z., Leong, K., Oberhauser, K., Altizer, S., Taylor, O., Vickerman, D., Calvert, W. H., Van Hook, T., Alonso-M., A., Malcolm, S. B., Owen, D. F., and Zalucki, M. P. (1995). On the dangers of interpopulational transfers of monarch butterflies. *Bioscience* **45,** 540–544.

Brower, L. P., Castilleja, G., Peralta, A., Lopez-Garcia, J., Bojorquez-Tapia, L., Diaz, S., Melgarejo, D., and Missrie, M. (2002). Quantitative changes in forest quality in a principal overwintering area of the monarch butterfly in Mexico: 1971 to 1999. *Conserv. Biol.* **16,** 346–359.

Gibo, D. L. (1986). Flight strategies of migrating monarch butterflies *(Danaus plexippus L.)* in southern Ontario. *In* "Insect Flight: Dispersal and Migration" (W. Danthanarayana, ed.), pp. 172–184. Springer-Verlag, Berlin.

Herman, W. S. (1993). Endocrinology of the monarch butterfly. *In* "Biology and Conservation of the Monarch Butterfly" (S. B. Malcolm and M. P. Zalucki, eds.), pp. 143–146. Los Angeles County Museum, Los Angeles.

Malcolm, S. B., Cockrell, B. J., and Brower, L. P. (1993). Spring recolonization of eastern North America by the monarch butterfly: Successive brood or single sweep migration? *In* "Biology and Conservation of the Monarch Butterfly" (S. B. Malcolm and M. P. Zalucki, eds.), pp. 253–267. Natural History Museum of Los Angeles County, Los Angeles.

Mouritsen, H., and Frost, B. J. (2002). Virtual migration in tethered flying monarch butterflies reveals their orientation mechanisms. *Proc. Natl. Acad. Sci. USA* **99,** 10162–10166.

Riley, C. V. (1871). Two of our common butterflies. Their natural history; with some general remarks on transformation and protective imitation as illustrated by them. *In* "Third Annual Report on the Noxious, Beneficial and Other Insects, of the State of Missouri" (C. V. Riley, ed.), pp. 142–175. Missouri State Board of Agriculture, Jefferson City.

Urquhart, F. A. (1987). "The Monarch Butterfly: International Traveler." Nelson–Hall, Chicago.

Wassenaar, L. I., and Hobson, K. A. (1998). Natal origins of migratory monarch butterflies at wintering colonies in Mexico: New isotopic evidence. *Proc. Natl. Acad. Sci. USA* **95,** 15436–15439.

Mosquitoes

Bruce F. Eldridge

University of California, Davis

Mosquitoes are small flying insects and are related to other members of the order Diptera, the "two-winged flies." The immature stages of larvae are aquatic and live in stagnant water sources in every biogeographic region of the world. Adult female mosquitoes of most species feed on blood of vertebrates, including humans, and this habit has resulted in great economic and public health significance for this group of insects.

There are well over 3000 species and subspecies of mosquitoes in the world. They occur in a variety of habitats, ranging from deserts at or below sea level to high mountain meadows at elevations of 3000 meters or more. Adult mosquitoes are terrestrial flying insects; immature stages are aquatic. Larvae and pupae of the various species can be found in ponds, ditches, puddles, swamps, marshes, water-filled rot holes of trees, rock pools, axils of plants, pools of melted snow, discarded tires, tin cans, and many other types of standing water. Some of the species are most active in the warmest part of the year, whereas others are adapted to cool temperatures. Many species of mosquitoes are rarely encountered and seldom pose a threat to the health or well-being of humans and domestic animals. However, other species are abundant, frequently encountered, and readily attack people, their pets, and their livestock. Some of these species are capable of transmitting microbial organisms that cause malaria and encephalitis and other severe diseases of humans and other vertebrates.

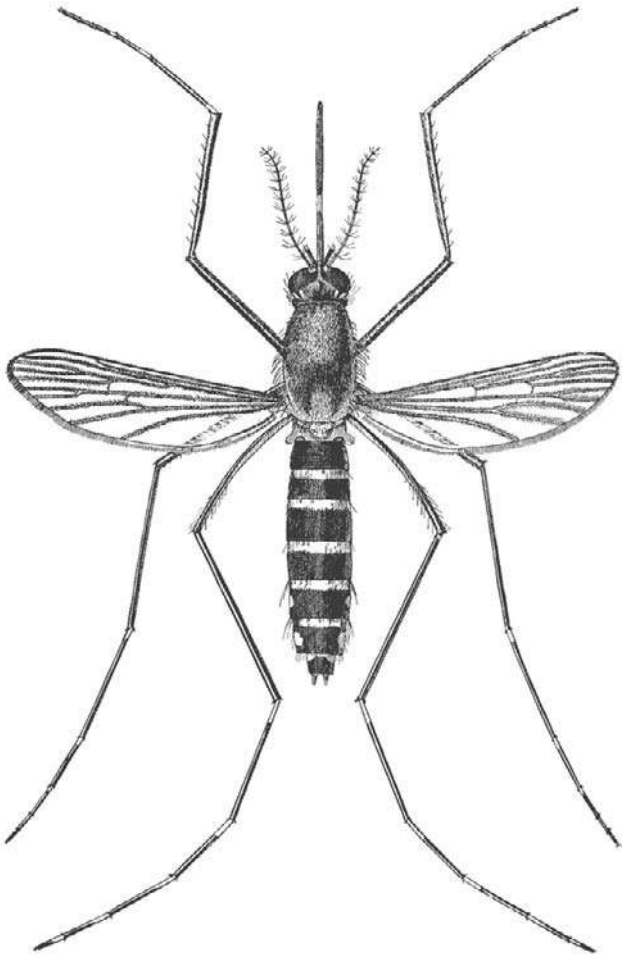

FIGURE 1 Adult female *Ochlerotatus taeniorhynchus,* a saltmarsh mosquito. (From King, Bradley, and McNeel, 1939, "The Mosquitoes of the Southeastern States," USDA Miscellaneous Publication No. 336.)

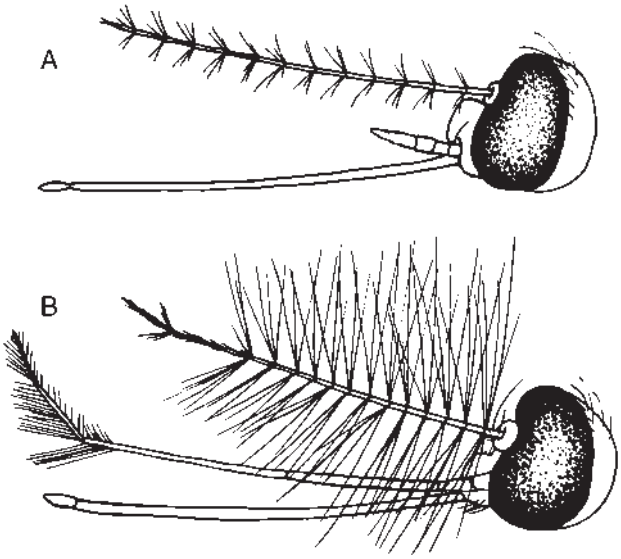

FIGURE 2 (A) Female and (B) male heads of adult *Aedes* mosquitoes. (From Gjullin and Eddy, 1972, "The Mosquitoes of the Northwestern States," USDA Technical Bulletin No. 1447.)

CLASSIFICATION AND IDENTIFICATION

Mosquitoes are classified into three subfamilies, with different characteristics in all of their life cycle stages. The species of importance from the standpoint of public health are contained in the subfamilies Anophelinae (referred to as anophelines) and Culicinae (referred to as culicines). Females of species in a third subfamily, Toxorhynchitinae, lack mouthparts adapted for sucking blood from vertebrates. The larvae of this subfamily are predaceous on other aquatic organisms and have been proposed as biological control agents of mosquito larvae.

Common genera of the Culicinae include *Culex, Aedes, Ochlerotatus* (formerly included in *Aedes*), *Psorophora, Mansonia, Haemagogus, Sabethes, Coquilletidia,* and *Culiseta.* Most species in the Anophelinae are contained in the genus *Anopheles.* The subfamily Toxorhynchitinae contains only the genus *Toxorhynchites.*

Mosquito adults are small flying midge-like insects. Most female mosquitoes can be differentiated from similar insects by the presence of a long slender proboscis, which is adapted for piercing skin and sucking blood, and long slender wings that are covered with small scales (Figs. 1 and 2A). Male mosquitoes have scale-covered wings, but their probosces are adapted for sucking plant juices and other sources of sugars (Fig. 2B). Most male mosquitoes can also be differentiated from females of the same species by their generally smaller size and by the presence of much longer and hairier maxillary palps. The immature stages of mosquitoes, the larvae (Fig. 3) and pupae (Fig. 4), vary in color from yellowish tan to black. Most mosquito larvae have a distinctive siphon, or air tube, at the rear of their bodies (Fig. 3B), but some species lack this tube (Fig. 3A).

Culicine larvae have an air tube extending from the posterior section of their body and in most species hang at rest from water surfaces at an angle of approximately 45°. Larvae of *Coquilletidia* and *Mansonia* have air tubes adapted for piercing submerged plants to obtain air for breathing. They are rarely found at water surfaces. Anopheline larvae lack an air tube and consequently rest parallel to water surfaces.

Culicine adult females have probosces developed for piercing the skin of vertebrates and sucking their blood. While feeding, their bodies are usually arranged somewhat parallel to the skin surface of their hosts. Anopheline adult females also have probosces adapted for piercing vertebrate skin, but they orient themselves at about a 45° angle while blood feeding.

The eggs of mosquitoes also vary (Fig. 5). Females of culicine species deposit single eggs *(Aedes, Psorophora),* boat-shaped rafts of 100 or more eggs *(Culex, Culiseta),* or clusters of eggs attached to floating plants *(Mansonia, Coquilletidia).* Anopheline eggs are also laid singly, but have elaborate floats extending to the sides of the eggs. Anopheline eggs are often found in clusters on water surfaces, forming interesting

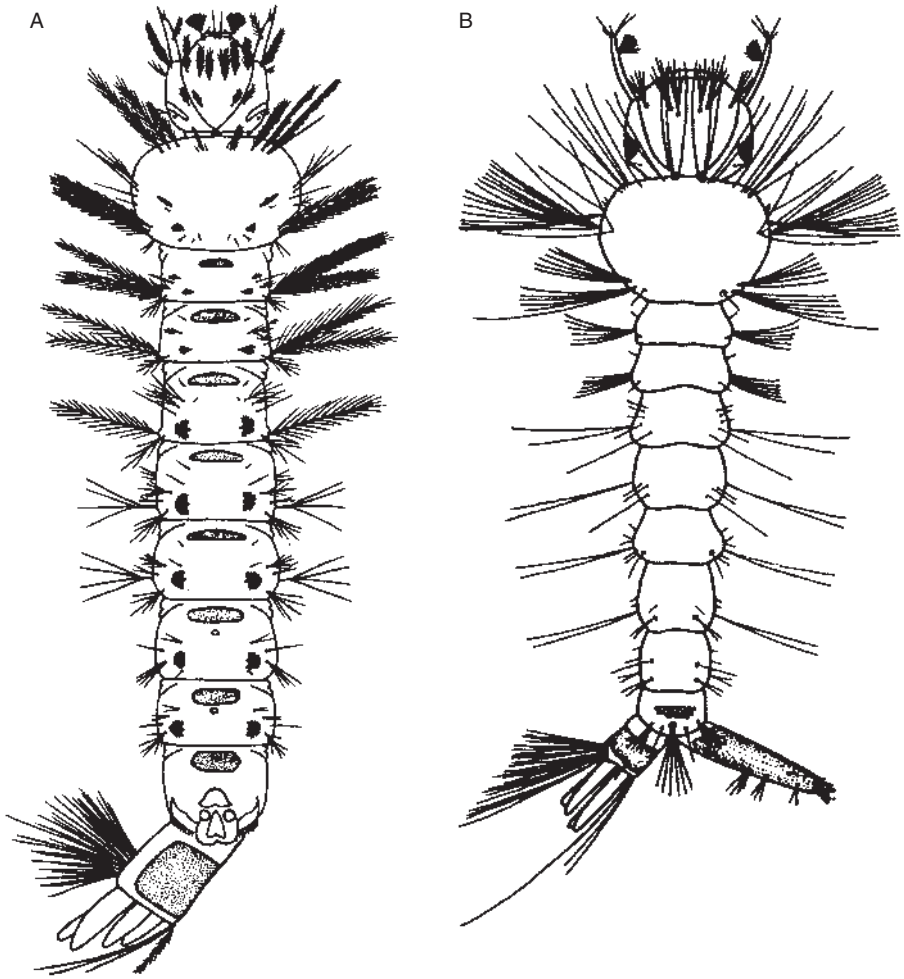

FIGURE 3 Larvae of mosquitoes. (A) An anopheline larva *(Anopheles quadrimaculatus).* (B) A culicine larva *(Culex quinquefasciatus).* (Illustrations by Harry D. Pratt, courtesy of U.S. Centers for Disease Control and Prevention.)

geometric patterns. *Toxorhynchites* eggs are also laid singly, usually on water surfaces.

LIFE CYCLE

Egg Stage

The egg-laying habits of female mosquitoes vary widely from species to species. Some female mosquitoes lay eggs on water surfaces (e.g., *Anopheles*), others lay single eggs on moist soil where later flooding is likely (e.g., *Aedes*). From eggs deposited on water surfaces, larvae usually hatch within a day or so, but from eggs laid on soil surfaces, larvae do not hatch until eggs are flooded, which may occur months, or even years, later. The environmental cues female mosquitoes use to find suitable sites for oviposition remain only partially known. Color, moisture, and volatile chemical stimulants appear to play a role in certain species. Efforts to explain the occurrence of various mosquito species in different aquatic habitats based strictly on oviposition cues have been unsuccessful.

Larval Stages

Small larvae that are nearly invisible to the naked eye hatch from eggs. Larvae molt three times to become fourth-stage larvae. Several days later, this larval form molts again to become a pupa. The time required for development of the larval stages depends on several factors, the most important of which is water temperature. Availability of food and larval density are also factors. Water temperature and food are inversely related to time of development; larval density is directly related.

The majority of mosquito species have larvae that are restricted to fresh water. However, larvae of a few species can develop under other conditions, e.g., brackish or salt water or water polluted with organic solids. Species with larvae adapted to salt water can maintain osmotic pressure within their bodies by drinking substantial amounts of water and by removing ions from their hemolymph through their Malpighian tubules and rectum. Generally, saline species can also develop in fresh water, but do not compete well with freshwater species. The

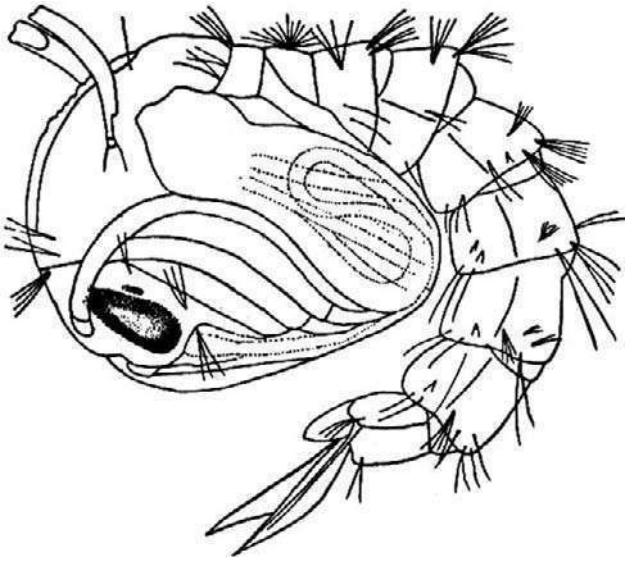

FIGURE 4 A mosquito pupa. (Original illustration from Howard, Dyar, and Knab, 1912, "The Mosquitoes of North and Central America and the West Indies." Reprinted from USDA Handbook No. 336, 1939.)

inverse is not true and consequently, various kinds of water usually have a characteristic mosquito fauna.

Adults

Adult mosquitoes emerge 1–2 days after the appearance of pupae, with males emerging first. In the summer, the entire life cycle, from egg to adult, may be completed in 10 days or less. Females feed on vertebrate blood for the development of eggs. This behavior by females is the single most important characteristic of mosquitoes from the human standpoint. Blood feeding in insects is believed to have evolved several times independently from ancestral forms adapted for sucking plant juices or for preying on other insects. The means by which female mosquitoes locate suitable hosts for blood feeding has been studied for many years, but there are still many unknown features to this behavior. The best explanation is that females are attracted by warmth, moisture, and carbon dioxide from hosts, but other factors are involved. There have been studies that have suggested that substances such as lactic acid, a component of human sweat, may serve as an attractant.

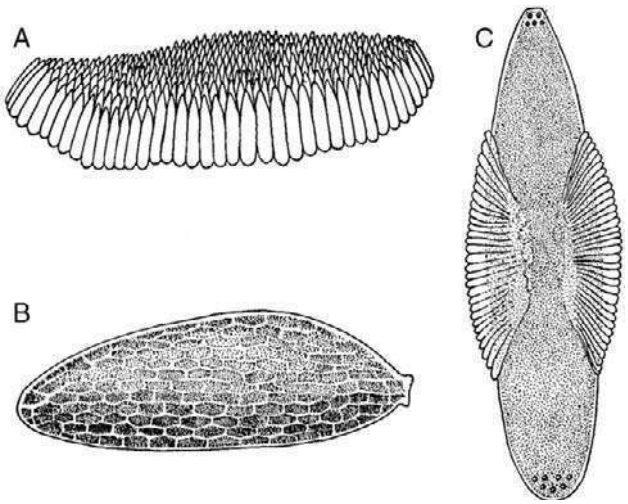

FIGURE 5 Eggs of mosquitoes. (A) Egg raft of *Culex* mosquito. (B) Single egg of *Aedes* mosquito. (C) Single egg of anopheline mosquito. (Original illustration from Howard, Dyar, and Knab, 1912, "The Mosquitoes of North and Central America and the West Indies." Reprinted from USDA Handbook No. 336, 1939.)

Ordinarily, a female mosquito cannot develop a batch of eggs unless she has taken a blood meal to obtain nourishment for ovarian development. However, some strains or individuals of several species can develop eggs without a blood meal, which is called autogeny. The nourishment for egg development is carried over from the larval stages, and consequently, only the first batch of eggs can develop in this way. The usual situation, in which a blood meal is required for the development of all batches of eggs in an individual female, is called anautogeny.

Some mosquitoes take blood only from certain groups of vertebrate animals. For example, *Culex pipiens,* the northern house mosquito, feeds almost entirely on birds. *Ochlerotatus sierrensis,* the western treehole mosquito, feeds only on mammals. *Culex tarsalis,* the encephalitis mosquito, feeds on birds and mammals (this dual host preference is one characteristic of an effective vector of disease pathogens). In the past, this relative host specificity has been called host preference. However, this term is not appropriate, because it ignores availability of hosts, host defensive behavior, and other factors unrelated to the mosquitoes themselves. The blood-feeding drive is controlled by neurohormones and can be induced artificially by treatment with juvenile hormone or one its analogs. This hormonal influence is why mosquitoes that have recently had a blood meal and are developing a batch of eggs do not usually seek another blood meal. However, multiple blood meals (more than one blood meal in a single gonotropic cycle) do occur at times in nature in some species.

Blood feeding by mosquitoes is a complex process. It is facilitated by the infection of saliva into the feeding wound of the vertebrate host. Saliva comes from organs in the thorax of mosquitoes called salivary glands. Saliva may contain a variety of substances, including chemicals that reduce clotting of vertebrate blood. Digestion of a blood meal usually takes 2–3 days, depending on the ambient temperature. The uptake of blood is accomplished by the action of muscular pumps in the head of female mosquitoes. Blood travels through the digestive tract of the mosquito into a structure called the midgut. After blood reaches the midgut it is soon surrounded by a thin sheath, the peritrophic membrane, that is secreted by cells at the front of the midgut. Digestion of the blood takes place within this structure.

Seasonal Development

Some species of mosquito have but a single generation per year (univoltine), whereas others have many (multivoltine), depending upon the length of the season favoring the activity of the adult stages. To avoid seasons of the year not favorable to adult activity (usually the winter), mosquitoes may have some kind of diapause mechanism. In *Aedes* and related genera, the diapause mechanism usually involves the egg stage. In temperate and subarctic zone *Aedes,* populations may survive winters as desiccation-resistant eggs, sometimes under the surface of snow or along river flood plains. The larvae then hatch in the spring after the eggs are flooded from melted snow or after flooding of the riverbanks.

Culex and *Anopheles* females usually survive unfavorable periods as diapausing or quiescent adult females. Male mosquitoes usually do not survive unfavorable periods, so it is necessary for insemination to occur before the onset of diapause.

Some mosquito species survive unfavorable periods as diapausing larvae (e.g., species of *Aedes, Anopheles, Culiseta*). Diapause can be variable in some species, depending upon the latitude at which they occur, with diapause occurring in the larval stage at warmer latitudes and in the egg stage at cooler ones.

There is considerable variation in the environmental and physiological control of diapause. In nearly all diapausing mosquitoes studied, diapause is triggered by exposure of one or more of the life cycle stages to daylength. In *Culex* species, and other mosquitoes that overwinter as adults, exposure of late-stage larvae and of pupae to short daily photophases occurring in autumn results in diapause in adult females. This diapause is manifested by lowered activity levels, inhibition of blood-feeding drive, and arrestment of follicle development in ovaries. In some *Aedes* species, the short autumn days experienced by females result in deposition of eggs that are in the diapause state. The larvae in these eggs do not hatch until a period of exposure to near-freezing temperatures lasting several months. In other species of *Aedes,* diapause results from exposure of the eggs themselves to short daylengths. Still other *Aedes* species have larvae that enter diapause triggered by their exposure to short daylengths.

As with other aspects of reproduction and development, diapause is controlled directly by neurohormones. Diapause can be induced in most diapausing species by exposure to juvenile hormone or one of its analogs.

Many tropical and subtropical species, such as *Aedes aegypti,* the yellow-fever mosquito, do not have a diapause mechanism. Still other tropical species have mechanisms for avoidance of hot, dry seasons, but these mechanisms have been little studied.

PUBLIC HEALTH AND VETERINARY IMPORTANCE

As discussed earlier, female mosquitoes of nearly all species require blood from vertebrate animals to develop their eggs, and many species bite people, their pets, and livestock for this purpose. The most important result of this behavior is the transmission of microorganisms that cause diseases such as malaria, filariasis, yellow fever, and dengue. These and other mosquitoborne diseases can have serious and sometimes fatal consequences in people. These diseases can also have an impact on livestock, pets, and wildlife. Even when no infectious disease pathogens are transmitted by mosquitoes, they can be a serious health problem to people and livestock. Biting of people by mosquitoes can result in secondary infections, allergic reactions, pain, irritation, redness, and itching. Mosquito biting of beef cattle can result in reductions in weight gains and in dairy cows, reduction in milk production.

The interactions between mosquito hosts and the pathogens they transmit are highly variable. Three basic types of transmission mechanisms are involved: (1) propagative transmission, in which the pathogen multiplies within the mosquito but does not undergo any changes in developmental form; (2) developmental transmission, in which the pathogen undergoes developmental changes, but does not multiply; and (3) propagative-developmental transmission (also called cyclodevelopmental transmission), in which the pathogen multiplies and undergoes changes in developmental forms. Transmission of the yellow-fever virus by the yellow-fever mosquito is an example of propagative transmission. The virus is taken up by a female mosquito from a viremic host during blood feeding, multiplies many times, and eventually infects the salivary glands of the host. When the female mosquito takes another blood meal, she may infect a new host by injection of saliva.

Some pathogens are transmitted to the offspring of female mosquitoes via infected eggs. This type of transmission is known as transovarial transmission.

Filarial worms, the cause of the disease filariasis (a type of which is called elephantiasis) in humans and other vertebrates, are transmitted by developmental transmission. In this example, very small immature forms of the worms, called microfilariae, occur in the blood of infected vertebrate hosts and are taken up by female mosquitoes in a blood meal. Within the mosquito, the filariae molt several times until they eventually become infectious larvae. These larvae migrate down the proboscis of the mosquito and enter the feeding wound caused by the mosquito during a subsequent blood feeding. Within the vertebrate host, these larvae may eventually develop into adult male and female worms that mate and produce microfilariae. It is the presence of large numbers of adult worms that results in the symptoms of filariasis.

Malarial parasites have a very complex life cycle, involving both multiplication of parasites and development of life cycle stages. Anopheline mosquitoes are the vectors of human malaria, and because the sexual stages and fertilization occur within mosquitoes, they are the definitive hosts. Parasite forms called microgametocytes (male sex cells) and macrogametocytes (female sex cells) occur in the peripheral blood of humans and are taken up by mosquitoes. Fertilization of the female cells by the male cells occurs within the gut of the vector

mosquito. After several life cycle changes, and multiplication of forms within cysts on the gut wall, forms of the parasite called sporozoites enter salivary glands of the mosquito and infect new hosts during blood feeding.

There are hundreds of types of microorganisms that are transmitted by mosquitoes to vertebrates that result in diseases. A few are extremely important worldwide because of their high incidence and the severity of their symptoms in humans.

Malaria is one of the most important diseases in the world. Several hundred million people are infected with malarial parasites, resulting in over 2 million fatalities annually, mostly in tropical countries in Africa and Asia. Malaria is especially serious in pregnant women and young children. Typically, more than a million children die each year from this disease. The economic development of a number of tropical countries is badly hindered by malaria because of the burden of chronic malaria infections in working-age men and women.

The virus disease known as dengue, transmitted mostly by the yellow-fever mosquito, is a rapidly expanding problem in the world and now is considered second in importance only to malaria among mosquitoborne diseases. The increase in global human travel resulting from expanded rapid air transportation has been paralleled by the increase in the number of viral strains causing dengue and the increase in the number of cases of a particularly serious form of the disease called dengue hemorrhagic fever. This form of the disease is most serious in children and is a significant cause of mortality.

Filariasis is a general term applied to infection of vertebrate animals by many different species of parasitic worms belonging to the superfamily Filaroidea. A form of mosquitoborne filariasis is called lymphatic filariasis because infection can cause impairment of the lymphatic system. Lymphatic filariasis is a chronic disease that can lead to the well-known disfigurement of humans called elephantiasis. Another type of filariasis called dog heartworm occurs in dogs, other canids (e.g., wolves and coyotes), and felids (e.g., cats). Heavy infections can result in large buildups of adult worms in the cardiopulmonary system and can be fatal.

Yellow fever, a virus disease, has virtually disappeared from the United States because of the availability of an extremely effective vaccine. This vaccine may provide lifelong immunity from a single inoculation. Unfortunately, the availability of the vaccine is limited on a worldwide basis and there are many unvaccinated people living in areas where the mosquito vector, *A. aegypti,* is common. Yellow fever is an extremely serious disease. There is no available treatment, and infections in humans are frequently fatal. Periodic epidemics continue to occur in various tropical countries. *A. aegypti* is common in urban and suburban areas of the tropics and subtropics. The larvae of this species occur in water in various types of artificial containers such as shallow wells, water urns, discarded containers, and tires. It is very difficult to control.

There are many other mosquitoborne diseases, several of them caused by viruses. Some of these viral diseases, such as Japanese encephalitis, La Crosse encephalitis, West Nile fever, Ross River disease, and Rift Valley fever, affect large numbers of people in parts of the world where they occur.

CONTROL OF MOSQUITOES

In most industrialized nations, mosquito control is done by government-supported agencies that are either components of health agencies or separate agencies organized specifically for that purpose. In the United States, states that have the most serious mosquito problems (e.g., New Jersey, Florida, Texas, Louisiana, California) have many such agencies. Some are small and have responsibility for mosquito abatement in a few hundred square kilometers, whereas the activities of others may encompass one or more entire counties. However, even in states that have many mosquito abatement districts, many people live in areas with no organized mosquito control. In underdeveloped areas of the world, organized mosquito control is rare except for scattered programs aimed at specific diseases such as malaria.

Most organized mosquito control is accomplished by searching out mosquito larvae in standing water and then treating the water with some kind of material that kills the larvae. Modern materials are highly specific for mosquitoes and ordinarily have little or no effect on other organisms. One such material is a bacterial product called Bti *(Bacillus thuringiensis israelensis)* that produces a toxin that kills only larvae of mosquitoes, black flies, and certain midges. Mosquito abatement agencies may also apply chemical pesticides to kill adult mosquitoes, but ordinarily only when adult populations become so high that they cause extreme annoyance to many people or when the threat of disease transmission to people is high. Therefore, the most common method for this is known as ultralow volume, or ULV. This approach involves using special equipment to spray extremely small volumes of small particles of highly concentrated insecticides. When used properly, it is a safe and highly specific method of mosquito control.

Control of irrigation water in agricultural areas to avoid excess runoff is an important mosquito control method, but in recent years elimination of small bodies of water that can serve as wildlife habitat has ceased to be a mosquito control option.

In the first half of the past century, elimination of bodies of temporary and permanent water (swamps, marshes, vernal pools) was an accepted form of mosquito control. Recent years have seen the realization that such habitats are valuable and irreplaceable components of the environment and that a variety of activities have resulted in the permanent loss of many of these wetland habitats. This loss has resulted in the development of mosquito management strategies that are much more ecologically sound. Considerable research has been conducted on management strategies that enhance wetland habitats while minimizing problems from mosquito breeding.

Biological Control

A method that is a preferred alternative to chemical control is the use of live organisms to control mosquitoes, either by predation or by infection. Mosquitofish (the common guppy) have been used for many years for this purpose, often with effective results. However, because mosquitofish are generalist predators, they must be used with great care to avoid damage to other aquatic organisms. Many other forms of biological control for mosquitoes have been tried, including other types of fishes, fungi, bacteria, nematode worms, flat worms, protozoan parasites, and predaceous insects (including some mosquitoes). Some of these organisms have been effective under special circumstances, but few of them have been adopted widely. Microbial organisms such as Bti and *Bacillus sphaericus* may be considered biological control agents, and these are used to great advantage in a variety of aquatic habitats.

Insecticides

At one time there were dozens of insecticides available for killing both adults and immature stages of mosquitoes. However, because of economics, primarily the costs involved in developing, testing, and registering new materials, and the development of resistance to insecticides by mosquitoes, the number of available materials is now down to a handful. A class of insecticides known as insect growth regulators has been highly effective and specific for mosquitoes, but the development of resistance to even these materials has clouded the future of these so-called third-generation pesticides. The best hope for circumventing resistance to pesticides is the use of a combination of approaches referred to as pesticide resistance management. Frequent testing for susceptibility in mosquito populations, alternation of pesticides, and avoidance of methods that result in the persistence of low dosages of pesticides are examples of this approach. Insecticide resistance is under genetic control, and the goal of insecticide resistance management is preservation of genes in mosquitoes associated with susceptibility.

Protection from Mosquito Bites

People living in areas lacking organized mosquito control must protect themselves from bites of mosquitoes by using a variety of strategies. Probably the most effective method of personal protection from mosquito bites is to avoid places where mosquito densities are high and to avoid being out-of-doors at times of the day when mosquito activity is at its highest. In mountainous areas, most species of mosquitoes bite during the morning and afternoon and often not at all during periods of darkness. In low-elevation areas some mosquitoes tend to bite at night, whereas others bite during the day. The species of mosquito present in a given area varies from place to place, and it is necessary to learn the activity patterns of the mosquitoes present to develop avoidance strategies.

If exposure to biting mosquitoes cannot be avoided, there are several ways to minimize discomfort. The most important of these is to reduce exposed skin surfaces by wearing hats, long trousers, and long-sleeved shirts. Some mosquitoes bite through light clothing, but the number of bites received is definitely reduced if most areas are covered. When mosquito densities become very high, application of a mosquito repellent may be needed to avoid bites. Currently, the only really effective repellents on the market are those that contain the material DEET. Skin repellents have some drawbacks. After application, they are effective for only about 4 h at the maximum. Other factors, such as wind, high temperature, high humidity, and sweating, reduce this time even further. When applying DEET, the material must be thoroughly applied to all exposed skin, including behind the ears. In recent years, longer lasting formulations of DEET have been developed by the incorporation of various additives such as lotions and polymers.

Many people have tried gadgets such as ultrasonic emitters, electric grids, aromatic plants, and even vitamins for mosquito protection. Research has shown that most such methods are of little or no value in repelling mosquitoes, but such devices continue to appear on the market. In some areas of the world incense coils are sold for avoidance of mosquitoes. They may afford protection within a short distance of the burning coils.

Bednets can provide excellent protection from mosquito bites at night if used properly. The use of bednets treated with insecticides has been shown to afford excellent protection from attack by malaria mosquitoes. When they are available, vaccines may protect humans from mosquito-borne disease (e.g., yellow fever) and prophylactic drugs may be used to avoid some diseases (e.g., malaria).

See Also the Following Articles

Aquatic Habitats • Blood Sucking • Dengue • Dog Heartworm • Malaria • Medical Entomology • Salivary Glands • Yellow Fever

Further Reading

Bates, M. (1954). "The Natural History of Mosquitoes." Macmillan Co., New York.

Clements, A. N. (1963). "The Physiology of Mosquitoes." Macmillan Co., New York.

Clements, A. N. (1992). "The Biology of Mosquitoes," Vol. 1. Chapman & Hall, New York.

Clements, A. N. (1992). "The Biology of Mosquitoes," Vol. 2. CAB Int., Oxon, U.K.

Darsie, R. F., Jr., and Ward, R. A. (1981). "Identification and Geographical Distribution of the Mosquitoes of North America, North of Mexico." American Mosquito Control Association, Fresno, CA.

Eldridge, B. F., and Edman, J. D. (eds.) (2000). "Medical Entomology." Kluwer Academic, Dordrecht.

Knight, K. L., and Stone, A. (1977). "A Catalog of the Mosquitoes of the World." Thomas Say Foundation, Entomol. Soc. Am., College Park, MD.

Service, M. W. (1993). "Mosquito Ecology: Field Sampling Methods," 2nd ed. Elsevier, London/New York.

Moth

see *Lepidoptera*

Mouthparts

R. F. Chapman
University of Arizona

The mouthparts of insects are structures surrounding the mouth that are involved in the mechanics of feeding and processing and manipulating the food so that it can be ingested. Although functionally equivalent to the jaws of vertebrates, they lie outside the mouth, not within a buccal cavity. Good basic accounts of insect mouthpart structure are to be found in most textbooks of entomology. The aim of this article is to supplement these basic accounts by briefly considering some of the variation associated with different feeding habits and different types of food. It also gives some information on the functioning of the mouthparts.

BITING AND CHEWING INSECTS

Insect mouthparts are derived from the appendages of four of the segments forming the insect head. They surround the mouth and are external to it, unlike the condition in vertebrates in which the teeth are within the oral cavity. The basic segmental character of the mouthparts is most apparent in insects that bite off fragments of food and then chew it before ingesting it (Fig. 1). Insects that do this are said to be "mandibulate" because the mandibles are relatively unmodified compared with those of fluid-feeding insects (see below). These are also commonly called biting mouthparts, although there is some risk of confusion with blood-sucking insects, such as mosquitoes, which bite! In this article the latter are distinguished as "piercing." The mandibulate arrangement occurs in the primitively wingless insects (Apterygota), in the cockroaches and grasshoppers and their allies, in larval and adult beetles (Coleoptera) and most Hymenoptera, and in caterpillars (larval Lepidoptera), among the more advanced groups of insects.

Immediately in front of the mouth is the labrum formed from the fusion of the appendages on either side of the labral segment. It comprises a flat sclerotized plate of cuticle continuous with the cuticle of the front of the head (clypeus). Its inner side (toward the mouth) is known as the epipharynx, and it is formed from membranous cuticle-bearing tracts of noninnervated hairs, all pointing toward the mouth. In grasshoppers, and probably in other insects with similar mouthparts, the hairs are easily wetted, whereas the other parts of the cuticle are water repellent. The hydrophilic hairs may serve to direct the flow of fluid from the food toward the mouth and also to groups of contact chemoreceptors (taste receptors) that occur just outside the mouth. Contact chemoreceptors also often occur along the distal edge of the labrum. At rest, the labrum presses back on the mandibles, which are immediately behind it, being held in this position by a rubber-like protein, called resilin, in its connection with the clypeus.

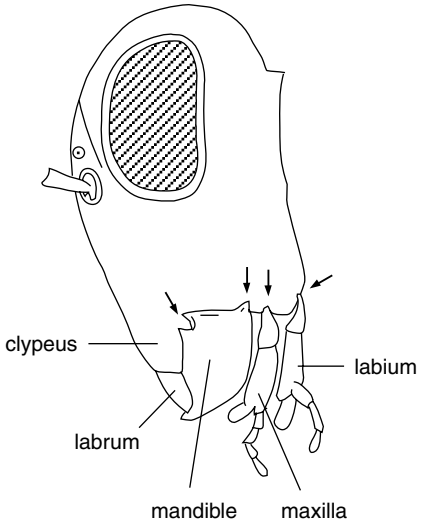

FIGURE 1 A lateral view of the head of a grasshopper showing the segmental arrangement of the mouthparts: labrum, mandible, maxilla, and labium. Arrows show the points of articulation (condyles) with the head capsule. The mandible has two condyles (dicondylic), the maxilla only one, and the labium one on each side (modified after Snodgrass, 1935, "Principles of Insect Morphology," McGraw-Hill).

The mandibles, one on each side, are hinged to the head capsule by one or two condyles. Archaeognatha have only one condyle (monocondylic), whereas Thysanura and all mandibulate pterygote insects have two (dicondylic). The change from one to two condyles represents a considerable evolutionary advance because it gives the mandibles a much firmer base and so facilitates feeding on hard materials. The mandibles of the two sides swing transversely to meet below or in front of the mouth, depending on the orientation of the head, and are opened and closed by a pair of muscles, one inserted on either side of the axis of mandibular attachment at the condyles. The opener muscle is called the abductor, whereas the closer is the adductor. The latter is the larger of the two because it provides the force necessary to bite into or through material. Both muscles arise on the cuticle at the top of the head and, in grasshoppers and caterpillars, the head capsule grows bigger to accommodate the increased size of the adductor muscle if the insect feeds on tough food.

The two mandibles are asymmetrical so that where they meet in the midline the cusps on the biting surface of the two sides fit between each other (Fig. 2). These cusps are extremely hard. In addition to being sclerotized like the hard cuticle elsewhere in the body, their cuticle contains zinc or manganese or, occasionally, iron, which are assumed to contribute to the hardness. The form of the biting cusps varies from species to

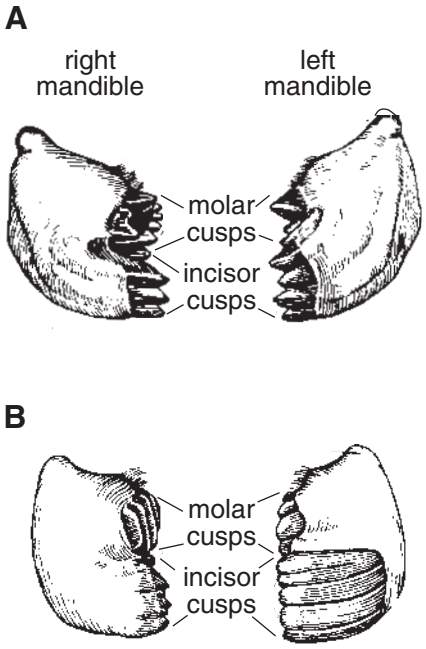

FIGURE 2 Mandibulate mouthparts. Mandibles, seen from in front with the labrum removed, of grasshoppers with different feeding habits. Notice that the mandibles of the two sides are asymmetrical. (A) A grasshopper that feeds on soft, broad-leaved plants. (B) A grasshopper that feeds on grasses. (Reproduced, with permission, from F. B. Iseley, 1944, Correlation between mandibular morphology and food specificity in grasshoppers. *Ann. Entomol. Soc. Am.* **37,** 47–67.)

species in relation to feeding habits in a way that, superficially, is comparable with the adaptations in the jaws of mammals. Predaceous tettigoniids, for example, have a sharply pointed cusp distally and powerful blade-like cusps more proximally that have some resemblance to the canine and carnassial teeth of carnivorous mammals and presumably serve similar functions of grasping and tearing the prey. Among grasshoppers, species feeding on soft, broad-leaved plants have small, sharply pointed cusps that cut the food into very small fragments. Grass feeders, on the other hand, have very long, chisel-edged incisor cusps distally with short, flattened molar cusps proximally, which can superficially be compared with the chisel-shaped incisors and grinding molars of mammalian herbivores. Other insects also exhibit food-related modifications of the mandibles. The cusps become worn down with use, especially if the insect is feeding on hard foods, and there is some evidence that this wearing down reduces the rate of food intake. The cusps can be renewed only at a molt, when new cuticle is formed.

In insects eating food that requires special treatment during ingestion, the mandibles may become highly modified. An example occurs in the larvae of pergine sawflies (Fig. 3). These Australian insects feed on *Eucalyptus* and related trees, the leaves of which are rich in essential oils. The insects sequester the oils in a diverticulum of the foregut and use them for defense. The mandibles are apparently adapted for separating the oils from the leaf tissue. Sticking out from the center of the mandible is a structure called the scopa mandibularis. It is covered by

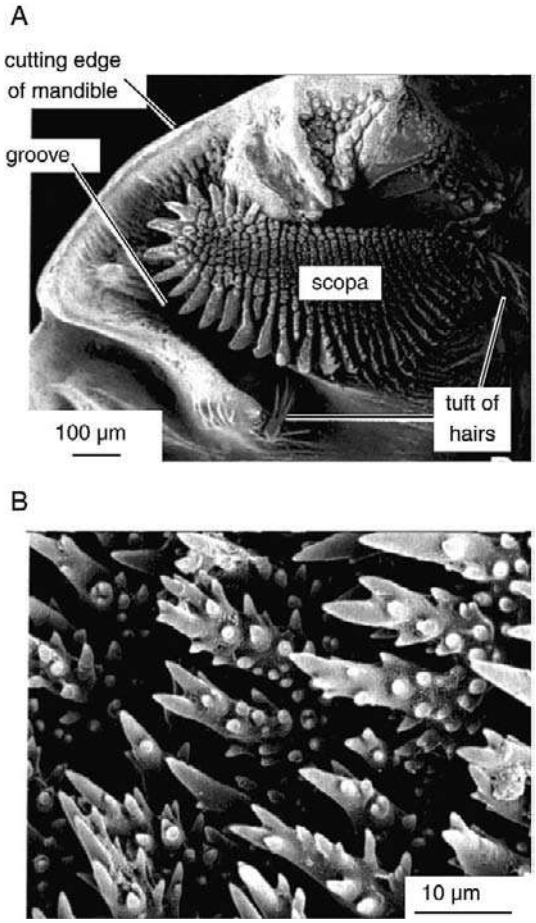

FIGURE 3 Mandibulate mouthparts. Specialized mandible of a pergid sawfly larva. These insects feed on *Eucalyptus* leaves containing large quantities of essential oils. The oils are apparently squeezed from the leaf tissue by the action of the scopa mandibularis (scopa) and conducted along the groove to a pharyngeal diverticulum where they are stored. (A) Surface view of the biting face of the right mandible showing the rows of spines forming the scopa mandibularis. (B) Spines of the scopa mandibularis, which presumably abrade the leaf surface to release the oils. (Reproduced, with permission, from S. Schmidt, G. H. Walter and C. J. Moore, 2000, Host plant adaptations in myrtaceous-feeding Pergid sawflies: Essential oils and the morphology and behaviour of *Pergagrapta* larvae (Hymenoptera, Symphyta, Pergidae). *Biol. J. Linn. Soc.* **70,** 15–26.)

rows of pointed setae and these, perhaps by scraping and shredding the leaves, seem to be involved in extracting the oils.

In the midline, immediately behind the mouth and probably also derived from the mandibular segment, is the hypopharynx. This structure is a lobe of mostly membranous cuticle but with rods of sclerotized cuticle to which muscles are attached. Like the epipharynx, it bears tracts of hairs pointing toward the mouth and these hairs probably help to move food toward the mouth as the hypopharynx is moved by its muscles.

Behind the mandibles are the maxillae, one on each side of the head. Each maxilla articulates with the head capsule by a single condyle so that it is extremely mobile. This high degree of movement allows the maxillae to manipulate food between the mandibles and move it toward the mouth. The laciniae at the distal ends of the maxillae are especially important for this

and they are usually curved and pointed with the tip hardened like the mandibles. The maxillary palps are leg-like structures often with three to five segments and they have an important sensory function. At the tip of each palp is an array of contact chemoreceptors; in a large grasshopper there may be as many as 400 chemosensilla on the tip of each palp. These receptors have an important role in food selection. Grasshoppers drum on a leaf surface with the palps before accepting or rejecting it as food, and they continue to drum at intervals during feeding. Chemoreceptors are also present on the galea, a distal lobe of the maxilla immediately lateral to the lacinea.

The labium is essentially similar in structure to the maxillae but with the appendages of the two sides fused together in the midline behind the hypopharynx. There is a single articulation with the head capsule on each side, which allows the labium to swing beneath the head in the vertical plane of the body. It provides a scoop that prevents food from spreading backward from between the mandibles. As with the maxilla there are two terminal extensions on each side, known as the glossa and paraglossa. There is also a leg-like labial palp with chemoreceptors at the tip.

The labium is uniquely developed in larvae of dragonflies and damselflies, forming their prey capture equipment. The form is basically similar to the labium of other insects except that the basal parts are lengthened and the palps are claw-like. It is sometimes called a labial mask because the distal parts cover the lower part of the face when the labium is folded beneath the head. The mask can be suddenly extended by hemolymph pressure, enabling the larva to capture prey a little distance in front of it without moving its body.

FLUID-FEEDING INSECTS

Many insects feed on liquid food and their mouthparts are modified to form a tube through which fluid can be drawn into the mouth and, often, another tube through which saliva is injected into the food so that it is digested to some extent before being ingested. In most fluid-feeding insects the basic segmental arrangement and appendicular form of the mouthparts are no longer obvious, but some predaceous larvae that feed on the body fluids of their hosts are mandibulate, with mandibles resembling those of insects feeding on solid food. These are larvae of lacewings and ant lions (Neuroptera), glowworms (Lampyridae), and dytiscid beetles. In all of them, the mandibles are sickle shaped with a groove along the inner edge. In the beetles, the two sides of the groove arch over to meet, or almost meet, so that a tube is formed. In ant lions, the lacinea of the maxilla is also sickle shaped and it fits in the mandibular groove to form an enclosed canal. These insects can pump the fluid contents of their prey into the foregut through the tubes.

In other fluid-feeding insects, the basic segmental arrangement of the mouthparts is not apparent and in many insects the mouthparts themselves are drawn out into long, slender structures called stylets. The food and salivary canals are formed in different ways in different insect groups. In Hemiptera, both canals are formed between the styliform maxillae, which interlock by a tongue-and-groove mechanism that permits them to slide lengthwise with respect to each other but prevents them from coming apart (Fig. 4). They are supported in a groove along the anterior margin of the elongate labium, which is often referred to as the rostrum. The food canal is formed between the maxillary galeae in Lepidoptera, but here the two sides are linked by a series of cuticular hooks and plates that hold the two sides together while allowing them to coil up beneath the head when not in use (Fig. 5). This device makes it possible for some lepidopterans to have an extremely long tongue, which would not be possible if the insect were unable to coil it. The longest examples are in the hawk moths, the Sphingidae. Many of these have a tongue 30 mm or more in length, but one species, *Cocytius cluentis,* from South America has a tongue 250 mm long! Lepidopterans have no salivary canal in the tongue because the nectar on which they feed does not require digestion before being ingested. Each galea contains an extension of the hemocoel and the proboscis is uncoiled by an increase in pressure generated at the base of each galea. A series of short muscles extends across the galea and these muscles are involved in coiling the proboscis beneath the head. There are contact chemoreceptors at the tips of the galeae and the axons from the sensory receptor cells combine to form a nerve, which also

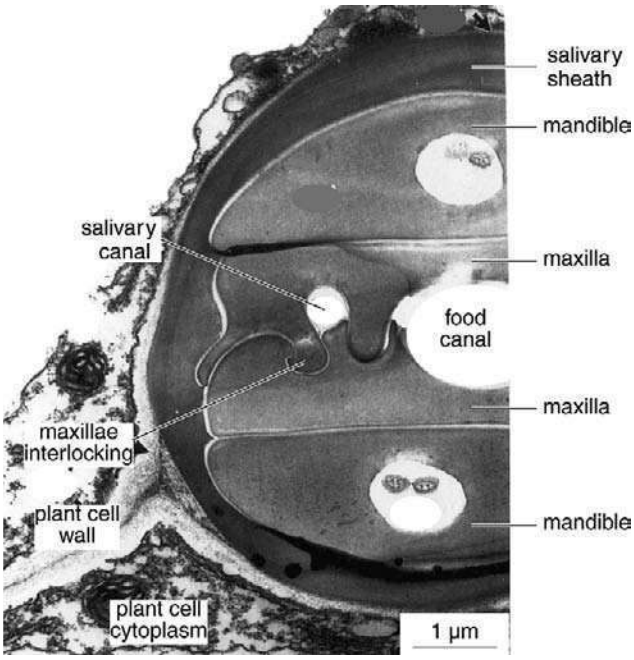

FIGURE 4 Piercing and sucking mouthparts of a hemipteran. Electron micrograph of a transverse section through part of the stylet bundle of an aphid within a leaf. Notice how the maxillary stylets interlock to form the food canal and the salivary canal. The stylets are surrounded by a sheath of solidified saliva that is produced as they penetrate the plant. The stylets are within the cell wall, which is seen in its normal form at lower left. (Reproduced with permission, from W. F. Tjallingii and T. H. Esch, 1993, Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. *Physiol. Entomol.* **18,** 317–328.)

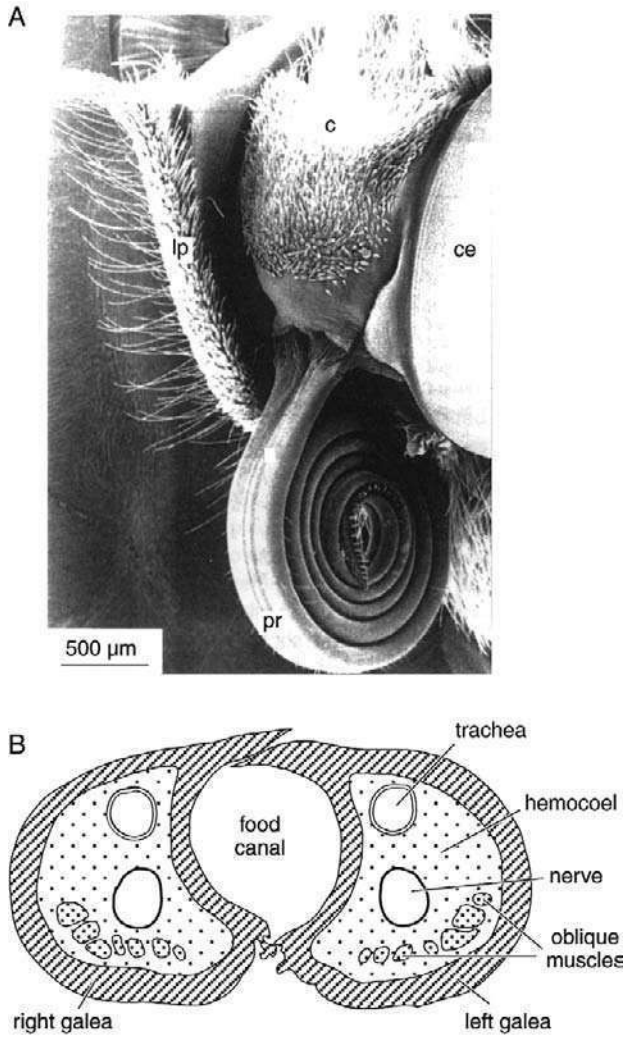

FIGURE 5 Sucking mouthparts of a butterfly. (A) Proboscis coiled beneath the head. The labial palp on the left (near) side has been removed. Abbreviations: c, clypeus; ce, compound eye; lp, labial palp; pr, proboscis. (Reprinted from *Int. J. Insect Morphol. Embryol.,* **27,** H. W. Krenn and C. M. Penz, Mouthparts of *Heliconius* butterflies (Lepidoptera: Nymphalidae): A search for anatomical adaptations to pollen-feeding behavior, 301–309, copyright 1998, with permission from Elsevier Science.) (B) Transverse section through the proboscis near the base. The galea of either side fit together to form the food canal. Each galea is blood filled and contains a nerve and trachea running the full length of the proboscis and short muscles (seen in oblique cross sections in the diagram) that run obliquely across the galea and are involved in coiling the proboscis. (Reproduced, with permission from Springer-Verlag, from H. W. Krenn, 1990, Functional morphology and movements of the proboscis of Lepidoptera. *Zoomorphology* **110,** 105–114. Copyright Springer-Verlag.)

contains motor axons to the muscles, running the length of each galea. Oxygen is supplied to these tissues via a longitudinal trachea.

Among the flies (Diptera), the feeding canal is a groove along the underside of the long labrum, closed behind by the other mouthparts, whereas the salivary canal is a narrow tube running through the styliform hypopharynx. The labium forms a sheath that encloses the stylets formed by the other mouthparts and is called the haustellum in the higher Diptera. In addition, in cyclorraphous flies the distal part of the labium is extended to form a flattened membranous lobe called the labellum. It is conspicuous in house flies and blow flies (Fig. 6). The ventral surface of each labellum is invaginated to form a series of channels that collect together medially where they make contact with the food canal in the labrum. The walls of the channels are supported by a series of incomplete rings of sclerotized cuticle. These rings prevent the channels from collapsing when suction is exerted by the pumps in the head and give the channels an appearance that is superficially like that of tracheae and so they are called pseudotracheae. The labellum, with the pseudotracheae, enables the fly to draw fluids from a relatively large surface. The channels open to the exterior via a narrow groove that is closed off during feeding except for occasional openings through which fluid can pass freely. Fleas have the food (blood) channel between a highly developed epipharynx and the two maxillae. There are two salivary canals, one in each maxilla.

The mouthparts of bees are unusual among the fluid-feeding insects. They retain normal mandibles that are used for wax and pollen manipulation, but are not involved in nectar feeding. The other mouthparts retain some semblance of the appearance in biting and chewing insects, but are elongate. The two glossae (parts of the labium) are fused together to form an elongate tongue with an open gutter posteriorly. The glossal tongue is surrounded by the lengthened and flattened galeae (of the maxillae) and labial palps. The food canal is formed by the space between the glossal tongue and the other components.

Lepidoptera, bees, and some flies feed from fluids, often nectar, that is present on a surface, but other fluid-feeding insects obtain their food from within plants or other animals and so must pierce the host tissues before being able to feed. This is true of all Hemiptera, fleas, and some flies. In Hemiptera, the mandibular stylets are the main piercing structures. The relatively stout labium does not enter the wound, but folds up beneath the insect as the mandibles and maxillae penetrate deeper into the host tissues. The stylets of aphids and coccids are very flexible and usually follow intercellular pathways that may be quite tortuous. The watery saliva of aphids contains a pectinase that degrades the pectin of the cell walls and facilitates movement of the stylets.

Among blood-sucking flies, the maxillae are the primary piercing organs of female mosquitoes (Fig. 7). They have recurved teeth distally that probably anchor the stylets in position in the wound so that when the retractor muscles contract they pull the down toward the host skin, pushing the labrum into the wound at the same time. Male mosquitoes do not feed on blood and, in many species, the piercing stylets are greatly reduced. Horse flies and deer flies (Tabanidae) have a completely different mechanism. Their somewhat elongate mandibles are articulated to the head capsule so that they move transversely, like the mandibles of biting and chewing insects. This scissor-like motion cuts through the skin of the host and the labrum and maxillae are forced into the wound. The labium

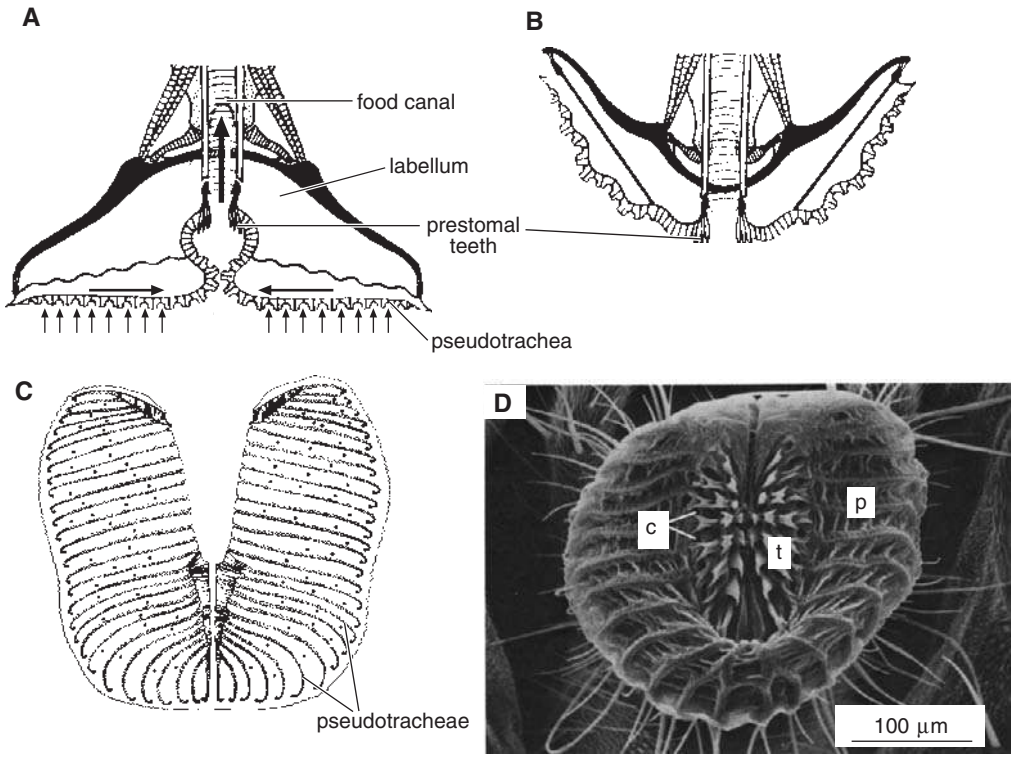

FIGURE 6 Sucking mouthparts of a blowfly. (A) Feeding by suction. With the labellar lobes spread flat on a surface, the openings of the pseudotracheae are brought into contact with fluid on the surface. Suction exerted by the pharyngeal pump draws liquid up the food canal and so through the pseudotracheal openings and along the pseudotracheae. Notice that the prestomal teeth are not exposed. Arrows show the direction of flow. (B) Feeding by rasping. By pulling back the labellar lobes, the prestomal teeth can be brought into contact with the substrate and are used to rasp at solid food. (A and B reproduced with permission from Cambridge University Press, from G. S. Graham-Smith, 1930, Further observations on the anatomy and function of the proboscis of the blowfly, *Calliphora erythrocephala* L. *Parasitology* **22,** 47–115.) (C) View of the ventral surface of the labellum showing the arrangement of pseudotracheae. The smaller branches join with major collecting trunks, which are functionally connected to the food canal. (Reproduced with permission from Wiley–Liss, Inc., a subsidiary of John Wiley & Sons, Inc., from M. Wilczek, 1967, The distribution and neuroanatomy of the labellar sense organs of the blowly *Phormia regina* Meigen. *J. Morphol.* **122,** 175–201.) (D) Ventral view of the labellum with the prestomal teeth everted for rasping. Abbreviations: c, openings of collecting channels into which the pseudotracheae open; p, pseudotracheae; t, prestomal teeth. (Reproduced with permission from Elsevier Science, from Smith, 1985.)

does not enter the wound and blood is taken directly into the food canal on the inner side of the labrum. Tsetse flies *(Glossina)* and stable flies *(Stomoxys)* penetrate the host tissues by a rasping movement of prestomal cuticular teeth on the labellar lobes (Fig. 8). House flies have similar, but much smaller teeth that they may use to rasp the surface of solid food, but in their blood-sucking relatives the teeth are stronger and are accompanied by banks of cuticular spines that form rasps. When these flies feed, the teeth and rasps are rapidly rotated round the tip of the labellum in a series of rasping movements that enable them to tear through the host's skin. For example, compare Fig. 8B, which depicts the rasps and spines on the inside of the labellar lobe and pointing downward, with Fig. 8C, in which they are on the outside and pointing upward. Contact chemoreceptors are also exposed as the teeth are moved round the tip of the labellar lobes so that they are in a position to detect blood as the host's capillaries are damaged.

Once the insect starts to feed, the properties of the fluid and dimensions of the food canal affect the rate of uptake and so the rate of nutrient intake. The more viscous a fluid, the more slowly it flows, so that although nectar containing high concentrations of sugars has more nutrients per unit volume, it also is taken up more slowly than a more dilute solution. The flow rate is also negatively correlated with the length of the proboscis, but positively correlated with the diameter of the food canal; the greater the diameter of the canal, the faster the fluid flows.

Movement of fluid into the gut is affected by three factors: the hydrostatic pressure of the fluid in the host organism, capillarity, and muscular activity. If the fluid is under high pressure, simply piercing the vessel containing it is sufficient for the fluid to be forced out, just as water gushes out from a burst water main. Phloem, the fluid carrying sugars and amino acids away from photosynthetic tissues to other parts of a plant, is under such positive pressure, up to 1 MPa, and, consequently, phloem-feeding insects, such as most aphids, have simply to penetrate a sieve tube and the phloem is forced through the food canal and into the gut. If the stylets are cut experimentally, phloem continues to ooze out, and this oozing provides a method for obtaining samples of phloem. Although vertebrate blood is under pressure, pressure in the blood capillaries is probably too low to play a major part in forcing blood into the insect's gut.

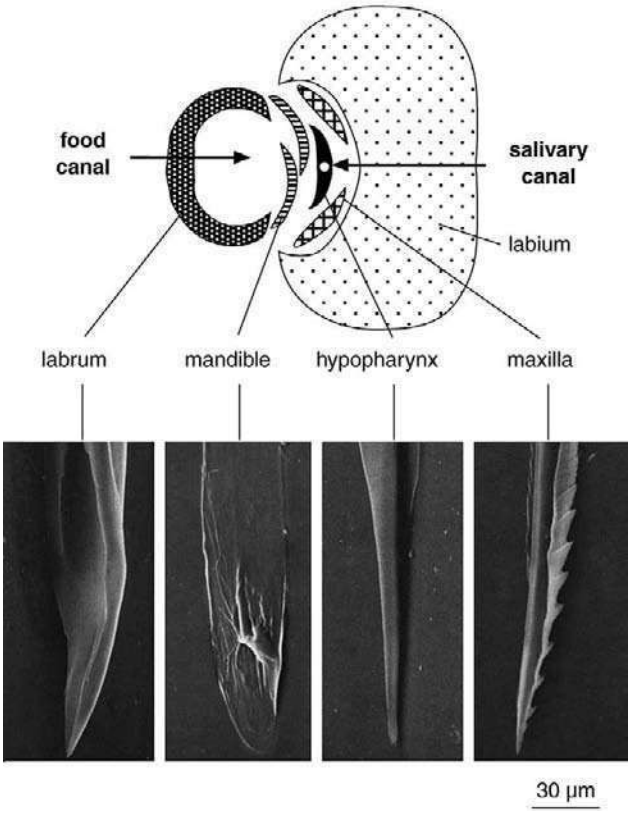

FIGURE 7 Piercing and sucking mouthparts of a female mosquito. Above is a transverse section through the proboscis. Below are electron micrographs of the tips of the stylets. (Reproduced, with permission from Elsevier Science, from Smith, 1985.)

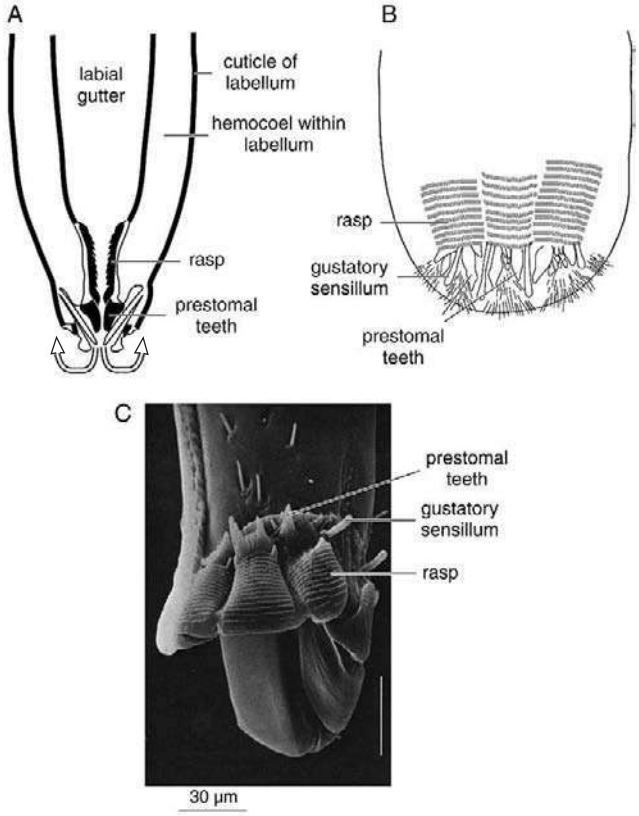

FIGURE 8 Piercing and sucking mouthparts of a tsetse fly. (A) Longitudinal section through the tip of the labellum at rest. The rasps are internal, and prestomal teeth are concealed within the opening. Arrows indicate the directions of movement of the rasps when the labellar lobes are everted. (B) Inside of one labellar lobe in the rest position. Each rasp is made up of rows of downwardly pointing cuticular spines. (A and B reproduced, with permission from Cambridge University Press, from B. Jobling, 1933, A revision of the structure of the head, mouth-part and salivary glands of *Glossina palpalis* Rob-Desv. *Parasitology* **24,** 449–490.) (C) Side view of the labellum with rasps everted. The rasps and prestomal teeth are now on the outside of the labellar lobe. Notice that the teeth now point upward. In moving from the position shown in (A) and (B), the pointed ends have scraped the skin of the host. This movement, and the anatomical arrangement, are basically similar to that seen in the blow fly (Fig. 6). (Reproduced, with permission from Elsevier Science, from Smith, 1985.)

Xylem, in contrast to phloem, is under high negative pressure which may exceed –1 MPa. [Xylem is the fluid imbibed by the roots of plants and drawn upward through xylem vessels as a result of water loss (transpiration) from the leaves.] Consequently, insects that feed on xylem require a powerful pump to overcome this negative pressure and draw the fluid into the mouth. Cicadas have a highly developed cibarial pump made obvious externally by the inflated clypeus. The cibarium is the space between the mouthparts, outside the mouth. In fluid-feeding insects, this space forms a continuum between the mouthparts and the mouth. Blood-sucking insects such as mosquitoes also have a cibarial pump, but it is less well developed than the pharyngeal pump formed by the first part of the foregut.

The importance of capillarity in insect feeding is not well understood. Insect cuticle, in general, tends to be water repellent but, if the cuticle lining the food canal in the mouthparts is wettable, capillarity might be important. In honey bees, when the glossal tongue is dipped into nectar, the fluid adheres to it, being held in place by hairs that project from the surface of the tongue. The glossa is drawn in and out between the folds of the galeae and labial palps and it is probable that capillarity is important in drawing the fluid toward the mouth.

See Also the Following Articles

Feeding Behavior • Rostrum

Further Reading

Chapman, R. F. (1995). Mechanics of food handling by fluid-feeding insects. *In* "Regulatory Mechanisms of Insect Feeding" (R. F. Chapman and G. de Boer, eds.), pp. 3–31. Chapman & Hall, New York,

Kingsolver, J. G., and Daniel, T. L. (1995). Mechanics of food handling by chewing insects. *In* "Regulatory Mechanisms of Insect Feeding" (R. F. Chapman and G. de Boer, eds.), pp. 32–73. Chapman & Hall, New York.

Labandeira, C. C. (1997). Insect mouthparts: Ascertaining the paleobiology of insect feeding strategies. *Annu. Rev. Ecol. Syst.* **28,** 153–193.

Smith, J. J. B. (1985). Feeding mechanisms. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 4, pp. 33–85. Pergamon Press, Oxford.

Movies, Insects in

May R. Berenbaum and Richard J. Leskosky
University of Illinois

The ubiquity of insects and the frequency of their interactions with humans virtually ensure that they will feature prominently in cultural contexts. Throughout history, insects have conspicuously appeared in a range of visual media, including painting, sculpture, printing, and engraving. Thus, with the advent of film in the late-19th century, insects were depicted in some of the earliest efforts; since then, they have made appearances in virtually every form of this modern medium. As the 21st century begins, insect images are common in film and television, and their role in cinema is firmly established. In fact, their impact on culture has been so pronounced that references to insect movies even serve on occasion as punch lines in jokes and cartoons, and the expression "big bug film" is widely recognized.

THE TAXONOMY OF CINEMA

What constitutes an insect in cinema is not necessarily consistent with scientific standards. In the taxonomy of cinema, any jointed-legged, segmented organism with an exoskeleton is likely to be classified as an insect, irrespective of how many legs or how few antennae it possesses. For example, in *Sherlock Holmes and the Spider Woman,* Holmes identifies a spider (used by the Spider Woman to dispatch her victims) as *Lycosa carnivora* from the Obongo River in Africa, the "deadliest insect known to science"; in actuality, spiders are classified as arachnids and not as insects at all. Taxonomic categories are also ill defined in cinema; in the film *Tarantula,* for example, the artificial-nutrient-enhanced giant spider is identified as being "from a species called Arachnida—a tarantula to be exact." The taxon Arachnida is not a species; rather, it is a class, containing thousands of species. In *Mimic,* the eminent entomologist Dr. Gates, mentor of a young scientist engaged in genetic engineering experiments with cockroaches, makes reference to the "Phylum Insecta." Again, "Insecta" is the name of the taxonomic unit called a "class"; the phylum to which insects belong is Arthropoda.

Insect morphology in the movies reflects the relatively sketchy familiarity most filmmakers have with entomological reality. As is the case for real-life insects, most movie insects have six legs, whereas movie arachnids often have eight. In general, movie insects also have the three characteristic body regions—head, thorax, and abdomen—that differentiate them from other arthropods. Even at the ordinal level, many morphological features are depicted with some degree of accuracy. Movie mantids can have raptorial forelegs (e.g., *The Deadly Mantis*), and movie lepidopterans (e.g., *Mothra,* the giant moth that attacked Tokyo in a series of Japanese films from the 1960s) possess scales. Large flat objects the size of pie plates scattered around the countryside (eventually identified as oversize scales) provide evidence of an enormous moth in *The Blood Beast Terror.*

Other aspects of insect anatomy, however, are not so accurately portrayed. Compound eyes are cause for some confusion; many films present an "insect-eye" view of a particular scene (usually a victim-to-be) through a Fresnel lens, to simulate what is imagined to be the image created by compound eyes (e.g., in *Empire of the Ants*). In reality, these images appear to insects to be more like mosaics than repeated images. In *Monster from Green Hell,* the compound eyes of the cosmic wasps roll in their sockets; real compound eyes are incapable of such motion. Antennae are also poorly understood anatomical features; on occasion, movie arachnids are equipped with a pair, even though antennae are lacking in real-life arachnids. Not surprisingly, mouthparts (whose intricacies in real insects are rarely visible to the naked eye) in movie arthropods often bear little resemblance to real arthropod mouthparts.

Insect physiology in movies often bears only a passing resemblance to the physiology of real arthropods. According to Dr. Elliot Jacobs, the entomologist in *Blue Monkey* who assists in attempting to control an outbreak of genetically engineered mutant cockroaches in a hospital, "Insects aren't like humans or animals. They're 80% water and muscle. They have very few internal organs." A recurring conceit in insect films is the violation of the constraint imposed by the ratio of surface area to volume—movie arthropods routinely grow to enormous size without suffering the limitations of tracheal respiration or ecdysis and sclerotization experienced by real-life arthropods. Nonetheless, there are physiological attributes of film arthropods that are reproduced with some degree of fidelity. Insect pheromones figure prominently in insect fear films (although they are not always identified as such; in *The Bees* they're called "pherones"). In *Empire of the Ants,* for example, giant ants use pheromones to enslave the local human population and to compel the humans to operate a sugar factory for them. The explanation provided for the response is that a pheromone "causes an obligatory response—did you hear that? Obligatory. It's a mind-bending substance that forces obedience. . . ." Although they have long been documented to exist in a wide range of organisms (including humans), pheromones rarely appear in science fiction films outside an entomological context.

As is the case with insect physiology and morphology, insect ecology takes on different dimensions in the movies. Life cycles are unorthodox and generally dramatically abbreviated by entomological standards. In *Mosquito,* for example, mutated mosquitoes, the offspring of normal mosquitoes that had consumed the blood of aliens in a crash-landed UFO, have a life cycle consisting only of egg and adult stages. In *Ticks,* full-grown ticks eclose from what seems to be a cocoon. Population dynamics differ as well. A number of movie arthropods seem to have a population size of one (as evidenced by the titles—e.g., *Tarantula, The Deadly Mantis*),

and reproduction does not seem to occur (at least over the 2-h span of the movie). At the other extreme, populations often build up to enormous sizes without depletion of any apparent food source. Bees blacken the sky in *The Bees* and *The Swarm* in a remarkably short period of time with no superabundance of nectar sources in evidence. It must be assumed that food utilization efficiencies of virtually all film arthropods are far higher than they are in real life because arthropods in films, giant or otherwise, rarely produce any frass (in *Beginning of the End,* giant grasshoppers that consume several tons of wheat in a 3-month period with little or no frass to show for it). In *Starship Troopers,* it is unclear what the giant arthropods living on a planet that is bereft of other life-forms eat to attain their large size. However, because they are alien insects, terrestrial biological standards may not necessarily be applicable.

Insect behavior in big bug films is often biologically mystifying. Screen insect predators and herbivores alike almost invariably announce their presence with an ear-piercing stridulating sound (e.g., *Them, The Deadly Mantis, Beginning of the End, Empire of the Ants*); in reality, such behavior would alert prey to danger and elicit escape or defensive behavior (which, on the part of humans in many films, involves machine guns and bazooka fire directed at the insect). For example, in *Beginning of the End,* a television newscaster updates viewers in Chicago on the Illinois National Guard's efforts against hordes of gigantic radiation-induced mutant grasshoppers descending on the city, reassuring them that "the one advantage our forces hold over the enemy is that they ALWAYS reveal their intention to attack. Before every attack the locusts send forth this warning in the form of a high-pitched screech. Now, this screech increases in intensity until it reaches ear-shattering proportions. And it's when this screech reaches its full intensity that the locust attacks." Such maladaptive behavior is unlikely to persist in nature.

INSECTS IN ANIMATED FILMS

Until the mid-20th century, insect representation in cinema was restricted largely to animated films. The small size of insects presented challenges to the standard equipment of the time that could not be met without either a disproportionate increase in cost or a decrease in visual quality. In animated films, however, technical limitations could be avoided; to create the illusion of a close-up, the animator can simply draw a larger image. In animated films, one or two frames are exposed at a time, and between exposures small changes are introduced; for example, one drawing may be substituted for another slightly different drawing or a puppet or clay model slightly repositioned. When the film is projected at normal speed, the image appears to move. An insect may even have inspired one film pioneer to become one of the first animators. Segundo de Chomon, a Spanish filmmaker of the late 19th and early 20th centuries, allegedly conceived of the animation process while shooting intertitles for a silent film and noticing that a fly, included on the footage exposed a frame at a time, appeared to move in a jerky fashion when the film was projected.

This is not to say that insects were not a challenge to animators. Because of their many moving body parts—six legs, two antennae, and from two to four moving wings—many animators simplified their drawing by reducing appendages. Thus, in animated films, insects may be depicted with four instead of six legs and spiders with six instead of eight. The first appearance of an insect in an animated cartoon was in a 1910 film by Winsor McCay titled *How a Mosquito Works*—the second American animated cartoon. Although McCay accurately portrayed his mosquitoes with six legs and two wings, in contrast with later animated films featuring insects with reduced appendages, even this early film contains many of the other conventions typically found in insect cartoons—an adversarial relationship between humans and insects, as well as the depiction of insect mouthparts as tools. McCay showed the film in vaudeville houses to large crowds and later returned to the use of insect characters with his 1921 film *Bug Vaudeville.*

Another very early example of puppet animation was provided by entomologist-turned-animator Wladislaw Starewicz. In attempting to film the mating behavior of stag beetles, Starewicz discovered that the hot lights used to illuminate his subjects caused them to stop moving altogether; accordingly, he killed and dismembered the beetles and wired their appendages back onto their carcasses, painstakingly repositioning them for sequential shots in the short film *The Fight of the Stag Beetles.* That film and its fictionalized sequel, *Beautiful Lucanida or the Bloody Fight of the Horned and the Whiskered,* proved to be quite popular with audiences. Starewicz expanded his efforts, eventually abandoning real insects and constructing puppets *de novo* for his later films with more complex plots (as in *Revenge of the Kinematograph Cameraman,* a story of love and betrayal among a variety of insect species).

Arguably the most well-known animated arthropod in animation was Jiminy Cricket, who initially appeared in a supporting role in the 1940 Walt Disney feature *Pinocchio.* Disney animators used a talking cricket, a minor character that appeared in the original Pinocchio story by Carlo Collodi, to unify disparate elements within the film. The character proved to be popular as a "voice of conscience" and appeared in several series of subsequent short subjects and educational films. Jiminy exemplifies the liberties taken with insect morphology by animators; although early sketches depicted the character with more insect-like features, the final film version, with its two arms and two legs, eyes with pupils, and morning coat and vest, resembles a dapper elf more than any arthropod.

Computer animation developed at a rapid pace during the 1980s and has proved particularly well suited to depicting insects. Modern methods of computer-generated imagery (CGI) have become particularly effective at creating shiny metallic surfaces and at joining slender rodlike structures to larger volumes—exactly what is needed to depict an insect's exoskeleton and multiple appendages. The first computer-animated insect was Wally-Bee, in the 1984 short film from

Pixar titled *The Adventures of Andre and Wally-Bee;* this film was the first computer-animated short film with a plot line. That CGI offers technological advantages over traditional animation is not to say that it has resulted in more realistic animated insects. *A Bug's Life* (1998) from Disney/Pixar continued to depict insects with anthropomorphized faces and four limbs to ensure audience empathy with the characters; the DreamWorks film *AntZ* (1998) gave its ants six legs but provided them with similarly humanized faces and raised the head and thorax into a vertical position, making them look like tiny centaurs. CGI is not limited to what is basically caricature, however; the otherwise live-action *Joe's Apartment* (1996) featured hundreds of computer-rendered cockroaches which were indistinguishable from the real thing, except for their ability to sing and dance.

INSECTS IN FEATURE FILMS

Big Bug Films

Frequent appearances by insects in live-action films are a relatively recent phenomenon in the history of film. For many years, the technical challenges of filming very small, largely untrainable, fast-moving creatures proved a disincentive for incorporating them into films. The pioneering efforts of special-effects genius Willis O'Brien, starting in the 1930s, and of his protégé Ray Harryhausen, as well as technical advances in the production of film stock and traveling matte techniques, gradually made the incorporation of insect images in film economically attractive, or at least reasonable. Moreover, competition for audiences, particularly with the rise of television, led the major film studios to increase investment in hitherto minor genres, such as science fiction. With bigger budgets, more elaborate effects became feasible.

The year 1954 was a watershed year; *Them!* was released by Warner Brothers Studios, featuring giant ants mutated by exposure to atomic testing in the Arizona desert (Fig. 1). The film, tapping into widespread fears of atomic power in the aftermath of World War II, was an enormous success, grossing more money for the studio that year than any other and winning an Academy Award for special effects. Its success is understandable in retrospect: its use of large mechanical models was innovative and dramatic, its screenplay was tight and well written, it featured several big-name actors of the era, and its subtext about invasion disrupting the fabric of American life played well to American fears of communist powers.

The "big bug films" inspired by the success of *Them!* were by and large lesser efforts. Many of these were the work of director/producer Bert I. Gordon, who made so many films with big animals that he was known as "Mr. Big" (a reference as well to his initials—B.I.G.); his big bug films included *Beginning of the End* (1957), featuring giant radiation-induced grasshoppers threatening to destroy Chicago, and *Empire of the Ants* (1977), about giant radioactive-waste-induced ants threatening a real estate development in Florida. Other notable titles of the fifties in the "big bug" genre (Table I) include *Tarantula, The Deadly Mantis, The Black Scorpion, Monster from Green Hell,* and *Earth vs the Spider.* The Japanese film industry did not embrace this genre until the 1960s but made up for the slow start in volume; the first Japanese big bug film, *Mothra,* featuring a giant radiation-induced moth, was released in 1962 and was followed by four sequels, in which Mothra appeared with other "big" science fiction stars such as Godzilla and Rodan.

FIGURE 1 Lobby poster from the science fiction classic *Them!* (1954), noted for its dramatic special effects and suspenseful screenplay. The firm depicted an attack on the city of Los Angeles by ferocious giant ants, which have been enlarged to greater than human size by the mutating effects of radiation exposure. Made during the height of the 1950s Red Scare, *Them!* attacted large audiences with its ability to link the imaginary threat of gigantic, murderous insects with America's very real fears of nuclear fallout, foreign invasion, and scientific manipulation of the natural world. (*Them!* © 1954 Warner Bros. Pictures, Inc. All rights reserved.)

Transformation/Metamorphosis Films

Metamorphosis is a characteristic of a substantial proportion of movie arthropods, although the process differs on screen.

TABLE I Live-Action Feature Films with Insects as Major Components

Year	Film
1938	*Yellow Jack*
1944	*Sherlock Holmes and the Spider Woman*
1944	*Once upon a Time*
1953	*Mesa of Lost Women*
1954	*Them!*
1954	*Naked Jungle*
1955	*Tarantula*
1955	*Panther Girl of the Kongo*
1957	*The Black Scorpion*
1957	*Beginning of the End*
1957	*Earth vs the Spider*
1957	*The Deadly Mantis*
1957	*Monster from Green Hell*
1958	*The Cosmic Monsters*
1958	*She Devil*
1958	*The Fly*
1959	*Return of the Fly*
1959	*The Brain Eaters*
1959	*Wasp Woman*
1962	*Mothra*
1964	*Godzilla vs The Thing*
1965	*Horrors of Spider Island*
1966	*The Deadly Bees*
1968	*Destroy All Monsters!*
1969	*The Blood Beast Terror* (aka *The Vampire Beast Craves Blood*)
1970	*Flesh Feast*
1971	*The Hellstrom Chronicle*
1971	*The Legend of Spider Forest*
1972	*Kiss of the Tarantula (Shudders)*
1973	*Invasion of the Bee Girls*
1974	*Phase IV*
1974	*Locusts*[a]
1974	*The Killer Bees*[a]
1975	*Bug*
1975	*The Giant Spider Invasion*
1975	*Food of the Gods*
1976	*The Savage Bees*[a]
1976	*Curse of the Black Widow*[a]
1977	*Empire of the Ants*
1977	*Exorcist II—The Heretic*
1977	*Ants: It Happened at Lakewood Manor*[a]
1977	*Kingdom of the Spiders*

continues

TABLE I (*Continued*)

Year	Film
1977	*Terror out of the Sky*[a]
1978	*Tarantulas: The Deadly Cargo*[a]
1978	*The Bees*
1978	*The Swarm*
1978	*Curse of the Black Widow*[a]
1980	*Island Claws* (aka *Night of the Claw*)
1982	*Creepshow*
1982	*Legend of Spider Forest*
1985	*Flicks*
1985	*Creepers* (aka *Phenomena*)
1986	*The Fly*
1987	*Blue Monkey*
1987	*The Nest*
1987	*Deep Space*
1989	*The Fly II*
1990	*Arachnophobia*
1991	*Meet the Applegates*
1991	*The Age of Insects*
1993	*Cronos*
1993	*Ticks*
1994	*Skeeters*
1995	*Mosquito* (aka *Nightswarm*)
1996	*Angels and Insects*
1996	*Wax, or the Discovery of Television among the Bees*
1996	*Joe's Apartment*
1996	*Wasp Woman*[a]
1997	*Starship Troopers*
1997	*Men in Black*
1997	*Mimic*
1998	*X-Files: The Movie*
1999	*Deadly Invasion: The Killer Bee Nightmare*[a]
1999	*Atomic Space Bug*
2000	*They Nest*[a]
2000	*Bug Blaster*[a]
2000	*Spiders* (aka *Cobwebs*)
2001	*Evolution*
2001	*Bug*
2001	*Mimic II: Hardshell*
2001	*Spiders II*
2002	*Men in Black II*
2002	*Spiderman*

[a] Made-for-television movie.

The transformation most frequently depicted in films is insect to human or human to insect, generally involving some form of exchange of body fluids—"*Drosophila* serum" in the case of *She Devil* (which allows the patient to transform herself at will from brunette to blonde), "spider hormones" in *Mesa of Lost Women,* "royal jelly" in *Wasp Woman,* and "DNA" in the 1986 remake of *The Fly*. Insects most likely appear frequently in films involving metamorphosis because of the shock value—the transformation of a human into a life-form radically different in appearance. Generally, transformations of humans into other animal forms in films involve magic or reincarnation *(The Shaggy Dog, The Shaggy D.A., Oh, Heavenly Dog, Lucky Dog)* or genetic predisposition *(Teen Wolf* and its sequel, *The Howling, Cat People)* rather than mediation by hormones (with the exception of the early films of Bela Lugosi, including *The Ape Man* and *Return of the Ape Man,* which involve serum exchanges between humans and apes).

In the 1980s, insect fear films acquired a new life with the release of David Cronenberg's *The Fly.* Although as scientifically as inaccurate as earlier efforts with respect to surface area/volume rules, it was generally regarded by critics as an artistic success, thematically depicting physical transformation leading to mental and emotional change. Although *The Fly II* (directed Chris Walas, special-effects artist on the earlier film) was not embraced as enthusiastically by critics, it nonetheless was perceived as more than just a horror film, with allegorical elements relating the physical and emotional

changes of adolescence with the metamorphic transformation of the protagonist. Despite the presence of redeeming intellectual content in these films, these films attracted considerable attention for their graphic special effects, far surpassing earlier efforts.

In 1983, the first report of successful genetic transformation of an insect *(Drosophila melanogaster)* was published, and by 1987 genetically engineered insects (specifically, mutant killer cockroaches) made their first appearance in a science fiction film, in the otherwise unremarkable film *The Nest.* Genetic engineering techniques advanced more quickly on screen than in real life; by 1997, in *Mimic,* the young entomologist Susan Tyler (Mira Sorvino) is able to incorporate termite and mantid DNA into cockroaches, with the goal of creating a "Judas bug" to bring contagion to the cockroach vectors of a human illness but instead unleashing a plague of six-foot-tall people-eating cockroaches in the subway system of New York City. In *Spiders,* unspecified alien DNA is incorporated into the titular arthropods to wreak havoc in a secret government laboratory.

"Social" Insect Films: Small Size, Large Numbers

One of the largest orders of real-life arthropods, the Hymenoptera, is in fact the most frequently depicted in insect fear films. There may be several reasons for this proportional similarity. For example, bees are relatively easily manipulated for the camera in comparison with other insects, can be produced commercially, and can be reared in enormous numbers with comparative ease. Perhaps an even more important factor, however, is their familiarity to the audience. Encounters with bees, ants, and wasps are part of the normal course of life for most moviegoers. Such an explanation also can account for the proliferation of films involving cockroaches, although these, too, share the practical advantage of ease of rearing in enormous numbers and affordability.

Films using footage of real insects engaging in more or less normal insect behaviors rose to prominence in the 1970s and included such efforts as *Phase IV,* featuring documentary-quality footage of ants, and *Bug,* featuring Madagascar hissing cockroaches (albeit engaged in some unusual behaviors, such as spelling out death threats with their bodies on the wall of a house). The appearance of so-called killer bees on a container ship in San Francisco harbor in 1974 may have inspired filmmakers to capitalize on a real threat—the introduction of African honey bees with a reputation for defensive behavior often lethal to animals and sometimes to humans. The films proved popular with filmmakers in part because audience members enter the theater with at least passing familiarity with the film's antagonists (in contrast with giant arthropods). As well, bees can be controlled chemically—by pheromones—to cluster or land in a particular spot and so are more easily manipulated for special effects. Five films were made about killer bee invasions between 1974 and 1978 (Table I), although none of them was particularly successful at the box office (surprisingly for the 1978 film *The Swarm,* with a screenplay by Arthur Herzog and a cast including such Academy Award-caliber actors as Henry Fonda and Michael Caine).

In many of the films featuring large numbers of small insects, ecological disruption is a recurring theme. Biomagnification, accumulation of toxins up a food chain, is the focus of several. In *Kingdom of the Spiders,* tarantulas take over a town and start consuming livestock because "DDT" destroyed the food chain and deprived them of their normal prey. Other films depicting altered food web dynamics as a result of pollution (radioactive and/or toxic waste) include *Skeeters* and *Empire of the Ants.* In *Ticks,* fertilizers and other chemicals used by illegal marijuana growers are encountered by ticks, which grow to enormous size and terrorize a group of inner-city teens in the woods on a wilderness survival trip.

Another ecological phenomenon of concern both in the movies and in real life is the accidental introduction of alien species (although in the movies these are more likely to be real aliens, from outer space, not just a foreign country). *Arachnophobia* depicts the fictional consequences of the accidental introduction of a South American spider species to the Pacific Northwest. The many killer bee movies pointedly make reference to the dangers of accidental importation of strains of bees into new habitats (although in *The Bees* their introduction is no accident; greedy cosmetics magnates import killer bees in the hope of producing large amounts of profitable royal jelly).

INSECTS IN DOCUMENTARY FILMS

Although educational shorts for school and extension markets often deal with entomological topics, documentary filmmaking, which combines information and art, has tended to ignore this area. For a long time, documentary filmmakers faced many of the same challenges faced by feature filmmakers with an interest in insects. Only until the latter half of the 20th century did developments in technology permit the capture of small moving objects (such as insects) on film in a compelling and effective manner. Yet another obstacle, particularly problematic for documentary filmmakers, was audience interest; whereas audiences could accept insects bent on destruction of the human race in science fiction or horror films, they generally showed considerably less interest in the accurate depiction of the lives of real-life insects. Animal documentary filmmakers have long had to accept the fact that audiences prefer drama to accuracy in depictions of nature. Walt Disney, with his groundbreaking *True Life Adventure* nature films made between 1948 and 1970, relied in many of his nature films on personification and anthropomorphism to make the animal subjects of studio films more appealing to audiences.

Arguably the first "documentary" films involving insects were the pioneering efforts of F. Percy Smith, who in 1912 created films aimed at illustrating the physical prowess of the common house fly. Smith enclosed a fly inside a dark box

equipped with a thin glass door at one end; the door in turn had a small opening into which was fitted a toothed wheel that was free to rotate. The fly, orienting to the light entering through the glass door at one end of the box, would move toward the light; when it encountered the glass door obstructing its escape, it was struck on the head by a tooth in the wheel which rotated as a consequence of the fly's movements. Eventually, via conditioning, the fly simply walked up the wheel, which would rotate, creating a treadmill and providing the photographer an opportunity to film the fly walking in place. Smith modified his approach to film flies outside the box, tethered in place, and in this way was able to obtain footage of them seemingly juggling dumbbells, corks, bits of vegetables, other flies, and sundry other objects. When the film was released newspaper reports accredited the cinematographer with strange powers, and the capacity to train house flies as others do circus animals.

Audience reluctance to accept insects for their own sake is the explanation for the peculiar framing device used in the first big-budget feature-length documentary about insects, *The Hellstrom Chronicle.* This film was originally conceived as a straightforward documentary and featured what was at the time state-of-the-art macrophotography that provided startling and dramatic close-ups of its arthropod subjects. The extraordinary inventiveness of cinematographer Ken Middleham led to spectacular images of insects engaged in a wide range of behaviors. However, the studio heads were unconvinced that a documentary about insects could bring in an audience and insisted on adding to the film a fictional storyline, about an academic, Dr. Nils Hellstrom (Lawrence Pressman), denied tenure because of his insistence that insects were bent on human destruction. As a result, the hybrid film was a commercial success as well as an artistic success of sorts (earning a Grand Prix de Technique award at the 1971 Cannes Film Festival for its remarkable images), although it was panned by critics, in part because of its sensationalistic tone.

A general awakening of the American public to environmental issues in the 1970s did little to inspire interest in insect biology, and insect documentaries have been few and far between since *The Hellstrom Chronicle.* Insects figured peripherally in the documentary *Cane Toads, An Unnatural History,* directed and written by Mark Lewis, and first shown in 1988. The cane toad *Bufo marinus* was deliberately introduced into North Queensland, Australia in 1935 to control *Lepidoderma albohirtum* (a beetle larva) and its relatives. Although the cane toads were ineffectual biocontrol agents, they were exceptionally effective colonizers, which now populate much of Queensland, northern New South Wales, and eastern Northern Territory, wreaking ecological and environmental havoc. The history of this ill-conceived biocontrol effort and its consequences are the subject of the documentary.

Microcosmos (1996) is similar to *The Hellstrom Chronicle* in that its success was due largely to quantum improvements in capturing insect images and behavior on film. Filmmakers Claude Nuridsany and Marie Perennou spent 15 years researching, 2 years designing equipment (including inventing a remote-controlled helicopter for aerial shots), 3 years shooting, and 6 months editing a masterpiece of insect cinematography. Yet the concept underlying the film was quite novel. Instead of a "superdocumentary" of amazing insect feats, the filmmakers settled on the idea of telling the story of a single summer day (albeit in reality filmed over a much longer period of time) in a field in the countryside of Aveyron, France (where they lived and worked). Their goal was to depict insects and other small creatures not as "small bloodthirsty robots" but rather as individuals with unique abilities. Instead of narrration, there was a simple introduction, 40 words, spoken by actress Kristin Scott Thomas. *Microcosmos* was well received by critics (although it failed to win a nomination for best documentary at the Academy Awards), and it performed respectably at the box office. Although some notable entomologists bemoaned the absence of voiceover and the loss of an opportunity to educate the public about the insect lives captured on film, the extraordinary images depicted on screen will likely set the standard for excellence in insect documentary filmmaking for years to come.

INSECT WRANGLERS AND SPECIAL EFFECTS

Because handling insects and other arthropods and eliciting appropriate behaviors from them on cue is beyond the experience and training of most directors, these responsibilities are frequently delegated to a specialized crew member known in the profession as an "insect wrangler" or "bug wrangler." Since the early 1960s, only a handful of individuals have engaged in this occupation in a conspicuous way. Some insect wranglers specialize in handling a narrow range of taxa. Norman Gary has been a bee wrangler for more than a quarter-century. Currently an emeritus professor at University of California at Davis, he served as a faculty member in bee biology from 1962 to 1994. His research interests have been in the area of bee behavior, and he has written or coauthored over 100 publications on bees. Since 1966, he has been a consultant for legal, industrial, film, and television productions about bees. His ingenuity in developing methods for manipulating bees and their behavior has led him to develop methods of narcotizing queens to facilitate instrumental insemination, as well as vacuum devices for handling, tagging, counting, confining, and otherwise handling bees. An abbreviated filmography for Gary includes *My Girl, Fried Green Tomatoes, Candyman, Beverly Hillbillies, Man of the House, X-Files, The Truth about Cats and Dogs, Leonard Part VI, A Walk in the Clouds,* and *Invasion of the Bee Girls.*

Another individual with an affinity for a particular taxon is Ray Mendez, who worked as an entomologist at the American Museum of Natural History in New York. Mendez, along with colleague David Brody, provided over 20,000 cockroaches for the film *Creepshow* in 1982; in 1996 Mendez wrangled 5000 live cockroaches and provided advice on animated and puppet cockroaches for the film *Joe's Apartment.*

Mendez is also an authority on naked mole rats and was featured in the documentary *Fast, Cheap and Out of Control.*

Steven R. Kutcher, a consulting entomologist in Arcadia, California, and part-time biology instructor at West Los Angeles College, is notable for the range of arthropods with which he has worked. Kutcher obtained a bachelor's degree in entomology at University of California at Davis and a master's degree in biology with an emphasis on insect behavior and ecology at California State University, Long Beach. Since 1976 he has been involved in arthropod wrangling for many movies and commercials. He has worked with a variety of arthropods, including spiders, yellowjackets, cockroaches, mealworms, grasshoppers, and several species of butterflies. Among his film credits are *Extremities, Exorcist II: The Heretic, Arachnophobia, Race the Sun, Jurassic Park,* and *Spiderman.* His unusual vocation has made him the focus of more than 100 print articles, and in 1990 his work was the subject of a short documentary by National Geographic.

INSECT FEAR FILM FESTIVALS

The idea of using insects in movies as a means of entomological outreach apparently dates back to the origins of the annual Insect Fear Film Festival at the University of Illinois at Urbana-Champaign. The first festival, brainchild of then assistant professor of entomology May Berenbaum, was held in March 1984. The goal of the festival has been to use insect fear films to draw in an audience and to use the films as a means for highlighting scientific misconceptions about insects. At each festival, two or three feature-length films are shown, interspersed with animated shorts. Before the festival begins, and between films, the audience is invited to see and handle a variety of live specimens as well as pinned specimens. Generally, the festivals are organized around themes, which have included female insects, noninsect arthropods, orthopteroids, social insects, cockroaches, flies, and mosquitoes.

Other events that have been held in conjunction with the festival included a thematically relevant blood drive, held in cooperation with Community Blood Services of Champaign, for the 1999 mosquito film festival. Attendance at these festivals can exceed 1000. Over the years, the festival has been featured in a wide range of media throughout the world.

Other insect fear film festivals per se are few in number; Iowa State University has conducted an Insect Horror Film Festival since 1985, and Washington State University has hosted its Insect Cinema Cult Classics festival since 1990. Insect films, however, have been elements of insect expo and public outreach efforts in many venues, including museums, science centers, and universities across the country.

There is one legitimate insect film festival in the traditional sense, in which films are submitted in competition and are judged and awarded prizes. FIFI, organized by l'Office pour les Insectes et leur Environnement du Languedoc-Roussillon (OPIE LR) and the the Regional Natural Park of Narbonne and the city of Narbonne, France, is a biannual international film festival dedicated to insects and other small animals. The FIFI, in its fourth year in 2001, is the result of a partnership with the Institute for Research and Development (IRD), the French National Center for Scientific Research (CNRS), the National Institute of Agronomic Research (INRA), the City of Sciences and Industry (Paris), the National Museum of Natural History (Paris), and the Agronomic University of Gembloux (Belgium). Its stated objectives are to increase the sensitivity of the media and the public to the ecological importance of continental invertebrates as well as to encourage and promote the making of films or videos dedicated to insects.

FUTURE OF INSECTS IN CINEMA

With the continuing development of CGI and the veritable explosion of such outlets for film as cable stations, satellite television, DVD and video markets, the future of insects and other arthropods in the movies looks assured. Arthropods will certainly continue to be objects of distaste and unease for audiences throughout the world and so will remain staples of horror films and certain types of science fiction adventure. Moreover, CGI and developments in macrophotography techniques ensure that insect images on screen will become increasingly sophisticated, although scriptwriting will almost assuredly remain as resolutely unrealistic as it has since the earliest days of insects in cinema.

See Also the Following Articles

Cultural Entomology • Folk Beliefs and Superstitions • Teaching Resources

Further Reading

Berenbaum, M. R. (1995). "Bugs in the System: Insects and Their Impact on Human Affairs." Addison Wesley, Reading, MA.

Berenbaum, M. R. (2000). See you in the movies. *Am. Entomol.* **46,** 210–212.

Berenbaum, M., and Leskosky, R. (1991). Mosquitoes in the movies. *Vector Control Bull. North Central States* **1(2),** 94–98.

Berenbaum, M. R., and Leskosky, R. J. (1992). Life history strategies and population biology in science fiction films. *Bull. Ecol. Soc. Am.* **73,** 236–240.

Brosnan, J. (1991). "The Primal Screen. A History of Science Fiction Films." Little Brown, Boston. (See Chap. 6, The Metaphor That Ate Tokyo: Monster Movies of the Fifties.)

Kottmeyer, M. S. (1997). Bugs invade! A cultural history of horror. *UFO Mag.* **12,** 20–25.

Leskosky, R. J. and Berenbaum, M. R. (1988). Insects in animated films; or, not all "bugs" are bunnies. *Bull. Entomol. Soc. Am.* **34,** 55–63.

Mertins, J. W. (1986). Arthropods on the screen. *Bull. Entomol. Soc. Am.,* **32,** 85–90.

Warren, W. (1982). "Keep Watching the Skies!" McFarland, Jefferson, NC.

Warren, W. (1986). "Keep Watching the Skies! II." McFarland, Jefferson, NC.

Musca domestica

see *House Fly*

Muscle System

Robert Josephson
University of California, Irvine

Muscle is the excitable, contractile tissue of animals that is responsible for movement and behavior. Although there is great variability in structure and performance among different insect muscles, many basic features of biochemical composition, ultrastructural organization, and contractile performance are common among insect muscles and indeed are similar between muscles of insects and those of vertebrates.

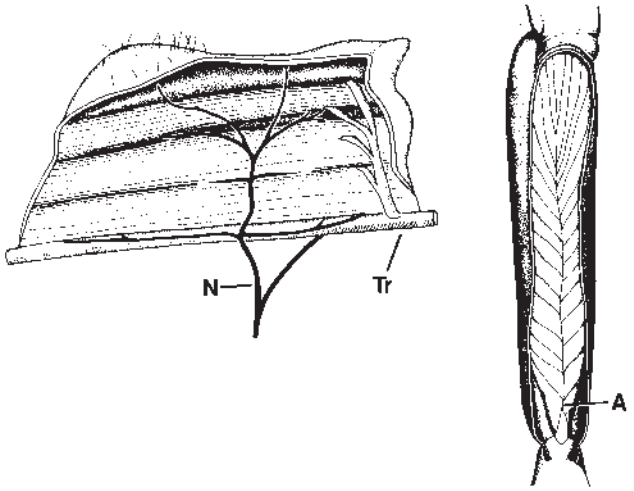

FIGURE 1 Muscle with parallel fibers (left) and one with pinnate fibers (right). The parallel-fibered muscle is the mesothoracic dorsal longitudinal flight muscle of the tettigoniid *Neoconocephalus robustus.* (Modified from Stokes, Josephson and Price, 1975, *J. Exp. Zool.* **194,** 379–407.) The dark structure coursing across the muscle surface is the motor nerve that innervates the muscle. The pinnately fibered muscle is the metathoracic extensor tibia of the cricket *Teleogryllus oceanicus.* (Modified from Donaldson and Josephson, 1981, *J. Comp. Neurol.* **196,** 735–742.) Abbreviations: N, motor nerve; Tr, trachea; A, apodeme.

MUSCLE STRUCTURE AND ULTRASTRUCTURE

Muscle Fibers and Fiber Bundles

The skeletal muscles of insects are bundles of elongate, multinucleate cells called muscle fibers. The fibers attach at each end to the exoskeleton. The muscles typically span joints of the exoskeleton and, when active, cause bending of the joint or stabilization of the joint against external forces. Skeletal muscles are the muscles of behavior, the muscles involved in posture and locomotion. In addition to skeletal muscles, insects contain visceral muscles that cause movement of the gut, Malpighian tubules, and parts of the reproductive system; there are also cardiac muscles that cause contraction of tissue sheets and vessels associated with the circulatory system. The visceral and cardiac muscle cells are typically small, spindle shaped, and with a single nucleus.

An individual insect contains many morphologically identifiable skeletal muscles. The large number of muscles is a consequence of the segmental organization of insects and the serial replication of parts associated with segmentation. Each of the wing-bearing segments of a cockroach contains about 50 separate muscles, an abdominal segment a somewhat smaller number. In a classic anatomical study, Lyonet, in 1762, noted that the larva of the goat moth, *Cossus,* contains three times the number of anatomically distinct skeletal muscles as does a human!

In most insect muscles the fibers lie parallel to one another, and when the muscle contracts it shortens along the long axis of the fiber bundle. Such muscles are spoken of as being parallel-fibered muscles (Fig. 1, left). In some muscles, in particular peripheral leg muscles, the fibers attach obliquely at one of their ends onto an internal, cuticular extension called an apodeme (Fig. 1, right). When these muscles are activated, the muscle as a whole shortens along the axis of the apodeme, oblique to the fiber axis. The oblique insertion of fibers onto the apodeme is remindful of the oblique junction between lateral filaments and the main shaft of a feather, hence muscles with an oblique fiber arrangement are called pinnate (L. *pinna* = feather). The force that a muscle can generate increases with increasing cross-sectional area. The pinnate arrangement of muscles increases the effective cross-sectional area and hence the force that the muscle can produce.

Filaments and Fibrils

Muscle shortening in insects as in other animals results from sliding movement between interdigitating thick and thin filaments contained within the muscle fibers. The force of contraction is a shearing force developed between these filaments. The thick filaments are made up largely of the protein myosin, the thin filaments of the protein actin. A single thick filament is composed of many individual myosin molecules and, similarly, a thin filament contains many actin molecules. Projections of the myosin molecules from the thick filaments toward the thin filaments, called cross-bridges, are the sites of interaction between the two and are the force generators for contraction.

The thick and thin filaments are grouped into longitudinal bundles called fibrils. The filaments within a fibril are grouped precisely, both longitudinally and transversely (Figs. 2 and 3). The thin filaments attach to and project from both sides of transverse structures called Z disks. The Z disks occur regularly along the length of the fibril. The interval from one Z disk to the next is called a sarcomere. The thick filaments lay side by side in the middle of the sarcomere. The sarcomere lengths in fibrils of fast muscles such as flight muscles are 2 to 4 μm; those in leg muscles, body wall muscles, and visceral muscles tend to be longer, up to 7 to 10 μm. The regular longitudinal arrangement of Z disks, thin filaments, and thick filaments creates a striped pattern along the length of a fibril

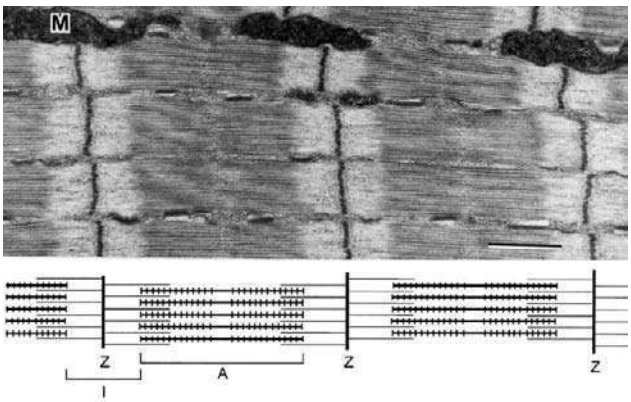

FIGURE 2 Origin of the transverse striations in skeletal muscle. The upper electron micrograph is a longitudinal section of a somewhat stretched fiber from the mesothoracic dorsal longitudinal muscle of the tettigoniid *Neoconocephalus ensiger.* The scale bar represents 1 μm. Abbreviations: M, mitochondrion; I, I band; Z, Z disk; A, A band.

(Figs. 2 and 3). The most obvious components of the striped pattern are (1) the Z disks; (2) the A bands, corresponding to that part of a sarcomere containing thick filaments; and (3) the I bands, corresponding to that part of a sarcomere without thick filaments. When a muscle shortens, the thin filaments slide toward the center of the sarcomere and the I bands become shorter. Because of their transverse banding pattern, muscles in insects (and skeletal muscle in vertebrates, which have a similar organization) are described as being striated muscles. The visceral muscles of insects are similar in function to vertebrate smooth muscles and in many ways similar in physiology as well. But although vertebrate smooth muscles lack striations, the visceral muscles of insects, like the skeletal muscles, are striated.

The thick filaments of insect muscles, and of vertebrate striated muscles, occur in a regular, hexagonal array. In vertebrate muscles a thick filament is surrounded by 6 thin filaments, each of which lies at the midpoint between three adjacent thick filaments (Fig. 4), and the overall ratio of thin to thick filaments is 2:1. In fast muscles of insects, for example flight muscles, there are also 6 thin filaments surrounding each thick filament, but these occur at the midpoint between two thick filaments and the thin-to-thick ratio is 3:1. In slower insect muscles, such as body wall muscles, the usual pattern is for each thick filament to be surrounded by a circle of up to 12 thin filaments.

The fibrils of insect muscles occur in two basic patterns, cylindrical and radial–lamellar (Figs. 3 and 5). In muscles with cylindrical fibrils the bundles of filaments forming the fibrils occur as elongate cylinders that are often polygonal in cross section. In radial–lamellar fibers the fibrils are ribbon-shaped structures arranged radially about the center of the fiber.

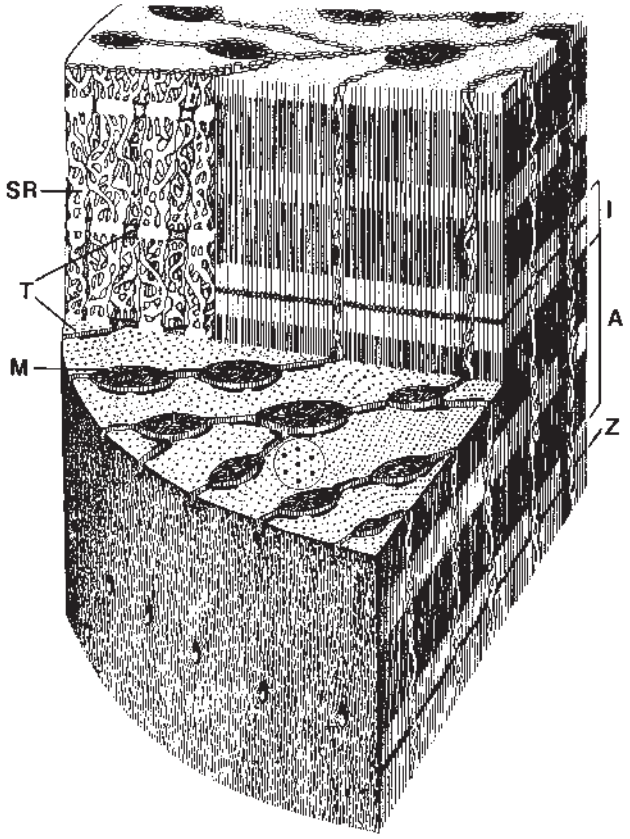

FIGURE 3 Structural organization of a fiber from an insect fast muscle. The drawing is based on electron micrographs from a tettigoniid singing muscle. The fibrils here are radial–lamellar. Abbreviations: A, A band; I, I band; M, mitochondrion; SR, sarcoplasmic reticulum; T, transverse tubule; Z, Z disk. (Modified from Josephson, 1975.)

Other Components

The cellular components of muscle fibers seen in electron micrographs fall into four functional groups. First are those structures directly involved in the generation of force and mechanical power. These structures are the thick and thin filaments that collectively form the fibrils. Second are those components involved in the control of contraction. The most obvious structures involved in the control of contraction are the transverse tubular system (T tubules) and the sarcoplasmic reticulum (SR). The T tubules are membrane-bound tubular structures oriented perpendicular to the fiber axis. The membrane of a T tubule is continuous with the surface membrane of the fiber, and the T tubule can be regarded as an inwardly directed

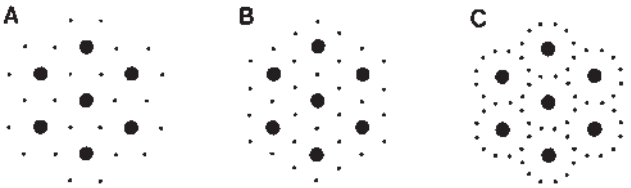

FIGURE 4 Organization of thick and thin filaments as seen in cross sections of fibers from (A) a vertebrate skeletal muscle, (B) a fast insect muscle, and (C) a slow insect muscle.

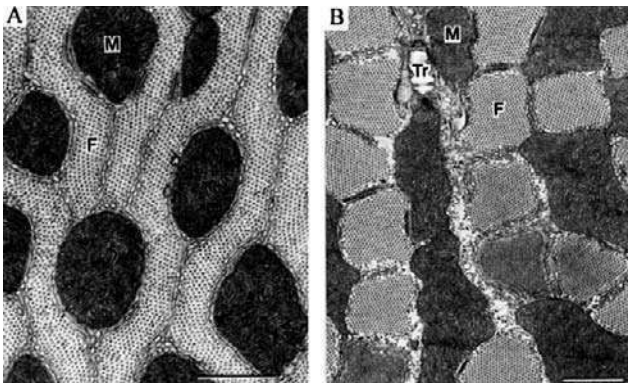

FIGURE 5 Transverse electron microscope sections through (A) a fiber with radial–lamellar fibrils (a flight muscle of the tettigoniid *Euconocephalus nasutus*) and (B) a fiber with columnar fibrils (from the tymbal muscle of the cicada *Abricta curvicoasta*). The scale bars indicate 1 μm. Abbreviations: M, mitochondrion; F, fibril; Tr, intracellular tracheole.

extension of the surface membrane. In most insect muscles there are two tubules per sarcomere, lying in the overlap areas between thick and thin filaments (Fig. 3), but in the fibers of some muscles there is a single, centrally located T tubule per sarcomere. Within the fiber the T tubules make specialized junctions with the SR, which is an internally closed, membrane-bound compartment within the fiber. The function of the T tubules and SR is considered further below. Other elements involved in the control of contraction are the surface membrane of the muscle fibers and membrane specializations at the sites at which nerve processes contact muscle fibers. Third are the structural elements of the metabolic power supply. These are the mitochondria, which provide ATP, and glycogen granules. ATP is the immediate energy source for contraction; glycogen is a stored fuel for cellular metabolism. It would be appropriate to include among the elements involved in metabolic power the tracheoles, the terminal portions of the gas-exchange system that ramify throughout muscle fibers, even though topologically tracheoles are external to and not really part of the muscle fibers. Fourth are the structures involved in long-term maintenance of muscle, specifically the many nuclei of the fibers.

The relative abundance of different cellular components in muscle is tightly correlated with the functional capacity of the muscle fibers. SR and T tubules are particularly abundant in muscles that can produce brief contractions, that is, in muscles in which the contractile apparatus can be rapidly activated and inactivated. Muscles capable of sustained activity at high power output are particularly well supplied with mitochondria and tracheolar endings. Mitochondria make up 30 to 40% of the muscle volume in wing muscle of active fliers and in sound-producing muscles that are active continuously and at high frequency. Such muscles are often pink, because of the cytochromes in the abundant mitochondria. It should be noted that hypertrophy of mitochondria, and of T-tubules and SR, is at the expense of myofibrillar volume, so fast and fatigue-resistant muscles are likely to be relatively weak.

Muscle Attachments

Skeletal muscles attach to the cuticle of the exoskeleton through specialized epidermal cells. The muscle fibers are joined to these cells by specialized junctions. called desmosomes. The terminal sarcomeres of the fibrils lack a final Z disk; instead, the thin filaments are attached to the muscle portion of the terminal desomsome through a band of what have been called junctional filaments. Visceral muscles are frequently joined to one another by desmosomes, and cardiac muscle fibers are joined by structures resembling the intercalary disks of vertebrate cardiac muscle.

INNERVATION AND ACTIVATION

There is an electrical potential across the surface membrane of a living, resting muscle fiber; the interior of the fiber is typically 30 to 70 mV electrically negative with respect to the extracellular solution. Nerve cells in the central nervous system send out long processes (motor axons) to the muscle fibers where they make specialized contacts termed synapses. A motor axon makes many synaptic contacts along the length of each muscle fiber that it innervates (multiterminal innervation), and a single muscle fiber may receive inputs from more than one motor axon (polyneuronal innervation). Impulses initiated in the central nervous system travel along the motor axons and cause the release of specific chemical signals (transmitters) from the motor axon terminals at the synapses. The transmitter released from the terminals of most motor axons leads to a reduction (depolarization) in the transmembrane potential of the muscle fiber in the vicinity of the nerve terminal. Muscle fiber depolarization initiates contraction of the fiber. Motor axons that depolarize muscle fibers and cause muscle contraction are called excitatory axons. Some axons, termed inhibitory axons, release transmitters that stabilize the transmembrane potential of the muscle fiber or even make it greater, thus antagonizing excitatory inputs. In addition to excitatory and inhibitory neural inputs, many muscles receive inputs from modulatory motor neurons, activity that releases chemicals that modify muscle performance, for example, increasing muscle force and work output or speeding relaxation.

Insects, like other arthropods, manage their muscles using relatively few motorneurons. Some major muscles, for example tymbal muscles of cicadas, are innervated by a single motor neuron. Many muscles receive 2 to 4 motor neurons. The largest number of motor neurons yet described for an insect muscle is 16, to the flexor muscle in the leg of a locust.

The processes linking membrane depolarization and contractile activation have been little studied in insect muscles, but the ultrastructure, biochemistry, and contractile performance of insect muscle are so similar to those of the far better studied frog, cat, and rodent muscles that one can predict with confidence that the basic principles worked out for vertebrate muscles apply to insects as well. The expected scheme is as

follows. Membrane depolarization spreads inwardly into the fiber along the T tubules. Depolarization of the T tubules, which are coupled to the SR through specialized junctions, leads to release of calcium from the SR. Released calcium reversibly binds to regulatory sites in the fibrils, turning on the contractile machinery. Relaxation occurs as the SR takes up the released calcium and reduces the calcium concentration in the cytoplasm below that needed for contractile activity.

MUSCLE MECHANICS

Muscle Force and Muscle Length

The muscle contraction initiated by a single stimulus, or by a single impulse in an innervating motor neuron, is termed a twitch; that evoked by repetitive input at a frequency high enough to maintain full activation of the muscle is termed a tetanus. A response in which a stimulated muscle develops force while held at constant length is called an isometric contraction. The isometric force generated by a muscle stimulated to contract in a tetanus is maximal at about the normal muscle length in the insect body and declines at longer and shorter lengths. The decline in force with increasing muscle length beyond the optimum is thought to be caused by a reduction in the overlap between the thick and the thin filaments and therefore in the number of myosin cross-bridges that can interact with the actin filaments. The decrease in force at short muscle lengths is probably a consequence of the thick filaments running into and being impeded by the Z disks, of collision of thin filaments in the middle of the sarcomere, and, at still shorter lengths, of overlap of thin filaments with portions of thick filaments of inappropriate polarity on the far side of the center of the sarcomere.

Some muscles in insects and elsewhere can shorten to a small fraction of their resting length, a response termed supercontraction. The capacity for supercontraction appears to involve modifications in the structure of the Z disk such that there is not a collision between the Z disks and the thick filaments at short muscle lengths. In the supercontracting muscles that have been examined, the Z disk becomes perforate at short muscle lengths and the thick filaments slide through the spaces in the disks.

The posterior, intersegmental, abdominal muscles of female locusts are of particular interest for the wide range of lengths over which they can operate. During oviposition, appendages on the end of the abdomen dig and pull the posterior abdomen down into a relatively deep hole. The intersegmental muscles become stretched to about nine times their resting length. During this stretch, called superextension, the Z disks become broken up into discontinuous, nonaligned elements to which the thin filaments are attached. Muscle contractility is not lost, and contraction of intersegmental muscles returns the abdomen to its normal length following oviposition. The latter part of the recovery may be supercontraction, for in the resting state the posterior intersegmental muscles are normally supercontracted, with thick filaments protruding through gaps in the Z disks.

Force, Shortening Velocity, and Power

There is an inverse relationship between the force on a muscle and the velocity with which it can shorten, a relationship conveniently expressed in a force–velocity plot (Fig. 6). To facilitate comparison of muscles of differing size, force in a force–velocity plot is usually expressed as stress (force per unit cross-sectional area) and shortening velocity as strain rate (shortening velocity per unit muscle length). Two points on a force–velocity plot are frequently used to characterize a muscle's contractile properties: the maximum isometric stress of the muscle (F_{max}, the intercept of the curve with the 0 velocity axis) and the maximum shortening velocity (V_{max}, the intercept of the curve with the 0 force axis). Values for the maximum force in insect muscles are generally 5 to 35 N cm^{-2} (N, Newton; 1 N is approximately the downward force exerted by a mass of 100 g in the gravitational field at the earth's surface). The few available measurements of the maximum shortening velocity for insect muscles, all from fast muscles, are on the order of 3 to 15 lengths/s.

The product of force and shortening velocity has dimensions of force × distance per time (work per time) and is the rate of doing work, i.e., the mechanical power output. The product of stress (force per area) and strain rate (shortening distance per second per unit muscle length) is the mechanical power per unit volume of muscle, which is readily convertible to power output per unit muscle mass. Thus each point on a force–velocity plot (or a plot of stress against strain rate) represents a power output. The power predicted from a force–velocity curve is the instantaneous power output. The peak instantaneous power is substantially greater than the sustainable power from a muscle, for during maintained activity a muscle goes through repeated contraction–relaxation cycles and therefore shortens and produces power for only part of the total time. For fast muscle, including fast insect muscles, the peak power output is 100 to 500 W kg^{-1} (1 W = 1 Joule s^{-1} = 1 N m s^{-1}).

Muscles in insects may be divided into synchronous muscles and asynchronous muscles on the basis of the relationship between the patterns of neural activation and of contraction (see below). The sustainable power available during repetitive, cyclic contraction has been determined for several synchronous insect muscles using the work loop approach, in which the muscle is subjected to length changes simulating those during normal activity and stimulated phasically during the length cycles. A plot of muscle force against muscle length for a full cycle produces a loop, the area of which is the net work output of the muscle for that cycle. The product of work per cycle and cycle frequency is the power output. The mechanical power available from synchronous flight muscles of several locusts and katydids and of a moth measured in this way ranges from 50 to 120 W kg^{-1} at normal operating temperature.

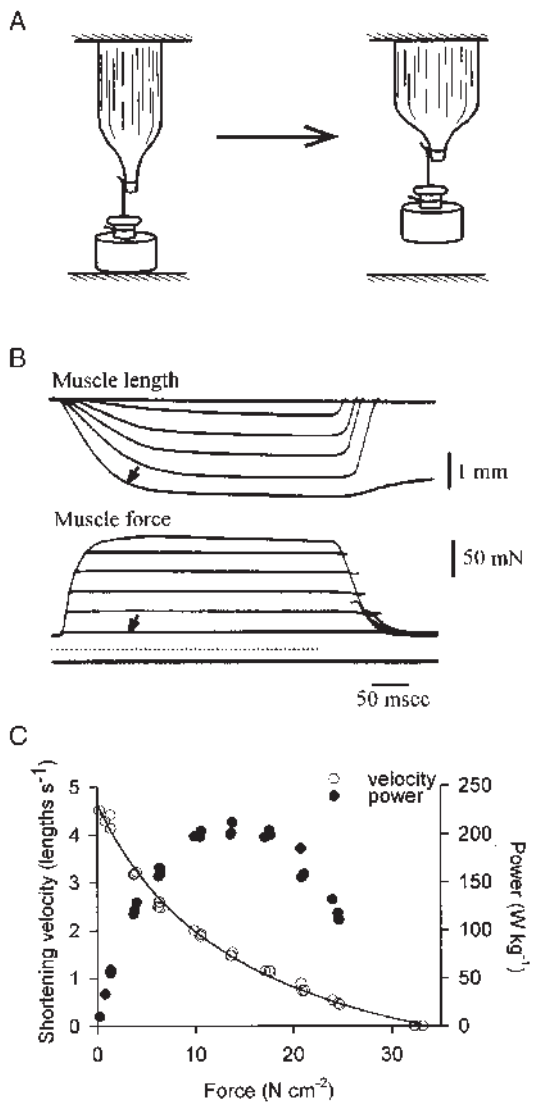

FIGURE 6 Relationships between muscle force, shortening velocity, and power output. (A) Elements of a method used to determine the relationship between force and shortening velocity. The muscle is attached to a load that is supported from below. When stimulated the muscle develops force without shortening until the force equals that of the load, following which the muscle shortens under constant load. (B) Results from an experiment examining force–velocity relations using a tettigoniid wing muscle. (Modified from Josephson, 1984, *J. Exp. Biol.* **108,** 77–96.) The lowest trace indicates the times at which the muscle was stimulated. The force trace marked by an arrow is the contraction with the smallest load of the series; the corresponding shortening trace, which has the shortest latency and the highest initial velocity, is similarly marked. (C) Force–velocity plot and a corresponding plot of power output for a wing muscle of the locust *Schistocerca americana,* 30°C (data provided by J. Malamud).

Asynchronous and Synchronous Muscles

Most insect muscles are like vertebrate skeletal muscles in that each contraction is initiated by depolarization of muscle fibers, and there is a 1:1 relationship between muscle electrical activity and muscle contraction. Such muscles may be termed synchronous muscles, reflecting the congruence between electrical and mechanical activity. The major flight muscles in several insect groups are different in that there is no synchrony between electrical and mechanical events. These muscles are known as asynchronous muscles. Neural input and fiber depolarization are needed to activate an asynchronous muscle, but when it is activated an asynchronous muscle can contract in an oscillatory manner if it is attached to a mechanically resonant load. The resonant loads for the flight muscles are the wings, which may be regarded as small, somewhat dampened tuning forks. The frequency of the oscillatory contraction is the mechanically resonant frequency of the load, which is greater than the neural input frequency required to keep the muscle fully activated. The contraction frequency of asynchronous wing muscles during flight is typically 3 to 10 times higher than the neural input frequency in each of the motorneurons activating the muscle. The main singing muscles in some but not all cicadas are asynchronous muscles; the resonant load here is the cuticular tymbal to which the muscle is attached and whose inward movement produces the sound pulses.

The features of asynchronous muscle that allow oscillatory contraction are stretch activation and shortening deactivation. When allowed to shorten rapidly an asynchronous muscle becomes deactivated, and while deactivated it can be stretched out to its original length, developing less force than it did while shortening. Stretching the muscle, in turn, reactivates it. Because of shortening deactivation, less work is required to restretch an asynchronous muscle than is produced by the muscle during shortening, and there is net work output when the muscle undergoes a shortening–lengthening cycle. It is this net work that is available to drive the wings and power flight.

Asynchronous muscles occur in several of the most successful insect groups. They power flight in beetles, flies, bees, and wasps and many of the true bugs. The distribution of asynchronous muscles among insect taxa suggests that this mode of muscle control has evolved independently as many as 7 to 10 times. It is likely that asynchronous muscle has been favored by evolution because it is more powerful and more efficient than is synchronous muscle for operation at the high frequencies characteristic of insect flight. It is more powerful, in part, because asynchronous control does not demand rapidity in the rate at which muscle is turned on and off by neural input. High-frequency contraction is achieved without hypertrophy of the sarcoplasmic reticulum, leaving more room in muscle fibers for fibrils, which are the power-producing component. It is more efficient because a relatively low-frequency neural input is needed to maintain full activation, which reduces the amount of calcium that is released and re-bound during activity and the associated metabolic costs of calcium cycling.

Are Insect Muscles Unusual as Motors?

In the minds of many people, insects are extraordinary athletes. One sometimes hears it said that if a person were as strong as an insect, he or she could carry enormous weights or leap over tall buildings. Such assertions are largely based

TABLE I Contractile Properties of a Locust Flight Muscle (Metathoracic Second Tergocoxal Muscle of *S. americana*) and the Frog Sartorius Muscle

	Locust	Frog
Twitch time course, ms		
Rise time	20	21
Onset to 50% relaxation	39	52
Tension, N cm^{-2}		
Twitch	17	5
Tetanic	30	25
Maximum shortening velocity, length s^{-1}	4.1	6.4
Power, W kg^{-1}		
Peak instantaneous	150	250
Cyclic, sustained	48	50

Note. Values were collected at 25°C or adjusted to the expected value at 25°C from measurements made at 20 or 30°C using an assumed Q_{10} of 2.

Locust data are from Malamud, Mizisin, and Josephson (1988, *J. Comp. Physiol. A* **162,** 827–835) and Malamud, unpublished; frog data are from Renaud and Stevens (1981, *Am. J. Physiol.* **240,** R301–R309), Rome (1983, *Physiol. Zool.* **56,** 33–40), and Stevens (1988, *J. Muscle Res. Cell Motil.* **9,** 329–333).

on incorrect application of principles of scaling. Consider, for example, jumping ability. A 1-g locust can develop enough power to lift its 1 g of mass to a height of about 1 m. A 70-kg person can develop enough power in a jump to lift his or her 70 kg to a height of 1 m. The work done is 1 g m for the locust, 70 kg m for the person, and the power required per mass of animal is the same.

The most often studied and certainly the most completely analyzed muscle for any animal is the frog sartorius muscle. The most complete body of information on contractile properties for an insect muscle is probably for the wing muscles of locusts, both *Schistocerca gregaria* and *S. americana.* The frog sartorius muscle is not the strongest or the fastest vertebrate muscle known, but it is a good representative of a fast vertebrate muscle. Similarly locust flight muscles are neither the strongest nor the fastest insect muscles, but they are reasonable representatives of fast insect muscles. Some of the contractile properties of frog and locust muscle are tabulated in Table I. The vertebrate muscle and the insect muscle are surprisingly similar in many of their contractile properties. The muscles of insects share the same capacities and are subject to the same limitations as are muscles elsewhere throughout the animal kingdom.

See Also the Following Articles

Flight • Walking and Jumping

Further Reading

Dudley, R. (2000). "The Biomechanics of Insect Flight: Form, Function, Evolution." Princeton University Press, Princeton, NJ.

Hoyle, G. (1983). "Muscles and Their Neural Control." Wiley, New York.

Josephson, R. K. (1975). Extensive and intensive factors determining the performance of striated muscle. *J. Exp. Zool.* **194,** 135–154.

Josephson, R. K. (1993). Contraction dynamics and power output of skeletal muscle. *Annu. Rev. Physiol.* **55,** 527–546.

Josephson, R. K., Malamud, J. G. and Stokes, D. R. (2000). Asynchronous muscle: A primer. *J. Exp. Biol.* **203,** 2713–2722.

Smith, D. S. (1984). The structure of insect muscles. *In* "Insect Ultrastructure" (R. C. King and H. Akai, eds.), Vol. 2, pp. 111–150. Plenum Press, New York.

Museums and Display Collections

Gordon M. Nishida
University of California, Berkeley

Insect collections are often oversimplified as either a precious and essential, rapidly developing treasure or a musty, dusty, moribund assemblage of archaic specimens. Those who now prefer the latter characterization have not been paying attention. Entomological collections and their curators are experiencing a steady, if not spectacular, evolution in the way collections operate, and perhaps the best is yet to come.

The term "insect collections" is often loosely used, and these collections usually contain other arthropods, including arachnids (e.g., spiders, mites), myriopods (e.g., centipedes, millipedes), and terrestrial crustaceans (e.g., sowbugs and pillbugs, amphipods). So most entomological collections are actually arthropod collections.

Collections of insects and related arthropods constitute an enormous resource for biological information and are an irreplaceable tapestry documenting Earth's entomological natural history. The numbers of specimens in collections are staggering, as is the amount of work yet to be accomplished just describing and cataloging millions of new species. Despite the revolutionary changes energizing these collections, still daunting are the myriads of problems, particularly financial, that continue to debilitate many collections' operations.

SCOPE OF THE WORLD'S INSECT COLLECTIONS

Insects are the largest, most diverse group of organisms in the world. Over 900,000 species have been described, and current estimates on the number still without names range as high as 30 million. That theoretical number is based on work in the New World tropics. However, more recent work testing that hypothesis, including the Old World tropics and measuring host specificity, now suggests a more modest but still astounding 4 to 6 million unnamed species. Reflecting the number of species potentially involved, insect collections maintain a large number of specimens. Based on numbers and extrapolations primarily from Arnett, Samuelson, and Nishida and adjusted for the last 8 years using an annual growth rate of 2% (average over a 20-year period), there are

conservatively 724 million specimens in entomological collections worldwide. This figure does not include the unknown number of specimens in private collections, nor does it include a significant number of unprocessed specimens. In 1991 Miller reported the numbers of specimens for entomological collections in the United States and Canada and also included figures on unprocessed specimens (i.e., backlog). In 1976 reports indicated that 26% of specimens were reported unprocessed, increasing to 28% in 1981 and 30% in 1986. If these percentages are extrapolated worldwide and added to the processed specimens, there may be nearly a billion arthropod specimens housed in entomological collections.

As might be expected, a great many collections house the rather large number of specimens that have been amassed. The insects and spider collections of one Web site (http://www.bishopmuseum.org/bishop/ento/codens-inst.html) lists 904 institutional or organizational insect collections. Despite the plethora of collections, the largest house a disproportionate number of the specimens. Table I lists the largest entomological collections, that is, those reporting over 5 million specimens in their holdings. The largest collections are found in Europe, North America, Australia, and New Zealand. The collections' locations are likely influenced by the historical origins of insect collections, centers for entomological research, cultural interests, and economic restrictions. These collections do not coincide geographically with the regions of highest insect biodiversity, which tend to be in the tropics.

The total number of specimens may appear excessive given the number of species described, but the total figure includes many undescribed species. Also, long series are required to study and account for the morphological, geographic, and seasonal variability within a species. The specimens are also widely spread throughout many collections, enabling multiple centers for study and providing insurance against loss of species representation in case of natural or man-made disaster.

TABLE I Largest Entomological Collections[a, b]

Collection	Country	Number of specimens
Muséum d'Histoire Naturelle	France (Paris)	30,000,000[b]
The Natural History Museum	United Kingdom (London)	30,000,000
Smithsonian Institution	United States (Washington, D.C.)	30,000,000
Zoologische Staatssammlung	Germany (Munich)	16,566,000
American Museum of Natural History	United States (New York)	16,204,000
Canadian National Collection	Canada (Ottawa)	15,000,000
Alexander Koenig Zoological Museum	Germany (Bonn)	14,000,000
Bernice P. Bishop Museum	United States (Hawaii, Honolulu)	13,250,000
Musée Royal de I'Afrique Centrale	Belgium (Tervuren)	10,510,000
Australian National Insect Collection	Australia (Canberra)	10,000,000
Museum für Naturlunde der Humboldt University	Germany (Berlin)	10,000,000
Zoologisch Museum, Universiteit van Amsterdam	Netherlands (Amsterdam)	9,685,000
Field Museum of Natural History	United States (Chicago)	9,000,000
Institut Royal des Sciences Naturelles	Belgium (Brussels)	8,000,000
Museum of Comparative Zoology	United States (Massachusetts, Cambridge)	7,601,000
California Academy of Sciences	United States (San Francisco)	7,000,000
Hungarian Natural History Museum	Hungary (Budapest)	6,700,000
New Zealand Arthropod Collection	New Zealand (Auckland)	6,560,000
Naturhistoriska Riksmuseet	Sweden (Stockholm)	6,500,000
Finnish Museum of Natural History	Finland (Helsinki)	6,500,000
Florida State Collection of Arthropods	United States (Florida, Gainesville)	6,500,000
Bohart Museum of Entomology	United States (California, Davis)	6,241,000
Naturhistorisches Museum Wien	Austria (Vienna)	6,000,000
National Museum of Natural History	Bulgaria (Sofia)	6,000,000
Zoological Museum, University of Copenhagen	Denmark (Copenhagen)	6,000,000
Illinois Natural History Survey	United States (Illinois, Champaign)	6,000,000
Carnegie Museum of Natural History	United States (Pennsylvania, Pittsburgh)	5,500,000
National Natuurhistorische Museum	Netherlands (Leiden)	5,200,000
South African National Collection of Insects	South Africa (Pretoria)	5,000,000
Staatliches Museum für Tierkunde	Germany (Dresden)	5,000,000
Los Angeles County Museum	United States (Los Angeles)	5,000,000

[a]From Arnett, R. A., Jr., Samuelson, G. A., and Nishida, G. M. (1993). "The Insect and Spider Collections of the World." 2nd ed. Sandhill Crane Press, Gainesville, FL and Miller, S. E. (1991). Entomological collections in the United States and Canada. *Am. Entomol.* Summer, 77–84.

[b]Collections reporting over 5 million specimens; does not include large Russian collections such as Leningrad whose holdings are not available.

[c]In 1993 Arnett, Samuelson, and Nishida reported an estimate of 100 million. However, this is generally considered to be in error, hence the more conservative figure provided here.

BEGINNINGS OF INSECT COLLECTIONS

The beginningss of insect collections are lost in unrecorded history. The Chinese used silkworms as early as 4700 B.C., honey bees by the fifth century, and scale insects by the 13th century. A treatise describing insects and their pharmaceutical properties had been published in China by at least A.D. 200. Cuneiform texts found in Mesopotamia dating to earlier than 669–626 B.C. contain systematically arranged names of insects. Aristotle (384–322 B.C.) studied insects and taught entomology. In A.D. 77 Pliny produced an encyclopedia that included entries on insects. Insect collections were no doubt made during antiquity, but no record of them has been found.

The early periods saw interest in insects primarily for their practical use or as pests. European explorations in the 15th and 16th centuries opened up shipping lanes for commerce and trade. The explorers and adventurers brought back hordes of items including insect novelties, piquing the interest of many. The first entomological collections were included in the cabinets of "curiosities" assembled by wealthy Renaissance families to show to friends and associates. These cabinets were to eventually lead to modern natural history museums.

The first compound microscopes were made by Hans and Zacharias Jansen in Holland in 1590. This major technological advance, coupled with the improvements made by Antony van Leeuwenhoek in the late 1600s, permitted the observation of tiny insects and their minute parts, advancing their study.

As the number of curiosity cabinets continued to grow in the 16th and 17th centuries, a means to organize their contents became necessary. John Ray (1628–1705) attempted a classification of insects in 1705, which was published posthumously in 1710. Carolus Linnaeus (Carl von Linné) (1707–1778) published the first edition of his *Systema Naturae* in 1735, the 10th edition, published in 1758, became the basis for modern insect classification.

SOURCES OF EARLY SPECIMENS

Linnaeus surrounded himself with students not only to carry on his work but also to provide specimens for study. Linnaeus was particularly interested in obtaining species of practical use, and he arranged for his students to go on voyages and encouraged them to send material from their travels. From 1744 to 1796, students sent back to Linnaeus and his associates material not only from Europe but from the Middle East, Africa, India, Asia, South America, Africa, and the Pacific (Cook's first two voyages).

Some of these students of Linnaeus eventually produced works of their own: for example, Fredrik Hasselquist's *Iter Palaestinum eller Resa til Heliga Landet* in 1762 and Pehr Forsskål's *Descriptiones animalium* in 1775. Other Linnaean students, for example, Johann Christian Fabricius (1745–1808), were extremely productive in discovering and describing new species, continuing the Linnaean traditions and fostering this era of cataloging.

Wealthy scholars and others amassed large collections that eventually wound up in institutional collections. Catherine the Great started Peter Simon Pallas on collecting and exploring travels throughout the Russian Empire between 1767 and 1810. A. M. F. J. Palisot de Beauvois described insects from Africa and America collected on his own travels from 1781 to 1797. Guillaume Antoine Olivier, sent on expedition to Turkey, Asia Minor, Persia, Egypt and the Mediterranean islands (1792–1798), later became a patron for other naturalists. Pierre François Marie Auguste Dejean (1780–1845), a soldier of fortune, collected in Austria and by exchange or purchase amassed the greatest collection of beetles in the world at that time. Thomas Say (1787–1834), the father of American entomology was appointed naturalist for Long's expeditions to the Far West and visited the Rocky Mountains and the sources of the St. Peters River in 1823. Victor Ivanovich Motschulsky (1810–1871) was a Russian military officer who traveled through Europe, the Caucasus, Siberia, the Kirghiz steppes, Egypt, India, the United States, and Panama. Thomas de Gray, Lord Walsingham (1843–1919), was an English nobleman wealthy enough to travel extensively (United States, North Africa, Europe) and also purchase specimens.

The 19th century was a fertile time for voyages and expeditions. Alexander von Humboldt visited the Spanish colonies of the American tropics between 1799 and 1804. Otto von Kotzebue sailed around the world from 1815 to 1818. On board was J. Friedrich von Eschscholtz who collected in California, Hawaii, the Philippines, Brazil, Chile, and other places. On a later voyage, again with Kotzebue, Eschscholtz amassed a large collection from the tropics, California, and Sitka (Alaska). Other important voyages for entomological specimens included those of the *Astrolabe* (1826–1829), the *Astrolabe* and the *Zelee* (1837–1840), the Swedish frigate *Eugenie* (1851–1853), and the Austrian frigate *Novara* (1857–1859). The United States Exploring Expedition (1838–1842) visited Madeira, Brazil, Chile, California, Oregon, Pacific Islands (including Hawaii, Australia, the Philippines, and Singapore), South Africa, and St. Helena. Sources of entomological collections mirrored the spread of empire and the pursuit of national interests.

Rather than circumnavigating, some targeted specific locations. For example, Henry Walter Bates spent the years 1851 to 1870 in South America, mostly in the Amazon, collecting and sending material back to England. Giacomo Doria (1840–1913), who founded the Genoa Museum, funded expeditions to areas particularly rich in insect diversity (Africa, Southeast Asia, New Guinea, Indian Ocean islands). Others such as Lionel Walter Rothschild (1868–1937) focused on specific groups. Lord Rothschild concentrated on butterflies and moths and funded expeditions to the far corners of the world; in 50 years' time he amassed the greatest personal collection ever (including 2.25 million butterflies and moths).

The amount of material grew exponentially, and as the Linnaean system became entrenched, secure and centrally

located places to store the reference collections and unprocessed materials were sought.

DEVELOPMENT OF FORMAL COLLECTIONS

The first natural history museum probably was that of Conrad Gessner a scholar of mid-16th-century Zurich. Very few present-day natural history museums were established before the mid-18th century. The Muséum National d'Histoire Naturelle in Paris was established in 1635 and was the first natural history museum established in the form we recognize today. Others were the Staatliches Museum für Tierkunde in Dresden established in 1650, the Zoologiceskii Instituti Zoologiceskii Muzei in Leningrad in 1727, the Zoologiska Museet in Lund in 1735, the Naturhistorisches Museum in Austria in 1748, and the British Museum in 1753. Scientific academies, beginning with the Accademia dei Lincei in Rome in 1603, fostered and often housed early collections. The Royal Society in London was founded in 1662, and the Académie Royale des Sciences of Paris in 1666. The Academy of Natural Sciences in Philadelphia, founded in 1812, is the oldest North American academy. The academies were later augmented by natural history societies that often performed similar functions. The First Aurelian Society (early insect collectors were known as "Aurelians") was founded in London in 1745. The Entomological Society of Philadelphia, established in 1859, was the first American entomological society.

The great explosion of natural history museums occurred in the latter part of 18th century and into the 19th century, with the continuation of exploration and collecting. Coincidentally, this proliferation of museum collections coincided with the earliest use of persistent poisons such as arsenic to protect biological specimens from damage by the pests that had destroyed many earlier collections (a Western discovery, but the Chinese had written about the use of arsenic and mercury for control of human parasites in A.D. 100–200). Perhaps the concentration of the entomological collections in temperate Europe and North America was a result not only of European influence but also of the climate, which likely was less favorable to potential pests of museum specimens than in the warmer regions of the world.

Growth continued through the 20th century with the exception of periods of global conflict. The 19th and 20th centuries saw increasing participation of institutions and government in organizing and funding expeditions. Expatriates and professionals on foreign assignment were also great sources of collections. The 20th century later saw a focus on regional faunas, opportunistic trips, and taxon-based initiatives. Institutions became less involved with collecting efforts and individuals increased their efforts, particularly to aid their research goals. Toward the end of the 1900s, cooperative efforts returned as the cost of fieldwork increased. Large national inventories such as the Instituto Nacional de Biodiversidad (INBio) established in 1989 in Costa Rica were undertaken, and ATBIs (all taxa biological inventories) were begun to inventory areas with a combination of high biodiversity and high threat of loss due to that biodiversity.

As the great expeditions filled the collections' coffers and the continuous additions provided more than enough work for taxonomists, collections also needed technological support to improve storage and retrieval capabilities.

TECHNOLOGICAL ADVANCES IN METHODOLOGIES USED BY COLLECTIONS

Technological advances, from the microscope to finding effective pesticides, had and still have a profound influence on the development of entomological collections. Although the basic methods of collections and specimen preparation have not changed radically, new innovations usually made the process quicker, more efficient, and more inclusive. Steady changes have taken place in the past three decades, and in retrospect, a major revolution has occurred.

Many innovations have been implemented, including the use of glass-topped drawers for specimen storage rather than just simple boxes. It is likely that insects were placed in containers to take advantage of the efficacy of pesticides by maintaining the specimens in a closed environment. Dried specimens are still prepared with pins, and paper labels are affixed to the pin beneath the specimen. However, the materials used have been much improved. Insect pins are of higher quality and are less likely to corrode.

Along with the advances in technology, the last few decades have seen a concerted effort to introduce materials conservation techniques into collections management. Recent advances in storage technology in insect collections include the development of the unit tray system. This is a system of topless boxes made of cardboard and with a material that functions as a pinning base at the bottom of the box. This innovation permits the rapid rearrangement of the collections as new studies modify the organization of a group. Placing like elements in a single unit tray (e.g., specimens of a species all collected in one area) also permits the use of a header card, or tray identifier, that permits rapid recognition and retrieval of information. The material first selected for tray bottoms was cork. Within several decades, however, cork was found to be unacceptable because its acidic nature affected the pins and labels and even perhaps the specimens. A new material was searched for, and most collections have settled on a cross-linked polyethylene product that is inert and pliable.

In the late 19th century, new methods helped streamline the papermaking process. Unfortunately, the new papers are slightly more acidic and tended to break down faster, with the ironic result that older 17th and 18th-century labels are more permanent. Labels have changed in both substance and content. Some early collections were not labeled at all to show the origin of specimens. As systematic work progressed and some species were found to be restricted in their distribution, labels were added. These labels were laboriously handwritten in ink and often would indicate only a country, an island, or

a region. As even more species of insects were discovered, classifications became larger, and more specimens were collected, labels became more specific, adding ecological data, collecting information, and details of the collecting locality. With the development of GPS (global positioning systems), collecting location coordinates are routinely being added to label data.

Hand-printed labels gave way to mass-printed forms. Typesetting took time, and so photoreduction methods were used to speed the process. Since the 1980s, the desktop computer has become a mainstay in collections work. Initial efforts to use the computer to generate labels were restricted to providing printed copies since the early printers did not have a font size small enough to fit all the necessary information onto the tiny specimen labels. As laserwriters developed, collections personnel began experimenting with producing labels on them. The first efforts were encouraging until it was discovered that the toner (laser cartridge "ink") did not stick to the paper used for labels in fluid (ethanol) collections. At this time, inkjet printers seem to be a useful alternative.

The last two examples serve to show why collections often embrace technology slowly. Yet despite the pitfalls inherent in adopting new technologies, most collections today are in much better shape for long-term survival than they were before. Other advances in storage and maintenance include new or retrofitted buildings for many collections, use of climate control to reduce fluctuations in temperature and humidity, and installation of compactor storage systems to make better use of space and improve access to collections. Repellents used in collections that are potential human health risks, such as naphthalene (usage introduced in the late 19th century) and paradichlorobenzene, are being replaced with freezing procedures and integrated pest management monitoring techniques such as sticky traps to intercept possible problems sooner.

Collecting methods have improved and can easily overwhelm the preparation capabilities of most present-day collection staffs. Mass collecting techniques took a great leap forward with the invention of a flight intercept trap by René Malaise in 1937 (prototyped in 1933). The Malaise trap is a tentlike structure placed in a position to intercept flying insects and have them self-collect in strategically placed containers. Today's arsenal of passive traps includes yellow pan traps, pitfall traps, Berlese samplers, and innumerable other specialty traps and methods, including modifications of the original Malaise trap. For those unwilling to wait for their specimens to come to them, proprietary devices (e.g., D-Vac) allow vacuuming of vegetation. Even more efficient is the use of pyrethrum fog to assemble vast quantities of specimens from tree canopies.

As collections techniques evolved, methods of studying and interpreting species have influenced collections also.

CHANGING ROLES OF COLLECTIONS

Modern entomological collections are rooted in Linnaean classification. The century following Linnaeus was devoted to describing and cataloging the massive amount of material gathered, but there were too many species and not enough taxonomists. Other disciplines began studies of arthropods, and many researchers had to develop taxonomic expertise on their own to have names for the organisms they were studying.

Around the mid-19th century, large-scale agriculture and expanding horticultural efforts underscored the importance of insects as pests and opened up an entirely new area for collections, namely, the role of identifying insect pests and establishing the authenticity of such identification. The use of parasitoid and predatory insects for biological control began a new type of biological exploration, sending professional entomologists around the world to find control agents, which were subsequently reared and released. Governments established sections of entomology in agricultural divisions and often associated them with national collections such as the U.S. Department of Agriculture and the Smithsonian Institution or the Commonwealth Agricultural Bureaux and the British Museum of Natural History.

Between 1897 and 1900, Ronald Ross and Patrick Manson discovered and experimentally proved that mosquitoes act as vectors in transmitting malaria, launching yet another era in collections development: the collection and study of medically important arthropods. Identification and research units were often established within military units or centers, such as the Walter Reed Army Medical Center in the United States.

Evolutionary biology has had a profound influence on the development of insect collections. Both Charles Darwin and Alfred Russel Wallace had been influenced by insects when they proposed their theory of natural selection in 1858. Further evidence for evolution was sought within insects, enhancing collections in the process. In the past 50 years, areas such as genetics, population ecology, and even bioprospecting have used existing insect collections and developed additional collections as adjuncts to their research. Insect collections have provided support for unexpected areas such as medical forensics. Within the last two decades, molecular biology has been asserting its influence on entomological collections, similarly using available specimens and gathering more, making use of DNA evidence to establish better understanding of species relationships.

New approaches to taxonomic studies have appeared in the decades since 1980. Phenetics (sometimes referred to as numerical taxonomy) bases classifications on overall similarities. Cladistics places emphasis for classification on branching points of a phylogeny. Evolutionary taxonomy adds degree of similarity to evolutionary origin. The advent of computers has helped the establishment of these systems by facilitating the manipulation of data. Ernst Mayr perhaps summed up the current status in insect classification best: "Taxonomy has been more active, more in ferment, in the last 50 years, than ever before in its history." Today these words are even more applicable.

As these new sciences developed, positions in classical taxonomy dropped steadily, many workers transferred to the

newer "cutting edge" sciences. An increase in efforts to stimulate appreciation of entomology was one of the responses to the perceived loss of positions and funding.

COLLECTIONS ON DISPLAY

Early collections were displayed in their entirety. In 1864 John Edward Gray of the British Museum proposed to store the collections away from public view, and Richard Owen eventually produced an "index collection" that became the model for display of collections specimens. This began the period of exhibit collections, with only a small portion of the material placed on view and the remaining material devoted to research. With little variation, this was the extent of entomological specimen exhibition for the next century.

Toward the latter half of the 20th century, new entomological exhibits with educational themes began appearing in an effort to interest the public. As operating costs rose and income stagnated, museums began adding an entertainment component to their exhibits in hopes of attracting more visitors, and to this end they allied themselves more closely with educational institutions. Exhibitors and collections staff quickly discovered that people respond positively to the display of live arthropods.

In 1976 the first insect zoos (e.g., Smithsonian Insect Zoo) appeared in North America. These living exhibits were extremely successful, spawning dozens of replicates. The first butterfly house debuted in 1976 on the British island of Guernsey; over a hundred soon followed throughout the world. Determining the exact number of living insect exhibits throughout the world is difficult because many zoological parks have an insect component that is not readily apparent. Both conventional and insect zoos have live specimens on display, but whereas the older style zoos maintain special enclosures and habitats for the animals, butterfly houses permit visitors to walk through the enclosure. These two innovations sparked a flurry of activity in arthropod husbandry. An indication of the popularity of living collections just in North America is the number (328) of organizations represented at the year 2000 Invertebrates in Captivity Conference sponsored by SASI (Sonoran Arthropod Studies Institute).

SHIFTING FUNDING

As museums began appearing in the late 1700s and 1800s, funding for collections shifted from the province of wealthy patrons, who were largely replaced by institutions supported by government or public funds. As governments centralized and became more active, particularly in the areas of medicine and agriculture, funding support shifted. Beginning in the late 1800s, agriculture became a major source of economic support, and indeed, modern land grant college collections owe their existence to agricultural funding. Today most large museums in the United States survive on a mix of revenue from endowment, gifts, income from visitors, grants from government or private sources, and direct support from government. In the rest of the world, funding for collections more likely comes from government sources, with less dependence on private sources. Traditional sources of funding such as agriculture or military research have waned recently, as has government spending in general. Though it is likely that funding sources such as biodiversity or conservation have not yet realized their full potential, many insect collections are now seeking supporting funds from private foundations as well as continued support from customary sources such as the National Science Foundation.

FUTURE OF ENTOMOLOGICAL COLLECTIONS

Inventory of the world's biodiversity is far from complete, and present work is being accomplished in the face of extraordinarily high rates of extinction. Despite the large holdings presently in collections, many more species probably have not been collected yet, or even discovered. Financing for large-scale inventory projects has been minimal at best. However, a revival of taxonomy is under way, with funding coming from more diverse sources than before and with entomological collections reasserting their roles. Interest remains in endangered species and interest in alien species and their impact is growing rapidly.

Collections have always been in the information business. Specimens are about data, and collections are in the business of brokering information and storing and retrieving data. Present information technologies are revolutionizing data-sharing capabilities, and many collections are furiously converting their older information retrieval systems (card files, specimen data, etc.) into electronic systems that can be used internally and shared internationally. Bioinformatics tools such as electronic catalogs [e.g., Biosystematic Database of World Diptera (http://www.sel.barc.usda.gov/diptera/biosys.htm) or Orthoptera Species File Online (http://viceroy.eeb.uconn.edu/Orthoptera)] are rapidly coming online. Large clearinghouses for biological information such as the National Biological Information Infrastructure (NBII) are available on the World Wide Web.

The global availability of this information will provide more opportunities for cooperative ventures and at the very least will make the process of scientific inquiry a lot less time-consuming and less costly. Perhaps even more intriguing is the potential to overlay information from other disciplines—for example, adding geographic capabilities to enhance understanding of species distributions, or plant data to further understand host associations—and provide a historical perspective to boot. Ambitious systems such as Species 2000 and the Global Biodiversity Information Facility will permit this type of knowledge synthesis and interoperability. Perhaps an opportunity to better understand human impact on the world's ecosystems from an arthropod perspective is finally at hand. This is an exciting time for insect museums.

See Also the Following Articles

Collection and Preservation • Entomological Societies • History of Entomology • Insect Zoos • Photography of Insects • Teaching Resources

Further Reading

American Association for the Advancement of Science. (2000). Bioinformatics for biodiversity: A Web registry. *Science* **289,** 2229–2440.

Arnett, R. A., Jr., Samuelson, G. A., and Nishida, G. M. (1993). "The Insect and Spider Collections of the World." 2nd ed. Sandhill Crane Press, Gainesville, FL.

Duckworth, W. D., Genoways, H. H., and Rose, C. L. (1993). "Preserving Natural Science Collections: Chronicle of Our Environmental Heritage." National Institute for the Conservation of Cultural Property, Washington, DC.

Ghiselin, M. T., Leviton, A. E. (eds.) (2000). "Cultures and Institutions of Natural History." Memoir 25. California Academy of Sciences, San Francisco.

Gilbert, P., and Hamilton, C. J. (1983). "Entomology. A Guide to Information Sources." Mansell, London.

Mayr, E. (1988). Recent historical developments. *In* "Prospects in Systematics" (D. L. Hawksworth, ed.), pp. 31–43. Clarendon Press, Oxford, U.K.

Miller, S. E. (1991). Entomological collections in the United States and Canada. *Am. Entomol.* Summer, 77–84.

Smith, R. F., Mittler, T. E., and C.N. Smith, (C. N. eds.), (1973). "History of Entomology." Annual Reviews, Palo Alto, CA.

N

Neosomy

Frank J. Radovsky
Oregon State University

Neosomy in an arthropod is a remarkable enlargement or formation of new external structure, or both, resulting from the secretion of new cuticle unrelated to a molt, during an active instar or in the adult. It is distinct from the more modest addition of cuticle that contributes to intermolt growth of many immature insects (notably larvae of Endopterygota) and that can be detected in some adult insects. Neosomy usually is associated with symbiosis and often with tachygenesis. *Physogastry* has been used overlappingly, but this term etymologically indicates abdominal swelling, and it usually has been defined as distension and not in relation to cuticular growth.

EXAMPLES

Neosomy is widely present in the Acari, including some parasitic larval mites and all stages (larva, nymph, adult female) of hard ticks (Ixodidae). Female ixodid ticks generally increase in volume about 100× or more, after first doubling cuticular thickness during the principal time they are attached to the host. In the chigger genus *Vatacarus* taken from the lungs of sea snakes, larval volume increases 1500× or more and neosomules (the new external structures) form as papillae that aid worm-like movement. All feeding is larval, and the adult casts larval and nymphal exuviae together when it emerges. In the Crustacea, some ectoparasitic copepods are also neosomatic.

Neosomy is seen in adults, primarily females, of some holometabolous insects. It occurs in females of four families of fleas. For example, *Tunga monositus,* embedded in the skin of rodents, grows more than 1000× until it no longer has any surface resemblance to a flea; addition of cuticle is centered in the second abdominal segment. *Tunga penetrans* (Fig. 1A), in mammals, including humans, increases in size less than other *Tunga.* Flies of the genus *Ascodipteron* in bats develop much as *Tunga* do.

Some termite- and ant-associated beetles and flies are neosomatic. Social parasites among staphylinid beetles mimic termites in order to be accepted by the colony. The termite-mimicking cuticle that grows from the abdomen of an initially normal-looking adult beetle has paired "legs" and "antennae" (Fig. 1B). Neosomatic cuticle usually grows in soft areas between sclerites, but it can also involve sclerotized parts in these beetles. Many queen termites, sometimes >12.5 cm in length, and some queen ants are neosomatic.

FUNCTIONAL SIGNIFICANCE

Combination of neosomy with symbiosis results from the abundance of food provided by a host or host colony. Neosomy in a termite or ant queen suggests that, although the association is intraspecific and hence not symbiotic, the

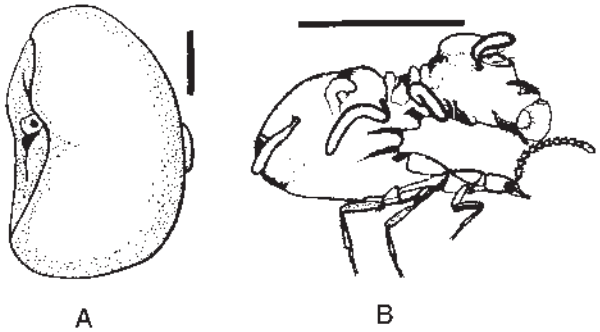

FIGURE 1 Female neosome, lateral view (scale bar, 1 mm). (A) *T. penetrans,* with part of head visible on left. (Reproduced, with permission, from Audy *et al.,* 1972.) (B) *Coatonachthodes ovambolandicus.* (Reproduced, with permission, from Kistner, 1979.)

queen's relationship to the colony as a whole parallels symbiosis functionally. A female ixodid tick or a queen termite, for example, has unusual access to food and produces many eggs compared with related groups.

Tachygenesis has a comparable effect in sheltering the species from risks. *T. penetrans* has two rather than the usual three instars of fleas. The more advanced *T. monositus* does not feed as a larva, and all stages except the female in the host are sheltered in a rodent burrow, nonfeeding, and quiescent (except the larva spinning its cocoon and the male mating). The mimicry seen in termite-associated beetles presumably could not be achieved other than by neosomy.

See Also the Following Article

Mites

Further Reading

Audy, J. R., Radovsky, F. J., and Vercammen-Grandjean, P. H. (1972). Neosomy: Radical intrastadial metamorphosis associated with arthropod symbioses. *J. Med. Entomol.* **9,** 487–494.

Kistner, D. H. (1979). Social and evolutionary significance of social insect symbionts. *In* "Social Insects" (H. R. Hermann, ed.), Vol. 1. Academic Press, New York.

Nijhout, H. F. (1994). "Insect Hormones." Princeton University Press, Princeton, NJ.

Rothschild, M. (1988). Giant polyploid cells in *Tunga monositus* (Siphonaptera: Tungidae). *Syst. Assoc. Spec.* **37,** 313–323.

Williams, C.M. (1980). Growth in insects. *In* "Insect Biology in the Future" (M. Locke and D. S. Smith, eds.). Academic Press, New York.

Neotropical African Bees

Orley R. Taylor
University of Kansas

Neotropical African bees are derived from an intentional introduction of bees from South Africa into Brazil. These bees are famous for their stinging behavior and they are known as "killer" or "Africanized" bees in the media.

The introduction from South Africa into Brazil of a subspecies of the common honey bee, *Apis mellifera,* resulted in a spectacular biological invasion that has had profound effects on agriculture, beekeeping, and human and animal health. Establishment of the African bees led to hybridization with bees of European subspecies maintained by beekeepers for honey production and pollination. The resulting apiary bees became extremely defensive, and honey production declined as beekeepers abandoned beekeeping. Stinging incidents that resulted in human and animal deaths soon led the press and the public to refer to these invaders as "killer bees." The perception that the feral African-derived bees were of hybrid origin led to the term "Africanized bees," which is a misnomer.

PREVIOUS INTRODUCTIONS

Honey bees *(A. mellifera)* are native to the Old World, and all honey bees present in the Americas are descendants of bees introduced from Europe, Africa, and the Middle East. Managed honey bees in the Americas are derived from at least five introduced European honey bee (EHB) subspecies belonging to two major lineages of *A. mellifera* (west European bees, *A. m. mellifera* and *iberica;* eastern European bees, *A. m. ligustica, carnica,* and *caucasica*). These temperate subspecies are notably successful in other temperate regions (e.g., Australia) and have been used with some success in tropical areas. However, they are not well adapted to tropical conditions and did not establish large self-sustaining feral populations in the tropics as they have done in subtropical and temperate regions. In many tropical areas, European bees would not persist without human assistance and, in most habitats (with the exception of regions with pronounced dry seasons) honey production was marginal. In 1956, to improve honey production in Brazil, Warwick Kerr intentionally introduced to Brazil *A. m. scutellata,* a subspecies belonging to the African lineage of *A. mellifera.* The introduced subspecies became established in southern Brazil in early 1957 and, because this bee was well adapted to tropical conditions, a large feral population soon developed and began to spread at rates of 100 to 300 miles per year. No other invading species has expanded so rapidly into new habitats.

DEFENSIVE BEHAVIOR

The Neotropical African honey bees (AHBs) are well known for their defensive behavior, and the deaths of hundreds, perhaps thousands, of people (and certainly thousands of domestic animals) have been attributed to these insects. These bees are not always defensive, but under certain conditions they will attack people and animals near their nests in massive numbers, inflicting hundreds and even thousands of stings. The venom (per bee) of AHBs is less toxic than that of EHBs; nevertheless, human victims of massive stinging require immediate medical attention to minimize lysis of blood cells, breakdown of muscle tissue, and kidney damage, which can result in acute kidney and multiple organ failure.

IMPACT ON BEEKEEPING

Beekeepers accustomed to dealing with relatively gentle and manageable EHBs were unable to adapt to the defensive behavior as AHBs advanced through the Americas, and many abandoned beekeeping. Honey production declined and many countries became honey importers rather than exporters. For example, in the Yucatan peninsula of Mexico, an area of intense beekeeping with both EHBs and native stingless bees (*Melipona* and *Trigona*), both types of beekeeping declined precipitously following the arrival of AHBs. In much of the Americas, as new beekeepers adapted to AHBs, beekeeping and honey production recovered, most notably in Brazil.

AHBs are now used for honey production in many areas of the Americas where EHBs were ineffective, and it appears that Kerr's goal of improving honey production in the New World tropics will be realized.

DIFFERENCES BETWEEN AFRICAN AND EUROPEAN HONEY BEES

Honey bees subspecies from tropical and temperate regions have evolved adaptations that are suitable for their respective environments. AHBs are smaller, and they have higher metabolic rates, more rapid development, reduced longevity, smaller nest sizes, greater brood production, and lower honey storage than EHBs. These traits combine to limit the ability of AHBs to overwinter in areas where the interval between first and last frost is longer than 3.5 months. In contrast, feral EHBs can exist in areas where this winter interval is 6 months. Adaptations to tropical conditions that give AHBs advantages include higher rates of swarming (reproduction) and the ability to abandon the nest (abscond) and move to new habitats under unfavorable conditions. Also affording AHBs a distinct competitive advantage in the tropics are these insects' abilities to find pollen and nectar, to increase their brood production under conditions in which EHBs are unable to do so, and to mount an intense nest defense that repels predators.

HYBRIDIZATION AND GENETIC DIFFERENCES

When two formerly isolated species or populations come into secondary contact there are four possible outcomes: coexistence with complete reproductive isolation, replacement of one population by the other, fusion of the two populations and complete mixing of the two gene pools (sometimes referred to as "dilution"), and establishment of a more or less permanent hybrid zone. The first scenario, coexistence with reproductive isolation between these biotypes, has not developed anywhere in the Americas. The second scenario seems to be the rule in nonmanaged populations: African-derived bees establish large feral populations and replace any resident European feral honey bees.

Gene flow between neotropical AHBs and EHBs seems to be strongly asymmetrical. AHBs have maintained their genetic integrity, in spite of hybridizing with EHBs, as they have expanded their distribution. Even after 45 years of interaction with EHBs, these bees are indistinguishable in behavior and so similar genetically to bees in the Transvaal of South Africa that it is more appropriate to refer to them as Neotropical African bees or as African-derived bees rather than Africanized bees. Low acquisition of EHB traits into the AHB population can be attributed to pre- and postzygotic isolating mechanisms (i.e., mate selection, queen developmental time, and hybrid dysfunction). For example, AHB queens mate predominantly with AHB drones even in the presence of large numbers of EHB drones. When AHB queens are inseminated with semen from drones of both types, the AHB queen progeny develops faster than the hybrids, assuring that in most cases the next queen would be an AHB rather than hybrid.

Colonies from backcrosses of F1 hybrid queens to either parental genotype have unusual metabolic patterns, low honey storage, and high rates of mortality. This finding suggests possible incompatibility of the nuclear and mitochondrial genomes of these biotypes. EHBs, in contrast, rapidly become Africanized, and nearly all traces of the EHB nuclear and mitochondrial genome disappear from the feral bee populations following the arrival of AHBs. Disappearance of the European traits seems to result from a lack of prereproductive isolation, which results in extensive mating by EHB queens with AHB drones. This is followed by a pattern of queen development that favors hybrid rather than EHB queens. Matings by these F1 queens to AHB drones results in colonies with low fitness and the eventual loss of EHB mitochondrial DNA from the population. Displacement of EHBs therefore seems to result, in part, in a type of "genetic capture" in which one form, *A. m. scutellata,* eliminates the others by hybridizing with their females. The genetic and population consequences of the interactions between *A. m. scutellata* and *A. mellifera* subspecies from Europe suggest that *A. m. scutellata* deserves the status of a semispecies.

HYBRID ZONES

A hybrid zone formed at the southern limit of AHBs in northern Argentina in the late 1960s and early 1970s. Although many anticipated that a similar zone would form in the United States, this has not happened. Coincident with the arrival of AHBs in the United States, the mite, *Varroa destructor,* an introduced brood parasite from Asia that kills EHB colonies, spread rapidly throughout the country eliminating feral EHB colonies. At present, feral EHBs, which usually are escaped swarms from managed apiaries, are transitory and persist only for a short time; this precludes the formation of a persistent hybrid zone. In Argentina, it is likely that *Varroa,* which arrived after the formation of the hybrid zone, has changed the dynamic of the interaction of both types of bee as well.

FUTURE

In the future, the only significant feral bee populations in the United States will consist of AHBs in the southwestern states and possibly Florida, until or unless EHBs acquire sufficient tolerance of *Varroa* mites to once again establish feral populations (Fig. 1). AHBs will not become more European through hybridization or selection and move further north. Barriers to gene flow into the AHB genome, coupled with selection against any AHBs with EHB genes that make them susceptible to *Varroa* mites, assures the continuation of a nearly separate Neotropical African bee genome. AHBs have many useful attributes but hesitancy on the part of beekeepers to work with these bees, partly because of familiarity with EHBs but also because of concerns of legal issues should their bees be linked

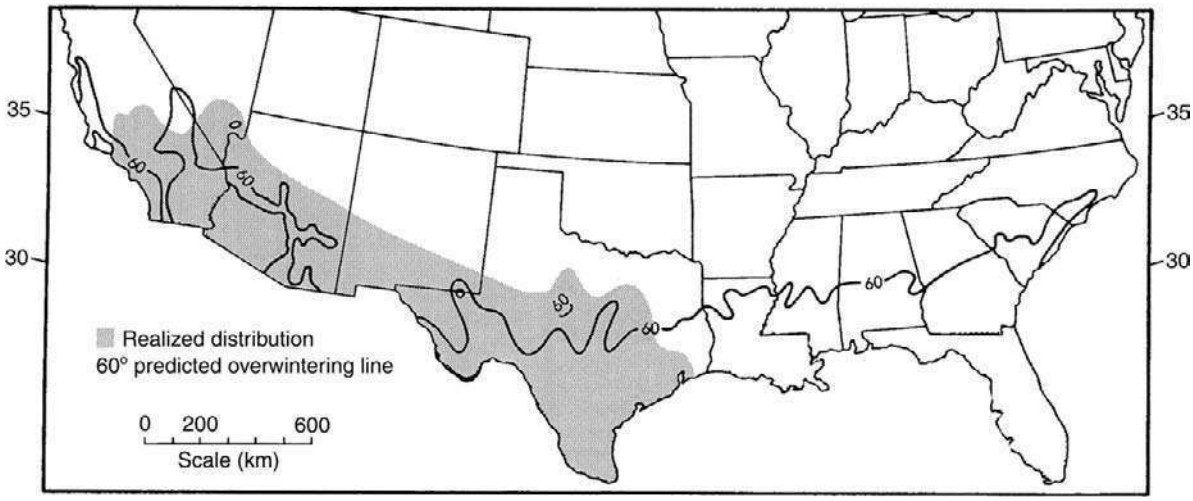

FIGURE 1 The known summer (shaded) distribution and predicted overwintering limit (60-degree line for January) of Neotropical African honey bees in the United States.

to stinging incidents, will keep American beekeepers from adapting this bee to apiculture.

See Also the Following Articles

Apis Spieces • Beekeeping • Introduced Insects • Venom

Further Reading

Hall, H. G., and Muralidharan, K. (1989). African honey bees spread as continuous maternal lineages. *Nature* **339,** 211–213.

Kerr, W. E., Del Rio, S., and De Barrionuevo, M. D. (1982). The southern limits of the distribution of the Africanized bee in South America. *Am. Bee J.* **122,** 196–198.

Michener, C. D. (1975). The Brazilian bee problem. *Annu. Rev. Entomol.* **20,** 399–416.

Taylor, O. R. (1985). African bees: Potential impact in the United States. *Bull. Entomol. Soc. Am.* **31,** 14–24.

Taylor, O. R., and Spivak, M. (1984). Climatic limits of tropical African honeybees in the Americas. *Bee World* **65,** 38–47.

Nervous System

see *Brain and Optic Lobes*

Nest Building

Robert L. Jeanne
University of Wisconsin

Insects manipulate their environment for a variety of purposes: to trap prey, attract mates, and provide shelter from the elements and protection from predators and parasites. Nests are a special category of environmental manipulation. A nest may be defined as any modification of the environment by adult insects that provides shelter for the rearing of their offspring. In most nest-building insects, the nests are simple excavations or small constructions that provide temporary protection for eggs or larvae, with or without the adult parent(s) in attendance to provide continuing parental care. In the numerous lineages of nest-building insects, an increase in parental care has generally been accompanied by the evolution of more elaborate nesting behavior. The trend has climaxed numerous times in the eusocial insects such as termites, ants, wasps, and bees, whose nests may be very large and architecturally complex and may house the colony for many years under homeostatically controlled physical conditions.

TAXONOMIC DISTRIBUTION OF NEST-BUILDING

Nest building has evolved in only a handful of mandibulate insect orders. In this brief survey, only a few selected examples of nest-building species are given for each.

Orthoptera

True nests have evolved in a few species of locusts and crickets. In the burrowing cricket, *Anurogryllus muticus* (Gryllidae), nesting behavior reaches the highest point found in the order. The female of this species excavates a brood chamber in the soil and then seals herself inside and lays her eggs. When the nymphs hatch, the mother feeds them with special trophic eggs and later with grass that she gathers outside and brings into the nest.

Coleoptera

A number of beetles manipulate the environment so as to provide shelter and/or food for their young. The female of the leaf roller, *Deporaus betulae* (Attelabidae), cuts across a leaf

along a precise trajectory, then rolls the leaf into a tube, inside of which she lays her eggs. The larvae feed on the inner layers of the leaf roll, while being protected by the outer layers. The dung beetles (Scarabaeidae) excavate nests in the ground and provision them with balls of dung rolled to the site. Some of the carrion beetles (Silphidae) form the body of a dead mouse or other small mammal into a ball, drop it into an excavated chamber, then lay eggs on it. In some species, the female remains in the nest and feeds the young larvae by regurgitation until they are large enough to feed on the carrion directly.

Embiidina

Both sexes of webspinners, adults as well as nymphs, produce silk from the swollen metatarsal glands in the forelegs, which they use to spin a network of galleries on tree trunks or in leaf litter. Not only do the galleries serve as a center for brood-rearing, but they also provide a shelter within which the family of webspinners grazes on bark, dead leaves, moss, or lichens.

Isoptera

All the termites are eusocial (reproductive division of labor, cooperation in brood care, and overlap of at least two generations capable of contributing labor to the colony) and all live in nests. In the more primitive species, the colony nests in the wood source it feeds on. Such "single-site nesting" is exemplified by the small colonies formed by species of Termopsidae, most genera of Kalotermitidae, and the less derived members of Rhinotermitidae. The colony spends its entire life in its log, the nest consisting simply of the irregular galleries excavated by the feeding termites. The "higher" termites, belonging to the family Termitidae and others, have evolved the ability to nest independent of their food source, in the soil or arboreally. Dissolving the identity between food source and nest freed these species to evolve larger colony size and to exploit a wider range of cellulose sources, including wood fragments of all sizes, grass, seeds, leaf litter, and humus.

Hymenoptera

Nesting behavior in the Hymenoptera is limited to three superfamilies: Sphecoidea, Vespoidea, and Apoidea. The ancestors of nest-building aculeates were nonnesting parasitoids of other arthropods. Nesting behavior probably got its start when a female parasitoid dragged her paralyzed prey into a crudely excavated nest in the ground and laid an egg on it, much as some sphecoids do today. Many of the solitary sphecoids and vespoids (sand wasps, digger wasps, spider wasps) excavate a subterranean nest and stock it with one or more paralyzed prey, on which an egg is laid, whereas others (mud daubers, potter wasps) construct aerial nests of mud. A few sphecids nest in hollow stems or other natural cavities.

Except for the parasitic "cuckoo bees," all bees (Apoidea) make nests. Most are solitary, the female excavating a nest in the ground or using hollow stems or other natural cavities. Carpenter bees excavate burrows in solid wood. Some solitary bees construct nests of resin or a mixture of resin and pebbles, leaf pulp, or mud on rocks, stems, or leaves.

Eusocial behavior has arisen in all three superfamilies. Nesting behavior, a prerequisite, was already established well in advance of the numerous origins of eusociality in these taxa. In the ants (Formicidae) and bees (Apidae), the evolution of eusocial behavior occurred in subterranean nests, while in the wasps (Sphecidae, Vespidae) it took place in constructed, aerial nests of naked brood cells. In each group, as social life became more elaborate, nests increased in size and complexity and adapted to new nesting sites. Although many species of ants nest in the ground, many others, especially in the tropics, construct arboreal nests or nest intimately with plants. Many of the eusocial bees construct their nests in cavities, whereas others construct aerial nests, either with combs exposed or enclosed in a heavy involucrum. With larger colony size in the wasps came the evolution of protective nest envelopes and/or the move to cavities in the soil or in trees.

NEST-BUILDING BEHAVIOR

Materials and Tools

For the majority of social species that excavate nests in soil or wood, the nest consists merely of the cavity left after the removal of material. In contrast, constructed nests, which have evolved in all four eusocial groups, require a combination of exogenous structural material and adhesive to bind the particles of material together. A variety of materials are used: termites use soil or wood particles cemented together with saliva and/or fecal material. Ants use wood or other vegetable fiber or mud. *Lasius fuliginosus,* for example, fills its nest cavity with an irregular carton meshwork glued together with honeydew. The matrix is strengthened by the penetration of the hyphae of a symbiotic fungus. Social wasps (Vespidae) are known as "paper wasps" because familiar species use wood pulp as a structural material, although many tropical species use plant hairs and some even use mud. The fibers are chewed and mixed with a proteinaceous secretion of the labial gland that dries into a plastic-like matrix, giving the finished carton strength and a modicum of water repellency. Wasps that build exposed, pedicellate combs construct the pedicel primarily of this secretion, giving it toughness and a dark, shiny appearance (Fig. 1). Honey bees are unusual in using wax, secreted by wax glands on the abdomen, rather than collected material, to construct their brood and storage cells. Other social bees also use wax, but mix it with exogenous materials, including pollen, plant resins (propolis), vegetable material, mud, or even feces. *Microstigmus* wasps (Sphecidae) produce silk from glands at the tip of the abdomen and use this to glue together the leaf pubescence from which their delicate nests are sculpted. Weaver ants (*Oecophylla* spp.) sew living leaves together with

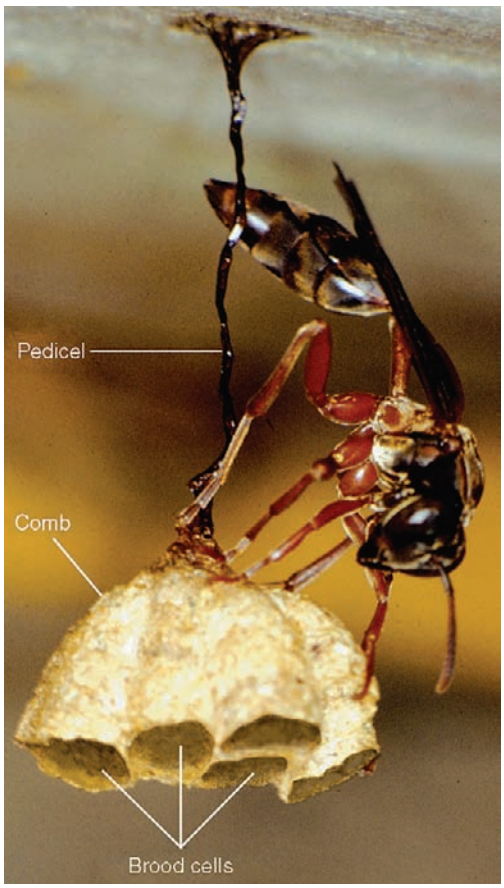

FIGURE 1 Newly founded nest of the social wasp, *Mischocyttarus drewseni,* from Brazil. The founding female is shown wiping an ant-repelling secretion (the gland opens at the base of the terminal abdominal sternite) onto the pedicel of the nest, where it reduces the likelihood that ants will discover the comb of brood cells while the queen is away on a foraging trip. (Reprinted, with permission, from R. L. Jeanne, Chemical defense of brood by a social wasp. *Science* **168,** 1465–1466. Copyright 1970 American Association for the Advancement of Science.)

strands of larval silk to create multiple arboreal nesting chambers in which the young are reared.

In all nest-building insects the mandibles are chisel and trowel, the primary tools used to excavate, collect, carry, and mix materials and shape them into the nest. Other tools are important in a few species: sand wasps use the legs to kick excavated sand out of the burrow, and paper wasps use the forelegs to help manipulate wads of nest material during chewing and mixing with oral secretion. Sensory feedback is critical for precision construction. Wasps use the antennae as calipers to control the size of brood cells in the comb. Honey bees measure brood cell diameter with the prothoracic tarsi, while they sense the thickness of wax in the cell walls via pressure receptors on the antennae.

Information Sources

The information required to construct the nest ultimately resides in the genome of individuals, not as a blueprint of the finished nest, but as a set of one or more kinds of construction acts combined with a set of decision rules. The decision rules determine the location and orientation of material added in relation to environmental cues that include gravity, the current structure of the nest, and the location and state of brood and food stores in the nest. It is the interaction of innate rules of behavior and feedback from external cues that results in the species-typical form of the nest.

The simplest nests of solitary species are constructed by following a linear (nonbranching) sequence of steps. A sand wasp digging her nest, for example, need only decide when to switch from extending the burrow to excavating a brood cell. In contrast, nonlinearity characterizes the construction behavior of all social insects. Rather than following a programmed linear sequence, workers make choices among several types of building behavior according to the current state of construction of the nest. Thus, a *Polistes* wasp can use her load of pulp to thicken the pedicel, lengthen a brood cell, initiate a new brood cell, or cover the silken cap of a cell containing a pupa. A social insect worker in a large colony may, in the course of her entire lifetime, perform only one or a small subset of the kinds of construction acts and decision rules in her species' repertory.

Social Organization of Building

In the eusocial insect colony, workers specialize on different elements of nest construction. Older *Polybia* wasp workers collect materials, some specializing in water, others in wood pulp. Back at the nest, these materials are turned over to younger workers, the builders, who keep the pulp moist with water as they add it to the appropriate places on the nest. The builders regulate the overall rate of activity, for it is they who have direct contact with the construction site and can determine the level of demand for materials. Foragers gain information about demand for their material as they seek builders to unload to.

Nest Architecture and Expansion

Termites and ants tend their brood in loose piles in nursery chambers. In contrast, eusocial bees and wasps rear their offspring individually in cylindrical cells (bumble bees rear several immatures per cell). In all but the simplest bee and wasp societies, brood cells are grouped into combs of various sizes, shapes, and orientations (Fig. 1). The most space-efficient way to close-pack cylindrical cells is to surround each cell with six others. Since adjoining cells share walls, this results in the familiar "honeycomb" pattern of hexagonal cells.

From the core of the nest outward, the typical arrangement is brood, then food storage areas, and finally the defensive structures. In the honey bee hive, for example, the central brood cells are surrounded by a concentric layer of pollen-storage cells and then an outer layer of honey-storage cells. The entrance to the nest cavity (or a hive box) is secured by guard bees against intrusion by predators and parasites. A similar arrangement is also seen in the nests of termites (Fig. 2) and wasps.

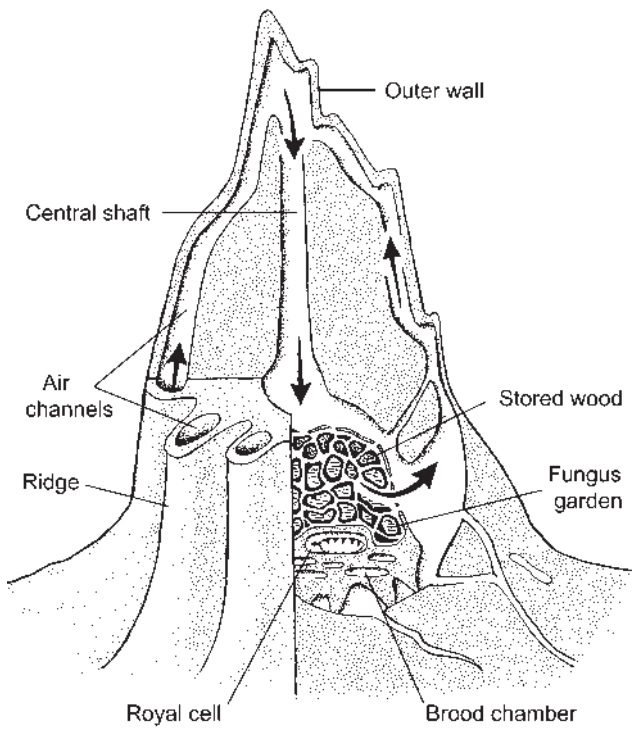

FIGURE 2 Simplified diagram of the nest of the termite, *Macrotermes bellicosus,* from the savanna of Ivory Coast. The front half of the nest is cut away, except for the lower left quarter, which shows the external surface. In the lower part of the mound, just at ground level, is the nest proper, a construction consisting of the central royal cell (containing the queen, king, and attending workers), surrounded by chambers containing brood, fungus gardens, and stored food. Surrounding this is the ridged outer nest, whose design enables it to function as a giant air conditioner. During the day in the dry season, ventilation within the nest is externally driven by the sun, which warms the air in the peripheral air channels, causing it to rise. This sets up a convective circulation within the mound (arrows). CO_2 produced in the central nest diffuses out through the walls of the ridges. Air temperatures are highest and CO_2 levels lowest in the upper portions of the peripheral air channels. Air temperature within the fungus gardens is kept within 29–31°C. (Illustration by Lee Clippard; based on Collins, 1979, and Korb and Linsenmair, 2000.)

Most social insects are able to expand their nests to accommodate colony growth. In some species, such as yellowjacket wasps, growth is continuous throughout the life the colony, whereas in others it occurs in bouts separated by periods of no growth. Honey bees expand the combs in the hive when there is a strong nectar flow coupled with a shortage of honey storage cells. In ants and termites, nest expansion may occur opportunistically after rains soften the soil.

FUNCTIONS OF NESTS

Social Functions

The nest is the information center of the colony. It is here that information is communicated about resource supply and demand and about the status of the queen. The nest itself is involved in the distribution of information about colony membership. Each colony has a unique mix of chemicals that labels every individual as belonging to that colony. This "colony odor" resides on the cuticle of each individual as well as in the nest material, and it has been shown for wasps and honey bees that newly emerging workers learn to recognize their colony odor from the nest.

Food Storage

Bumble bees construct specialized wax pots in which they store honey and pollen during periods of good foraging. Stingless bees and honey bees store pollen and enough honey to sustain the adult population through the unfavorable season. The "honey wasps" (*Brachygastra* spp.) of the Neotropics also store large amounts of honey in their brood cells for the same purpose. Many other species of social wasps store enough honey as droplets in empty brood cells to get the colony through several days of poor foraging. Desert seed harvester ants stockpile seeds in chambers in their nests, and honey ants (*Myrmecocystus* and others) store large amounts of honey in the crops of specialized workers called repletes. The fungus ants and higher termites in the subfamily Macrotermitinae (Termitidae) grow specialized fungus as food in chambers in their nests (Fig. 2).

Defense

Nests often incorporate or accommodate some means of defense against natural enemies. The broods of small, newly initiated colonies are especially vulnerable when the founding queen must leave the nest to forage. Wasps in the genera *Polistes* and *Mischocyttarus* suspend their uncovered combs from a narrow, tough pedicel, which they coat with an ant-repelling secretion produced by an exocrine gland at the base of the terminal sternite (Fig. 1). Among the swarm-founding wasps of the tropics are several species that surround the access to the nest with "ant traps" made of carton bristles several millimeters long, each tipped with a sticky droplet.

Bees and wasps that nest in cavities or construct protective outer covers reduce access by ants and parasitoids to a narrow entrance that can be guarded by a few defending workers. Some stingless bees cover the entrance tube with sticky propolis as a barrier against ants, whereas others pull the soft, waxy tube closed each night. The outer layer of the involucrum of arboreal nests of *Trigona corvina* and *T. spinipes* is thin and easily broken by an intruder, allowing defending bees to swarm out through passageways in the tough, inner layer and launch an attack.

Homeostasis

Social species that form small colonies can exert little control over temperature, humidity, or atmospheric gas concentration in the nest, but some compensate by placing their nests in favorable microhabitats. By building their nests where sun-warmed air collects, such as under eaves on the east and south sides of outbuildings in sunny locations, *Polistes* wasps

at higher latitudes achieve shorter egg–adult development times than they would at ambient temperatures.

By virtue of a larger metabolizing biomass and lower nest surface/volume ratios, social species with larger colonies are better able to regulate nest conditions, and nest architecture is often adapted to enhancing homeostatic control. Large colonies produce considerable amounts of metabolic heat, raising nest core temperature well above ambient. Thick nest cavity walls or insulating envelope reduce the loss of this heat to the environment. By combining metabolic heating with evaporative cooling, honey bees can regulate the temperature in the core of the nest to within half a degree of 35°C, even if the outside temperature is many degrees lower or higher. The multiple layers of paper envelope of yellowjacket wasps *(Vespula)* enclose dead air spaces that insulate the nest against heat loss, enabling the colony to maintain steady nest temperatures well above ambient.

Subterranean nests of termites and ants have less control over the temperature in the chambers of their nests, but they can construct the nest to take advantage of solar heating. Some ground-nesting ants of temperate regions excavate chambers under flat rocks lying on sunny ground. As the rock warms in the sun during the day, heat is conducted downward to the ground below. By moving the brood up into these warm but moist chambers during the day, the ants accelerate the development of their immatures. At night, as the rock loses its heat, the brood is moved down to relatively warmer chambers deeper in the soil.

Some ants in higher latitudes build honeycombed mounds of soil or plant detritus. In some of the *Formica* species these can be over 2 m in height. The sun warms the mound to several degrees above ambient, and the colony incubates pupae by moving them up into chambers in the mound during the day.

The most spectacular examples of homeostatic control of nest conditions are the large epigeal (aboveground) mounds built by termites in the savannas of the tropics. Homeostatic mechanisms vary across species, habitat, season, and even time of day. One example is *Macrotermes bellicosus,* found on the savannas of western Africa, whose colonies can reach 2 million workers living inside large, cathedral-like towers that are 3 m or more in height. Air circulation within the mound during the day in the dry season is shown in Fig. 2.

See Also the Following Articles

Ants • Beekeeping • Homeostasis • Isoptera • Parental Care • Sociality • Wasps

Further Reading

Abe, T., Bignell, D. E., and Higashi, M. (2000). "Termites: Evolution, Sociality, Symbioses, Ecology." Kluwer, Boston.

Collins, N. M. (1979). The nests of *Macrotermes bellicosus* (Smeathman) from Mokwa, Nigeria. *Insect. Soc.* **26,** 240–246.

Edgerly, J. S. (1997). Life beneath silk walls: A review of the primitively social Embiidina. *In* "The Evolution of Social Behavior in Insects and Arachnids" (J. C. Choe and B. J. Crespi, eds.). Cambridge University Press, Cambridge, UK.

Evans, H. E. (1958). The evolution of social life in wasps. *In* "Proceedings of the 10th International Congress of Entomology," pp. 449–457.

Evans, H. E. (1966). "The Comparative Ethology and Evolution of the Sand Wasps." Harvard Univ. Press, Cambridge, MA.

Hansell, M. H. (1984). "Animal Architecture and Building Behaviour." Longman, New York.

Hölldobler, B., and Wilson, E. O. (1990). "The Ants." Harvard University Press, Cambridge, MA.

Korb, J., and Linsenmair, K. E. (2000). Ventilation of termite mounds: New results require a new model. *Behav. Ecol.* **11,** 486–494.

Michener, C. D. (2000). "The Bees of the World." Johns Hopkins University Press, Baltimore.

Ross, K. G., and Matthews, R. W. (1991). "The Social Biology of Wasps." Cornell University Press, Ithaca, NY.

Seeley, T. D. (1995). "The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies." Harvard University Press, Cambridge, MA.

Shellman-Reeve, J. S. (1997). The spectrum of eusociality in termites. *In* "The Evolution of Social Behavior in Insects and Arachnids" (J. C. Choe and B. J. Crespi, eds.). Cambridge University Press, Cambridge, UK.

Wilson, E. O. (1971). "The Insect Societies." Harvard University Press, Cambridge, MA.

Neuropeptides

Miriam Altstein
Agricultural Research Organization, Israel

Neuropeptides (Nps) are extracellular chemical messengers, found throughout the animal kingdom, forming a most structurally and functionally diverse group of compounds. Nps have been very well conserved during the course of evolution, indicating their major role as regulators of physiological processes. Nps may act as neurotransmitters, neurohormones, or neuromodulators and, in the hierarchy of entities that regulate endogenous biochemical control functions, the Np messengers rank the highest. The original definition of Nps covered small molecules (< 50 amino acids) of a peptidic nature, synthesized in specialized nerve cells termed neurosecretory cells (NSC) and released from their axon terminals, either into the intracellular space of an adjacent cell (nerve, endocrine, or nonendocrine) or into the circulatory system. Nps released into intracellular spaces, affect proximal effector sites; those that enter the general circulation reach peripheral organs, where their activity can be manifested either directly by activation of a distal target organ or indirectly via signals to nonneuronal internal secretory glands. In recent years, during which many nonneuronal tissues have been found to produce the same peptides as neural tissues, the Np concept has been widened to include peptides that serve to integrate the brain and other tissues for the maintenance of normal physiology, homeostasis, and behavioral patterns.

In insects, Nps were found to regulate a long list of physiological and behavioral processes during development, reproduction, and senescence, and to maintain growth, homeostasis,

osmoregulation, water balance, metabolism, and visceral activities. In the past two decades, a large number of insect Nps have been identified, some of which are similar in structure to vertebrate Nps. The study of insect Nps is diverse and multidisciplinary. It integrates cellular and molecular studies of the basic principles of Np action (e.g., biosynthesis, posttranslational processing, release, transport, activation of the target cell, and degradation), chemical approaches for their identification and characterization, immunochemical studies for anatomical localization, and physiological, behavioral, and pharmacological approaches to study their roles in the physiology of organisms. This article describes the distribution and localization of the insect neuroendocrine system, lists the various Np families, and considers the cellular and molecular basic steps of Np action, providing insights into the common properties of the large number of Nps presently identified and the approaches taken to study their regulatory functions.

NP RESEARCH IN HISTORICAL PERSPECTIVE

The concept of neuroendocrine control dates back to the beginning of the 20th century (1922) when Stephan Kopeč first suggested that metamorphosis in insects is regulated by brain factors that are released into the hemolymph. Further progress in the field came from the studies of Berta and Ernst Scharrer, who introduced the basic concepts of neurosecretion and NSC, and described the similarities between the retrocerebral complex in insects and the hypothalamic–hypophysial system in vertebrates. Although neurosecretion was first observed in insects, invertebrate neuropeptide research lagged behind the vertebrate studies, mainly because of low availability of biological material and the lack of sensitive techniques for isolation, sequencing, and synthesis of peptides. The development of chemical, biochemical, and genetic engineering technologies as well as the growing awareness of the major role Nps play in the physiology of organisms, stimulated active interest in insect Np studies, and indeed, in 1975 Starratt and Brown released their pioneering publication announcing the initial determination of a primary structure of the insect Np proctolin. Since then, nearly 150 insect Nps have been reported in the literature, most of which have been isolated from cockroaches (e.g., *Leucophaea maderae, Periplaneta americana, Diploptera punctata*), locusts (e.g., *Locusta migratoria, Schistocerca gregaria*), moths (e.g., *Manduca sexta, Bombyx mori,* and various Heliothinae species), and the fruit fly, *Drosophila melanogaster.*

ANATOMY OF THE NEUROENDOCRINE SYSTEM IN INSECTS

The cellular distribution of Nps in insects has been mapped by histochemical and immunocytochemical techniques as well as by *in situ* hybridization studies. On the basis of this work it was found that the main localization of NSC and their release sites are in the brain–subesophageal ganglion

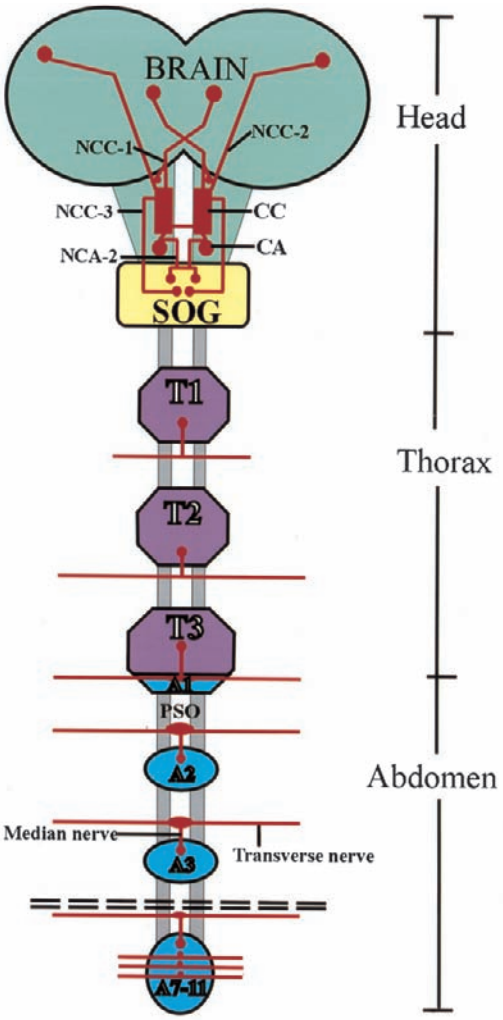

FIGURE 1 Schematic representation of the major neurohemal release sites of the central nervous system of insects. Most of the thoracic and abdominal release sites have not yet been fully characterized. CC, corpora cardiaca; CA, corpora allata; SOG, subesophageal ganglion; NCC, nervus corporis cardiaci; NCA, nervus corporis allati; PSO, perisympathetic organs; T1–T3, thoracic ganglia; A1–A11, abdominal ganglia. [From Predel, R., and Eckert, M. (2000). Nemosecretion: peptidergic systems in insects. *Naturwissenschaften,* **87,** 343–350. © Springer-Verlag GmbH & Co. KG, Heidelberg.]

(SOG)–corpora cardiaca–corpora allata complex (Fig. 1). This complex comprises six clusters of NSC: a pair of medial NSC that originates in the pars intercerebralis (PI), a pair of lateral NSC that originates in the protocerebral region of the brain, and a pair of NSC that originates in the SOG. These six clusters form axon bundles termed nervi corporis cardiaci (NCC): NCC1, NCC2, and NCC3, respectively. Each of the nerve bundles terminates in a pair of retrocerebral neurohemal glands termed corpora cardiaca (CC). The nerve terminals form the storage lobe of the CC through which Nps are released into the circulatory system. Additional neuroendocrine (intrinsic) cells are present in another lobe of the CC glands (the glandular lobe), which is the site of synthesis and release of other Nps. Two additional clusters of

NSC that originate in the SOG extend axons that form the nervi corporis allati 2 (NCA2), which terminate in another pair of endocrine glands (of non-nervous-tissue origin) termed corpora allata (CA) and form another neurohemal region. Another pair of dorsolateral NSC that originate in the PI region of the brain, extend axons that terminate in the CA. The CC and CA are adjacent glands, partially fused with the ventral wall of the aorta, which enables release into the circulatory system of the Nps that are synthesized in the glandular cells, as well as those from the brain/SOG nerve terminals.

Another much smaller neuroendocrine structure consists of the segmentally arranged perisympathetic organs (PSO), which serve as the storage and release site of Nps produced in the ventral nerve cord. Most of the Nps that have been detected in the PSO are not homologous with those found in the retrocerebral complex. A newly discovered endocrine structure is the epitracheal system, which consists of segmentally arranged nerve cells located at the trachea near the spiracles that form the epitracheal glands (EGs). The system produces two blood-borne Nps that trigger pre-ecdysis and ecdysis behavior: pre-ecdysis triggering hormone (PETH) and ecdysis triggering hormone (ETH). Other NSC are distributed in the insect body: in the central nervous system, in the sympathetic nervous system and the peripheral nervous system, on the aorta, and at the ampullae of the antennal heart. Most of these release sites are situated in well-circulated regions of the body. Np-containing cells have also been reported in visceral organs such as the gut, oviduct, accessory glands, and even hemocytes. Raabe has provided a detailed description of the insect neuroendocrine anatomical structure.

NP FAMILIES

Categorization of Nps is usually based on their main action or the one for which a given Np is best known. The major groups of Nps involved in development, reproduction, homeostasis, myotropic activity, and coloration are as follows.

Developmental Nps: The main Nps in this category are the allatotropins/allatostatins, which stimulate/inhibit synthesis of juvenile hormones by the CA; PETH, ETH, crustacean cardioactive peptide (CCAP), and eclosion hormone (EH), which are involved in controlling pre-ecdysis and ecdysis behavior; prothoracicotropic hormone (PTTH), which stimulates molting by initiating biosynthesis and release of ecdysone by the prothoracic gland and diapause hormone that arrests development.

Reproductive Nps: This family includes the ovary-maturating peptide (OMP) and egg development neurosecretory hormone (EDNH), which stimulate egg development; oostatic hormone (OH), which inhibits maturation of ovaries; trypsin modulatory oostatic factor (TMOF), which regulates egg development by modulating trypsin biosynthesis in the gut; neuroparsin, which affects gonad activity; PTTH, which affects egg development; and pheromone biosynthesis activating neuropeptide (PBAN), which elicits sex pheromone biosynthesis in female moths.

Homeostatic Nps: The homeostatic group includes adipokinetic hormone (AKH), hypertrehalosaemic peptides, bombyxin, ion transport peptide (ITP), and other insulin-related peptides that control fat, carbohydrate, and protein metabolism. Additional members of the family are the diuretic and antidiuretic peptides, which are involved in ion and water balances.

Myotropic Nps: This family is one of the largest Np families in insects. It includes peptides such as proctolin and cardiostimulatory peptides, myokinins, sulfakinins, pyrokinins, myotropins, tachykinins, periviscerokinins, accessory gland and midgut myotropins, myoinhibitory peptides, and FMRFamide-related peptides.

Chromatotropic Nps: Members of this family include melanization and reddish coloration hormone (MRCH), pigment-dispersing hormone (PDH), and corazonin, a cardioactive peptide that has recently been found to exhibit dark pigmentation properties.

Recent immunocytochemical studies have indicated the presence of insect Nps that are comparable to vertebrate Nps. Their functions, however, have not been discovered. Most of the Np families in insects have not yet been detected in vertebrates.

Most of the above-mentioned insect Nps have been characterized, their amino acid sequences have been determined, and their cDNA and genes have been cloned from various insect species. The studies revealed that some Nps may occur in multiple forms (e.g., AKH; allatostatins; myotropic and FMRF-related peptides), a well-known phenomenon among invertebrate Nps. The multiple peptide forms are often encoded by the same gene and result from repeated internal gene duplication and subsequent diversification. Np diversity can also result from a duplication of the whole gene and subsequent mutations. Many Nps elicit more than one biological response in the same or different insect species, and several biological activities may be regulated, in the same insect species, by more than one peptide. A detailed, well documented, review on the structural, biochemical, and physiological characterization of insect Nps has been presented by Gäde.

NOMENCLATURE

The nomenclature of insect Nps is usually based on two primary characteristics: the Np source, which is indicated by the first two letters of the genus name (with the first letter capitalized) and the first letter of the species name, and the first-reported or the major biological function. For example, a peptide isolated from *Helicoverpa zea,* which was first reported to have a pheromonotropic activity, would be designated Hez-PBAN. A detailed explanation of insect peptide nomenclature was presented by Raina and Gäde.

CELLULAR AND MOLECULAR ASPECTS OF NP ACTIVITY

The cellular and molecular components of Np activity include biosynthesis, release, transport, activation of the target cell, and degradation of the Np to terminate its action. Nps are synthesized as large precursor polypeptide chains (termed preprohormones) that include a signal sequence that is removed by an endopeptidase during translation. The remaining peptide chain, the prohormone, is transported to the Golgi apparatus, where it is packed into secretory vesicles in which it is further processed proteolytically into smaller fragments by prohormone-converting enzymes to yield biologically active (and inactive) peptides. During the transit through the Golgi network, the precursors may be subjected to posttranslational modifications such as glycosylation, phosphorylation, sulfation, or hydroxylation. Gene expression is regulated by a complex series of factors controlled by other Nps and neurotransmitters. Nps are secreted by a regulatory secretory pathway in which peptides, stored in secretory vesicles, are released in response to secretagogues. Secretion of Nps is usually Ca^{2+} dependent and, unlike neurotransmitters, Nps are not recycled at the nerve terminals but are newly synthesized in the cell body. Upon secretion (either to the circulatory system via neurohemal organs or to the intracellular space of adjacent cells), Nps reach their target organ, where they activate the target cells by binding to cell surface proteins (termed receptors) and exciting second-messenger systems, thus initiating a variety of cellular responses. At the end of the activation, Nps dissociate from the receptor and are rapidly inactivated by peptidases present in the plasma, or the intercellular space or in the target cell membrane. All the above-mentioned events are common to all Nps, regardless of their origin or biological function. Strand has provided a detailed summary of these processes.

FUTURE PROSPECTS

Despite the enormous amount of structural information that has been accumulated on insect Nps, our understanding of their mode of action is very limited and rudimentary. It is still necessary to develop additional *in vivo* bioassays that involve whole organisms and to create novel tools and technologies to unravel the complex coordination of the many Nps involved in the regulation of the physiological processes and behavioral patterns. There is also a need to study the cellular and molecular factors that underlie their activity, and to discover their pathways of synthesis and release and their targets of action. These issues are being addressed with the help of highly advanced molecular biology and genetic engineering techniques (e.g., gene cloning and expression, *in situ* hybridization, gene transfer, gene knockouts, site-directed mutagenesis), immunocytochemical techniques, and advanced chemical (e.g., liquid chromatography in combination with mass spectrometry), immunochemical, and biochemical methods. Of great importance is the study of Np receptors and the development of selective agonists and antagonists, which may serve not only as research tools but also as a basis for the design of novel, environmentally friendly insect control agents. Altstein has summarized a novel approach to the exploitation of this avenue.

See Also the Following Articles

Brain and Optic Lobes • Development, Hormonal Control of • Juvenile Hormone

Further Reading

Altstein, M. (2001). Insect neuropeptide antagonists. *Biopolymers (Pept. Sci.)* **60,** 460–473.

Gäde, G. (1997). The explosion of structural information on insect neuropeptides. *In* "Progress in the Chemistry of Organic Natural Products," (W. Herz *et al.,* eds.) Vol. 71, p. 128. Springer-Verlag, New York.

Kopeč, S. (1922). Studies on the necessity of the brain for the inception of insect metamorphosis. *Biol. Bull.* **42,** 323–342.

Raabe, M. (1989). "Recent Developments in Insect Neurohormones." Plenum Press, New York.

Raina, A. K., and Gäde, G. (1988). Insect peptide nomenclature. *Insect Biochem.* **18,** 785–787.

Scharrer, B., and Scharrer, E. (1944). Neurosecretion. IV: Comparison between the intercerebralis–cardiacum–allatum system of the insects and the hypothalamo-hypophysial system of the vertebrates. *Biol. Bull.* **87,** 242–251.

Starratt, A. N., and Brown, B. E. (1975). Structure of the pentapeptide proctulin, a proposed neurotransmitter in insects. *Life Sci.* **17,** 1253–1256.

Strand, F. L. (1999). "Neuropeptides. Regulators of Physiological Processes." MIT Press, Cambridge, MA.

Neuroptera
(Lacewings, Antlions)

Catherine A. Tauber and Maurice J. Tauber
Cornell University

Gilberto S. Albuquerque
Universidade Estadual do Norte Fluminense, Brazil

The Neuroptera, also known as Planipennia, is one of the oldest insect orders with complete metamorphosis; it includes the green and brown lacewings, antlions, owlflies, dustywings, mantidflies, and allies. Although it is a relatively small order, with about 6000 species distributed among 17 families, its members occur in a variety of habitats throughout the world and their habits are diverse and interesting. Largely because of their lacey and colorful wings, delicate bodies, and fascinating biology, neuropteran adults are attractive to both biologists and laypersons (see Fig. 1).

Neuropteran larvae, which are less noticeable than adults, have received much less attention. Unlike the adults, which may or may not be predaceous, almost all neuropteran larvae are predaceous; they feed on a variety of soft-bodied arthropods. Because of this predaceous habit, several families

FIGURE 1 Neuropteran adult in the family Crocidae. Note the colorful forewings, long slender hind wings, and delicate body. (Photograph by E. S. Ross, California Academy of Sciences.)

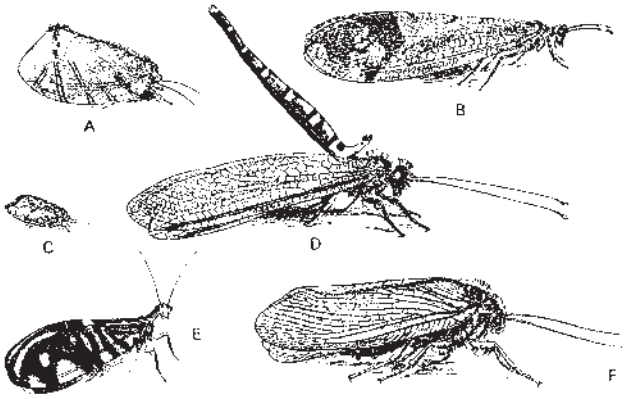

FIGURE 2 Adult Neuroptera. (A) Psychopsidae; (B) Myrmeleontidae; (C) Hemerobiidae; (D) Ascalaphidae; (E) Osmylidae; (F) Ithonidae. (Reproduced, with permission, from New, 1991, © CSIRO Australia.)

(primarily Chrysopidae, Hemerobiidae, and Coniopterygidae) are very useful in the natural, biological, and integrated control of many economically significant insect pests. But, despite their actual and potential importance, they have received less emphasis than other groups, such as the predaceous lady beetles.

The name "Neuroptera" is rooted in two Greek words: "neuron" meaning "sinew" and "pteron" meaning "wing." It refers to the netlike arrangement of the veins and crossveins of the wings and comes from an old usage of the word "nervation," meaning "strengthening by sinews." The name does not refer to nerves, as suggested by many authors; the order received its name long before nerves were recognized as tissue.

CHARACTERISTICS

Typically, neuropteran adults are soft-bodied and have four membranous wings that are similar in size, structure, and venation and that are held roof-like over the body while at rest (Fig. 2). Branches of the veins are generally bifurcated at the margins of the wings. The adults have chewing mouthparts, large lateral eyes, and multiarticulate antennae that are usually filiform (threadlike) or moniliform (with beadlike segments); except for one family (Osmylidae), neuropteran adults lack ocelli. The mesothorax and metathorax are similar in structure, and the abdomen is cylindrical and without cerci.

Unlike megalopteran and raphidiopteran larvae, neuropteran larvae differ distinctly from their adult stages. Larval mandibles and maxillae are usually elongate, slender, and modified for sucking, whereas maxillary palpi are absent. The larval thorax bears walking legs, and the one-segmented tarsus usually ends in two claws that function in locomotion. The terminal adhesive disks of the abdomen also aid locomotion. Like the adult abdomen, the larval abdomen does not have cerci.

Neuropteran pupae are enclosed within silken cocoons and they have characteristics that help distinguish them from the pupae of other insects. For example, they are exarate, i.e., the legs and wings are free from the body, the abdomen is moveable; as a result the pupa itself is capable of limited locomotion. Also, they are decticous, i.e., they have strong mandibles that are used to open the cocoon during emergence.

Neuroptera possess some interesting cytological characteristics. Although some variation occurs, almost all neuropteran species studied thus far have XX/XY sex determination. However, the sex chromosomes of most neuropteran species have an unusual type of pairing, called distance pairing, in which the chromosomes do not align to form bivalents during meiosis; rather, they are pulled from within the spindle to stabilized positions at the poles. This characteristic form of meiosis is shared with Raphidioptera. In many neuropteran taxa, meiosis occurs early in development, e.g., during the last instar or the pupal stage. In some taxa, the adult males have completely degenerate testes; mature sperm bundles are stored in seminal vesicles.

FOSSIL RECORD AND GEOGRAPHIC DISTRIBUTION

Although the fossil record of the Neuroptera is small and fragmentary, ancient (extinct) neuropteran families have been traced back with certainty to the Lower Permian in Kansas and the Upper Permian in Australia and Russia. The affinities of these archaic forms to modern taxa are unknown.

Modern families (e.g., Nymphidae, Psychopsidae, Chrysopidae, Osmylidae) appear in mid-Mesozoic fossils; probably the earliest and most diverse of these fossils are allied to the Psychopsidae. The largest known neuropteran, a psychopsid-like lacewing with a wingspan of 24 cm and large conspicuous eyespots on the wings, existed in the Jurassic. Recognizable examples of berothids and coniopterygids were also present during the Lower Jurassic. Cretaceous amber includes specimens from an array of modern families, including Berothidae, Mantispidae, Sisyridae, Chrysopidae, and Hemerobiidae, and the diverse Neuroptera found within Baltic amber can be placed in modern families.

Today, the order Neuroptera is distributed worldwide, with the exclusion of Antarctica. Europe, North America,

and Asia have rich neuropteran faunas as do southern Africa, South America, and Australia. Australia probably has the broadest diversity; it lacks representatives in only two families (Dilaridae and Polystoechotidae) and most of the presumed archaic families are represented there (e.g., Nevrorthidae, Ithonidae). In contrast to this abundance and diversity, New Zealand and the South Pacific islands have only meager neuropteran faunas. Remarkably, Hawaii seems unique in that it is the only island group where complexes of endemic species (Hemerobiidae and Chrysopidae) have evolved.

EVOLUTIONARY RELATIONSHIPS WITH OTHER ORDERS

Taken together, the three orders, Megaloptera (dobsonflies, alderflies), Raphidioptera (snakeflies), and Neuroptera (Neuroptera *sensu stricto,* Planipennia) (lacewings, antlions, dustywings, and allies) form the superorder Neuropterida.

Because of its ancient fossil record and the generalized body structure of its larvae and adults, this superorder is considered to be among the most primitive of the Holometabola (insects with complete metamorphosis). Significant morphological and molecular evidence indicates that this superorder has a sister relationship with the Coleoptera.

A number of synapomorphic (shared, relatively derived or specialized) characteristics distinguish the Neuroptera as a monophyletic order that is separate from Megaloptera and Raphidioptera; notably, almost all of these distinguishing features occur in the larvae. For example, megalopteran and raphidiopteran larvae have biting–chewing mouthparts and the mouth opens anteriorly. In contrast, the mouthparts of neuropteran larvae are suctorial and consist of elongate and pointed mandibles and maxillae whose adjacent grooved surfaces form a feeding tube. The mouth, instead of opening anteriorly, connects to the feeding tubes at the sides of the head.

Other larval characteristics distinguish the Neuroptera from the other two neuropteridan orders. Neuropteran larvae do not have contiguous intestinal tracts. Rather, the midgut and hindgut remain separate until pupation. As the larva feeds, feces accumulate in the midgut, and only after metamorphosis from the larval stage to the adult, during which the midgut and hindgut become connected, does the newly emerged adult expel the feces in the form of a meconial pellet. Neuropteran larvae use the hindgut and associated structures (Malphigian tubules) to produce silken cocoons within which pupation occurs. In contrast, megalopteran and raphidiopteran larvae have contiguous intestines. Moreover, they do not form cocoons; rather they pupate within earthen chambers or wooden cells.

Another synapomorphy of neuropteran larvae is an articulated, neck-like cervix. This contrasts with the ribbon-like cervical sclerite of megalopteran and raphidiopteran larvae.

Two recent cladistic analyses provide conflicting results concerning the relationships among the three neuropteridan orders. One study, based on morphological characteristics, indicates that contrary to traditional thought, the Megaloptera and Neuroptera are more closely related to each other than either is to the Raphidioptera. The subsequent cladistic analysis, which was based on both morphological and molecular data, yielded a trichotomous relationship among the three neuropteridan orders. The relationships remain unresolved; but, in this article we chose to use the first study; it yielded dichotomous relationships among the three neuropteridan orders and a fairly well supported sister relationship between Megaloptera and Neuroptera (Fig. 3).

EVOLUTIONARY RELATIONSHIPS WITHIN NEUROPTERA

The order Neuroptera encompasses 17 families that currently fall into three more or less well-supported suborders (see below). Several strong larval synapomorphies support the hypothesis of a sister relationship between Myrmeleontiformia and Hemerobiiformia. The Nevrorthiformia appears to be ancestral.

Order Neuroptera (=Neuroptera *sensu stricto,* Planipennia)
- Suborder Nevrorthiformia (=Neurorthiformia)
  - Family Nevrorthidae (=Neurorthidae) (11 species)
- Suborder Myrmeleontiformia
  - Family Psychopsidae (silky lacewings) (26 species)
  - Family Nemopteridae (spoon-winged lacewings) (100 species)
  - Family Crocidae (thread-winged lacewings) (50 species)
  - Family Nymphidae (including Myiodactylidae of some authors) (split-footed lacewings) (35 species)
  - Family Myrmeleontidae (including Stilbopterygidae of some authors) (antlions) (2100 species)
  - Family Ascalaphidae (owlflies) (400 species)
- Suborder Hemerobiiformia
  - Family Ithonidae (including Rapismatidae of some authors) (moth lacewings) (32 species)
  - Family Polystoechotidae (giant lacewings) (4 species)
  - Family Chrysopidae (green lacewings) (1200 species)
  - Family Osmylidae (160 species)
  - Family Hemerobiidae (brown lacewings) (550 species)
  - Family Coniopterygidae (including Brucheiseridae of some authors) (dustywings) (450 species)
  - Family Sisyridae (spongillaflies) (50 species)
  - Family Dilaridae (pleasing lacewings) (50 species)
  - Family Mantispidae (mantidflies) (400 species)
  - Family Berothidae (including Rhachiberothidae of some authors) (beaded lacewings) (115 species)

SUBORDER NEVRORTHIFORMIA

The Nevrorthiformia (=Neurorthiformia, which is a misspelling in the literature) contains a single, very small family, Nevrorthidae (=Neurorthidae, a misspelling). This family shares several characteristics with the Megaloptera and is considered basal among the Neuroptera; e.g., its aquatic larvae have distinctive head structures and Malphigian tubules

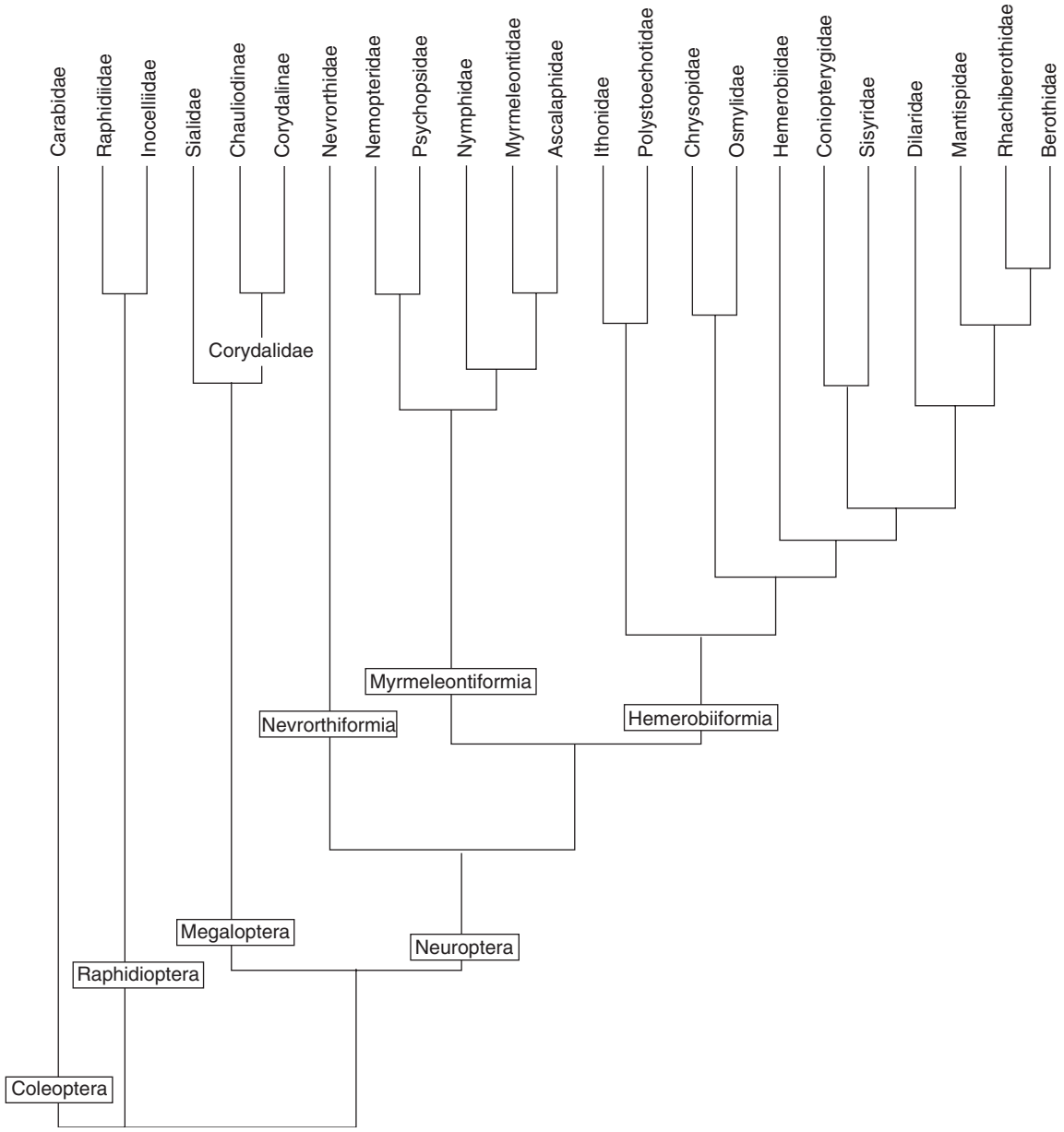

FIGURE 3 Phylogenetic relationships within Neuropterida (as proposed by Aspöck *et al.,* 2001). Note the sister relationship between Neuroptera and Megaloptera, the composition of the three neuropteran suborders, and the relationships of the families within the suborders.

that lie free distally (i.e., without cryptonephry). Thus, the Nevrorthiformia emerges as the basal neuropteran suborder.

Among the Neuropterida, a terrestrial lifestyle without cryptonephry is considered plesiomorphic (primitive); this lifestyle is typical of Raphidioptera. The evolution of an aquatic larva (without cryptonephry) is considered a synapomorphy that supports the sister-relationship between the megalopteran and the neuropteran lineages. Apparently, the aquatic lifestyle was retained in the Nevrorthiformia, whereas a reversal to the terrestrial lifestyle characterizes the remainder of the Neuroptera (the Myrmeleontiformia and Hemerobiiformia). Moreover, in these two groups (Myrmeleontiformia and Hemerobiiformia), cryptonephry arose (i.e., all of the Malphigian tubules except two are fused to the hindgut). The only exception is found in Sisyridae, which falls within the Hemerobiiformia; here, the apparent partial loss of cryptonephry (all but one Malphigian tubule are free of the rectum) may be regarded as an evolutionary reversal associated with the secondary evolution of an aquatic larva. The single remaining fused Malphigian tubule may be evidence of ancestral cryptonephry.

Nevrorthidae

Eleven species are known from three genera of Nevrorthidae; these occur in the Mediterranean Region, eastern Asia, and Australia. Nevrorthid adults are small (forewing length 6–10

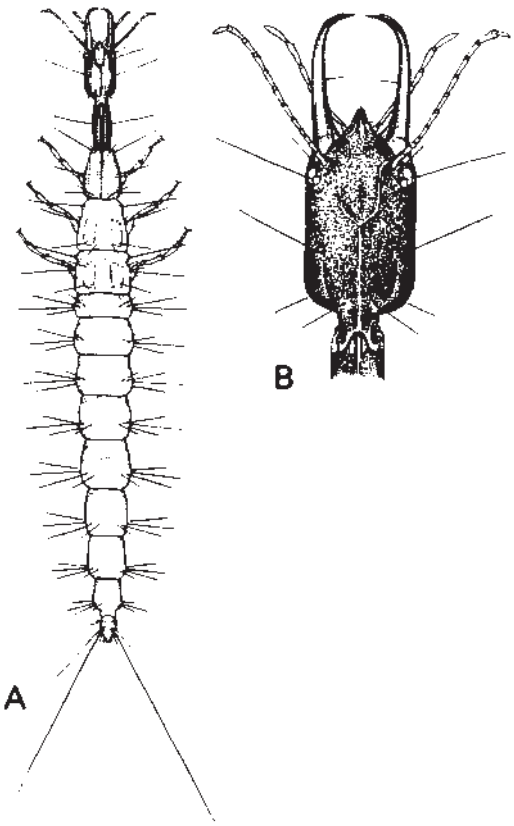

FIGURE 4 Nevrorthidae. (A) Larval body; (B) larval head. (Reproduced with permission, from New, 1991, © CSIRO Australia.)

mm) and delicate, and although they resemble sisyrids, they have unique, defining characteristics in the head structure and male and female genitalia.

Nevrorthid larvae, like sisyrid larvae, occur in aquatic habitats, but morphologically they are very distinct. For example, unlike sisyrids, they lack abdominal gills and their mouthparts are robust and curved inward (not needle-like) (Fig. 4).

Little is known about the biology or behavior of nevrorthid adults or larvae. Larvae of a southern European species have been collected on the stony bottoms of cold, swiftly moving mountain streams; those of an Australian species are believed to live in moist litter. Adults are usually found near streams or in wet, forested areas.

SUBORDER MYRMELEONTIFORMIA

Although its constituent members have changed over the years, the suborder Myrmeleontiformia has long been recognized as a valid grouping. Currently, it is composed of six families: Psychopsidae, Nemopteridae, Crocidae, Nymphidae, Myrmeleontidae, and Ascalaphidae.

Although myrmeleontiform adults are morphologically diverse and without clearly defined synapomorphies, the larvae of the six families share many morphological and biological characteristics. Primary among these are a head capsule with a highly sclerotized tentorium and a prementum that resembles a segment of the labial palp. Generally, myrmeleontiform larvae ambush, rather than pursue, their prey, and the head capsules and jaws are modified into a "trap"-like mechanism that can close very quickly. Their long, robust, inwardly curved mandibles are frequently toothed and constricted rather than enlarged basally. The maxillae are lance-like, and the head capsule is robust, quadrate, or cordate (Fig. 5). The body form is short, broad, and powerful. The antennae are short and have only 10 to 12 segments, with a thick scape and a narrow distal portion. There are two tarsal claws and the empodium is absent (except from psychopsid larvae).

A sister relationship between Myrmeleontiformia and Hemerobiiformia appears to be well supported by a suite of presumed apomorphies in both male and female genitalia and in the larvae (e.g., terrestrial lifestyle, cryptonephry).

Within the Myrmeleontiformia, the Nymphidae, Myrmeleontidae, and Ascalaphidae form a reasonably well-defined group, within which the Myrmeleontidae and Ascalaphidae are very closely related and seem to have a sister relationship. In fact, there are probably no absolute criteria for separating owlfly and antlion larvae or adults because most of the distinguishing traits are shared by some members of the other family. Nevertheless, in the majority of myrmeleontid and ascalaphid species throughout the world, both adults and larvae exhibit the suite of characters that are typical of one family or the other, and so it is prudent to retain the two families until a thorough cladistic analysis is completed.

Currently, thoughts differ as to whether the Psychopsidae is the basal group of the Myrmeleontiformia or has a sister relationship with the Nemopteridae or Crocidae. Detailed studies of additional species from all three families will help resolve this issue.

Psychopsidae (Silky Lacewings)

The Psychopsidae forms a small family of approximately 26 species within five genera. They are large, attractive, and moth-like insects (forewing length 10–35 mm). The family is restricted to sub-Saharan Africa, southeast Asia, and Australia. However, fossils are known from North America, Europe, Asia, and Australia.

Very little is known about psychopsid biology. Eggs are unstalked and attached to the substrate. They are laid singly or in groups on the bark of trees, and they are covered with a secretion that is presumably derived from plant material or sand but whose origin is unknown. Larvae of Australian species are found under loose bark (Fig. 5A). There are three instars and the life cycle may take 2 years. The cocoon has two layers of silk. Adults have been collected in river valleys.

Nemopteridae (Spoon-Winged or Ribbon-Winged Lacewings)

Until recently, this small family of approximately 100 species included species now comprising the Crocidae. Both groups

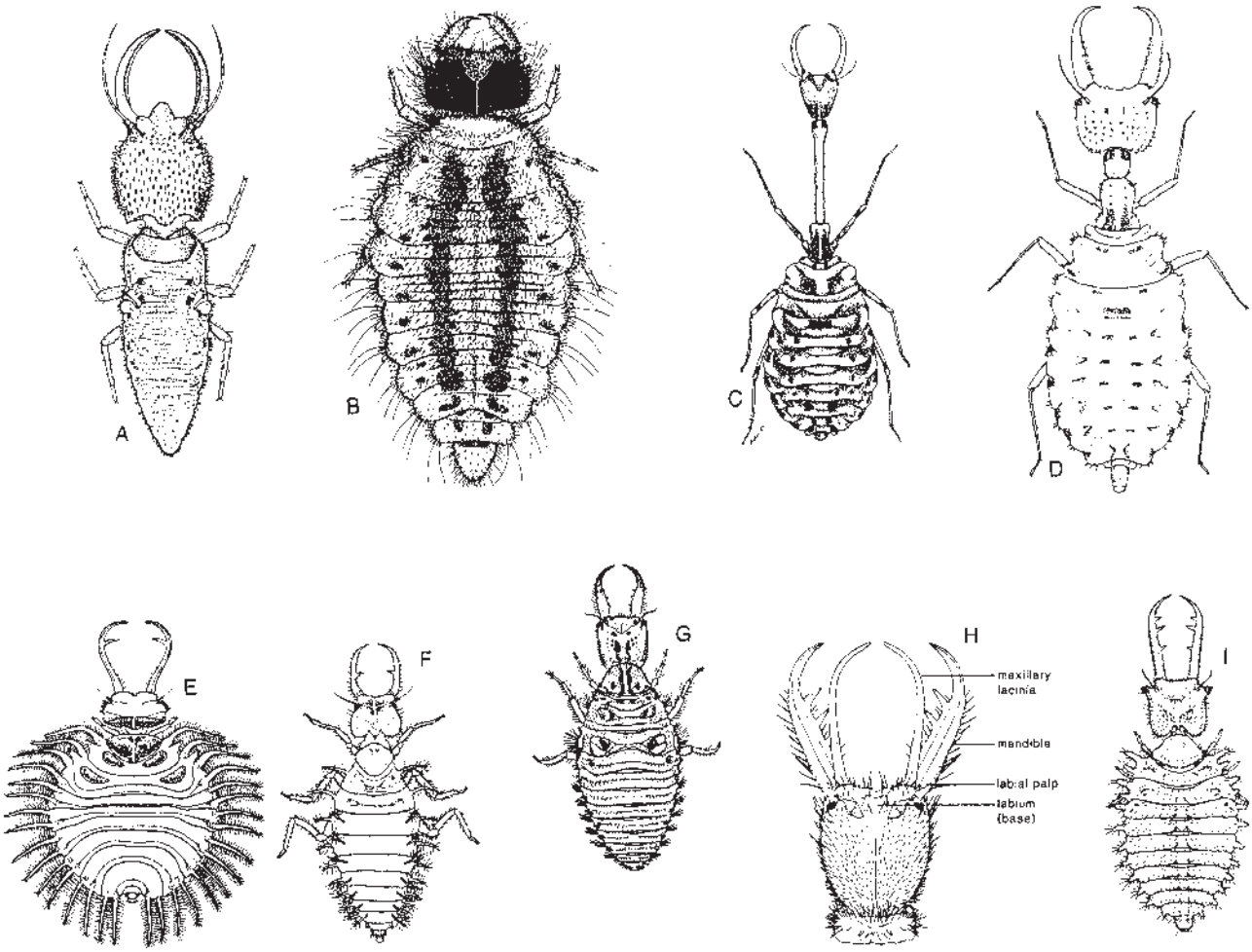

FIGURE 5 Myrmeleontiform larvae. (A) Psychopsidae; (B) Nemopteridae; (C) Crocidae; (D) Crocidae; (E) Nymphidae (Myiodactylinae); (F) Nymphidae (Nymphinae); (G) Myrmeleontidae; (H) Myrmeleontidae; (I) Ascalaphidae. (A, D–G, and I reproduced with permission, from New, 1991, © CSIRO Australia. B and C reproduced with permission, from Mansell, 1996. H modified from Tauber, 1991.)

are extraordinary looking lacewings that are easily distinguished by their remarkably elongate and narrow hind wings and by a very short metathorax. Nevertheless, there are morphological and ecological differences between the two groups that are consistent with family status. Externally, nemopterid adults are distinguished from crocid adults by their large size (forewing length 13–35 mm) and hind wings that are ribbon-like, with or without dilation, but with distinct dark areas. Nemopterid larvae are myrmeleontid-like (Fig. 5B). In contrast, adults of Crocidae have smaller bodies and their hind wings are thread-like, white, and rarely with dark shading. Crocid larvae are very unusual, especially in the elongation and sclerotization of the cervix (Figs. 5C and 5D).

The Nemopteridae occurs in the arid and desert zones of the southern borders of the West Palaearctic and West Oriental Regions, and in dry areas of the Neotropical, Afrotropical, and Australian Regions. The world's richest nemopterid fauna is concentrated in southern Africa (>60% of the species).

Although nemopterid larvae have not been collected in ant nests, myrmecophily is strongly suspected. Nemopterid eggs are unstalked, small, spherical, or ovoid, and laid singly in the sand or soil. They have a micropyle and an oviruptor (an internal, toothed structure that tears open the chorion during hatching), but the surface of the chorion is smooth and seems to lack the aeropyles that are present on crocid eggs. The eggs may be very hard and lack adhesives, and it is believed that they are harvested and introduced into ant colonies by foraging ants. Granivorous ants apparently collect the eggs.

Larvae may also be harvested by ants and those of at least some nemopterid species live in or very near ant nests. Apparently, the larvae avoid detection by the ants through chemical mimicry and also through the covering of sand that adheres to their bodies. Young larvae may burrow with the front legs and push with the other legs and they may climb to the soil surface. Mature third instars become very rotund and movement is slow and awkward. Ants appear to be the main food and it is thought that toxic materials are injected into prey, because prey become immobilized immediately after attack. Larvae can withstand long periods (up to several months) without food, and they require few prey items to

complete their development. Apparently they have very low rates of metabolism. Pupation occurs within a silken cocoon.

Adults are mostly crepuscular or nocturnal, and they feed on pollen. The modified hind wings appear to have a defensive function against aerial predators; they provide camouflage and crypsis or they give the illusion of greater size. Like myrmeleontid males, some nemopterid males have membranous sacs between tergites 5 and 6 that may emit pheromones. The elongated rostrum probably functions in collecting pollen and nectar. Adults occur during a short period of the year, perhaps in synchrony with the ephemeral burst of flowers in their inhospitable habitat.

Crocidae (Thread-Winged Lacewings)

Until very recently, this small family of approximately 50 species was included in the Nemopteridae. The two families share elongate hind wings, but crocids are distinguished by the thread-like hind wings (Fig. 1) and several larval characteristics. The distribution of Crocidae overlaps that of the Nemopteridae: arid and desert zones on the southern borders of the West Palaearctic and West Oriental Regions and dry areas from the Neotropical, Afrotropical, and Australian Regions. Crocid adults are medium to large sized (forewing length 7–15 mm).

The unstalked eggs of Crocidae differ markedly from nemopterid eggs: they have a sponge-like micropyle, the chorion has aeropyles, and no oviruptor is present. The larvae of some crocid species have elongated cervical regions (Fig. 5C); these species are frequently associated with caves or dwellings. Species that have a shorter cervix usually live in detritus, under rocks or in crevices (Fig. 5D). Crocid larvae are largely "sit-and-wait" predators and they occur beneath sand or soil surfaces; however, some may pursue prey. In both cases, they are able to survive for long periods without food, but little is known of the larval diet. Pupation occurs within a silken cocoon that incorporates sand or debris externally. Adults are either crepuscular or diurnal, and they may feed on pollen and/or nectar. The hind wings apparently play sensory and stabilizing functions, which enable the lacewings to detect vertical and horizontal surfaces and to fly in confined spaces, such as caves. They may also function in mate attraction and courtship.

Nymphidae (Split-Footed Lacewings)

This small neuropteran family is restricted to the Australian Region (Australia, New Guinea, and nearby islands); it contains about 35 species in seven genera. Currently the Nymphidae includes two well-defined lineages, the subfamilies Myiodactylinae and Nymphinae. Nymphid adults are large (forewing length 18 to >40 mm).

Nymphid eggs are laid on slender, filamentous stalks. In Myiodactylinae, the stalks are either pendant or looped, so that the egg contacts the substrate. In Nymphinae, the eggs are arranged in intricate patterns. In both subfamilies, the stalk may be coated with beads of liquid that may serve nutritional and/or defensive functions. Hatching occurs via an oviruptor. Myiodactyline larvae are very flat, and the margins of their bodies have long scoli (Fig. 5E); they are green and arboreal, and they rest on the surface of leaves with their jaws at an angle of ~180°. Nymphine larvae (Fig. 5F) occur in litter or on the bark of trees where they are camouflaged by the debris they carry or by the markings on the body. Cocoons probably are spun in sand. Adults of one species form large aggregations, but the function (e.g., mating, defense) of the aggregations is not known. Some species occur in association with acacias. Adults produce an odor from eversible abdominal glands and copulation involves enlarged and elaborate male terminalia that presumably have a grasping function.

Myrmeleontidae (Antlions, Doodlebugs)

With approximately 2100 species in 300 genera, the antlions constitute the largest neuropteran family. Members of this family have intrigued naturalists from the earliest times; imaginative accounts were summarized by W. M. Wheeler in his 1930 book, *Demons of the Dust.* Most people know antlions because the larvae of some species have pit-building habits. In truth, both the name "antlion" and the assumption that all antlions construct pits are misleading; myrmeleontid larvae do not feed exclusively on ants and most do not construct pits. Adults are slender bodied and medium sized to large (forewing length 10–70 mm).

Four subfamilies of Myrmeleontidae are generally recognized: Myrmeleoninae, Palparinae, Acanthaclisinae, and Stilbopteryginae (formerly a separate family, Stilbopterygidae). Larval morphological (Figs. 5G and 5H) and biological characteristics are crucial in the classification of the family, especially at the tribal level but also for many genera.

This family has a worldwide distribution, notably in the arid and semiarid areas of subtropical and tropical Africa, Australia, Asia, and the Americas. They inhabit open woodlands, scrub grasslands, and dry sandy areas. Efforts have been made in South Africa to assess the taxonomic richness of the large myrmeleontid fauna and to help conserve it.

Myrmeleontid eggs are unstalked and relatively large; they are laid singly in open areas or tree holes, under bushes, in caves, under rock overhangings, or in areas sheltered by buildings. The eggs are covered with a glandular secretion that may facilitate adhesion of sand or soil particles, and they lack an oviruptor.

The larvae of most species appear to be sit-and-wait predators. In most taxa they live beneath the soil surface, on trees, in tree holes, under stones, or in debris. Larvae of a very small number of genera construct pits in sand or soil where they capture insects that fall into the pits. For some of these species, pit architecture and the influence of prey availability on pit size, location, and relocation, as well as on larval growth itself, have been studied extensively.

Although the larvae of some myrmeleontid species can travel quickly across the surface of the sand, others have slow, creeping movements or fast backward movements under the

sand that are aided by the forward-directed terminal segments (fused tibia–tarsus) of the hind legs. These patterns of behavior have led to the common name, "doodlebugs." Ingestion is accomplished by the injection of digestive enzymes from the midgut into the prey, followed, after several minutes, by sucking. The regurgitated gastric juice is not mixed with the contents of the crop or the midgut. Instead, it seems to be extruded from the space between the peritrophic membrane and the epithelium of the midgut; then contractions of the crop's muscular system transport the fluid forward through a fold in the wall of the crop.

Larvae pass through three instars. In some species, larval development can be protracted over several years depending upon the availability of prey; overwintering occurs in the larval stage. Univoltine or semivoltine life cycles and the synchrony of adult emergence are maintained by photoperiodic and thermal responses during early and late larval stages and by the physiological effect of pupal body size. Larvae generally spin a single-walled cocoon; a double-walled cocoon occurs in one unusual South American species.

Adults are largely nocturnal and presumed predaceous. Their flight resembles that of a damselfly. Sexual communication involves the extrusion of "hair pencils" and abdominal glands or sacs on the male (analogous to those found in lepidopteran males), as well as the release of volatile substances from thoracic glands in both males and females.

Ascalaphidae (Owlflies)

The owlflies constitute a medium-sized neuropteran family of about 400 species assigned to 65 genera. There are three subfamilies: one with bisected eyes (Ascalaphinae) and two with entire eyes: Haplogleniinae and Albardiinae. The subfamily Albardiinae consists of a single, very unusual, large-bodied Brazilian species.

The two major subfamilies (Ascalaphinae and Haplogleniinae) are widely distributed in warm regions of the world, but of the two only the Haplogleniinae occurs in Australia. Ascalaphids inhabit grasslands and warm dry woodlands. Most species are nocturnal or crepuscular, but some Eurasian species are diurnal and have pigmented wings that resemble those of butterflies. Adults are relatively large (forewing length 15–60 mm).

Clusters of 20 to 75 large, unstalked eggs are laid on twigs in spirals or rows. Individual eggs are reported to have two micropyles, but lack an oviruptor. Females of many species place small, modified eggs (repagula) on or around egg batches; these repagula reportedly have defensive and nutritional functions—they serve to repel predators and/or provide food for newly hatched larvae. The two major subfamilies in the New World possess this habit, but it is absent from Old World and Australian ascalaphids.

Newly hatched larvae often remain together near the egg cluster for a week or more before dispersing. Larvae are either terrestrial (in the soil or litter) or arboreal (on leaves or tree trunks), and most seem to be sit-and-wait predators (Fig. 5I). Characteristically, the larvae hold their jaws open at very wide angles; some New World species are able to open their jaws beyond 270°. When prey makes contact with the larva, the jaws can close very rapidly and larvae can take relatively large prey. Considerable evidence shows that the larvae paralyze their prey with toxins from the midgut, not from glands. As in other myrmeleontiform larvae, there are three instars.

Larvae of some species that live in the soil or sand camouflage themselves with sand grains or debris. Such behavior shares features with the "camouflaging" behavior of chrysopid larvae, but ascalaphid larvae use their flexible foretarsi, rather than the jaws, to place material on their dorsa. A thick mat of tangled threads anchors the debris to the dorsal surface.

Second and third instars resist starvation well, and development may be extended for 1 or 2 years. Because ascalaphid adults typically occur at very specific times of the year, diapause probably intervenes in some larval stages. Photoperiod or other factors may be involved in regulating the occurrence of diapause, but the responses and mechanisms have not been studied. Pupation occurs within silken cocoons that are spun either on the ground or on trees, sometimes incorporated with sand or debris.

Adults remain motionless in a characteristic head-downward position for most of the day; flight is restricted to a relatively short period around dusk and is preceded by ~10 min of muscle-warming via wing vibration. Compared to that of other neuropterans, the flight of ascalaphids is very strong and agile and similar to that of dragonflies. Adults attack and feed on large numbers of flying insects (e.g., caddisfly adults); prey capture and mating occur on the wing.

SUBORDER HEMEROBIIFORMIA

The Hemerobiiformia is currently considered to be a monophyletic but heterogeneous grouping having a sister relationship with the Myrmeleontiformia. It contains 10 families whose larvae are diverse in lifestyles and structure (Fig. 6), but which share a number of distinguishing features. For example, hemerobiiform larvae have a head capsule whose posteroventral region is composed primarily of the maxillae, cardines that are elongate, and a cervix that is cushion-like.

Within the Hemerobiiformia, two groupings emerge. First, Ithonidae + Polystoechotidae appear as a basal sister group. Among the Neuroptera, ithonid and polystoechotid larvae are unique in that they feed by sucking on plant tissues. Given the close relationship between *Rapisma,* which has been designated a separate family (Rapismatidae), and *Adamsiana,* a newly described genus of Ithonidae from Honduras, it is difficult to justify retaining the family Rapismatidae. Thus, we include both genera in the Ithonidae while awaiting discovery and study of their larvae and a thorough phylogenetic analysis of the Neuroptera.

The second hemerobiiform grouping harbors the remaining families, all of which are carnivorous in the larval stages. Within

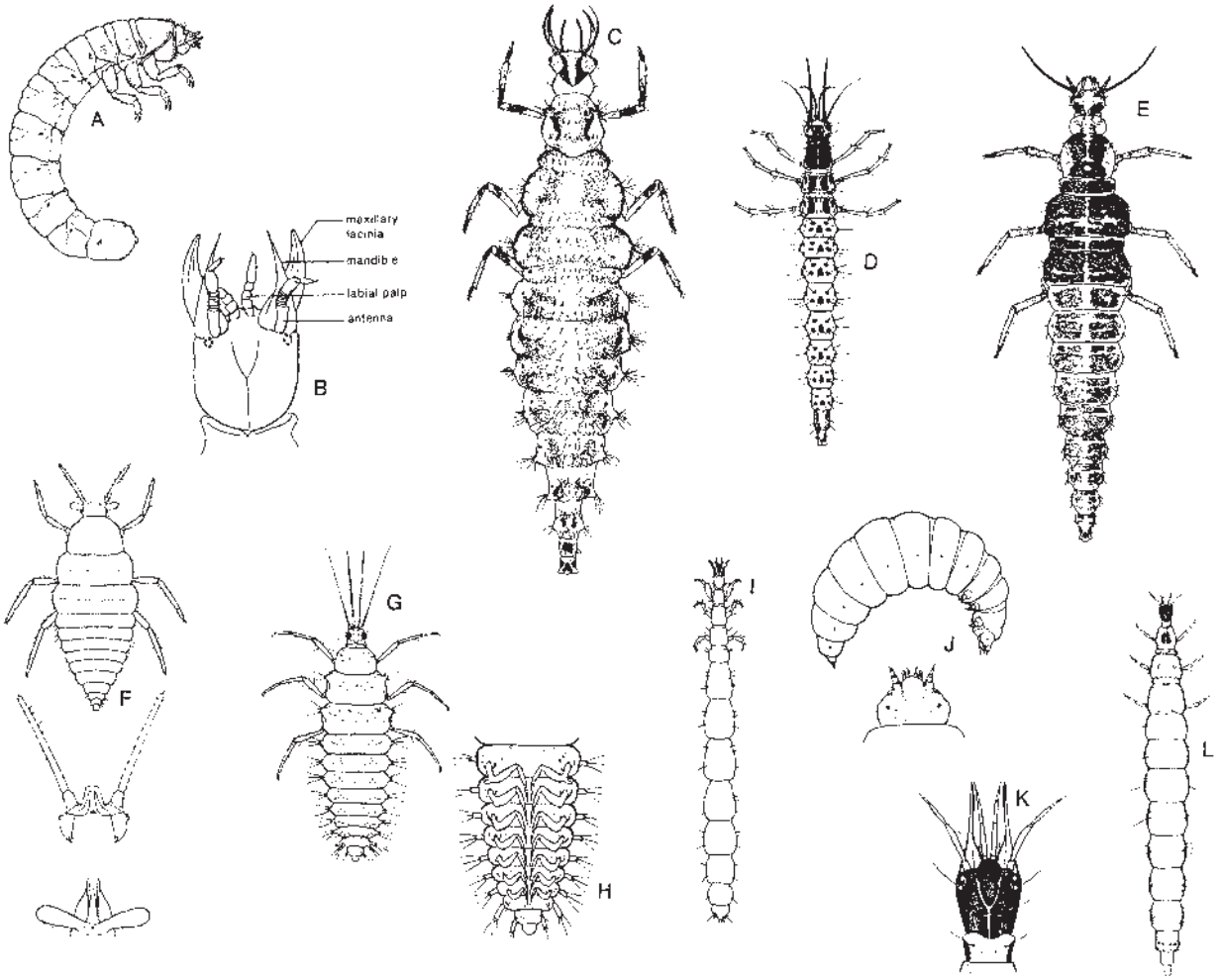

FIGURE 6 Hemerobiiform larvae. (A) Ithonidae; (B) Polystoechotidae; (C) Chrysopidae; (D) Osmylidae; (E) Hemerobiidae; (F) Coniopterygidae; (G) Sisyridae; (H) Sisyridae (venter of abdomen; note gills); (I) Dilaridae; (J) Mantispidae; (K) Berothidae; (L) Berothidae. (A, D, F–H, and J–L reproduced, with permission, from New, 1991, © CSIRO Australia. B and I modified from Tauber, 1991. C reproduced from Tauber, 1974. E reproduced, with permission, from Tauber and Krakauer, 1997.)

this grouping, the Chrysopidae and Osmylidae appear to form a sister group, although the evidence for this relationship is not strong. In turn, these two families emerge as a sister group to the poorly defined group of six families that remain. The reasonably well-defined dilarid lineage [Dilaridae, Mantispidae, and Berothidae (including the Rhachiberothinae)] apparently has a sister relationship with the Coniopterygidae–Sisyridae sister group, but the relationship is not strongly supported. Finally, the Hemerobiidae emerge as a sister-group to the dilarid + (coniopterygid + sisyrid) lineages, but again the relationship is tentative.

Ithonidae (Moth Lacewings)

This very small family includes approximately 32 species in three genera from Australia, southern Asia, southwestern United States, Mexico, and Central America. Adults are large and moth-like (forewing length 15–30 mm); they share a variety of plesiomorphic characters, but few apomorphic characters are known. The larvae are subterranean and scarabaeiform (Fig. 6A). The abdomen is very large and swollen; the legs are short and fossorial. The short mandibles curve inward and slightly upward. The maxillae are broad and robust; the mandibles are narrow. Eyes are absent.

Little is known about ithonid biology. Unstalked eggs are laid singly in the soil, where their adhesive surface gathers soil and sand particles. The larvae of one species in Australia are associated with the roots of *Eucalyptus* trees and a North American species occurs near creosote bushes. The specific food source of these larvae (plant, mycorrhizae or other fungi, or associated herbivores) is not known. In one Australian species, five instars have been demonstrated. Larval development may take 2 or more years; apparently, mature larvae undergo diapause. Pupation occurs within silken cocoons in the soil and adults emerge synchronously in large numbers, usually following a period of rainfall. Males emerge first and form aggregations that attract females; adults live for only a few days.

FIGURE 7 Despite their common name, green lacewings, adults of many species of Chrysopidae, like this *Nothochrysa* adult, are darkly colored. (Photograph by E. S. Ross, California Academy of Sciences.)

Polystoechotidae (Giant Lacewings)

The Polystoechotidae constitutes an extremely small family (four species) of lacewings that are characterized by a very large body and long wings (forewing length 15–40 mm). The adults resemble ithonids, but they have larger wings, more complex wing venation, and genital differences. This family is restricted to the New World.

Eggs are like those of ithonids: large, unstalked, and covered with sticky material. Only the first instar of *Polystoechotes* is known. Unlike the scarabaeiform ithonid larvae, this larva is hemerobiid-like: it is elongate and has relatively long legs, a short multisegmented antenna, and robust jaws (Fig. 6B). Little is known about polystoechotid biology; phytophagy is suspected but not confirmed. It is noteworthy that populations of one species in North America have experienced large reductions in their numbers since the 1900s. Thus, lack of knowledge concerning this family is especially unfortunate.

Chrysopidae (Green Lacewings)

Chrysopidae, with its approximately 1200 recognized species, is one of the largest families of Neuroptera, second only to the Myrmeleontidae. The larvae of many chrysopid species feed on insect and mite pests of agricultural crops or horticultural plantings, and because of their importance in biological control, chrysopids are the most frequently studied of the Neuroptera.

Adults are medium-sized to large, delicate insects with four subequal wings (forewing length 6–35 mm) and relatively long filiform antennae. In most species the adults are green with large golden eyes, but some species have black, brown, or reddish adults (Fig. 7). Larvae vary in shape and habits; some are voracious, active, and free-living general predators (thus the name "aphis-lions") (Fig. 6C); others are slower moving, cryptic, trash-carrying predators that are associated with specific types of prey or habitats; still others live in ant nests where they feed on the inhabitants.

Currently, the Chrysopidae comprises three subfamilies (Nothochrysinae, Apochrysinae, and Chrysopinae), all of which are based strictly on adult morphological characteristics and none of which are well defined. Comparative studies of the larval morphology and biology are needed to clarify the systematics of the group and to facilitate the identification and use of these predators in biological control.

The Nothochrysinae includes only a small number (9) of genera; it is believed to be the basal chrysopid lineage. Defining characteristics may occur in the larval stages, but larvae from very few genera are known. Apochrysinae and Chrysopinae are probably not monophyletic. Apochrysinae has about 13 genera that are based largely on somewhat variable characters in the wing venation; this subfamily contains the largest and visually most spectacular green lacewings. The larvae of one apochrysine species have been described, but distinguishing subfamilial traits were not apparent. The large subfamily Chrysopinae encompasses over 97% of the known chrysopid species; it includes about 60 genera distributed among four poorly defined tribes. The tribe Chrysopini contains almost all of the lacewings of economic importance.

As a group, the Chrysopidae is cosmopolitan; similarly, all of the subfamilies are widely distributed. Nevertheless, many of the genera are restricted to small regions of the earth. For example, most genera of Nothochrysinae are endemic to very small geographic ranges; many species are known solely from a very few specimens. Among the Apochrysinae, one genus occurs only in South Australia, two only in Central and South America, and one only in the Oriental Region. Two other genera are widespread. The chrysopine genera range from cosmopolitan to narrowly endemic.

There are two basic larval forms: debris-carrying and naked. The debris carriers construct and carry large packets of exogenous material (e.g., plant parts, exuviae, waxy secretions, or remains of prey) on their dorsal surfaces. Usually, they have humped (gibbous) abdomens with rows of hooked setae and long thoracic tubercles bearing numerous, long setae. In contrast, naked larvae have more or less flat abdomens and short, straight setae. The thoracic tubercles are reduced in size and also bear short setae. In the most extreme cases, the lateral tubercles are absent. Given the wide range of morphological and behavioral variation among chrysopid larvae, it is clear that inclusion of all life stages is crucial for advancing the systematics of the family. However, except for the European and Japanese faunas for which larvae of approximately 80% of the species are described, the world's chrysopid larvae are poorly known.

Typically, chrysopid eggs are laid at the end of long stalks, singly, in groups, or in clusters, with the stalks loosely or tightly intertwined. The egg stalks can be naked or they may bear oily droplets; the droplets may contain nutrients or defensive substances that protect the egg or the newly hatched larva from natural enemies.

Chrysopid larvae feed on a variety of soft-bodied arthropods; they may be generalist predators or they may have very strong association with a particular type of prey. For example, prey specialization in *Chrysopa* is based on a number of intrinsic and extrinsic factors, including maternal oviposition behav-

ior, larval size, morphology and behavior, prey influence on life-history traits, responses to natural enemies that are associated with specific prey, and phenology.

Adults of some chrysopid genera are predaceous. Those of other genera take honeydew and pollen; in this group the dorsal crop diverticulum has numerous tracheae and is filled with symbiotic yeast that aid in digestion. Chrysopid adults are not strong flyers; nevertheless, they are known to move considerable distances with the wind. Adults of some species emit foul-smelling defensive odors when they are disturbed.

Overwintering may occur in the larval, prepupal, or adult stages; the overwintering stage is a generic characteristic. Usually a photoperiodically induced diapause is involved.

Chrysopine lacewings have two types of hearing. The "ear" (tympanal organ) is located at the base of the radial vein in each forewing. It is the smallest tympanal organ known, and functions in the perception of ultrasound signals produced by bats that prey on small flying insects. When a lacewing perceives ultrasound signals that are emitted at low rates (1–50 pulses per second), flight ceases; this response causes the lacewing to begin falling. As the bat continues to approach, its signal increases in frequency; the high-frequency signal causes the lacewing suddenly to flip its wings open and fly, thus aiding its escape. The second type of hearing, the perception of low-frequency, substrate-borne sounds that are emitted during courtship, is accomplished through scolopidial organs in the legs. Such sounds are an integral part of courtship in a number of *Chrysoperla* species; both species-specific and geographic variations in the production of these sounds appear to be considerable.

The endemic complex of green lacewings on the Hawaiian Islands belonging to the genus *Anomalochrysa* has evolved several unique characteristics and exhibits an extraordinary range of variation in morphology and behavior. For example, unlike any other known chrysopids, *Anomalochrysa* females lay sessile (unstalked) eggs, either singly or in batches. Larvae range in body form from fusiform with greatly reduced lateral tubercles and few, short setae, to flattened with well-developed lateral tubercles and numerous, robust, long setae. In continental lineages, such broad variation is found only among genera. In some species, adults or larvae are very bright and colorful; in others they are cryptic and polymorphic. Males and females may produce conspicuously loud clicking sounds during courtship and mating; how these sounds are perceived is unknown.

A number of species in the genus *Chrysoperla* are mass reared and released for use in the biological control of agricultural and horticultural pests. Among those in North America are *Chrysoperla carnea* and *Chrysoperla rufilabris.* These species possess a number of characteristics that are excellent for mass rearing, including the ability of adults to use artificial diets for reproduction and to be stored for long periods without significant loss of reproductive potential and the ability of larvae to feed on artificial or factitious prey. Larvae of *Ceraeochrysa* species, which are trash carriers, share many of these traits and are also excellent candidates for mass production and release. They have the added advantage of being camouflaged and thus protected from their own natural enemies, e.g., ants. The role and efficacy of lacewings in pest management are beginning to be evaluated quantitatively under field conditions, but additional studies in a variety of crops are needed.

Osmylidae

Although the evidence for a relationship is weak, this family of 160 species may be the sister taxon of Chrysopidae. Eight poorly defined subfamilies of osmylids are recognized, but their systematics needs considerable reassessment.

Osmylids are slender, moderate-sized lacewings (forewing length 15–30 mm), with broad pigmented wings. The family is distributed over much of the Old World. Five of the subfamilies occur in Australia and two in South America, but apparently osmylids are absent from North America.

Little is known of osmylid biology. Elongate, knobbed, unstalked eggs are laid with their sides attached to foliage. Larvae live under stones or at the water–land interface near streams or under the loose bark of trees. Osmylid larvae have long slender stylets like those found in sisyrids (and berothids) (Fig. 6D), but unlike sisyrids, they lack gills and breathe through thoracic and abdominal spiracles.

Hemerobiidae (Brown Lacewings)

The third largest neuropteran family, with approximately 550 species, Hemerobiidae constitutes a cosmopolitan clade that is relatively well known and easily recognized. Adults are generally small (forewing length 3–18 mm), brown, and inconspicuous. The approximately 27 extant genera of hemerobiids fall into 10 reasonably well-defined subfamilies. The Carobiinae and Psychobiellinae each consist of one genus that is restricted to the Australian Region. The Hemerobiinae, Sympherobiinae, Notiobiellinae, and Microminae each include 3 to 5 genera; all four subfamilies are cosmopolitan, although some of the small genera that they encompass have very restricted distributions. The Drepanacrinae and Drepanepteryginae each contain three genera with restricted distributions, and the Megalominae comprises one genus with broad distribution. The most recently described subfamily, Adelphohemerobiinae, consists of one genus that is known only from South America.

Despite the fact that hemerobiid larvae offer a rich suite of traits for phylogenetic analysis, the larvae of only nine genera (from 7 of the 10 subfamilies) have been described. There are three instars. In the first instar, body setation is sparse, and trumpet-shaped empodia are present between the tarsal claws. Second and third instars are similar to each other except in size; they may have numerous short setae, and their empodia are short (Fig. 6E).

Mainly because the systematics of the family was neglected until recently, the life cycles of relatively few hemerobiid genera are known, and the groups that have been studied occur

largely in the Northern Hemisphere. In general, hemerobiid eggs are sessile (unstalked) and laid singly or in clusters. Hatching is accomplished by means of an oviruptor.

Larvae prey upon a variety of small, soft-bodied arthropods and eggs. Little is known about the range of larval diet or its specificity, but some species show strong association with a particular type of plant or habitat. Pupation occurs within thinly spun cocoons. Pupae have a peculiar set of hooks on the dorsum of the abdomen; their function is unknown. Most species seem to be predaceous in the adult stage, but there are records of extensive honeydew feeding by adults. Life cycles range from univoltine to multivoltine, but for most taxa the overwintering stage is unknown.

Flightlessness has evolved several times in the Hemerobiidae; it is largely confined to species that occur on islands or are restricted to isolated mountains. In flightless forms, the hind wings are greatly reduced or absent, or the forewings are hardened or fused. Modifications associated with flightlessness are probably most extremely exhibited in the endemic Hawaiian *Micromus.* Spectacular sculpturing of the wings also occurs in winged (and flighted) endemic Hawaiian *Micromus* species.

Many species of Hemerobiidae are believed to be important natural enemies of insect pests on agricultural and horticultural crops or in forests. Hemerobiids often are active at relatively low temperatures; thus they can be useful as biological control agents in temperate regions early in the season when other natural enemies remain inactive.

Coniopterygidae (Dustywings)

Because of their smallness and cryptic nature, coniopterygids are generally overlooked and thus considered rare. However, with approximately 450 species, the Coniopterygidae constitutes a relatively large family and is perhaps the best known systematically. Although they clearly belong within the Neuroptera, coniopterygids differ in a number of ways from other neuropteran families. Previously, they were considered the sole family of a separate primitive suborder (superfamily), the Coniopterygoidea. However, a recent cladistic analysis provides some evidence that the Coniopterygidae and Sisyridae may form a derived sister group within the Hemerobiiformia.

The Coniopterygidae is generally a very homogeneous family characterized by very small adults (forewing length 2–5 mm) with bodies covered by white waxy ("dusty") secretions. The secretions originate from hypodermal wax glands on the sternites and tergites of the abdomen and are spread over the body by the hindlegs. Other than in the coniopterygids, such glands are found only in the homopterans Aleurodina and Coccina. This similarity represents a remarkable example of convergent evolution especially because coniopterygids frequently are associated with these waxy homopterans.

Currently, the Coniopterygidae contains three well-defined, probably monophyletic subfamilies: Coniopteryginae, Aleuropteryginae, and Brucheiserinae. Both the Coniopteryginae and the Aleuropteryginae are large groups with cosmopolitan distributions. They are distinguished from the Brucheiserinae by their predominantly longitudinal wing venation, whereas the Brucheiserinae have highly unusual reticulate wing venation. Brucheiserinae is known only from the neotropics and its larvae are not described. Some authors have considered it to comprise a separate family (Brucheiseridae), but this distinction is probably not justified. In this regard, discovery of the larvae is likely to prove very valuable.

The life histories of very few coniopterygid species are known. Both larvae and adults occur on trees and bushes (sometimes on low vegetation). Many species appear to be associated with specific types of vegetation, and this habit may indicate a degree of prey specialization. Eggs are unstalked and laid near prey; there are three or four instars (Fig. 6F).

Adults and larvae are predaceous on small, soft-bodied arthropods (aphids, scales, mites); adults probably also feed on honeydew and perhaps pollen. Flat cocoons with double walls are spun on foliage or tree trunks. Adults are usually active at dusk or at night. Life cycles and overwintering stages vary (prepupae within cocoons, free-living second instars) and have not been well studied.

Many species of coniopterygids are considered important natural biological control agents, but unfortunately, their role has not been evaluated. Other species have great potential for use in classical biological control, as well as in commercial mass production and release. To date, their potential remains undeveloped.

Sisyridae (Spongillaflies)

The Sisyridae constitutes a small but cosmopolitan family that contains about 50 species in four genera: *Climacia,* which is restricted to the New World; *Sisyra,* which is cosmopolitan; and *Sisyrina* and *Sisyrella,* which are small Australian and Asian genera. This is the only neuropteran family with truly aquatic larvae. Sisyrid larvae are believed to feed exclusively on freshwater sponges, and they are unique among the Neuroptera in having segmented abdominal gills that function in breathing (Fig. 6H). Adults are dull colored and relatively small (forewing length 4–10 mm). They closely resemble brown lacewings, but the simple, open venation of the forewing and the branching pattern of the radial sector distinguish them.

The sessile eggs of sisyrids are laid singly or in groups on objects that overhang water. A flat layer of silk covers the eggs. Hatching is aided by an oviruptor, and subsequently the neonate larvae walk or drop to the water where they seek a sponge colony.

Sisyrid larvae probe sponges by means of the long, flexible mouthparts (Fig. 6G). Only one of the Malphigian tubules is attached distally to the rectum, a condition that is probably related to the aquatic lifestyle. After feeding and development, the mature larvae swim to the shore, attach to objects close to the water, and spin a double-layered cocoon within which they pupate. Adults forage on nectar, pollen, algae, fungi, aphids, and mites.

Dilaridae (Pleasing Lacewings)

This small family of approximately 50 species has well-defined affinities with the Berothidae and Mantispidae. It is divided into two subfamilies: Dilarinae, which is confined to the Old World, and Nallachiinae, which is restricted to the New World (with one species known from South Africa). It is one of the few neuropteran families that does not occur in the Australian region.

Adults resemble small, delicate hemerobiids (forewing length 3–16 mm in males and 5–22 mm in females). But, they are differentiated by ocelli-like tubercles on the head of both sexes (functional ocelli are absent), a long ovipositor in females, and pectinate antennae in males.

Dilarid eggs are elongate and unstalked. Those of *Nallachus* are laid in association with dead trees. The larvae of *Nallachus* inhabit the galleries of insects in decaying logs or the area beneath the tightly adhering bark of erect, recently dead trees (Fig. 6I). Larvae of *Dilar* have been found in the soil.

The larval diet is not known, but it does not appear to be highly specific. Development probably takes 1 year. Larvae may undergo supranumerary molts; i.e., if undernourished, they may continue to molt as many as 12 times. However, those that did so under laboratory conditions did not metamorphose successfully. Pupation occurs within cocoons.

Mantispidae (Mantidflies)

With approximately 400 species, Mantispidae is the largest family in the dilarid lineage (Dilaridae, Berothidae, Mantispidae). Adults are recognized by raptorial forelegs that resemble those of a mantid (Fig. 8); their simple, subequal wings are narrow and elongate and they have a distinct pterostigma and chrysopid-like venation. They are moderate-sized to large lacewings (forewing length 5–30 mm). Larvae are similar to those of the Berothidae in that they are hypermetamorphic; however, in the case of the mantispids the first instars are campodeiform and the second and third instars are grub-like (Fig. 6J). The mantispids differ from the Rhachiberothinae in their wing venation, terminalia, and larval characteristics, but they are of similar size and also have raptorial forelegs.

FIGURE 8 Mantispid adult in the genus *Plega*. Note the mantid-like raptorial forelegs and narrow forewings. (Photograph by E. S. Ross, California Academy of Sciences.)

The family contains four apparently monophyletic subfamilies: Symphrasinae, Drepanicinae, Calomantispinae, and Mantispinae. Symphrasinae encompasses a large, diverse assemblage of species that are distributed throughout South America and southern North America. Drepanicinae is a smaller subfamily that occurs in restricted areas within South America and mainland Australia. Calomantispinae, another small subfamily, has a disjunct distribution: eastern Australia (including Tasmania) and southern North America. Finally, the large subfamily Mantispinae ranges between 50°N and 45°S throughout much of the world. The biology and immatures of Drepanicinae and Calomantispinae remain largely unknown, whereas those of several genera in Symphrasinae and Mantispinae have been studied.

Larvae in the subfamily Mantispinae usually inhabit the egg sacs of spiders where they feed upon the contents, although some may be subterranean predators or possibly generalist predators. Numerous (200–2000) stalked eggs are laid randomly (and sometimes communally) on leaves and wooden structures.

The newly hatched campodeiform larvae find their hosts (spider eggs) via one of two methods. Either they actively seek a previously constructed spider egg sac that they enter through direct penetration or they climb onto a female spider and enter the egg sac as the female builds it. Those that board spiders can feed on the hemolymph of their host, but they do not molt until they enter an egg sac. If the campodeiform larva attaches to a male spider, it may transfer to a female during copulation. After an egg sac is found, the larva feeds on the contents (predation) and undergoes hypermetamorphosis.

The mature larva spins a cocoon, and pupation occurs within the larval skin, within the cocoon. Adults are predaceous and they are active during the day or night. Overwintering in some species occurs in the first instar, and there may be one to several generations per year.

Several species within the Symphrasinae have been reared from nests of aculeate Hymenoptera or reared in the laboratory on larvae or pupae of Lepidoptera, Coleoptera, and Diptera. First instars may find their hosts by attaching to an adult bee or wasp and moving into the cell when the egg is laid. Subsequently, they feed on a single host (parasitism), have the typical mantispid hypermetamorphosis, and adhere to their host with a sticky, yellow secretion.

Adult *Climaciella* can be highly polymorphic, with each of several morphs mimicking a different species of polistine wasp. The proportion of the different morphs may vary depending on the number and aggressiveness of the various wasp species at each locality.

The larval habits of the Drepanicinae are unknown; Calomantispine larvae may be generalist predators.

Berothidae (Beaded Lacewings)

The Berothidae is a small family of approximately 115 species in four more or less distinct subfamilies: Rhachiberothinae, Berothinae, Nosybinae, and Cyrenoberothinae. Adults typi-

cally are small to medium-sized (forewing length 6–15 mm) lacewings with brown wings and bodies. Frequently, the outer margin of the forewing is deeply incised. Larvae are associated with termite nests (Figs. 6K and 6L). The family is cosmopolitan and occurs predominantly in warm temperate, tropical, and subtropical areas; the greatest number of species is found in Africa.

One subfamily of berothids, the Rhachiberothinae, differs considerably from the others. Its adults have raptorial forelegs that are very similar to those of mantispids. As a result, some authors consider them to be a separate family. However, because of the close similarity between the few known larvae of Rhachiberothinae and the few known larvae of other Berothidae, we treat the Rhachiberothinae as a specialized subfamily of the Berothidae, while awaiting further study.

Little is known of the biology of the Berothidae except for the North American genus *Lomamyia* (Berothinae). Eggs of three subfamilies are laid in clusters attached to one or several long, silken stalks; those of the fourth subfamily, Rhachiberothinae, are sessile. *Lomamyia* eggs are laid on dead trees or logs that contain colonies of subterranean termites. First instars are mobile and after entering the termite colony they feed and molt. The second instar does not feed, but hangs immobile from the roof of the termite tunnel. Third instars resume feeding on termites. They subdue their prey with an allomone that is emitted from the tip of the abdomen and/or a neurotoxin that is injected through the mouthparts. Pupation occurs within silken cocoons that are spun in the termite nest. Adults are primarily nocturnal. As many as three generations can occur per year; the prepupal stage overwinters.

See Also the Following Articles

Hypermetamorphosis • Megaloptera • Raphidioptera

Further Reading

Aspöck, U., Plant, J. D., and Nemeschkal, H. L. (2001). Cladistic analysis of Neuroptera and their systematic position within Neuropterida (Insecta: Holometabola: Neuropterida: Neuroptera). *Syst. Entomol.* **26,** 73–86.

Brooks, S. J., and Barnard, P. C. (1990). The green lacewings of the world: A generic review (Neuroptera: Chrysopidae). *Bull. Br. Mus. Nat. Hist.* **59,** 117–286.

Brushwein, J. R. (1987). Bionomics of *Lomamyia hamata* (Neuroptera: Berothidae). *Ann. Entomol. Soc. Am.* **80,** 671–679.

Canard, M., Séméria, Y., and New, T. R. (1984). "Biology of Chrysopidae." Junk, The Hague.

Hagen, K. S., Mills, N. J., Gordh, G., and McMurty, J. A. (1999). Terrestrial arthropod predators of insect and mite pests. *In* "Handbook of Biological Control" (T. S. Bellows and T. W. Fisher, eds.), pp. 383–503. Academic Press, San Diego.

Mansell, M. W. (1996). Unique morphological and biological attributes: The keys to success in Nemopteridae (Insecta: Neuroptera). *In* "Pure and Applied Research in Neuropterology." Proceedings of the Fifth International Symposium on Neuropterology, Cairo (M. Canard, H. Aspöck, and M. W. Mansell, eds.), pp. 171–180. M. Canard, Toulouse.

McEwen, P., New, T. R., and Whittington, A. E. (eds.). (2001). "Lacewings in the Crop Environment." Cambridge University Press, Cambridge, UK.

Meinander, M. (1990). The Coniopterygidae (Neuroptera, Planipennia). A check-list of the species of the world, descriptions of new species and other new data. *Acta Zool. Fennica* **189,** 1–95.

Monserrat, V. J. (1996). Larval stages of European Nemopterinae, with systematic considerations on the family Nemopteridae (Insecta, Neuroptera). *Dtsch. Entomol. Z.* **43,** 99–121.

New, T. R. (1991). Neuroptera. *In* "The Insects of Australia," Vol. I, pp. 525–542. Melbourne University Press, Carlton, Victoria.

Oswald, J. D. (1993). Revision and cladistic analysis of the world genera of the family Hemerobiidae (Insecta: Neuroptera). *J. N. Y. Entomol. Soc.* **101,** 143–299.

Stange, L. A., and Miller, R. B. (1990). Classification of the Myrmeleontidae based on larvae (Insecta: Neuroptera). *In* "Advances in Neuropterology." Proceedings of the Third International Symposium on Neuropterology, Berg en Dal, Kruger National Park, Republic of South Africa (M. W. Mansell and H. Aspöck, eds.), pp. 151–169. Department of Agricultural Development, Directorate of Agricultural Information, Pretoria.

Tauber, C. A. (1974). Systematics of North American chrysopid larvae: *Chrysopa carnea* group (Neuroptera). *Can. Entomol.* **106,** 1133–1153.

Tauber, C. A. (1991). Neuroptera. *In* "Immature Insects" (F. W. Stehr, ed.), Vol. 2, pp. 126–143. Kendall–Hunt, Dubuque, IA.

Tauber, C. A., and Krakauer, A. H. (1997). Larval characteristics and generic placement of endemic Hawaiian hemerobiids (Neuroptera). *Pacific Sci.* **51,** 413–423.

Tauber, M. J., Tauber, C. A., Daane, K. M., and Hagen, K. S. (2000). Commercialization of predators: Recent lessons from green lacewings (Neuroptera: Chrysopidae: *Chrysoperla*). *Am. Entomol.* **46,** 26–38.

Nomenclature and Classification, Principles of

F. Christian Thompson

Systematic Entomology Laboratory, U.S. Department of Agriculture

Classification has two meanings in English: the process by which things are grouped into classes by shared characters and the arrangement of those classes. Identification is the process of observing characters and thereby classifying things. Biological classifications are arrangements of organisms. The ability to classify is common to all animals, for to survive animals must group other organisms into at least three classes: Those to be eaten, those to be avoided, and those to associate with, especially members of their own class. For scientists, classification is formalized into a nested or hierarchical set of hypotheses: hypotheses of characters, groups (taxa), and relationships among the groups. Individual specimens of organisms are observed and characteristics noted. So, for example, we may observe that some people are black, others yellow or white, and conclude as Linnaeus did that there are different groups of humans *(Homo sapiens)*. This is a hypothesis that skin color is a useful character. Further testing of this character hypothesis has shown that skin color among humans does not delimit natural groups; hence we reject skin color as a character for humans as well as those groups this character defines. Color, however, is a very useful character for classifying many other groups. Then there is the hypothesis of a group. Groups

of biological organisms are called taxa (taxon, singular) and these taxa are hierarchically arranged in our classifications. Taxa in a classification have rank, with the basic (basal, bottom) rank being designated as species. Some of the higher ranks are genus, tribe, family, order, class, phylum, and kingdom.

Nomenclature is a system of names along with the procedures for creating and maintaining that system. Classification, in its second definition, is the structure for nomenclature, being the model on which names are arranged. Names form the essential language of biology and are the means that we use to communicate about our science. To avoid a Tower of Babel, a common system of nomenclature is required, especially an effective, efficient, system that has a minimal cost.

NAMES AND CLASSIFICATION

Names are tags. Tags are words, short sequences of symbols (letters) used in place of something complex, which would require many more words to describe. Hence, tags save time and space. Instead of a long description, we use a short tag. A scientific name differs from a common name in that the scientific name is a unique tag. In nonscientific languages, such as English, there may be multiple tags (common names) for the same organism. For example, imagine the various words in English that are used to describe *H. sapiens.* In computer (database) jargon, data elements that are used to index information are termed keys, and keys that are unique are called primary keys. Scientific names are primary keys. The word "key" has another meaning in English, which is "something that unlocks something." Scientific names are those critical keys that unlock biosystematic information, which is all that we know about living organisms. To repeat: scientific names are tags that replace descriptions of objects or, more precisely, concepts based on objects (specimens). Scientific names are unique within a classification, there being only one valid scientific name for a particular concept, and each concept has only one valid scientific name.

Scientific names are more than just primary keys to information. They represent hypotheses. To most systematists, this is a trivial characteristic that is usually forgotten and thereby becomes a source of confusion later. To most users, this is an unknown characteristic that prevents them from obtaining the full value from scientific names. If a scientific name were only a unique key used for storing and retrieving information, it would be just like a social security number. *H. sapiens* is another unique key used to store and retrieve information about humans, but that key also places that information into a hierarchical classification. Hierarchical classifications allow for the storage at each node of the hierarchy of the information common to the subordinate nodes. Hence, redundant data, which would be spread throughout a nonhierarchical system, are eliminated. Biological classifications, however, do more than just hierarchically store information. If one accepts a single common (unique) history for life (phylogeny) and agrees that our biological classifications reflect this common history in their hierarchical arrangement, then biological classifications allow for prediction, namely that some information stored at a lower hierarchical node may belong to a higher node; that is, may be common to all members of the more inclusive group. Such predictions take the following form: if some members of a group share a characteristic that is unknown for other members of the same group, then that characteristic is likely to be common to all members of the group. So scientific names are tags, unique keys, hierarchical nodes, and phylogenetic hypotheses. Thus systematists pack a lot of information into their names and users can get a lot from them.

Scientific names are hypotheses, not proven facts. Systematists may and frequently do disagree about hypotheses. Hypotheses, which in systematics range from what is a character to what is the classification that best reflects the history of life, are always prone to falsification, hence to change. Disagreements about classification can arise from differences in paradigm and/or information. Systematists use different approaches to construct classifications, such as cladistic versus phylogenetic versus phenetic methods. Given the same set of data that underlies a given hierarchy, cladists will derive classifications different from those derived by pheneticists (Fig. 1). Even among cladists, there can be differences about the rank (genus, family, order) and thereby the hierarchical groups used. These are disagreements based on paradigm. There can be disagreement about the hypotheses that underlie the information used to construct the classifications, such as what are the characters. And disagreement can arise among systematists because individuals use different information. While disagreements will affect the ability to predict, they need not affect the ability to retrieve information.

The desirable attribute that must be preserved to ensure complete access to information across multiple classifications is uniqueness. Our scientific nomenclature must guarantee that any scientific name that is used in any classification is unique among all classifications. This can be assured by having two primary keys. Unfortunately, having two keys increases the overhead of our information systems. So most systematists and *all* users want to avoid this problem by mandating that there be only *one* classification. Although in theory there is only one correct classification, as there was only one history of life, in reality there have been multiple classifications in the past, there may be multiple classifications in use today, and there will be multiple classifications in the future. That is the price of scientific progress, of the increase in our knowledge of the world. If information is to be retrieved across time—that is, if we want to extract information stored under obsolete classifications, and if we want to avoid dictating "the correct" classification—then we need a nomenclatural system that supports two unique keys.

The two keys for our language of biodiversity are the valid name and the original name. The valid name is the correct name for a concept (taxon) within a classification; the original name is the valid name in the classification in which it was

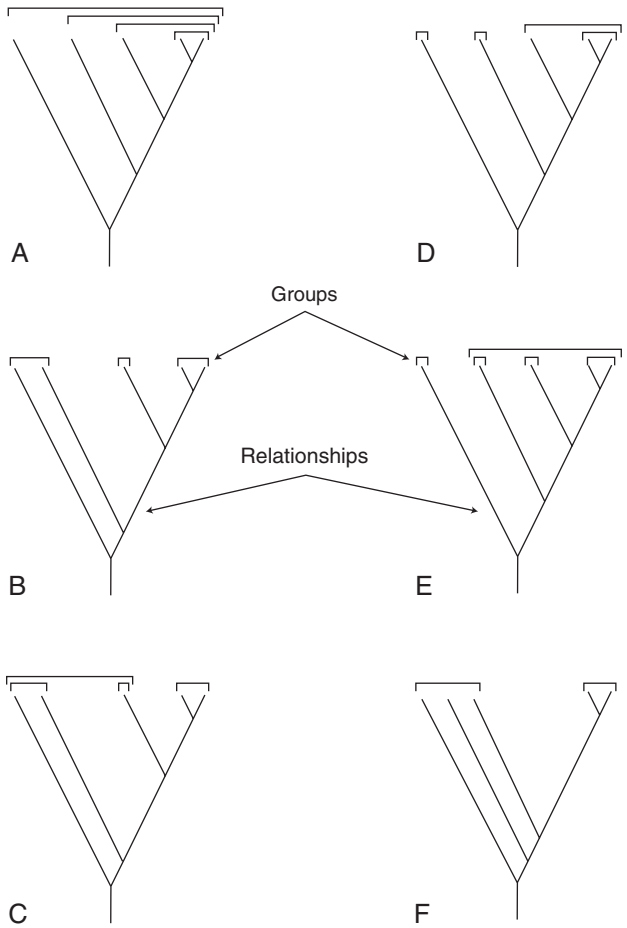

FIGURE 1 Multiple classifications for identical cladistic hypotheses. The brackets along the top of the trees indicate the groups recognized. Cladistic classifications are shown for trees A, D, and E and phenetic classifications for B, C, and D.

proposed. Valid names may be different among classifications, but the original name is invariant across all classifications (Table I). Valid names are the best names to use because they provide the full value of scientific names. These are the names that provide a basis for prediction. The original name is useful only for information retrieval across multiple classifications. Although valid and original names may be and frequently are the same, users must know the differences between them. Specifically, they need to know that a valid name is a powerful inference tool, that a valid name provides for prediction about unknown attributes of the organism that bears the name. But they must understand that there may be multiple valid names in the literature and/or in use and that valid names represent hypotheses that may change as our knowledge is tested and improved. So most importantly, if there are multiple valid names in use, then there are conflicting scientific hypotheses being advocated, and users must select the name that best serves their purpose. If users do not want to decide, do not want to use classifications to organize and synthesize their information, then they may use the original name to index their information, being assured that it will always be a unique key.

TABLE I Multiple Classifications (rows) and Primary Keys (columns) to Information

Year	Valid name	Original name	Authority
1776	*Musca balteata*	*Musca balteata*	De Geer
1822	*Syrphus balteatus*	*Musca balteata*	Meigen
1843	*Scaeva balteata*	*Musca balteata*	Zetterstedt
1917	*Episyrphus balteatus*	*Musca balteata*	Matsumura
1930	*Epistrophe balteata*	*Musca balteata*	Sack
1950	*Stenosyrphus balteatus*	*Musca balteata*	Fluke
Today	*Episyrphus balteatus*	*Musca balteata*	Vockeroth

There are other problems today with our classifications: synonymy, having two names for the same concept, and homonymy, having the same name for different concepts. These problems are, however, largely the result of ignorance. If we knew all names and their types and could agree on what are species, then by applying the rules of nomenclature we could immediately eliminate all synonymy and homonymy problems. Homonymy is eliminated by the rule of uniqueness. Synonymy is addressed by the rules of typification, which tie a physical instance of a concept to a name, and is resolved by logic of circumscription and the convention of priority (or usage). The name of a concept is the name affixed to one and only one of the types that falls within its circumscription (Fig. 1). The name used is determined by which name is the oldest (priority) or most widely used (usage). The specific rules for resolving homonymy and synonymy, as well as for the proper formation and documentation of names are our codes of nomenclature. These rules, however, do not address the problem of multiple classifications, nor can they establish order under conditions of ignorance of the universe of applicable names and their typification.

There is one final problem: the species problem. This is the problem of what is the basic unit of information and/or data. There is also the question of what species are and whether species are real or hypotheses. Species may be a category (rank) in classifications or a unit of information. The best current review on these questions is by Wheeler and Meier, but for nomenclature the species (or more precisely the species group, which includes the subspecies category) is considered to be a basic unit of information. The problem is that the species is not a data element. The species is not an indivisible unit, but consists of information, that is, data derived from specimens that have been identified as belonging to that species. Mistakes can be made during this identification, which is after all another hypothesis. Information is ultimately not derived from species, but from specimens. Biological information management really begins with specimen data management. The problems of specimen-based data management are not intractable but are readily addressed by the use of unique identifiers, such as bar codes, another form of unique keys.

The species problem is also one of circumscription, the definition of the limits of a taxon. A group with the same name and type may be more or less inclusive depending on the char-

acters used to define its limits. Zoologists differ from botanists in not considering circumscription to be a problem, since minimally all identically named taxa have at least some characteristics in common. The problem of how much is held in common, therefore, is best resolved by enumeration of the included taxa or specimens. The history of circumscription can be tracked by use of an additional key that uniquely identifies the person who defined the limits and the date of that action. Sufficient for our purposes is to know that specimen-based data will always be summarized into species-based information units and that all species-based information should be specimen based.

PARADIGMS AND CLASSIFICATIONS

The information that is embedded in nomenclature comes from the classification used. As noted, classifications consist of hierarchically nested groups of taxa, with the basic unit being the taxa ranked as species. Paradigms are theories about scientific knowledge and its organization. The first classifications developed by Linnaeus and Fabricius were largely based on Aristotelian essentialism/typology. Things were grouped together because they shared the essences of the group, which is the type. Later, when evolution was articulated as a paradigm, classifications were based on phylogeny, which is the genealogical hypothesis of relationship. More recently, when computers began to appear, classifications were proposed on the basis of statistical measures of overall or phenetic similarity. Finally, different ways of deciphering phylogeny were developed, and so, different ways of translating phylogenetic information into a hierarchical classification were proposed (phylogenetic vs cladistic methods). Over the past half-century, much has been written about the relative merits of phenetics, evolutionary systematics, and cladistics, but the inescapable conclusion for predictive and, therefore, maximally informative classifications, is that the cladistic paradigm is mandatory. Schuh provides a good summary of the arguments for cladistic classifications.

Regardless of the paradigm followed, all approaches leave unsolved the problem of how to translate the result of taxonomic analysis, be it a tree or a branching diagram of overall similarities, into a hierarchical classification. There are only two approaches to the translation of an analysis into a classification: subordination or sequencing. For subordination, each clade/branch becomes a recognized (named) taxon and a rank indicator provides a key to the relative level of subordination. Subordination works best when the phylogeny/branching diagram is balanced, that is, when each branching point divides the remaining terminal taxa into equal sized groups. For example (Fig. 2, Table II), 8 species could be clustered into 4 genera and 2 subfamilies, whereas a fully pectinate analysis would yield 7 genera clustered into 5 named ranks (subfamily, infrafamily, supertribe, tribe, subtribe). For sequencing, only the terminal clades/branches are recognized, but their order is preserved and suitably indicated to encode their sequential level of subordination. This method is highly efficient for analyses that result in pectinate trees. The pectinate example could be reduced to 7 sequenced genera. Sequencing does not work when the analysis is balanced. Given that most analyses are neither fully balanced nor fully pectinate, a mixture of subordination and sequencing should be used as long as the classification properly identifies which methodology was used for each portion. Wiley provides a full set of conventions to deal with these issues as well as others that involve the placement of fossil groups *(plesion)* or groups of uncertain or changeable position *(sedis mutabilis)* or unknown relationships *(incertae sedis).*

Beyond the translation of a taxonomical analysis into a hierarchical classification, another challenge remains, that is, what groups to formally name and what ranks to assign to

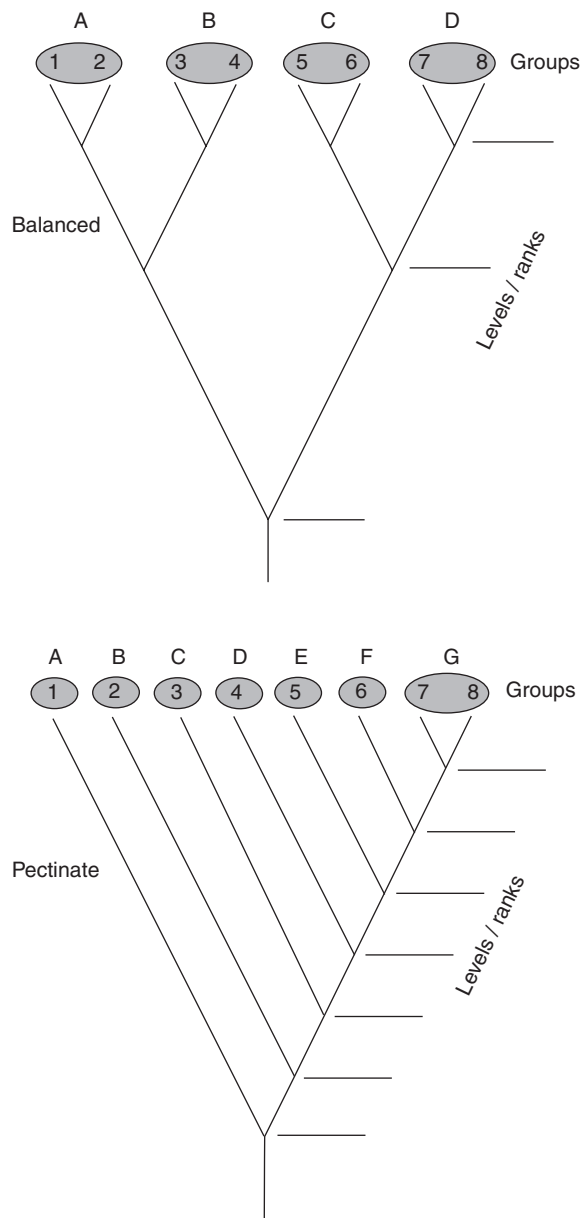

FIGURE 2 (Top) Balanced and (bottom) pectinate cladistic hypotheses. See Table II for the different classifications that result from these hypotheses.

TABLE II Classification: Sequencing and Subordination

Subordinated classification for Fig. 2A balanced analysis	Subordinated classification for Fig. 2B pectinate analysis	Sequenced classification for Fig. 2B pectinate analysis
Family A-idae	Family A-idae	Family A-idae
Subfamily A-inae	Subfamily A-inae	Genus A
Genus A	Genus A	Genus B
Genus B	Subfamily B-inae	Genus C
Subfamily C-inae	Infrafamily B-ites	Genus D
Genus C	Genus B	Genus E
Genus D	Infrafamily C-ites	Genus F
	Supertribe C-idi	Genus G
	Genus C	
	Supertribe D-idi	
	Tribe D-ini	
	Genus D	
	Tribe E-ini	
	Subtribe E-ina	
	Genus E	
	Subtribe F-ina	
	Genus F	
	Genus G	

those named groups. Obviously, when a group is fully resolved taxonomically, there could be as many named groups as there are terminal taxa. No school of taxonomy insists on naming all of them, but other than that there is no method nor any consensus among taxonomists on what taxa to name. This general problem is usually referred to by the names of the extreme views on either side, the "splitters" and the "lumpers," or those who would recognize many groups versus those who would recognize only a few. The merit of splitting is that the more taxa named, the more hierarchical information is embedded into those names and classification itself. Unfortunately that also leads to a loss of utility inasmuch as less information is summarized in each taxon.

Consider birds, the best-known group of organisms. Some 9700 species are clustered into 204 families and 2004 genera. Their scientific nomenclature is largely meaningless to many users, such as bird-watchers. For bird-watchers, common names, which more closely follow the original Linnaean classification, such as ducks *(Anas)* or hawks *(Falco)* or hummingbirds *(Trochilus),* are more meaningful groups than the oversplit genera. On the other hand, mosquitoes, some 3500 species, are clustered into only 34 genera. The important disease vectors, such as *Anopheles* for malaria and *Aedes* for yellow fever and dengue, remain large groups where the scientific name and common name are the same and are useful to doctors, public health workers, and other entomologists. The problem of the appropriate rank for groups recognized is similar. Naturally, splitters must have a greater series of rank indicators to express their fully named hierarchies. So, although there are relatively few species of birds, they are clustered into a large number of families (204), whereas flies (order Diptera) comprise 16 times as many species clustered into fewer families (142)!

The ranking issue also brings with it the question of equivalency. Obviously a family of birds is not an equivalent unit of biodiversity or of anything else in comparison to a family of flies. Rank equivalence is an important issue because many biologists want to make comparisons across different groups of organisms. Biological comparison should never be made on the basis of taxonomical categories above the rank of species. For example, studies that base conclusions on the circumstance that one treatment or niche has more families, than another are totally meaningless because the units being compared are not equivalent. Biological comparison should be made only on the basis of cladistically defined sister-group relationships, since sister groups are of equal age.

The entomologist Willi Hennig proposed in 1966 an objective method for assigning ranks that also allowed for biological comparisons: rank should reflect the hypothesized age of origin of the taxon. His suggestion has been rejected by all on the ground that the approach would cause a major upheaval in the traditional ranks of groups. For example, humans, placed in a separate kingdom by some (Psychozoa by Huxley in 1957), would be clustered among the apes and lemurs as nothing more than a species group. For entomology, some of the larger groups, like Coleoptera, Diptera, and Lepidoptera, which go back to Aristotle, would change in rank if not content. So after more than 2000 years of using those concepts, no one wants to split up the groups or change their rank.

These issues of classifications are largely ignored by working taxonomists, most of whom focus on their specialty and do not concern themselves with global classifications. Entomologists generally do not care how birds are classified, nor do beetle workers even worry about how flies are classified. Entomologists also tend to take a pragmatic, utilitarian approach, such that conservative ranks and grouping are used among mosquitoes and other economically important insects. In summary, a few general guidelines should be followed:

1. Only monophyletic taxa should be recognized and named.
2. Subordination or sequencing should be used as is most appropriate given the analysis and always should be annotated.
3. "Empty" taxa should never be named (i.e., if a family contains only a single genus, there is no a need to name a subfamily or tribe simply because these ranks are used elsewhere in the classification).
4. The fewer taxa named, the more useful the classification generally will be to nonspecialists.
5. Traditional groups and ranks should be preserved where possible.

CODES OF NOMENCLATURE

Because names are critical for communication and information retrieval, nomenclature needs to be universal, precise, and accurate. Universality requires that the same methodology be used by all and that methodology ensure stable nomenclature over time. Precision requires that only one result be derived from an individual application of the rules of nomenclature. Accuracy requires that names be consistently and precisely tied to the hypotheses they denote. The International Code of Zoological Nomenclature (ICZN) ensures the implementation of these basic functions in our scientific names and classification. This is achieved through a series of rules organized into chapters and articles.

Stability of nomenclature should not be confused with stability of taxonomic hypotheses (taxa) and classifications. As knowledge improves and more characters are discovered and analyzed, resulting in improved understanding of relationships among organisms, taxa and classifications will change. So, as more is known about the history of life, old Aristotelian groups like reptiles will be replaced by better defined ones, and the name Reptilia will drop from our classifications. But in other well-characterized groups, such as spiders (order Araneae) or flies (order Diptera), which have proven to be natural, the names shall remain unchanged in our classification.

The current ICZN is the product of a long evolution that began with the system of binominal nomenclature introduced by Carolus Linnaeus, a Swedish professor of natural history. This system was the direct result of an earlier government biodiversity project. The Swedish crown had some far-flung possessions and wanted to know what use could be made of them. Linnaeus was sent to investigate, to survey what today is called biodiversity, and to write a report characterizing his findings with recommendations on how to use them. At the time, there was only a binary system of nomenclature: one word for the genus, with the species being described by a series of adjectives. Given the diversity Linnaeus found, he did not want to waste time repeating long strings of adjectives that were required to characterize the biodiversity. So, because the base characterizations were in his flora of Sweden, he used a combination of the genus name and single word (an epithet) for each species to form a unique key to those descriptions (Stearn gives more details).

The system was an immediate success. Linnaeus codified the system, built and maintained a universal information database for all names (his *Systema Naturae,* 10th edition in 1758), and trained a cadre of students to carry on his work. The students dispersed and converted others. But since there could be only one master, Linnaeus, they divided nature up. There was to be no more *Systema Naturae.* For entomology, the student in charge was Johann Christian Fabricius. Fabricius defined his principles in his *Philosophia entomologica* and produced a series of *Systemae* for insects, the last comprehensive one being published in 1792 to 1794.

For the next 50 or so years, there was a significant increase in the number of animals discovered, described, and named, but little concern for nomenclature, which became muddled. This led a group of British zoologists, in 1843, to propose a formal set of rules, now known as the Strickland code, from which the modern ICZN evolved. After their effort, there was another half-century of new codes being proposed for various groups of animals (birds, fossils, insects) or nationalities (English, French, German). This proliferation led to zoologists joining forces and working toward an international code for all animals, resulting in the establishment of the International Commission on Zoological Nomenclature in 1895 and *Règles Internationales de la Nomenclature Zoologique* in 1905. Although a few entomologists (e.g., Banks and Caudell in 1912) continued to work on a specialized code for insect names, these development efforts were quickly abandoned.

For the next half-century, the *Règles* and the commission operated well, but clearly improvements were needed. So after the World War II, the task of revision began. After a series of international meetings, the American entomologist J. C. Bradley, then president of the commission, wrote out a draft that in 1962 became the second edition. The next edition, in 1985, and the current one, in 1999, were largely the work of Curtis W. Sabrosky, an American entomologist, David Ride, an Australian mammalogist, and Richard Melville, a British paleontologist.

The challenge in writing codes of nomenclature is making a set of rules that demand the best practices of taxonomists today, but also preserve the names created by past workers. Hence, to accommodate the work of the past, a code makes general provisions followed by a series of exceptions qualified by dates. Also, in zoology, there are two options for preserving history. A provision can be made in the code to solve a problem, or an appeal can be made to the commission to set aside the code to preserve an old name. Changing the code affects all occurrences of a problem, but a ruling of the commission applies only to a particular occurrence. In the past the commission frequently took many years to rule on cases. Hence, for Sabrosky and others, changing provisions of the code became the preferred method of addressing problems of old names. Hence, the current ICZN has many clauses that exist only to make old names available and to preserve their

customary usage. Unfortunately, in adding to the ICZN provisions of these kinds, Sabrosky and others frequently created unforeseen problems. So the ICZN must be used carefully, since for almost every positive statement there is usually an exception. Other linguistic constructions may also be confusing, such as the frequent use of the phrase "as such." For example, this phrase, used in Article 1.3.2, requires subsequent workers to decipher the intent of the original author. If the original author knowingly was describing an aberrant specimen, then the scientific name does not enter into nomenclature, but if the author thought the specimen was typical of the taxon being described, then the name must be considered to be available for use in nomenclature. Finally, the ICZN uses a number of specialized terms or special definitions for words; these are all covered in its glossary.

INTERNATIONAL CODE OF ZOOLOGICAL NOMENCLATURE

The current 4th edition of the ICZN consists of a preamble, a series of 18 chapters comprising 90 articles, recommendations and examples, and a glossary. The book also contains a preface, an introduction, and three appendices (the first two are general recommendations and the last is the Constitution of the International Commission on Zoological Nomenclature). The text is in both English and French, and there is a single combined index. This book, bounded in green, is the "official" edition published by the commission, but other official versions have been approved and published in different languages.

The preamble sets the objectives and basic principles of the code: "to promote stability and universality" of names and to ensure that the name of each taxon is unique. The preamble also declares that taxonomy is independent of nomenclature. The articles are the definitive rules, with examples of how they are applied in specific cases as well as recommendations of appropriate practices. The glossary defines each term used so that the rules can be interpreted consistently. The following is a summary of the code by major topics (in parentheses after the topic are the articles covered).

Scope (Arts. 1–3) The scope of zoological nomenclature is restricted to names for animals published starting in 1758, the date of the 10th edition of Linnaeus' *Systema Naturae.* Here, the ICZN uses the verb "deemed" to declare for nomenclature that *Systema Naturae* was published on 1 January and before Clerck's *Aranei Svecici* (spiders of Sweden), neither of which is true, since Clerck's work was actually published in 1757! Exclusions are also listed, such as hypothetical concepts. That simply means that if the Loch Ness monster is not real, then its name *Nessiteras rhombopteryx* Scott, is not a name covered by the ICZN.

Publication (Arts. 7–9), Dates (Arts. 21–22), and Authorship (Arts. 50–51) Although zoological nomenclature is a language for communication, the names that are regulated are those "published." Since taxonomy is based on some 250 years of work, the definition of publication used by the ICZN is based on printed works. For names and nomenclatural acts to be within the coverage of the ICZN, they must have been first published in a printed work in numerous copies available to the public and for the permanent scientific record. This definition excludes some printed works, such as daily newspapers, which are not published for the permanent and scientific record. The ICZN rules exclude the evolving digital world, such as the Internet. This assures that all users read the same material in determining what are the appropriate names. The ICZN does accept new digital media such as CD-ROM or DVD disks that are "stamped out," not printed. The ICZN provides rules to determine who is the author of these works and their dates of publication.

Names (Arts. 4–6) and Their Formation [Arts. 11 (11.2–11.3, 11.7–11.9), 25–49] Beyond falling within the scope of zoological nomenclature and having been published, names must be properly formed. They must be written with the Latin alphabet, and they must agree with various other requirements, many of which reflect the origin of the system at a time when all scholarly works were written in Latin. The ICZN groups scientific names into three kinds: family-group names, the names of taxa above the genus and species; genus-group names, the names for groups of species that form the first part of the binomen; and species-group names, the specific (epithet) names. The ICZN prescribes that names of higher taxa, such as superfamilies (-oidea), families (-idae), subfamilies (-inae), tribes (-ini), and subtribes (-ina), have specific suffixes; that generic and subgeneric names have gender and be nouns or be treated as such; and that specific names (epithets) be either nouns (and invariant) or adjectives (and whose ending agrees with the gender of the generic name with which it is combined).

Availability (Arts. 10, 12–20) Given that all the foregoing conditions are fulfilled, names and nomenclatural acts must meet additional requirements if they are to be made available for consideration under the ICZN and if they are to be held to be valid. The distinction between available and valid is critical. A valid name is the correct name to be used for a taxon, that is, a hypothesis of a group; an available name is any name that meets the requirements of the ICZN.

The additional requirements that names must follow to be available are as follows: (1) they must be formed as part of the system of binominal nomenclature, and (2) they must represent taxa that are considered to be valid. How taxa are defined is further regulated in a series of articles that are applied according to time. Before 1930 the standards were simple, such as attaching a previously unpublished name to an illustration, but current requirements are much more rigorous. For example, one must explicitly declare that a new name is being proposed, fully document the hypothesis (a description that "purports" to differentiate the taxon from all others), and designate the type for the name (which for a species is usually a dead specimen); if extant, the specimen must be deposited in a named, bona fide collection, or plans for such deposition must be furnished. (see Arts. 13.1.1, 16.1, and 16.4).

Validity (Arts. 23–24) A key problem for nomenclature is the existence of two or more names for the same taxon, for only one name can be considered to be correct or valid. This reflects the more general problem in science of who is to be given the credit for new ideas when multiple people claim to have arrived at them first. The general principle involved is that of priority—credit should go to the first person who published the idea. This principle was set forth by Henry Oldenburg, the first secretary of the Royal Society and editor of its *Philosophical Transactions* (1664–1677). Unfortunately, sometimes priority fails because the first to publish may have been forgotten, and someone else has become recognized as the first. In scientific nomenclature, this means that one name may become so familiar to many people that to change it to a name that is less widely known would cause instability. Hence, the ICZN provides a saving clause to allow for a widely used and familiar name to be preserved as the valid name when an older, obscure name is rediscovered. So, the ICZN provides a statement of the principle of priority (Art. 23.1), how it is to be applied in various situations, and, finally, when the principle should not be used (reversal of precedence, Art. 23.9).

Homonymy (Arts. 52–60) Another key problem of nomenclature occurs when two or more names that are the same apply to different taxa. This is known as homonymy. Because some names consisting of different Latin letters may mean the same thing, the ICZN defines "same": for example, the epithet pairs *microdon* and *mikrodon,* and *litoralis* and *littoralis* are considered to be the same (Art. 58). Then the ICZN dictates how homonymy is to be resolved: the senior name is to be retained, and the junior name is to be replaced; but there are exceptions as explained in Articles 52 to 60.

Typification (Arts. 61–76) Since names are only tags from scientific hypotheses, the question of whether two or more names are synonyms involves both nomenclature and taxonomy. Taxonomy in this regard can be defined as the circumscription of character space (Fig. 3), that is, the definition of taxa. Nomenclature is then the name, the designation of types for the name, and the rules for selecting among multiple types (see earlier remarks under Validity). Thus, a nominal taxon is only a name and a type; a taxon, the hypothesis, includes at least one nominal taxon. For species-group names, the type is a specimen (holotype, neotype, or lectotype) or a group of specimens (syntypes), which is or are the ultimate source of character information. For genus-group names, the type is a species name (nominal taxon), and for family-group names the type is a genus-group name. The determination of the type/genus of a family-group name is self-evident because that genus is basis for the family-group name itself (the type of the family Muscidae is the genus *Musca*). For both genus-group and species-group names, the types are designated either by the original author of the name or by subsequent workers.

An author may declare that a particular species/specimen is the type (original designation/holotype), may include only one species/specimen in the taxon (monotypy/holotypy) or may for genus-group names include a species-group name that is the same as the genus-group name (absolute or Linnaean tautonymy). If the type is not clearly fixed in the original publication, the ICZN provides rules for determining what species/specimens may be designated the type by subsequent workers.

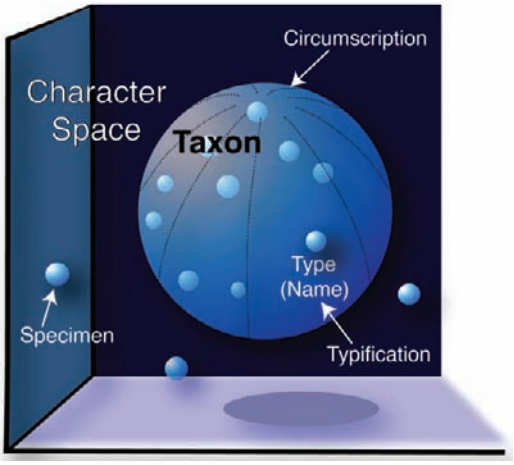

FIGURE 3 Circumscription and typification. Axes represent characters; small spheres represent specimens plotted against those axes; the large sphere represents the circumscription of a taxon; one specimen sphere is a type and its location determines the name for the taxon.

For genus-group names, all species included in a newly defined genus are eligible to be designated as the type species. But this applies only to genus-group names proposed before 1931, for as noted earlier, after 1930 type designation became a requirement of availability. If no species were originally included, then those first subsequently included are to be considered. To subsequently designate a type species, a worker merely, but unambiguously, declares one of these originally included species as the type.

For species-group names, when no type is designated in the original publication, all specimens upon which the author based the name (including specimens not seen by the author but referred to by bibliographic citation) are eligible to be designated lectotype and are called syntypes until such a lectotype is designated.

Collectively all specimens that are either holo-, lecto-, neo-, or syntypes are termed primary types. Other specimens studied by the author may be termed paratypes (or one may be an allotype if of a sex different from the holotype), but these secondary types have no nomenclatural significance. When all primary types are no longer extant (lost or destroyed), a subsequent worker may designate any specimen as the neotype to objectively define the nominal species-group taxon. Naturally, there are recommendations and restrictions about which specimen may make a more appropriate neotype. Provisions are also made for types that have been misidentified; that is, the characters used by an author to define a taxon do not agree with those of the nominal type. When this happens, workers are free to select as type either one that agrees with the characters or the named type.

Exceptions (Arts. 78, 80–83) and Registration (Art. 79) The stated objectives of the ICZN are to promote stability and universality in the names of animals, and thus all its provisions must be subservient to these goals. Hence, the ICZN provides means by which any provision of the ICZN (except those that deal with its authority and exception handling) can be set aside in a particular situation. These articles outline how one appeals to the commission and how the commission arrives at its opinion, which is then published in its *Bulletin of Zoological Nomenclature.* These plenary powers and specific powers are restricted to specific cases, usually involving only a few names or works. The ICZN has provided a procedure to rule on a whole set of names at once and has created a *List of Available Names in Zoology.* This specific power allows international groups of zoologists to propose a set of names that may be approved by the commission, thereby fixing all the relevant details about those names (their spelling, authorship, place and date of publication, and typification) and giving those names precedence over other names. Names not on the list are thereafter excluded from zoological nomenclature.

Authority (Arts. 77, 84–90) The last section of the ICZN includes a series of rules explaining the derivation and perpetuation of its authority, as well as the various regulations governing the particular edition of the ICZN. These state that the ICZN is prepared by the International Commission on Zoological Nomenclature with the participation of the zoological community under the authority of a single organization (originally the International Congresses of Zoology, now the International Union of Biological Sciences). For the future, authority can be delegated to other international organizations as specified. The effective date of the fourth edition is given as January 1, 2000, and all previous editions no longer have any force.

OTHER CODES

Currently there are five different codes of nomenclature in use for organisms-one each for plants, cultivated plants, bacteria, viruses, and animals. All these codes address the same problem, the need for universal, stable, and precise sets of names for organisms, and all are similar in their methodology. However, differences are significant and can cause difficulties for those developing databases that cover all life. Hence, in the early 1990s, an effort was undertaken to develop a single code of nomenclature for all life. Meetings were held and a draft BioCode was published, but nothing further has happened.

All the codes in use today are based on and have their origin in the binominal system established by Linnaeus. Although the Linnaean system has evolved from its topological roots into one adapted to the Darwinian evolutionary model, some believe that the system cannot fully express human knowledge about the cladistic relationships among organisms. A PhyloCode has been proposed to address these perceived failures. Unfortunately, the PhyloCode adds more uncertainty, since names for clades can be based on three different methods for defining groups, and clade names have no rank, which means that virtually all information content is lost (as discussed by Benton, Forey, and Platnick).

INTERNATIONAL COMMISSION ON ZOOLOGICAL NOMENCLATURE

Zoologists realized early on that no code of nomenclature could be perfect, always able to resolve all situations in a manner that promotes stability and universality in scientific names. Hence, zoologists established an international group of specialists with the authority not only to develop and maintain the ICZN, but to make rulings on specific names and situations. What the commission is empowered to do has been outlined in this article and is covered in Articles 78 to 81. How the commission operates is set in Articles 77 and 83 and in its constitution, which is published as Appendix 3 of the ICZN. To appeal to the commission, a worker prepares a proposal and submits it to the commission. The proposal is then published in the *Bulletin of Zoological Nomenclature* for public comment. After 6 months, the commission may vote on the proposal, and the ruling will later be published in the *Bulletin.* Each proposal is treated separately on its own merits. The commission does not set precedents or follow case law.

CONCLUSIONS

No modern science places as much emphasis on priority as taxonomy and nomenclature. This emphasis requires specialists to be familiar with at least a century of published work and sometimes 300 years' worth. Some may question the value of such a long view, in as much as most sciences look back only a decade or so. Beyond the moral and ethical considerations, however, much can be gained by understanding the past. Recognition of taxa is an innate ability of humans. Ernst Mayr once noted that the primitive natives of New Guinea knew and had names for 137 of the 138 local species of birds that took western scientists years to formally describe. So, previous workers who failed to generate cladistic classifications and were not aware of the proper names for their taxa may well have recognized and characterized natural groups. So, one tries to understand how one's predecessors, who looked at the same organisms, decided how to organize their observations into taxa. Thus all who want to truly master nomenclature and classification are well advised to examine carefully what their precedessors did, appreciating what Newton once wrote: "If I have seen further, it is by standing upon the shoulders of Giants."

See Also the Following Articles

Biodiversity • History of Entomology • Museums and Display Collections • Phylogeny

Further Reading

Banks, N., and Caudall, A. N. (1912). The entomological code—a code of nomenclature for use in entomology. Washington, DC.

Benton, M. J. (2000). Stems, nodes, crown clades, and rank-free lists: is Linnaeus dead? *Biol. Rev.* **75,** 633–648.
Forey, P. L. (2001). The PhyloCode: description and commentary. *Bull. Zool. Nomencl.* **58,** 81–96.
Hennig. W. (1966). "Phylogenetic Systematics." University of Illinois Press, Urbana.
Huxley, J. S. (1957). The three types of evolutionary process. *Nature* **180,** 454–455.
International Commission on Zoological Nomenclature. (1999). "International Code of Zoological Nomenclature," 4th ed. International Trust for Zoological Nomenclature, London.
Mayr, E., Linsley, E. G., and Usinger, R. L. (1953). "Methods and Principles of Systematic Zoology." McGraw Hill, New York.
Melville, R. V. (1995). "Towards Stability in the Names of Animals. A History of the International Commission on Zoological Nomenclature 1895–1995. International Trust for Zoological Nomenclature, London.
Schuh, R. T. (2000). "Biological Systematics. Principles and Applications." Cornell University Press, Ithaca, NY.
Stearn, W. T. (1957). An introduction to the *Species Plantarum* and cognate botanical works of Carl Linnaeus. *In* "C. Linnaeus, *Species Plantarum.*" [A facsimile of the first edition of 1753]. The Ray Society, London.
Thompson, F. C. (1996). Names: the keys to biodiversity. *In* "BioDiversity II" (Reaka-Kudla, M. L., Wilson, D. E. and Wilson, E. O. eds.), pp. 199–212. Joseph Henry Press, Washington, DC.
Wheeler, Q. D., and Meier, R. (eds.). (2000). "Species Concepts, Phylogenetic Theory: A Debate." Coumbia University Press, New York.
Wiley, E. O. (1981). "Phylogenetics. The Theory and Practice of Phylogenetic Systematics." Wiley, New York.

Nutrition

S. N. Thompson
University of California, Riverside
S. J. Simpson
University of Oxford

The nutrition of animals reflects their heterotrophic character and thus focuses on the need for animals to obtain many preformed organic substances that they lack the ability to synthesize from simpler carbon sources. Nutrients then are environmental factors that connect and intersect an animal's physiology and ecology. In the broadest sense, nutrition refers to the taking in and processing of substances that fuel the organism's energy needs for growth, maintenance, and reproduction. A strict definition of nutrition, however, is not possible and any specific view of nutrition depends on perspective. Regarding nutritional requirements, insects share much in common with other animals and nutritional studies with insects have contributed significantly to our general understanding of nutrition.

HISTORICAL OVERVIEW

Early interest in insect nutrition aimed at understanding the dietary requirements of insects, an investigative focus often called dietetics, that is, identifying and characterizing the food that insects eat to satisfy their nutritional needs, as well as the feeding behaviors and sensory physiology associated with obtaining those foods. A principal purpose of this research was to rear insects in the laboratory on artificial foods, thereby avoiding the often costly effort necessary to maintain natural foods for this purpose. Among the first successes was E. Bogdanow's 1908 report on rearing of a dipteran, *Calliphora vomitoria,* on a diet containing meat extract, starch, and peptone. Similar accounts followed, describing the rearing of many other insect species from a variety of taxa. Numerous scientists are associated with this early work, including S. Beck, R. Craig, G. Fraenkel, W. Trager, and B. Uvarov. Subsequent research attempted to refine the diets by using more purified natural products or even pure chemicals and nutrients. The effort was greatly aided by new information gleaned from research on the basic nutritional requirements of insects. A major advance was the discovery by E. Hobson in 1935 that insects require cholesterol. This clearly distinguished the nutrition of insects from that of most mammals and other vertebrate animals, whose nutritional requirements were, at that time, far better known. Further, the discovery quickly led to significant advances in the study of insect dietetics and nutrition.

ARTIFICIAL SYNTHETIC DIETS

Despite the diversity of insect dietary habits, insect nutritionists have been remarkably successful at developing artificial diets and rearing programs employing artificial diets for insects. Various compilations of insect diets suggest that several hundred species can be reared partially, or through their entire life cycles, without their natural foodstuffs, which was based on a collective effort of over 1000 scientific contributors. A variety of literature is now available describing various applications of this nutritional technology. A recent catalogue of a well-known commercial concern lists for sale artificial diets for over 50 species of insects, and artificial diets have been employed for mass culture of several insects. A U.S. Department of Agriculture laboratory in Phoenix, Arizona, for example, is currently rearing, on a completely artificial diet, 6 to 7 million pink bollworms, *Pectinophora gossypiella,* daily, for use in an autocidal pest management control program in California.

QUALITATIVE NUTRITIONAL REQUIREMENTS

The nutritional requirements of insects were at first inferred from knowledge of the chemical compositions of natural foodstuffs. On this basis, insects were categorized as carnivorous, omnivorous, or phytophagous, with the appropriate inferences regarding the nutritional content of the animals, plants, and other foods eaten. Studies of food utilization, and the analysis and comparison of foodstuff eaten and excreted or unabsorbed, allowed further assessment of the relative importance of various

nutrients for insects with different feeding habits. A more detailed and complete understanding of insect nutritional requirements finally emerged following the successful development of artificial diets, particularly chemically defined diets. The essentiality of individual nutrients could then be established by dietary deletion, the sequential elimination of individual chemicals, potentially nutritious, from a diet. This advancement involved many scientists now recognized for their contributions to the foundation of insect nutrition—J. S. Barlow, R. H. Dadd, S. Friedman, H. T. Gordon, H. L. House, J. G. Rodriquez, T. Ito, E. S. Vanderzant, G. P. Waldbauer, and others.

A nutrient is deemed *essential* if, when deleted, further growth, development, and/or reproduction is prevented. Almost all insects have a common set of essential nutritional requirements. Nutrients demonstrated to be necessary for optimizing growth, development, and/or reproduction are *required* nutrients. The value of specific nutrients, however, often depends on total dietary content. Dietary carbohydrate, for example, is often unnecessary, but may be required or even become essential for providing energy in the absence of dietary fat, or protein, in excess of the amount required for normal growth. The following summarizes the essential qualitative requirements of insects.

Nitrogen

Insects consume and utilize a wide variety of proteins to satisfy their nutritional requirements for amino acids. Of the commonly occurring 20 amino acids that comprise most proteins, 10 are not synthesized by insects, or most other animals, and are essential nutrients. These include arginine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine/tyrosine, threonine, tryptophan, and valine. Beyond the bulk requirement for protein synthesis, these essential amino acids serve a variety of additional physiological functions of particular note in insects and other invertebrates. Arginine, for example is a precursor for the principal invertebrate muscle phosphagen phosphoargininine. Tyrosine is important for production of phenolic and quinone metabolites that are critical components for cross-linking of protein during sclerotization.

Although the other 10 commonly occurring proteinaceous amino acids are generally nonessential, they are nevertheless required to some degree for normal growth and development, because few insects will develop on diets containing only the 10 essentials. Moreover, many insects will do poorly on diets containing amino acids as the sole source of nitrogen and require protein or a mixture of protein and amino acid for normal growth and development.

Insects can generally synthesize their own nucleosides, nucleotides, and nucleic acids, although several species, particularly various dipterans, have been shown to benefit by inclusion of nucleic acid or some constituents. In a few cases, nucleic acid is considered essential for completion of development.

Vitamins

Vitamins, particularly water-soluble B vitamins, including biotin (vitamin H), folic acid (B11), niacin, pantothenic acid, pyridoxine (B6), riboflavin (B2), and thiamine (B1), or close chemical relatives of these, are essential nutrients for all insects. These serve principally as precursors for the nearly universally needed coenzymes of intermediary metabolism. Regarding the fat-soluble vitamins of other animals, only tocopherol (E) and retinol (A) have proven beneficial for reproduction and sight, respectively, by some insects. Tocopherol also plays an important role as a lipid antioxidant.

Carbohydrate

Carbohydrate nutrients, although often required as an energy source, are rarely essential. An exception may be the case of some adult lepidopterans that feed solely on plant nectars. Indeed, some of these insects are thought to be capable of digesting sucrose alone. The same may hold true for some adult dipterans and hymenopterans. In addition to their role as nutrients, sugars, particularly sucrose, are powerful phagostimulants, without which some insects feed poorly or not at all.

Sterols

Insects also have an essential requirement for a dietary sterol. Cholesterol appears to be widely acceptable, but a number of other sterols, particularly β-sitosterol and other plant sterols, can also serve. Among a few exceptions is the interesting example of *Drosophila pachea,* the senita cactus fly, which requires Δ7 dietary sterols, metabolic derivatives of schottenol, a Δ7 sterol found in its natural food, the senita cactus. In addition to the bulk requirement for sterol utilization in membrane formation, sterol is also important for the production of ecdysone and other molting hormones in insects.

Fatty Acids

Fatty acids are not essential for most insects, but several mosquitoes and some lepidopterans require a polyunsaturated fatty acid. This requirement is associated with one of the few nutritional disease syndromes—the "crumpled wings" syndrome—in which absence of a polyunsaturated fatty acid results in adult insects that fail to fully expand their wings and are thus unable to fly. The chemical nature of this requirement is poorly understood. In the case of mosquitoes, arachidonic acid or some closely related fatty acids with an ω-6 double bond is essential. In those Lepidoptera requiring a polyunsaturated fatty acid, however, some species utilize an ω-6 fatty acid, whereas others utilize ω-3 fatty acids such as α-linolenic acid. Moreover, in those Lepidoptera requiring an ω-6 fatty acid, α-linoleic is generally preferred, and arachidonic acid is unsuitable. Nothing is known of the physiological basis for this requirement, but in the mosquitoes it may be linked

to the synthesis of prostaglandins, local regulators that target a wide variety of cellular functions in vertebrate animals.

Inorganics

The complex of mineral ions nutritionally essential for other animals is likewise essential for insects. Here the balance is often dramatically different from the well-known salt requirements established for mammals. Many insects, for example, require much greater proportions of potassium, magnesium, and phosphate relative to sodium, calcium, and chloride.

Ascorbic Acid and Other Water-Soluble Growth Factors

L-Ascorbic acid, vitamin C in vertebrate animals, is an essential growth factor for many phytophagous insects. In its absence, these insects generally fail to grow and/or develop. The pattern of this requirement among insects is thus similar to that of the higher animals of which species that have adapted to a diet of fruit and/or vegetables have apparently lost the ability to synthesize ascorbic acid. In contrast to the vitamin C requirement of vertebrates, ascorbic acid is required by insects in relatively large amounts, although this may in part reflect the necessity for a high antioxidant activity in synthetic artificial diets employed for testing. Moreover, unlike vertebrates requiring L-ascorbic acid, some insects utilize dehydroascorbate as effectively as L-ascorbate, as well as use the D geometric forms of closely related lactones, although these were generally not as effective. Although there is little understood about the role of ascorbic acid in insects, beyond its potential antioxidant action, it may play the same role as in vertebrates, that is, as a factor necessary for enzyme activities involving hydroxylation. Deficiency in insects is often associated with abnormalities of molting, possibly due to the absence of ascorbic acid effects on diphenyloxidases, perhaps in a manner analogous to its action on the synthesis of serotonin from tryptophan.

The lipogenic growth factors, choline and inositol, principal components of phospholipids, are required by many insects. The essentialness of these remains in question, although choline is likely an essential nutrient for most insects.

Other Unique Essential Nutrients

In addition to the above essential requirements, several nutrients may be uniquely essential for some insects. Several mosquitoes and one tachinid species, for example, appear unable to synthesize asparagine and this is an essential dietary amino acid. An essential requirement for proline by several taxonomically disparate insect species may be related to a limited activity of the urea cycle. For unknown reasons, glutamic acid or aspartic acids, normally nonessential, have been reported as essential for a few insects.

A unique and essential requirement for carnitine is known for several tenebrionid beetles, for which this usually nonessential nutrient is called vitamin B_T. Normally derived from choline, carnitine plays an important role in fatty acid transport.

Essential and nonessential nutrients are required in specific amounts, but as implied above, optimal levels of individual nutrients often depend on the dietary concentrations of other nutrients. Early studies with artificial diets purported to formulate optimal levels of nutrients, based on the relative concentrations found in natural foodstuffs. Diets employing these concentrations of nutrients often produced adequate results, but subsequent studies have demonstrated that the quantitative aspects of insect nutrition were far more complex than that approach suggested. Moreover, ecological, behavioral, and physiological factors are also important for optimal nutrition.

QUANTITATIVE NUTRITION

As we have seen, if an insect is to survive, grow, and reproduce, it must ingest several dozen different types of nutrient molecule. These molecules come packaged in varying amounts and ratios within foods, along with nonnutritive (and sometimes toxic) compounds. Foods in turn are distributed through time and space, and their finding, ingestion, and processing engender metabolic and ecological costs. Added to that, the nutritional needs of insects change with age, stage of development, reproductive status, etc. Matching the multidimensional and changing nutritional demands of the insect against the complex and changing composition of the nutritional environment has posed one of the greatest challenges to evolution—and to scientists who study insect nutrition. The problem is particularly difficult for herbivorous insects, for which the nutritional composition of host plants may be highly variable and there is the added complexity of secondary plant metabolites serving as antifeedants and toxins. There is a growing realization, however, that predators too may face considerable variation in food quality and may therefore have to regulate their intake and use of multiple nutrients rather than relying on more general food properties.

A powerful way of defining and exploring nutritional regulation that has arisen from work on insects has been to represent the animal, its intake and growth requirements, and the foods in its environment using multinutrient plots. This "geometric framework" has enabled the identification of the key elements in complex nutritional systems and the quantification of the interactions among them. These include interactions among the multiple constituents of the food as well as between behavioral and physiological regulatory mechanisms. The resulting descriptions provide a powerful means to study the mechanisms, ecology, and evolution of nutritional systems.

Quantifying Intake Requirements

Estimating the intake requirements (intake target) is a primary aim of any nutritional study. One way to do this is to allow the insect to demonstrate whether it is able to regulate its nutrient intake and if so, in which nutrient

dimensions. A well-documented study system is the locust, which has been shown to regulate its intake of both protein and carbohydrate under several challenges:

1. Pairs of complementary foods. When locusts were provided with one of four complementary food pairings (28:14 or 14:7% protein:% digestible carbohydrate vs either 14:28 or 7:14) they adjusted the amount and ratio of the two foods eaten to maintain a constant point of protein and carbohydrate intake (Fig. 1, top).
2. Food dilution. When locusts were given one of five foods containing a near-optimal ratio of protein to digestible carbohydrate (1:1), but diluted up to fivefold with indigestible cellulose (35:35, 28:28, 21:21, 14:14 or 7:7% protein:% digestible carbohydrate), they adjusted their consumption across all dilutions to maintain a constant point of nutrient intake (Fig. 1, middle).
3. Food frequencies. When two complementary foods (31:11 and 7:35) were provided in relative abundances of 1:3, 2:2, or 3:1 dishes of one versus the other food type, locusts precisely selected a point of protein-to-carbohydrate intake by adjusting their distribution of consumption between dishes (Fig. 1, bottom).

These remarkable feats of homeostasis were found to extend to regulation of salt versus macronutrient intake. Moreover, other studies indicate that such capabilities are by no means restricted to locusts.

Changes in Intake Requirements with Time

The intake requirements of insects are not static. They change with recent nutritional experience and level of activity. For instance, locusts and caterpillars select a protein-rich food following a short experience (only one meal in the case of the locust) of a protein-depleted food and show a similar preference for carbohydrate-rich food after a 4-h period of carbohydrate deprivation. Nutrient requirements also vary with stage of development, reproduction, and diapause. Over a longer time scale, nutritional requirements evolve to track changing nutritional environments and life histories.

Mechanisms Regulating Intake

Regulating nutrient intake requires two sources of information, the first being the composition of the food and the second the nutritional state of the animal. The responsiveness of an animal to a food of given composition should reflect, through feedbacks, the animal's nutritional state. Insects are able to taste certain key nutrients, notably sugars, amino acids, salts, and water. Studies on locusts, caterpillars, and blow flies have shown that the responsiveness of taste receptors to sugars and amino acids varies with nutritional state, as represented by concentrations of these nutrients in the hemolymph. Such nutrient-specific feedbacks enable insects to make sophisticated behavioral decisions about

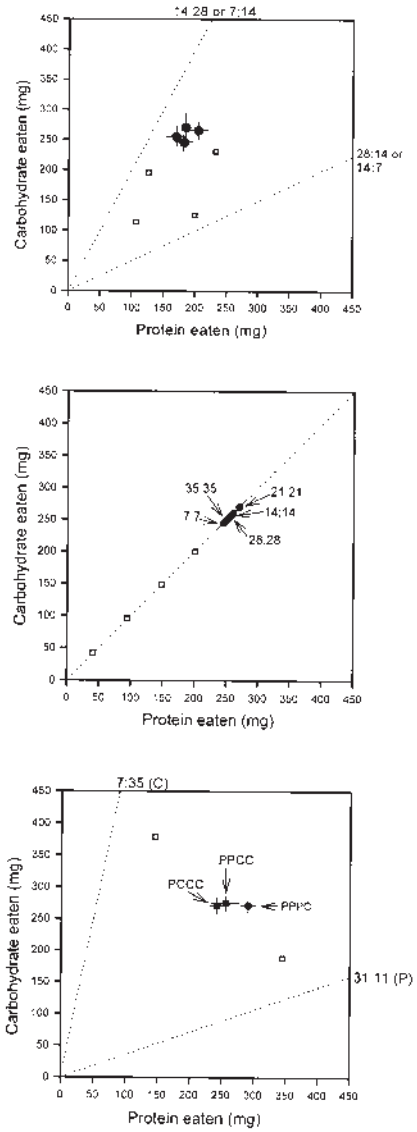

FIGURE 1 Three experiments indicating regulation of intake to a point ("intake target") in a carbohydrate–protein plane. The plots are of the form used in the geometric framework. Foods are shown as lines ("rails") radiating out from the origin. An insect feeding from a single food is constrained to move along that food rail in intake space, while two foods provide the opportunity to move anywhere in between by food mixing. In the top plot locusts were provided with one of four food choices (14:28 or 7:14 % protein:% carbohydrate vs 28:14 or 14:7) and altered relative amounts eaten from the two foods, thus reaching the same place in nutrient space. Open symbols indicate where insects would have arrived were they to have eaten indiscriminately between the two foods provided. The middle shows locusts that were given one of five foods containing a 1:1 ratio of protein to digestible carbohydrate, but diluted to various degrees with indigestible cellulose. Those on food 7:7 ingested five times more food than those on 35:35, thus achieving the same intake of both protein and digestible carbohydrate. Open symbols show points of intake were insects not to have compensated for dilution but eaten the same amount of food on each treatment. Locusts in the experiment shown at the bottom were given four food dishes, containing either 7:35 % protein:% carbohydrate (food C) or 31:11 (food P). The two food types were provided at different frequencies (all four dishes contained C, all four contained P, or two contained C and two P). Locusts adjusted the amounts eaten from each dish and regulated nutrient intake. Open symbols indicate points of intake were locusts to have distributed feeding equally among dishes and not shown frequency-dependent food selection.

what foods to eat to regulate nutrient intake. In addition, learning of various sorts, including aversion, learned specific appetites, and induced neophobic responses, also plays an important role in regulating food selection in insects. An impressive example is the ability of locusts to learn to associate the odor of a food with its protein content and to be attracted by odors previously paired with high-protein food, but only when in a state of protein deficit.

Nutrient Balancing on Suboptimal Foods

If an insect is restricted to suboptimal foods and is unable to reach its intake target, it must balance undereating some nutrients against overeating others, reaching some "point of compromise." Such a situation exposes whether and how the mechanisms regulating intake of different nutrient groups interact. A simple means of exploring the interactions between regulatory systems for different nutrients is to provide insects with one of an array of foods of varying composition and to measure intake. Collectively, the resulting points of nutrient intake across the array of foods form a pattern that describes the relationship between the mechanisms regulating intake of the nutrients concerned.

Various such relationships have been described to date in insects. The simplest outcome is that in which the insect abandons regulation of one nutrient when forced to balance it against regulation of another. For instance, locusts regulated macronutrient (protein and carbohydrate) intake and let salt intake vary passively when fed single foods containing suboptimal salt levels, even though they regulated both salt and the macronutrients when allowed to switch between complementary foods. In other cases, one nutrient does not overwhelm the other. When locusts were fed foods varying in protein and carbohydrate content they balanced over- and undereating the two nutrients, with the precise balancing rule depending on the species of locust. The grass-feeding specialist *Locusta migratoria* was less able to overeat unbalanced foods to gain more of the limiting nutrient than was the host-plant generalist *Schistocerca gregaria* (Fig. 2). It appears that the generalist species is better able to capitalize on excess ingested nutrients than is the specialist.

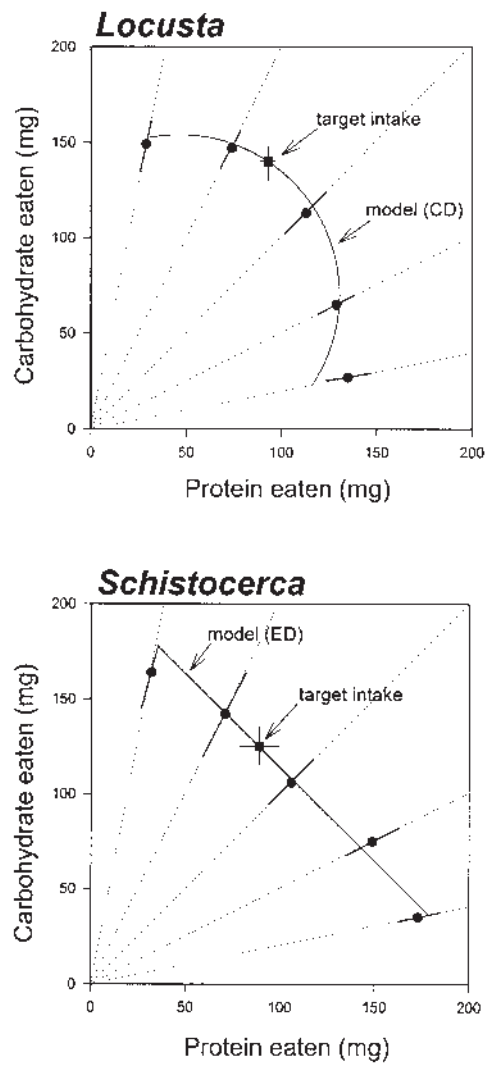

FIGURE 2 Arrays of nutrient intake shown by locusts when fed single foods and hence forced to balance their intake of two nutrient groups. The two plots show how the grass-specialist *L. migratoria* and the host-plant generalist *S. gregaria* have different balancing rules for protein and carbohydrate. *L. migratoria* minimized the distance from the intake target in nutrient space when unable to reach its intake target (the CD rule), while in comparison *S. gregaria* ate more of the nutritionally unbalanced foods (the ED rule).

Regulation of Growth and Metabolism

Whereas an insect may be constrained from reaching its intake requirements by available foods, it may still be able to regulate postingestive processing and thus achieve its growth target (Fig. 3). Regulation of growth and body composition involves differentially using ingested nutrients: ridding nutrients in excess and conserving those in deficit. Locusts are extremely effective at regulating growth, excreting excess ingested nitrogen, and respiring excess ingested carbohydrate or converting it to lipid. Deamination of excess ingested protein, principally by oxidation or transamination to the corresponding keto acid, may serve to augment limiting carbohydrate for energetic purposes.

NUTRIENT–ALLELOCHEMICAL INTERACTIONS

As well as primary nutrients, the foods of many animals contain harmful or unpalatable nonnutritive materials. This is especially the case for plants, which contain an abundance of defensive secondary metabolites (often termed allelochemicals). In some instances, insects make use of such compounds either as cues for host-plant recognition or as resources in their own right for defense or communication. In the main, however, allelochemicals are an impediment to nutrient regulation. A key point is that the effectiveness of allelochemicals as antinutritive compounds depends on the nutritional context in which they occur in the food. For example, locusts are not affected adversely by the presence of tannic acid in their food, even up

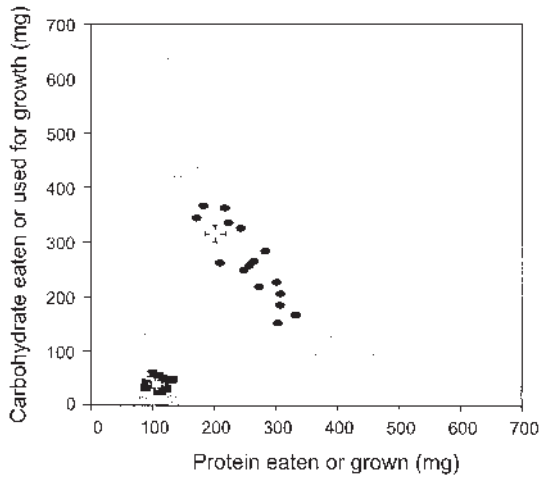

FIGURE 3 Patterns of intake and growth in locusts fed 1 of 25 diets varying in protein and digestible carbohydrate content. Intake data are across the entire fifth stadium. Note that protein- and carbohydrate-derived growth was regulated despite insects on the different diets having eaten widely different amounts of protein and carbohydrate. Locusts on all but the most extremely unbalanced diets regulated growth to a high degree in both nutrient dimensions. Regulation of growth involved differential utilization of ingested protein and carbohydrate. On diets with a very low protein-to-carbohydrate ratio (upright triangles), development was extended and body lipid content was high, whereas on foods with a very high protein-to-carbohydrate ratio (inverted triangles), there was no lipid deposition. Insects able to regulate their intake by selecting between complementary foods are indicated with the open square symbol.

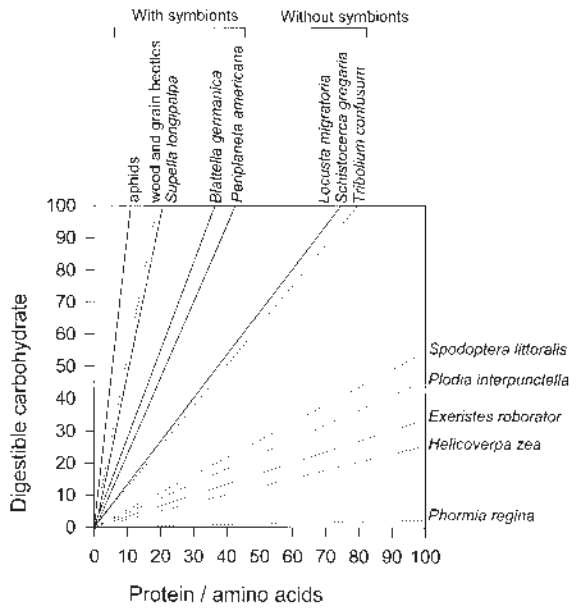

FIGURE 4 General effect of symbionts on nutrient intake. Data summarize the protein (or amino acid)-to-carbohydrate ratios in diets that support good larval development in a selection of insect species. Note how insects with microbial endosymbionts develop best on diets with a lower protein (or amino acid)-to-carbohydrate ratio (steep rails) than do other species. Dotted lines indicate endopterygote species while unbroken lines are for exopterygotes.

to 10% dry weight, if the protein and carbohydrate ratio and concentration are near optimal. When foods contain less than an optimal protein-to-carbohydrate ratio, tannic acid serves as a powerful feeding deterrent and thus causes high mortality and extended development. However, at higher than optimal protein-to-carbohydrate ratios tannic acid does not reduce intake but instead results in high mortality by disrupting protein utilization.

MICROBIAL ASSOCIATIONS

Symbionts, principally actinomycete fungi and bacteria, play a critical role in insect nutrition, enabling many species to develop normally on foods of limited nutritional value. Well-known examples of such foods include wood, blood, phloem, and plant litter. Many insects would quickly perish in the absence of these symbiotic relationships. Symbionts often provide nutrients directly as a result of synthetic capabilities that their insect hosts lack and/or allow, through the production of gut enzymes, insects to digest otherwise indigestible foodstuff. Alternatively, the symbiont itself may serve as food. Many symbiotic relationships are casual, involving ectosymbionts, usually comprising a rich gut microflora. Endosymbiosis is also common, and many insects have developed highly specialized anatomical and behavioral features for optimizing the benefits of the relationship and for efficiently transmitting their symbionts between generations.

Among the most significant endosymbioses are those involving mycetocytes and bacteriocytes, host cells specialized for housing symbiotic microorganisms—fungi or bacteria. Often these polyploid cells are associated with the midgut epithelium, although in the case of fungal-infected cells, aggregates called mycetomes may sometimes be found in the hemocoel. The genomes of the symbiont and the host cell are closely coordinated, forming a functional unit known as the symbiocosm. Such an endosymbiont is found in the rice weevil *Sitophilus oryzae,* in which it is referred to as the SOPE or *S. oryzae* principal endosymbiont. This bacterium (Family Enterobacteriaceae), whose expression is partly regulated by the host, occurs in the cytoplasm of the bacteriocyte (2×10^3 bacteria/host cell) and is known to be critically important in the insect's biology. It is, for example, the source of several vitamins, including riboflavin, pantothenic acid, and biotin. Moreover, the presence of the symbiont alters the balance of amino acid metabolism and mitochondrial phosphorylation, thereby affecting flight ability and performance.

The impact of symbionts on the quantitative nutrition of insects was recently made apparent by an analysis of the optimal dietary protein:carbohydrate ratio for 117 insect species. Insects housing symbionts displayed very steep intake target rails (Fig. 4), strongly suggesting that symbionts add considerably to the nitrogen nutrition of such species and, moreover, that this may generally be the case.

CONCLUSION

A thorough knowledge of insect nutrition is essential for understanding the biology of insects. The study of insect nutrition has recently undergone a metamorphosis, in that information gleaned from earlier investigations that focused principally on basic nutritional requirements and rearing technology is now being applied to understanding the feeding strategies, nutritional ecology, and evolution of insects. Nutritional physiology and biochemistry are also advancing. The neurological bases for food selection and the role of biogenic amines in regulating food choice are beginning to be understood. The chemical composition of the hemolymph is now recognized as a dynamic indicator of nutritional status, affecting food selection and nutrient intake. The metabolic responses of insects to altered nutritional status and the effects of fat-body metabolism on hemolymph composition are also being investigated. Future studies employing multidisciplinary approaches will continue to unravel the mysteries of insect nutrition and its consequences and significance to insect biology.

See Also the Following Articles

Digestion • Feeding Behavior • Metabolism • Plant–Insect Interactions • Rearing of Insects • Symbionts, Bacterial

Further Reading

Amakawa, T. (2001). Effects of age and blood sugar levels on the proboscis extension of the blow fly *Phormia regina. J. Insect Physiol.* **47,** 195–203.

Anderson, T. E., and Leppla, N. C. (1989). "Advances in Insect Rearing for Research and Pest Management." Westview Press, Boulder, CO.

Bernays, E. A., and Chapman, R. F. (1994). "Host Plant Selection in Phytophagous Insects." Chapman & Hall, New York.

Chapman, R. F. (1995). Chemosensory regulation of feeding. *In* "Chemical Ecology" (R. T. Cardé and W. J. Bell, eds.), pp. 101–136. Chapman & Hall, New York.

Dadd, R. H. (1985). Nutrition: Organisms. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 4, pp. 313–390. Pergamon, London.

Douglas, A. E. (1995). "Symbiotic Interactions." Oxford University Press, Oxford.

Heddi, A. Grenier, A.-M., Khatchadourian, C., Charles, H., and Nardon, P. (1999). Four intracellular genomes direct weevil biology: Nuclear, mitochondrial, principal endosymbiont, and *Wolbachia. Proc. Natl. Acad. Sci. USA* **96,** 6814–6819.

Raubenheimer, D., and Simpson, S. J. (1999). Integrating nutrition: A geometrical approach. *Entomol. Exp. Appl.* **91,** 67–82.

Reinecke, J. P. (1985). Nutrition: Artificial diets. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 4, pp. 391–419. Pergamon, London.

Simpson, S. J., and Raubenheimer, D. (2000). The hungry locust. *Adv. Stud. Behav.* **29,** 1–44.

Simpson, S. J., and Raubenheimer, D. (2001). The geometric analysis of nutrient–allelochemical interactions: A case study using locusts. *Ecology* **82,** 422–439.

Simpson, S. J., Raubenheimer, D., and Chambers, P. G. (1995). The mechanisms of nutritional homeostasis. *In* "Regulatory Mechanisms of Insect Feeding" (R. F. Chapman and De Boer, eds.), pp. 251–276. Chapman & Hall, New York.

Singh, P., and Moore, R. F. (1985). "Handbook of Insect Rearing," Vols. I and II. Elsevier, Amsterdam.

Slansky, F. (1992). Allelochemical–nutrient interactions in herbivore nutritional ecology. *In* "Herbivores: Their Interactions with Secondary Plant Metabolites" (G. A. Rosenthal and M. R. Berenbaum, eds.), Vol. 2, pp. 135–174. Academic Press, New York.

Thompson, S. N., and Redak, R. A. (2000). Interactions of dietary protein and carbohydrate determine blood sugar level and regulate nutrient selection in the insect *Manduca sexta* L. *Biochim. Biophys. Acta* **1523,** 91–102.

O

Ocelli and Stemmata

Frederick W. Stehr
Michigan State University

The number, size, and arrangement of ocelli, or simple eyes, are important diagnostic characters for many larvae. Until relatively recently, the simple eyes of both larvae and adults have been termed "ocelli" (ocellus), although it has been recognized that these are two different groups that are innervated from different parts of the brain.

OCELLI

One group of simple eyes, found in adult insects and larvae of nonholometabolans, is termed "dorsal ocelli" or simply "ocelli." These structures are innervated dorsally from the protocerebrum between the optic lobes. There are basically four dorsal ocelli, but one pair is fused to form the median ocellus; thus, there are typically three ocelli located near the midline of the head, but the number varies from zero to three (eight in Collembola). Their function is apparently visual, but knowledge is sparse about precisely what they see and how they interact with the compound eyes.

STEMMATA

The second group of simple eyes, formerly termed "lateral ocelli" but now termed "stemmata" (singular stemma), is found in the larvae of Holometabola. Stemmata are innervated laterally from the optic lobes, and typically there is a group on each side of the head. The number of stemmata is variable, ranging from zero to seven, and the number and arrangement can be diagnostic. They are most highly developed in externally feeding larvae such as caterpillars, sawfly larvae, and predaceous larvae, and are frequently less well developed, fewer in number, or absent in larvae found in concealed situations. Stemmata are used as horizon detectors, but the sharpness of their perception is no doubt limited.

See Also the Following Article

Eyes and Vision

Further Reading

Gilbert, C. (1994). Form and function of stemmata in larvae of holometabolous insects. *Annu. Rev. Entomol.* **39,** 323–349.

Paulus, H. F. (1881). Eyestructure and the monophyly of the Arthropoda. *In* "Arthropod Phylogeny" (A. P. Gupta, ed.), pp. 299–383. Van Nostrand Reinhold, New York.

Snodgrass, R. E. (1935). "Principles of Insect Morphology." McGraw-Hill, New York.

Stehr, Frederick W. (ed.). (1987). "Immature Insects," Vol. 1 Kendall/Hunt, Dubuque, IA.

Odonata

(Dragonflies, Damselflies)

K. J. Tennessen
Tennessee Valley Authority

Odonata (dragonflies) are paleopterous, exopterygote aquatic insects, related to the Ephemeroptera (mayflies). Dragonfly adults are predaceous, relatively long-lived insects. Their large compound eyes, strong chewing mouthparts, long legs, and unparalleled flight capabilities are ideal adaptations for catching and consuming insect prey. Although adult dragonflies have mastered the air, the larvae are aquatic and are usually much longer lived. Adaptation to an underwater existence has resulted

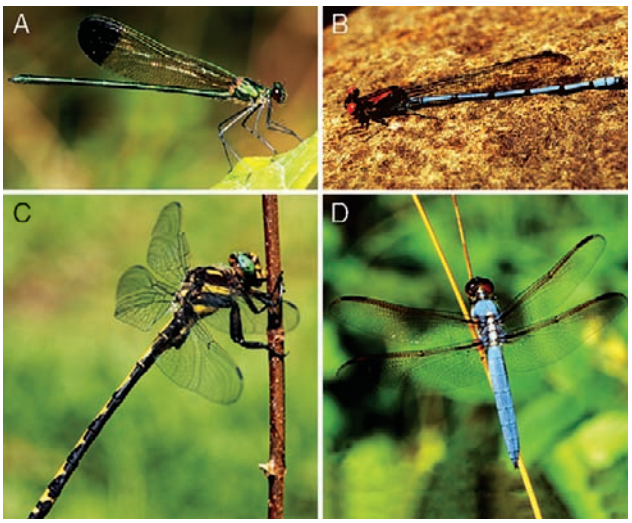

FIGURE 1 Examples of Zygoptera: (A) *Calopteryx dimidiata* and (B) *Argia* sp. (Photographs by R. S. Krotzer and K. J. Tennessen, respectively.) Examples of Anisoptera: (C) *Cordulegaster obliqua* and (D) *Libellula flavida.* (Photographs by R. S. Krotzer.)

in striking differences in form among larvae, whereas adults are much more uniform in shape.

Dragonflies are quite harmless insects; they do not sting and will try to bite humans only when held captive. However, they are hosts of trematodes (flukes in the family Lecithodendriidae) in Southeast Asia, and when eaten raw, they can be a source of infection in humans (by ingestion of Metacercariae). On the whole, dragonflies are considered to be beneficial insects for several reasons. In both larval and adult stages, they feed on many insects that are pests of humans and domestic animals, such as mosquitoes (Culicidae), deer-flies (Tabanidae), blackflies (Simuliidae), and other Diptera. They are important components of aquatic food webs and are used as indicators of ecological health of streams and lakes; in some areas, larvae are used as fish bait or as food. They make good subjects for behavioral and ecological studies, and poets and visual artists are often inspired by their beauty and behavior. Because some species are quite large and many are beautifully colored (Fig. 1), dragonflies have become fairly popular with the public. The relatively large size and distinctive color patterns allow the identification of many species of Odonata in the field, especially through binoculars. The recent appearance of field guides such as Dunkle's will facilitate natural history and behavioral studies also. As dragonflies become more popular, they may join butterflies as "ambassadors" for insect conservation and appreciation.

Dragonflies are a warm-water-adapted group that probably originated in a tropical environment, as evidenced by lower present-day species diversity in cooler climates (i.e., with increasing latitude and increasing altitude). At least 75% of the world fauna is tropically distributed, although a few genera (*Aeshna, Somatochlora,* and *Leucorrhinia*) have diversified in cooler climates, and their centers of distribution are located at higher latitudes and altitudes. Most high-latitude species in these genera are centered in the warmer parts of their geographic ranges, but a few species (e.g., in genus *Somatochlora*) are distributed mainly north of 60°.

PALEONTOLOGY

Fossil wings of several types of predatory, dragonfly-like insects from the Carboniferous (about 320 mya) have been found. The Eugeropteridae of the mid-Carboniferous are the most archaic members of the superorder Odonatoidea known. They had prothoracic winglets, but they also had pleated wings and a basal wing complex that included a cell in the shape of a parallelogram that allowed changes in camber of the wings and therefore maneuverable flight. Apparently the early odonatoids radiated and flourished throughout the Permian, when the large landmass Pangaea was still intact. They include the broad-winged Protodonata and Protanisoptera, and the petiolate-winged Protozygoptera. These primitive species lacked one or more of the diagnostic wing characters of all extant Odonata, including the modern basal wing complex (arculus and triangular or quadrangular conformation of veins), nodus, and pterostigma (see later) that together provide for highly maneuverable, swift flight. Some of these extinct "dragonflies" were probably the largest insects ever to have existed; for example, *Meganeuropsis* had a wing span of about 75 cm. The archaic odonatoids persisted until the Permo-Triassic extinction, a span of roughly 70 million years.

Although the phylogenetic relationships of the Paleozoic representatives are still poorly understood, the Odonatoidea as a whole almost unquestionably form a monophyletic group. Modern Odonata probably did not stem directly from these early odonatoids. Instead, it is probable that the ancestor of modern Odonata was similar to some of the extinct Jurassic groups (e.g., the Tarsophlebiidae, with nonpetiolate wings) previously placed in the "Anisozygoptera," now known to be a nonmonophyletic grouping. Therefore, the broad wings of Anisoptera and the petiolate wings of Zygoptera are equally derived characters, which probably evolved in the Jurassic. Anisoptera first appear in the fossil record in the Jurassic (150 mya), whereas Zygoptera do not appear until the Cretaceous period (120 mya).

Another puzzling question concerning Odonata paleontology is the appearance of aquatic larvae. The earliest fossil evidence of an aquatic larval existence is from the Mesozoic, and it has been suggested that they were semiaquatic before that period. However, the sister group of the Odonata, the Ephemeroptera, show evidence of larvae with gills in the early Permian at least 270 mya; no fossil mayfly larvae have been regarded as terrestrial.

SYSTEMATICS

Although systematics of Odonata is relatively advanced compared with most other insect orders, classification at the

TABLE I Odonata Families of the World[a]

Family	Distribution by continent	Genera	Species
Zygoptera			
Amphipterygidae	Africa, Asia, North and South America	4	9
Calopterygidae	Africa, Asia, Europe, North and South America	15	158
Chlorocyphidae	Africa, Asia	15	133
Chorismagrionidae	Australia	1	1
Coenagrionidae	Worldwide	82	1067
Dicteriadidae	South America	2	2
Diphlebiidae	Asia, Australia	2	9
Euphaeidae	Asia	8	63
Hemiphlebiidae	Australia	1	1
Isostictidae	Australia	12	45
Lestidae	Worldwide	9	155
Lestoideidae	Australia	1	4
Megapodagrionidae	Africa, Asia, Australia, North and South America	43	255
Perilestidae	Africa, North and South America	3	20
Platycnemididae	Africa, Asia, Australia, Europe	23	198
Platystictidae	Asia, North and South America	5	159
Polythoridae	North and South America	8	56
Protoneuridae	Africa, Asia, Australia, North and South America	24	240
Pseudostigmatidae	North and South America	5	18
Synlestidae	Africa, Asia, Australia, North America	7	33
Anisoptera			
Aeshnidae	Worldwide	49	411
Austropetaliidae	Australia, South America	7	15
Chlorogomphidae	Asia	2	40
Cordulegastridae	Asia, Europe, North America	4	51
Corduliidae	Worldwide	44	358
Epiophlebiidae	Asia	1	2
Gomphidae	Worldwide	91	952
Libellulidae	Worldwide	140	962
Neopetaliidae	South America	1	1
Petaluridae	Asia, Australia, North and South America	5	11
Synthemistidae	Australia	7	42
Total		621	5471

[a]North America includes Central America and the West Indies. Based in part on Davis, D. A. L., and Tobin, P. (1984). "The Dragonflies of the World," Vol. 1, "Zygoptera, Anisozygoptera." Societas Internationalis Odonatologica Rapid Communications (suppl.), No. 3, Utrecht; Davis, D.A.L., and Tobin, P. (1985). "The Dragonflies of the World," Vol. 2, "Anisoptera." Societas Internationalis Odonatiologica Rapid Communications (suppl.), No. 5, Utrecht; and Schorr, M., Lindeboom. M., and Paulson, D. (2000). List of Odonata of the world. http//www.ups.edu/biology/museum/worldodonates.html.

suborder and family level is still controversial. Until recently, three extant suborders were accepted, the Anisoptera, Zygoptera, and Anisozygoptera, with the latter group being represented by only two extant species of the Asian Epiophlebiidae. However, recent analyses indicate that "Anisozygoptera" is not a monophyletic group and that all taxa originally placed in that group are extinct. Therefore, the two living Epiophlebiidae are considered Anisoptera.

Classification at the family level also has been unstable. At present, 31 families are generally accepted (Table I), although there is some disagreement with this arrangement, especially within the Zygoptera. It is likely that DNA methods will create changes in classification and in the understanding of taxonomic relationships. Worldwide nearly 5500 species of Odonata have been described. The two suborders (Anisoptera and Zygoptera) have approximately equal numbers of known species. The rate of species description has remained fairly constant throughout recent decades (average nearly 350 new species per decade in the 20th century), an indication that the order is far from being completely known.

Larvae are much less well known than adults, especially in the tropics. However, larvae of nearly all the 427 North American species and 120 European species have been discovered, and much knowledge exists on their habitat requirements and life histories. Larvae of approximately 25% of the more than 1200 South American species have been described.

CHARACTERIZATION AND MORPHOLOGY

Characterization of the Order

The order Odonata is characterized by a prognathous head with chewing mouthparts, large compound eyes, three ocelli, small bristlelike antennae, a small prothorax, the meso- and

metathoracic segments fused into a large pterothorax, relatively long legs with three-segmented tarsi, two pairs of elongate wings, elongate abdomen, accessory male genitalia including the intromittent organ (not homologous with the penis of other insects) on the venter of the second abdominal segment, and one-segmented cerci. The odonate pterothorax is unusual in several features: (1) the bases of the legs are crowded forward, an arrangement conducive not to walking but to grasping; (2) the sternal sclerites make up most of the lateral and dorsal walls, with the mesepisterna meeting dorsally to form a middorsal carina; and (3) the wing bases are positioned posteriorly, and the tergal sclerites are extremely reduced. All these thoracic modifications facilitate strong flight and the pursuit and handling of prey. The huge flight muscles are connected directly to the bases of the wings. Therefore, the front and hind wings can be moved independently. The wing beat rate is relatively slow in comparison to neopterous insects (20 to 40 beats s^{-1} vs nearly 1000 beats s^{-1}), but dragonflies can fly almost as fast and as agilely as any other insect. Although most Zygoptera are relatively slow fliers, they can navigate precisely among stems and tangles of vegetation. The larvae are unique among Insecta in possessing an elongate, hinged labium that is folded under the head when not in use; it may be as long as one third of the larval body when extended to capture prey.

FIGURE 3 Anisoptera adult.

External Morphology

The suborders differ in two major characters. In Zygoptera (Fig. 2), the head is wider than the thorax and the hind wings are similar to the forewings in basal width and the orientation of the quadrangles. In Anisoptera (Fig. 3), the head is not wider than the thorax in dorsal view and the hind wings are wider at the base than are the forewings; also, the triangle of the forewing lies perpendicular to the long axis of the wing, whereas it usually lies parallel to the long axis in the hind wing. Larvae of the two suborders differ in that Zygoptera have three elongate, platelike or saclike anal gills and two comparatively short cerci (Fig. 4A,B), whereas Anisoptera have five pointed anal appendages (Fig. 4C,D) and an internal rectal gill chamber. Although the term "dragonfly" typically is used for the entire order, some authors restrict this term to the Anisoptera and use "damselfly" for the more slender Zygoptera.

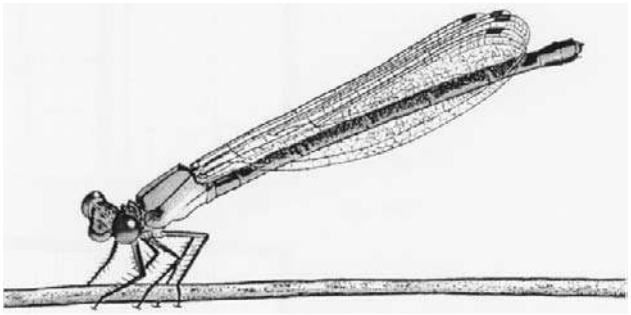

FIGURE 2 Zygoptera adult.

The wings of Odonata are richly veined to support the wing membrane. Wing venation is extremely important in systematics, especially at the family level. The beautiful colors of the body are produced by pigments under the cuticle and by diffraction of light by the cuticle. Females are often less strikingly colored than males of the same species, particularly in the large families Libellulidae and Coenagrionidae. Coloring may also be produced by pruinescence, a white or bluish white exudate of the hypodermis, which forms with sexual maturation, especially in males. Larvae are usually darkly colored, probably an adaptation to the colors of their microhabitat. In contrast to adults, larval body shape is highly variable, ranging from slender cylinders to nearly flat circles, and undoubtedly reflects specific habits of foraging and escaping predation. Larval antennae are more developed than in the adults. The most distinguishing characteristic of larvae is the prey-capturing labium, which is highly variable in its morphology and ranges from flat to cup- or spoon-shaped and from very elongate to short and wide. The palpal lobes and prementum are armed with a highly variable number of raptorial setae, though these may be absent.

BIOLOGY

Life Cycle

Odonata are hemimetabolous. The larvae have external wing sheaths, and although there is no pupal stage, the larva in its final stage differs greatly in form from the adult. Several developing adult structures, such as the labium, often can be

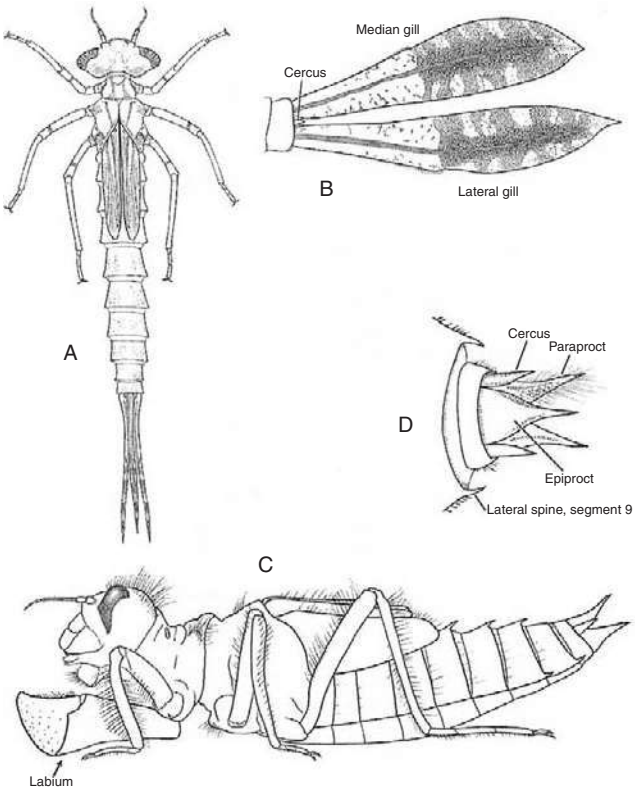

FIGURE 4 Odonata larvae: (A) Zygoptera larva in dorsal view, (B) anal gills of Zygoptera in lateral view, (C) Anisoptera larva in lateral view, and (D) anal pyramid of Anisoptera in dorsal view.

seen at this stage through the larval integument. Of the three life stages (egg, larva, adult), the larvae show the greatest diversity in functional morphology.

EGG STAGE Eggs are laid in or above permanent or temporary water bodies. The Zygoptera and a few Anisoptera families (Aeshnidae, Petaluridae) lay their eggs in plant tissue (endophytic oviposition); most Anisoptera families lay their eggs in open water (exophytic oviposition), although some species attach their eggs to plant tissue. Eggs are spindle shaped in endophytic species and usually round to ellipsoidal in exophytic species; they range in size from about 0.23 × 0.48 mm to 0.60 × 1.40 mm. Eggs are fertilized as they pass through the female's vagina during oviposition, and embryogenesis begins immediately after the eggs are laid. Fertilized eggs change from creamy white or light gray to a light or reddish brown or dark gray within the first 24 h. However, eggs of some tropical species are brightly colored (e.g., blue, green, pink) throughout the egg stage. Embryos that undergo direct development hatch within 5 to 60 days, whereas those undergoing delayed development (diapause) hatch between 80 and 200 days after oviposition depending on temperature.

The hatching process begins several hours before actual eclosion. Peristaltic movements of the esophagus result in swallowing of amniotic fluid, which causes water to enter the egg through the micropyles. In Zygoptera, increasing pressure within the egg causes the chorion to rupture, usually along curved lines of weakness. Continual swallowing and abdominal distension move the embryo forward, pushing the head forward into a chamber formed by the vitelline membrane. The embryo then swallows water, bursting the vitelline membrane, and the first instar slips most of the way out the chorion, but typically the tip of its abdomen remains inside the egg. In Anisoptera, a sclerotized frontal crest, called an "egg burster," is used to produce a longitudinal slit in the chorion, through which the larva exits the egg.

LARVAL STAGE The first instar, known also as the prolarva, is extremely brief in duration. It does not feed, and the legs are seldom functional. It may last for a few seconds to a few hours, depending on whether the egg was deposited in or out of water. In most species, the prolarva molts to the second instar while the tip of the abdomen is still within the egg. In species that hatch above the waterline, the prolarva drops out of the egg and reaches the water primarily by jumping. Second instars retain some yolk in their midgut to provide nutrition for a day or so, allowing them to become adept at their predatory habits. They are usually fairly pale, becoming darker in succeeding instars, usually matching their microhabitat.

The number of instars is highly variable within the order, ranging from 9 to 17. Most species have 11 to 13 instars, but even siblings treated identically can undergo a different number of molts. Instar classification is especially difficult with field-collected larvae, except that the first three and last three instars usually can be determined.

Duration of each instar is also highly variable and is dependent on species, temperature, and food availability. The final instar is the longest in duration, lasting as little as 5 days in rapidly developing species to a year or more in others. Growth occurs immediately after each molt for about an hour, while the integument is still pliable. Growth ratios (proportional increases in linear dimension from one instar to the next) for Odonata range from 1.2 to 1.3 and are usually very close to the average for Hemimetabola (1.27). Certain body structures change with successive molts; for example, the number of antennal segments and the number of palpal and premental setae increase, dorsal protuberances on the head of early instars disappear, the compound eyes become larger, color patterns develop, and rudiments of the sexual gonapophyses appear. Wing pads usually appear during the middle instars, and grow more rapidly than any other body part.

Most species of Odonata have one or two generations per year, but many are semivoltine. A higher percentage of species are multivoltine in the tropics than in temperate latitudes; many temperate-centered Anisoptera take 4 to 6 years to complete one generation. Odonate life cycles can be classified as either regulated or unregulated. In the tropics, life cycles are regulated by alternating wet and dry seasons, whereas in temperate zones they are regulated by alternating warm and cold seasons. In regulated types of life cycles, the dry season is

usually passed as prereproductive adults or as eggs, whereas the cold season is passed as larvae or as diapausing eggs. Species that occupy ephemeral habitats undergo rapid larval development, but they may or may not be multivoltine, depending on other environmental conditions (combinations of photoperiod and temperature). In continuously available habitats, life cycles are unregulated.

Although a few species of Odonata are known to have terrestrial larvae (e.g., *Megalagrion* in Hawaii), the vast majority require fresh water for functions such as respiration and feeding and to prevent desiccation. Likewise, a few species occupy brackish water habitats (the libellulid *Erythrodiplax berenice* can tolerate truly saline conditions and occupies coastal marshes but not the open ocean).

Respiration is largely through the integument, augmented by an internal rectal chamber in Anisoptera and by external anal and rarely lateral abdominal gills in Zygoptera. Anisoptera larvae can be readily observed breathing: as the abdomen enlarges, water is taken in through the anal opening; contraction of the abdominal muscles forces water out. The rectal epithelium of Anisoptera is developed into a specialized, richly tracheated branchial basket, into which oxygenated water is drawn by pumping action. Different families, and different genera within families, can differ greatly in tolerance of water low in dissolved oxygen. For example, many species of Libellulidae can thrive in low dissolved oxygen levels. In the family Gomphidae, larvae of *Aphylla* bury deep in the soft substrate of lake and pond habitats where oxygen levels are low, whereas species in the genus *Ophiogomphus* lie shallow in sand/gravel substrates typically in swift, highly aerated streams. The latter group appears to have narrow environmental requirements compared with a species such as *Dromogomphus spinosus,* which can occupy fast or slow streams and lakes; it even colonizes newly formed ponds.

Shortly after hatching and throughout their larval life, dragonflies must capture living prey and escape predation, often they are at or near the top of the aquatic food web. Prey includes many kinds of Diptera, but probably all other kinds of aquatic insect are consumed (e.g., mayflies, heteropterans, caddisflies) as well as many other invertebrates (protozoans, oligochetes, crustaceans, mollusks). Some species prey on small vertebrates such as larval fish and amphibians. Prey are detected usually by their movement, either tactually or visually. Motionless prey, such as snails, may be detected visually by recognition of their shape. Tactile detection is more important in earlier instars, vision becoming more keen after the first few molts. Larvae either stalk or ambush their prey. The prehensile, protractile labium (Fig. 4C), unique to the Odonata, strikes within milliseconds to capture prey. The labium grasps the prey and brings it to the mandibles, where it is chewed or engulfed whole. Because odonates are generalist predators, the potential for greatly affecting a particular prey population is low. Predators of dragonfly larvae include fish, frogs, a few reptiles and birds, other odonates, aquatic beetles, and heteropterans. Escape mechanisms include reduction of movement, even feigning death and seeking cover. When a predator grasps a zygopteran by a gill or leg, the appendage may be autotomized and later regenerated. Anisoptera are also capable of leg autotomy, or when grasped by the head or thorax, may make stabbing movements with the abdomen, using the sharp tips of the anal appendages to deter a predator.

In general, habitats range from streams of all sizes to seeps, bogs, ponds, lakes, swamps, and marshes. Some families, such as the Calopterygidae and Cordulegastridae, are restricted to flowing waters. Within the broad habitat categories, various species occupy different microhabitats. For example, elongate, cylindrical larvae, such as aeshnids and coenagrionids, usually cling to stems and sticks, whereas more flattened species, such as gomphids and libellulids, dwell near or in the bottom. A number of specialists occupy phytotelmata, which includes leaf axils (such as provided by bromeliads), rotten holes in tree trunks, bamboo internodes, and depressions at the bases of large trees, mostly in tropical areas. There is great diversity in larval form and behavior that allows dragonflies to occupy different types of habitat. Four categories of larvae (claspers, hiders, sprawlers, and burrowers) are based on how the microhabitat is occupied, although many species could be put in more than one of these categories, and larvae may move from one type of microhabitat to another, depending on age and season. Categorization is based mainly on leg shape and how the legs are used to situate the larva in its resting position.

Claspers, which cling to rocks, stems, or logs, have stout curved legs and cylindrical abdomens. Hiders are less elongate and conceal themselves among dead leaves or other debris by using strong legs; they usually have many stout setae to accumulate mud particles for better concealment. Sprawlers usually are flattened dorsoventrally and lie flat at the water–substrate interface. Burrowers, which may be semicylindrical or flattened, dig into the substrate to hide; the tip of the abdomen is often elongated to protrude above the substrate for respiration. Some burrowers propel themselves through the water by forcing water out of the anus, a form of jet propulsion (e.g., *Progomphus* can move several inches horizontally with one pump). A very few burrowing larvae make an actual burrow (e.g., some Petaluridae). The function of dorsal protuberances and lateral spines on the abdomen, common especially in Anisoptera, is not clear, although it is likely they serve multiple functions such as aiding in concealment and defense against attackers. Experimental evidence indicates that larvae grow longer abdominal spines in the presence of fish predators than larvae raised in the absence of such predators, suggesting that habitat shifts may be responsible for some of the differences observed within and between closely related taxa.

EMERGENCE Late in the final instar, when the larva is completing development but with a week or more to go before metamorphosing to the adult stage, the wing pads begin to thicken. The final instar usually does not leave the water until the day it is to metamorphose. At this time, the pharate adult is encased in the exuviae of the final instar. The main require-

ment for the dragonfly at this time is to find proper support structure, as it must cling tightly for metamorphosis to proceed successfully. The emergence support can include any physical object, ranging from the horizontal, terrestrial substrate adjacent to the water body to upright (vertical or inclined) objects such as rocks, plants, or synthetic structures. Shortly after leaving the water and securing a support grip, the dragonfly splits the integument along the middorsal line of the thorax and begins to push the adult thorax through the narrow opening. A slit in the integument of the head then appears, and the head and thorax push out. Shortly afterward, the legs, wings, and anterior portion of the abdomen appear. The dragonfly then remains motionless with legs folded for 10 to 20 min, hanging downward if oriented vertically, or protruding upright if oriented horizontally. When this apparent rest period is over, the legs are extended to grasp the exuviae and the rest of the abdomen is quickly withdrawn. Usually the cloudy wings then expand, followed by lengthening of the abdomen, each process taking about 15 min. As the wings become clear, drops of water are emitted from the anus and the abdomen becomes more slender, slowly taking final adult shape. The wings are suddenly spread out horizontally, and begin to vibrate. The full emergence process lasts from about 30 min to 2 h, ending with the maiden flight.

Some species emerge at night, apparently to escape predation, although many emerge at dawn or in full daylight. Some species have relatively synchronized emergence, all individuals within a population emerging within a day to about 2 weeks of each other. Other species are unsynchronized, emerging throughout the warmer seasons. Periodic exuviae collections can reveal such trends in emergence curves and population size. In nearly all species studied, the ratio of males to females is close to 1:1, although usually a slightly higher percentage of males occurs in Zygoptera, contrasted by a slightly higher percentage of females in Anisoptera.

Adult Behavior

The two main phases of adult life are the prereproductive (or maturation) period and the reproductive period. The prereproductive period lasts from the completion of emergence to the onset of sexual maturity. A brief postreproductive period, after reproductive capacity has passed, is sometimes observable.

PREREPRODUCTIVE PERIOD Upon reaching safety following the maiden flight, dragonflies remain in a teneral condition for about a day, during which time colors begin to develop and the integument begins to harden. The prereproductive phase then lasts from 2 days to several months, depending on species and environmental conditions; females usually take slightly longer to mature than males. During the prereproductive period, the gonads mature, the thoracic musculature becomes fully developed for agile flight (necessary for attaining a mate), weight increases, and mature colors are attained (for sex and maturity recognition in some species).

REPRODUCTIVE PERIOD Males and females of nearly all species mate with more than one individual, although monogamy has been reported in the coenagrionid genus *Ischnura.*

Male and Female Encounters Vision is the main sense used by Odonata for finding mates. The sexes usually meet at or near the aquatic oviposition site. Males that occupy a fixed territory can be categorized either as "fliers," patrolling continuously along the proper habitat, or as "perchers," making short defensive flights from a convenient perch. Territories are maintained by flying at or even clashing with intruders, and pursuit often results in both individuals leaving the habitat temporarily. Resident males usually return quickly to their territory. Mating usually occurs when a female arrives at the water, although males will attempt to acquire females several hundred meters from an oviposition site.

Recognition of Conspecifics Males recognize females mainly by color, color pattern, body shape, and flight style. Males of most species directly pursue and attempt to grasp any female that comes within sight, and if successful in achieving tandem, will attempt immediately to initiate copulation. In such species, males sometimes take heterospecific females into tandem.

Courtship has been described in a few families (Calopterygidae, Chlorocyphidae, Euphaeidae, Hemiphlebiidae, Platycnemididae, Libellulidae). In some species within these families, some males (but not all) present color and/or posture displays to induce a female to copulate. For example, in *Perithemis tenera* (a small North American libellulid with sexual wing color dimorphism), males establish territories around oviposition sites that consist of some sort of vegetation protruding from the water surface. On detecting a female near his site, a male will fly toward her and follow, moving from side to side. He then turns and flies to the oviposition site. The female may or may not follow him, depending on suitability of the site. If acceptable, she will follow, whereupon the male hovers over the site, fluttering his wings. The female slows her wing beat frequency and may even perch on the site. At this signal, the male initiates tandem linkage and copulation follows. When females are unreceptive to a male's approach, they perform distinctive refusal behaviors. Female Anisoptera usually fly very rapidly or erratically to escape, although some simply hide from males. Female Zygoptera usually show refusal by remaining perched and spreading their wings, sometimes also raising the abdomen or curving down the posterior portion of the abdomen. Males respond to such displays often by leaving, although some still attempt to achieve tandem.

Tandem Linkage and Sperm Transfer The postures adopted by Odonata in male–female tandem linkage and in copulation are unique. Tandem linkage, a necessary precursor of copulation, is initiated by the male. Males land on and seize females usually in flight (Fig. 5A), although they may grasp perched females also. The male, after landing on the female's pterothorax, brings his abdomen up and forward so that he can grasp the female's head or thorax with his anal

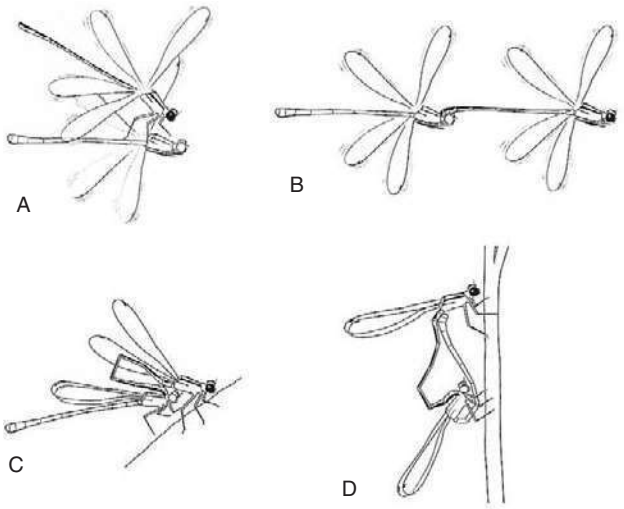

FIGURE 5 Mating sequence of Odonata (male with dark markings, female pale): (A) male grasping female, (B) tandem linkage, (C) intramale sperm translocation, and (D) copulatory wheel.

appendages; he then lets go with his legs and straightens his body, thus having achieved tandem linkage (Fig. 5B). In most Anisoptera, the male appendages fit over the dorsum and rear of the female head, whereas in Zygoptera they fit over the dorsum of the female prothorax. In the anisopteran family Aeshnidae, the male cerci fit tightly on the rear of the female head but also touch the anterior part of the prothorax; in some Zygoptera, the male cerci touch the anterior portion of the pterothorax. Shortly after achieving tandem, the male transfers sperm from the gonopore on abdominal segment 9 to the penis on segment 2 (Fig. 5C); this act is called intramale sperm translocation. Males sometimes translocate sperm before acquiring a female. The tandem pair either copulates in flight (most Anisoptera) or flies to a perch to copulate (Zygoptera and several Anisoptera families). The female swings her abdomen forward from underneath so that her genital aperture engages the venter of segment 2 of the male; the pair is now in the copulation wheel (Fig. 5D). Sperm are then transferred from the male's intromittent organ to the female's sperm storage organ. Copulation is usually extremely brief in flight (3–20 s) but can last from a few minutes to over an hour in perched pairs.

The male intromittent organ of Odonata is designed not only to inject sperm into the female, but also to remove or displace sperm of previous males. Jonathan Waage's discovery of this dual function led to evolutionary understanding of the whole suite of reproductive phenomena, from sexual selection to sperm competition and mate guarding. Sperm displacement may be achieved by removal, repositioning, or dilution. The structure and shape of the penis are vital to the mechanism employed: penes with backwardly directed barbs or hooks remove sperm from the bursa copulatrix, whereas those that are rounded pack sperm. Because most eggs are fertilized with sperm from the most recent insemination as they pass the female's fertilization pore during oviposition, the last male to copulate with a female is most likely to leave progeny. This phenomenon, termed sperm precedence, explains why males guard females after mating with them. By such postcopulatory association, a male protects his genetic investment by preventing other males from overtaking the female and by inducing the female to deposit (and therefore fertilize with his sperm) most of the eggs she is carrying.

Oviposition When copulation is terminated, pairs may break tandem linkage or remain together. Oviposition usually takes place shortly after copulation regardless of whether the tandem linkage is maintained, although females do lay eggs at times when males are not at oviposition sites. Ovipositing late in the day is a fairly common way for females to avoid interference from males. Males of many species guard females with which they mate, either by maintaining tandem contact with or by remaining near them, or both. Attempts to guard females are not always successful, as intruding males sometimes grasp and copulate with guarded females. Typically, guarded females oviposit more rapidly than those unguarded; their fitness thereby is enhanced.

Contact guarding Males of many Zygoptera remain in tandem with their mated female while she oviposits, even when she submerges. Males of many Coenagrionidae project vertically in the air with legs and wings folded, their only support provided by the grip of their anal appendages on the female thorax. In many Libellulidae, male and female fly in tandem low over the oviposition site, the male lowering his abdomen to cause the end of the female's abdomen to dip into the water and release eggs.

Noncontact guarding When males do not maintain females in tandem, they guard females by flying or perching nearby and warding off any intruding males. Males display toward and chase intruding males; male-to-male body and wing clashes may ensue. Some males guard multiple females. When male density is high, intensity of guarding ovipositing females probably increases in most species, but it has been reported to decrease in at least one species. In many Zygoptera, the male takes the female in tandem to the oviposition site, then releases her and either does or does not guard her. In some Trameinae, a subfamily of Libellulidae, the male releases the female as they fly in tandem over the oviposition site; she drops down to the water surface and releases a few eggs, then she flies back upward and the male takes her back into tandem. These latter two behaviors illustrate combinations of contact and noncontact guarding.

FORAGING Odonata feed on living prey throughout their adult life. When foraging, dragonflies can be categorized as "perchers" or "fliers." Perchers spend much of their time stationary, making short flights from perches to capture prey and then perching to consume it. In contrast, fliers are on the wing for a large part of their feeding activity,

capturing their prey in the air and swallowing small prey while in flight; they do, however, perch to consume larger prey. The "flier" mode requires much more energy, but fliers are more opportunistic feeders, able to forage later in the day. Perchers typically capture most of their prey during midday. This dichotomy of foraging styles results because fliers generate more body heat than perchers and can therefore remain active at lower air temperatures.

The major stimulus for detecting prey is movement. Odonata have very large eyes with many ommatidia, a specialization for detection of movement. However, a few species take stationary prey, apparently recognizing the prey by its shape. Prey such as small flying insects may be captured directly with the mouthparts, but the legs are also used for subduing certain types of prey. Odonata are typically generalists with few exceptions. Diptera, especially mosquitoes and midges, are a major component of the adult diet. One analysis found that Chironomidae constituted a significantly higher proportion of the gut contents than did Culicidae, probably reflecting differences in the flying and perching habits of midges and mosquitoes. Some species take mainly large prey, such as Lepidoptera and Odonata. For example, the large North American gomphid *Hagenius brevistylus* often has been observed feeding on other Anisoptera and has been dubbed the "dragonhunter." Members of the Neotropical family Pseudostigmatidae are specialist feeders. They glean small spiders in the rain forest by vertically searching trees, hovering near webs found on leaf tips, then flying directly up to the webs and snatching the spiders from their perches.

THERMOREGULATION Although insects are basically ectothermic, large species are able to generate body heat or adopt body positions to absorb sunlight and are able to maintain this heat gain via certain behavioral mechanisms. Odonata are among those with ecto- and endothermic thermoregulatory capabilities. The two basic behavioral styles, fliers and perchers, use different strategies to prolong activity under less than optimal ambient temperatures. Under cool conditions, fliers warm the thorax by wing-whirring (endothermy), whereas perchers expose as much of the surface area as possible to solar radiation. In some species of perchers, hairs on the thorax serve as insulators, or the wings may be deflected downward to insulate the thorax. Such species are among the first to appear in the spring at higher latitudes. Under very warm or hot conditions, perchers remain stationary longer and posture their bodies to absorb less solar radiation. A common posture is the obelisk position, in which the abdomen is raised to expose the minimum surface area to the sun and the wings are lowered to reflect sunlight away from the thorax. Fliers generally become inactive during midday and hang up in the shade. However, some species of Libellulidae glide, and some species of Aeshnidae are able to continue flying by shunting warm blood from the thorax to the abdomen, where excess heat is dissipated. By prolonging activity at the breeding site, dragonflies increase their chances of obtaining mates, feeding, and escaping predation.

Dispersal

Most flight involves small-scale, intrahabitat movements for immediate needs (feeding, finding mates, and escaping predators) that directly affect individual survival and reproductive success. Such flights usually result in dispersal distances up to a few hundred meters. Large-scale flight resulting in interhabitat displacement is regarded as migratory flight. Corbet defined migration as "spatial displacement that typically entails part or all of a population leaving the habitat where emergence took place and moving to a new habitat in which reproduction ensues." These dispersal movements also have consequences for survival and reproductive success.

In examples of migration thus far elucidated for tropical species, migration is a means of overcoming drought in the area where the species developed. For example, in ephemeral lentic habitats in Africa, the aeshnid *Hemianax ephippiger* develops rapidly (within 60–90 days) and upon emergence flies with rain-developing systems several hundred kilometers to areas that will receive the rainfall, as far north as Europe. There they feed, mate, and lay eggs in newly filled water bodies. Temperate species migrate to circumvent cold temperatures. For example, the wide-ranging aeshnid *Anax junius* emerges early in the year in southern North America, and arrives in the northern United States and southern Canada during warm periods as early as March and April. These immigrants mate and lay eggs in shallow lentic habitats; their progeny complete development in late summer or fall. The second generation then flies south; large numbers of "green darners" have been observed flying overhead as late as mid-November. Migratory flights may be made up of several species, usually from the families Aeshnidae and Libellulidae. Very few Zygoptera are known to migrate, and all recorded so far are in the subfamily Ischnurinae (Coenagrionidae).

ODONATA SURVIVAL IN A CHANGING WORLD

Habitat creation, loss, and alteration are the major causes of odonate population changes. Some odonate species have increased their geographic ranges and population numbers in response to man-made changes in habitat. For example, in the United States many pond-dwelling Libellulidae that were historically centered in the eastern part of the country have moved far west of the Mississippi River with advent of irrigation. Furthermore, exotic species can immigrate when gravid females ride tropical storms or can be introduced as eggs and larvae with the aquarium trade, their ranges thereby increasing dramatically. Alternately, many riverine and wetland species have undoubtedly declined because of habitat degradation and drainage changes. For example, some riverine Gomphidae are extremely rare, but could be protected by conserving the remaining habitat (e.g., in the eastern United States, *Ophiogomphus edmundo* is known from three localities, and *Gomphus sandrius* is known from seven localities). Only one species in the United States has

federal protection status as a threatened and endangered species, Hine's emerald *(Somatochlora hineana).* Many species in this genus are locally distributed, inhabiting lakes and bogs in different stages of succession, and their populations depend greatly on the availability of the proper microhabitat.

There is awareness of the threat to dragonfly diversity and populations of sensitive species in most countries, especially in Europe and Japan, where protection efforts are designed to heal or prevent damaged ecosystems. In tropical areas, however, where diversity is highest and incompletely known, habitat destruction continues at alarming rates. There has been some effort toward habitat conservation, as national parks and preserves have been established in many tropical countries. For example, in Thailand, almost all remaining forest areas are protected by parks, wildlife sanctuaries, and a ban on logging; this effort amounts to nearly 15% of the total land area. In other countries, the situation is less promising, and even preserves afford no insurance against habitat alteration. Odonata have existed for many millions of years, undoubtedly surviving small and massive extinction episodes; however, the magnitude of present-day environmental change may be without parallel.

See Also the Following Articles

Aquatic Habitats • Fossil Record • Mating Behaviors • Sexual Selection • Thermoregulation

Further Reading

Alcock, J. (1994). Postinsemination associations between males and females in insects: The mateguarding hypothesis. *Annu. Rev. Entomol.* **39,** 1–21.

Arnquist, G., and Johansson, F. (1998). Ontogenetic reaction norms of predator-induced defensive morphology in dragonfly larvae. *Ecology* **79,** 1847–1858.

Corbet, P. S. (1980). Biology of Odonata. *Annu. Rev. Entomol.* **25,** 189–217.

Corbet, P. S. (1999). "Dragonflies: Behavior and Ecology of Odonata." Cornell University Press, Ithaca, NY.

Davies, D. A. L., and Tobin, P. (1984). "The Dragonflies of the World: A Systematic List of the Extant Species of Odonata," Vol. 1, "Zygoptera, Anisozygoptera." Societas Internationalis Odonatologica Rapid Communications (Suppl.) No. 3, Utrecht.

Davies, D. A. L., and Tobin, P. (1985). "The Dragonflies of the World: A Systematic List of Extant Species of Odonata." Vol. 2, "Anisoptera." Societos Internationalis Odonatologica Rapid Communications (suppl.) No. 5, Utrecht.

Dunkle, Sidney W. (2000). "Dragonflies Through Binoculars: A Field Guide to Dragonflies of North America." Oxford University Press, New York.

Fincke, O. M. (1987). Female monogamy in the damselfly *Ischnura verticalis* Say (Zygoptera: Coenagrionidae). *Odonatologica* **16,** 129–143.

Jacobs, M. E. (1955). Studies on territorialism and sexual selection in dragonflies. *Ecology* **36,** 566–586.

Schorr, M., Lindeboom, M., and Paulson, D. (2000). List of Odonata of the world. Available at http//www.ups.edu/biology/museum/worldodonates.html

Tennessen, K. J. (1997). Aquatic insect resource management. *In* "Aquatic Fauna in Peril: The Southeastern Perspective" (G. W. Benz and D. E. Collins, eds.). Special Publication 1, Southeast Aquatic Research Institute, Lenz Design & Communications, Decatur, GA.

Tillyard, R. J. (1917). "The Biology of Dragonflies (Odonata or Paraneuroptera)." Cambridge University Press, Oxford, U.K.

Waage, J. K. (1979). Dual function of the damselfly penis: Sperm removal and transfer. *Science* **203,** 916–918.

Wootton, R. J. (1988). The historical ecology of aquatic insects: An overview. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **62,** 477–492.

Orientation

Ring T. Cardé

University of California, Riverside

Orientation refers to the way in which organisms direct their course of movement—not to their body or positional orientation per se, although clearly, if an organism is to aim its course, it must align its body's long axis with its intended track. Although orientation maneuvers thus involve an organism's orientation of body position, this article is concerned mainly with movements in location that range in scale from millimeters to thousands of kilometers. For example, a minute parasitoid wasp may walk to a resource, such as a prospective host, only millimeters away. On the other hand, movements of insects also can cover considerable distances, such as the long-distance migration of several thousands of kilometers that is undertaken in autumn by monarch butterflies *(Danaus plexippus)* flying from northeastern North America to their overwintering site in central Mexico. The mechanisms that insects and other organisms use to accomplish such feats are enormously variable. Moment-to-moment steering typically relies on simultaneous inputs from multiple sensory modalities, such as chemical cues, light, and wind. Most organisms use some stored information about very recent encounters with these cues and the organism's past position. In many parasitic and social Hymenoptera, learned information, including spatial maps and landmarks, plays a crucial role in these insects "knowing" either where they have been or their destination.

CLASSIFICATION OF ORIENTATION MANEUVERS

The modern system of categorizing orientation mechanisms by their forms of locomotion and their presumed sensory inputs dates to Fraenkel and Gunn's *The Orientation of Animals,* first published in 1940. These authors' classification of maneuvers relies on two distinctive patterns of movement. The first kind of maneuver is termed a kinesis (pl. kineses); it is defined as an undirected response in which the body's long axis exhibits no consistent relationship to the direction of the stimulus and the direction of locomotion is random. If a gradient of stimulus intensity regulates either the frequency of turns or the amount of turning per unit of time, the reaction is termed a klinokinesis. If the gradient of stimulus intensity regulates either speed or the frequency of locomotion, then the reaction is termed an orthokinesis. Both kinds of kinesis require a minimum of one sensory detector to monitor stimulus intensity, although multiple detector systems (such as paired antennae) are the norm. Randomly directed movement would seem to be an ineffective means for moving toward or away from a stimulus. Klinokinetic maneuvers can facilitate movement either toward or away from a stimulus gradient as illustrated in Fig. 1 where simply turning more frequently in bright light

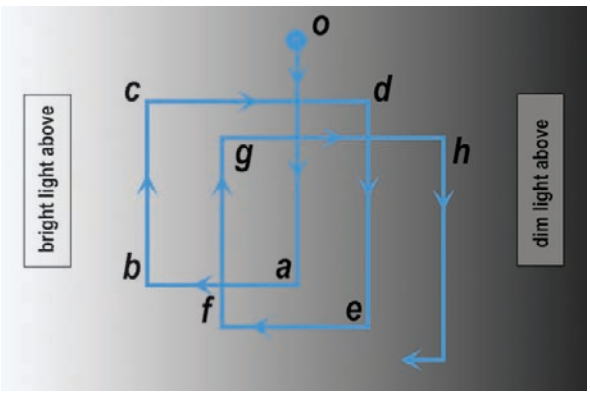

FIGURE 1 Track that could be expected from a hypothetical animal that always changes its direction by 90° to the right at a rate that is dependent on light intensity. The animal starts at point *O*. Turns occur more frequently when the animal is in bright light. Because of this tendency, the path *cd* is longer than *ab*. Although this is a simplistic representation of a klinokinetic reaction (in reality, turns will vary in angle and may be random to the right or left), it nonetheless demonstrates how klinokinesis can result in an animal orienting along a shallow stimulus gradient. [Modified from Fraenckel, F., and Gunn, L. L. (1940).]

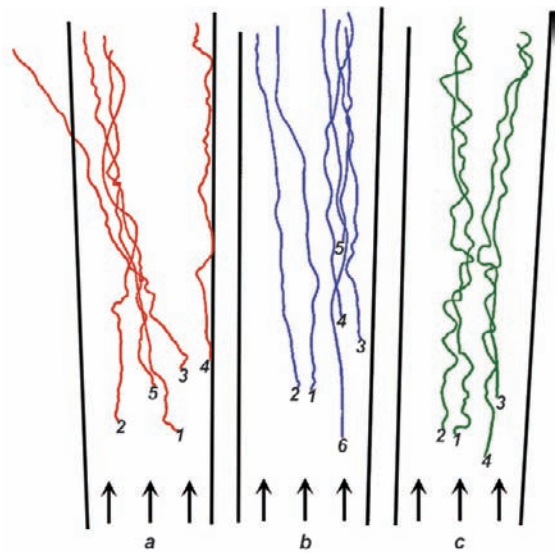

FIGURE 2 Courses (viewed from above) of blowfly *(Lucilia)* larvae crawling away from a light source (arrows depict direction of horizontal rays of light). The three individual larvae recorded are denoted "a" through "c" in different colors, and the tracks taken in repeated trials are indicated by numbers. Individuals in "a" and "b" seemed to deviate to the left in some trials and to the right in others. The track of larva "c" is represented in more detail than in the other tracks, showing alternating right and left movements of the head. Although J. Loeb stated that larvae "move as though they were impaled on a ray of light which passed through their medial plane," the larvae clearly have some variability in their paths. Based on the detailed head movements of the track of larva "c," larvae seem to orient by transverse klinokinesis. [Modified from Mast, S. O. (1911). "Light and the Behavior of Organisms." J. Wiley, New York.]

inevitably results in motion by the organism away from the bright light.

The second kind of maneuver is called a taxis (pl. taxes); it is defined as a directed reaction in which the organism's long body axis is aligned with the stimulus and movement is more or less directed toward or away from the stimulus. In klinotaxis the organism has available two strategies for sampling the intensity of the stimulus. In transverse klinotaxis the sampling occurs by moving the entire body or a part of it from side to side along the path. Alternatively, in longitudinal klinotaxis the organism samples intensity successively along its path. Both forms of klinotaxis require only a single detector capable of measuring stimulus intensity. A classic transverse klinotactic reaction is the movement of blow fly larvae *(Lucilia)* away from light (Fig. 2). A close examination of the movement of larvae along their path reveals that although their tracks are nearly straight, their heads move from side to side. A similar pattern is seen in ants following a pheromone trail (Fig. 3).

Tropotaxis, in contrast, relies on a paired detector system (such as the antennae); by balancing the stimulus intensity on the two sides of the organism, the heading can be aligned with a relatively steep stimulus gradient. Honey bee workers *(Apis),* for example, can center their body's long axis between balanced inputs of odor to each antenna. Telotaxis is a "direct" form of orientation in which the stimulus intensity is processed by a receptor system that has an array of directionally sensitive receptors, so that setting of a course toward a stimulus involves the relatively simple navigational task of holding a certain part of the receptor array in alignment with the stimulus. Sometimes termed "goal orientation," telotaxis is known only for orientation along a beam of light.

In klino-, tropo-, and telotactic reactions, the organism's long body axis is aligned with the direction of the stimulus, such as a beam of light or a fairly steep gradient of odor. In menotaxis, orientation of the long body axis is at a fixed angle to the stimulus, and course setting is maintained by stimulation of the sensory apparatus in a manner similar to telotaxis. Menotaxis is commonly called "compass orientation," and some still unresolved form of it is used by monarch butterflies to head toward their wintering grounds, and by honey bees in their dance language. A menotactic reaction also seems to be responsible for the attraction to lights seen in many moths and other nocturnal insects. The assumption is that moths follow a straightened-out path at night by using celestial cues as a menotactic guide. When they encounter an artificial point source of light, they attempt to maintain the same angle with respect to the fixed point as in menotaxis, but in so doing they inevitably spiral toward the light source. Such a spiral path is indeed seen in the approach of some insects to a light.

CLASSIFICATION OF SENSORY INPUTS

In describing how organisms orient, it is common to create terms that combine the kinds of environmental cue used in orientation with the form of taxis or kinesis. Common prefixes used include anemo (wind), chemo (odor or taste), mechano (pressure), phono (sound), photo (light), rheo (water flow) and scoto (darkness). The *Lucilia* maggot moving away from light

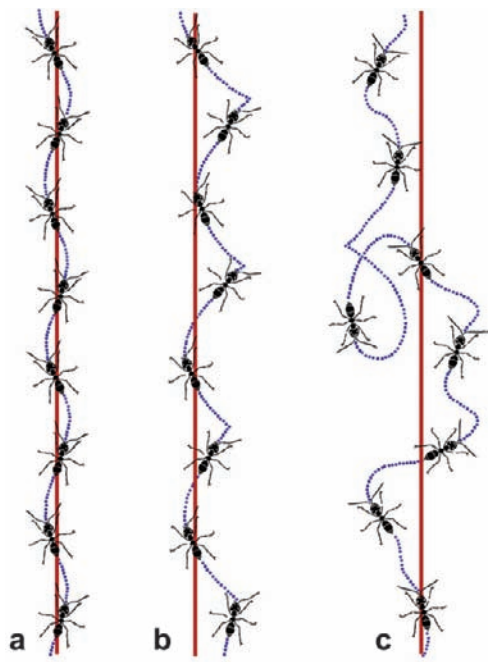

FIGURE 3 Trail following in worker ants *(Lasius fuliginonus)* in relatively still air. The straight red line marks the centerline of an odor trail. The dotted blue line denotes the path of the ant. The ant in "a" swings right and left across the trail, presumably bringing one antenna to an area of discernibly lower concentration before turning in the opposite direction. In "b," the ant's left antenna has been removed and she overcorrects her course to the right. In "c," the antennae were crossed and then glued, but the ant is able to orient along the trail, albeit with difficulty, aided partially by a light compass reaction. [Modified from Hangartner, W. (1967). *Z. Vgl. Physiol.* **57,** 103–136.]

can be said, for example, to be a navigating by a negative transverse photoklinokinesis. It is obvious that the seeming precision of this classification scheme makes for an unwieldy terminology. This deficiency was noted in 1984 by Bell and Cardé, who highlighted the need for "more practical, functionally related terms" that are not "teleological, poorly defined, non-probabilistic and difficult to spell." A related problem is that these terms tend to define the reactions so precisely that they can dictate the boundaries of experimental investigations, such that these may either neglect the integration of mechanisms or fail to consider mechanisms that do not fall within these constructs.

These classifications also neglect the role of internally stored information. The system of taxes and kineses assumes that an animal steers its path entirely in reference to the position of the external stimuli. However, the maneuvers can involve as well some self-steering that is governed by stored information about the animal's previous path. Such information can be classified as either idiothetic (i.e., information that is internally stored) or allothetic (i.e., information that is external, such as visual features of the environment). Kineses are, for example, clearly self-steered, whereas transverse and longitudinal klinokineses are partially self-steered, and menotaxis is not self-steered.

ALTERNATIVE CLASSIFICATIONS

"Attraction" and "aggregation" are often used to categorize orientation, but these terms describe end points of orientation and tell us little about the preceding maneuvers. Attraction and aggregation nonetheless remain widely used to describe the consequences of taxes and kineses. Pheromones, for example, are often labeled as attraction pheromones or aggregation pheromones.

In another approach to classifying orientation mechanisms, Jander emphasized in 1975 the distinction between the two broad categories of information used in orientation: information that is based on immediate sensory processing (for extrinsic or exokinetic orientation) and information that is stored centrally (for intrinsic or endokinetic orientation). Information that is stored may be subdivided into memory and that which is genetically determined. Jander also stressed the importance of ecology in studying orientation, and so his other categories included positional orientation (either staying in place or exhibiting locomotion), object orientation (movements with respect to the spatial position of either resources or sources of stress), topographic or home-range orientation (learned spatial orientation), and geographic orientation (migration over considerable distances).

Bell disavowed the time-tested system of taxes and kineses in his 1991 synthesis of foraging behavior. In analyzing the vast literature on foraging movements, Bell advocated describing the kinds of locomotory paths that were observed, the kinds of information available to mediate the motor output, and the presumed guidance system. Bell eschewed terming any of these maneuvers "taxis" or "kinesis." Despite such attempts to devise a new terminology, however, the system of taxes and kineses is likely to remain prevalent for some time because no clear alternative has emerged.

ODOR-INDUCED OPTOMOTOR ANEMOTAXIS

Among the best-studied orientation systems are those that enable organisms to locate upwind resources by flying along a plume of odor to the odor's source. Examples of such maneuvers include male moths flying over distances of hundreds and perhaps thousands of meters to a pheromone-releasing female, tsetse flies and mosquitoes flying over tens and perhaps hundreds of meters to a prospective vertebrate host, and parasitoid wasps flying over several meters to their intended host. All these reactions are mediated by odors that are released by the resource and form an odor plume as they are carried downwind. It is crucial to note that maneuvers cannot be governed by orienting to a gradient of odor. A gradient that would be sufficiently steep for such directional information exists only within a meter or less of the odor's source. Instead, insects and other organisms orient by advancing upwind when they encounter an above-threshold concentration of the odor linked to the resource.

The nonintuitive mechanism permitting in-flight anemotaxis is the optomotor response. An organism immersed in air

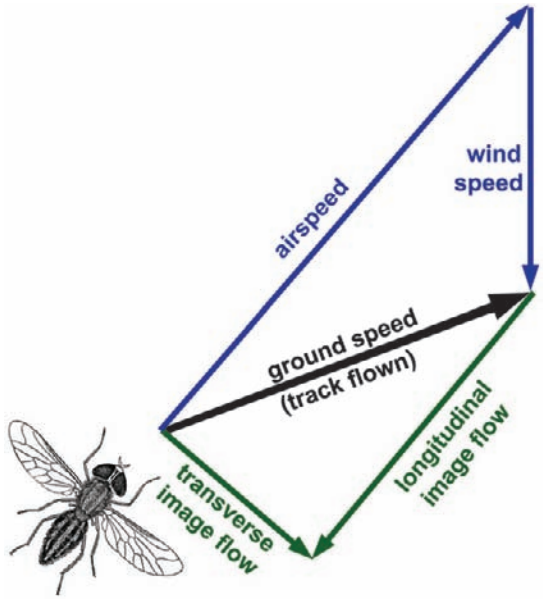

FIGURE 4 The relationship between a flying insect's body heading and the track taken when flying at an angle to the wind. The vectors depict the wind direction and velocity and the fly's direction and velocity along the track. The image flow the fly sees directly below has both longitudinal and transverse components, and therefore flows obliquely across the fly's eyes. When the fly heads directly upwind, the image flow is longitudinal, i.e., front to rear.

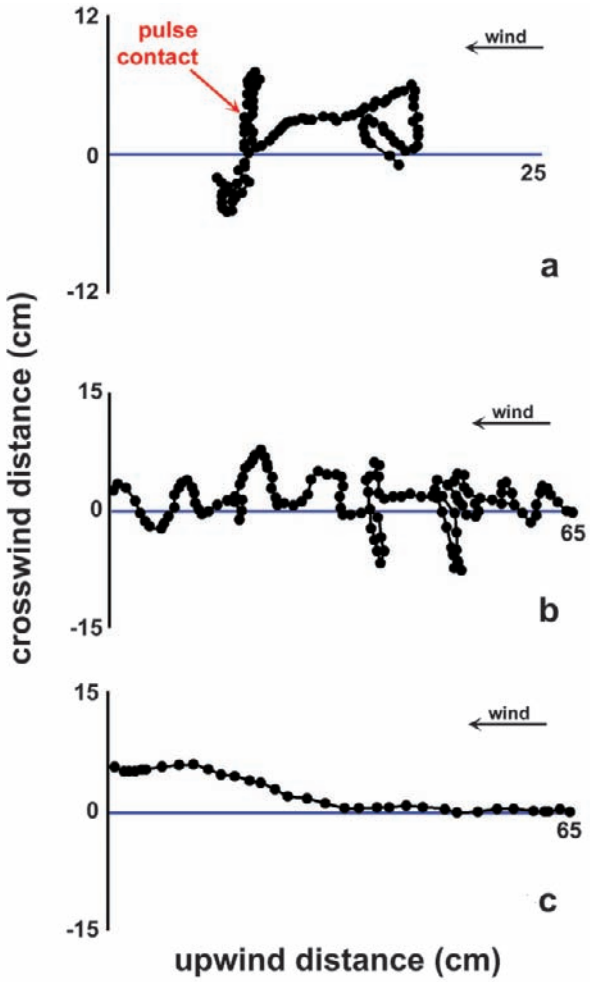

FIGURE 5 Top view of flight tracks of males of the moth *Cadra cautella* flying toward a source of a sex pheromone. The moth is traveling from left to right. The dots represent the moth's position every 1/30th of a second. The blue line shows the time-averaged centerline of the plume. Track "a" shows the path of a moth after intercepting a pulse of pheromone. [Redrawn from Mafra-Neto and Cardé (1994), *Nature* **369,** 142–144.] About 200 ms after intercepting a pulse, the male surges upwind. Track "b" shows a male flying along a very narrow ribbon plume of pheromone, sporadically encountering puffs of pheromone. Track "c" shows a male flying along a wide turbulent plume of pheromone, encountering many filaments of pheromone per second. [Redrawn from Mafra-Neto and Cardé (1995), *Physiological Entomology* **20,** 117–133.]

or water cannot discern the direction of the flow of these fluids by mechanosensory input, although it can use mechanosensory information to gauge its airspeed or water speed (i.e., its movement relative to the fluid flow). Instead, it detects its displacement relative to its ground position by literally seeing how the flow alters its path. For example, if an organism's long body axis is aligned directly with the fluid flow and the organism is making progress along the plume, then the image directly below or above the organism flows from front to rear. If the organism is moving at an angle to the fluid flow, then the image flows obliquely across the eyes (Fig. 4).

Optomotor anemotaxis was first demonstrated experimentally in 1939 by John Kennedy, working in England with the yellow fever mosquito, *Aedes aegypti*. Kennedy's ingenious wind tunnel used a movable pattern, projected onto the tunnel's floor, to manipulate the visual feedback a flying mosquito would experience. He was able to show that the upwind flight of females induced by carbon dioxide (the activating ingredient in human breath) was governed by the optomotor reaction. When the projected floor pattern was moved in the same direction as the airflow, mosquitoes decreased their airspeed, apparently perceiving by visual feedback from below that their airspeed had increased; conversely, when the image flow was reversed to the opposite direction, mosquitoes immediately increased their airspeed. Mosquitoes regulate their airspeed to maintain a relatively constant rate of longitudinal image flow. These simple manipulations verified that the mosquitoes' perception of movement relative to their visual surroundings dictates their airspeed, rather than some form of mechanosensory feedback.

The task an insect faces in finding an upwind source of odor, however, is much more complicated than simply flying upwind when an appropriate odor is encountered. Because turbulent forces cause the plume to meander and undulate, the direction of the plume's long axis is not always aligned with the upwind direction. Thus, an insect proceeding upwind often exits the plume. Thus many insects "lose" the plume and then "cast"; that is, they stop moving upwind and instead move back and forth lateral to the direction of wind flow. If they recontact the plume during such to-and-fro maneuvers, upwind flight may resume. A further difficultly is that turbulent forces fragment the plume's internal structure. Plumes

therefore have patchy distributions of odor, so that even within the plume's overall boundaries, insects encounter filaments of odor interspersed with gaps of clean air. Filaments can be encountered many times a second; for moths, whenever the encounters with filaments of odor are frequent, say above 10 Hz, flight can be aimed rapidly upwind, but when the encounter rate falls below 5 Hz, the flight can exhibit a much more substantial crosswind component (Fig. 5).

Odor-induced, optomotor anemotaxis as used by flying insects exemplifies several points common to orientation mechanisms. Several kinds of input (here visual, mechanosensory, olfactory) and self-steering all contribute to course setting and motor output. It is the integration of information that allows organisms to set their course.

Several related situations illustrate the variations on this theme. A flying insect also can orient along a plume of odor for moderate distances by flying a course that is set upon takeoff. After detecting the odor, the insect gauges upwind direction by mechanoreceptors and takes off due upwind. This is called an "aim-and-shoot" reaction, and the straight-line course is maintained by using visual cues perceived ahead to set the course. Flight continues as long as the insect remains in the plume. Tsetse flies (*Glossina* spp.) are thought sometimes to use aim-and-shoot upon takeoff, but other observations support a conventional optomotor anemotaxis maneuver during flight. Tsetse flies may shuttle between these two orientation strategies. Walking insects use a nonoptomotor version of anemotaxis. Upon detection of odor, the insect simply heads upwind, using mechanosensory input to determine wind direction.

CONCLUSION

Taxes and kineses remain the principal organizing system for understanding and investigating how insects and other organisms "know where to go." Discovering how these maneuvers work—what information is extracted, how it is processed, and the nature of guidance systems—remains an active area of inquiry.

See Also the Following Articles

Dance Language • Host Seeking • Magnetic Sense • Migration • Monarchs • Mosquitoes • Pheromones • Tsetse Fly

Further Reading

Bell, W. J. (1991). "Searching Behaviour: The Behavioural Ecology of Finding Resources." Chapman & Hall, London.

Bell, W. J., and Tobin, T. R. (1982). Chemo-orientation. *Biol. Rev. Cambr. Philos. Soc.* **57,** 219–260.

Bell, W. J., and Cardé, R. T. (eds.). (1984). "Chemical Ecology of Insects." Chapman & Hall, London.

Cardé, R. T. (1996). Odour plumes and odour-mediated flight in insects. *In:* "Olfaction in Mosquito–Host Interactions." Ciba Foundation. Symposium. 200, pp. 54–70. Wiley, Chichester, U.K.

Fraenckel, G. S., and Gunn, D. L. (1961). "The Orientation of Animals. Kineses, Taxes and Compass Reactions," 2nd ed. Dover, New York.

Jander, R. (1975). Ecological aspects of spatial orientation. *Annu. Rev. Ecol. Syst.* **6,** 171–188.

Kennedy, J. S. (1986). Some current issues in orientation to odour sources. *In:* "Mechanisms in Insect Olfaction" (T. L. Payne, M. C. Birch, and C. E. J. Kennedy, eds.), pp. 00–00. Clarendon Press, Oxford, U.K.

Schöne, H. (1984). "Spatial Orientation. The Spatial Control of Behavior in Animals and Man." Princeton Univresity Press, Princeton, NJ.

Orthoptera

(Grasshoppers, Locusts, Katydids, Crickets)

D. C. F. Rentz
California Academy of Sciences

You Ning Su
Australian National University

Orthoptera is considered here in the restricted sense; that is, we are not including the cockroaches, mantids, and stick insects, and these orders are covered elsewhere. The Orthoptera include terrestrial insects commonly known as short-horned grasshoppers, katydids, bush crickets, crickets, and locusts. The adult size range is from a few millimeters to some of the largest living insects, with bodies over 11.5 cm in length and wing spans more than 22 cm. Orthopterans occur all over the world except in the coldest parts of the earth's surface. They are best developed in the tropics, especially the New World tropics. In terms of numbers they are among the most common insects and are an important component of the fauna in most parts of the world. The order is readily identified by the characteristic hind legs that are developed for jumping. Summer nights are often dominated by the songs of many species in several families. Locusts are among the world's most important economic insects, and many species cause devastation in many parts of the world. Orthopterans are mentioned in biblical writings and in the earliest of Chinese literature. In recent times they have been important elements in the development of several fields of biology. Biological lifestyles in the Orthoptera include phytophilous (leaf-living), geophilous (living on and in the ground), cavernicolous (living in caves), and myrmecophilous (living with ants). Species can be diurnal and nocturnal. More than 20,000 species are known, but it is estimated that this figure may double when a thorough census has been made of uncollected regions of the globe.

CLASSIFICATION

There are several disparate classifications of orthopteroid insects that are used simultaneously, depending on preference. There has been an overall escalation of rank of categories in recent years above the tribal level. One of the extremes of these views was expressed by Dirsh, who created 10 orders from what was

TABLE I Some Characters Used to Separate the Two Suborders of the Order Orthoptera

Character	Ensifera	Caelifera
Antenna	More than 30 segments	Less than 30 segments
Auditory structure (when present)	On foretibia	On first abdominal tergite
Alary stridutatory structures (when present)	Forewings specialzed with file and scraper	Forewings modified laterally and ventrally
Ovipositor	Elongate, sword like or sickle-shaped	Short and stub-like
Molting	Skin usually eaten	Skin never eaten

traditionally considered one! Kevan provides a synthesis of the classifications to 1982. The advent of cladistics and molecular phylogeny has spawned additional changes to orthopteroid classification, and these are noted in the discussions of the respective groups. In general, except where noted, the conservative approach will be adopted here, since this is a period of flux and there is no consensus on which classification to adopt. The Orthoptera usually are divided into two suborders: the Ensifera (long-horned Orthoptera) (Table I) and the Caelifera (short-horned Orthoptera). The ensiferans are considered to be the more ancient group, with fossils dating from the Carboniferous, whereas caeliferans are known only from as far back as the Triassic (Fig. 1A).

FEATURES OF THE ORDER

General Comments

Orthoptera have been popular subjects for the behaviorist. Much has been done of an interdisciplinary nature relating to natural and sexual selection, signaling behavior, acoustic and vibrational communication, and displays.

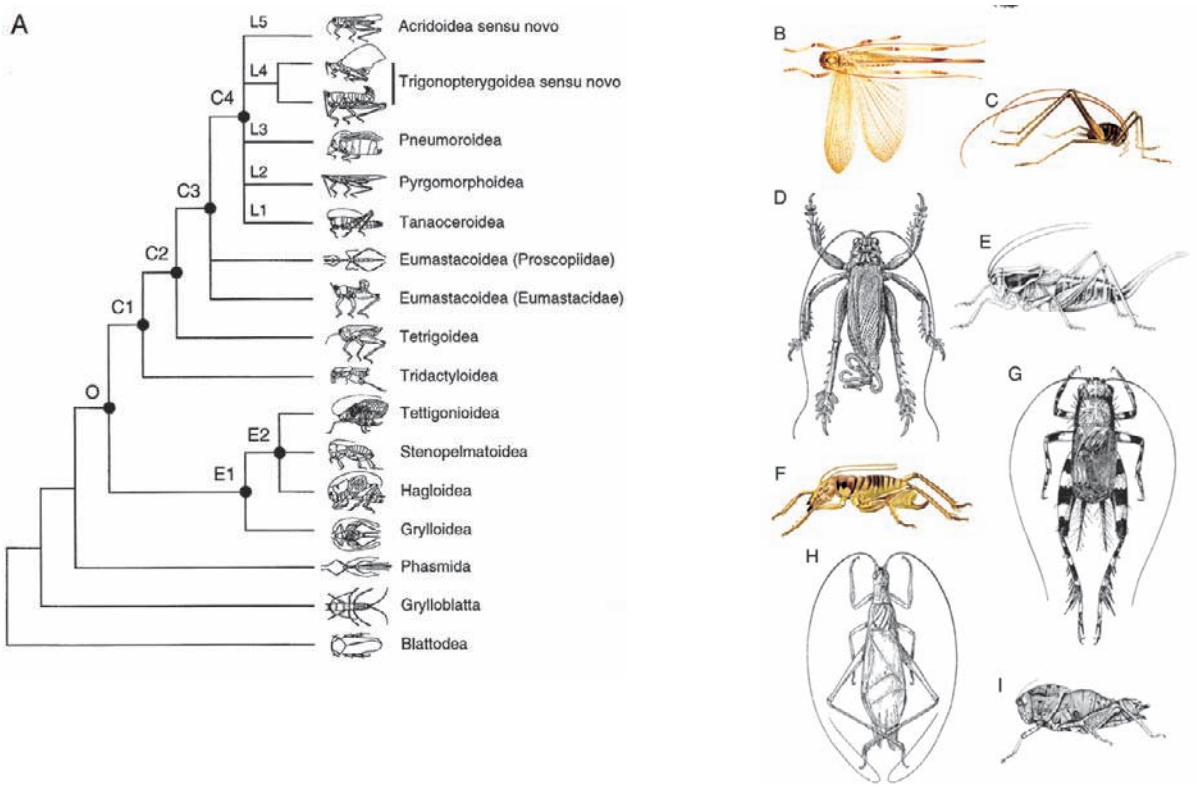

FIGURE 1 (A) Proposed phylogeny of the orthopteroid insects based on molecular studies. Note the arrangement of the Caelifera (C1). [From Flook, P. K., *et al.* (1999). *Syst. Biol.* **48,** Fig. 2, with permission of P. K. Flook.] (B) *Xanthogryllacris punctipennis,* an example of a gryllacridid with patterned wings. (C) *Hadenoecus puteanus,* a camel cricket illustrating the typical humpbacked appearance. [B and C modified from "Genera Insectorum." (1937). Vol. 206.] (D) *Schizodactylus monstrosus,* an unusual orthopteroid. [Modified from Karny, H. H. (1929). *Lignan Sci. J.*] (E) *Apote notabilis,* a large North American tettigoniine. [From "Genera Insectorum." (1908). Vol. 72.] (F) *Henicus prodigiosus,* a wingless South African henicine. [Modified from "Genera Insectorum." (1937). Vol. 206.] (G) *Dianemobius fascipes.* [From Chopard, L. (1969). "Fauna of India and the Adjacent Countries."] (H) *Oecanthus pellucens,* a typical oecanthine, male. [From Chopard, L. (1943). "Orthoptèroides de L'Afrique du Nord." Paris.] (I) *Stolliana sabulosa,* female wingless species. [From "Genera Insectorum." (1916). Vol. 170.]

Many orthopterans have excellent sight and hearing and are wary and difficult to approach. Others appear sluggish and rely on a number of cryptic strategies and distasteful properties for protection. For example, if seized, most species will kick out with their spiny hind legs and regurgitate acrid contents of the crop. Because orthopterans are fed upon by a vast number of vertebrates and many other insects, such behavior is not entirely successful. If a grasshopper, cricket, or katydid is seized by the hind leg, the leg will generally be lost at a point between the femur and trochanter. A number of unrelated species will squirt or discharge repugnatorial secretions from intersegmental glands.

The majority of orthopteran species are phytophagous, feeding on the foliage of higher plants. A number feed on roots and others on fungi. Many species are predaceous, while others are omnivorous. Relatively few species have been studied for the purpose of discerning their feeding activities, but some are very highly specialized, feeding only on seeds, pollen, or nectar or flowers of certain plant types. The fore and middle legs are used by some predators to form a "clap-trap" to catch unwary insects.

Orthopterans are renowned as examples of forms of crypsis. They span the range resembling leaves, twigs, bark, stones, or flowers. These appearances are meant to deceive vertebrate predators. To achieve this, various parts of the body are modified, and these combine with camouflaging colors and patterns, accompanied by the appropriate behavior. Mimicry of other insects abounds in the orthopteroids. Mimicry in the nymphal stages is often based on very different models from the adults. In a few examples, males have a different mimic model from females. The distasteful properties involved with aposematic coloration are exhibited in many orthopterans in a wide range of families and genera. The normally associated behavioral traits of sluggish behavior, conspicuous situations, and gregariousness are also seen. Flash or frightening colors are exhibited in a number of species. Several unrelated species in different parts of the world bury themselves in sand overnight or for short periods during the day when danger threatens. Many species overwinter in cracks or under leaves in leaf litter. A number of katydids and grasshoppers are semiaquatic. Some swim or skate over the surface film, and even nonaquatic species can swim freely in midwater. These species have subtle structural adaptations for aquatic life.

The majority of orthopterans are cryptozoic, lacking bright colors and patterns. They rely on pale, drab, and dull colors to conceal their presence. Cavernicolous species have pigmentation associated with their particular lifestyle in their habitat. The more highly specialized types lack pigmentation and have thin integument, as well as extraordinarily long antennae and long, delicate appendages. Extremes in structural modification are types that remain larviform as adults, having reduced eyes and antennae and nonjumping hind legs, but powerful burrowing adaptations. Some small crickets are dorsoventrally flattened and wingless and are myrmecophilous.

Reproduction

A diversity of attitudes is adopted in courtship and copulation. Sperm are transferred mostly in spermatophores, and much has been written regarding male investment, female selection of mates, and sperm competition. In the Ensifera, the spermatophore is transferred at copulation, and the vesicle usually remains attached externally to the vulva; it is eaten by the female and serves an important nutrient function in the development of the fertilized embryos. In the Acridoidea several small spermatophores may be inserted into the female tract, or the tubular part of the single spermatophore may penetrate the spermathecal duct while the vesicle remains in the phallus. Some other grasshoppers (the Australian endemic Morabinae of the Eumastacidae) produce no spermatophore, and their sperm is delivered directly to the bursa copulatrix, not the spermathecal duct.

Orthoptera eggs are most often laid in soil, but many other media are used. Many Ensifera insert them singly or in small numbers, into stems, leaves, or roots, or cement them to stems, twigs, or bark. One Australian species coats the exposed egg laid on a twig with soil particles, apparently to protect the egg from the elements and possibly from parasites. Acridoidea lay their eggs in oothecae or pods, in groups of more than 10 to 200, in pithy stems, in soft portions of dead timber, at the bases of grass tufts, or in animal dung; a number of species cement eggs to water plants below the surface. Burrowing forms deposit eggs in special chambers, lay them in the sides of chambers, or place the eggs around roots.

The ovipositor is often highly modified. In the Ensifera it is mostly a needlelike spear that is inserted into the substrate, usually the soil, wood, or other plant material. Many species lay specially hardened eggs that are glued to twigs or leaves. Many select a thick leaf, and with alternating penetrations of a highly modified, laterally compressed ovipositor, insert a disk-like egg into the leaf's edge. In Caelifera, the ovipositor is relatively similar among the wide range of species. It is short and "pronged," and it penetrates the substrate by opening, closing, and extending the abdomen. In the Acridoidea, the abdomen is frequently stretched to twice its normal length during oviposition.

Development

A newly hatched immature, or nymph, is enclosed in the embryonic cuticle and is called a pronymph or vermiform larva. At hatching the eggshell is fractured by pulsations of an extrusible structure known as the cervical ampulla, which is part of the dorsal membrane of the neck. This work is assisted by the cutting action of a ridge or row of teeth positioned at the front of the head. The ampulla is also involved in the emergence of the pronymph from the egg repository in the soil or plant tissue and in the intermediate molt by which the embryonic cuticle is cast when the insect is freed.

Nymphs differ from adults outwardly mainly in the reduced development of wings and reproductive organs, the less developed pronotum, and a reduced number of antennal segments. Nymphs undergo several molts, usually fewer in males than females. The intermediate stages of the developmental process are called instars. Molting occurs between instars and usually takes place at night under the protection of darkness or early morning, when the relative humidity is favorable for such processes. The wing buds and terminal reproductive structures increase with each molt. Antennal segments are added at molting. Many entomologists have great difficulty in attempting to determine whether a short-winged grasshopper is a nymph of a mature individual of a species. In actuality, it is a relatively easy determination but requires a little experience: grasshoppers have two pairs of wings; the first pair is the tegmina, and the second pair the wings. In immature grasshoppers, the wings are lateral with the costal margin positioned ventral, as in the adult, but in usually the penultimate molt, they rotate and assume a reversed, more dorsal position in which the costal margin is dorsal and the morphological ventral surface is external with the hind wing overlapping the tegmen.

REVIEW OF TAXA

Suborder Ensifera

Crickets and katydids (bush crickets, long-horned grasshoppers, and relatives) comprise the suborder Ensifera. This is the classical division of the order, but in the system using Grylloptera as a higher taxon, this group is treated as an order. Ensiferans share important characters with members of the other suborder, the caeliferans, such as the biting–chewing mouthparts, the modified pronotum, "leaping" hind legs, the similarities in wing venation and shape, and the sound-producing capacities (stridulation) of males.

More than 10,000 species are known in about 2000 genera with four infraorders, recognized here as superfamilies: Stenopelmatoidea (Gryllacridoidea), Tettigonioidea, and Grylloidea (including the Schizodactyloidea). The classification we follow is illustrated in Fig. 1A.

STENOPELMATOIDEA (GRYLLACRIDOIDEA, IN PART) Variously known as raspy crickets, leaf-rolling crickets, wood crickets, Jerusalem and sand crickets, king crickets, cave and camel crickets, and other names, the Stenopelmatoidea form a distinctive, conspicuous group with many species quite large and presenting a ferocious appearance. Most, if not all, are nocturnal, with all needing to seek shelter during daylight because of the threat of desiccation due to the thin integument. When the Grylloptera is recognized as a distinct order, this taxon with the Tettigoniidea constitutes the suborder Tettigonioidea. Because of their nocturnal and subterranean habits, most species are light colored, with brown or grays dominating the color scheme. Cave-dwelling species are often pale or white: green forms are very rare.

Species occupy most habitats, including the driest deserts and the wettest rain forests. The majority of species seem to be concentrated in the Old World, especially in the Southern Hemisphere. A few are considered to be minor crop pests, many attract public attention because of their size, and several are considered to be endangered by reason of habitat deprivation or loss due to introduced organisms like rats. The group comprises about 1000 species, but there are many undescribed species known from many parts of the world, especially Australia.

Gryllacrididae Recently considered as a subfamily of the Stenopelmatoidea, this taxon has been recognized as a family for more than 50 years. Gryllacridids are known by a variety of names. They are best known as raspy or leaf-rolling crickets but some are called king crickets or tree crickets. Most instars and adults spin threadlike "silk" from the mouthparts and use this material in reinforcing burrows and tying leaves and detritus together to form shelters. The integument of the body is soft and pliable (Fig. 1B), and the legs are adorned with many spines, some on the hind legs being modified for digging. They range in size from about 5 to 75 mm in body length. This family is best represented in the Old World. The majority of species are probably in Australia, with other regions of the Southern Hemisphere contributing considerable numbers.

Of the few species of gryllacridids that have been studied, all have been found to have peculiar life histories. The majority of these studies come from Australia, where the insects comprise a notable portion of the orthopteran fauna at a given locality. Species may construct burrows or tie leaves and detritus together, forming individual enclosures in which they reside during the day. Others live in the hollow branches or twigs and others assume a commensal lifestyle with termites.

Raspy crickets feed on a wide range of material. Some are specific seed eaters, others predaceous, and others specialize on flowers or fruits. This latter activity can be of economic concern when ripe fruit is chewed and when cut flowers or orchids are damaged by feeding activities. At times when gryllacridids inadvertently enter houses, curtains and draperies may be ruined by their chewing and clipping as bits of material are tied together with silk to make shelters.

Raspy crickets produce sounds in more than one way. All species possess a femoroabdominal stridulatory mechanism featuring a hind femur that is rough on the inside, with a pattern or shagreening of dorsal surface. This roughened area is rubbed against a pattern of pegs or modified hairs on the side of one or more abdominal tergites. These stridulatory mechanisms seem to show species-specific patterns and have been used to distinguish species. Vibrational sounds are produced, often by both sexes, during courtship activities. Depending on the species, sounds are generated by rhythmic drumming of the abdomen on the substrate or by "stomping" of the hind feet against the substrate. Some species use use combinations of drumming, stomping, and rasping during courtship. The fast-paced drumming is quite audible to the human ear for a short distance.

Surprisingly, none of these insects possess any obvious organs that would enable them to hear the sounds they produce. Perhaps, they detect the vibrations through sensory hairs on the pads of the tarsi.

Rhaphidophoridae Camel, cave, and sand-treader crickets are fairly similar in appearance (Fig. 1C). An extinct subfamily is known from amber inclusions. All members are apterous, but some can produce sounds by rubbing the inner faces of the hind femora against the opposing side of the abdomen and by rhythmic drumming of the abdomen against a substrate, be it the ground, a twig, or a branch. All have a humpbacked body structure with very long hind legs and antennae. Some of the sand-dwelling species have the hind legs modified into "sand baskets" for digging. Some groups are wholly confined to caves and others are obligate burrowers; the majority, however, live in leaf litter of dark crevices, where they spend the daylight hours and emerge only on humid nights to feed on detritus and leaf litter. Some feed on fungi, and there is at least one record of a cave-dwelling species that feeds on newly hatched birds. One camel cricket is cosmopolitan in its distribution, being moved in commerce, and is said to be a pest of greenhouses. About 350 species are known, with most species coming from the Indo-Australian area and Polynesia.

Schizodactylidae The splay-footed crickets are among the most bizarre of orthopteroids (Fig. 1D). There are fully winged as well as apterous species. They are broadly expanded by the possession of Lobelike or digitiform processes, which enable the crickets to run across dry sandy surfaces with efficiency. One interesting feature of the group is that members that have been studied have lower chromosome numbers than most typical orthopteroids. The splay-footed crickets are primarily predaceous, but one is considered to be a minor crop pest. They occur in parts of India, Myanmar, southwest Asia, and South Africa.

Anostostomatidae (Stenopelmatidae, in part of authors) The Anostostomatidae is a relatively recently proposed taxon accommodating a large range of genera formerly included in the Stenopelmatidae. Included here are the giant king crickets of Australia, New Zealand, and southern Africa. The odd genus *Cooloola* from Australia and a variety of peculiar genera from western North America have been placed here, along with an odd entity from Chile.

This family includes the Parktown prawn, *Libanasidus vittatus,* which lives in the Johannesburg area, where it is very well known. Females measure in excess of 6 cm, and most people fear them. They enter houses and beds of sleeping residents at night, and their scuttling and kicking is a common cause of concern. Their vile-smelling feces, exuded when disturbed, contribute to the unpleasantness. Myths have arisen regarding this species: some people feel it is an alien invader or the result of a freak mutation. The notoriety of this cricket has been a useful vehicle for educating the public about the biodiversity crisis. One of the most interesting features of this cricket is the tusklike mandibles of males. In this respect, they are similar to those of the New Zealand weta. The mandibles of *L. vittatus* serve in digging burrows as well as in battles with other males.

The New Zealand tree weta of the Deinacridinae are often large (8 cm or more in length), secretive, and aggressive species. The biology of *Hemideina* spp. is very well known. Some live in tree holes in galleries initially made by beetles or moth larvae. Others live in natural crevices and cavities. They prefer living trees to dead wood. These insects are herbivorous, feeding on leaves, flowers, and fruit. They may scavenge recently killed invertebrates. They remain arboreal except for a short period when females descend to the ground to oviposit. Some species have a unique size polymorphism related to social behavior. There is allometric growth of the head and jaws, much as the Australian king crickets. Certain males have extra instars to gain the larger head and mandibles. There is a dominance hierarchy, with the largest megacephalic males commanding the most desirable resources, which include galleries and mature females. Smaller males spend more time defending their galleries. These males live in galleries with smaller apertures and have been advantaged in dealing with alien predators such as rats. Tree wetas are smooth and shiny, with contrasting bands on the abdomen. Other New Zealand weta live on the ground. Several species raise their hind legs vertical to the substrate when alarmed. With a female weighing 50 g and measuring 7.5 cm in body length, this is one of the more formidable insects in defense.

All members of the Anostostomatidae are vulnerable to alien predators. The large size of the adults of many species may be effective in battles with rats, for example, but other stages from eggs to moderate-sized nymphs are extremely vulnerable. As a result, several species are threatened with extinction.

Stenopelmatidae This family is now considered by at least one authority to comprise five subfamilies, including the Gryllacridinae and Schizodactylinae. However, these groups are so different from typical stenopelmatids in morphology and biology that they are considered here as separate families. The remaining three subfamilies include the well-known Stenopelmatinae, comprising the Jerusalem crickets. This subfamily will be included here. The largest genus, *Stenopelmatus,* occurs in North and Central America. These insects are often known locally as "potato bugs" because they have been dug up in garden or potato fields. The derivation of the common name of the group, Jerusalem crickets, is shrouded in mystery. These insects have behavior patterns similar to those described for several anostostomatids.

There may be 100 species of the genus *Stenopelmatus,* but only a handful have been described. They are all very similar in overall appearance and differ only in their sexual drumming activities. Sympatric species have different drumming patterns. Both sexes and nymphs can produce rhythmic sounds by drumming their abdomens against the sides of their burrows or on the surface of the ground. Some of the sounds are audible from 20 m distance. Surprisingly, the nymphs produce species-specific drumming patterns in their later instars. The production of these sounds may serve

to keep the species together and may be effective in areas where there are sympatric species.

Copulation is distinctive in this group. In at least one species, the male rolls on his side and, if receptive, so does the female. Mating will not occur unless both partners "roll." Then, facing in opposite directions, the male grabs one of the hind tibiae of the female, not damaging her, positions his hind tarsi near her coxae, and curls his abdomen between his and her hind legs toward her subgenital plate. After several minutes, the male's telescoping abdomen nears the female's subgenital plate, and he grasps her with his hooks. With this anchor he everts the phallic lobes and attacks a spermatophore with a larger spermatophylax. In many ensiferans the spermatophylax is eaten, providing a source of nutrition for the mother and her eggs, the Jerusalem cricket female, however, does not consume the organ. Sometimes, though, the female consumes the male after mating; males offer no resistance to cannibalistic females.

TETTIGONIIOIDEA This is the largest superfamily, with more than 6000 species. Katydids (bush crickets or longhorned grasshoppers) can be of economic concern at times. Some flying species can swarm in the manner of locusts. With the "proper" environmental conditions, even flightless species can build in numbers that can affect crops and cause serious losses. The most notable example in the latter category is the Mormon cricket *(Anabrus simplex)* of western North America. It damages a variety of crops and rangeland plants. In Australia, Asia, and Africa, meadow katydids of the genus *Conocephalus* can build in numbers and move in large swarms. On a local level, many species cause damage to rice and crops from time to time.

On the other hand, the same species are often predaceous, feeding on eggs and larvae of other more important crop pests. In tropical regions rice is attacked by the copiphorine *Euconocephalus* sp. Phaneropterines seem to provide the majority of species that damage crops. *Scudderia* spp. in North America and *Caedicia* spp. in Australia both feed on the developing fruits and new leaves of citrus varieties. *Phaneroptera* species feed on a wide range of crops as well as citrus and coffee, kapok, and mimosas. Another phaneropterine, *Holochlora pygmaea,* feeds only on tea. *Ducetia* species feed on rice on two continents. *Elimaea chloris* feeds on a range of economic plants, the most notable of which are soybean, sugarcane, tobacco, and tea. The pseudophylline *Chloracris prasina* damages cacao, dadap, and rubber by its ovipositional habits. The widespread tropical mecopodine *Mecopoda elongata* feeds on beans, betel, cassava, castor, dadap, maize, potato, rice, sorghum, and tobacco. There are abundant literature references to "sexavae" or coconut treehoppers of two genera, *Sexava* and *Segestidia.* They are major pests of coconut and oil palm. They also feed on banana, karuka, manila hemp, and sago palm. With agriculture ever expanding to more remote areas, additional species not usually associated with economic damage can be expected to cause economic problems.

Katydids are widely distributed throughout the world, but the majority of species can be expected in tropical regions, especially the New World tropics. Many species are arboreal or bush dwelling. Some live in reeds or grasses, and many live on the ground. A few species can be found in alpine regions far above treeline. Most species exhibit cryptic behavior, especially during the daytime when they are inactive. Other species are aposematic and display warning colors. These species are primarily diurnal in their habits. Immature stages (nymphs) of some species mimic wasps, ants, beetles, bugs, and spiders. Ancestors of katydids probably were predaceous, but this role is minor among the extant forms. The majority feed on foliage, flowers, and seeds or are omnivorous.

Haglidae (Prophalangopsidae of authors) The ambidextrous crickets or hump-winged crickets comprise a single living family that is divided into two subfamilies. One subfamily, the Prophalangopsinae, consists solely of the species *Prophalangopsis obscura,* represented by a unique specimen, which is recorded from northern India. The specimen is fully winged and appears to resemble a mixture of a tettigoniid and a gryllacridid. The second subfamily, the Cyrtophyllitinae, is represented by two living genera. All species are ground dwelling and live in coniferous forests, where they ascend tree boles after dark to broadcast their loud stridulations.

Tettigoniidae The katydids, bush crickets, and longhorned grasshoppers differ from the Haglidae by having a more advanced stridulatory specialization. Many subfamilies are recognized, with several having been escalated to family rank as a result of recent investigations. Following the pattern adopted at the beginning of this article, these controversial changes are not recognized here.

The "primitive" katydids have been considered to be those in which the antennal attachment to the head is low on the head, that is, below the halfway point of the eyes. But this placement is open to interpretation, and it is often difficult to determine just where the critical position is. As a result some "primitive" katydids may not be primitive at all. Some of the more notable subfamilies are highlighted here.

The Bradyporinae is represented by a single genus found in the eastern Mediterranean as far as Iran. The Ephipperinae are represented by eight genera, mostly of Mediterranean origin. The Hetrodinae contribute 13 genera in Africa that often occur in aggregations. In some species both sexes can stridulate. They are characterized by a spiny appearance. The Acridoxeninae or dead-leaf katydids, are remarkable insects. The group is represented by a single genus with two species in equatorial West Africa, where they are apparently found on spiny plants.

The Phaneropterinae, or leaf katydids, bush katydids, or lyre bush crickets, is the largest assemblage of genera in the family. It has been accorded family status by some investigators. About 2000 species are known, distributed throughout the world, but the majority of species are in tropical climes. Many species are involved in mimicry complexes. Most nymphs resemble foliage or plant parts, but many are mimics of ants, bugs, spiders or cicindelid ground beetles.

The Pseudophyllinae, true katydids or bark crickets, include giant fully winged, leaflike species (with wings spanning more

than 20 cm) and smaller, micropterous ones. Most species are splendid examples of twig, foliage, and bark mimicry. With some species that resemble leaves, transparent holes and irregular, "bitten" pieces, and even colors resembling fungus or lichens are not uncommon in this group. About 1000 species are known in 250 genera with the majority in the New World tropics. The species that provided the name "katydid" is a member of this subfamily. All known species are phytophagous, and all known species use the often large falcate ovipositor to deposit eggs in plant tissue, either in bark, twigs, or dead wood. Some species have unusual lifestyles. A Mexican species of *Pterophylla* has been observed crossing a stream underwater. A species of the southeast Asian genus *Callimenellus* occurs in marine littoral rock crevices.

The Microtettigoniinae are represented by a single genus, *Microtettigonia* from southern Australia. These are minute (males as small as 5 mm) micropterous diurnal katydids that are extraordinarily fast moving. They live in grasses and lilies, and they seem to feed on floral parts.

The Conocephalinae is a large cosmopolitan group that comprises at least four tribes with more than 1000 species. The Gondwanan-distributed Coniungopterini is represented by three genera: two occur in Australia and New Guinea, and a third is found in Chile. The genus *Conocephalus* is represented by more than 50 species in Australia alone. They are small, agile katydids with a relatively broad, blunt fastigium. They are similar in appearance and habit and often occur in large numbers. Some species are diurnal, others nocturnal, and still others are active both day and night. *Conocephalus* species can be of economic concern at times. At least one species has been distributed through commerce.

Other genera have a much smaller number of species and are widely distributed. The Copiphorini are often associated with grasses, where their slender, bladelike appearance renders them almost invisible as they perch on the stems. The species associated with grasses and reeds feed largely on the floral parts of the host plants, preferring seeds. The mandibles are unusually strong, and this is an adaptation for seed predation. The buzzing calls are familiar sounds to most residents and visitors to appropriate habitats on all continents, but few people have ever seen the producers of the sounds they frequently hear because of the insects' secretive and cryptic habits. Thus many attribute their sounds to cicadas. The other subfamily, the Agraeciini is a disparate assemblage that most likely comprises a number of higher taxa. Members of this group have an extraordinary size range, with some of the world's most robust species represented.

The Mecopodinae or Kutsuwa bush crickets are usually large (some species have wingspans > 20 cm), and most species are fully winged and resemble either dead or living leaves to a remarkable degree. Others are short winged or wingless in at least one sex. Females sometimes stridulate. Most species are confined to the Old World. The subfamily is represented by two tribes, the Mecopodini and Moristini (Sexavini). Members of three genera are called coconut treehoppers because of the damage they cause to coconuts. Some species of this subfamily are kept in cages in Asia for the songs they produce.

The Phyllophorinae, or giant leaf katydids, are among the largest of all tettigoniids, with wingspans great than 25 cm. Some 10 genera and 60 species are known, mostly from the rain forests of Indo-Malaysia, New Guinea, and northern Australia. They are related to the mecopodines and the phaneropterines.

The Phasmodinae, or stick katydids, are wholly confined to southwestern Australia. They occur in winter and spring and are largely gone as the hot days of summer approach. They are very elongate and wingless in both sexes. They can cause economic concern when they feed on wildflowers in plantings in parks adjacent to natural areas. The remarkable resemblance to Phasmatodea is one of the most striking examples of convergence in the orthopteroid insects.

The Zaprochilinae, or pollen- and nectar-feeding katydids, are represented by four genera comprising 17 species from Australia. They are usually gray, setose, soft-bodied insects, and some species are fully winged, with the tegmina held at an angle and rolled; some species are micropterous in males and wingless in females. *Kawanaphila* species have been important study organisms in studies of sexual selection.

The Tettigoniinae, comprising shield-backed katydids, wart-biters, or great green grasshoppers, is one of the largest and most widespread of tettigoniid subfamilies, with the majority of representatives in temperate climes of both hemispheres. The common name, shield-backed katydids, is derived from the development of the pronotum, which is often extreme. While most are predaceous or specialized feeders, the Mormon cricket *(Anabrus simplex),* the Coulee cricket *(Peranabrus scabricollis)* of North America, and *Decticoides brevipennis* of Africa can occur in large numbers during favorable years, causing enormous damage to crops. The wart-biter *(Decticus verrucivorus)* was used during the Middle Ages to "cure" warts by allowing the aggressive insects to bite them off. Perhaps a substance in the insect's saliva contributed to the cure. The subfamily is diverse in form, ranging from among the smallest of tettigoniids to some of the largest. Fully winged and micropterous (Fig. 1E) species are known.

The Onconotinae are unusual micropterous katydids living in shrubbery. The Saginae is a small but characteristic subfamily of voracious predators or large insects. Four genera are represented, all from the Old World. *Saga* species can enter a cataleptic state characterized by complete immobility lasting some 20 min followed by a slow recovery. The function of this behavior is unknown.

The Austrosaginae comprises five genera, some of which use the powerful mandibles to crack seeds and feed on fruits. Listroscelidinae is a disparate assemblage of genera, many of which will probably be moved to other subfamilies when the group is fully revised. All species appear to be predaceous, some nocturnal and others diurnal. Some of the world's most spectacular katydids are members of this subfamily. Many are capable of delivering a painful bite when handled. The center of distribution for this group seems to be Australia.

The Meconematinae, or oak crickets or swayers, are small greenish-yellow katydids that are mostly diurnal; their distribution is tropical or subtropical. The notable exception is *Meconema thalassinum,* which occurs in Europe and Asia and has been introduced into the eastern United States. These delicate katydids have large eyes on the top of an unusual heart-shaped head. The few observations that have been made record the katydids catching their small prey on the "jump," in midair. This is not the manner used by phisidine katydids to secure their prey. The meconematines have been confused with the Phisidini of the Listroscelidinae, which, although somewhat similar in appearance, have an entirely different lifestyle. The phisidine species are nocturnal predators and have typical listroscelidine eggs. Eggs of meconematines are incredibly large for the size of the katydid and can be seen through the thin integument of the female. They are laid in decaying wood or other plant tissue, such as galls made by wasps, with the flattened, hard cap protruding from the substrate. The Tympanophorinae, or timbrel bush crickets, comprise a single genus confined to temperate Australia. The tegmina are unique in that some species, bear four kinds of stridulatory teeth on an expanded "rib." There is evidence that soft structures on the rib may provide openings for a lubricant or a substance released from a chamber or reservoir beneath it. This feature is probably associated with the unusual reproductive biology of species. In other tettigoniids, females are drawn to stridulating males, or answer them with their own calls, and the pairs are eventually brought together. The situation must be different in the tympanophorines, since the females are flightless and incapable of producing sounds. Male stridulatory behavior is unusual: they sing from perches for a short time and then fly 30 m or more to another perch and continue their song. How a flightless female can attract the attention of the transient male is unknown. All tympanophorine species are nocturnal and predaceous. They use the short forelegs to capture small-insect prey.

GRYLLOIDEA The true crickets and mole crickets range in size from less than 1 mm to more than 6 cm. Most species possess "ground colors," and few green forms are known. This is an adaptation to living on or in the ground.

The Grylloidea can be divided into a few very unequal families. More than 350 genera are known, encompassing 3000 species that comprise the Gryllidae, only about six genera and some 70 species comprise the Gryllotalpidae, and only a few species are known in the Myrmecophilidae. Although the majority of species are tropical, large numbers occur in all the temperate parts of the world. All terrestrial habitats seem to be inhabited except the highest mountain peaks. A few species can "skate" on water surfaces, and several live in mangrove swamps, where they use the stems to submerge themselves in saltwater when danger threatens. Many species burrow deep into the ground and seldom emerge. Others are blind, lack pigmentation, and live deep in subterranean caves. A number of species are commensal, living with rodents, ants, termites, and even mankind. Eggs are laid singly and deposited in the ground or in plant tissue such as decaying wood or grass stems.

Many crickets are crop pests, and population explosions sporadically occur with devastating results. Invasions of human habitations by crickets cause angst owing to the interminable chirping. Crickets around the world, have been known to ravage foods and furnishings.

Several radical classifications have raised "classical" subfamilies to familial status. However, the boundaries between these taxa are often unclear and a broader, more classical view is adopted here.

Gryllidae The true crickets comprise the principal family of the superfamily. Body size can range from less than 5 mm to more than 50 mm. The family has been divided into a number of subfamilies, and various classifications elevate some of these taxa to family level. The Gryllinae contain the crickets known to almost everyone. There are more than 550 species in more than 75 genera known from all parts of the world. Field crickets and house crickets belong to this group. Almost all species live on the ground, and some construct elaborate burrows that are often modified to amplify sound production. The house cricket, *Acheta domesticus,* can be considered to be domesticated insect. It is used in commerce as a food for mammals, birds, and reptiles, and seems unable to exist for any length of time in nature.

Several cricket species are kept, mostly in Asian countries, for the songs they produce. They are the source of a rich folklore and are regaled in poetry and song. Fighting crickets are an important part of the social scene in many Asian countries. A number of the subfamilies are highlighted shortly. The short-tailed crickets, a group that is quite prominent in Asian culture, are notable for the extensive galleries they make. There are brood chambers where the young are looked after. Several species cause damage to crops by feeding or destroying the roots of the plants. Some authors have accorded them a subfamily of their own, Brachytrupinae, but the differences between them and the Gryllinae are small, and they are probably best regarded as a separate tribe, the Brachytrupini.

The Nemobiinae, or pygmy or dwarf crickets, comprise more than 200 species and, as the name suggests, are small. While most forms are winged, there are many apterous species only a few millimeters in body length (Fig. 1G) that can easily be mistaken for nymphs of other species. Several genera exist in marine habitats. A number of genera in several parts of the world are adapted for life deep in caves and lava tubes. The majority of species live on the ground, often in moist habitats. Their numbers can be incredible in certain situations. Both diurnal and nocturnal species are known.

The Trigonidiinae, or leaf-running or sword-tailed crickets, are small, ground-dwelling or bush-dwelling crickets that are usually diurnal. They are active and often brightly colored. Some have a metallic sheen. Many species mimic spiders, ants, or wasps to mask their edibility.

Eggs are deposited in stems and bark. More than 300 species are known in more than 25 genera. Many genera are cosmopolitan or at least occur on more than one continent.

Preferred habitats are in rank vegetation, including shrubs, grasses, and small trees, but several genera occur in leaf litter. One genus with long tibial spurs can skate over water.

The Eneopterinae or bush crickets are small to medium-sized slender crickets. There is a range of color in this group depending on the habitats occupied. About 100 genera with more than 500 species are known, with the majority of species in the tribe Podoscirtini. Most of the many endemic crickets of Hawaii are in this subfamily. Most examples are from the tropics, with only a few making it to the temperate climes. These crickets usually live in shrubs, trees, herbaceous vegetation, or grasses. A small number can be found in leaf litter. Virtually none are of economic significance.

The Phalangopsinae, or spider crickets, are long-legged, rather flimsy crickets that are often gregarious in habit. The subfamily comprises more than 60 genera and 300 species, with the majority in the Old World tropics. Some species occur on the ground or in tree holes or decaying logs. Large numbers live in caves, crevices, or in the cavities created by large animals. Similar habitats afforded by dams, bridges, and buildings attract these crickets. True cave crickets belong to this subfamily. The cheerful songs of oriental species are prized, and these crickets are often kept in cages for such reasons.

The Sclerogryllinae, or stiff-winged crickets, are represented by only a single genus and several species. These crickets occur in Africa and southern and eastern Asia, where they live in leaf litter. The Pteroplistinae, or feather-winged crickets, are represented by three genera and only seven species. The Cachoplistinae are sometimes called beetle crickets because members of at least one species strongly resemble small beetles.

The Mogoplistinae, or scale crickets, are a widespread and often common assemblage of small flattened species, all of which are covered with minute scales. About 15 genera are known, comprising nearly 200 species. They are very widespread, with the majority of species in the Old World tropics. Many species live on the leaf surfaces, and there is an abundance of species that live in leaf litter. A single European species is considered to be intertidal. One tribe is known to be associated with rodent burrows. Both nocturnal and diurnal species are known. They are mostly small, ranging from 1.0 to 20 mm. The beautiful tones and sequences of the male song are valued in the Orient, and several species are kept in cages and sold in market for their songs.

The Oecanthinae, or tree crickets, are often given familial status. This subfamily has a worldwide distribution, but only seven genera and many species are known. Their nocturnal calls are well known to many people, although the insects themselves are seldom seen by humans. Some males utilize holes in leaves to amplify and direct their songs. They are kept also in cages in Asian countries because of the melodious songs they produce. Tree crickets are both beneficial and important local crop pests. Nymphs are often predaceous, feeding on a variety of insects, including aphids and scales. At times females oviposit in developing fruit and have been involved in transmitting plant diseases. Tree crickets have a characteristic appearance (Fig. 1H) and are usually pale greenish white, ranging from 1.0 to 1.5 cm in length.

Myrmecophilidae The Myrmecophilidae, or ant crickets, are minute, scale-covered, wingless crickets. There are five genera in this small family, with about 50 species found throughout the world, although most are known from tropical and subtropical regions. They are associated with ants, often living in the nests with them or following them along their trails. Some species have a wide range of hosts; others are known from only one species. Their commensal relations are unclear, but they seem to be unable to live independently for any period of time. A few species exist with termites. Some species seem to be parthenogenetic. Eggs are relatively large for the size of the female producing them. They are robust crickets that live in association with ants.

Gryllotalpidae The mole crickets comprise five genera with about 60 species worldwide. Most species are found in the cosmopolitan genus *Gryllotalpa.* Mole crickets use the extraordinarily developed forelegs for digging deep, permanent galleries and foraging for plant roots. Some mole crickets are known to collect seeds and store them in larders in circular chambers underground for future use. Some species brood eggs in chambers, and in many species both sexes stridulate. The calling songs are often of short duration and very loud. The horn-shaped entrance chamber of the burrow is used differently by different species to increase the male's acoustical output. Most species are herbivorous, but a few are carnivorous. Several cause major damage to crops by feeding on roots, on seedlings, or both.

Suborder Caelifera

The short-horned grasshoppers and locusts and their relatives comprise the large and well-known suborder Caelifera. Rarely, this group is treated as a separate order.

Most caeliferans are diurnal, but increasingly investigators are discovering that many species are active both day and night. Males of many species stridulate in bright sunshine; rarely, those of others sing on warm nights. Although primarily tropical, many species occur in all parts of the world. Some are semiaquatic, and only a few are true burrowers. Almost all species feed on plant material, but many feed on dead members of their own or other species. None are commensal. Aposematic coloration is a feature among many species.

Copulation in the Caelifera is rather uniform, with the male clinging to the back of the female for a considerable period of time with the spermatophore elongated and occupying the female's genital tract. There is no visible external spermatophylax. Eggs are laid in pods or oothecae that are enclosed in a reticulate membrane and covered by a foamy secretion that dries out eventually. Eggs of most species are deposited directly into the ground, but others lay in plant tissue or in cracks in bark.

Many species are disposed to gregarious behavior and swarming, with locusts offering the best example of this behavior. All true locusts are placed in the Caelifera. However, many

other species not properly designated locusts can become extraordinarily numerous and cause great damage to crops and other vegetation. Many species are eaten in Asia and Africa, usually fried. Many others are dangerously poisonous, and deaths have been recorded from eating these insects in Africa. After fires, the predominant colors among nymphs are dark, often black. This characteristic, called fire melanism, is found in almost all grasshoppers in all areas. The time span for these changes is extremely short. However, some surviving adults that become very dark after a fire produce nymphs that show little or no melanism.

The more than 2000 genera and 11,000 species comprising the Caelifera are arranged in superfamilies, suborders, or infraorders depending on the classification followed. Here they are regarded as superfamilies and are as follows (Fig. 1): Acridoidea, Eumastacoidea, Pyrgomorphoidea, Tanaoceroidea, Pneumoroidea, Trigonopterygoidea, Tetrigoidea, and Tridactyloidea.

ACRIDOIDEA *(SENSU LATO)* The grasshoppers and locusts comprise the largest superfamily of the suborder Caelifera. More than 8000 species in more than 1500 genera are known worldwide. Adults range in size from less than 1 cm to more than 25 cm.

Many grasshoppers have a disparity in size between males and females, with the former being much smaller than the latter. In many species males ride "piggyback" on females, not always copulating but "protecting" their prize and defending their potential progeny from other males with similar intent.

Grasshoppers have a short, almost invisible ovipositor. The eggs are laid in a pod covered by a protective coating. Many species oviposit on grasses, wood, and other plant tissue. A North American species has been observed ovipositing in dry buffalo dung. A number of species, especially locusts, utilize "lekking sites" to oviposit. Large numbers of females lay in the same place year after year. Knowledge of such behavior can be valuable in planning the control of pest species. Many grasshoppers take advantage of walking tracks or nonpaved roads for such activity. In oviposition, the female uses muscular contractions to extend the abdomen into the ground, often telescoping many times its length. The digging process also is performed by contractions and is aided by the teeth at the end of the abdomen. In tropical species the eggs can hatch in 3 to 4 weeks. In temperate species the eggs enter a period of diapause and hatch later. In some species, only a portion of the eggs hatch the following year, with the remainder hatching over a period of years. This serves as a "safety valve," ensuring that the species survives in periods of unfavorable weather. Upon hatching, the vermiform larva wriggles to the surface of the soil and upon reaching the surface, molts into the miniature form that will eventually become the adult. Nymphs undergo a series of approximately six molts, each one resembling more and more the adult in color pattern and shape. Gregarious grasshoppers illustrate synchronized molting, with all nymphs shedding their skins within hours of one another.

Locusts are species that occasionally form dense migratory swarms. These are often so large that they cross oceans but most occur in inland regions. Worldwide, more than 20 species from several subfamilies of Acrididae form migratory locust swarms.

The number of families depends on the classification followed. The scheme followed here (Fig. 1A) reflects the latest molecular attempts to establish a phylogenetically based classification based on defensible evidence.

Pamphagidae The Pamphagidae comprises generally larger, sluggish grasshoppers often resembling stones and bark. They have an array of body shapes (Fig. 1I). The family has an Old World distribution with the majority of species occurring in Africa and a few in Europe and Asia. It is absent from Australasia.

Lentulidae The lentulid grasshoppers have been called "nymphlike" grasshoppers and range in size from 8 to 25 mm in body length. These (Fig. 2A) grasshoppers are wingless. They occur in southern, eastern, and central Africa. Most species live in bushes, and at least one species causes feeding damage to nursery stock. One species prevents a weedy shrub from becoming a pest.

Pauliniidae The Pauliniidae, or aquatic grasshoppers, are small to medium-sized species with a smooth body integument. This family is known only from South America. Although other grasshoppers are aquatic, these grasshoppers are more wholly aquatic than any others. They can skate on the surface or dive and swim beneath it. Eggs are laid on the submerged parts of water plants. Their terrestrial behavior seems to be mostly nocturnal. They feed on aquatic plants and one species, *Paulinia acuminata,* has been introduced into Africa for the control of *Salvinia.*

Tristiridae The Tristiridae, or Andean wingless grasshoppers, are small to moderate in size. The body shape is variable, but the integument is always wrinkled and the color brown or grayish. The family has three subfamilies found only in the mountains of Cordilleran and Patagonian South America. They are found at high altitudes (2800–3000 m). Their protective coloration renders them almost invisible on pebbles and gravel, but aside from that nothing is known of their biology.

Ommexechidae The Ommexechidae are called South American toad-hoppers. The family includes 12 genera with some 30 species, all from South America. They live on the ground, inhabiting "coarse" vegetation in dry, sandy, or stony areas. They are of no economic importance.

Romaleidae The Romaleidae, or lubber grasshoppers, are regarded as a subfamily of the Acrididae in most older works. Recent molecular investigations, however, show that full family status is more appropriate. These grasshoppers are moderately large to very large in size and are often very colorful. The family has about 200 species in three subfamilies with the majority of species found in the Americas. A small number of examples occur in the Old World, in India, Afghanistan, Iran, and eastern Africa. They are notably absent from Africa and Australia. They occur in many habitats from desert to tropical rain forest. Some live on the ground, closely resembling stones

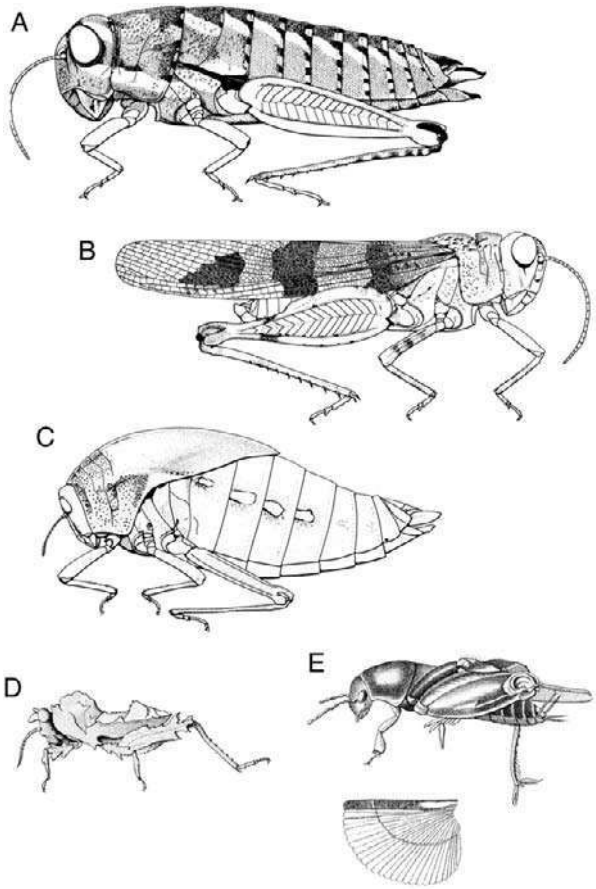

FIGURE 2 (A) A lentulid grasshopper, *Lentula callani* from Africa. [From Dirsh, V. M. (1975). "Classification of Acridomorphid Insects." Classey, Faringdon, U.K.] (B) *Oedipoda miniata,* now considered to be in the Acridinae, but many experts prefer to include it in a separate subfamily, Oedipodinae, with many other species. [From Dirsh, V. M. (1975). "Classification of Acridomorphid Insects." Classey, Faringdon, U.K.] (C) Flying gooseberry, *Bullacris unicolor,* male; tegmina and wings absent from females. [From Dirsh, V. M. (1975). "Classification of Acridomorphid Insects." Classey, Faringdon, U.K.] (D) *Cota saxosa,* a Peruvian species. Note flanges on legs and pronotum. [From "Genera Insectorum." (1906). Vol. 48.] (E) *Bruntridactylus tartarus.* Note fan-shaped hind wing. [From Saussure, H. (1874). "Voyage au Turkestan, Orthoptères."]

or "toads"; others are found in bushes, and still others high in trees. Many species show aposematic coloration, if not in the adult stage, then as nymphs. Some species can produce a hissing sound when disturbed, and bubbles can be produced from spiracles when the insects are greatly disturbed. In many aspects of both appearance and behavior, these grasshoppers resemble the pyrgomorphs. Few romaleines, however, are associated with grasses or sedges. The lubber grasshopper, *Brachystola magna,* of the southwestern United States, can be a serious road hazard when killed by automobiles. Their crushed bodies have caused cars to skid and accidents have occurred. The largest of grasshoppers occur in this family. The tropical American genera *Titanacris* and *Tropidacris* normally feed on vines, but at times they can become pests of plantations by defoliating plantation trees. They are called "giant locusts," but they are not really locusts at all because they do not form swarms.

Lathiceridae The Lathiceridae have been called desert gravel-hoppers. They are another peculiar small group known only from Africa. There are only four genera and five species, and nothing is known of their biology except that they burrow in sand and spend considerable time underground. They strongly resemble pebbles.

Pamphagodidae The Pamphagodidae (Charilaidae) are called the twin-keeled grasshoppers because of the parallel longitudinal carinae in the surface of the pronotum. There are four genera, comprising only five species, in this very small group from Africa.

Acrididae The Acrididae, or true locusts or grasshoppers, comprise the largest family of the suborder, and a full range of most of the features already described can be found here. Species range in size from less than 5 mm to more than 12 cm. The body form reflects grasshopper's role in the habitat. Short, stout, stone- or toadlike species are known, as well as stem- or twiglike grasshoppers. The colors span the complete range for the order, with browns or earth colors predominating.

Grasshoppers occur everywhere that orthopteroids are found. Many of the higher groups have restricted distributions, but the majority of species are tropical. Until recently, around 17 subfamilies were recognized, with more than 1500 genera. The characteristic songs are produced by males to attract females. Since these grasshoppers are primarily diurnal, color and behavior play important roles in species recognition. Some species have more than a dozen movements necessary for mate recognition. Elaborate colors and patterns and modifications of body pairs are associated with this behavior. But not all courtship occurs during the day. Some groups mate under the cover of darkness and rely on chemical clues to find mates. With these species, color is mostly protective, and bright colors are associated with aposematic features. Grasshoppers have a worldwide distribution and extend into some of the very cold regions of north and south. In this respect they occupy more territory than the Ensifera. Only a few subfamilies can be highlighted here.

The Oxyinae include many species associated with water and grasslands. *Oxya japonica*, one of the most destructive grasshoppers in rice in Southeast Asia, does only minimal damage in Australia. *Bermiella acuta* from Australia occurs on sedges, rushes, and grasses near the water and has adaptations for aquatic life such as dense, water-excluding patches of hairs on the distal abdominal sterna and tegmina, and an "air chamber" formed by the doming of the costal area of the tegmen over the first abdominal spiracle.

The Gomphocerinae are an important component of the acridid grasshoppers. The songs of stridulating males are a summertime characteristic in the meadows of the Northern Hemisphere. Several species of the same genus may occupy similar habitats and have similar feeding habitats. The calling songs of the respective species are different and serve as isolating mechanisms. Females are attracted only to the males that perform the calls of their own species.

The Catantopinae, or spur-throated grasshoppers, comprise many species (84% of the fauna in Australia). Some are very colorful. Many catantopines produce no audible sound, but some perform drumming actions on their host plants with the hind legs, thereby announcing the presence of a mate without the need of acoustical amplification. The responding members of the opposite sex merely move to the point of the drummer to consummate the union. A few other species produce a soft sound by rubbing the mandibles together. This can serve two functions. Some do this only when grasped, a startling reaction that might cause a would-be predator to drop the grasshopper. Others produce a similar sound from perches in shrubbery, obviously as an attractant to potential mates.

The Cyrtacanthacridinae, or large spur-throated grasshoppers, comprise some of the world's most important locust pests. *Valanga irregularis* from Australasia is one of the world's largest grasshoppers. It feeds on the leaves of trees and shrubs and can cause damage to fruit, nut, and plantation trees. Grasshoppers and locusts of the genus *Schistocerca* are members of this subfamily.

The Acridinae is a large group (Fig. 2B) with a worldwide distribution. Several locusts are included. Many are colorful and have brightly colored hind wings and slanted heads. This group contains some of the most phylogenetically advanced grasshoppers. Australia's most important locusts pest, *Chortoicetes terminifera,* is a member of this subfamily. The plague locust, *Locusta migratoria*, is also a member of the Acridinae.

TANAOCEROIDEA This superfamily includes a single family, the Tanaoceridae. Once considered to be related to both the eumastacids and xyronotids, the family Tanaoceridae comprises two genera with a few species from the deserts of western North America and Mexico. Members of *Mohavacris* have been found on sagebrush (*Artemisia tridentata*), where they closely resemble the bark of the thick stems. *Tanaocerus* species occur on the ground or on shrubbery. Both genera are nocturnal and are active on the cold nights of winter. Eggs are apparently laid in the ground and hatch in early autumn.

PYRGOMORPHOIDEA This superfamily now includes only the Pyrgomorphidae, comprising a most diverse assemblage of genera. The size and shape of the grasshoppers is extremely variable, but the head has a characteristic fastigial furrow. The family comprises some 30 tribes, and two, at the most, subfamilies are recognized. The family is mostly tropical, with the greatest number of species being from Africa and Madagascar, but the greatest diversity of genera is from the Australasia. In the Americas, the family is represented only in Mexico.

Pyrgomorphs live on bushes, herbage, grasses, and sedges and on soil and sand. Although many are camouflaged and show adaptation in body shape with twigs, grasses, leaves, and so on, many are stunning examples of aposematic coloration. They perform impressive displays when annoyed or threatened. Several species eject irritant fluids or a foamy froth when irritated. Not only are some of these substances toxic to mammal, birds, and reptiles, they are poisonous to humans. The majority of species oviposit in the ground like other grasshoppers, but some have been observed ovipositing in rotting logs, epiphytes and, most likely, the soil in trees caused by the presence of epiphytes. A few species are facultatively aquatic. Many species are gregarious, especially in the juvenile stages, and move together in the manner of locusts. At times they cause damage to crops but, in general, they are not pests.

PNEUMOROIDEA The bladder hoppers, or flying gooseberries, comprise a single, most peculiar family, the Pneumoridae. They are usually large with some species reaching 10 cm or more in length. The body shape of males is characteristic (Fig. 2C), that of females more resembling a normal grasshopper. The major part of the abdomen in males forms a huge, inflated, resonating chamber that is highly translucent. Females produce sounds but in a different way. Biological observations made more than 200 years ago indicate that males produce a loud noise after dark from shrubs in dry habitats. The sound, with its deep resonance, is often mistaken that of for a larger animal such as a bullfrog. There are two subfamilies with nine genera and some 20 species. They are all confined to southern and eastern Africa. They are of no economic importance.

EUMASTACOIDEA The Eumastaciodea consists of two families, the Proscopiidae and the Eumastacidae. Both have short antennae and an angular head that often appears too large for the body. Most species are wingless, but there are winged species that often resemble damselflies. The proscopiids, or false stick insects, are readily distinguished by elongate and twiglike appearance. They have an exaggerated head that appears too large for the body. All species are wingless, and some can be quite large (2.5–16.5 cm). Sixteen genera are all confined to South America.

The Eumactacidae, or monkey grasshoppers, is a much larger group. Many subfamilies have been recognized with hundreds of species. Some of these taxa have been regarded as separate families in some publications. Most species are small, with the body seldom exceeding 4.5 cm. Many are wingless, but there are many Old World forms that are fully winged and readily fly. Most species have a characteristic way, shared with the proscopiids, of sitting, exposed, with the hind legs splayed akimbo. They are diurnal and readily fly in the sun, and they resemble damselflies in several respects. Eumastacids feed on a variety of plant types ranging from grasses and sedges to desert shrubs and ferns in Old World rain forests. Probably the majority of species are nocturnal.

The Australian endemic subfamily, Morabinae, comprises 41 genera with 243 species. They are elongate, matchsticklike plain brown or green grasshoppers that do not sit with legs akimbo. They live on a wide variety of plants and are often very localized in their distribution. A fossil species of the extant genus *Erucius,* of the Eruciinae, has been discovered in Oligocene deposits in western North America. Present-day species in this genus are found in Malaysia and the Philippines.

TRIGONOPTERYGOIDEA This superfamily comprises two families: Trigonopterygidae and the Xyronotidae. The Trigonopterygidae are now not believed to be related to the pyrgomorphs or pneumorids (Fig. 1A). They have been called broad-leaf bushhoppers. There are two subfamilies. The group is confined to Asia, and several of the species occur in Borneo, where they live on the ground in dead leaves they resemble.

The Xyronotidae, or razor-backed bushhoppers, comprise a single genus with only two species with no known relatives. These grasshoppers live on the ground in leaf litter and have continuous generations. They are found only in Mexico.

TETRIGOIDEA This suborder contains the grouse locusts or pygmy grasshoppers. These are small usually gray, black, or mottled grasshoppers seldom exceeding 20 mm in body length (Fig. 2D). Depending on the classification, there are two families, or one family and many subfamilies. About 1000 species are known in 185 genera, with the majority of species in the Tetriginae. Most species live on the ground, most often on moist ground or along streams and ponds, where they feed on algae and diatoms. Members of the tribe Scelimini have members that are fully aquatic and can swim effectively underwater. In tropical climes some species are arboreal and live among lichens and mosses in tree buttresses or even higher in the canopy. Eggs of the terrestrial species are laid in the soil and bear a peculiar terminal filament that is directed upward when the eggs are laid. They are mostly of little economic concern, although some species are said to feed on rice.

TRIDACTYLOIDEA This small group contains some of the most bizarre of the orthopteroids. The pygmy mole crickets, pygmy sand crickets and mud crickets, and sand gropers are peculiar in a many respects. They range in size from less than 4 mm to more than 80 mm. Three families are recognized. The Tridactylidae and Rhipipterygidae are more closely related to one another than to the Cylindrachetidae.

Tridactylidae and Rhipipterygidae The tridactylids and rhipipterygids are small, usually variegated black, yellowish, or reddish minute cricketlike insects (Fig. 2E). The two families comprise about 210 species in 11 genera.

Most of the tridactylids and rhipipterygids are associated with damp habitats. They seem to be gregarious, and those that live in these situations can construct "nests" out of mud and debris to spend the night or to "hibernate." Many are active swimmers. There is another group that is arboreal, living on leaf surfaces in tropical climes, where they feed on the rain of particulate matter from the canopy. Many of these species have wasplike color patterns and wasplike "jerky" movements. Others living in the same habitats are dark blue and have white tips to the antennae.

Cylindrachetidae The Cylindrachetidae or sand gropers have an elongate or wormlike appearance. There are nine known species in three genera. They reflect a Gondwanan distribution, with species in Australia, New Guinea, and Patagonia. At least one species causes considerable damage to wheat in Western Australia, where large populations build up in the soil and feed on the root of the plants.

See Also the Following Articles

Crickets • Locusts

Further Reading

Bailey, W. J., and Rentz, D. C. F. (eds.). (1990). "The Tettigoniidae: Biology, Systematics and Evolution." Crawford House Press, Bathurst, Australia.

Chapman, R. F., and Joern, A. (eds.). (1990). "Biology of Grasshoppers." Wiley, New York.

Dirsh, V. M. (1975). "Classification of Acridomorphid Insects." E. W. Classey, Faringdon, U.K.

Field, L. (2001). "The Biology of Wetas, King Crickets and Their Allies." CAB International, Wallingford, Oxon, U.K.

Flook, P. K., Klee, S., and Rowell, C. H. F. (1999). Combined molecular phylogenetic analysis of the Orthoptera (Arthropoda, Insecta) and implications for their higher systematics. *Syst. Biol.,* **48,** 233–253.

Gwynne, D. T. (2001). "Katydids and Bush-crickets: Reproductive Behavior and Evolution of the Tettigoniidae." Cornell University Press, Ithaca, NY.

Naumann, I. D. (ed.). (1991). "The Insects of Australia, a Textbook for Students and Research Workers," Vols. 1, 2. Melbourne University Press, Carlton, Australia.

Preston-Mafham, K. (1990). "Grasshoppers and Mantids of the World." Blandford, London.

Rentz, D. C. F. (1996). "Grasshopper Country: The Abundant Orthopteroid Insects of Australia." New South Wales University Press, Sydney, Australia.

Ovarioles

Diana E. Wheeler

University of Arizona

Ovarioles are egg-producing tubules that are the fundamental units of ovaries in female insects. The number of ovarioles in each ovary is typically 4–8, but varies widely depending on the particular insect and its ecology.

GENERAL ARRANGEMENT AND STRUCTURE

Each ovariole is a tube in which oocytes form at one end and complete development as they reach the other. The terminal filament and the germarium, which contains germ cells, are at the distal end. Ovarioles may have one of several topological arrangements within an ovary. In some species ovarioles join the end of an oviduct radially around a central point. In others, ovarioles arise in single file off the oviduct, like teeth on a comb.

NUMBER

The number of ovarioles per ovary varies with taxon, size, and life history. All Lepidopteran females have four ovarioles,

but many groups tend to be more variable, both within and across species. Variability in ovariole number is particularly spectacular in social insects. Obligately sterile workers in ants can lack ovarioles entirely, and the most fecund queens in ants and termites have about 1200 ovarioles per ovary.

DEVELOPMENTAL ORIGIN

The period during which ovarioles form varies widely in insects, ranging from embryonic development in aphids to the pupal stage in flies. In some taxa, the number of ovarioles can be adjusted based on environmental factors. In *Drosophila,* for example, ovarioles form during the pupal period. This timing provides the opportunity for the number of ovarioles constructed to be adjusted based on previous diet and temperature. In honey bees, however, ovarioles form in early larval development. The number of ovarioles formed is at first the same in future queens and workers. In workers, however, most ovarioles undergo cell death, whereas those in developing queens persist.

SOMATIC TISSUE AROUND DEVELOPING OOCYTES

Each ovariole is made up of both somatic and germ cells. The somatic tissue includes a tubular sheath surrounding all the developing eggs as well as follicle cells around each oocyte. The sheath consists of inner and outer layers. The outer sheath is an open network of cells, sometimes containing muscle. The tissue is rich in lipids and glycogen and is metabolically active. Even so, there is no evidence of direct involvement in oocyte development. Tracheoles form part of the outer sheath but do not penetrate below it. The outer sheath can be important in sequestering bacterial symbionts that will be passed on to offspring. The inner sheath is a layer of extracellular matrix. In addition to physical support, the inner sheath can function as a sieve.

A layer of follicle cells surrounds each developing oocyte. Follicle cells are very active metabolically, contributing a variety of materials essential to developing eggs. During yolk uptake, follicle cells can separate slightly, allowing vitellogenin-laden hemolymph to contact the oocyte surface directly. The oocyte then can take up vitellogenin and other nutrients. As egg development nears completion, follicle cells secrete the eggshell, which consists of vitelline envelope and layers of chorion.

PATTERNS OF OOCYTE DEVELOPMENT

Ovarioles can be categorized based on how oocytes are produced from stem germ cells (Fig. 1). A stem germ cell produces two daughter cells when it divides; one remains a stem cell and the other becomes a cystoblast. Most commonly, cystoblasts undergo rounds of division but remain connected by intercellular bridges. This process is called cluster formation.

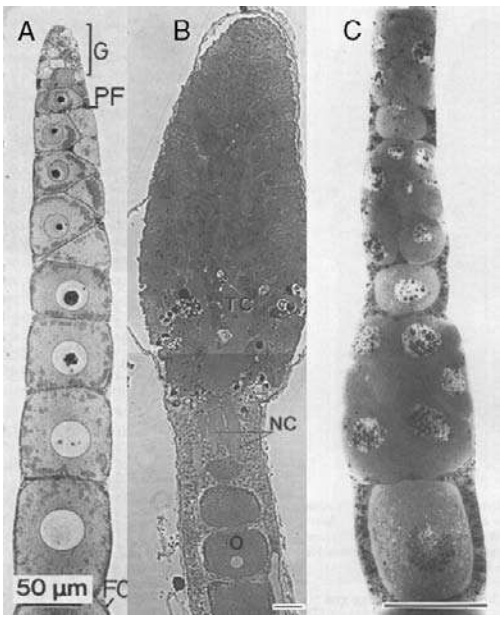

FIGURE 1 Patterns of oocyte development. (A) Panoistic ovary of an apterygote insect, *Lepisma saccharina.* G, germarium; FC, follicle cell. (Reprinted from R. C. King and J. Büning, 1985, The origin and functioning of insect oocytes and nurse cells. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," Vol. 1, pp. 37–82, with permission from Elsevier Science.) (B) Teletrophic ovary of a heteropteran insect, *Dysdercus intermedius.* The nutritive cords (NC) extending from the nurse cells to the oocyte are clearly visible. (C) Polytrophic ovary of a carabid beetle, *Nebria brevicollis.* (B and C reproduced, with kind permission from Kluwer Academic Publishers, from J. Büning, 1994, Figs. 3.35 and 3.57.)

Generally, only one of the daughter cells in a cluster becomes an oocyte, while the remainder degenerate or become nurse cells. The type of oocyte development in which an oocyte is connected to sister cells that contribute to the contents of the egg is termed meroistic.

Stem cells in ovarioles of the most primitive insects do not undergo cluster formation. Instead, each cystoblast develops into an oocyte. Such ovarioles are called panoistic. Insect orders that have primitively panoistic ovaries are Archeognatha, Thysanura, Odonata, Embioptera, and most of the orthopteroid orders. Plecoptera shows an interesting intermediate condition, in which cluster formation occurs, but all the daughter cells become oocytes. Panoistic ovaries are also found in more recently derived insect groups, in which the design has evolved secondarily, and these taxa include Siphonaptera, Strepsiptera, and Thysanoptera.

Meroistic ovaries can be subdivided into two types based on the location of accessory germ cells relative to the oocyte. In polytrophic ovaries, accessory germ cells maintain short connections and accompany the oocyte as it moves down the ovariole. In telotrophic ovaries, accessory tissue remains at the distal end of the ovariole.

In polytrophic ovaries, follicle cells encapsulate both the nurse cells and the oocyte. Nurse cells generally are polyploid and express genes whose products are transported to the

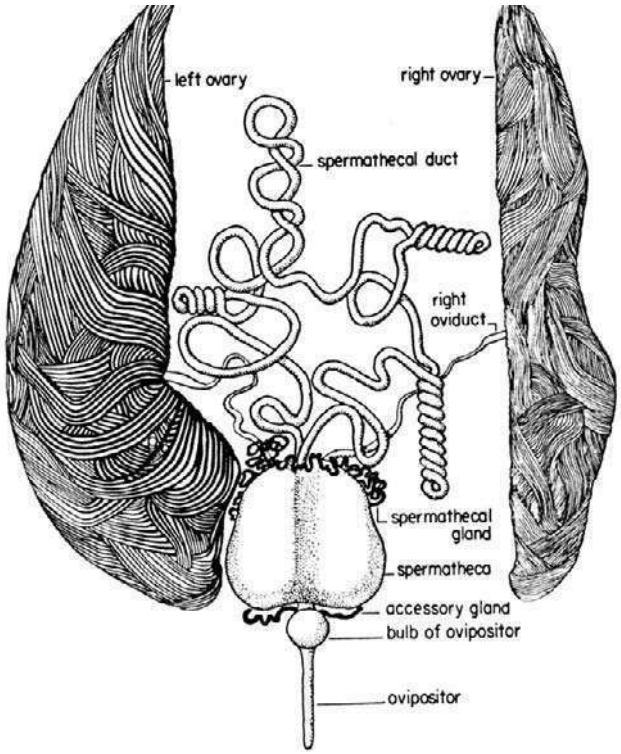

FIGURE 2 An army ant queen *(Eciton burchelli)* shows extreme development of ovarioles in both number and length. [Reproduced from H. R. Hagan, 1954, The reproductive system of the army-ant queen, *Eciton (Eciton).* American Museum Novitates No. 1663. Courtesy of the American Museum of Natural History.]

growing egg. Trichoptera, Lepidoptera, Hymenoptera, Diptera, adephagan Coleoptera, Mecoptera, and Neuroptera all have polytrophic ovaries.

In telotrophic ovarioles, accessory nurse cells remain in the anterior part of the ovariole and so, as an oocyte moves away, cords of oocyte tissue lengthen to maintain the connection. Teletrophic ovaries have apparently evolved independently several times. In Heteroptera, Coleoptera (Polyphaga), Raphidioptera, and Megaloptera, they evolved from polytrophic ovaries. Telotrophic Ephemeroptera evolved directly from panoistic ancestors.

Ovariole architecture is related to both phylogeny and life history. Both panoistic and meroistic ovarioles can support very rapid egg production as illustrated by the more than 40,000 eggs/day produced by both the panoistic ovarioles of the most fecund termite queens and the polytrophic ovarioles of the most fecund ants (Fig. 2). An important consequence of panoistic vs meroistic eggs is duration of embryonic development. The design of meroistic ovarioles facilitates the production of eggs well-stocked with ribosomes, mRNA, tRNA, and other macromolecules. Such eggs generally develop rapidly, as exemplified by *Drosophila melanogaster.* In contrast, autonomous oocytes produced by panoistic ovaries require longer periods to complete embryogenesis.

See Also the Following Articles

Egg Coverings • Reproduction, Female • Vitellogenesis

Further Reading

Büning, J. (1994). "The Insect Ovary." Chapman & Hall. London.

Chapman, R. F. (1998). "The Insects: Structure and Function," pp. 298–312. Cambridge University Press, Cambridge, UK.

Engelmann, F. (1970). "The Physiology of Insect Reproduction," pp. 4–7 and 45–49. Pergamon, Oxford.

Snodgrass, R. E. (1935). "Principles of Insect Morphology," pp. 552–561. McGraw–Hill, New York.

Stys, P., and Bilinski, S. (1990). Ovariole types and the phylogeny of hexapods. *Biol. Rev.* **65,** 401–429.

Wigglesworth, V. B. (1939). "The Principles of Insect Physiology," pp. 376–377. Methuen, London.

Oviposition Behavior

Marc J. Klowden
University of Idaho

Oviposition behavior comprises one of the final steps in insect reproduction. It involves the deposition of the mature egg outside the body of the female and includes a series of behavioral and physiological events that begin with the movement of the egg through the oviduct and end with the placement of the egg on a substrate that will support the development of the larva. Specialized behaviors and structures on the female allow her to place the eggs within a protected environment during oviposition.

OVULATION

Insect eggs develop within the ovaries, the reproductive structures of the female that are composed of tapering units called ovarioles. The oocytes differentiate from germ cells within the germarium of the ovariole, and as they begin their downward movement in the ovariole they are first completely surrounded by a monolayer of follicle cells (Fig. 1). These follicle cells are involved in the transport from the hemolymph into the oocyte of substances that are stored for later use and nourish the embryo during its development. Nurse cells may also be present to provide the oocyte with other maternal contributions, such as messenger RNA and mitochondria. The nurse cells subsequently degenerate, and the bulk of the yolk proteins deposited in the oocyte cytoplasm must then cross the layer of follicle cells after their synthesis in the fat body. During a later stage of development, the follicle cells also synthesize the eggshell (or chorion), which provides protection and waterproofing once the egg has been laid. After deposition of the chorion, the follicle cells degenerate, leaving the chorion as the outermost egg layer. With the follicle cells no longer on the outside of the egg, the egg is free to move out of the ovariole

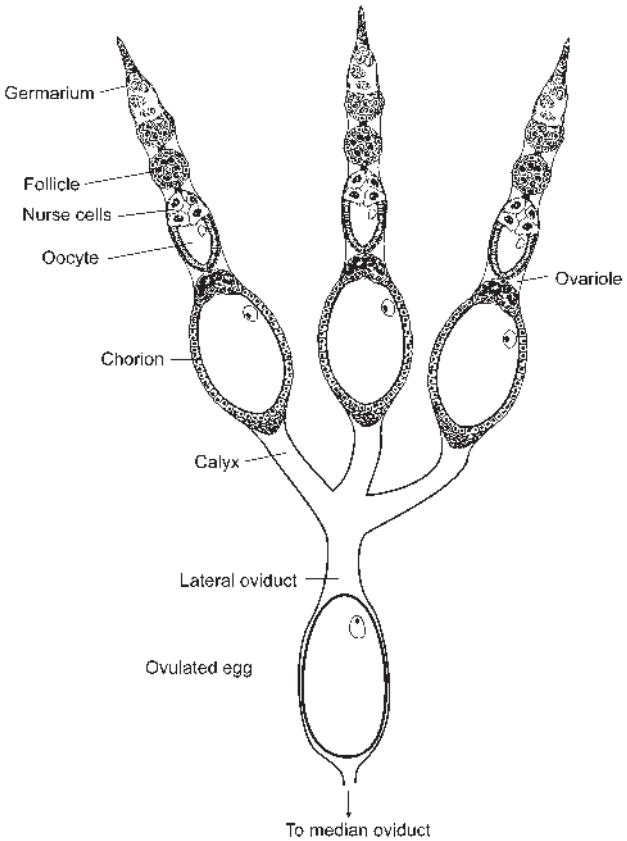

FIGURE 1 A cluster of three ovarioles that make up the larger ovary. Each has a calyx that connects to the lateral oviduct. Stem cells give rise to oocytes within the germarium, and as the oocyte descends within the ovariole, it becomes enclosed by follicle cells and deposits yolk into its cytoplasm. Upon degeneration of the follicle cells, the egg is free to move into the lateral oviduct in the process of ovulation. [Modified from Schwalm, F. E. (1988). "Insect Morphogenesis." Karger, Bosel.]

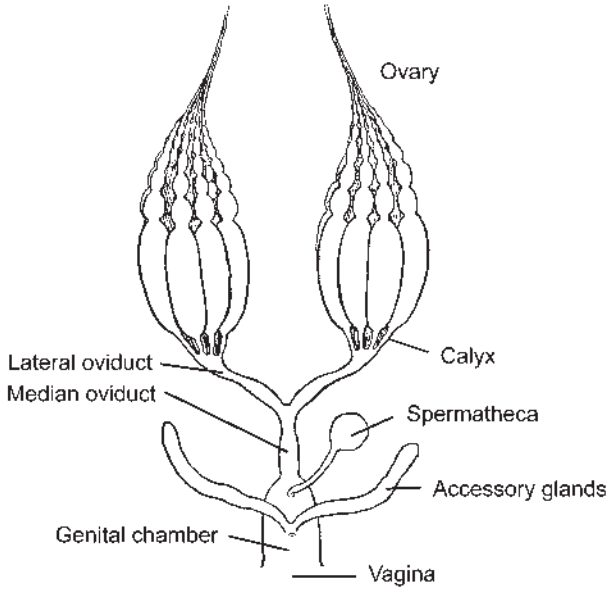

FIGURE 2 The female reproductive system, consisting of the ovaries that contain the ovarioles. As the egg moves down the median oviduct, it is fertilized by sperm released from the spermatheca. The egg is oviposited when it is deposited outside the body. [Modified from Snodgrass, R. E. (1935). "Principles of Insect Morphology." McGraw-Hill, New York.]

and into the oviduct, under the impetus of contractions of muscles in the oviduct walls. This movement of the egg to the outside of the ovariole is termed ovulation. Ovulation of the egg must occur before oviposition can take place.

The muscular contractions of the ovariole and oviduct that propel the egg through the reproductive tract are coordinated by hormones called myotropins. Myotropins are secreted by neurosecretory cells in the brain once the central nervous system has received the physiological confirmation that mating has occurred and that the eggs are mature. For example, in the bloodsucking bug *Rhodnius prolixus,* ovulation is initiated by a myotropin that is released only after the spermatheca has been filled with enough sperm and male accessory gland substances to produce a factor that induces the maturing eggs to begin producing 20-hydroxyecdysone. In the tsetse fly, *Glossina,* ovulation is initiated when a mature egg is present in the ovariole and the female has mated, but the stimulus from mating is not released until the female has received prolonged mechanical stimuli associated with copulation in addition to substances from the male accessory glands. Even several short copulatory experiences that fail to transfer sperm can initiate ovulation if their total duration is as long as a single successful copulation.

FERTILIZATION

As the egg passes through the median oviduct, it is fertilized by sperm already stored within the female's capsulelike spermatheca (Fig. 2). Sensory receptors within the oviduct are mechanically stimulated by the distension the egg creates as it descends; motor neurons activate the muscular walls of the spermathecal duct that allow the sperm to exit the spermatheca.

In hymenopterans, where sex determination occurs by haplodiploidy, the process of sperm release is controlled more directly by the female. Eggs that are fertilized develop into female workers, but those that are not fertilized develop into male drones. The workers ultimately control whether the queen fertilizes the egg; they determine the size of the cells for rearing the larvae into which the queen deposits the egg. Before the egg is laid, the queen evaluates each cell with the sensory receptors on her ovipositor, opening her spermathecal duct only when placing an egg into the larger cell for female rearing.

The sperm enter the egg through the micropyle, a modification of the normally impermeable chorion. Just prior to fertilization, the oocyte has been arrested in metaphase of the first meiotic division, but shortly after the sperm has penetrated the egg and oviposition has occured, the oocyte completes its meiosis. The meiotic division results in a haploid oocyte nucleus and three polar nuclei that inhabit the periplasm at the

periphery of the egg. The oocyte nucleus, surrounded by an island of cytoplasm, then moves to the interior of the egg, where it meets the sperm that has already entered. Syngamy, the union of sperm and egg, occurs at the interior. In some insects whose eggs develop by parthenogenesis (i.e., without the fertilization by sperm), a haploid polar nucleus combines with the haploid oocyte nucleus to restore the diploid number without requiring fertilization by male gametes.

OVIPOSITION

After ovulation and fertilization, the eggs are usually deposited outside the female's body during oviposition. The eggs move down the common oviduct by peristaltic waves of muscle contractions and out of the body through the ovipositor. The movement of the egg downward through the oviducts is facilitated by backwardly directed scales inside the oviduct, which act like a ratchet mechanism, allowing the egg to move in only one direction, down toward the genital opening.

Modified dermal glands known as female accessory glands may also be present on the common oviduct. These glands produce cement that allows the deposited eggs to be glued together or attached to the substrate. In some insects that retain their eggs after the young hatch, such as *Glossina* the accessory glands produce a nutritive secretion that nourishes the larvae during their entire larval period. In cockroaches and mantids, the female accessory glands produce the hardened egg case, or ootheca.

The spermathecae are used for the storage of sperm in the inseminated female. Also ectodermal in origin, the spermatheca generally opens into the common oviduct and releases sperm as the fully formed egg passes by. The spermathecal duct may contain glycogen deposits that can serve as an energy source for the sperm as they pass through to the egg. Nearest its opening to the outside, the common oviduct may be modified into a genital chamber that can be used to incubate eggs internally. The bursa copulatrix is an additional pouch within the chamber that is present in some insects into which sperm may first be deposited after mating. The sperm leave the bursa and then move into the spermatheca, where they are permanently stored. The bursa may have a series of toothlike structures that disrupt the spermatophore, the vessel in which the sperm are contained in more primitive insects, and facilitate their release. It may also secrete chemical signals into the hemolymph when it is filled with sperm to signal to the female that mating has occurred.

The control of oviposition in the cockroach *Spodromantis* is a good example of the integration of environmental and physiological information that must occur during egg laying. This insect normally lays its eggs at the beginning of the photophase. The brain integrates the information it receives about the photoperiod and the presence of mature oocytes and triggers the release of an oviposition-stimulating hormone that activates the ovipositor, ovariole, and oviduct muscles. As the insect probes the substrate with its ovipositor, tactile sensations from sensilla are sent to the terminal abdominal ganglion, which controls further movements of the egg and secretions by the accessory glands.

In most species, ovulation and oviposition may be part of a continuum, but in others they may be separate events in which the eggs are retained for a variable period between ovulation and oviposition. Some species of ovoviviparous cockroaches retain in a brood sac eggs that have been ovulated until the female finds a suitable place to lay them. The nymphs hatch inside the female. The viviparous tsetse retain their larvae after they have hatched from the egg and provide them during the larval stage with food that is secreted from their accessory glands. "Larviposition" occurs toward the end of the larval stage, and after oviposition the larvae pupate in the soil.

Most insects lack specialized structures for the deposition of eggs, but in some species the terminal segments of the abdomen form a telescoping ovipositor that can be used to deposit eggs into crevices or even, when the tip is modified into sawlike blades, into hardened substrates such as wood. This ovipositor may be derived from the modified appendages just described or from the modified terminal abdominal segments themselves. Special muscles within the ovipositor allow it to engage in superextension, lengthening considerably to reach hidden sites into which the eggs can be placed. Both the tips of the abdomen and the ovipositor may contain sensory receptors that can provide the female with information about the nature of the oviposition substrate. Sensory cues that are evaluated by the receptors are important in locating an oviposition site and in initiating oviposition once the site has been found.

Prior to depositing her eggs, the female must engage in behaviors that will bring her to an environment that is suitable for larval development. This is especially necessary for the eggs of holometabolous insects, since the larvae are very different from the adults. Being relatively immobile, the larvae depend on the adult to locate a site on which they can develop. For example, adult mosquitoes are terrestrial but their larvae must develop in water. Although the adults are generally attracted to stimuli from vertebrate hosts for blood, when they carry mature eggs they become more sensitive to the stimuli from potential oviposition sites. The presence of mature eggs appears to trigger a physiological switch that changes the behavior of the female, attracting her to oviposition sites suitable for her eggs. In the absence of mature eggs, the female is more attracted to a host for a blood meal that would support another batch of eggs.

Lepidopterans undergo a sequence of behaviors, including searching, orientation, encounter, landing, evaluation of the surface, and acceptance. Stimuli received during each phase are integrated by the central nervous system, which processes the sensory information from sensory receptors on the antennae, tarsi, and ovipositor. Visual cues such as plant color and shape are important during searching behavior. Volatile chemicals produced by the plant initiate the orientation and encounter behaviors that attract the female from a distance of several meters. Visual cues are once again important in landing behavior, with the involvement of chemical cues that act as attractants, arrestants, or repellents. Once on the host plant, the female

evaluates the physical and chemical cues from the plant surface and often begins drumming her front legs on the leaf surface, perhaps to provide more chemical information about the plant to tarsal sensory receptors. By using the total sensory input to the central nervous system, the female is able to recognize the suitability of the substrate for her eggs before they are laid.

The torsalo, *Dermatobia hominis,* has an unusual way of finding an oviposition site. The gravid female captures another bloodsucking fly and glues her eggs to the underside of the carrier's body. When the carrier seeks out a vertebrate host for a blood meal, the larva hatches and burrows through the skin of the host. The larva develops in the vertebrate host, exits prior to the pupal stage, and pupates, in the soil.

See Also the Following Articles

Accessory Glands • Egg Coverings • Reproduction, Female: Hormonal Control of

Further Reading

Austin, A. D. and Browning, T. O. (1981). A mechanism for movement of eggs along insect ovipositors. *Int. J. Insect Morphol. Embryol.* **10,** 93–108.

Curtin, T. J., and Jones, J. C. (1961). The mechanism of ovulation and oviposition in *Aedes aegypti. Ann. Entomol. Soc. Am.* **54,** 298–313.

Davey, K. G. (1985). The female reproductive tract. *In* "Comprehensive Insect Physiology Biochemistry, and Pharmacology (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 1, pp. 1–14. Pergamon Press, Oxford, U.K.

Hinton, H. E. (1981). "The Biology of Insect Eggs," Vol. 1. Pergamon Press, Oxford, U.K.

Mesnier, M. (1984). Patterns of laying behaviour and control of oviposition in insects: Further experiments on *Sphodromantis lineola* (Dictyoptera). *Int. J. Inverteb. Rep. Dev.* **7,** 23–32.

Mesnier, M. (1985). Origin and release sites of a hormone stimulating oviposition in the stick insect *Clitumnus extradentatus. J. Insect Physiol.* **31,** 299–306.

Ramaswamy, S. B. (1988). Host finding by moths: Sensory modalities and behaviours. *J. Insect Physiol.* **34,** 235–249.

Renwick, J. A. A., and Chew, F. S. (1994). Oviposition behavior in Lepidoptera. *Annu. Rev. Entomol.* **39,** 377–400.

Schwalm, F. E. (1988). "Insect Morphogenesis." Karger, Basel.

Snodgrass, R. E. (1935). "Principles of Insect Morphology." McGraw-Hill, New York.

Sugawara, T., and Loher, W. (1986). Oviposition behaviour of the cricket *Teleogryllus commodus:* Observation of external and internal events. *J. Insect Physiol.* **32,** 179–188.

Thompson, J. N., and Pellmyr, O. (1991). Evolution of oviposition and host preference in Lepidoptera. *Annu. Rev. Entomol.* **36,** 65–89.

P

Parasitoids

Nick Mills
University of California, Berkeley

Parasitoids are insects with a parasitic larval stage that develops by feeding on the body of a single host insect or other arthropod. Feeding by the larval parasitoid invariably results in the death of its host, and the resulting adult parasitoid is a free-living insect. Thus parasitoids occupy an intermediate position between predators and true parasites; in contrast to parasites they kill their host like a predator, but in contrast to predators they require only a single host to complete their development, as is the case for parasites.

Nearly 10% of described insect species are parasitoids and, as they often belong to poorly known groups of insects, it has been suggested that they are more likely to represent 20–25% of all insect species. Since the first accounts of insect parasitism at the beginning of the 18th century, parasitoids have attracted considerable attention because of their potential to regulate the abundance of insect hosts in both natural and managed ecosystems. Numerous insect pests have been effectively controlled through releases of insect parasitoids in biological control programs around the world. Although not always applicable or successful, biological control continues to provide some dramatic examples of sustained long-term control of invasive insect pests. As a result of the research conducted through biological control programs, parasitoids have also become important model organisms for the study of ecology, behavior, and evolution.

ORIGIN AND DIVERSITY OF PARASITISM

About 74% of the known parasitoid species belong to the parasitic Hymenoptera, an arbitrary division of the suborder Apocrita of the order Hymenoptera. It is generally believed that parasitism evolved just once in the Hymenoptera and that the Apocrita together with the sawfly family Orrusidae form a holophyletic group that includes all of the known parasitic wasps. Although the larvae of some species may initially feed on microorganisms in the tunnels of wood-boring insects, the young larvae of others feed externally and subsequently internally on larvae of the borers themselves. Many horntails and wood wasps (sawfly superfamily Siricoidea) carry symbiotic fungi in pouches located at the base of the ovipositor (mycangia) that are inoculated at oviposition, allowing the larvae to feed on infected and partially digested wood. However, not all siricoids carry fungal symbionts and some may have evolved to kill those that did and subsequently feed on the more nutritious dead insect rather than the wood. This trait is seen among the present-day orrusids.

Within the parasitic Hymenoptera there are close to 64,000 described species of parasitoids in 10 superfamilies that are, with rare exception, exclusively parasitic (Table I). However, parasitoids are also known from five other insect orders. In contrast to the Hymenoptera, parasitism has evolved repeatedly within these orders, from a fungal-feeding (e.g., Rhipiphoridae), dead-organism-feeding (e.g., Phoridae, Sarcophagidae), predatory (e.g., Carabidae, dipteran families), or phytophagous (e.g., Lepidoptera) ancestor. For example, parasitism is estimated to have occurred 21 times in the Diptera, producing nearly 15,000 described parasitoid species, and 15 times in the Coleoptera, producing a further 3400 parasitoid species (Table I). Parasitism is unusual and rare among the remaining three orders.

PARASITOID LIFESTYLES

Parasitoids are frequently categorized as having larvae that are either ectoparasitic (external feeders) or endoparasitic (internal feeders) and either solitary (one per host) or gregarious (several per host) in their development. However, a more useful categorization of the lifestyles of parasitoids is the dichotomy between idiobiosis and koinobiosis. Idiobionts paralyze and/or

TABLE I The Main Superfamilies of Hymenoptera and Families of Other Insect Orders That Contain Parasitoids, with an Indication of either the Total Described Species for Taxa That Are Exclusively Parasitic or the Number of Described Parasitoid Species for Taxa (Followed by *) That Include Many Nonparasitic Species

Order	Superfamily/family	World species
Hymenoptera	Orussoidea	75
	Trigonaloidea	70
	Evanioidea	1,050
	Cynipoidea	2,335*
	Chalcidoidea	17,500*
	Proctotrupoidea	6,135
	Ceraphronoidea	250
	Ichneumonoidea	25,000
	Chrysidoidea	5,850*
	Vespoidea	5,500*
Diptera	Cecidomyiidae	6*
	Acroceridae	475
	Bombylidae	3,000
	Nemestrinidae	300
	Phoridae	300*
	Pipunculidae	600
	Conopidae	800
	Sarcophagidae	1,250*
	Tachinidae	8,200
Coleoptera	Carabidae	470*
	Staphylinidae	500*
	Rhipiphoridae	400
	Meloidae	2,000*
	Mengeidae	10*
Lepidoptera	Pyralidae	1*
	Epipyropidae	10*
Neuroptera	Mantispidae	50*
Trichoptera		1*

Note. Data from Godfray (1994).

arrest the development of a host at oviposition, providing their larvae with an immobilized static resource on which to feed. In contrast, koinobionts allow the host to continue to feed and/or develop after oviposition, such that their larvae feed on an active host that is killed only at a later stage. Although ecto- and endoparasitism are distinct modes of parasitism, the separation of idiobiont and koinobiont lifestyles is less clear. There should be no problem in correctly categorizing parasitoids that attack an active stage of the host life cycle (larva or adult) or those that allow a transition between life cycle stages of the parasitized host (egg–prepupal and pupal–adult parasitoids). However, the distinction is not always so obvious for parasitoids that attack and complete their development in an inactive stage of the life cycle (egg or pupa).

Idiobionts

Idiobiont parasitoids frequently attack hosts that are concealed in plant tissues or exposed hosts that provide some other form of physical protection, such as the scale covering of armored scale insects. Thus, many have long ovipositors to reach their concealed hosts and strong mandibles to escape from concealed locations. The majority of idiobionts that attack concealed hosts have larvae that are ectoparasitic and feed on hosts in the later stages of their development. An ovipositing adult first injects a venom into the host, to induce temporary or permanent paralysis, and then oviposits on or near to the immobilized host. In a few cases (some Bethylidae, Braconidae, and Eulophidae in the Hymenoptera), the parent female remains with the parasitized host either to defend her young offspring against competitors or hyperparasitoids or to ensure continued paralysis of the host. Ectoparasitic idiobionts are often long lived, feeding for somatic maintenance on honeydew, nectar, or other plant exudates. They are synovigenic, meaning that they continue to mature eggs throughout adult life and produce large yolk-rich eggs by using nutrients gained from host feeding (see below). Consequently, ectoparasitic idiobionts have a relatively low rate of host attack and a low fecundity. The hatching larvae are protected from desiccation by the host concealment, but develop continuously and rapidly to consume the host before it is attacked by scavengers or microbial decay.

Endoparasitic egg and pupal parasitoids are also considered to be idiobionts, although as noted above, this categorization is less clear. In this case the hosts are frequently exposed rather than concealed, and in place of paralysis, development of the host appears to be arrested by secretions either from the ovipositing adult (egg parasitoids) or from the young larva (pupal parasitoids). Endoparasitism of host eggs is facilitated by the lack of an immune defense, but it is not known how pupal idiobionts avoid such defenses in their hosts.

The absence of an intimate association between the juvenile stages of an ectoparasitic idiobiont and its host allows this group of parasitoids to have a broader host range, using a variety of hosts that share a common habitat or host plant. Some endoparasitic idiobionts are more restricted in their host range, however, perhaps because of a need for more specific cues in host recognition or for detoxification of the chemical defenses of exposed hosts.

Koinobionts

All dipteran parasitoids, the majority of coleopteran parasitoids, and many hymenopteran parasitoids adopt a koinobiont way of life. Koinobiont parasitoids attack both exposed and concealed hosts, and the majority of species have endoparasitic larvae. Endoparasitic koinobionts attack a broad range of developmental stages of their hosts. Access to concealed hosts is facilitated either by attacking the more accessible egg or young larval stage of their host (hymenopteran koinobionts only) or by production of a free-living first instar that can complete the location of a suitable host.

Koinobionts typically oviposit into the body of their hosts with minimal disruption to the normal activity of the host. The hemocytic immune response of the host can result in the encapsulation of parasitoid eggs and larvae and is a significant obstacle to endoparasitism. This response can be overcome by (1)

avoidance, by placing eggs in host tissues such as the salivary gland or nerve ganglia; (2) evasion, by producing an egg with a fibrous coat or a coating of proteins that are not recognized as foreign to the host; (3) suppression, by injecting immunosuppressive polydnaviruses or virus-like particles at oviposition; or (4) subversion, by allowing host hemocytes to form a sheath around the developing larva that is attached to the host tracheal system to avoid asphyxiation. Koinobionts are also mostly synovigenic, but differ from idiobionts in having a much greater fecundity, with the majority of their eggs produced early in adult life. Koinobiont eggs are relatively small and enrichment occurs after oviposition through absorption of nutrients from the host hemolymph. The greater emphasis on early reproduction is often associated with a shorter adult life, although longevity is greatly influenced by access to a sugar food source.

Many koinobionts exhibit protracted larval development, remaining as a first instar within the host until the latter has retreated to its pupation site. This both allows the host larva to complete its development without suffering debilitating effects from parasitism and permits the parasitoid to remain in its most competitive stage, because first instars bear strong defensive mandibles, until the host has gained maximum size as a resource for parasitoid larval development. Thus, koinobiont parasitoids are often more specialized in their host range than idiobionts, because of their more intimate relationship with an actively feeding or reproducing life stage of the host and the need to overcome its immune defenses.

BEHAVIOR AND INTERACTIONS

Host Feeding

Many idiobiont adult females acquire nutrients from the host by feeding on the body fluids that exude from wounds inflicted by the ovipositor, a process known as host feeding. Some small gregarious parasitoids are able to host feed on the same host individual used for oviposition, but in most cases host feeding is destructive and can be responsible for substantial levels of host mortality. Destructive host feeders select smaller host individuals for host feeding, because they are unsuitable for parasitism, and then use larger hosts for oviposition.

Interactions among Parasitoids

Because the majority of host insects are attacked by several different parasitoid species, a variety of trophic and competitive interactions occur among them. Hyperparasitism is a trophic interaction that occurs when a secondary parasitoid parasitizes a primary parasitoid. Competitive interactions include superparasitism and multiparasitism. Superparasitism occurs with more than one oviposition by one or more individuals of the same parasitoid species into the same host individual. The resulting intraspecific competition between parasitoid larvae leads to the death of all but one individual, in the case of solitary parasitoids, and to male bias in the sex ratio and reduced adult size among the progeny of gregarious parasitoids. Multiparasitism is the corresponding interspecific competition that results from oviposition by two or more different parasitoid species in the same host individual. The outcome of multiparasitism is often indeterminate; it can favor the species that attacked first but it can also be fixed, with a strong competitor being the victor whatever the sequence of attack. A particularly interesting form of the latter is cleptoparasitism, in which the success of host location by a cleptoparasitoid is facilitated by its response to chemical markers used by an inferior competitor to avoid reattacking a previously parasitized host. The cleptoparasitoid is able to steal the host from its inferior competitor by having an aggressive first instar that is able to kill the original occupant of the host.

Clutch Size and Sex Ratio

One of the most important "decisions" for a gregarious parasitoid is how many eggs to lay on a particular host. Clutch size increases with the size and quality of a host, but decreases as the rate of host encounters increases and often decreases with the age of the parasitoid. Gregarious parasitoids adjust clutch size to match the quality and frequency of hosts encountered, thereby balancing the need to maximize reproductive output and to minimize intraspecific competition among the larval brood. Because hymenopteran parasitoids use haplodiploid reproduction (males develop from unfertilized haploid eggs, females from fertilized diploid eggs), parent females can choose the sex of their offspring. Solitary parasitoids tend to allocate male eggs to small or low-quality hosts and female eggs to large or high-quality hosts, but typically produce a balanced sex ratio. In contrast, female bias is frequent in the sex ratio of gregarious parasitoid broods in which the probability of sibling mating is high and sons compete within broods for mates. Thus, local mate competition within broods of gregarious hymenopteran parasitoids leads to the allocation of just enough sons to be able to mate effectively with all the daughters in the brood.

PARASITOID COMMUNITIES

Parasitoids can utilize hosts in a variety of ways to support the development of their progeny. For example, a lepidopteran host such as the gypsy moth *(Lymantria dispar)* is attacked at different stages through its life cycle by a series of 13 species in six parasitoid guilds (Fig. 1). Very few terrestrial insects escape the attention of parasitoids, exceptions being the few taxa that are either too small (e.g., Adelgidae) or too well defended (e.g., *Dactylopius* mealybugs). However, the parasitoid load, or number of parasitoid species supported by a host species, varies tremendously. For example, the average number of parasitoid species associated with an aphid is 1.7 in comparison with 12.4 for bivoltine Lepidoptera. In contrast, aquatic insects support relatively few parasitoids and are most vulnerable during the nonaquatic stages of their life cycle. It is notable, however, that a few mostly very small parasitoids

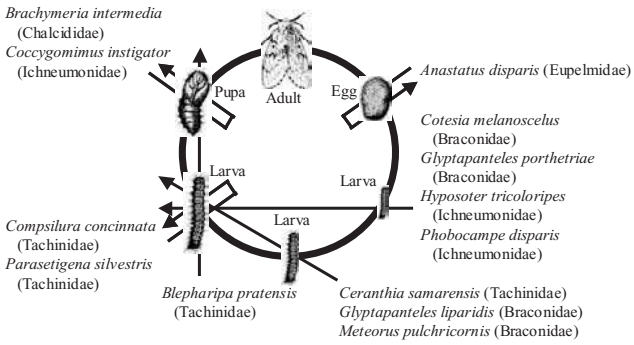

FIGURE 1 The parasitoid assemblage associated with the gypsy moth *(L. dispar)* as it passes through its life cycle in Eurasia. Arrows indicate the host stages attacked and killed by the six different guilds of parasitoid species.

are known to be truly aquatic, using their legs or wings to swim through the water to locate submerged host eggs.

Parasitoid load is determined by a combination of phylogeny, feeding niche, abundance, and chemical defense. The absence of a pupal stage greatly reduces the parasitoid load of a host, but even among the holometabolous insects, beetles consistently support far fewer parasitoids than moths, indicating the importance of host phylogeny. The feeding niche of a larval host also has an important influence on parasitoid load, being greatest among those hosts that have restricted mobility and poor protection (e.g., leafminers, casebearers) and reduced either by greater mobility (e.g., external feeders) or by greater protection (e.g., borers). The greater the abundance of a host, the greater its parasitoid load, because a number of less specialized parasitoids and even incidental species are able to make use of an abundant resource. Then, finally, the sequestration of defensive plant chemicals by externally feeding host larvae appears to offer a further line of defense that can lead to a reduction in parasitoid load.

It has frequently been suggested that parasitoid diversity declines from temperate zones to the tropics. Such a pattern is evident for the very species-rich superfamily the Ichneumonoidea (Ichneumonidae and Braconidae), but is not upheld among the Chalcidoidea. Nonetheless, the decline, or absence of an increase, in overall parasitoid diversity in the tropics is an interesting anomaly in comparison to the increase in diversity of their insect hosts in the tropics. An increased level of predation (notably by ants) in the tropics, a general reduction in the abundance of each host species due to fragmentation of resources, and a greater availability of plant-based chemical defenses for host insects to use for protection may all help to account for this anomaly.

See Also the Following Articles

Biological Control • Gypsy Moth • Host Seeking, by Parasitoids • Population Ecology

Further Reading

Eggleton, P., and Belshaw, R. (1992). Insect parasitoids: An evolutionary overview. *Philos. Trans. R. Soc. London B* **337,** 1–20.

Feener, D. H. (1997). Diptera as parasitoids. *Annu. Rev. Entomol.* **42,** 73–97.

Gauld, I. D., and Hanson, P. E. (1995). The biology of Hymenoptera. *In* "The Hymenoptera of Costa Rica" (P. E. Hanson and I. D. Gauld, eds.). Oxford University Press, Oxford, UK.

Godfray, H. C. J. (1994). "Parasitoids: Behavioral and Evolutionary Ecology." Princeton University Press, Princeton, NJ.

Gordh, G., Legner, E. F., and Caltagirone, L. E. (1999). Biology of parasitic Hymenoptera. *In* "Handbook of Biological Control" (T. S. Bellows and T. W. Fisher, eds.). Academic Press, San Diego.

Mills, N. J. (1994). Parasitoid guilds: A comparative analysis of the parasitoid communities of tortricids and weevils. *In* "Parasitoid Community Ecology" (B. A. Hawkins and W. Sheehan, eds.). Oxford University Press, Oxford, UK.

Quicke, D. L. J. (1998). "Parasitic Wasps." Chapman & Hall, London.

Strand, M. R. (1995). Immunological basis for compatibility in parasitoid–host relationships. *Annu. Rev. Entomol.* **40,** 31–56.

Whitfield, J. B. (1998). Phylogeny and evolution of host–parasitoid interactions in Hymenoptera. *Annu. Rev. Entomol.* **43,** 129–151.

Parental Care

Michelle Pellissier Scott
University of New Hampshire

Parental care in insects ranges from covering eggs with a protective coating to remaining to feed and protect young to forming eusocial societies with life-long associations of parents and offspring and alloparental care. Most commonly care is provided solely by females; males rarely care for eggs and young alone, but in some cases males and females form long associations to rear young to adulthood (Table I). Parental care has evolved independently many times, and examples do not follow along phylogenetic lines. This trait is widespread taxonomically and is most developed in Hemiptera (true bugs), Thysanoptera (thrips), Embioptera (webspinners), Coleoptera (beetles), Hymenoptera (ants, bees, and wasps), and Isoptera (termites).

FORMS OF PARENTAL CARE

Maternal Care

The most rudimentary form of maternal care is provided by females that incorporate toxins into their eggs, oviposit them in protected places, or cover their eggs with a hard shell or waxlike compound before abandoning them (Table I). For example, embiopteran webspinner females *(Antipalurai urichi)* cover their eggs with layers of macerated bark and other substrate materials and silk to protect them from hymenopteran parasites. Webspinner maternal care is more complex and extensive than egg protection, and these females construct silk galleries and remain with their eggs and nymphs.

Many species of insects guard their young against predators by using chemicals or defensive behaviors. Care

TABLE I Types of Parental Care by Insects

	Number of orders[a]		
Behavior[b]	Maternal care	Paternal care	Biparental care
EP	6	0	0
EC	12	2	3
YC	9	2	4
YP	6	0	3
OV	10		
PZ		9	

[a]Number of orders in which parents exhibit the behaviors listed.

[b]Key: EP, either or both parents cover eggs before abandonment; EC, either or both parents remain and guard eggs; YC, either or both parents remain with young and care for them; YP, either or both parents provision young or regurgitate food to them; OV, females extend development internally and are ovoviviparous, larviparous, or viviparous; PZ, males provide prezygotic investment.

may end when young hatch, or it may extend until larvae or nymphs are mature. For example eggplant lace bugs *(Gargaphia solani)* guard their eggs and gregarious nymphs until maturity; if a predator approaches, the female rushes at it, fanning her wings.

A second major function of maternal care is to facilitate feeding. A plant-feeding membracid bug, *Umbonia crassicornis,* cuts slits in the bark with her ovipositor and remains with the nymphs, actively maintaining feeding aggregations until the young reach adulthood. Parental care often comprises a suite of adaptations of multiple behaviors that serve multiple functions. A well-cited example of this behavior is the female salt-marsh beetle, *Bledius spectabalis,* which maintains a burrow shaped in a way that prevents flooding during high tide. She also provisions the young with algae, prevents mold, and protects the vulnerable first instars from attack by parasitic wasps.

In some insects, maternal investment takes the form of a period of internal development. Among insects, cockroaches carry this investment to the extreme and show the entire range of reproductive modes and maternal care. The oviparous German cockroach *Blatella germanica* carries her egg sac externally until nymphs hatch, and *B. vaga* produces maternal secretions that her neonates feed on briefly. In ovoviviparous species, eggs develop inside the body of the mother and have sufficient yolk to complete development. The viviparous cockroach *Diploptera punctata* displays a remarkable form of parental care. Females undergo a 60-day "pregnancy" during which a highly nutritious milk, secreted from the walls of the brood sac, is ingested orally by the developing young. At birth young are in an advanced state of development, and care is terminated shortly after birth.

Paternal Care

Exclusive paternal care of eggs or larvae is restricted to about 100 species of insects, almost all within the Hemiptera. For example, in a giant water bug, *Abedus herberti,* females adhere their eggs to the wingcovers of a male, which then stops feeding and instead spends his time aerating the eggs at the water surface, protecting them from predators until young emerge. Males of the subsocial spider-hunting wasp, *Trypoxylon superbum,* are an unusual example from the Hymenoptera. After the female has provisioned and sealed the cells, males remain to guard nests against parasitism and ant predation.

Indirect paternal contributions to offspring are widespread in a number of taxa. Males may invest in offspring with nutritional offerings to the female in the form of nuptial gifts of captured prey items or even their own bodies. They may transfer proteins or protective substances in a spermatophore. Male katydids are excellent examples because they provide a spermatophore during copulation that may be as much as 40% of their body mass; spermatophore nutrients have been shown to be important to the reproductive success of females. Male arctiid moths, *Utetheisa ornatrix,* provide a different sort of indirect paternal contribution when they transfer protective pyrrolizidine alkaloids to females during mating. These alkaloids are passed to the eggs, which are then unappealing to predators.

Biparental Care

Biparental care of offspring is restricted to beetles, termites, and cockroaches, and may include earwigs. It can be very elaborate and extensive. For example, the woodroach, *Cryptocercus punctulatus,* and all termites form life-long family associations. Male and female construct and guard an extensive tunnel system or a nest, and they protect and facilitate feeding of young until the offspring reach maturity. Woodroaches care for a single brood for 3 or more years, feeding them on hindgut secretions containing symbiotic fauna necessary to digest their wood diet. In many of the "higher termites," (e.g., Rhinotermitidae and the Termitidae), few or no workers or soldiers reproduce; rather, they remain as alloparents. Task specializations based on morphology and sex is strongly expressed. Primary reproductives may live 20 years or more, whereas workers and soldiers often live less than a year.

In another well-studied example, male and female burying beetles cooperate to bury and prepare small vertebrate carcasses to serve as the food source for their young. Both parents treat the carcass with preservative anal and oral secretions; both regurgitate semidigested protein to the begging larvae (Fig. 1). In the burying beetle, *Nicrophorus orbicollis,* males commonly remain in the nest until larvae are half-grown and the carcass is substantially consumed; females remain until larval development is complete and may even accompany larvae during the wandering stage. As with most species with biparental care, male and female burying beetles do not have exclusive, specialized tasks. When both parents are present, females feed larvae more often; but if the male becomes a single parent, he compensates for the loss of a mate with increased feeding rates. However, in another species with biparental

FIGURE 1 Female *N. orbicollis* regurgitates to larvae while the male, in the background, keeps the carcass free of fungi. (Photograph courtesy of Mark W. Moffett, University of California, Berkeley.)

care, the dung beetle, *Cephalodesmius armiger,* there are some task specializations. Male and female form a permanent pair-bond to rear one brood in a subterranean nest. Males forage outside for plant material that females process into brood balls into which they lay a single egg. Males continue to forage, and females enlarge the brood ball as the larvae grow.

EVOLUTION OF PARENTAL CARE

E. O. Wilson identified four ecological pressures that select for parental care in insects: (1) stable, structured environments such as wood, (2) stressful environments such as tidal salt marshes, (3) the need for an unpredictable but valuable resource such as carrion, and (4) high predation pressure. Parental care is often associated with territoriality and spatial fidelity. It is common when there is something that parents are able to do to increase the survival of young. Thus, wood is an abundant resource and supplies both food and protection, but it is difficult for immature insects to access and requires gut symbionts to use. Foliage- and sap-feeding insects may also be considered to live in a stable environment. In many of these species parents defend young and may facilitate feeding. At the other extreme, stressful environments may require parental care if young are to survive at all. However, the evolution of parental care involves a suite of changes, including greater dependence on the part of the young and the loss of adaptations for independence from parents. Many environments, both stressful and stable, harbor parental and nonparental members of a family.

Valuable but unpredictable resources (carrion or dung) or resources that are hard to acquire (a tunnel system) are especially potent selective forces for parental care. Both involve the construction of elaborate nests that represent a substantial investment. These resources are associated with the production of fewer and smaller clutches of larger young. Burying beetles, dung beetles, and the woodroach produce on average only a single brood in a lifetime. Biparental care is associated with these species both because male assistance in guarding and provisioning can often greatly increase the survival of the young and because there are few additional mating/breeding opportunities. The improved survival of his young offsets the male parent's potential gain from leaving and searching for rare carrion or females when most individuals in the population are mated.

Exclusive paternal care is rare in insects because external fertilization is rare. Internal fertilization both reduces the certainty of paternity if females have multiple mates and disassociates the father from his offspring when they appear. These factors discourage paternal care and encourage maternal care in other taxa. Exclusive paternal care is associated with fidelity to a nest site, including a "nest" of his own wingcovers, and with the ability of the male to guard the clutch of multiple females. Under these conditions, a male does not forgo additional matings; in fact, the demonstration of paternal behavior may increase his attractiveness to females.

Thus, insects not only show a variety of forms of parental care but also many demonstrate behavioral plasticity that allows them to adjust the level or form of care to changing circumstances, such as the loss of a mate. As more examples come to light and familiar ones are better studied, it becomes clear that the functions of care are also varied and complex even within a single species. This variety is not surprising, since parental care has evolved independently many times in insects in response to different selective forces.

See Also the Following Articles

Defensive Behavior • Division of Labor in Social Insects • Dung Beetls • Egg Coverings • Mating Behaviors • Nest Building • Spermatophore

Further Reading

Choe, J. C., and Crespi, B. J. (1997). "Social Behavior in Insects and Arachnids." Cambridge University Press, Cambridge, U.K.

Clutton-Brock, T. H. (1991). "Parental Care." Princeton University Press, Princeton, NJ.

Scott, M. P. (1998). The ecology and behavior of burying beetles. *Annu. Rev. Entomol.* **43,** 595–618.

Tallamy, D. W., and Wood, T. K. (1986). Convergent patterns in subsocial insects. *Annu. Rev. Entomol.* **31,** 369–390.

Trumbo, S. T. (1996). Parental care in invertebrates. *In* "Parental Care: Evolution, Mechanisms, and Adaptive Significance" (J. S. Rosenblatt and C. T. Snowdon, eds.), pp. 3–51. Academic Press, San Diego.

Wilson, E. O. (1971). "Insect Societies." Belknap Press, Cambridge, MA.

Zeh, D. W., and Smith, R. L. (1985). Paternal investment by terrestrial arthropods. *Am. Zool.* **25,** 785–805.

Parthenogenesis

Lawrence R. Kirkendall
University of Bergen, Norway

Benjamin Normark
University of Massachusetts

Most insects, like most other animals and plants, reproduce sexually. Each gamete (egg or sperm) contains one complete set of chromosomes, and the fusion of a sperm and an egg results in a zygote, which then develops into a new individual. In some insects, however (as in many other groups), offspring can be produced in a way that circumvents mating. Parthenogenesis (*partheno,* virgin, + *genesis,* to give birth, from *gen,* to be produced) is the development of an egg cell into a new individual without fertilization.

Insects have always played a central role in our understanding of parthenogenesis; Charles Bonnet first demonstrated the occurrence of this phenomenon by careful experiments with isolated females of three species of aphids, in the 1740s. During the next hundred years, further experiments with aphids and with bagworm moths, drone bees, and silkworm moths verified the reality of virgin birth, though the term "parthenogenesis" did not come into common usage until the 1850s. Parthenogenesis has fascinated biologists since it was first described, in part because it involves doing away with mating and with males and in part because clonal reproduction challenges cherished assumptions about the necessity of genetic variability in nature.

Over the past century, we have come to realize that there are numerous reproductive systems in insects that deviate from outbreeding sexual reproduction. Modern discussions of parthenogenesis adhere less strictly to the criterion of fertilization (or lack thereof), focusing instead on whether genomes are passed on intact. Whereas sexually produced offspring may be highly variable and have unique combinations of genes, parthenogenetically produced offspring typically have a genotype identical to that of their mother or only slightly different from hers.

Parthenogenesis has dramatic consequences for individuals, populations, and species. It allows females to: (1) pass along their successful genotypes to all of their offspring; (2) produce only daughters, maximizing the rate of increase; and (3) eliminate the need for finding (or being found by) a mate. Sexual reproduction, in contrast, results in: (1) offspring being different from each other and from their mothers; (2) production of 50% males, which cannot themselves produce offspring; and (3) an inability to reproduce without males being locally present and without diverting a certain amount of time and energy to the mating process.

Specialized Terms

apomixis Parthenogenesis in which eggs are produced without meiosis.

automixis Parthenogenesis in which eggs undergo meiosis.

diploid Having two complete sets of chromosomes (like a typical adult animal).

haplodiploidy (arrhenotoky) A genetic system in which unfertilized, haploid eggs develop into males and fertilized, diploid eggs develop into females.

haploid Having one complete set of chromosomes (like a typical egg or sperm cell).

parthenogenesis Reproduction without fertilization (in the sense of fusion of sperm and egg nuclei).

polyploid Having more than two complete sets of chromosomes (e.g., triploid, having three sets, tetraploid, having four sets).

pseudogamy A form of sperm-dependent parthenogenesis in which eggs require activation by entry of sperm, but only maternal chromosomes are expressed and passed on.

thelytoky Parthenogenesis in which only female offspring are produced.

tychoparthenogenesis The rare or occasional production of eggs that start developing without having been fertilized.

FORMS OF PARTHENOGENETIC REPRODUCTION

In most species of insects, every individual develops from a diploid zygote formed by the union of haploid egg with haploid sperm. But there are many alternative life cycles in which some or all individuals develop from cells that are not zygotes in this sense, and most of these life cycles come under the heading of "parthenogenesis." Thus, insect parthenogenesis encompasses much of the diversity of insect life cycles.

Sex of Offspring

Eggs laid by a virgin female may develop into other females (thelytoky), into individuals of both sexes (deuterotoky), or into males (arrhenotoky). Arrhenotoky and deuterotoky are only partially parthenogenetic systems, because males still occur. Only thelytoky is potentially a completely parthenogenetic system, and the term "parthenogenesis" is often used to refer specifically to thelytoky.

Facultative or obligate thelytoky occurs sporadically but is found in over 80 families of the superclass Hexapoda and is also scattered throughout the mites (Acari; we use mites as a collective term for all acarines, including ticks). Generally, thelytoky occurs as scattered instances in hexapods, though there are several families of mites (in the suborder Oribatida) that are strictly thelytokous.

Thelytoky is found in most orders of hexapods but the orders with the highest frequency of strictly thelytokous species are Thysanoptera, Psocoptera, Hemiptera (*sensu lato,* with thelytoky concentrated in the "homopterous" suborder Sternorrhyncha), and Phasmatodea. The largest insect orders (Lepidoptera, Diptera, Coleoptera, Hymenoptera) have a low incidence of thelytoky overall, but very high rates of thelytoky in some families, such as weevils (Coleoptera: Curculionidae), bagworm moths (Lepidoptera: Psychidae), and chironomid midges (Diptera: Chironomidae) (Diptera). Thelytoky has never been reported from several species-poor orders (Protura, Diplura, Zoraptera, Grylloblattodea, Megaloptera, Raphidioptera, Mecoptera), from the strictly parasitic orders (Siphonaptera, Phthiraptera), or from Plecoptera, and there have been only dubious reports of thelytoky in Demaptera, Neuroptera, and Strepsiptera.

Deuterotoky differs from thelytoky only in that males as well as females are produced from unfertilized eggs. There are a few insect life cycles in which deuterotoky is a normal feature: cynipid wasps, micromalthid beetles, and a few cecidomyiid flies. Reports of deuterotoky in several groups of mites have been proposed, but none have yet been confirmed. However, the boundary between thelytoky and deuterotoky is sometimes unclear; intensive study of parasitic wasp species thought to reproduce by thelytoky often turns up occasional males, which may or may not be able to mate. The offspring of matings with such males are thelytokous females. It is unclear whether an otherwise thelytokous species with rare males (which may or not be sexually competent) should be classified as deuterotokous.

In arrhenotoky, fertilized eggs develop into females while unfertilized eggs become males. Males are haploid (the only known exception occurs in the scale insect *Lecanium putmani*) and the females diploid; this is why arrhenotoky is commonly referred to as "haplodiploidy." In two orders (Hymenoptera and Thysanoptera), haplodiploidy is the only known sexual system. Haplodiploidy also occurs in some scale insects (Hemiptera: Margarodidae), whiteflies (Homoptera: Aleurodidae), some bark beetles (Curculionidae), and the bizarre beetles of the monotypic family Micromalthidae.

Arrhenotokous males arise parthenogenetically; indeed, the production of drones by queen bees prevented from mating was one of the earliest experimental demonstrations of parthenogenesis (by Dzierzon, in the 1840s). However, a haplodiploid species as a whole reproduces sexually, and the genotype of a parthenogenetically produced male is a unique recombinant product of meiosis. Thus, although parthenogenesis occurs in a haplodiploid species, clonal reproduction does not. Because males must always mate to produce offspring, under arrhenotoky there are never two successive parthenogenetic generations. Genetically equivalent to haplodiploidy is the production of male offspring in lineages that engage in paternal genome elimination (also known as pseudoarrhenotoky), in which the paternal genome in males is genetically inert and is ultimately eliminated entirely, such that males pass on only their maternally inherited genes. Pseudoarrhenotoky occurs in some mites (Acari), in most scale insects (Hemiptera: Coccoidea), in fungus gnats (Diptera: Sciaridae and Cecidomyiidae), and in one group of bark beetles (Curculionidae, Scolytinae, *Hypothenemus*). Because arrhenotoky and pseudoarrhenotoky are essentially obligately sexual systems of reproduction that have little in common with the departures from sexuality represented by thelytokous systems, this entry focuses primarily on thelytoky.

Occurrence of Meiosis

The most fundamental feature of the canonical eukaryote life cycle is the alternation between meiosis, in which the DNA content of a diploid nucleus is halved, and syngamy, in which two haploid nuclei (usually, from a sperm and an egg) fuse to form a new diploid nucleus. This alteration is commonly called "the sexual cycle," or simply "sex." But since the word sex has even more different meanings than the word parthenogenesis, it is less confusing to use the technical term mixis to refer to this cycle. In many insects, the ancestral mictic cycle has been replaced by apomixis, in which there is neither meiosis nor syngamy; instead, new individuals arise from mitotically produced cells that are genetically identical (except for new mutations) to the parent that produced them. "Vegetative" reproduction in insects is in the form of embryos undergoing fission, a process known as polyembryony. As many as 2500 individuals can develop in this way from a single egg. Polyembryony in insects occurs regularly in parasitoids in a few families of Hymenoptera and in Strepsiptera and has been reported to occur sporadically in grasshoppers. Polyembryonically produced insects will be genetically identical to each other, but not to their parents.

Apomixis is a form of parthenogenesis, but the term parthenogenesis encompasses mictic cycles as well, in which an alteration of gamete production or of the early stages of embryo development allows the production of offspring with the same ploidy as their parents; if syngamy occurs, it is a fusion of two maternally derived nuclei. Thus, in automixis, a normal meiosis occurs in the course of oogenesis but the resulting offspring have only one parent; the contrasting term is amphimixis, in which gametes from two different parents join in syngamy. Automixis itself includes a diverse group of systems, which vary as to how and when diploidy is restored; the exact mechanism of diploidization determines whether homozygosity is enforced in every generation or heterozygosity is maintained for at least some time (Fig. 1).

In premeiotic doubling, chromosomes are replicated prior to meiosis, so that the reduction–division process that is meiosis results in a diploid egg with the same chromosome complement as in the parent. *Warramaba virgo* (an acridid grasshopper) and the pseudogamous triploid ptinid beetle *Ptinus clavipes* are the only known insect examples. Maternal ploidy (in the following examples, diploidy) can also be restored by several postmeiotic means, all of which result in instant homozygosity at all loci.

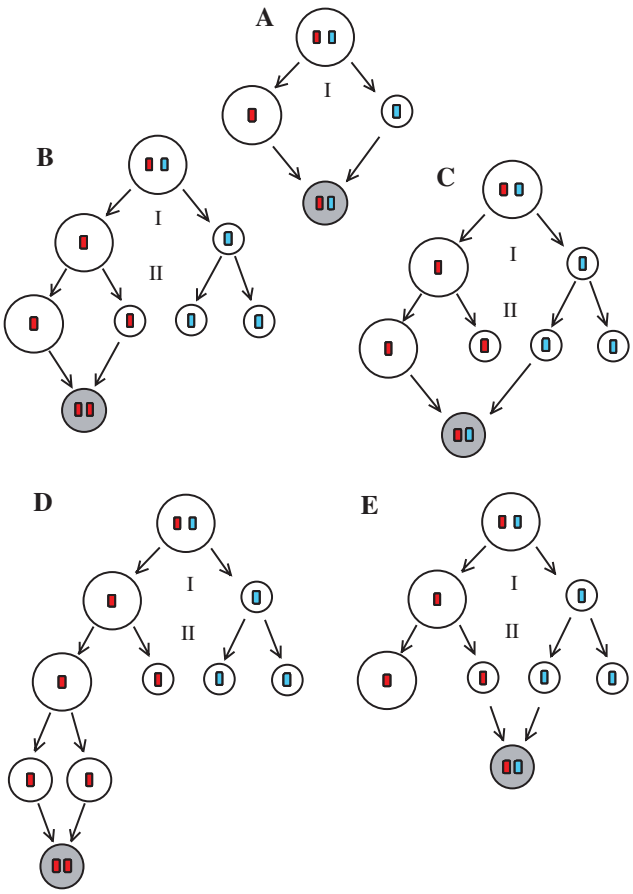

FIGURE 1 The meiotic process in five common forms of automictic parthenogenesis. One set of homologous chromosomes is followed up to the point of the restoration of maternal ploidy (the shaded circle), for each form; "I" indicates the first meiotic division, "II" the second. (A) The two products of the first meiotic division fuse. (B) The second polar nucleus fuses with the egg nucleus, an automictic process called "terminal fusion." (C) A derivative of the first polar body fuses with the egg nucleus. (D) The products of a division of the egg nucleus fuse, "gamete duplication." (E) The two central polar nuclei fuse, "central fusion." Note that in A, C, and E, the zygote has the same chromosome constitution as the mother; heterozygosity is maintained indefinitely in A, but C and E allow heterozygosity to gradually decay through crossing-over. In B and D, the zygote consists of duplicated chromosomes, producing instant homozygosity. (Modified from Suomalainen *et al.,* 1987.)

In a process sometimes referred to as gamete duplication, the haploid egg nucleus replicates but the two products fuse. Examples are found in the Phasmatodea, Hemiptera (one aleurodid, various coccids), Lepidoptera (occasional in silkworm moths), Diptera (three *Drosophila* species), and one Hymenoptera (a cynipid wasp). In terminal fusion, syngamy occurs between genetically identical cells, the egg nucleus, and the second polar nucleus. Examples are found in Orthoptera (various tychoparthenogenetic tettigid grasshoppers), Homoptera (four coccid species from three different genera), Thysanoptera (one *Heliothrips* species), Diptera (the same three *Drosophila* species as above, one or two *Lonchoptera* species), and Hymenoptera (two tenthredinids, one aphelinid). Central fusion represents a postmeiotic diploidization process in which all maternal chromosomes (and hence heterozygosity) are retained. It is characterized by the fusion of the two central polar nuclei, and examples include Lepidoptera (*Solenobia triquetella,* a well-studied psychid moth), Diptera (the parthenogenetic *Drosophila* spp. and *Lonchoptera* spp.), and Hymenoptera (an ichneumonid wasp *Venturia canescens* and occasionally in the cape honey bee *Apis mellifera capensis*). A final postmeiotic process is the fusion of the egg nucleus with a product of the first polar nucleus, again restoring the original chromosome complement, a process that occurs in a collembolan and in Lepidoptera (four psychid moth species from three genera).

Although the above-mentioned forms of automixis are considered to be parthenogenetic in the traditional sense, an exception is self-fertilization by an individual that produces both eggs and sperm (a hermaphrodite). Hermaphroditism in insects has been found only in three species of *Icerya* scale insects, all of which are known to self-fertilize.

Obligate, Cyclical, and Facultative Parthenogenesis

Modes of reproduction can be thought of as arrayed along a continuum, in which the faithfulness of mother-to-egg reproduction of a chromosome or set of chromosomes varies from about 50% (outbreeding sexuality) to about 100% (obligate apomictic parthenogenesis). Many of the intermediate modes of reproduction fall under the (again heterogeneous) categories of cyclic or facultative parthenogenesis. Cyclical parthenogenesis involves an alternation of one generation of sexual reproduction with one or more generations of parthenogenetic reproduction. The term facultative parthenogenesis potentially covers a great many possible life cycles, but it implies that a given individual can reproduce either sexually or asexually. Within the category of facultative parthenogenesis, an important subcategory is tychoparthenogenesis, in which eggs are typically fertilized, but if a clutch of eggs is left unfertilized for sufficiently long, a small proportion of them will begin development. The proportion of eggs that develop successfully to adults is usually extremely low; for example, for caught-in-the-wild *Drosophila mercatorum,* the chance of a female developing from an unfertilized egg who is herself capable of parthenogenesis is 1/10,000. Rare overall in insects, mites, and ticks, this capability is rather common in some polyneopteran orders (in orthopterans, cockroaches, stick insects, and mantids) and also occurs in psocopterans, lepidopterans, and dipterans. Where the cytology is known, in all cases oogenesis is automictic. Tychoparthenogenesis provides evidence for genetic variability for parthenogenesis within normally amphimictic species. Artificial selection experiments with drosophilids and stick insects have been able to increase by up to a thousandfold the proportion of eggs that develop parthenogenetically.

The categories of cyclical and facultative parthenogenesis intergrade when the number of parthenogenetic generations intervening between sexual ones is variable, as in certain

cecidomyiid midges, all aphids, and micromalthid beetles, all of which switch from parthenogenesis to sexuality in response to environmental cues. However, there is never more than one sexual generation before the switch back to parthenogenesis occurs. Only cynipid gall wasps are regularly cyclical in the sense that there are never two successive parthenogenetic generations, although in cynipids, as in every other group with cyclic or facultative parthenogenesis, obligately parthenogenetic lineages have arisen multiple times.

Occurrence of Mating

Thelytoky is characterized by absence of mating; indeed, the first indication that an insect species might be thelytokous is the observation that no males exist in collections. In pseudogamy (also called gynogenesis), mating with males of the same or a different species is necessary; egg development is activated by contact with or penetration by sperm. After entry into the egg, paternal chromosomes degenerate, and only maternal chromosomes survive in the offspring. This process can be considered pseudofertilization because offspring develop clonally, and all will be female, and is thus genetically a form of parthenogenesis. Pseudogamy is difficult to observe, since mating does occur; it is usually discovered by noting that mated females regularly produce all-female broods. In insects, pseudogamy is found only in Coleoptera (two origins in *Ips* bark beetles and one species of ptinid beetle), two species of Lepidoptera, and one species each of Collembola, Orthoptera, and Homoptera. Possible cases of pseudogamy in acarines have not been confirmed.

In hybridogenesis, the paternal genome is excluded during oogenesis such that ova have only maternally derived chromosomes. The intact maternal genome is paired in each new generation with sperm from males of a sexually reproducing host species. Such hemiclonal reproduction was first discovered in vertebrates and to date is known in only one group of insects, stick insects of the genus *Bacillus.* Although classed along with pseudogamy as "sperm-dependent parthenogenesis," it is actually a hybrid form of reproduction; it combines sexual reproduction, with respect to spermatogenesis, with nonrecombinant, clonal maternal genomes.

Finally, regular inbreeding between close relatives (such as self-fertilization or brother–sister mating) has genetic consequences virtually identical to those of some forms of automictic parthenogenesis: extreme inbreeding should result in homozygous clones. Because of this, some evolutionary biologists include extreme inbreeding under a broader definition of parthenogenesis, despite the involvement of mating. The similarity to automictic parthenogenesis is enhanced by the fact that species with regular sibling mating usually produce only one or a few males per brood, such that populations are nearly all-female. The taxonomic distribution of extreme inbreeding is not precisely known; it has evolved frequently in bark beetles (Scolytinae), occurs sporadically in Hymenoptera (in which it has been intensively studied in parasitic wasps, but also is found in wasps, bees, and a few ants), eusocial Thysanoptera, and mites.

ORIGINS AND GENETICS OF PARTHENOGENESIS

Relatively little is known of the details of how parthenogens arise or of the genetic changes necessary to subvert the mictic cycle. Parthenogenesis seems often to have resulted from the genetic disturbances that accompany the intrusion of foreign chromosomes; all vertebrate parthenogens appear to have arisen from interspecific hybridization, as shown by studies of chromosomes, protein variation, and DNA sequences. A hybrid origin for thelytoky has also been shown for some of the best studied cases in insects, *Warramaba* grasshoppers and *Bacillus* stick insects (nonhybrid thelytokes also occur, in the latter). Experimental hybridization between presumed parental species has generated parthenogenetic forms indistinguishable from naturally occurring ones, in apomictic triploid pseudogamous *Muellerianella* delphacid leafhoppers. Interspecific hybridization in *Oncopeltus* milkweed bugs also has produced pseudogamy.

However, at least three lines of evidence argue that parthenogenesis can evolve without hybridization, in at least some groups. In species with facultative parthenogenesis (such as tychoparthenogenesis), parthenogenesis clearly arises without interspecific hybridization. Also, clonal lineages that are genetically and morphologically similar or nearly identical to known sexual species have presumably originated without interspecific matings (e.g., thelytokous "races" of aphids and cynipid wasps that occur in otherwise cyclically parthenogenetic species). Finally, some species with thelytokous races are taxonomically isolated; *Bromius obscurus* exists as diploid bisexuals in North America but as triploid apomicts in Europe and is the only species in its genus.

There is a third way in which parthenogens arise. Recently, the intracellular parasitic proteobacterium *Wolbachia pipientis* has been shown to induce automictic thelytoky in various wasps, certain thrips, and some mites, all of which are haplodiploid. *Wolbachia* causes gamete duplication in unfertilized eggs, leading them to develop as diploid females rather than haploid males. Since *Wolbachia* is not universally present in these haplodiploid groups, thelytoky has evolved both with and without the help of these microorganisms.

Intensive cytological and genetic investigations frequently uncover the presence of multiple mechanisms for parthenogenetic reproduction in a species or group of species. The wingless stick insect genus *Bacillus* provides an excellent example. Endemic to the Mediterranean region, the genus includes two sexual species, *B. rossius* (which also has thelytokous females) and *B. grandii* (strictly bisexual), and a thelytokous lineage known as *B. atticus.* Where two or three species occur in sympatry (as on the island of Sicily), hybridization occurs, which has resulted in the production of two hybridogens, the diploid automictic *B. whitei* and the trihybrid apomictic *B. lynceorum.* Each *Bacillus* hybrid uses a different egg maturation

process; however, they share the common cytological feature of an intrameiotic extra doubling of DNA, resulting in four-stranded chromosomes, which enables the meiotic process to produce balanced chromosome complements in gametes.

Examples of multiple mechanisms within single species are provided by the extraordinary life cycles of *Micromalthus debilis* beetles and cecidomyiid midges *Heteropeza pygmaea, Miastor metralaos,* and *Mycophila speyeri* (all living in dead wood). In each of these species there are four different kinds of reproductive females: an adult female; a thelytokous paedogenetic larva (i.e., a reproductively mature larva), an arrhenotokous paedogenetic larva, and a deuterotokous paedogenetic larva.

A few studies have revealed something of the genetic basis for parthenogenesis, and it is likely that it will prove as varied as are the mechanisms of parthenogenesis. In *Rhopalosiphum padi* (the bird cherry-oat aphid), obligate parthenogenesis appears to be determined by a single locus and is recessive to cyclical parthenogenesis. In contrast, the predisposition for parthenogenesis in *D. mercatorum* was induced by genes at a number of independent loci. Some lineages of *R. padi* reproduce largely by parthenogenesis but do produce some males (although no sexual females). There is some evidence that in several cases genes carried by these males have "converted" cyclically parthenogenetic lineages of *R. padi* into obligate parthenogens.

CONSEQUENCES OF PARTHENOGENESIS: GENETIC VARIATION

Parthenogenetic insect lineages usually consist of a variety of genetically distinct clones, each of which was derived independently from the ancestral sexual populations. Clones may vary in ploidy and in their genetic composition, and as a result parthenogenetic populations can exhibit considerable diversity in morphology, behavior, and life history. Although much less than the variability within an outbreeding amphimictic population (in which each individual is genetically unique), the clonal polymorphism of thelytokous forms implies a potential for adaptation to a wider range of ecological conditions than would be possible for an invariant population.

PATTERNS IN PARTHENOGENESIS: BIOGEOGRAPHY AND ECOLOGY

Many parthenogenetic hexapods and mites, whether mictic or amictic, are common and abundant. In many cases, clonal forms are more widespread than their closest sexual relatives. Frequently, parthenogens have geographic distributions different from those of the sexual taxa to which they are most closely related, a phenomenon known as geographic parthenogenesis. Geographic parthenogenesis in hexapods (but not acarines) often takes the form of parthenogens being closer to the poles (a latitudinal gradient) and at higher altitudes. Thus, the bisexual forms of several European weevils (e.g., *Otiorhynchus scaber, Polydrosus mollis*) and *Solenobia* bagworm moths are found only in a few alpine sites thought to have been glacial refugia, while thelytokous forms of the same species are widespread in central and northern Europe. Psocoptera species with both thelytokous and bisexual forms follow a similar pattern in North America.

A number of ecological patterns have been elucidated, to describe the distribution of parthenogenesis in hexapods and mites. For example, parthenogenesis seems to be associated with low dispersal capabilities (with winglessness in Phasmatodea, Orthoptera, and Lepidoptera) and disturbed or ephemeral habitats (parthenogens often being categorized as "weedy"). Among mites, clusters of thelytokous taxa (including entire families and genera of "Endeostigmatida," Mesostigmatida, Prostigmatida, and Oribatida) are strongly associated with soil dwelling (particularly with stable soil horizons), and thelytokous (vs nonthelytokous) oribatids are strongly overrepresented on oceanic islands. In Collembola, too, there is a strong association between soil dwelling and thelytoky, and the only cockroach with thelytokous races *(Pycnoscelis indicus)* is a burrower.

CONCLUSION

Parthenogenesis encompasses a variety of reproductive systems and is often considered synonymous with "clonal reproduction." Indeed, the central feature of thelytokous parthenogenetic reproduction is that maternal genomes are normally passed on intact through a series of all-female broods. It is important to emphasize, however, that there are forms of automictic parthenogenesis in which recombination is possible and that in pseudogamous parthenogenesis, mating is necessary even though reproduction is essentially clonal.

Parthenogenetic reproduction requires a mechanism to circumvent the normal halving of ploidy that results from gametogenesis. In insects, many mechanisms for the preservation or restoration of diploidy have evolved. Either meiosis is eliminated (apomixis) or diploidy is restored (automixis) during or after meiosis. Apomixis and some forms of automixis result in maintenance of heterozygosity, whereas other forms of automixis result in instant homozygosity. Far from being a reproductive curiosity, parthenogenesis has arisen in most insect groups, and many parthenogenetic species are both abundant and widespread. Thelytokous and pseudogamous taxa are so ecologically successful that we cannot simply view them as reproductive aberrations. However, because parthenogenesis leads instantly, or relatively quickly, to lineages of genetically identical individuals, parthenogens should be depauparate in genetic variation relative to comparable sexual forms. Given the success of clonal forms, we cannot always assume that genetic variability is vitally important in nature. We are encouraged, then, to search for patterns in the occurrence of parthenogenesis that might explain when and why it is successful.

The study of parthenogenesis can illuminate one of the central problems in biology, that of explaining the ubiquity

of sex and recombination, by revealing when and where genetic variability is important in nature. Insects and mites, because of their short life cycles and often large population numbers, are ideal organisms for studying the evolutionary and ecological consequences of parthenogenesis.

See Also the Following Articles

Genetic Variation • Polyembryony • Sex Determination • Wolbachia

Further Reading

Bell, G. (1982). "The Masterpiece of Nature: The Evolution and Genetics of Sexuality." Croom Helm, London.

Beukeboom, L. W., and Vrijenhoek, R. C. (1998). Evolutionary genetics and ecology of sperm-dependent pathenogenesis. *J. Evol. Biol.* **11,** 755–782.

Normark, B. B. (2003). Evolution of alternative genetic systems in insects. *Annu. Rev. Entomol.* **48,** 397–423.

Parker, E. D., Jr. (2000). Geographic parthenogenesis in terrestrial invertebrates: Generalist or specialist clones? *In* "Progress in Asexual Propagation and Reproductive Strategies" (R. N. Hughes, ed.). Oxford–IBH, Oxford.

Suomalainen, E., Saura, A., and Lokki, J. (1987). "Cytology and Evolution in Parthenogenesis." CRC Press, Boca Raton, FL.

Wrensch, D. L., and Ebbert, M. A. (eds.) (1993). "Evolution and Diversity of Sex Ratio in Insects and Mites." Chapman & Hall, New York.

Pathogens of Insects

Brian A. Federici
University of California, Riverside

Pathogens are viruses or microorganisms that cause disease. Like all other organisms, insects are susceptible to a variety of diseases caused by pathogens. Many of these pathogens cause diseases that are acute and fatal and therefore are used as models to study processes of infection and pathogenesis as well as to control populations of insects that are pests or vectors of plant and animal diseases. Generally, insect pathogens have a relatively narrow host range and thus are considered to be more environmentally friendly than synthetic chemical insecticides. The pathogens that cause disease in insects fall into four main groups: viruses, bacteria, fungi, and protozoa. This article discusses the primary biological properties of each of these pathogen groups, with specific emphasis on how these pathogens have been used to benefit humans.

VIRUSES

Viruses are obligate intracellular parasites, meaning that they can reproduce only in living cells and are composed in the simplest form of a nucleic acid, either DNA or RNA, and a protein shell referred to as the capsid. More complex viruses also contain a lipoprotein envelope. Insect viruses can be cultured in living hosts (i.e., *in vivo*) or in cultured insect cells *(in vitro).* In general, insect viruses are divided into two broad nontaxonomic categories, the occluded viruses and the nonoccluded viruses. Occluded viruses are so named because after formation in infected cells, the mature virus particles (virions) are occluded within a protein matrix, forming paracrystalline bodies that are generically referred to as either inclusion or occlusion bodies. In the nonoccluded viruses, the virions occur freely or occasionally form paracrystalline arrays of virions that are also known as inclusion bodies. These, however, have no occlusion body protein interspersed among the virions.

The five most commonly encountered types of insect viruses are iridoviruses, cytoplasmic polyhedrosis viruses, entomopoxviruses, ascoviruses, and baculoviruses.

Iridoviruses

Nonoccluded viruses with a linear double-stranded DNA genome, the iridoviruses (family Iridovirdae) produce large, enveloped, icosahedral virions (125–200 nm) that replicate in the cytoplasm of a wide range of tissues in infected hosts. Virions form paracrystalline arrays in infected tissues, imparting an iridescent hue to infected hosts, from which the name of this virus group is derived. Over 30 types are known, and these have been most commonly reported from larval stages of Diptera larvae, such as mosquito larvae, as well as from larvae of Coleoptera and Lepidoptera. Generally, the iridoviruses occur very broadly, and they are known from other invertebrates, such as isopods, as well as from certain vertebrates including frogs and fish. Observations of natural occurrence in host field populations suggest that one host range of each type is quite narrow, although in the laboratory iridoviruses are easy to transmit from one insect species to another by inoculation. Prevalence and mortality rates in natural populations of host insects are typically less than 1%.

Cytoplasmic Polyhedrosis Viruses

The cytoplasmic polyhedrosis viruses (family Reoviridae) are occluded double-stranded RNA viruses with a genome divided into 9 or 10 segments of RNA. These viruses, commonly referred to as CPVs, cause a chronic disease and reproduce only in the stomach of insects, where typically they form large (ca. 0.5–2 μm) polyhedral to spherical occlusion bodies in the cytoplasm of midgut epithelial cells. Infection in early instars retards growth and development, extending the larval phase by weeks. The disease is often fatal. In advanced stages of disease, the infected midgut is white rather than translucent brown because large numbers of polyhedra have accumulated there. This virus type is relatively common among lepidopterous insects and among dipterous insects of the suborder Nematocera (e.g., mosquitoes, blackflies, midges). CPVs are typically easy to transmit by feeding to species that belong to the same family

of the host from which they were isolated, and thus the host range of this virus type quite broad.

Entomopoxviruses

The entomopoxviruses (family Poxviridae) are occluded double-stranded DNA viruses that produce large, enveloped virions (150 nm × 300 nm) that replicate in the cytoplasm of a wide range of tissues in most hosts, causing an acute, fatal disease. Occlusion bodies vary from being oval to spindle shaped and generally occlude 100 or more virions. These viruses have been most commonly reported from coleopterans, from which there are over 30 isolates, but they are also known from lepidopterous, dipterous (midges), and orthopterous (grasshoppers) insects. This virus type is easily transmitted by feeding, although where the experimental host range of individual isolates has been tested, it has been found to be relatively narrow, generally being restricted to closely related species. Insect poxviruses are related to vertebrate poxviruses, such as the variola virus, the etiological agent of smallpox, and they may be the evolutionary source of the vertebrate poxviruses.

Ascoviruses

The ascoviruses are a new family of DNA viruses (family Ascoviridae) at present known only from larvae of species in the lepidopteran family Noctuidae, where they have been reported from several common pest species such as the cabbage looper, cotton budworm, corn earworm, and fall armyworm. Ascoviruses cause a chronic, fatal disease of larvae. The virions of ascoviruses are large (130 nm × 400 nm), enveloped, and reniform to bacilliform in shape; they exhibit complex symmetry and contain a linear, double-stranded DNA genome. During the course of ascovirus disease, large numbers of virion-containing vesicles accumulate in the blood of infected caterpillars, changing its color from translucent green to milky white. These virion-containing vesicles are formed by a unique developmental sequence in which each infected host cell cleaves into a cluster of vesicles as virion assembly proceeds.

An interesting ascovirus feature is that transmission from host to host depends on vectoring by female endoparasitic wasps. Ascoviruses are extremely difficult to transmit by feeding, with typical infection rates averaging less than 15% even when larvae are fed thousands of vesicles in a single dose. In contrast, infection rates for caterpillars injected with as few as 10 virion-containing vesicles are typically greater than 90%, and experiments with parasitic wasps show that these insects can transmit ascoviruses.

Baculoviruses

Baculoviruses (family Baculoviridae) are large, enveloped, double-stranded, occluded DNA viruses. These viruses are divided into two main types, commonly known as the nuclear polyhedrosis viruses (NPVs) and the granulosis viruses (GVs). Both NPVs and GVs are highly infectious by feeding, and in some insect species periodically cause epizootics, or widespread outbreaks of disease, that result in significant (> 90%) declines in caterpillar populations.

NUCLEAR POLYHEDROSIS VIRUSES The NPVs (Fig. 1) are known from a wide range of insect orders but have been most commonly reported by far from lepidopterous insects, from which well over 500 isolates are known. Many of these are different viruses (i.e., viral species). NPVs replicate in the nuclei of cells, generally causing an acute fatal disease. The virions are large (80–200 nm × 280 nm) and consist of one

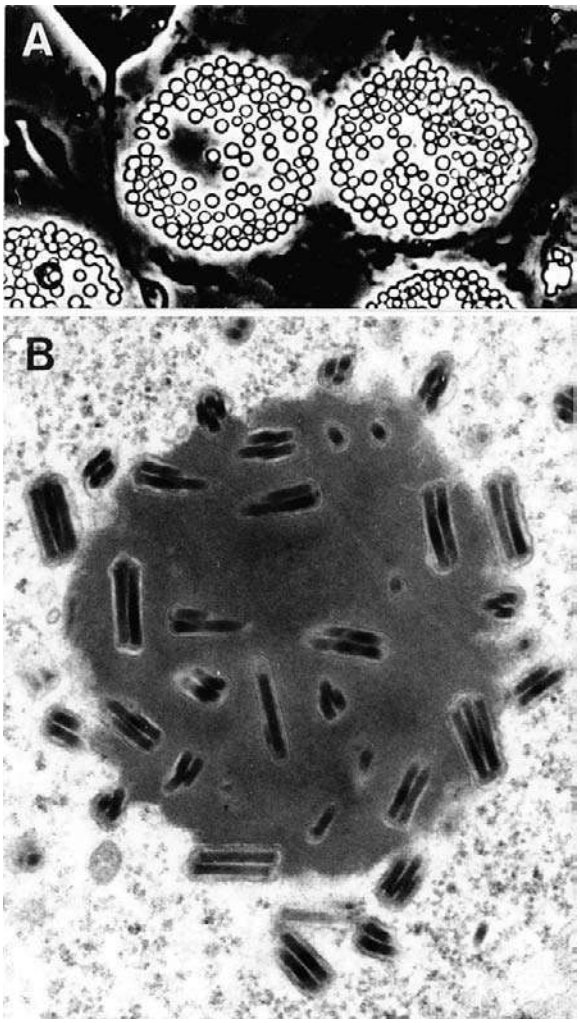

FIGURE 1 Nuclear polyhedrosis virus polyhedra. (A) Wet mount preparation viewed with phase microscopy showing refractile polyhedra in two infected nuclei. (B) Transmission electron micrograph through a single polyhedron showing the enveloped rod-shaped virions, characteristic of NPVs, occluded within the polyhedral matrix. Upon ingestion, this matrix dissolves in the insect midgut, and the virions invade the host through midgut microvilli.

or more rod-shaped nucleocapsids with a double-stranded circular DNA genome enclosed in an envelope. The occlusion bodies of NPVs are referred to commonly as polyhedra because typically their shape is polyhedral. Polyhedra are large (ca. 0.5–2 μm) and form in the nuclei, where each occludes as many as several hundred virions. The NPVs of lepidopterous insects infect a range of host tissues, but those of other orders are typically restricted to the midgut epithelium. Some NPVs have a very narrow host range and may replicate efficiently only in a single species, whereas others, such as the AcMNPV (i.e., the NPV of the alfalfa looper, *Autographa californica*), have a relatively broad host range and are capable of infecting species in other genera.

GRANULOSIS VIRUSES The GVs, of which over 100 isolates are known, are closely related to the NPVs but differ from the latter in several important respects. The virions of GVs are similar to those of NPVs but contain only one nucleocapsid per envelope. GVs are known only from lepidopterous insects. Like NPVs, they initially replicate in the cell nucleus, but pathogenesis involves early lysis of the nucleus (as virions begin to assemble), which in the NPVs occurs only after most polyhedra have formed. After the nucleus has lysed, GV replication continues throughout the cell, which now consists of a mixture of cytoplasm and nucleoplasm. When completely assembled, the virions are occluded individually in small (200 nm × 600 nm) occlusion bodies referred as granules. Many GVs primarily infect the fat body, whereas others have a broader tissue tropism and replicate throughout the epidermis, tracheal matrix, and fat body. One, the GV of the grapeleaf skeletonizer *Harrisina brillians*, is unusual in that it replicates only in the midgut epithelium.

Use of Viruses as Insect Control Agents

The best example of the use of a virus as an insect control agent is the use of the NPV of the European spruce sawfly, *Gilpinia hercyniae*, as a classical biological control agent. The European spruce sawfly was introduced into eastern Canada from northern Europe around the turn of the century and had become a severe forest pest by the 1930s. Hymenopteran parasitoids were introduced from Europe in the mid-1930s as part of a biological control effort, and inadvertently along with these came the NPV, which was first detected in 1936. Natural epizootics caused by the virus began in 1938, by which time the sawfly had spread over 31,000 km^2. Most sawfly populations were reduced to below economic threshold levels by 1943 and remain under natural control today, the control being effected by a combination of the NPV, which accounts for more than 90% of the control, and the parasitoids.

Although viruses, particularly NPVs, are frequently associated with rapid declines in the populations of important lepidopterous and hymenopterous (sawfly) pests, *G. hercyniae* NPV is the only example of a virus that has proven effective as a classical biological control agent. Another putative baculovirus, the "nonoccluded" baculovirus of the palm rhinoceros beetle, *Oryctes rhinoceros*, has been a quasi-classical biological control success in that once introduced into populations, can yield control for several years, but ultimately it dissipates and must be reapplied. Moreover, augmentative seasonal introductions have been effective only rarely and are not well documented. Thus, the control potential of most viruses is best evaluated by assessing their utility as microbial insecticides. From this perspective, the iridoviruses are essentially useless because of their poor infectivity by feeding. Cytoplasmic polyhedrosis viruses are not much better because, although highly infectious by feeding, the disease they cause is chronic. CPVs have, however, been useful in some situations, such as for suppression of the pine caterpillar, *Dendrolimus spectabilis*, in Japan. Ascoviruses and entomopoxviruses have not been developed as control agents for any insect owing to lack of efficacy.

For several reasons, the viruses most commonly used or considered as microbial insecticides in industrialized as well as less developed countries are the NPVs. First, NPVs are common in and easily isolated from pest populations. In addition, production in their hosts is cheap and easy, and the technology for formulation and application is simple and adaptable to standard pesticide application methods. Most NPVs, however, are narrow in their host range, infecting only a few closely related species. Furthermore, although several can be grown *in vitro* in small to moderate volumes (ca. 20- to 300-liter cell cultures), no fermentation technology currently exists for their mass production on a scale that would permit repeated applications to hundreds of thousands of acres, which is possible with *Bacillus thuringiensis* (Bt) chemical insecticides. These two key limitations have been major disincentives for the commercial development of NPVs, especially in industrialized countries.

Despite these drawbacks, several NPVs have been registered as microbial insecticides even though the market size for most is small. And registered or not, several are used in many less developed countries, particularly for control of lepidopteran pests of field and vegetable crops. Moreover, over the past decade there has been renewed interest in developing NPVs because recombinant DNA technology offers potential for improving the efficacy of these viruses.

Other Uses

In addition to the use of NPVs in insect control, one baculovirus, the AcMNPV noted earlier, has been developed as an expression vector for producing a large number of foreign proteins *in vitro.* This expression system takes advantage of the strong polyhedrin promoter system, which in the wild-type viruses produces large amounts of the polyhedria used to occlude virions. By substituting foreign genes for the polyhedrin gene, it is possible to synthesize in insect cell cultures large quantities of foreign proteins, such as the capsid proteins of viruses that attack the vertebrates used for vaccine development.

BACTERIA

Bacteria are relatively simple unicellular microorganisms that lack internal organelles such as a nucleus and mitochondria, and reproduce by binary fission. With a few exceptions, most of those that cause disease in insects grow readily on a wide variety of inexpensive substrates, a characteristic that greatly facilitates their mass production. A variety of bacteria are capable of causing diseases in insects, but those that have received the most study are spore-forming bacilli (family Bacillaceae), especially *B. thuringiensis.* Many subspecies of Bt are used as bacterial insecticides and as a source of genes for insecticidal proteins added to make transgenic plants resistant to insect attack, especially attack by caterpillars and beetles. The other bacterial insect pathogens that have received various degrees of study are *B. sphaericus, Paenibacillus popilliae,* and *P. larvae,* the latter being the etiological agent of foulbrood, an important disease of honey bee larvae, *Serratia entomophila* and *S. marcescens.* Several of these, in order of importance, are discussed here to represent the diversity of bacteria that cause disease insects.

Bacillus thuringiensis

B. thuringiensis is a complex of bacterial subspecies that occur commonly in such habitats as soil, leaf litter, on the surfaces of leaves, in insect feces, and as a part of the flora in the midguts of many insect species. Bts are characterized by the production of a parasporal body during sporulation that contains one or more protein endotoxins in a crystalline form (Fig. 2). Many of these are highly insecticidal to certain insect species. These endotoxins are actually protoxins activated by proteolytic cleavage in the insect midgut after ingestion. The activated toxins destroy midgut epithelial cells, killing sensitive insects within a day or two of ingestion. In insects species only moderately sensitive to the toxins, such as *Spodoptera* species (caterpillars commonly known as armyworms), the spore contributes to pathogenesis. Bt also produces other insecticidal compounds including β-exotoxin, zwittermicin A, and vegetative insecticidal proteins (Vips).

The most widely used Bt is the HD1 isolate of *B. thuringiensis* subsp. *kurstaki* (Btk), an isolate that produces four major endotoxin proteins packaged into the crystalline parasporal body (Fig. 2B). This isolate is the active ingredient in numerous commercially available bacterial insecticides used to control lepidopterous pests in field and vegetable crops, and in forests. Another successful Bt is the ONR60A isolate of *B. thuringiensis* subsp. *israelensis* (Bti), which is highly toxic to the larvae of many mosquito and blackfly species. This isolate also produces a parasporal body that contains four major endotoxins (Fig. 2B), but these are different from those that occur in Btk. Several commercial products based on Bti are available and are used to control both nuisance and vector mosquitoes and blackflies. A third isolate of Bt that has been developed commercially is the DSM2803 isolate of *B. thuringiensis* subsp. *morrisoni* (pathovar *tenebrionis*). This

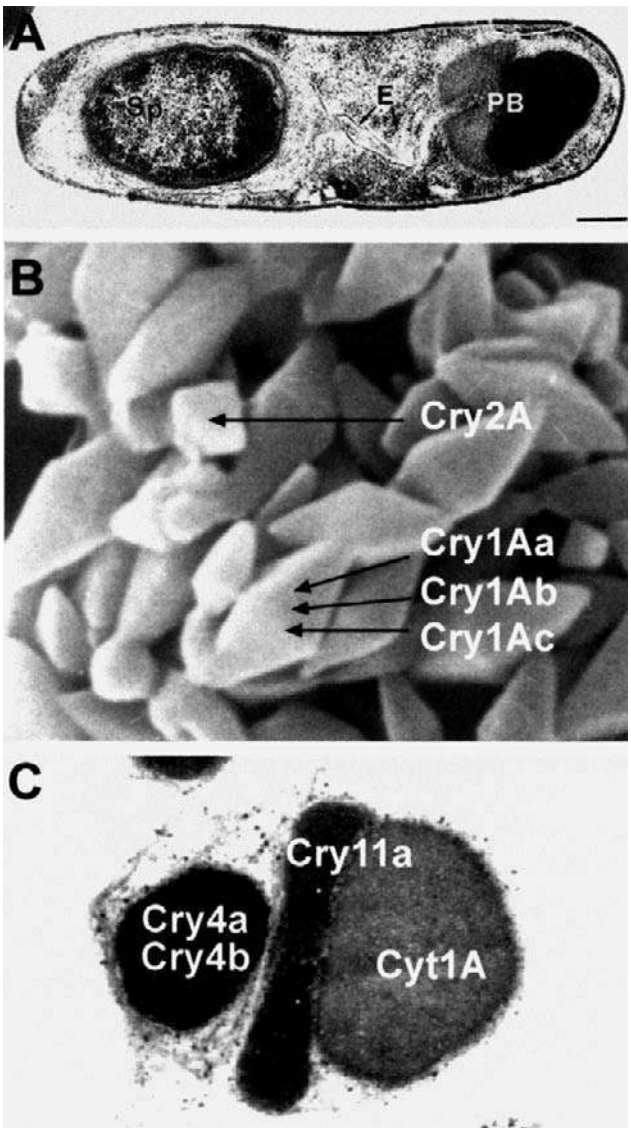

FIGURE 2 Sporulating cell of *Bacillus thuringiensis* and insecticidal parasporal bodies. (A) Transmission electron micrograph through a cell of *B. thuringiensis* subsp. *israelensis* illustrating a developing spore (Sp) and endotoxin-containing parasporal body (PB) outside the exoporium membrane (E). Bar, 250 nm. (B) Scanning electron micrograph of parasporal bodies (crystals) of *B. thuringiensis* subsp. *kurstaki,* a subspecies used widely to control caterpillar pests. The bipyramidal crystals contain three endotoxins (Cry1Aa, Cry1Ab, and Cry1Ac), whereas the smaller cuboidal crystal contains a single endotoxin (Cry2A). The bipyramidal crystals contain three endotoxin proteins (Cry1Aa, Cry1Ab, and Cry1Ac), and the cuboidal crystal has an additional toxin (Cry2A). This toxin complexity accounts for the broad spectrum of activity of many isolates of *B. thuringiensis* subsp. *kurstaki.* (C) Transmission electron micrograph of a parasporal body of *B. turingiensis* subsp. *israelensis* used widely to control the larvae of mosquitoes and blackflies. This parasporal body is also composed of four major endotoxins, a large semispherical inclusion containing Cyt1Aa, a dense spehrical body that apparently contains the Cry4Aa and Cry4Ba proteins, and a bar-shaped body that contains Cry11Aa. The endotoxin inclusions of this subspecies are held together by an envelope of unknown composition. This parasporal body has the highest specific toxicity of known Bt species, and this is due to synergistic interactions between the Cyt1Aa and Cry proteins as well as synergistic interactions among the Cry proteins. Bt endotoxins act by destroying the insect midgut epithelium (stomach).

isolate produces a cuboidal parasporal body toxic to many coleopterous insects and is used commercially to control several beetle pests.

All the above-mentioned isolates are essentially used as bacterial insecticides, applied as needed. A variety of formulations are available, including emulsifiable concentrates, wettable powders, and granules, for use against different pests in a variety of habitats. On a worldwide basis, millions of hectares are treated annually with products based on Bt. Recent estimates indicate the worldwide market is about $80 to $100 million and growing. Although most use is currently as a bacterial insecticide, plants have been engineered to produce Bt proteins for resistance to insects, and this use will probably surpass the use of Bt insecticides during the first decade of this century.

Bacillus sphaericus

Since the mid-1960s it has been known that many isolates of *B. sphaericus* (Bs) are toxic to certain mosquito species. Over the past three decades, three isolates have been evaluated for their mosquito control potential, 1593 from Indonesia, 2297 from Sri Lanka, and 2362 from Nigeria. The 1593 and 2297 isolates were obtained form soil and water samples at mosquito breeding sites, whereas 1593 was isolated from a dead adult blackfly.

Like Bt, Bs acquires its toxicity as the result of protein endotoxins that are produced during sporulation and assembled into parasporal bodies. Bs is unusual in that the main toxin is a binary toxin, (i.e., composed of two protein subunits). These are proteolytically activated in the mosquito midgut to release peptides having molecular masses of, respectively, 43 and 39 kDa, that associate to form the binary toxin, with the former protein constituting the binding domain, and the latter the toxin domain. The toxins bind to microvilli of the midgut epithelium, causing hypertrophy and lysis of cells, destroying the midgut and killing the mosquito larva.

Paenibacillus popilliae

P. popilliae is an highly fastidious bacterium that is the primary etiological agent of the so-called milky diseases of scarab larvae. These insects are the immature stages of beetles, such as the Japanese beetle, *Popillia japonica,* that are important grass and plant pests belonging to the coleopteran family Scarabaeidae. The term "milky disease" is derived from the opaque white color that characterizes diseased larvae and results from the accumulation of sporulating bacteria in larval hemolymph (blood). The disease is initiated when grubs feeding on the roots of grasses or other plants ingest the bacterial spores. The spores germinate in the midgut and vegetative cells invade the midgut epithelium, where they grow and reproduce, changing in form as they progress toward invasion of the hemocoel (body cavity). After passing through the basement membrane of the midgut, the bacteria colonize the blood over a period of several weeks and sporulate, reaching populations of 100,000,000 cells per milliliter. For larvae that ingest a sufficient number of spores early in development, the disease is fatal. Dead larvae in essence become foci of spores that serve as a source of infection for up to 30 years.

Despite decades of research, suitable media for the growth and mass production of *P. popilliae in vitro* have not been developed. Thus, the technical material (i.e. spores) used in commercial formulations is produced in living, field-collected scarab larvae. Nevertheless, a small but steady market remains for *P. popilliae* in the United States because of serious problems due to scarab larvae, such as damage to turf grass by larvae of the Japanese beetle.

Serratia entomophila

A novel bacterium named. *S. entomophila* causes amber disease in the grass grub, *Costelystra zealandica,* an important pest of pastures in New Zealand, and has been developed as a biological control agent for this pest. This bacterium adheres to the chitinous intima of the foregut, where it grows extensively, eventually causing the larvae to develop an amber color; the result of infection is death. The bacterium is easily grown and mass-produced *in vitro* and can now be grown to densities as high as 4×10^{10} cells ml^{-1}. Successful mass production of *S. entomophila* led to its rapid commercialization. It is now used to treat infested pastures in New Zealand at a rate of one liter of product per hectare. Liquid formulations of this living, non-spore-forming bacterium are applied with subsurface application equipment. The rapid development and commercialization of the bacterium, even though the use is rather restricted, shows how microbials can be successful in niche markets, where there are few alternatives, and mass production methods, the most critical factor, are available.

FUNGI

The fungi constitute a large and diverse group of eukaryotic organisms distinguished from others by the presence of a cell wall, as in plants, but lacking chloroplasts and thus the ability to carry out photosynthesis. Fungi live either as saprophytes or as parasites of plants and animals, and require organic food for growth, obtained by absorption from the substrates on which they live. The vegetative phase, known as a thallus, can be either unicellular, as in yeasts, or multicellular and filamentous, forming a mycelium, the latter being characteristic for most of the fungi that attack insects. During vegetative growth, the mycelium consists primarily of hyphae, which may be septate or nonseptate, and these grow throughout the substrate to acquire nutrients. Reproduction can be sexual or asexual, and during this phase the mycelium produces specialized structures such as motile spores, sporangia, and conidia, typically the agents by which fungi infect insects. Fungi usually grow best under wet or moist conditions, and those that are saprophytic as well as many of the parasitic species are easily cultured on artificial media.

The fungi are divided into five major subdivisions, and these reflect the evolution of the biology of fungi from an aquatic to terrestrial habitats. For example, species of the genera *Coelomomyces* and *Lagenidium* (subdivision Mastigomycotina) are aquatic and produce motile zoospores during reproduction, whereas members of the genera *Metarhizium* and *Beauveria* (subdivision Deuteromycotina) are terrestrial and reproduce and disseminate via nonmotile conidia.

Unlike most other pathogens, fungi usually infect insects by active penetration through the cuticle. The typical life cycle begins when a spore, either a motile spore or a conidium, lands on the cuticle of an insect. Soon after, under suitable conditions, the spore germinates, producing a germ tube that grows and penetrates down through the cuticle into the hemocoel. Once in the hemolymph, the fungus colonizes the insect. Hyphal bodies bud off from the penetrant hyphae and either continue to grow and divide in a yeastlike manner or elongate, forming hyphae that grow throughout the insect body. Complete colonization of the body typically requires 7 to 10 days, after which the insect dies. Some fungi produce peptide toxins during vegetative growth, and in these strains death can occur within 48 h. Subsequently, if conditions are favorable, which generally means an ambient relative humidity of greater than 90% in the immediate vicinity of the dead insect, the mycelium will form reproductive structures and spores, thereby completing the life cycle. Depending on the type of fungus and species, these will be produced either internally or externally as motile spores, resistant spores, sporangia, or conidia.

Fungi are one of the most common types of pathogen observed to cause disease in insects in the field. Moreover, outbreaks of fungal diseases under favorable conditions often lead to spectacular epizootics that decimate populations of specific insects over areas as large as several hundred square kilometers. As a result, there has been interest in using fungi to control insects for well over a century; the first efforts, in Russia in the late 1880s, used *Metarhizium anisopliae* to control the wheat cockchafer *Anisoplia austriaca.* Though there have been numerous attempts since then to develop fungi as commercial microbial insecticides, very few of these efforts have met with success. Thus, at present barely a handful of commercially available fungal insecticides are available for use in industrialized countries, and true commercial success has remained elusive. On the other hand, in developing countries (e.g., Brazil and China), "cottage industry" technology like that used to produce viruses has been turned to the production fungi such as *M. anisopliae* and *Beauveria bassiana.* A quasi-commerical product Boverin, developed and used in Russia for control of the Colorado potato beetle, proved ineffective in the United States. Current efforts to find alternatives to chemical insecticides have intensified research on fungi, with the aim of identifying new isolates or improving existing strains through molecular genetic manipulation. Researchers hope to obtain products that will prove more successful as either classical biological control agents or mycoinsecticides. The subsections that follow summarize briefly the critical biological features of selected fungi to illustrate the advantages and disadvantages of these as control agents.

Aquatic Fungi

Aquatic fungi of two types that attack mosquito larvae have received considerable study: species of *Coelomomyces* (class Chytridiomycetes: order Blastocladiales) and *Lagenidium giganteum* (class Oomycetes: order Lagenidiales).

The genus *Coelomomyces* comprises over 80 species of obligately parasitic fungi that have a complex life cycle involving an alternation of sexual (gametophytic) and asexual (sporophytic) generations. The sexual phase parasitizes a microcrustacean host, typically a copepod, whereas the asexual generation develops, with rare exception, in mosquito larvae. In the life cycle, a biflagellate zygospore invades the hemocoel of a mosquito larva, where it produces a sporophyte that colonizes the body and forms resistant sporangia. The larva dies and subsequently the sporangia undergo meiosis, producing uniflagellate meiospores that invade the hemocoel of a copepod host, where a gametophyte develops. At maturation, the gametophyte cleaves, forming thousands of uniflagellate gametes. Cleavage results in death of the copepod and escape of the gametes, which complete the life cycle by fusing to biflagellate zygospores, which then seek out another mosquito host. The life cycles of these fungi are highly adapted to those of their hosts. Moreover, as obligate parasites these fungi are very fastidious in their nutritional requirements, and as a result no species of *Coelomomyces* has been cultured *in vitro.*

Coelomomyces, the largest genus of insect-parasitic fungi, has been reported worldwide from numerous mosquito species, many of which are vectors of important diseases such as malaria and filiariasis. In some of these species, *Anopheles gambiae* in Africa, for example, epizootics caused in some areas by *Coelomomyces* kill greater than 95% of the larval populations. Such epizootics led to efforts to develop several species as biological control agents. For several reasons, however, these efforts were discontinued. One important factor was the discovery that the life cycle requires a second host for completion. Also contributing were the inability to culture these fungi *in vitro* and the development of Bti as a bacterial larvicide for mosquitoes.

Alhough it is unlikely that *Coelomomyces* fungi will be developed as a biological control agents, interest remains in developing *L. giganteum.* This oomycete fungus is easily cultured on artificial media and does not require an alternate host. In the life cycle, a motile zoospore invades a mosquito larva through the cuticle. Once within the hemocoel, the fungus colonizes the body over a period of 2 to 3 days, producing an extensive mycelium consisting largely of nonseptate hyphae. Toward the end of growth, the hyphae become septate, and out of each segment an exit tube forms which grows back out through the cuticle and forms zoosporangia at the tip. Zoospores quickly differentiate in these,

exiting through an apical pore to seek out a new substrate. In addition to this asexual cycle, thick-walled resistant sexual oospores can be formed in the mosquito cadaver.

Terrestrial Fungi

The fungi that have received the most attention for use in biological control are terrestrial fungi, with most emphasis placed on the development of selected species of hyphomycetes such as *M. anisopliae* and *B. bassiana* for use as microbial insecticides. In addition, the more specific and nutritionally fastidious entomophthoraceous fungi continue to receive attention, but for their potential use as classical biological control agents rather than as microbial insecticides. Representative examples of these terrestrial fungi are discussed in the subsections that follow.

ENTOMOPHTHORALES These fungi comprise a large order of the class Zygomycetes that contains numerous genera, many species of which are commonly found parasitizing insects and other arthropods. The fungi routinely cause localized and sometimes widespread epizootics in populations of hemipterous and homopterous insects, particularly aphids and leafhoppers, but also in insects of other types such as grasshoppers, flies, beetle larvae, and caterpillars. In addition, a few species of the genus *Conidiobolus* are able to cause mycoses in some mammals, including humans. Apart from these few species, most of the entomophthoraceous fungi are highly specific, obligate parasites of insects and therefore their use for biological control poses no threat to nontarget organisms. As with *Coelomomyces,* however, the complex nutritional requirements, which have thus far prevented mass production *in vitro,* and high degree of host specificity, make these fungi poor candidates for development as microbial insecticides. Moreover, the conidia are very fragile, providing a challenge to formulation, and the resistant spores, like the oospores of *L. giganteum,* are difficult to germinate in a predictable manner. Nevertheless, there is evidence that if cultural practices in crop production are modified, these fungi can provide effective insect control where they occur naturally, and through the introduction of foreign strains and species (i.e., a classical biological control approach).

The most important genera found attacking insects in the field are *Conidiobolus* (aphids), *Erynia* (aphids), *Entomophthora* (aphids), *Zoophthora* (aphids, caterpillars, beetles), and *Entomophaga* (grasshoppers, caterpillars). Although many species of these genera cause epizootics and have received considerable study, none really seems to have much potential for development as a commercial microbial insecticide. On the other hand, cultural control, classical biological control, and environmental monitoring methods continue to show promise for using entomophthoraceous fungi for insect control. For example, the introduction of *Erynia radicans* from Israel into Australia to control the spotted alfalfa aphid, *Therioaphis maculata,* has proven a classical biological control success.

A recent example of apparent classical biological control can he found in the natural outbreaks of *Entomophaga miamiaga* in larval populations of the gypsy moth, *Lymantria dispar,* an important pest of deciduous forests throughout several states comprising the middle Atlantic and New England regions of the United States. Outbreaks of *E. miamiaga* have reduced larval populations to below economic thresholds, and the fungus is spreading westward naturally, and with human assistance, to gypsy moth populations established in other states. The source of this fungus is Japan, although it is not clear when the fungus causing present outbreaks of disease first appeared in the United States. The fungus was purposely introduced into the United States around the turn of the century but seems not to have become established at that time. Then in the late 1980s, outbreaks of *E. miamiaga* began to occur in Connecticut and New York, and later in Virginia. In areas where it has established, given sufficient rainfall, the fungus seems to be capable of keeping the gypsy moth population below defoliation levels. It will require another 10 years of evaluation to determine whether this is a valid instance of classical biological control by a fungus.

CLASS HYPHOMYCETES The hyphomycete fungi belong to the fungal subdivision Deuteromycotina (imperfect fungi), a grouping erected to accommodate fungi for which the sexual phase (perfect state) has been lost or remains unknown. This group contains the fungal species that most workers consider to have the best potential for development as microbial insecticides, *B. bassiana* and *M. anisopliae,* the agents of, respectively, the white and green muscardine diseases of insects. Unlike the fungi already discussed, these two species have very broad host ranges and probably are capable of infecting insects of most orders.

With respect to the general life cycle of these fungi, the process of invasion, colonization of the insect body, and formation of conidiophores and conidia is similar to that described for the other fungi. During invasion and colonization, some fungal species produce peptide toxins that quicken host death. The infectious agent is the conidium (Fig. 3), and the taxonomy for the hyphomycetes is based primarily on the morphology of the reproductive structures, particularly the conidiophores and the conidia. Most of the hyphomycete fungi used or under development grow well on a variety of artificial media, and this attribute, along with their ability to infect insects via the cuticle, favors commercial development. In the "cottage industry" commercial operations in Brazil, China, and the former Soviet Union, solid or semisolid substrates are used for production, and the primary ingredients are grain or grain hulls.

In general, the development of *B. bassiana* and *M. anisopliae* is being targeted for control of insects that live in cooler and moist environments, such as beetle larvae in soil and planthoppers on rice, though the former species is also being evaluated against whiteflies in glasshouses, as well as grasshoppers, especially locusts, in field crops. In addition to these two species,

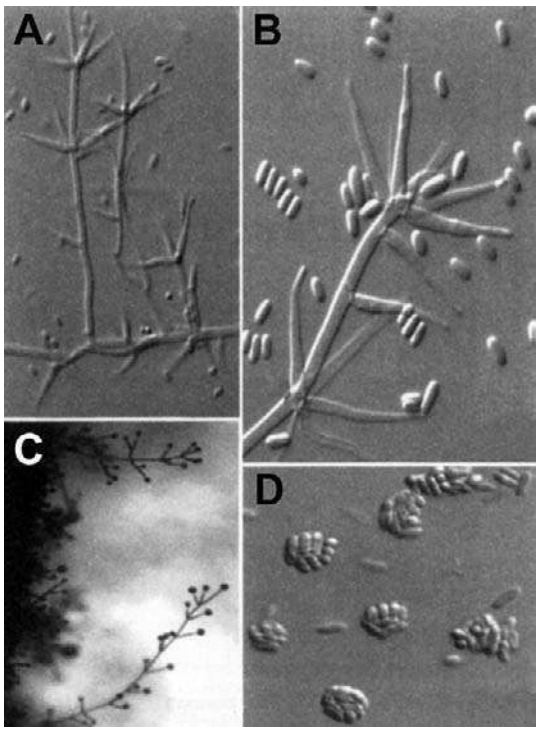

FIGURE 3 Typical reproductive structures of deuteromycete (imperfect) fungi. (A–D) Wet mount preparation of conidia-generating cells and conidia of *Verticillium lecanii,* which commonly attacks aphids and whiteflies. The conidia visible as free conidia and conidial clusters in (B) and (D) are the principal infective units. When these come in contact with an insect host, they germinate and penetrate the body, forming a mycelium that colonizes the insect over a period of several days. When conditions are appropriate, typically meaning high relative humidity, hyphae penetrate back out through the cuticle, producing conidiophores, the visible branched structures in these panels (A–C), which form reproductive conidia at their tips. (Photomicrographs courtesy of Richard A. Humber, U.S. Department of Agriculture).

several species with much narrower host ranges are considered to have potential for development, including *Paecilomyces fumoso-rosea* (for whiteflies), *Verticillium lecanii* (for aphids and whiteflies in glasshouses), *Hirutella thompsonii* (for mites), and *Nomurea rileyi* (for noctuid caterpillars).

With these apparent advantages, it is natural to ask why none of the hyphomycete fungi have been commercially successful as microbial insecticides in developed countries. There are several reasons related to their biological properties. First and foremost is that the production of conidia or mycelial fragments that are used as the active ingredient of formulations is not cost-effective because too much material is required to allow the achievement of an acceptable level of control. In addition to the problem of inefficient yields, the formulations are bulky, and preservation of fungal viability beyond a few months is low because the conidia are fragile. In mosquito and blackfly control, similar constraints apply. In addition, the discovery of cost-effective strains of *B. thuringiensis* and *B. sphaericus* has generally eliminated imperfect fungi, as well as many other microorganisms, for consideration as biological control agents for these important nuisance and vector insects.

In developing countries, *B. bassiana* and *M. anisopliae* have been used in some crops with considerable success. For example, in China *B. bassiana* has been used to control the European corn borer, *Ostrinia nubilalis,* in maize. The fungus is produced in large covered pits on maize stalks. In Brazil, a preparation of *M. anisopliae* known as Metaquino has been used for many years on sugarcane plantations and in pastures to control the spittlebug, *Mahanarva posticata.* Fungal conidia are produced in sealed plastic bags on rice. Figures indicate that as much as 50,000 ha is treated annually, and reductions in spittlebug populations are sufficient to keep populations below damaging levels. In the South Pacific, *M. anisopliae* has also been used to assist control of the rhinoceros beetle, *Orycetes rhinocerous,* a serious pests of coconut palms. Application of conidia at a rate of 50 g per square meter of soil yielded 80% larval mortality and improved cococnut yields by 25%. While these are examples of local successes, their applicability to agricultural production in developed countries is questionable.

PROTOZOA

Protozoa is a general term applied to a large and diverse group of eukaryotic unicellular motile microorganisms that belong to what is now known as the kingdom Protista. Members of this kingdom can be free living and saprophytic, commensal, symbiotic, or parasitic. The cell contains a variety of organelles, but no cell wall, and cells vary greatly in size and shape among different species. Feeding is by ingestion or more typically by adsorption, and vegetative reproduction is by binary or multiple fission. Sexual reproduction, often useful for taxonomy, can be very complex, but asexual reproduction occurs as well. Many protozoa produce a resistant spore stage that is also used in taxonomy. Divided into a series of phyla based primarily on mode of locomotion and structure of locomotory organelles, the kingdom includes the Sarcomastigophora (flagellates and amoebae), Apicomplexa (sporozoa), Microspora (microsporidia), Acetospora (haplosporidia, now thought to be a type of parasitic alga), and Ciliophora (ciliates). Protozoa of some types, such as the free-living amoebae and ciliates, are easily cultured *in vitro,* whereas many of the obligate intracellular parasites have not yet been grown outside cells.

As might be expected from such a large and diverse group of organisms, many species of protozoans are associated with insects, and the biology of these associations covers the gamut from being symbiotic to parasitic. Those that are parasitic have the general feature of causing diseases that are chronic. Many of the parasitic types, especially the microsporidia, build up slowly in insect populations, eventually causing epizootics that lead to rapid declines in populations of specific species. These epizootics attracted interest in the possibility of using protozoa to control pest insects, and over the past several decades numerous studies have been aimed at evaluating this potential. In general, these studies have shown that protozoa hold little potential for use as fast-acting microbial insecticides because of the chronic nature of the diseases they cause and because commercially suitable

methods for mass production are lacking. However, as with the entomophthoraceous fungi, the possibility exists that protozoans, particularly microsporidia, may be useful as classical biological control agents. Clear examples of the effectiveness of such strategies remain to be demonstrated.

The life cycles and biologies that occur among the various protozoa that attack insects are too diverse in relation to their pest control potential for even a few to be covered here. Instead, the group with the most potential—the microsporidia—is described in terms of general biology and possible use in insect control.

General Biology of Microsporidia

The microsporidia (phylum Microspora) are the most common and best studied of the protozoans that cause important diseases of insects. Well over 800 species are known, and most of these have been described from insects. Microsporidia have been most commonly described from insects of the orders Coleoptera, Lepidoptera, Diptera, and Orthoptera, but they are also known from other orders and probably occur in all. The epizootics in insect populations caused by protozoa are usually due to microsporidia. All microsporidia are obligate intracellular parasites and are unusual in that they lack mitochondria. In addition, they produce spores that are distinguished from the spores of organisms of all other known types by the presence of a polar filament (Fig. 4), a long coiled tube inside the spore used to infect hosts with the sporoplasm.

The typical microsporidian life cycle begins with the ingestion of the spore by a susceptible insect. Once inside the midgut, the polar filament everts, rapidly injecting the sporoplasm into host tissue. The sporoplasm is unicellular but may be uni- or binucleate. Upon entry into the cytoplasm of a host cell (e.g., the fat body in many species of insects), the sporoplasm forms a plasmodium (meront), which undergoes numerous cycles of vegetative growth (merogony). During these, the cells multiply extensively, dividing by binary or multiple fission and spreading to other cells, and, in many species, to other tissues of the host. After several mergonic cycles, the microsporidian undergoes sporulation. This consists of two major phases, sporogony—a terminal reproductive division committed to sporulation—and spore morphogenesis. In the sexual phase of reproduction, meiosis occurs early during sporogony. The spores, which in general measure several micrometers in diameter and length, have a thick wall and are highly, refractile when viewed by phase microscopy. The disease often lasts for several weeks during which billions of spores may accumulate in the tissues of a single infected host.

FIGURE 4 Representative microsporidian spore: transmission electron micrograph through a uninucleate spore of *Amblyospora abserrati* from a larva of the mosquito *Ochlerotatus abserratus.* The circular structures on each side of the spore are cross sections through the polar filament that is used to inject the contents of the spore into the mosquito body after ingestion and activation of the spore. (Photomicrograph courtesy of Dr. Theodore Andreadis, Connecticut Agricultural Experiment Station.)

Microsporidian systematics is based on the size and structure of the spores, life cycles, and host associations. In addition to transmission by ingestion, many microsporidia are transmitted vertically from adult females to larvae via the egg (transovarially). With respect to host range, some species are species specific, whereas others occur in many species of the same family or order, and some can be transmitted to insects of different orders.

Microsporidia as Biological Control Agents

Naturally occurring epizootics caused by microsporidia are periodically very effective in significantly reducing insect pest populations. The problem is that these epizootics cannot be predicted with any degree of accuracy, nor can they be relied upon for adequate control, even though many of the conditions that facilitate their occurrence are known. The epizootics caused by *Nosema pyrausta* in populations of the European corn borer often a classic example of this unreliability. These epizootics are useful when they occur, but because this often happens too late to prevent economic damage, reliance on *N. pyrausta* alone is insufficient. Thus, efforts have been directed toward developing methods for amplifying spore loads in the field through inundative releases, in essence using microsporidia as microbial insecticides.

Because they are obligate intracellular parasites that lack mitochondria, microsporidia cannot be grown on artificial media. Several species have been grown, however, in established insect cell lines, although this is not practical for field use. For field application, whether for microbial insecticide trials or for introductions into populations, spores are grown in living hosts. With such methods the yield can be quite high (10^9–10^{10} spores per host). These yields in terms of the number of larvae that must be grown to treat a hectare and infect most of the target population are comparable to the requirements for nuclear polyhedrosis viruses. Thus, if the microsporidia could cause acute diseases, they would be on an equal footing with many of the NPVs. However, the diseases are chronic, and even if a high percentage of the target pest population is infected, there all too often is little, if any, crop protection. In fact, if advanced instars such as thirds and fourths are treated, the larvae may live longer and

cause greater crop damage than if the fields were left untreated. Thus, microsporidia are not useful as microbial insecticides.

There is now a general realization that microsporidia and other protozoans have virtually no potential for use as microbial insecticides. They may, however, be useful as population management tools.

See Also the Following Articles

Biological Control • Cell Culture • Genetically Modified Plants • Integrated Pest Management

Further Reading

Adams, J. R., and Bonami, J. R. (eds.) (1991). "Atlas of Invertebrate Viruses." CRC Press, Boca Raton, FL.

Anthony, D. W., and Comps, M. (1991). Iridoviridae. *In* "Atlas of Invertebrate Viruses" (J. R. Adams and J. R. Bonamri, eds.), pp. 55–86. CRC Press, Boca Raton, FL.

Brooks, W. M. (1988). Entomogenous protozoa. *In* "Handbook of Natural Pesticides. Microbial Pesticides," Part A, "Entomogenous Protozoa and Fungi" (C. M. Ignoffo and N. B. Mandava, eds.), pp. 1–149. CRC Press, Boca Raton, FL.

Federici, B. A. (1998). *Bacillus thuringiensis* in biological control. *In* "Handbook of Biological Control" (T. S. Bellows, G. Gordh, and T. W. Fisher, eds.), pp. 519–529. Academic Press, San Diego, CA.

Fuxa, J. R., and Tanada, Y. (eds). (1987) "Epizootiology of Insect Diseases." Wiley, New York.

Entwistle, P. F., Cory, J. S., Bailey, M. J., and Higgs, S. (eds.) (1993). "*Bacillus thuringiensis,* an Environmental Biopesticide: Theory and Practice." Wiley Chicester, U.K.

Hajek, A. E., Elkinton, J. S., and Witcosky, J. J. (1996). Introduction and spread of the fungal pathogen *Entomophaga miamiaga* (Zygomycetes: Entomophthorales) along the leading edge of gypsy moth (Lepidoptera: Lymantriidae) spread. *Environ. Entomol.* **25,** 1235–1247.

Henry, J. (1990) Control of insects by protozoa. *In* "New Directions in Biological Control" (R. Baker and P. Dunn, eds.), pp. 161–176. Liss, New York.

Hunter-Fujita, F. R., Entwistle, P. E. Evans, H. F., and Crook, N. E. (eds.) (1998). "Insect Viruses and Pest Management." Wiley, Chichester, U. K.

Lacey, L. A. (1997). "Manual of Techniques in Insect Pathology." Academic Press, San Diego, CA.

McCoy, C. W., Samson, R. A., and Boucias, D. G. (1988). Entomogenous fungi. *In* "Handbook of Natural Pesticides, Microbial Pesticides," Part A, "Entomogenous Protozoa and Fungi" (C. M. Ignoffo and N. B. Mandava, eds.), pp. 151–236. CRC Press, Boca Raton, FL.

Miller, L. K. (ed.). (1997). "The Baculoviruses." Plenum Press, New York.

Tanada, Y., and Kaya, H. K. (1993). "Insect Pathology." Academic Press, San Diego, CA.

Phasmida

(Stick and Leaf Insects)

Erich H. Tilgner

University of Georgia

Phasmida are nocturnal exopterygote insects. They exhibit a variety of unpredictable and bizarre shapes. Some look like twigs or tree bark (Figs. 1, 2) and may seem to be covered by lichens or moss. Others are indistinguishable from living or dead leaves (Fig. 3), mimicking even leaf veins and mildew spots to perfection. Phasmida are large, and a few are remarkable for their gigantic size. The longest insect species in the world is *Phobaeticus kirbyi* from Borneo, with one documented female measuring 55 cm in length. Phasmida inhabit tropical, subtropical, and temperate forests, savannas, grasslands, and chaparral; their diversity is highest in the tropics.

FIGURE 1 A female of *Oxyartes spinosissimus,* a lichen mimic.

PHYLOGENY AND CLASSIFICATION

Over 3000 species of Phasmida have been described. The genus *Timema* from the western United States is considered to

FIGURE 2 A male and female of *Aplopus* sp. in the act of mating. The smaller male is hanging off the back of the female.

FIGURE 3 A female of the walking leaf *Phyllium bioculatum.*

be sister group to the remainder of the order, which is referred to as the Euphasmida. *Timema* are small, wingless, and cryptically colored. Euphasmida are larger, winged or wingless, usually possessing an elongated mesothorax, and are stereotyped as stick or leaf insects (Fig. 4). *Timema* have no fossil record. The oldest Euphasmida fossils date to the middle Eocene, 44 to 49 mya. Oligocene and Miocene fossils are known from Florissant shale, Baltic, and Dominican Republic amber.

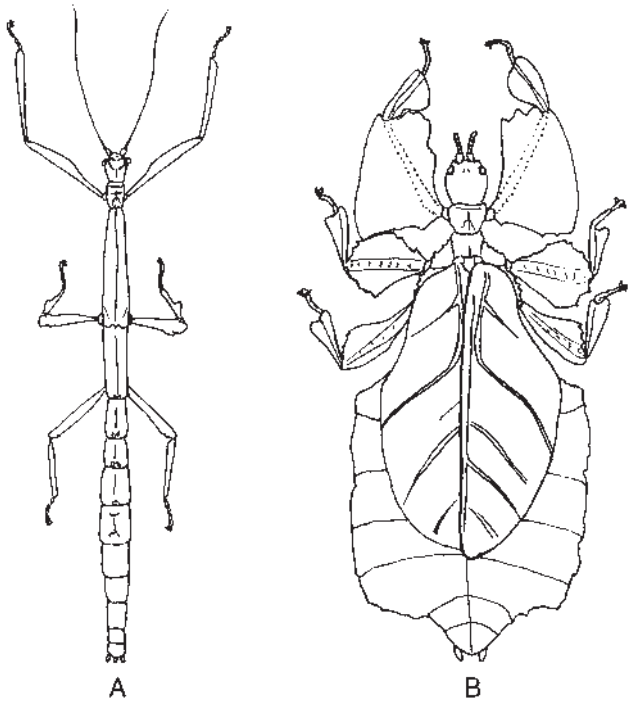

FIGURE 4 Female Euphasmida: (A) *Phenacephorus auriculatus* and (B) *P. bioculatum.*

The taxonomy of the order is problematic. No workable classification scheme exists, and those that are available are not based on phylogenetic relationships. Assignment to a category such as family, tribe, even suborder provides so little information that it is almost meaningless. This is in contrast other insect orders, such as Coleoptera, where a suborder, or family-level identification, say, provides a wealth of biological information about the specimen. In spite of the lack of an acceptable classification, the fauna of a few areas, (e.g., Europe, Malaysia, Borneo, Japan, United States, Canada, New Zealand) have been sufficiently studied to permit tentative identification of species by nonspecialists.

BIOLOGY

Sexual dimorphism is extreme in the Phasmida, and it is difficult to associate the sexes unless mating adults are found under natural conditions, or if males and females are obtained from the rearing of eggs in captivity. Reproduction is usually sexual, but many species are parthenogenetic. Eggs resemble plant seeds, are laid singly, and are either dropped, flicked, buried, glued to a surface, or riveted to a leaf. Some species rely on ants to disperse them. After successive molts, nymphs regenerate limbs lost by autotomy (the purposeful shedding of appendages). The entire life cycle takes from several months to several years depending on the species. Phasmida feed primarily on flowering plants, but a few eat either gymnosperms or ferns. The primary defense against predation is crypsis. Secondary defenses can include catalepsy (e.g., death feigning), startle displays, or the ejection of an irritating spray fired from a pair of prothoracic exocrine glands.

See Also the Following Articles

Crypsis • Parthenogenesis

Further Reading

Brock, P. D. (1999). The amazing world of stick and leaf-insects (R. Fry, ed.). Cravitz Printing, Essex, U.K.

Key, K. H. L. (1991). Phasmatodea (stick-insects). *In* "The Insects of Australia," Vol. I (CSIRO, ed.). Melbourne University Press, Carlton, Australia.

Tilgner, E. (2001). Fossil record of Phasmida (Insecta: Neoptera). *Insect System. Evol.* **31,** 473–480.

Pheromones

Ring T. Cardé and Jocelyn G. Millar
University of California, Riverside

Pheromones are chemical messages that induce a behavioral reaction or developmental process among individuals of the same species. The term is derived from the Greek for "carrier of excitation" and was coined in 1959 by the German

biochemist Peter Karlson and the Swiss entomologist Martin Lüscher during their investigations of the chemicals that regulate caste development in termites. In 1963 E. O. Wilson and W. H. Bossert of Harvard University formally distinguished two classes of pheromones. Releaser pheromones are messages that induce an immediate behavioral reaction in the receiver. The kinds of behavioral response evoked in insects are incredibly diverse, and they include alarm, defense, aggregation, attraction, kin and colony recognition, marking of territories and egg deposition sites, mating behaviors, recruitment, trail following, and even thermoregulation. In contrast, primer pheromones cause a physiological change in the receiver, such as development of a particular caste or sexual maturation, which eventually modifies the organism's behavior.

All pheromones fall under the broader umbrella classification of semiochemicals—chemicals that are involved in communication. The two major classes of semiochemicals besides pheromones are allomones and kairomones. These are solely interspecific cues, in contrast to pheromones, which are always intraspecific cues. Allomones are chemicals that provide some advantage to the emitter (e.g., defensive secretions), whereas kairomones are signals that confer an advantage to the receiver (e.g., emanations used by a parasite to locate a host). This article only touches on the diversity and complexity of pheromone-mediated behaviors and developmental changes in insects. Communication among social insects, especially among ants, bees, wasps, and termites, involves a highly sophisticated pheromonal language, in which the interpretation of the individual chemical constituents or "words" depends on their particular combinations, ratios, concentrations, and even order of presentation. Context, that is, the recent experiences of the receiver and its physiological state, is all-important in response.

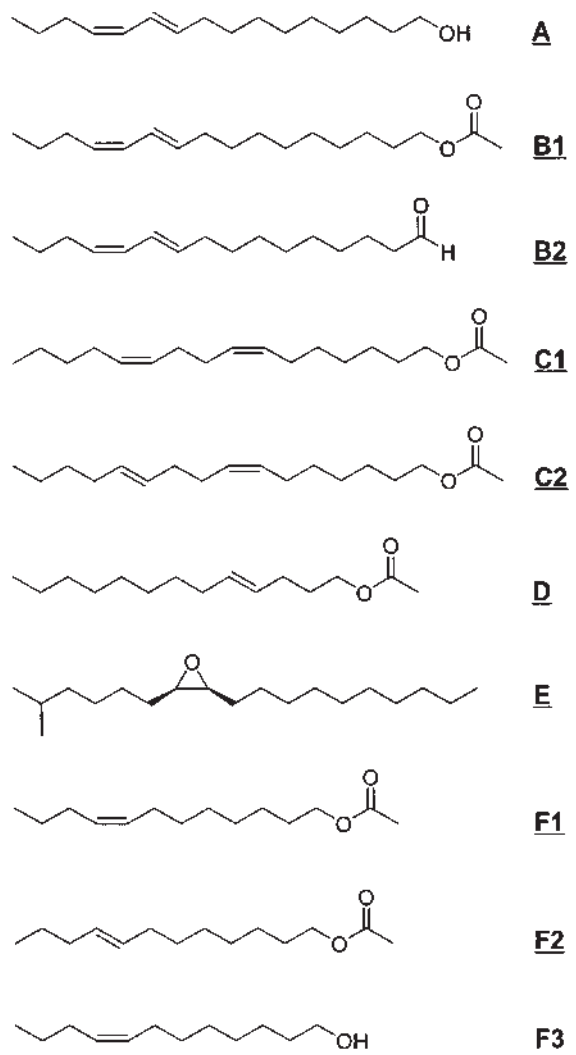

FIGURE 1 Representative structures of lepidopteran (moth) pheromones. (A) *Bombyx mori,* the commercial silkworm (Bombycidae), (B) *Hemileuca electra* (Saturniidae), (C) *Pectinophora gossypiella,* pink bollworm (Gelechiidae), (D) *Keiferia lycopersicella,* tomato pinworm, (Gelechiidae), (E) *Lymantria dispar,* gypsy moth (Lymantriidae), and (F) *Grapholita molesta,* oriental fruit moth (Tortricidae).

SEX ATTRACTANT PHEROMONES OF LEPIDOPTERA

The first definitive evidence of pheromone communication dates to experiments performed with the great peacock moth *Saturnia pyri* by the French naturalist Jean-Henri Fabre in the 1870s. Fabre sequestered a female moth in a screened cage following her morning emergence, to permit her wings to expand and harden. That evening more than 40 male moths arrived at Fabre's study, "eager to pay their respects to their marriageable bride born that morning." Further observations showed that cages that had housed virgin females also were attractive; this and other observations led Fabre to conclude that "effluvia of extreme subtlety" mediated attraction. Nearly 90 years would pass before microanalytical techniques would permit identification of the minute quantities of pheromone involved.

The first pheromone to be identified was the sex attractant pheromone of *Bombyx mori,* the commercial silkworm. This silkworm is an entirely domesticated species that is no longer capable of flight; its female-emitted pheromone (Fig. 1) induces upwind walking and courtship behaviors in conspecific males. The German biochemist Adolf Butenandt (who received the Nobel Prize in 1939 for his work identifying the human sex hormones) spent more than two decades in this quest. In 1959 he identified (*E*10,*Z*12)-hexadecadienol as the single compound causing upwind walking and copulatory attempts and named it "bombykol." To provide some perspective on this remarkable achievement, Butenandt and his coworkers extracted a half million female moths, finally isolating a few milligrams of the pure pheromone. Today's modern methods of isolation and characterization (especially coupled gas chromatography and mass spectrometry) were not yet available.

The amount of pheromone that is secreted from or is present in a pheromone-producing gland varies enormously

with species and, to some extent, behavioral function. Sex attractant pheromones may be present in microgram, nanogram, and even picogram (10^{-12} g) quantities per individual. Microanalytical techniques are now so advanced that identifications on occasion can be made from either gland extracts or airborne collections from a few individuals, and with just nanogram or even lower quantities of natural chemical. Even the always tedious behavioral bioassays, long used to monitor for behaviorally active components of gland extracts and airborne collections, have been largely supplanted by using a living insect antenna as a detector. The electroantennogram (EAG) was pioneered in the mid-1950s by Dietrich Schneider, working at the Max Planck Institute near Munich, Germany. In the 1970s, Wendell Roelofs of Cornell University adapted this assay to speed up identifications. A moth antenna was used to monitor which fractions separated by gas chromatography contained the active components. Later applications mounted the EAG apparatus at the outflow of a gas chromatograph column, and this living detector indicated the presence (and the retention times) of compounds that were likely to be behaviorally active by means of an electrical signal elicited by the interaction of pheromone and the antennal receptors. These advances allowed chemists to zero in quickly on the compounds present in crude extracts that were most likely to comprise the pheromone.

The sensitivity of a male silkworm to bombykol is legendary. It has been investigated by recording the electrical response of individual sensory hairs on their antennae (each antenna is equipped with 40,000 such hairs) and by monitoring a single male's change from quiescence to wing fanning and upwind walking. The estimates are astonishing: one bombykol molecule is sufficient to induce the firing of an individual receptor, and a behavioral response can be evoked with only 200 molecules ($\sim 10^{-19}$ g!).

Pheromone structures now have been described for several hundred species of moths. Nearly all these pheromones induce upwind flight by the male to the pheromone-releasing female. The majority of known structures for moth pheromones (examples in Fig. 1) are hydrocarbon chains, usually 10 to 18 carbons in length, with 1 to 3 double bonds and a terminal acetate, alcohol, or aldehyde. Less common structural motifs in moth pheromones include epoxides, ketones, and hydrocarbons with one or more double bonds or methyl branches; chain lengths known so far range from C_{10} to C_{23}. Many pheromones, such as those of the moths *Hemileuca electra* and *Grapholita molesta,* comprise blends of two, three, or even more components. Specificity of the chemical message is accomplished in many species by females emitting and males responding to precise ratios of their pheromone blend. For example, for males of *G. molesta,* the ratio of the *(Z)*-8- and *(E)*-8-dodecenyl acetate components must be very close to the 95:5 mix produced by the female for maximum attraction. The use of blends and in some species precise ratios allows many closely related moth species to have "exclusive" communication channels, even though they share some components of their respective blends. Other strategies for partitioning of the communication channel include restricting sexual activity to specific times of the day or night (Fig. 2).

FIGURE 2 Female of the day-active saturniid moth *Hemileuca electra* exposing her pheromone gland, located at the tip of her abdomen. Such pheromone-releasing behavior, termed "calling," and the male's mate-finding behaviors typically occur at set times of the day or night. In the Mojave Desert of California *H. electra* calls from midmorning to early afternoon; the closely related species *H. burnsi,* which shares pheromone components with *H. electra,* calls from midafternoon to dusk. Without exclusive times for mating activities, these species would cross-attract. (Photograph courtesy of Chris Conlan.)

Pheromones of other types are produced by males of many moths and facilitate close-range recognition and acceptance by the female. In a few species the sexual roles are reversed, with male moths being the attractive sex and recruiting females. Many male butterflies also use pheromones in courtship, disseminating an "aphrodisiac" scent from scales on their wings or, in some butterflies, from specialized paired brushes located at the tip of the abdomen. However, butterflies do not attract mates with long-distance pheromones; instead they rely on visual cues for mate finding.

ATTRACTANT AND AGGREGATION PHEROMONES

Although long-distance communication by attractant pheromones is well established in nearly all moth lineages, pheromones are widely used by many insect groups in mate finding. Such messages are categorized as either sex attractant pheromones, if one sex attracts the other (as in moths), or aggregation pheromones, if both sexes are attracted. Feeding on a plant host and release of aggregation pheromones typically are linked, and mating often occurs in such aggregations. Therefore aggregation pheromones can play a multifunctional role. Representative structures (Fig. 3) of sex attractants of insects other than moths include those of the cockroach, aphid, scale insect, caddisfly, sawfly, beetle, and true fruit fly. The chemistries of these messages are diverse, as are the locations of the glands responsible for their production.

BARK BEETLE PHEROMONES

Bark and ambrosia beetles (Scolytidae) use pheromones to facilitate colonization of host trees (aggregation) and to attract

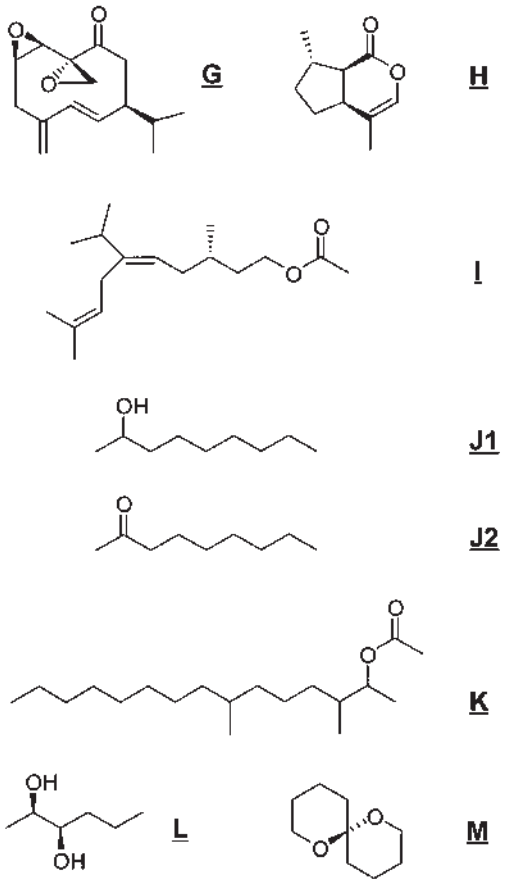

FIGURE 3 Examples of sex attractant pheromones of insects from nonlepidopteran orders: (G) periplanone B, from the American cockroach, *Periplaneta americana,* (H) nepetalactone, a pheromone component of a number of aphid species, (I) pheromone of California yellow scale, *Aonidiella citrina,* (J1 and J2) pheromone components of caddisflies, (K) pheromone component of diprionid sawflies, (L) (2*R*,3*R*)-2,3-hexanediol, a pheromone component of the cerambycid beetle *Hylotrupes bajulus,* and (M) *(R)*-1,7-dioxaspiro[5,5]decane, from the olive fruit fly *Dacus oleae.*

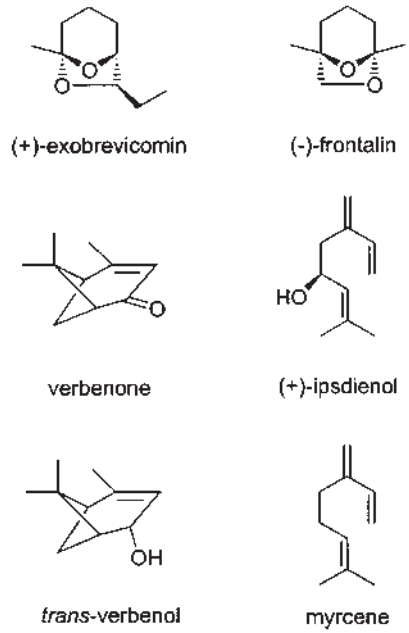

FIGURE 4 Pheromone components of *Dendroctonus brevicomis,* the western pine beetle. Myrcene is emitted by the beetle's principal host, ponderosa pine *(Pinus ponderosa).*

mates. Many scolytid species must attack a tree en masse if they are to overwhelm the tree's defense, which consists of exuding sap into the tunnel that each beetle bores. The first beetles to arrive may identify the host by means of chemicals emitted by the host tree itself; as they bore into the tree, they release pheromones and increase emission of tree chemicals that together attract both male and female beetles. David Wood of the University of California at Berkeley and Robert Silverstein, then at Stanford Research Institute, worked out these intricate interactions in *Dendroctonus brevicomis,* the western pine beetle. Infestations begin when females are attracted to their principal host, ponderosa pine, by myrcene, a monoterpene the tree emits as a defensive compound (allomone) when injured or stressed, and by the tree's silhouette. As the "pioneer" females bore into the tree's bark, they release their pheromone, (+)-*exo*-brevicomin (Fig. 4), which is augmented by increased release of myrcene from the host tree. Males are attracted and, after one enters the female's tunnel, he emits (–)-frontalin. The combination of host-, female-, and male-released volatiles attracts many more males and females, ensuring that the tree's sticky sap defense will be insufficient. Once males and females have mated, they alter their chemical message: males and females emit verbenone and *trans*-verbenol and males release (+)-ipsdienol (Fig. 4). Together, these three chemicals interrupt further attraction of males and females, thereby helping to regulate the level of infestation and avoiding overexploitation of the tree.

PHEROMONES OF SOCIAL INSECTS

Pheromones mediate many activities of social insects, including defense of the colony, recruitment to food, recognition of individuals and nestmates, and regulation of caste development. The exocrine glands that produce the various pheromones are dispersed throughout the body, as exemplified by those of leafcutter ants (Fig. 5).

Alarm and Defense

Charles Butler was the first to describe the behavioral effects of an insect alarm signal. He recognized in 1609 that the stinger of a honey bee *(Apis)* worker impaled on human skin or clothing attracts more bees to sting that site. The multifunctional role of this signal is shown by the reactions to the same chemical signal of guard honey bees at the entrance to their hive. The presence of an intruder can cause a guard bee to release alarm pheromone from her sting chamber; she disseminates this message into the hive by wing fanning, thereby summoning many bees. These alerted bees seem to be "agitated," with rapid movements and mandibles

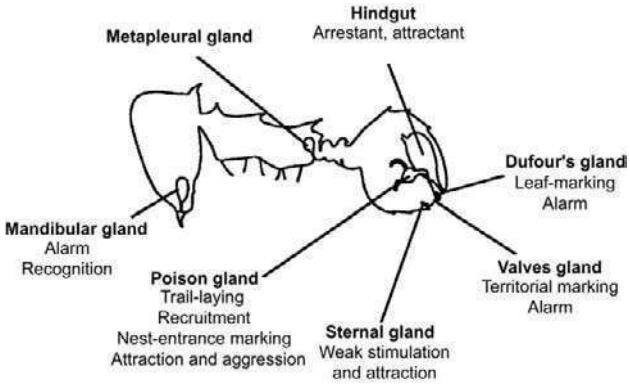

FIGURE 5 Location and function of pheromone-producing glands in leafcutter ants. [After Howse *et al.* (1998).]

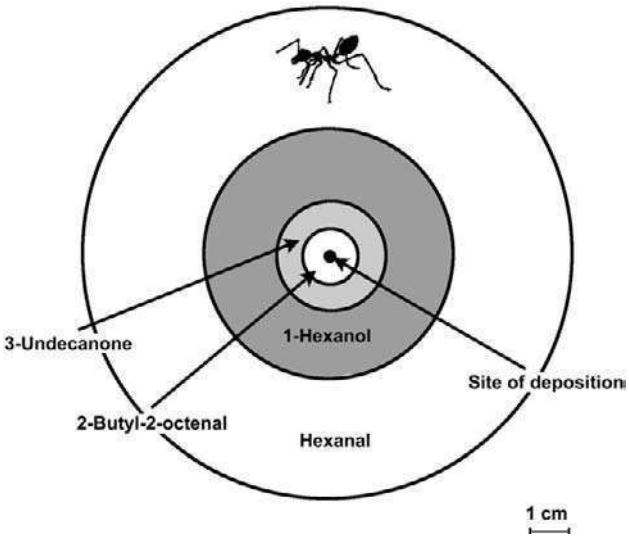

FIGURE 6 Active spaces in still air of the principal components from the mandibular gland of the weaver ant, *Oecophylla longinoda.* The pheromone has been deposited in the center and the boundaries of behavioral activity of each component 20 seconds later are represented. [After Howse *et al.* (1998).]

agape, poised to defend their colony. The sting chamber has more than 20 known pheromone components, most of which induce either alerting or stinging. The other behavioral reactions evoked by components of the sting chamber include lowering the number of foragers departing the hive and repelling foragers that have arrived at a food source. The main constituent of this mixture-evoked alerting and, to a lesser extent, stinging is isopentyl acetate, and the amount per worker changes dramatically with the behavioral task that bees are performing. In the first few days of an adult worker's life, when she is confined to housekeeping and brood-tending tasks inside the hive, essentially no isopentyl acetate is present; the amount rises to 4 to 5 μg at several weeks of age, when she either assumes guard duties or starts foraging outside the hive; for the remainder of her life as a forager, the amount falls to approximately 2 μg per stinger. Stinging itself seems to be released by many components of the sting chamber, including isopentyl acetate, and especially *n*-butyl acetate and 1-pentanol. As with many reactions of social Hymenoptera, the context in which the signal is released is crucial to the kinds of behavior evoked. Context must be taken into account in interpreting behavior and in devising diagnostic behavioral bioassays. For assays of honey bee stinging, for example, one standardized procedure is to provide a vibration of set intensity to the hive, followed by presentation in front of the hive entrance of a swinging target such as a cotton ball or piece of leather containing a candidate alarm pheromone. The number of bees attacking and stinging the target is used to score the level of response.

Defensive behaviors of social insects are quite varied and difficult to categorize into mutually exclusive behaviors. Detection of an alarm pheromone by an ant, for example, can cause it to splay its mandibles, raise its head, bite, and spray an odoriferous and irritating defensive secretion toward a perceived enemy. Unlike attractant and aggregation pheromones, which typically are carried downwind in a turbulent wind flow, alarm pheromones are often released either in relatively still air within the confines of the colony or at ground level, where wind is attenuated. In such situations, molecular diffusion largely or exclusively governs the distribution of pheromone, and the resulting concentration gradient of pheromone supplies potentially useful information about the direction toward the source of alarm. Thus, ants may run toward the source of pheromone (essentially up the odor gradient) or, at lower concentrations farther away from the odor source, movements may seem to be undirected with respect to the odor source. There are many kinds of defensive reactions and, in leafcutter ants, several glandular sources for alarm pheromones (Fig. 5).

An instructive example is provided by the defensive reactions evoked by some of the more than 30 mandibular gland components of the weaver ant, *Oecophylla longinoda,* worked out in considerable detail by John Bradshaw, Philip Howse, and Ray Baker at the University of Southampton. If a droplet of mandibular gland secretion is daubed onto a flat surface in still air, the volatile pheromone diffuses outward at a rate that is dependent on its molecular weight and its diffusion coefficient. A region in which the concentration of pheromone is above the minimum required to elicit a particular behavioral reaction is termed an "active space." The active spaces of each of the four principal components of the mandibular gland secretion 20 s after its deposition are shown in Fig. 6. The sequence of defensive activities seems to be ordered by proximity to the odor source. At the outer limit of the active space, worker ants encounter only hexanal above threshold levels. Ants show heightened levels of running with open mandibles, but their trajectory is not aimed toward the odor's source. Ants that enter the 1-hexanol region, however, move up the odor gradient toward the odor's source. Once they have reached the active space of 3-undecanone, this compound further facilitates orientation and lowers the threshold for biting, as does 2-butyl-2-octenal. Together these

four compounds ensure that the ants are recruited to the site at which the alarm pheromone was released and that they attack an adversary that has been marked with this secretion. Mandibular gland components involved in defense, including other active constituents in addition to the four main constituents, vary within a colony among castes and even between major and minor workers. There also is substantial variation in a given caste among colonies, suggesting that different colonies may have unique defensive codes.

Trail Following and Recruitment

Social insects use trails of varying permanence to exploit food resources and sometimes for colony movement and relocation. E. O. Wilson's exhaustive study of what governs the persistence of the food trail of the fire ant, *Solenopsis invicta,* provides an example of how such systems function. A foraging worker that has encountered a suitable food source lays down a chemical trail by dragging its stinger sporadically along the ground as it returns to the nest. The trail pheromone is a mixture of farnesenes [mainly *(Z,E)*-α-farnesene] from their Dufour's gland, and at any given moment each ant contains only about a nanogram of trail pheromone. The trail from one individual does not persist for long—the active space falls below threshold in less than 2 min, and its effective length is not much more than a meter. *Solenopsis* can even adjust the amount of pheromone deposited on the trail by altering how firmly it drags its sting. The amount of pheromone on the trail is regulated by three factors: the number of ants returning, the proportion of ants laying a trail, and the amount that each ant contributes to the trail. When the food is gone or an ant cannot reach the food source because of other ants, any ant that cannot feed simply does not reinforce the trail.

The number of ants recruited to leave the nest for foraging is a direct function of the quantity of trail pheromone released by a returning forager: to induce nestmates to leave the nest and forage along the trail, the returning forager releases much more pheromone than is found along the trail itself. These simple rules followed by individuals allow mass recruitment, a sophisticated system whereby one group of ants transfers information about the quality of a food source some distance away to another group of ants. The seeming disadvantage of the impermanence of such trails is actually a useful feature of the system that permits ants to match the number of foragers to the quantity of the resource.

Trail communication also is used in relocation of *Solenopsis* colonies. Scouts that have discovered a suitable nest location lay down a trail that other workers follow to the same site. If the location is indeed deemed favorable by new workers inspecting the site, these workers add pheromone to the trail upon their return trip to the nest; this leads to an exponential increase in traffic. Eventually the brood is transferred to the new nest, and the queen follows. Trails close to the nest of some ant species can be relatively permanent, lasting days, and these are called trunk trails. The constituents of the trail pheromones are known for many ant groups, and they are produced from a variety of glandular sources, including the Dufour's, poison, and sternal glands, and the hindgut. Trail pheromones also are widely used by termites in foraging activities. Chemical trails are also important to regulating foraging activities of colonial tent caterpillars (Lasiocampidae). Caterpillars that have located food on distant branches add pheromone to the silken trail on their return trip to the silk nest at which they spend their nonfeeding time. These marked paths are then followed preferentially by future foragers.

Queen Pheromone of the Honey Bee

As emphasized for alarm pheromones, a pheromone can communicate many meanings depending on context. The queen pheromone of the honey bee *(Apis mellifera)* exemplifies this principle, which has been termed "pheromonal parsimony" by Murray Blum of the University of Georgia. The queen pheromone is produced by the queen's mandibular glands and its five known components are 9-oxo-*(E)*-2-decenoic acid, (+)- and (–)-9-hydrox-*(E)*-2-decenoic acid, methyl *p*-hydroxybenzoate, and 4-hydroxy-3-methoxyphenylethanol. The queen produces about 500 μg of this mixture daily, most of which is picked up by the continually changing retinue of a dozen or so workers that constantly groom the queen. Trophallaxis (interchange of food), antennation, and grooming among colony members in turn disperse queen pheromone throughout the colony. The releasing functions of the queen pheromone include the "retinue" behavior (attendance and grooming of the queen—these behaviors require all five components), suppression of construction of new queen cells, and a delay in swarming.

Outside the hive, swarming bees without a queen are attracted to a source of 9-oxo-*(E)*-2-decenoic acid, but they will not form a cluster without the addition of 9-hydrox-*(E)*-2-decenoic acid. Drones (males) are attracted to virgin queens flying 10 or so meters above ground level by her release of 9-oxo-*(E)*-2-decenoic acid. Perhaps the most dramatic effect of the queen pheromone is in its governance of colony productivity. Without queen or queen pheromone in the colony, many workers remain idle. Queen pheromone stimulates comb construction, brood rearing, foraging, and food storage. Artificial application of queen pheromone increases all these activities. The queen pheromone also has a clear primer effect inasmuch as it inhibits development of the workers' ovaries. In the absence of the queen (and the queen pheromone), egg production is triggered in up to one-quarter of the workers, and these "false queens" and other workers in turn produce some queen pheromone.

Pheromones of Honey Bee Workers

Honey bee workers produce diverse messages from a number of exocrine glands. The Nasonov gland (which queens and

drones lack) is situated on the seventh abdominal tergite. A worker exposes this gland by flexing its abdomen, usually while wing fanning and elevating its abdomen. The secretion contains mainly geraniol, geranial and geranic acid, and these influence foraging, marking and, when coupled with queen pheromone, clustering. Nasonov pheromone also is important to swarming. After a swarm has departed the hive, the first workers to arrive at a clustering site expose their Nasonov gland, and the scent attracts other flying workers. Nasonov secretions also are used in "house hunting." A scout that has found a potentially suitable nest site returns to the swarm and communicates direction and distance by the dance language. Scouts release Nasonov pheromone at the site, thereby helping to attract more bees to evaluate its suitability. Nasnonov pheromone also is released by bees fanning at the hive's entry; this odor (probably mixed with odors from the hive) seems to aid disoriented foragers in finding the hive's entrance.

The mandibular gland produces mainly 2-heptanone, which is released by guard bees as an alerting pheromone, and possibly to mark an intruder. This compound, perhaps in combination with other components of the mandibular gland, also may be used to mark flowers that are no longer productive, thereby improving foraging efficiency. Other pheromones labeled "footprint" pheromones mark the nest entrance, and a thermoregulatory pheromone causes nurse bees that are incubating pupae to raise their body temperature by means of muscle contractions. Capping of brood cells is induced by a pheromone consisting of mixture of four fatty acid esters. The examples of pheromonal communication in *A. mellifera* considered here provide a glimpse into the pervasiveness of pheromonal communication in this insect, and the diversity of reactions that can be mediated by pheromonal messages.

Termite Pheromones

The development of castes in termites seems to be governed by complex interactions between juvenile hormone, pheromones, and environmental conditions such as food availability and time of year. For example, in a colony of *Kalotermes* (a "lower termite"), the absence of a king and queen in the colony causes development of replacement (supplementary) reproductives from pseudergate workers, but with the establishment of a reproductive pair (or more), they secrete pheromones that induce pseudergates to eat the excess of reproductives. A queen-produced pheromone inhibits female pseudergates from becoming reproductively competent, and a male-produced pheromone similarly inhibits male pseudergates from becoming reproductive. In the absence of a queen, the king secretes a pheromone that stimulates production of females. The proportion of soldiers in the colony is also regulated by similar interactions. The identity of the pheromones that modulate the proportion of castes in a colony remains unknown.

WHEN PHEROMONES BECOME KAIROMONES

It is also worth noting that parasitoids and predators have coevolved to exploit and manipulate the pheromones of their prey. For example, a group of clerid beetles uses the pheromonal signals of bark beetles to locate and invade the tunnels of their prey in the bark and cambium layers of conifers. Similarly, pentatomid bug species frequently suffer high levels of parasitism from parasitic flies from several families, or from specialist wasp predators. It has been unequivocally demonstrated that these parasitoids use the bugs' pheromones as kairomonal cues to locate hosts. The flies and wasps are attracted specifically to components of their host's pheromone blend. For both the predatory clerids attacking bark beetles and the fly and wasp species attacking pentatomid bugs, the attraction to the pheromones of their prey can be so strong that traps baited with the prey pheromones actually catch more of the parasitoids or predators than the target species.

However, illicit use of the pheromones of prey can go well beyond simply eavesdropping on pheromonal signals. In a fascinating example of coevolution, bolas spiders in the genus *Mastophora* (and several other genera) produce the pheromonal signals of their prey to lure the prey close enough to be caught by a swinging, sticky thread of silk. The prey are male noctuid moths responding to copies of the female pheromone. Even more extraordinary, there is evidence to suggest that within an hour or so, the spiders can change the pheromone lures they produce, to maximally exploit the different times of flight of prey species that respond to differing pheromone blends.

A variety of insects that live inside and sometimes parasitize social insects such as ants and termites also have developed the ability to aggressively mimic the pheromonal signals of their hosts. For example, several staphylinid beetle species live inside termite nests, where they receive all their food from their hosts, and are groomed and cared for by their hosts as though they were indeed termite brood. The chemical cues that both prevent the termites from recognizing the inquilines and induce the feeding and grooming behaviors closely mimic the true pheromones used by the termites for these functions. In an even more aggressive example, the larvae of some syrphid flies are obligate predators on the brood of their ant hosts. The fly larvae produce a blend of cuticular hydrocarbons that closely matches the hydrocarbon profile of the host's brood, effectively camouflaging the fly. The camouflage is so good that if the nest is attacked, the worker ants will carry the fly larvae to safety as though they were indeed ant brood.

APPLICATION OF PHEROMONES IN PEST MANAGEMENT

Insect pheromones have proven useful in pest control. Most of these applications use synthetic copies of pheromones that mediate either attraction or aggregation. Compounds are formulated in protective matrices or reservoirs that emit the

pheromone over weeks or months. Pheromone-baited traps are used to detect exotic invaders, to decide whether pest levels are sufficient to warrant intervention, and to time the application of conventional insecticides or other control measures. For example, the spread of the gypsy moth *(Lymantria dispar)* in the United States is monitored with inexpensive pheromone-baited traps whose sticky internal surface ensnares males. Approximately 350,000 traps are deployed yearly to determine the extent of spread of the European strain of the gypsy moth from the eastern portions of the United States to the Midwest and South, or, especially along the west coast, to signal the occasional invasion of the Asian gypsy moth strain.

Pheromones also are used for direct population control. The tomato pinworm *(Keiferia lycopersicella),* for example, is a devastating pest of tomatoes in Mexico, largely because this moth is highly resistant to insecticides. Application of "cocktails" (mixtures of two or more insecticides) as many as 40 times during a crop cycle may not prevent complete crop loss. Ideally, however, if a tomato field is blanketed with plastic dispensers that release micrograms per hour of synthetic pheromone, the emitted pheromone will interfere with normal mate-finding activities of males, even if the pinworm population is initially at high density. Just how mating disruption works is not fully established, but efficacy likely involves the additive effects of habituation of responsiveness (a presumed central nervous system phenomenon) and some competition for the male's attention between the natural emitters, the females, and the numerous sources of synthetic pheromone. The amount of synthetic pheromone needed to disrupt mate finding is quite small: typical application rates are several grams per hectare per week, and nearly all pheromones are nontoxic and nonpersistent. This technique is now commonly used to control several dozen moth species.

CONCLUSION

Pheromones are a dominant form of communication in most insects, and the messages conveyed serve myriad behavioral and physiological functions. Current microanalytical techniques permit identification of many of these messages, even though they occur in minuscule quantities. Studies expanding our understanding of the chemistry of these messages continue, although in many pheromone systems our ability to characterize the chemicals produced is in advance of our progress in understanding the evoked behaviors, particularly among the complex communication systems of social insects. Current frontiers of investigation in insect pheromones include establishing how genes control biosynthesis, how these chemicals are transduced into an electrical signal in the receptor cells of the responder, and how the signals are processed in the brain, leading to a behavioral output. Only a few structures of primer pheromones have been elucidated, and consequently much remains to be learned about their mode of action.

See Also the Following Articles

Antennae • Caste • Chemoreception • Defensive Behavior • Orientation • Recruitment Communication

Further Reading

Agnosta, W. C. (1992). "Chemical Communication: The Language of Pheromones." Scientific American Library, New York.

Cardé, R. T., and Bell, W. J. (eds.). (1995). "Chemical Ecology of Insects," Vol. 2. Chapman & Hall, London.

Cardé, R. T., and Minks, A. K. (1995). Control of moth pests by mating disruption: Successes and constraints. *Annu. Rev. Entomol.* **40,** 559–585.

Cardé, R. T., and Minks, A. K. (eds.). (1996). "Insect Pheromones. New Directions." Chapman & Hall, New York.

Eisner, T., and Meinwald, J. (eds.). (1996). "Chemical Ecology. The Chemistry of Biotic Interaction." National Academy of Science Press, Washington, DC.

Francke, W., and Schulz, S. (1999). Pheromones. *In* "Comprehensive Natural Products Chemistry," Vol. 8 "Miscellaneous Natural Products Including Marine Natural Products, Pheromones, Plant Hormones, and Aspects of Ecology." (K. Mori, ed.), pp. 197–261. Pergamon Press 5, Amsterdam.

Free, J. B. (1987). "Pheromones of Social Bees." Chapman & Hall, London.

Hansson, B. S. (ed.). (1999). "Insect Olfaction." Springer-Verlag, Berlin.

Hardie, J., and Minks, A. K. (1999). "Pheromones of Non-Lepidopteran Insects Associated with Agricultural Plants." CABI Publishing, Wallingford, Oxon, U.K.

Holldöbler, B., and Wilson, E. O. (1990). "The Ants." Bellknap Press of Harvard University Press, Cambridge, MA.

Howse, P., Stevens I., and Jones, O. (1998). "Insect Pheromones and Their Use in Pest Management." Chapman & Hall, London.

Vander Meer, R. K., Breed, M. D., Winston M. L., and Espelie, K. E. (eds.). (1998). "Pheromone Communication in Social Insects." Westview Press, Boulder, CO.

Phoresy

Marilyn A. Houck
Texas Tech University

Phoresy is a special kind of commensal relationship in which one organism (the phoretic or phoront) attaches to another (the host) for a limited time period to enhance dispersal of the phoront from the natal (or birth) habitat, resulting in colonization of a new and potentially better habitat. In addition to transport, the phoretic host may incidentally provide substrate, shelter, and even some indirect defense or protection for the phoront, but the strict definition of phoresy excludes any direct physiological benefit during transit. For example, the host does not provide the phoretic with food while in transit nor does it contribute to the ontogeny (development) of the phoretic during transit. If feeding does occur, the more appropriate term to describe this relationship would be parasitism. Although phoresy is not a form of parasitism, phoresy can eventually extend into a parasitic association (see below). Alternatively, if the host receives a benefit from its passenger the

relationship is again not phoresy, but a form of mutualism. Thus, this discrete definition of phoresy separates phoresy from all other forms of symbiotic interactions.

The term phoresy is uniformly applied throughout the plant and animal kingdoms and does not exclusively apply to interactions of, or with, insects. Seeds that hitchhike on fur and pants are phoretic. The remora fish (Echeniedea: *Remora remora*) has a dorsal fin modified into a sucker that allows it to attach to the sides of larger fish and turtles, using them for phoretic transport. However, if the remora also snatches pulverized leftovers created during feeding by the host, this act violates the strict definition of phoresy. It soon becomes clear that it can be difficult to distinguish phoresy from other forms of symbiosis. Detailed aspects of behavior, natural history, and physiology are essential to a firm understanding of interspecific associations.

PHORESY AMONG INSECTS AND ARACHNIDS

Phoresy among arthropods has been recognized since at least the mid-1700s. It was formalized and defined by Lesne in 1896 as "those cases in which the transport host serves its passenger as a vehicle." The etymology of the word unites the perspective of both the phoretic (*phor* (Gr.) = thief) and the host (*phoras* (Gr.) = bearing) within the interaction and has a counterpart to most forms of human transport: a "boat" in aquatic phoretic interactions, a "bus" in terrestrial settings, and an "aircraft" in aerial transport. Despite the medium being traversed, there are general principles that seem to be consistent among phoretic relationships, but these are not inclusive and exceptions occur: (1) the phoretic is usually much smaller than its host; (2) there are often several phoretic passengers on any individual host, and mass transit is not uncommon; (3) the phoretic does not have an effective means of independent dispersal (e.g., wings or oars), whereas the host usually is quite mobile; (4) phoresy has played a role in dispersal for a long time and evidence of phoretic relationships can be found in Baltic and Dominican amber from as far back as the early Tertiary (40 mya); (5) phoresy is a response to degradation of an ephemeral habitat (transient habitats available for relatively short periods of time, e.g., beach wrack, dung, etc.) or depletion of a limited resource; (6) phoretics dismount from the host when a suitable new habitat is encountered by the host, indicating that the phoretic has sensory recognition and interpretation of some element in that habitat; (7) usually only one member of a complex life cycle participates in the phoresy and the other members in the life cycle are not phoretic; in such cases, the phoretic may come from among any of the life stages (adult, nymph, or larva); (8) phoresy may be highly coevolved and stenoxenic when both the phoretic and the host are trophic specialists or very indiscriminant, in which case many "buses" may lead to alternate and appropriate habitats; (9) phoresy may be obligate (required) or facultative (occurring under some conditions); (10) phoretics may have little morphological adaptation specific to the attachment to the host (e.g., hold on with mouthparts or clasp with legs) and some have extensive modifications specific to attachment (e.g., extensive sucker plates or highly modified grasping legs); (11) enhanced by wind currents, phoresy is effective across impressive distances and there is even evidence of transoceanic voyages; (12) phoresy may be continual or seasonal, period, or cyclical; (13) the phoretic and the host may come from very different branches of the "tree of life," as divergent as humans and plants, or from within related lineages (e.g., different insect orders); and (14) phoresy has originated independently several times within one lineage, in some instances (e.g., Meloidae or the blister beetles).

Most insect orders have members that participate in phoresy; however, the Diptera and Coleoptera form some of the most extensive phoretic associations with vertebrates, other insects, and mites. They can participate in phoresy as phoronts, as well as phoretic hosts. An interesting example of an insect as a phoretic host is the case of a common phoretic nematode, *Pelodera coarctata,* and its dung beetle host, *Aphodius.* As a dung pat deteriorates and dries, a special resistant phoretic nematode larva is produced that attaches to visiting dung beetles. The phoretic nematodes remain in a dormant state on the beetles until the beetles arrive at a fresh dung pat. Then the nematodes emerge, become active, and begin a new population of free-living nematodes.

Pseudoscorpions (arachnids that looks like small scorpions) are notorious phoronts, found on an impressively large array of insect hosts: Diptera, Hymenoptera, Coleoptera, Odonata, Orthoptera, Heteroptera, Lepidoptera, Trichoptera, harvestmen (Opiliones), spiders, birds, and even small mammals. Pseudoscorpions attach by the chela, or venomous pedipalps, and hold on tightly enough to prevent being brushed off or blown off the host during transit. In one interesting phoretic interaction, a neriid fly that began as a phoretic host for a species of pseudoscoprion then becomes a postdispersal meal at the end of the journey: a case of turning the bus into a lunch wagon at the point of destination.

However, by far the most impressive radiation of phoretic associations occurs among the mites (Fig. 1). Intense selection pressure results in phoresy when organisms are of such extremely small size and they do not possess wings for dispersal. Small body size allows mites to exploit limited ephemeral resources that would be too small to be useful to larger organisms (e.g., a dead snail or nectar within a single flower). Because these resources are small, they degrade quickly and disappear rapidly. And, there may be large distances between them.

To survive, mites must travel to a better resource. Thus, they spend their lives tracking transient habitats. Establishing phoretic relationships with other organisms traveling among the same kinds of habitats gives them more rapid and direct access to a potentially better future and enhances their chances of survival.

Dung beetles, for example, thrive on dung, but as the dung dries and turns to soil it is no longer useful to the dung beetles. As the beetles depart, mites using the dung patties climb on board and hitch a ride to the next site. The journey

FIGURE 1 Example of a mite–beetle phoretic relationship. The mites on the head and body of the *Nicrophorus* beetle are likely *Poecilochirus* (Mesostigmata: Parasitidae), which feed on nematodes in the beetles' nest chamber. (Photographed in Carrer County, Minnesota, by Raphael Carter.)

would be perilous if these soft-bodied mites had to depend on walking the distance to their next meal. As the phoretic association progresses, the mites becomes increasingly dependent on those organisms that provide the most direct route to the best habitats.

EVOLUTION OF PARASITISM FROM PHORESY

Phoresy may begin with unrelated organisms moving independently among shared habitats. A relationship between a phoretic and a potential host can be established when some of the members of the population incidentally and randomly climb on board and are then delivered fortuitously to a habitat suitable for population growth. If encounters are repeated and consistent, mites develop cues to the most productive of these associations. Successful relationships become even more specialized and eventually stenoxenic. In some instances, phoretic association can become an intermediate step that grades into parasitism when the phoretic finds a way to get a meal, as well as transport, from the host.

A well-documented case of a phoretic relationship becoming parasitic is that of the mite *Hemisarcoptes* and the coccinellid beetle *Chilocorus*. Both feed on diaspid scale insects and the association originated as a phoretic relationship. However, coccinellid beetles are reflex bleeders and *Hemisarcoptes* has become adapted to the reflexed alkaloid toxins in the hemolymph and now requires it to molt and complete its development. Because feeding and completion of ontogenesis are part of the contribution of the host, this relationship has graded into parasitism. Other related members of the same mite family (Hemisarcoptidae), which use other phoretic hosts, remain phoretic. This is good evidence that phoresy can be an end point and that it can also progress to other forms of symbiosis (e.g., parasitism). Phoresy thus benefits the individual, but it can also act to enhance the diversity and complexity of community interactions within and among habitats.

See Also the Following Articles

Hypermetamorphosis • Migration

Further Reading

Bologna, M. A., and Pinto, J. D. (2001). Phylogenetic studies of Meloidae (Coleoptera), with emphasis on the evolution of phoresy. *Syst. Entomol.* **26,** 33–72.

Holte, A. E., Houck, M. A., and Collie, N. L. (2001). Potential role of parasitism in the evolution of mutualism in astigmatid mites: *Hemisarcoptes* as a model. *Exp. Appl. Acarol.* **1720,** 1–11.

Houck, M. A., and O'Connor, B. M. (1991). Ecological and evolutionary significance of phoresy in the Astigmata. *Annu. Rev. Entomol.* **36,** 611–636.

Klepzig, K. D., Moser, J. C., Lombardero, F. J., Hofstetter, R. W., and Ayres, M. P. (2001). Symbiosis and competition: Complex interactions among beetles, fungi, and mites. *Symbiosis* **30,** 83–96.

Poinar, G. O., Ćurčić, B. P. M., and Cokendolpher, J. (1998). Arthropod phoresy involving pseudoscorpions in the past and present. *Acta Arachnol.* **47,** 79–96.

Schwarz, H. H., and Huck, K. (1997). Phoretic mites use of flowers to transfer between foraging bumblebees. *Insectes Soc.* **44,** 303–310.

Zeh, D. W., and Zeh, J. A. (1992). On the function of harlequin beetle-riding in the pseudoscorpion, *Cordylochernes scorpioides* (Pseudoscorpionida: Chernietidae). *J. Arachnol.* **20,** 47–51.

Photography of Insects

Mark W. Moffett

University of California, Berkeley

Quality photography of small animals in the field depends on rapid response to opportunities; this requisite is often impeded by the need to constantly root around in overstocked camera bags for the "right" piece of equipment. This article describes the author's personal approach to insect photography, which is to improvise with as little gear as possible. If it doesn't fit comfortably in your backpack, why have it at all? This philosophy may not be relevant to the intended audience of most insect photography essays, namely, the museum curator, studio photographer, artist, or laboratory scientist, who have time to "set up" a picture. Specific brands are not considered here. In part this is because the market is in flux with the ongoing technological shift to autofocus and digital cameras and, in part, because the quality of lenses produced by any of the major camera lines today is sufficient to produce excellent photographs. Expensive gear is not a necessity for the best results.

For all but the largest insects or insect-produced structures such as nests, capturing insects on film typically requires macrophotography, which is photography at a magnification of 1:1 ("life size") or greater, which is the focus of this article. At 1:1, subjects are the same size on the film as they are in life. That is, the image of a 30-mm-long beetle measures 30 mm in length on the film negative or slide, which means it fills

most of the frame in a photograph on 35-mm film. For a 3-mm beetle to appear just as large (i.e., so that it is likewise 30 mm in length on the film) requires 10 times life-size magnification, which is typically described as "×10," or 10:1. The measurements are taken from the original film exposed in the camera, so prints made from a negative or slide would enlarge the subject beyond this size. For example, a 35-cm-wide print of the 3-mm-long beetle at ×10 on 35-mm film shows the beetle 30 cm in length, or 100 times its natural size.

One of the wonders of a well-executed insect photograph is that the subject's original size is forgotten: the small beetle is just as imposing as the large one. This actually makes photography a wonderful medium for entomologists. Not realizing this, photographers often concentrate on insects that are impressive merely for their size. In fact, when not photographed on the photographer's hand, many large insects can look small on film. Similarly, elephants may be spectacular in life, but on film we can only make one appear *smaller* than it really is. I recall my mentor at National Geographic, editor Mary G. Smith, taking her first look at my first "macro" images in 1986 (Fig. 1), which were photographs of marauder ants killing prey. She exclaimed that they reminded her of the movie "Terminator," even though the ants looming over doomed prey in my pictures were just a few millimeters long. A criterion for a good insect photograph is that viewers are as surprised as Mary when a subject's size is revealed.

FIGURE 1 An 8-mm-long major worker ("soldier") marauder ant, *Pheidologeton diversus,* killing a centipede that has been pinned down by numerous smaller minor worker ants in Malaysia.

LENSES

Although regular lenses can be used in combination with other equipment to produce images of life size or greater, for example, by adding extension tubes or magnifying filters, the results can be inferior to the photographs made with standard "macro lenses." Although the term "macro" has been watered down in recent years by its application to lenses that focus relatively closely to the subject, true macro lenses focus all the way to 1:2 or 1:1 (sometimes an extension tube may be required for 1:1). These macro lenses are the sharpest, most optically corrected lenses produced by most companies and thus can be a wise investment. Many of them (typically those with focal lengths of about 50, 100, or sometimes 200 mm) can also be used as regular lenses, in which case they completely replace any other lens of a similar focal length. It should be noted that some of the optical precision of macro lenses relates to their having a "flat field" of focus. This characteristic is seldom important when working with live insects, unless perhaps one is in the habit of photographing straight down on flat insects living on the kitchen floor. Thus, when used with care, some less expensive lenses may be just as good in the field as true macro lenses, even when used at magnifications as high as 1:1.

A constant problem in insect photography is the distance from the front of the lens to the subject. With a macro lens set to a high magnification this distance may be only a few millimeters, so it takes practice not to disturb the insect when preparing to take a picture. In this regard stalking an insect is little different from stalking a leopard or a deer. As one moves the camera until the quarry comes into focus, one can learn to recognize through the viewfinder the moment when an insect detects the photographer. Each insect can require a different stalking technique.

If a short working distance causes the subject to be knocked or startled, the photographer should shift to a longer focal-length lens. A 200-mm macro lens provides a much greater working distance than a 50-mm macro set to the same magnification. Yet the longer lens is likely not to be quite as sharp, is harder to hold steady enough to focus precisely, requires heftier (and more unwieldy) brackets to hold the flashes, and needs more extension tubing to achieve magnifications beyond 1:1. If one can handle its short working distances, a 50-mm lens is therefore the ideal macro, but the 100-mm lens comes a very close second.

To achieve magnifications beyond life size with standard macro lenses, extension tubes are placed between a lens and the camera mount. Bellows are a more flexible alternative to tubes, because they provide every conceivable magnification between some upper and lower limit, but they are unwieldy and fragile

FIGURE 2 Soldiers of the aphid *Pseudoregma* jamming their needle-like "horns" into the cuticle of a syrphid fly larva that was attacking their colony in Japan. Photographed at ×14 magnification.

in the field. The fixed lengths of the tubes are seldom a liability outdoors where being able to quickly frame and shoot insects is more important than precise cropping. In fact, tubes shorter than about 25 mm generally have little value with macro lenses because their effect on magnification is so slight.

In addition to standard macro lenses some manufacturers offer lenses designed exclusively for macrophotography (for example, the ×1 to ×5 zoom lens made by Canon). Most of these lenses are built for high magnifications. An article by the author on soldier aphids in 1989 (Fig. 2) contains images between ×5 and ×20 on the film taken with a hand-held camera in the field. Whereas a standard macro lens can be used with only two or three extension tubes before working distance becomes too small to be practical or image quality declines, with some of these specialized lenses it may be possible to use several tubes at once.

Rotating the lens barrel either manually or with autofocus accomplishes focusing in normal photography, but it changes the magnification of an image in macrophotography, making it difficult to achieve the desired results. Autofocus therefore should be disengaged. Instead, focusing must be done as a manual, multistep process. Begin by deciding what magnification is desired, based on a subject's size. Select the lens and extension tubes needed to achieve that magnification. Then aim the camera toward the subject, rocking slowly in and out until the subject appears in focus.

Photographs at the highest magnifications require a steady hand, a sharp eye, and a lot of practice. Each photographer needs to understand his or her own limits from experience. The problem is not only in seeing and composing the shot but also in the limited depth of field at higher magnifications. The depth of field is the depth (from front to back) that appears in focus within an image. All the significant parts of the image must fall within this area for a photograph to make sense. At times the author has used a half meter of extension tubing in the field, and in these situations the depth of field can be the length of a paramecium! To increase the chance of success, take a few images after focusing on the subject by rocking back and forth slightly to subtly change the plane of focus each time.

As in all high-magnification photography, automatic film advance comes in handy not so much to allow rapid-fire picture taking but to allow one to keep a subject correctly framed and in focus while these frame-to-frame shifts in the film plane are made. Another aid is to use a flashlight, even at midday (illuminate crevices and other small shadows or the shadow of the lens itself). Using Velcro, tape, or glue, attach a small penlight to either the lens or one of the flashes so that it can be aimed at the subject. Camping stores have many flashlights that can work. It is best to bring in a camera and try out specific designs.

FLASHES

It is possible, but seldom advisable, to photograph a large insect with natural light using a tripod. With sufficient photographic skill, any image of an insect can be improved by the addition of flashes. Flashes provide a consistent and high quality of light, though flashes with colder (bluer) light can be improved by leaving a "warming gel," like a Kodak CC10Y filter, taped over the flash window (one layer of frosted scotch tape can by itself sometimes do the trick). Flashes freeze any motion of the subject or camera—including most importantly any motion caused by the trembling of the photographer's hands—and do so far more effectively than a tripod (particularly at high magnification). Furthermore, in the time it takes to set up a tripod, most insect subjects will have moved on to greener pastures. Flashes allow one to move in quickly and to constantly adjust the angle of approach as the action unfolds. For these reasons, the author seldom carries a tripod and even then never for macro work. Nonetheless, views of large insects in their environment, often at a high aperature with a wide-angle lens focused on a close subject, can be taken using natural light and a tripod, though with ground- or bark-dwelling species the camera may be steadied sufficiently by pressing it firmly against the stable substrate.

Not only can flash light increase the quality of any insect photograph, but when used with skill, two flashes are always an improvement over one flash, and two regular flash units are always an improvement over any method using a ring flash. The second flash serves as a "fill light," that is, it is weaker (usually one-half the strength or one stop weaker) than the

other, "main flash." A fill flash results in photographs with nicely defined, pleasant shadows and lots of three-dimensional information about the subject. The fill flash can be a weaker flash model, but often it is the same kind of flash placed at the same distance as the primary flash, but set at a lower power output or filtered to reduce the light (a sheet of tissue or artist's tracing paper over the fill-flash head can do the trick). The flashes should be aimed toward the subject at about 45° from the axis of the lens and positioned anywhere from 90 to 120° apart around the lens barrel. For insects on reflective surfaces such as flowers and bright leaves, the light bounces off the surfaces and "fills" shadows somewhat, reducing the necessity of a fill flash. For shooting under these conditions, a single flash method as described by Shaw may suffice.

Ring flashes present problems with insects. The lower part of the ring impedes one from shooting low to the ground, blocking the most dramatic views. Also, the ring projects forward beyond the front of the lens all around the circumference, reducing the working distance. Further, because their light circles the lens and is directed forward rather than angled at the insect, ring flashes reduce the image's three-dimensional content, making a subject look relatively shapeless and flat (ring lights were developed for flat objects such as stamps). This problem can be partially remedied by taping over parts of the ring. On the positive side, ring flashes provide good color saturation for photographers unsure of their photographic technique, especially when they are used in combination with a fill light.

The greatest stumbling block to great macrophotography is the scarcity of good two-flash systems. Some flash brackets have been marketed, but most are bulky (making it hard to photograph in the tight corners, where many insects dwell), are difficult to hold steady for long periods because of their weight, and are easily knocked out of position in the field. The author makes his own systems, from various flashes, power packs, brackets, cords, and slaves. Because flashes can be used in their nonautomatic (manual) settings, items from various camera brands can be mixed to achieve the lightest, most compact results with the flashes at optimal positions. Camera brands must be mixed with care, for example, by taping over flash shoe connections if flashes of one brand are used with a camera body of another brand. No design is perfect, but some are better than others, and all of them can be modified.

The position of the flashes is most critical. When a small flash head is even a few centimeters away from the insect, viewed from the subject's position it appears as a point of light—much as the sun (although it is in fact a huge disc) appears as a point in the sky because of its distance from us. This kind of "point source" of light causes the most intensely illuminated spots to be "burned out" (too bright) and casts deep shadows that no film can handle (as occurs with any sunlit object on a cloudless day). To avoid such problems flashes should be placed as close to the subject as comfortably possible. So positioned, the lights resemble the light boxes used in studio photography for portraits. What is desired, in fact, is to replicate such a studio in miniature. One can even add a third light traditional in portrait studios, the "hair light," that is aimed from behind to define the edges of the subject, but for most field photographers this light is often unnecessary.

SHARPNESS AND EXPOSURE

Achieving a sharp, correctly exposed macro image usually takes knowledge that comes from testing the system being used. Exposure tables and other textbook information are seldom accurate and are no substitute for judging for oneself what is pleasing. Tests that are done carefully the first time will never have to be done again—good results are guaranteed.

Light meters are not always effective with many macro scenes, which often include objects that can be either black or brightly lit depending on slight changes in camera position relative to the subject and its background. For this reason, manual exposure techniques provide greater accuracy, but if a camera meter is preferred, these problems in a scene must be recognized and exposure bracketed accordingly. The best flashes for manual exposure work have multiple manual settings, that is, full power, half-power, quarter-power, and so on.

To test both the flashes and the lens and to develop a technique, select a fine-grained slide film (many with an ISO of 100 or less work well). Slides allow one to accurately gauge the exposure and sharpness of the images. As a test subject, put a dead insect of a kind likely to be photographed on an 18% gray card (available at most major photography stores). Then take a series of test photographs, recording magnification, flash position, flash power, and f-stop for each frame of the film as follows:

1. Set the lens so that it gives a certain magnification (say, 1:1) or has a certain number of tubes that can be remembered (say, one 25-mm tube).
2. Select the power settings of the two flashes (perhaps put one on full power and the fill flash on half-power).
3. Take a series of photographs of the insect on the gray card at a standard angle (say about 45° from the horizontal), the first at the lens' minimum aperture (say, f32) and then at one-f-stop intervals below that, down to f8.
4. Select another set of power settings for the flashes, say making both of them one stop weaker (that is, one at half-power and the other at quarter-power).
5. Repeat steps 2 and 3 with weaker flash power settings.

The developed slides should be checked for image detail (such as the texture of the gray card and the sculpturing on the insect) and exposure (have the original gray card on hand to see if the brightness of the slide matches the gray of the card). Pick the combination of flash powers and f-stop that gives the most pleasing result (see below). That result might be, for example, f22 with the flashes set at half- and quarter-power (but if f22 looks a bit too dark and f16 looks too light, record the intermediate setting as correct, i.e., f18). Write the settings down and use them thereafter for that magnification

(although, as in any photographic situation, one can bracket slightly, for example, by opening up to f16 when the subject or its background is very dark). If it turns out the test photographs are all dark, move the flashes closer or purchase stronger units. If they are all too bright (or if the correct exposure occurs at an f-stop that has too little depth of field, as is explained below), weaken both flashes and test again.

Results from the first magnification can be used as the starting point for testing other magnifications, because results tend not to differ radically from one magnification to the next. For example, try a series of magnifications such as ×2, ×3, and ×6 (or, if preferred, the same lens with 50, 100, and 200 mm of lens tubing). After working out the correct exposure for each magnification, write up an exposure table and tape it to the back of the flash heads for reference.

Most people choose images that are overexposed (i.e., too bright), which for macrophotography means that valuable flash battery power has been wasted in producing too much light. If in doubt, choose a slightly darker image over a slightly bright one. To correctly judge exposure and image quality use a color-corrected (5500K) light table and a photographer's loupe. For slide film the highlights (bright parts of the image) should not be entirely burnt out except perhaps for tiny areas, and the colors over most of the image should be saturated (richly hued and not faded by strong light). Meanwhile, the darkest parts of an image should be inky black, but not so much as to lose detail by rendering an image blotchy. If burnt areas or blotchiness occur, try repositioning the flashes.

The tests may show the classic trade-off in macrophotography between depth of field and image quality. Thus, even though most macro lenses tend to close down to 32 and this f-stop provides the most depth of field, to produce sharp images it is best to open the aperture at least one stop from this setting (in this example, f22). The best quality—highest resolution and contrast—may actually occur at a stop lower that that (e.g., at f16), but the difference may be marginal enough so that the best choice is to use f22 because of the greater depth of field it provides. If a lot of extension tubing is added to a lens, image quality for the same f-stop setting on the lens may drop further, perhaps to an aperture of f11. This adjustment is a problem because the depth of field also declines as extension tubes are added or magnification is otherwise increased. Therefore the highest magnification attainable by a lens depends on the accuracy of the photographer with focusing and the optical limits of that lens. At some point it becomes necessary to purchase a lens better designed for the magnification in question.

TECHNIQUE

The most difficult subjects require considerable patience and a lot of time—sometimes a hundred attempts for every usable image. Ways of improving one's chances can be found, such as having on hand a supply of food items in photographing predation, but such techniques must be used with care. Usually, the best image results from capturing the animal in the act of normal behavior in its normal habitat. This image is "best" not necessarily because of its technical perfection (on the contrary, the gritty realism of a slightly imperfect image sometimes enhances its drama, as is often true in photojournalism), but because of its accuracy. For example, a knowledgeable person might detect that the prey provided to the subject for a photograph is not a species that it normally would find and catch. Even more egregious is the refrigeration of specimens to slow them down for a picture. Despite their stiff exoskeletons, insects express themselves by subtle postures and actions (see Fig. 1). To the expert, a staged picture of a chilled insect appears as unnatural as one of a frozen human being.

With time and experience, one may want to attempt a photoessay or lecture on a particular subject. To hold a viewer's interest, try to incorporate a variety of compositions and magnifications. A critical overview of nature photojournalism was provided by the author in a 1995 article.

See Also the Following Articles

Collection and Preservation • Cultural Entomology • Museums and Display Collections

Further Reading

Constant, A. R. (2000). "Close-Up Photography." Focal Press, Woburn, MA.

Davies, P. W. (1998). "The Complete Guide to Close-Up and Macro Photography." David & Charles, Devon.

Moffett, M. W. (1986). Marauders of the jungle floor. *Natl. Geographic* **170,** 272–286.

Moffett, M. W. (1989). Samurai aphids: Survival under siege. *Natl. Geographic* **176,** 406–422.

Moffett, M. W. (1995). Essay. *In* "Magnificent Moments: The World's Greatest Wildlife Photographs" (G. H. Harrison, ed.), pp. 106–111. Willow Creek Press, Minocqua, WI.

Shaw, J. (2000). "Nature Photography Field Guide." Amphoto, New York.

Sholik, S., and Eggers, R. (2000). "Macro and Close-Up Photography Handbook." Amherst Media, Buffalo, NY.

Phthiraptera
(Chewing and Sucking Lice)

Ronald A. Hellenthal
University of Notre Dame

Roger D. Price
University of Minnesota

Phthiraptera are obligatory, lifelong ectoparasites of birds and mammals. They are hemimetabolous and wingless, with dorsoventrally flattened bodies, three pairs of well-developed legs, and a single- or double-segmented tarsus usually with one or two claws, but occasionally with claws

absent. Ocelli are absent, eyes are reduced or absent, and antennae are short with three to five segments. Body length of adults ranges from 0.3 to 12 mm.

EVOLUTION AND DIVERSITY

Over 4900 louse species are recognized. Lice formerly were classified in two orders, Mallophaga (chewing lice) and Anoplura (sucking lice). However, they now are combined into the order Phthiraptera, with the former Mallophaga comprising three suborders (Amblycera, Ischnocera, and Rhynchophthirina) and the Anoplura a fourth. Although the suborders show substantial differences, they are believed to form a hemipteroid monophyletic unit whose sister group is the Psocoptera (psocids or booklice). Chewing lice have mandibulate mouthparts and probably evolved on birds. They are thought to have fed initially on skin and feathers, with some groups ultimately expanding their diets to include tissue fluids and blood. Some chewing lice eventually made a transition from birds to mammals and some Ricinidae (Amblycera) have mouthparts adapted to pierce host skin.

Sucking lice are restricted to mammalian hosts and have unusual piercing–sucking mouthparts that consist of three stylets, probably derived from the fused maxillae, hypopharynx, and labium. The stylets, retracted into the head when not in use, are everted to pierce skin and feed on host blood. Salivary secretions injected into the host include an anticoagulant to prevent blood coagulation during feeding.

Lice are recorded from approximately 5000 bird and mammal species, encompassing fewer than 30% of mammal species and about 36% of bird species. As many as 50% of the extant louse species remain undescribed. Lice are found on all bird orders but are not known to be associated with the mammalian order Monotremata (monotremes), the marsupial orders Microbiotheria (the monito del monte) and Notoryctemorhpia (marsupial "moles"), Chiroptera (bats), Cetacea (whales, dolphins, and porpoises), Sirenia (dugong, sea cow, and manatees), or Pholidota (scaly anteaters).

LIFE HISTORY

Details of louse life history are known for only a relatively few species, especially those of economic importance. Eggs (also known as nits) typically are cemented to hairs or feathers of the host and all life stages are confined to a single host. Some notable exceptions to this egg-laying behavior include the human body louse, *Pediculus humanus humanus* (Anoplura: Pediculidae), which attaches its eggs to clothing fiber and spends most of its time between feedings on clothing rather than on the host itself, and several genera in the chewing louse family Menoponidae (Amblycera) that spend their entire life cycles and deposit eggs inside the quills of primary and secondary feathers.

Immatures emerge from eggs by exerting pressure on an operculum at the free end of the egg. The timing and duration of the egg, three instars, and adult stages vary among species and may be affected by environmental temperature and humidity. Nymphs of the human head louse *(P. humanus capitis)* take from 6 to 9 days to hatch and reach the adult stage in about 1 week. Adults remaining on the host live about 1 month, with each female laying as many as eight eggs per day. Immatures and adults normally feed twice each day and cannot survive off of the host for more than a few days.

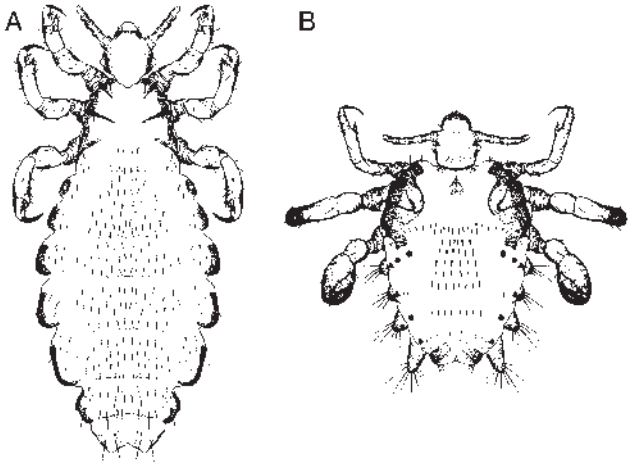

FIGURE 1 Lice of humans (dorsal view of adult females). (A) Head louse, *P. humanus capitis* (Anoplura: Pediculidae), adult body length 2.1–3.3 mm (body lice look similar to head lice, but tend to be slightly larger). (B) Crab louse, *P. pubis* (Anoplura: Pthiridae), average adult body length about 1.75 mm for females and 1.25 mm for males. (Illustrations adapted from Ferris, 1951, courtesy of the Pacific Coast Entomological Society and Stanford University Press.)

Dispersal of lice typically requires close contact between hosts, although immature lice have been found attached to flying insects that move between hosts, and phoresy may play a role in the dispersal of some species. The human body louse is dispersed by sharing infested clothing, the head louse (Fig. 1A) by personal contact or sharing combs and brushes, and the crab louse, *Pthirus pubis* (Anoplura: Pthiridae) (Fig. 1B), by sexual contact or, less commonly, through infested bed linens, towels, or clothing.

LOUSE–HOST SPECIFICITY AND COSPECIATION

Many lice show a high degree of host specificity, causing them to play an important role in studies of host–parasite cospeciation. For example, the 36 species of pocket gophers (Rodentia: Geomyidae) are parasitized by 122 species and subspecies of *Geomydoecus* and *Thomomydoecus* chewing lice (Ischnocera: Trichodectidae), many of which have distributions that conform closely to host–subspecies groups that are genetically similar. However, host specificity is not universal because many louse species occur on more than a single host species. *Anatoecus dentatus* and *A. icterodes* in the Philopteridae (Ischnocera), for example, each occur on more than 60 different species of ducks, geese, and swans (Anseriformes: Anatidae).

Many host species are parasitized by more than one louse species. For example, a single subspecies of brown tinamou, *Crypturellus obsoletus punensis* (Tinamiformes: Tinamidae), is infested by 11 species of lice, representing 10 genera in two families. When multiple louse species infest a single host individual, the different species may congregate in different areas of the body. In mammalian lice, habitat specialization often is associated with hair diameter. For example, the crab louse tends to be confined to the relatively coarse hairs associated with the genital areas, the face, and the underarms, whereas the head louse tends to be associated with the smaller diameter hairs of the scalp.

ECONOMIC IMPORTANCE

The economic importance of most louse species is unknown, although they are known to cause irritation, inflammation, and itching and serve as vectors of diseases and other parasites. High louse densities often are found on weakened or sick hosts, especially on birds with damaged bills or those unable to perform normal grooming activities. Lice infest a number of domestic animals including poultry, domestic dogs and cats, cattle, sheep, goats, horses, pigs, and rabbits. Animals in zoos and laboratory colonies of rats and mice also may become infested. Economic loss to the livestock and poultry industry in the United States caused by lice has been estimated at more than $550,000,000 per year, with about two-thirds of this resulting from weight and egg production losses in poultry caused by chewing louse infestations. More than $350,000,000 is spent annually on control of lice on humans.

The human body louse serves as the vector of epidemic typhus and epidemic relapsing fever and as an occasional vector of murine typhus. Epidemic louse-borne typhus, a rickettsial disease caused by *Rickettsia prowazekii,* is an important scourge of humans often associated with wars and disasters. It still is endemic in cold areas of Africa, Asia, and Central and South America. Zinsser gives a remarkable account of the impact of this disease on human history. For example, the disastrous failure of Napoleon's Grand Armée to conquer Russia during the campaign of 1812 is attributed largely to the effects of epidemic louse-borne typhus fever combined with malnutrition, dysentery, and exposure. The effectiveness of the pesticide DDT was demonstrated in 1943 during World War II in the control of lice responsible for an outbreak of typhus fever after the Allied bombing of Naples, Italy.

The name typhus is derived from Greek *typhos,* meaning stupor or fever. Infected humans experience high fever often accompanied by severe headaches, mental confusion, chills, coughing, and muscular pain. A rash generally appears on the 5th to 6th day and usually spreads to much of the body except the face, palms, and soles of the feet. The mortality rate usually is 1 to 20%, but can be much greater under epidemic conditions in populations that are weakened by malnutrition or other diseases. Transmission is by contamination of wounds with louse feces rather than through the bite or feeding of the louse. Brill–Zinsser disease is a relatively mild form of typhus that can remain viable in humans for years and can serve as a reservoir of the disease and the source of future epidemics. Although not associated with disease transmission, outbreaks of the human head louse are epidemic throughout the world, with treatment complicated by the development of louse resistance to some commonly used chemical control agents.

CLASSIFICATION AND HOST ASSOCIATIONS

Amblycera is regarded as the most primitive louse suborder and the Rhynchophthirina and Anoplura are thought to be most advanced. The suborder Amblycera (Fig. 2A) includes 6 families: Boopiidae (mostly on marsupials, with one species infesting dogs and another the cassowary), Gyropidae (mostly on rodents, but with one genus occurring on New World monkeys and another on peccaries), Laemobothriidae (found on six bird orders), Menoponidae (widely distributed on birds), Ricinidae (on passerines and hummingbirds), and Trimenoponidae (on marsupials and rodents). The suborder Ischnocera (Fig. 2B) includes 2 families: Philopteridae (widely distributed on birds, with two species on primates, one on lemurs and the other on the indri) and Trichodectidae (on seven mammalian orders). The suborder Rhynchophthirina (Fig. 2C) contains a single family, Haematomyzidae, found on elephants, wart hogs, and the red river hog. The suborder Anoplura (Fig. 2D) includes 15 families, all restricted to mammalian hosts: Echinophthiriidae (on seals, sea lions, the walrus, and the North American river otter), Enderleinellidae (on squirrels), Haematopinidae (on horses and their relatives, pigs, cattle, and deer), Hamophthiriidae (on flying lemurs), Hoplopleuridae (on 8 families of rodents, a few shrews, and one species of pika), Hybophthiridae (on the aardvark), Linognathidae (on deer, cattle and their relatives, camels, and dogs and their relatives), Microthoraciidae (on the llama, alpaca, guanaco, and dromedary), Neolinognathidae (on elephant shrews), Pecaroecidae (on peccaries), Pedicinidae (on Old World monkeys), Pediculidae (on humans, chimpanzees, and New World monkeys), Polyplacidae (on 13 families of rodents, shrews, tree shrews, hares, and 5 families of primates), Pthiridae (on humans and gorillas), and Ratemiidae (on horses and their relatives).

There is disagreement as to whether the human head louse should be recognized as a distinct species *(Pediculus capitis)* or as a subspecies of *P. humanus (P. humanus capitis).* Those favoring its recognition as a separate species cite differences in behavior, size, coloration, and ability to transmit disease. Those favoring subspecies recognition note that populations of head and body lice are separable only statistically on the basis of minor differences in body size; that coloration in human lice is highly variable, with lice often taking on the color of their surroundings; and that hybridization between head and body lice has been demonstrated in the laboratory and is thought to occur in nature.

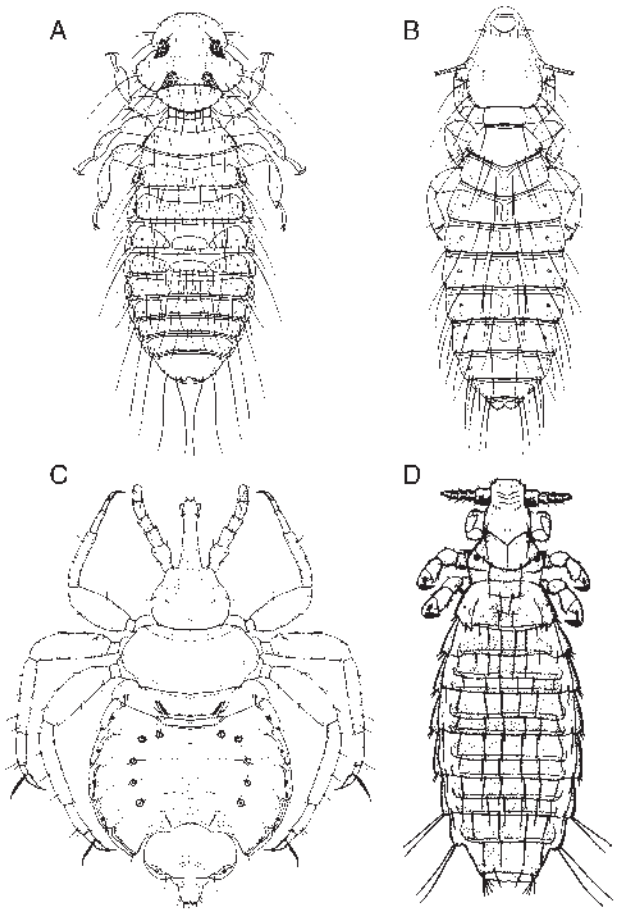

FIGURE 2 Louse suborders (dorsal view of adults). (A) Amblycera: *Colpocephalum fregili* (Menoponidae), male from the red-billed chough *(Pyrrhocorax pyrrhocorax).* (Adapted from Price and Beer, 1965, *Proc. Entomol. Soc. Wash.* **67,** 7–14.) (B) Ischnocera: *Quadraceps crassipedalis* (Philopteridae), female from the least seed-snipe *(Thinocorus rumicivorus).* (Adapted from Emerson and Price, 1985, *Proc. Entomol. Soc. Wash.* **87,** 395–401.) (C) Rhynchophthirina: *Haematomyzus porci* (Haematomyzidae), male from the red river hog *(Potamochoerus porcus).* (Adapted from Emerson and Price, 1988, *Proc. Entomol. Soc. Wash.* **90,** 338–342.) (D) Anoplura: *Polyplax spinulosa* (Polyplacidae), female of the spiny rat louse from the Norway rat *(Rattus norvegicus).* (Adapted from Kellogg and Ferris, 1915, Leland Stanford Junior University Publication, courtesy of Stanford University Press.)

See Also the Following Articles

Lice, Human • Medical Entomology • Psocoptera • Salivary Glands • Veterinary Entomology

Further Reading

Durden, L. A., and Musser, G. G. (1994). The sucking lice (Insecta, Anoplura) of the world: A taxonomic checklist with records of mammalian hosts and geographical distributions. *Bull. Am. Mus. Nat. Hist.* **218,** 1–90.

Ferris, G. F. (1951). The sucking lice. *Mem. Pacific Coast Entomol. Soc.* **1,** 1–320.

Hellenthal, R. A., and Price, R. D. (1991). Biosystematics of the chewing lice of pocket gophers. *Annu. Rev. Entomol.* **36,** 185–203.

Hopkins, G. H. E., and Clay, T. (1952). "A Check List of the Genera & Species of Mallophaga." Br. Mus. (Nat. Hist.), London.

Kim, K. C., Pratt, H. D., and Stojanovich, C. J. (1986). "The Sucking Lice of North America: An Illustrated Manual for Identification." Pennsylvania State University Press, College Park, PA.

Peterson, R. K. D. (1995). Insects, disease, and military history: The Napoleonic campaigns and historical perceptions. *Am. Entomol.* **41,** 147–160.

Price, M. A., and Graham, O. H. (1997). "Chewing and Sucking Lice as Parasites of Mammals and Birds." USDA Technical Bulletin No. 1849.

Zinsser, H. (1934). "Rats, Lice and History: A Study in Biography." Little, Brown, Boston.

Phylogeny of Insects

Peter S. Cranston and Penny J. Gullan
University of California, Davis

Ideas concerning the phylogenetic relationships among the major taxa of arthropods, and the included insects, are dynamic. Although there is a single evolutionary history, efforts to uncover this phylogeny vary between different researchers, techniques, and character systems studied. No technique or character system alone can guarantee that it reveals the true relationships of the studied taxa; in actuality, convergent (homoplastic) similarity that confuses relationships is common to all data. The evidence behind traditional systems, representing perhaps the thorough understanding of a single character system rather than an integration of all knowledge, sometimes cannot withstand detailed scrutiny. Molecular sequence data often appear to overturn previous ideas derived from morphological interpretation, but may be misleading because of undersampling, unrecognized sampling of alternative gene duplicates (paralogs), and/or inappropriate analyses. In this article, the different sources of evidence for the phylogenies that we have chosen to portray are assessed critically. Well-founded and less well founded traditional, even refuted, relationships are discussed, and where resolution appears to be lacking, this lack is identified.

RELATIONSHIPS OF THE HEXAPODA TO OTHER ARTHROPODA

Insects belong to arguably the most successful major lineage of the phylum Arthropoda, the joint-legged animals. This clade comprises myriapods (centipedes, millipedes, and their relatives), chelicerates (horseshoe crabs and arachnids), crustaceans (crabs, shrimps, and relatives), and hexapods (the 6-legged arthropods, Insecta and their relatives). Lobopods (onychophorans) sometimes have been included, but now almost universally are considered to lie among likely sister groups outside Arthropoda. Although traditionally each major arthropod group has been considered monophyletic, most have been suspected of nonmonophyly by at least a few investigators. Results of molecular analyses have provided frequent challenges, particularly in suggesting the possible paraphyly of myriapods and of crustaceans. Even if considered monophyletic, estimation of interrelationships has been a most

Specialized Terms

apomorphy (-ic) A feature of an organism in the derived state, contrasted with an alternative one in the ancestral (primitive) state—a **plesiomorphy** (-ic). For example, with the character of forewing development, the sclerotized elytron is an apomorphy for Coleoptera, and the alternative, a conventional flying forewing, is a plesiomorphy at this level of comparison.

cladogram Diagramatic illustration of the branching sequence of purported relationships of organisms, based on distribution of shared derived features (synapomorphies).

monophyletic Referring to a taxonomic group (called a clade) that contains all descendants derived from a single ancestor and recognized by the possession of a shared derived feature(s). For example, the clade Diptera is monophyletic, recognized by shared derived development of the hind wing as a haltere (balancing organ).

paraphyletic Referring to a taxonomic group (called a grade) derived from a single ancestor but not containing all descendants; grades share ancestral features (e.g., Mecoptera relative to Siphonaptera).

polyphyletic Referring to a taxonomic group derived from more than one ancestor and recognized by the possession of one or more features evolved convergently. For example, if the primitively wingless silverfish were united with secondarily wingless grasshoppers, beetles, and flies, the resulting group would be polyphyletic.

sister groups Species or monophyletic groups that arose from the stem species of a monophyletic group by a singular, identical splitting event. For example, the Lepidoptera and Trichoptera are sister groups; they shared a common ancestor that gave rise to no other lineage.

synapomorphy (-ic) A derived state shared among the members of a monophyletic group, in contrast to a **symplesiomorphy**—a shared ancestral (plesiomorphic) state from which phylogenetic relationships cannot be inferred.

taxon (pl. taxa) The general name for a taxonomic group at any rank.

taxonomic rank The classificatory level in the taxonomic hierarchy, e.g., species, genus, family, order. No rank is absolute, and comparisons between ranks of different organisms are inexact or even misleading; despite this, traditional ranks used for insects—notably orders and families—have useful didactic and synoptic value.

contentious issue in biology, with almost every possible higher level relationship finding some support. A once-influential "Mantonian" view proposed three groups of arthropods, each of which was derived from a different nonarthropod group, namely Uniramia (lobopods, myriapods, and insects, united by single-branched legs), Crustacea, and Chelicerata. More recent morphological and molecular studies reject this hypothesis, proposing instead monophyly of arthropodization, but postulated internal relationships are diverse. Part of Manton's Uniramia group—Atelocerata (also known as Tracheata), comprising myriapods and hexapods—finds support from morphological features including the occurrence of a tracheal system, Malpighian tubules, unbranched limbs, eversible coxal vesicles, postantennal organs, and anterior tentorial arms, lack of any evidence of the second antenna of crustaceans or a homologous structure, and the mandible comprising a complete limb rather than the limb base of the crustacean mandible. Proponents of this relationship saw Crustacea either grouping with chelicerates and extinct trilobites, separate from Atelocerata, or forming its sister group, in a clade called Mandibulata. Among all these schemes, the closest relatives of Hexapoda were proposed to be with, or possibly within, Myriapoda.

In contrast, novel and some rediscovered shared morphological features, including some from the nervous system (e.g., brain structure, neuroblast formation, and axon development), the visual system (e.g., fine structure of the ommatidia, optic nerves), and the process and control of development, especially segmentation, argued for a close relationship of Hexapoda to Crustacea, termed Pancrustacea, and exclusion of myriapods. Furthermore, all analyzed molecular sequence data with adequate signal to resolve relationships support Pancrustacea and not Atelocerata. As more nuclear, mitochondrial gene order and protein-encoding gene data have been examined for an ever-wider set of taxa, little or no support has been found for any of the possible groupings alternative to Pancrustacea. This does not imply that such analyses all identify Pancrustacea—sometimes certain "problematic" taxa have had to be removed, even from sparsely sampled data sets, and evidently certain genes do not retain strong phylogenetic signal from very old radiations.

If molecular-derived relationships are correct, features understood previously to infer monophyly of Atelocerata must be reconsidered. Postantennal organs occur in Hexapoda, but only in Collembola and Protura, and are suggested to be convergent with the organs in Myriapoda. Shared absence of features (such as lack of second antenna) cannot be taken as positive evidence of relationship. Malpighian tubules also are present (surely convergently) in arachnids, and evidence for homology between their structure in hexapods and myriapods remains inadequately studied. Coxal vesicles are not developed in all clades and may not be homologous in Myriapoda and those Hexapoda possessing these structures. Thus, morphological characters traditionally used to support Atelocerata include states that may be nonhomologous and convergently acquired through terrestriality, not distributed

across all included taxa, or inadequately surveyed across the immense morphological diversity of the arthropods. A major finding from molecular embryology, that the developmental expression of a homeotic gene *(Dll—Distal-less)* in the mandible of studied insects was the same as in sampled crustaceans, refutes the independent derivation of hexapod mandibles from those of crustaceans. This developmental homology for mandibles substantiates an earlier morphological hypothesis and undermines Manton's argument for arthropod polyphyly. In summary, data derived from the neural, visual, and developmental systems, even though sampled across relatively few taxa, appears to reflect more accurately phylogeny than much of the earlier external morphological studies.

The question remains as to whether part or all of the Crustacea constitute the sister group to Hexapoda. Morphology generally supports a monophyletic Crustacea, but inferences from some molecular data imply paraphyly, including a suggestion that Malacostraca alone are sister to Hexapoda (see below). Combined morphological and molecular data support both Crustacea and Pancrustacea monophyly, and Crustacea monophyly is thus preferred.

THE EXTANT HEXAPODA

Hexapoda (ranked usually as a superclass) contains all 6-legged arthropods; diagnosis includes possession of a unique tagmosis, namely specialization of successive body segments that more or less unite to form three sections or tagmata: head, thorax, and abdomen. The head is composed of a pregnathal region (often considered to be 3 segments) and 3 gnathal segments bearing mandibles, maxillae, and labium, respectively; the eyes are variously developed and sometimes absent. The thorax comprises 3 segments each of which bears one pair of legs, and each thoracic leg has a maximum of 6 segments in extant forms, but was primitively 11-segmented with up to 5 exites (outer appendages of the leg), a coxal endite (an inner appendage of the leg), and 2 terminal claws. Primitively the abdomen has 11 segments plus a telson or homolog; abdominal limbs, if present, are smaller and weaker than those of the thorax and primitively occurred on all segments except the 10th.

Basal hexapods undoubtedly include taxa whose ancestors were wingless and terrestrial. This grouping is not monophyletic, being based entirely on evident symplesiomorphies or otherwise doubtfully derived characters. Included groups, treated as orders, are Protura, Collembola, Diplura, Archaeognatha, and Zygentoma (Thysanura). True Insecta are the Archaeognatha, the Zygentoma, and the huge radiation of Pterygota (primary winged hexapods). Because Insecta is treated as a class, the successively more distant sister groups Diplura and Collembola (with or without Protura) are of equal rank.

Relationships among the component taxa of Hexapoda are uncertain, although the cladogram shown in Fig. 1 and the classification presented in the sections that follow reflect a current synthetic view. Traditionally, Collembola, Protura,

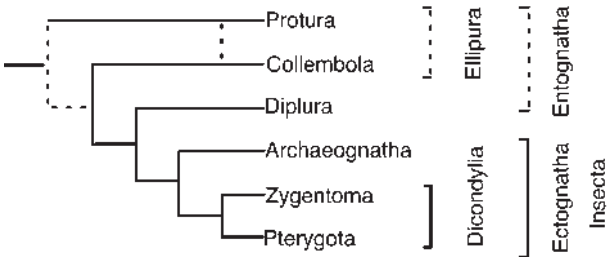

FIGURE 1 Cladogram depicting relationships among, and inferred classification of, higher ranked Hexapoda (six-legged Arthopoda). Dashed lines indicate uncertainty in cladogram or paraphyly in classification.

and Diplura were grouped as "Entognatha," based on the apparently similar morphology of the mouthparts. The mouthparts of Insecta (Archaeognatha + Zygentoma + Pterygota) are exposed (ectognathous), whereas those of Entognatha are enclosed in folds of the head. However, two different types of entognathy now are recognized, one shared by Collembola and Protura and the second found only in Diplura. Other morphological evidence indicates that Diplura may be closer to Insecta than to other entognathans and thus Entognatha may be paraphyletic (as indicated by the broken line in Fig. 1).

Protura (Proturans)

Proturans are small, delicate, elongate, mostly unpigmented hexapods, lacking eyes and antennae, with entognathous mouthparts consisting of slender mandibles and maxillae that slightly protrude from the mouth cavity. Maxillary and labial palps are present. The thorax is poorly differentiated from the 12-segmented abdomen. Legs are 5-segmented. A gonopore lies between segments 11 and 12, and the anus is terminal. Cerci are absent. Larval development is anamorphic, that is, with segments added posteriorly during development. Protura either is sister to Collembola, forming Ellipura in a weakly supported relationship based on similarity of the entognathous mouthparts and lack of cerci, or is sister to all remaining Hexapoda.

Collembola (Springtails)

Collembolans are minute to small and soft bodied, often with rudimentary eyes or ocelli. The antennae are 4- to 6-segmented. The mouthparts are entognathous, consisting predominantly of elongate maxillae and mandibles enclosed by lateral folds of head and lacking maxillary and labial palps. The legs are 4-segmented. The abdomen is 6-segmented with a sucker-like ventral tube, a retaining hook, and a furcula (forked jumping organ) on segments 1, 3, and 4, respectively. A gonopore is present on segment 5, the anus on segment 6. Cerci are absent. Larval development is, epimorphic, that is, with segment number constant through development. Collembola form either the sister group to Protura comprising Ellipura or a more strongly supported relationship as sister to Diplura + Insecta.

Diplura (Diplurans)

Diplurans are small to medium sized, mostly unpigmented, possessing long, moniliform antennae (like a string of beads), but lacking eyes. The mouthparts are entognathous, with tips of well-developed mandibles and maxillae protruding from the mouth cavity and maxillary and labial palps reduced. The thorax is poorly differentiated from the 10-segmented abdomen. The legs are 5-segmented and some abdominal segments have small styles and protrusible vesicles. A gonopore lies between segments 8 and 9, the anus is terminal. Cerci are filiform to forceps-like. The tracheal system is relatively well developed, whereas it is absent or poorly developed in other entognath groups. Larval development is epimorphic. Diplura forms the sister group to Insecta.

Class Insecta (True Insects)

Insects range from minute to large (0.2 to 300 mm in length) and are very variable in appearance. They typically have ocelli and compound eyes, at least in adults, and the mouthparts are exposed (ectognathous), with the maxillary and labial palps usually well developed. The thorax is variably developed in immature stages, but distinct in adults with degree of development dependent on the presence of wings. Thoracic legs have more than 5 segments. The abdomen is primitively 11-segmented with the gonopore nearly always on segment 8 in the female and segment 9 in the male. Cerci are primitively present. Gas exchange is predominantly tracheal with spiracles present on both the thorax and the abdomen, but variably reduced or absent (e.g., in many immature stages). Larval/nymphal development is epimorphic, that is, with the number of body segments constant during development.

The insects may be divided into two groups. Monocondylia is represented by just one small order, Archaeognatha, in which each mandible has a single posterior articulation with the head, whereas Dicondylia (Fig. 1), which contains the overwhelming majority of species, is characterized by mandibles with secondary anterior articulation in addition to the primary posterior one. The once traditional group Apterygota comprising the primarily wingless taxa Archaeognatha + Zygentoma is paraphyletic and rejected (Fig. 2).

ARCHAEOGNATHA (ARCHAEOGNATHANS, BRISTLETAILS) Archaeognathans are medium-sized, elongate cylindrical apterygotes, with some 500 species in two extant families. The head bears 3 ocelli and large compound eyes that are in contact medially. The antennae are multisegmented; the mouthparts project ventrally and can be partially retracted into the head and include elongate mandibles, one with 2 neighboring condyli each, and elongate 7-segmented maxillary palps. Often coxae II and III or III of legs bear a coxal stylet, tarsi 2- to 3-segmented. The abdomen, which continues in an even contour from the humped thorax, bears ventral muscle-containing styles (representing reduced limbs) on segments 2 to 9 and generally one or two pairs of eversible vesicles medial to the styles on segments 1 to 7. Cerci are multisegmented and shorter than the median caudal appendage. Sperm transfer is indirect via sperm droplets attached to a silken line or by stalked spermatophores. Development occurs without change in body form.

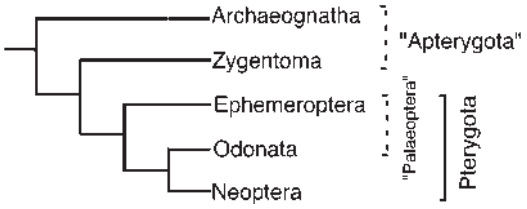

FIGURE 2 Cladogram depicting relationships among, and inferred classification of, higher ranked Insecta. Dashed lines indicate paraphyly in classification.

The two families of recent Archaeognatha, Machilidae and Meinertellidae, form an undoubted monophyletic group, whose position at the base of Ectognatha and as sister group to Dicondylia (Zygentoma + Pterygota) should be carefully investigated (Figs. 1 and 2).

ZYGENTOMA (THYSANURA, SILVERFISH) Zygentomans (thysanurans) are medium-sized, dorsoventrally flattened apterygotes with almost 400 species in four extant families. Eyes and ocelli are present, reduced, or absent and the antennae multisegmented. The mouthparts are ventrally to slightly forward projecting and include a special form of double-articulated (dicondylous) mandibles and 5-segmented maxillary palps. The abdomen continues the even contour of the thorax and includes ventral muscle-containing styles (representing reduced limbs) on at least segments 7 to 9, sometimes on 2 to 9, with eversible vesicles medial to the styles on some segments. Cerci are multisegmented and subequal to the length of the median caudal appendage. Sperm transfer is indirect via a spermatophore that the female picks up from the substrate. Development occurs without change in body form.

PTERYGOTA Pterygotes are winged or secondarily apterous insects, in which the thoracic segments of adults are usually large with the meso- and metathorax variably united to form a pterothorax. The lateral regions of the thorax are well developed. The 8 to 11 abdominal segments lack styles and vesicular appendages, and only most Ephemeroptera have a median terminal filament. The spiracles primarily have a muscular closing apparatus. Mating is by copulation. Metamorphosis is hemi- to holometabolous, with no adult ecdysis, except for the ephemeropteran subimago (subadult).

Informal Grouping "Palaeoptera" Palaeopteran wings are unable to be folded against the body at rest because articulation is via axillary plates that are fused with veins. Extant orders typically have triadic veins (paired main veins with intercalated longitudinal veins of convexity/concavity opposite to that of the adjacent main veins) and a network of crossveins. This wing venation and articulation, substantiated

by paleontological studies on similar features, has suggested that Odonata and Ephemeroptera form a monophyletic group, Palaeoptera, which is a sister group to Neoptera, which contains all remaining extant and primarily winged orders. However, reassessment of morphology of extant basal lineages and much existing molecular evidence appear to reject a monophyletic Palaeoptera. Here Ephemeroptera is treated as sister to Odonata + Neoptera, with Odonata alone as the sister group to Neoptera, giving a higher classification of Pterygota into three Divisions.

Ephemeroptera (Mayflies) Ephemeroptera has a fossil record dating back to the Carboniferous and is represented today by a few thousand species. In addition to their palaeopteran features, mayflies display a number of unique characteristics including the nonfunctional, strongly reduced adult mouthparts, the presence of just one axillary plate in the wing articulation, a hypertrophied costal brace, and male forelegs modified for grasping the female during copulatory flight. The retention of a subimago (subadult stage) is unique. Nymphs (larvae) are aquatic and the mandible articulation, which is intermediate between monocondyly and the dicondylous ball and socket joint of all higher Insecta, may be diagnostic. Historic contraction of ephemeropteran diversity and remnant high levels of homoplasy render phylogenetic reconstruction difficult. Ephemeroptera traditionally has been divided into two suborders: Schistonota (with nymphal forewing pads separate from each other for over half their length), containing the superfamilies Baetoidea, Heptagenioidea, Leptophlebioidea, and Ephemeroidea, and Pannota ("fused back"—with more extensively fused forewing pads), containing Ephemerelloidea and Caenoidea. Recent studies suggest that this concept of Schistonota is paraphyletic. Families Baetiscidae and Prosopistomatidae, whose nymphs have unusually developed thoracic shields, have been withdrawn from the Caenoidea and are placed now in the suborder Catapacea.

Odonata (Dragonflies and Damselflies) Odonates have palaeopteran as well as many additional unique features, including the presence of two axillary plates (humeral and posterior axillary) in the wing articulation and many features associated with specialized copulatory behavior, including possession of secondary copulatory apparatus on ventral segments 2 and 3 of the male and the formation of a tandem wheel during copulation. The immature stages are aquatic and possess a highly modified prehensile labium for catching prey.

Traditonally odonatologists have recognized three groups of taxa, Zygoptera (damselflies), Anisozygoptera, and Anisoptera (dragonflies), generally ranked as suborders. Assessment of their monophyly or paraphyly relies very much on characters derived from the very complex wing venation, but homology of these features within the odonates and between other insects has been substantially prejudiced by prior phylogenetic ideas. Thus, the Comstock and Needham wing-vein naming system implies that the common ancestor of modern Odonata was anisopteran and the zygopteran venation arrived by reduction. In contrast, the Tillyard system implied that Zygoptera is a grade on the way to Anisozygoptera, which itself is a grade on the way to Anisoptera. A well-supported view, including information from the substantial fossil record, has Zygoptera probably paraphyletic, Anisozygoptera undoubtedly paraphyletic, and Anisoptera as the monophyletic sister group to some extinct anisozygopterans.

Zygoptera contains three broad suprafamilial groupings, the Coenagrionoidea, Lestoidea, and Calopterygoidea. The first is centered around the family Coenagrionidae, but the group is made paraphyletic by excision of Eurasian Platycnemididae (for their dilated tibiae), South American Pseudostigmatidae (for their huge size), Platystictidae (for a narrow wing), and Protoneuridae and Isostictidae (for wings narrowed in a different way). Those latter two families differ only in the degree of wing narrowing, and the narrower one, Australasian Isostictidae, often is treated as a subfamily of the less narrow one, Protoneuridae.

The Calopterygoidea usually is regarded as centered around Calopterygidae. This specious family is made paraphyletic by excision of Chlorocyphidae (whose frons has a unique shape), Dicteriadidae (with legs lacking setae), Euphaeidae (a Southeast Asian variant on Calopterygidae, with abdominal gills in the larvae), and South American Polythoridae, with brightly colored wings.

Lestoidea contains the families Lestidae and Synlestidae, which clearly are related to each other. Perilestidae may belong here, and the enigmatic Hemiphlebiidae may be sister to this grouping. The quite different-looking Megapodagrionidae are related in some way to Amphipterygidae and Lestoideidae (the latter being an unstable mix of small genera allocated to either Lestoidea or Calopterygoidea).

Among Anisoptera four major lineages can be recognized, but their relationships to each other are obscure. Three aeshnid families, Aeshnidae, Neopetaliidae (evidently a subset of Aeshnidae), and Cordulegastridae (aeshnids with a secondarily elongate ovipositor), form a clade. The small (10 species) but very distinct Petaluridae forms a distinctive group. Gomphidae forms a large family all on its own. The superfamily Libelluloidea traditionally is divided into two large families, Cordulidae and Libellulidae, but the limits of each division are unclear, and no single character separates them. Chlorogomphidae, Macromiidae, and Synthemistidae are small, local "families" often separated out as near the corduliids.

Sister to Anisoptera is the minor suborder Anisozygoptera containing one extant genus with two species.

Neoptera Neopteran insects diagnostically have wings capable of being folded back against their abdomen when at rest, with wing articulation deriving from separate movable sclerites in the wing base and wing venation with fewer (or lacking completely) triadic veins and mostly lacking anastomosing (joining) crossveins.

The phylogeny (and hence classification) of the neopteran orders is still the subject of debate, mainly concerning (a) the placement of many extinct orders described only from fossils of variably adequate preservation, (b) the relationships among

the Polyneoptera (orthopteroid and plecopteroid orders), and (c) the relationships of the highly derived Strepsiptera.

However, the summary that follows reflects one possibility among current interpretations, based on both morphology and molecules. No single or combined data set provides unambiguous resolution of insect order-level phylogeny and there are several areas of controversy (such as the position of the Strepsiptera) arising from both inadequate data (insufficient or inappropriate taxon sampling) and character conflict within existing data. In the absence of a robust phylogeny, ranking is somewhat subjective and "informal" ranks abound.

A group of 11 orders termed the orthopteroid–plecopteroid assemblage (if monophyly is uncertain) or Polyneoptera (if monophyletic) is considered to be sister to the remaining Neoptera. The remaining neopterans can be divided readily into two monophyletic groups, namely Paraneoptera and Endopterygota (Holometabola). These three clades may be given the rank of subdivision.

Polyneoptera (or Orthopteroid–Plecopteroid Assemblage of Basal Neopteran Orders) [Isoptera, Blattodea, Mantodea, Dermaptera, Grylloblattaria (Grylloblattodia), Plecoptera, Orthoptera, Phasmatodea, Embiidina (Embioptera), Zoraptera, Mantophasmatodea] The relationships of the basal neopteran orders are poorly resolved with several, often contradictory, relationships being suggested by morphology. The 11 included orders may be monophyletic, based on the shared presence of tarsal plantulae (lacking only in Zoraptera) and limited, but increasing, molecular information. Within Polyneoptera only the grouping comprising Blattodea (cockroaches), Isoptera (termites), and Mantodea (mantids)—the Dictyoptera (Fig. 3)—is robust. Although each of these three orders is distinctive, features of the head skeleton (perforated tentorium), mouthparts (paraglossal musculature), digestive system (toothed proventriculus), and female genitalia (shortened ovipositor above a large subgenital plate) demonstrate monophyly of Dictyoptera, substantiated by nearly all molecular analyses. However, as seen below, views on the internal relationships are changing. Dermaptera (earwigs) is sister to Dictyoptera, and Grylloblattarea (rock crawlers; now apterous, but with winged fossils) may be sister to this grouping.

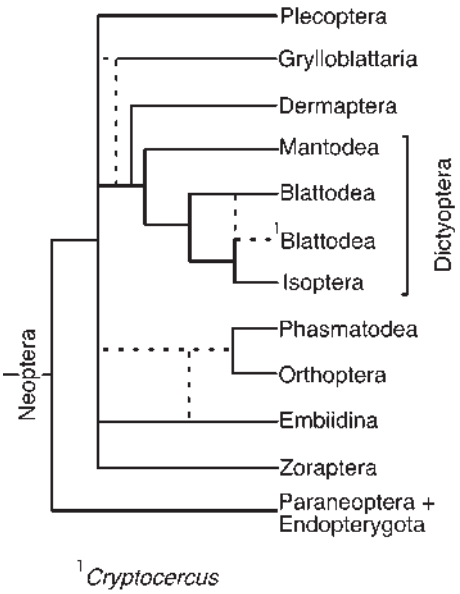

FIGURE 3 Cladogram depicting relationships among, and inferred classification of, orders of Neoptera: Polyneoptera. Dashed lines indicate uncertainty in cladogram. Mantophasmatodea not included.

Some molecular data suggest that Orthoptera (crickets, katydids, grasshoppers, locusts, etc.), Phasmatodea (stick insects or phasmids), and Embiidina (webspinners) may be closely related, forming Orthopteroidea in the sense of Hennig. The relationships of Plecoptera (stoneflies), orthopteroids, Zoraptera (zorapterans), and the recently discovered Mantophasmatodea to one another and to the above groupings are less well understood.

Isoptera (Termites, White Ants). Isoptera forms a small order of eusocial hemimetabolous neopterans, with some 2600 described species, living socially with polymorphic caste systems of reproductives, workers, and soldiers. The mouthparts are typically blattoid, being mandibulate but varying between castes, with some soldiers having bizarre development of mandibles or a nasus (snout). The compound eyes are frequently reduced, the antennae are long and multisegmented, and the fore- and hind wings are generally similar and membranous and have restricted venation. *Mastotermes* (Mastotermitidae) has complex wing venation and a broad hind-wing anal lobe and is exceptional among termites in that the female has a reduced blattoid-type ovipositor. The male external genitalia are weakly developed and symmetrical, in contrast to the complex, symmetrical genitalia of Blattodea and Mantodea.

Isopteran relationships are somewhat controversial, although they have always been considered to belong in Dictyoptera close to Blattodea. Recent studies that include the structure of the proventriculus and molecular sequence data suggest that termites may have arisen from within the cockroaches, thereby rendering that group paraphyletic. Under this scenario, the (wingless) wood roaches of North America and eastern Asia (genus *Cryptocercus*) form the sister group to Isoptera. This contrasts with alternative suggestions that the semisociality (parental care and transfer of symbiotic gut flagellates between generations) of *Cryptocercus* was convergent with certain features of termite sociality and independently originated within the true cockroaches. These two contrasting views are shown in Fig. 4. The social system and general morphology of *Mastotermes* suggests a cockroach-like condition, and most phylogenies place this group as sister to remaining extant Isoptera. Of considerable interest is the wide distribution and species richness of Mastotermitidae in Cretaceous times, compared with the reduced diversity of the extant family, which comprises just one species in northern Australia.

Blattodea (Cockroaches). Blattodea contains over 3500 species in at least eight families worldwide. They are hemimetabolous, dorsoventrally flattened insects with filiform, multisegmented antennae, and mandibulate, ventrally projecting mouthparts.

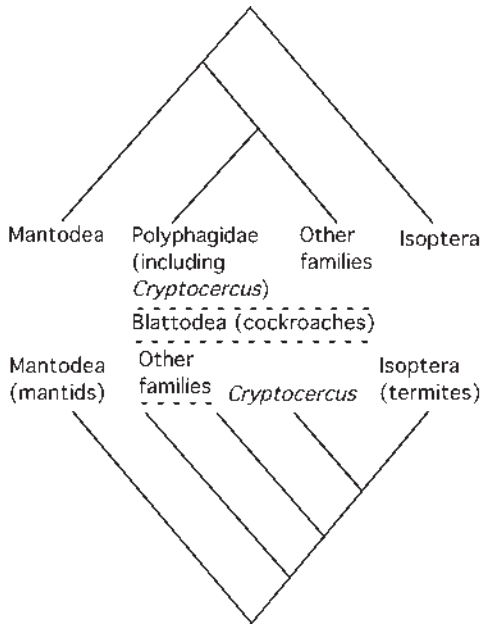

FIGURE 4 Cladogram depicting alternative relationships among Dictyoptera. Dashed lines indicate paraphyly in classification.

The prothorax has an enlarged, shield-like pronotum, often covering the head; the meso- and metathorax are rectangular and subequal. The forewings are sclerotized as tegmina, which protect the membranous hind wings folded fan-like at rest beneath the tegmina and characterized by many vein branches and a large anal lobe; wings are often reduced. Often the legs are spiny and the tarsi are 5-segmented. The abdomen has 10 visible segments, with a subgenital plate (sternum 9), bearing in the male well-developed asymmetrical genitalia, with one or two styles, and concealing the reduced 11th segment. Cerci have 1 or, usually, many segments; the female ovipositor valves are concealed beneath tergum 10.

Although long considered an order (and hence monophyletic) convincing evidence has been produced that the termites arose from within the cockroaches, and the "order" thus is rendered paraphyletic. The sister group of the Isoptera may be *Cryptocercus,* a social cockroach (Fig. 4). Blatellidae and Blaberidae form a derived sister grouping.

Mantodea (Mantids). Mantodea contains some 2000 species in eight families. They are hemimetabolous predators, with males generally smaller than females. The small, triangular head is mobile, with slender antennae, large, widely separated eyes, and mandibulate mouthparts. The prothorax is narrow and elongate, with the meso- and metathorax shorter. The forewings form leathery tegmina with a reduced anal area; the hind wings are broad and membranous, with long unbranched veins and many crossveins, but often are reduced or absent. The forelegs are raptorial, whereas the mid- and hind legs are elongate for walking. The abdomen has a visible 10th segment, bearing variably segmented cerci. The ovipositor is predominantly internal and the external male genitalia are asymmetrical.

Mantodea forms the sister group to Blattodea + Isoptera and shares many features with Blattodea, such as strong direct flight muscles and weak indirect (longitudinal) flight muscles, asymmetrical male genitalia, and multisegmented cerci. Derived features of Mantodea relative to Blattodea involve modifications associated with predation, including leg morphology, an elongate prothorax, and characteristics of visual predation, namely the mobile head with large, separated eyes. Internal relationships of Mantodea are uncertain and little studied.

Grylloblattaria (Notoptera, Grylloblattodea) (Grylloblattids or Rock Crawlers). Grylloblattaria contains one family (Grylloblattidae) with 20 species, restricted to western North America and central to eastern Asia and particularly adapted to cold and high elevations. Grylloblattids are moderate-sized, soft-bodied insects with anteriorly projecting mandibulate mouthparts and compound eyes that are either reduced or absent. The antennae are multisegmented and the mouthparts mandibulate. The quadrate prothorax is larger than the meso- or metathorax, and the wings are absent. The legs are adapted for running, with large coxae and 5-segmented tarsi. There are 10 visible abdominal segments with rudiments of segment 11, including 5- to 9-segmented cerci. The female has a short ovipositor, and the male genitalia are asymmetrical.

The phylogenetic placement of Grylloblattaria is controversial, as they are generally argued to be relicts that either "bridge the cockroaches and orthopterans" or are "primitive among orthopteroids." The antennal musculature resembles that of mantids and embiids, mandibular musculature resembles that of Dictyoptera, and the maxillary muscles resemble those of Dermaptera. Embryologically grylloblattids appear closest to the orthopteroids. The only molecular phylogenetic study that included a grylloblattid implied a sister group relationship to Dictyoptera, instead of one lying more basal in the Neoptera as is implied by the morphology. However, sampling in this analysis lacked some important plesiomorphic taxa, such as *Cryptocercus, Mastotermes,* and Embiidina. A tentative relationship of Grylloblattaria as sister to Dermaptera + Dictyoptera remains a favored hypothesis.

Dermaptera (Earwigs). Dermaptera is a worldwide order, modest in size, with some 10 families and about 1800 species. Adult earwigs are elongate and dorsoventrally flattened with mandibulate, forward-projecting mouthparts, compound eyes ranging from large to absent, no ocelli, and short annulate antennae. The tarsi are 3-segmented with a short second tarsomere. Many species are apterous or, if winged, the forewings are small, leathery, and smooth, forming unveined tegmina, and the hind wings are large, membranous, semicircular, and dominated by an anal fan of radiating vein branches connected by crossveins; when at rest, the hind wings are folded fan-like and then longitudinally, protruding slightly from beneath the tegmina.

The five species of suborder Arixeniina are commensals or ectoparasites of bats in Southeast Asia. A few species of semiparasites of African rodents have been placed in a suborder,

Hemimerina. These earwigs are blind, are apterous, have rod-like forceps, and exhibit pseudoplacental viviparity. Recent morphological study of Hemimerina suggests derivation from within Forficulina, rendering that suborder paraphyletic. The relationships of Arixeniina to more "typical" earwigs (Forficulina) are uninvestigated. Within Forficulina, only four (Karshiellidae, Apachyidae, Chelisochidae, and Forficulidae) of eight families proposed appear to be supported by synapomorphies. Other families may not be monophyletic, as much weight has been placed on plesiomorphies, especially of the penis specifically and the genitalia more generally, or homoplasies (convergences) in furcula form and wing reduction.

A sister group relationship to Dictyoptera is well supported on morphology, including many features of the wing venation.

Plecoptera (Stoneflies). Plecoptera is a minor order of some 16 families, predominantly living in temperate and cool areas. The adult is mandibulate with filiform antennae, bulging compound eyes, 2 or 3 ocelli, and subequal thoracic segments. The fore- and hind wings are membranous and similar except that the hind wings are broader; when folded, the wings partly wrap the abdomen and extend beyond the abdominal apex; aptery and brachyptery are frequent. The abdomen is soft and visibly 10-segmented, although remnants of segments 11 and 12 are present, including cerci. Nymphs have many (up to 33) aquatic instars, which have fully developed mandibulate mouthparts, and wing pads first become visible when the young are half-grown.

Monophyly of the order is supported by few morphological characters, including in the adult the looping and partial fusion of gonads and male seminal vesicles and the absence of an ovipositor. In nymphs the presence of strong, oblique, ventrolongitudinal muscles running intersegmentally and allowing lateral undulating swimming and the probably widespread "cercus heart," an accessory circulatory organ associated with posterior abdominal gills, support the monophyly of the order. Gills may be present in nymphal Plecoptera on almost any part of the body or may be absent, causing problems of homology of gills between families and between Plecoptera and other orders. Whether Plecoptera are derived from an aquatic or terrestrial ancestor is debatable.

The phylogenetic position of Plecoptera is certainly among "lower Neoptera," possibly as sister group to the remainder of Neoptera. However, some molecular and combined molecular plus morphological evidence tends to support a more derived position, including as sister to (i) Embiidina or, more likely, (ii) Dermaptera + Dictyoptera.

Internal relationships have been proposed as two predominantly disjunct suborders, the austral Antarctoperlaria and northern Arctoperlaria. The monophyly of Antarctoperlaria is argued based on the unique sternal depressor muscle of the fore trochanter, lack of the usual tergal depressor, and presence of floriform chloride cells, which may have a sensory function. Some of the included taxa are the large-sized Eustheniidae and Diamphipnoidae, the Gripopterygidae, and the Austroperlidae—all families with a Southern Hemisphere "Gondwanan"-type distribution. Recent molecular studies support this clade.

The sister group Arctoperlaria lacks defining morphology, but is united by a variety of mechanisms associated with drumming (sound production) used in mate-finding. The component families Scopuridae, Taeniopterygidae, Capniidae, Leuctridae, and Nemouridae (including Notonemouridae) are essentially of the Northern Hemisphere with a lesser radiation of Notonemouridae into the Southern Hemisphere. Molecular studies suggest the paraphyly of Arctoperlaria, with most elements of Notonemouridae forming the sister group to the remainder of the families. Relationships among extant Plecoptera are proving important in hypothesizing the origins of wings from "thoracic gills" and in tracing the possible development of aerial flight from surface flapping, with legs trailing on the water surface, and forms of gliding.

Zoraptera (Zorapterans). Zoraptera is one of the smallest and probably the least known pterygote order. Zorapterans are small, rather termite-like insects, found worldwide in tropical and warm temperate regions except Australia. Their morphology is simple, with biting, generalized mouthparts, including 5-segmented maxillary palps and 3-segmented labial palps. Sometimes both sexes are apterous, and in alate forms the hind wings are smaller than the forewings; the wings are shed as in ants and termites. Wing venation is highly specialized and reduced.

Traditionally the order contained only one family (Zorotypidae) and one genus *(Zorotypus),* but has been expanded to include seven genera delimited predominantly on wing venation. The phylogenetic position of Zoraptera based on morphology has been controversial, ranging through membership of the hemipteroid orders, sister to Isoptera, an orthopteroid, or a blattoid. Analysis of major wing structures and musculature imply that Zoraptera belong in the blattoid lineage. Although the wing shape and venation resemble those of narrow-winged Isoptera, cephalic and abdominal characters indicate an early divergence from the blattoid stock, prior to the divergence of Dermaptera and much before the origin of the Dictyoptera lineage.

Orthoptera (Grasshoppers, Locusts, Katydids, Crickets). Orthopterans belong to at least 30 families and more than 20,000 species, and most are medium-sized to large insects with hind legs enlarged for jumping (saltation). The compound eyes are well developed, the antennae are elongate and multisegmented, and the prothorax is large with a shield-like pronotum curving downward laterally. The forewings form narrow, leathery tegmina, and the hind wings are broad, with numerous longitudinal and crossveins, folded beneath the tegmina by pleating; aptery and brachyptery are frequent. The abdomen has 8 or 9 annular visible segments, with the 2 or 3 terminal segments reduced, and 1-segmented cerci. The female has a well-developed ovipositor formed from highly modified abdominal appendages.

Virtually all morphological evidence, and much of the newer molecular data suggest that the Orthoptera form the

sister group to Phasmatodea. Some authors have united the orders, but the different wing-bud development, different egg morphology, and lack of auditory organs in phasmatids suggest separation. Molecular evidence indicates that Embiidina may be sister to the orthopteran–phasmatid clade, but the support for this relationship is weak.

The division of Orthoptera into two monophyletic suborders, Caelifera (grasshoppers and locusts—predominantly day-active, fast-moving, visually acute, terrestrial herbivores) and Ensifera (katydids and crickets—often night active, camouflaged or mimetic, predators, omnivores, or phytophages), is supported on morphological and molecular evidence. Grylloidea probably is the sister group (but highly divergent, with a long branch separation) of the remaining ensiferan taxa, Tettigonioidea, Hagloidea, and Stenopelmatoidea. On grounds of some molecular and morphological data Tettigoniidae and Haglidae form a monophyletic group, sister to Stenopelmatidae and relatives (mormon crickets, wetas, cooloola monsters, and the like), but alternative analyses suggest different relationships, and conservatively, an unresolved group is perhaps appropriate at this stage.

In Caelifera a well-supported recent proposal for four superfamilies, namely [Tridactyloidea (Tetragoidea (Eumastacoidea + "higher Caelifera"))] reconciles molecular evidence with certain earlier suggestions from morphology. The major grouping of acridoid grasshoppers (Acridoidea) lies in the unnamed clade "higher Caelifera," which also includes the superfamilies Tanaoceroidea, Pyrgomorphoidea, Pneumoroidea, and Trigonopterygoidea.

Phasmatodea (Phasmatids, Phasmids, Stick Insects or Walking Sticks). Phasmatodea are a worldwide, predominantly tropical order of more than 2500 species of hemimetabolous insects, conventionally classified in three families (although some workers raise many subfamilies to family rank). Body shapes are variations on elongate cylindrical, and stick-like or flattened, or often leaf-like. The mouthparts are mandibulate. The compound eyes are relatively small and placed anterolaterally, with ocelli only in winged species and often only in males. The antennae are short to long, with 8 to 100 segments. The prothorax is small, and mesothorax and metathorax are long in wingless species and shorter in apterous ones. The wings, when present, are functional in males, often reduced in females, but with many species apterous in both sexes; the forewings form short leathery tegmina, whereas the hind wings are broad with a network of numerous crossveins and with the anterior margin toughened to protect the folded wing. The legs are elongate, slender, and adapted for walking, with 5-segmented tarsi. The abdomen is 11-segmented, with segment 11 often forming a concealed supra-anal plate in males or a more obvious segment in females.

Phasmatodea are sister to Orthoptera in the orthopteroid assemblage. Novel support for this grouping comes from the dorsal position of the cell body of salivary neuron 1 in the subesophageal ganglion and presence of serotonin in salivary neuron 2. Phasmatodea are distinguished from the Orthoptera by their body shape, asymmetrical male genitalia, proventricular structure, and lack of rotation of nymphal wing pads during development.

Embiidina (Embioptera) (Embiids, Webspinners). Embiidina comprise some 200 described species (perhaps up to an order of magnitude more remain undescribed) in at least eight families. The body is elongate, cylindrical, and somewhat flattened in males. The head has kidney-shaped compound eyes that are larger in males than in females and lacks ocelli. The antennae are multisegmented and the mandibulate mouthparts project forward (prognathy). All females and some males are apterous, but if present, the wings are characteristically soft and flexible, with blood sinus veins stiffened for flight by blood pressure. The legs are short, with 3-segmented tarsi, and the basal segment of the fore tarsi is swollen because it contains silk glands. The hind femora are swollen by strong tibial muscles. The abdomen is 10-segmented with rudiments of segment 11 and with 2-segmented cerci. The female external genitalia are simple (no ovipositor), and those of the males are complex and asymmetrical.

Embiids are undoubtedly monophyletic based on, *inter alia,* the ability to produce silk from unicellular glands in the anterior basal tarsus. They have a general morphological resemblance to Plecoptera based on reduced phallomeres, a trochantin–episternal sulcus, separate coxopleuron, and premental lobes. However, molecular evidence suggests a closer relationship to Orthoptera and Phasmatodea; they also have some similarity to the Dermaptera, notably deriving from their prognathy. Internal relationships among the described higher taxa of Embidiina suggest that the prevailing classification includes many nonmonophyletic groups. Evidently much further study is needed to understand relationships within Embiidina and among it and other neopterans.

Mantophasmatodea. Mantophasmatodea has been recognized recently for a species in Baltic amber and museum specimens representing two species from southwest and east Africa and from freshly collected material from Namibia. The taxon cannot be placed within any of the existing orders and initial estimates of relationships are unclear. Some resemblances to Grylloblattarea and Phasmatodea are evident, but more study, including molecular sequencing, is required.

Paraneoptera (Acercaria or Hemipteroid Assemblage) This group contains Psocoptera + Phthiraptera, Thysanoptera, and Hemiptera and is defined by derived features of the mouthparts, including the slender, elongate maxillary lacinia separated from the stipes, and the swollen postclypeus containing an enlarged cibarium (sucking pump), and the reduction in tarsomere number to 3 or less.

Within Paraneoptera, the monophyletic superorder Psocodea contains Phthiraptera (parasitic lice) and Psocoptera (book lice). Although Phthiraptera is monophyletic, the clade arose from within Psocoptera, rendering that group paraphyletic. Although sperm morphology and some molecular sequence data imply the relationship Hemiptera (Psocodea + Thysanoptera), a grouping of Thysanoptera + Hemiptera

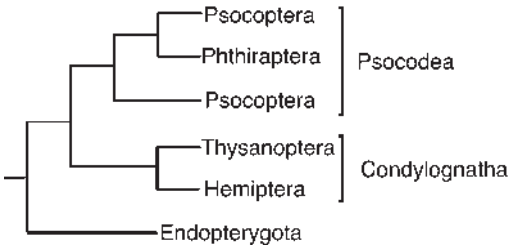

FIGURE 5 Cladogram depicting relationships among, and inferred classification of, Paraneoptera.

(superorder Condylognatha) is supported by head and mouthpart synapomorphies, including the stylet mouthparts, features of the wing base, and the sclerotized ring between antennal flagellomeres. Condylognatha thus forms the sister group to Psocodea (Fig. 5).

Psocoptera (Plant Lice). Psocoptera is a worldwide order of cryptic small insects, with a large, mobile head, bulbous postclypeus, and membranous wings held roof-like over the abdomen. Some 3000 species are described in 35 families. Evidently Pscoptera belong with Phthiraptera in a monophyletic clade, Psocodea. However, Psocoptera is rendered paraphyletic by a postulated relationship of Phthiraptera to the psocopteran family Liposcelidae. Internal relationships in Psocoptera are poorly known.

Phthiraptera (Lice). Phthirapterans are wingless obligate ectoparasites of birds and mammals, lacking any free-living stage, with some 3000 species in some 17 families. Monophyly of, and relationships among, the traditional suborders Anoplura, Amblycera, Ischnocera, and Rhyncophthirina are poorly understood and nearly all possible arrangements have been proposed. The latter three suborders have been treated as a monophyletic Mallophaga (biting and chewing lice) based on their feeding mode and morphology, in contrast to the piercing and blood-feeding Anoplura. Cladistic analysis of morphology has disputed mallophagan monophyly, suggesting the relationship Amblycera [Ischnocera (Anoplura + Rhyncophthirina)]. The only molecular data adduced thus far found unequivocal monophyly only for Amblycera, with a placement that rendered Ischnocera paraphyletic. The data neither supported nor refuted mallophagan monophyly. Resolution of these issues is important in estimation of degree of cospeciation between lice and their bird and mammal hosts.

Thysanoptera (Thrips). Thysanoptera is a worldwide order of just over 5000 species in nine families. The development of thrips is intermediate between hemi- and holometabolous. Their head is elongate and the mouthparts are unique in that the maxillary laciniae formed grooved stylets, the right mandible is atrophied, but the left mandible forms a stylet; all three stylets together form the feeding apparatus. The tarsi are 1- or 2-segmented, and the pretarsus has an apical protrusible adhesive ariolum (bladder or vesicle). Female thrips are diploid, whereas males (if present) are haploid, being produced from unfertilized eggs.

Molecular evidence supports one of the traditional morphological divisions of the Thysanoptera into two suborders, Terebrantia and Tubulifera, containing the sole, speciose, family Phlaeothripidae. Terebrantia includes three speciose families, Thripidae, Heterothripidae, and Aeolothripidae, and five smaller families. Relationships among families in Terebrantia are poorly resolved, although phylogenies are being generated at lower levels concerning aspects of the evolution of sociality, especially the origins of gall-inducing thrips and of "soldier" castes in Australian gall-inducing Thripidae.

Hemiptera (Bugs, Cicadas, Leafhoppers, Planthoppers, Spittlebugs, Aphids, Psylloids, Scale Insects, Whiteflies, Moss Bugs). Hemiptera is the largest of the non-endopterygote orders, with more than 50,000 species in approximately 100 families. Hemipteran mouthparts diagnostically have the mandibles and maxillae modified as needle-like stylets, lying in a beak-like, grooved labium, collectively forming a rostrum or proboscis within which the stylet bundle contains two canals, one delivering saliva and the other taking up fluid. Palps are lacking. The thorax usually has a large prothorax and mesothorax and a small metathorax. Both pairs of wings often have reduced venation, some species are apterous, and male scale insects have only one pair of wings. Legs often possess complex pretarsal adhesive structures. Cerci are lacking.

Hemiptera and Thysanoptera are sister groups within Paraneoptera. Hemiptera used to be divided into two groups, Heteroptera (true bugs) and "Homoptera" (cicadas, leafhoppers, planthoppers, spittlebugs, aphids, psylloids, scale insects, and whiteflies), treated variously as suborders or orders. All homopterans are terrestrial plant feeders and many share a common biology of producing honeydew and being ant attended. However, although possessing defining features, such as wings held roof-like over the abdomen, forewings in the form of a tegmina of uniform texture, and a rostrum arising ventrally close the anterior of the thorax, "Homoptera" represents a paraphyletic grade rather than a clade. This view is supported by reinterpreted morphological data and by cladistic analysis of nucleotide sequences from the nuclear small subunit ribosomal RNA gene (also called 18S rRNA). These data also suggest a much more complicated pattern of relationships among the higher groups of hemipterans (Fig. 6).

The ranking of the various clades is much disputed and thus the more stable superfamily and family names have been used here. Four suborders appear warranted on phylogenetic grounds: Archaeorrhyncha, Clypeorrhyncha, and Prosorrhyncha collectively form the Euhemiptera, which is the sister group to the fourth suborder, Sternorrhyncha. The latter contains the aphids, psyllids, scale insects, and whiteflies, which are characterized principally by their possession of a particular kind of gut filter chamber, a rostrum that appears to arise between the bases of their front legs, and, if winged, the absence of vannus and vannal fold in the hind wing. Some relationships among Euhemiptera are unsettled. A monophyletic Auchenorrhyncha, morphologically defined by their possession of a tymbal acoustic system, an aristate antennal

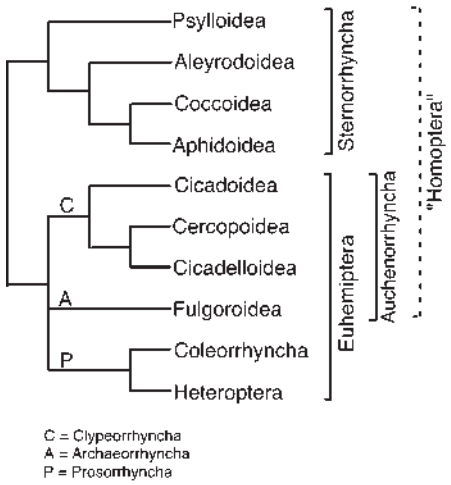

FIGURE 6 Cladogram depicting relationships among, and inferred classification of, Hemiptera. Dashed line indicates paraphyly in classification.

flagellum, and reduction of the proximal median plate in the wing base, contains two suborders, Archaeorrhyncha (planthoppers, often called Fulgoromorpha) and Clypeorrhyncha (cicadas, leafhoppers, and spittlebugs, often called Cicadomorpha). Molecular data tend to refute this monophyly, implying that Archaeorrhyncha is closer to Heteroptera, but relationships depend upon sampling and more traditional arrangements are minimally less parsimonious.

Heteroptera (true bugs, including assassin bugs, backswimmers, lace bugs, stink bugs, waterstriders, and others) and its sister group, variously called Coleorrhyncha, Peloridioidea, or Peloridiomorpha and containing only one family, Peloridiidae, or moss bugs, form the suborder Prosorrhyncha. Although small, cryptic, and rarely collected, moss bugs have generated considerable phylogenetic interest due to their combination of ancestral and derived hemipteran features and their exclusively "relictual" Gondwanan distribution. Heteropteran diversity is distributed among some 75 families, forming the largest hemipteran clade. Heteroptera is most easily diagnosed by the presence of metapleural scent glands, and monophyly is never disputed.

Endopterygota (Coleoptera, Neuroptera, Megaloptera, Raphidioptera, Hymenoptera, Trichoptera, Lepidoptera, Mecoptera, Siphonaptera, Diptera, Strepsiptera) Endopterygota comprise insects with immature (larval) instars that are very different from their respective adults. The adult wings and genitalia are internalized in their preadult expression, developing in imaginal discs that are evaginated at the penultimate molt. Larvae lack true ocelli. The "resting stage" or pupa is nonfeeding and precedes an often active pharate ("cloaked" in pupal cuticle) adult. Unique derived features are less evident in the adults than in immature stages, but the clade is consistently recovered from morphological, molecular, and combined analyses.

Two or three groups currently are proposed among the endopterygotes, of which one of the strongest is a sister group relationship termed Amphiesmenoptera between the Trichoptera (caddisflies) and Lepidoptera (butterflies and moths). A plausible scenario of an ancestral amphiesmenopteran taxon has a larva living in damp soil among liverworts and mosses followed by radiation into water (Trichoptera) or into terrestriality and phytophagy (Lepidoptera).

A second (usually) strongly supported relationship is between three orders, Neuroptera, Megaloptera, and Rhaphidioptera—Neuropterida (sometimes treated as a group of ordinal rank)—showing a sister group relationship to Coleoptera.

A third, postulated relationship—Antliophora—unites Diptera (true flies), Siphonaptera (fleas), and Mecoptera (scorpionflies and hangingflies). Debate contines about the relationships of these taxa, particularly concerning the relationships of Siphonaptera. Fleas were considered sister to Diptera, but molecular and novel anatomical evidence increasingly points to a relationship with a curious-looking mecopteran, *Boreus.*

Strepsiptera is phylogenetically enigmatic, but resemblance of their first-instar larvae (called triungulins) to certain Coleoptera, notably parasitic Rhipiphoridae, and some wing base features have been cited as indicative of relationship. This placement is becoming less likely, as molecular evidence (and haltere development) suggests a link between Strepsiptera and Diptera. Strepsiptera is highly derived in both morphological and molecular evolution, and thus possesses few features shared with any other taxon. The long-isolated evolution of the genome can create a problem known as "long-branch attraction," in which nucleotide sequences may converge by chance events alone with those of an unrelated taxon with a similarly long evolution, for the strepsipteran notably with Diptera. The issue remains unresolved.

The positions of two major orders of endopterygotes, Coleoptera and Hymenoptera, remain to be considered. Several positions have been proposed for Coleoptera but current evidence derived from female genitalia and ambivalent evidence from eye structure supports a sister group relationship to Neuropterida. This grouping forms the sister to the remaining Endopterygota in most analyses. Hymenoptera may be the sister to Antliophora + Amphiesmenoptera; the many highly derived features of adults, and reductions in larvae, limit morphological justification for this position.

Within the limits of uncertainty, the relationships within Endopterygota are summarized in Fig. 7.

Coleoptera (Beetles). Coleoptera undoubtedly lie among the basal Endopterygota. The major synapomorphic feature of Coleoptera is the development of the forewings as sclerotized rigid elytra, which extend to cover some or many of the abdominal segments and beneath which the propulsive hind wings are elaborately folded when at rest. Some molecular studies show Coleoptera polyphyletic or paraphyletic with respect to some or all of the Neuropterida. However, this is impossible to reconcile with the morphological support for coleopteran monophyly, and a sister group relationship to Neuropterida is accepted.

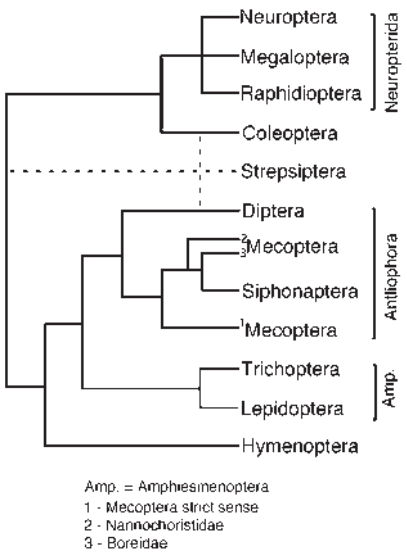

FIGURE 7 Cladogram depicting relationships among, and inferred classification of, Endopterygota. Dashed lines indicate three possible alternative placements of Strepsiptera.

Within Coleoptera, four modern lineages (treated as suborders) are recognized: Archostemata, Adephaga, Polyphaga, and Myxophaga. Archostemata includes only the small families Ommatidae, Crowsoniellidae, Cupedidae, and Micromalthidae and forms the sister group to the remaining extant Coleoptera. The few known larvae are woodminers with a scerotized ligula and a large mola on each mandible. Adults have movable hind coxae with usually visible trochantins and 5 (not 6) ventral abdominal plates (ventrites) but share with Myxophaga and Adephaga wing-folding features (apex spirally rolled, major transverse fold crossing vein MP), absence of cervical sclerites, and the external prothoracic pleuron. In contrast to Myxophaga, the pretarsus and tarsus are unfused.

Adephaga is diverse, second in size only to Polyphaga, and includes ground beetles, tiger beetles, whirligigs, predaceous diving beetles, and wrinkled bark beetles, among others. Larval mouthparts are adapted for liquid-feeding, with a fused labrum and no mandibular mola. Adults have the notopleural sutures visible on the prothorax and have 6 visible abdominal sterna with the first three fused into a single ventrite, which is divided by the hind coxae. Pygidial defense glands are widespread in adults. The most speciose included family is Carabidae, or ground beetles, with a predominantly predaceous feeding habit, but Adephaga also includes the aquatic families, Haliplidae and Noteridae, which are algivorous, and the mycophagous Rhysodidae. Morphology has suggested that Adephaga is sister group to the combined Myxophaga and Polyphaga, but molecular data (18S rDNA) suggest Adephaga as sister to Polyphaga, with Myxophaga sister to them.

Myxophaga is a clade of small, primarily riparian aquatic beetles, comprising the families Lepiceridae, Torridincolidae, Hydroscaphidae, and Microsporidae, united by the synapomorphic fusion of the pretarsus and tarsus. The 3-segmented larval antenna, 5-segmented larval legs with a single pretarsal claw, fusion of trochantin with the pleuron, and ventrite structure support a sister group relationship of Myxophaga with the Polyphaga. This has been challenged by some workers, notably because wing venation and folding provide evidence for (Polyphaga (Archostemata (Myxophaga + Adephaga))).

Polyphaga contains the majority (>90% of species) of beetles, with about 300,000 described species. The suborder includes rove beetles (Staphylinoidea), scarabs and stag beetles (Scarabaeoidea), metallic wood-boring beetles (Buprestoidea), click beetles and fireflies (Elateroidea), as well as diverse Cucujiformia, including fungus beetles, grain beetles, ladybird beetles, darkling beetles, blister beetles, longhorned beetles, leaf beetles, and weevils. The prothoracic pleuron is not externally visible, but is fused with the trochantin and remnant internally as a "cryptopleuron." Thus one suture between the notum and the sternum is visible in the prothorax in polyphagans, whereas two sutures (the sternopleural and notopleural) often are visible externally in other suborders (unless secondary fusion between the sclerites obfuscates the sutures, as in *Micromalthus*). The transverse fold of the hind wing never crosses MP, cervical sclerites are present, and hind coxae are mobile and do not divide the first ventrite. Female polyphagan beetles have telotrophic ovarioles, which is a derived condition within beetles.

The internal classification of Polyphaga involves several superfamilies or series, whose constituents are relatively stable, although some smaller families (whose rank even is disputed) are allocated to different clades by different authors. Large superfamilies include Hydrophiloidea, Staphylinoidea, Scarabaeoidea, Buprestoidea, Byrrhoidea, Elateroidea, Bostrichoidea, and the grouping Cucujiformia. This latter includes the vast majority of phytophagous (plant-eating) beetles, united by cryptonephric Malpighian tubules of the normal type, a cone ommatidium with open rhabdom, and lack of functional spiracles on the eighth abdominal segment. Constituent superfamilies of Cucujiformia are Cleroidea, Cucujoidea, Tenebrionoidea, Chrysomeloidea, and Curculionoidea. Evidently adoption of a phytophagous lifestyle correlates with speciosity in beetles, with Cucujiformia, especially weevils (Curculionoidea), forming a major radiation.

Neuropterida or Neuropteroid Orders: Megaloptera (Alderflies, Dobsonflies, Fishflies), Raphidioptera (Snakeflies), and Neuroptera (Lacewings, Antlions, Owlflies). Neuropterida comprise three small orders with holometabolous development, with approximately 6000 species of Neuroptera in about 20 families, 300 species of Megaloptera in 2 widely recognized families, and 200 species of Raphidioptera in 2 families. Adults have multisegmented antennae, large, separated eyes, and mandibulate mouthparts. The prothorax may be larger than the meso- and metathorax, which are about equal in size. Legs sometimes are modified for predation. The fore- and hind wings are quite similar in shape and venation, with folded wings often extending beyond the abdomen. The abdomen lacks cerci.

Megalopterans are predatory only in the aquatic larval stage; although adults have strong mandibles, they are not used in feeding. Adults closely resemble neuropterans, except for the presence of an anal fold in the hind wing. Raphidiopterans are terrestrial predators as adults and larvae. The adult is mantid-like, with an elongate prothorax, and the head is mobile and used to strike, snake-like, at prey. The larval head is large and forwardly directed. Many adult neuropterans are predators and have wings typically characterized by numerous crossveins and "twigging" at the ends of veins. Neuropteran larvae usually are active predators with slender, elongate mandibles and maxillae combined to form piercing and sucking mouthparts.

Megaloptera, Raphidioptera, and Neuroptera may treated as separate orders or united in Neuropterida, or Raphidioptera may be included in Megaloptera. There is little doubt that Neuropterida is monophyletic, with new support from wing base morphology. This latter feature also supplements data supporting the long-held view that Neuropterida forms a sister group to Coleoptera. Each component appears monophyletic, although a doubt remains concerning megalopteran monophyly. There remains uncertainty about internal relationships, which traditionally have Megaloptera and Raphidioptera as sister groups. Recent reanalyses with some new character suites have postulated Megaloptera as sister to Neuroptera and proposed a novel scenario of the plesiomorphy of aquatic larvae (all Megaloptera and Sisyridae in Neuroptera) in Neuropterida.

Strepsiptera. Strepsiptera form an enigmatic order of nearly 400 species of highly modified endoparasites, most commonly of Hemiptera and Hymenoptera, and show extreme sexual dimorphism. The male has a large head with bulging eyes comprising few large facets and lacks ocelli; the antenna are flabellate or branched, with 4 to 7 segments; the forewings are stubby and lack veins, whereas the hind wings are broadly fan-shaped, with few radiating veins; the legs lack trochanters and often also claws. Females are either coccoid-like or larviform, wingless, and usually retained in a pharate (cloaked) state, protruding from the host. The first instar is a triungulin, without antennae and mandibles, but with three pairs of thoracic legs; subsequent instars are maggot-like, lacking mouthparts or appendages. The pupa, which has immovable mandibles but appendages free from its body, develops within a puparium formed from the last instar.

The phylogenetic position of Strepsiptera has been subject to much speculation because modifications associated with their endoparasitic lifestyle mean that few characteristics are shared with possible relatives. In having posteromotor flight (only metathoracic wings) they resemble Coleoptera, but other attributes traditionally argued to be synapomorphous with Coleoptera are suspect or mistaken. The forewing-derived halteres of strepsipterans are gyroscopic organs of equilibrium with the same functional role as the halteres of Diptera, although the latter are derived from the hind wing. Molecular sequence studies indicate that Strepsiptera possibly is a sister group to Diptera, and some tantalizing information from developmental biology suggests that wings and halteres might be "reverse-expressed" on meso- and metathoracic segments.

Mecoptera (Scorpionflies, Hangingflies). Mecopterans are holometabolous insects comprising about 500 known species in nine families, with common names associated with the two largest families—Bittacidae (hangingflies) and Panorpidae (scorpionflies). Adults have an elongate ventrally projecting rostrum, containing elongate, slender mandibles and maxillae, and an elongate labium. The eyes are large and separated, the antennae filiform and multisegmented. The fore- and hind wings are narrow, similar in size, shape, and venation, but often are reduced or absent. The legs may be modified for predation. Larvae have a heavily sclerotized head capsule, are mandibulate, and may have compound eyes comprising 3 to 30 ocelli (absent in Panorpidae, indistinct in Nannochoristidae). The thoracic segments are about equal and have short thoracic legs with fused tibia and tarsus and a single claw; prolegs usually are present on abdominal segments 1 to 8, and the terminal segment (10) has either paired hooks or a suction disk. The pupa is immobile, mandibulate, and with appendages free.

Although some adult Mecoptera resemble neuropterans, strong evidence supports a relationship to Diptera. Intriguing recent morphological studies, plus robust evidence from molecular sequences, suggest that Siphonaptera arise from within Mecoptera, as a sister group to the "snow fleas" (Boreidae). The phylogenetic position of Nannochoristidae, the southern hemisphere mecopteran taxon currently treated as being of subfamily rank, has a significant bearing on internal relationships within Antliophora. Molecular evidence suggests that it lies as sister to Boreidae + Siphonaptera and therefore is of rank equivalent to the boreids, the fleas, and the residue of Mecoptera—and logically each should be treated as an order or Siphonaptera should be reduced in rank within Mecoptera.

Siphonaptera (Fleas). Siphonaptera is a highly modified order of holometabolous insects, comprising some 2400 species, all of which are bilaterally compressed, apterous ectoparasites. The mouthparts are specialized for piercing and sucking, lack mandibles, but have an upaired labral stylet and two elongate serrate, lacinial stylets that together lie within a maxillary sheath. A salivary pump injects saliva into the wound, and cibarial and pharyngeal pumps suck up the blood meal. Fleas lack compound eyes and the antennae lie in deep lateral grooves. The body is armed with many posteriorly directed setae and spines, some of which form combs, especially on the head and anterior thorax. The metathorax houses very large muscles associated with the long and strong hind legs, which are used to power the prodigious leaps made by these insects.

After early suggestions that the fleas arose from a mecopteran, the weight of evidence suggested that they formed the sister group to Diptera. However, increasing molecular and novel morphological evidence now points to a sister group relationship to a subordinate component of Mecoptera, specifically Boreidae (snow fleas). Internal relationships of the fleas are under study, and preliminary results imply that monophyly of many families is uncertain.

Diptera (True Flies). Diptera is a major order of insects, with perhaps as many as a quarter of a million species in some 120 families. Dipterans are holometabolous and readily recognized by the development of hind (metathoracic) wings as balancers, or halteres (halters), and in the larval stages by lack of true legs and the often maggot-like appearance. Venation of the fore (mesothoracic) flying wings ranges from complex to extremely simple. Mouthparts range from biting-and-sucking (e.g., biting midges and mosquitoes) to "lapping"-type with a pseudotracheate labella functioning as a sponge (e.g., house flies). Dipteran larvae lack true legs, although various kinds of locomotory apparatus range from unsegmented pseudolegs to creeping welts on maggots. The larval head capsule ranges from complete, through partially undeveloped, to complete absence of a maggot head, retaining only the sclerotized mandibles ("mouth hooks") and supporting structures.

Traditionally Diptera had two suborders, Nematocera (crane flies, midges, mosquitos, and gnats), with a slender, multisegmented antennal flagellum, and the heavier built Brachycera ("higher flies," including hover flies, blow flies, and dung flies), with shorter, stouter, and fewer-segmented antenna. However, Brachycera is sister to only part of "Nematocera," and thus Nematocera is paraphyletic.

Internal relationships among Diptera are becoming better understood, although with some notable exceptions. Ideas concerning basal Diptera are inconsistent: traditionally Tipulidae (or Tipulomorpha if subordinate groups are given family rank, but nonetheless undoubtedly monophyletic) is a basal clade, particularly on evidence from the adult wing and other morphology. Such an arrangement is difficult to reconcile with the much more derived larva, in which the head capsule is variably reduced. Furthermore, some molecular evidence casts doubt on the basal position of the crane flies, but as yet does not produce a robust estimate for any alternative basal grouping. Alternative views based on morphology have suggested that the relictual family Tanyderidae, with complex ("primitive") wing venation, lies close to the base of the order. In this instance support comes both from the tanyderid larval morphology and from the putative placement in Psychodomorpha, considered a probable basal clade.

There is strong support for Culicomorpha, comprising mosquitoes (Culicidae) and their relatives (Corethrellidae, Chaoboridae, Dixidae) and their sister group the blackflies, midges, and relatives (Simuliidae, Thaumaleidae, Ceratopogonidae, Chironomidae), and for Bibionomorpha, comprising the fungus gnats (Mycetophilidae *sensu lato*), Bibionidae, Anisopodidae, and possibly Cecidomyiidae. However, in both groups internal relationships remain a matter of debate, which molecular evidence may help to resolve.

Monophyly of Brachycera, comprising higher flies, is established by features including, in the larva, the posterior elongation of the head into the prothorax, the divided mandible, and the loss of premandible and, in the adult, the 8 or fewer antennal flagellomeres, 2 or fewer palp segments, and separation of the male genitalia into 2 parts (epandrium and hypandrium). Possible relationships of Brachycera include sister to Psychodomorpha or even to Culicomorpha (molecular data only) but strong support is provided for sister taxon to the Bibiomorpha or to subordinate Anisopodidae. Brachycera contains four equivalent groups with internally unresolved relationships: Tabanomorpha (with brush on larval mandible and larval head retractile), Stratiomyomorpha (with larval cuticle calcified and pupation in last instar exuviae), Xylophagomorpha (with distinctive elongate, conical, strongly sclerotized larval head capsule and abdomen posteriorly ending in sclerotized plate with terminal hooks), and Muscomorpha (adults with tibial spurs absent, flagellum with no more than 4 flagellomeres, and female cercus single segmented). This latter speciose group contains Asiloidea (robber flies, bee flies, and relatives) and Eremoneura (empidoids and Cyclorrhapha). Eremoneura is a strongly supported clade based on wing venation (loss or fusion of vein M_4 and closure of anal cell before margin), presence of ocellar setae, unitary palp, and several genital characters, plus larval features including maxillary reduction and presence of only three instars.

Cyclorrhaphans, united by their pupation within a puparium formed by the last instar skin, include a heterogeneous aschizan group comprising Phoridae and Syrphidae (hover flies) and the Schizophora, defined by the presence of a balloon-like ptilinum that everts from the frons to assist the adult escape from the puparium. Higher flies include the ecologically very diverse acalypterates and blow flies and relatives (Calypteratae), treated here as sister groups (Fig. 8), but with alternative views suggested.

Hymenoptera (Wasps, Bees, Ants, Sawflies, and Wood Wasps). Hymenoptera contains at least 100,000 described species of holometabolous neopterans, varying from minute (e.g., Trichogrammatidae) to large-sized (0.15–120 mm in length) and slender (e.g., many Ichneumonidae) to robust (e.g., certain bees). The hymenopteran head has mouthparts ventrally directed to forward projecting, ranging from

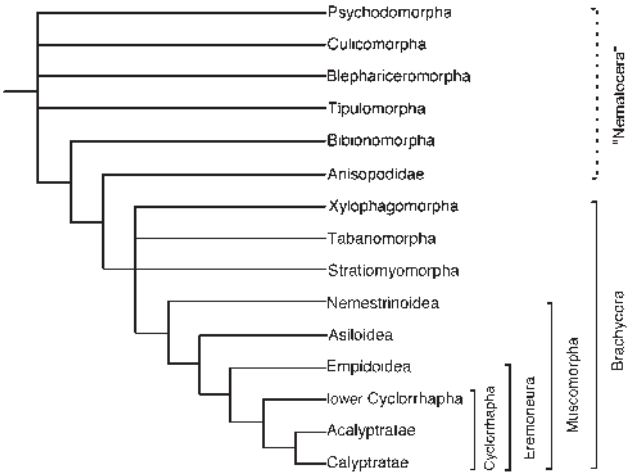

FIGURE 8 Cladogram depicting relationships among, and inferred classification of, Diptera. Dashed line indicates paraphyly in classification.

generalized mandibulate ("Symphyta"—woodwasps and sawflies) to sucking and chewing, with mandibles often used for killing and handling prey and, in Apocrita (ants, bees, and wasps), in defense and nest building. The compound eyes often are large; the antennae are long, multisegmented, and often prominently held forwardly or recurved dorsally. "Symphyta" has a conventional 3-segmented thorax, but in Apocrita the propodeum, abdominal segment 1, is included with the thorax into a mesosoma or, in ants, the alitrunk. The wing venation is relatively complete in large sawflies but is reduced in Apocrita in correlation with body size, such that very small species of 1 to 2 mm have only one divided vein or none. The hind wing has rows of hooks (hamuli) along the leading edge that couple with the hind margin of the forewing in flight. In Apocrita, the second abdominal segment (and sometimes also third) forms a constriction, or petiole. Hymenopteran female genitalia include an ovipositor, comprising 3 valves and 2 major basal sclerites, which may be long and highly mobile, allowing valves to be directed vertically between legs. The ovipositor of aculeate Hymenoptera is modified as a sting associated with a venom apparatus.

Symphytan larvae are eruciform (caterpillar-like) with 3 pairs of thoracic legs bearing apical claws and with some abdominal legs; most are phytophagous. Apocritan larvae are apodous, with the head capsule frequently reduced but with prominent strong mandibles; larvae may vary greatly in morphology during development (heteromorphosis). Apocritan larvae have diverse feeding habits and may be parasitic, gall-inducing, or fed with prey or nectar and pollen by their parent or, if a social species, other colony members. Adult hymenopterans mostly feed on nectar or honeydew; only a few consume other insects.

Hymenoptera is considered to form the sister group to Amphiesmenoptera (Trichoptera + Lepidoptera) + Antliophora (Diptera + Mecoptera/Siphonaptera), although a more basal position in the Holometabola has been advocated. Hymenoptera often are treated as containing two suborders, Symphyta (wood wasps and sawflies) and Apocrita (wasps, bees, and ants). However, Apocrita appears to be sister to one family of symphytan only, the Orussidae, and thus "Symphyta" form a basal paraphyletic group, whose basalmost clade is the Xyeloidea. This is sister to a monophyletic tenthrinoid (sawfly) clade, in turn sister to weakly supported Pamphiliodea [Cephoidea (possible grade Siricidae and relatives (Orussidae + Apocrita))] (Fig. 9).

Within Apocrita, aculeate (Aculeata) and parasitic (Parasitica or terebrant) wasp groups were considered each to be monophyletic, but Parasitica evidently is rendered paraphyletic by Aculeates originating from somewhere within Parasitica. Some traditional groupings also are nonmonophyletic, including Proctotrupoidea, but Proctotrupomorpha, comprising superfamilies Proctotrupoidea, Chalcidoidea, Platygasteroidea, and Cynipoidea, appears to be monophyletic. From morphology, Ichneumonoidea were argued to be sister to Aculeata, but molecular data and reanalysis refute this. The

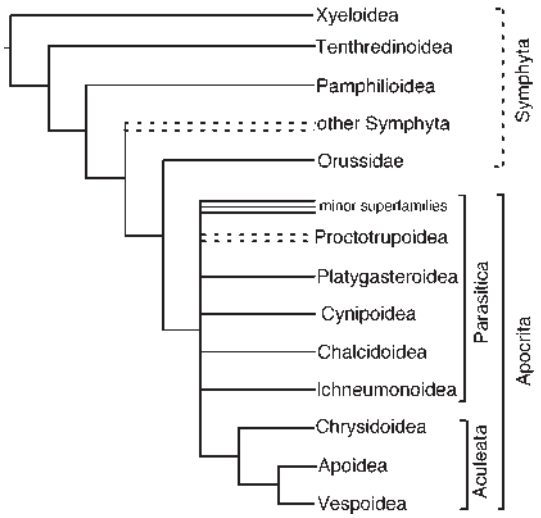

FIGURE 9 Cladogram depicting relationships among, and inferred classification of, Hymenoptera. Only major superfamilies are depicted. Dashed lines indicate uncertainty in cladogram or paraphyly in classification.

monophyletic aculeates, defined by their distinctive ovipositor construction and loss of female cerci, comprise Chrysidoidea (Vespoidea + Apoidea). Internal relationships of aculeates, including vespids (wasps), formicids (ants), and apids (bees), and the monophyly of subordinate groups are under scrutiny. Apidae evidently arose from within Sphecidae (digger wasps), but the precise relationships of another significant group of aculeates, Formicidae (ants), within Vespoidea is less well established.

Trichoptera (Caddisflies). Trichoptera contains about 45 extant families containing some 10,000 described species, with estimates of undescribed (mostly Southeast Asian) species diversity some four- to fivefold higher. Trichoptera are holometabolous; the moth-like adult has reduced mouthparts lacking any proboscis, but with 3- to 5-segmented maxillary palps and 3-segmented labial palps. The antennae are multisegmented and filiform and often as long as the wings. The compound eyes are large, and there are 2 or 3 ocelli. The wings are haired or less often scaled and differentiated from all but the most basal Lepidoptera by the looped anal veins in the forewing and absence of a discal cell. The larva is aquatic, has fully developed mouthparts, has 3 pairs of thoracic legs (each with at least 5 segments), and lacks the ventral prolegs characteristic of lepidopteran larvae. The abdomen terminates in hook-bearing prolegs. The tracheal system is closed and associated with tracheal gills on most abdominal segments. The pupa is also aquatic, enclosed in a retreat often made of silk, and possesses functional mandibles to aid in emergence from the sealed case.

Amphiesmenoptera (Trichoptera + Lepidoptera) is now unchallenged, despite earlier suggestions that Trichoptera arose from within Lepidoptera. Proposed internal relationships within the Trichoptera range from stable and well supported to unstable and anecdotal. Monophyly of the suborder

Annulipalpia in its strictest sense, comprising families Hydropsychidae, Polycentropodidae, Philopotamidae, and some close relatives, is well supported by larval and adult morphology—including the presence of an annulate apical segment of both adult maxillary and larval palp, absence of male phallic parameres, presence of papillae lateral to the female cerci, and, in the larva, presence of elongate anal hooks and reduced abdominal tergite 10.

The monophyly of the case-making Integripalpia (comprising families Phryganeidae, Limnephilidae, Leptoceridae, Sericostomatidae, and relatives) is supported *inter alia* by the absence of the *m* crossvein; hind wings broader than forewings, especially in the anal area; the female lacking both segment 11 and cerci; and larval character states including the usual complete sclerotization of the mesonotum, hind legs with lateral projection, lateral and middorsal humps on abdominal segment 1, and short and stout anal hooks.

Monophyly of a third putative suborder, Spicipalpia, is more contentious. Defined for the Glossosomatidae, Hydroptilidae, and Rhyacophilidae (and perhaps the Hydrobiosidae), the proposed uniting features are the spiculate apex of the adult maxillary and labial palps, ovoid second segment of the maxillary palp, and eversible oviscapt (egg-laying appendage). Morphological and molecular evidence fail to confirm Spicipalpia monophyly, unless at least Hydroptilidae is removed. All possible relationships between Annulipalpia, Integripalpia, and Spicipalpia have been proposed, often associated with scenarios concerning the evolution of case-making. An early idea that Annulipalpia are sister to a paraphyletic Spicipalpia + monophyletic Integripalpia finds support from some morphological and molecular data.

At lower phylogenetic levels, several interfamily relationships have been explored and congruent findings made, for example, concerning the families of Sericostomatoidea, including several from the landmasses formerly constituting Gondwana, including South Africa. In contrast, relationships of taxa associated with the family Leptoceridae vary dramatically between researchers. Forthcoming molecular evidence may be expected to assist in resolution of some issues and perhaps confuse otherwise robust relationships.

Lepidoptera (Moths and Butterflies). Lepidoptera, with some 140,000 described species in 70 families, is one of the major orders of Holometabola. Adults range from very small to large, with wings always covered in scales. The head bears a long, coiled proboscis formed from greatly elongated maxillary galeae; large labial palps usually are present, but other mouthparts are absent, except that mandibles are present primitively in some groups. The compound eyes are large, and ocelli frequently are present. The multisegmented antennae are often pectinate in moths and knobbed or clubbed in butterflies (Papilionoidea + Hedyloidea + Hesperioidea). The wings are completely covered with a double layer of scales (flattened modified macrotrichia), and the hind and forewings are linked by either frenulum, jugum, or simple overlap. Lepidopteran larvae have a sclerotized, head capsule with mandibulate mouthparts, usually 6 lateral ocelli, and short 3-segmented antennae. The thoracic legs are 5-segmented with single claws, and the abdomen is 10-segmented with short prolegs on some segments. Silk-gland products are extruded from a characteristic spinneret at the median apex of the labial prementum. The pupa sometimes is contained within a silken cocoon.

The basal radiation of this large order is considered well enough resolved to serve as a test for the ability of particular molecules to recover phylogenetic signal. Although over 98% of the species of Lepidoptera belong in Ditrysia (Fig. 10), the morphological diversity is concentrated in a small nonditrysian group. Three of the four suborders are species-poor (Micropterigidae, Agathiphagidae, Heterobathmiidae), lie sequentially at the base of the Lepidoptera, and lack the synapomorphy of the megadiverse suborder Glossata, namely the characteristically developed coiled proboscis formed from the fused galea. The highly speciose Glossata contains a comb-like basal series of species-poor taxa and a subordinate clade (Neolepidoptera) defined by the possession in the larva (caterpillar) of abdominal prolegs with muscles and apical crochets (hooklets). Much of the glossatan diversity is found in the derived Ditrysia, defined by the unique two genital openings in the female, one the ostium bursae on sternite 8, the other the genitalia proper on sternites 9 and 10. Additionally the wing coupling is always frenulate or amplexiform and not jugate, and the wing venation tends to be heteroneuran (with venation dissimilar between fore- and hind wings). Trends in the evolution of Ditrysia include elaboration of the proboscis (haustellum) and the reduction or loss of maxillary palpi. Relationships between the smaller superfamilies (not shown in Fig. 10) and the few highly diverse ones (Tineoidea, Gelechioidea, Tortricoidea, Pyraloidea, Noctuoidea, and Geometroidea) are not well understood and susceptible to change. However, one of the best supported relationships in Ditrysia is the grouping of Hesperioidea (skippers) and Papilionoidea (butterflies), united by their clubbed, dilate antennae, lack of frenulum in the wing, and large humeral lobe

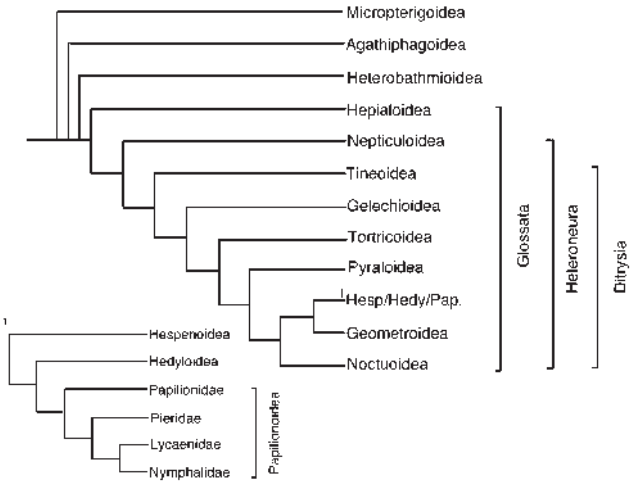

FIGURE 10 Cladogram depicting relationships among, and inferred classification of, Lepidoptera. Only selected superfamilies are depicted.

on the hind wing. To this has been added the neotropical Hedyloidea to form the monophyletic grouping we know as the butterflies.

See Also the Following Articles

Arthropoda and Related Groups • Biodiversity • Biogeographical Patterns • Fossil Record • Insecta, Overview

Further Reading

Bitsch, C., and Bitsch, J. (2000). The phylogenetic interrelationships of the higher taxa of apterygote hexapods. *Zool. Scripta* **29(2),** 131–156.

Kristensen, N. P. (1999). Phylogeny of endopterygote insects, the most successful lineage of living organisms. *Eur. J. Entomol.* **96(3),** 237–253.

Lyal, C. H. C. (1985). Phylogeny and classification of the Psocodea, with particular reference to the lice (Psocodea: Phthiraptera). *Syst. Entomol.* **10,** 145–165.

Ouvrard, D., Campbell, B. C., Bourgoin, T., and Chan, K. L. (2000). 18S rRNA secondary structure and phylogenetic position of Peloridiidae (Insecta: Hemiptera). *Mol. Phylogenet. Evol.* **16,** 403–417.

Ronquist, F., *et al.* (1999). Phylogeny of the Hymenoptera: A cladistic reanalysis of Rasnitsyn's (1988) data. *Zool. Scripta* **28,** 13–50.

Schultz, J. W., and Regier, J. C. (2000). Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean-hexapod clade. *Proc. R. Soc. Biol. Sci. B* **267(1447),** 1011–1019.

Sorensen, J. T., Campbell, B. C., Gill, R. J., and Steffen-Campbell, J. D. (1995). Non-monophyly of Auchenorrhyncha ("Homoptera"), based upon 18S rDNA phylogeny: Eco-evolutionary and cladistic implications within pre-Heteropterodea Hemiptera (s.l.) and a proposal for new monophyletic suborders. *Pan-Pacific Entomol.* **71,** 31–60.

Vilhelsen, L. (2000). Before the wasp-waist: Comparative anatomy and phylogenetic implications of the skeleto-musculature of the thoraco-abdominal boundary region in basal Hymenoptera (Insecta). *Zoomorphology* **119,** 185–221.

von Dohlen, C. D., and Moran, N. A. (1995). Molecular phylogeny of the Homoptera: A paraphyletic taxon. *J. Mol. Evol.* **41,** 211–223.

Physical Control of Insect Pests

Charles Vincent and Bernard Panneton

Agriculture and Agri-Food Canada, Quebec

Physical control is one of the four main approaches to crop protection against insects; the other three are the chemical, biological, and biopesticide approaches. From a theoretical and a technical point of view, all of these approaches have limits that make them more or less suitable against a given pest control target. In practice, the relative merits of each approach are also weighted against numerous factors before an actual decision is made regarding the most appropriate method to implement. A majority of agricultural commodities are protected using chemical control but ideally, all components and technologies should be blended optimally and harmoniously into an integrated pest management (IPM) program.

PHYSICAL CONTROL METHODS

Definition, Context, and Literature

Physical control methods in crop protection comprise all pest management techniques that rely on the use of physical processes to damage, kill, or induce behavioral changes in target organisms. The primary action may have a direct impact, for example, when insects are killed immediately by mechanical shocks. In other instances, the desired effect is attained through stress responses.

Various physical control methods have been used throughout the long history of plant protection (Table I). With the rapid advances that have occurred in the physical, chemical, and biological sciences since the late 19th century, agriculture has been transformed from a strictly empirical activity, largely based on tradition and aimed primarily at staying off famine, to a quantitative form of agriculture focused on producing a certain amount of food. During this transition, which has been sustained at an increasing rate over the past 50 years, physical control methods have been set aside because of the tremendous success of chemical control. It is only natural that some people should view the use of physical control methods as a step backward to those distant ancestral practices. Thanks to technological advances and greater precision in the implementation of such methods, physical control now has all the necessary attributes for incorporation into IPM strategies.

Use in Agricultural Production

The different methods of physical control used against crop pests have some common characteristics. One of the characteristics that differentiates physical tactics from the other control methods (Table I) is the absence of persistence. In almost every case, the effect of a treatment is limited to the period of application. When treatment stops, the stressor disappears immediately or dissipates quickly. From the standpoint of exercising control over the treatment and its secondary effects, the absence of a residual action is an advantage. However, this characteristic can also be regarded as a drawback, because the treatment may have to be repeated every few days to control crop pests that emerge and are active for a few days or a few weeks. In such cases, persistent chemicals constitute a much more convenient approach, although they are often undesirable from an environmental standpoint.

In addition to being restricted to the time of application, the impact of a physical control method is limited spatially. Mechanical, pneumatic, electrical, and thermal energies are dissipated locally over a distance of up to a few meters from the site of application. Electromagnetic radiation, which propagates over considerable distances and is subject to numerous restrictions (reserved frequency bands, maximum power, absence of interference), is an exception. Some pesticides have the unfortunate characteristic of dispersing over considerable distances. Similarly, many biological control agents can disperse or become dispersed beyond the treatment area.

TABLE I Comparison of Control Methods for Crop Protection

Characteristic	Method		
	Chemical	Biological	Physical
Advent	20th century	20th century	With agriculture
Registration	Required	A few cases	Never
Supporting sciences	Analytical chemistry, chemical synthesis, biology	Biology, biotechnology, ecology	Engineering (mechanical, electrical, electronic), biology
Scientific references	Very abundant	Abundant	Few
Residual action (residues and persistence)	Yes (variable)	Yes, if reproduction occurs	Negligible
Possibility of combining with another method	Yes (sometimes difficult with biological methods)	Yes	Yes
Active or passive method	Active	Active	Active and passive
Application to field crops	High level	Low level	Low to moderate level
Application to crops with a high profit margin per hectare	High	Moderate to high	Moderate to high
Safety for crop	Moderate to high (phytotoxicity)	High	High (passive) Low (active)
Labor requirements	Low	High	Medium to high
Work rate (hectares treated per hour)	High	Variable	Low (active) High (passive)
Site of action	Photosynthetic system, nervous system (few genes involved)	Systems allowing adaptation to biotic stresses	Systems allowing adaptation to abiotic stresses
Environmental or toxicological requirements, safety	High and costly	Moderate (e.g., virus)	Low (exception: electromagnetic radiation)
Geographic impact	Drift, run-off, evaporation, food chain	Colonization of nontarget habitats by parasites or predators	Restricted to area treated (exception: electromagnetic radiation)
Energy requirements	High for production	Low	Low (passive) High (active)
Machinery required	Ground or aerial sprayer	Little or none	Many types of equipment, few machines are suited to more than one purpose

Note. Reproduced, with permission, from Vincent, C., Panneton, B., and Fleurat-Lessard, F. (eds.) (2001). "Physical Control Methods in Plant Protection." Springer/INRA. Copyright Springer-Verlag.

Compared with traditional chemical control, present methods of physical control are more labor intensive and often time consuming (Table I). This drawback is one of the main reasons physical control techniques have had little success in penetrating the field crop market. Given these circumstances, only crops with a high profit margin per hectare represent an obvious market for physical control methods. From the viewpoint of implementation, physical methods compare favorably with biological methods (other than biopesticides), which often entail labor-intensive field observations and are difficult to apply in a field-crop setting.

Modes of Action and Classification

There are two basic types of physical control, i.e., active and passive (Table II). Active methods use some form of energy to destroy, injure, or induce stress in crop pests or to remove them from the environment. This type of approach has an effect at the time of application, with virtually no residual action. Passive methods, in contrast, cause changes in the environment and have a more lasting effect. Physical methods of control also can be classified according to the mode of energy use: thermal shock (heat), electromagnetic radiation (microwaves, infrared and radiofrequencies), mechanical shock, and pneumatic control (blowing or vacuuming tools) (Table II). Mechanical barriers to keep pests out, combined with physical suppression techniques, are the cornerstone of the approach adopted by industrialized countries to replace methyl bromide. Other techniques such as diatomaceous earth, hydrophilic particle films, sticky traps, and oils also are passive techniques.

Various applications that use thermal shock for crop protection in the field have been developed and research is in progress. This type of approach is based on the premise that the commodity or crop to be protected will be less sensitive than the target pest to an abrupt change in temperature. Research on thermal sensitivity thresholds and physiological reactions to short-duration thermal stresses is central to the development of control methods based on thermal shock.

Several avenues have been explored for applying the different forms of electromagnetic radiation as a tool for controlling

TABLE II Classes, Families, and Examples of Physical Control Methods

Class, family	Method	Examples of target insects	Situation	Comments
Passive				
	Physical barriers	Several mosquito species of medical importance, Colorado potato beetle, chinch bugs	Pr, PH	Efficacy limited to area
	Trapping	Apple maggot	Pr	In apple orchards
	Mulching	Aphids	Pr	
	Spatial barrier	Screwworm		
	Inert dusts	Stored-product insects	Pr	Including diatomaceous earth, silica dust
	Thin films (kaolin)	Codling moth, leafroller, mites	Pr	
	Mineral and edible oils	Phytophagous mites, scales	Pr, PH	
	Adhesives	Ants, cockroaches, flies		Glues, petroleum jelly, creosote
	Slippery surfaces	Ants, cockroaches		Made of Fluon, Teflon, or dust
	Fences	Chinch bugs, Colorado potato, beetle, cutworm	Pr	
	Windbreak	Aphid vectors		
	Wrapping	Gypsy moth, forest tent caterpillar, insects of transformed products		
	Organic mulch	Melon worm, Colorado potato beetle		
	Inorganic mulch	Aphid vectors, potato tuberworm		
	Trench	Chinch bugs, Colorado potato beetle,		
	Water	Chinch bugs		
	Modified atmosphere		PH	Modulating CO_2
	Inert gases	Stored-product insects	PH	
Active				
Mechanical	Mechanical impacts			
	Pneumatic (vacuuming, blowing)	Colorado potato beetle, *Lygus* spp.	Pr	
	Dislodging	Plum curculio	Pr	In apple orchards
	Leaf shredding	Spotted tentiform leafminer	Pr	
	Disturbing	Stored-product insects	Pr	
	Forced air	House flies		Building entrances
Thermal	Flaming	Colorado potato beetle	Pr	
	Hot/cool air	Stored-product insects	PH	
	Hot water–steam		PH	
	Rapid freezing	Stored-product insects	PH	
	Postharvest chilling	Stored-product insects	PH	
	IR heating		PH	
Electromagnetic	Microwave		Pr, PH	Also in museums
	Radiofrequencies		PH	
	Ionizing radiation		PH	
	UV and visible light	Mosquitoes		In combination with electricity
Other	Flooding	Cranberry insects	Pr	

Note. Pr, preharvest control; PH, postharvest control.

insects. Nonionizing electromagnetic radiation kills insects by raising their internal temperature. The utilization of radio, microwave, and infrared frequencies is based on a principle similar to that of thermal shock methods except that, with applications involving electromagnetic radiation, the transfer of energy occurs without using a heat transfer fluid. Technologies that harness electromagnetic radiation are often too expensive for use in the field. Furthermore, existing regulations restrict the available frequency bands, either for reasons of user and environmental safety or because certain frequency bands have been set aside for specific applications that do not tolerate interference (e.g., microwave-based landing guidance systems for aircraft).

There is a wide variety of physical barriers used as physical control techniques. The technology associated with physical barriers can be applied in the field or in greenhouses. Barriers used in the field can take several forms (trenches, vertical nets, etc.) and can be deployed on a range of scales to protect a complete field, a crop row, or a group of plants. Passive methods should be used whenever possible, because they extend the length of the treatment. For example, plastic-lined trenches along field boundaries trap Colorado potato beetles during the whole migration period.

Pneumatic control consists in using an airstream to dislodge insect pests. Insects that are removed by vacuum pressure are killed when they pass through the moving parts of the blower (mechanical shock). After being dislodged by a blowing device, individuals of some insect species are injured and die because they are unable to climb back onto the host plant. Other

machines are equipped with a device for collecting the dislodged insects, which are subsequently killed. Sound knowledge of the target insect's behavior is necessary in order to enhance the effectiveness of this type of approach as exemplified by the management of tarnished plant bug as a pest of strawberries.

Like any pest management approach, physical methods have strengths and weaknesses, and some of them are likely to have secondary effects on nontarget organisms, e.g., pollinators of strawberry. In an IPM context, the decision to use a physical control tactic must therefore be made on a case-by-case basis according to the same criteria as in decision-making regarding the appropriateness of pesticide applications: efficacy, cost-effectiveness, and undesirable effects. In addition, no physical control technique is sufficient on its own for all pest control treatments in a given crop.

Postharvest Physical Control

Postharverst situations offer ideal opportunities to research and implement physical control methods. In long-term storage of nonperishable agricultural commodities (seeds and grain, dried fruits, by-products, dried and dehydrated plants, spices, herbs, coffee, cocoa), the most serious losses are due to the action of insects and mites or the spoliage by certain microorganisms (e.g., fungi). Chemical control using persistent insecticides is the most commonly used approach for preventing damage to grain and seeds by insect pests. Some of the benefits of this strategy are low cost, ease of implementation, and protection that lasts for several months, until the quantity of active residues falls below the lethal threshold for the target species.

Physical control applications developed for postharvest treatments have focused on procedures for controlling physical conditions in stocks of stored grain (temperature and water content), thermal or mechanical shock, the establishment of extreme conditions for insect pests (anaerobiosis, pressure, and modified atmospheres), the use of abrasive or dehydrating minerals, and the erection of physical barriers to keep insects out. Postharvest control approaches are essentially based on passive methods, with the exception of thermal and mechanical shock treatments. In postharvest situations, most physical methods are suitable solely for prevention against pests and hence cannot be compared with classical chemical control. A thorough knowledge of integrated pest management strategies is required, since these techniques afford no protection following application, unlike persistent insecticides.

Practical use of physical control necessitates multiple verifications and supporting data, particularly in relation to secondary effects on the quality of treated products (for example, the germinating power of malting barley or the baking quality of bread-making wheat). Nonetheless, the prospect of registration reviews for several pesticides (e.g., methyl bromide) that are currently used in postharvest protection of raw foodstuffs or in processing and storage facilities for intermediate or finished products has revived interest in postharvest physical control. Furthermore, because the use of persistent contact insecticides on processed food products (e.g., wheat, semolina, dried fruit) is prohibited worldwide, the industry has to rely exclusively on fumigants or physical procedures to eradicate insects in such products. Fumigants are used only marginally in many countries such as France, the United Kingdom, and Canada, and their use is likely to decline further when methyl bromide production and use are phased out, sometime between 2002 and 2005. In view of this situation, physical control is the only strategy likely to ensure successful control of insect and mite pests in postharvest food stocks. Physical methods hold promise as a complement to chemical pesticides and a means of moving away from excessive reliance on chemical control. Furthermore, the most accessible physical methods for this pest management sector, such as dry heat treatments or airtight storage using inert gases, should help to diminish the secondary risk of spoilage of raw foodstuffs by storage molds.

GENERAL CONSIDERATIONS

Physical control deserves to be recognized as a well-defined area of expertise as is the case for biological control. This recognition is bound to come as the quest for alternatives to chemical pesticides intensifies. Although physical control went out of use with the advent of chemically based pest control methods in the middle of the 20th century, the limitations of pesticide use, coupled with the difficulties of implementing biological control, have created a crossroads for the renewed development of physical control.

Most physical methods of control can be used in a crop protection program incorporating both chemical and biological controls. A potential problem occurs when physical barriers are still in place during chemical or biological treatments. Chemical and biological methods are sometimes incompatible, particularly in production systems eschewing chemical pesticides. In the latter case, only biological and physical methods can be applied. Although not economically significant at present, organic crops represent a growing market segment. This is a niche that will definitely provide leverage for the development of physical control measures.

The regulatory framework for physical control differs markedly from that for agrochemical products (Table I). First, many physical techniques are subject to rules concerning their use (i.e., the registration process), which are designed to protect users and the general public. Sometimes, such as with the use of propane gas, specialized training is required. The use of electromagnetic radiation is constrained by telecommunications regulations, some of which stem from international agreements. In the case of microwave energy, for instance, only a handful of frequencies has been set aside for industrial, scientific, and medical applications. With regard to the regulatory framework for physical control technologies, it is completely defined *a priori.* In short, the equipment employed must meet the applicable standards (mostly related to user safety). With chemically or biologically based methods, the

difficulty in anticipating secondary effects precludes the establishment of comprehensive specifications that would be known *a priori.* This explains the need for increasingly costly test protocols designed to evaluate pesticide safety from the standpoint of human health as well as ecotoxicology.

A number of factors tend to complicate the implementation of physical control methods, and physical tactics cannot be readily compared with crop protection systems based solely on the use of an agricultural sprayer to apply pesticides in liquid form. For agricultural operations, this system entails low variable costs and fixed costs. In contrast, the equipment used for physical control is often very specific: cultivators for weeds, vacuuming device for Colorado potato beetles, and so on (Table I). Very few physical control tools offer the operational versatility that would allow them to perform several types of crop protection operations. Integration efforts, such as research aiming to design burners for use in controlling Colorado potato beetles, killing weeds, and performing top-killing, are needed to enable physical control tools to penetrate the crop protection market.

See Also the Following Articles

Agricultural Entomology • Biological Control • Colorado Potato Beetle • Integrated Pest Management

Further Reading

Banks, H. J. (1976). Physical control of insects—Recent developments. *J. Aust. Entomol. Soc.* **15,** 89–100.

Bostanian, N. J., Vincent, C., Chouinard, G., and Racette, G. (1999). Managing apple maggot, *Rhagoletis pomonella* (Diptera: Tephritidae), by perimeter trapping. *Phytoprotection* **80,** 21–33.

Chiasson, H., Vincent, C., and de Oliveira, D. (1997). Effect of an insect vacuum device on strawberry pollinators. *Acta Horticult.* **437,** 373–377.

Davidson, N. A., Dibble, J. E., Flint, M. L., Marer, P. J., and Guye, A. (1991). "Managing Insects and Mites with Spray Oils." Publication No. 3347, IPM Education and Publications, University of California, Oakland.

Fields, P. G., and Muir, W. E. (1995). Physical control. *In* "Integrated Pest Management of Insects in Stored Products" (B. Subramanyam and D. W. Hagstrum, eds.), pp. 195–222. Dekker, New York.

Glenn, D. M., Puterka, G. M., VanderZwet, T., Byers, R. E., and Feldhake, C. (1999). Hydrophilic particle films: A new paradigm for supression of arthropod pests and plant diseases. *J. Econ. Entomol.* **92,** 759–771.

Hallman, G. J., and Denlinger, D. L. (eds.) (1998). "Temperature Sensitivity in Insects and Application in Integrated Pest Management." Westview Press, Boulder, CO.

Jayas, D., White, N. D. G., and Muir, W. E. (1995). "Stored-Grain Ecosystems." Dekker, New York.

Korunic, Z. (1998). Diatomaceous earths, a group of natural insecticides. *J. Stored Prod. Res.* **34,** 87–97.

Lurie, S. (1998). Postharvest heat treatments. *Post Harvest Biol. Technol.* **14,** 257–269.

Nelson, S. O. (1996). A review and assessment of microwave emergy for soil treatment to control pests. *Trans. ASAE* **39,** 281–289.

Oseto, C. Y. (2000). Physical control of insects. *In* "Insect Pest Management, Techniques for Environmental Protection" (J. E. Rechcigl and N. A. Rechcigl, eds.), pp. 25–100. Lewis, Boca Raton, FL.

Panneton, B., Vincent, C., and Fleurat-Lessard, F. (2001). Plant protection and physical control methods: The need to protect crop plants. *In* "Physical Control Methods in Plant Protection" (C. Vincent, B. Panneton, and F. Fleurat-Lessard, eds.), pp. 9–32. Springer-INRA, Heidelberg.

Vincent, C. (2002). Pneumatic control of agricultural insect pests. *In* "Encyclopedia of Pest Management" (D. Pimentel, ed.), pp. 639–641. Dekker, New York.

Vincent, C., and Chagnon, R. (2000). Vacuuming tarnished plant bug on strawberry: A bench study of operational parameters versus insect behavior. *Entomol. Exp. Appl.* **97,** 347–354.

Vincent, C., Panneton, B., and Fleurat-Lessard, F. (eds.) (2001). "Physical Control Methods in Plant Protection." Springer/INRA, Heildelberg. [Translated from Vincent, C., Panneton, B., and Fleurat-Lessard, F. (eds.) (2000). "La Lutte Physique en Phytoprotection." Editions INRA, Paris.]

Vincent, C., Hallman, G., Panneton, B., and Fluerat-Lessard, F. (2003). Management of agricultural insects with physical control methods. *Annu. Rev. Entomol.* **48,** 261–281.

Phytophagous Insects

Elizabeth A. Bernays

University of Arizona

Phytophagous insects are generally considered to be those that feed on green plants. They include species that attack roots, stems, leaves, flowers, and fruits, either as larvae or as adults or in both stages. Leaf feeders may be external (exophytic) or they may mine the tissues, sometimes even specializing on a particular cell type. Typically nectar and pollen feeders are not included. "Phytophagous" is often synonymous with "herbivorous," although the latter term is sometimes restricted to those species feeding on herbs (i.e., herbaceous plants). Commonly, species that use only one plant genus or species are called monophagous and species that use plants within a tribe or family are called oligophagous. The term stenophagous is less commonly used and includes both of these. Polyphagous species are those that use plants in several to many plant families.

DIVERSITY

Phytophagous insects are highly diverse and the total species number is at least 500,000. This represents about 25% of known multicellular animals. There are phytophagous insect species in the majority of insect orders, including Orthoptera, Lepidoptera, Coleoptera, Heteroptera, Hymenoptera, and Diptera but there are very large differences in numbers of species in the different groups (Table I). All green plants are eaten by one or more species of phytophagous insects.

Major differences occur among orders in the ways in which hosts are selected and the life forms that use plants. Thus, grasshoppers usually lay eggs in the soil so that the nymphal stages must select the plant resource. In the Orthoptera and Hemiptera, which are hemimetabolous, the nymphal and adult stages generally have similar feeding habits. By contrast, among Lepidoptera and Hymenoptera adults commonly feed on pollen or nectar while only the larvae are phytophagous, and among Coleoptera the larvae

TABLE I Approximate Proportions of Species of Phytophagous Insects from Different Insect Orders

Order	Common example	Proportion of species in order that are phytophagous	Proportion of all phytophagous insect species
Orthoptera	Grasshoppers	95	3
Heteroptera	Bugs	90	14
Hymenoptera	Sawflies	11	9
Diptera	Flies	29	15
Lepidoptera	Moths and butterflies	99	26
Coleoptera	Beetles	35	28

and adults of leaf beetles have similar habits, while larvae of wood borers often have adults with different feeding habits. Thus, there is a dichotomy (with some notable exceptions) between species in which the mother chooses for her offspring and species in which all stages are independent.

The great diversity of insects feeding on plants is matched by a remarkable diversity of lifestyles, mouthparts and gut morphological adaptations to the food eaten, cuticular morphology and coloration adapted for crypsis or aposematism, and behavioral adaptations for use of particular plants and escape from natural enemies. Many of the remarkable pictures of insects in popular journals involve surprisingly effective visual and behavioral crypsis.

Insects feeding on plants had their origin early in the history of life on land, with all the major orders that feed on plants today being present 300 million years ago (with the possible exception of Lepidoptera). This means that partitioning of food resources had occurred then, because there were spore feeders, sap suckers, and gall makers as well as miners and external feeders. This can be determined from information provided by fossils of insects and damaged plants and fossilized remains of insect feces. The diversity of insect mouthparts required for the different feeding guilds was established well before the appearance of angiosperms 200 mya, but diversification of families, genera, and species has continued unabated ever since. The diversification is now known to be clearly related to angiosperm diversity and believed to be related largely to the diversity of plant secondary metabolites. These chemicals provide the signals for acceptance or rejection of potential host plants.

Mouthparts and Feeding

Although each of the major phytophagous orders of insects has distinctive biting and chewing or sucking mouthparts, the structures are highly diversified to handle every type of physical problem. For example, caterpillar species feeding on different plants can often be recognized by their mandibular morphology, and species feeding on physically similar plant parts tend to have similar morphology, whether they have a common ancestry or not. Among grasshoppers (Acrididae), those that feed only on grasses have highly specialized mandibles with incisor regions for snipping through the tough parallel veins and molar regions for grinding the tissue. These highly characteristic mandibles have evolved independently at least eight times during the evolution of grasshoppers. It is probable that the details of mouthpart structure evolve quite quickly to suit changing diet, because it has been found in certain seed-sucking bugs (e.g., Jadera) that beak length has evolved to suit different fruit sizes within the past 100 years. Such rapid evolution may reflect the need to process food efficiently to maximize growth as well as the need to ingest food very rapidly to minimize predation risk.

Unless they utilize such structures as seeds and pollen, most phytophagous insects deal with the low protein levels characteristic of much plant tissue by eating relatively large amounts, and the gut throughput rates are high, with food in some cases taking only a couple of hours to pass through the digestive system. Some supplement their diets with carnivory or use symbionts to upgrade the levels of essential amino acids. Aphids, for example, which feed on phloem, tend to be particularly short of certain essential amino acids such as tryptophan, and their symbiotic bacteria commonly have multiple copies of genes involved in making tryptophan, so that the aphid obtains its requirements with the help of the symbionts.

Unlike many vertebrate herbivores, insects that feed on plant tissues often do not make nutritional use of the cellulose, which makes up a large proportion of plant bulk. This may partly reflect the fact that for phytophagous insects, protein is more likely to be limiting than carbohydrate. As heterotherms (animals whose body temperature varies with that of the environment) they are not concerned with the use of diets that are high in calories for maintenance of body temperature. Those species that do use cellulose, such as beetles feeding on wood or termites feeding on dead and decaying plants, often have symbionts that break down the cellulose, releasing nonprotein amino acids as well as sugars. It is possible that the digestion of cellulose in these cases is a mechanism for releasing nonprotein amino acids that are bound to cellulose but that can be converted to useful amino acids for the insects by the resident symbionts.

Apart from the need to obtain sufficient quantities of major nutrients such as protein, insects (like other animals) often require nutrients in suitable ratios. For example, the proportions of protein and carbohydrate required for maximal growth vary among taxa; if a particular resource is limited in one respect the balance can be improved behaviorally. Thus, individual insects with a choice of high-protein/low-carbohydrate food

and low-protein/high-carbohydrate food can eat a mixture of both. Studies indicate that species in different taxonomic groups are able to select among the foods available to optimize the balance ingested. This ability depends on a variety of mechanisms, including nutrient feedbacks that influence taste receptor sensitivity to sugars and amino acids, a tendency to move away from a food that has recently proved unsuitable, and a tendency to select foods with new and different flavors following experience on an unsuitable food. In addition, it has been demonstrated that insects can learn to avoid unsuitable food and to associate particular odors with high-quality foods.

THE IMPORTANCE OF PLANT SECONDARY METABOLITES

The great diversity of secondary metabolites in plants profoundly affects the behavior of phytophagous insects and, thus, the evolution of that behavior. These compounds may be repellent, or deterrent after contact. In addition, the deleterious postingestive effects of them enable insects to learn to reject a plant. Such food aversion learning has been demonstrated particularly in grasshoppers.

Many plant secondary metabolites serve as relatively nonspecific attractants or feeding stimulants for insects that feed on plants, although more commonly one or several particular compounds in a plant species serve as highly specific attractants or stimulants for feeding/oviposition by insects adapted to that plant. Thus, specialist phytophagous insects generally have a genetic predisposition to accept plants with a particular chemical or suite of chemicals present, the so-called sign stimuli; indeed, the sign stimuli may act as valuable signals in the sense of improving the speed of decision-making and discrimination by these insects. Some examples of sign stimuli are shown in Table II.

Apart from Orthoptera (specifically grasshoppers), the majority of phytophagous insect species tend to be specialists.

TABLE II Examples of Sign Stimuli That Are Particularly Important in Host Recognition by Phytophagous Insects

Insect	Diet breadth	Chemicals
Junonia coenia, buckeye butterfly	Several families	Iridoid glycosides
Plutella xylostella, diamondback moth	Family Brassicaceae	Glucosinolates
Pieris rapae, imported cabbage worm	Family Brassicaceae	Glucosinolates
Uresiphita reversalis, genista caterpillar	Tribe Genisteae	Quinolizidine alkaloids
Delia antiqua, onion maggot	Genus *Allium*	Disulfides
Chrysolina brunsvicensis, beetle	Genus *Hypericum*	Hypericin (quinone)
Plagioderma versicolora, willow beetle	Genus *Salix*	Salicin (phenolic glycoside)

That is, they feed on just a few species, genera, or tribes of plants. It is common among Lepidoptera (moths and butterflies), for example, to find species that feed on plants in one family, one tribe, one genus, or one species of plant. It appears that the degree of specialization is related to the occurrence of one or a few chemicals that characterize that plant group. In addition, the narrower the host range, the more likely it is that non-host chemicals will repel or deter the insects at relatively low concentrations. Many of these chemicals that reduce feeding or oviposition behavior are not noxious if ingested, suggesting that their role in these cases is more as signals of non-hosts than as signals of toxicity in specifically evolved plant defenses. Nonetheless, there are situations in which plants probably evolved high concentrations of particular chemicals in response to the attack of insects. Insects, over time, would be likely to evolve countermeasures. Thus one can envisage, as many have done, that a chemical arms race between plants and phytophagous insects has occurred (and is occurring).

Diet breadth variation occurs in all phytophagous insect orders, and there is evidence from molecular and other studies that evolution of diet breadth can occur in either direction. What drives these changes has been a subject of much controversy. Included in the hypotheses are the following: arms race coevolution in which specialists are in some way more successful, sequential evolution of insects in which species benefit from the use of specific plant signals to improve behavioral efficiency, adoption of specific plant hosts from which specialists may sequester high levels of particular protective chemicals, and the selection pressure of parasites and predators that involve adoption of limited host plant species as a means of better avoiding attack (for example, by enhanced visual or chemical crypsis).

The study of host plant selection by phytophagous insects has been important in theories of resource use and whether it should be flexible in ecological or evolutionary time. For example, a change in host use may involve a change in specificity (how many different resources are acceptable) or a change in preference (which of a limited number of available resources is ranked highest). A change in specificity could result from a simple change in gustatory or olfactory sensitivity to many plant secondary metabolites or a change in the central nervous system affecting the relative importance of negative inputs from chemoreceptors. A change in preference, however, probably involves a change in receptor conformation or proportions of receptors with different conformations, at the level of the sensillum.

VARIATION IN HOST USE

Advances in how evolutionary changes occur or have occurred are being studied by examining genetic variation currently seen within populations of particular species, geographic variation that occurs in host specificity or preference, and historical changes in host use together with the physiological mechanisms underlying them. Experiential change can

influence host preference, in turn altering the nature of selection pressure on those individuals and their offspring.

Host shifts or increases in host range have provided models of evolutionary change including the study of sympatric speciation, whereby populations of an insect species become associated with different hosts. Because many species mate on or near their hosts, gene flow between populations using different hosts drops and speciation becomes possible. An example of this kind of divergence is with the apple maggot *Rhagoletis pomonella* in North America. Some populations moved from the ancestral *Crataegus* to apple and, currently, populations are diverging and evolving additional differences.

The interaction of phytophagous insects and plants is greatly influenced by predators and parasites. This is because they are major mortality factors for the phytophagous insects and yet use the plants themselves for sources of nectar and places to shelter, as well as places where they find their prey and hosts. In many cases, they are attracted to the odors of plants that are being attacked by plant-feeding species. The phytophagous insects that can remain visually or chemically cryptic by being specialists, or sequester plant toxins and become warningly colored, will be selected for by these natural enemies.

AGRICULTURE

The study of phytophagous insects has been very important in agriculture. Probably from ancient times, humans have selected varieties of crop plants that are minimally attacked by insects, and in the last 100 years breeding programs have been important in specifically increasing plant resistance. For example, resistance in rice to the rice brown planthopper *(Nilaparvata lugens)* and resistance in wheat to the greenbug *(Schizaphis graminum)* have resulted from dedicated research effort. Today, genes that express resistance factors against particular insect pests are inserted into some crops. For example, a toxic protein from the insect disease bacterium *Bacillus thuringiensis* can be produced in plants genetically modified by introduction of the bacterial gene.

Research in all areas of the biology of phytophagous insects has found application in agriculture. For example, behavioral studies of attractants has led to the use of traps for specific pests, and the study of antifeedants has had use in development of such materials in crop protection. Plant resistance is sometimes indirect, with plants expressing characteristics that favor natural enemies of the phytophagous species. For example, some crop plant varieties have hairs distributed at such densities that the parasitoids of pest whiteflies are slowed down in their running on a plant to a speed that improves host recognition. In other crops, surfaces lacking a wax bloom enable small predators to run more easily and thus efficiently find small caterpillars.

See Also the Following Articles

Coevolution • Digestion • Feeding Behavior • Genetically Modified Plants • Host Seeking, for Plants • Mouthparts • Symbionts Aiding Digestion

Further Reading

Bernays, E. A. (ed.) (1989–1993). "Insect–Plant Interactions," Vols. I to V. CRC Press, Boca Raton, FL.

Bernays, E. A. (1998). Evolution of feeding behavior in insect herbivores. *BioScience* **48,** 35–44.

Bernays, E. A. (2001). Neural limitations in phytophagous insects: Implications for diet breadth and evolution of host affiliation. *Annu. Rev. Entomol.* **46,** 703–727.

Bernays, E. A., and Chapman, R. F. (1994). "Host-Plant Selection by Phytophagous Insects." Chapman & Hall, New York.

Farrell, B. D., Mitter, C., and Futuyma, D. J. (1992). Diversification at the insect–plant interface. *BioScience* **42,** 34–42.

Fritz, R. S., and Simms, E. L. (eds.) (1992). "Plant Resistance to Herbivores and Pathogens." Chicago University Press, Chicago.

Panda, N., and Khush, G. S. (1995). "Host Plant Resistance to Insects." CAB Int. Wallingford, UK.

Price, P. W., Lewinsohn, T. M., Wilson Fernandes, G., and Benson, W. W. (eds.) (1991). "Plant–Animal Interactions." Wiley–Interscience, New York.

Rosenthal, G. A., and Berenbaum, M. R. (eds.) (1991). "Herbivores: Their Interactions with Secondary Plant Metabolites," 2nd ed., Vols. I and II. Academic Press, New York.

Schoonhoven, L. M., Jermy, T., and van Loon, J. J. A. (1998). "Insect–Plant Biology: From Physiology to Evolution." Chapman & Hall, London.

Stamp, N. E., and Casey, T. M. (eds.) (1993). "Caterpillars: Ecological and Evolutionary Constraints on Foraging." Chapman & Hall, New York.

Strong, L., Lawton, J. H., and Southwood, R. (1984). "Insects on Plants: Community Patterns and Mechanisms." Harvard University Press, Cambridge, MA.

Phytotoxemia

Alexander H. Purcell
University of California, Berkeley

The feeding of insects and mites, especially those that suck fluids from plants, can cause symptoms of distress such as yellowing (chlorosis), silvering, bronzing or necrosis of foliage, wilting or discoloration of shoots, and malformation of leaves, stems, roots, fruits, and other plant organs or tissues. Phytotoxemia (or phytotoxicity) is the production of plant symptoms of distress caused by the reaction of plants to chemicals (toxins) produced by insect feeding.

DEFINITION AND IMPORTANCE

Insects damage plants in many ways, but the most common type of damage is the removal of plant tissues, as in the familiar examples of caterpillars, beetles, and grasshoppers, whose feeding creates noticeable holes or even the wholesale removal of leaves, fruits, or other plant parts. Minute arthropods such as mites and thrips remove plant parts on a smaller scale by evacuating the contents of individual plant cells by their feeding. In contrast to damage caused by the removal of plant tissue, a phytotoxemia or phytotoxic reaction is the reaction of a plant to a chemical toxin introduced by insect feeding. Sucking insects in the orders

Hemiptera (part of which has been called Homoptera in the past), such as plant bugs, aphids, leafhoppers, and psyllids, are the best known insects that cause phytotoxicity by their feeding. Examples of external phytotoxins include chemical air pollutants and pesticides. Abnormal plant growths or deformities called galls can be caused by certain insects.

The symptoms or appearance of insect phytoxemias are often very similar to symptoms caused by infections with plant viruses or other infectious microorganisms that cause plant disease. Phytotoxemias can be distinguished from diseases caused by pathogens by the following criteria: (1) the capacity of the insect to produce symptoms is characteristic of the species, not restricted to certain individuals within the species; (2) the incidence and severity of symptoms are directly related to the number of insects (even where a single insect may cause symptoms) and the amount of time spent feeding; and (3) plants recover at least partly from further development of symptoms when the insects are removed. In contrast, plant viruses transmitted by insect vectors are not present in all individuals in a vector species; each individual must acquire or inherit the virus. Furthermore, plants develop symptoms of viral disease as a result of infection introduced by a single vector insect that might have transmitted the virus to the plant. Finally, plants will not recover from virus diseases when the vectors are removed from the plant.

The appearance of a phytotoxemia can be small, localized lesions of discolored or dead (necrotic) tissue; the formation of cork, scabs, or pits; premature fall of leaves or fruits; or curling or malformation of plant organs. These are often symptoms of plant diseases caused by pathogenic microorganisms and thus may be easily confused with symptoms caused by pathogens. Symptoms may be systemic, that is, far removed from the site of insect feeding. For example, leaves may be discolored or distorted by phytotoxic feeding in roots or stems.

EXAMPLES AND CAUSES OF INSECT PHYTOTOXICITY

The causes of phytoxemias are not well understood. Mechanical damage to plant tissues caused by sucking insects is not as easy to observe as for chewing insects because the damage occurs at the cellular level and is internal. For example, hopperburn is a condition caused in a variety of plants by small leafhoppers (family Cicadellidae) in the genus *Empoasca.* The most widely studied type of hopperburn is in alfalfa and is caused by the potato leafhopper *Empoasca fabae.* Laceration of the vascular tissues of the plant disrupts the flow of water and plant sap to leaves, but there is also evidence that chemicals within the leafhoppers' saliva induce yellowing and necrosis typical of hopperburn. Simulation of leafhopper feeding with a fine needle causes some but not all symptoms of hopperburn. However, introducing crushed salivary glands of the potato leafhopper more closely simulates hopperburn. Other species of *Empoasca* cause hopperburn in beans in South America *(E. kreameri)* and grapes in Europe and North Africa *(E. lybica).* The phytotoxicity of the leafhopper *Sophonia rufofascia,* which

FIGURE 1 The myrtle bushes on the left are damaged by the feeding of the leafhopper *S. rufofascia* (inset). (Photograph used with permission of Vincent Jones, Washington State University. Insect photograph by Walter Nagamini, Hawaii Department of Agriculture.)

became established in Hawaii in the 1980s, caused extensive damage to many native plants in Hawaii, threatening the continued survival of some rare species. Its phytotoxicity is similar in many respects to hopperburn: halted growth and necrosis in some plants, yellowing of leaf tissues, and stunting of new growth in other plants (Fig. 1).

The grape phylloxera *(Daktulosphaira vitifoliae)* is a devastating pest of grapevines that ruined millions of hectares of European vineyards in the 19th century and continues today to be a serious pest of vines worldwide. The aphid-like phylloxera have a complex life cycle, with some forms (gallicoles) causing galls on leaves and other forms (radicoles) causing galls and necrosis in roots. Phylloxera damage to roots kills European vines unless they are grafted onto rootstocks of wild grape species native to North America or onto hybrids of tolerant wild species (Fig. 2). Resistant grape species may have fewer phylloxera on their roots than European grapes, but it is their tolerance of grape phylloxera infestation without damage that is the main basis of their usefulness in avoiding phylloxera phytotoxicity. Feeding causes small roots to develop deformities and also kills root cells just beneath the bark. Soil-inhabiting fungi that invade roots may promote the decline of phylloxera-damaged vines.

The feeding of the silverleaf whitefly *(Bemisia argentifolii)* causes silvering of the leaves of squash plants and reduces fruit size and color. Its feeding also causes tomatoes to ripen more slowly than normal. Unidentified chemical components of whitefly saliva seem to be involved, but the plant's reaction is genetically controlled. The potato psyllid *(Paratrioza cockerelli)* causes a condition of potato and tomato called psyllid yellows, which, in addition to causing foliage of affected plants to become yellowish, reduces the size and quality of the fruits. The immature stages of potato psyllid rather than adults cause the most damage. Because individual psyllids may vary markedly in their ability to cause damage and only a small number of psyllids are sufficient to cause symptoms, a virus was suspected

FIGURE 2 Dead grapevines attest to the severity of phylloxera feeding causing damage to roots. The healthy vines next to the declining vines have only the initial stages of phylloxera damage to roots. (Photograph by A. H. Purcell.)

to be involved in causing psyllid yellows. However, other criteria, such as the direct relationship between damage and the number of psyllids feeding on a plant and the disappearance of symptoms after removing the psyllids, provide evidence against involvement of a virus or other pathogen.

Other insects that cause phytotoxemias include species of aphids, mealybugs, planthoppers, treehoppers, spittlebugs, mirid bugs, coreid bugs, and stink bugs. Blister or gall mites (family Eriophyiidae) are also well known for inducing abnormal growth in many uncultivated plants as well as ornamentals and some crop plants.

See Also the Following Articles

Gallmaking • Plant Diseases and Insects • Salivary Glands

Further Reading

Carter, W. (1973). "Insects in Relation to Plant Disease," 2nd ed. Wiley, New York.

Cohen, S., Duffus, J. E., and Liu, H. Y. (1992). A new *Bemisia tabaci* biotype in the southwestern United States and its role in silverleaf of squash and transmission of lettuce infectious yellow virus. *Phytopathology* **82,** 86–90.

Dixon, A. F. G. (1994). Insect-induced phytotoxemias: Damage, tumors, and galls. *In* "Current Topics in Vector Research," (K. F. Harris, ed.). Springer-Verlag, Berlin.

Granett, J., Walker, M. A., Koesis, L., and Omer, A. D. (2001). Biology and management of grape phylloxera. *Annu. Rev. Entomol.* **46,** 387–412.

Jones, V. P., Anderson-Wong, P., Follett, P. A., Yang, P., Westcot, D. M., Hu, J. S., and Ullman, D. E. (2000). Feeding damage of the introduced leafhopper *Sophonia rufofascia* (Homoptera: Cicadellidae) to plants in forests and watersheds of the Hawaiian Islands. *Environ. Entomol.* **29,** 171–180.

Plant Diseases and Insects

Alexander H. Purcell

University of California, Berkeley

Insects and mites can cause plant diseases directly through feeding or can transport and inoculate viruses and microorganisms such as bacteria and fungi that cause plant disease. Insects are the most important vectors of plant viruses and the main or sole means of spread of many plant pathogens. Plant diseases spread by insect vectors can be crucially important to the profitable production of some crops. Insecticidal or biological control of vectors often does not control the diseases spread by some vectors; consequently, physiological and ecological relationships between vectors and the pathogens they transmit are important to understand.

The direct damage to plants caused by insect feeding (herbivory) generally is considered to be in a separate category from plant disease. However, some insect feeding causes plant responses (phytotoxic reactions or gall formations) that are very similar in appearance to plant diseases caused by microorganisms and may be difficult to distinguish from diseases caused by microbial pathogens such as viruses or fungi.

VECTOR TRANSMISSION OF PLANT VIRUSES

Arthropod transmission of plant viruses illustrates the complexity and variety of relationships between plant pathogens and the arthropods that transport and introduce viruses to the plants. Because plants and plant viruses cannot move by themselves, plant viruses usually have mobile vectors. Most plant virus vectors are insects, mites, or nematodes. Arthropods are not thought to be important in the spread of the smallest plant pathogens—viroids (infectious small ribonucleic acids or RNAs); these are transmitted by gall mites (family Eriophyidae). However, they are important in the transmission of viruses. Viruses consist of protein-coated nucleic acids [either RNA or deoxyribonucleic acid (DNA)] that provide the genetic information for host cells to generate new copies of the virus (replication). In some cases, the viral coat contains lipids or glycoproteins.

Vector transmission of a pathogen is usually characterized with respect to vector efficiency or competence. Efficiency is usually estimated on the basis of how likely transmission is to occur during each opportunity that a vector has for transmission. Usually, this is estimated by determining how many individuals of a particular species are able to transmit a pathogen to plants during a given time interval. An important aspect of transmission efficiency is that only a single insect species is known to transmit some viruses. Such viruses are said to be highly vector specific. Other viruses have less vector specificity; for example, many species of insects within a family or subfamily may be vectors of a particular virus. Such

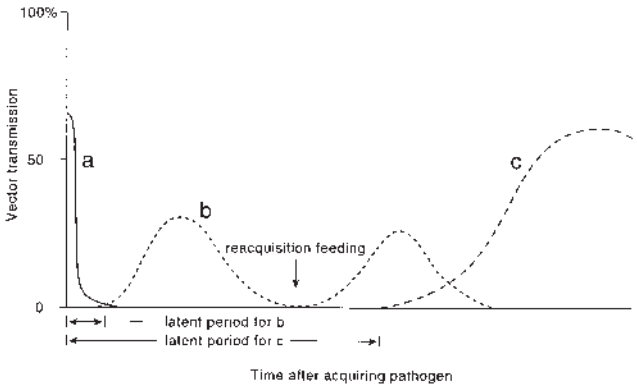

FIGURE 1 Vector transmission efficiency changes over time after acquisition. (a) Nonpersistent transmission; (b) persistent transmission; (c) persistent (over weeks), circulative, or propagative transmission. The latent periods for b and c are indicated. (Reproduced from Daly, Doyen, and Purcell, 1998, "Introduction to Insect Biology and Diversity," with permission of Oxford University Press, Inc., New York.)

viruses are classified as having group specificity or low vector specificity, respectively.

Transmission efficiency often changes dramatically over time. For example, vectors such as the green peach aphid *(Myzus persicae)* transmit viruses such as potato virus Y (PVY) most efficiently within seconds after acquiring the virus from a plant. In addition, after only a few minutes or at most hours of feeding on plants, the aphids (family Aphididae) no longer transmit the virus to plants. In fact, even the aphid's efficiency of acquiring PVY varies with time. Aphids that are fed only for a brief interval on the virus-infected source plant more frequently transmit virus than do aphids that are fed for longer intervals on the source plant. Aphid transmission of PVY is said to be nonpersistent, meaning that virus transmission rapidly declines after acquisition (Fig. 1a), although air-borne aphids lose transmission efficiency at much slower rates than aphids that are feeding on plants or even probing an inert surface such as glass. Nonpersistently transmitted viruses generally have low vector specificity; that is, many aphid species can transmit them. At the other extreme, the green peach aphid transmits potato leaf roll virus (PLRV) only after an interval of many hours or even days after it acquires PLRV from feeding on virus-infected plants; it then continues to transmit for many days (persistent transmission). Only a few aphid species can transmit PLRV. The time required between the vector's acquiring the virus and its successful introduction (inoculation) of the virus to a plant is called a latent period (LP).

Changes in vector transmission efficiency over time (Fig. 1) provide clues as to the nature of the relationship between the vector and the virus and help to explain how vectors transmit the virus. For example, the persistent type of transmission typified by aphid transmission of PLVR often results from the virus having to circulate within its aphid vector before it can be transmitted, explaining the delay or LP required for transmission. A plant virus, such as the aphid-transmitted lettuce necrotic yellows virus (LNYV), that must multiply within a vector before it can be transmitted will typically have a median or average LP of days or even weeks. LNYV is also persistently transmitted by its aphid vectors after the LP is completed.

In contrast, nonpersistently transmitted viruses such as PVY seem to be carried in or on the needle-like mouthparts of its aphid vectors. Many such viruses produce a viral-encoded polypeptide or small protein "helper factor" that is thought to act as a bridge to aid the attachment of the virus to the aphid vector's mouthparts. The helper factor of one virus may also act as a helper factor for a different virus. Certain viruses may require the presence of another virus in the same plant to be transmitted to another plant. It is not known if the assisting virus or an extraviral helper factor provided by the assisting virus is what aids transmission of the dependent virus. For example, rice tungro disease is caused by the rice tungro bacilliform virus (RTBV), which can be transmitted only by its vector, the rice green leafhopper (*Nephotettix cincticeps,* family Cicadellidae), along with another virus, the associated rice tungro spherical virus (RTSV). By itself, RTBV can cause tungro disease but cannot be transmitted to other plants, and RTSV by itself does not cause a disease in rice.

Some viruses, such as maize chlorotic dwarf virus (MCDV), persist for only hours to days in the blackfaced leafhopper, *Graminella nigrifrons,* vector and are classified as semipersistently transmitted. The shedding of the vector's exoskeleton during molting from one growth stage to another halts the transmission of nonpersistently and semipersistently transmitted viruses. Because the lining of the foregut is shed during molting, PVY and MCDV are thought to be transmitted to plants from a location within the mouthparts or foregut of the vector. Electron microscopy and the labeling of viruses with fluorescently or colloidal gold-tagged antibodies that bind to specific viral proteins have been used to identify areas where viruses accumulate or attach within the foregut. The same approach can be used for viruses that circulate within the vector's body cavity.

The circulative viruses, such as luteoviruses, that are transmitted by aphids and gemini viruses that are transmitted by leafhoppers or whiteflies do not appear to multiply within their vectors; instead, their transmission is thought to entail efficient methods of viral uptake and translocation within the vector's hemocoel (body cavity). Luteoviruses appear to be taken up by a process of endocytosis, in which virus particles are engulfed in a portion of the external cell membrane of intestinal cells, transported into the cell, and expelled from the cell into the body cavity. By processes that are less well understood, viruses can penetrate the membranes surrounding the salivary glands of the vector. Virus particles then enter plants as a result of the vector's salivation while feeding on plants. Luteoviurses such as PLRV can bind to the major protein (symbionin) associated with bacteria (called symbionts) that that live in specialized tissues within aphids and provide required nutritients to their aphid hosts (Fig. 2). The attachment of virus particles to the symbionin molecules may aid in the efficient circulation and persistence of virus from its entry via the aphid's digestive tract

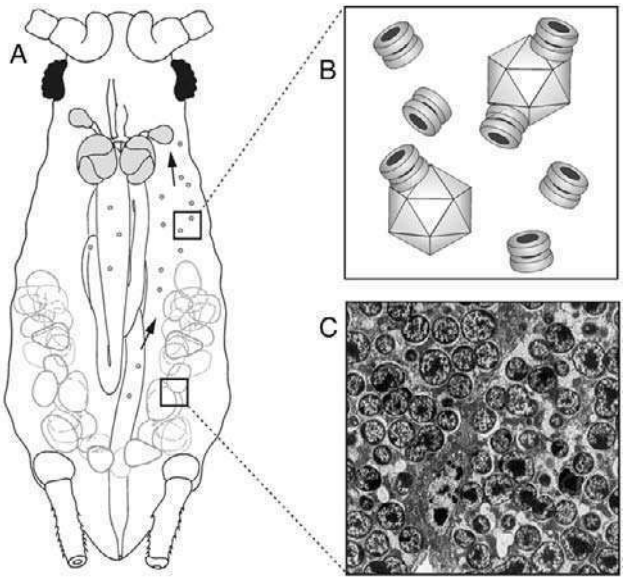

FIGURE 2 The symbionin protein of symbiotic bacteria within aphids binds luteoviruses in the hemolymph of aphids carrying the virus. [From Van den Heuvel *et al.,* (1999) *Trends in Microbiology* **7,** 71–76. Reproduced with permission from Elsevier Science.]

to its entering the accessory lobes of the aphid's salivary glands. Disruption of the symbionts with antibiotic chemicals greatly reduces aphid transmission efficiency of the luteovirus.

Experimental manipulations can result in a "heterologous" virus made up of the genetic information (DNA) of one virus encased in the protein coat of another virus. A heterologous virus consisting of the DNA of a whitefly-transmitted virus encased in the coat of a leafhopper-transmitted virus can be transmitted by leafhoppers from plants that contain the heterologous virus to a new plant. However, once the virus begins to replicate in the plant inoculated by the leafhopper, it constructs the proper coat protein directed by the DNA of the whitefly-transmitted virus. Leafhoppers no longer can transmit this virus from plant to plant; only whiteflies (family Aleyrodidae) transmit it. Experiments of this sort demonstrate that the viral protein coat, not the viral genome, determined the vector specificity of the virus. The protein coat probably does this by mediating passage of viral particles through the midgut and the salivary glands of the leafhopper, even though the viral DNA encoded whitefly-transmitted virus. Once the whitefly-transmitted virus replicated and produced its corresponding protein coat, only whiteflies could transmit it from plant to plant.

Some of the persistently transmitted plant viruses, such as the reoviruses, replicate within their insect vectors. It is remarkable that the same virus can subvert the genetic and protein processing systems of both plant and animal cells for viral replication. Although vectors normally acquire viruses by feeding on virus-infected plants, some plant viruses can invade the developing eggs or embryos within a female vector insect. The rice dwarf virus (RDV) is an example of a reovirus that passes from virus-infected female rice green leafhoppers to their offspring by this route. Invariably these transovarially transmitted viruses, such as RDV, multiply within their vectors, yet most viruses that multiply in vectors are not transmitted transovarially. The tomato spotted wilt virus (TSWV) is unusual because it can be acquired only by immature stages of thrips. An adult thrip can transmit this virus only if it fed as a nymph on a plant with TSWV. Thus, the vectors of TSWV have vector specificity not only for particular vector species but also for the immature stages, at least in the acquisition phase.

Insects that feed on plants' vascular tissues (xylem and phloem) appear to be the most common vectors of plant viruses. The greatest number of insect vectors of plant viruses are sucking insects in the order Hemiptera. Within the Hemiptera, aphids (Aphididae) transmit the greatest number of different plant viruses, followed by whiteflies (Aleyrodidae), leafhoppers (Cicadellidae), and planthoppers (Fulgoroidea). Mealybugs (Pseudococcidae) and various other hemipteran, families have species that are virus vectors as well. Thrips (order Thysanoptera) transmit only a few viruses, but these can be of great economic importance worldwide. Mandibulate, or chewing insects, mostly beetles (order Coleoptera), transmit a relatively small number but varied types of plant viruses. Among the mites, the minute bud or gall mites (family Eriophyidae) are the most important virus vectors.

ARTHROPOD VECTORS OF BACTERIAL PATHOGENS

Unlike viruses, most bacterial (prokaryotic) diseases of plants do not require insects as vectors, relying instead on rain, wind, soil, seed dispersal, or other means of transport and entry to plants. However, insect vectors do contribute to the spread of some bacterial pathogens of plants. Fire blight is an important bacterial disease of pome fruits, such as pears and apples, in which flower-visiting insects may have an important role in disseminating the causal bacterium *(Erwinia amylovora)* among blossoms. Insects are not essential, however, for fire blight to spread within plants once the bacteria are established, and there is little vector specificity among flower-visiting insects. Bacteria that rot potatoes *(Erwinia caratovora)* may be transported from infested potato tubers to uninfested tubers by flies whose maggots feed on plant roots or seeds beneath the soil. There is much greater vector specificity in corn flea beetle transmission of the bacterium *(Erwinia stewartii)* that causes Stewart's wilt of corn and in cucumber beetle transmission of the bacterium *(Erwinia tracheiphila)* that causes cucurbit wilt, which is an important disease of melons, squash, and cucumbers. The bacteria enter feeding wounds made by the beetle vectors, but not much is known of how the beetles introduce the bacteria into plants. Overwintering adult beetles provide an important way for these bacteria to survive the winter season without host plants.

Some bacterial pathogens are specialized parasites of plant vascular systems and require insect vectors for plant-to-plant movement and to enter and infect plants. These bacterial

pathogens are specialized for vector transmission and for living in plant vascular systems. Examples are the mollicutes (bacteria that lack a rigid cell wall) that live exclusively in the nutrient-rich phloem tissues. A few bacterial pathogens with rigid cell walls, such as the bacterium that causes citrus greening disease, also specialize in living within plant phloem sap. The citrus greening bacterium is transmitted by psyllids (superfamily Psylloidea). The mollicute plant pathogens include phytoplasmas and spiroplasmas. Most of the helical-shaped spiroplasma pathogens of plants, such as the spiroplasma that causes citrus stubborn disease *(Spiroplasma citri)* and the corn stunt spiroplasma *(Spiroplasma kunkelii),* can be cultured on artificial media. So far, none of the phytoplasma (formerly known as mycoplasma-like organism) plant pathogens have been cultured. Examples of economically important phytoplasmas are aster yellows phytoplasma in lettuce, carrot, celery, and other flower and vegetable crops and X-disease phytoplasma in stone fruits such as peach or cherry. Lethal yellowing disease of palms has been a major factor in killing coconut palms in Africa and the Caribbean. Both phytoplasmas and spiroplasmas are more specialized for parasitizing insects rather than plants because they can successfully colonize and, more importantly, can be transmitted by only a few species of insects. The most important vectors of mollicute plant pathogens are leafhoppers and planthoppers, but psyllids are an important third group of Hemiptera that are vectors. The pear psylla *(Cacopsylla pyicola)* transmits the pear decline phytoplasma, which causes the widespread pear decline disease. Typically, only one or a few species of insects within one of these families have been shown to transmit any particular mollicute. In contrast to their high degree of vector specificity, phytoplasmas and spiroplasmas can parasitize a typically wide range of plant species if the vectors can feed successfully on the plants.

Transmission appears to require that the mollicutes be taken up by vector feeding, penetrate the gut and multiply within the vector's body cavity, enter the salivary glands, and be expelled with saliva during vector feeding into functioning phloem tissues. Thus, it not surprising that vector transmission of various phytoplasmas or spiroplasmas requires a latent period ranging from 1 to over 4 weeks. The length of the latent period may be very sensitive to temperature, probably because the mollicutes must multiply within the vector for transmission to occur and multiplication is temperature sensitive.

Vector-borne bacterial species that parasitize the water-conducting part of the plant's vascular system (xylem) are less numerous but cause some important plant diseases. One such pathogen is *Xylella fastidiosa,* best known as the cause of Pierce's disease of grapes, but other strains of this bacterium cause important other diseases of citrus, coffee, peach, and other crop and forest plants. Sucking insects in several families that feed primarily on xylem sap are *Xylella* vectors. This includes sharpshooter leafhoppers in the subfamily Cicadellinae of the leafhopper family Cicadellidae and spittlebugs (family Cercopidae). Vectors appear to transmit the bacterium from their foregut without any required latent period, but continue to transmit for weeks or even months as adults. An immature vector (nymph) stops transmitting after molting its exoskeleton. Sumatra disease of cloves in Indonesia, caused by the xylem sap-inhabiting bacterium *Pseudomonas syzygii,* is spread by tube-building spittlebugs (family Machaerotidae), which are also xylem sap feeders.

Arthropods and Fungal Plant Diseases

The fungi are the most varied, common, and important plant pathogens, but the great majority of fungal pathogens do not require mobile vectors such as insects or mites. Instead fungal pathogens disperse to plants mainly in wind, rain, or soil. A large variety of fungi colonize plant wounds, including those made by arthropod feeding. However, some fungi are specialized for transmission by insect vectors.

Dutch elm disease is the best known example of a fungal disease of plants transmitted by an insect vector. The causal fungus, *Ophiostoma ulmi,* grows into a spore-bearing fungal mass (mycelium) under the bark and into the water-conducting woody tissues of elms. Adult bark beetles, such as the European elm bark beetle *(Scolytus multistriatus),* are especially attracted to distressed elms or freshly cut elm logs. The adult beetles excavate a tunnel by feeding beneath the bark and deposit eggs along the tunnel. Beetle larvae hatch from the eggs, tunnel farther under the bark, pupate, and then emerge as adults the following year. The Dutch elm disease fungus grows throughout brood chambers excavated by the beetles and produces sticky spores that attach to the body and mouthparts of the adult beetles that bore out of the bark to exit the tree. The beetles transmit the fungal spores to wounds they inflict while feeding on elm twigs. The fungus gradually spreads from the point of infection into the larger branches of the tree and then to the tree's trunk, where its action on the woody tissues eventually kills the tree. In very cold climates of North America, the native elm bark beetle *(Hylurgopinus rufipes)* is the main Dutch elm disease vector. Its transmission of fungal spores to elms leads to more rapid development of disease because it principally feeds on the trunk and large branches of the elm tree rather than small branches. Oak wilt disease, caused by the fungus *Ceratocystis fagacearum,* is spread by sap beetles (family Nitidulidae) that vector the spores from oozing cankers on diseased trees to fresh wounds on trees to which these beetles are attracted.

Some insects can create wounds, which fungi can then colonize without transport by the insects. Yet, even though insects in these cases are not vectors of the fungi, they can be important in determining how severe fungal infestation becomes. An example is a variety of fungi that can colonize the feeding wounds of caterpillars that feed on maize or peanuts. Some of these fungi (notably *Aspergillus* species) can produce powerful toxins, called aflatoxins, that sicken or even kill animals that are sensitive to the toxins. Preventing insect damage to grain in the field or in storage is an important step in preventing high levels of aflatoxins.

Insects as Vectors of Trypanosomes and Nematodes

Plant diseases caused by trypanosomes are not as well understood as trypanosome diseases of animals and humans, such as sleeping sickness. Trypanosomes are protozoans of variable body shape, which depends on their developmental stage and environment. Most insect-associated trypanosomes have stages that are elongated or leaf-like and are propelled by a centrally attached flagellum. Milkweed bugs (family Lygaeidae) transmit trypanosomes to milkweeds, in which they harmlessly occupy the interconnected latex system. A variety of other plant-parasitic trypanosomes inhabit the phloem systems of their plant hosts, causing severe disease. Phloem necrosis disease of coffee in northern South America and heartrot of palms are spread by sucking bugs in the family Pentatomidae and other related families.

Wood-boring beetles are vectors of pinewood nematodes *(Bursaphelenchus xylophilus)* that cause a severe disease of conifers in Asia and, more recently, North America. The juvenile nematodes enter the tracheae (breathing tubes) of adult long horned beetles (family Cerambycidae), as the beetles emerge from the dead trees in which they breed. As the beetles bore into new trees, the nematodes disperse from the beetles into the tree's woody tissues, causing blockage of the water-conducting system.

Control of Diseases Spread by Arthropods

The most obvious first step in controlling diseases caused by insect-borne pathogens might seem to be the elimination of vectors with insecticides. Although they are very effective in some situations, insecticides usually are not the best tools for control of most vector-borne pathogens of crops. The most effective approaches combine multiple methods and integrated approaches.

Sanitation to eliminate nearby sources of the pathogen (usually diseased plants) reduces the number of infective vectors near the crop to be protected. Physical isolation to prevent disease spread may be achieved in some cases by growing susceptible crops as little as 100 m or so from infected sources, but normally much greater isolation or separation is required. Therefore, area-wide cooperation may be necessary for sanitation to control some diseases.

The use of virus-free plants is probably the most widespread method of preventing virus spread. This is especially important for perennial plants such as fruit trees or plants propagated from vegetative cuttings, such as potatoes, strawberries, or sugarcane. Heat therapy or antiviral chemical treatments may be used to produce virus-free new plant growth that can be grafted or rooted to create virus-free plants for nursery propagation. As new infections of virus appear in trees, the trees can be removed; for some diseases, such as swollen shoot disease of cacao, trees within a specified radius of a newly diseased tree are also removed. Some viruses are transmitted via seeds from infected plants, and control may be based on planting virus-free seed. Lettuce mosaic virus (LMV), for example, is best controlled by using seed with less than 1 in 10,000 seedlings infected with LMV via seed. Above this threshold, aphid spread of the nonpersistently transmitted LMV will be economically damaging. The production of virus-free seed may require growing the seed crop in isolated areas that are otherwise free of the crop and the viruses of concern. Insuring virus-free foundation plants or seed to seed producers or nurseries is an important service usually provided by government or grower cooperatives.

Establishing a period of the year that is completely free of the targeted crop may reduce virus spread by breaking the transmission cycle. For example, this approach has proven effective for yellows viruses of sugar beet and celery mosaic virus, both of which are aphid transmitted. The growing of these crops for several months during the year is prohibited on an area-wide basis to prevent the carryover of virus in crop plants from one season to another. New fast-maturing varieties of rice that allowed multiple crops per year rather than a traditional single crop introduced new problems with long-established viruses because virus-infected crops coexisted next to newly planted fields. The solution was to have at least one period of the year free of all rice crops. Peak infective periods can be avoided for some virus diseases by planting after peak vector flight periods if the late planting still produces a profitable crop.

Removing diseased plants as soon as symptoms appear is an important step in preventing further spread of Dutch elm disease. The bark beetle vectors of the fungus that causes the disease are attracted to weakened trees, so removing diseased elms reduces populations of the beetles as well as reducing the percentage of beetles carrying fungal spores. Sanitation may also limit the spread of phytoplasma diseases of trees, such as X-disease of stone fruits like cherry.

The effects of the crop environment on vector flight behavior or plant choice may be effective in slowing virus spread, even for nonpersistently transmitted viruses. For example, reflective plastic sheeting used as a mulch (ground cover) repels aphids from landing on melon crops. The resulting delay in virus infection usually prevents the virus from reducing average fruit quality or yield. Sprays of 1% emulsions of paraffin oils on peppers or tomatoes reduces aphid transmission of nonpersistently transmitted viruses. Plants must be sprayed frequently with oil sprays to cover new growth because the oil directly affects the inoculation and acquisition of virus by aphids.

Insecticides generally are most effective in controlling disease spread where vector-borne pathogens are persistently transmitted or are spread mostly within the crop (termed secondary spread) rather than being carried into the crop from outside sources (primary spread) or where the most important vectors reproduce on the crop. Insecticides are usually not effective against nonpersistently transmitted viruses unless they quickly reduce or inhibit vector probing on treated plants. Some pyrethroid insecticides can rapidly intoxicate aphids after they land on plants, reducing even nonpersistent virus transmission. Examples of successful insecticidal control of vectors that achieved economic control of viruses are the persistently

transmitted barley yellow dwarf luteovirus in grain crops, potato leaf roll virus, and leafhopper-transmitted beet curly top virus. These are all viruses that are acquired or inoculated into plants only during relatively long feeding probes by aphids. Insecticides also reduce the spread of the leafhopper-transmitted aster yellows phytoplasma if vector numbers are not too high.

Genetically based plant resistance to pathogens or tolerance of infection without loss of yield provides the basis for the most successful control programs for vector-borne plant pathogens. A drawback is that breeding resistant plant varieties that are commercially acceptable has proven to be difficult or impossible to achieve for some crops. Molecular methods of introducing novel genes for resistance to viruses directly into crop plants promise to provide resistance to virus diseases for which no genetic resistance has yet been discovered.

See Also the Following Articles

Borers • Gallmaking and Insects • Genetic Engineering • Phytotoxemia

Further Reading

Camargo, E. P., and Wallace, F. G. (1994). Vectors of plant parasites of the genus *Phytomonas* (Protozoa, Zoomastigophorea, Kinetoplastida). *Adv. Dis. Vector Res.* **10,** 333–359.

Carter, W. (1973). "Insects in Relation to Plant Disease," 2nd ed. Wiley, New York.

Gray, S., and Banerjee, N. (1999). Mechanisms of arthropod transmission of plant and animal viruses. *Microbiol. Mol. Biol. Rev.* **63,** 128–48.

Hadidi, A., Khetarpal, R. K., and Koganezawa, H. (eds.) (1998). "Plant Virus Disease Control." APS Press, St. Paul, MN.

Harris, K. F., and Maramorosch, K. (eds.) (1982). "Pathogens, Vectors, and Plant Diseases: Approaches to Control." Academic Press, New York.

Matthews, R. E. F. (1992). "Fundamentals of Plant Virology." Academic Press, New York.

Perring, T. M., Gruenhagen, N. M., and Farrar, C. A. (1999). Management of plant viral diseases through chemical control of insect vectors. *Annu. Rev. Entomol.* **44,** 4457–4481.

Pirone, T. P., and Blanc, S. (1996). Helper-dependent vector transmission of plant viruses. *Annu. Rev. Phytopathol.* **34,** 227–247.

Purcell, A. H. (1989). Homopteran transmission of xylem-inhabiting bacteria. *Adv. Dis. Vector Res.* **6,** 243–266

Purcell, A. H. (1998). Insects and microbes. *In* "Introduction to Insect Biology and Diversity" (H. V. Daly, J. T. Doyen, and A. H. Purcell, eds.), Chap. 13. Oxford University. Press, New York.

Van den Heuvel, J. F. J. M., Hogenhout, S. A., and van der Wilk, F. (1999). Recognition and receptors in virus transmission by arthropods. *Trends Microbiol.* **7,** 71–76.

Plant–Insect Interactions

J. Mark Scriber

Michigan State University

The interactions of plants and their herbivores center upon the primary (nutritional) and secondary (allelochemical) composition of plants. The fundamental limitations of insect use of plants as food involve nutrients (nitrogen/protein, waters, lipids, and various minerals) as well as various classes of secondary chemical defenses (including alkaloids, cyanogenic glycosides, glucosinolates, terpenoids, phenolics, phytoecdysteroids, and polyacetates). The location ("findability") and utilization (suitability) of plant parts as insect food may depend on phenotypic variations induced by previous herbivores and microbes, as well as a wide array of interactions involving environmental factors such as nutrient availability, light regime, water, temperature, carbon dioxide, and various pollutants. Seasonal changes in plant growth, reproduction, and chemical/physical defenses also are important. Natural declines during plant maturation in the concentration of many low-molecular-weight allelochemicals (often called qualitative defenses) are contrasted with other higher molecular weight chemicals (such as tannins, lignins, and fiber; sometimes called quantitative defenses). In leaves of plants, a general pattern of decline in the concentrations of total nitrogen, water, and many qualitative chemical defenses usually accompanies leaf maturation. The phytochemical suitability of leaves for insect herbivores (the leaf-chewing guild in particular) also has genetically based biochemical variation. Together these factors affect the physiological and ecological suitability of the plant for supporting herbivore feeding, growth, survival, and reproduction.

HOST PLANT RESISTANCE AND INSECT COUNTERADAPTATIONS

In natural terrestrial communities, approximately 10% of the annual plant production on average is consumed by herbivores, a percentage that is generally greater than the plant biomass allocated to reproduction. In addition to the well-known defensive structures of thorns, barbs, spines, trichomes, hairiness or fuzziness, and physical toughness, plants possess a large array of chemicals that defend against the herbivore and pathogen enemies. Artificial selection also has produced insect-resistant genotypes and cultivars of crop plants, which has helped reduce the reliance on broad-spectrum synthetic pesticides. Plant breeders have recently been able to use techniques of molecular biology to incorporate new arrays of biochemical or microbial "pesticides" for plant defenses that have not been previously evolved by the plants (but that may occur naturally in the plant environment or even on the leaf phylloplane).

Insects may respond to these secondary plant products physiologically (e.g., by sequestration or enhanced excretion rates), biochemically with resistance (e.g., via target site insensitivity or enzymatic detoxification), or behaviorally (e.g., by reducing exposure or consumption by changes in feeding behavior). Many of these phytochemicals may be used by adapted herbivores that sequester the bioactive compounds in their wings or other body parts where they may serve a protective function from herbivore enemies (such as in distasteful models with aposematic or warning coloration) and often in various insect mimicry complexes.

PLANT DEFENSE AND HERBIVORE OUTBREAK THEORY

An understanding of the full array of potential insect–plant interactions is probably beyond comprehension. The geographic, altitudinal, and seasonal variations in plant chemistry in even a single species and the associated responses of insect herbivores (each species with its own geographical and genetic variation) make the task truly daunting. However, out of this complexity, ecologists have attempted to identify general patterns and organizing principles.

It is especially useful to examine the development of a series of general explanatory hypotheses or proposed models of insect herbivory, population dynamics, and plant defenses. These hypotheses have been mostly generated since the 1970s and are not mutually exclusive and have often built upon the theories of their predecessors. Details of the historical development of plant defense theory were nicely summarized in 1997 by Price.

Climatic Release Hypothesis

Outbreaks of insect herbivores on plants following periods of atypically warm, dry weather are numerous and have been documented for nearly a century. Up to the 1950s, herbivorous insect life table analysis suggested bad weather, lack of food, or lack of their natural enemies (predators, parasites, or diseases) were the primary insect population regulators. The indirect effects of the climate and abiotic environment as mediated through changes in host plant nutritional quality will likely be of increased significance in the near future given increasing concentrations of certain atmospheric gases (e.g., carbon dioxide), acid rain, global warming, and increased pollution.

Plant Stress Hypothesis

The connection between warm, dry weather and insect population eruptions, combined with similar outbreaks on plants in poor, dry soils, led to the hypotheses that water stress in plants may affect the availability of soluble nitrogen (especially for those in the plant-sucking or sap-feeding guilds). Nitrogen generally limits insect herbivores and population growth rates.

Plant Apparency Hypothesis

Following study of insect herbivores on oak trees in England and those on the cabbage family (Cruciferae) in North America, P. P. Feeny noted the divergent patterns of chemical defense used by these two plant types. The tree leaves were composed of relatively high concentrations (2.5–5% dry weight) of compounds believed to be digestibility reducers (tannins, lignans, resins, cellulose, silica). In contrast, herbaceous crucifers (forbs) had low concentrations (usually lower than 1%) of biosynthetically "less expensive" toxins (such as mustard oil glycosides or other low-molecular-weight chemicals such as alkaloids, cyanogenic glycosides, and coumaric acids). Herbivore food was prevalent and predictable (i.e., apparent) in trees and mature plant leaves containing these convergent digestibility reducers, or quantitative defenses or hurdles to herbivores, whereas the food resources for forb-feeding herbivores seemed less predictable, with divergent, qualitative (toxic) barriers in annual plants/herbs and very early immature (i.e., unapparent) plant parts. The "bound-to-be-found" trees were late successional, frequently in pure stands, large, and long-lived, with large amounts of general chemical defenses (effective against specialists as well as generalists) that acted in a dose-dependent (i.e., quantitative) manner. In contrast, the annual plants were short-lived, hidden from enemies in space and in time (unapparent), and defended by small quantities of qualitative allelochemicals such as mustard oils that repel nonspecialized insects (but are not effective against adapted herbivores) and that are effective at very low levels (basically not dose dependent).

The explanatory value of the apparency concept has been questioned because of the difficulty in its quantification. Most plants and plant parts have a dynamic continuum of both quantitative and qualitative chemical defenses, as well as phenological changes in the nutritional quality of leaves (as indexed, for example, by leaf water and total leaf nitrogen concentrations). Tannins were also shown not to be the general dose-dependent, digestibility-reducing chemicals they were originally believed to be. Instead, tannins evoke a wide variety of physiological effects such as increased mortality, decreased consumption rates, histopathological effects in the gut, and elevated metabolic costs for insect herbivores.

Induced-Defenses Hypothesis

The occurrence of phytochemical induction with leaf damage has been observed since the late 1970s. Plant-to-plant chemical communications and plant-to-insect parasite/predators have subsequently been included as multitrophic-level chemical synomones (plant volatiles that benefit both the sender and its receiver). Herbivore-damaged plants have been shown to provide carnivorous enemies of insect herbivores with important volatile chemical cues. These chemicals are detectable from a distance and aid natural enemies in locating suitable herbivore prey. However, plant volatiles from particular herbivores have not conclusively been documented to be consistently capable of providing critical species-specific herbivore information to the predator/parasites. Nevertheless, it is clear that multitrophic-level interactions are directly and indirectly influenced by damage-induced chemical responses of plants.

Expanded Hypotheses on Resource Availability and Plant Vigor for Defense against Herbivores

Although the value of the apparency hypotheses was weakened by difficulties in empirically assessing apparency, it nonetheless stimulated a great deal of research. Some of this resulted in promising alternatives relating to a causal relationship between

"resource availability" and plant allocations to antiherbivore defense.

Reduced nitrogen availability for plants usually resulted in reduced nitrogen-based defenses (usually toxins), but not necessarily in reduced carbon-based defenses (digestibility reducers). Thus, on nutrient-poor or late-successional sites, it was suggested that the inherently slow growth rates of plants may select for more carbon-based herbivore defenses (e.g., resins or phenolics) that could be reduced with fertilization with nitrogen (with corresponding increases in the nutritional value for herbivores). Low carbon conditions (e.g., shade and reduced photosynthesis) may result in slow growth despite high nitrogen, which could then be used for N-based defenses (e.g., alkaloids, cyanogenic glycosides).

These hypotheses about the carbon-nutrient balance and resource availability for slow-growing and fast-growing plants received support from many researchers in the 1980s. In many instances, predictions of the apparency, resource availability, and growth–differentiation balance hypotheses are in agreement; for example, trends in the types of chemical defense of early successional plants and late successional plant communities are congruent. However, equally apparent plants in resource-rich and resource-poor environments suggest that the resource availability hypothesis may have greater explanatory power than the apparency concept because the fast growers in resource-rich environments (nutrients and light) seem to support high herbivory (and may be predisposed to rapidly recyclable defenses such as alkaloids or other toxins) that can contrast with slow growers in the tropics, temperate zone, and arctic communities. This preference of herbivores for fast-growing plants has led to the suggestion that interactions may relate most simply to "plant vigor" (as a general hypothesis) to explain not only persistent differences in plant defenses between species, but also quality differences within a single plant species, genotype, or even individual.

In summary, despite the different roles of an overwhelming diversity of secondary plant chemicals, the fundamental limitations on herbivore growth rates seem to relate to the nutritional suitability of the insect food. Different leaf water and leaf nitrogen contents for different plant tissues correlate well with insect growth rates and efficiencies for most guilds and hundreds of different herbivore species. These indices of resistance can be induced by herbivory itself, as well as being constitutive.

Voltinism-Suitability Hypothesis

In addition to nutritional value, plant tissue suitability to an insect herbivore also depends on the degree of physiological and behavioral adaptations to a variety of plant secondary chemicals as well to leaf water, nitrogen, lipids, vitamins, and minerals. The question even arises as to whether low concentrations of nutrients may sometimes actually serve as plant defenses. At certain latitudes or in geographically localized cold pockets, seasonal thermal unit constraints (degree days as a resource) can determine whether an additional insect herbivore generation is feasible in any given year, depending on its selection of the most nutritional host plant species, which varies with the timing of leaf bud break and phenological (seasonal) patterns, which differ at various locations. Thus, abiotic factors have been shown to affect host-plant choice (acceptability) and host-plant suitability for herbivores. High nitrogen and high water content generally reflect the most rapid leaf and cell growth and presumably plant vigor as well. Insect growth performance usually correlates well with these plant quality indices.

The range of host plants utilized at a given latitude (or local climatic zone) may be the result of natural selection in relation to these abiotic factors. Thermal constraints for the summer growing season (as in Alaska or in similar localized cold pockets in the continental United States) contrast ecologically with thermally relaxed zones (i.e., where choice of either excellent or poor host-plant species or leaves does not influence the possibility of an extra generation per season). The difficulty in the voltinism-suitability model is that a good host typically is more than just the plant species, more than its allelochemical acceptability, more than its nutritional suitability, and more than the abiotic thermal regime in which the herbivores are trying to optimize their growth and survival. The biotic community of natural enemies (e.g., enemy-free space as a resource) must also be considered as a critical determinant of the real ecological/evolutionary suitability. The relative roles in plant defense played by natural enemies, weather-induced stress, herbivore-induced stress, and various abiotic factors remain complex, with many unique and dynamic variations on the suitability hypothesis.

EVOLUTIONARY HISTORY AND PHYTOCHEMICAL FUTURE

Coevolutionary or reciprocal changes between plants and insects are the foundation of numerous phytochemical defense theories. However, there is surprisingly little direct evidence that insects select for plant phytochemical defenses. Most insect–plant interactions are diffuse without mutual counteradaptations. They will be, at best, a geographic mosaic with isolated and dynamic hot spots. Additional studies of different geographical populations (with and without herbivore selection pressure), plant and herbivore genetic analyses, phytochemical dynamics in relation to abiotic factors, and historical biogeography all seem to be warranted and critically needed. The ecologically enigmatic problem seems to be that our understanding of both plant resistance and insect counteradaptations ultimately depends on the identification of specific molecular pathways and an elucidation of the relative roles of genetic and environmentally induced variation between interactive populations.

INSECT HERBIVORES CAN BENEFIT PLANTS

It has been generally accepted that insect herbivory results in plant tissue damage that is detrimental for plant growth,

survival, or reproduction. It has been argued that this is not true in all situations and that insect herbivory may sometimes be beneficial for plant productivity. Plant responses to herbivory are determined by many different habitat, plant, nutrient, and herbivore specifics that are biologically variable.

The frass fall and uneaten leaf pieces that reach the ground from insect herbivores release nutrients for plant growth throughout the season and delay plant leaf senescence. The carbohydrates in aphid honeydew drippings can stimulate soil microbes and enhance nitrogen fixation, which benefits the adjacent plants. Such enhancement of nutrient cycling by herbivore feeding can be important in regulation of ecosystem productivity, especially in grasslands and forests. Although many crops and other plants can sustain 30 to 40% defoliation with little obvious impact on production, such feeding, if repeated annually or if on flowers and/or seeds, can be much more damaging.

Insects provide direct nutritional benefit to some plants. Carnivorous plants that "digest" insects and use chemicals from their bodies for nutrients (especially nitrogen) are represented by many species, including pitcher plants, bladderworts, Venus flytraps, and sundews. These carnivorous plants are often found in soils that are low in available nitrogen, which may have been an important selection pressure for the evolution of these botanical life history traits.

SUMMARY

Insect–plant interactions involve a wide array of biotic and abiotic environmental influences as well as geographical and temporal variations built upon the diverse genetic foundations and inducible phenotypic plasticity of species, populations, and individuals. It therefore seems very appealing when theories arise that seem to have predictive power for these complex interactions. Such complexity is amplified when the variety of insect feeding guilds and variations in response of different plant parts and tissues are considered. The relationships between normal phenological changes in plant leaf (or part) composition throughout the growing season, carbon-nutrient stress, mineral nutrition, plant vigor, phytochemical induction of resistance in damaged/diseased leaves, increases in certain atmospheric gases, global warming, metabolism of different forms of carbon, and annual versus perennial growth forms need coordinated biocomplexity studies. With such knowledge, the suitability of such plant tissues for insect and other herbivores (or the resistance of plants to their enemies) may become much more predictable.

See Also the Following Articles

Honey • Host Seeking, for Plants • Nutrition

Further Reading

Ayers, M. P., Clausen, T. P., MacLean, S. F., Redman, A. M., and Reichardt, P. B. (1997). Diversity of structure and antiherbivore activity in condensed tannins. *Ecology* **78,** 1696–1712.

Bernays, E. A. (1989–1994). "Insect–Plant Interactions," Vols. 1–5. CRC Press, Boca Raton, FL.

Dicke, M. (1999). Are herbivore-induced plant volatiles reliable indicators of herbivore identity to foraging carnivorous arthropods? *Entomol. Exp. Appl.* **91,** 131–142.

Fritz, R. S., and Simms, E. (1992). "Plant Resistance to Herbivores and Pathogens: Ecology, Evolution, and Genetics." University of Chicago Press, Chicago.

Herms, D. A., and Mattson, W. J. (1992). The dilemma of plants: To grow or to defend. *Q. Rev. Biol.* **67,** 283–335.

Karban, R., and Baldwin, I. T. (1997). "Induced Responses to Herbivory." University of Chicago Press, Chicago.

Price, P. W. (1997). "Insect Ecology," 3rd ed. Wiley, New York.

Price, P. W., Lewinsohn, T. M., Fernandes, G. W., and Benson, W. W. (1991). "Plant–Animal Interactions: Evolutionary Ecology in Tropical and Temperate Regions." Wiley, New York.

Rosenthal, G. A., and Berenbaum, M. R. (1991 and 1992). "Herbivores: Their Interactions with Secondary Plant Metabolites," Vols. 1–2. Academic Press, New York.

Schoonhoven, L. M., Jermy, T., and van Loon, J. J. A. (1998). "Insect–Plant Biology: From Physiology to Evolution." Chapman & Hall, New York.

Scriber, J. M. (1984). Insect–plant interactions: Host plant suitability. *In* "The Chemical Ecology of Insects" (W. J. Bell and R. T. Cardé, eds.), pp. 159–202. Chapman & Hall, London.

Scriber, J. M. (2002). Latitudinal and local geographic mosaics in host plant preferences as shaped by thermal units and voltinism in *Papilio* spp. (Lepidoptera). *Eur. J. Entomol.* **99,** 225–239.

Scriber, J. M., and Lederhouse, R. C. (1992). The thermal environment as a resource dictating geographic patters of feeding specialization of insect herbivores. *In* "Effects of Resource Distribution on Animal–Plant Interactions" (M. R. Hunter, T. Ohgushi, and P. W. Price, eds.), pp. 429–466. Academic Press, New York.

Scriber, J. M., and Slansky, F., Jr. (1981). The nutritional ecology of immature insects. *Annu. Rev. Entomol.* **26,** 183–211.

Slansky, F., and Rodriguez, J. G. (1987). "Nutritional Ecology of Insects, Mites, Spiders, and Related Invertebrates." Wiley, New York.

Slansky, F., and Scriber, J. M. (1985). Food consumption and utilization. "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 4, pp. 87–163. Permagon, Oxford.

Plecoptera
(Stoneflies)

Kenneth W. Stewart
University of North Texas

Stoneflies comprise a hemimetabolous order of 16 families and more than 2000 species of aquatic insects distributed on all continents except Antarctica and most major islands except notably Cuba, Fiji, Hawaii, and New Caledonia. They are primarily associated with running water, where nymphs inhabit mineral or organic substrates of streambeds, and the winged adults rest throughout their seasonal lives in streamside microhabitats such as rocks, moss, debris, leaf packs, and riparian vegetation. A few species occur in waveswept substrates of cold alpine and boreal lakes or in intermittent streams.

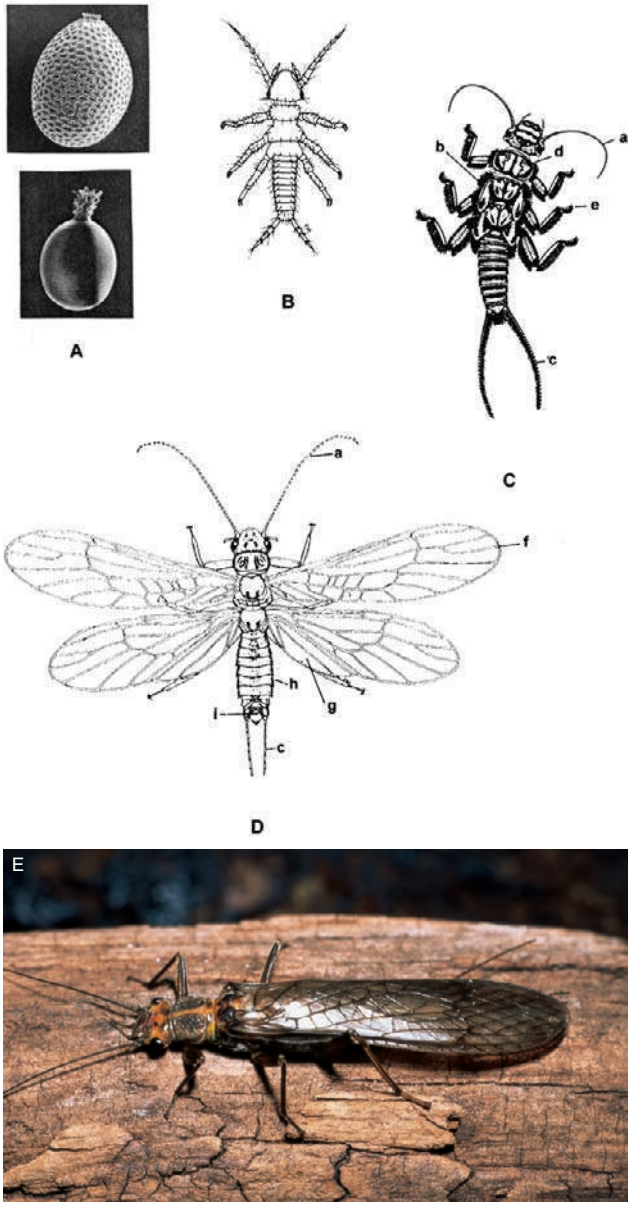

FIGURE 1 Life stages of stoneflies. (A) Eggs. (B) Hatchling (first instar). (C) Late instar. (D) Adult male. a, antennae; b, gills; c, cerci; d, prothorax; e, tarsal claws; f, forewing; g, hindwing anal area; h, abdomen; i, epiproct of male genitalia; (E) Adult of the stonefly *Calineuria californica* (Banks). (Photograph by B. P. Stark.)

Stonefly adults (Figs. 1D and 1E) are variable in size from about 5 to 50 mm and in color from black to green or yellow, often marked with distinctive light or dark patterns. The aquatic adult of one species known from the depths of Lake Tahoe *(Capnia lacustra)* and a few other species are apterous (wingless), but most adults are winged. The wings of males and females of some species, or particular populations of a species, are shortened (brachypterous) and they do not fly, but the typical condition is of two pairs of wings as long as or longer than the abdomen (macropterous). As the ordinal name (Plecoptera = folded wings) describes, the hind wings typically have an expanded posterior (anal) lobe that folds longitudinally under the main wing (Fig. 1D).

Stoneflies are relatively slow, somewhat awkward fliers that typically fly short distances to disperse, to search for mates, or, for females, to deposit eggs.

Adults (Fig. 1D) have ten abdominal segments; the genitalia of males are distinctive at the generic and species levels and consist mainly of various external manifestations of the ninth and tenth segments, such as paired hooks, lobes (paraprocts), or sclerotized stylets and in some taxa a median probe (epiproct) of various shapes. During copulation, the copulatory organ (aedeagus), normally inside the abdominal cavity, is everted ventrally from the genital opening on the ninth sternum. The external female genitalia consists of a flap-like subgenital plate covering the genital opening on the eighth abdominal sternum, a structure that the male grasps or holds with his hooks or lobes during copulation. The wings, 3-segmented tarsus, genitalia, and a pair of usually multisegmented tails (cerci), arising from the 10th abdominal segment, generally characterize stonefly adults. Nymphs (or larvae) (Figs. 1B and 1C) may or may not generally resemble their adults. They are gill-less or have diagnostic simple or branched tracheal gills arising from different parts of the body such as near mouthparts, ventral head, thorax, coxae, or abdomen, and they always have multisegmented tails (cerci). Stubs or the basal remnants of gills are retained as vestigial structures in the adults of some taxa and aid in their identification to family and genus. The long, multisegmented nymphal cerci become reduced to fewer segments in some adults or to a single segment in adult males of the families Leuctridae, Nemouridae, and some Taeniopterygidae.

TAXONOMY AND GENERAL DISTRIBUTION

The Plecoptera is divided into the two suborders Arctoperlaria and Antarctoperlaria. The Arctoperlaria are distributed in the Northern Hemisphere, except the family Notonemouridae, which occurs only in southern South America, southern Africa, Madagascar, Australia, Tasmania, and New Zealand, and some genera of Perlidae such as *Anacroneuria* and *Neoperla* that have moved south across the equator in recent times, perhaps from 15 to 30 mya. The Arctoperlaria is further divided into the group Euholognatha (containing six families: Capniidae, Leuctridae, Nemouridae, Notonemouridae, Taeniopterygidae, and Scopuridae) and the group Systellognatha (containing six families: Chloroperlidae, Peltoperlidae, Perlidae, Pteronarcyidae, and Styloperlidae). The Euholognatha have mouthparts adapted for herbivory (scrapers, grazers, collector–gatherers, shredders, gougers, and detritivores), including molariform mandibles, and its species occur with few exceptions in streams of various sizes. The Systellognatha, except Peltoperlidae and Pteronarcyidae whose mouthparts are similar to those of Euholognatha because of convergent evolution to herbivorous food habit, have mouthparts mainly adapted for predation, including sharp-cusped mandibles and toothed lacinia for grasping and holding prey. The systellognathan families

Perlidae and Peltperlidae have very few species in arctic and subarctic streams.

The suborder Antarctoperlaria, as the name implies, is restricted in distribution to the Southern Hemisphere. In some areas, recently invading genera of Arctoperlaria, such as *Anacroneuria* (Perlidae) in South America and *Neoperla* (Perlidae) in Africa, have outcompeted them. The suborder contains four families: Austroperlidae, Diamphipnoidae, Eustheniidae, and Gripopterygidae. The Austroperlidae and Gripopterygidae live in a wide variety of habitats and the Diamphipnoidae and Eustheniidae are restricted to relict populations in the southern Neotropical and Australian regions. Each of the 16 families of Plecoptera has unique combinations of wing venational, genital, gill, and other characteristics.

ECOLOGICAL IMPORTANCE

Stoneflies are integral and important food web components of most stream ecosystems throughout the world and therefore are almost exclusively beneficial insects. The various taxa have radiated to use virtually every type of food and substrate habitat resource available to them. The nymphs are variously detritivores, herbivores, insectivores, or omnivores, and in some species the diets of nymphs shift from detritivory or herbivory, through omnivory, to strict insectivory as development proceeds. In turn, they become food for larger insectivores and fishes and are therefore important in the energy dynamics of stream food webs. Particular taxa are usually associated with particular stream microhabitats, such as the interspaces of loose gravel or cobble substrates, leaf packs, detritus, debris, or logs. The nymphs of numerous species have evolved to coexist in a relatively harmonious, noncompetitive way in given stream ecosystems by partitioning their food, space, and time resources. Most stonefly species require relatively undisturbed conditions of the streams they historically inhabit and therefore are important biological indicators of stream water quality. They constitute the "P" component of one of the major biomonitoring indexes of "clean water species," called the EPT Index ("E" for Ephemeroptera, "T" for Trichoptera), used for assessing water quality and degree of stream disturbance by humans.

LIFE HISTORY AND BEHAVIORS

Adults

Adult stoneflies (Figs. 1D and 1E) usually emerge during the night from nymphs that have crawled out of the water onto objects projecting from streams or on the stream bank, such as entrained leaf packs, logs or debris, rocks, or riparian vegetation. Some species of Euholognatha are black and emerge under ice or snow cover in winter. A light colored (teneral), clumped-winged, soft adult emerges from the last instar skin during the molt through a split in the head and dorsal thoracic segments. Males of many species emerge a day or more before females, so that they are present and searching when females appear. After some degree of hardening, both sexes become cryptic, hiding in crevices or vegetation during inactive periods, usually during the day, and becoming active in mate-finding and other activities typically at night. Adults of Systellognatha have reduced mouthparts and typically do not feed, but may ingest liquids or nectar. Adults of Euholognatha feed variously on algae, lichens, flower pollen and nectar, or soft fruits. For most species of stoneflies, the details of transformation, dispersal, feeding inactivity periods, and longevity are unknown. Generally, adults live only one to a few weeks and are mostly actively engaged during that time in the reproductive activities of mate-finding, copulation, and oviposition.

Communication, Mate-Finding, and Mating

The primary method of communication for locating mates in the Northern Hemisphere stonefly suborder Arctoperlaria is vibrational signaling through substrates. The vibrations are produced by tapping or rubbing the abdomen on the substrate or by body tremulations transferred to the substrate. The signals of most insects that use this method consist of simple volleys of evenly spaced vibrations, but stoneflies have evolved a much more diverse and complex system of vibrational communication than is known for other insect groups. The currently accepted paradigm of how this behavior, generally known as drumming, evolved suggests that the ancestral method of signal production was percussion and that signals were monophasic volleys of evenly spaced drumbeats. Natural selection favored increasingly complex signals, leading to greater specificity of communication and mate-finding among species and possibly increased capability for sexual selection to measure reproductive fitness. An increasing complexity of particularly male call signals may have evolved through the following three behaviors: (1) more sophisticated signaling methods, sometimes associated with specialized coevolving ventral abdominal structures, (2) rhythmic patterning of signals, and (3) possible use of selected natural substrates for signal transmission. The result has been that current species signal variously by percussion, scraping, or rubbing the abdomen on the substrate (abdominal–substrate stridulation) or tremulation (vibrations produced by push-ups or rocking motions of the body without abdominal contact with the substrate). Signal rhythms are species-specific and vary from evenly or unevenly spaced monophasic volleys to variously spaced diphasic or grouped signals.

The entire mating system of stoneflies involves communication as well as aggregation and movement behaviors of both sexes, the actual copulation, and postmating behaviors. The typical system in Arctoperlaria involves the following complex of behaviors: (1) initial aggregation of sexes at encounter sites near streams, (2) calling by males with species-specific signals during ranging search, (3) duet establishment by virgin females answering the male call if effective vibrational

communication distance from her is achieved, (4) a localized search by the male in a "triangulation" or other pattern for the now stationary female while both continue dueting, and (5) almost immediate mating after the male locates and contacts the female. Males are polygamous and presumably continue calling and searching during their short reproductive lives.

Typically, mated and unguarded females reject subsequent male advances by raising and curving their abdomens. Southern Hemisphere stoneflies of the suborder Antarctoperlaria have never been documented to drum; therefore, their communication–search system for mate-finding is unknown, but they may have evolved a highly specific encounter site aggregation behavior that enables sufficient mate-locating ability without vibrational or other forms of intersexual communication.

Mating in stoneflies involves the male mounting the female, curving his abdomen around her left or right side, and engaging the subgenital plate, pulling it down with his external genitalia. This effectively matches her genital opening beneath the plate with a dorsal position between his cerci where the aedeagus will project. His aedeagus is everted from beneath the ninth sternum and expands backward and upward between his cerci into the female. Sperm are usually conveyed into the female by this intromittent aedeagus, but in some species sperm are conveyed through a hollow male epiproct or are externally deposited onto the female opening to be subsequently aspirated into the bursa (vagina) by telescoping movements of her abdomen.

Eggs and Oviposition

The eggs of stoneflies (Fig. 1A) vary considerably in size, shape, and details of chorionic (eggshell) ornamentation and sculpturing. Commonly, eggs are spindle-shaped, but they also may be spherical, flattened or three-sided. Frequently, an anterior collar is present and the shell surface may be smooth or ornamented either with ridges or the hexagonal pattern of impressions formed by the cells lining the ovarian chambers where the eggs are produced. Micropyles (sperm entrance holes) penetrate the chorion completely and may have associated surface grooves or ornate projections that serve as sperm guides. Actual penetration and fertilization by sperm of the egg is, as in most insects, delayed until the eggs are being stored or passed through the oviduct just prior to oviposition. The sperm are stored in the female spermathecum between copulation and fertilization. There also may be shallow pores leading to elaborate respiratory networks within the chorion. Eggs have sticky membranous or gelatinous surface coverings and sometimes filament-like projections with hooked tips that swell and help the eggs attach to substrates under water close to the selectively optimal sites where females deposit them.

Eggs are deposited by females in pellets or masses, each containing numerous eggs, that the females hold on the subgenital plate. They release the egg masses directly into the water by splashing into the surface during an oviposition flight, or by contacting shallow water while running near the shore, or by dropping eggs from the air while flying over water. Females of some Capniidae are also known to completely submerge themselves and crawl along the bottom and scrape the egg mass off onto the substrate.

In most species, embryonic development proceeds directly and is complete within 3 to 4 weeks. In other species, particularly those adapted to intermittent streams or streams subjected to extremes in temperature, embryonic development may be arrested for from 3 months to 1 or more years, and hatching is thus delayed until environmental conditions are favorable for nymphal survival.

Nymphs

Hatchlings (Fig. 1B) emerge from the egg by pushing on the chorion with the front of their head. The shell breaks into two halves or splits leaving a hinged cap and opening through which the first instar crawls out. The first instars are unpigmented with few body hairs, have fewer than 12 antennal and 6 cercal segments, and gills and wingpads are absent, reduced, or represented only by knobs or stubs. Little is known about the food, habitat, or behavior of hatchling and early instar stoneflies. The few species that have been studied feed mainly on detritus or the microflora–fauna on the surface of decomposing leaves.

Nymphs progressively develop and grow through about 10 to 24 sizes (instars). Full development of particular species requires from 4 months to 3 to 4 years. During this time there is a molt between each instar, addition of antennal and cercal segments, usually the addition of body hairs, a progressive increase in size of wingpads, and appearance and development of gills (if present in particular taxa) and characteristic pigment patterns. Growth of a particular species may be sustained at an even pace or seasonal, with fast and slow stages. In temperate climates growth is generally rapid during spring and fall and slowed or arrested (diapause) during extreme temperatures in summer or winter. But, interestingly, a number of euholognathan species, particularly in the families Capniidae and Taeniopterygidae, have adapted to experience their major growth in late fall and winter and emerge as adults during winter on ice or snow or in early spring during ice breakup. Completion of development and subsequent emergence as adults in temperate climates, therefore, may occur during any season, depending on altitude, latitude, and species.

Stoneflies have diversified their food habits such that different species fill about every conceivable major food niche in streams. Some species are herbivore–detritivores throughout their development, some are insectivores throughout development, and some experience an ontogenic (developmental) shift from herbivory–detritivory through omnivory and finally to carnivory. Characteristics of mandibles give a clue to food habits. The mandibles of herbivores have molariform surfaces or scraping ridges and those of carnivores sharp teeth for grasping and tearing. The food of carnivorous species consists primarily

of the other aquatic insects of their communities such as midge larvae (Chironomidae), mayfly nymphs (Ephemeroptera), caddisfly larvae (Trichoptera), and occasionally the smaller nymphs of other stoneflies. Nymphs may be opportunists or in some cases are very selective for the size, behavior, and taxa of their prey. Prey are captured, grasped by the head with the lacinia and mandibles, and usually swallowed whole, headfirst.

Particular species of nymphs are found in certain types or sizes of streams at particular latitudes or elevations and often specific microhabitats. Rare and endemic species have very specific biological and physical requirements and therefore continue to exist only in pristine or little-disturbed sections of streams that in many instances are now found only in remote areas or in, or adjacent to, protected national parks or preserves. Many species have broader requirements and are more widespread or ubiquitous over large areas of continents in a wide variety of habitats. Only a few species are tolerant of the conditions of streams disturbed by siltation, alteration of natural temperature regimes, or chemical pollution. Stoneflies depend on substrates as a place in which to live. Slender species live in the interspaces of gravel, cobble, or vegetable debris such as leaf packs. Partitioning of microhabitats, and consequent microdistribution, is characteristic of most stonefly assemblages in a given stream. Most species live in the surface layers of a streambed, but a few live deep in loose mineral substrates such as glacial till and sometimes in the water-filled spaces of such substrates for considerable distances deep and lateral from the margins of the surface stream. Nymphs are sometimes found drifting in the water column of streams. This results from being dislodged by some physical disturbance or entering the water column as a behavioral means of dispersal, using the flow of water. Drifting enables nymphs to escape predators or move to a less populated habitat where food and/or space resources are more available or of higher quality.

Life Cycles

With few exceptions, a full generation of the egg, nymph, and adult stages of a stonefly species requires 1 to 4 years. For one-year (univoltine) cycles, the nymphal growth portion may be "fast," requiring only 4 to 7 months, or "slow," requiring nearly a full year. The fast type is characteristic of species that diapause for variable times, up to 8 months during warm or dry periods of streams; the slow type characterizes species whose nymphal stage requires about 11 months. Species requiring more than 1 year for a generation are termed semivoltine. Those that live in intermittent streams may diapause in the egg stage during drought periods for more than a year, then have a fast growing nymph for only 4 to 6 months. Semivoltine species living in cold streams may have a short egg stage, with nymphs requiring 2 or more years to develop, or a year-long diapausing egg stage, with nymphs requiring 1 to 3 years to develop. The life cycles and resource requirements of stoneflies are important considerations for developing stream management strategies.

See Also the Following Articles

Aquatic Habitats • Pollution • Swimming, Stream Insects • Vibrational Communication

Further Reading

Resh, V. H., and Rosenberg, D. M. (1984). "The Ecology of Aquatic Insects." Praeger, New York.

Stark, B. P., Szczytko, S. W., and Nelson, C. R. (1998). "American Stoneflies: A Photographic Guide to the Plecoptera." Caddis Press, Columbus, Oh.

Stewart, K. W. (1993). Theoretical considerations of mate-finding and other adult behaviors of Plecoptera. *Aquat. Insects* **16,** 95–104.

Stewart, K. W. (1997). Vibrational communication in insects: Epidomy in the language of stoneflies? *Am. Entomol.* **43,** 81–91.

Stewart, K. W., and Harper, P. P. (1996). Plecoptera. *In* "An Introduction to the Aquatic Insects of North America" (R. W. Merritt and K. W. Cummins, eds.). Kendall/Hunt, Dubuque, IA.

Stewart, K. W., and Stark, B. P. (1988). Nymphs of North American stonefly genera (Plecoptera). *Entomol. Soc. Am.* **12,** 1–460.

Zwick, P. (2000). Phylogenetic system and zoogeography of the Plecoptera. *Annu. Rev. Entomol.* **45,** 709–746.

Pollination and Pollinators

Gordon W. Frankie
University of California, Berkeley

Robbin W. Thorp
University of California, Davis

Pollination in its most basic sense is the transfer of pollen from the male sex organ (anther) to the receptive portion of the female sex organ (stigma) in flowering plants (Fig. 1).

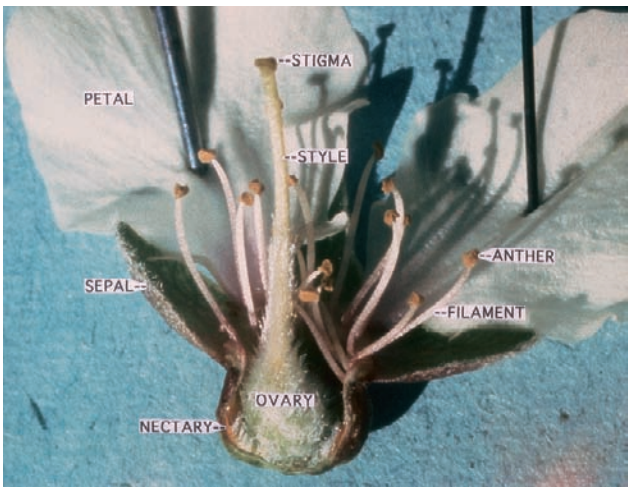

FIGURE 1 Floral parts of an almond blossom. The petals are color signals, the male stamen (anther and filament) and the female pistil (stigma, style, and ovary) are the reproductive parts. Nectar from the nectary and pollen from the anthers are food rewards to pollinators.

If the transfer is successful, it leads to fertilization, production of seed, and reproduction of the plant. This process often involves some sort of external vector such as wind, water, or animals. Some flowering plants may reproduce without the aid of pollen vectors, using mechanisms such as vegetative reproduction, apomixis, or automatic selfing. But our concern in this article is with animal vectors of pollen. Many kinds of animals may perform the ecological service known as pollination, including birds, bats, and some nonflying mammals. However, the dominant group of pollinators is the insects, especially bees.

ENTOMOPHILY

Many flowering plants are adapted to insects as pollinators and provide primary rewards that attract and keep the pollen vectors returning to flowers. Rewards include nectar, pollen, lipid secretions, food bodies, scents, resins, and material for nest building. In additon to primary rewards, most insect-pollinated flowers also produce a number of cues or signals that distinguish them from other species and promote the ease by which an insect can relocate a rewarding flower, thereby encouraging the insect to move pollen from flower to flower of the same species. These signals include odors, colors, shapes, textures, and tastes. These signals are often combined into patterns that have been recognized as syndromes related to the type of pollen vector. For example, a typical butterfly-pollinated flower that would be red, have little odor, possess a landing platform, have nectar hidden at the base of a deep tube that can be reached by the long coiled butterfly proboscis, have nectar that would be high in amino acids, and have flowers that would be open during the day. In contrast, a typical hawk moth-pollinated flower would be white, have a strong sweet odor, lack a landing platform, have long stamens with freely swinging anthers, and would bloom nocturnally. Nectar position and composition would be similar to those of a butterfly flower. All these characteristics are well suited to the hovering flight and extremely long proboscis of a hawk moth (Fig. 2).

Pollinator syndromes are perhaps most readily distinguished in the tropics. However, not all flowers are easily classified in a pollinator syndrome. Many flowers are visited by guilds of visitors that include diverse taxa of insects. For these reasons some pollination biologists see little value in the use of pollination syndromes even for teaching. However, even when using a pollination syndrome approach, it is important to distinguish which taxa among the visitor guilds are actually effective pollinators of a flowering plant.

Many insect taxa visit flowers and thus are potential pollinators; however, only a few taxa are of prime importance. Minor groups include those Orthoptera that feed on pollen and some Heteroptera that visit flowers for nectar or those that use flowers as sites that attract prey items (e.g., Phymatidae, Reduviidae). Thrips (Thysanoptera) are common on flowers and often feed on pollen and other flower tissues. They may do more damage than good, but their positive contribution as pollen vectors is understudied. The major groups of insects that pollinate plants belong to the four largest orders of insects: Coleoptera (beetles), Diptera (flies), Lepidoptera (butterflies and moths), and Hymenoptera (bees, ants, and wasps).

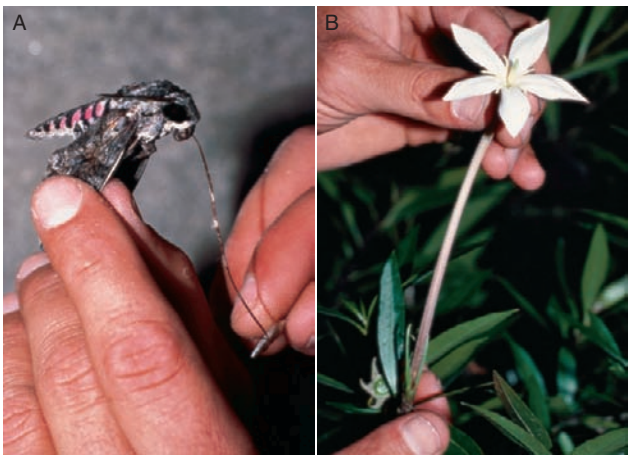

FIGURE 2 (A) Uncoiled proboscis (tongue) of a hawk moth, Costa Rica. (B) Tubular flower of *Lindenia rivalis* (Rubiaceae), pollinated at night by hawk moths, Costa Rica.

Coleoptera

Beetles are often considered "mess and soil" pollinators in that while rumaging around in flowers feeding on pollen and other flower parts they pick up pollen on their bodies that is transferred to other flowers on subsequent visits. This type includes many that destroy some flowers by feeding on them (e.g., Scarabaeidae, Meloidae), but in the process others get pollinated. Some beetles are associated with pollination of some "primitive" flowers and have been considered responsible for pollination and diversification of early flowering plants (Angiospermae). Beetles of the family Nitidulidae feed on specialized food bodies on anther tips of the spice bush *Calycanthus* (Fig. 3).

FIGURE 3 White-tipped "food bodies" on anthers of *Calycanthus* (Calycanthaceae) that attract beetles (Nitidulidae) that pollinate the flower.

Diptera

Diverse flies, including male mosquitoes, various midges, carrion flies, pollen-feeding Syrphidae, and long-tongued nectar feeders (e.g., Bombyliidae, Acroceridae, Nemastrinidae), pollinate flowering plants. Pollination by flies is greatly understudied and underrated. Although flowers of pipevines are considered classical "trap" flowers that imprison flies with inward directed hairs until the flower has released pollen on them, flowers of the California pipevine, *Aristolochia californica,* exhibit a different mechanism involving a reward to retain flies until pollen is released. Midges of the family Mycetophilidae are the primary visitors and pollinators of *A. californica.* When flowers of *A. californica* first open, the stigma is receptive and a dark ring of glandular trichomes encircles the outer wall of the flower at the level of the stigma. Flies descend through the hooded entrance to the lower bowl and are attracted to the area of the stigma and trichomes by a light window. Flies feed on the trichome surface, contacting the stigma and depositing pollen from previous flower visits. The stigma closes, the trichomes wilt, and the anthers shed pollen into the bowl and onto the flies. Because there is no more food available, flies exit the flower and seek another, thereby pollinating the next flower.

Lepidoptera

Most adult butterflies and hawk moths are well-known flower visitors. Other day-flying moths (e.g., *Schinia* of the Noctuidae, *Adela,* Incurvariidae) and nocturnal moths pollinate while settling or perching on flowers during feeding or oviposition. A very specialized relationship exists between yucca moths (*Tegiticula,* Incurvariidae) and their host yucca flowers (*Yucca,* Liliaceae). Female *Tegiticula* enter *Yucca* flowers, collect pollen into a ball in specialized maxillary palps, move to the apex of the pistil where pollen is deposited on the stigma, and oviposit into the base of the pistil where seeds will develop from the pollination behavior. Larvae of the moth develop in the fruit pod, feeding on a portion of the seeds. Thus, both insect and plant benefit from this highly mutualistic association; the plant gets pollinated and produces seed, some of which goes to producing new moths.

Hymenoptera

Some sawflies and parasitoid wasps feed on pollen, especially on open shallow flowers. Classical mutualism occurs with the pollination association of fig wasps (Agaonidae) and their floral hosts, figs (*Ficus,* Moraceae). Similar to the situation with yucca moths and yucca, a portion of seeds in fig host flowers provides nourishment for development of the pollinating wasps. The story is often more complex and involves more than one generation of wasp per year and more than one host.

Other aculeate wasps, both social and solitary, augment insect prey diets with nectar. One genus of solitary wasps, *Pseudomasaris* (Masaridini, Vespidae), is completely dependent on flowers for nectar and pollen as food for their young. Ants may visit flowers but because their metathoracic glands secrete mold inhibitors they inhibit pollen germination and are unlikely pollen vectors. Bees (superfamily Apoidea) are the single most important taxon of pollinating insects with 20,000 to 30,000 species worldwide. Bees are derived from wasps and highly adapted to gathering pollen as brood food and nectar for flight fuel.

EXPLOITATION RELATIONSHIPS

Not all pollination relationships are mutualistic, i.e., beneficial for both partners. Some are based on deceit or robbery, in which only one partner benefits and the other may even be injured.

Thievery

Insect visitors to flowers may obtain the food items they seek without transferring pollen in the process. Some insects with mouthparts too short to reach nectar sequestered in the bottom of long tubes or spurs are able to penetrate the nectar-bearing structures with strong mandibles or maxillae. Such behavior is well documented for bumble bees, such as *Bombus occidentalis* in western North America or the related *Bombus terrestris* of Europe, when they encounter long-tubed flowers. This behavior is commonly exhibited by carpenter bees, especially in the tropics (Fig. 4). Insects that are mismatched in size with the flowers they visit may be effective gleaners of pollen from the anthers, but rarely if ever contact the stigmas in the flowers they visit. These thieves often scavenge pollen from flowers adapted to other types of pollinators. For example, the evening primrose of the southwestern deserts of North America are typically adapted for pollination by night-flying hawk moths, but they are visited early in the morning after they have opened by solitary ground-nesting bees of the genus *Andrena* for pollen. In fact, these bees have become so completely adapted to collecting this source of pollen, and their seasonal synchrony

FIGURE 4 Female carpenter bee (*Xylocopa tabaniformis orpifex,* Apidae) robbing nectar from base of a California fuchsia (*Epilobium canum,* Onagraceae) flower.

and the morphology of their pollen transport structures are so specialized, that they visit no other plants for pollen. So although the bees specialize on these flowers, they are not effective pollinators because they are small enough that they rarely contact the stigmas with the pollen they are collecting.

Floral Deceit

Many floral deceit mechanisms in flowers take advantage of basic behaviors and instincts in insects, especially feeding, mating, and oviposition. Some flowers are green or brownish rather than being colorful and have putrid or rotten meat aromas rather than sweet odors. These are highly attractive to carrion flies seeking source foods rich in amino acids and suitable as sites for oviposition and rearing of their young. Other flowers mimic females of bees or wasps in such fine details of form, color, odors, and texture that male insects actually attempt to mate with these models; in the process they pick up and distribute pollen from one flower to another. This process is called pseudocopulation because it relies on mating attempts by male wasps or bees. The flowers bloom during the brief time when male insects are on the wing before females of the species emerge. During this period, flowers are the only potential sources of "mates" for the male insects. Once females of the hymenopteran species emerge, the floral mimics are forsaken by males for the real female.

CROP POLLINATION

Economic Value

The most recognized benefit of pollinators to humanity is their value as pollinators of many of the crop plants that we use for food and fiber. The principal pollinator managed for crop pollination has been and currently is the honey bee, *Apis mellifera.* Although calculating the value of pollination by honey bees is far from exact, the most recent estimate of the annual value of increased production of crops contributed by honey bee colonies rented for pollination in the United States is over $14.5 billion, on average, over 1996 to 1998. The "free" contribution from native pollinators, especially other bee species, is even less measurable, but attempts are under way to estimate their value to sustainable agriculture farms in central California.

Exemplar Crops

Calfornia is perhaps the leading state in rentals of honey bee colonies for crop pollination. In large measure this is the result of the continuing increase in the acreage of almond, which has increased from about 36,000 hectares in the mid-1960s to over 200,000 hectares in 2001. At the recommended five to eight honey bee colonies per hectare, more than twice as many commercial colonies as exist in the state are required to accomodate the demand. Thus, there is a mass movement of colonies into California each year from as far as the Dakotas and Texas and beyond to pollinate the crop.

Alfalfa is another crop traditionally pollinated by honey bees that was widespread in California in the mid-1960s. However, the honey bee is not an effective pollinator of alfalfa over much of the crop's range. When two more efficient alternative pollinators (the introduced alfalfa leafcutting bee, *Megachile rotundata,* and the native alkali bee, *Nomia melanderi*) came under management for pollination of the alfalfa seed crop, much of the production shifted to the Pacific Northwest. Only the more southern areas of California continued to produce alfalfa seed solely with honey bees. Currently even some of these areas are augmented with alfalfa leafcutting bees.

Crop Pollinators Other Than Honey Bees

Although honey bees are readily available, easily transportable in large quantities, and generalist pollinators, they are not universal pollinators. There are some crop flowers, such as figs, that require insect pollination (specialized wasps) and cannot be pollinated by honey bees. There are other crop flowers that can be pollinated by honey bees but for which honey bees are not the most effective pollinators; these include crops such as alfalfa, squash, and greenhouse tomatoes. Another concern about excess reliance on a single pollinator for a wide variety of crops has been the widespread decimation of feral honey bees by the "vampire" mite, *Varroa;* increased cost of treating colonies to maintain healthy pollinating units; and reduction in numbers of beekeepers and colonies available for pollination. Warnings about this overdependence on honey bees and the general decline of pollinators because of factors such as loss of habitat and pesticides were issued in 1996 in a landmark publication by Buchmann and Nabhan.

One of the first insects introduced into North America specifically to pollinate a crop was the fig wasp, *Blastophaga psenes,* for production of edible Smyrna figs in southern California in 1899. Attempts to produce edible figs in California in the late 1800s failed until it was recognized that wasps from the wild ancestral caprifig, *Ficus carica,* were required. Edible figs contain predominantly pistilate flowers; pollen from male flowers of the caprifig is vectored by fig wasps. The growing of caprifigs containing introduced fig wasps, harvesting the fruits with a new generation of fig wasps, and then hanging them in baskets in trees of edible figs became a common practice in southern California, a process called "caprification."

Other insects that have been used for commercial pollination on a small scale include various flies, especially for breeding hybrid seed crops in cages by seed companies. Most of these have been muscoid flies, the pupae of which are readily available from insectaries. Results of large-scale open-field trials using carrion or other baits to attract flies have been equivocal for crop pollination.

Various species of non-*Apis* bees have been and are being studied for their management potential for pollination of

crops. In the late 1950s, studies were begun to manage two bee species (*M. rotundata* and *N. melanderi*) that are more effective than honey bees as pollinators of alfalfa for seed production.

More recently mason bees in the genus *Osmia* have been studied for pollination of crops in North America and Europe. *Osmia lignaria propinqua,* referred to as the "blue orchard bee," has been successfully managed to pollinate tree fruits in western North America. It is a cavity nester, like the alfalfa leafcutting bee, but uses mud partitions to create brood cells. Many of the management techniques were adapted from those used for *Megachile,* but modified to accomodate specific life history and behavioral attributes, including early spring activity and a single generation per year. Successes in managing other species of *Osmia* include *O. cornifrons* in Japan, *O. cornuta* in Spain, and *O. rufa* in Britain and France. In North America, *Osmia* are being studied for pollination of blueberries *(O. ribifloris)* and clovers *(O. sanrafaelae).*

In the late 1980s, major breakthroughs in year-round production of bumble bee colonies completely altered and expanded hothouse production of tomatoes. Tomato flowers require "buzz" pollination (i.e, vibration of flowers to release pollen from the apical pores of their specialized anthers) and bumble bees are much more effective at this than are honey bees or humans who hand pollinate with vibrating tools. This led to extremely large-scale movements of bumble bee colonies and queens, primarily *B. terrestris,* from central Europe, New Zealand, and Israel to many nations throughout the world. Some of this trafficking was unnecessary because closely related species were available for use at some locales. In Japan, environmental concerns have been expressed over the thousands of imported *B. terrestris* colonies and the subsequent establishment of this species outside the greenhouse environment. Males of this species will mate with queens of local species and produce viable offspring. Similar environmental concerns are being raised in other countries. In Canada and the United States, importation of *B. terrestris* was not sanctioned, but local bumble bees have been successfully reared and used in hothouse tomato production. East of the 100th meridian (a line that runs from central North Dakota to central Texas), *B. impatiens* was the species of choice and west of this line, *B. occidentalis.* These bees were used in their respective areas of distribution until 1998, when a disease outbreak was reported in the western species. Since then, the eastern *B. impatiens* has been imported into all western states, again causing concern in some areas that establishment outside its normal range may produce environmental damage.

Although pollinators are generally considered beneficial insects, importations to new areas should be done with considerable care to avoid environmental risks, such as introduction of disease organisms, nectar thieving, decreased pollination of nontarget native plants, enhanced pollination of introduced weeds, and genetic contamination of and competition for food and/or nest sites with native pollinators. The local fauna should be studied and searched for candidate species that could be suitably managed before any exotic species are introduced. Rearing technology is already available for several species of cavity-nesting bees, for bumble bees, and for some soil-nesting species. These methods may be modified and applied to local species that show promise for solving difficult pollination problems.

If it is deemed necessary to introduce a new pollinator, these should be thoroughly screened for biotic enemies (e.g., parasites, disease organisms) before being introduced. They should be monitored after release in the new environment to determine their efficiency in pollinating the target crop and to detect any adverse enviromental effects.

POLLINATOR DECLINE

Pollination biologists have worked for years studying the behavioral, ecological, and evolutionary intricacies of animal–plant pollination relationships. With the environmental movement of the early 1970s came awareness that mistreatment of the environment could also negatively affect these unique relationships. However, very few biologists sounded the alarm at that time that trouble was brewing for pollination relationships and pollinators, especially at the ecosystem or landscape level.

In the early to mid-1990s, the issue of pollination/pollinator problems was again brought to the attention of the biological community and the informed public, through the Island Press publication of *The Forgotten Pollinators* in 1996 by Stephen Buchmann and Gary Nabhan. Although the book contains anecdotal accounts of pollinator problems, the message was clear—everyone concerned with the pollination of plants needed to pay serious attention to what appeared to be an emerging picture of global pollinator decline. A subsequently important publication by Allen-Wardell *et al.* in 1998 pointed out the potential threat of pollinator decline to the human food supply. Adding to the general concern for declining pollinators was the fact that European honey bees in the New World tropics, and several western and southwestern U.S. states, were being systematically replaced with Africanized honey bees, and that all honey bees were being attacked and significantly reduced in numbers by two species of parasitic mites. Continued careless use of pesticides and bacterial infections were also cited as factors in the decline of honey bees.

The above trends have led to many scientific conferences worldwide to address the issue of pollinator decline and its potential consequences to crop plants and to native wildland plants. The first important global conference was held in 1998 in São Paulo, Brazil, where numerous relevant issues on pollinator decline were discussed by more than 60 pollinator/pollination professionals representing several New and Old World countries. The São Paulo meetings and subsequent meetings in other parts of the world have put into motion new research directions in the field, which are listed below.

1. Documenting pollinator decline. One of the main recommendations emerging from São Paulo was to seek

quantification of pollinator declines through careful case history studies. Long-term monitoring of pollinators and comparative assessments at specific study areas were recommended approaches for gathering the needed data on decline. This has proved to be quite a challenge because of natural fluctuations in pollinator populations and the lack of long-term baseline data. Where decline has been detected, loss of habitat is suggested as the main cause.

2. Causes of decline and restoration of pollinators. Habitat loss is a convenient general explanation for the cause of decline, but detailed information is needed on the precise factors causing decline. Further, specific causes of decline are not always obvious, and this becomes an important issue when considering projects to restore or establish pollinators and their required resources (often diverse) to an area.

3. More research on non-honey-bee pollinators. Honey bees can be likened to a monoculture in agriculture; overdependence on a single organism in agriculture can lead to disastrous results when natural mortality factors get out of balance. Such was the case with blight on a large portion of the U.S. corn crop several years ago. Because honey bees are now showing great vulnerability to two species of parasitic mites, more research on these parasites and on other bees (especially native solitary species) is being conducted.

4. Conservation of pollinators. Much has been written on conserving pollinators and especially bees. Unfortunately, there is little evidence to suggest that conservation recommendations, which would result in measurable increases in pollinator numbers, have been put into action. Many possibilities exist for increasing pollinators through manipulations of preferred food plants, nonfloral plant products (e.g., resins), and planned efforts to increase nesting sites and other requisites such as alternate food plants for moths, beetles, wasps, flies, bats, etc.

5. Increasing awareness of pollinator services. The conservation of pollinators and calling attention to vital services provided by pollinators become issues of information transfer that biologists must address. They are the only professionals who know the needs and fragilities of small organisms such as bees, flies, beetles, and nocturnal organisms such as moths and bats. Pollinator/pollination professionals will need to form closer working relationships with policy-makers, land stewards, and a wide variety of government and nongovernmental organizations to realize future successes in the management of pollinators.

There are at least two courses of action that pollinator biologists could pursue now to assist declining pollinator populations. First, they can collaborate with other biologists who are also concerned about decline of their specific organisms (e.g., birds, mammals) and habitat. Building a coalition of concerned biologists with integrated management plans for habitat protection for several threatened species could be effective if land stewards, associated with the habitat, were receptive and willing to participate in some way as stakeholders in the project. Second, biologists could also work in a variety of ways toward conserving areas known to naturally harbor healthy populations of pollinators, preferably several types. Biologists are aware, through years of field experience, which areas have good diversity and abundance of, for example, bees and moths.

TROPICAL POLLINATION

The literature is filled with fascinating case histories of individual tropical plants and their pollinators. More recently, researchers have been investigating pollination systems involving groups of prominent pollinators (e.g., bees, bats, moths, hummingbirds) and their plants in major tropical life zones. A few larger, long-term studies from the New and Old World tropics have also provided the first community pollination patterns for a high percentage of the representative plant life forms. These latter studies are particularly helpful in elucidating diversity and frequency of pollination systems for conservation work, as well as for interesting comparisons with temperate environments.

In Table I, pollination systems of two lowland forest sites and one midelevation cloud forest (1200–1800 m) in Costa Rica are compared and contrasted with one lowland forest site in Malaysia. From about 40 to 70% of the surveyed plant species in each of the four sites were pollinated by bees. Although less frequent than bees, birds were important pollinators in Costa Rica's cloud forest and wet forest and in the Malaysian forest. Beetles were important in the Costa Rican wet forest and in Malaysia; moths were important in all three Neotropical forests. The four sites had numerous plant species that were visited (and pollinated) by a variety of general insects.

The importance of animals, especially insects, as pollen vectors of tropical plants was clearly demonstrated in a classic paper by Bawa in 1974, in which he reported that most tree species in a lowland dry forest of Costa Rica were obliged to outcross. Through controlled pollinations, he demonstrated that a high percentage of the tree species tested were self-incompatible (incapable of self-pollination) or dioecious (having separate male and female plants of a species). Until that time, many biologists believed that self-pollination was probably the rule in tropical forests. Flower-visiting animals, and especially insects, were not viewed as capable travelers between widely distributed tropical plants. Subsequent studies on interplant movements and foraging patterns of these animals have substantiated their capacities to move among flowering plant species at levels required to produce abundant fruit crops.

Some general, but limited, comparisons of tropical versus temperate pollination systems are possible with the information available in the literature. First, the flora of many low- to midelevation temperate habitats is mostly pollinated by bees. Estimates vary widely from 70 to more than 90%, depending on locality. Flies and Lepidoptera may also be important in

TABLE I Percentages of Pollination Systems Represented in Four Tropical Forests

	Costa Rica			Malaysia
	Dry forest[a]	Cloud forest[b]	Wet forest[c]	Dipterocarp forest[d]
Bee	69.5%	44.4%	38.4%	50.4%
Moth	7.3	6.2	8	1.1
Bat	2.4	3.1	3.6	1.5
Bird	2.2	9.6	14.9	7
Beetle	1.1	2.4	12.7	20.7
Wasp (large)	0.2	1.6	2.5	0
Butterfly	0.9	1.9	4.3	2.2
Fly	0	0.2	1.8	0
Fig wasp	0.4	0.5	?	0
Mammal (arboreal)	0	0.1	0	0.4
General insect	15.3[e]	20.4[e]	11.2[e]	13.7
Wind	0.9	5.2	2.5	0
Miscellaneous	0	4.5	0	3

[a]*N* = 465 plant species. Source: Frankie *et al.* (2003).
[b]*N* = 1100 plant species. Source: Frankie *et al.* (2003).
[c]*N* = 276 plant species. Source: Kress and Beach (1994).
[d]*N* = 270 plant species. Source: Momose *et al.* (1998).
[e]Several plant species listed in this category may prove to be primarily bee pollinated.

the pollination of temperate plants. Second, the diversity of pollination systems is comparatively lower in temperate environments because most lack, for example, the more specialized bird, bat, beetle, and fig wasp systems. Third, when lowland tropical and temperate forests are compared, the high diversity of tropical trees and their dependence on animal pollination become immediately apparent. Temperate forests have relatively low tree species diversity and most, such as conifers, oaks, willows, elms, and maples, are wind pollinated. In contrast, wind pollination is rare in tropical forests (Table I).

SPECIALIZED VS GENERALIZED POLLINATION

As mentioned previously, there is ongoing controversy about specialized versus generalized pollination systems and the associated concept of pollination syndromes that propose to characterize a plant as to a particular pollinator type. There can be little doubt that some plants have highly specialized systems that can be easily characterized by floral morphology and behavior alone, such as fig flowers and fig wasps, certain orchids and their specific bee relationships, and long, white-tubed fragrant flowers that open at night and are pollinated by long-tongued hawk moths (Fig. 2). There are, however, many instances in which flowers attract a wide variety of visitor types, making characterization of pollination syndromes difficult. For example, there are examples of "large-bee flowers" that are regularly visited by small bees, butterflies, and wasps; hawk moths that visit "bat flowers"; bees that visit both "hawk moth flowers" and "bat flowers" the morning after. Further, there are flower types that attract a wide diversity of insect visitors (Table I). Who are the pollinators and who are the visitors to these flower types? Do many or most plants have the option of being pollinated by a variety of potential vectors?

The answers to these questions will be forthcoming through carefully planned experimental studies that include evaluations of all visitors and their capacity to transport pollen on their body parts, as well as between plants. Floral behavior must be studied simultaneously, especially with regard to breeding system and period of stigmatic receptivity. These case studies should take much of the speculation and guesswork out of pollinators, visitors, and pollination ecology.

CHEMICAL ECOLOGY AND POLLINATION

Chemical communication between flowering plants and their pollinators and between conspecific pollinators in relation to floral resources is commonplace in many natural communities. Flowers release a variety of odors that attract pollinators and other visitors. Some of these are sweet and highly fragrant, as with many "moth flowers" and some "bee flowers." Some are unpleasant to humans but highly attractive to flies and beetles. Other floral fragrances fall somewhere between fragrant and unpleasant such as musky odors associated with some flowers that attract a variety of visitor types.

In the New World tropics, chemicals are emitted from certain orchid species that attract only males of the Eugolssini tribe of bees, better known as orchid bees. These orchids and "their" bees have evolved a unique relationship in which chemical substances are scratched from special regions of the orchid flowers by male bees. The compounds are collected in special leg glands that have unique apertures on the hind legs. According to theory, the scratched compounds are metabolized

and transformed into chemical messages or pheromones that male bees use for attracting females for mating. Some male orchid bees form elaborate leks or mating rituals, which they use to lure females. During the process of scratching chemicals, orchids belonging to genera such as *Catesetum, Cycnoches,* and *Stanhopea* have cleverly evolved elaborate mechanisms for attaching or actually gluing a pollen packet (or pollinia) on the bee for transport to the next orchid, thereby effecting cross pollination. Each orchid species glues its pollinia on a characteristic location of the bee. Where orchid diversity is rich, it is common to see some bees with pollinia on their heads, other bees with pollinia on their thoraces or abdomens, and still other bees with pollinia on more than one body location, indicating their visits to more than one orchid species.

There is also chemical communication among members of some social and solitary bee species. Some honey bee species scent mark flowers, and this informs others of the same species of a very recently visited flower. The mark also serves to alert a bee that it has just visited a particular flower that it marked, thereby conserving its energy and time. This behavior has also been observed frequently in large carpenter bees *(Xylocopa);* chemicals of the mandibular and mesosomal glands were found responsible for the scents. Stingless bees in the Neotropics regularly mark flowers and nearby vegetation in establishing a scent trail back to inform the nest where a good floral source can be located. Some bumble bees *(Bombus)* also scent mark flowers, making these flowers less attractive to other foraging bumble bees.

CONCLUDING REMARKS

Despite years of study and an enormous literature, we still have much to learn about pollination and pollinators. The field of study is particularly challenging today because of the many questions surrounding pollination/pollinator relationships in human-impacted environments. Many questions (and problems) will require new approaches and methods and will need to be better integrated with societal needs and structures.

Suggested studies for the future include: (a) more detailed work on chemical relationships between pollinators, visitors, and flowers; (b) more attention paid to actively conserving, protecting, and restoring pollinators at local, regional, national, and international levels; (c) improving methods for managing pollinators for production of human food crops; (d) development and transfer of information on pollination and pollinators to a wide variety of new audiences such as policy/decision-makers, government agencies, nongovernmental conservation organizations, and managers; and (e) developing monitoring methods to gauge effects of global warming on pollinators and the plants they pollinate.

See Also the Following Articles

Apis Species • Beekeeping • Conservation • Neotropical African Bees

Further Reading

Allen-Wardell, G., *et al.* (1998). The potential consequences of pollinator declines on the conservation of biodiversity and stability of food crop yields. *Conserv. Biol.* **12,** 8–17.

Bawa, K. S. (1974). Breeding systems of tree species of a lowland tropical community. *Evolution* **28,** 85–92.

Buchmann, S. L., and Nabhan, G. P. (1996). "The Forgotten Pollinators." Island Press, Covelo, CA.

Daily, G. C. (ed.) (1997). "Nature's Services: Societal Dependence on Natural Ecosystems." Island Press, Covelo, CA.

Delaplane, K. S., and Mayer, D. F. (2000). Crop pollination by bees. CAB Int., Tucson.

Frankie, G. W., *et al.* (2003). Flowering phenology and pollination systems diversity in the seasonal dry forest. *In* "Biodiversity Conservation in Costa Rica: Learning the Lessons in a Seasonal Dry Forest" (G. W. Frankie, A. Mata, and S. B. Vinson, eds.). University of California Press, Berkeley.

Free, J. B. (1993). "Insect Pollination of Crops," 2nd ed. Academic Press, San Diego.

Kress, W. J., and Beach, J. H. (1994). Flowering plant reproductive systems. *In* "La Selva, Ecology and Natural History of a Neotropical Rain Forest" (L. A. McDade *et al.,* eds.). University of Chicago Press, Chicago/London.

Matheson, A., Buchmann, S. L., O'Toole, C., Westrich, P., and Williams, I. H. (eds.) (1996). "The Conservation of Bees." Academic Press, San Diego. [Linnean Soc. Symp. Ser. No. 18].

Michener, C. D. (2000). "Bees of the World." Johns Hopkins University Press, Baltimore.

Momose, K., Yumoto, T., Nagamitsu, T., Kato, M., Nagamasu, H., Sakai, S., Harrison, R. D., Itioka, T., Hamid, A. A., and Inoue, T. (1998). Pollination biology in a lowland Dipterocarp forest in Sarawak, Malaysia. I. Characteristics of plant-pollinator community in a lowland Dipterocarp forest. *Am. J. Bot.* **85,** 1477–1501.

Morse, R. A., and Calderone, N. W. (2000). The value of honey bees as pollinators of U.S. crops in 2000. *Bee Cult.* March 2000, 1–15.

O'Toole, C., and Raw, A. (1991). "Bees of the World." Facts on File (http://www.factsonfile.com/).

Proctor, M., Yeo, P., and Lack, A. (1996). "The Natural History of Pollination." Timber Press, Portland, OR.

Roubik, D. W. (1989). "Ecology and Natural History of Tropical Bees." Cambridge University Press, Cambridge, UK. [Cambridge Tropical Biol. Ser.]

Roubik, D. W. (ed.) (1995). "Pollination of Cultivated Plants in the Tropics." FAO, Rome. [Agric. Services Bull. No. 118].

Pollution, Insect Response to

David M. Rosenberg
Freshwater Institute, Winnipeg, Canada

Vincent H. Resh
University of California, Berkeley

Pollution is essentially the wrong substance, in the wrong place, in the wrong concentration, at the wrong time. More formally, pollution can be defined as the introduction of human-made substances (or natural substances released by

humans) and forms of energy into the environment that are likely to damage ecosystems or their constituents, amenities, or structures.

Insects, like other living organisms, are affected by pollution. However, insects are also used to assess the effects of pollution as surrogates or representatives of the larger assemblages of organisms in communities and ecosystems. We refer to insects in this latter role as biomonitoring agents.

Pollution can be caused by a variety of substances and activities: sewage and other organic enrichment, fertilizers (e.g., nutrients such as phosphorus), siltation (e.g., from erosion), pesticides (e.g., herbicides, fungicides, insecticides), metals (e.g., cadmium, mercury, selenium), organic compounds (polychlorinated biphenyls, polycyclic aromatic hydrocarbons, industrial atmospheric emissions (e.g., sulfur dioxide and NO_x, which are precursors to acid rain; greenhouse gases such as carbon dioxide and methane), radiation (as in the Chernobyl disaster), heat (e.g., thermal pollution from power plants), and habitat destruction (e.g., clear-cutting, stream channelization, reservoir creation). Disturbances that have consequences similar to those of human activities can also be caused by natural events such as volcanic eruptions or forest fires, but the products of these disturbances do not fit our definition because they are not deliberately introduced by humans.

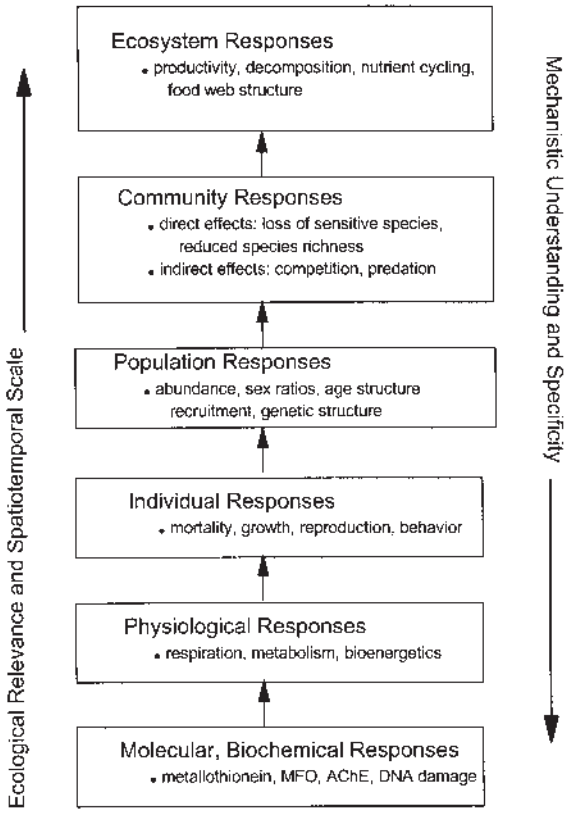

FIGURE 1 Relationship between ecological relevance, spatiotemporal scale, mechanistic understanding, and specificity across levels of biological organization. MFO, mixed-function oxidase; AChE, acetylcholinesterase; DNA, deoxyribonucleic acid. (Reproduced from W. H. Clements, 2000, Integrating effects of contaminants across levels of biological organization: An overview. *J. Aquat. Ecosyst. Stress Recovery* **7**, 113–116, with kind permission from Kluwer Academic Publishers.)

POLLUTION EFFECTS ON INSECTS

The effects of pollution on insects occur at a variety of spatial and temporal scales. For example, effects can occur at the molecular level in fractions of seconds (i.e., biochemical effects), at the ecosystem level over several decades, and at various scales in between these extremes (Fig. 1). The most easily detected responses are at the level of the individual (as in bioassays), in which evaluation is often based on whether an insect lives or dies when exposed to a contaminant. However, population and community levels are generally used when effects are examined in nature (Fig. 1). Within a population, we can see shifts occurring in the frequency of organisms of different sizes. For example, early instars in a population seem to be more susceptible than later instars because early instars have higher surface-to-volume ratios, they are more active, and they have thinner cuticles. Eggs, pupae, and diapausing insects are usually more resistant stages.

Common community-level changes observed with the onset of pollution involve decreases in species richness (i.e., the total number of species), decreases in species evenness (i.e., the distribution of numbers of individuals among species), and alterations in species composition. Severe sewage pollution in fresh waters presents a good example of the kinds of changes that can occur. With sewage input, the total number of species decreases immediately downstream of a sewage source and species evenness decreases because the remaining tolerant organisms proliferate. Also, pollution-tolerant organisms such as dronefly maggots *(Eristalis tenax),* some species of midge (Chironomidae) larvae, and aquatic earthworms replace less tolerant ones, and these tolerant organisms can reach extremely high densities.

Alteration of feeding-group structure is another response to pollution. For example, the removal of a riparian zone surrounding a stream can lead to increased production of algae because of more open conditions and greater sunlight reaching the stream substrate, a decline in the leaves falling into the stream, and an increase in grazing insects (feeding on algae), which replace the insects that would normally feed on leaves.

More subtle community-level effects are also possible. For example, the suppression of parasites by air pollutants may enhance outbreaks of forest insect pests whose populations are normally held in check by the parasites. These indirect effects (Fig. 1) are often more difficult to detect than the direct effects discussed above.

Natural variability is a problem in evaluating the effects of pollution on insects. Only the accrual of long-term baseline data on insect species present and their abundances, and the use of an experimental approach to establish causation, can lead to understanding of the effects of pollutants on insects.

USE OF INSECTS IN BIOMONITORING

The responses of insects to pollution can be used in biomonitoring of air, land, and water quality. The most extensive biomonitoring has been developed in aquatic habitats, perhaps because aquatic insects are directly exposed to water pollution and their responses are easy to measure.

Insects present a number of advantages for biomonitoring: (1) they are ubiquitous, so they are exposed to pollution in many different habitats; (2) the large number of species offers a range of responses; (3) the sedentary nature of many insects allows spatial analysis of pollutant effects; and (4) their long life cycles allow temporal analysis of pollutant effects. Unlike relying on instantaneous measurement of physical and chemical variables, the use of living organisms, like insects, provides a temporal integration of pollutant effects over their life span.

Biomonitoring can be done at a variety of spatial and temporal scales (Fig. 1). The smallest scale is biochemical; for example, exposure of stonefly species to the pesticide fenitrothion can be measured by depression of acetylcholinesterase activity in the stonefly head. The largest scale is the ecosystem, in which measures of processes such as productivity and decomposition can indicate pollutant stress. However, biomonitoring is most commonly done at individual, population (or species assemblage), and community levels. Common examples of biomonitoring at the individual level include the use of insects as sentinel organisms and for measuring morphological deformities. The use of morphological deformities to measure pollutant effects is more common in freshwater than in terrestrial habitats. The head capsules of midge larvae are frequently used for this purpose. Bioassays often involve exposure of individuals to potential pollutants to detect sublethal responses (e.g., diminished growth or fecundity) or lethal responses (e.g., mortality).

Biotic indices are popular ways to summarize information and assess pollution at the population or species assemblage levels. For example, indices of the trophic status of lakes have been developed using midges, and an index of acid stress is available that uses a group of macroinvertebrate organisms, each of which has a different sensitivity to acidification.

However, biomonitoring is most frequently done at the community level, and measures vary from simple taxa richness to biotic indices that use the whole community (e.g., Family Biotic Index) to complex multivariate statistics that use the "reference condition" approach.

Bioassays are widely used in determining toxicity of industrial by-products before and after they are released into the environment and can be a valuable adjunct to field biomonitoring. They can be used to answer questions such as (1) At what concentration does a pollutant become toxic to a certain insect species? (2) What are the physical and chemical conditions under which the pollutant is most harmful? (3) Which stage of the life cycle is most susceptible to the pollutant? (4) What are the effects of acute versus chronic exposure? (5) Which species are most susceptible to the pollutant? Bioassays can be as simple or as complex as needed (e.g., single versus multiple species, static or flowthrough setup, laboratory beaker versus field mesocosm).

Biomonitoring using insects can involve volunteers as well as professionals. Work with insects is usually labor intensive, and volunteers can provide much of this labor. In fact, volunteer biomonitoring programs involving school classes and local community groups are currently very popular throughout the world.

In conclusion, although pollutants have affected living organisms, habitat destruction may be more important in determining whether organisms can recover. For example, it may be possible to clean air and water of their pollutants but recovery of ecosystems is impossible if the habitat remains impaired.

See Also the Following Articles

Aquatic Habitats • Conservation • Forest Habitats • Grassland Habitats • Greenhouse Gases, Global Warming, and Insects • Soil Habitats • Urban Habitats

Further Reading

Fjellheim, A., and Raddum, G. G. (1990). Acid precipitation: Biological monitoring of streams and lakes. *Sci. Total Environ.* **96,** 57–66.

Flannagan, J. F., Lockhart, W. L., Cobb, D. G., and Metner, D. (1978). Stonefly (Plecoptera) head cholinesterase as an indicator of exposure to fenitrothion. *Manit. Entomol.* **12,** 42–48.

Heliövaara, K., and Väisänen, R. (1993). "Insects and Pollution." CRC Press, Boca Raton, FL.

Hellawell, J. M. (1986). "Biological Indicators of Freshwater Pollution and Environmental Management." Elsevier, London.

Hilsenhoff, W. L. (1988). Rapid field assessment of organic pollution with a family-level biotic index. *J. North Am. Benthol. Soc.* **7,** 65–68.

Kerr, M., Ely, E., Lee, V., and Desbonnet, A. (1994). "National Directory of Volunteer Environmental Monitoring Programs," 4th ed. U.S. Environmental Protection Agency and Rhode Island Sea Grant, University of Rhode Island, Narragansett. [EPA 841-B-94-001]

Reynoldson, T. B., Norris, R. H., Resh, V. H., Day, K. E., and Rosenberg, D. M. (1997). The reference condition: A comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *J. North Am. Benthol. Soc.* **16,** 833–852.

Rosenberg, D. M., and Resh, V. H. (eds.) (1993). "Freshwater Biomonitoring and Benthic Macroinvertebrates." Chapman & Hall, New York.

Saether, O. A. (1979). Chironomid communities as water quality indicators. *Holarctic Ecol.* **2,** 65–74.

Polyembryony

Michael R. Strand
University of Georgia

Polyembryony is a form of clonal development in which a single egg produces two or more genetically identical

offspring. Many animals are sporadically polyembryonic, including humans, who occasionally give birth to identical twins. However, a few groups of parasites (some cestodes, treamatodes, and insects), colonial aquatic invertebrates (oligochaetes, bryzoans), and mammals (armadillos) are obligately polyembryonic. Among these obligately polyembryonic animals, insects produce the largest broods, with some species generating more than 3000 offspring per egg. There are hundreds and perhaps thousands of polyembryonic insects, but all known species occur in either the order Hymenoptera (bees, wasps, and ants) or the order Strepsiptera. Several species of polyembryonic wasps have been studied in detail, whereas little is known about polyembryonic strepsipterans.

In the Hymenoptera, polyembryony occurs in selected genera from four families: the Braconidae, Platygasteridae, Encyrtidae, and Dryinidae. Polyembryony in these groups clearly arose from ancestors that produced only a single offspring per egg (monoembryonic), because the most basal members of these large families are all monoembryonic. The phylogenetic distance between these families also indicates that polyembryony has evolved independently at least four times in the Hymenoptera. Despite multiple independent origins, all polyembryonic wasps share the common biology of being parasitoids that lay their eggs into the egg or larval (nymphal) stage of their insect hosts. After oviposition, the egg of a polyembryonic wasp develops into a single embryo. This embryo then proliferates into an assemblage of embryos called a polygerm or polymorula. The majority of embryos in the polymorula develop into larvae when the host molts to its final instar. These so-called reproductive larvae consume the host, pupate, and emerge as adult wasps that seek mates (if male) or new hosts (if female). In some species from the family Encyrtidae, however, a small proportion of the embryos in the polymorula develop into what are called precocious larvae. Precocious larvae are morphologically distinct from reproductive larvae and function as a sterile soldier caste that defends their reproductive-caste siblings from competitors. Embryonic development of polyembryonic wasps differs in several respects from other insects. The presence of a caste system in some polyembryonic wasps also offers insights into conditions favoring the evolution of extreme reproductive altruism.

LIFE HISTORY AND DEVELOPMENT OF POLYEMBRYONIC WASPS

Polyembryony in insects was first described by the Italian naturalist F. Caldani almost 200 years ago. The development of polyembryonic wasps has been most thoroughly studied in the encyrtid wasp *Copidosoma floridanum,* which parasitizes the eggs of plusiine moths such as the cabbage looper *Trichoplusia ni* (Fig. 1A). Unlike terrestrial insects, whose eggs contain large amounts of yolk and are enclosed in a rigid chorion, *C. floridanum* eggs contain no yolk and are surrounded by a thin chorion (Fig. 1B). After oviposition, the egg undergoes complete (holoblastic) cleavage to form a single embryo that ruptures out of its chorion to continue development unconstrained by the eggshell. This embryo is called the primary morula and consists of approximately 200 embryonic cells surrounded by an extraembryonic membrane of polar body origin (Figs. 1C and 1D). After the host emerges from its egg, the cells of the primary morula continue to divide inside the host larva to form additional embryonic masses that consist of embyronic cells surrounded by the extraembryonic membrane. These embryonic masses are called secondary morulae and the entire mass is called a polymorula (Fig. 1E). All morulae formed during this period of proliferation result from the invagination of the extraembryonic membrane and subsequent partitioning of embryonic cells (Figs. 1F–1H). The process of membrane invagination and cell partitioning repeats itself an indeterminate number of times during the first through the fourth instars of the host caterpillar. This results in an increasing number of secondary morulae per host, but each morula contains progressively fewer embryonic cells. For example, from 5 to 10 secondary morulae that each contain several hundred embryonic cells are present in a first-instar host, whereas an average total of 1200 morulae that each contain about 20 cells are present in a fourth-instar host.

A few morulae in the polymorula differentiate during the host's first through fourth instar and develop into precocious larvae (Fig. 1I). One or two precocious larvae develop when the host is a first-instar larva and from 2 to 10 precocious larvae develop during the host's second through fourth instar. As a result, the number of precocious larvae increases progressively as the host caterpillar grows. An average total of 40 precocious larvae develop, which equals approximately 4% of the total number of larvae produced per host. The remaining embryos always differentiate in the host's fifth instar and develop into reproductive larvae (Fig. 1J). Reproductive larvae rapidly consume the host and then pupate inside the remnant cuticle, forming a mummy (Fig. 1K). In contrast, precocious larvae never molt and die from desiccation after the host is consumed by their reproductive-caste siblings. Although no detailed embryological information is available outside *C. floridanum,* descriptions of polyembryonic braconids, platygasterids, and dryinids suggest that these wasps also lay small, yolkless eggs that undergo complete cleavage to form a single embryo. This embryo then proliferates into additional embryos. However, polyembryonic wasps in these families produce only embryos that develop into reproductive larvae (i.e., no precocious larvae or caste system exists).

REGULATION OF POLYEMBRYONIC DEVELOPMENT

Caste formation in social insects like ants, bees, or termites is usually mediated by environmental factors (photoperiod, crowding, pheromones, nutrition) acting on endocrine physiology. Environmentally induced alterations in hormonal

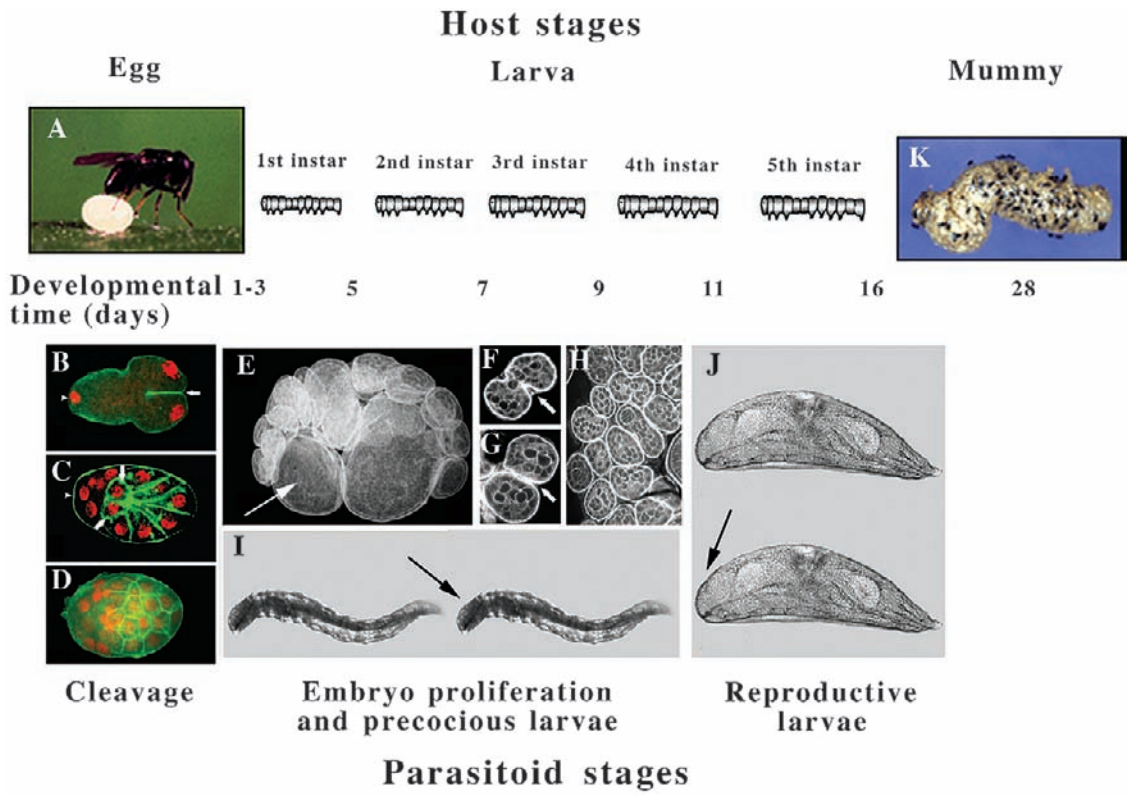

FIGURE 1 Development of the polyembryonic wasp *C. floridanum.* The top shows the egg, larval, and mummy stages of the host, the middle shows developmental time from egg laying to emergence of adult wasps from the mummy, and the bottom shows key events during embryogenesis of the wasp. (A) The female wasp oviposits into the egg stage of the host, *T. ni.* (B) Soon after oviposition, the 50-μm-long wasp egg undergoes the first cleavage that results in two equal-sized blastomeres at the posterior (arrow) and a polar cell containing the polar nucleus at the anterior (arrowhead). (C, D) The syncytial extraembryonic membrane (arrowhead) derived from the polar cell begins to envelop the blastomeres that continue to divide (arrows). Soon after envelopment begins, the embryo ruptures out of the chorion, and by 24 h after oviposition the primary morula has formed. (E) At the end of the host's egg stage and throughout the first through fourth instar, embryo proliferation occurs to produce an increasing number of secondary morulae. Some embryos during this period undergo morphogenesis (arrow) and develop into precocious larvae. (F, G) Additional secondary morulae are formed during proliferation by ingrowth of the extraembryonic membrane (arrows) surrounding an embryo, which partitions the enclosed embryonic cells and results in formation of two embryos. (H) More than 1000 secondary morulae are formed in this manner by the host's fourth instar. (I) Precocious larvae formed during the host first through fourth instar are serepentine in shape and possess a well sclerotized head (arrow). (J) Reproductive larvae that emerge during the host fifth instar are rounded in shape and have a very weakly sclerotized head (arrow). These larvae consume the host at the end of the fifth instar and pupate inside the host cuticle to form a mummy. (K) One-millimeter-long adult wasps emerge from the mummy in approximately 28 days.

state in turn affect the development and/or behavior of individuals in the colony to produce distinct castes. In contrast, insect embryogenesis is primarily regulated endogenously by the coordinated expression of maternal and zygotic factors known as patterning genes. Signaling molecules and determinants associated with specific cell lineages also play a significant role in regulating embryogenesis of other animals. Within this framework, caste formation and embryogenesis in polyembryonic wasps could be regulated by both environmental signals from the host and endogenous factors in embryos.

Polyembryonic wasps differ from other caste-forming insects in that both castes develop from the same egg, coexist in the same environment (the host), and develop in proximity to one another in the polymorula. Nonetheless, the development of precocious larvae during the host first through fourth instar and reproductive larvae in the host fifth instar suggests that host environment may influence caste formation in *C. floridanum.* Molting and metamorphosis of the host, like in most insects, are regulated by the steroid ecdysone and sesquiterpenoid juvenile hormone (JH). A rise in the ecdysteroid titer stimulates the host to molt, whereas the JH titer influences the type of molt that occurs. *T. ni* larvae parasitized by *C. floridanum* molt and develop almost identical to normal, unparasitized caterpillars. The obvious exception is that *C. floridanum* reproductive larvae consume the parasitized larva at the end of its fifth larval instar, whereas unparasitized *T. ni* pupate after the fifth instar. The fact that host JH titers are elevated in early larval instars and decline in the final instar suggests that embryos develop into precocious larvae when the host JH titer is elevated, whereas embryos develop into reproductive larvae under conditions of low host JH titer.

However, a number of different experiments collectively indicate that neither JH, ecdysone, nor other environmental factors are the primary factors that regulate morula proliferation, morphogenesis, or the proportion of embryos that develop into each caste. Instead, these events appear to be regulated

by endogenous factors in individual embryos. One class of endogenous factors involved in regulation of morphogenesis contains the same genes that control pattern formation in all insects. Most insects (*Drosophila* is the best studied insect in this regard) lay eggs that are rich in yolk and that have a rigid chorion that protects the embryo from injury and desiccation. At the beginning of embryogenesis, *Drosophila* and most other insects undergo a syncytial phase of development whereby nuclear divisions occur without cell division to produce thousands of nuclei in a single-celled embryo. This syncytial phase is crucial for axial patterning as it allows RNAs of the maternal coordinate genes *bicoid* and *nanos* to be properly distributed within the egg. These maternal RNAs are then translated to form a gradient of regulatory proteins that control where the head, thorax, and abdomen of the embryo will develop. Localization of these factors in the posterior of the egg is also coupled to localization of another set of determinants that specify germ cells, which will give rise to sperm or eggs when the insect becomes an adult. In effect, the entire body plan and reproductive capacity of the future insect are determined immediately after the egg is laid. After syncytial cleavage, insect embryos cellularize to form a blastoderm. Subsequent specification of individual thoracic and abdominal segments is then regulated by a cascade of zygotic genes (*gap, pair-rule, segment polarity,* and *Hox* genes).

The embryos of polyembryonic wasps differ from those of *Drosophila* and most other insects in several ways. First, the eggs of polyembryonic wasps lack yolk and a rigid chorion, and embryogenesis proceeds after the embryo emerges from the eggshell. Second, the polarity of the egg is not directly linked to the axial polarity of future larvae, given that the primary morula proliferates to form thousands of embryos before morphogenesis. Since each embryo is randomly oriented within the polymorula, axial polarity must be reestablished after proliferation. Third, polyembryonic wasp eggs lack a syncytium and instead cellularize at the four-cell stage. Despite these differences, many of the genes that regulate pattern formation in other insects are also operative in polyembryonic wasps. However, all of these patterning genes are expressed after embryos begin morphogenesis and none of them are expressed during embryo proliferation, when the number and types (precocious or reproductive) of embryos increase. This suggests that patterning genes have not been co-opted to perform these novel functions in polyembryonic wasps. However, other factors that could play a role in regulating proliferation and caste fate are genes that control germ cell specification and certain cell signaling molecules that establish pattern in tissues such as imaginal discs.

THE EVOLUTION OF POLYEMBRYONY

The idea that polyembryony has evolved several times in the Hymenoptera but has not arisen in any other insects outside the Strepsiptera suggests that wasps possess unique preadaptations that have favored the evolution of this unusual form of development. The Hymenoptera exhibit incredible diversity in life history (Fig. 2). Most hymenopterans are in the suborder Apocrita, which consists of free-living species (pollen-nectar feeders and predators) and parasitoids (Fig. 2). Typical free-living apocritans such as ants and bees are restricted to a single group known collectively as the Aculeata (Chrysidoidea, Vespoidea, and Apoidea), whereas the other major superfamilies comprise primarily parasitoids (Fig. 2).

Parasitoids exhibit two different developmental strategies. Some species develop as ectoparasitoids that lay their eggs externally on hosts and feed as larvae by rasping a hole through the host's cuticle. Other species develop as endoparasitoids that inject their eggs into the hemocoel of the host where the larvae feed on blood or tissues. Phylogenetic analyses indicate that all aprocritans likely evolved from an ectoparasitic ancestor related to contemporary orussoids (Fig. 2).

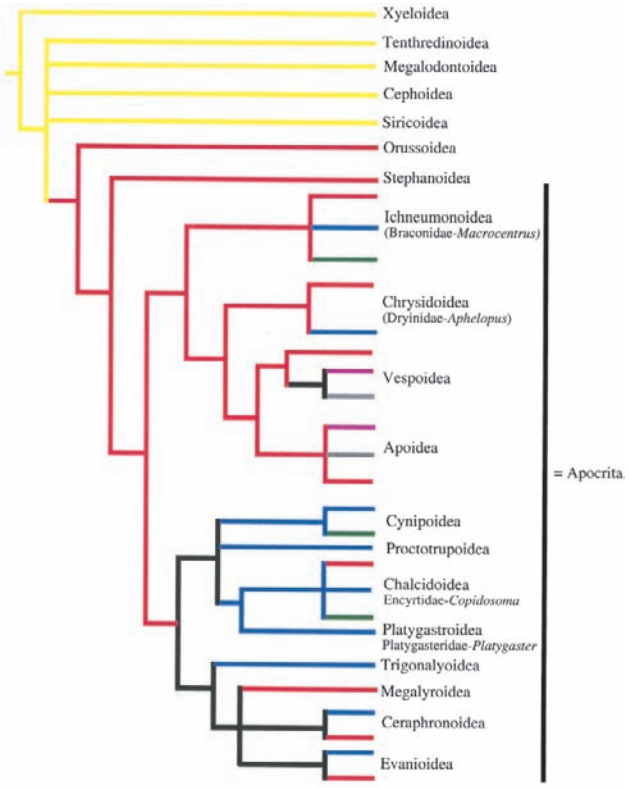

FIGURE 2 Consensus phylogeny and the evolution of different life histories in the order Hymenoptera. The listed life histories are herbivores (phytophagous) (yellow), ectoparasitoids (red), endoparasitoids (blue), gall formers (green), pollen-nectar feeders (magenta), and predators (gray). Unknown relationships are indicated in black. The specific families and genera in which polyembryony has evolved are indicated under the appropriate superfamily. The most primitive hymenopterans include several superfamilies of herbivores commonly known as sawflies. The advanced hymenopterans (Apocrita) form a monophyletic assemblage derived from an ectoparasitoid ancestor likely similar to orrusoids, which parasitize wood-infesting beetles. Endoparasitism and other life histories have thereafter arisen multiple times in different superfamilies from ectoparasitic ancestors. Among endoparasitic lineages, polyembryony is known to have evolved four times. (Figure developed from the consensus phylogenies and discussion of Whitfield, 1998).

Thereafter, endoparasitism evolved independently at least eight times from different ectoparasitic ancestors. The four hymenopteran families that contain polyembryonic species are in different superfamilies (Ichneumonoidea, Platygastroidae, Chalcidoidea, and Vespoidea), and the nearest relatives to each polyembryonic taxon are monoembryonic endoparasitoids. Thus, polyembryony itself has evolved at least four times and in each case this form of development arose from a monoembryonic, endoparasitic ancestor (Fig. 2).

The correlation between endoparasitism and the subsequent evolution of polyembryony is unlikely to be coincidental. Free-living insects such as *Drosophila,* ants, or honey bees develop in a terrestrial environment independent of the parent. Adaptations for survival include a thick chorion to protect the embryo and an abundant yolk source to supply nutrients for development. Ectoparasitoids encounter the same environmental conditions, but endoparasitoids develop in the nutrient-rich blood of another insect where protection from desiccation and prepackaging of a yolk source are no longer required. Unconstrained by the need for a prepackaged source of nutrition (yolk) and a chorion, the potential exists for embryos of endoparasitoids to shift away from syncytial cleavage and toward early cellularization and complete cleavage. Unconstrained by a rigid chorion, the embryos of endoparasitoids also have the potential to increase significantly in volume during embryogenesis. If endoparasitism is an important preadaptation for the ultimate evolution of polyembryony, the loss of yolk and alterations in early development associated with polyembryonic wasp eggs should also occur in at least some lineages of monoembryonic endoparasitoids. This in fact is seen in several different families of endoparasitic Hymenoptera, which supports the idea that endoparasitism has selectively favored alterations in early development essential for the evolution of polyembryony.

THE FUNCTION AND EVOLUTION OF A LARVAL CASTE SYSTEM

The evolution of specialized castes has occurred most often in species that live in groups comprising closely related individuals and that occupy resource-rich but defensible resources. These conditions exist among well-known eusocial insects such as ants and termites as well as among less studied caste-forming groups such as gall-making thrips and aphids. These conditions also exist among polyembryonic wasps that propagate clonally inside the nutrient-rich but defensible environment of the host. Reproductive larvae are so named because they are the only offspring that molt and develop into adult wasps. However, the insects parasitized by polyembryonic wasps are also commonly parasitized by other parasitoids, resulting in frequent opportunities for interspecific competition to occur. Cruz reported in 1981 that precocious larvae function as soldiers that kill interspecific competitors by piercing their cuticle with their mandibles. This altruistic behavior is advantageous, because precocious larvae increase their own fitness by ensuring the survival of their reproductive siblings. Since selection acts at the level of both the individual and the brood, the ratio of investment in reproductive and precocious larvae would also be predicted to vary depending on how large or small the threat from competitors might be. Such adaptive phenotypic plasticity occurs in *C. floridanum,* in which the proportion of embryos developing into soldiers changes from 4% in hosts that are not attacked by another parasitoid to 24% in hosts that are attacked by a competitor.

A second, more complex function of precocious larvae is that in some species, they kill male siblings and significantly distort the sex ratio of adult wasps that emerge from the host. Like most Hymenoptera, polyembryonic wasps are haplodiploid, which means that male offspring develop from unfertilized eggs and female offspring develop from fertilized eggs. *C. floridanum* female wasps usually lay two eggs per host (one male and one female) and as a consequence most broods are composed of both sexes. However, precocious larvae develop almost exclusively from female eggs such that soldiers are fully related to their sisters but not to their brothers. Moreover, female precocious larvae attack their brothers while still embryos, but not their sisters, resulting in most of the adult wasps (>95%) that emerge from the host being females. These females then mate with the surviving male wasps before dispersing to find new hosts. The tendency to kill brothers but not sisters likely arose as a consequence of genetic conflict between siblings. In effect, the host is a finite resource and precocious larvae make available more resources for their reproductive sisters to consume by killing their brothers. However, a few males almost always survive and these individuals are able to mate with most of their sisters as well as disperse and mate with females from other hosts.

See Also the Following Articles

Development, Hormonal Control of • *Embryogenesis* • *Parasitoids*

Further Reading

Baehrecke, E. H., Aiken, J. M., Dover, J. M., and Strand, M. R. (1993). Ecdysteroid induction of embryonic morphogenesis in a parasitic wasp. *Dev. Biol.* **158,** 275–287.

Cruz, Y. P. (1981). A sterile defender morph in a polyembryonic hymenopterous parasite. *Nature* **294,** 446–447.

Gilbert, S. F. (1994). "Developmental Biology." Sinauer, Sunderland, MA.

Grbic', M. (2000). "Alien" wasps and the evolution of development. *BioEssays* **22,** 920–932.

Grbic', M., Ode, P. J., and Strand, M. R. (1992). Sibling rivalry and brood sex ratios in polyembryonic wasps. *Nature* **360,** 254–256.

Grbic', M., and Strand, M. R. (1998). Shifts in the life history of parasitic wasps correlate with pronounced alterations in early development. *Proc. Natl. Acad. Sci. USA* **95,** 1097–1101.

Harvey, J. A., Corley, L. S., and Strand, M. R. (2000). Competition induces adaptive shifts in caste ratios of a polyembryonic wasp. *Nature* **406,** 183–186.

Ivanova-Kasas, O. M. (1972). Polyembryony in insects. *In* "Developmental Systems, Insects" (S. J. Counce and C. H. Waddington, eds.), Vol. 1. Academic Press, New York.

Nijhout, F. (1994). "Insect Endocrinology." Princeton University Press, Princeton, NJ.

Riddiford, L. M. (1985). Hormone action at the cellular level. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.). Pergamon, Oxford.

St Johnston, D., and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68,** 201–219.

Strand, M. R. (2000). Developmental traits and life-history evolution in parasitoids. *In* "Parasitoid Population Biology" (M. E. Hochberg and A. R. Ives, eds.). Princeton University Press, Princeton, NJ.

Strand, M. R., and Grbic', M. (1997). The development and evolution of polyembryonic insects. *Curr. Top. Dev. Biol.* **35,** 121–158.

Whitfield, J. B. (1998). Phylogeny and evolution of host–parasitoid interactions in Hymenoptera. *Annu. Rev. Entomol.* **43,** 129–151.

Population Ecology

Joseph S. Elkinton
University of Massachusetts

Population ecology deals with questions related to the density or number of individuals of a species in a habitat or location. Insect population ecologists try to understand why population densities of some insects fluctuate dramatically, but others show little variation in density. They want to know why some species are rare, whereas others are common. They seek to understand the mixture of factors affecting mortality and fecundity that together determine the characteristic density of a species and whether the density will decline or increase. Applied population ecologists try to understand the conditions that allow densities of pest insects to reach damaging levels, and they seek to modify those conditions to keep the densities below those levels. To isolate the effects of particular variables from the multitude of factors influencing population growth and survival, population ecologists apply a range of tools to understand population systems, including gathering data on density, fecundity, and mortality from natural populations, construction of mathematical models of population interactions, and experimental manipulation of populations in the field.

ESTIMATING INSECT ABUNDANCE

Research on the population ecology of an insect typically begins with the development of sampling techniques to determine the abundance of the insect in the habitat under study. When the habitat has definable boundaries, we can estimate the number of individuals in the population. More frequently, we estimate density or number of individuals per unit of habitat. Sampling methods vary tremendously with habitat type. Specialized sampling techniques have been developed for insects that live in the soil, in the air, on foliage, and on vertebrate hosts. A large scientific literature exists on the methods developed for each of these habitats.

One can make several general distinctions about different methods of sampling insects. One of these is to distinguish between absolute and relative measures of density. Absolute measures quantify numbers per unit area or volume of habitat. Examples are numbers of grasshoppers per square meter in quadrat samples or numbers of Collembola per cubic centimeter of soil. Relative measures express numbers per sample unit, such as numbers of insects captured in a sweep-net sample or numbers of male moths captured in pheromone-baited traps. Relative measures may or may not reflect absolute density. They are frequently much easier to obtain than absolute measures, but they may be influenced by many factors (e.g., air temperature) that affect the activity of the insect and its likelihood of being captured. Consequently, much research is needed to be sure that relative measures give an accurate indication of differences in absolute population density. In addition, there are population indices, which are indirect measures of density, such as measures of defoliation.

Another general characteristic of sample methods is that some, such as sticky traps or pheromone-baited traps, obtain data continuously over an extended interval of time, whereas others obtain an essentially instantaneous "snapshot" of a population at a particular moment. Examples of the latter include sweep-net samples, quadrat samples, and insecticide knockdown samples. Such instantaneous samples are susceptible to the effects of time of day or weather conditions at the time of sampling, which in turn may influence the number captured. Continuous samples may also be influenced by weather conditions but, because capturing occurs over an extended interval, these effects are averaged over a range of such conditions.

Any investigator planning a sampling program must think carefully about how the samples will be selected. Consultations with statisticians with expertise in experimental design prior to collection of data is always wise, as is preliminary sampling or pilot testing to estimate the expected amount of sample variability. Most sampling schemes may be classified as random, systematic, or stratified random designs. In random sampling, the sample units are chosen or placed at randomly selected locations in the sample area. Random selection is typically done by choosing coordinate points from a table or list of random numbers. In a systematic design, sample units are placed at regular intervals across the sample universe (e.g., every 20 m or every 10th plant). Systematic samples are frequently much easier to conduct, and they ensure that the samples are distributed evenly across the sample universe. However, systematic samples violate the requirement that samples be selected independently and at random from the sample universe, imposed by most statistical analyses. Whether violation of this requirement leads to erroneous conclusions in any particular system is usually debatable. A reasonable compromise between the two approaches is a stratified random or randomized block design in which the sample universe is divided into regularly spaced subunits or blocks and samples are selected at random from each block. Differences in density between subunits caused by edge effects or density gradients across the field can be detected with appropriate statistical analysis.

Mark–recapture techniques are sometimes used to estimate insect abundance. Individuals are marked with a permanent mark that does not significantly alter the behavior of the insect or affect its survival. Marked individuals are counted and released into a population. Subsequently, a sample of individuals is collected from the population and the proportion *(m/n)* of marked individuals in the sample is recorded. Nearly all mark–recapture methods rest on the basic assumption that *m/n* of the sample is an estimate of the proportion of marked individuals in the larger population *M/N.* If one assumes that *M,* the number of marked individuals in the population, is equal to the number released, then the population size *N* can be calculated directly: $N = M(n/m)$. This calculation is known as the Lincoln index or Peterson estimate. As a rule, it cannot be assumed that *M* is equal to the number of marked individuals released because some of them will have died or emigrated in the interval between release and recapture. Several fairly elaborate techniques involving multiple rounds of mark and recapture have been developed to estimate *M* when it is unknown.

Applied ecologists frequently use sequential sampling to classify populations into density categories such as high (requiring treatment) or low (requiring no treatment). The basic idea is to minimize the cost of the sampling effort by collecting only the number of samples needed to make this decision with sufficient statistical reliability. Development of a sequential sampling scheme for a particular insect and crop or habitat requires research to define the population density that forms the dividing line between the high and low categories, a definition of the degree of allowable risk for making an incorrect classification, and an estimate of the degree of clumping or patchiness that typifies the species.

POPULATION GROWTH AND LIMITS TO GROWTH

A fundamental property of the population dynamics of all species is that the number or density of individuals will grow at an ever-increasing rate when conditions are favorable. The simplest example of such growth is illustrated by the replication of single-celled organisms. If a bacterium divides every hour, a colony that began with one individual would grow to 2, 4, 8, 16, …, 2^t, where *t* is the number of hours or replications. With insects, the rate of replication with each generation is potentially much faster. The house fly, *Musca domestica,* for example, has generation time of approximately 2 weeks and lays more than 200 eggs. If half those eggs were females and all survived to maturity, the population would increase by more than a hundred fold each generation. By the end of one year the population would have increased to more than 10^{52} individuals (a mass of flies much larger than the earth). Even if mortality were much higher, (e.g., only 1 fly in 10 survived to maturity) the end result would be similar but would take longer to achieve. Mathematically, we refer to this process as geometric or exponential growth, and for

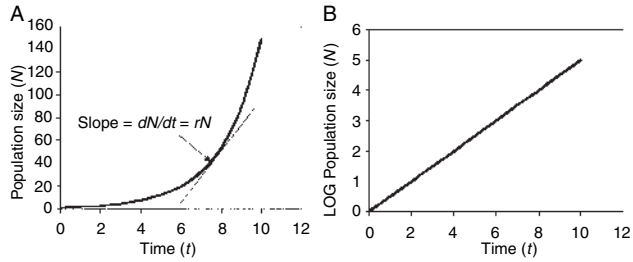

FIGURE 1 (A) Exponential growth of a population yields a straight line when density *(N)* is plotted versus time *(t),* as in (B), on a logarithmic scale.

insects that reproduce continuously, it is typically expressed with the following equation (Fig. 1):

$$\frac{dN}{dt} = rN \qquad \text{or} \qquad \frac{dN}{dt}\frac{1}{N} = r$$

Here *N* is the population size or density, *dN/dt* is the growth rate (the change in density per unit time), and the constant *r* is the instantaneous per capita rate of increase. The parameter *r* equals instantaneous per-capita birthrate minus the instantaneous per-capita death rate. When the rates of birth and death are equal, $r = 0$ and the population ceases to grow. When the death rate exceeds the birthrate, *r* is negative and the population declines.

The notion of exponential population growth was first expressed in 1798 by T. R. Malthus, who observed that human populations increased exponentially, whereas food production increased in a linear fashion. As a result, Malthus predicted that mankind would be doomed forever to a life close to the edge of starvation. Malthus did not foresee the large increase in agricultural production that accompanied the Industrial Revolution, but he may yet prove to be correct. The ideas of Malthus were fundamental to the development of Darwin's theory of evolution: all organisms must struggle to survive because all species produce more offspring than can survive.

It is obvious that no population can continue growing forever; sooner or later, it will reach a density above which individuals can no longer obtain the resources they need to survive. This density is known as the carrying capacity of the environment. For different species in different habitats, the carrying capacity will be determined by competition for particular resources. For desert plants, water is typically the limiting resource. For many animals, food supply determines the carrying capacity. As a population expands toward the carrying capacity, the rate of growth slows down. This process is typically represented by the following logistic equation (Fig. 2), which was first applied to population growth by P. F Verhulst in 1838 and independently by R. Pearl and L. J. Reed in 1920:

$$\frac{dN}{dt} = rN - \frac{rN^2}{K}$$

Here *K* is the carrying capacity and *r* and *N* retain their earlier definitions. The first term *(rN)* represents exponential

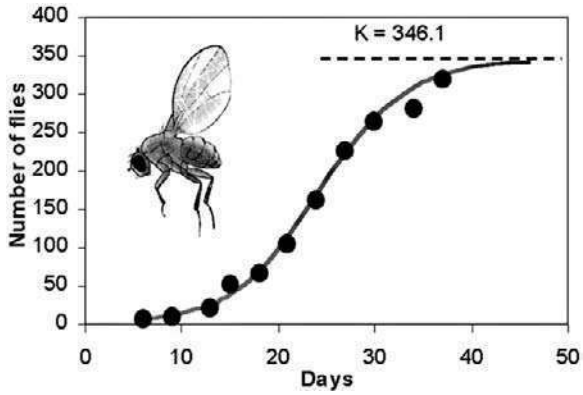

FIGURE 2 Logistic population growth model fit to data from laboratory colony of fruit flies. [Adapted from Krebs (2001) and Pearl, R. (1925). "Biology of Population Growth." Alfred A. Knopf, New York.]

growth. The effect of the second term (rN^2/K), often called environmental resistance, increases as N becomes large. As N approaches K, the rate of growth (dN/dt) approaches zero.

There are a number of assumptions inherent in the use of the logistic equation to represent population growth. The first is that unless disturbed, population density will approach the carrying capacity (K) and then remain stable about it. In actuality, most populations fluctuate in density, even populations kept in the laboratory under constant environmental conditions. Another assumption is that the shape of the curve is symmetrical above and below the midpoint. In fact, few population systems, even in the laboratory, follow the trajectory predicted by the logistic equation. Rather, the importance of the equation is its contribution to theoretical ecology. It captures the most basic processes of population dynamics: exponential growth and the effects of factors that limit growth. Variations of the logistic equation have been explored by many individuals; indeed, it is the foundation of a large body of work in theoretical population ecology. A. J. Lotka in 1925 and V. Volterra in 1926 extended the logistic to describe both competition between species and predator–prey interactions. More recent applications include models of food webs and interactions between many species in a community.

REGULATION OF POPULATION DENSITY

Although no one doubts that competition for resources confers an upper limit on the growth of all populations, it seems clear that many populations of insects persist at densities far below any obvious carrying capacity. Furthermore, the densities of most species fluctuate within a fairly narrow range of values. For a population to remain at constant density, the birthrate must equal death rate. Each individual must on average replace itself with one surviving offspring. Indeed, for any species to persist over evolutionary time, the average birthrate must equal the average death rate, although these quantities may vary considerably from year to year. Organisms that experience high mortality compensate by producing lots of young. For this fundamental reason, most population ecologists believe that most populations are stabilized by factors that are called density dependent. Such factors influence birthrates or death rates in a way that varies systematically with density, such that populations converge to densities at which rates of birth and death are equal and the density is at equilibrium. Such factors act as a negative feedback system that is analogous to the regulation of room temperature by a thermostat. If densities rise above the equilibrium value, the death rate exceeds the birthrate and the population returns to equilibrium. If densities fall below the equilibrium value, the birthrate exceeds the death rate and the population increases.

Predators, pathogens, and parasitoids usually cause mortality to insects that is density dependent. The proportion or percentage of the population killed by these factors varies systematically with density. An increase in the proportion dying with increasing density is called positive density dependence; a decrease in the proportion dying with increasing density is termed negative or inverse density dependence. A mortality factor is density independent when the proportion killed varies in ways that are unrelated to population density. Many abiotic factors, such as mortality due to subfreezing temperatures, act in ways that are density independent. Although many insect population ecologists focus on sources of mortality, density-dependent changes in fecundity may either lead to changes in density or serve to stabilize densities. Certainly, competition for resources is a density-dependent process that will stabilize a population at the carrying capacity, if other factors do not intervene at lower densities.

For more than 50 years ecologists have debated whether population densities of most species are stabilized by such density-dependent factors. L. O. Howard and W. E. Fiske in 1911 were the first to articulate the idea that populations cannot long persist unless they contain at least one density-dependent factor that causes the average fecundity to balance the average mortality. Other early proponents of this idea in the 1950s were A. J. Nicholson and D. Lack. In contrast, H. G. Andrewartha and L. C. Birch argued in 1954 that most populations are not held at equilibrium density. Rather, densities merely fluctuate. In their view, most species avoid extinction because they comprise what we now call metapopulations, that is, a series of subpopulations that are linked by dispersal but whose densities fluctuate independently of one another. Extinction of subpopulations occurs quite frequently, but these are recolonized by individuals dispersing from other subpopulations, thereby allowing the species to persist indefinitely over the entire region. In recent years metapopulation dynamics has been explored by way of computer simulations, and these have revealed that in the absence of density dependence, such systems eventually go extinct.

The debate about the ubiquity of density-dependent processes has persisted despite the efforts of various ecologists to terminate the discussion either because it was bankrupt or because they deemed that prevalence of density dependence

was too obvious to deny. Many ecologists have insisted that no conclusion can be made regarding the existence of density dependence in a population system unless its action can be demonstrated statistically in data collected from the populations. This has proved difficult to achieve. Until recently, adequate methods for detecting density dependence in population systems have been lacking, and earlier methods are now known to be statistically flawed. Several new methods were proposed in the decade following 1985, most of which involved a variety of computer-based resampling procedures, such as the bootstrap.

LIFE TABLES

Population ecologists typically summarize their collected mortality data in life tables, which express the proportion of the insects that survive to successive instars or life stages. The proportion dying in each instar or life stage can be partitioned into components caused by different agents of mortality. Techniques for quantifying mortality in this way were pioneered in the 1960s by R. F. Morris, who worked on spruce budworm, *Choristoneura fumiferana,* in Canada, and by G. C Varley and G. R. Gradwell in England, who studied the winter moth *Operophtera brumata,* a defoliator of oak trees in Europe. In 1960 Varley and Gradwell reported a 13-year study of winter moth survival to successive life stages at one site. They showed that predation on pupae was density dependent and was responsible for maintaining a low-density equilibrium. Their studies illustrated the benefits of collecting data on density, fecundity, and mortality from the same populations over many generations. They developed a graphical procedure known as key-factor analysis to identify the life stages or causes of mortality most responsible for observed fluctuations in density. In the winter moth system, the key factor was overwintering mortality. Subsequent work offered analytic procedures for key-factor analysis, but others, such as E. Kuno and T. Royama, noted statistical shortcomings; as a result this procedure is rarely used.

Mortality from parasitoids and disease is typically measured by collecting a sample of hosts and dissecting them, or rearing them to determine the proportion that is parasitized or infected. This proportion, however, may or may not accurately reflect the total proportion of hosts attacked over a life stage or generation, since attack rates vary over time with the number of hosts available to attack and the number of adult parasitoids searching. In 1992 J. T. S. Bellows and colleagues reviewed the techniques that have been developed to convert proportion infected in samples to the summary values appropriate for life tables or for comparisons of parasitism between populations or treatments in a study.

PREDATORS, PATHOGENS, AND PARASITOIDS

Nearly all insect species are attacked by a suite of natural enemies and, as indicated earlier, these frequently act in a density-dependent manner that serves to maintain populations at densities well below the carrying capacity determined by resource limitation. Different natural enemies will attack different life stages. Predators are mobile organisms that feed on many prey. Predators of insects include many vertebrate and invertebrate species. Parasitoids are insects whose immature stages develop on or within a single host individual, usually killing it in the process. The majority of parasitoids belong to the order Hymenoptera or Diptera. Pathogens are microorganisms including viruses, bacteria, microsporidia, or fungi that infect a host and cause disease. Biological control focuses on the use of natural enemies to maintain the densities of pest insects at levels below that at which they cause damage.

Density-dependent predation or parasitism may arise from two different sources: the numerical response and the functional response. The numerical response is an increase in the density or number of predators or parasitoids in response to increasing prey density. The numerical response can arise from increased reproduction or survival of predator or parasitoid offspring induced by increases in prey availability, or it can arise from an aggregative response whereby predators and parasitoids are attracted to sites with high densities of prey.

The functional response is an increase in the number of prey taken per predator or parasitoid at increasing prey density. Important contributions to the understanding of the functional response were made by C. S. Holling beginning in 1959. In laboratory experiments, Holling presented individual predators with different numbers of prey. He showed that the number of prey consumed over a specified time interval increased with number of prey available, but at a decreasing rate towards an upper maximum (Fig. 3A). This effect is caused by an upper limit in the predator's capacity for consumption and by the increasing proportion of time

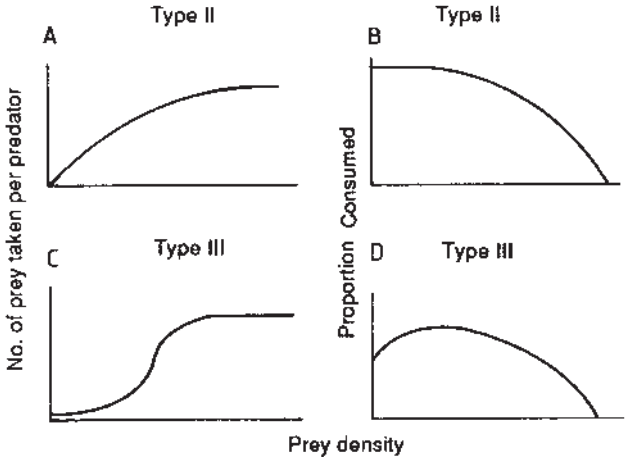

FIGURE 3 Number of prey consumed per predator and the corresponding proportion consumed for type II (A, B) and type III (C, D) functional responses. [Adapted from Holling, C. S. (1965). *Mem. Entomo. Soc. Can.* **45,** 3–60.]

devoted to handling the large number of prey at the expense of time spent searching for prey. Above this limit, further increases in prey density will not cause higher consumption. The proportion of prey consumed plotted versus prey density steadily declines (Fig. 3B), illustrating that the functional response is inherently inverse density dependent. Without a numerical response, predators and parasitoids are unlikely to stabilize a host population. Further work by Holling showed that under some important conditions, the functional response can lead to positive density-dependent predation. Whenever increases in prey density result in some change in the foraging behavior of the predator or parasitoid, such that foraging becomes more efficient or efforts are concentrated on the particular prey species, the number taken will accelerate with increasing host density (Fig. 3C), and the proportion taken will increase (Fig. 3D) over the lowest range of prey densities. Holling termed this a type III functional response in contrast to type II, which is the continuous decline in proportion taken evident whenever there is no change in foraging behavior in response to changes in prey density (Fig. 3A,B). Holling demonstrated a type III response for shrews foraging for sawfly pupae. He envisioned type III responses to be characteristic of vertebrate predators, which have a relatively high capacity for learning and behavioral change. However, the type III functional response has subsequently been demonstrated in many insect predators and parasitoids.

Specialist or monophagous natural enemies are those that attack a single host species. Oligophagous natural enemies restrict their attacks to a closely related group of species. Generalist or polyphagous natural enemies attack a wide range of host species. The distinction is important because generalists and specialists typically respond very differently to changes in host density. Specialists are most likely to exhibit a numerical response to changes in density of their prey because they depend on no other food sources and their seasonal development is closely linked with that of their prey. Generalists may exhibit little or no numerical response, because they depend on many types of prey and may shift from one prey to another, depending on which species are available. In fact, it is very common for many natural enemies, especially generalists, to exhibit inverse density dependence, wherein mortality declines as prey density increases. Such mortality cannot stabilize prey densities unless accompanied by a numerical response.

Many systems exhibit complex density dependence; for example, mortality from particular natural enemies may switch from positive to negative as host density increases. Thus, bird predation on forest-dwelling caterpillars that is density dependent at the lowest density may shift to inverse density dependence as the densities exceed the capacities of the predators to respond numerically and the functional response approaches the upper limit of prey consumption. Under such conditions, the prey densities may "escape" into an outbreak phase, which is characteristic of a few species. Outbreak populations are typically subject to a different suite of density-dependent factors, such as viral diseases and starvation, which become major sources of mortality only when densities are high. These factors can maintain populations at a high-density equilibrium, but more frequently they cause the collapse of populations back down to a low-density, endemic phase. In contrast, generalist predators, which might consume the majority of prey individuals at low density, are likely to consume a tiny fraction of the population at high density even though they are attacking the same or higher numbers of prey individuals at these high densities.

In 1976 T. R. E. Southwood and H. N. Comins proposed a "synoptic model" as a general feature of insects that occasionally go into an outbreak phase (Fig. 4). Earlier expressions of this idea can be found in the writings of R. F. Morris and R. M. Campbell. The model is depicted by plotting R_0, the net reproductive rate, against density. At intermediate densities in the "natural enemy ravine," density-dependent mortality caused by natural enemies maintains the population at equilibrium (R_0=1). The natural enemy ravine separates two "ridges," one at high and one at low density, where mortality is lower and population densities increase. At very high density, other mortality factors such as starvation and disease cause the populations to collapse. At the extremes of low density, an "Allé effect" comes into play, caused by the failure of individuals to find mates and reproduce. Populations in this range decline inexorably to extinction. Such low densities are infrequent in most natural populations.

POPULATION MODELS

Theoretical ecologists have developed numerous population models to study the effects of natural enemies on their prey. Much insight has been gained from simple mathematical expressions that relate prey density to that of changes in density of specialist predators or parasitoids. One approach, pioneered by A. J. Lotka and V. Volterra in the 1920s, entailed a simple modification to the logistic equation by

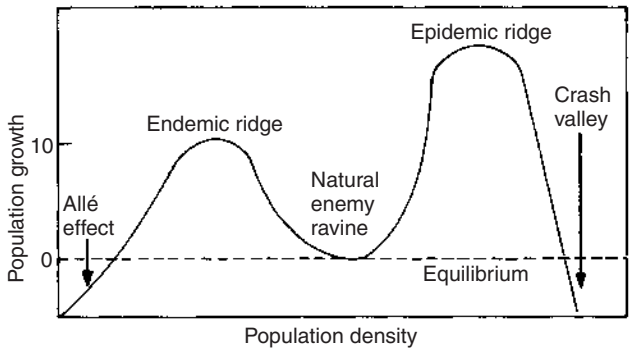

FIGURE 4 Synoptic model of complex density dependence. [Adapted from Southwood, T. R. E., and Comins, H. N. (1976). *J. Anim. Ecol.* **45,** 949–965, British Ecological Society.]

adding a term that represents prey consumption on densities of both host and predator:

$$\frac{dN}{dt} = r_1N - k_1PN$$

and

$$\frac{dP}{dt} = -r_2P + k_2PN$$

where N and P are the respective densities of host and predator, and the rates of population growth are given by dN/dt and dP/dt. In the first equation, the first term represents exponential growth of the host (r_1N) in the absence of the predator, whereas in the second equation, the first term represents exponential decline in the predator ($-r_2P$) in the absence of the host. The second term in each equation represents the effects of predation, which is determined by the encounter rate of host and predator and is proportional to PN. The constant k_1 is a measure of the ability of prey to escape predators, and k_2 is a measure of the ability of predators to capture the prey. The model predicts a predator–prey oscillation (Fig. 5A). The changes in density of the predator or parasitoid lag behind those of its host. The highest rates of attack on the host occur at peak predator density, which is observed after the host population density has declined. Plotting percent mortality from the predator or parasitoid against host density would not reveal a consistent pattern of positive density dependence, even though the predator or parasitoid is clearly maintaining the host within a narrow range of density. This type of response is termed delayed density dependence. In the 1990s, P. Turchin and others developed new statistical methods based on time-series analysis for detecting delayed density dependence in annual census data. Laboratory studies of predator–prey interaction frequently show such predator–prey oscillations (Fig. 5B).

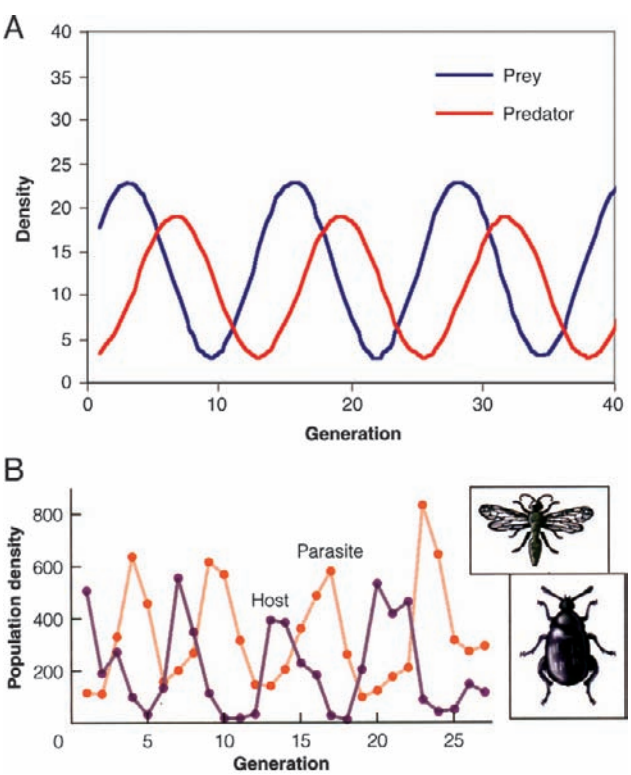

FIGURE 5 (A) Predator–prey oscillation as predicted by a Lotka–Volterra model and (B) Host–parasitoid oscillation of the Azoki bean weevil in laboratory culture. [From Krebs (2001), adapted from Utida, S. (1957). *Ecology* **38,** 442–449.]

A different class of models appropriate for parasitoids and host populations with discrete generations was initiated by W. R. Thompson in 1924 and by A. J. Nicholson and V. A. Bailey in 1935. These models were difference equations, in contrast to the differential equations of the Lotka–Volterra type. The general form of the model expresses host or prey density N in generation $t + 1$ as follows:

$$N_{t+1} = \lambda N_t f(N_t P_t)$$

Here λ is the rate of increase per generation of the host in the absence of parasitism and $f(N_t,P_t)$ is the proportion of hosts surviving parasitism in the preceding generation (t). Similarly, the number of parasitoids in the next generation P_{t+1} is given by

$$P_{t+1} = cN_t[1 - f(N_t,P_t)]$$

where $(1 - f(N_t,P_t)$ is the proportion of hosts attacked by parasitoids in generation t and c is the number of surviving parasite progeny produced per parasitized host. The notation $f(N_t,P_t)$ stands for any function of N_t and P_t. Variations in the model involve incorporating different factors into $f(N_t,P_t)$. The simplest version for $f(N_t,P_t)$, proposed by Nicholson and Bailey, assumes that all hosts are equally likely to be attacked and that parasitoids search at "random," such that the proportion of hosts that escape is given by the zero term of the Poisson distribution. This model predicts that hosts and parasitoids will experience density oscillations of ever-increasing amplitude until both go extinct. Obviously, this does not occur routinely in nature.

In the 1960s, M. P. Hassell, R. M. May, and colleagues began an exploration, extending over several decades, of the various factors that would stabilize host–parasitoid interactions in models of this type. These factors included mutual interference of parasitoids, patchiness of hosts or parasitoid attacks, and variation in host susceptibility. For example, Fig. 6 illustrates how an increase in the magnitude of mutual interference between parasitoids will reduce parasitoid efficiency at high host density and thus stabilize the system. When mutual interference is low (Fig. 6A,B), or absent altogether ($m = 0$, as in the original Nicholson–Bailey version), the model oscillates with increasing amplitude and both host and parasitoid go extinct. Other factors such as patchiness in parasite attacks have a similar effect.

The models described so far are relatively simple. They contain a small number of parameters or variables and leave out much of the known biology of the host and its natural

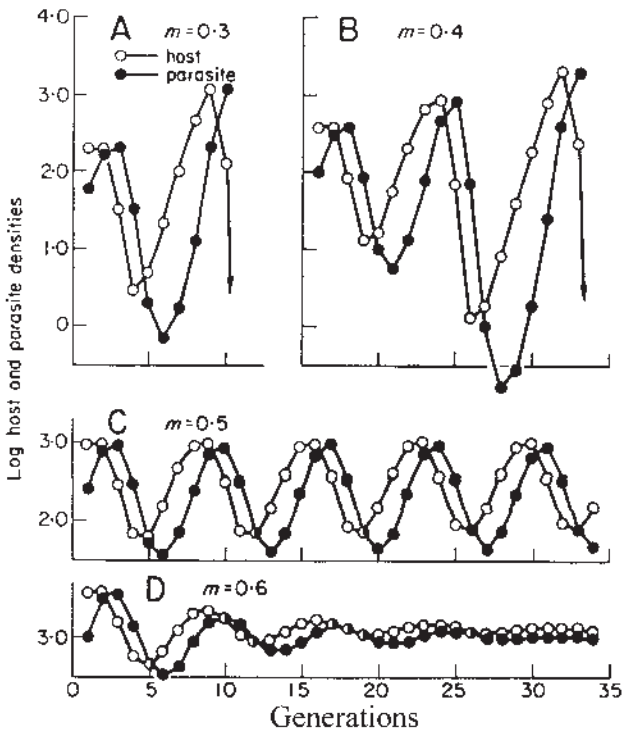

FIGURE 6 Stabilizing effects of mutual interference *(m)* between parasitoids on models of the Nicholson–Bailey type. [From Varley *et al.* (1973); adapted by permission from Hassell, M. P., and Varley, G. C. (1969). *Nature* **223,** 1133–1137, copyright 1969 Macmillan Publishers Ltd.]

enemies. Theoretical ecologists focus on such models because they can be analyzed by a variety of mathematical tools and can be used to address questions of general ecological significance. They hope that the models capture the essential features of the systems they represent. Applied ecologists, in contrast, are often drawn to more complex models, because they wish to understand the complex interplay of environmental and biotic variables that account for the density fluctuations of particular species of interest. With modern computers, there is virtually no limit to the complexity that can be built into such models, but this does not mean that the resulting simulations will necessarily be useful or revealing. Many highly complex models constructed in the 1970s, when high-speed computers first became widely available, were abandoned because they failed to accurately predict the behavior of the systems they represented and were too complicated to understand. Many ecologists have advocated models of intermediate complexity.

COMPETITION AND TROPHIC INTERACTIONS

When resources are limiting, intraspecific competition prevents growth of a population beyond the carrying capacity of the habitat. Interspecific competition occurs when two or more species compete for the same resource. Ecologists have long appreciated the importance of interspecific competition in shaping community structure and the interaction between species. Models of competition developed in the 1920s by Lotka and Volterra, analogous to those for predation, predicted that when two species compete for the same resource, under most conditions one species will win out and the other will go extinct. This phenomenon, which became known as competitive exclusion, was readily demonstrated in laboratory populations by G. F. Gause in the 1930s with microorganisms, and by T. Park and associates in the 1950s with flour beetles (*Tribolium* spp.). Coexistence occurs only when species do not overlap completely in their use of resources. To put it another way, two species cannot coexist if they occupy the same niche. Clear examples of competitive exclusion of insects in the field are harder to come by. Most examples involve introductions of insects to regions outside their native range. For example, the Argentine ant *Linepithema humile* has been introduced to many regions of the world and has outcompeted and excluded many native ant species. In California around 1900, biological control introductions of *Aphytis* parasitoids of California red scale, *Aonidiella aurantii,* resulted in the establishment on citrus of *Aphytis chrysomphali.* After further introductions in the 1940s and 1950s, this species was displaced by *A. lingnanensis,* which in turn was displaced by *A. melinus* in the hot interior, but not the coastal, regions of southern California.

In many other systems, however, different insect species coexist, even though they appear to use the same resource in the same way. For example, in 1981 J. H. Lawton and D. R. Strong examined the various coexisting herbivores of bracken fern and found no evidence for competition. A famous experiment in the rocky intertidal habitat reported by R. T. Paine in 1974 suggested a general explanation. The intertidal community consists of various species of filter feeders (e.g., barnacles) and other invertebrates that coexisted, even though they competed intensely for space on the rock surfaces. Paine removed starfish, the top predator in this system, and as a result, one species (a mussel) outcompeted and excluded all the other invertebrates and took over the site. Paine termed the starfish a keystone predator, meaning that it had dominant effect on the number of species present in the entire community. The implication was that such predators in many communities prevent competitive exclusion among their prey by maintaining densities below the level at which competition would cause one species to predominate.

Population and community ecologists frequently portray the interaction of organisms in a community in terms of trophic levels or food webs. At the bottom trophic level are the primary producers, the green plants that use photosynthesis to sequester energy from the sun. The next trophic level is occupied by herbivores that feed on the plants. Carnivores that feed on the herbivores or on other carnivores occupy higher trophic levels. A typical food web involving insects on collard plants is illustrated in Fig. 7.

An influential paper by N. G. Hairston and colleagues in 1960 proclaimed that the importance of competition for resources, as opposed to regulation by natural enemies, varies with trophic level. These authors proposed that densities of primary producers are typically governed by competition for

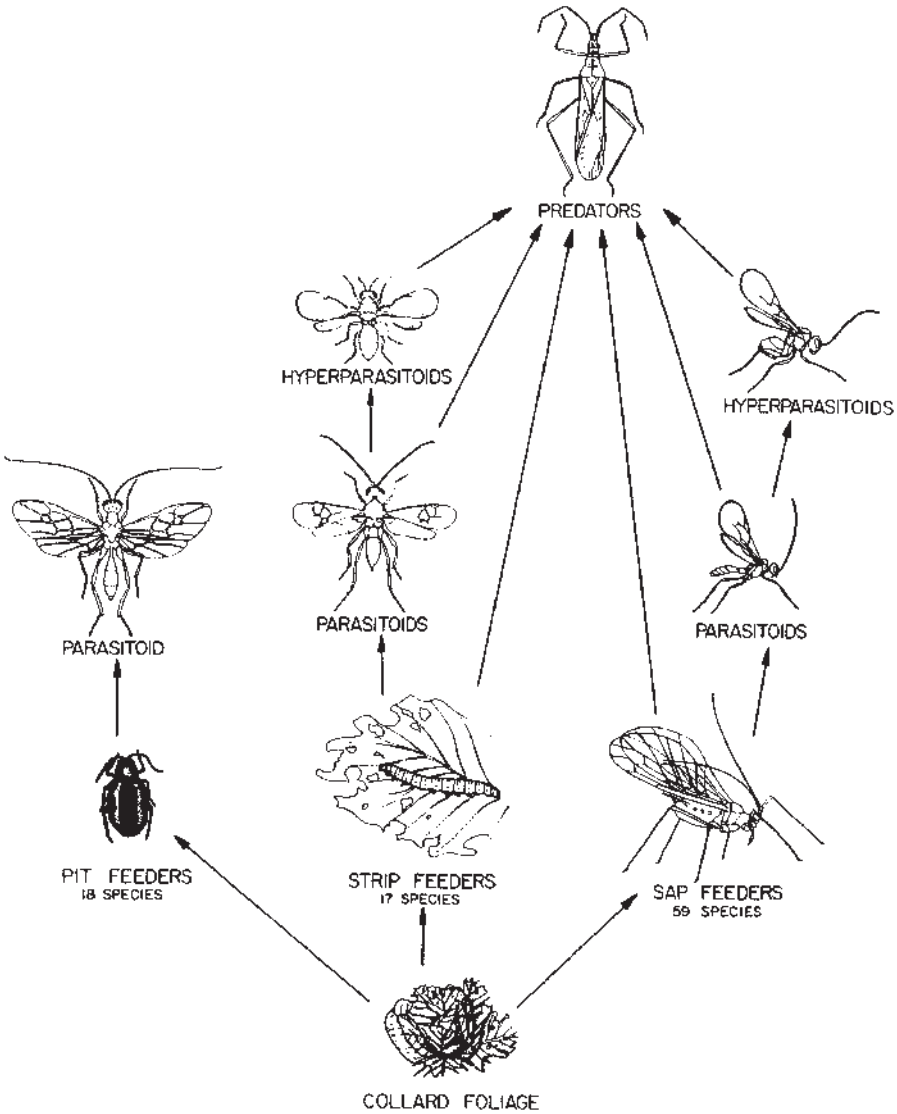

FIGURE 7 Schematic of arthropod food web in collards. [From Price, P. W. (1984). *In* "Ecological Entomology" (Huffaker, C.B., and Rabb, R. L., eds.), pp. 20–50. Wiley Interscience, NY; adapted from Root, R. B. (1973). *Ecological Monographs* **43,** 95–124.]

light, nutrients, or water. Herbivores, on the other hand, are typically regulated by their natural enemies at densities well below carrying capacity, so that competition for resources is unimportant. In contrast, carnivores typically limit the densities of their prey and thus are regulated by competition for resources. Regulation of population density by natural enemies has been termed "top-down" control. Regulation by way of competition for resources or by other interactions between an animal and its food supply has been termed "bottom-up" control.

The generalizations of Hairston and colleagues were widely criticized and many obvious exceptions exist, but variations on this idea have persisted. More recent versions have included effects of several levels of carnivores and other factors such as community productivity or environmental stress. Experimental manipulations, particularly in aquatic systems, have provided support for the general idea in the form of trophic cascades. For example, in 1990 M. E. Power reported the experimental exclusion of top carnivores (large fish) from a food web in a California river. The results were a 10-fold increase in smaller fish and predatory insects that in turn caused an 80% decrease in herbivores (larval chironomid midges) and a 3- to 120-fold increase in the primary producers (algae). These results were the opposite of those predicted by Hairston and coworkers because there were four trophic levels, instead of three. Omnivores that feed simultaneously on several trophic levels complicate our understanding of most natural food webs. Research in several agricultural crops shows that intraguild predation can cause counterintuitive effects of predators on herbivorous insects. For example, in 2001 W. E. Snyder and A. R. Ives showed that carabid predators in alfalfa feed on pea aphids. Since, however, they feed even more heavily on the

aphid "mummies" that contain parasitoids of pea aphids, their net impact is to increase aphid densities.

The relative importance of top-down versus bottom-up regulation of herbivores has been debated for several decades. A review of population studies by D. R. Strong and colleagues in 1984 concurred with Hairston and colleagues that competition among most insect herbivores was rare, because herbivore densities were typically kept far below carrying capacity by natural enemies. More recent reviews (e.g., by R. Denno *et al.* in 1995) have challenged this view. One issue is that many plants have sophisticated chemical defenses and many plant parts, such as foliage, are nutrient poor. Many herbivores consume specialized plant parts that are either nutrient rich (e.g., seeds) or are poorly defended (e.g., new leaves). Competition for these resources may be intense. Furthermore, plants influence both the survival and fecundity of herbivores in many subtle ways. For example, all herbivores disperse and must locate suitable host plants; failure to find host plants may be a dominant source of mortality and may vary from one habitat to another. Spring-feeding foliage feeders must synchronize their emergence with that of host foliage; failure to synchronize may cause mortality or may reduce fecundity because of consumption of inadequate or poor-quality foliage. Such factors are rarely documented in life table studies. For most insect herbivores, a mixture of top-down and bottom-up forces determines density.

TEMPORAL PATTERNS OF FLUCTUATION

Density fluctuations, whether they are large or small, may occur at erratic or at regular intervals. Simple population models of predators and prey or laboratory studies, as already discussed, often exhibit regular oscillations. Evidence, however, for regular oscillations in nature is rare. The density of most species fluctuates erratically (Fig 8). This is not surprising because the density of most species is influenced by weather conditions in a multitude of ways. Except for obvious regularity of seasonal variation, weather conditions vary from week to week and year to year in ways that are mostly erratic. Furthermore, most species are influenced by a large suite of natural enemies. Predator–prey oscillations, such as those that occur in simple theoretical models or simple laboratory systems, become erratic and irregular when many natural enemies are involved.

Although the density fluctuations of most species are erratic, there are a few that exhibit regular cycles of abundance, similar to what is observed in laboratory populations or in theoretical models. For many others it is not obvious whether the fluctuations are or are not regular. For example, which of the insects in Fig. 8 have a regular component in their fluctuations, and what is the cycle period? Time-series analysis provides the statistical tool to answer these questions. Perhaps the most famous example among insects of regular cycles is that of the larch budmoth, *Zeiraphera diniana,* an insect that periodically defoliates larch forests in the European Alps (Fig. 9). The defoliator reaches peak density at 8- to 9-year intervals and seems to increase or decrease exponentially in the intervals between (linear on a log scale, as in Fig 1B). The changes in budmoth density were associated with increases in both parasitism and mortality from a virus disease in what looked like a classic case of delayed density dependence. In 1981 R. M. Anderson and R. M. May proposed a host–pathogen model for this system that has served as a template for similar models of other insect host–pathogen systems. Anderson and May believed that the pathogen alone could account for the regular cycles. Other

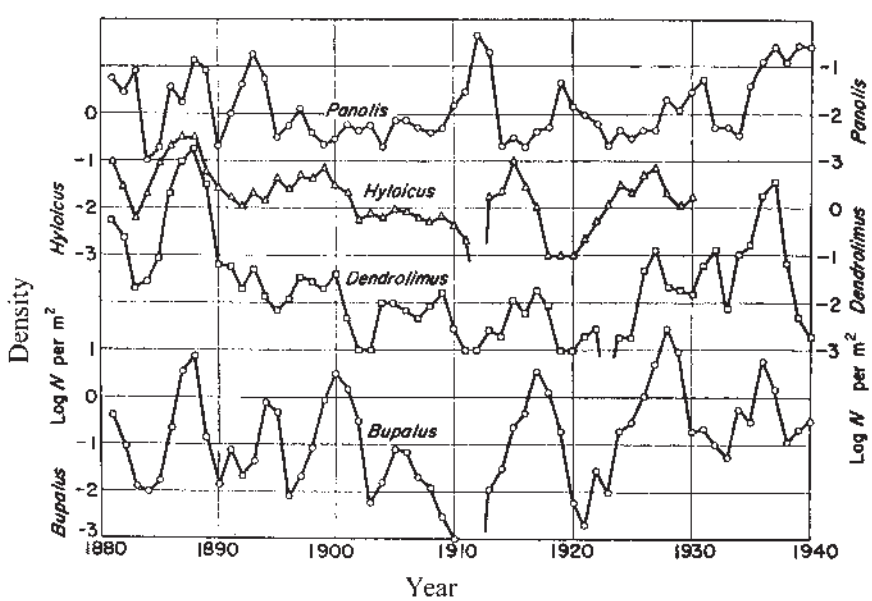

FIGURE 8 Sixty years of density fluctuations (plotted on a log scale) of four forest Lepidoptera in Europe. [From Varley *et al.* (1973); adapted from Schwerdtfeger, F. (1935). *Z Forst-u. Jagdw.* **67,** 15–38, 85–104, 449–482, 513–540 and Schwerdtfeger, F. (1941). *Z. Angew. Ent.* **28,** 254–303.]

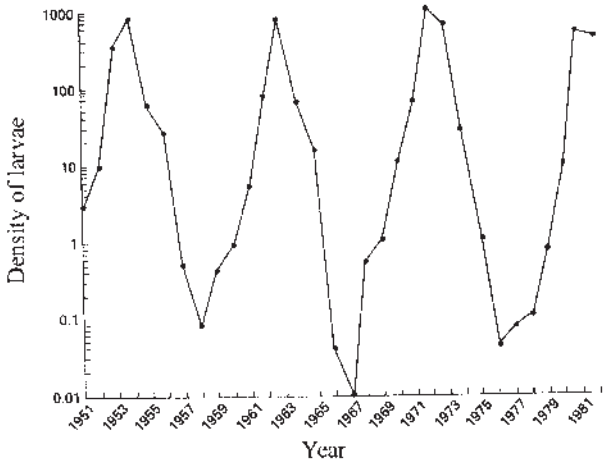

FIGURE 9 Regular cycles of larch budmoth density in the European Alps. [From Speight *et al.* (1999); adapted from Baltensweiler (1964). *Canad. Entomol.* **96,** 792–800.]

data suggested that effects of defoliation on host plant quality, as well as genetic variation in the budmoth itself, may also account for the density oscillations. This famous example illustrates the difficulty in distinguishing cause from effect in the confluence of bottom-up and top-down forces that account for cyclic behavior in some population systems.

Prior to the 1970s, ecologists assumed that erratic fluctuation evident in most populations was caused by erratic weather conditions or other sources of randomness. They assumed that the density fluctuations embodied in mathematical models or in simple laboratory populations would have simple dynamics. Densities of such systems would either remain at equilibrium or exhibit regular oscillations about the equilibrium value, as in all the examples given earlier (Figs. 5, 6). The pioneering work on deterministic chaos by R. M. May in 1974 taught us otherwise. May studied the behavior of versions of the logistic equation appropriate for organisms with discrete or nonoverlapping generations (Fig. 10). There is a family of such models, of which the simplest is

$$X_{t+1} = aX_t - aX_t^2$$

where a is reproductive rate analogous to r of the continuous logistic, and X_t is density in generation t as a proportion of the carrying capacity K (i.e., $X_t = N_t/K$). When the reproductive rate was low (Fig. 10A), the population remains at equilibrium. At higher values of a, the system alternates between high and low values, a pattern known as a 2-point cycle (Fig. 10B). As a increases further, the system shifts to 4-point (Fig 10C), and then 8-point, …, 2^n cycles. Above a = 3.57 however, an entirely new behavior, the chaotic regime, prevails, and densities fluctuate erratically without ever repeating themselves (Fig. 10D). These results had profound implications. They suggested that the erratic behavior characteristic of most natural populations (e.g., Fig. 8) might be due to the inherent mathematical properties of the interaction with natural enemies and other density-dependent effects, instead of resulting from random forces such as weather.

There ensued an effort to determine whether natural populations were indeed chaotic. Early studies concluded that most populations were not chaotic. These were based on attempts to fit natural populations to simple models and then to see whether the values of model parameters representing, for example, density dependence, time delays, or reproductive rate would elicit the expectation of chaotic behavior. The problem was that the conclusion depended on the particular

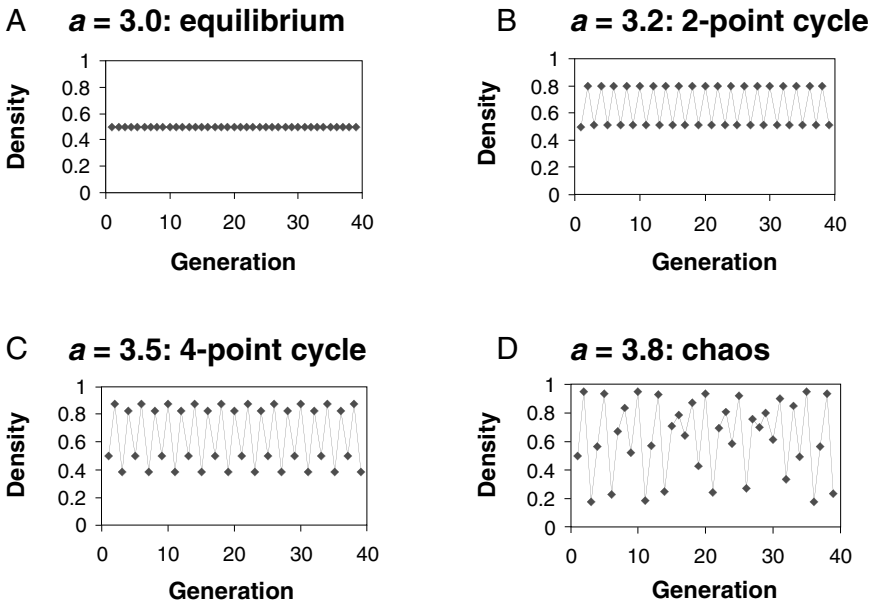

FIGURE 10 Density of a hypothetical population plotted versus time for a version of the logistic equation appropriate for insects with discrete generations under different values of the reproductive rate a illustrating (A) a steady state, (B) a 2-point cycle, (C) a 4-point cycle, and (D) chaos. [Adapted from May, R. M. (1976). *Nature* **261,** 459–467.]

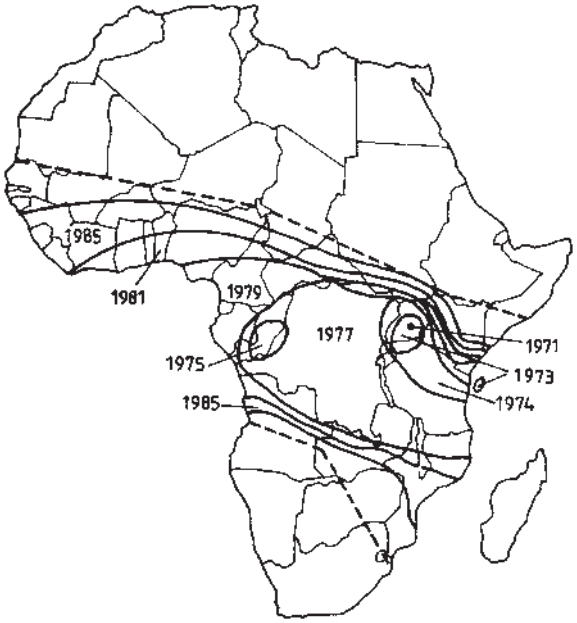

FIGURE 11 Spread of the cassava green mite from its point of introduction near Kampala in 1971. [Reproduced from Yaninek, J. S., de Moraes, G. J., and Markham, R. H. (1989). "Handbook on the Cassava Green Mite *(Mononychellus tanajoa)* in Africa: A Guide to Its Biology and Procedures for Implementing Classical Biological Control." International Institute of Tropical Agriculture, London.]

model used, which was always a simplistic abstraction of the inevitably complicated dynamics of real populations. Subsequent investigators offered techniques that were more general and free of assuming particular population models. Applications of these techniques have indicated that some population systems are chaotic, but most are not.

SPATIAL PROCESSES

Species that are introduced into new habitats will spread from the point of introduction to surrounding areas in a pattern that frequently resembles concentric rings (Fig. 11). Theoretical population biologists, beginning with J. G. Skellam in 1951, have developed models of population spread that couple processes representing population growth with that of dispersal. Dispersal is typically modeled as a random movement process that is analogous to molecular diffusion. These models typically predict that the leading edge of the infestation forms a traveling wave that moves across the landscape at a constant speed in any given direction. As illustrated in Fig. 11, speed of movement will in fact vary markedly in different directions with differences in prevailing winds or with the presence of barriers such as the Sahara Desert.

Density fluctuations of different populations of the same species frequently exhibit spatial synchrony, meaning that densities go up and down together over a large region, as illustrated by gypsy moth *Lymantria dispar,* in the northeastern United States (Fig. 12). Since there are several possible causes of spatial synchrony, including dispersal of animals between populations, increases in density in one population will produce emigrants that trigger increases in surrounding populations. By far the most common cause of spatial synchrony, however, is the Moran effect, named after the individual who studied the synchrony of lynxes and hares in boreal Canada. In 1953 P. A. P. Moran showed mathematically that the density fluctuations of two or more populations will synchronize, provided they are governed by the same factors, such as the same natural enemies and are also influenced by some common factor such as the weather, which varies from year to year in the same way on a regional

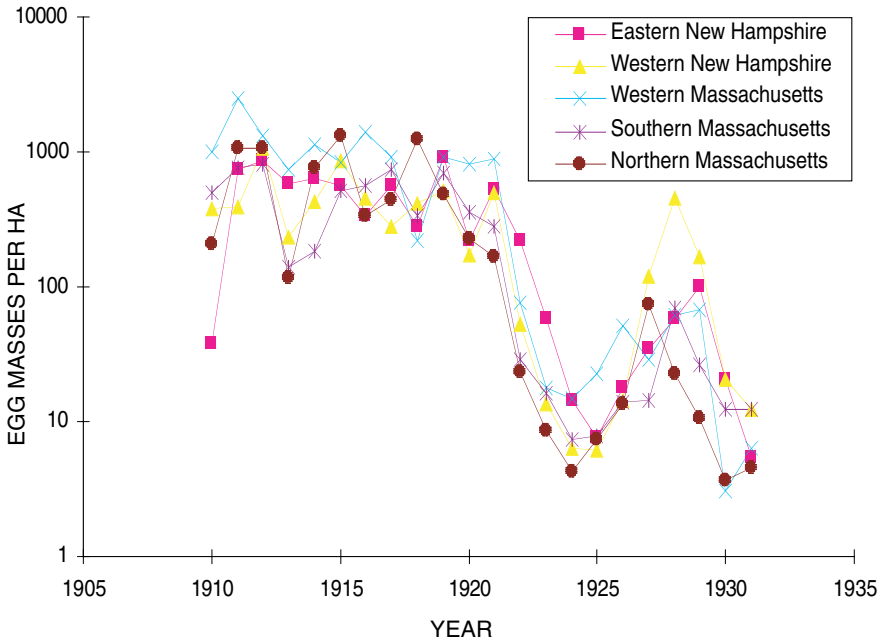

FIGURE 12 Spatial synchrony of gypsy moth populations in the northeastern United States. [Adapted from Campbell, R. W. (1981). *In* "The Gypsy Moth: Research Toward Integrated Pest Management" (Doane, C.C., and McManus, M.L., eds.), *USDA For. Serv. Tech. Bull. 1584,* pp. 65–86.]

spatial scale. The important point is that although variations in weather may cause spatial synchrony, we should not infer that weather is also responsible for fluctuations in density, which might typically be caused by interactions with predators, pathogens, and parasitoids.

CONCLUSION

Population ecology forms the theoretical foundations upon which management of insects is based. Much has been learned from the development of mathematical models of population systems, studies of laboratory populations, analyses of data from natural populations, and experimental manipulation of populations in the field. Despite decades of research, we are still far from understanding the population dynamics of most insects, even those such as gypsy moth or spruce budworm, which have been studied by several generations of researchers. Although various individuals have proposed theories to account for the dynamics of these species and have collected supporting data, rigorous demonstrations that these theories are correct elude us. For most insect species we have neither theories nor data to account for their population dynamics. The reason for this state of affairs is that it is extremely difficult to accurately measure the density of natural populations, especially when densities are low, and even more difficult to measure the impact of various factors causing mortality or variation in fecundity. Each insect and its natural enemies are embedded in a web of interactions whose intricacies must be disentangled. To understand how these intricacies vary over time, such information must be gathered for decades and coupled with experimental manipulations, that are often prohibitively expensive to conduct. The financial and human resources necessary to conduct such research are rarely assembled.

See Also the Following Articles

Biological Control • Insectivorous Vertebrates • Parasitoids • Pathogens of Insects • Predation

Further Reading

Begon, M. (1979). "Investigating Animal Abundance: Capture–Recapture for Biologists." University Park Press, Baltimore, MD.

Begon, M., Harper, J. L., and Townsend, C. R. (1996). "Ecology: Individuals, Populations and Communities," 3rd ed. Blackwell Science, Oxford, U.K.

Begon, M., Mortimer, M., and Thompson, D. J. (1996). "Population Ecology: A Unified Study of Animals and Plants," 3rd ed. Blackwell Science, Oxford, U.K.

Bellows, J. T. S., Van Driesche, R. G., and Elkinton, J. S. (1992). Life-table construction and analysis in the evaluation of natural enemies. *Annu. Rev. Entomol.* **37,** 587–614.

Denno, R., McClure, M. S., and Ott, J. R. (1995). Interspecific interactions in phytophagous insects: competition reexamined and resurrected. *Annu. Rev. Entomol.* **40,** 297–331.

Gilman, M., and Hails, R. (1996). "Ecological Modeling: Putting Theory into Practice." Blackwell Science, Oxford, U.K.

Hassell, M. P. (2001). "The Spatial and Temporal Dynamics of Host-Parasitoid Interactions." Oxford University Press, Oxford, U.K.

Krebs, C. J. (2001). "Ecology: The Experimental Analysis of Distribution and Abundance," 5th ed. Benjamin Cummings, San Francisco, CA.

Krebs, C. J. (2001). 'Ecological Methodology," 2nd ed. Benjamin/Cummings, Menlo Park, CA.

Price, P. W. (1997). "Insect Ecology," 3rd ed. Wiley, New York.

Royama, T. (1992). "Analytical Population Dynamics." Chapman & Hall, New York.

Seber, G. A. F. (1982). "The Estimation of Animal Abundance and Related Parameter," 2nd ed. Griffin, London.

Southwood, T. R. E., and Henderson, P. A. (2001). "Ecological Methods," 3rd ed. Blackwell Science, London.

Speight, M. R., Hunter, M. D., and Watt, A. D. (1999). "Ecology of Insects: Concepts and Applications." Blackwell Science, Oxford, U.K.

Strong, D. R., Lawton, J. H., and Southwood, T. R. E. (1984). "Insects on Plants. Community Patterns and Mechanisms." Harvard University Press, Cambridge, MA.

Turchin, P. (1995). Population regulation: old arguments and a new synthesis. *In* "Population Dynamics: New approaches and Synthesis" (N. Cappuccino and P. W. Price, eds.), pp. 19–41. Academic Press, San Diego, CA.

Varley, G. C., Gradwell, G. R., and Hassell, M. P. (1973). "Insect Population Ecology." Blackwell Science, Oxford, U.K.

Praying Mantid

see *Mantodea*

Predation/Predatory Insects

Ronald M. Weseloh

Connecticut Agricultural Experiment Station

Predators are animals that must consume more than one individual of another animal to complete their life cycle. Under this definition, the predatory habit occurs in a wide range of insect groups. Some orders, such as Odonata and Neuroptera, are wholly or predominately predaceous. Other orders contain large numbers of species that are predatory, such as Hemiptera, Coleoptera, Mecoptera, Diptera, and Hymenoptera. Even a few species in the orders Ephemeroptera, Orthoptera, Plecoptera, Thysanoptera, Trichoptera, and Lepidoptera are predaceous. This article emphasizes the diverse ways in which predaceous insects live.

MODES OF PREDATION

Hunting

Many insect predators carry out active hunting. The adults of tiger beetles (Cicindellidae), ground beetles (Carabidae), and many ants (Formicidae) actively run over the ground to capture prey. Many walking or crawling predators, including ladybird beetles (Coccinellidae), lacewing larvae (Chrysopidae), and syrphid fly larvae (Syrphidae), feed on sedentary insects such as

aphids and scale insects. Other insect predators are agile flyers and actively snatch insects out of the air. Dragonflies and damselflies (Odonata) have large eyes and strong wings and feed mainly on mosquitoes and other small flying insects. Robber flies (Asilidae) also catch their prey on the wing, but will often tackle and subdue insects at least their own size, even bees and wasps. Their legs are very strong and they have piercing mouthparts to suck up prey juices. Other insects, such as many social wasps (Vespidae), pluck insects from the ground or vegetation while flying.

Many predaceous insects live in fresh water and pursue prey by swimming. The giant water bugs (Belostomatidae) are excellent swimmers and often capture and subdue small fish and tadpoles. Backswimmers (Notonectidae) hang ventral side up at the water surface and dart after aquatic organisms using powerful, oar-like hind legs. The diving water beetles (Dytiscidae) are also voracious predators that actively capture prey under water. Water striders (Gerridae) are true bugs that skim over the water surface and prey on small organisms that fall from above. Whirligig beetles (Gyrinidae) also hunt on the water film, but sometimes dive below the surface. Each eye is a double structure, with half adapted to seeing above the surface and the other half below.

Other active hunters are less visible. A large number of beetle and fly larvae, as well as immature insects from other orders, prey on small organisms by burrowing through soil, wet ground, rotting logs, or other vegetation. The immatures of many horse flies and deer flies (Tabanidae), for instance, are predaceous on organisms in moist soil and water. The larvae of some common midges (Chironomidae) feed on other organisms in the mud at the bottom of ponds or along its margins, as do the larvae of some crane flies (Tipulidae). Click beetle larvae (wireworms, Elateridae) are generally plant or detritus feeders, but some are predatory in soil or rotting logs.

Stalking and Ambush

Many predaceous insects subdue active prey, but not by hunting. These are the ambush predators or slow-moving stalkers. Many have special adaptations to efficiently capture prey. Praying mantids (Mantidae) have large, grasping forelegs, superbly fitted for grasping and holding prey, and remain motionless on vegetation until an insect comes near enough that they can strike out with their forelegs. A number of the true bugs also capture prey this way, including assassin bugs (Reduviidae), Nabidae, and ambush bugs (Phymatidae). Dragonfly and damselfly nymphs are successful ambushers that live in water. Their labium, or lower lip, is modified as a grasping tool that can be rapidly shot out to impale an insect, tadpole, or small fish. Tiger beetle adults are hunters, but their larvae dig vertical burrows in the ground and wait with their heads plugging the entrance. Such larvae grab small organisms walking near the burrow. They have back spines that help anchor them so they are not pulled out. After capture, the prey is dragged to the bottom of the burrow and devoured.

Trapping

Insects that trap their food are similar to ambush predators, but go a step further and use inanimate materials to help snare their prey. The ant lions (Myrmeleontidae) are the larvae of insects in the order Neuroptera that as adults look somewhat like damselflies. Larvae burrow into sand and form conical pits. Small, walking organisms, such as ants, fall into these pits and have difficulty escaping because of the loose material of the walls. In addition, the ant lion at the bottom actively tosses its head to remove sand dislodged by the victim. When the prey falls to the bottom, it is grabbed in the long mandibles of the predator, dragged under, and sucked dry. Some snipe fly larvae (Rhagionidae) build similar pits and capture prey the same way.

A few larvae of fungus gnats (Mycetophilidae) capture prey by secreting webs of mucilaginous materials in moist locations, such as beneath rocks, under bark, or in caves. In New Zealand caves, individuals of *Arachnocampa luminosa* not only spin webs that hang from cave ceilings, but also glow to attract adults of chironomid midges that breed in the water below. The colonies of the gnats are tourist attractions, and the ceilings of the caves are covered with pinpricks of light reminiscent of starry nights.

Provisioning

Solitary wasps in the families Vespidae, Sphecidae, and related families are often provisioners—the adult females capture prey that they paralyze with a sting and place the prey in a closed cell along with an egg. The larva that hatches feeds on the enclosed prey until fully developed. Different provisioner species specialize on different prey and construct different kinds of cells. For instance, the mud-dauber sphecid wasp, *Sceliphron caementarium,* constructs cells out of mud and provisions these with spiders. *Ammophila,* another sphecid, constructs cells by burrowing into the ground and stocks them with caterpillars. Others may burrow into wood or plant stems. Prey species vary from aphids to grasshoppers, true bugs, caterpillars, or spiders. Such provisioning wasps as the cicada killer, *Sphecius speciosus,* stretch the definition of predator. The female usually places one paralyzed cicada per underground cell, but sometimes she puts in two. Thus, the resulting progeny are sometimes parasitoids that can complete development on one host individual and sometimes predators that consume more than one prey.

Host Feeding

This is another behavior that blurs the line between parasitoids and predators. Host-feeders are female adult hymenopterous parasitoids, especially of the family Pteromalidae, that feed on host fluids oozing out of oviposition wounds. Such females host-feed to obtain protein for egg maturation. In many cases the host is parasitized as well. But some parasitoids have developed a habit of host-feeding without oviposition, and in some cases the host (which in this case becomes the prey) may

die from excessive feeding. Some parasitoids even use fast-drying fluids secreted from the base of their ovipositors to construct "feeding tubes" from a concealed prey to the surface through which host fluids can be drawn up.

Mass Foraging

There are many social insects, primarily wasps and ants, that actively hunt prey for food. Most of these do so as individuals, capturing small prey and returning directly to the nest. But some ants are able to overcome vertebrate and large invertebrate prey. These are the mass foragers, the army ants and driver ants. These ants move over the ground in large swarms or columns and subdue prey, such as large scorpions and lizards, that other ants cannot. They also devour many small arthropods and deplete an area of prey so rapidly that they must frequently move from one nest site to another. The army ants of the New World tropics and the driver ants of Africa have been depicted in fiction as very dangerous predators that may even threaten humans. While it is possible that penned animals may be killed, the ants are not generally dangerous. The swarm front or hunting column moves forward only relatively slowly, and it is easy to step out of the way. In fact, a swarm of army ants removes vermin as they raid through a house and may be beneficial, so long as the house is unoccupied while the ants are there.

Deception and Predation

Some insect predators use deceptive practices. Firefly (Lampiridae) adults use bioluminescence to attract mates. Each species has a set of recognition flashes that are exchanged between flying males and the sedentary and often wingless females. If the sequence is correct, a male is guided to a receptive female. However, the females of some species mimic the recognition flashes of other species and so lure fooled males to their death.

Perhaps some of the most sophisticated deceptive predation practices occur among the insects that live in or near ant nests. These "ant guests" or "myrmecophiles" may survive by stealing food from ants by soliciting, but many are also predators. To be successful they must deceive the ants into accepting or at least tolerating them. Some of these predators live on the periphery of nests and prey on foragers that they encounter. The reduviid *Acanthaspis concinnula* fools ants by placing the bodies of its prey on its back. Another reduviid, *Ptilocerus ochraceus,* secretes "tranquilizing" attractive substances from trichomes on its abdomen that apparently subdue ants long enough for the bug to kill them.

Other insects, particularly beetles of the family Staphylinidae, live in the brood chambers of ants, where, among other things, they prey on ant larvae. Most of these predators apparently use chemical subterfuge to gain entry and stay in the brood chambers. Apparently, the chemicals mimic recognition substances produced by the ant brood. In fact, ants may pay more attention to the larvae of the beetles than they do to their own larvae. Some lycaenid butterfly larvae also have this habit. Early instars of these caterpillars live on plants and secrete attractive substances that ants collect. Larger caterpillars move to the brood chambers where they prey on ant larvae.

BEHAVIORAL ECOLOGY AND PREDATORS

Behavioral ecology is a field in which actions and movements of individual organisms are studied in order to understand interactions and impacts between organisms and their environment. Predaceous insects are often used as subjects in behavioral ecological studies because they are numerous, generally have a limited number of responses, and are small enough to be easily manipulated and yet large enough for easy observation. For instance, numerous studies have investigated how the number of prey eaten by a predator varies as prey numbers change. These functional responses usually show an increased tendency of predators to devour more prey as the number of prey individuals increases until the predators become satiated, so that the number of prey eaten increases rapidly at first as prey numbers increase and then becomes steady. However, backswimmers (Notonectidae), which feed on a variety of prey, may switch from one prey type to another as the latter become more abundant. In such a case, the functional response is S shaped, at first starting out slowly and then increasing rapidly until slowing again at high prey numbers. Such S-shaped, or sigmoid, functional responses are usually more typical of vertebrate predators than insects.

An important aspect of behavioral ecology has to do with searching by organisms in areas that contain various amounts and quality of food. Many predators, such as ladybird beetles (Coccinellidae), aggregate in places containing many prey. Finding the most profitable way in which a predator can divide its time between encountering prey within an area of prey abundance and searching for other concentrations of prey is an exercise in optimal foraging theory. Predators such as backswimmers (Notonectidae) have been found to do this rather well. The actual time they spend foraging in an area varies with prey abundance in a way that is close to optimal.

Ants have been used to investigate the kind of prey searching called center place foraging. In this situation, an organism departs and returns to a fixed location, often a nest, during a foraging session. Studies of ant foraging show that a center place forager often restricts its searching to only part of the available area, foraging in different areas at different times. Also, such foragers often do not search near the nest, but only at some distance from it. Both these characteristics may increase the efficiency of foraging by decreasing searching overlap with nest mates.

ECONOMIC IMPORTANCE OF PREDATORS

While most biological control work has involved parasitoids, some predators have been used as well. In fact, the first really successful classical biological control project was the control of

the cottony cushion scale, *Icerya purchasi,* by the vedalia beetle (Coccinellidae), *Rodolia cardinalis.* Other coccinellids are important predators of aphids and mites. Ground beetles are important generalist predators, and at least one of these *(Calosoma sychophanta)* is an effective predator of the gypsy moth. Other predators used in biological control practice include the true bugs *Orius* (Anthocoridae) and *Geocoris* (Lygaeidae), clerid beetles, green lacewings, syrphid flies, and marsh flies (Sciomyzidae). On the other hand, some robber flies are voracious predators of bees and are considered to be economically harmful. Also, because of their propensity to attack nontarget prey, generalist predators are not considered good prospects for use in programs that introduce exotic natural enemies into new areas.

Insect predators may be more important to our well-being than is commonly thought. Because they often leave no identifiable remains after devouring prey, the extent of predation is hard to quantify. It is probably true that insect predators help keep the majority of potential pest organisms under control. Without them, the world likely would be quite different.

See Also the Following Articles

Biological Control • Feeding Behavior • Insectivorous Vertebrates

Further Reading

Bell, W. J. (1990). Searching behavior patterns in insects. *Annu. Rev. Entomol.* **35,** 447–467.

Caltagirone, L. E., and Doutt, R. L. (1989). The history of the vedalia beetle importation to California and its impact on the development of biological control. *Annu. Rev. Entomol.* **34,** 1–16.

Clausen, C. P. (1940). "Entomphagous Insects." Hafner, New York.

Cohen, A. C. (1995). Extra-oral digestion in predaceous terrestrial Arthropoda. *Annu. Rev. Entomol.* **40,** 85–103.

Dixon, A. F. G. (2000). "Insect Predator–Prey Dynamics: Ladybird Beetles and Biological Control." Cambridge University Press, Cambridge, U.K.

Hölldobler, B., and Wilson, E. O. (1990). "The Ants." Harvard University Press, Cambridge, MA.

Lövei, G. L., and Sunderland, K. D. (1996). Ecology and behavior of ground beetles (Coleoptera: Carabidae). *Annu. Rev. Entomol.* **41,** 231–256.

Obrycki, J. J., and Kring, T. J. (1998). Predaceous Coccinellidae in biological control. *Annu. Rev. Entomol.* **43,** 295–321.

Pearson, D. L. (1988). Biology of tiger beetles. *Annu. Rev. Entomol.* **33,** 123–147.

Prosorrhyncha

(Heteroptera and Coleorrhyncha)

Carl W. Schaefer

University of Connecticut

The Hemipteran suborder Prosorrhyncha comprises two groups, the Coleorrhyncha (family Peloridiidae) and the Heteroptera. Like all hemipterans, members of these two groups have elongated mouthparts for sucking fluids; unlike all other hemipterans, many heteropterans ("true bugs") suck the fluids of animals—mostly of arthropods but in a few instances of vertebrates. Heteroptera contains 8 infraorders, 79 families, and about 38,000 described species. The families range from the very small (one species) to the very large (> 10,000 species), and heteropterans as a group live in all habitable (and many apparently uninhabitable) parts of the world, from deserts to forest canopies, and on and below the sea. Heteropterans feed on all parts of plants (including on fungi and freshwater algae); some are sufficiently serious predators of pests to be biocontrol agents; and a few (bed bugs, triatomine bugs) feed on human blood. As a whole, Heteroptera is the most diverse of all hemimetabolous insect groups. Nevertheless, we know amazingly little about heteropteran biology and ecology, and the more we learn, the greater this diversity is shown to be.

CLASSIFICATION

The classification of the order Hemiptera is confused, sometimes defies the available evidence, and always arouses controversy. That there *is* an order Hemiptera, few doubt; moreover, that the group of true bugs, Heteroptera, is a phylogenetically valid taxon, no one doubts. Questions arise, however, concerning what had been considered the other suborder, Homoptera, a group that some (mostly homopterists) elevate to ordinal rank; indeed, some homopterists elevate groups within Homoptera to ordinal rank. However, recent work has shown that Homoptera and some of its subordinate groups are paraphyletic (and may not even be monophyletic). As a result, confusion reigns.

The order Hemiptera is characterized by being hemimetabolic (three life stages: egg, nymph, adult), having wings (with a few exceptions), and especially in possessing elongated mouthparts ("beak," or "rostrum") designed for the piercing of plants or animals and the sucking up of their juices. The Heteroptera differ from the majority of other hemipterans. In these other insects (mostly homopterans) the forewings are usually opaque (hence, "Homoptera"), but they are half opaque and half membranous in Heteroptera (again, hence the name). The two groups differ also in the apparent location of the mouthparts, which arise from the ventral surface of the head in both groups but, in Homoptera, arise from the back of the head (sometimes appear to arise from the thorax) and, in Heteroptera, arise from the front of the head. In addition, all homopterans, but not all heteropterans, are herbivorous. Nearly all nonheteropterans, except coleorrhynchans, have a filter chamber, which allows water ingested with the dilute plant juices to be "short-circuited" in the digestive system for rapid elimination. Heteropterans also differ from homopterans in that many feed on animal juices, a few indeed on the blood of vertebrates (including humans). It is likely that animal-feeding characterized the early heteropterans, although this view is controversial.

There is some confusion over the meaning and extent of "Hemiptera" and over just what a "bug" is. To some, "Hemiptera" refers only to Heteroptera, which is treated as an order the equivalent of Homoptera. To others, including most if not all students of Heteroptera, "Hemiptera" is an order with two suborders, Heteroptera and Homoptera. The features (i.e., synapomorphies) that the two groups share seem to be far more basic and significant than the two groups' differences, a circumstance that is best represented by treating all members of the two groups, Heteroptera and Homoptera, whatever this latter group may in fact contain, as belonging to an order, Hemiptera.

Because the group known as "Homoptera" (= non-heteropteran) is either paraphyletic or not monophyletic, the higher classification of Hemiptera is in flux. One consequence of the recent analysis of these relationships (based particularly on molecular evidence) has been the recognition of a small formerly homopteran group, Coleorrhyncha, as the sister group of Heteroptera. These two—Coleorrhyncha and Heteroptera—have been placed by some as a single hemipteran suborder, Prosorrhyncha. Although this idea was based on molecular evidence (and therefore likely to be discounted by some), morphological evidence also supports it. This article considers the hemipteran suborder Prosorrhyncha. Because the Coleorryncha is very small and of limited distribution, the focus is on Heteroptera.

Coleorrhyncha

The Coleorrhyncha is a group that contains the single family Peloridiidae, which had once been treated as a major homopteran group (Coleorrhyncha) and then as a group equal in taxonomic status to Heteroptera and Homoptera *(sensu antiquo)* and perhaps "bridging the evolutionary gap" between them.

The dozen or so species of Peloridiidae have a classic Gondwana distribution; they occur in southern South America, New Zealand, and Australia and are associated with southern beech trees *(Nothofagus),* where they feed probably on moss (like idiostolids, mentioned later under Pentatomomorpha). They are small (2–5 mm long) and flattened, and many have expansions on the head and/or thorax which, semitransparent, are quite beautiful.

Heteroptera

All insects are not bugs. Heteropterans, however, are indeed bugs, and so are often called "true bugs" (i.e., "real" bugs, not "faithful" ones). The word *bug* derives from the Middle English word *bugge,* meaning a "spirit" or "ghost." If one awakens in the morning with red itching welts, one clearly has been visited in the night by a malevolent wraith or spirit—that is, by a bugge. This, of course, is correct: one has indeed been visited by a bug, by a bed bug *(Cimex lectularius),* which has fed and fled, leaving behind a red welt and a mystery. Bed bugs are "bugs" and so, by extension, are all their relatives—all heteropterans. (The original meaning of *bugge* is retained in English in such words as *bugbear* and *bugaboo,* and in the name of the insect-repelling plant, bugbane. And the scientific name of the bed bug, *Cimex lectularius,* means "bed bug," *cimex* being the Latin for *bug,* and *lectularius* describing a small couch or bed. The word "cimex," by the way, was sometimes used by the Romans as a derogatory epithet.)

The suborder Heteroptera is perhaps the largest of the hemimetabolous groups (with about 38,000 species, and probably considerably more are undescribed), and also among the most diverse. The group's diversity is reflected in the variety of organisms fed upon, the variety of habitats lived in, and the variety of habits; these varieties are in turn reflected in the variety of shapes, sizes, and ornamentation found among the Heteroptera. Many heteropterans are aquatic or semiaquatic, and in this feature alone they differ from other hemimetabolans. Many are predaceous (upon other arthropods) and some are hematophagous (upon vertebrate blood), and in this feature too they differ.

Like all hemipterans, heteropterans have elongated mouthparts for piercing organisms and the drawing up of those organisms' juices; predaceous bugs reduce to juice the soft internal organs of their prey. The organisms fed upon include all terrestrial groups of plants (except perhaps algae; but Corixidae include freshwater algae in their diet), other arthropods, snails (fed upon by some giant water bugs), and of course vertebrate blood.

Heteroptera is further distinguished by having scent glands. These occur on the abdominal dorsum in immatures and on the metathoracic pleura in adults. Nymphs lack the thoracic glands, but the abdominal glands occur in adults usually as nonfunctional scars; however, considerable evidence has accumulated showing that in many heteropteran groups one or more of the abdominal glands are functional in adults as well as in nymphs. Another distinctive feature of heteropterans is the scutellum ("little shield"). This triangular structure (broad base anterior) is a modification of the mesonotum; it may be quite small relative to the body, or quite large. In a few families and in a few members of a very few other families, the scutellum is large enough to cover the wings and the entire dorsum of the abdomen.

Regarding the general account of heteroperan biology that follows, and also the more specific accounts of the different groups, it cannot be too heavily emphasized very little is known about most heteropterans: many of the generalities given here are based on small samples, and what we do not know far outweighs what we do know. The reader is urged to take all general statements, if not with a pillar, then with several grains of salt.

Predaceous bugs, and those feeding on vertebrate blood, are not particularly host specific, although some degree of specificity may be imposed upon them by their habitats: bat bugs, living in bat caves, are restricted to feeding on bats, although given the chance they may feed on other warm-blooded

vertebrates; many triatomine bugs live in or with their vertebrate hosts and are perforce restricted to them. A few predaceous bugs are more clearly host specific: the bed bug (everybody's favorite heteropteran) lives only with humans and does poorly when fed other kinds of blood; certain minute flower bugs (Anthocoridae) appear to specialize on certain species of scale insects (although here the specificity may be habitat driven, not prey driven); and some predaceous mirids prefer certain species of lace bugs as prey. There are few other examples.

The host specificity of herbivorous heteropterans is in general greater than that of predaceous ones. Not all herbivorous bugs are host specific, and among those that are, the specificity is sometimes at the species–species, or species–genus level (bug–plant), but more commonly at the family group–family group level (e.g., Alydinae and Plataspidae on Leguminosae; Rhopalidae: Serinethinae on Sapindaceae). Many bugs, and family-level groups of bugs, are oligo- or polyphagous. Of course, the heteropterans that compete with humans for their food and are agricultural pests are those that are either host specific on crop species or, although polyphagous, can build up large populations quickly on crops.

Measuring host specificity can be impeded by the willingness of bugs to probe nonhost plants, either to test their suitability as food or merely to get water. Too many host plant records are of this sort; the best records are those of the feeding of immatures, since immatures, being wingless, cannot go elsewhere to better their lot.

Most heteropterans (perhaps two-thirds) are herbivorous. Many pierce plants to their circulatory systems and feed on the contents of phloem or (less often) xylem; a few very small heteropterans feed on plant cell contents. These heteropterans thus feed not unlike most homopterans. However, plant sap is low in nutrients, and many heteropterans feed on reproductive tissues such as flowers, ovules, and ripe and unripe seeds, which are richer in nitrogen; another group of heteropterans (Miridae and some Tingidae) feeds on somatic tissues that, to the delectation of the bugs, mobilize nitrogen when the host is wounded. A very few heteropterans feed on bryophytes, and a few Tingidae are the only heteropterans to induce plant galls. Many herbivorous bugs, especially in the infraorder Pentatomomorpha, contain bacterial symbionts in gastric ceca or mycetomes; these presumably supply necessary trace nutrients.

As mentioned earlier, several heteropterans may become pests on agricultural crops. In the Old World tropics, cotton stainers (Pyrrhocoridae) feed on developing cotton bolls and may allow entrance of destructive pathogens; the bugs' excreta also stain the cotton (hence the bugs' name). Sunn pests, species of the genera *Eurygaster* and *Aelia* (Scutelleridae and Pentatomidae), are among the most serious pests of small-grain crops (especially wheat) in a broad band of countries from eastern Europe through the Middle East into western Asia. Chinch bugs similarly afflict wheat in North America, and a close relative attacks lawn and golf course grasses in Florida. Lygus bugs (several species), a leaf-footed bug (Coreidae), and southern green stinkbugs (Pentatomidae) feed on a variety of vegetables and fruits and are pests worldwide (lygus especially in the north-temperate regions). Species of *Riptortus* and *Neomegalotomus* (Alydidae) may become pests of legume crops (Fig. 1). Several burrower bugs (Cydnidae) are pests of many crops and range grasses, upon whose roots they feed. The red pumpkin bug (Dinidoridae) may seriously damage cucurbit crops in Asia. There are many other major pests, and very many other minor or regionally restricted ones; nearly 1000 are discussed in a recent book by Schaefer and Panizzi. As the foregoing list suggests, most pests specialize on the plant family to which the crop belongs. A few (e.g., lygus, southern green stink bug, and the leaf-footed bug *Leptoglossus gonagra*) are exceptions, feeding on many unrelated plants and quickly achieving large populations.

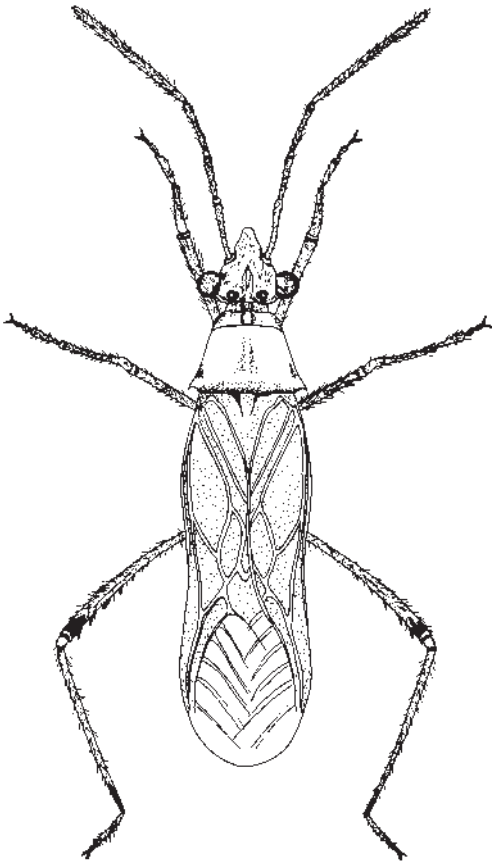

FIGURE 1 An alydid *(Neomegalotomus parvus),* a soybean pest in Brazil.

Only a few heteropterans are vectors of plant diseases; one important group is Piesmatidae, one of whose members transmits a serious viral disease to sugar beets. Overall, however, heteropterans are far less important than homopterans as vectors of plant diseases, perhaps because bugs, unlike many homopterans, do not feed directly in plant cells.

The few bloodsucking bugs may cause damage by withdrawing excessive amounts of blood and thus weakening people who may, in addition, be weakened by disease, or frail in health because very old or very young. Bed bug

populations may increase to the point where such damage may occur.

Of the two important groups of blood-sucking bugs, the bed and bat bugs (Cimicidae) do not transmit disease pathogens, although there is some slight evidence that the hepatitis B virus may be transmitted by bed bugs in southern Africa. On the other hand, triatomine bugs (Reduviidae, subfamily Triatominae) are vectors of *Trypanosoma cruzi. T. cruzi* causes Chagas' disease, a debilitating and often fatal disease of the New World tropics and subtropics (a handful of cases reported in the United States); major international efforts are now under way to eliminate this disease by eliminating the places where the bugs hide. These efforts have been successful in many areas of South America.

Although most heteropterans having an economic impact on humans are harmful, some are beneficial. Chief among the latter are the predaceous bugs that feed on insect pests. Because predators as a rule are not host specific, there are no one-to-one relationships of predator to prey in Heteroptera or in any other group considered for biological control. Nevertheless, many groups of heteropterans have been studied for, and some successfully used in, biological control programs: asopine pentatomids (Pentatomidae, subfamily Asopinae) against many pests, especially garden pests (indeed, an artificial aggregation pheromone has been developed to "call in" these predators); certain minute flower bugs (Anthocoridae) to combat stored-product pests; some predaceous mirids on vegetable and greenhouse crops; and an array of predaceous bugs in field crops. The one exception—remarkably, considering its wide distribution, its variety and numbers of species, and the size (and presumably appetite) of its members—is the Reduviidae. The biological control potential of this family has been surprisingly little explored. It should also be mentioned that the value of very small predators—Anthocoridae, small mirids, early instars of other predators—is unknown; these feed on arthropod eggs and small instars, whose absence goes unnoticed because their presence was overlooked in the first place.

In 1998 Cohen showed that predaceous bugs feed differently than had been thought: instead of merely sucking up the fluids inside a prey, these bugs inject salivary enzymes that digest all soft tissue; the digested material is then sucked up. This discovery has profound implications for biological control: a single predator will be satiated by a far smaller number of prey than had been believed, and will therefore be less effective in controlling pests. This realization is only slowly making its way into the biological control community.

Of less importance is the control of harmful plants, weeds, by plant-feeding heteropterans: Some tingids have been used in various parts of the world to control (not very successfully) the invasive plant lantana; cactus bugs (*Chelinidea,* Coreidae) have been tested—with small success—against invasive cactus in Australia; and several other bugs have been and are being tested for possible control of other weeds. So far, no major plant pest has been controlled by a heteropteran, but the exhaustive study of biology and ecology necessary for success in such an endeavor is only now being undertaken.

Detailed accounts of economically important heteropterans are provided in the book by Schaefer and Panizzi.

COURTSHIP, MATING, OVIPOSITION Unlike feeding, the sexual lives of heteropterans show little variety. There is little courtship, although males of several species may fight (usually briefly) for the attention of a female. The sexes are brought together initially probably by sex pheromones and, in a few species, by sex-neutral aggregation pheromones; however, only a tiny handful of species, all of them land bugs, has been shown to have pheromones of both types. Mating occurs end to end or with the male directly or diagonally across the female; the position seems to vary with the family group or even the infraorder of the bugs. Mating may be brief or long; if long, the larger female may move about and even feed, while the hapless male scrambles along behind, walking backward.

Mating by bed bugs and some relatives in Anthocoridae and Nabidae, is (to quote from another context) nasty, brutish, and short. The male mounts the female and, with a scimitar-shaped clasper, pierces the side of her abdomen and deposits sperm; in primitive species this traumatic insemination (as it is called) may occur in various parts of the abdomen, allowing sperm to enter the abdominal cavity; in more advanced species (including the bed bug itself), sperm is deposited in a special patch of tissue. Such insemination leaves small healed wounds in the female's cuticle; counting these, the number of times an individual has been mated can be determined. Sperm then progresses to a sperm storage organ at the base of the common oviduct.

The internal reproductive structures are very conservative—that is, they resemble those of very many other insects. There is a pair of ovaries or testes, each almost always consisting of seven ovarioles or testicular follicles. Paired ducts lead from them to a common duct, with which may also be associated various glands and (in females) organs for sperm storage. The external genitalia are more complex, consisting in the male of a genital capsule ("pygophore") containing a pair of claspers and the intromittent organ and, in the female, of a genital chamber into which sperm is deposited and a complex ovipositor designed in some bugs for the laying of eggs on surfaces ("platelike ovipositor") and, in others, for inserting eggs into crevices or actually into slits in vegetation ("laciniate ovipositor").

As is true of many other groups, the external genitalia of male heteropterans are usually more complex and varied than those of the female. In Heteroptera as in other groups, genitalia provide very useful characters for taxonomic and phylogenetic studies.

The egg is often the stage that survives inhospitable seasons (winter, dry season), and so eggs are laid where they may best be protected from the elements as well as from predators and parasites. Eggs are laid also where food will be available, because the immatures, wingless (of course), cannot seek food themselves. The latter requirement presents certain difficulties to predaceous bugs, which cannot be certain where prey will

occur. Perhaps this accounts for the habitat specificity (resulting in a certain degree of prey specificity) of some predators, as mentioned earlier: predators' eggs are laid on plants on which certain prey species are host specific; the predator then in turns appears itself to be host specific, on the herbivore. Perhaps.

The egg itself varies considerably in shape from group to group; it may be round or oval, and in stinkbugs especially the eggs are barrel shaped. Each egg bears small openings through which it was fertilized, and many have other small openings for oxygen intake. Carbon dioxide is released through the shell (chorion). The few eggs that have been studied have a complex architecture of the chorion, presumably related both to gas exchange and to maintaining an appropriate humidity balance.

Eggs are laid singly or in small batches (rarely more than 100), and if in batches may be laid apparently at random or in tidy rows, sometimes more than a single egg deep. They may be deposited on surfaces (of leaves, stems or trunks, ground, stones, etc.) or placed into crevices; the crevices may in some groups be created by the female's ovipositor. The eggs of some mirids and tingids absorb water in the spring, and thus nymphs hatch just as the plant becomes suitable for feeding.

The females of some heteropteran groups (some Tingidae, many—all?—Acanthsomatidae; a few others) guard their eggs and the early immature stages. Guarded eggs have been shown to be less heavily parasitized, and guarded nymphs less heavily preyed upon, than unguarded eggs and nymphs. A single female will lay several batches of eggs. In a few groups [some giant water bugs (Belostomatidae), at least one coreid], females lay eggs on the backs of males. Belostomatid males then aerate the eggs, sweeping fresh water over them; without such aeration, the eggs might die of oxygen lack and/or of fungal infection. In some species each new batch requires a fresh mating; in bloodsucking species each new batch requires a blood meal.

Hatching occurs sometimes through areas of weakness in the egg's chorion; not all species' eggs have such areas. In some groups (Pentatomoidea, some Coreoidea, and others?) the new hatchling (first instars) do not feed, although they may take water, and pentatomoids may suck up symbionts from the emptied eggs. Nearly all heteropterans have five immature stages (instars) (those that do not, have four; very rarely are there six), and except for blood-feeding bugs, the stimulus to molt from one instar to the next is unknown. Blood-feeding bugs require the stimulation (manifested as a distension of the abdomen) of a blood-meal to molt.

As the instars grow larger at each successive molt, the wing pads develop; as these become more and more distinct, one can determine the instar by the degree of their development.

The instars (except sometimes the first one) feed for the most part on the same food as their adults. Exceptions include predaceous bugs, whose smaller instars may take smaller prey (although a group of small instars may attack a single large slow prey, like a caterpillar); and some herbivores, whose instars do not feed on the plants chosen by than their adults, probably because of seasonal differences in the plants' availability, and/or because some plants may provide nutrition to the young and nutrients needed for reproduction to the adult.

Development (egg to egg) in temperate climes usually takes a year, the egg overwintering or sometimes a mated female (less often, other stages). In regions of less predictable or less severely contrasting seasons, there may be several—rarely, many—generations in a year. Even when there is only one annual generation, however, populations of some pest heteropterans can build up over a very few years.

A distinctive feature is the possession by some heteropterans of a pair of small (micro- or *m*-) chromosomes. The sex-determining mechanism is XY, although XO occurs in some groups and, in others (especially the bed bugs, Cimicidae), multiple X chromosomes may occur. The chromosomes are holocentric, and numerical doubling of the autosomal complement has occurred in several groups.

DISPERSAL As with all living things, dispersal is an important aspect of heteropterans' lives, necessary for the leaving of places no longer suitable and for the discovery of places as yet unexploited; it is necessary, too, for the finding of nonsibling mates. The wingless immatures depend upon the mother to place them where resources—biotic and abiotic—are suitable for development.

Dispersal for mate finding appears to be far more frequent among bugs than dispersal for the discovery of new habitats, although of course it may often serve both purposes. One important exception is the dispersal of cotton stainers (*Dysdercus* spp., Pyrrhocoridae) from areas where high populations have depleted food to new areas, which may include cotton fields. After the move, females resorb the flight muscles, converting the products into eggs; thus such dispersal occurs but once. (Evidence suggests that not all *Dysdercus* spp. do this.) Another exception may be the moonlit flights of some giant water bugs (Belostomatidae) in Africa; it is not clear that these flights correspond to drying of habitat. Like other insect groups, Heteroptera contains some groups that are drawn to light traps (white or black light) and some groups that are not; no one has listed these groups, much less sought a correlation between being attracted to light and having a propensity to disperse.

Of course, for many heteropterans resources are replenished: for predators, new prey is continually being produced; and many smaller herbivores feed on annual plants that appear and disappear too rapidly for their populations to be seriously reduced. In addition, many herbivorous bugs feed on several different species and turn easily from one host to another. Thus for these bugs, dispersal may at times be necessary, but it is not a constant need.

Far better documented for far more heteropteran species is movement in response to sex pheromones, often produced by both sexes. In addition, aggregation pheromones occur in some species. This type of dispersal is of course a "pull" dispersal, not a "push" dispersal: the insect is drawn toward

something rather than being driven away. Nevertheless, like the messengers in *Through the Looking-Glass* (one to fetch and one to carry), movement toward a conspecific may draw one also to fresher resources. The pheromone of at least one beneficial predaceous bug has been synthesized, patented, and marketed to draw these bugs into users' gardens.

ENEMIES AND DEFENSE Heteropterans have the same enemies as other organisms: predators that would eat them from the outside and parasites that would eat them from the inside. The major predators—again like other insects'—are insects; vertebrates seem to be less important than they are for some other insect groups. Bugs may be protected from vertebrate predators by the secretions of their metathoracic scent glands, which also (it has been suggested) protect bugs from ants; the experimental work demonstrating these possibilities is scant and mostly anecdotal.

Many bugs are also protected from visually orienting vertebrates by being bad tasting and warningly colored. Aposematic coloration occurs in nearly every group of land bug as an arresting contrast of black and red, yellow, orange, or white; the similarity of color and pattern has little to do with genetic or phylogenetic relatedness, and much to do with the strictures placed on color by limited metabolic pathways, as well as the limits placed on pattern and design by the strictures of small body size. Thus unrelated aposematic bugs may resemble one another merely because the possibilities for variety are few. However, as in other groups of animals, both Batesian and Müllerian mimicry occur in Heteroptera: the edible look like the inedible (Batesian) and the inedible look like other inedibles (Müllerian). Mimetic assemblages occur, although they have been little studied. One comprises some cotton stainer relatives (Pyrrhocoridae) and their predators, certain Reduviidae. It has been suggested that the reduviids can thus slip unnoticed upon their prey ("aggressive mimicry"). This seems outrageously unlikely: more probably, the group as a whole benefits from a collective aposematicism, from which the pyrrhocorids benefit more than they lose from being preyed upon.

Few creatures like to eat ants, and it is not surprising that ant mimicry has arisen often in the Heteroptera. The immatures of some Alydidae resemble specific species of local ants; and as the immatures become larger, they resemble increasingly larger species of ants. One assumes that looking like ants is good; no one has tested the idea.

None of these defenses (except perhaps scent) is especially effective against other insects. More effective (guessing here) are the thickened cuticles of many heteropterans, the coriaceous hemelytra, as well, perhaps, as the enlarged scutellum of several bugs.

Body shape may also play a part: the round sleek surfaces of Canopidae, Plataspidae, Megaridiidae, and a few other heteropterans offer the jaws of small predators (especially mandibulate insects) little purchase. And the many small spines of coreid nymphs (and others), and the single large spine of adult Cyrtocoridae and some Podopinae (Pentatomidae), must also prove to be unpleasant surprises to vertebrate predators.

It is difficult to know whether heteropterans are more or less parasitized than other insects. The eggs of certain species are often well parasitized by wasps, and the immatures and occasionally the adults by tachinid files. Two indirect bits of evidence suggest that parasitizing bugs is an old and successful way of life: certain family groups of parasites have evolved that parasitize only heteropterans (Diptera: Tachinidae: Phasiinae; several genera of Hymenoptera). And parental care has evolved several times in Heteroptera, to ward off potential parasites (and predators, but these seem to be a lesser evil).

It has been carefully shown that guarding by parents (usually but not always females) protects certain lace bugs (Tingidae), coreids, and many acanthosomatids from parasitism and predation; and the eggs placed upon the backs of some giant water bugs (Belostomatidae) are protected from fungal parasitism by the water currents generated by the male's rowing movements.

THE INFRAORDERS AND FAMILIES OF HETEROPTERA Cladistic work on the morphology of heteropterans has resulted in a new classification into eight infraorders: Enicocephalomorpha, Dipsocoromorpha, Gerromorpha, Nepomorpha, Leptopodomorpha, Cimicomorpha, Pentatomomorpha, and Aradomorpha. In a commonly accepted (but not thoroughly tested) analysis, each of these infraorders is the sister group of those succeeding. Aradomorpha, the most recently proposed, is not universally accepted.

To some extent, these groupings correspond to earlier ones: Preceding this arrangement, for example, Heteroptera was divided into Hydrocorisae (= Nepomorpha, with some adjustments), Amphibicorisae (= Gerromorpha, with some adjustments), and Geocorisae (the remainder, with some adjustments): "water," "amphibious," and "terrestrial bugs," respectively. The last was divided into Cimicomorpha and Pentatomomorpha, names retained in the new classification with the same family composition (with some adjustments in Cimicomorpha); Aradomorpha was included in the earlier Pentatomomorpha. Each -morpha name is based on a genus in the infraorder.

The following treatment of infraorders and families of Heteroptera, is presented, once again, with the admonition to recognize that little is known in this area and often details are generalized to an entire family from what is known about a tiny fraction of its members.

Enicocephalomorpha, Dipsocoromorpha These small infraorders are the most primitive in Heteroptera. The latter is usually assumed to have arisen from the former, but there is some suggestion that each may have arisen separately from one or more related preheteropteran ancestors. Each is small: the two families of Enicocephalomorpha, Aenictopecheidae and Enicocephalidae, contain about 20 and 400 described species, respectively; however, it is probable that several thousand species remain to be described.

The same is true of the Dipsocoromorpha, whose five families are Ceratocombidae (50 described species), Dipsocoridae (30), Schizopteridae (120), Hypsipteryigdae (3), and Stemocryptidae (1). As in Enicocephalomorpha, there remain to be described many hundreds, if not thousands, of dipsocoromorphan species.

Members of both groups are small (Enicocephalomorpha, 2–16 mm long; Dipsocoromorpha, perhaps the smallest of the heteropterans, 0.4–4 mm long). They are all predaceous (as far as is known) and live in or on the soil surface, in leaf litter, or in the interstices of mosses and similar low-growing plants. Their small size, and their life in a habitat of remarkably small interest to biologists or ecologists, render them apparently rare, but actually locally abundant yet nearly wholly unknown.

This is unfortunate, because not only are these bugs of great ecological interest (as predators—and perhaps scavengers—in an ancient and poorly known habitat), but because many occur worldwide and study would yield valuable biogeographic information. Moreover, since they are the most primitive of heteropterans, careful study should reveal much about the origins of this group, as well as its relationships to certain homopteran groups: some enicocephalomorphans share some genitalic characteristics with some auchenorrhynchous homopterans; and Aenictopecheidae is defined wholly by plesiomorphies.

Some enicocephalids swarm, either in single-sex or mixed-sex swarms, and have been confused with midges! This statement sums up our knowledge of enicocephalomorphan reproductive biology!

Gerromorpha Members of this infraorder are the semiaquatic bugs whose members live near or on water, but never in it. Several members of several gerromorphan families live near or on the surface of the sea and are among the only heteropterans that are, if imperfectly, marine; halobatines (a subfamily of the Gerridae) may occur many kilometers from land. These gerromorphans are *semi*aquatic bugs—they do not get wet, except occasionally; and indeed a few live not where it is wet, but where it is merely damp. Like other water-associated animals (e.g., Nepomorpha), gerromorphans are usually dark above and pale below.

All are predaceous, those living on the water surface feeding on land arthropods that fall onto and are caught by the water's surface film; there is some evidence that these bugs may also feed on aquatic crustaceans that are seized upon rising to the water's surface. Prey is located by many bugs by ripples caused by its struggling. The bugs probably secrete something unpleasant because they themselves are only occasionally preyed upon by fish.

Many gerromorphans are fully or partly winged, some are wingless, and populations of some are pterygopolymorphic (some members wingless, some winged), a phenomenon that may also vary seasonally. The hormonal and ecophysiological bases for pterygopolymorphism has been studied in a few gerrid species. This variety, as well as variety in diapause physiology, is doubtless related to the temporary nature of the habitat, which may dry up or freeze at various locations and times of the year. In temperate regions, many gerromorphans overwinter in debris away from, but close to, their water habitat.

Because of their ubiquity, and especially because of their tight adaptations to an unusual way of life, gerromorphans (especially gerrids) are being used more and more to work out questions in communication and reproductive evolutionary ecology (including courtship). These efforts are helped of course by the wide interest in aquatic ecology, and by the excellent phylogenetic foundations provided by Andersen in 1982 and much recent work as well.

Being semiaquatic predators, a few gerromorophans have been studied for their ability to control rice pests and the aquatic larvae of biting flies; in each situation, success is at best limited, probably because rice pests do not fall into the water often enough and aquatic larvae do not spend much time at the water's surface.

There are eight families: Mesoveliidae (35–40 species), Hebridae (150), Paraphrynoveliidae (2), Macroveliidae (3), Hermatobatidae (8), Veliidae (720), Gerridae (620), and Hydrometridae (110–115). Several of these families—or components of them—were once contained within a more exclusive Gerridae or Veliidae, but the thorough phylogenetic studies just mentioned have yielded a firm and stable systematic classification.

The insects range in size from very small (1 mm long) to the aptly named *Gigantometra* (Gerridae), which is up to 36 mm long. Most gerromorphans are at the small end of the range.

Members of the family Gerridae occur worldwide and are among the most frequently seen and frequently admired inhabitants in—or rather on—bodies of fresh water. As they skate about on the surface, their shadows are patterned dark on submerged objects in the water body. This aesthetic appeal, and the bugs' ability to remain on the water, not in it, have earned them various names, such as "water bugs," "pond skaters," "wherrymen" (in Britain), and even "Jesus bugs."

Their shape is characteristic and easily recognized: the head extends somewhat forward of the eyes, the abdomen is often truncate and its segments compressed, and the middle and hind legs are greatly elongated, sometimes looking as if they had been designed for a larger bug. All legs are used in moving on the water's surface, and the front legs are short, often stout, and used too for prey capture. The claws of all legs are subapical (not apical, as is usual). Probably this is an adaptation for moving on water surface films—something they do with remarkable speed and agility.

Much work has been done on communication and courtship/reproduction behavior in a few gerrid species: communication is via wave patterns of surface waves and ripples created by the bugs; these ripples apparently serve several functions, including location and courting of potential mates, defining territories, and perhaps even recognizing conspecifics for the avoidance of cannibalism. The extent to which other gerromorphans use similar signals is unknown, although there is some evidence that a veliid does, and it seems reasonable that other surface skaters would also.

Veliidae (riffle bugs, broad-shouldered bugs) is the largest family; the genus *Rhagovelia* alone contains more than 200 species. Like gerrids' claws, veliids' claws are subapical; the legs are not relatively so long (not greatly surpassing the body), the abdomen is more often elongate, the pronotum is larger (covering the rest of the thoracic terga), and the middle legs' claws of some veliids are modified into a fanlike structure for pushing against the water's surface. In general, the veliids are smaller than gerrids: 1–10 mm vs 1.5–35 mm in length.

Like gerrids, veliids for the most part occur on the surface of ponds or of the quiet stretches of moving water; also like a few gerrids, some veliids live on ocean surfaces, and others are nearly terrestrial, occurring in damp situations or intertidally, or in mangrove swamps. Veliids too feed on smaller animals trapped in the surface film.

Again like gerrids, veliids can move very quickly, and they move even faster when they secrete from the mouthparts a fluid that lowers the surface tension and causes them to "scoot" away.

Veliids may form large assemblages of very many individuals, sometimes almost seeming to cover a stretch of water. What brings and keeps the bugs together is unknown, nor is the advantage of these assemblages clear.

Members of Hebridae are the velvet water bugs, so called because of their coating of fine (presumably hydrofuge) hairs. Quite small (1.2–3.5 mm long), they are not often collected and live in damp places such as the edges of water bodies, in moss, or in and among waterside vegetation. The family occurs worldwide. Hebrids look somewhat like small veliids, both being somewhat broader and thus relatively shorter than gerrids.

As the names suggest (in English and Latinized Greek), water treaders or marsh treaders (Hydrometridae) walk with measured tread on surfaces close to the water's edge, and sometimes even on the surface itself. They are classic sit-and-wait predators, motionless for long periods until some small prey comes near. Hydrometrids are elongate insects, small to moderate in size (2.5–20 mm or so long); a distinctive feature is the position of the eyes about halfway along the length of the head. The genus *Hydrometra,* with 80 to 85 species, is found worldwide, but the rest of the family is tropical.

Mesoveliidae may be the most primitive family in the infraorder, possessing as it does a number of primitive character states and sharing only a few advanced states with other gerromorphans; indeed, the group has at times been excluded from the infraorder, although its inclusion there is now firmly based. Like so many gerromorphans, mesoveliids are small and pterygopolymorphic (some species); *Mesovelia* occurs worldwide, but most of the remaining genera are localized. Mesoveliids vary in their habitats: some (many *Mesovelia*) are active on the water surface, but others live in damp habitats some distance from water; a few are cavernicolous, and a few are intertidal.

The other three families (Paraphrynoveliidae, Macroveliidae, and Hermatobatidae) are small in number of species (113 species in all) and in size (1.5–2.5, 2.5–5.5 and 2.5–4 mm long, respectively). The first has so far been found only in southern Africa; the second only in the New World; and members of the third live near coral reefs, where they remain in air bubbles during high tide and feed during low.

Nepomorpha Most of these are the truly aquatic bugs—that is, bugs that actually occur under water and get wet; the group is the same as the earlier Hydrocorisae (water bugs). Most are predaceous, but corixids are, as a group, more omnivorous. Although nepomorphans live much of their lives in water, they also are active aerial dispersers. It is hard to imagine, therefore, why most of these bugs lack ocelli: even if unnecessary below the water's surface, ocelli ought to be useful in the air (e.g., the riparian families have ocelli). Water bugs range from small (1 mm long, some Helotrephidae) to huge, the largest of the Heteroptera (112 mm long, some Belostomatidae). Like gerromorphans, and for the same reason, most nepomorphans are much paler ventrally than dorsally. Unlike gerromorphans, most (not all) nepomorphans are streamlined, fusiform; those that are not (Gelastocoridae, Ochteridae) are riparian, not wholly aquatic. The antennae of nepomorphans are short and often concealed; hence an earlier name, "Cryptocerata" ("hidden horns").

Living as most nepomorphans do, under water, the problem of breathing arises and has been solved in several ways. Two families are nearly terrestrial: Gelastocoridae and Ochteridae live on the edges of streams and breathe like terrestrial bugs. Belostomatids and nepids have extrusible (belostomatids) or permanently exserted (nepids) "airstraps" or siphons, which can be thrust above the surface. Members of the other families rise to the surface periodically to trap air in hydrofuge hairs; oxygen is then taken into the body (much carbon dioxide here as in most animals is released through the body wall); in some groups and to some extent, plastron respiration occurs (i.e., as the oxygen tension in the air bubble drops, oxygen is drawn into the bubble directly from the water). Plastron respiration is so efficient in Aphelocheiridae that these bugs may remain permanently submerged.

Because of these bugs' importance as fish food (and sometimes as human food), because they occur in aquatic habitats, which have always held a fascination for ecologists, biologists, and folks in general, and because some are so large, nepomorphans have probably been better studied as a group than any other infraorder. The literature on the systematics, biology, and ecology of the group and its member is very large.

There are nine families: Corixidae (at least 600 species), Nepidae (225 species), Belostomatidae (150), Naucoridae (nearly 500), Notonectidae (350), Pleidae (40), Helotrephidae (44–120 species: authorities differ), Ochteridae (50–55), and Gelastocoridae (100). All but the last two live below the water's surface; Ochteridae and Gelastocoridae are riparian (ripicolous).

Corixids (family Corixidae) are the water boatmen, found commonly in ponds, lakes, and (less often) streams throughout the world. More than 600 species make this the largest nepomorphan family. Aspects of their morphology differ so greatly from those of other bugs that Corixidae has

sometimes been placed in its own group, separate from other heteropterans. In particular, the labium of the mouthparts is broad and fused to the head, and the foretarsi are modified (enlarged, somewhat flattened, with an array of long setae) for food gathering. Like other nepomorphans, the hind legs are flattened and hairy, for swimming; in particular, corixids look superficially like notonectids, but corixids are flatter and swim rightside up. Under magnification, the dorsum of many corixids has many fine horizontal zigzag and anastomosing pale lines; this feature too distinguishes these bugs from notonectids and, indeed, from other bugs.

Water boatmen are 2.5 to 15 mm long and occur throughout the world; they live in aquatic habitats of all kinds, including the intertidal. Unlike other nepomorphans, corixids probably feed mostly on algae and (one suspects other small bits of living or dead organic matter). Some are carnivorous, however; indeed, probably many are (only a few have been studied) and are important predators in waters containing few others (e.g., acidified or saline inland waters). Excellent fliers over long distances, water boatmen are often among the first animal colonizers of new aquatic habitats and have even been captured landing on the shiny roof of an automobile in the desert far from the nearest water.

Corixids are of some small importance in feeding on aquatic larvae of noxious dipterans; they are of greater importance as food for commercially valuable fish (and in some cultures, for humans). In addition, they have been used as indicators of organic pollution.

The function of stridulation has been more thoroughly studied in Corixidae than in any other heteropteran group. (The so-called stridulitrum on the male's abdomen is not sound producing but probably aids in clasping the female.)

Superficially similar to corixids are the Notonectidae, backswimmers [wherrymen (cf. also Gerridae), boat flies] which, however, swim upside down; notonectids swim ventral-side down because that is where the buoyant air bubble is. Backswimmers, like water boatmen, are fusiform and also have flattened hirsute hind legs for swimming. However, they have typically heteropteran mouthparts, are wholly predaceous, lack modified foretarsi, and never have the pale irregular linear markings of many corixids. They range from 5 to 15 mm in length. Moreover, unlike corixids, notonectids bite viciously and painfully, as many a hydrobiologist has learned. The largest genus, *Notonecta,* occurs worldwide, and the other genera are more localized, although sometimes within entire continents.

Like corixids, notonectids are excellent fliers and are also early colonizers of new bodies of water. Some notonectids also may use sound in courtship and mating, but the evidence here is far more slight than that for corixids. Backswimmers inhabit nearly all types of water, although records from saline or near-saline waters are scarce.

These bugs are eager predators and they prey near the water's surface, whence they may drive their prey, a procedure aided by the bugs' buoyancy. They feed on other arthropods (and occasionally fish: see later), including small crustaceans and the larvae and pupae of blood-feeding flies. Mosquito larvae and pupae are favored by some notonectids, but over long periods they do not provide consistent biological control; however, female mosquitoes may recognize notonectid-infested ponds and refuse to lay eggs there.

Backswimmers are themselves food for several kinds of fish (good), but may also feed upon fish larvae (bad); however, neither being fed upon nor feeding is of great economic importance.

Pleidae (pygmy backswimmers) and Helotrephidae (no common name) resemble very small (both are about 1–4 mm long) notonectids; however, pleids lack the flattened hirsute hind legs of notonectids (and helotrephids). Both are oval to globular and very streamlined; helotrephids indeed have head and thorax fused. Like backswimmers, these bugs (both families) swim upside down (although they can at times swim rightside up), and they presumably get oxygen and prey in much the same way as do notonectids. Pleids occur almost exclusively in still waters, permanent or temporary; helotrephids live in almost any kind of water, still, or running, permanent or temporary, hot springs, or water seeps; one species can even wait in the desert for ephemeral waters. Members of both families may be of some small value in controlling mosquito larvae. Pleidae is wholly, and Helotrephidae mostly, tropical.

The giant water bugs comprise the Belostomatidae; here are the largest of Heteroptera, ranging from 9 to 112 mm in length. All are predaceous, and the largest may be important pests in fish hatcheries; one group of species feeds on snails in Africa and is of some value in suppressing populations of snails that carry bilharzia (schistosomiasis); the report of a giant water bug attacking baby ducklings may (or may not) be a rural legend. Although active swimmers, belostomatids usually wait on submerged materials to capture prey swimming by. The bugs are attracted to lights, which explains a common name, "electric light bugs."

The belostomatid body is broadly oval and somewhat flattened; the forelegs are modified for grasping, and one claw is reduced. The eighth abdominal tergum is modified laterally into a pair of extrusible airstraps, which can be thrust above the water's surface and through which air is moved to an air bubble below the wings. In several subfamilies several antennal segments bear fingerlike projections.

Giant water bugs are well known for the remarkably painful "bite" (actually, stab) they can inflict. But they are even better known because females of one subfamily (Belostomatinae) deposit their eggs on the backs of males. The males care for the eggs by swishing water over them; this aerates the eggs and (it has been shown) prevents fungal growth.

These bugs occur throughout the world and are the most diverse in the tropics. In some cultures giant water bugs are dried and eaten; in parts of Asia they are a delicacy.

Closely related to Belostomatidae is Nepidae, the water scorpions; they range in length from 15 to 50 mm. Nepines (subfamily Nepinae) look rather like small belostomatids

except for the terminal air siphon, which may be more than twice the body's length. Ranatrinae also has a long air siphon, and its body is long and very narrow, nearly round in cross section. Like the antennal segments of belostomatids, those of nepids have projections; unlike belostomatids, nepids have one-segmented tarsi.

Water scorpions do not swim actively; they are, rather, "sit-and-wait" predators, often sitting patiently on vegetation in quiet ponds until appropriate prey wanders past. They grab the prey swiftly with the forelegs, rather like aquatic preying mantids.

Like Belostomatidae, Nepidae occurs worldwide but is particularly abundant in the tropics.

The toad bugs, Gelastocoridae (6–15 mm long), are riparian, not truly aquatic; they look (at least to the extremely untrained eye) like tiny toads, and they jump. The body is rough-textured and lumpy ("warty"); the eyes large, and the forefemora and foretibiae modified for prey capture. They occur throughout the world, and most species are tropical.

Ochterids (Ochteridae), or velvety shore bugs, are also riparian, living along the margins of water bodies. Unlike most other nepomorphans, ochterids have antennae that are visible (although small); the bugs are 4 to 9 mm long; and are mostly but not wholly tropical. With their visible antennae, suboval shape, and often mottled color patterns, ochterids look rather like saldids (see later, under Leptopodomorpha), a leptopodomorphan group that indeed is ecologically replaced by ochterids in the tropics.

Naucoridae, which contain the creeping water bugs, or saucer bugs, are 5 to 20 mm long and elongate-oval, with hidden antennae and enlarged forefemora. They look like small belostomatids; however, naucorids lack airstraps, their eyes are relatively larger than belostomatids', and their dorsal surface is sometimes mottled.

Two naucorid subfamilies, Aphelocheirinae and Potamocorinae, are often treated as separate families in the recent literature, partly because they have small but visible antennae and other nonnaucorid features. The former subfamily (60 species) ranges in length from 3.5 to 12 mm and contains the best plastron breathers in Heteroptera. So effective in trapping and keeping the air bubble are their hydrofuge hairs that these bugs can remain their entire lives below water, using the bubble as a physical gill. This ability allows aphelocheirines to live deep in the water, a niche unavailable to other aquatic bugs which must surface at least occasionally to replenish their air supply. Aphelocheirinae, like Nepinae, are primarily tropical, but with significant representation in the Holarctic (Nepinae) or Palearctic (Aphelocheirinae) region.

The eight potamocorine species are all Neotropical and look like very small (2.5–3 mm long) naucorids except for the visible if small antennae. Nothing is known of their biology.

Leptopodomorpha With one exception, the families of this infraorder have few species and are poorly known. The family groups in Leptopodomorpha have been arranged variously by various authors. This article follows Schuh and Slater, who in 1995 listed four families: Saldidae (265+ species), Aepophilidae (1), Omaniidae (5), and Leptopodidae (about 40). Most leptopodomorphs live near water, in damp places, and members of the Omaniidae and Aepophilidae are intertidal; many leptopodids, however, live in quite dry habitats. These bugs are small (about 1–7.8 mm long), with large eyes (except Aepophilidae) and a broad head. All are predaceous (as far as is known).

The largest and best known family is Saldidae, whose members range in length from 2 to 7.8 mm. Known as shore bugs, saldids occur in damp areas near water; they are excellent jumpers, fly readily, and are more difficult to collect than they at first appear to be. The eyes are indented medially (kidney-shaped), and the dorsum of many species is mottled dark and light. The male has a grasping structure on the side of the abdomen. Shore bugs occur mostly in temperate regions (somewhat unusual for heteropterans) and may occur at fairly high elevations or quite far north (e.g., Alaska); a few are intertidal or found in salt marshes.

Aepophilidae and Omaniidae are wholly intertidal. Indeed, the single aepophilid species lives below the water, breathing (like Aphelocheiridae) via plastron respiration with an air bubble permanently trapped in a dense mat of hydrofuge hairs. Omaniids seek prey on exposed rocks at low tide; at high tide they wait in crevices, probably tapping air bubbles for oxygen. Members of both families are very small (2 and 1–2 mm long, respectively), and presumably feed on even smaller intertidal invertebrates. Aepophilids have quite small eyes, as do aphelocheirids, suggesting that a life permanently under erratically moving water is not sight based.

Aepophilus bonnairei occurs in Europe south to coastal North Africa. Omaniids are known from the Red Sea, south to Aldabra on the African coast, and east to Samoa; it seems likely that they occur on the Indian Ocean coast, and on coasts to the east.

Species of Leptopodidae are about the same size as saldids, ranging from about 2 to 7 mm in length. These bugs vary in shape, and males have a grasping organ similar to but different from that of saldids. Most species are tropical, and many occur near water, but others are in quite dry habitats. Some have been collected in ant lion pits, which suggests they may be scavengers.

Cimicomorpha This is by far largest of the heteropteran infraorders, containing as it does Miridae and Reduviidae, the two largest families. Within the infraorder's 14 families is the widest variety of foods and of habitats: primitively predaceous on other arthropods, three major groups have moved to feeding on vertebrate blood, and others to feeding on plants. Habitats of various cimicomorphans range from human homes and caves to the webs of spiders, embiids, and psocids; one reduviid "fishes" for termites and another is "led" to its food by ants.

The group has long been recognized (before the current classification of Heteroptera) but is held together only loosely by shared advanced character states, not all of which are

possessed by all members. The head *usually* bears some long-socketed hairs (trichobothria) and is *usually* prominent and extends directly forward; the forewing's membrane *usually* has several closed cells, the claws *usually* lack pads between them; and the eggs *often* have a characteristic type and arrangement of microstructures. However the sperm storage organ of other heteropterans (the spermatheca) is in cimicomorphs nonfunctional or either greatly reduced or absent. The cimicomorphan families are united by this last character. Moreover, since each family is more or less phylogenetically related to another, the first and last are joined less by their relationship to each other and more by their relationship one at time to those in between (rather like a group of people in a line holding hands).

With its 6700 species, Reduviidae is the second largest of the Heteropteran families; it is also perhaps the most diverse, as is suggested by the 25, or 22, or 21, or 29, or 23 subfamilies—and the confusion over the number! Suggested by the number is the systematic confusion at the higher taxonomic levels, a situation similar to that in the Pentatomidae (another large family) and—until recently—in the Lygaeidae (yet another).

Reduviids, also called assassin bugs, for the fierce way some attack prey, vary greatly in shape and in size, ranging in length from 3–4 mm to 40+ mm. All (or most) have a short, stubby, slightly curved beak, designed for being stabbed downward through the cuticle of arthropod prey; well-developed eyes (sight-orienting predators); forefemora often enlarged and with spines for prey capture; forewing membrane with two closed cells; glandular setae and a spongy pad on the foretibiae, presumably to help in grasping; paired first-abdominal glands (Brindley's glands) whose secretion apparently wards off enemies (but, then, what is the function of the metathoracic scent glands?); and a stridulitrum on the prosternum whose plectrum is on the beak, also presumably to frighten would-be vertebrate predators. Not all reduviids—even not all reduviid subfamilies and tribes—have all these characters, but most species are two to four times longer than wide, with a large head, bulging eyes, and a stout beak. Most assassin bugs are easily recognized.

All reduviids are predaceous, the great majority on other arthropods. There is some degree of specialization: Ectrichodiinae (645 species) on millipedes, Peiratinae (47 species) on hard-bodied prey like grasshoppers and beetles, Harpactocorinae (2060 species) on soft-bodied prey like grubs and caterpillars, Reduviinae (980 species) on social insects, Emesinae (920 species) on flies, and other subfamilies with usually lesser degrees of specialization (but again it must be emphasized: much generalization here is based on scant data).

One group, the subfamily Triatominae, feeds upon the blood of vertebrates (mostly warm-blooded vertebrates). These are the only bugs responsible for a human disease: Chagas' disease is a serious trypanosomiasis in the neotropics, costing many deaths every year. Triatomines that live near or in human dwellings are vectors of the protozoan parasite *Trypanosoma cruzi.* An intense cooperative effort by Latin American countries and several other national and international agencies is gradually reducing the range of this disease.

Triatomines are large (50–20 mm long, 5–8 mm wide) bugs, usually brown or patterned light and dark brown. They occur throughout the New World tropics and subtropics (including well into the United States) and the "wild" (sylvatic) ones feed on small mammals (often burrowing rodents) and (some species) on birds. Domestic triatomines live in the thatch or in litter on the floor of dwellings. One genus *(Linshcosteus)* occurs only in India; a species complex of *Triatoma* occurs in Malaysia, Indonesia, New Guinea, and surrounding areas; *T. rubrofasciata* is pantropical (and may be the ancestor of the species in the second group). The possibility that Triatominae is paraphyletic (or polyphyletic)—that its members are descended from several different reduviid groups and independently evolved blood feeding—is being investigated.

Certain reduviid species and smaller groups feed rather strangely: An Indian reduviine *(Acanthaspis siva)* hunts honey bees at their nests, and some neotropical apiomerines, which wait near resin sites for certain bees to arrive to harvest nest-constructing resin, may even release a kairomone to attract the bees, which are eaten; fossil evidence indicates that the latter relationship goes back at least 25 million years. The Indian harpactocorine *Lophocephala guerini* feeds on the juices emanating from cow feces (or perhaps on arthropods found within?) and is led to this food by ants, which in turn feed upon the bugs' feces. Many emesines live unentangled in spider or psocid webs, and they feed on their hosts or on prey trapped therein. *Salyavata variegata* (Salyavatinae, neotropical) waits by crevices in termite nests and disguises itself with bits of the nest itself; it eats termites that emerge to repair the crevice and even uses the remains of fed-upon termites as bait to lure more termites out.

Most reduviids actively seek prey and seize it. Some wait where prey is sure to be found (e.g., termite or bee nests); others wait more patiently for less specific prey. Once seized, prey may be fed upon at once or, often, dragged to a secluded spot. Nymphal reduviids may engage in communal feeding: several bugs feeding on a single prey. This occurs in other predaceous heteropterans too, and probably the pooling of saliva with its digestive enzymes is of benefit to all (except the prey). Several reduviids resemble other heteropterans (e.g., alydids, pyrrhocorids) and may be associated with them. Some have thought this a ruse to lure these others to their doom (see earlier comments on this "aggressive mimicry").

Reduviids are abundant in most habitats. They are large and presumably eat a lot. They are aggressive predators. It is therefore surprising that their possibilities as biological control agents have been so little studied (except—recently—in southern India, as described by Ambrose).

Phymatidae is a family often included in Reduviidae, and as often treated separately. Also known as ambush bugs, phymatids lie in wait for their prey, grabbing them with ferociously enlarged forelegs. In the United States, a common ambush bug is yellowish and lurks in late-summer yellow

flowers, waiting for hapless bees and other insects seeking nectar where "rosae … sera moror" (to twist Horace a bit).

The small family Pachynomidae contains 15 described species of tropical predators. They look like certain nabids and share characters with several different cimicomorphan groups; recently they have been placed as the sibling group of Reduviidae. They and another family, Velocipedidae, are the only cimicomorphans with ventral abdominal trichobothria (long-socketed hairs, regularly arranged), a feature found more commonly in Pentatomomorpha. Pachynomids range from 3 to 15 mm in length. Nothing is known of their biology—not even the immatures (that they are predaceous is inferred from the thick curved beak).

The family Miridae [Capsidae (Britain), plant bugs], with its 10,000 species, contains nearly one-third of all heteropterans. So vast is the family and so diverse in habits and habitats, that it cannot be covered here. Luckily, Wheeler's excellent book on the family's biology is available.

All but one group of mirids lack ocelli (hence the German name *Blindwanzen,* "blind bugs"); all are relatively small (1.5–15 mm long), delicate in appearance, usually brown or greenish but often brightly and contrastingly colored, and range in shape from elongated to round to myrmecomorphic. Many (most?) have setae on their surfaces, a "break" in the forewing where leathery part (corium) and membranous part meet, one or two closed membranal cells, and asymmetrical genitalia in the male.

The primitive groups are predaceous, reflecting the predacity of their infraorder (Cimicomorpha). Many mirids are herbivorous, however, although even many of these have reverted to predacity ("secondarily predaceous"). Overall, the percentage of predaceous (and scavenging) mirids is quite high; and many others are omnivorous, supplementing a plant diet with arthropod prey, or vice versa. Moreover, many mirids are cannibalistic. In general, then, a great many mirids are opportunistic feeders and, therefore, may be crop pests under one set of circumstances and useful control agents (of weeds or pests) under other circumstances. However, mirids are small and feed on small prey (eggs, small arthropods). Thus they are rarely seen feeding, and their impact on natural ecosystems and on agroecosystems, although great, is much underestimated (this of course is true of all small predaceous arthropods).

Many mirids are somewhat host specific, at least at the bug-species to plant-genus/tribe level. The bugs feed, usually intracellularly, on the reproductive parts of plants and on new, still-growing somatic tissues. The bugs' enzymes cause the plant to mobilize nitrogen to these feeding wounds, which the bugs suck up. Many mirids feed on annual plants, and often on plants that appear only briefly (e.g., early successional plants); the brief presence of such plants, and their great diversity, may help explain the diversity of mirids: this type of association would seem to promote speciation.

Some of the predaceous mirids also appear to be host specific, but the specificity would seem to be more one of habitat than of prey: associated with a particular habitat (rhododendron plants, tree bark), the mirids perforce feed on prey found only there (cf. Anthocoridae). Members of one subfamily, Cylapinae, occur with fungi, and although they may feed on arthropods also found there, at least one species feeds directly on the fungus.

Because of their numbers and diversity, mirids occur throughout the world in all manner of habitats (except aquatic, but some are found in salt marshes). Probably a majority of plant species is fed upon by mirids, and their populations may become high locally; mirids are therefore often the most frequently collected of heteropterans.

As mentioned, the family is very large; one genus *(Phytocoris)* contains more than 600 species. Eight subfamilies are accepted (but not universally), but it is not surprising that there is less agreement about the number and validity of tribes. Much work is needed on cladistic analyses and higher group systematics of Miridae. And, because of the variety in food preferences, in habitats, in degrees of food and habitat specificity, and in other aspects of their biology, biological and ecological data should be used in these analyses.

Related to the mirids are the Tingidae, known as lace bugs because of the elaborate expansions of the thoraces (and sometimes heads and abdomens) of many of these bugs. Although small (1.5–10 mm long), these are among the loveliest of Heteroptera; some vaguely resemble large coleorrhynchans, but there is no phylogenetic relationship.

Because of their often bizarre shape and strange (and attractive) outgrowths, it is not easy to characterize lace bugs. The head is small or of moderate size, the body (when shorn of outgrowths) usually oval. In most tingids the expansions of the thorax conceal the scutellum and, in some, a pronotal "hood" may partly cover the head; some species are coleopteroid (forewings completely leathery or hard, and closely appressed; the bug looks beetlelike).

The family, with nearly 2000 species, is distributed worldwide; all tingids are herbivorous, even members of a small subfamily that live in ants' nests and (probably) feed on rootlets that penetrate therein. As a group, lace bugs prefer woody plants (but there are many exceptions) and are quite host specific, either at the plant-specific or -generic level. Species in one genus *(Acalypta)* are among the few heteropterans that feed on mosses; and members of two other genera (*Copium* and *Paracopium*) are the only heteropterans that induce gall formation in their host plants. Tingids' preference for woody (perennial) plants sets them apart from mirids (which seem to prefer annuals) and perhaps partly explains why there are so many more of the latter.

This feeding preference of lace bugs also explains why so many are pests of ornamental and fruit-bearing shrubs and small trees, where they feed mostly on somatic tissue. Large populations can arise quickly, the buildup aided in part by various defensive mechanisms of the nymphs. These range from maternal care through several defensive pheromones and kairomones, to an array of sharp or defensive-liquid-producing setae on the body. Another strategy is the laying of

small batches of eggs in different places, a form of bet hedging (this of course occurs in many other groups). These population buildups often lead to major problems localized in time and space.

Maternal care has been carefully worked out in the genus *Gargaphia,* especially *G. solani.* In this species several females may oviposit in a single cluster, which they then communally guard, as well as the resulting nymphs. This opens up the possibility of "cheating": a female that lays eggs in the cluster but does not help guard it can thus devote more time and resources to producing more eggs.

One other family of Cimicormorpha is herbivorous: the 19 species of Thaumastocoridae feed on a variety of plants, but many feed on palms, and the name palm bugs has been given to the family. These are small (2–5 mm long) insects, broadly to narrowly oval, and flattened. The external genitalia are much reduced: the ovipositor is gone in females, one or both parameres are gone in males; the genital capsule is highly asymmetrical. Like tingids, these bugs feed on the cells of plants' somatic tissues (usually leaves), leaving behind pale spots. Populations occasionally build upon ornamental plants to the point of visible damage. One such culprit is the royal palm bug, *Xylocoris luteolus,* which may at times cause unsightly damage to a decorative palm.

The approximately 400 species of Nabidae are about 8 to 12 mm in length and most are slight in appearance, which perhaps accounts for the derivation of their name, damsel bugs. All are predaceous, but some will probe plants, probably for water (none survive on plant food alone). The majority of species seem to prefer seeking prey on low vegetation or in fields, and for this reason nabids are an important component of agroecosystems. Nabids also fly readily and disperse well. As a result, several species are cosmopolitan. One genus *(Arachnocoris)* lives in spiderwebs, probably stealing prey.

Most nabids are brown or light brown, although a few are black and red. The males of most species have a specialized group of tibal setae which, upon being rubbed across part of the genital capsule, spread an attractant pheromone produced in rectal glands.

In many species copulation is normal, but in some a form of internal traumatic insemination (cf. Cimicidae) occurs, wherein the male aedeagus pierces the wall of the female's genital chamber and deposits sperm in her hemocoel.

Two small cimicomorph families, Medocostidae (one species) and Velocipedidae (discussed earlier), have been placed in Nabidae from time to time. The arguments for excluding them are slightly more persuasive than the arguments for including them (more persuasive for the first than the second); and the evidence for their phylogenetic relationships is equally insecure.

The single medocostid species is about 9 mm length, elongated oval, and presumably predaceous. It has an unusually long fourth rostal (beak) segment and lives under bark in western and central Africa.

Three other small families are Plokiophilidae (a half-dozen or so species), Microphysidae (25–30 species), and Joppeicidae (one species). The first appears to be related to Anthocoridae–Cimicidae, but the relationships of the other two are not clear, their affinities with other groups consisting mostly of character-state losses. All are small (1–3 mm long) and predaceous. Of these, the most interesting are the plokiophilid bugs, which live in webs, one species in those of embiids, where they feed on eggs or dead embiids. Other species live in spiderwebs (like some nabids), where they feed alongside the spider on trapped prey. Microphysids are found on tree trunks, and the joppeicid in dry situations.

Closely related to one another are Polyctenidae (bat bugs) and Cimicidae (bed bugs, although most—like all polyctenids—are ectoparasites of bats). Both groups are well adapted as ectoparasites of vertebrates: They are wingless and flattened, the latter an adaptation both for slipping through the hair or feathers of a host and for expansion of the body during engorgement. Cimicids are temporary ectoparasites, moving onto the host only to feed.

As noted earlier, it is cimicids, the bed bugs, that have given their name to heteropterans: true bugs. The 100 or so species are temporarily ectoparasitic on bats, birds (mostly those that roost in groups), and humans. Of the human bed bugs, *C. lectularius* (worldwide, especially in cooler regions) and *C. hemipterus* (tropics), the former is among the very few arthropods associated *only* with humans (like the house fly and the head/body louse). *C. hemipterus,* the tropical bed bug, may also be found on bats, but it seems to prefer humans.

These human bed bugs are small (5 mm long), round to broadly oval, flat (except when engorged with blood), and brown; they are well known, to the extent that although no one claims to have seen one, everyone knows someone who has. Because bed bugs are secretive and feed at night, they are not uncommon but are rarely seen, spending the day hidden in crevices in dwellings and in bedclothes. Despite their ubiquity, they usually do little harm. They spread no pathogen (although they have been implicated in transmission of hepatitis B virus in southern Africa, the evidence offered is at best equivocal), and for the most part they are annoying rather than harmful.

Cimicids that feed on birds can cause harm: some feed on poultry and may take enough blood to decrease egg and meat production; and swallow bugs (genus *Oecacious*) may cause serious damage to populations of these attractive and useful birds.

It is in the human bed bug that traumatic insemination has been studied in the most detail. The male punctures the venter of the female's abdomen with his scimitar-shaped left clasper. Sperm are deposited either directly into the hemocoel (more primitive cimicids, some noncimicids) or into a patch of tissue specialized for their reception. In either case, the sperm then travel (for part of their journey, from cell to cell) to the base of the ovarioles, where they are stored; this journey cannot occur if the sperm are not activated by an agent in the seminal fluid. Eggs are fertilized as they pass the storage depot. This process has not been worked out for the

majority of traumatically inseminated species but is probably the same in broad outline.

Unlike bed bugs, polyctenids live permanently on their bat hosts and are so well adapted to this way of life that they closely resemble not cimicids but bat flies (Streblidae, Nycteribiidae). They are eyeless and have claws and stiff setae "designed" as in other ectoparasites to catch on fur and prevent them from being captured and dragged away by the host. Bat bugs have also carried further the traumatic insemination of their relatives: the female is mated before molting to adult, through the base of the right midcoxa into the hemocoel, whence the sperm move close to the oviducts. Here (apparently) fertilization takes place, and here too occurs not only embryogenesis but hatching. Nymphs are then "born," and because some of these births may occur before the female is adult, they represent a form of pedogenesis (i.e., reproduction by immature stages). The embryos are nourished in part directly from the oviduct wall, and the entire process has therefore been termed pseudoplacental viviparity, a phenomenon that occurs in other groups of insects (cf. the tsetse fly). (It is not clear that mating *must* occur before molting; presumably an unmated adult female can be mated.)

Bat bugs are 3 to 5 mm in length and occur in the tropics. As a group they show some host specificity, and it would be interesting to compare their host associations with those of streblids and nycteribiids.

Anthocoridae contains 500 to 600 species and here includes the subfamilies Lyctocorinae and Lasiochilinae, sometimes treated as separate families. Anthocorids are sometimes called minute pirate bugs, perhaps because they are predaceous and some of the more common ones are picturesquely black and white (like the Jolly Roger). Anthocorids' eyes are prominent, suggesting they orient to prey visually "up close"; some evidence indicates they are initially attracted by the odor of the leaves on which prey are feeding. The left paramere of the male is in many anthocorids modified for traumatic insemination of the female.

Although small (1.5–6 mm long), anthocorids are efficient predators on small arthropods and upon the eggs and early instars of larger arthropods. They are also abundant in many habitats, and for these reasons are quite useful in biological control. In Europe especially, anthocorids are used for control of pests in such confined growing areas as greenhouses. Some members of the Lyctocorinae live in dry areas and have proven useful in the control of stored-food pests.

Many anthocorid species seem to be habitat specific. One impressive example is *Elatophilus,* whose species are found on the trunks of north temperate pines, where they feed on *Matsucoccus* scale insects; the specificity appears to be one of habitat, not prey. Other anthocorids live on the trunks of other trees, feeding there on small homopterans and sometimes bark beetles.

Also known as flower bugs (again perhaps for their daintily contrasting colors), a few anthocorids—mostly of the large and widespread genus *Orius*—live on annual plants; but this is not the most important habitat of anthocorids. Most species (as far as is known: not very far) live on tree trunks, some in leaf and other ground litter and the bases of grass clumps, and some in natural accumulations of seed or berries (and in the nests of rodents that have carried these foods in—perhaps evolutionarily en route to living where grain is stored). Because of the considerable value of these bugs in biological control, the literature on their biology is vast. Unfortunately, the literature is concerned with relatively few species, and the biologies of most are wholly unknown. It is known that some species (perhaps many?) supplement their diet with pollen; and at least one may be entirely herbivorous.

Pentatomomorpha This is the second largest of the heteropteran infraorders. Unlike the others, it (and Aradomorpha) is primitively herbivorous: that is, the ancestor, or the earliest members, of the infraorder, were themselves almost certainly feeders on plant juices. With few exceptions, members of Pentatomomorpha feed on plants. The exceptions are certain Rhyparochromidae that feed on vertebrate blood; members of Geocoridae, which prey on small arthropods as well as plants; a very few Pyrrhocoridae that may be either obligate or facultative predators (it is not always clear which); and Pentatomidae of the subfamily Asopinae, all of whose members are predaceous. For the most part, pentatomomorphs feed on the nitrogen-rich reproductive parts of plants: flowers, ovules, and ripe and unripe seeds. Many, however, feed on plant somatic tissues.

Pentatomomorphs range from the very small (some Lygaeoidea) to the largest of the land bugs. The beak is long, extending usually at least to the mesosternum and sometimes beyond the abdomen; the abdomen's venter bears a regular number of regularly arranged socketed setae. Associated with the metathoracic scent gland opening is one or more raised structures, and the shape is rounded to narrowly oval (rarely long and thin).

Pentatomomorpha contains five superfamilies—six if one includes Aradoidea. (The recent elevation of this group to infraorder level has not met with universal acceptance, and in this article "Aradoidea" is treated as the infraorder Aradomorpha).

The Idiostoloidea contains two families, Idiostolidae (four species) and Henicoridae (one species); the latter was until recently considered to be a subfamily of Lygaeidae (in the broad sense, as discussed shortly). In some features (e.g., abundance of trichobothria), Idiostoloidea appears to be the most primitive of the pentatomomorphan superfamilies; however, in Henry's analysis, this group is the sister group of the families formerly included in Lygaeidae. The biology of Idiostoloidea is not known, although specimens have been captured on the ground, not up on plants. Idiostolids are associated with the moss and litter of southern beech *(Nothofagus)* stands, and may indeed feed on mosses. The superfamily has a Gondwana distribution occurring in southern South America and in Australia (see also Coleorrhyncha, earlier).

Until recently most members of the superfamily Lygaeoidea were classified in the family Lygaeidae. However, it had long been recognized that the members of this family were not all closely related and that indeed Lygaeidae was paraphyletic and, probably, polyphyletic: that is, Lygaeidae did not include all descendents from the common ancestor, and indeed different members of Lygaeidae were descended from several different ancestors. In 1997 T. J. Henry applied cladistic analysis to Lygaeidae, as well as to related families, and presented a new classification of the members of the former Lygaeidae. This classification, which has been generally (but not universally) accepted, is used here.

The family Lygaeidae (*sensu* Henry) now consists of 15 families: Artheneidae, Blissidae, Cryptorhamphidae, Cymidae, Geocoridae, Heterogastridae, Lygaeidae *(sensu strictu),* Ninidae, Oxycarenidae, Pachygronthidae, Rhyparochromidae, Berytidae, Colobathristidae, Malcidae, and Piesmatidae. The last four had already been treated as separate families in the old Lygaeoidea. The superfamily contains roughly 4400 species, most of them in the former "Lygaeidae."

Lygaeoids are relatively small heteropterans (1–20 mm long); the largest are the milkweed bugs, several of which are common in North America. Milkweed bugs are Lygaeidae *(sensu strictu),* and many are warningly colored because they are distasteful. Of greater importance are the chinch bugs (Blissidae); all feed on grasses, and several are serious pests of graminaceous crops (wheat, sugarcane, rangeland and lawn grasses, etc.). Several other families are of some economic importance (Geocoridae, one of the few facultatively predaceous groups in Pentatomomorpha, in biological control; some Colobathristidae on sugarcane, Fig. 2).

Members of the largest lygaeoid family, Rhyparochromidae, with nearly half the superfamily's species, feed on fallen seeds; for this reason they are called seed bugs, a name that by extension is often applied to lygaeoids generally. Members of other lygaeoid families feed on seeds still attached to plants, and others feed directly on somatic tissues (mostly leaves, sometimes in vascular tissue, sometimes in plant cells).

Members of the rhyparochromine tribe Cleradini, and one species in the tribe Udeocorini, feed on vertebrate blood, and it has been suggested they might be able to vector *T. cruzi,* the pathogen of Chagas disease (see Reduviidae).

Of the five families in the superfamily Coreoidea (Coreidae, Alydidae, Rhopalidae, Stenocephalidae, and Hyocephalidae), the first is by far the largest in number of species (1800–1900) and in size; indeed, members of the neotropical coreid genus *Thasus* may be the largest land heteropterans known (Fig. 3); up to 40 mm in length, being exceeded in this respect only by some giant water bugs (Belostomatidae). The other families are smaller in numbers and size: Alydidae (about 250 species), Rhopalidae (somewhat more than 200), Stenocephalidae (about 30), and Hyocephalidae (3). All are phytophagous.

Members of Coreidae are sometimes called leaf-footed bugs because several species, occurring in several apparently unrelated tribes, have hind tibial expansions (and sometimes antennal expansions) as adults, as nymphs, or as both; their function is obscure, although they may serve to deflect the attacks of birds. Males of several groups of coreids have enlarged hind femora, which are used to defend territories from other males. The nymphs of several species are gregarious, and it would be interesting to learn whether this gregariousness occurs in species whose males are territorial. The coreid head is small relative to the body (cf. Alydidae), and the bugs' length ranges from about 6 to 40 mm.

Members of the family feed up on plants, some species on ripening seeds and other reproductive parts, and many species on the juices from vascular tissue. Varying degrees of host specialization occur in the family, from mono- through oligo- to polyphagy. Some species feed on cucurbits (hence another common name, squash bugs). Some unrelated coreid groups (Pseudophloeinae, and three coreine tribes) are among the few heteropteran family groups to have broached the defenses of the Leguminosae and to have radiated upon these plants. Several of these bugs are serious pests of legume crops (e.g., pigeon pea) in the Old World tropics.

FIGURE 2 A lygaeid, Wekiu bug, *Nysius wekiuicola,* feeding on Calliphoridae. (Photograph by Peter Oboyski.)

FIGURE 3 Mesquite bug (adult), *Thasus neocalifornicus.* (Photograph by John H. Acorn.)

FIGURE 4 Broad-headed bug (nymph), probably *Megalotomus quinquespinosus.* Note ant mimicry. (Photograph by John H. Acorn.)

FIGURE 5 Box elder bugs (adult and nymphs), *Boisea trivittatus.* (Photograph by John H. Acorn.)

The males of one species, *Phyllomorpha laciniata,* are heavily spined and, somewhat like some giant water bug males, carry among the dorsal spines their females' eggs.

Alydidae—called broad-headed bugs because the head is relatively wide (it also resembles an ant's)—comprises three groups. The subfamily Alydinae is another group specializing on legumes, upon whose ripening seeds the bugs feed. Several species are pests of legume crops (e.g., soybean, pigeon pea) in the tropics. Immature alydines mimic ants, often successfully enough to fool heteropterists; some species mimic small species of ants as early instars and other larger ant species when older (Fig. 4).

Some members of the second alydid group (Micrelytrinae: Micrelytrini) mimic ants both as adults and as nymphs. The biology of this pantropical tribe is almost completely unknown, although one species may become abundant on range grasses in northern South America. It has been speculated that many micrelytrines are grass feeders (on grass seeds?), as are members of the other tribe (Leptocorisini) of Micrelytrinae.

All members of Leptocorisini feed on grasses, as far as is known. Some species (genera *Leptocorisa, Stenocoris*) are elongated and are often serious pests on rice in the Old World. Leptocorisines feed on the rice panicles and, when rice of the appropriate stage is not available, rice bugs feed on wild grasses found nearby. These grasses thus serve as a reservoir for the rice pests; some leptocorisines prefer these grasses to rice, however.

Members of Rhopalidae are called scentless plant bugs because all are herbivorous and the external opening of the metathoracic scent gland is small and placed more ventrally (hence harder to see) than are the openings of other bugs. Rhopalids are not scentless, however, and indeed the scents of a few species have been analyzed.

The subfamily Serinethinae (60–65 species, mostly tropical and subtropical) includes the box elder bug (Fig. 5), which in North America sometimes seeks warmth in houses in late fall. Serinethines feed on sapindalian plants and occasionally become minor pests on these ornamental plants. The 150 or so members of Rhopalinae are more drab than the red and black Serinethinae, and far more of them occur in the Holarctic region. They feed on a variety of plants and rarely become pests even locally. At least one species has helped biologically control a weed, velvetleaf, in the United States.

Stenocephalidae occurs in the Old World tropics, with a few members in the Palearctic region; the single species on the Galápagos Islands may have been introduced by ship from Africa. Stenocephalids feed on euphorbs, as far as is known. The family is of interest because it possesses certain features of Coreoidea (many) and of Lygaeoidea (ovipositor type, egg type); the phylogenetic significance is unclear.

The three species of Hyocephalidae are all Australian. The family is related to Stenocephalidae, and both families seem to be primitive in the Coreoidea. Little is known of hyocephalid biology except that some occur under stones, where they apparently feed on fallen seeds. Both stenocephalids and hyocephalids are relatively large (8–15 mm long) and slender.

Pentatomoidea is a well-defined superfamily, although there is some disagreement (not to say confusion) about the higher classification of the groups within: some heteropterists treat some groups as tribes, subfamilies, or even families, and others treat them differently. The taxonomic limits of Pentatomoidea are clear and agreed upon; and there is general agreement that the various tribes, and so on are indeed worthy of suprageneric rank: the level of that rank is sometimes argued. A forthcoming catalog of the largest family, Pentatomidae (by D. A. Rider), will help settle some of these questions.

In general, pentatomoids are larger than many other bugs. Most (but not all) have five-segmented antennae (hence the name); most other heteropterans have four. Most are also wider than many other heteropterans, and, ranging in length from about 4 to 30 mm, many pentatomoids appear somewhat "squat." Quite a few are very brightly colored. All except one subfamily (Pentatomidae: Asopinae) are herbivorous; the

FIGURE 6 Adult predaceous pentatomids *(Eocanthecona furcellata)* feeding collectively on a caterpillar *(Eupterote mollifera)* in southern India.

FIGURE 7 *Coleotichus blackburniae* (Scutelleridae). (Photograph by Peter Oboyski.)

superfamily contains general feeders and quite host-specific ones. Within both groups are some serious pests of crops, and asopines have shown some success in biological control.

The largest family, Pentatomidae, whose approximately 4500 species make it slightly larger than Lygaeidae (in the old sense, as discussed earlier), occurs abundantly throughout the world. One species, the southern green stinkbug, *Nezara viridula,* is cosmopolitan, feeds on just about anything green, is a major pest in some areas and a minor one in many others, and in Asia has several striking color varieties. This species has been so widespread for so long that its place of origin remained unknown until recently: it probably arose in or near Ethiopia.

Pentatomids are for the most part relatively large, ranging in length from 8 to 20 mm. They are also herbivorous, with the exception of one subfamily (Fig. 6). Most of the few pentatomids whose feeding habits are known are polyphagous; some feed more narrowly. Members of one group, related to *Aelia,* feed on grasses; and several species of *Aelia* itself are serious pests of small-grain crops (Sunn pests, as discussed shortly in connection with "Scutelleridae"). Pentatomids are among the most serious pests of soybean worldwide: when a new area opens up to soybean production (e.g., southern Brazil several years ago, central Brazil a few years ago), local legume-feeding or polyphagous pentatomid species rapidly become soybean pests. Because of the family's size and the number of crops its members feed on, Pentatomidae is one of the most important heteropteran families.

The subfamily Asopinae is predaceous, a fact easily recognized by its members' short stubby beak, well designed for stabbing into prey. Asopines are useful in biological control, and one *(Perillus bioculatus)* specializes on the Colorado potato beetle and related chrysomelids; the red of this bug is derived from ingested pigment of the beetle prey.

Many pentatomids are green or (more often) brown, perhaps as camouflage on plants and ground (some feed up on plants and hibernate or estivate on the ground). However, quite a few are brightly and contrastingly colored and are presumably aposematic. Many pentatomids live on tree trunks and indeed may be the most common group of herbivorous bugs to use this habitat; it is not clear what they feed on, perhaps mosses and other epiphytes and perhaps (occasionally) these bugs can penetrate through to the tree's cambium.

Most pentatomids are somewhat broad and flattened, but some (like the grass feeders) are more or less elongated. Some members of Podopinae are almost globose and may bear thornlike spines dorsally (see also Cyrtocoridae).

Various degrees of parental care have arisen in this family (or have been retained from parentally caring ancestors; the point is controversial) and in other pentatomoid families (and elsewhere, rarely). For the most part, this care protects eggs and early instars from parasitism and predation. In many species, however, this effort denies the female an opportunity to feed and, as a result, she produces fewer eggs or batches of eggs than her less caring sisters. The significance of this trade-off remains to be worked out for most species.

Scutelleridae contains 400 to 500 species in which the scutellum is greatly extended to cover the abdomen (hence the common name, shield bugs); these bugs range from 5 to 20 mm in length and are somewhat more globose than pentatomids (although some of the more spectacular ones are broadly elongate). Several species are brilliantly colored (one genus is *Chrysocoris,* or golden bug), either in solid sometimes iridescent colors (blues, violets, reds) (Fig. 7) or in bold patterns of spots and stripes. These are among the most beautiful of heteropterans and, in some cultures, considered to be among the most tasty. Most however are drab brown or tan, the color sometimes patterned.

Despite their relatively large size and often striking appearance, shield bugs remain poorly known biologically, and very poorly known ecologically. The females of a few care for their young, as do those of a few other pentatomoid families. All scutellerids are herbivorous, and a few are pests of several crops. One group is particularly important.

Several species of *Eurygaster* are very serious pests of small-grain crops in a broad swath from eastern Europe through the Middle East (where they are the most serious) into eastern Asia. In the spring, these Sunn pests move from upland wild plants down to the fields with the young grain;

there they feed first on the young shoots and then on seeds. In late summer or early fall, they move back into the hills and mountains, sometimes undertaking a trip of many miles. Crops may be completely destroyed over a very large expanse.

The family is represented throughout the world, but members of each of the four subfamilies are concentrated in different continents, with a few representatives in one or two other continents (the historical biogeography of the family is needed; the pretty ones are all tropical or subtropical).

Acanthosomatids (Acanthosomatidae; there is no common name) look much like pentatomids, but are somewhat larger (7–20 mm long) and more elongated. Their tarsi have only two segments, and the bugs have several other advanced features. The latest revision (1974) recognizes about 45 genera; there are about 200 species, although no one has listed or cataloged them. As far as is known, acanthosomatids feed on woody plants (trees and shrubs). Maternal care by the females of several species has been thoroughly studied; as in Pentatomidae, protection is provided eggs and early instars against predators (chiefly ants) and parasites; the body-jerking and wing-fanning movements are so strong that small predators may actually be hurled from the leaf.

The 250 species of Tessaratomidae are in general larger (15–26 mm long), and paleotropical (a few species in neotropics). The second abdominal spiracle is exposed, and a projection from the mesosternum extends anteriorly sometimes beyond the tip of the beak. One tessaratomid, the bronze orange bug *(Musgraveia sulciventris),* is a frequent pest on citrus in Australia.

Possibly related to this family is Dinidoridae, whose 85 to 90 members are also large (10–30 mm long) and robust, and whose distribution is roughly that of Tessaratomidae. Here too the second abdominal spiracle is exposed (although this may in both families be plesiomorphic); and often the pro- and mesosternum bear a midline groove. Dinidorids in general prefer cucurbits as food, and one species, the red pumpkin bug, *Aspongopus* (formerly *Coridius*) *janus,* is a pest in India on cucurbit crops.

Urostylinae and Saileriolinae, the two subfamilies of Urostylidae, share several features but look very different and differ considerably in general appearance and in other morphological features; most urostylids are Asian. Urostylines (80–90 species) are elongated, (4–15 mm long) and look more like coreids than pentatomoids. Saileriolines are small (2.5–3.8 mm long), more pentatomoid in shape, and with the anterior part of the forewing (the corium) less "leathery" than in most heteropterans. Some urostylines occasionally attack ornamental trees; in great numbers they may become a problem. One early (and successful!) control method was the use of gunpowder.

Cydnidae (burrower bugs) is the only heteropteran group most of whose members live below ground, sometimes several feet below the surface. Most cydnids are dark brown or black, 1.5 to 25 mm in length (most in the range of 5–12 mm), and have a wing stridulitrum. Their smooth bodies, flattened heads, and strong forelegs are adapted to digging; so also probably is the coxal "comb" (not a "coxcomb"), an array of stout setae perhaps used for cleaning soil particles from the antennae. The foretibiae of one subfamily (Scaptocorinae) are greatly developed for digging. There are four subfamilies: Sehirinae (60 species), Thyreocorinae (5), Corimelaeninae (200), and Parastrachiinae (2). All live above ground. Females of the Sehirinae and Parastrachiinae care for their eggs and early instars. The Thyreocorinae and Corimelaeninae have sometimes been grouped as a single family; and the two *Parastrachia* species are large and brightly red and black and should certainly be elevated to family rank.

The remaining 350 species are subterranean, feeding on roots. They may at times become serious, although localized, pests of both crops and rangeland grasses.

Cydnidae may be phylogenetically close to the "base" of the Pentatomoidea and may also be related to the group of small families described next.

Within the Pentatomoidea is a number of small families, mostly restricted to the neotropics [Cyrtocoridae (11 species), Canopidae (8), Megarididae (16), Phloeidae (3)] or to Australia [Aphylidae (2 species), and Lestoniidae (2)]. Plataspidae (500 species, a very rough estimate) is primarily paleotropical, but a few species occur in the Palearctic region. Characteristic of many of these families are small size and a rounded globose body, with forewings and/or scutellum enlarged (and often ornamented) to cover the entire abdomen. The cyrtocorid scutellum bears a stout impressive spine. Some (all?) may be related to Cydnidae and to Scutelleridae and as, Schuh and Slater say, the phylogeny and biogeography of these groups need to be worked out. Of phylogenetic importance too is the primitive family Thaumastellidae, whose few members live in southern Africa.

Aradomorpha This infraorder contains two families, Aradidae (1800–2000 species, worldwide) and Termitaphididae (9 species, pantropical). These families were included in Pentatomomorpha, from which Aradomorpha probably evolved.

Aradomorphans are very flattened, and (most striking) their feeding stylets are very long, very narrow, and stored coiled within the head. Termitaphidids are small (2 to 3 mm long), wingless, eyeless, and ovipositorless; they live in termites' nests and probably feed, like aradids, on fungal mycelia.

Most flat bugs (Aradidae) are larger (3–11 mm long) and live under the bark of dead or dying trees, a habitat to which their flat body and shortened legs and antennae fit them; other aradids live in the litter of the forest floor; and a few live in termites' nests, or vertebrates' burrows. All these places are closely confined and thus both temperature and humidity vary little, conditions conducive to the growth of fungi, within whose long tubular mycelia the flat bugs feed. This way of life is unique to aradids, which have successfully exploited it (as witness the many aradid species).

One species is an exception: *Aradus cinnamomeus* (actually, a three-species complex) feeds on the circulatory

tissues of living pines (and occasionally *Larix*), and in central and eastern Europe can become a serious pest. Some other aradids also may feed on tree fluids, but the frequency of such feeding, the number of species that so feed, and the phylogenetic significance of such feeding are all unexplored.

ENVOI

Heteroptera not only comprises the most species of any hemimetabolous group (except perhaps the paraphyletic "Homoptera") but contains the most biological, structural, and ecological diversity. It is a measure of this diversity, this evolutionary versatility, that so many ecological niches are filled by bugs. From the Alaskan cold to the webs of embiids to the bottoms of streams and the surface of the sea, and to the beds of people, heteropterans live where few other insect groups (and no other hemimetabolans) occur. Moreover, a vast number of heteropterans remain to be described, many from unusual habitats (tree canopies, leaf litter); and the biologies of most heteropterans remain to be worked out. The diversity we see, although great, is less than the diversity to be discovered. Why this group of insects should be so diverse cannot be answered here. But that it is so diverse explains the group's great and continuing fascination.

See Also the Following Articles

Aposematic Coloration • Aquatic Habitats • Auchenorrhyncha • Plant–Insect Interactions • Rostrum • Sternorrhyncha

Further Reading

Ambrose, D. P. (1999). "Assassin Bugs." Science Publishers, Enfield, NH.

Andersen, N. M. (1982). "The Semiaquatic Bugs (Hemiptera, Gerromorpha). Phylogeny, Adaptations, Biogeography and Classification." Entomonograph 3. Scandinavian Science Press, Klampenborg, Denmark.

Blatchley, W. S. (1926). "Heteroptera or True Bugs of Eastern North America." Nature Publishing, Indianapolis, IN.

Butler, E. A. (1923). "Biology of the British Hemiptera-Heteroptera." H.F. & G, Witherby, London.

Cohen, A. C. (1998). Solid-to-liquid-feeding: the inside(s) story on extra-oral digestion in predaceous Arthropoda. *Am. Entomol.* **44,** 103–116.

Dolling, W. R. (1991). "The Hemiptera." Oxford University Press, Oxford, U.K.

Henry, T. J. (1997). Phylogenetic analysis of family groups within the infraorder Pentatomomorpha (Hemiptera: Heteroptera), with emphasis on the Lygaeoidea. *Ann. Entomol. Soc. Am.* **90,** 275–301.

Polhemus, J. T. (1996). Aquatic and semi-aquatic Hemiptera. *In* "An Introduction to the Aquatic Insects of North America" (R. W. Merritt and K. W. Cummings, eds.), Chap. 15. Kendall/Hunt, Dubuque, IA.

Schaefer, C. W., and Panizzi, A. R. (2000). "Heteroptera of Economic Importance." CRC Press, Boca Raton, FL.

Schuh, R. T., and Slater, J. A. (1995). "True Bugs of the World (Hemiptera: Heteroptera)." Cornell University Press, Ithaca, NY.

Slater, J. A., and Baranowski, R. M. (1978). "How to Know the True Bugs." Brown, Dubuque, IA.

Weber, H. (1930). "Biologie der Hemipteren: eine Naturgeschichte der Schnabelkerfe." Springer-Verlag, Berlin.

Wheeler, A. G., Jr. (2001). "Biology of the Plant Bugs (Hemiptera: Miridae): Pests, Predators, Opportunists." Cornell University Press, Ithaca, NY.

Protura

Robert T. Allen
Paris, Arkansas

Members of the order Protura are small (usually < 1.0 mm), pale to white arthropods that live in the soil and ground litter debris. Because of these characteristics, the group was not discovered and recognized until 1907, long after almost all the other insect orders had been described and classified. Two notable Italian entomologists, F. Silvestri and A. Berlese, recognized the group at about the same time. In 1907 Silvestri described the order Protura and created the first family, Acerentomidae. This work was quickly followed by the description of a second family, Eosentomidae, and a number of new species by Berlese in 1908. In 1909 a comprehensive monograph on the order was published by Berlese. After almost 100 years of study by many workers, over 600 species have been described and placed in two suborders and eight families (Table I).

Since their discovery and characterization, the Protura have always been recognized as a very ancient group of arthropods that evolved early in the history of the phylum. Early workers placed the group with the Collembola and

TABLE I The Order Protura

Suborder Acerentomoidea
Family Fujientomidae
Family Hesperentomidae
Subfamily Hesperentominae
Subfamily Huhentominae
Family Protentomidae
Subfamily Condellinae
Subfamily Proentominae
Family Acerentomidae
Subfamily Acerentulinae
Subfamily Tuxenentulinae
Subfamily Acerentominae
Family Berberentomidae
Subfamily Proacerelinae
Subfamily Berberentulinae
Subfamily Silvestridinae
Family Acerellidae
Subfamily Alaskaentominae
Subfamily Acerellinae
Subfamily Nipponentominae
Suborder Eosentomoidea
Family Eosentomidae
Subfamily Isoentominae
Subfamily Eosentominae
Subfamily Anisentominae
Subfamily Antelientominae
Family Sinentomidae

Source: After Yin, W. Y. (1984). A new idea on phylogeny of Protura with approach to its origin and systematic position. *Sci. Sin. (B)* **27,** 149–160.

FIGURE 1 Scanning electron microscopy of a proturan (family Acerentomidae) found in subtropical rain forest litter from Lamington National Park, Queensland, Australia. (Photomicrograph courtesy of D. E. Walter.)

Diplura and referred to the three orders as the "pouched head" or entognathous orders because the mouthparts were enclosed by the sides of the head capsule. Most recently, the Collembola and Protura have been removed from this association and placed in the suborder Ellipurata (some workers consider the group to be a class), standing between the class Symphyla and the true insects or hexapod orders.

CLASSIFICATION

The Protura may be readily recognized by the following characters: antennae and eyes absent, head conical, mouthparts enclosed by the sides of the head, first pair of legs with numerous sensory organs, and first three abdominal segments with paired styli (Fig. 1).

COLLECTING AND SPECIMEN PREPARATION

Protura are most readily collected by means of Berlese or Tullgren funnels. Flotation techniques have also been used to extract specimens from soil samples. Specimens may be collected in 70% ethanol and held indefinitely. For identification, specimens must be mounted on microscope slides. In properly mounted specimens, a compound microscope equipped with phase contrast may be used to study characters at 1000 ×, under oil immersion.

IDENTIFICATION

For North American forms, the keys published by Copeland and Imadate in 1990 are indispensable for generic identification. Once the genus has been correctly identified, it is necessary to refer to a number of articles describing individual species in that genus. General works published during the latter half of the 20th century by Tuxen, Nosek, Imadate, and Houston will also be helpful.

BIOLOGY

After insemination by males, the females lay their eggs in the soil or litter. Protura prelarvae hatch in 8 to 12 days under favorable conditions. The abdomen of prelarvae consists of nine segments, and the mouthparts and thoracic and abdominal appendages are not fully developed. The prelarva molts in 2 to 3 days and changes to the larva I stage. During this stage the appendages and setae on the body develop more fully. After molting a second time, larva II adds an abdominal segment (10-segmented abdomen) and the chaetotaxi (specifically arranged hairs on the body) continue to develop. The next state, termed the maturus junior, is characterized by the addition of the last two abdominal segments (12-segmented abdomen), but the genitalia remain undeveloped. The preimago is the fifth-instar and quickly molts to become the imago or adult form. Adults are characterized by the 12-segmented abdomen and fully developed genitalia in both sexes.

It is now known that Protura can be among the most abundant arthropods inhabiting soil and litter. As many as 150 individuals representing several species have been collected in forest soil samples in eastern North America. They may occur in soil to a depth of 0.5 m or more.

The diet of the Protura is uncertain. Some species have been observed feeding on mycorrhizal fungi; other possible food sources are unknown. Other aspects concerning the biology and ecology of these animals have not been studied.

See Also the Following Article

Diplura

Further Reading

Bernard, E. C. (1990). New species, clarifications, and changes in status within *Eosentomon* Berlese (Hexapoda: Protura: Eosentomidae) from the United States. *Proc. Biol. Soc. Wash.* **103,** 861–890.

Copeland, T. P. (1962). A taxonomic treatment of *Eosentomon* Berlese (Protura) of east Tennessee. Ph. D. dissertation, University of Tennessee, Knoxville.

Copeland, T. P. and Imadate, G. (1990). Insecta: Protura. *In* "Soil Biology Guide" (D. I. Dindal, ed.), pp. 914–933. Wiley, New York.

Houston, W. W. K. (1994). Zoological Catalogue of Australia 22, Protura. CSIRO, Information Service, East Melbourne, Victoria 3002, Australia.

Imadate, G. (1974). "Fauna Japan, *Protura.*" Kelgaku Publishing, Tokyo.

Nosek, J. (1973). "The European Protura." Muséum d'Histoire Naturelle, Geneva.

Tuxen, S. L. (1944). "The Protura." Hermann, Paris.

Yin, W. Y. (1984). A new idea on phylogeny of Protura with approach to its origin and systematic position. *Sci. Sin. (B)* **27,** 149–160.

Psocoptera
(Psocids, Booklice)

Edward L. Mockford

Illinois State University

Psocoptera (Corrodentia, Copeognatha) constitutes an order of neopterous, exopterygote insects commonly called psocids, barklice, or booklice. Their closest relatives are the

FIGURE 1 Psocid, *Graphopsocus cruciatus.* (Photograph courtesy of Ken Gray Collection, Entomology Department, Oregon State University.)

Phthiraptera or true lice. Psocids are small and soft-bodied and therefore have received little attention from collectors. Only recently, when several dozens of species have been found in stored products, and the tropical forests have proved to harbor a highly diverse fauna, has greater interest been focused on this group.

EXTERNAL ANATOMY

Adult psocids range from about 1 to 10 mm in body length. Most adults are fully winged with the forewings longer and more complexly veined than their hind wings (Fig. 1). The forewings at rest generally exceed the tip of the abdomen. Antennae are long and slender, in the larger forms tending to be longer than the body. The head is rounded with compound eyes often large and bulging but sometimes greatly reduced. The postclypeus (Fig. 2) is usually swollen to accommodate the well-developed cibarial pump muscles. Often the postclypeus bears pigmented chevron marks (Fig. 2) indicating the attachment points of these muscles. Mouthparts are of the chewing type with large, well-developed mandibles. The maxilla contains an elongate, slender lacinia. The hypopharynx includes structures unique to these insects that are important in maintaining water balance. In fully winged forms the prothorax is frequently reduced while the meso- and metanota are swollen to accommodate the flight muscles. Wing venation tends to be relatively simple. Legs tend to be slender, with the hind femur somewhat swollen in some forms, allowing short hops or initiation of flight by jumping. In the small booklice of the genus *Liposcelis* (Fig. 3), the greatly swollen hind femora are thought to facilitate the ability to change course rapidly and crawl backward. In many winged adult psocids, the hind coxae each bear a rasp and tympanum; the two structures together are called the coxal organ. This is thought to be a stridulatory organ, although as yet no sound has been detected from it.

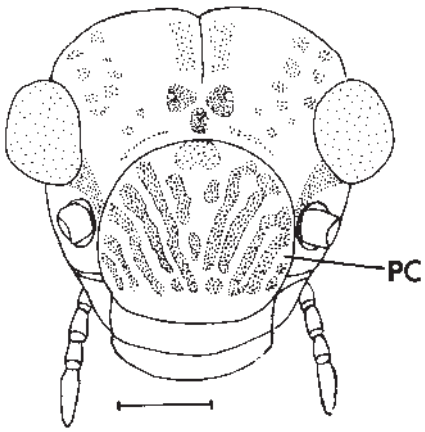

FIGURE 2 Face of a psocid, *Aaroniella badonneli,* showing prominent postclypeus (PC) with chevron marks. Scale, 0.2 mm.

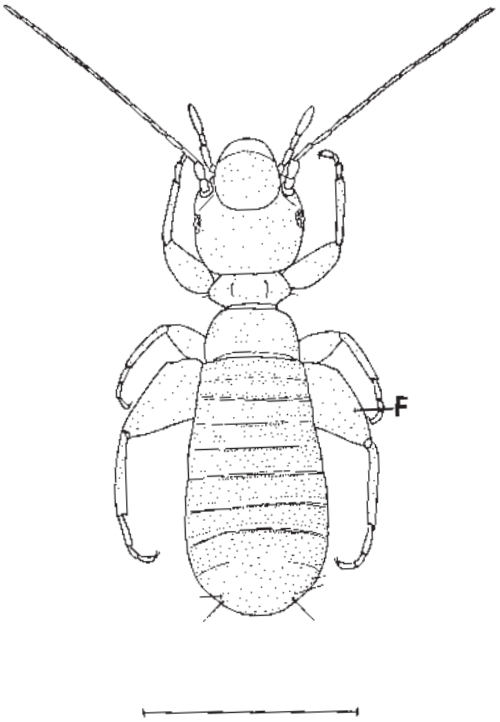

FIGURE 3 Habitus of a booklouse, *Liposcelis corrodens,* showing enlarged hind femur (F) and neotenic features: absence of wings and ocelli and reduced compound eyes. Scale, 0.5 mm.

The abdomen consists of 11 segments. The first 7 are usually membranous, although the terga of the first 2 may be sclerotized and fused together. The clunium, formed by fused segments 8 to 10, is the bearer of external genitalic structures. Distal to the clunium are three semimembranous flaps guarding the anus—the dorsal epiproct and the lateral paraprocts. These may be modified in various ways and probably are involved with copulation.

LIFE CYCLE

Most psocids are oviparous. In various taxa, eggs may be laid singly or in groups and either bare or with a covering of encrusting material, webbing, or both. Eclosion (emergence

of the pronymph from the egg) is facilitated by means of an oviruptor, a blade-like or saw-like structure on the head of the pronymph. By means of a slight rocking motion, imbibing of air, and hemolymphatic pressure on the head, the insect pushes its oviruptor against the thin egg cuticle, creating a slit through which it exits. Immediately it molts again, casting off the pronymphal exuviae (the cast cuticle), which remain partially caught in the exit slit of the egg.

Eggs of some species undergo a winter diapause and in the Northern Hemisphere this is brought about by the mother's perception of decreasing day length in late summer.

Nymphs undergo five to six instars, requiring 4 to 6 weeks to reach maturity. Adults remain teneral for a day or two prior to engaging in sexual activity. Females of some sexual species produce a sex pheromone that, at a short distance (1–2 cm), brings about immediate, rapid search behavior in mature males—involving them running about and vibrating their wings. This reaction may be elicited by a piece of substrate on which an "advertising" female has been standing or even by the tips of a pair of forceps that have just held such a piece of substrate. Behavior after contact is tremendously varied. Frequently a male "antennates" a female, i.e., touches his antennae to hers and then crawls over her back and forces himself under her body from in front. The copulatory structures then grasp together, after which the male may stay under the female or move—180° to face opposite the female (suborder Trogiomorpha) or back on top of the female (family Lachesillidae). Copulation lasts only a few seconds in many groups of the suborder Psocomorpha, although in the genus *Lachesilla* it generally requires some 35 min. In the suborder Trogiomorpha, it may last up to 2 h during which individual sperm pass through the slender spermathecal duct, then clump together, and become enclosed in "spermatophores" in the spermathecal sac. This is probably the sequence of events in any form in which the spermathecal duct is slender and spermatophores are seen in the spermathecal sac.

Oviposition may start within a day or 2 of copulation. Under favorable conditions, oviposition may continue over a period of 2 months.

EVOLUTION TOWARD NEOTENY

Although most adult psocids are fully winged, a variety of levels of wing reduction are seen throughout the order. Associated with wing reduction are reductions in several other structures: compound eyes, ocelli, paraproctal sensorium, and ctenidia (comb-based setae) of the basal hind tarsomere. The resulting adult appears nymph-like in these characters, and these changes can thus be regarded as neotenic. Several patterns of change are seen, which can be summarized as completely wingless, *Liposcelis* (Fig. 3); with very short scale-like or button-like wings, *Trogium, Cerobassi;* wing development variable in both sexes, *Rhyopsocus, Lachesilla* (some species); with sexual dimorphism in wing development and males wingless or nearly so, females variable in wing development, *Embidopsocus, Psoquilla, Archipsocus;* sexual dimorphism and males fully winged, females variable in wing development, *Peripsocus* (some species), *Valenzuela* (some species); and sexual dimophism with males fully winged, females wingless or nearly so, *Camelopsocus, Mesopsocus* (most species), *Reuterella, Lachesilla* (some species; Fig. 4). Except for the first two categories, these reductions have occurred repeatedly in the evolution of the order. Thus, they appear to offer advantages in energy conservation and keeping the organism closely associated with its habitat.

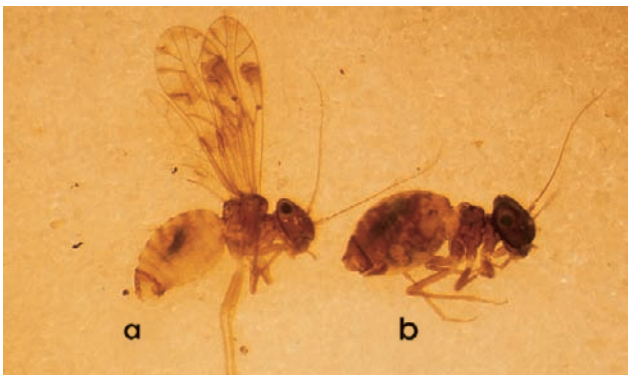

FIGURE 4 Male (a) and female (b) of a psocid, *Lachesilla pallida,* showing extreme sexual dimorphism in wing development.

HABITATS AND FEEDING PREFERENCES

Psocids live in a great variety of habitats, including living leaves, especially of monocotyledonous plants and conifers; dead foliage of all plants, both hanging and in ground litter; trunks and branches of trees and shrubs (open surfaces and under bark); rock surfaces; and human dwellings. In general, psocids feed on algae, fungi, lichens, small eggs of insects, particles of organic debris, and dead bodies of insects. Some seem to be strict alga feeders (some Psocidae, Myopsocidae, Peripsocidae), and it is likely that they are somewhat specific in their choice of algae. There are also some strict lichen feeders, and some specificity has been observed among them. Some of the (micro)fungus feeders are not so specific. In culture, the corticolous lepidopsocid *Echmepteryx hageni* takes lichens, algae, yeasts, and pollen and occasionally nibbles on its own eggs when these are not well concealed. The species of leaf dwellers are primarily microfungus feeders, whereas the species living on open surfaces of tree trunks and branches and on rock outcrops are primarily lichen and alga feeders.

ECONOMIC IMPORTANCE

The booklice, genus *Liposcelis,* are frequently household pests and may be among the causative agents of asthmatic reactions. Of greater monetary importance is their tendency to enter and reproduce rapidly in food storage and food

processing facilities, where they may render the products unfit for human consumption. Control methods emphasize sanitation and reduction of relative humidity.

See Also the Following Article

Phthiraptera

Further Reading

Eertmoed, G. (1978). Embryonic diapause in the psocid *Peripsocus quadrifasciatus:* Photoperiod, temperature, ontogeny and geographic variation. *Physiol. Entomol.* **3,** 197–206.

Glinyanaya, Y. I. (1975). The importance of day length in the regulation of seasonal cycles and diapause in some Psocoptera. *Entomol. Rev.* (transl. of *Entomologicheskoye Obsreniye*) **54,** 10–13.

Mockford, E. L. (1991). Psocids (Psocoptera). *In* "Insect and Mite Pests in Food, an Illustrated Key" (J. R. Gorham, ed.). U.S. Department of Agriculture and U.S. Department of Health and Human Services, Washington, DC.

Mockford, E. L. (1993). "Flora and Fauna Handbook," No. 10, "North American Psocoptera (Insecta)." Sandhill Crane Press, Gainesville, FL/Leiden.

New, T. R. (1987). Biology of the Psocoptera. *Orient. Insects* **21,** 1–109.

Smithers, C. N. (1967). A catalog of the Psocoptera of the world. *Aust. Zool.* **14,** 1–145.

Smithers, C. N., and Lienhard, C. (1992). A revised bibliography of the Psocoptera (Arthropoda: Insecta). *Tech. Rep. Aust. Mus.* **6,** 1–86.

Pterygota

Pterygota is a subclass of the class Insecta of the phylum Arthropoda. It contains two divisions: the Exopterygota and the Endopterygota. Most extant orders of insects are pterygotes.

Puddling Behavior

Scott R. Smedley
Trinity College, Connecticut

Puddling is a behavior in which insects, chiefly adult lepidopterans, drink from mud puddles or moist soil. Although broadly distributed geographically, puddling is particularly spectacular in the tropics where numerous individual butterflies representing multiple species often gather to imbibe from damp ground along river banks. The behavior generally is strongly sex biased, with typically only males participating. Puddling is associated with both the insects' nutritional ecology and their reproductive biology. In certain moths, the behavior is rather herculean, as individuals imbibe several hundred times their body mass in puddle fluid in a single drinking bout!

GENERAL BACKGROUND

Puddling is widely distributed taxonomically among the Lepidoptera, occurring diurnally in butterflies (5 families reported) as well as nocturnally in moths (at least 12 families). Puddling takes place in tropical, temperate, and boreal ecological regions. Interestingly, in light of the discussion of sodium below, puddling may be less common in coastal areas. While studying butterflies puddling in Amazonia in the 1860s, H. W. Bates, the British naturalist best known for his work on insect mimicry, was the first to note that the behavior is strikingly male biased. Subsequently, numerous studies have confirmed that this is generally true, although there are certain species in which females often exhibit the behavior. In addition to lepidopterans, nocturnal aggregations of leafhoppers (Homoptera: Cicadellidae), also predominately male, drink from moist soil surrounding puddles.

NUTRITIONAL ECOLOGY

Insect herbivores experience limited access to certain nutrients because of the paucity of these materials in their diet. Sodium as well as proteins and amino acids are substances of physiological importance that are not readily available from the foliar diets of larval lepidopterans. Puddling is a means by which adults augment the larval intake of these scarce materials. Since the 1910s, beginning with the British entomologist E. B. Poulton, there was speculation that puddling might allow lepidopterans to obtain sodium. The demonstration in the 1970s that sodium was a necessary stimulus to evoke puddling in the tiger swallowtail *(Papilio glaucus)* supported the notion. This and subsequent studies with butterfly species showed that when presented an array of various ionic solutions, the insects preferred those with sodium as the cation, and among sodium solutions, those with higher concentrations were favored. In the 1990s it was established that puddling indeed leads to sodium uptake: while drinking from mud puddles, the notodontid moth *Gluphisia septentrionis* showed a gain of sodium, matched evenly (on a molar basis) by a loss of potassium. Lepidopterans typically have an abundance of potassium, because the mineral is plentiful in the larval diet and consequently in the adult body.

Puddling can further result in the acquisition of nitrogenous nutrients. Tiger swallowtails drinking from moist soil laced with tritiated glycine and leucine incorporated these labeled amino acids into body proteins. Studies of Malaysian butterfly communities show that representatives of certain families (Papilionidae and Pieridae) preferentially visit sodium sources, while members of other families (Lycaenidae, Nymphalidae, and Hesperiidae) typically prefer a protein source, suggesting that nutritional needs may vary among taxa. Other drinking behaviors of moths and butterflies are also likely specializations to procure sodium or nitrogenous substances and thus may be functionally related to puddling. These insects often imbibe fluids that arise from vertebrates, including urine, feces, perspiration, blood, and lachrymal secretions.

REPRODUCTIVE BIOLOGY

Puddling is linked to lepidopteran reproduction. The male *Gluphisia* moth sequesters about 17 μg of sodium through puddling. Approximately 10 μg of this puddle-derived material is conveyed nuptially to the female, apparently via the spermatophore. About 5 μg of the transferred sodium is incorporated into the eggs. A nuptial transfer of sodium is also known from the European skipper *(Thymelicus lineola),* another puddling species. In this butterfly, there is apparently no endowment of eggs with male-derived sodium, but access to the mineral does enhance male mating success. To date, no studies have directly examined whether male Lepidoptera contribute puddle-derived amino acids to their mates. However, such a bestowal is quite plausible, given that males of certain nonpuddling lepidopteran species transfer amino acids via the spermatophore. The observed relationship between paternal consumption of an amino acid solution and increased egg viability in the tiger swallowtail may reflect such a nuptial transfer.

EXTREME PUDDLING BEHAVIOR

When puddling, a number of butterfly and moth species pump fluid through the digestive tract, emitting droplets from the anus. This behavior is displayed in extreme form by the male *G. septentrionis,* which forcibly releases anal jets at approximately 3-s intervals while imbibing from mud puddles. In this insect's quest for sodium, drinking often persists for hours, resulting in the passage of immense volumes that can amount to over 600 times the moth's body mass. Coupled to this behavior is a sexual dimorphism of the ileum (anterior hindgut). The male's ileum is longer and wider than that of the nonpuddling female. Numerous villi are present in the male ileum, but are virtually absent in the female's. Thus the surface area of the male ileum is nearly 20 times that of the female, making this enteric region a likely site for the observed sodium absorption.

See Also the Following Articles

Digestion • Digestive System • Feeding Behavior • Lepidoptera • Nutrition

Further Reading

Adler, P. H. (1982). Soil- and puddle-visiting habits of moths. *J. Lepidopterists' Soc.* **36,** 161–173.

Adler, P. H., and Pearson, D. L. (1982). Why do male butterflies visit mud puddles? *Can. J. Zool.* **60,** 322–325.

Arms, K., Feeny, P., and Lederhouse, R. (1974). Sodium: Stimulus for puddling behavior by tiger swallowtail butterflies, *Papilio glaucus. Science* **185,** 372–374.

Beck, J., Mühlenberg, E., and Fiedler, K. (1999) Mud-puddling behavior in tropical butterflies: In search of proteins or minerals? *Oecologia* **119,** 140–148.

Boggs, C. L., and Jackson, L. A. (1991). Mud puddling by butterflies is not a simple matter. *Ecol. Entomol.* **16,** 123–127.

Lederhouse, R. C., Ayers, M. P., and Scriber, J. M. (1990). Adult nutrition affects male virility in *Papilio glaucus* L. *Funct. Ecol.* **4,** 743–751.

Pivnick, K. A., and McNeil, J. N. (1987). Puddling in butterflies: Sodium affects reproductive success in *Thymelicus lineola. Physiol. Entomol.* **12,** 461–472.

Smedley, S. R., and Eisner, T. (1995). Sodium uptake by puddling in a moth. *Science* **270,** 1816–1818.

Smedley, S. R., and Eisner, T. (1996). Sodium: A male moth's gift to its offspring. *Proc. Natl. Acad. Sci. USA* **93,** 809–813.

Pupa and Puparium

Frederick W. Stehr

Michigan State University

A pupa is the stage in the development of holometabolous insects between the mature larva and the adult wherein major morphological reorganization takes place. Pupation usually occurs in a protected location (in a cell or cocoon), but in some groups, such as many butterflies, the pupa (chrysalis) is suspended openly and is usually well camouflaged by its shape and color.

There are two basic kinds of pupae, exarate (Fig. 1) and obtect (Fig. 2). An exarate pupa has free appendages. An

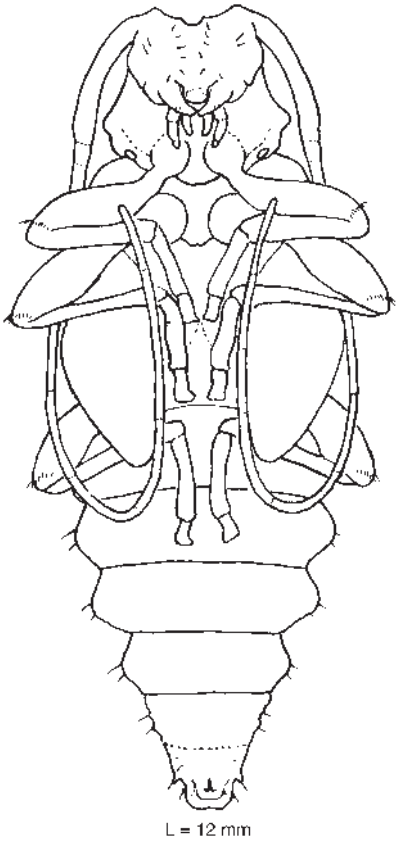

FIGURE 1 Exarate pupa of long-horned wood-boring beetle (Cerambycidae), showing free appendages (ventral). [From Peterson, A. (1948). "Larvae of Insects," Vol. 1. With permission of Jon A. Peterson.]

FIGURE 2 Obtect pupa of the monarch butterfly, *Danaus plexippus,* with appendages adhering to the body wall (lateral). The ready-to-emerge adult can be seen through the cuticle. (Photograph by Larry J. West, Mason, MI.)

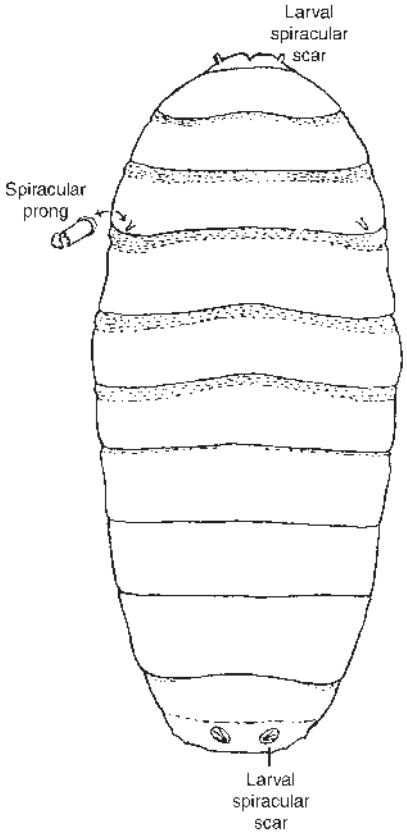

FIGURE 3 Puparium of a higher fly, formed from the last larval skin, showing the larval spiracular scars and pupal spiracular prongs that anchor the exarate pupa within (dorsal). [From Peterson, A. (1948). "Larvae of Insects," Vol. 1. With permission of Jon A. Peterson.]

obtect pupa has the appendages adhering to the body wall. Most Lepidoptera, most lower Diptera, some chrysomelid and staphylinid beetles, and many chalcidoid Hymenoptera have obtect pupae; nearly all other pupae are exarate.

BEHAVIORS

Most pupae are inactive, their body movements often limited to the abdominal segments. However, pupae in some groups are capable of locomotion, and some have functional mandibles that enable them to cut their way out of the pupal cell, cocoon, or chamber. These active pupae are sometimes referred to as pharate adults because the adult is enclosed in the pupal cuticle.

DESCRIPTION

Pupae may have articulated mandibles (decticous) or nonarticulated mandibles (adecticous). Decticous pupae are capable of chewing their way out of cells or cocoons and may be active. They occur in the Mecoptera (scorpionflies), Megaloptera (dobsonflies and fishflies), Neuroptera (lacewings, ant lions, and relatives), Raphidioptera (snakeflies), Trichoptera (caddisflies), and primitive Lepidoptera (Micropterigidae).

In the higher Diptera (Muscomorpha: Schizophora), the exarate pupa (also called a coarctate pupa) is enclosed in a delicate membrane within the hardened and barrel-shaped last (third) larval skin termed the puparium (puparia) (Fig. 3), which surrounds and protects the pupa. The adult fly emerges by forcing the end of the puparium off with the ptilinum, a membranous eversible pouch between the eyes and above the antennae that is expanded outward by blood pressure from the abdomen and thorax. After emergence, the ptilinum is withdrawn into the head, resulting in the frontal suture that arches over the antennae. If the puparium is buried in the soil, the ptilinum may also be used to help the adult fly crawl or force its way to the surface.

A prepupa is the last instar that has completed feeding. It may wander in search of a pupation site, but generally it becomes nonmobile before pupation. It is easily observed in Lepidoptera, wherein the mature larva becomes shortened and the prolegs and crochets become progressively retracted before pupation.

See Also the Following Articles
Chrysalis • Cocoon

Further Reading

Stehr, F. W. (ed.). (1987). "Immature Insects," Vol. 1. Kendall/Hunt, Dubuque, IA.

Queen

see *Caste*

R

Raphidioptera
(Snakeflies)

Ulrike Aspöck
Naturhistorisches Museum Wien

Horst Aspöck
University of Vienna

The Raphidioptera is a small order of winged, holometabolous insects. Adults (Fig. 1) have an elongate pronotum and two pairs of subequal wings of about 5 to 20 mm in length. Females have an elongated ovipositor. Larvae (Fig. 2) are terrestrial, living under bark or in detritus, and have biting mouthparts. Pupae are decticous (with articulated mandibles).

Raphidioptera is presumed to be the sister group of Megaloptera + Neuroptera and comprises two families: the Raphidiidae with 185 described species and the Inocelliidae with 21.

The Raphidioptera is distributed throughout the Holarctic region, except for the northern and eastern parts of North America; the southernmost records are from Mexico, northwest Africa, northern India, Indochina, and Taiwan. They are restricted to woodland habitats and occur in almost all Holarctic types of forests and forest-like habitats. In southern parts of their distribution they live mainly at high altitudes, up to about 3000 m.

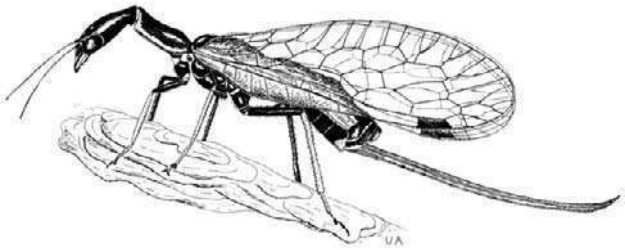

FIGURE 1 Female adult of *Turcoraphidia acerba* (Raphidioptera: Raphidiidae) from Anatolia. Length of forewing 8.5 mm.

HISTORY OF RESEARCH

Snakeflies first appear in the literature in 1735, when Linnaeus described an insect that he called Raphidia. By 1800 only 3 species had been described; in the 19th century snakeflies were known from southern Europe, Anatolia, and North America and, by the beginning of the 20th century, also from northern and central Asia and from northern Africa. By 1900, 31 species were known; in 1950, there were 60. At the beginning of the 1960s, a worldwide search for snakeflies and

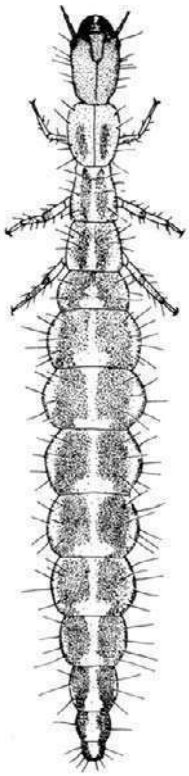

FIGURE 2 Larva of *Indianoinocellia mayana* (Raphidioptera: Inocelliidae) from Mexico. Length of body 22.5 mm. (Reproduced, with permission, from U. Aspöck and H. Aspöck, 1996.)

revision of the order was started and about 140 species were discovered and described between 1960 and 2000. Many species were reared from the egg, so that preimaginal stages and biologies of the majority of species are documented.

CHARACTERIZATION

An extremely flexible head characterizes the adults. It is prognathous, elongate, flat, and strongly sclerotized and may be broad or tapering basally. The large compound eyes are situated laterally. Inocelliidae lack dorsal ocelli. The mouthparts are of the biting type. The prothorax is remarkably elongated, particularly in Raphidiidae, and very mobile (hence the common names snakeflies and camelneck flies). All three pairs of legs are similar and cursorial. Both pairs of the elongate subequal wings are membraneous, and the venation is simple, with few crossveins. The abdomen consists of 10 (visible) segments. Terminal sclerites constituting the external genitalia are extremely complex in males; in females they are equipped with a long ovipositor.

The larvae are elongate and flattened, with a prognathous head, biting mouthparts, and five to seven lateral stemmata. The head and the only moderately elongate prothorax are strongly sclerotized, and the 10-segmented abdomen is of soft consistency.

BIOLOGY

Adult Raphidiidae are day-active entomophagous insects, preying on soft-bodied arthropods and pollen; the natural food of Inocelliidae is virtually unknown, but in captivity they feed on artificial diets and pollen has very rarely been found in their gut. All stages of larvae of both families are entomophagous, feeding on a variety of soft-bodied arthropods; the spectrum of prey is, however, considerably different in bark-inhabiting larvae on one hand and in larvae living in the soil on the other.

There is a long courtship before mating, including a highly sophisticated cleaning behavior with legs and antennae. Two positions of copulation have been observed: a "dragging position" in Raphidiidae, in which the male hangs head first from the female and is carried by her, and a "tandem position" in Inocelliidae, in which the male crawls under the female attaching his head in fixed connection to the fifth sternite of the female. Copulation lasts from a few minutes to $1\frac{1}{2}$ hours in Raphidiidae, but up to 3 hours in Inocelliidae. Spermatophores have been observed and studied only in Raphidiidae.

The egg stage lasts a few days to 3 weeks. The number of instars varies around 10 to 11 and may reach 15 or more. The larval period lasts 1 year in a few species, in most species it is 2 or 3 years, and under experimental conditions it may be up to 6 years. The prepupal stage lasts a few days. In the majority of species, pupation takes place in spring and lasts a few days to 3 weeks. In some species pupation starts in summer or autumn and the pupal stage lasts several (up to 10) months; in a few others pupation starts in summer and adults hatch in late summer after a short pupal stage. Hibernation thus usually takes place in the last larval stage, the penultimate stage, or the pupa, but never as eggs, prepupae, or adults. The pharate adult (the active pupa) is very mobile.

Snakeflies need a period of low temperature (around 0°C) to induce pupation or hatching of the imago. Larvae that are continuously kept at room temperature will usually not pupate but become prothetelous, i.e., they develop pupal or imaginal characters, such as compound eyes, wingpads, and appendages on the abdomen, and may live for years.

Parasites, parasitoids, and hyperparasites are a frequent phenomenon among Raphidioptera. Hymenoptera are of considerable significance as parasitoids of larvae, and species of the genus *Nemeritis* (Ichneumonidae) comprise 90 to 95% of them; other ichneumonids, braconids, and chalcidids contribute about 1%.

Snakeflies are effective predators; all larval stages of all species of both families, and at least the adults of Raphidiidae, feed on (mainly soft-bodied) arthropods. Snakeflies are believed to be rare insects. This is true for many species and many regions, but a number of species often occur in large numbers. There have been several attempts to use snakeflies as biological control agents: an unidentified North American species was introduced in Australia and New Zealand 100 years ago, but did not become established. The use of snakeflies for future biological control efforts is, however, hampered by the long developmental period and narrow food preference of these insects.

DISTRIBUTION AND BIOGEOGRAPHY

Extant Raphidioptera are confined to the Northern Hemisphere and almost exclusively to the Holarctic region. In Central America the southernmost records are from high altitudes at the Mexican–Guatemalan border. In Africa snakeflies have been found only in arboreal regions north of the Sahara, and in Asia the southernmost records are from altitudes above 900 m in transition areas from the Palaearctic to the Oriental Region in northern India, Myanmar, and northern Thailand. The northern and eastern parts of North America lack snakeflies. Almost all species are restricted to very limited areas of a refugal nature. Only three species represent a Eurosiberian type of distribution occurring throughout northern Asia to central and northern Europe, and a few Nearctic species with distribution centers in the southwest have reached Canada.

SYSTEMATICS, TAXONOMY, AND FOSSILS

The order Raphidioptera is a relic group of "living fossils" that comprises two extant families and 206 described species.

Taxonomy of adults has been well established. Because shape and coloration of body structures and of wing venation are highly variable, the most powerful and reliable tool is the morphology of the genital sclerites, in particular of the males. Taxonomy of larvae, mainly based on patterns and coloration of the abdomen, still remains difficult because of the similarity of related species.

The fossils lead to the conclusion that there was an enormous biodiversity of Raphidioptera in the Mesozoic. The majority of species and genera (in several families) are known from Jurassic and Cretaceous deposits. The rich and diverse Raphidioptera fauna of the Mesozoic died out at the end of the Cretaceous, probably due to the K/T event (that is the worldwide catastrophe resulting from an asteroid of about 10 km diameter that hit our planet) 65 mya and its climatic consequences. In particular, all snakeflies of tropical climates disappeared and apparently only the few representatives adapted to a cold climate survived. Fossil snakeflies from Tertiary layers, as well as from Baltic amber, belong to the two extant families.

See Also the Following Articles

Megaloptera • Neuroptera

Further Reading

Aspöck, H. (1998). Distribution and biogeography of the order Raphidioptera: Updated facts and a new hypothesis. *Acta Zool. Fenn.* **209,** 33–44.

Aspöck, H., and Aspöck, U. (1991). Raphidioptera (snakeflies, camelneckflies). *In* "The Insects of Australia. A Textbook for Students and Research Workers" (I. D. Naumann *et al.,* eds.), 2nd ed., Vol. I, pp. 521–524. Melbourne University Press, Melbourne.

Aspöck, H., Aspöck. U., and Rausch, H. (1991). "The Raphidioptera of the World. A Monograph of the Systematics, Taxonomy, Biology, Ecology and Chorology of the Extant Raphidioptera of the World, with a Summarizing Account of the Extinct Raphidioptera (Insecta: Neuropteroidea)." Goecke & Evers, Krefeld, Germany. [In German]

Aspöck, U., and Aspöck, H. (1996). Raphidioptera. *In* "Biodiversidad, Taxonomía Biogeografía de Artrópodos de México: Hacia una Síntesis de Su Conocimiento" (J. L. Bousquets, A. N. García Aldrete, and E. G. Soriano, eds.), pp. 277–286. Universidad Nacional Autónoma de México, México, D.F.

Aspöck, U., Aspöck, H., and Rausch, H. (1992). Extant southern boundaries of the distribution of the order Raphidioptera in America (Insecta: Neuropteroidea). *Entomol. Gen.* **17,** 169–184. [In German].

Aspöck, U., Aspöck, H., and Rausch, H. (1994). New species of the family Raphidiidae from Mexico and evidence of a spermatophore in the order Raphidioptera (Insecta: Neuropteroidea). *Entomol. Gen.* **18,** 145–163. [In German].

Aspöck, U., Plant, J. D., and Nemeschkal, H. L. (2001). Cladistic analysis of Neuroptera and their systematic position within Neuropterida (Insecta: Holometabola: Neuropterida: Neuroptera). *Syst. Entomol.* **26,** 73–86.

Kovarik, P. W., Burke, H. R., and Agnew, C. W. (1991). Development and behavior of a snakefly, Raphidia bicolor Albarda (Neuroptera: Raphidiidae). *Southwest. Entomol.* **16,** 353–364.

Tauber, C. A. (1991). Order Raphidioptera. *In* "Immature Insects" (F. W. Stehr, ed.), Vol. 2, pp. 123–125. Kendall–Hunt, Dubuque, IA.

Rearing of Insects

Norman C. Leppla
University of Florida

The goal of organized insect rearing is to provide reliable, affordable sources of high-quality insects for their many important uses. We are now able to rear literally thousands of insect species through multiple generations and many more for part of their life cycles. The greatest difficulty is to provide fresh host material or to develop a diet that is nutritionally complete and induces feeding, especially for parasites or parasitoids. Precautions must be taken to start colonies with an adequate number of uncontaminated specimens and maintain them with very limited levels of mortality. These measures help to limit genetic bottlenecks caused by inbreeding, competition with other species, and disease epizootics. As production levels and the number of species increase, rearing must be organized into systems with discrete operations or activities. These operations incorporate the rearing procedures necessary for each stage of the insect's life cycle plus maintenance of supplies, equipment, and facilities. Facilities must be designed and constructed to contain and maintain the insects under specified conditions. Problems encountered in established insect rearing systems invariably are caused by inattention to procedural details or lack of environmental control. For the foreseeable future, advances in insect rearing will be focused primarily on culturing new species, natural enemies, and pest species that have been genetically modified for pest management.

PURPOSES OF REARING INSECTS

Insects are reared for many reasons that may not be obvious to the nonentomologist. Certainly, insects are aesthetically pleasing and, therefore, reared by those who appreciate their many shapes, colors, features, and behavior patterns. This appreciation may be shared by means of personal collections, static displays, insect zoos, butterfly houses, and even household pets. Many of these and similar experiences would not be possible without insect rearing.

Insects are reared for a wide range of primarily agricultural and medical applications, an unusual example being the recently publicized biological warfare against plants that produce cocaine. Chemical insecticides are developed by using laboratory colonies of insects to mechanically screen massive numbers of candidates, more than 1 million insecticides per year in some instances. Similarly, plants are screened for resistance to insects or the disease organisms they transmit. Nontarget insects and plants are tested as possible hosts before nonindigenous natural enemies are released into nature. Insects have been reared, marked, and released to understand their orientation, dispersal, and migration. The cells of insects are used to study physiological processes such as reproduction, ontogeny, growth, aging, and cold tolerance. The fruit fly *Drosophila melanogaster* is the "white rat" of the geneticists and the nerves of large cockroaches are used by sensory physiologists. Students learn morphology by dissecting lubber grasshoppers, *Romalea* spp., and other insects reared in the laboratory. A variety of educational subjects involve insect colonies, including the process of insect identification, principles of insect taxonomy and systematics, and engineering aspects of insect flight. A somewhat gruesome but effective practice is the postoperative surgical use of the blow

fly maggots *Lucilia sericata* and *Phormia regina* to maintain clean wounds.

The major use for laboratory-reared insects is in biologically based methods of pest management. One of these methods, biological control, is the rearing and release of parasitoids, predators, and pathogens to suppress pest insects. A global industry has developed to rear and sell these natural enemies (www.anbp.org). Autocidal control is accomplished using the sterile-insect technique or inherited sterility. Male insects are reproductively sterilized with γ-irradiation or chemicals and released by the thousands per hectare to mate with wild females. This ensures that most of the wild females mate with sterile males and do not produce offspring. Inherited sterility is a variation using partially sterile males to induce sterility in subsequent generations. Genetic control takes advantage of altered genes to disrupt the insect's physiology and behavior. In the near future, we may use mass-reared insects to widely distribute aberrant endosymbionts via paratransgenic strains.

Insects have been reared to enhance wild and domestic populations, particularly for improving their products and benefits. Silk and honey production are obvious examples but pollination and pest management are equally important. Elaborate strain development methods have been used for silkworms, honey bees, screwworms (autocidal control), parasitic wasps (biological control), and others. Field insectary populations of expensive, showy butterflies and beetles have been used to augment natural populations that may become depleted. Eventually, the use of reared insects in conservation may include protection of rare and endangered species and repopulation, as is currently practiced with birds and predators.

KINDS OF INSECTS THAT ARE REARED

Insect rearing may be partial (egg to larva or nymph and adult) or complete (egg to egg). Virtually any free-living insect can be collected as an egg, larva, or nymph and maintained through stages in metamorphoses until it becomes an adult. However, this often requires considerable knowledge about the species' habitat and natural food. Re-creation of soil or aquatic environments, symbiotic relationships, and specialized foods can make rearing difficult. Some insects undergo temperature- and photoperiod-dependent diapause or require host plant cues to terminate multiyear cycles. Trophallactic feeding may necessitate maintenance of an entire colony, as in termites, ants, and other social insects. Because of these and other peculiar life history characteristics, the easiest immature insects to rear are relatively small, multivoltine (more than one generation per year), plant-feeding, terrestrial species with wide host ranges and no unusual environmental requirements. Species that infest common crops, landscape plants, or stored products are particularly suited to artificial rearing.

Complete rearing of an insect for one or more generations is complicated by the mating and oviposition requirements of the adults. Species-specific temperature, humidity, amounts of space, light characteristics, photoperiod, population size, food, oviposition stimulants and substrates, and other environmental conditions all must be understood and provided. Insects may swarm and couple in flight, form mating aggregations on host plants, orient to each other by means of pheromones or auditory signals, transfer spermatophores (sperm packets), engage in postmating female guarding to protect their paternity, and perform other unimaginable actions to produce another generation. Fortunately, most of the species we rear for multiple generations have less complicated requirements; these are primarily butterflies and moths (Lepidoptera, 300+ species have been reared), beetles (Coleoptera, 200+ species), flies (Diptera, nearly 200 species), bugs (Heteroptera, <100 species), and bees and wasps (Hymenoptera, <100 species). Grasshoppers and katydids (Orthoptera), lacewings (Neuroptera), cockroaches (Blattodea), termites (Isoptera), and fleas (Siphanoptera) are also reared but in reduced numbers (roughly 10–20 species). Many more species undoubtedly could be reared using the techniques developed for their close relatives.

NATURAL AND ARTIFICIAL DIETS

Immature herbivorous (plant-feeding) insects can often be reared by feeding them clean, fresh-cut, or potted versions of the plant material on which they feed in nature. The roots, stems, leaves, flowers, or fruit must be readily available or grown in sufficient quantities. Examples are larval silkworms raised on mulberry leaves, boll weevils on cotton squares and bolls, tropical fruit flies on papaya, and monarch butterflies on milkweed leaves. Similarly, medical and veterinary insects, i.e., adult mosquitoes and biting flies, are fed on their hosts or suitable surrogates, such as rodents, sheep, goats, or pigs.

Substitute plants can also be used to maintain herbivorous insects that are adaptable, usually readily available human and animal food. Green beans can be used for plant bugs, lettuce for grasshoppers, dry dog food for cockroaches, and cow manure for house flies. Rearing natural enemies of plant- or animal-feeding insects is considerably more difficult because three trophic levels must be synchronized: the plant or animal, the insect host, and the natural enemy.

Artificial diets have been developed to simplify and improve the rearing of both plant- and animal-feeding insects. Henry Richardson's development in 1932 of a bran, alfalfa meal, yeast, and diamalt formula for rearing house flies eliminated the objectionable mess and odor of cow manure. A commercial diet made of wheat bran (33.3%), alfalfa meal (26.7%), and brewer's grain (40%) is now available for rearing house fly larvae (CSMA medium; Chemical Specialties Manufacturer's Association, Ralston-Purina, St. Louis, MO). Another historical advancement was M. H. Haydak's 1936 grain flower, milk powder, honey, and glycerine diet for stored-grain insects, such as mealworms and flower moths. Gelled diets were developed for rearing insect larvae that require large quantities of contained water in their diets. The first was Pearl's 1926 diet for *Drosophila* spp., followed by Botger's

1942 larval medium for the European corn borer, *Ostrinia nubilalis.* Interestingly, these and subsequent diets have been gelled with agar, a polysaccharide derived from seaweed that was previously developed for use in bacteriology by Robert Koch in the late 1800s. Agar remains the standard gelling agent for culturing both microorganisms and insects; however, its cost and requirement for heating have led to the development of alternative materials: polysaccharides (industrial gums, cellulose, pectin, plant starches), heteropolysaccharides (carrageenan, sodium and calcium alginate), scleroproteins (gelatins, animal albuminoids), starch polymers (polyacrylonitrile graft copolymers), crude fibrous plant products, and waxes. Ground plant fibers, such as sugarcane bagasse (pulp remaining after the sugar is extracted), corncobs, and carrot powder are used to rear tropical fruit fly larvae. The awful stench of using a mixture of dried blood, milk, and yeast in water for rearing screwworm larvae was virtually eliminated by a starch polymer-gelled diet developed primarily by David Taylor of the USDA, Agricultural Research Service (ARS), in 1988. A practical artificial diet (primarily of beef liver, ground beef, and sucrose) for predaceous insects has recently been perfected and patented by Allen Cohen, also with the ARS.

INITIATION AND MAINTENANCE OF COLONIES

Insect colonies are initiated from field-collected specimens or samples from previously established colonies. Any developmental stage can be used to start a colony, but surface-sterilized eggs are generally preferred because they are durable, easy to ship, and less likely than other stages to carry a pathogen. However, eggs may be difficult to find in nature and often larvae suffer high levels of mortality as first instars because they are not yet adapted to the laboratory. It is generally advisable to use late instars, hold them in individual containers for parasitoid and pathogen screening, combine the adults in mating cages with a suitable oviposition substrate, and collect and treat the eggs before colonization. From either source, field or insectary, the degree of success achieved over multiple generations will depend on the quality of the colonized insects and the skill with which they are reared. Many species that can be colonized and reared for multiple generations are much larger, healthier (free of malnutrition, pathogens, parasitoids, and predators), uniform in growth, and more active and fertile than those in nature. Special consideration must be given to rearing insects that are required to behave normally, particularly those destined for release to suppress wild populations. Fruit flies, screwworms, and other species that have been mass reared for 20 or more generations in isolation may no longer interact and mate with their target populations in nature. To avoid this so-called "strain deterioration," insectary populations must be recolonized or infused periodically using specimens collected from the targeted wild population. Yields will be relatively low for several generations in a new colony destined for mass production, unless a previously isolated strain has been adapted to the insectary in anticipation of its use for colonization. Infusion is accomplished best by holding field-collected larvae, obtaining adults, and combining their eggs with those of the mass-reared colony, as explained below. A large number of existing new relative insects must survive to ensure that the colony has been infused. Also, genetic bottlenecks (sources of selective mortality) must be avoided because these will hasten the selection of insects that no longer behave normally.

Colony starts can be obtained by contacting entomologists who publish on species of interest or are involved in large-scale, biologically based pest management programs. In the 1980s the Entomological Society of America published lists of *Arthropod Species in Culture* and specialized directories still exist for *Drosophila* strains and other species. *Arthropod Species in Culture* listed colonies of the following taxonomic orders (number of species, colonies): Acari (41, 77), Anoplura (1, 1), Coleoptera (78, 266), Diptera (168, 301), Heteroptera (90, 206), Hymenoptera (119, 169), Lepidoptera (101, 308), Mallophaga (3, 3), Neuroptera (2, 2), Orthoptera (64, 203), Siphonaptera (2, 7), and Thysanura (4, 11). *Suppliers of Beneficial Organisms in North America* is maintained by Charles Hunter of the California Environmental Protection Agency (http://www.cdpr.ca.gov). This publication lists more than 130 species of beneficial organisms available from 142 suppliers. The most popular species offered for sale were the green lacewing, *Chrysoperla carnea* (65 suppliers); brown lacewing, *C. rufilabris* (54); mealybug destroyer, *Cryptolaemus montrouzieri* (52); whitefly parasitoid, *Encarsia formosa* (54); convergent ladybeetle, *Hippodamia convergens* (56); predaceous mite, *Phytoseiulus persimilis* (54); and egg parasitoid, *Trichogramma pretiosum* (77). There are also many commercial sources of insects, particularly for classroom education, including Carolina Biological Supply, Entomos, and Combined Scientific Supplies. Pioneers in supplying diets for research are BioServe and Southland Products. It is easy to colonize expensive pet food insects, such as mealworms and crickets, that can be purchased at local stores. Insect strains are not patented like plant varieties, so their use is not restricted.

CATEGORIES OF INSECT REARING

There are three distinct approaches to rearing insects: single species, multiple species, and mass rearing. In single-species rearing, immature stages are usually fed on host plants or animals, although artificial diets may be substituted. Seminatural environments and oviposition substrates are duplicated from nature and all rearing operations can be performed by a single individual. Multiple-species rearing is usually accomplished in centralized facilities to support research. There is economy of scale in these rearing operations, i.e., diet preparation, egg treatment, larval rearing, harvesting of pupae, and adult colony maintenance can be combined for similar species. Multiple-species rearing is common in research laboratories, such as those used for insecticide or transgenic crop development, and is typically performed by a small staff. Mass rearing

involves a single species produced for biologically based pest management that is reared in factory-like facilities with controlled environments, artificial diets and oviposition substrates, mechanized equipment, and operations performed by work units. The largest facilities are used to rear the screwworm and Mediterranean fruit fly. Although based on single-species rearing, these three approaches are distinct in design and implementation, not merely multiplications of scale.

BASIC INSECT REARING OPERATIONS

Regardless of the size of the colony, rearing procedures are organized into operations based primarily on the species' life history. At some level of complexity, insect rearing operations include inventory, acquisition, and storage of supplies; diet preparation and containerization; egg collection and treatment; larval or nymphal development; pupal or adult recovery and distribution; adult colony maintenance; quality control; and facility and equipment maintenance. Depending on the species, these operations are subdivided into procedures for which space is allocated and personnel are trained and assigned. "Traffic patterns" are established in the facility that flow from relatively clean areas to those that are potentially contaminated, i.e., diet preparation through larval development and diet disposal. Diet and eggs are usually handled more carefully until they are sealed in clean rearing containers. Isolating the larvae, pupae, and adults in containers protects them from contamination and prevents workers from being exposed to potentially harmful microorganisms and allergens. Rearing operations are performed in synchronized sequences, so they can be interfaced at the most appropriate times.

REQUIREMENTS FOR INSECT REARING FACILITIES

Insect rearing facilities have evolved from "field insectaries" with outdoor temperature and humidity to controlled labratories that contain insects under high security. Field insectaries provide adequate environments for rearing insects under seminatural conditions on host plants or animals. However, pathogens, parasitoids, and predators are not controlled, and workers are exposed to potentially dangerous pathogens and allergens. Conversely, temperature, humidity, air quality and quantity, light quality and photoperiod, sanitation, and security are closely maintained in laboratory insectaries. Insects and supplies are carefully screened for contaminants before they are admitted and human access and exposure are limited. All openings to a laboratory insectary are sealed or filtered to prevent insects from entering or leaving, particularly in quarantine facilities. Quarantine and containment facilities have strict construction requirements and operational protocols. Laboratory insectaries must be well insulated and have highly filtered, recirculated air to be efficient and cost effective.

PROBLEMS ENCOUNTERED IN MAINTAINING INSECT COLONIES

Virtually all severe insect rearing problems result from failure to perform standard operating procedures or defective environmental controls. Once established, rearing operations become routine and individual procedures are easy to forget. For example, a dietary ingredient may be omitted or destroyed by overheating, eggs may be accidentally desiccated following surface sterilization (exchorionation), closely related species may be mixed unintentionally, or larval densities may be more or less than required per container. Additionally, dietary ingredients can deteriorate after prolonged storage. These kinds of problems can be detected and corrected before a colony is lost. However, temporary loss of temperature control can destroy all of the insects. Pathogens can also be devastating, although diseased insects are usually confined to certain containers that can be eliminated before others are contaminated. Parasitoids and predators must be detected and similarly discarded. Genetic deterioration from inbreeding and genetic drift has been blamed for declines in insect colonies, but this is not a typical problem in large colonies. Insect rearing is generally safe for humans unless they are hypersensitive to insect proteins, react to the physical irritation of insects or their body parts, work with insects that sting or bite, or expose themselves to toxic substances used in the operations.

THE FUTURE OF INSECT REARING

Insect rearing is in transition, along with the science and technology it supports. Curiosity about the natural history of insects and their rearing is increasing steadily as people enjoy ecotourism, butterfly houses and gardens, insect zoos, and educational products based on insects. Insects are commonly used as baits for fishing and, in certain countries, have become popular pets. Rearing is becoming more important as field collection of insects is restricted to preserve habitat biodiversity and protect germplasm ownership. This is analogous to the collection of orchids and other showy organisms. Insects are no longer major sources of natural products, such as silk, wax, or dyes, and their use as human food is very limited. However, bird-watching and exotic pet ownership have become extremely popular in affluent countries, causing a significant increase in the use of insects as animal food. Rearing insects to produce bioactive substances remains a research support activity in the fields of biology and medicine. Living as well as dead insects are used increasingly in classroom education.

Agricultural uses for insects have expanded dramatically during the past 40 to 50 years in the development and support of new pest management technologies. Every major company that produces chemical insecticides or pest-resistant plants maintains a multispecies insect rearing facility. Although insecticides have provided effective insect control at individual farms and residences, overall losses to insects

have not declined and we need new options for sustainable pest management. Large federal and state entomological research laboratories and university entomology departments also have laboratory insectaries, although the current trend is toward decentralization and outsourcing. The sterile insect technique, pioneered by E. F. Knipling of the United States Department of Agriculture, has proven too expensive, unless used on an area-wide basis with stringent regulatory controls, as in the screwworm, *Cochliomyia hominivorax,* and Mediterranean fruit fly, *Ceratitis capitata,* eradication programs. Advances in insect rearing have enabled the eradication of these species and others from vast geographical areas. As a result, area-wide approaches to pest management are increasing, along with new methods for using genetically modified organisms. Reliable, cost-effective rearing will be required for these technologies to be successful.

Major advances in insect rearing are currently being made in support of augmentative biological control. Natural enemies are reared and released to prevent rather than cure pest problems, and they rarely have unacceptable nontarget effects. Another major advantage in using natural enemies is that they do not induce the pest resistance that eventually makes insecticides ineffective. We now have efficient host-rearing systems for many parasitoids and new artificial diets for predators that have greatly increased shelf life. Advancements are also being made in the mechanization of rearing operations, large-scale release technology, and rearing of newly discovered natural enemies. A global industry has developed during the past 10 years to ensure the quality of natural enemies and expand their use. Markets are increasing in organic food production, ornamental and vegetable greenhouse crops, urban pest management, filth fly control, home gardening, and other specialized areas. Biological control will continue its expansion as insect resistance to chemical insecticides increases, worker protection and food safety regulations are enforced, people avoid real or perceived environmental contamination, and the efficacy of natural enemies improves.

Insect rearing will play a critical role in the future of entomology. Insects will always be appreciated for their intrinsic value, used as a source of useful materials, and produced as food for wildlife. However, they will become more important for pest prevention in natural areas, crops, and buildings and on human and animal wastes. Other uses will include the production of transgenic biological control agents, autocidal agents (sterile insect technique, paratransgenesis), and new species and strains for biological control. Insects will be reared to restore and supplement insect populations in nature; control pests over vast, low value-per-acre lands; and eliminate chemical insecticides in specialized cropping systems.

Insect rearing can be enhanced most by learning more about the natural history of insects, emphasizing their ecology, behavior, and systematics. This knowledge can be used to create artificial diets and environments that separate species from limiting factors, biotic (pathogens, parasitoids, and predators) and abiotic (temperature, humidity, and light). As information is gained, it must be published, communicated at meetings, and widely distributed to help advance the field. The Entomological Society of America and International Organization for Biological Control have been particularly helpful in publicizing insect rearing information. Unfortunately, however, insect rearing is often considered a support activity most appropriately described in the methods sections of articles on other subjects. This makes the information difficult to retrieve and has led to relatively obscure publications on insect rearing. Another major limitation has been a general lack of formal education and training in insect rearing. However, Frank Davis conducts a popular annual short course on the subject at Mississippi State University and David Dame covers the subject in his biannual FAO, International Atomic Energy Agency short course at the University of Florida. It is essential for us to continue discovering, documenting, and sharing insect rearing knowledge and preserving the rich history of this field.

See Also the Following Articles

Biological Control • Cell Culture • Genetic Engineering • Medicine, Insects in • Nutrition • Pathogens of Insects • Sterile Insect Technique

Further Reading

Anderson, T. E., and Leppla, N. C. (eds.) (1992). "Advances in Insect Rearing for Research and Pest Management." Westview Press, Boulder, CO.

Edwards, D. R., Leppla, N. C., and Dickerson, W. A. (1987). "Arthropod Species in Culture." Entomol. Soc. Am. College Park, MD.

Grenier, S., Greany, P. D., and Cohen, A. C. (1994). Potential for mass release of insect parasitoids and predators through development of artificial culture techniques. *In* "Pest Management in the Subtropics: Biological Control—A Florida Perspective" (D. Rosen, F. D. Bennett, and J. L. Capinera, eds.). Intercept, Andover, MA.

King, E. G., and Leppla, N. C. (eds.) (1984). "Advances and Challenges in Insect Rearing." USDA, ARS, U.S. Gov't Printing Office, Washington, DC.

Leppla, N. C., and Ashley, T. R. (eds.) (1978). "Facilities for Insect Research and Production." USDA Technical Bull. No. 1576.

Needham, J. G., Galtsoff, P. S., Lutz, F. E., and Welch, P. S. (eds.) (1937). "Culture Methods for Invertebrate Animals." Comstock, Ithaca, NY.

Peterson, A. (1964). "Entomological Techniques: How to Work with Insects." Entomological Reprint Specialists, Los Angeles.

Poinar, G. O., and Thomas, G. M. (1978). "Diagnostic Manual for the Identification of Insect Pathogens." Plenum, New York.

Ridgway, R. L., Hoffmann, M. P., Inscoe, M. N., and Glenister, C. S. (eds.) (1998). "Mass-Reared Natural Enemies: Application, Regulation, and Needs." Entomol. Soc. Am., Lanham, MD.

Singh, P. (1977). "Artificial Diets for Insects, Mites, and Spiders." Plenum, New York.

Singh, P., and Moore, R. R. (eds.) (1985). "Handbook of Insect Rearing," Vols. 1 and 2. Elsevier, Amsterdam.

Smith, C. N. (ed.) (1966). "Insect Colonization and Mass Production." Academic Press, New York.

Villiard, P. (1969). "Moths and How to Rear Them." Funk & Wagnalls, New York.

Recruitment Communication

James F. A. Traniello
Boston University

In the social insects, the daily demands of colonial life are often met cooperatively by directing workers to sites where work is required, such as an energetically rewarding food source, an alternate nest site, or the site of a territorial dispute. The behavior that mobilizes nestmates is termed recruitment, and it is a type of communication commonly used by social insects to accomplish work. The mechanisms that mediate recruitment communication can be understood by analyzing the behaviors of individuals and the signals they produce. Numerous studies have described how pheromones, sometimes in conjunction with nonchemical signals, underscore insect social integration. Although systems as "linguistically" elaborate as the honey bee waggle dance have evolved to communicate information about the location of food sources or nest sites, it is pheromones that serve the primary signaling role in wingless social species such as ants and termites.

The majority of research on recruitment behavior concerns communication during foraging and defense. Workers serve as scouts, searching for food sources and patrolling the territory of a colony; when food is located or a competitor from an alien colony is identified, nestmates are recruited. Scouts that have located food or a territorial intrusion return to the nest, laying a chemical trail and sometimes performing behavioral displays to alert nestmates, which then leave the nest and orient along the scout's trail to the target area. Recruited nestmates, in turn, may continue the process by reinforcing the trail. Regulatory mechanisms turn off recruitment when the food has been harvested or the threat no longer exists.

Ecology influences the evolution of recruitment communication, and the adaptiveness of recruitment behavior can be studied in reference to patterns of food distribution, predation, and competition. Foraging behavior and space-use patterns are the result of community-level interactions such as interference competition, and recruitment communication is one behavioral mechanism that mediates interactions between sympatric species. Recruitment signals, in turn, are generally trail pheromones that "excite" nestmates and orient them to a locus of activity. The physical characteristics of trail (e.g., how long it lasts as a signal) and the response it induces reflect the ecological function of the pheromone.

This article considers recruitment behavior in social insects, focusing primarily on ants and termites, groups that frequently employ these communication systems and for which the greatest level of ethological and ecological understanding has been achieved.

PHYSIOLOGY AND BEHAVIOR OF RECRUITMENT

Exocrine Gland Sources of Recruitment Pheromones; Trail-Laying and Trail-Following Behavior

Recruitment pheromones are discharged from exocrine glands, which are anatomical structures often specialized for their synthesis and secretion. Since the first identification of the source of the trail pheromone in the fire ant by E. O. Wilson more than four decades ago, ants have served as excellent models for the study of the organization of recruitment. Thanks to the research of Bert Hölldobler and Hiltrud Engel-Siegel, among others, the structure of ant exocrine glands and the function of their secretions have been described in detail for many species. In ants, the accessory glands to the sting (the Dufour's gland and the poison gland), the pygidial gland and sternal glands, the hindgut, the rectal gland, and the tibial and tarsal glands are known to produce substances that serve to recruit nestmates (Fig. 1). In ants in the subfamily Formicinae, the hindgut is the source of trail pheromone, which is emitted through the acidopore located at the tip of the gaster. In another diverse group of ants, the subfamily Myrmicinae, the poison gland and the Dufour's gland secrete recruitment chemicals. Other ants rely on a variety of glands to produce trail substances. In termites, only one source of trail pheromones, the sternal gland, has been described. The structure of the sternal gland, which is composed of modified epidermal cells, varies in different genera.

The secretions of the trail-substance-producing exocrine glands are deposited as a worker travels from a target area to the nest, or vice versa. In some ant species (e.g., myrmicine ants) the sting, serving as a conduit for the secretions of exocrine glands, is extruded and dragged over the substrate to release trail pheromones. Chemical trails may be deposited continuously or as a series of point sources between the nest and the target area, sometimes in conjunction with other secretions. The cuticle may be adapted as an applicator for the secretions of exocrine glands beneath. In termites, the sternal gland is pressed against the substrate, and sensory structures monitor contact between sternites and the substrate to regulate pheromone deposition. Some termites mark areas around the nest entrance or a food source by dotting the tip of the abdomen and laying directional trails using the sternal gland. There is significant convergence in the trail-laying behaviors of ant and termite species.

The spatial processing of the information encoded in an odor trail is by tropotaxis, which is a sampling of the trail pheromone by means of the paired antennae. Antennal chemoreceptors perceive variation in pheromone concentration along the trail's semiellipsoidal active space, the area in which the concentration of the pheromone is sufficient to elicit following behavior. Theoretically, diffusion yields a gradient of odor molecules; pheromone concentration is highest at the point of application (the centerline of the trail) and symmetrically decreases on either side, defining the boundaries of the active space. The odor gradient is sampled by the antennae as

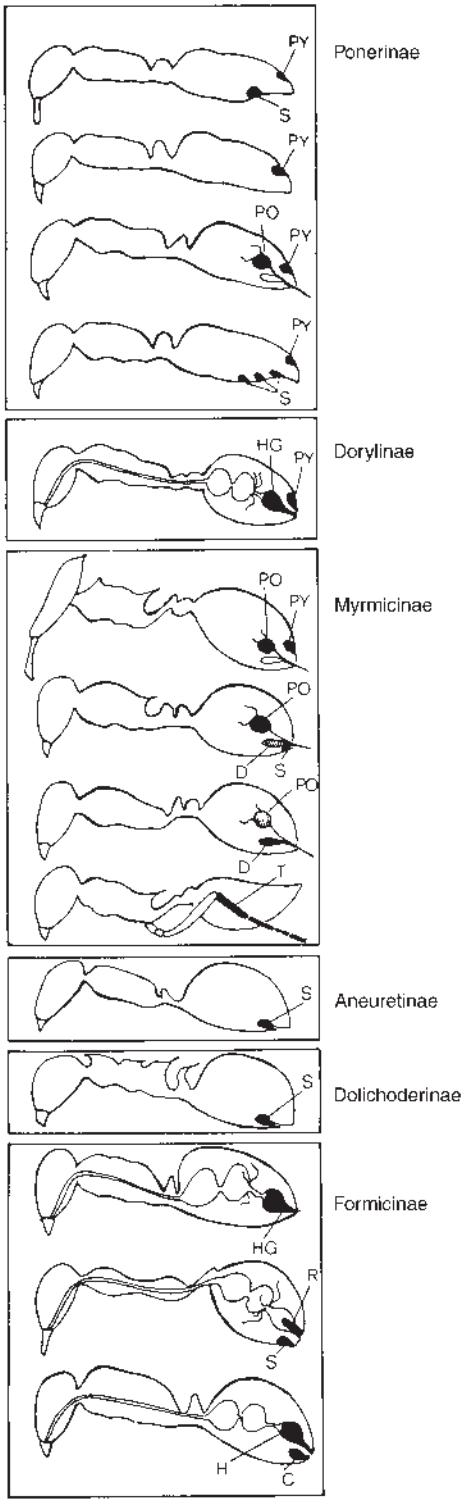

FIGURE 1 Exocrine gland sources of trail pheromones in the different subfamilies of ants: C, cloacal gland; D, Dufour's gland; HG, hindgut; PO, poison gland; PY, pygidial gland; R, rectal gland; S, sternal gland; T, tibial gland. [From Hölldobler, B., and Wilson, E. O. (1990). "The Ants," p. 268. Belknap Press of Harvard University, Cambridge, MA. Originally published in Hölldobler, B. (1984). Evolution of insect communication. *In* "Insect Communication" (Symposium of the Royal Entomological Society of London) (T. Lewis, ed.), pp. 349–377. Academic Press, London. Reprinted with permission of Harvard University Press and Prof. Hölldobler.]

the insect travels along the trail. When lower concentrations are sensed at the lateral edges of the active space, opposing movements are made so that position within the active space is maintained.

Determining the Source and Behavioral Effects of Trail Pheromones

Understanding the glandular sources and chemistry of trail substances is at the heart of the study of recruitment communication. The nature of the bioassay, the behavioral test used to determine the effectiveness of different substances as recruitment or trail pheromones, is critically important. The bioassays used in trail pheromone isolation and identification must distinguish among the range of behavioral responses involved in trail communication. It must be noted that trail pheromones can have both recruitment and orientation effects. A recruitment pheromone induces nestmates to leave the nest to travel to a work site. An orientation pheromone has no such stimulatory effect, but it can serve as a chemical "guide" for worker traffic. Trail substances, if they have a recruitment effect, will stimulate nestmates to leave the nest or otherwise alter their task performance in the context of a current need. In some ants, nestmates are recruited with a motor display delivered in the nest by a recruiting worker. The trail substance in this case does not have the ability to draw ants out from the nest; rather, it is used as an orientation cue by nestmates that have contacted a recruiting ant. Some trail pheromones can alone elicit both excitation and orientation in the absence of any other behavioral display or stimulus. If an artificial trail (one prepared from a solvent extract of the appropriate exocrine gland) is drawn out from the nest entrance and ants leave the nest to follow it, a recruitment effect has been demonstrated. If the artificial trail cannot induce inactive workers to leave the nest, yet the trail is able to orient workers alerted by either a motor display or some other trail chemical that has an alerting property, an orientation effect is occurring. Careful dissection of the kinds and sequences of behaviors in the recruitment process and detailed chemical analyses have revealed that several pheromone constituents may control a number of behaviors associated with recruitment and trail following.

TRAIL CHEMISTRY AND RECRUITMENT BEHAVIOR

Social insects may mix the secretions of different exocrine glands to induce recruitment and trail-following behaviors, or the chemical output of a single gland may be composed of more than one substance, each having a distinct role in releasing behavior. Trail communication can therefore be a multisource phenomenon or a process that involves a series of chemical homologues produced in the same exocrine gland. For example, the Dufour's gland of the fire ant is the source of a trail pheromone that induces both recruitment and orientation behaviors in workers. Dufour's gland chemistry is varied: the constituents of this gland's secretion regulate

different behaviors, which have been called "subcategories" of trail following. These chemicals include recruitment primers, synergists, and orientation inducers. Primer and inducer substances together release recruitment and orientation behaviors. *(Z,E)-a*-Farnesene is the principal trail orientation component isolated from the Dufour's gland. Another chemical fraction acts together with *(Z,E)-a*-farnesene to increase the effectiveness of the mixture in inducing trail communication. Homofarnesenes of presently unknown function and an orientation inducer present in the secretion also increase trail following. In the ant *Myrmica,* homofarnesenes in the Dufour's gland may be added to 3-ethyl-2,5-dimethylpyrazine (EDMP), which is the poison gland trail pheromone, as part of a multicomponent trail system.

Similar findings have been made in other ant species. Pure chemicals seem to induce lower responses than gland extracts, indicating the importance of the naturally occurring chemical mixtures in trail communication. Constituents present in different ratios sometimes show synergistic effects. Artificial trails prepared from extracts of the poison gland of the harvester ant *Pogonomyrmex badius* have a recruitment effect lasting approximately 20 mins, whereas artificial trails prepared from Dufour's gland secretions and aged for longer periods of time have elicited orientation responses. In the ant *Leptogenys diminuta* the poison gland and pygidial gland produce (3*R*,4*S*)-4-methyl-3-heptanol and isogeraniol, respectively, to regulate orientation and recruitment. Similarly, pheromone blends are known to mediate alarm communication, including alarm–recruitment systems involved in defense.

The sternal gland secretions of termites stimulate recruitment and may have highly durable orientation effects. Single chemicals such as *(E)*-6-neocembrene A and dodecatrienol have been isolated from whole-body extracts, but termite sternal gland secretions probably are more elaborate mixtures of pheromones that have different functions. Researchers have described "recruitment" and "basic" trails in termites; "basic" trails have only an orientation effect. The sternal gland secretion of *Nasutitermes costalis,* for example, can induce recruitment (drawing undisturbed soldiers and workers from the nest) and can orient searching and/or homing termites. Sternal gland material collected from trails aged for more than 20 years can orient, but not recruit, termites. Thus although the chemical that regulates recruitment dissipates in minutes, the orientation component of the secretion is a remarkably stable pheromone.

The persistent orientation components of a trail substance can "channel" foragers away from neighboring nests to minimize aggressive confrontations and can also serve as territorial recognition cues or as an initial guide for naive foragers. Some species of desert ants have trunk trail systems (a network of trails emanating from the nest entrance and arborizing at their distal portions) marked with Dufour's gland secretions composed of durable blends of hydrocarbons that are specific for species, populations, and colonies. In other ant species, different glands may produce trail chemicals with different behavioral effects. The ecological significance of trail structure in termites is not well understood, although apparently foraging galleries divide foraging space to increase the efficiency of harvesting food.

Trail Pheromone Specificity

Pheromone specificity is achieved through chemical mixtures and molecular structure. Early studies in ants suggested that trail substances were highly species specific, but results of more recent work do not support such a conclusion. One striking example of this lack of specificity is that ants in as many as six different genera in the subfamily Myrmicinae use the same trail pheromone, EDMP. And dodecatrienol, a trail pheromone in the termites *Reticulitermes virginicus, R. speratus,* and *Coptotermes formosanus,* provides a nonspecific orientation cue in these species, in other species of *Reticulitermes,* and in a cluster of geographically and phylogenetically diverse termites. *(E)*-6-Cembrene A, which has been isolated from whole-body extracts of the Australian *Nasutitermes exitiosus,* can induce orientation in African nasutitermitines, rhinotermitids, and to a lower degree in African macrotermitines (fungus-growing termites). At present, it is challenging to explain the significance of trail pheromone specificity in termites.

Metabolic end products, dietary differences, and genetics may produce variation in the chemistry of trail pheromones. In myrmicine ants, pyrazines and farnesenes are shared by different species and genera; these chemical constituents, present in small quantities and serving no function in one species, may have a prominent role in another. Biochemical variation may provide a substrate for evolution to act on in the selection of trail pheromones.

There have been many analyses of the level of specificity of trail substances, but the ecological significance of specificity is to a great extent unknown. Research on the behavioral ecology of trail pheromones indicates that chemical specificity may play a role in community structure. For example, competition may have selected for differences in the trail communication systems of sympatric species. Indeed, variation in recruitment communication systems in desert ants have been correlated with resource use and competition. Desert harvester ants forage as individuals on evenly distributed seed resources and recruit nestmates to cooperate in collecting seeds from dense patches. Different foraging strategies may be adapted to the exploitation of resources with different density distributions, serving as a mechanism of resource partitioning in granivorous ant communities. Foraging systems and their pheromonal regulatory mechanisms may also enhance food defense and retrieval, thus reducing interference competition.

The behavioral mechanisms of recruitment that are the basis for enhanced competitive ability may be associated with trail pheromone chemistry and response specificity. The harvester ants *Pogonomyrmex rugosus* and *P. barbatus,* which are very similar ecologically, have trunk trail systems that divert groups of foragers away from each other. Their trunk routes,

which are composed of Dufour's gland and are colony specific, suggest chemical differentiation resulting from competition. The specificity of the trunk routes may also give an advantage in territorial defense if fighting ability is greater when ants are on their own territory. Other ant species are known to mark trails with persistent colony-specific pheromones.

Colony specificity in the trail pheromones of termites has rarely been examined. In *Trinervitermes bettonianus,* workers do not discriminate their own trails from those of neighboring colonies, although their trail pheromones and the trail substances of other sympatric termites appear to be species specific. Some termites distinguish between intra- and interspecific competitors. Workers deposit rectal fluid and sternal gland secretions on trails, a mixture that may encode colony identity.

Surprisingly, the chemical trails of some ants may also encode information about the identity of the individual that laid the trail. Individually specific trail markings are used during nest emigration by workers of the ant *Pachycondyla tesserinoda.* Individually specific trails, used for food recruitment, have also been described in the ant *Leptothorax affinis.* Individuality in a trail pheromone could provide a finely tuned mechanism of directional discrimination and maintain the path fidelity of individual foragers.

If ants can decipher the chemical code of another species' trail substance, they may be able to exploit information about the location of food sources. Interspecific trail-following is uncommon but is known in ants in parabiotic associations (i.e., different species of ants living together in a compound nest). In these species, foragers lay and follow their own chemical trails but are also able to interpret the trail pheromones of other species and to exploit their food discoveries.

Regulation of Recruitment and Foraging Activity

Trail pheromones communicate information about food quality to nestmates that have not had direct experience with a food source. Trail-laying behavior regulates pheromone concentration (i.e., the amount of pheromone deposited on a trail), which in turn controls a colony's response. After deposition on the substrate, a trail pheromone diffuses. The chemical properties of the pheromone both determine the spatial and temporal structure of its active space and regulate foraging activity at the colony level. The concentration of trail pheromone mediates communication between groups of individuals, those that have fed and any potential foragers within the nest. In fire ants, the continuity of the sting trail, measured by causing a scout ant to walk over a smoked glass slide (removing soot from the slide where the ant's sting is dragged) depends upon the concentration of a sugar solution offered as a food source. The more concentrated and thus rewarding the solution, the greater the extent to which the sting is extruded and dragged continuously over the substrate, and the greater the number of workers that will lay such a trail after they have contacted the food source (Fig. 2). The distance between the food source and the nest can also affect the rate at which food is retrieved, and thus the profitability of the colony's foraging.

FIGURE 2 Fire ants deposit trail pheromones by discharging the contents of the Dufour's gland through the extruded sting. The continuity of the trail is made visible when a worker walks over a treated glass slide, removing soot where her body contacts the substrate. Marks made by the tarsi and hairs on the tip of the gaster are visible. Increase in the continuity of the sting trail can be seen. [From Hölldobler, B. and Wilson, E. O. (1990). "The Ants," p. 270. Belknap Press of Harvard University, Cambridge, MA. Originally published in Hölldobler, B. (1970). Chemische Verständigung im Insektenstaat am Beispiel der Hautflugler (Hymenoptera). *Unschau* **70(21),** 663–669. Reprinted with permission of Harvard University Press and Prof. Hölldobler.]

The regulatory mechanism underlying foraging organization is called mass communication. Foraging is initiated when scouts locate new food, determine its quality, and deposit trail pheromone according to the food's energetic value and the colony's nutritional needs. This trail induces recruitment in nestmates, which repeat the cycle of food quality assessment and trail phermone deposition until the food source has been depleted or the colony satiated. The entire process is regulated by the concentration and evaporation of the trail pheromone. If the recruitment trail is not reinforced, the trail substance evaporates and foraging rate decays over time until food collection ends. Different ant and termite species show variations on the theme of mass communication that likely are associated with a the foraging ecology of individual species.

DEFENSIVE RECRUITMENT AND DIVISION OF LABOR BETWEEN CASTES

Social insects can be the most important competitors and predators of other social insects. Army ants, for example, exert significant predation pressure on social wasps and ground-dwelling ant species, favoring the evolution of adaptive systems of predator recognition and response. Some responses involve a division of labor among castes. Castes are groups of individuals specialized for a given set of tasks; they may have specific functions related to foraging or colony defense, and different

castes may be recruited according to their specialization. Upon encountering a particularly important competitor or predator, ants that are patrolling the territory of a colony may prempt an attack through a caste-specific alarm– recruitment system. For example, the ant species *Pheidole dentata* and *Solenopsis geminata* are sympatric (i.e., occur together) in the southeastern United States and use similar nest sites. *S. geminata,* the native fire ant, as well as *S. invicta,* the imported fire ant, may attack colonies of *P. dentata* and destroy the colony (Fig. 3). Because fire ants are fierce predators of *P. dentata,* it is important to quickly respond to a potential threat. In *P. dentata,* minor workers usually care for brood, maintain the nest, and forage, whereas major workers, which have large heads and mandibles, have a significant role in colony defense. When minor workers encounter as few as one or two fire ants, they return to the nest, laying pheromone trails to recruit minor and major workers. Major workers recognize the threat as emanating from fire ants through the odor of the predator carried on the messengers' bodies, and together with the excitatory behavior and trail substances deposited by minor workers, are recruited to the site at which the enemy has been located (Fig. 4). Because majors are recruited to respond defensively following contact with fire ants but not other ant species, this defensive recruitment system is said to have "enemy specification."

Species of termites that forage above the ground face a higher degree of predation than species that forage in the confines of subterranean gallery systems, and the social architecture of these ecologically different groups of species seems to have undergone adaptive modification. For example, species whose foragers harvest food above the ground have a high proportion of soldiers that use chemical defense in combating predators. Species that nest and seek food below ground have relatively low investment in soldiers, which rely on mechanical defense and are not involved in foraging. The recruitment communication systems that organize foraging in aboveground species appear to be adapted to reduce predation during the time period between the discovery of food and the construction of covered protective galleries, when termites are

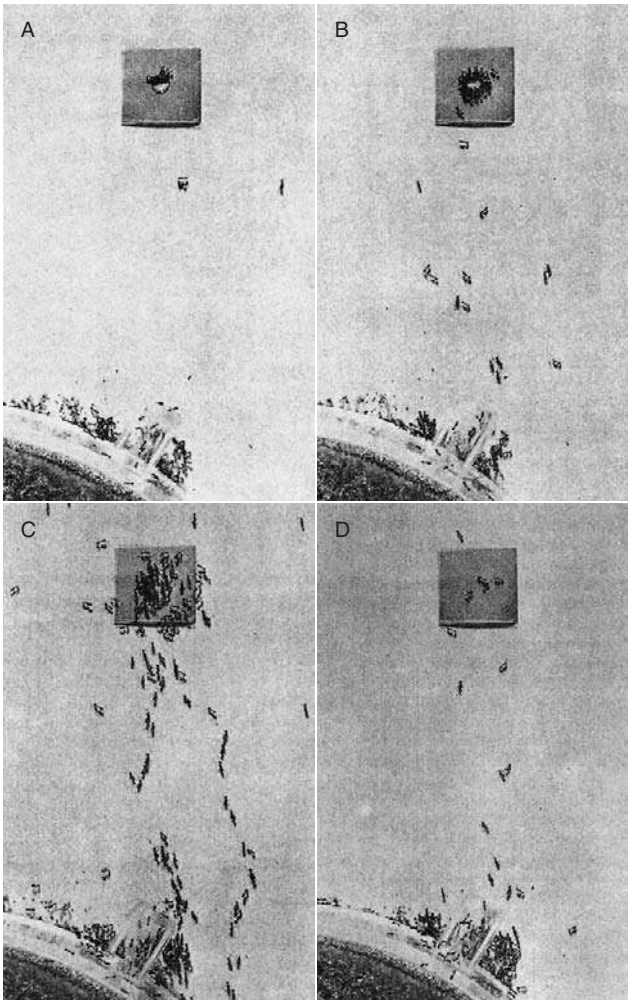

FIGURE 3 Recruitment and feeding behavior in a laboratory colony of fire ants. (A) Scout ants feeding at a drop of sucrose. (B) Ants initially recruited from the nest by trail-laying workers that had fed now feed and lay trails to the nest. (C) Recruitment increases as more ants arrive at the food and contribute pheromone to the trail. (D) The food is depleted, and new recruits arriving at the food do not lay trails on their return trip to the nest. Foraging now stops because the trail pheromone has evaporated. [From Wilson, E. O. (1963). Pheromones. *Sci. Am.* May, p. 110. Reprinted with the permission of Miriam F. Rothman for the estate of the photographer, Sol Mednick.]

FIGURE 4 Enemy specification in the recruitment behavior of the ant *P. dentata.* In response to contact with fire ants (light shading), major workers are recruited to defend the colony by attacking with their well-developed mandibles. [Reprinted with permission from Wilson, E. O. (1976). The organization of colony defense in the ant *Pheidole dentata* Mayr. *Behav. Ecol. Sociobiol.* **1,** 66. © Springer-Verlag GmbH & Co. KG.]

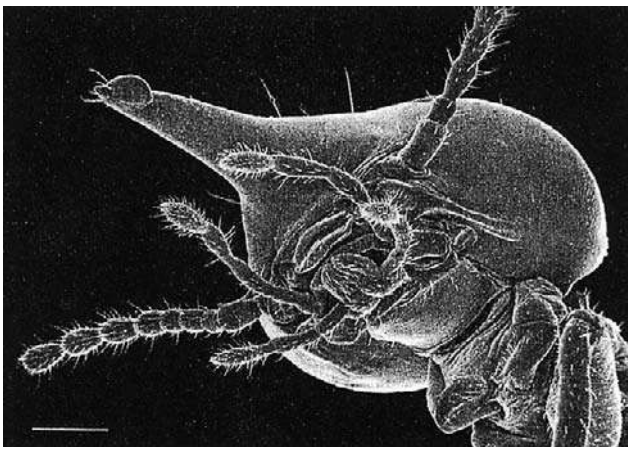

FIGURE 5 Head of a soldier of *Nasutitermes.* The mandibles are vestigial and the head is shaped to discharge defensive secretions. Scale bar = 0.25 mm. [Reprinted with permission from Eisner, T. *et al.* (1976). The organization of colony defense in the termite *N. exitiosus. J. Comp. Physiol.* **90,** Fig. 1. © Springer-Verlag GmbH & Co. KG.]

exposed and are vulnerable to predation by ants. In *Nasutitermes,* for example, the most diverse genus of the higher termites, there are worker and soldier castes: the ampule-headed soldiers are highly modified for chemical defense but also have an important role in organizing foraging (Fig. 5). Soldiers of the neotropical *N. costalis* scout in groups for food sources, and upon locating food communicate its location to nestmates. Groups of soldiers of *N. costalis* move in amoeboid fashion from the nest, ends of foraging galleries, and currently used food sources, recruiting other soldiers with sternal gland secretions as they slowly explore the environment. Parties of soldiers search as groups along trails in different areas, but all trails generally coalesce when a food source is located; subsequently additional soldiers and then workers are recruited.

There are three phases of foraging organization in *N. costalis,* each involving recruitment communication within and between castes. First, soldiers search for and discover new food sources and communicate information about their location to other soldiers. Next, workers are recruited in large numbers. Finally, the recruitment of workers increases further and soldiers flank both sides of the foraging trail to protect the more vulnerable workers traveling between the nest and the food. This pattern of soldier and worker recruitment shows that soldiers, which themselves do not feed directly, are nevertheless scouts, which assess food quality and recruit workers that will harvest food. Soldiers first recruit other soldiers to ensure an adequate defense at the food source and then communicate with workers to begin food collection. This intercaste communication is chemical; both soldiers and workers have a sternal gland, which produces a trail pheromone that induces the recruitment of soldiers and workers depending upon the quantity of pheromone deposited. The sternal gland secretion is not caste specific, but the volume of the sternal gland varies in soldiers and workers (large workers have a significantly larger sternal gland volume than soldiers). Caste differences in pheromone perception and/or responsiveness to the sternal gland pheromone appear to regulate the prominent division of labor during foraging organization in nasute termites.

Nasute termites also have an alarm–recruitment response. When disturbed, soldiers discharge frontal gland terpenoid secretions; soldiers nearby then are recruited over short distances. Frontal gland and sternal gland secretions function in defensive recruitment and cause soldiers to remain in an area where a disturbance has been signaled. Similar foraging and defensive recruitment communication systems have been identified in other termites.

Defensive recruitment and the recognition of specific enemies occur in some termites. *N. costalis* responds defensively to the presence of a single soldier or worker of an alien conspecific colony by recruiting large third-instar workers, which attack the intruders with their mandibles. This response is not seen if other species of termites are encountered.

THE EVOLUTION OF RECRUITMENT COMMUNICATION

Ethological Models

Ethologists have long hypothesized that the origins of behavior can be traced from comparisons of similar actions in groups of closely related species. The history of recruitment behavior and chemical trail communication has been examined in ants with such an approach. The ancestral condition is thought to involve a behavior called tandem running. In this mode of recruitment, a single nestmate is led "in tandem" from the nest to a new nest or food source: a "leader" guides a "follower" to a target area. Tandem running involves motor displays that initiate pairing (Fig. 6) and surface pheromones and other exocrine gland secretions to maintain the communicative tie while the ants move pairwise, outbound from the nest. This type of recruitment communication is considered to be basal because it commonly occurs in ponerine species, which are themselves ancestral in the evolutionary history of ants.

In chemical mass communication, excitatory and orienting information is contained in the structure of the trail substance, and no other signals are required to control recruitment activity. In tandem running, behavioral displays alert recruitees to the need for their assistance; chemical signals on the body of the recruiter, as well as physical contact between the members of the tandem pair, provide directional guidance. In the most derived state of recruitment communication, all information required to regulate group action is encoded by pheromones. This type of trail communication is characteristic of many species in the majority of ant subfamilies and in the termites appears to be the only method of trail communication.

The origin of trail communication in termites has received relatively little attention. The secretions of the sternal gland are known to inhibit fungal growth; thus the ancestral function of this gland may have been the production and deposition

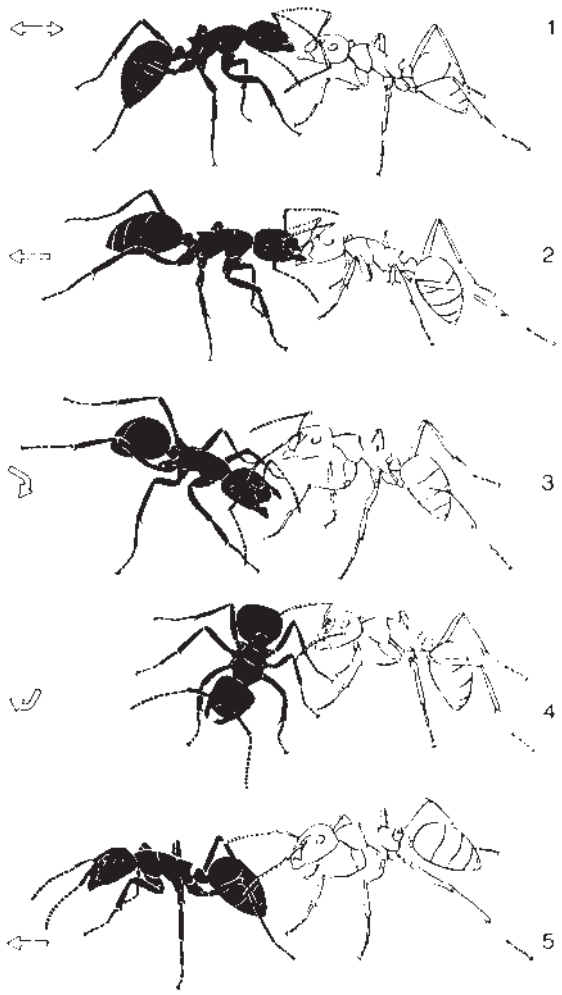

FIGURE 6 Tandem running in ants. A recruiter (black) contacts a nestmate (white) and shows a jerking behavior, pulling the nestmate by the mandibles to invite her to follow (1, 2). The recruiter then turns and offers the gaster (3, 4). When the recruited ant contacts the gaster and hind legs of the recruiter, the pair move in tandem to a new nest site (5). Arrows illustrate the direction of motion of the recruiter. [From Hölldobler, B. *et al.* (1974). Communication by tandem running in the ant *Camponotus sericeus. J. Comp. Physiol.* **90,** 105–127, © Springer-Verlag GmbH & Co. KG.]

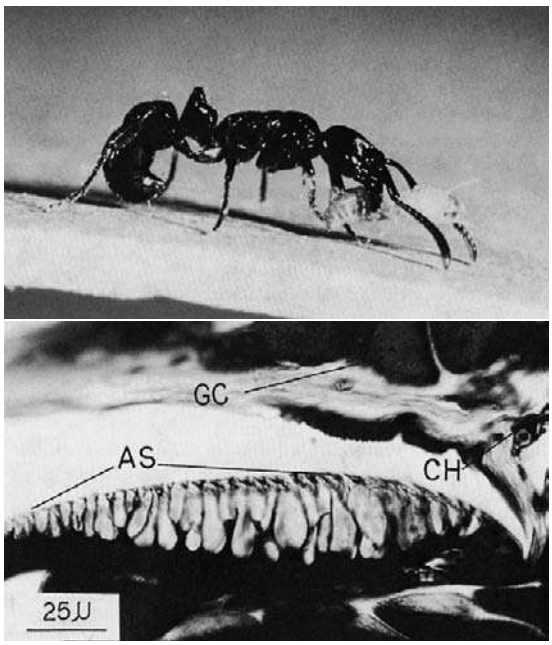

FIGURE 7 The pygidial gland and trail laying. (A) The ant *P. laevigata* applies the pygidial gland, located dorsally, to the substrate to lay a trail. (B) Structure of the gland: AS, cuticular applicator; GC, gland cells, CH, gland channels in the intersegmental membrane. Scale bar = 2.5 mm. [Modified from Hölldobler, B., and Traniello, J. (1980). The pygidial gland and chemical recruitment communication in *Pachycondyla* (= *Termitipone*) *laevigata. J. Chem. Ecol.* **6,** pp. 886, 887, with permission of Kluwer Academic/Plenum Publisher.]

of antibiotics to control microbial growth in the nest. Recruitment communication became more elaborate with the separation of nesting and feeding ecology and the evolution of a sterile worker caste.

Resource Distribution and the Evolution of Recruitment Communication

Some species of ants have no recruitment communication; these species offer important insights into the role of food distribution in the evolution of recruitment behavior. The relative significance of phylogeny and ecology can be separated in some of these species because the absence of recruitment communication has been noted in ants in derived subfamilies. The ant *Cataglyphis bicolor,* a formicine species that nests in the ground and forages on the arid salt pans of North Africa, does not mobilize nestmates to food sources greater in size than what a single worker can transport. Recruitment does not occur naturally and cannot be induced experimentally. The diet of this ant is primarily composed of arthropods that have desiccated; the distribution of these prey is unpredictable in space and time, suggesting that the ecology of prey distribution has been the major selective force for individual retrieval without the possibility of recruitment.

Among the basal ponerine ants, the same point concerning food distribution and phylogeny can be made. Ponerine ants in the genus *Pachycondyla* generally show no food recruitment, employing tandem running only during nest emigration. These ants feed on randomly distributed individual prey. Yet chemical recruitment behavior occurs in species that utilize clumped food resources, such as *P. laevigata,* an obligate termite predator. *P. obscuricornis* huntresses collect prey as individuals but do not have food recruitment communication. Nest relocations involve tandem running; secretions from the pygidial gland, which is located dorsally beneath the seventh tergite on the gaster, hold the tandem pair leader and her follower together as they make their way to the new nest. In *P. laevigata,* an obligate termite predator, scout ants initiate the recruitment of nestmates after only a single termite has been found. Workers of this species apply the potent secretions of their pygidial gland to the substrate by curling

the gaster forward ventrally and dragging the dorsal surface (Fig. 7). One trail-laying ant can induce the formation of a foraging column of several dozen workers. Shifts in diet in the genus therefore appear to have selected for the changes in trail-laying behavior and the chemistry of the pygidial gland secretion. Another basal ant that has well-developed recruitment communication is the ambyloponine *Onychomyrmex,* which is a specialist on large prey and has an army ant-like life cycle.

Resource distribution is associated with the use of chemical communication during foraging in desert seed-harvesting ants. *Pogonomyrmex rugosus* and *P. barbatus* occur sympatrically and feed on seed clumps. These ants have well-developed trail communication. On the other hand, *P. maricopa* collects scattered seeds, primarily through individual retrieval, and has a comparatively narrow diet breadth and a relatively weak recruitment response to seed patches.

RECRUITMENT, COMPETITION, AND COOPERATIVE RETRIEVAL Recruitment communication allows the diet of a species to be expanded to include food items greater in mass than a single forager's load size limit. This is accomplished by recruiting nestmates to help transport prey. In this way, small-bodied ants can cooperate in prey retrieval, sometimes greatly increasing diet breadth. Recruitment communication is flexible; the number of ants in the cooperative retrieval group can be adjusted to prey size and thus allow a colony to efficiently collect resources of different sizes. There may be an energetic advantage to cooperative foraging, reflected in reduced search and retrieval costs and the caloric benefit of successfully acquiring large prey. Recruitment may also improve competitive success; workers recruited to a contested resource can enhance prey defense as well as assist in transport. Short-range recruitment, mediated by acoustical and/or chemical signals, may be used to attract nearby ants to a food find. Such is the case in the foraging organization of the desert ant *Aphaenogaster cockerelli.* Following short-range recruitment, a small group of recruits will cooperatively transport the food to the nest to minimize competition with sympatric ants.

The recruitment response of a colony is influenced by a number of ecological factors that can alter the way in which resources are partitioned between species. Recruitment varies with temperature and food item size, and each species in a community may have different thermal and prey size optima. This will cause competitive ability to vary among species and may result in some species "specializing" on prey of a certain size while foraging within a given temperature range. In open-field north temperate ant communities, the small-bodied *Monomorium minimum* is more tolerant of higher temperatures and can use a mass recruitment pheromone from the Dufour's gland to recruit rapidly to large prey, which workers defend with a chemical repellent originating in the poison gland. *M. minimum* foragers also dissect prey more rapidly than sympatric ants. Temperature preferences, recruitment, prey defense, and dissection together represent a foraging strategy that allows this ant to successfully retrieve large prey.

While the benefit of recruitment is diet expansion, its costs lie in the time required to assemble a cooperative retrieval group. During this time, competing ant species may discover the prey and interfere with its exploitation. Indeed, the probability of interference competition from sympatric ant species can increase from 0% to 100% when prey items exceed the upper size limit for individual worker carriage. The ability to move prey is an important factor in decreasing losses to competitors.

SELF-ORGANIZATION AND RECRUITMENT BEHAVIOR

Theories of foraging strategy note the significance of time, energy, and food profitability in the evolution of feeding behavior. Social insect colonies adjust their foraging (i.e., regulate the number of workers feeding at a given food source) according to food profitability, the risk of worker loss to predation, and competition. Recruitment communication, through the physical properties of trail pheromones, allows foraging adjustments to be made adaptively. It is at the colony level that such adjustments in worker feeding behavior, or decisions, are made. The concept of self-organization has been used to explain how collective action at the level of the colony can be the result of simple rules followed by individual workers. The complex foraging activities of a colony thus result from the interactions of simple workers, whose behavior is determined by trail pheromone concentration.

Self-organization has been used to explain the foraging behavior of mass-recruiting species such as fire ants, which have large colonies composed of workers with limited behavioral repertoires. In addition, their communication systems rely almost entirely on the deposition and diffusion of trail pheromones. Computer simulations and mathematical models have been used to describe the mechanisms involved in dividing foragers between two food sources, as well as the structure of raids of army ants and the rotation of foraging columns of desert seed-harvesting ants. Self-organization theory has also attempted to explain why mass-recruiting ant species have larger colony size: the greater complexity of larger colonies and their efficiency of colony operations may be dependent upon the individual simplicity of the workers that make up these social groups. Individual simplicity may allow greater flexibility in colony-level behavior than would be possible in a social group composed of complex individuals. This reinforces the notion that social insect colonies have a decentralized rather than hierarchical system of control. It has long been known that the queen does not "hand down" orders to workers to direct colony activites. The idea of self-organization can provide detailed descriptions of how efficiency is achieved through decentralization and the chemical regulation of worker behavior.

See Also the Following Articles

Ants • Caste • Dance Language • Division of Labor • Hymenoptera • Isoptera • Orientation • Pheromones • Sociality

Further Reading

Detrain, C., Pasteels, J. M., and Deneubourg, J.-L. (1999). "Information Processing in Social Insects." Birkhäuser-Verlag, Basel.

Hölldobler, B., and Wilson, E. O. (1990). "The Ants." Belknap Press of Harvard University, Cambridge, MA.

Traniello, J. F. A., and Leuthold, R. (2000). The behavioral ecology of foraging in termites. *In* "Termites: Evolution, Sociality, Symbioses, Ecology" (T. Abe, M. Higashi, and D. Bignell, eds.), pp. 141–168. Kluwer, The Hague.

Traniello, J. F. A., and Robson, S. (1995). Trail and territorial pheromones in the social insects. *In* "Chemical Ecology of Insects," Vol. 2 (R. T. Cardé and W. J. Bell, eds.), pp. 241–285. Chapman & Hall, New York.

Regulatory Entomology

Robert V. Dowell

California Department of Food and Agriculture

Regulatory entomology is concerned with preventing the unwanted movement of insects and related invertebrates from areas where they occur to areas where they do not occur. Nonnative species have caused untold amounts of damage to our crops, homes, environments, and persons. Officials in regional, state, or national governments achieve this goal with a five-component prevention program of exclusion, detection, eradication, identification, and public awareness. Exclusion is the primary prevention level. It uses phytosanitary regulations, including quarantines and other legal actions, to prevent the introduction of the quarantine pests. Detection is the secondary prevention level, which uses visual surveys or traps to find infestations of quarantine pests that have penetrated the exclusion barriers. Identification provides the name of the invading organism and information about its biology. It also provides intelligence to assist exclusion and eradication activities. Eradication is the tertiary prevention level. Eradication is the elimination of populations of quarantine pests found by the detection program. Public awareness strives to make the citizens aware of the regulatory laws concerning the importation of quarantine pests and items and enlists their cooperation in making the prevention efforts successful. The components work together much like a ladder with exclusion, detection, and eradication being the rungs, which are held together by identification and public awareness. Working together, these components can prevent the unwanted immigration of quarantine pests.

PEST EXCLUSION

The basic premise of pest exclusion is that it is better to live without pests than to live with them. History provides numerous examples of nonnative species that have caused extensive damage when brought into new areas. Examples of invasive species that have caused damage in nonnative areas include the Formosan termite *(Coptotermes formosanus)* in Hawaii and Louisiana, gypsy moth *(Lymantria dispar)* in the eastern United States, opuntia cactus (*Opuntia* spp.) and rabbits *(Oryctolagus cuniculus)* in Australia, brown tree snake *(Boiga irregularis)* in the Pacific region, American cockroach *(Periplaneta americana)* throughout most of the world, and Colorado potato beetle *(Leptinotarsa decemlinata)* in Europe. Pest exclusion uses phytosanitary regulations, including quarantines and other legal actions, to prevent the immigration of quarantine pests. Quarantines prohibit the movement of the quarantine pest or plants or other items known to be infested or liable to be infested from crossing the quarantine boundaries.

Quarantines may identify specific organisms (e.g., San Jose scale, *Quadraspiditotus perniciosus*) or groups of organisms (e.g., fruit flies in the family Tephritidae). Pest insects attack human food, fiber, shelter, or persons; carry diseases of people or other organisms, including domestic livestock or native biota; or are pests of human or natural environments (Table I).

Plants or other regulated articles are prohibited from crossing the quarantine boundaries unless they meet compliance procedures specified in the phytosanitary regulations that render them free of the quarantine pest. Compliance agreements and permits are authorizations to move the prohibited items under conditions prescribed by the quarantine issuer. Quarantine compliance may be performed at origin, the preferred option, or at destination, which is the less preferred option. The compliance requirements vary from those designed to keep the pest out of the regulated items, such as growing nursery stock in insect-proof screenhouses or applying pesticide sprays while the plants are growing, to those designed to kill any pests associated with the regulated items prior to or during shipment, including fumigation, hot/cold treatments, or irradiation (Table II).

Permits are also used to regulate the importation and holding of live insects for exhibit or research by insect zoos, butterfly houses, public or private research facilities and universities, and private citizens. The regulations that allow the

TABLE I Examples of Quarantine Pests at Local, State, and National Levels

Scientific name	Common name	Reason
Aedes albopictus	Asian tiger mosquito	Human disease vector
Anoplophora malasiaca	White-spotted longicorn bettle	Wood-boring beetle
C. formosanus	Formosan termite	Dwelling pest
Dacus ciliatus	Lesser pumpkin fly	Fruit pest
Danaus plexippus	Monarch butterfly	Butterfly disease carrier
Diaphorina citri	Citrus psyllid	Citrus disease vector
Le. decemlineata	Colorado potato beetle	Vegetable defoliator
Ly. dispar	Gypsy moth	Forest defoliator
Q. perniciosus	San Jose scale	Fruit tree/ornamental pest
Trogoderma granarium	Khapra beetle	Grain pest

TABLE II Examples of Accepted Measures Used to Meet Phytosanitary Regulations

Pest	Origin	Measure used
Wood-boring beetles in dunnage	Asia	Fumigation
Fruit flies (Tephritidae)	Various	Vapor heat, fumigation, cold treatment, pesticide sprays
Thrips (Thysanoptera) on flowers	Hawaii	Irradiation
Caribbean fruit fly *(Anastrepha suspensa)*	Florida	Systems approach
Plum curculio *(Conotrachelus nenuphar)*	Eastern United States	Controlled atmosphere
Corn stalk borers	Eastern United States	Hot water dip
Japanese beetle *(Popillia japonica)* in nursery stock	Eastern United States	Pesticide drench of pots, plants grown in insect-proof screenhouses

insects to be held safely are specified in the permit issued to the person or entity requesting permission to import or hold live insects that are subject to regulation. Another form of pest exclusion uses the continual release of sterile insects throughout an area of high risk of invasion by that pest to preclude establishment. These sterile insects mate with wild adults as they emerge, producing infertile eggs. This preventative release approach stops the invading pest from developing infestations that may require eradication. This preventative release approach is being used against the Mexican fruit fly *(Anastrepha ludens)* along the Mexico–California border and against the Mediterranean fruit fly *(Ceratitis capitata)* in the greater Los Angeles area of California, along the Peru–Chile border, and along the Mexico–Guatemala border.

PEST DETECTION

Pest detection programs serve two purposes: to ensure that a political entity is free of the quarantine pest(s) regulated by their phytosanitary regulations and to find infestations of quarantine pests that have penetrated the exclusion barrier. There are two basic forms of pest detection: traps and visual surveys. The use of traps has many advantages and traps are the preferred method of detection of insect pests. However, not all pests respond to traps or lures. For these pests visual surveys, with their own advantages and disadvantages, are used (Table III).

TABLE III Advantages and Disadvantages of Traps and Visual Surveys Used to Detect Quarantine Pests

Advantages	Disadvantages
Traps	
Can be highly attractive, may detect pest at low densities	Not all pests can be trapped
Operate continuously when in field	Often limited to small number of species
Can be left in field for days to months	
One person can operate many traps	
Can cover large area quickly	
Visual surveys	
Can detect many pests at once	Labor intensive
Can detect any pest insect	Operate only when people are in field
	Cover limited area slowly

Traps may be completely visual, such as yellow traps used to catch aphids and whiteflies, but most often they contain one or more chemicals attractive to the target species (Table IV). These chemical lures include sex pheromones, pheromone precursors, parapheromones (attractant not of botanical origin or a pheromone precursor that attracts males), food lures, chemicals of unknown action, and combinations of visual traps and lures.

Visual surveys can look for the pest itself, such as the white garden snail *(Theba pisana);* the damage the pest can cause, such as that caused by the Asian longhorn beetle *(Anoplophora glabripennis);* or the abode of the pest, such as the galls of the balsam gall midge *(Paradiplosis tumifex).* Visual surveys may use food lures to attract and hold the pest as is done with the red imported fire ant *(Solenopsis invicta).*

Detection efforts are concentrated in those areas where the target pest might enter and where the pest may become established. In general, large urban areas have a greater number of invading exotic species than rural or agricultural areas. This is the result of people directly bringing in exotic invertebrates in the fruit, flowers, and other items that they obtain while traveling and the importation of large volumes of goods (food, plants, etc.) that are needed to service large population centers (Table V).

If a pest is found during a detection program, a delimitation survey is used to determine whether an infestation exists

TABLE IV Examples of Lures Used in Pest Detection Programs

Lure	Type	Pest
Disparlure	Pheromone	Gypsy moth
Methyl eugenol	Pheromone precursor	*Bactrocera* fruit fly species
Trimedlure	Parapheromone	*Ceratitis* spp. fruit flies
Cuelure	Unknown	*Bacterocera* fruit fly species
Cracked grain	Food	Khapra beetle
Decaying protein in water	Food	*Anastrepha* fruit fly species
Ammonium carbonate and yellow traps	Food lure/foliage mimic	*Rhagoletis* fruit fly species

TABLE V Examples of Locales in Which Detection Efforts Are Conducted

Locale	Pest
Urban areas	Many pests, including fruit flies
Campgrounds	Gypsy moth
Corn fields	European corn borer *(Ostrinia nubilalis)*
Slate yards	Snails and lygaeid bugs
Ports	Wood-boring beetles
Airports	Japanese beetle
Nurseries	Many pests, including scale insects
Almond orchards	Red imported fire ant *(S. invicta)* in palletized loads of honey bee colonies
Cotton fields	Boll weevil *(Anthonomus grandis grandis)*

and if so, its physical boundaries. Unlike detection efforts that try to uniformly cover an area, delimitation efforts strongly target the area immediately around where the pest was found, with decreasing effort as one moves away from that point. For example, Mediterranean fruit fly detection in urban areas of California deploys a uniform 4 Jackson traps baited with Trimedlure per square kilometer. For delimitation, 100 Trimedlure-baited Jackson traps are deployed in the 2.56 km^2 area (core area) centered on the site where the first fly was found. An additional 50, 25, and 20 Trimedlure-baited Jackson traps per 2.56 km^2 are deployed respectively in three 1.6-km-wide rings around the core area.

Delimitation surveys are conducted until an infestation is confirmed and a decision about further actions is made or until it is determined that no infestation exists. The latter requires that no more target insects be found for a time period equal to several generations, typically two to three, of the pest. At the end of the delimitation effort, a trapping or visual survey program resumes at the detection level.

Delimitation efforts using visual surveys follow the same format, with a greater effort expended in the area immediately around the site where the pest was detected and diminishing effort as one moves away from that point.

PEST IDENTIFICATION

An important part of any detection program is the rapid and accurate identification of organisms that are found. Professional scientists provide these identifications and background information on the biology of the organism, which are used to help decide what, if any actions, may be taken against the pest. These professionals also provide information used to develop phytosanitary regulations and quarantine pest detection and delimitation programs.

PEST ERADICATION

Eradication of infestations of exotic pests is the most controversial aspect of regulatory entomology. Eradication is designed to eliminate a pest from a proscribed area, usually within a given time. Eradication programs are conducted at the point where the pest is found, not necessarily where it may do the most damage. Eradication programs using pesticide sprays in urban settings often generate considerable public outcry and opposition. Programs using nonpesticidal tactics or less intrusive methods of pesticidal applications engender little to no public concern. Thus the aerial application of pesticides or the extensive use of ground applications of pesticides in urban areas is strongly contested by the public, whereas mass trapping or male annihilation programs in the same areas are ignored. In general, successful eradication programs have the following components: the organism poses a clear-cut threat, an effective detection technique is available—usually not only visual surveys, the organism is limited in its distribution in the newly invaded area, continuous natural invasion of the organism cannot occur, and effective techniques exist to reduce the target population below the point at which reproduction can occur (Table VI).

TABLE VI Examples of the Eradication Tactics Used against Invertebrate Pests

Pest	Tactic[a]
Japanese beetle	Cover sprays of pesticides
Olive fly *(Bactrocera oleae)*	Pesticide and bait sprays
Mediterranean fruit fly	Use of sterile males
Melon fly *(Bactrocera cucurbitae)*	Mass trapping using Cuelure and an insecticide
Oriental fruit fly *(Bactrocera dorsalis)*	Male annihilation using methyl eugenol and an insecticide
Asian longhorned beetle	Host removal
Gypsy moth	Cover sprays of microbial insecticides
Boll weevil/pink bollworm *(A. grandis grandis/Pectinophora gossypiella)*	Cultural controls such as specific plow-down dates

[a]One or more tactics are often used against a pest.

Eradication procedures are applied to the target population for several life cycles of the organism beyond the last individual found. Posttreatment monitoring at delimitation levels is conducted for at least one additional life cycle of the pest. If no further individuals of the target organism are found within the treated area, eradication is considered to have been successful.

PUBLIC OUTREACH

Public outreach is conducted in a number of venues. Among the most frequently encountered are the forms that are required to be filled out before leaving a plane that has landed at a foreign airport or crossing a border. Information flyers handed out during eradication programs, on cruise ships, and at other sites are another venue through which the public is made aware of the exclusion efforts conducted on their behalf. Regular inspections of products moving into or

out of an area keep professional importers and exporters aware of the quarantine regulations of their trading partners.

The five-component pest exclusion program described above, with modifications for local weather, topography, target species, and so on, is used by all countries that adhere to the International Standards for Phytosanitary Standards published by the Food and Agriculture Organization of the United Nations.

See Also the Following Articles

Agricultural Entomology • Extension Entomology • Fire Ants • Gypsy Moth • Sterile Insect Technique

Further Reading

Collard, S. B., III (1996). "Alien Invaders." Grolier Press, Danbury, CT.

Dowell, R. V., Siddiqui, I. A., Meyer, F., and Spaugy, E. L. (2000). Mediterranean fruit fly preventative release programme in southern California. *In* "Area-wide Control of Fruit Flies and Other Pests" (K.-H. Tan, ed.), pp. 369–376. Penerbit Universiti Sains Malaysia, Pulau, Pinang.

Hancock, D. L., Osborne, R., Boughton, S., and Gleeson, P. (2000). Eradication of *Bactrocera papayae* (Diptera: Tephritidae) by male annihilation and protein baiting in Queensland, Australia. *In* "Area-wide Control of Fruit Flies and Other Pests" (K.-H. Tan, ed.), pp. 381–388. Penerbit Universiti Sains Malaysia, Pulau, Pinang.

"Harmful Non-indigenous Species in the United States" (1993). Office of Technology Assessment, U.S. Govt. Printing Office, Washington, DC. [OTA-F-565].

"Insect Trapping Guide" (1998). California Department of Food and Agriculture (www.cdfa.ca.gov).

"International Standards for Phytosanitary Measures" (2000). Food and Agriculture Organization of the United Nations (www.fao.org).

Pegram, R. G., Gersabeck, E. F., Wilson, D. D., and Hansen, J. W. (2000). Progress in the eradication of *Amblyomma variegatum* Fabricius, 1794 (Ixodoidea, Ixodidae) from the Caribbean. *In* "Area-wide Control of Fruit Flies and Other Pests" (K.-H. Tan, ed.), pp. 123–130. Penerbit Universiti Sains Malaysia, Pulau, Pinang.

"Plant Quarantine Manual" (1998). California Department of Food and Agriculture (www.cdfa.ca.gov).

"Quarantine Pests for Europe" (1997). CAB Int. and European and Mediterranean Plant Protection Organization, University Press, Cambridge, UK.

Villasenor, A., Carrillo, J., Zavala, J., Stewart, J., Lira, C., and Reyes, J. (2000). Current progress in the medfly program Mexico–Guatemala. *In* "Area-wide Control of Fruit Flies and Other Pests" (K.-H. Tan, ed.), pp. 361–368. Penerbit Universiti Sains Malaysia, Pulau, Pinang.

Wyss, J. H. (2000). Screw-worm eradication in the Americas—Overview. *In* "Area-wide Control of Fruit Flies and Other Pests" (K.-H. Tan, ed.), pp. 79–86. Penerbit Universiti Sains Malaysia, Pulau, Pinang.

Reproduction, Female

Diana E. Wheeler
University of Arizona

In female insects, reproduction generally involves producing yolky eggs, mating, and then laying fertilized eggs. Across the diversity of insects, however, different ways of reproducing illustrate an astounding variation in this simple series of events as well as divergence from it. In the most extreme examples, females can reproduce without supplying eggs with yolk, without mating, and even without laying eggs.

Female reproduction has been one of the most intensively studied aspects of insect biology in the past 50 years for two reasons. First, frequent confrontations between humans and insects in the arenas of agriculture and health make understanding insect reproduction of great practical importance. Production of the next generation of insects has several steps that are centered on the female. To reproduce, females need to make eggs or provide their embryos with nutrition in other ways. Once made, females must find an appropriate spot to deposit their eggs. For entomologists concerned with problem insects, these steps offer opportunities to disrupt reproduction and reduce the number of insects in the next generation. Second, the diversity of ways that insects reproduce provides a rich source of material for discovering the underlying rules of biology. For example, the extraordinary effectiveness of female insects in converting resources into eggs led to their use as an intensively studied model system. The process by which yolk is taken up into insect eggs serves as a model for how cells take up large molecules from the surrounding environment.

STRUCTURE OF FEMALE REPRODUCTIVE SYSTEMS

Female insects can make eggs, receive sperm, store sperm, manipulate sperm from different males, and lay eggs. Their reproductive systems are made up of a pair of ovaries, accessory glands, one or more spermathecae, and ducts connecting these parts. Ovaries make eggs, and accessory glands produce substances to help package and lay the eggs. Spermathecae store sperm for varying periods of time and, along with portions of the oviducts, can control sperm use. The ducts and spermathecae are lined with cuticle.

The ovaries are made up of a number of egg tubes, called ovarioles. The number of ovarioles varies with the type of insect, its size, and its particular life history. Clearly, the number of ovarioles and the number of eggs that can be produced by each set an upper limit to the total number of eggs, or young, an insect can produce. The rate that eggs can develop is also influenced by ovariole design. In meroistic ovaries, the eggs-to-be divide repeatedly and most of the daughter cells become helper cells for a single oocyte in the cluster. In panoistic ovaries, each egg-to-be produced by stem germ cells develops into an oocyte; there are no helper cells from the germ line. Production of eggs by panoistic ovaries tends to be slower than that by meroistic ovaries.

Accessory glands or glandular parts of the oviducts produce a variety of substances for sperm maintenance, transport, and fertilization, as well as for protection of eggs. They can produce glue and protective substances for coating eggs or tough coverings for a batch of eggs called oothecae.

Spermathecae are tubes or sacs in which sperm can be stored between the time of mating and the time an egg is fertilized. Paternity testing of insects has revealed that some, and probably many, female insects use the spermatheca and various ducts to control or bias sperm used in favor of some males over others.

COMMON AND DIVERSE MODES OF REPRODUCTION

The most common mode of reproduction in insects is by yolked eggs, fertilized internally, that are laid outside the body. However, other modes of reproduction are not uncommon. Insects with unusual and even unique modes of reproduction are interesting as examples of extreme biological forms and processes. Modes of reproduction vary in three important aspects: whether eggs are fertilized, whether eggs are provisioned, and where embryonic development takes place.

First, reproduction without fertilization, or parthenogenesis, is common in insects as a normal means of reproduction in addition to or instead of sexual reproduction. For example, the system of sex determination in the Hymenoptera relies on parthenogenetic production of males. Virtually the entire order produces males from unfertilized haploid eggs and female from fertilized diploid eggs. This type of sex determination, called haplodiploidy, is also found in some Sternorryncha, Thysanoptera, and Coleoptera.

In another common type of parthenogenesis, germ stem cells in the ovary do not go through meiosis before they start development. As a result, offspring are clones of the mothers, having a full copy of her genes. Aphids are the premier example of this type of parthenogenesis. In a third type, meiosis occurs but the diploid number of chromosomes is restored by fusion of two nuclei. This type of parthenogenesis has been particularly well studied in stick insects (Phasmida). A fourth type requires mating, but fertilization is not completed. Sperm is necessary for development to begin, but the male's genes are discarded. Sperm-dependent parthenogenesis is found in some bark beetles.

A second important factor that defines the mode of development is the amount of yolk material provided to the egg. Amount of yolk can vary from none to more than enough to complete development. When eggs are laid without provisioning, nutrition must be obtained from elsewhere. Females of some insects, such as aphids, can keep developing embryos in their bodies and supply them directly with needed nutrients. In contrast, some parasitic and parasitoid wasps (e.g., Trigonalidae, Braconidae) and flies (Tachnidae) produce tiny eggs lacking yolk and lay them inside other insects that can provide for them.

A third important descriptor of developmental mode is the site of embryonic development. Most eggs are laid outside the mother's body but, in ovoviparous insects, eggs can be retained inside the body of the mother where the embryos develop. At hatching, they are released to the outside world. In true viviparity, embryos also develop in the mother's body but there is no intermediate egg stage. The site of development of parasitoids is similar to that in viviparous insects, in that it is inside another insect. Development of parasitoids, however, takes place in the bodies of host insects, rather than in the mother.

MANAGEMENT OF NUTRIENTS FOR EGGS

In the common mode of reproduction by yolked eggs, female insects accumulate large amounts of macronutrients, especially protein and fat. Lipids are usually derived from carbohydrates in the diet and are not generally in short supply. Amino acids, particularly essential amino acids, can be limiting. Therefore these are a particularly important part of the yolk.

Eggs may be provisioned with nutrients obtained in either the larval or the adult stage or both. Insects that do not feed as adults and have their lifetime's egg production completed when they become adults can draw only on larval nutrients. Even insects that feed after eclosion can use excess larval nutrients to provision eggs.

Mosquitoes and other blood feeding Diptera provide examples of the often interlocking roles of larval and adult feeding. Female mosquitoes feed on nectar and vertebrate blood to obtain nutrients for egg production. In some mosquitoes, however, food eaten during the larval stage supports the production of at least some eggs. The ability to produce eggs without blood feeding is called autogeny. Some autogenous species can mature their eggs only this way and have lost the ability to feed on hosts. Other species are more flexible and can use leftover reserves, if they are available, but can feed on blood immediately if they are not. In addition, when species occur over a broad geographical area, they can be locally adapted. The pitcherplant mosquito, *Wyeomyia smithii,* is completely autogenous in the northeastern United States, where larval densities are low and food more abundant, but must feed on blood in the southeast, where larval resources are more scarce.

Aspects of both larval and adult environments can favor autogeny. A larval environment that offers more consistent resources than the adult one will favor obligate autogeny, whereas more predictably abundant food in the adult environment will favor obligate blood feeding. In the Northern Hemisphere, autogeny becomes more common toward the arctic, where host vertebrates are less abundant. Short-term variability in nutritional resources, either in a patchy spatial environment or over time, can make physiological flexibility, termed facultative autogeny, a better strategy than obligate autogeny.

Males can be an important source of nutrients especially when the female's resources are limited. Nutritional contributions are especially conspicuous in Orthoptera, which transfer large spermatophores that can weigh over 20% of the male's body weight. In a variety of insects, proteins transferred to females during mating have been found in their eggs, ovaries, blood, and various body parts.

TRANSFER OF SPERM AND POSTMATING MANAGEMENT

Fertilization of eggs is the focus of sexual selection. Natural selection biases survival toward individuals that are the most successful in their environment. Sexual selection biases its rewards toward those individuals that enhance the success of their own genotype in the next generation through any aspect of mating and subsequent fertilization.

Classically, males compete with each other for access to females, and females can choose to mate with them or not. Because insects have internal fertilization, there are many possibilities for females to manipulate sperm after mating. When females mate multiple times, male competition can take place between sperm in the female reproductive tract. Overlooked for many years, it is now known that the female may also have control over which male's sperm will fertilize her eggs. Mechanisms that females use to bias paternity include active elimination of sperm, digestion of sperm, lack of sperm transport to the spermatheca, and decreased use of some sperm batches in fertilization. Postmating female choice is often called cryptic female choice, although it is no more cryptic than postmating competition between sperm.

Females are generally believed to choose their mates based on features that indicate their quality as parents. For example, a choice can be made based on behaviors such as gifts of food or some other indicator of resources, on the amount of sperm transferred, or on the concentration of a protective chemical given to the female. In addition to features that indicate success under the rules of natural and sexual selection, females may also choose sperm or mates based on genotypes that are most complementary with their own.

Highly social insects, especially those in perennial colonies, provide an interesting variation in female reproductive strategies. For example, to produce enough worker insects, queens must have a large supply of sperm. Only a small proportion of total sperm is used to produce new queens and males; most of it is used to make more workers. In social Hymenoptera, which includes ants, bees, and wasps, males can be produced without sperm because males develop from unfertilized eggs. As far as is known, new queens mate only before starting or joining a colony and must, therefore, at that time store as much sperm as they will need for the remainder of their lives. One important factor that affects colony longevity is the amount of sperm available for fertilizing eggs. Queens of large leafcutting ants *(Atta)* mate many times and store as many as 250 million sperm. In contrast, termites do not have to store large amounts of sperm because termite colonies are headed by both a queen and a king, and mating takes place throughout their lives.

Ants, the most speciose of the social insects, have a wide range of colony sizes and rates of worker production. Across the group, the number of sperm stored is correlated with the number of ovarioles, which suggests that there is a cost to long-term sperm storage and that sperm storage is matched to the lifetime needs of a successful colony queen.

OVIPOSITION

The culmination of insect reproduction is typically the oviposition of mature, fertilized eggs in an environment that will support their development. Eggs can be placed on surfaces, in crevices, in soil, and in animal or plant tissue, and accessory glands produce secretions used to protect eggs. Females often have a structure called an ovipositor, which is made up of modified appendages on the last abdominal segments and serves as a tool for penetrating substrates. For example, many grasshoppers use their ovipositors to dig into soil, where they lay eggs in a frothy matrix. Also, some parasitic wasps have long, needle-like ovipositors that can drill through wood to reach host insects.

Oviposition can be linked physiologically to prior events in egg maturation. For example, insects that mature eggs throughout the adult stage can often adjust their egg production based on available oviposition sites as well as on available nutrition. When oviposition sites are rare, unlaid eggs can inhibit the hormonal control network that guides egg maturation. Delayed oviposition can go further than slowing egg production and lead to resorption of eggs, which reallocates resources away from reproduction and toward survival.

Oviposition sites also provide sensory cues that can stimulate further egg maturation. For example, egg production in newly eclosed females in the diamondback moth, *Plutella xylostella,* is accelerated by the presence of single volatile components of host cabbage plants. Ovarian response to oviposition and related cues can reduce the opportunity costs females suffer when the supply of completed eggs does not match the availability of oviposition sites.

Protection of eggs, particularly those laid in clusters, can be extended by parental care. For example, egg guarding reduces the risk of eggs being used by parasitoids or eaten by predators.

See Also the Following Articles

Accessory Glands • Eggs • Ovarioles • Parthenogenesis • Sexual Selection • Spermatheca • Vibrational Communication • Vitellogenesis

Further Reading

Chapman, R. F. (1998). "The Insects: Structure and Function," pp. 298–312. Cambridge University Press, Cambridge, UK.

Eberhard, W. G. (1996). "Female Control: Sexual Selection by Cryptic Female Choice." Princeton University Press, Princeton, NJ.

Engelmann, F. (1999). Reproduction in insects. *In* "Ecological Entomology" (C. B. Huffaker and A. P. Guttierez, eds.), 2nd ed., pp. 123–158. Wiley, New York.

Leather, S. R., and Hardie, J. (eds.) (1995). "Insect Reproduction." CRC Press, Boca Raton, FL.

Nijhout, H. F. (1994). "Insect Hormones," pp. 147–156. Princeton University Press, Princeton, NJ.

Papaj, D. R. (2000). Ovarian dynamics and host use. *Annu. Rev. Entomol.* **45,** 423–448.

Thornhill, R., and Alcock, J. (2001). "The Evolution of Insect Mating Systems." iUniverse.com.

White, M. J. D. (1973). "Animal Cytology and Evolution." Cambridge University Press, London.

Reproduction, Female: Hormonal Control of

Diana E. Wheeler
University of Arizona

Female reproduction has been a major focus of entomological research for the past century, driven by the need to control populations of insect pests. The central process in female reproduction in insects, the production of eggs, is hormonally regulated. To reproduce successfully, females must coordinate egg production with other aspects of reproduction such as dispersal, the availability of resources, and selection of mates and oviposition sites. Environmental signals are effectively translated into physiological processes by networks of hormonal signals.

Egg development in insects has become a model experimental system studied to understand the general principles of stage-, sex-, and tissue-specific responses to hormones. In insects, juvenile hormone (JH) and ecdysone typically play important roles in orchestrating egg development. The development of improved analytical methods has led to the elucidation of the roles of other key hormones, particularly a variety of neurosecretory hormones. Common themes in the hormonal control of egg production are becoming clearer, as are important differences between insect groups.

PATTERNS OF FEMALE REPRODUCTION

To reproduce successfully, female insects must coordinate feeding, mating, and locating places to lay their eggs. Mating and egg laying almost always take place in the adult stage, but female insects vary widely in the manner in which they feed and accumulate nutrients for egg production. The accumulation of nutrients for eggs takes place during larval or nymphal development, as well as during adult life.

Eggs are generally filled with bulk nutrients to support embryonic development, and so it is not surprising that the timing of egg production is often tightly linked to the timing of feeding. The earlier nutrients required for egg production are eaten, the earlier egg production can begin. Starting egg production during the immature stage and completing it before eclosion allows female insects having very short adult lives to focus on mating and laying eggs. In fact, adults in many insects do not feed at all and even lack mouthparts. Mayflies are an example. At the other end of the spectrum, insects may take only sufficient nutrients during the larval stage to support their own larval or nymphal development. As adults, they must compensate for the lack of stored reserves by additional feeding. Female mosquitoes that emerge ready to find a blood meal are one example of this pattern. The relative advantages of these contrasting strategies depend on the opportunities for acquiring nutrients, as well as other ecological factors that exist in both the immature and adult habitats. When egg development is telescoped into preadult stages, hormonal signals that control larval or nymphal development and metamorphosis must be coordinated with those of ovarian and reproductive development. Not surprisingly, setting the timing of egg development to different points relative to preadult development requires different hormonal controls.

Female insects that feed as adults differ in the size and number of their meals. At one extreme, females feed fairly continuously. At the other extreme, females take very large meals and have long periods of fasting between them. In blood-feeding insects particularly, a single blood meal can supply the bulk of nutrients necessary for a batch of eggs. Hormones signal the results of feeding to ensure that egg development is matched to an adequate supply of nutrients.

Egg development begins in the germarium, when stem germ cells produce daughter cells that become oocytes. Depending on the type of egg development, oocytes occur alone or in association with sister nurse cells, and follicle cells surround either the oocyte or the oocyte–nurse cells complex. Both nurse cells and follicle cells make and transfer materials important for future embryonic development to developing oocytes. In addition, vitellogenins, fats, and carbohydrates are all taken up into the egg, mostly from the blood. Most or all of the vitellogenin is made by the fat body, a complex organ analogous to the vertebrate liver. Fat body cells release yolk proteins into the blood, from which they are taken up by rapidly growing eggs. When uptake of nutrients is complete, the follicle cells secrete egg coverings in preparation for oviposition.

Hormonal control of egg production is well understood in only a few species. Nevertheless, enough is known about a wide variety of species to infer some general patterns. Four check points in ovarian development are commonly regulated by hormonal signals: the formation of new oocytes by stem cells in the germarium, initial growth of the oocyte, vitellogenin synthesis by the fat body, and vitellogenin uptake by the oocyte. The hormonal signals that break these checkpoints reflect various aspects of the external environment and internal conditions.

HORMONES

The two major hormones that control female reproduction, ecdysone and JH, also control preadult development and metamorphosis. In adult females, ecdysone is produced by ovaries, rather than by the prothoracic glands as in nymphs and larvae. JH is produced by a pair of glands called the corpora allata in both preadult and adult stages. The corpora allata are located near the brain and are connected to it by tracts of neurosecretory cells. Neurosecretory hormones made in the brain and in other parts of the central nervous system are also important controlling factors. Nerves are also part of control networks, especially in relaying sensory information to the brain cells, which affect the hormone-producing organs.

EXAMPLES OF HORMONAL CONTROL OF EGG PRODUCTION

Hemimetabolous Insects That Feed Continuously

Migratory locusts, *Locusta migratoria,* feed almost continuously if they can, and females produce and lay eggs in batches. The importance of JH in controlling egg production in this species has long been recognized. Shortly after eclosion, JH levels rise and stimulate synthesis of vitellogenin by the fat body. Mating and plant odors also affect JH level through neurosecretory cells in the brain. JH alone, however, is insufficient to complete a batch of eggs. An ovary-maturing parsin (OMP), which is a neurohormone, is required, in addition to JH, to stimulate sufficient vitellogenin synthesis and uptake.

In the viviparous Pacific beetle cockroach, *Diploptera punctata,* JH also regulates vitellogenin synthesis and uptake. Signals associated with mating and pregnancy result in increased and decreased levels of JH, respectively. Mating stimulates the release of JH through mechanical stretch of the reproductive tract. The signal is transmitted to the brain by nerves. After mating, females produce a set of eggs and retain them in a pouch off the oviduct called the brood chamber until embryonic development is complete. After the hatching and deposition of nymphs, the female can reproduce again. As in many insects, lack of sufficient nutrients causes JH levels to fall, which delays production of the next batch of eggs.

Hemimetabolous, Blood-Feeding Insects

The bloodsucking bug *Rhodnius prolixus* takes blood meals throughout its life. For nymphs, blood meals are required for growth and molting; for adults they are needed for egg production. In adult females, abdominal stretching associated with feeding stimulates the release of a peptide hormone from the thoracic ganglion. As a result, the corpora allata release JH, which causes vitellogenin synthesis by the fat body and vitellogenin uptake by growing oocytes. Only the terminal, largest oocyte in each ovariole develops, however. Nerves that stretch as the ovaries grow secrete a neurohormone, called an oostatic hormone, that prevents smaller oocytes from taking up vitellogenin. The result is a synchronously produced batch of eggs. As the eggs mature, the ovary secretes ecdysone, which triggers the release from the brain of a neurohormone that stimulates contractions for laying the eggs.

Lepidoptera: Holometabolous Insects That Vary in Timing of Egg Production

Lepidoptera are holometabolous insects, which means that preadults (larvae) and adults differ greatly in form, function, and ecology. Larvae are specialized for feeding, and adults are specialized for dispersal and reproduction. Despite the tidy appearance of such discrete stages, an extensive part of egg production can occur before the adult stage. Enough Lepidoptera species have been studied to allow some appreciation of the variation in egg production patterns across the group. Development and egg production are orchestrated by the same toolbox of hormones, so it is not surprising to find that Lepidoptera in which development and egg production occur simultaneously have control networks different from those in which these processes are sequential.

At one end of the spectrum, females complete egg production before eclosion. Adults of these species do not feed; they mate immediately after eclosion and are short-lived. Examples are silkworms *(Bombyx mori)* and gypsy moths *(Lymantria dispar).* JH does not stimulate egg production and can even inhibit some aspects of it. Pulses of ecdysone associated with pupal development apparently stimulate vitellogenin synthesis and uptake.

In some moths, only part of egg production is completed before eclosion. Examples include pyralid moths such as the Indian meal moth *(Plodia interpunctella)* and the southwestern corn borer *(Diatraea grandiosella).* Synthesis and uptake of vitellogenin take place before the molt to adult, and egg coverings are added afterward by the follicle cells. The portion of egg development that occurs after eclosion can be controlled by JH.

Finally, in many Lepidoptera, such as the monarch butterfly *(Danaus plexippus),* development and egg production are sequential and nonoverlapping, and hormones regulate egg development during the adult stage. In these insects, JH typically stimulates both vitellogenin synthesis and uptake.

Holometabolous, Blood-Feeding Insects

Hormonal control of egg production has been studied extensively in *Aedes aegypti,* the yellow fever mosquito, spurred by the ability of these and other mosquitoes to transmit diseases. Most mosquitoes must drink blood to obtain sufficient amino acids to make eggs. Females in a few species of mosquitoes have the ability to carry over reserves from the larval stage to make part or all of entire first batch without blood feeding.

Eclosion of female mosquitoes is usually followed closely by an increase in JH titer. The rise in JH induces target tissues to become responsive to hormonal signals that occur later. Specifically, the ovaries become sensitive to ovarian ecdysiotropic hormone (OEH), the primary oocyte grows slightly, and the fat body becomes responsive to ecdysone. The posteclosion peak of JH also causes behavioral changes: females search first for mates and then for hosts. Egg development is blocked until a blood meal can be taken. This checkpoint ensures that the intensive metabolic activity required for making eggs will not occur unless enough nutrients are available. When the mosquito has found a sufficiently large blood meal or series of meals, OEH is released into the blood by neurosecretory cells in the brain. OEH stimulates the ovary to produce ecdysone, which then stimulates the fat body to synthesize vitellogenin and the ovary to take it up. The ecdysone peak also causes new, secondary oocytes to separate from germaria, the first step in preparing for the next cycle of

egg production. Growth of the primary oocyte ends at about 36 h after blood feeding. JH begins to rise again, and this stimulates the secondary oocytes to grow slightly and become ready for the next round of egg production. Finally, the eggshell is produced, the eggs are laid, and the mosquito is then ready to seek another blood meal.

Stimulation of vitellogenin synthesis by ecdysone has been studied at the molecular level. Analyses of the nucleotide sequence upstream of the vitellogenin gene show several broad regions of control. First, closest to the gene, is a binding site for ecdysone bound to its receptor complex. Binding to this site is necessary for expression. Second, more distally, are binding sites for two transcription factors, E74 and E75, that are known to be expressed quickly in response to an ecdysone signal. Therefore, the regulatory region of the gene has both direct and indirect interactions with ecdysone.

Holometabolous, Feeding Continuously

In contrast to mosquitoes, the fruit fly *Drosophila melanogaster* feeds continuously, and eggs develop serially as resources become available. The chain of oocytes in each ovariole includes a range of developmental stages. A balance between JH and ecdysone levels controls how many eggs produced.

In *D. melanogaster,* yolk proteins are made not only by the fat body but also by ovarian follicle cells. JH and ecdysone stimulate synthesis in the fat body, but only JH stimulates follicle cells.

JH appears to enhance egg maturation, whereas ecdysone inhibits it. The antagonism in the effects of the two hormones can be seen by manipulating the amount of food and mating. When food is withheld, JH level drops relative to ecdysone levels, and oocytes on the verge of taking up yolk proteins die. In addition, formation of new oocytes is reduced as more die than progress. Providing sufficient food again increases JH levels and reduces cell death. Mating enhances oocyte progression. Sex peptide, a hormonelike substance in seminal fluid, stimulates a rise in JH. Therefore, the increase in egg production seen after mating seems to act through the same hormonal mechanism used in nutrition. In this way, a sexually mature virgin female will not produce new eggs if she lacks either food or a mate. The two points of cell death, at separation of oocytes from the germarium and at the initiation of yolk protein uptake, correspond with similar checkpoints in mosquitoes.

HORMONAL LINKS BETWEEN EGG PRODUCTION, RECEPTIVITY, OVIPOSITION, AND PARENTAL CARE

Hormonal control of egg production in insects can be linked to other activities or states related to reproduction such as receptivity to mating, feeding, oviposition, and migration. In mosquitoes and *D. melanogaster,* receptivity and feeding are associated with an increase in JH level, which triggers early events in the production of eggs. In other insects, however, JH can be inhibitory or have no effect.

Oviposition, too, can be under hormonal control. Oviposition by *L. migratoria* requires extreme extension of the abdomen. Muscles in the appropriate segments respond to JH by altering their contractile properties so that they can function when greatly extended.

Adult insects are often specialized for migration in addition to reproduction. These two functions can create a conflict between protein needs for flight muscles and eggs, and between lipid and carbohydrate needs for flight fuel and eggs. Hormonal coordination can minimize such physiological conflicts. Both the large milkweed bug, *Oncopeltus fasciatus,* and the boll weevil, *Anthonomus grandis,* use JH to stimulate egg production and long-duration flights. Stimulation of flight apparently requires low levels of hormone, whereas higher levels stimulate oogenesis but inhibit flight. High levels of JH in some insect species lead to both yolk uptake and breakdown of flight muscle.

Burying beetles *(Nicrophorus orbicollis)* use carcasses of small vertebrates for their own food, for an oviposition site, and as food for their offspring. Immediately after females have found a suitable dead body, their JH levels rise and egg production is stimulated. JH cannot, however, stimulate yolk uptake in the absence of a carcass, indicating that other factors are important in integrating the discovery of the food source and oviposition site with egg production. Juvenile hormone surges even higher in females when the larvae hatch. Egg production is not stimulated by this rise, which is necessary to stimulate and maintain parental care of developing larvae.

SUMMARY

Female reproduction in insects is controlled in diverse ways. The most fundamental process, egg production, is linked to other aspects of reproduction and reproductive behaviors by hormonal controls. Common themes are the central roles of juvenile hormone and ecdysone, with control points modulated by neurohormones. Hormonal control of egg production is relatively well understood in a few insects because of its use as a model system for studying hormone action. The integration of hormonal controls with other aspects of reproduction is more complex and is less well understood.

See Also the Following Articles

Ecdysteroids • Juvenile Hormone • Mating Behaviors • Migration • Vitellogenesis

Further Reading

Archives of Insect Biochemistry and Physiology, **35 (4)** (1997). Special issue: Juvenile Hormone Revisited: Dynamics of Its Involvement in Reproductive Physiology and Behavior of Insects.

Chapman, R. F. (1998). "The Insects: Structure and Function," pp. 298–312. Chapman & Hall. London.

Engelmann, F. (1970). "The Physiology of Insect Reproduction," pp. 4–7, 45–49. Pergamon Press. Oxford, U.K.

Hagedorn, H. H. (1985). The role of ecdysteroids in reproduction. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," Vol, 8 (G. A. Kerkut and L. I. Gilbert, eds.), pp. 205–262. Pergamon Press, Oxford, U.K.

Koeppe, J. K., Fuchs, M., Chen, T. T., Hunt, L. M., Kovalick, G. E., and Briers, T. (1985). The role of juvenile hormones in reproduction. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," Vol, 8 (G. A. Kerkut and L. I. Gilbert, eds.), pp. 166–204. Pergamon Press, Oxford, U.K.

Nijhout, H. F. (1994). "Insect Hormones," pp. 147–156. Princeton University Press, Princeton NJ.

Raikhel, A. S. (2002). Vitellogenesis of disease vectors, from physiology to genes. *In* "Biology of Disease Vectors." 2nd ed. Harcourt Academic.

Soller, M., Bownes, M., and Kubli, E. (1999). Control of oocyte maturation in sexually mature *Drosophila* females. *Dev. Biol.* **208,** 337–351.

Wheeler, D. E. (1996). The role of nourishment in oogenesis. *Annu. Rev. Entomol.* **41,** 407–431.

Wyatt, G. R., and Davey, K. G. (1996). Cellular and molecular actions of juvenile hormone. II. Roles of juvenile hormones in adult insects. *Adv. Insect Physiol.* **26,** 1–155.

Reproduction, Male

Marc J. Klowden
University of Idaho

Insect reproduction is accomplished with both enormous numbers and an elegant simplicity. In partnership with the female reproductive system, the male reproductive system of insects produces haploid gametes, the spermatozoa, that fertilize the haploid female gamete, the oocyte. Sperm develop in the testes and, when mixed with secretions from accessory glands, travel through the female genital tract when the insects mate, to be stored for later egg fertilization.

STRUCTURE OF THE MALE REPRODUCTIVE SYSTEM

The basic component of the male reproductive system is the testis (Fig. 1A), suspended in the body cavity by tracheae and fat body. The more primitive apterygote insects have a single testis, and in some lepidopterans the two maturing testes are secondarily fused into one structure during the later stages of larval development, although the ducts leading from them remain separate. However, most male insects have a pair of testes, inside of which are sperm tubes or follicles that are enclosed within a membranous sac. The follicles connect to the vas deferens by the vas efferens (Fig. 1B), and the two tubular vasa deferentia connect to a median ejaculatory duct that leads to the outside. A portion of the vas deferens is often enlarged to form the seminal vesicle, which stores the sperm before they are discharged into the female. The seminal vesicles have glandular linings that secrete nutrients for nourishment

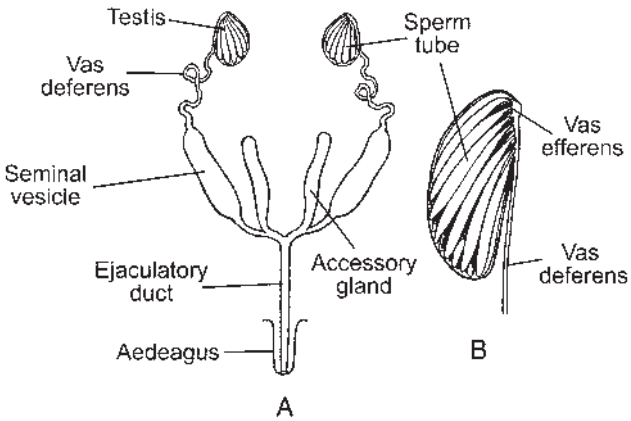

FIGURE 1 (A) The generalized structure of the male insect reproductive system. (B) A testis containing sperm tubes. (Modified from Snodgrass, 1935.)

and maintenance of the sperm. The ejaculatory duct is derived from an invagination of the epidermal cells during development and, as a result, has a cuticular lining. The terminal portion of the ejaculatory duct may be sclerotized to form the intromittent organ, the aedeagus. The remainder of the male reproductive system is derived from embryonic mesoderm, except for the germ cells, or spermatogonia, which descend from the primordial pole cells very early during embryogenesis.

PRODUCTION OF SPERMATOZOA

Within the testes are numerous tubular follicles that contain the developing male gametes. These follicles are in turn enclosed by a peritoneal sheath. The follicle number varies considerably; there is only 1 follicle in some coleopterans and over 100 in some orthopterans. At the anterior region of the follicle within the germarium are the spermatogonia, the undifferentiated germ cells, which are interspersed with somatic cells (Fig. 2). As the spermatogonia divide and move down the follicle, they become enclosed by the somatic tissue to form cysts. The development of the spermatozoa occurs within these cysts, which elongate as development proceeds. As more cysts are synthesized, they move the more mature ones posteriorly so that a range of cysts that contains gametes in progressive developmental stages is produced.

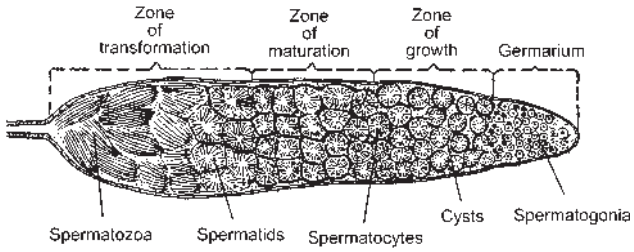

FIGURE 2 Longitudinal section of one sperm tube showing the development of spermatozoa from spermatogonia within cysts. (Modified from Snodgrass, 1935.)

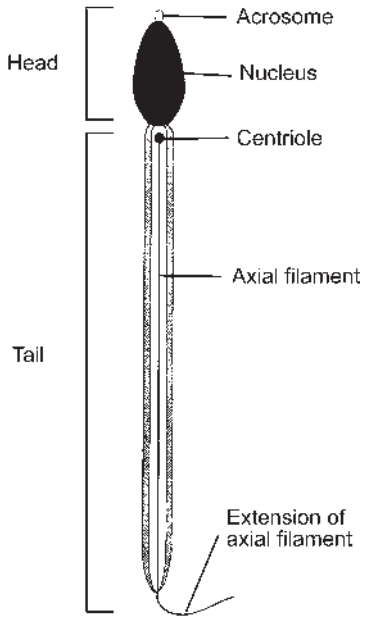

FIGURE 3 A generalized insect spermatozoon. (Modified, with permission, from Breland *et al.,* 1968.)

In the zone of growth (Fig. 2), a spermatogonium within its cyst divides mitotically six to eight times, ultimately resulting in 64 to 256 diploid spermatocytes, each of which remains connected by cytoplasmic bridges, or ring canals. Farther down in the zone of maturation, the spermatocytes undergo meiotic division to become haploid spermatids. At the posterior end of the cyst, the spermatids complete a developmental transformation known as spermiogenesis to become the spermatozoa, with heads that contain the nuclear material and elongated flagella for motility. Finally, the spermatozoa break through the cyst and migrate to the seminal vesicles where they remain as a reservoir of sperm until the insect copulates with a female.

Insect spermatozoa have elongated heads and especially long tails (Fig. 3). Within the head are the nucleus and the acrosome, which contains the enzyme acrosin, which is associated with the penetration of the plasma membrane of the egg for fertilization. Insect sperm are unique in that they are the only cells of these animals that bear a flagellum. Within the flagellum is an axial filament that develops from the centriole and consists of an array of microtubules in a 9+9+2 arrangement, with nine outer accessory tubules, nine inner doublets, and a pair of central tubules. One member of each inner doublet also bears the protein dynein, which breaks down ATP and converts the energy released into flagellar movement. At the sides of the axoneme are elongated crystalline structures, the Nebenkern, that are derived from mitochondria.

Another feature of the reproductive tract of male insects is a pair of male accessory glands. These glands can open into either the vas deferens or the ejaculatory duct. Those arising from the vas deferens during development are mesodermal in origin, whereas in some insects in which they originate from the ejaculatory duct they have an ectodermal origin. The male accessory glands serve a variety of functions, including producing the seminal fluid that serves as a transport, activation, and nourishing medium for sperm; acting as a vaginal mating plug that temporarily blocks sperm from another male from entering; and forming spermatophores, which are proteinaceous secretions of the glands that enclose the sperm and protect them while they are being transferred to the female.

The male accessory glands also produce substances that affect the behavior and physiology of the female. During insemination, peptides produced by the accessory glands that are contained in the semen are transported to various sites within the female. They are frequently involved in triggering her reproductive processes, including oogenesis and oviposition, or terminating subsequent female mating behavior so as to prevent her from acquiring the sperm of another male. Male accessory gland substances are frequently responsible for monogamy in females.

EXTERNAL GENITALIA AND MALE COURTSHIP

The external genitalia of males show an extreme morphological divergence. Indeed, their shapes often serve as the only reliable basis for species identification in taxonomic keys. In species of which the females mate often, there may be selective pressure for males to develop genitalia that compete for more effective ways to inseminate the females. There is a selective pressure for the evolution of male genitalia that offer males the greatest success; males with genitalia that more effectively transfer sperm are more likely to have their genetic contributions remain in the gene pool. Because copulation may last for a long period, males must have a way of avoiding detachment by other males that are also attempting to mate. Male insects have thus evolved a diverse array of grasping structures that allow the male to remain attached to the female; these devices include claspers, spines on the aedeagus, and even the use of mandibles to grasp a portion of the female's body. Structures have also evolved for the removal of existing sperm from a prior mating. For example, the aedeagus of the damselfly male scoops out the sperm present in the female before he adds his own. He thus eliminates any competition among sperm that may already be present in the female. Some male bedbugs inseminate the female by piercing her cuticle and depositing sperm directly into the body cavity.

Male reproductive behavior typically involves first searching for a female and, once she is found, precopulatory behaviors that increase her receptivity for copulation. This may include the presentation of a nuptial gift that occupies the female and increases her nutritional state. Once copulation is initiated, the actual transfer of sperm may require from a few minutes to a few hours. To protect their genetic investment, some males remain with the female for a variable interval after insemination has taken place. Male termites guard their mates for life. Some male scarab beetles gather dung for the female to lay her eggs on or help her to build a burrow and even remain inside the burrow with her for months.

Male paternal investment in offspring is most highly expressed in the giant water bugs of the family Belastomatidae.

Females lay their eggs on the backs of partially submerged males, interrupted only to mate repeatedly with the male. The males carry the eggs on their backs until the young hatch, actively maintaining their own position in the water in order to aerate the eggs.

See Also the Following Articles

Accessory Glands • Embryogenesis • Mating Behaviors • Spermatophore

Further Reading

Ben-Ari, E. T. (1999). Paternity battles. How males compete for fatherhood via sperm competition. *BioScience* **49,** 951–956.

Breland, O. P., Eddleman, C. D., and Biesele, J. J. (1968). Studies of insect spermatozoa. I. *Entomol. News* **79,** 197–216.

Davey, K. G. (1985). The male reproductive tract. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), Vol. 1, pp. 1–14. Pergamon, Oxford.

Jamieson, B. G. M. (1987). "The Ultrastructure and Phylogeny of Insect Spermatozoa." Cambridge University Press, Cambridge, UK.

Phillips, D. M. (1970). Insect sperm: Their structure and morphogenesis. *J. Cell Biol.* **44,** 243–277.

Tallamy, D. W. (2001). Evolution of exclusive paternal care in arthropods. *Annu. Rev. Entomol.* **46,** 139–165.

Thornhill, R., and Alcock, J. (1983). "The Evolution of Insect Mating Systems." Harvard University Press, Cambridge, MA.

Reproduction, Male: Hormonal Control of

Marc J. Klowden
University of Idaho

The endocrine control of reproduction in female insects has been relatively well studied, perhaps because egg production is often cyclical and the experimental manipulation of hormones at key developmental periods has a direct effect on the number of offspring the females produce. This has not been the case for the males of most insect species, in which the production of sperm may begin early during the larval, nymphal, and pupal stages and often continues throughout adult life. Given the vastly different hormonal conditions that exist during the immature and adult periods, a unifying scheme for the control of male spermatogenesis has not been forthcoming. There are only a few examples of hormones controlling the reproductive processes of male insects.

PRODUCTION OF SPERM

The male gametes are produced in the paired testes from the spermatogonia. As the spermatogonia divide and move down the sperm tube, they become surrounded by tissue to form cysts. Their subsequent development to mature spermatozoa occurs within these cysts. First, the spermatogonia divide mitotically to produce diploid spermatocytes, and these spermatocytes undergo a meiotic division to become haploid spermatids. The spermatids then undergo a developmental transformation to become differentiated spermatozoa that ultimately break through the cyst and migrate to the seminal vesicles where they remain until the male copulates with a female.

The development of the testes during metamorphosis responds to the insect hormone 20-hydroxyecdysone (20HE). The synthesis of proteins by the testes of newly emerged adult *Tenebrio molitor* beetles is dependent on their exposure to this hormone.

Both 20HE and juvenile hormone (JH) have been implicated in the regulation of spermatozoa within the testes. High levels of 20HE increase the rate at which the spermatogonia undergo mitotic divisions to form spermatocytes, but this increase is abolished by high levels of JH. When spermatogenesis begins during the larval period, as in some Lepidoptera, the spermatocytes initiate their meiotic divisions but are arrested at prophase until the end of the larval period is reached. In many holometabolous insects, a peak of 20HE occurs before the end of the larval period that induces the "wandering" behavior that allows the larva to find a secluded spot in which to pupate. A postwandering peak of 20HE unblocks the meiotic division of the spermatocytes and allows the cells to proceed to metaphase. In some insects, JH has also been found to accelerate spermatogenesis.

When spermatogenesis begins during the larval stage, as it does in most Lepidoptera, it is interrupted if the insect diapauses as a larva or pupa. However, it resumes once diapause has been completed. The interruption is not caused by a pause in developmental activity but rather from the lysis of developing gametes before they become mature. The renewal of spermatogenesis when diapause has been terminated occurs when the titers of 20HE increase.

The developmental sequence in which spermatozoa develop from spermatids involves an elongation of both the nucleus and the flagellum. In Lepidoptera, the nuclear elongation is triggered by the declining concentrations of 20HE, but the elongation of the flagellum appears to be independent of any hormone.

Lepidopteran males produce two different types of sperm. Eupyrene sperm have a nucleus and fertilize the eggs, but apyrene sperm are anucleate. Both types of sperm are transferred to the female spermatheca after copulation. The function of the apyrene sperm is not well understood but they may provide nutrients required by the eupyrene sperm or aid in the migration of the eupyrene sperm within the female genital tract. The developmental pathway leading to the differentiation of apyrene sperm is regulated by a humoral factor that is present at the time of pupation.

ACCESSORY GLANDS

The accessory glands of the male reproductive system produce semen, accessory structures such as spermatophores, and various

peptides that regulate female behavior and physiology. The interaction between JH and 20-hydroxyecdysone during postembryonic development regulates the development and differentiation of the glands. Juvenile hormone alone may also control the synthesis of those specific proteins that are transferred to the female. The accumulation of some secretory peptides in the glands that are enhanced by JH may be either enhanced or inhibited by the simultaneous presence of 20-hydroxyecdysone. In the German cockroach, *Blattella germanica,* the activity of the corpora allata, which is the source of JH, declines during the formation and transfer of the spermatophore and may thus initiate a new cycle of male accessory gland maturation.

See Also the Following Articles

Accessory Glands • Ecdysteroids • Juvenile Hormone • Spermatophore

Further Reading

Chapman, R. F. (1998). "The Insects: Structure and Function," 4th ed. Cambridge University Press, Cambridge, UK.

Dumser, J. B. (1980). The regulation of spermatogenesis in insects. *Annu. Rev. Entomol.* **25,** 341–369.

Friedländer, M. (1997). Control of eupyrene–apyrene sperm dimorphism in Lepidoptera. *J. Insect Physiol.* **43,** 1085–1092.

Gillott, C., and Gaines, S. (1992). Endocrine regulation of male accessory gland development and activity. *Can. Entomol.* **124,** 871–886.

Klowden, M. J. (2002). "Physiological Systems in Insects." Academic Press, San Diego.

Research Tools, Insects as

Kipling W. Will
University of California, Berkeley

In addition to serving as classic experimental laboratory animals (e.g., *Drosophila* flies in genetics, *Periplaneta* cockroaches in neurophysiology, *Manduca* moths and *Schistocerca* grasshoppers in physiology), insects have been essential to the formulation and testing of many general theorems in ecology and evolutionary biology. Scientists seek to develop general synthetic theories in biology, just as in physics and chemistry, that provide answers to questions of how and why things are as they are. Perhaps more importantly, these generalities make predictions that allow us to test what we think we know. The use of data to constantly reevaluate our theories divides empirical science from personal belief systems and popular metaphysics. The latter two are concerned with understanding the fundamental nature of all reality and often are based on abstract elements. In contrast, hypotheses in mainstream science are typically explanations of the processes that exist in nature and are tested against empirical observations.

To develop these testable hypotheses in science a constellation of data from model systems is needed, especially in biology. Insects are used as many of these model systems. In part this is because they are so abundant and species rich that, in terms of diversity, they make up the bulk of terrestrial species and, in many habitats, the greatest number of individuals. Because of this dominance, comprehensive biological hypotheses must account for insects if the hypotheses are to be generally accepted. Insects provide the numerous observations that are essential for developing and supporting broadly applicable hypotheses explaining the diversity and distribution of life on earth.

Insects are a magnificent source of observations. The sheer number of units for study at all levels—species, populations, and individuals—provide the repeated patterns of variation that provoke questions and provide data for hypothesis testing. Also, insects usually have a relatively short life cycle, often more than one generation per year. This allows scientists to make multiple observations of all life stages of a species in a relatively short time period. Insect species can be widespread, but typically they are localized, making it easy to accumulate distributional data for at least the more conspicuous taxa. Certainly there is an important human factor as to why insects are so important in the study of biology. Insects delight us with their forms and behaviors and seem to embody all that fascinates humans about the natural world—beauty, diversity, mystery, and perhaps most of all discovery. Once a biologist, or any naturalist, is exposed to the wonders of insects they are usually hooked for life.

Biology has benefitted greatly from both reductionist and integrative research. The reductionist program attempts to minimize the number variables in the study system and identify causal mechanisms. Stunning success has been achieved using *Drosophila* as a laboratory animal to investigate developmental and genetic systems and to discover basic mechanisms from which inferences about general principles in biology are made. An integrative or synthetic approach is also essential in biology. Insects have been crucial model organisms, providing some of the most important advances in synthetic theory.

THEORETICAL WORK IN THE 19TH CENTURY

Evolution, or the theory of natural selection, is the most influential of all biological theories, and the two men who codified the basic mechanisms of descent with modification were dedicated observers and collectors of insects. Charles Darwin and Alfred Wallace shared what Wallace referred to as a "child-like" passion for beetles; Wallace even suggested this may have been a common thread that helped to lead both of them to arrive independently at similar conclusions about the evolutionary origin of species. Both present colorful stories of collecting insects. Wallace wrote wonderful passages on "one good day's work" collecting in Borneo, recalling species by species those collected and those that escaped one day, to be pursued the next. Darwin recalls with great passion his beetle collecting and the unfading thrill of discovering rare or new species. Wallace earned much of his livelihood collecting and providing specimens to museums and private collectors. He

sold thousands of specimens at about 2 cents each to fund his tropical expeditions. For both Darwin and Wallace, attention to details necessary for separating species, subspecies, and varieties of insects was fundamental to developing their ideas. The diversity of forms and sheer reproductive output of insects provided examples that cultivated in their minds the theory of natural selection.

Wallace traveled with another entomologist and great naturalist, Henry Bates. In 1842 Wallace and Bates went to the Amazon to explore and to collect insects. These explorations and Bates' collections (Wallace's were unfortunately lost in a ship fire) became incredibly valuable in terms of insights into natural history and evolution. For 11 years, Bates collected insects, primarily butterflies and beetles, that were and remain a source of awe and study material for students of insects and users of European museum collections. Bates readily accepted Darwin's and Wallace's ideas of natural selection as the mechanism of evolution and went on to develop his theory of mimicry. Known as Batesian mimicry, this concept stemmed from his experience with tropical butterflies. This theory is widely applied throughout biology as an explanation for the similar and convergent appearance of some organisms.

THEORETICAL WORK IN THE 20TH CENTURY

Willi Hennig is best known for developing a coherent theory for phylogenetic systematics, a field of research that investigates and presents relationships among taxa. Hennig's theoretical works form the core of modern cladistic methods (use of shared derived characteristics to elucidate sister group relationships of taxa). Hennig was also the foremost authority on flies (Diptera) and produced many publications, including his series of publications on maggots (dipterous larvae), which became the standard work on the subject. Throughout his classic work *Phylogenetic Systematics* he relied on insect examples. Today our ideas about how to develop hypotheses of relationships for animals in an evolutionary scheme and how to classify them are largely based on theories and methods developed with insects as models.

Biogeography is a major field of biology that strives to understand the spatial relationships of organisms and looks at both historical and contemporary questions regarding biodiversity. Darwin, Wallace, and Hennig were all prominent contributors to this field, each drawing on ample observations from the insect world. Wallace was particularly influential in developing ideas that are still important in biogeographical studies. Biogeographical regions of the earth presented by Wallace, which were modified from Philip Sclater's previously published scheme, are still a standard part of describing the geographic distribution of animals. Most prominent is Wallace's observation of a distinct change in fauna between Bali and Lombok in the East Indies, known as Wallace's line. Wallace drew heavily on his knowledge of insect life histories, dispersal abilities, and distribution of some conspicuous insects (beetles and butterflies) to develop his biogeographical ideas.

A significant change from thinking about biogeography only in terms of evolutionary and historical scenarios to looking at ecological dynamics began in the 1960s. Robert MacArthur and E. O. Wilson published the equilibrium theory of island biogeography, a model based on land area and distance from source populations that explained how newly available islands could become populated with plants and animals, ultimately coming to a point of equilibrium in terms of species number. This became one of the most influential works in the field. Noted biologist and entomologist Ed Wilson is an ant systematist and he used these insects to support the development of this theory.

Wilson's contributions to biology are many, but some of the best known and most controversial are the ideas presented in sociobiology. In general this field seeks explanations of social behavior in animals based on common biological and evolutionary concepts. Largely this synthesis is based on his knowledge of ants, animals that are truly social. Many other insect groups, including beetles, butterflies, termites, dragonflies, and bees, just to name a few, are exemplar taxa used to illustrate concepts of sociobiology. Sociobiology is broad and interdisciplinary; insects are a vital part of the hypotheses that span many fields of inquiry.

LOOKING FORWARD IN THE 21ST CENTURY

Whether through reductionist approaches or the development of synthetic theories of biology, it is clear that the study of and passion for insects is an incredibly important part of our understanding of the world we live in and realizing our place in the natural system. Just as insects have proven to be essential in developing theories using the integrative approach, they will continue to be prominent in studies at all levels of organization from the molecular to the ecosystem.

See Also the Following Articles

Biodiversity • Drosophila melanogaster • Industrial Melanism • Mimicry

Further Reading

Bates, H. W. (1862). Contributions to an insect fauna of the Amazon valley: Lepidoptera: Heliconidae. *Trans. Linn. Soc. London* **23,** 495–566.

Darwin, C. (1859). "The Origin of Species by Means of Natural Selection: Or, the Preservation of Favoured Races in the Struggle for Life." Murray, London.

Darwin, C., and Wallace, A. R. (1858). On the tendency of varieties to depart indefinitely from the original type; and on the perpetuation of varieties and species by natural means of selection. *J. Proc. Linn. Soc. Zool.* **3(9),** 53–62.

Dethier, V. G. (1962). "To Know a Fly." Holden–Day, San Francisco.

Hennig, W. (1966). "Phylogenetic Systematics." University of Illinois Press, Urbana. [Translated by D. Dwight Davis and Rainer Zangerl]

MacArthur, R. H., and Wilson, E. O. (1967). "The Theory of Island Biogeography." Monographs in Population Biology, No. 1. Princeton University Press, Princeton, NJ.

Rogers, B. T., and Kaufman, T. C. (1997). Structure of the insect head in ontogeny and phylogeny: A view from *Drosophila. In* "International Review of Cytology" (K. W. Jeon, ed.). Academic Press, San Diego.

Sclater, P. L. (1858). On the general geographical distribution of the members of the class Aves. *J. Linn. Soc. Zool.* **2,** 130–145.
Statzner, B., Hildrew, A. G., and Resh, V. H. (2001). Species traits and environmental constraints: Entomological research and the history of Ecological Theory. *Annu. Rev. Entomol.* **46,** 291–316.
Wallace, A. R. (1855). Concerning collecting dated 8 April 1855, Si Munjon Coal Works, Borneo. *Zoologist* **13,** 4803–4807. [Letter]
Wallace, A. R. (1876). "The Geographical Distribution of Animals, with a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface." Macmillan & Co., London.
Wilson, E. O. (1975). "Sociobiology: The New Synthesis." Harvard University Press, Cambridge, MA.

Respiratory System

Jon F. Harrison
Arizona State University

The primary goals of the insect respiratory system are to deliver oxygen from the air to the tissues and to transport carbon dioxide from the tissues to air. Gases are transported through the tracheal system by both diffusion and convection, with the relative importance of these two mechanisms varying across and within species. Aquatic and endoparasitic insects exchange gases by a variety of mechanisms; most have tracheae and tracheoles within thin-walled appendages that function as gills. Within individual insects, the structure of the tracheal system can be altered dramatically during ontogeny and in response to rearing conditions, such as low oxygen. With a given tracheal structure, the ability of the tracheal system to transport gases can be modulated dramatically by varying spiracular opening, ventilation, and the fluid level in the tracheoles. Control of flexibility in tracheal gas exchange capacity depends strongly on neuroendocrine control of muscles that drive convection or control spiracular opening.

DEVELOPMENT

In most insects, the tracheal system first appears in the embryo. In general, the size of the tracheal system increases with age in order to support the increased gas-exchange needs of the larger insect. However, major changes in tracheal structure, including changes in spiracle number and tracheal system organization, can occur at each molt and during the pupal period for endopterygote insects. During the molts, the cuticular lining of the trachea is drawn out of the spiracle with the old integument.

Changes in tracheal system structure are not limited to molting periods, because the tracheoles can change structure within an instar. In the event of injury or oxygen deprivation, local tracheoles grow and increase in branching. If no undamaged tracheoles are nearby, damaged tissues produce cytoplasmic threads that extend toward and attach to healthy tracheoles. These threads then contract, dragging the tracheole and its respective trachea to the region of oxygen-deficient tissue.

Both tracheoles and trachea are fluid-filled in newly hatched insects, and fluid fills the space between the old and the new trachea at each molt. Usually, this fluid is replaced with gas shortly after hatching or molting. In some cases, the spiracles must be open to the air for gas-filling of the tracheae to occur, suggesting that gas-filling occurs as fluid is actively absorbed by the tracheal and tracheole epithelia, with air entering through the spiracle to replace the absorbed fluid. However, in aquatic insects that lack spiracles, the tracheae also become gas-filled, indicating that this gas can be generated by the tissues or hemolymph.

MECHANISMS OF GAS EXCHANGE

Insect gas exchange occurs in a series of steps. Oxygen molecules first enter the insect via the spiracle, then proceed down the branching tracheae to the tracheoles. The terminal tips of the tracheoles are sometimes fluid-filled, so at this point gas transport may occur in a liquid medium rather than air. Oxygen then must move across the tracheolar walls, through the hemolymph, across the plasma membranes of the cells, and finally through the cytoplasm to the mitochondria. Carbon dioxide generally follows a reverse path.

Diffusion

Diffusion is the passive movement of molecules down their concentration gradient, driven by random molecular motions. Because oxygen is transported to the tissues as a gas and the diffusion rate of oxygen is much more rapid in air than in water, the insect tracheal system is capable of high rates of gas exchange by diffusion. Consumption of oxygen by the tissues lowers internal oxygen levels, creating a concentration gradient from air to tissues that drives oxygen through the tracheae. The converse occurs for carbon dioxide. The final steps of oxygen delivery, from the tracheoles to the mitochondria, may occur by diffusion in all insects, because diffusion operates rapidly over micrometer distances. In the initial steps of oxygen delivery (across the spiracles, through the tracheae), the importance of diffusion is more variable. Simple diffusive gas exchange through the tracheae and spiracles likely occurs in some pupae, as washout rates of inert gases are similar to those predicted from their diffusion coefficients. Additionally, in a variety of insects, no ventilatory movements have been discerned, which may mean that these insects exchange gases by diffusion through their tracheae and spiracles.

Convection

Convection is the bulk movement of a fluid (gas or liquid) driven by pressure. Differential air pressures can drive gas movement through the tracheae and spiracles at much higher rates and over longer distances than diffusion. In many insects, well-coordinated actions of muscles and spiracles produce regulated convective air flow through the tracheae and spiracles.

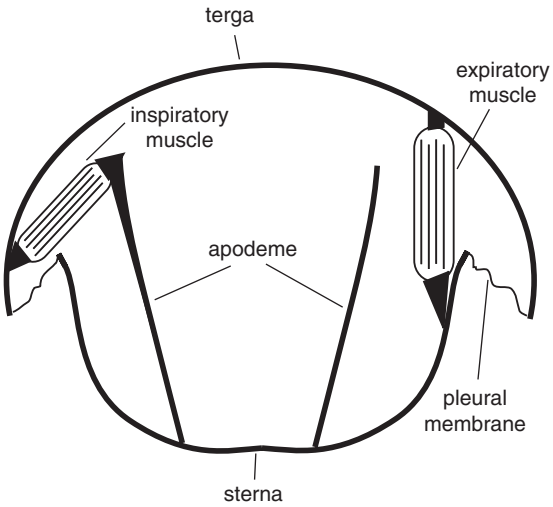

FIGURE 1 Respiratory muscles that drive the dorsoventral movements of the abdomen during abdominal pumping in a grasshopper. (Adapted from Hustert, 1975, and Chapman, 1998, with permission of Cambridge University Press.)

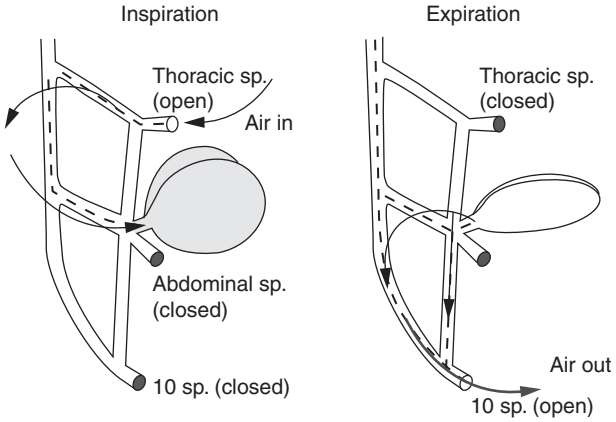

FIGURE 2 Unidirectional air flow during abdominal pumping in a grasshopper. During inspiration, air flows in through open thoracic spiracles (sp), along the longitudinal trachea, and into the air sacs. At low metabolic rates, air flows out only through the 10th abdominal spiracles; in more active animals, air flows out all abdominal spiracles.

Convection can be driven by a variety of mechanisms in insects. Most commonly, convection is driven by contractions of respiratory muscles attached to the body wall, which produce increases or decreases in body volume, causing compressible portions of the tracheal system to inflate or deflate.

One common mechanism by which insects accomplish convective air flow through the trachea is abdominal pumping. Expiratory muscles connect the ventral and dorsal cuticular plates and also span adjacent abdominal segments. When they contract, they pull the dorsal terga and ventral sterna together (Fig. 1) and the tip of the abdomen inward as the cuticular plates slide over the flexible pleural and intersegmental membranes. When the body volume decreases, the cuticle pushes on hemolymph and tissues, which in turn compress the collapsible air sacs. Inspiration may be passive and result from cuticular elasticity. Alternatively, contractions of inspiratory muscles attached to tall sternal apodomes and the lower edge of the terga can lift the terga relative to the sterna and expand abdominal volume and the air sacs (Fig. 1). Similar movements of the terga and sterna produce abdominal pumping in Orthoptera and in adult Odonata, Hymenoptera, and Diptera, whereas in adult Coleoptera and Heteroptera only the terga moves. In adult Lepidoptera, Orthoptera (Tettigoniidae), Neuroptera, and Trichoptera, movements of the sterna, terga, and lateral pleural membranes all contribute to abdominal pumping.

In many of the insects that have been examined, abdominal pumping is coordinated with spiracular opening in a manner that produces unidirectional air flow through the longitudinal tracheal trunks. During inspiration, abdominal spiracles are closed and air flows in through open thoracic spiracles (Fig. 2). During expiration, air flows out open abdominal spiracles, while the thoracic spiracles are closed. There is evidence for unidirectional air flow in adult cockroaches (Blattodea), grasshoppers (Orthoptera) and mantids (Mantodea), honey bees (Hymenoptera), and dragonflies (Odonata). In some pupae, abdominal pumping is coordinated with the opening of one or a few "master spiracles" that exchange all gases.

Muscles that have the primary purpose of driving hemolymph circulation, such as the heart and ventral diaphragm, also play a role in creating convective ventilation. Pumping of the heart and accessory muscles at the base of the wings, antennae, and legs pushes hemolymph and air into these appendages. In a variety of adult Lepidoptera, Diptera, and Hymenoptera, the heart occasionally reverses pumping direction, shifting hemolymph from the thorax to the abdomen. As the hemolymph accumulates in one body compartment, it compresses air sacs in that compartment, causing convective air flow through tracheae and spiracles. Similar hemolymph transfers between body compartments coupled to ventilation may occur because of intermittent heart activity in bees (Hymenoptera).

DISCONTINUOUS GAS EXCHANGE

For many insects, especially those that are highly active, the spiracles close for only brief periods, if at all, and oxygen and carbon dioxide are exchanged relatively continuously. However, many insects exhibit discontinuous gas exchange. During discontinuous gas exchange, periods of spiracular closure (in which no or reduced gas exchange occurs) alternate with periods of spiracular opening (and greatly increased gas exchange; Fig. 3).

Discontinuous gas exchange has been most extensively studied in ants and diapausing lepidopteran pupae. In lepidopteran pupae, spiracular openings can be separated by hours. After the burst of gas exchange, the spiracles hermetically seal (closed phase). While the spiracles are closed, oxygen is consumed from within the tracheal system (Fig. 3). The oxygen removed from the tracheal air space is not completely replaced

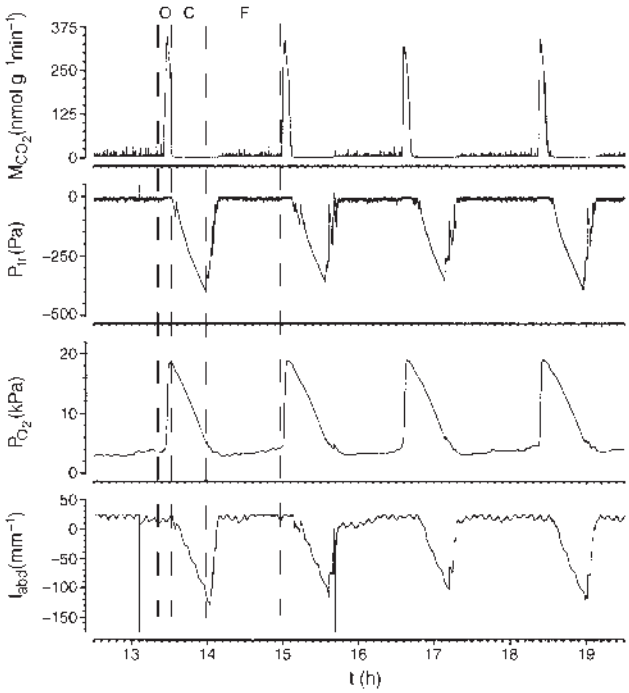

FIGURE 3 Four cycles of discontinuous gas exchange in pupal *Attacus* (Lepidoptera). From top to bottom: carbon dioxide emission (MCO_2), tracheal pressure (P_{tr}), tracheal oxygen concentration (Po_2), and abdominal length (I_{abd}). For the first cycle, the open (O), closed (C), and flutter (F) phases are indicated. (Reproduced with permission from Hetz, 1994).

by carbon dioxide primarily because a large fraction of the carbon dioxide produced dissolves in the tissues and hemolymph (carbon dioxide, in contrast to oxygen, is highly soluble in biological fluids). Therefore, pressures fall below atmospheric within the tracheal system and abdominal length is reduced (Fig. 3). When internal oxygen tensions reach a low threshold, the spiracles begin to open slightly at high frequency (flutter phase), and tracheal pressures rise to near-atmospheric levels. During the flutter phase, high-frequency, but minutely subatmospheric, air pressures allow the animal to convectively take in oxygen with minimal emission of carbon dioxide. The spiracles remain sufficiently closed so that internal oxygen levels remain low (Fig. 3). Carbon dioxide accumulates throughout the closed and fluttering phases, eventually triggering the spiracular open phase, when tracheal oxygen levels are restored to near-atmospheric concentrations.

Discontinuous gas exchange patterns and mechanisms are quite variable among and within species. Many insects exhibit discontinuous gas exchange only when at rest or during diapause. However, some highly active insects, such as running ants, also exchange gases discontinuously. When metabolic rates increase (at higher temperatures, or during activity), the time between spiracular phases tends to decrease. In contrast to the pattern diagrammed in Fig. 3, significant carbon dioxide is lost during the flutter phase of many insects, suggesting a predominance of diffusive gas exchange. Some species use abdominal pumping to enhance gas exchange during the open phase. The ecological and evolutionary significance of this variation is unclear. Although many researchers have hypothesized that discontinuous gas exchange functions to reduce respiratory water loss, most recent tests of this hypothesis have failed to support it.

AQUATIC AND ENDOPARASITIC INSECTS

Insects can obtain oxygen while living in fluid environments and are common in fresh waters, in brackish estuaries, and as endoparasites. Aquatic and endoparasitic insects generally retain an internal, air-filled tracheal system. Maddrell has suggested that the lack of insects in deep-water environments occurs because the air-filled, buoyant tracheal system precludes insects from being able to dive deeply and avoid predators.

Many aquatic and endoparasitic species have anatomical features that allow them to feed under water (or within the host) while maintaining contact with air. For example, in many aquatic dipteran larvae such as mosquitoes, the posterior spiracles are surrounded by water-repellent hairs and are kept in the air while the animals feed in a head-down position. In water scorpions (Heteroptera: Nepidae), spiracles are located on the end of a long tube which is extended up to the water surface. Similarly, endoparasitic insects such as chalcid (Hymenoptera) larvae and tachinid (Diptera) larvae connect to the air using posterior spiracles inserted through the host's integument or tracheal system.

Some beetles (Coleoptera) and true bugs (Heteroptera) use their hairs or wings to carry air bubbles adjacent to their spiracles when they dive. These air bubbles serve as oxygen stores and as temporary gas exchange structures. Oxygen is removed from the air bubble by the diving insect, causing oxygen concentrations in the air bubble to fall below that in the surrounding water and oxygen to move from the water into the air bubble by diffusion. The removal of oxygen from the bubble also causes the nitrogen concentration in the air bubble to rise above that in the surrounding water, so nitrogen leaves the bubble by diffusion. Eventually, when all the nitrogen leaves, the bubble disappears and the insect must return to the surface.

Many aquatic and endoparasitic insects never access air and must obtain oxygen directly from water or the blood of the host. Some aquatic insects that rarely visit the surface obtain oxygen from the water by having specialized structures called plastrons which hold a thin film of air on the outside of their body. These insects have a thick (as many as 2 million hairs per square millimeter) layer of short water-repellent hairs that resists wetting and does not collapse under pressure. Spiracles open directly into the air space of the plastron. Plastrons are thought to behave as gas exchange structures in a manner similar to that of air bubbles. However, because the dense hairs make the plastron incompressible, oxygen removal lowers both the total pressure and the oxygen concentration in the air of the plastron. Thus nitrogen levels are likely to remain near those in water, and the air and gas exchange function of the plastron can persist indefinitely.

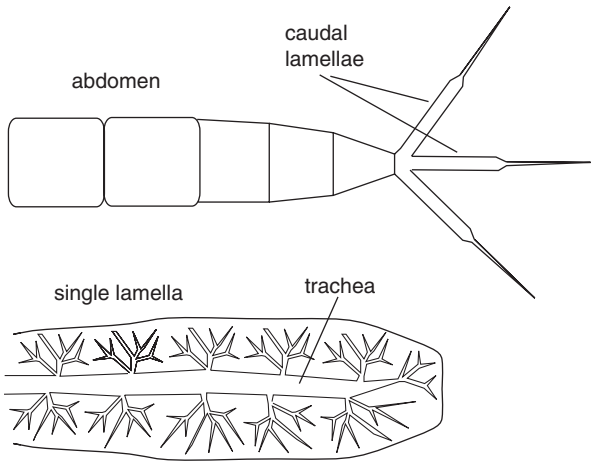

FIGURE 4 Caudal lamellae, which function as tracheal gills. (Modified with the permission of Cambridge University Press, from Chapman, 1998).

The majority of aquatic or endoparasitic insects that do not access air lack spiracles, so that oxygen must be transported by diffusion from the water, across the cuticle, and then into the tracheae. A common feature of these insects is tracheal gills, leaf-like structures of thin cuticle containing many tracheae and tracheoles, which increase the available surface area for obtaining oxygen from water (Fig. 4). Gills may be on the abdominal tip (Diptera, Odonata: Zygoptera), laterally along the abdomen (Ephemeroptera, Trichoptera), or within the rectum (Diptera, Odonata: Anisoptera).

As oxygen is absorbed across a gill, oxygen level in the boundary layer of water over the gill can become low, slowing diffusive influx of oxygen. To overcome this problem, muscles drive rhythmic waving of abdominal gills, forcing fresh, oxygenated water to flow over the gills, thereby reducing boundary layer thickness. For similar reasons, rectal gills are ventilated by alternate contractions and relaxations of the rectum, which drive water in and out of the anus. Such ventilatory movements are enhanced in water containing low levels of oxygen.

Finally, some aquatic and endoparasitic insects have evolved the use of respiratory pigments to enhance oxygen storage and delivery. Some *Chironomus* larvae (Diptera) have hemoglobin in their hemolymph. Many of these species live in mud under stagnant water with low oxygen content, and the possession of hemoglobin with a very high affinity for oxygen allows these insects to obtain, store, and transfer oxygen at very low oxygen levels. Larval *Gasterophilus* (Diptera), an internal parasite in the hypoxic stomach lumen of horses, also possesses a hemoglobin with very high affinity for oxygen. In early instars, the hemoglobin of *Gasterophilus* is in the hemolymph, and so it may function in oxygen transport through the blood, whereas in later instars hemoglobin is concentrated in cells that are surrounded with tracheoles. It is thought that oxygen availability in the stomach varies with the presence of swallowed air bubbles and that possession of hemoglobin in cells allows *Gasterophilus* to store oxygen until the next air bubble appears. The diving bug *Anisops* (Heteroptera) stores hemoglobin in hemoglobin cells surrounded with tracheae. *Anisops* obtains its oxygen from the air, carrying an air bubble under its wings when diving. The hemoglobin of *Anisops* has a lower affinity for oxygen than that of *Chironomus* or *Gasterophilus,* and it functions in a manner similar to that of hemoglobins in the blood of diving mammals, becoming oxygenated at the surface and releasing oxygen during the dive. Possession of hemoglobin allows *Anisops* to sustain a fivefold longer dive.

PLASTICITY IN TRACHEAL STRUCTURE AND FUNCTION

An insect's need for oxygen and production of carbon dioxide vary tremendously, as metabolic rate varies strongly with transitions from rest to flight, unfed to fed state, or diapause to growth. When metabolic rates increase, there is an increased need for oxygen uptake and carbon dioxide removal. Insects also require a higher capacity to exchange gases when they are exposed to low oxygen conditions such as hypoxic waters or high altitudes.

In the short term, increased gas exchange with a set tracheal system structure can be accomplished by at least four nonexclusive mechanisms. First, insects can simply tolerate lower internal oxygen and higher internal carbon dioxide concentrations. The increased gas concentration gradients will enhance gas exchange by either diffusion or convection. Second, insects can increase diffusive gas exchange by opening the spiracles to a greater degree or for longer time periods. Third, insects can increase convective ventilation through the spiracles and tracheal system (terrestrial insects) or over the gills (aquatic insects) by mechanisms such as abdominal pumping or gill waving. Fourth, insects can reduce fluid levels in the tracheoles, enhancing diffusive oxygen delivery within the tissues because of the faster diffusion of oxygen through air than through water. In the longer term, alterations in tracheal morphology (e.g., tracheal diameter, number of tracheoles) can strongly affect gas exchange capacity.

Locomotion

Locomotion requires muscular activity and increased ATP turnover, which increases the organism's need for oxygen. Running insects increase their gas exchange rates by 2- to 10-fold relative to resting conditions. Flight is usually associated with larger increases in gas exchange, especially in insects that maintain thoracic temperatures near 40°C, such as bees and dragonflies. These insect athletes have oxygen consumption rates among the highest known in the animal kingdom, 2 to 100 times resting values.

Relatively little is known of the mechanisms by which insects increase gas exchange during terrestrial locomotion. In a cockroach (Blattodea) and a grasshopper (Orthoptera) internal carbon dioxide levels rise while oxygen levels fall

during running and hopping, respectively. In the two-striped grasshopper *Melanoplus bivittatus,* abdominal pumping rates are low during hopping, but do increase relative to resting rates afterward, increasing convection and restoring tracheal gases to normal, resting levels. Even though there is no evidence for increased abdominal pumping during hopping, increased convective ventilation may occur due to pressure fluctuations associated with cuticular deformations associated with jumping.

During flight, convective gas exchange can be increased by thoracic autoventilation and abdominal pumping. Thoracic autoventilation occurs when flight muscle contractions cause compression of the air sacs within the thorax, producing strong convective ventilation. Such increases in thoracic autoventilation have been shown for grasshoppers (Orthoptera); Cerambycidae, Elateridae, and Anthribidae (Coleoptera); moths (Lepidoptera); and dragonflies (Odonata). Another method for increasing gas exchange has been shown for the gigantic cerambycid beetle *(Petrognatha gigas),* in which wind pressure generated by forward flight drives convective air flow through the major tracheal trunks. Abdominal pumping is considered the major mechanism by which convective ventilation is increased during flight in large Hymenoptera, Diptera, and Coleaptera (Scarabaeidae and Buprestidae), and abdominal pumping supplements autoventilation during flight in dragonflies (Odonata) and grasshoppers (Orthoptera). At least in one insect studied (a moth), oxygen concentrations in active flight muscles are maintained at levels similar to those in resting animals, suggesting that increases in the capacity of the tracheal system to conduct gases match increased tissue needs for oxygen during flight.

During Hypoxia

Insects may encounter low environmental oxygen availability in a number of locations, including hypoxic waters, burrows, feeding sites within large dense structures such as granaries, or high altitudes. Terrestrial insects are generally quite good at coping with hypoxia and, usually, can maintain resting metabolic rates down to atmospheric oxygen levels of 1 to 5 kpa (normal oxygen concentration is 21 kpa). The large safety margin for oxygen delivery in resting terrestrial insects probably reflects the fact that tracheal systems must be designed to allow the much higher rates of gas exchange during activity. In support of this hypothesis, metabolic rates during flight are generally more sensitive to hypoxia, with flight metabolism being inhibited at 8 kpa in hovering honey bees and at 10 kpa in tethered flying flies and being stimulated by hyperoxia in a dragonfly. Growth rates are inhibited by relatively mild hypoxia (10 kpa), at least in larval mealworms (Coleoptera) and fruit flies (Diptera), suggesting that such moderate hypoxia limits some physiological processes even in nonlocomoting insects.

Insects respond to hypoxia by increasing tracheal gas exchange capacity. In a variety of insects, hypoxia induces spiracular opening. In ants and lepidopteran pupae exhibiting discontinuous ventilation, exposure to hypoxia increases the frequency and duration of spiracular opening. In adult grasshoppers, exposure to hypoxia causes a strong increase in convective ventilation, mostly accomplished by an increase in the frequency of abdominal pumping. Aquatic insects with gills generally increase convective water flow past the gills in response to hypoxia, by increasing the frequency of their gill beating (Ephemeroptera), body undulations (Plecoptera), or rectal pumping (Odonata). Aquatic insects exposed to hypoxia often move to faster flowing water, to the water surface, or even into the air.

Insects can exhibit changes in tracheal system structure in response to longer term exposure to hypoxia. For example, the transverse tracheae of mealworms (Coleoptera) increase in size when reared under conditions of low oxygen availability. Tracheole growth and branching are stimulated by hypoxia in the epidermis of *Rhodnius* (Heteroptera). Similarly, in some immature Ephemeroptera, gill area is inversely proportional to environmental oxygen levels.

CONTROL OF RESPIRATORY FUNCTION

The spiracular muscles are primarily controlled by nerves from the central nervous system. The motor neurons to the spiracular muscles arise from ganglia in the same segment or that immediately anterior. In dragonflies, *Periplaneta* cockroaches, and *Schistocerca* grasshoppers, the closer muscles are innervated by two motor nerves that branch from the median nerve, sending an axon to each side of the animal, so that both spiracles on a segment receive similar neural input. In the prothoracic spiracle of *Schistocerca,* two motor nerves innervate the opener muscle from the prothoracic ganglia, and one motor nerve arrives from the mesothoracic ganglia. Increasing action potential frequencies in these nerves stimulate increasing muscle activity, resulting in gradations in the magnitude and duration of spiracular closing.

Current information suggests that for spiracles with two muscles, effects of carbon dioxide or oxygen are mediated centrally. When high levels of carbon dioxide or low levels of oxygen are applied to the ventral ganglia or head, the frequency of action potentials to the closer muscles decreases, and the frequency of action potentials to the opener muscles increases, increasing spiracular opening. Temperature increases or flight also affect the frequency of action potentials to these spiracular muscles.

In contrast, data for spiracles with only a closer muscle suggest that carbon dioxide acts peripherally, while oxygen acts centrally. Central application of hypoxic gases stimulates a fall in the action potential frequencies to the closer muscle, whereas carbon dioxide has little effect. However, when carbon dioxide is applied directly on the spiracular muscle, there is a fall in the muscle membrane depolarization resulting from nerve stimulation, a decrease in closer muscle tension, and increased spiracular opening. Spiracle opening is stimulated by a rise in carbon dioxide but not a decrease in extracellular pH, suggesting that the carbon dioxide effect is not mediated by pH changes.

The rhythmic abdominal pumping movements that drive convective ventilation in many insects are initiated by central pattern generators in the metathoracic ganglia or the first abdominal ganglia. These rhythmic ventilatory bursts can be demonstrated in nerve cords isolated *in vitro.* The motor output from these central pattern generators passes down the ventral nerve cord via interneurons and stimulates rhythmical sequences of respiratory muscle activity.

In grasshoppers (Orthoptera) and probably many other insects, the rate of abdominal pumping can be altered by sensory output from stretch receptors in the abdomen, by chemosensors located in the thorax and head, and by feed-forward control. Increasing carbon dioxide or decreasing oxygen concentrations in the tracheae stimulate the frequency of abdominal pumping. Depression of carbon dioxide below normal levels or elevation of oxygen above normal levels inhibits abdominal pumping, indicating homeostatic regulation of internal gas levels. Feed-forward control of abdominal pumping occurs when neural centers that control flight muscles also stimulate the central pattern generators, increasing the rate of abdominal pumping.

In a series of classic experiments, V. B. Wigglesworth demonstrated that the terminal ends of the tracheoles can contain fluid that disappears in response to hypoxia or activity. Changes in the fluid levels in the tracheoles have the potential to strongly affect the ability of the tracheoles to conduct gases, which would provide a highly significant control mechanism for the tracheal system. Further experiments by Wigglesworth suggested that the withdrawal of fluid from the tracheoles could be due to elevations in hemolymph osmotic pressure.

In summary, the tracheal respiratory system of insects is a dynamic system, capable of a tremendous range of function and fine control. This light-weight, adaptable, high-capacity respiratory system is certainly one of the major traits that underlie the ecological and evolutionary success of insects.

See Also the Following Articles

Aquatic Habitats • Flight • Hemolymph • Muscle System • Tracheal System

Further Reading

Chapman, R. F. (1998). "The Insects: Structure and Function." Cambridge University Press, Cambridge, UK.

Harrison, J. F. (2001). Insect acid–base physiology. *Annu. Rev. Entomol.* **46,** 221–250.

Harrison, J. F., and Roberts, S. P. (2000). Flight respiration and energetics. *Annu. Rev. Physiol.* **62,** 179–206.

Hetz, S. (1994). "Untersuchungen zu Atmung, Kreislauf und Säure-Basen-Regulation an Puppen der Tropischen Schmetterlingsgattungen *Ornithoptera, Troides* und *Attacus.*" Friedrich-Alexander-Universität Erlangen–Nürnberg, Erlangen. [Ph.D. dissertation]

Hustert, R. (1975). Neuromuscular coordination and proprioceptive control of rhythmical abdominal ventilation in intact *Locusta migratoria migratorioides. J. Comp. Physiol.* **97,** 159–179.

Kestler, P. (1985). Respiration and respiratory water loss. *In* "Environmental Physiology and Biochemistry" (K. H. Hoffman, ed.). Springer-Verlag, Berlin.

Komai, Y. (1998). Augmented respiration in a flying insect. *J. Exp. Biol.* **201,** 2359–2366.

Lighton, J. R. B. (1996). Discontinuous gas exchange in insects. *Annu. Rev. Entomol.* **41,** 309–324.

Maddrell, S. H. P. (1998). Why are there no insects in the open sea? *J. Exp. Biol.* **201,** 2461–2464.

Manning, G., and Krasnow, M. A. (1993). Development of the *Drosophila* tracheal system. *In* "The Development of *Drosophila melanogaster*" (M. Bate and A. M. Arias, eds.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mill, P. J. (1985). Structure and physiology of the respiratory system. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology," pp. 517–593. Pergamon Press, New York.

Slama, K. (1988). A new look at insect respiration. *Biol. Bull.* **175,** 289–300.

Wasserthal, L. T. (1996). Interaction of circulation and tracheal ventilation in holometabolous insects. *Adv. Insect Physiol.* **26,** 297–351.

Wigglesworth, V. B. (1983). The physiology of insect tracheoles. *Adv. Insect Physiol.* **17,** 85–149.

River Blindness

Vincent H. Resh

University of California, Berkeley

Onchocerciasis, or river blindness, is a nonfatal human disease that affects the skin and vision and can ultimately lead to blindness in infected persons. The name river blindness reflects both the place where the disease is most common and severe, and its ultimate outcome.

The disease is caused by a filarial (threadlike) worm in the phylum Nematoda, *Onchocerca volvulus.* The disease is found in 37 countries, 30 of which are in Africa, 6 in the Americas, and 1 in the Arabian Peninsula; however, Africa, and particularly sub-Saharan West Africa, is by far the most affected area in terms of clinical manifestations of the disease, number of affected persons, and widespread occurrence.

The worm is most commonly transmitted to humans in Africa by black flies in the *Simulium damnosum* species complex, which contains about 40 different forms, several of which are distinct species. Other species of *Simulium* are vectors in other parts of the world. Larval and pupal stages of these species occur attached to near-surface substrates in fast-flowing rivers. Because the adults are capable of long flights (>200 km), the aquatic stages are considered the most vulnerable for control.

In humans, the disease takes three progressive forms: first, skin lesions and intense, often violent, itching occur; second, painless nodules containing adult worms form where tissues are thin over bones, such as pelvis, ribs, and scalp, and these nodules are the sites of reproduction for the worms; and third, eye lesions form that can lead to blindness. The third phase can be devastating to communities; 10% of the people in vast areas can be blind, with levels of adults being blind in individual villages sometimes exceeding 30%. This disease is most prevalent among the rural villages that occur in some of the poorest areas of the poorest countries of the world. In these

areas, subsistence (whether through agriculture or freshwater fishing) is difficult to achieve; with large portions of the adults being blind, the burden of the disease is devastating to these communities.

The life cycle involves both humans and black flies, and there are no known animal reservoirs of this disease. First, a biting, black fly female sucks blood from a human host; if that person is infected, she may also suck up some microfilaria (which are best thought of as worm embryos) of the parasite. Interestingly, evidence indicates that the worms migrate to the skin surface during the day, when black flies bite, which increases their chance of uptake by the biting females. Second, the microfilaria move from the fly's stomach to its thoracic muscles, where they pass through three larval stages (L_1, L_2, L_3), and the infective L_3 larva moves into the black fly vector's head and mouthparts. Third, in the course of biting, L_3 larvae are transferred from female black flies to humans. Fourth, the larval worm matures to an adult (which can be 40–45 cm in length) and adults mate (in the nodules that form), producing millions of microfilaria (each about 3 mm in length) that can be picked up from the skin of that human by a biting black fly. The death and disintegration of microfilaria result in an inflammatory reaction, which ultimately leads to itching, visual impairment, and eventually blindness.

The control of river blindness in West Africa has been one of the success stories of public health and economic development of the 20th century (Fig. 1). In 1974, the Onchocerciasis Control Programme in West Africa (OCP) began, with its objective being to eliminate onchocerciasis as a disease of public health importance and as an obstacle to socioeconomic development in this region. Eleven countries participated in this program, in a geographical region ranging from Senegal in the west to Benin in the east. Under the jurisdiction of the World Health Organization, and funded through the World Bank and 20 donor countries and organizations, the strategy was to interupt transmission of the blinding strain of the worm *O. volvulus* for a period of about 12 years, which is the life span of the adult worm. This was done by aerial application of seven different selective insecticides, weekly, to rivers infested with black flies. Because the female black fly lays batches of eggs at or near the water surface in fast-flowing water, these were the sites sprayed with insecticides. Much of the current understanding of the environmental impact of pesticides on aquatic ecosystems comes from techniques developed as part of this control effort. Beginning in the late 1980s, an ivermectin drug (a microfilaricide originally used as a veterinary product for the treatment of dog heartworm) was also distributed annually to people living in infected areas.

The success of the OCP is especially impressive when one considers the scope of the project: Onchocerciasis was eradicated in most parts of 11 different countries; at times, over 50,000 km of rivers had to be sprayed weekly for more than 10 years; no permanent environmental damage resulted from the spraying activities because of the increased reliance on insecticides that selectively targeted the black fly vectors;

FIGURE 1 Blind adults being led through villages were a common sight in West Africa before the control of river blindness. Photograph courtesy of The World Health Organization (WHO/TDR/W. Imber).

the drugs were distributed to almost 7 million people by over 22,000 community volunteers; millions of children have been born that will be free of this disease; the drugs were donated to the program without charge by the producer of them, Merck; the drug distribution program has recently been expanded to cover 19 more countries (and 157 million people); and plans include using this same system for dispensing drugs against lymphatic filariasis (elephantiasis) and pesticide-impregnated bed nets against malaria. The OCP ended in 2002; perhaps its most significant accomplishment is not only that the disease is under control, but also that the previously abandoned valleys, now free of river blindness, grow food for 17 million people in areas inhabited by the poorest of the poor, the people truly living at the "end of the road."

See Also the Following Article

Medical Entomology

Further Reading

Crosskey, R. W. (1990). "The Natural History of Blackflies." Wiley, Chichester, UK.

Crossky, R. W. (1993). Blackflies *(Simuliidae). In* "Medical Insects and Arachnids" (R. P. Lane and R. W. Crossky, eds.). Chapman & Hall, London.

Davies, J. B. (1994). Sixty years of Ochocerciasis vector control: A chronological summary with comments on eradication, reinvasion, and insecticide resistance. *Annu. Rev. Entomol.* **39,** 23–45.

Hougard, J. M., Yaméogo, L., Sekétéli, A., Boaten, B., and Dadzie, K. Y. (1997). Twenty-two years of blackfly control in the Ochocerciasis Control Programme in West Africa. *Parasitol. Today* **13,** 425–431.

Rostrum

R. F. Chapman
University of Arizona

The term rostrum is derived from a Latin word meaning a beak or snout and it is somewhat loosely used in entomology to refer to development of the head or mouthparts, which have some resemblance to a beak.

In Heteroptera, the mouthparts are modified for piercing and sucking. The mandibles and maxillae are slender stylets that pierce the host tissues. They are supported by the labium (rostrum) and slide in a groove that runs along its anterior face, being concealed within it when the insect is not feeding. The rostrum has between one and five segments. It is three-segmented in Reduviidae and four-segmented in Coreidae and Pentatomidae. In psyllids, aleyrodids, aphids, and coccids it appears to arise from between the anterior coxae, giving rise to the name Sternorrhyncha for the suborder. The rostrum does not enter host tissue as the stylets penetrate and commonly telescopes within the lower part of the head or becomes folded beneath the head, with its tip still surrounding the stylets (Fig. 1). Most reduviids have a ridged file between the forelegs against which they can rub the tip of the rostrum to produce an irregular pattern of sound that probably functions to deter potential predators but may also serve as an alarm signal.

In various other insect groups, the anterior or ventral part of the head capsule is elongate to form a rostrum. Among the weevils (Curculionidae), the frons and vertex are lengthened dorsally, and the biting and chewing mouthparts are at the extreme tip of the forwardly extending rostrum. In many species, the female has a longer rostrum than the male and in some the terminal mouthparts permit the insect to bore a hole in which eggs are placed. A rostrum is also characteristic of most Mecoptera (scorpion flies), but here it points downward rather than forward and the elongate basal sclerites of the mouthparts contribute to it. The rostrum of Cyclorrhaphous Diptera is also a ventral extension of the head that bears the maxillary palps anteriorly and the rest of the mouthparts ventrally.

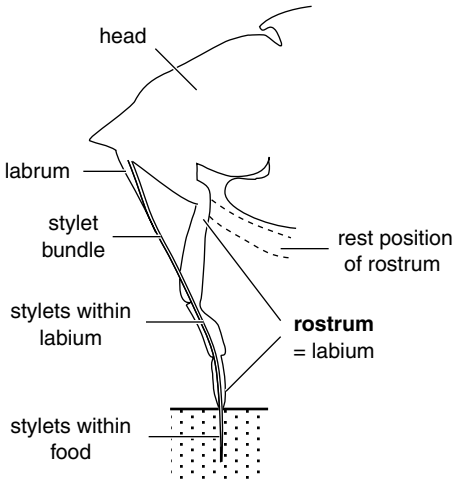

FIGURE 1 Diagram of the rostrum of a heteropteran. It is segmented and as the stylets enter the host tissues it folds beneath the head, still holding the stylets at its tip. When the insect is not feeding, the rostrum is held parallel with the body in the position shown by the dashed line. (Reproduced, with permission, from H. Weber, 1930, "Biologie der Hemipteren." Springer-Verlag, Berlin.)

Among termites, the rostrum of soldier termites in the subfamily Nasutitermitinae is different from these others. It is formed as a pointed projection of the frons and is independent of the mouthparts. Unlike the soldiers of most other termites, nasute termites have poorly developed mandibles but carry out their defensive function by ejecting a sticky spray from the opening at the tip of the rostrum.

See Also the Following Articles

Blood Sucking • Feeding Behavior • Mouthparts

Further Reading

Chapman, R. F. (1998). "The Insects: Structure and Function," 4th ed. Cambridge University Press, Cambridge, UK.

Royal Jelly

Eva Crane
International Bee Research Association

Royal jelly is secreted by the hypopharyngeal gland located below the pharynx in the head of young worker honey bees (brood food glands). The term royal jelly was probably first used by François Huber in Switzerland in 1792, as *gelée royale.* Young *Apis mellifera* worker bees secrete royal jelly into a queen cell, an extra large brood cell in which a female larva is fed on special food and develops into a queen (a sexually reproductive female). The workers then seal the larva inside, and it develops into a pupa and then an adult queen. A worker or drone larva receives food similar to royal jelly only for the first 3 days, and thereafter modified less rich food, which includes pollen and honey. A queen larva consumes about 25% more food than a worker larva, and its weight increases by 1300 times in 6 days.

COMPOSITION OF *A. MELLIFERA* ROYAL JELLY

Royal jelly, like beeswax, is a secretion of worker bees and as such has a more constant composition than honey or pollen, which are derived from plants. Of the components of royal jelly, the most interesting are perhaps 10-hydroxy-2-decenoic acid and vitamins in the B complex, which include thiamine, riboflavin, niacin, pyridoxine, pantothenic acid, biotin, and folic acid. Royal jelly contains very little vitamin C, and vitamins A and E are absent or nearly so; D and K are probably also absent.

PRODUCTION OF *A. MELLIFERA* ROYAL JELLY BY HUMANS

Well before the 20th century, the dramatic effect of royal jelly on female larvae that developed into queens created much interest. In the 1950s royal jelly was the first of the newer

hive products to be exploited. One factor that stimulated beekeepers to move toward royal jelly production was that it could be accomplished in areas where plant sources of nectar and pollen were inadequate for profitable honey production. Provided colonies were fed sufficient sugar syrup and pollen, they could be managed so that they produced royal jelly for the beekeeper. However, to produce it, skilled operators must carry out a sequence of labor-intensive procedures tied to a rigorous timetable.

Each hive used may be organized as follows. The queen of each colony is confined in a lower brood box by a queen excluder placed over it. Above the excluder is a box of honey combs, and above that the box in which the royal jelly will be produced. This contains framed combs of honey and pollen and also unsealed brood to attract nurse bees. Some of the frames of comb are replaced by frames containing three cross-bars, each supporting on its underside about 15 "queen cups," synthetic shallow wax cups with an opening on the bottom, similar to the cells that bees build when they start to rear queens. An operator transfers a worker larva 18–24 h of age from a worker comb into each queen cup, using a "grafting tool" that is rather like an insect pin bent into the required shape and mounted on a handle. There are a hundred or more of the queen cups in the box.

On 3 successive days a frame of "grafted" cells is placed in the top brood box. On day 4 the cells in the first frame contain the maximum amount of royal jelly (e.g., 235 mg each), so the operator removes the frame and extracts the royal jelly from the cups in it by aspiration. Fresh larvae are grafted into them, and the frame is reinserted in the hive, thus restarting the cycle.

The extracted royal jelly is strained to remove any wax and refrigerated as soon as possible; it can be kept for up to a year at 2°C.

PROPERTIES AND USES

Royal jelly shows wide-spectrum activity against bacteria (although none against fungi). It has been reported in at least one study to have bactericidal activity on *Bacillus metiens; Escherichia coli; Mycobacterium tuberculosis,* which causes tuberculosis; *Proteus vulgaris; Staphylococcus aureus;* and a *Streptococcus* species.

Royal jelly achieved, and has maintained, a place on the world market as a specialized dietary supplement and in cosmetics; by 1990 the annual production was probably between 500 and 600 tonnes. Some scientists have documented a reported feeling of general well-being in humans after consumption of royal jelly, thus justifying its use in anorexia, emaciation, or loss of muscular strength; it has also been recommended for older people. Many of the reported effects of royal jelly on humans and other mammals may, however, be produced much less expensively by other substances.

A queen honey bee differs from a worker in her length of life (several years instead of several weeks or months) and in her reproductive ability. Some people therefore believe that royal jelly can also prolong the life of humans, as well as improving their vigor and sex life. However, these properties have not been substantiated.

WORLD TRADE

Royal jelly production was already proving profitable in France in 1953, and 1.5 tonnes a year was being produced by 1958. Production was established in Cuba before 1957, and in the early 1960s 12 other countries were reported to produce it, including France and Japan (1.5 tonnes each), Canada, Israel, Taiwan, and Korea. The price at that time was from U.S. $220 to $500 per kilogram. By 1984 China and Taiwan were producing 400 and 234 tonnes a year, respectively, and the world price had dropped to U.S. $70 per kilogram. Eastern Asia is still the main center of the world's royal jelly production, and Japan is both a large producer and a large importer and consumer.

See Also the Following Articles

Apis Species • Commercial Products from Insects

Further Reading

Crane, E. (1990). "Bees and Beekeeping: Science, Practice and World Resources," Chap. 14. Heinemann Newnes, Oxford.

Mizrahi, A., and Lensky, Y. (1997). "Bee Products: Properties, Applications, and Apitherapy." Plenum, New York.

Schmidt, J. O., and Buchmann, S. L. (1992). Other products of the hive. *In* "The Hive and the Honey Bee" (J. Graham, ed.), Chap. 22. Dadant, Hamilton, IL.

S

Salivary Glands

Gregory P. Walker
University of California, Riverside

Salivary glands are glands associated with the mouth or oral cavity and produce secretions (saliva) that are mixed with the food during feeding and are ingested along with the food. Among insects, there are four pairs of glands associated with the mouth or oral cavity, although all four generally are not present in the same insect. Each of these is associated with and is named after its associated mouthpart: the mandibular, maxillary, hypopharyngeal, and labial glands. The presence or absence of each of these glands varies among insect species, and among holometabolous insects, a given gland may be present in only one life stage. For example, in Lepidoptera, mandibular glands occur in the larval stage, but not in the adult. The four glands serve a variety of functions, and the same gland may serve different functions in different species or even in different life stages of the same species. Within a given species, one or more of these four pairs of glands usually function as salivary glands.

GLANDS ASSOCIATED WITH THE MOUTH

Mandibular glands, found in many insects, function as the main salivary glands in Lepidoptera larvae. A common function of mandibular glands in social Hymenoptera is to produce pheromones such as alarm pheromones in ants and honey bees, and the queen substance in queen honey bees. Function can vary with age. For example, the mandibular glands of older honey bee workers (which perform mostly foraging tasks) produce an alarm pheromone, whereas the mandibular glands of young workers (which perform mostly nursing tasks) produce secretions, called royal jelly, that are fed to larvae in differential quantities to control whether a given larva will develop into a queen or worker. Mandibular glands of stingless bees serve in defense and produce a burning sensation when ejected onto the victim.

Maxillary glands occur in Protura, Collembola, some Heteroptera, and some larval Neuroptera and Hymenoptera. They are believed to provide secretions to lubricate the mouthparts, and thus serve one of the salivary functions.

Hypopharyngeal glands occur in the Hymenoptera, and in honey bee workers they produce secretions that are fed to the larvae (a substance different from royal jelly). Hypopharyngeal glands are vestigial in honey bee queens and absent in males. The hypopharyngeal glands of honey bee workers also produce invertase, a common salivary enzyme that hydrolyzes sucrose, and another enzyme that oxidizes glucose to an acid, which is believed to serve as a preservative in honey.

Labial glands occur in the great majority of insect orders (an important exception is the Coleoptera), and most commonly function as salivary glands, but in the larval stage of some groups of silk-producing insects, such as Lepidoptera, Trichoptera, and Hymenoptera, labial glands are the silk-producing organs either throughout the larval stage or at its end, just before pupation. In the Psocoptera, adults have two pairs of labial glands, one pair functioning as silk glands and the other pair functioning as salivary glands. In the primitive orders Collembola and Thysanura, which lack Malpighian tubules (the usual excretory organs of insects), the labial glands function as excretory organs.

STRUCTURE AND FUNCTION OF SALIVARY GLANDS

The salivary glands of most insects are labial glands, which are the focus of this section. Labial salivary glands have been examined in detail in relatively few insect species, and there is great variation among the species examined (Fig. 1). This is not surprising, considering the great variation in mode of feeding

(e.g., chewing, piercing-sucking, non-piercing-sucking, sponging, etc.) and types of food consumed by different insect species.

General Description

Several aspects of structure and function are common to most or all variations of insect labial salivary glands. The glands occur in pairs, and the ducts from each gland usually join to form a single common duct that opens to the oral cavity at a single orifice (Fig. 1). Even though the glands originate in the labial segment, the orifice usually occurs just behind or on the hypopharynx, and the glands often extend back into the thorax and even as far back as the abdomen (e.g., Fig. 1A,B).

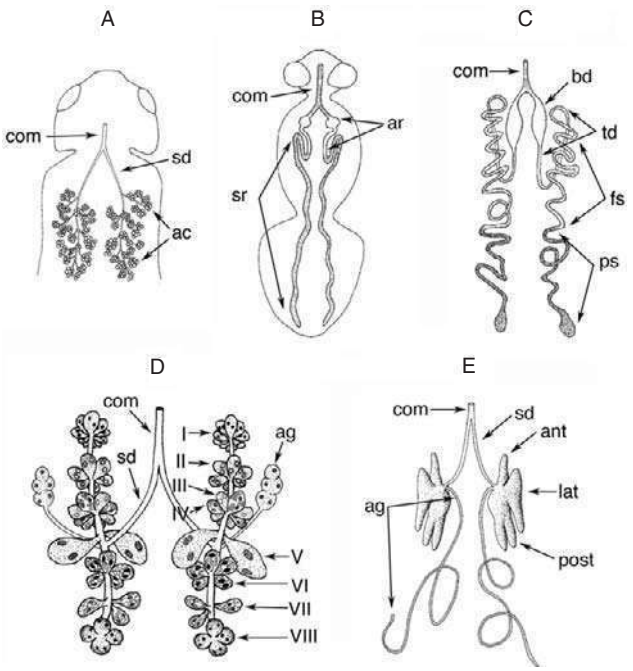

FIGURE 1 Salivary glands of representative insects. (A) The locust, *Locusta.* [After Chapman R. F. (1998). "The Insects: Structure and Function." 4th ed., Fig. 2.16. With permission of Cambridge University Press.] (B) Adult blowfly, *Calliphora.* [After Berridge, M. J., and Prince, W. T. (1971). The electrical response of isolated salivary glands during stimulation wtih 5-hydroxytryptamine and cyclic AMP. *Philos. Trans. R. Soc. Lond. B* **262,** 111–120. Courtesy of the Royal Society of London.] (C) The tobacco hornworm, moth *M. sexta.* [After Leslie. R. A., and Robertson, H. A. (1973). The structure of the salivary gland of the moth *(Manduca sexta). Z. Zellforsch. Mikrosk. Anat.* **146,** 553–564. Copyright Springer-Verlag GmbH & Co. KG. Used with permission.] (D) The beet leafhopper, *C. tenellus,* showing cell types I through VIII. [Modified after Wayadande, A. C., Baker, G. R., and Fletcher, J. (1997). Comparative ultrastructure of the salivary glands of two phyopathogen vectors, the beet leafhopper, *Circulifer tenellus* (Baker) and the corn leafhopper, *Dalbulus maidis* DeLong and Wolcott (Homoptera: Cicadellidae). *Int. J. Insect Morphol. Embryol.* **26,** 113–120.] (E) The large milkweed bug, *O. fasciatus.* [Modified after Miles, P. W. (1967). The physiological division of labour in the salivary glands of *Oncopeltus fasciatus* (Dall.) (Heteroptera; Lygaeidae). *Aust. J. Biol. Sci.* **20,** 785–797.] Abbreviations: ac, acini; ag, accessory gland; ant, anterior lobe; ar, absorptive region; bd, bulbous duct region; com, common salivary duct; fs, fluid secretion region; lat, lateral lobe; post, posterior lobe; ps, protein secretion region; sd, salivary duct; sr, secretory region; td, thin duct region.

The glands are suspended in the hemocoel and are constantly bathed in hemolymph. The glands generally have at least two regions: a secretory region and a reabsorptive region. Generally, the lumen of the salivary duct is lined with cuticle, at least at the end closest to its opening.

The secretory region produces the primary saliva. The major component of saliva is water. Water is transported from the hemolymph across cells of the salivary gland and into the lumen of the gland. Movement of water from the blood to the gland lumen is accomplished by active transport of potassium or sodium ions from the hemolymph to the lumen, causing water to move from the hemolymph to the lumen down an osmotic gradient. Cells responsible for water transport generally have deep infoldings of the cell membrane and/or dense microvilli on the side of the cell adjacent to the lumen of the gland. This serves to greatly increase the cell's luminal surface area, and also serves to enclose very narrow extracellular spaces into which ions are pumped. The enclosed nature of the spaces helps contain the ions to keep their concentration high, thus facilitating the osmotic movement of water from the cell into the space. The infoldings and microvilli usually are associated with abundant mitochondria to provide the energy for the ion pumps. The secretory region of the gland also synthesizes proteins, such as salivary enzymes, and other organic components of the saliva. Cells responsible for secretion of these components generally possess extensive endoplasmic reticulum, Golgi bodies, and secretory granules that synthesize and transport (intracellularly) the secretions. There may be one or several different types of cell in the secretory region. It should be noted that salivary components are not necessarily produced by the salivary glands themselves but may be produced elsewhere in the body and transported to the salivary glands via the hemolymph.

The reabsorptive region of the salivary gland reabsorbs potassium or sodium ions from the saliva and transports them back into the hemolymph. As a result, potassium and sodium ions are conserved, and the saliva is usually hypotonic to the hemolymph. Reabsorptive cells often have infoldings, especially on their basal side (hemolymph side), to increase surface area. These infoldings, however, tend not to be tightly enclosed (unlike the lumen side of water-secreting cells in the secretory region), to facilitate movement of secreted ions into the hemolymph and away from the cells, thus reducing the osmotic gradient, which would cause the cells to lose water. Reabsorptive cells also have abundant mitochondria to power the active transport of ions from saliva to hemolymph.

Acinuous Salivary Glands

The salivary glands of many insects are composed of clusters of acini, or saclike glandular structures (Fig. 1A) comprising at least two types of cell. Each acinus empties into a cuticle-lined duct, and the ducts of different acini fuse, eventually forming a common duct that leads to an opening just behind or on the hypopharynx. The anatomy of acinuous salivary

glands varies greatly among different insects. The acinuous glands of locusts and cockroaches have been particularly well studied and are described next.

LOCUSTS AND COCKROACHES The gross anatomy of locust and cockroach salivary glands is illustrated in Fig. 1A. The acini are the secretory region of the salivary gland and consist of peripheral cells and central cells. The peripheral cells have deep, microvilli-lined invaginations that are contiguous with extracellular canaliculi that open into the gland's duct. The peripheral cells transport water from the hemolymph to the canaliculi, and from there, the water empties into the duct. Additionally in locusts, the peripheral cells seem to synthesize and secrete other salivary components. The central cells of both locusts and cockroaches synthesize salivary enzymes and other salivary components, and secrete them into the salivary duct. The walls of the salivary ducts of locusts and cockroaches are one cell thick, and the lumen is lined with a thin cuticle. The salivary duct contains the reabsorptive region of the gland, and in some cockroaches, a small part of the duct adjacent to the acinus is secretory.

Tubular Salivary Glands

Some groups of insects, such as Lepidoptera, Diptera, and Siphonaptera, possess tubular salivary glands. Generally, the walls of tubular salivary glands are one cell thick and the lumen is lined with thin cuticle. Tubular salivary glands are divided into several regions along the length of the gland. The number of regions and anatomical details of the regions vary among insect groups. Two examples of tubular salivary glands are described.

BLOWFLY ADULTS There are a pair of tubular salivary glands, each with three regions (Fig. 1B). The apical region is the longest and is the secretory region; the shorter middle region is the reabsorptive region; and the short common duct at the proximal end opens to the exterior of the body on the hypopharynx.

The blowfly's secretory region has a long distal section located mostly in the abdomen and a shorter proximal section located in the thorax. Cells in the distal section serve a dual function: they move water from the hemolymph into the lumen of the gland, and they synthesize and secrete salivary enzymes and other salivary components. The cell surface adjacent to the lumen of the gland's central duct encloses extensive canaliculi into which the cells secrete their products. Secretions then move from the canaliculi to the gland's duct. Cells in the proximal section are generally similar in appearance to those in the distal section, but they do not contain secretory granules. Thus, they do not seem to secrete enzymes and probably secrete only water and ions. The reabsorptive region of the blowfly's salivary glands consists of a single cell type that is believed to be responsible for reabsorbing ions from the saliva back to the hemolymph. Finally, the common duct of the blowfly's salivary glands comprises highly flattened cells that seem to play no role in the secretion or reabsorption of any salivary components. This region of the salivary gland is very short and opens on the hypopharynx.

SPHINGID MOTH In the tobacco hornworm, *Manduca sexta,* adults have a pair of tubular salivary glands that are divided into four regions (Fig. 1C). In the apical region, proteinaceous material is synthesized in the extensive rough endoplasmic reticulum and Golgi bodies and is stored in large vacuoles before eventual release into the lumen of the gland. The second region transports water from the hemolymph to the lumen. Cells in this region have the characteristic structure of water-secreting cells, to accommodate what is believed to be the primary function of this region; but in addition, the presence of rough endoplasmic reticulum and Golgi bodies suggests that these cells may secrete more than just water. The third and fourth regions, called the thin duct and bulbous duct, both seem to have a reabsorptive function, moving ions back from the saliva and into the hemolymph. Cells in these two regions differ in the structure of their surface adjacent to the lumen, but the reason for the difference is unknown. After the fourth region, the right and left glands fuse forming the common duct. Cells of the common duct are unspecialized and probably play no role in saliva production.

Salivary Glands of Hemiptera

The most complex insect salivary glands that have been studied occur in the Hemiptera. This complexity is undoubtedly related to the piercing-sucking mode of feeding in this taxon, where saliva is injected into the food substrate via a specialized salivary canal in the elongate maxillary stylets. In this mode of feeding, solid substrates must be pierced and then the food, which is often initially solid, must be liquefied before ingestion through the maxillary food canal. These processes depend greatly on a multitude of salivary components that serve different functions. Consequently, there are usually many distinct types of secretory cells in the same gland, each producing different salivary components. Furthermore, many phytophagous hemipterans produce two distinct types of saliva at different times in the feeding process: sheath saliva and watery saliva. Sheath saliva consists mostly of lipoprotein and is secreted incrementally as the stylets advance through the plant tissue. It gels shortly after secretion and forms a continuous solid sheath around the stylets. As a consequence, only the stylet tips come in direct contact with the plant tissue; the shaft of the stylet bundle is encased by the sheath. Watery saliva, as the name implies, is dilute and does not gel. It contains mostly water and various enzymes.

In general, salivary glands of the Hemiptera are divided into two main parts, the principal gland and the accessory gland. The principal gland is often subdivided into two or more

lobes. The principal and accessory glands are served by their respective ducts, and the ducts from each fuse to form a lateral salivary duct. The lateral ducts from each side fuse to form a median duct that leads to the salivary pump (described later in this article). There is considerable variation in the salivary glands among the Hemiptera, even within the same family. Two examples are described.

BEET LEAFHOPPER The principal salivary glands of the beet leafhopper, *Circulifer tenellus,* are divided into an anterior and posterior lobe, each served by its own duct (Fig. 1D). The ducts from each lobe fuse to form the lateral salivary duct. The accessory gland is not subdivided, and its duct joins the others near the point where they fuse to form the lateral duct. The principal glands have eight different cell types that are arranged in rings around the duct of each lobe, five in the anterior lobe and three in the posterior lobe. All eight cell types are secretory, possessing abundant endoplasmic reticulum and/or secretory granules, and all have intracellular canaliculi that come in contact with the salivary duct. Details of the cells' fine structure and the staining properties of the secretory granules differ sufficiently to indicate that each cell type produces different components of the saliva. Accessory gland cells have some features typical of water-secreting cells but also have abundant endoplasmic reticulum, Golgi bodies, and secretory granules, suggesting that they secrete water and other salivary components.

MILKWEED BUG As in most hemipterans in the suborder Prosorrhyncha, the principal salivary glands of the large milkweed bug, *Oncopeltus fasciatus,* (Fig. 1E) are divided into discrete lobes that are much more compact than the lobes just described for the beet leafhopper. The lobes comprise a mass of cells with a distinct glandular lumen. The accessory glands and the three lobes of the principal glands each secrete their own salivary components. The anterior and lateral lobes secrete two different components that mix to form the salivary stylet sheath (described earlier). The lateral lobe secretes the bulk of the sheath saliva protein, including most of the components that form the hydrogen bonds that solidify the salivary sheath, while the anterior lobe secretes the components that form most of the disulfide bonds. The posterior lobe produces salivary digestive enzymes, such as amylase and esterase. The accessory gland supplies the bulk of the water in the saliva, as well as polyphenoloxidase and possibly mucoid substances.

CONTROL OF SECRETION, INNERVATION

Secretion of saliva generally is stimulated either by direct innervation or by neurohormonal factors that are released into the blood by secretory neurons. Innervation of the salivary glands varies considerably among different insects. Innervation can come from the subesophageal ganglion, thoracic ganglia, the stomatogastric nervous system, the median–transverse nervous system, or a combination of these.

Two neurotransmitters, serotonin and dopamine, are commonly found in neurons innervating the salivary glands, and each may stimulate different aspects of salivation. For example, in the American cockroach, serotonin induces secretion of proteinaceous saliva, whereas dopamine induces secretion of nonproteinaceous saliva. Other neurotransmitters occur in neurons innervating salivary glands, but their roles are not understood. The diversity of neurotransmitters suggests that control of salivation is a complex process.

The salivary glands of some insects such as blow flies lack direct innervation, and salivation is induced by one or more neurohormonal factors released into the blood. One of these factors seems to be serotonin.

STRUCTURES ASSOCIATED WITH SALIVARY GLANDS

Salivary Reservoirs

Some insects, such as cockroaches, have a pair of distensible salivary reservoirs for storage of saliva. In cockroaches, each reservoir has its own duct, which joins with the duct of its associated salivary gland. The combined gland/reservoir ducts from each side fuse to form the common salivary duct leading to an opening on the hypopharynx. A valve near the orifice of the common duct on the hypopharynx opens and closes as the hypopharynx is raised and lowered during feeding, thus controlling release of saliva. When the insect is not feeding, the hypopharynx is in a lowered position, closing the valve, and saliva produced by the salivary glands then backs up into the reservoirs where it is stored.

Salivary Pumps

Many insects with piercing-sucking mouthparts inject saliva into their food for various purposes. Often a pumping mechanism is used to inject the saliva through elongate hypodermic needle-like mouthparts. In many Hemiptera, the pump is a hollow chamber near the hypopharynx and is referred to as a salivary pump or salivary syringe.

FUNCTIONS OF SALIVA

General

Perhaps the most fundamental and ubiquitous function of saliva in insects is lubrication of the mouthparts and lubrication of the food bolus to assist its transport through the foregut. Lubrication can be achieved primarily by water, the most abundant constituent in saliva. Water in the saliva also can dissolve components in the food, such as sugars, which then become detectable by chemoreceptors on the mouthparts. Thus, saliva also can aid in food recognition.

The most common class of organic constituents of saliva consists of digestive enzymes, such as amylase, invertase, various proteases, and lipases. In many insects with chewing

mouthparts, salivary enzymes are mixed with the food during chewing and swallowing, and initiate digestion. However, the midgut usually is the main site of production and secretion of digestive enzymes, and salivary enzymes in these insects generally play only a secondary role in digestion. In other insects, salivary digestive enzymes provide the main digestive function. This is especially common in piercing-sucking insects, such as many Heteroptera, in which digestive enzymes are injected into the food. The enzymes then break down and liquefy the food, and the digested, liquefied food is sucked up through the mouthparts.

Predators

Predaceous insects that have piercing-sucking mouthparts often capture and eat prey that are as big or even bigger than themselves. Large prey size is not nearly as common in predators with chewing mouthparts. This is because many predators with piercing-sucking mouthparts use the mouthparts to inject a salivary toxin into the prey, which enables them to subdue prey without having to be large and strong enough to physically overpower them. Thus, in these insects, saliva assists prey capture and gives the insects a potentially larger range of prey than their chewing mouthpart counterparts. Venoms in predaceous piercing-sucking insects often are accompanied by salivary hyaluronidase, which breaks down hyaluronic acid, an important "intercellular cement" in insects. Hyaluronidase is believed to serve as a spreading agent for toxins (as well as for digestive enzymes), assisting their penetration between cells by breaking down the intercellular cement.

Blood Feeders

Salivary components of blood-feeding insects serve several functions. These have been best studied in mosquitoes, other biting flies, and kissing bugs. Blood feeding is the most dangerous time in the lives of these insects, and survival is greatly enhanced by the ability to complete the task quickly and escape before being detected by the host. Salivary enzymes help to shorten the time it takes to acquire a blood meal. Blood vessels occupy only a small volume of skin tissue, and thus locating blood with the mouthparts can take considerable time. During probing, many blood feeders damage capillaries or tiny blood vessels by random movement of the stylets, and small subcutaneous pools of blood (hematomas) form in the vicinity of the damaged vessels. This increases the volume of the blood in the skin, and thus increases the probability that the stylets will locate a blood source, reducing the time required to locate blood. Successful formation of hematomas is greatly facilitated by factors that inhibit blood clotting, and consequently, the most common salivary components in these insects are factors that inhibit clotting. Clotting comprises two general processes, platelet aggregation and coagulation. In the small vessels used by most blood-feeding insects, platelet aggregation is the more important of the two, being capable of plugging a damaged vessel in a matter of seconds. One of the compounds that initiates platelet aggregation is ADP, and the saliva of many different blood-feeding insects contains apyrase, which is an enzyme that breaks down ADP. The importance of apyrase is demonstrated in mosquitoes, where the time required to complete a blood meal is directly dependent on the amount of apyrase injected by the mosquito. In addition to salivary components that interfere with platelet aggregation, some blood-feeding insects have salivary components such as antithrombins that inhibit coagulation. Also, substances that inhibit vasoconstriction occur in some blood-feeding insects, thus inhibiting the host's attempt to restrict the flow of blood to the feeding site.

Anticoagulants also serve another function: they prevent blood from coagulating in the food canal. Coagulation would clog the food canal and lead to starvation and death. Finally, antihistamines in the saliva of some blood feeders may act as anti-inflammatory agents and reduce the probability that the feeding insect will be detected by the host.

Herbivorous Hemipterans

Many or most phytophagous hemipterans produce two kinds of saliva, sheath saliva and watery saliva, which were described briefly earlier. Sheath saliva forms a continuous solid sheath around the stylets, and several functions for it have been proposed. One likely function is to reduce friction between the stylets and plant tissue, facilitating advancement and withdrawal of the stylets. Another function may be to shield the moving stylets from plant cells, and thus avoid triggering a defensive response by the plant that could include hypersensitive reactions or release of plant defensive chemicals. This may be especially important in sternorrhynchan hemipterans like aphids and whiteflies, which feed primarily on sap from phloem sieve elements that lie deep in plant tissue. These insects carefully weave their stylets between and around plant cells from the plant surface to the sieve elements, and successful extraction of sap from the sieve elements may be dependent on avoiding the triggering of plant defensive responses during penetration to the sieve elements. In addition to mechanically shielding the stylets from plant cells to avoid plant defenses, salivary sheaths contain the enzyme polyphenoloxidase, which has been proposed to serve the function of oxidizing plant defensive chemicals and converting them to more harmless forms.

Watery saliva contains assorted enzymes and other components that vary among species. Some, like proteases and amylases, serve a digestive function, breaking down insoluble plant constituents into soluble forms that can be ingested through the stylet food canal. Others, like pectinases, break down pectin, which is the "intercellular cement" in plants, and loosen the adhesion between adjacent cells. For hemipterans like aphids and whiteflies, whose stylets penetrate between cells until they reach their actual ingestion site (phloem sieve

elements), pectinase facilitates the penetration of stylets between cells by loosening the intercellular cement. After penetrating a phloem sieve element, aphids and whiteflies inject saliva into it, an action that is believed to make the sieve element more suitable for ingestion and to prevent the sieve element from clogging.

The water component of watery saliva also can have a critical role other than simply serving as a carrier for enzymes. Many phytophagous hemipterans (especially phytophagous Heteroptera) feed by a method known as "lacerate and flush." In this feeding method, the stylets are inserted into the plant, and a pocket of cells beneath the surface is liquefied by the combined action of digestive enzymes in the saliva and mechanical laceration by repeated thrusts of the stylets. Once the pocket of cells has been liquefied, the nutrient-rich liquid is flushed out of the pocket by copious secretion of watery saliva and sucked up through the stylet food canal.

Gall Formers

Plant galls are produced by several groups of phytophagous insects that occur especially, but not exclusively, in the dipteran family Cecidomyiidae and the hymenopteran family Cynipidae, as well as many species of sternorrhynchan Hemiptera and some Thysanoptera. The abnormal tissue and cell growth characteristic of plant galls is caused by secretions from the insect, usually salivary secretions, that mimic plant growth hormones or serve as molecular signals that redirect plant cell growth from its normal course to an abnormal form that serves the needs of the gall maker.

Construction of Shelters and Webs

Silk is produced by many insects for a variety of functions such as construction of pupal cocoons in many Lepidoptera, Hymenoptera, and Siphonaptera, and construction of larval retreats or food-gathering nets in most Trichoptera larvae and some chironomid larvae. Many Psocoptera use silk to construct sheetlike shelters under which they aggregate and also use silk to attach their eggs to the substrate. Weaver ants use silk to tie together leaves to construct their arboreal nests, but interestingly, only larvae produce silk; so to weave leaves together with silk, the workers hold the larvae in their jaws and use them as silk dispensers. Not all insects that produce silk do so with their labial glands, but all the examples just cited do, representing some of the diverse uses of these specialized salivary secretions.

Mucoid secretions are produced by the salivary glands of several groups of Diptera. Some Diptera, such as *Drosophila,* produce salivary mucopolysaccharides that glue the puparium to the substrate. In some sciarids and fungus gnats (Mycetophilidae), mucoid secretions are used as a "slime trail" to facilitate larval locomotion, much like terrestrial snails. Larvae of some predaceous fungus gnats use these sticky mucoid salivary secretions to capture prey. The most fascinating of these are the New Zealand glow worms, which construct a silken retreat from which they dangle silk threads that are covered with sticky mucoproteins to trap prey. The larvae reside in their silken retreats, and at night, they produce light by bioluminescence to attract nocturnal flying insects to their traplines.

Trophallaxis

Trophallaxis is the exchange of food between two individuals. The food exchanged may be salivary secretions or regurgitated gut contents. Larvae of many ants and wasps are dependent on adults to feed them. In exchange for being given food by the adult, the larvae of many species secrete a salivary fluid that is greedily consumed by the adult that provided the food. This stimulus for adults to give up food to a larva may have played a role in the evolution of eusociality in the Hymenoptera, in which adult females readily feed larvae that are not even their own offspring.

Others

Larval warble flies bore their way through the subcutaneous tissues of their mammalian hosts. To facilitate movement through the host's tissues, they secrete a salivary collagenase, which breaks down collagen, a main component of connective tissue.

The saliva of some moths contains an enzyme called cocoonase that weakens the silk cocoon. It is produced by the newly eclosed moth to assist its escape from the pupal cocoon.

See Also the Following Articles

Blood Sucking • Digestion • Feeding Behavior • Gallmaking • Mouthparts • Silk Production

Further Reading

Ali, D. W. (1997). The aminergic and peptidergic innervation of insect salivary glands. *J. Exp. Biol.* **200,** 1941–1949.

Berridge, M. J., and Prince, W. T. (1971). The electrical response of isolated salivary glands during stimulation with 5-hydroxytryptamine and cyclic AMP. *Philos. Trans. R. Soc. Lond. B* **262,** 111–120.

Chapman, R. F. (1998). "The Insects. Structure and Function." 4th ed., pp. 30–34. Cambridge University Press, Cambridge, U.K.

House, C. R., and Ginsborg, B. L. (1985). Salivary gland. *In* "Comprehensive Insect Physiology, Biochemistry, and Pharmacology," Vol. 11, "Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), pp. 195–224. Pergamon Press, Oxford, U.K.

Kendall, M. D. (1969). The fine structure of the salivary glands of the desert locust, *Schistocerca gregaria* Forskål. *Z. Zellforsch. Mikrosk. Anat.* **98,** 399–420.

Lehane, M. J. (1991). "Biology of Blood-Sucking Insects." HarperCollins/Academic Press, London.

Leslie, R. A., and Robertson, H. A. (1973). The structure of the salivary gland of the moth *(Manduca sexta). Z. Zellforsch. Mikrosk. Anat.* **146,** 553–564.

Miles, P. W. (1967). The physiological division of labour in the salivary glands of *Oncopeltus fasciatus* (Dall.) (Heteroptera: Lygaeidae). *Aust. J. Biol. Sci.* **20,** 785–797.

Miles, P. W. (1972). The saliva of Hemiptera. *Adv. Insect Physiol.* **9,** 183–255.

Miles, P. W. (1999). Aphid saliva. *Biol. Rev.* **74,** 41–85.

Oschman, J. L., and Berridge, M. J. (1970). Structural and functional aspects of salivary fluid secretion in *Calliphora*. *Tissue Cell* **2,** 281–310.
Ribeiro, J. M. C. (1987). Role of saliva in blood-feeding by arthropods. *Annu. Rev. Entomol.* **32,** 463–478.
Sutherland, D. J., and Chillseyzn, J. M. (1968). Function and operation of the cockroach salivary reservoir. *J. Insect Physiol.* **14,** 21–31.
Wayadande, A. C., Baker, G. R., and Fletcher, J. (1997). Comparative ultrastructure of the salivary glands of two phytopathogen vectors, the beet leafhopper, *Circulifer tenellus* (Baker), and the corn leafhopper, *Dalbulus maidis* DeLong and Wolcott (Homoptera: Cicadellidae). *Int. J. Insect Morphol. Embryol.* **26,** 113–120.

Scale Insect

see *Sternorrhyncha*

Scales and Setae

Shaun L. Winterton
North Carolina State University

Setae are multicellular protuberances on the arthropod cuticle used primarily for mechanoreception. In all groups of arthropods and especially insects, the role of the setae has evolved from simple mechanoreception to various other functions, including defense, locomotion, prey capture, pheromone dispersal, sexual display, preening, and camouflage. Setae are often highly modified, with one common modification being that they may be flattened into a broad, plate-like scale. This article specifically examines the tremendous diversity in shape, structure, and function of setae and scales used by different insect groups and at different life stages. This diversity in setae type is then characterized according to their four major functions in the biology of the insect: mechanoreception, camouflage, defense, and pheromone dispersal.

STRUCTURAL MORPHOLOGY

The terminology applied to scales and setae has historically been confused, with numerous, often interchangeable, terms used by researchers to describe the various types of structures (e.g., hairs, bristles, trichiae, aculei, chetae). Each seta, or trichoid sensillium, is a multicellular protuberance with specifically differentiated cells, with the most diagnostic landmark being its socket. The term trichoid sensillum applies to chemoreceptors as well as mechanoreceptors. The term hair is incorrect when applied to insects because hairs are morphologically different from setae and are considered a characteristic of mammals and not arthropods. Trichobothria is a term occasionally used to describe large, nontapering seta often associated with the mouthparts and genitalia. Other types of cuticular protuberances not considered here include the various minute projections of the cuticle (e.g., pruinescence, velutum, pollen), collectively termed microtrichia, and large sclerotized extensions of the cuticle (e.g., spines, tubercles).

Most setae form the covering of the body surface, which in very high densities is called a pile. Usually a mechanoreceptor, a typical seta, is composed of four cells: (1) A sensory cell, which innervates the seta, is surrounded by a (2) thecogen cell, which acts as an auxiliary cell by secreting the dendrite sheath. During seta development the (3) trichogen cell secretes the tapered or scale-like protuberance, whereas the (4) tormogen cell is responsible for formation of the socket. While most setae develop as mechanoreceptors, some will become secondarily noninnervated and lose their sensory function. These setae may take on a different role (such as aerodynamics, sexual display, preening) in the functioning of the insect.

Scales are modified setae that have a flattened blade with longitudinal ridges, sometimes with serrate edges (Figs. 1A and 1B). Scales are usually inclined relative to the cuticle, overlapping each other when present in sufficient densities. Each scale has a narrow pedicel, and the enlarged blade may be either gradually attenuated to an apex or truncated.

Scales are found in most groups of insects and serve a variety of functions. In true flies (Diptera), scales are not common but in certain species can be found on the wings (e.g., mosquito, *Anopheles annulipes*—Culicidae), legs (e.g., *Trichopoda* spp.—Tachinidae), or body (e.g., *Metatrichia* spp.—Scenopinidae). Scales are also found on the elytra of various beetle (Coleoptera) families (e.g., Elateridae, Curculeonidae, Buprestidae), but the function of such scales in Coleoptera and in Diptera is unknown. Scales on the body and wings of Lepidoptera have been studied in greater detail and are known to function in cryptic coloration, thermal regulation (e.g., butterfly, *Colias* spp.), and aposematism (warning coloration). In butterflies, wing scales contribute to lift during flight but not to drag, thus enabling them to glide for longer periods.

Lepidopteran wing scales have a specialized structure and are divided into two types. Primitive-type scales, found in nonditrysian Lepidoptera, are solid with longitudinal ridges (Fig. 1C). Normal-type scales, found in the ditrysian Lepidoptera, are composed of superior and inferior lamellae, with an internal lumen subdivided by internal supports called trabeculae (Fig. 1D). Whereas the inferior lamella is smooth, the superior lamella is usually covered with longitudinal ridges, interconnecting transverse ridges (flutes), and/or perforations (windows). Some groups of nonditrysian Lepidoptera have a mixture of the two types, with normal-type scales layered over the primitive-type scales.

Wing color in Lepidoptera is produced by wing scales. Individual scales are usually a single color which may be generated by any one or a combination of pterins (red, yellow, white), melanins (black), flavonoids (white), carotenoids (blue, yellow), papilliochromes (cream-yellow), and ommachromes

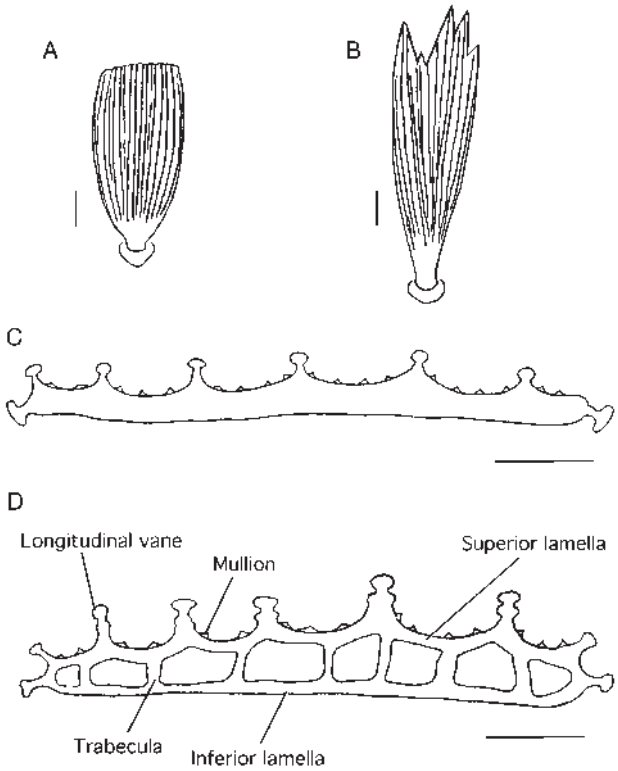

FIGURE 1 Lepidopteran scales. (A) Truncated wing scale, (B) serrate body scale, (C) cross section of primitive-type wing scale, (D) cross section of normal-type wing scale. Scale line length, 0.1 mm.

(red). Scales also produce color by diffraction; regular longitudinal ridge spacing produces various diffraction colors, whereas irregular spacing produces overlap in color spectra and is perceived as white light. Interference colors, from which light is reflected from a series of superimposed surfaces separated by distances equivalent to light wavelengths, result in selective phasing in and out of various colors. This type of color is important for producing bright color hues in adult butterflies (e.g., *Morpho* spp.) and is produced from a series of longitudinal vanes along the scale blade. Each vane is supported by a series of thickenings, or mullions, which act in concert with other vanes to form a series of reflecting surfaces and produce interference colors.

FUNCTIONAL TYPES

The functions of setae are sensory (e.g., touch, taste, and smell), but in many insects the original sensory function of some setae has been discontinued though cell death that creates a noninnervated seta. These setae often have more specialized roles, sometimes with great morphological modification. There is incredible diversity in the shape, structure, and function of both the seta and its morphological derivative, the scale. A detailed examination of almost any insect will reveal a myriad of seta or scale types, often with vastly differing roles in the biological functioning of the organism. The following classification is clearly artificial and categorizes the diversity of setae and scales into functional types for ease of discussion.

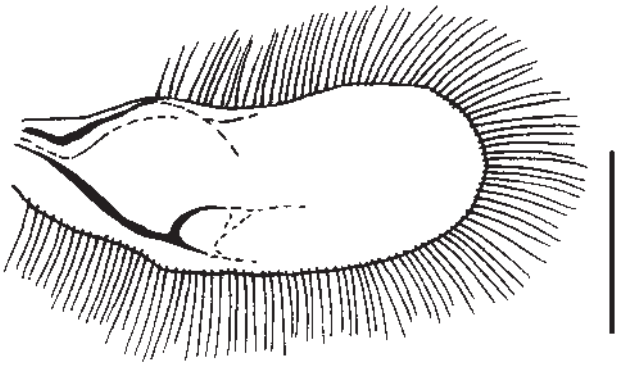

FIGURE 2 Right forewing of *Acrotrichis* sp. (Ptilidae: Coleoptera). Scale line, 0.5 mm.

Mechanical

LOCOMOTION Elongate setae are used by numerous insect groups for flight and swimming. Elongation of marginal setae is an efficient mechanism to increase the surface area of locomotory structures (e.g., wing, leg), with only a slight increase in weight or developmental investment (Fig. 2). Such elongation of setae along wing margins is common only in very small sized insects, as the degree of elongation is finite and insufficient to generate adequate lift for larger insects. Groups of insects that use this mechanism for flight include all Thysanoptera and various microhymenopterans (e.g., *Mymar*—Mymaridae), microlepidoptera (e.g., *Macarostola*—Gracillaridae), and Coleoptera (e.g., *Acrotrichis*—Ptilidae) (Fig. 2). Aquatic insects that are active swimmers often have paddle-shaped leg segments fringed with elongated setae for added propulsion through water. Examples include various families of Heteroptera (e.g., Naucoridae, Corixidae, and Notonectidae) and Coleoptera (e.g., Hygrobiidae, Haliplidae, Dytiscidae, and Gyrinidae).

PLASTRON To extend their duration under water, some aquatic insects use a plastron, or air bubble trapped by fine hydrofuge setae. The setae are used to hold a large air bubble by surface tension. The plastron enables them to live permanently submerged in water because oxygen passes by diffusion from the water into the plastron. Examples of plastron use by this method include the beetle families Elmidae and Hydrophilidae.

FOOD GATHERING AND PREY CAPTURE In predatory insects, enlarged setae are an economical substitute for teeth or spines for holding prey. Stiff enlarged setae are present along the inner margin of mandibles of some larval Nymphidae and Myrmeleontidae (Neuroptera). Similar setae on the inner margin of fore femora of adult Leptopodidae (Heteroptera) are used for the same purpose. Aquatic insects may use

elongated rows of setae around the mouth parts as filters to trap food particles in flowing water (e.g., *Coloburiscoides* sp.—Coloburiscidae: Ephemeroptera) or use setae-fringed foretarsi to sieve for prey in detrital ooze (e.g., *Agraptocorixa*—Corixidae: Hemiptera). Bees (Hymenoptera) commonly have enlarged hind tibiae and basitarsi covered with brush-like setae (scopa), which are used to carry pollen.

WING COUPLING Some insect groups use specialized setae to couple the fore- and hind wings together, so ensuring synchronous wing beats during flight. The basic pattern, which occurs in Choristidae (Mecoptera), is composed of retinacular setae along the jugal margin of the forewing interlocking with frenular setae along the basicostal margin of the hind wing. More advanced forms are found in Lepidoptera, in which the frenulum may be a large single seta or multiple slender setae. In Trichoptera, a row of large setae is present along the costal margin or the subcostal vein of the hind wing, which engages either the jugal lobe or a ventral ridge in the anal field.

DIGGING Fossorial insects often have stout setae arranged in rows on the leg segments for digging. A comb of long setae (pecten) is commonly present on the foretarsi of ground-nesting aculeate wasps (e.g., *Bembix*—Sphecidae). Enlarged and thickened setae-like structures borne on acanthophorites are used for digging during oviposition by females of various asiloid Diptera (e.g., Therevidae, Apioceridae).

Camouflage Aids

Highly modified setae are commonly used by insect larvae that camouflage themselves by carrying soil, feces, and/or trash particles on their body. Some of the most elaborate of these may be found on the bodies of myrmeleontoid Neuroptera (i.e., Ascalaphidae, Nemopteridae, Psychopsidae, Nymphidae, and Myrmeleontidae). Other than entangling camouflage materials, the biological function of these highly modified setae, called dolichasters, is unknown. Dolichasters may be simple, scale shaped or highly ornate, star or cup shaped or recurved hooks (Fig. 3). Ascalaphid and nymphid lacewing larvae may also have abdominal extensions called scoli (Fig. 4), on which rows of enlarged scales or setae are used to entangle camouflaging materials.

Defensive

Lepidopteran larvae of several families (e.g., Arctiidae, Notodontidae, Thaumetopoeidae) use specialized urticating (irritating) setae as a defense against potential predators. Urticating setae are modified setae with a poison cell associated with the trichogen cell; the former discharges venom when the tip of the seta is broken off. The urticating setae may be very long and scattered over the body surface or short and positioned on the apex of a scoli. In Limacodidae, the stinging

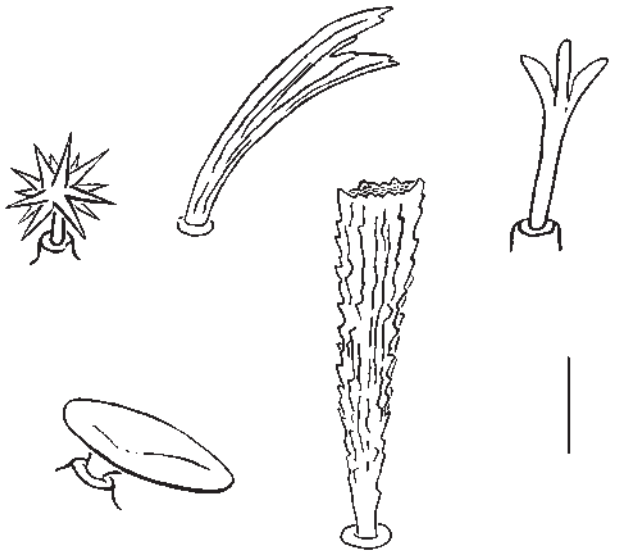

FIGURE 3 Dolichasters from head and thorax of unidentified species of Ascalaphidae (Neuroptera). Scale line, 0.2 mm.

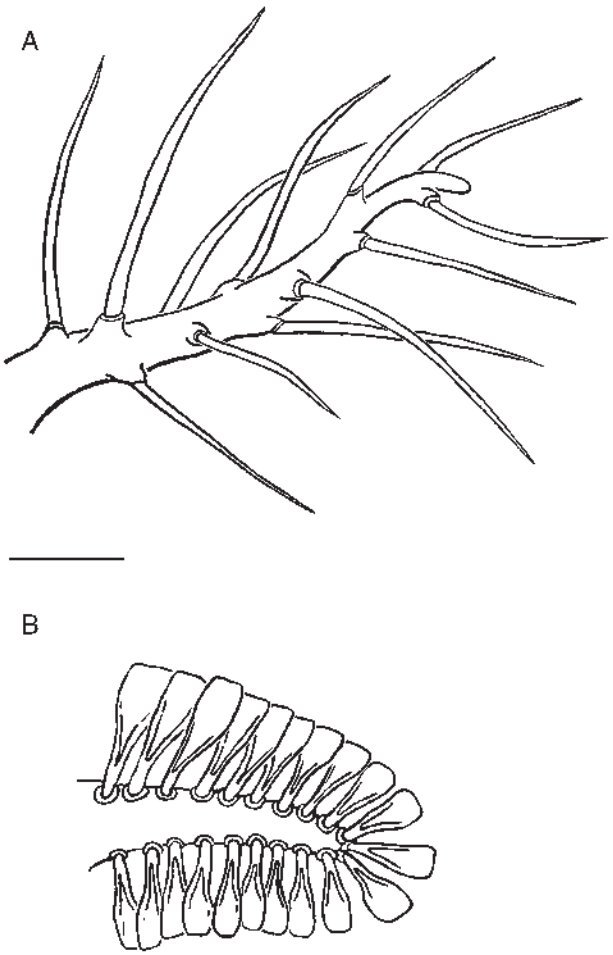

FIGURE 4 Lateral scoli on abdomen of neuropteran larvae. (A) *Osmylops* sp. (Nymphidae), (B) unidentified ascalaphid. Scale line, 1.0 mm.

setae are short and positioned at the apices of often brightly colored, eversible tufts that are extended when the individual is disturbed. Some adults and/or larvae of the coleopteran families Cleridae, Trogossitidae, and Dermestidae are often clothed in long fine setae, which may be recurved. In addition to a sensory function, these setae may be used defensively to obstruct or entangle potential predators or parasitoids.

Secretory Substrate

Setae are commonly used by insects as a substrate of high surface area from which a pheromone is dispersed by evaporation. Adult males of some nemopterid (Neuroptera) genera use distinctive tufts of fine setae (bulla) on the hind margins of the fore- or hind wings to disperse pheromone. Males of the antlion tribe Acanthoclisinae (Myrmeleontidae) have eversible sacs (hair pencils) covered with fine setae for a similar purpose, and many male butterflies use specialized setae (androconia) located in basal depressions along the forewing veins called androconial organs. During courtship pheromone is dispersed from the hairs, often with elaborate "calling" behavior, which can consist of hovering and diving at the prospective female and fanning pheromone in her direction.

See Also the Following Articles

Antennae • Chemoreception • Coloration • Flight • Mechanoreception

Further Reading

Chapman, R. F. (1998). "The Insects: Structure and Function," 4th ed. Cambridge University Press, Cambridge, UK.

Crouau, Y. (1997). Comparison of crustacean and insect mechanoreceptive setae. *Int. J. Insect Morphol. Embryol.* **26,** 181–190.

CSIRO (ed.) (1991). "The Insects of Australia: A Textbook for Students and Research Workers." Melbourne University Press, Cartlon, Australia.

Ferris, G. F. (1934). Setae. *Can. Entomol.* **66,** 145–150.

Keil, T. A., and Steinbrecht, R. A. (1984). Mechanosensitive and olfactory sensilla of insects. *In* "Insect Ultrastructure" (R. C. King and H. Akai, eds), Vol. 2. Plenum, New York.

Richards, A. G. (1951). "The Integument of Arthropods." University of Minnesota Press, Minneapolis.

Richards, A. G., and Richards, P. A. (1979). The cuticular protuberances of insects. *Int. J. Insect Morphol. Embryol.* **8,** 143–158.

Snodgrass, R. E. (1935). "Principles of Insect Morphology." McGraw–Hill, New York.

Scorpions

Stanley C. Williams
San Francisco State University

Scorpions are among the most recognizable groups of arthropods. Their highly segmented body plan is uniquely subdivided into a leg-bearing prosoma, a broad seven-segmented preabdomen (mesosoma), and a narrow, five-segmented postabdomen (metasoma) that terminates in a bulbous stinging organ, the telson. The fossil record indicates scorpions were the first arthropods to occupy the terrestrial environment. A prominent member of the class Arachnida, scorpions survived periods of mass extinctions and today occupy a prominent position in arthropod communities. This gives them the distinction of being the oldest surviving group of terrestrial arthropods.

EVOLUTION

Scorpions date back to the Silurian, some 400 mya. The oldest taxa, *Palaeophonus* spp., are strikingly similar in structure to modern forms in their extensive, characteristic segmentation, appendages, and general body form. Their body was organized into four regions: a prosoma (cephalothorax); a broad mesosoma ("preabdomen" containing digestive and reproductive organs); a taillike metasoma (slender "postabdomen" composed of five ringlike segments); and a terminal segment, the telson. *Palaeophonus* differed from modern scorpions in not having pretarsal claws on walking legs and in having a blunt ending of the telson. The absorptive membranes of the book lungs might have been eversible, forming a gill-like respiratory structure. This suggests that *Palaeophonus* might have existed in aquatic and land habitats.

The fossil record indicates that scorpions underwent a rapid adaptive radiation that led to structural diversity, especially in the segmentation of the ventral prosoma, the development of pretarsal claws, and the formation of a sharp venom-delivering structure on the telson.

MORPHOLOGICAL CHARACTERISTICS

Scorpions are readily distinguished by their unique morphology (Fig. 1). Their prosoma consists of six body segments that are covered dorsally by an unsegmented carapace. Each of these segments bears a pair of characteristic segmented appendages. The first pair, the chelicerae, are small, chelate or scissorlike, and serve as the mouthparts. The second appendages are large, conspicuous pedipalps that terminate in a strong scissorlike chela used to capture and immobilize prey. The next four pairs of appendages are the walking legs, each of which terminates in a pair of distinctive pretarsal claws, the ungues. Along with a single pair of median ocelli, there are two groups of lateral ocelli located at the anterolateral corners of the carapace. Each group of lateral ocelli consists of zero to four facets, evidently derived from a primitive compound eye. The mesosoma is attached broadly to the prosoma, lacks appendages other than the sensory pair of comblike pectines, and houses the digestive and reproductive organs. Expansion of intersegmental membranes permits meal engorgement during feeding and allows for increase in body volume to accommodate internal development of their embryos. The metasoma is composed of five narrow segments forming a freely articulating, taillike

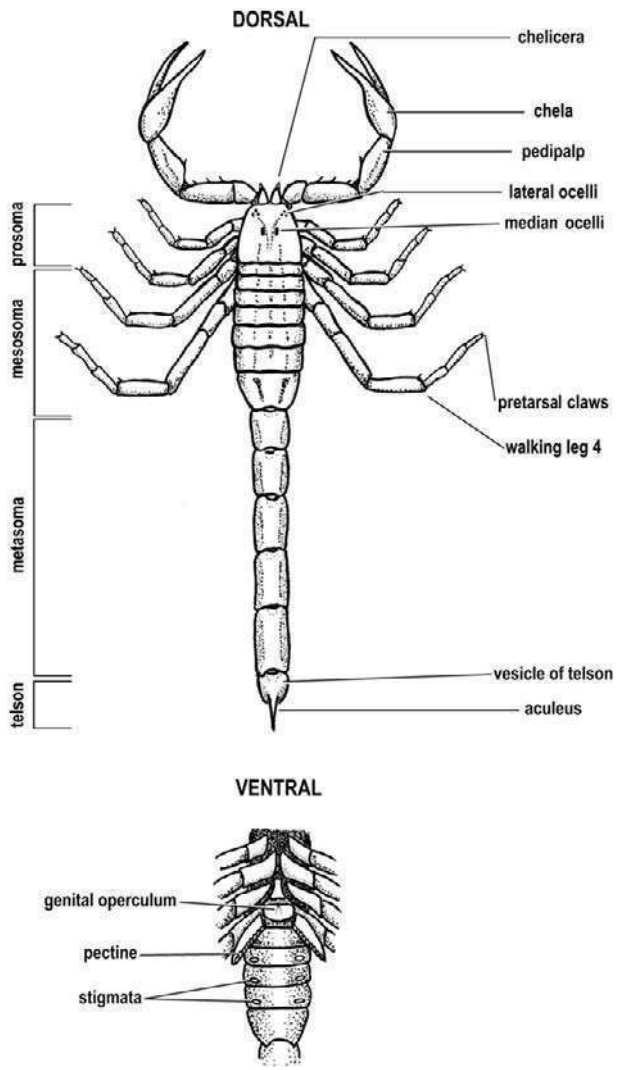

FIGURE 1 Dorsal and partial ventral views of scorpion showing the distinctive body segmentation, tagmosis, and morphological features. (Drawing by K. A. Justus.)

body region. A single unsegmented telson freely articulates with the metasoma and is composed of a swollen vesicle that terminates in a sharply pointed aculeus or "sting." Paired venom glands are housed in the vesicle, and the associated aculeus injects the venom into the victim (Fig. 2). Scorpions all over the world have the same basic body structure, which is surprising considering the antiquity of the extant families.

ANATOMICAL ADAPTATIONS

Scorpions demonstrate basically the same organ systems characteristic of most arachnids. These systems do show adaptations that support their successful lifestyle as opportunistic ambush predators. The digestive system begins with a tiny mouth located below the base of the chelicerae. Its small diameter allows intake of food only in fluid form. Preoral enzymes are secreted by salivary glands directly into the body of the prey, and the liquefied, preorally digested tissues are ingested

FIGURE 2 A giant hairy scorpion *(Hadrurus concolorous)* in alert posture. Notice the pedipalps positioned forward with their chelae open. The metasoma is in flexed position, ready to strike. The body is balanced over and supported by the walking legs.

through a sucking action. This is facilitated by a pressing action of the gnathobases at the inner bases of the pedipalps.

Large, expansive midgut ceca facilitate rapid intake and storage of nutrients. Because particulate matter is not ingested with a meal, fecal wastes are minimal. Nitrogenous wastes are excreted by means of Malpighian tubules that deposit wastes in the form of crystalline guanine directly into the digestive tube for subsequent elimination. Some species have also been found to excrete nitrogenous wastes in the form of uric acid and xanthine. Scorpions have no tracheal system but instead use four pairs of book lungs to respire. Air passes through four pairs of tiny stigmata, which are opened and closed by membranous opercula, in the mesosomal sterna. Inspired air then enters the lamellar spaces of the membranous book lungs. This limiting mode of oxygen uptake is correlated with the sedentary lifestyle and metabolic simplicity characteristic of most scorpions. A ventral nerve cord with highly metameric ganglia characterizes the nervous system. At the head end is found the subesophageal ganglion and supraesophageal brain. The circulatory system consists of a pulsing muscular tube, called the dorsal vessel, and colorless blood called hemolymph.

The reproductive system has organs arranged in a somewhat segmented, fishnetlike form within the mesosoma. The male has a paraxial organ that contains a pair of hemispermatophores that fuse to form a sperm-carrying spermatophore used for indirect insemination of a female during the courtship dance. The female system has an inner lumen that exits the body through the ventrally located genital aperture, which is covered by a valvelike genital operculum on the mesosoma. Multiple oocytes in different stages of maturation may be seen on the surface of the reproductive tube. Fertilization is internal, and the female carries the developing embryos until birthing as larvae.

SYSTEMATICS

Substantial numbers of undescribed species are still being found. This largely results from the nocturnal and secretive habits of

scorpions, which make them difficult to find. Approximately 1286 species of living scorpions are currently recognized and are assigned to 17 families and 157 genera (Table I). The higher classification of scorpions has been the subject of periodic reevaluation over the past two decades. The trend has been to recognize more families and genera and to base the resulting classification on phylogenetic relationships as suggested by cladistic analyses. The numbers of extant scorpion families and species are surprisingly low considering their success as a group, their broad distribution, their diverse habitats, and their antiquity.

DISTRIBUTION AND BIODIVERSITY

Scorpions are conspicuously common in tropical and subtropical regions throughout the world. They also range into the more temperate regions of both the Northern and Southern Hemispheres. In North America they range as far north as British Columbia, Alberta, and Saskatchewan, Canada (approximately 52° N). In South America, scorpions range south into Chile and Argentina (approximately 50° S).

Scorpions show wide-ranging adaptations to different elevations. In Baja California, Mexico, scorpions are found in the intertidal and beach habitats. In the White Mountains of California they are well established in elevations up to 2130 m. Scorpion species show definite habitat preferences and are normally found in patchy distributions. Although they are often thought of as tropical, scorpions seem to reach their highest biotic diversity in the arid regions of more temperate latitudes. Most scorpion habitats are characterized by a modest diversity ranging from one to five sympatric species. The greatest regional diversity of scorpions is reported from Baja California, where 61 species and 11 genera are found. The higher latitudes and higher elevations have more limited diversity, often with only a single species represented in a habitat.

BIOLOGY

General Life Cycle Strategy

Scorpions often are not noticed, even when abundant. They are mainly nocturnal, secretive animals and usually remain inactive, hidden in their retreats, except when feeding, mating, or disturbed. For many scorpions, such as *Hadrurus* and *Vaejovis,* adults are the stage usually encountered; juveniles tend to minimize their exposure in the environment. This pattern is most conspicuous in scorpions found in more temperate areas and at higher elevations. In other species, such as *Centruroides* spp., and some *Paruroctonus* spp. and *Smeringerus* spp., however, juvenile instars are commonly encountered.

Life Cycle

The scorpion life cycle is a simple one: the first instar larva, five nymphal instars, and the adult. The larva differs morphologically from other instars in lacking the typical unguicular claws at the tips of the walking legs, lack of a sharp aculeus or "sting," lack of dark pigmentation and sclerotization of the exoskeleton, and lack of effective locomotor ability. Behaviorally, the larva is unique in that it remains on the mother's back and does not travel alone or feed. The sedentary larva molts to the second instar nymph on the mother's back at which time it

TABLE I Higher Classification of Contemporary Scorpions

Family	Number of genera	Number of species	Common genera
Bothriuridae	12	90	*Bothriurus, Brachistostermis, Centromachetes, Cercophonius, Lisposoma, Orobothriurus, Urophonius*
Buthidae	72	528	*Ananteris, Androctonus, Babycurus, Buthacus, Buthus, Centruroides, Compsobuthus, Hottentotta, Isometrus, Leiurus, Lychas, Mesobuthus, Orthochirus, Parabuthus, Rhopalurus, Tityus, Uroplectes*
Chactidae	11	132	*Broteochactas, Brotheas, Chactas, Nullibrotheas, Teuthraustes*
Chaerilidae	1	21	*Chaerilus*
Diplocentridae	8	76	*Didymocentrus, Diplocentrus, Heteronebo, Nebo, Oiclus*
Euscorpiidae	4	14	*Euscorpius, Megacormus, Plesiochactas, Troglocormus*
Hemiscorpiidae	2	7	*Habibiella, Hemiscorpius*
Heteroscorpionidae	1	2	*Heteroscorpion*
Ischnuridae	8	56	*Cheloctonus, Hadogenes, Iomachus, Liocheles, Opisthacanthus*
Iuridae	6	21	*Anuroctonus, Carboctonus, Hadruroides, Hadrurus, Iurus*
Microcharmidae	2	6	*Akentrobuthus, Microcharmus*
Scorpionidae	7	130	*Heterometrus, Opistophthalmus, Pandinus, Scorpio*
Scorpiopidae	6	27	*Alloscorpiops, Euscorpiops, Neoscorpiops, Scorpiops*
Superstitioniidae	4	9	*Alacran, Sotanochactas, Superstitionia, Typhlochactas*
Troglotayosicidae	2	2	*Belisarius, Troglotayosicus*
Urodacidae	1	19	*Urodacus*
Vaejovidae	10	146	*Paruroctonus, Pseudouroctonus, Serradigitus, Uroctonus, Vaejovis*

shows increase in linear dimensions, has pretarsal claws and a sharp aculeus, and becomes increasingly physically active. At this time the second instar normally leaves the mother's back, disperses from the mother's shelter, locates a new shelter, and begins its own independent life. After periods of feeding and growth, nymphs periodically molt to their successive instars, until adulthood is reached. Adults differ from the nymphs in being sexually mature, but they often show little morphological difference except in their larger size and sometimes in the proportion of body parts such as the larger pedipalp chela and elongation of the metasoma. Adults may also show distinctive sexual dimorphism in body proportions.

Feeding

Ambush predation is the main means of prey capture. For example, some burrowing species such as *Anuroctonus phaiodactylus* quietly wait inside their burrow entrance and ambush prey that enters or passes by their burrow. Many species of scorpions, when hungry, leave their protective shelter during nocturnal hours and take a stationary position in their environment, which may be on the substrate surface or in vegetation. Here they will remain motionless until an unsuspecting prey ventures close, at which time it is grasped with the strong pedipalp chela and stung, if necessary. Other scorpions such as *Centruroides* spp. often range great distances from their diurnal shelters in search for prey. At such times they may actively stalk their prey. If a scorpion is unsuccessful it will return to its shelter and resume its predation behavior the next night, and the night after, until successful.

After successful prey capture, a scorpion draws the immobilized prey close to its oral cavity and exudes preoral digestive fluids that digest the prey's tissues internally. During ingestion the chelicerae also shred prey tissues. The resultant fluid is sucked into the oral region, and any remaining fluids are pressed from the prey remains by the gnathobases of the inner pedipalp coxae. Up to 97% of the body mass of a prey may be ingested. A satiated scorpion then returns to its shelter and will normally not be seen until hungry again, which may be from 2 weeks to a month or two. Its effective mode of predation, habit of engorgement feeding, and low metabolic rate result in a lifestyle that requires minimal prey, minimal exposure of the scorpion, and survival when prey are not readily available. Scorpions in the laboratory may go for as long as 12 months between meals. However, in nature, one feeding every 2 to 6 weeks is probably more common. Almost any animal that a scorpion can catch and immobilize is a potential prey, so scorpions feed on a variety of insects and spiders. Predation on centipedes, other scorpions, lizards, and small snakes has also been observed.

Reproduction

The scorpions' reproductive mode is one of the most ancient ones used by terrestrial arthropods. It involves sexual reproduction, with insemination of the female by the male following a ritualistic courtship dance. Internal fertilization is complicated by the lack of an intromittent organ in the male. Fertilization takes place indirectly by a spermatophore deposited by the male.

Pheromones bring potential mates together and initiate courtship. When two potential mates encounter each other, they grasp each other by their pedipalpal chelae and undergo a forward and backward "dancelike" behavior, the *"promenade à deux,"* which may be over in few minutes or last for hours. At the end of this dance, the male emits a sclerotized spermatophore onto the substrate and draws the female over it, at which time she takes up the sperm packet in her gonopore. The two then part and the female normally returns to her shelter. Depending on the taxa, gestation takes from approximately 2 or 3 months to nearly a year. Development is internal and results in the birth of a dozen to over a hundred larvae, which quickly ascend to the mother's back and remain there for the remainder of their first stadium. Parturition takes from an hour to about 3 days, depending on the species and number of offspring. Some species seem to have several birth cycles each year, whereas others seem to have only one per year. Parthenogenesis has been reported in *Tityus serrulatus* and *Liocheles australasiae.* Field observations of *Liocheles* suggest that some natural populations might have parthenogenesis as their primary means of reproduction.

Habitats

Scorpions have been roughly categorized as "ground" versus "bark" dwellers. Ground dwellers construct burrows in the ground, seek shelter in rocky substrate, invade the burrows of other organisms, or occupy protective spaces under or within ground surface debris. Bark dwellers are commonly found sheltered under bark, in bromeliads, in vegetation, in residential thatch, or in other plant material. They are often climbers and may primarily occupy forest canopies. During some seasons, forest-dwelling forms may migrate to the ground to seek shelter under rocks or surface debris. This group has members that invade human habitations and can be seen climbing walls and ceilings at night. They may reside in thatching and in the mortar spaces between bricks and rocks of walls. Most ground scorpions are solitary, although some bark scorpions such as *Centruroides* may be found in aggregations during certain times of the year. *Anuroctonus phaiodactylus* constructs a burrow as a young nymph and occupies it throughout its life, which may last for several years. It seldom ventures far from its burrow and is intolerant of other scorpions in its burrow. Many of the ground-dwelling species probably spend most of their life in their permanent burrow. At time of courtship, however, mature males abandon their burrows or other territory and become more nomadic in search of mates. Scorpions are thigmotactic, seeking closely fitting shelters. The availability of suitable shelters is believed to limit the population size of some species.

Scorpions are found in a variety of habitats from desert to mesic, and they thrive in highly xeric environments. They are common in tropical and subtropical regions, and many species

extend into the temperate latitudes. Most scorpions prefer well-drained habitats, but some such as *Centruroides* species thrive in riparian and tropical rain forest environments, and seem to survive drowning by climbing trees during times of flooding. Burrowers vary greatly in their choice of habitats and behavior: *A. phaiodactylus* requires soil with good drainage and adequate compaction to support permanent burrows; *Vejovoidus longiunguis* constructs burrows in unconsolidated soil of sand dunes; species of *Serradigitus* do not burrow but live in crevices, talus slopes, and rock fractures; *Superstitionia donensis* may occupy a burrow or may be found in a simple cell constructed under a rock, animal dung, or vegetation; *Centruroides thorellii* resides in bromeliads within tropical forests; *Vaejovis littoralis* is found in the intertidal wrack zone, where it continually migrates ahead of the changing tides; *Centruroides exilicauda* seeks shelter under bark of trees, but may also be found under surface debris or in rock crevices, and is a common invader of human habitations; and *Uroctonus mordax* is commonly found in wet habitats, where it may reside under mosses and rotting logs, and in soil crevices. Almost any habitat that provides adequate shelter and prey is suitable for some kind of scorpion. As a result, these arthropods are globally widely distributed.

Predators and Parasites

Although scorpions do not have many natural enemies, they are an attractive prey for a few predators. Their rich nutrient content and numerical abundance contribute to their vulnerability as prey. Predators that hunt during the nocturnal hours are particularly effective in feeding on scorpions. Most capture scorpions while they are exposed on the substrate surface, but some, such as coyotes, will dig them up. Some 124 vertebrate and 26 invertebrate predators have been reported to feed on scorpions. Elf owls *(Micrathene),* burrowing owls *(Speotyto),* barn owls *(Tyto),* grasshopper mice *(Onychomys),* coyotes, bats, desert shrews *(Noteosorex),* and a variety of lizards are among the most conspicuous scorpion predators. Parasitism of scorpions is not common; reported endoparasites include larvae of a tachinid fly *(Spilochaetosoma californicum),* mermithid nematode larvae, and the larvae of a sarcophagid fly *(Sarcodexia sternodontis).* About eight species of mites have been reported as ectoparasites of scorpions, but mite parasitism is not commonly observed.

VENOMS

Scorpions are universally recognized because of their venoms and conspicuous stinging apparatus. All species are venomous, but only a few are harmful to humans. The venom is produced in a pair of glands located in the vesicle of the telson and collects in the lumen of each gland, flowing through a simple duct that terminates near the sharp tip of the aculeus. The venom is delivered during a rapid thrust of the metasoma that results in the penetration of the victim's skin by the sharp aculeus. Contraction of muscles surrounding each venom gland discharges the venom.

The venoms are complex substances composed of water, a number of low molecular weight proteins (neurotoxins), and various organic compounds such as oligopeptides, nucleotides, amino acids, mucus, and cellular debris. A common fraction of venoms function as simple irritants, often causing a sharp burning sensation. Other fractions may cause inflammatory responses resulting in edema. In highly toxic venoms, systemic components may result in neurological symptoms such as convulsion followed by death. The neurotoxic effects seem to be the result of multiple interactions of certain toxins with voltage-dependent ion channels of excitable cell membranes, such as those of neurons. Venom interactions may include membrane depolarization, repetitive firing, prolonged action potentials, and massive release of neurotransmitters, most importantly from the adrenal medulla.

The cumulative effects of a venom in envenomated animals are complex and varied. Envenomation particularly may affect skeletal muscles, the cardiovascular system, lungs, visceral smooth muscle, uterus, and glands. The severe convulsions that may be associated with the sting of more toxic species (e.g., *Centruroides* spp.) are of particular concern. The venom of each species has a unique composition that explains the many varied venom reactions observed. Curiously, each scorpion venom contains a number of different toxic components.

Scorpion venoms are toxic to a variety of organisms, including arthropods and vertebrates, but are often nontoxic to other scorpions. The hemolymph may have the capacity for neutralizing the toxic components in the venom from other scorpions.

The number of reported human deaths from scorpion stings ranges from several hundred to several thousand annually. Stings are considered to be more serious when they happen to children, the elderly, and those in poor health. The number of reported deaths has been declining because of better treatments, development of antivenins, the reduction of human contact with scorpions in critical areas, and public health awareness. Because of unreliable reporting in developing countries, however, the reported incidence of scorpion stings and mortality is probably greatly underestimated.

Venomous scorpions of concern to humans are found throughout the world, but some of the more severe problems are in the more arid regions. Of the 1286 known species, only a dozen or so cause significant health problems for humans. The species of greatest medical concern are members of the family Buthidae and belong to the genera *Androctonus, Buthacus, Buthus, Centruroides, Leiurus, Mesobuthus, Parabuthus,* and *Tityus.*

METHODS OF OBSERVATION AND STUDY

Sampling and Collecting

Scorpion assessments and collections were traditionally made by inspecting trees and by looking under rocks, trash piles, and other surface debris. The discovery that scorpions fluoresce

under ultraviolet light [with black light blue (BLB) bulb] led to the use of the "ultraviolet method" of scorpion detection and study. Equipped with a portable ultraviolet light, an investigator can walk through a habitat at night, counting and observing scorpions undisturbed in their natural state. With such a lamp a scorpion can be readily detected on a sand dune at a distance of about 15 m.

Care and Maintenance of Captive Specimens

Scorpions are easily kept in captivity and require minimal care. They may be housed in any closed container, such as a terrarium, plastic box, jar, or plastic bag. If soil is provided, a rock or other surface cover should be added for shelter because scorpions are thigmotactic and seek close-fitting shelters. A small amount of water needs to be provided periodically. Scorpions can be maintained at room temperature and should not be exposed to freezing temperatures, especially the tropical forms. Most live insects may be supplied for food, but crickets are particularly well accepted. Generally one cricket every 2 to 4 weeks is adequate feeding for a moderate-sized scorpion.

As a rule, scorpions remain inactive in their shelters during the day and will show normal activity (if any) at night. Lack of physical activity is common and typical of most scorpions. Excessive activity often indicates a lack of food or water, excess light, excessive temperature, or other disturbance. Covering the rearing chamber with red cellophane readily simulates nocturnal environmental conditions. Most scorpions will live for 1 to 5 years in captivity.

See Also the Following Articles

Arthropoda and Related Groups • Spiders • Venom

Further Reading

Baerg, W. J. (1961). Scorpions: biology and effect of their venom. Agricultural Experimental Station Bulletin **649.** University of Arkansas, Fayetteville.

Brownell, P., and Polis, G. (eds.). (2001). "Scorpion Biology and Research." Oxford University Press, Oxford, U.K.

Fet, V., Sissom, W. D., Lowe, G., and Braunwalder, M. E. (2000). "Catalog of the Scorpions of the World (1758–1998)." New York Entomological Society, New York.

Keegan, H. L. (1980). "Scorpions of medical importance." University Press of Mississippi, Jackson.

Polis, G. A. (ed.). (1990). "The Biology of Scorpions." Stanford University Press, Stanford, CA.

Prendini, L. (2000). Phylogeny and classification of the superfamily Scorpionoidea Latreille 1802 (Chelicerata, Scorpiones): an exemplar approach. *Cladistics* **16,** 1–78.

Rubio, M. (2000). "Scorpions: A Complete Pet Owner's Manual." Barron's Educational Series, Hauppauge, NY.

Stoops, E. D., and Martin, J. L. (1995). "Scorpions and Venomous Insects of the Southwest." Golden West Publishers, Phoenix, AZ.

Vosjoli, P. de. (1991). "Arachnomania: The General Care and Maintenance of Tarantulas and Scorpions." Advanced Vivarium Systems, Lakeside, CA.

Williams, S. C. (1980). Scorpions of Baja California, Mexico, and adjacent islands. *Occas. Pap. Calif. Acad. Sci.* **135,** 1–127.

Williams, S. C. (1987). Scorpion bionomics. *Annu. Rev. Entomol.* **32,** 275–295.

Segmentation

Nipam H. Patel
University of Chicago

Segmentation (the repetition of body units along the anterior–posterior axis) is a fundamental property of all insects; indeed, it is an obvious character of all arthropods. Insect segments are clearly visible as reiterated patterns visible in the exoskeleton, but repeating patterns are present in internal structures such as muscles, neurons, and tracheae, as well. Through genetic and molecular approaches in the dipteran fruit fly, *Drosophila melanogaster,* the mechanisms of segmentation in this insect are now understood in great detail. Additional experiments indicate that some aspects of the *Drosophila* mechanisms are conserved in all insects, and others have undergone extensive evolutionary changes.

PATTERN OF SEGMENTS

In virtually all insect embryos, larvae (where present), and adults, the pattern of segmentation in the thorax and abdomen is clearly visible. Segments are usually separated by grooves, called segmental grooves, that lie at the boundary between each pair of segments. The configuration of segments is also often characterized by repeating patterns of pigmentation and elaborations, such as denticles or hairs on the exoskeleton. Internally, segmentation is reflected in repeating patterns within the nervous system, musculature, and tracheal system. Segments of the head and the terminal abdominal regions are sometimes more difficult to recognize. It appears that all insects are composed of six head segments, called the antennal, ocular, intercalary, mandibular, maxillary, and labial segments (going progressively from anterior to posterior), although some authors have suggested the existence of a seventh segment at the anterior of this pattern. The thorax is always composed of three segments (T1–T3), and the abdomen is generally composed of 11 segments (A1–A11) (Fig. 1), the most posterior abdominal segments (A10 and A11) are often fused during later development but usually can be detected separately during embryogenesis.

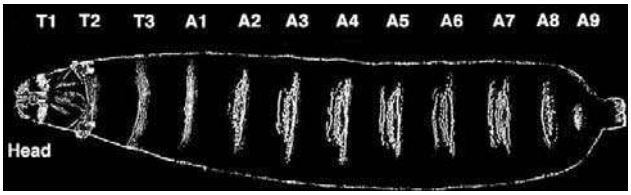

FIGURE 1 The larval segments of *Drosophila.* The pattern of segments is clearly revealed by the pattern of denticles (hairlike projections) on the ventral surface of the larvae. The segments of the head are involuted inside, and the final abdominal segments are not visible on the surface.

MECHANISMS OF SEGMENTATION IN *DROSOPHILA*

Studies of insect segmentation through experimental manipulations, such as ligation, ablation, centrifugation, and transplantation, have a rich history and continue to provide important insights into the mechanisms of segmentation. Beginning about 30 years ago, however, understanding of these mechanisms was rapidly accelerated by the genetic mutant screens in the fruit fly. Indeed, these screens uncovered such fundamental principles of biological pattern formation that the geneticists who carried them out, Edward Lewis, Eric Wieschaus, and Christiane Nüsslein-Volhard, were awarded the Nobel Prize in medicine and physiology for their contributions to the genetic analysis of tagmosis (regionalization) and segmentation in *Drosophila.*

The genetic analysis of segmentation was quickly supplemented with molecular and biochemical studies that have provided detailed knowledge of how segments are generated along the anterior–posterior axis during *Drosophila* embryogenesis. As it turns out, the mutants that effect this process can be grouped into several specific categories that act in a hierarchical manner to sequentially subdivide the embryo into smaller and smaller units, ultimately establishing the pattern of larval segments we see in *Drosophila* (Fig. 1). The mutant classes include maternal effect mutations, which disrupt the anterior or posterior halves of the embryo; gap mutations, which eliminate several contiguous segments; pair-rule mutations, which delete regions in a two-segment periodicity; and segment polarity mutations, which cause deletions and duplications in every segment.

Studies of the maternal effect mutations show that the process of segmentation actually begins during oogenesis, when the female localizes specific messenger RNAs (mRNAs) at either the posterior or anterior end of the developing egg. For example, bicoid mRNA is localized to the anterior end of the egg and forms a gradient of protein in the egg once it has been fertilized (with the highest concentration of bicoid protein at the anterior end). Mothers lacking functional bicoid gene form embryos in which the anterior segments are missing. A reciprocal gradient of the nanos protein is also formed, and *nanos* mutants are missing the more posterior regions of their body. The formation of these gradients is possible because the early development of *Drosophila* is syncytial, with no cell membranes between the nuclei of the early embryo. These gradients of information act to control the expression of the various zygotic gap genes, which come on in individual, well-defined broad regions along the anterior–posterior axis of the embryo.

The interaction of the gap genes (which are all transcription factors) generates the first periodic patterns in the embryo. These periodic patterns are stripes of the pair-rule genes. In *Drosophila,* most of the pair-rule genes display a pattern of seven stripes in the so-called cellular blastoderm stage of the embryo (at about 2.5 h after fertilization, when cell membranes form between the nuclei transforming the embryo from a syncytial to cellular blastoderm). The pair-rule genes are again generally transcription factors that regulate the final level of the hierarchy, the segment polarity genes. These genes are expressed in patterns of segmental stripes and include not just transcription factors, but also various receptors, ligands, and enzymes that are used in cell–cell communication and to refine and maintain the pattern of segments that has been elaborated.

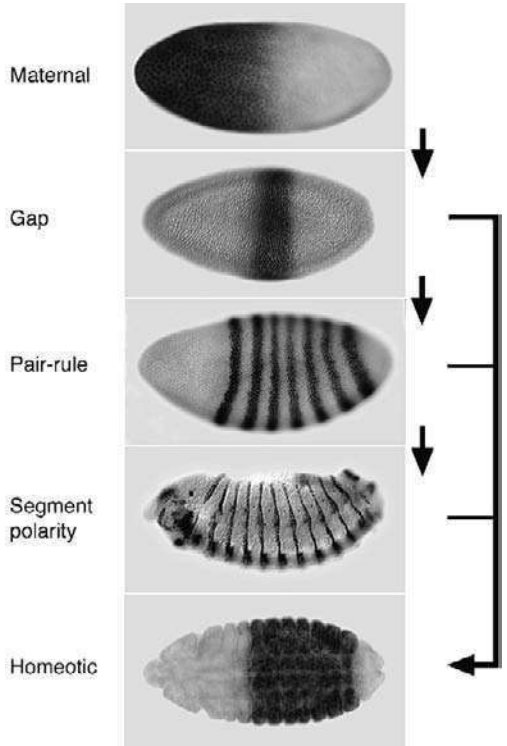

FIGURE 2 The segmentation hierarchy in *Drosophila.* The hierarchy is composed of the sequential expression of maternal, gap, pair-rule, and segment polarity genes. An example of the expression pattern of a single member of each class is shown here. The homeotic genes act to give regionalization to the segments and are primarily controlled by the gap genes, with some input from the pair-rule and segment polarity genes.

A final category of genes, the homeotic genes, give identity to the segments. Mutations in homeotic genes result in the transformation of one or more segments into the identity of adjacent segments, and the homeotic genes are primarily regulated directly by the gap genes, although pair-rule and segment polarity genes also have some control on the precise boundaries of homeotic gene expression. All the homeotic genes encode a family of closely related transcription factors and are organized into two complexes on one of the chromosomes of *Drosophila.* Figure 2 shows the expression pattern of representatives of each of the foregoing classes of genes.

RELATIONSHIP OF *DROSOPHILA* SEGMENTATION TO SEGMENTATION IN OTHER INSECTS

Although the genetic analysis of segmentation in *Drosophila* provided an invaluable insight into the mechanisms of pattern formation, earlier manipulative studies in a variety of insects

suggested that some aspects of segmentation differ among the various insects and that indeed *Drosophila* might be somewhat unusual in its mechanisms of segmentation. *Drosophila* is classified as a long-germ insect because various manipulative experiments showed that pattern formation was achieved very rapidly across the entire length of the embryo all at once and without the need for growth. At the molecular level, this is reflected in the nearly simultaneous appearance of pair-rule gene stripes in the *Drosophila* blastoderm. Many other insects, however, display a short or intermediate germ mode of development. In these insect embryos, only the most anterior segments are present in the blastoderm (prior to gastrulation), and more posterior segments are added only as the embryo elongates at later stages. Insects such as the red flour beetle, *Tribolium castaneum,* and the grasshopper *Schistocerca americana* display this type of development.

This progressive addition of segments is supported by comparative studies of gene expression. In both *Tribolium* and *Schistocerca,* the homologues of *Drosophila* segment polarity genes are expressed in stripes as in *Drosophila,* but the stripes appear sequentially as the embryos grow. In *Tribolium,* all the homologues of *Drosophila* pair-rule genes studied so far are expressed in the same pattern as in *Drosophila* (and prior to the expression of segment polarity genes), but again the stripes appear sequentially over time as the embryo elongates. In the more evolutionarily distant embryos of *Schistocerca,* however, the expression pattern of pair-rule gene homologues differs; indeed, some of them are not even expressed in a pattern of stripes. This suggests that extensive evolutionary alterations have occurred at this step of the segmentation hierarchy, although these changes still result in a conserved output of segment polarity gene expression.

The earliest steps of pattern formation are probably even more labile during insect evolution. This is not surprising, given how variable early embryogenesis can be in insects. For example, it is difficult to imagine how a gradient of bicoid protein can form in grasshopper embryos, given that the entire thorax and abdomen arises as a result of cell proliferation well after the blastoderm stage. Recent studies suggest that the bicoid gene, a key component of the maternal gradient *Drosophila* system, evolved somewhere within the dipteran lineage. Apparently extensive modifications have occurred in the segmentation system in different insect lineages, and these changes may reflect adaptive changes in the speed and patterns of oogenesis and early embryogenesis in different insect groups. Nevertheless, the overall logic of the *Drosophila* segmentation hierarchy has been conserved, not just in all insects, but possibly in all arthropods.

RELATIONSHIP TO SEGMENTATION IN OTHER ANIMAL PHYLA

Remarkably, many of the genes involved in *Drosophila* segmentation and regionalization are well conserved throughout animal evolution. In particular, the homeotic genes, which control segment identity, are conserved in both structure and function between flies and vertebrates. Some changes in the expression patterns of these homeotic genes within insects, however, seem to be responsible for some evolutionary alterations in segment morphology within this group. Homologues of *Drosophila* segment polarity and pair-rule genes are also well conserved, and usually these proteins still play similar biochemical roles, but in different developmental contexts. For example, the segment polarity gene *hedgehog* is used in many pattern formation steps in vertebrates, such as patterning the dorsal–ventral axis of the neural tube, but it has no known function in vertebrate segmentation. Recently, however, it has been found that the mouse and chicken homologue of a *Drosophila* pair-rule gene, *hairy,* is involved in vertebrate segmentation. Thus, there is still considerable debate about the evolutionary origins of segmentation in arthropods, annelids, and vertebrates, with some believing that segmentation is homologous between these groups and other believing that segmentation has evolved independently in these different animal lineages.

See Also the Following Articles

Drosophila melanogaster • *Embryogenesis*

Further Reading

Akam, M. (1987). The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101,** 1–22.

Brown, S., Fellers, J., Shippy, T., Denell, R., Stauber, M., and Schmidt-Ott, U. (2001). A strategy for mapping bicoid on the phylogenetic tree. *Curr. Biol.* **11,** R43–R44.

Davis, G. K., and Patel, N. H. (1999). The origin and evolution of segmentation. *Trends Genet.* **15,** M68–M72.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila. Nature* **287,** 795–801.

Palmeirim, I., Henrique, D., Ish-Horowicz, D., and Pourquie, O. (1997). Avian *hairy* gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell,* **91,** 639–648.

Patel, N. H. (1994). Developmental evolution: Insights from studies of insect segmentation. *Science* **266,** 581–590.

Sander, K. (1976). Specification of the basic body pattern in insect embryogenesis. *Adv. Insect Physiol.* **12,** 125–238.

Sericulture

Satoshi Takeda

National Institute of Agrobiological Sciences, Ibaraki, Japan

Sericulture is an industry that is characterized by a two-step process, the cultivation of mulberry trees and the rearing of silkworms on mulberry leaves to produce cocoons. A cocoon is an oval- to football-shaped object made by a mature silkworm larva by spinning silk proteins; the silkworm larva develops into a pupa inside it. Silkworms are monophagous insects, feeding only on mulberry leaves (Moraceae, genus *Morus*). Because the mulberry leaves must be fresh, it is difficult to transport them over long distances or store them for

long periods. This has resulted in the rearing of silkworms and cultivation of mulberry trees generally forming a single enterprise. Mulberry tree cultivation starts with the production of mulberry seedlings, followed by mulberry tree training, cultivation, harvesting, and insect pest control. Silkworm rearing includes preservation of silkworm eggs, management of rearing rooms, handling of rearing equipment, prevention of silkworm diseases, supplying mulberry leaves, and collecting mature larvae to transfer to the cocooning frame.

ORIGINS OF SERICULTURE

It is impossible to document when sericulture began. The silkworm, *Bombyx mori,* now has no wild populations; it is a completely domesticated insect. The oldest written record of sericulture is the Chinese silkworm book *Can-jing,* which states that the queen of the Huang-Di empire started silkworm rearing. The Huang-Di era was around 2650 B.C., but sericulture must have been carried out in China in even earlier times. From China, sericulture spread via the "Silk Road." In the East, it was introduced into the Korean Peninsula and from there to Japan in about the 3rd century B.C. In the West, it spread to Central Asia and India and from India to Persia. Sericulture is thought to have reached Europe in 550 A.D., when silkworm eggs were presented to the Roman Emperor of the East. Silk was an important trade item along the Silk Road, where it was exchanged for its weight in gold. Commerce in silk along the Silk Road also made a major contribution to the exchange of Eastern and Western culture.

THE WORLD'S SERICULTURE INDUSTRY: PAST AND PRESENT

The state of cocoon production worldwide in 1997 compared to 1930 indicates that total global cocoon production, 617,910 tons in 1930 and 620,000 tons in 1997, was almost the same; the cocoon-producing countries, however, have changed considerably. In 1930, Japan ranked first, with a yield of 382,850 tons, and accounted for 62% of global production. China (then the Republic of China), was second with 129,528 tons, 21% of the world total; Italy ranked third with 53,348 tons (8.6%) and the Soviet Union fourth with 15,300 tons (2.5%). At that time cocoons were produced almost everywhere in the world where the mulberry could be cultivated.

In contrast, in 1997 China had by far the greatest production, 423,000 tons, with 68.2% of global production. India ranked second with 127,000 tons (20.5%), followed by Uzbekistan, Brazil, Thailand, Vietnam, and North Korea. Cocoon production in Japan, which accounted for 62% of world production in 1930, dropped precipitously after World War II and today is 2500 tons, a mere 0.4% of the world total. The dramatic fall in cocoon production in Japan was the result of soaring labor and production costs and low cocoon prices compared with other agricultural products. Although an aging population of sericulture workers and a shortage of replacements were factors, the primary cause of the decline in recent years has been the development of large differences in cocoon prices between Japan and other cocoon-producing countries, such as China and Brazil.

CURRENT STATUS OF SERICULTURE IN DIFFERENT PARTS OF THE WORLD

From the 1980s to the 1990s the sericulture industry became concentrated in Asia. Brazil is the sole country outside Asia in which export-quality cocoons are produced. Although sericulture thrived in Europe during the first half of the 20th century, particularly in Italy and France, only a very small scale production remains in Eastern European countries, such as Bulgaria and Romania. The current status of sericulture in the principal cocoon-producing countries is briefly described below.

In China, the provinces of Szechuan, Jiangsu/Chekiang, and Goangdong are the three great sericulture regions. Cocoons of the wild Chinese oak silkworm, *Antheraea pernyi,* add to the production of silk from *B. mori.* About 50,000 tons a year of wild silkworm cocoons are produced by outdoor rearing in mountains and forests, chiefly in the northeast.

Sericulture suited to each of its regions is carried out nationwide in India, the second largest cocoon-producing country in the world, in which production has increased sharply in recent years. In the north of India, temperate-region sericulture is conducted with bivoltine varieties of silkworm, and in the south it is being carried out with polyvoltine varieties or hybrids between polyvoltine and bivoltine varieties. Many species of wild silkworm, including the Tassar silkworm *(Antheraea mylytta)* and the Muga silkworm *(Antheraea assamensis),* are used for sericulture in the northern regions of India. However, sericulture of wild silkworm species remains a manual industry from rearing to reeling, not like that of *B. mori.*

In Southeast Asia, traditional sericulture industries in Thailand, Vietnam, and Laos use tropical polyvoltine varieties suitable for clothing. In central Asia, Uzbekistan is the major sericulture country ranking third among cocoon-producing countries. There, univoltine silkworm varieties are used, and sericulture is conducted as a sideline industry to cotton growing. The Brazilian silk-reeling industry was started by Italian immigrants, and introduction of technology by Japanese silk-reeling companies has enabled the production of high-quality raw silk thread and silk, and Brazil is now a raw silk thread and silk exporter on par with China. In Japan, cocoon production amounts to less than 1%; however, Japan consumes about 25% of global production, ranking second to China.

MANAGEMENT OF SILKWORM EGGS

Management of silkworm eggs is one of the most important sericultural techniques. In tropical regions where the mulberry leaves are available all year, the management of silkworm eggs is not particularly important. Polyvoltine strains used in the tropics hatch year round and can be reared anytime. However,

because univoltine and bivoltine silkworms are usually reared in temperate and subtropical regions, larval hatch must be coordinated with the season when mulberry leaves are available. Once the univoltine silkworms are reared, the larvae generally do not hatch from the eggs until early spring of the next year. Such eggs are called "hibernating eggs." Artificial hibernation activates larvae to hatch by incubating hibernating eggs at about 10°C. Hydrochloric acid treatment can be used to induce hatching on demand.

SILKWORM REARING

Rearing silkworms consists of a series of tasks that includes harvesting and transporting mulberry leaves, supplying the mulberry leaf, cleaning the rearing beds, mounting the larvae so they can spin cocoons, and collecting and shipping the cocoons. The goal of silkworm rearing is to produce many high-quality cocoons while economizing on labor and material. The temperature range in which silkworms can be reared is 7 to 40°C, but practical rearing occurs in the 20–30°C range. Larval period rearing is roughly divided into the young silkworm period (first to third instar) and the grown silkworm period (fourth to fifth instar), and the fundamentals of rearing of each of them are different.

Young Silkworm Period

The growth rate is very high during the young silkworm period, especially during the first instar, but this period is characterized by higher susceptibility to bacterial pathogens and malnutrition. The young silkworm period is also physiologically characterized by a low quantity of food ingested, but a high rate of digestion. The amount of mulberry leaves ingested by young silkworms, i.e., first to third instar, is only about 2% of the amount ingested during the entire larval period. The mulberry leaves provided in this period are finely minced.

The basic principles of young silkworm rearing are to provide relatively high temperatures and high humidity. A temperature range of 26 to 28°C is ideal, and as the instars proceed, temperature is gradually reduced. The ideal humidity is 75 to 90% and is lowered as the instars progress. In advanced sericulture countries, rearing of the young silkworms is generally carried out in cooperative rearing facilities.

Grown Silkworm Period

During the fourth instar silkworms ingest 10% of the total amount of mulberry leaves taken, and during the fifth instar they ingest approximately 88%. Because the fifth instar is the period when the silk proteins for cocoons are actively biosynthesized, an adequate supply of mulberry leaves must be provided. When silkworms are reared in temperate regions, at high temperatures of 30°C or more, disease or failure to spin silk can occur. When fifth instars are mature they start to spin. Silkworms discharge a great deal of fluid outside their bodies with the silk proteins, in addition to that from defecation and urination, during the cocoon production, or mounting, period.

FIGURE 1 Silkworm larvae (fifth instar) reared on artificial diet.

REARING ON AN ARTIFICIAL DIET

In 1960, an artificial diet for silkworms was developed that could successfully sustain them from the first to the fifth instar (Fig. 1). The growth of the silkworms and the cocoon size obtained, however, were considerably inferior to those of silkworms reared on mulberry leaves and therefore this diet was inadequate for actual use in sericulture. Following many improvements the use of artificial diets during the young silkworm period has become widespread, and in 1990 the rate of young silkworm rearing on artificial diets in Japan exceeded 40%. Japan is is the only country in which artificial diets are used for cooperative young silkworm rearing.

See Also the Following Articles

Bombyx mori • Commercial Products from Insects • Rearing of Insects • Silk Production

Further Reading

Hazama, K. (1995). "Sericulture in the Tropics." Assoc. for Int. Cooperation in Agriculture and Forestry, Tokyo.

Japanese Society of Sericultural Science (ed) (1992). "Fundamental Sericulture." Dainihon Sanshikai, Tokyo. [In Japanese]

Kuribayashi, S. (1998). Production and utilization of silkworms. *In* "Insect Resources in Asia" (Japan Int. Res. Center for Agricultural Science, ed.). Assoc. Agric. and Forestry Services, Tokyo. [In Japanese.]

Tajima, Y. (1978). "The Silkworm: An Important Laboratory Tool." Kodansha, Tokyo.

Sex Determination

Michael F. Antolin and Adam D. Henk
Colorado State University

Sex determination depends upon molecular switches that signal whether the male or the female sex-differentiating pathway will be followed during development; it can be trig-

gered by genetic, epigenetic, or environmental cues. Insects display sexual dimorphism, in which males and female differ in form, behavior, and/or physiology. Although sex determination and developmental pathways leading to two distinct sexes are universal within insects, the primary signals that trigger sex determination are highly diverse and differ between groups. The primary signal for sex determination can entail genetic signaling, epigenetic signaling via maternally expressed genes or genomic imprinting of genes, or cytoplasmic factors like B chromosomes and bacterial infections. The primary signals can act alone or in combination. Molecular genetic details of sex determination and sex differentiation are known mainly from *Drosophila,* and these are described below. Comparison of sex determination of insects, nematode worms, and mammals points to a similar genetic mechanism that underlies sex determination: a cascade of gene expression, with alternative splicing of key genes and intermediate genes, leading to alternative splicing of a double-switch gene that ultimately controls differentiation of males and females. In insects, it appears that key genes and intermediate genes early in the cascade are unique to each group of insects, but genetic pathways for controlling sex differentiation after the double switch appear to be the same in all insects. Overall, sex determination in insects is highly variable among groups, and in some groups the mechanisms of sex determination differ between populations of a single species (e.g., the house fly *Musca* and the midge *Chironomus*).

SEX-DETERMINING SIGNALS

Patterns of sex determination have been explored since chromosomes were first described in the late 1800s. By the early 1900s, it was discovered that many insects have distinct chromosomes in males and females; the work is best described in the comprehensive 1973 book by M. J. D. White, *Animal Cytology and Evolution.* A generality about sex-determining mechanisms is that they are tremendously variable among insect orders, although considerable variation also can be found within genera or even within species. Regardless, some broad categories of the primary signal for sex determination can be identified.

The most common pattern of chromosomal sex determination is for females to be the homogametic sex, with two copies of one sex chromosome (XX), whereas males are the heterogametic sex, with two different sex chromosomes (XY). Some insect groups (e.g., Lepidoptera, Trichoptera) have the opposite pattern, in which males are the homogametic sex (ZZ) and females are heterogametic (ZW). Y chromosomes are usually much smaller than X and do not successfully recombine with the X. Unlike mammals, few sex-determining genes are located on Y chromosomes. Thus, the Y chromosome has been lost entirely in numerous orders of insects, leading to a system in which females are XX but males are haploid for sex chromosomes (denoted XO). A molecular genetic mechanism interacting with chromosomal sex determination has been described in *Drosophila,* and it entails a "molecular counting" mechanism that assesses the ratio of X chromosomes to autosomes early in development. Thus, chromosomal sex determination may generally depend on this type of genic balance, but whether this mechanism is a general feature of insects or is unique to *Drosophila* is unknown.

Sex chromosomes vary greatly in number and size in insects. Translocations and duplications of sex chromosomes are common and fall into two general categories. First, in groups with XO males, small parts of the X chromosomes may be duplicated and form what are called "neo-Y" chromosomes. Second, entire X and Y chromosomes may be duplicated, so that some species have sexes that are homogametic or heterogametic, but contain multiple copies of the sex chromosomes. For instance, in the oriental rat flea, *Xenopsylla cheopis,* males are X_1X_2Y, whereas females are $X_1X_1X_2X_2$, although in other insect groups it is not uncommon for there to be up to five copies of the X chromosome. Multiple sex chromosomes like these would change the balance between sex chromosomes and autosomes and thus bring into doubt the generality of genetic balance as a sex-determining signal.

Many insects have no identifiable sex chromosomes, but still have specific genetic loci encoding genes that act as key sex-determining signals. A common pattern is to have a dominant sex-determining factor that specifies male development for individuals with an *M*/+ genotype, whereas females have a +/+ genotype. In this regard, *M* represents a key gene in a system similar to heterogametic sex determination, but lacking sex chromosomes that differ in size. House flies *(Musca domestica)* demonstrate further variations of the same theme, in which intermediate genes change the sexual phenotype. Another dominant genetic factor, *F,* interacts with rare *M/M* genotypes to produce females with an *M/M F*/+ genotype. Individuals with an *M/M* +/+ genotype develop as males. Finally, some house fly populations have epigenetic sex determination via a maternally expressed gene. Here, a gene found in females determines whether they will produce only female or only male progeny, regardless of the genotype of their male mates.

Haplodiploid sex determination (sometimes called arrhenotoky—the production of males) is found in all Hymenoptera and Thysanoptera and in some Homoptera and Coleoptera. It provides an interesting study of combinations of epigenetic and genetic sex determination. In haplodiploidy, females arise from fertilized eggs and therefore have the diploid complement of chromosomes, whereas males parthenogenetically develop from unfertilized eggs and are haploid. Recent evidence from the wasp *Nasonia* indicates that haplodiploidy can be explained by an epigenetic switch: genomic imprinting of sex-determining factors during oogenesis. Imprinting renders some female-determining genes inactive in eggs during development. If the genomic imprint is removed after fertilization, fertilized eggs develop as females. Unfertilized eggs would still have inactive female-determining genes and would become male. Haplodiploidy represents a case of a key sex-determining signal under epigenetic, rather than purely genetic, control.

Many Hymenoptera have an additional mechanism that overlays the epigenetic control of haplodiploidy: a single-locus genetic sex-determining system called complementary sex determination (CSD). CSD is found in sawflies, and in most apocritan groups (bees, ants, vespid wasps, ichneumonid wasps, braconid wasps), but not in the Parasitica. In this system, sex is determined by an individual's genotype at a single genetic locus, and the locus is known to have a large number of complementary alleles. As in all haplodiploids, males arise from unfertilized eggs and are haploid. If an egg is fertilized, however, it will develop as a female only if egg and sperm carry two different (complementary) alleles of the sex-determining locus. Thus, CSD of females depends upon heterozygosity at the sex locus. If egg and sperm have the same allele, the resulting diploid individual is homozygous at the sex locus and develops as a diploid male. Diploid males in Hymenoptera are usually inviable or sterile, so that inbreeding in species with complementary sex determination has severe consequences. The genetic mechanism underlying complementary sex determination remains to be discovered.

Haplodiploidy (arrhenotoky) differs from the case of paternal genome loss in some Homoptera (scale insects) that have haploid males. Under paternal genome loss, both males and females arise from fertilized eggs. However, eggs that lose the paternal complement of chromosomes early in development become male and are haploid.

Sex determination in insects also can be influenced by factors passed vertically through an egg's cytoplasm. The most common appear to be by infection with the bacterium *Wolbachia,* an intracellular parasite that is widespread in arthropods. In many cases, *Wolbachia* infections selectively kill one sex early in development and skew sex ratios. *Wolbachia* also induce unisexual parthenogenesis (theletoky) in some Hymenoptera. One case of sex determination being modified by *Wolbachia* is known from insects, that of the Asian corn borer *Ostrinia furnacalis.* Supernumerary chromosomes also can be passed via the cytoplasm. In the wasp *Nasonia vitripennis,* a B chromosome called paternal sex ratio *(psr)* causes condensation of chromosomes shortly after fertilization of eggs. Eggs fertilized by *psr* sperm develop as haploid males rather than as females, and subsequently these males continue to transmit the *psr* B chromosome to future generations.

Environmental cues affect sex determination, especially in reptiles and fish, although few cases are currently known from insects. A temperature-sensitive allele of the intermediate sex-determining gene *transformer* (see below) has been described in *Drosophila,* in which high temperature leads to development of males.

GENE CASCADES IN SEX-DETERMINING PATHWAYS

Despite the variety of sex-determining mechanisms described above, they have in common that both sex determination and sex differentiation ultimately depend on a hierarchical cascade of gene expression during development, with numerous switches in pathways that eventually lead to somatic sexual differentiation and germ-line differentiation. Here we describe the hierarchy as it is known from the careful work done on *Drosophila.* The sex-determining "trigger" for male development (XY) is an imbalance between genes on the X chromosome and one of the autosomes (X:A = 0.5), whereas females (XX) have an equal ratio of genes on the X chromosome and autosomes (X:A = 1.0). A molecular counting mechanism controls early transcription of *sexlethal (Sxl),* the key gene in the sex-determining cascade in *Drosophila* that controls sex determination and dosage compensation. Intermediate genes, especially *transformer (tra),* interact with one another to control expression of the double-switch gene *doublesex (dsx).* This pathway controls somatic sexual differentiation of male and female morphology and behavior, in addition to germ-line differentiation leading to mature ovaries and testes.

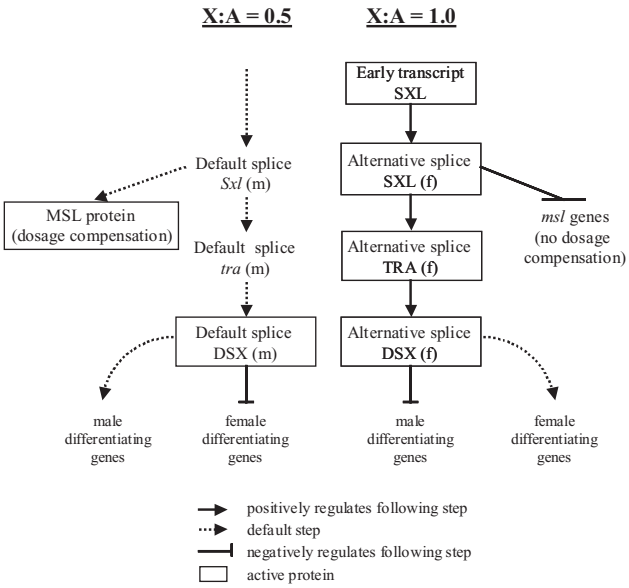

FIGURE 1 A simplified diagram of heterogametic (XX:XY) sex determination in *D. melanogaster,* showing the main genes and proteins contributing to the cascade of gene expression leading to development of males and females. This is the best understood and most fully explored genetic system of sex determination in insects, but is probably unique to *D. melanogaster* and closely related species. Names of genes are in lowercase italics; names of proteins are in uppercase.

In brief, the cascades work as follows (Fig. 1). In females (X:A = 1.0), the molecular counting mechanism activates the early transcription of *Sxl* mRNA. SXL protein produced by this mRNA regulates the splicing of subsequent *Sxl* mRNAs to produce the active female-specific form of SXL (f) protein. Active SXL (f) protein regulates splicing of *tra* mRNA to produce female-specific TRA (f) protein. In turn, TRA (f) protein interacts with other proteins (not shown) to regulate the splicing of *dsx* mRNA, generating the DSX (f) form. DSX (f) protein negatively regulates the genes responsible for male somatic sexual differentiation, but allows genes respon-

sible for somatic sexual differentiation to produce a morphologically female fly.

In male *Drosophila* (X:A = 0.5), the molecular counting mechanism prevents transcription of the *Sxl* gene early in development. In the next step, the absence of SXL protein results in unspliced *Sxl (m)* mRNA. This form includes an exon with stop codons embedded in its sequence, so that *Sxl (m)* transcripts make no functional protein. In turn, transcripts of the *tra* gene also are not spliced, again resulting in nonfunctional protein because of internal stop codons. Without TRA protein, there is no alterative splicing of *dsx,* and DSX (m) protein is made. This pathway is the default, because of the absence of functional proteins from key and intermediate genes early in the pathway. Without these proteins, transcription of the double-switch gene *dsx* results in male-specific DSX (m) protein. DSX (m) negatively regulates the genes responsible for female somatic sexual differentiation, but allows genes responsible for somatic sexual differentiation to produce a morphologically male fly.

Because male and female *D. melanogaster* have different numbers of X chromosomes (males have one X and females have two), a mechanism is required to compensate for the difference in the number of genes present in the two genotypes. This mechanism is known as dosage compensation. In male *Drosophila,* a group of genes collectively known as male-specific lethal *(msl)* genes are active and increase transcription/translation of genes on the single X chromosome. Thus, genes on the single X chromosome in males generate approximately twice the gene products, which are approximately equal to those from genes on the two X chromosomes in the female fly. In female *Drosophila* the SXL (f) protein prevents *msl* genes from acting, and there is no subsequent increase in transcription and translation of genes on the X chromosome in females.

A fairly large number of genes also control germ-line sex differentiation and the production of egg and sperm. The regulation of the genes responsible for germ-line sex differentiation is poorly understood. Although different forms of DSX protein appear to affect the morphology of the sex cells in *Drosophila,* which gene products comprise the "switch" that determines the fate of germ cells as either egg or sperm remain to be identified.

Insects other than *Drosophila* do not use exactly the same genetic pathways for sex determination. Not only are sex-determining triggers different, but for many insects that have been studied some or all of the genes described above either do not exist at all or do not perform the same function. In addition to *Drosophila,* the gene *Sxl* has been isolated from four other Diptera (the fruit fly *Bactrocrera,* Mediterranean fruit fly *Ceratitis,* phorid fly *Megaselia,* house fly *Musca*), but in none of these does SXL protein differ between males and females. Thus *Sxl* is not involved in sex determination in these insects, and since no homolog of *tra* has been found in any other insects outside of *Drosophila,* it is unlikely that the same genetic pathway for sex determination exists in other insects. On the other hand, the gene *dsx* appears to be conserved in structure and function in *Bactrocera, Ceratitis, Megaselia,* and the silkworm *Bombyx mori* (Lepidoptera). Apparent homologs of *dsx* can also be found in the nematode *Caenorhabditis (mab-3)* and in mammals *(DMT1),* and in each they have sex-specific functions in somatic sexual differentiation and/or germ-line differentiation. The control of sex determination is highly variable in insects, but control of sex differentiation leading to somatic sexual differentiation and germ-line differentiation in males and females appears to be conserved, with the possibility that genes within pathways of sex differentiation of all metazoans have common origins.

See Also the Following Articles

Chromosomes • Drosophila melanogaster • Sociality • Wolbachia

Further Reading

Baker, B. S. (1989). Sex in flies: The splice of life. *Nature* **340,** 521–524.

Beukeboom, L. W. (1995). Sex determination in Hymenoptera: A need for genetic and molecular studies. *BioEssays* **17,** 813–817.

Blackman, R. L. (1995). Sex determination in insects. *In* "Insect Reproduction" (S.R. Leather and J. Hardie, eds.). CRC Press, Boca Raton, FL.

Bull, J. J. (1983). "Evolution of Sex Determining Mechanisms." Benjamin–Cummings, Menlo Park, CA.

Cline, T. W., and Meyer, B. J. (1996). Vive la difference: Males vs females in flies vs worms. *Annu. Rev. Genet.* **30,** 637–702.

Hodgkin, J. (1990). Sex determination compared in *Drosophila* and *Caenorhabditis. Nature* **344,** 721–728.

Marin, I., and Baker, B. S. (1998). The evolutionary dynamics of sex determination. *Science* **281,** 1990–1994.

Schütt, C., and Nöthiger, R. (2000). Structure, function, and evolution of sex-determining systems in Dipteran insects. *Development* **127,** 667–677.

Stouthamer, R., Breeuwer, J. A. J., and Hurst, G. D. D. (1999). *Wolbachia pipientis:* Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* **53,** 71–102.

Werren, J. H., and Beukeboom, L. W. (1998). Sex determination, sex ratios, and genetic conflict. *Annu. Rev. Ecol. Syst.* **29,** 233–261.

White, M. J. D. (1973). "Animal Cytology and Evolution." Cambridge University Press, London.

Sexual Selection

Kenneth Y. Kaneshiro

University of Hawaii

Sexual selection depends on the success of certain individuals over others of the same sex, in relation to the propagation of the species; while natural selection depends on the success of both sexes, at all ages, in relation to the general conditions of life. Sexual selection is a struggle between individuals of one sex, generally the males, for the possession of the other sex. The result is not death to the unsuccessful competitor, but few or no offspring….

—Charles Darwin, 1859

HISTORICAL PERSPECTIVES

Darwin proposed the concept of sexual selection to explain the sexually dimorphic characters he observed in a wide diversity of organisms. He introduced the concept in *On the Origin of Species* in 1859 and further expanded on his theory in *The Descent of Man and Selection in Relation to Sex* in 1871. The development of the theory has been rich in controversies and so continues to the present, primarily because of the uncertainty over how much of the sexual dimorphism present in animals is the result of natural selection as opposed to sexual selection.

Notable theoreticians and evolutionary biologists severely criticized Darwin's theory of sexual selection and argued that sexual selection was less important than natural selection in bringing about evolutionary changes. Until recently, most evolutionary biologists accepted the notion that natural selection is the most dominant force in the evolutionary history of the species. The ubiquity of this view is well understood when one considers the notion that populations could ill afford the incapacity to adjust to the effects of their external environment. Even Darwin stated that "sexual selection will also be dominated by natural selection tending towards the general welfare of the species." Nevertheless, Darwin stood firm in his ideas and stated that his "conviction of the power of sexual selection remains unshaken."

THEORETICAL CONSIDERATIONS

Anisogamy and Parental Investment

The evolutionary origins of differences between the male and female sexes lie in what is referred to as anisogamy, which is the difference in gamete size between the two sexes. Females generally produce a few but large gametes, whereas males produce large numbers of much smaller gametes. Furthermore, such sex bias in gamete size is also characteristic of the initial stages of the evolutionary differences in parental investment, because the production of large gametes requires greater energetic investment by the female than the smaller male gametes. In most species, the male's investment ends with fertilization, while that of females may continue throughout the developmental period of the zygote, embryo, and other immature stages of the organism. Consequently, the limited number of gametes and the greater parental investment of the females result in an asymmetry in the degree of sexual selection between the two sexes. One might expect females to be generally less sexually selected but more selective in their choice of males (but see later), with males being under stronger sexual selection, hence attempting to mate with as many females as possible. It is thought that the males with the best territories, the most elaborate structures, the most attractive displays (i.e., secondary sexual traits), and so on will mate more frequently and therefore will leave more offspring. In terms of "investment" then, since it costs females much more to produce fewer large gametes than it does males, which produce lots of small gametes, females must "choose" mates that will produce the highest quality offspring (genetically).

Competition for Mates

It has been suggested that Darwin's concept of sexual selection may be subdivided into two aspects: "intrasexual selection," which involves competition between members of the same sex for individuals of the opposite sex, and "epigamic selection," which is the preferential choice of mates by one sex relative to the other. Intrasexual selection is most evident in males and perhaps can be best illustrated in species that display lek mating behavior. In such species, males compete among themselves for territories to which receptive females are attracted strictly for the purpose of mating (i.e., the territories do not provide other resources such as food or egg-laying substrates). In some species, a dominance hierarchy is established among the males that may be participants in a lek system and as a rule, the males able to occupy and defend the "best" territories will have access to most of the females that arrive at the lek. Epigamic selection, which is also referred to as "intersexual selection," is primarily an attribute of females, since the female sex ultimately discriminates among males and exercises choice of mating partners. Curiously, in most lek species, even though the most dominant males are successful in defending what seem to be the best territories in the lek system (as evidenced by the number of females that arrive at the territories of such males), males nevertheless perform an elaborate courtship ritual before mating occurs. Sometimes a female indeed rejects a dominant male's courtship and mating attempts, moves to an adjacent territory, and eventually mates with a subordinate male. Such occurrences illustrate that although fierce intrasexual selection among males can occur in competition for mating territories, epigamic selection based on female choice is the ultimate determinant of mating success among males, and thus intrasexual selection is not necessarily linked or correlated with epigamic selection. That is, a dominant male that is able to occupy and defend a prime mating territory, must, nevertheless perform courtship displays, which may or may not be acceptable to receptive females.

Runaway Sexual Selection Model

Darwin himself recognized a serious gap in his ideas of sexual selection and stated: "Our difficulty in regard to sexual selection lies in understanding how it is that the males which conquer other males, or those which prove the most attractive to the females, leave a greater number of offspring to inherit their superiority than their beaten and less attractive rivals. Unless this result does follow, the characters which give to certain males an advantage over others could not be perfected and augmented through sexual selection." Critics ridiculed Darwin's ideas and sarcastically implied that Darwin would not be able to prove his theory.

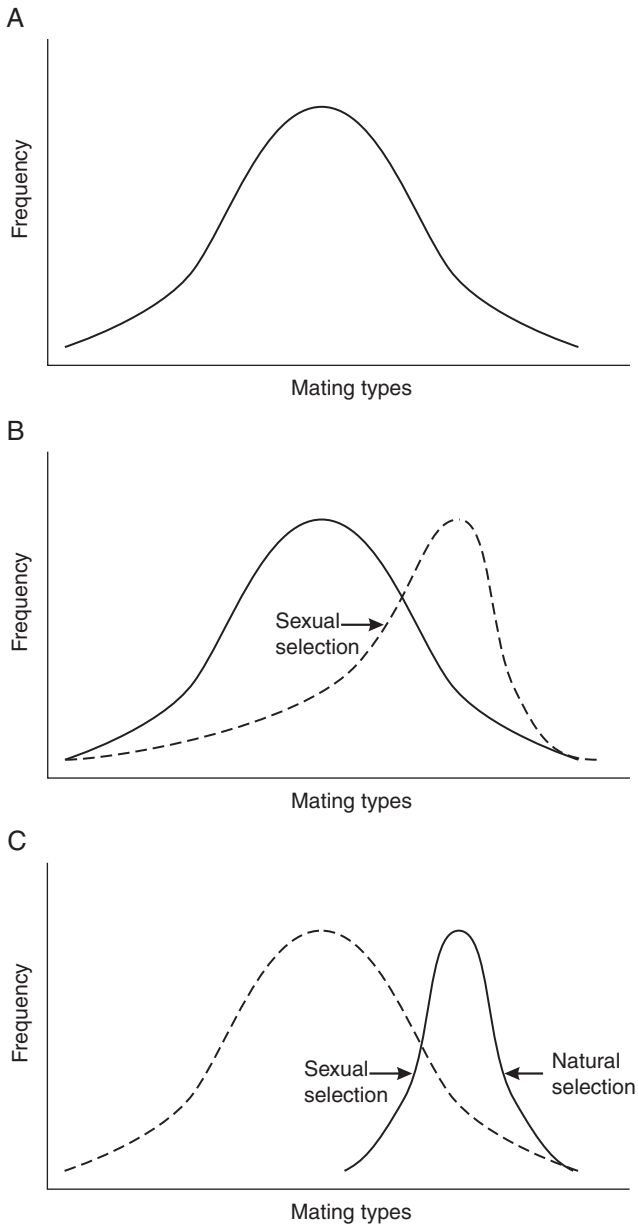

FIGURE 1 The runaway sexual selection model: (A) distribution of mating types segregating in the population, (B) strong sexual selection and elaboration of male trait that confer mating success, and (C) opposing forces of natural selection, which select for optimum phenotype for a particular environment.

The oftentimes sarcastic attacks against Darwin's sexual selection theory challenged theoreticians to develop alternative hypotheses for the evolution of female preference for, and the evolution of, secondary sexual traits in males. The most notable of these hypotheses is the "runaway selection" model, according to which it was inferred that the evolution of a sexually dimorphic character in males could result in a correlated response in the female's preference for that character. The model predicted that female choice and male characteristics influenced by sexual selection would coevolve very rapidly in an interbreeding population. The runaway selection model presumed that two forces act to counterbalance the runaway process of sexual selection. That is, female preference for a certain male character, whether morphological or behavioral, tends to select for extreme forms of that character until natural selection exerts its forces to maintain the optimum male phenotype that is best able to survive in its particular environment. In an article published in 1972, Ernst Mayr stated that "natural selection will surely come into play as soon as sexual selection leads to the production of excesses that significantly lower the fitness of the species."

The paradox of the runaway sexual selection model is that the opposing forces of sexual selection for elaboration of conspicuous male traits and natural selection, which maintains the optimum condition for the particular environment, result in reduced genetic variation for such male characters (Fig. 1). However, without genetic variation, selection can no longer occur; and unless secondary sexual characters either are linked to or are pleiotropic effects of other components of fitness, such conspicuous characters would be energetically costly to produce, and individuals possessing such traits would be in greater danger of predation.

The Differential Selection Model

In 1989, in an attempt to provide an alternative explanation for the maintenance of secondary sexual characters observed in males, Kaneshiro proposed a model based on his research on the mating behavior of Hawaiian *Drosophila* species. He suggested that there is a range of male mating types segregating within an interbreeding population. That is, some males in the population are highly successful in mating and often accomplish the majority of the matings in the population. Other males are less successful and in fact may not mate at all, despite being given the opportunity to do so with several receptive females. Similarly, among females, there are those that exhibit higher levels of mating receptivity thresholds and strongly discriminate against most of the males in the population. In the same population, there are also females that exhibit lower receptivity thresholds and seem to accept the courtship overtures of most of the males in the population.

Observations of mating experiments of Hawaiian *Drosophila* both in the laboratory and in the field indicate that the most likely matings are between males that are most successful in satisfying the courtship requirements of females and females that are not extremely choosy in selecting mating partners. The genetic correlation between these two behavioral phenotypes in the two sexes (i.e., highly successful males and less choosy females) generates the entire range of mating types in both sexes in subsequent generations (Fig. 2). In this model then, there is *differential selection* for opposite ends of the mating distributions in the two sexes and therefore, sexual selection itself serves as the stabilizing force in maintaining a balanced polymorphism in the mating system of the population. Hence, it is not necessary to invoke the forces of natural selection to counterbalance the elaboration of secondary sexual characters as required in the runaway sexual selection model.

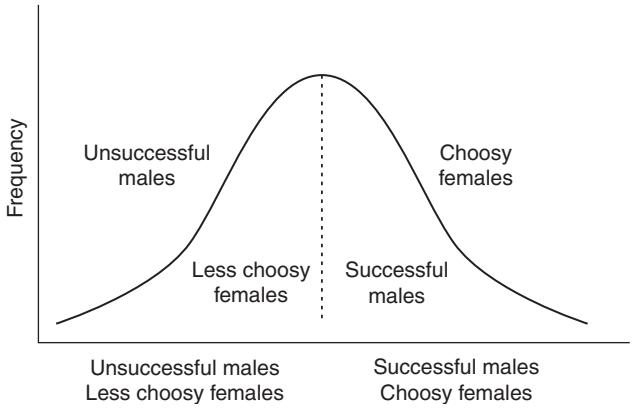

FIGURE 2 Differential sexual selection model, which shows that matings between less choosy females and successful males generate the entire range of mating types in each successive generation.

FIGURE 3 Males of the native Hawaiian *Drosophila, D. heteroneura,* in head-to-head, wingtip-to-wingtip posture, defending a territory to which females are attracted solely for the purpose of mating. The "hammerhead" shape of the males appears to be a result of sexual selection for this behavioral display among males in a lek mating system.

SEXUAL SELECTION IN INSECTS

Different patterns of sexual selection that have been observed can be illustrated by descriptions of sexual selection in a few insect species. This is certainly not a comprehensive review, since space constraints allow only a very few examples to be discussed. Several books present a more in-depth discussion of sexual selection systems in insects, and the reader is referred to such treatments.

Lek Behavior in Hawaiian *Drosophila* and Other Dipteran Species

In an article published in 1997, Shelly and Whittier list the following criteria in defining a lek species: "(1) the absence of male parental care, with males contributing nothing but gametes to the female; (2) the existence of a mating arena in which most matings occur; (3) male territories that contain no resources vital to females; and (4) the opportunity for females to freely select a mate in the arena." These authors also recognize two types of male mating aggregations involved in lek mating species: *substrate-based,* where males spend most of the time perched at stations or territories to which females are attracted for mating (although mating may occur in the air), and *aerial aggregations* or *swarms,* in which males are usually in continuous flight and matings are typically initiated in the air.

In the mating system of the Hawaiian Drosophilidae, males take up station either on the underside of leaves, one male to a leaf, or on branches of understory shrubs. Resident males aggressively defend territories from intruding males with fierce battles (Fig. 3), which in some species resemble sumo wrestling bouts. In the lek mating system of the Mediterranean fruit fly, *Ceratitis capitata* (Diptera: Tephritidae), males occupy and aggressively defend the underside of leaves of host plants as mating territories to which females are attracted solely for the purpose of mating. Once a female has arrived at the mating arena, the males of both the Mediterranean fruit fly and the drosophilids display complex courtship behaviors, which may or may not result in acceptance by the female. This means that successful aggressive behavior is not necessarily linked to or correlated with successful mating behavior. That is, a male that is successful in occupying and defending prime mating territories still must perform a courtship sequence that is acceptable by the females to ensure mating success. In some cases, the alpha male (i.e., the male that is most successful in defending a prime mating territory) is not successful in mating despite numerous visits by receptive females. Thus, at least for the drosophilids and tephritid fruit flies, although success in intrasexual selection among males may mean increased opportunity to encounter receptive females, the ultimate criterion for successful mating is epigamic selection based on female choice (Fig. 4).

Resource-Based Sexual Selection in the Scorpionfly

The mating system of the scorpionfly, *Hylobittacus apicalis* (Mecoptera: Bittacidae), provides another excellent example of intrasexual and epigamic sexual selection. Here intrasexual selection is evidenced by the competition among males for arthropod prey, which are offered to females as nuptial gifts. Males able to acquire such resources will have greater access to receptive females. It has also been shown that the size of the nuptial prey plays an important role in epigamic selection. Females apparently have the ability to evaluate the males based on the size of the nuptial gift, on which they feed during copulation. Females prefer males with a large prey versus males that offer smaller prey. Such preference results in an increased number of eggs oviposited as well as the survivorship of the female. Essentially, size of prey is correlated with the length of the copulation. It was shown that at least 5 min of mating is required before any sperm is transferred to the female and that copulation duration ranging between 5 to 20 min is directly correlated with the number of sperm transferred. Thus, when a male offers and the female accepts a large nuptial gift,

FIGURE 4 Courtship display of a native Hawaiian *Drosophila, D. clavisetae.* While in a "scorpion-like" posture, the male produces a pheromone bubble and wafts (a row of apically flattened bristles near the terminal end of the abdomen serves as a "fan") the chemical stimulant toward the female by vibrating his abdomen in an up-and-down motion toward the female. The courtship "dance" also involves visual cues (striking white face and wing markings are sexually dimorphic characters), tactile cues (female "kisses" the extended proboscis of the male), as well as acoustical cues. Despite the complex courtship displays and considerable effort on the part of the male, sexual selection via female choice (i.e., epigamic selection) plays a powerful role in determining mating success.

the male is provided an opportunity to transfer a complete quantity of sperm. In some cases, if the prey is large enough, the males may end the mating with the first female to be able to use the prey as a gift for a second female. Furthermore, a male that mates for 20 min or longer introduces a substance into the female's reproductive tract that renders the female sexually nonreceptive and stimulates oviposition. Females that mate for less than 20 min because of small prey size will continue to seek males with larger prey.

Thus, *H. apicalis* females may either reject males with small prey or mate with them for only a brief period, feeding on the prey but receiving few or no sperm. Females may then seek males that can offer larger prey, which will increase the likelihood of a sperm transfer adequate to fertilize her full complement of eggs. Such female preference for males that can offer larger, nutrient-rich prey, which influences the number of offspring produced, has clearly shaped the evolution of male behavioral components in this species. Since insect prey is a limited resource, as evidenced by the small number of males (~ 2–0%) that are carrying prey at any one moment, there is competition among males in securing prey. Males that are successful in securing a gift are not guaranteed to reproduce, since females discriminate against prey-bearing males that offer small prey. On the other hand, males that are able to secure sufficiently large prey may be able to mate with more than one female.

Female choice for prey-bearing males has apparently resulted in the evolution of other male behaviors. For example, it has been shown that males foraging for prey are more prone to predation and that males are 2.3 times more likely to be trapped in spider webs than females. Because hunting for prey can be hazardous, *H. apicalis* males can employ an alternative tactic of securing prey, namely, thievery. Some males fly to calling males and mimic female behavior—for example, perching next to a prey-bearing male and waiting for the gift to be presented, as if to a real female. If the ruse works, the thief will snatch the prey and use it to attract and feed a female without having to spend energy in a risky search of prey of his own. Thus, it would seem that sexual selection has played an important role not only in the evolution of nuptial gift giving in males but also in the development of female-mimicking behavior and prey stealing as an alternative strategy to minimize predation pressure.

THE ROLE OF SEXUAL SELECTION IN SPECIATION

Two decades ago, Mayr remarked: "Speciation…now appears as the key problem of evolution. It is remarkable how many problems of evolution cannot be fully understood until speciation is understood…." Over the past 20 years, there has been renewed interest in the process of speciation, initially to address arguments advanced by the proponents of the punctuated equilibrium model of macroevolution but also to revisit the definition of a species. At least two books address questions of speciation and the evolutionary processes of species as populational units of biological diversity. It is not the intent of this chapter to discuss the various models of speciation. *Speciation and Its Consequences* edited by Otte and Endler, published in 1989, and *Speciation and the Recognition Concept* edited by Lambert and Spencer, published in 1995, offer comprehensive reviews of this topic.

In 1997 Hampton Carson suggested that when demographic circumstances force populations to pass through a size bottleneck, the sexual selection system is temporarily disorganized by genetic recombination and the consequences of small population effects. As the population builds back up to restore efficient mate choice, novel genetic recombinants may be generated that in turn may lead to "novel selective change over a relatively small number of generations…" and that "the new population may thus be driven to undergo progressive genetic change under sexual selection."

Based on results of mating experiments on Hawaiian *Drosophila* species, Kaneshiro formulated a hypothesis, that may provide an intuitive explanation for what Carson referred to as the "driver of genetic change." Kaneshiro suggested that under conditions of drastic population reduction, there is even stronger selection for less choosy females in the population, since highly discriminant females may never encounter males that are able to satisfy their courtship requirements. If the bottleneck condition persists over a few generations, there will be a shift in the distribution of mating types in the population until a significant increase in frequency of less choosy females in the population has occurred. (Fig. 5). Such a shift in the distribution of mating types may be accompanied by a corresponding shift in the gene frequencies in the population, resulting in the destabilization of the coadapted genetic systems that had evolved in the population. The resulting destabilized

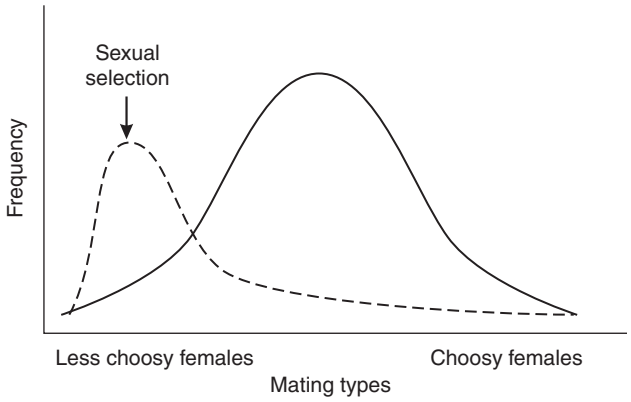

FIGURE 5 Under conditions of reduced population, there is strong selection for less choosy females, since choosy females may never encounter males able to satisfy their courtship requirements.

genetic environment thus sets the stage for genetic changes conducive to the speciation process. That is, the breakup of coadapted sets of genes allows the generation of novel genetic recombinants, some of which may be better adapted to the environmental conditions that led to population reduction. Thus, the dynamics of the sexual selection system can play an extremely important role in the speciation process as well as in the maintenance of genetic variability in the population on which the forces of natural selection can operate.

In the evolution of island biota such as that of the Hawaiian Islands, the most likely mode of speciation is what evolutionary biologists call founder event speciation. For insect groups such as the Hawaiian Drosophilidae, the most probable scenario is that a single fertilized female may be blown to an adjacent island, there to establish a new colony. For a number of generations, during the initial stages of colonization when the population size is small, there would be strong selection for females that are less choosy in selecting mates, once again, because females that are too choosy may never encounter males able to satisfy their courtship requirements. Therefore, within a few generations, there would be a shift in the distribution of mating types toward an increased frequency of less choosy females, resulting in a significant shift in gene frequencies followed by a destabilization of the coadapted genetic system. Genetic recombinants better adapted to the new habitat may be generated and strongly selected, especially if these genotypes are linked or correlated with the genotypes of the less choosy females. Thus, at least during the initial stages of colonization immediately following the founder event, the dynamics of sexual selection may play a significant role in generating a genetic environment in the population that is conducive to the formation of new species.

THE ROLE OF SEXUAL SELECTION AND NATURAL HYBRIDIZATION

It has been suggested that the dynamics of the sexual selection process also provides the opportunity for occasional natural hybridization and permits the "leakage" of small amounts of genetic material from a related species without destroying the integrity of the separate gene pools. When the population is small and there is a shift in the distribution of mating types toward an increase in frequency of less choosy females, such females in the population may occasionally accept males of a related species, which may be less susceptible to the environmental stress conditions. Genes that may be resistant to the conditions responsible for reducing the population size may be strongly selected, especially if such genes are linked to the genotypes of the less choosy females. Therefore, natural hybridization as permitted by the sexual selection process may do more than play an important role in replenishing genetic variability that may be lost owing to drift when population size is small; it may also provide a process by which resistant genes are transmitted between populations.

CONCLUDING REMARKS

The differential sexual selection model discussed here and its role in the processes of speciation may provide a possible explanation of how genetic variability may be not only maintained but actually enhanced. The generation of novel genetic recombinants and the selection for genotypes that are better adapted to changing environmental conditions are enhanced by the sexual selection system in the population, especially when subjected to population bottlenecks. It is suggested that sexual selection is a dynamic process, influenced by density-dependent factors, that enables populations to overcome environmental conditions that result in drastic reduction in population size. Shifts in the distribution of mating types when population sizes are small can generate a genetic environment producing novel recombinants able to respond to the environmental stress imposed on the population. Sexual selection influences the levels of genetic variability generated via novel genetic recombinants resulting from a reorganization of the genome, and via natural hybridization that is permitted by the sexual selection model. Thus, the biology of small populations and the dynamics of the sexual selection process are important aspects of evolutionary biology and the biology of insect populations in general. A better understanding of the role of sexual selection would provide important insights into the mechanisms of species formation as well as enabling us to develop more effective management programs, not only for pest populations but also for the conservation of rare and endangered insects.

See Also the Following Article

Mating Behaviors

Further Reading

Blum, M. S. and Blum, N. A. (1979). "Sexual Selection and Reproductive Competition in Insects." Academic Press, New York.

Carson, H. L. (1997). Sexual selection: A driver of genetic change in Hawaiian *Drosophila*. *J. Hered.* **88,** 343–352.

Choe, J. C., and Crespi, B. J. (1997). "Mating Systems in Insects and Arachnids." Cambridge University Press, Cambridge, U. K.

Darwin, C. (1859). "On the Origin of Species." Murray, London.
Darwin, C. (1871). "The Descent of Man and Selection in Relation to Sex." Murray, London.
Kaneshiro, K. Y. (1989). The dynamics of sexual selection and founder effects in species formation. *In* "Genetics, Speciation and the Founder Principle" (L. V. Giddings, K. Y. Kaneshiro, and W. W. Anderson, eds.), pp. 279–296. Oxford University Press, New York.
Kaneshiro, K. Y. (1995). Evolution, speciation, and the genetic structure of island populations. *In* "Islands: Biological Diversity and Ecosystem Function" (P. Vitousek, H. Adsersen, and L. Loope, eds.), pp. 23–33. Springer-Verlag, New York.
Lambert, D. L., and Spencer, H. G. (1995). "Speciation and the Recognition Concept." Johns Hopkins University Press, Baltimore.
Mayr, E. (1942). "Systematics and the Origin of Species." Columbia University Press, New York.
Mayr, E. (1972). Sexual selection and natural selection. *In* "Sexual Selection and the Descent of Man, 1871–1971" (B. Campbell, ed.), pp. 87–104. Aldine Press, Chicago.
Mayr, E. (1982). Processes of speciation in animals. *In* "Mechanisms of Speciation" (C. Barigozzi, ed.), pp. 1–19. Liss, New York.
Otte, D., and Endler, J. A. (1989). "Speciation and Its Consequences." Sinauer Sunderland, MA.
Shelly, T. E., and Whittier, T. S. (1997). Lek behavior in insects. *In* "Mating Systems in Insects and Arachnids" (J. C. Choe and B. J. Crespi, eds.), pp. 273–293. Cambridge University Press, Cambridge, U.K.
Thornhill, R., and Alcock, J. (1983). "The Evolution of Insect Mating Systems." Harvard University Press, Cambridge, MA.

Silk Moth

see *Bombyx mori*

Silk Production

Catherine L. Craig
Harvard University/Tufts University

Silks are highly expressed proteins produced by insects and spiders. The ability to produce silk proteins has evolved multiple times in the arthropods via two different pathways. Silks secreted via cuticular glands are produced by either immature or adult insects regardless of developmental mode; silk secretion via systemic and dedicated silk glands is present only in larval insects whose development is holometabolous. The first use of silks by insects was for reproduction; later uses were for protection and foraging.

THE MULTIPLE EVOLUTIONARY PATHWAYS OF SILKS

Phylogenetic comparison across insect taxa suggests that the ability to secrete fibrous proteins is a primitive feature of the hexapods, hence ancestral to all insect taxa. More specifically, silk production evolved first among adult hexapods, and silks were used during reproduction (Fig. 1). Among larvae, silk production evolved sporadically (Fig. 2). All the derived, holometabolous insect larvae produce silk proteins, which they use for protection.

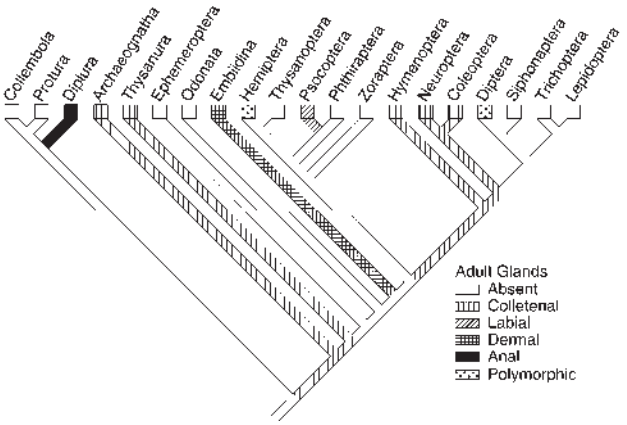

FIGURE 1 Silks produced by hexapod adults. Silk production among hexapod adults is less common than among larvae and all silks are secreted from glands that evolved for some other primary purpose. Most frequently, adult insects produce fibrous proteins in colleterial glands. (Reprinted, with permission, from the *Annual Review of Entomology,* Vol. 45, © 2000, by Annual Reviews, www.annualreviews.org.)

The evolution of silk-producing glands is equally complex. At least five taxa of adult insects produce silks in colleterial glands, and colleterial silk production has been lost and regained at least once. Among immatures, silk production in Malpighian tubules evolved early in the Ephemeroptera, then was lost and later reappeared sporadically (Neuroptera, Coleoptera, Hymenoptera) across the insect orders. Labial silk glands evolved only twice, once in the Psocoptera and later in the derived, holometabolous insect larvae (Fig. 2).

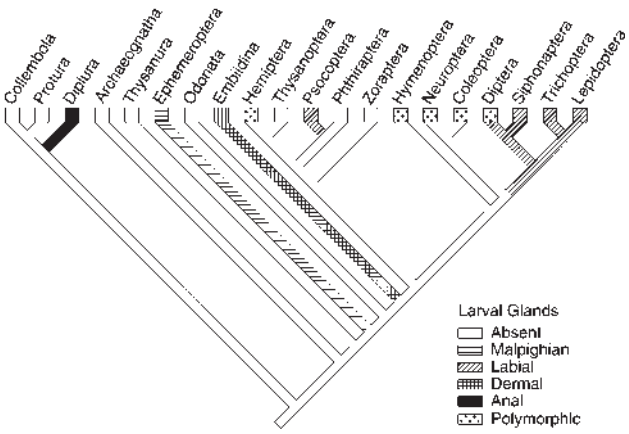

FIGURE 2 Phylogenetic analysis of silks produced by hexapod larvae. Silk production is common among hexapod larvae. The most prolific silk producers, the Embiidina and Lepidoptera, have evolved dedicated silk glands. (Reprinted, with permission from the *Annual Review of Entomology,* Vol. 45, © 2000, by Annual Reviews, www.annualreviews.org.)

Only the Embiidina and the Diplura have the ability to produce silks throughout their life. The Diplura produce silks in their anal glands, and the Embiidina draw silks from specialized dermal glands.

SILK PROTEINS AS SECONDARY PRODUCTS OF OTHER SECRETORY SYSTEMS

For most insect groups, silks evolved as secondary products of insect reproductive and excretory systems and were first used for reproductive purposes. For example, males in the primitive Apterygote order Thysanura (the bristletails) produce protein secretions in accessory genital glands. The secretions are drawn into threads used to restrict females during mating. Female mantids (Mantoidea: Mantidae) produce colleterial gland silks they use to protect their eggs, as does the beetle *Hydrophilus piceus* (Hydrophilidae). Green lacewings, *Chrysopa flava* (Neuroptera: Chrysopidae), produce colleterial gland silks from which they construct an egg sac stalk. Colleterial gland silks are characterized as having either a cross-β or α-helical configuration (Figs. 3 and 4). Except for those produced by male Thysanura, none of the colleterial gland silks are delivered through a specialized spinning organ. Furthermore, only adult insects produce silks in their colleterial or accessory genital glands, perhaps a reflection of their uses and their link to reproductive physiology.

Silks as secondary secretions are also produced in some insect excretory systems, primarily by larval insects in the Hymenoptera, Coleoptera, and Neuroptera. In most cases their silks are used for protection. Larvae of the phylogenetically primitive stingless bee *Plebeia droryana* and the bumble bee, *Bombus atratus,* bind cocoon silks (produced in modified labial glands) with Malpighian tubule secretions. In contrast, *Hypera postica* (Coleoptera) produce cocoons from Malpighian silks (cross-β conformation) that are characterized by loosely fabricated networks of coarse brown fibers. Larvae in the phylogenetically primitive neuropteran taxa Megaloptera and Raphidioidea produce Malpighian silks as well, but their conformations have not been identified. Insects in the Neuroptera produce silks both as larvae and adult but they are products of different organs. For example, Malpighian silks are produced by larval chrysopids while adults produce colleterial gland silks. Like most colleterial gland silks (except for those of the Thysanura), none of the silks produced in insect Malpighian systems are delivered through a specialized spinning organ.

Orthopterans in the superfamily Gryllacridoidea (not included in the final phylogenetic analysis because of polytomy) are unique among in the Orthoptera in that they produce labial gland secretions throughout their lives. Although, like many fibrous proteins, their molecular structure has not been determined, they have been identified as silks on the basis of their general appearance. *Camptonotus carolinensis* uses silks to roll and bind leaves together; *Cnemotettix* sp. constructs a vertical, cylindrical burrow in which it "sews" small packets of sand grains that it fits into the burrow ceiling or wall; and *Bothriogryllacris brevicauda* uses silks to tie a pebble or soil cap to seal off its burrow and protect it against desiccation.

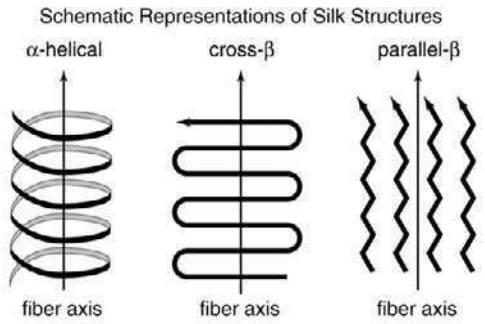

FIGURE 3 Schematic representations of silk structures. (A) The primary unit of an α-helical protein is a polypeptide chain that assumes the conformation of a rodlike coil stabilized by hydrogen bonds that form within the silk backbone. (B) β-Pleated silks are composed of polypeptide chains that are aligned side by side in a parallel or antiparallel configuration along the fiber axis. (C) Cross-β pleated sheet silks are a subset of the parallel fibroins whose polypeptide strands are oriented perpendicular to the fiber axis.

SILK PROTEINS AS PRIMARY PRODUCTS OF SILK-SPECIFIC GLANDS

Silks produced in dedicated silk glands are used mainly for protection. Character reconstruction shows that except for the silks produced by Siphonaptera, all labial gland silks are some type of parallel-β fibroin. Furthermore, except for the Embiidina and a few Psocoptera, the only insects that produce labial gland silks are insects whose development is holometabolous. Labial gland silks have evolved in some of the hemipteroid insects and insects in the Psocoptera have silk glands similar to those of caterpillars. In particular, Psocidae cover their eggs with silks, Philotarsidae spin webs singly or gregariously in small groups, and Archipsocidae spin extensive webs covering trunks and branches of large trees. Representatives of all endopterygote insect orders (Hymenoptera, Diptera, Siphonaptera, Trichoptera, and Lepidoptera), are characterized by the ability to produce labial gland silks.

SUMMARY

All silk proteins are produced in cells whose lineage is ectodermal. Most insects produce silk proteins in glands that have been modified from some other function. Silks evolved first for reproductive purposes and were produced in the anal and colleterial glands. Later, silks were used for protection and were derived from larval Malpighian tubules and labial glands. Although more ancestral insects produce silks in colleterial glands and Malpighian tubules that are characterized by an α-helical or cross-β conformation, holometabolous larvae produce labial gland silks characterized by parallel-β conformations. Only the paurometabolous Embiidina are able to produce silks characterized by a parallel-β conformation throughout their life (Fig. 4).

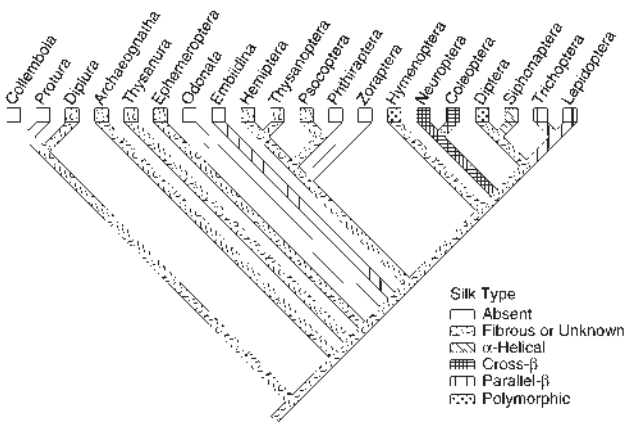

FIGURE 4 Structural configuration of silks produced by hexapods. In most cases the structural configuration of silks produced by hexapods has not been determined. The Hymenoptera produce the greatest diversity of silks. Parallel-β silks evolve only in hexapods that are able to produce both their peculiar molecular sequence and also a means of shearing the protein upon secretion.

All of the parallel-β silks are pulled (spun), in contrast to the α-helical and cross-β silks, which are deposited, secreted, or ejected. Among insects, labial gland silks evolved late in the hexapod clade, and these glands are found only among the larvae of insects characterized by holometabolous development. The tarsal silk glands of the Embiidina represent a unique evolutionary event in the history of the hexapods. All hexapod silk delivery systems or "spinnerets" are simple cuticular hardenings or pores.

This article was adapted, with permission, from the *Annual Review of Entomology,* Volume 42, © 1997, by Annual Reviews: www.annualreviews.org.

See Also the Following Articles

Accessory Glands • Bombyx mori • Embiidina • Salivary Glands • Sericulture • Spiders

Further Reading

Craig, Catherine L. (1997). Evolution of arthropod silks. *Annu. Rev. Entomol.* **42,** 231–267.

Lucas, F., and Rudall, K. M. (1968). Extracellular fibrous proteins: the silks. *In* "Comprehensive Biochemistry." (M. Florkin and E. H. Stotz, eds.), pp. 475–558. Elsevier, Amsterdam.

Rudall, K. M., and Kenchington, W. (1971). Arthropod silks: the problem of fibrous proteins in animal tissues. *Annu. Rev. Entomol.* **16,** 73–96.

Sehnal, F., and Akai, H. (1990). Insect silk glands: their types, development and function, and effects of environmental factors and morphogenetic hormones on them. *Int. J. Insect Morphol. Embryol.* **19,** 79–132.

Stehr, F. W. (1987). "Immature Insects." Kendall/Hunt, Dubuque, IA.

Silverfish

see *Zygentoma*

Siphonaptera
(Fleas)

Michael W. Hastriter and Michael F. Whiting
Brigham Young University

The Siphonaptera are laterally compressed, wingless, holometabolous insects. The order contains approximately 2575 species. All species are parasitic in the adult stage and possess mouthparts modified for piercing and sucking, highly modified combs and setae on their body and legs, and legs that are modified for jumping. Some species are vectors of disease, and current research is providing important insights into flea phylogeny and evolution.

EARLY RESEARCH ON FLEAS

Fewer than 100 species of fleas had been described prior to the end of the 19th century. In 1898 it was demonstrated that fleas were capable of transmitting plague organisms and, later, murine typhus, which stimulated a frenzy of flea studies. In the early 20th century, A. C. Oudemans, Julius Wagner, Nathan C. Rothschild, and Karl Jordan made major contributions. Karl Jordan is recognized as "the father of flea systematics." From 1953 to 1971, G. H. E. Hopkins and M. Rothschild published a comprehensive five-volume series on flea systematics, and three additional companion volumes were published for the remaining families by Mardon in 1981; Traub, Rothschild, and Haddow in 1983; and Smit in 1987. Robert Traub described 153 flea taxa and provided critical insights into flea phylogeny, convergent evolution, and zoogeography. Today there are only a few flea specialists with knowledge of the global flea fauna.

FOSSIL RECORD

Fossils of fleas are extremely rare. Three extinct species *(Palaeopsylla baltica, Palaeopsylla dissimilis, Palaeopsylla klebsiana)* have been reported from Baltic amber and a single specimen *(Pulex larimerius)* has been reported from Dominican amber. The former three species belong to an extant Palaearctic genus, whereas the latter belongs to an extant genus in which five species are confined to the Nearctic and Neotropical regions and a sixth is cosmopolitan *(Pulex irritans).*

ZOOGEOGRAPHY

Current distribution of flea families suggests an ancient origin, possibly in the Jurassic period (>140 mya). In general, families endemic to the boreal continents (North America, Europe, Asia) are predominantly Ceratophyllidae, Hystrichopsyllidae, Leptopsyllidae, Vemipsyllidae (Holoarctic), Coptopsyllidae

(Palaearctic), and Ancistropsyllidae (Oriental), whereas families endemic to the southern continents (Australia, South America, Africa, Antarctica) include Malacopsyllidae, Rhopalopsyllidae (Neotropical), Stephanocircidae, Pygiopsyllidae (Australian and Neotropical), and Xiphiopsyllidae and Chimaeropsyllidae (Ethiopian). The families Ctenophthalmidae, Ischnopsyllidae, and Pulicidae are more broadly distributed in both the Northern and the Southern Hemispheres.

CLASSIFICATION

The approximately 2575 species recognized in the order belong to 244 genera. The most recent review of the order by Lewis in 1998 recognizes 15 families. Several classifications published in recent years have substantially conflicting treatments of superfamilial relationships.

PHYLOGENY

Siphonaptera is a neglected holometabolous insect order from a phylogenetic standpoint. The major obstacle in deciphering flea phylogeny has been their extreme morphological specializations associated with ectoparasitism, which hamper the systematist from homologizing characters across taxa. The majority of characters used for species diagnoses are of limited use for phylogenetic reconstruction. Siphonaptera appears to have many instances of parallel reductions and modifications associated with multiple invasions of similar hosts, further obscuring homology. Recent molecular analyses have provided some insight into flea phylogeny. Ceratophylloidea (Ceratophyllidae + Leptopsyllidae + Ischnopsyllidae) is a monophyletic group, and the families Pulicidae, Ceratophyllidae, and Ischnopsyllidae are also each monophyletic. However, Leptopsyllidae and Ctenophthalmidae are paraphyletic assemblages. It is still unclear which flea group is most basal.

EVOLUTION

Traditional phylogenetic hypotheses favored an origin of fleas from the Diptera or Mecoptera but did not pinpoint the actual sister group. Current evidence suggests that the closest living relative to the fleas is a single mecopteran family, the snow fleas (Boreidae). This relationship is supported by the presence of unusual proventricular spines, a suite of ovariole characters (shared only by fleas and boreids), DNA sequence data, and other morphological and molecular characters. This new phylogenetic hypothesis presents a novel scenario for the evolution of fleas from a mecopteran ancestor. In Boreidae, females are wingless and males have the forewing reduced and modified into hooks, used for grasping the female during copulation. In addition, boreids have the ability to jump in a manner that appears similar to that of fleas. Boreids emerge as adults only during winter months and are closely associated with mosses, and, like many other winter insects, the reduction and the loss of wings reduce the body surface area and are probably an adaptation to the cold. The ability of boreids to jump facilitates movement on soft, fluffy snow and is also probably an adaptation to this extreme environment. When the boreid–flea ancestor shifted from a snowy, mossy habitat to the nest of a host animal, it would have lost its wings and acquired the ability to jump. These adaptations allowed fleas to move efficiently through the host's fur and between hosts. As further adaptations to a parasitic life, modifications to the primitive flea include lateral flattening, the development of sucking mouthparts, and the development of elaborate combs and setae.

MEDICAL AND VETERINARY SIGNIFICANCE

Fleas are capable of transmitting disease organisms to humans, including plague *(Yersinia pestis)* and murine typhus *(Rickettsia typhi).* They have also been implicated in the transmission of Q fever *(Coxiella burnetti),* tularemia *(Franciscella tularensis),* listeriosis *(Listeria monocytogenes),* salmonellosis, cat scratch fever *(Bartonella henselae)*, and possibly Carrion's disease *(Bartonella bacilliformis).* Several tapeworms occasionally infect humans who accidentally ingest fleas containing the intermediate stage (cysticercoid) of the tapeworms, including the double-pored dog tapeworm *(Dipylidium caninum)* and several tapeworms of rats *(Hymenolepis diminuta, H. nana).* Fleas also transmit a protozoan blood parasite *(Trypanosoma lewisi)* among rodents and a filarial worm *(Dipetalonema reconditum)* among dogs. Fleas are notorious for their biting and burrowing habits that cause intense itching and annoyance.

BIOLOGY

Detailed bionomic studies are available for only a small number of known flea taxa, so the following concepts are based on generalities found within the order.

Life Cycle

Fleas display four stadia (egg, larva, pupa, and adult) (Figs. 1A–1D). The number and size of eggs and where they are laid differ from species to species, as does the time interval the larvae require to hatch. There are generally three instars. The exceptional neosomic *Tunga monositus* has only two larval stages. Larvae are ca. 1.5 to 10 mm in length, vermiform (worm-like), eyeless, legless, and without prominent features. Prior to pupation, the third instar ceases feeding and clears its gut contents by defecation. Silk glands present in this last instar are used to spin a cocoon prior to transformation into a quiescent exarate (appendages not "glued" to the body) pupa. The emergence of adults from the cocoon requires specific stimuli (e.g., temperature, vibrations) dependent on the adaptations of the individual species. Fleas may be univoltine or multivoltine. The optimum interval for development from egg to adult ranges from 3 weeks to more than a year. Adults may live for only a few weeks to more than 3 years.

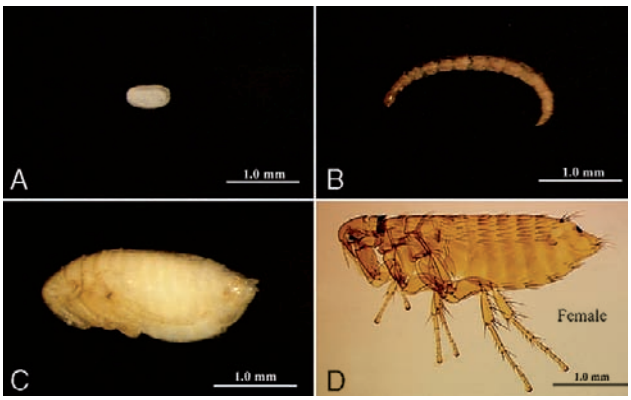

FIGURE 1 Life stages of a flea: (A) egg, (B) larva, (C) pupa, and (D) adult (a female is shown).

Reproduction

Males are smaller than females, have longer antennae, frequently have a dorsal groove on the occipital area of head, have less reticular sculpturing on sternum II, and possess an elaborate aedeagus and modified terminal segments. The interantennal groove or falx frequently divides the preantennal and postantennal areas of the head in males, while this division is often absent in females of the same species. The details of copulation are understood in only a few species of fleas. In general, the male assumes a position under the female, clasps the ventral sternites of the female with his antennae (equipped with suction cup-line organelles), then clasps the female terminal segments (sternum VII) with the tergum IX (clasper), inserts the aedeagus, and transfers sperm by extrusion of long paired penis rods that penetrate the bursa copulatrix and duct of the spermatheca into the spermatheca. Spermatozoa may be stored for long periods of time in the bulga (head) of the spermatheca. The hilla (tail) of the spermatheca of many species has an external muscle extended from its apex to the bulga. When an egg passes though the oviduct, the muscle contracts, pumping the sperm from the bulga through the duct of the spermatheca to the oviduct, where fertilization occurs through a micropyle in the surface of the egg. Depending on the species, the eggs may be laid on the host, in the nest, or randomly in the environment. The reproductive cycles of *Cediopsylla* and *Spilopsyllus* are linked to the hormonal estrous cycle of their rabbit hosts. Other species, such as *Glaciopsyllus antarcticus,* have their cycle synchronized with the seasonal migratory and nesting patterns of their avian hosts.

Host Specificity

Mammals harbor about 95% of the known flea species, with the remaining species living on birds. The seasonal and geographical distribution and host specificity of fleas are largely determined by specific requirements of the larvae. Each species has evolved a narrow range of tolerance for temperature and humidity, enabling exploitation of virtually every environment on the globe where avian or mammalian hosts flourish. These include the Antarctic *(Glaciopsyllus);* extreme xeric conditions of the world's greatest deserts (*Nosopsyllus, Hopkinipsylla, Xenopsylla* spp.); altitudinal extremes of the Andean, Himalayan, and Rocky mountains *(Amphalius, Ctenophyllus, Geusibia, Ochotonobius, Craneopsylla, Plocopsylla, Cleopsylla, Stephanocircus);* tropical rainforests *(Polygenis, Rhopalopsyllus,* Pygiopsyllidae*);* and coastal marine environments *(Notiopsylla, Paraspsyllus).* Although many species are catholic in their host preferences, the larval development is usually limited by the microenvironment of the individual host's nest. Those demonstrating the least host specificity include those free-living taxa whose immature stages can tolerate wide ranges of temperature, humidity, and nutritional requirements *(Ctenocephalides, Pulex).*

Nutritional Requirements

Most flea species require a blood meal prior to oogenesis and spermatogenesis. With the exception of some species whose reproductive cycle is synchronized with that of their host, there are few studies on the nutritional requirements of adult fleas. Although the larvae of several fleas have been noted to burrow into dead host tissues, those of *Uropsylla tasmanica* are the only truly parasitic flea larvae that burrow into and feed on the tissues of their live hosts. Some larvae have been documented to feed directly on fresh blood as it is excreted from the anus of feeding adult fleas, but most scavenge on the dried blood residues or animal dandruff and excreta. A few species are predaceous on other small organisms within the nest.

Neosomy

Neosomy is the formation or enlargement of a morphological structure from secretion of new cuticle during an active stadium of a group that normally changes form by molting. Additional cuticle forms in neosomes, accommodating expansion during enlargement (1000-fold) of gravid females that are capable of laying more than 1000 eggs. Neosomy occurs only among females because the male's principal function is to mate with embedded females *in situ.* Neosomatic growth occurs in the burrowing fleas *Neotunga euloidea* and all species of *Tunga* (Pulicidae). Fleas that do not become embedded, but attach for long periods of time, and also display neosomatic growth include members of *Chaetopsylla, Dorcadia, Vermipsylla* (Vermipsyllidae), *Malacopsylla* (Malacopsyllidae), and *Hectopsylla* (Pulicidae). Neosomatic fleas attach only once and do not feed repeatedly.

ANATOMY OF ADULT FLEAS

The morphological diversity demonstrated among families, and even among species within the same genus, is extraordinary. A glossary of terms published in 1971 by Rothschild and

Traub provides the standard for the interpretations of structures by systematists. The male aedeagus and the female seventh sternite and spermatheca are important anatomical features used in determination of species and subspecies. Study of morphological structures with the aid of a compound microscope requires mounting of fleas on glass microscope slides. Traditional techniques include puncturing the area of abdominal sterna II and III with a minuten pin, soaking for 24 h in 10% potassium hydroxide, transferring to distilled water and gently compressing the flea's abdomen to evaluate dissolved soft tissues, dehydrating in a series of ethanol solutions (70%, 80%, 95%, absolute) for 30 min each, clarifying the chitin for 15 to 20 min in methyl salicylate, transferring to xylene for a minimum of 1 h, and mounting in Canada balsam using a coverslip (No. 1 thickness). Fleas should be arranged with their heads to the right and their legs away from the technician. Thus, when viewed with a compound microscope, specimens will appear in an acceptable anatomical position with head to the left and legs toward the observer. Dissections are frequently required to permit observation of otherwise hidden structures.

Head

The preantennal and postantennal regions of the head capsule may or may not be divided by a groove or ridge that connects the antennal fossae (depressions) that lay on each side of the head. The antennae are usually longer in males than in females and the males of many species are capable of extending both antennae vertically above the dorsal margin of the head. The eye (when present) is always simple (single fascicle) and located anterior to the antennal fossa on each side of the head; it may or may not be pigmented and is often vestigial or absent. Setae arrangements on the head frequently provide important features in distinguishing taxa.

Thorax

The extraordinary jumping ability of fleas is reflected in the development of the pleural arch, which is formed from a fusion of the ventral end of the notal ridge and the transverse ridge of the metanotal area. The notal ridge is homologous with the ridge that divides the wing-bearing notum into an anterior scutal region and a posterior scutellar region of winged insects. The pleural arch is considered homologous with the pleural wing hinge of winged insects, which is composed of a rubber-like protein called resilin. The chitinous pleural arch of fleas (also composed of resilin) has shifted laterally in fleas from the more dorsally located pleural wing hinges of flying insects. This unique elastic material enhances the jumping ability of fleas. Strong jumpers have a hyperdeveloped pleural arch atop the pleural rod of the metathorax. Fleas that live their lives in the nest have little need to jump because they crawl onto the host in the nest, feed, and crawl off. The pleural arch is poorly developed or vestigial in nest fleas and they are infrequently collected on hosts outside the nest. Many taxa have no ctenidial combs on the dorsal thoracic segments, some possess pronotal combs that extend over the mesothorax, and a few genera have marginal ctenidial combs or spinelets on the metathorax.

Legs

The two front legs are more elongate than the more robust mid- and hind pairs and are articulated beneath the genal process (frequently adorned with teeth) to facilitate forward locomotion in the pelage or feathers of the host. The lengths of individual tarsal segments and the arrangements of bristles on the ventral and lateral aspects of the terminal tarsal segments are frequently important taxonomic characters.

Abdomen

The abdomen comprises 10 tergites and 9 sternites. In the genitalia of the male (Figs. 2A and 2C), the highly modified terga VIII and IX and sterna VIII and IX are used as the principal structures to distinguish many taxa. In general, the 9th tergite

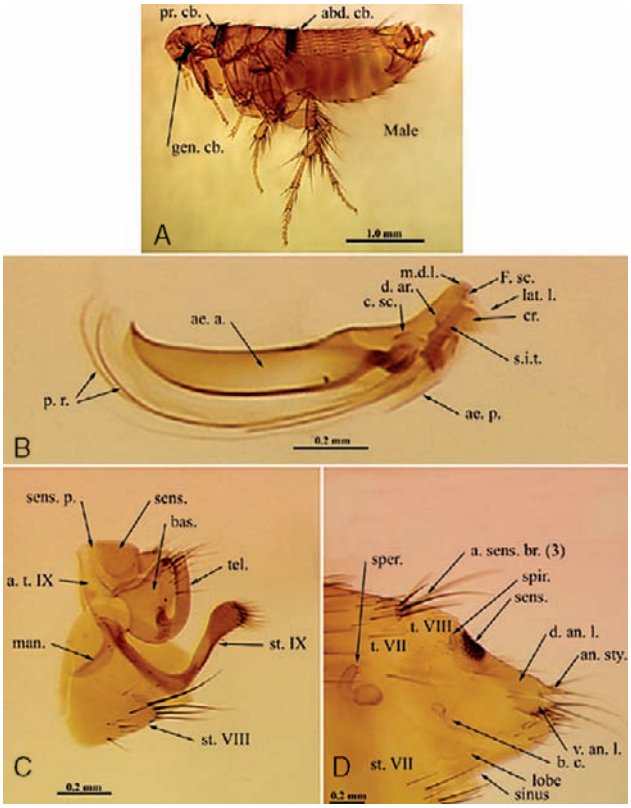

FIGURE 2 (A) Adult male. (B) Male aedeagus. (C) Male terminalia. (D) Female terminalia. Abbreviations: abd.cb., abdominal comb; ae.a., aedeagal apodeme; ae.p., aedeagal pouch; an.sty., anal stylet; bas., basimere; b.c., bursa copulatrix; c.sc., crescent sclerite; cr., crochet; d.an.l., dorsal anal lobe; d.ar., dorsal armature; F.sc., Ford's sclerite; gen.cb., genal comb; lat.l., lateral lobe; man., manubrium; m.d.l., median dorsal lobe; p.r., penis rods; pr.cb., pronotal comb; sens., sensilium; sens.p., sensilial plate; s.i.t., sclerotized inner tube; tel., telomere; v.an.l., ventral anal lobe.

is divided into an immovable process, or basimere; an articulated movable process, or telomere; and a manubrium. Various arrangements of lobes and processes adorn the basimere and are important features to distinguish taxonomic groupings. The 8th tergite is either small and obscure or greatly enlarged to ensheath the 9th tergite and the aedeagus. The 8th sternite is greatly expanded in some groups and reduced or pencil-like in others. An important taxonomic feature of most species is the 9th sternite. The aedeagus is a highly complex structure used extensively in the systematics of the order. The taxonomic significance of the terminal lobes (median dorsal lobe, lateral lobe, ventral lobe), crochet, sclerotized inner tube, crescent sclerite, penis rods, aedeagal apodeme, dorsal armature, and Ford's sclerite are of particular importance. The female (Figs. 1D and 2D) 8th tergite is variably shaped and frequently enlarged to cover the less developed ventral terminal segments. The caudal margin of the 7th sternite is modified in many species and is an important taxonomic character. The various lobes and sinuses of the 7th sternite are thought to be important anatomical features used during copulation. Nine genera possess two spermathecae, each with a duct leading to the perula of the bursa copulatrix. The remainder have but one spermatheca, although a remnant blind duct leading from the bursa copulatrix can be seen in most of them. The bursa copulatrix is often sclerotized and it and the spermatheca distinguish many species. The anal opening passes between the dorsal and the ventral anal lobes, each considered the 10th tergite and sternite, respectively. The 10th tergite usually bears an anal stylet with a variable number and arrangement of apical setae, a character used to distinguish species.

ANATOMY OF LARVAE

The diversity of morphological characters parallels those of adult fleas. Important structures include the shape of the egg tooth (first instar only); chaetotaxy and sensory structures of the head capsule, antenna, and antennal mound; cuticular sensilla of the dorsal plates of the thorax and abdomen; number and length of body setae; sculpturing of the integument; and arrangement of the anal comb and other setae on the 10th abdominal segment.

CONTROL OF FLEAS

Control is often necessary when fleas parasitize livestock and pets, become potential vectors of disease, or are biting people. Household sanitation such as thorough vacuuming of indoor areas and washing of pet bedding removes many of the immature stages (eggs, larvae, and pupae). Insect growth regulators (methoprene and pyriproxyfen) that inhibit the development of immature stages are used in conjunction with organophosphates (chlorpyrifos and diazinon) to interrupt the developmental cycle. Dog and cat flea collars and the use of systemic insecticides add additional preventive measures. Control of potential plague vectors in rural or sylvatic environments requires more drastic measures. Wild mammals such as rats, prairie dogs, ground squirrels, and chipmunks must be controlled around human habitations and recreation sites frequented by people. In such situations it is necessary to apply Carbamates (carbaryl) to burrows and rodent harborages prior to control of animal populations. Failure to control fleas prior to implementing animal control endangers humans that may come into contact with these potentially disease-infected fleas.

See Also the Following Articles

Bubonic Plague • Cat Fleas • Veterinary Entomology

Further Reading

Hopkins, G. H. E., and Rothschild, M. (1953–1971). "An Illustrated Catalogue of the Rothschild Collection of Fleas (Siphonaptera) in the British Museum (Natural History)," Vol. I–V. British Museum (Natural History), London.

Lewis. (1998). Resume of the Siphonaptera (Insecta) of the World. *J. Med. Entomol.* **35,** 377–389.

Mardon, D. K. (1981). Pygiopsyllidae, Vol. VI of "An Illustrated Catalogue of the Rothschild Collection of Fleas (Siphonaptera) in the British Museum (Natural History)." British Museum (Natural History), London.

Rothschild, M., and Traub, R. (1971). "A Revised Glossary of Terms Used in the Taxonomy and Morphology of Fleas." British Museum (Natural History), London.

Smit, F. G. A. M. (1987). Malacopsyllidae and Rhopalopsyllidae, Vol. VII of "An Illustrated Catalogue of the Rothschild Collection of Fleas (Siphonaptera) in the British Museum (Natural History)." British Museum (Natural History), London.

Traub, R., Rothschild, M., and Haddow, J. (1983). "The Rothschild Collection of Fleas. The Ceratophyllidae: Key to the Genera and Host Relationships." University Press, Cambridge, UK.

Whiting, M. F. (2001). Mecoptera is paraphyletic: Multiple genes and a phylogeny for Mecoptera and Siphonaptera. *Zool. Scripta.* **31,** 93–104.

Snakefly

see *Raphidioptera*

Sociality

James E. Zablotny
U. S. Department of Agriculture

Insect sociality refers to populations of insects that use some form of social behavior. Eusocial insects, for example, are true social species that possess the following three characteristics: cooperative brood care, overlap of two or more generations with offspring assisting with brood care, and reproductive division of labor. In a broad sense, the term "social" is often

used for pre- and subsocial insects that have fewer than three of the characteristics of eusociality. In a strict sense, a social insect is usually understood to mean a representative of the ants, termites, vespoid wasps, and social bees, the true eusocial insects. Social insects are an important component of Earth's biological diversity. E. O. Wilson has noted that eusocial insects (including termites, ants, bees, and wasps) make up a significant proportion (75%) of the world's insect biomass. The evolutionary and ecological success of social insects is evident in both tropical and temperate ecosystems.

Sociality is a life history strategy that has evolved multiple times within and among diverse and distantly related insect taxa. Because of large population sizes, social insects collectively consume more energy and biomass than vertebrate animals. In tropical ecosystems, ants and termites play a major role in nutrient cycling and soil turnover. In 1990 B. Hölldobler and Wilson noted that in tropical and temperate habitats, the amount of soil turnover attributable to ants and termites far outweighs the cycling done by earthworms. In the region of the Brazilian Amazon, ants release a significant quantity of atmospheric formic acid, which contributes to natural acidification of rainwater and weathering of limestone-based rock formations.

In eusocial insects, the ability to reproduce has been lost in worker and guard castes. Sterile workers care for the brood, provide nourishment, maintain environmental conditions (e.g., by thermoregulation) inside the nest, and remove waste. The female reproductives or gynes generally outlive several generations of the offspring and are protected and tended by workers in the confines of the nest. The presence of reproductive division of labor is vividly portrayed in social insects, with reproduction restricted to the queen, and nest or colony support being provided by the sterile worker and guard castes. With these three characteristics operating, an insect society is a highly organized, cooperative group of organisms that functions like a superorganism. Social and physiological homeostasis of the insect society is maintained through the integration of caste members through pheromone, tactile, auditory, and sometimes visual communication.

The evolutionary patterns within the Hymenoptera are complex, and eusociality has evolved over eight times within the social bees. The Isoptera (termites) is the only insect order that is exclusively eusocial. Eusociality is rare but present in some aphids (Heteroptera: Cerataphidini) and thrips (Thysanoptera). In aculeate Hymenoptera, many examples of eusocial species are present and include all ants (Formicidae), one group of sphecoid wasps (Crabronini), vespoid wasps (Vespidae, Masaridae, and Eumenidae), and the Apoidea or social bees (Apidae: Bombinae, Apinae, Meliponinae, and Euglossinae).

DIVERSITY IN INSECT SOCIALITY

In 1923 W. M. Wheeler published one of the first contemporary treatments of insect sociality. In the 1960s and 1970s, usable classification schemes for insect sociality evolved from the works of C. D. Michener and E. O. Wilson. Wilson's 1971 treatment of the social insects streamlined and established a standard vocabulary for describing the degrees of insect sociality, ranging from solitary to eusocial, with several different grades or intermediate forms (Table I).

TABLE I Synoptic Overview of the Degrees and Characteristics of Sociality Derived from Wilson

Degree of sociality	Characteristics[a,b]		
	Overlapping generations	Reproductive castes	Cooperative brood care
Quasi-social	–	–	+
Semisocial	–	+	+
Eusocial	+	+	+

[a]Plus sign (+) denotes presence and minus sign (–) indicates an absence of a characteristic.

[b]Solitary, subsocial, and communal insects do not have overlapping generations, reproductive castes, or cooperative brood care.

Presocial Insects: Subsociality

Most insects lead solitary lives, with many species forming simple aggregations during mating or at commonly used food sources. Sociality appears in insects as a departure from a solitary life history strategy. Presocial insects have one to two characteristics of sociality. One subgroup of presocial insects includes insects that provide limited parental care to their offspring. Labeled as subsocial, these species protect or shelter their eggs, nymphs, or larvae; eggs and larvae may be given parental care in an enclosed nest.

Most subsocial insects provide care of offspring in the absence of a nest or shelter. Many examples of subsociality are known from the true bugs (Heteroptera). For example, in some belostomatid genera *(Belostoma, Abedus)* the female lays her eggs on the male's back, and the male carries the eggs until nymphs hatch. Males of the belostomatid *Lethocerus americanus* assist the female with active guarding of the eggs that are laid on aquatic vegetation. Parental care of eggs is also known in water striders (Gerridae) and leaf-footed bugs (Coreidae). Active guarding of eggs and newly hatched nymphs is known in the Acanthosomatidae, Pentatomidae, Tingidae, Reduviidae, and Scutellaridae. Usually, the female guards the clutch and covers the first instars when a predator approaches. The nymphs often orient themselves toward the parent to facilitate concealment.

More examples of behaviorally advanced stages of offspring care can be found in the orders Heteroptera and Coleoptera. Treehoppers (Heteroptera, suborder Auchenorryncha; Membracidae) also exhibit parental behaviors toward early instars. R. Cocroft, for example, observed female thornbugs *(Umbonia crassicornis)* straddling egg clutches to prevent parasitism and creating feeding slits on plant stems for hatchlings. Additionally, the mother actively herds wandering nymphs to

keep them near or around the feeding slits and emits alarm calls to alert offspring to a predator's presence.

Within the Coleoptera, highly developed forms of parental behavior and nest making have been discovered in the Silphidae and in the dung-feeding Scarabaeidae. Known as burying beetles, species of *Nicrophorus* (Coleoptera: Silphidae) provide provisions and actively feed newly hatched offspring through regurgitation. After discovering a small animal carcass, *Nicrophorus* parents process the carcass into a congealed ball of putrefying flesh. After burying the provision, the female chews a shallow depression into the mass, where she deposits her eggs. To solicit parental feeding, the larvae raise the anterior portions of their bodies and wave their thoracic legs. The female parent then regurgitates a small droplet of food material, which is devoured by a larva. In some *Nicrophorus* species, males also participate in larval feeding but less frequently than the female. Parental feeding in *Nicrophorus* is essential for development of the young: removal of the female parent inhibits metamorphosis to the adult.

G. Eickwort and Wilson categorize most bark and ambrosia beetles as subsocial or colonial insects. Members of the Scolytidae and Curculionidae (Platypodinae) have elaborate life history strategies that enable them to evade a host tree's defense mechanism. The scolytid beetle *Xyleborus* produces communal brood chambers where all life stages reside. In similar manner to Hymenoptera, these xyleborines also have haplodiploid sex determination. Another similarity between xyleborines and eusocial Hymenoptera is the presence of a biased sex ratio in favor of females. Female offspring assist in cultivating ambrosia fungi, which provide the cellulase necessary for digestion of wood fibers.

Parasocial Insects: Communal Societies

"Parasocial" refers to species that exhibit communal, quasi-social, and semisocial behavioral patterns with some form of generation interaction between generations. Communal species include insects of the same generation that share a nest site without any form of collective offspring care. A communal life history strategy is not just an aggregation: it may involve complex forms of communication and offer some distinct advantages over a solitary life history strategy. Communal aggregations enhance accessibility to food materials that are difficult to process, facilitate thermoregulation, and offer protection from predation. Examples of communal social behavior in bees can be found in the Andrenidae, Megachilidae, and several subfamilies of the Halictidae. Multiple females construct a large, composite nest, but each female tends only to her own nest cells. This produces a nest having a single nest opening that is more easily defended by multiple female bees. Many different species of beetles have communal life history strategies. Some species of scolytid bark beetles are exemplary communal insects.

Communal behaviors in larval insects can be found in tent-making caterpillars (Lepidoptera: Lasiocampidae) and some of the sulfur butterflies (Lepidoptera: Pieridae), silk moths (Lepidoptera: Saturniidae), nymphalid butterflies (Lepidoptera: Nymphalidae), and tussock moths (Lepidoptera: Lymantriidae). Batch laying of eggs predisposes some larval insects for communal life history strategies, but other species disperse from a common oviposition site to lead solitary lives. Although complex social interactions between group members are limited in many species of caterpillars, some species utilize central-place foraging that requires an extensive communication network within the caterpillar community. Central-place foraging is not tied universally to web making, but instead is scattered throughout the lineages of Lepidoptera. In 1925 F. Balfour-Brown described a model for the evolution of elaborate caterpillar web dwellings from a mat of silk threads or carpet web. This model suggests that carpet webs, the ancestral condition, further evolved with caterpillars that used feeding webs and later to those that used domicile or home webs.

In the jack pine sawfly, *Neodiprion pratti banksinae* (Hymenoptera: Neodiprionidae), communal feeding facilitates the procurement of nutrients from tough waxy pine needles. The collective efforts of many larvae chewing over the same area on a pine needle reduces the amount of time needed to break through the cuticle. Other species of larval sawflies and Lepidoptera use group defense postures in which the aggregation moves in a synchronized manner to thwart a predator. The sawflies *Neodiprion sertifer* and *Diprion pini* use synchronized head jerking and twitching to repel both egg-laying ichneumonid wasps and birds. To be effective, these protective behaviors require a coordinated group response. Similar defense strategies are also known in the woolly or eriosomatid aphids in which a coordinated waving of waxy filaments may intimidate predators.

Parasocial Insects: Quasi-Social Societies

Unlike communal insects, quasi-social insects have some cooperative brood care while sharing a common nest site. Female quasi-social bees participate in nest guarding and assist with building and stocking nest cells with provisions. In 1967, R. B. Roberts and C. H. Dodson discovered that multiple female orchid bees (Apidae: Euglossinae) often occupy a shared nest site, with the number of females present usually exceeding the cells available for egg laying. This finding suggests that euglossine bee societies have a form of cooperative brood care. Temporary quasi-sociality is also known in a few species of polistine wasps with age polyethism. Young females begin their adult lives as foragers and assist with nest provisioning. Later on, these individuals become egg layers in the common nest site.

Parasocial Insects: Semisocial Societies

C. D. Michener has stated that semisocial insects have an additional characteristic of sociality over the quasi-social species in that the former also have some form of reproductive caste that is cared for by a worker caste. Variation in ovary matu-

ration in spring halictine bee populations suggests that the semisocial condition can arise spontaneously within a mixed assemblage of reproductives of different stages of maturity. Similar patterns of seasonal semisocial behavior occur in some *Polistes* wasps.

Eusocial Insects

The fundamental difference between semisocial and eusocial insects is that eusocial insects have overlapping generations that feature cooperative brood care by the offspring. Eusociality has evolved multiple times within insects and first appeared in the termites. Eusociality in lower termites (Hodotermitidae, Termopsidae, Kalotermitidae, Mastotermitidae, Rhinotermitidae) is associated with trophallaxis and enables all individuals of the nest to be inoculated with gut symbionts necessary for the digestion of cellulose. Trophallaxis also serves to transfer chemical communication in lower and higher termites. Higher termites such as the Termitidae consume sources of cellulose other than wood, lack flagellate endosymbionts, and use anal trophallaxis for transferring chemical cues.

N. Lin and C. D. Michener proposed two distinct pathways of character evolution in social insects. Most bees are believed to have evolved eusociality through a parasocial sequence. Ancestral social bees used a communal social strategy with cooperative nest construction. Later generations switched from a communal to a quasi-social strategy incorporating cooperative brood care within an annual life cycle. The semisocial stage followed the quasi-social stage, with a well-defined worker caste providing cooperative brood care. Fully evolved eusocial species then acquired the characteristic of overlapping generations.

Ants, termites, vespid wasps, and some social bees are believed to have used a subsocial sequence for the evolution of eusociality. The subsocial sequence for the evolution of eusociality is based primarily on the degree of brood care by the female with respect to the age of the offspring. In basal or ancestral taxa, brood care was given only to early instar offspring, with the mother abandoning the young well before maturity. In stage two, brood care is extended to the time during which the young reach maturity. In the final sequence of events, some of the brood become permanent, sterile workers and provide cooperative brood care to the female parent.

Eusociality has also evolved in some gall-forming thrips and aphids. These insects have recently been classified as eusocial because of the presence of a sterile guard caste. However, this form of restricted eusociality differs in important ways from the eusocial condition of Hymenoptera. The guard caste is the only nonreproductive caste in aphid and thrips societies, and the guards do not care for the young. In aphids, this is partly because aphid offspring are precocious like small versions of the adult, they are motile and able to feed themselves.

B. Crespi discovered eusociality in several species of phlaeothripine gall-forming thrips. In these eusocial thrips, the first generations of adults develop into sterile soldiers and are used for defending the gall against invasive species such as ants. These Australian thrips reside in *Acacia* leaf galls or leaf mines and utilize social behaviors for defense of the domicile. The guards also assist with domicile construction. Morphological specialization associated with the guard caste in eusocial thrips includes enlarged prothoracic legs that are used to attack invaders. The prothoracic femora may be expanded and adorned with robust spines. Mortality in the soldier caste is high.

S. Aoki along with Y. Ito have shown that soldier castes are also present in the Cerataphidini or bamboo aphids (Hormaphididae) and in the gall-forming pemphigid aphids. Ito found that some of the Cerataphidini produce a sterile guard caste that defends the offspring. In *Ceratovacuna* and *Pseudoregma,* members of the soldier caste are adorned with sharply pointed spines on the head and are used for colony defense on secondary hosts. In *Pemphigus obesinymphae,* sterile guards are also used for protecting the domicile. The guards are larger than the nymphs and kick at invading predators with their long legs. In social aphids, the guard caste never assists with brood care but serves only to protect the colony from predation. Parthenogenesis is a reproductive mode employed by aphids, but how the soldier caste is determined is unknown.

In 1992 D. S. Kent and J. A. Simpson discovered a eusocial ambrosia beetle (Coleoptera: Curculionidae). A member of the Platypodinae, *Austroplatypus incompertus* lives within heartwood galleries of *Eucalyptus.* Platypodine beetles like *A. incomperus* feed on ambrosia fungi, and nonreproductive females tend the galleries and process fungal mycelia. Kent and Simpson found that established colonies older than 2 years were composed of a single fertilized female, and up to five unfertilized females, which helped maintain the gallery structure and provide protection from predators. The presence of overlapping generations, reproductive/nonreproductive castes, and cooperative brood care within a shared domicile provide evidence for eusociality in these ambrosia beetles.

Within the Hymenoptera, eusociality is found in all ants (Formicidae), vespid wasps (Vespidae), and some bees (Apidae). Members of the subfamilies Bombinae, Meliponinae, and Apinae are eusocial. Sex determination in Hymenoptera follows Dzierzon's rule, first shown by J. Dzierzon in 1845 with honey bees and now termed haplodiploidy. Male progeny are produced from unfertilized eggs through a process called arrhenotoky. In many species of social Hymenoptera, males can be produced at the appropriate time from unmated females within the colony. Workers will sometimes undergo facultative thelytoky to produce diploid eggs. Automictic parthenogenesis produces the diploid eggs in queenless hives of *A. mellifera capensis.* This form of thelytoky is uncommon and is restricted to certain subspecies of the honey bee.

COLONY SIZE AND LONGEVITY IN EUSOCIAL INSECTS

Insect societies may range in size from fewer than 10 to over 3 million adults. Some species of termites (families Termitidae,

Mastotermitidae, and Rhinotermitidae) have colonies with millions of members. Colonies of vespid wasps can have 100 to over 5000 members; colonies of the Allegheny mound ant, *Formica exsectoides,* can exceed 40,000; and nests of *A. mellifera* usually contain between 20,000 and 80,000 individuals.

Species that build nests located on or in the ground tend to have more colony members than species that use more restrictive nest sites. Ground-nesting vespid wasps, for example, usually have larger colony sizes than wasps that build arboreal carton nests. Species that have a large colony size generally tend to have more complex social behaviors than species with small colonies. The correlation of increased colony size with highly evolved social behavior suggests that complex behavioral integration is necessary for colony survival.

Although most data gathered on longevity have been derived from laboratory populations, the female reproductive or queen of some eusocial insects can live for more than 10 years. Well-protected inside the nest, most queens are tended by a court of workers and insulated from the hazards of everyday life. Workers have significantly shorter life spans and are quickly replaced with newer offspring. Attrition is high among workers, and approximately 1% of the worker caste of the army ant *Eciton hamatum* is lost daily. Unlike the queen, workers are exposed to many hazards, and activity outside the nest contributes greatly to worker mortality.

ROLE OF ALTRUISM IN INSECT SOCIALITY

Kin selection is the form of natural selection essential for explaining how the characteristics of eusociality evolved. In 1964 W. D. Hamilton published a mathematical model explaining the economic basis for self-sacrifice as a result of altruism. Hamilton's rule, $C/B < r$, suggests that the ratio of the cost C of sacrifice between two closely related individuals to the benefit B of the receiver r must be smaller than the degree of relatedness between the two individuals. Thus, the inclusive fitness of a group of related individuals is the amount of individual and kin selection. The application of Hamilton's rule to social Hymenoptera, which follow Dzierzon's rule with haploid males and diploid females, is illustrated by the following example. A female offspring inherits 50% of its genome from its father and 25% of its genome from its mother. Hamilton's rule predicts that sisters produced from the same mating should tend to sacrifice themselves for each other because they share 75% of their genetic material with each other and 50% of their genetic material with their mother. Thus, the benefit for an altruistic behavior does not have to be high to spread throughout a population.

In 1977 H. E. Evans proposed that selection factors other than kin selection could explain the evolution of altruism and social behaviors in Hymenoptera. Extrinsic factors such as nest parasitism are just as plausible as kin selection for the evolution of altruism. Evans noted that in general, polygynous (multiple queens per nest) wasp species are more common than monogynous (single egg-laying queen per nest) species. Evans hypothesizes that polygynous nesting strategies are ubiquitous because they are necessary for protecting the progeny and provisions of the extended family group from nest parasitism. Likewise, polygynous nests can provide many foragers for provisioning the nest as well.

Although altruism can be detrimental to the individual, advantages to the group or colony are many. Most social insects utilize group defensive strategies when the nest or colony is under attack. Aggressive defense strategies observed in social Hymenoptera, social aphids, and termites demonstrate that defensive behaviors are usually detrimental to the defenders but facilitate survival of the reproductives and brood within the nest. Self-sacrifice occurs frequently in social insects. In honey bees, for example, the barbed sting of the worker bee eviscerates and kills the worker after stinging, suggesting that the cost of altruism to the individual is high but trivial to entire colony. In termites, guards will position themselves closest to the source of danger when provoked, whereas other colony stimuli are generally ignored.

Although queen adoption is most often associated with colony multiplication, Wilson believes that the genetic advantages for this strategy are clear. Increasing the number of queens in an isolated colony serves to increase the effective population size while incorporating more altruistic traits into the colony. However, polygyny in ants is not restricted to rare, isolated populations, but can also be found among ubiquitous ant populations as well.

CASTES IN SOCIAL INSECTS

The caste system and polyethism determine the division of labor in insect societies. Castes divide colony members into distinct functional roles within the insect society. Specialized caste members conduct such tasks as defense, brood care, nest construction, thermoregulation, provisioning, and egg laying. Morphological specialization or caste polyethism is a key feature of caste structure. The role of a caste member may also change with age (age polyethism).

Morphological specialization associated with caste polyethism is often the result of allometry and usually is tightly coupled with behavioral changes. Allometry or differential growth rates can provide dramatic changes in morphology with simple change in relative growth rates (Fig. 1). Allometric growth provides termite or ant guard castes with hypertrophied mandibles that are used for defense. The head capsule, which may also be enlarged relative to other body parts, accommodates the mandibular muscles used for biting and dismembering enemies.

Overall change in body size also can be important in defining the queen caste. In most eusocial insects, the queen is larger than nonreproductives. The abdomen, which houses the ovaries, may be grotesquely enlarged in mated termite queens and some ant queens. Such enlargement (physogastry) also is known in replete ants, which gather and store honeydew in the crop.

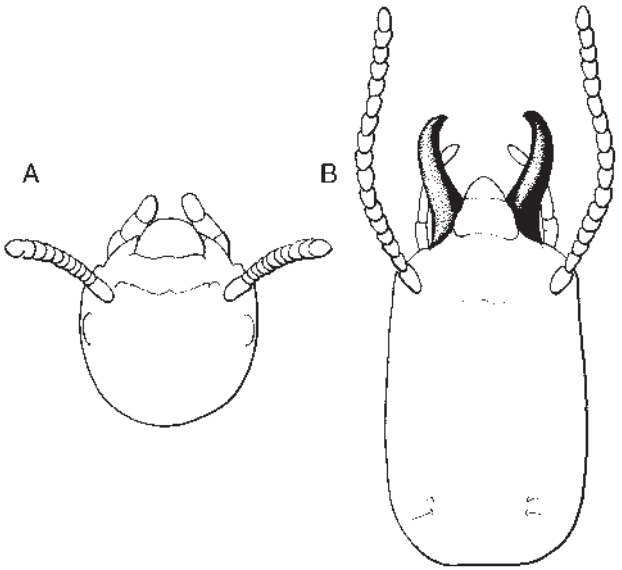

FIGURE 1 Cranial morphological differences in worker and soldier or guard termite castes. In *Reticulotermes flavipes,* workers (A) have short chewing mandibles concealed beneath the clypeus and soldiers (B) have an elongated, enlarged head with prominent mandibles.

The castes of typical social insects can be categorized according to the following classification scheme. Within social Hymenoptera, guards serve as nest defenders and are equipped with well-developed mandibles and potent stings. Workers, the most ubiquitous caste, help with nest construction and repair, thermoregulation, provisioning, and brood and queen care. The queen makes up the reproductive caste, and her egg-laying activity provides the fecundity for maintaining the colony. Most of the members of an insect society are workers, and high numbers of this caste are necessary for maintaining the physiological homeostasis of the domicile.

Many differences are evident between the caste system of termites and social Hymenoptera. In lower termites, the second instars function like workers and are actively involved in nest maintenance. Because, at least in part, of their hemimetabolous development, larval termites are active and also take care of themselves as well. Unless destined to become guards, most apterous termite larvae remain as workers throughout their lives. Termite castes differ from social Hymenoptera castes in that they contain both sexes, with reproduction restricted to the fertilized queen (Table II). Once established, termite colonies can replenish reproductives from several distinct supplementary reproductive castes. Within the Termitidae, adultoid reproductives can become functional reproductives and replace lost reproductives. Nymphoid reproductives (larvae that possess short wing pads) can become fully functional supplementary reproductives. Ergatoid reproductives are neotenic, apterous individuals with pigmented exoskeletons. Inhibitory pheromones released by functional reproductives prevent the supplementary reproductives from attaining functionality.

In honey bees, morphological and behavioral differences in castes are regulated by endocrine control of development, dietary differences of larval workers and queens, and queen pheromonal conrol. Hormonal developmental control is associated with the activity of juvenile hormone (JH) on target tissue sensitivity to ecdysteroids. This modulates or controls reproductive tissue differentiation in developing larvae. In addition, JH may inhibit the release of ecdysteroid and determine the duration of the last instar feeding stage. In developing queens, JH may inhibit ovariole cell death, a feature common in developing workers. Other hormones also influence ovariole development in larval queens. Makisterone A, an ecdysteroid secreted by the prothoracic glands, promotes ovariole differentiation.

Dietary differences are correlated with reproductive caste development in honey bees. Honey bee worker larvae are fed pollen and honey, whereas queen larvae are fed royal jelly. The hypopharyngeal and mandibular glands of workers produce royal jelly. The production and feeding of royal jelly to select larvae is associated with lowered amounts of queen substance, reduced queen vitality, or queen death; the construction of queen cells is also debilitating to the queen. Different proportions of proteins, lipids, carbohydrates, and trace compounds are apparent in honey and royal jelly, but the active ingredient in royal jelly responsible for queen development is still unknown.

Pheromonal control is also evident in caste formation in honey bees. Queen mandibular gland secretion reduces the amount of JH in developing worker larvae, stimulates worker foraging, promotes queen tending, and inhibits swarming

TABLE II Generalized Classification Scheme of Caste Structure in Termites Derived from Wilson

Caste	Morphological features	Roles
Larvae[a]	Usually apterous (wingless) and lack reproductive and guard caste structures	May develop into workers or guards
	Brachypterous larvae (wing pads present) are known as nymphs	May develop into functional reproductives
Worker	Absence of compound eyes and ocelli; apterous, with reduced pterothorax, well-developed jaws	Nest construction, sanitation, provisioning, brood care
Pseudergate	Regressed nymph, found only in lower termites	May develop into secondary reproductives, soldiers
Soldier/guard	Enlarged mandibles, hypertrophied head glands, head	Defense
Primary reproductives[b]	Queen or male derived from alate adults	Colony founders

[a]Immature termites known as larvae are further grouped by the presence or absence of distinct wing pads.
[b]Primary reproductives become dealate reproductives after shedding their wings.

behavior. Through modulation of JH titers in developing workers, queen substance inhibits ovariole development.

Examples of age polyethism are frequent in social Hymenoptera. In worker honey bees and ants, temporal changes in maintenance behaviors are associated with age. Young workers take care of the queen and feed the larvae, whereas older workers conduct most of their activities outside the nest. Behavioral changes in honey bees also correlate with the thickness of the mandibular, hypopharyngeal, postcerebral, wax, and thoracic glands. The nursing period of worker bees coincides with the maximum development of mandibular and hypopharyngeal glands at between 5 and 10 days of age.

COMMUNICATION IN EUSOCIAL INSECTS

The functional integrity of an insect society is established through complex forms of communication. Communication pathways, either chemical, visual, or acoustic, can be used to convey alarm, simple and multiple attraction for assembly, recruitment to a new domicile or food source, grooming, trophallaxis and exchange of food material, facilitation or inhibition of group or colony effects, nestmate recognition, and caste determination. Trophallaxis and exchange of food material also function as mechanisms for the transfer of chemical information between members of a eusocial insect colony. Sensory cues of many different types are used for communication in social insects, and chemical and acoustic cues have been discovered in termites and social Hymenoptera. Visual communication is limited but, according to Karl von Frisch, plays an important role in the waggle dance of honey bees, in which the bees indicate the location of a food resource through a series of stereotypical movements. Chemical communication is believed to be important for maintaining caste structure, nestmate recognition, and establishment of the queen in an insect society. Chemical signals have an immediate effect that typically ends immediately after the chemical has diffused away from the sensory receptors. Some chemical and acoustic cues have also been implicated for initiating defensive and flight behaviors in social insects.

Queen Dominance and Pheromones

Queen dominance in higher social insects is essential for controlling nestmates and is maintained with a variety of chemical cues in social Hymenoptera. In honey bees such as *A. mellifera,* the mandibular glands and Koschevnikov's gland produce the components of queen substance. In *A. mellifera,* mandibular gland secretion is a pheromone blend composed mostly of 8-hydroxyoctanoic, *(E)*-9-oxo-2-decanoic (9-ODA), *(E)*-9-hydroxy-2-decanoic, 10-hydroxydecanoic, and *(E)*-10-hydroxy-2-decanoic acids. As a newly emerged queen ages and her attractiveness to workers increases, the relative concentrations of these organic acids change. Worker attractiveness is maximized after the queen becomes mated. Although 9-ODA alone does not cause worker court formation, when 9-ODA is supplied with mandibular gland secretion, court formation occurs. The pheromone 9-ODA is also associated with the inhibition of queen-rearing behaviors and gonad development in worker bees. A possible mode of action of 9-ODA is to reduce the quantity of gonadotrophic hormone so that ovary development is inhibited in worker bees.

Court attraction, in which is a group of workers gather around and tend the queen, is controlled via glandular secretion. Koschevnikov's gland secretions from mated queens are correlated with court attraction. This gland plus mandibular gland secretions may serve to provide an olfactory stimulus for court formation from workers at short distances from the queen. Virgin queens do not form worker court aggregations as readily as do mated queens, which suggests that Koschevnikov gland secretions alone may not be responsible for court formation. Abdominal tergite gland secretions also work as a contact pheromone and facilitate court formation after workers touch the queen's abdominal tergites with their antennae.

Colony Odor

Nest odor is believed to be important for nestmate recognition in social insects. Sources of nest odor are not clearly defined, but a combination of genetic and environmental factors may be important sources of colony odor. Environmental odors from shared foods like nectar or pollen may contribute to the uniqueness of an individual colony's odor. Likewise, food transfer in honey bees may facilitate the spread of environmental odors to all members of the hive or nest. The waxy cuticle of honey bees may also trap environmental odors from the atmosphere of the nest. Cuticular hydrocarbons have been implicated as a means for nestmate recognition in *Polistes* wasps (Hymenoptera: Vespidae).

Queen odors are genetic factors that are believed to be a major component of colony odor. M. D. Breed provided some evidence to support this hypothesis. In 1981 Breed published some research findings suggesting that small groups of worker honey bees are more likely to accept a sister queen over an unrelated queen. Further research has demonstrated that resident bees are able to detect and recognize foreign queens raised under the same environmental conditions. This provides additional support for the presence of queen odor as a component of colony odor.

Trail Pheromones

Trail pheromones are used for recruitment, for marking pathways to resources, and for indicating resource richness. Often released with alarm pheromones, trail pheromones enable guards to aggregate around a nest invader. Furthermore, short-lived and persistent trail pheromones are known. In ants, trail pheromones are produced in the hindgut, Dufour's gland, the venom gland, and the tibial, tarsal, and abdominal glands. In some myrmecine and ponerine ants, the venom gland is the source of trail odors. Odor trails produced by termites originate

from the abdominal sternal glands. In *Zootermopsis nevadensis,* sternal gland secretions deposited on a trail from a breached or damaged nest wall to the interior of the nest draws workers toward the area for nest repair.

Odor trails emitted by forest-inhabiting bees provide three-dimensional information on resource location. Some large stingless bees of the genus *Trigona* produce strong odor chemicals from enlarged mandibular glands. *Trigona* forage in tropical forests and utilize mandibular gland secretions to cue in other foragers to nectar sources. Foraging *Trigona* will release more trail pheromone near a food source, thus indicating proximity to the food source for other foragers. The trail pheromone components of *T. subterranea* include both *E*- and *Z*-citrals. Differences in pheromone composition between species may reduce interspecific trail following in stingless bees.

Alarm Pheromones

Alarm pheromones are volatile organic molecules of low molecular weight that diffuse rapidly, forming a concentration gradient away from the signal source. Alarm pheromones initiate arousal, defensive, and assembly behaviors. Alarm pheromones are often not species specific in social Hymenoptera and termites, but the same pheromone often initiates different behaviors in different species. For example, the mandibular gland secretion 4-methyl-3-heptanone is associated with digging behavior in *Pogonomyrmex* species, but in high concentration causes repelling behaviors in the Texas leafcutting ant *Atta texana.*

Contextual responses to alarm pheromones are correlated with the proximity of the nest to the signal's source. Flight occurs when alarm pheromones are perceived far from the nearest nest opening. When these substances are released near the nest, colony members become defensive. Aggregation, defensive postures, and frenzied excitement are often observed in agitated social insects.

Alarm pheromones are often coupled with other glandular secretions such as trail pheromones. Most alarm pheromones originate from the exocrine mandibular gland. Mandibular gland secretions produce a variety of responses that include excitement, attraction, and threat posturing in different species of ants. Other glands associated with alarm signals in ants include Dufour's gland and the anal gland. In honey bees, isoamyl acetate is secreted by the lining of the sting pouch and generates attraction and investigation behaviors. After stinging, honey bees leave the sting and associated glandular components in the adversary's skin, which releases more volatile alarm pheromones for attracting more defenders to the vicinity. The mandibular gland in honey bees secretes 2-heptanone, which initiates alarm behaviors.

The cephalic gland secretions found in termite soldiers are believed to initiate aggressive excitement behaviors in other soldiers and alarm in nonsoldier nest members. Known organic compounds produced by termite cephalic glands include limonene, terpinolene, and α- and β-pinenes.

Acoustic Communication

Termites and ants also use some acoustic alarm signals, and some vespid wasps employ sound to alert offspring to the presence of provisions. Acoustic communication in social insects is conveyed through the substrate, not through the air. Acousticosensory organs present on the legs of social insects detect vibrations transmitted through the substrate.

Ants use subgenual organs and campaniform sensilla located on the legs to detect acoustic signals. Among carpenter ants, the subgenual organ is sensitive to frequencies ranging from 1.5 to 3 kHz. In the leafcutting ant, *Atta cephalotes,* campaniform sensilla function as a sound-detecting organ. Sensilla campaniformae are located near the distal end of the trochanter in *A. cephalotes.* The sensitivity of campaniform sensilla to sound waves varies from the anterior to posterior legs, with the anterior legs being the most sensitive to vibration.

Acoustic signals in ants are generated by stridulation or by tapping the gastral segments on the substrate. Stridulatory files are found universally in pseudomyrmecines, in over 80% of all studied myrmecines, and in nearly half of all studied ponerine ants. Stridulatory files when present are located at the junction of the abdomen (fourth abdominal tergite) with the postpetiolus. Buried or confined ants use stridulatory files to generate acoustic signals through the substrate for initiating digging behaviors in other workers.

Formicine ants are believed to have secondarily lost their stridulatory files and none use different forms of auditory communication. *Campanotus* (carpenter ants) use a combination of mandibular and gastral tapping when disturbed. Response to these alarm cues depends on the state of arousal in the signal receiver. In some cases, the tapping behavior causes some ants to lie motionless, presumably to make them less visible to predators, while more agitated ants orient toward the source of the signal and approach it. Outside the nest, tapping enhances the stimuli associated with disturbance and functions as a danger alarm.

Similar tapping behaviors have also evolved in some vespid wasps but not as an alarm cue. Gastral vibration is believed to signal food or provisioning to larvae. The abdomen or gaster is rapidly tapped on the surface of the comb when provisions are brought into the nest. In *Vespa tropica,* foundresses tap their legs on the nest comb to signal the presence of provisions to hungry larvae. In this species, the sounds produced by leg tapping are loud and audible to humans standing at least 1 m from the nest.

In termites, sounds are generated by rapidly tapping the head or abdomen against the substrate. Head banging in *Zootermopsis* is audible as a rustling sound and is believed to function either as a substrate-transmitted alarm call or as a defensive behavior meant to scare intruders away from the nest. Research on the response of termite subgenual organs to the vibrations generated by head tapping behavior (1 kHz) suggest that these acousticosensory organs are highly sensitive to that frequency.

INQUILINES AND NEST PARASITES OF SOCIAL INSECTS

Eusocial insect nests often harbor guests that take advantage of the food resources and microclimates inside the nest chambers. Arthropod ectosymbionts of social insects are numerous and include some species of mites, spiders, millipedes, isopods, and insects. Most orders of insects contain some ectosymbiotic species. Some nest invaders, such as wax moths (Lepidoptera: Pyralidae), feed on the nest structure of honey bees, whereas others may feed on the young, stored provisions, or scavenge from the debris.

Some nest owners receive a benefit (mutualism) for harboring another species. Trophobionts like aphids (Heteroptera: Aphididae), scale insects (Heteroptera: Coccidae), and mealybugs (Heteroptera: Pseudococcidae) are often protected and sheltered by ants. In return, ants receive from these insects honeydew, a sugary fluid, while providing protection and shelter to these guests. Some have suggested that the posterior morphology of aphids mimics the head region of an ant, with honeydew secretion simulating the transfer of food between ant workers.

A diverse group of guests is usually encountered with army ant colonies. R. D. Akre and C. W. Rettenmeyer studied the association between ectosymbiotic staphylinid beetles and army ants. These beetles have evolved two distinct behavioral strategies to gain access to the resources provided by army ant colonies. Specialized species such as *Ecitomorpha* and *Ecitiosus* are mimetic forms that live within the bivouacs of army ants. They are highly integrated into army ant societies and die if removed from the confines of the colony. Often associated with the larvae or booty, these staphylinid beetles tend to move together within an emigrating column of ants. Phoresy is common, with army ant workers transporting and moving these staphylinids to new bivouacs. *Microdonia* uses a generalized strategy for exploiting army ants. The association between generalists and their host ants is less integrated than specialized species. They usually frequent the periphery of an army ant colony and will often attack wounded or dying ants. The ants will sometimes attack and attempt to drive these insects away from the nest area.

Evolved mechanisms that enable symbiotic associations between symphiles and their hosts may include behavioral and chemical mimicry. Many symphiles can attract hosts and through appeasement gain access to the nest. B. Hölldobler was one of the first researchers to discover that scents are important for attraction of worker ants to larval Staphylinidiae. Some researchers believe that trichome secretions have an intoxicating effect and cause disorientation in host species. The reduviid bug *Ptilocerus ochraceus,* an ant predator, uses a similarly acting substance to attract and paralyze *Dolichoderus* ants. This bug presents its abdominal trichomes to an ant, which begins to lick the trichome hairs leading to intoxication.

The variety of interactions between different species of ants is extensive. Interactions range from coexistence to true parasitic relationships. In plesiobiosis, distantly related ant species can coexist with minimal interactions between closely neighboring nests. Some ant species rely on cleptobiosis or thievery to acquire food or refuse from nests of ants of other species. Cleptobiotic species reside in separate domiciles from their hosts. Lestobiosis differs from cleptobiosis in that one species will actually invade the host's nest to prey on the young or food cache. For example, ants of the genus *Carebara* nest within the walls of termite nests and are believed to prey on termites.

Mixed species colonies usually involve some form of social parasitism. Temporary social parasites were first recognized by Wheeler in 1904. Ants in the *Formica microgyna* group utilize this type of life history strategy. A newly fertilized queen enters a host colony and coerces the host workers to take over the nest through assassination of the host queen. Gradually, the parasitic queen's offspring replace the host workers through attrition.

Slave-making ants invade a host nest and steal pupae for incorporation in their own nests. After eclosion, the slaves work as foragers and as nest constructors, and conduct brood care for the slave makers. Slave makers parasitize other ant species that are close relatives. Some species of *Polyergus* use *Formica* species for slaves, *Formica* species of the *sanguinea* group typically use other *Formica* species as slaves.

Permanent parasitism or inquilinism occurs entirely inside the host's nest. In some species, workers are present but have limited behavioral roles within the host's nest. For example, *Teleutomyrmex schneideri* no longer has a worker caste and represents a highly refined form of inquilinism. *T. schneideri* lives its entire life within the nest of *Tetramorium caespitum,* a close relative of *T. schneideri.* The queens of *T. schneideri* are smaller than the queens of *T. caespitum* and are morphologically patterned to ride on the backs of the host queen. *Teleutomyrmex* queens release attractants, from cuticular glands of the thorax and petiole and are eagerly tended by the host's workers.

See Also the Following Articles

Caste • Colonies • Division of Labor • Hymenoptera • Isoptera • Parental Care • Pheromones • Sex Determination

Further Reading

Akre, R. D., and Rettenmeyer, C. W. (1966). Behavior of Staphylinidae associated with army ants (Formicidae: Ecitoninii.) *J. Kans. Entomol. Soc.* **39(4),** 745–782.

Aoki, S. (1987). Evolution of sterile soldiers in aphids. *In* "Animal Societies: Theories and Facts" (Y. Ito, J. L. Brown, and J. Kikkawa, eds.), pp. 53–65. Japan Science Society Press, Tokyo.

Breed, M. D. (1981). Individual recognition and learning of queen odors by worker honey bees. *Proc. Natl. Acad. Sci. USA* **78,** 2635–2637.

Cocroft, R. B. (1999). Offspring–parent communication in a subsocial treehopper (Hemiptera: Membracidae: *Umbonia crassicornis*). *Behaviour* **136(1),** 1–21.

Crespi, B. (1992). Eusociality in Australian gall thrips. *Nature* **359,** 724.

Eickwort, G. C. (1981). Presocial insects. *In* "Social Insects," Vol. II. (H. R. Hermann, ed.), pp. 199–280. New York, Academic Press.

Evans, H. E. (1977). Extrinsic versus intrinsic factors in the evolution of insect society. *BioScience* **27,** 613–617.

Frisch, K. V. (1967). Honey bees: Do they use direction and distance information provided by their dancers? *Science* **158,** 1072–1076.

Hamilton, W. D. (1964). The genetical evolution of social behavior, I and II. *J. Theor. Biol.* **7,** 1–52.
Hölldobler, B., and Wilson, E. O. (1990). "The Ants." Belknap Press of Harvard University Press, Cambridge, MA.
Ito, Y. (1989). The evolutionary biology of sterile soldiers in aphids. *Trends Ecol. Evol.* **4,** 69–73.
Kent, D. S., and Simpson, J. A. (1992). Eusociality in the bark beetle *Australoplatypus incompertus* (Coleoptera; Curculionidae). *Naturwissenschaften* **79,** 86–87.
Lin, N., and Michener, C. D. (1972). Evolution of sociality in insects. *Q. Rev. Biol.* **47,** 131–159.
Michener, C. D. (1974). "The Social Behavior of Bees: A Comparative Study." Harvard University Press, Cambridge, MA.
Roberts, R. B., and Dodson, C. H. (1967). Nesting biology of two communal bees, *Euglossa imperialis* and *Euglossa ignita* (Hymenoptera: Apidae), including description of larvae. *Ann. Entomol. Soc. Am.* **60(5),** 1007–1014.
Wheeler, W. M. (1904). A new type of social parasitism among ants. *Bull. Am. Mus. Nat. His.* **20(30),** 347–375.
Wheeler, W. M. (1923). "Social Life Among the Insects." Harcourt Brace, New York.
Wilson, E. O. (1971). "The Insect Societies." Belknap Press of Harvard University Press, Cambridge, MA.

Soil Habitats

Patricia J. Vittum
University of Massachusetts

Many basic and applied studies in insect ecology have considered responses of insect populations to their physical or chemical environment. For insects that live above ground, the mechanisms of behavioral responses to environmental factors often are directly observable. However, behavioral responses of soil-inhabiting insects are much more difficult to observe and quantify. Soil texture and structure can have a direct impact on arthropod behavior and adaptations. Field studies of soil insects often quantify the consequences of behavior, whereas the actual behaviors are only inferred. There have been studies of ecological and physiological adaptations of several soil arthropods that are not considered to be agricultural pests but have certain life history features that make them amenable to investigation. However, most insects that are agricultural pests as soil-inhabiting immature forms (e.g., wireworms or rootworms in corn, scarab grubs in turf grass) often have been studied in detail only in their more accessible, or observable, adult form. This is in part because of the logistical challenge of observing the movements and other behaviors of the soil-bound larvae.

CHARACTERIZATION OF THE SOIL ENVIRONMENT

Soil Solids

Most soils have a complex structure, consisting of solids, liquids, and atmospheric gases. The solid components (e.g., sand or clay) constitute much of the soil matrix in bulk. Inorganic, or mineral, particles range in size from clays (< 0.002 mm in diameter) to silts (0.002–0.05 mm) to sands (0.05–2 mm) to pieces of gravel (> 2 mm). The proportion of different sized particles, or soil texture, determines to a large extent the physical properties and appearance of a soil, as well as its ability to supply chemical nutrients to plants. The soil texture also has an impact on the soil arthropod population. For example, small soil arthropods sometimes find it difficult to move in heavy clay or tightly compacted soils.

Soil organic matter consists of an accumulation of partially disintegrated and decomposed plant and animal residues that have been broken down and resynthesized by microorganisms in the soil. Although organic matter usually does not constitute more than 6% by weight in topsoil (and even less in subsoils), it binds mineral particles into slightly larger granules that produce loose, crumbly soils that can hold more water than their mineral counterparts. Organic matter also is a primary source of energy for a variety of soil organisms, including many arthropods.

Pore Spaces

A typical loam soil will consist of roughly 50% soil solids (a combination of sand, silt, and clay) and 50% pore spaces and water. The size and distribution of pore spaces will depend on the size and shape of the mineral particles, as well as the activity of microorganisms. A predominantly clay soil will usually have very small pore spaces because the clay particles are very small and can pack together effectively. A sandy soil will tend to have much larger pore spaces because the sand grains are more irregular in shape and do not compact as readily. Atmospheric gases (most notably, oxygen and carbon dioxide) also occupy pore spaces and can move passively through the soil profile, depending on surface conditions.

Pore spaces are further characterized as micropores (< 0.06 mm) and macropores (> 0.06 mm). Macropores tend to allow movement of air and percolating water very readily, whereas micropores are the first to be filled with water in a moist field soil and do not permit much air movement into or out of the pores.

Convection can enhance the exchange of gases within a soil or between a soil and the atmosphere above. Aeration (movement of oxygen and other gases) near the surface of a soil occurs most readily in the presence of large, interconnected pore networks or channels. The rate of aeration is influenced by changes in barometric pressure, temperature gradients, and wind gusts. Temperature, relative humidity, surface texture, and continuity of soil pores affect the diffusion of gases into and out of soil.

Soil Water

Water (and dissolved minerals) accumulates in pore spaces and moves vertically through the soil profile if the surface input (e.g., rain or irrigation) exceeds the rate at which any vegetation absorbs the water. Water moves most readily through soils that have well-spaced, interconnecting macropores, but it also fills

many of the micropores closest to the surface. Since water displaces air in the pore spaces, soils that are saturated with water cannot retain atmospheric gases that are critical to plant growth.

Soil moisture tension determines how much water remains in the soil at equilibrium and is a function of the sizes and volumes of pore spaces (matric potential), the presence of solutes in the soil (osmotic potential), and gravity. When the surface of a soil dries following an extended period of dry weather (further enhanced by low humidity or steady winds), water can also be drawn back toward the surface through capillary action.

The relative concentrations of oxygen, carbon dioxide, and water vapor often are considerably different in soil pore spaces (soil atmosphere) and in the open air. Oxygen concentrations tend to be lower and CO_2 concentrations tend to be higher in the soil atmosphere as a result of plant and animal respiration and biochemical soil processes. Relative humidity tends to be relatively high, particularly in soils used to produce agricultural crops.

Soil Temperature

Surface temperatures of soils often fluctuate at least as much as ambient air temperatures, but the difference between the daily maximum and minimum temperature decreases as depth in the soil increases. There is a delay of maximum and minimum soil temperatures compared with the overlying surface temperatures, correlated with depth. Seasonally the maximum and minimum temperatures occur in the warmest and coolest seasons, respectively, in the upper soil layers. In contrast, the highest temperatures occur in early winter and the lowest temperatures occur in midsummer, at 7 m depth. Soil insects and other arthropods often move vertically in direct response to soil temperature, moving downward in late autumn to avoid freezing temperatures on the surface and returning to the surface in the spring to resume feeding.

FORMS OF LIFE IN THE SOIL

A vertical cross section of a soil profile reveals several distinct layers. The lowest layer, or horizon, is the C horizon, which consists of unweathered rock. The B horizon contains weathered, rough mineral soil with small deposits of humus. The A horizon normally has fine mineral soil interspersed with organic matter. The O horizon is a layer of plant debris lying on the surface of the mineral soil. The thickness of this layer depends on the amount of vegetation that is deposited seasonally and annually, and on the amount of degradation that occurs as a result of soil organisms. Within the O horizon are several layers, including (from the top) leaf litter, a fermentation layer, and a humus layer. The humus layer often merges into humus-enriched topsoil.

Soils and the overlying organic layer are not homogeneous, but rather are stratified. Similarly, the arthropods that live in these regions are grouped into different life-forms that have adapted to the various conditions that exist in the soil.

Euedaphons

Euedaphic soil arthropods inhabit the lowest soil layers, generally moving within the soil pore system. These arthropods usually are small and are characterized by a round or wormlike body form. The body size corresponds to that of the pore system, and extremities are often reduced. Because they cannot escape predators, many euedaphons generate and release toxic or defensive substances. Most euedaphons are photophobic and either lack eyes or have eyes that have degenerated considerably. They tend to have well-developed mechano- or chemosensitive organs, which compensate for their poor or nonexistent vision. Arthropods occurring in the euedaphon include several species of proturans, diplurans, and symphylans, as well as a few oribatid mites.

Epedaphons

Epedaphic arthropods live on the soil surface and in leaf litter. They are not well adapted to the conditions found in the soil pore system (e.g., high relative humidity, restricted gas exchange, restricted mobility). They are represented by many different body forms, usually are strongly pigmented, and often are dorsoventrally flattened. They have well-developed sensory organs and are highly mobile. Arthropods occurring in the epedaphon include oribatid mites (Oribatei), springtails (Collembola), ectobiid cockroaches, several cricket species, and several predatory beetles, including rove beetles (Staphylinidae) and ground beetles (Carabidae).

Hemiedaphons

The hemiedaphon group represents a transitory form of life, enabling some epedaphic or atmobiotic arthropods to occupy burrows in the soil. Hemiedaphic arthropods have the ability to excavate through soil by means of modified mouthparts or fossorial legs, and often they can enlarge existing cracks and pores. Several insect taxa have adopted a hemiedaphic life for a variety of reasons: to dig channels and then wait for surface prey to fall into the pit, to burrow through the soil hunting for small epedaphic arthropods, to avoid temperature or moisture extremes on the surface, or to feed on roots of plants. Hemiedaphic arthropods include earwigs (Dermaptera), field crickets and mole crickets, tiger beetles (Cicandellidae), and white grubs (scarab larvae).

ADAPTIVE STRATEGIES OF SOIL ARTHROPODS

Response to Soil Texture

Organisms in the soil (and leaf litter) community play very different roles, based in part on their size. Organisms (e.g., protozoa, bacteria, and some nematodes) that exist in water films, often in soil micropores, have resource requirements and defense needs that differ from those of organisms able to move in and out of soil pores independently. Similarly, a soil

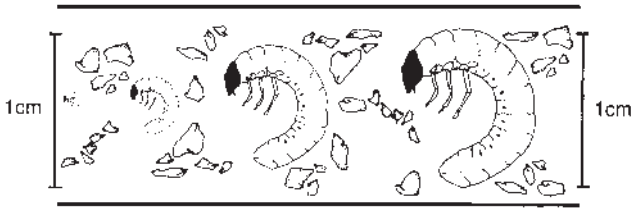

FIGURE 1 Schematic representation of the impact of soil particle size on movement of a soil insect. Japanese beetle larvae (neonate, late first instar, mid second instar, and mid third instar) shown in a square centimeter of typical loamy sand soil. [Drawing by Stephen L. Thomas, University of Massachusetts, adapted from Villani, M. G., and Wright, R. J. (1990). Environmental influences on soil macroarthropod behavior in agricultural systems. *Annu. Rev. Entomol.* **35,** 249–269.]

macroarthropod, as it grows, will perceive the soil matrix differently (Fig. 1). A neonate exists functionally as a microarthropod, able to move only within existing pores in the soil. Thus its ability to move is a function of the porosity of the soil (including the size of the pores and their continuity). As the arthropod grows, less of the pore space is available for free movement. At this point pore space is less important in impeding movement than gross soil structure (impacting insect movement among soil aggregates) and aggregate density (movement through the aggregates). Plant root activity, surface cover, traffic and other sources of compaction, and density of soil organisms all impact aggregate formation. As the insect grows, less of the soil pore space is available to the insect for free movement, but it may take advantage of preexisting soil channels created by soil macrofauna such as earthworms, other arthropods, or small vertebrates. Water-filled soil pores can inhibit movement.

Response to Temperature

Several soil insect species demonstrate a seasonal pattern of vertical movement associated with soil temperature. In temperate climates, many soil macroarthropods move downward in late autumn to avoid freezing and return to the upper soil layers in spring. Some species move away from the soil surface in the middle of summer, in part to avoid high soil temperatures. Species-specific responses to soil temperatures influence periods of feeding activity and may enable similar species to occupy slightly different niches (e.g., wireworms, scarab grubs).

Response to Moisture

Some hemiedaphic arthropods have developmental stages that cannot tolerate extremes in moisture, and many of those stages are very sensitive to soil moisture levels. For example, grubs of several scarab species that had been held in dry soils moved upward almost immediately after moisture was applied to the surface. Several macroarthropods, including some wireworm species, alter the soil environment by creating semipermanent earthen cells. These temporary cavities enable the insects to create nearly saturated chambers in the soil, greatly reducing moisture loss from evaporation. Numerous desert arthropods create earthen cells in which they can aestivate when soil moisture is extremely low (or temperature is extremely high).

Usually, eggs and pupae of insects are most resistant to moisture loss and least able to escape undesirable conditions. Larvae and adults are often mobile stages and may be able to move away or alter their behavior to minimize the impact of adverse moisture extremes. For example, scarab larvae can move downward in the soil profile to seek moister (and cooler) conditions during periods of heat or drought stress. Heavily sclerotized soil arthropods may be less vulnerable to cuticular moisture loss than are less sclerotized forms, such as grubs or maggots.

Locomotion

Edaphic arthropods move through soil to locate food, to escape predators, or to escape unfavorable abiotic conditions. Adaptations for movement in soil depend on soil type, particle size, pore size, and soil density, among other things. Although many epedaphic species use their bodies as wedges, euedaphic species tend to dig through soil, using their legs and mandibles as shovels. Legs of edaphic insects often are highly modified to facilitate digging or burrowing through soil. At least one leg segment is likely to be enlarged, specially shaped, and edged with spines or lobes to create functional spades. Expanded tibiae or femora provide increased surface area, providing improved leverage when the insect is moving soil. For example, mole crickets (Orthoptera: Gryllotalpidae) use their greatly enhanced fossorial forelegs to burrow through soils very rapidly.

Many euedaphic arthropods lack obvious modifications for digging, but are able to move through soil by fitting between particles or by moving in existing burrows or crevices. Body shapes may be flattened or cylindrical, but in general antennae and legs tend to be reduced or absent. Millipedes serve as an example of the variety of adaptations that have arisen over time. Some species act as bulldozers with long bodies, many legs (for purchase against soil particles), and broad heads; other species have shorter bodies, fewer but longer legs, and tapered heads that allow the millipedes to wedge through small spaces in the leaf litter or upper layers of soil. Still other species of millipedes have very tapered anteriors and compressible bodies that can be used to widen crevices. The heavily sclerotized elytra and terga on adult beetles can provide a very effective protective shield when an insect is pushing aside soil particles or layers of leaf litter.

Host Finding

Much of the behavioral research that has been conducted over the past 50 years has centered on atmobiotic insects, in part because until recently it was virtually impossible to observe soil insects *in situ* without disturbing them. Destructive sampling techniques enabled researchers to make quantitative assessments but revealed very little information about how or why soil insects exhibited certain behaviors.

For many hemiedaphic insects, host finding begins with choices made by the mobile adult female as she seeks oviposition sites. For example, a female corn rootworm beetle *(Diabrotica)* may oviposit in a field where corn is growing, but if the field is planted to a different (nonhost) crop the following growing season, the neonates that emerge in the spring may have to move relatively large distances to reach a suitable host. Similarly, emerging scarab larvae must locate suitable roots and begin feeding within 24 to 48 h of eclosion, and overwintering grubs must relocate to suitable host plants when returning to the upper soil layers in the spring.

There are many plant-derived chemicals that elicit responses in insects in general, including host plant extracts that initiate host-searching behaviors or avoidance mechanisms. Many of the most intensively studied phytochemicals are produced in leaves or stems, but soil insects are more likely to respond to chemicals produced in the root zone. Some of these compounds are quite specific and elicit responses (e.g., host seeking) from limited taxa. Others, like carbon dioxide, are not species specific and influence a wide range of soil insects. Chemicals produced in the soil (typically as root extracts) may diffuse over relatively large distances, but the diffusion rate depends on soil moisture, texture, and compaction.

Defensive Adaptations

Edaphic arthropods produce a variety of compounds that can function as contact toxins, repellents, or irritants. Many different taxa use similar biosynthetic pathways to produce closely related compounds. For example, several soil arthropod groups, including some millipedes and centipedes, as well as some chrysomelid larvae, produce hydrogen cyanide. The glands that produce this nonselective toxin apparently are not homologous, suggesting that the capability has evolved more than once.

Some edaphic arthropods have developed chemical defenses against deep-soil predators that use mechanoreceptors and chemoreceptors to provide cues to locate their prey. Others evade predators by running or jumping. Epedaphic Collembola (springtails) that live in leaf litter are cryptically colored and have a highly evolved mechanism that enables the insect to jump away from a disturbance virtually instantaneously. In contrast, Collembola that live wholly within the soil (and thus are constrained by soil particles) tend to be smaller and paler, and to have less well developed jumping mechanisms than epedaphic species. Instead, the euedaphic springtails secrete noxious fluids that protect them against many predators.

Interactions of Soil Insects with Chemical Control Agents

Edaphic insects that are considered to be pests in production agriculture or the green industry are often much more difficult to "control" than their atmobiotic counterparts. One of the greatest challenges is achieving adequate contact of an insecticide with the target insect. Many insecticides dissipate or degrade relatively quickly (before they reach the soil) or are adsorbed to soil particles. Most soil insecticides (and other pesticides) remain in the top 5 to 10 cm of the soil, where they are subject to microbial degradation. Soil factors such as pH, organic matter, moisture, temperature, and microbial community diversity will have a direct impact on the mobility and persistence of a soil insecticide. Insecticides that are highly mobile in soil may be ineffective because they move beyond the target zone too rapidly.

Many soil insects can detect the presence of insecticides and other chemicals and will initiate avoidance behavior (e.g., moving away from the soil zone in which the chemical is detected). Abiotic factors, such as soil moisture or temperature, that induce a target insect to move as little as 1 cm further into the soil profile may place target insects beyond the effective "range" of some chemical control agents. Often the manipulation of irrigation apparatus or the use of application equipment that incorporates a control agent directly into the soil at the desired depth can enhance the efficacy of a soil insecticide.

IMPACT OF SOIL ON PATHOGENS, PREDATORS, AND PARASITOIDS

Pathogens

The upper layers of most soils, as well as leaf litter, support active microbial communities. Many of these organisms are natural decomposers, breaking down plant and animal tissues that ultimately become part of the organic matter in the underlying soil. Interestingly, many edaphic arthropod populations are nearly free of disease. Several studies have demonstrated that secretions from many different edaphic arthropods inhibit vegetative growth or suppress the germination of pathogenic organisms.

Edaphic arthropods that are highly susceptible to pathogenic organisms in laboratory tests are rarely infected in the field, suggesting that there may be a critical behavioral component. For example, several soil insects, including earwigs (Dermaptera), mole crickets, and ants, tend and lick eggs. This behavior may remove fungal spores or bacteria or inhibit their germination. In addition, some laboratory studies utilize artificially dense populations of the target arthropod, which may enhance the spread of pathogens from one organism to another.

Nevertheless, several insect pathogens occur naturally in soils. For these pathogens to induce an epizootic, three conditions must be met. A susceptible host must be present, with host susceptibility being governed by population density, species composition, presence or absence of other stress factors, and behavioral responses to the pathogen. A pathogen must be present, with a suitable level of virulence and persistence. Finally, the environment must support both the host and the pathogen. Soil conditions, such as low temperature or high moisture levels, may stress the target insect population, predisposing individuals to infection and ultimately leading to population declines.

Several microbial insecticides were identified and developed during the last half of the 20th century. All were found in naturally occurring epizootics in the soil and were subsequently commercialized. Microbial insecticides are passively mobile because an insect that comes in contact with the microbial product may move some distance from the initial point of contact before it dies. The behavior of the target insect can be important to the spread of the pathogen, particularly when normal (or pathogen-induced) behaviors result in movement of individuals beyond their normal range. Most microbial insecticides are able to replicate within the host.

In some instances, edaphic arthropods are able to detect and avoid fungal pathogens in soil. In 1994 Villani *et al.* conducted a series of microcosm studies proving that when incorporated into soil, mycelial formulations of *Metarhizium anisopliae,* a naturally occurring soil fungus, repell third-instar Japanese beetle *(Popillia japonica)* grubs for as long as 3 weeks. Similar responses have been observed with tawny mole crickets and subterranean termites.

Examples of microbial insect pathogens include bacteria (e.g., *Bacillus popilliae, B. thuringiensis, Serratia marcescens, S. entomophilas*), fungi (e.g., *M. anisopliae, Beauveria bassiana, Fusarisum* spp., *Penicillium* spp., and *Aspergillus* spp.), protozoa (e.g., *Ovavesicula popilliae*), and various rickettsia (bacteria-like) organisms. Many of these have been developed commercially, with varying degrees of success.

External factors relating to soil condition may modify arthropod behavior. Localized flooding may force edaphic insects to move to unsaturated soils, whereas after several weeks without rain these insects may seek moister (usually lower) ground. These localized migrations may result in contacts between populations of edaphic insects and pockets of pathogen activity that otherwise might not have occurred. Disruptions in social behavior may increase or decrease infection rates. For example, studies have demonstrated that applications of sublethal rates of certain soil insecticides, such as imidacloprid, greatly increase the pathogenicity of some pathogens, including *B. bassiana* and some entomopathogenic nematodes (see next section). Apparently the sublethal exposure of the insecticide disrupts normal social behavior in the termite colony, including grooming, trophallaxis (food exchange), and construction of tunnels.

Predators and Parasitoids

Insect predators are mobile and self-replicating. Their effectiveness depends on the interaction of the soil environment with both the agent and the host. The initial contact and subsequent spread through the population depend on temporal and spatial overlap. The predator must be well adapted to the soil conditions. In particular, the predator must be able to move through the soil and to respond to host cues. Some predators are edaphic arthropods, such as predatory beetles or spiders.

Entomopathogenic nematodes can also be considered to be predators, since they move through the soil in search of host insects. Some species move actively, whereas others are more passive, ambushing prey as it moves nearby. Entomopathogenic nematodes enter the host through natural openings, such as the mouth, spiracles, or anus. Nematodes in the genera *Heterorhabditis* and *Steinernema* carry a pathogenic bacterium, which is released in the body cavity of the host. The bacteria multiply within the host and produce toxins that kill the host rapidly. Although several nematodes species are available commercially, their efficacy in field conditions has been inconsistent, in part because the nematodes are extremely sensitive to temperature and soil moisture levels.

Several insects have evolved as parasitoids of edaphic arthropods. Host finding is presumed to be much more complex than for atmobiotic hosts because any host volatiles must move through the soil matrix. The parasitoid also must be able to move through the soil structure to reach the host. Studies attempting to evaluate the specific host-finding behaviors of these parasitoids must include consideration of soil texture, moisture, and temperature, as well as root zone exudates and sensitivity to movement-induced vibrations.

SAMPLING TECHNIQUES FOR SOIL INSECTS

One of the greatest challenges facing soil insect ecologists is the need to develop techniques of following insect movement and feeding behavior *in situ* while minimizing disturbance of the soil system. Until recently, researchers relied on the "snapshot" approach of collecting soil samples and sampling destructively, to determine how many soil insects might be present in the sample. This approach provided limited quantitative assessments but could not provide information over an extended period of time.

One technique that seems to be ideally suited to observing soil insect behavior is radiography, which has been used to trace the movement of scarab grubs and mole crickets in turfgrass, wireworms in corn, and onion maggots in onions, among other arenas. Plastic boxes of varying dimensions are filled with soil, and virtually any edaphic arthropod can be introduced to the microcosm and observed without disturbance. The technique has been expanded since Villani and Wright first described some of the applications in 1988. It has been used to investigate the response of soil insects to the presence of pathogens, to study the movement of two different species in a confined space, and to conduct preliminary basic observations of species behavior in soil. Understanding of insect movement in soil has expanded tremendously as a result of radiographic observations.

See Also the Following Articles

Collembola • Japanese Beetle • June Beetles • Isoptera

Further Reading

Brady, N. C. (1990). "The Nature and Properties of Soils." 9th ed. Macmillan, New York.

Eisenbas, G., and Wichard, W. (1987). "Atlas on the Biology of Soil Arthropods." Springer-Verlag, Berlin.

Hillel, D. (1980). "Fundamentals of Soil Physics." Academic Press, New York.
Villani, M. G., and Wright, R. J. (1988). Use of radiography in behavioral studies of turfgrass-infesting scarab grub species (Coleoptera: Scarabaeidae). *Bull. Entomol. Soc. Am.* **34,** 132–144.
Villani, M. G., and Wright, R. J. (1990). Environmental influences on soil macroarthropod behavior in agricultural systems. *Annu. Rev. Entomol.* **35,** 249–269.
Villani, M. G., Krueger, S. R., Schroeder, P. C., Consolie, F., and Consolie, N. H. (1994). Soil application effects of *Metarhizium anisopliae* on Japanese beetle (Coleoptera: Scarabaeidae) behavior and survival in turfgrass microcosms. *Environ. Entomol.* **23,** 502–513.
Villani, M. G., Allee, L. L., Diaz, A., and Robbins, P. S. (1999). Adaptive strategies of edaphic arthropods. *Annu. Rev. Entomol.* **44,** 233–256.

Sound Production

see *Hearing*

Spermatheca

Marc J. Klowden
University of Idaho

The spermatheca is a special pouch in the female in which spermatozoa are stored and maintained after mating. In more advanced insects, the male deposits seminal fluid or a spermatophore in the bursa copulatrix, and the sperm move from the bursa into the spermatheca. The sperm often remain viable within the female's spermatheca for her entire life, which may be as long as 3 to 4 years for social insects such as the honey bee queen. The sperm are released from the spermatheca only when eggs pass down the oviduct so fertilization may occur just before the eggs are laid.

SPERMATHECAL STRUCTURE

A female insect typically has only one spermatheca, but some coleopterans have two, and there are three spermathecae in some dipterans. The spermathecae vary considerably in their overall structure from insect to insect, but as outgrowths of the median oviduct, they are all ectodermal in origin and produce a cuticular lining. They generally arise from the median oviduct near or on the genital chamber (Fig. 1). The spermathecal sac (or receptaculum seminis) is connected to the genital chamber by a secretory duct (or ductus seminalis) through which the sperm are discharged. There also may be specialized secretory cells associated with the capsule that produce secretions that maintain the viability of the sperm and activate them. The sperm move quite actively within the spermathecal sac, and the secretions of the secretory cells are assumed to provide nourishment to them. Secretions of the spermatheca may also be directed into the hemocoel. In the blood-sucking bug, *Rhodnius prolixus,* a factor produced by the sperm-filled spermatheca is responsible for the increased rate of egg development that is associated with mating.

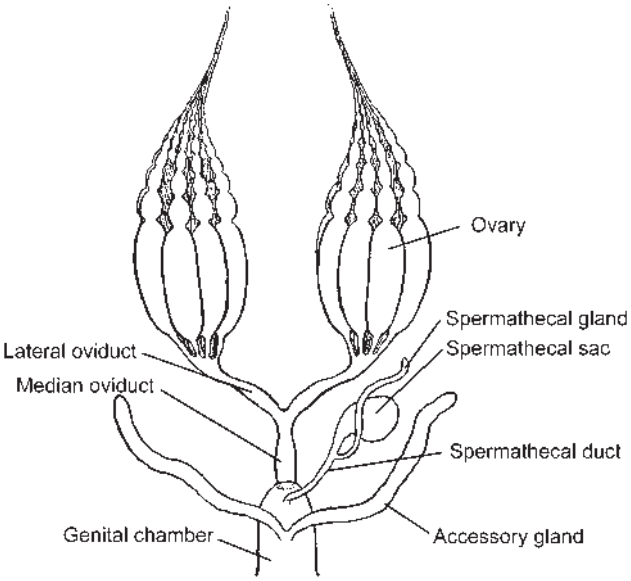

FIGURE 1 The generalized structure of the female reproductive tract, showing the location of the spermatheca. [After Snodgrass, R. E. (1935) "Principles of Insect Morphology." McGraw-Hill, New York.]

UTILIZATION OF SPERM

The spermatheca may be innervated by nerves that branch from the terminal abdominal ganglion and control the release of sperm. When sensory receptors on the oviduct are activated by the presence of an egg, a reflex arc through the terminal abdominal ganglion activates a motor neuron that triggers the contraction of muscles surrounding the spermathecal sac. Sperm are squeezed through the spermathecal duct to the oviduct, where the egg is fertilized before it passes out of the body.

There may not always be a random utilization of sperm present in the female spermatheca. When sequential matings occur, although the sperm from different males may be mixed within the storage organ, the sperm from one mating may be used preferentially. Most often there is precedence in the utilization of sperm so that the last sperm to enter are the first ones to leave, increasing the probability that a second mating will be used to fertilize the eggs. To ensure paternity in second matings to an even greater degree, male insects have evolved some novel mechanisms of sperm displacement. For example, the male of the damselfly, *Calopteryx maculata,* uses its scooplike penis both to transfer its own sperm to the female and also to remove from the spermatheca sperm from a previous mating. An unusual method of insemination and sperm storage occurs in some members of the heteropteran superfamily Cimicoidea, which includes the bedbugs. Males

puncture the cuticle of the female, injecting the sperm into a specialized tissue mass there. The sperm may then move to the ovaries or remain within the tissue mass for storage, as if in a spermatheca.

See Also the Following Articles

Accessory Glands • Reproduction, Female

Further Reading

Clark, J., and Lange, A. B. (2000). The neural control of spermathecal contractions in the locust, *Locusta migratoria. J. Insect Physiol.* **46,** 191–201.

Davey, K. G. (1985). The female reproductive tract. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology," Vol. 1. (G. A. Kerkut and L. I. Gilbert, eds.), pp. 15–36. Pergamon Press, Oxford, U.K.

Okelo, O. (1979). Mechanisms of sperm release from the receptaculum seminis of *Schistocerca vaga* Scudder (Orthoptera: Acrididae). *Int. J. Invertebr. Reprod.* **1,** 121–131.

Spermatophore

Marc J. Klowden
University of Idaho

In the more primitive insect groups, males synthesize specialized sperm carriers called spermatophores to protect the sperm during their transfer to the female genital tract. Insects are believed to have descended from aquatic ancestors whose males released their sperm directly into the water. In adaptation to a terrestrial existence, however, a more precise method of mating was required to prevent the sperm from desiccating. The spermatophore thus represents an initial adaptation for life on land that protects the male gametes until they are within the female reproductive tract. In more advanced insects, the sperm are transferred directly in seminal fluid and the spermatophore is no longer produced.

SPERMATOPHORES IN APTERYGOTE INSECTS

The insect spermatophore is synthesized by the male accessory glands. It consists of a viscous secretion that is shaped by the internal structures of either the male or the female once it has been inserted into the female reproductive tract. A change in pH may account for the transition from a liquid secretion by the male accessory glands to a more solid, gel-like mass. The sperm are contained within the saclike ampulla of the spermatophore before it is transferred to the female.

The structure of the spermatophore and the mechanism of transfer differ among the insect orders. In many primitive apterygote insects, such as the collembolans, the spermatophore consists of a drop of sperm placed in a simple sac at the end of a stalk on the ground. Females must find the spermatophore themselves and actively take it up into the genital tract. The behavior is more complex in other species of Collembola, where the male may actively manipulate the female and direct her to the spermatophore he has already deposited. These males have not developed the modifications necessary to grasp the female for conventional mating. Similarly, some Thysanura males deposit an unstalked spermatophore on the ground, but place a wall of silken threads around it to signal to the female that the spermatophore is nearby. Because all these insects inhabit humid soil environments, there is little chance for the sperm to dry out before the females find them and take them up.

SPERMATOPHORES IN PTERYGOTE INSECTS

In pterygote insects that produce a spermatophore, transfer to the female is more direct. Odonate males deposit their spermatophore from their genital opening at the tip of the abdomen to the secondary genitalia on the anterior segments of the abdomen. During copulation, the male grasps the female with the tip of his abdomen while her abdomen loops forward to receive the sperm from his secondary genitalia. In many orthopterans, the spermatophore is inserted into the female genital tract with a long spermatophore bulb that protrudes from the female genitalia. The female may eat the exposed bulb once the sperm have left, deriving the considerable nutritional investment that the male has made. Some males present a nuptial gift during and shortly after copulation to distract the female and prevent her from eating the spermatophore before the sperm have left it. In the cricket *Aceta domesticus,* the physical coupling of the male and female is followed by the insertion of the spermatophore into the female genital tract and the attachment of its long tube to the female's ovipositor. The sperm travel into the spermatheca of the female, and after 30 to 40 min, the flasklike spermatophore is dislodged.

In the more advanced pterygotes, the males form the spermatophore in a more recently evolved structure in the female genital tract, the bursa copulatrix, and the sperm then move to the spermatheca(e) of the female for ultimate storage and utilization. In some Lepidoptera, the bursa contains spines that rupture the spermatophore and allow the sperm to escape. Once the sperm have escaped, the spermatophore may be digested by enzymes secreted into the bursa, and its raw materials exploited by the female for egg production. Male *Anopheles* mosquitoes produce a modified spermatophore that serves as a mating plug to temporarily block the genital tract and prevent the female from mating with other males.

The metabolic costs of synthesizing a spermatophore may be one reason for its ultimate disappearance in many of the higher insect orders. Accompanying the loss of the spermatophore has often been the development of a male intromittent organ that is capable of placing the sperm directly into the bursa copulatrix or the spermatheca(e), making the spermatophore unnecessary.

See Also the Following Articles

Mating Behaviors • Reproduction, Male

Further Reading

Davey, K. G. (1960). The evolution of spermatophores in insects. *Proc. R. Entomol. Soc. Lond.* **35A,** 107–113.

Davey, K. G. (1985). The male reproductive tract. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology" (G. A. Kerkut and L. I. Gilbert, eds.), pp 1–14. Pergamon Press, Oxford, U.K.

Khalifa, A. (1949). The mechanism of insemination and the mode of action of the spermatophore in *Gryllus domesticus. Q. J. Microsc. Sci.* **90,** 281–292.

Proctor, H. C. (1998). Indirect sperm transfer in arthropods: Behavioral and evolutionary trends. *Annu. Rev. Entomol.* **43,** 153–174.

Schaller, F. (1971). Indirect sperm transfer by soil arthropods. *Annu. Rev. Entomol.* **16,** 407–446.

Spiders

Rosemary G. Gillespie and Joseph C. Spagna
University of California, Berkeley

For many, spiders are a cause of fear and a source of revulsion. Even to entomologists, spiders have often been thought of as a mere annoyance, filling nets and pitfall traps meant for insect quarry. It is therefore surprising to learn that spiders have held a prominent role in traditional cultures for centuries. Indeed, the terms "arachnid" and "arachnology" come from Greek mythology: A young maiden, Arachne, dared to challenge the great goddess Athena to a spinning contest. Athena wove a remarkable tapestry, yet that spun by Arachne was far superior in its intricacy and perfection. Infuriated that a mere mortal could spin such a masterpiece, Athena turned Arachne into a spider, condemned to a life of perpetual spinning. In native American culture, it was a spider that brought fire to the Cherokee people; the Navajo told of Spider Woman (Na ashje'ii 'Asdzáá), who came from the "first world" and taught the women how to weave; according to Pueblo legend, Spider Woman was at the core of creation; and the Sioux Indians use the "dream catcher," spun by a spider, to capture the good dreams of life. Spiders are also prominent in African culture, as illustrated in the well-known stories of "Anansi, the Trickster." Rather more recently, Scottish legend tells of King Robert the Bruce, whose observations of a spider inspired him to persevere, going on to conquer the English. The reasoning behind some of the tales can be obscure. For example, in the far-flung Micronesian island of Palau, the ability of women to perform natural childbirth is attributed to a spider. With such a diversity of lore, it is clear that spiders have been held in high regard across a global spectrum of cultures and for a very long time.

Spiders, like insects, belong to the phylum Arthropoda, but they are in the class Arachnida, which includes the orders Acari (ticks and mites), Scorpiones, Pseudoscorpiones, Opiliones (harvestmen or daddy-long-legs), and several less common orders. Arachnids are only distantly related to the other major terrestrial arthropod group, the insects, and represent a separate evolutionary transition from marine to terrestrial living, because their closest living relatives are thought to be the marine horseshoe crabs (xiphosurans) and sea spiders (pycnogonids). Together, these marine and terrestrial orders are called the chelicerate orders based on the structure of their mouthparts, in contrast to the orders to which insects and crustaceans belong. Recent molecular and morphological evidence points to the Ambylpygi, or tailless whip scorpions, as the group sharing the most recent common ancestor with spiders, with the other arachnid orders more distantly related. Spiders can easily be distinguished from other arachnids by their lack of visible segmentation and the marked constriction between the prosoma and the opisthosoma, dividing the body into cephalothorax and abdomen, respectively.

Although much less diverse than insects in habits and morphology, spiders, which are in the order Araneae, nonetheless occupy nearly all terrestrial environments and can be found wherever there are other terrestrial arthropods to prey upon. Research into spider biology, particularly the diversity of silks, webs, and venoms, together with the associated ecology and behaviors, has increased greatly in recent decades. Moreover, phylogenetic advances are beginning to provide the context for comparisons between spider taxa and between spiders, other arachnids, and other terrestrial arthropods.

EXTERNAL ANATOMY

The body of all arachnids is divided into an anterior prosoma and posterior opisthosoma, and in spiders these major divisions are referred to as the cephalothorax and abdomen (Fig. 1). These are connected by a narrow stalk, the pedicel. The cephalothorax is covered dorsally by the carapace and ventrally by the sternum and labium. Unlike in insects and most other arachnids, the segmentation of the spider body is not visible externally (except in two primitive lineages) and the cuticle is relatively soft, particularly on the abdomen.

At the front end of the carapace are the simple eyes, or ocelli, usually in four pairs, but in some groups one or more pairs may be reduced or absent. The eight simple eyes are usually arranged in two rows of four, though each of these rows may be curved in such a way that individual eye pairs seem to form their own rows, and in some species one, two, three, or even all four pairs of eyes may be lost. The anterior eye row is closer to the chelicerae, whereas the posterior eye row is farther back on the cephalothorax. Each eye row consists of median and lateral eye pairs, so that each pair can be identified both by row and by position within that row. For example, the anterior median eyes (AMEs) are the central pair, closest to the chelicerae. Below the front margin of the ocular area is the clypeus, from which the two chelicerae extend downward. Each chelicera consists of a stout basal section, from the outer corner of which articulates a narrower distal section, the fang. Behind the mouth ventrally is a second pair of mouthparts,

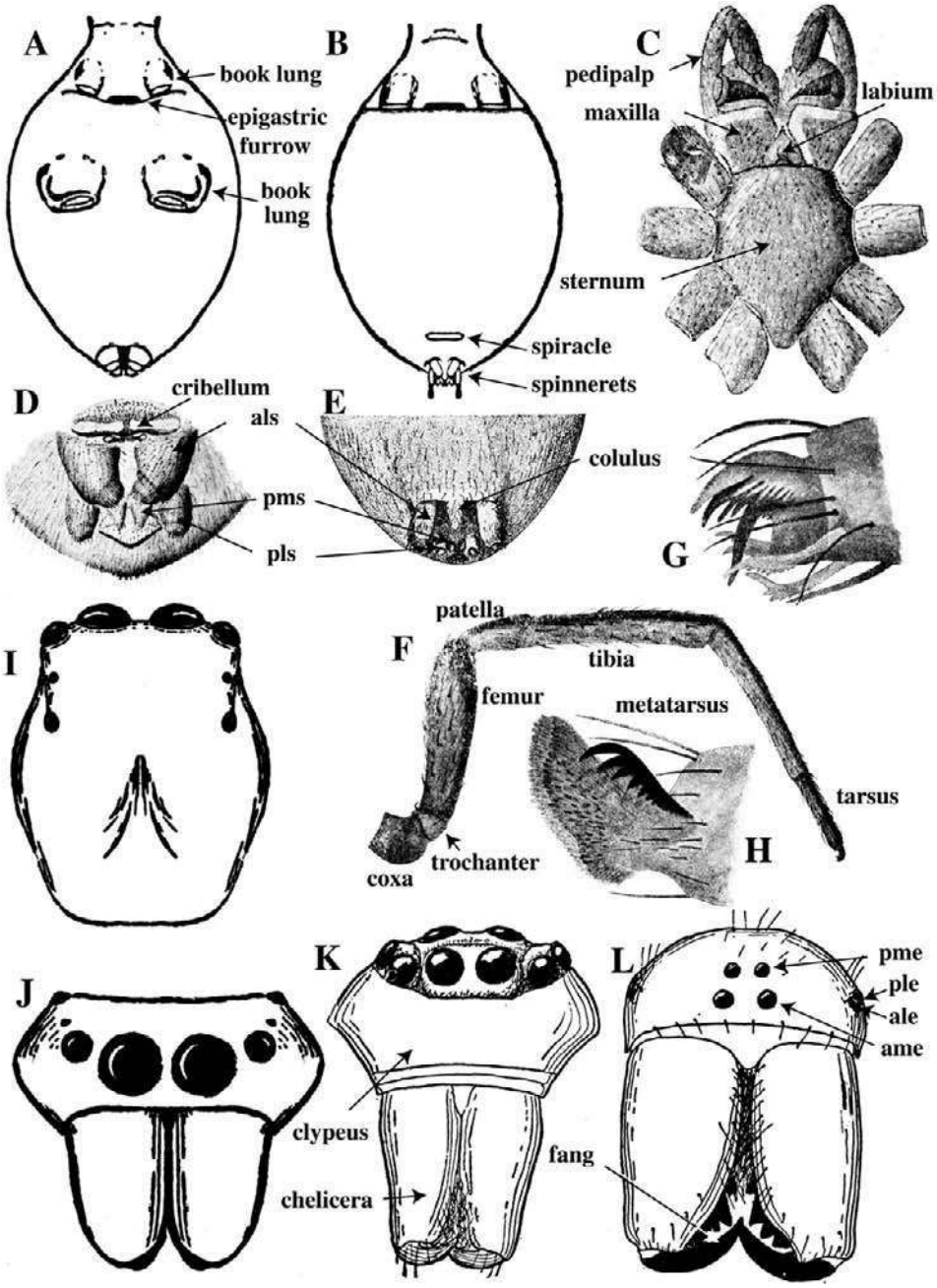

FIGURE 1 External anatomy of a spider. (A–B) Ventral view of abdomen, showing (A) four pairs of book lungs, and (B) two pairs of book lungs. (C) Ventral view of prosoma. (D) Ventral view of female cribellate spider, showing cribellum. (E) Ventral view of female ecribellate spider, showing colulus. (F) Lateral view of leg. (G) Tarsus of leg of three-clawed spider. (H) Tarsus of leg of two-clawed spider and claw tuft. (I) Dorsal view of prosoma (cephalothorax) of jumping spider. (J–L) Anterior view of carapace and chelicerae of (J) jumping spider, (K) tangle-web spider, and (L) orb-web spider. ale, anterior lateral eyes; als, anterior lateral spinnerets; ame, anterior median eyes; ple, posterior lateral eyes; pls, posterior lateral spinnerets; pme, posterior median eyes; pms, posterior median spinnerets.

leg-like in appearance, the pedipalps. In mature males, these are modified for secondary sperm transfer and appear swollen, facilitating recognition of the spider's sex.

Extending laterally from the ventral cephalothorax are four pairs of legs, each consisting of seven segments: coxa, trochanter, femur, patella, tibia, metatarsus, and tarsus. The legs are generally covered with hairs and often have a great diversity of spines, bristles, and scales. These outgrowths serve a variety of functions, including mechanical and chemical sensory functions (see below). Each leg terminates in two or three tarsal claws. In three-clawed spiders, there are usually two larger paired claws and a smaller unpaired median claw, whereas in two-clawed spiders, the median claw is often replaced by a tuft of dense, stiff hairs, called the claw tuft.

A spider's abdomen is carried behind the cephalothorax. The abdomen may be globose or elongate in appearance and

is sometimes covered in hairs or scales similar to those found on the legs. The respiratory book lungs (one or two pairs) open at the anterior end of the abdominal venter. They can be seen externally as patches of hairless cuticle on the venter of the spider, adjacent to the genital opening. In mature female spiders, the genital area is found nestled between the book lungs, the form varying from a simple slit or pair of holes to a complex sclerotized copulatory plate, the epigynum.

At the posterior end of the abdomen are the spinnerets—usually in three pairs, the anterior, posterior, and median

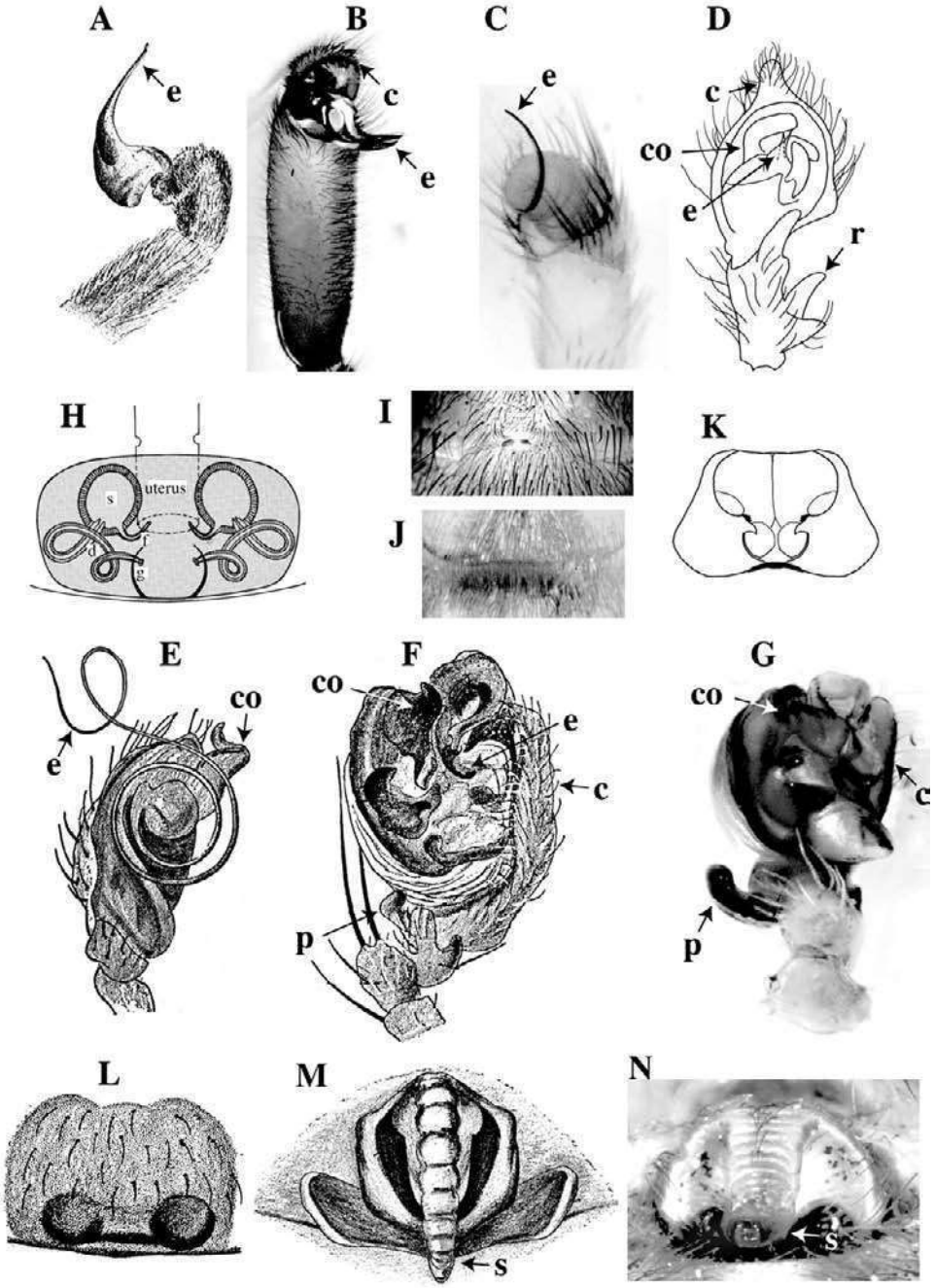

FIGURE 2 Spider genitalia. (A–G) Comparison of male palpal structures showing the change in complexity in different lineages of spiders (see Fig. 5 for major lineages). Lineages represented are (A) *Eurypelma* sp. (Mygalomorphae), (B) *Atypoides* sp. (Mygalomorphae), (C) *Kukulcania* sp. (Haplogynae), (D) *Hololena* sp. (RTA clade), (E) *Theridion spirale* (Orbicularia), (F) *Araneus gigas* (Orbicularia), (G) *Araneus gemma* (Orbicularia). Abbreviations of structures: c, cymbium; co, conductor; e, embolus, p, paracymbium; r, retrolateral tibial apophysis. (H) Diagram of epigynum, ventral view. Abbreviations of structures: d, coiled ducts; f, fertilization duct; g, genital opening; s, seminal receptacle. (Reprinted, with permission, from Fœlix, 1996, Fig. 135, © Oxford University Press.) (I–N) Comparison of female genital structures. (I) *Antrodiatus* sp. (Mygalomorphae), (J) *Kukulcania* sp. (Haplogynae) showing simple slit opening of seminal receptacles (no epigynum), (K) *Hololena* sp. (RTA clade) showing sclerotized epigynum, (L) *Theridion spirale* (Orbicularia), (M) *Araneus gigas* (Orbicularia), (N) *Araneus* sp. (Orbicularia) showing elaborate sclerotized epigynum and scape(s). Panels A, E, F, L, and M, adapted from J. H. Comstock, (1912), "The Spider Book," Doubleday, Page, Garden City, NY. Panels B, C, G, I, J, and N, photographs (microscopic, automontage) taken by the authors. Panels D and K drawn by the authors.

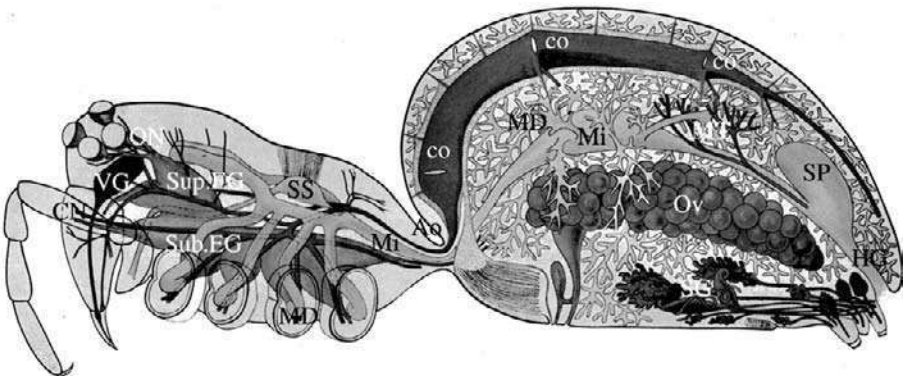

FIGURE 3 Internal anatomy of a female spider (adapted from J. H. Comstock, (1912), "The Spider Book"). Ao, aorta; BL, book lung; co, cardiac ostia of the heart; CN, cheliceral nerve; (P), pharynx (behind the esophageal ganglion); (E), esophagus (behind the esophageal ganglion); HG, hindgut; Mi, midgut; MD, midgut diverticula; MT, Malpighian tubules, Oc, ocular area; ON, optic nerves; Ov, ovaries; SG, silk glands; SS, sucking stomach; Sup. EG, supraesophageal ganglion; Sub. EG, subesophageal ganglion; SP, stercoral pocket; VG, venom gland.

pairs—and the anal tubercle. The median spinneret pair is often obscured by the larger anterior and posterior pairs, and the entire complex of spinnerets may be surrounded by a sclerotized ring. Each spinneret pair has its own complement of silk spigots that extrude silks for specific functions. In addition, many spiders have a cribellum, or spinning plate, found adjacent to the standard spinneret complement that they use to produce an ultrafine looped silk for prey capture. In others, the cribellum has been reduced to a small vestigial lobe, called the colulus, or lost completely.

The sex organs of male spiders are the palps (Figs. 2A–2G). The cymbium, or modified tarsus of the mature male palp, is hollowed out to contain the copulatory organ. The basal appendage of the cymbium is called the paracymbium. The bulb of the male's palp opens through a spine-shaped apparatus, which ends in a fine tube, the embolus. The male palp is generally less complex in primitive spiders (haplogynae) relative to more advanced spiders (entelegynae). More complex palps have hard sclerotized parts (sclerites, which include the conductor, embolus, tegulum, etc.) and soft parts (hematodochae), both of which can carry prominences, or apophyses. At rest, most structures are folded away, with the delicate sclerites protected. During mating, the hematodochae are filled using hydraulic pressure and the sclerites extended, as the male inserts his palp into the female epigynum. The intromittent portion of the palp must navigate through the complex ducts of the female to achieve sperm transfer, while other sclerotized projections on the palp help to lever the embolus into position during copulation. The actual copulation has been referred to as a "lock and key" mechanism because of the often complementary species-specific form of the genital structures.

The female genitalia (Figs. 2H–2N) of primitive spiders comprise a ventral fold within which a hidden opening leads into a single duct, or egg tube. This tube serves for sperm insemination and egg laying. In more advanced spiders (entelegynae), there are two openings, the egg tube and the genital tube, both hidden by an external structure, the epigynum. The epigynum is sclerotized and becomes visible upon maturity. The complex structures on the female epigynum guide the male palpal bulb through a convoluted set of internal ducts into the female's genital tube, which leads to the sperm pocket where the sperm is stored.

INTERNAL ANATOMY

The central nervous system of the spider is located in the cephalothorax and consists of two main ganglia, the larger subesophageal ganglion and the smaller, more anterior supraesophageal ganglion, sometimes referred to as the "brain" of the spider. The two ganglia are divided horizontally by the esophagus, and nerves radiate from both, forming the peripheral nervous system. The supraesophageal ganglion connects to the cheliceral and optic nerves, while the subesophageal ganglion connects to the peripheral nerves of the palps, legs, and abdomen (Fig. 3).

Paired venom glands occupy the upper portion of the chelicera in all spiders except the Uloboridae and, in many spiders, extend well into the cephalothorax about midway between the eyes and the supraesophageal ganglion. The digestive tract consists of a pharynx, esophagus, sucking stomach, and the beginning of the midgut in the cephalothorax, with the rest of the midgut, hindgut, and anus located in the abdomen.

The major respiratory organs are in the lower abdomen and are called "book lungs" because they resemble stacked sheets of paper. More primitive spiders have two pairs of book lungs, others have one pair and may also have a set of tubular tracheae. Within spiders, there has been a sequence of replacement of book lungs by tracheae, apparently in response to problems of circulation, with more active spiders having variably elaborate tracheal systems (Table I). The heart is a tube-like organ, suspended by muscles and ligaments along the dorsal midline of the abdomen, with multiple ostia that serve as valves to keep the blood flowing in one direction. The heart pumps the hemolymph forward through the aorta into the cephalothorax. Silk glands are numerous and may fill up to a third of the volume of the abdomen. In more advanced spiders they can have varied functions (Table II). The gonads consist

TABLE I Relative Compositions of Book Lungs versus Tracheae in Different Groups of Spiders

Second abdominal segment	Third abdominal segment	Example
Book lungs	Book lungs	Mesothelae, Orthognatha
Book lungs	Short tube tracheae	*Filistata*
Book lungs	Long tube tracheae	*Dysdera*
Book lungs	Long entapophyseal tracheae	*Argyroneta*
Book lungs	Short tube tracheae and short entapophyseal tracheae	Araneidae, Lycosidae
Book lungs	Short tube tracheae and long entapophyseal tracheae	*Cryphoeca*
Book lungs	No respiratory organs	Pholcidae
Sieve tracheae	Long tube tracheae	Caponiidae
Sieve tracheae	No respiratory organs	Symphytognathidae
Tube tracheae	Long tube tracheae	*Telema*

Note. The more primitive groups (Mesothelae, Orthognatha) retain book lungs on both second and third abdominal segments. More advanced groups show loss of book lungs and/or replacement by tracheae.

of paired, coiled, tubular testes in males and paired ovaries in which the follicles may appear grape-like in females.

PHYSIOLOGY

Feeding and Venoms

Upon capture, a spider sinks its fangs into the body of its prey. At this point, the spider must paralyze, or otherwise restrict the movement, of the prey rapidly before it can eat it. Nearly all spiders use venom to incapacitate their prey. When envenomating prey, muscles surrounding the venom glands contract, forcing venom out of the glands and through ducts that carry the venom to the tips of the fangs and into the prey. Spider venoms are toxic cocktails of polypeptides and proteolytic enzymes that are quite effective for paralyzing the spider's (usually insect) prey. Because most of the polypeptides have evolved to act on nervous systems of arthropods, which use glutamate as a neurotransmitter, most spider venoms have little effect on vertebrates, but there are exceptions. Some venoms, like those of the widow spiders *(Latrodectus),* contain components that broadly affect vertebrate nervous systems. In humans, a black widow bite, which may go unnoticed when it happens, can result in several days of pain, muscle spasms, abdominal cramping, and weakness. More serious symptoms may include respiratory difficulty and hypertension. Deaths are rare, however, and antivenins are available in emergency situations. Other species, such as the brown recluse spider *(Loxosceles reclusa),* produce venoms that cause tissue death (necrotism) at the site of the wound. While these may not cause systemic effects like the black widow toxins, the wounds produced may take weeks or months to heal, and infection is a serious risk.

TABLE II Silk Gland Types, Silk Uses, and Location of Spigots within Spinnerets

Gland type	Silk uses	Spigot locations
Major ampullate glands	Dragline, web frame	Anterior
Minor ampullate glands	Dragline reinforcement	Median
Aciniform glands	Swathing silk, sperm web, egg sac outer wall	Median, posterior
Cylindrical (or tubuliform) glands	Cocoon silk	Median, posterior
Aggregate glands	Glue for sticky spiral	Posterior
Flagelliform glands	Core of sticky spiral	Posterior
Piriform glands	Attachment disc silk	Anterior

Digestion and Excretion

Spider digestion begins outside its body. Once the spider has disabled its prey with venom, silk, or both, it extrudes digestive enzymes into the prey and then, using negative pressure from its sucking stomach, reingests the soup of digestive enzymes and partially digested food. This is repeated until all of the prey's soft tissue has been consumed. Once the liquefied food has passed through the sucking stomach, it enters the midgut where nutrient absorption takes place, with secretory cells producing digestive enzymes and resorptive cells absorbing food into vacuoles.

Waste is concentrated in the cytoplasm as the nearly insoluble products guanine, adenine, hypoxanthine, and uric acid. These products are collected via the Malpighian tubules and moved into the stercoral pocket, which empties through the hindgut and anus. Other excretory tissues in spiders include the coxal glands, which appear to be involved in water balance, and the large nephrocyte cells that concentrate metabolites.

Circulation and Respiration

Although spiders have well-defined blood vessels, they lack capillaries and have few veins; thus, their circulatory system is basically open. The heart is suspended in the pericardial sinus, and blood enters the heart through paired slits called ostia, which open when the heart is at rest. The heart primarily pumps hemolymph from the abdomen forward through the aorta into the cephalothorax, supplying oxygen to the central nervous system and the skeletal muscles. Once depleted of oxygen, the fluid passes into two sinuses, which lead to the base of the abdomen where the fluid is reoxygenated by the

book lungs and (if present) by tubular tracheae before pressure pulls it through pulmonary veins back into the pericardial sinus.

Unlike most insects, spider hemolymph has an oxygen-carrying pigment, hemocyanin. Hemocyanin is structurally similar to hemoglobin, but instead of iron it uses copper as its oxygen-binding metal, which can make spider hemolymph appear bluish green. Compared to hemoglobin, hemocyanin is less efficient (~5%) in oxygen transportation and is not concentrated in specialized cells.

NEUROBIOLOGY

The detection of touch and sound, mediated by vibrations, is the primary sense of spiders, although other senses may be well developed in some groups. These senses are discussed in detail in a 1985 book by F. G. Barth.

Touch and Sound

Spiders are notably hairy, and most of the various hairs on their bodies function as mechanoreceptors, sensing movement and vibrations from both the spider's substrate (which may be the web, the ground, or vegetation on which the spider is situated) and the surrounding air. They also serve as touch receptors. The many stout hairs on the legs, cephalothorax, and abdomen, as well as the finer, more upright trichobothria, found only on the legs, are triply innervated and are involved in the localization and identification of potential prey. Additional mechanoreceptors are the slit sensilla and a variety of other proprioceptors, which respond to stresses in the exoskeleton caused by external vibration or by the spider's own movements. The slit sensilla are found all over the exoskeleton, but may be arranged in groups to form "lyriform organs." Together with the trichobothria, the slit sensilla/lyriform organs may be the functional equivalent of spider "ears." Other proprioceptors include internal ganglia at leg joints and hairs that respond to joint movement.

Taste and Smell

Chemoreceptive hairs are localized mainly on the tarsi of the front legs, and in the palps, although some chemoreception may take place at the mouthparts. The hairs involved in chemoreception resemble tactile hairs, but are open-ended and S-shaped. Spiders respond to sex pheromones of conspecifics, and it is also likely that they use chemoreception to identify potential prey, enemies, and environmental change.

Sight

The "simple" eyes, or ocelli, have a single cuticular lens. The back of the eye contains the retina, which consists of visual cells (including the light-sensitive rhabdomeres) and pigment cells. The main eyes (always the AMEs) consist of a lens, vitreous body, and retina (which contains the visual cells), while the secondary eyes may also have a light-reflecting surface, the tapetum, behind the retina, which causes the well-known eye shine of spiders at night. The visual acuity of spiders depends on the shape of the lens and the number of rhabdomeres. For most species vision is poorly developed. Most web-building species rely almost exclusively on touch, with vision used to detect light and dark and (in a few species) direction of polarized light. However, vision is well developed in jumping spiders, with the AMEs having high visual acuity (acting like the fovea of the human eye), the remaining eyes having lower acuity but a broader field of vision and functioning for peripheral vision.

SILKS AND SPIDER WEBS

Spiders are unique among arthropods in their use of silks at all stages of their lives. Silks are produced in the abdomen in specialized silk glands, each of which yields a different kind of silk. The general structure of a silk gland is a tail area that secretes the liquid silk proteins into a sac-like lumen, or storage area. The lumen empties into a duct leading out to the spinneret spigots. The ducts are important in silk production because their tapered shape helps to orient the molecules relative to the axis of the thread to maximize strength, whereas the cells surrounding the duct draw off water from the oriented protein and turn them into solid silk fibers. The ducts of the different silk glands terminate at specific spigots in the spinnerets (Table II).

The various silks produced by spiders show different combinations of remarkable material properties, including high tensile strength, extensibility reminiscent of rubber, and resistance to decay. These properties have led to research in the pursuit of silk genes and techniques for spinning the translated products of these genes; genetically engineered spider silk may soon be found in parachutes, bulletproof vests, car bumpers, artificial ligaments, etc. At the molecular level, spider silks are a family of proteins made up primarily of a subset of amino acids (alanine, glycine, serine, proline, glutamine, and tyrosine make up over 75% of the composition of characterized spider silks) arranged in a highly repetitive manner. The smallest repeating units of two to six amino acids are strung together into larger "tandem repeats." This arrangement is thought to form secondary structures that determine the kind of silk produced. The overall structure can be compared to a composite material with stiff, stress-resistant crystals interspersed in an extensible, energy-absorbing matrix.

The most obvious use of spider silk is in web construction, but it is employed for many other purposes in the life of a spider, including the following:

Ballooning

Spiders are capable of producing silk as soon as they emerge from the egg, and two behaviors, the manufacture of a brood web and ballooning, are characteristic of newly hatched juveniles, or spiderlings. A brood web is made by some spiderlings upon emergence and for the earliest instars can serve as a communal nest for the young to catch prey together. Ballooning generally occurs once dispersal in the spiderlings has been initiated by developmental or environmental cues.

The spiderlings let out a silk thread, which produces enough drag to catch the wind and carry the spider off, sometimes for great distances.

Dragline

Nearly all spiders trail a dragline behind them during locomotion. The dragline is extruded from the ampullate glands as the spider walks along or moves in its web, stopping occasionally to anchor the line with an attachment disc secreted by the piriform glands. The jumping spiders (Salticidae) may make seemingly reckless jumps, but are almost invariably anchored with a dragline, so a missed leap is seldom fatal. Dragline diameters vary from several hundred nanometers to several micrometers, depending on the size of the animal being supported. Draglines are stronger than any other known natural fiber, with the added ability to stretch 15 to 30% beyond their original length before breaking.

Egg Sacs

All spiders use silk to protect their eggs. The silk covering of the egg mass may range from a few strands to a thick covering with multiple silks. The resulting silken cocoon can take on a huge variety of forms depending on the spider and the habitat.

Sperm Webs

Silk is also used to make the male sperm web. Shortly after the final molt to maturity, a male spider makes a small web (sometimes just one or a few strands of silk), upon which he deposits sperm from the abdomen. He then places the tip of his palp into the sperm, which is drawn through the palp's opening into the sperm duct, where it is stored.

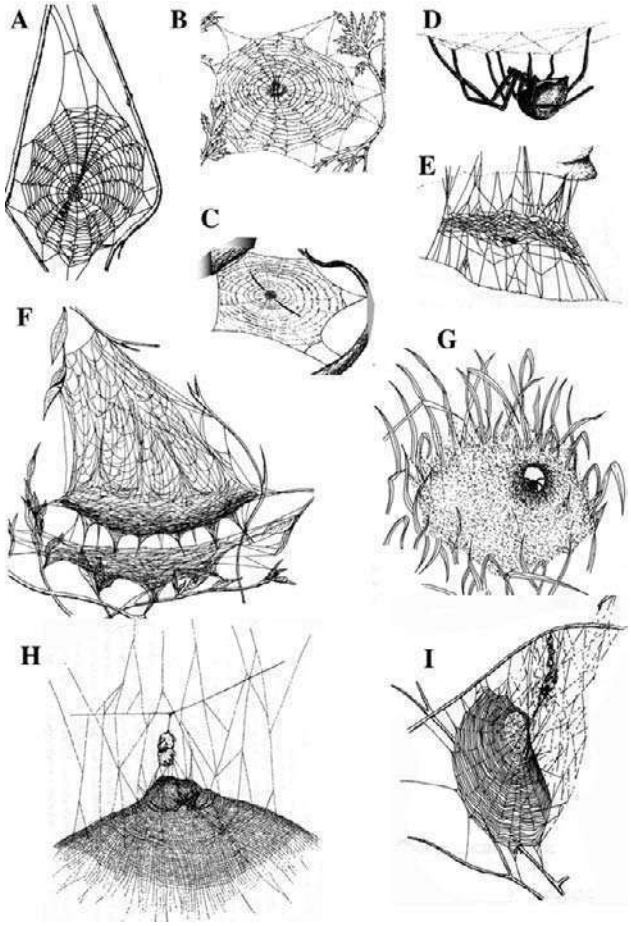

FIGURE 4 Web structures. (A) Typical orb web (Araneidae). (B) Orb web of species of Tetragnathidae. (C) Cribellate orb web (Uloboriodae). (D) Tangle web of black widow (Theridiidae). (E) Tangle web of *Steatoda* sp. (Theridiidae). (F) Sheet web of *Frontinella* sp. (Linyphiidae). (G) Funnel web of *Agelenopsis* sp. (Agelenidae). (Reproduced, with permission, from Wise, 1993; illustrations by Gilbert H. Wise.)

Capture Webs

The most remarkable use of spider silk is in the construction of snares for catching prey. In webs, silks can function both directly as a mechanical trap, to stop or slow potential prey, and indirectly as an extension of the spider's sensory apparatus, alerting it to the trapped prey. A number of spiders can produce sticky capture webs, which come in two forms. The "hackled" or woolly silk of cribellate spiders is sticky (akin to Velcro) because of its fine fibers. The silk is formed by rapid combing (by the calamistrum on the rear leg) of a silk produced from the cribellum, a field of fine openings in front of the spinnerets. In contrast, the web silk of ecribellate spiders usually has sticky globules, secreted by the aggregate glands, arranged along its length.

The type of web that spiders produce is often characteristic of a family, with forms including the two-dimensional orb, tangle (cobweb), sheet, and funnel webs (Fig. 4). Each of these webs has its variants as well; for example, orb webs can be oriented either vertically or horizontally or may be reduced, sometimes to a single section of a complete orb. The impressive body of literature on the form, function, and evolution of spider webs was reviewed by Eberhard in 1990.

Despite the fact that all webs serve the same basic function and despite the common characterization of spiders as generalist predators, the form of the silk of certain webs can be specialized for capture of a small subset of potential prey types. For example, many comb-footed spiders (family Theridiidae) make tangle webs with viscid threads extended to the substrate below, the last centimeter or two of which are coated with a sticky substance ("gum-foot lines") and serve to snare cursorial prey. More specialized is the bolas spider, an orb weaver whose "web" consists of a single strand of silk with a sticky droplet (bolas) at the end, which it uses to catch moths. In addition to these differences in silk form between web types, silks often have different functions within webs. For example, silk from the flagelliform glands makes up the core fibers of the capture spiral in orb webs and has properties quite different from those of the silks that form the web frame and radial elements. The remarkable property

of this capture-spiral silk is its ability to stretch; it may more than double its length while absorbing the kinetic energy of a flying insect. In addition, the capture spiral recovers tension slowly, which prevents prey items from being flung back out of the web.

Web Decorations

One of the most conspicuous features of the webs of some orb spinners is the presence of a stabilimentum, a prominent silk line, cross, and/or spiral, at the center of the orb. The function of the stabilimentum is not clear, though it may serve to camouflage the spider, startle predators, or protect the integrity of the web from accidental damage.

RELATIONSHIPS AND TAXONOMY

Understanding of the relationships among different spider groups and between spiders and other arthropods has increased dramatically in the past 2 decades. Some portions of the spider tree remain unresolved, and additional morphological and molecular study will be needed to settle these uncertainties. Statements of phylogenetic relatedness between spider groups are based on a review of morphological data by Coddington and Levi in 1991, and all groups mentioned can be found on the summary cladogram (Fig. 5), unless otherwise noted.

Spiders are divided into three suborders, the Mesothelae, Mygalomorphae, and Araneomophae, and 106 families. The Mesothelae (1 family, Liphistiidae, two genera, 40 species) are considered the most primitive of all living spiders, based on their external visible segmentation and location and number of spinnerets (all four pairs present, without a cribellum). The Mygalomorphae, recognized by the articulation of the chelicerae parallel to the body (paraxial), are also considered primitive; they are generally stout-bodied and include the trap-door spiders and the impressively large tarantulas (family Theraphosidae) among their 15 families. Araneomorphs, the "true" spiders, represent over 90% of spider diversity. They can be distinguished from the more primitive spiders by the sideways (diaxial) articulation of their chelicerae. Paraxial and diaxial cheliceral orientations are also referred to as orthognath and labidognath, respectively. The most basal lineage of the Araneomorphae is the Paleocribellatae, consisting of 1 family, the Hypochilidae, in which the body plan is a mosaic of primitive and derived characters.

The rest of the araneomorphs belong to one of two clades, the haplogynae and the entelegynae, distinguished on the basis of the complexity of the female genitalia (Fig. 2). In the haplogynae, the smaller group, the external female genitalia consist of a simple opening, or gonopore, tucked between the book lungs, with a single duct serving both the copulatory and the fertilization functions. Two well-known haplogyne families are the spindly legged Pholcidae (or daddy-long-legs spiders), common around and inside houses, and the Sicariidae, which may also occur in and around houses, and include the much-feared brown recluse spiders *(L. reclusa),* whose venom may cause necrotic wounds in humans. In the entelegynae, the female genitalia are more elaborate, with separate fertilization and copulatory ducts and sclerotized epigynum. The most diverse and well-known entelegyne spider families belong to

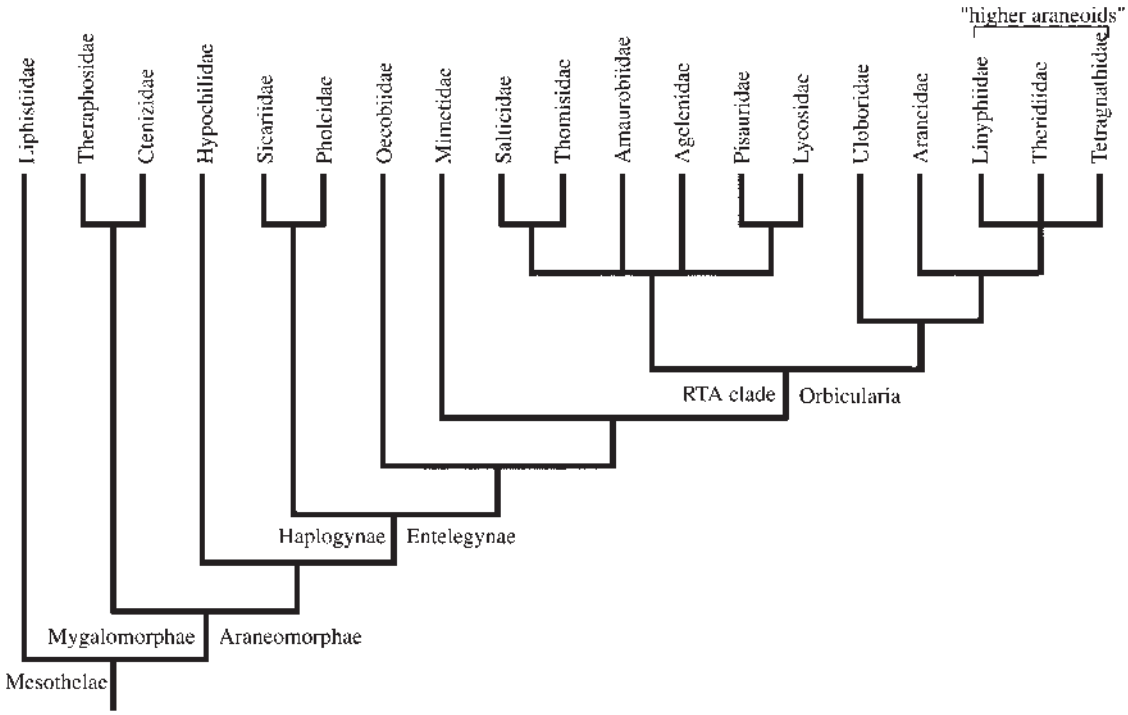

FIGURE 5 Phylogenetic relationships of major spider lineages, with exemplar families. (After Coddington and Levi, 1991.)

the RTA clade, so named because of a retrolateral tibial apophysis on the male's palp, and the orbicularia, which includes the orb-web builders and their relatives. The RTA clade is a huge, ecologically diverse group of spiders and includes about one-third of all described spider species. Many are web builders, for example the funnel-web spiders (Agelenidae). However, some of the most successful lineages have shifted to non-web-building hunting strategies, including the jumping spiders (Salticidae), in which the median tarsal claw is replaced by a tuft of tarsal hairs that facilitates a cursorial lifestyle. Salticidae also have greatly enlarged anterior median eyes, giving them the best vision of all spiders. They use this vision to track prey and carry out unique jumping attacks with great accuracy and in mate choice when faced with flamboyant courtship displays. The crab spiders (Thomisidae) are another family of two-clawed spiders with an interesting variation on the typical spider sit-and-wait game. Thomisids hide in the petals of flowers, where their bright colors often disguise them, and use their raptorial forelegs to capture visiting pollinators. Wolf spiders and their relatives (Lycosoidea) are also cursorial hunters, though they retain all three tarsal claws.

The Orbicularia includes some of the best known web spinners. In particular, the orb web is often considered to be the "classic" spiderweb—a two-dimensional sphere consisting of a frame, with radii projecting outward, and a spiral of sticky silk wrapped surrounding the center. There are three families primarily responsible for these orb webs, the cribellate Uloboridae and the Araneidae and Tetragnathidae, which have lost the cribellum. The orb design, once thought to be the "pinnacle" of web evolution, is now thought to have served as a point of departure for some successful groups (the "higher araneoids") that make webs of quite different designs. The comb-footed spiders (Theridiidae) make seemingly disorganized cobwebs, often with viscid gum-foot lines as mentioned above. On the other hand, "bowl-and-doily" or money spiders (Linyphiidae) make dome-like sheets from which they hang on the lower surface; the viscid silk in these spiders dries up after being produced and serves to cement together the different layers of the sheet.

COURTSHIP, REPRODUCTIVE BEHAVIOR, AND GROWTH

Courtship

The first step in courtship involves maturation of the male spider, at which stage the palps have become modified and swollen for sperm storage and transfer. Once the palps are charged with sperm, the male sets out to find a receptive female. However, spiders are assiduous predators, and the male must overcome this propensity of the female if he is to mate. There is a huge array of courtship strategies that allow the male to approach. Most spiders employ some kind of vibratory communication during courtship. Among web-building spiders, the male often locates himself on the edge of the female's web and gently plucks. It often takes hours, even days, until the female becomes receptive. Wolf spiders (Lycosidae), although they do not have webs, make use of vibrational cues during courtship. Vibrations generated from stridulating organs are usually transmitted through the leaf litter. Visual cues, although used by wolf spiders, are most complex among jumping spiders (Salticidae). Male jumping spiders generally communicate courtship by performing a variably elaborate dance for the female, waving their often brightly colored legs and body to show off iridescent plumes. If successful, the male can extend his forelegs to touch the female before climbing on top of her.

In a number of species the female must adopt a state of complete immobility before the male can initiate copulation. In other species, the male waits until the final molt and can mate with the female immediately. In certain species the female is very much larger than the male and appears hardly to notice that the male is either approaching or copulating with her. Other males secure the female by wrapping her in silk prior to copulation. Species in the genus *Tetragnatha* have an unusual way of mating whereby the cheliceral fang of the female becomes wedged against a dorsal notch on the cheliceral surface of the male. The male then locks the female in place by closing his fangs over hers.

Copulation

Copulation involves the injection of sperm from the male's palps into the seminal receptacles of the female, with the palps being inserted alternately into the epigynum. In some species, the palp breaks off and seals the epigynum. Once copulation has been completed, the male must escape from the female before her brief period of receptivity ends. If the male fails to escape in time, he can be caught and consumed by the female, although more commonly he escapes to mate again. However, the life span of a mature male is generally short, and many do not eat at all. Female spiders, on the other hand, are able to store sperm and so can produce fertile eggs long after copulation. Accordingly, females may survive for some time before egg laying is complete, and in many species they survive longer, to care for their offspring.

Egg Sacs

Some time after mating, the female will deposit an egg sac. The eggs are always laid within a cocoon of silk. However, like courtship, there is a diverse array of egg sac types and behaviors that spiders use to protect their eggs. In most spiders, the female spins a layer of silk into which she deposits her eggs. She then covers the eggs with more silk. The covering may be scant, as in the daddy-long-legs spiders (Pholcidae). In other spiders the sac is very thick, with multiple silks, soft inside but tough and water resistant outside, as in many spiders of the family Theridiidae (e.g., *Latrodectus*). Egg sacs are also produced in a variety of colors, from white to yellow, green,

or black, in pastel or vivid shades. Textures may vary from papery smooth to tufted and furry and shapes from flat to round or angular, with connections to the substrate ranging from a tight fixture to loosely suspended or pendulous. These colors and shades generally match the substrate on which the egg sac is laid, serving to camouflage the developing eggs as protection against predation and parasitism.

Parental Care

Parental care varies tremendously among spiders. Some females abandon their egg sac immediately after it has been laid although, even in these species, the female selects specific sites for deposition of the egg sac. For example, it can be plastered on a twig (e.g., *Tetragnatha*), placed under a leaf or stone, wrapped in a leaf, or suspended in the web or retreat, either with a stalk (e.g., *Argyrodes*) or without. However, extended parental care has been documented in a number of spiders and has broad implications with regard to the origin of social behavior.

EGG SAC DEFENSE A number of female spiders guard their eggs closely until (or after) hatching. Such guarding is common among jumping, crab, sac, and ground spiders, and a variety of others, and can be critical for egg survival in, for example, the green lynx spider *Peucetia viridans* (Oxyopidae) and the Hawaiian happy face spider *Theridion grallator* (Theridiidae). In several groups of spiders, the females carry their eggs with them wherever they go. In particular, wolf spiders (Lycosidae) carry their egg sac attached to their spinnerets; fishing (Pisauridae) and giant crab (Heteropodidae) spiders carry their egg sac under the sternum; daddy-long-legs spiders (Pholcidae) carry their egg sac held in the chelicerae.

CARE OF SPIDERLINGS Maternal care in spiders, when present, often terminates after hatching, and the young disperse. However, female wolf spiders carry their young on their abdomen once they have hatched from the egg sac; fishing spiders build a "nursery web," a large tent-like structure in which the spiderlings live while the mother stands guard. Among a few comb-footed and other spiders, care of the emergent spiderlings can be developed to an extraordinary degree. This is true of the communal comb-footed spiders *Theridion saxatile, T. sisyphium, T. grallator,* and *Anelosimus studiosus.* Females of the latter two species defend the egg sac aggressively and then capture prey for, and even feed, the young, which are unable to capture prey on their own. Providing the young with food appears to be the primary function of brood care once the spiderlings emerge from the egg sac. In addition to simply securing prey, a mother may feed her offspring by regurgitation or by laying "trophic eggs." In some species she may even feed herself to her offspring, a process known as "matriphagy," and in at least one species, *Amaurobius ferox,* the mother is known to expedite this process herself.

Dispersal

One of the most intriguing aspects of spiders is their dispersal. When spiderlings hatch, they are generally aggregated. However, on the first day with suitable wind speeds, they will frequently move up to the highest point they can find (e.g., the tip of a grass stalk) and let out silk which catches in the wind. As the spider lets out more silk, the pull of the wind on the silk becomes sufficient to allow the spiderlings to become airborne. Spiderlings can travel tremendous distances by ballooning and are frequently the dominant component of aerial plankton, although the family groups represented vary with locality: Linyphiidae comprise much of the fauna above land areas that have been examined, whereas Tetragnathidae dominate over oceanic areas. Because of their capacity to balloon, spiders are often the first to colonize unoccupied land masses, whether cleared land or new islands in the middle of the ocean.

Growth

Spiders, like insects, have a rigid exoskeleton and must molt to grow, with three to nine instars (stages between molts) before reaching maturity. Spiders do not metamorphose: A first-instar spiderling looks similar to, though smaller than, an adult. Most spiders live for about a year, though mygalomorphs may live for 30 years.

SOCIALITY

Although most spiders are solitary and highly intolerant of others, there are several species that exhibit some form of social behavior, ranging from aggregations at a certain life stage to prolonged maternal care and even quasisociality or true sociality. Considerable controversy surrounds the origins of quasisocial and true social behavior in spiders, discussed in chapters of a book by Choe and Crespi published in 1997. It may have evolved through coloniality and the development of aggregations around an abundant resource. Alternatively, it may have arisen through extension of brood care into later instars. Many studies have examined the biology of social theridiids, including the formation of colonies and sex ratios in true social species.

Colonial Spiders

Some spiders (e.g., several species in the families Araneidae, Uloboridae, and Tetragnathidae) live in colonies, usually around an abundant food source. The benefits from such behavior include sharing silk support lines for prey capture, sharing prey that is not overall in short supply, and communication.

Subsocial Spiders

In subsocial spiders, there is both maternal brood care beyond the first few developmental instars that is typical of most

spider species and an extended phase of tolerance among young within the maternal nest. However, these species live solitarily as adults. True sociality in spiders is thought to have evolved via the "subsocial pathway" by a prolongation of an early tolerance phase without dispersal. The subsocial route to sociality entails decreasing the genetic variance within breeding groups as a result of families staying in proximity.

True Sociality

True sociality appears to have arisen independently multiple times in spiders. Social spiders, unlike eusocial insects, have no castes that are morphologically different or sterile and most individuals within colonies reproduce. In addition, social spiders show a female-biased sex ratio and high population turnover and inbreeding. Their breeding colonies are closed and new colonies are formed by splitting an existing colony, by a swarm of related females, or by single gravid females. High levels of inbreeding and relatedness among females bias the sex ratio toward the dispersing sex. So, unlike most spiders, males do not appear to disperse large distances in social species. Gene flow between established colonies is rare, and colonies show a high degree of genetic structure. It seems likely that sociality may be maintained by behavioral preadaptations that lead to tolerance and cooperation among colony members on the one hand and population structure on the other.

CAMOUFLAGE AND MIMICRY

Crypsis

Spiders have many enemies. In addition to other spiders, perhaps the most important predators are birds and spider-hunting wasps, both of which possess high visual acuity and color vision. Spiders can have specific adaptations for matching background colors (crypsis) such as flowers/leaves, grass/twigs, bark, underleaf surfaces, and the ground. In some spiders, the crypsis is extraordinarily close and can vary between individuals on different backgrounds. There are, for example, several species that are variable in color but always seem to match the bark on which they are living. Selection for crypsis in similar types of habitat has led to the repeated evolution of similar colorations in unrelated species. For example, spiders that live under leaves in the tropics are generally translucent yellow, often with dark leg joints that disrupt the outline of the legs. During the day these spiders lie flat against the underside of the leaf, thereby reducing shadows and becoming highly cryptic against the light transmitted through the leaf. Five unrelated spider species living in the same Hawaiian forests exhibit these adaptations for underleaf crypsis, as do at least four species from the forests of Panama. Within a single lineage, repeated evolution of similar cryptic coloration in different species has been found in a radiation of *Tetragnatha* spiders in Hawaii.

Disruptive Coloration

In some spiders the characteristic shape of the body is concealed, for example by bold, juxtaposed colors. These kinds of colors tend to be found where the propensity to wander over different backgrounds while searching for prey might preclude true crypsis. Crypsis *per se* may also be difficult in diurnal orb-spinning spiders. Some of these have developed mimetic resemblances to dead leaves and sticks but others have apparently adopted disruptive coloration.

Mimicry

Most studies of mimicry in spiders have been concerned with the imitation of ants, thought to be a form of Batesian mimicry. The spiders may gain some protection from predators through their resemblance to aggressive or unpalatable ants. The topic of ant mimicry has recently been reviewed by Cushing, in 1997. Spiders may also mimic a range of other organisms, alive or dead, and inanimate objects. For example, many *Cyclosa* spp. (Araneidae) build vertical "sticks" of prey remains within the web but leave a gap in the center, which is filled by the spider itself. Crab spiders (Thomisidae) mimic the color of flower heads very precisely and prey on pollinators that approach. Some spiders resemble bird droppings, which are attractive to insects. In the garden spider *Argiope argentata,* the visibility of both the contrastingly colored ventral and the UV-reflecting dorsal side of the opisthosoma may increase insect prey caught in the web.

Apostatic Coloration

At least some predators can develop search images, concentrating their search effort on more common forms of individuals of any species. This can result in polymorphism within a prey species. One of the best examples of this is found in *Theridion grallator* (Theridiidae) (Fig. 6). Within this species, there is a remarkable diversity of color forms, yet the frequency of color forms is similar in different populations, apparently maintained by bird predation.

HABITAT SELECTION

How do spiders select a site in which to live? Many studies have demonstrated that there are clear associations between spider abundance and the structural diversity of the habitat, climatic regime, and prey availability. Most spiders are considered to be sit-and-wait predators, spending much time in locating a suitable site in which to wait and remaining there until its quality deteriorates. The time they remain at a site depends on their investment at the site. For species that do not spin webs, the investment is only the time spent finding the site. Ecribellate orb-web spiders can regain most of their resource investment at a site by ingesting the web before abandoning the site. Among other web-building spiders, however, the

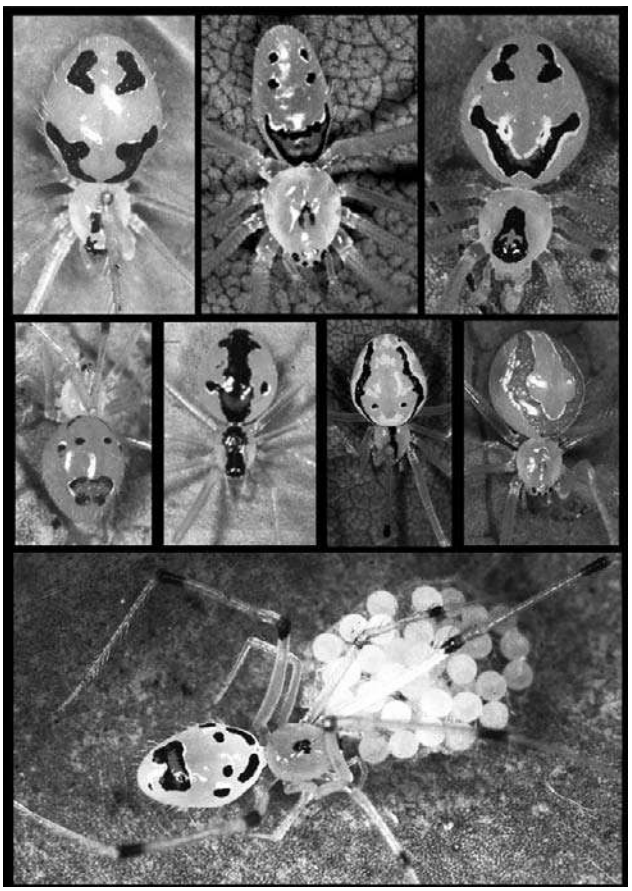

FIGURE 6 Some representative color morphs of the Hawaiian happy face spider *Theridion grallator.*

investment at a site cannot be regained; it can only be decreased through reduction in web size. Movement from a site may be dictated by disturbance or web destruction, microclimate change, growth of the spider relative to structural requirements for web construction, and/or prey capture success.

In selecting a site, it is often not clear what aspect of the environment it is to which the spider is responding. It is possible for hunting spiders to move directly to their feeding site, at least when they can perceive air- or substrate-borne vibrations. Web-building spiders are generally confined to their web for prey detection so that movement from a web site is generally not directed, and to select a site they must "sample" a location by building a web. However, the main vibration receptors of web spiders (trichobothria) are basically the same as those of wandering spiders. Accordingly, web-building spiders may also be capable of perceiving vibrations mediated by the air or the substrate.

PREDATION

Prey Capture

For web-building spiders, prey are usually captured upon being trapped in the sticky threads. The spider will generally respond to the vibration of an insect caught in the threads by shaking the web, causing the insect to become more firmly stuck and allowing the spider to locate the position of the insect. The spider may then approach the insect, often doing so very rapidly, and wrap it in silk of various amounts depending on the size and strength of the insect. This approach and attack behavior varies, with some web-building species leaping upon and biting the prey even before it becomes firmly secured in the web. More specialized strategies have evolved in a number of web builders. For example, the ogre-faced spider *(Deinopis)* holds its cribellate web, orb-like in form, between the two front pairs of legs; it uses the web much like a net, hurling it over prey that come close. Comb-footed spiders may use their gum-foot lines to catch crawling prey; the unsuspecting prey may either dislodge the tensed lines, which springs them into the tangle of the web, or they get reeled in by the alerted spider.

For spiders that do not build webs, prey may be detected by air- or substrate-borne vibrational cues or, alternatively, at least in the case of jumping spiders, by visual cues. Cursorial spiders often jump on their prey without wrapping it in silk. However, certain groups have elaborate mechanisms for securing their prey. In particular, the spitting spiders (Scytodidae) have a high-domed cephalothorax, which contains venom glands connected to a posterior gland that secretes a sticky silk. The spider creeps up on its prey and, by rapid contractions of the muscles in the prosoma, ejects sticky silk and venom over the prey, which is thus immobilized. One species of long-jawed spider in Hawaii has long tarsal claws that it uses to impale insects directly from the air. Other specialized mechanisms of prey capture are discussed below.

Diet

Spiders are exclusively carnivorous, although they are usually regarded as generalist predators, taking prey as they are encountered. However, their diet is dictated by the habitat in which they select to live, and habitat specialization can greatly restrict the dietary repertoire of many species. Even within a given habitat, species may specialize on prey, either through choice or because they exhibit specialized predatory behaviors. For example, species in the genera *Dipoena* and *Euryopis* (Theridiidae) are specialized for feeding on ants, while those in *Dysdera* (Dysderidae) appear to specialize on isopods, and pirate spiders (Mimetidae) and some *Argyrodes* (Theridiidae) feed on other spiders.

Kleptoparasitism

Kleptoparasitic spiders steal food from the webs of other, unrelated and usually larger web-building, spiders, which are also potential predators. They may form groups of up to 50 individuals around the web of a "host" spider and glean small insects from the periphery of the host spider's web, eat the host's silk, steal food bundles previously wrapped and left in the web by the hosts, or even approach a feeding host spider

and then feed, undetected, next to the host spider. The most diverse collection of obligate kleptoparasites is found in the spider genus *Argyrodes* (Theridiidae), which contains at least 200 species, about 100 kleptoparasitic.

Aggressive Mimicry

Aggressive mimicry is mimicry that enhances the access of spiders to their prey. Perhaps the most interesting case of aggressive mimicry is shown by certain jumping spiders that can mimic the vibratory signals of a fly by plucking with their legs and palps on the web of other spiders. This may attract the occupant to the jumping spider close enough to be attacked. The best-studied species in this regard is *Portia fimbriata* (Salticidae) from Queensland. This spider preys on web builders with poor eyesight, which allows the araneophagic *Portia* to drum out a pattern of signals on the web that imitates the vibrations of the normal prey of the web spider. In contrast to the vibrational mimicry of *Portia,* the bolas spider (Araneidae) emits odors that mimic the pheromone of female moths to attract male moths to the vicinity. As soon as a moth flies within range, the spider launches its sticky bolas on a silk thread; if successful, the bolas hits and sticks to the moth, and the spider reels in its prey.

DIVERSITY AND CONSERVATION

There are about 34,000 described species of spiders, though the actual number of species has been estimated at around 170,000. In common with insects, species diversity is highest in tropical regions, where knowledge of the biota is least. Also like other arthropod groups, spiders exhibit spectacular examples of adaptive radiation and local endemism in isolated situations, for example the genus *Habronattus* (Salticidae) on the North American mountaintop "sky islands" and the genus *Tetragnatha* (Tetragnathidae) in the Hawaiian Islands. However, in almost every ecosystem, the number and diversity of spiders make them important in the management of ecosystems, whether natural or disturbed, agricultural or urban.

Importance and Conservation in Agricultural Systems

Spiders have been known to be important agents of insect pest control for hundreds of years. It is also well known now that chemical pesticides kill not only the insect pests, but also their major predators, in particular spiders. A review by Riechert and Lockley in 1984 brought attention to spiders as potential agents of biological pest control. Manipulations of habitat structure have resulted in increased spider densities and concomitant decrease in insect pests. The effectiveness of spiders is often enhanced by the ability of the web to kill prey even in the absence of the spider and because many species practice wasteful killing (i.e., they kill more insects than they consume). Control of insect pests by spiders appears to be achieved most efficiently by using an assemblage of spiders (rather than specific species) and incorporating natural refuges in and around the crops. The particular methods that are best for enhancing the role of spiders in pest control in agriculture are: (i) conservation tillage and retention of stubble, preserving weeds and mulching; (ii) intercropping, provision of native vegetation within crops; and (iii) agroforestry, planting a combination of trees and food crops.

Importance and Conservation in Natural Ecosystems

We have little understanding of the precise role that spiders play in a given ecosystem. However, we do know that spiders are the dominant invertebrate predator in most terrestrial ecosystems and are also important as a food resource for other predators. Accordingly, they must inevitably play a key role in community dynamics. Studies on birds in natural areas have shown that spiders can comprise a significant proportion of the diet of many species, including a number of endangered birds. Spiders also have utilitarian value and have been used as model organisms for research in ecology, behavior, and communication. Moreover, there has been much recent attention on their potential for providing silk for materials science and supplying venom for medical and insecticide research.

Clearly, there is a huge need to augment taxonomic information on spiders, particularly in tropical areas, and to understand patterns of biodiversity. Pursuing this knowledge is imperative as we recognize the obvious importance of spiders in ecosystems, both as prey and as predators. Moreover, given the intrigue that spiders have long held in human society, such an undertaking can only lead to greater heights of fascination.

See Also the Following Articles

Daddy-Long-Legs • Mimicry • Predation • Silk Production • Venom

Further Reading

Barth, F. G. (1985). "Neurobiology of the Arachnids." Springer-Verlag, Berlin.

Choe, J., and Crespi, B. (1997). "Evolution of Sociality," Vol. II of "Social Competition and Cooperation among Insects and Arachnids." Cambridge University Press, Cambridge, UK.

Coddington, J. A., and Levi, H. W. (1991). Systematics and evolution of spiders Araneae. *Annu. Rev. Ecol. Syst.* **22,** 565–592.

Cushing, P. E. (1997). Myrmecomorphy and myrmecophily in spiders: A review. *Fla. Entomol.* **80(2),** 165–193.

Eberhard, W. G. (1990). Function and phylogeny of spider webs. *Annu. Rev. Ecol. Syst.* **21,** 341–372.

Fœlix, R. F. (1996). "Biology of Spiders." Oxford University Press, Oxford.

Jackson, R. R., and Pollard, S. D. (1996). Predatory behavior of jumping spiders. *Annu. Rev. Entomol.* **41,** 287–308.

Kaston, B. J. (1978). "How to Know the Spiders." Brown, Dubuque, IA.

Nentwig, W. (ed.) (1987). "Ecophysiology of Spiders." Springer-Verlag, Berlin.

Oxford, G. S., and Gillespie, R. G. (1998). Color polymorphisms in spiders. *Annu. Rev. Entomol.* **43,** 619–643.

Oxford, G. S., and Gillespie, R. G. (2001). Portraits of evolution: Studies of coloration in Hawaiian spiders. *Bioscience* **51,** 521–528.

Riechert, S. E., and Harp, J. M. (1986). Nutritional ecology of spiders. *In* "Nutritional Ecology of Insects, Mites, Spiders and Related Invertebrates" (F. Slansky and J. G. Rodriguez, eds.), pp. 645–672. Wiley, New York.

Riechert, S. E., and Lockley, T. (1984). Spiders as biological control agents. *Annu. Rev. Entomol.* **29,** 299–320.
Roth, V. D. (1993). "Spider Genera of North America," 3rd ed. Am. Arachnol. Soc., New York.
Shear, W. A. (ed.) (1986). "Spiders, Webs, Behavior, and Evolution." Stanford University Press, Stanford, CA.
Skerl, K. L., and Gillespie, R. G. (1999). Spiders in conservation: Tools, targets and other topics. *J. Insect Conservat.* **3(4),** 249–250.
Wise, D. H. (1993). "Spiders in Ecological Webs." Cambridge University Press, Cambridge.
Witt, P. N., and Rovner, J. S. (eds.) (1982). "Spider Communication—Mechanisms and Ecological Significance." Princeton University Press, Princeton, New Jersey.

Springtail

see *Collembola*

Stamps, Insects and

Charles V. Covell Jr.
University of Louisville

One of the major recent developments in philately (stamp collecting) is the popularity of "topical" or "thematic" collecting. The collector looks for anything that relates to the chosen theme. Popular examples are flowers, railroads, birds, sports, stamps-on-stamps, space, and insects. In a recent survey, butterflies and other insects on stamps ranked as the sixth most popular topic.

As of 1991, over 5000 stamps were issued by over 300 countries depicting over 1300 insect species (Fig. 1). Therefore specialization in types of insect being collected might be desirable, as it is with insect collecting itself. Subsets that could be selected as a specialty include butterflies and/or moths, apiculture, sericulture, malaria and other aspects of medical entomology, applied entomology, and noninsect arthropods. A "taxonomic" collection (in which the insect is the only real design) might be restricted to identifiable and named insects. Some, however, prefer to collect stamps with fanciful depictions, such as butterfly carnival costumes, a butterfly in the movie *Bambi* (on Bambi's nose), or a "fly speck" insect in the beak of a bird. Still other philatelists specialize in mint (unused) stamps or canceled (used) ones, including "CTOs," or stamps "canceled to order." The CTOs are stamps with neat cancellations applied before the stamps are released, and the cancels are normally in one corner to avoid obscuring much of the stamp design. Many philatelists frown on CTOs because they are obviously prepared for philatelists and are not used in carrying the mail.

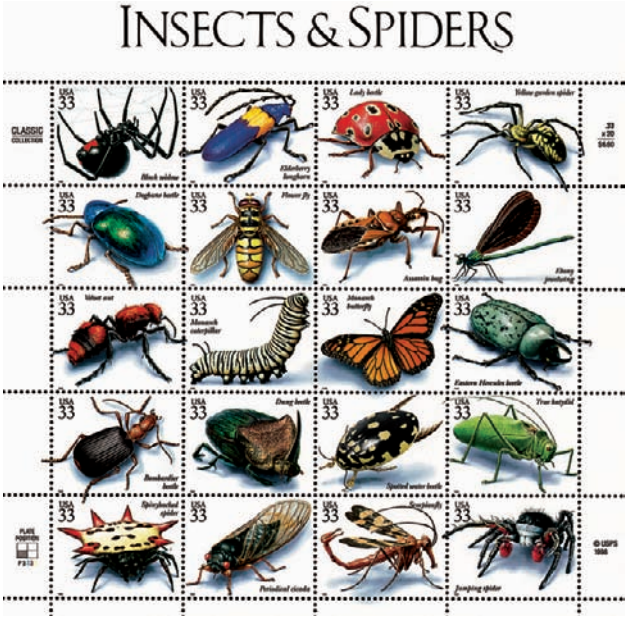

FIGURE 1 This pane of 20 different 33-cent stamps showing U.S. insects and spiders was issued on October 1, 1999 (Scott Catalog item 3351a-1).

PHILATELIC ELEMENTS AND TERMINOLOGY

Philately consists of a number of "elements" (a term used in exhibiting). Mint stamps, used stamps, and CTOs are just the beginning. Other related postal items that can be collected and exhibited are as follows:

Cancel The marking applied in the post office to show the stamp has been used.

Cover An envelope with stamps and cancel attached.

First-day cover (FDC) Cover canceled on the first day of the stamp's issue, often with special cancels and a cachet, (or illustration on envelope created to enhance interest and value) (Fig. 2).

Favor cancel Cancel applied for a collector, not to carry mail.

Piece Fragment of a cover, especially a large one such as from a package, but including stamp and cancel.

FIGURE 2 First-day cover for U.S. 4-cent malaria eradication stamp issued on March 30, 1962. Design of envelope (cachet) and special first-day cancel make the "FDC" a favorite collectible (Scott Catalog item 1194).

Meter cancel A machine-applied indication of postage paid, often accompanied to the left by a picture or slogan and used in place of stamps.

Proof A special impression of the stamp's design used to check quality of design and color, with die proofs usually one color and often signed by the artist.

Essay Original artwork submitted to the stamp-issuing authority for consideration for adoption (these are rare and usually unique; essays for designs that were never printed as stamps are especially rare).

Block Symmetrical grouping of stamps, still attached to each other, with a block of four being the most common.

Plate block Block with the selvedge (unprinted border) containing a plate serial number printed in the original pane of stamps.

Revenues Stamps issued for fiscal purposes (postage stamps are sometimes applied to receipts in place of revenue stamps—an uncommon element to find).

Sheet A full pane of stamps, usually surrounded by a border (selvedge), which is the way stamps usually arrive at a post office.

Miniature sheet A special sheet of usually two to several stamps; not a normal full pane; it may have special related pictures in the selvedge.

Postal stationery Postal cards and letter sheets with stamps printed on them.

Souvenir sheet One or more stamps inside a printed border that may be a continuation of the vignette (picture) of the stamp itself.

Unified design sheet A pane in which each stamp is a small part of an overall picture depicted by the sheet.

GETTING STARTED

The first thing a beginner should do is get some help from someone who has experience in collecting stamps. The rudiments of philately, however, do not take long to learn. Knowledge of the terms just defined, and how to handle the different items is a good starting place. There are books and periodicals available as well. In North America, *Linn's Stamp News,* a weekly newspaper, has pages of ads as well as columns on how to "do" philately. People at a local stamp club and/or stamp shop also can be helpful.

Knowledge about stamps and other postal subjects can be found in specialized books. Stamp values (mint and used) in North America usually follow the *Scott Standard Postage Stamp Catalog,* illustrated and revised yearly.

The APS (American Philatelic Society), which publishes a monthly magazine, *American Philatelist,* has departments including library and expertization services. The ATA (American Topical Association) is for thematic collectors and publishes the bimonthly *Topical Time.* This organization is divided into specialty groups. The Biology Unit, which publishes a newsletter called "Biophilately," serves, among others, insect philatelists. The Philatelic Lepidopterists' Association publishes a newsletter called "The Philatelic Aurelian" (an aurelian is a butterfly collector), which features updates on new butterfly and moth issues (identifiable ones only). Finally, by using key words "stamps" or "philately," one can find many sources of material and information via any search engine on the World Wide Web.

A collector of insect stamps might wish to subscribe to a new issues service run by a dealer who mails new issues to subscribers each month. Such services are listed in the ad section of *Linn's,* and many philatelists are glad to recommend their favorites. New issue dealers very likely can supply older stamps from their stocks of previous issues. There are many stamp club shows and occasional bourses (shows without exhibits, with dealers only), where collectors can fill checklists from dealers' stocks and buy supplies—and meet with other philatelists.

Insect philately keeps alive the thrill of the hunt and the love of the beauty and order of an insect collection. However, for many participants this kind of collection takes less storage space, needs no fumigating, and can keep giving enjoyment when entomologists' more active years of insect collecting are behind them.

See Also the Following Articles

Commercial Products from Insects • Cultural Entomology

Further Reading

American Philatelic Society, P.O. Box 8000, College Station, PA 16803. Web site: www.stamps.org.

American Topical Association, P.O. Box 50820, Albuquerque, NM 87181–0820. Email: atastamps@juno.com.

Coles, A. and Phipps, T. (1991). "Collect Butterflies and Other Insects on Stamps." Stanley Gibbons Thematic Catalogue. Stanley Gibbons Ltd., London.

Hamel, D. R. (1991). "Atlas of Insects on Stamps of the World." Tico Press, Falls Church, VA.

"Scott Standard Postage Stamp Catalog." (revised annually). Scott Publishing, Sidney, OH.

Williams, L. N. (1990). "Fundamentals of Philately." American Philatelic Society, College Station, PA.

Sterile Insect Technique

Jorge Hendrichs

Joint Food and Agriculture Organization/International Atomic Energy Division, Vienna Division, Vienna

Alan Robinson

International Atomic Energy Agency Laboratories, Seibersdorf, Austria

The sterile insect technique (SIT) is a biologically based method for the control of insect pests. Female insects inseminated by released, radiation-sterilized males do not reproduce, and repeated releases of the sterilized insects lead

to a reduction in the pest population. Effective control using SIT is achieved when it is applied systematically on an area-wide basis over multiple generations. SIT is species specific, nonpolluting, and resistance free. Since the original concept was developed in the United States in the 1940s and 1950s, SIT has been used successfully for screwworm flies, tsetse flies, fruit flies, and moths. Technical developments in behavioral ecology, mass rearing, strain improvement, global information systems, aerial release, and monitoring systems, combined with economies of scale and a growing demand for low-pesticide agricultural products in local and international trade, have increased the use of SIT in area-wide integrated pest management programs. These programs have decreased insecticide use against some major pest insects and thereby facilitated the efficacy of biological control agents against secondary insect pests.

HISTORY OF SIT

In the late 1930s E. F. Knipling, working with *Cochliomyia hominivorax,* a screwworm fly causing myasis in livestock and humans, developed the concept of using insects to control themselves (i.e., autocidal control) by reducing their reproductive capacity. Two developments helped to translate this concept into reality. The first was a breakthrough in the development of mass rearing procedures for this insect and the second was the identification of an efficient method to sterilize insects. In 1916 G. A. Runner demonstrated that X rays could sterilize the cigarette beetle, *Lasioderma serricorne,* and in the 1920s, H. J. Muller showed that *Drosophila* fruit flies could be sexually sterilized by radiation-induced dominant lethal mutations. Knipling and colleagues confirmed that low radiation doses given to mature pupae and adults of both sexes of the screwworm sterilized them without excessive damage to their mating ability or diminishing their longevity. The first decisive test that confirmed the feasibility of SIT was implemented in 1954 on the island of Curacao, 64 km off the coast of Venezuela. The success of the experiment so impressed the cattle industry in Florida that in 1957 members persuaded the U.S. Congress to fund a statewide eradication program. A rearing facility with a capacity of producing 50 to 75 million sterile flies per week was established in 1958 at Sebring, Florida, with the result that the last endemic screwworm case was recorded in Florida in 1959. Based on this success, the program was expanded, with new rearing facilities at Mission, Texas, and eventually in Tuxtla Gutierrez, Chiapas, Mexico. Release of sterile screwworms eradicated this pest from North America and more recently from Central America (Fig. 1). Currently, a permanent "sterile fly barrier" consisting of the aerial release of 40 million sterile flies per week in eastern Panama protects Central and North America from reinvasion.

PRINCIPLE OF SIT

The SIT involves industrial-scale mass production of insects in large biofactories. Insects for release are exposed to a precise dose of γ-radiation that is sufficient to induce sterility without significantly impairing the ability of the treated insects to fly, mate, and transfer sperm. The treated insects are released over the target area on a sustained basis and in sufficient numbers to achieve appropriate sterile-to-wild insect overflooding ratios. This tactic substantially reduces the proportion of fertile matings and results in the decline of

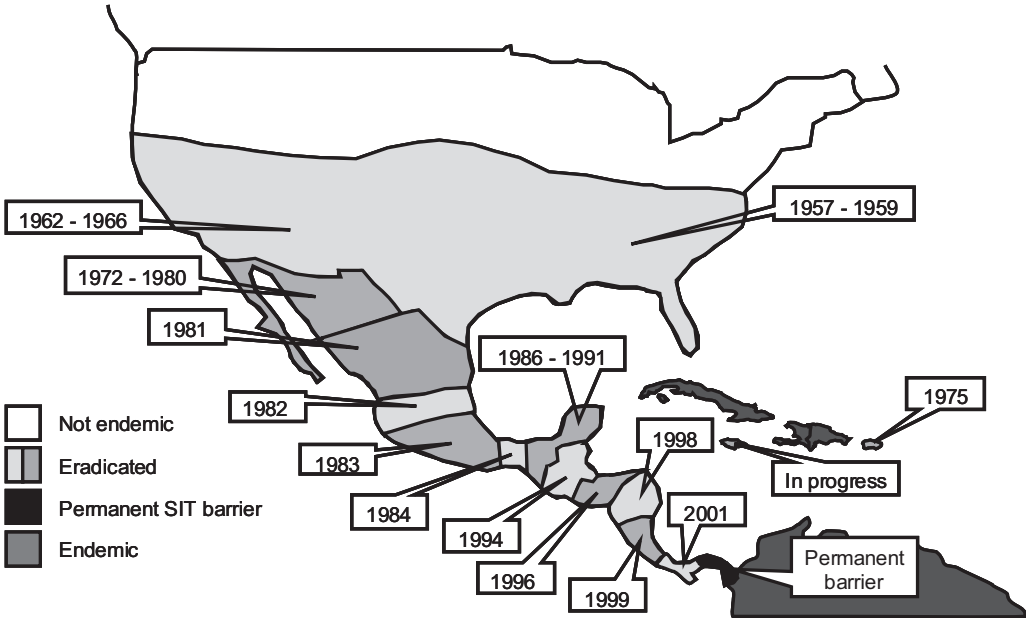

FIGURE 1 Progressive shift of eradication zones over time in the screwworm SIT program in North and Central America. [Reprinted from Robinson, A. S. (2002). *Mutation Res.* **511,** 113–132, with permission.]

the field population. The SIT stands or falls on the ability of the released sterile males to transfer functional sperm to the females that succeeds in fertilizing the eggs. The number of times a female mates is irrelevant, providing that the sperm from the sterile male is competitive with the sperm from the wild male. Knipling devised a simple mathematical model to describe the SIT. Despite some limitations of the model, it provides a valuable set of equations to demonstrate the principles of SIT. As the wild population declines and the number of released insects remains constant, both the proportion of sterile matings and the rate of suppression increase, i.e., the efficiency of SIT is inversely density dependent, and once a wild population begins to decline, the rate at which this occurs increases with time. Also, compared with classical biological control, which involves the introduction of exotic control agents, SIT has important advantages: (1) autocidal control is species specific, (2) no exotic species are involved, and (3) no new strains of the target pest can become established because the released insects are sterile.

REQUIREMENTS FOR SIT APPLICATION

Not Stand-Alone Technology

The use of any single technique in isolation is unlikely to be successful in an insect control program and integrated approaches are now favored. This caveat applies especially to the use of the SIT because insect populations in the field are often of such magnitude that it is impractical to consider releasing sufficient sterile insects to cause the initial population decline. Because the strength of the SIT is at low population densities, to be effective it relies on a variety of presuppression activities (such as application of insecticides), depending on the particular insect species targeted. An alternative, and frequently used, strategy is to apply the SIT when the natural population is at low density because of adverse climatic conditions.

Area-Wide

The effectiveness of integrating compatible pest control methods is increased by coordinated implementation over large contiguous areas that targets the total insect population rather than the protection of individual crops or animal herds. This approach is critical to the SIT because already-mated females that move into an area under treatment are largely unaffected by the presence of sterile males. The area-wide concept has to be seen, not just in terms of the individual farmer or even of the agricultural community, but in terms of the whole ecosystem, including human activity, in the treatment area.

Selection of Target Species

Most insect pest species reproduce by sexual reproduction and therefore, in theory, are amenable to this type of control strategy. However, in reality the technique is used on a selective basis. Probably the first criterion is that the target species is a key pest in its particular ecosystem. It also makes very little sense to develop SIT for a species for which effective, acceptable conventional controls are available. As the technique relies on the release of large numbers of adult sterile insects into the environment, it is important that this stage not be the damaging one. For example, use of SIT against house flies or cockroaches would probably not be acceptable.

Knowledge of Population Dynamics, Ecology, and Mating System

SIT compromises the natural reproduction system of the target species, and it is therefore very important that the relevant population dynamics are well understood in terms of density dependence, life-table parameters, and so on. In addition, the spatial and temporal ecological parameters of the target population need to be defined and the population size must be estimated. It is also essential that the natural mating system is well understood, so that the mating behavior of the released males can be monitored to ensure that no significant changes in the behavior of released insects accumulate during mass rearing.

Colonization, Mass Rearing, and Sterilization

An efficient, economical, and biologically acceptable method for the mass production of competitive insects is essential. Mass rearing of large numbers of insects over many generations in large biofactories imposes major selection and adaptation forces on all aspects of the behavior of the insect. These forces inevitably work to the detriment of the competitiveness of the insect in the field. The art of mass rearing is to find an acceptable compromise between the efficiency of mass production and the overall quality of the mass-produced individuals. Beyond a certain point it is not possible to compensate for the poor quality of insects by simply releasing more of them into the environment. The final and most critical treatment that the insects undergo before release is sterilization. This process uses ionizing radiation and must be subject to stringent quality control protocols to ensure that all the released insects are irradiated and that the required dose is applied.

Quality Control

As in any industrial system, quality control in terms of both the process (mass rearing) and the product (sterile insects) plays a major role in determining the success of an SIT program. For most programs a detailed set of standard operating procedures is in place, which provide some confidence as to the quality of the insects being produced. The use of mating tests under seminatural conditions to ensure continued compatibility of released insects with the wild population is essential.

STRATEGIC APPLICATIONS OF SIT

Eradication

SIT has the unique capability of eradicating populations of pest insects providing that certain ecological and regulatory conditions are met. The inverse density-dependent action of the SIT is very effective in removing the last individuals from the target population. Eradication is often assumed not to be an option if the target population is not isolated. To some extent this is true, but in many cases isolation can be created by making use of natural barriers to pest distribution and by applying the technique on an area-wide basis. The screwworm eradication program relied on this approach with the population being eradicated in a "rolling carpet" approach (Fig. 1). The maintenance of the pest-free status following eradication requires adequate quarantine measures and continued surveillance. SIT is also one of the most powerful tools to eliminate incipient outbreaks of introduced exotic invasive species.

Containment

The very benign environmental impact of SIT means that sterile insects can be released preventively into an area on a continuous basis to avoid the movement of a pest into a sensitive agricultural area. Sterile insects can also be used in a barrier approach to protect an eradicated area from reinvasion.

Suppression

The concept of using sterile insects as a tool in integrated suppression programs to replace the use of insecticides is gaining increasing acceptance as the cost of sterile insects is decreasing and concerns about the environment and food safety are increasing. The use of SIT for suppression removes the need for expensive and difficult to maintain quarantine regulations and the continuous nature of a suppression program will stimulate the commercialization of the SIT.

OPERATIONAL PROGRAMS

In the past decades, SIT has been applied in over 20 countries on five continents to control some 15 insect species that account for an important part of agricultural losses worldwide. Table I provides an overview of the mass rearing facilities currently in operation in support of SIT programs against four main groups of pest insects: fruit flies, screwworms, tsetse, and moths.

TABLE I Sterile Insect Mass Rearing Facilities in the World

Country	Location	Species	Weekly capacity (millions)
		Fruit flies	
Argentina	Mendoza	*Ceratitis capitata*	100–250[a]
Australia	Perth	*C. capitata*	5–15[a]
Chile	Arica	*C. capitata*	50–75[a]
Guatemala	El Pino	*C. capitata*	1000–1500[a]
Mexico	Metapa	*C. capitata*	500–600
Portugal	Madeira	*C. capitata*	35–60[a]
South Africa	Stellenbosch	*C. capitata*	5–10[a]
United States	Hawaii	*C. capitata*	210–450
Australia	Campden	*Bactrocera tryoni*	15–25
Japan	Okinoawa	*B. cucurbitae*	70–200
Mexico	Metapa	*Anastrepha ludens*	150–250
Mexico	Metapa	*A. obliqua*	25–50
Mexico	Metapa	*A. serpentina*	5–10
Philippines	Quezon City	*B. philippinensis*	10–20
Thailand	Pathumtanee	*B. dorsalis*	15–40
United States	Gainesville	*A. suspensa*	10–50
United States	Mission	*A. ludens*	15–40
		Screwworms	
Mexico	Tuxtla	*Cochliomyia hominivorax*	120–500
		Moths	
Canada	Okanagan	*Cydia pomonella*	12–15
United States	Phoenix	*Pectinophora gossypiella*	70–90
		Tsetse (1000s)	
Tanzania	Tanga	*Glossina austeni*	60–100[a]
Burkina Faso	Bobo-Dioulasso	*G. palpalis gambiensis*	40–50[a]

[a]Sterile male-only production, otherwise 50% sterile males and 50% sterile females. [Modified from Hendrichs, J. (1995). *J. Appl. Entomol.* **119,** 371–377.]

Fruit Flies

Few insects have a greater impact on international marketing and world trade in agricultural produce than tephritid fruit flies. Maintaining or achieving a pest-free status has therefore been the goal of a number of large operational programs. Mediterranean fruit fly, *Ceratitis capitata,* ("medfly"), was removed from areas it had invaded in southern Mexico, and it is now close to 20 years that a sterile fly barrier across Guatemala has maintained its medfly-free status. Similarly, Chile eliminated medfly from northern Chile and since the mid-1990s has supported a sterile fly barrier in Peru's two southernmost valleys. The melon fly, *Bactrocera cucurbitae,* was eradicated from southern Japan and the Queensland fruit fly, *B. tryoni,* from Western Australia and some *Anastrepha* species from northern Mexico. Rather than having to eradicate outbreaks of exotic medflies, often on an annual basis, California and Florida maintain permanent medfly preventive release programs in high-risk areas, which is less costly and politically more acceptable than aerial insecticide applications. Increasingly, SIT is also being applied for area-wide suppression purposes without quarantines, with pilot programs in progress against medfly in Israel/Jordan, Portugal, South Africa, and Tunisia and against *Bactrocera* fruit flies in Philippines and Thailand.

Screwworms

In addition to the current barrier against *C. hominivorax* across the Darien region in eastern Panama, an eradication project is under way in Jamaica, and others are under assessment for Cuba and Hispaniola. As part of an Australian contingency planning to maintain Australia free of the Old World screwworm, *Chrysomya bezziana* (OWS), a pilot mass rearing facility was recently established in Malaysia. Application of SIT against OWS is also being considered in the Persian Gulf region, to where it has spread over the past decade, causing large outbreaks in Iraq, and from where it is threatening to invade the Near East and Mediterranean regions.

Moths

Field programs are in progress for prevention or suppression of pink bollworm, *Pectinophora gossypiella,* in various cotton areas in California and for codling moth, *Cydia pomonella,* suppression in apple production areas in British Columbia, Canada. In the northeastern United States, a number of successful pilot tests were carried out against the gypsy moth, *Lymantria dispar.* In addition, the SIT package is in various stages of development against other moth pests in various parts of the world, targeting diamondback moth, false codling moth, cotton bollworm, date moth, sugarcane borer, and peach borer. For moth pests there is considerable potential for expanding the implementation of inherited sterility suppression programs, which have shown a significant synergistic interaction between this type of SIT and biological control. Inherited sterility refers to the release of moths which have been given a substerilizing dose of radiation.

Tsetse

Tsetse flies and the diseases they transmit (sleeping sickness in humans and Nagana or animal trypanosomosis in livestock) prevent the establishment of sustainable agricultural systems in many areas of great agricultural potential in sub-Saharan Africa. Following pilot SIT projects in Burkina Faso and Nigeria, and breakthroughs in mass rearing and aerial release technology, an area-wide program succeeded in 1996 in eradicating *Glossina austeni* from the island of Zanzibar. This success has led to the consideration of similar programs for selected areas in mainland Africa.

ECONOMIC BENEFITS

Direct benefits of screwworm eradication to the North and Central American livestock industries are estimated to be over $1,500,000,000/year, compared with a total investment over half a century of close to $1 billion. Mexico protects a fruit and vegetable export market of over $1 billion/year through the annual investment of ca. $10 million. Achieving the medfly-free status has been estimated to have opened markets for Chile's fruit exports of up to $500 million. When implemented on an area-wide basis and with economies of scale in the mass rearing process, the application of SIT for suppression purposes is cost competitive with conventional control, in addition to its environmental benefits.

FUTURE TRENDS

The increasing constraints being applied to the use of insecticides, especially for food crops, will enhance the need for alternative control strategies for both native and introduced pest insects. SIT can be an increasingly important component of these strategies, especially when used in pest suppression programs. The expanding use of the SIT will be sustainable only when improvements in the efficiency of the technique continue to be made. This requires substantial investment in research and development related to all components of the technology, from improved insect diets to more sensitive and informative pest monitoring systems. The expansion of the use of the SIT to include suppression will be a major focus in the future, as it will provide a great opportunity for the commercialization of parts of the technology. This could range from customers simply purchasing sterile flies and developing their own release and monitoring systems to the purchase of a complete SIT package. For the medfly, these developments are imminent. Along with any commercialization of the technology, quality control procedures of both the process and the product will need to be reinforced. For the product, i.e., the sterile insect, routine protocols will have to be developed to measure the mating success of the males with females in the field. SIT programs will also

benefit tremendously if methods can be developed that enable only male insects to be reared as required. For the medfly this genetic sexing technology is now available, leading to major improvements in program efficiency. As far as the process is concerned, more appropriate artificial diets for insect adults and larvae could make a major contribution through both improved economies and improved insect quality.

In many insect species, transgenic technology has been developed and there is great interest in using it to improve the SIT. A first application could be in replacing the use of fluorescent dyes by a transgenic marker. Suitable markers are available and are currently undergoing evaluation in several pest species. A second application is the use of transgenic technology to develop molecular methods for genetic sexing. A caveat to the use of any transgenic technology will be regulatory approval; however, because the SIT uses sterile insects there is no risk of vertical transmission of the transgene.

See Also the Following Articles

Agricultural Entomology • Biotechnology and Insects • Integrated Pest Management • Rearing of Insects • Tsetse Fly

Further Reading

Hendrichs, J. (2000). Use of the sterile insect technique against key insect pests. *Sustainable Dev. Int.* **2,** 75–79.

Knipling, E. F. (1979). Use of insects for self-destruction. *In* "The Basic Principles of Insect Population Suppression and Management." USDA, Washington, DC. [Agricultural Handbook No. 512]

Krafsur, E. S. (1998). Sterile insect technique for suppressing and eradicating insect populations: 55 years and counting. *J. Agric. Entomol.* **15,** 303–317.

Tan, K.-H. (2000). "Area-Wide Control of Fruit Flies and Other Insect Pests." Pernerbit Universiti Sains Malaysia, Paulau, Pinang.

USDA National Agricultural Library. (2000). "The Screwworm Eradication Collection." Special Collections. www.nal.usda.gov/speccoll/screwworm/.

Vreysen, M. J. B., Saleh, K. M., Ali, M. Y., Abdulla, A. M., Zhu, Z.-R., Juma, K. G., Dyck, A., Msangi, A. R., Mkonyi, P. A., and Feldmann, H. U. (2000). *Glossina austeni* (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. *J. Econ. Entomol.* **93,** 123–135.

Sternorrhyncha

(Jumping Plant Lice, Whiteflies, Aphids, and Scale Insects)

Penny J. Gullan
University of California, Davis

Jon H. Martin
The Natural History Museum, London

The Sternorrhyncha is one of the suborders of the order Hemiptera; it comprises some 16,000 described species. It contains four major groups, all entirely phytophagous and usually recognized as superfamilies: the Psylloidea (psylloids or jumping plant lice), Aleyrodoidea (whiteflies), Aphidoidea (aphids), and Coccoidea (scale insects or coccoids). The name Sternorrhyncha (from the Greek *sternon,* meaning chest, and *rhynchos,* meaning nose or snout) refers to the location of the mouthparts on the underside of the insect, between the bases of the front legs, although sometimes the mouthparts are lacking in the adult.

OVERVIEW OF STERNORRHYNCHA

Insects belonging to the Hemiptera are unique in having their mouthparts forming a rostrum that comprises mandibles and maxillae modified as needle- or thread-like stylets lying in a grooved labium. Two pairs of stylets interlock to form two canals, one delivering saliva and the other taking up plant or animal fluid. The Sternorrhyncha is widely accepted as the sister group to the rest of the Hemiptera. It is a well-defined group diagnosed by the position of the rostrum and the presence of only one or two segments on each tarsus (the part of the leg most distant from the body and bearing one or two claws); in most other hemipterans there are three tarsal segments.

Sternorrhynchans use their stylets to probe plant tissues intracellularly (into or through plant cells) or intercellularly (between plant cells). The tips of the stylets always enter cells at the site of ingestion, which is often phloem-sieve elements. Generally, stylet penetration is accompanied by secretion of a solidifying saliva that forms a sheath around the stylets. Other hemipterans mostly probe intracellularly, may or may not secrete salivary sheaths, and ingest from a wider variety of plant or animal tissues. Most sternorrhynchans are phloem feeders, and thus have a diet rich in carbohydrates (sugars) and deficient in amino acids and other nitrogenous compounds. Generally, there is an intimate association with intracellular bacteria, called endosymbionts, which are housed in special tissue (bacteriomes or mycetomes) and contribute nutrition to the insect host.

Sternorrhynchan excreta is often a sticky, sugary liquid called honeydew that often contaminates foliage; it serves as a substrate for the growth of black sooty mold fungi that can impede photosynthesis and reduce plant vigor. Honeydew often attracts ants that may protect the sternorrhynchans from their natural enemies, especially predatory and parasitic insects.

Adult jumping plant lice, whiteflies, and many adult aphids have two pairs of wings. All adult female scale insects and most adult female aphids are wingless (apterous). Adult male scale insects usually resemble small flies (order Diptera) in having the hind wings reduced to small balancers (halteres), although sometimes the halteres are absent; in a few scale insect species males are apterous. The absence of wings in adult females of scale insects and many aphids means that the immature stages (nymphs) of these groups can be difficult to distinguish from their adults.

Evolutionary interpretations of morphology have suggested either that whiteflies are sister to aphids + scale insects, with

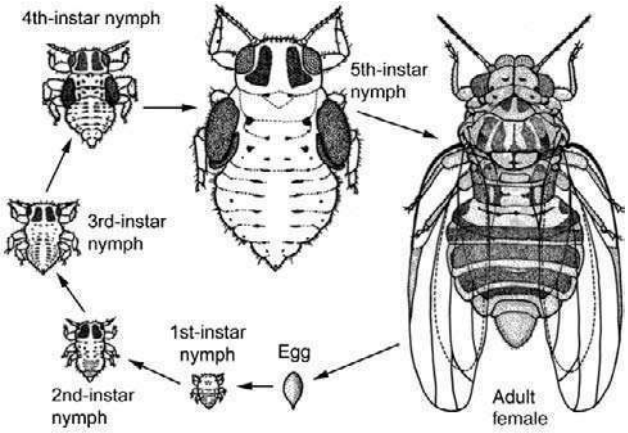

FIGURE 1 The life cycle of *Cardiaspina fiscella* (Psylloidea: Psyllidae). [Adapted with permission from K. L. Taylor, 1962, *Aust. J. Zool.* **10(2),** 310.]

psylloids sister to all three, or that whiteflies + psylloids are sister to aphids + scale insects. Phylogenetic analysis of nucleotide sequences (of small subunit rDNA, also called 18S rDNA) supports the former hypothesis that Psylloidea is the sister group to the other sternorrhynchans. Some of the morphological traits shared by Aleyrodoidea and Psylloidea are ancestral features (plesiomorphies), such as two, similar-sized, tarsal segments (the first tarsal segment is reduced or absent in aphids and coccoids) and good jumping ability. Also both groups have pedunculate eggs, i.e., one end of the egg has a short stalk or narrow extension (Fig. 1). Whiteflies, aphids, and scale insects all have Malpighian tubules (filamentous excretory organs) reduced in number or absent and their gut often has a well-developed filter chamber that allows most of the water in ingested sap to bypass the absorptive part of the midgut. In contrast, psylloids have four Malpighian tubules (an ancestral feature found in other Hemiptera) and possess only a rudimentary filter chamber.

PSYLLOIDEA (JUMPING PLANT LICE)

Worldwide, there are about 2500 described species of psylloids or jumping plant lice. Sometimes the common name "psyllids" is applied to all jumping plant lice but this name may cause confusion because it is also the common name for one family of psylloids, the Psyllidae. The greatest abundance and species richness of psylloids are found in tropical and south temperate regions, where many new species await discovery. In the Northern Hemisphere, for example, fewer than 100 species occur in Britain and about 300 species are known from North America. Adult psylloids are small, ranging in length from 0.2 to 8.0 mm, and superficially resemble miniature cicadas or small leafhoppers, but are distinguished by their multisegmented antennae. Their common name derives from the ability of adults to jump backward when disturbed. All psylloids feed by sucking sap, mostly of woody dicotyledonous plants. They ingest sap from a variety of tissue sources, including phloem, xylem, and mesophyll parenchyma, and thus are not phloem specialists like most other sternorrhynchans. Psylloid nymphs are either free living, gall inducing, or lerp forming; lerps are manufactured sugary and starchy covers or scales under which the nymphs live. Some species of jumping plant lice are pests of culivated plants, with damage resulting from loss of plant sap, toxins in the saliva, and/or disease caused by transmitted microorganisms.

Evolution and Classification

Historically, the classification of psylloids was based on the well-known, but relatively species-poor, Holarctic fauna. Expanding knowledge of the diverse tropical and south temperate faunas has led to better understanding of relationships and hence a more natural classification. The system used here is based on that of White and Hodkinson presented in 1985, with emendations suggested by D. Burckhardt and D. Hollis. In this system, six families of extant psylloid species are recognized (with each of the first four families listed here having 10 or fewer genera): Calophyidae, Carsidaridae, Homotomidae, Phacopteronidae, Psyllidae (includes the Aphalaridae, Liviidae, and Spondyliaspididae, recognized by some authors; about 120 genera), and Triozidae (over 40 genera).

Many species of the predominantly Neotropical and tropical/subtropical Asian family Calophyidae feed on plants in the Anacardiaceae. Carsidaridae is almost entirely pantropical in distribution and is restricted to the related plant families Bombacaceae, Malvaceae, and Sterculiaceae. The Homotomidae is predominantly from the Old World tropics and feeds almost exclusively on figs (Moraceae: *Ficus*). The Phacopteronidae occurs in tropical Asia, Africa, and the Americas and many species feed on plants in the Meliaceae. The Psyllidae is the largest family, but it is heterogeneous morphologically. A large number of psyllids are associated with legumes (Fabaceae), although the speciose subfamily Spondyliaspidinae specializes on *Eucalyptus* (Myrtaceae). Many species of Spondyliaspidinae produce lerps or induce leaf galls that they plug with lerp substance. The Triozidae is cosmopolitan and has wide-ranging host preferences.

The fossil record of the Psylloidea extends back into the Permian (>270 mya). The families Cicadopsyllidae, Liadopsyllidae, and Protopsyllidiidae are known only as fossils, and the extant families probably diversified in the Cretaceous. Relationships among psylloid higher groups (families and subfamilies) are largely unresolved.

Life History

Reproduction is always sexual. There are seven life stages (Fig. 1): the egg, which is stalked; five instars; and the adult, with the sexes in about equal numbers. Eggs are either completely inserted into plant tissue or inserted by just the stalk (peduncle) through which water is absorbed. Depending on species, each female may lay as few as 20 to more than 1000 eggs. In cold temperate regions, psylloids frequently have only

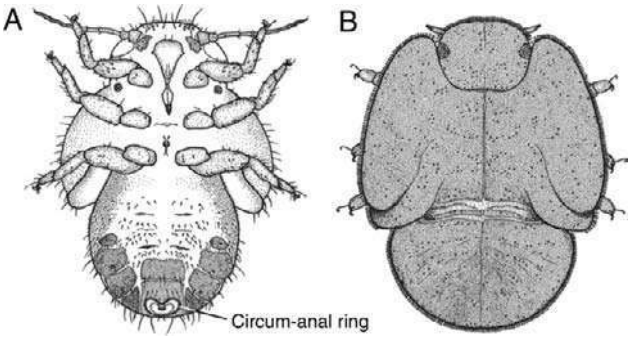

FIGURE 2 Nymphs of jumping plant lice (Psylloidea). (A) The underside, or venter, of a nymph of *Padaukia kino* (Psyllidae). (Illustration by J. H. Martin; reproduced, with permission, from D. Hollis and J. H. Martin, 1993, *Bull. Entomol. Res.* **83,** 207.) (B) The upper side, or dorsum, of a nymph of *Trioza remota* (Triozidae). [Illustration by J. H. Martin; reproduced, with permission, from W. R. Dolling, 1991, "The Hemiptera," p. 171, British Museum (Natural History), London.]

a single generation per year (univoltine). In warmer climates, development continues year round, often with several to many generations per year. Species with multiple generations per year (polyvoltine) are more likely to have population outbreaks than univoltine or bivoltine species.

Morphology

Identification is based primarily on adult characteristics, with males being most important for species-level determinations, although features of the female terminalia are sometimes used. The fifth instar can materially aid identification in many groups. Adult males and females are morphologically similar, except that males are often smaller and have different genitalia and thus different abdominal shapes. In a few species the sexes are color dimorphic, and there is often a seasonal color dimorphism. Body coloration or the pattern of markings can be diagnostic of species. The size and shape of the head and its appendages provide useful taxonomic characters. Immature psylloids (Fig. 2) are usually characteristically flattened dorsoventrally and many secrete wax.

Behavior and Ecology

Adult psylloids disperse over short distances by jumping or flying, but many species may travel long distances on air currents. The adults of a number of species are known to stridulate as a method of mate attraction. Mating behavior is often rudimentary, but precopulatory darting movements and wing flicking have been observed in some species of Spondylaspidinae.

The habits of nymphs vary among the different taxonomic groups. Nymphs of many species are naked, although often covered in waxy body secretions, and live exposed on the shoots of their host plant. Others induce galls of various complexity, from simple pits in leaves to round or elongate structures that totally enclose the nymphs. A third lifestyle is for the nymphs to develop on foliage but protected under scale-like covers called lerps. Each lerp is formed of substances eliminated from the insect's gut (a special kind of liquid "feces" that is different from honeydew) or of a mixture of "feces" intermingled with wax filaments (exuded from pores in the anal plate). The "feces" of at least some species are composed of starch, especially amylose and amylopectin, synthesized in the insect's gut. The insect builds the lerp by moving its abdomen back and forth while squeezing out the liquid excreta (like glue from a tube), which rapidly hardens to form a lacework of filaments. Lerps have various shapes and sizes, from simple cones to fans or lacy shell-like structures, and often are of diagnostic use. They vary in color from white to cream or brown and may be opaque or transparent. Some lerp-formers (e.g., *Glycaspis* spp.) produce copious honeydew, whereas others (e.g., *Cardiaspina* spp.) produce very little or none.

Most species are specialist feeders as nymphs, with many species restricted to one plant genus or even a single species and often to particular host structures (leaves, new shoots, etc.) or growth stages (either young or mature foliage). Almost all psylloids develop on dicots, a few on monocots or conifers, but none on ferns. In cool-temperate climates, some species overwinter as adults on evergreens, returning to their true hosts in spring, when eggs are laid. Host selection appears to involve taste rather than olfaction. Adults frequently are less discriminating than nymphs and thus host plants are defined as those on which nymphs can complete their development to adulthood. Nymphal feeding frequently damages the host plant (e.g., premature leaf senescence), whereas adult feeding usually causes little injury. Large natural increases, or outbreaks, in psylloid populations appear to result from changes in the suitability of their food supply, rather than from reduction in numbers of predators or parasitoids. Many psylloid species have a mutualistic association with honeydew-seeking ants.

Notable Pest Species and Their Control

Some species of jumping plant lice have become pests of plantation trees. For example, since the mid-1990s, several species of Australian Psyllidae have become pests of South American plantations of *Eucalyptus.* In particular, the eucalypt shoot psyllid *Blastopsylla occidentalis,* the blue gum psyllid *Ctenarytaina eucalypti,* and *Ctenarytaina spatulata* have caused considerable damage. In western North America, Australian psylloid pests appeared in the early 1980s, with *B. occidentalis, C. eucalypti,* the tristania psyllid *Ctenarytaina longicauda,* the red gum lerp psyllid *Glycaspis brimblecombei,* and the lemon gum lerp psyllid *Eucalyptolyma maideni* all established in California.

Two of the world's most serious citrus pests are the Asian citrus psylla (or psyllid), *Diaphorina citri* (Psyllidae), and the African citrus psylla, *Trioza erytreae* (Triozidae). Damage to citrus derives from direct feeding (removing sap), development of sooty mold on honeydew, and, most importantly, from transmission of phloem-limited bacteria (*Liberobacter africanum* and *L. asiaticum*) that cause a disease called citrus greening.

The most characteristic symptom is a yellow-green mottling associated with dieback of shoots and leaves, followed by subsequent deterioration of fruit quality. Adults and nymphs of both psylloid species can transmit both diseases.

Other than the citrus pests, the most economically important psylloids of temperate and subtropical areas feed on cultivated apple and pear trees. *Cacopsylla mali* (Psyllidae) and related species causes various kinds of damage to apples. Pears also host a suite of *Cacopsylla* species, especially *Ca. pyricola,* which can transmit the virus that causes pear decline.

Trioza species (Triozidae) occur as pests on persimmons, cultivated fig, and carrots. *Paratrioza cockerelli* (Triozidae) feeds on potatoes, tomatoes, and other Solanaceae; nymphal saliva contains phytotoxins that cause "psyllid yellows" in potato (reducing tuber growth) and sometimes in tomato. The leucaena louse, *Heteropsylla cubana* (Psyllidae), which is native to the Caribbean, has become a pest of the tropical leguminous tree *Leucaena leucocephala* in several parts of the world where it is variously planted for timber, fuel, animal fodder, or shade for other crops and or because it fixes nitrogen and tolerates drought.

Recorded parasitoids are mostly psylloid specialists, but there is little evidence of parasitoid–host specificity. Worldwide a number of wasp species, especially from the families Encrytidae (e.g., *Psyllaephagus* spp.) and Eulophidae (e.g., *Tamarixia* spp.), are parasitic on psylloids. In Europe, cecidomyiid flies (*Endopsylla*) are parasitic on some adult psylloid pests (e.g., *Cacopsylla* spp.).

ALEYRODOIDEA (WHITEFLIES)

Members of the Aleyrodoidea derive their common name from the powdery, white waxy secretion preened over the body and wings of most adults. The Greek root *aleuro-,* found in many whitefly names, means flour. Adults of both sexes are tiny, delicate, and free flying. They have a wingspan of up to 4 mm, but usually about 2 mm, and resemble minute moths. They are a familiar sight to many home gardeners because, when infestations are heavy, the adults will fly up en masse if disturbed from the foliage of favored host plants. The group occurs worldwide, although few whiteflies occur in the cooler temperate regions. About 1450 species are described, with perhaps two or three times as many species awaiting collection and formal taxonomic study. All adults and nymphs feed by ingesting phloem sap and several species are serious plant pests.

Evolution and Classification

All whitefly species belong to one family, Aleyrodidae, which is divided into just two extant subfamilies, the Aleurodicinae and Aleyrodinae. The Aleurodicinae contains 17 genera and just over 100 described species, mostly from the Neotropics (Central and South America and the Caribbean). All other species (in more than 110 genera) belong to the Aleyrodinae, which has a mostly pantropical distribution. These two subfamilies are defined on both adult and nymphal features, whereas species and genera are diagnosed mainly or entirely on characters of the fourth instar, usually known as a puparium. The Aleurodicinae generally have larger adults with more complex wing venation than those of the Aleyrodinae (Fig. 3).

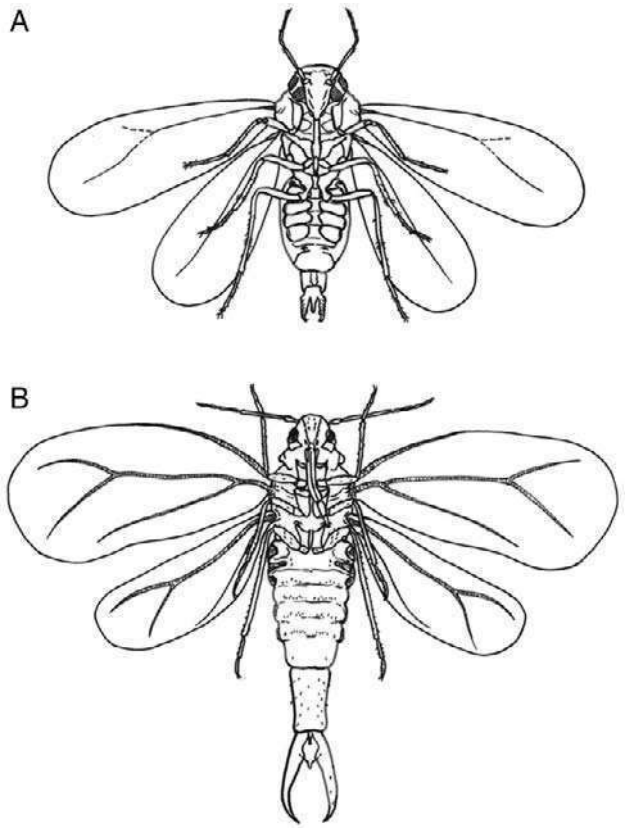

FIGURE 3 Adult male whiteflies (Aleyrodoidea: Aleyrodidae), ventral views. (A) *Trialeurodes vittata,* subfamily Aleyrodinae. (Unpublished illustration by R. J. Gill, adapted with permission.) (B) *Aleurodicus* sp., subfamily Aleurodicinae. [Adapted with permission from R. J. Gill, 1990, "Whiteflies: Their Bionomics, Pest Status and Control" (D. Gerling, ed.), p. 31, Intercept Ltd., Andover, UK.]

Whiteflies are rarely fossilized because they have small and delicate bodies. The known fossils are preserved mostly in Cretaceous and Tertiary amber with one record from the Upper Jurassic. The two modern subfamilies probably diverged during the Cretaceous, whereas the Aleyrodoidea may have originated in the Late Permian or even earlier.

Life History

Reproduction is usually sexual, with fertilized eggs producing females and unfertilized eggs producing males (arrhenotoky or arrhenotokous parthenogenesis); unmated females lay only haploid eggs. A few species produce only females (thelotoky or thelotokous parthenogenesis). There are six life stages (Fig. 4): the egg stage; the first instar, in which the nymph is often called a crawler; the second and third instars, in which the nymphs are sessile; the fourth instar or last nymphal stage, at the end

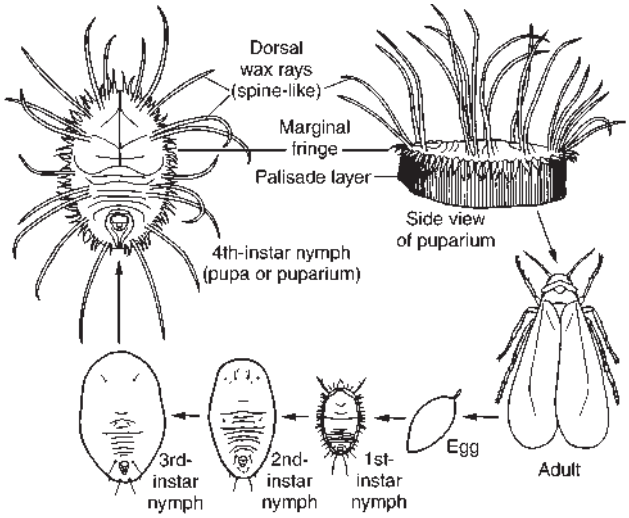

FIGURE 4 Life cycle of the greenhouse whitefly *T. vaporariorum* (Aleyrodoidea: Aleyrodidae). [Adapted with permission from R. J. Gill, 1990, "Whiteflies: Their Bionomics, Pest Status and Control" (D. Gerling, ed.), p. 15, Intercept Ltd., Andover, UK.]

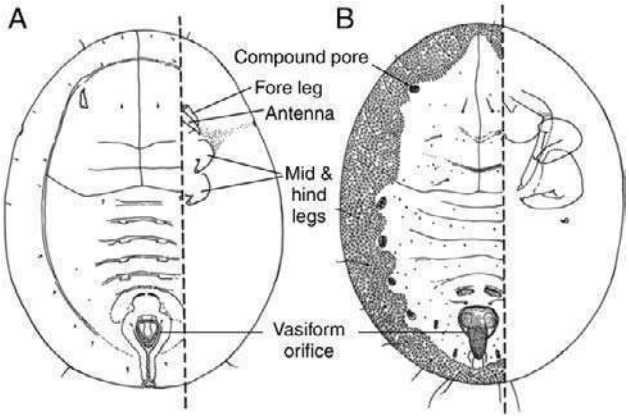

FIGURE 5 Whitefly puparia, the last nymphal stage (Aleyrodoidea: Aleyrodidae), with dorsum drawn to left and venter to right. (A) *Aleurolobus marlatti,* subfamily Aleyrodinae. [Reproduced, with permission, from J. H. Martin, 1999, "The Whitefly Fauna of Australia (*Sternorrhyncha:* Aleyrodidae): A Taxonomic Account and Identification Guide," p. 172, CSIRO, Canberra.] (B) *Aleurodicus cocois,* subfamily Aleurodicinae. (Adapted with permission from J. H. Martin, 1997, *Mem. Mus. Victoria* **56,** 126.)

of which the insect is usually referred to as a "pupa," or a "puparium," (Fig. 5) although it is not a true pupa as seen in holometabolous insects as there is no molt to a totally non-feeding, "resting" stage; and, finally, the adult (or imaginal) stage. After the emergence of the adult, the empty cuticle (exuviae) is often called a "pupal case."

Whitefly eggs are usually attached to the underside of leaves by a short stalk, the pedicel, through which water is absorbed from the plant. Eggs are often laid in circles or arcs, and egg batches are frequently conspicuous because of a dusting of white wax. First instars usually settle adjacent to the eggs from which they hatched. In most species, development occurs on the underside of leaves. After the first molt the nymphs become immobile and cannot move to a new site if food quality deteriorates. During the fourth instar, feeding ceases and the insect transforms into the winged adult.

Species such as the greenhouse whitefly (*Trialeurodes vaporariorum*) can breed continually in greenhouses or indoors. Those whitefly species that have economic impact have several to many generations per year, but many species are thought to have only one or two generations annually.

Morphology

Historically, puparial rather than adult characteristics have been used to recognize species and genera, although adults do display many taxonomically useful characteristics. Seasonal dimorphism, as seen in many aphid species, is uncommon but occurs in the puparia of a few temperate species. Male (Fig. 3) and female whiteflies are similar in appearance except for their genitalia, but males are often smaller than females of their own species. The body and wings are dusted with powdery wax emanating from large, paired wax plates on the underside of the abdomen and thus adults most commonly appear whitish or grayish even if the body color is yellow, brown, or red or if they are darkly sclerotized. The male genitalia (the claspers or parameres and the intromittent organ or aedeagus) are more informative taxonomically than the female terminalia (ovipositor and associated structures).

Nymphs of all stages (Fig. 4) superficially resemble scale insects (Coccoidea) and frequently have ornate wax secretions in later instars. First instars have functional legs, but nymphs of the subsequent three instars have vestigial appendages and are immobile. The anus of all nymphal stages (and adults) opens in a special pit called the vasiform orifice (Fig. 5), located dorsally near the posterior end of the abdomen. It comprises a depression with a sclerotized border, a dorsal flap (the operculum), and a tongue-shaped structure (the lingula). The anus opens under the operculum, which covers at least the base of the lingula; each droplet of anal excreta (honeydew) that accumulates is catapulted away by an upward flick of the lingula, which is the only mobile body part in instars II to IV.

Puparia are mostly oval or elongate-oval, 0.5 to 2.0 mm in length, with the body margin either smooth or variously sculptured. Color varies from transparent to white, brown, or black. Waxes also contribute to the puparial appearance and may be transparent, translucent, iridescent, or opaque, usually colored white or grayish; mechanical color (iridescence) may be blue, turquoise, bottle-green, or pearly white. Transparent wax occurs in a thin layer over the body or as a marginal fringe and/or dorsal "spines." Opaque wax may occur in powdery deposits, loose mats, or filaments, or be defined as tufts or rays, and is exuded from a variety of pores and tubes in the cuticle.

Behavior and Ecology

Adult males of a few whitefly species, especially *Bemisia tabaci* and *T. vaporariorum,* have been observed to display courtship

behavior prior to mating, including abdominal oscillations that result in acoustic signals caused by substrate-borne vibrations. Females may produce sex pheromones to attract males. Adult whiteflies will fly short distances, if disturbed from their host plant. They also undertake longer migratory flights, which are air-current dependent because whiteflies are weak fliers. They have complex host-finding and host-orientation behaviors, involving at least attraction to particular colors, especially yellow or yellow-green.

Whiteflies feed from plant vascular tissue and all feeding stages produce copious quantities of honeydew, yet relatively few species are ant attended. Most whitefly species appear to be oligophagous, with few known to be monophagous; however, polyphagous species are the best documented because they are most likely to be pests. Whitefly host plants are almost entirely flowering plants (angiosperms), especially woody dicots; relatively few species of whiteflies are found on herbs, grasses, ferns, or palms.

The morphology of the puparia of some species has been shown to vary depending on the species of plant that is acting as host, probably because of differences in the nature of the plant surfaces. For example, the size, number, and position of setae of the puparia can be correlated with leaf hairiness in some species (e.g., *B. tabaci*). These phenotypes of the same species often look very different from each other and this phenomenon can confound identification. This variation is especially a problem among polyphagous species in the genera *Bemisia* and *Trialeurodes.*

Notable Pest Species and Their Control

The feeding activity of many whiteflies damages their host plants and some species are known or believed to transmit diseases, especially those caused by viruses. Feeding can cause excessive sap loss as well as physiological changes in host tissues, such as leaf discoloration, wilting and premature shedding, and fouling by honeydew. The most serious aleyrodid pests are those of orchards (especially citrus), greenhouse crops, and vegetable field crops, and a few pest species are almost cosmopolitan in distribution. The past 2 decades have seen an increase in the pest status of certain whiteflies, probably because of the development of aggressive strains that are more fecund, are more efficient virus vectors, and/or have broader host ranges. Successful whitefly control depends upon an appropriate integrated pest management program in which natural enemies are protected or augmented. A number of groups of parasitic wasps (particularly Eulophidae and Aphelindiae) specialize on whiteflies, and useful predators include some ladybird beetles (Coccinellidae). Whiteflies are attacked also by pathogenic fungi, especially *Aschersonia* species.

One of the most important pests is the polyphagous tobacco or sweetpotato whitefly, *B. tabaci,* which occurs predominantly in tropical and subtropical regions and on protected crops elsewhere. It feeds on numerous fiber, food, and ornamental plants, and damage is exacerbated by its ability to vector more than 70 different viruses. *B. tabaci* exists in many biotypes or strains, and one virulent form, biotype B, sometimes is recognized controversially as a separate species, named *Bemisia argentifolii* (the silverleaf whitefly). Effective biological control of *Bemisia* whiteflies is possible using host-specific parasitoids, such as wasps of *Encarsia* and *Eretmocerus* species (Aphelinidae).

The polyphagous greenhouse whitefly, *T. vaporariorum,* occurs worldwide in greenhouses and on indoor plants; the aphelinid wasp parasitoid *Encarsia formosa* is an important biological control agent. *Aleurodicus destructor* and spiraling whitefly, *A. dispersus,* are polyphagous pests with abundant wax secretions covering the puparia. Spiraling whiteflies are parasitized by wasps of *Aleuroctonus vittatus* (Eulophidae) and *Encarsia* species, including *E. guadeloupae* (Aphelindiae), which can reduce the whitefly populations. *Siphoninus phillyreae,* perhaps of Middle Eastern origin, and *Parabemisia myricae* from Japan are pests of fruit trees and have extended their ranges into the Mediterranean region and parts of the United States. The Indian species *Vasdavidius* (formerly *Aleurocybotus*) *indicus* now attacks rice in West Africa. The woolly whitefly, *Aleurothrixus floccosus,* which is endemic to the Americas, has been introduced accidentally into Africa and Southeast Asia, where it is a pest, especially of citrus. *Dialeurodes citri* and *Singhiella* (formerly *Dialeurodes*) *citrifolii* are widespread pests of citrus, with *D. citri* attacking a wider range of other plants than does *S. citrifolii.*

APHIDOIDEA (APHIDS)

Aphids are small soft-bodied insects, ranging from 1 to 8 mm in length. They are usually found living in aggregations on rapidly growing parts of their host plants and occur predominantly in the northern temperate regions of the world. More than 4400 species are known. The life cycles of aphids are frequently complex and usually include parthenogenetic (or asexual), but often also sexual, reproduction; many species display host alternation in which cyclical parthenogenesis is combined with the obligate use of two unrelated host plants. Also many aphids produce either eggs or living young at different parts of the cycle. Adults may be winged (alatae) or wingless (apterae). Most aphids can increase their population size rapidly because of parthenogenetic production of live young combined with "generational telescoping," whereby a mother aphid carries both her daughters and their daughters (i.e., embryos within embryos). Aphid feeding activities can have deleterious effects on their host plants mainly via sap removal and/or virus transmission.

Evolution and Classification

The Aphidoidea contains three families: Phylloxeridae, Adelgidae, and Aphididae. Strictly speaking, the common name of the entire group should be "aphidoids," but "aphids" is used almost universally as a collective name. Very occasionally the name "aphids" has been applied to just members of

the Aphididae, which should more properly be referred to as "aphidids."

The Phylloxeridae contains about 75 species, in eight genera, divided between two tribes. The Phylloxerini has seven genera, feeding mostly on oaks (Fagaceae: *Quercus*), or hickory and pecans (Juglandaceae: *Carya*), but with one pest species on grape vines (Vitaceae: *Vitis*). The Phylloxerinini has a single genus associated with the willow family (Salicaceae). The Phylloxeridae is probably the oldest extant family, although only one fossil (*Palaeophylloxera* from the Lower Miocene) is known. Many phylloxerid species induce galls on their host plants; only a few species are host alternating.

Conifer woolly aphids, the Adelgidae, belong to two genera and about 50 species and are entirely Holarctic in distribution. Adelgids induce galls on their primary host, spruce (*Picea*), and move to other conifers as alternate hosts.

The largest family, Aphididae, has been divided into a number of subfamilies and tribes, but different authors frequently use different classifications. The 10 subfamilies listed here are those recognized in the monographs by R. L. Blackman and V. F. Eastop: Anoeciinae (34 species), Aphidinae (over 2700 species), Calaphidinae (including the Drepanosiphinae, Thelaxinae, and several other groups treated as subfamilies by some other authors; 400 species), Chaitophorinae (164 species), Eriostomatinae (formerly Pemphiginae; 319 species), Greenideinae (151 species), Hormaphidinae (176 species), Lachninae (355 species), Mindarinae (5 species), and Phloeomyzinae (1 species).

The oldest aphid fossils are from the Triassic (at ca. 230 mya) but aphids may have originated in the Permian. Phylogenetic analysis of molecular data suggests that aphids underwent a rapid radiation into the current tribes after shifting from gymnosperms to angiosperms some time during the Upper Cretaceous. Furthermore, the ancestral aphid probably had a simple life cycle with host alternation evolving independently in each of the families and perhaps several times within the Aphididae.

Life History

There are six stages in the life history of an individual aphid: the egg or embryonic stage, four instars, and the adult. Aphids have evolved a range of annual or biennial life cycles and other adaptive strategies that often vary within as well as among species. This complexity can complicate the study of aphid biology. Furthermore, aphidologists have developed a special nomenclature for the different life stages and types. Life cycles are called holocyclic (Fig. 6) if a sexually reproducing generation is present or anholocyclic if the sexual generation is absent. A typical complete cycle (holocycle) consists of a single generation of sexual morphs (sexuales) and several to many generations of only parthenogenetic females. In the sexual generation, males mate with females to produce fertilized eggs, which are all female. Each egg gives rise to a wingless "foundress" or "stem mother" (fundatrix) that gives rise to a lineage of parthenogenetic females. The parthenogenetic descendants of a single foundress are called collectively a clone and are identical genetically but may be of different

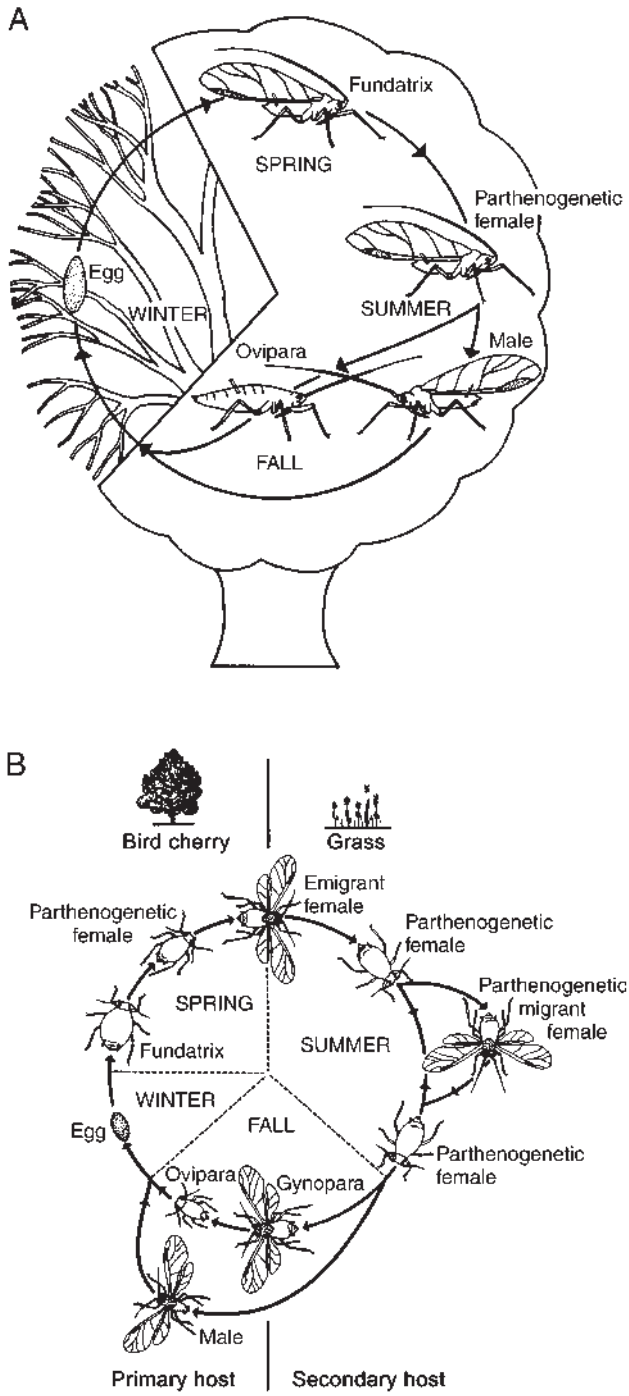

FIGURE 6 Life cycles of aphids. (A) The holocyclic life cycle of the monoecious sycamore aphid, *Drepanosiphum platanoidis.* (B) The holocyclic life cycle of the heteroecious bird cherry–oat aphid, *Rhopalosiphum padi.* Fundatrix, foundress, a wingless parthenogenetic female that produces live offspring rather than eggs; gynopara, migrant that flies back to the primary host where she produces oviparae; ovipara, wingless, sexual egg-laying female that mates with a male. (Reproduced with permission from A. F. G. Dixon, 1973, "Biology of Aphids," pp. 8 and 9, Edward Arnold Ltd., London.)

phenotypes (morphs). The adults may be apterae or alatae, depending on environmental conditions. In Aphididae, parthenogenetic females (viviparae) give birth to live young, whereas the sexual females (oviparae) lay eggs. In Adelgidae and Phylloxeridae, females lay eggs during both the sexual and the asexual phases of reproduction. The phenomenon of cyclical parthenogenesis is a key feature of aphid biology. Some aphids have lost, or can facultatively lose, the sexual part of the life cycle (a condition then called anholocycly). It is common for aphids to have 15 to 20 generations per year, with up to 40 in tropical climates.

Most aphids are monoecious (Fig. 6A), i.e., they undergo all phases of their life cycle on one host plant or on a small number of closely related hosts. About 10% of aphids are heteroecious (Fig. 6B), having more complex life cycles involving host alternation (heteroecy). The sexual morphs mate and the oviparae lay eggs on the primary host (often a woody plant); however, a regular move occurs to another, unrelated plant, the secondary host (which may be either herbaceous or woody), on which the parthenogenetic generations live. The aphids must return to their primary host for the next sexual generation. Host alternation may occur as part of a 1-year (annual) life cycle, as in many Aphididae, or as a 2-year (biennial) life cycle, as in Adelgidae. In the Aphididae, in temperate climes, alternation is frequently obligate, overcoming the twin problems of poor sap flow in woody primary hosts in summer and the death of many herbs in winter.

Morphology

Aphids are highly polymorphic, with most species occurring in several different forms or morphs. In individuals destined to be winged as adults, the wing buds are usually apparent after the second nymphal molt. Most common aphid species are soft-bodied and green in color, but dark and brightly colored species and a few hard-bodied species also occur. Nymphs superficially resemble adults except that they are smaller in size and never have wings. Aphids usually have two prominent, tube-like structures on the posterior dorsum of the abdomen, called siphunculi or cornicles (Fig. 7), which can discharge defensive lipids and alarm pheromones. In some species, these structures are reduced to pores or are absent. Another structure unique to the aphids is a posterior projection on the tip of the abdomen called the cauda, equivalent to the lingula in whiteflies. Protective wax secretions are common.

Behavior and Ecology

Both adults and nymphs can travel short distances by walking, but dispersal of alatae occurs by both active flight and passive long-distance movement on air currents. Apterae and nymphs are frequently transported by attendant ants. In sexual morphs, males are attracted to females by sex pheromones released from glands on the female's legs. Altruistic behavior has been reported in the few aphid species that have a sterile soldier "caste," which is a special kind of first or second instar that defends the family group against competitors and natural enemies.

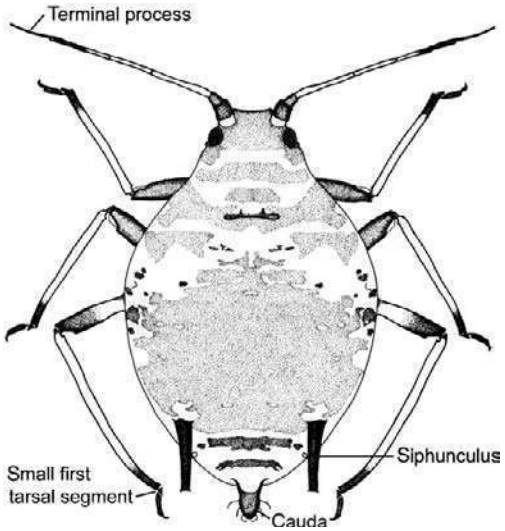

FIGURE 7 An apterous female cowpea aphid, *Aphis craccivora* (Aphidoidea: Aphididae). [Illustration by J. H. Martin; reproduced with permission from W. R. Dolling, 1991, "The Hemiptera," p. 193, British Museum (Natural History), London.]

Aphids generally live in aggregations on the buds, stems, and/or leaves of their host plants. Some species induce galls and a few others live underground on roots. Almost all aphids are phloem feeders, producing copious quantities of honeydew, and thus many species are attended by ants. Some aphids feed from both phloem and parenchyma tissue, and adelgids are largely parenchyma feeders. In general, aphids are quite host specific, with each genus being associated with a particular host-plant family and each species with one genus or species of host plant or at least closely related plant genera. In host-alternating species, however, the primary and secondary hosts are usually unrelated and the specificity to the primary host is generally higher than to the secondary host(s). Pest aphids, especially those of agricultural crops, often feed on plants in a number of unrelated families.

Notable Pest Species and Their Control

Infestations of aphids can grow rapidly to enormous size and cause plant debilitation through nutrient deprivation, but even small numbers of aphids can transmit viruses to uninfected plants. Pest aphids often are more polyphagous than nonpest species and frequently are species exotic to the area where they are of most economic concern. Cosmopolitan pest species are frequently anholocyclic, i.e., reproducing continually by parthenogenesis on their crop hosts, and thus are particular pests in the tropics.

One of the most notorious aphidoid pests is the grape phylloxera, *Daktulosphaira vitifoliae* (Phylloxeridae), which induces galls on *Vitis* species, including the viticulturally important *V. vinifera.* Damage in vineyards results mainly

from the rotting of roots when infestations are heavy, rather than sap loss per se. The main method of control is the use of resistant grape rootstocks.

Genera of Aphididae that contain significant pests of crops include *Aphis, Brevicoryne, Macrosiphum, Myzus,* and *Rhopalosiphum;* the species listed below all transmit plant viruses. There are a number of serious pest species of *Aphis,* including the cowpea aphid *A. craccivora* (Fig. 7), bean aphid *A. fabae,* and melon or cotton aphid *A. gossypii.* The cabbage aphid, *B. brassicae,* attacks members of the Cruciferae. The potato aphid *Ma. euphorbiae* and the rose aphid *Ma. rosae* both use roses as primary hosts. The polyphagous shallot aphid *My. ascalonicus* appears to be exclusively anholocyclic. The primary hosts of black cherry aphid *My. cerasi* and green peach aphid *My. persicae* are *Prunus* fruit trees; secondary hosts include a range of economically important plants. The corn leaf aphid *R. maidis* is a worldwide pest of cereal crops, particularly maize, sorghum, and barley, and is habitually anholocyclic in most places. Important tree pests include *Pineus* species (Adelgidae) on pines; *Cerataphis fransseni* (Hormaphidinae) on coconut and other palms; *Cinara* species (Lachninae) on a wide range of conifers; *Dysaphis, Myzus,* and *Brachycaudus* species (Aphidinae), many of which utilize rosaceous trees as their primary hosts, causing unsightly leaf curling/galling in spring; and *Eriosoma lanigerum* (Eriosomatinae), which develops woolly colonies on apple branches and trunks.

The most common aphidophagous predators are ladybird beetles (Coccinellidae), lacewings (Neuroptera), some hover flies (Syrphidae), a few gall midges (Cecidomyiidae, especially the widespread *Aphidoletes aphidimyza*), and certain predatory bugs (Anthocoridae). Wasps of the large family Aphidiidae and several genera of Aphelinidae are endophagous parasitoids of Aphididae; Adelgidae and Phylloxeridae are not parasitized.

COCCOIDEA (SCALE INSECTS)

The scale insects (also called coccoids) occur worldwide. They are mostly small (less than 5 mm in length) and often cryptic in habit. There are estimated to be about 7500 described species. Many scale insects are economically important pests of agriculture, horticulture, and forestry. Male scale insects display complete metamorphosis, whereas female development is neotenous (adults resemble nymphs). A number of taxa display remarkable diversity in their genetic systems (e.g., parthenogenesis, hermaphroditism, and paternal genome elimination) as well as in chromosome number, sperm structure, and types of endosymbioses. The name "scale insect" derives both from the frequent presence of a protective covering or "scale" and from the appearance of many of the female insects themselves. Most species produce a waxy secretion that covers the body either as a structure detached from the body (a scale or test) or as a substance that adheres to the body surface. Some coccoid species have been used as sources of candle wax, lacquers such as shellac, or dyes. Scale insects are more diverse in terms of major evolutionary lineages (families), species richness, and morphology than any of the other sternorrhynchan groups.

Evolution and Classification

Scale insects have been assigned variously to 20 or more families, but the family-level classification is controversial. Often the Coccoidea is divided into two major, informal groups, the archaeococcoids (or archaeococcids) and the neococcoids (or neococcids). The extant archaeococcoids comprise the Margarodidae *sensu lato* (with over 400 species and sometimes treated as several families), Ortheziidae (ensign scales; about 155 species), Carayonemidae (4 species), Phenacoleachiidae (2 species), and perhaps the genus *Puto* (about 60 species, sometimes placed in their own family, Putoidae, or otherwise in Pseudococcidae). Collectively, the above archaeococcoid families number only approximately 80 genera and 600 species. Some of the morphological features that define the archaeococcoids occur more widely in the Hemiptera, and monophyly of archaeococcoids is uncertain.

The neococcoids, which comprise all of the other extant families (usually 17 are recognized) and most of the species of scale insects (about 7000), are a monophyletic group characterized by derived features, including a chromosome system involving paternal genome elimination, needle-like apical setae on the labium, and loss of abdominal spiracles. Among neococcoids, most families (with the exception of the Eriococcidae) are well characterized morphologically. In contrast, relationships among families are largely unknown or not supported well by available data. The three largest families of neococcoids, in order of size, are the Diaspididae (armored scales; about 2400 species), Pseudococcidae (mealybugs; about 2000 species), and Coccidae (soft scales; over 1000 species). The other neococcoid families are the Eriococcidae (felt scales; about 550 species); Asterolecaniidae (over 200 species), Lecanodiaspididae (about 80 species), and Cerococcidae (about 70 species) (these 3 families collectively are called pit scales); Kerriidae (Tachardiidae) (lac insects; about 100 species); Kermesidae (gall-like scales; about 90 species); Aclerdidae (flat grass scales; about 50 species); Conchaspididae (about 30 species); Halimococcidae (about 20 species); Stictococcidae (16 species); Beesoniidae (9 species); Dactylopiidae (cochineal insects; 9 species); Micrococcidae (8 species); and Phoenicococcidae (1 species).

The oldest fossil scale insects are from the Lower Cretaceous but the group is at least of Triassic and probably of Permian age. The earliest radiation involved the archaeococcoids, with the neococcoids apparently diversifying in conjunction with flowering plants.

Life History

The reproductive repertoire of scale insects includes hermaphroditism, seven kinds of parthenogenesis, and six major types of sexual chromosome systems. In the vast majority of coccoid

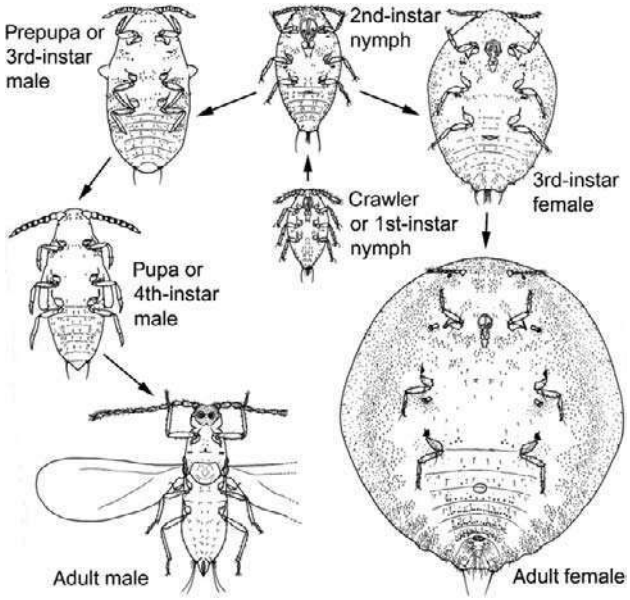

FIGURE 8 The life cycle of the Australian *Acacia*-feeding mealybug *Melanococcus albizziae* (Coccoidea: Pseudococcidae). (Partially adapted, with permission, from G. S. Farrell, 1990, *Mem. Mus. Victoria* **51,** 49–64.)

species, the males are functionally (because of inactivation or elimination of paternal chromosomes), and sometimes actually, haploid.

Each individual female scale insect has four or five growth stages (Fig. 8): the egg, two or three immature (nymphal) instars, and the adult (imaginal stage). The female lays eggs (ovipary) either in a cavity under her body or in a waxy covering (ovisac; Fig. 9E) that may be attached to her body, or she retains the eggs in her reproductive tract until the young are ready to hatch (ovovivipary). The mobile first instars, called crawlers, are the main dispersal agents for Coccoidea; other immature instars generally are sessile. Adult females may live for an extended period (months to several years).

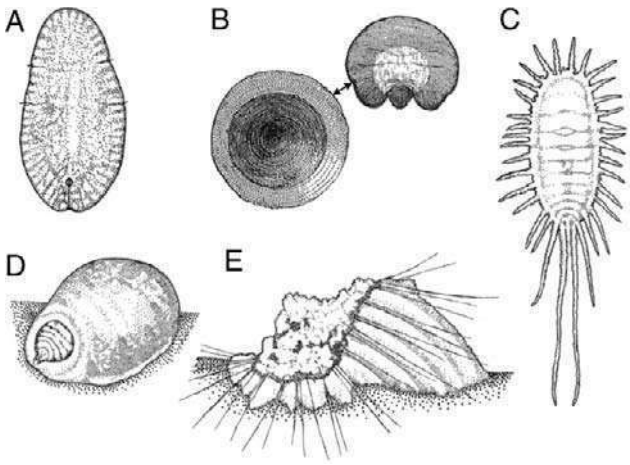

FIGURE 9 The morphological diversity of scale insects (Coccoidea). (A) An adult female of the soft brown scale, *Coccus hesperidum* (Coccidae). (B) Adult female (right) and scale cover of a California red scale, *Aonidiella aurantii* (Diaspididae). (C) An adult female of the longtailed mealybug, *Pseudococcus longispinus* (Pseudococcidae). (D) Adult female of a gum tree scale, *Eriococcus coriaceus* (Eriococcidae), inside her waxy test with just her posterior abdomen visible. (E) The hermaphroditic adult of the cottony cushion scale *Icerya purchasi* with ovisac (Margarodidae). (Illustration by P. J. Gullan.)

After hatching from the egg, male scale insects have a total of four immature or preimaginal instars including their own specially derived form of a complete metamorphosis (holometaboly) involving one or two pupa-like stages (Fig. 8). These are called the prepupa and pupa and develop either under a scale cover or test or inside a waxy cocoon or test that is produced by the second instar. Neither the pupa-like instars nor the adult males feed; adult males are short-lived (at most a few days) and have limited time to seek out the sedentary females for mating.

Host-alternating life cycles, as seen in some aphids, are unknown. The number of annual generations varies among and often within species and ranges from fewer than one to up to seven or eight per year. Annual life cycles are common in cool-temperate regions and numerous annual generations, typical of many aphids, do not occur in scale insects.

Morphology

Morphologically, scale insects are among the most unusual of insects. There is a very marked dimorphism between the adult male and the adult female and identification is based almost entirely on the long-lived adult female. Adult females (Fig. 9) are sac-like without a well-defined head, thorax, or abdomen, and they are neotenic, that is, the adult resembles an immature individual. Adult females are wingless, may or may not have legs, but usually have well-developed mouthparts. The anus opens posteriorly and usually is surrounded by a chitinized ring and often anal plates (Fig. 9A) or anal lobes flank the opening. Species that have a scale cover (e.g., armored scales; Fig. 9B) usually incorporate in it the dorsal part of the cast-off cuticle from the previous growth stage. If the body wax is of the adhering type, then it may vary from a thin translucent sheet (many soft scales) to a thick, wet mass (e.g., wax scales), a cottony secretion (e.g., some margarodids) or a powdery, white dusting (many mealybugs; Fig. 9C). Species that produce waxy tests (e.g., felt scales; Fig. 9D) often also have a dusting of powdery body wax. Wax is produced by epidermal glands and is exuded from a great variety of cuticular pores and ducts and sometimes also from glandular setae. Typically the body also is covered in setae that may be hair-like, spine-like, or other shapes, often with several types present on each species.

Adult males resemble small, delicate flies (Fig. 8) and have a distinct head, thorax, and abdomen. Most have a pair of membranous forewings and a pair of vestigial hind wings (balancers or halteres), although adult males of some species are wingless. The mouthparts of the prepupa, pupa, and adult males are reduced or absent.

Crawlers (Fig. 8) are usually ovoid or elongate ovoid and flattened dorsoventrally. Their antennae, legs, and mouthparts are well developed and various kinds of pores, ducts, and setae

are present on the body and its appendages; often a fringe of setae surrounds the body margin.

Behavior and Ecology

The first instars either seek suitable feeding sites on the natal host plant or disperse on the wind; some crawlers display behaviors that increase their chances of becoming airborne. Adult males probably locate their sessile conspecific females using sex pheromones but the presence of these chemicals has been demonstrated experimentally for very few species.

Scale insects primarily feed from either the phloem or the parenchyma, and their host associations range from monophagous to polyphagous. Sap removal is the main cause of plant damage, but a few species (especially of mealybugs and armored scales) also transmit plant pathogens and/or toxins that may further reduce plant vigor and eventually kill the host. Furthermore, most scale insects (except armored scales and a few others) produce honeydew and are ant attended; associations are usually facultative but a number of obligate ant–coccoid relationships have been described.

Cochineal scales (*Dactylopius* spp.) and certain mealybugs and lac insects have been used as biological control agents for particular noxious weeds; e.g., cochineal insects can assist with the control of prickly pear cacti, *Opuntia* species.

Notable Pest Species and Their Control

Some scale insects are serious plant pests, especially of perennial agricultural plants. They can cause damage to nut and fruit trees, forest or plantation trees, glasshouse plants, woody ornamentals, house plants, and sometimes sugar cane and even grass in lawns. Pests are usually either polyphagous (e.g., certain wax scales, *Ceroplastes* species; the pink hibiscus mealybug *Maconellicoccus hirsutus;* and the cottony cushion scale *Icerya purchasi*) or oligophagous (e.g., the beech scale *Cryptococcus fagisuga* on beech trees or *Matsucoccus* species on pines). The cryptic habits and small size of most scale insects mean that they may not be detected until plant damage is substantial. Also, if populations on plants are low, they can be notoriously difficult to detect during quarantine inspections. Most pest scales belong to the Diaspididae (e.g., California red scale *Aonidiella aurantii* on citrus, Boisduval scale *Diaspis boisduvalii* on orchids, euonymus scale *Unaspis euonymi* on ornamentals, and black pineleaf scale *Nuculaspis californica* on pines), Coccidae (e.g., certain polyphagous *Ceroplastes, Coccus, Pulvinaria,* and *Saissetia* species), and Pseudococcidae (e.g., many species of *Dysmicoccus, Pseudococcus,* and *Planococcus* on a wide range of plants), but a few significant pests belong to other families, such as the Margarodidae (especially polyphagous *Icerya* species), Eriococcidae (e.g., certain *Eriococcus* species on ornamental plants), and Asterolecaniidae (e.g., *Bambusaspis bambusae* on bamboo).

The most important predators of scale insects are ladybird beetles (Coccinellidae; especially *Rodolia, Chilocorus,* and *Cryptolaemus*). The main parasitoids of scale insects are chalcidoid wasps, especially species of Aphelinidae (e.g., species of *Aphytis, Encarsia,* and *Coccophagus*) and Encrytidae (e.g., species of *Anagyrus, Leptomastix,* and *Metaphycus*), although some scale insects are attacked by flies that may be either parasitic (e.g., Cryptochetidae) or egg predators (e.g., a few Cecidomyiidae).

See Also the Following Articles

Auchenorryncha • Cicadas • Honeydew • Parthenogenesis • Prosorrhyncha • Rostrum • Symbionts Aiding Digestion

Further Reading

Ben-Dov, Y., Miller, D. R., and Gibson, G. A. P. (2002). ScaleNet. http://www.sel.barc.usda.gov/scalenet/scalenet.htm.

Blackman, R. L., and Eastop, V. F. (1994). "Aphids on the World's Trees: An Identification and Information Guide." CAB Int., Wallingford, UK.

Blackman, R. L., and Eastop, V. F. (2000). "Aphids on the World's Crops: An Identification and Information Guide," 2nd ed. Wiley, Chichester.

Burckhardt, D. (1994). Psylloid pests of temperate and subtropical crop and ornamental plants (Hemiptera, Psylloidea): A review. *Entomologia* **2,** 173–186.

Byrne, D. N., and Bellows, T. S., Jr. (1991). Whitefly biology. *Annu. Rev. Entomol.* **36,** 431–457.

Carver, M., Gross, G. F., and Woodward, T. E. (1991). Hemiptera. *In* "The Insects of Australia: A Textbook for Students and Research Workers" (CSIRO, ed.), 2nd ed., Vol. I, pp. 429–509. Melbourne University Press, Carlton.

Gerling, D. (ed.) (1990). "Whiteflies: Their Bionomics, Pest Status and Management." Intercept, Andover, UK.

Gullan, P. J., and Kosztarab, M. (1997). Adaptations in scale insects. *Annu. Rev. Entomol.* **42,** 23–50.

Minks, A. K., and Harrewijn P. (eds.) (1987–1989). "Aphids: Their Biology, Natural Enemies and Control," Vols. 2A, 2B, and 2C of "World Crop Pests." Elsevier, Amsterdam.

White, I. M., and Hodkinson, I. D. (1985). Nymphal taxonomy and systematics of the Psylloidea (Homoptera). *Bull. Br. Mus. (Nat. Hist.), Entomol.* **50(2),** 153–301.

Stonefly

see *Plecoptera*

Stored Products as Habitats

Rudy Plarre

Federal German Institute for Materials Research and Testing

Wendell E. Burkholder

University of Wisconsin

The protection of stored food against attack by insects and molds has been a necessity since the dawn of agriculture,

FIGURE 1 Traditional storage facility for grain in China.

or perhaps since humans began to manufacture primitive clothing and tools for daily use from plant or animal material. Protection of stored products, thus, has a tradition of tens of thousands of years and is documented in the earliest records of history.

Stored food and other products of plant or animal origin provide a habitat with shelter, food, and breeding sites for insects (Fig. 1). Worldwide losses from these postharvest insects and other food pests are estimated to be from a few percent in developed countries to 20% in an average year in some developing countries. Under certain storage conditions, 50 to 100% of food is lost.

Several hundred insect species are associated with stored products. Most are beetles (Coleoptera) or moths (Lepidoptera), with their parasitoids being wasps (Hymenoptera). However, there are only about 100 species of economic importance. These can be divided into two main categories: species infesting dry botanicals, mainly but not exclusively grain seeds and their products, and scavenger species infesting animal products (meat, furs, wool, feathers, etc.). Very few pest insects can feed and breed successfully on both plant and animal material.

ECOLOGY OF STORED-PRODUCT INSECTS

The storage environment can be regarded as a human-made, simply structured, unstable ecosystem. All trophic levels of a typical ecosystem are represented. However, the system is not self-sustaining, and there is a unidirectional flow of energy and biomass.

Trophic Levels in a Storage Environment

The harvested and stored parts of the plants or the processed material are regarded as the producers (or the first trophic level) and contain the energy that was captured by the crops in the field. The primary consumers of the second trophic level in stored cereals are insects that infest the intact grain kernel. Most numerically dominant are the grain weevils of the genus *Sitophilus* (Curculionidae) (Fig. 2), the grain borers *Rhyzopertha dominica* (Fig. 3) and *Prostephanus truncatus* (Bostrychidae), and the grain moth *Sitotroga cereallela* (Gelechiidae). Primary consumers provide routes of entry for secondary consumers, such as *Tribolium* spp., that feed on the waste products or on debris, the frass, provided by the primary consumers. However, because damaged grain and grain dust are always present, secondary consumers can also exist without primary consumers.

The oxidation processes of the stored commodity and the

FIGURE 2 *Sitophilus zeamais,* maize weevil, adults on maize. (Photograph from Agriculture and Agri-Food Canada.)

FIGURE 3 *Rhyzopertha dominica,* lesser grain borer, adults on wheat. (Photograph from Agriculture and Agri-Food Canada.)

respiratory activity of primary and secondary consumers generate heat and moisture that affect the microclimate. "Hot spots" may occur, which favor microbial growth. Certain microbial fungi further enhance conditions favorable to primary and secondary consumers. At this stage secondary and tertiary consumers become difficult to separate. Finally, the way is paved for the third trophic level of decomposing fungi, yeasts, and bacteria.

Predators and parasitoids of stored product insects are frequently encountered in highly infested commodities. Some of them are very closely associated with prey or host. Others are less specific: most parasitic wasps, for example, parasitize a number of different host species.

Carnivorous or scavenger species are typically found on stored commodities of animal origin such as dry meat or fish, wool, hair, and furs or products made out of these such as clothing or carpets. For example, dermestid beetles of the genera *Dermestes, Anthrenus, Attagenus,* or *Trogoderma* and the tineid moths *Tineola bisselliella* and *Tinea pellionella* are the most common insects in households. Scavengers feed on dead insect parts as well, which is why they are often found in stored products of plant origin that are infested with phytophagous insects.

Because of the weblike interactions of the different trophic levels, it is difficult to place insect species in specific categories. Their roles may be granivore, carnivore, fungivore, predator, parasite, and scavenger. These roles, together with the microflora and the environmental factors temperature and humidity, result in ongoing fluctuations of physical parameters, species succession, changes in population densities, and continuous degradation of the stored commodity.

Feeding Habits and Damage

Primary and secondary insect pests of stored foodstuff are polyphagous, with a general preference for either dry plant material or material of animal origin. Under a different perspective, these pests can also be classified into internal and external feeders. Internal feeders are species that penetrate the intact outer coating of seeds either as feeding or ovipositing adults or as freshly hatched larvae. External feeders graze on the surface of stored food stuff, on already damaged material, and on debris. Nevertheless, they almost all have a cryptic feeding habit that is probably a result of their negative phototaxis. The larvae of the Indian meal moth, *Plodia interpunctella,* the most common stored-product moth worldwide and also frequently encountered in households, prefers undisturbed and darker places wherever suitable food is available.

The most obvious damage caused by stored-product insects is the destruction of food destined for human consumption. In addition to their capability to cause tremendous losses, insects contaminate the commodity with frass, feces, exuviae, and corpses. The accumulation of filth may lead to the rejection of the stored material. Except for some tenebrionid grain beetles infected with a certain tapeworm, stored-product insects have not been shown to transmit human pathogens or diseases. However, some of their by-products can be highly allergenic to humans.

Environmental Factors

Temperature and relative humidity usually determine the limits of insect infestations in stored products. The range of temperature in which stored-product insects are active lies between 8 and 41°C. Reproduction and larval development above or below these limits are not possible, and the individuals can survive extremes only in diapausing or hibernation stages. Some species are more cold-hardy or heat tolerant than others. Relative humidity also is important because it directly influences the moisture content of the foodstuff. The relative humidities to which pest insects have adapted range from near 0 to close to 100%, with typical optimal conditions being from 60 to 75%. Temperature and relative humidity of the microclimate can be actively influenced through the metabolic activity of aggregated individuals. Light is less important than temperature and humidity as a factor influencing the survival of stored-product insects. However, most larvae have a negative phototaxis and therefore are most likely to be found in dark, less disturbed areas.

BEHAVIOR OF STORED-PRODUCT INSECTS

In addition to physical parameters, the behavior of stored-product insects is greatly influenced by semiochemicals. Living organisms, grain, microflora, insects, and mites, as well as decaying matter, produce an array of volatiles that combine with the oxygen, nitrogen, and carbon dioxide of the system. For example, enzymatic reactions in stored seeds produce unsaturated fatty acids that are highly attractive to a number of pests. Catabolism of decaying plant and animal matter enables pest insects to locate suitable feeding and breeding sites. Fungi have a distinct odor that is attractive to fungivorous species but repellent to others at high concentrations.

Intra- and interspecific communication is directed by a number of pheromones and allelochemicals. For example, spacing and oviposition pheromones are found among bruchids marking already infested seeds. These compounds are repellent to conspecifics. Spacing pheromones of certain species not only act on conspecifics but also repel nonconspecifics and thus act as allomones.

Sex pheromones, nearly always produced by the female, attract mates. These pheromones comprise specific compounds or a highly specific combination of compounds. Aggregation pheromones, typically found in species with long-lived adults, are produced by males and attract conspecifics of both sexes. Certain compounds of some aggregation pheromones are not species specific, and there is cross-attraction, especially among closely related species. In most cases, foods synergize the effect

of aggregation pheromones. Although aggregation pheromones increase the likelihood that an individual will find a mate among aggregated conspecifics, the role of these substances as mating pheromones is not necessarily obvious.

Several parasitoids and predators make use of volatiles to locate hosts and/or prey, and under certain circumstances pheromone components may act as allelochemicals. Specialists that parasitize or prey on a single host species, a very narrow range of related host species, or specific developmental stages are attracted to specific kairomones emitted by their host or prey.

ORIGIN OF STORED-PRODUCT INSECTS

The origin of stored-product insects can be examined from three viewpoints: first, from a practical or physical point of view, asking for the natural reservoirs from which the insects spread into the storage system and seeking to learn how they invade a particular commodity; second, from a phylogenetic point of view, examining the ecological and behavioral preadaptations that made it possible for insects to shift from an original or primary habitat to a relatively recent, human-made habitat; and finally, from a historical point of view, searching for the geographic origins of the various insect species that infest stored commodities, which today are considered to be distributed worldwide.

Preadaptations and Natural Habitats

Insects species capable of surviving in stored commodities generally have some biological strategies in common. Species infesting stored products are not monophagous; rather, stored-product insects are food opportunists. For example, the cigarette beetle, *Lasioderma serricorne,* and the drugstore beetle, *Stegobium paniceum,* are perhaps the most ubiquitous of all stored-product pests, but all others also feed and breed on a variety of products. Almost all species found in human-made storage systems also can be found outdoors.

Natural breeding sites of stored-product insects can be sheltered niches where seeds and nesting material were gathered by rodents, birds, or other insects. Dermestids, ptinids (spider beetles), and tineids (clothes moths) are frequently found in empty rodent or bird nests where they feed on hairs, feathers, feces, pellets regurgitated by birds of prey, and food debris brought in or left by the previous occupants. Carrion is a food source as well for some species.

Stored-product insects typically have the ability to survive a long time without food. When environmental conditions are favorable and resources are available, stored-product insects reproduce over prolonged periods and have a rapid population growth rate, exploiting resources maximally. Dispersal is regulated by population density, and most stored-product insects are good fliers, which is essential for locating patchily distributed food and breeding habitats. Therefore, the following preadaptations help insect species to exploit the resources of stored-product systems: a range of natural reservoirs, a range of food habits, ability to survive in the absence of food, a wide range of tolerance for physical factors in the environment, and a high reproductive rate at optimal conditions.

Additionally important for consumers of dry foodstuffs is the ability to conserve water. Pest insects have to be either tolerant of low humidity and moisture content, able to recycle metabolic water, or able to influence and alter microclimatic conditions. Several species, in which the adults feed, tend to aggregate at spots suitable for feeding and breeding. Such aggregation is not incidental but a result of male-produced pheromones that are attractive to conspecific males and females. This behavior raises the question of the evolutionary advantage for males producing pheromone that attracts not only potential mates (conspecific females) but also potential competitors (conspecific males). It is thought that male-produced aggregation pheromones originally served to attract a mate. If this is correct, the use of the pheromone by other males of the same species then evolved secondarily, presumably allowing them to locate resources and assembled females. This would result in a fitness gain only for the intruding males.

An alternative explanation for the production of aggregation pheromones could be that the primary role of the male-produced aggregation pheromone was to attract and congregate conspecifics to overcome host defenses or unsuitable climatic conditions. Because males and females can serve equally well in this respect, the pheromone emitter benefits from attracting beetles of either sex.

To infest dry foodstuffs or stored commodities successfully, stored-product beetles certainly do not have to overcome host defense systems. However, in natural habitats, such as under bark or in wood, stored-product beetles most likely benefit from, and rely on, aggregation pheromones to weaken the host and make the habitat more suitable. This behavior can be observed in non-stored-product species closely related to postharvest pests in the Tenebrionidae, Curculionidae, and Bostrychidae, which partly feed and breed on physiologically active host plant parts. Where seeds and other harvested products are usually stored under low temperatures and rather dry conditions, the aggregation pheromones may serve to congregate groups that alter the microclimate. Such areas of higher temperatures and humidity may provide favorable conditions for insect development.

Not all insects that infest dry foodstuff have aggregation pheromones. In those that do not, there is a tendency of the larvae to aggregate. For many species, development is maximized at an optimal density of individuals. Individual fitness declines as the population shrinks or enlarges. For example, the eggs of many bruchids are deposited in clutches. Ovipositing females in the pyralids are guided to breeding sites by mandibular gland secretions produced by conspecific larvae. (A larval pheromone from mandibular glands is responsible for an optimal spacing, thus regulating population density.) Many aggregation pheromones also serve as guides to ovipositing females.

Temporary Changes in Habitats

The evolutionary origin of most stored-product insects remains speculative. However, autecological studies indicate a more or less xylophilous origin, under bark or in decaying wood, for most species. Omnivorous and partly predatory feeding habits could have allowed those species to invade animal nests in the same habitat or close by. These nests and the materials within could have provided good sources of otherwise scarce carbohydrate- and protein-enriched resources. In the same way, at the beginning of agriculture more than 10,000 years ago, insects could have gained access to storage bins in granaries constructed of infested timber. This kind of behavior can be observed in *Prostephanus truncatus,* the larger grain borer, which can survive in a variety of wooden building materials often used for simple granaries in tropical and subtropical countries. When the new harvest is brought in, this beetle readily invades the stored commodity, returning to its xylophilous habit when the alternative enriched food source declines. Other insect pests associated with stored products also frequently commute between the stored commodity and their cryptic life in cracks and crevices of the storage facility, where they feed on debris.

Geographic Origin and Worldwide Distribution

Today most stored-product insects are worldwide in distribution, and it is difficult to trace species to specific geographic origins with certainty. Global trade of food and other commodities, and the past lack of quarantine regulations, dispersed most storage pests with the products they had infested. Despite their cosmopolitan distribution, not all storage insects are established worldwide or occur outside warehouses or granaries. This is especially true in the colder parts of the world. *Callosobruchus maculatus,* for example, does not survive winters in the Northern Hemisphere unless inside heated storage facilities. For some species it is impossible for climatic reasons to successfully survive as nonsynanthropic populations.

The association of certain insects with humans most likely began in prehistoric times, probably contemporaneous with intensified agriculture and the domestication of animals. Phylogenetic, ecological, and historical evidence suggests that the majority of stored-product insects originated in the tropical and subtropical parts of the world. These centers of origin were also the centers of Neolithic agriculture.

PROTECTING STORED PRODUCTS AGAINST PEST INSECTS

Some of the most modern approaches to keeping stored food from spoiling are based on historical and traditional methods modified by today's technology. The storage of commodities under low oxygen atmospheres, to slow down the biological decay mechanisms, was in principle already used in ancient underground and hermetically sealed storage facilities of the Old World. The same effect occurs in modern grain facilities that are fumigated with modified atmospheres enriched with nitrogen and carbon dioxide. The ancient Egyptians intermixed ash with grain, which acted as abrasives on insects' cuticle, causing desiccation. Today, inert dusts on a base of diatoms are applied, and by a slightly different mode of action these additives cause the insects to desiccate. The Egyptians added specific herbs and spices to the commodities to repel pest insects. Sulphur was one of the first fumigants for stored product protection. The insecticidal property of burning sulfur was known in the New Kingdom around 1,000 B.C. Toxic fumigants have the great advantage over contact insecticides of fully penetrating the commodity and therefore also killing internal infestations. However, the application of fumigants and pesticides is not risk free. Environmental concerns and the potential danger to human health have led to a steady reduction of registered fumigants and contact insecticides. Incorrect handling of pesticides can lead to control failures and resistance.

In the past, eradication of pests was the primary objective, and increasingly high food quality standards allow no, or only little, contamination of stored commodities. On the other hand, consumers demand foodstuff that is free of chemical residues. A zero pest tolerance, however, can rarely be achieved without the application of insecticides. Today's pest control programs must therefore address health-related, ecological, and economic concerns. Modern pest control strategies in stored product protection focus on integrated pest management (IPM). The concept of IPM is based on using threshold levels to make control decisions and on the combination of physical, biological, and chemical pest control measures. One threshold level, the economic injury level, is defined as the pest infestation level at which the loss due to pests is equal to the costs of available control measures. The threshold is not a fixed parameter, but varies with the control options, the infested commodity, and the infesting pest species. Fundamental requirements for a successful application of an IPM program are detailed understanding of the mode of action of any control measure and specific knowledge of the pest's biology.

See Also the Following Articles

Agricultural Entomology • Integrated Pest Management

Further Reading

Anonymous. (1979). "Proceedings of the Second International Working Conference on Stored-Product Entomology," Ibadan, Nigeria, 1978. Technical Assistance Bureau Agency for International Development, Washington DC.

Anonymous. (1984). "Proceedings of the Third International Working Conference on Stored-Product Entomology," Manhattan, KS, 1983. Kansas State University, Manhattan, KS.

Donahaye, E., and Navarro, S. (eds.). (1987). "Proceedings of the Fourth International Working Conference on Stored-Product Protection," Tel Aviv, Israel, 1986. Moar–Wallach Press, Caspit, Jerusalem.

Fleurat-Lesssard, F., and Ducom, P. (eds.). (1991). "Proceedings of the Fifth International Working Conference on Stored-Product Protection," Bordeaux, France 1990, Vols. I–III.

Gorham, J. R. (1991). "Ecology and Management of Food-Industry Pests." Association of Official Analytical Chemists, Arlington, VA.
Highley, E., Wright, E. J., Banks, H. J., and Champ, B. R. (eds.) (1994). "Proceedings of the Sixth International Working Conference on Stored-Product Protection," Canberra, Australia, 1994. CAB International.
Hinton, H. E. (1945). "A Monograph of the Beetles Associated with Stored Products," Vol. I. British Museum, London.
Jayas, D. S., White, N. D. G., and Muir, W. E. (1995). "Stored-Grain Ecosystems." Dekker, New York.
Subramanyam, B., and Hagstrum, D. W. (1996). "Integrated Management of Insects in Stored Products." Dekker, New York.
Subramanyam, B., and Hagstrum, D. W. (2000). "Alternatives to Pesticides in Stored-Product IPM." Kluwer Academic Publishers, Dordrecht, Netherlands.
Zuxun, J., Quan, L., Yongsheng, L., Xianchang, T., and Lianghua, G. (eds.) (1999). "Proceedings of the Seventh International Working Conference on Stored-Product Protection," Beijing, China, 1998. Sichuan Publishing House of Science and Technology, Chengdu.

Strepsiptera

Michael F. Whiting
Brigham Young University

Strepsiptera (twisted-winged parasites) is a cosmopolitan order of small insects (males, 1–7 mm; females, 2–30 mm) that are obligate insect endoparasites. The order comprises ~550 species placed within eight extant and one extinct family. Strepsiptera derive their common name from the male front wing, which is haltere-like, and early workers considered it to be twisted in appearance when dried specimens were examined. All members of this group spend the majority of their life cycle as internal parasites of other insects and consequently have a highly specialized morphology, extreme sexual dimorphism, and a unique biology.

BIOLOGY

The adult male strepsipteran is free living and winged, whereas the adult female is entirely parasitic within the host, with the exception of one family (Mengenillidae) in which the female last instar leaves the host to pupate externally. Strepsiptera parasitize species from seven insect orders: Zygentoma, Orthoptera, Blattodea, Mantodea, Heteroptera, Hymenoptera, and Diptera. In one family (Myrmecolacidae) the males are known to parasitize ant hosts, whereas the females are parasites of Orthoptera. The life cycle of most strepsipteran species is unknown, and only a few species have been studied in detail.

The first instar is free living, emerges from the female brood canal, and disperses to the surrounding vegetation in search of a new insect host. In some species, the larvae have long abdominal setae that are used to propel them on to a new host. These larvae may be carried back to the nest of a social host to begin parasitizing the host's larvae or may simply spring onto the early instars of a nonholometabolous host. In *Stylops pacificus,* the abdominal setae are short and the larvae are transported to the nest inside of the crop of the bee *Andrena complex,* where it will begin parasitizing a single egg of the bee.

Once a first instar enters a new host, it molts into an apodous larva and feeds transcutaneously from the host's hemolymph. *Elenchus tenuicornis* undergoes two additional molts prior to pupation; apolysis is not followed by ecdysis in strepsipteran larvae, such that the cuticle of the second to the fourth instar forms the puparium. In Mengenillidae, both sexes leave the insect host to pupate externally. The remaining Strepsiptera pupate in the host with a portion of the head and thorax extruded from the host's cuticle (Fig. 1C).

Males will emerge from the puparium and seek out a female for mating. Virgin females release a pheromone to attract males, and in Stylopidia the male copulates by rupturing the female's brood canal opening between her head and prothorax in a process referred to as hypodermic insemination. Adult males live only a few hours, which makes them difficult to collect, although they are occasionally found at lights or may be lured to a virgin female kept in a cage in the field. Adult males do not feed and their mouthparts are partly modified into sensory structures. After insemination, females may live for a few weeks to a few months. In *Xenos peckii,* the female is typically inseminated in the fall and she remains in the wasp host until spring, when the planidial larvae emerge. The female may be an enormous endoparasite relative to the size of its host. For example, *X. peckii* may fill up to 90% of the abdominal cavity of the wasp host. It is relatively common to find a host that has been parasitized by multiple males and/or females.

MORPHOLOGY

First instars bear stemmata (simple, single-lens eyes), mandibles, maxillae, and a labium. Antennae are absent. Abdominal tergites are typically smooth, the abdominal sternites are serrate, and the legs are filiform with slender tarsi (Fig. 1A). In Mengenillidae, the pupae are free living and have legs, mouth, and a segmented abdomen. In the remaining Strepsiptera, the puparium is tanned and male pupae possess compound eyes, abdominal and thoracic segmentation, and developing wings, legs, and antennae (Fig. 1B).

The adult male has large compound eyes, with ommatidia separated by cuticle and/or setae, giving the eye a blackberry appearance. Antennae are flabellate and the mandibles are conical, except in Corioxenidae, which entirely lack mandibles (Fig. 1F). The first pair of wings is reduced, clubbed, and morphologically similar to the halteres of Diptera. The hind wings are enlarged and venation is reduced, and these wings are used for flight (Fig. 1G). The prothorax is small and saddle-shaped and the metathorax is large and bears the principal flight muscles. The trochanter is fused with the femora in the fore- and middle legs, the hind legs lack free

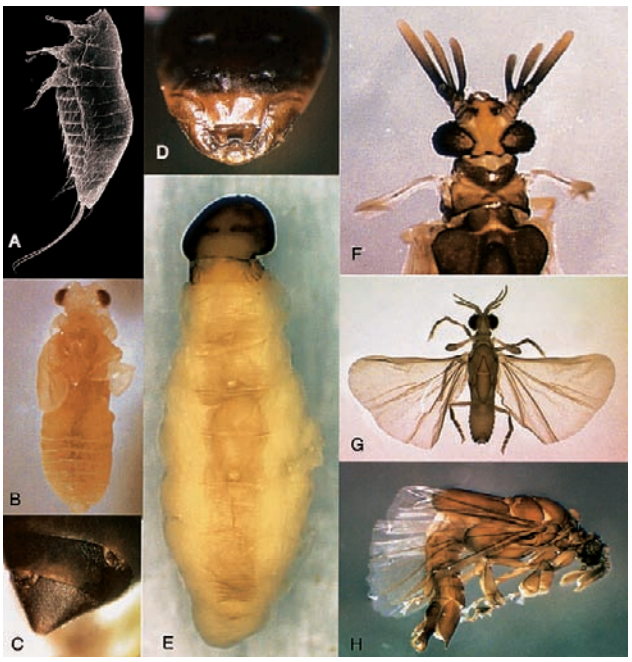

FIGURE 1 (A) First instar of *Mengenilla chobauti.* (B) Male pupa of *Xenos peckii.* (C) Cephalothorax of female *X. peckii* extruded from the abdominal tergites of the wasp host *Polistes fuscatus.* (D) Close-up of cephalothorax of female *X. peckii.* (E) Female *Xenos vesparum.* (F) Head and thorax of *M. chobauti.* (G) Male *Triozocera bedfordiensis.* (H) Male *Coriophagus rieki.*

coxae, and the tarsi have two to five segments and may lack a tarsal claw (Fig. 1H).

The adult female lacks wings, legs, and external genitalia, and only rudiments of the mouthparts, antennae, and eyes remain (Figs. 1D and 1E). The thorax and head are fused to form a heavily sclerotized cephalothorax, and this is the portion of the female that is extruded from between the abdominal segments of its host. The abdomen is large, segmentation is reduced, and the cuticle is unsclerotized, which allows nutrients to pass from the host to the developing embryos. The tracheation and the nervous system are reduced, and the reproductive system is highly modified, allowing the production of vast quantities of eggs that are freely distributed throughout the abdominal cavity.

CLASSIFICATION

The Strepsiptera are divided into two major groups: the Stylopidia and the Mengenillidia. Mengenillidia comprises one extinct family (Mengeidae) and one extant family (Mengenillidae). Mengenillidia is considered to be the most primitive strepsipteran group because females are free living, with rudimentary legs, antennae, and a single genital opening, and the males have robust mandibles and a hind wing with an elongated and sturdy MA vein.

Stylopidia comprises seven families: Corioxenidae, Halictophagidae, Callipharixenidae, Bohartillidae, Elenchidae, Myrmecolacidae, and Stylopidae. The monophyly of Stylopidia is supported by the fact that females are found in the host, the female has multiple genital openings, and the hind wing in the male has only residues of the MA_1 vein. Recent research on the phylogenetic relationships among strepsipteran families, based on the first instars, suggests that the Corioxenidae is the most basal family within the Stylopidia and that Stylopidae is paraphyletic with the Xeninae and Stylopinae placed in different groups.

PHYLOGENY AND EVOLUTION

The phylogenetic position of Strepsiptera relative to other insect orders has been one of the most contentious issues in insect systematics, and controversy still exists as to their phylogenetic affinity. There have been four phylogenetic hypotheses presented: (1) placement as a group somewhere within Coleoptera, as suggested by the superficial similarities in their parasitic lifestyle and morphology with certain beetle groups; (2) placement as sister group to Coleoptera, as suggested by the powering of flight via the hind wing; (3) placement outside of Holometabola, as suggested by the presence of larval stemmata; and (4) placement as sister group to Diptera, as suggested by DNA sequence data. The latter hypothesis has been the most controversial in recent years and different interpretations of the molecular data have been presented. However, if Strepsiptera are indeed sister group to Diptera, then this gives rise to some intriguing possibilities in insect evolution. Functionally, the halteres of Diptera and Strepsiptera appear identical, both highly specialized to serve as a gyroscopic balancing mechanism during flight. Therefore, it is possible that the mesothoracic halteres of Strepsiptera were derived via a specialized mutation in the genes controlling the development of the metathoracic halteres in Diptera, essentially causing the wings of the dipteran to switch positions to produce the type of wings as observed in Strepsiptera.

See Also the Following Article

Fossil Record

Further Reading

Crowson, R. A. (1960). The phylogeny of Coleoptera. *Annu. Rev. Entomol.* **5,** 111–134.

Kathirithamby, J. (1989). Review of the order Strepsiptera. *Syst. Entomol.* **14,** 41–92.

Kinzelbach, R. K. (1971). Morphologische Befunde and Facherfluglern und ihre phylogenetische bedeutung (Insecta: Strepsiptera). *Zoologica* **119,** 1–256.

Kristensen, N. P. (1991). Phylogeny of extant hexapods. *In* "The Insects of Australia: A Textbook for Students and Research Workers" (P. B. C. I. D. Naumann, J. F. Lawrence, E. S. Nielsen, J. P. Spradberry, R. W. Taylor, M. J. Whitten, and M. J. Littlejohn, eds.), 2nd ed., pp. 125–190. CSIRO, Melbourne University Press, Melbourne.

Linsley, E. G., and Macswain, J. W. (1957). Observations on the habits of *Stylops pacifica* Bohart. *Entomology* **11,** 395–430.

Pix, W., Nalbach, G., and Zeil, J. (1993). Strepsipteran forewings are haltere-like organs of equilibrium. *Naturwissenschaften* **80,** 371–374.

Pohl, H. (2002). Phylogeny of the Strepsiptera based on morphological data of the first instar. *Zool. Scripta* **31,** 123–134.

Whiting, M. F., Carpenter, J. C., Wheeler, Q. D., and Wheeler, W. C. (1997). The Strepsiptera problem: Phylogeny of the Holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.* **46,** 1–68.

Young, G. R. (1987). Notes on the life history of *Stichotrema dallatorreanum* Hofeneder (Strepsiptera: Myrmecolacidae), a parasite of *Segestes decoratus* Redtenbacher (Orthoptera: Tettigoniidae) from Papua New Guinea. *Gen. Appl. Entomol.* **19,** 57–64.

Swimming, Lake Insects

Werner Nachtigall
Universität der Saarlandes

Insects of many types, such as beetles (Coleoptera), true bugs (Heteroptera), and fly larvae (Diptera), can be observed swimming in ponds and lakes. Of these, the water beetles in the family Dytiscidae are reputed to be the best swimmers. The trunks of their bodies are well adapted to flow, and they generate thrust by executing synchronous power strokes with their flattened rear legs, which bear two rows of "swimming" hairs.

FLOW ADAPTATION

Measurements on flow adaptation have been carried out on trunks of several Dytiscidae, especially with the large European water beetle *Dytiscus marginalis.*

The technical term coefficient of drag, or c_d, is used as an indicator of flow adaptation. A small coefficient indicates that a beetle with a given frontal area and at a given swimming speed generates little drag; that is, it is well adapted to flow. However, one cannot readily compare the drag created by objects moving at different speeds because the coefficient of drag is a function of the Reynolds number (i.e., of its swimming speed in water and its body length). For certain ranges of Reynolds numbers, there are sufficient data available, ranging from the most streamlined bodies (e.g., drop-shaped trunks; lowest c_d values) to the objects that produce the greatest drag (e.g., parachute forms; highest c_d values). Figure 1A shows the trunks of four *Dytiscus* species viewed from above, from the side, and from the front (the front projection shows the frontal area *A*). Figure 1B shows the measured c_d within the spectrum of coefficients that are possible for the range of Reynolds number (Re) obtained from swimming water beetles ($10^3 < \mathrm{Re} < 10^4$). The coefficients of the beetle fluctuate between 0.38 for *D. latissimus* and 0.43 for *D. pisanus,* with the smallest possible value being 0.2 or slightly less, and the highest possible value approximately 1.4. Compared with the possible spectrum, the beetles' trunks seem to be well adapted to flow, although they do not reach the extreme values technically possible. Low-drag technical constructions are unstable; in the presence of oblique flow

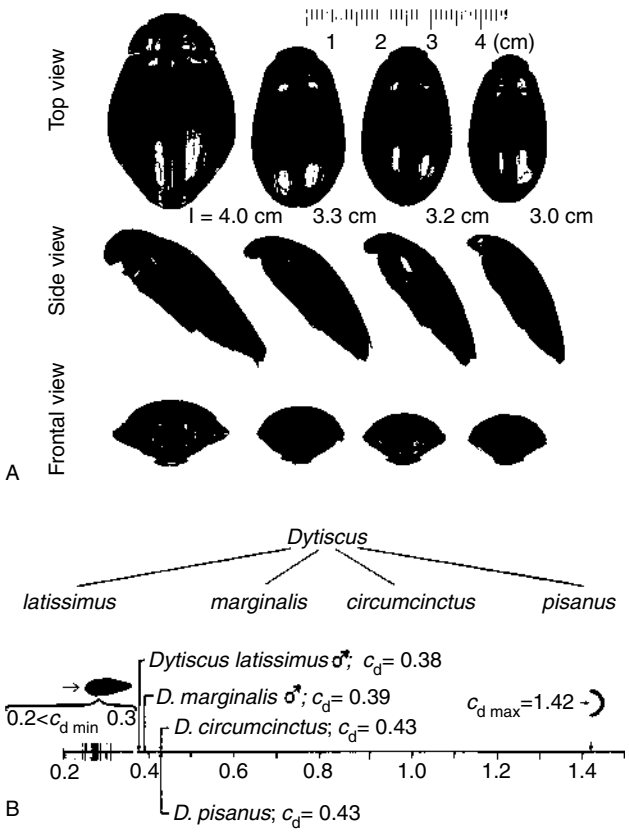

FIGURE 1 (A) Trunks from four large species of *Dytiscus* viewed from above, from the side, and from the front. (B) Classifying the four c_d values of these beetles at zero angle of attack (cf. Fig. 2) within the spectrum of possible coefficients of drag in the range $10^3 < \mathrm{Re} < 10^4$. [After Nachtigall, W. (1986).]

they turn immediately broadside, and stabilizing surfaces preventing this behavior would increase drag. Enlarged prothoracic and elytra edges in large *Dytiscus* beetles (extreme in *D. latissimus*) serve as stabilizing surfaces for swimming by damping oscillations around the longitudinal and lateral axes and by creating stabilizing moments. If one reduces these stabilizing surfaces along the edges, drag at almost all angles of attack between trunk and flow is perceptibly reduced. But the trunk then is distinctly unstable. These two contradictory demands have led to the evolution of an optimal shape with good swimming stability and good flow adaption.

Geometrically similar, 10-times-enlarged models of large *Dytiscus* beetles have been used to measure the pressure distribution along the dorsal midline of these trunks, with numerous pressure holes. A positive pressure dent appears on the head–prothoracic region (and also on the abdomen), whereas the remaining upper abdominal side is under negative pressure (Fig. 2). Thus, a *Dytiscus* trunk is very similar to a small-span wing profile and should therefore create a certain dynamic lift during fast swimming.

Because the drag coefficient of bodies increases with smaller Reynolds numbers, one may expect that the smallest, most slowly moving Dytiscidae (e.g., *Hyphydrus, Bidessus*) are characterized by relatively high drag coefficient values. Thus,

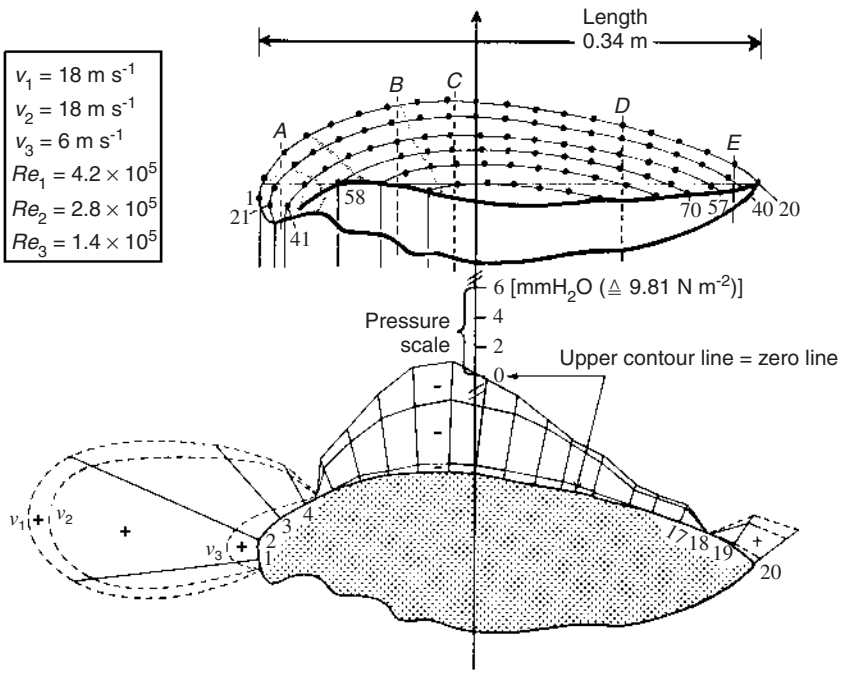

FIGURE 2 Pressure distribution in the median plane of a model of *D. marginalis* (male) enlarged 10 times, measured in a wind tunnel: dashed lines, pressure above atmospheric ("pressure"); solid lines, subatmospheric pressure ("suction"). [After Nachtigall, W. (1986).]

they seem to be less well adapted to flow simply by their smallness (i.e., by their low Reynolds numbers). From flow mechanics, we know that streamlining is less effective at lower Reynolds numbers. Indeed, these small beetles are characterized by blunt body forms that may have biological advantages (more effective space for the organs). Because these beetles are energetically better equipped, this inherent physically disadvantage may not be of primary importance.

FUNCTIONAL MORPHOLOGY

Leg Flattening and Position of Swimming Appendages

The functional formation of a swimming leg is easy to understand because of the simple mechanical demands made upon it. A swimming leg should produce strong thrust (i.e., a force component in the swimming direction) during a rowing stroke (i.e., a power stroke), and little counterthrust while being drawn forward (i.e., a recovery stroke). This action is aided by morphological and kinematic adaptations.

At a given speed the hydrodynamic drag force F_d is proportional to the leg surface A. Therefore, A must be as large as possible during rowing stroke, and as small as possible when drawn forward. This is achieved by flattening the leg segments and by means of thrust-inducing processes such as swimming hairs (e.g., *Dytiscus, Corixa*) or swimming platelets *(Gyrinus).* During the rowing stroke, the flat broadside of the leg is brought perpendicular to flow, and the hairs and platelets spread out automatically under the pressure of flow, thus increasing the rowing surface. When the leg is drawn forward, it is turned so that the flattened surface lies parallel to flow, and the rowing appendages are folded together and pressed against the leg by the flow; their additional surface disappears completely (Fig. 3).

Leg oscillation is an angular movement. Therefore, the drag, which is created by an element of surface, is proportional to the square of its rotating radius r, or its distance from the coxa–trochanter joint. A component of the

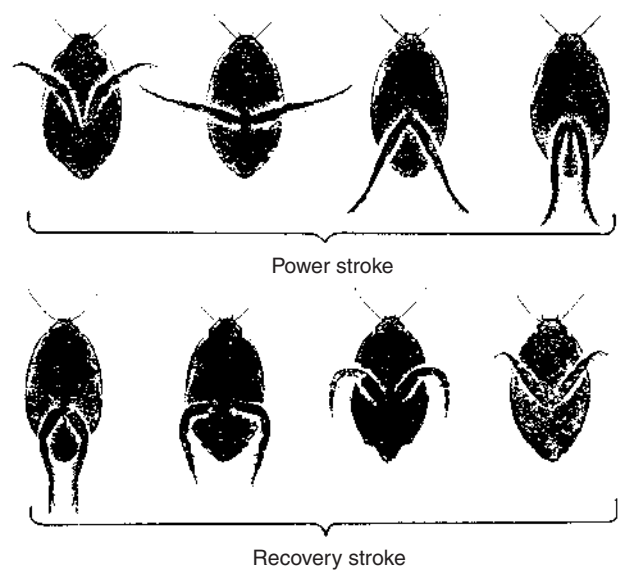

FIGURE 3 Stroke phases of the hind legs of *A. sulcatus* during power stroke and recovery stroke. [After Nachtigall, W. (1986).]

counterforce of this drag, parallel to the median, is the thrust force. The thrust-creating area during a rowing stroke should lie as far out as possible (*r* large) during the power stroke, and as far in as possible (*r* small) during the recovery stroke. Thus, the broad area and its hairs can be called a thrust-creating area. This is achieved principally because the distal leg segments, the tibia and the tarsus, are greatly enlarged and carry thrust-creating swimming hairs or swimming platelets, and because the legs are outstretched during stroke and then bent and drawn forward as close as possible to the median.

Stroke Areas and Radii of Rotation

The leg segments in *Gyrinus* have been flattened so extensively that their surface area is approximately five times that of a round leg; because of the simultaneously increasing coefficient of drag, a hydrodynamic force roughly eight times larger is created. During the power stroke, the leg segments spread out like playing cards, thus increasing their size by roughly 1.6 times over their projection area when not spread. The supplementary surface areas of the swimming platelets increase the stroke surface in *Gyrinus* even more: by roughly 230% in the tibia and roughly 130% in the tarsus.

In *Acilius,* the flattened tarsus positioned very distally occupies 80% of the total stroke area. Its mean rotational radius is also roughly 80% of the leg length. In *Gyrinus,* the corresponding values are approximately 40 and 80% for the tarsus and approximately 50 and 60% for the tibia.

During the recovery stroke, *Acilius* tilts the broad side of its leg parallel to the flow and presents the narrow edge of its leg to the flow. At the same time, the swimming hairs and swimming platelets are flattened and lie close to the leg. Because the exterior leg sections lie more or less parallel to the median during the first phase of drawing forward, its area of projection in the direction of flow is smaller yet. In *Gyrinus,* this projection area in proportion to the area of the power stroke is reduced to 1/13 for the middle legs and 1/16 for the hind legs. The tarsus from *Gyrinus* collapses completely and disappears partly into a groove in the tibia, which in turn slips partly into a hollow in the femur, so that the leg surface area is dramatically reduced by roughly 70% of the leg surface during the recovery stroke.

By folding the legs when they are drawn forward nestled against the body, *Gyrinus* reduces the mean rotational distance relative to that of the power stroke from 60% to approximately 35% of total leg length. The values for the tarsus alone are 80 and 50%.

Two additional phenomena should augment swimming efficiency, although their precise contribution remains to be measured. First, pressure drag is created mainly during the power stroke, and frictional drag is probably when the legs are drawn forward. Second, the leg segments, insofar as they nestle against the ventral surface with their broad side when drawn forward, move within the boundary layer of the trunk, thus creating less drag than when moving through free water.

All results presented here are for "standing water," that is, for the very first strokes after starting. During steady swimming, the relations are more complicated because the leg interferes with flowing water so that the speed of the leg's rowing surface relative to the surrounding water changes. Small fishes are propelled by the rowing motions of their pectoral (or breast) fins in almost the same manner as the water beetles are propelled by rowing with their hind legs. It has been shown in these small fishes that interference of the rowing surfaces with flowing water does not play a major role.

See Also the Following Articles

Aquatic Habitats • Walking and Jumping

Further Reading

Blake, R. W. (1983). "Fish Locomotion." Cambridge University Press, Cambridge, U.K.

Hughes, G. (1958). The co-ordination of insect movements. III. Swimming in *Dytiscus, Hydrophilus* and a dragonfly nymph. *J. Exp. Biol. London* **35,** 567–583.

Nachtigall, W. (1986). Bewegungsphysiologie. Laufen, Schwimmen, Fliegen. Handbuch der Zoologie, Bd. 4, Arthropoda: Insecta. Gruyter, Berlin.

Swimming, Stream Insects

Bernhard Statzner
CNRS-Université Lyon 1, France

The physical world of an insect living in water or air is ruled by the fluid velocity relative to its body. Thus, in principle, it should not matter whether an insect swims fast in still water or slowly against the current of flowing water. However, the flow in streams is physically so harsh that it often constrains the movements of stream insects. Imagine a mountain stream with its typical current velocity of 50 cm s^{-1} and a 1-cm-long mayfly larva swimming upstream. To move upstream relative to the stream bottom, this larva must propel itself at more than 50 body lengths s^{-1}, which is a difficult accomplishment. Therefore, stream insects rarely swim and live predominantly at or in the surface layer of the stream bottom, where friction between the moving water and the solid bottom material generates extremely complex flow patterns. As a consequence, the aquatic stages of stream insects move in a physical world that is much more complicated than that of lake insects.

LIFE IN FLOW NEAR THE STREAM BOTTOM

Streams offer a mosaic of flow conditions, including areas with (almost) still water. Where a stream runs over a rough (e.g., stony) bottom, flow can be laminar or turbulent (or

transitional between these two). The difference between the laminar and the turbulent flow regime is easily understood by watching the smoke released by a cigarette in a calm place. Close to the tip, the ascending smoke travels in an orderly manner on a narrow path as a laminar flow, until suddenly the flow becomes turbulent and travels whirling on a wider path. The physical formulas for these regimes are quite different; thus, these two flow regimes constitute different physical worlds.

The Reynolds number *Re* indicates the flow regime for a particular flow situation through $Re = l \times U/\upsilon$, where l is a length dimension (e.g., the body length of an insect), U is the relative velocity to the insect body, and υ is the kinematic viscosity of the water. Low *Re* indicates laminar and high *Re* turbulent regime. Therefore, in general, small and slow (relative to the flow) stream insects (having a low $l \times U$) experience laminar flow in which viscous forces matter, whereas large and fast ones experience turbulent flow, where inertial forces predominate.

However, near the bottom of a stream things are a bit more complicated. When water runs over a stone, the water in contact with the stone surface does not slip relative to the surface (the no-slip condition) and a velocity gradient develops above the surface (e.g., as in Fig. 1c). Near to the surface, the flow can be laminar ("laminar sublayer"). If the stone is embedded among other stones of similar size (Fig. 1a) and the water is shallow relative to the stone size, turbulent eddies may extend to the stone surface and disrupt the formation of the laminar sublayer. As a consequence, velocity varies tremendously over short periods near the stone surface (Fig. 1b), as well as around insects living at that surface. To maneuver in that type of flow, a 1-cm-long insect must deal with velocity changes from 5 to more than 40 body lengths s^{-1}, and this within a fraction of a second. The body vibrations of black flies or elongated caddisflies staying at the surface of a stone in a shallow stream nicely illustrate that these insects have a very "shaky" life.

In natural streams, isolated larger bottom elements that rise far into the water column (Fig. 1a) are often very abundant. Above the surfaces of these elements, the no-slip condition produces a steep velocity gradient (Fig. 1c). Flow is generally laminar nearest to the surface (low standard deviation of the velocity in the laminar sublayer: Fig. 1c), whereas it is turbulent in the remaining part of the steep velocity gradient (elevated standard deviations in Fig. 1c). At a given location, the geometry (e.g., the thickness of the laminar sublayer) of this steep gradient changes with changing flow. That is, discharge variation in a stream causes temporal variability of the physical conditions at that location. At a given flow, the geometry of this gradient varies spatially along the slightly curved bottom element. The changes of the isovel (line of equal velocity) pattern along the curvature (Fig. 1d) demonstrate considerable spatial variability of the physical conditions over the distance of a few millimeters. As a consequence of these temporal and spatial patterns, a relatively sessile black fly larva must deal with temporal physical variation caused by changing stream discharge, whereas a relatively mobile mayfly larva must deal with spatial physical variation when moving at the bottom surface.

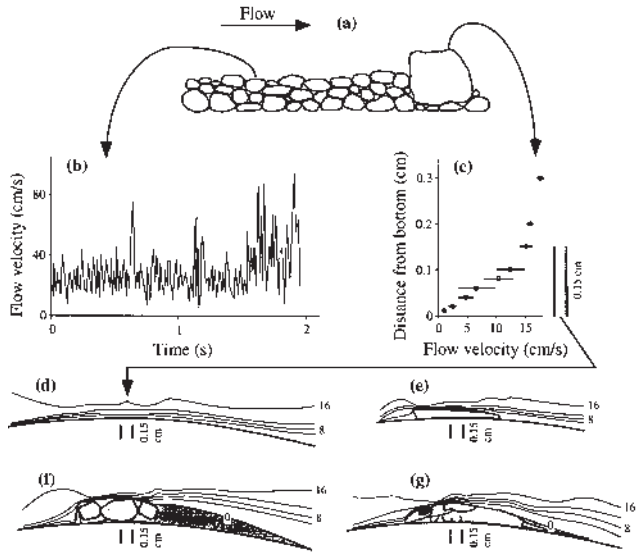

FIGURE 1 Flow patterns near the stream bottom and near three fully grown bottom-dwelling stream insects. (a) Profile of a stream bottom. (b) Short-time velocity variation measured 0.2 cm above the surface of a natural stone that was embedded among other stones of similar size [cf. (a)] in a natural stream riffle. (c)–(g) Flow patterns measured above a curved bottom in a laboratory flume, simulating natural conditions that prevail above bottom elements (large stones, rocks, wooden logs) rising distinctly above neighboring bottom elements [cf. (a)]. (c) Mean and standard deviation of the velocity at different height above the highest point of the curved bottom; the column indicates the height of the layer having a steep velocity gradient. (d) Profile of the curved bottom and isovels [lines of equal velocities, drawn for velocity steps of 4 cm s^{-1} for the same flow as in (c)]; the column indicates the height of the layer having a steep velocity gradient above the highest point of the curved bottom (cf. c). (e)–(g) Changes of the isovel pattern shown in (d) by the cases of two caddisflies [*Micrasema* in (e), *Silo* in (f)] and a dorsoventrally flattened mayfly [*Ecdyonurus* in (g)] [note that (e)–(g) have the same scale as (d) and that velocity fluctuated around 0 cm s^{-1} below the zero isovel]. [Part (b) redrawn and simplified after Hart, D. D., and Finelli, C. M. (1999). Physical–biological coupling in streams: the pervasive effects of flow on benthic organisms. *Annu. Rev. Ecol. Syst.* **30,** 363–395, with permission from Annual Review of Ecology & Systematics, Vol. 30 (1999) by Annual Reviews (www.annualreviews.org); part (g) redrawn and simplified after Statzner, B., and Holm, T. F. (1982). Morphological adaptations of benthic invertebrates to stream flow—an old question studied by means of a new technique (Laser Doppler Anemometry). *Oecologia* **53,** 290–292, and parts (c)–(f) redrawn and simplified after Statzner, B., and Holm, T. F. (1989). Morphological adaptation of shape to flow: microcurrents around lotic macroinvertebrates with known Reynolds number at quasi-natural flow conditions. *Oecologia* **78,** 145–157; (c)–(g) with permission of Springer-Verlag GmbH & Co. KG.]

Whether sessile or mobile, stream insects living in steep velocity gradients can modify the flow patterns to a considerable extent (cf. Fig. 1d and Fig. 1e–g). Frequently, these stream insects are so large (e.g., in Fig. 1f, g) that their lower part is in the laminar flow, their intermediate part is in the layer of elevated turbulence, and their upper part extends beyond the steep velocity gradient. These insects live with different body parts in three different physical "worlds," and

the relative importance of these different worlds experienced by their bodies changes when the insects move. Thus, the flow conditions near larger bottom elements are also physically harsh for stream insects because they variably affect factors of physiological relevance for the insects (e.g., facility of respiration, drag, or lift force).

MOVEMENTS IN THE STREAM ENVIRONMENT

The surface of the stream bottom provides essential resources (e.g., food, oxygen-rich water). To exploit these resources, insects have developed many different strategies that relate to movements at and in the stream bottom, or in the free-flowing water column.

Movements at and in the Stream Bottom

A strategy frequently used to deal with the physical harshness at the stream bottom is temporal avoidance. Particularly, mayfly and stonefly larvae move vertically among the interstices of the stream bottom, to the top of the bottom surface and back into deeper bottom layers that offer shelter from the surface flow. Such movement produces daily and/or seasonal patterns in the presence of these larvae at the bottom surface. However, the dissolved oxygen concentration of the water interferes with the use of interstitial flow shelters. If the oxygen in the interstices drops below a species-specific critical value, the physiological means of regulatory oxygen consumption (gill movements or undulatory body movements) become insufficient to meet respiratory needs. Thus, larvae leave the interstices and crawl to current-exposed locations at the front faces or tops of stones, where oxygen concentrations are usually higher and the elevated flow increases the renewal rate of oxygen at respiratory surfaces.

Another means of staying or moving in the flow near the bottom surface is safe fixation. Often long, curved tarsal claws that enable a good attachment to the bottom surface are found in stream insects (e.g., adult riffle beetles). In addition, hooks or claws situated near the end of the abdomen may be used for bottom attachment. Stream-dwelling larvae of caddisflies, moths, and midges use silk for such attachment. Within the true midges and the caddisflies, many species use silk to build tubelike larval retreats that are fixed at the bottom surface. Caseless polycentropodid and hydropsychid caddis larvae crawling over coarse bottom material glue a silken safety thread in a zigzag line on their path. Other caddisfly larvae that build cases temporarily secure these cases with silk. Finally, black fly larvae spin tiny silk carpets, fix them at the bottom surface, and then attach their posterior abdominal hooks to these carpets.

The stream insects that colonize the most extreme physical conditions are the larvae of the net-winged midges. They can live at rock surfaces where velocities exceed 2 m s^{-1}, fixing themselves to the rock with six ventral suckers. The larvae can precisely adapt their locomotion to a given situation because they move by releasing one or more suckers at a time, having at least two suckers firmly attached to the rock at any moment. This highly specialized locomotory system enables the larvae to move straight upflow at a relative velocity (to the rock surface) of ≈ 0.03 body length s^{-1}, which seems to be a rather poor performance. However, the larvae achieve this movement against water velocities of 2 m s^{-1}. Therefore, their physically relevant speed corresponds to ≈ 300 body lengths s^{-1}, which is an extraordinary performance (to achieve equivalent performance, a human would have to swim 100 m in less than 0.2 s).

Claws, silk, suckers, and other means not detailed here enable stream insects to maneuver actively at or in the stream bottom. Daily movements tend to carry the insects upstream because their body (even of dead nymphs and exuviae) is oriented upstream. Thereby, the larvae of mayflies, stoneflies, and caseless caddisflies may move upstream (mostly at night) several meters per day. In comparison, cased caddisfly larvae often move relatively little during periods of normal flow. Other movements at or in the bottom have seasonal patterns. The nymphs of some mayfly or stonefly species crawl bankward prior to emergence, and old larvae of some limnephilid caddisflies even move toward land in early summer, either to feed on semiaquatic plants or to aestivate outside the stream. In streams that freeze down to the bottom, true midges, dance flies, and some caddisflies can overwinter in the frozen habitat, whereas all other insects actively move away when facing an advancing freezing front. Among the mayflies, the nymphs of many species crawl bankward when the water level of the streams rises during springtime. If the adjacent floodplain of a stream becomes inundated, the nymphs may continue their movements toward the floodplain, where most nymphal growth and development can take place. Other mayfly species not only move toward the stream banks but continue to move upstream near the shoreline at a speed of about 100 m day^{-1}. By following the shoreline of the main stream that bends toward the tributaries, the nymphs move into the tributaries and finally reach marshy areas drained by the tributaries. The nymphs complete their development in these marshy areas.

Active movements of stream insects are also caused by occasionally occurring, extreme hydrological events—floods and droughts. During floods, insects may move toward deeper bottom layers to find flow shelter, but evidence for such behavior is equivocal. A giant water bug living in desert streams has a successful response to flooding. During periods of heavy rainfall that often precede flash floods in these streams, adult and juvenile water bugs move toward the banks, leave the stream, and crawl up to about 20 m over land toward sheltered areas (from where they return to the stream after 24 h). Some studies report that stream insects burrow deeper into the humid sediments to avoid effects of drying, whereas other studies do not confirm such movements. With sinking water level in a stream, insects may also crawl away over land (adult beetles), or they crawl up- or downstream to places where water remains. The speed of insects during such movements can be rather high. For example, the larvae of a cased limnephilid caddisfly crawl downstream at about 12 m h^{-1}.

Finally, the aerial females of a few species of stream insects crawl down on solid objects through the water surface and fix their eggs on the submerged surface of the object, or they oviposit in an aquatic host insect. The first behavior has been primarily observed in mayflies, caddisflies, and blackflies, the second in parasitic wasps.

Movements in the Free-Flowing Water Column

Drifting downstream with the flow is the typical movement of stream insects in the water column. Physically, a passively drifting insect barely moves because the water velocity relative to its body is almost zero. However, relative to the stream bottom it travels at the speed of the flow.

The drift of stream insects is perhaps one of the most frequently studied topics in stream ecology. Given that both the diversity of stream insects and the diversity of running water conditions across the continents are very high, there is evidence for almost any conceivable drift response. For example, the following factors may increase, decrease, or have no effect on the natural drift of stream insects: sun- or moonlight; current velocity; stream discharge; type of the bottom substrate; turbidity, oxygen concentration, ion concentration, or temperature of the water; organic matter; food; predators; microbial pathogens; larger parasites; molting process; benthic density; and age or behavior of the drifting species. Thus, it is difficult to identify clear patterns in the drift of stream insects.

Typically, stream insects drift during the night. This drift may be caused by accidental dislodgement of the insects through the current or by deliberate, active entries into the drift. Active entries occur when a habitat patch is overcrowded and resources (e.g., food, space, preferred flow conditions) are lacking. Approximately 10 to 30% of the insect population of a stream reach may drift during one night, and they may travel between 2 and 20 m during one drift movement. However, caddisfly larvae that build heavy cases from gravel typically have distinctly lower drift rates and shorter drift distances. Insects that produce silk may also have shorter drift distances. Hydropsychid and polycentropodid caddisfly larvae, and black fly larvae, can fix a silk thread to the stream bottom, actively enter into the drift and, prolonging the thread by spinning, rope themselves downstream over several centimeters until they resettle at the bottom. Polycentropodid caddis larvae that have lost contact with the bottom often spin "anchor" threads of 30 cm in length. When such a thread adheres to the stream bottom, the larva is able to return to the bottom by climbing down along the thread. Another means of affecting drift distance is used by drifting larvae of mayflies, stoneflies, and caseless caddisflies. They modify their body posture, thereby either decreasing or increasing their sinking speed, and thus the distance traveled in the water column.

The intensity of everyday drift events usually varies across seasons, either because of seasonal changes in factors affecting the insect drift or because of the seasonal occurrence of a particular developmental stage that drifts more than other stages. A typical example for the first cause is the seasonal change of temperature in temperate streams and the related effects on the insect drift. A typical example for the second cause is found in hydropsychid caddisflies that may have newly hatched first instars that drift at much higher rates than later instars.

Occasional natural events such as floods and droughts, or unnatural ones such as pesticide or pollution spills, also increase drift. During such events and in contrast to the everyday drift, insects also drift during the daytime, and the distances traveled are longer. When discharge increases, and thus near-bottom velocity, many stream insects are dislodged. Among the sessile forms that fix themselves firmly to the stream bottom, those sitting on submerged wood or leaf litter may travel extremely far downstream on dislodged litter pieces that drift for long distances. In contrast, when discharge and thus near-bottom velocity decreases during a drought, many insects release their grasp on the stream bottom and swim upward until they are caught and transported by the free flow in the water column.

Another cause for massive upward swimming of insects from the bottom is relatively low oxygen concentration in the water, a condition typical of streams with organic pollution. Drifting insects may swim, either to avoid sinking so that they stay longer in the water column or to return faster to the stream bottom. Finally, some mayfly larvae swim over short distances when they encounter predators or aggressive conspecifics. These larvae are such good swimmers that they can travel against the current just above the stream bottom, where the velocity is reduced. The larvae of most insect groups swim by body undulations.

An exceptional type of movement among the stream insects is the diving of aerial females to reach submerged oviposition sites. Female black flies can dive through thin water layers flowing over rocks to get foothold at the rock surface, where they affix the eggs. Similarly, female caddisflies dive vertically and swim to oviposition sites below inclined submerged stones.

See Also the Following Articles

Aquatic Habitats • Walking and Jumping

Further Reading

Brittain, J. E., and Eikeland, T. J. (1988). Invertebrate drift—A review. *Hydrobiologia* **166,** 77–93.

Brunke, M., and Gonser, T. (1997). The ecological significance of exchange processes between rivers and groundwater. *Freshwater Biol.* **37,** 1–33.

Frutiger, A. (1998). Walking on suckers—New insights into the locomotory behavior of larval net-winged midges (Diptera: Blephariceridae). *J. N. Am. Benthol. Soc.* **17,** 104–120.

Hart, D. D., and Finelli, C. M. (1999). Physical–biological coupling in streams: The pervasive effects of flow on benthic organisms. *Annu. Rev. Ecol. Syst.* **30,** 363–395.

Koehl, M. A. R. (1996). When does morphology matter? *Annu. Rev. Ecol. Syst.* **27,** 501–542.

Otto, C., and Sjöström, P. (1986). Behaviour of drifting insect larvae. *Hydrobiologia* **131,** 77–86.

Söderström, O. (1987). Upstream movements of invertebrates in running waters—A review. *Arch. Hydrobiol.* **111,** 197–208.

Statzner, B., and Holm, T. F. (1982). Morphological adaptations of benthic invertebrates to stream flow—An old question studied by means of a new technique (Laser Doppler Anemometry). *Oecologia* **53,** 290–292.
Statzner, B., and Holm, T. F. (1989). Morphological adaptation of shape to flow: Microcurrents around lotic macroinvertebrates with known Reynolds numbers at quasi-natural flow conditions. *Oecologia* **78,** 145–157.
Statzner, B., Gore, J. A., and Resh, V. H. (1988). Hydraulic stream ecology: Observed patterns and potential applications. *J. N. Am. Benthol. Soc.* **7,** 307–360.
Vogel, S. (1994). "Life in Moving Fluids." 2nd ed. Princeton University Press, Princeton, NJ.
Williams, D. D., and Williams, N. E. (1993). The upstream/downstream movement paradox of lotic invertebrates: Quantitative evidence from a Welsh mountain stream. *Freshwater Biol.* **30,** 199–218.

Symbionts Aiding Digestion

Andreas Brune
Universität Konstanz, Germany

The interactions of insects with microorganisms range from the cultivation of fungus gardens to intimate associations in which bacteria are housed either within special organs (mycetomes) or intracellularly in dedicated mycetocytes. Many of these associations have nutritional implications; this article will focus on microbial symbionts that colonize the intestinal tract and are directly involved in digestion.

Although insects produce a wide variety of digestive enzymes, many species harbor an intestinal microbiota that converts a substantial portion of the dietary components to fermentation products before they are resorbed by the intestinal epithelia. Interestingly, the occurrence of such associations is always correlated with a dietary specialization, which indicates that the symbiosis provides metabolic capacities that are normally not available to the host.

Digestive symbioses are most common among insects feeding on wood or other lignified plant materials. The most prominent example is that of wood-feeding termites (Isoptera), which represent the only group of insects whose interactions with intestinal microorganisms has been systematically studied. Here, they serve to illustrate principles that most likely govern also other cases of symbiotic associations in which detailed information on the gut microbiota and its function in digestion is lacking.

BIOCHEMICAL BASIS OF LIGNOCELLULOSE DIGESTION

Why would an insect need help in digesting its food, and which benefits would be gained from sharing this resource with an intestinal microbiota? For xylophagous (wood-feeding) insects, the answers are found in the chemical structure and composition of their fiber-rich, low-nutrient diet.

Structure and Composition of Lignocellulose

The major components of plant cell walls are cellulose and hemicelluloses. Although the glycosidic linkages between the sugar subunits of these structural polysaccharides are relatively easy to hydrolyze, the sheer size of the molecules, their primary and secondary structures, and the intimate physical association of the different components, especially in lignified cell walls, impart a considerable recalcitrance to enzymatic digestion.

In cellulose, the linear β-(1→4)-linked polyglucose chains are arranged in microfibrils with a highly ordered, mostly crystalline structure. Cellulose depolymerization is catalyzed by endoglucanases (endo-β-1,4-glucanases), which cleave randomly within the extremely long polyglucose chains, and by exoglucanases (exo-β-1,4-cellobiohydrolases and β-glucosidases), which require free nonreducing ends for their catalytic activity. Because native cellulose contains only one nonreducing end per many thousand glucose units, and because the action of endoglucanases is restricted to the amorphous regions of the microfibrils, an efficient cellulolytic system requires the synergistic action of both types of enzymes.

Enzymatic activity of cellulases is further impeded by the insolubility of cellulose, and depolymerization is usually the kinetically limiting step even in the digestion of pure cellulose. In the cell wall, the cellulose fibrils are intimately associated with hemicelluloses, which comprise a wide variety of homopolymers and heteropolymers of different sugars and sugar acids. Hemicellulose chains are often branched and lack the highly ordered structure of cellulose. The variety of primary structures requires an equal variety of digestive enzymes, and it is not astonishing that the ability to degrade hemicelluloses is a prerequisite for an efficient cellulose degradation.

Sound wood is most difficult to digest because, through lignification, the polysaccharides of the secondary plant cell wall are embedded in an amorphous resin of phenolic polymers, which provide an efficient barrier to enzymatic attack of the polysaccharides. The lignin macromolecule itself is extremely recalcitrant to degradation since the bonds between the subunits cannot be hydrolyzed.

Roles of Symbionts in Digestion

For the reasons outlined above, the degradation of plant cell walls requires the synergistic action of many different enzymes and, in the case of lignified substrates, also a mechanism to break up the lignocellulose complex. The most efficient cellulose and hemicellulose degraders in nature are microorganisms, i.e., bacteria, protozoa, and fungi. Fungi and certain filamentous bacteria (actinomycetes) are also the only organisms that have developed a strategy for the chemical breakdown of lignin. Not surprisingly, insects and other animals have made use of these capacities by using microorganisms as symbionts in the digestion of lignocellulosic food.

In addition to being difficult to degrade, lignocellulose is an extremely nutrient-poor substrate. Whereas nonlignified

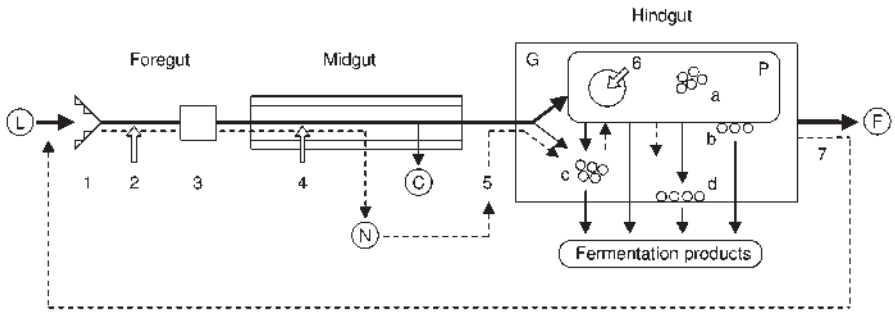

FIGURE 1 Major events in the symbiotic digestion of lignocellulose by wood-feeding lower termites. The bold lines show the path of the insoluble material, the lignin-rich residues of which are released as feces, whereas the thinner lines represent soluble degradation products that are eventually resorbed by the host. The dashed lines indicate the cycling of nitrogenous compounds. Hollow arrows mark the sites where cellulolytic enzymes are secreted. Lowercase letters refer to the different groups of bacteria, which are either endobionts (a) or epibionts (b) of the protozoa, suspended in the gut lumen (c), or attached to the gut wall (d). The scheme has been simplified for the sake of clarity; not all possible interactions are shown. Further details are given in the text. L, lignocellulose; C, carbohydrates; N, nitrogenous compounds; F, fecal matter; G, gut lumen; P, protozoa; 1, mandibles; 2, salivary glands; 3, proventriculus; 4, midgut epithelium; 5, Malpighian tubules; 6, phagosomes; 7, proctodeal feeding.

plant cell walls contain some structural proteins, the C-to-N ratio of sound wood is up to 100-fold higher than that of the insect body. Moreover, a lignocellulosic diet typically lacks most of the essential nutrients required by the animal, such as amino acids, vitamins, and sterols.

In contrast to higher animals, many microorganisms are capable of fixing dinitrogen, assimilating nitrate and ammonia, and synthesizing those amino acids and vitamins essential for the host. Many animals, including insects, have developed means of exploiting these biosynthetic capacities of microorganisms, which include—in the simplest case—the digestion of the intestinal symbionts.

LIGNOCELLULOSE DIGESTION IN TERMITES

The symbiotic digestion of lignocellulose by termites is a complex series of events involving both the host and its gut microbiota, which comprises prokaryotic symbionts (bacteria and archaea) and, at least in lower termites, also protozoa and possibly fungi (yeasts). While the events in foregut and midgut are mainly the result of host activities, the digestive processes in the hindgut are largely controlled by the symbionts (Fig. 1). Many aspects of lignocellulose digestion are common to all termite species, yet there are several noteworthy differences between the phylogenetically lower and higher taxa.

Host-Related Processes

One of the key contributions of the termite to wood digestion is the scraping action of the mandibles and the grinding activity of the proventriculus located at the end of the foregut, which results in the comminution of wood particles to a microscopic size. This is not only a prerequisite for the ingestion of the wood particles by symbiotic protozoa, but it also mechanically destroys the lignin–carbohydrate complexes, which creates an enormous surface area for the digestive enzymes provided by host and symbionts, thereby relieving much of the kinetic limitations of cellulose digestion.

In all insects, the digesta are exposed to a variety of digestive enzymes secreted by the salivary glands and the midgut epithelium. The complement of enzymes released by termites has not been systematically studied, but evidence is accumulating that it comprises enzymes necessary for the digestion of polysaccharides, microbial biomass, and other organic components of the diet and even includes cellulolytic activities contributed by the host (see below). Although experimental data are lacking, one can safely assume that—as in other insects—most of the easily digestible material has been mobilized and resorbed at the end of the midgut.

Insects lacking a pronounced gut microbiota usually have a short hindgut that serves mainly in recovering water and useful electrolytes from the residual material before the feces are voided. The digestive tract of termites, however, is characterized by one or more proctodeal enlargements and may reach enormous dimensions in length and volume (Fig. 2a). The dilatations increase the residence time of the digesta,

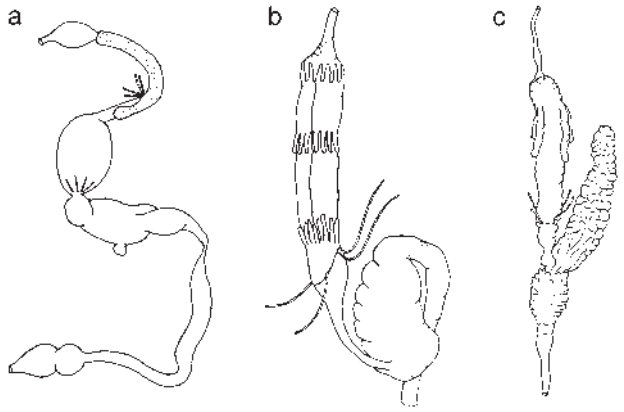

FIGURE 2 Intestinal tracts of insects with fermentation chambers harboring symbiotic microorganisms. (a) *Thoracotermes macrothorax* (Isoptera: Termitidae). (b) *Tipula flaveolineata* (Diptera: Tipulidae). (Reproduced from Buchner, 1928, "Holznahrung und Symbiose," J. Springer, Berlin.) (c) *Potosia cuprea* (Coleoptera: Scarabaeidae). (Reproduced from Werner, 1926, *Z. Morph. Ökol. Tiere* **6,** 150–206. Not drawn to scale.)

thereby prolonging the exposure to the activities of the intestinal microbiota. Moreover, the increased diameter also affects the oxygen status of the "fermentation chambers" because it reduces the proportion of the gut volume rendered oxic by the diffusion of oxygen into the periphery, thus providing a favorable environment for the oxygen-sensitive microbiota and reducing the inevitable loss of fermentation products to aerobic processes (see below).

Fiber-Digesting Symbionts

In lower termites, the solid food particles entering the hindgut are immediately phagocytized by the intestinal protozoa (Fig. 3a). These oxygen-sensitive flagellates, which make up a large fraction of the hindgut volume, are essential for wood digestion and represent a major source of cellulolytic and xylanolytic activities in the hindgut. It appears that the different flagellate species are nutritionally specialized and each species might fill a specific niche in symbiotic digestion.

The hindguts of all families of lower termites are packed with flagellates, whereas the phylogenetically higher termites (family Termitidae) contain a largely prokaryotic microbiota. It was initially assumed that other symbionts took over the cellulolytic function of the protozoa in the course of termite evolution. In the case of the fungus-cultivating termites (subfamily Macrotermitinae), which cultivate basidiomycete fungi (*Termitomyces* spp.) on predigested food in "fungus gardens" located within their nests, the key activities of the symbiotic partner in the fungus combs are extensive delignification and conversion of lignocellulose to fungal biomass. It has been proposed that fungal cellulases ingested by the termites together with the comb material are essential for cellulose digestion in the gut, but this "acquired enzyme hypothesis" is controversial.

All other higher termites do not cultivate fungus gardens, and their intestinal tracts harbor cellulolytic activities that are probably neither produced by symbiotic bacteria nor the result of ingested enzymes. The persisting dogma that higher animals do not produce cellulases has been unequivocally refuted by the demonstration of endoglucanase genes in the termite genome and their expression in the cells of the midgut epithelium and in the salivary glands.

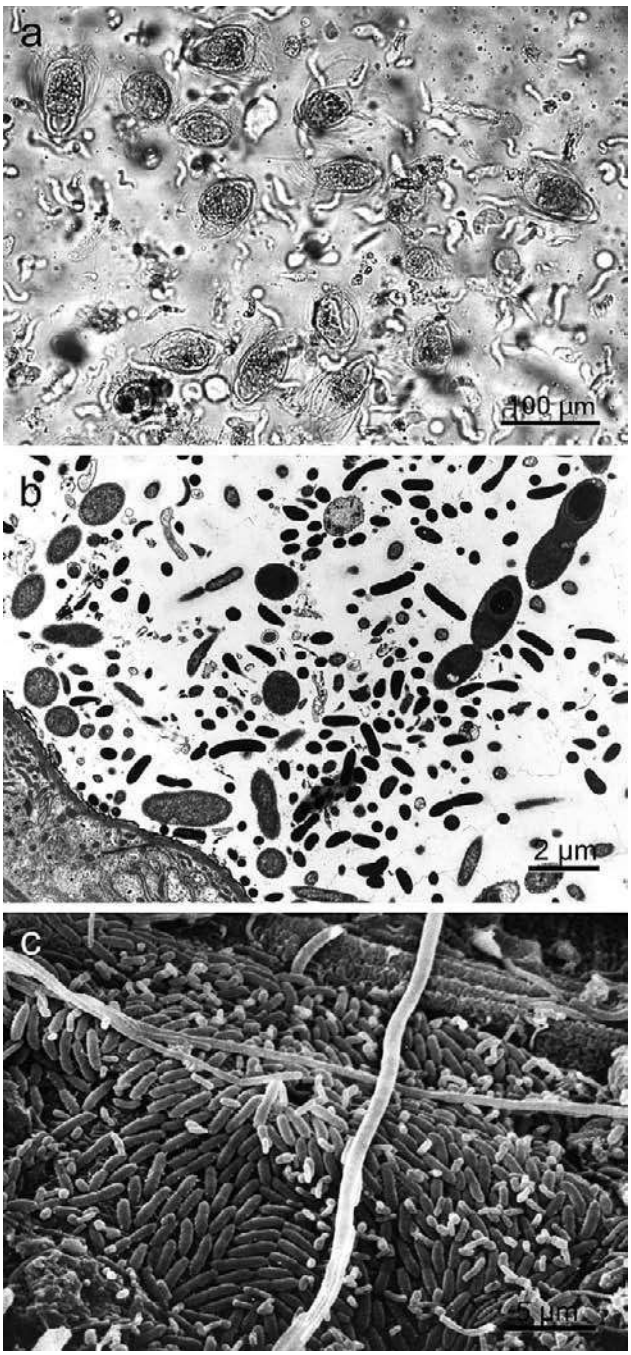

FIGURE 3 Examples of microbial symbionts in the hindgut of *Reticulitermes flavipes* (Isoptera: Rhinotermitidae), a wood-feeding lower termite. (a) Preparation of anaerobic protozoa from the hindgut of a worker larva, showing the large hypermastigote flagellate *Trichonympha agilis,* filled with wood particles, and numerous smaller flagellates (mainly oxymonads, *Dinenympha* spp.). Differential interference contrast photomicrograph taken by U. Stingl. (b) Transverse section through the peripheral hindgut, showing the diverse bacterial microbiota associated with the thin cuticle of the hindgut wall (bottom left). Transmission electron micrograph provided by J. A. Breznak. (c) Preparation of the hindgut wall, showing the dense colonization of the cuticle by numerous rod-shaped and filamentous bacterial morphotypes. Scanning electron micrograph provided by J. A. Breznak.

Noncellulolytic Microbiota

Phylogenetic characterization of the prokaryotic microbiota of termite guts, which was boosted by the advent of cultivation-independent methods during the past decade, has revealed an enormous diversity. Although most of these symbionts do not take part in fiber degradation, they seem to play other important roles in digestion and in the nitrogen metabolism of the hindgut.

METABOLIC INTERACTIONS The original concept of the hindgut metabolism in lower termites assumed that the anaerobic flagellates depolymerize the polysaccharides to

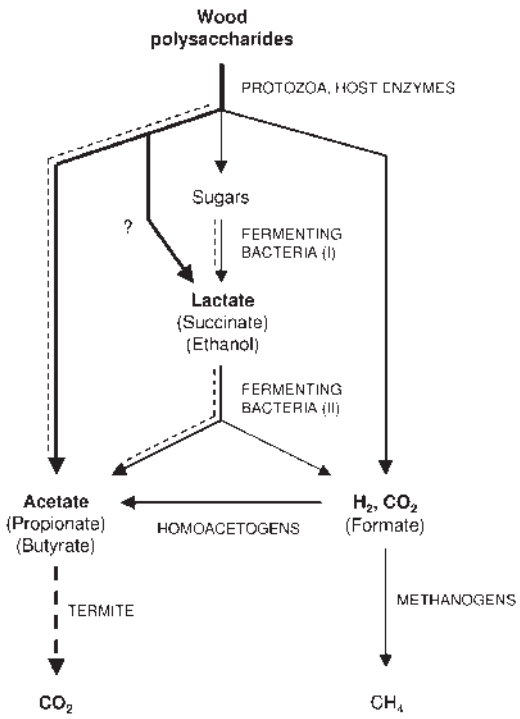

FIGURE 4 Schematic presentation of the metabolic processes involved in the fermentative degradation of polysaccharides in the hindgut of the termite *R. flavipes*. The dashed lines indicate metabolic fluxes that seem to be strongly influenced by the continuous influx of oxygen into the gut periphery.

sugar units, which are then fermented to acetate, hydrogen, and CO_2 as the major products. Recent results indicate that the spectrum of fermentation products released by the diverse protozoa is much wider, giving rise to a variety of intermediates that form the substrates of the endobiotic or epibiotic bacteria colonizing the protozoa. In addition, metabolites released by the protozoa or originating directly from the midgut support numerous bacteria located in the lumen or attached to the wall of the hindgut and are eventually converted to a range of short-chain fatty acids (mostly acetate, propionate, butyrate) that accumulate in the hindgut fluid and that are eventually resorbed by the hindgut epithelium (Fig. 1).

Although it is still difficult to assign functions to individual members of the gut microbial community and to localize exactly the microbial populations involved in these reactions, the major roles of different functional groups are beginning to take shape (Fig. 4). At least two metabolically different groups of prokaryotes are involved in the oxidation of the hydrogen or one-carbon compounds produced in the microbial fermentations: (1) methanogens, which reduce CO_2 to methane, and (2) homoacetogens, which reduce CO_2 to acetate.

Methanogenesis and reductive acetogenesis occur in the hindgut of all termites, although in most wood-feeding species, reductive acetogenesis prevails over methanogenesis as a hydrogen sink and substantially increases the pool of short-chain fatty acids available to the host. In the fungus-cultivating and soil-feeding termites, however, methanogenesis may amount to a considerable portion of the total respiratory activity of the insect. Although methane production is of little use for the host, methane emission acquires global significance due to the enormous biomass of these groups in tropical rainforests and savannahs.

THE ROLE OF OXYGEN The steep gradients of oxygen between oxic gut epithelium and anoxic gut contents drives a continuous influx of oxygen into the gut. Microsensor measurements have shown that oxygen may penetrate 50 to 200 μm into the periphery of the hindgut, leaving only the central portion of the dilated compartments anoxic. The small insect guts have an enormous surface-to-volume ratio, which lends much greater importance to all surface-related processes than in larger animals. Because oxygen removal in the gut periphery is fueled by the fermentative processes in the hindgut lumen, the maintenance of anoxia is not a trivial issue, and there must be a lower size limit for arthropods with a symbiotic digestion.

The bacteria and protozoa colonizing the gut periphery, especially those directly associated with the gut epithelium (Figs. 3b and 3c), have to be specifically adapted to the presence of oxygen at low levels and may even use oxygen as an electron acceptor. Radiotracer analysis of the *in situ* metabolism in the hindgut of the termite *Reticulitermes flavipes* has demonstrated that the high oxygen fluxes significantly influence the metabolic processes in the hindgut (Fig. 4), but the details of this scenario remain to be studied.

Nitrogen Economy of Termites

Wood-feeding termites have developed several adaptations that help to compensate for the low nitrogen content of their diet, all of which involve the biochemical capacities of their gut microbiota. The most important strategy is a combination of conservation and recycling and is reminiscent of the application of organic fertilizers in agriculture.

Like in any other insect, uric acid and urea, the waste products of nucleic acid and protein metabolism, are secreted into the digesta via the Malpighian tubules at the midgut–hindgut junction. However, they are not voided with the feces, but are readily mineralized by the hindgut microbiota. The resulting ammonia is assimilated into microbial biomass; it remains to be clarified whether the intestinal protozoa can also assimilate ammonia directly or acquire combined nitrogen by phagocytosis and digestion of other microbial symbionts.

The nitrogen cycle is closed by the digestion of microbial cells. Because termites cannot access the microbes in the hindgut directly, worker larvae solicit hindgut contents from their nestmates. This behavior, which has been termed proctodeal trophallaxis and is unique to this group of social insects, increases in frequency with the level of nitrogen limitation. Digestion of the hindgut contents and resorption of the nitrogenous products probably take place in the foregut

and midgut (Fig. 1). The efficiency of nitrogen conservation within the colony is increased further by the consumption of exuviae and dead individuals by nestmates.

While nitrogen recycling creates high ammonia concentrations in the hindgut, which allow the maintenance of an active gut microbiota and thus ensure high rates of carbon mineralization, the low nitrogen content of the food still limits the growth of a termite colony severely. However, the presence of bacteria among the hindgut microbiota capable of fixing atmospheric nitrogen may contribute considerably to increasing the nitrogen pool. It has been estimated that the nitrogenase activity in certain *Nasutitermes* termites would be sufficient to double the nitrogen content of a colony within a few years, and stable isotope analysis has revealed that 30 to 60% of the nitrogen in *Neotermes koshunensis* workers is derived via this pathway.

Although most lower termites are strictly xylophagous, the evolutionarily higher termites in the family Termitidae show an enormously successful dietary diversification. The fungus-cultivating Macrotermitinae, which are specialized on degradation of nitrogen-poor plant litter, probably recycle nitrogen within the colony by exploiting the ammonium-assimilating capacities of the fungus. The diet of the other subfamilies (Nasutitermitinae, Termitinae, and Apicotermitinae) ranges from sound wood to lignocellulosic plant materials in different stages of humification, including soil and animal dung. The C-to-N ratio of the diet decreases with increasing humification, which suggests that microbial biomass in the organic matter forms a potential source of nutrition. The extreme gut alkalinity in humivorous termites, which may reach pH 12 in the anterior proctodeal compartment of soil-feeding Termitinae, may have allowed the specialization on nitrogen-rich but recalcitrant diets, because it helps to release fermentable substrates from the ingested humic substances that would otherwise be protected from microbial degradation in soil.

DIGESTIVE SYMBIOSES IN OTHER INSECTS

The occurrence of a specific, autochthonous gut microbiota among insects remains to be systematically studied, but sufficient evidence for the presence of a digestive symbiosis has accumulated for representatives of several insect orders.

Phytophagous insects feeding on protein-rich plant material (e.g., caterpillars) usually have a relatively undifferentiated intestinal tract and, because of the rapid gut passage, digest little or no cellulose. Many xylophagous, detritivorous, and humivorous insect larvae, however, possess hindgut dilations that are missing in closely related species with a different feeding habit. The most prominent examples are among the Coleoptera (e.g., Scarabaeidae) and among the Diptera (e.g., Tipulidae) (Figs. 2b and 2c).

Scarabaeids and tipulids have an actively fermenting gut microbiota, including cellulolytic and hemicellulolytic bacteria and, in the former, also methanogenic archaea, which are attached to brush-like chitinous structures. Also, the guts of omnivorous cockroaches contain a largely prokaryotic microbiota of bacteria and methanogenic archaea, especially when maintained on a fiber-rich diet; many species also harbor anaerobic ciliates with methanogenic endosymbionts in their hindguts. Fiber-degrading protozoa, however, which are so prominent among the Isoptera, are found elsewhere only in *Cryptocercus punctulatus* (Cryptocercidae), which is not astonishing because these wood-feeding cockroaches share a common ancestor with the termites.

Reports on the Orthoptera are somewhat contradictory. The bacteria in the gut of locusts (family Acrididae) have been considered commensals because their absence in germ-free cultures had no obvious effect on the host, although a recent study documented that key components of the pheromone responsible for aggregation of *Schistocerca gregaria* are produced by its gut microbiota. The situation is different in crickets (Gryllidae), which benefit from the presence of a gut microbiota when raised on a fiber-rich diet. In the hindgut of *Achaeta domesticus,* the density of microorganisms is even higher than that in termites, and there are brush-like supports for the attachment of bacteria that resemble those in scarab beetle larvae.

It is very likely that insects other than termites access protein and recycle nitrogen via digestion by microbial symbionts. Proctodeal feeding is a form of social behavior that is restricted to the termites and the wood-feeding cockroach, *C. punctulatus,* but theoretically the consumption of feces would also allow access to the microbial protein. A special adaptation to digestion by symbionts seems to be present in Scarabaeidae, in which a reflux of hindgut contents into the alkaline midgut has been observed. Alkaline midguts are encountered in many humivorous, detritivorous, and coprophagous dipteran and coleopteran larvae and may be an analogy to the extreme alkalinity in the anterior hindgut of soil-feeding termites.

Newly hatched first instars do not possess a gut microbiota, and all larvae have to be reinoculated after each molting because the hindgut intima together with the complete contents is shed in this process. However, many insect larvae eat their exuviae directly after molting, and first instars may pick up their intestinal symbionts together with the food from their environment. Because the oxygen-sensitive flagellates in lower termites do not form resistant cysts, proctodeal trophallaxis is essential for the transfer of the microbial symbionts between nestmates, and it has been speculated that the evolution of sociality and symbiotic digestion in termites might have proceeded hand in hand. This is supported by the strict host specificity of the different flagellates and their apparent coevolution with their hosts.

MUTUALISTS VS COMMENSALS

In the symbiosis between lower termites and fiber-digesting flagellates, both partners are indispensable and the mutual advantage is obvious. However, there are many symbionts in the guts of termites and other insects whose presence—to the

best of our knowledge—appears to be of no advantage to the host. When the benefit of the association is unidirectional, a symbiont is classified as a commensal, and the host might even benefit from its elimination.

However, it is also not unlikely that in such cases we are merely lacking insight into the symbiont's role in the symbiosis. An example might be the methanogenic archaea, which are regularly encountered in the hindguts of termites, scarab beetle larvae, and many cockroaches. Although methane, their sole metabolic product, cannot be utilized by the host, methanogenic archaea seem to be an integral part of the gut microbiota. They colonize the hindgut intima or the intestinal protozoa and are often attached to cuticular spines, which apparently represent specific attachment sites for methanogenic symbionts.

Unfortunately, it is often impossible to eliminate a specific member of the intestinal microbial community selectively, and even when possible, it is difficult to distinguish between the direct and the indirect consequences of their elimination. The only way to gain insight into the complicated microbe–microbe and microbe–host interactions in the intestinal tracts of insects is probably in treating the gut microbiota as a whole.

It might also be worth speculating on the problems that would be created by a general absence of specific gut symbionts and on the efforts involved in keeping a gut sterile. Microorganisms are continuously incorporated with the food, and while a rapid passage of the digesta controls microbial growth by washout (as long as attachment is prevented), the slow gut passage, necessary also for the digestion of lignocellulose by host enzymes, would allow the uncontrolled proliferation of foreign microorganisms. Perhaps the promotion of specific symbionts excludes colonization by potential pathogens.

As long as the access of oxygen to the dilated hindgut compartments is limited, the energy expenditure for any microbial symbionts remains relatively small. A fermentative degradation of carbohydrates to acetate releases only about 10% of the free energy contained in the substrates, and a fraction of this energy is returned as nutritionally valuable microbial biomass. Together with the added value of metabolic properties like nitrogen fixation, ammonia assimilation, and the provision of vitamins, the advantages for the host may be well worth the investment.

CONCLUSION

The digestive symbioses of insects with their intestinal microbiota allow the insects to overcome the severe kinetic limitation of lignocellulose digestion and the nutritional restrictions imposed by this special diet. Although the situation is somewhat reminiscent of the digestive symbiosis encountered in ruminants, it is important to consider the small size of insects, which affects especially the oxygen status of the gut. From a microbiological point of view, insect guts are not simply anoxic fermentors, but axially and radially structured environments with physicochemically distinct microhabitats. We are just beginning to understand the complex interactions within the intestinal microbial communities of termites. Further efforts targeted at identifying the microbial symbionts and their metabolic potentials in this and other, hitherto understudied, groups of insects are sorely needed.

See Also the Following Articles

Digestive System • Isoptera • Nutrition

Further Reading

Bignell, D. E. (2000). Introduction to symbiosis. *In* "Termites: Evolution, Sociality, Symbiosis, Ecology" (T. Abe, D. E. Bignell, and M. Higashi, eds.), pp. 189–208. Kluwer Academic, Dordrecht.

Breznak, J. A. (2000). Ecology of prokaryotic microbes in the guts of wood- and litter-feeding termites. *In* "Termites: Evolution, Sociality, Symbiosis, Ecology" (T. Abe, D. E. Bignell, and M. Higashi, eds.), pp. 209–231. Kluwer Academic, Dordrecht.

Breznak, J. A., and Brune, A. (1994). Role of microorganisms in the digestion of lignocellulose by termites. *Annu. Rev. Entomol.* **39,** 453–487.

Brune, A. (1998). Termite guts: The world's smallest bioreactors. *Trends Biotechnol.* **16,** 16–21.

Brune, A., and Friedrich, M. (2000). Microecology of the termite gut: Structure and function on a microscale. *Curr. Opin. Microbiol.* **3,** 263–269.

Hackstein, J. H. P., and Stumm, C. K. (1994). Methane production in terrestrial arthropods. *Proc. Natl. Acad. Sci. USA* **91,** 5441–5445.

Kane, M. D. (1997). Microbial fermentation in insect guts. *In* "Gastrointestinal Microbiology" (R. I. Mackie, and B. A. White, eds.), pp. 231–265. Chapman & Hall, New York.

Radek, R. (1999). Flagellates, bacteria, and fungi associated with termites: Diversity and function in nutrition—A review. *Ecotropica* **5,** 183–196.

Slaytor, M. (2000). Energy metabolism in the termite and its gut microbiota. *In* "Termites: Evolution, Sociality, Symbiosis, Ecology" (T. Abe, D. E. Bignell, and M. Higashi, eds.), pp. 307–332. Kluwer Academic, Dordrecht.

Watanabe, H., and Tokuda, G. (2001). Animal cellulases. *Cell. Mol. Life Sci.* **58,** 1167–1178.

Symbionts, Bacterial

Michael E. N. Majerus
University of Cambridge

Many species of insect are host to symbiotic microorganisms. These symbionts have traditionally been classified as mutualists (i.e., beneficial), parasitic (harmful), or commensal (neutral). A variety of microorganisms among these symbionts are vertically transmitted, i.e., they are inherited. The survival of these heritable symbionts is almost totally dependent on the success of their host. It follows that these symbionts should evolve characteristics that increase host survival. However, some so-called "ultraselfish" symbionts manipulate their hosts to the symbiont's benefit even when this is to the detriment of the host. The strategies used by such microorganisms include sterilization of noncarriers, feminiza-

tion of genetic males, induction of asexual reproduction, and biasing host sex ratios in favor of females by killing male hosts. The most common of these reproductive parasites are the ultraselfish bacteria.

ULTRASELFISH, MATERNALLY INHERITED BACTERIA

Inherited bacteria that live in the cytoplasm of host cells are transmitted to subsequent generations through the female line. This is because female gametes are heavily resourced with cytoplasm, whereas sperm contain negligible amounts. A maternally inherited bacterium can thus increase its prevalence either by increasing the fitness of infected compared with uninfected hosts or by biasing the sex ratio in favor of females. The phenomenon of cytoplasmic incompatibility (CI), which involves the sterilization of noncarriers, follows the first of these options. Three other strategies, feminization, parthenogenesis induction (PI), and male-killing (M-k), skew populations in favor of females.

CYTOPLASMIC INCOMPATIBILITY

Cytoplasmic incompatibility was first noted when some crosses between *Culex pipiens* mosquitoes from European, Asian, and American populations failed to produce offspring or produced offspring only when the crosses were carried out in one direction. Observed patterns of reproductive successes and failures demonstrated maternal inheritance (usually 95–100% efficiency), indicating the possible involvement of some cytoplasmic factor, hence the name given to the phenomenon. A bacterium of the genus *Wolbachia* was found to be responsible. Subsequently, CI has been reported from many insect orders.

Cytoplasmic incompatibility occurs when a male carrying a CI *Wolbachia* mates with a female that does not bear the same *Wolbachia.* The bacteria allegedly secrete a chemical into the sperm of their host that kills zygotes formed within the female parent if they do not bear the *Wolbachia.* In essence then, an uninfected female that mates with an infected male is rendered sterile thereafter. Other mating combinations are not affected.

CI *Wolbachia* can give rise to two patterns of reproductive failure. If one host population has the *Wolbachia,* whereas the other lacks it, matings fail in only one direction; i.e., the incompatibility is unidirectional. If, however, each population harbors a different strain of CI *Wolbachia,* matings fail in both directions and the incompatibility is said to be bidirectional. For example, in the fruit fly *Drosophila simulans,* both unidirectional and bidirectional incompatibility have been recorded. Indeed, some CI host species carry more than one strain of CI *Wolbachia,* with individuals in some populations hosting two strains concurrently.

Host Range and Identity of the Agents Causing CI

Cytoplasmic incompatibility is commonly reported when crosses of insects from geographically distinct populations are made. It has been reported most often from species of Diptera. However, Coleoptera, Heteroptera, Lepidoptera, and Hymenoptera have also been found to exhibit the phenomenon and other orders will certainly be added to the list in the near future.

The CI phenotype can be cured by treatment with antibiotics or by exposing infected strains to high temperatures (≈37°C) for several days. This suggests that the agent responsible for CI is a bacterium and in every case in which the agent responsible has been identified it is a bacterium of the *Wolbachia pipientis* complex. This complex represents a widespread group of α-Proteobacteria, whose members are known to infect hosts from all the major orders of insects as well as some other arthropods (mites, spiders, isopods) and phyla (nematodes). Not all *Wolbachia* infections result in CI. The bacterium is known to manipulate hosts in a variety of other ways and some appear to have no overt effect on their hosts, whereas some are beneficial. Estimates from molecular surveys using *Wolbachia*-specific probes suggest that between 10 and 20% of insect species harbor this bacterium.

Phylogenetic analysis has shown that *Wolbachia* in taxonomically very different hosts may be more similar to one another than *Wolbachia* in very closely related hosts, suggesting that *Wolbachia* may transfer between host species.

Mechanisms of Incompatibility

The precise mechanics of incompatibility are not well understood, partly because the mechanism of incompatibility appears to vary between host species. In *C. pipiens,* incompatibility results from failure of sperm bearing CI *Wolbachia* to fuse correctly with female gametes that lack the same *Wolbachia.* In fruit flies *(Drosophila)* and moths *(Ephestia kuehniella),* embryo development appears to be suppressed at an early stage. The situation in the jewel wasp, *Nasonia vitripennis,* is a little more definite. Here *Wolbachia* interferes with condensation of the paternal chromosome set during the first mitotic cell division of the embryo so that the paternal chromosome set is lost. All progeny that get the *Wolbachia* thus carry just the maternal chromosome set, are thus haploid, and develop as males. Here CI results in a male-biased sex ratio.

The Population Dynamics of CI

Wolbachia that cause CI spread through host populations because infected females have an advantage over uninfected females, which are rendered sterile once they have mated with an infected male. The precise dynamics of CI are influenced by several factors, including (1) the vertical transmission of the bacterium, (2) the fitness costs or benefits

it imposes on its host and possibly horizontal transmission, and (3) host suppression systems.

In general, the prevalence levels of CI *Wolbachia* in infected populations are high and often approach 100%. Few studies of the population dynamics of CI have been conducted. The most intensive studies involve drosophilid hosts. For example, in *D. simulans,* in California, rapid spread of a CI strain of *Wolbachia* was monitored during the 1980s and the 1990s. A theoretical model, based upon the levels of incompatibility, the vertical transmission efficiency of the bacterium, and the assumption that bearing the symbiont had no direct fitness effect on its bearer, produced theoretical predictions close to the observed rates of increase.

FEMINIZATION

Microbes and Sex Determination

Microorganisms of two types are known to cause feminization of their hosts. One type, microsporidians and other simple protists, as yet are known to infect only crustaceans. The other type, bacteria of the genus *Wolbachia,* causes feminization in both crustaceans and insects. In essence, what both of these groups of microorganisms do is turn genetic males into females.

The best studied cases of feminization involve crustaceans, such as the woodlouse, *Armadillium vulgare.* However, many aspects of the feminization system in crustaceans are probably similar to those in insects, of which two species of moth, *Ostrinia furnacalis* and *Ostrinia scapulalis* harbor feminizing *Wolbachia.*

Sex determination in *A. vulgare* and in the Lepidoptera is normally based upon sex chromosomes; in the Lepidoptera females are the heterogametic sex, carrying two different sex chromosomes (WZ), and males are homogametic.

In *O. furnacalis* all-female strains were first reported in 1998. A second example of feminization has been reported in *O. scapulalis* subsequently. Antibiotic treatment has shown the female biases to be tetracycline sensitive, and molecular analysis has revealed a *Wolbachia.* In both moths this bacterium is taxonomically similar to that causing feminization in *A. vulgare.*

In parallel to *A. vulgare,* cure of feminizer lines with antibiotics led to a strong male bias in progeny of cured females. These moths are ancestrally female heterogametic, with the result that feminized ZZ males, when cured of the bacterium and mated by normal (ZZ) males, can produce only male offspring. It is perhaps no coincidence that in the examples of *Wolbachia*-induced feminization, the host is ancestrally female heterogametic, and further instances of *Wolbachia*-induced feminization in the Lepidoptera and other insect taxa with heterogametic females are likely to be discovered in the near future.

Mechanism of Feminization

The mechanism of feminization in the two lepidopteran examples has not been determined. However, in *A. vulgare,* the difference between developing into a male and developing into a female appears to depend on the activity of a single gene that blocks the expression of one or more genes that cause the differentiation of the androgenic gland.

PARTHENOGENESIS INDUCERS

Microbe-Induced Parthenogenesis in the Hymenoptera

Much of the biotic world indulges in sexual reproduction. However, some organisms reproduce asexually, either ancestrally or secondarily when asexuality has evolved from sexuality. Secondary asexual reproduction, or parthenogenesis, occurs with new individuals arising from unfertilized eggs.

Various types of parthenogenesis exist. Normal haplodiploid species, in which females result from fertilized eggs and males from unfertilized eggs, are called arrhenotokous. In deuterotoky, both males and females arise from unfertilized eggs, whereas thelytoky involves the production of only female offspring without fertilization. The production of only female offspring, without the need for males, raises the possible involvement of cytoplasmic symbionts whose interests are favored by female hosts, which can vertically transmit them. This involvement is realized in a number of haplodiploid Hymenoptera and Thysanoptera.

The first report of bacterially induced thelytoky was in the parasitoid hymenopteran *Trichogramma pretiosum.* Here, administration of various antibiotics or of high temperatures to thelytokous females led to the production of sons among the offspring. This led to the deduction that some sort of cytoplasmic microorganism, probably a bacterium, caused thelytoky in this species. Microscopic examination revealed a bacterium present in eggs of some thelytokous lines, but not in antibiotic-treated, temperature-treated, or arrhenotokous lines. Molecular genetic analysis revealed the presence of a *Wolbachia.*

Evidence of microbe-induced parthenogenesis has been revealed subsequently in over 30 species of parasitoid wasp, in the predatory thrip *Franklinothrips vespiformis,* and in the colembollan *Folsomia candida,* as well as in some species of mite (*Bryobia* spp.). In all cases in which the microbe involved has been identified, it is a *Wolbachia.* The *Wolbachia* involved do not form a single taxonomic group within the *Wolbachia,* but are intermixed with *Wolbachia* that affect hosts in other ways (e.g., CI and male-killing). This suggests that PI has evolved independently several times in the *Wolbachia.* Alternatively, the genes that cause PI may be horizontally transferred between *Wolbachia,* through DNA exchange involving virus-like particles or plasmids. A third possibility is that the same *Wolbachia* causes different effects in different hosts, for example, parthenogenesis in haplodiploids and CI or male-killing in diploids. Phylogenetic analysis of a number of PI *Wolbachia,* particularly in the genus *Trichogramma,* has provided evidence for interspecific horizontal transmission of the symbiont.

Mechanism of Parthenogenesis Induction

Investigations of parthenogenetic organisms have shown variation in the way in which diploid offspring can result from unfertilized eggs. This diversity can be divided into two basic groups: either meiosis is modified with the diploid number of chromosomes being retained or, following meiosis, the diploid chromosome number is restored after the formation of a single pronucleus, usually by the fusion of two haploid mitotic products. In the examples of *Wolbachia*-induced parthenogenesis that have been studied microscopically, the latter of these two routes applies. In infected females, both sets of chromosomes migrate to the same pole during the first mitotic division of the meiotic product. This results in a single diploid mitotic product. The ways in which *Wolbachia* affect the behavior of host chromosomes during the first mitotic division are currently not known.

Dynamics of *Wolbachia* Infection

In most Hymenoptera with PI *Wolbachia,* thelytoky has been fixed. However, in many *Trichogramma* species both thelytokous and arrhenotokous individuals occur together. The reasons for polymorphism in some species and monomorphism in others are as yet not clear, but incomplete vertical transmission, negative-fitness effects of *Wolbachia* on their host, and suppresser genes may all have an influence.

MALE-KILLERS

Types of Male-Killers

Male-killing is perhaps the least sophisticated of the mechanisms of sex ratio distortion practiced by inherited symbionts. The basic mechanism involved is for the symbiont to kill male but not female hosts. Among arthropods, M-k has been reported from five orders of insects (Hemiptera, Coleoptera, Lepidoptera, Hymenoptera, and Diptera) and two species of mite. In addition, perhaps because M-k is easily evolved, M-k's themselves are taxonomically diverse, with α- and γ-Proteobacteria, Mycoplasma, Flavobacteria, and Microsporidia being recorded. Two M-k strategies have been reported. The first strategy, which involves killing male hosts late in their development (typically in the final larval instar), allows symbionts to replicate within their host, so that large numbers of individuals are released from the host's corpse. Here the population dynamics of the M-k depend on both horizontal and vertical transmission. Reported late M-k's are confined to Microsporidia infecting mosquitoes.

The second strategy involves killing males early in their development, typically during embryogenesis. In killing its host, the M-k dies, but in doing so it increases the fitness of copies of itself in female hosts. The death of the M-k in male hosts means that early M-k does not involve significant levels of horizontal transmission. The increase in the fitness of infected females resulting from M-k is called fitness compensation, because for a M-k to invade a host population, infected females must have higher fitness than uninfected females to compensate for the death of the infected males. The nature of cytoplasmic bacterial reproduction means that many or all progeny of an infected mother will carry copies of the symbiont that are virtually identical by descent. If by killing male hosts the symbiont thereby increases the fitness of the dead males' sisters, a fitness benefit will also accrue to "exact" copies of the male-killer within these females.

Two advantages of M-k have been identified. First, the probability of mating with close relations, specifically brothers, will be reduced, thus avoiding the harmful effects of inbreeding. An infected female whose brothers have been killed has no siblings to mate with. The second advantage involves resource reallocation. If the resources that would have been consumed by sons of an infected female become available specifically to female siblings these females will benefit.

Early M-k has been recorded from many insect hosts, and to date, all known early M-k's are bacteria. A diverse array of bacteria is involved, suggesting that M-k has evolved independently many times.

The distribution of M-k's among insects is not random. Certain groups have behavioral traits and ecologies that provide a context in which high resource benefits are obtained by daughters of infected females specifically as a result of the death of males. In these instances, the resources made available to females from the death of males become preferentially available to infected compared with uninfected females.

Hot spots for male-killing are known in the milkweed bugs (Hemiptera), nymphalid butterflies (particularly of the genus *Acraea*), and ladybug beetles (Coccinellidae).

Early Male-Killing in Ladybugs

Female-biased sex ratios were first reported in some families of the ladybug *Adalia bipunctata* in Russia in 1947. The female-biased sex ratio trait was shown to be maternally inherited. More recent work has demonstrated that the trait is curable with both antibiotics and heat treatment, suggesting that a bacterial agent is responsible. Male-killers have been discovered in eight other species of ladybug. In each case, the female sex ratio bias is associated with the death of approximately 50% of eggs before they hatch (Fig. 1). This leads to the interpretation that the eggs of infected females have a normal sex ratio at fertilization, but that the secondary sex ratio is biased toward females because male embryos die before hatching.

Experimental work has shown that female larvae from mothers bearing a M-k gain a considerable nutritional advantage from consuming the contents of their dead brothers' eggs. This means that they are larger when they disperse from their egg clutch than larvae from uninfected mothers, and they have longer to find and subdue their first prey before dying of starvation.

Detailed examination of resource reallocation in coccinellids has provided a set of criteria that are predicted to allow the

FIGURE 1 An egg clutch laid by a male-killer-infected female *Harmonia axyridis.* Approximately half of the larvae (the females) have hatched and the neonates have begun consuming the soma of the dead male eggs.

invasion of M-k's. The ladybug must feed on an ephemeral and often limited food, such as aphids. Neonate larval mortality from starvation must be significant. The ladybug must also lay eggs in clutches and neonate larvae must consume unhatched eggs in their clutch. Lack of one or more of these criteria should make a ladybug immune to M-k invasion. All ladybugs known to harbor M-k's possess all the relevant criteria.

The Mechanism of Male-Killing

Little is known about of the mechanism by which bacteria discriminate between male and female hosts and kill only the former. It is known that the sex determination systems of species that host M-k's vary. So, for example, ladybugs have XX females and XY males, butterflies of the genera *Acraea* and *Danaus* have ZZ males and ZW females, and the hymenopteran parasitoid *N. vitripennis* has typically arrhenotokous haplodiploid sex determination, yet all have M-k's. Thus M-k's do not all detect that they are in males by identifying a specific region of a male-specific chromosome.

Consequences of Population Sex Ratio Distortion

Rather more is known about the evolutionary effects of M-k's on their hosts. The prevalence of M-k's varies from less than 5% to over 95% in different insect groups. In the multicolored Asian ladybug *Harmonia axyridis,* the prevalence of a male-killing *Spiroplasma* is closely correlated to the population sex ratio. At low prevalence, the population sex ratio deviates little from 50% of each sex. However, at high prevalence (50% of females infected) over two-thirds of the population are female. Similarly female-biased populations of *Ad. bipunctata* have been reported. More spectacularly, in the butterflies *Acraea encedon* and *Ac. encedana,* populations in which females comprise over 95% occur. Such highly biased sex ratios have the potential to impact on the evolution of reproductive strategies. This potential may be the cause of a number of unusual features of the courtship and copulatory behavior of these species.

First, both *Ad. bipunctata* and *H. axyridis* exhibit a variety of genetically determined mate choice patterns, including, unusually, weak male choice of females. Furthermore, in *Ad. bipunctata,* the investment that a male makes in an individual copulation decreases as the ratio of females to males increases. Male *Ad. bipunctata* may transfer one, two, or three spermatophores (sperm packages) during a single copulation. In populations with a 1:1 sex ratio, males transfer, on average, just under two spermatophores. Conversely, in populations approaching two females for each male, males rarely transfer more than a single spermatophore (average 1.2).

In highly female-biased populations of the butterflies *Ac. encedon* and *Ac. encedana,* many females die without mating. This suggests that males are limiting. The result has been the evolution of sex-role-reversed behaviors. The most striking of these is of female lekking behavior. Male leks, which are resource-lacking arenas where males congregate and compete for mates, are known in many taxa. In *Ac. encedon* and *Ac. encedana,* it is virgin females that congregate at resource-lacking sites. Within these arenas females compete for matings. Furthermore, there is some evidence to suggest that males preferentially mate with females that are free from male-killer infection.

IMPLICATIONS AND APPLICATIONS

Ultraselfish inherited bacteria adopt a strategy of harming their host to aid their own persistence and spread. Early reports of such symbionts, which caused abnormalities in host sex ratios or reproductive patterns, were long thought of as evolutionary oddities. Over the past 2 decades, that view has been changed as molecular genetic advances have revealed that many insects are host to these reproductive manipulators. Surveys have shown that 10 to 20% of insect species harbor bacteria of a single genus, *Wolbachia.* It is likely that similar surveys directed toward other clades of inherited bacteria will lead to the conclusion that species of insects that host antagonist bacteria are in the majority.

Invasion by ultraselfish inherited bacteria can have far-reaching effects on the evolution of hosts. It has already been shown that the selection pressures imposed by these cytoplasmic symbionts have led to the evolution of nuclear suppresser genes in some insects. Such genes provide evidence of the intragenomic conflict that results from the maternal inheritance of these microbes. Invasion and spread of any maternally inherited symbiont through a host population will also cause an immediate decrease in the diversity of host cytoplasmic organelles, such as mitochondria, for they are inherited together. The spread of a symbiont, therefore, pulls the organelles of its original female host through the host population with it. Further effects of CI, PI, feminizing, and M-k bacteria undoubtedly await discovery. Given the antagonistic interactions between deleterious symbionts and their

hosts, it is difficult to believe that a complete understanding of the biology of any insect species can be achieved without consideration of the symbionts it carries.

Deleterious inherited bacteria, such as *Wolbachia,* currently have a high scientific profile. This is at least partly because they may have a future role in pest control. Several strategies of use can be envisaged. Endosymbionts that cause female biases in host species that are already used as biological control agents, such as parasitoid wasps or aphidophagous coccinellids, may be used to increase the impact of the control agent or reduce costs of a control program. Alternatively, mass release of CI agents could be used to cause sterility in recipient populations, much as mass releases of males sterilized by radiation have been used to control *Cochliomyia hominivorax.* A third strategy would be to use these bacteria to transfer genes. If a symbiont bearing "useful" genes were likely to spread through a target population to fixation, as is likely of a CI agent, the useful genes need be introduced by only a relatively small initial release.

The study of ultraselfish inherited bacteria in insects has blossomed over the past decade. The focus on this group of extraordinary parasites is likely to increase as the breadth of their influence on host evolution becomes more clearly understood.

See Also the Following Articles

Genetic Engineering • Parthenogenesis • Reproduction, Female • Reproduction, Male • Sex Determination • Wolbachia

Further Reading

Hoffmann, A. A., and Turelli, M. (1997). Cytoplasmic incompatibility in insects. *In* "Influential Passengers" (S. L. O'Neill, A. A. Hoffmann, and J. H. Werren, eds.), pp. 42–80. Oxford University Press, Oxford.

Hurst, G. D. D., Hurst, L. D., and Majerus, M. E. N. (1997). Cytoplasmic sex-ratio distorters. *In* "Influential Passengers" (S. L. O'Neill, A. A. Hoffmann, and J. H. Werren, eds.), pp. 125–154. Oxford University Press, Oxford.

Jiggins, F. M., Hurst, G. D. D., and Majerus, M. E. N. (1999). Sex ratio distorting *Wolbachia* causes sex role reversal in its butterfly host. *Proc. R. Soc. London. B.* **267,** 69–73.

Majerus, M. E. N., and Hurst, G. D. D. (1997). Ladybirds as a model system for the study of male-killing symbionts. *Entomophaga* **42,** 13–20.

Rigaud, T. (1997). Inherited microorganisms and sex determination of arthropod hosts. *In* "Influential Passengers" (S. L. O'Neill, A. A. Hoffmann, and J. H. Werren, eds.), pp. 81–101. Oxford University Press, Oxford.

Stouthamer, R. (1997). *Wolbachia*-induced parthenogenesis. *In* "Influential Passengers" (S. L. O'Neill, A. A. Hoffmann, and J. H. Werren, eds.), pp. 102–124. Oxford University Press, Oxford.

Werren, J. H., and O'Neill, S. L. (1997). The evolution of heritable symbionts. *In* "Influential Passengers" (S. L. O'Neill, A. A. Hoffmann, and J. H. Werren, eds.), pp. 1–41. Oxford University Press, Oxford.

Systematics

see *Nomenclature and Classification*

T

Taste

see *Chemoreception*

Taxonomy

see *Nomenclature and Classification*

Teaching Resources

John H. Acorn and Felix A. H. Sperling
University of Alberta, Canada

In the classroom, insects are perennially popular subjects. Small enough and odd enough to be both simple to manage and intriguing to students, they are sometimes taught about for their own sake and sometimes used as models for the teaching of concepts in biology as a whole. For those who teach about insects in the entomological sense, the best resources have always been a good entomological library, a selection of live insects in culture, a well-maintained collection of preserved specimens, and somewhere to take students to see live insects behaving in a natural fashion. These are still the basics, along with contagious enthusiasm and genuine personal knowledge of the subject. For those who use insects as conceptual models, there are a few standard species (cockroaches, crickets, fruit flies, and others) for which methods of culturing and experiments for demonstrating various principles are well established and generally available.

In recent years, two changes in social attitudes have strongly affected entomological teaching. First, globalization, especially through the Internet, has allowed people to pursue their interests at any time and place, with consequent opportunities to compare information and to communicate on a global stage. Second, localization has also occurred, at least in the sense of a surge in pride in the natural features and biodiversity that can be found in one's own backyard. One might argue that this is a bit of a reaction to globalization as well. The combination of these two seemingly contradictory movements has meant that a deluge of information and resources is available and relevant at varying scales. However, this information must be filtered carefully, because much of what is available now, especially through the Internet, is often not peer reviewed nor generated by trained entomologists.

WEB SITES

Web sites are the most important recent development in entomological teaching. They take various forms, such as clearinghouses to other sites, museum databases, exercises for instruction at various levels, and taxon-based compilations and/or species accounts. The following list of Web sites offers a general overview of what is available, as well as providing links to other sites. Keep in mind, however, that the Web is an ever-changing thing and that sites come and go, as well as change address. In general, institutionally affiliated Web sites provide the most reliable information, but this is not always so. Luckily, entomological Web sites as a group have been less affected by pseudoscience and the fringe element than have many other areas in cyberspace.

- Coleoptera Home Page: http://www.coleoptera.org/. This site is an excellent starting point for Web resources about beetles.
- Electronic Resources on Lepidoptera: http://www.chebucto.ns.ca/Environment/NHR/lepidoptera.html. This site provides many links to Web resources for the study of butterflies and moths.

- Entomology Index of Internet Resources: http://www.ent.iastate.edu/List/. The hub of the virtual entomological wheel, this site has links to other sites on just about every entomological topic imaginable.
- Ephemeroptera Galactica: http://www.famu.org/mayfly/. "Everything you ever wanted to know about mayflies" is at this site.
- Insect Physiology Online: http://lamar.colostate.edu/~insects/. This site is "an interactive, referenced teaching resource" for insect physiology.
- Introduction to the Odonata: http://www.ucmp.berkeley.edu/arthropoda/uniramia/odonatoidea.html. This site is a fine starting place for resources on damselflies and dragonflies.
- NeuroWeb: http://entowww.tamu.edu/research/neuropterida/neuroweb.html. "The neuropterist's home page," with links and information about lacewings, ant lions, and the like, is at this site.
- Orthoptera Species File Online: http://viceroy.eeb.uconn.edu/Orthoptera. This site is "a taxonomic database of the world's orthopteroid insects."
- Crickets in the Classroom: http://members.attcanada.ca/~ecade/cricket_n_classroom.html. This is a good site for those who would like to use crickets to demonstrate various biological principles.
- Social Insects Web: http://research.amnh.org/entomology/social_insects/. A starting point for resources on social hymenopterans, this site has an emphasis on the ants.
- The Diptera Site: http://www.sel.barc.usda.gov/Diptera/. This site is a clearing house for studies of flies, from "young dipterists" to current research.
- The Tree of Life: http://ag.arizona.edu/tree/phylogeny.html. This is a superb site that not only provides information on as many types of living things as possible, but also places this information in an evolutionary context through an explicitly phylogenetic structure. Insects are particularly well covered here.
- UF Book of Insect Records: http://ufbir.ifas.ufl.edu/. This site is a fascinating compilation of subject matter, from "fastest flier" to "least specific sucker of vertebrate blood."

ELECTRONIC MAILING LISTS

Posting a question to a mailing list, or a listserve, is often the quickest way to answer an obscure question. Keep in mind, however, that experts on the list may resent the implication that they are being used to do your work for you. Therefore, before posting a question, make a serious attempt to find the answer in the standard sources first. You may have to subscribe to the list in order to post a question, or you can have a current subscriber post the question on your behalf. Your posting may also be archived, and as a result you may be quoted some time later on. Professionalism is therefore a wise path to follow. A good way to learn how to ask questions appropriately is to visit the Insect Question Page at http://www.ent.iastate.edu/mailinglist/bugnet/question.html.

OUTREACH PROJECTS

A number of outreach projects have entomological themes. A good example is Monarch Watch, a program of the University of Kansas that calls upon schools to help in the monitoring of monarch butterflies *(Danaus plexippus)* and that in return provides educational materials. Other examples of "citizen science" include the Canadian Nature Federation's Ladybug Survey, which may be explored at http://cnf.ca/beetle/guide.html.

Various institutions and organizations combine a Web presence with an opportunity for school visits. One prominent example in the United States is the Young Entomologists Society (http://members.aol.com/YESbugs/mainmenu.html), which offers programs for schools and groups, workshops for educators, a mobile exhibit including live insects, and an education center, located in Michigan. Many universities also have the equivalent of insect hotlines, or extension entomologists whose duties include answering agricultural and urban entomology questions.

BOOKS AND LIBRARIES

The Internet is a wonderful tool, but the library is still the rightful place for scholarly information in entomology. Books and journals are for the most part much more carefully written, reviewed, and designed than their Web-based counterparts, and therefore they are the guardians of mainstream entomological knowledge.

Here we will not cite particular faunal guides, but it is worth noting that there are interesting trends apparent in recently published works. For some groups, formerly obsolete treatments are being updated or replaced on a broad geographic scale. An example is *Dragonflies of North America* by J. G. Needham, Michael L. May, and Minter J. Westfall (2000, Scientific Publishers, Gainesville, FL). In other instances, local faunas are being treated separately and in greater detail, often for the first time (for example *The Butterflies of West Virginia and Their Caterpillars,* Thomas J. Allen, 1997, University of Pittsburgh Press). Still other books are the results of citizen-science survey projects, such as *The Millennium Atlas of Butterflies in Britain and Ireland* (James Asher, Martin Warren, Richard Fox, Paul Harding, Gail Jeffcoate, and Stephen Jeffcoate, 2001, Oxford University Press). One faunal treatment, *The Insects of Australia* (listed below) goes beyond geography and has become a classic of entomology in general.

For the study of most groups of insects, the traditional methods of collecting and preserving specimens are promoted, but for some, and especially for butterflies, many recent books advocate either a hands-off approach to insect study or catching and releasing the insects one finds. This trend seems directly related to geography—dense populations of more-or-less urbanized people, sensitive to conservation issues, are likely to support "nonconsumptive" entomology, despite a lack of evidence that this is a legitimate conservation concern, especially compared with the effects of habitat loss and introduced species.

The following is a list of 25 books, written in English, that are clearly entomological classics. Some are impressive summaries of one aspect of entomological science, and some are masterfully written explorations of particular topics. With them, an instructor should be able to quickly answer most entomological questions. A good entomology library will contain most of them, but the odds are that every entomologist has at least a few of these volumes close at hand.

- *Bugs in the System: Insects and Their Impact on Human Affairs.* M. R. Berenbaum, 1995, Addison–Wesley, Reading, MA.
- *Life on a Little Known Planet,* rev. ed. H. E. Evans, 1993, Lyons & Burford, New York.
- *Man Eating Bugs.* P. Menzel and F. D'Aluisio, 1998, Ten Speed Press, Berkeley, CA.
- *The Forgotten Pollinators.* S. L. Buchmann and G. P. Nabhan, 1996, Island Press, Covelo, CA.
- *The Hot-Blooded Insects: Strategies and Mechanisms of Thermoregulation.* B. Heinrich, 1993, Harvard University Press, Cambridge, MA.
- *The Insect Societies.* E. O. Wilson, 1971, Belknap Press of Harvard University Press, Cambridge, MA.
- *The Evolution of Insect Mating Systems.* R. Thornhill and J. Alcock, 1983, Harvard University Press. Cambridge, MA.
- *Principles of Insect Morphology.* R. E. Snodgrass, reprinted 1993, Cornell University Press, Ithaca, NY.
- *The Insects: Structure and Function,* 4th ed. R. F. Chapman, 1998, Cambridge University Press, Cambridge, UK.
- *Entomology,* 2nd ed. C. Gillott, 1995, Plenum, New York.
- *The Torre-Bueno Glossary of Entomology,* rev. ed. J. R. de la Torre-Bueno, 1989, New York Entomological Society, American Museum of Natural History, New York.
- *The Encyclopedia of Insects.* C. O'Toole, ed., 1995, Facts on File, Inc. (online).
- *Insects of Australia,* 2nd ed., Vols. I and II. CSIRO, 1991, Melbourne University Press.
- *An Introduction to the Study of Insects,* 6th ed. D. J. Borror, C. A. Triplehorn, and N. F. Johnson, 1992, Harcourt Brace, New York.
- *Aquatic Insects of North America,* 3rd ed. R. W. Merritt and K. W. Cummins, 1996, Kendall/Hunt, Dubuque, IA.
- *Immature Insects,* Vols. 1 and 2. F. W. Stehr, 1987 and 1991, Kendall/Hunt, Dubugue, IA.
- *Biology of Spiders.* R. F. Foelix, 1982, Harvard University Press, Cambridge, MA.
- *Dragonflies: Behavior and Ecology of Odonata.* Philip S. Corbet, 1999, Cornell University Press, Ithaca, NY.
- *The Biology of the Coleoptera.* R. A. Crowson, 1981, Academic Press, San Diego.
- *Hymenoptera of the World: An Identification Guide to Families.* Henri Goulet and John T. Huber, eds., 1993, Agriculture Canada Publication 1894/E, Ottawa, Ontario.
- *The Ants.* B. Hölldobler and E. O. Wilson, 1990, Belknap Press of Harvard University Press, Cambridge, MA.
- *The Bees of the World.* C. D. Michener, 2000, Johns Hopkins University Press, Baltimore.
- *Manual of Nearctic Diptera,* Vols. 1 and 2. J. F. McAlpine, ed., 1981 and 1987, Agriculture Canada Monographs 27 and 28, Ottawa, Ontario.
- *The Lepidoptera: Form, Function and Diversity.* Malcolm J. Scoble, 1995, Oxford University Press.
- *Illustrated Encyclopedia of the Butterfly World.* P. Smart, 1996, Salamander Books, London.

TELEVISION AND VIDEO

A number of excellent television programs, as well as regular series, have featured insects as their subject matter. Some of these are available on videotape or DVD. In general, video narration is not peer-reviewed and is notorious for inaccuracies. Nonetheless, the value of video in education is difficult to overestimate, and sometimes a game of "spot the mistakes" can be a learning experience in itself, especially when watching "B movies."

In terms of television series, *Insectia* is a 13-part, half-hour series hosted by Georges Brossard, the founder of the Montreal Insectarium, and it features many exotic locations and extralarge insects. *Alien Empire* is a 3-part series of 1-hour programs produced by the Public Broadcasting System in the United States. It is especially memorable for its use of miniature cameras and close-focusing wide-angle optics that are able to literally track alongside running and flying insects. *Acorn the Nature Nut* is a 91-part, half-hour series, with 26 episodes having entomological themes. It is hosted by Canadian entomologist John Acorn (coauthor of this article) and combines humor, music, and reliable entomological information.

A recent, well-received film on insects is *Microcosmos,* by Claude Nuridsany and Marie Pérennou. It features superb cinematography of the insects of southern France, and has a companion book by the same name (1997, Stewart, Tabori & Chang, New York).

SUPPLIERS

The following suppliers are keenly aware of the needs of entomology teaching. Although they deal with a mix of research customers, hobbyists, and teachers, their catalogs provide a good overview of the sorts of materials available for sale in an entomological context. Field equipment can enhance field-trip experience, whereas models, charts, and properly designed lab apparatus can greatly improve the quality of teaching in the laboratory. This list is incomplete, but the following suppliers have a long history and good reputation in the field.

- Bio Quip Products, 17803 La Salle Avenue, Gardena, CA 90248-3602. Phone: (310) 324-0620. Fax: (310) 324-7931. E-mail: bioquip@aol.com. Web site: http://www.bioquip.com. Bio Quip has been the main supplier of entomological equipment and books in North America for over 55 years. Their selection of educational items is very good.

• Watkins & Doncaster., P.O. Box 5, Cranbrook, Kent TN18 5Ez, England. Phone: +4401580 753133. Fax: +4401580 754054. E-mail: sales@watdon.com. Web site: http://www.watdon.com. The British equivalent of Bio Quip, with similar items in the catalog and a history dating back to 1874.

• Acorn Naturalists, 17821 East 17th Street, No. 103, Tustin, CA 92780-2138. Phone: (800) 422-8886 (Canada and U.S.A.), (714) 838-4888 (international). Fax: (800) 452-2802 (Canada and U.S.A.), (714) 838-5869 (international). E-mail: EMailAcorn@aol.com. Web site: http://www.acornnaturalists.com. Good selection of educational materials, especially for grade school teaching.

• Classey Books, Oxford House, Marlborough Street, Faringdon, Oxfordshire SN7 7JP, England. Phone: +44(0)1367 244700. Fax: +44(0)1367 244800. E-mail: info@classeybooks.com. Web site: http://www.classeybooks.com. Since 1949, E. W. Classey Ltd. has carried an excellent selection of entomology books, both new and used.

• Carolina Biological Supply, 2700 York Road, Burlington, NC 27215-3398. Phone: (800) 334-5551 (U.S.A. only), (800) 387-2474 (Canada only), (336) 584-0381 (international), Fax: (800) 222-7112 (U.S.A. only), (800) 374-6714 (Canada only). E-mail: carolina@carolina.com. Web site: http://www.carolina.com. Long-established general source for science and technology teaching materials.

• Ward's Natural Science Establishment, 5100 West Henrietta Road, P.O. Box 92912, Rochester, NY 14692-9012. Canadian address: 397 Vansickle Road, St. Catherine's, Ontario L2S 3T5. Phone: (800) 635-8439 (U.S.A. only), (800) 387-7822 (Canada only), (716) 359-2502 (international). Fax: (800) 635-8439 (U.S.A. only), (905) 984-5952 (Canada only), (716) 334-6174 (international). Web site: http://www.wardsci.com. Another long-standing source for general science teaching materials.

See Also the Following Articles

Collection and Preservation • Entomological Societies • Insect Zoos • Movies, Insects in • Museums and Display Collections • Photography

Temperature, Effects on Development and Growth

František Sehnal, Oldřich Nedvěd, and Vladimír Košťál

Institute of Entomology, Academy of Sciences, Czech Republic

The body temperature of insects, as in other ectothermic organisms, is linked to changes in the ambient temperature. Temperature fluctuations are small in environments such as the tropical rainforest, caves, and some aquatic habitats. In most habitats, however, the seasonal and diurnal temperature oscillations are considerable. For example, insect body temperature can change abruptly by 10°C or more when exposure to direct sunlight is followed by the shade of a cloud. The way of life with fluctuating body temperatures is called heterothermy; this is in contrast to the homeothermy of endothermic organisms, such as birds and mammals, which regulate their body temperature, partly by generating endogenous heat.

DEVELOPMENTAL PARAMETERS

Every organism is adapted to a set temperature range. A general preference for high temperatures is referred to as thermophily, and the inclination to low temperatures is psychrophily. Insects living in warm climates or parasitizing warm-blooded animals are thermophilic, whereas those dwelling in soil are usually psychrophilic. Temperature preferences may change during development; for example, aquatic insects inhabit cold mountain streams during their immature stages but fly in warm air as adults. Temperature fluctuations within the species-specific physiological range determine the rate of development and often exert other physiological effects.

Development and reproduction occur at physiological temperatures that are delimited by an upper and a lower developmental threshold (UDT and LDT, respectively). Within this range, there is an optimal temperature for rapid development. The dependence on temperature can be expressed as a metabolic rate or as a developmental rate. The metabolic rate (MR) reflects the velocity of the energy-supplying biochemical processes and can be measured as oxygen consumption, carbon dioxide production, or heat generation. Many enzymatic reactions and the total body metabolism increase exponentially over a broader temperature range than is the span of physiological temperatures. Metabolic increase is usually two- to threefold with temperature elevation by 10°C and can be expressed by $MR = e^{a+kT}$, where a and k are constants and T is temperature.

Metabolic rate determines the developmental rate (DR), which is a reciprocal value of the developmental time (DT), DR = 1/DT. Measuring developmental times, such as duration of larval development, length of the reproductive period, or expanse of the entire life cycle, requires maintenance of defined conditions (notably temperature and nutrition). This is difficult to do for long periods of time and this is why DT and DR values are usually established for individual developmental stages and then recalculated for the entire life cycle. Circadian rhythmicity of some processes, for example, the synchrony of hatching or adult emergence at a certain time of day, complicates DT assessments.

In contrast to the exponential rise of the metabolic rate, the increase in DR within the physiological temperature range is linear. When a series of temperature and corresponding DR values is plotted in a graph, a straight line so obtained crosses

the *x* axis at the theoretical LDT point. Close to this point, the relationship between developmental rate and temperature ceases to be linear, and the straight line is bent into a sigmoidal curve. Actual LDT is therefore somewhat lower then predicted. At the upper temperature range, the DR slows down before it reaches a maximum at the optimal temperature. After the maximum, DR sharply drops and at UDT the development is discontinued.

Developmental time depends on the effective temperature, i.e., temperature value above LDT (actual temperature *T* minus LDT). The constant product of effective temperature and developmental time is called the sum of effective temperatures (SET) and represents the heat required for the completion of a particular developmental stage. SET is conveniently expressed as the number of degree days. For example, a SET value of 100 degree days means that development at 5°C above LDT lasts 20 days and at 10°C above LDT 10 days. When the insects develop at fluctuating temperatures, the average temperature above LDT and the length of time when the temperature surpasses LDT are considered for each day, and the number of degree days established in this way is summed. Developmental stage is completed when the summation reaches the SET value.

A temporal drop in temperature below LDT is associated with developmental block and is counted as 0. Usually, there is no "negative developmental rate" at temperatures below LDT, i.e., no delay of development is observed after transfer to an effective temperature. Natural temperature fluctuations, however, may have a signaling effect and influence the SET value in some species, and this possibility must be checked experimentally.

The LDT and SET values are species-specific population characteristics. The LDT values are similar for all developmental stages of a given species, even when they develop in diverse seasons and experience disparate temperature fluctuations. The stability of LDT is manifested as developmental thermal isometry, i.e., the percentage of time spent in a particular stage at any constant physiological temperature is a stable fraction of the entire developmental time. The LDT and SET values established in the laboratory enable prediction of the course of development in the field. Control of many insect pests in agriculture and forestry largely relies on such predictions. For example, on the basis of the LDT and SET data for the codling moth *(Cydia pomonella)* we can predict, on the basis of daily temperature measurements in an orchard, the time of the first egg deposition and time insecticide sprays accordingly.

Both LDT and, especially, SET may vary between geographical populations because of adjustments to local climatic conditions. Insects that have spread to temperate zones from the tropical regions often maintain a high LDT and can reproduce and develop only in the hot season, spending most of the year in a state of dormancy. The survival in cold is made possible by increased cold hardiness, a parameter that seems to be more plastic than LDT.

DEVELOPMENTAL ARRESTS

Most insects must overcome long periods of adverse conditions when food is wanting and temperature remains outside the physiological limits. Insects cease development and reproduction but, if the temperature does not reach lethal extremes, they remain capable of resuming these processes as soon as the conditions become favorable. The state of easily reversible, directly temperature-dependent developmental arrest is known as quiescence. It is typical of insects adapted to relatively short periods of unfavorable, nonlethal circumstances, and it is usually associated with temperature acclimation.

Insects that will be exposed to severely hostile conditions that can last for many months enter a programmed developmental arrest, diapause. Diapause occurs in anticipation of a season in which the insect could not survive in the active state. It is induced by environmental signals acting before the adverse conditions set it, sometimes on a much earlier developmental stage and exceptionally on the parental generation.

Seasonal changes in the environment are specific for each latitude, altitude, and habitat, but always correlate with changes in the photoperiod, i.e., the length of day versus the length of night. Photoperiodic changes therefore provide ideal signal for the advent of unsuitable conditions. However, temperature can shift the diapause-inducing photoperiod response over a broad range. For example, 50% of caterpillars of *Acronycta rumicis* are induced to enter pupal diapause at a daylength of 19 h at 15°C, but at 16 h at 25°C. Low temperature normally enhances the effect of short photoperiod and high temperature enhances the effect of long photoperiod. Daily fluctuations of temperature, the thermoperiod, can induce diapause in a few species kept in constant darkness. On the other hand, high temperature can abolish the diapause-inducing effect of a short photoperiod. For most insects, the diapause is facultative and there is a variation in the critical photoperiod/thermoperiod at which each individual enters diapause.

Once induced, diapause is not terminated immediately after the diapause-inducing conditions disappear. A certain time must elapse, during which neurohormonal regulations return to the pattern supporting development and reproduction. The mechanisms controlling diapause termination are not known. The length of time in diapause depends on its "depth" and on environmental conditions, especially temperature, to which the diapausing insects are exposed. In the overwintering insects, diapause is often shortest at temperatures around 5°C (Fig. 1) and the photoperiod is irrelevant. Due to exposure to low temperatures in late fall, the overwintering insects terminate diapause in early winter, and the resumption of their development or reproduction is then halted only by a direct effect of low temperature: diapause turns into quiescence.

ALTERNATIVE DEVELOPMENTAL PATHWAYS

In certain insects, temperatures within the physiological range affect the course of non-diapause development. For example,

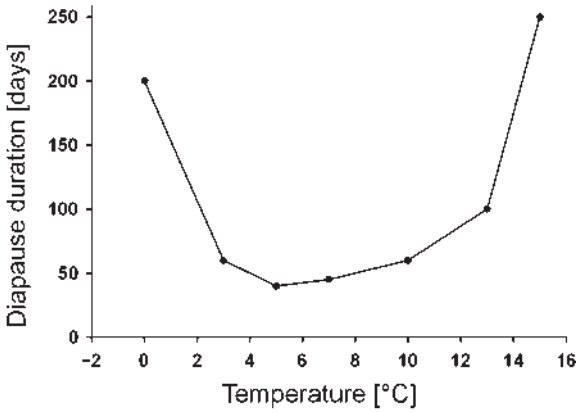

FIGURE 1 The influence of constant temperature on the termination of imaginal diapause in the chrysomelid beetle *Atrachya menetriesi.* The diapause is completed (50% of individuals can resume development) most rapidly at 5°C. Based on data of Ando (1983).

larvae of the yellow mealworm, *Tenebrio molitor,* develop in 11–15 instars at 25°C and in 15–23 instars at 30°C. Less dramatic changes in the number of molts and the growth rate were noted in a number of insects. Abnormal temperatures can also dissociate the onset of metamorphosis from the body size at which it normally occurs. In the wax moth, *Galleria mellonella,* placing newly ecdysed larvae of the last or penultimate instar on melting ice induces an additional larval molt. To cause such anomalies, the temperature must alter the secretion of hormones that control specific developmental events.

Some insect species occur in more than one form and their alternation depends on temperature. For instance, development of young caterpillars of *Colias eurytheme* at 18°C leads to mainly yellow, and development at 27–32°C to orange, butterflies. The spring and summer forms of some other butterflies are well known examples of seasonal dimorphism in which one form is linked to diapause.

LETHALITY AT EXTREME TEMPERATURES

A general response of insects to temperatures just below their LDT or above their UDT is the cessation of development and reproduction while the insects remain active and feed. The larvae may slowly grow and the adults accumulate reserves and to some extent undergo gonadal maturation. These processes are terminated at more extreme temperatures when the insects begin to die.

During cooling, the metabolic rate and motility gradually decrease. At a certain temperature, the neural and muscular activities are impaired and the insect lapses into cold stupor. The metabolic rate of such immobile insects continues to decline with decreasing temperature. The stupor point is as high as 12°C in some tropical insects and the honey bees, around 5°C in many temperate species, near 0°C in most overwintering insects, and below the freezing point in species living in very cold areas.

The nature of chill injuries is little understood. Desiccation and nutrient depletion during the cold-induced starvation are certainly incompatible with long-term survival, but death usually occurs earlier and is probably the result of damaging effects at cellular level. The loss of cell membrane fluidity, imperfect protein functions (enzymatic activities, transport, signaling, etc.), and the resulting asynchrony of the life-supporting processes cause metabolic disorders. For example, the ion pumps in the cell membrane become inefficient and sodium concentration in cytoplasm increases, while the potassium ions flow out into the hemolymph.

The upper temperature extremes are also lethal. Gradual warming past UDT, which is for many species around 35°C but is never sharply delimited, increases the metabolic rate, loss of water, and motility. At a certain temperatures, usually around 40°C, the water loss, and thereby the evaporative body cooling, increases sharply. The spiracles are wide open and the melting of cuticular lipids permits evaporation through the body surface. After some time at such a high temperature, the losses of water and nutrients lead to exhaustion, manifested as a rapid decrease of motility and a drop of transpiration. If this state is brief, it can be reversed. The temperature at which it occurs is the upper lethal threshold. A gradual temperature increase to this threshold may cause heat stupor.

Survival at temperatures above the threshold is a function of temperature and length of exposure. Warming to the absolute upper lethal temperature, which is usually around 50–55°C, causes irreversible tissue damage, and even a short exposure is lethal.

ACCLIMATION AND TEMPERATURE TOLERANCE

The survival at extreme temperatures is improved after an acclimation. Shortening of the photoperiod in late autumn and early winter usually acts synergistically with descending temperature in triggering a seasonal cold acclimation. An exposure to low temperature alone is often insufficient for full cold acclimation and successful winter survival because some physiological adjustments (for instance down-regulation of the ice nucleators and enhancement of cryoprotectant biosynthesis) require a preceding switch to the diapause developmental mode and this is controlled by the photoperiod. In other insects, cold acclimation is associated with temperature-dependent quiescence and the photoperiod is irrelevant.

Cold acclimation is a complex adjustment involving profound changes at the organismic and tissue levels. Accumulation of low-molecular-weight cryoprotective polyols and, in some insects, also synthesis of antifreeze proteins are characteristic features of cold acclimation. Other physiological changes include changes in cuticular lipids, increased fluidity of phospholipids in the cell membranes, conformation changes of some proteins, possibly production of alternative enzymes with activity optima at lower temperatures, and synthesis of heat-shock proteins. Acclimation may also encompass changes in morphology (e.g., cold-acclimated lacewings turn from

green to reddish brown) and behavior (e.g., formation of cocoons with higher resistance to desiccation and ice penetration, voiding the gut to get rid of ice nucleators, and seeking dry places to prevent ice inoculation from the surroundings).

The temperature at which insects freeze is relatively low even without cold acclimation. Insects contain only low amounts of nucleators needed for the ice crystal formation and are therefore capable of considerable supercooling. The nonacclimated individuals freeze at –10 to –15°C, while those with accumulated cryoprotectants and antifreeze proteins can be supercooled to temperatures below –20°C. All cryoprotective compounds disappear from the organism after a certain time. The length of this deacclimation process also depends on temperature, but details have not been examined.

Changes associated with heat acclimation are little known. Synthesis of heat-shock proteins, which were discovered in *Drosophila* exposed to elevated temperature, is a very general response not only to heat but also to cold and various other types of stress. Several types of heat-shock proteins are known from organisms ranging from bacteria to plants. It is believed that they are chaperones enabling or protecting functional protein conformations. Temperature or another stressing factor acts as a signal inducing their synthesis at transcriptional level. Accumulation in the cells of "unfolded" proteins is believed to be the common intracellular message triggering this transcription.

See Also the Following Articles

Aestivation • Cold/Heat Protection • Diapause • Fat Body • Greenhouse Gases, Global Warming, and Insects • Growth, Individual • Homeostasis • Thermoregulation

Further Reading

Ando, Y. (1983). Diapause and geographic variation in a leaf beetle. *In* "Diapause and Life Cycle Strategies in Insects" (V. K. Brown and I. Hodek, eds.), Vol. 23 of Series Entomologica, pp. 127–141. Junk, The Hague.

Beck, D. S. (1983). Insect thermoperiodism. *Annu. Rev. Entomol.* **28,** 91–108.

Browse, J., and Xin, Z. (2001). Temperature sensing and cold acclimation. *Curr. Opin. Plant Biol.* **4,** 241–246.

Danilevsky, A. S. (1965). "Photoperiodism and Seasonal Development of Insects." Oliver & Boyd, London.

Danks, H. V. (1987). Insect dormancy: An ecological perspective. *Biol. Surv. Can. Monogr. Ser. 1,* Ottawa.

Lee, R. E., Jr., and Denlinger, D. L. (1991). "Insects at Low Temperature." Chapman & Hall, New York/London.

Leather, S. R., Walters, K. F. A., and Bale, J. S. (1993). "The Ecology of Insect Overwintering." Cambridge University Press, Cambridge, UK.

Lindquist, S., and Craig, E. A. (1988). The heat shock proteins. *Annu. Rev. Genet.* **22,** 631–677.

Lozina-Lozinskii, L. K. (1974). "Studies in Cryobiology." Wiley, New York.

Morimoto, R. I., Tissieres, A., and Georgopoulos, C. (eds.) (1994). "The Biology of Heat Shock Proteins and Molecular Chaperones." Cold Spring Harbor Laboratory Press, Cold Sprimg Harbor, NY.

Smith, A. U. (1961). "Biological Effects of Freezing and Supercooling." Arnold, London.

Vannier, G. (1994). The thermobiological limits of some freezing-intolerant insects: The supercooling and the thermostupor points. *Acta Oecol.* **15,** 31–42.

Termite

see *Isoptera*

Terrestrial Insects

see *Soil Habitats*

Thermoregulation

Bernd Heinrich
University of Vermont

In insects, as in other animals, body temperature strongly affects the rate of energy expenditure, the rate at which food can be located and harvested, growth, the facility with which mates can be acquired and predators avoided, and sometimes also the susceptibility to disease organisms. Thermoregulation refers to the ability to regulate that body temperature which best serves survival and reproduction, and it encompasses numerous conflicting constraints and selective pressures. In insects, major considerations involve body mass and access to either external or internal heat. Thermoregulation operates through behavior, physiology, and morphology. For the most part, insects are too small to be able to appreciably elevate, or regulate, their body temperature by internal heat production, although some are large enough and that, coupled with their high flight metabolism, could easily cause them to overheat. In numerous insects, elaborate mechanisms of thermoregulation have evolved both for heating and for cooling the body that possibly rival those of the typically endothermic vertebrates.

ENDOTHERMY IN FLYING INSECTS

Insects arose on earth at least 350 mya in the Devonian Period of the Paleozoic Era. Little is known about the earliest forms, except that originally they must have been crawlers, not flyers, and their bodies assumed approximately the temperature of the immediate surroundings to which they adapted. This holds true even when the immediate surroundings are quite frigid. The adult form of a flightless midge (*Diamesa* sp.) walks on glacier ice even when its body temperature is chilled to –16°C. It is so sensitive to heat that, when taken from its natural environment and held in one's hand, it is killed by the warmth of one's skin. However, there are insects that maintain quite specific and high body temperatures. Some species of

sphinx moths, for example, have thick insulating fur and normally maintain a thoracic temperature near 46°C during flight over a wide range of ambient temperatures. To these moths, our own normal body temperature of 37°C is almost cool. An insect's head and abdominal temperatures are for the most part unregulated.

In the same way that the motor heats up when a car burns fuel, heat is released as an inevitable by-product of cellular metabolism whenever muscle contracts. Close to 94% of the energy expended by muscles during contraction is degraded to heat, while approximately 6% appears as mechanical force on the wings. Insect flight is one of the most energetically demanding activities known, and thus most insects produce more heat per unit muscle mass when they fly than almost any organism on earth. Most insects exist under conditions somewhere in between the cold-blooded crawler and the hot-blooded flyer, but these extremes show us what is possible, and they thus offer us a remarkable window into thermal adaptation from an evolutionary perspective.

The ability of birds and mammals to regulate body temperature at one set point, specifically 37–41°C, has long been considered proof of sophistication and phylogenetic advancement relative to animals whose body temperature varies with that of their environment. Deviation of body temperature from the set point of 37–41°C is, in birds and mammals, often associated with illness and was once thought to be caused by a failure of the thermoregulatory system. We now know that both increases and decreases in body temperature can be and often are adaptive responses. Both responses are often sophisticated physiological mechanisms that involve more thermoregulation rather than less, albeit the body is kept at a more appropriate temperature for specific conditions.

Small insects have much lower body temperatures in flight than large insects, not because they produce less heat—in fact, they may have *higher* rates of heat production than larger insects, but because they have much greater conductance because of their large relative surface area. In bees, for example, only the large species heat up in flight and generate an appreciable elevation of body temperature even though metabolic cost of flight per unit weight declines approximately 230% for a 10-fold increase in mass. A mosquito in flight maintains only a tiny (<1°C) gradient between thoracic and ambient temperature, despite prodigious amounts of heat production. A blow fly *(Calliphora vicia)* may heat up 5°C, and a honey bee heats up its thorax about 15°C. Having a much larger thorax, and hence a smaller relative surface area, means that the internally generated heat during flight is not lost by convection at the same rate that it is produced until a much higher temperature gradient has been generated.

Large insects—those that inevitably generate a high body temperature during continuous flight—must be biochemically adapted to operate their flight muscles at the high temperatures experienced. Temperature is important for mechanical efficiency; at low muscle temperature there is partial overlap in contractions of the up- and downstroke muscles, the dorsoventral and dorsal longitudinal muscles; the two sets of muscles then work against each other rather than working to move the wings.

Numerous other moths—such as most microlepidoptera and geometrids (inchworms) and some ctenuchids and arctiids—are small or weak flyers that do not heat up but fly at low air temperature. They fly at muscle temperatures much lower even than those at which the large-bodied, small-winged (and hot-blooded) sphinx moths generate zero power. Evolution has acted strongly to tailor the flight motor's capacity for maximum power output for much lower ranges of operating temperatures. For example, the geometrid *Operophtera bruceata* can gain sufficient power to fly at 0°C. (Nevertheless, its capacity to do so is only partially the result of muscle physiology.)

The basis for the evolution of differences between species arising from a common ancestor is variation among individuals. Variation was present in the past, and for many traits, variation is maintained even now. For example, in *Colias* (sulphur) butterflies, the gene locus for phosphoglucose isomerase, one of the enzymes involved in energy metabolism in these butterflies, changes in allele frequency with season and habitat temperature. This suggests that natural selection is occurring even over very short (that is, seasonal) time spans. The different enzyme alleles have different thermal stabilities, and heterozygotes are thought to have an advantage in an environment of rapidly fluctuating temperatures inasmuch as the individuals heterozygotic for this locus fly over a range of temperatures broader than that of individuals of other genotypes.

When a seasonally changing temperature environment, which is the rule, can select for heterozygosity, then one might expect that an environment of constantly high or low temperature, which is the exception, should lead to the fixation of an appropriate genotype. Hence, selection in terms of gene-frequency changes would not normally be present for our inspection in more constant environments, in which appropriate genotypes would already have been selected long ago to adapt to the average temperature. The specific thoracic temperature that is maintained by regulation is "chosen" by evolution probably because it is the temperature most readily regulated for maximum activity over a range of prevailing environmental conditions.

WARM-UP BY SHIVERING

During his classic studies of honey bee communication, Karl von Frisch noted that bees often interrupted flight for a few minutes when they were returning to the hive heavily laden with nectar. He presumed they stopped "to rest," but we now know they were stopping to work: to raise their thoracic temperature. They most likely stopped flight because it was a cold day and they had cooled convectively. Bees are able to raise their thoracic temperature by shivering, which can work their flight muscles harder than flight itself does. Von Frisch could not have known any of this, because shivering and thermoregulation by individual insects was unknown in the

1960s, nor is shivering externally visible in bees even if one looks very closely.

Like the maintenance of an elevated body temperature by internal heat production in flight, physiological warm-up is found in all large, active flyers among the dragonflies (Odonata), moths and butterflies (Lepidoptera), katydids (Orthoptera), cicadas (Clypeorrhyncha or Homoptera), flies (Diptera), beetles (Coleoptera), and wasps and bees (Hymenoptera). That is, it is found from some of the earliest forms, the Odonata, to the most evolutionarily highly derived, the Diptera, Coleoptera, and Hymenoptera. It is not found in the small and therefore nonendothermic members of the same groups. Because no insects shiver except those that then also heat up from flight metabolism, it seems reasonable to conclude that the evolution of shivering behaviors is related to the evolution of flight but is unrelated to the insect's place on the phylogenic tree.

During preflight warm-up, there are synchronous contractions of groups of muscles that normally contract alternately in flight. That is, the main wing-depressor muscles, the dorsal longitudinal muscles, are excited simultaneously—in other words, in synchrony—with the dorsoventral wing elevator muscles. The neural activation pattern of thoracic flight muscles needs to be and is already very labile for flight control, and to add shivering when flight behavior has already evolved is probably a very minor evolutionary step. Physiological warm-up in its most basic form is like the idling of an engine; the engine "evolved" to propel the car, not to warm it up. Once present, the heat-producing flight muscle system required only a slight modification of neuronal activation patterns and, in the more sophisticated models, also the addition of the biological equivalent of a clutch—a mechanism to disengage the wings in the same way that an automotive clutch disengages the car's wheels. Some insects, such as dragonflies and moths, do have visible external wing vibrations that were originally called "wing whirring," these were once thought to pump air into the animal before the true function was elucidated.

The zenith of the shivering response of any hot-blooded animal (vertebrate as well as invertebrate) belongs to some bees. Honey bees and bumble bees have a physiological sophistication either not existing or not yet observed in other insects, and they exploit shivering behavior to an unprecedented extent and in a variety of ways. Like flies and beetles, bees are "myogenic" flyers in which the wing-beat cycle runs in part on automatic; as the downstroke muscles contract they stretch the upstroke muscles. This stretching *by itself* causes the upstroke muscles to contract. The downstroke of the wing therefore automatically causes the upstroke muscles to contract and vice versa, in a repeating cycle that is sparked by neural commands that are at a much lower frequency than the wing beats and are no longer specific to a single wing beat. (This system permits some of the smallest insects, such as midges, to achieve the unprecedented coordination required for wing-stroke cycles of over 1000 beats per second.) But the stretching of opposing muscle groups that maintains the myogenic contraction cycle can occur only if the wings are actually beating—namely, during flight. When the wings are not in use, as when they are folded back dorsally and the clutch-like wing hinge is engaged, then the muscle groups are in near tetanus. That is the reason for the bees' shivering.

Very vigorous shivering in bees is physically dampened even more by yet another mechanism. One of the two sets of opposite-acting muscles is activated (and hence contracted) slightly more than the other. Because the opposing muscles act like weights forcing down each side of a seesaw, the added force on one set of muscles prevents the "seesaw" from working (and the wings from "vibrating" back and forth).

WARM-UP BY BASKING

Like warm-up by shivering, warm-up by basking occurs in all major orders of insects that have fast flyers large enough to heat up from their flight metabolism. In its simplest form, behavioral warm-up is merely heat-seeking. A basking insect usually takes specific postures that simultaneously maximize solar input and minimize convective heat loss. Heat input is maximized by exposing the maximum surface area to the sun, while convective heat loss is minimized by using body parts (such as the spread wings) as baffles to retard air movement around the body. Orienting the body parallel to the air stream (as a wingless insect might do) would reduce the effect of convective cooling, but orienting the body perpendicular to the sun's rays to facilitate heating should take precedence, because no heat loss can be minimized until heat is first gained. Grasshoppers, beetles, and flies use these basking methods. For some dragonflies and butterflies the wings are especially important during warm-up in their role of reducing convective cooling.

Behaviorally distinct types of basking have been described in butterflies, although some (tropical) butterflies do not bask at all. In one type, called "lateral basking," the butterfly closes its wings dorsally and then tilts to present either the right or the left wing and body surface to the sun. The lower portions of the wings wrap around the body and touch it, and warming the lower wing portions in sunshine then causes heat to be conducted directly through them and into the body.

Many species of small-bodied butterflies, primarily pierids and some lycaenids (which commonly fly in breezy mountain meadows) bask by opening their wings partially in a V so that the body is directly available to the sun's rays at the bottom of the V. The wings then serve as convection baffles to reduce cooling in the breezy environment. "Dorsal baskers" hug a solid substrate, such as the ground, and pull their wings down around them. They thereby expose the dorsal body surface to the sun while simultaneously capturing heat from the sun-heated substrate.

HEAT LOSS MECHANISMS

Insects near the size of a honey bee (approximately 200 mg) or larger may experience body temperatures during forced flight exercise that are potentially lethal to them. Alternately,

metabolic heating may inhibit continuous flight at relatively modest ambient thermal conditions (of air temperature and solar radiation), unless one or more of the following mechanisms for heat loss are activated.

Harnessing Convection

The rate of convective heat loss from a body is determined by the conductance of the body, i.e., its intrinsic rate of heat loss (which is a function of body size, shape, and insulation). Conductance, in turn, is a function of the wind speed (a hot body cools more quickly in wind than in still air—meteorologists call this the "wind-chill factor"). However, no convective heat loss is possible, regardless of conductance, if body and ambient temperature are equal, and at any one conductance and wind speed the amount of heat loss is directly proportional to the temperature difference between the body and the ambient surroundings.

Small insects that are endothermic in flight are sufficiently air-cooled such that they almost never reach the potentially dangerous high-temperature ceiling of near 45°C that is common to most animal tissues at normal atmospheric pressures. These small insects thus have no need of a specialized cooling system: they lose sufficient heat passively. Theoretically, larger insects could cool themselves by increasing flight speed and thus increasing convective heat loss, but flying faster would generally increase metabolic heat production, which would cancel out the increased heat loss, unless the internally generated heat is redistributed.

Heat Radiators

A radiator is a device that increases the surface area of a body or object so that more heat can be transmitted to the surrounding environment by convection. In some radiators, a fluid with a high heat capacity (like water, blood, or other liquid) circulates by means of a pump and transfers heat from its source to the radiator site for the heat loss. That is how a car engine is cooled.

Radiators are utilized by many large insects from the very diverse orders Lepidoptera, Odonata, Dipera, and Hymenoptera. The animals have a fluid-transfer cooling mechanism that dissipates heat through an abdominal radiator, while small members of the same groups that are not strong or continuous flyers lack the heat-transfer response. When we humans exercise in the heat, blood is pumped to the skin or extremities to facilitate heat loss, but this is done at the expense of pumping blood and oxygen to the muscles instead. Therefore, work capacity is compromised. Insects, on the other hand, do not need to compromise aerobic work capacity at higher air temperatures because of thermoregulation. In insects, the total separation of respiratory and heat-transfer functions makes it possible for them to continue working, even when the fluid flow is interrupted.

The "radiator tube" that conducts the hemolymph to the abdominal heat radiator in insects also serves as a pump, which operates by peristaltic contractions along its entire length. Although sphinx moths in whom surgery has rendered this "heart" inoperative (by tying it shut) can still fly until reaching near-lethal thoracic temperatures, removal of their insulating layer of thoracic scales makes continuous flight again possible.

Evaporative Cooling

One of the extraordinary examples of an evaporative cooling mechanism specifically for thermoregulation is that found in the workers of honey bees, *Apis mellifera,* and yellowjackets, *Vespula* spp. These insects use the head as a radiator, but they do so with a difference. Nectar-gathering honey bees normally fly with flight-motor temperatures near 15°C above air temperature. They are capable of the astounding feat of flying even at ambient temperatures near 45°C while maintaining the thorax at the same or only slightly lower temperature. They do so by regurgitating nectar from the honeycrop, and while the nectar is held on the mouthparts and the head, water evaporates from it. Because of the physical contact between the head and the thorax, thermoregulation of one effectively results in thermoregulation of the other. Thus, the head is cooled by evaporation of water until there is a large temperature difference between the head and the metabolically heated thorax, at which point heat from the thorax follows the temperature gradient and is transmitted to the head. Head temperature is actively regulated, with thoracic temperature passively following, because artificial heating of the thorax alone does not result in the heat-dissipation response so long as head temperature remains low. However, artificial heating of the head (as with a narrow beam of light from a heat lamp) almost immediately results in nectar regurgitation and evaporative cooling, even while thoracic temperature is still (momentarily) low.

Some insects cool evaporatively from the back. In the hot Australian deserts, the larvae of the sawfly *Perga dorsalis,* in response to solar heat stress, first raise their abdomen to the sun to shade the body and to increase convective heat loss. In an emergency, when this response is insufficient, they also emit rectal fluid and spread it over their ventral surface to cool themselves evaporatively.

Diceroprocta apache of the Sonoran desert of the southwestern United States employs a third evaporative cooling mechanism, this one analogous to sweating. These cicadas are plant-sap feeders, and despite living in a dry environment, they have access to a large fluid supply by inserting their sucking mouthparts into the xylem of deep-rooted shrubs, such as mesquite. They thus indirectly tap water from deep underground stores. Cicadas sing when ambient temperatures in the shade reach 40°C, and the repetitive contractions of their tymbal muscles result in internal heat production that adds to the already considerable external heat load.

Body temperature during this exercise in the heat is reduced to tolerable levels by evaporative cooling from fluid shed

through large pores distributed over their dorsal body surfaces. The release of this fluid, and the consequent evaporative cooling, occurs only in response to very high body temperature. Most insects, especially those of desert environments, are instead highly resistant to water loss when alive, and upon death, there results an immediate *increase* in water loss as the spiracles are no longer actively maintained shut. Killing of the cicada, in contrast, immediately stops the sweating response, therefore showing that it is under metabolic control. The cooling response is mediated, ironically enough, by aspirin-like substances produced in their bodies in response to heat stress.

MORPHOLOGY AND THERMOREGULATION

Aside from physiology, various aspects of insects' morphology come into play in their thermoregulating responses.

Insulation

Many hot-blooded insects that regulate their body temperature have, like their endothermic vertebrate counterparts, bodies wholly or at least partially covered with insulation. One type of insulation is derived from air sacs. Insects already have air sacs used for breathing, and still air is, next to a vacuum, the best possible insulator. Many insects of various orders have air sacs between the thorax and the abdomen that greatly retard the leakage of heat into the abdomen. But large-bodied dragonflies have gone one step further: their air sacs surround the thoracic flight motor. The other two types of insulation are derived from exterior cuticular structures.

Lepidopterans are covered with a layer of thin overlapping scales, which are especially noticeable in coloring the wings. Rather than remaining flat and colorful for visual signaling as in butterflies, they have become long and thin to form a thick insulating body (thoracic) pile or fur coat in many moths. This coating of pile is so effective as insulation it more than halves the rate of heat loss, or doubles the temperature excess, hence permitting flight at much lower air temperatures. Endothermic insects with pile now fly in many northern areas and at times of the year at which they would otherwise be excluded. Conversely, insects from tropical environments have no or only sparse pile covering.

A covering of setae, small hair-like projections from the cuticle, is the third source of insect insulation. Setae have various functions and numerous independent evolutionary origins. Within the Hymenoptera, only the northernmost large bees, the bumble bees, have a heavily insulated flight motor. However, even honey bees have a layer of short insulating pile on the thorax, which aids them on cool mornings and at high elevations. Bees inhabiting the tropics and hot deserts do not have a covering of pile dense enough to provide appreciable insulation. But even tropical bees cannot get along totally without setae, because they use these projections to trap pollen from flowers. Some wasps, in contrast, live in the same northern areas that bumble bees do, but they do not rely on pollen for protein. Instead, many are predators on fast-flying insects, and they are glabrous. Perhaps an advantage of fuel economy in flight, or perhaps the necessity for fast flight, has assumed more importance than drag-inducing insulation for thermoregulatory control.

Color

In a few instances, an insect's color has a functional significance in thermal balance during basking. For example, lateral-basking sulphur butterflies (*Colias* spp., usually yellow or white) found in cool environments (such as mountaintops or high latitudes) or seasons (early spring) tend to have dark wing undersides. These darker individuals are able to heat up the thorax slightly faster than lighter congeners, which buys them additional flight time when basking is needed to prepare for flight.

Although color can have a slight thermal advantage, it is more often subservient to other needs, such as the need to evade predators. Not surprisingly, insects that inhabit open ground often match their background in color and thus are highly camouflaged.

Stilts and Parasols

A beetle walking on hot desert sand might experience temperatures that could kill it in a minute or less. Just a few millimeters above the ground, however, the hot, ground-hugging air layer is disrupted and mixed with cooler air from above. If we were shrunk to Lilliputian size and forced to live where a few millimeters' difference in elevation could mean the difference between life and death, we would find some way to lift ourselves above the searing heat. Numerous ground-dwelling beetles living on hot sands do just that. Tiger beetles (Cicindelidae) begin to stand tall when sand temperatures exceed 40°C. Aside from extending their jointed legs to stand taller, some beetles, like *Stenocara phalangium* (Tenebrionidae) from the Namib Desert of southern Africa, have evolved very long stilt-like legs that allow them to avoid overheating by both avoiding the heat at ground level and losing some of the solar heat by convection through fast running.

Another option is to use one body part to shade another. For example, some beetles reduce their absorption of external heat from direct solar radiation by having an air space beneath the elytra that insulates the abdomen from direct solar radiation.

Countercurrent and Alternating-Current Heat Exchanges

At low air temperatures when the abdomen is cool, the flight motor could cool precipitously if the hemolymph carried heat away from the thorax to be dissipated from the abdomen. Two mechanisms, however, normally prevent this potential problem of thoracic cooling. The first is a temporary reduction or elimination of the circulation: to retard heat loss.

Another mechanism, one more subtle than cardiac arrest, helps some insects prevent heat leakage from thorax to abdomen. Proof of that is seen in honey bee workers and Cuculliinae winter moths, which never show appreciable increases in abdominal temperature, even as the flight motor stays hot. An examination of their circulatory anatomy explains the mystery: they harness countercurrent heat transfer.

A countercurrent implies two separate currents flowing next to each other but in opposite directions, as through the petiole between thorax and abdomen in insects. If the fluid in one current is of a higher temperature than that of the other, then heat (which is not confined by the vessel walls) will passively flow "downhill," from high to low temperature, across these walls. Thus, if the hot blood leaving the thorax flows around the vessel in close proximity to cool blood entering it from the abdomen, as in most bees, then heat exchange is inevitable. At least some of the heat from the thorax will be recycled back into the thorax because the incoming blood is heated by the outgoing blood. In honey bees and winter moths, countercurrent heat exchange is greatly enhanced by prolonging the area for that potential heat exchange to occur, as the aorta in the petiole is lengthened and (in honey bees) convoluted into loops.

Bumble bees and northern vespine wasps have a very much different and seemingly less efficient countercurrent heat exchange circulatory anatomy than honey bees and winter moths (Fig. 1). This situation may seem counterintuitive because they live in cold climates, some species even inhabiting the High Arctic. They might thus be expected to have even better countercurrent heat exchangers than honey bees, which are of temperate and tropical origin. Instead of having more loops for countercurrent heat exchange, they have none! Nevertheless, their anatomy can also be understood in terms of thermal strategy, but as it relates to their social system.

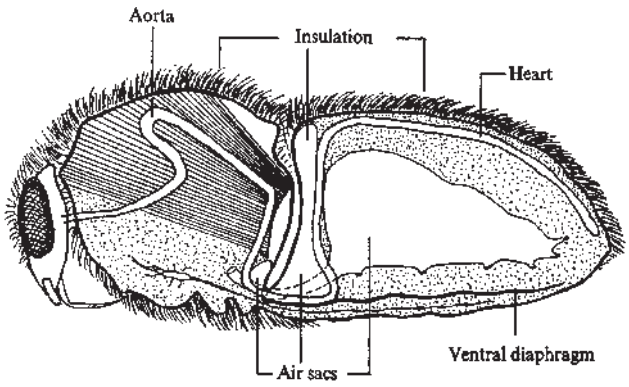

FIGURE 1 Anatomy of a bumble bee *(Bombus)* relevant to thermoregulation. The thorax is insulated with pile that reduces the rate of convective heat loss. The ventor of the abdomen is lightly insulated or uninsulated when the bee presses her abdomen onto brood to be heated. Hemolymph (blood) is pumped anteriorly by the heart, from the abdomen into the thorax. When dissipating heat from the working muscles in the thorax, the blood enters the aorta from the heart in pulses. Each pulse of cool blood from the abdominal heart into the thoracic aorta alternates with a pulse of warm blood entering the abdomen to the thermal window. In this way, countercurrent heat flow (into blood returning to the thorax) is minimized and heat flow (into the abdomen) is maximized. [Reproduced, by permission of Oxford University Press, from B. Heinrich, (1976), *J. Exp. Biol.* **64,** 561–585.]

Bumble bee and wasp queens start their colonies very early in the spring; each new queen attempts this task alone, as an individual. To this individual bee or wasp, time is of the essence, for she must complete the whole colony cycle within a single growing season. A queen's first priority, then, is to rear a group of helpers. Temperatures when and where she builds her nest may be near 0°C, and if the brood were left at that temperature it might take years for them to develop to adults—provided they could withstand the freezing temperature. Even in the High Arctic, however, the queens of *Bombus polaris* can produce a batch of workers in about 2 weeks, as can other bumble bees and *Vespula* wasps. Both bees and wasps accomplish these feats by incubating the brood from the egg to the pupal stage. The queens perch upon their brood clump—consisting of eggs, larvae, and/or pupae—and they press their abdomen upon the brood, much as a hen incubates her eggs with her belly. Only the abdomen provides a smooth surface for contact, but only the thorax produces heat by way of intense shivering by the flight muscles. No incubation, and hence social life, would be possible for these insects in a cold environment if, like honey bees, they were incapable of transferring heat from the source of its production into the abdomen that provides the smooth tight contact with the brood.

The bumble bee's aorta is long enough to permit moderate heat exchange and hence retention of heat in the thorax, but it is short and straight enough so that a physiological mechanism can be activated that shunts the fluid and heat through, effectively eliminating countercurrent heat exchange.

Countercurrent heat exchangers in vertebrate animals can be bypassed by rerouting the blood into an alternate (generally external) channel. That is why our own veins seem to pop out when we are active in the heat. Such rerouting of the blood from internal to external channels is not possible, however, in insects with open circulatory systems lacking veins and capillaries. Instead, in bumble bees there is a physiological solution for heat loss in the presence of a countercurrent heat exchange anatomy that serves the same purpose as an alternate blood channel. In the bumble bee, this consists of an *alternating-current* flow of blood. To shunt heat past the heat exchanger and into the abdomen, the bee lifts a small valve that allows a pulse of warm blood to enter the abdomen, and in the fraction of a second after the warm blood enters the abdomen, she then squirts a bolus of cool blood into the thorax. And so it goes back and forth, hot and cold pulses of hemolymph passing alternately through the heat exchange area in the bee's waist. The essential point is that although the blood is not rerouted into a different channel, it is instead temporarily "chopped" into alternating pulses in the same channels. This is the opposite of countercurrent flow because instead of recovering heat from the thorax, the system acts to remove it, in this case into the abdomen.

The pumping of hemolymph by the heart and the ventral diaphragm is also aided by in–out pumping movements of the

whole abdomen, which otherwise function only for moving gas in and out of the thorax; the in–out telescoping movements of the abdomen are synchronous with the heart beats and the ventral diaphragm beats, and they cause pressure changes that facilitate hemolymph flow in precise alternating currents.

THERMAL ARMS RACES

Against Predators

On a summer day in the Sahara Desert in Algeria, as the sun rises and begins to heat the sands that have been cold at night, an abundance of insect life is forced to retreat to cool underground refuge. Those unfortunate ones caught out in the heat become disoriented and, moving frantically, they heat up even more, then they die. The sun keeps rising, and sand temperatures begin to exceed 46°C. The desert lizards, *Accanthodactylus dumerili,* continue to hunt the incapacitated prey and any ants they can find. But they now dash quickly across the sand, and when they stop and stand, they alternately lift their feet to prevent burning them. Meanwhile, long-legged silver ants, *Cataglyphis bombycina,* avoid the lizards by remaining in their burrows under the sand. However, they are poised to leave, waiting for the sand temperatures to heat up even more, until it reaches about 60°C, when the temperature of the air at ant height is about 46.5°C. Temperature "testers" among them lurk at the nest entrance. At the right moment, they signal the time to come out by releasing pheromones from their mandibular secretions. The rest of the colony then rushes out into the field to forage safely, until they too must retire back to their underground shelters—when they experience air temperatures of 53.6°C, which is just a fraction of a degree below their thermal death point.

In the southwestern deserts of the United States near Phoenix, Arizona, the desert or Apache cicada, *D. apache,* also engages in a thermal arms race against vertebrate predators. These cicadas are active at the hottest time of the year, and even then they wait until the high midday temperatures of 44°C (in the shade) near noon to be most active, when the cicada-killing wasps and birds are forced to retire from the heat.

In the deserts of southern California, the grasshopper *Trimerotropis pallidipennis* endures heat rather than regulating heat loss like the sweating cicada. By blending in with the background of the desert floor, it hides to escape bird and lizard predators. Normally grasshoppers that inhabit the ground stilt high above that substrate when it becomes heated to very high temperatures in sunshine. But to remain camouflaged it is imperative for *T. pallidipennis* to crouch down onto the searing hot ground. When that ground heats to near 60°C in sunshine, the duration of time that a grasshopper can remain hidden is limited by how high a body temperature it can tolerate. *T. pallidipennis* has evolved to tolerate the extraordinary high body temperature of 50°C and can thus escape into the sanctuary of sunlight, where a predator such as a lizard or bird cannot hunt.

It is probably rare that insects escape predators by seeking out low temperatures. Possibly the best candidates are the Cuculliinae, a subfamily of the generally endothermic Noctuidae or owlet moths. The Cuculliinae are a northern circumpolar group of moths, and in northern New England they may fly during any month of the winter when temperatures reach 0–10°C and when most of their bat and bird predators have left. During flight, cuculliinines have flight-motor temperatures near 30–35°C, as do other moths of their size and wing loading. Unlike all other moths, however, the cuculliinines can begin to shiver at the extraordinarily low muscle temperature of 0°C, and they continue shivering to warm up all the way to 35°C.

The giant hornets, *Vespa mandarinia,* attack honey bee colonies. During a typical giant hornet attack, a lone hornet forager first captures bees at the periphery of the bees' nest. After several successful foraging trips to the beehive, the hornet deposits a marking pheromone at the hive entrance from the van der Vecht gland at the tip of the abdomen. This pheromone attracts other hornets from the home nest, and then the slaughter phase of the hornet attack begins: 30,000 bees can be killed in 3 h by a group of 30 to 40 hornets. Subsequently the hornets may occupy the hive itself, and then they carry off the bees' larvae and pupae to feed to their own young.

The above happens when hornets attack colonies of the introduced European honey bee, *A. mellifera,* but the Japanese honey bee, *A. cerana,* has evolved an effective counterstrategy to the hornets' mass invasion. With the latter, those unfortunate hornets that are recruited by the pheromone and then try to enter the hive are met and killed by heat as hundreds of bees envelop each wasp into a tight ball. The interior of these bee balls quickly rises to 47°C, killing the hornet but not the bees, whose upper lethal temperature is 48 to 50°C.

Against Competitors

Contest competition or fighting over food is rare in insects, but at least two species of African dung beetles, *Scarabaeus laevistriatus* and *Kheper nigroaeneus,* engage in combat over dung balls that they make to feed on and/or to serve as sexual attractants. An elevated thoracic temperature plays a crucial role in these contests on the ground. The more a beetle shivers to keep warm (with its flight muscles), the higher the temperature of the leg muscles adjacent to the flight muscles in the thorax and the faster its legs can move and construct the dung into balls and roll it away. Endothermy thus aids in the scramble competition for food, and it reduces the duration of exposure to predators. Additionally, hot beetles have the edge in contest competitions over dung balls made by other beetles; in fights over dung balls, hot beetles almost invariably defeat cooler ones, often despite a large size disadvantage.

Mate Competition

For large insects, endothermic heat production is a requisite for flight, and it is during flight that other activities,

including foraging, oviposition, and predator escape, as opportunity and necessity dictate, may occur. Hence, to find a direct effect of body temperature on mating success specifically, one must examine a mating behavior that is not already tightly linked with some other temperature-dependent activity. Singing in some species is a good candidate. Singing is one activity that serves only for mate attraction, and in katydids and cidadas only males sing and the females remain silent. The vigor of this singing activity is associated with and dependent on thermoregulation. Katydids, *Neoconocephalus robustus,* warm up for their ear-shattering mating concerts by shivering, bringing flight-muscle temperatures above 30°C. Males of the Malaysian green bush cricket, *Hexacentrus unicolor,* sing from dusk until well into the night, and before they sing, they prepare themselves by shivering to achieve thoracic temperatures near 37°C. At thoracic temperatures of 37 to 38°C, the males are able to achieve the extraordinarily fast wing movements of up to about 400 vibrations per second.

The dragonfly *Libellula pulchella* demonstrates both the importance of body temperature for mating success and the trade-offs required for maximizing power output as an insect matures. The young, nonreproductive adults of this species are sit-and-wait predators that typically fly with relatively low thoracic temperature. Their flight-muscle performance does not peak at any one temperature; instead, performance is uniformly spread over a wide range of low thoracic temperatures. In contrast, sexually mature males engage in nearly continuous flight in intense territorial contests. At such times, they generate a very high thoracic temperature, and they regulate that thoracic temperature precisely and within only 2.5°C from their upper lethal temperature. Thus, muscle performance of the sexually mature males is narrowly specialized relative to that of young adults that do not engage in strenuous battle.

SOCIAL THERMOREGULATION

Many of the social insects regulate the temperature of their nests in coordinated behavioral and physiological responses involving the adult nest inhabitants. Nest temperature regulation functions primarily to maintain activity and to keep the otherwise thermally labile larvae at the proper temperature for rapid growth. Thermoregulation allows social insects to rapidly build up large nest populations and to inhabit environments where they could not otherwise exist.

Nest Site

One of the first requirements for effective nest temperature regulation is the choice of an appropriate nest site. Typically, northern ants nest in the open, often under solar-heated rocks, or they make solar-heated mounds; many termites also nest to maximize exposure to solar radiation. Honey bees, that live in northern temperate climates require enclosed nest sites such as tree cavities, whereas a variety of other more tropical bees have open and exposed nests.

Nest Construction

Northern vespine wasps enclose their nests in multiple layers of paper that insulate the nest contents. Some termites and ants construct nests so located and constructed as to maximize solar heating in the morning and evening and to minimize overheating at noon. Nests may be constructed so that air circulation and heat transfer are enhanced for thermoregulation.

Behavior and Physiology

Ants regulate the temperature of their brood by carrying it to those parts of the nest with suitable temperatures. Both honey bees and vespine wasps regulate the temperature of the nest, especially near the brood. They fan to circulate air and carry off heat when nest overheating is imminent, and if temperatures continue to increase they carry in water and sprinkle it on the combs for evaporative cooling results. At low temperatures, such as during winter and in swarm clusters outside the hive, the bees crowd together tightly as air temperatures drop, thereby trapping heat inside. As air temperatures rise, the bees on the cluster start to disperse, the cluster loosens, and heat from the interior is released. In hives containing both honeycomb and comb with brood, the bees preferentially cluster around the brood. Brood temperature is maintained near 36°C in hives that may be subjected to air temperatures as low as –50°C and as high as 50°C, provided the bees have access to honey as an energy source for heat production in the cold and to water for evaporative cooling in the heat.

See Also the Following Articles

Cold/Heat Protection • Diapause • Dormancy • Flight • Hibernation • Temperature, Effects on Development and Growth

Further Reading

Esch, H., and Goller, F. (1991). Neural control of honeybee fibrillar muscles during shivering and flight. *J. Exp. Biol.* **159,** 419–431.

Heinrich, B. (1971). Temperature regulation of the sphinx moth, *Manduca sexta.* II. *J. Exp. Biol.* **54,** 141–166.

Heinrich, B. (1993). "The Hot-Blooded Insects: Mechanisms and Evolution of Thermoregulation." Harvard University Press, Cambridge, MA.

Heinrich, B., and Esch, H. (1994). Thermoregulation in bees. *Am. Sci.* **82,** 164–170.

Kammer, A. E., and Heinrich, B. (1972). Neural control of bumblebee fibrillar muscle during shivering. *J. Comp. Physiol.* **78,** 337–345.

May, M. L. (1982). Heat exchange and endothermy in Protodonata. *Evolution* **36,** 1051–1058.

Seeley, T. D., and Heinich, B. (1981). Regulation in the nests of social insects. *In* "Insect Thermoregulation," (B. Heinrich, ed.). Wiley, New York.

Toolson, F. C. (1987). Water prolifigy as an adaptation to hot deserts: Water loss rats and evaporative cooling in the Sonoran Desert cicada, *Diceroprocta apache* (Homoptera: Cicadidae). *Physiol. Zool.* **60,** 379–385.

Wehner, R., March, A. C., and Wehner, S. (1992). Desert ants on a thermal tightrope. *Nature* **357,** 585–587.

Thrips

see *Thysanoptera*

Thysanoptera

Laurence Mound
CSIRO Entomology, Canberra, Australia

The 5000 described species of Thysanoptera, the thrips, exhibit a wide range of biologies. About 50% feed only on fungi, with most of these feeding on hyphae but some on spores. Of the remainder, approximately equal numbers feed either in flowers or on green leaves; a few are obligate predators on other small arthropods. Several opportunist species are crop pests, causing feeding damage and vectoring tospoviruses, but sometimes acting as beneficials by feeding on other pest arthropods (Fig. 1).

Thrips have unique asymmetric mouthparts involving only one mandible, a life history that is intermediate between those of the hemi- and the holometabola, and a haplodiploid sex control system, and many species exhibit complex behavioral patterns including leking, fighting, and eusociality.

THYSANOPTERA STRUCTURE

Larvae and adults have only a left mandible. Their paired maxillary stylets (lacinia) are coadapted to form a feeding tube with a single central channel and a subterminal aperture. These feeding stylets emerge through a mouth cone that points either downward or backward. The dorsal surface of the head is symmetrical, but the ventral surface is asymmetrical, reflecting the absence of the right mandible. The head bears a pair of antennae commonly with seven or eight segments, although the plesiomorphic (or ancestral) number is presumably nine, and various species have segments fused to produce lower numbers.

Paired compound eyes are usually well developed, although reduced to less than 10 ommatidia in some wingless species. Winged, but not wingless, adults have three ocelli between the compound eyes. The pronotum commonly has a regular number of major setae, five in Phlaeothripidae but usually only two in Thripidae. The legs of adults lack typical insect tarsal claws, but each tarsus has an eversible bladder-like arolium.

Members of the two suborders of Thysanoptera, Terebrantia and Tubulifera, differ from each other considerably in structure. The forewings of adult Tubulifera lack longitudinal veins, have a smooth surface, and bear nonarticulating fringing cilia that insert directly into the wing membrane. The abdominal tergites of these species bear one or two pairs of sigmoid wing-holding setae, under which the wings lie flat on top of each other when at rest. Moreover, the 10th abdominal segment is tubular, with the anus terminal but the genital opening at the base of the tube, the female's ovipositor being an eversible, chute-like, structure.

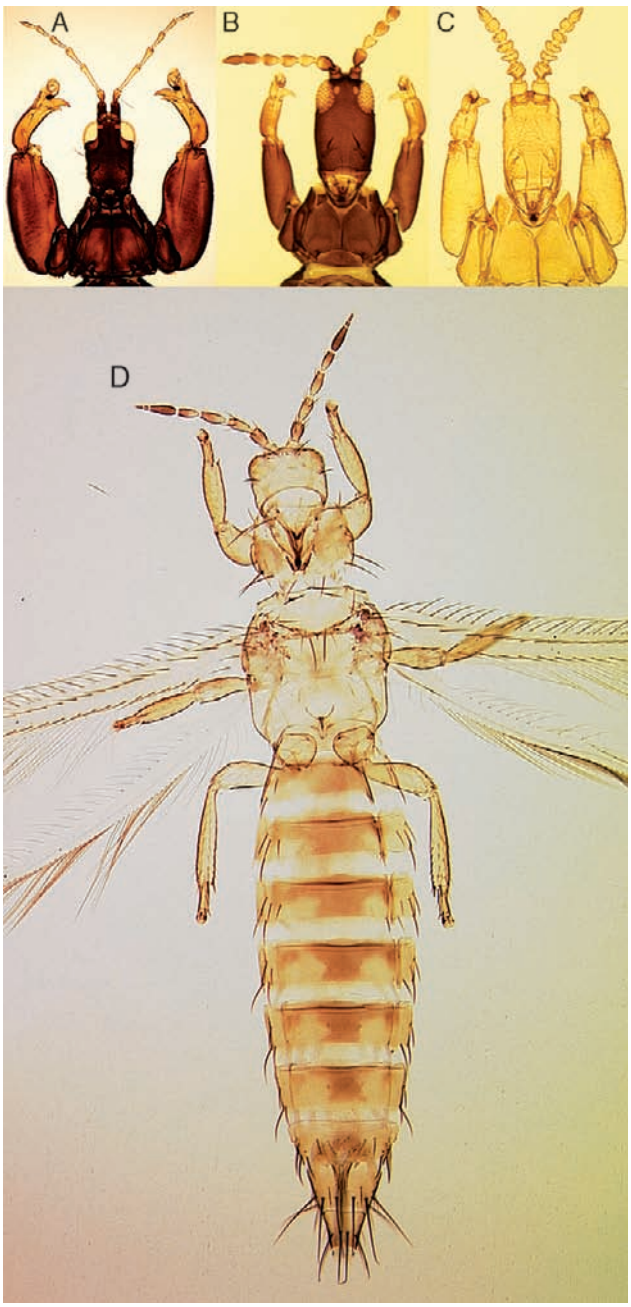

FIGURE 1 Thysanoptera diversity. (A) *Lichanothrips pulchra* female, which creates a domicile by gluing together pairs of Acacia leaves. (B) *Oncothrips waterhousei* foundress, which induces leaf gall on Acacia. (C) *O. waterhousei* first-generation female, which functions as a soldier to defend a gall. (D) Western flower thrips *(Frankliniella occidentalis),* one of the world's major insect pests.

In the species of Terebrantia, in contrast, the forewings have two longitudinal veins, the wing surface is covered in

microtrichia, and the fringing cilia are inserted into sockets that are figure-8-shaped. The wings lie parallel to each other on the abdomen when at rest, with the cilia of the posterior margins in the midline, but the wing-holding mechanisms vary considerably between species. The 10th abdominal segment of Terebrantia species is incomplete ventrally, and the ovipositor comprises four saw-like blades that are used to insert an egg into plant tissue.

FAMILY CLASSIFICATION

The suborder Tubulifera is usually considered to include a single family of about 3000 species, the Phlaeothripidae, although aberrant species, or groups of species, within this suborder are sometimes segregated into small families. The Terebrantia is a more diverse group of 2000 species and currently includes eight families.

Phlaeothripidae

About 600 of the 3000 species in the family Phlaeothripidae feed by ingesting whole fungus spores, and these species comprise the subfamily Idolothripinae. These are some of the largest thrips, with body sizes up to 15 mm, and the males are commonly considerably larger than the females, with prominent tubercles and large foretarsal teeth. Moreover, these large thrips commonly exhibit patterns of allometry, such that the largest males differ considerably in structure and appearance from the smallest in the same population, and these size differences are reflected in their behavior.

The remaining members of the family constitute the Phlaeothripinae, most species being grouped into three ill-defined lineages. The *Haplothrips* lineage includes those Phlaeothripidae most commonly seen in the north-temperate zone, where they are common in the flowers of Asteraceae and Poaceae, including cereal crops. In the warmer parts of the world, members of the *Liothrips* lineage are more abundant, feeding on the leaves of shrubs and trees and commonly causing them to distort, to roll, or to form discrete galls. Many species and genera have been described in this group, but there are few good studies on biology, host specificity, or structural variation, so that species recognition and generic classifications remain unsatisfactory. Species in the third major group, the *Phlaeothrips* lineage, feed on fungal hyphae on dead branches or in leaf litter. Again, many exhibit complex allometry in males, form colonies, and are subsocial. The complexity of variation in structure within and between sexes is such that males and females can sometimes not be recognized as belonging to the same species unless observed alive within a single colony.

Thripidae

In contrast to the other Terebrantia, adults of all 1700 thripid species have slender emergent, simple or forked, sense cones located on the third and fourth segments of the antennae. The greenhouse thrips, *Heliothrips haemorroidalis,* is placed in the subfamily Panchaetothripinae, together with about 120 related species that have a body and legs heavily reticulate. Similar small numbers of species are placed in two further subfamilies, Dendrothripinae and Sericothripinae, but most thripids are placed in the Thripinae. This subfamily includes most of those flower-living insects that are commonly recognized as thrips, particularly the 450 members of two genera, *Thrips* and *Frankliniella,* many of which are important crop pests (Fig. 1D).

Aeolothripidae

The 200 species in this family all have nine-segmented antennae with at least the last three segments closely joined, and the sensoria on the third and fourth segments are almost always linear along these segments. Aeolothripids are relatively large thrips, with the forewings broadly rounded at the apex and commonly banded black and white. Although they are probably all predatory on other arthropods, the common flower-living members of the genus *Aeolothrips* also feed on plant tissue, the only obligate predators being tropical species in the genera *Franklinothrips, Mymarothrips,* and possibly *Stomatothrips.*

Melanthripidae

The 60 species in this family are sometimes referred to the Aeolothripidae, but they are all flower feeders, not predators. In contrast to aeolothripids, all nine antennal segments are distinct and bear transverse rows of microtrichia, and the sensoria on the third and fourth segments are linear around the apex. Moreover, remnants of an eighth sternite are visible on the abdomen, as in Merothripidae.

Heterothripidae

The 70 species recognized in this family are found only in the New World. The adults all have antennae with nine segments, and the sensoria on the third and fourth segments are continuous around the apex. Little is known about the biology of most of these species, but they probably all breed in flowers, some being known to be host-specific, and a wide range of plant families is involved.

Minor Families

The remaining four families in the Terebrantia include a total of fewer than 30 species. A single, widespread, tropical species is placed in the Uzelothripidae. This is presumed to be fungus-feeding and has a remarkable whip-like terminal antennal segment. The Adiheterothripidae includes two species from the western United States and four species from the flowers of date palms between the eastern Mediterranean and India. The Fauriellidae includes one species from California, two from southeastern Africa, and two from southern Europe. The family Merothripidae includes fewer than 20 species, mostly from

South America, all of which feed on fungus on dead twigs and leaves. These are minute thrips, and they retain structural character states that are presumed to be closest to the ancestral states of all Thysanoptera.

LIFE HISTORY

Reproduction in thrips is haplodiploid, which involves males having one-half the number of chromosomes of females and developing from unfertilized eggs. Despite this, several thrips species can produce females from unfertilized eggs, a process known as thelytoky. In a few common and widespread species, such as the greenhouse thrips *H. haemorroidalis,* males are rare or, as in the introduced basswood thrips, *Thrips calcaratus,* unknown. Larvae usually hatch within a few days, but in some of the larger Idolothripinae eggs develop while still in the abdomen of a female and larvae hatch soon after the eggs are laid.

There are two larval stages, both of which feed actively for about 2 to 5 days. In Terebrantia there are then two pupal stages with the antennal segmentation reduced or absent and the mouth parts nonfunctional. Wing rudiments can be seen in the first, the propupa, but are much longer in the second, the pupa. Tubulifera species are even more remarkable in that they all have three pupal stages.

The site of pupation varies greatly among species. Some Thripidae and Phlaeothripidae pupate on leaves in association with their larvae, but more commonly pupae are found in leaf litter or on the trunks of trees. The pupae of Aeolothripidae, Heterothripidae, and even a few Thripidae are contained within a pupal cocoon that is spun by the second instar. Larvae in these species have stout tubercles near the posterior end of the abdomen that presumably assist burrowing into the soil.

Species that are host-specific within the flowers of particular plants are usually univoltine. Related polyphagous species often breed continuously as long as conditions remain suitable. Overwintering stages are usually pupae, or adults, but the citrus thrips, *Scirtothrips citri,* overwinters as eggs in leaves and shoots.

FEEDING

Little difference is evident between phytophagous, predatory, and fungus-feeding species in the structure of their feeding stylets. The two maxillary stylets fit together along their length with a tongue-and-groove system, and their apices are linked with slender finger-like projections around a subapical feeding aperture. Only in the spore-feeding Idolothripinae are the stylets clearly adapted to the food consumed, being exceptionally broad to facilitate the ingestion of whole spores.

The only obvious difference in the stylets between species is their length. Among the Terebrantia, the maxillary stylets are short and restricted to the mouth cone, but in many Phlaeothripidae the stylets are deeply retracted into the head, sometimes as far as the eyes, and lie close together along the midline. In a few species the stylets are longer than the total body length and are coiled within the head.

When feeding, a thrips initially makes a hole by extruding the single mandible, a solid needle-like structure. The mandible is then withdrawn, and the maxillary stylets are inserted into the food source through the hole, saliva being pumped into the tissues and the resultant fluid being pumped back into the thrips' crop. In contrast to the independent salivary and feeding channels of an aphid or similar bug, a thrips has a feeding tube with only a single channel.

In those thrips species that transmit tospovirus diseases to plants, the virus is acquired only by a first, or early second, instar. The virus is taken up from an infected cell when a larva feeds and passes into the gut and through the gut wall. Ultimately, the virus reaches the salivary glands of the adult, from where it is then reinjected into a plant. Although an adult may take a tospovirus into its gut, the virus is not able to pass through the gut wall and reach the salivary glands, so that no adult can acquire and then transmit a tospovirus. Viruliferous adult thrips must have acquired the virus when they were larvae.

HOST RELATIONSHIPS

Almost nothing is known of host specificity in fungus-feeding thrips. In Europe, certain species of the phlaeothripine genus *Hoplothrips* are associated with *Stereum* fungi on dead branches of trees. Those tropical species of Idolothripinae that have exceptionally broad stylets presumably feed on larger fungal spores than related species with slightly less broad stylets. Such details of thrips natural history are, regrettably, little studied. Even among phytophagous species, precise host relationships have rarely been established. In Europe, the host plants of most species in the genus *Haplothrips* are known—many living only in the flowers of a particular species of Asteraceae—but the plants on which the many tropical species of this genus breed remain unknown. Similarly, the host plant of not 1 of the 43 endemic North American species of the genus *Thrips* has been determined. A major reason for this lack of sound biological information about thrips species lies in the dispersive activity of adults, these sometimes being found in very large numbers on plants on which they cannot breed.

Despite the inadequacy of field studies, some host-plant relationships are well established. For example, the Palearctic species of *Odontothrips* breed only in flowers of the family Papilionaceae, and the Old World tropical species of *Megalurothrips* also breed in such flowers. Similarly, *Dichromothrips* species live on Orchidaceae in the Old World; *Projectothrips* species live only in the flowers of *Pandanus,* the screw pines of the Old World tropics; and all four species of the adiheterothripid genus *Holarthrothrips* live only in the flowers of the date palm, *Phoenix dactylifera.* The Poaceae has a particularly rich fauna of Thripidae, with *Aptinothrips* and *Limothrips* being specific to grasses in the Palearctic; *Stenchaetothrips, Fulmekiola,* and *Bregmatothrips* specific to grasses in the tropics; and *Chirothrips* and *Arorathrips* breeding in grass flowers in many parts of the world.

In contrast to these host associations at genus level, some thrips genera show an entirely different pattern of host exploitation, with each species using as host some entirely unrelated plant. In the New World genus *Echinothrips,* one species lives on a species of *Selaginella* (Lycopsida), another lives on the needles of *Tsuga* (Pinaceae), a third lives on various soft-leaved plants in Euphorbiaceae and Balsaminaceae, and a fourth is polyphagous with no clear host associations.

Host plant data are particularly weak in the Phlaeothripidae, many species being known only from single samples, particularly in the tropics. However, in Australia one suite of about 200 species of Phlaeothripinae is associated only with *Acacia,* and this suite of species appears to represent a single evolutionary lineage. This close association between a thrips lineage and one plant genus is remarkable, particularly so when the richness of the Australian flora and the complete absence of Phlaeothripinae from the leaves of any of the 900 species of *Eucalyptus,* the second largest plant genus in Australia, are considered.

FLIGHT AND DISTRIBUTION PATTERNS

Adults of both phytophagous and mycophagous species can be observed in warm weather climbing to the tops of plants such as grasses or climbing up dead twigs. They spread their wings, and in Terebrantia the marginal cilia are combed from a parked position parallel to the wing margin into a flight position at right angles to the wing. Many species then actively jump into the air and fly vertically upward. Despite this activity, the dispersal of thrips is presumably determined primarily by air currents, and even wingless individuals are dispersed widely by the wind. Wingless species on the mountains of southeastern Australia are recorded dispersing to the northern part of South Island, New Zealand, a distance of more than 1600 km across the Tasman Sea. Such dispersal is probably not entirely fortuitous, but is a function of the behavior of particular species. Some species, such as *Frankliniella schultzei* in Australia, seem to be particularly prone to long-distance migration flights. In contrast, the extensive dispersal of the western flower thrips, *Frankliniella occidentalis,* around the world is probably due mainly to the horticultural trade.

The worldwide distributions of many thrips species are primarily the result of human trading patterns. For example, *Chirothrips* species pupate within the glumes of grass florets and have thus been distributed around the world in commercial grass seed. Other grass thrips were widely distributed in hay used to feed animals on sailing ships during the period of colonial expansion. Thus some members of the European genera *Aptinothrips* and *Limothrips* can now be found in temperate zones all over the world, including mountains in tropical countries. Similarly, orchids, bananas, and sugarcane, all of which are transported and planted from plant parts, not seeds, have been accompanied around the world by their pest thrips species. The vast increase in the use of air transport by the horticultural trade since 1980 is expanding the distributions of pest thrips around the world.

Despite these disrupted patterns of thrips distribution, there remain several natural distribution patterns that are of interest. The family Heterothripidae is confined to the New World, and most species of Merothripidae are restricted to that area. In contrast, the genus *Thrips,* with more than 270 species worldwide, has no species native to the Americas south of Mexico, and the genus *Frankliniella,* with more than 180 species, includes very few that have a natural distribution anywhere outside the New World. Presumably these two advanced genera of Thripinae evolved at about the time that the American continent separated from Europe; the presence of the Heterothripidae only in the New World may also suggest a relatively recent origin for that family.

Within the Melanthripidae, the genus *Dorythrips* has three species in southern South America and two in Western Australia, and *Cranothrips* has one species in South Africa and several in Australia. Within the Aeolothripidae, the Cycadothripinae and Dactuliothripinae are considered sister groups. Species of the first live only in Australia, in the cones of *Macrozamia* cycads, whereas the second is found only along the western side of the American continent in the flowers of various plants. Among the more advanced Aeolothripidae, the two genera *Aeolothrips* and *Desmothrips* are ecological and morphological counterparts of each other, the first restricted to the Holarctic, the second to Australia.

Among the Phlaeothripidae, geographical patterns of distribution are less clear. Some smaller genera are restricted either to the Western or to the Eastern Hemisphere, but larger genera are more widely distributed. The leaf-feeding members of the genus *Liothrips* are found throughout the tropics, including the Pacific region, although they have been little studied in the Neotropics. The large species of the genus *Elaphrothrips,* all of which feed on fungal spores, are found widely throughout the tropics, being replaced east of Wallace's Line by members of the genus *Mecynothrips.*

BEHAVIOR

In Terebrantia, males are usually much smaller than females, whereas in Tubulifera males are commonly much larger than females. These differences are related to different patterns of behavior. In fungus-feeding phlaeothripids, it is not uncommon for a male to defend a female and the egg mass that she produces or, alternatively, for a male to defend a single egg mass to which various females contribute after first mating with him. These strategies lead to competitive behavior between individual males, involving flicking with the abdomen to displace a rival or even stabbing with foretarsal teeth to kill a rival. Moreover, while large males are involved in such competitive activities, smaller males may sneak-mate. Clearly there is a balance of advantages, between developing into a small male on a restricted food supply and requiring more food and developing a larger body.

Competition between males is possibly a plesiotypic behavioral trait in Thysanoptera, because in the basal clades Merothripidae and Aeolothripidae males of some species are polymorphic and presumably competitive. In *Merothrips* and *Cycadothrips* species the largest and smallest males differ considerably in body size and in the strength of their forelegs and abdominal setae. Male competitiveness also occurs in some Thripidae, including the pest species *F. occidentalis.* Males of Kelly's citrus thrips, *Pezothrips kellyanus,* a pest of citrus in Australia and the Mediterranean, form leks on ripe lemons, and females are attracted to these male aggregations for mating.

THRIPS DOMICILES

The term "domicile" is used to include both leaf galls that are induced by thrips and the shelters that many Australian Phlaeothripinae construct by fixing leaves together with glue or silk. Gall induction by thrips, mainly by species of Phlaeothripinae, is widespread in tropical countries although inadequately recorded in the Neotropics. Galls range from simple rolled leaves containing a few thrips to highly contorted masses of leaf tissue enclosing up to 10,000 adults and larvae. In most gall-inducing thrips from the Oriental Region there is little sexual dimorphism, whereas gall thrips on *Acacia* and *Casuarina* trees in Australia exhibit considerable variation both between sexes and between long- and short-wing morphs.

In some phlaeothripines on *Acacia* in Australia, the gall foundress is a fully winged female, but the eggs she lays first develop into short-winged adults of both sexes. These adults sometimes have reduced reproduction and act as soldiers to defend a gall while the foundress produces a second and larger generation that become winged adults. This behavioral strategy falls within the definition of eusociality.

The habit of domicile construction by thrips is recorded only from Australia, in a suite of species on *Acacia.* At least 30 species are now known to form such shelters using a secretion from the anus. In some of these species the *Acacia* leaves, or more precisely phyllodes, are glued together in pairs at an angle, and the thrips breed within the space created by a ring of glue and the two phyllode surfaces. In other species the secretion is more silken in form, and this silk is used to sew together two or more phyllodes enclosing a small space within which the thrips breed. At least two species are known to use this silken material to weave a tent on one surface of a phyllode within which to breed.

These Australian thrips domiciles, whether galls or constructs, are evidently of great value in ecosystems with high temperatures and low humidities. As a result, a range of kleptoparasitic species has evolved, each of which has a different method of trying to usurp a domicile. These methods of driving out the original inhabitants range from a frontal assault with sharp foretarsal teeth to a porcupine-like action of the abdomen which bears many stout setae. A few Australian *Acacia* thrips species seem to have evolved into true inquilines, in that they breed within the colony of a domicile producer without unduly disturbing the original inhabitants.

POLLINATION

Thrips are sometimes abundant in flowers, their bodies often bear large numbers of pollen grains, and they can fly actively between flowers. Despite this, their function as pollinators, even their presence, is frequently overlooked; no less than Charles Darwin complained of thrips interfering with his experiments on larger pollinators! Thrips have been demonstrated to be pollinators in a wide range of flowers: heather plants in the north of the Northern Hemisphere, dipterocarp trees in Malaysia, a rain forest tree in eastern Australia—*Wilkiea huegeliana* (Monimiaceae), the Panama rubber tree—*Castilla elastica* (Moraceae), and several *Macrozamia* cycad species in Australia.

Although a thrips individual does not carry a large pollen load, the large number of thrips in each flower, each one carrying between 10 and 50 pollen grains, is enough to deliver pollen to many stigmatic surfaces. Despite this, thrips probably remain the most underestimated of all flower pollinators, the majority of botanists failing to see these small insects, let alone consider their significance.

PEST SPECIES

In general, the only thrips that are noticed by nonspecialists are pest species. Most pest thrips are members of the Thripidae, although particular phlaeothripid species cause leaf damage on a wide range of plants in the warmer parts of the world, including decorative *Ficus* trees, black pepper vines, and olives. In the Northern Hemisphere cereal crops are also damaged by a phlaeothripid species. In contrast there are many thripid species that attack cultivated plants, and some of these cause serious economic losses.

Citrus production in both California and southern Africa can suffer considerable losses because of downgrading of scarred fruit on which thrips have been feeding. The value of a nectarine can be seriously reduced through a single thrips larva feeding on the fruit when it is young. Cucumbers, capsicums, and strawberries are badly distorted at times because of the feeding activity of thrips. Roses, carnations, and chrysanthemum flowers can be devalued through thrips feeding damage, and table grapes burst and become fungal-infected through thrips oviposition scars.

Worldwide, there are four species of thripids that are particularly significant as pests: the onion thrips *(Thrips tabaci),* the melon thrips *(Thrips palmi),* the tomato thrips *(Frankliniella schultzei),* and the western flower thrips *(F. occidentalis).* Feeding by each of these species can cause severe damage on some crops, but the most serious damage associated with them is due to the tospoviruses they can transmit. More than 12 of these viruses have been described, and although they cause damage only to plants it is evident from their molecular structure that they are members of the animal virus family Bunyaviridae. The origin of the plant infections remains unknown, but each tospovirus is dependent for its

continued existence on being transmitted from one plant to another by one or more of 10 thrips species.

See Also the Following Articles

Gallmaking • Mouthparts • Plant Diseases and Insects • Rostrum

Further Reading

Crespi, B. J., and Mound, L. A. (1997). Ecology and evolution of social behaviour among Australian gall thrips and their allies. *In* "Evolution of Social Behaviour in Insects and Arachnids" (J. Choe, and B. J. Crespi, eds.), pp. 166–180. Cambridge University Press, Campridge, UK.

Lewis, T. (ed.) (1997). "Thrips as Crop Pests." CAB Int., Wallingford, UK.

Moritz, G., Morris, C. D., and Mound, L. A. (2001). ThripsID—Pest thrips of the world. An interactive identification and information system. CD-ROM published by ACIAR and CSIRO Publishing, Australia.

Mound, L. A., and Marullo, R. (1996). The thrips of Central and South America: An introduction. *Mem. Entomol. Int.* **6,** 1–488.

Thysanura

see *Archaeognatha; Zygentoma*

Ticks

Daniel E. Sonenshine
Old Dominion University

Ticks comprise a distinct group of exclusively blood-feeding ectoparasites familiar to most people in virtually all regions of the world. Ticks transmit a greater variety of disease-causing pathogenic agents than any group of arthropods, including protozoan, viral, bacterial, and even fungal pathogens. An example is Lyme disease (LD), which is now the most important vector-borne disease of humans in the United States, Europe, and Asia. In numerous countries in tropical and subtropical regions of the world, tick-borne diseases of livestock such as babesiosis, theileriosis, and heartwater, have made it difficult or impossible to raise domestic animals for food or animal products. Ticks also can cause irritating or even fatal injury to humans and animals because of paralysis, toxicity, or severe allergic reactions to their bites.

BODY STRUCTURE

The tick body is organized into two major body regions, the anterior capitulum, bearing the mouthparts, and the idiosoma, which bears the four pairs of walking legs (Figs. 1, 2). There is no head, and the highly fused body is not divided into a thorax and an abdomen. The capitulum contains the toothed hypostome, which embeds the tick into the host's skin and also contains the food canal for blood imbibition, the chelicerae, delicate paired appendages that cut into the skin, and the four-segmented paired palps that provide important sensory information for host identification. In argasid ticks, the capitulum is recessed under an anterior extension of the body. The remainder of the body contains the genital pore, anus, and spiracles, which are visible on the ventral surface. In ixodid ticks (so-called hard ticks), a prominent platelike scutum is found on the dorsal surface. Argasid ticks (so-called soft ticks) are similar to the ixodid ticks but lack a scutum, and the body cuticle is leathery.

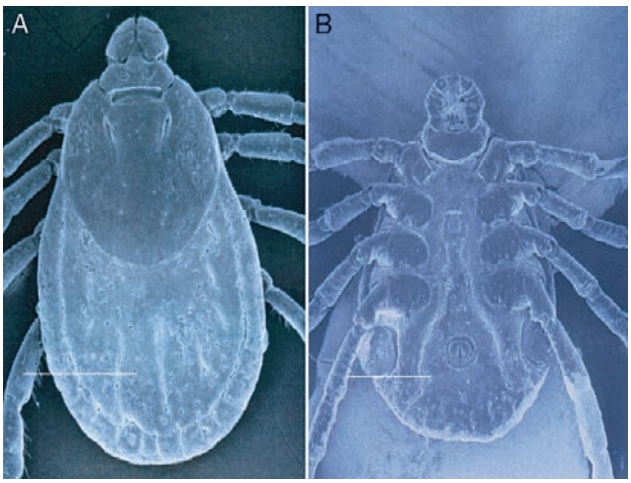

FIGURE 1 Scanning electron micrographs of a representative adult female ixodid tick, *D. variabilis:* (A) dorsal view and (B) ventral view. Scale = 1 mm.

The interior of the tick body is a simple, open cavity called the hemocoel that is filled with a circulating fluid, the hemolymph, which bathes the internal organs. In ticks, as in other terrestrial arthropods, there is no hemoglobin, and the hemolymph does not function in oxygen transport. Most of the body interior is occupied by the midgut, the largest internal organ of the tick body, which consists of a central saclike stomach and several lateral diverticuli. Other prominent internal organs are the paired salivary glands, which

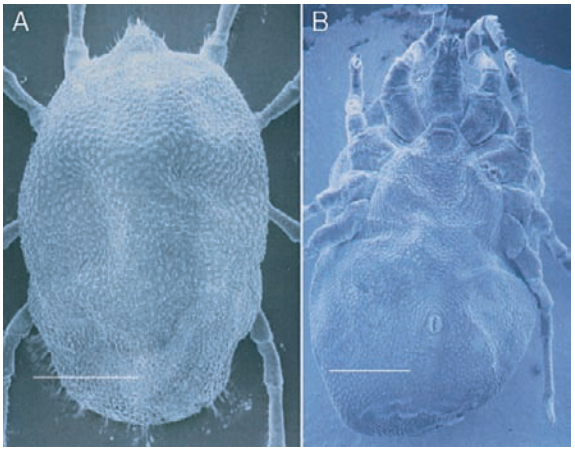

FIGURE 2 Scanning electron micrographs of a representative adult female argasid tick. *O. parkeri:* (A) dorsal view and (B) ventral view. Scale = 1 mm.

appear as white grapelike clusters, and the reproductive organs. In females, these are the ovary, paired oviducts, uterus, and seminal receptacle, and the vagina that connects the system to the genital pore. During feeding, the ovary enlarges and becomes distended with large, brown or amber-colored eggs. In males, the reproductive system consists of the testis, vasa deferentia, and seminal vesicle, and the ejaculatory duct, which is connected to the genital pore. Much of the system is obscured by the large, white multilobed accessory gland. This gland provides the components for the saclike spermatophore that the male tick uses to transfer its sperm to the female. Also present are numerous tracheal tubes, connected to the marginal spiracles, that provide the respiratory system, and the Malpighian tubules and rectal sac, connected to the anal pore, for waste elimination. Argasid ticks have a pair of coxal glands that excrete via the coxal pores excess water and salts accumulated during feeding. The fused central nervous system, the synganglion, is located in the body above the genital pore.

SYSTEMATIC RELATIONSHIPS

Ticks are classified with the class Arachnida, the group that contains the familiar spiders and scorpions. Arachnids have chelicerae, which are appendages with pincerlike or scissorlike cutting edges, instead of mandibles. There is no head or thorax such as occurs in insects. There are no antennae. Ticks are grouped together with the mites in the subclass Acari. Ticks constitute a distinct suborder, the Ixodida, within the acarine order Parasitiformes. The Ixodida contains three families, the Ixodidae, Argasidae, and Nuttalliellidae. The Ixodidae or hard ticks are by far the largest of the different families of ticks, with approximately 650 species. Ixodid ticks have three active life stages, including a single nymphal stage. Hard ticks contain most of the important disease vectors and pest species that plague livestock and wildlife. The Ixodidae are further subdivided into the Prostriata, represented by the single genus *Ixodes,* which is easily recognized by the anterior anal groove, and the Metastriata, which include the remaining 13 genera, in which the anal groove is posterior to the anal aperture. The Ixodidae include the ticks that transmit the agents of LD, Rocky Mountain spotted fever, bouteonneuse fever, babesiosis and theileriosis of livestock, and most of the other tick-borne disease-causing agents.

The Argasidae comprise the soft ticks, with their leathery, highly flexible cuticle. There are approximately 170 species divided into five genera (four according to some authorities). Soft ticks also have three active life stages, but most species have multiple nymphal stages before they develop into adults. Ticks of the genus *Ornithodoros* (> 100 species) have a leathery cuticle with innumerable small elevations known as mammillae and a rounded body margin. Ticks of the genus *Argas* (~ 58 species) have a flattened lateral margin marked by a sutural line. The leathery cuticle bears small ridges and folds in a rectangular pattern, each with a pit at the center of these buttonlike enclosures. Except for species that transmit the spirochetes that cause relapsing fever, most soft ticks are not important in the transmission of disease.

The third family of ticks is the monospecific family Nuttalliellidae, represented by only one species, *Nuttalliella namaqua,* in southern Africa. It contains structures characteristic of both other tick families. Its highly wrinkled cuticle with pits and elevations resembles the cuticle of argasid ticks, but its dorsal pseudoscutum resembles the scutum of the ixodid ticks.

Ticks are an ancient group of specialized acarines that were already well developed during the Mesozoic era (i.e., the era of the dinosaurs). A larval tick found in amber in New Jersey, was dated between 90 and 94 mya (i.e., during the Upper Cretaceous period). Although exhibiting some unusual characteristics in its setal arrangements, the tick was readily characterized as a member of the genus *Carios,* a genus that exists today. This finding suggests that this genus at least (and perhaps other argasid ticks) has not changed very much in many millions of years.

TICK BIOLOGY: LIFE CYCLES, FEEDING BEHAVIOR, DEVELOPMENT, AND REPRODUCTION

The tick life cycle comprises the egg and three active stages, namely, larva, nymph, and adult. There is only a single nymphal instar in the ixodid tick life cycle, but varying numbers of nymphal stages may occur in the argasid tick life cycles. All ticks feed on blood during some or all stages. Most species are three-host ticks; that is, each stage attacks hosts, feeds, and detaches before developing into the next life cycle stage. Adult ticks seek hosts, feed, and, in the case of engorged females, drop off to lay their eggs. Ticks can survive for long periods between blood meals. Consequently, when feeding is delayed, the life cycle may be extended for years or, in the case of some argasids, for a decade or longer. There are major differences between the life cycles of the Ixodidae and the Argasidae, as discussed next.

Life Cycles of Ixodid Ticks

The ixodid ticks feed slowly, from several days to as long as 2 weeks. Immature and adult ticks each take a blood meal, except for the nonfeeding males of some species. After crawling onto their hosts, these ticks embed their mouthparts into the host skin and secrete cement from their salivary glands into and around the wound site to anchor themselves. The cement binds the ticks firmly in place and makes them very difficult to remove. During blood feeding, the ticks secrete potent anticoagulants and anti-inflammatory agents, which suppress host wound healing and facilitate blood flow. As the ticks feed, new cuticle is synthesized to accommodate the enormous blood meals the animals consume, often 10 to 100 times their original body weight. Females feed only once. Mating occurs during feeding, although ticks of the genus *Ixodes* may also

mate prior to host attachment. In the metastriate Ixodidae (i.e., ixodids other than ticks of the genus *Ixodes*), mating occurs within a few days after the commencement of feeding and is regulated by sex pheromones, including the volatile 2,6-dichlorophenol and the nonvolatile cholesteryl esters on the body surface. Following mating, females suck blood rapidly (24–48 h) and swell enormously, whereupon the replete females drop from their hosts, find a sheltered location, and lay thousands of eggs. An example is the American dog tick, *Dermacentor variabilis,* which typically lays more than 5000 eggs. Following oviposition, the female dies. In contrast to the females, males swell only slightly during feeding. However, they can mate many times, feeding between matings. In certain species of *Ixodes,* mating may occur either on or off the host (e.g., the blacklegged tick, *Ixodes scapularis,* or the sheep tick, *I. ricinus*). In some nest-inhabiting *Ixodes* species, the males have vestigial hypostomes. These ticks always mate off the host.

Once oviposition has been completed, the larvae hatch in the thousands and begin host-seeking activity. Except for the nest-inhabiting species, the larvae disperse into the vegetation, where they come in contact with passing animals. Once they have attached to a host, the larvae embed themselves into the host skin, form a feeding pool, and engorge in the manner already described. Feeding usually takes 2 to 4 days, whereupon the engorged larvae drop from their hosts to molt on the ground. Molting usually occurs in some sheltered microhabitat such as soil or leaf litter, or in host nests. After molting, nymphal and adult ticks must seek another host and feed. However, more than 90% of the life cycle is spent off the host. When host seeking and feeding occur in all three parasitic stages, the pattern is termed a three-host life cycle.

A few ixodid species exhibit a two-host or one-host life cycle. For example, in the camel tick *Hyalomma dromedarii,* both larvae and nymphs feed on the same host (two-host life cycle), and in the cattle tick *Boophilus annulatus,* all stages feed, molt, and even mate on the same host (one-host life cycle).

Life Cycles of Argasid Ticks

Feeding is very rapid among the argosid ticks. Once they have crawled onto a host, the ticks embed their mouthparts in the same manner as their ixodid relatives, but without secreting cement. Bloodsucking commences quickly and, as feeding progresses, the bloated ticks excrete copious quantities of a clear, colorless coxal fluid (some times this occurs soon after feeding). By eliminating excess water and salts via the coxal fluid, the ticks can concentrate their blood meals and adjust their internal water balance. The ticks expand to about 5 to 10 times their original body weight, depending on the ability of the cuticle to stretch. Following feeding, often completed within as little as 30 to 60 min, the replete ticks drop off to molt or, if female, to lay eggs. Argasid females take repeated small blood meals and lay small batches of eggs (typically < 500 eggs in a batch) after each feeding (multiple gonotrophic cycles). The interval between feedings is typically several months but may be up to several years, depending on host availability. Mating usually occurs off the host. Because of the multiple nymphal instars (six or even seven in some species), argasid ticks often live for many years. In addition, these ticks are highly resistant to starvation, an advantage that can extend their longevity even further. As a result, the entire life cycle may take from 10 to 20 years.

Following oviposition and hatching, most argasid tick larvae seek hosts, feed rapidly, and molt to the first nymphal instar. These nymphs seek hosts again, feed rapidly, and molt to the second nymphal instar. Subsequently, the life cycle varies considerably, leading to additional nymphal instars or proceeding directly to the adult stage. As a rule, males emerge earlier than females and have fewer nymphal stages. In some argasids, especially bat parasites, the larvae remain attached to their hosts for many days, feeding slowly, just like their ixodid tick relatives, and then molt twice without additional feeding. Thereafter, the life cycle resembles the typical argasid pattern. Another unusual species is *Otobius megnini,* which has only a single nymphal stage. Neither the males nor the females feed, and the females lay eggs without having had a blood meal (i.e., autogeny).

ECOLOGY

Most ticks are exophiles (i.e., nonnidicolous, living exposed in the open environment rather than in shelters). Most ticks live in forests, savannahs, brush, grassy meadows, or under stones, crevices, or even in sand in semidesert environments. Others, however, are nidicoles, surviving in caves, burrows, houses, cracks, and crevices where their hosts obtain shelter. This habit is characteristic of most argasids and many species of the genus *Ixodes.*

Seasonal Activity and Host-Seeking Behavior

Exophilous ticks are active during certain periods of the year when climatic conditions are suitable for development and reproduction. During this seasonal activity period, they attack and feed on suitable animals. This is known as host-seeking behavior. At other times, ticks remain in diapause (i.e., a state of reduced metabolic activity). In temperate and subpolar regions, the seasonal activity period is regulated by ambient temperature, changing photoperiod, and incident solar energy. Tick seasonal activity usually commences with the onset of warmer weather and increasing daylength. In what is termed the ambush strategy, hungry ticks climb on the vegetation to varying heights, depending on life stage (e.g., adults climb the highest) and cling to any passing animals. In some species, the ticks emerge from their shelters and run toward their hosts when they detect animal odors (or, rarely, noise) from animals nearby. This so-called hunter strategy is useful in arid habitats, where there is little vegetation or source of other moist, protective covering. Argasid ticks normally do not exhibit seasonal activity, since they live in proximity to their hosts in nests,

burrows, or other shelters. However, in some species specific for migratory birds or bats, host-seeking activity is synchronized with the period of the year when these hosts return to reoccupy their nests.

For many ixodid ticks that occur in temperate or subarctic regions, seasonal activity begins in the spring. In *D. variabilis,* for example, larvae that survived the winter begin to feed on small mammals. Activity accelerates rapidly as increasing numbers of larvae, stimulated by rising soil temperatures and lengthening photoperiods, emerge to attack these animals, reaching the seasonal peak within a few weeks. Feeding by nymphal and adult ticks follows soon afterward, with the adult peak in early summer. In the southern parts of its range, the tick's entire life cycle, from eggs to ovipositing females, is completed in one year. In the northern parts of the *D. variabilis* range, larval emergence is delayed until late spring. Moreover, the cooler soil temperatures and shorter daylengths delay molting of fed ticks. As a result, adults emerge from fed nymphs in late summer or early fall, when soil temperatures and incident solar radiation are declining, and this results in a two-year life cycle. Thus, both a one-year and a two-year life cycle can occur because of variations in climactic conditions within this wide-ranging species.

In *I. scapularis,* larvae and nymphs feed in the spring and summer, as in *D. variabilis,* whereas adults are active in the fall and early spring. However, the order of larval and nymphal feeding is the opposite of the dog tick. Nymphal ticks emerge from their overwintering diapause in spring or early summer, depending on the region of the United States where they occur. Larvae appear next, typically a month or two after the nymphal peak. Meanwhile, fed nymphs molt over the summer, but the newly emerged adults delay host-seeking activity until the cooler months of the fall or early winter. This life cycle pattern enables nymphs infected with *Borellia burgdorferi* to infect mice, providing thereby a reservoir of infected hosts to infect the next generation of larval ticks. The implications of these different tick life cycle patterns for the survival and transmission of zoonotic diseases are discussed further in the sections on specific diseases.

Occasionally, ticks are active only during the winter months. An example is the winter tick, *D. albipictus,* a one-host tick that feeds on horses, deer, elk, moose, and other large ungulates. In this case, larvae commence host-seeking activity in late summer or early fall. Larvae and nymphs feed and molt on the same hosts and the resulting adults reattach, feed, and mate. Feeding and development require many weeks for completion, even though these activities occur on the same host and, as a result, the adults are often found on their ungulate hosts in winter or early spring. In *D. albipictus* declining photoperiod stimulates feeding activity, just the opposite of the pattern seen in *D. variabilis.*

Host Specificity

Host-seeking activity is strongly influenced by the availability of hosts and host selection behavior. All ticks species exhibit varying degrees of host specificity; most (> 85%) exhibit relatively strict host specificity. At one extreme are the argasid ticks that feed exclusively on bats (e.g., ticks of the genus *Antricola* and certain species of *Ornithodoros*). For example, during his graduate student years, Sonenshine colonized the bat tick, *Ornithodoros kelleyi,* in the laboratory. To feed the ticks, he had to maintain a colony of bats collected from limestone caves or attics of old buildings. No other hosts would do. Similarly, the cattle ticks *Boophilus microplus* and *B. annulatus* feed solely on cattle and, when available, on white-tailed deer. Other species exhibit limited host specificity (e.g., *D. variabilis*). Larvae and nymphs of *D. variabilis* feed on a wide range of small mammals (e.g., white-footed mice, meadow voles, etc.), but never on carnivores, ungulates, humans, or other large mammals. In contrast, adults of this species feed on medium-sized and large mammals, including humans (although they can be induced to feed on rodents when confined in capsules). Finally, at the opposite extreme of the specificity spectrum are the opportunistic species that feed on hosts of virtually all types, (e.g., *I. scapularis* and *I. ricinus*). Immatures feed on lizards, birds, and small, medium-sized, and large mammals, including deer and humans. Adults feed on medium-sized and large mammals, including humans. Although the range of confirmed hosts is astounding, even these opportunistic ticks have preferred hosts (e.g., mice for the immatures; deer, sheep, and other mammals for the adults). Host specificity also is strongly influenced by ecological adaptations, so that ticks adapted to a particular habitat in a given region of the world will encounter only vertebrates adapted to the same habitat.

As tick–host associations evolved, ticks gradually developed the ability to facilitate long-term feeding by evading or suppressing host homeostatic systems. For example, *I. scapularis* saliva contains pharmacologically active compounds that suppress edema and inflammation in their hosts while enhancing vasodilation. This leads to greater blood flow into the wound site without the pain and intense itching sensation so characteristic of the bites of mosquitoes or biting flies. These adaptations are most effective for the hosts encountered most frequently by each tick species, so-called preferred hosts, but less effective for uncommon hosts.

Survival between Blood Meals

One of the most remarkable aspects of tick biology is the ability to survive for long periods between blood meals. The tick's midgut serves as a storage organ where the blood meal is digested slowly over long periods. Among the argasids, individuals may survive for several years without feeding while waiting for the occasional wandering hosts that enter their secluded shelters. Among the exophilous ixodids, survival periods are much shorter, but even these ticks may survive for up to one year between blood meals. According to a study by Needham and Teel in 1991, ticks spend more than 90% of their life history off the host.

Ticks also must conserve body water to survive while they wait for hosts. While questing (i.e., perching for attack) on short stems, blades of grass, or other vegetation, ticks are exposed to desiccating conditions that can become life threatening within a few days or weeks, depending on the species. Among the desiccation-intolerant *I. scapularis* and *I. ricinus,* which are adapted to cool, humid forest habitats, desiccated individuals retreat to the forest floor or rotting vegetation at the base of a meadow. In these nearly saturated humid microenvironments, they can restore their water balance by a process known as atmospheric sorption, in which the partially desiccated ticks salivate salt-rich secretions onto their hypostomes. This hygroscopic secretion collects moisture, which is sucked back into the body. Since, however, the process demands a considerable expenditure of energy, the number of cycles of desiccation and sorption is limited as the tick's age. Other species, such as the relatively desiccation-tolerant *Hyalomma asiasticum,* can survive for long periods in the semidesert habitats in central Asia, where it waits for passing camels and other large ungulates. Nidicolous ticks, sheltering in caves, burrows, or other protected microenvironments, are subject to less stressful conditions during the long wait between hosts. These ticks exhibit behavioral patterns that restrict their distribution to these sheltered locations.

REPRESENTATIVE TICK-BORNE DISEASES

In view of the exceptionally large variety of diseases caused by tick-borne pathogens and injurious substances, this section is limited to a brief description of several representative examples, with primary emphasis on development of the infection in the tick and tick vector ecology. For a more extensive review, the reader may wish to consult books by Sonenshine and Strickland, or review articles, for information on the specific diseases. Table I lists some representative tick-borne diseases affecting humans and animals.

Lyme Disease

The most common tick-borne disease affecting human health in the world today, LD occurs throughout most of the United States and southern Canada, Europe and northern Asia. The disease is caused by *B. burgdorferi (sensu latu),* a type of bacterium known as a spirochete. *B. burgdorferi* is the causative agent of LD in most of the United States. A second genospecies, *B. lonestari,* was isolated from lone star ticks in the southeastern United States, but the relationship of these spirochetes to LD in humans is uncertain. In Europe, LD is caused by *B. afzellii* and *B. garinii* as well as *B. burgdorferi.*

In humans, LD results from the bite of an infected tick, either a nymph or adult of the genus *Ixodes.* The tick must have remained attached for several days to allow for the bacteria to travel from the tick's midgut to its salivary glands and into the wound site. Symptoms begin several days to several weeks later. Onset of illness is characterized by mild, flulike fever and, in most patients, a reddish skin rash, known as the erythema migrans (EM). The typical EM rash is a gradually expanding circular or elliptical lesion with a red margin and clear center, at least 5 cm or more in diameter, and often near the site of the tick bite. Some patients show multiple EM rashes. If left untreated in this early stage, the fever abates, the rash fades, and the patient may recover without any further symptoms. Often, however, the bacteria remain in people's bodies and spread into the nervous system and joints, where they cause the long-lasting secondary symptoms of chronic LD. Late manifestations include

TABLE I Representative Tick-Borne Diseases, Their Causative Agents, Tick Vectors, and Reservoir Hosts

Disease	Causative agent	Primary tick vector species	Affected hosts	Major clinical symptoms
Protozoan				
Human babesiosis	*Babesia microti, B. divergens, B. gibsoni*	*Ixodes scapularis, I. ricinus*	Humans	Malaria-like fevers, myalgia, arthralgia, nausea, sweating
Bovine babesiosis	*B. bigemina*	*Boophilus annulatus, B. microplus,* others	Cattle	Hemoglobinuria (redwater) fever, death
East Coast fever	*Theileria parva*	*Rhipicephalus appendiculatus*	Cattle, buffalo	Fever, lymphadenopathy, pulmonary edema
Tropical theileriosis	*T. annulata*	*Hyalomma anatolicum*	Cattle, horses	Fever, lymphodenopathy, pulmonary edema
Feline cytauxzoonosis	*Cytauxzoon felis*	*Dermacentor variabilis*	Cats	Fever, emaciation, splenomegaly, death
Bacterial, extracellular				
Tularemia	*Francisella tularensis*	*Haemaphysalis leporispalustria,* other tick species	Humans, various other mammals	Fever, headache, pustular, ulcerated papulae; pneumonia, pleuritis, rash; however few deaths
Lyme disease	*Borrelia burgdorferi, B. afzelii, B. garinii, I. persulcatus,* others	*Ixodes scapularis, I. ricinus, I. pacificus,*	Humans, dogs, cats, domestic animals	Initial phase: fever, EM rash Chronic phase: arthritis, neurologic symptoms

(*Continues*)

TABLE I *(Continued)*

Disease	Causative agent	Primary tick vector species	Affected hosts	Major clinical symptoms
Tick-borne relapsing fever	*Borrelia* spp.	*Ornithodoros* spp.	Humans	Intermittent fevers, chills, fatigue, myalgia, arthralgia; generally mild illness; death rare
Avian spirochetosis	*B. anserina*	*Argas persicus*	Birds	Fever, death
Epizootic bovine abortion abortion	Unknown, possibly *B. coriaceae*	*Ornithodoros coriaceus*	Cattle, deer	Fever, spontaneous abortion
Bacterial, intracellular (Rickettsiales)				
Rocky Mountain fever	*Rickettsia rickettsii*	*Dermacentar variabilis, D. andersoni,* others	Humans	High fevers, spotted whole-body rash
Boutonneuse fever[a]	*Rickettsia conorii*	*R. sanguineus, D. reticulatus,* others	Humans	High fever, rash, ulceration at bite site (eschar)
Human monocytic ehrlichiosis (HME)	*Ehrlichia chaffeensis*	*Amblyomma americanum, D. variabilis*	Humans	Fever, rash (sometimes), muscle aches, joint aches
Human granulocytic ehrlichiosis (HGE)	*Ehrlichia phagocytophilia*	*Ixodes scapularis, I. pacificus, I. ricinus*	Humans	Fever, rash (sometimes), muscle aches, joint aches
Tick-borne fever (sheep pyemia)	*E. phagocytophilia*	*I. ricinus*	Sheep	Fever, weight loss, reduced milk production, abortion
Canine ehrlichiosis	*E. canis, E. ewingli, E. phagocytophilia*	*R. sanguineus, I. ricinus, A. americanum,* others	Dogs	Fever, loss of appetite, weight loss, apathy, death in severe cases
Heartwater	*Cowdria ruminantium*	*Amblyomma hebraeum, A. variegatum,* others	Ruminants	Fever, "pedaling behavior", prostration, coma, death
Anaplasmosis	*Anaplasmosa marginale, A. centrale, A. ovis*	*D. andersoni, D. occidentalis, R. sanguineus,* others	Cattle, sheep, other ruminants	Fever, anemia, death
Q fever	*Coxiella burnettii*	Many tick species	Humans, large domestic livestock	Low-grade fever, sweating, sore throat, pneumonia, severe frontal headache, myalgia, photophobia
Arborviruses				
Tick-borne encephalitis	*Flavivirus*[b]	*I. ricinus, I. persulcatus*	Humans, carnivores	Fever, headache, encephalitis, meningitis, paralysis; death in severe cases
Powassan encephalitis	*Flavivirus*[c]	*Ixodes, Dermacentor, Haemaphysalis* spp.	Rodents, hares, etc.	Fever, headache, encephalitis, neurological symptoms, brain damage, death
Colorado tick fever	*Coltivirus*[c]	*D. andersoni*	Rodents, humans, domestic animals	Biphasic fever, headache, muscle aches, joint pain
Crimean–Congo hemorrhagic fever	*Nairovirus*[d]	*Hyalomma m. marginatum, H. m. rufipes,* others	Hares, humans, small mammals, others	Fever, chills, headache, internal bleeding, rashes; death in severe cases
Louping ill	*Flavivirus*[b]	*Ixodes ricinus*	Sheep	Fever, erratic, louping gait, loss of motor encephalitis, death
African swine fever	*Iridovirus*	*Ornithodoros moubata porcinus, O. erraticus*	Domestic pigs, wild boars, warthogs	Fever, internal damage; death in most cases
Tick-caused diseases				
Tick paralysis	Tick proteins	*I. holocyclus, I. rubicundus, D. variabilis, D. andersoni,* others	Cattle, sheep, humans, other mammals	Ascending paralysis, loss of motor control, no fever; death
Tick bite allergies	Tick proteins	*Argas reflexus, O. cariaceus, I. pacificus,* etc.	Humans	Nausea, vomiting, diarrhea, irregular pulse, shocklike symptoms; rarely death
Sweating sickness (and other tick toxicosis)	Tick proteins	*H. truncatum, O. savignyi, O. lahorensis, A. persicus*	Cattle, sheep, others	Fever, sweating, anorexia, tearing, salivation; high mortality

[a]Also known as Mediterranean spotted fever.
[b]Family Flaviviridae.
[c]Family Reoviridae.
[d]Family Bungaviridae.

arthritis, especially in the knee joints, which frequently spreads to different joints (migratory polyarthritis). Damage to the synovial membranes is a distinctive feature of this type of arthritis because of invasion of the synovial fluid by the spirochetes. Neurological symptoms include nerve pain (peripheral neuropathy), various types of palsy caused by nerve damage (e.g., Bell's palsy), and central nervous system disorders. Many patients suffering from some or all of these symptoms simultaneously also complain of severe fatigue. In Europe, another chronic feature of late LD is acrodermatitis chronica atrophicans, a condition in which the skin atrophies and peels. Chronic LD is greatly feared in regions where it is endemic because, among other reasons, it can persist for many years, despite treatment with antibiotics.

Only a limited variety of ticks of the genus *Ixodes* are competent vectors of LD spirochetes. Other biting arthropods may acquire these bacteria, but they cannot transmit the bacteria when they feed again. In the eastern and central United States, the only proven vector to humans is *I. scapularis.* In the western part of the country the vector is the western blacklegged tick *I. pacificus.* However, other species of the so-called *I. ricinus* complex that feed solely on wildlife contribute to maintaining the disease in nature; examples include *I. spinipalpis (I. neotomae)* in the west and *I. dentatus* in the east. The sheep tick, *I. ricinus,* is the primary vector in Europe and western Asia, while the taiga tick, *I. persulcatus,* is the primary vector from eastern Europe across most of northern Asia. Other species of the *I. ricinus* complex (e.g., *I. ovatus*), also serve as efficient enzootic vectors.

The cycle of *B. burgdorferi* development in the tick begins with ingestion of an infectious blood meal by the larvae as they feed on a mouse or other reservoir host. Spirochetes survive in the midgut diverticula but are not disseminated until the tick molts to the nymphal stage and feeds again. Following the influx of fresh host blood, the spirochetes surviving in the midgut pass between the cells of the gut wall and disseminate to the salivary glands and other internal organs. However, in females, spirochetal invasion of the ovary damages the developing oocytes, which prevents efficient transovarial transmission. As a result, transmission is by the transstadial route. As the tick feeds, spirochetes escape with the saliva and are introduced to the wound site of the new host. Transmission from nymph to adult also is common.

In the eastern United States, the risk of acquiring LD is greatest in the late spring or early summer, because of the peculiar nature of the LD cycle in nature. Nymphal ticks infected with *B. burgdorferi* emerge from their overwintering diapause to attack hosts, especially white-footed mice and other small mammals, as well as humans. This feeding pattern ensures the presence soon afterwards of numerous infected hosts when the larvae emerge, spreading the *B. burgdorferi* infection to vast numbers of these tiny ticks. Fed larvae molt in late summer or early fall, but the resultant nymphs diapause rather than commence feeding. Thus, the infection is perpetuated from one year to the next. Meanwhile, adults that emerge from fed nymphs delay feeding until the fall, whereupon they seek white-tailed deer and other large mammals (including humans). This pattern results in a 2-year cycle throughout most of the tick's geographic range.

Although the vector ticks are opportunistic feeders and will attack most vertebrates, only a limited variety of hosts are competent reservoirs of *B. burgdorferi.* Only competent reservoirs, animals capable of maintaining spirochetes within their tissues for prolonged periods, can infect ticks that feed on them. Examples of competent reservoirs include the white-footed mouse, the dusky-footed wood rat, the California kangaroo rat, and several other rodents in North America, and such animals as the common shrew, bank vole, wood mouse, yellow-necked field mouse, and hares and pheasants in Europe. Several species of birds also are reservoir competent and play an important enzootiologic role by dispersing infected ticks over considerable distances, thereby establishing new foci of infection. In contrast, some of the animals that are excellent hosts of vector ticks, such as white-tailed deer and western fence lizards, destroy invading spirochetes and, therefore, are not reservoirs. Such animals serve as amplifying hosts and are critically important for the expansion and spread of the tick populations, but they play no direct role in the perpetuation of the infection.

In the United States, LD has increased more than 1.7 times since it was first designated a reportable disease in 1991. By 1999, 16,273 cases that met the U.S. Centers for Disease Control and Prevention definition had been reported, for an overall incidence of 6.0 per 100,000. Most cases were reported from the northeastern, mid-Atlantic, and north central United States. Other important foci are in northern Wisconsin, northern Minnesota, and northern California. In Europe, important foci occur in the Scandinavian countries and in Germany, Poland, the Czech Republic, and Russia.

Domestic animals are also susceptible to infection with *Borrelia* pathogens. Dogs, cats, cattle, horses, and possibly other livestock and companion animals were found to be infected with *B. burgdorferi.* High seroprevalence rates occur in hyperendemic areas such as the northeastern United States, where many dogs show typical symptoms of chronic LD, especially lameness in one or more legs, fever, and fatigue.

Rocky Mountain Spotted Fever

Rocky Mountain spotted fever (RMSF) occurs throughout almost the entire United States, southern Canada, and Mexico and, to a lesser extent, in South America. This disease is caused by a tiny intracellular bacterium, the rickettsia *Rickettsia rickettsii.* People become ill with RMSF following the bite of an infected adult tick. Once in the human host, the rickettsia multiply profusely in the epithelial linings of the capillaries, arterioles, and venules. Vessels hemorrhage which, in the dermis of the skin, leads to the characteristic red spots. These innumerable reddish lesions, raised above the skin surface (maculopapular rash), coalesce to form the characteristic spotted

rash. Patients also develop high fever, severe headaches, nausea, joint and muscle pain, photophobia, and other symptoms. Unless treated with antibiotics, some patients die and others suffer irreversible injury.

In the United States and southern Canada, the primary vectors of RMSF are the Rocky Mountain wood tick, *Dermacentor andersoni,* in the west and *D. variabilis* in the east. Adults of these tick species readily attack humans. Other tick species (e.g., the rabbit ticks *Haemaphysalis leporispalustris* and *Ixodes dentatus*) transmit the rickettsia when they feed on rabbits and birds, thereby contributing to the maintenance of the zoonosis in nature, but these latter species do not bite humans. The disease survives in the natural environment in overwintering (i.e., diapausing) ticks, but not in the reservoir hosts. When rickettsia-infected larvae emerge from diapause in the spring, they infect mice and other susceptible rodents on which they feed. Other, uninfected larvae feeding on the rickettsemic animals acquire the infection, and the disease spreads rapidly in the tick population. Nymphal ticks spread the infection further as they feed again on other small mammals. Adult ticks seek larger animals (e.g., dogs, raccoons, etc.) as well as humans. Rickettsia are passed to subsequent generations of ticks by transovarial transmission. Ticks also harbor nonpathogenic species such as *R. montana, R. belli, R. rhipicephali, R. parkeri,* and other as yet unnamed rickettsia, complicating attempts to measure the incidence of RMSF in nature.

In the United States, about 600 to 800 cases of RMSF have been reported yearly since 1985, with an estimated annual incidence between 0.24 to 0.32 per 100,000 population. Most cases now occur east of the Mississippi River, with the highest concentration in the south-central and southeastern states, especially along the Atlantic coast. Cases tend to occur in foci in rural areas and suburban communities near major population centers. Since RMSF is a seasonal disease, the frequency of cases follows the seasonal activity pattern of the adults, with highest frequency in July and August in the southern United States, but greater frequency in May and June in the northeastern part of the country.

A closely related disease, boutonneuse fever (Mediterranean spotted fever), caused by *R. conorii,* occurs in southern Europe, North Africa, and Asia. These rickettsia are transmitted to humans by the bite of the brown dog tick, *Rhipicephalus sanguineus.* Although generally similar in its symptoms to RMSF, this illness is distinguished by the formation of a black ulcer, the eschar, at the wound site where the infected tick had attached.

Ehrlichiosis

Another rickettsial disease that has emerged into increasing prominence in recent years is ehrlichiosis. The causative agents are known as ehrlichiae, tiny obligate intracellular organisms that invade the white blood cells of humans and animals. Ehrlichiae develop within cytoplasmic vacuoles in different blood cell types such as monocytes, granulocytes, lymphocytes, or even platelets, depending on the species. Patients develop an acute illness with high fever and severe headaches, as well as aching muscles and joints. In contrast to RMSF, a rash is not common (20–30% of patients) and usually does not involve the palms or soles. Although severe cases can occur and may result in death, most cases are relatively mild. In the southeastern United States, most cases are caused by *Ehrlichia chaffeensis,* which invades the monocytic leukocytes and causes leukopenia (abnormal loss of white blood cells) among other symptoms. This disease is now known as human monocytic ehrlichiosis (HME). The primary vector for *E. chaffeensis* is the lone star tick, *Amblyomma americanum.* These ehrlichiae develop primarily in the monocytes. Some cases are caused by *E. ewingii,* also transmitted by lone star ticks. In the northern and western United States, the disease is caused by a related ehrlichia, *E. phagocytophila (E. equi),* which also causes illness in livestock. However, like *E. ewingii,* these ehrlichiae prefer the granulocytic leukocytes and, for this reason, the disease is known as human granulocytic ehrlichiosis (HGE). The primary vector in the northern United States is *I. scapularis.* In California, the primary vector for humans is *I. pacificus.* In Europe, the pathogen is transmitted to humans by the bites of the *I. ricinus. E. phagocytophila,* believed to be the most widespread of the various human-infecting ehrlichiae, has been reported in many countries of Europe and northern Asia. Ehrlichiosis also affects dogs, sheep, and other animals, being caused by different ehrlichiae.

Tick-Borne Encephalitis

Tick-borne encephalitis (TBE) is caused by viruses of the family Flaviviridae. In Europe and northern Asia, the illness in humans is manifested by high, often biphasic, fever and headache, followed soon afterward by inflammation of the brain (encephalitis) and meninges (meningitis). Some patients develop muscle weakness or paralysis, especially in the right shoulder muscles. In the Far East, case fatality rates are relatively high (up to 54%), whereas the disease in Europe is considerably milder. In Europe and the Far East, TBE viruses are transmitted by *I. ricinus* and by the taiga tick, *I. persulcatus.* Other tick species that feed on wild animals may amplify viral infection. Rodents and insectivores are the chief reservoir hosts. The disease is endemic in central, northern, and eastern Europe, where thousands of cases are reported each year. In India, a similar disease known as Kyasanur Forest disease has been responsible for numerous cases of human illness. In North America, a *Flavirus* of the TBE complex causes a disease known Powassan encephalitis, a serious illness that has caused death in some patients and permanent nerve damage in others. The primary vector to humans in the east is a little-known tick, *I. cookei,* a common parasite of woodchucks, foxes, and raccoons; in the west it is transmitted by the *D. andersoni.*

Babesiosis and Theileriosis

Two genera of tick-borne protozoa, *Babesia* and *Theileria,* cause severe illness and even death in domestic livestock and wildlife throughout most of the world. Tick-borne babesiosis also can cause illness in humans. *B. bigemina, B. bovis,* and *B. divergens* are examples of tick-borne protozoans that can cause babesiosis, a serious and often fatal illness in cattle prevalent in Mexico, parts of South America, Africa, and elsewhere. The clinical course of the disease is characterized by high fever and dehydration. Sick animals stop feeding, become lethargic, and show labored breathing. Anemia and bloody urine (so-called redwater fever) is a consequence of the massive destruction of red blood cells. Although some animals recover naturally, most of the diseased individuals gradually become comatose and die. Mortality estimates for infections with *B. bigemina* are 30%; for *B. bovis,* 70 to 80%. The causative agents, which are introduced by the bites of infected ticks, develop as meronts in the erythrocytes and multiply by multiple fission or, in the case of *B. bigemina,* by binary fission. Subsequently, huge numbers of erythrocytes are destroyed, leading to the symptoms noted. Eventually, some parasites invade leukocytes, where they are transformed into piroplasms, the stage that will infect ticks. Piroplasms remain in an arrested state of development until ingested by a suitable tick vector. In the tick's digestive tract, the erythrocytes and leukocytes of the host blood are lysed, releasing the parasites. The meronts are destroyed, but the piroplasms develop into ray-shaped gamonts (strahlenkorpers) that fuse with other, similar gamonts to form the diploid zygotes. These parasites are transformed into cell-penetrating ookinetes that traverse the tick's midgut epithelium and seek out the ovaries. In female *Boophilus* ticks, numerous oocytes are invaded by these ookinetes. When the ticks mate and lay eggs, the ookinetes proliferate in the developing embryos. Eventually, as the embryos mature, the ookinetes move to the salivary glands of the young larvae, where they form the invasive stage known as sporozoites, ready to infect cattle. Thus, a single blood meal in a female tick can result in thousands of disease-infected larval progeny, a phenomenon known as transovarial transmission. As a result, numerous animals or even an entire herd of cattle can be infected in a single tick generation.

Human babesiosis is a malaria-like illness caused by *B. microti, B. gibsoni, B. major,* or *B. divergens.* In the northeastern and midwestern United States, the dominant pathogen afflicting humans is *B. microti,* transmitted by the bite of *I. scapularis.* In Europe, human babesiosis may be caused by *B. microti,* or *B. divergens,* transmitted by *I. ricinus* and/or other ixodids. Although the life cycle in the human body is similar to that of the bovine babesias, development in the tick is quite different. When blood infected with *B. microti* is ingested by immature deer ticks, the invasive stages of the parasite transform into cell-penetrating sporokinetes, which, aided by a cell-piercing arrowhead organelle, traverse the midgut epithelium and find their way to the salivary glands. When the fed ticks molt to the next feeding stage and attack new human or animal hosts, the sporozoites are injected with the tick's saliva.

Babesiosis also afflicts other animals. In dogs, *B. canis,* transmitted by the bites of the *R. sanguineus,* causes severe illness and may be fatal. Other babesias cause illness in horses, sheep, and other animals.

Protozoan parasites closely related to the babesias, *Theileria parva* and *T. annulata* cause serious or fatal illness in cattle and other ungulates. *T. parva,* transmitted by the bites of the African brown ear tick, *Rhipicephalus appendiculatus,* causes a disease in southern Africa known as East Coast fever. *T. annulata,* transmitted by the bites of ticks of the genus *Hyalomma,* causes a disease known as tropical theileriosis. It is much more widespread, ranging from North Africa across western and central Asia to the Indian subcontinent. The life cycle of both protozoan parasites is generally similar to that of the babesias, but with several notable exceptions. *Theileria* parasites multiply in the leukocytes instead of the erythrocytes. Following destruction of the host cells, the *Theileria* progeny infect other leukocytes, thereby damaging the host's immune system, while others invade erythrocytes to form piroplasms. The latter are the stage that will infect ticks. These piroplasms remain in an arrested state of development until ingested by the tick vector. When vector-competent ticks feed on *Theileria*-infected cattle, the piroplasms are released from the disintegrating erythrocytes and transform into gamonts similar to those seen in the babesias. Following gamont fusion and fertilization, the zygotes develop into motile kinetes that cross the midgut epithelium and travel to the salivary glands. Because only the salivary glands are infected, pathogen transmission occurs only when the infected larval or nymphal ticks molt and feed again in the next life stage on susceptible hosts (transstadial transmission).

Tick Paralysis and Tick Bite Allergies

In addition to transmission of infectious microbes, ticks may cause paralysis, allergies, and severe toxic reactions in their hosts. In humans, this affliction is manifested by a gradually ascending paralysis, beginning with the loss of sensation and motor coordination in the legs, abdominal muscles, back muscles and, eventually, the diaphragm and intercostal muscles. Unless the tick responsible for these symptoms is found and removed, the patient will undergo respiratory failure and die. Most patients recover rapidly after tick removal, but those with advanced symptoms may require several weeks for complete recovery.

Tick paralysis also affects animals, especially large ungulates. In Australia, a single *Ixodes holocyclus* female may be sufficient to cause the death of a full-grown cow or bull. Thousands of cattle died in the western United States and Canada because of paralysis induced by the bites of *D. andersoni.* Other than finding and removing all the feeding female ticks, there is no known cure. Some ticks also cause severe allergic and toxic reactions. A notable example is the African tampan *Ornithodoros savignyi,* the bite of which may cause cattle to salivate, tremble, gnash their teeth, become disoriented, and

even die. A similar condition, known as sweating sickness, is caused by ticks of the genus *Hyalomma,* especially *H. truncatum.* The illness is characterized by fever, loss of appetite, sweating, lachrymation, salivation and, usually, death. However, there is no paralysis.

CONTROL

Although difficult to estimate in precise monetary figures, the economic importance of ticks is enormous. This is primarily because of losses caused by tick-transmitted diseases affecting livestock. In addition, ticks cause injury from severe blood loss, reduced milk production, and damage to the hides. Worldwide, losses to the cattle industry alone were estimated to be more than $7 billion. Perhaps the most important economically are the cattle ticks, especially *Boophilus annulatus, B. decoloratus,* and *B. microplus,* because of the disease agents they transmit (*Babesia* spp.). Other economically important species are the bont ticks, *Amblyomma variegatum* and *A. hebraeum,* which are vectors of the causative agent of heartwater, *Cowdria ruminantium.*

Tick control is done almost exclusively with acaricides: poisonous compounds such as chlorinated hydrocarbons (e.g., DDT), organophosphorus compounds (e.g., diazinon), carbamates (e.g., carbaryl), and pyrethroids (e.g., permethrin, flumethrin). Animals are treated with acaricides by either bathing or spraying them with mixtures of liquids containing the specific toxicant. Acaricides also can be delivered as pour-ons or spot-ons, highly concentrated mixtures containing surfactants (i.e., spreading agents) that enable the product to disperse naturally throughout the animal's hair coat or as dusts, consisting of mixtures of talc and the acaricide. For long-lasting effects, acaricides are incorporated into plastic collars, ear tags, or even tail tags, which gradually emit the toxic chemicals over a long period of time. Also promising is the tick decoy, in which tail tags are impregnated with the tick's natural pheromone as well as the acaricide. These "tick decoys" demonstrated excellent efficacy (> 95%) for up to 3 months when applied to cattle in Zimbabwe. Other promising innovations for treating tick-infested animals include self-medicating devices, such as the "four-poster" developed by the U.S. Department of Agriculture for treating white-tailed deer, and antitick vaccines. Self-medicating applicators release an acaricide from a reservoir when host animals insert their heads into the device to retrieve corn or other bait. This is perhaps the first practical method that has been developed for controlling ticks on wild animals without killing or injuring them. Antitick vaccines work by immunizing livestock animals such as beef cattle against proteins in the tick's digestive tract, thereby killing the ticks as they attempt to feed. Further scientific advances may make it possible to reduce the current dependence on toxic chemicals to control ticks on domestic animals, pets and even wildlife.

Hikers, campers, and other people entering tick-infested habitats should take appropriate precautions to avoid ticks and the risk of acquiring tick-borne diseases. Each person should wear long trousers, tuck the trouser ends into boots or shoes, and draw the socks over the trousers. The protective barrier should be secured with tape and treated with a repellent—for example, one of the many products that contain diethyl toluamide (DEET). Upon returning from the trip, each person should decontaminate the field clothes and carefully inspect skin and hair for ticks that may have attached unnoticed. Early detection and removal of attached ticks will minimize the risk of transmission of disease-causing agents. Unattached ticks, if any, may be destroyed and discarded. However, attached ticks should be removed carefully with tweezers (forceps) and retained intact if needed for future examination in case of illness.

See Also the Following Articles

Arthropoda and Related Groups • Medical Entomology • Mites • Veterinary Entomology • Zoonoses, Arthropod-Borne

Further Reading

Anonymous. (2001). Lyme disease—United States, 1999. *MMWR* **50,** 181–185.

Hoogstraal, H., and Aeschlimann, A. (1982). Tick–host specificity. *Bull. Entomol. Suisse* **55,** 135–238.

Jaenson, T. G. T. (1991). The epidemiology of Lyme borreliosis. *Parasitol. Today* **7,** 39–45.

Klompen, H., and Grimaldi, D. (2001). First Mesozoic record of a parasitiform mite: A larval argasid tick in Cretaceous amber (Acari: Ixodida: Argasidae). *Ann. Entomol. Soc. Am.* **94(1),** 10–15.

Lane, R. S., Piesman, J., and Burgdorfer, W. (1991). Lyme borreliosis: Relation of its causative agent to its vectors and hosts in North America and Europe. *Annu. Rev. Entomol.* **36,** 587–609.

Needham, G. R., and Teel, P. D. (1991). Off-host physiological ecology of ixodid ticks. *Annu. Rev. Entomol.* **36,** 659–681.

Norval, R. A. I., Sonenshine, D. E., Allan, S. A., and Burridge, M. J. (1996). Efficacy of pheromone-acaricide impregnated tail-tag decoys for control of bont ticks, *Amblyomma hebraeum* on cattle in Zimbabwe. *Exp. Appl. Acarol.* **20,** 31–46.

Pound, J. M., Miller, J. A., George, J. E., and Lemeilleur, C. A. (2000). The '4-poster' passive topical treatment device to apply acaricide for controlling ticks (Acari: Ixodidae) feeding on white-tailed deer. *J. Med. Entomol.* **37,** 588–594.

Ribeiro, J. M. C. (1987). Role of saliva in blood feeding by arthropods. *Annu. Rev. Entomol.* **32,** 463–478.

Sonenshine, D. E. (1993). "Biology of Ticks." Oxford University Press, New York.

Sonenshine, D. E., Hamilton, J. G., Phillips, J. S., and Lusby, W. R. (1991). Mounting sex pheromone: its role in regulation of mate recognition in the ixodidae. *In* "Modern Acarology," Vol. 1 (F. Dusbabek and V. Bukva, eds.), pp. 71–78. Academia, Prague, and SPB Publishing, the Hague.

Strickland, G. T. (ed.). (2000). "Hunter's Tropical Medicine and Emerging Infectious Diseases." Saunders, Philadelphia.

Telford, S. R., III, and Spielman, A. (1989). Competence of a rabbit-feeding *Ixodes* (Acari: Ixodidae) as a vector of the Lyme disease spirochete. *J. Med. Entomol.* **26,** 118–121.

Touch

see *Mechanoreception*

Tracheal System

Jon F. Harrison
Arizona State University

In contrast to most vertebrates, crustaceans, mollusks, and many types of worm, insects do not use respiratory pigments in blood to transport oxygen. Instead, insects have a tracheal respiratory system in which oxygen and carbon dioxide travel primarily through air-filled tubes called tracheae. Usually the tracheal system penetrates the cuticle via closable valves called spiracles and ends near or within the tissues in tiny tubes called tracheoles (Fig. 1). The tracheae primarily serve as pipes that transport gases between the spiracles and the tracheoles, whereas the thin-walled tracheoles are thought to be the main sites of gas exchange with the tissues. In many insects, dilations of the tracheae form thin-walled air sacs or tracheal lungs that serve as bellows for enhancing the flow of gases through the tracheal system or directly deliver oxygen to tissues. The organization of the tracheal system varies dramatically among insects, with spiracle number ranging from zero to 20 and with tracheal branching patterns varying widely across species, between body regions, and during the developmental stages of holometabolous insects.

SPIRACLES

Spiracles are the openings of the tracheal system on the integument of the insect. A few apterygote insects lack valves in their spiracles and therefore have tracheae that are always open to the environment. However, spiracles generally have valvelike structures that allow them to close (Fig. 2). Spiracles range tremendously in area from 27 mm^2 for the second spiracle of the giant *Petrognatha* beetle to openings of only a few micrometers. Often the outside of the spiracle is fringed with hairs or covered with a filter to help resist entry of dust, water, or parasites.

The valve that closes the spiracle takes various forms but can appear as two lips of cuticle or as a bar of cuticle that pinches the trachea shut (Fig. 2). The movement of the valve usually depends on the action of muscles that insert on or near it, and elastic elements in the cuticle. Some thoracic spiracles have only a closer muscle, and elastic elements tend to open the spiracular valve (Fig 2A). Often the valve is recessed behind the atrium, an enclosed space behind the external opening (Fig. 2B). The most common situation (true for all abdominal spiracles studied to date) is for the spiracle to be recessed behind an atrium and to have both opener and closer muscles (Fig. 2B).

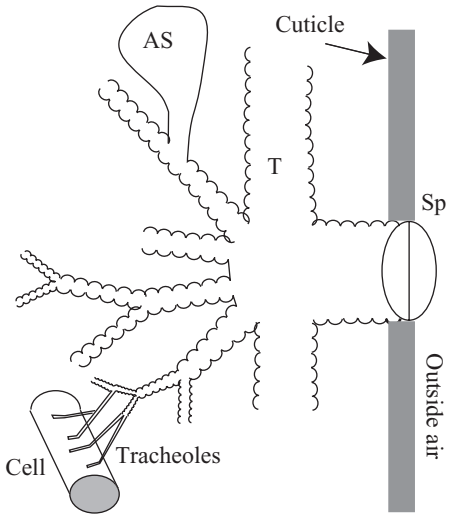

FIGURE 1 Overview of the tracheal system. Spiracles (Sp) in the integument connect to tracheae (T), which branch repeatedly within the insect, eventually leading to the tracheoles, which are located near most cells. In some insects, soft, collapsible air sacs (AS) occur.

TRACHEAE

Tracheae, the largest tubes of the tracheal system, range in diameter from a few millimeters to about 1 μm. The tracheal wall (Fig. 3A) contains an epithelial cell layer that secretes a

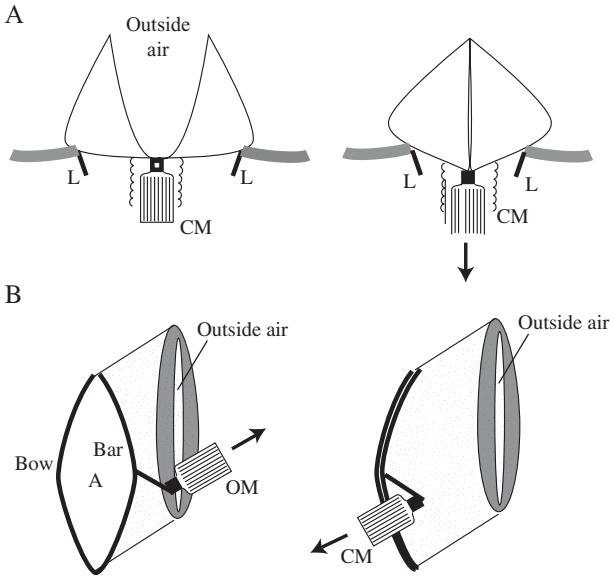

FIGURE 2 Spiracular structure and function. (A) Schematic side view of the locust metathoracic spiracle, a one-muscle spiracle with an external valve. [After Chapman, R. F. (1998). "The Insects: Structure and Function," Fig. 17.11c. Cambridge University Press, Cambridge, U.K., with permission of the publisher and the author.] When the closer muscle (CM) contracts, the lips of the valve come together. When the closer muscle relaxes, ligaments (L) attached to the cuticle passively pull the valve lips apart. (B) Schematic view from the inside of the animal of the abdominal spiracle of a butterfly, a two-muscle spiracle with an internal valve. [After Schmitz, A., and Wasserthal, L. T. (1999). Comparative morphology of the spiracles of the Papilionidae, Sphingidae, and Saturnicdae (Insecta: Lepidoptera). *Int. J. Insect Morphol.* **28,** 13–26, © 1999 with permission of Elsevier Science and the authors.] When the opener muscle (OM) contracts, it pulls on a lever that pulls the bar away from the bow, opening the atrium (A). When the closer muscle (CM) contracts, it pulls the bar against the bow, collapsing the walls of the atrium together.

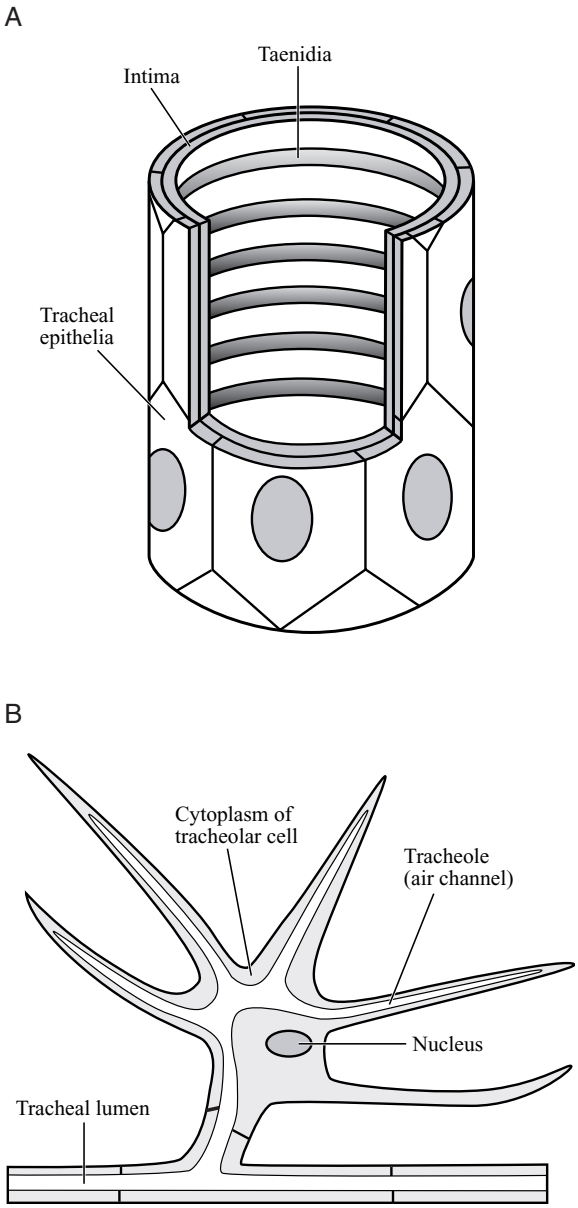

FIGURE 3 (A) Structure of a trachea. The basement membrane that encircles the tracheal epithelia has been omitted. (B) Tracheolar cell connected to a trachea.

basement membrane that forms the outermost layer of the tracheal wall. Internally, the tracheal cells secrete the intima, which contains protein and chitin fibers. The intima forms spiral folds known as taenidia. In general, it is thought that the taenidia prevent the trachea from collapsing inward, yet allow the trachea to flex. However, in the tracheae of some insects, the taenidia are widely spaced within the tracheal wall, and the intima is quite flexible, with the result that such tracheae are compressible. Because tracheal walls are relatively thick, oxygen is not believed to pass out of the larger tracheae in appreciable amounts. Thus, the primary function of the trachea is to transport gases between the tracheoles and the spiracles. However,

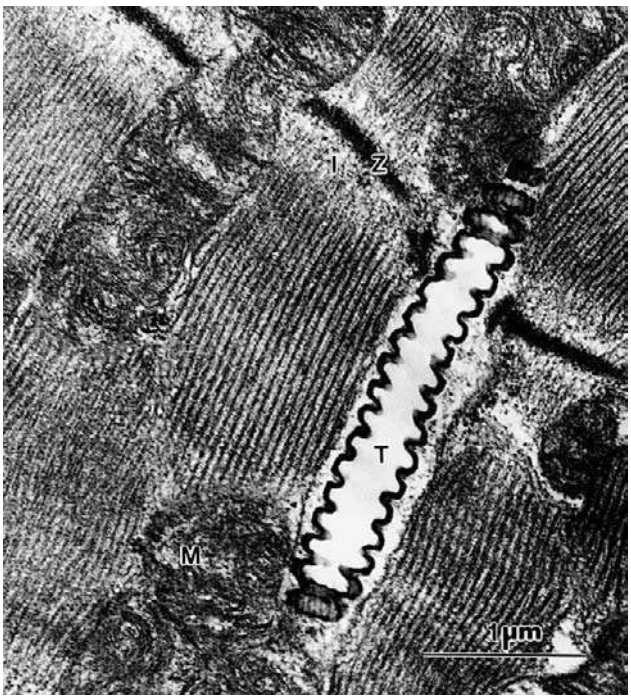

FIGURE 4 Electron micrograph showing a tracheole (T) within the flight muscle of an American locust. M, mitochondria; Z, Z line; I, I-band. Scale bar = 1 micron. (Image courtesy of Scott D. Kirkton, Arizona State University.)

because carbon dioxide can diffuse through tissue more easily than oxygen, it is possible that a significant portion of the carbon dioxide emitted may pass through the tracheal walls.

TRACHEOLES

Tracheoles, the small tubes that form the terminal endings of the tracheal system, range in size from 1 to 0.1 μm in diameter. Tracheoles are formed within single tracheolar cells, (Fig. 3B). The tracheolar cells have many branching processes, some of which contain an air-filled channel (the tracheole) that connects to the air-filled lumen of the trachea (Fig. 3B). Tracheole walls are capable of transporting oxygen at high rates by diffusion because they are thin (usually < 0.1 μm), and their ratio of surface area to volume is very large. Thus, it is likely that the tracheoles are the major site of gas exchanges between the tissues and the tracheal system.

Tracheoles are particularly dense in metabolically active tissues such as flight muscle. Most tracheoles occur outside the cells of the insect's body, but sometimes in histological sections they appear to be within cells, particularly in flight muscle (Fig. 4). These tracheoles are believed to enter the flight muscle via infoldings of the muscle plasma membrane, and so they are actually extracellular. The high tracheolar densities and penetration of flight muscle cells by tracheoles allows flying insects to achieve oxygen consumption rates that are among the highest in the animal kingdom.

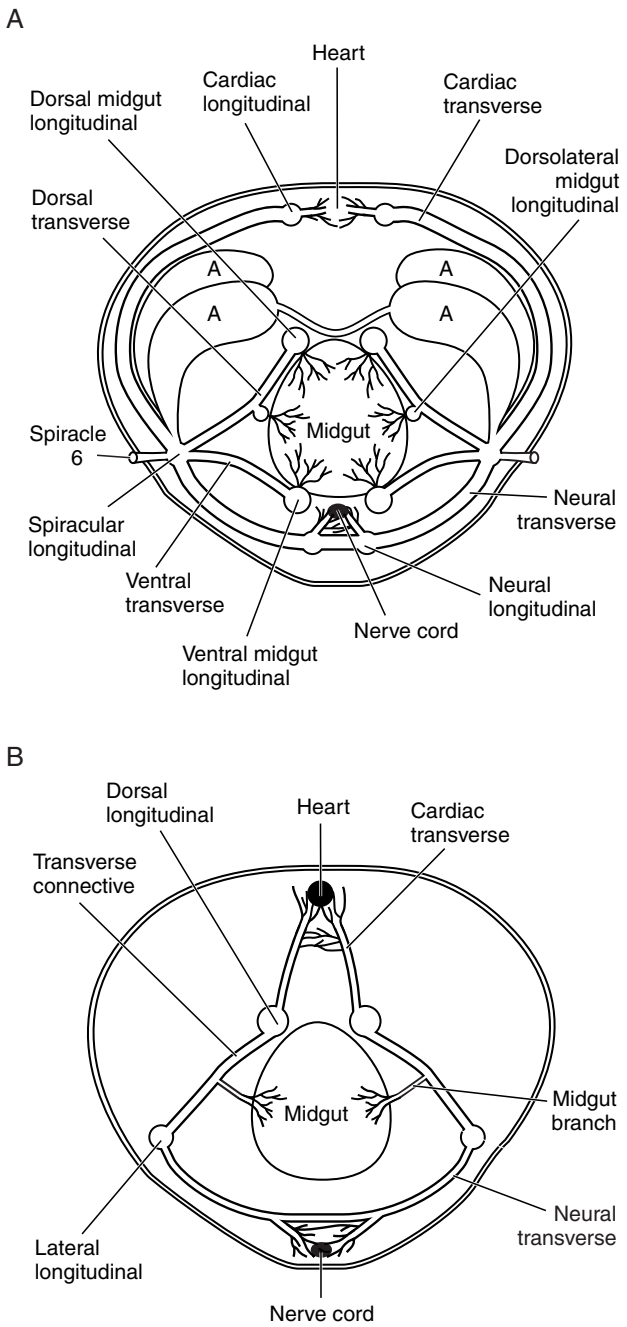

FIGURE 5 (A) Cross-sectional view of the tracheal system at the sixth abdominal segment of an adult grasshopper *(Schistocerca americana),* an insect with air sacs and 10 paired spiracles. (B) Cross-sectional view of the tracheal system at the sixth abdominal segment of an embryonic or larval fruit fly *(Drosophila melanogaster),* an insect without air sacs and with only the first and tenth spiracles functional. In both insects, longitudinal tracheal trunks run the length of the animal, connecting the thoracic and abdominal segments, but longitudinal trunks are much more numerous in the grasshopper.

AIR SACS, AERIFEROUS TRACHEAE, AND TRACHEAL LUNGS

In active insects, particularly those with a rigid cuticle, portions of the tracheal system are enlarged to form thin-walled air sacs. These air sacs lack taenidia and so collapse and expand with variation in hemolymph pressure. Air sacs are common in adult flying insects of many orders, including Diptera, Hymenoptera, Lepidoptera, Odonata, and Orthoptera. These air sacs serve as bellows to enhance the movement of air through the tracheal system. For insects that have them, air sacs often account for the majority of the volume of the tracheal system. Of insects that lack air sacs (such as cockroaches or Dictyoptera), many have slightly dilated, compressible tracheae that may also function as bellows.

Tracheal lungs, or aeriferous tracheae, are similar to air sacs in that they are thin-walled expansions of the tracheal system with few taenidia that float freely in the hemolymph. Their primary function is believed to be oxygen delivery to tissues that lack tracheoles, such as hematocytes, and some ovaries and Malpighian tubules. In the absence of tracheoles, the oxygen must diffuse through the thin wall of the tracheal lung, through the hemolymph, and to the tissues.

ORGANIZATION

The organization of the tracheal system varies substantially among insects, at least partially in correlation with species ecology. One obvious variant is spiracle number. In some aquatic insects, lack of spiracles prevents water entry to the tracheae. Many dipteran larvae have only one pair of spiracles on their abdominal tip, which they extend into the air as they feed head-down in water. In contrast, many terrestrial insects have a full 10 pairs of spiracles. Insects also vary in the number of air sacs, the location and size of tracheal branches, and the number of longitudinal tracheae.

A glimpse of the cross-species variety can be gained by comparing the organization of the major tracheae in the sixth abdominal segment of an adult grasshopper and a larval fruit fly (Fig. 5). Grasshoppers have a spiracle on every segment, whereas fly larvae have only functional spiracles on the head and abdominal tip. In the grasshopper, transverse tracheae lead from the spiracles to longitudinal tracheae that run along the heart, ventral nerve cord, and several positions on the gut. Air sacs are prominent. In contrast, in flies, two large dorsal longitudinal trunks connect the two functional spiracles on each side, with branches extending from the dorsal trunk to the tissues and one other longitudinal trachea. Air sacs are absent.

Tracheal organization also varies with location within the insect. In most insects, the head segments lack spiracles, and so tracheae usually extend forward from the prothoracic spiracles to supply the tissues in the head; similar elaborations of the tracheae occur in the posterior direction in the tenth segment. Structures that extend away from the main body of the insect, such as legs, antennae, and wings, also lack spiracles and usually have long tracheae extending from the spiracle of the local body segment. Antennae, legs, and wings all are known to have accessory pulsatile structures that pump hemolymph and help circulate air through their long tracheae. The tracheal organization in the thorax of flying insects is very different

from that in the abdomen, since the thoraces often contain extensive air sacs associated with the flight muscles.

Tracheal organization can also be strongly modified by developmental stage, especially in holometabolous insects. For example, many dipterans transform from aquatic larvae with one or two pairs of spiracles and no air sacs to adults with 10 pairs of spiracles and extensive air sacs, a transition that enables high oxygen delivery and flight. Clearly, flexibility in tracheal structure and function is a key trait in the success, adaptability, and diversity of insects.

See Also the Following Articles

Anatomy • Respiratory System

Further Reading

Chapman, R. F. (1998). "The Insects: Structure and Function." Cambridge University Press, Cambridge, U.K.

Locke, M. (1998). Caterpillars have evolved lungs for hemocyte gas exchange. *Journal of Insect Physiol.* **44,** 1–20.

Manning, G., and Krasnow, M. A. (1993). Development of the *Drosophila* tracheal system. "The Development of *Drosophila melanogaster*" (M. Bate and A. M. Arias, eds.). Cold Springs Harbor Laboratory Press, Plainview, NY.

Nikam, T. B., and Khole, V. V. (1989). "Insect Spiracular Systems." Ellis Horwood Entomology and Nematology, Chichester, U.K.

Pass, G. (2000). Accessory pulsatile organs: Evolutionary innovations in insects. *Annu. Rev. Entomol.* **45,** 495–518.

Schmitz, A., and Perry, S. F. (1999). Stereological determination of tracheal volume and diffusing capacity of the tracheal walls in the stick insect *Carausius morosus* (Phasmatodea, Lonchodidae). *Physiol. Biochem. Zool.* **72,** 205–218.

Schmitz, A., and Wasserthal, L. T. (1999). Comparative morphology of the spiracles of the Papilionidae, Sphingidae, and Saturniidae (Insecta: Lepidoptera). *Int. J. Insect Morphol.* **28,** 13–26.

Snodgrass, R. E. (1993). "Principles of Insect Morphology." Cornell University Press, Ithaca, NY.

Wasserthal, L. T. (1998). The open hemolymph system of Holometabola and its relation to the tracheal space. *In* "Microscopic Anatomy of Invertebrates," Vol. 11b, "Insecta," (F. W. Harrison and M. Locke, eds.), pp. 583–620. Wiley-Liss, New York.

Wigglesworth, V. B. (1983). The physiology of insect tracheoles. *Adv. Insect Physiol.* **17,** 85–149.

Trichoptera
(Caddisflies)

John C. Morse
Clemson University

Trichoptera, or caddisflies, are holometabolous insects closely related to Lepidoptera, or moths. However, unlike most moths, their eggs, larvae, and pupae are usually found in or very near fresh water, and adults are aerial, usually not far from their aquatic habitats (Fig. 1). The Trichoptera include more species than any of the other primarily aquatic orders of insects. This high species diversity is correlated with an unusually broad range of ecological specialization. Immature caddisflies occur in almost every type of freshwater habitat and on every continent except Antarctica; they are often one of the most abundant insect orders in streams and ponds. Larvae of many species use silk to construct portable cases of various shapes and materials to serve as physical protection, camou-

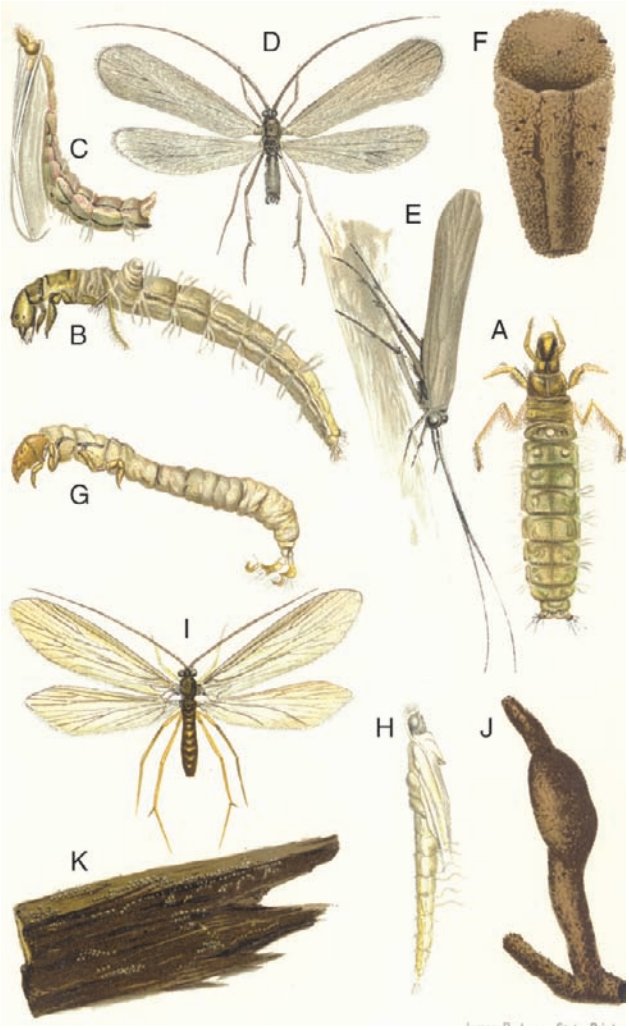

FIGURE 1 Caddisfly larvae, pupal, imagoes, cases, and eggs: (A) dorsal view of larva of *Molanna cinerea,* ×4; (B) lateral view of larva of *M. cinerea,* ×5; (C) lateral view of the pupa of *M. cinerea,* ×4.5; (D) dorsal view of imago of *M. cinerea,* ×4; (E) accustomed resting position of the imago of *M. cinerea;* (F) ventral view of the flat larval case of *M. cinerea,* ×2; (G) lateral view of larva of *Phylocentropus lucidus,* showing the very long anal prolegs and the absence of gill filaments, ×5; (H) lateral view of pupa of *P. lucidus,* ×6; (I) dorsal view of imago of *P. lucidus,* ×3.5; (J) larval case of *P. lucidus,* tube composed of sand and silk, two-layered, enlargement near the end containing the pupa; (K) eggs laid by *P. lucidus* female on a stick protruding from the water in a breeding cage. [From Betten, C. (1901). Order Trichoptera, caddisflies. "Aquatic Insects of the Adirondacks: A Study Conducted at the Entomologic Field Station, Saranac Inn, N.Y, under the Direction of Ephraim Porter Felt D.Sc. State Entomologist. (J. G. Needham and C. Betten, eds.) pp. 561–573 and plates 13,15, 30–34. *Bull. N.Y. State Mus.* **47,** 383–612, plates 1–36.]

TABLE I Extant Families of Trichoptera, with Primary Larval Feeding Strategies, Distributions, and Approximate Numbers of Named Species and Subspecies

Family	Larval feeding strategy[a]	Distribution[b]							Number of species[c]
		AT	AU	EP	NA	NT	OL	WP	
Annulipalpia									
Stenopsychidae	cf	X	X	X		X	X		93
Philopotamidae	cf	X	X	X	X	X	X	X	911
Hydropsychidae	cf, pe	X	X	X	X	X	X	X	1462
Dipseudopsidae	cf	X	X	X	X		X	X	165
Polycentropodidae	cf, pe	X	X	X	X	X	X	X	600
Ecnomidae	cf	X	X	X	X	X	X	X	345
Xiphocentronidae	cg	X		X	X	X	X	X	130
Psychomyiidae	cg	X	X	X	X		X	X	395
"Spicipalpia"									
Rhyacophilidae	pe	X	X	X	X	X	X	X	685
Hydrobiosidae	pe		X	X	X	X	X	X	375
Glossosomatidae	sc	X	X	X	X	X	X	X	522
Hydroptilidae	ph, sc, cg	X	X	X	X	X	X	X	1677
Integripalpia									
Oeconesidae	sd		X						19
Brachycentridae	cf, cg, sh			X	X		X	X	111
Phryganopsychidae	sd			X			X		3
Lepidostomatidae	sd	X	X	X	X	X	X	X	391
Kokiriidae	pe		X			X			8
Plectrotarsidae	sd		X						5
Phryganeidae	sh, pe			X	X		X	X	79
Goeridae	sc, cg	X	X	X	X		X	X	163
Uenoidae	sc, cg			X	X		X	X	79
Apataniidae	sc, cg			X	X		X	X	186
Limnephilidae	sd, cg, sc		X	X	X	X	X	X	873
Tasimiidae	sc		X			X			9
Odontoceridae	sd		X	X	X	X	X	X	106
Atriplectididae	sd	X	X			X			5
Philorheithridae	pe		X			X			21
Molannidae	sc, cg, pe			X	X		X	X	32
Calamoceratidae	sd, sc,	X	X	X	X	X	X	X	124
Leptoceridae	cg, sh, pe	X	X	X	X	X	X	X	1567
Sericostomatidae	sd	X		X	X	X	X	X	97
Beraeidae	cg	X		X	X			X	52
Anomalopsychidae	sc					X			22
Helicopsychidae	sc	X	X	X	X	X	X	X	197
Chathamiidae	sh		X						5
Helicophidae	cg, sh		X			X			21
Calocidae	cg		X						20
Conoesucidae	sc, sd, sh		X						42
Antipodoeciidae	sc		X						1
Barbarochthonidae	sh, sc	X							1
Hydrosalpingidae	sc, sd, sh	X							1
Limnocentropodidae	pe			X			X		15
Petrothrincidae	sc	X							6
Pisuliidae	sd	X							15
Rossianidae	sc, sh				X				2

[a]Abbreviations: cf, collector-filterers; cg, collector-gatherers: pe, predator-engulfers; ph, piercer-herbivores; pp, predator-piercers; sc, scrapers; sd, shredder-detritivores; sh, shredder-herbivores. (From Merritt and Cummins, 1996. Some larval feeding strategy data from M. Winterbourne, University of Canterbury, Christchurch, New Zealand, and F. C. de Moor, Albany Museum, Grahamstown, South Africa.)

[b]Abbreviations: AT, Afrotropical; AU, Australasian; EP, East Palearctic; NA, Nearctic; NT, Neotropical; OL, Oriental; WP, West Palearctic.

[c]The total number of these known species and subspecies is 11,638.

flage, or aids in respiration. Others make stationary retreats of silk for similar purposes or to serve as food-gathering structures. The variety of feeding strategies employed by caddisflies is as great as for the highly diverse freshwater Diptera, or true flies (Table I). Because of their diversity and density in most clean, freshwater ecosystems, the significance of caddisflies for processing nutrients and transferring energy is often great. Their importance as food for predators such as fish is emphasized by fly-fishing enthusiasts who imitate them with artificial lures. The different caddisfly species are variously sensitive to changes in environmental conditions, such that the diversity of the order is commonly used in part as a measure of pollution.

DIVERSITY AND PHYLOGENETIC RELATIONSHIPS OF TRICHOPTERA

More than 11,000 species of caddisflies have been described globally (Table I), with more than 1300 reported for America north of Mexico. The world species are in 504 genera in 45 families. The most species have been described in the micro-caddisfly family, Hydroptilidae, with nearly 1700 species. Other well-represented families globally include the long-horned caddisflies (Leptoceridae, > 1500 species), the common netspinner caddisflies (Hydropsychidae, > 1400 species), and the northern caddisflies (Limnephilidae, almost 900 species). The highest known species diversity and the greatest density of species occurs in the Oriental biogeographical region (> 3700 species, with 1.6 species per kilohectare).

The Trichoptera and Lepidoptera are generally considered to be sister orders, together constituting the Amphiesmenoptera, with earliest fossils dating from the Permian. The families of tube-case-making caddisflies are generally considered to constitute a monophyletic suborder Integripalpia and the net-spinning caddisfly families a monophyletic suborder Annulipalpia (Fig. 2). The relationships of the remaining caddisfly families (Glossosomatidae, Hydrobiosidae, Hydroptilidae, and Rhyacophilidae, collectively the "Spicipalpia") are yet unknown but are apparently near the base of the caddisfly phylogeny. The phylogenetic relationships among genera and species of caddisflies have been studied for most extant families, mostly with morphological characters visible with a light microscope. The phylogeny of Trichoptera species continues to be investigated at all categorical scales by Kjer and others, using both morphological and molecular evidence and computer-managed algorithms. These hypothetical relationships have proven useful in numerous comparative studies, suggesting hypotheses for research of case- and retreat-making behaviors, feeding strategies, historical biogeography, mating behaviors, and other aspects of caddisfly biology.

Triassic fossils of species of Philopotamidae, Prorhyacophilidae, and Necrotauliidae are thought to represent the oldest known Trichoptera. The family Philopotamidae includes modern species, but the other two families are extinct.

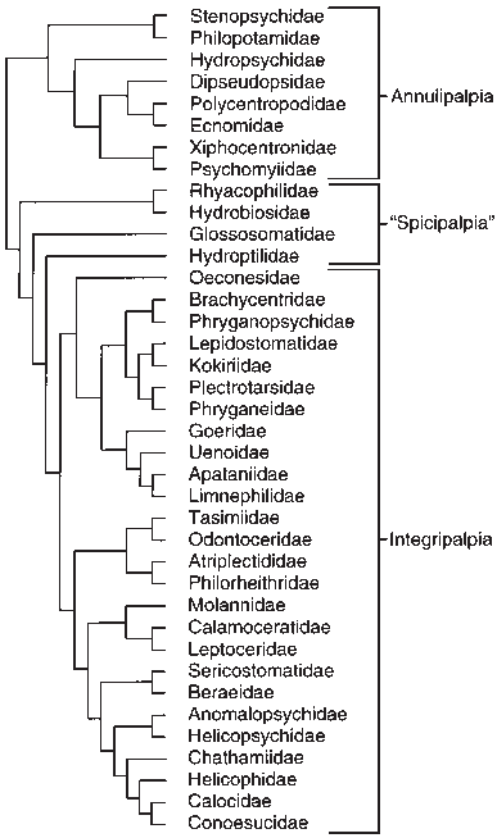

FIGURE 2 Phylogeny of extant families of Trichoptera; families not analyzed by Kjer *et al.* include Antipodoeciidae, Barbarochthonidae, Hydrosalpingidae, Limnocentropodidae, Petrothrincidae, Pisuliidae, and Rossianidae. [After Kjer, K. M., Blahnik, R. J., and Holzental, R. W. (2001). Phylogeny of Tricoptera (caddisflies): characterization of signal and noise within multiple data sets. *Syst. Biol.* **50(6),** 781–816.]

LIFE HISTORIES OF TRICHOPTERA

Life Cycles

Among the relatively few species of caddisflies for which life histories are known, most exhibit a univoltine, or one-year, life cycle, with some species having more than one generation per year and some with one generation every 2 or 3 years. In general, longer life cycles are found in the higher latitudes and shorter life cycles closer to the equator. A few species have been shown to have more than one cohort in the same locality, with different segments of the same population having their life cycles synchronized apart from the synchrony of one or more other segments. The time of the year during which particular phases of the life cycle occur is distinctive for the species, with most species reaching adulthood in temperate regions during the warmer months of the year.

Eggs

Eggs are embedded in a sticky, gelatinous polysaccharide called spumalin, forming an egg mass. Round or elliptical

eggs are generally laid in egg masses of particular shapes with 12 to 636 eggs reported per egg mass and with a single female depositing one to several egg masses. Egg masses may be spherical, pyramidal, butterfly shaped, or some other configuration, or they may be arranged in a flat spiral or as a string of beads. Because of the spumalin, little is known about the surface structure of caddisfly eggs. Eggs may diapause for several months before larvae hatch, especially during winter, but larvae generally hatch within a few days depending on temperature. If the egg mass is out of water, the first instars of some species may remain in the spumalin matrix until it is inundated; for other species whose egg masses are laid on plants or rocks overhanging the water, young larvae drip out of the spumalin into the water.

Larvae

Larvae usually undergo five instars before pupation, although a few species have a generally indeterminate number of instars greater than five. The larval stage usually is the longest stage of the life cycle, with development completed during a period of 2 months to nearly 2 years. The shape of the larva differs usually with the family or genus and depends in part on the habits and feeding strategy peculiar to that taxon. Eruciform larvae, typical of the Integripalpia, are cylindrical, with hypognathous mouthparts oriented at a right angle with the body axis and with the posterior end of the abdomen blunt and having thick anal prolegs fused with segment IX; these live in portable tubular cases and feed generally as shredders or scrapers or collector-gatherers.

Campodeiform larvae, typical of the Annulipalpia and unplaced primitive families, have a tapered shape anteriorly and posteriorly, with prognathous mouthparts nearly aligned with the body axis; the posterior end of the abdomen is slender, and there are slender, independent or semi-independent anal prolegs. These larvae live in silk retreats or roam freely in search of food. Larvae of microcaddisflies (family Hydroptilidae) typically undergo hypermetamorphosis, with the first four instars campodeiform and free-living and the fifth instar with a much-enlarged abdomen and living in a purse case.

Pupae

The last instar builds a shelter or modifies the larval case to serve as a shelter. Modification of a larval case usually means simply sealing the ends with silk closure membranes that are perforated to allow movement of water through the shelter. Larvae may then line the shelter with silk; a few families (Glossosomatidae, Hydrobiosidae, Rhyacophilidae) spin a semipermeable cocoon inside the shelter. The larva may then diapause for a few months as a prepupa inside the shelter, but usually it proceeds immediately to molt into the pupal stage, usually retaining the exuviae of the last instar inside the shelter. The pupal appendages are not fused with the body, as they are in some Lepidoptera and Diptera, but are merely folded against the body, with the wings and antennae wrapped around the sides and venter of the body.

Special structures of the pupa include long mandibles for keeping debris from blocking anterior closure membrane perforations and for cutting the closure membrane or cocoon at the time of emergence, setose caudal processes for keeping debris from blocking posterior closure membrane perforations, and small plates of hooks dorsally on the abdomen for maneuvering the pupa inside the silk-lined shelter or cocoon. Respiration is assisted by undulating movements of the pupa, forcing water through the shelter and cocoon from anterior to posterior ends. The pupa usually completes development to the adult stage in two or three weeks.

Emergence

When the pupa is ready to transform to the adult stage, it uses its long mandibles to cut through the shelter's anterior closure membrane, swims to the water surface, and emerges from the pupal exuviae at the surface in open water or on some emergent object. In the molt from pupa to adult, the period between apolysis and ecdysis is unusually long, allowing body sclerites to complete much of their sclerotization before the adult emerges. In this way, the adult caddisfly is able to fly quickly from the surface of the water, away from predatory fish, usually completing the escape from the shelter and the water surface in less than a minute. The teneral adult then rests on shoreside vegetation or rock until sclerotization is complete.

Adult

Nearly all adult caddisflies are capable of flight, with only a few species having short, flightless wings. They are secretive, hiding among shoreside vegetation or rocks most of the time, with periods of activity in the day or night specific for the different species. Adults of different species live for a few days to several months. Those that have long adult lives have an especially well-developed haustellum for imbibing liquids; apparently they ingest water or nectar for sustenance. Mating behavior for some species is mediated by sexual pheromones, with females emitting odors that attract males. Males of some day-flying species also exhibit distinctive flight patterns or swarms that attract females. Mating is accomplished while standing on shoreside substrate in an end-to-end orientation with tips of the wings overlapping; the male inferior appendages, or "claspers," hold the end of the female abdomen while he inserts his phallus and releases a spermatophore into her spermatheca. The female then lays her egg mass. It may be laid on a rock or vegetation overhanging the water, from which first instars may drip into the water upon hatching. It may be laid on underwater substrates by a female that crawls or swims beneath the water surface. A few species may lay an egg mass in a protected spot in a dry depression that will later be filled with water.

HABITATS AND DISTRIBUTION OF TRICHOPTERA

Habitats

Immature stages of caddisflies may live in a wide range of habitats with fast to slow current speed or in standing water, but the habitat preferences of the individual caddisfly species usually are quite restricted. Most species require relatively clean, cool water with high concentration of dissolved oxygen. In these habitats, larvae may be found in accumulations of dead leaves and sticks, on or in woody debris, on the tops or bottoms or sides of stones, burrowing in sand or silt, in the crevices among gravel, on submerged or floating parts of living plants, or on other relatively stable debris. A few species complete their entire life cycles away from water, with immature stages crawling among leaves on the forest floor or clinging to vertical rock faces. Immature stages of a few species develop in brackish water, and those of species of the family Chathamiidae apparently all develop only in ocean surf on the shores of New Zealand, at least one of them in a complex mutualistic relationship with starfish.

Distribution

Caddisflies are distributed throughout the world, with species known from every continent except Antarctica. The greatest known species diversity is in the Oriental biogeographic region, where the density of species per hectare is nearly twice that of the next most diverse region.

CASE AND RETREAT MAKING BY TRICHOPTERA

Case Making

Caddisfly larvae have long been appreciated for the beautiful and complex cases that many of them build. These usually are not attached to a substrate but are carried by the larva as it crawls or swims in its habitat. Larvae build cases from a wide range of mineral and plant materials, with the type of material and the shape of the case often recognizable in the field for the genus or even for the species. Mineral building materials may include sand grains or small stones, which may be all of one size or may be of two or more sizes organized in a species- or genus-specific pattern, for example with larger "ballast" stones laterally. The larva selects a mineral particle of preferred size and shape and applies it to the anterior edge of the case with silk extruded from the spinneret in its mouth. Examples of plant building materials include living or dead pieces of leaves or wood that have been shaped by the larva before their attachment to the anterior edge of the case with silk. The preferred plant material may be very specific, such as algae, mosses, liverworts, or rootlets or leaves of particular vascular plants growing in the water, or pine needles, sticks, or leaves that fall from shoreside vegetation. These materials may be oriented longitudinally or horizontally; longitudinal materials may be organized in rings or in a spiral; horizontal materials may be interlocked at corners of a two-, three-, or four-sided case or may be wrapped in a cylindrical case. The larvae of a few species make their cases by hollowing the axis from a stick or piece of wood. Some cases are made entirely from silk or may incorporate pieces of freshwater sponge or abandoned shells of freshwater mollusks. At least one marine species builds its case with bits of coral.

The cases of caddisflies are often classified as tube cases (most Integripalpia), saddle cases or tortoise cases (Glossosomatidae), or purse cases (Hydroptilidae). A tube case is more or less cylindrical, surrounding the larva at approximately the same distance from all sides, although the shape of the case, especially externally, may vary greatly from one species to another. Thus, a tube case may be externally cylindrical or four-sided or three-sided or flattened; it may have lateral and/or anterodorsal flanges; or it may be straight or curved or coiled into a tight spiral like a snail shell. A tube case almost always has recognizable anterior and posterior ends. Pupation occurs within the tube case after it has been fastened to stationary substrate and the ends have been sealed with silk membranes (except for holes left for water circulation).

A saddle case is more or less domelike, oval on its flat side, always of mineral materials, sometimes with larger stones laterally. A transverse strap of fine sand grains connects the longer sides of the oval ventrally, beneath the larva, leaving anterior and posterior openings on the ventral side, from which the head and anus of the larva protrude interchangeably. This ventral strap is removed and the dome is fastened to stable substrate by the larva as it prepares to pupate under the dome.

A purse case typically consists of two flat sheets of silk (often including sand or algae) that have been sewn together at their edges, leaving the ends open for the head and anus to protrude interchangeably. Purse cases sometimes are cemented to the substrate. Pupation occurs within the purse case after the case has been fastened to stationary substrate and the ends have been sealed with silk membranes.

Retreat Making

About half the species of caddisflies do not build cases, but instead spin silken retreats that are stationary, fastened to the substrate. These retreats often are modified to assist with capturing food, usually by filtering it from moving water. The shape of a retreat usually is characteristic of a family or genus and may appear on wood or stone substrate—for example, as a flat sheet over a shallow depression on the substrate; a long, sinuate tunnel or a short covering that incorporates bits of mineral or plant material; a fingerlike net on the undersides of stones; or a bag of silk suspended from substrate in the current. Some species construct a silk-lined vertical tube in sandy soil. Pupation occurs either in the silken retreat or in a special pupation chamber constructed of silk and plant or mineral substrate by the larva.

Free-Living Larvae

Larvae of species of Rhyacophilidae and Hydrobiosidae are free-living predators, roaming about the substrate in search of prey. Pupation occurs under a domelike pupation chamber constructed of silk and plant or mineral substrate by the larva.

FEEDING STRATEGIES OF TRICHOPTERA

Although some long-lived adults may imbibe water or nectar, most nutrients are acquired by larvae. Larvae feed in a variety of ways, with a greater diversity of feeding strategies than any other group of insects having a comparable number of species. Early instars of most species and late instars of many species are collector-gatherers, picking small bits of loose organic material from the substrate. Many other species are collector-filterers, using silken nets (usually) or hairy legs to strain small bits of organic matter from moving water. Shredding herbivores chew pieces of leaves of living plants, often using some of the same leaves in their cases. Shredding detritivores gouge rotting wood or cut pieces of dead leaves that have been "conditioned" by fungi and bacteria, getting most of their nutrition from digestion of the fungi and bacteria. Scrapers graze on the algae, fungi, and associated organic material ("periphyton") that is attached to the surfaces of stones and plant material exposed to sunlight. Predators either chase their prey or lie in wait for it to wander nearby; prey items typically include other insects, microcrustaceans, and annelids; the predators either swallow their prey whole or they chew them in bits. Some caddisfly larvae may eat the flesh of dead animals, such as dead fish, as these become available. Some species of long-horned caddisflies feed facultatively or obligatorily on freshwater sponge, ingesting whole pieces of the sponge, including the spicules, which accumulate in the gut.

IMPORTANCE OF TRICHOPTERA

Caddisflies are one of the major groups of macroinvertebrates in freshwater ecosystems both in terms of species diversity and of density, especially in relatively unpolluted waterways. For this reason, they are significant contributors in the processing of nutrients. On one hand, collecting-gathering and collecting-filtering and scraping caddisflies help concentrate the nutrients of fine particulate organic matter into their own bodies, making the nutrients available to invertebrate and vertebrate predators in the food web. On the other hand, shredding herbivores and shredding detritivores and predators help to break coarse organic materials into small particles, including feces, that can then be used by many animals that are able to ingest only fine bits of organic material.

Food Resources

Because of the many different feeding strategies and habitat preferences of this diverse order, nearly every conceivable food resource is processed by caddisflies. Because their populations are so large, they process significant amounts of these resources for the benefit of the other animals in the ecosystem. There is even evidence that scraping caddisflies help to increase the production rate of the attached algae on which they feed, much as mowing the grass in a lawn or pruning new growth of fruit trees stimulates increased growth and production in those plants. The role of caddisflies in the food web is appreciated very well by sport fishing enthusiasts, who tie imitations of larval, pupal, and adult caddisflies on hooks to entice their game fish to bite a hook. The more those who fish learn about the species diversity, behavior, and biology of caddisflies, the more likely they are to succeed in tricking the fish to take the hook.

Pollution Tolerance

Although caddisflies generally will not tolerate even moderate levels of pollution, the range of tolerances is wide among the various species of caddisflies. For this reason and because of the usual high species diversity and density of caddisflies in unpolluted surface waters, communities of Trichoptera and other macroinvertebrates are often used to detect the presence of pollution. The occurrence of several of the less-tolerant species and high densities of large numbers of species of caddisflies suggest that the water is relatively unpolluted. Pollution may be detected with this technique more reliably and more cheaply than with chemical analyses. Once it has been established that a given waterway is polluted, follow-up analyses may then be attempted to discover the specific polluting substances or microorganisms and their concentrations. Finally, equipped with those data, land managers, engineers, economists, politicians, and other responsible decision makers may be better able to determine appropriate mitigating measures to reduce or eliminate the pollution.

See Also the Following Articles

Aquatic Habitats • Pollution, Insect Response to

Further Reading

Barbour, M. T., Gerritsen, J., Snyder, B. D., and Stribling, J. B. (1999). "Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish." 2nd ed. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Holzenthal, R. W., and Blahnik, R. J. (1997). Tree of Life: Trichoptera. http//phylogeny.arizona.edu/tree/eukaryotes/animals/arthropoda/hexapoda/trichoptera/trichoptera.html, updated 20 February 2001.

Kjer, K. M., Blahnik, R. J., and Holzental, R. W. (2001). Phylogeny of Trichoptera (caddisflies): Characterization of signal and noise within multiple datasets. *Syst. Biol.* **50(6),** 781–816.

LaFontaine, G. (1989). "Caddisflies: A Major Study of One of the Most Important Aquatic Insects—Entomology, Fly Tying, and Proven Fishing Techniques." Lyons Press, New York, NY.

Merritt, R. W., and Cummins, K. W., (eds.) (1996). "An Introduction to the Aquatic Insects of North America." 3rd ed., Kendall/Hunt, Dubuque, IA.

Morse, J. C. (1997). Phylogeny of Trichoptera. *Annu. Rev. Entomol.* **42,** 427–450.

Morse, J. C., (ed.) (1999). Trichoptera World Checklist. http://entweb.clemson.edu/database/trichopt/index.htm, effective 22 May 1999, updated 28 July 2000, 8 January 2001.
Schmid, F. (1998). "The Insects and Arachnids of Canada," Part 7: "Genera of the Trichoptera of Canada and Adjoining or Adjacent United States." NRC Research Press, Ottawa.
Wiggins, G. B. (1996). "Larvae of the North American Caddisfly Genera (Trichoptera)." 2nd ed. University of Toronto Press, Toronto.

Tsetse Fly

Stephen G. A. Leak
International Trypanotolerance Centre, The Gambia

Tsetse flies belong to the single genus *Glossina,* in the family Glossinidae of the order Diptera. They are found only in sub-Saharan Africa, Yemen, and Saudi Arabia, infesting 38 countries and occupying about 11 million square kilometers. As cyclical vectors of protozoan parasites of the genus *Trypanosoma,* they are of major economic and biological importance. Trypanosomosis (sleeping sickness in humans) is a major constraint to livestock production and is a threat to millions of people in Africa. Within the genus, 31 species and subspecies of tsetse have been identified. Fossil *Glossina,* found in Florissant shales in Colorado, date back to the Oligocene, indicating a wider original distribution.

The genus is split into three subgenera or groups: *Austenina,* (*fusca* group), *Nemorhina* (*palpalis* group), and *Glossina* (*morsitans* group). The forest-dwelling *fusca* group are the most primitive. Classification is based largely on morphology of the genitalia; however, the three groups also differ ecologically. Differences in the cuticular alkanes of their sex pheromones can also be used for classification and for estimating "genetic" distance between species.

FIGURE 1 An adult tsetse fly in the process of feeding. (Photograph courtesy of the International Livestock Research Institute, Graphic Arts Unit, Nairobi.)

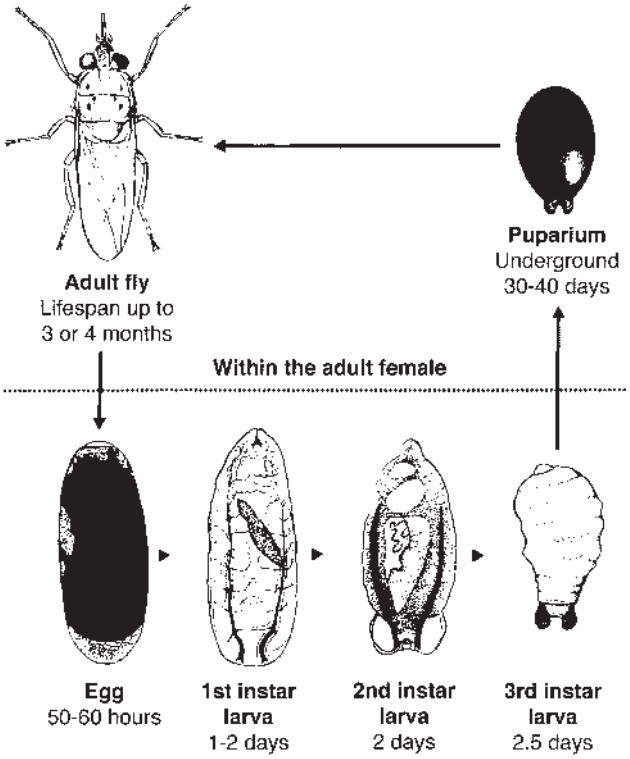

FIGURE 2 Life cycle of the tsetse fly.

LIFE CYCLE AND REPRODUCTION

An adult female produces one egg at a time, from which a first instar hatches in the uterus, where it is supplied with nutrients. After a period of development and molting in the uterus, a third instar is deposited on the ground. One full-grown larva is produced every 9 or 10 days, which then pupates in light or sandy soil. The adult emerges after a puparial period of around 30 days, depending on the ambient temperature. This process, termed adenotrophic viviparity, results in a very low rate of reproduction.

Spermatogenesis occurs during the puparial stage of males and ceases after eclosion; maturation of adult males takes approximately 3 days. Substances synthesized in the male accessory glands are transferred to the female hemolymph during copulation and stimulate receptivity of female tsetse. Unfed females are unwilling to copulate, and receptivity declines with age or after mating. Most female tsetse are successfully inseminated even at very low population densities and need to mate only once in their lifetime, although multiple mating does occur. Male flies seem to rely on visual attraction to find mates.

Larval Nutrition and Development

In pregnant females, proteins synthesized by highly efficient "milk" glands provide nourishment to the developing larva. Milk is secreted cyclically, with a peak in early pregnancy, which continues from the time of larval hatch until partu-

rition. Some components of the milk probably originate from fat-body secretions, and the milk contains relatively constant proportions of protein and lipid plus phospholipids, cholesterol, triglycerides, and tyrosine. The latter may be necessary for tanning larval and adult cuticles. At eclosion of the first instar, the milk is rich in acidic lipids, which are later replaced by proteins.

The anterodorsal mouth of the first instar is blocked by a chitinized median tooth. This "egg-tooth," lost with the skin of the first instar, is used to puncture the eggshell (chorion), which then splits along the dorsal side, allowing the first instar to hatch.

Larviposition and Pupariation

Unlike most dipteran larvae, which commit to metamorphosis early in the third instar, the third instar tsetse may pupariate only when fully grown. Shortly after larviposition, larvae exhibit a strong photonegative response and attempt to burrow. Uric acid from the anus then spreads over the larva and may protect it from predation. Periodicity of larviposition is affected by temperature, and the tsetse may respond to seasonal temperature cycles such that larviposition occurs in the most favorable conditions. The female fly becomes very active a few hours before parturition, presumably while searching for a suitable larviposition site. Both tactile and visual responses have a role in breeding site location. Females tend to choose darker resting sites as pregnancy progresses and when temperatures increase. The interlarval period depends on maternal nutrition and temperature conditions and varies according to species, localities, and seasons, from 20 to 90 days. Survival of puparia is dependent on relative humidity, which, together with temperature, influences selection of breeding and larviposition sites. Survival of emerging adults depends on the existence of sufficient fat reserves to support them until their first blood meal.

The adult emerges from the puparium using the "ptilinum," a saclike structure folded in a frontal cavity of the head. Rhythmic contractions of thoracic and abdominal muscles help the adult to break out of the puparium and reach the soil surface. Females emerge 1 to 2 days before males. Immediately after eclosion, the wings, abdomen, and thorax are highly compressed, although the legs are already full sized. The newly emerged fly has to expand by about 90%, using the cibarial pump to suck air into the gut, and must then feed to complete growth of the flight muscles and endocuticle.

BLOOD MEAL DIGESTION AND UTILIZATION

During feeding, tsetse discharge saliva containing a powerful anticoagulant antithrombin enzyme. The quantity of saliva secreted increases as tsetse become hungrier and may contain trypanosomes in infected flies. Two platelet aggregation inhibitors identified in saliva may also have immunosuppressive, anti-inflammatory properties. Digestion of blood proteins takes place in the posterior section of the midgut and involves six enzymes, which convert proteins to peptides and free amino acids.

Water Balance and Excretion

After feeding, the increased weight of the fly hampers flight and increases vulnerability to predators. There is therefore a rapid reduction of the water content of the blood meal from 79% to about 55% within 3 h after feeding. A blood meal in early pregnancy provides females with nutrients that are stored for larval development in late pregnancy.

Pheromones and Hormones

Sex recognition pheromones of tsetse are saturated, methyl-branched hydrocarbons found in cuticular waxes and are apparently unique to each tsetse species. They trigger mating behavior in males at ultraclose range or upon contact with baited decoys. The pheromone is released from the cuticle via unicellular glands on the legs and spreads over the external body surface by diffusion and grooming behavior. It is present on the pharate adult female about 2 days before emergence from the puparium and remains throughout life. A larval pheromone may be released from the anal orifice of tsetse larvae. The pheromone seems to affect choice of larviposition site, resulting in aggregation of larvae in breeding sites.

Flight Metabolism

The energy required for flight is produced from proline and is sustained by utilization of lipid reserves. Proline is depleted during flight and cannot rapidly be replenished; therefore, tsetse are limited to short periods of high-speed flight, at between 6.5 and 7.5 m s^{-1} (0.4 km min^{-1}) in open country. Proline is synthesized from lipids in fat bodies and oxidized to alanine in the muscles under hormonal control. Daily flight duration in male flies is about 15 min in the hot season and more than twice as long in the cold season. If fat reserves fall to about 6% of total dry body mass, the fly may die of starvation; such nutritional stress also results in production of small puparia.

Vision and Sense Organs

The compound eyes of both sexes are similar, with a specialized zone of greater visual acuity in the forward-pointing region. Males, however, have a wider region of binocular overlap, which may enhance detection of females. The visual function is consistent with fast flight, detection of drift due to low wind speeds, mating chases, and discrimination of cryptic hosts at high light intensities.

For mating behavior to be initiated, both mechano- and chemoreceptors must be stimulated. The chemoreceptors are

thought to be basiconic sensilla on the tibiae and tarsi. Other, hair-shaped chemo- and mechanoreceptor sensilla are found on the front side of the wings.

DISTRIBUTION AND HABITATS

The general distribution of tsetse flies, determined principally by climate, is influenced by altitude, by vegetation, and by the presence of suitable host animals. The limit of distribution is closely correlated with the "tropical savanna (summer rain) climate," which follows the > 508 mm annual rainfall line, whereas the "tropical rain forest (equatorial) climate" controls the habitats of the *fusca* and *palpalis* groups. The surrounding savannas are the habitat of the *morsitans* group. The southern limits of *Glossina* distribution in Africa lie north of a line drawn from Benguela, in Angola, to Durban, in South Africa. The northern limits are roughly a line from Dakar in Senegal across to Ethiopia and Mogadishu in Somalia on the east coast.

The distribution and abundance of *morsitans* group tsetse often correspond to those of wild animals. In West Africa, *G. longipalpis* inhabits forest islands with plenty of shade and moves to forest edges only during seasons of high rainfall. *G. austeni* is confined to the coastal zone of East Africa, from Somalia to South Africa, including the island of Zanzibar until it was eradicated from the island in 1997.

Species of the *palpalis* group occur in drainage systems leading to the Atlantic Ocean or to the Mediterranean, not those draining into the Indian Ocean. *G. palpalis* is found close to water, in gallery forests, and it cannot tolerate the wide range of climatic conditions occurring in the savanna belt, where it is restricted to watercourses or "forest islands."

Species of the *fusca* group are mostly found in forests of West and central Africa. Humans are likely to have an adverse effect on *fusca* group tsetse populations through forest clearing and hunting.

Human activity has an important influence on tsetse distribution and abundance; humans can scare away or kill potential hosts, or destroy the vegetation forming the flies' habitat through agricultural or other development. Most African countries, particularly those in tsetse-infested areas, have low human population densities; however, Nigeria, Africa's most densely populated country, has a population of 89 to 100 million, equivalent to 108 people km^{-2}. Nigeria developed rapidly, and as a result, up to two-thirds of the potential *G. m. submorsitans* population of Nigeria may have been suppressed. *G. m. submorsitans* occurs in areas with human population densities ranging from 0 to 15 km^{-2}, occasionally in areas of 15 to 40 km^{-2}, but never when the population exceeds 40 km^{-2}. Most sub-Saharan African countries have densities below 40 km^{-2}. Flies of the *palpalis* group, particularly *G. tachinoides,* are much less affected by human settlement, possibly because they are able to adapt from a preference for feeding on wild mammals and reptiles, to feeding mainly on humans and their domestic animals. Tsetse flies can adapt to peridomestic habitats, and *G. tachinoides* commonly follows domestic pigs in villages and utilizes human-made larviposition sites such as clumps of oil palm, cola nut, and banana trees.

HOSTS AND FEEDING BEHAVIOR

Potential hosts are recognized from visual and olfactory stimuli, which, together with mechanical stimulation, activate tsetse and initiate host-oriented responses. Approach to a stationary host is by upwind flight, modulated by olfactory stimuli, flight speed significantly reduced when a fly enters an odor plume. Final orientation toward the host is visual. Heat stimulation after a fly has landed on the host may then cause a probing response and subsequent feeding. Both endogenous and exogenous factors influence host-seeking behavior. Endogenous factors include a circadian rhythm of activity, as well as the level of starvation, age, sex, and pregnancy status of the flies. Exogenous factors include temperature, vapor pressure deficit, and visual, mechanical, and olfactory stimuli.

Olfactory Stimuli

Odor attractants are of three types: those associated with animal breath, such as acetone, octenol, and CO_2; those associated with urine, such as the phenols; and those associated with skin secretions, such as sebum.

Feeding Process

As a tsetse fly lands on a host, the labium is enclosed between the palps. Heat is the prime stimulus leading to probing and subsequent feeding by tsetse. Responsiveness increases with hunger until the final stages of starvation. As the fly starts to probe, the labium moves from the palps, to an angle of 90° to the skin. While the labella rests on the skin, the teeth on the inner surface are everted and penetrate it. At the same time, saliva is excreted from the hypopharynx. Normally, the labellar teeth lacerate capillaries, resulting in a hemorrhage, which is sucked into the labrum. When the fly stops sucking, a small pool of blood forms. If blood is not found, the fly withdraws the labium partially and makes a new penetration. Unmated female tsetse take smaller meals than mated ones either because the gut cannot be distended to the same extent or because of lower metabolic demands.

Host Effects and Preferences

Nonrandom feeding patterns may result from responses of hosts to tsetse attack, such as tail-flicking or skin-rippling reactions. Although normally unable to discriminate between potential hosts at long-range, the upright habit of humans, and lactic acid secretions in the skin, result in visual and olfactory repellency, including inhibited landing responses. Zebras may be protected from being fed on by biting flies,

including tsetse, by their striped pattern. Tsetse are less attracted to stripes than to solid colors and seem to avoid horizontally striped objects. Zebras usually have a combination of vertical and horizontal stripes, but the horizontal stripes are on the lower part of the animal, where tsetse normally feeds.

Natural Hosts

There is a close correspondence between ecological niches of common hosts and tsetse habitats, and overlapping habitats are important in determining feeding patterns, in addition to behavioral characteristics of the host.

Tsetse will change to alternate hosts if their usual host(s) become unavailable. For example, *G. longipennis,* which once fed mainly on rhinoceros, adapted to feeding on other large animals as rhinoceros became scarce. Tsetse select host species rather than simply feeding on the most common available host.

Wild pigs (Suidae), particularly warthogs, are important hosts for *morsitans* group tsetse; however, the feeding efficiency of tsetse visiting a single warthog could be as low as 12 to 18%. Between 26 and 31% of tsetse landed on the head region of live adult warthogs, apparently because of a visual response to a dark patch produced by the preorbital glands of mature warthogs. In East Africa Suidae can form about half the food supply of *morsitans* group tsetse in areas with wildlife. Most other feeds are from ruminants, particularly bushbuck.

The *palpalis* group, generally inhabiting riverine or lacustrine vegetation, have a close ecological association with important hosts such as crocodiles and monitor lizards. They are, however, opportunistic feeders, and many mammals, including humans and reptiles, that enter their habitat may be fed upon. In peridomestic situations in the Guinean "forest savanna mosaic" region of West Africa, domestic pigs are an important source of blood meals for *G. palpalis* and *G. tachinoides;* wild Suidae, humans, and small ruminants are the next most common hosts. In the forest zone of Côte d'Ivoire the feeding habits of *G. palpalis* vary with the availability of wild hosts and activities of humans. Around villages, nearly all feeds are from domestic pigs, whereas in plantations, more feeds are from humans. In human trypanosomosis foci outside the edges of villages, *palpalis* group flies may feed equally on humans and antelopes, especially bushbuck. *G. tachinoides* can survive in close association with humans in the virtual absence of wild mammals and reptiles. It will readily feed on domestic pigs or cattle, but feeds have rarely been identified from domestic sheep and goats even when these were common.

RESTING BEHAVIOR

Tsetse rest for most of the day; making use of sites that provide protection from extreme temperatures and from predators. The sites may also provide vantage points from which to seek hosts, and their location can change according to time of day, season, and hunger stage of the fly. Recently engorged tsetse fly poorly and rest low down on tree trunks; the height above ground level of the resting site increases in relation to the fly's nutritional state.

During the daytime, high temperatures of exposed resting sites result in tsetse moving to rest in sites such as the boles of large trees, where they squeeze into the fissures of the bark, at heights generally less than 0.3 m from the ground. Otherwise, flies hide in rot holes, often quite high up, in big tree trunks.

At dusk, tsetse flies generally move upward, to spend the night on leaves or small twigs, possibly to avoid predators. The choice of nocturnal resting sites is influenced by seasonal effects on the physical condition of the vegetation, such as leaf fall. A reverse migration seems to take place about an hour after sunrise, possibly in response to light intensity falling below a critical level. This rapid migration may occur at different temperatures in different seasons.

Activity

The diurnal pattern of tsetse activity may differ between species and according to sex, hunger, pregnancy, and nutritional state of the population. The pattern may also differ for the same species of tsetse in different localities. Natural activity of tsetse flies can be environmentally stimulated or spontaneous, but it has an underlying endogenous circadian rhythm modified by temperature. Intensive activity occurs for periods of less than a minute. The *morsitans* group tsetse are mostly active early in the morning and/or late in the afternoon.

DEVELOPMENT OF TRYPANOSOMES IN TSETSE

Fly species differ in their capacity to transmit trypanosomes, and individual fly genotypes also vary and affect susceptibility to trypanosome infection. For example, *salmon* mutants of *G. m. morsitans* appear to be better vectors of trypanosomes, perhaps owing the metabolism of tryptophan, which is essential for trypanosomes. Metabolism in *salmon* mutants is affected so that tryptophan accumulates, predisposing the fly to trypanosome infection. Flies of the *morsitans* group are mostly good vectors of all trypanosome species. Species of the *palpalis* group seem to be poor vectors of *Trypanosoma congolense* but efficient vectors of some stocks of *T. vivax* and can be important vectors of human infective trypanosomes. Species of the *fusca* group can be effective vectors of *T. congolense* and *T. vivax* but poor vectors of *Trypanozoon* trypanosomes. Trypanosome infection rates in tsetse are generally low and can depend on the age of the fly at the time of the infective feed. Wild male tsetse can achieve a life span of almost 5 months, but this is probably unusual. Teneral flies taking an infective blood meal on the first day of life, or soon after emergence, are the most easily infected. The prevalence of mature infections appears to increase with age for *Nannomonas* and *Duttonella* group trypanosomes in many tsetse species. Susceptibility to trypanosome infection may be age dependent, but the actual age is less critical than whether the fly has taken a previous, uninfected feed.

SLEEPING SICKNESS

Trypanosomes were first known to infect humans after being detected in the blood of a steamboat captain in the Gambia in 1901. The parasite was identified and named *Trypanosoma gambiense.* Sleeping sickness currently occurs in about 200 distinct disease foci in Africa. These foci place between 35 and 55 million people at risk, although only about 3 million of them are under surveillance and relatively few new cases are diagnosed annually. Human sleeping sickness was largely under control in the 1960s, but its incidence then increased, and in 1994 it was estimated that there were 150,000 cases in the Democratic Republic of the Congo. Other estimates suggest that around 300,000 people are infected in Africa, many of whom will die because of lack of treatment.

Identity and Origins of the Parasite

The two parasites causing human sleeping sickness, *T. brucei gambiense* and *T. b. rhodesiense,* are morphologically indistinguishable, and early diagnosis was based on clinical signs of the disease. Sleeping sickness caused by *T. b. rhodesiense* is an acute disease, with death occurring after only a few months, whereas with *T. b. gambiense* death may not occur for several years. *T. b. gambiense* occurs in West and central Africa and is transmitted predominantly by *palpalis* group tsetse, while *T. b. rhodesiense* occurs in southern and East Africa and is transmitted by *morsitans* group flies.

Although generally accepted that sleeping sickness trypanosomes are derived from *T. b. brucei* (infecting wild and domestic animals and morphologically indistinguishable from the human infective parasites), there are several theories on the evolution of species infecting humans. First, the two types of sleeping sickness may be caused by the same parasite, the disease simply being more acute when the trypanosome becomes adapted to certain species of wild animals. Chronic disease occurs when the trypanosome is adapted to humans as its principal host. Second, *T. b. rhodesiense* may have evolved in the late Miocene or Pliocene, when hominids were exposed to *Trypanozoon* trypanosomes in their savanna habitat. Subsequently, *T. b. gambiense* may have evolved from *T. b. rhodesiense* when hominids invaded forests and became hosts of *palpalis* group tsetse. Third, the two subspecies may have evolved independently, from *T. b. brucei* or a common ancestral species. A fourth theory is that after *T. b. gambiense* spread to savanna areas of southeast Africa, *T. b. rhodesiense* evolved from it and subsequently spread northward. An alternative theory is that *T. b. brucei* gave rise first to *T. b. rhodesiense* and then to *T. b. gambiense* and that the two diseases evolved in humans from an animal infection to an anthropozoonosis *(T. b. rhodesiense),* and then to a pure anthroponosis *(T. b. gambiense).*

Molecular characterization of *T. b. rhodesiense* suggests that it had polyphyletic origins. It is widely accepted that *T. b. gambiense* sleeping sickness spread from West to central Africa. Areas of West Africa where sleeping sickness occurred and the Nile basin are ecologically similar, with a continuous corridor of Sudano-Guinean climate and vegetation. *T. b. rhodesiense* then would have spread northward either from Central Africa through East Africa or from the Zambezi valley in Zambia, increasing in virulence the further north it progressed. Contradictory evidence from biochemical and molecular characterizations suggests that strains of *T. rhodesiense* from Zambia, Kenya, and Uganda have independent origins.

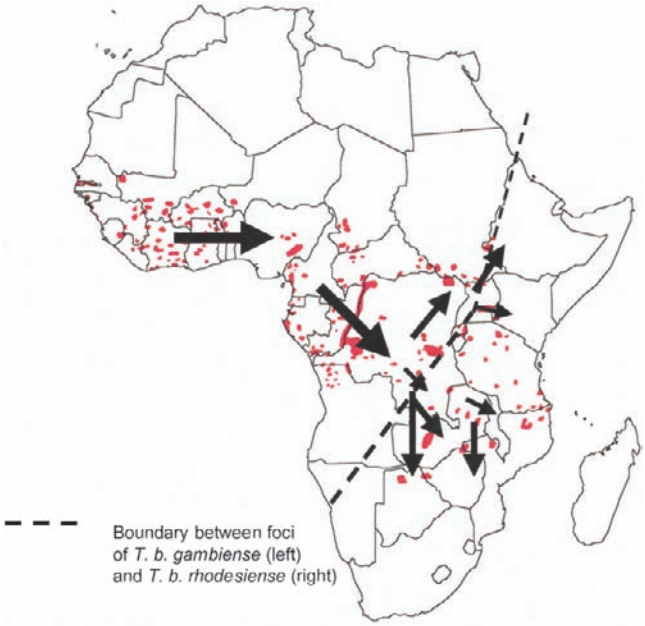

FIGURE 3 Distribution of human sleeping sickness foci in Africa. Dashed line indicates boundary between foci of *T. b. gambiense* (west) and *T. b. rhodesiense* (east). Arrows denote the probable direction of spread of the two types.

Reservoir Hosts

Until recently, one of the difficulties in proving the existence of reservoir hosts for *T. b. rhodesiense* and *T. b. gambiense* arose from problems in subspecies identification. It is now possible to use molecular genetic and biochemical methods to identify human infective forms. Consequently, there is now evidence of a range of wild and domestic animal host reservoirs. The latter are likely to be less important in the epidemiology of sleeping sickness due to *T. b. gambiense* than for *T. b. rhodesiense.* Many animals can be experimentally infected with either subspecies, and wild and domestic animals can maintain the "*rhodesiense*" disease endemically.

See Also the Following Articles

Medical Entomology • Veterinary Entomology • Zoonoses, Arthropod-Borne

Further Reading

Beard, C. B., O'Neill, S. L., Tesh, R. B., Richards, F. F., and Askoy, S. (1993). Modification of arthropod vector competence via symbiotic bacteria. *Parasitol. Today* **9,** 179–183.

Brady, J., Packer, M. J., and Gibson, G. (1990). Odour plume shape and host finding by tsetse. *Insect Sci. Appl.* **11,** 377–384.

Carlson, D. A., Milstrey, S. K., and Narang, S. K. (1993). Classification of tsetse flies *Glossina* spp. (Diptera: Glossinidae) by gas chromatographic analysis of cuticular components. *Bull. Entomol. Res.* **83,** 507–515.

Cockerell, T. D. A. (1919). New species of North American fossil beetles, cockroaches, and tsetse flies. *Proc. U.S. Nal. Mus.* **54,** 2337, 301–313.

Colvin, J., and Gibson, G. (1992). Host seeking behavior and management of tsetse. *Annu. Rev. Entomol.* **37,** 21–40.

Gooding, R. H. (1984). Tsetse genetics: A review. *Quaest. Entomol.* **20,** 89–128.

Jackson, C. H. N. (1946). An artificially isolated generation of tsetse-flies (Diptera). *Bull. Entomol. Res.* **32,** 291–299.

Langley, P. A., Pimley, R. W., and Carlson, D. A. (1975). Sex recognition pheromone in tsetse fly *Glossina morsitans. Nature* **254,** 51–53.

Leak, S. G. A. (1998). "Tsetse Biology and Ecology: Their Role in the Epidemiology and Control of Trypanosomosis." CABI, Wallingford, Oxon, Vit.

Maudlin, I. (1991). Transmission of African trypanosomiasis: interactions among tsetse immune system, symbionts, and parasites. *Adv. Dis. Vector Res.* **7,** 117–148.

Mulligan, H. W. (ed.) (1970). "The African Trypanosomiases." Ministry of Overseas Development and George Allen & Unwin, London.

O'Neill, S. L., Gooding, R. H., and Aksoy, S. (1993). Phylogenetically distant symbiotic micro-organisms reside in *Glossina* midgut and ovary tissues. *Med. Vet. Entomol.* **7,** 377–383.

Welburn, S. C., and Maudlin, I. (1992). The nature of the teneral state in *Glossina* and its role in the acquisition of trypanosome infection in tsetse. *Ann. Trop. Med. Parasit.* **86,** 529–536.

Zdárek, J., and Denlinger, D. L. (1993). Metamorphosis behaviour and regulation in tsetse flies (*Glossina* spp.) (Diptera: Glossinidae): A review. *Bull. Entomol. Res.* **83,** 447–461.

U

Urban Habitats

Michael K. Rust
University of California, Riverside

Urban environments are created from natural or agricultural ecosystems that have been disturbed by human activities, typically by the construction of towns and cities. The urban biotic environment can be divided into plant and animal communities that surround houses and buildings, occur indoors or in close contact with humans, or directly affect the structure. Human activities give rise to special conditions and environments that are frequently exploited by a select group of insects and arthropods. For example, about 25 species of the some 3500 known cockroaches are considered to be domicilary pests. Of these 25 species, only about 5 are considered to be serious urban pests. Similarly, of the approximately 550 species of ants found in the United States, only 30 species commonly infest homes, and only 10 of these are of major importance. Urban landscapes are truly special ecological environments and niches.

Urban environments are similar in developing and developed countries worldwide in that natural systems have been disturbed by human activities. The rate at which urbanization is occurring in undeveloped areas of Africa, South America, and Asia is almost incomprehensible. By the year 2025, 6.3 to 8 billion people will inhabit less developed regions of the world compared with 1.2 to 1.5 billion people that will inhabit the developed regions. Between 1960 and 1990, the number of cities in China, Indonesia, and Brazil with more than 500,000 residents doubled; in India, the number tripled. By the year 2010, about 57% of the world's population will be living in urban centers, and in Brazil, it will be close to 89%. Because of the rapid growth of urban cities in developing countries with tropical climates, urban insect pests, especially those that vector human and animal disease pathogens, are a critically important emerging problem.

DEVELOPMENT OF MEGACITIES

A newly emerging feature of countries with large population centers is the development of mega-urban cities. Megacities typically incorporate two or more large urban centers linked by transportation. Rural zones between such megacenters are characterized by commercialization of agriculture, expansion of transport systems, and employment shift from farming to other activities, accompanied by migration to the cities. The number and size of these megacities in developing countries will continue to grow, and by 2015 it is projected that there will be some 23 megacities with more than 10 million inhabitants (Table I). Tremendous amounts of resources and commerce will be necessary to support the large populations and the waste and garbage they will produce. It is these conditions that are conducive to insect pest problems such as mosquitoes, flies, cockroaches, fleas, termites, and wood-destroying pests. In addition, crowding and waste lead to vertebrate pest problems such as rats, mice, and pigeons.

Urban centers require the movement of goods and food to maintain their large human populations. Human movement and transportation of commerce provide a unique and rapid dissemination of insects and arthropods between centers. At the turn of the 20th century, it took over a week for ships to sail from Europe to North America. Today aircraft make the same journey in less than 7 h. The potential exists for rapidly spreading insects and the diseases they vector from one continent to another. Three recent examples of urban invasive species spread by commerce are presented.

URBAN INVASIVE SPECIES

Asian Tiger Mosquito

TABLE I Emerging Megacities of the World with More Than 10 Million Inhabitants (population in millions)

1975		2000		2015	
City	Population	City	Population	City	Population
Tokyo	19.8	Tokyo	26.4	Tokyo	26.4
New York	15.4	Mexico City	18.1	Bombay	26.1
Shanghai	11.4	Bombay	18.1	Lagos	23.2
Mexico City	11.2	São Paulo	17.8	Dhaka	21.1
São Paulo	10.0	Shanghai	17.0	São Paulo	20.4
		New York	16.6	Karachi	19.2
		Lagos	13.4	Mexico City	19.2
		Los Angeles	13.1	Shanghai	19.1
		Calcutta	12.9	New York	17.4
		Buenos Aires	12.6	Jakarta	17.3
		Dhaka	12.3	Calcutta	17.3
		Karachi	11.8	Delhi	16.8
		Delhi	11.7	Metro Manila	14.8
		Jakarta	11.0	Los Angeles	14.1
		Osaka	11.0	Buenos Aires	14.1
		Metro Manila	10.9	Cairo	13.8
		Beijing	10.8	Istanbul	12.5
		Rio de Janeiro	10.6	Beijing	12.3
		Cairo	10.6	Rio de Janeiro	11.9
				Osaka	11.0
				Tianjin	10.7
				Hyderabad	10.5
				Bangkok	10.1

Source: United Nations Population Division. "World Urbanization Prospects," 1999 rev. UN, New York.

It is thought that the Asian tiger mosquito, *Aedes albopictus,* entered the United States in Houston, Texas, in 1985 in used automobile tires imported from northern Asia. *Ae. albopictus* is a major human biting pest species throughout its range. In addition, it has been experimentally shown to be a competent vector of a number of viruses including western equine encephalitis, dengue, Japanese encephalitis, and yellow fever. To highlight the impact of commerce on the potential movement of insect pests, over 11 million used tires from 58 different countries entered the United States from 1978 to 1985. *Ae. albopictus* also shows the importance of local transport and the interstate highways. In 1987, of the 92 counties infested with this species, 64 were on an interstate highway, nearly twice the number that would be expected if their spread were not related to these road systems.

Since its introduction, this mosquito has spread to 25 states, mostly throughout the southeastern United States but as far north as Iowa, Illinois, and Ohio (Fig. 1). The female's propensity to lay eggs in bottles, cans, tires, rain gutters, flower pots, watering cans, children's swimming pools, and containers makes this mosquito ideally suited for urban environments.

Formosan Subterranean Termite

The Formosan subterranean termite, *Coptotermes formosanus,* has spread by marine commerce to semitropical and tropical seaports throughout the world. Once ashore, railroad ties provide a rapid method of moving and establishing large colonies of termites. In the 1960s, *C. formosanus* was discovered in Texas, Louisiana, and South Carolina. By 2000, it had spread to nine other southern states, and to California and Hawaii. *C. formosanus* is characterized by extremely large populations of aggressive foragers that can do extensive damage to structures. It is the most import urban pest in Hawaii and has been estimated to cause about $300 million damage a year in New Orleans, threatening nearly all the buildings of the historic French Quarter.

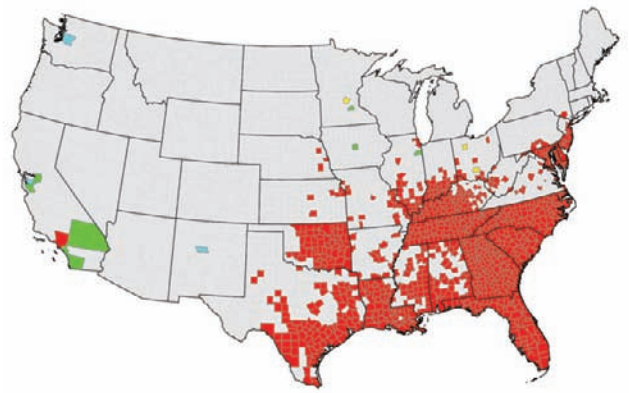

FIGURE 1 Current distribution of *Ae. albopictus* in the United States: red, positive counties; yellow, negative/eradicated/disappeared; blue, intercepted (never established); green, current status unknown. [From Moore, C. G. (1999). *Aedes albopictus* in the United States: Current status and prospects for further spread. *Am. J. Mosq. Control Assoc.* **15(2),** 221–227.]

FIGURE 2 A male and a female *An. glabripennis* on a poplar *(Populus alba)* tree. (Photograph by Baode Wang, taken on July 4, 1999 in Gansu province, China. U.S. Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantines Programs, Otis Plant Protection Laboratory.)

Asian Long-Horned Beetle

The Asian long-horned beetle, *Anoplophora glabripennis* (Fig. 2) has the potential of becoming a serious pest of ornamental trees found in urban landscapes. Infestations of Asian long-horned beetle were first discovered in 1996 in Brooklyn. Another infestation was found in Chicago in 1998. It is believed that the beetle entered the United States in solid wood packing material from China. The beetle has been found in warehouses in some 26 locations thoroughout the United States. This beetle poses a serious threat to many different hardwood trees that are common in urban landscapes, including sugar, silver and red maple, birch, poplar, willow, ash, and black locust. The cost of suppressing the beetle in New York has already reached $4 million.

ADDITIONAL URBAN PESTS

The vast majority of insects, however, are not capable of adapting to the artificial and unnatural conditions created by humans in urban environments. A few pests such as the German cockroach, cat flea, and house fly have readily adapted, have been able to survive in close proximity to humans, and now are cosmopolitan. Many urban insect pests such as the German cockroach, oriental cockroach, confused flour beetle, and granary weevil have lost their ability to fly and rely on human activities for their dispersal. Other species such as the Argentine ant and the pharaoh ant do not swarm and rely on budding for reproduction and commerce for transport. Many of the indoor pests of food and fabric have continuous development and do not require special dormancy or hibernation to complete development. These insects thrive at temperatures ranging from 18 to 35°C and relative humidity (RH) ranging from 50 to 90%, which are ranges typically found within human dwellings.

In addition to human occupants, domesticated pets may be a source of pests such as fleas, ticks, and mites. The cat flea, *Ctenocephalides felis,* is the most important ectoparasite of domesticated cats and dogs worldwide. From its probable origin in Africa, the cat flea has spread worldwide. In most urban environments, the larval development requirements of *C. felis* (RH > 75%, 21–32°C) are found in restricted microhabitats such as crawl spaces under homes, pet houses and bedding, utility rooms, and bedrooms. Interestingly, surveys of feral animals such as opossums indicate that *C. felis* is encountered only on animals frequenting urban settings.

NEED FOR URBAN PEST MANAGEMENT

The need for urban pest management in and around structures can be directly attributed to the dislike of insects and arthropods by humans. Some pests such as cockroaches, fleas, flies, and mosquitoes represent potential health risks. Others such as termites and wood-destroying pests can cause serious economic damage to structures. Pantry and fabric pests cause damage to foods, fabrics, and items made with furs, skins, and feathers. Many pests such as earwigs (Dermaptera), crickets (Orthoptera), pillbugs (Isopoda), sowbugs (Isopoda), and millipedes (class: Diplopodia) are occasional intruders, but their suppression often requires remedial action. The control of urban pests has created a large and important industry in the United States and Europe; in the United States alone about 1 million structures are treated each year for termite damage, repairs, and treatments at a cost of about $520 million.

The costs of treatment and repair by the pest control industry in the United States are estimated at about $2 billion. Consumers' annual expenditures on over-the-counter pest control remedies probably are at least as much. In developing countries, there will be an increasing need for pest control to protect the health and welfare of the inhabitants. With the increasing levels of international trade and tourism urban pest management will take on special importance in the 21st century.

See Also the Following Articles

Blattodea • Cat Fleas • Introduced Insects • Isoptera • Stored Products as Habitats

Further Reading

Anon. (1992) "Long-Range World Population Projections: Two Centuries of Population Growth, 1950–2150." United Nations Publication ST/ESA/SER.A/125. UN, New York.

Ebeling, W. (1975). "Urban Entomology." University of California Press, Division of Agricultural Sciences, Berkeley.

Moore, C. G. (1999). *Aedes albopictus* in the United States: Current status and prospects for further spread. *J. Am. Mosq. Control Assoc.* **15(2),** 221–227.

Robinson, W. H. (1996). "Urban Entomology. Insect and Mite Pests in the Human Environment." Chapman & Hall, London.

United Nations Population Division. (1999) "World urbanization prospects: The 1999 Revision." UN, New York.

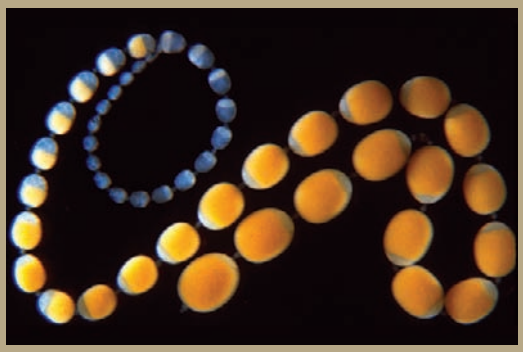

V

Venom

Justin O. Schmidt
Southwestern Biological Institute, Tucson

Venoms are biologically active liquids delivered into or onto another organism through a piercing structure such as a mouthpart or sting apparatus and often resulting in pain, tissue damage, paralysis, or death in the target organism. Insects contain more venomous species than all other groups of terrestrial animals combined. Within insects, venoms evolved numerous times to occupy positions in groups as diverse as ants and wasps to biting bugs, robber flies, and stinging caterpillars. Venoms empower insects with unique properties that enable them to achieve far greater effects or impacts on target organisms than would be expected for such small animals. For example, a tiny wasp weighing far less than a milligram can rapidly paralyze a caterpillar hundreds of times its own size. Likewise, a single attacking honey bee can send a million times larger human into panicked flight.

Venoms provide the biological means for insects to break free from many ecological restraints, thereby expanding their opportunities to exploit the world around them. One could argue that the omnipresence of ants in today's world is the result of venom combined with a social structure in ancestral ant lineages. Parasitic braconid and ichneumonid wasps, whose species number more than 100,000, with probably several times that number undescribed, owe their successful speciation, in part, to the ability of their venoms to facilitate use of a variety of host organisms. Honey bees (*Apis* spp.) have become dominant pollinators throughout much of the world, in large part because their venomous stings have enabled them to defend against a multitude of large, often intelligent predators. Without their effective defensive venoms, honey bees could not prevent the plunder of their large stores of honey and pollen, which they must maintain to overcome long seasonal periods of harsh conditions. The biological roles, operation, and properties of insect venoms provide dazzling illustrations of the successful adaptation of insects to their environment.

BIOLOGICAL ROLES OF VENOMS

Venoms serve two primary roles—defense and prey capture. In most lineages, the ancestral role, from which a defensive role evolved secondarily, was the use of venom for prey capture. The venoms of many taxa retain primary roles of prey capture. Examples are numerous and include many families or superfamilies of bugs, beetles, flies, neuropterans, and especially parasitoid and solitary aculeate wasps. Many familiar insects have taken venom use another step—to defend against predators. The mere fact that insects such as honey bees, yellowjackets, and hornets are so universally well known attests to the effectiveness of their defensive venoms. Defensive venoms are present in all social wasps, many ant species (including fire and harvester ants), some solitary bees and wasps, a few bugs, and some spiny caterpillars. In many of these species, the role of venom for prey capture has been entirely replaced by a role in defense. In caterpillars, which never used venom for prey capture, the role evolved independently. Some species retain both roles, with the spider wasps (Pompilidae) as conspicuous examples.

EVOLUTION OF VENOMS

Most venoms evolved from preexisting digestive or reproductive systems. Exceptions are the lepidopteran caterpillars, for which the origin of their venomous spines is unclear. The venoms of all other non-hymenopteran insects evolved from salivary or gastric secretions associated with the mouthparts and digestion. In some taxa, the venom toxins are salivary secretions injected into the prey. In others, the toxins are regurgitated from the gut into the prey through piercing mouthparts. Asilid flies use the latter form of venom delivery.

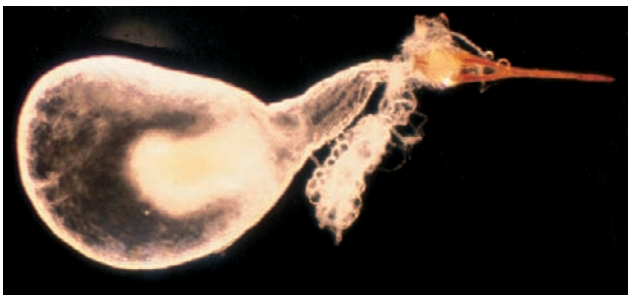

FIGURE 1 Fire ant *(S. invicta)* sting apparatus including the stinger, a derived ovipositor, with attached large venom sac and small Dufour's gland, which are derived accessory reproductive glands. (Photograph by J. O. Schmidt.)

The Hymenoptera evolved an entirely different venom system. In Hymenoptera, the stinger evolved through modification of the female ovipositor, and venom evolved from the accessory reproductive glands (Figs. 1 and 2). Consequently, only female Hymenoptera can sting. The lack of ability to sting and defend a colony against large predators is considered a primary reason for the failure of males, often called drones, of honey bees and social Hymenoptera to meaningfully defend their colonies against predatory assaults or to help the colony via foraging.

VENOMS USED FOR PREY CAPTURE

Two basic types of venom are used for prey capture: those that kill outright and those that only paralyze or alter prey physiologically. Prey paralysis is important for many insects that provide prey for their young. Prey that is paralyzed but still living does not spoil rapidly, giving larvae time to feed and grow. Insects that kill their prey with venom tend to consume it immediately, though exceptions to this rule exist. Wasps that use their venoms to provide a living host for their larvae and oviposit one or more eggs in or on the host are often called parasitoids. The venoms of parasitoids span a paralysis range from no paralysis to temporary paralysis to permanent paralysis. Venoms causing no paralysis, or short-acting paralysis, are found in many species of wasps in which the parasitized host continues feeding and growing until just before it dies, all the while harboring and protecting the parasitoid larvae from environmental and predatory risks. The venoms of these wasps often contain mutualistic viruses and other components that not only help protect the oviposited eggs from host encapsulation and immune defenses, but also alter the hormone balance inside the host to prevent molting to the next stage of development. If the host is in a protected environment, such as exists for wax moth larvae in combs of a dead honey bee colony, an extremely potent venom that causes total and permanent paralysis, like that of a parasitoid wasp such as *Bracon hebetor,* can be beneficial.

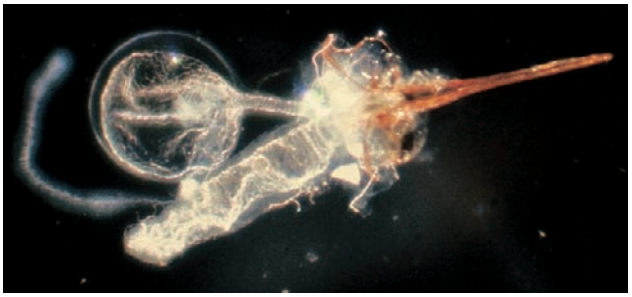

FIGURE 2 *Pogonomyrmex badius* harvester ant sting apparatus including the round venom sac with attached free filament that collects materials from the hemolymph for venom synthesis, and the tubular Dufour's gland. (Photograph by J. O. Schmidt.)

Hunting wasps are another group of solitary wasps that use venom for prey capture. These wasps include familiar sphecid, eumenid, and spider wasps plus a variety of other families of aculeate wasps. Like classical parasitoid wasps, hunting wasps possess venoms with a large activity span. For example, the sphecid *Larra analis* uses venom to paralyze its mole cricket hosts temporarily. At the other extreme, represented by the beewolf *Philanthus triangulum* and tarantula hawks (*Pepsis* spp.), the hunting wasps paralyze the honey bee and tarantula prey permanently. No evidence for injection of viruses or factors altering the host's endocrine or immune systems exists for hunting wasps. The majority of parasitoid and hunting wasp venoms are of little value for defense against predators. In many groups, the wasps cannot, or do not, even attempt to sting predators; in others, the pain and activity of the venom, at least to humans, is trivial. Some exceptions exist: most pompilid wasps, some eumenid and bethylid wasps, and a very few sphecid wasps can deliver painful stings; however, none of these venoms is meaningfully toxic to mammals.

DEFENSIVE VENOMS

Because of the extreme size difference between vertebrates and insects, insects are vulnerable to attack by vertebrate predators. Once such a small creature is captured by a large predator, typical defenses such as kicking, scratching, and biting are all but useless. A very few defenses, particularly potent venomous stings and allomones, can neutralize the predator's size advantage and blunt attacks. A key feature promoting the efficacy of defensive venoms is the ability of the sting apparatus to bypass the predator's skin and external defensive barriers and inject venom directly into sensitive tissues. Venoms function as defenses in several ways. First, they cause pain. Second, they can cause body or tissue damage, or death. Beyond these primary activities, venom can have defensive value based on its nasty taste, and, possibly, on its ability to induce allergic reactions. Honey bee venom illustrates the value of a repugnant taste of a venom: presumably birds and other vertebrates reject worker bees while readily devouring drones because of the ability of workers to sting. For example, western king birds selectively prey on drone honey bees, which they manipulate entirely with their hard bill until the prey can be quickly swallowed, head first. The handling and speed of consumption suggested to the author

that a king bird could consume worker bees without being stung. Why then were workers rejected? A suspected, but unconfirmed, explanation was revealed upon consumption of whole frozen drone honey bees and abdomens of workers. Drones were generally palatable. In contrast, worker abdomens were noisome, possessing a bitter and "hydrolytic" taste, a taste confirmed (by sampling of an isolated venom sac) to be venom.

No direct evidence confirms a defensive value for venom allergy. However, three lines of reasoning suggest that such a value might exist. First, the 1 to 4% incidence of allergic reaction to insect stings in the general human population is disproportionately high compared with allergy to other proteinaceous substances. Second, observations of social interactions among human groups in which one individual experiences an allergic reaction to a sting reveal that fear of the culprit stinging insect is shared among the whole group. Primates and other intelligent social mammals, major potential predators of social insects, likely also learn to avoid the stinging insects by observation of the plight of a group member. Third, an allergic reaction impairs an individual's ability to avoid its own predators, to obtain food, and to reproduce as effectively as if no reaction had occurred. Consequently, the genes from an individual that attacks stinging insects and suffers a reaction are not as likely to be passed to future generations. Such "genetic learning" would be a consequence of allergic reactions.

Pain, the most notorious property of an insect sting, is key for instant defensive value of venom. Pain is the biological signal that indicates to an organism that bodily damage has occurred, is occurring, or is about to occur (Fig. 3). Consequently, pain is a signal that must be heeded and acted upon if an animal is to enhance its own survival and fitness. Pain acts to stop an attacking predator and to cause it at least momentarily to assess its own risk situation. In the meantime, the stinging insect gains valuable time for escape or further defense. Because pain cannot be measured easily or precisely, exact values to use for comparing the effects of different venoms on potential predators are unobtainable. To provide a means for comparison, Schmidt developed a pain index, based on a scale of 0 to 4. Zero indicates that the stinging insect is too small or otherwise unable to pierce the human skin; 2 is the central value, as represented by the pain of a honey bee sting; and 4, the top of the scale, is typically described as causing immediate, excruciating, totally debilitating pain that completely eliminates the ability of the stung individual to continue to act in a normal fashion. Table I lists some of the pain values for common insects.

Pain, in itself, is not an outright defense—it can be a bluff, mimicking damage and deceiving a would-be predator into believing that it has been injured. Unless confirmation of injury follows, an intelligent animal quickly learns to ignore the pain and continue with its attack. Beekeepers provide a familiar example of the principle—upon learning that bee sting pain does not translate into actual damage or risk, they continue their activities, often receiving scores of stings in a day. If damage did occur to beekeepers, the profession would have become extinct long ago. A minority of beekeepers do experience damage in the form of large local or allergic reactions, and these individuals usually abandon the profession. To avoid the limiting problem associated with pain, venoms of many insects have evolved a step further to the stage of "truth in advertising": pain is the advertising; toxicity is the truth. If the

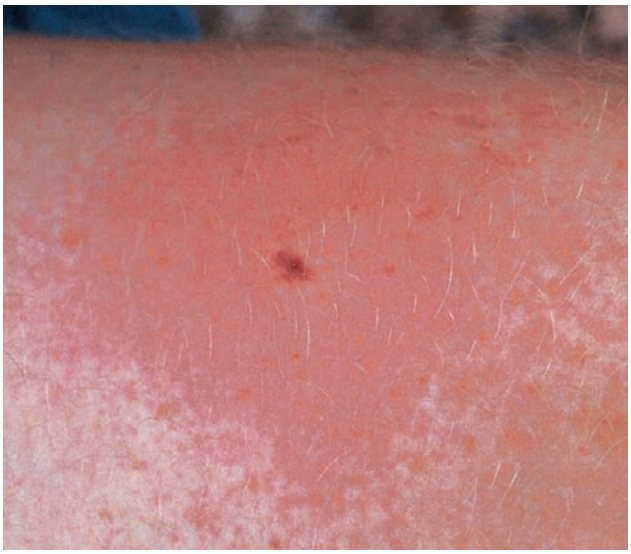

FIGURE 3 Painful local reaction illustrating central tissue damage and surrounding inflammation induced by a honey bee sting. (Photograph by J. O. Schmidt.)

TABLE I Pain Rating and Lethality of Common Insect Venoms[a]

Species	Common name	Pain rating[b]	Lethality[c]
Lasioglossum spp.	Sweat bee	1	n/a
Sceliphron caementarium	Mud dauber wasp	1	n/a
Solenopsis invicta	Fire ant	1	n/a
Sphecius grandis	Cicada killer wasp	1–2	46
Myrmecia gulosa	Bull ant	1–2	0.18
Dinoponera gigantea	Giant ant	2	12
Xylocopa californica	Carpenter bee	2	26
Dolichovespula maculata	Baldfaced hornet	2	6
Bombus impatiens	Bumble bee	2	12
Apis mellifera	Honey bee	2	2.8
Vespula germanica	German yellowjacket	2	2.8
Polistes canadensis	Red paper wasp	2–3	2.4
Dasymutilla klugii	Velvet-ant wasp	3	71
Pogonomyrmex maricopa	Harvester ant	3	0.125
Pepsis formosa	Tarantula hawk wasp	4	65
Paraponera clavata	Bullet ant	4	1.5

[a]For a more complete list, see Schmidt (1990).
[b]Based on a scale ranging from 0 to 4; 4 is the highest score.
[c]Measured as milligrams of venom per kilogram of body weight in mice; the smaller the number value, the more potent the venom.

pain signal is associated with damage, then predators cannot easily bypass the pain signal. Venom toxicity is most easily measured in terms of its ability to cause death in an animal.

Table I also rates the lethality of representative insect venoms. Often, but not always, increased pain parallels increased lethality, hence the tandem evolution of both "advertising" and "truth." Notable venoms on the list are those of the harvester ants *(Pogonomyrmex),* the most lethal insect venoms in the world, which cause rather nasty intense pain that lasts for 4 to 8 h, and bullet ants *(Paraponera),* the world's most painful venomous insects, which also have a respectable lethality. Bull ants *(Myrmecia)* and tarantula hawks *(Pepsis)* present curious cases: bull ants are rather lethal, yet they induce relatively little pain; tarantula hawks cause a most intense pain, equaling that of bullet ants, for 2 to 3 min, yet they possess no vertebrate lethality. The explanation for this discrepancy between pain and lethality is unclear for bull ants, but for tarantula hawks the explanation might be that *Pepsis* must permanently paralyze its spider prey without killing them.

ECOLOGY AND EVOLUTION

Insect venoms have played an enormous role in ecology and evolution. For parasitic and hunting wasps, venoms have provided the means to expand the menu of hosts as food sources. This in turn allowed, and continues to allow, isolation of breeding populations and rapid speciation of the wasps. Via their increased impact, wasps alter the foraging and defensive behaviors of their host species, with a ripple effect insofar as other predators and parasitoids are influenced. The influence of defensive venoms is even greater. The defensive potency of venoms provided venomous insects the freedom to enhance further their defensive strengths through evolution of secondary defenses of aposematism and mimicry. Warning colors, sounds, or chemical signals used in conjunction with the honest threats of damage from the sting provide venomous insects with the ability not only to defend against potential predators, but also to reduce risks of being attacked. After one or a few species develop aposematic signals, other venomous insects can become Müllerian mimics of the signals. The system opens to "cheaters"—which are not noxious—to become Batesian mimics of the venomous models.

For most species, freedom from one or more predators, or reduced predatory pressure, provides wonderful behavioral and evolutionary opportunities. Secretive species may expand their niches into more conspicuous environments or to increased periods of time, thereby reaping the benefits of increased quantity and variety of resources. Examples of this are brightly colored bumble bees, social wasps, and spider wasps that can forage during daylight hours in open environments, free from most threats from predatory birds and lizards. Likewise, venomous ants can exploit daylight periods, open areas, or conspicuous areas in the vegetation.

All other effects of venoms pale in comparison to their impact on the evolution and maintenance of sociality in the Hymenoptera. Social wasps, bees, and ants account for a huge percentage of the individuals and biomass of most ecosystems, especially tropical systems. Social wasps are important predators of other arthropods, social bees are responsible for much of the pollination of plants, and ants are major predators and influences on other animals. Insect sociality evolved through kin selection in which related presocial individuals that cooperated had higher inclusive fitness (the successful passing on of their genes to the next generation) than did individuals that did not cooperate. In Hymenoptera, the evolution of sociality, and its maintenance, was possible, in large part, because of defensive venoms. Social insects and their nests full of defenseless brood and/or food stores represent potential bonanzas for predators capable of exploiting them. As colony size increases, larger and more determined predators will attempt to exploit a colony's brood or food cache. Venoms provided the means, both proximate and ultimate, for colony defense against destruction. Once established, sociality in Hymenoptera could progress from small simple social structure to highly organized structures via the evolution of increasingly potent venoms and other defenses. Populous, highly organized social species dominate and rule most environments.

See Also the Following Articles

Bee Products • Chemical Defense • Mimicry • Predation • Sociality • Wasps

Further Reading

Bettini, S. (ed.). (1978). "Arthropod Venoms," Vol. 48 *In* "Handbook of Experimental Pharmacology," (especially Chaps. 18–26, pp. 489–894). Springer-Verlag, New York.

Piek, T. (ed.). (1986). "Venoms of the Hymenoptera." Academic Press, London.

Quicke, D. L. J. (1997). "Parasitic Wasps," pp. 221–242. Chapman & Hall, London.

Schmidt, J. O. (1982). Biochemistry of insect venoms. *Annu. Rev. Entomol.* **27,** 339–368.

Schmidt, J. O. (1990). Hymenoptera venoms: Striving toward the ultimate defense against vertebrates. *In* "Insect Defenses: Adaptive Mechanisms and Strategies of Prey and Predators" (D. L. Evans and J. O. Schmidt, eds.), pp. 387–419. State University of New York Press, Albany.

Veterinary Entomology

Bradley A. Mullens
University of California, Riverside

Veterinary entomology deals with arthropod pests and vectors of disease agents to livestock, poultry, pets, and wildlife. It is allied with the fields of medical entomology, parasitology, animal sciences, veterinary medicine, and epidemiology. The main pests of veterinary concern are sucking and biting lice, biting flies, nonbiting muscoid flies, bot flies, fleas, and Acari (mites and ticks).

ARTHROPOD GROUPS

Arthropods that affect animals can be categorized by the intimacy of their host association, and these range from permanent ectoparasites to pests that contact the vertebrate only briefly once every few days.

Permanent Ectoparasites

Some arthropods, such as lice and many parasitic mites, complete their entire life cycle on the host. All stages of sucking lice (Pthiraptera, suborder Anoplura) are mammal parasites and feed on blood, whereas biting lice (suborders Amblycera and Ischnocera) use either mammal or bird hosts, feeding on skin, hair, and feather debris. Lice tend to be abundant in cool weather or on animals stressed by poor nutrition or overcrowding. Many lice are specific to one or a few closely related hosts and cannot survive more than one to a few days away from the host. Transmission from host to host is mostly by direct contact.

Parasitic mites (Acari, suborders Mesostigmata, Acaridida, and Actinedida) are found on most groups of birds and mammals. Like lice, most are specific to one host species or a small group of related species. Several genera comprise what are commonly known as "mange mites." *Sarcoptes scabei,* which causes sarcoptic mange, burrows at the surface of the dermis and exists in a number of races that generally are host specific to swine, dogs, and so on. *Demodex* mites, causing demodectic mange, live in follicles and can be important especially in immunocompromised hosts. *Chorioptes* and *Psoroptes,* causing chorioptic and psoroptic mange, include species of considerable importance for cattle, and wild and domestic sheep. The latter mites complete their development at the skin surface and do not actually burrow in the skin, although mites often are covered by scabs and are frequently called scab mites. *Ornithonyssus* mites are blood feeders and are especially important for wild and domestic birds, where populations may reach many thousands per host (Fig. 1). They occupy fur and feathers, travelling to the skin surface to feed regularly.

Some of the more advanced flies are also permanent parasites, and a number of species in three families hardly resemble flies at all because they have secondarily lost their wings (apterous). Members of the dipteran families Streblidae and Nycteribiidae live on bats, whereas members of the Hippoboscidae parasitize various birds and mammals. Some hippoboscids are economically important, such as the sheep ked, *Mallophagus ovinus.* In all these families, the adult female nurtures a single larva within her body until it is mature; this is a very unusual pattern for insects. After the mature larva exits the female, it promptly pupates on the host.

Semipermanent Ectoparasites

The semipermanent ectoparasitic arthropods do not complete the entire life cycle on the host, but they do spend at least several days at a time on a vertebrate. The hard ticks (Ixodidae) attach to feed for several days in each of the life stages. Although some, like the cattle tick *Boophilus,* complete the entire life cycle on a single host, the most abundant and widespread species tend to use a separate host for each stage. In this case the engorged tick falls off, molts, and then finds a new host by crawling up on vegetation and waiting for a passing vertebrate, to which it attaches. This activity is known as questing. Often the larva hatches from an egg and attaches to a small host such as a rodent, whereas the nymph (the stage after the larval molt) or adult may attach to a larger host such as a deer. Examples

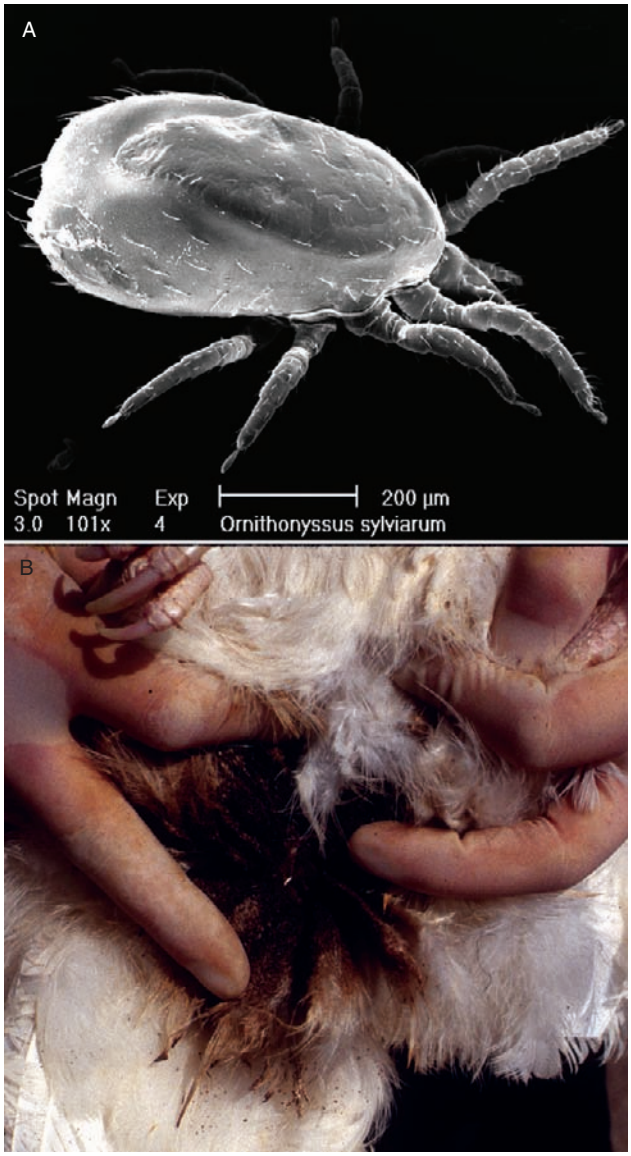

FIGURE 1 (A) Scanning electron micrograph of the northern fowl mite, *O. sylviarum,* a permanent ectoparasite of many birds, including domestic chickens. (B) The vent of a chicken, showing the blackened feathers typical of a heavy infestation. This hen has over 20,000 mites. [Part (A) courtesy of Jeb Owen, University of California, Riverside.]

of this include the American dog tick *Dermacentor variabilis.* Adult female hard ticks take a large blood meal, produce a single large batch of eggs (typically several thousand), and then die.

Fleas (Siphonaptera) generally are on a host for most or all of the adult stage, and feed on blood. About 94% of flea species live on mammals, and the rest on birds. Flea eggs fall from the host pelage into a nest environment, where the larvae feed on organic debris and sometimes on excess blood produced by the adults. Fleas thus are often lacking on hosts that do not return to long-term bedding areas or nests.

Bot flies include important species in the dipteran families Oestridae, Gasterophilidae, and Cuteribridae. They spend nearly the entire year as immatures within the vertebrate's body. Eggs, often laid on hairs, hatch and enter the host body.

Horse bot larvae *(Gasterophilus)* attach to the wall of the gastrointestinal tract for several months before they pass from the host with feces to pupate in soil. Cattle grubs *(Hypoderma)* migrate through the lining of the esophagus or spinal cord, depending on the species, and eventually form a cyst in the back. After a final period of maturation, larvae exit the cyst and fall into the soil to pupate. Bot fly adults lack functional mouthparts and depend entirely on reserves from the larval stage to sustain them for several days; in this brief time they must find mates and hosts. The invasion of vertebrate tissues by fly larvae is known as myiasis.

Certain other flies also are semipermanent parasites. Among muscoid Diptera, the horn fly, *Haematobia irritans,* is on cattle most of its adult life, where adults of both sexes feed 20 to 30 times daily. They leave the host to disperse and to lay eggs in very fresh dung, and then return to the host.

Occasional Parasites

The broad category of occasional parasites includes a range of arthropods. The most intimate host associations in this group include the soft ticks (Argasidae) and some blood-feeding mites such as the bird parasite *Dermanyssus gallinae.* In both cases, nymphs or adults hide in or near nest areas, sheltered in cracks and crevices or under debris; thus, they are closely associated with animals, although they contact them only periodically. They leave the hiding places, often at night, to feed for periods of 5 to 30 min, returning to the nest to digest the blood meal. Some soft ticks can withstand dry conditions and years without feeding.

Many serious biting fly pests spend the immature period away from the host, exploiting an entirely different resource base. It is common for larvae to feed on detritus in wet habitats, whereas the adults use plant nectar for energy and blood for egg development. For example, larvae of stable flies live in rotting vegetation, blackflies in running water, horse flies and biting midges in swampy mud, or mosquitoes in ponded or slowly moving water. When the adults emerge, they take blood meals at intervals of 1 to 4 days, but they typically are in contact with the host for only a few minutes at a time.

FIGURE 2 The stable fly, *Stomoxys calcitrans,* may attack animals in high numbers: there are approximately 200 flies feeding on the front leg of this bull. This species is illustrative of the occasional parasites because the adult flies blood-feed for only a few minutes every day or two, and larvae are in rotting vegetation away from the host animal.

They leave the host to digest the blood in some sheltered resting location. For the higher Diptera (suborder Cyclorrhapha), such as stable flies (Fig. 2) or tsetse flies, both sexes feed on blood, and multiple blood meals usually are needed to develop a batch of eggs (or, for tsetse, a single mature larva). In the lower Diptera (suborders Brachycera and Nematocera) such as horse flies, blackflies, or mosquitoes, only females take blood, and most species require only a single large blood meal to develop an entire batch of eggs numbering 50 to 300. Some other pests in this general category do not feed on blood but visit the host to take meals of tears or other protein-rich secretions that are also used to develop eggs. A good example is the face fly, *Musca autumnalis.*

HOW ARTHROPODS CAUSE DAMAGE TO ANIMALS

There are several basic ways in which arthropods cause damage to animals, and the different mechanisms interact to impact agricultural production. Arthropod damage to plant crops often is evident to consumers, who react with disgust to cabbage leaves damaged by loopers, corn earworm larvae on an ear of corn, or scale insects on citrus fruit. In contrast, arthropod damage to animal production tends to be hidden from the consumer because the product is purchased in the form of jugs of milk or wrapped packages of butchered meat. Nevertheless, losses are serious for producers, and costs are passed on to consumers.

Loss of Blood and Tissue Fluids

Many arthropods ingest blood, usually for egg development. Impact of blood loss on the animal reflects the style of feeding and the number of arthropods. For example, mosquitoes canulate a vessel and hosts generally lose only the blood mosquitoes ingest. In contrast, not only do biting flies that macerate the capillary beds of the skin to feed from blood pools tend to inflict more painful bites, but the bites themselves lead to a larger quantity of blood loss per feeding insect. For example, horse flies (Tabanidae) may directly remove over 200 ml day^{-1} from a host in a pasture, but blood continues to run from the wound for a period of time, often being fed on by other flies. Many hard ticks (Ixodidae) increase in weight by 100-fold or so as they feed for 7 to 10 days. Females of very large species of ticks may contain over a milliliter of blood at a time, and hundreds may be attached to a single host animal. Blood or fluids such as lymph are metabolically "expensive" for a vertebrate to produce. Such loss is reflected in reduced feed conversion efficiency, which occurs when animals eat more for a given yield of meat or eggs. Insect feeding also causes significantly lower weight gains or milk production. Losses of 10 to 20% feed conversion, 0.1 to 0.2 kg day^{-1} weight gain (cattle), and 5 to 10% loss in milk yield are not uncommon for animals under heavy attack.

Pain and Interference with Activities

Pain and irritation caused by arthropod attack force animals to alter their feeding or activity patterns and to engage in a number of sometimes vigorous behaviors to defend themselves. There may be economic loss as well, since animals are not feeding normally and must expend energy that might otherwise be directed toward growth or reproduction. For example, stable flies *(Stomoxys)* and face flies, as well as other biting flies, can cause animals to retreat into groups for refuge. In groups, insect attack rates usually are lower per host (the herd dilution effect), particularly for the animals that occupy the interior of an aggregation. Animals also may enter woods or bodies of water in an apparent attempt to escape insects. Cattle pursued by adults of cattle grubs *(Hypoderma)* experience no immediate pain from the flies, which cannot bite. The female flies are trying merely to deposit eggs on the cattle hair at the base of the legs. Still, cattle exhibit an interesting, stereotypical behavior known as "gadding" (Fig. 3). The animals run at full speed with tails raised straight into the air, which expends energy and may cause accidental injury.

Although pests such as house flies may not cause direct losses to the animals, they are produced near animal operations and thus are a veterinary entomology problem. Excessive numbers ("excessive" admittedly is a subjective term) of nuisance arthropods cause great annoyance to people living nearby, and thus constitute serious public relations and legal problems for producers. Public health agencies can close facilities unable or unwilling to mitigate such a problem.

FIGURE 3 "Gadding" behavior (note the tail held up in the air) by a calf being attacked by cattle grub flies (*Hypoderma* spp.). Vertebrate host behavior can be altered by parasites. (Photograph courtesy of Dr. Jerry Weintraub, Agriculture Canada.)

Allergic Responses to Saliva

Blood-feeding arthropods possess a potent arsenal of chemicals in their saliva to maintain blood flow (vasodilators, anticoagulants) and sometimes have anesthetics to reduce host defensive response. Like humans, animals can develop allergies to these compounds. Horses commonly react to biting midge *(Culicoides)* feeding with an allergic reaction called Queensland itch or sweet itch, resulting in skin inflammation and hair loss. Mass emergences of the blackflies *Simulium arcticum* in Canada and *Cnephia pecuarum* in the valley of the southern Mississippi River have resulted in the deaths of livestock, probably from allergic responses as well as blood loss. Larvae of sheep blowfly *(Lucilia)* feed near the skin surface, especially where the wool is wet, and can contribute to a toxic shock–type syndrome fatal to infested sheep. Pets may develop serious allergies to fleas, with resulting hair loss and other symptoms.

Product Damage

Arthropods sometimes cause direct damage to parts of the animal desired by people. For example, cattle grub *(Hypoderma)* larvae form large cysts in the backs of cattle. They cut a hole in the skin to breathe, and this skin is the thickest on the animal. Although the holes heal after the larva exits, the scarred skin is less valuable for leather. The presence of larvae also can affect the quality of the meat in this area of the animal, which is the part where steaks come from, and damaged meat sometimes must be trimmed at the slaughterhouse. Mites such as *Psoroptes* and *Sarcoptes,* as well as many lice, often result in irritation, rubbing, and gross loss or damage to hair and wool.

Cosmetic damage, including rashes or minor hair loss, can be predictably serious to the owner of a pet or a show animal. However, cosmetic damage also can cause losses in animal agriculture out of proportion to actual damage. An example

of this is the condition "gotch ear" in cattle caused by the Gulf Coast tick *Amblyomma maculatum* in the southern United States. Damaged ear cartilage is cosmetic, but causes the animals to be placed in an "odd lot," with per-pound prices of $0.05 to 0.10 less than undamaged cattle.

Restricted Trade

Many pests have distinctive distributions, and preventing movement or dispersal into new areas is of paramount importance. Cattle ticks *(Boophilus)* and screwworm *(Cochliomyia hominivorax)* were eradicated from the southern United States in the 20th century, but they persist in Central and South America. The U.S. habitat obviously is still suitable. Without complex systems of animal quarantine, treatment, and examination, it is certain they would reestablish in the United States.

Exotic arthropods pose a great threat either as direct pests or vectors of disease agents such as those that cause heartwater or African swine fever. The Office International des Epizooties lists diseases of risk for animals worldwide, and one of those on List A (greatest risk) is bluetongue. This viral disease of ruminants, such as cattle and sheep, is transmitted by biting midges, and is endemic in the United States. Trade restrictions from bluetongue cost the U.S. cattle industry many millions of dollars annually, even though cattle themselves do not usually develop obvious disease. The reason is that some major trading partners (e.g., western Europe) lack bluetongue, and their agricultural authorities fear an impact on their sheep industries from accidental importation.

Diseases

A number of serious animal disease agents are transmitted by arthropods. The worst of these are tropical, and they cause death and heavy production losses in the affected countries. African trypanosomiasis causes a wasting-type disease known as nagana in animals and sleeping sickness in humans, and *Theileria parva,* called East Coast fever, can cause 90 to 100% mortality in affected cattle in eastern Africa. People in developed countries tend to underestimate the true value of animals in the developing world. Animals are vital there not only for protein-rich food, but for draft and transportation purposes, and as wealth. They are the basis of many pastoral peoples' economies, and the economic impact of some of these animal diseases can far exceed the impact of similar, serious human pathogens.

Temperate zones also have some rather important arthropod-transmitted animal disease agents, including *Anaplasma,* dog heartworm, and equine infectious anemia virus. In the United States and Europe, the direct effects of arthropods on animal production generally exceed losses caused by arthropod-transmitted diseases. However, the role of wild animals as natural reservoirs of pathogens that incidentally infect people is very important in both temperate and tropical zones. Diseases that cycle naturally in animal populations and occasionally infect people are called zoonoses. Zoonoses comprise some of the more notorious arthropod-related human health problems. They include plague (maintained in rodents and transmitted by fleas), Lyme disease (maintained in rodents and transmitted by ticks), and St. Louis encephalitis (maintained in birds and transmitted by mosquitoes). Previously unknown tick-borne ehrlichioses (caused by intracellular bacteria-like organisms in the genus *Ehrlichia*) have been recently discovered infecting humans in the United States. They are zoonotic in origin and typify a category of "emerging" human diseases that is now of great interest in the medical community.

See Also the Following Articles

Bubonic Plague • Dog Heartworm • Medical Entomology • Mites • Tsetse Fly • Zoonoses, Arthropod-Borne

Further Reading

Harwood, R. F., and James, M. T. (1979). "Entomology in Human and Animal Health." Macmillan, New York.

Kettle, D. S. (1995). "Medical and Veterinary Entomology." 2nd ed. CAB International, Wallingford, Dxon, U.K.

Lancaster, J. L., and Meisch, M. V. (1986). "Arthropods in Livestock and Poultry Production." Ellis Horwood, Chichester, U.K.

Mullen, G., and Durden, L. (eds.). (2002). "Medical and Veterinary Entomology." Academic Press, San Diego, CA.

Wall, R. and Shearer, D. (1997). "Veterinary Entomology." Chapman & Hall, London.

Williams, R. E., Hall, R. D., Broce, A. B., and Scholl, P. J. (1985). "Livestock Entomology." Wiley, New York.

Vibrational Communication

Andrej Čokl and Meta Virant-Doberlet

National Institute of Biology, Slovenia

Many insects communicate by vibratory signals that are produced directly by a body part or, indirectly, as a by-product of some other activity. Insects emit vibratory signals in connection with aggression, distress, calling, courtship, rivalry, and other specific behaviors, among other functions enabling mate location and recognition. This article presents mechanisms, structures, and signals of vibratory communication in relation to behavior.

MECHANISMS OF VIBRATION PRODUCTION

Insects produce vibratory signals by percussion, vibration of a body part, tymbal mechanisms, or stridulation. Percussion is a very common because of the hard exoskeleton, which enables either percussion of two body parts or striking against a substrate. The percussive structures are in most cases relatively simple. Book lice (Psocoptera) and stoneflies (Plecoptera) tap their abdomens against the substrate, Orthoptera use their legs, and termites and beetles drum with their heads. Low-

frequency signals also can be emitted by vibration of some body part(s). For example, male and female chloropid flies *Lipara* communicate over distances of more than 2 m on a reed with signals produced by vibration of the abdomen. Lacewings (Chrysopidae) oscillate the abdomen without touching the substrate, and in this way shake the stem or leaf on which they are standing with the low-frequency component of the broadband signal being used for communication. Abdominal muscles are involved in song production also in the Hawaiian fly *Drosophila sylvestris.* Bees produce vibratory signals by thorax vibrations; "begging" signals are transmitted to the substrate through the legs and "tooting" and "quacking" signals directly by pressing the thorax to the substrate.

Tymbal-like mechanisms are used to produce vibratory signals in most cicadas, small plant bugs, and hoppers. Because there are no special resonant air sacs behind the tymbal, the carrier frequency remains low and suitable for transmission through a substrate. Many insects produce vibratory signals by friction of two body parts moved one across another. For example, ants stridulate by means of a file on the dorsal surface of the first abdominal segment and a plectrum on the posterior edge of the metathorax or postpetiole. The substrate-borne component of the audible signal is used for communication. Burrower bugs (Cydnidae) communicate with substrate-borne signals produced with the tergal plate, which functions simultaneously as tymbal, as a file with its latero-frontal surface, or as a plectrum for the alary stridulitrum on the postcubital vein.

Vibratory signals are produced by different insect activities. An insect singing on or close to a substrate produces signals with airborne and substrate-borne components, and many of them use both simultaneously or alternatively. Fruit flies communicate with near-field sound radiated from wings and with substrate-borne signals, which are simultaneously produced by muscles driving the wings. Soil buzz vibrations and sounds produced by solitary bees and wasps are used as cues for localization.

TRANSMISSION OF VIBRATIONS

An insect standing on a substrate may be represented by a model of a mass on six springs. The bodies of treehoppers, spiders, and even fiddler crabs respond to substrate vibration with resonance at lower frequencies and attenuation at higher frequencies. Legs transmit body vibrations to the substrate and are the seat of most sensitive vibrational receptors. Transmission over the legs depends on their stiffness, and the response properties of the body–legs oscillating system varies in different species with posture changes. Signal amplitude decreases during transmission to the substrate, mainly because the body also vibrates during singing in a horizontal plane.

Plants are widely used as transmission channels for insect vibrational songs. Although quasi-longitudinal waves and waves of some other types cannot be excluded, communication signals are carried by bending or flexural waves with only a little frictional loss in energy and particle motion, both in the longitudinal direction and in a plane perpendicular to the direction of wave propagation and to the surface. Their propagation velocity varies little with the structure's mechanical properties and is proportional to the square root of the structure's radius and the square root of the frequency. Vibratory signals reflect mainly at the root and top of the plant and may travel back and forth several times. In standing wave conditions with a complicated and frequency-dependent pattern, it is a better strategy for communication to use broadband and not pure tone or narrowband signals.

There are many examples that demonstrate how well the spectra of vibratory signals fit the filtering properties of host substrates. In plants, lowest attenuation with distance occurs with signals of frequencies around 100 Hz. Spectra of the songs of the southern green stink-bug, *Nezara viridula,* have a narrow dominant frequency peak between 80 and 160 Hz (Fig. 1) and the main energy of the broadband "small cicadas" and cydnid signals is emitted at lower frequencies. Carpenter ants, *Camponotus herculeanum,* live in tree trunks from which the soft spring wood has been eaten out, leaving thin, lignified lamellae. The alarm signals, produced by drumming the head and gaster against the substrate, are particularly suited for transmission through the nest but not through to the outside. "Begging signals," emitted by bees following a dancer, are used to solicit food samples. The 1-μm peak amplitude of the signals lies slightly above the threshold for the bee "freezing response," and the signal dominant frequency around 320 Hz lies in the range of the best signal-to-noise ratio of the bee comb.

RECEPTOR MECHANISMS

The most sensitive and specialized receptor for substrate vibrations, the subgenual organ, is derived evolutionarily from chordotonal organs in the body and appendage joints. The organ lies in insects in all the six legs close to the main tracheal trunk and almost completely occludes the dorsal blood sinus of the proximal tibia. The organ is supported proximally by the nerves and distally by accessory cells that form a ligament attached most often to the cuticle of the leg. The structural unit of the organ is the scolopidium, each of which is composed of one or more bipolar sensory cell(s) and two accessory cells, the scolopale and the cap (attachment) cell.

In Orthoptera, the subgenual organ is accompanied by other scolopal organs of which the crista acoustica, supported by sound-transmitting structures, represents a very sensitive receptor organ for airborne sound. In the American cockroach, *Periplaneta americana,* the subgenual and the neighboring Nebenorgan (Fig. 2) are most sensitive to substrate vibrations. The distal organ, which surrounds the hemolymph, responds to variations in hemolymph pressure. Scolopidia of the subgenual organ are organized in a fan-shaped manner and converge radially toward a single insertion point so that vibrations of the substrate cause tension changes. The organ is most sensitive in the frequency range between 1000 and 5000 Hz, with threshold amplitudes around 0.1 nm. Scolopidia of the Nebenorgan

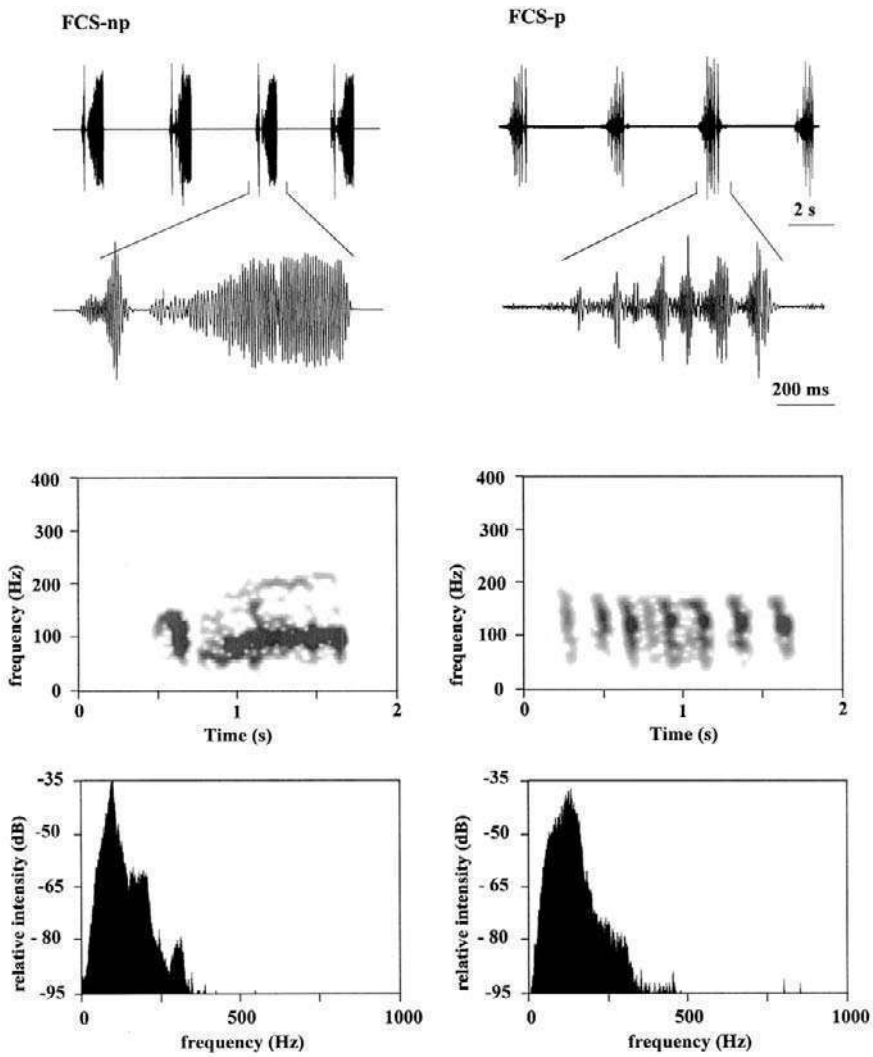

FIGURE 1 Oscillograms (top), sonograms (middle), and power spectra (bottom) of the nonpulsed (left) and of the pulsed (right) type of *N. viridula* female calling song. [After Čokl, A., Virant-Doberlet, M., and Stritih, N. (2000). The structure and function of songs emitted by southern green stinkbugs from Brazil, Florida, Italy, and Slovenia. *Physiol. Entomol.* **25,** 196–205, Blackwell Science Ltd., with permission.]

are stretched perpendicular to the long axis of the tibia so that the organ reacts to tension changes within the scolopale caused by vibration of the cuticular walls.

Subgenual organs of different morphology have been described in other insect groups. In the southern green stinkbug, the subgenual organ is proximally attached to the epithelium of the tibial wall, whereas the two scolopidia with the cilia and the flat and thin ligament are stretched out in the hemolymph of the blood channel. The subgenual organ in the legs of the lacewing *Chrysoperla carnea* (Fig. 3) is composed of only three scolopidia. The cell bodies of the three cap cells form a lenslike part of the organ, the velum, which distally divides the blood channel in two separate parts. Scolopidia are attached to the middle of the velum and extend to the dorsal leg wall. The hollow cone-shaped bee subgenual organ with approximately 40 scolopidia is connected at two points to the cuticle and at two points to the membrane bag surrounding the organ and the membrane lining the tracheal wall. The organ oscillates with the hemolymph, and the sensory cells respond to displacements of the organ relative to the leg.

Mechanoreceptors like campaniform sensilla, joint chordotonal organs, and Johnston's organ respond to substrate vibrations with lower sensitivity, preferentially in the frequency range below 100 Hz.

VIBRATORY SIGNALS AND INSECT BEHAVIOR

Insects emit vibratory signals in connection with aggression, distress, calling, courting, rivalry, and many other behaviors. In many habitats, vibrational communication represents the only useful way to exchange information. Social insects like bees, ants, and termites live in nests where the possibility of communication by signals of other modalities (other than airborne chemicals) is limited. Insects, whose small body size

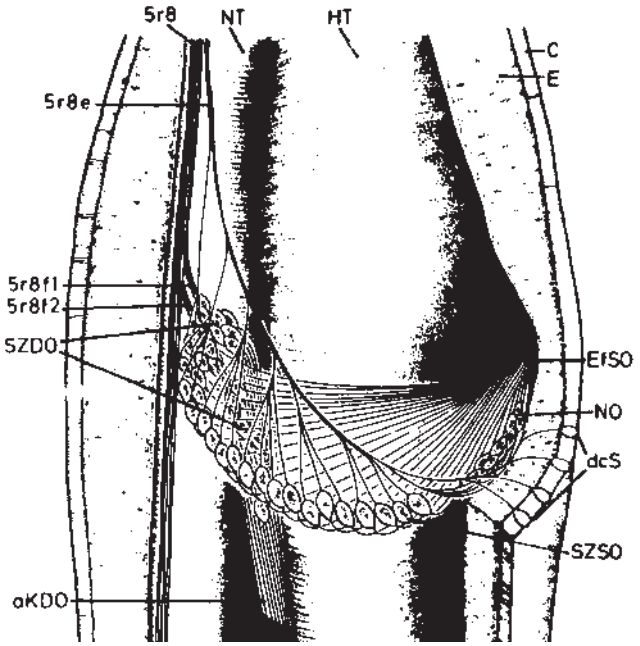

FIGURE 2 Subgenual organs of the cockroach *P. americana.* The left tibia is opened to show the spatial arrangement of the scolopidial organs. aKDO, accessory cap cells of the distal organ; C, cuticle; dcS, distal campaniform sensilla; E, epidermis; EfSO, terminal fila of the subgenual organ; HT, main trachea; NO, Nebenorgan; NT, small trachea; SZDO, sensory cells of the distal organ; SZSO, sensory cells of the subgenual organ; 5r8e, nerve innervating the subgenual organs and the campaniform sensilla. [After Schnorbus, H. (1971). The subgenual organs of *Periplaneta Americana:* histology and thresholds for vibration stimuli. *Vgl. Physiol.* **71**, 16. Springer-Verlag GmbH & Co. KG, with permission.]

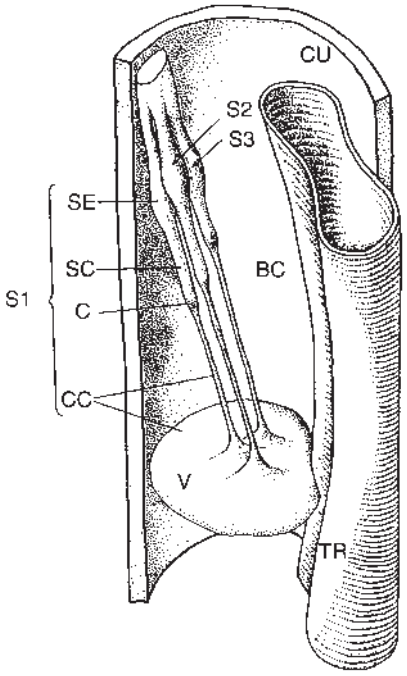

FIGURE 3 Three-dimensional reconstruction of the subgenual organ of the left middle tibia of the lacewing *C. carnea:* BC, blood channel; C, cap; CC, cap cell; CU, cuticle; S1, S2, S3, three scolopidia; SC, scolopale cell; SE, sensory cell, TR, trachea; V, velum. [After Devetak, D., and Pabst M. A. (1994). Structure of the subgenual organ in the green lacewing, *Chrysoperla carnea. Tissue and Cell* **26(2),** 249–257, Harcourt Inc., with permission.]

does not allow efficient low-frequency sound radiation, can use vibrational communication without attracting predators or parasites as they would by singing. Many insects use the vibratory component of the emitted airborne signals to improve signal discrimination and recognition.

On the other hand, vibrational communication through plants is disturbed by environmental factors such as wind and raindrops. For example, wind induces low-frequency vibrations that are accompanied in apple (but not banana) leaves with broad-band vibrations of spectra up to 25 kHz. Raindrops falling on banana leaves produce vibrations up to 1000 Hz; those on apple leaves cause vibrations composed of an irregular high-frequency phase and a regular low-frequency phase. Signal-to-noise ratio is enhanced by the highly ordered temporal structure of songs with narrower spectra with dominant frequencies, usually above the low-frequency noise level, as well as by band-pass filtering properties of the subgenual organ.

The time pattern of a vibration pulse series usually carries more specificity information than its spectral structure. Generally, a rapid signal divergence occurs in sympatric taxa, and song similarity is expected only in allopatric or allochronic species. In *Chrysoperla,* tremulation songs represent the best cue for species identification within the *carnea* group. Within the same region, the songs of different cryptic species differ structurally and functionally more than in pairs of species from North America and Eurasia. The vibratory song repertoire of sympatric Californian species *N. viridula* and *Acrosternum hilare* are different, and interspecific mating has not been observed. In Japan, interspecific mating was described between the sympatric species *N. viridula* and *N. antennata,* although songs, which correspond to definite behavioral features, are different.

Vibratory signals are used by insects also as cues for mate, prey, or enemy localization. In *N. viridula,* male directional movement on a plant results from female calling song signals. Leafcutting ant, *Atta cephalotes*, workers stridulate when they cut leaves, and nearby ants respond by orienting toward the source of the vibrations and join in the leafcutting. The response of the heteropteran predator *Podisus maculiventris* to vibrational signals produced by a common prey species demonstrates that these predators are capable of using substrate-borne vibrations as cues for prey location. The male katydid, *Conocephalus nigropleurum,* shakes its body to produce vibratory signals that attract females during courtship and mating.

Central nervous resolution of stimulation time differences and amplitude gradients, which occur when vibration travels through a substrate from one leg to another, is the mechanism that enables directionality in larger insects. For example, the response of the body of the treehopper *Umbonia crassicornis*

relative to the substrate reveals resonance at lower frequencies and attenuation at higher frequencies. The transfer functions measured on the body differ substantially depending on whether the stimulus originates in front or behind, indicating that directional information is available in the mechanical response of the body to substrate vibration. Comparison of signal amplitudes between receptors of the front and back legs might enable vibration localization in smaller insects.

See Also the Following Articles

Hearing • Mechanoreception

Further Reading

Bailey W. J. (1991). "Acoustic Behaviour of Insects. An Evolutionary Perspective." Chapman & Hall, London.

Barth, F. G. (1998). The vibrational sense of spiders. *In* "Comparative Hearing: Insects (R. R. Hoy, A. N. Popper, and R. R. Fay, eds.), pp. 00–00. Springer-Verlag, New York.

Devetak, D. (1998). Detection of substrate vibration in Neuropteroidea: a review. *Acta Zool Fennica* **209,** 87–94.

Dusenbery, D. B. (1992). "Sensory Ecology: How Organisms Acquire and Respond to Information." Freeman, New York.

Ewing, A. W. (1989). "Arthropod Bioacoustics: Neurobiology and Behaviour." Edinburgh University Press, Edinburgh, U.K.

Gogala, M. (1985). Vibration producing structures and songs of terrestrial Heteroptera as systematic character. *Biol. Vesn.* **32,** 19–36.

Henry, C. S. (1984). The sexual behavior of green lacewings. *In* "Biology of Chrysopidae." (M. Canard, Y. Semeria, and T. R. New, eds.), pp. 00–00 Junk Publishers, The Hague.

Kalmring, K., and Elsner, N. (1985). "Acoustic and Vibrational Communication in Insects." Verlag Paul Parey, Berlin.

Markl, H. (1983). Vibrational communication. *In* "Neuroethology and Behavioral Physiology. Roots and Growing Points." (F. Huber, and H. Markl, eds.), pp. 00–00. Springer-Verlag, Berlin.

Michelsen, A., Fink, F., Gogala, M., and Traue, D. (1982). Plants as transmission channels for insect vibrational songs. *Behav. Ecol. Sociobiol.* **11,** 269–281.

Shaw, S. R. (1994). Detection of airborne sound by a cockroach "vibration detector": a possible missing link in insect auditory evolution. *J. Exp. Biol.* **193,** 13–47.

Vision

see *Eyes and Vision*

Vitellogenesis

William H. Telfer

University of Pennsylvania

Vitellogenesis is the process by which yolk accumulates in the cytoplasm of an ovarian oocyte. It is one of the final stages of egg formation, occurring just prior to deposition of the chorion. Studies on vitellogenesis have focused during the last 50 years on the major protein yolk precursor, vitellogenin, its synthesis in the fat body, its transport to the oocyte, its sequestration by receptor-mediated endocytosis, and the developmental and hormonal mechanisms that control these processes.

VITELLOGENIN SECRETION BY THE FAT BODY

Before vitellogenesis can begin, the female fat body must transform from a tissue that supports molting, metamorphosis, and intermolt metabolism to one that can secrete vitellogenin and other yolk precursors. When the fat body has reached a requisite stage of maturity and hormonal stimulation, vitellogenin genes are transcribed that encode a polypeptide chain whose molecular weight in many insects is over 200,000. During transport through the endoplasmic reticulum and Golgi bodies, this polypeptide is clipped by endoproteases that remove a secretory signal from its N-terminal end and divide it into shorter polypeptide subunits. The latter differ in number from one in some Hymenoptera to nine in some Hemimetabola. As it moves through the secretory pathway, the complex of subunits is conjugated at sequence-specific sites with high-mannose oligosaccharides, phosphate, and lipids.

Amino acid sequences determined for vitellogenins from at least seven orders of insects are sufficiently similar to indicate a common ancestry. They are members of a protein superfamily that also includes the yolk proteins of vertebrates and nematodes, lipid transport proteins in the blood of both vertebrates and insects, and receptor proteins involved in the endocytosis of lipoproteins.

The yolk proteins of cyclorrhaphan Diptera are exceptional in that their amino acid sequences resemble those of a family of digestive enzymes, the vertebrate lipases. They nevertheless behave like conventional vitellogenins in being synthesized and conjugated in the female fat body, and deposited in cytoplasmic vesicles by the oocyte after endocytosis. A notable difference from conventional vitellogenins is that the several yolk polypeptides in a species are encoded by separate genes, rather than being proteolytic fragments of one gene product.

Other fat body products may supplement vitellogenin in the yolk. The eggs of several Lepidoptera contain lipophorin, a hemolymph protein whose functions include the delivery of lipids from the fat body to other tissues. In the yellow fever mosquito, *Aedes aegypti,* two proproteases are synthesized in the fat body and deposited in the eggs, where they are converted to active proteases during embryogenesis. Endocytosis of hemolymph proteins that bind iron, calcium, heme groups, or biliverdin concentrates these ligands in the yolk of select species. Some of these supplementary proteins are secreted in synchrony with and under the same hormonal control as vitellogenin. Others, like lipophorin, serve somatic functions that require their presence in the hemolymph of males as well.

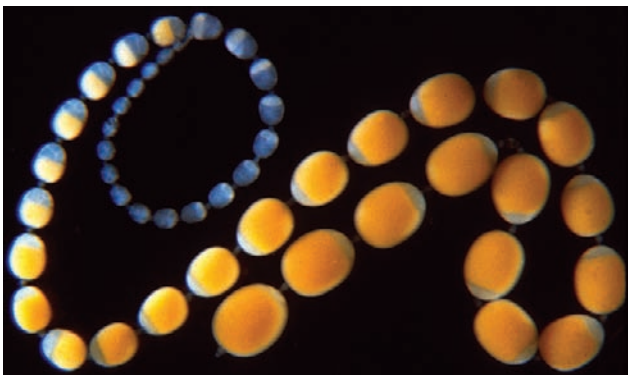

FIGURE 1 Reflected light micrograph of a vitellogenic ovariole from the saturniid moth *H. cecropia.* The largest 36 follicles are vitellogenic. The yellow mass in each follicle is a yolk-filled oocyte, whose opacity is due to light scattering by yolk particles. The yellow color is due to carotenoids carried by vitellogenin and lipophorin, the two most abundant yolk proteins in saturniids. Nurse cells form a transparent cap at one end of each follicle. The epithelium of follicle cells surrounding each oocyte and its nurse cells is too thin to be readily visible, except where it forms a connection between successive follicles. Scale: largest follicle is about 1.9 mm long.

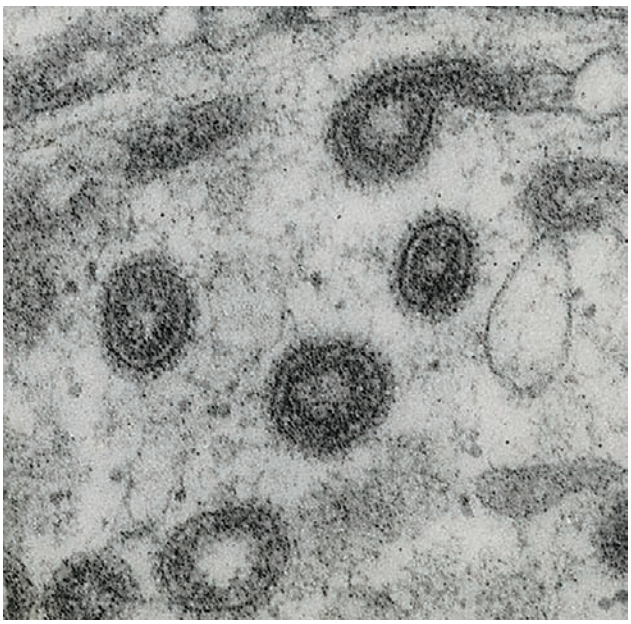

FIGURE 2 Electron micrograph of several endocytotic vesicles near the surface of a vitellogenic oocyte from *H. cecropia.* The bristlelike outer coat is the clathrin lattice that generates the force used to bend coated surface membranes into cytoplasmic vesicles. Receptor-bound vitellogenin is included in the thick layer of granular material that lines each vesicle. An incipient vesicle on the upper right is still attached to an infolding of the cell membrane. The average diameter of coated endocytotic vesicles is about 0.15 mm.

VITELLOGENIC FUNCTIONS IN OVARIAN FOLLICLES

In the ovaries, the morphological unit of vitellogenesis is a follicle—a single oocyte surrounded by an epithelium of somatic cells (the follicle cells) and associated in many insects with a set of modified germ cells (the nurse cells) (Fig. 1). All three cell types are necessary for vitellogenesis, but they contribute to it in very different ways.

Follicles begin to form during or shortly after metamorphosis. They are produced in linear chains termed ovarioles. Within each ovariole is a developmental gradient of follicles (Fig. 1), with the most mature one lying close to the beginning of the oviduct. A common pattern among cyclic egg producers is for only one follicle at a time in each ovariole to form yolk. The penultimate follicle is retarded in its development until the next reproductive cycle. In noncyclic insects, many follicles in each ovariole may simultaneously form yolk (Fig. 1).

Follicle Cells, Patency, and Secondary Yolk Proteins

At the onset of vitellogenesis, the follicle cells develop a system of intercellular spaces that give proteins from the hemolymph access to the surface of the oocyte. In the bug *Rhodnius prolixus,* patency has been attributed to cellular protrusions whose cytoskeletal elongation pushes neighboring follicle cells apart. In the cecropia moth, *Hyalophora cecropia,* osmotic shrinkage of the follicle cells is crucial. Whichever mechanism applies, the intercellular spaces are under tight developmental control: they arise at the onset of vitellogenesis and close when it terminates.

In addition, the follicle cells of many insects secrete proteins that are endocytosed by the oocyte along with vitellogenin. These products may resemble vitellogenin in size, antigenicity, and amino acid sequence; examples occur in Thysanura, Heteroptera, and Coleoptera. A similar relationship holds for the yolk proteins of the cyclorrhaphan Diptera. In Lepidoptera, follicle cell products are instead lipaselike sequences reminiscent of the yolk proteins of the Cyclorrhapha. A few exceptions are known in which either fat body or follicle cells but not both secrete precursors for the protein yolk.

Finally, the follicle cells connect to the oocyte during vitellogenesis via gap junctions that permit direct cytoplasm-to-cytoplasm transfer of ion currents and small organic molecules. These junctions have the potential to function in intercellular exchange of signaling substances such as cyclic nucleotides and calcium ions.

Nurse Cells and the Origin of Egg Cytoplasm

In all holometabolous and a few hemimetabolous orders, the oocytes connect via cytoplasmic bridges to nurse cells (Fig. 1). The bridges are wide enough to permit passage of ribosomes and membranous organelles such as mitochondria. Once believed to be the site of yolk production, nurse cells are now known to be the primary source of egg cytoplasm. They support vitellogenesis by providing the oocyte with the ribosomes, transcripts, and metabolic machinery needed to synthesize the receptors, structural proteins, and enzymes necessary for yolk deposition. In Orthoptera, Blattodea, and other Hemimetabola

that lack nurse cells, the requisite transcripts are produced instead within the oocyte's own nucleus by amplified nucleoli and lampbrush chromosomes.

Receptor-Mediated Endocytosis in the Oocyte

The surface of the vitellogenic oocyte contains receptors that can selectively bind vitellogenin and other yolk precursors after they have penetrated the spaces between the follicle cells. The receptors form transmembrane associations with clathrin lattices on the cytoplasmic side of the membrane. The membranes containing these complexes then fold inward to form endocytotic vesicles (Fig. 2). In subsequent processing steps, vitellogenin dissociates from its receptors within the vesicles, and the clathrin is released from the lattices on the outside. The denuded vesicles transfer their cargo of yolk precursors to neighboring yolk bodies by membrane fusion. Receptors, clathrin lattices, and extra membrane are recycled to the oocyte surface for more rounds of endocytosis. In some insects, vitellogenin is later modified in the yolk to a less soluble storage form known as vitellin.

In the other two classes of yolk particles, glycogen is synthesized inside the oocyte from hemolymph-derived sugars, whereas lipid droplets are assembled from precursors carried from the fat body to the oocyte by lipophorin. Because the lipid droplets contain primarily triacylglycerols and lipophorin transports primarily diacylglycerols, enzymatic conversions must take place during the transfer.

CONTROL BY JUVENILE HORMONE AND ECDYSONE

In many insects, juvenile hormone secreted by the corpora allata stimulates the fat body of adult females to initiate the synthesis of vitellogenin. It may also promote patency of the follicle cells, hence making vitellogenin in the hemolymph available to the oocyte for endocytosis. Juvenile hormone, whose power as an inhibitor of metamorphosis requires that its secretion be reduced during pupation and adult development, has thus evolved new kinds of role in adults: it has become an effector of the neuroendocrine networks that synchronize vitellogenesis with feeding, the photoperiod, and mating.

Variations occur on this theme. In species such as stick insects, and native and domestic species of silk moths that complete egg formation during adult development, juvenile hormone is not required for vitellogenesis. In its classical role as an inhibitor of metamorphosis, the hormone may even prevent the fat body and ovaries from completing their essential previtellogenic development.

Another kind of variation occurs among Diptera, whose yolk protein synthesis by fat body is triggered by ecdysone. Juvenile hormone may still be required, but here it promotes posteclosion development of the fat body to a stage capable of responding to ecdysone. In the yellow fever mosquito, the ovaries themselves were shown to secrete ecdysone in response to EDNH (egg development neurosecretory hormone), a brain hormone released from the corpora cardiaca following a blood meal.

See Also the Following Articles

Eggs • Egg Coverings • Reproduction, Female: Hormonal Control of

Further Reading

Hagedorn, H., Maddison, D., and Tu, Z. (1998). The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. *Adv. Insect Physiol.* **27,** 335–384.

Raikhel, A., and Dhadialla, T. (1992). Accumulation of yolk proteins in insect oocytes. *Annu. Rev. Entomol.* **37,** 217–251.

Sappington, T., and Raikhel, A. (1998). Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem. Mol. Biol.* **28,** 277–300.

Telfer, W., Huebner, E., and Smith, D. (1982). The cell biology of vitellogenic follicles in *Hyalophora* and *Rhodnius. In* "Insect Ultrastructure" (R. King and H. Akai, eds.), pp. 118–149. Plenum Press, New York.

Telfer, W (2002). Insect yolk proteins: a progress report. *In* "Reproductive Biology of Invertebrates," Vol. 12 of "Recent Progess in Vitellogenesis" (A. Raikhel and T. Sappington, eds), pp. 33–71. Wiley, Chichester, U.K.

Wyatt, G. (1991) Gene regulation in insect reproduction. *Invert. Reprod. Dev.* **20,** 1–35.

Wyatt, G., and Davey, K. (1996) Cellular and molecular actions of juvenile hormone. II: Roles of juvenile hormones in adult insects. *Adv. Insect Physiol.* **26,** 1–155.

W

Walking and Jumping

Roy E. Ritzmann
Case Western Reserve University

Sasha N. Zill
Marshall University

The agility of insects certainly contributes to their reign as the most successful creatures in the animal kingdom. Few if any forms of terrain present an insurmountable barrier to all insects. By evolving variations on a basic body plan, they have achieved remarkable dexterity in a wide range of environmental niches. We find insects that walk slowly over floors, scurry under rocks, climb up walls and over ceilings, or jump over barriers that if scaled to human dimensions would represent achievements unattainable by the most accomplished athletes. These abilities have attracted the attention of engineers who study insect locomotion as inspiration for legged robotic devices.

Although these creatures are often described as "simple systems," a close examination of their abilities reveals mechanisms that are elegant and not really simplistic. Their capabilities represent remarkable combinations of mechanical principles, neural control, and sensory input leading to efficient movement of leg joints. Although many aspects of these systems are economical in design and have been studied for many decades, they are only now beginning to be understood.

The problems inherent in insect walking and jumping encompass issues ranging from biomechanics to both central and peripheral neurobiological factors, as well as force development in muscle. To illustrate these points, we will examine the leg structures and neural circuits that produce walking and running in the cockroach. However, a variety of other insects have also been studied extensively by neurobiologists and engineers, including stick insects, grasshoppers, and crickets.

To understand how an insect walks, we must subdivide the process. First, we must understand the movements of the legs and their constituent joints, then we can begin to look at how muscles generate these movements and the circuits within the central nervous system that control motor activity. Over most walking speeds, sense organs of the limbs contribute to motor control by providing detailed information about the position, velocity, and the forces occurring in each leg. Therefore, we must also investigate the role of sensory input in the control of movement. Even with all this information, we will understand only how an animal can walk on a horizontal surface. But the aspects of walking in insects that are truly remarkable, and attract them to robotics engineers, are the abilities to run over and around obstacles, climb up walls, and jump over barriers. Often these adaptations occur through subtle changes in the basic pattern of locomotory behavior.

LEG MOVEMENTS

The walking movements of insects were accurately described in 1887 by Morgan who, in an era in which galloping of horses held the public attention, wrote a letter to *Nature* entitled "The Beetle in Motion" that recounted the coordination seen in a tripod gait. In more detailed accounts in the mid-20th century, Hughes and Wilson noted that insects walk by moving their six legs in reproducible patterns. Each leg alternates between a stance phase, when the tarsus (foot) is on the ground and the animal is pushed forward, and a swing phase, when the tarsus is moved forward through the air. At slow speeds, the legs follow a metachronal pattern, moving from the hind legs to middle and then to front legs on either side. However, to move more rapidly, some legs must be moved at the same time and therefore the insect shifts into a modification of the metachronal pattern called the tripod gait. Here the front and rear legs on one side of the animal move as a unit with the middle leg on the opposite side (Fig. 1A, B). This tripod alternates between swing and stance with the tripod made up of

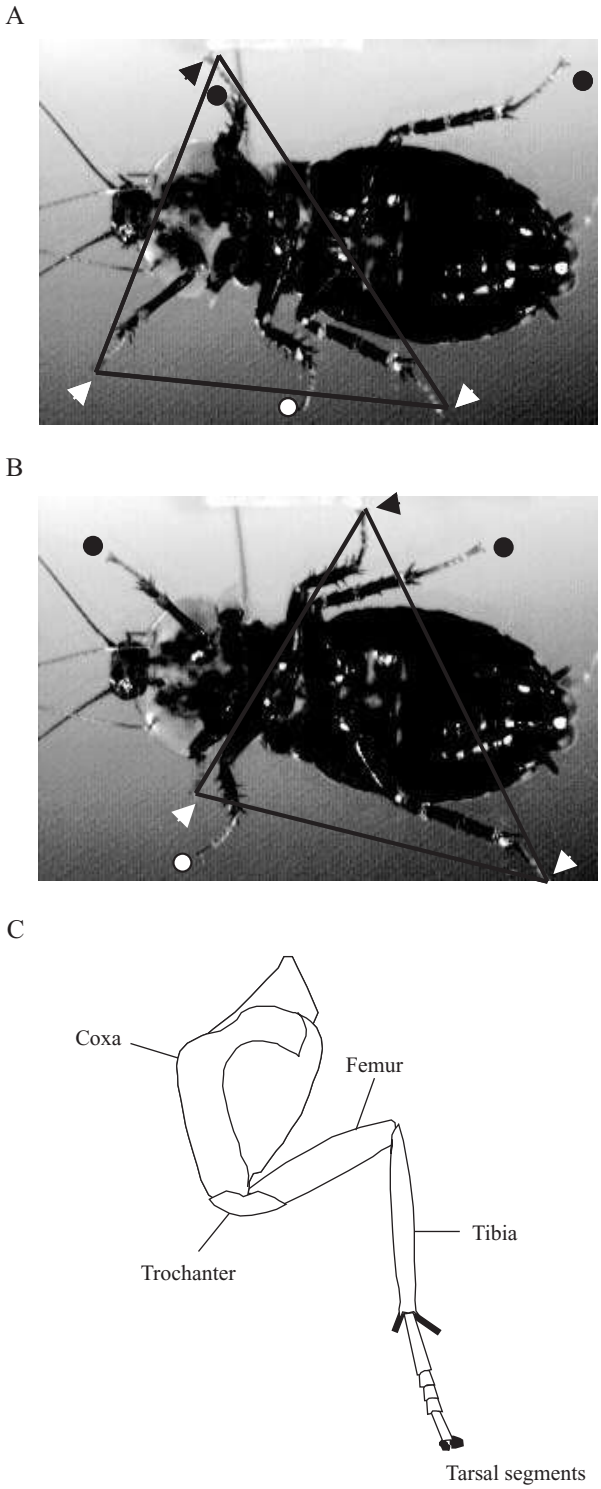

FIGURE 1 Description of leg segments and movements in tripod gait. (A and B) Pictures from a high-speed video record of a cockroach walking on a lightly oiled plate. These images are taken at the beginning and end of one leg cycle. The legs forming one tripod (animal's right front and rear legs and left middle leg) are indicated with triangles; the legs forming the other tripod are designated with circles. Lines connect the triangles to clearly indicate the tripod. Note that in (A) the left front and middle tarsi are very close together, almost overlapping. (C) Diagram showing the segments that make up a typical cockroach leg.

the remaining legs. The tripod gait is very stable, because at most speeds the animal's center of mass remains within the base of support. However, at very high speeds, many insects make dynamic postural adjustments to stay upright. Remarkably, at exceptionally high speeds the American cockroach *Periplaneta americana* has been seen to rise up on its hind legs and run with a bipedal gait. This gait is not statically stable (the cockroach would fall if it stopped), but represents a balance of forces that produces dynamic stability.

Each leg is made up of segments that are similar from leg to leg but differ in dimensions. From the most proximal to distal location, the leg segments are the coxa, trochanter, femur, and tibia, and a series of tarsal segments ending in a retractable claw (Fig. 1C). In the cockroach, the most important joints for walking are the coxa–trochanter (CTr) joint and the femur–tibia (FTi) joint. The CTr joint actually moves the femur relative to the coxa because the trochanter–femur (TrF) joint makes only small movements. Although flexion of the TrF joint can effectively rotate the tarsus, during many movements it acts mechanically as a fused joint. In other insects, relative proportions of these leg segments are changed to match the needs of specialized forms of locomotion. Thus, the hind leg of a locust has a very short coxa and a long muscular femur, making a powerful jumping leg.

Although the legs within a tripod move their feet as a unit, the joint movements and resulting forces are unique for each pair of legs. In cockroaches, the hind legs are specialized to propel the animal forward in walking. To accomplish this the CTr joint and the FT joint move in near synchrony (Fig. 1A). This action allows these rotary joints to direct the movements of the tarsi (feet) in a line nearly parallel to the long axis of the animal's body. The middle legs make similar movements, but with smaller joint [CTr] excursions. The orientation of this leg causes it to first brake and then accelerate forward movements of the animal.

The front legs make very different movements. Unlike the other two pairs of legs, the front legs make much greater use of the body–coxa (BC) joint, which attaches the leg to the thorax. This joint has three degrees of freedom, similar to the ball-and-socket joint in a human shoulder. Movement of the BC joint swings the front legs far forward much like an extending human arm. The CTr and FT joints then move out of phase with each other as the foot is drawn back toward the body and then is pushed backward. The resulting ground reaction forces slow the forward movement of the body and keep the animal from losing control. Clearly, the neural control of this leg is much different from control of the other two pairs of legs. Nevertheless, the sum of the ground reaction forces of all three pairs of legs is similar to that seen in the bipedal leg movements of a human.

Cockroaches can run extremely rapidly (up to 25 steps per second), ranking them as the fastest terrestrial animal (in steps/per second). Running, which follows wind stimulation of the cerci (abdominal appendages) or tactile stimulation of antennae or body cuticle, is used to escape from predators. Although

fast running shows a number of similarities to walking, it occurs so rapidly that it may actually be a separate behavior. Passive elastic structures in the legs can play an important role in generating such rapid locomotion, prompting cockroach running to be modeled as a mass on a spring.

MOTOR CONTROL OF LEG MUSCLE

The muscular anatomy of insect legs follows a proximal-to-distal arrangement that makes very good biomechanical sense. The largest and most powerful muscles are proximal or closer to the body. The muscles are smaller in more distal segments of the leg. With the most powerful muscles placed near the body, inertial effects in the distal part of the limb are reduced, allowing for more controlled movement. Also, the movements of the proximal joints act on the tarsus or foot through the relatively large lever arm of the intervening leg segments to the tarsus or foot. Thus, for example, relatively small movements at the BC joint can greatly alter the orientation of the tarsal end point.

The most distal segment, the tarsus, is actually made up of a series of segments ending in a retractable claw that have again remarkable mobility. However, as with human fingers, the muscle for these segments is found in more proximal segments and imparts its movements via a long apodeme that serves the role of a tendon. Indeed, although the claw can be engaged by this muscle, there is no antagonistic muscle. Rather the claw is disengaged and lifted from the substrate by a remarkably efficient elastic protein (resilin) found in the joints of the tarsal segments. The resilin acts like a spring on a screen door; is stretched when the claw is engaged and causes the tarsus to be lifted automatically when the muscle is relaxed.

NEUROMUSCULAR SYSTEM

The neuromuscular arrangement of insects provides distinct advantages for motor analysis. Unlike vertebrate systems in which muscle cells fire action potentials, in insects most muscle fibers produce graded potentials when motorneurons fire action potentials. Variation in tension in a vertebrate muscle requires recruitment of more or fewer motor neurons. However, in arthropods, simply altering the frequency of action potentials in a single motor neuron can control tension. Furthermore, insect muscles are innervated by very few motor neurons. Often, especially in stance phase muscles (the muscles that extend the leg while the tarsus contacts the ground), there are only two motor neurons serving a range of muscles. These can be readily distinguished as either fast or slow motor neurons depending on the types of muscle contraction they produce. Slow motor neurons need a series of action potentials to generate significant movement, whereas fast muscle generates a reasonable twitch with a single muscle potential. Typically, extracellular recordings indicate that fast motor neurons generate larger action potentials. This neuromuscular arrangement allows one to use electromyogram (EMG) electrodes to record muscle activity extracellularly and often to know exactly which motor neuron is observed.

The leg movements that occur during walking on a horizontal surface are associated with typical patterns of motor activity. In both the middle and hind legs of cockroaches, CTr extension is generated by a burst of activity in the slow depressor motor neuron (Ds). At faster speeds, one or more muscle potentials from the fast depressor motor neuron (Df) occurs at specific times in the leg cycle. These actions generate the stance phase of the leg cycle and alternate with activity from several flexor motor neurons that produce the swing phase. The simultaneous extension at the FT joint is generated by the slow extensor of the tibia (SETi) motor neuron and the fast extensor of the tibia (FETi), and is again opposed by activity in several flexor motor neurons that generate the swing phase of that joint. As expected, the motor patterns of the front legs are more complicated, matching the joint movements described earlier.

The cockroach controls speed of walking by altering the frequency of motor action potentials. Motor frequency is positively correlated with stepping frequency and joint velocity. Thus, the animal can move faster by increasing the frequency of Ds and SETi in each leg. It can turn by increasing frequency in one or more legs of the tripod while decreasing the frequency in the leg or legs located on the opposite side. This change creates stronger forces on the outer leg of the turn and weaker lateral forces on the inner leg, thereby turning the animal.

PATTERN GENERATION

Although both stance and swing phase motor neurons increase burst duration with increasing walking speed, the change in stance phase motor activity is much greater. This observation led to a model for insect walking referred to as the flexor burst generator. In this model, a set of interneurons within the central nervous system called a central pattern generator controls the flexor motor neurons directly and the stance phase motor neurons indirectly through inhibition from swing phase interneurons.

Investigators have recorded traces from oscillatory neurons in cockroaches and stick insects that could be part of such a central pattern generation circuit. For example, in cockroach, neurons have been found that undergo membrane potential oscillations in time with levator motor activity. Moreover, stimulation of these cells activates the flexor motor neurons. Thus, these cells could be part of the flexor burst pattern generator. However, they by no means represent the entire circuit. Other work has strongly suggested that there are separate burst generators that activate motorneurons for each joint in the leg. This would provide for considerable flexibility. The circuits that produce walking could be used in other rhythmic behaviors, such as righting (when an animal falls over and uses its legs to regain an upright posture) or grooming, by changing the coupling between the oscillators.

LEG SENSORS

Even where central pattern generation circuits contribute to control of locomotion, sensory systems still play important roles in generating behaviorally meaningful motor activity. This concept was particularly well demonstrated in the locust flight system. Locust preparations that have been isolated from sensory structures in the periphery (deafferented) are capable of generating bursts of activity that could move the wings in an appropriate manner. However, the frequency of the "wingbeat" cycle is much lower than normal and the pattern quickly dies out, whereas in the intact animal, flight can go on for very long periods of time. If appropriately timed activity from wing stretch receptors is added in, a more normal flight pattern occurs that is maintained for long periods of time. Thus, the central circuits are capable of generating part of the cyclical pattern but not a complete, behaviorally relevant one.

Similarly, normal walking behaviors in all insects that have been studied require input from sensory structures on the legs. However, in rapid running movements associated with escape, feedback from sensory inputs probably is too slow to influence walking within a single step, and the animal may be running "on autopilot."

The sensors that play an important role in walking provide precise bits of feedback information to the walking system through structures that are most often found in the same locations in all three pairs of legs (Fig. 2). Internal sensors such as the chordotonal organs span the joints and monitor joint angle. Hair plates (HP) are positioned near joints where adjacent segments will touch during maximal flexion. Campaniform sensilla (CS) are located in strategic locations near joints and muscle insertions and detect strain in the cuticle. These structures are made up of sensory nerve endings positioned in sockets with a flexible dome. As strain increases in the surrounding cuticle, the dome is deformed and activates the sensory neuron.

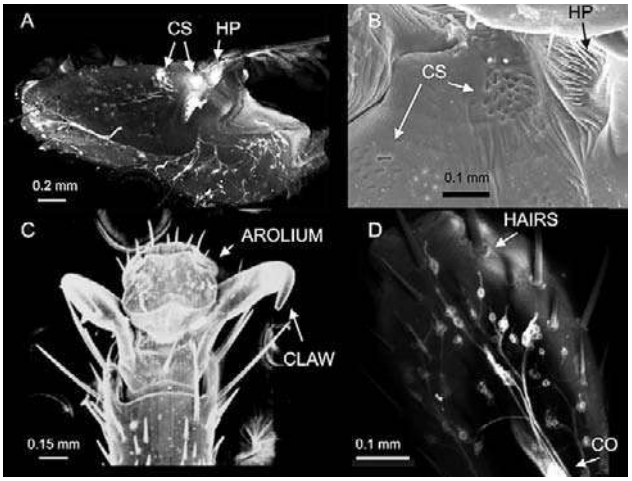

FIGURE 2 Sensory and cuticular structures of cockroach leg (A) View of the trochanter in a confocal microscope in which the main leg nerve was infiltrated with fluorescent dye (diI) that diffuses in membranes and stains sensory neurons when it is applied to peripheral nerves. The sensory neurons of campaniform sensilla (CS) and the hair plate (HP) are filled with the dye and fluoresce brightly. (B) Scanning electron micrograph of the region containing the sense organs showing cuticular caps of campaniform sensilla and long hairs of hair plate. (C) Scanning electron micrograph of the distal end of the tarsus (foot), showing the hooklike claws and the arolium (an adhesive pad between the claws) (D) View of the dorsal side of the claw shows numerous sensory neurons stained with diI that innervate hairs as well as a portion of the chordotonal organ (CO). [Photomicrographs (A) and (D) by Faith Frazier, (B) by David Neff, and (C) by Laura Quimby, Marshall University School of Medicine.]

Each peripheral sensor provides information that is used at discrete phases of the step cycle as the insect walks. Although joint angle detectors such as the femoral chordotonal organ are active throughout the ranges of joint movement, they often reach their highest levels of firing in positions of maximum joint flexion or extension. In walking of stick insects, these receptors trigger a change from stance to swing phase when joint angles reach the extreme joint position. Hair plates monitor the resulting joint movements in swing and can aid in the initiation of stance. Campaniform sensilla are activated during stance as forces are exerted upon the leg and the load of the insect's body is supported. They then act to increase the activity in the extensor motor neurons to rapidly generate forces in the leg. The greater extensor activation in each stance leg propels the animal forward more efficiently and also allows other legs to be lifted in swing. The elevated stance activity continues until the chordotonal organs trigger the switch to swing phase.

Beyond helping to set the normal locomotor pattern, these sensors are, of course, readily available to make rapid adjustments to the motor pattern as the animal moves away from horizontal surfaces and climbs inclines or vertical surfaces, or even walks on ceilings and negotiates obstacles in its path. The increases in load experienced on each leg as the animal climbs a surface ultimately translates into an altered gait. For example, as a cockroach walks on a ceiling, legs hold on to the surface for greater periods of time. Metachronal gaits are used when the animal is inverted, and individual legs make very rapid swing movements before reattaching to the substrate.

Arrays of neurons with diverse properties receive sensory inputs and could mediate these effects. Within each thoracic ganglion, the sensory neurons from these leg sensors project to populations of spiking local interneurons (interneurons that are isolated to a single ganglion). The projections of many receptors follow a topographic pattern that reflects the location of the sensors in the leg. The spiking interneurons project both to motor neurons and to another set of local interneurons called nonspiking interneurons. These interneurons never generate action potentials. Rather, they act through graded potentials to control and coordinate motor activity to various leg muscles within a thoracic segment. Coordination among joints of each leg (intraleg coordination) occurs through interactions in these local circuits.

Finally, spiking interganglionic interneurons project between thoracic ganglia, influencing local circuits for adjacent legs. Through these circuits the sensory activity directs appropriate

adjustments in individual muscles, coordinates segments within a single leg, and influences the coordination between legs.

INTERLEG COORDINATION

To generate an effective tripod gait, insects must coordinate movements of joints not only within a single leg but also between legs. This interleg coordination on the surface appears to be a daunting task. The tarsi of the three legs that make up an effective tripod must move in synchrony. However, as described earlier, the legs of each segment make different movements. The patterns of the motor neurons that control those movements also differ in each leg. Thus, the animal must coordinate the tarsi through joint movements and neural activities that are specific to each pair of legs.

Observations on stick insects suggest that these potentially complex connections can be functionally formulated as a set of fairly straightforward rules indicating that events in one leg that are readily detected by leg sensors influence the actions of other legs in discrete ways. For example, one rule is that when one leg is in swing, the adjacent anterior leg cannot enter into swing phase. Alternatively, the start of the stance phase promotes the start of swing in adjacent anterior legs and in the contralateral leg of the same segment. More subtle influences also exist to account for transient changes. For example, an increased load on one leg prolongs the power stroke on the contralateral leg of the same segment. This would result in coactivation of legs when the animal is encountering increased resistance, such as during climbing. The specific neural connections that underlie these rules are an exciting new area of research.

NEGOTIATING OBSTACLES

The remarkable agility that attracts interest from roboticists is seen when insects are faced with an obstacle. Most insects effortlessly negotiate barriers that would pose considerable control problems in robots. This problem is only now beginning to be investigated, since climbing over obstacles is a more transient event than horizontal walking and there is considerable variability in the strategies used. However, some common themes are evident. In climbing over a block, for example, insects must move their center of mass (CoM) upward to surmount the obstacle. For cockroaches, the strategies used to accomplish this task vary with the size of the obstacle and the speed that the insect is moving at when the block is encountered. The front legs are normally lifted up fairly high during normal walking. If the block that is encountered is lower than the normal front leg trajectory, the insect hardly needs to change its movement at all. The front leg will be placed on top of the obstacle and pushed down to move the animal's CoM upward as a natural consequence of the encounter with the block. Larger blocks can be negotiated in similar ways when the animal is moving at faster speeds by taking advantage of the inertia developed by body movements.

At slower speeds, encounters with larger blocks require the animal to first rear upward to get its front legs on top of the block and initiate the climb. Locusts negotiate such blocks by using an elevator reflex, in which the front leg bumps into the front surface of the block repeatedly while moving upward until the top of the block is reached. However, in cockroaches, the front leg typically reaches the top of the block in a single movement, without even touching the front surface.

For very large obstacles, the insect must actually climb up the front face of the object. Here the insect must attach itself as if the face of the block were a wall. This problem is solved by a combination of claws and pads that adhere to various substrates. Cockroach claws, which are remarkably similar to the claws of a cat, are normally held up during horizontal walking. However, when a cockroach is walking up a vertical surface, the claw is pulled down by a muscle in the tibia, which attaches to the claw through a long apodeme. There is no return muscle. Rather, as we have described, the claw pops back up passively with the aid of strategically placed resilin ligaments. For smoother surfaces, including those as polished as glass, some cockroaches have cuticular pads between their claws that can adhere to the substrate for stability but be readily lifted when needed in walking.

JUMPING

Many insects have developed efficient jumping behaviors to move long distances with a single jump. The jump requires a powerful and rapidly accelerating movement of the jumping leg. Orthopterans such as grasshoppers and locusts have very large hind legs that can generate powerful jumping movements. However, even the large femur extensor muscles of these jumping legs cannot generate the quick extension necessary for an efficient jump without some mechanical modification within the leg structure for storing energy. In locusts, the tendon of the tibial flexor muscle moves over a stop that allows the extensor to contract without moving the leg when the muscles are coactivated. As a result, a considerable amount of energy is stored in the mechanical distortion of the femur, tibia, and extensor tendon (like the bow of an archer). Inhibition of the tibial flexor muscle releases locking mechanisms and produces a very rapid and powerful extension that propels the animal upward. The timing of these events is critical, and neural circuits have been identified in association with kicking movements that provide the appropriate control.

Fleas generate remarkable jumps relative to their tiny size. Again, a proportionally very large hind leg is used to generate the jumping movement, with a hook on the hind leg preventing movement until a large isometric force has been achieved. The isometric force is stored in strategically located resilin ligaments and released during the jump.

One of the most remarkable jumping strategies in insects does not even involve legs. Click beetles can jump from a standstill to four times their body length by rapidly accelerating the joint between the prothoracic and mesothoracic segments

of the body. Here again a mechanical stop prevents movement until large isometric force has been achieved, this time in thoracic muscles, and then is released suddenly to shoot the animal upward.

ROBOTIC DEVICES INSPIRED BY INSECT WALKING

The efficiency and agility of walking insects as well as specialized strategies such as jumping have not been lost on robotics engineers. Insects have provided popular models for legged robots because the hexapod gait is statically stable. That is, the CoM remains within the base of support of at least one tripod at all times. Thus, the control for a robot need not actively maintain balance during horizontal walking. However, to capture the agility of the insects, postural adjustments, reflexes, efficient leg design, and sensors must be incorporated into the robotic designs. This realization has generated collaborations between engineers and biologists to study insects and then incorporate newly discovered control and mechanical properties into more efficient robots. The efforts to develop and control these new robots in turn provide new insights into how the insect controls its own movements.

See Also the Following Articles

Anatomy • Flight • Legs • Wings

Further Reading

Bässler, U., and Büschges, A. (1998). Pattern generation for stick insect walking movements—Multisensory control of a locomotor program. *Brain Res.* **27,** 65–88.

Burrows, M. (1996). "The Neurobiology of an Insect Brain." Oxford University Press, Oxford, U.K.

Camhi, J. M. (1980). The escape system of the cockroach. *Sci. Am.* **243(6),** 158–172.

Cruse, H. (1990). What mechanisms coordinate leg movement in walking arthropods? *Trends Neurosci.* **13,** 15–21.

Delcomyn, F. (1985). Walking and running. *In* "Comprehensive Insect Physiology, Biochemistry and Pharmacology: Nervous System: Structure and Motor Function." (G. A. Kerkut and L. I. Gilbert, eds.), pp. 439–466. Pergamon Press, Oxford, U.K.

Heitler, W. J., and Burrows, M. (1977). The locust jump. I: The motor programme. *J. Exp. Biol.* **66,** 203–219.

Hughes, G. M. (1952). The co-ordination of insect movements. I: The walking movements of insects. *J. Exp. Biol.* **29,** 267–284.

Morgan, C. L. (1887). The beetle in motion. *Nature* **35,** 7.

Pearson, K. (1976). The control of walking. *Sci. Am.* **235,** 72–86.

Ritzmann, R. E., Quinn, R. D., Watson, J. T. and Zill, S. N. (2000). Insect walking and biorobotics: A relationship with mutual benefits. *BioScience* **50(1),** 23–33.

Watson, J. T., and Ritzmann, R. E. (1998). Leg kinematics and muscle activity during treadmill running in the cockroach, *Blaberus discoidalis.* I: Slow running. *J. Comp. Physiol. A* **182,** 11–22.

Wilson, D. M. (1966). Insect walking. *Annu. Rev. Entomol.* **11,** 103–123.

Zill, S. N. (1990). Mechanoreceptors: Exteroceptors and proprioceptors. *In* "Cockroaches as Models for Neurobiology: Applications in Biomedical Research" (I. Huber, E. P. Masler, and B. R. Rao, eds.), pp. 247–267. CRC Press, Boca Raton, FL.

Zill, S. N., and Seyfarth, E.-A. (1996). Exoskeletal sensors for walking. *Sci. Am.* **275(1),** 86–90.

Walking Sticks

see *Phasmida*

Wasps

Justin O. Schmidt
Southwestern Biological Institute, Tucson

Few entomological words evoke a more vivid image and response than the word "wasp." Wasps live on all continents inhabited by people, and nearly everyone is familiar with their colorful presence and habits. The root of their name is the word "webh" (Low German)—to weave, a reference to the nests constructed by these insects. In entomological literature, "wasp" is often applied to all members of the Apocrita suborder of Hymenoptera except ants and bees. However, in popular usage a wasp is any social species in the family Vespidae, especially in the genera *Vespula, Dolichovespula,* and *Vespa,* which are viewed as irritating, plus a few conspicuous solitary hunting wasps. In this article, emphasis is according to popular usage.

TAXONOMY

In the broad sense, wasps include an enormous number of parasitoid species in the families Ichneumonidae, Braconidae, Chalcididae, and many other insects referred to as "parasitic wasps," plus solitary or hunting wasps, including sphecid wasps (Sphecidae), spider wasps (Pompilidae), potter wasps (Vespidae: Eumeninae and others), scoliid wasps (Scoliidae), velvet ants (Mutillidae), and others. In terms of sheer numbers of species, these groups far outnumber those of the social wasps. In terms of success, however, social wasps have earned due respect as major influences and elements of most ecosystems.

The total number of social wasps in the world is slightly over 900 species divided into three subfamilies: the Stenogastrinae, or hover wasps; the Polistinae, or paper wasps; and the Vespinae, or wasps, hornets, and yellowjackets (Fig. 1). Stenogastrines live in the South Indo-Pacific region, and comprise just over 60 species of long slender adults in six genera; their small, inconspicuous nests often are attached to roots or twigs under overhangs of banks. They provide their larvae with insects, often taken from the webs of spiders, hence, the term "hover wasp." The Polistinae are by far the most speciose of the three groups of social wasps: approximately 800 species are divided into 27 genera, with a cosmopolitan distribution except for the colder regions of Eurasia and North and South America. The Vespinae contains just over 60 species broken into four genera. *Vespa,* the true hornets, inhabit temperate and tropical Eurasia

FIGURE 1 Wasps. Clockwise from upper left: *Polistes exclamans,* larvae and capped pupae of *Vespa mandarinia, Vespula maculifrons* workers stinging a leather target—note the loss of the sting apparatus (sting autotomy) in the leather, *Agelaia myrmecophila* carving "meatballs" from a dead mammal. (Photographs by J. O. Schmidt.)

and northern Africa, number about 23 species and are the largest of the wasps. *Vespula,* called yellowjackets or the true wasps, are a Northern Hemisphere genus of about 22 species that are present from the Arctic to the northwestern fringe of Africa, to India, and into Central America. The 15 species of *Dolichovespula,* sometimes called aerial yellowjackets because many species construct round paper nests above ground, occupy the same range as *Vespula* but are absent in extreme Southeast Asia and Central America plus most of Mexico. The smallest genus of vespines, *Provespa,* consists of three pale nocturnal species limited to forested areas in tropical Southeast Asia.

BIOLOGY AND LIFE HISTORY STRATEGIES

Social wasps are hunters and scavengers that prey on a variety of arthropod and animal protein sources. They are well equipped for this role, possessing large, powerful cutting mandibles for capturing, subduing, and processing prey, large eyes for detecting potential prey, and the ability to fly and hover in pursuit of food. Although well known for their stinging ability, wasps do not capture or subdue prey with the sting. Their venom is used entirely for defense, primarily against vertebrate potential predators. Indeed, all investigated wasp venoms are toxic, painful, and effective against vertebrates, but rather inactive and slow in affecting insects unless the sting is delivered near a ganglionic center of the nervous system. Moreover, wasps have little need to sting prey because they are equipped with powerful mandibles that can be used to chew wood fibers from dead trees or cut through tough insect nets in a matter of minutes.

Prey of wasps is varied. Paper wasps *(Polistes)* are specialists on caterpillars, which they locate by visually and olfactorally searching vegetation likely to harbor them. Once located, the prey is quickly subdued, cut into manageable pieces that are chewed into a "meatball," and carried back to the nest to be fed to the larvae. Yellowjackets and hornets tend to have a broader diet than paper wasps and will capture a variety of arthropod prey including flies, spiders, caterpillars, and an assortment of other groups. House flies *(Musca domestica)* and other flies comprise major prey items of some species. Some species also scavenge for prey, removing insects captured in spiderwebs, carving flesh from dead animals such as rodents, and even removing insects freshly smashed on radiators and grilles of cars. The scavenging habits of some species have earned them the distinction of being considered to be pests at picnics, outdoor events, and around garbage cans.

Prey of the swarm-founding epiponine wasps in the subfamily Polistinae is less well characterized than prey of other wasps. Like other wasps, the epiponines often remove wings, legs, and the head of prey and chew it into a pulp before returning to the nest. Epiponines take a variety of prey, including flies, caterpillars, leafhoppers, and other insects, and many species are likely to specialize on certain taxa, or to prefer particular parts of the habitat, such as open areas, forest canopies, or thick vegetation. Some species in the genus *Agelaia* are well known for removing pieces of flesh from large dead animals and have been given the common name of "vulture wasps."

Adult wasps usually obtain energy for flight and general metabolism from sugar sources. Nectar from flowers, honeydew from aphids and other homopterous insects, and sweet fluids from fruit all can be food sources. In most species, these sugars are supplemented by the sugary trophallactic secretions produced by larvae in response to solicitation by adults.

All wasps construct multicell nests from plant fibers (Fig. 2). The most common materials for making nests are wood fibers scraped from weathered dead wood from trees or twigs, but other materials including rotten wood, fibers from living plant leaves, and stems are sometimes used. In most cases, these fibers are strengthened with salivary secretions during preparation and application to the nest. After application of the fibers, abdominal secretions may be added to further strengthen the material and help repel ants.

Wasp life cycles consist of colony initiation, growth and expansion, production of reproductives, reproductive dispersal, and colony decline. In most taxa the cycle is determinate and lasts less than a year, with complete dissolution of the society at the end. Some taxa have an indeterminate cycle in which the colony population simply decreases after a major reproductive event before again beginning the growth and expansion cycle in the same nest, often with new queens. Some indeterminate species can remain in a nest for years, with a record of 30 years.

Colony founding can be independent, or by swarm founding. In independent founding, the colony is initiated by one or more queens without the aid of workers; in swarm founding, colony initiation is accomplished by a swarm of many workers plus reproductive queens. Independent founding can be by a single queen (haplometrosis) or by several queens joining their efforts to initiate the colony (pleometrosis). Most yellowjackets and hornets are haplometrotic, although *Vespa affinis* sometimes is pleometrotic with multiple foundresses in a colony, and *Provespa,* a greater departure from the rule, repro-

FIGURE 2 Wasp nests. Clockwise from upper left: *Polistes snelleni, Vespa simillima, Apoica pallens*—note aposematic color and alignment of abdomens of adults on nest, *Parachartergus fraternus,* cryptic nest of *Metapolybia aztecoides* attached to a *Bursera simaruba* tree trunk, *Polybia simillima*—workers are in the defensive attack position on nest. (Photographs by J. O. Schmidt.)

duces by swarming. *Polistes* and *Mischocyttarus* independently found colonies with a mix of both pleometrotic and haplometrotic colony founding. Single founding queens have the advantages of absence of competition for egg laying from other queens and a potentially greater number of reproductive offspring per foundress. Disadvantages include greater risks from ants and other predators or parasites attacking the unattended nest when the foundress is foraging, and often a much lower success rate in establishing a nest. In addition to better protection of the nest, advantages for multiple foundresses include faster construction of the nest, more reliable prey capture for the larvae, and increased colony survival in the event of the loss of a foundress. These trade-offs of better survival and growth often prove advantageous for foundresses to join others. Both within individual species and across the genus as a whole, pleometrotic founding tends to be more prevalent at lower latitudes than haplometrotic founding, perhaps in part because of the more intense pressure from predators and parasites in warmer environments.

Reproduction and colony founding by swarms occurs in all epiponine wasps plus *Provespa* and some *Ropalidia.* During swarm founding, hundreds (or even thousands) of workers with young queens leave the parent nest and move to a new location to construct a new nest and reproductive unit. This form of reproduction, though similar in many ways to that of honey bees, differs in several details including the general presence of many more queens in wasp swarms. Clear advantages of swarm founding over independent founding are the availability of a large worker force to construct the nest quickly and the opportunity for task specialization by individuals within the swarm. Reproductive individuals in the swarm need not forage as must their independent founding counterparts, and individual workers can specialize in different tasks such as collecting nectar, prey, fiber for nest construction, or water for cooling and mixing with fiber for nest construction. Multiple workers also ensure that a large defending force is always available, should intruders appear.

WASPS AND PEOPLE

Nuisance

The general public is not aware of the enormous beneficial role wasps play in controlling the populations of flies, caterpillars, and other often unwanted insects. It could be argued that even wasp scavenging on dead animals might provide competition for flies. These beneficial aspects of wasps are overshadowed by the view that they are nuisances, pests, and threats. Only a very few species of *Vespula,* particularly *V. germanica, V. vulgaris, V. maculifrons, V. squamosa,* and a few others, are attracted to human food and activities. Particularly in the fall, these species will readily forage around outdoor dining areas, areas where food is processed or openly available, or where garbage is present. At these locations, their bright yellow- or white-on-black color patterns make them conspicuous, as does their buzzing and foraging action. At least in many western

cultures, an almost innate swatting or flapping reflex is exhibited by people when they detect a buzzing insect, irrespective of whether the insect is a fly or a wasp. The origin of this reaction is uncertain: Is it fear of wasp stings, dislike of flies, or a general dislike of any insect that approaches one's body? In any event, wasps, by their very nature, are perhaps the insect world's masters at stimulating human aversive swatting, in the process gaining a solid reputation for being an irritating nuisance.

Wasp nests themselves are often viewed as an encroachment on human space and aesthetics. Were it not for the habit of many species of *Polistes* to build their exposed comb nests under eaves or inside man-made structures, their presence likely would never be noticed. These nests with their exposed adult wasps, especially when near entrances to buildings, tend to generate fear and dislike on the part of the humans passing through the area. This often results in an unfortunate consequence for the wasps—the destruction of their nests and sometimes the adults. Aerial yellowjackets, especially *Dolichovespula arenaria* and *D. maculata* in North America, build large conspicuous gray, nearly spherical nests attached under eaves or to the sides of buildings. Their large size and busy activities tend to elicit an even less favorable reaction by people for these wasps than for *Polistes.*

The mere existence of wasps is seen as a nuisance by much of the human populace. As long as neither the wasps nor their nests are apparent, they do not come into conflict with people. However, should a hidden *Polistes* nest in vegetation, or a *Vespula, Dolichovespula,* or *Vespa* nest in the ground, a tree, or within a building be discovered, the nest is either avoided or destroyed. Such discoveries are usually beneficial neither to the wasps nor to the people involved.

Tramp and Pest Status

Tramp species are those that tend to be able to disperse via the intentional or inadvertent help of humans to areas where they are not native, and to establish large successful populations in the new areas. Numerous examples can be found among the cockroaches, flies, rodents, domestic animals, and others. Among the wasps, only a relatively few species are successful tramps: for example, the German yellowjacket *(Vespula germanica),* the common wasp *(Vespula vulgaris),* the western yellowjacket *(Vespula pennsylvanica),* and the European paper wasp *Polistes dominulus. Vespa crabro,* the European hornet, has also been introduced into eastern North America, but it seems to be neither particularly abundant nor a pest in its new location. The others, once released from many of the controlling constraints placed on their populations by predators, parasites, competitors, and possibly pathogens and diseases, tended to expand their populations to enormous numbers and become pests.

The German yellowjacket is classic in this regard: it is successfully established in North America, New Zealand, Australia, South Africa, and several other locations. In New Zealand, where no other social wasps previously existed, it expanded its niche and population to such densities that it constitutes a major pest and threat to much of the native fauna of the country. In addition to its dense populations, *V. germanica* evolved secondary polygyny in which some colonies do not disintegrate at the end of the season, but instead many mated queens remain in (or return to) the nest and the brood rearing cycle resumes—although this time the colony can become huge, with many tens of thousands of individuals. In North America, *V. germanica* rapidly expanded its population and range to include most of the northern part of the United States and the populated areas of California. Its success at the apparent expense of some native species of *Vespula* is likely a result of its aggressive foraging behavior and its affinity for nesting in human structures. The walls of buildings provide many benefits for the wasps: warmth during the late fall and early winter, shelter from moisture and many predators, and proximity to garbage and other food sources. This proximity to people and their food allows colonies to remain active longer during the year, to grow larger, and to produce more reproductives. It also brings them into direct conflict with people and makes them a major pest in some areas.

P. dominulus, unlike *V. germanica,* is not particularly defensive to people near its nest and tends to be less threatening in appearance. In recent years, it successfully expanded its range in North America to include much of the Midwest, where it can be present in enormous numbers. Also, its tendency to be less choosy in nest site locations causes annoyance beyond what was traditionally elicited by a *Polistes* species.

Stings, Venom, and Medical Risks

At the sight or the mention of a wasp, simultaneous thoughts enter the minds of many people—dislike and sting. The immediate human association of wasp with "sting" demonstrates the success of the wasp sting as a defense against large predators. Wasp stings hurt; and wasp stings provide reinforcement to the pain in the form of local swelling, redness, itching, and tenderness. Social wasps, like many other social insects, live in immobile colonies with many edible and nutritious young, offering an incentive for large predators to attempt to overpower the wasps and consume the brood. The sting with its associated venom is a nearly perfect defensive system to counter attacks by large predators: the sting can penetrate the tough skin of the predator, and the venom is then delivered into richly innervated living tissue, where it causes immediate pain and tissue damage. The consequence of wasps possessing their venomous sting is that few vertebrate predators can successfully attack wasp nests: the exceptions are a few specialists that have secondarily evolved means to exploit wasps and their nests. The effectiveness of wasp stings is not lost on people and forms much of the basis for human aversion to wasps. Until recently, this worked to the benefit of wasps, but now with modern protective clothing, devices, insecticides, and other means to attack wasps, the defense is less effective against people. Nevertheless, general human fear and dislike of wasps continue undiminished.

Wasps can pose a minor medical risk to people hypersensitive to their stings. Overall estimates of the incidence of allergic reactions to insect stings range from about 0.5 to 2% of the population, with slightly over half the individuals being sensitive to wasp stings (the rest are primarily sensitive to honey bee and fire ant stings). Of this wasp-sensitive population of approximately 3 million in the United States, about 20 actually die each year. The remaining 99.999+% of the hypersensitive population suffer at most generalized reactions involving the skin or the respiratory or cardiovascular systems. For people who suffer severe allergic reactions, particularly those in which breathing becomes difficult, or fainting or other signs of blood pressure drop occur, preventive approaches in the form of personal epinephrine injectors or immunotherapy are available. For the rest of the populace, wasps present little risk beyond the affront they cause with their stinging.

See Also the Following Articles

Caste • Colonies • Defensive Behavior • Hymenoptera • Nest Building • Venom

Further Reading

Itô, Y. (1993). "Behaviour and Social Evolution of Wasps." Oxford University Press, Oxford, UK.

O'Neill, K. M. (2001). "Solitary Wasps: Behavior and Natural History." Cornell University Press, Ithaca, NY.

Ross, K. G. and Matthews, R. W. (eds.). (1991). "The Social Biology of Wasps." Cornell University Press, Ithaca, NY.

Schmidt, J. O. (1990). Hymenoptera venoms: striving toward the ultimate defense against vertebrates. *In* "Insect Defenses: Adaptive Mechanisms and Strategies of Prey and Predators" (D. L. Evans and J. O. Schmidt, eds.), pp. 387–419. State University of New York Press, Albany.

Turillazzi, S., and West-Eberhard, M. J. (1996). "Natural History and Evolution of Paper-Wasps." Oxford University Press, Oxford, U.K.

Water and Ion Balance, Hormonal Control of

Thomas M. Clark
Indiana University, South Bend

Insects are found in habitats ranging from the driest deserts to aquatic habitats of diverse ionic composition. Their survival depends on the ability to keep their tissues and cells moist, and to regulate the composition of their body fluids. Body fluid composition is challenged whenever exchanges of materials with the environment take place, and such exchanges are unavoidable. Routes of gain of materials include ingestion through the mouth, osmotic gain of water, active uptake of ions, and diffusion across the body surface. Materials are lost through the excretory system and in the feces, as well as by osmotic, evaporative, or diffusive losses across the body surface. During feeding or drinking, an insect may rapidly ingest a large volume of material that is significantly different in composition from the body fluids. Insects respond to perturbations in body water composition by using a combination of hormonal and autonomous mechanisms to vary the amount and composition of fluid excreted by the excretory system.

STRUCTURE AND FUNCTION OF THE EXCRETORY SYSTEM

The excretory system consists of the Malpighian tubules and the hindgut (Fig. 1). The Malpighian tubules form the primary urine by active secretion of ions into the lumen. Water follows passively by osmosis, drawing with it small hemolymph solutes. This primary urine is then modified as it passes through more proximal regions of the Malpighian tubules and through the hindgut. The Malpighian tubules and the hindgut each frequently consists of several functionally distinct regions and/or distinct cell types, and each region contributes by specific mechanisms to the overall formation and processing of the urine.

Water balance can be adjusted by altering the rate of formation and/or composition of primary urine, and by changing the activity of mechanisms that modify the primary urine. Fluid secreted by Malpighian tubules can be passed forward into the midgut and reabsorbed, or passed back through the hindgut for modification prior to elimination. Evidence suggests that Malpighian tubule, midgut, and hindgut transport processes are each regulated by hormones.

REGULATION OF THE EXCRETORY SYSTEM

A number of chemical factors have been isolated from neural tissue of many insect species that alter fluid and/or ion transport rates of particular regions of the excretory system. Only rarely have these factors been demonstrated to act as hormones, and

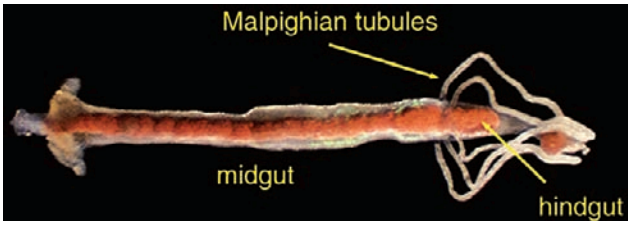

FIGURE 1 The larval mosquito gut and excretory system. The five Malpighian tubules form the primary urine by active secretion of ions, with water following the resulting osmotic gradient. Fluid formed by the Malpighian tubules enters the alimentary canal at the midgut–hindgut junction. This primary urine can be either shunted forward into the midgut and reabsorbed, or passed backward into the hindgut for modification and eventual elimination. Evidence suggests that all three regions, the midgut, Malpighian tubules, and hindgut, contribute to water and ion balance and are under hormonal control. [Modified from Clark, T. M. *et al.* (1999). The anterior and posterior "stomach" regions of the larval *Aedes aegypti* midgut: regional specialization of ion transport and stimulation by 5-hydroxytryptamine. *J. Exp. Biol.* **202,** 247–252. With permission of the Company of Biologists Limited.]

their specific roles in regulation of water and ion balance remain largely speculative. Nevertheless, available evidence suggests that regulation of water and ion balance involves several different hormonal factors, each with a specific mechanism of action and acting on only a portion of the excretory system or even a specific group of cells within a region. At least some regions seem to be regulated by multiple hormones, with combinations of hormones producing responses that are often distinct from the actions of the individual hormones acting alone. The actions of the distinct regions and cell types within the excretory system are apparently coordinated by hormones to produce final excreta of appropriate composition. Evidence to date suggests that hormones may regulate several related parameters, including hemolymph volume, hemolymph ionic composition, and clearance of metabolic wastes and toxins from the hemolymph.

Hormonal Regulation of Malpighian Tubule Transport

Factors hypothesized or demonstrated to play a hormonal role in regulation of Malpighian tubule function include peptides and the biogenic amine serotonin. The best studied of the peptides are the kinins and the diuretic peptides that are related to corticotropin-releasing factor (CRF).

Kinins are small (6- to 15-amino-acid) peptides with a C-terminal amide. Those identified to date have similar C-terminal sequences of Phe-Xxx-Yyy-Trp-Gly-amide, where Xxx can be Asn, Ser, His, Phe, or Tyr and Yyy can be Pro, Ser, or Ala. The kinins have been isolated from several insect species but have not been detected in animals other than insects. A single species may contain several kinins. To date, a hormonal role for a kinin has been demonstrated in only the cricket *Acheta domesticus* (see later).

The CRF-related diuretic peptides range from 30 to 46 amino acids in length and show structural homology to a family of peptides found in vertebrates that includes corticotropin releasing factor (hence the name), sauvagine, and urotensin. Available evidence suggests that a single species produces more than one CRF-related diuretic peptide. To date, a hormonal role for a CRF-related diuretic peptide has been demonstrated conclusively in only one species, *Locusta migratoria* (see later).

Serotonin stimulates Malpighian tubule secretion in a number of insect species and is known to hormonally regulate excretory function in the bloodsucking bug *Rhodnius prolixus* and in larval mosquitoes *(Aedes aegypti).* In *Rhodnius* and most likely in mosquitoes as well, serotonin stimulates both uptake of fluid across the midgut into the hemolymph and removal of that fluid from the hemolymph by the Malpighian tubules. Thus the neurotransmitter acts to coordinate midgut and excretory function to remove excess fluid and ions from the body following ingestion.

Malpighian tubules of at least some insect species appear to be further regulated by unidentified antidiuretic factors. In addition, short-term hormonal adjustments in Malpighian tubule transport rates in response to immediate challenges can be superimposed on longer term increases and decreases in basal and maximal transport capacities. This process may be mediated at least in part by prostaglandins.

Hormonal Regulation of Hindgut Transport

Hormones acting on the hindgut appear to stimulate recovery of fluid and ions from the primary urine, leading to cycling of hemolymph through the excretory system and clearance of wastes from the hemolymph. The majority of our information about regulation of hindgut transport comes from studies on the locust. The locust hindgut consists of two distinct regions, a more anterior ileum and a more posterior rectum. Two peptide factors isolated from the corpus cardiacum are ion transport peptide (ITP), acting on the ileum, and chloride transport stimulating hormone (CTSH), acting on the rectum. The amino acid sequence of these peptides has been determined. Other factors that stimulate water reabsorption by the hindgut include neuroparsins A and B. To date, none of the factors acting on hindgut transport have been demonstrated in hemolymph.

Hormonal Regulation of Midgut Transport

Regulation of midgut transport has received little attention. Hemolymph serotonin stimulates midgut transport mechanisms in *Rhodnius* and in larval mosquitoes. Generation by larval lepidopterans of a highly alkaline midgut lumen, creating conditions favorable for digestion and nutrient absorption, seems to be regulated by an unknown endocrine or paracrine signal. Several peptides inhibit ion transport across the lepidopteran midgut. Evidence for regulation of more than one midgut transport mechanism in diverse insect species suggests that the regulation of midgut transport may be more important in the control of water and ion balance than has been assumed.

BIOLOGICAL ROLES OF HORMONAL REGULATION OF EXCRETORY FUNCTION

Orthoptera: Grasshoppers

When food is available, locusts ingest considerable volumes of fluid with their food and produce copious, moist feces. When deprived of food and water, the feces produced are very dry. During times of limited water availability, Malpighian tubules secrete fluid at a low rate, and much of the secreted water is recovered in the hindgut. This recovery of ions and water seems to be mediated by a hormone or hormones (ITP and CTSH, mentioned earlier). Upon feeding, a CRF-related diuretic hormone appears in the hemolymph, stimulating fluid secretion by the Malpighian tubules. Presumably, antidiuretic factors acting on the hindgut are no longer released under these conditions, and thus elimination of excess water and ions (especially KCl) becomes possible. Hormones acting on the hindgut also seem to be involved in regulation of hemolymph pH.

Orthoptera: Crickets

Crickets possess Malpighian tubules consisting of distinct distal, mid, and proximal regions. The distal and midtubule regions at least may be regulated independently, and apparently secretion rates of each region can be either stimulated or inhibited. Little is known about the biological roles of the majority of the factors that alter tubule function *in vitro.* Achetakinin (a member of the kinin family of peptides) appears in hemolymph at levels sufficient to stimulate Malpighian tubule secretion in response to starvation, at a time when the insect is conserving water, rather than during feeding or water loading. These data suggest that achetakinin may act to stimulate fluid cycling through the excretory system, leading to clearance of materials from the hemolymph, rather than stimulating diuresis.

Heteroptera: Rhodnius

Most of our information concerning regulation of water balance in Heteroptera comes from studies on the bloodsucking bug *Rhodnius.* The physiology of this animal is unusual in that *Rhodnius* seldom feeds, and when it does it may ingest a volume of blood up to 12 times its own body mass. At this time, serotonin released into the hemolymph causes the cuticle to become more plastic (allowing accommodation of the meal), stimulates anterior midgut fluid and ion absorption, and acts synergistically with a diuretic peptide to cause a 1000-fold increase in Malpighian tubule secretion rates. This leads to rapid removal from the midgut lumen and hemolymph of the unusable fluid and ion load ingested with the meal.

Diptera: Mosquitoes

Much of our information about hormonal regulation of water and ion balance among Diptera comes from studies on larval and adult mosquitoes. Larval mosquitoes are aquatic, but the adults are terrestrial. The adult female of most species seeks out and consumes large blood meals to provide proteins for successful egg production. Larvae and adults thus have very different challenges to water and ion balance.

Larval *A. aegypti* ingest the medium more rapidly when food is available and increase their drinking rates as the ambient salinity increases. Serotonin levels in hemolymph increase in response to increased salinity but do not change with feeding status. As drinking rates increase with increasing ambient salinity, serotonin seems to act to stimulate absorption of ingested fluid from the midgut and secretion of fluid by the Malpighian tubules, leading to clearance of material from the hemolymph. For this strategy to work, at least some of the secreted fluid must be recovered in the hindgut, suggesting that hindgut transport is also regulated.

Adult females of most mosquito species survive for days at a time, feeding only on plant juices and other fluids, then consume within a few minutes a blood meal that may triple their body mass. Even while feeding, they begin to eliminate the bulk of the fluid and ionic load ingested with the meal. This response appears to be hormonally mediated, with CRF-related diuretic peptides, kinins, and serotonin all stimulating Malpighian tubule secretion. In part because of the small size and limited hemolymph volume of these insects, it is not known which of these potential hormones are involved in the response to feeding.

Lepidoptera: Butterflies and Moths

Lepidopteran larvae are able to adjust to diets of different water content by regulating the amount of water lost in the feces. This regulation seems to be under hormonal control. Both CRF-related diuretic peptides and kinins stimulate Malpighian tubule secretion *in vitro,* but they seem to have very different biological roles. When injected into caterpillars, CRF-related diuretic peptide (Mas-DH) has an antidiuretic effect: it stimulates fluid (and presumably ion) recovery from the cryptonephric complex. In contrast, kinins act to increase the fluid content of the feces, leading to marked reduction in weight gain by injected larvae. Allatotropins and a group of small peptides with an amino terminus amino acid sequence of Phe-Leu-Arg-Phe-NH_2 (FLRFamides) have been found to inhibit ion transport across the midgut epithelium, but the physiological significance of this response in terms of water and ion balance is not known.

Upon emerging from the chrysalis, butterflies discharge a considerable volume of fluid containing metabolic wastes accumulated during the pupal stage. This diuresis is regulated by unidentified hormones.

Desert Beetles

Beetles of the family Tenebrionidae are generally very tolerant of dry conditions, to the extent that larvae of at least some species can remove water vapor from unsaturated air across the rectum. Despite this, factors that stimulate Malpighian tubule fluid secretion to high rates have been found in some of the most drought-tolerant species of this family. When injected into the animals, these factors stimulate fluid secretion by Malpighian tubules; but rather than being passed back through the hindgut and eliminated, the secreted fluid is passed forward into the midgut. The water and valuable hemolymph components are recovered, leaving toxins and metabolic wastes in the gut where they can be eliminated. These actions have led to the suggestion that factors regulating excretory function often are more properly called "clearance hormones" instead of "diuretic hormones."

See Also the Following Articles

Excretion • Hemolymph

Further Reading

Chung, J. S., Goldsworthy, G. J., and Coast, G. M. (1994). Haemolymph and tissue titres of achetakinins in the house cricket *Acheta domesticus:* Effect of starvation and dehydration. *J. Exp. Biol.* **193,** 307–319.

Clark, T. M., and Bradley, T. J. (1997). Malpighian tubules of larval *Aedes aegypti* are hormonally stimulated by 5-hydroxytryptamine in response to increased salinity. *Arch. Insect Biochem. Physiol.* **34,** 123–141.

Coast, G. M. (1998). Insect diuretic peptides: Structures, evolution and actions. *Am. Zool.* **38(3),** 442–449.

Nicolson, S. W. (1991). Diuresis or clearance: Is there a physiological role for the "diuretic hormone" of the desert beetle *Onymacris? J. Insect Physiol.* **37(6),** 447–452.

Patel, M., Hayes, T. K., and Coast, G. M. (1995). Evidence for the hormonal function of a CRF-related diuretic peptide (Locusta-DP) in *Locusta migratoria. J. Exp. Biol.* **198,** 793–804.

Petzel, D. H., and Stanley-Samuelson, D. W. (1992). Inhibition of eicosanoid biosynthesis modulates basal fluid secretion in the Malpighian tubules of the yellow fever mosquito *(Aedes aegypti). J. Insect Physiol.* **38(1),** 1–8.

Phillips, J. E. (1983). Endocrine control of salt and water balance: Excretion. *In* "Endocrinology of Insects" (R. G. H. Downer and H. Laufer, eds.), pp. 411–425. A. R. Liss, New York.

Spring, J. H. (1990). Endocrine regulation of diuresis in insects. *J. Insect Physiol.* **36(1),** 13–22.

Weevil

see *Boll Weevil*

Wings

Robin J. Wootton
University of Exeter, United Kingdom

Insects are the only invertebrates to have developed the power of flight, and their wings are unique. Unlike those of birds and bats, they are not modified limbs and so have no internal muscles; they consist predominantly of an extracellular material—cuticle. Typical wings nonetheless need to perform—semiautomatically—the complex movements and cyclic changes in attitude and shape associated with flapping flight, and these necessities strongly influence their structural design. The latter shows huge variation within the class Insecta, and this is made greater by the wide range of secondary functions that the wings have evolved to perform, sometimes instead of flight, sometimes in addition to it. While great progress has been made in interpreting wing structures in functional terms, many questions still remain to be answered.

ORIGIN

We have no fossil record of the earliest winged insects nor of the structures from which wings evolved. Majority opinion currently favors their origin from articulated lateral outgrowths—"exites"—of an obsolete basal segment of the legs, long since incorporated in the thoracic wall. Similar, though not homologous, structures are found on the coxae and abdominal segments of some apterygote insects. The articulated gill plates of juvenile Ephemeroptera may actually be wing homologues. Those of some fossil mayfly nymphs indeed resemble tiny wings, though this may simply reflect the fact that both are adapted for accelerating fluid.

The origin of flight is still controversial. Two conflicting theories both have significant support; in 2001 they, and their implications on the later development of flight, were reviewed and compared in a chapter by Wootton in *Insect Movement: Mechanisms and Consequences.* The first, and more traditional, theory is supported by some experimental evidence and theoretical modeling and proposes that powered flight arose in tree-dwelling insects, through parachuting and gliding stages. These "protopterygotes" would have had winglets, probably movable, on most body segments. Selection would favor enlargement and improvement of winglets close to the center of mass of the body, and reduction and loss of the remainder, together with the evolution of an effective mechanism for flapping and twisting the wings. The second hypothesis suggests that flight arose in semiaquatic insects, initially using their winglets to skim on the surface film of water bodies, as do many modern adult Plecoptera and Ephemeroptera. Selection would favor enlargement and improvement of the thoracic winglets to the point at which they could generate sufficient upward force for the insects to leave the surface and fly.

STRUCTURE

Although some Carboniferous and Permian insects had short, apparently movable winglets on the first thoracic segment, the wings of modern insects are borne on the second and third segments only. They articulate with the back (tergum) and sides (pleura) of the thoracic segments via the axillae: complex, three-dimensional mechanisms of stiff and compliant cuticle through which the muscular forces of the thorax are transmitted to the wings and which control the relative movements of the wings' basal components. The axillary structures vary greatly in detail, but those of most insects can be referred to a common plan of up to four so-called axillary sclerites and three other sclerotized plates, the humeral and the proximal and distal median plates, linked together by hinges and broader expanses of soft cuticle (Fig. 1). The axillary structures of the Ephemeroptera and Odonata, orders which appear primitively incapable of folding their wings back over the abdomen, differ from the typical pattern of wing-folding (neopterous) insects and also from each other. Various attempts have been made to homologize these with the neopterous pattern, which is clearly influenced by the need to fold; but no consensus has been reached.

Orthodox wings themselves consist of cuticular membrane supported by, and continuous with, a framework of veins. Both veins and membrane are double structures, formed by juxtaposition of the cuticle of the dorsal and ventral sides as the

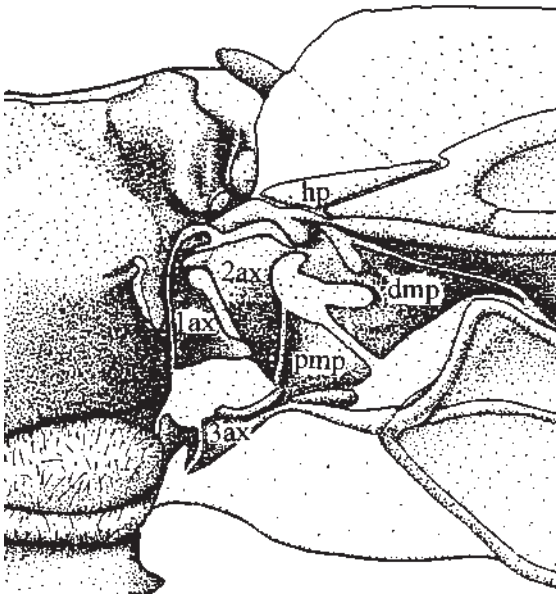

FIGURE 1 Forewing axilla and wing base of the desert locust, *Schistocerca gregaria.* 1ax, first axillary sclerite; 2ax, second axillary sclerite; 3ax, third axillary sclerite; dmp, distal median plate; hp, humeral plate; pmp, proximal median plate. (Modified with permission from Wootton, 1979.)

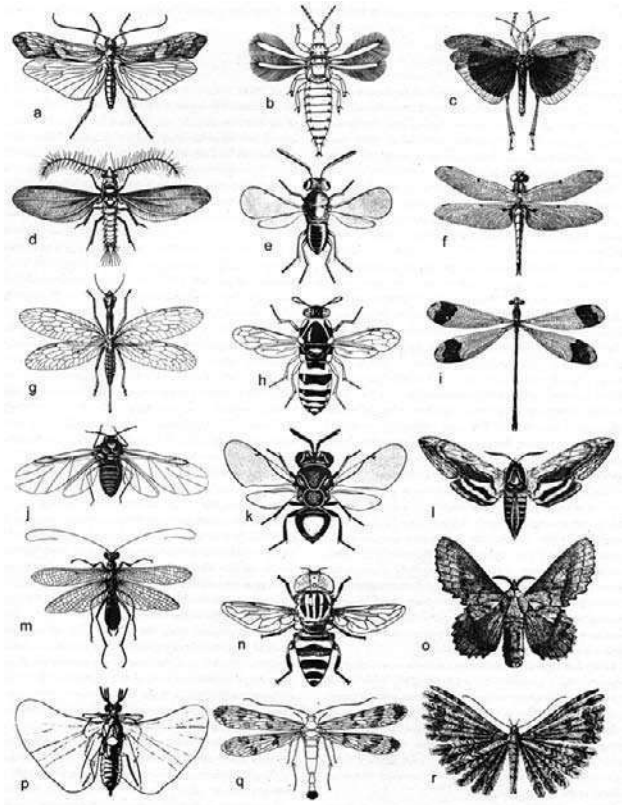

FIGURE 2 The variety of insect wings. The wing spans extend over two orders of magnitude, from ca. 1 mm in the thrips (b) to 100 mm in the damselfly (i). (a) Caddis (Trichoptera), *Limnophilus rhombicus.* (b) Thrips (Thysanoptera), *Liothrips oleae.* (c) Grasshopper (Orthoptera), *Dissosteira carolina.* (d) Male coccid (Homoptera), *Icerya purchasi.* (e) Parasitoid wasp (Hymenoptera), *Coccophagus tschirchii.* (f) Dragonfly (Odonata), *Libellula quadrimaculata.* (g) Snake fly (Megaloptera), *Raphidia adanata.* (h) Wasp (Hymenoptera), *Celonites abbreviatus.* (i) Damselfly (Odonata), *Megaloprepus coerulatus.* (j) Aphid (Homoptera), *Eriosoma lanigerum.* (k) Perilampid wasp (Hymenoptera), *Perilampus chrysopae.* (l) Hawk moth (Lepidoptera), *Hyloicus ligustri.* (m) Mantis (Mantodea), *Mantoida brunneriana.* (n) Hover fly (Diptera) *Lathyrophthalmus quibquelineatus.* (o) Lasiocampid moth (Lepidoptera), *Gastropacha quercifolia.* (p) Male stylopid (Strepsiptera), *Eoxeonos laboulbenei.* (q) Scorpionfly (Mecoptera), *Panorpa communis.* (r) Plume moth (Lepidoptera), *Orneodes cymodactyla.* (Reproduced with permission from W. Nachtigall, 1974.)

wing increases in area, thins, and stiffens after the final molt. The veins are typically tubular and contain hemolymph, and major vein branches usually also carry branches of the tracheal system. The hemolymph circulates or moves tidally within the veins and serves to maintain the moisture content of the cuticle, which would otherwise become stiff and brittle. Little or no epidermal material appears to remain in the expanded, sclerotized adult wing.

The longitudinal veins run distally from the wing base and many branch as the wing broadens along the span. They are usually to some extent linked by crossveins. Together they form a supporting and conducting framework, which may be structurally very complex (Figs. 2c, 2f, 2i, and 2m) or relatively simple (Figs. 2d, 2e, 2j, 2k, and 2p).

Even the simplest wings are far more complex than conventional illustrations indicate. Most have considerable relief, which adds substantially to their rigidity. The primitive condition may well have been a radially pleated structure, with the longitudinal veins alternately occupying the ridges and troughs of the pleats. This is found in many early fossil insects and is still the case in Odonata and Ephemeroptera. Some pleating indeed persists in parts of most insect wings, though it may be absent elsewhere. Many wings show an arched, or cambered, cross section, which again enhances rigidity to bending and introduces other important properties, which will be discussed later.

Further features, often omitted from illustrations, but with great mechanical significance, are lines and bands of flexibility within the wing and the axilla. One can conveniently distinguish between flexion lines, whose primary function is the facilitation and control of wing deformations in flight, and fold lines, whose primary role is in folding the wings to and from the resting position. The distinction is not absolute: there may well be some limited deformation along flexion lines during wing folding and along fold lines in flight. Some other lines of flexible cuticle, for example in the distal parts of some hymenopteran and dipteran wings, seem to fit neither category and are best interpreted as adaptations to allow the wing tip to crumple reversibly on impact with obstacles. The veins themselves vary considerably in diameter, wall thickness, and shape of their cross section, all factors that strongly influence their local resistance to bending and help determine the pattern of deformation in flight. Figure 3 shows some details of contrasting vein morphology. Many crossveins have a characteristic annulated form (Figs. 3f and 3g), a means of reconciling flexibility with the need to maintain an open

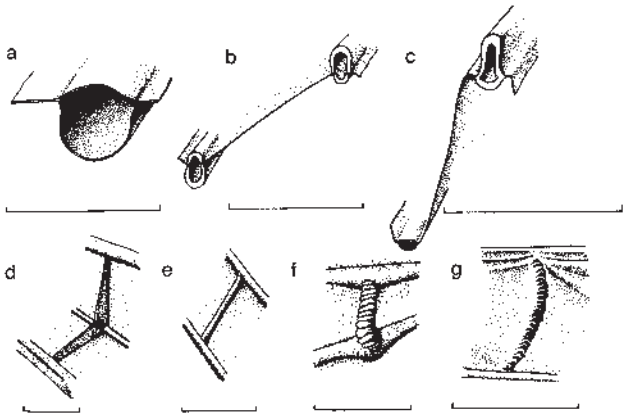

FIGURE 3 Details of veins. (a–c) Transverse sections of longitudinal veins. (a) Forewing subcosta of a butterfly, *Papilio rumanzowia* (Lepidoptera). (b) Adjacent veins in the radial sector field of *Calopteryx splendens* (Odonata). (c) Ridge and trough veins from the hind wing fan of *Schistocerca gregaria* (Orthoptera). (d–g) Crossveins. (d) Angle-bracket-like crossveins from the leading edge spar of *C. splendens* (Odonata). (e) Crossvein from the remigium of the wing of *C. splendens.* (f) Flexible crossvein from *Eristalis tenax* (Diptera). (g) Flexible crossvein from the hind wing fan of *S. gregaria.* Scale bars, 0.5 mm. (Modified with permission from R. Wootton, 1992.)

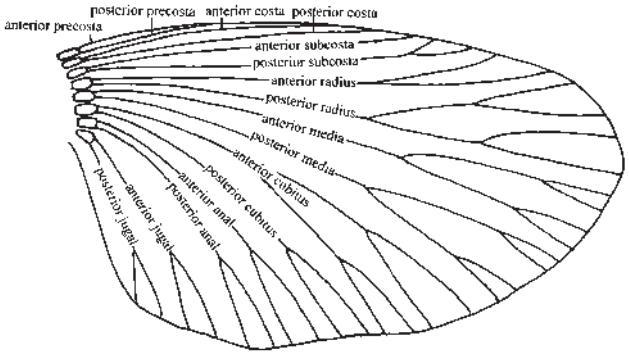

FIGURE 4 Wing vein nomenclature. The system of Kukalová-Peck. (Modified from Kukalová-Peck, 1991.)

section, as in flexible drinking straws and some hoses and flexible pipes. Elsewhere, crossveins can be stout and rigid and can even form high-relief angle brackets, as in the leading edge spar of dragonfly wings (Fig. 3d). Where longitudinal or crossveins are crossed by flexion lines and fold lines they may be locally interrupted, thinned, or annulated.

The membrane itself can be thickened and sclerotized, either over substantial areas, as in the protective or semiprotective forewings of Coleoptera, Dermaptera, Heteroptera, Orthoptera, and Blattodea, or in more restricted lines or zones. In the latter case it may not always be easy to distinguish the result from veins, especially since some true veins are not tubular, but are simply thickened, sclerotized grooves.

The texture of the membrane is also variable. Under the microscope, areas that undergo appreciable stretching in flight may appear crumpled like crêpe paper, as in the hind wing fan of locusts, or as an expanse of hollow papillae, as in those of some Heteroptera. Microtrichia (small unarticulated hairs) are widespread, and the wing membrane of many insects, particularly among Holometabola, bears articulated hairs (e.g., Trichoptera) or scales (e.g., Lepidoptera, Psocoptera, mosquitoes). The scales are sometimes extraordinarily complex in structure, providing both insulation and color, including some spectacular physical colors. This complexity is particularly remarkable when one considers that each scale is produced by a single cell.

The veins too may bear scales, hairs, and other ornamentation, including airflow- and vibration-sensitive hairs and sense organs that detect distortions of the cuticle in flight.

In several orders the fore- and hind (meso- and metathoracic) wings are in physical contact during the wing stroke, the posterior part of the forewing overlying the anterior part of the hind and in many cases physically linked to it by a coupling mechanism. Some mechanisms may in fact provide sensory information about the relative positions of the wings rather than rigid mechanical coupling, but in Hymenoptera the hind wing leading edge bears a row of hooks that firmly clasp a fold of the forewing; in Heteroptera the hind wing edge is held by hairs in a groove in the forewing, and in many Homoptera the wings are held together by interlinking grooves.

HOMOLOGY AND TERMINOLOGY

Despite occasional claims to the contrary, there is no good evidence that wings arose more than once within the insects, and it is generally accepted that homologous wing characters should be recognizable throughout the class. In fact this is anything but straightforward. In the course of evolution some veins have degenerated, or their identities have become obscured by fusion. Vein branches have been gained or lost, have become realigned, have come to resemble crossveins. Linear, vein-like membrane thickenings have arisen and are readily mistaken for true veins. The homologies of flexion lines and fold lines, and the extent to which they can be used as landmarks in vein identification, are also far from obvious.

Vein Nomenclature

Unfortunately, several conflicting systems are in use. All derive principally from a scheme established in the late 19th century by J. H. Comstock and J. G. Needham after a series of comparative studies of wing tracheal supply, but the situation has become confused by later additions and alterations. Two widely used schemes are illustrated here. Figure 4, adapted from Kukalová-Peck, shows all the longitudinal veins of which any trace remains in extinct, as well as living, insects and is an attempt at an overall ground plan for the winged insects. It proposes that veins were initially paired, the anterior member of each pair being convex and situated on a ridge of a pleated wing, the posterior member concave and situated in a trough. They are the anterior and posterior precosta, costa, subcosta, radius, media, cubitus, anal, and jugal. This scheme is consis-

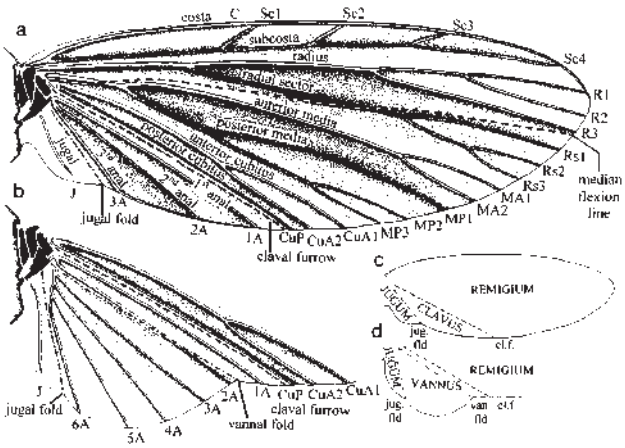

FIGURE 5 Wing nomenclature. A more traditional system, after Wootton, showing veins, some flexion and fold lines, and the main areas of the wing. (a, c) A forewing or a hind wing lacking an expanded anojugal area. (b, d) Cubital and anojugal region of hind wing with an expanded anojugal area. (Reproduced with permission from Wootton, 1979).

tent and logical and is increasingly widely followed. In most modern insects, however, several of the most anterior veins are obscure or absent, and many entomologists tend to ignore these, or indeed to question their existence, and to follow a scheme closer to that of Comstock and Needham. Figure 5 illustrates one such system, showing (Fig. 5a) a diagrammatic wing, (Fig. 5b) the expanded anal fan (vannus) of some hind wings, and (Figs. 5c, and 5d) the names in common use for specific regions of the wing. Kukalová-Peck's precostal veins are omitted, as are her posterior costa and anterior subcosta. The anteriormost veins are hence the convex, unpaired costa and the concave subcosta. Kukalová-Peck's posterior radius is called by the older name of radial sector. The number of anal veins is regarded as unfixed.

Both diagrams show most veins to be branched, as is often the case, but we do not in fact know what was the primitive number of branches for any vein. Crossveins are omitted from the diagrams. Where comparatively few are present in a wing, these are often named by the longitudinal veins that they connect, with the symbols for these veins given in lowercase type.

Regions of the Wing

In the wings of a neopterous (wing-folding) insect distinct areas can be recognized, delimited by flexion lines or fold lines. The main, anterior area, between the leading edge and the claval furrow, is the remigium. In wings without an expanded anal fan, a narrower area between the claval furrow and the jugal fold is the clavus. The remaining, usually small area, which lies posterior to and inboard of the jugal fold, is the jugum. In the hind wings of many forms the area between the claval furrow and the jugal fold is expanded into a broad anal fan, or vannus, which sometimes, notably in Orthoptera, Phasmida, and Blattodea, has fold lines within it and folds into a pleated fan.

Flexion Lines and Fold Lines

Flexion lines and fold lines have received comparatively little attention and indeed are often omitted from diagrams. In neopterous insects, one flexion line and one fold line are particularly widespread. The claval furrow is a flexion line that typically lies posterior to the posterior cubitus and allows the remigium to hinge upward or downward relative to the clavus. The jugal fold is continuous with the basal hinge of the wing and is involved in both flapping and folding.

Two other flexion lines are common, but their positions relative to other wing landmarks are not constant, and they may have arisen independently several times. The median, or remigial flexion line, runs longitudinally from the wing base, often between the radial and the median vein systems. It allows the profile of the wing to be altered in flight and may influence its resistance to transverse bending. Transverse flexion lines, by contrast, facilitate ventral bending across the span. A transverse flexion line occurs in several orders, crossing the wing along a curved or irregular, sometimes oblique, path from the costa to the posterior margin.

DIVERSITY

The wings of insects are very diverse indeed—far more so than those of birds or bats. They vary in proportions; the relative sizes of fore- and hind wings; relief, texture, and ornamentation; thickness; venational richness; the details of veins, fold lines, and flexion lines; and the extent and manner in which they are coupled in flight. Figure 2 illustrates some of this diversity.

It is widely assumed, though without direct evidence, that fore- and hind wings were originally similar in size and shape. This is true today of zygopterous Odonata (Fig. 2i); of many Neuroptera, Megaloptera (Fig. 2g), Mecoptera (Fig. 2q), Trichoptera, and primitive Lepidoptera; of Thysanoptera (Fig. 2b), Embioptera, most Isoptera, and a few Plecoptera (Chloroperlidae). In the last two orders at least the condition is clearly secondary. Some Permian fossil Ephemeroptera and Heteroptera also had similar fore- and hind wings, as did some members of several extinct orders.

However, some of the earliest insect fossils already show relative enlargement or reduction of one wing pair, and one or the other applies to most insects today. In anisopterous Odonata (Fig. 2f); Orthoptera (Fig. 2c) and winged Phasmida; Blattodea; one family of Isoptera (Mastotermitidae) and Mantodea (Fig. 2m); most Plecoptera, Dermaptera, Trichoptera (Fig. 2a), and Coleoptera; and some Heteroptera and male Strepsiptera (Fig. 2p), the hind wing is significantly broader than the fore and has become the principal lifting surface. In Ephemeroptera, Psocoptera, some Heteroptera (Fig. 2j) and Neuroptera, many Lepidoptera (Figs. 2l, 2o, and 2r), most Hymenoptera (Figs. 2e, 2h, and 2k), and all Diptera (Fig. 2n) the converse is true; the forewings have greater area and usually greater length than the hind. In all but the Odonata, which do not fold their wings, it is necessary for the hind wings to fold up along two

or more radiating fold lines to lie beneath the forewings at rest. The forewings themselves are often shorter than the hind, and in some cases have become thickened, providing some protection for the abdomen and the more fragile hind wings. When moderately thickened, as in many Orthoptera, Blattodea, Mantodea, and some Homoptera, the forewings are often referred to as tegmina (singular tegmen). When thickened to the extent that their flight function is almost lost or wholly so, as in Coleoptera and Dermaptera, they are known as elytra (singular elytron). The typical forewings of Heteroptera, with the basal part thickened but a membranous tip, are known as hemielytra.

Forewings adapted for protection tend to be less effective in flight and are usually accompanied by disproportionately large, elaborately folding hind wings with expanded anal fans or else by total loss of flight capability. Where the forewings are significantly shorter than the hind, e.g., in some Orthoptera, winged Phasmida, Dermaptera, and Coleoptera, two different solutions have been adopted to meet the problem of stowing the hind wings at rest. In Orthoptera, e.g., Gryllotalpidae, Tridactylidae, and Tetrigidae, and in Phasmida, the longitudinally folded hind wings project from beneath the tegmina along the back of the abdomen. In Phasmida and Tridactylidae, the narrow remigium of the hind wing itself is thickened and protects the more delicate fan. However, in Dermaptera and Coleoptera and in several genera of Blattodea the hind wings fold transversely as well as longitudinally, in a variety of complex and mechanically fascinating ways.

Some reductions and modifications are extreme. Among Ephemeroptera, the Caenidae and some Baetidae have lost the hind wings entirely and fly with the forewings alone; the same is true of male Coccidae (Homoptera, Fig. 2d). The hind wings of Diptera are modified as gyroscopic sense organs, the halteres, and the forewings of male Strepsiptera (Fig. 2p) are apparently similarly adapted. The hind wings of nemopterid Neuroptera are elongate plumes or filaments. In the nemopterid species that have been investigated they do not flap, but are spread out behind the insect, apparently acting as physical stabilizers. Similar long, more or less slender hind wings occur in several families of butterflies and moths, but in some species at least they flap in phase with the forewings, sending ripple-like waves along their "tails."

The wings of many small moths, flies, beetles, and wasps are fringed with long hairs, which appear to operate aerodynamically as if they were a continuous membrane. This is taken to extremes in the tiny feathery wings of Thysanoptera (Fig. 2b), ptiliid beetles, nymphomyiid Diptera, and mymarid Hymenoptera, of which most of the lifting surface consists of long hairs, arising from a rod-like rachis. The emarginated wings of orneodid moths (Fig. 2r) consist of a radiating fan of similar feather-like structures, and the pterophorid moths approach this condition, though with fewer plumes.

The forewings of many male Orthoptera in the suborder Ensifera are adapted for sound production and amplification, with distorted venation enclosing large areas of membrane—the "mirrors." These regions, and their function, are often retained in species whose wings are otherwise reduced. The forewings of some noctuid moths are also adapted for sound production.

Reduction (brachyptery) and loss (aptery) of both wing pairs have happened in many orders. It is particularly frequent in Orthoptera, Blattodea, and Hemiptera, and sometimes occurs polymorphically within a species. The adaptive significance of brachyptery is seldom clear. Short forewings are sometimes retained, apparently as protective structures, as are the elytra of many flightless beetles that lack hind wings.

FUNCTIONS AND DESIGN

The primary function of wings is flight. The wings operate as flexible aerofoils, oscillating as first-order levers over a process of the pleuron, which acts as a fulcrum. They need to be rigid enough to resist unwanted deformation under the forces they receive from the air and the inertial forces from the repeated accelerations and decelerations of flapping, but locally flexible enough to allow the cyclic twisting and bending deformations that are necessary to generate the forces to propel the insect and support its weight in air. Because they have no internal muscles, their shape from instant to instant needs to be controlled partly remotely though the axillae, partly automatically by their own structure: the distribution of rigid and flexible elements, of relief, veins and flexion lines within the wings themselves.

The Wing Beat

To interpret these aspects of wing morphology it is necessary to understand the nature of the wing beat. Simply to beat wings symmetrically up and down would be aerodynamically ineffective, because the vertical forces would tend to cancel out. It necessary to introduce some asymmetry to the wing beat, in the path of the wings and/or their shape and attitude. An abundance of high-speed still and cine photographs of a wide range of insects shows that the downstroke has many features common to all. The wings beat downward, and often rather forward relative to the insect's body axis, and are usually relatively straight along their length, with a rather curved section and a slight nose-down twist along the span. In the upstroke, however, the shape of the wings varies greatly, both between species and according to how the insect is at that moment performing.

In virtually all species the wings twist backward (supinate) to some extent at the bottom of the downstroke and continue supinated throughout the upstroke, twisting forward (pronating) again at the top. The extent to which they can twist varies greatly, as was illustrated by Wootton in 2000. The wings of most Diptera, and of zygopterous Odonata, which have slender bases and operate in individual pairs, are able to twist almost onto their backs. The camber becomes reversed, and they are then capable of generating upward, weight-supporting forces on the upstroke as well as the downstroke. This allows them to fly slowly and to hover. At the other

extreme, locusts, and probably most other flying Orthoptera, can twist their wings only slightly and get the necessary stroke asymmetry by limited forewing twisting, but mainly by pulling the hind wings fully forward in the downstroke and partly retracting them for the upstroke. Pulling the hind wing forward extends the fan, and the radiating veins within it become compressed and curved like the spokes of an umbrella, giving the wing an aerodynamically effective cambered section, which is lost as the wing relaxes into the upstroke. Only the downstroke generates significant weight support, and the insects can fly only forward, at a relatively high speed.

Between these two extremes fall many insects whose wings are capable of varying amounts of torsion and which hence display a range of versatility and maneuverability in flight. Broad wings and fore- and hind-wing coupling both tend to limit active twisting, but many wings show a degree of passive torsion along their span, so that the distal part is significantly supinated for the upstroke and can generate useful lift. This is in many insects—Plecoptera, sialid Megaloptera, Mecoptera, many Hemiptera and Hymenoptera, some Trichoptera and Lepidoptera—enhanced by a degree of ventral bending along a more or less marked transverse flexion line. Where such a flexion line runs obliquely across the wing, ventral bending serves also to twist the distal part of the wing, enhancing upstroke lift generation and increasing flight versatility.

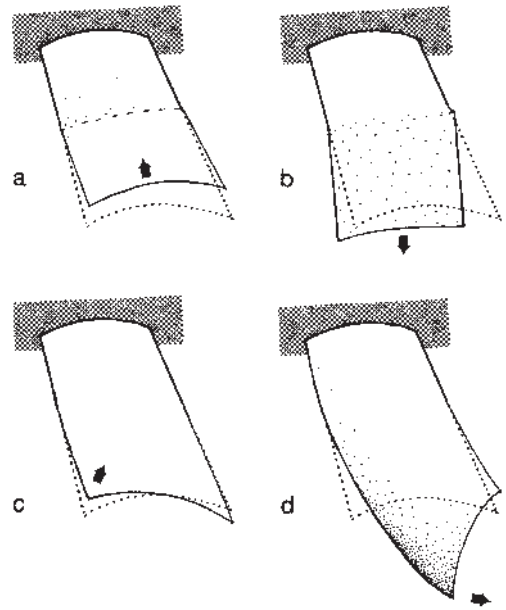

FIGURE 6 Mechanical behavior of cambered plates. (a, b) Cambered plates, fixed at one end, are more easily bent by a force from the convex side than from the concave side. (c, d) Such plates are more readily twisted by a force applied from the convex side behind the torsional axis than from the concave side.

The Importance of Wing Camber

A cambered cross section makes a wing rigid to both bending and torsion, but asymmetrically so. A thin, dorsally convex wing is far more resistant to a bending force applied from below (Fig. 6a) than to one from above (Fig. 6b), as the latter tends to make the edges buckle outward and flatten the section. Furthermore if the force is applied behind the axis along which the wing naturally twists, the latter twists far more easily if the force is from above (Fig. 6d) than from below (Fig. 6c). These effects are exploited by many insects to favor ventral bending and supinatory twisting for the upstroke, whether or not a transverse flexion line is present. There is evidence, moreover, that some insects can actively alter the height of the camber, and hence the wing's rigidity, by longitudinal bending along the median flexion line, and this may prove to be the latter's principal function.

Automatic Mechanisms

The mechanisms so far described illustrate that the control of wing shape in flight is to a considerable extent automatic: determined by the local distribution of flexibility and rigidity within the structure, moderated remotely by the controlling muscles at the wing base. Nowhere is this automation better demonstrated than in the wings of dragonflies (Odonata, Suborder Anisoptera).

These agile aerial predators have long complex wings (Figs. 2f and 2i), extensively corrugated, with several richly branched longitudinal veins and an abundance of crossveins. The principal support is provided by the costa, subcosta, and radius, which are linked together in the basal part of the wing by high-relief, angle-bracket-shaped crossveins (Fig. 3d) into a spar with a V-shaped cross section, rigid to both bending and torsion. About halfway along the wing the subcosta terminates at the "nodus," a high-relief structure consisting of a strong transverse strut and a patch of flexible cuticle that interrupts the costa. Beyond the nodus the costa and radius form with the first branch of the radial sector a more shallow spar with an inverted V-shaped section, supporting the distal part of the wing but compliant to supinatory twisting. The crossveins here, and elsewhere in the wing, are slender, allowing considerable deformation, so that the flapping wing twists forward and backward about its leading edge spar, like the sail of a tacking dinghy. The nodus serves to reinforce the wing and to minimize stress concentrations at the transition point between the two, mechanically very different, sections of the leading edge spar. The extreme twisting, essential to its versatile, high-performance flight, is assisted by an unusually flexible trailing edge; but it is essential that this latter should not be allowed to swing up like a flag to flutter in the airflow, which would be aerodynamically useless and potentially damaging to the wing. This is automatically prevented by two features: a dense spot called the pterostigma, near the end of the wing tip, which is believed to act as a regulating counterweight to the swinging wing, and a curious, three-dimensional formation of veins near the wing base—the triangle and supratriangle—which automatically levers the trailing edge downward in direct

response to the aerodynamic forces that the wing generates. This degree of automation in a propulsive appendage is unique in the animal kingdom and has no obvious parallels in technology.

Wing Folding

All modern insects except the Odonata and Ephemeroptera have the ability to fold their wings back over the abdomen, an adaptation that increases their ability to move freely over a substrate and to penetrate small spaces. Simple folding back is achieved mainly by muscular rotation of the third axillary sclerite and folding of the wing base along the basal hinge between the sclerite and the side of the thorax and along the jugal fold line. In hind wings with a large vannus (anal fan) there may be many extra fold lines lying parallel to the radiating veins, so that the wing folds like a pleated fan. The hind wings of Dermaptera, Coleoptera, and the cockroach *Diploptera,* however, fold up transversely as well as longitudinally; in some cases with extraordinary complexity. Because the wings have no internal muscles, folding and unfolding need to be achieved remotely, by mechanisms operated from the wing base and by the action of other body parts, combined with elasticity within the wing. Hydraulic mechanisms have sometimes been postulated, but recent work suggests that these are unnecessary. The unfolding of beetle wings appears to be achieved by active, scissor-like movements of the bases of the main supporting veins, which set in action a cascade of interconnected, simple opening mechanisms similar to those used in pop-up books and three-dimensional greeting cards; assisted in places by elasticity in the cuticle. Refolding the wings involves the reversal of these mechanisms, assisted by movements of the abdomen. Dermaptera, by contrast, use their abdomens to unfold the hind wings by means of a complex array of similar mechanisms, and refolding is largely elastic.

Secondary Functions

Wings have assumed a range of secondary functions. The sound-producing and -amplifying forewings of many Orthoptera, the protective elytra and tegmina of several orders, and the sensory halteres of Diptera and Strepsiptera have already been mentioned. The wings of Odonata and many Lepidoptera assist in thermoregulation. Members of several orders—Orthoptera, Phasmida, Mantodea, Odonata, Diptera, Lepidoptera—use theirs in active signaling: defensive, territorial, and sexual. Wings with warning patterns and colors are found particularly among Coleoptera, Lepidoptera, Orthoptera, and Heteroptera. The use of wings in camouflage is especially widespread in Orthoptera, Mantodea, Heteroptera, and Lepidoptera and may be reflected in their shape as well as color and pattern. Those of leaf mimics may be unusually broad and may bear petiole-like protuberances and emarginations resembling insect damage. It is often hard to determine whether particular wing characters are flight adaptations or have only secondary functions.

Unsolved Problems

Many wing characters are still poorly understood. We know far too little about the aerodynamic significance of the great variety of wing shapes in, for example, Lepidoptera, and of surface structures like hairs, spines, and scales. While the principles underlying venation pattern and the layout and orientation of flexion lines are broadly clear, the details are often quite obscure. Much work, linking aerodynamics, functional morphology and structural mechanics, remains to be done.

See Also the Following Articles

Anatomy • Exoskeleton • Flight • Fossil Record • Integument • Tracheal System

Further Reading

Brodsky, A. K. (1994). "The Evolution of Insect Flight." Oxford University Press, Oxford/New York.

Dudley, R. (2000). "The Biomechanics of Insect Flight." Princeton University. Press, Princeton, NJ.

Kukalová-Peck, J. (1991). Fossil history and the evolution of hexapod structures. *In* "The Insects of Australia" (CSIRO, eds.), 2nd ed., Vol. 1. Cornell University Press, Ithaca, NY.

Grodnitsky, D. L. (1999). "Form and Function of Insect Wings." Johns Hopkins University Press, Baltimore/London.

Nachtigall, W. (1974). "Insects in Flight." Allen & Unwin, London.

Wootton, R. J. (1979). Function, homology and terminology in insect wings. *Syst. Entomol.* **4,** 81–93.

Wootton, R. J. (1981). Support and deformability in insect wings. *J. Zool. London* **193,** 447–468.

Wootton, R. J. (1991). The functional morphology of the wings of Odonata. *Adv. Odonatol.* **5,** 153–169.

Wootton, R. J. (1992). Functional morphology of insect wings. *Annu. Rev. Entomol.* **37,** 113–140.

Wootton, R. J. (2001). How insect wings evolved. *In* "Insect Movement: Mechanisms and Consequences" (I. Woiwod, D. R. Reynolds, and C. D. Thomas, eds.). CAB Int., Wallingford/New York.

Wootton, R. J., Evans, K. E., Herbert, R., and Smith, C. W. (2000). The hind wing of the desert locust (*Schistocerca gregaria* Forskål). 1. Functional morphology and mode of operation. *J. Exp. Biol.* **203,** 2933–2943.

Wolbachia

Richard Stouthamer

University of California, Riverside

In the heyday of light-microscopy studies, intracellular bacteria were discovered in many organisms. In 1924 M. Hertig and S. B. Wolbach observed bacteria in the cells of a mosquito *(Culex pipiens).* Hertig formally described them in 1936 as *Wolbachia pipientis* in honor of his collaborator in the earlier work. *Wolbachia* was placed in the Rickkettsiales, which belong to the α-Proteobacteria. Its closest relatives are several pathogenetic bacteria (*Cowdria* and *Ehrlichia*) that are transmitted by arthropods to vertebrate hosts. After their discovery little

was reported on these bacteria until Yen and Barr in 1971 discovered that they were involved in causing the death of mosquito eggs when sperm of infected males fertilized the eggs of *Wolbachia*-free females. Since that time the interest in *Wolbachia* has increased rapidly because *Wolbachia* appears to be very common in arthropods (up to 70% of all insects) and causes a number of unusual manipulations of their host's reproduction. These manipulations are interesting both from an academic point of view (evolution of sex, speciation, conflict between different sets of genes within one organism) and from an applied point of view (Can these bacteria be used to manipulate their host for our benefit?).

WOLBACHIA'S HOST MANIPULATIONS

Wolbachia is generally passed on only from an infected female to her offspring; infected males do not pass on their infection to their offspring. This can be understood because insect eggs are large and can harbor bacteria, whereas sperm are small and carry very little cytoplasm in which bacteria can be transported. Such asymmetry in the transmission ability of their host explains many of the effects that *Wolbachia* has. An extreme case of this host manipulation is the induction of parthenogenesis (a type of asexual reproduction) in wasps by *Wolbachia.* Infected female wasps produce only daughters; in addition, they do not have to mate to produce daughters. Some wasp species consist entirely of infected females. In this case *Wolbachia* has manipulated its host into producing only females that can transmit the bacterium. In addition no eggs are "wasted" on the production of males, which are a dead end for *Wolbachia.* Other manipulations will be discussed below.

Cytoplasmic Incompatibility

One of the best known and possibly the most common effect of *Wolbachia* is cytoplasmic incompatibility (CI). Several different forms of CI are known. In its simplest form, CI causes the death of embryos when a sperm from an infected male fertilizes the egg of an uninfected female. Consequently, in populations in which infected and uninfected individuals occur together, all possible crosses of infected females result in infected offspring. Uninfected offspring are produced only when an uninfected male mates with an uninfected female, whereas all the offspring of uninfected females that have mated with infected males die. This difference in offspring production causes the CI infection to rapidly spread through the population. In the fruit fly *Drosophila simulans,* the infection in California spread at a rate of approximately 100 km per year.

The underlying mechanism of CI is not yet understood, but it is believed that sperm of infected males are somehow modified during spermatogenesis so that only in eggs infected with the same type of *Wolbachia* can this sperm function properly or be "rescued." This system of modification (mod) of the sperm and rescue (res) in the eggs is used to understand the possible CI types. For instance if in one species two different *Wolbachia* infections exist, this could lead to two-way incompatibility because the modified sperm is not rescued by the eggs infected with the other type of *Wolbachia.* This leads to the death of all the embryos resulting from intertype crosses. Therefore two infections within one species can cause genetic isolation of the two different subpopulations. This isolation may speed up speciation.

Not only is it possible that different populations within a species are infected with a different *Wolbachia,* but superinfections with several *Wolbachia* types are also common. Individuals infected with two different *Wolbachia* may be incompatible with individuals that are infected with only one of the two types. Sperm from doubly infected males may not be rescued in eggs that are singly infected. In these situations the double infection should spread through the population, displacing the single infection.

The types of CI *Wolbachia* that have been discussed so far can all be classified as modifiers (mod+) and rescuers (res+). If we assume that both these functions are energy costly to maintain, one can imagine that once a complete population is infected with the mod+, res+ variant, *Wolbachia* variants that do not modify the sperm any longer (mod–, res+) will spread relative to the modifying form (mod+, res+). Finally, once all of the individuals carry the mod–, res+ variant, the population becomes vulnerable to invasion by mod–, res– variants. Mod–, res– host populations either may lose their infection or can be invaded by new mod+, res+ variants. Examples of mod+, res+, mod–, res+, and mod–, res– are all known. Mod+, res– *Wolbachia* do not exist because the males would be incompatible with the females of their own population.

CI *Wolbachia* is thought to be promising in the spread of beneficial traits through a population. The idea is to genetically modify a *Wolbachia* with genes that inhibit the transmission of disease-causing organisms (malaria, dengue fever, etc.) through the insects that vector the disease. If this gene were inserted into a CI *Wolbachia,* the cytoplasmic incompatibility caused by the *Wolbachia* would lead to the spread of the bacterium through the vector population and all infected insects would be unable to transmit the disease. *Wolbachia* appears to be present in many different parts of the insect body. Consequently the gene products inhibiting the disease organisms may also be delivered to various insect tissues.

Parthenogenesis-Inducing *Wolbachia*

Wolbachia inducing parthenogenesis in their hosts are known from many wasps and some thrips (Fig. 1). Infected females can produce daughters both from fertilized and from unfertilized eggs. The cytogenetic mechanism through which *Wolbachia* cause unfertilized eggs to grow into females is known for a number of parasitoid wasps. *Wolbachia* exploits the normal sex determination in wasps. In this system, males arise from unfertilized eggs, which contain only one set of chromosomes, whereas females arise from fertilized eggs, containing two sets

FIGURE 1 PI *Wolbachia* visible as light dots in a 4,6-diamidino-2-phenylindole-stained egg of the parasitoid wasp *Trichogramma kaykai.* The *Wolbachia* are concentrated around the area of the egg from which the future eggs of the wasp develop. Photograph by Merijn Salverda.

of chromosomes. In unfertilized, infected eggs containing only a single set of chromosomes, *Wolbachia* causes a doubling in the number of chromosomes. This is accomplished through a modification of the first mitotic division in the egg, causing the egg to develop into a female.

In most populations in which parthenogenesis-inducing (PI) *Wolbachia* are known, the entire population consists of females. Feeding antibiotics to these females causes them to produce male offspring. In most instances, these males are not able to successfully inseminate the females of their own line. It is assumed that the populations have been parthenogenetic for such a long time that mutations have accumulated in the part of the genome involved with sexual reproduction. An exception to this situation is found in the tiny parasitoid wasps of the genus *Trichogramma* in which, in many populations, both infected and uninfected individuals coexist. In these populations, males that are the offspring of antibiotic-treated mothers can successfully mate with infected females.

PI *Wolbachia* are of interest from the applied point of view as well. Many parasitoid wasps are reared on a large scale in insectaries for release in the field to control pest insects. Only female wasps are effective biological control agents because (1) they lay their eggs in or on the pest insect, (2) the wasp larvae will subsequently eat up the host insect, and (3) new wasps emerge from the remains of the pest insect. PI-*Wolbachia*-infected females are cheaper to produce than normal sexual wasps. No production effort is wasted on rearing of males that are worthless for biocontrol.

Feminizing *Wolbachia*

In many woodlice species (pill bugs, Crustacea, Isopoda), a *Wolbachia* is found that causes infected genetic males to develop into functional females. The feminizing *Wolbachia* accomplishes this through the infection of the male-hormone-producing gland. Without this hormone, genetic males develop into females. The spread of such an infection through a population can result in a severe shortage of males. Because males are needed for reproduction, this has led to counteradaptations by the nuclear genes in such a way that males will be produced even if the individual is infected. The effects of the *Wolbachia* and the counteradaptations against the feminization have led to a dynamic system of sex determination in some species of woodlice.

Feminizing *Wolbachia* have also been discovered in a moth (*Ostrinia scapularis:* Noctuidae). Infected individuals grow into females, but how the *Wolbachia* accomplishes this transition is not known.

Male-Killing *Wolbachia*

Many different microorganisms have evolved methods of killing only male offspring of infected mothers. *Wolbachia* have this ability in several lady beetles and butterflies. The advantage of killing male offspring to *Wolbachia* requires that the infected daughters somehow benefit from the death of their brothers. In the case of the lady beetles this is most easily explained. Lady beetles generally lay their eggs in clusters, and the survival of the larvae is greatly enhanced if they find food soon after hatching. If an infected mother has laid a clutch of eggs the daughters emerge and then consume the eggs containing their dead brothers as their first meal. Because males cannot pass on the *Wolbachia,* the enhanced survival of the infected females benefits the *Wolbachia.* The presence of the male-killing *Wolbachia* can lead to extreme sex ratios in the population and just like in the case of the feminizing *Wolbachia* counteradaptations are expected.

Male-killing *Wolbachia* occur with a high frequency in several species of the butterflies of the genus *Acraea* occurring in East Africa. This has led in some populations to extremely female-biased sex ratios and a concurrent change in behavior of the butterflies. Normally males form leks where females would "choose" the male to mate with, but in these female-biased populations females form leks where males "choose" their partners. Little is known about the mechanism that allows these *Wolbachia* to kill exclusively infected male eggs.

PHYLOGENY OF *WOLBACHIA*

The phylogeny of *Wolbachia* has been studied extensively using several *Wolbachia* genes. Initially the 16S ribosomal DNA sequence was used to derive the relationship between the different *Wolbachia.* From this work it became clear that within the insect *Wolbachia* two groups exist, later named A and B, that had diverged from each other approximately 60 mya. Later several other genes were used to get a finer scale picture of the relationship of the different *Wolbachia* to each other. From all this work two main conclusions can be drawn: closely related *Wolbachia* can have quite different effects on their host

and closely related hosts can carry quite different *Wolbachia.* To understand the distribution of the *Wolbachia* over the different hosts it is clear that transfer of *Wolbachia* between species must happen quite frequently over evolutionary time. The fact that the different host manipulations of *Wolbachia* can all be found in closely related *Wolbachia* must be the result of a relatively simple evolution of one effect into the other, for instance from CI to PI, or a more likely hypothesis is that the phylogenies based on different *Wolbachia* genes do not reflect the phylogeny of the genes involved in causing the host manipulation. If the genes causing the effects are found on phages that can jump from one *Wolbachia* to the next, the observed phylogenetic pattern can easily be explained. A potential candidate is the phage WO discovered by Masui *et al.* in 2000, which shows evidence of frequent horizontal transfer between different *Wolbachia.* More information on the genes involved in the different effects of *Wolbachia* will become available once its genome is published.

See Also the Following Articles

Genetic Engineering • *Sex Determination* • *Symbionts, Bacterial*

Further Reading

Dobson, S. L., Bourtzis, K., Braig, H. R., Jones, B. F., Zhou, W., Rousset, F, and O'Neill, S. L. (1999). *Wolbachia* infections are distributed throughout insect somatic and germline tissues. *Insect Biochem. Mol. Biol.* **29,** 153–160.

Hertig, M. (1936). The rickettsia, *Wolbachia pipiens* (gen. et sp.n.) and associated inclusions of the mosquito, *Culex pipiens. Parasitology* **28,** 453–486.

Jiggins, F. M., Hurst, G. D. D., and Majerus, M. E. N. (2000). Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. *Proc. R. Soc. Biol. Sci. B* **267,** 69–73.

Masui, S., Kamoda, S., Sasaki, T., and Ishikawa, H. (2000). Distribution and evolution of bacteriophage WO in *Wolbachia,* the endosymbiont causing sexual alterations in arthropods. *J. Mol. Evol.* **51,** 491–497.

O'Neill, S. L., Hoffman, A. A., and Werren, J. H. (eds.) (1997). "Influential Passengers: Inherited Microorganisms and Arthropod Reproduction." Oxford University Press, London.

Stouthamer, R., Breeuwer, J. A. J., and Hurst, G. D. D. (1999). *Wolbachia pipientis:* Manipulators of invertebrate reproduction. *Annu. Rev. Microbiol.* **53,** 71–102.

Turelli, M., and Hoffmann, A. A. (1991). Rapid spread of an inherited incompatibility factor in California *Drosophila. Nature* **353,** 440–442.

Werren, J. H., Zhang, W., and Guo, L. R. (1995). Evolution and phylogeny of *Wolbachia:* Reproductive parasites of arthropods. *Proc. R. Soc. London B* **261,** 55–63.

Yen, J. H., and Barr, A. R. (1971). New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens. Nature* **232,** 657–658.

Worker

see *Caste*

Y

Yellow Fever

Thomas P. Monath
Acambis Inc., Cambridge, Massachusetts

Yellow fever virus is the prototype of the family Flaviviridae, which includes approximately 70 single-stranded RNA viruses, the majority of which are transmitted by mosquitoes or ticks. Yellow fever virus is transmitted principally by insects (mosquitoes), but ticks *(Amblyomma variegatum)* may play a secondary and minor role in Africa.

DISEASE AND MEDICAL IMPACT

The disease caused by the yellow fever virus is a severe hemorrhagic fever, characterized by high viremia (virus in the blood), hepatic, renal, and myocardial injury, hemorrhage, and case-fatality rates of 20 to 50%. Today, the disease occurs in tropical Africa and South America. It is estimated that up to 200,000 cases occur annually but far fewer are reported officially. Between 1990 and 1999, 11,297 cases and 2648 deaths were reported; 83% of these cases occurred in Africa. The highest incidence was in Nigeria, where epidemics occurred between 1986 and 1994. Recent epidemics have affected Cameroon (1990), Kenya (1992), Ghana (1993–1994, 1996), Liberia (1995, 1998, 2000), Guinea (2000), Gabon (1994), Senegal (1995, 1996), and Benin (1996). In South America, 1939 cases and 941 deaths were reported between 1990 and 1999, for an average of about 200 cases per year. Bolivia and Peru had the highest incidence, reflecting low vaccination coverage. In Brazil, an increase in yellow fever activity occurred over a wide area between 1998 and 2000.

ROLE OF INSECTS IN TRANSMISSION

The enzootic transmission cycle involves tree-hole-breeding mosquitoes such as *Aemagogus janthinomys* (South America), *Aedes africanus* (Africa), and nonhuman primates. Infection of mosquitoes begins after ingestion of blood containing a threshold concentration of virus (~ 3.5 $\log_{10}$ ml^{-1}), resulting in infection of the midgut epithelium. The virus is released from the midgut into the hemolymph and spreads to other tissues, notably the reproductive tract and salivary glands. Seven to 10 days is required between ingestion of virus and virus secretion in saliva (the extrinsic incubation period), after which the female mosquito is capable of transmitting virus to a susceptible host. Vertical transmission of virus occurs from the female mosquito to her progeny and from congenitally infected males to females during copulation. Virus in the egg stage provides a mechanism for virus survival over the dry season when adult mosquito activity and horizontal transmission abate.

In tropical South America, the virus is transmitted by *Haemagogus* mosquitoes (principally *Hg. janthinomys*) between monkeys in the rain forest canopy and from monkeys to humans. Transmission peaks during months of high rainfall, humidity, and temperature (January–March), corresponding to the activity of *Haemagogus* mosquitoes. It is during occupational activities, such as forest clearing, lumbering, and road construction that humans acquire the infection ("jungle" yellow fever) from mosquitoes that previously had fed on viremic monkeys; most patients are young adult males. The density of *Haemagogus* and risk of human exposure are relatively low, and human cases occur in a sporadic fashion. Other mosquitoes involved in jungle yellow fever in South America include *Hg. leucocalaenus* and *Sabethes chloropterus.* In the early 1950s an epizootic extended to Central America, where other species were implicated *(Hg. lucifer, Hg. equinus, Hg.*

iridicolor, Hg. mesodentatus). Prior to the 1940s epidemics of yellow fever due to bites from *Ae. aegypti* occurred in the Americas. *Ae. aegypti* breeds in and around houses, principally in artificial containers, and is responsible for direct interhuman transmission of virus ("urban" yellow fever). Age and sex distributions are different from jungle yellow fever, because children and females are affected. Most areas of South America had effectively eliminated *Ae. aegypti* during eradication efforts (1930s–1960s), but the continent was reinvaded in the 1970s, and the vector is now widespread. The first outbreak of urban yellow fever since 1954 occurred in Santa Cruz, Bolivia, in 1997–1998. The wide distribution of *Ae. aegypti* (including areas juxtaposed to the yellow fever endemic zone) and the rise in air travel have increased the risk of reemergence of urban yellow fever in the Americas.

In Africa, the enzootic cycle involves tree-hole-breeding *Aedes* mosquitoes, which transmit virus between monkeys, from monkeys to humans, and between humans. In West Africa, cases appear during the middle of the rainy season (August) and peak during the early dry season (October), the period of maximum vector longevity. The principal enzootic vector is *Ae. africanus;* this species is responsible for transmission between monkeys in the forest zone and in gallery forests. During the rainy season, *Ae. africanus* and other vector species [*Ae. furcifer, Ae. taylori, Ae. luteocephalus, Ae. metallicus, Ae. vittatus,* and (in many parts of East Africa) *Ae. simpsoni*] reach high densities in the humid savanna zone (the "zone of emergence" of yellow fever). The prevalence of immunity in the human population accumulates rapidly with age, and children are at highest risk of infection and illness. The domestic vector *Ae. aegypti* breeds in receptacles widely used by humans for water storage where piped water supplies are unavailable. Where this mosquito is involved in virus transmission, the disease may occur in the dry season in both rural and urban areas. *Ae. aegypti* has been the principal vector in many recent human epidemics in West Africa. Because the density and biting rates of *Ae. aegypti* and sylvatic *Aedes* vectors are high, human infection rates during epidemics may reach 20–30%.

Rainfall and temperature influence vector abundance and transmission rates and the potential for yellow fever epidemics. Prolongation of the rainy season, reflected by vegetational indices in satellite images, predict a risk of yellow fever epidemics. Increasing temperature shortens the extrinsic incubation period of yellow fever virus in the mosquito, resulting in a significantly increased rate of transmission. Even brief exposure to high temperatures (e.g., in a sunlit forest clearing) can have this effect. Warm temperature also increases biting and reproductive rates of *Ae. aegypti.*

Ae. aegypti populations differ genetically in vector competence for yellow fever virus. High vector densities may be required for virus transmission by relatively insusceptible populations, such as occur in West Africa.

CONTROL STRATEGIES

The principal method for prevention and control is human vaccination. A safe and effective live, attenuated vaccine (yellow fever 17D vaccine) has been in wide use since World War II. The vaccine is incorporated into routine childhood vaccination programs in South America and some parts of Africa. The occurrence of human disease reflects incomplete coverage of the population, because vaccine immunity is probably lifelong after a single dose.

In the Americas, control of *Ae. aegypti* through larval source reduction was successful in reducing urban yellow fever in the early decades of the 20th century. The increasing size and complexity of urban areas, the proliferation of detritus and objects such as automobile tires that collect rainwater and breed mosquitoes, and increasing costs and competition with other health care priorities have limited the impact of mosquito control efforts. Emergency vector control using space sprays to kill infected female vectors could undoubtedly be used to contain urban yellow fever outbreaks, but the measure has never been evaluated specifically for this purpose. Spraying of large areas of forest and around villages by air- or ground-operated equipment to contain sylvatic yellow fever has been evaluated, with mixed success.

See Also the Following Articles

Medical Entomology • Mosquitoes • Zoonoses, Arthropod-Borne

Further Reading

Digoutte, J.-P., Cornet, M., Deubel, V., and Downs, W. G. (1995). Yellow fever. *In* "Exotic Virus Infections." (J. S. Porterfield, ed.), pp. 67–102. Chapman & Hall, London.

Miller, B. R., Monath, T. P., Tabachnick, W. J., *et al.* (1989). Epidemic yellow fever caused by an incompetent mosquito vector. *Trop. Med. Parasitol.* **40,** 396–399.

Monath, T. P. (1988). Yellow fever. *In* Monath, T. P. (ed.). "The Arboviruses: Ecology and Epidemiology," Vol. V (T. P. Monath, ed.), pp. 139–231. CRC Press, Boca Raton, FL.

Monath, T. P., Craven, R. B., Adjukiewicz, A., *et al.* (1980). Yellow fever in the Gambia, 1978–1979: epidemiologic aspects with observations on the occurrence of Orungo virus infections. *Am. J. Trop. Med. Hyg.* **29,** 912–928.

Robertson, S. E., Hull, B. P., Tomori, O., *et al.* (1996). Yellow fever. A decade of reemergence. Reprinted in *J. Am. Med. Assoc.* **276,** 11562.

Tabachnik, W. J., Wallis, G. P., Aitken, T. H. G., *et al.* (1985). Oral infection of *Aedes aegypti* with yellow fever virus: geographic variation and genetic considerations. *Am. J. Trop. Med. Hyg.* **34,** 1219–1227.

Trapido, H., and Galindo, P. (1957). Mosquitoes associated with sylvan yellow fever near Almirante, Panama. *Am. J. Trop. Med. Hyg.* **6,** 114–135.

Yellowjacket

see *Wasps*

Z

Zoonoses, Arthropod-Borne

Robert S. Lane
University of California, Berkeley

The term zoonosis (plural, zoonoses) is derived from the Greek roots for animal *(zoon)* and disease *(osis)* and was coined initially to characterize any infection of lower animals that could be transmitted to humans. The Food and Agriculture Organization and the World Health Organization (FAO/WHO) broadened the definition during the 1950s to include infections that are transmitted in either direction between humans and other vertebrate animals. Thus, zoonoses are those diseases and infections for which the agents are naturally transmitted between vertebrate animals and humans. Zoonoses can be transmitted to humans by direct contact with an infectious vertebrate host or a fomite (i.e., an inanimate object such as an article of clothing that can harbor an agent); by ingestion of contaminated water, food, or other organic matter; by inhalation; and by arthropod vectors. This article is devoted exclusively to zoonotic agents that are transmitted from other vertebrates to humans by arthropods. It does not include nonzoonotic disease agents that are transmitted solely from person-to-person by arthropods (e.g., malaria). The terms arthropod vector versus vector, and arthropod-borne versus vector-borne, are used interchangeably.

CHARACTERISTICS

Viruses, prions, and bacteria, which include the rickettsiae, and fungi, protozoa, and helminths can serve as zoonotic disease agents. Taylor and Woolhouse reported in 2000 that 832 (49%) of 1709 infectious organisms known to be pathogenic for humans are zoonotic in origin. These include 507 viruses and prions (30%), 541 bacteria (32%), 309 fungi (18%), 66 protozoa (4%), and 286 (17%) helminths. Furthermore, 156 of these agents produce diseases that are considered to be emerging, and 73% of the emerging pathogens are zoonotic.

Many zoonotic agents are transmitted to humans by means of arthropods, particularly insects or ticks. Indeed, two of the three contemporary internationally quarantinable diseases are vector-borne zoonoses transmitted either by mosquitoes (yellow fever) or fleas (plague). Plague also is transmitted by other routes, but the bacterial agent *(Yersinia pestis)* is maintained primarily in natural cycles involving rodents and their associated fleas. In the tropics and subtropics, the most important arthropod-borne zoonotic disease agents are transmitted by flies (Diptera) such as mosquitoes, phlebotomine sand flies, and tsetse flies, whereas in temperate regions ticks transmit over 90% of the vector-borne human pathogens.

Humans are "dead-end" hosts for most zoonotic agents, i.e., secondary cases do not occur following human infection. Notable exceptions are yellow fever, plague, and Chagas disease. Humans infected with yellow fever virus can serve as a source of infection for uninfected mosquitoes that feed on them, individuals manifesting primary or secondary pneumonic plague potentially can transmit *Y. pestis* to other persons by the respiratory route, and people infected with the trypanosome causing Chagas disease can infect triatomine bugs that bite them.

Several factors that affect the ability of zoonotic agents to produce disease in people are the route of transmission (portal of entry), the genetically determined invasiveness (pathogenicity) of the infecting organism, and the age and immune status of an exposed individual. Two North American mosquito-borne viral encephalitides tend to produce more severe disease in disparate age groups, i.e., children infected with western equine encephalomyelitis (WEE) virus experience graver illness than adults, whereas the reverse is true for individuals infected with St. Louis encephalitis virus. Zoonoses may cause comparable harm to humans and some lower animals, or they may adversely affect one group or the other in a dispropor-

tionate manner. Thus, certain of the mosquito-borne encephalitic viruses (e.g., eastern equine encephalitis, venezuelan equine encephalitis, WEE) can be highly pathogenic for both humans and horses.

Most arthropod-borne zoonotic infections can be characterized by their focal distribution within particular geographic landscapes. A natural focus, also known as a nidus, is an area where the complex interplay of various biotic and abiotic environmental factors ensures the temporal persistence of a zoonotic agent. This concept, referred to as "landscape epidemiology," was developed into a doctrine by the Russian scientist E. N. Pavlovsky in 1939. Accordingly, foci exist under certain conditions of macro- and microclimate, vegetation, and soil in localities in which suitable arthropod vectors and reservoir hosts of the disease agent also exist.

Furthermore, foci can be as distinct and limited geographically as the burrow system of a small mammal or a tick-infested cottage, or they can be of the diffuse variety in which case they encompass portions of much broader territories such as coniferous or mixed hardwood forests. For instance, individual cases or localized outbreaks of relapsing fever in California usually are associated with rodent- and tick-infested cabins at higher elevations. In such foci, the spirochete *Borrelia hermsii* occasionally is transmitted to sleeping individuals by the bite of the rapid-feeding, nocturnally active soft-tick vector, *Ornithodoros hermsi.* In contrast, northern Californian foci of the Lyme disease spirochete *Borrelia burgdorferi,* although present in several major habitat types, seem to be most intense in certain leaf-litter areas within mixed hardwood forests where infection prevalences in nymphal *Ixodes pacificus* ticks sometimes reach 20 to 40%.

VECTORS

The phylum Arthropoda contains six classes of varying medical importance but only two of them, the Insecta and Arachnida, are of paramount importance as transmitters of zoonotic agents. Mites, ticks, spiders, and scorpions comprise the Arachnida, but only certain species of mites and particularly ticks are capable of transmitting infectious organisms to humans.

Insects that transmit zoonotic agents belong principally to three orders, namely, the Siphonaptera (fleas), Heteroptera (true bugs), and especially the Diptera (flies). Important families of flies that transmit zoonotic agents are the Culicidae (mosquitoes), Glossinidae (tsetse flies), Psychodidae (phlebotomine sand flies), Ceratopogonidae (biting midges), and Tabanidae (deer flies and horse flies). Other dipteran families, like those comprising the filth flies generally and the house fly in particular (Muscidae), may mechanically contaminate human foodstuffs with bacteria, protozoans, or other agents that can cause gastrointestinal illnesses.

Among the arachnids, ticks overwhelmingly outweigh the mites in importance as vectors of zoonotic agents because of their universal blood-sucking habit. With rare exceptions, all three trophic stages (larva, nymph, adult) of ticks must ingest a blood meal to complete development and to ensure reproductive success. By comparison, the majority of mites are free living and nonparasitic. In the United States, approximately 10% of the 83 described tick species transmit one or more viruses, bacteria, or protozoan parasites to people with any regularity. One commonality shared by many arachnid and insectan vectors is that they feed intermittently upon their hosts and therefore spend more than about 95% of their entire life cycle off the host.

CLASSIFICATION

Zoonoses were classified by the FAO/WHO in 1967 according to whether their primary reservoir hosts are humans or lower animals and with regard to the type of life cycle of the infecting organism. The latter classification, which is based upon shared epidemiologic features, is more instructive than one based solely on reservoir hosts. In it, zoonoses have been categorized as direct zoonoses, cyclozoonoses, metazoonoses, or saprozoonoses. However, only two of these categories involve transmission by arthropods. Direct zoonoses are transmitted, in part, from an infected to a susceptible vertebrate host mechanically by a vector, whereas metazoonoses are transmitted biologically by vectors.

TRANSMISSION CYCLES

Arthropod-borne zoonotic agents are maintained in transmission cycles of variable complexity. Nevertheless, four components are evident in all such cycles: the agent itself, one or more efficient arthropod vectors and primary reservoir hosts, and a permissive environment. A reservoir host is a vertebrate that is readily infected with the agent, is capable of maintaining the agent within its tissues for an extended period, and can serve as a source of infection for uninfected vectors that feed on it while it is in an infectious state. In reality, arthropods also contribute to the maintenance of zoonotic agents to various degrees; some, like certain hard-tick vectors of spotted fever group rickettsiae or mosquito vectors of some viruses, may be considered reservoirs as well. That is, they can maintain the agent in their tissues for months or even years, efficiently pass it from one trophic stage to the next (transstadial transmission), and eventually transmit it from one generation to the next via the eggs of infected females (transovarial transmission). In such cases, the reservoir of infection may best be considered polyhostal because both the vertebrate host(s) and the vector(s) help to maintain the cycle of infection.

Arthropods may transmit zoonotic agents from an infected vertebrate to a susceptible one either mechanically or biologically. Of these, biological transmission is much more prevalent than mechanical transmission among most groups of arthropod vectors. Mechanical transmission occurs when an agent adheres externally on the mouthparts, legs, or other bodily regions of a vector and then is transported directly or indirectly to a vertebrate, e.g., by means of contaminated foodstuffs or

by specific inoculation into the skin or bloodstream by the bite of contaminated mouthparts. In mechanical transmission, the agent does not require the arthropod to complete an essential part of its life cycle, and the arthropod/agent relationship is of an accidental nature.

In biological transmission, the vector plays an indispensable role in the life cycle of the agent. Biological transmission takes three forms: cyclodevelopmental, cyclopropagative, and propagative. In cyclodevelopmental transmission, the agent undergoes cyclical changes within the internal tissues of the vector but does not multiply (e.g., mosquito transmission of the filariid nematode *Wuchereria bancrofti*). In cyclopropagative transmission, the agent undergoes cyclical development and multiplication in the arthropod's body (e.g., malaria plasmodia in anopheline mosquitoes or babesial piroplasms in their tick vectors). In propagative transmission, the agent multiplies but undergoes no cyclical development within the vector's body (e.g., most bacteria and viruses).

Regardless of the specific mode of biological transmission, the vector usually transmits the agent anteriorally via its salivary secretions while ingesting a blood meal from a vertebrate host. In a few instances, the agents are transmitted posteriorly when infectious feces are deposited on the host near the bite wound (e.g., feces of triatomine bugs infected with *Trypanosoma cruzi,* the causative agent of Chagas disease).

PUBLIC HEALTH IMPORTANCE

The public health and economic impact of the arthropod-borne zoonoses is immense, particularly in developing countries in tropical or subtropical regions. Losses include morbidity and mortality among humans and livestock and the resultant direct and indirect economic effects to affected individuals and to society at large. For instance, the African trypanosomiases still are among the most devastating diseases afflicting humans and livestock. African sleeping sickness is estimated to have killed up to 500,000 people between 1896 and 1906, and one epidemic near Lake Victoria in Uganda claimed about 200,000 lives. Today, it is estimated that over 300,000 new cases of the disease are contracted annually and that 60 million people in 36 countries of sub-Saharan Africa and an even greater number of livestock are at risk. In addition to African trypanosomiasis, other zoonotic and nonzoonotic vector-borne diseases such as malaria, dengue, yellow fever, filariasis, leishmaniasis, plague, and louse-borne typhus caused more human morbidity and mortality than all other causes from the 17th to the early 20th centuries, especially in the tropics. Since the 1970s, there has been a resurgence of many of these diseases and the emergence of others, notably tick-borne diseases like Lyme disease, ehrlichiosis, and babesiosis in temperate regions.

Certain population groups traditionally have been at elevated risk of acquiring vector-borne zoonotic infections. Agricultural and forestry workers, hunters, field naturalists, park rangers, wildlife biologists, woodcutters, and tourists represent just a few of the many groups that are vulnerable to such infections. In the northeastern United States, suburbanites may be at considerable risk for contracting Lyme disease in the peridomestic environment because of the presence of host-seeking, spirochete-laden *Ixodes scapularis* ticks on lawns and in adjacent shrubbery or forested areas. Similarly, in the western United States, the risk of human plague increased in peridomestic settings during the late 20th century as human populations encroached into formerly unpopulated foci of *Y. pestis.*

Apart from the human misery and loss of life attributable to zoonoses, the significant costs associated with the occurrence of individual cases or outbreaks of a few specific arthropod-borne diseases in the United States illustrate their economic impact. Following a single human plague case at Plumas–Eureka State Park in Plumas County, California, in 1976, attendance at that park declined markedly for the remainder of the year. Moreover, a benefit–cost analysis of bubonic plague surveillance and control at that park and another nearby campground revealed that the costs incurred when human plague was contracted at either recreational area averaged $52,000. In 1966, an outbreak of St. Louis encephalitis virus, a mosquito-borne flavivirus, in Dallas, Texas, was estimated to cost nearly $800,000 in terms of morbidity, mortality, patient treatment, and control activities. In total, 172 presumptive or serologically confirmed cases with 20 fatalities were reported during that outbreak. In Massachuetts, the average total cost per case of eastern encephalitis, a mosquito-borne alphavirus, was determined to be $21,000 for a transient case versus a lifetime cost of nearly $3 million for persons suffering severe residual sequelae.

Finally, in contrast to the tropics where dipterans (flies) rank first as transmitters of vector-borne disease agents, ticks are unsurpassed in importance as primary vectors in temperate regions. Lyme disease alone currently accounts for over 95% of all cases of vector-borne diseases reported annually in the United States. In 1999, Lyme disease and Rocky Mountain spotted fever composed 99.5% of the total number of indigenously acquired vector-borne disease cases reported that year; mosquito- and flea-borne infections accounted for the remaining 0.5%. These computations for 1999 exclude numerous cases (probably a few hundred) of several tick-borne diseases (e.g., Colorado tick fever, human granulocytic and monocytic ehrlichioses, and relapsing fever) and 62 cases of the newly introduced mosquito-borne West Nile virus because none of these zoonoses is a nationally notifiable disease. Likewise, tick-borne diseases predominate throughout much of Europe and in parts of Asia, especially in the former Soviet Union.

In conclusion, zoonoses comprise about one-half of all infectious organisms known to cause disease in humans. Knowledge of the specific environmental factors and vectors that maintain and distribute arthropod-borne zoonotic agents is essential before locally effective control methods can be developed and implemented. This information is especially valuable because relatively few vaccines are available for personal protection against vector-borne infections.

See Also the Following Articles

Bubonic Plague • Dengue • Medical Entomology • Mosquitoes • Ticks • Tsetse Fly • Veterinary Entomology • Yellow Fever

Further Reading

Acha, P. N., and Szyfres, B. (1987). "Zoonoses and Communicable Diseases Common to Man and Animals," 2nd ed. Pan Am. Health Org., Washington, DC. [Scientific Pub. No. 503]

Centers for Disease Control and Prevention (1999). Recommendations for the use of Lyme disease vaccine: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Suppl.* **48(RR-7),** 1–25.

Centers for Disease Control and Prevention (2001). Summary of notifiable diseases, United States, 1999. *MMWR* **48(53),** 1–101.

Gubler, D. J. (1998). Resurgent vector-borne diseases as a global health problem. *Emerg. Infect. Dis.* **4,** 442–450.

Harwood, R. F., and James, M. T. (1979). "Entomology in Human and Animal Health," 7th ed. Macmillan Co., New York.

Hugh-Jones, M. E., Hubbert, W. T., and Hagstad, H. V. (eds.) (2000). "Zoonoses: Recognition, Control, and Prevention." Iowa State University Press, Ames.

Kimsey, S. W., Carpenter, T. E., Pappaioanou, M., and Lusk, E. (1985). Benefit–cost analysis of bubonic plague surveillance and control at two campgrounds in California, USA. *J. Med. Entomol.* **22,** 499–506.

Lawyer, P. G., and Perkins, P. V. (2000). Leishmaniasis and trypanosomiasis. *In* "Medical Entomology: A Textbook on Public Health and Veterinary Problems Caused by Arthropods" (B. F. Eldridge and J. D. Edman, eds.), Chap. 8. Kluwer Academic, Dordrecht.

Pavlovsky, E. N. (1966). "Natural Nidality of Transmissible Diseases with Special Reference to the Landscape Epidemiology of Zooanthroponoses." University of Illinois Press, Urbana.

Reeves, W. C. (1990). Epidemiology and control of mosquito-borne arboviruses in California, 1943–1987. Calif. Mosquito Vector Control Assoc., Sacramento.

Schwab, P. M. (1968). Economic cost of St. Louis encephalitis epidemic in Dallas, Texas, 1966. *Public Health Rep.* **83,** 860–866.

Tälleklint-Eisen, L., and Lane, R. S. (1999). Variation in the density of questing *Ixodes pacificus* (Acari: Ixodidae) nymphs infected with *Borrelia burgdorferi* at different spatial scales in California. *J. Parasitol.* **85,** 824–831.

Taylor, L. H., and Woolhouse, M. E. J. (2000). Zoonoses and the risk of disease emergence. *In* "International Conference on Emerging Infectious Diseases, Atlanta, Georgia, Program and Abstracts Book," late breaker poster session, Board 122, p. 14.

Villari, P., Spielman, A., Komar, N., McDowell, M., and Timperi, R. J. (1995). The economic burden imposed by a residual case of eastern encephalitis. *Am. J. Trop. Med. Hyg.* **52,** 8–13.

World Health Organization (1967). Third Report of a Joint WHO/FAO Expert Committee on Zoonoses. *WHO Tech. Rep. Ser.* **378.**

Zoraptera

Michael S. Engel
University of Kansas

The Zoraptera ("zorapterans," "angel insects," or less appropriately "soil lice") are one of the smallest insect orders with only 32 living species and 6 more known from fossils. Zorapterans are rarely observed; indeed many entomologists have never seen them alive. There is a single family (Zorotypidae) and only two genera, *Zorotypus* and the fossil genus †*Xenozorotypus* from the Middle/Late Cretaceous. Some authors have at times attempted to divide living species of *Zorotypus* into multiple genera or subgenera, but none of these have successfully characterized natural (i.e., monophyletic) groups. Owing to the paucity of species and the great morphological homogeneity of the order, the recognition of numerous living genera is currently unwarranted.

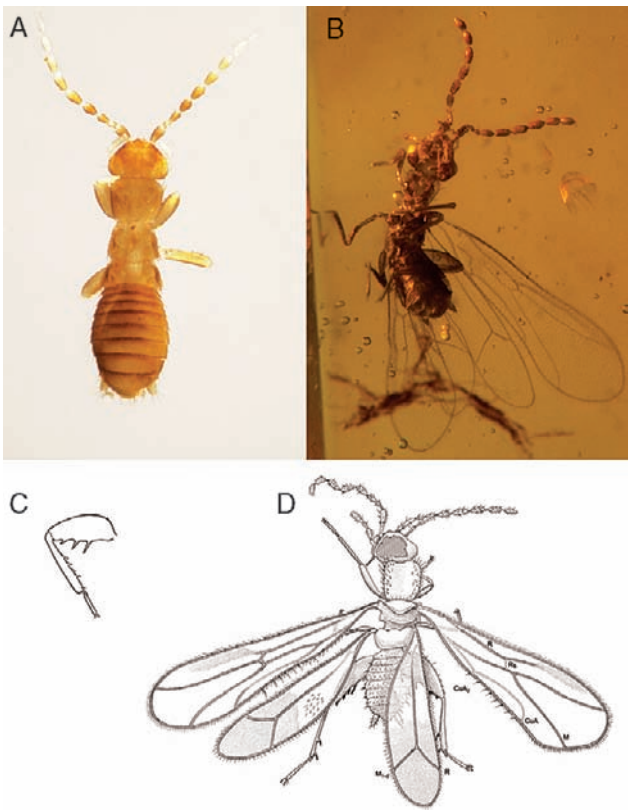

FIGURE 1 (A) Adult female of *Zorotypus huxleyi* (most legs curled under body), blind–apterous morph. (B) Alate female of †*Z. goeleti* in Miocene amber from the Dominican Republic. (C) Hind leg of *Z. huxleyi* depicting femoral and tibial spines. (D) Alate of †*Z. nascimbenei* in Cretaceous amber from Myanmar, with wing veins labeled. (Reproduced, courtesy of the American Museum of Natural History, from Engel and Grimaldi, 2002.)

Zorapterans superficially resemble termites (Isoptera) and bark lice (Psocoptera) (Figs. 1A, 1B, and 1D) and they have at times been classified near or within these groups. Individuals are typically less than 4 mm in body length (excluding their antennae) and live in small colonies. The most readily noticeable traits of the order are the expanded hind femora that bear stout spines on their ventral surface (Fig. 1C), the presence of only two tarsomeres (the basalmost of which is greatly reduced in size; Fig. 1C), and paddle-shaped wings with reduced venation (Figs. 1B and 1D) that are shed by an ill-defined basal fracture.

† Extinct group.

ORIGIN AND EVOLUTION

Although zorapterans are relatively minute both in size and in species diversity, they have weighed heavily on the minds of evolutionary biologists. The order has been at one time or another classified in a bewildering array of phylogenetic positions, at times being considered relatives of the termites, roaches (Blattaria), mantises (Mantodea), webspinners (Embiidina), or bark lice; sister to Paraneoptera; or even allied to the Holometabola. The most well supported hypothesis is that they are polyneopterans and the living sister to the webspinners, both orders sharing a unique development of the musculature in the hind legs, among several other important traits. The Zoraptera–Embiidina likely occurred during the early Mesozoic, perhaps as long ago as the early Triassic, while the ancestor of both groups diverged from stoneflies (Plecoptera) perhaps sometime in the Carboniferous. The Zoraptera are clearly ancient as evidenced by the presence of extinct species of the living genus *Zorotypus* in Middle/Late Cretaceous amber (e.g., Fig. 1D). These extinct Mesozoic species already show the development of the dual-morphs, with blind–apterous and eyed–winged individuals. Only the extinct genus †*Xenozorotypus* shows a departure in morphology (albeit not nearly as dramatically as the anatomical diversity found in other orders) from all other zorapterans by the primitive retention of additional wing veins and the structure of spines on the hind tibia.

DISTRIBUTION

Species principally occur in tropical habitats throughout the world although four species occur north of the Tropic of Cancer (i.e., north of 23.5°N): two in North America and two in Tibet. Similarly, the known fossil zorapterans have all been discovered in ambers that were formed in warm-tropical paleoclimates. The northernmost areas of distribution for *Z. hubbardi* in North America are likely a result of human activity. Most northerly colonies are formed in sawdust piles (rather than natural logs) where warmth from the decaying material can sustain a colony for a few seasons. Thus, once a pile is no longer suitable dispersal to new sawdust piles or into natural logs is required or, conversely, those northernmost populations repeatedly become extinct and are maintained over the long term by perpetual reintroductions from more southern populations. Once believed to be highly endemic as a result of an inability to disperse, several *Zorotypus* species are increasingly being recognized over larger geographic ranges (e.g., *Z. hubbardi* in south-central and southeastern United States), evidence of an ability to disperse to some degree so as to maintain specific integrity.

GENERAL BIOLOGY

The biology of zorapterans is poorly documented. Only three species have been studied and only one, *Z. hubbardi,* generated a sizeable literature. Species are gregarious with colonies consisting of 15 to 120 individuals in relatively ephemeral, subcortical habitats. Colonies occur in logs (or man-made sawdust piles) but only after wood decay is well progressed. Zorapterans feed principally on fungal hyphae and spores, although they can also be predatory, victimizing nematodes, mites, and other minute arthropods such as springtails (Collembola). The number of nymphal instars is estimated to be either four or five. Captive individuals live up to 110 days.

Individuals occur in two morphs. Blind, wingless individuals predominate during the general life of a colony, whereas dispersive individuals (with eyes and wings) are generally rare and are produced as resources become depleted and the colony becomes crowded. Winged females are more common than winged males, suggesting that females perhaps mate prior to dispersal. After arriving at a new log, individuals shed their wings and dealated individuals are often found in young colonies. Wide distributions for some species as well as the presence of species endemic to distantly isolated islands such as Christmas Island, Fiji, and Hawaii are evidence for their dispersal abilities.

BEHAVIOR

Although *Zorotypus* species are gregarious and live in small colonies, they evidently do not distinguish between individuals from the same colony and introduced vagrants. Isolated individuals do not survive.

Zorapterans spend much of their time grooming either themselves or other individuals. Grooming may be a way of removing fungal pathogens and this may be important for their colonial lifestyle. Most of the movements in the grooming repertoire are found in other insect orders but some are unique to zorapterans. Most notable are the movements associated with the cleaning of the posterior sterna, the cerci, and the genital–anal area using their mouthparts. This complex action involves the raising of the body on a four-point stance, with both anterior and posterior ends of the body bent downward (perpendicular with the substrate) to meet under the insect and between the middle and the hind legs.

Females mate multiple times (with up to eight mates) and with multiple males (up to three multiple matings). Prior to copulation, zorapteran species have complex courtship behaviors which appear to be species specific. The process is begun by the male who strokes the female with his antennae. If the female reciprocates with similar antennal signals, the male initiates courtship displays during which his head and a portion of his neck are extended. If the female approaches, the male secretes fluids from a cephalic gland, which she ingests while the male and female bend their abdomens toward one another. Copulation is initiated by the linking of the male to the female via a hook located medially on his 10th abdominal tergum (females have a small groove into which the hook is placed on their eighth sternum). Once the hook is coupled with the female, the male inverts himself, assuming an upside-down position and facing away from the female.

ECONOMIC IMPORTANCE

Although zorapteran colonies at times can be found in wood or sawdust piles at lumber yards, owing to their relatively minute colony sizes, general scarcity, and apparent minimal ecological impact, it is unlikely that they will ever be considered of any economic importance. If the fragmentary reports of wood ingestion by zorapterans are confirmed as well as studies of their gut fauna (bacteria, protozoa, etc.) undertaken, then the economic role of *Zorotypus* might be reevaluated.

See Also the Following Articles

Isoptera • Psocoptera

Further Reading

Choe, J. C. (1992). Zoraptera of Panama, with a review of the morphology, systematics, and biology of the order. *In* "Insects of Panama and Mesoamerica: Selected Studies" (D. Quintero and A. Aiello, eds.). Oxford University Press, Oxford.

Choe, J. C. (1995). Courtship feeding and repeated mating in *Zorotypus barberi* (Insecta: Zoraptera). *Anim. Behav.* **49,** 1511–1520.

Choe, J. C. (1997). The evolution of mating systems in the Zoraptera: Mating variations and sexual conflicts. *In* "The Evolution of Mating Systems in Insects and Arachnids" (J. C. Choe and B. J. Crespi, eds.). Cambridge University Press, Cambridge, UK.

Engel, M. S. (2000). A new *Zorotypus* from Peru, with notes on related neotropical species (Zoraptera: Zorotypidae). *J. Kansas Entomol. Soc.* **73,** 11–20.

Engel, M. S., and Grimaldi, D. A. (2000). A winged *Zorotypus* in Miocene amber from the Dominican Republic (Zoraptera: Zorotypidae), with discussion on relationships of and within the order. *Acta Geol. Hisp.* **35,** 149–164.

Engel, M. S., and Grimaldi, D. A. (2002). The first Mesozoic Zoraptera (Insecta). *Am. Mus. Novitates* **3362,** 1–20.

Gurney, A. B. (1938). A synopsis of the order Zoraptera, with notes on the biology of *Zorotypus hubbardi* Caudell. *Proc. Entomol. Soc. Wash.* **40,** 57–87.

Silvestri, F. (1913). Descrizione di un nuovo ordine di insetti. *Bol. Lab. Zool. Gen. Agric. Portici* **7,** 193–209.

Zygentoma
(Thysanura, Silverfish)

Helmut Sturm

University of Hildesheim, Germany

The Zygentoma (Thysanura *sensu stricto*) are medium-sized apterygote insects with a body length from 5 to 30 mm. Their body is flattened and the eyes are small or absent. There are no ocelli except for the Lepidothrichidae, which has three ocelli. The flagellate (whiplike) antennae can be short (e.g., in Ateluridae) or much longer than the body (Fig. 1). The mouthparts are ectognathous. The mandibles are dicondylic (have two rotation points) and the maxillary palps have five segments that are of normal length. Nearly all parts of the body and appendices bear bristles of different length and structure. Scales are present on the thoracic segments of Lepismatidae and Ateluridae, but lacking in the three other families.

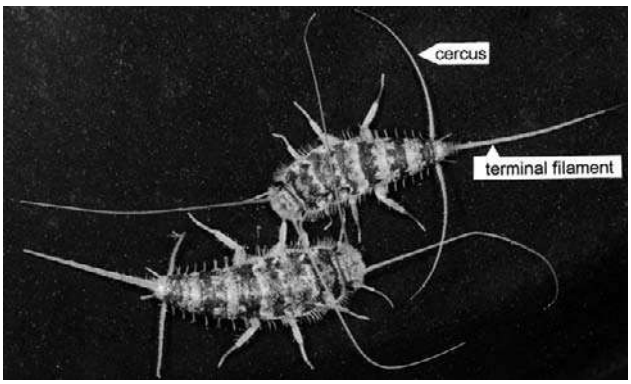

FIGURE 1 Male (above) and female of *Thermobia domestica* during premating (passing-by behavior), vertical view. In this phase the antennae and the cerci are in contact; body length about 12 mm.

All Zygentoma are able to run quickly, but they cannot jump as do the Archaeognatha. The coxae of all legs are large and flattened. The penis of males is situated on abdominal segment 9; on the hind end, three caudal appendages are present as in the Archaeognatha. Molting continues in adults. Some species of the family Lepismatidae (e.g., *Lepisma saccharina, Thermobia domestica, Acrotelsa collaris*) prefer to live in human houses.

FOSSIL RECORD, SYSTEMATICS, BIOGEOGRAPHY

The oldest fossils that are clearly Zygentoma were found in Mesozoic layers (Lower Cretaceous) of Brazil. From the Cenozoic on, there are many Zygentoma fossils, and most of them are amber inclusions. Nearly all of them are similar to extant taxa (e.g., *Ctenolepisma*, *Nicoletia*, Lepidothrichidae). However there are representatives of apterygote Hexapoda from the Paleozoic (e.g., Monura and Cercopodata) that cannot be included in extant orders.

The Zygentoma comprise five families and some 400 species in about 90 genera.

1. **Lepismatidae** This largest family includes the well-known silverfishes, some of which occupy human houses. This family is the richest in genera (more than 21) and species (> 200).
2. **Nicoletiidae** Relatively small and slim, and without dark pigment, eyes, and scales, they live in the upper layers of soils within the humus layer or beneath stones. Some species or populations reproduce parthenogenetically. More than 10 genera and more than 30 species are known.
3. **Ateluridae** Small, blind, and of droplike form, apparently all species in this family are commensal with ants or termites. There are more than 10 genera and more than 30 species.
4. **Lepidothrichidae** There are two species: *Lepidothrix pilifera,* which is known from Baltic amber, and the extant

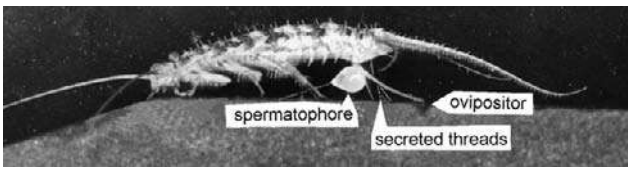

FIGURE 2 Female of *T. domestica,* dorsal view. The female is picking up the spermatophore with her ovipositor. The spermatophore was deposited on secreted threads spun, a short time before, by the male.

Tricholepidion gertschi, which occurs in forests of northern California (Fig. 3).

5. **Maindroniidae** There is one genus and three species from the Middle East and Chile; their habitat is somewhat similar to that of the Nicoletiidae.

The Lepismatidae and Nicoletiidae are distributed worldwide. Some Lepismatidae are household pests: *L. saccharina* in northern Europe; *A. collaris* in the tropics; and *T. domestica* in the United States (Figs. 1 and 2). In southern Europe, *L. saccharina* and other lepismatid species also live outside houses. In Africa, Lepismatidae occur in areas with sand dunes. Nicoletiidae are found up to 3000 m elevation in Colombia and enter the northern parts of Europe via transport of potted plants to greenhouses. The distribution of Ateluridae coincides fairly well with that of ants and termites. The distribution of Lepidothrichidae and Maindroniidae appears quite restricted.

FIGURE 3 Male of *T. gertschi* (Lepidothrichidae), vertical view, body length about 14 mm. This species is probably a relic and the most archaic within the Zygentoma. It is found only in the forests of Northern California. The animal is seen here on rotting wood, its natural environment.

Their relatively flat body form enables the Zygentoma to enter small fissures. The Nicoletiidae are especially well adapted to the upper layers of the soil, whereas Lepismatidae generally rest in fissures during the day and are active during the night. If the relative humidity is above 50 to 60%, lepismatidae can absorb atmospheric water by means of the anal sac (part of the hindgut).

PHYSIOLOGY AND DEVELOPMENT

As in Archaeognatha and in other orders of apterygote insects, the cuticle of Zygentoma is mainly thin and flexible. The mandibles present an intermediate form between monocondylic and a typical dicondylic structure. The Lepismatidae eat plant material and detritus; *L. saccharina* is able to digest cellulose by the cellulase produced in its midgut.

Development has been studied mainly on Lepismatidae (especially *L. saccharina, T. domestica,* and *Ctenolepisma* spp.). The eggs of *T. domestica* are of oval form with a largest diameter of about 1.2 mm, and are laid underground. At a temperature of 37°C, they need 12 to 16 days until eclosion. In *L. saccharina,* egg development lasts about 35 days at temperatures near 20°C. The first free-living developmental stage possesses an egg tooth on the front of the head, and the first three stages lack scales. The molting intervals are dependent on age, temperature, loss of scales and/or appendages, species, and even population. At temperatures of about 40°C *T. domestica* molts in 9- to 11-day intervals. The life span of *L. saccharina* is up to 3 years, whereas that of *Thermobia* is about 2 years.

EVOLUTIONARY ASPECTS

With flattened form, the Zygentoma are adapted to small fissures in soil, detritus, or in houses. The Lepismatidae leave those places during night to find food, favorable conditions (e.g., high relative humidity), and members of the same species for mating. In contrast with the Archaeognatha, their food use is broad. Lepismatidae have been spread widely by humans. For example, of the 13 lepismatid species in the United States, only three are believed to be native. In comparison to the Archaeognatha, the Zygentoma show a great adaptability in terms of the different habitats they live in and the astonishing adaptation of the Ateluridae, which live with ants and termites. They possess some morphological structures that correspond to the Pterygota: a special form of mandibles; maxillary palps smaller than thoracic legs; an abdomen with tracheal commisures and connections; and an ovipositor base with well-developed gonangulum. Apparently, however, they are not the ancestors of the Pterygota. Because of the crosswise joining of abdominal terga with correspondent coxites (a unique apomorphy within the Hexapoda), the possession of 11 abdominal segments, styli on abdominal segments, eversible vesicles,

three abdominal appendages (filum terminale and two cerci), and the structure of the ovipositor, the Archaeognatha and Zygentoma are obviously linked. However the "jump" from Apterygota to Pterygota probably was from a line separate from the Thysanura.

See Also the Following Article

Archaeognatha

Further Reading

Smith, G. B. (1995). Thysanura (Zygentoma). *In* "CSIRO, Zoological Catalog of Australia," Vol. 23, pp. 6–18. Commonwealth Scientific and Industrial Research Organization, Canberra.

Sturm, H. (1997). The mating behavior of *Tricholepidion gertschi* Wygodzinsky, 1961 (Lepidotrichidae, Zygentoma) and its comparison with the behaviour of other "Apterygota." *Pedobiologia* **41,** 44–49.

Wygodzinsky, P. W. (1972). A review of the silverfish (Lepismatidae, Thysanura) of the United States and the Caribbean area. *Am. Mus. Novit.* **2481,** 1–26.

This Page Intentionally Left Blank

GLOSSARY

The following glossary defines approximately 800 specialized terms that are found in the articles in the *Encyclopedia of Insects.* The authors of the Encyclopedia have defined these words as they are used in their articles, for the benefit of readers who may not be familiar with the technical vocabulary of insect study. As the name indicates, the Glossary is a compendium of the terminology of this particular encyclopedia. It is not a general dictionary and therefore does not include all possible entomological terms that a reader might encounter in other publications.

A

accessory gland Any secondary gland of a glandular system.

accessory pulsatile organs (APOs) Very small muscular pumps and veins that direct hemolymph into appendages such as wings and antennae.

acclimation A process of adjustment to survival at temperatures that are normally lethal.

acetylcholinesterase (AChE) The enzyme responsible for breaking down the neurotransmitter **acetylcholine (ACh)** at nerve synapses, thereby preventing hyperexcitation of cholinergic pathways in the nervous system. This enzyme is the major target of organophosphate and carbamate insecticides.

acrosome The anterior portion of the differentiated spermatozoa that assists in the penetration of the egg membrane.

across-fiber pattern code An often complex code in which each of several cell types sends a message about a single stimulus to higher processing centers. The stimulus may contain one or more than one type of chemical.

aculeate Any member of a group of families that includes the familiar stinging ants, bees, and social and hunting wasps (e.g., sphecids, eumenids, scoliids, pompilid spider wasps). Aculeates are largely characterized by their stinging apparatus, which is derived from the ovipositor of their ancestors but is no longer used for laying eggs.

acylurea A class of insect growth regulators, including diflubenzuron.

adaptive immunity A host defense mechanism that is specific for an epitope; it involves B and T cells and exhibits memory.

adaptive radiation Diversification of species within a lineage associated with ecological change.

adipocytes (trophocytes) The major cell type of the insect fat body that accumulates and stores reserve nutrients such as lipids, glycogen, and proteins used primarily during molting, metamorphosis, flight, and oocyte formation. These cells are actively involved in intermediary metabolism and play a crucial role in regulating blood proteins (enzymes), lipids, and carbohydrates. A fundamental function of adult adipocytes is the synthesis and release of vitellogenins (yolk proteins) taken up from the hemolymph by developing oocytes.

adipokinetic hormone (ADK) A hormone involved in promoting fat metabolism.

aedeagus The sclerotized terminal portion of the male genital tract that is inserted into the female during insemination.

aerobic Using molecular oxygen as an electron acceptor; pertains to organisms or metabolic processes.

agent An organism that can infect a larger organism in which it may cause a specific disease.

alary muscles Muscles of the dorsal diaphragm that sometimes participate in hemolymph movement in and around the pericardial sinus of the abdomen.

alata The parthenogenetic winged morph of viviparae, specialized for migration; among dioecious aphids, alatae termed emigrants fly from the primary to the secondary host each spring.

alate **1.** Winged; having wings. **2.** An adult individual with fully developed wings.

allele Form of a gene. Genes are considered **allelic** when they occur in the same position on members of a chromosome pair.

allelochemical A nonnutritional chemical produced by one species that affects the growth, health, behavior, or reproduction of another species, e.g., phytochemicals that function in plants as defenses against herbivores or pathogens.

allergen An antigen that induces an allergic reaction.

allometric growth A pattern of growth in which a body part grows at a rate different from that of other body parts or the body as a whole. **Allometry** can be measured within an individual over the course of its development, and also between individuals. **Isometry** is a special case of allometry in which the relative sizes of two or more structures are constant within or among individuals.

allomone An allelochemical that benefits the producing organism and harms the receiving organism, such as chemicals produced by a stink bug to protect it against natural enemies.

alloparent An individual that assists the parents in care of the young.

allozyme A particular variant of an enzyme.

alternative splicing The ability to create two or more types of proteins from the same (eukaryotic) gene, by selective inclusion or exclusion of exons from mature messenger RNA, whereby regulatory interactions between genes result in great diversity of proteins from the same group of genes and alternative developmental pathways.

altruism An organism's behavior that does not directly benefit itself but contributes to the welfare of others within the family group. Associated with kin selection in eusocial insects.

ametabolous Characterized by no significant metamorphosis. The immature body form does not differ from the adult, except for external genitalia, and the number of stages is variable.

anaerobic Not using molecular oxygen as an electron acceptor; pertains to organisms or metabolic processes.

anaerobic flagellates The heterogeneous assembly of phylogenetically ancient protozoa (archaezoa) that is found exclusively in the hindguts of termites and wood-feeding cockroaches.

anchialine Relating to surface and underground habitats associated with coastal ponds that have only underground connections to the sea.

anemotaxis Directed orientation with respect to wind, usually involving walking or flying in an upwind direction.

anholocycly The asexual reproduction of parthenogenetic viviparous aphids, wherein oviparae, males, eggs, and fundatrices are omitted in a nonrecombinant lifestyle.

anisogamy A condition in which the female sex cells (ova) are larger than the male sex cells (sperm).

Anopheles A genus of mosquitoes (order Diptera, family Culicidae) that contains 422 species, of which 40 frequently may be involved in the transmission of human malaria.

anopheline mosquito A mosquito classified in the family Culicidae, subfamily Anophelinae. The important genus *Anopheles* is in this subfamily.

anoxic Devoid of molecular oxygen; describes the oxygen status of an environment.

antifeedant A chemical, often toxic, that prevents or reduces feeding.

ant-tending The activity of ants resulting in the consumption of honeydew produced by the tended insects. Also, **ant attendance**.

aphidophagous Feeding on aphids.

apitherapy Medicinal use of the honey bee or its products.

Apocrita The Hymenoptera with a wasp waist (i.e., the sister group to the Orussidae), which comprises the majority of species in the order.

apolysis Separation of new and old cuticle caused by release of ecdysteroids at the beginning of each molt.

apomixis Parthenogenesis in which eggs are produced without meiosis.

apomorphous Describing a novel character evolved from a preexisting character.

apomorphy (-ic) A feature of an organism in the derived state, contrasted with an alternative one in the ancestral (primitive) state—a plesiomorphy (-ic). For example, with the character of forewing development, the sclerotized elytron is an apomorphy for Coleoptera, and the alternative, a conventional flying forewing, is a plesiomorphy at this level of comparison.

aposematic signals Conspicuous visual, acoustic, olfactory, or behavioral warnings displayed by an animal to advertise its noxious or dangerous properties. Aposematic colors are contrasting patterns of reds, whites, oranges, yellows, or blacks; sounds are usually generalized hisses or stridulatory squeaks that have low complexity and a wide range of frequency emissions; odors are strongly apparent to most organisms and are distinct from common odors; behaviors are obvious jerks, wing flicks, or accentuated movements.

apostatic selection Evolutionary selection force that confers protective advantage to individuals within a population that have rare color or pattern relative to the others.

aptera The parthenogenetic wingless morph of viviparae, specializing in reproduction.

apterous **1.** Wingless; without wings. **2.** Describing an individual lacking wings altogether.

aptery A state of winglessness.

apterygote **1.** Primitively wingless. **2.** Insect orders without wings.

arboreal nest A nest built in a tree or shrub.

archedictyon The primitive original vein network characterizing the wings of many of the most recent insect fossils.

arolium Inflatable pad on the tarsus of thrips.

arrhenotoky The process of producing males from unfertilized eggs.

arthropod vector An arthropod, such as an insect or tick, that can biologically or mechanically transmit an animal disease agent from one vertebrate to another.

artificial honeydew A sugar-containing solution sprayed onto plants with the aim of attracting natural enemies of homopteran insect pests.

artificial selection Selective breeding under laboratory conditions to change a trait. Such selection may be unidirectional or bidirectional; in bidirectional selection some metric is both increased and decreased by appropriate breeding.

arylphorins Storage proteins, rich in aromatic amino acids, that are synthesized in the fat body, stored in the hemolymph, and utilized for cuticle synthesis in connection with a molt.

atmobiotic Living above ground.

ATPase An enzyme that cleaves adenosine triphosphate (ATP) and uses the energy derived from this cleavage to drive further chemical reactions. With respect to excretion, ATPase refers to proteins that drive ion transport across cell membranes.

autocidal control The use of insects to control wild-type populations of the same species.

autoecious See MONOECIOUS.

autogeny The ability to produce eggs without feeding as an adult. Species that can reproduce only in this way have **obligate autogeny**; species that can produce eggs with or without feeding have **facultative autogeny**. The opposite condition is called **anautogeny**.

automixis Parthenogenesis in which eggs undergo meiosis.

B

bacillus Bacterium with a rod shape.

balancer chromosome A chromosome with multiple inversions that suppresses recombination.

ballooning Aeronautical dispersal by means of air currents acting on strands of silk.

Batesian mimic A palatable insect that gains protection by resembling another species that is unpalatable to predators.

Batesian mimicry A parasitic relationship, named after Henry Walter Bates (1862), in which an undefended (e.g., palatable) species copies the signal of poisonous or otherwise defended species in order to deceive the receiver.

behavioral ecology The study of movements and activities of animals as they relate to interrelations with other organisms and the environment.

behavioral fixity Principle stating that the behavior, ecology, and climatic preferences of fossil organisms will be similar to what is found in their present-day descendants at the generic level; used in reconstructing the behavior and ecological preferences of extinct organisms.

benthos Bottom-living organisms in streams or lakes, including biota that ranges from bacteria to algae to invertebrates to fish.

Bergmann's rule The principle that individuals within a species should grow larger in size at higher latitudes (and altitudes) to reduce heat losses by increasing the ratio of surface area to volume.

bioassay (biological assay) The measurement of a biological response (e.g., mortality) to different concentrations of a biologically active compound, using standardized conditions in the laboratory or field.

biodiversity (biological diversity) The variation in all living organisms from gene to ecosystems.

biological amplification The increase in concentration of chemical compounds in higher trophic levels (e.g. predators) of a food web.

biological control The use of living organisms, parasitoids, predators, pathogens, and competitors to suppress pest populations below levels that would occur naturally. Biological control programs are intended to suppress invasive species through importation of specialized natural enemies that have evolved with a pest in its native range.

bioluminescence The production of light by living organisms.

biomonitoring The systematic use of living organisms or their responses to determine the quality of the environment.

bionomics Study of the life history and habits of a species. Often used as a synonym of *ecology*.

biotype A distinct strain of a species.

bivoltine Describing a life cycle with two generations a year.

blastoderm A one-cell-thick layer of cells surrounding the yolk.

blastomeres The embryonic cells produced during the process of cleavage.

book lungs Respiratory pouches of arachnids, filled with closely packed sheets of tissue to provide maximum surface for aeration.

Brachycera The so-called "short-horned" flies, characterized generally by short, three-segmented antennae, the last segment of which is either stylate or aristate, and by larvae that usually have a reduced head capsule.

brachyptery (-ous) The condition of having shortened wings, whose function in flight is lost.

brood The immature members of a colony collectively, including eggs, nymphs, larvae, and pupae.

brood cell A chamber or pocket built to house one or more immature individuals.

bursa copulatrix An accessory structure in the female reproductive tract of some insects that temporarily accepts the spermatophore or sperm from the male.

C

cambium The thin layer of cells beneath the bark of woody plants that gives rise to new cells (phloem and xylem) and is therefore responsible for diameter growth.

cAMP Adenosine 3′,5′-cyclic monophosphate. A compound produced by the enzyme adenylate cyclase upon binding of a hormone to certain receptors located on the cell surface. cAMP carries the information from the hormone into the cell where the response occurs.

campodeiform An insect larva with an elongate body shape.

canaliculus *plural,* **canaliculi**. A small channel.

carapace A hard dorsal covering of the cephalothorax in Arachnida.

carboxylesterase A hydrolytic enzyme that can cleave compounds containing a carboxylester group (COOR).

carrying capacity The maximum population size or density of a species that can be sustained indefinitely in a given environment.

Ca^{2+} signaling Changes in the intracellular calcium (Ca^{2+}) concentration, through the release of Ca^{2+} ions from intracellular stores by the opening of ligand-gated ion channels or the entry of Ca^{2+} ions into the cell through different types of Ca^{2+}-selective channels located in the plasma membrane, that play a role in the regulation of various cellular processes, including cell metabolism, gene expression, cytoskeletal dynamics, and neurotransmission.

caste Any set of individuals in a colony that are morphologically distinct and/or that perform a specialized task.

cauda A specialized process on the tip of the abdomen; used to flick away honeydew waste.

cavernicole An organism that lives in caves and other cave-like subterranean habitats.

ceca Outpocketings of the midgut epithelium. The cecal cells are often differentiated from the other midgut cells both histologically and functionally.

cecidium *plural,* **cecidia**. Technical term for plant gall, an abnormal growth of plant tissue under the influence of a parasitic organism; thousands of insect species induce these growths on a variety of plant species and plant parts.

cecidogenesis The mechanism of instigating physiological changes in host plant tissue to form a gall; this begins with a chemical stimulus passed from insect to plant either at the time of oviposition or as the growing larvae feed upon the plant.

cecidology The study of plant galls and gallmaking.

center of mass The point in a body or system of bodies about which mass is evenly distributed.

center place foraging The process whereby an animal searches for a resource (usually food) starting from a more or less permanent location to which the animal returns after searching is complete.

cephalothorax United head and thorax of Arachnida and Crustacea.

cercus *plural,* **cerci**. Simple or segmented appendages of the 11th abdominal segment of insects that usually act as sensory organs; often called "forceps."

chaetotaxy The arrangement and nomenclature of the setae or bristles on the exoskeleton of an insect.

chelicerae *singular,* **chelicera**. Jaws, each comprising a large basal part and fang.

chemoautotrophic Relating to the ability of an organism to obtain its nourishment through the breakdown of inorganic chemical compounds rather than through photosynthesis.

chemoreceptor Either the chemosensillum (preferably called the "sensillum") or the receptor protein on the sensory cell membrane.

chemotransduction The process leading from a chemical recognition signal (molecular interaction) to a depolarization of a sensory dendrite. This involves an amplification step that allows a small number of stimulus molecules to yield a large cellular response, thus increasing sensitivity.

chitin A polysaccharide that serves as the major fibrous component in insect cuticle.

chitin synthase A membrane-bound enzyme that polymerizes *N*-acetyl-D-glucosamine units into a straight-chain amino sugar homopolymer. The protein has a large number of transmembrane segments that may be involved in translocation of the chitin polymer across the cell membrane.

chlorinated hydrocarbons Organic compounds containing chlorine, hydrogen, and occasionally oxygen and sulfur. The first widely used synthetic organic insecticides.

chorion A complex set of eggshell layers produced by follicle cells toward the end of egg maturation.

clade A hypothesized monophyletic group, i.e., a group of related taxa that includes all descendants of a common ancestor.

cladogram A diagrammatic illustration of the branching sequence of purported relationships of organisms, based on distribution of shared derived features (synapomorphies).

clathrin lattice A honeycomblike structure that adheres to the cytoplasmic face of a cellular membrane, where it folds inward and forms a vesicle.

clavus In a neopterous wing without an expanded anal fan, the area between the claval furrow and the jugal fold.

cleptoparasite An insect that feeds on the food resources or uses the nest of another species.

coadapted genetic system The entire set of interacting genetic factors of an organism.

coelopulse system A part of the nervous system that drives the extracardiac pulsations.

cold hardening An increase in cold tolerance, usually achieved by exposure to a moderately low temperature.

cold hardiness The ability to survive subzero temperatures.

colleterial glands In females, accessory glands that secrete fibrous proteins to fasten eggs to a support; in males, accessory genital glands.

colulus A slender or pointed appendage immediately in front of the spinnerets of some spiders; in others it is greatly reduced or seemingly missing; a homologue of anterior median spinnerets or cribellum.

comb A layer of more or less regularly arranged, closely packed brood cells.

commensalism A form of intimate interaction between members of two different species. One partner of a commensal relationship benefits from the association but the other partner neither benefits nor is harmed during the interaction.

community Populations of different species that coexist and interact with each other in a defined area.

competition An interaction between two or more organisms or species that utilize the same resources, in which the presence of one reduces the birthrate or increases the death rate of the other.

competitive exclusion The elimination of one species from an area of habitat due to competitive interaction with another species.

complementary foods Foods that individually are nutritionally unbalanced, but when combined can provide a balanced diet.

complete (holoblastic) cleavage A coordinated series of mitotic divisions wherein the entire volume of the egg cytoplasm is divided into numerous smaller cells that are fully separated from one another and that together comprise a morula.

compound eye A photoreceptive unit functioning through combined action of two or more ommatidia, each consisting of a convex corneal lens, variably shaped crystalline cone, distal pigment-screening cells, composite rhabdome, and retinular cells.

conductor Semimembranous structure in male palp that supports and guides the embolus in insemination.

condyle Any process by means of which an appendage is articulated into a pan or cavity.

convection The bulk movement of a fluid (either gas or liquid) driven by a physical pressure gradient.

copal Semifossilized (subfossil) resin ranging from several years to thousands of years old with a melting point under 150°C, a hardness of 1 to 2 on the Mohs scale, and a surface soluble to organic solvents.

copulation The period or process during which an intromittent organ (almost always of the male) introduces gametes into a reproductive tract of the opposite sex.

cordon sanitaire Literally, sanitary rope; a line of guards stationed to prevent communication with an infected district.

cornicles See SIPHUNCULI.

corpora allata *singular,* **corpus allatum**. The pair of endocrine glands that make juvenile hormone.

corpora cardiaca *singular,* **corpus cardiacum**. A pair of endocrine glands closely associated with the aorta just behind the brain and connected to it by nerves.

coupling A term for processes and structures that convey mechanical events in the environment to the sensitive plasma membrane of a mechanosensory neuron.

courtship Traditionally viewed as communication between the sexes that brings about successful copulation. However, given that communication during copulation may influence fertilization success, courtship can in theory occur anytime during mating.

coxites Flat appendages on the abdominal sterna, often bearing styli and exsertile vesicles.

critical daylength The daylength marking the transition between a daylength that elicits diapause and one that promotes nondiapause development.

critically endangered Synonymous with endangered; a designation used by the International Union for Conservation of Nature and Natural Resources.

critically imperiled Synonymous with endangered; a designation used by the U.S. Natural Heritage Program.

cryoprotectants Low-molecular-weight polyols and sugars that act as classic

antifreezes and thus function to depress the insect's supercooling point.

crypsis Camouflage of color and pattern or of odor such that the individual blends into the background environment and is not readily detected by predators.

cryptobiosis Complete cessation of active life processes that gives hardiness to extreme conditions of heat, cold, or drought.

cryptonephric complex A structure in which the distal ends of the Malpighian tubules are closely associated with the rectum, and the structure is enclosed within a perinephric space by a perinephric membrane.

culicine mosquito In general terminology, a mosquito classified in the family Culicidae, subfamily Culicinae. In strict usage, a mosquito classified in the tribe Culicini of the subgenus.

cursorial Adapted for running.

cuticle The outer covering of the insect that serves as its skeleton to which the muscles attach. It is secreted by the epidermis and must be shed periodically to allow for continued growth and for the change of form at metamorphosis.

cyclic AMP A cyclic derivative of adenosine monophosphate that is synthesized from ATP by adenylyl cyclase. An intracellular second messenger involved in the regulation or modulation of ion channels, protein kinase activity, and gene expression.

Cyclorrhapha A subgroup of Brachycera characterized by a flagellum of composite first flagellomere and three aristomeres, pupation within the cuticle of the last instar (i.e., puparium), and a life cycle including three instars.

cystoblast Egg precursor cell produced mitotically by division of a germ stem cell into another stem cell and the cystoblast.

cytochrome P450 monoxygenases A ubiquitous group of enzymes involved in the NAPDH-mediated oxidation and metabolism of a broad range of endogenous and exogenous substrates.

cytokine A chemical mediator, secreted by blood cells, that acts at close range on other immunocompetent cells.

cytopathology Pathology of disease exhibited at the cellular level.

D

daytime dark firefly A member of the family Lampyridae that is sexually active in the daytime and is without light organs as an adult (e.g., *Ellychnia*, *Pyropyga*, some *Lucidota*).

dealate **1.** Describing individuals that have shed their wings in termite and ant reproductive castes. **2.** An adult individual that has shed its wings.

debridement Removal of dead or contaminated tissue from a wound.

decomposer An organism that feeds on dead organic material.

dendrite The part of a chemosensitive cell that bears the receptor proteins. It is in close communication with the environment via pores in the sensillum cuticle.

density The number of individuals of a species per unit of habitat.

density dependence The tendency for the birthrates or death rates to change as density of a population increases or decreases.

density independence Factors operating in population regulation that are not related to population density (e.g., climate).

desmosomes Special contact zones between the cytoplasmic membranes of epidermal cells.

determinate growth A pattern of growth in which both growth and molting cease upon reaching the adult stage.

detoxification The fact or process of reducing or removing toxicity. Once insecticides have entered the body, they may be metabolically altered by enzymes in insects as well as other organisms. Through this process of detoxification, such alterations generally produce less toxic products called metabolites.

detritus Organic matter of unidentifiable origin.

detritus feeder An organism that feeds by eating soil or soil particles, from which nutrients are extracted by its digestive system.

diapause A programmed state of dormancy or arrest of development mediated by the neuroendocrine system, usually occurring in a species-specific stage of the life cycle. It may be induced obligatorily or facultatively in response to seasonal cues (day length, temperature, food quality, etc.).

dicondylic Having two condyles.

diffusion The movement of a molecule from high to low concentration driven by random thermal motion.

diflubenzuron An acylurea insect growth regulator that disrupts formation of normal cuticle via inhibition of chitin synthesis.

digestibility coefficients Coefficients that measure an organism's food input and waste output and thus allow one to estimate how much food was digested.

digestion The biochemical process by which food is dissolved and converted into compounds that can be absorbed by the intestinal epithelia.

dioecious A host-alternating aphid life cycle requiring primary and secondary hosts; synonym: *heteroecious*.

diploid Having two complete sets of chromosomes (like a typical adult animal).

disruptive coloration Patterns of coloration that can blur the outlines of a potential prey's body.

dominance hierarchy A social order of dominance established among individuals of one sex, usually males, by aggressive or other behavioral displays.

dominant (genetic) The stronger of a pair of alleles, expressed as fully when in single dose (i.e., heterozygous) as it is when present in double dose (i.e., homozygous). The opposite of recessive.

dorsal diaphragm A complete or fenestrated membrane separating the dorsal vessel in its pericardial sinus from the alimentary canal in its perivisceral sinus.

dorsal vessel The principal hemolymph pump in the body, consisting of a heart in the abdomen and an aorta in the thorax and head regions.

dosage compensation Differential regulation of genes found in different numbers (dosages) within cells of males and females, usually because of heterogametic (XY) sex determination.

ductus seminalis A duct connecting the spermathecal sac with the median oviduct.

Dyar's law An empirical observation that indicates a geometric progression in head width in successive instars of most holometabolous larvae, as proposed by H. G. Dyar in 1890.

Dzierzon's rule The original term for haplodiploid sex determination in honey bees, as defined by J. Dzierzon in 1845.

E

ecdysis The emergence of an insect from its old cuticle during the final stages of a molt.

ecdysone In the context of hormonal control of female reproduction, one of a group of hormones called ecdysteroids. Synthesized by the prothoracic gland, it is the precursor to 20-hydroxyecdysone.

ecdysteroids Steroid molting hormones of insects that are synthesized in prothoracic glands of larvae and in the reproductive organs of adults.

eclosion The molting of a new adult insect from the cuticle of its preceding stage.

ecological potential The ability of a species to adjust to environmental factors.

economic loss For an agricultural crop, the fraction of the realizable yield that is lost to the combined effect of all pests and the physical forces of the environment.

ecosystem A community of organisms and their interaction with the physical environment.

ectognathous Having exserted mouthparts.

ectoparasite A parasite that lives on the surface or within the skin of its host.

ectophagous hyperparasitoid Hyperparasitoid having larval offspring that feed externally on the primary parasitoid host.

ectotherm An organism whose body temperature is strongly influenced by the ambient temperature.

ectothermy The fact of having a body temperature roughly equal to the ambient temperature.

edaphon An organism living in soils.

effective (physiological) temperature Temperature permitting development; the sum of effective temperatures equal to the number of degree days needed for completing a certain developmental period.

elateriform An insect larva with a cylindrical body shape.

elytra *singular,* **elytron**. The anterior leathery or chitinous wings of beetles, serving as coverings to the hind wings, commonly meeting in a straight line down the middle of the dorsum in repose.

embolus A structure in the male palp containing the terminal part of the ejaculatory duct and its opening; it may be small, long, whip-like, or coiled; sometimes divided into several structures.

embryonic diapause The cessation of development in the embryo stage with metabolism markedly reduced.

encapsulation The response of host blood cells, known as plasmatocytes, to the presence of a parasitoid egg or larvae, which results in the formation of a multilayered capsule that causes the death of the parasitoid through asphyxiation.

encoding The creation of a train of action potentials from a sensory receptor potential.

endangered Referring to species that face a high risk of extinction in the near future unless action is taken to protect them; a designation used by the U.S. Fish and Wildlife Service and International Union for Conservation of Nature and Natural Resources. (A formal listing as endangered or threatened by the U.S. Fish and Wildlife Service is the only designation that provides legal protection to species that are at risk of extinction.)

endemic Restricted to a certain region or part of a region; describing a species found naturally in one area and nowhere else; such species can be **neoendemics**, frequently formed through adaptive radiation, or **paleoendemics**, formed through genetic isolation from source populations.

endocuticle The flexible inner layer of procuticle. The endocuticle is not stiffened or cross-linked and is usually resorbed, and its components are recycled with each molt.

endocytosis The formation of vacuoles that transport fluids and solutes into the cytoplasm of a cell.

endoparasite A parasite that lives within the body of its host.

endophagous hyperparasitoid A hyperparasitoid having larval offspring that feed internally inside the primary parasitoid microwasp host.

endoplasmic reticulum A system of cytoplasmic vesicles whose functions include collecting newly synthesized proteins destined for secretion from the cell.

endopterygote A complete metamorphosis with egg, larval, pupal, and adult stages.

endosymbiosis An association of organism in which one organism lives within the other; ants, termites, and other wood-feeding insects contain **endosymbiotic** bacteria that digest cellulose and make it available as food.

endotherm An organism able to keep its body temperature at a level that is more or less independent of the ambient temperature.

endothermy The fact of having a body temperature that is primarily controlled by endogenous mechanisms rather than by the ambient temperature.

endoxyly The fact or condition of living within wood.

enhancer A DNA sequence that recognizes certain transcription factors to stimulate transcription of nearby genes.

enhancer trapping The identification and cloning of enhancer regions through the nearby insertion of a transposable element containing a reporter gene that becomes regulated by the enhancer.

entomopathogenic Causing disease or death in insects.

entomophobia An unnatural and unreasonable fear of arthropods.

envelope A sheath or carton surrounding the combs of a social wasp nest.

epedaphon An organism living on the soil surface and in leaf litter.

epicuticle The thin (but chemically complex) outermost layer of cuticle. The epicuticle is the first layer to be formed and thereafter delimits the space into which the inner procuticle may be secreted.

epigynum A more or less complicated apparatus for storing spermatozoa, immediately in front of the opening of internal reproductive organs of female spiders.

epitope A site on an antigen that is recognized by an antibody or T-cell receptor.

epizootic Describing an outbreak of disease in an animal population.

epoxide hydrolase An enzyme that inactivates juvenile hormone through hydrolysis of the epoxide substituent.

eradication The application of phytosanitary measures to eliminate a pest from an area.

eruciform An insect larva having a caterpillar-like shape.

essential amino acids Amino acids that contribute to protein and cannot be synthesized de novo by most animals, usually resulting in a dietary requirement for such amino acids.

etiological agent The causative agent of a disease, such as a pathogen, a chemical, or a genetic mutation.

euedaphon An organism living wholly within the soil, often well below the surface.

eusocial Fully social; a term applied to some species of insects and other animals

that live in groups for at least part of their life cycle. Eusocial insects exhibit cooperative offspring care and an overlap of adult generations in their groups, and some individuals (workers) forgo reproduction to help others reproduce.

eutrophic Describing a pond, lake, or other body of water containing a rich supply of plant nutrients.

exocrine secretion An exudate discharged to the exterior from a specialized or exocrine gland that functions as an interspecific (allomone) or intraspecific agent (pheromone).

exocuticle The rigid and sclerotized (chemically stiffened) outer layer of some procuticles. Because of its chemical cross-linking, the exocuticle cannot be resorbed during the molting cycle, but its stability makes it a good candidate for structural color production.

exopterygote A condition in which wings develop externally and increase in size with each molt.

exoskeleton The external skeletal structure typically composed of chitin and often complexed with protein and strengthened with calcium carbonate.

extracardiac pulsations Minute contractions of the abdominal body wall muscles that cause pressure pulses in the hemolymph, which moves hemolymph around organs and aids in ventilation.

F

facultative diapause A form of diapause that occurs in response to specific environmental cues that have been received.

facultative hyperparasitoid Hyperparasitoid having progeny that can develop as either a primary parasitoid or a secondary parasitoid.

fang A claw-like part of each chelicera, usually with a groove down which venom and digestive juices can flow into prey.

fat body A structure within the cockroach body that contains stored nutrients, uric acid, and endosymbiotic bacteria, each housed in a different cell type.

feeding guild A group of organisms that exploit a resource in a similar manner (e.g., sap-feeding, leaf-chewing, leaf-mining).

filarial nematode Characteristically long, thin unsegmented roundworms that live everywhere in vertebrates, with the exception of the lumen of the digestive tract.

filter chamber Anastomoses of the foregut and proximal rectum in many (but not all) honeydew-producing insects.

filter feeding The process of acquiring food by straining small particles from the air or water column.

flagellum The third (most distal) segment of an insect antenna.

flash pattern The species-typical unit of light emission of "lightning bug" fireflies that is repeated at somewhat regular time intervals by advertising, mate-seeking males; commonly it is the flashed entity that stimulates the female response flash.

flexion lines Lines or bands of flexible cuticle within the wing, primarily adapted for deformability in flight.

focus (nidus) A place in the environment where ecological conditions ensure the perpetuation of a zoonotic disease agent for an extended period of time.

fold lines Lines or bands of flexible cuticle, primarily adapted for folding the wing at rest.

follicle cells Cells not derived from germ cells that cover oocytes and contribute to their development.

follicles Sperm tubes in the testes in which spermatogenesis occurs.

formulation The form in which an insecticide is applied; the most common formulations are sprays, dusts, and baits.

fossorial Of a limb, body part, etc., adapted for digging or burrowing through soil.

founder event speciation Mode of speciation generally associated with island biota: a small number of individuals, perhaps even a single fertilized female, colonize and become established on an adjacent island or in a geographically isolated habitat.

fragment island An island formed as a result of separation from a larger landmass.

frass Solid excrement from an insect, often containing undigested plant or wood parts.

frequency-dependent selection Selection that is positively or negatively dependent on the frequency of phenotypes in the population. It is negative when phenotypes that are rare are favored (leading to balanced polymorphisms) and positive when phenotypes that are rare are disfavored (leading to stabilizing selection and monomorphism).

fumigant A volatile chemical that acts as a poisonous gas in a confined area.

functional response The relationship between the quantity of food available to an organism and the quantity actually eaten.

fundatrix The parthenogenetic morph that hatches in the spring from the recombinant egg and gives rise to the viviparae; synonym *stem mother*.

G

GABA γ-Aminobutyric acid, a naturally produced inhibitory neurotransmitter that acts on its specific receptor, the **GABA receptor**, to suppress excitation by opening a chloride channel.

gallmaker The organism that induces the host plant to form the gall structure, which includes an excess of protein-rich parenchyma upon which the gallmaker can feed.

gametocytes Infective haploid blood stage of the malaria parasite for mosquitoes acquired during blood feeding. Sexual union of micro- and macrogametes occurs in the midgut of the definitive mosquito host.

gap genes Genes expressed in broad domains in the *Drosophila* embryo; mutations in these genes cause large deletions of the body plan.

gene The unit of heredity.

gene tagging The identification and cloning of a gene based on the insertion of a transposable element that inactivates it or alters its expression.

genetic marker A gene that enables one to detect a transgenic individual. The marker gene is located within the gene vector, usually a transposable element, and its expression indicates that the vector has inserted it into the target DNA.

genetic polymorphism The occurrence together of two or more discontinuous and heritable forms of a species at such frequencies that the rarest of them may not be maintained merely by recurrent mutation.

genetic sexing The selective production of one sex over the other as a direct consequence of a genetic difference between them.

genetic transformation A stable, heritable change in genotype caused by the incorporation of foreign DNA into the genome.

genome The entire nucleotide sequence of an organism, including the entire set of genes.

genomics The comparative study of genomes, the entire ensemble of an organism's genetic material.

genotype All or part of the genetic composition of an individual or population.

geomagnetic field A dipolar magnetic field surrounding the earth with poles in the vicinity of the earth's poles. The field's lines of force are approximately horizontal in the vicinity of the equator and tilt increasingly upward toward the South Pole and increasingly downward toward the North Pole.

geomagnetic vector The description of a local geomagnetic field comprising the local total strength of the field, the dip angle (inclination) of its lines of magnetic force relative to the horizontal and the usually small deviation (declination) of these lines of force from true north.

germ anlage The cluster of cells that forms the embryo proper; the germ anlage represents the ventral surface of the embryo.

germarium The distal region of an ovariole, consisting of germ stem cells, young cystoblasts, and precursor to follicle tissue.

germ band In the blastula, a thickened band of cells that forms the future embryo. The germ band consists of a head, three gnathal, three thoracic, and eight to ten abdominal segments; it has been referred to as the "phylotypic" stage for insects.

germ cells Cells that are ancestors of oocytes.

germ-line differentiation Specification early in the development of testes (spermatogenesis) or ovaries (oogenesis), from a small group of cells in embryos, and subsequent production of mature sperm and eggs.

glowworm A term historically and currently applied to the north European glowworm firefly, *Lampyris noctiluca*, but for more than a century used as a general term for worm-like or larviform glowing organisms, including lampyrid and phengodid larvae and larviform females and fungus gnat larvae.

glowworm firefly A "primitive" member of the family Lampyridae that uses glow signals for sexual communication, typically with glowing females attracting nonluminescent males (e.g., *Microphotus, Pleotomus*).

gluconeogenesis A metabolic pathway responsible for the synthesis of glucose from amino acids, lactate, and glycerol.

glutathione S-transferases (GSTs) Enzymes that catalyze the metabolism of a range of substrates after their conjugation with the endogenous tripeptide glutathione.

glycolysis The principal metabolic process responsible for oxidation of glucose to pyruvate during cellular respiration.

gnathal Referring to the feeding appendages, including the mandible and maxillae.

Golgi bodies Membrane-bound cellular organelles associated with the endoplasmic reticulum; cellular products produced by the endoplasmic reticulum are bound by Golgi bodies possibly for further modification and/or transport.

Gondwanan distribution Having a geographic distribution on more than one of the southern continents that were once united to form Gondwanaland (a supercontinent precursor to current landmasses).

G protein-coupled receptors Integral membrane proteins that constitute a large family of neurotransmitter, hormone, or olfactory receptors. Characterized by seven transmembrane regions. When agonists bind to these receptors, trimeric GTP-binding (G) proteins are activated that then regulate the activity of intracellular secondary effectors, which change intracellular concentrations of second messengers or ion channel activity.

gregarious Liking to be in a crowd. In locusts, this implies that the insects tend to group together spontaneously. Locusts of the **gregarious phase** tend to be very active and to have conspicuous coloration—black and yellow or black and orange.

gross pathology Pathology of disease exhibited at the level of the whole animal, generally observed externally.

gyne Female reproductives in an insect society.

gynopara The specialized winged vivipara that returns to the primary host in the fall and produces the ovipara.

H

haltere The modified metathoracic wings of the dipteran adult, which are sense organs concerned with the maintenance of stability in flight.

hamulus *plural,* **hamuli**. Hook-like setae that can allow wings to be temporarily attached.

haplodiploid Describing organisms that have one sex haploid and the other diploid. **Haplodiploidy** is based upon a sex-determining mechanism by which females develop from fertilized eggs and functional males from unfertilized eggs.

haploid Having one complete set of chromosomes (like a typical egg or sperm cell).

heartbeat reversal During immobile stages or behaviors (e.g., pupal stages), contraction of the dorsal vessel in characteristic peristaltic waves that travel in the anterograde (to the front) or retrograde direction.

heat-shock proteins Stress proteins that are highly expressed in response to heat and other forms of stress and contribute to both high- and low-temperature tolerance.

hematophagous Feeding on vertebrate blood as a food source.

hemidesmosomes Structures that connect the basal cytoplasmic membrane to the basal lamina.

hemiedaphon An organism that completes part of its development in the soil but spends the remainder of its life above ground.

hemimetabolous Describing immature forms that closely resemble the adult form and pass through a fixed number of juvenile instars; the last instar exhibits wing buds and incomplete metamorphosis.

hemocoel(e) The open body cavity of insects and other arthropods through which the hemolymph (blood) circulates.

hemocytes Insect blood cells. The main classes are plasmatocytes, granulocytes, spherule cells, lamellocytes, oenocytoids, and crystal cells.

hemoglobin An oxygen-carrying pigment of the blood possessing a high oxygen affinity and occurring in some aquatic insects inhabiting low oxygen environments.

hemolymph The blood of insects that bathes all tissues via an open circulatory system powered by an open tubular heart. It transports all the nutrients and hormones to the cells and removes the cellular wastes.

heterochrony The precocious or delayed appearance of certain features in the ontogenic development.

heteroecious See DIOECIOUS.

heterogametic sex The sex that carries sex chromosomes of different types and thus produces gametes of two types with

respect to the sex chromosome they carry. In humans, males are the heterogametic sex, carrying one X and one Y chromosome.

heterogeny Alternation of generations in a single insect species, wherein one generation includes both sexes while the other generation includes only females; among gallmaking insects, these alternate generations induce very different galls.

heterothermy (poikilothermy) The condition of having fluctuating body temperature.

heterotrophic Referring to a life condition in which an organism obtains organic chemicals in a preformed state by consuming other organisms or their byproducts.

heterozygote advantage A situation in which the fitness of individuals carrying two different alleles of a particular gene exceeds that of those carrying two alleles of the same type.

hexamerins Abundant proteins in insect plasma, whose main function is storage of amino acids for later use.

Hexapoda A superclass of arthropods containing the Insecta, Collembola, Protura, and Diplura.

higher termites Descriptive term for a family (Termitidae) of Isoptera that lack gut protozoa, instead having bacteria.

histopathology Characteristics of disease observed in specific tissues.

holoblastic cleavage Cleavage in which the entire egg is divided.

holocycly The seasonally cyclic sexual and asexual reproduction of aphids, in which sexual morphs produce a genetically recombinant overwintering egg.

holometabolous Having complete metamorphosis; having four life stages consisting of egg, larva, pupa, and adult.

homeostasis The maintenance of a steady physiological state by means of self-regulation through feedback.

homeotic genes Genes whose mutations cause transformation of segment identity.

homogametic sex The sex that carries sex chromosomes that are the same. In humans, females are the homogametic sex, carrying two X chromosomes.

homologous Describing or referring to structures or organs derived by evolution from the same ancestral structure.

homology The existence of structures that have a common origin.

homozygote An individual who bears two copies of the same allele at a given locus.

horizon (soil) A layer of soil, roughly parallel to the soil surface, that differs in properties and characteristics from adjacent layers above or below it.

horizontal transmission The transmission of a gene or genetic element from one organism to another by any mechanism other than that normally involved in the inheritance of genetic material by offspring from parents.

host plant The organism upon which a gall grows and from the tissues of which the gall is formed.

humoral immunity Immunity conferred by non-cell-mediated defense mechanisms, with the active molecules being plasma-borne.

humus The relatively stable portion of soil organic matter that remains after the major portions of plant or animal residues have decomposed.

20-hydroxyecdysone The most common biologically active steroid in insects.

hypermetamorphosis Development involving more than one larval stage.

hyperparasitoid A parasitoid that attacks another species that is itself a parasitoid, usually when the latter is feeding with its host. Such species may be detrimental to insect biological control programs.

hypognathous Having mouthparts directed ventrally.

I

ice nucleator An agent that facilitates the organization of water molecules into crystals and thus promotes ice formation.

idiobiont A parasitoid whose host is not allowed to develop further after parasitization (e.g., egg parasitoids that complete their entire development within a host egg, larval parasitoids that paralyze their hosts permanently at the time of oviposition).

imaginal discs Clusters of undifferentiated embryonic cells in holometabolous insects that proliferate during larval stages and then differentiate during the pupal stage upon induction by ecdysteroids in the absence of juvenile hormone.

immature Referring to an egg, larval, or pupal stage.

imperiled Synonymous with threatened; a designation used by the U.S. Natural Heritage Program.

indeterminate growth A pattern of growth in which an insect continues to molt after reaching the adult stage.

indirect fertilization The transfer of sperm by mean of an externally laid spermatophore.

induced chemical defense Increased production or mobilization of plant allelochemicals in response to some cue (commonly leaf damage) that indicates a high probability of future attack and that is usually absent in unstressed and undamaged plants.

infective juvenile (J3) The third-stage juvenile or filariform larva found within the biting mouthparts of the arthropod intermediate host.

inflammation A localized tissue response to an antigen that is characterized by pain, redness, swelling, and the influx of white blood cells.

injury In this context, the mechanical or physiological effect on a plant of an insect's feeding, oviposition, excretion, or nesting and sheltering behavior.

innate immunity A nonspecific host defense reaction, such as inflammation, to an infectious agent.

inoculative biological control A form of augmentative biological control in which the goal is to establish reproducing populations of natural enemies at the start of the crop.

inquilinism A replacement name for permanent parasitism. **Inquilines** are social parasites that reside entirely within the nest of their host.

insectaries Rearing facilities in which parasitoids and predators are mass-produced, usually for sale as agents for augmentative biological control programs.

insecticide targets Physiologically or biochemically important molecules (such as those comprising ion channels, receptors, enzymes, proteins) and structural molecules (such as chitin) that, when disrupted by insecticide chemicals, cause damage or death to pest insects.

***in situ* hybridization** A technique in which small fragments of DNA are labeled with radioactive or chemiluminescent compounds and used as probes to localize genes or other segments of DNA.

instar The growth stage of an insect between two successive molts.

intake target The optimal nutrient intake over a given period in development that can be represented as a point in multidimensional nutrient space.

integrated Coordinated or harmonious use of multiple approaches to control single or multiple pests.

intensity Intensity of a vibratory signal can be expressed as a displacement (i.e., a vector quantity that specifies the change of position of a body), as velocity (i.e., a vector quantity that specifies rate of change of displacement), or as acceleration (i.e., a vector quantity that specifies rate of change of velocity).

interkernel space An air-filled space between grain kernels.

intermolt A period of feeding and growth; it begins with ecdysis from previous stage and ends with apolysis.

Inter-Tropical Convergence A zone extending across Africa and Arabia along which northerly winds meet (converge) with those blowing from the southwest. Rising air currents along the convergence lead to rain if the converging winds carry moisture, and, if rain occurs at all in the arid regions of North Africa, it is most likely to be along the convergence.

intragenomic conflict A situation in which the action of one gene, by increasing its chances of transmission, is in conflict with the interests of other parts of the genome.

inundative biological control A form of augmentative biological control in which natural enemies are released in large numbers throughout the cropping period, with no expectation that the released biological control agents will establish; pest control is expected from the individuals actually released, not their progeny.

involucrum A sheath of resin and wax surrounding the brood chamber in the nest of most stingless honey bees.

isometric growth A pattern of growth in which a body part grows at the same rate as other body parts or the body as a whole.

ivermectin An oral antiparasitic drug also effective against some ectoparasites.

J

Johnston's organ A mechanosensory organ located in the pedicel that responds to changes in antennal position.

jugum The wing area posterior to the jugal fold.

juvenile hormone (JH) One of the two major insect developmental hormones (the other is ecdysteroid or molting hormone). It is produced by the corpora allata under control of the brain. During larval stages JH maintains larval characters. In the adult it influences egg development and other traits, including migration.

juvenile hormone esterase An enzyme that inactivates juvenile hormone through cleavage of the methyl ester bond, yielding juvenile hormone acid.

K

kairomone A chemical that is produced by one organism conveying information to another organism of a different species; it is advantageous to the recipient but detrimental to the producer of the chemical.

karst A landform created by solution of a substrate, usually limestone and other carbonates; characterized by subsurface drainage and fissures, sinkholes, underground streams, and caverns.

key-factor analysis An analytic procedure for identifying the main causes of observed fluctuations in population density.

keystone predator A predator whose activities (or lack of same) determine the species composition of an entire community.

kin selection The principle that reproduction by an individual's relatives can increase that individual's genetic representation in future generations, above and beyond its own specific reproductive output. Therefore, adaptations can evolve that favor relatives' survival and reproduction, as well as personal fitness.

kleptoparasite An individual or species that steals the food resource secured by another. This may be by taking the food away, as with some sphecid wasps that enter unguarded nest burrows of conspecifics and remove the prey items another individual has stowed there; or it may involve laying an egg on the food in the nest of another wasp, not necessarily a conspecific, which the kleptoparasite's larva then consumes.

klinokinesis An undirected change in the turning movements of an organism resulting from the intensity of a stimulus.

koinobiont A lifestyle in which a parasitoid allows its host to continue to feed and/or develop after oviposition, such that its larvae feed on an active host that is killed at a later stage. For example, many endoparasitic larval parasitoids lay their egg into a young host larva (or even egg) but do not complete development until the host has grown, and maybe even pupated.

L

labeled-line code A simple sensory code by which one cell or a single cell type sends a unique message to the higher processing centers.

labellum The apical part of the adult proboscis, comprising the modified labial palpi.

labium The "lower lip" of the insect head.

labrum The "upper lip" of the insect head.

lacinia The distal median lobe of the insect maxilla.

Lamarckian evolution A theory proposed by Jean Baptiste Lamarck, describing the passing on of an organism's characteristics to succeeding generations as the result of environmental influence on the organism during its lifetime; a predecessor to the Darwinian theory of evolution.

larviposition The deposition of living larvae that have already hatched inside the female.

latent period The time interval between when a vector acquires a pathogen and when the vector is able to inoculate the pathogen into a susceptible host.

lek behavior A mating system in which assemblages of males defend territories to which females are attracted solely for the purpose of mating.

lerp Sugary and waxy formations made by larval psyllids for their protection.

life cycle The sequence of events in the life of an insect from hatching; immature development up to adult emergence.

life table A summary of the survival rates of individuals in a population to each life stage or age category.

lightning bug firefly A member of the family Lampyridae whose adults use flashes, flickers, or other rapidly controlled bioluminescent emissions for sexual signaling (e.g., *Photinus*, *Photuris*).

lignocellulosic diet A diet consisting of wood or other lignified plant materials, whether sound or in different stages of humification.

lineage A clade, or monophyletic group; the group is defined on the basis of definitive derived traits.

lipid A chemically diverse group of molecules that are insoluble in water and other polar solvents.

lipophorin A plasma protein in insect hemolymph that transports lipids between tissues.

locus A site on a chromosome; the term is sometimes also used to refer to the gene itself.

logistic growth Population growth that is influenced by the carrying capacity of the environment.

lower termites A descriptive term for families (Mastotermitidae, Kalotermitidae, Termopsidae, Hodotermitidae, Rhinotermitidae, Serritermitidae) of Isoptera that have symbiotic intestinal gut protozoa.

luciferase The generic name given to certain biological catalysts of bioluminescence, which are enzymes that facilitate a light-emitting reaction.

luciferin The generic name given to substrate molecules that are oxidized in many light-emitting reactions.

lumen The open center of a duct.

luminescence The emission of light energy by molecules involving shifting of subatomic particles, contrasting with incandescence.

M

macroinvertebrate A collective term for aquatic insects and other arthropods without backbones and generally visible to the unaided eye.

macrophyte A large aquatic plant that can occur above, below, or on the water surface.

maggot therapy Therapeutic myiasis; introducing live fly larvae into wounds to treat them.

magnetite A common crystallized ferromagnetic mineral of iron oxide (Fe_3O_4). Submicroscopic particles of magnetite (<100 nm) have been localized in cells of various organisms, from bacteria to pigeons. These particles may implement the sensing of local geomagnetic vectors.

major/minor morphs Forms that result from dimorphism in the size of beetles and the horns and other ornaments on the head and thorax of males.

malaria paroxysm A clinical attack of malaria in a human host; associated with the liberation of parasites from the red blood cells, featuring cold (shivering, lasting <1 h), hot (fever as high as 41°C, 2–6 h), and sweating (fever breaks, temperature drops rapidly to or below normal) stages.

male accessory glands Secretory glands associated with the male reproductive tract that produce seminal fluid and the structural components of the spermatophore.

Malpighian tubules Long, narrow outpocketings of the gut; the site of primary urine formation in most insects.

mandibulate Having mandibles adapted for chewing or biting.

mating disruption The application of a formulated pheromone to a crop in order to interfere with mate finding by a pest insect.

matric potential The portion of the total soil water potential due to the attractive forces between water and soil solids.

meconium An accumulation of waste products from the larval stage.

melanin A black pigment, toxic to parasites and pathogens; formed from the precursors tyrosine and DOPA via the phenoloxidase reaction.

meristem Undifferentiated tissue capable of active cell division and differentiation into specialized tissues.

meroblastic cleavage Cleavage in which only the nuclear material of the egg is divided.

metabolic rate The rate at which an organism converts chemical energy (fuels) into heat. Increased metabolic rate is correlated with greater oxygen consumption and with greater production of carbon dioxide, water, and heat.

metamere (somite) A true body segment derived during embryonic development.

metapopulation A set of local populations linked together by dispersal.

microfilaria The first-stage juvenile in the developmental life history of dog heartworm found in the host's blood.

microtubules Proteinaceous tubular cytoplasmic structures that are part of the cytoskeleton and are involved in force transmission and movements of cell organelles.

microvilli Fingerlike projections of cells, bound by the cell membrane and extending outward into extracellular space; usually associated with absorptive or secretory regions to increase the surface area of the cell.

microwasp A wasp that is 4 to 5 mm in length or smaller.

midgut A region of the insect gut that lies between the foregut and the hindgut. The midgut is derived from embryonic endoderm. The midgut is the site of food digestion and nutrient absorption in insects.

mimicry The close resemblance of insects to different species with the resultant gain of protection by being mistaken for something else that is dangerous, unpalatable, toxic, etc.

mimicry ring A group of coexisting species that use the same signal, for example, a group of unrelated butterflies bearing the same warning coloration. A mimicry ring typically involves one or several models that evolved the signal first and one or several mimics that were later selected to use the same signal as the models.

mirror A term for the membranous area whose border is formed by the posterior cubitus vein on the right forewing, serving as a sounding board during stridulation.

mitochondria Cellular organelles that, among other functions, produce adenosine triphosphate (ATP), the primary energy source for cellular function.

mode of action A term for the mechanism by which an insecticide affects an insect.

mola Thickened grinding and crushing region at the inner edge of the mandibular base.

mollicute A member of a class of bacteria that lacks a cell wall.

molt The period during which an insect synthesizes new cuticle and other structures appropriate for the next developmental stage; it begins with ecdysteroid-induced apolysis and ends with ecdysis of old cuticle.

monocondylic Having one condyle (especially on the mandible).

monoecious A nonhost-alternating aphid life cycle that requires a single host; synonym: *autoecious*.

monogyny The presence in the nest of a single functional queen.

monomorphic/polymorphic Describing the distribution of body shapes and/or sizes among the members of a social group. If body size variation clusters around a single average value, typically approximating a normal or bell-shaped distribution, and body shapes are isometric, then group members are said to be monomorphic. In polymorphism there are two or more discrete (nonoverlapping) size classes of individuals, or individuals exhibiting differences in body shape.

monophagous Specialized to feed and develop on one host plant.

monophyletic Referring to a taxonomic group (called a clade) that contains all descendants derived from a single ancestor and recognized by the possession of a shared derived feature(s). For example, the clade Diptera is monophyletic, recognized by shared derived development of the hind wing as a haltere (balancing organ).

monophyly The classification or status of being monophyletic (see above).

morphogenesis The phase of embryogenesis in which specific tissues, organs, and structures develop.

Müllerian mimicry A mutualistic relationship, named after Fritz Müller (1879), in which several prey species, or co-mimics, with anti-predator defenses (e.g., bad taste, toxicity, venom) direct the same signal to the receiver. Despite being mutualistic, Müllerian mimicry is not necessarily coevolutionary and co-mimics do not necessarily benefit equally from the mimicry.

multivoltine Describing species that have two or more generations per year. Those that have a single generation per year are called *univoltine*.

mutation Any change in genetic material, though the term usually is used to refer to an error in replication of a nucleotide sequence.

mutualism An interspecific relationship in which both partners (**symbionts**) benefit from the interaction.

mya Millions of years ago.

mycangium An invagination of the cuticle within which microorganisms are held, facilitating their transfer to uninoculated host substrate.

mycetocytes A specialized form of fat body cells harboring symbiontic bacteria that might contribute essential nutrients. Mycetocytes are in proximity to urocytes in the fat body lobe, suggesting a biochemical interaction between the two types of cell.

mycetophagous Describing an organism that is fungus feeding.

myiasis The invasion of living tissues by fly larvae (maggots).

myrmecophiles Organisms, generally insects other than ants, that live in or around ant nests and that exploit resources from the ants by begging, scavenging, or predation.

N

naiad A term for the larva of the aquatic, hemimetabolous insect orders.

nasute A soldier in the family Termitidae that has a well-developed horn-like median projection (**nasus**) and reduced mandibles. A defensive fluid may be produced and ejected. Soldiers for some species in the family Rhiontermitidae may also produce a defensive fluid from a fontanelle gland on the top of the head, but do not have the horn-like median projection.

natural control The level of control produced by natural enemies that occur normally without any active management by humans.

Nematocera The paraphyletic or grade-level group of so-called "long-horned" flies or "lower Diptera," characterized generally by long, multisegmented antennae and by larvae with a well-developed, sclerotized head capsule.

neonates Newly hatched young.

neopterous Having a structure that permits folding the wings at rest down to the sides of the body.

neosome An organism altered by neosomy.

neosomule A new external structure resulting from neosomy.

neosomy The ability to produce new cuticle without molting.

neotenic An individual that retains juvenile morphological features after having reached sexual maturity.

neoteny A condition in which nymphal or larval structures are carried over into the adult and the corresponding adult structures are suppressed.

nest Any modification of the environment by adult insects that provides shelter for the rearing of their offspring.

neurohormone A small organic or peptidergic substance that is produced in neurosecretory cells; released into the hemolymph at special regions called neurohemal organs and transported to target tissues with the hemolymph.

neuromodulator A neuroactive substance that is released by synaptic terminals. It simultaneously acts on large numbers of cells in the proximity of the releasing cell and modifies the properties of synaptic transmission and the properties of target cells.

neurosecretory cell A neuron that produces one or more peptides or small proteins (neuropeptides), stores the peptide(s) at the axon terminal, and then releases it into the hemolymph in response to excitatory neural input.

neurotransmitter A chemical substance that is released from the presynaptic endings of a neuron. It transmits information across the synaptic cleft to specific receptors located on the surface of postsynaptic cells.

niche The environmental limits within which a species can survive and reproduce.

nit Informal term for a louse egg.

nontarget impact Mortality or injury that may occur to neutral or beneficial species as a consequence of the use of natural enemies in biological control programs or insecticide applications.

nuclear receptors A family of soluble proteins that are mobilized by steroid hormones to coordinate gene expression through direct binding to DNA.

numerical response An increase in the number of predators or parasitoids in response to an increase in prey density.

nutritional rail A means of representing a food in a nutrient space as a line which moves from the origin into the space.

nymph The larval stage of hemimetabolous insect orders.

O

obligate hyperparasitoid A hyperparasitoid that is always a secondary parasitoid; that is, its progeny can develop only on or in a primary parasitoid wasp host.

obligate parasite A parasite that cannot complete development or reproduce without obtaining nourishment from a host animal.

obligatory diapause A form of diapause that is genetically programmed to occur at a specific developmental stage regardless of the environmental conditions that prevail.

ocellus *plural,* **ocelli**. A simple eye consisting of photosensitive cells and sometimes a single, beadlike lens; multiple ocelli may be present, but their action is not coordinated.

odor binding protein (OBP) A protein that binds and transports an odor molecule to the receptor protein on the sensory cell dendrite. It is needed to convey the water-insoluble odor molecule across the watery lumen of the sensillum.

oenocytes Special cells associated with epidermis and tracheae. They are involved in synthesis of waxes, such as beeswax and waxes used for waterproofing cuticle and eggshells.

official control A term for the suppression, containment, or eradication of a pest population by a plant protection organization.

oligogyny The presence in a nest of two or more functional queens.

oligophagous Feeding on a few, often related plant taxa, such as different species in one genus or a few genera in one family.

oligosaccharide A molecule composed of 4 to 20 monosaccharide sugar units.

oligotrophic Describing a pond, lake, or other body of water that is deficient in plant nutrients.

ommatidium *plural,* **ommatidia**. An individual unit of the compound eye.

oocyte A female gamete or developing egg.

oogenesis The formation and development of the eggs in the female. It includes egg maturation and yolk formation (vitellogenesis).

ootheca A case or other covering of an egg mass that has been laid.

optimal foraging theory A theory that attempts to document the most efficient ways (in terms of evolutionary fitness) for an organism to find food or other resources.

optimal temperature The temperature at which any body performance or activity is most efficient.

orthokinesis An undirected change in speed of movement of an organism resulting from the intensity of a stimulus.

Orthorrhapha A probable paraphyletic subgroup of Brachycera in which the pupae are not enclosed in puparium and in which the sclerotized portions of the larval head capsule are exposed externally.

osmoregulation Control of osmotic pressure or salt and water balance.

osmotic potential The portion of the total soil water potential due to the presence of solutes in soil water.

ostia *singular,* **ostium**. Valves that can be permanently open or flapped to allow uni- or bidirectional flow of hemolymph into and out of dorsal vessel and accessory pulsatile organs.

ovariole An element of the insect ovary that consists of a tapering tubule in which oocytes are produced.

ovipara The sexual female produced by the viviparae (or gynoparae) in the fall; oviparae mate with males on the primary host, producing the overwintering genetically recombinant egg.

oviparity Egg laying as the normal means of reproduction.

ovipary Alternate term for oviparity (egg laying).

oviposition The act of laying eggs; the passage of an egg from the median oviduct to outside the insect's body.

ovipositor The organ by which eggs are deposited.

ovoviviparity The production of well-developed eggs that hatch inside the mother's body.

ovulation The passage of an egg from the ovariole into the oviduct.

oxic Containing a certain amount of molecular oxygen; describes the oxygen status of an environment.

P

pair formation The stage before mating during which the sexually receptive male and female are attracted to one another.

pair-rule genes Genes expressed in seven stripes in the *Drosophila* embryo; mutations in these genes cause deletions in a two-segment periodicity.

Paleoptera Insect orders that have direct flight muscles and lack the ability to fold their wings over their back while at rest.

paleosymbiosis Fossil evidence of associations between two different species; this includes **paleoinquilinism** (two organisms in the same niche but neither benefits nor is harmed by the other), **paleocommensalism** (one organism benefits, neither is harmed), **paleomutalism** (both organisms benefit, neither is harmed), and **paleoparasitism** (one organism benefits by taking nourishment at the expense of the other, which is harmed, often killed).

palp/palpal organs Variably complex structures found in the terminal part of adult male palp.

pandemic A disease that affects populations in many countries.

Pangaea The single continent that existed about 220 mya; about 135 mya, this protocontinent broke up and drifted apart into two pieces, Gondwana to the south and Laurasia to the north.

paracellular permeability The ability of substances to travel from one side of an epithelium to the other by moving through the spaces between cells.

paracrine Describing chemical messengers that act on cells in the immediate area without entering the blood.

parameres pairs of annulated gonapophyses on abdominal segments VIII and IX or IX.

Paraneoptera A superordinal group of insects related to the Holometabola and consisting of the orders Hemiptera, Thysanoptera, Psocoptera, and Phthiraptera.

paraphyletic Referring to a taxonomic group containing a single ancestor and some, but not all, of its descendants.

paraphyly The status of a group that includes some, but not all, the descendants of a single common ancestor.

parasite An organism living in a relationship with another organism in which it benefits but the other does not (and typically is harmed).

parasitoid A specialized form of parasite that develops in or on a host, eventually killing it. Larval parasitoids require one host to complete development. In addition to parasitism, adult female parasitoids can kill hosts through host feeding. Most hosts of parasitoids are other insects, and most parasitoids are in the orders Hymenoptera and Diptera

paratransgenesis The process of genetically modifying endosymbionts, which results in the malnutrition and death of the insect that requires them.

parental care Behavior undertaken by a parent organism that is beneficial to offspring; e.g., in some beetle species the mother remains underground tending the brood mass until development of her offspring is complete.

parental investment Any behavioral or metabolic investment by a parent that increases the individual offspring's survival at the expense of the parent's ability to benefit other offspring.

parthenogenesis Reproduction in which eggs develop without fertilization by a male gamete.

pathogen An organism that causes disease. Viruses, bacteria, microsporidia, and fungi can be pathogens.

pattern formation The process by which embryonic cells form ordered spatial arrangements of differentiated tissues; during insect embryogenesis, patterning begins with specification of the anterior–posterior and dorsal–ventral axes and ends with the formation of the characteristic structures associated with each segment of the head, thorax, and abdomen.

pattern generation Control of repeating movements such as walking, wing flapping, swimming, and breathing. The neural circuits that control these behaviors include neurons within the central nervous system that are often influenced by sensory inputs from the periphery.

pedicel The second segment of an insect antenna.

pedogenesis Reproduction by an anatomically immature life stage.

pedomorphosis An evolutionary change in which adults of a later species retain

some of the characteristics of juveniles of the ancestral forms.

perimicrovillar membranes In hemipteran midgut cells, lipoprotein membranes that ensheath the microvillar membranes like glove fingers and extend toward the luminal compartment; they are apparently involved in amino acid absorption from dilute diets.

peritrophic membrane A filmlike anatomical structure made of protein and chitin that separates food from midgut cells; it is believed to have evolved from an ancestral mucus and combines the protection function of this mucus with several roles in digestion associated with midgut compartmentalization.

permissive host A host insect species in which a parasite or pathogenic organism establishes a successful infection. Species or host strains in which the development of the invader is thwarted by hostimmunological defenses are **nonpermissive**.

pest A general term for any species whose activities interfere in some way with human health, comfort, or profit. In **integrated pest management (IPM)**, a pest can be an insect or other animal, a plant, or a microbial pathogen.

petiole The slender abdominal segment of Hymenoptera.

phagocyte A cell that is able to engulf foreign bodies.

phagocytosis The process of "cell eating," in which small particles or microorganisms are engulfed by blood cells.

phagostimulant A chemical that induces and maintains feeding.

phagostimulate To promote (stimulate) feeding, usually by a nutrient.

pharate The early period of an instar in which a new cuticle has been formed, and the insect is still enclosed in the cuticle of the former instar.

phenology The periodicity of life cycle stages in relationship to seasonal occurrences in a given region.

phenotype The outward appearance of an individual, being determined by its genotype and the influence of the environment on the expression of this genotype.

pheromone Chemical message released by an individual that induces either a behavioral reaction or a developmental process in other individuals of the same species.

phloem Active, food-conducting tissue of the inner bark of trees and other woody plants, which with age becomes outer bark. Food is stored in this tissue and is also conducted to the roots for storage.

phosphorylation Transient, reversible posttranslational modification of proteins in which the terminal phosphate group of ATP is transferred to specific residues of a polypeptide by kinases and often alters the properties of the protein.

photoperiod The length of "day," i.e., **photophase**, and "night," i.e., **scotophase**, usually during a 24-h period.

phreatic Of or relating to groundwater, specifically the saturated zone below the water table.

phylogenetic Incorporating the evolutionary history and expressed in ancestor-descendant relationships of a species or taxonomic group.

physiogenesis A physiological (still unknown) process in the diapause stage that leads to the termination of diapause; sometimes termed *diapause development*.

phytosanitary regulations Official rules to prevent introduction and/or spread of pests by regulating the production, movement, or existence of commodities or other articles, or the normal activities of persons.

planidium The first instar of some hymenopteran parasitoids. **Planidial** larvae can move about actively because they have hardened integuments and spines.

plasma The liquid portion of hemolymph, consisting of water and dissolved solutes.

plasmid An independent, stable, self-replicating piece of DNA in bacterial cells that is not a part of the normal cell genome. Plasmids are commonly used for cloning.

Plasmodium A genus (order Coccidiida, family Plasmodiidae) that contains 100 species of blood parasites of vertebrates, of which *Plasmodium malariae*, *P. vivax*, *P. falciparum*, and *P. ovale* cause malaria in humans.

plastron A structure by which a layer of air is held close to the body and by which gas exchange occurs with surrounding water.

plesiomorphic Referring to the ancestral condition of a character state.

pneumonic Spread from the lungs by coughing and spitting.

poikilothermy See HETEROTHERMY.

polydnavirus A viruslike entity comprising multiple circular DNA molecules in a protein coat. Polydnaviruses are encoded by the genomes of some parasitic wasps (some ichneumonids and braconids), produced in the female reproductive tract, and injected into hosts, where they are expressed and help overcome the host's immune defenses.

polyethism Caste designation or division of labor in social insects. In **caste polyethism**, morphological specialization involves different functional castes, whereas **age polyethism** refers to a change in behavior related to age.

polygenic character (trait) A trait controlled by the integrated action of multiple independent genes.

polygyny The presence in a nest of multiple, functional queens.

polymerization A process by which small molecules (**monomers**) combine chemically to produce a large network of molecules, or **polymers**; one of the main processes by which resin becomes amber.

polymorphism The existence of two or more genotypes for a given trait within a population.

Polyneoptera A superordinal group of insects consisting of the Plecoptera, Embiidina, Zoraptera, Orthoptera, Phasmida, Dermaptera, Grylloblattodea, Isoptera, Mantodea, and Blattaria.

polyphagous Feeding on many plant species from a range of families.

polyphenism The occurrence of alternate phenotypes caused by environmental cues and usually mediated through hormonal signaling. Polyphenisms can occur sequentially as a part of a developmental sequence (larva, pupa, and adult) or alternatively as either seasonal morphs or castes in social insects.

polyphyletic Referring to a taxonomic group derived from more than one ancestor and recognized by the possession of one or more features evolved convergently. For example, if the primitively wingless silverfish were united with secondarily wingless grasshoppers, beetles, and flies, the resulting group would be polyphyletic.

polyploid Having more than two complete sets of chromosomes (e.g., triploid, having three sets; tetraploid, having four sets).

polytene chromosomes Giant chromosomes arising from replication without mitosis, allowing visualization of banding patterns and puffs indicative of transcriptional activity.

population A particular group of interbreeding individuals more or less separated in time or space from other groups of the same species.

preadaptation The existence of a preexisting anatomical structure, physiological process, or behavior pattern that makes new forms of evolutionary adaptations more likely.

predator An organism that attacks and eats another; a true predator is distinct from a parasitoid in that it does not develop on or within the host and usually kills and consumes many prey during its lifetime.

primary host A term for an aphid's overwintering host, where egg deposition occurs in a dioecious life cycle; usually a deciduous woody angiosperm that was the ancestral host before host alternation was adopted.

primary sex ratio The sex ratio at fertilization or zygote formation.

proctodeal trophallaxis The feeding on microbe-rich hindgut fluid solicited from nestmates; found only in termites and wood-feeding cockroaches.

procuticle The inner layer of cuticle, beneath the epicuticle. In stiffened (sclerotized) cuticle, it typically consists of exo- and endocuticle, but in flexible cuticle, the exocuticle is not present.

prognathous Having mouthparts directed anteriorly.

prokaryotes A collective name for all microorganisms without a true nucleus, comprising the Archaea (formerly Archaebacteria) and the Bacteria (formerly Eubacteria), which form two phylogenetically independent lineages next to the Eukarya (also, eukaryotes).

prolegs Unsegmented, false legs, as found in many larval insects.

pronation A classification of wings; i.e., twisting, anterior margin downward, posterior margin upward.

propagule The initial group of individuals that found an introduced population.

propolis Resins collected from plants and used in the nest by honey bees.

prostheca Any of a variety of structures—rigid or flexible, simple or complex, lobe-like or setose—arising from the mesal surface of the mandible just distal of the mola.

prothoracic gland A diffuse endocrine organ in the thoracic area of insects that produces ecdysone, the precursor of 20-hydroxyecdysone.

pseudergate A caste in some species in the families Termopsidae, Kalotermitidae, and Rhiontermitidae; a nonreproducing and nonsoldier caste that is developmentally very plastic and can do "worker" tasks within colonies.

pseudogamy A form of sperm-dependent parthenogenesis in which eggs require activation by entry of sperm, but only maternal chromosomes are expressed and passed on.

pseudokarst A landform mimicking karst in form but created by volcanism or erosion; like karst, drainage is primarily subterranean.

pterygote Describing insect orders with wings.

puparium The sclerotized exuvium of the last instar, within which the pupa (intermediate developmental stage between larva and adult) is formed.

pyrethroid An organic synthetic insecticide with a structure based on that of pyrethrum, which is a natural botanical insecticide.

Q

quarantine A secure importation facility designed to prevent the unintentional release of natural enemy species into new geographic regions after their discovery during foreign exploration.

quarantine pest A pest of potential economic importance to a given area endangered thereby and not yet present there, or present but not widely distributed and being officially controlled.

quiescence A forced state of dormancy imposed by an immediate effect of environmental conditions that prevent the normal activity.

R

receptaculum seminis The spermathecal sac in which sperm are stored.

recessive referring to an allele that is not expressed phenotypically when present in a heterozygote, but only when in a homozygote. The opposite of dominant.

rectal pads Thickened regions of the rectal epithelium in which the cells are specialized for active transport of solutes across the epithelial wall.

reference condition The environmental condition that is representative of a group of minimally disturbed sites organized by selected physical, chemical, and biological characteristics.

reflex bleeding The utilization of blood (hemolymph), sometimes fortified with deterrent allomones, as a defensive secretion in adversarial contexts. **Reflex bleeders** exude hemolymph from weak sutures in the chitin (e.g., at femoral joints). Used by beetles to defend themselves against attack from ants and other predators.

refuge A location in or near crop fields in which natural enemies are protected from pesticides and provided with resources necessary for their survival and reproduction.

remigium The area of the wing of a neopterous insect lying anterior to the claval furrow.

reproductive success An estimate of evolutionary fitness used to measure sexual selection. Measures can include success in obtaining matings with different numbers of the opposite sex as well as fertilization success (by males) when females mate and store sperm from several males.

required nutrients Nutrients, including essential nutrients, that enhance or optimize growth, development, and/or reproduction.

reservoir A term for an animal, usually a vertebrate, that is infected with an animal disease agent, can maintain the agent in its tissues for a prolonged period, is fed upon abundantly by one or more efficient vectors, and is capable of serving as a source of infection for such arthropods. Some vectors that transmit agents transstadially, transovarially, or by both mechanisms also serve as reservoirs.

resilin A very elastic protein found in insects that is particularly associated with joints. Resilin is rubberlike and can store energy and can recoil to produce efficient movement.

resin A water-insoluble exudate emitted from parenchymal cells of various plants, both angiosperms and gymnosperms; it is composed mainly of a mixture of terpenoids, acids, alcohols, and carbohydrates and can be molded by hand.

resonance A condition such that the frequency of an applied force is equal to the natural frequency of a system; the frequency of vibration of a system when the amplitude of vibration is a maximum.

respiration The collection of metabolic pathways responsible for the oxidation of glucose and fatty acids, with the production of energy involving an electron transport chain.

rhabdome The rod-like structures that collectively comprise the "retina" of an ommatidium.

RNA interference The ability of double-stranded RNA to inactivate the expression of the homologous cellular gene.

S

salivary glands In mosquitoes and other blood-feeding dipterans, glands located in the thorax of both immature and adult stages. In adult females, the glands are associated with blood feeding. Certain pathogens causing diseases in vertebrates are injected into feeding wounds during blood feeding, thus resulting in a type of transmission known as salivarian.

sapwood The soft wood just beneath the inner bark of a tree.

scabreiform An insect larva that is grub shaped.

scape The first (most proximal) segment of an insect antenna.

sclerotization A chemical process in which insect cuticle and other materials are rendered hard, insoluble, dehydrated, and resistant to degradation by the oxidative incorporation of phenolic compounds.

scolopale A "sense rod," or minute rod-like structure around the distal end of a sensory neuron.

search imagery The ability of predators to detect cryptic prey more efficiently by learning to search for potential prey of a specific visual appearance based on color, pattern, size, movement, or location within the environment, often with concurrent emphasis on search in specific habitats and reduction of the search rate.

secondary compound A toxic or digestibility-reducing compounds synthesized primarily for defense against herbivores. Not involved in the primary metabolic pathways of plants.

secondary host A term for an aphid's summer and usually herbaceous host in dioecious life cycles; in secondarily monoecious aphids it may be the host that remained after the aphid's ancestors abandoned the primary host.

secondary metabolite A plant chemical product that is not essential to the basic metabolism of the plant, e.g., phenolic glycosides. Contrast with primary metabolites such as nucleic acids, lipids, proteins, carbohydrates, or ribulose diphosphate. Secondary metabolites may or may not function as allelochemicals.

secondary sex ratio A change in the sex ratio of a family or population that occurs after fertilization but prior to reproductive maturity of the individuals involved.

secondary sexual traits Morphological and behavioral differences between the sexes that relate to structures other than the reproductive organs and gametes.

second messenger An intracellular substance, such as Ca^{2+}, cyclic AMP, inositol-1,4,5-trisphosphate, that modifies or modulates cellular responses. Its concentration changes in response to activation of G protein-coupled receptors.

segment polarity genes Genes expressed in segmental stripes in the *Drosophila* embryo; mutations in these genes cause deletions and duplications in every segment.

selectivity The property of given insecticides to selectively affect the target pest species, sparing other organisms collectively called nontarget species.

semiochemical A chemical produced by an organism that has a communication effect on another individual of the same or a different species.

semivoltine Describing a life cycle with one generation every two years.

sensillum *plural,* **sensilla**. An organ used for sensing. For olfactory and gustatory (taste) sensilla, cuticular modifications, accessory cells, and sensory cells combine to make a functional sensillum.

sensitive Referring to species for which population viability is a concern; a designation used by the U.S. Forest Service.

sentinel organism An organism that accumulates pollutants from its surroundings and/or food and is used in tissue analysis to provide an indirect estimate of prevailing environmental concentrations of these substances.

septicemic Characterized by the presence and growth of pathogenic infectious agents in the blood.

serial homology Similarities between repeated structures in different segments of an organism, caused by shared origin in development.

sericulture The industry in which mulberry plants are cultivated and used to raise silkworms, the cocoons of which are used to produce silk.

serotinous cones Cones that remain closed and on the tree for one or more years after seed maturation. High temperatures melt the resin that holds the cones closed, causing them to open rapidly.

sesquiterpenoid A substituted 15-carbon terpene derived from three isoprene units.

setae *singular,* **seta**. Slender, rigid, bristle-like hairs extending from the cuticle.

sex differentiation The genetic pathway that leads to male or female development, including regulation of somatic sexual differentiation, germ-line differentiation, and (possibly) dosage compensation.

sex role reversal A situation that occurs in species in which the sex ratio of individuals available for reproduction is female biased, such that males are the limiting sex. This leads to female competition for males and male choice of females.

sexual dimorphism The development of drastically different morphology in the male and female of a species.

sexually dimorphic Having morphological differences between males and females of the same species.

sexual selection Darwinian selection for traits—including behavioral traits—that promote success in competition for mates or the best mates (operating through discrimination of mates or direct competition). For males—the sex typically subject to the more intense sexual selection—the competition is ultimately for fertilizations.

signal Any trait borne by some individual organism, called a sender, and perceived as information by another organism, called a receiver. The receiver usually displays a signal-dependent response (otherwise the signal is meaningless).

signal system The program/format of signaling by which males and females identify and reach each other; basic systems are system I, a stationary individual broadcasts a signal, the other sex approaches, and system II, one individual broadcasts a signal, the other responds with a signal, the first approaches the second, which typically remains stationary.

silk protein A class of proteins constituting the cocoons produced by the silkworm family (Bombycidae). Silk protein is usually composed of two proteins, fibroin and sericin.

siphunculi A pair of specialized pore-bearing structures on the posterior third of the abdomen; involved in production

of aphid alarm pheromones; synonym: cornicles.

sister groups Species or monophyletic groups that arose from the stem species of a monophyletic group by a singular, identical splitting event. For example, the Lepidoptera and Trichoptera are sister groups; they shared a common ancestor that gave rise to no other lineage.

social Living in association and/or cooperation with others of the same species. The term **social insect** is applied broadly to include presocial and eusocial insect species.

solitarious In isolation; not grouped. Locusts in the **solitarious phase** do not swarm. The opposite of *gregarious*.

solubility A measure of the amount of a gas that dissolves in a fluid when exposed to a defined pressure of that gas in the air. Carbon dioxide is approximately 1000 times more soluble in water than oxygen.

somatic cells A collective term for all cells except gametes.

somatic sexual differentiation Specification during development of body parts, including male and female sex organs, and other sex-specific morphological and/or behavioral differences.

spermatheca A sperm storage organ in female insects that dispenses sperm as the eggs pass through the oviduct.

spermatogenesis The production of spermatozoa from germinal cells.

spermatophore An encapsulated package of sperm that is passed during mating and can include proteins that are eaten or absorbed by the female.

sperm competition Competition between sperm from two or more males, within a single female, for fertilization of the eggs.

sperm precedence The increased likelihood that sperm from a particular mating will be used for fertilizing an egg.

spinnerets Finger-like abdominal appendages of spiders containing numerous spigots through which silk is extruded.

spiracle An opening in the insect integument that connects the tracheal system to the air.

sporozoite Infective stage of the malaria parasite for the human host, inoculated with saliva by *Anopheles* females during blood feeding.

stadium The interval between molts.

stance The phase of the leg cycle movement in which the foot, or tarsus, is in contact with the ground and extension of the leg joints acts on the movement of the body relative to the substrate.

station-keeping Activities and movements that keep an animal within its home range or enable it to return to its home range.

stem mother See FUNDATRIX.

stenoxenic phoresy Relationship restricted to only one host genus or species.

sterile insect technique A process in which insects are reared in massive numbers, sterilized, and released to prevent normal mating in target populations.

steroid A lipid containing a 17-carbon nucleus of fused rings: three cyclohexane rings and one cyclopentane ring.

stridulating organ A file and scraper structure for sound production; may be variously located on chelicerae, palps, legs, abdomen, and carapace.

stygobite An aquatic troglobite.

stylet bundle A group of stylets (see below) that penetrates the food substrate as a functional unit.

stylets Thin, elongated mouthparts of piercing–sucking insects that penetrate into the food substrate.

subimago The winged, terrestrial instar of Ephemeroptera that molts to the reproductively mature adult.

subsocial Describing a social system in which adults protect and/or feed their own offspring for some period of time after birth, although typically the parent leaves or dies before the offspring become adults.

subterranean nest A nest below the surface of the ground.

sun compass The ability of an organism to use the sun to maintain a constant direction. This requires a "biological clock" to allow compensation for the passage of the sun across the sky.

supercooling The absence of freezing at or below the normal freezing point of water.

supercooling point The subzero temperature at which body fluids freeze, also known as the **temperature of crystallization**.

supergene A group of tightly linked genes on a chromosome, functioning as an integrated unit and segregating like a single gene. The genetic subunits usually have a more restricted phenotypic effect than the integrated supergene.

supination A classification of wings; i.e., twisting, posterior margin downward, anterior margin upward.

swarm founding The initiation and founding of a colony by a group of workers along with one or more queens.

swing The phase of the leg movement cycle in which the foot, or tarsus, is raised from the ground and repositioned for the next stance phase.

sylvatic Living or found in forests or woods.

symbiont An organism living in intimate association with another dissimilar organism. Symbiotic associations can be mutualistic, neutral, or parasitic.

symbiosis A condition in which a plant or animal is intimately associated with another organism of a different species; the living together of two taxonomically distinct organisms.

sympatric Describing two or more species inhabiting the same or overlapping geographic areas.

synanthropic Associated with humans or their dwellings; living in close association with humans.

synapomorphy (-ic) A derived state shared among the members of a monophyletic group, in contrast to a **symplesiomorphy**, a shared ancestral (plesiomorphic) state from which phylogenetic relationships cannot be inferred.

syncytial cleavage The condition of all the cleavage nuclei containing a common cytoplasm.

syndrome In genetics, an association of traits functioning together. To be adaptive and acted upon by natural selection, the traits must be influenced by the same genes (i.e., they must be genetically correlated).

synomone A chemical signal with a beneficial effect to the sender and to the receiver of the message, as in a floral scent which indicates a nectar source to a pollinator and thereby facilitates pollination.

synovigenic A form of reproduction in which an adult female continues to produce and to mature eggs throughout adult life.

systematics The study of the relationships and classifications and naming of sorts of organisms. The terms *taxonomy* and *systematics* are often used interchangeably.

T

tachygenesis Abbreviation of the developmental cycle, by suppression of one or more instars or their feeding requirements.

tagma *plural,* **tagmata**. A body region consisting of metameres grouped or fused to perform similar functions (e.g., head, thorax, abdomen).

taxa *singular,* **taxon**. A collective term for all the taxonomic groups of organisms within a higher taxonomic group (a genus, family, tribe, etc.).

taxonomic Having to do with the classification and naming of organisms.

taxonomic rank A classificatory level in the taxonomic hierarchy, e.g., species, genus, family, order. No rank is absolute, and comparisons between ranks of different organisms are inexact or even misleading; despite this, traditional ranks used for insects—notably orders and families—have useful didactic and synoptic value.

taxonomy The procedure or discipline of classifying organisms.

tegmen *plural,* **tegmina**. A parchment-like forewing characteristic of the Orthoptera and related orders. The cuticle is somewhat thickened, in association with protection.

temperature threshold The limiting (minimal or maximal) temperature at which growth and development will occur.

teneral Describing a condition in which the cuticle is soft and usually pale immediately after molting.

tentorium A bracing structure inside the head formed by invaginations of cuticle.

testis The organ in the male that produces sperm.

thelytoky Parthenogenesis in which only females are produced; the production of females from unfertilized eggs.

thermoperiod Daily temperature fluctuations.

thermosensitivity A decrease in tolerance to high temperature resulting from prior exposure to high temperature.

thermotolerance An increase in tolerance to high temperature, usually achieved by exposure to a moderately high temperature.

thigmotactic Contact orientation, as in insects that inhabit crevices or cracks.

threatened Referring to species that face a high risk of becoming endangered in the near future unless action is taken to protect them; a designation used by the U.S. Fish and Wildlife Service and the International Union for Conservation of Nature and Natural Resources.

tidal flow (of hemolymph) The reciprocal exchange of oxygen and hemolymph in some adult insects, especially where hemolymph volume is low.

trachea An air-filled branching tube that transports gases between tissues and the environment, formed by a cylinder of cells (the tracheal epithelium).

tracheal gill A body appendage with a thin cuticle and containing many trachea, used for gas exchange by many aquatic insects.

tracheal lung Thin-walled, expanded tracheae, thought to be specialized for oxygen delivery to cells lacking tracheoles such as hematocytes.

tracheole A small, blind-ended, terminal tube of the tracheal system. Each cell receives oxygen and expels carbon dioxide by means of its own tracheole.

transduction The conversion of mechanical displacement at a nerve cell membrane into a receptor current that changes the membrane potential.

transformation A process by which the genetic makeup of an organism is altered by the incorporation of foreign DNA.

transgenic Describing an organism with foreign genes incorporated into its genome by recombinant DNA techniques.

transovarial transmission The transmission of an animal disease agent from an infected female arthropod to its progeny via the egg stage.

transposable elements (TEs) Mobile segments of DNA that are able to replicate themselves and insert copies into new locations in the genome.

transstadial transmission Passage of an animal disease agent from one life stage of an infected arthropod to the next stage during the transstadial molt, e.g., from an infected larval tick to the nymphal stage.

trichomes Specialized tufts of hair on some insects that serve to dispense chemical secretions from underlying glands.

tripod gait A pattern of leg movements used by insects in which the front and rear legs on one side move in conjunction with the middle leg on the opposite side, forming a stable tripod of support. The remaining three legs also form a tripod, and motion occurs when the two sets alternate their motions.

triungulin The campodeiform first instar of Meloidae, Rhipiphoridae, and Micromalthidae that molts to a less active, feeding second instar. Unlike later instars, triungulins have well-developed legs and move around actively.

troglobite A species obligately adapted to live only in subterranean habitats and unable to survive in surface environments.

troglomorphic Relating to the suite of morphological, physiological, and behavioral adaptations that are characteristic of troglobites.

troglophile A species able to live and reproduce in caves but also able to survive in surface habitats.

trogloxene A species habitually roosting or regularly visiting caves for food or shelter but unable to complete its life cycle underground.

trophallaxis The exchange by oral transfer of alimentary liquid among members of a colony of social insects. Typically, this liquid serves to feed and unite members of a colony.

trophic cascade An interaction of trophic levels in a community, such that changes in density of a species at one level have multiple influences on the density of species at other trophic levels.

trophic cells Specialized cells containing a reserve of fat or other nutritive substance.

trophic egg An egg, usually degenerate in form and inviable, that is fed to members of the colony.

trophic level A functional classification of organisms in a community based on what they feed on. The first trophic level consists of primary producers, mostly green plants that obtain energy from the sun. The second trophic level consists of herbivores, organisms that feed on plants. The third trophic level consists of carnivores that feed on the herbivores, and so on.

trypanosome A type of protozoan that has a leaf-like motile stage.

tubular body A microtubule-based structure found in the distal sensory dendrites of many mechanoreceptor neurons.

tychoparthenogenesis The rare or occasional production of eggs that begin developing without having been fertilized.

U

ultraselfish Describing a strategy used by symbionts whose spread and maintenance occur despite and because they cause damage to the individual in which they occur.

univoltine Describing species that have a single generation per year. Those that have two or more are called *multivoltine*.

unprofitability The character of a prey that gives no net reward to the predator once consumed, leading to learned or evolved avoidance. Examples are toxicity, toughness, difficult handling, and difficult/costly capture.

urocytes Cells that accumulate uric acid either as a form of nitrogen waste product or as reserves to be used as a nitrogen source by the mycetocytes.

urogomphi Sclerotized, paired dorsal processes that project from the posterior margin of the ninth larval tergite.

USP Ultraspiracle, a nuclear receptor that forms a dimer with the ecdysteroid receptor to regulate gene expression.

V

vannus The expanded fan-like area of some hind wings, posterior to the claval furrow.

vas deferens The sperm duct leading away from the testis.

vector Any agent that transports a microorganism from one host to another.

vectorial capacity The epidemiological efficiency of *Anopheles* host species in transmitting malaria parasites, expressed as new infections per infection per day, based on mathematical relationships among the mosquito biological characteristics of daily survival, blood meal host-feeding pattern and frequency, and susceptibility to parasite infection.

veins The cuticular, usually tubular, rods that support the wing.

venation The pattern of veins within a wing.

ventral diaphragm A complete or fenestrated membrane always associated with the ventral nerve cord of insects and defining a perivisceral sinus when present.

veriform Describing an insect larva that is worm shaped.

vertical transmission The transmission of genetic elements from parents to progeny.

vesicle An extensible organ.

vitelline membrane The innermost layer of eggshell secreted by the follicle cells; sometimes considered to be the first layer of the chorion.

vitellogenic cycle A cycle of oocyte development associated with the synthesis of **vitellogenin**, a protein that contributes most of the protein in yolk. In some insects, such as mosquitoes, **vitellogenesis** (vitellogenin production) is limited by lack of protein and occurs only following a highly proteinaceous meal, such as blood.

vivipara The parthenogenetic winged or wingless morphs that bear live young; synonym *virginopara*.

viviparity Reproduction by giving birth to live offspring rather than eggs.

voltage-gated sodium channel A large transmembrane protein that regulates the flow of sodium ions across axonal membranes and mediates the rising phase of action potentials.

vulnerable Facing a risk of extinction in the medium-term future; a designation used by the International Union for Conservation of Nature and Natural Resources.

W

Wallace's Line A term for the junction between the Oriental and the Australian zoogeographical regions at the straits between Bali and Lombok.

wasp waist The strong constriction between the first and second abdominal segments, present in virtually all members of the Apocrita. The thorax and first abdominal segment in these insects comprise the **mesosoma**, and the rest of the abdomen, the **metasoma**.

wild type An organism that has no visible mutant phenotype.

X

xylem Aplant conductive tissue interior to the cambial layer, composed of several cell types, which transports water from the roots to other sites in the plant and also serves as structural support.

xylophilous Preferring wood.

Y

yolk An accumulation of protein, lipid, and glycogen particles in the cytoplasm of an egg. The particles are assembled during vitellogenesis and consumed during embryonic development.

Z

zoonosis *plural,* **zoonoses**. A disease of vertebrates communicable from animal to animal and from animal to human.

zoonotic Transmitting disease between vertebrate animals and between animals and humans.

COOPERATIVE EXTENSION SERVICE
UNIVERSITY OF KENTUCKY COLLEGE OF AGRICULTURE, FOOD AND ENVIRONMENT

Plant Pathology & Forestry Fact Sheet

PPFS-GEN-14
FORFS-16-01

Don't Eat Those Wild Mushrooms

...unless you know what you are doing!

Ellen Crocker, *Postdoctoral Scholar, Department of Forestry*
Nicole Ward Gauthier, *Extension Plant Pathologist, Department of Plant Pathology*

Mushrooms are strange and wonderful things – some are beautiful, some are ugly, some are delicious, and some are deadly. Mushroom hunting is a fun and rewarding hobby that can turn a hike through local woods into a puzzle-solving adventure. Many people are drawn to mushroom hunting and the potential to forage for food. Unfortunately, there is a dark side to mushroom foraging: poisoning. Each year, wild mushrooms lead to numerous illnesses and even a few deaths.

While the threat of mushroom poisoning is real, most mushroom-caused illnesses can be prevented if collectors are well-informed and cautious. This fact sheet provides introductory information regarding mushroom safety.

When Are Mushrooms Dangerous to Consume?

Poisonous mushrooms

Some mushrooms contain toxins that directly cause poisoning. Different mushroom toxins can cause different symptoms—ranging from mild stomach distress to liver failure and death. A few mushrooms, like the appropriately named Destroying Angel or Death Cap (both *Amanita* spp.)(Figure 1) contain potentially lethal toxins and kill a few Americans every year.

Figure 1. Two poisonous mushrooms that can cause liver failure and death, the Destroying Angel (A) and Death Cap (B).

Several other mushroom toxins can cause an intense upset stomach. One example is the Jack-O'-Lantern mushroom (*Omphalotus illudens*), which is sometimes confused with the edible Chanterelle (*Cantharellus* spp.). Despite its similar appearance and appealing odor, consumption of the Jack-O'-Lantern results in stomach pain, vomiting, and diarrhea. This is why it's important to carefully identify each mushroom, rather than relying upon superficial appearances.

Environmental toxins

Mushrooms can also absorb toxins from the environment, so it is important to know the history of the site from which specimens were collected. If site history is unknown (e.g., lawn that may have been recently treated with pesticides), do not consume mushrooms from the area.

Preparation, interactions, and allergies

Even mushrooms that don't contain toxins can sometimes cause illness. Many mushrooms, like Morels and some that are commonly found in grocery stores, must be properly cooked to avoid illness. For example, popular Shiitake mushrooms (*Lentinula edodes*) can cause a severe rash in people who are allergic to it in a raw or poorly-cooked state.

Compounds in some mushrooms may also react with other foods or drugs. For example, some Inky Caps (*Coprinopsis* spp.) are tasty but poisonous when combined with alcohol. When consumed up to 5 days before drinking alcohol, these mushrooms can cause sweating, nausea, vomiting, and even heart problems. Other mushrooms are also known to adversely react with various medications, drugs, and alcohol.

Poorly stored mushrooms

Surprisingly, most illnesses associated with consumption of edible mushrooms are a result of microbial spoilage or decaying mushrooms. Bacteria can contaminate mushrooms during collection, transportation, or storage. See safety tip 6, below, to avoid illness from contaminants in poorly stored mushrooms.

7 Tips for Safe Mushroom Collecting

1 Learn to identify mushrooms

Like driving a car, mushroom foraging can be dangerous without training, patience, and practice. Those who are interested in mushroom foraging should first learn to accurately identify mushrooms (Figure 2). Poisonous mushrooms can look similar

Figure 2. Knowing the parts of a mushroom and their characteristics are critical to proper identification.

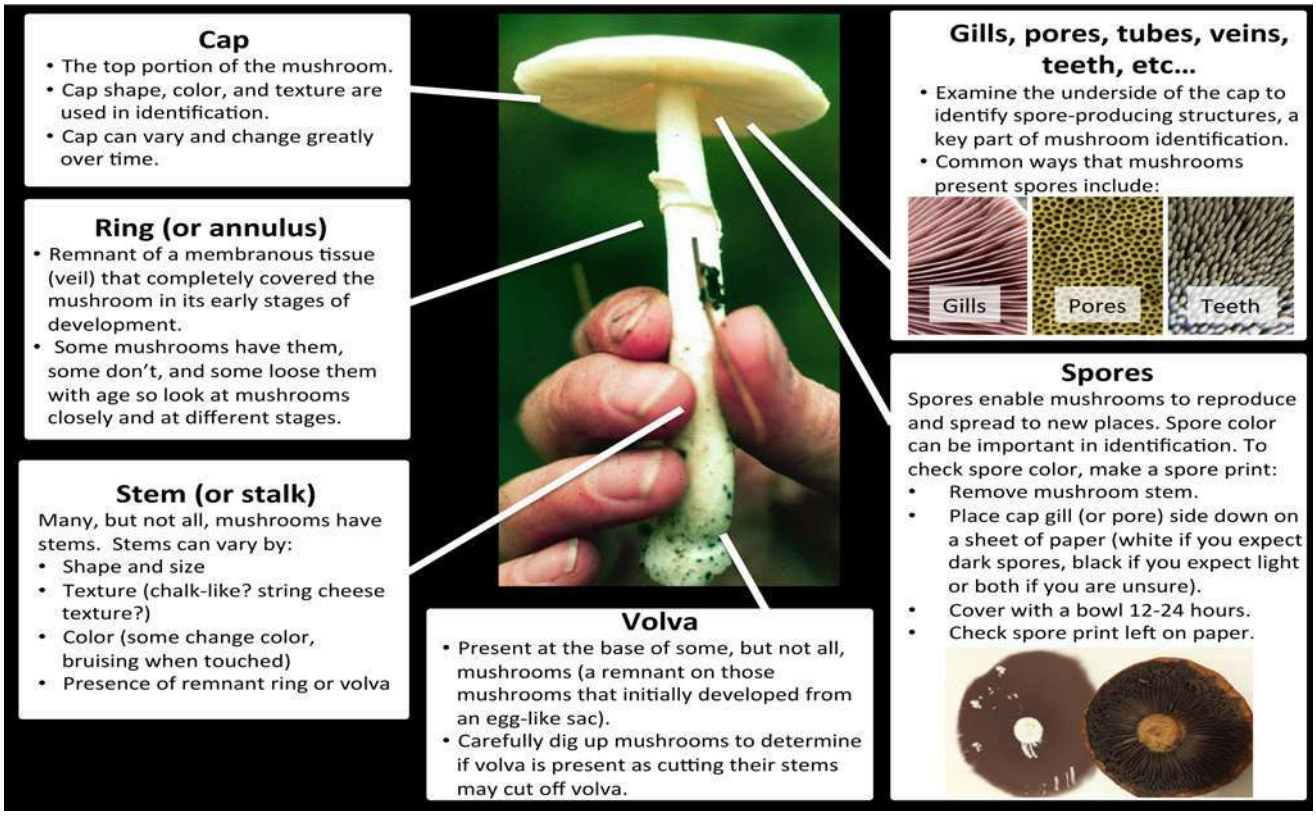

to edible mushrooms at certain life stages, so careful observation and identification of each mushroom is critical. Train with a plant pathologist or mycologist, and refer to reputable identification guides. Do not hesitate to ask for help or to confirm identification with experts. Be sure to also learn the poisonous species in the area while learning the edible ones. A great way to learn mushrooms is to join a local mushroom club. Note: if an expert identifies a sample as an edible species, it does not necessarily mean that others growing nearby are the same species.

2 Identify ALL the mushrooms to be eaten

The same mushroom species can look different at different life stages or when growing in different sites. Thus, it is important to identify each mushroom each time any are collected. Once the key characteristics of a favorite mushroom are known, foragers should look for those identifiable features during each collection.

3 Start small

When experimenting with a new mushroom, eat only a little in case of adverse reaction. Wait a few days before eating more. Research the preparation and safety of each mushroom species, as well as food and drug interactions. In addition, always store a few samples of mushrooms that have been eaten in the refrigerator for professional identification…in case there is a need for medical attention.

4 Know mushroom look-alikes

While mushroom identification is notoriously difficult, the good news is that it is relatively easy to avoid the most dangerous mushrooms. Collect and consume only those mushrooms that have been positively identified as edible, and know all of the similar but poisonous look-alikes (Figure 3). This greatly minimizes risk of poisoning.

5 Do not trust folklore

Many common hunting anecdotes are dangerously incorrect. Do not trust folklore regarding identification of poisonous mushrooms. For example, **not all** poisonous mushrooms have pointed caps, and **not all** all white mushrooms are safe to eat! Unfortunately, there are no shortcuts to determining whether a mushroom is poisonous. Science-based information and experience are critical.

6 Store mushrooms properly

Whether storing supermarket mushrooms or those collected from the wild, take these steps to reduce contaminants:

- Cut off or clean any dirty portions (once the mushroom is identified as edible). Stalk bases and dirty (soiled) portions may contain bacteria that can grow during transport or storage.
- Store mushrooms in paper (not plastic) bags. Plastic bags create humid conditions where bacteria and other contaminants thrive.
- Keep mushrooms cool; refrigerate immediately after collecting.
- Before preparing mushrooms for consumption, closely inspect and confirm that they are clean and fresh. Discard old mushrooms and any that appear damaged, wilted, or slimy.

7 When in doubt, throw it out

As the old saying goes: "There are old mushroom hunters and there are bold mushroom hunters. But there are no old, bold mushroom hunters." When in doubt, don't risk it—remember that delicious edibles can be purchased from many grocery stores, and questionable mushrooms are not worth the potential consequences or the worry.

If A Potentially Dangerous Mushroom Has Been Eaten…

SEEK MEDICAL ATTENTION IMMEDIATELY!

If safety tip 3, above, was followed, some small samples have been saved for identification. This can make the difference between life and death. For example, without medical attention, the survival rate after consumption of toxic *Amanita* mushrooms is approximately 50 percent; but with medical treatment, survival rates increase to around 90 percent.

Follow the Law While Collecting

Remember to only collect mushrooms where permitted to do so. Check with parks and public areas regarding regulations and apply for permits, if needed. Do not hunt on private property without permission.

Chanterelle v. Jack-O'-Lantern?

Traits	Chanterelle	Jack-O'-Lantern (poisonous)
Gills?	Close but not true gills. Instead, has smooth gill-like ridges (look melted and are more difficult to remove from cap than normal gills) under the cap to the stem.	Yes. These mushrooms have sharp and non-forking gills that go down the stem of the mushroom a little from the cap.
Growth pattern?	• Typically solitary or in pairs. • Typically growing out of soil.	• Typically in big clusters. • Typically out of woody material (tree base or roots).
Color in stem?	Cut stem is usually white inside.	Cut stem is usually orange inside.

Puffball v. young Amanita?

Traits	Puffball	Amanita (some very poisonous)
Form?	A ball shape (may be oblong or miss-shaped, can vary widely in appearance, even having a stem or stalk.	Mature mushroom has a cap and stalk, but at early stage covered in sack and appears egg-like.
Slice it in half? (top to bottom)	• Inside you will only find flesh or spores, depending on how old the puffball is. (Only eat them when young, no spores). • If it is gooey inside, it might be a slime mold or immature stinkhorn, not for eating.	If the egg-like button stage is cut in half, a mushroom form is evident. 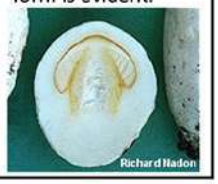

Shaggy Mane or Shaggy Parasol v. Green-Spored Parasol?

Traits	Shaggy Mane or Shaggy Parasol	Green-Spored Parasol (poisonous)
Spores?	• Black (Shaggy Mane) or white (Shaggy Parasol) spores.* *Many other mushrooms have white or black spores. On its own, this information is not diagnostic, just helpful for distinguishing from the Green-Spored Parasol.	• Green (make a spore print since the spores may not be evident by examining the gills of young mushrooms, which are white but color with age).

Morel v. False Morel?

Traits	Morel	False Morel (poisonous)
Caps?	• Typically uniform shape (cap usually longer than stem). • Inward pits and ridges. • Cap attaches at base to stem.	• Can have an irregular shape or bulge outwards. • Wavy, lobed, or wrinkled (brain-like). • Cap attached at top but not around stem.
Inside? (cut open top to bottom)	• Hollow from top (cap) to bottom (stem), looking like a rubber mold.	• Stem has cottony fibers inside. • Parts may be hollow (air pockets) but typically not symmetrical.

FIGURE 3. BEFORE FORAGING FOR MUSHROOMS, KNOW THE LOOK-ALIKES OF COMMON EDIBLES AND THE KEY TRAITS THAT DISTINGUISH THE TWO.

Additional Resources

On the Internet

- Field Guide to Common Macrofungi in Eastern Forests and Their Ecosystem Function, GTR NRS-79 (USDA)
http://www.fs.fed.us/nrs/pubs/gtr/gtr_nrs79.pdf

- Mushrooming (Missouri Department of Conservation)
http://mdc.mo.gov/discover-nature/outdoor-recreation/mushrooming

- MycoKey – The Mycological Identification Site
http://www.mycokey.com/newMycoKeySite/MycoKeyIdentQuick.html

- MushroomExpert.com
http://mushroomexpert.com/

- North American Mycological Society
http://www.namyco.org/
Note: includes a list of regional mushroom clubs

- Various online forums and groups for mushroom identification, for example,
http://mushroomobserver.org/
Note: Be very careful with these sites; as with all social media, there is a risk of unsolicited commentary and misinformation. However, many of these groups are filled with knowledgeable and enthusiastic mushroom hunters. While these groups and forums can be helpful tools, they should not be an exclusive source of identification. information.

In print

- *A Field Guide to Southern Mushrooms*. Nancy Smith Weber and Alexander H. Smith. 1988. The University of Michigan Press: Ann Arbor, Michigan. 280pp.

- *Field Guide to Mushrooms*. G. Lincoff. National Audubon Society. 1981. Knopf Publishers. 928 pp.

- *Mushrooms Demystified*. David Arora. 1986. Ten Speed Press: Berkeley, CA. 958 pp.

- *Mushrooms of Northeastern North America*. Bessette, Bessette and Fischer. 1997. Syracuse University Press. 582 pp.

- *Mushrooms of West Virginia and the Central Appalachians*. William C. Roody. 2003. University Press of Kentucky: Lexington, KY. 536 pp.

- *North American Mushrooms: A Field Guide to Edible and Inedible Fungi*. O. Miller and H. Miller. 2006. Falcon. 592p.

February 2016

Acknowledgements

Thanks to Kathie Hodge (Associate Professor, Plant Pathology and Plant-Microbe Biology, Cornell University) and Chad Niman (Primary Forest Products Specialist, KY Forestry Extension, University of Kentucky) for helpful comments on drafts of this article.

Photos credits--Kathie Hodge, Cornell University (1a); Ellen Crocker, University of Kentucky (2-gills, teeth & spore print); Richard Nadon, mushroomexpert.com (3-young Amanita cross-section); Bugwood.org: Norman D. Davis (1b), Joseph O'Brien, USDA-Forest Service (1a, 2-pores, 3-Amanita), USDA Forest Service, Northeastern Area (2-Amanita, 3-Chanterelle), Robert L. Anderson, USDA Forest Service (3-jack o' lantern), Curtis E. Young, The Ohio State University (3-puffball, 3-green-spored parasol spores), Joseph LaForest, University of Georgia (3-Amanita), George Hudler, Cornell University (3-shaggy mane), Gerald Holmes, California Polytechnic State University at San Luis Obispo (3-green-spored parasol), Chris Evans, University of Illinois (3-morel), Dave Powell, USDA-Forest Service, retired (3-false morel)

Educational programs of the Kentucky Cooperative Extension Service serve all people regardless of race, color, age, sex, religion, disability, or national origin.

United States
Department of
Agriculture

Forest Service

**Northern
Research Station**

General Technical
Report NRS-79

Field Guide to Common Macrofungi in Eastern Forests and Their Ecosystem Functions

Michael E. Ostry
Neil A. Anderson
Joseph G. O'Brien

INTRODUCTION: ABOUT THIS GUIDE

This guide is intended to serve as a quick reference to selected, common macrofungi (fungi with large fruit bodies such as mushrooms, brackets, or conks) frequently encountered in four broad forest ecosystems in the Midwest and Northeast: aspen-birch, northern hardwoods, lowland conifers, and upland conifers. Although these fungi are most common in the ecosystems we list them in, many can be found associated with tree species in multiple ecosystems. We provide brief identifying characteristics of the selected mushrooms to allow you to identify some down to the species level and others to the genus or group to which they belong. Former scientific names are provided in parentheses. Also included in each mushroom description are details about its ecosystem function, season of fruiting, edibility, and other characteristics.

Although we provide information about edibility in this guide, DO NOT eat any mushroom unless you are absolutely certain of its identity: many mushroom species look alike and some species are highly poisonous. Many mushrooms can be identified only by examining the color of spore prints or by examining spores and tissues under a microscope. As mushrooms age, changes in their shape, color, and general appearance make it necessary to examine several individuals for their distinguishing features.

For additional information on other species of macrofungi, serious mushroom hunters may wish to consult any of the excellent illustrated guides and detailed keys available (see Suggested References at the end of this guide). Several useful mycological Web sites with images and descriptions of fungi are available and a few of these are also listed.

Mushroom Basics

Fungi are important organisms that serve many vital functions in forest ecosystems including decomposition (Fig. 1), nutrient cycling, symbiotic relationships with trees and other plants, biological control of other fungi, and as the causal agents of diseases in plants and animals. Mushrooms are sources of food for wildlife (Figs. 2, 3), and fungi that cause decay in living trees are beneficial to many species of birds and mammals (Figs. 4, 5). Less than 5 percent of the estimated 1.5 million species of fungi have been described, and their exact roles and interactions in ecosystems are largely unknown.

Mike Ostry. U.S. Forest Service

Figure 1.—Mossy Maze Polypore (*Cerrena unicolor* [*Daedalea unicolor*]). Wood decay fungi are critical in nutrient cycling and increasing soil fertility.

Figure 2.—Hollow Stem Larch Suillus (*Suillus cavipes*).

Mike Ostry. U.S. Forest Service

Figure 3.—Emetic Russula (*Russula emetica*) stored on branches of black spruce by squirrels.

Macrofungi are distinguished from other fungi by their fruiting structures (fruit bodies bearing spores) that we know as mushrooms. Mushrooms with gills, the most common, produce spores that range from white to pink and shades of yellow to brown to black. Most mushrooms produce spores on gills that increase the spore-bearing surface on the underside of the cap. Other mushrooms, such as the Boletes, produce their spores in elongated tubes, and the hedgehog mushrooms produce spores on elongated spines.

Most of the fungus biomass consists of the largely unseen mass of interwoven threadlike hyphae growing in plant tissues and in the soil. Annual variation in the timing and production of the aboveground mushrooms is largely influenced by temperature and precipitation.

The most commonly encountered macrofungi in our woodlands throughout the year are the wood-decaying bracket and conk fungi. These fungi, found on the stems of dead and living trees, produce their spores in small, rigid tubes in leathery-woody fruit bodies that are annual or perennial. The perennial species produce a new layer of tubes to the enlarging fruit body each year.

Mike Ostry. U.S. Forest Service

Figure 4.—Signs of woodpecker activity on aspen decayed by True Tinder Conk (*Fomes fomentarius*) and other decay fungi.

Mike Ostry. U.S. Forest Service

Figure 5.—Cavity in maple decayed by Mossy Maple Polypore (*Oxyporous populinus* [*Fomes connatus*]) used by a squirrel to cache acorns.

Mushroom species form new clones when two compatible spores of the same species germinate and grow together. However, most mushroom spores are dispersed, germinate, and contribute genetic variation to established clones in soil and wood. In nature, many mushrooms and bracket fungi may look alike, but they do not interbreed and thus are distinct biological species. Their growth on different hosts or physical separation from each other over time has made them genetically incompatible.

The parts of a mushroom important for identifying groups and species of fungi are shown in Figure 6. Species of *Amanita* are common, and some are deadly poisonous. Because they possess key identifying parts, we use an *Amanita* to illustrate the key structures of a mushroom.

Young mushrooms are called buttons or the egg stage and contain the preformed cap and stalk. As the mushroom grows, the cap breaks through the egg's universal veil, the stalk elongates and the cap expands

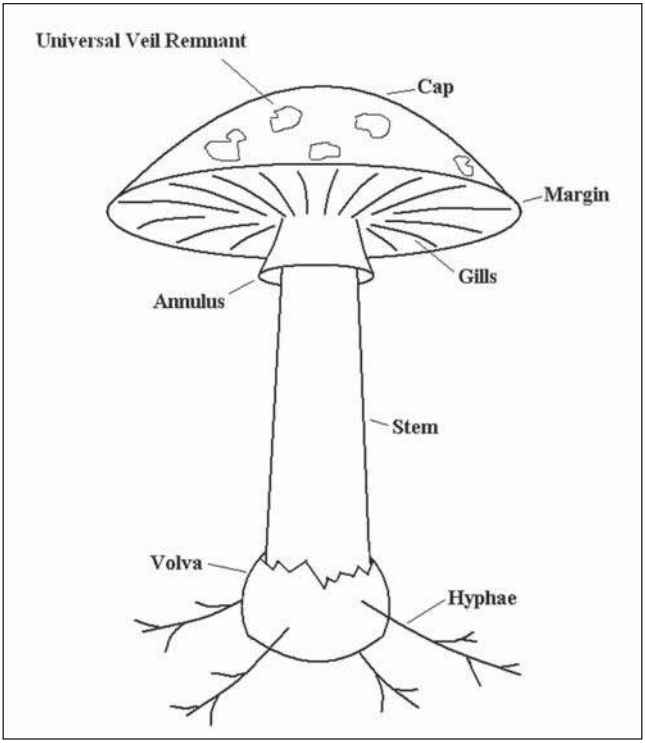

Figure 6.—Identifying parts of an *Amanita* mushroom.
Drawing by Melanie Moore, U.S. Forest Service.

like an umbrella. The secondary veil protecting the gills and spores is broken by the expanding cap, and remnants of this veil form a ring (annulus) on the stem, also referred to as the stipe or stalk. *Amanita* mushrooms also have a cup (volva) at the base of the stalk, often within the soil layer. **Therefore, mushrooms should always be dug, not picked, in order to detect this cup feature of a potentially poisonous mushroom.** In addition to the ring on the stalk and the basal cup, white gills that are free from the stalk and a white spore print distinguish *Amanita* mushrooms from other species.

The fungi illustrated in this guide serve critical ecological functions, and their roles as symbionts, in litter and wood decay, and as pathogens are described. An important beneficial function of many

macrofungi is the relationship with forest tree roots in the uptake of nutrients and water and in the protection of the tree roots from pathogenic fungi and nematodes. Strands (hyphae) of the fungus form a dense layer (mantle) around the fine roots of trees and extend out into the surrounding soil. This root-fungus association is called mycorrhizae and benefits both the fungus and the tree.

Pathogenic fungi such as the root and butt rot fungi illustrated in this guide can be damaging, but they also provide important ecological services through nutrient cycling and development of forest structure and wildlife habitat. Distinguishing the potential positive effects from the negative effects of these fungi will enable woodland managers and owners to make informed management decisions based on their objectives.

ASPEN-BIRCH ECOSYSTEM

Mike Ostry, U.S. Forest Service

Aspen-birch

Fly Agaric *Amanita muscaria*

Identification: Cap yellow to orange with white scales that are remnants of the universal veil; white gills free from stalk; white veil; volva (cup) consisting of 2-3 scaly rings on stalk above bulbous base

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with hardwoods and conifers

Edibility: Poisonous

Fungal note: This fungus forms fairy rings that grow radially 3-5 inches every year.

Mike Ostry, U.S. Forest Service

Amanita muscaria

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Destroying Angel *Amanita virosa, A. verna, A. bisporigera*

Identification: Cap white, smooth; white gills free from stalk; bulbous base; white veil

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with hardwoods and conifers

Edibility: Highly poisonous and often fatal

Fungal note: These three mushrooms can only be distinguished from each other by their spore characteristics; collectively, they cause 95 percent of fatal mushroom poisonings.

Mike Ostry, U.S. Forest Service

Amanita virosa

The Omnipresent Laccaria *Laccaria bicolor*

Identification: Cap colors vary from yellow to buff to orange to lilac, waxy, fibrous; stalk often twisted

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with aspen, spruce, and pine of all ages

Edibility: Good

Fungal note: This is one of the most common mushrooms on upland sites and was the first mycorrhizal fungus to have its entire genome sequenced. *Laccaria longipes* is common with black spruce in bogs.

Neil A. Anderson, University of Minnesota; used with permission

Laccaria bicolor

Mike Ostry, U.S. Forest Service

Laccaria longipes

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Aspen Bolete — *Leccinum aurantiacum, L. insigne*

Identification: Cap red-brown, dry; flesh white turning red then blue-gray when bruised; stalk with brown-black scales called scabers

Season of fruiting: Late summer-fall

Ecosystem function: Mycorrhizal with aspen

Edibility: Edible

Fungal note: Genera of boletes are distinguished from each other by spore color, linear or random arrangement of the tubes on their lower surface, and type of ornamentation on their stalks.

Mike Ostry, U.S. Forest Service

Leccinum aurantiacum

Birch Bolete *Leccinum scabrum*

Identification: Cap gray-brown to yellow-brown; flesh white, not staining when bruised; stalk with brown-black scales called scabers

Season of fruiting: Late summer-fall

Ecosystem function: Mycorrhizal with birch

Edibility: Edible

Fungal note: Boletes are among the most sought after edible mushrooms and are ecologically important as tree symbionts.

Joseph O'Brien, U.S. Forest Service

Leccinum scabrum

Oyster Mushroom *Pleurotus populinus* (*P. ostreatus*)

Identification: Cap white-pale tan; stem usually lateral or absent; gills white and run down the stem; spore print white; found only on aspen

Season of fruiting: Summer-fall

Ecosystem function: Sapwood rotter

Edibility: Choice

Fungal note: Three closely related species are known: *P. populinus* is found on aspen; *P. pulmonarius* (*P. sapidus*) is found on hardwoods other than aspen and has a lilac color spore print; *P. ostreatus* has a white spore print and is also found on hardwoods other than aspen, often in riparian areas.

Mike Ostry, U.S. Forest Service

Pleurotus populinus

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Artist's Conk — *Ganoderma applanatum* (*Fomes applanatus*)

Identification: Shelf-like, hard gray-brown zonate upper surface; white lower surface that turns brown when scratched

Season of fruiting: Perennial

Ecosystem function: Causes a white stem and butt rot of hardwoods

Edibility: Inedible

Fungal note: The most common perennial wood decay fungus of dead and dying hardwood trees. A single conk can produce 1.25 billion spores each hour for 5-6 months each year.

Mike Ostry, U.S. Forest Service

Ganoderma applanatum

DO NOT eat any mushroom unless you are absolutely certain of its identity.

False Tinder Conk — *Phellinus tremulae* (*Fomes ignarius*)

Identification: Hoof-shaped, gray-black hard conk with a brown margin

Season of fruiting: Perennial

Ecosystem function: Causes a white trunk rot of aspen

Edibility: Inedible

Fungal note: This fungus causes more wood volume loss than any other aspen pathogen; however, the resulting soft wood of affected stems is beneficial for cavity-nesting wildlife. On average, decay extends 8 feet above and 5 feet below an individual conk.

Mike Ostry, U.S. Forest Service

Phellinus tremulae

Mike Ostry, U.S. Forest Service

Cross section of aspen stem near a conk of False Tinder Conk (*P. tremulae*) revealing a column of soft, decayed wood that benefits cavity-nesting birds and animals.

Mike Ostry, U.S. Forest Service

Cavities excavated by woodpeckers in aspen affected by *P. tremulae*.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

True Tinder Conk *Fomes fomentarius*

Identification: Hoof-shaped, gray, hard conk

Season of fruiting: Perennial

Ecosystem function: Causes a wood rot, common on dead birch

Edibility: Inedible

Fungal note: The felt-like inner layer makes excellent tinder. This material, called "amadou," has also been used as a substitute for matches after soaking it in solutions of potassium or sodium nitrate and then drying it.

Mike Ostry, U.S. Forest Service

Fomes fomentarius

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Birch Polypore *Piptoporus betulinus* (*Polyporus betulinus*)

Identification: Circular, round, shelf-like, white to brown

Season of fruiting: Annual

Ecosystem function: Causes a brown cubical wood rot, common on dead birch trees

Edibility: Tough, inedible unless very young

Fungal note: The inner material of the conk can be used as fire tinder when dry.

Mike Ostry, U.S. Forest Service

Piptoporus betulinus

Multicolor Gill Polypore *Lenzites betulina*

Identification: Fruit body leathery, hairy with alternating bands of gray, yellow, and brown; undulating, gray gills

Season of fruiting: Summer-fall

Ecosystem function: White sapwood rot of dead birch and other hardwoods

Edibility: Inedible

Fungal note: Fruit bodies are white when young, turning gray with age, often with green algae on the surface.

Joseph O'Brien, U.S. Forest Service

Lenzites betulina, bottom view.

Morel (Sponge Mushroom) *Morchella esculenta*

Identification: Cap resembles an inverted pine cone with ridges and deep pits, gray-cream-yellow; stem white-cream and hollow

Season of fruiting: Brief (2-3 weeks) in spring

Ecosystem function: Litter and wood decay; found on the ground among aspen and many other hardwood species, spruce, and pine of all ages

Edibility: Choice

Fungal note: One of the most sought after edible mushrooms.

Neil A. Anderson, University of Minnesota; used with permission

Morchella esculenta

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Honey Mushroom *Armillaria gallica*

Identification: Cap tan to golden yellow; prominent ring on stem; white spore print; black "shoestring" cords (rhizomorphs) that transport food to growing hyphae

Season of fruiting: Fall

Ecosystem function: Root and butt rot capable of killing trees, especially stressed trees, creating root rot pockets resulting in canopy gaps

Edibility: Choice

Fungal note: An individual clone of this fungus, 15.4 ha in size and estimated to be 1,500 years old, was identified in northern Michigan. The mushroom *Entoloma abortivum* parasitizes fruit bodies of *Armillaria* turning them into misshapen Abortive *Entoloma* mushrooms.

Joseph O'Brien, U.S. Forest Service

Armillaria gallica

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Joseph O'Brien, U.S. Forest Service

Armillaria sp. with characteristic ring (annulus) on the stems.

Joseph O'Brien, U.S. Forest Service

Abortive *Entoloma* fruit body resulting from *Armillaria* mushrooms parasitized by *Entoloma abortivum*.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

NORTHERN HARDWOOD ECOSYSTEM

Joseph O'Brien, U.S. Forest Service

Northern hardwoods

Giant Puffball *Calvatia gigantea*

Identification: Softball-soccer ball in size; white leathery skin when young turning yellow-tan when mature

Season of fruiting: Late summer-fall

Ecosystem function: Mycorrhizal

Edibility: Edible when young

Fungal note: Giant puffballs 30.5 cm in diameter can produce 7 trillion or more spores that are perfectly adapted to wind dissemination. In calm air, spores fall at a rate of 0.5 mm per second.

Joseph O'Brien, U.S. Forest Service

Calvatia gigantea

Mike Ostry, U.S. Forest Service

Calvatia gigantea

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Bear's Head Tooth *Hericium coralloides*

Identification: From a single stem, the fruit body branches into clusters of snow-white spines that point down and bear the spores of the fungus on their outer surface. Spines darken to yellow or brown with age.

Season of fruiting: Late summer-fall

Ecosystem function: Decay of hardwood logs

Edibility: Choice

Fungal note: This fungus can be pickled, marinated, or fried.

Joseph O'Brien, U.S. Forest Service

Hericium coralloides

Scaly Pholiota *Pholiota squarrosa*

Identification: Cap dry, yellow-pale tan with brown scales; gills yellow-light brown, brown spores; stalk with a veil forming a ring, scales present below but not above ring

Season of fruiting: Summer-fall

Ecosystem function: Wood rotter of hardwoods and conifers

Edibility: Not recommended

Fungal note: Common butt rotter of living aspen and birch as well as down aspen logs. Often found in large clusters. *P. squarrosoides* is another very similar *Pholiota* species that is frequently found.

Mike Ostry, U.S. Forest Service

Pholiota squarrosa

Milk-White Toothed Polypore *Irpex lacteus* (*Polyporus tulipiferae*)

Identification: White, crust-like, flat to substrate, pores breaking into teeth

Season of fruiting: Spring-fall

Ecosystem function: White rot of hardwoods

Edibility: Inedible

Fungal note: This fungus is very common on dead branches of hardwood trees.

Joseph O'Brien, U.S. Forest Service

Irpex lacteus This specimen has discolored to yellow-brown with age.

Violet Polypore *Trichaptum biforme* (*Hirschioporus pargamenus, Polyporus pargamenus*)

Identification: Fruit bodies thin, leathery, with zones of various colors and a violet pore surface only on the fruit body margin that breaks into teeth with age; often covering large areas of dead trees.

Season of fruiting: Spring-fall

Ecosystem function: White pocket rot of hardwoods, very common on dead aspen; a very similar species, *T. abietinum*, occurs on conifers

Edibility: Inedible

Fungal note: One of the most common decay fungi in the U.S. The sporocarps are often covered with green algal growth.

Joseph O'Brien, U.S. Forest Service

Trichaptum biforme

Joseph O'Brien, U.S. Forest Service

Trichaptum biforme, upper and lower surface

Mike Ostry, U.S. Forest Service

Early growth form of Violet Polypore (*Trichaptum biforme*) on the lower surface of a fallen aspen stem.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Smoky Polypore — *Bjerkandera adusta* (*Polyporus adustus*)

Identification: Clusters, small, white to grayish, velvety caps, pore surface gray to black

Season of fruiting: Spring-fall

Ecosystem function: White sapwood rot of dead hardwood trees

Edibility: Inedible

Fungal note: Fruit bodies can revive after long periods of drought. Pores of fruit body are very small (5-7 pores per millimeter).

Mike Ostry, U.S. Forest Service

Bjerkandera adusta

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Mike Ostry, U.S. Forest Service

Bjerkandera adusta

Mike Ostry, U.S. Forest Service

Bjerkandera adusta

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Common Split Gill *Schizophyllum commune*

Identification: Clusters of leathery, whitish gray, fan-shaped gilled fruit bodies

Season of fruiting: Perennial

Ecosystem function: White sapwood rot of living and dead hardwood trees

Edibility: Inedible

Fungal note: Spores of this fungus were obtained from fruit bodies after 50 years of dry storage. Each gill is split into two halves that curl in dry weather to protect the spore-bearing surface.

Mike Ostry, U.S. Forest Service

Schizophyllum commune

Mike Ostry, U.S. Forest Service

Schizophyllum commune, lower gill surface.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Hen of the Woods *Grifola frondosa* (*Polyporus frondosus*)

Identification: Large, dull white to gray, solitary fruit bodies with overlapping shelves on the ground near stumps or at the base of living hardwood trees

Season of fruiting: Late summer-fall

Ecosystem function: White butt rot of hardwoods

Edibility: Choice

Fungal note: Always found growing on the ground.

Neil A. Anderson, University of Minnesota; used with permission

Grifola frondosa

Maze Bracket *Daedalea quercina*

Identification: Gray to light brown, leathery, shelf with mazelike lower surface

Season of fruiting: Spring-fall

Ecosystem function: Brown heart rot of oaks

Edibility: Inedible

Fungal note: This fungus is not found west of the Mississippi River.

Joseph O'Brien, U.S. Forest Service

Daedalea quercina, bottom view.

Annual Shelf Fungus — *Phellinus gilvus* (*Polyporus gilvus*)

Identification: Leathery, yellow to brown shelf, yellow-brown interior

Season of fruiting: Summer-fall

Ecosystem function: White sapwood decay and occasionally heart rot

Edibility: Inedible

Fungal note: Common on red oak and other hardwood trees.

Neil A. Anderson, University of Minnesota; used with permission

Phellinus gilvus

Hoof Conk

Phellinus everhartii (*Fomes everhartii*)

Identification: Woody, hoof-shaped, brown to black and crusty upper surface, rusty brown interior

Season of fruiting: Perennial

Ecosystem function: White heart rot

Edibility: Inedible

Fungal note: Common on oaks, this fungus can cause large economic losses.

Joseph O'Brien, U.S. Forest Service

Phellinus everhartii

Diamond Polypore — *Polyporus alveolaris* (*Favolus alveolaris*)

Identification: Fruit body cream to orange or reddish brown; short lateral stalk, white to buff color; large diamond-shaped tubes

Season of fruiting: Spring-early summer on dead hardwood branches

Ecosystem function: White rot

Joseph O'Brien, U.S. Forest Service

Polyporus alveolaris, bottom view.

Edibility: Edible when young

Fungal note: Can cause decay when wood is at low moisture content.

Joseph O'Brien, U.S. Forest Service

Polyporus alveolaris, top view.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Dryad's Saddle *Polyporus squamosus*

Identification: Fan-shaped with central stalk, white-yellow with brown scales, white pore surface

Season of fruiting: Spring-fall

Ecosystem function: White heart rot of hardwood trees

Edibility: Inedible

Fungal note: It was estimated that a single fruit body of this fungus could produce 100 billion spores.

Mike Ostry, U.S. Forest Service

Polyporus squamosus

Mike Ostry, U.S. Forest Service

Wildlife cavity in elm with heart rot caused by Dryad's Saddle (*Polyporus squamosus*).

Cinnabar-Red Polypore *Pycnoporus cinnabarinus* (*Polyporus cinnabarinus*)

Identification: Orange-red, broadly attached leathery cap

Season of fruiting: Summer-fall

Ecosystem function: White sapwood rot of dead hardwoods

Edibility: Inedible

Fungal note: Some fruit bodies can produce spores into the second and third years.

Joseph O'Brien, U.S. Forest Service

Polyporus cinnabarinus, top view.

Joseph O'Brien, U.S. Forest Service

Polyporus cinnabarinus, bottom view.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Turkey Tail

Trametes versicolor (*Coriolus versicolor, Polyporus versicolor*)

Identification: Cap thin, leathery bracket-like; surface velvet-like with concentric bands of brown-red-yellow-gray-blue colors; pores white-yellow

Season of fruiting: Spring-fall

Ecosystem function: Causes a white rot of hardwood trees and logs

Edibility: Inedible

Fungal note: Wood decayed by this fungus often has black zone lines where different clones of this species meet but do not exchange genetic material. The zone lines produce beautiful patterns in turned vases and other objects made with the affected wood, known as spalted wood.

Mike Ostry, U.S. Forest Service

Trametes versicolor

Weeping Polypore *Ischnoderma resinosum* (*Polyporus resinosus*)

Identification: Clusters of shelf-like fruit bodies; surface dark brown and velvety with a broad white margin; amber drops of a watery fluid on the surface when fresh

Season of fruiting: Summer-fall

Ecosystem function: Decay of hardwoods, causes a white rot of sapwood and heartwood that causes the annual rings to separate

Edibility: Inedible

Fungal note: The pores of older fruit bodies break up into tooth-like spines and the entire fruit body becomes brittle. The fruit body has an anise-like odor. A very similar form of this species occurs on conifers.

Joseph O'Brien, U.S. Forest Service

Ischnoderma resinosum

Coral-Like Jelly Fungus *Tremellodendron pallidum*

Identification: Fruit body resembling coral with white, leathery, flattened upright branches found on the ground in hardwood and conifer stands

Season of fruiting: Summer-fall

Ecosystem function: Decay of litter

Edibility: Inedible

Fungal note: The spores of the true coral fungi develop on structures (basidia) on the exterior of their branches while spores of *Tremellodendron pallidum* develop on basidia within the branches.

Mike Ostry, U.S. Forest Service

Tremellodendron pallidum

Northern Tooth *Climacodon septentrionalis* (*Steccherinum septentrionale*)

Identification: Overlapping yellowish-white annual shelves with toothed undersides found on living hardwoods, especially maples

Season of fruiting: Late summer-fall

Ecosystem function: Spongy heart rot

Edibility: Inedible

Fungal note: This fungus fruits only occasionally on individual trees, and its teeth can reach 10-15 mm in length.

Joseph O'Brien, U.S. Forest Service

Climacodon septentrionalis

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Inky Caps — *Coprinus, Coprinellus, Coprinopsis* spp.

Identification: Cap conical in shape, tissue autodigests from the gills and cap margin into a black liquid containing black spores. *Coprinus comatus* (shaggy mane) has a large, white, scaly columnar cap; *Coprinellus micaceus* (mica cap) has a brown cap with mica-like particles; *Coprinopsis atramentaria* has a light gray-brown cap and occurs in clusters of 3 or more.

Season of fruiting: Summer-fall

Ecosystem function: This group of fungi fruits on buried, decayed woody debris

Edibility: Edible

Fungal note: *Coprinopsis atramentaria* and probably other related species contain coprine, a toxin that interacts with alcohol when ingested and causes severe nausea.

Joseph O'Brien, U.S. Forest Service

Coprinus comatus

DO NOT eat any mushroom unless you are absolutely certain of its identity.

UPLAND CONIFER ECOSYSTEM

Mike Ostry, U.S. Forest Service

Upland conifer

Slippery Jack Bolete *Suillus luteus*

Identification: Cap smooth, sticky red-brown; flesh white; tube openings radiate out from stalk in a linear pattern

Season of fruiting: Late summer-fall

Ecosystem function: Mycorrhizal with red pine of all ages

Edibility: Edible after removing skin of cap

Fungal note: Boletes are important food for insect larvae, invertebrates, turtles, snails, slugs, and many mammals, especially squirrels who often store the mushrooms in trees. Another bolete, *S. brevipes*, often found with jack pine, has such a short stalk that it looks like the cap is resting directly on the ground.

Mike Ostry, U.S. Forest Service

Suillus luteus, top and bottom views.

White Pine Bolete *Suillus americanus*

Identification: Cap yellow with red streaks, smooth; flesh yellow; tube openings radiate out from stalk in a linear pattern

Season of fruiting: Late summer-fall

Ecosystem function: Mycorrhizal only with white pine

Edibility: Edible

Fungal note: In mixed plantings of red and white pine, this mushroom will be found only in association with white pine. *S. luteus* will be found fruiting under red pine usually at the same time or within 1-2 weeks of *S. americanus*.

Mike Ostry, U.S. Forest Service

Suillus americanus, top and bottom views.

The King Bolete *Boletus edulis*

Identification: Cap cream-brown to reddish brown; tube openings random resembling a sponge; flesh white-yellow; stalk white-ivory with fine lines (reticulations) forming a net

Season of fruiting: Late summer-fall

Ecosystem function: Mycorrhizal with pine, spruce, oak, and birch

Edibility: Choice

Fungal note: Research suggests there are many different strains of this species and some may have anti-cancer properties.

Mike Ostry. U.S. Forest Service

Boletus edulis

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Mike Ostry, U.S. Forest Service

Boletus edulis

Mike Ostry, U.S. Forest Service

Boletus edulis

DO NOT eat any mushroom unless you are absolutely certain of its identity.

False Morel *Gyromitra esculenta*

Identification: Cap red-brown, irregular, brain-like; stalk white-yellow

Season of fruiting: Spring

Ecosystem function: Litter fungus in red and jack pine stands

Edibility: Poisonous; fumes while boiling this fungus can be toxic

Fungal note: This fungus is reported to produce the compound mono methyl hydrazine, found in rocket fuel.

Mike Ostry, U.S. Forest Service

Gyromitra esculenta

Golden Chanterelle *Cantharellus cibarius*

Identification: Cap yellow to orange, funnel-shape; gills shallow, yellow, blunt, and run down the yellow stalk

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with pine and upland hardwoods

Edibility: Choice

Fungal note: The *Cantharellus* mushrooms are known worldwide as chantherelles and are some of the very best edible mushrooms. Chantherelles are always found growing from soil, unlike false chantherelles (*Hygrophoropsis aurantiaca*) that are found on woody debris.

Joseph O'Brien, U.S. Forest Service

Cantharellus cibarius

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Lilac Cort *Cortinarius traganeus*

Identification: Cap light lilac in color; annulus is curtain-like (individual threads are distinct)

Season of fruiting: Late summer-fall

Ecosystem function: Mycorrhizal with conifers

Edibility: Highly poisonous. No species in this genus should be eaten because some contain a deadly toxin.

Fungal note: The curtain-like annulus covering the gill surface is a distinctive trait of this genus.

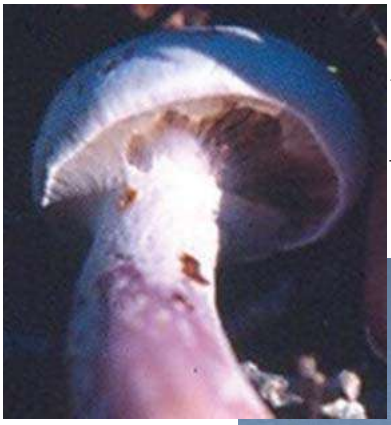

Neil A. Anderson, University of Minnesota; used with permission

Cortinarius traganeus

Mike Ostry, U.S. Forest Service

Curtain on young *Cortinarius.*

DO NOT eat any mushroom unless you are absolutely certain of its identity.

False Chanterelle *Hygrophoropsis aurantiaca*

Identification: Cap orange to orangish-brown, shallow, velvety, funnel-shaped; gills attached to stem; flesh waxy

Season of fruiting: Late summer-fall

Ecosystem function: Decay of woody debris

Edibility: Not recommended

Fungal note: Often mistaken for the true chanterelle (*Cantharellus cibarius*), but the true chanterelle is a soil fungus and does not grow on woody debris.

Mike Ostry, U.S. Forest Service

Hygrophoropsis aurantiaca, top and bottom views.

False Truffles *Rhizopogon* spp.

Identification: Mature fruit body the color and shape of a small russet potato with a chambered interior, white when young, form below or at the soil surface

Season of fruiting: Fall

Ecosystem function: One of the most important mycorrhizal species with red pine of all ages

Edibility: Inedible

Fungal note: More than 200 species of *Rhizopogon* have been described. They are eaten and inadvertently spread by many wildlife species.

Mike Ostry, U.S. Forest Service

Rhizopogon sp.

Witches Hat *Hygrocybe conica* (*Hygrophorous conicus*)

Identification: Cap cone-shaped with a definite peak when young, golden yellow-orange or red, sticky when wet; gills are waxy, white to olive yellow, and almost free from the stalk that is often twisted, hollow, striated and the same color as the cap

Season of fruiting: Summer-fall

Ecosystem function: Decays litter in conifer and hardwood stands

Edibility: Not recommended

Fungal note: These brilliantly colored mushrooms have waxy gills that are triangular in cross section.

Neil A. Anderson, University of Minnesota; used with permission

Hygrocybe conica

Pine Conk *Phellinus pini* (*Fomes pini*)

Identification: Shelf-like, tough, red-brown to brown-black

Season of fruiting: Perennial

Ecosystem function: Causes white pocket rot of living pine

Edibility: Inedible

Fungal note: This fungus causes more decay of living pines than any other fungus but does not decay wood in service such as poles, posts, and structural timbers.

Joseph O'Brien, U.S. Forest Service

Phellinus pini

The Red Band Fungus *Fomitopsis pinicola* (*Fomes pinicola*)

Identification: Brown-black, crusty fruit body with white-red margin and yellow-brown lower pore surface

Season of fruiting: Perennial on conifers and hardwoods

Ecosystem function: Common on dead trees and logs causing a brown rot

Edibility: Inedible

Fungal note: Several biological species of this fungus are known.

Mike Ostry, U.S. Forest Service

Fomitopsis pinicola

Conifer Parchment *Phlebiopsis gigantea* (*Peniophora gigantea*)

Identification: Thin, white-tan crust on stumps and logs of pine that still have bark on them

Season of fruiting: Perennial

Ecosystem function: Early colonizer of conifer sapwood

Edibility: Inedible

Fungal note: This is the world's best known biological control fungus. Conidia naturally disseminated or purposely applied onto freshly cut pine stumps will prevent decay by *Heterobasidon annosum*. Widely used in Europe, its use is not yet approved in the U.S.

Mike Ostry, U.S. Forest Service

Phlebiopsis gigantea on pine stump.

Velvet Top Fungus *Phaeolus schweinitzii* (*Polyporus schweinitzii*)

Identification: Cap is a shallow funnel with a central stalk when decaying roots or in the form of a bracket when decaying standing trees, stumps, and logs. Color ranges from yellow-brown to dark red-brown, hairy with concentric ridges. Pores form a maze when young, becoming toothed with age.

Season of fruiting: Summer-fall

Ecosystem function: Decays the heartwood of living and dead red pine

Edibility: Inedible

Fungal note: Mature red pines affected by this fungus are commonly wind thrown.

Mike Ostry, U.S. Forest Service

Phaeolus schweinitzii

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Club Coral *Clavariadelphus ligula*

Identification: Club-shaped, smooth, orange-dull yellow

Season of fruiting: Summer-fall

Ecosystem function: Decay of pine litter

Edibility: Inedible

Fungal note: A similar fungus, *C. pistillaris*, is found decaying litter in hardwood stands.

Mike Ostry, U.S. Forest Service

Clavariadelphus ligula

Honey Mushroom *Armillaria solidipes* (*A. ostoyae*)

Identification: Cap golden yellow; prominent ring on stem; black shoe-string cords (rhizomorphs) under bark of infected trees or in the soil

Season of fruiting: Late summer-fall

Ecosystem function: Causes a root and butt rot of pine

Edibility: Edible

Fungal note: The genus *Armillaria* is complex and contains 10 biological species that have restricted geographical distributions and vegetation associations. Species can be distinguished only by using laboratory techniques. All species are luminescent, often glowing in patches of decayed root or stem tissue. Clones of *Armillaria* several hundred acres in size have been found in the western U.S.

Mike Ostry, U.S. Forest Service

Armillaria solidipes

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Truffle Eater *Cordyceps ophioglossoides*

Identification: Club-shaped; yellow to olive-brown; yellow threads extending down into the soil where it parasitizes the fungus *Elaphomyces granulatus* (deer truffle)

Season of fruiting: Fall

Ecosystem function: Parasite of the deer truffle (a dark brown sphere with thick walls and a solid black interior found underground) that is mycorrhizal with jack and red pine. Other *Cordyceps* species are parasites of insect larvae and aboveground plant feeding aphids.

Edibility: Inedible

Fungal note: From a 1-m^2 sample area in a Minnesota jack pine stand, it was estimated that there were about 410,000 deer truffles per ha. These truffles are fed upon by many mammal species.

Mike Ostry, U.S. Forest Service

Cordyceps ophioglossoides the truffle eater. Several deer truffles (Elaphomyces granulatus) were dug nearby and placed in the foreground.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Conifer-Base Polypore *Heterobasidion irregulare* (*H. annosum, Fomes annosus*)

Identification: Fruit body small, white, "popcorn-like," later lying flat (resupinate) or shelf-like at base of trunks or on stumps; upper surface dark brown to black, hairy, becoming smooth with a hard crust; pore surface white-yellow

Season of fruiting: Perennial

Ecosystem function: Causes a spongy rot of sapwood and heartwood

Edibility: Inedible

Fungal note: *H. annosum* is a species complex with pine, spruce, or fir hosts. *Phlebiopsis gigantea* (conifer parchment) is used as a natural biological control of *H. annosum* in Europe when commercial formulations are applied to fresh stumps of pine when stands are thinned.

Mike Ostry, U.S. Forest Service

Heterobasidion irregulare at base of tree.

U.S. Forest Service

Heterobasidion irregulare, pore surface.

Joseph O'Brien, U.S. Forest Service

Heterobasidion irregulare, on young white pine.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Sulfur Shelf — *Laetiporus sulphureus* (*Polyporus sulphureus*)

Identification: Multiple clusters of yellow-orange shelves growing on wood, soft, fleshy when young, turning hard when mature

Season of fruiting: Summer-fall

Ecosystem function: Causes a brown cubical rot of living and dead hardwood and conifer trees

Edibility: Edible when young

Fungal note: This fungus, also called chicken of the woods, is very common on red oaks. A similar-looking species, *L. cincinnatus*, grows on the roots of infected trees.

Joseph O'Brien, U.S. Forest Service

Laetiporus sulphureus

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Joseph O'Brien, U.S. Forest Service

Laetiporus sulphureus

Joseph O'Brien, U.S. Forest Service

Laetiporus sulphureus

DO NOT eat any mushroom unless you are absolutely certain of its identity.

LOWLAND CONIFER ECOSYSTEM

Mike Ostry, U.S. Forest Service

Lowland conifer

Hollow Stem Larch Suillus *Suillus cavipes*

Identification: Cap surface dark red-brown with dense hair; pore surface white-pale yellow with tubes radiating out from a hollow stem

Season of fruiting: Fall

Ecosystem function: Mycorrhizal with tamarack in bogs

Edibility: Choice

Fungal note: Squirrels often cache this species in trees (Fig. 3).

Mike Ostry, U.S. Forest Service

Suillus cavipes, top and bottom views. Note hollow stem.

Short-Stemmed Russula *Russula brevipes*

Identification: Cap white-yellow, funnel-shaped; alternating long and short gills extending down the stalk

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with hardwoods, pine, and black spruce

Edibility: Edible, said to be choice if colonized by the orange fungus *Hypomyces lactifluorum* (bottom image)

Fungal note: Large groups of this mushroom can be overlooked because they are often partially covered by soil and leaf litter.

Joseph O'Brien, U.S. Forest Service

Russula brevipes

Mike Ostry, U.S. Forest Service

Russula brevipes parasitized by *Hypomyces lactifluorum.*

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Swamp Death Angel *Amanita brunnescens*

Identification: Cap light brown; veil and bulb present; white gills free from stalk

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with black spruce and tamarack in bogs

Edibility: Poisonous

Fungal note: *Amanita* mushrooms as a group are the most poisonous, accounting for almost all of the deaths caused by mushroom poisonings in the United States.

Mike Ostry, U.S. Forest Service

Amanita brunnescens

Larch Suillus *Suillus grevillei*

Identification: Cap shiny, bright red-brown, smooth, sticky; lower surface yellow; prominent veil on stalk

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with upland tamarack

Edibility: Edible

Fungal note: An attractive, robust mushroom found only near tamarack.

Mike Ostry, U.S. Forest Service

Suillus grevillei

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Tent Stakes *Gomphidius glutinosus*

Identification: Cap dark brown, sticky; gills run down the thick stalk; lower surface white but turning black when spores are released

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with white spruce and other conifers

Edibility: Edible

Fungal note: Mushrooms in this group are also called slime caps.

Mike Ostry, U.S. Forest Service

Gomphidius glutinosus

Hedgehog Mushroom *Hydnum repandum* (*Dentinum repandum*)

Identification: Cap buff-tan-dull orange with white-yellow teeth on the underside

Season of fruiting: Summer-fall

Ecosystem function: Litter decay hardwood and conifer stands

Edibility: Edible

Fungal note: Spores are produced on the outside surface of the downward pointing teeth.

Mike Ostry, U.S. Forest Service

Hydnum repandum, top view.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Mike Ostry, U.S. Forest Service

Hydnum repandum, bottom view.

Joseph O'Brien, U.S. Forest Service

Hydnum repandum, lower surface.

Milky Caps *Lactarius volemus*

Identification: Cap rounded, center often depressed; all members of this group contain a latex that is exuded when the gills are cut

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with conifer and hardwood trees

Edibility: Not recommended; mushrooms with a latex that turns yellow or lilac color are poisonous

Fungal note: The edible *L. deliciosus*, found in conifer and mixed conifer-hardwood stands, has an orange cap that becomes stained green when bruised and contains a yellow-orange latex.

Mike Ostry, U.S. Forest Service

Lactarius volemus, top and bottom views. Note liquid latex on cut gill surface.

Emetic Russula *Russula emetica*

Identification: Cap smooth, bright red when fresh; evenly spaced white gills; stalk dull white and hollow

Season of fruiting: Summer-fall

Ecosystem function: Mycorrhizal with conifer trees

Edibility: Mildly poisonous

Fungal note: This species can be found in deep moss in bogs.

Mike Ostry, U.S. Forest Service

Russula emetica

Mike Ostry, U.S. Forest Service

Russula emetica, bottom view.

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Yellow-Red Gill Polypore

Gloeophyllum sepiarium (*Lenzites sepiaria*)

Identification: Wooly, reddish brown shelf; yellowish-brown gills

Season of fruiting: Summer-fall

Ecosystem function: Brown cubical rot of conifers

Edibility: Inedible

Fungal note: This fungus can also decay coniferous wood products. *Gloeophyllum sepiarium* has both gills and pores and is thought to be a connecting link between the gill and pore fungi.

Joseph O'Brien, U.S. Forest Service

Gloeophyllum sepiarium

Hairy Cushion

Onnia tomentosa (*Inonotus tomentosus*, *Polyporus tomentosus*)

Identification: Brown to yellow, hairy-velvety, funnel-shaped or shelf-like on the ground at the base of trees

Season of fruiting: Summer-fall

Ecosystem function: Root and butt rot of pines and white and black spruce. A closely related fungus *Onnia circinatum* has been considered a variety of the hairy cushion. It causes a white pocket root and butt rot of pines and spruce. The species differ mainly by the shape of sterile bristle-like structures (setae) in their spore-producing areas.

Edibility: Inedible

Fungal note: Spread by root contact, this fungus causes stand openings.

Joseph O'Brien, U.S. Forest Service

Onnia tomentosa

DO NOT eat any mushroom unless you are absolutely certain of its identity.

Coral Fungus *Clavicorona pyxidata*

Identification: Multiple branched stalks, white-yellow, tips of branches forming a crown

Season of fruiting: Late spring-summer

Ecosystem function: Completes the breakdown of decayed wood

Edibility: Good when fresh

Fungal note: Spores form on the upright stalks of coral fungi and are thus unprotected from the elements compared to gill or pore fungi.

Mike Ostry, U.S. Forest Service

Clavicorona pyxidata

Mushroom

This article is about fungi. For use in food, see Edible mushroom. For other uses, see Mushroom (disambiguation).
"Toadstool" redirects here. For other uses, see Toadstool (disambiguation).

A **mushroom** (or **toadstool**) is the fleshy, spore-

Amanita muscaria, *commonly known as "fly agaric"*

bearing fruiting body of a fungus, typically produced above ground on soil or on its food source.

The standard for the name "mushroom" is the cultivated white button mushroom, *Agaricus bisporus*; hence the word "mushroom" is most often applied to those fungi (Basidiomycota, Agaricomycetes) that have a stem (stipe), a cap (pileus), and gills (lamellae, sing. lamella) on the underside of the cap. These gills produce microscopic spores that help the fungus spread across the ground or its occupant surface.

"Mushroom" describes a variety of gilled fungi, with or without stems, and the term is used even more generally, to describe both the fleshy fruiting bodies of some Ascomycota and the woody or leathery fruiting bodies of some Basidiomycota, depending upon the context of the word.

Forms deviating from the standard morphology usually have more specific names, such as "bolete", "puffball", "stinkhorn", and "morel", and gilled mushrooms themselves are often called "agarics" in reference to their similarity to *Agaricus* or their place Agaricales. By extension, the term "mushroom" can also designate the entire fungus when in culture; the thallus (called a mycelium) of species forming the fruiting bodies called mushrooms; or the species itself.

1 Identification

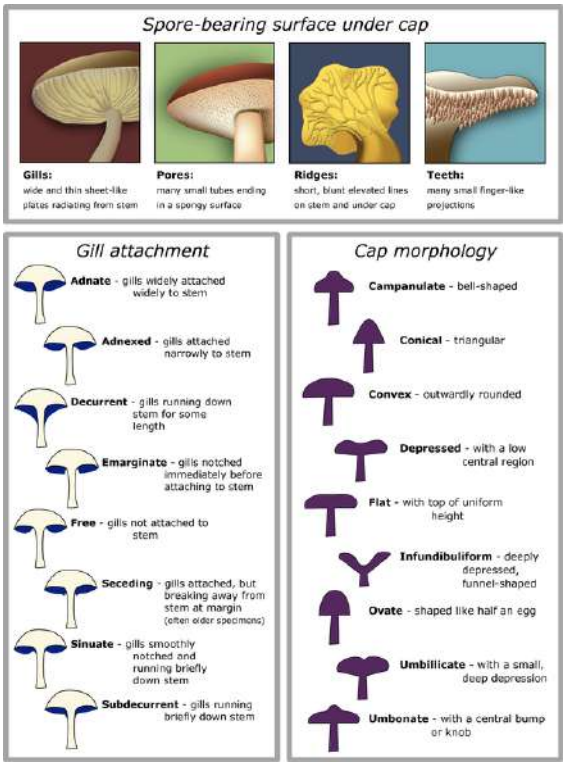

Morphological characteristics of the caps of mushrooms

Identifying mushrooms requires a basic understanding of their macroscopic structure. Most are Basidiomycetes and gilled. Their spores, called basidiospores, are produced on the gills and fall in a fine rain of powder from under the caps as a result. At the microscopic level the basidiospores are shot off basidia and then fall between the gills in the dead air space. As a result, for most mush-

rooms, if the cap is cut off and placed gill-side-down overnight, a powdery impression reflecting the shape of the gills (or pores, or spines, etc.) is formed (when the fruit body is sporulating). The color of the powdery print, called a spore print, is used to help classify mushrooms and can help to identify them. Spore print colors include white (most common), brown, black, purple-brown, pink, yellow, and creamy, but almost never blue, green, or red.[1]

While modern identification of mushrooms is quickly becoming molecular, the standard methods for identification are still used by most and have developed into a fine art harking back to medieval times and the Victorian era, combined with microscopic examination. The presence of juices upon breaking, bruising reactions, odors, tastes, shades of color, habitat, habit, and season are all considered by both amateur and professional mycologists. Tasting and smelling mushrooms carries its own hazards because of poisons and allergens. Chemical tests are also used for some genera.[2]

In general, identification to genus can often be accomplished in the field using a local mushroom guide. Identification to species, however, requires more effort; one must remember that a mushroom develops from a button stage into a mature structure, and only the latter can provide certain characteristics needed for the identification of the species. However, over-mature specimens lose features and cease producing spores. Many novices have mistaken humid water marks on paper for white spore prints, or discolored paper from oozing liquids on lamella edges for colored spored prints.

2 Classification

Main articles: Sporocarp (fungi), Basidiocarp and Ascocarp

Typical mushrooms are the fruit bodies of members of

Trametes versicolor, *a polypore mushroom*

the order Agaricales, whose type genus is *Agaricus* and type species is the field mushroom, *Agaricus campestris*. However, in modern molecularly defined classifications, not all members of the order Agaricales produce mushroom fruit bodies, and many other gilled fungi, collectively called mushrooms, occur in other orders of the class Agaricomycetes. For example, chanterelles are in the Cantharellales, false chanterelles such as *Gomphus* are in the Gomphales, milk-cap mushrooms (*Lactarius*, *Lactifluus*) and russulas (*Russula*), as well as *Lentinellus*, are in the Russulales, while the tough, leathery genera *Lentinus* and *Panus* are among the Polyporales, but *Neolentinus* is in the Gloeophyllales, and the little pin-mushroom genus, *Rickenella*, along with similar genera, are in the Hymenochaetales.

Within the main body of mushrooms, in the Agaricales, are common fungi like the common fairy-ring mushroom, shiitake, enoki, oyster mushrooms, fly agarics and other amanitas, magic mushrooms like species of *Psilocybe*, paddy straw mushrooms, shaggy manes, etc.

An atypical mushroom is the lobster mushroom, which is a deformed, cooked-lobster-colored parasitized fruitbody of a *Russula* or *Lactarius*, colored and deformed by the mycoparasitic Ascomycete *Hypomyces lactifluorum*.[3]

Other mushrooms are not gilled, so the term "mushroom" is loosely used, and giving a full account of their classifications is difficult. Some have pores underneath (and are usually called boletes), others have spines, such as the hedgehog mushroom and other tooth fungi, and so on. "Mushroom" has been used for polypores, puffballs, jelly fungi, coral fungi, bracket fungi, stinkhorns, and cup fungi. Thus, the term is more one of common application to macroscopic fungal fruiting bodies than one having precise taxonomic meaning. Approximately 14,000 species of mushrooms are described.[4]

3 Etymology

The terms "mushroom" and "toadstool" go back centuries and were never precisely defined, nor was there consensus on application. The term "toadstool" was often, but not exclusively, applied to poisonous mushrooms or to those that have the classic umbrella-like cap-and-stem form. Between 1400 and 1600 AD, the terms *tadstoles, frogstooles, frogge stoles, tadstooles, tode stoles, toodys hatte, paddockstool, puddockstool, paddocstol, toadstoole, and paddockstooles* sometimes were used synonymously with *mushrom, mushrum, muscheron, mousheroms, mussheron, or musserouns*.[6]

The word has apparent analogies in Dutch *padde(n)stoel* (toad-stool/chair, mushroom) and German *Krötenschwamm* (toad-fungus, alt. word for panther cap). In German folklore and old fairy tales, toads are often depicted sitting on toadstool mushrooms and catching, with their tongues, the flies that are said to be drawn to the *Fliegenpilz*, a German name for the toadstool, meaning "flies' mushroom". This is how the mushroom

Amanita muscaria*, the most easily recognised "toadstool", is frequently depicted in fairy stories and on greeting cards. It is often associated with gnomes.*[5]

got another of its names, *Krötenstuhl* (a less-used German name for the mushroom), literally translating to "toad-stool".

The term "mushroom" and its variations may have been derived from the French word *mousseron* in reference to moss (*mousse*). The toadstool's connection to toads may be direct, in reference to some species of poisonous toad,[7] or may just be a case of phonosemantic matching from the German word.[8] However, delineation between edible and poisonous fungi is not clear-cut, so a "mushroom" may be edible, poisonous, or unpalatable. The term "toadstool" is nowadays used in storytelling when referring to poisonous or suspect mushrooms. The classic example of a toadstool is *Amanita muscaria*.

Cultural or social phobias of mushrooms and fungi may be related. The term "fungophobia" was coined by William Delisle Hay of England, who noted a national superstition or fear of "toadstools".[9][10] He described the "fungus-hunter" as being contemptible and detailed the larger demographic's attitude toward mushrooms as "abnormal, worthless, or inexplicable".[10] Fungophobia spread to the United States and Australia, where it was inherited from England.[10][11] The underlying cause of a cultural fungophobia may also be related to the exaggerated importance placed on the few deadly and poisonous mushrooms found in the region of that culture.[12] In these regions, mushrooms were also sometimes regarded as magic or satanic, their fruiting bodies appearing quickly overnight from underground. Some believed they were the Devil's fruit, and others that mushroom rings were magical portals.

4 Morphology

Amanita jacksonii *buttons emerging from their universal veils*

The blue gills of Lactarius indigo*, a milk-cap mushroom*

A mushroom develops from a nodule, or pinhead, less than two millimeters in diameter, called a primordium, which is typically found on or near the surface of the substrate. It is formed within the mycelium, the mass of threadlike hyphae that make up the fungus. The primordium enlarges into a roundish structure of interwoven hyphae roughly resembling an egg, called a "button". The button has a cottony roll of mycelium, the universal veil, that surrounds the developing fruit body. As the egg expands, the universal veil ruptures and may remain as a cup, or volva, at the base of the stalk, or as warts or volval patches on the cap. Many mushrooms lack a universal veil, therefore they do not have either a volva or

volval patches. Often, a second layer of tissue, the partial veil, covers the bladelike gills that bear spores. As the cap expands, the veil breaks, and remnants of the partial veil may remain as a ring, or annulus, around the middle of the stalk or as fragments hanging from the margin of the cap. The ring may be skirt-like as in some species of *Amanita*, collar-like as in many species of *Lepiota*, or merely the faint remnants of a cortina (a partial veil composed of filaments resembling a spiderweb), which is typical of the genus *Cortinarius*. Mushrooms lacking partial veils do not form an annulus.[13]

The stalk (also called the stipe, or stem) may be central and support the cap in the middle, or it may be off-center and/or lateral, as in species of *Pleurotus* and *Panus*. In other mushrooms, a stalk may be absent, as in the polypores that form shelf-like brackets. Puffballs lack a stalk, but may have a supporting base. Other mushrooms, such as truffles, jellies, earthstars, and bird's nests, usually do not have stalks, and a specialized mycological vocabulary exists to describe their parts.

The way the gills attach to the top of the stalk is an important feature of mushroom morphology. Mushrooms in the genera *Agaricus*, *Amanita*, *Lepiota* and *Pluteus*, among others, have free gills that do not extend to the top of the stalk. Others have decurrent gills that extend down the stalk, as in the genera *Omphalotus* and *Pleurotus*. There are a great number of variations between the extremes of free and decurrent, collectively called attached gills. Finer distinctions are often made to distinguish the types of attached gills: adnate gills, which adjoin squarely to the stalk; notched gills, which are notched where they join the top of the stalk; adnexed gills, which curve upward to meet the stalk, and so on. These distinctions between attached gills are sometimes difficult to interpret, since gill attachment may change as the mushroom matures, or with different environmental conditions.[14]

4.1 Microscopic features

Morchella elata *asci viewed with phase contrast microscopy*

A hymenium is a layer of microscopic spore-bearing cells that covers the surface of gills. In the nongilled mushrooms, the hymenium lines the inner surfaces of the tubes of boletes and polypores, or covers the teeth of spine fungi and the branches of corals. In the Ascomycota, spores develop within microscopic elongated, sac-like cells called asci, which typically contain eight spores in each ascus. The Discomycetes, which contain the cup, sponge, brain, and some club-like fungi, develop an exposed layer of asci, as on the inner surfaces of cup fungi or within the pits of morels. The Pyrenomycetes, tiny dark-colored fungi that live on a wide range of substrates including soil, dung, leaf litter, and decaying wood, as well as other fungi, produce minute, flask-shaped structures called perithecia, within which the asci develop.[15]

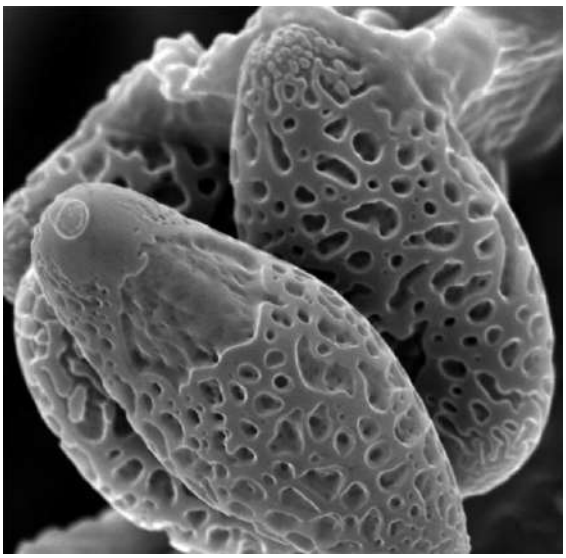

Austroboletus mutabilis *spores viewed using electron microscopy*

In the Basidiomycetes, usually four spores develop on the tips of thin projections called sterigmata, which extend from club-shaped cells called a basidia. The fertile portion of the Gasteromycetes, called a gleba, may become powdery as in the puffballs or slimy as in the stinkhorns. Interspersed among the asci are threadlike sterile cells called paraphyses. Similar structures called cystidia often occur within the hymenium of the Basidiomycota. Many types of cystidia exist, and assessing their presence, shape, and size is often used to verify the identification of a mushroom.[15]

The most important microscopic feature for identification of mushrooms is the spores. Their color, shape, size, attachment, ornamentation, and reaction to chemical tests often can be the crux of an identification. A spore often has a protrusion at one end, called an apiculus, which is the point of attachment to the basidium, termed the apical germ pore, from which the hypha emerges when the spore germinates.[15]

Agaricus bitorquis *mushroom emerging through asphalt concrete in summer*

*Yellow flower pot mushrooms (*Leucocoprinus birnbaumii*) at various states of development*

5 Growth

Many species of mushrooms seemingly appear overnight, growing or expanding rapidly. This phenomenon is the source of several common expressions in the English language including "to mushroom" or "mushrooming" (expanding rapidly in size or scope) and "to pop up like a mushroom" (to appear unexpectedly and quickly). In reality all species of mushrooms take several days to form primordial mushroom fruit bodies, though they do expand rapidly by the absorption of fluids.

The cultivated mushroom as well as the common field mushroom initially form a minute fruiting body, referred to as the pin stage because of their small size. Slightly expanded they are called buttons, once again because of the relative size and shape. Once such stages are formed, the mushroom can rapidly pull in water from its mycelium and expand, mainly by inflating preformed cells that took several days to form in the primordia.

Similarly, there are even more ephemeral mushrooms, like *Parasola plicatilis* (formerly *Coprinus plicatlis*), that literally appear overnight and may disappear by late afternoon on a hot day after rainfall.[16] The primordia form at ground level in lawns in humid spaces under the thatch and after heavy rainfall or in dewy conditions balloon to full size in a few hours, release spores, and then collapse. They "mushroom" to full size.

Not all mushrooms expand overnight; some grow very slowly and add tissue to their fruitbodies by growing from the edges of the colony or by inserting hyphae. For example, *Pleurotus nebrodensis* grows slowly, and because of this combined with human collection, it is now critically endangered.[17]

Though mushroom fruiting bodies are short-lived, the underlying mycelium can itself be long-lived and massive. A colony of *Armillaria solidipes* (formerly known as *Armillaria ostoyae*) in Malheur National Forest in the United States is estimated to be 2,400 years old, possibly older, and spans an estimated 2,200 acres (8.9 km^2). Most of the fungus is underground and in decaying wood or dying tree roots in the form of white mycelia combined with black shoelace-like rhizomorphs that bridge colonized separated woody substrates.[18]

It has been suggested the electrical stimulus of a lightning bolt striking mycelia in logs accelerates the production of mushrooms.[19]

6 Nutrition

Mushrooms are a low-calorie food eaten cooked, raw or as a garnish to a meal. In a 100 g (3.5 ounce) serving, mushrooms are an excellent source (higher than 20% of the daily value, DV) of B vitamins, such as riboflavin, niacin and pantothenic acid, an excellent source of the essential minerals, selenium (37% DV) and copper (25% DV), and a good source (10-19% DV) of phosphorus and potassium. Fat, carbohydrate and calorie content are low, with absence of vitamin C and sodium. There are 27 calories in a typical serving of fresh mushrooms (table).

When exposed to ultraviolet (UV) light even after harvesting,[20] natural ergosterols in mushrooms produce vitamin D_2,[21] a process now used to supply fresh vitamin D mushrooms for the functional food grocery market.

In a comprehensive safety assessment of producing vitamin D in fresh mushrooms, researchers showed that artificial UV light technologies were equally effective for vitamin D production as in mushrooms exposed to natural sunlight, and that UV light has a long record of safe use for production of vitamin D in food.[22]

7 Human use

Further information: Ethnomycology

The Agaricus bisporus*, one of the most widely cultivated and popular mushrooms in the world*

7.1 Edible mushrooms

Main articles: Edible mushroom, Mushroom hunting and Fungiculture

Mushrooms are used extensively in cooking, in many cuisines (notably Chinese, Korean, European, and Japanese). Though neither meat nor vegetable, mushrooms are known as the "meat" of the vegetable world.[23]

Most mushrooms sold in supermarkets have been commercially grown on mushroom farms. The most popular of these, *Agaricus bisporus*, is considered safe for most people to eat because it is grown in controlled, sterilized environments. Several varieties of *A. bisporus* are grown commercially, including whites, crimini, and portobello. Other cultivated species now available at many grocers include shiitake, maitake or hen-of-the-woods, oyster, and enoki. In recent years, increasing affluence in developing countries has led to a considerable growth in interest in mushroom cultivation, which is now seen as a potentially important economic activity for small farmers.[24]

Mushroom and truffle output in 2005

A number of species of mushrooms are poisonous; although some resemble certain edible species, consuming them could be fatal. Eating mushrooms gathered in the wild is risky and should only be undertaken by individuals knowledgeable in mushroom identification. Common best practice is for wild mushroom pickers to focus on collecting a small number of visually distinctive, edible mushroom species that cannot be easily confused with poisonous varieties. *A. bisporus* contains carcinogens called hydrazines, the most abundant of which is agaritine. However, the carcinogens are destroyed by moderate heat when cooking.[25]

More generally, and particularly with gilled mushrooms, separating edible from poisonous species requires meticulous attention to detail; there is no single trait by which all toxic mushrooms can be identified, nor one by which all edible mushrooms can be identified. Additionally, even edible mushrooms may produce allergic reactions in susceptible individuals, from a mild asthmatic response to severe anaphylactic shock.[26][27]

People who collect mushrooms for consumption are known as mycophagists,[28] and the act of collecting them for such is known as mushroom hunting, or simply "mushrooming".

China is the world's largest edible mushroom producer.[29] The country produces about half of all cultivated mushrooms, and around 2.7 kilograms (6.0 lb) of mushrooms are consumed per person per year by over a billion people.[30]

7.2 Toxic mushrooms

Main article: Mushroom poisoning

Many mushroom species produce secondary metabolites

Young Amanita phalloides*, "death cap" mushrooms*

that can be toxic, mind-altering, antibiotic, antiviral, or bioluminescent. Although there are only a small number of deadly species, several others can cause particularly severe and unpleasant symptoms. Toxicity likely plays a role in protecting the function of the basidiocarp: the mycelium has expended considerable energy and protoplasmic material to develop a structure to efficiently distribute its spores. One defense against consumption and premature destruction is the evolution of chemicals that render the mushroom inedible, either causing the consumer to vomit the meal (see emetics), or to learn to avoid consumption altogether. In addition, due to the propensity of mushrooms to absorb heavy metals, includ-

ing those that are radioactive, European mushrooms may, to date, include toxicity from the 1986 Chernobyl disaster and continue to be studied.[31][32][33][34][35]

7.3 Psychoactive mushrooms

Psilocybe zapotecorum, *a hallucinogenic mushroom*

Mushrooms with psychoactive properties have long played a role in various native medicine traditions in cultures all around the world. They have been used as sacrament in rituals aimed at mental and physical healing, and to facilitate visionary states. One such ritual is the *velada* ceremony. A practitioner of traditional mushroom use is the *shaman* or *curandera* (priest-healer).[36]

Psilocybin mushrooms possess psychedelic properties. Commonly known as “magic mushrooms” or "'shrooms", they are openly available in smart shops in many parts of the world, or on the black market in those countries that have outlawed their sale. Psilocybin mushrooms have been reported as facilitating profound and life-changing insights often described as mystical experiences. Recent scientific work has supported these claims, as well as the long-lasting effects of such induced spiritual experiences.[37]

Psilocybin, a naturally occurring chemical in certain psychedelic mushrooms such as *Psilocybe cubensis*, is being studied for its ability to help people suffering from psychological disorders, such as obsessive-compulsive disorder. Minute amounts have been reported to stop cluster and migraine headaches.[38] A double-blind study, done by the Johns Hopkins Hospital, showed psychedelic mushrooms could provide people an experience with substantial personal meaning and spiritual significance. In the study, one third of the subjects reported ingestion of psychedelic mushrooms was the single most spiritually significant event of their lives. Over two-thirds reported it among their five most meaningful and spiritually significant events. On the other hand, one-third of the subjects reported extreme anxiety. However, the anxiety went away after a short period of time.[39][40] Psilocybin mushrooms have also shown to be successful in treating addiction, specifically with alcohol and cigarettes.[41]

A few species in the *Amanita* genus, most recognizably *A. muscaria*, but also *A. pantherina*, among others, contain the psychoactive compound muscimol. The muscimol-containing chemotaxonomic group of *Amanitas* contains no amatoxins or phallotoxins, and as such are not hepatoxic, though if not properly cured will be nonlethally neurotoxic due to the presence of ibotenic acid. The *Amanita* intoxication is similar to Z-drugs in that it includes CNS depressant and sedative-hypnotic effects, but also dissociation and delirium in high doses.

7.4 Medicinal properties

Main article: Medicinal mushrooms

Some mushrooms or extracts are used or studied as

Ganoderma lucidum

possible treatments for diseases, such as cardiovascular disorders.[42] Some mushroom materials, including polysaccharides, glycoproteins and proteoglycans are under basic research for their potential to modulate immune system responses and inhibit tumor growth,[43] whereas other isolates show potential antiviral, antibacterial, antiparasitic, anti-inflammatory, and antidiabetic properties in preliminary studies.[44] Currently, several extracts have widespread use in Japan, Korea and China, as adjuncts to radiation treatments and chemotherapy,[45][46] even though clinical evidence of efficacy in humans has not been confirmed.

Historically, mushrooms have long been thought to hold medicinal value, especially in traditional Chinese medicine.[47] They have been studied in modern medical research since the 1960s, where most studies use extracts, rather than whole mushrooms. Only a few specific extracts have been tested for efficacy in laboratory research. Polysaccharide-K and lentinan are among extracts best understood from *in vitro* research, animal models such as mice, or early-stage human pilot studies.[46]

Preliminary experiments show glucan-containing mushroom extracts may affect function of the

innate and adaptive immune systems, functioning as bioresponse modulators.[46] In some countries, extracts of polysaccharide-K, schizophyllan, polysaccharide peptide, or lentinan are government-registered adjuvant cancer therapies.[45][48]

As of June 2014, whole mushrooms or mushroom ingredients are being studied in 32 human clinical trials registered with the US National Institutes of Health for their potential effects on a variety of diseases and normal physiological conditions, including vitamin D deficiency, cancer, bone metabolism, glaucoma, immune functions and inflammatory bowel disease.[49]

7.5 Other uses

A tinder fungus, Fomes fomentarius

Mushrooms can be used for dyeing wool and other natural fibers. The chromophores of mushroom dyes are organic compounds and produce strong and vivid colors, and all colors of the spectrum can be achieved with mushroom dyes. Before the invention of synthetic dyes, mushrooms were the source of many textile dyes.[50]

Some fungi, types of polypores loosely called mushrooms, have been used as fire starters (known as tinder fungi).

Mushrooms and other fungi play a role in the development of new biological remediation techniques (e.g., using mycorrhizae to spur plant growth) and filtration technologies (e.g. using fungi to lower bacterial levels in contaminated water).[51]

8 References

[1] Dickinson C, Lucas J. (1982). *VNR Color Dictionary of Mushrooms*. Van Nostrand Reinhold. pp. 9–11. ISBN 978-0-442-21998-7.

[2] Ammirati *et al.*, 1985, pp. 40–41.

[3] Volk T. (2001). "*Hypomyces lactifluorum*, the lobster mushroom". *Fungus of the Month*. University of Wisconsin-La Crosse, Department of Biology. Retrieved 2008-10-13.

[4] Miles PG, Chang S-T. (2004). *Mushrooms: Cultivation, Nutritional Value, Medicinal Effect, and Environmental Impact*. Boca Raton, Florida: CRC Press. ISBN 0-8493-1043-1.

[5] Harding, Patrick (2008). *Mushroom Miscellany*. HarperCollins. p. 149. ISBN 978-0-00-728464-1.

[6] Ramsbottom J. (1954). *Mushrooms & Toadstools: a study of the activities of fungi*. London: Collins.

[7] "Botany". Ontarioprofessionals.com. 2009-03-26. Retrieved 2010-05-30.

[8] "Yahoo! Babel Fish – Text Translation and Web Page Translation". Babelfish.yahoo.com. Archived from the original on July 18, 2011. Retrieved 2010-05-30.

[9] Hay, William Deslisle (1887). "An Elementary Text-Book of British Fungi". London, S. Sonnenschein, Lowrey. pp. 6–7.

[10] Arora, David (1986). *Mushrooms Demystified, A Comprehensive Guide to the Fleshy Fungi*. Ten Speed Press. pp. 1–3. ISBN 978-0-89815-169-5.

[11] Leschyn, Wade. "Identifying Mushrooms". Peninsula Mycological Circle. Retrieved 2012-01-02.

[12] Hunter, Jessica. "The Mushroom Hunt". Synergy Magazine. Retrieved 2012-01-02.

[13] Stuntz *et al.*, 1978, pp. 12–13.

[14] Stuntz *et al.*, 1978, pp. 28–29.

[15] Ammirati *et al.*, 1985, pp. 25–34.

[16] Nelson N. (2006-08-13). "*Parasola plicatilis*". Retrieved 2008-10-13.

[17] Venturella, G. 2006. *Pleurotus nebrodensis*. In: IUCN. 2009. IUCN Red List of Threatened Species. Version 2009.1. http://www.iucnredlist.org/apps/redlist/details/full/61597/0 Downloaded on 15 October 2009.

[18] Dodge SR. "And the Humongous Fungus Race Continues". US Forest Service: Pacific Northwest Research Station. Retrieved 2011-02-28.

[19] "IEEE Xplore – Development of an Automatic Electrical Stimulator for Mushroom Sawdust Bottle". Ieeexplore.ieee.org. 2005-06-17. doi:10.1109/PPC.2005.300675. Retrieved 2014-01-24.

[20] Kalaras, M. D.; Beelman, R. B.; Elias, R. J. (2012). "Effects of postharvest pulsed UV light treatment of white button mushrooms (Agaricus bisporus) on vitamin D2 content and quality attributes". *Journal of Agricultural and Food Chemistry* **60** (1): 220–5. doi:10.1021/jf203825e. PMID 22132934.

[21] Koyyalamudi SR, Jeong SC, Song CH, Cho KY, Pang G. (2009). "Vitamin D2 formation and bioavailability from *Agaricus bisporus* button mushrooms treated with ultraviolet irradiation" (PDF). *Journal of Agricultural and Food Chemistry* **57** (8): 3351–5. doi:10.1021/jf803908q. PMID 19281276.

200 Proofs Earth is Not a Spinning Ball

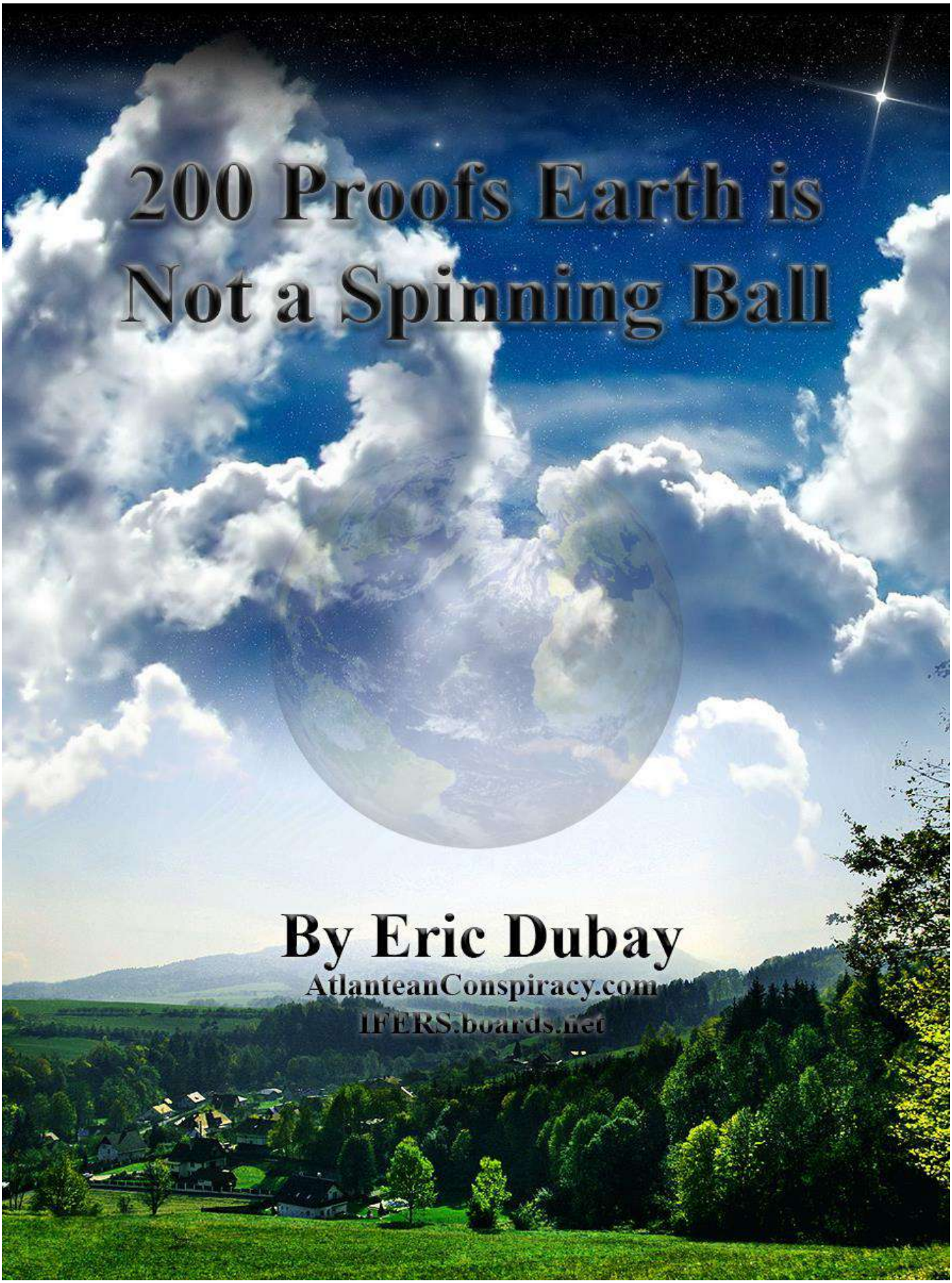

By Eric Dubay

AtlanteanConspiracy.com

IFERS.boards.net

1) The horizon always appears perfectly flat 360 degrees around the observer regardless of altitude. All amateur balloon, rocket, plane and drone footage show a completely flat horizon over 20+ miles high. Only NASA and other government "space agencies" show curvature in their fake CGI photos/videos.

2) The horizon always rises to the eye level of the observer as altitude is gained, so you never have to look down to see it. If Earth were in fact a globe, no matter how large, as you ascended the horizon would stay fixed and the observer / camera would have to tilt looking down further and further to see it.

3) The natural physics of water is to find and maintain its level. If Earth were a giant sphere tilted, wobbling and hurdling through infinite space then truly flat, consistently level surfaces would not exist here. But since Earth is in fact an extended flat plane, this fundamental physical property of fluids finding and remaining level is consistent with experience and common sense.

4) Rivers run down to sea-level finding the easiest course, North, South, East, West and all other intermediary directions over the Earth at the same time. If Earth were truly a spinning ball then many of these rivers would be impossibly flowing uphill, for example the Mississippi in its 3000 miles would have to ascend 11 miles before reaching the Gulf of Mexico.

5) One portion of the Nile River flows for a thousand miles with a fall of only one foot. Parts of the West African Congo, according to the supposed inclination and movement of the ball-Earth, would be sometimes running uphill and sometimes down. This would also be the case for the Parana, Paraguay and other long rivers.

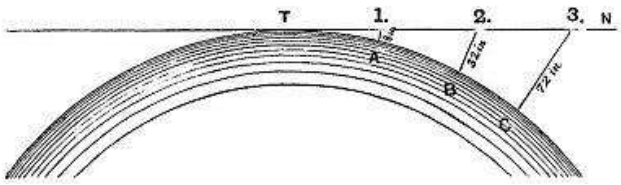

6) If Earth were a ball 25,000 miles in circumference as NASA and modern astronomy claim, spherical trigonometry dictates the surface of all standing water must curve downward an easily measurable 8 inches per mile multiplied by the square of the distance. This means along a 6 mile channel of standing water, the Earth would dip 6 feet on either end from the central peak. Every time such experiments have been conducted, however, standing water has proven to be perfectly level.

7) Surveyors, engineers and architects are never required to factor the supposed curvature of the Earth into their projects. Canals, railways, bridges and tunnels for example are always cut and laid horizontally, often over hundreds of miles without any allowance for curvature.

8) The Suez Canal connecting the Mediterranean with the Red Sea is 100 miles long without any locks making the water an uninterrupted continuation of the two seas. When constructed, the Earth's supposed curvature was not taken into account, it was dug along a horizontal datum line 26 feet below sea-level, passing through several lakes from one sea to the other, with the datum line and water's surface running perfectly parallel over the 100 miles.

9) Engineer, W. Winckler was published in the Earth Review regarding the Earth's supposed curvature, stating, "*As an engineer of many years standing, I saw that this absurd allowance is only permitted in school books. No engineer would dream of allowing anything of the kind. I have projected many miles of railways and many more of canals and the allowance has not even been thought of, much less allowed for. This allowance for curvature means this - that it is 8" for the first mile of a canal, and increasing at the ratio by the square of the distance in miles; thus a small navigable canal for boats, say 30 miles long, will have, by the above rule an allowance for curvature of 600 feet. Think of that and then please credit engineers as not being quite such fools. Nothing of the sort is allowed. We no more think of allowing 600 feet for a line of 30 miles of railway or canal, than of wasting our time trying to square the circle*"

10) The London and Northwestern Railway forms a straight line 180 miles long between London and Liverpool. The railroad's highest point, midway at Birmingham station, is only 240 feet above sea-level. If the world were actually a globe, however, curving 8 inches per mile squared, the 180 mile stretch of rail would form an arc with the center point at Birmingham raising over a mile, a full 5,400 feet above London and Liverpool.

11) A surveyor and engineer of thirty years published in the Birmingham Weekly Mercury stated, "*I am thoroughly acquainted with the theory and practice of civil engineering. However bigoted some of our professors may be in the theory of surveying according to the prescribed rules, yet it is well known amongst us that such theoretical measurements are INCAPABLE OF ANY PRACTICAL ILLUSTRATION. All our locomotives are designed to run on what may be regarded as TRUE LEVELS or FLATS. There are, of course, partial inclines or gradients here and there, but they are always accurately defined and must be carefully traversed. But anything approaching to eight inches in the mile, increasing as the square of the distance, COULD NOT BE WORKED BY ANY ENGINE THAT WAS EVER YET CONSTRUCTED. Taking one station with another all over England and Scotland, it may be stated that all the platforms are ON THE SAME RELATIVE LEVEL. The distance between Eastern and Western coasts of England may be set down as 300 miles. If the prescribed curvature was indeed as represented, the central stations at Rugby or Warwick ought to be close upon three miles higher than a chord drawn from the two extremities. If such was the case there is not a driver or stoker within the Kingdom that would be found to take charge of the train. We can only laugh at those of your readers who seriously give us credit for such venturesome exploits, as running trains round spherical curves. Horizontal curves on levels are dangerous enough, vertical curves would be a thousand times worse, and with our rolling stock constructed as at present* ***physically impossible.***"

12) The Manchester Ship Canal Company published in the Earth Review stated, "*It is customary in Railway and Canal constructions for all levels to be referred to a datum which is nominally horizontal and is so shown on all sections.* ***It is not the practice in laying out Public Works to make allowances for the curvature of the earth***

13) In a 19th century French experiment by M. M. Biot and Arago a powerful lamp with good reflectors was placed on the summit of Desierto las Palmas in Spain and able to be seen all the way from Camprey on the Island of Iviza. Since the elevation of the two points were identical and the distance between covered nearly 100 miles, if Earth were a

ball 25,000 miles in circumference, the light should have been more than 6600 feet, a mile and a quarter, below the line of sight!

14) The Lieutenant-Colonel Portlock experiment used oxy-hydrogen Drummond's lights and heliostats to reflect the sun's rays across stations set up across 108 miles of St. George's Channel. If the Earth were actually a ball 25,000 miles in circumference, Portlock's light should have remained hidden under a mile and a half of curvature.

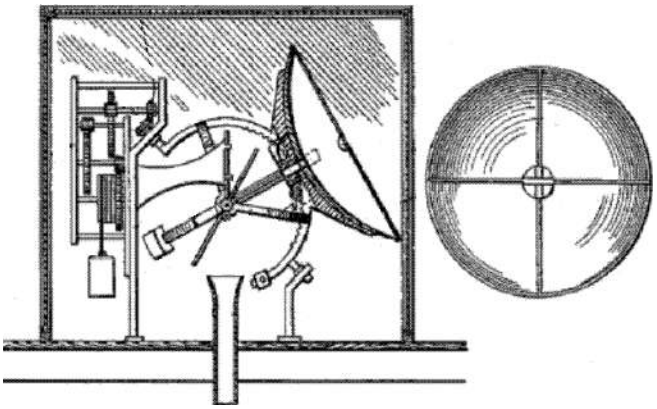

15) If the Earth were truly a sphere 25,000 miles in circumference, airplane pilots would have to constantly correct their altitudes downwards so as to not fly straight off into "outer space;" a pilot wishing to simply maintain their altitude at a typical cruising speed of 500 mph, would have to constantly dip their nose downwards and descend 2,777 feet (over half a mile) every minute! Otherwise, without compensation, in one hour's time the pilot would find themselves 31.5 miles higher than expected.

16) The experiment known as "Airy's Failure" proved that the stars move relative to a stationary Earth and not the other way around. By first filling a telescope with water to slow down the speed of light inside, then calculating the tilt necessary to get the starlight directly down the tube, Airy failed to prove the heliocentric theory since the starlight was already coming in the correct angle with no change necessary, and instead proved the geocentric model correct.

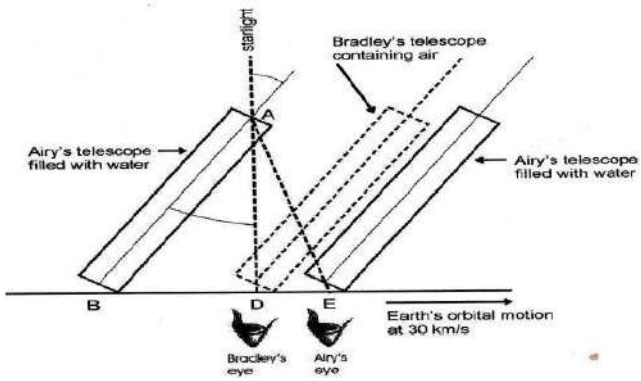

17) "Olber's Paradox" states that if there were billions of stars which are suns the night sky would be filled completely with light. As Edgar Allen Poe said, "*Were the succession of stars endless, then the background of the sky would present us a uniform luminosity, since there could exist absolutely no point, in all that background, at which would not exist a star*." In fact Olber's "Paradox" is no more a paradox than George Airy's experiment was a "failure." Both are actually excellent refutations of the heliocentric spinning ball model.

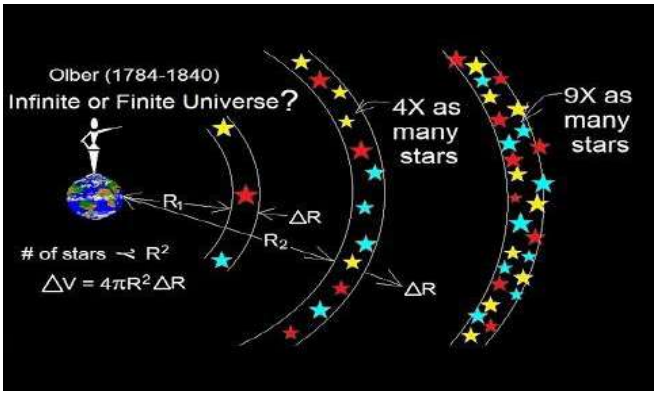

18) The Michelson-Morley and Sagnac experiments attempted to measure the change in speed of light due to Earth's assumed motion through space. After measuring in every possible different direction in various locations they failed to detect any significant change whatsoever, again proving the stationary geocentric model.

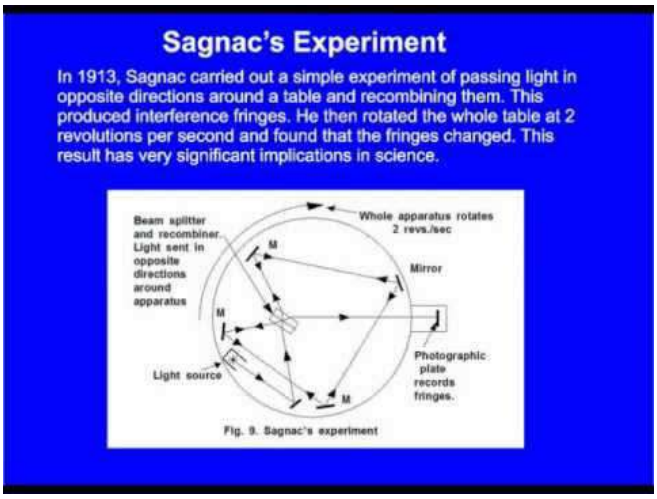

19) Tycho Brahe famously argued against the heliocentric theory in his time, positing that if the Earth revolved around

the Sun, the change in relative position of the stars after 6 months orbital motion could not fail to be seen. He argued that the stars should seem to separate as we approach and come together as we recede. In actual fact, however, after 190,000,000 miles of supposed orbit around the Sun, not a single inch of parallax can be detected in the stars, proving we have not moved at all.

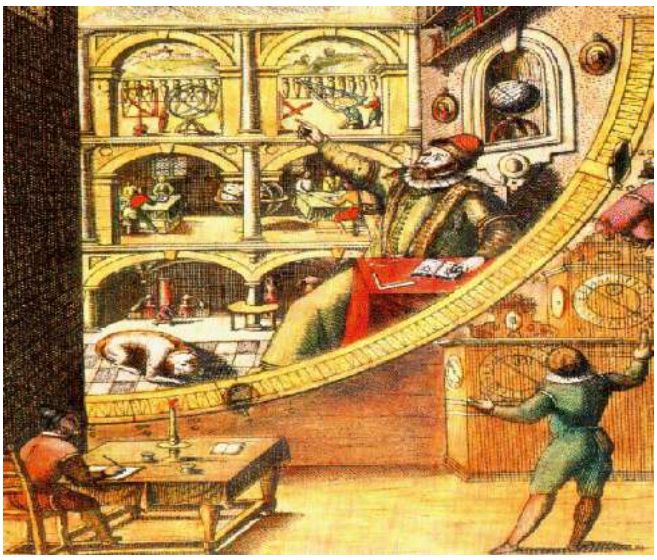

20) If Earth were truly constantly spinning Eastwards at over 1000mph, vertically-fired cannonballs and other projectiles should fall significantly due west. In actual fact, however, whenever this has been tested, vertically-fired cannonballs shoot upwards an average of 14 seconds ascending, 14 seconds descending, and fall back to the ground no more than 2 feet away from the cannon, often directly back into the muzzle.

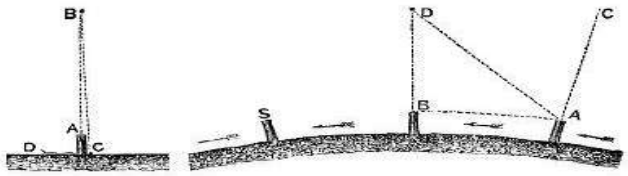

21) If the Earth were truly constantly spinning Eastwards at over 1000mph, helicopters and hot-air balloons should be able to simply hover over the surface of the Earth and wait for their destinations to come to them!

22) If Earth were truly constantly spinning Eastwards at over 1000mph, during the Red Bull stratosphere dive, Felix Baumgartner, spending 3 hours ascending over New Mexico, should have landed 2500 miles West into the Pacific Ocean but instead landed a few dozen miles East of the take-off point.

23) Ball-believers often claim "gravity" magically and inexplicably drags the entire lower-atmosphere of the Earth in perfect synchronization up to some undetermined height where this progressively faster spinning atmosphere gives way to the non-spinning, non-gravitized, non-atmosphere of infinite vacuum space. Such non-sensical theories are debunked, however, by rain, fireworks, birds, bugs, clouds, smoke, planes and projectiles all of which would behave very differently if both the ball-Earth and its atmosphere were constantly spinning Eastwards at 1000mph.

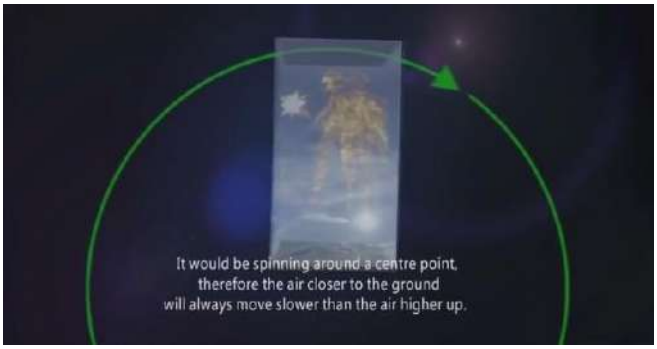

24) If Earth and its atmosphere were constantly spinning eastwards over 1000mph then North/South facing cannons should establish a control while East-firing cannonballs should fall significantly farther than all others while West-firing cannonballs should fall significantly closer. In actual fact, however, regardless of which direction cannons are fired, the distance covered is always the same.

25) If Earth and its atmosphere were constantly spinning eastwards over 1000mph, then the average commercial airliner traveling 500mph should never be able to reach its Eastward destinations before they come speeding up from behind! Likewise Westward destinations should be arrived at thrice the speed, but this is not the case.

26) Quoting "Heaven and Earth" by Gabrielle Henriet, "*If flying had been invented at the time of Copernicus, there is no doubt that he would have soon realized that his contention regarding the rotation of the earth was wrong, on account of the relation existing between the speed of an aircraft and that of the earth's rotation. If the earth rotates, as it is said, at 1,000 miles an hour, and a plane flies in the same direction at only 500 miles, it is obvious that its place of destination will be farther removed every minute. On the other hand, if flying took place in the direction opposite to that of the rotation, a distance of 1,500 miles would be covered in one hour, instead of 500, since the speed of the rotation is to be added to that of the plane. It could also be pointed out that such a flying speed of 1,000 miles an hour, which is supposed to be that of the earth's rotation, has*

recently been achieved, so that an aircraft flying at this rate in the same direction as that of the rotation could not cover any ground at all. It would remain suspended in mid-air over the spot from which it took off, since both speeds are equal."

27) If Earth and its atmosphere were constantly spinning Eastwards over 1000mph, landing airplanes on such fast-moving runways which face all manner of directions North, South, East, West and otherwise would be practically impossible, yet in reality such fictional concerns are completely negligible.

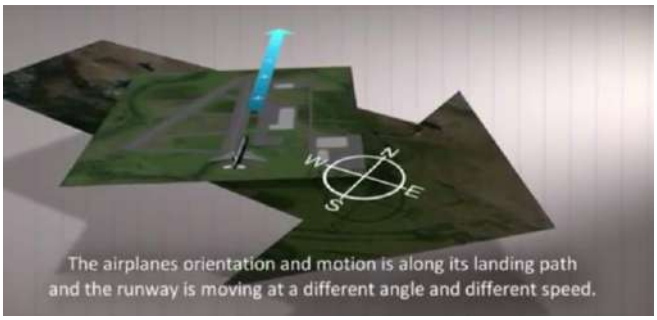

28) If the Earth and its atmosphere were constantly spinning Eastwards over 1000mph, then clouds, wind and weather patterns could not casually and unpredictably go every which way, with clouds often travelling in opposing directions at varying altitudes simultaneously.

29) If the Earth and its atmosphere were constantly spinning Eastwards over 1000mph, this should somewhere somehow be seen, heard, felt or measured by someone, yet no one in history has ever experienced this alleged Eastward motion; meanwhile, however, we can hear, feel and experimentally measure even the slightest Westward breeze.

30) In his book "South Sea Voyages," Arctic and Antarctic explorer Sir James Clarke Ross, described his experience on the night of November 27th, 1839 and his conclusion that the Earth must be motionless: *"The sky being very clear … it enabled us to observe the higher stratum of clouds to be moving in an exactly opposite direction to that of the wind--a circumstance which is frequently recorded in our meteorological journal both in the north-east and south-east trades, and has also often been observed by former voyagers. Captain Basil Hall witnessed it from the summit of the Peak of Teneriffe; and Count Strzelechi, on ascending the volcanic mountain of Kiranea, in Owhyhee, reached at 4000 feet an elevation above that of the trade wind, and experienced the influence of an opposite current of air of a different hygrometric and thermometric condition … Count Strzelechi further informed me of the following seemingly anomalous circumstance--that at the height of 6000 feet he found the current of air blowing at right angles to both the lower strata, also of a different hygrometric and thermometric condition, but warmer than the inter-stratum.* ***Such a state of the atmosphere is compatible only with the fact which other evidence has demonstrated, that the earth is at rest."***

31) Quoting "Zetetic Cosmogeny" Thomas Winships states: "*Let 'imagination' picture to the mind what force air would have which was set in motion by a spherical body of 8,000 miles in diameter, which in one hour was spinning round 1,000 mph, rushing through space at 65,000 mph and gyrating across the heavens? Then let 'conjecture' endeavor to discover whether the inhabitants on such a globe could keep their hair on? If the earth-globe rotates on its axis at the terrific rate of 1,000 miles an hour, such an immense mass would of necessity cause a tremendous rush of wind in the space it occupied. The wind would go all one way, and anything like clouds which got 'within the sphere of influence' of the rotating sphere, would have to go the same way.* ***The fact that the earth is at rest is proved by kite flying***."

32) If "gravity" is credited with being a force strong enough to hold the world's oceans, buildings, people and atmosphere stuck to the surface of a rapidly spinning ball, then it is impossible for "gravity" to also simultaneously be weak enough to allow little birds, bugs, and planes to take-off and travel freely unabated in any direction.

33) If "gravity" is credited with being a force strong enough to curve the massive expanse of oceans around a globular Earth, it would be impossible for fish and other creatures to swim through such forcefully held water.

34) Ship captains in navigating great distances at sea never need to factor the supposed curvature of the Earth into their calculations. Both Plane Sailing and Great Circle Sailing, the most popular navigation methods, use plane, not spherical trigonometry, making all mathematical

calculations on the assumption that the Earth is perfectly flat. If the Earth were in fact a sphere, such an errant assumption would lead to constant glaring inaccuracies. Plane Sailing has worked perfectly fine in both theory and practice for thousands of years, however, and plane trigonometry has time and again proven **more** accurate than spherical trigonometry in determining distances across the oceans.

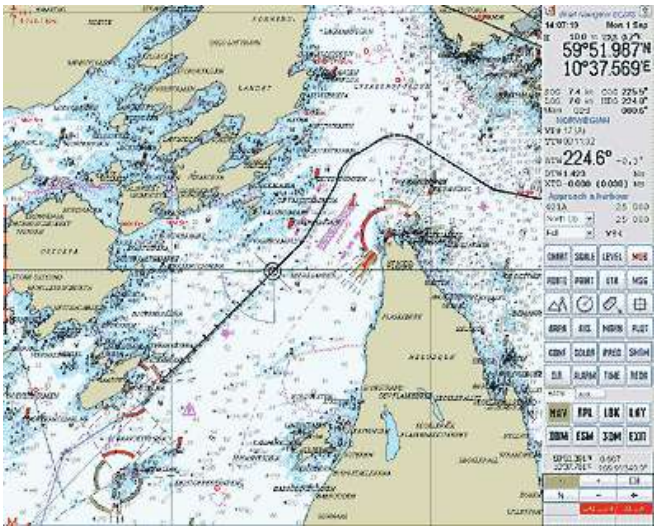

35) If the Earth were truly a globe, then every line of latitude south of the equator would have to measure a gradually smaller and smaller circumference the farther South travelled. If, however, the Earth is an extended plane, then every line of latitude south of the equator should measure a gradually larger and larger circumference the farther South travelled. The fact that many captains navigating south of the equator assuming the globular theory have found themselves drastically out of reckoning, moreso the farther South travelled, testifies to the fact that the Earth is not a ball.

36) During Captain James Clark Ross's voyages around the Antarctic circumference, he often wrote in his journal perplexed at how they routinely found themselves out of accordance with their charts, stating that they found themselves an average of 12-16 miles outside their reckoning every day, later on further south as much as 29 miles.

37) Lieutenant Charles Wilkes commanded a United States Navy exploration expedition to the Antarctic from 1838 to 1842, and in his journals also mentioned being consistently east of his reckoning, sometimes over 20 miles in less than 18 hours.

38) To quote Reverend Thomas Milner, "*In the southern hemisphere, navigators to India have often fancied themselves east of the Cape when still west, and have been driven ashore on the African coast, which, according to their reckoning, lay behind them. This misfortune happened to a fine frigate, the Challenger, in 1845. How came Her Majesty's Ship 'Conqueror,' to be lost? How have so many other noble vessels, perfectly sound, perfectly manned, perfectly navigated, been wrecked in calm weather, not only in dark night, or in a fog, but in broad daylight and sunshine - in the former case upon the coasts, in the latter, upon sunken rocks - from being 'out of reckoning?'*" The simple answer is that Earth is not a ball.

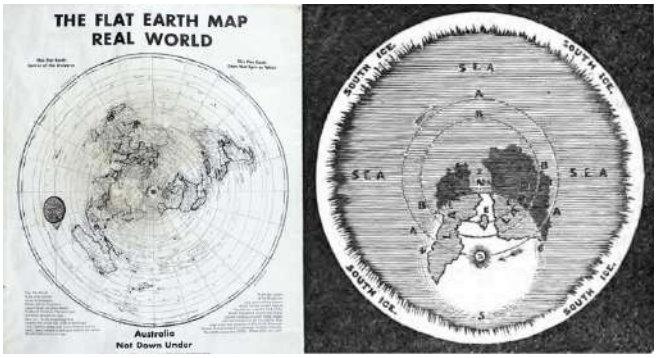

39) Practical distance measurements taken from "The Australian Handbook, Almanack, Shippers' and Importers' Directory" state that the straight line distance between Sydney and Nelson is 1550 statute miles. Their given difference in longitude is 22 degrees 2'14". Therefore if 22 degrees 2'14" out of 360 is 1550 miles, the entirety would measure 25,182 miles. This is not only larger than the ball-Earth is said to be at the equator, but a whole 4262 miles greater than it would be at Sydney's southern latitude on a globe of said proportions.

40) From near Cape Horn, Chile to Port Philip in Melbourne, Australia the distance is 10,500 miles, or 143 degrees of longitude away. Factoring in the remaining degrees to 360 makes for a total distance of 26,430 miles

around this particular latitude, which is over 1500 miles wider than Earth is supposed to be at the equator, and many more thousands of miles wider than it is supposed to be at such Southern latitudes.

41) Similar calculations made from the Cape of Good Hope, South Africa to Melbourne, Australia at an average latitude of 35.5 degrees South, have given an approximate figure of over 25,000 miles, which is again equal to or greater than the Earth's supposed greatest circumference at the equator. Calculations from Sydney, Australia to Wellington, New Zealand at an average of 37.5 degrees South have given an approximate circumference of 25,500 miles, greater still! According to the ball-Earth theory, the circumference of the Earth at 37.5 degrees Southern latitude should be only 19,757 statute miles, almost six thousand miles less than such practical measurements.

42) In the ball-Earth model Antarctica is an ice continent which covers the bottom of the ball from 78 degrees South latitude to 90 and is therefore not more than 12,000 miles in circumference. Many early explorers including Captian Cook and James Clark Ross, however, in attempting Antarctic circumnavigation took 3 to 4 years and clocked 50-60,000 miles around. The British ship Challenger also made an indirect but complete circumnavigation of Antarctica traversing 69,000 miles. This is entirely inconsistent with the ball model.

43) If Earth was a ball there are several flights in the Southern hemisphere which would have their quickest, straightest path over the Antarctic continent such as Santiago, Chile to Sydney, Australia. Instead of taking the shortest, quickest route in a straight line over Antarctica, all such flights detour all manner of directions away from Antarctica instead claiming the temperatures too cold for airplane travel! Considering the fact that there are plenty of flights to/from/over Antarctica, and NASA claims to have technology keeping them in conditions far colder (and far hotter) than any experienced on Earth, such an excuse is clearly just an excuse, and these flights aren't made because they are impossible.

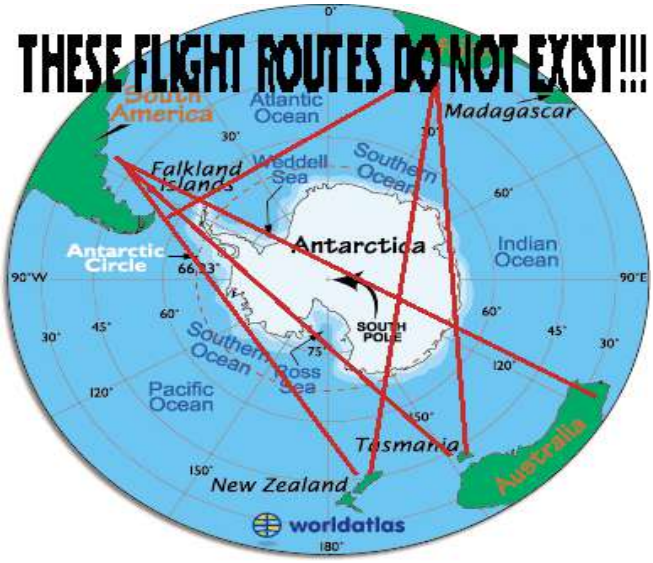

44) If Earth was a ball, and Antarctica was too cold to fly over, the only logical way to fly from Sydney to Santiago would be a straight shot over the Pacific staying in the Southern hemisphere the entire way. Re-fueling could be done in New Zealand or other Southern hemisphere destinations along the way if absolutely necessary. In actual fact, however, Santiago-Sydney flights go into the Northern hemisphere making stop-overs at LAX and other North American airports before continuing back down to the Southern hemisphere. Such ridiculously wayward detours make no sense on the globe but make perfect sense and form nearly straight lines when shown on a flat Earth map.

45) On a ball-Earth, Johannesburg, South Africa to Perth, Australia should be a straight shot over the Indian Ocean with convenient re-fueling possibilities on Mauritus or Madagascar. In actual practice, however, most Johannesburg to Perth flights curiously stop over either in Dubai, Hong Kong or Malaysia all of which make no sense

on the ball, but are completely understandable when mapped on a flat Earth.

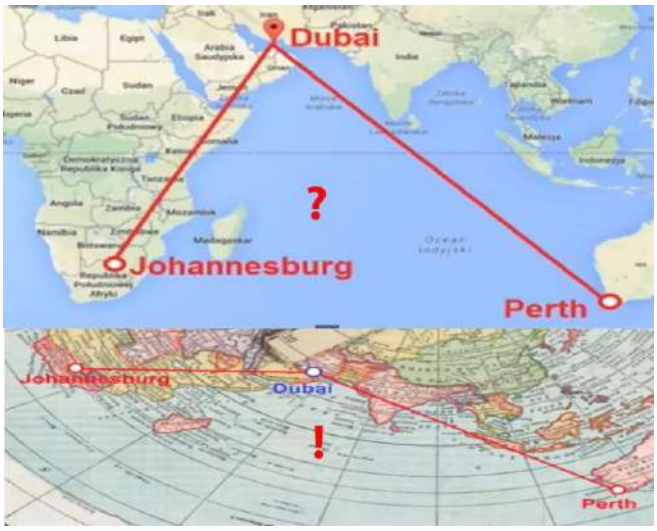

46) On a ball-Earth Cape Town, South Africa to Buenos Aries, Argentina should be a straight shot over the Atlantic following the same line of latitude across, but instead every flight goes to connecting locations in the Northern hemisphere first, stopping over anywhere from London to Turkey to Dubai. Once again these make absolutely no sense on the globe but are completely understandable options when mapped on a flat Earth.

47) On a ball-Earth Johannesburg, South Africa to Sao Paolo, Brazil should be a quick straight shot along the 25th Southern latitude, but instead nearly every flight makes a re-fueling stop at the 50th degree North latitude in London first! The only reason such a ridiculous stop-over works in reality is because the Earth is flat.

48) On a ball-Earth Santiago, Chile to Johannesburg, South Africa should be an easy flight all taking place below the Tropic of Capricorn in the Southern hemisphere, yet every listed flight makes a curious re-fueling stop in Senegal near the Tropic of Cancer in the North hemisphere first! When mapped on a flat Earth the reason why is clear to see, however, Senegal is actually directly in a straight-line path half-way between the two.

49) If Earth were a spinning ball heated by a Sun 93 million miles away, it would be impossible to have simultaneously sweltering summers in Africa while just a few thousand miles away bone-chilling frozen Arctic/Antarctic winters experiencing little to no heat from the Sun whatsoever. If the heat from the Sun traveled 93,000,000 miles to the Sahara desert, it is absurd to assert that another 4,000 miles (0.00004%) further to Antarctica would completely negate such sweltering heat resulting in such drastic differences.

50) If the Earth were truly a globe, the Arctic and Antarctic polar regions and areas of comparable latitude North and South of the equator should share similar conditions and characteristics such as comparable temperatures, seasonal changes, length of daylight, plant and animal life. In reality, however, the Arctic/Antarctic regions and areas of comparable latitude North/South of the equator differ greatly in many ways entirely inconsistent with the ball model and entirely consistent with the flat model.

51) Antarctica is by far the coldest place on Earth with an average annual temperature of approximately -57 degrees Fahrenheit, and a record low of -135.8! The average annual temperature at the North Pole, however, is a comparatively warm 4 degrees. Throughout the year, temperatures in the Antarctic vary less than half the amount at comparable Arctic latitudes. The Northern Arctic region enjoys moderately warm summers and manageable winters, whereas the Southern Antarctic region never even warms enough to melt the perpetual snow and ice. On a tilting, wobbling, ball-Earth spinning uniformly around the Sun, Arctic and Antarctic temperatures and seasons should not vary so greatly.

52) Iceland at 65 degrees North latitude is home to 870 species of native plants and abundant various animal life. Compare this with the Isle of Georgia at just 54 degrees South latitude where there are only 18 species of native plants and animal life is almost non-existent. The same latitude as Canada or England in the North where dense forests of various tall trees abound, the infamous Captain Cook wrote of Georgia that he was unable to find a single shrub large enough to make a toothpick! Cook wrote, "*Not a tree was to be seen. The lands which lie to the south are doomed by nature to perpetual frigidness - never to feel the warmth of the sun's rays; whose horrible and savage aspect I have not words to describe. Even marine life is sparse in certain tracts of vast extent, and the sea-bird is seldom observed flying over such lonely wastes. The contrasts between the limits of organic life in Arctic and Antarctic zones is very remarkable and significant.*"

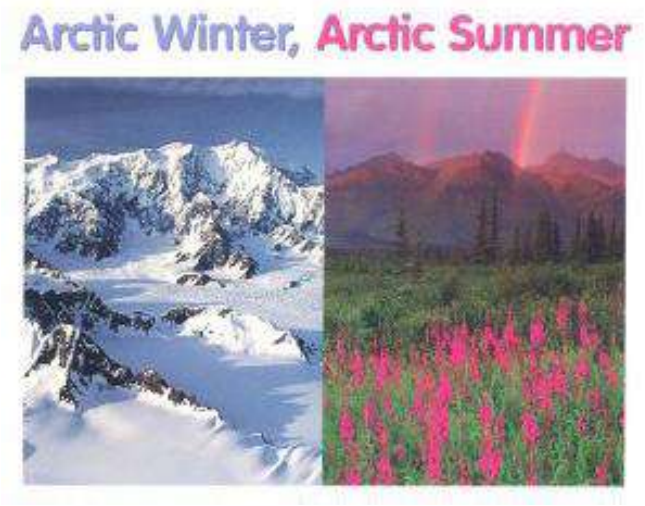

53) At places of comparable latitude North and South, the Sun behaves very differently than it would on a spinning ball Earth but precisely how it should on a flat Earth. For example, the longest summer days North of the equator are much longer than those South of the equator, and the shortest winter days North of the equator are much shorter than the shortest South of the equator. This is inexplicable on a uniformly spinning, wobbling ball Earth but fits exactly on the flat model with the Sun traveling circles over and around the Earth from Tropic to Tropic.

54) At places of comparable latitude North and South, dawn and dusk happen very differently than they would on a spinning ball, but precisely how they should on a flat Earth. In the North dawn and dusk come slowly and last far longer than in the South where they come and go very quickly. Certain places in the North twilight can last for over an hour while at comparable Southern latitudes within a few minutes the sunlight completely disappears. This is inexplicable on a uniformly spinning, wobbling ball Earth but is exactly what is expected on a flat Earth with the Sun traveling faster, wider circles over the South and slower, narrower circles over the North.

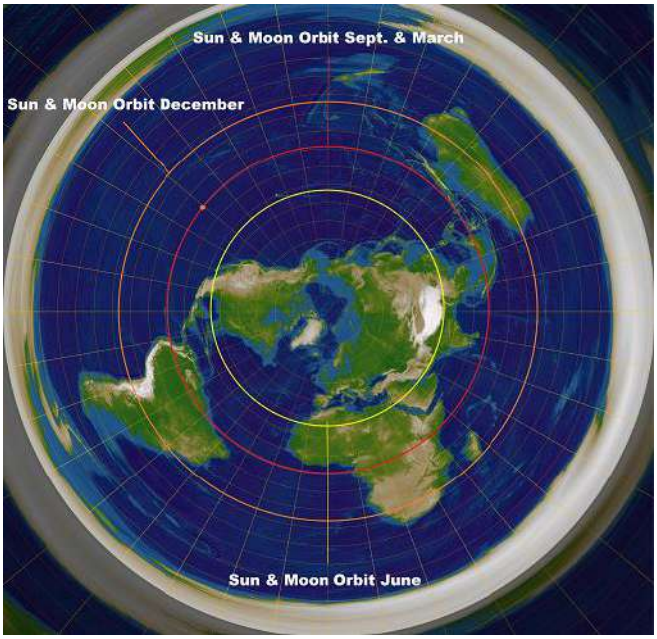

55) If the Sun circles over and around the Earth every 24 hours, steadily travelling from Tropic to Tropic every 6 months, it follows that the Northern, central region would annually receive far more heat and sunlight than the Southern circumferential region. Since the Sun must sweep over the larger Southern region in the same 24 hours it has to pass over the smaller Northern region, its passage must necessarily be proportionally faster as well. This perfectly explains the differences in Arctic/Antarctic temperatures, seasons, length of daylight, plant and animal life; this is why the Antarctic morning dawn and evening twilight are very abrupt compared with the North; and this explains why many midsummer Arctic nights the Sun does not set at all!

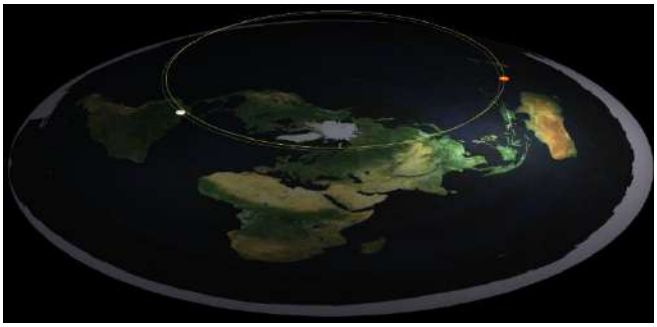

56) The "Midnight Sun" is an Arctic phenomenon occurring annually during the summer solstice where for several days straight an observer significantly far enough north can watch the Sun traveling circles over-head, rising and falling in the sky throughout the day, but never fully setting for upwards of 72+ hours! If the Earth were actually a spinning globe revolving around the Sun, the only place such a phenomenon as the Midnight Sun could be observed would be at the poles. Any other vantage point from 89 degrees latitude downwards could never, regardless of any tilt or inclination, see the Sun for 24 hours straight. To see the Sun for an entire revolution on a spinning globe at a point other than the poles, you would have to be looking through miles and miles of land and sea for part of the revolution!

57) The establishment claims the Midnight Sun IS experienced in Antarctica but they conveniently do not have any uncut videos showing this, nor do they allow independent explorers to travel to Antarctica during the winter solstice to verify or refute these claims. Conversely, there are dozens of uncut videos publicly available showing the Arctic Midnight Sun and it has been verified beyond any shadow of a doubt.

58) The Royal Belgian Geographical Society in their "Expedition Antarctique Belge," recorded that during the most severe part of the Antarctic winter, from 71 degrees South latitude onwards, the sun sets on May 17th and is not seen above the horizon again until July 21st! This is completely at odds with the ball-Earth theory, but easily explained by the flat-Earth model. The Midnight Sun is seen from high altitudes in extreme Northern latitudes during Arctic summer because the Sun, at its inner-most cycle, is circling tightly enough around the polar center that it remains visible above the horizon for someone at such a vantage point. Likewise, in extreme Southern latitudes during Arctic summer, the Sun completely disappears from view for over 2 months because there at the Northern Tropic, at the inner-most arc of its boomerang journey, the Sun is circling the Northern center too tightly to be seen from the Southern circumference.

59) Quoting Gabrielle Henriet, "*The theory of the rotation of the earth may once and for all be definitely disposed of as impracticable by pointing out the following inadvertence. It is said that the rotation takes twenty-four hours and that its speed is uniform, in which case, necessarily, days and nights should have an identical duration of twelve hours each all the year round. The sun should invariably rise in the morning and set in the evening at the same hours, with the result that it would be the equinox every day from the 1st of January to the 31st of December. One should stop and reflect on this before saying that the earth has a movement of rotation. How does the system of gravitation account for the seasonal variations in the lengths of days and nights if the earth rotates at a uniform speed in twenty-four hours!?*"

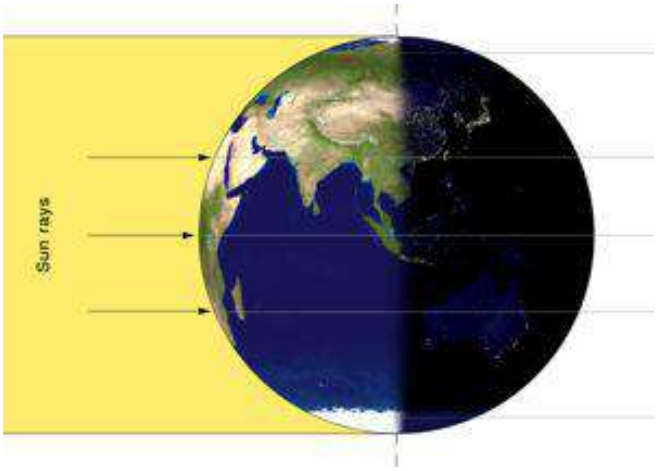

60) Anyone can prove the sea-horizon perfectly straight and the entire Earth perfectly flat using nothing more than a level, tripods and a wooden plank. At any altitude above sea-level, simply fix a 6-12 foot long, smooth, leveled board edgewise upon tripods and observe the skyline from eye-level behind it. The distant horizon will always align perfectly parallel with the upper edge of the board. Furthermore, if you move in a half-circle from one end of the board to the other whilst observing the skyline over the upper edge, you will be able to trace a clear, flat 10-20 miles depending on your altitude. This would be impossible if the Earth were a globe 25,000 miles in circumference; the horizon would align over the center of the board but then gradually, noticeably decline towards the extremities. Just ten miles on each side would necessitate an easily visible curvature of 66.6 feet from each end to the center.

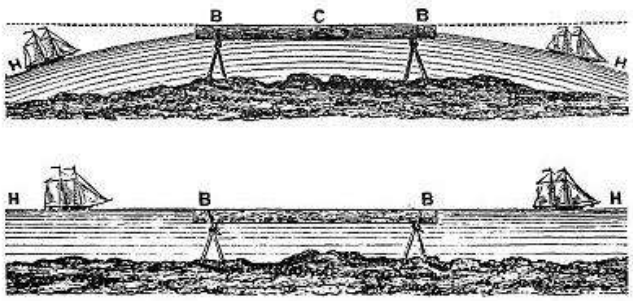

61) If the Earth were actually a big ball 25,000 miles in circumference, the horizon would be noticeably curved even at sea-level, and everything on or approaching the horizon

would appear to tilt backwards slightly from your perspective. Distant buildings along the horizon would all look like leaning towers of Piza falling away from the observer. A hot-air balloon taking off then drifting steadily away from you, on a ball-Earth would slowly and constantly appear to lean back more and more the farther away it flew, the bottom of the basket coming gradually into view as the top of the balloon disappears from sight. In reality, however, buildings, balloons, trees, people, anything and everything at right angles to the ground/horizon remains so regardless the distance or height of the observer.

62) Samuel Rowbotham's experiments at the Old Bedford Level proved conclusively the canal's water to be completely flat over a 6 mile stretch. First he stood in the canal with his telescope held 8 inches above the surface of the water, then his friend in a boat with a 5 foot tall flag sailed the 6 miles away. If Earth were a ball 25,000 miles in circumference the 6 mile stretch of water should have comprised an arc exactly 6 feet high in the middle, so the entire boat and flag should have ultimately disappeared, when in fact the entire boat and flag remained visible at the same height for the entire journey.

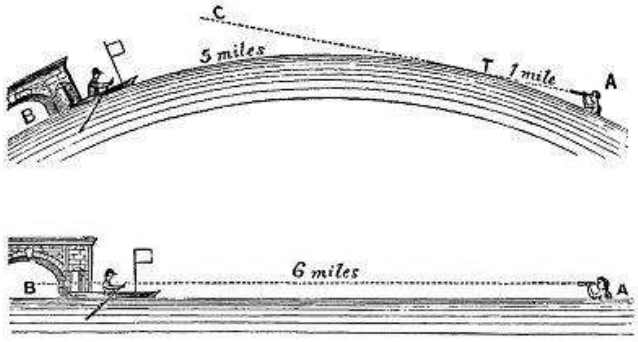

63) In a second experiment Dr. Rowbotham affixed flags 5 feet high along the shoreline, one at every mile marker. Then using his telescope mounted at 5 feet just behind the first flag looked over the tops of all 6 flags which lined up in a perfectly straight line. If the Earth were a ball 25,000 miles in circumference the flags should have progressively dipped down after the first establishing line of sight, the second would have descended 8 inches, 32 inches for the third, 6 feet for the fourth, 10 feet 8 inches for the fifth, and 16 feet 8 inches for the sixth.

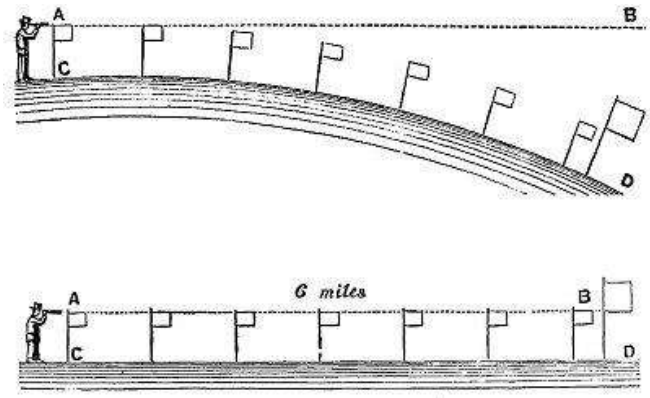

64) Quoting "Earth Not a Globe!" by Samuel Rowbotham, "*It is known that the horizon at sea, whatever distance it may extend to the right and left of the observer on land, always appears as a straight line. The following experiment has been tried in various parts of the country. At Brighton, on a rising ground near the race course, two poles were fixed in the earth six yards apart, and directly opposite the sea. Between these poles a line was tightly stretched parallel to the horizon. From the center of the line the view embraced not less than 20 miles on each side making a distance of 40 miles. A vessel was observed sailing directly westwards; the line cut the rigging a little above the bulwarks, which it did for several hours or until the vessel had sailed the whole distance of 40 miles. The ship coming into view from the east would have to ascend an inclined plane for 20 miles until it arrived at the center of the arc, whence it would have to descend for the same distance. The square of 20 miles multiplied by 8 inches gives 266 feet as the amount the vessel would be below the line at the beginning and at the end of the 40 miles.*"

65) Also Quoting Dr. Rowbotham, "*On the shore near Waterloo, a few miles to the north of Liverpool, a good telescope was fixed, at an elevation of 6 feet above the water. It was directed to a large steamer, just leaving the River Mersey, and sailing out to Dublin. Gradually the mast-head of the receding vessel came nearer to the horizon, until, at length, after more than four hours had elapsed, it disappeared. The ordinary rate of sailing of the Dublin steamers was fully eight miles an hour; so that the vessel would be, at least, thirty-two miles distant when the mast-head came to the horizon. The 6 feet of elevation of the telescope would require three miles to be deducted for convexity, which would leave twenty-nine miles, the square of which, multiplied by 8 inches, gives 560 feet; deducting 80 feet for the height of the main-mast, and we find that, according to the doctrine of rotundity, the mast-head of the outward bound steamer should have been 480 feet below the horizon. Many other experiments of this kind have been made upon sea-going steamers, and always with results entirely incompatible with the theory that the earth is a globe.*"

66) Dr. Rowbotham conducted several other experiments using telescopes, spirit levels, sextants and "theodolites," special precision instruments used for measuring angles in

horizontal or vertical planes. By positioning them at equal heights aimed at each other successively he proved over and over the Earth to be perfectly flat for miles without a single inch of curvature. His findings caused quite a stir in the scientific community and thanks to 30 years of his efforts, the shape of the Earth became a hot topic of debate around the turn of the nineteenth century.

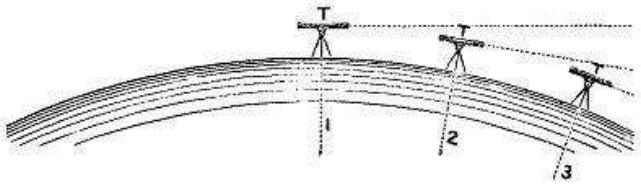

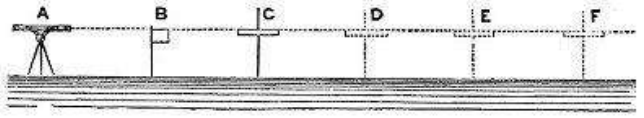

67) The distance across the Irish Sea from the Isle of Man's Douglas Harbor to Great Orm's Head in North Wales is 60 miles. If the Earth was a globe then the surface of the water between them would form a 60 mile arc, the center towering 1944 feet higher than the coastlines at either end. It is well-known and easily verifiable, however, that on a clear day, from a modest altitude of 100 feet, the Great Orm's Head is visible from Douglas Harbor. This would be completely impossible on a globe of 25,000 miles. Assuming the 100 foot altitude causes the horizon to appear approximately 13 miles off, the 47 miles remaining means the Welsh coastline should still fall an impossible 1472 feet below the line of sight!

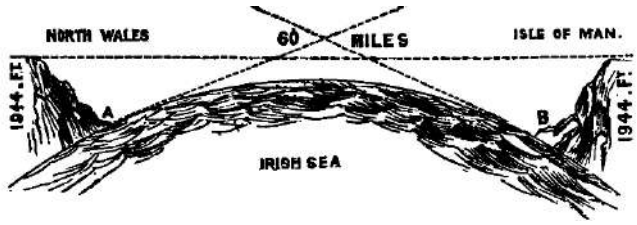

68) The Philadelphia skyline is clearly visible from Apple Pie Hill in the New Jersey Pine Barrens 40 miles away. If Earth were a ball 25,000 miles in circumference, factoring in the 205 foot elevation of Apple Pie Hill, the Philly skyline should remain well-hidden beyond 335 feet of curvature.

69) The New York City skyline is clearly visible from Harriman State Park's Bear Mountain 60 miles away. If Earth were a ball 25,000 miles in circumference, viewing from Bear Mountain's 1,283 foot summit, the Pythagorean Theorem determining distance to the horizon being 1.23 times the square root of the height in feet, the NYC skyline should be invisible behind 170 feet of curved Earth.

70) From Washington's Rock in New Jersey, at just a 400 foot elevation, it is possible on a clear day to see the skylines of both New York and Philadelphia in opposite directions at the same time covering a total distance of 120 miles! If Earth were a ball 25,000 miles in circumference, both of these skylines should be hidden behind over 800 feet of Earth's curvature.

71) It is often possible to see the Chicago skyline from sea-level 60 miles away across Lake Michigan. In 2015 after photographer Joshua Nowicki photographed this phenomenon several news channels quickly claimed his picture to be a "superior mirage," an atmospheric anomaly caused by temperature inversion. While these certainly do occur, the skyline in question was facing right-side up and clearly seen unlike a hazy illusory mirage, and on a ball-Earth 25,000 miles in circumference should be 2,400 feet below the horizon.

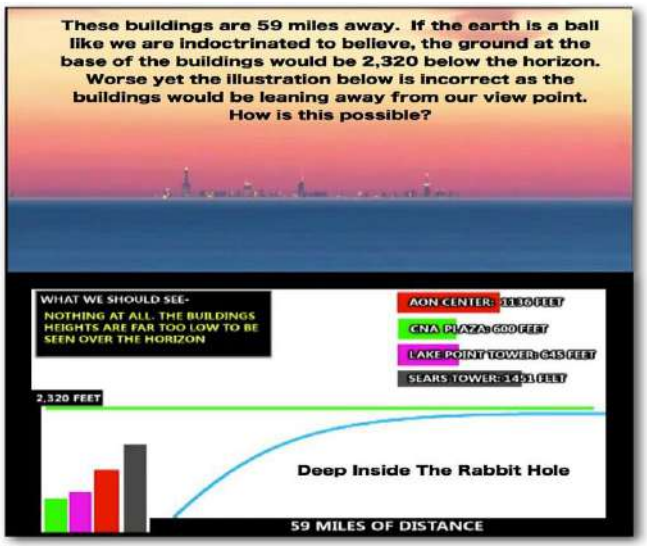

72) October 16, 1854 the Times newspaper reported the Queen's visit to Great Grimsby from Hull recording they were able to see the 300 foot tall dock tower from 70 miles away. On a ball-Earth 25,000 miles in circumference, factoring their 10 foot elevation above the water and the tower's 300 foot height, at 70 miles away the dock tower should have remained an entire 2,600 feet below the horizon.

73) In 1872 Capt. Gibson and crewmates, sailing the ship "Thomas Wood" from China to London, reported seeing the entirety of St. Helena Island on a clear day from 75 miles away. Factoring in their height during measurement on a ball-Earth 25,000 miles in circumference, it was found the island should have been 3,650 feet below their line of sight.

74) From Genoa, Italy at a height of just 70 feet above sea-level, the island of Gorgona can often be seen 81 miles away. If Earth were a ball 25,000 miles in circumference, Gorgona should be hidden beyond 3,332 feet of curvature.

75) From Genoa, Italy at a height of just 70 feet above sea-level, the island of Corsica can often be seen 99 miles away. If Earth were a ball 25,000 miles in circumference, Corsica should fall 5,245 feet, almost an entire mile below the horizon.

76) From Genoa, Italy 70 feet above sea-level, the island of Capraia 102 miles away can often be seen as well. If Earth were a ball 25,000 miles in circumference, Capraia should always remain hidden behind 5,605 feet, over a mile of supposed curvature.

77) Also from Genoa, on bright clear days, the island of Elba can be seen an incredible 125 miles away! If Earth were a ball 25,000 miles in circumference, Elba should be forever invisible behind 8770 feet of curvature.

78) From Anchorage, Alaska at an elevation of 102 feet, on clear days Mount Foraker can be seen with the naked eye 120 miles away. If Earth were a ball 25,000 miles in circumference, Mount Foraker's 17,400 summit should be leaning back away from the observer covered by 7,719 feet of curved Earth. In reality, however, the entire mountain can be quite easily seen standing straight from base to summit.

79) From Anchorage, Alaska at an elevation of 102 feet, on clear days Mount McKinley can be seen with the naked eye from 130 miles away. If Earth were a ball 25,000 miles in circumference, Mount McKinley's 20,320 foot summit should be leaning back away from the observer and almost half covered by 9,220 feet of curved Earth. In reality, however, the entire mountain can be quite easily seen standing straight from base to summit.

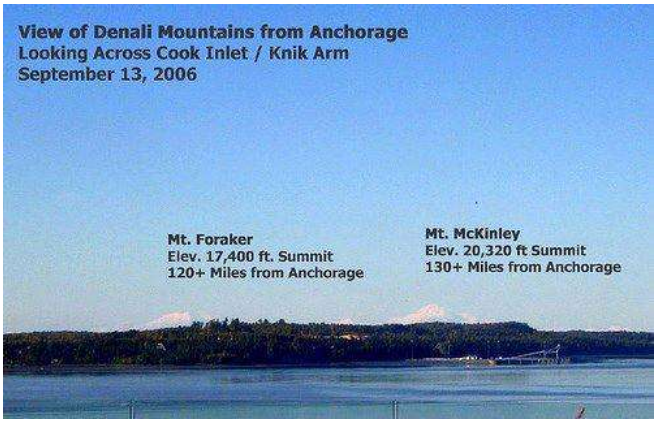

80) In Chambers' Journal, February 1895, a sailor near Mauritius in the Indian Ocean reported having seen a vessel which turned out to be an incredible 200 miles away! The incident caused much heated debate in nautical circles at the time, gaining further confirmation in Aden, Yemen where another witness reported seeing a missing Bombay steamer

from 200 miles away. He correctly stated the precise appearance, location and direction of the steamer all later corroborated and confirmed correct by those onboard. Such sightings are absolutely inexplicable if the Earth were actually a ball 25,000 miles around, as ships 200 miles distant would have to fall approximately 5 miles below line of sight!

81) The distance from which various lighthouse lights around the world are visible at sea far exceeds what could be found on a ball-Earth 25,000 miles in circumference. For example, the Dunkerque Light in southern France at an altitude of 194 feet is visible from a boat (10 feet above sea-level) 28 miles away. Spherical trigonometry dictates that if the Earth was a globe with the given curvature of 8 inches per mile squared, this light should be hidden 190 feet below the horizon.

82) The Port Nicholson Light in New Zealand is 420 feet above sea-level and visible from 35 miles away where it should be 220 feet below the horizon.

83) The Egerö Light in Norway is 154 feet above high-water and visible from 28 statute miles where it should be 230 feet below the horizon.

84) The Light at Madras, on the Esplanade, is 132 feet high and visible from 28 miles away, where it should be 250 feet below the line of sight.

85) The Cordonan Light on the west coast of France is 207 feet high and visible from 31 miles away, where it should be 280 feet below the line of sight.

86) The light at Cape Bonavista, Newfoundland is 150 feet above sea-level and visible at 35 miles, where it should be 491 feet below the horizon.

87) The lighthouse steeple of St. Botolph's Parish Church in Boston is 290 feet tall and visible from over 40 miles away, where it should be hidden a full 800 feet below the horizon!

88) The Isle of Wight lighthouse in England is 180 feet high and can be seen up to 42 miles away, a distance at which modern astronomers say the light should fall 996 feet below line of sight.

89) The Cape L'Agulhas lighthouse in South Africa is 33 feet high, 238 feet above sea level, and can be seen for over 50 miles. If the world were a globe, this light would fall 1,400 feet below an observer's line of sight.

90) The Statue of Liberty in New York stands 326 feet above sea level and on a clear day can be seen as far as 60 miles away. If the Earth were a globe, that would put Lady Liberty at an impossible 2,074 feet below the horizon.

91) The lighthouse at Port Said, Egypt, at an elevation of only 60 feet has been seen an astonishing 58 miles away, where, according to modern astronomy it should be 2,182 feet below the line of sight!

92) The Notre Dame Antwerp spire stands 403 feet high from the foot of the tower with Strasburg measuring 468 feet above sea level. With the aid of a telescope, ships can be distinguished on the horizon and captains declare they can see the cathedral spire from an amazing 150 miles away. If the Earth were a globe, however, at that distance the spire should be an entire mile, 5,280 feet below the horizon!

93) The St. George's Channel between Holyhead and Kingstown Harbor near Dublin is 60 miles across. When half-way across a ferry passenger will notice behind them

the light on Holyhead pier as well as in front of them the Poolbeg light in Dublin Bay. The Holyhead Pier light is 44 feet high, while the Poolbeg lighthouse 68 feet, therefore a vessel in the middle of the channel, 30 miles from either side standing on a deck 24 feet above the water, can clearly see both lights. On a ball Earth 25,000 miles in circumference, however, both lights should be hidden well below both horizons by over 300 feet!

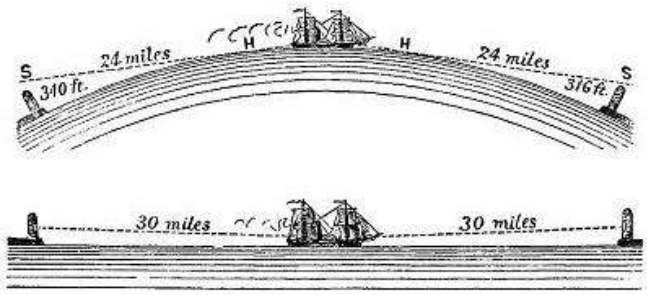

94) From the highland near Portsmouth Harbor in Hampshire, England looking across Spithead to the Isle of Wight, the entire base of the island, where water and land come together composes a perfectly straight line 22 statute miles long. According to the ball-Earth theory, the Isle of Wight should decline 80 feet from the center on each side to account for the necessary curvature. The cross-hairs of a good theodolite directed there, however, have repeatedly shown the land and water line to be perfectly level.

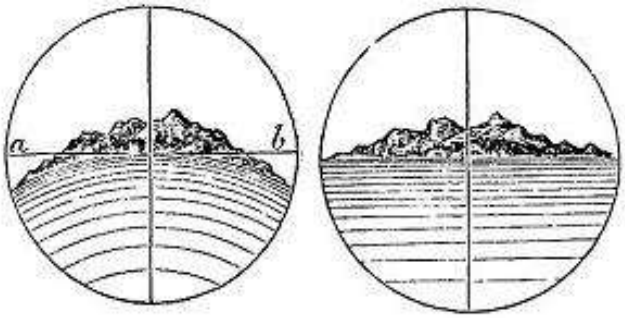

95) On a clear day from the highland near Douglas Harbor on the Isle of Man, the whole length of the coast of North Wales is often plainly visible to the naked eye. From the Point of Ayr at the mouth of the River Dee to Holyhead comprises a 50 mile stretch which has also been repeatedly found to be perfectly horizontal. If the Earth actually had curvature of 8 inches per mile squared, as NASA and modern astronomy claim, the 50 mile length of Welsh coast seen along the horizon in Liverpool Bay would have to decline from the center-point an easily detectable 416 feet on each side!

96) From "100 Proofs the Earth is Not a Globe" by William Carpenter, "*If we take a journey down the Chesapeake Bay, by night, we shall see the 'light' exhibited at Sharpe's Island for an hour before the steamer gets to it. We may take up a position on the deck so that the rail of the vessel's side will be in a line with the 'light' and in the line of sight; and we shall find that in the whole journey the light won't vary in the slightest degree in its apparent elevation. But, say that a distance of thirteen miles has been traversed, the astronomers' theory of 'curvature' demands a difference (one way or the other!) in the apparent elevation of the light, of 112 feet 8 inches! Since, however, there is not a difference of 100 hair's breadths, we have a plain proof that the water of the Chesapeake Bay is not curved, which is a proof that the Earth is not a globe.*"

97) NASA and modern astronomy say the Earth is a giant ball tilted back, wobbling and spinning 1,000 mph around its central axis, traveling 67,000 mph circles around the Sun, spiraling 500,000 mph around the Milky Way, while the entire galaxy rockets a ridiculous 670,000,000 mph through the Universe, with all of these motions originating from an alleged "Big Bang" cosmogenic explosion 14 billion years ago. That's a grand total of 670,568,000 mph in several different directions we're all supposedly speeding along at simultaneously, yet no one has ever seen, felt, heard, measured or proven a single one of these motions to exist whatsoever.

98) NASA and modern astronomy say Polaris, the North Pole star, is somewhere between 323-434 light years, or about 2 quadrillion miles, away from us! Firstly, note that is between 1,938,000,000,000,000 - 2,604,000,000,000,000 miles making a difference of 666,000,000,000,000 (over six hundred trillion) miles! If modern astronomy cannot even agree on the distance to stars within **hundreds of trillions** of miles, perhaps their "science" is flawed and their theory needs re-examining. However, even granting them their obscurely distant stars, it is impossible for heliocentrists to explain how Polaris manages to always remain perfectly aligned straight above the North Pole throughout Earth's various alleged tilting, wobbling, rotating and revolving motions.

99) Viewed from a ball-Earth, Polaris, situated directly over the North Pole, should not be visible anywhere in the Southern hemisphere. For Polaris to be seen from the Southern hemisphere of a globular Earth, the observer would have to be somehow looking "through the globe," and miles of land and sea would have to be transparent. Polaris can be seen, however, up to over 20 degrees South latitude.

100) If Earth were a ball, the Southern Cross and other Southern constellations would all be visible at the same time

from every longitude on the same latitude as is the case in the North with Polaris and its surrounding constellations. Ursa Major/Minor and many others can be seen from every Northern meridian simultaneously whereas in the South, constellations like the Southern Cross cannot. This proves the Southern hemisphere is not "turned under" as in the ball-Earth model, but simply stretching further outwards away from the Northern center-point as in the flat Earth model.

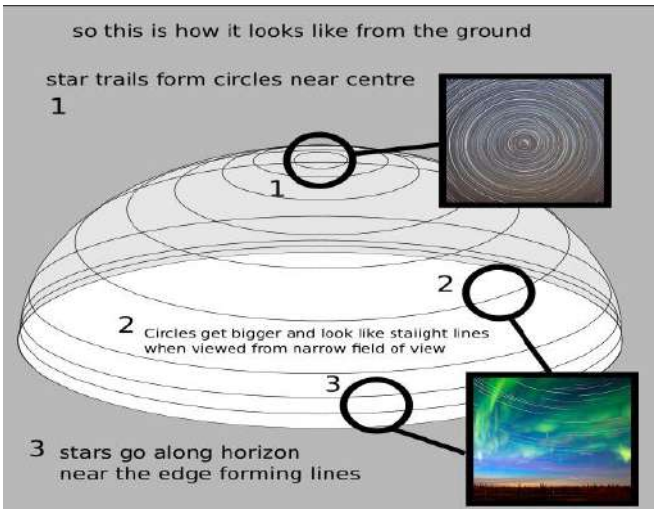

101) Sigma Octantis is claimed to be a Southern central pole star similar to Polaris, around which the Southern hemisphere stars all rotate around the opposite direction. Unlike Polaris, however, Sigma Octantis can NOT be seen simultaneously from every point along the same latitude, it is NOT central but allegedly 1 degree off-center, it is NOT motionless, and in fact cannot be seen at all using publicly available telescopes! There is legitimate speculation regarding whether Sigma Octantis even exists. Either way, the direction in which stars move overhead is based on perspective and the exact direction you're facing, not which hemisphere you are in.

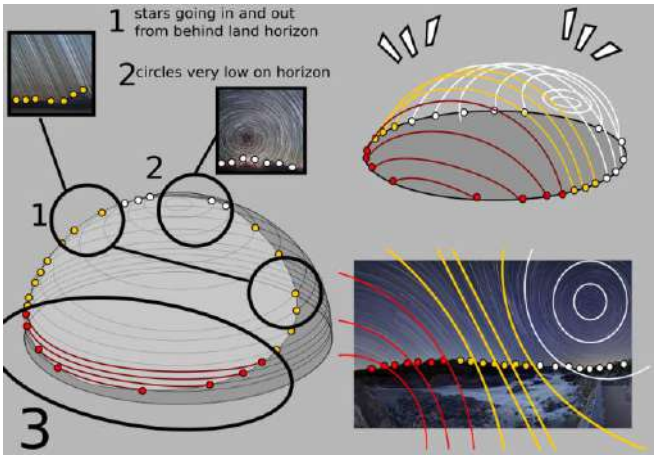

102) Some heliocentrists have tried to suggest that the Pole Star's gradual declination overhead as an observer travels southwards is proof of a globular Earth. Far from it, the declination of the Pole Star or any other object is simply a result of the Law of Perspective on plane (flat) surfaces. The Law of Perspective dictates that the angle and height at which an object is seen diminishes the farther one recedes from the object, until at a certain point the line of sight and the seemingly uprising surface of the Earth converges to a vanishing point (i.e. the horizon line) beyond which the object is invisible. In the ball-Earth model the horizon is claimed to be the curvature of the Earth, whereas in reality, the horizon is known to be simply the vanishing line of perspective based on the strength of your eyes, instruments, weather and altitude.

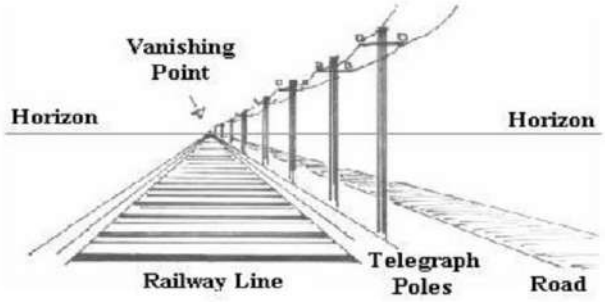

103) There are several constellations which can be seen from far greater distances over the face of the Earth than should be possible if the world were a rotating, revolving, wobbling ball. For instance, Ursa Major, very close to Polaris, can be seen from 90 degrees North latitude (the North Pole) all the way down to 30 degrees South latitude. For this to be possible on a ball-Earth the Southern observers would have to be seeing through hundreds or thousands of miles of bulging Earth to the Northern sky.

104) The constellation Vulpecula can be seen from 90 degrees North latitude, all the way to 55 degrees South latitude. Taurus, Pisces and Leo can be seen from 90 degrees North all the way to 65 degrees South. An observer on a ball-Earth, regardless of any tilt or inclination, should not logically be able to see this far.

105) Aquarius and Libra can be seen from 65 degrees North to 90 degrees South! The constellation Virgo is visible from 80 degrees North down to 80 degrees South, and Orion can be seen from 85 degrees North all the way to 75 degrees South latitude! These are all only possible because the "hemispheres" are not spheres at all but concentric circles of latitude extending outwards from the central North Pole with the stars rotating over and around.

106) The so-called "South Pole" is simply an arbitrary point along the Antarctic ice marked with a red and white barbershop pole topped with a metal ball-Earth. This ceremonial South Pole is admittedly and provably NOT the actual South Pole, however, because the actual South Pole could be demonstrably confirmed with the aid of a compass showing North to be 360 degrees around the observer. Since this feat has never been achieved, the model remains pure theory, along with the establishment's excuse that the geomagnetic poles supposedly constantly move around making verification of their claims impossible.

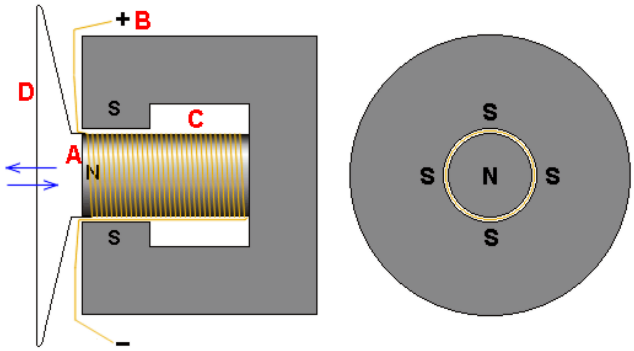

107) Ring magnets of the kind found in loudspeakers have a central North pole with the opposite "South" pole actually being all points along the outer circumference. This perfectly demonstrates the magnetism of our flat Earth, whereas the alleged source of magnetism in the ball-Earth model is emitted from a hypothetical molten magnetic core in the center of the ball which they claim conveniently causes both poles to constantly move thus evading independent verification at their two "ceremonial poles." In reality the deepest drilling operation in history, the Russian Kola Ultradeep, managed to get only 8 miles down, so the entire ball-Earth model taught in schools showing a crust, outer-mantle, inner-mantle, outer-core and inner-core layers are all purely speculation as we have never penetrated through beyond the crust.

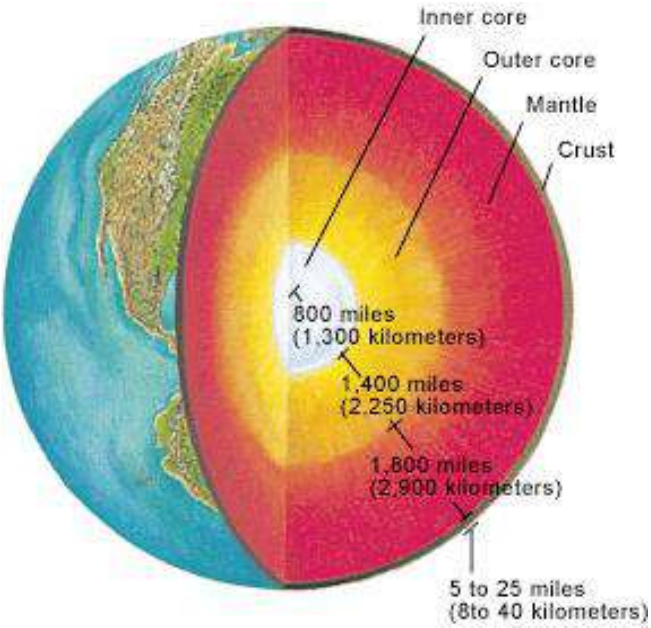

108) The mariner's compass is an impossible and non-sensical instrument for use on a ball-Earth. It simultaneously points North and South over a flat surface, yet claims to be pin-pointing two constantly moving geomagnetic poles at opposite ends of a spinning sphere originating from a hypothetical molten metal core. If compass needles were actually drawn to the North pole of a globe, the opposing "South" needle would actually be pointing up and off into outer-space.

109) There are no fixed "East" or "West" points just as there is no fixed "South." The North central Pole is the only proven fixed point on our flat Earth, with South being all straight lines outwards from the pole, East and West being concentric circles at constant right angles 90 degrees from the pole. A westerly circumnavigation of Earth is thus going around with Polaris continually on your right, while an easterly circumnavigation is going around with Polaris always at your left.

110) Magellan and others' East/West circumnavigations of Earth are often quoted as proof of the ball model. In actual fact, however, sailing or flying at rights angles to the North pole and eventually returning to one's original location is no more difficult or mysterious than doing so on a globe. Just as an architect's compass can place its center-point on a flat piece of paper and trace a circle either way around the "pole," so can a ship or plane circumnavigate a flat-Earth.

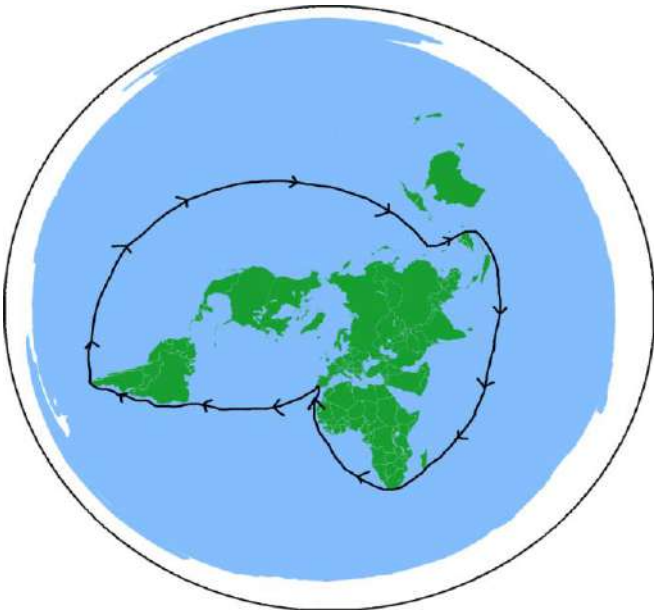

111) Since the North Pole and Antarctica are covered in ice and guarded "no-fly" zones, no ships or planes have ever been known to circumnavigate the Earth in North/South

directions. The only kind of circumnavigation which could not happen on a flat-Earth is North/Southbound, which is likely the very reason for the heavily-enforced flight restrictions. The fact that there has yet to be a single verified North/South circumnavigation of Earth serves as standing proof the world is not a ball.

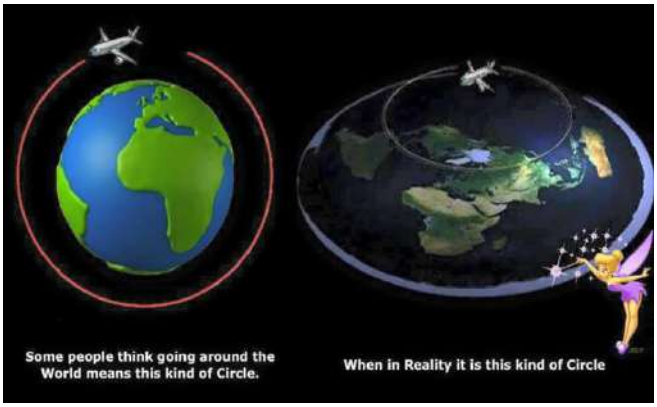

112) The Sun brings noon to every time-zone as it passes directly over-head every 15 degree demarcation point, 24 times per day in its circular path over and around the Earth. If time-zones were instead caused by the uniform spinning of the ball-Earth around the Sun, every 6 months as Earth found itself on the opposite side of the Sun, clocks all over Earth would have to flip 12 hours, day would be night and night would be day.

113) The idea that people are standing, ships are sailing and planes are flying upside down on certain parts of Earth while others tilted at 90 degrees and all other impossible angles is complete absurdity. The idea that a man digging a hole straight down could eventually reach sky on the other side is ludicrous. Common sense tells every free-thinking person correctly that there truly is an "up" and "down" in nature, unlike the "everything is relative" rhetoric of the Newtonian/Einsteinian paradigm.

114) Quoting, "On the False Wisdom of the Philosophers" by Lacantius, "*A sphere where people on the other side live with their feet above their heads, where rain, snow and hail fall upwards, where trees and crops grow upside-down and the sky is lower than the ground? The ancient wonder of the hanging gardens of Babylon dwindle into nothing in comparison to the fields, seas, towns and mountains that pagan philosophers believe to be hanging from the earth without support!*"

115) The existing laws of density and buoyancy perfectly explained the physics of falling objects long before knighted Freemason "Sir" Isaac Newton bestowed his theory of "gravity" upon the world. It is a fact that objects placed in denser mediums rise up while objects placed in less dense mediums sink down. To fit with the heliocentric model which has no up or down, Newton instead claimed objects are attracted to large masses and fall towards the center. Not a single experiment in history, however, has shown an object massive enough to, by virtue of its mass alone, cause other smaller masses to be attracted to it as Newton claims "gravity" does with Earth, the Sun, Moon, Stars and Planets.

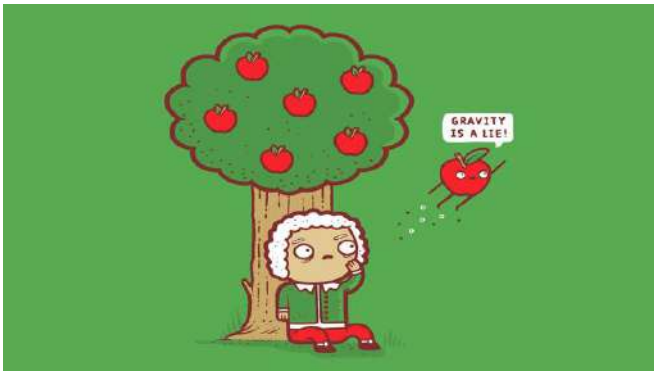

116) There has also never been a single experiment in history showing an object massive enough to, by virtue of its mass alone, cause another smaller mass to orbit around it. The magic theory of gravity allows for oceans, buildings and people to remain forever stuck to the underside of a spinning ball while simultaneously causing objects like the Moon and satellites to remain locked in perpetual circular orbits around the Earth. If these were both true then people should be able to jump up and start orbiting circles around the Earth, or the Moon should have long ago been sucked into the Earth. Neither of these theories have ever been experimentally verified and their alleged results are mutually exclusive.

117) Newton also theorized and it is now commonly taught that the Earth's ocean tides are caused by gravitational lunar attraction. If the Moon is only 2,160 miles in diameter and

the Earth 8,000 miles, however, using their own math and "law," it follows that the Earth is 87 times more massive and therefore the larger object should attract the smaller to it, and not the other way around. If the Earth's greater gravity is what keeps the Moon in orbit, it is impossible for the Moon's lesser gravity to supersede the Earth's gravity, especially at Earth's sea-level, where its gravitational attraction would even further out-trump the Moon's. And if the Moon's gravity truly did supersede the Earth's causing the tides to be drawn towards it, there should be nothing to stop them from continuing onwards and upwards towards their great attractor.

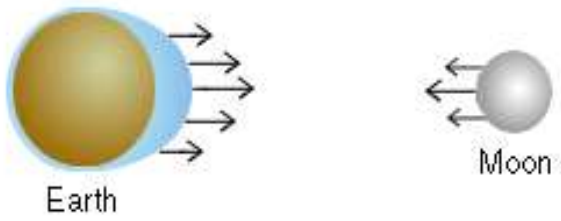

118) Furthermore, the velocity and path of the Moon are uniform and should therefore exert a uniform influence on the Earth's tides, when in actuality the Earth's tides vary greatly and do not follow the Moon. Earth's lakes, ponds, marshes and other inland bodies of water also inexplicably remain forever outside the Moon's gravitational grasp! If "gravity" was truly drawing Earth's oceans up to it, all lakes, ponds and other bodies of standing water should certainly have tides as well.

119) It is claimed that the other planets are spheres and so therefore Earth must also be a sphere. Firstly, Earth is a "plane" not a "planet," so the shape of these "planets" in the sky have no bearing on the shape of the Earth beneath our feet. Secondly, these "planets" have been known for thousands of years around the world as the "wandering stars" since they differ from the other fixed stars in their relative motions only. When looked at with an unprejudiced naked-eye or through a telescope, the fixed and wandering stars appear as luminous discs of light, NOT spherical terra firma. The pictures and videos shown by NASA of spherical terra firma planets are all clearly fake computer-generated images, and NOT photographs.

120) The etymology of the word "planet" actually comes from late Old English planete, from Old French planete (Modern French planète), from Latin planeta, from Greek planetes, from (asteres) planetai "wandering (stars)," from planasthai "to wander," of unknown origin, possibly from PIE *pele "flat, to spread" or notion of "spread out." And Plane (n) "flat surface," c. 1600, from Latin planum "flat surface, plane, level, plain," planus "flat, level, even, plain, clear." They just added a "t" to our Earth plane and everyone bought it.

121) When you observe the Sun and Moon you see two equally-sized equidistant circles tracing similar paths at similar speeds around a flat, stationary Earth. The "experts" at NASA, however, claim your common sense every day experience is false on all counts! To begin with, they say the Earth is not flat but a big ball; not stationary but spinning around 19 miles per second; they say the Sun does not revolve around the Earth as it appears, but Earth revolves around the Sun; the Moon, on the other hand, does revolve around the Earth, though not East to West as it appears, rather West to East; and the Sun is actually 400 times larger than the Moon and 400 times farther away! You can clearly see they are the same size and distance, you can see the Earth is flat, you can feel the Earth is stationary, but according to the gospel of modern astronomy, you are wrong and a simpleton worthy of endless ridicule if you dare to trust your own eyes and experience.

122) Quoting Allen Daves, "*If the Government or NASA had said to you that the Earth is stationary, imagine that. And then imagine we are trying to convince people that 'no, no it's not stationary, it's moving forward at 32 times rifle bullet speed and spinning at 1,000 miles per hour.' We would be laughed at! We would have so many people telling us 'you are crazy, the Earth is not moving!' We would be ridiculed for having no scientific backing for this convoluted moving Earth theory. And not only that but then people would say, 'oh then how do you explain a fixed, calm atmosphere and the Sun's observable movement, how do you explain that?' Imagine saying to people, 'no, no, the atmosphere is moving also but is somehow magically velcroed to the moving-Earth. The reason is not simply because the Earth is stationary.' So what we are actually doing is what makes sense. We are saying that the moving-Earth theory is nonsense. The stationary-Earth theory makes sense and we are being ridiculed. You've got to picture it being the other way around to realize just how RIDICULOUS this situation is. This theory from the*

Government and NASA that the Earth is rotating and orbiting and leaning over and wobbling is absolute nonsense and yet people are clinging to it, tightly, like a teddy bear. They just can't bring themselves to face the possibility that the Earth is stationary though ALL the evidence shows it: we feel no movement, the atmosphere hasn't been blown away, we see the Sun move from East-to-West, everything can be explained by a motionless Earth without bringing in all these assumptions to cover up previous assumptions gone bad."

123) Heliocentrists' astronomical figures always sound perfectly precise, but they have historically been notorious for regularly and drastically changing them to suit their various models. For instance, in his time Copernicus calculated the Sun's distance from Earth to be 3,391,200 miles. The next century Johannes Kepler decided it was actually 12,376,800 miles away. Issac Newton once said, "*It matters not whether we reckon it 28 or 54 million miles distant for either would do just as well!*" How scientific!? Benjamin Martin calculated between 81 and 82 million miles, Thomas Dilworth claimed 93,726,900 miles, John Hind stated positively 95,298,260 miles, Benjamin Gould said more than 96 million miles, and Christian Mayer thought it was more than 104 million! Flat-Earthers throughout the ages, conversely, have used sextants and plane trigonometry to make such calculations and found the Sun and Moon both to be only about 32 miles in diameter and less than a few thousand miles from Earth.

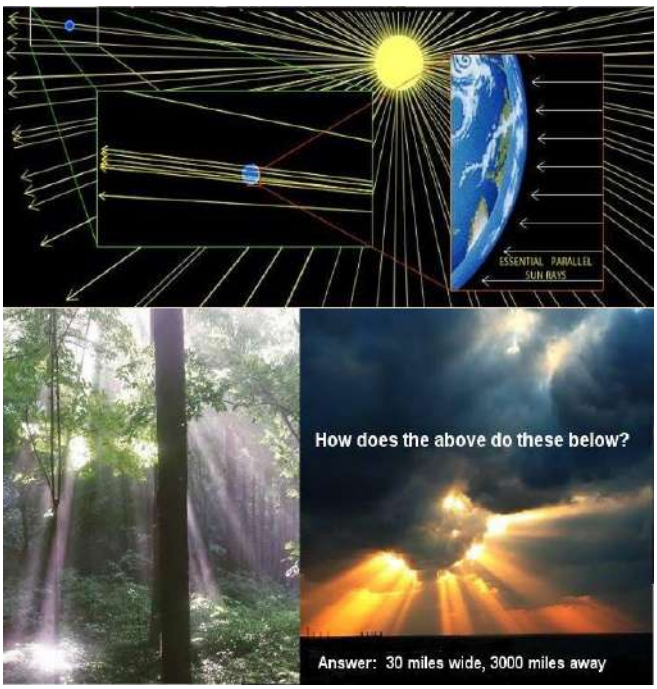

124) Amateur balloon footage taken above the clouds has provided stunning visual proof that the Sun cannot be millions of miles away. In several shots you can see a clear hot-spot reflecting on the clouds directly below the Sun's spotlight-like influence. If the Sun were actually millions of miles away such a small, localized hot-spot could not occur.

125) Another proof the Sun is not millions of miles away is found by tracing the angle of sun-rays back to their source above the clouds. There are thousands of pictures showing how sunlight comes down through cloud-cover at a variance of converging angles. The area of convergence is of course the Sun, and is clearly NOT millions of miles away, but rather relatively close to Earth just above the clouds.

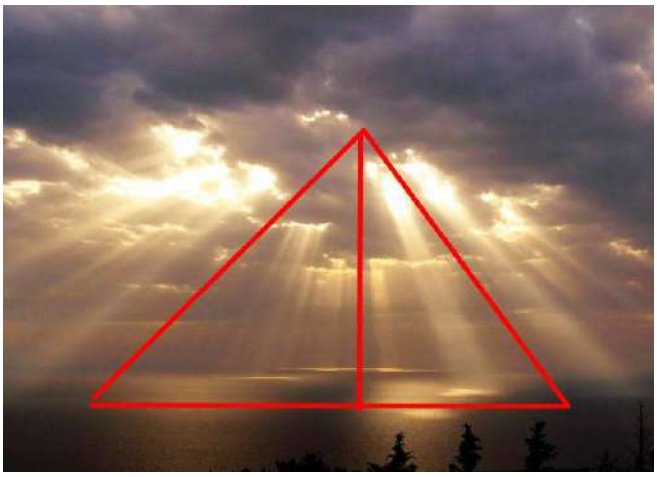

126) The Sun's annual journey from tropic to tropic, solstice to solstice, is what determines the length and character of days, nights and seasons. This is why equatorial regions experience almost year-round summer and heat while higher latitudes North and especially South experience more distinct seasons with harsh winters. The heliocentric model claims seasons change based on the ball-Earth's alleged "axial tilt" and "elliptical orbit" around the Sun, yet their flawed current model places us closest to the Sun (91,400,000 miles) in January when its actually winter, and farthest from the Sun (94,500,000 miles) in July when its actually summer throughout most of the Earth.

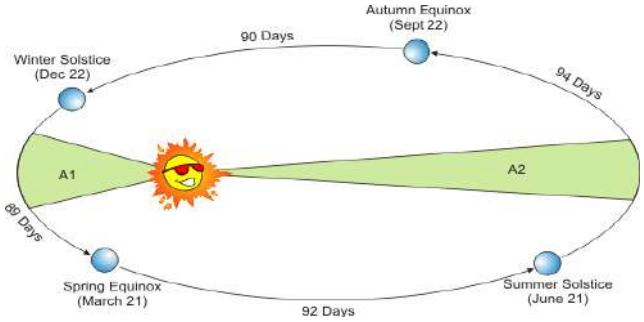

127) The fact that the Sun and Moon's reflections on water always form a straight line path from the horizon to the observer proves the Earth is not a ball. If Earth's surface was curved it would be impossible for the reflected light to curve over the ball from horizon to observer.

128) There are huge centuries-old stone sundials and moondials all over the world which still tell the time now down to the minute as perfectly as the day they were made. If the Earth, Sun and Moon were truly subject to the number of contradictory revolving, rotating, wobbling and spiraling motions claimed by modern astronomy, it would be impossible for these monuments to so accurately tell time without constant adjustment.

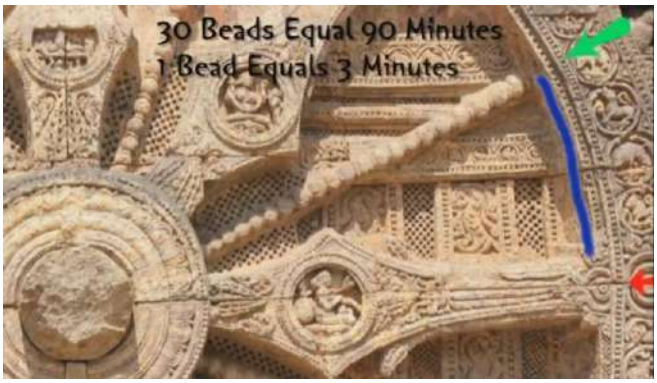

129) To quote William Carpenter, "*Why, in the name of common sense, should observers have to fix their telescopes on solid stone bases so that they should not move a hair's-breadth, - if the Earth on which they fix them moves at the rate of nineteen miles in a second? Indeed, to believe that 'six thousand million million million tons' is 'rolling, surging, flying, darting on through space for ever' with a velocity compared with which a shot from a cannon is a 'very slow coach,' with such unerring accuracy that a telescope fixed on granite pillars in an observatory will not enable a lynx-eyed astronomer to detect a variation in its onward motion of the thousandth part of a hair's-breadth is to conceive a miracle compared with which all the miracles on record put together would sink into utter insignificance. Since we can, (in middle north latitudes), see the North Star, on looking out of a window that faces it - and out of the very same corner of the very same pane of glass in the very same window - all the year round, it is proof enough for any man in his senses that we have made no motion at all and that the Earth is not a globe*."

130) From "Earth Not a Globe!" by Samuel Rowbotham, *"Take two carefully-bored metallic tubes, not less than six feet in length, and place them one yard asunder, on the opposite sides of a wooden frame, or a solid block of wood or stone: so adjust them that their centres or axes of vision shall be perfectly parallel to each other. Now, direct them to the plane of some notable fixed star, a few seconds previous to its meridian time. Let an observer be stationed at each tube and the moment the star appears in the first tube let a loud knock or other signal be given, to be repeated by the observer at the second tube when he first sees the same star. A distinct period of time will elapse between the signals given. The signals will follow each other in very rapid succession, but still, the time between is sufficient to show that the same star is not visible at the same moment by two parallel lines of sight when only one yard asunder. A slight inclination of the second tube towards the first tube would be required for the star to be seen through both tubes at the same instant. Let the tubes remain in their position for six months; at the end of which time the same observation or experiment will produce the same results--the star will be visible at the same meridian time, without the slightest alteration being required in the direction of the tubes: from which it is concluded that if the earth had moved one single yard in an orbit through space, there would at least be observed the slight inclination of the tube which the difference in position of one yard had previously required. But as no such difference in the direction of the tube is required,* ***the conclusion is unavoidable, that in six months a given meridian upon the earth's surface does not move a single yard, and therefore, that the earth has not the slightest degree of orbital motion****."*

131) NASA and modern astronomy maintain that the Moon is a solid, spherical, Earth-like habitation which man has actually flown to and set foot on. They claim the Moon is a non-luminescent planetoid which receives and reflects all its light from the Sun. The reality is, however, that the Moon is observably not a solid body, it is clearly circular, but not spherical, and not in any way an Earth-like planetoid which humans could set foot on. In fact, the Moon has been proven largely transparent and completely self-luminescent, shining with its own unique light.

132) The Sun's light is golden, warm, drying, preservative and antiseptic, while the Moon's light is silver, cool, damp, putrefying and septic. The Sun's rays decrease the combustion of a bonfire, while the Moon's rays increase combustion. Plant and animal substances exposed to sunlight quickly dry, shrink, coagulate, and lose the tendency to decompose and putrify; grapes and other fruits become solid, partially candied and preserved like raisins, dates, and prunes; animal flesh coagulates, loses its volatile gaseous constituents, becomes firm, dry, and slow to decay. When exposed to moonlight, however, plant and animal substances tend to show symptoms of putrefaction and decay. This proves that Sun and Moon light are different, unique, and opposites as they are in the geocentric flat model.

133) In direct sunlight a thermometer will read higher than another thermometer placed in the shade, but in full, direct moonlight a thermometer will read lower than another placed in the shade. If the Sun's light is collected in a large lens and thrown to a focus point it can create significant heat, while the Moon's light collected similarly creates no heat. In the "Lancet Medical Journal," from March 14th, 1856, particulars are given of several experiments which proved the Moon's rays when concentrated can actually reduce the temperature upon a thermometer more than eight degrees. So sunlight and moonlight clearly have altogether different properties.

134) Furthermore the Moon itself cannot physically be both a spherical body and a reflector of the Sun's light. Reflectors must be flat or concave for light rays to have any angle of incidence; If a reflector's surface is convex then every ray of light points in a direct line with the radius perpendicular to the surface resulting in no reflection.

135) Not only is the Moon clearly self-luminescent, shining its own unique light, but it is also largely transparent. When the waxing or waning Moon is visible during the day it is possible to see the blue sky right through the Moon. And on a clear night, during a waxing or waning cycle, it is even possible to occasionally see stars and "planets" directly through the surface of the Moon! The Royal Astronomical Society has on record many such occurrences throughout history which all defy the heliocentric model.

136) Many people think that modern astronomy's ability to accurately predict lunar and solar eclipses is a result and proof positive of the heliocentric theory of the universe. The fact of the matter however is that eclipses have been accurately predicted by cultures worldwide for thousands of years before the "heliocentric ball-Earth" was even a glimmer in Copernicus' imagination. Ptolemy in the 1st century A.D. accurately predicted eclipses for six hundred years on the basis of a flat, stationary Earth with equal precision as anyone living today. All the way back in 600 B.C. Thales accurately predicted an eclipse which ended the war between the Medes and Lydians. Eclipses happen regularly with precision in 18 year cycles, so regardless of geocentric or heliocentric, flat or globe Earth cosmologies, eclipses can be accurately calculated independent of such factors.

137) Another assumption and supposed proof of Earth's shape, heliocentrists claim that lunar eclipses are caused by the shadow of the ball-Earth occulting the Moon. They claim the Sun, Earth, and Moon spheres perfectly align like three billiard balls in a row so that the Sun's light casts the Earth's shadow onto the Moon. Unfortunately for heliocentrists, this explanation is rendered completely invalid due to the fact that lunar eclipses have happened and continue to happen regularly when both the Sun and Moon are still visible together above the horizon! For the Sun's light to be casting Earth's shadow onto the Moon, the three bodies must be aligned in a straight 180 degree syzygy, but as early as the time of Pliny, there are records of lunar

eclipses happening while both the Sun and Moon are visible in the sky. Therefore the eclipsor of the Moon cannot be the Earth/Earth's shadow and some other explanation must be sought.

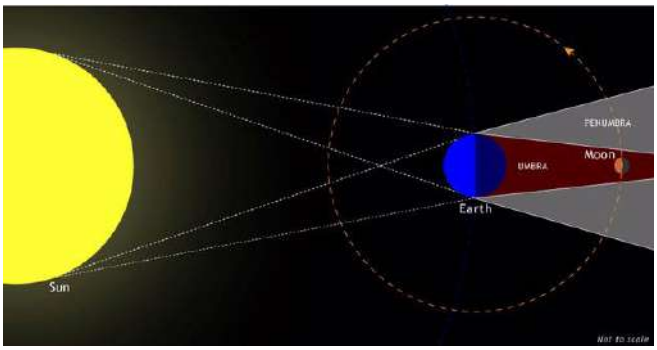

138) Another favorite "proof" of ball-Earthers is the appearance from an observer on shore of ships' hulls being obfuscated by the water and disappearing from view when sailing away towards the horizon. Their claim is that ships' hulls disappear before their mast-heads because the ship is beginning its declination around the convex curvature of the ball-Earth. Once again, however, their hasty conclusion is drawn from a faulty premise, namely that only on a ball-Earth could this phenomenon occur. The fact of the matter is that the Law of Perspective on plane surfaces dictates and necessitates the exact same occurrence. For example a girl wearing a dress walking away towards the horizon will appear to sink into the Earth the farther away she walks. Her feet will disappear from view first and the distance between the ground and the bottom of her dress will gradually diminish until after about half a mile it seems like her dress is touching the ground as she walks on invisible legs. Such is the case on plane surfaces, the lowest parts of objects receding from a given point of observation necessarily disappear before the highest.

139) Not only is the disappearance of ship's hulls explained by the Law of Perspective on flat surfaces, it is proven undeniably true with the aid of a good telescope. If you watch a ship sailing away into the horizon with the naked eye until its hull has completely disappeared from view under the supposed "curvature of the Earth," then look through a telescope, you will notice the entire ship quickly zooms back into view, hull and all, proving that the disappearance was caused by the Law of Perspective, not by a wall of curved water! This also proves that the horizon is simply the vanishing line of perspective from your point of view, NOT the alleged "curvature" of Earth.

140) Foucault's Pendulums are often quoted as proof of a rotating Earth but upon closer investigation prove the opposite. To begin with, Foucault's pendulums do not uniformly swing in any one direction. Sometimes they rotate clockwise and sometimes counter-clockwise, sometimes they fail to rotate and sometimes they rotate far too much. The behavior of the pendulum actually depends on 1) the initial force beginning its swing and, 2) the ball-and-socket joint used which most-readily facilitates circular motion over any other. The supposed rotation of the Earth is completely inconsequential and irrelevant to the pendulum's swing. If the alleged constant rotation of the Earth affected pendulums in any way, then there should be no need to manually start pendulums in motion. If the Earth's diurnal rotation caused the 360 degree uniform diurnal rotation of pendulums, then there should not exist a stationary pendulum anywhere on Earth!

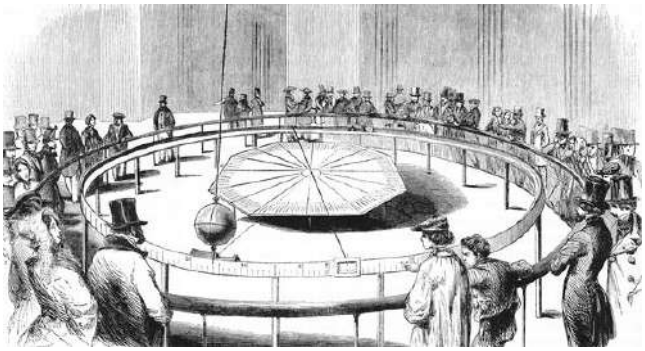

141) The "Coriolis Effect" is often said to cause sinks and toilet bowls in the Northern Hemisphere to drain spinning in one direction while in the Southern Hemisphere causing them to spin the opposite way, thus providing proof of the spinning ball-Earth. Once again, however, just like Foucault's Pendulums spinning either which way, sinks and toilets in the Northern and Southern hemispheres do not consistently spin in any one direction! Sinks and toilets in the very same household are often found to spin opposite directions, depending entirely upon the shape of the basin and the angle of the water's entry, not the supposed rotation of the Earth.

142) People claim that if the Earth were flat, they should be able to use a telescope and see clear across the oceans! This is absurd, however, as the air is full of precipitation especially over the oceans, and especially at the lowest, densest layer of atmosphere is NOT transparent. Picture the blurry haze over roads on hot, humid days. Even the best telescope will blur out long before you could see across an ocean. You can, however, use a telescope to zoom in

MUCH more of our flat Earth than would be possible on a ball 25,000 miles in circumference.

143) People claim that if the Earth were flat, with the Sun circling over and around us, we should be able to see the Sun from everywhere all over the Earth, and there should be daylight even at night-time. Since the Sun is NOT 93 million miles away but rather just a few thousand and shining down like a spotlight, once it has moved significantly far enough away from your location it becomes invisible beyond the horizon and daylight slowly fades until it completely disappears. If the Sun were 93 million miles away and the Earth a spinning ball, the transition from daylight to night would instead be almost instantaneous as you passed the terminator line.

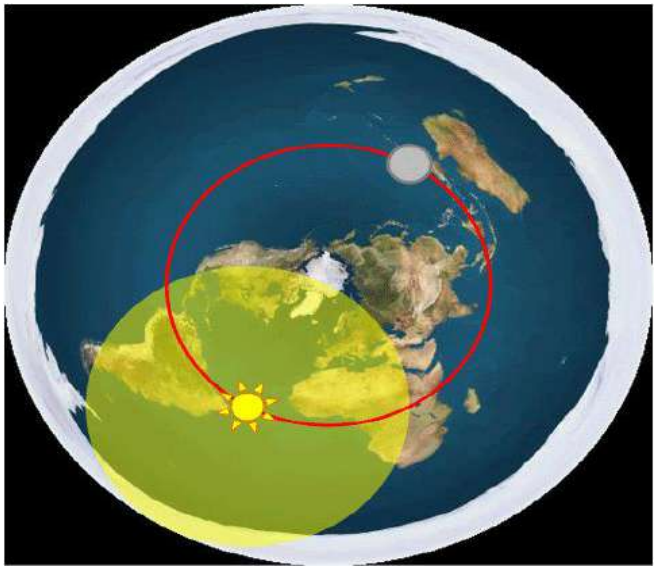

144) Pictures of the Moon appearing upside-down in the Southern hemisphere and right-side up in the North are often cited as proof of the ball-Earth, but once again, upon closer inspection, provide another proof of the flat model. In fact, time-lapse photography shows the Moon itself turns clockwise like a wheel as it circles over and around the Earth. You can find pictures of the Moon at 360 degrees of various inclination from all over the Earth simply depending on where and when the picture was taken.

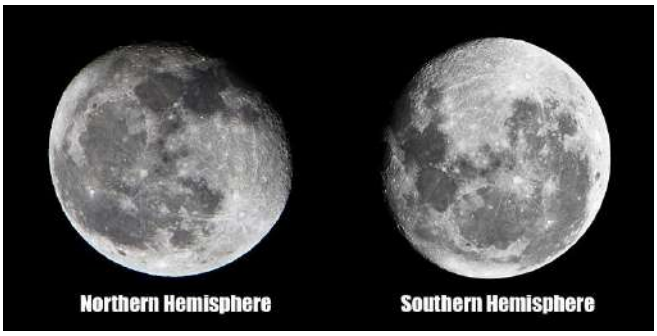

145) Heliocentrists believe the Moon is a ball, even though its appearance is clearly that of a flat luminous disc. We only ever see the same one face (albeit at various inclinations) of the Moon, yet it is claimed that there is another "dark side of the Moon" which remains hidden. NASA states the Moon spins opposite the spin of the Earth in such a perfectly synchronized way that the motions cancel each other out so we will conveniently never be able to observe the supposed dark-side of the Moon outside of their terrible fake CGI images. The fact of the matter is, however, if the Moon were a sphere, observers in Antarctica would see a different face from those at the equator, yet they do not – just the same flat face rotated at various degrees.

146) The ball-Earth model claims the Moon orbits around the Earth once every 28 days, yet it is plain for anyone to see that the Moon orbits around the Earth every single day! The Moon's orbit is slightly slower than the Sun's, but follows the Sun's same path from Tropic to Tropic, solstice to solstice, making a full circle over the Earth in just under 25 hours.

147) The ball-Earth model claims the Sun is precisely 400 times larger than the Moon and 400 times further away from Earth making them "falsely" appear exactly the same size. Once again, the ball model asks us to accept as coincidence something that cannot be explained other than by natural design. The Sun and the Moon occupy the same amount of space in the sky and have been measured with sextants to be of equal size and equal distance, so claiming otherwise is against our eyes, experience, experiments and common sense.

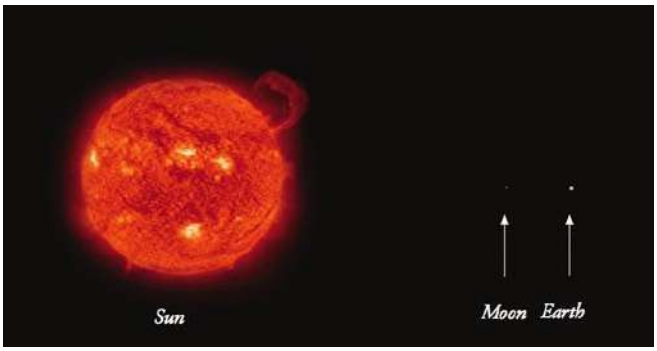

148) Quoting "Earth Not a Globe!" by Samuel Rowbotham, "*It is found by observation that the stars come to the meridian about four minutes earlier every twenty-four hours than the sun, taking the solar time as the standard. This makes 120 minutes every thirty days, and twenty-four hours in the year. Hence all the constellations have passed before or in advance of the sun in that time. This is the simple fact as observed in nature, but the theory of rotundity and*

motion on axes and in an orbit has no place for it. Visible truth must be ignored, because this theory stands in the way, and prevents its votaries from understanding it."

149) Throughout thousands of years the same constellations have remained fixed in their same patterns without moving out of position whatsoever. If the Earth were a big ball spinning around a bigger Sun spinning around a bigger galaxy shooting off from the Biggest Bang as NASA claims, it is impossible that the constellations would remain so fixed. Based on their model, we should, in fact, have an entirely different night sky every single night and never repeat exactly the same star pattern twice.

150) If Earth were a spinning ball it would be impossible to photograph star-trail time-lapses turning perfect circles around Polaris anywhere but the North Pole. At all other vantage points the stars would be seen to travel more or less horizontally across the observer's horizon due to the alleged 1000mph motion beneath their feet. In reality, however, Polaris's surrounding stars can always be photographed turning perfect circles around the central star all the way down to the Tropic of Capricorn.

151) If Earth were a spinning ball revolving around the Sun it would actually be impossible for star-trail photos to show perfect circles even at the North Pole! Since the Earth is also allegedly moving 67,000mph around the Sun, the Sun moving 500,000mph around the Milky Way, and the entire galaxy going 670,000,000mph, these four contradictory motions would make star-trail time-lapses all show irregular curved lines.

152) In 2003, three University Geography professors collaborated in an experiment to prove that the state of Kansas is indeed actually flatter than a pancake! Using topigraphical geodetic surveys covering over 80,000 square miles it was determined that Kansas has a flatness ratio of 0.9997 over the entire state while the average pancake, precisely measured using a confocal laser microscope comes in at 0.957, making Kansas thereby literally flatter than a pancake.

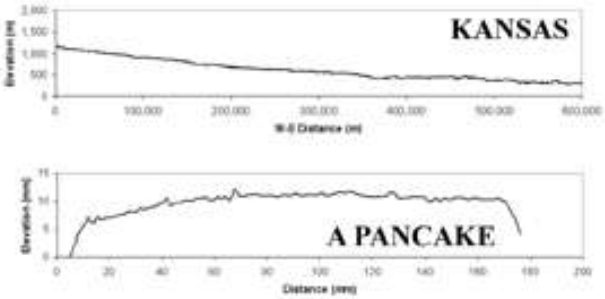

153) Quoting Reverend Thomas Milner's "Atlas of Physical Geography," we find that, "*Vast areas exhibit a perfectly dead level, scarcely a rise existing through 1,500 miles from the Carpathians to the Urals. South of the Baltic the country is so flat that a prevailing north wind will drive the waters of the Stattiner Haf into the mouth of the Oder, and give the river a backward flow 30 or 40 miles. The plains of Venezuela and New Granada, in South America chiefly on the left of the Orinoco, are termed Ilanos, or level fields. Often in the space of 270 square miles the surface does not vary a single foot. The Amazon only falls 12 feet in the last 700 miles of its course; the La Plata has only a descent of one thirty-third of an inch a mile.*"

154) The Felix Baumgartner Red Bull dive outside camera shows the same amount of "curvature of Earth" from surface-level to jump-height proving it to be a deceiving fish-eyed wide-angle lens, while the inside regular camera shows a perfectly flat horizon, eye level at 128,000 feet, which is only consistent with a flat plane.

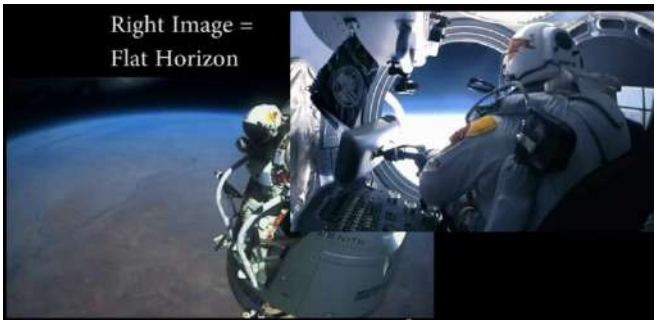

155) Some people claim to have seen the curvature of the Earth out their airplane windows. The glass used in all commercial airplanes, however, is curved to remain flush

with the fuselage. This creates a slight effect mixed with confirmation bias people mistake for being the alleged curvature of the Earth. In actuality, the fact that you can see the horizon at eye-level at 35,000 feet out both port/starboard windows proves the Earth is flat. If the Earth were a ball, no matter how big, the horizon would stay exactly where it was and you would have to look DOWN further and further to see the horizon at all. Looking straight out the window at 35,000 feet you should see nothing but "outer-space" from the port and starboard windows, as the Earth/horizon are supposed to be BELOW you. If they are visible at eye level outside both side windows, it's because the Earth is flat!

156) People also claim to see curvature in Go Pro or other high altitude camera footage of the horizon. While it is true that the horizon often appears convex in such footage, it just as often appears concave or flat depending on the tilt/movement of the camera. The effect is simply a distortion due to wide-angle lenses. In lens-corrected and footage taken without wide-angle technology, all amateur high-altitude horizon shots appear perfectly flat.

157) If "gravity" magically dragged the atmosphere along with the spinning ball Earth, that would mean the atmosphere near the equator would be spinning around at over 1000mph, the atmosphere over the mid-latitudes would be spinning around 500mph, and gradually slower down to the poles where the atmosphere would be unaffected at 0mph. In reality, however, the atmosphere at every point on Earth is equally unaffected by this alleged force, as it has never been measured or calculated and proven non-existent by the ability of airplanes to fly unabated in any direction without experiencing any such atmospheric changes.

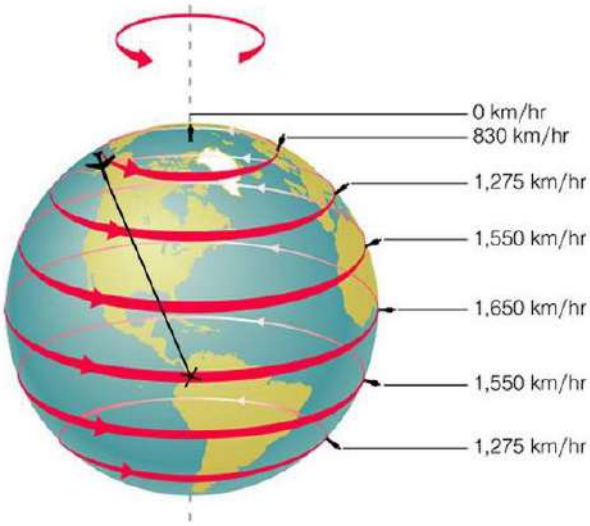

158) If "gravity" magically dragged the atmosphere along with the spinning ball Earth, that would mean the higher the altitude, the faster the spinning atmosphere would have to be turning around the center of rotation. In reality, however, if this were happening then rain and fireworks would behave entirely differently as they fell down through progressively slower and slower spinning atmosphere. Hot-air balloons would also be forced steadily faster Eastwards as they ascended through the ever increasing atmospheric speeds.

159) If there were progressively faster and faster spinning atmosphere the higher the altitude that would mean it would have to abruptly end at some key altitude where the fastest layer of gravitized spinning atmosphere meets the supposed non-gravitized non-spinning non-atmosphere of infinite vacuum space! NASA has never mentioned what altitude this impossible feat allegedly happens, but it is easily philosophically refuted by the simple fact that vacuums cannot exist connected to non-vacuums while maintaining the properties of a vacuum – not to mention, the effect such a transition would have on a rocket "space ship" would be disastrous.

160) It is impossible for rockets or any type of jet propulsion engines to work in the alleged non-atmosphere of vacuum space because without air/atmosphere to push against there is nothing to propel the vehicle forwards. Instead the rockets and shuttles would be sent spinning around their own axis uncontrollably in all directions like a gyroscope. It would be impossible to fly to the Moon or go in any direction whatsoever, especially if "gravity" were real and constantly sucking you towards the closest densest body.

161) If Earth were really a ball, there would be no reason to use rockets for flying into "outer-space" anyway because simply flying an airplane straight at any altitude for long enough should and would send you off into outer-space. To prevent their airplanes from flying tangent to the ball-Earth, pilots would have to constantly course-correct downwards, or else within just a few hours the average commercial airliner traveling 500mph would find themselves lost in "outer-space." The fact that this never happens, artificial horizons remain level at pilot's desired altitudes and do NOT require constant downwards adjustments, proves the Earth is not a ball.

162) All NASA and other "space agencies" rocket launches never go straight up. Every rocket forms a parabolic curve, peaks out, and inevitably starts falling back to Earth. The rockets which are declared "successful" are those few which don't explode or start falling too soon but make it out of range of spectator view before crashing down into restricted waters and recovered. There is no magic altitude where rockets or anything else can simply go up, up, up and then suddenly just start "free-floating" in space. This is all a science-fiction illusion created by wires, green-screens, dark pools, some permed hair and Zero-G planes.

163) NASA and other space agencies have been caught time and again with air bubbles forming and floating off in their official "outer-space" footage. Astronauts have also been caught using scuba-space-gear, kicking their legs to move, and astronaut Luca Parmitano even almost drowned when water started filling up his helmet while allegedly on a "space-walk." It is admitted that astronauts train for their "space-walks" in under-water training facilities like NASA's "Neutral Buoyancy Lab," but what is obvious from their "space bubbles," and other blunders is that all official "space-walk" footage is also fake and filmed under-water.

164) Analysis of many interior videos from the "International Space Station," have shown the use of camera-tricks such as green-screens, harnesses and even wildly permed hair to achieve a zero-gravity type effect. Footage of astronauts seemingly floating in the zero-gravity of their "space station" is indistinguishable from "vomit comet" Zero-G airplane footage. By flying parabolic maneuvers this Zero-G floating effect can be achieved over and over again then edited together. For longer uncut shots, NASA has been caught using simple wires and green screen technology.

165) NASA claims one can observe the International Space Station pass by overhead proving its existence, yet analysis of the "ISS" seen through zoom cameras proves it to be some type of hologram/drone, not a physical floating space-base. As you can see in my documentary "ISS Hoax," when zooming in/out, the "ISS" dramatically and impossibly changes shape and color, displaying a prismatic rainbow

flat earth advanced

The Moon
Function & Cataclysm

mark knight

FLAT EARTH ADVANCED – THE MOON

Function & Cataclysm

Mark Knight www.waykiwayki.com Full Moon, July, 2015

Contents

1. Introduction 3
2. The Moon is upon a Flat Earth 4
3. The Moon Lights Itself 22
4. The Moon is Not Transparent 27
5. The Moon is Flat Not Spherical 34
6. Disinformation 41
7. Sacred Text, Esoterica, and Myth 42
8. Eclipses 45
9. The Moon is an Old Malfunctioning Sun Thesis - Introduction 64
10. The Moon is an Old Malfunctioning Sun Thesis - Advanced 74
Afterword 92

"A truth's initial commotion is directly proportional to how deeply the lie was believed. It wasn't the world being round that agitated people, but that the world wasn't flat. When a well-packaged web of lies has been sold gradually to the masses over generations, the truth will seem utterly preposterous and its speaker a raving lunatic." Dresden James

"Once you eliminate the impossible, whatever remains, no matter how improbable, must be the truth." Arthur Conan Doyle

"The further a society drifts from truth, the more it will hate those that speak it." George Orwell

For dad, who passed away when writing this, thanks for so much.

For the boys, maybe the incredulous slander allowed the space for this research.

1. Introduction

Many people have woken up to the Earth being a flat plane, and this awakening will gain more momentum in the coming years. What still remains a mystery to many within the Flat Earth community is the Moon, because it is probably the toughest nut to crack in the basket.

This offering aims to clearly present what is logical and known about the Moon, plus information, myth, and intuition around some of the Moon's behavior and history - much of which the Flat Earth community remains divided and disparate upon….amidst a bombardment of disinformation, ego, and confusion.

The Moon has always been an enigma; it has always been worshipped throughout time, and its energy has always been harnessed by mystics and the like, therefore it deserves its own publication within the growing Flat Earth community…..and luckily there are still some who prefer the refined, literal experience instead of a YouTube clip (at times awash with ghastly music, amateurism, subscriber addiction, and haste).

The book's structure will *layer up*, so when reading a passage and you feel you require more to be explained, be patient, all is revealed in its most relevant section. The book´s structure is designed this way. Some of the earlier chapters, plus the simplistic writing style on some of the later deeper subjects are deliberately implemented to make this book accessible to a wider range of people.

There is some information in the latter pages not yet disclosed anywhere else, and I see it as important work. For me, personally, the book is partly the fruit yielded from a mystical dream puzzle, a puzzle that started during a fifty day, silent fast meditation, within a hermetic cocoon up a volcano many years ago. I come from more of a mysticism and astrological background, so apologies if you are seeking a 100% *pure-rational-science* offering throughout.

I ask one thing: slightly try to let go of labels and preconceived ideas about this mysterious light.

2. The Moon is upon a Flat Earth

Model

Let us look at the Moon upon the Flat Earth model before we prove this *has* to be the case.

The Moon moves around us in *spiraling-circles* above the stationary Flat Earth, moving from tropic to tropic over the course of six months, in an almost identical trajectory to the Sun.

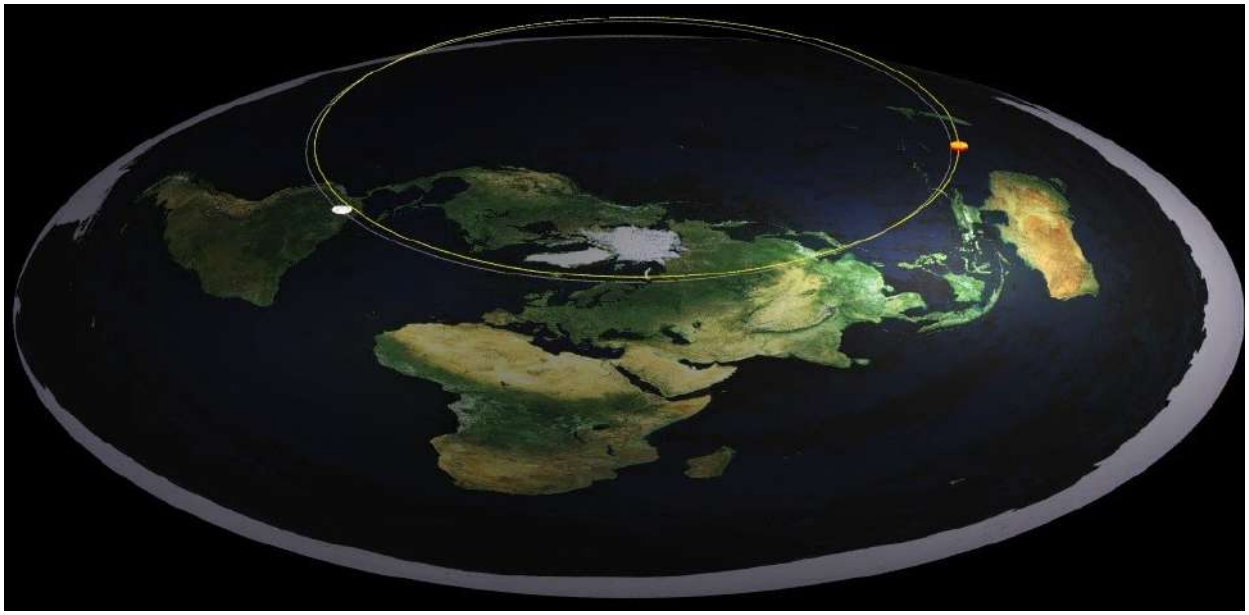

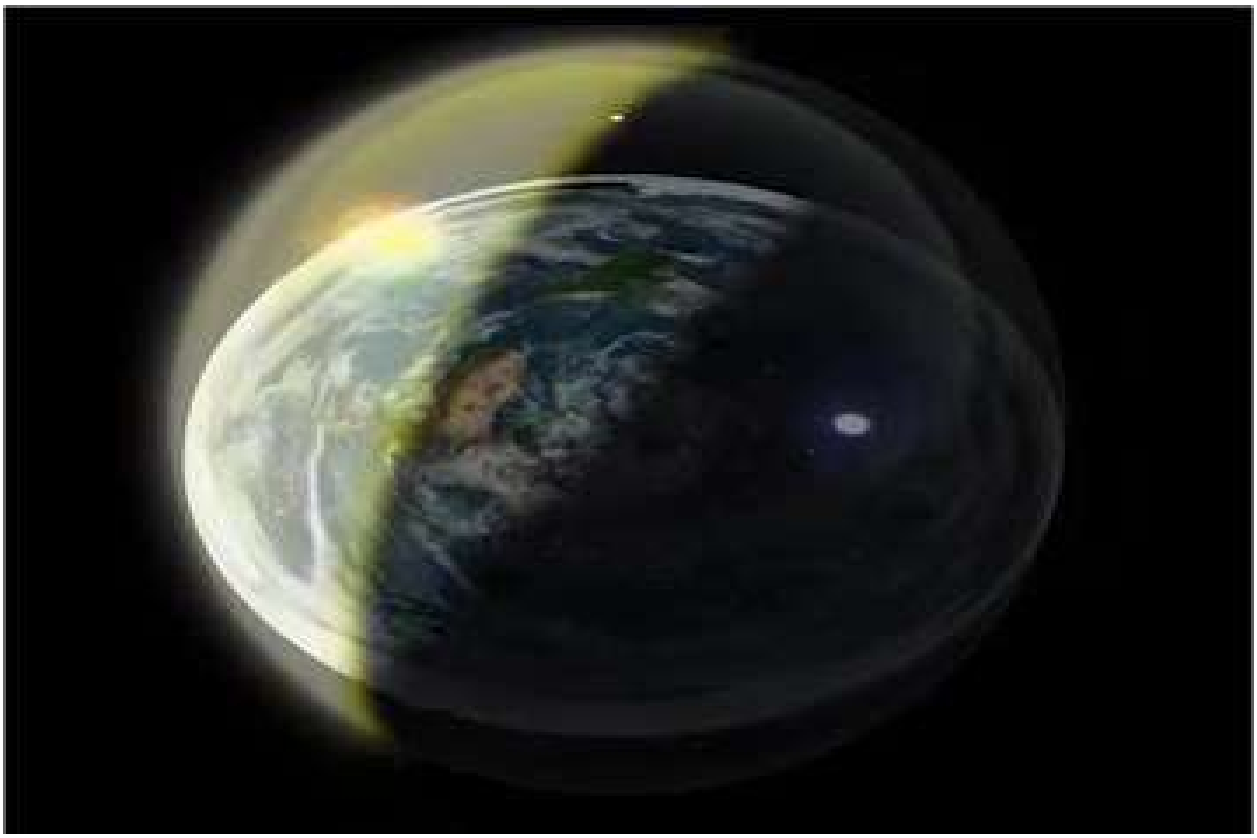

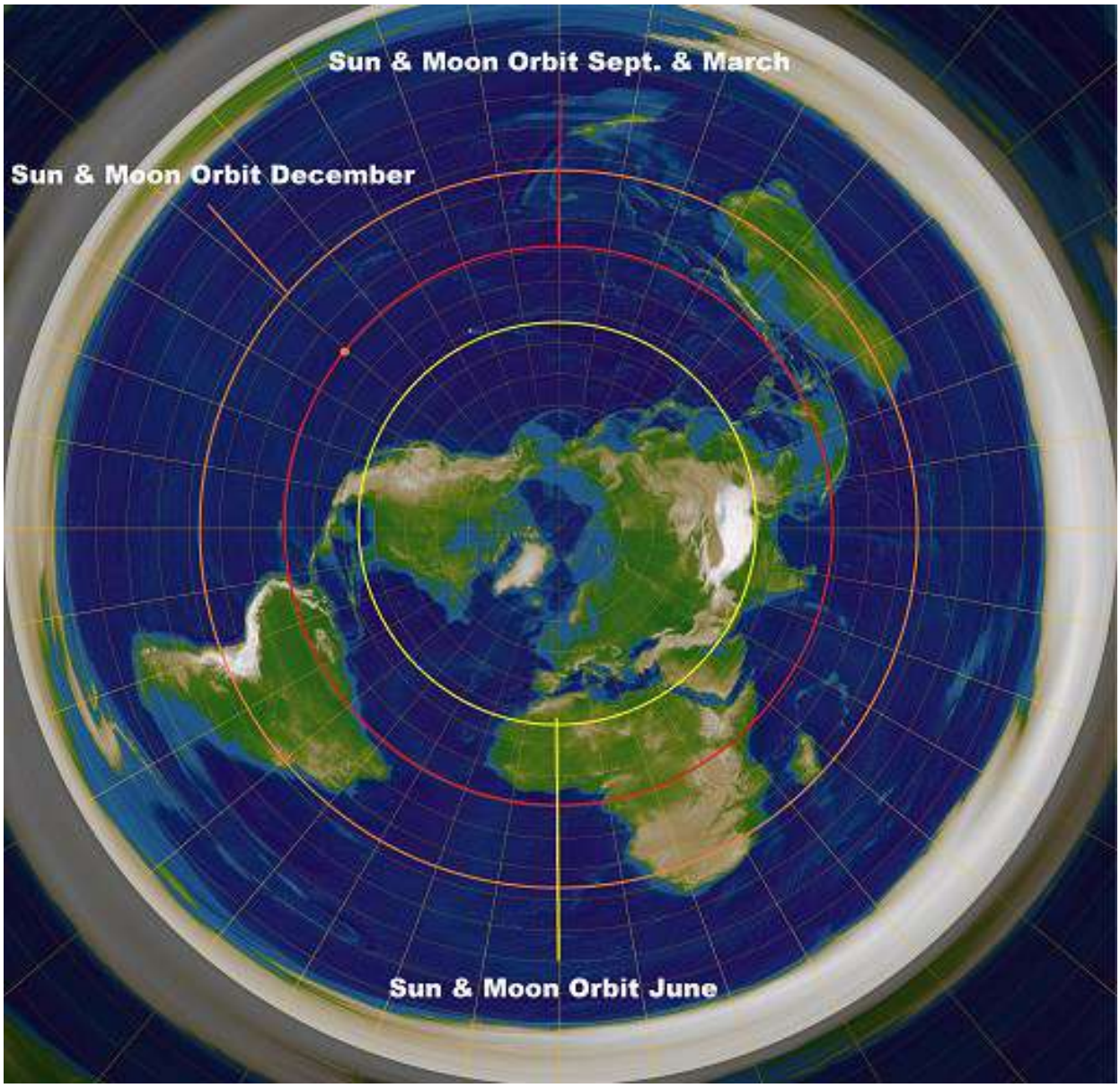

One Face

All humans on Earth can only ever see one face of the Moon, and it appears with the face rotated - with the degree depending on location. The Moon covers the full 360 degrees of rotation. Mainstream science simplifies this fact and states the Moon is *right-side up* north of the equator and *upside down* south of the equator. This isn't true, it appears rotated to different observers in different locations.

If the Earth and Moon were both balls, then humans at the far north would see a different *face* of the Moon than those in the far south, with some *overlap*.

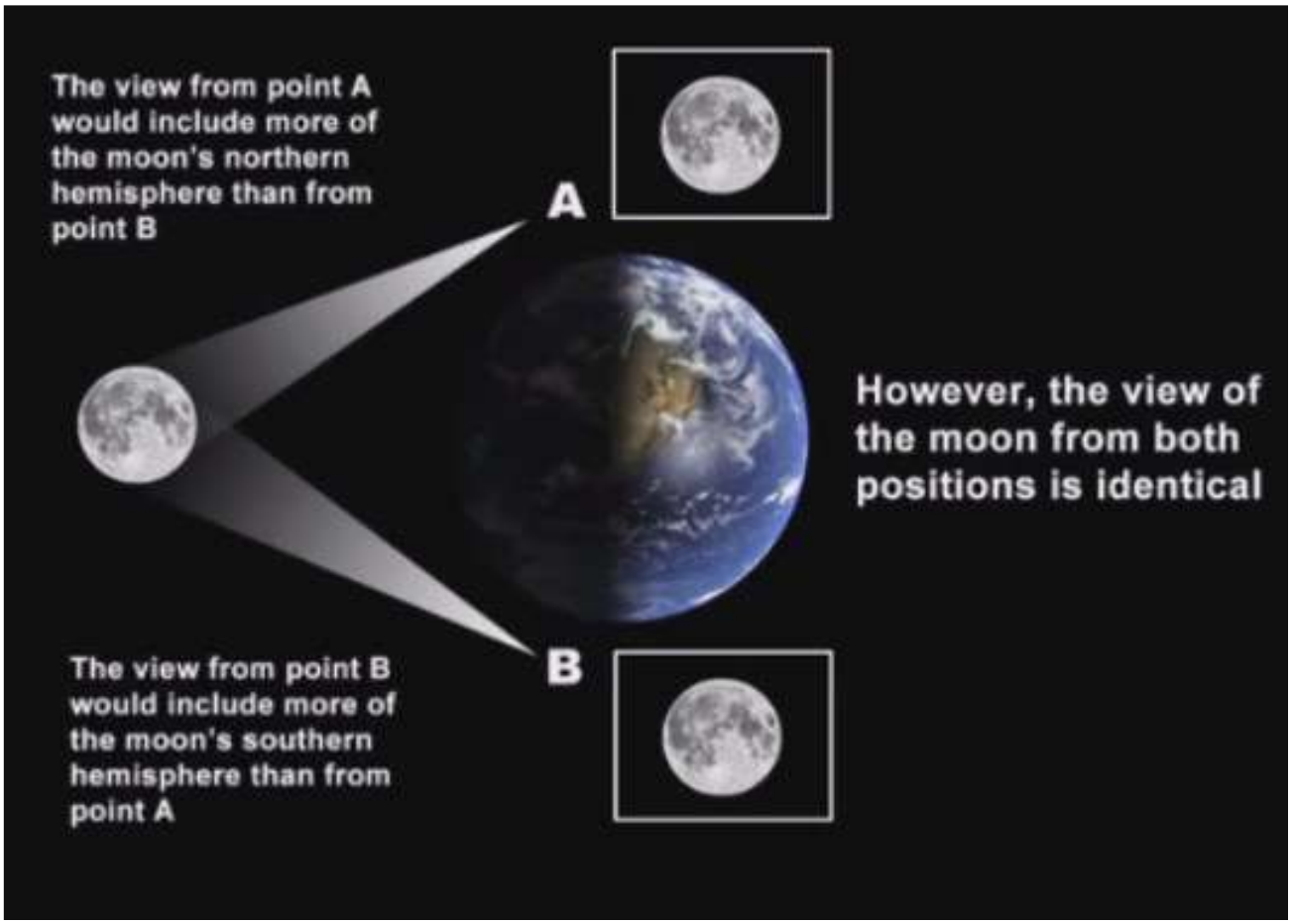
The view from point A would include more of the moon's northern hemisphere than from point B
A
However, the view of the moon from both positions is identical
B
The view from point B would include more of the moon's southern hemisphere than from point A

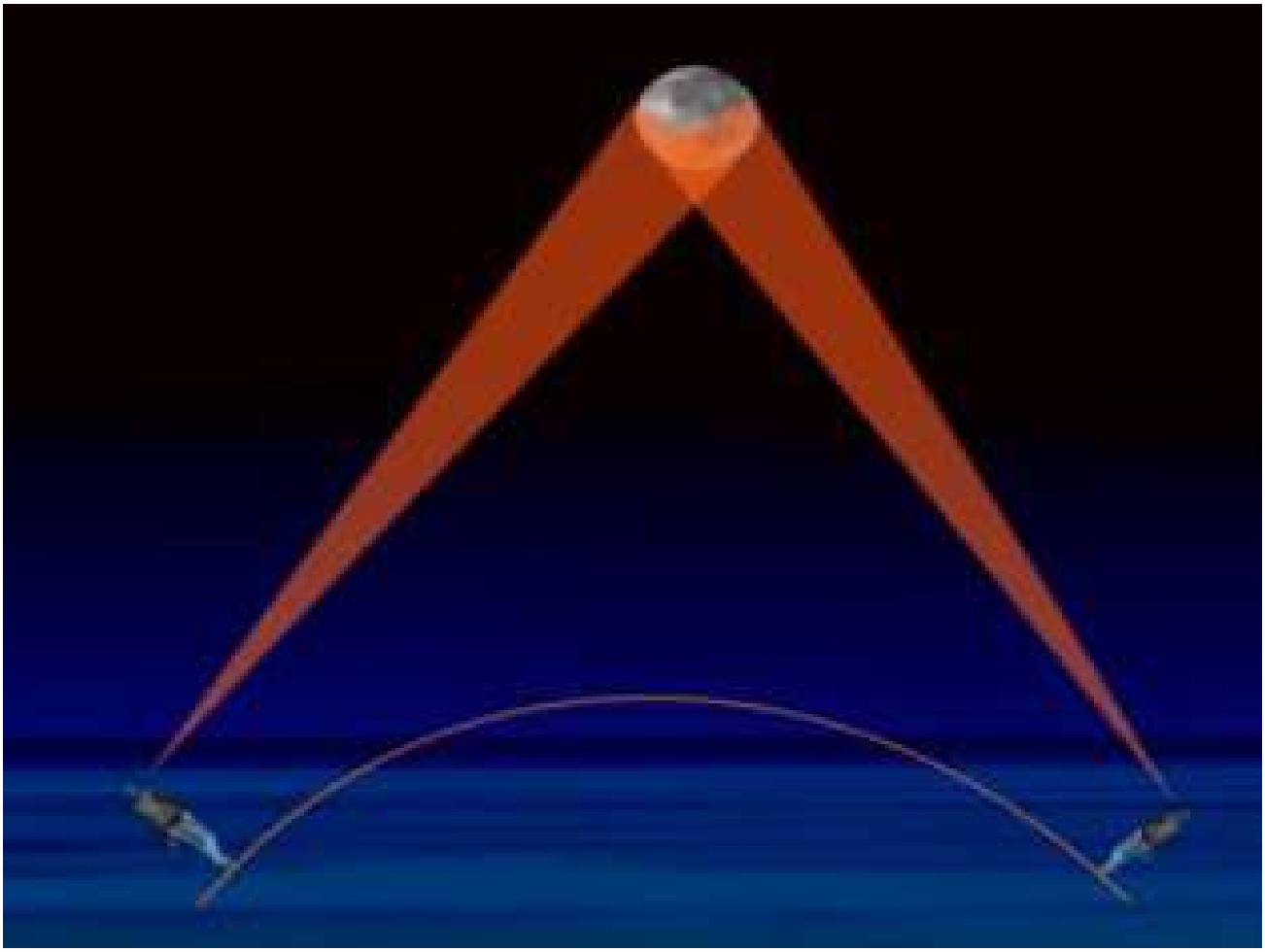

The only way we can all see the same face is upon a Flat Earth.

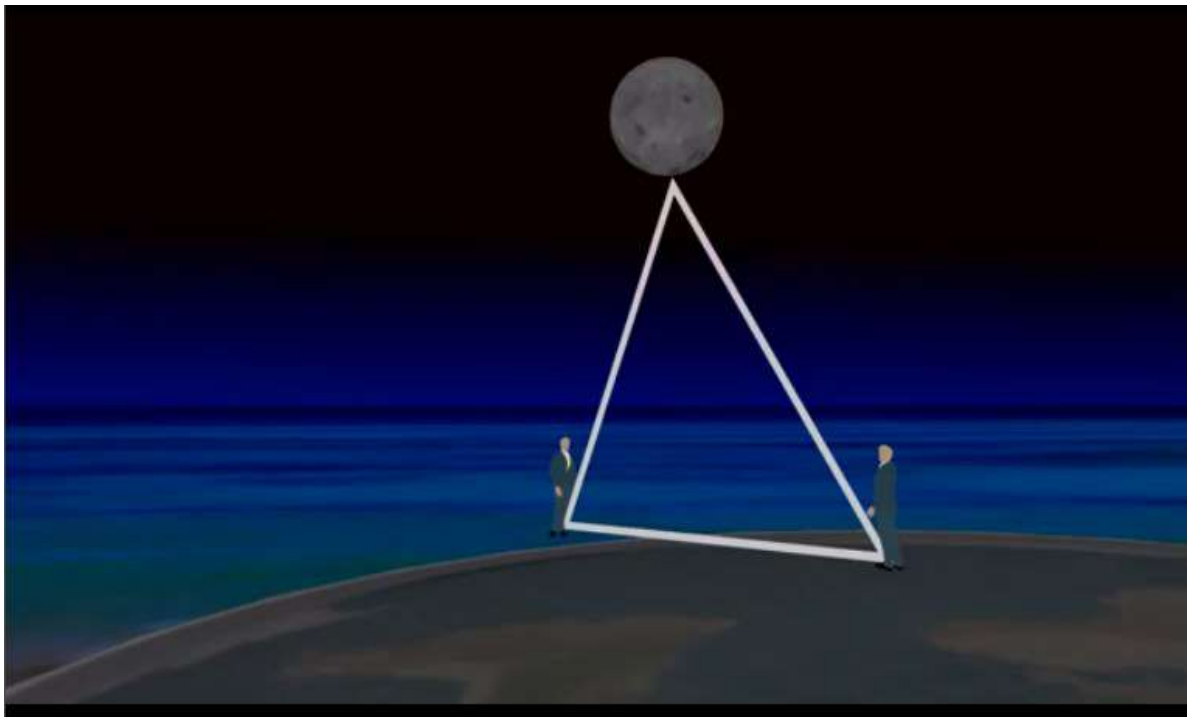

One can clearly see the Moon moving away from the observer, to the vanishing point of a human´s eye. Eyes are spherical objects that focus light on the lens at the back, the horizon has nothing to do with the shape of the world and everything to do with distance. People are fooled into thinking the horizon is the edge of the curve.

Phases

We are told the Moon's phases (the 29.53 day rhythm of different degrees of illumination) are from Sunlight via the Sun´s angle to the Moon, but this cannot be the case. Below is the mainstream model.

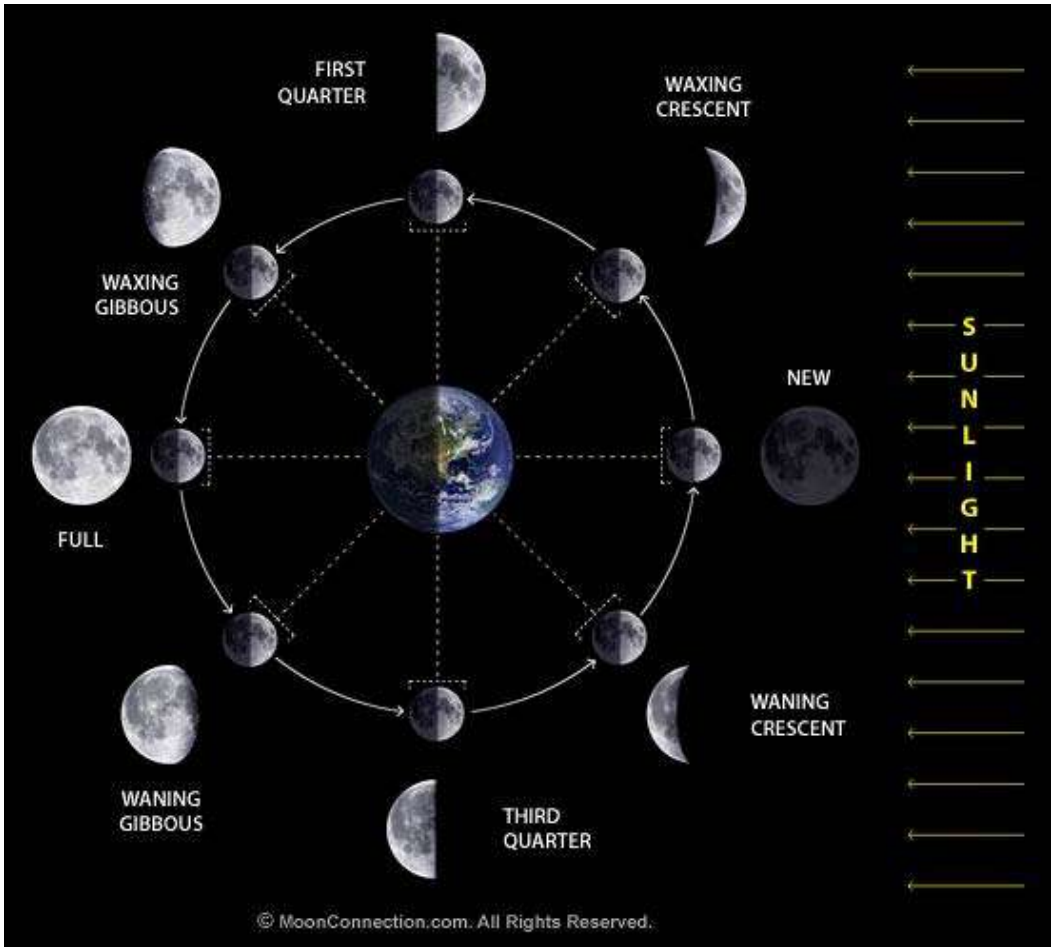

But when it is New Moon in day in the ball model, we should see a black silhouette, we don't though, we see nothing. We should see a silhouette in the ball model because the Sun is so much bigger than the Moon and behind it.

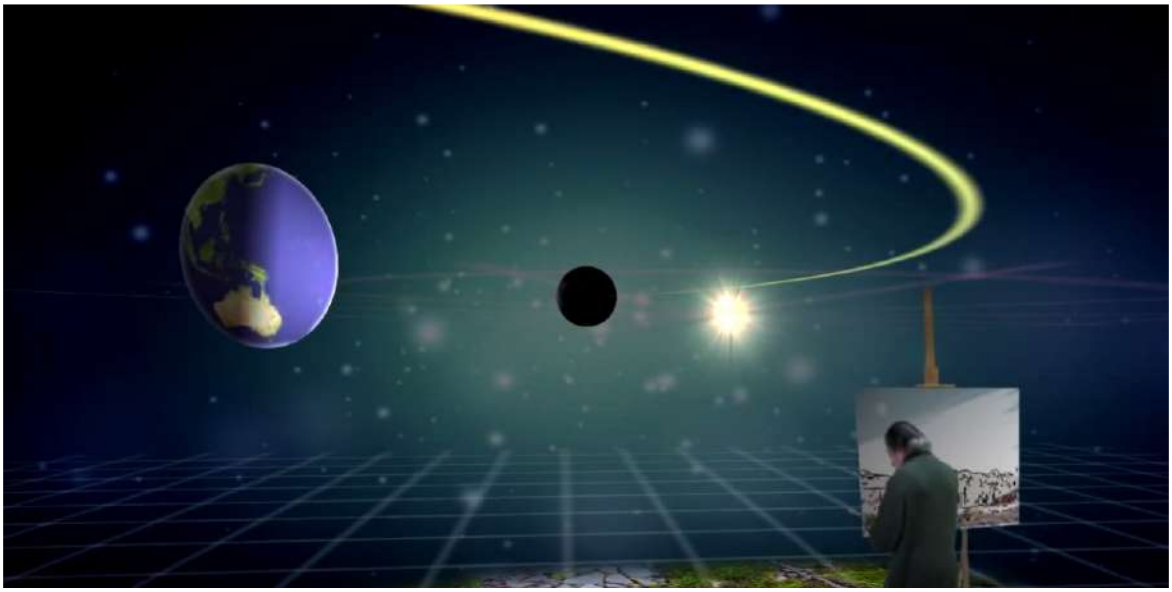

Also we can see a nearly Full Moon in the day.

This is not possible on the ball Earth, as during the day, the earth is facing the Sun (red below). The Earth would be in the way of someone trying to see the Moon in the daytime.

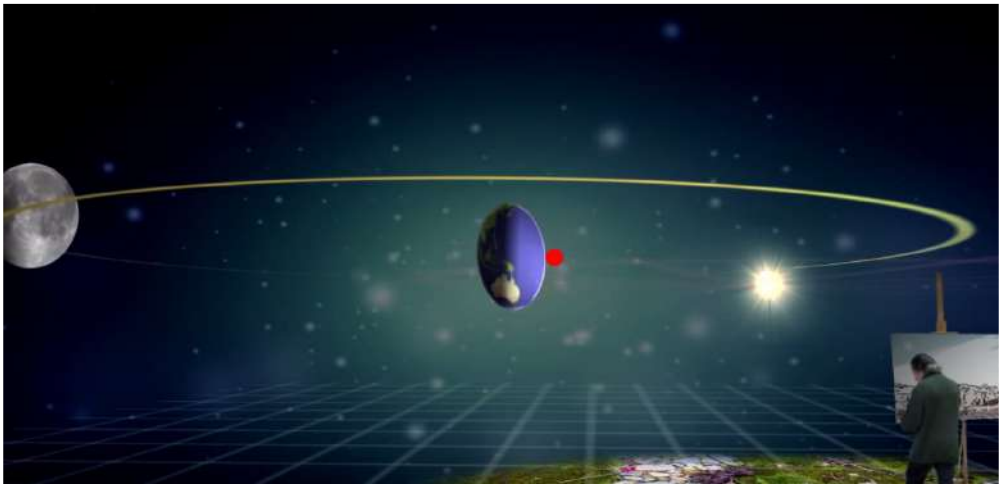

Also we can sometimes see the Sun and Moon in the sky at the same time (for some hours in the days just before and just after the Full Moon). There is nothing between the Sun and Moon, so the Moon should be 100% lit in the ball model. There is nothing obstructing the *massive* Sun's rays, but the Moon appears only four fifths or so illuminated.

If the Moon *was* orbiting the Earth, and both were orbiting the Sun, then six months later the Moon's phases should have swapped over – and each lunar month, the phases should appear different.

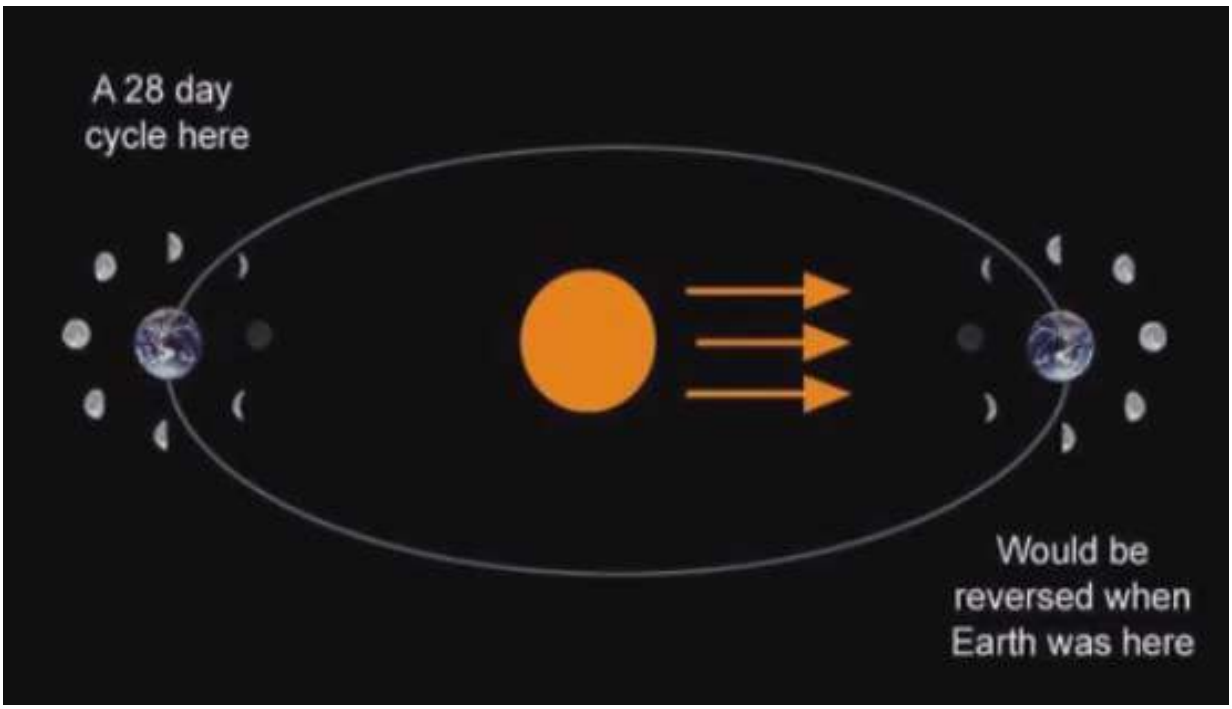

We are told that the Sun gives the Moon its light for the phases, but if this were the case, during each lunar month, the Moon would be at a different angle to the Sun. It would be impossible for the phases to remain the same each lunar month....but they do.

Study the two images below to get a full understanding of this. These two images both present the same mainstream paradox: a) the Moon does what it does with its phases each month. b) The Moon goes around the Sun with the Earth. One of these has to be wrong, but we are told that both are true

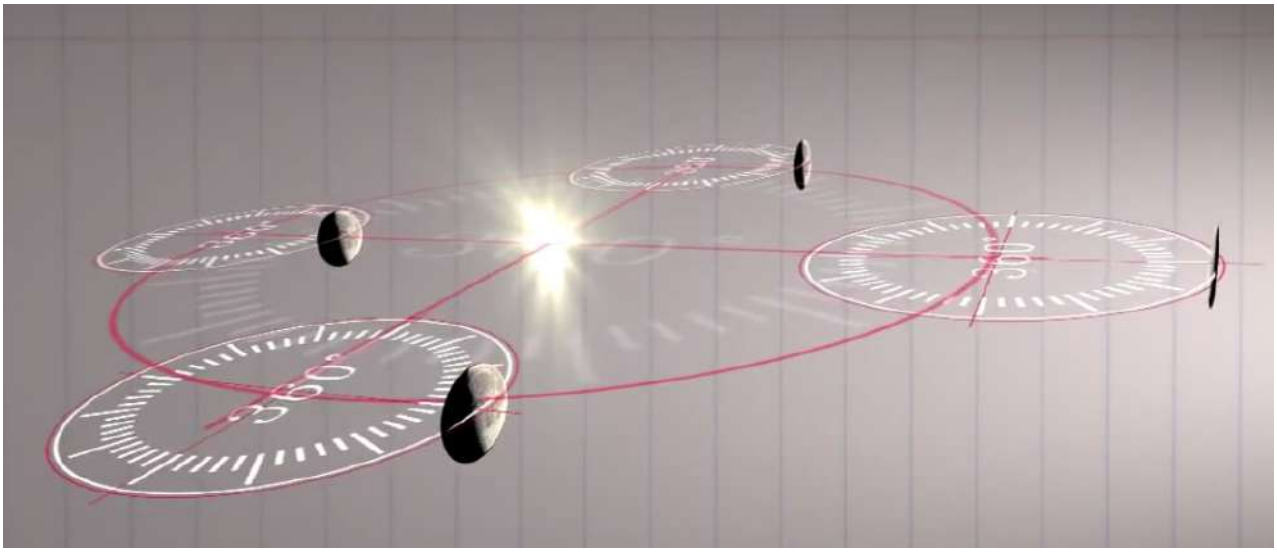

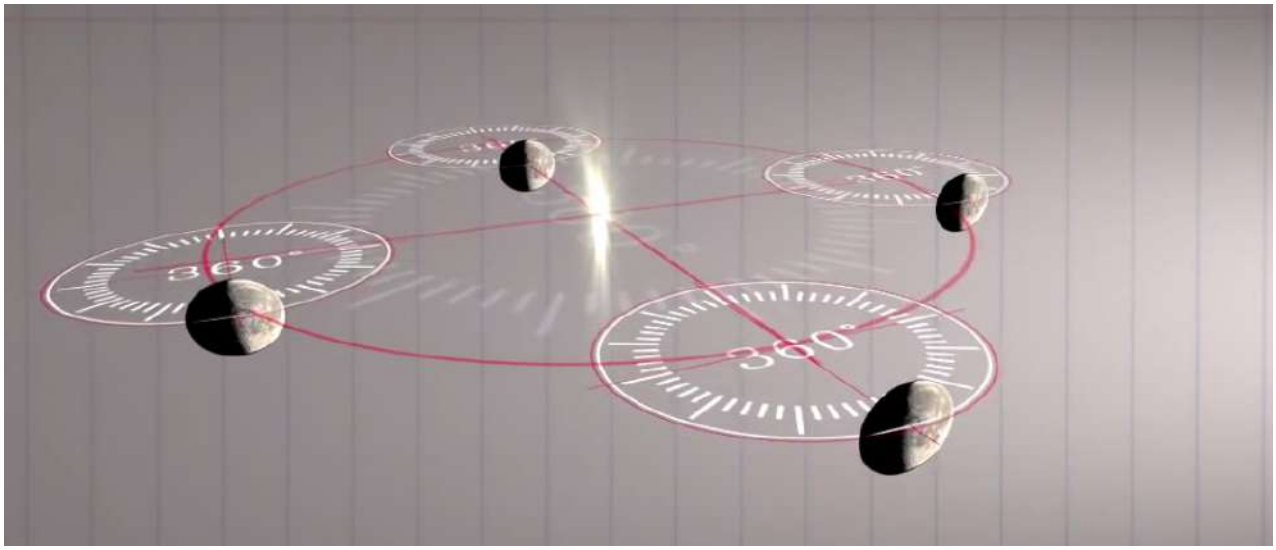

You can see from the two images above that the Moon should have a different light reflection from the Sun at different locations around the Sun, but at the same time and location of its own lunar month.

It should actually act like this below, but it doesn't, helping to prove the ball Earth model is a hoax.

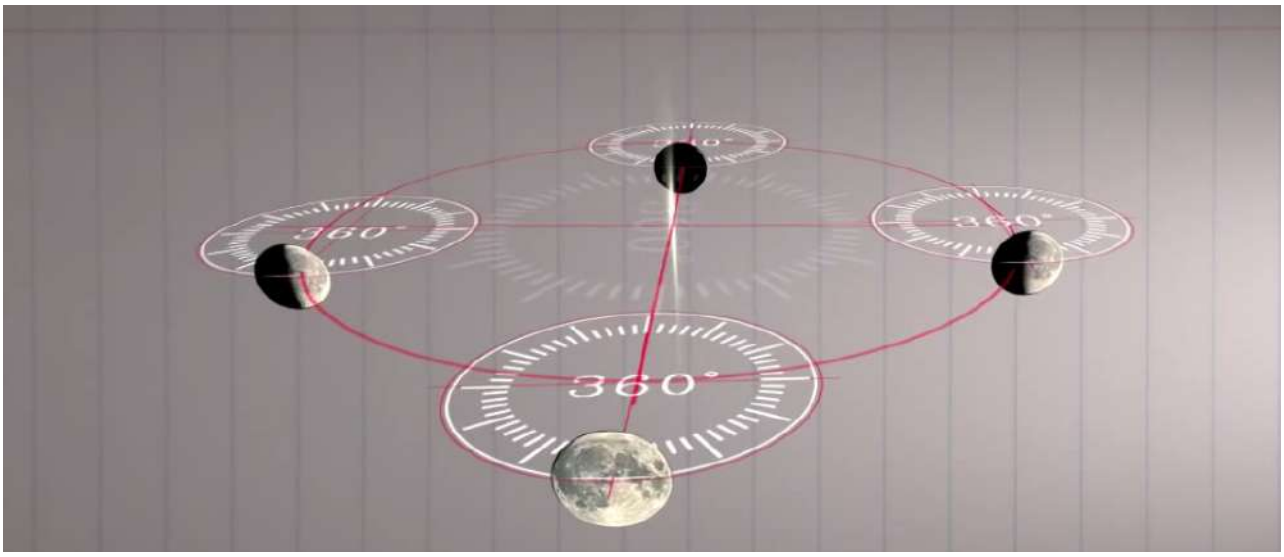

This might need some contemplation, a pen and paper, or possibly a re-read. But it is indisputable.

Reflection

The Moon cannot be above a ball or the reflection on the ocean wouldn't reach to the person on the shore, but it does reach an observer on the shore.

If the Moon were above a ball, the reflection on the ocean would behave much differently. See below where we have pieced in the (phantom) curvature.

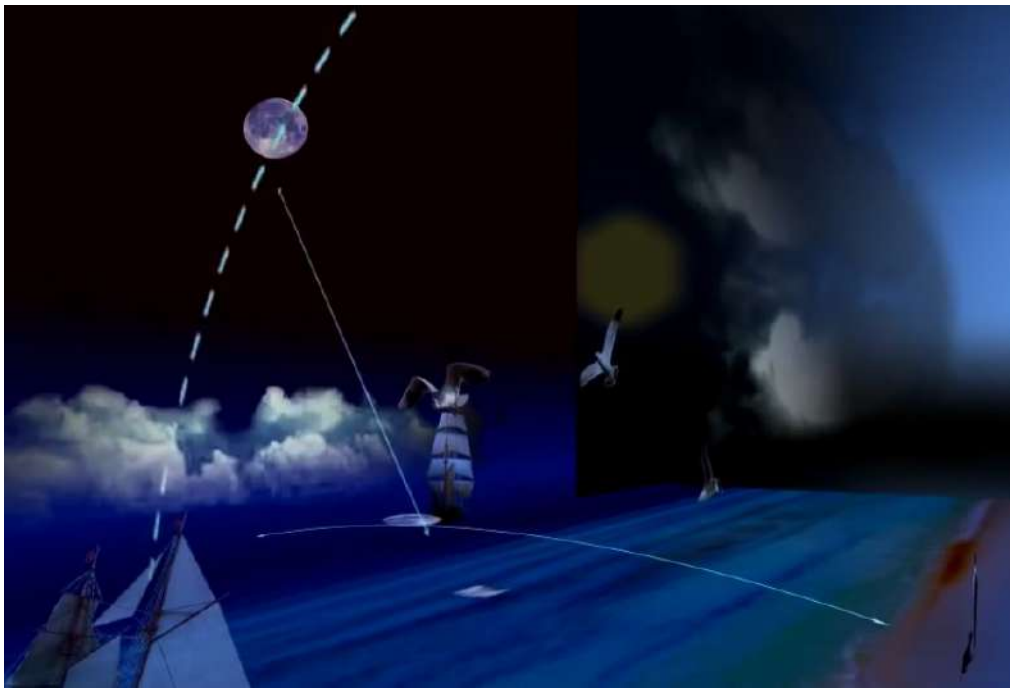

Tides

We are told that the gravitational pull of the Moon *alone* creates the tides, but this is also wrong.

If the Moon was 2,160 miles in diameter and the Earth 8,000 miles in diameter, the Earth would be 87 times larger, and would attract the smaller mass (the Moon) to *it*, not the other way.

We are told the Earth's stronger gravity keeps the Moon in orbit, so how can the Moon's weaker gravity override Earth's gravity at Earth's lowest points (sea level and below sea level).

If the magical error-ridden force called gravity existed, all trees would look like willow trees, and all flowers would wilt in an 'n' shape. But that is for another time.

The Earth's tides vary greatly. For example, near the Isle of Wight, there are four tides a day. Some beaches have tides which go out far, while others not so much. The mainstream map of tides is not uniform, and would not be caused by a single body moving harmoniously across from east to west.

"If the Moon lifted up the water, it is evident that near the land, the water would be drawn away and low instead of high tide caused. Again, the velocity and path of the Moon are uniform, and it follows that if she exerted any influence on the Earth, that influence could only be a uniform influence. But the tides are not uniform. At Port Natal the rise and fall is about 6 feet, while at Beira, about 600 miles up the coast, the rise and fall is 26 feet. This effectually settles the matter that the Moon has no influence on the tides. Tides are caused by the gentle and gradual rise and fall of the Earth on the bosom of the mighty deep. In inland lakes, there are no tides; which also proves that the Moon cannot attract either the Earth or water to cause tides. But the fact that the basin of the lake is on the Earth which rests on the waters of the deep shows that no tides are possible, as the waters of the lakes together with the Earth rise and fall, and thus the tides at the coast are caused; while there are no tides on waters unconnected with the sea." -Thomas Winship, "ZeteticCosmogeny" (130-131)

“It is affirmed that the intensity of attraction increases with proximity, and vice versa. How, then, when the waters are drawn up by the Moon from their bed, and away from the Earth's attraction,--which at that greater distance from the centre is considerably diminished, while that of the Moon is proportionately increased--is it possible that all the waters acted on should be prevented leaving the Earth and flying away to the Moon? If the Moon has power of attraction sufficient to lift the waters of the Earth at all, even a single inch from their deepest receptacles, where the Earth's attraction is much the greater, there is nothing in the theory of attraction of gravitation to prevent her taking to herself all the waters which come within her influence. Let the smaller body once overcome the power of the larger, and the power of the smaller becomes greater than when it first operated, because the matter acted on is nearer to it. Proximity is greater, and therefore power is greater … How then can the waters of the ocean immediately underneath the Moon flow towards the shores, and so cause a flood tide? Water flows, it is said, through the law of gravity, or attraction of the Earth's centre; is it possible then for the Moon, having once overcome the power of the Earth, to let go her hold upon the waters, through the influence of a power which she has conquered, and which therefore, is less than her own? … The above and other difficulties which exist in connection with the explanation of the tides afforded by the Newtonian system, have led many, including Sir Isaac Newton himself, to admit that such explanation is the least satisfactory portion of the ‘theory of gravitation.’ Thus we have been carried forward by the sheer force of evidence to the conclusion that the tides of the sea do not arise from the attraction of the Moon, but simply from the rising and falling of the floating Earth in the waters of the ‘great deep.’ That calmness which is found to exist at the bottom of the great seas could not be possible if the waters were alternately raised by the Moon and pulled down by the Earth.” -Dr. Samuel Rowbotham, “Zetetic Astronomy, Earth Not a Globe!” (159-175)

“Even Sir Isaac Newton himself confessed that the explanation of the Moon's action on the Tides was the least satisfactory part of his theory of Gravitation. This theory asserts that the larger object attracts the smaller, and the mass of the Moon being reckoned as only one-eighth of that of the Earth, it follows that, if, by the presumed force of Gravitation, the Earth revolves round the Sun, much more, for the same reason, should the Moon do so likewise, instead of which that willful orb still continues to go round our world. Tides vary greatly in height, owing chiefly to the different configurations of the adjoining lands. At Chepstow it rises to 60 feet, at Portishead to 50, while at Dublin Bay it is but 1 2, and at Wexford only 5 feet … That the Earth itself has a slight tremulous motion may be seen in the movement of the spirit-level, even when fixed as steadily as possible, and that the sea has a fluctuation may be witnessed by the oscillation of an anchored ship in the calmest day of summer. By what means the tides are so regularly affected is at present only conjectured; possibly it may be by atmospheric pressure on the waters of the Great Deep, and perhaps even the Moon itself, as suggested by the late Dr. Rowbotham, may influence the atmosphere, increasing or diminishing its barometric pressure, and indirectly the rise and fall of the Earth in the waters.” -David Wardlaw Scott, “Terra Firma” (259-260)

the Sun and Moon together create the tides. The biggest tides are when the Sun and Moon are both above the horizon and one near to the other.

The combination of the Sun and Moon locations most probably create electromagnetic currents. Tides only happen in the oceans, never in creeks, rivers or marshes. The ocean is made up of salt water, and salt is a conductor of electricity, and therefore has an EMF. Salt molecules are made of sodium ions and chlorine ions, and an ion is an atom that has an electrical charge.

Here´s a sea temperature map upon the Flat Earth, it's also possible this has a bearing on tides too.

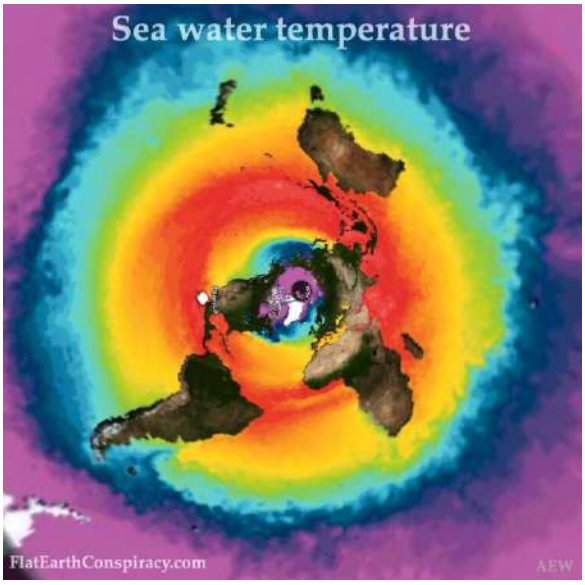

"I've noticed that when the Sun is over say the Atlantic then it's low tide, and when the Moon is over the Atlantic it is high tide, in my research into electromagnetic energy is that the Sun is very energetic to humans, animals and even gives food to plants by Chlorophyll. And the Moon is the negative energy as it putrefies plants etc, so in studying I've noticed it may be affected by this dual positive/negative together and they work in harmony with the ions in the water and electromagnetic fields." Anonymous.

When we look at tides on the Flat Earth model things make more sense.

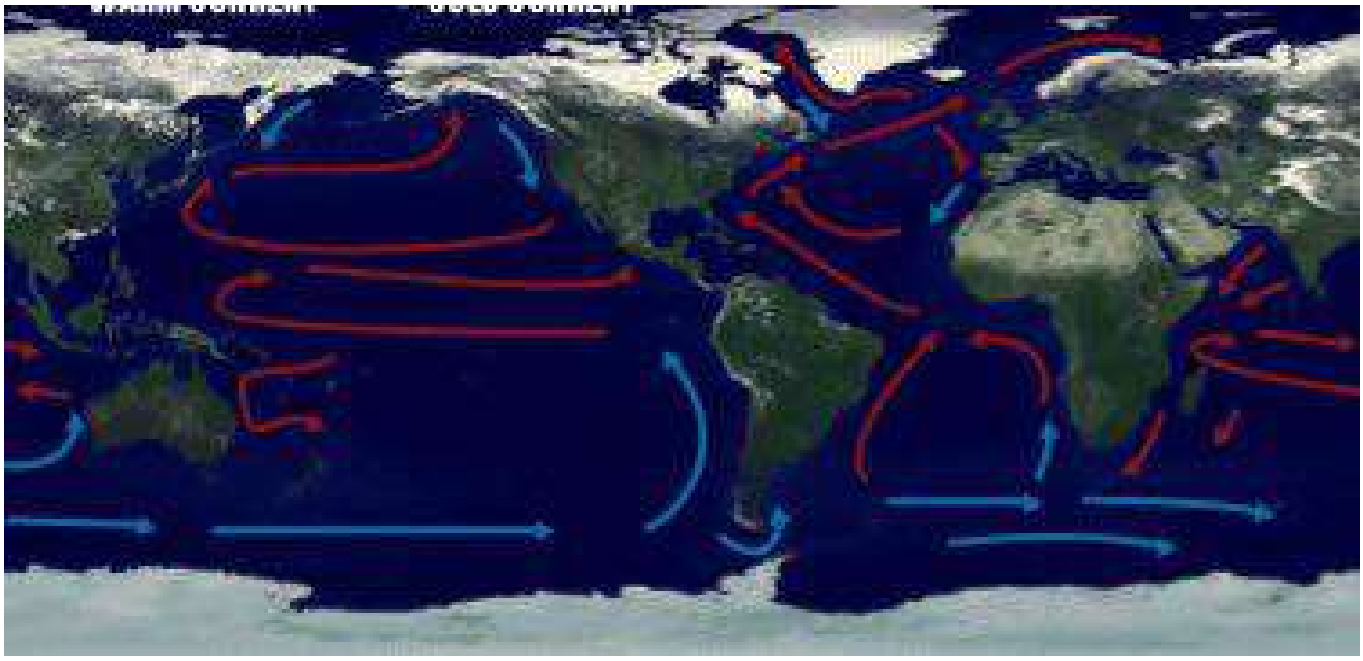

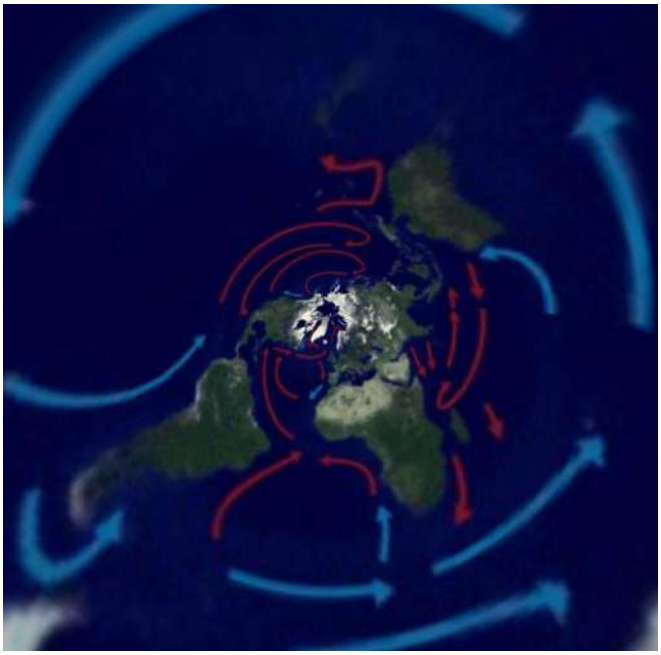

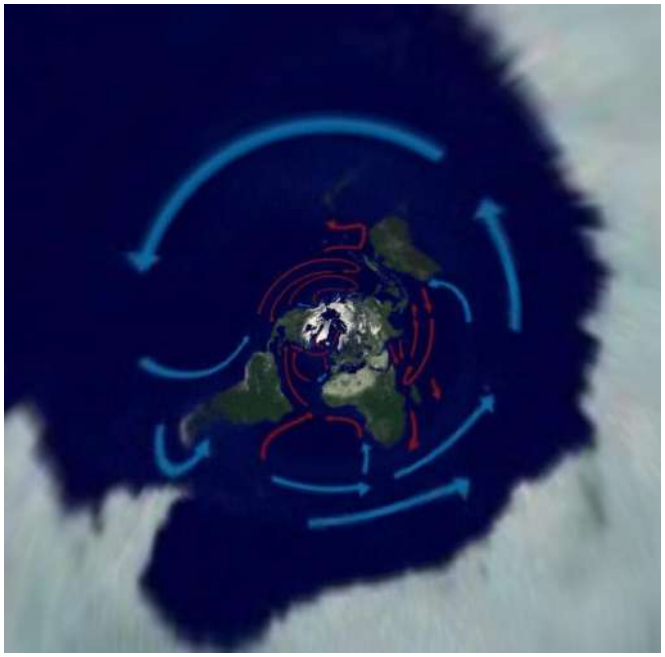

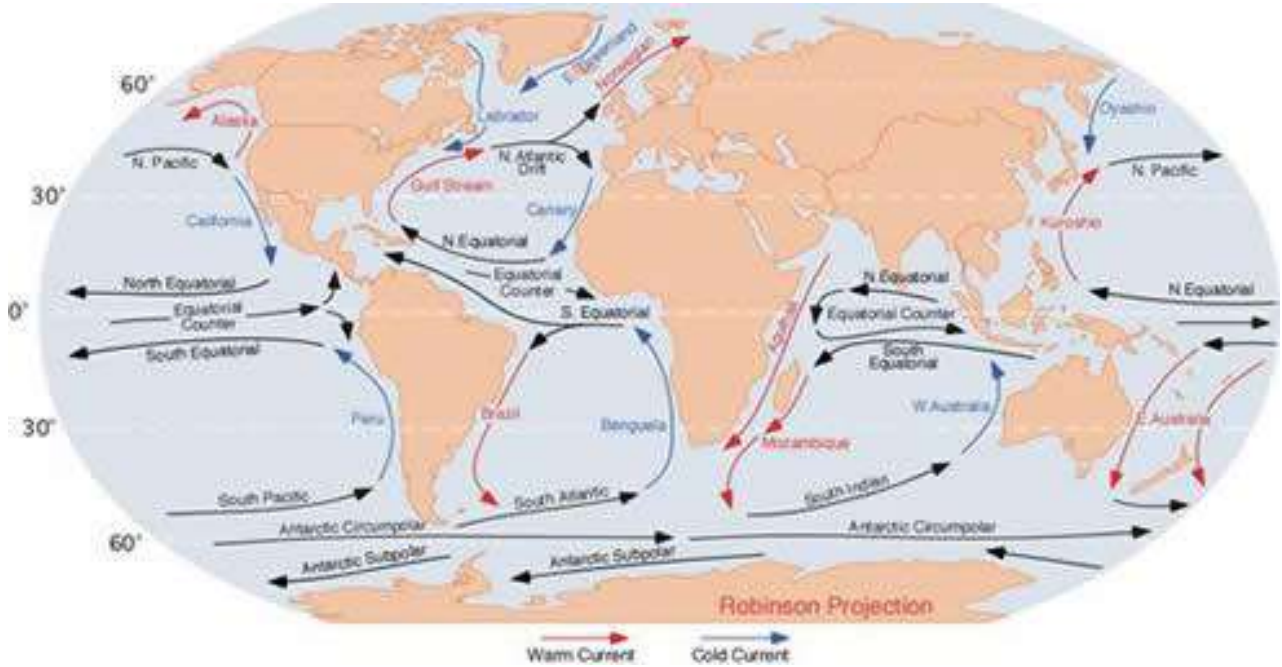
60°
30°
0°
30°
60°
Alaska
N. Pacific
California
North Equatorial
Equatorial Counter
South Equatorial
Peru
South Pacific
Antarctic Circumpolar
Antarctic Subpolar
Labrador
N. Atlantic Drift
Gulf Stream
Canary
N. Equatorial
Equatorial Counter
S. Equatorial
Brazil
Benguela
South Atlantic
Oyashio
N. Pacific
N. Equatorial
Equatorial Counter
South Equatorial
W. Australia
Mozambique
South Indian
Antarctic Circumpolar
Antarctic Subpolar
Robinson Projection
Warm Current
Cold Current

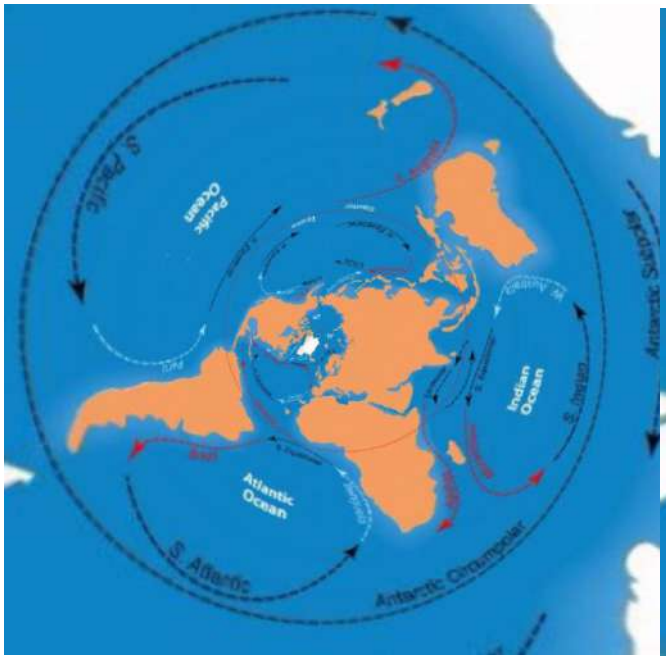
S. Pacific
Pacific Ocean
Indian Ocean
S. Indian
Atlantic Ocean
S. Atlantic
Antarctic Circumpolar
Antarctic Subpolar

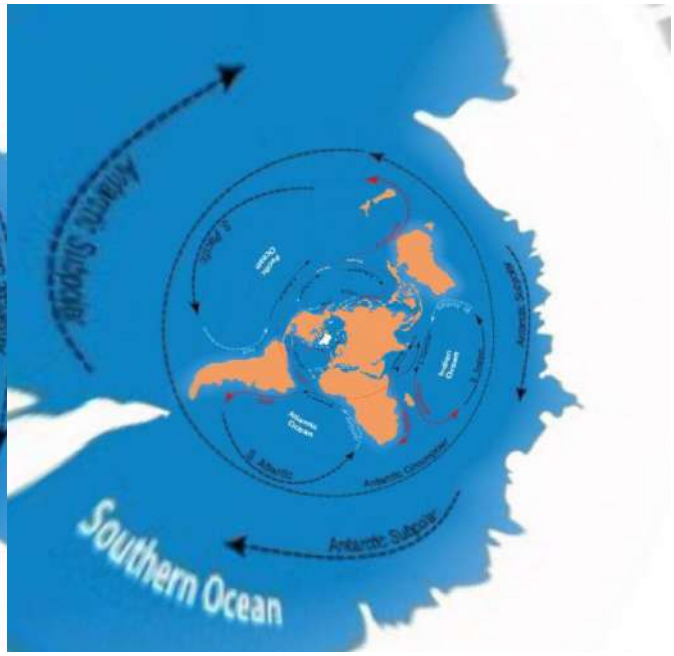
Antarctic Subpolar
Southern Ocean

Size, Distance, and Speed

The Moon is thirty two or so miles wide, as is the Sun.

The Moon is around 3000 miles away. Do you really think you can see details on something 235,000 miles away, when you can't see details on a mountain seventy miles away on a clear day? I have a theory that the Moon could be as near as 700 miles away, but I would need a sextant to do my own calculations, but I don't have one. One thing's for sure, it is much closer than we are told. Currently, I cannot go against the '3000 miles-away distance'.

The Moon is very similar in distance from Earth as the Sun is, but is a bit closer to Earth as the Moon appears in front of the Sun on a Solar Eclipse. The Moon is anything up to a few hundred miles nearer to Earth than the Sun is, but the altitude of the Moon (and/or the Sun too) varies. How do we know this? We can witness total Solar Eclipses and annular Solar Eclipses (as we will see later).

The Moon travels around sixty to seventy minutes a day slower than the Sun in a twenty-four hour period.

The Moon's movement (forward motion) is not the *forward* route through the zodiac as we see it in the sky, the Moon´s travel is the other way as the backdrop of the zodiac spins/moves faster (as viewed from a human eye). This is also the same rule for the Sun.

The constellations go much faster than the Moon as viewed from a human eye (so the Moon appears to go through a zodiac constellation each 2.5 days or so in reverse - but this is seen in astrology as normal / forward).

The constellations only go slightly faster than the Sun, as viewed from the human eye, so it spends a longer time in each constellation -- about a month.

Here are images from Stellarium, showing the movement of the Moon from a baseline in southern Spain, then forty-eight hours later, and then another forty-eight hours later.

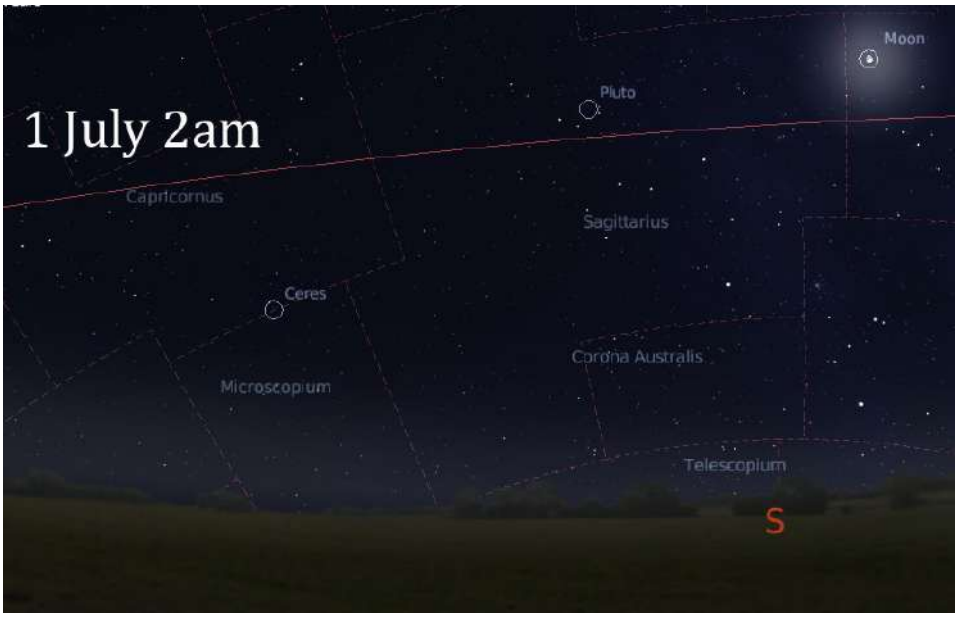

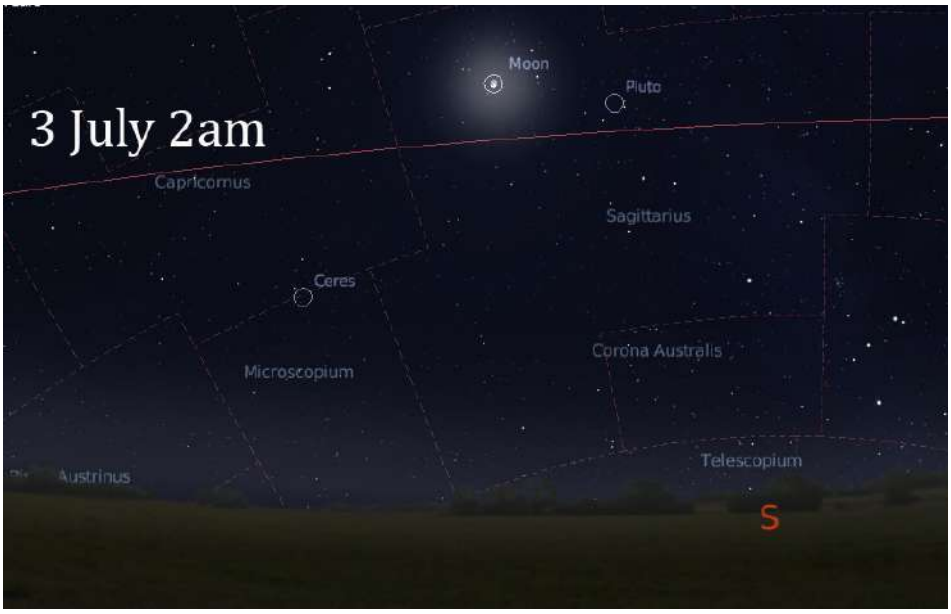

The Moon *really is* moving as Stellarium suggests (essentially ball Earth software) – it's line in the sky is correct.

A New Moon is basically when the Sun overtakes the Moon, as it moves quicker. A *Lunar Overtaken*.

A Full Moon is basically when the Moon is as far away from the Sun as it can be within the *current* lunar cycle, but the Sun is slowly catching up and now moving into the same semi-circle. A *Lunar Run but Can´t hide*.

Nodes

This image is presented as a fact, with the Lunar Nodes around the globe Earth, the cause being due to an approximate 5% tilt of the Moon's orbits.

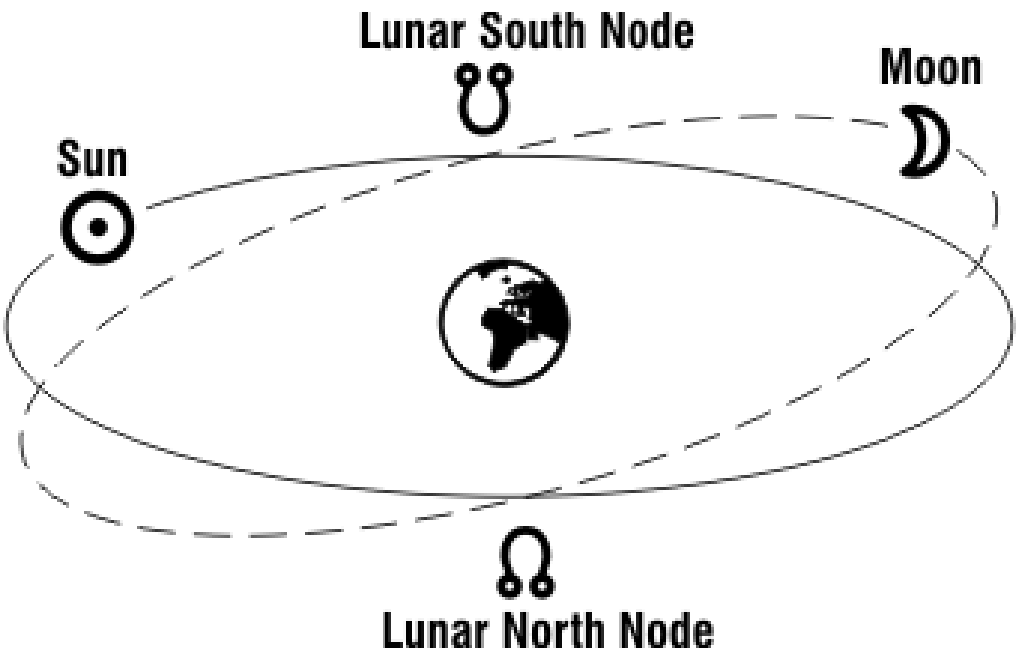

Within the Flat Earth model, this is simply explained; we look up and we see the Moon moving above and below the Sun´s *apparent path*, which is known as the ecliptic in ball Earth terms. In Flat Earth terms, *the ecliptic is just a line that shows the Sun´s path*, nothing more.

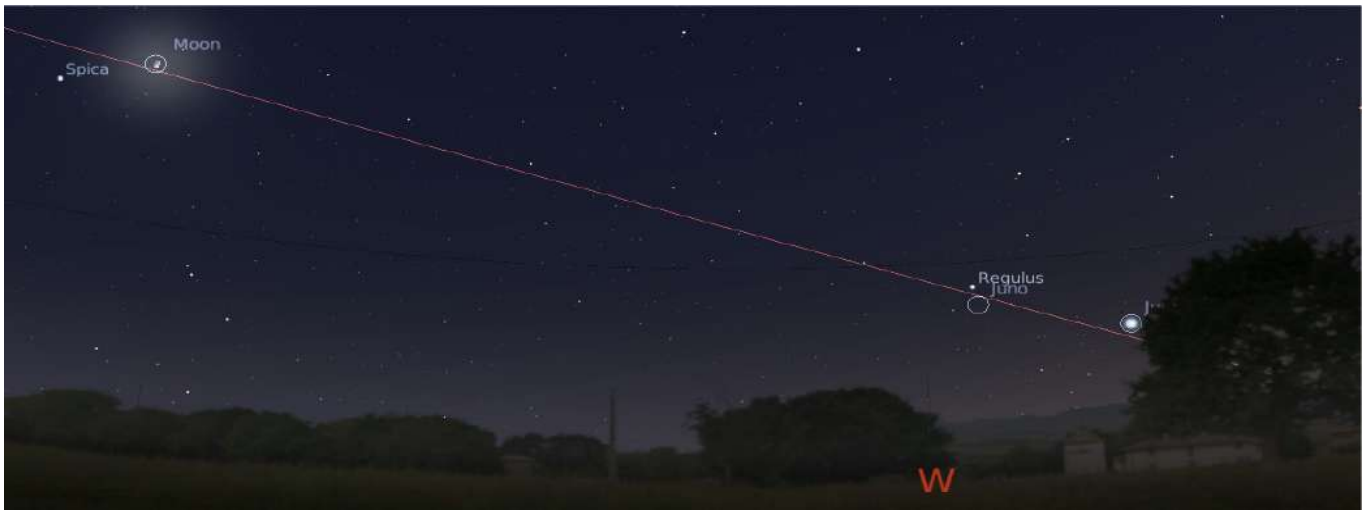

So over the course of time (twice a month), the Moon crosses above and then below the Sun´s trajectory line (see image above where the Moon is above the line). Where the Moon crosses it, this is a Node (in astrological terms). The Moon spends half the time above this line, and half the time below this line (from a human eye on our Flat Earth). The Nodes are intersections.

It´s important you understand and remember this simple Nodes' explanation for the eclipse chapter much later on. We will use the term FE-Node from now on to avoid confusion.

The Moon *at* a Node is related to eclipses, and to the past and future in astrology, the South Node natal placement of a person is what one brings into this incarnation. The North Node natal placement is where one can evolve and learn the most, where one *should* be heading.

NASA

Why are there no photos of the Moon from one of the thousands of satellites we supposedly have in space? Or any photos from the ISS, supposedly it is manned all the time, but no photos of the Moon? The Hubble telescope hasn't taken any photos of the Moon, as it is apparently on high-zoom mode. NASA has also made some errors with the Moon. Here is the Moon being supposedly filmed from The ISS.

For sure this is a really poor effort at Photoshop – a white circle - maybe the license ran out and they had to use MS Paint?

Here is a famous NASA image below on the left, the Earth from the Moon.

But this is wrong, the Earth is supposed to be four times bigger than the Moon, but the Earth appears about the same size the Moon does from Earth. The image on the left should have looked more like the one on the right if the ball model were true.

3. The Moon Lights Itself

We are told that the Moon reflects Sunlight, and this is the only reason for its llumination. Here are some facts.

- Sunlight is gold and yellow, Moonlight is white and blue.
- The Sun's light is drying and preserving while the Moon's light is damp and putrefactive.
- The Sun's rays can put out a common fire while the Moon's light will increase combustion.
- The light of the Sun falling upon certain chemicals produces a change of colour such as photographic processes. The light of the Moon fails to produce the same effect.
- The Sun's light when concentrated by a number of mirrors, or a large lens produces a focus that is entirely non-luminous, but the heat is so great that metallic and alkaline substances are quickly fused; Earth and mineral compounds almost immediately vitrified; and all animals and vegetable structures and burned and destroyed in a few seconds.
- The Moon's light when concentrated produces a brilliant focus, so luminous that it's difficult to look upon it, and yet there is no increase of temperature.

- In direct Sunlight a thermometer will read higher than another thermometer placed in the shade, but in full, direct Moonlight a thermometer will read lower than another placed in Moon shade.

.

"The Sun's light, when concentrated by a number of plane or concave mirrors throwing the light to the same point; or by a large burning lens, produces a black or non-luminous focus, in which the heat is so intense that metallic and alkaline substances are quickly fused; Earthy and mineral compounds almost immediately vitrified; and all animal and vegetable structures in a few seconds decomposed, burned up and destroyed. The Moon's light concentrated in the above manner produces a focus so brilliant and luminous that it is difficult to look upon it; yet there is no increase of temperature. In the focus of Sun-light there is great heat but no light. In that of the Moon's light there is great light but no heat." -Dr. Samuel Rowbotham, "Zetetic Astronomy, Earth Not a Globe!" (144)

"Light which is reflected must necessarily be of the same character as that which causes the reflection, but the light of the Moon is altogether different from the light of the Sun, therefore the light of the Moon is not reflected from the Sun." -David Wardlaw Scott, "Terra Firma" (151-2)

Close to the New Moon phase on either side, one can sometimes see most of the Moon faded whilst a slither is illuminated.

So what is illuminating this more faded part of the Moon above? It is dark up there, we should not be able to see this portion if only the Moon´s edge was lit by the Sun.

When we see Moon at night, it is lit up, illuminated, but when we *see* the Moon in day, even with the Sun in the sky too, it is not lit, but a matt cold stony grey. The Sun is lit in day and gives light. So why not the Moon if it is reflecting the Sunlight? Why only at night? If I shine a torch on a rock it lights up in day.

What we are told from repeaters (teachers and scientists) doesn't make sense.

Phases of the Moon are just different portions of the Moon´s self-illumination, but the phases *are* highly probably *linked* to the Sun.

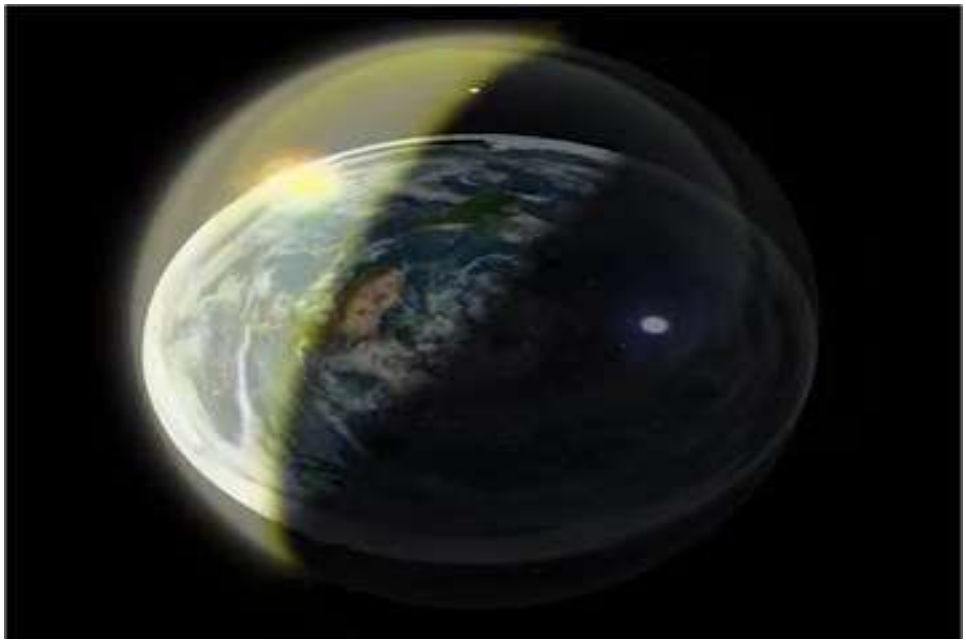

Note: The Sun revolves above us once every 24 hours, and the Moon revolves above us about once every 25.25 hours; the Sun chases the Moon.

The phases of the Moon *are* related to (actually, caused by) the relative positions of the Moon and Sun in the Flat Earth sky. For example, New Moon always occurs when the Sun and Moon

are very close together in the sky. Full Moon always occurs (image above) when the Sun and Moon are at opposite positions in the sky - which is why a Full Moon rises about the time of Sunset, and sets about the time of Sunrise. First and Last Quarters occur when the Sun and Moon are about 90 degrees apart in the sky.

On the day of the New Moon (image below is a day or so before the New Moon. Remember that the Sun is chasing the Moon so to speak), the Moon rises when the Sun rises. It sets when the Sun sets. It crosses the sky with the Sun during the day. That's why we can't see the New Moon in the sky. It´s too close to the Sun's glare to be visible, but the Moon stops illuminating and it doesn't shine.

Then a day or two later, the Moon reappears, in the west after Sunset. Then it's a slim waxing crescent visible only briefly after Sunset – what some call a young Moon.

A waxing crescent - almost a New Moon - is seen only very briefly after Sunset. This is because at waxing crescent, the Sun rises immediately before the Moon rises, and then the Sun sets immediately before the Moon sets. Therefore, right after the Sun sets and the Sun's glare fades in dusk, the very slim crescent (nearly a New Moon) becomes visible, and then the Moon shortly thereafter sets...and can't be seen for the rest of the night. If the Sun's glare wasn't so bright during the day, you could see the Moon with this slight crescent all day long. But you can't. However, at dusk, when the Sun's glare fades over the western horizon, the slim crescent appears! (It doesn't appear because it comes up above the horizon, but only because the Sun's brightness fades into the night). But watch the waxing crescent into night, and you will see it quickly set below the western horizon, and it won't be seen again until the following dusk.

"A waning crescent" - almost a New Moon - is seen only very briefly before Sun rise. This is because at waning crescent, the Moon rises immediately before the Sunrise, and then the Moon sets immediately before the Sun sets. As a result, right before Sunrise, in early dawn, the slim crescent Moon (almost a New Moon) rises in the East. It appears as it rises above the Eastern horizon, before the Sun rises. However, you can't see it very long, either, because then the Sun rises, and the Sun's glare is so bright, it overwhelms the slim glimmer of the waning crescent! However, it is still there...and is there in the sky all day long! And finally, the Moon sets first over the western horizon in dusk, not to be seen. And then the Sun sets...to a Moonless night almost all night until right before dawn, when the waning crescent rises to be seen again."

Maybe the Sun drains the power of the Moon, and when the Moon is close to the Sun it loses its power (illumination)? The closer to the Sun the Moon gets, the more power (self-illumination) it loses, and when it is very close it loses its power (self-illumination) completely, and becomes what we call a New Moon. Maybe there is a handshake of energy between the Sun and Moon´s energy, be that spiritual-pranic energy or electromagnetic energy.

Many though think there is another object in the sky.

I will go into these notions deeper in later chapters, but let us digress and note one thing here:

Venus and Mercury also have phases – Mainstream science tells us again that these occur due to the Sunlight's angle too (which we now know is impossible).

Venus and Mercury are both, at times, in front of the Sun and behind the Moon – this makes them very small, and also makes them around the same distance from us as the Sun and Moon. Maybe Venus and Mercury's altitude varies slightly, or the Sun and Moon´s altitude varies slightly, or maybe all of these object´s altitudes vary slightly. One or more of these have to be true.

Maybe Venus and Mercury also *self illuminate* according to their distance from the Sun?

Note that the Maya were obsessed with Venus transit cycles for their cataclysm calendars.

***

Once in solitude in a meditative state many years ago I watched the Full Moon all night, the Sun was nowhere to be seen. I felt then intuitively that the Sun doesn't illuminate the Moon.

To conclude this chapter, remember this for later: If the Moon self illuminates, then it also self de-illuminates....they are both the same thing. So a waxing Moon, the clever, good people of the Flat Earth community believe, is the Moon self illuminating, as they rightly should, but these people cannot believe or think that the Moon self de-illuminates on a waning cycle? Remember this notation for later for it will become important.

From what we know thus far, the Moon´s phases have nothing to do with shadows from the Earth or shadows from different angles of Sunlight. The Moon self illuminates.

We were crazy and a little stupid to ever think anything different.

4. The Moon is Not Transparent

Let us note firstly, that what science calls the Moon's valleys and mountains, are really nothing more than small surface irregularities that are sometimes enhanced by light and shadow, and by our perspective from the Earth's surface. One could rename these valleys and mountains to wrinkles and scars, and it would be a million times closer to the truth than what NASA tells us. As I said in the introduction, try and stay away from subconscious labels that were delivered to you at a young age. The craters on the other hand are real, and are caused by impacts with smaller and falling celestial bodies.

Many in the Flat Earth community think the Moon is semi-transparent at times.

This is largely based upon three things; blue sky phenomena, star sightings, and Islamic symbolism. Let's dissect and debunk each of these.

Many say that as one can see the blue sky *through* the Moon sometimes, therefore it´s transparent. This argument falls down in many ways.

The blue color of the atmosphere is only fifteen to twenty kilometers thick, and you can clearly see this in all of the high-altitude weather balloon videos. When they go over thirty kilometers, that blue layer is not there anymore, and when the camera points up, it’s just black, and when the camera points down it can be seen that the blue layer below is only above the Earth's surface in a thin layer. These images are from around 120,000 feet, and the blue layer is below, and the Moon is higher.

One modern researcher claimed:

"The Moon is most certainly self-illuminated, but not a "ball" or "solid." You can easily see through the Moon most afternoons. You can see blue sky right through the supposed "craters" on the Moon!"

The *craters* of the Moon are just darker than the rest of the Moon's surface, this is why the blue haze has more of an impact on the color of the *craters*. But it's just a simple and obvious illusion.

Jump the gun sidenote: There are just some portions of the Moon that have a dimmer light, the possible *malfunctioning* (we will come onto this later) is greater in those sections because of stronger impacts during cataclysm.

The so-called *blue craters* just look darker than the rest of the Moon's surface, the same as they look at night - they don't shine as brightly as the rest of the Moon's surface.

This thin blue layer of atmosphere is probably some sort of prana or chi, though mainstream science says the blue layer is caused by Sunlight scattering more blue (shorter smaller waves) than any other colour. This is known as *Rayleigh Scattering* (basically a type of refraction), and for me their evidence is flakey, at best.

Stars in front of the Moon

There have been many sightings of a light (star and/or planet) passing in front of the Moon.

"That the Moon is not a perfectly opaque body, but a crystallized substance, is shown from the fact that when a few hours old or even at quarter we can through the unilluminated portion see the light shining on the other side. Stars have also been observed through her surface!" -J. Atkinson, "Earth Review Magazine

On March 7th, 1794, four astronomers (three in Norwich, one in London) wrote in "The Philosophical Transactions of the Royal Astronomical Society" that they "saw a star in the dark part of the Moon, which had not then attained the first quadrature; and from the representations which are given the star must have appeared very far advanced upon the disc." Sir James South of the Royal Observatory in Kensington wrote in a letter to the Times newspaper April 7, 1848, that, "On the 15th of March, 1848, when the Moon was seven and a half days old, I never saw her unillumined disc so beautifully. On my first looking into the telescope a star of about the 7th magnitude was some minutes of a degree distant from the Moon's dark limb. I saw that its occultation by the Moon was inevitable … The star, instead of disappearing the moment the Moon's edge came in contact with it, apparently glided on the Moon's dark face, as if it had been seen through a transparent Moon; or, as if a star were between me and the Moon … I have seen a similar apparent projection several times … The cause of this phenomenon is involved in impenetrable mystery." In the monthly notices of the Royal Astronomical Society for June 8, 1860, Thomas Gaunt stated that the "Occultation of Jupiter by the Moon, on the 24th of May, 1860, was seen with an achromatic of 3.3 inches aperture, 50 inches focus; the immersion with a power of 50, and the emersion with a power of 70. At the immersion I could not see the dark limb of the Moon until the planet appeared to touch it, and then only to the extent of the diameter of the planet; but what I was most struck with was the appearance on the Moon as it passed over the planet. It appeared as though the planet was a dark object, and glided on to the Moon instead of behind it; and the appearance continued until the planet was hid, when I suddenly lost the dark limb of the Moon altogether." I have personally also seen stars through the edge of the waxing/waning Moon. It actually happens fairly often; if you are diligent and specifically observing for the phenomenon on starry nights you can occasionally see it even with the naked eye." Zetetic Astronomy, Earth Not a Globe!" (337)

"On the 15th of March, 1848, when the Moon was seven and a half days old, I never saw her unillumined disc so beautifully. . . . On my first looking into the telescope a star of about the 7th magnitude was some minutes of a degree distant from the Moon's dark limb. I saw that its occultation by the Moon was inevitable. . . . The star, instead of disappearing the moment the Moon's edge came in contact with it, apparently glided on the Moon's dark face, as if it had been seen through a transparent Moon; or, as if a star were between me and the Moon. . . ." Zetetic Astronomy

This is an artist's impression of the phenomena.

Islamic Symbolism

Some point to the Symbol for Islam as backing up the notion that the Moon is transparent.

Note that the five pointed star is often Venus in much esoterica.

The origin of this symbol is deep in Mesopotamia, with the *Crescent of Sin* (Nanna to the Sumerians) and the *Star of Ishtar* (Inanna to the Sumerians).

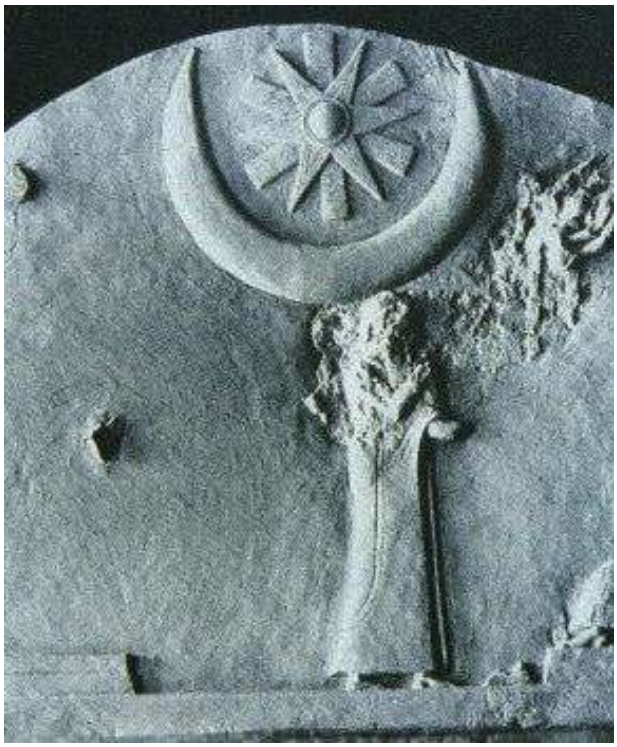

First Century coin.

***

There is nowhere near enough evidence to suggest that the Moon is *sometimes* transparent.

My answer to this rare phenomena is this:

Venus (and possibly Mercury) sometimes (very rarely) pass in front of the Moon:

- They are both a similar distance away from the Earth as they both pass in front of the Sun sometimes. In this theory Venus and Mercury would obviously have a pattern in which they slowly came a bit nearer to the Earth (in front of the Moon), then slowly moved a bit further out again (behind the Moon as is normally the case). This is logical because Venus and Mercury sometimes (rarely) pass in front of the Sun, but mostly behind it. Islamic scholars and astronomers state that on the dawn of June 10, 640 A.D., Venus was conjunct the Sun, around the time Mohammed apparently had a chinwag with Gabriel.

  Below are images of transits in front of the Sun, Venus on the left, Mercury on the Right.

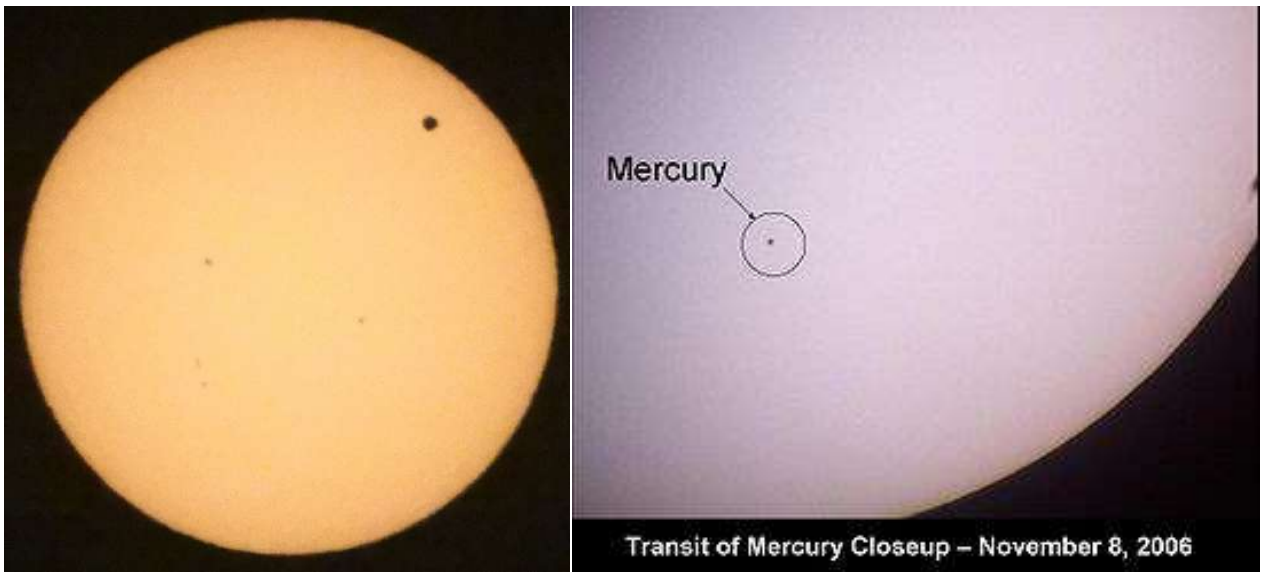

- Venus transits occur in a pattern that generally repeats every 243 years, with pairs of transits eight years apart separated by long gaps of 121.5 years and 105.5 years. Each transit lasts a few hours. Funnily enough, it appears as a black dot, like the Moon should at New Moon!

- Mercury is known to transit in front of the Sun much more frequently than Venus, about thirteen or fourteen times each century. Each transit lasts a few hours. Funnily enough it appears as a black dot, like the Moon should at New Moon! Also Mercury's November transits happen every 7, 13, or 33 years, and Mercury's May transits every 13 or 33 years. Occult numbers galore.

- Both Venus and Mercury transits appear differently in different parts of the world, this helps prove that this is all going on not too far away.

- If Venus and Mercury pass behind and in front of the Sun in the Flat Earth model, it´s highly logical that sometimes they will pass in front of the Moon. In the Flat Earth model the Sun and Moon are both around 3000 miles away, with the Moon slightly closer.

- As Venus is much brighter than Mercury, the star in the Moon just has to be Venus.

***

To conclude, the Moon is not transparent, it is opaque yes, but transparent, no

5. The Moon is Flat Not Spherical

Flat Moon Proof

This is a subject I've had trouble with and have gone around in many circles. I know many others have issue with this too.

First off, through a telescope it *appears* to be a sphere to the human eye and mind, but this doesn't mean it actually is. First off we´ve been told all our lives it is a sphere so our subconscious will project sphere-*ness* onto it.

We saw before how Sunlight and Moonlight clearly have different properties, and furthermore the Moon itself cannot physically be both a spherical body and reflect the Sun's light. Reflectors must be flat or concave for light rays to have any angle of incidence; if a reflector's surface is convex then every ray of light points in a direct line with the radius perpendicular to the surface resulting in no reflection.

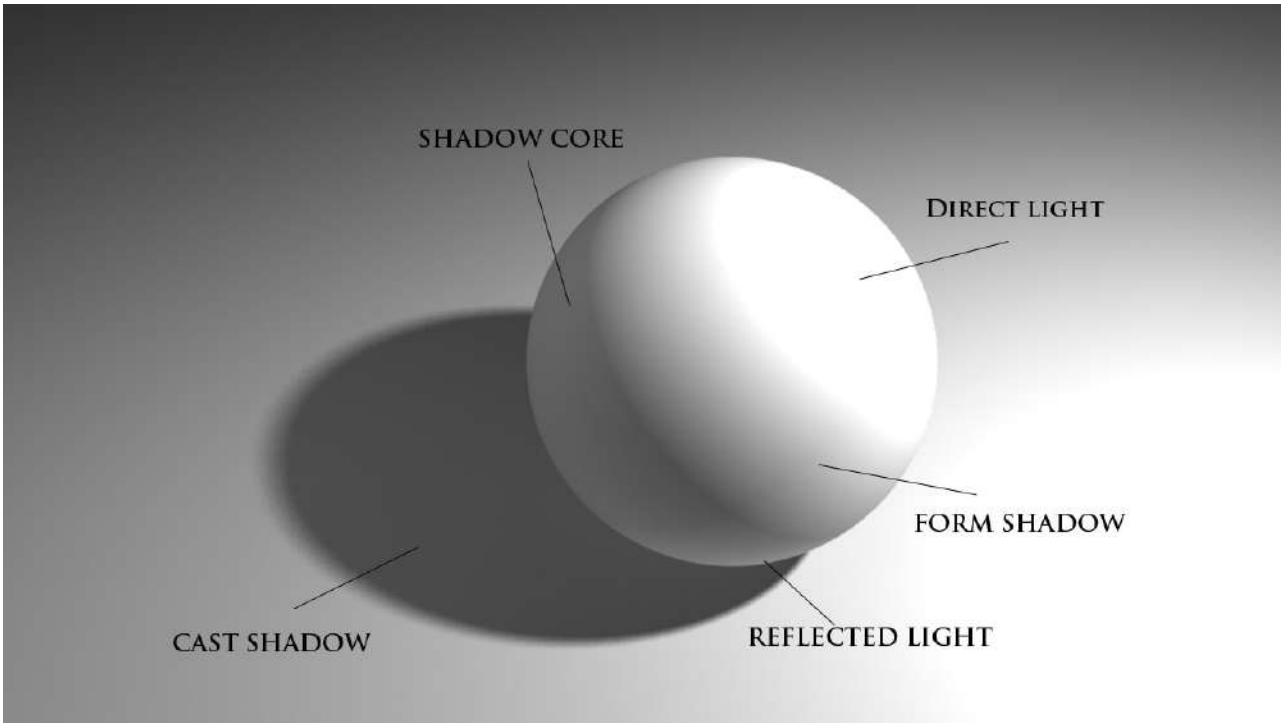

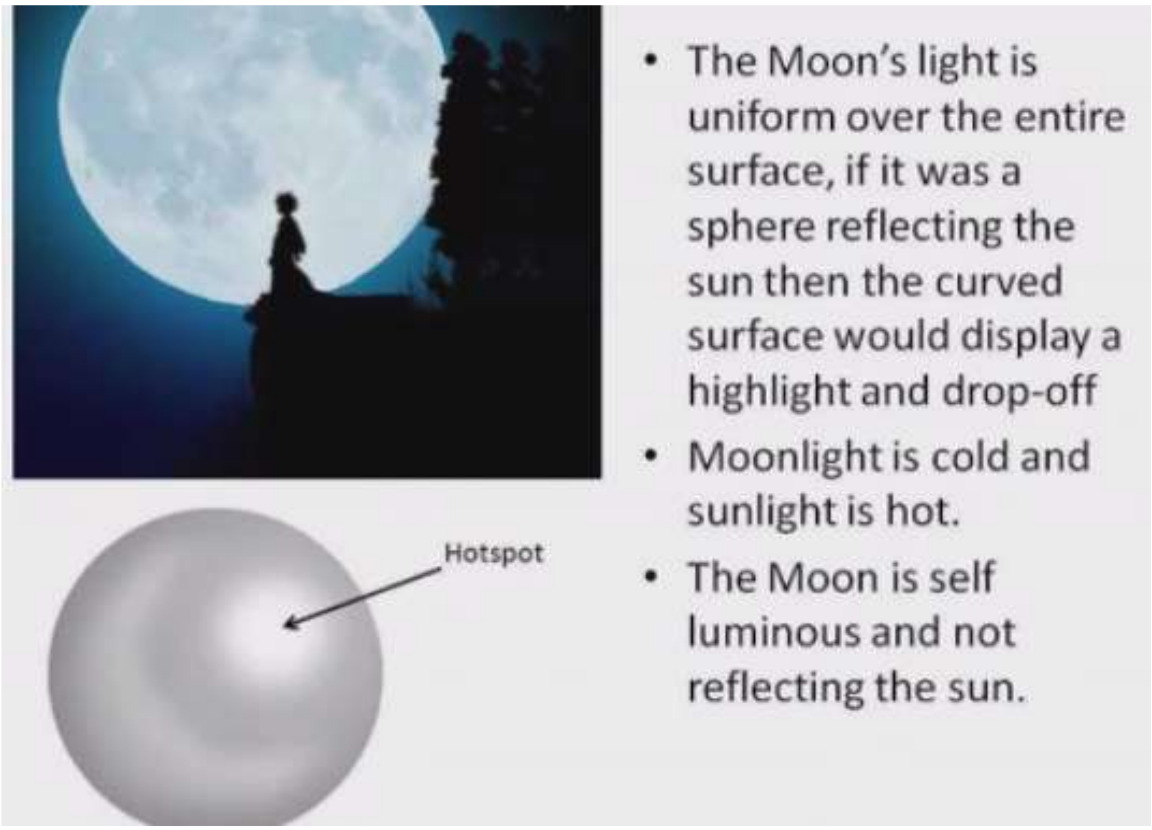

"Again, if the Moon is a sphere, which it is declared to be, how can its surface reflect the light of the Sun? If her surface was a mass of polished silver, it could not reflect from more than a mere point! Let a silvered glass ball or globe of considerable size be held before a lamp or fire of any magnitude, and it will be seen that instead of the whole surface reflecting light, there will be a very small portion only illuminated. But the Moon's whole surface is brilliantly illuminated! A condition or effect utterly impossible if it be spherical." -Dr. Samuel Rowbotham, "Earth Not a Globe, 2nd Edition" (97)

We also have from the Greek legends, the chariot wheels symbolizing the Sun and Moon, and In ancient Egypt a boat is transporting the Sun disk. The ancients always referred to the Sun and Moon as discs, they never said they were spheres.

NASA perpetuate the sphere lie and tell us there is a *dark side* to the Moon that no one ever sees, and that they filmed it from photo-mapping. The thing with this time-lapse *cartoon* movie they concocted is the clouds on the Earth don't move in two *days* of Earth spinning…..

….also in the two day time lapse video, the Moon's phase would not have changed so much, plus, the Moon doesn't seem to move in its own *fake orbit*.

Why no photo of the Moon´s *sphere* in 50 years of space travel? Didn´t they land there?

So it's flat –but this raises two problems.

Problem One. Flat, but how?

When the Moon is retreating from me in Europe (*setting* in common terms), those on the east coast of America would soon see it approaching, but they would see the same face. The Moon isn't a disc that flips 180 degrees when nobody's looking, so how do we all see the same face?

The model has to be as such; the Moon is flat and facing down. This is the only way we can all see the same and only face.

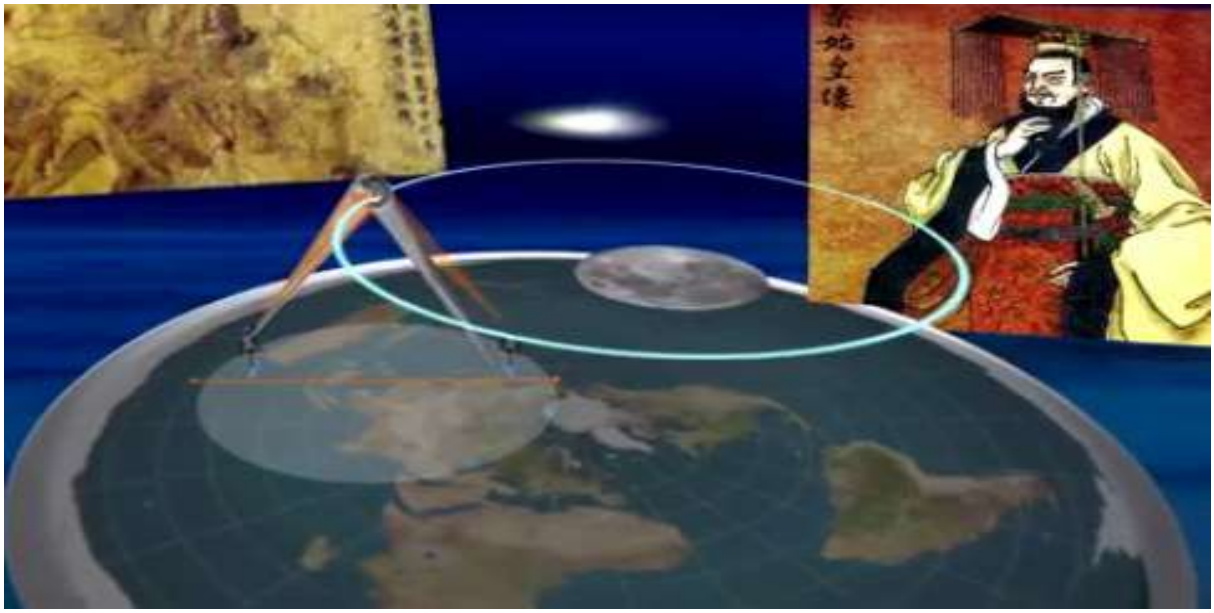

But this raises problem two, and it´s a big problem…..

Problem Two - The Ellipse Issue

If the Moon is flat and facing down, some would say we should not see it sink behind the horizon. But we should, due to our perspective vision and our eyes´ vanishing point.

As stated before, eyes are spherical objects that focus light on the lens at the back. *The horizon* has nothing to do with the shape of the world and everything to do with distance. People are fooled into thinking the horizon is the edge of the curve.

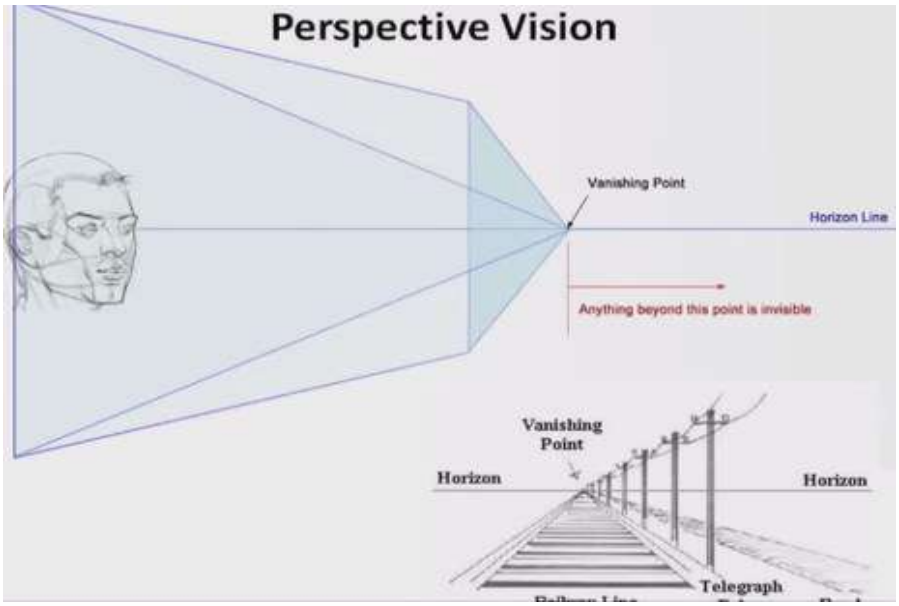

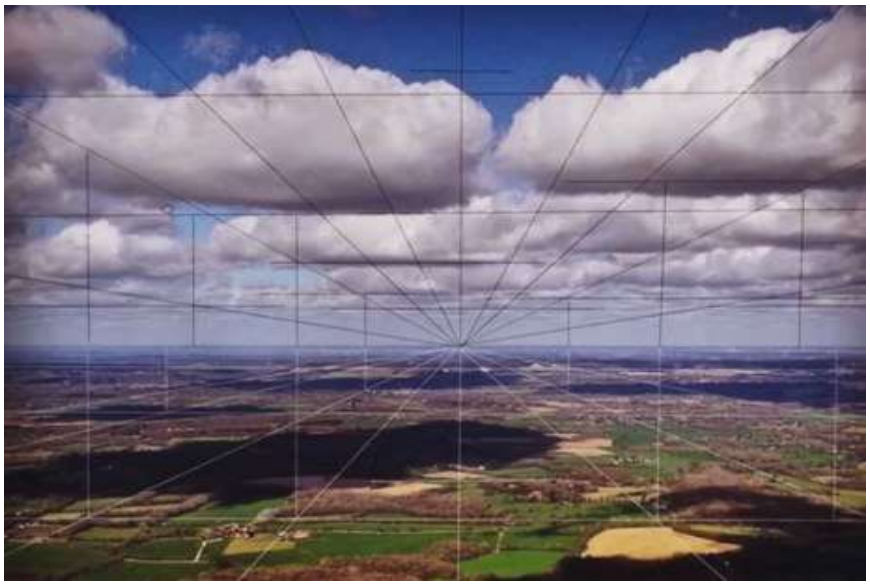

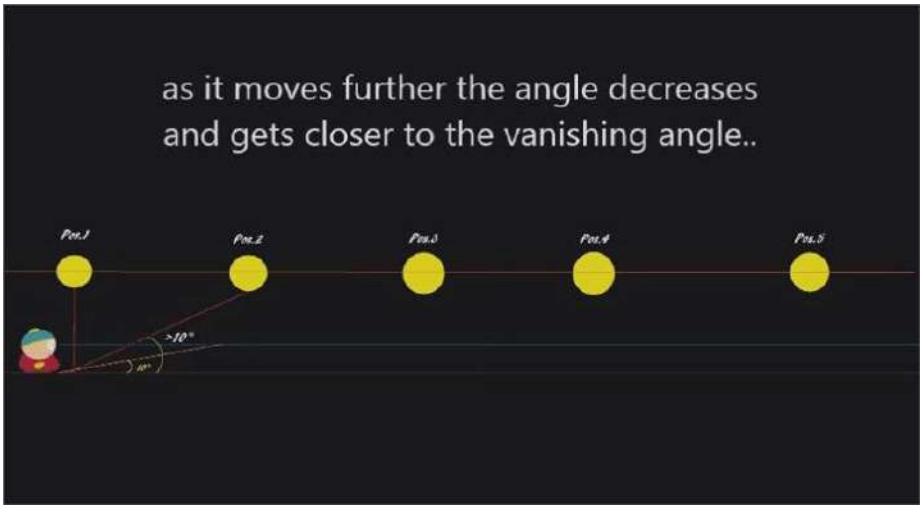

But the problem we find is that the Moon does not slowly appear to our eyes as an ellipse during its departure (increase in distance from the eye), like all other flat face down circles appear to.

It appears to us as a circle, as though it is vertical in relation to our eyes.

Surely within the flat face down model it should turn into an ellipse?

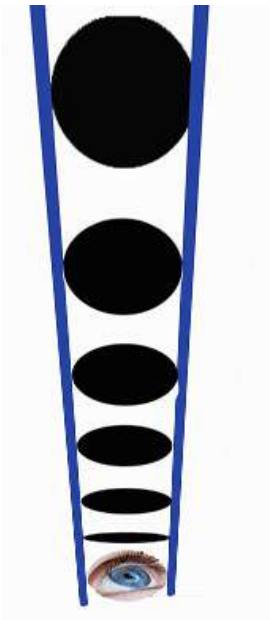

I held mugs over my head and moved them, and made white paper circles of different sizes and got someone to move them away from me, and each crude test (with terrible scale) showed them *set* but also turns into an ellipse.

Now the only reasons for the Moon not appearing to human eyes as an ellipse I can conjure up are:

- Some sort of refraction or reflection through a lens of some kind. Some sort of lens that always reflects the Moon to the facing observer. Some sort of transparent and dense medium. Let us note that a layer of density does exists at around 100km up.
    - Note1: Sprites (plasma discharge) are only below 100km.
    - Note2: Elves (large halos of plasma discharge) are only above 100km

- Note3: The Northern lights start at 100km altitude.
- Note4: Temperature changes around 100km.
- Note5: There is no evidence I've seen to suggest and human technology has gone above 100km.
- Note6: Some see this layer as glass (see Libyan Glass), but this can´t be the case as this glass would be found at every meteor site (see my rift video for more on this).

- In mythology, there is mention of a transparent medium that is located in front of the Sun, like a shield. From the Nordic legends of Svalinn mythology I quote, "The Earth is protected from the full heat of the Sun by the shield Svalinn, which is placed before Sól."

- The Firmament´s dome ceiling (which contains the meteorites and arcs above the stars) has a role to play in how we see the Moon with our eyes. (I see this ceiling as being a thick layer made up from metals, rock, ice, and glass).

- The Moon sits within a layer of liquid of some kind. This may sound strange but when one studies *the star in a jar* (sonoluminescence), it is also something we cannot rule out too quickly. Note that the Moon Goddess Pamuya from the Native American Hopi meant *water Moon*, and the Moon Goddess Lasya from Tibet was a beauty who carried a mirror. Let us note that Genesis sees the model as a Flat Earth with water above the firmament, but yes, the Moon is within / below the firmament ceiling.

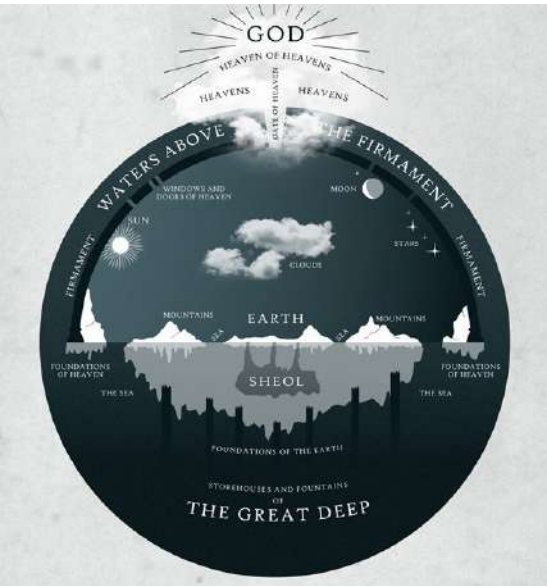

I am not the only one that thinks the firmament *may* have a role to play in how the Moon appears to our eyes when it approaches and departs.

"As already stated, the surface irregularities which were thought to exist on the imaginary masses called planets, are those of the dome of the sky, such as they are seen through the transparent discs. The so called mountains, craters and depressions of the Moon are details of the structure of the dome. It may be remarked in this connection that the astronomers at the Mount Palomar Observatory in America have recently reported that striking alterations have taken place on the surface of the Moon. Gigantic craters and fissures of more than five hundred kilometers are said to have appeared; and

ranges of mountains as important as the Alps have disappeared without leaving a trace; but all those supposed alterations correspond clearly to the successive structural features of the dome which are progressively revealed by the luminous and transparent disc of the Moon as it moves on."
Gabrielle Henriet

If anyone has any other ideas or angles on the ellipse problem, then I am all ears. It just has to be flat and it has to be face down for the solid reasons presented.

6. Disinformation

The Moon has more crazy theories banded about than almost any other subject in our reality.

Here is a quick list, and a quick debunk so we can move on.

- Lunar waves proving it's a hologram – this is from digital cameras *refreshing*.
- It's a projection – From where or what?
- It's a hologram – Then how can its light affect plants, chemicals, animals, and people?
- Spaceship Moon theory – Also known as the Vasin-Shcerbakov theory. This is based on physical a-**lie-**ns, which don't exist (though inter-dimensional entities exist).
- Nazi Space Station on the other side of the Moon – The technology still does not exist to go to the Moon, due to radiation and temperature constraints. I've yet to see enough evidence to suggest that space even exists.
- Tiamat collision theories – Wild conjecture with no evidence or logic.
- Hollow Moon theories – Wild conjecture with no evidence or logic from Hollow Head theorists.
- The Moon is Reflection off of the Firmament – Wild conjecture with no evidence of logic.
- David Icke's Moon Matrix Theory, where reptilians use the Moon as a base, and control humanity by blocking off frequencies – Icke is a classic disinformation agent; Build up a large following by slagging off banks and police for years, and then send everyone the wrong way. This is a man who preaches freedom for $70 a ticket - but he lives in a large, middle class house in a conservative area, and happily pays council tax and corporate bills.

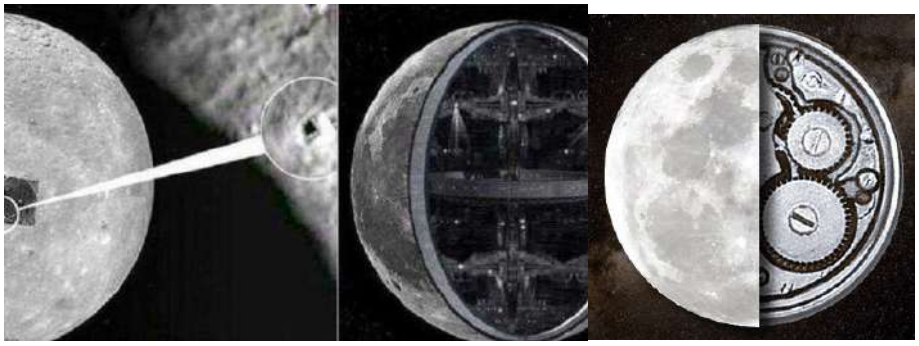

7. Sacred Text, Esoterica, and Myth

The following chapters will contain much more myths and archetypes to back up some claims, so it's good we define a base for this first.

The Moon is awash with mystique, myth, superstition, and esoterica, here´s just a small run down.

The Moon has over a hundred deities from all around the world, dating back as far as around 2700BC.

At Full Moon the pineal gland opens more, energy is more *fizzy,* and ritual and ceremony pack more punch. The spiritual veil is thinner so to speak. Do ayahuasca or a meditation retreat on a New Moon and then on a Full Moon, I guarantee you will experience the big difference in energy and flavor. Or go to a rough nightclub New Moon and Full Moon and witness the difference – it´s a fact there are more accident and emergency admittances around Full Moon. Most humans have dense energy and live in the five senses, and are disconnected from the Moon´s energetic waxing and waning cycles.

The New Moon is like the start of a wave, the Full Moon is like the crashing of the wave, and the Waning Moon is like the decay of a wave retreating.

A healthy woman in her power has her menstrual cycle during or near Full Moon.

It is more beneficial, esoterically, to fast the days before a New Moon.

In astrological terms, the Moon is ones soul, emotion, and subconscious, and is home in the sing of Cancer (feeling, untrusting, subconscious, intuition). I would like to note here that for me, astrology is the cycles of the luminaries that are part of the clockwork, a clockwork that fires down upon us different rays of energy in multitudes of combinations (aspects). This clockwork also allows different soul types to be born, each with their past life karma and emotions attached to the soul matrix.

In the Tarot, the Moon Major Arcana Trump signifies the subconscious, intuition, the past, and illusions. It is a card of deception, and in readings it can mean *projecting fear into your present and your future, based on past experiences*. (Remember this for the last couple of chapters).

Dog - Our trained, domesticated nature. Wolf - Our wild, primitive instincts. Pond - The primordial swamp. Path - Mankind's journey from unconsciousness to spirituality. Towers - illusion of safety. Crayfish/Crab - symbol of astrological Cancer. Distant Mountains - Moving towards the unknown. Yod´s falling - spiritual help.

In the Qabblah, the Moon is the 9th Sephiroth (linking to the Sun 6, Venus 8, and Mercury 7) and also the Moon also appears on the Path of Qoph along with the dreamy Pisces energy.

Let's look at sacred scriptures regarding the Moon, starting with the Bible:

Genesis 1: 14 And God said, Let there be lights in the firmament of the heaven to divide the day from the night; and let them be for signs, and for seasons, and for days, and
years: 15 And let them be for lights in the firmament of the heaven to give light upon the Earth: and it was so. 16 And God made two great lights; the greater light to rule the day [the Sun], and the lesser light to rule the night [the Moon]: he made the stars also. 17 And God set them in the firmament of the heaven to give light upon the Earth, 18 And to rule over the day and over the night, and to divide the light from the darkness: and God saw that it was good.

Matthew 24: 29 Immediately after the tribulation of those days shall the Sun be darkened, and the Moon shall not give her light, and the stars shall fall from heaven, and the powers of the heavens shall be shaken:

Mark 13: 24 But in those days, after that tribulation, the Sun shall be darkened, and the Moon shall not give her light.

Isaiah 13: 10 For the stars of heaven and the constellations thereof shall not give their light: the Sun shall be darkened in his going forth, and the Moon shall not cause her light to shine.

Isaiah 30: 25 And there shall be upon every high mountain, and upon every high hill, rivers and streams of waters in the day of the great slaughter, when the towers fall. 26 Moreover the light of the Moon shall be as the light of the Sun, and the light of the Sun shall be sevenfold, as the light of seven days, in the day that the Lord bindeth up the breach of his people, and healeth the stroke of their wound.

Jeremiah 31: 35 Thus saith the Lord, which giveth the Sun for a light by day, and the ordinances of the Moon and of the stars for a light by night, which divideth the sea when the waves thereof roar; The Lord of hosts is his name:

Ezekiel 32: 7 And when I shall put thee out, I will cover the heaven, and make the stars thereof dark; I will cover the Sun with a cloud, and the Moon shall not give her light.

Revelation 21: 23 And the city had no need of the Sun, neither of the Moon, to shine in it, for the glory of God did lighten it, and the Lamb is the light thereof.

Isaiah 60: 19 The Sun shall be no more thy light by day; neither for brightness shall the Moon give light unto thee: but the Lord shall be unto thee an everlasting light, and thy God thy glory. 20 Thy Sun shall no more go down; neither shall thy Moon withdraw itself: for the Lord shall be thine everlasting light, and the days of thy mourning shall be ended.

The Book of Enoch states:

Chapter 28 states “These are the two great luminaries, whose orbs are as the orbs of heaven; and the dimensions of both are equal” and “Uriel likewise showed me another regulation, when light is poured into the Moon, how it is poured into it from the Sun.”

Chapter 71 states “In which gates also the Moon rises and sets; and I beheld the conductors of the stars, among those who precede them; six gates were at the rising, and six at the setting of the Sun.”

The Quran states:

He is the One who rendered the Sun radiant, and the Moon a light, and He designed its phases that you may learn to count the years and to calculate." (Qur'an 10:5)

"Most blessed in the One who placed constellations in the sky, and placed in it a lamp, and a shining Moon." (Qur'an 25:61)

"He designed the Moon therein to be a light, and placed the Sun to be a lamp." (Qur'an 71:16)

Vedic book, The Holy Gita states:

24 Those who know the Supreme Brahman attain that Supreme by passing away from the world during the influence of the fiery god, in the light, at an auspicious moment of the day, during the fortnight of the waxing Moon, or during the six months when the Sun travels in the north.

25 The mystic who passes away from this world during the smoke, the night, the fortnight of the waning Moon, or the six months when the Sun passes to the south reaches the Moon planet but again comes back.

***

Lunar comes from the Latin word Luna, meaning Moon. The Roman goddess of the Moon is called Luna (Selene in Greek mythology). Luna was the Roman incarnation of Selene, the Greek goddess personifying the Moon. (The word Selene likely comes from a Greek root selas, for

brightness). The adjective lunar is still used today when referring to the activities and features of the Moon. But in Antiquity, lunar was also often employed in reference to both a woman's menstrual cycle and to insanity, (i.e., lunacy, lunatic).

What is the myth of Selene?

1. Selene, the goddess of the Moon, falls in love with Endymion and unites with him, bearing him fifty daughters. She so loves him she cannot bear the thought of his eventual death, so she uses her magic of the night to put him into a deep sleep forever. This way she may see him, unchanging, for all eternity.
2. The beautiful boy Endymion was granted eternal youth and immortality by Zeus and placed in a state of eternal slumber in a cave near the peak of Lydian Mount Latmos. There his heavenly bride Selene descended to consort with him in the night.

The cave could be the dark rift for anyone following my other work.

8. Eclipses

We will break this up into the two different eclipses we can experience; the solar and Lunar Eclipses. What we have been taught about eclipses from a young age is based upon the ball Earth model so is therefore needs to be scrubbed, and we have to start again with a blank canvas.

The word eclipse comes from a Greek word meaning "abandonment." Quite literally, an eclipse was seen as the Sun (Solar Eclipse) or Moon (Lunar Eclipse) abandoning the Earth.

It is a myth that only modern science can accurately predict eclipses, this has been going on for millennia in various cultures. Thales accurately predicted an eclipse which ended the war between the Lydians and Medes all the way back in 600 B.C. Saros astrology can predict eclipses and this goes back to Chaldean and Babylonian astronomers.

Remember that eclipses occur on the Flat Earth model when the Moon is near one of the two FE-Nodes. To produce an eclipse, the Sun must also be near a FE-Node at that time: the same FE-Node for a Solar Eclipse, or the opposite FE-Node for a Lunar Eclipse.

Solar Eclipse

Note that NASA have not taken any photos or videos of a Solar Eclipse from space, since the 1960's. This is because satellites and spacecraft are hoaxes, and nothing has gone past 100km in altitude.

NASA with all their money and *spaceships* just give us a diagram and cartoon style CGI (images below), so we need to forget all of this hogwash, and go back to ancient myth.

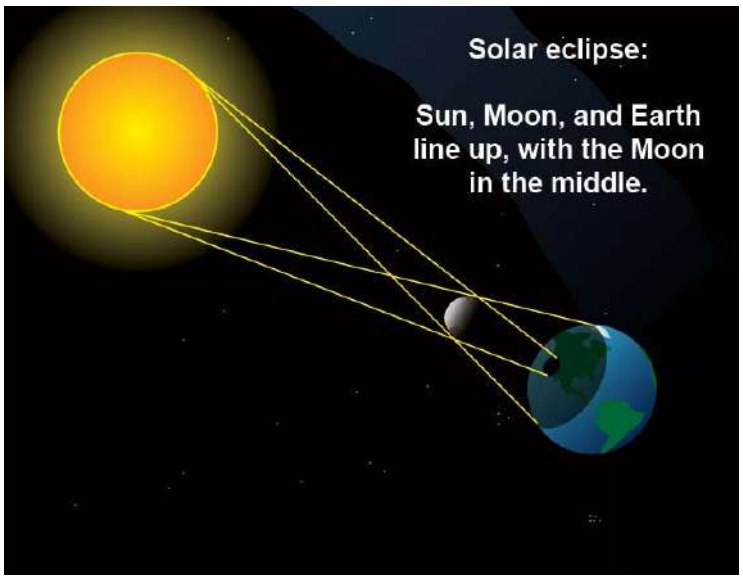

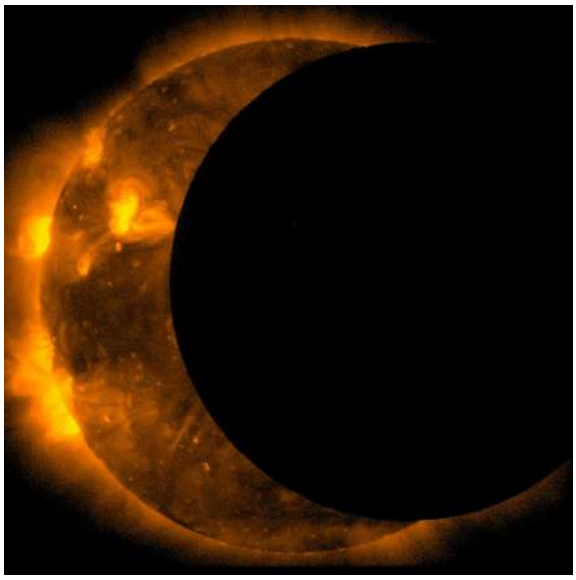

The Solar Eclipse has caused fear, and been associated with myths, legends and superstitions throughout history. Even today, an eclipse of the Sun is considered a bad omen in many cultures, but we will elaborate on this later.

Note: Esoterically and astrologically I would say never to make a big life decision near or on a Solar Eclipse (or even during the solar and Lunar Eclipse *fortnight*) unless you really know what you are doing with bringing the energy down and harnessing it.

In Vietnam, people believed that a giant frog was devouring the Sun, while Viking cultures blamed wolves for eating the Sun. In ancient China, a celestial dragon was thought to lunch on the Sun.

Norse mythology explained that eclipses were the doing of sky wolves, or *warg*, which would chase and eat both the Sun and Moon.

According to ancient Vedic mythology, the deity Rahu, a God of Darkness, attempted to sneak a taste of an elixir of immortality. The Sun and Moon told the god Vishnu about Rahu's crime. Vishnu sliced off Rahu's head as the demon was drinking, so Rahu's head became immortal. In rage and frustration, Rahu's head continues to chase the Sun and Moon, occasionally swallowing them. Because he has no body, however, the Moon and Sun disappear only momentarily, and fall out the bottom of his head. Rahu is also known as The Black Sun or Shadow Planet.

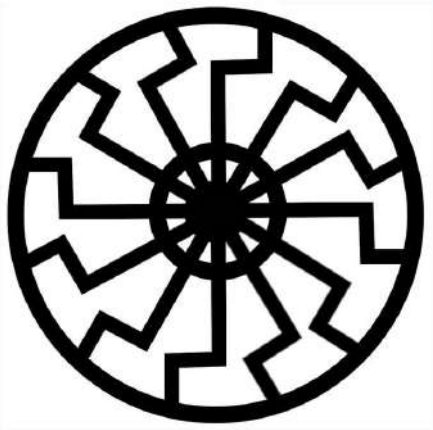

There are scores more myths from around the world regarding the Solar Eclipse, but we will focus on Rahu from Vedic lore more than the others.

***

Remember in the first Chapter we spoke of FE-Nodes; where the Moon crosses the Sun´s trajectory in a given Lunar Month. A Solar Eclipse only occurs when there is a New Moon *and* the Sun and Moon are both at the same FE-Node at the same time. This happens about twice a year on average.

There are three types of Solar Eclipse.

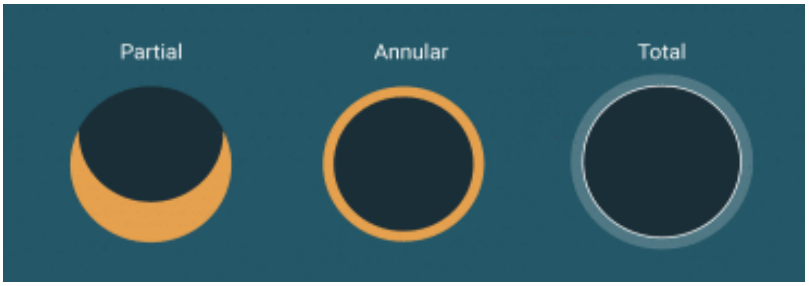

Partial

Annular

Total

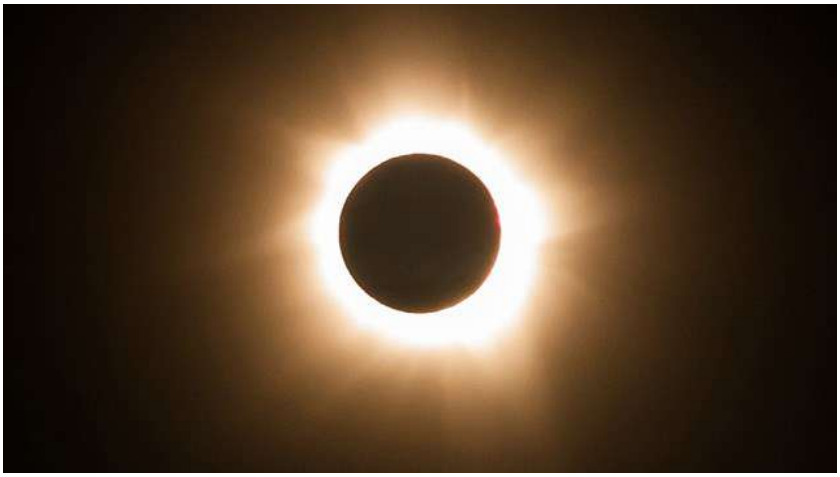

The total and annular Solar Eclipses prove that either the Sun or Moon, or both, vary their altitude from the Earth.

Totality during a full Solar eclipse can only be seen from a very small area on Earth. This area is usually about 100 miles (160 km) wide and 10,000 miles (16,100 km) long. Areas outside this track may be able to see a partial eclipse of the Sun, and most areas will see nothing. These facts all fit very well with the Flat Earth model, and distance and size of the Sun and Moon (that we presented earlier).

Have you ever noticed the path of a Solar Eclipse is always a *thin stripped curve* on the ball model? Place this on the Flat Earth and it´s simply the path of the Sun at the time of the eclipse.

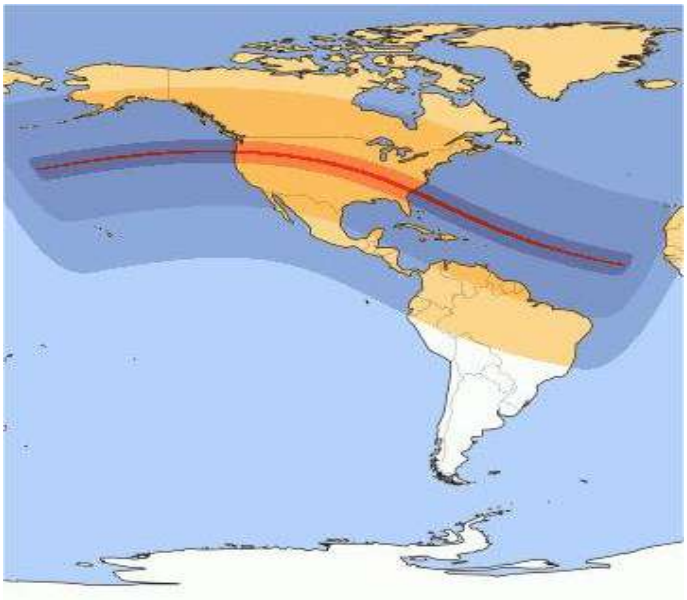

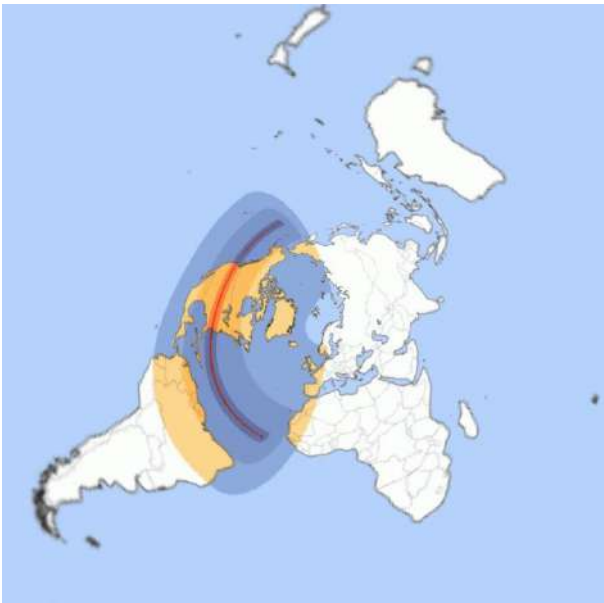

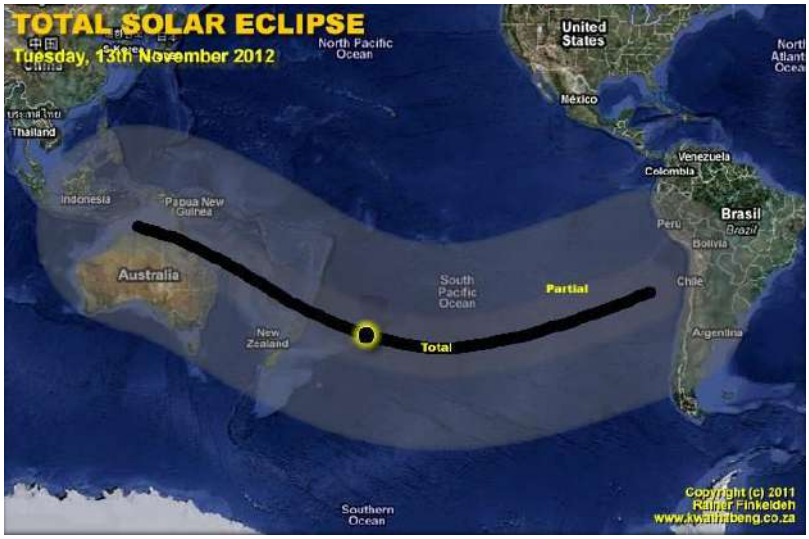

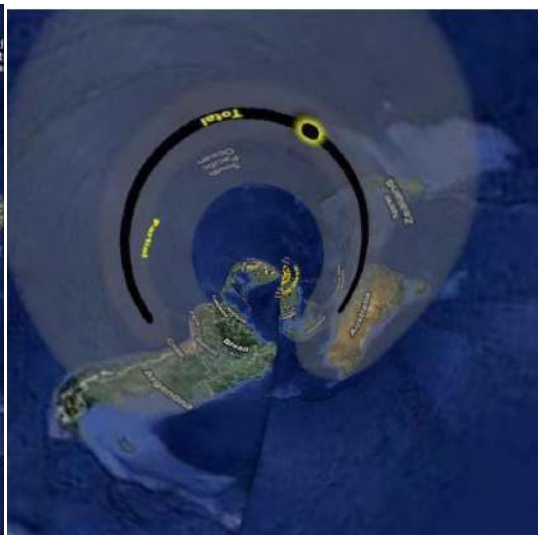

Below are all of the Solar Eclipse paths from 1996 to 2020:

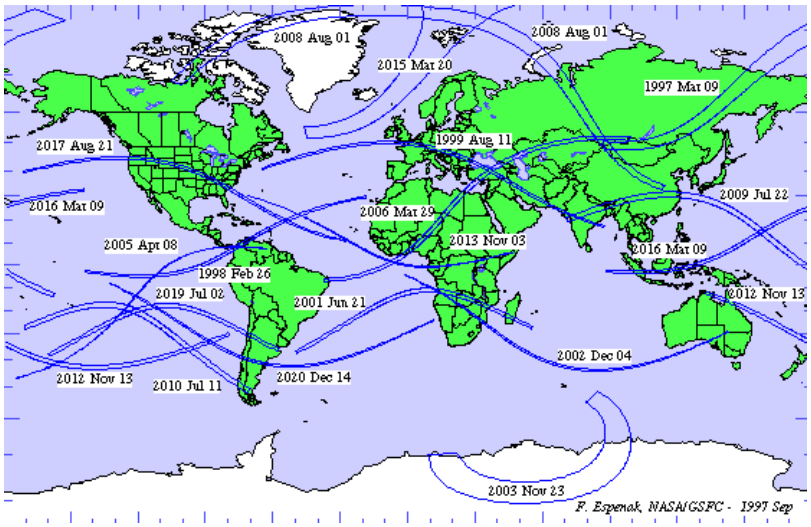

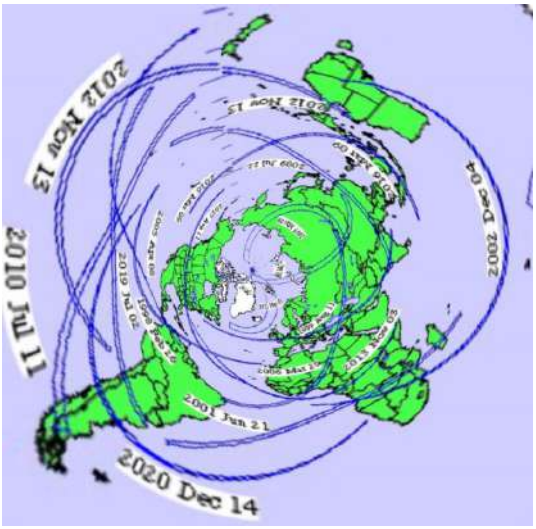

***

So what causes the Solar Eclipse?

Many within the Flat Earth community, many believe there is another object, like a Black Sun (Rahu), one driver for this is that the Moon cannot eclipse the Sun because it´s semi-transparent - but we debunked this earlier. Some even think this *extra body in the sky* creates the phases. I quote from a Flat Earth public forum:

> "Rahu seems to occasionally try to swallow the Sun and the Moon creating solar and Lunar Eclipses but it seems like every month Rahu attempts to swallow the Moon very slowly before the Moon gradually comes out the other end meaning that Rahu either has a hole in the end of is simply blocking to Sun. I wonder if Rahu is an intelligent entity or a dark piece of matter."

> "Solar eclipses aren't caused by the Moon, but by a 'shadow plane-t' known in ancient India as Rahu."

This is good conjecture, has some rationality behind it, but is wrong, and I will present why.

We never have a Solar Eclipse when the Moon has just been visible in the sky. The Solar Eclipses only occur when the Moon is a New Moon. E.g. when the Moon is always very near the Sun.

A total Solar Eclipse lasts for only a few minutes at any given place, so what does that tells us?

It seems that the altitude of the Sun and Moon vary slightly, giving us things like Venus Transits, Mercury Transits, and the different Solar Eclipses. Therefore, we see the width of the path of totality differs from one eclipse to another. This width can even change during a single Solar Eclipse. The track of a Solar Eclipse (path of totality) is a very narrow path over the Earth's surface.

The Solar Eclipses have all the characteristics of an obstructing object passing in front of the Sun (the narrow shadow on the Earth, the small duration, and the eclipse shape).

During a Solar Eclipse, the Sun is indeed obstructed by a round celestial body which I am sure *is* the Moon.

Annular Solar Eclipses are complete proof that a *fully* round object is passing in front of the Sun.

The annular eclipse is just like the Venus transit. The only difference is that the body is much larger.

So could this third Sun (Black Sun, Rahu) be responsible for the annular Solar Eclipse? If we analyze it we see that it doesn't hold weight, because I repeat, the Moon is near the Sun in the days before the eclipse, and it´s getting closer and closer. Plus, in the day of the Solar Eclipse it

is almost exactly near the Sun and during the eclipse it is passing in front of it. There is simply nowhere else the Moon can be! Note: We will go further into Rahu in the next section.

These images are from The Miloslav Druckmüller Institute of Mathematics, Faculty of Mechanical Engineering Brno University of Technology, Czech Republic. I have no reason to believe that these are fake, and the obstructing body does looks very much like the Moon.

Lunar Eclipse

The Lunar Eclipse is a tougher nut to crack, and this book will start going a bit deeper now, so sit down and strap in. This section will be written slowly and clearly, even with some repetition, because the concept has never been written about or talked of before.

Again, NASA has since the 1960´s have taken no photos or videos from space of a Lunar Eclipse. This is because satellites and spacecraft are a hoax, and nothing has gone past 100km in altitude.

NASA with all their money and *spaceships* just give us a diagram and CGI (below), so we need to forget all of this hogwash, and go back to ancient myth.

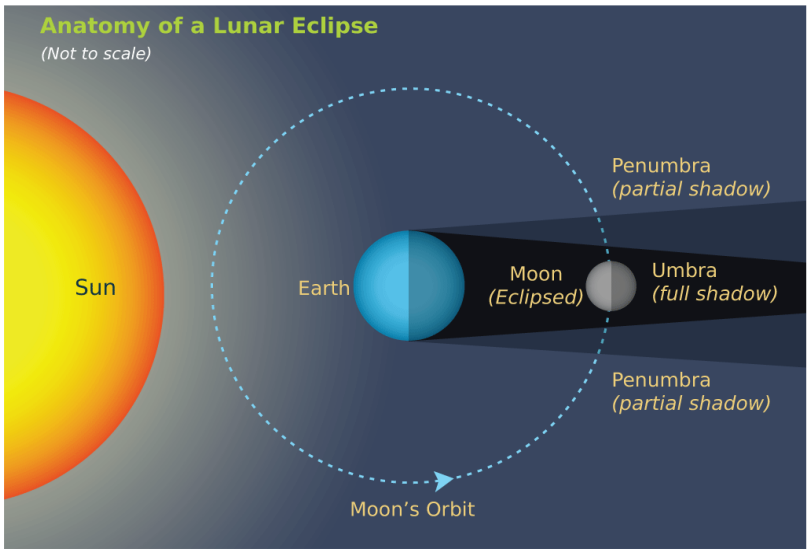

***

In the *Edda*, an ancient collection of Scandinavian poetry, embodying the national mythology, Managarmer is the monster who sometimes swallows up the Moon, and stains the heaven and the air with blood.

"One of the most terrible phenomena to heathens was an *eclipse* of the Sun or Moon, which they associated with a destruction of all things and the end of the world. I may safely assume that the same superstitious notions and practices attend eclipses among nations ancient and modern. The Indian belief is that a serpent eats up the Sun and Moon when they arc eclipsed, or a demon devours them. To this day the Hindus consider that a giant lays hold of the luminaries and tries to swallow them. The Chinese call the Solar Eclipse zhishi (solis devoratio), the lunar yueshi (lunæ devoratio), and ascribe

them both to the machinations of a dragon. Nearly all the populations of Northern Asia hold the same opinion. The Finns of Europe, the Lithuanians, and the Moors in Africa, have a similar belief."
Grimm´s Teutonic Mythology.

"An eclipse of the Moon filled them with dismay; they supposed the planet was *natua*, or under the influence of the spell of some evil spirit that was destroying it. Hence they repaired to the temple, and offered prayers for the Moon's release."
Mariner´s Natives of the Tonga Islands

"The common notion amongst ignorant Mahometans is, that an eclipse is caused by some evil spirit catching hold of the Sun or Moon. On such occasions, in Eastern towns, the whole population assembles with pots, pans, and other equally rude instruments of music, and, with the aid of their lungs, make a din and turmoil which might suffice to drive away a whole army of evil spirits, even at so great a distance."
Discoveries in the Ruins of Nineveh and Babylon, 1853

There are scores more myth about the Lunar Eclipse, but they all seem to state the same two things. *Firstly*, a Lunar Eclipse is believed to be devoured by some evil demon, or by wolves or dogs. This is the superstitious vagary of the Hindus, the Chinese, Asians, Europeans, Africans, Americans, and Polynesians. *Secondly*, a Lunar Eclipse is the precursor of some dreadful calamity to the inhabitants of the Earth. These notions are also traceable in every quarter of the Earth.

Rahu and Ketu are mentioned in the Vedas, and their meanings since Vedic times have not remained static, as we will present in due course. Commonly they are known as invisible celestial bodies (shadow planets, Black Suns) that interact with the Sun and Moon. These demons are said to wreak vengeance on the Sun and Moon whenever they come near. The head of this great demon snake is known as Rahu and his tail is known as Ketu.

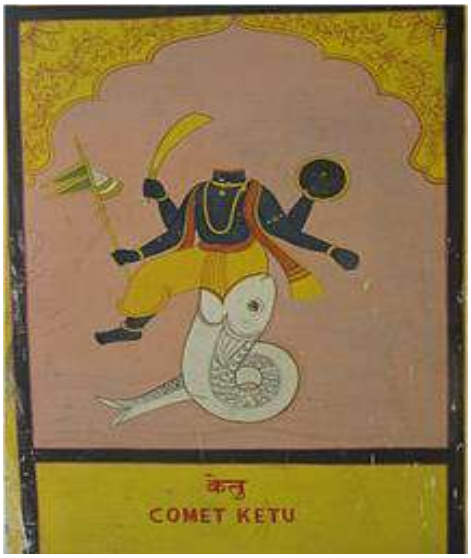

(Note the word comet below Ketu for later).

The Mahabharata from India (Bhismaparva 13.39-45) uses both Svarbhanu and Rahu as interchangeable names. Rahu is a graha, 12000 yojanas in diameter, bigger than both the Moon (11000 yojanas) and the Sun (10000 yojanas). Rahu had to be bigger than the Sun and the Moon so that it could grab them. Note that the term 'graha' here carries the sense of a *grabber* and not that of a body in orbit.

In more modern times they became simply deities for the Lunar Nodes. Note: The Lunar Eclipse only occurs when the Moon is Full, and on an opposite FE-Node to the Sun.

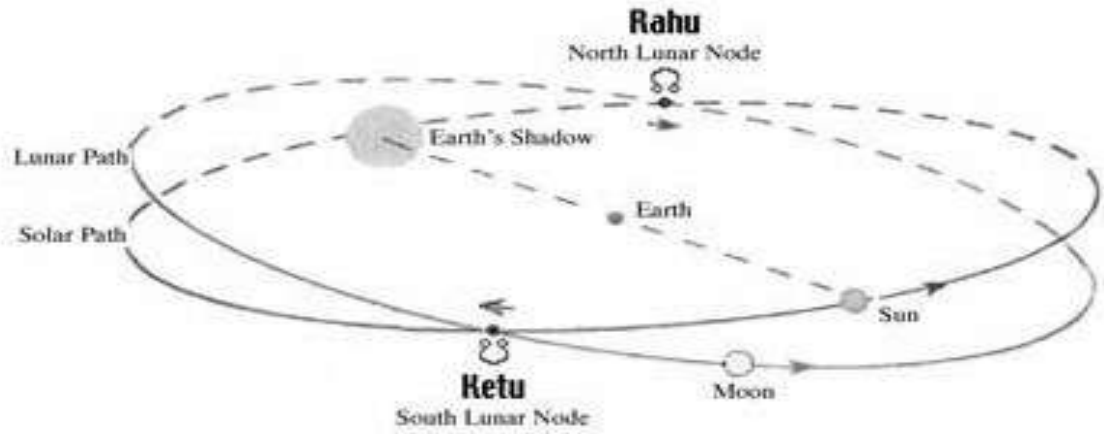

(See the Nodes section in early chapter for the simple FE-Node explanation).

***

Mainstream science tells us that the Lunar Eclipse occurs when the Moon and Sun are on opposite Nodes, and then the spherical Earth casts a shadow onto the spherical Moon from the Sunlight behind the Earth – from a spherical Sun. This can be experienced as a partial Lunar Eclipse, or a Total Lunar Eclipse, or a faint Penumbral Eclipse – each depending on the accuracy of the three spheres lining up.

Mainstream science says this doesn´t happen every month because the three spherical objects are not lined up fully (Not accurate enough placement at the Nodes).

In the Flat Earth model this mainstream indoctrination is all hogwash, as we have proven in earlier chapters. They also say it only happens on a Full Moon and at night. This *is* true within the Flat Earth Model.

In recent years, the term and label *Blood Moon* has become popular when referring to Total Lunar Eclipses, due to the Moon turning ***fiery*** red in the middle stage of the Lunar eclipse.

The mainstream model falls down on so many levels as you should already know by now.

Here is a video of a Lunar Eclipse occurring with the Sun rising, ignore the globehead in the video as his eyes are under hypnosis from his controlled mind.

Here are a few quotes to add some more weight.

"The Newtonian hypothesis involves the necessity of the Sun, in the case of a Lunar Eclipse, being on the opposite side of a globular Earth, to cast its shadow on the Moon: but, since eclipses of the Moon have taken place with both the Sun and the Moon above the horizon, it follows that it cannot be the shadow of the Earth that eclipses the Moon, and that the theory is a blunder." -William Carpenter, "100 Proofs the Earth is Not a Globe" (57)

"That the eclipsor of the Moon is a shadow at all is assumption--no proof whatever is offered. That the Moon receives her light from the Sun, and that therefore her surface is darkened by the Earth intercepting the Sun's light, is not proved. It is not proved that the Earth moves in an orbit round the Sun, and therefore, by being in different positions, conjunction of Sun, Earth, and Moon, 'Day some-times occur.' The contrary has been clearly proved--that the Moon is not eclipsed by a shadow; that she is self-luminous, and not merely a reflector of solar light, and therefore could not possibly be obscured or eclipsed by a shadow from any object whatever; and that the Earth is devoid of motion, either on axes or in an orbit through space. Hence to call that an argument for the Earth's rotundity, where every necessary proposition is only assumed, and in relation to which direct and practical evidence to the contrary is abundant, is to stultify the judgment and every other reasoning faculty." -Dr. Samuel Rowbotham, "Zetetic Astronomy, Earth Not a Globe!" (301)

"According to the globular theory, a Lunar Eclipse occurs when the Sun, Earth, and Moon are in a direct line; but it is on record that since about the fifteenth century over fifty eclipses have occurred while both Sun and Moon have been visible above the horizon." -F.H. Cook, "The Terrestrial Plane"

The Greenwich Royal Observatory recorded that "during the Lunar Eclipses of July 17th, 1590, November 3rd, 1648, June 16th, 1666, and May 26th, 1668 the Moon rose eclipsed whilst the Sun was still above the horizon." McCulluch's Geography recorded that "on September 20th, 1717 and April 20th, 1837 the Moon appeared to rise eclipsed before the Sun had set." Sir Henry Holland also noted in his "Recollections of Past Life" the April 20th, 1837 phenomena where "the Moon rose eclipsed before the Sun set." The Daily Telegraph recorded it happening again on January 17th, 1870, then again in July of the same year, and it continues to happen during Lunar Eclipses to this day.

"As before stated, the Earth is proved to be a plane, without motion, and always several hundred miles below the Sun and Moon, and cannot, by any known possibility come between them. It cannot therefore intercept the light of the Sun, and throw its own shadow upon the Moon. If such a thing were a natural possibility, how could the Moon continue to shine during the whole or any considerable part of the period of its passage through the dark shadow of the Earth? Refraction, or what has been called "Earth light," will not aid in the explanation; because the light of the Moon is at such times "like the

glowing heat of firer tinged with deep red." "Reddish is not the word to express it, it was red--red hot." "The reddish light made it, seem to be on fire." "It looked like a fire smouldering in its ashes." "Its tint was that of red-hot copper." The Sun light is of an entirely different colour to that of the eclipsed Moon; and it is contrary to known optical principles to say that light when refracted or reflected, or both simultaneously, is thereby changed in colour. If a light of a given colour is seen through a great depth of a comparatively dense medium, as the Sun is often seen in winter through the fog and vapour of the atmosphere, it appears of a different colour, and generally of such as that which the Moon so often gives during a total eclipse; but a shadow cannot produce any such effect, as it is, in fact, not an entity at all, but simply the absence of light." Zetetic Astronomy, Samuel Birley Rowbotham, 1881

***

So again, we have to go back to a blank canvas, and really note what we see with our eyes regarding the Lunar Eclipse.

We see the Moon de-illuminate, go fiery red, and then illuminate. Sometimes it de-illuminates fully, sometimes only a little – depending on its accuracy to the Sun at the other FE-Node.

The Lunar Eclipse event cannot be *from* the Sun´s light in any way, whether it is direct Sunlight, Sunlight refracting around the Earth´s atmosphere, or Sunlight somehow bending around the Earth´s atmosphere. Why? Because we are on a Flat Earth Model.

So, what *is* going on? What is making the Moon turn black and red? And why does everyone assume it is a shadow of some kind?

Again, we can somewhat crudely place a Lunar Eclipse viewing area onto the Flat Earth model (note: the dark blue and red signifies 100% visible, and the light blue and white signifies 0% visible). This is from a recent Lunar Eclipse.

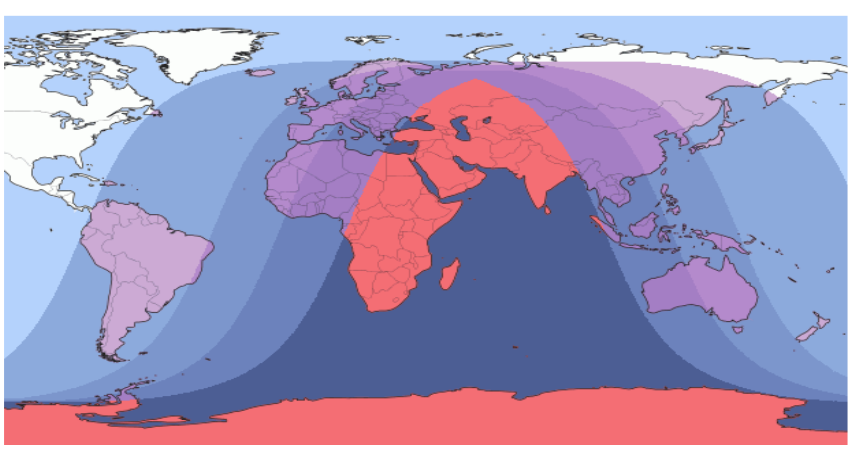

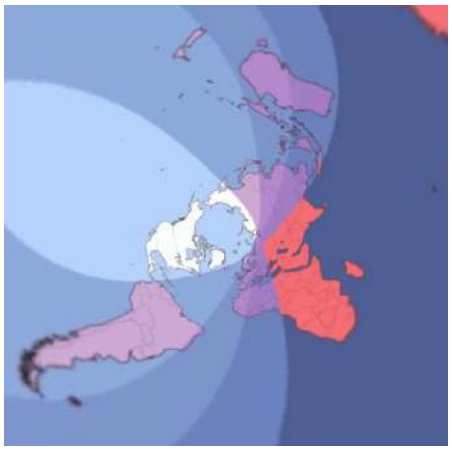

Hopefully you have noticed that there is a massive viewing area of the Lunar Eclipse compared to the Solar Eclipse. This is important as we shall present.

***

So what is truly going on?

It is true that there are three Lunar events we can experience; a total, partial, and a faint de-illumination / shadowing of the Moon, with the Moon turning a fiery blood red in the middle of a total de-illumination / shadowing.

Lunar eclipses look the same to all who can see them, and are seen at the same time, even if these people are thousands of miles apart. This is different to the Solar Eclipse.

Lunar eclipses have all the characteristics of *the source* of light losing its power and changing *its own* color. Before you close the book, please hold on and read further, I will take it slowly.

A Lunar Eclipse is seen wherever it´s night time. If it were the case that the Moon is obstructed by another celestial body (Rahu, Ketu, Shadow Planet, Mount Meru, Black Sun, Lilith) we would have seen the same phenomena as in the case of the Solar Eclipse – only visible in a small location.

A Lunar Eclipse may be viewed from any place on Earth that is in darkness (at night). Lunar eclipses last for a few hours, much longer than the minutes of the Solar Eclipse.

They are completely different events.

As the Lunar Eclipses are seen over a massive land space and at the same intensity, it would mean that any obstructing body would be huge (many times bigger than the Moon), and if this were the case, then the stars which are situated in a large portion of the sky, surrounding the Moon would be eclipsed as well. But that is not the case, and only the Moon is affected.

Note here I am a firm believer that all the stars are above the Moon, but the wandering stars (plane-ts) are much lower and some *may* come very rarely in front of the Moon – see transparency chapter.

If it were Mount Meru at the North Pole creating the shadow (another theory being banded about), or the Black Sun, Rahu etc, it simply would be too huge.

YouTube videos and forums that spout these sorts of diagrams have not thought it through enough, because this would affect our sight to the Moon and sky at other times of the month, even daily it would cause a visual experience that we just don´t witness.

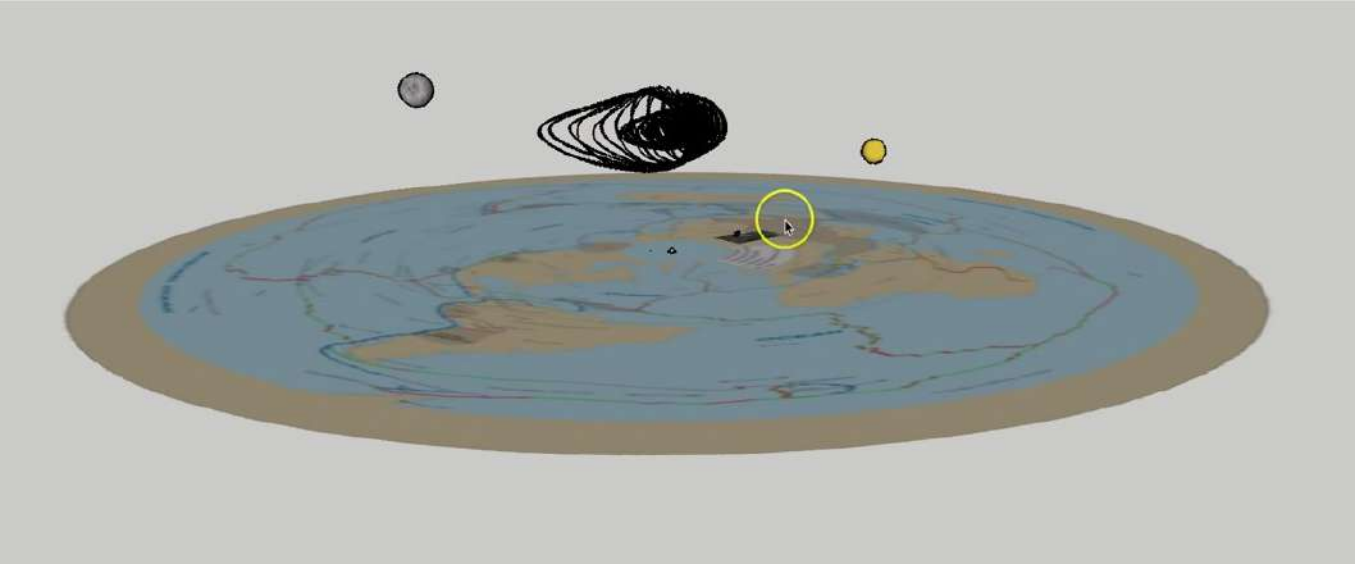

The change in color during a Lunar Eclipse is seen simultaneously by people in different continents, and because of this, the obstruction theory is impossible. Rahu and Ketu we will sort out soon enough, they *do* hold relevance to the Sun and Moon, but we must proceed with the Lunar Eclipse explanation.

This event only occurs at a Full Moon, why?

Since the Earth is flat, we already know that the phases of the Moon are not caused by the shadow of the Earth or from the Sun´s light.

But Lunar Eclipses never happen during the other phases of the Moon, no partial, penumbral, or total eclipse of the Moon occurs. There is no Moon eclipse when the Moon is not full. Why is this?

There are no eclipses because during the other phases of the Moon, the Moon *is* actually *eclipsed* but not by another celestial body, **but by itself, it just loses its power**, and it does exactly the same thing when it´s a Full Moon sometimes, when the Sun and Moon are both at opposite FE-Nodes, when the Sun has more of an energetic link to it. An aspect in astrological terms.

Only when it illuminates and de-illuminates in the space of a few hours do people call it a Lunar Eclipse. When the Moon illuminates and de-illuminates more slowly, people call it phases.

During the phases' period, which last a long time, the Moon is not obstructed by anything, and it illuminates and de-illuminates. During the Lunar Eclipse period, which lasts a shorter time, the Moon is not obstructed by anything, and it also illuminates and de-illuminates.

The Moon also turns a fiery, blood red; this is it sparking, trying to regain its power. It has its link to the Sun in geometric alignment at this time via the FE-Nodes, and it wants to regain its power. It is fully illuminated (Full Moon), aligned to the Sun, and it is the best and only time for it to spark / jolt / surge it´s power. It´s the time when it´s energetically linked to the Sun in a geometric aspect, and highly probably at exactly the same height. (We saw earlier the Moon and/or Sun´s altitude has to vary slightly).

What´s this about the Moon losing power, regaining power, surging in power sparking in power?

The next chapter will go into this thesis alone as for sure it needs some weight to back up this claim, for now we must continue with the Lunar Eclipse.

***

Because both the Solar and Lunar events are both labeled Eclipses, people automatically think they are similar events, but it´s misleading. They are completely different events.

Another major key in all of this is what happens to the Moon just before and just after it turns red. A small portion of the Moon starts to illuminate much brighter than it was the previous minutes, hours, and even days.

The first illuminating portion of the Moon after the red colour of the Moon shines really bright.

The last illuminating portion of the Moon just before the red colour of the Moon also shines really bright.

Before the red experience, this bright illuminated portion gets bigger and brighter, then the whole Moon gets illuminated and then it turns red. It is like it is rebooting itself, or surging in power, or energetically handshaking with the Sun across at the other FE-Node.

An obstructing object (Rahu etc) simply cannot do this.

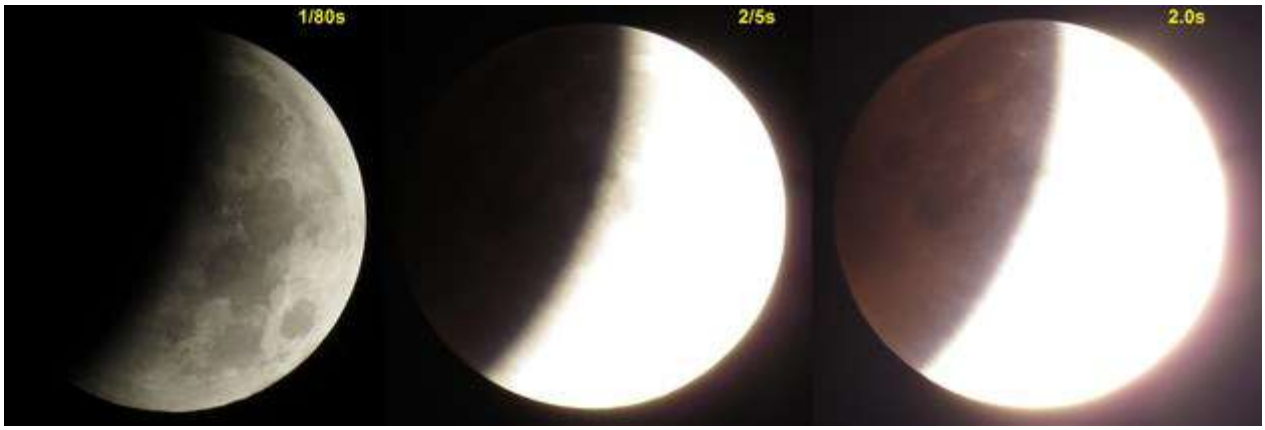

Watch these two time lapse videos of the Lunar Event, and let go of what you´ve been taught. What is really happening? Watch passively with your eyes and not your mind. Video1 Video2.

The Moon is losing power (or preparing energy, retracting), and then sparks (surges, pulses) red, then regains power. This could be seen as the Moon simply malfunctioning, akin to a broken electronic gadget. It looks like a malfunctioning spotlight that has short-circuited, and it´s blinking / flickering. It seems to be a small reset of an already malfunctioning device.

Remember, this event only happens on a Full Moon (already powered up) and when the Sun is in a geometric alignment, and with an FE-Node alignment (energy can be shared between the Sun and Moon, akin to a super tight and accurate astrological opposition aspect).

Also, really think about these images below, and the labels you have used all your life. On the left is a photo of the Moon in phase, and on the right is a photo of a Lunar Eclipse.

There is no difference in the core of both events; the Moon just de-illuminates in both instances.

***

Lunar eclipses usually don´t occur in any specific order. However, every once in a while, four total Lunar Eclipses happen in a row. This is called a Lunar Tetrad. These total Lunar Eclipses happen 6 months apart, having at least six full Moons between two total Lunar Eclipses in a tetrad.

At the time of writing we are coming up to the final Total Lunar Eclipse within the current Tetrad, funnily enough, a Tetrad that contains links to Jewish festivals. There is much information - Cern relevance, plus Hollywood symbolism stating a meteorite or comet may hit the Earth. Pick one from disinformation, top shelf fear porn, an illuminati joke, an illuminati controlled event, or something actually coming. We will see. It should be mentioned here though that myth all around the ancient world saw the red Moon as a sign of cataclysm.

9. The Moon is an Old Malfunctioning Sun Thesis - Introduction

Let go of old notions and *feel* the initial logic.

- The Moon is the same size as the Sun.
- The Moon is at nearly the same altitude as the Sun.
- The Moon travels in almost the same trajectory as the Sun (a five degree, skewed tilt difference, wherein the Moon crosses the Sun´s trajectory twice a month in what we have called FE-Nodes).
- The Moon nearly travels at the same speed as the Sun, but is slightly slower as it´s lost its power and has malfunctioned. The Sun catches it and overtakes it once a lunar month at the New Moon.
- The Moon self illuminates and self de-illuminates for phases.
- The Moon self illuminates and self de-illuminates for Lunar Eclipses.
- The Moon turns a fiery blood red when powered up (Full Moon) when at the same time holds a geometric aspect to the Sun (strangely labeled a Lunar Eclipse).
- The Sun has been known to turn the same very fiery red too (explained below).
- Both affect plants and animals.....and all life on the planet.
- Both have evidence of being electric in their nature.
- Both send intelligent light waveforms to the ground of the Earth.

Obviously we need much more information than the above. So let's go to old books, scriptures, and the ancient myths once again, and include the related Sun information before we start to tie things together with more weight.

It is written in scripture that the Sun we know now also turned red, blood red. The Sun we know now also lost *its power* and turned red.

"In AD 542, according to Hector Boetius, "The Sun appeared about noondays, **all wholly of a bloody colour**. The element appeared full of bright stars to every man's sight, continually, for the space of two days together."
Comet/Asteroid Impacts and Human Society: An Interdisciplinary Approach, Peter T. Bobrowsky, Hans Rickman – 2007

"On April 23, 1547, in England, France and Germany, the Sun appeared for three days as if it were suffused by blood while at the same time many stars were visible at noon."
Cosmos: A Sketch of a Physical Description of the Universe, Vol4, Alexander von Humboldt

"On the forenoon of October 11th, 1520, an eclipse of the Sun was expected. At eight seconds past ten a.m. The Sun, having then reached the altitude of 42°, began to lose its brightness, and gradually continued so to do, **changing to a dark red colour**, without any cloud intervening that could be perceived. No part of the body of the Sun was hid, but the whole appeared as when seen through a thick smoke, till it passed the altitude of 44½°, after which it recovered its former lustre."
Discoveries in the South Sea, p. 39, by Captain James Burney.

So we see that the stars near the Sun were visible with the Blood-Sun too, meaning the sky was not obstructed by anything. It seems the Sun had some problem, it lost its power and it turned red.....just like the Moon does.

Below are some ancient examples of there not even being a Moon in the sky.

The period when the Earth was Moonless is probably the most remote recollection of mankind. Democritus and Anaxagoras taught that there was a time when the Earth was without the Moon.(1)

Aristotle wrote that Arcadia in Greece, before being inhabited by the Hellenes, had a population of Pelasgians, and that these aborigines occupied the land already before there was a Moon in the sky above the Earth; for this reason they were called Proselenes.(2).

Apollonius of Rhodes mentioned the time "when not all the orbs were yet in the heavens, before the Danai and Deukalion races came into existence, and only the Arcadians lived, of whom it is said that they dwelt on mountains and fed on acorns, before there was a Moon." (3)

Plutarch wrote in The Roman Questions: "There were Arcadians of Evander's following, the so-called pre-Lunar people."(4)

Similarly wrote Ovid: "The Arcadians are said to have possessed their land before the birth of Jove, and the folk is older than the Moon." (5)

Hippolytus refers to a legend that "Arcadia brought forth Pelasgus, of greater antiquity than the Moon."(6)

Lucian in his Astrology says that “the Arcadians affirm in their folly that they are older than the Moon.”(7)

Censorinus also alludes to a time in the past when there was no Moon in the sky.(8)

Some allusions to the time before there was a Moon may be found also in the Scriptures. In Job 25:5 the grandeur of the Lord who “Makes peace in the heights” is praised and the time is mentioned “before [there was] a Moon and it did not shine.” Also in Psalm 72:5 it is said: “Thou wast feared since [the time of] the Sun and before [the time of] the Moon, a generation of generations.” A “generation of generations” means a very long time. Of course, it is of no use to counter this psalm with the myth of the first chapter of Genesis, a tale brought down from exotic and later sources.

The memory of a world without a Moon lives in oral traditions among Indigenous tribes. The Indians of the Bogota highlands in the eastern Cordilleras of Colombia relate some of their tribal reminiscences to the time before there was a Moon. “In the earliest times, when the Moon was not yet in the heavens,” say the tribesmen of Chibchas.(9)

In the memory of mankind, no Moon accompanied the Earth.(10)

Since human beings already peopled the Earth, it is improbable that the Moon sprang from it: there must have existed a solid lithosphere, not a liquid Earth. Thus while I do not claim to know the origin of the Moon, I find it more probable that the Moon was captured by the Earth. Such an event would have occurred as a catastrophe.(11)

If the Moon’s formation took place away from the Earth,(12) its composition may be quite different.

There is no evidence to suggest whether the Moon was a planet, a satellite of another planet, or a comet at the time of its capture by the Earth. Whatever atmosphere it may have had(13) was pulled away by the Earth, by other contacting bodies, or dissipated in some other way.

References for the above claims and quotes:.

1. Hippolytus, Refutatio Omnium Haeresium V. ii.
2. Aristotle, fr. 591 (ed. V. Rose [Teubner:Tuebingen, 1886]). Cf. Pauly’s Realencyclopaedie der classischen Altertumswissenschaft, article “Mond” ; H. Roscher, Lexicon d. griech. und roemisch. Mythologie, article “Proselenes.”
3. Argonautica IV.264.
4. Plutarch, Moralia, transl. by F. C. Babbit, sect. 76.
5. Fasti, transl. by Sir J. Frazer, II. 290.
6. Refutatio Omnium Haeresium V. ii.
7. Lucian, Astrology, transl. by A. M. Harmon (1936), p. 367, par. 26.
8. Liber de die natali 19; also scholium on Aristophanes’ Clouds, line 398.
9. A. von Humboldt, Vues des Cordillères (1816), English transl.: Researches Concerning the Institutions and Monuments of the Ancient Inhabitants of America, (1814), vol. I, p. 87; cf. H. Fischer, In mondener Welt (1930), p. 145.
10. [In addition to the sources cited above, cf. The Nihongi Chronicles of Japan (I.ii, in Transactions and Proceedings of the Japanese Society, vol. I [1896]) which recount how “Heaven and Earth . . . produced the Moon-god.” The Kalevala of the Finns recalls a time “when the Moon was placed in orbit.” (Rune III.35)]
11. [Cf. The effects of such an event on the Earth’s rotation calculated by H. Gerstenkorn in Zeitschrift fuer Astrophysik, 36 (1955), p. 245; cf. idem, in Mantles of the Earth and the Terrestrial Planets, S. K. Runcorn ed., (New York, 1967); also idem in Icarus 9 (1968), p. 394.]
12 .[Cf. H. Alfven and G. Arrhenius, “Two Alternatives for the History of the Moon,” Science 165 (1969), 11ff.; S. F. Singer and L. W. Banderman, “Where was the Moon Formed?” Science 170 (1970), 438-439: “.
13. [Cf. Yu. B. Chernyak, “On Recent Lunar Atmosphere,” Nature, 273 (15 June, 1978), pp. 497ff.

"And the Sun, and the Moon, and the stars were now all established. Yet was not the Sun then in the beginning the same as now; his heat wanted force, and he was but as a reflection in a mirror; verily, say the histories, **not at all the same Sun as that of to-day. Nevertheless he dried up and warmed the surface of the Earth**, and answered many good ends."
Bancroft, H. H., "The Native Races, vol. 3-Myths and Languages" P51. (Central American myth)

"Immediately on the death of the gods the Sun began his motion in the heavens; and a man called Tecuzistecatl, or Tezcociztecatl, who, when Nanahuatzin leaped into the fire, had retired into a cave, now **emerged from his concealment as the Moon**."
Same book as above, P62. (Central American myth)

"The present is the Fourth Age. To it appear to belong the falling of the goddess-born flint from heaven, the birth of the sixteen hundred heroes from that flint, the birth of mankind from the bone brought from hades, **the transformation of Nanahuatzin into the Sun, the transformation of Tezcatecatl into the Moon**, and the death of the sixteen hundred heroes or gods. It is called the Sun of Fire, and is to be ended by a universal conflagration." Same book as above, P64 . (Central American myth)

Jain cosmology records tell us that there were two Sun and two Moon revolving over Jambudvip (Earth), as do other myths.

"This is the age of the fifth Sun..... Eventually, **two bright Suns rose in the sky**. Angry that Tecciztecatl continues to follow Nanauatl, the other gods throw a rabbit at him, dimming the Sun and leaving an imprint of a rabbit on his face. This is why the Aztecs say there is a rabbit in the Moon."
Welker, Glen. Aztec Creation Story. Indigenous People's Literature

"And it is the existence of two rival celestial bodies that, as we will see further, is an important motif of theMaya and, in particular, Lacandon mythology,as well as the need to subsequently eliminate one of them.In the Central Mexican cosmological story, one of the angry gods threw a rabbit in the face of the coward Tecuciztécatl, causing him to lose his brilliance**. Tecu- ciztécatl thus became the false Sun with weak bril-liance – the Moon**"
www.academia.edu/9523131/THE_MAYA_MYTH_ABOUT_TWO_SUNS

If you think these are crazy old quotes, google recent events of humans witnessing two Suns in the sky (see images below). Some witnesses claim that one Sun came from behind the one we know.

Moving on…..The powers that be invented the term, *Sun dogs* for a sky phenomena seen many times throughout history. Therefore most of *this* grouping of eye witness accounts from the past and recent times are ignored. Maybe people didn't see reflections, they saw parts of our reality that they were not supposed to see?

"My beloved children, I want to tell you that on the day after the departure of our brothers Kuntz and Michel, on a Friday, we saw three Suns in the sky for a good long time, about an hour, as well as two rainbows. These had their backs turned toward each other, almost touching in the middle, and their ends pointed away from each other. And this I, Jakob, saw with my own eyes, and many brothers and sisters saw it with me. After a while the two Suns and rainbows disappeared, and only the one Sun remained. **Even though the other two Suns were not as bright as the one, they were clearly visible**. I feel this was no small miracle."
Jacob Hatter, Brotherly Faithfulness: Epistles from a Time of Persecution

And let us not forget how Hollywood love to put things in our faces.

There have also been reports of two Moons in recent centuries, The first major claim of another Moon of Earth was made by French astronomer Frédéric Petit, director of the Toulouse Observatory, who in 1846 announced that he had discovered a second Moon in an elliptical orbit around Earth. It was claimed to have also been reported by Lebon and Dassier at Toulouse, and by Larivière at Artenac Observatory, during the early evening of March 21, 1846. Also In 1898 Hamburg scientist Dr. Georg Waltemath made a similar claim, and in 1918, astrologer Walter Gornold made a similar claim. Some said this was Lilith, but this has not been proved.

Is there other evidence of the Sun losing its power more recently?

There is a story of a cataclysm in the 6th century few really know of. Between the years 535 and 536, a series of major global climatic events took place that could easily be described as a global cataclysm with catastrophic consequences:

"There was a sign from the Sun, the like of which had never been seen and reported before. The Sun became dark and its darkness lasted for 18 months. Each day, it shone for about four hours, and still this light was only a feeble shadow. Everyone declared that the Sun would never recover its full light again."
John of Ephesus, Church Leader, Historiae Ecclesiasticae

"… during this year a most dread portent took place. For the Sun gave forth its light without brightness...and it seemed exceedingly like the Sun in eclipse, for the beams it shed were not clear."
History of the Wars: The Vandalic War.Procopius, who lived between 500 and 565 AD

The sixth century writer, Zacharias of Mytilene, who authored a chronicle that contains a section referring to the 'Dark Sun' for the period of 535/536 AD:

"The Sun began to be darkened by day and the Moon by night, while the ocean was tumultuous with spray from the 24th of March in this year till the 24th of June in the following year... And, as the winter was a severe one, so much so that from the large and unwonted quantity of snow the birds perished... There was distress... among men... from the evil things."
Zacharias of Mytilene (Chronicle, 9.19, 10.1)

In the mid-sixth century, Europe and Asia experienced the most severe and protracted episode of cooling of the last 2,000 years.
Claire Bates, Mail Online

The year 1816 became known as the year without summer with extremely cool temperatures all around the Earth, and snow in rare places in the summer.

NASA brings up excuses for the malfunctioning Sun in the above events by saying there was an asteroid or volcano eruption, accompanied by some trademark CGI, but don´t be fooled, the Sun lost its power.

Another clue, which helps to prove that the Moon was once a Sun, though now at different states of health and life, is that both affect the tides (as we mentioned in an earlier chapter).

Eric Dollard (called a modern day Tesla), says that the current Sun is losing power. He says the Sun burns no energy, actually burns nothing, and has no fusion, but is more a convertor of electromagnetic light. A giant Tesla lamp, so to speak.

So, the Sun is not just a burning celestial body that's emanating heat like a camp fire, because if it were it would have been hotter on mountain peaks or at 20 miles' altitude, for example. But that is not the case, it gets very cold as we go up to 100km. We know this for a fact.

So then, the light from the Sun reacts with the dense atmosphere at ground level and it produces heat.
In interactions with matter, visible light primarily acts to elevate electrons to higher energy levels, thus we see that visible light and gamma rays and microwaves are really the same things. They are all electromagnetic radiation; they just differ in their wavelengths.
Most of the electromagnetic radiation from the Sun is in the form of visible light. Light is made up of waves of different frequencies. These frequencies are interpreted by our brain as colors. Infrared waves and ultraviolet waves are two types of waves from the Sun that we cannot see. So because the atmosphere is so dense at ground levels the microwaves from the Sun generate heat by generating electron excitation in the air gases.
The same phenomena happens also to living organisms. We all know the heating effect that Sunlight has on our skin. The denser an object is - the greater is the heating effect.
This type of heat transfer can be observed on Sunny days. Your face will feel warm when you are standing in the Sun. The Sunlight is absorbed by your face, and warms your face without warming the air around you at the same level. The energy from the Sun that is absorbed by your face is called radiant energy or radiation. Radiation is the transfer of this heat energy by electromagnetic waves.

So the light that comes from the Sun is complex and intelligently designed with many frequencies.

Because the Moon is a malfunctioning Sun, the frequencies are changed, and only a fraction of the spectrum is emanated because of the dimmed power, so the effect it has on the Earth's surface has changed.

Please read the online article, "Ultraviolet Spectral Radiant Energy Reflected from the Moon" by Ralph Stair and Russell Johnston, from 1953.
Note: Some systems may not open this link due to a Microsoft bug, if this is the case, copy the below link, apologies;
http://nvlpubs.nist.gov/nistpubs/jres/51/jresv51n2p81_A1b.pdf

In the above article we see that some researchers discovered that the Moon has some of the wavelengths found in the Sun. Of course they interpreted the data from the round Earth perspective, and they incorrectly believed that the frequencies they were recording were reflected by the Moon. But since the Moon is self-illuminated, the ultraviolet spectral energy was not reflected from the sun, and instead was produced by the Moon itself.

We all know for sure that the Sun has a huge role in sustaining life on earth, and nobody can contradict its importance. But what about the Moon?
Samuel Rowbotham presented some good proofs based upon observations, and also upon experiments; that the moonlight is detrimental to life on earth.
And it's exactly what we can expect from a malfunctioning device, its life sustaining functions are not present anymore and the light it emits now is not so beneficial for the life on earth, since the majority of the life giving frequencies and wavelengths are missing.

***

The light emanated by the Sun contains multiple wavelengths, and visible light is only a small part of it.

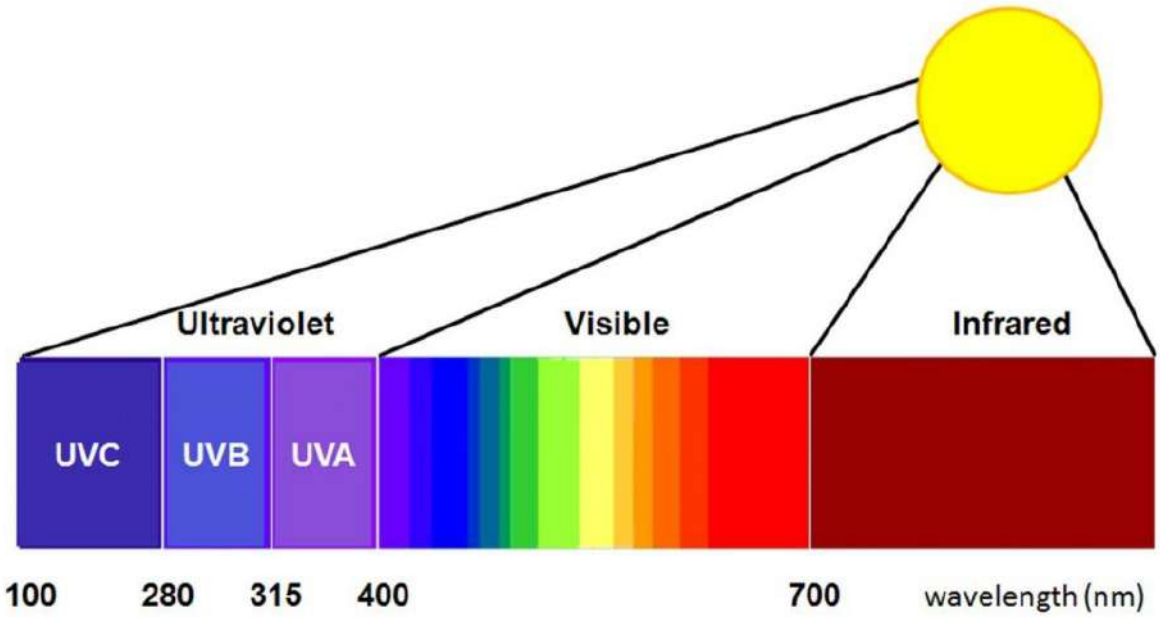

The range of electromagnetic energy emitted by the Sun is known as the solar spectrum and many researchers have found information encoded in the Sunlight.

Sunlight is intelligent, it´s not a huge burning body like NASA wants us to believe. This eludes to the fact that it could have inter-dimensional controllers, or human manipulators via something like chemtrails.

Therefore, the Sun must either have some sort of pranic energy or electrical plasma energy.

Know this too: All solar flare information is presented to us through NASA satellites; it can all be scrubbed. Spots on the Sun exist, but flares have to be ruled out for now.

Ask yourself if you can really know for sure if this is a photo of the Moon or the Sun?

***

But how did the old malfunctioning Sun initially occur? How did the new Sun arrive? What was the catalyst for this event?

It seems that during a cataclysm the sky mechanism changed somehow.

The Moon, as we know it from the perspective of this book, is an ancient Sun that lost its power during the last major cataclysm, and is now malfunctioning. This takes us to the last chapter.

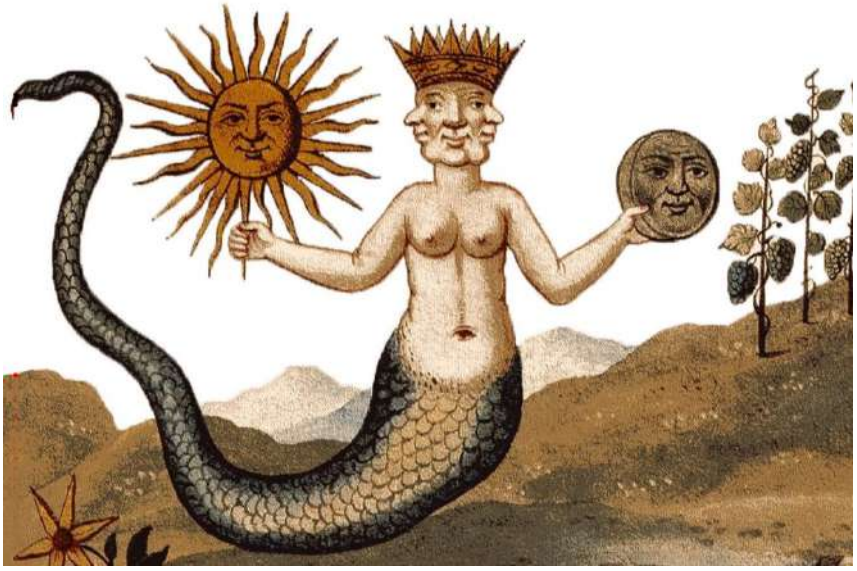

10. The Moon is an Old Malfunctioning Sun Thesis - Advanced

To backup such a bold statement that the Moon is an old malfunctioning Sun, I will provide more information from all around the world. This chapter will only include a small percentage of what I have actually found, for a whole book could be written about this subject alone.

So for a Sun to lose its power and malfunction, and a new Sun to appear, what event could create (or coincide with) this occurrence?

A Cataclysm, and most know the last big one occurred between 10,000-10,000BC, although some say they occur, cyclic, every 5000 years.

I already have a related long article written here, and a one-hour video here. Neither are absolutely necessary before this chapter, but either or both will help to back up claims of the Moon and Sun being centre stage in past cataclysms, from a Flat Earth perspective.

These two offerings elude to the Dark Rift being near us, and a new Sun (and/or destroyer) coming out of the dark rift / milky way, also known as *the Virgin Birth*, *the Cosmic Birth, the Cave, The Gate to the Underworld, the Great rift, the Black Road, the Road to the Underworld, a Crevice in the Branches of the Cosmic Tree, the Mouth of the Cosmic Monster, and the Birth Canal of the Cosmic Mother.*

Cataclysmic Myth Involving the Sun and Moon

There are tales all around the world of the Sun´s light dying out, days of darkness, a new Sun, old Suns (Maya we are in the 5th Sun, Chinese we are in the 10th Sun, Toltec we are in the 3rd), the Sun standing still, the Sun causing destruction by fire, there being no Sun, Moon or stars in the sky, the Sun dimming, the Moon falling, the Sun falling, the stars falling, the Sun rising in the west and setting in the east, the Sun travelling too fast, too slow and the Sun and Moon travelling off their normal paths.

Here is a good example of some of myths and scripture to add some initial weight.

"**The third Sun** (or era) is called Quia-Tonatiuh (Toltec Deity), son of rain, because there fell a rain of fire; all which existed burned; and there fell a rain of gravel. The sandstone, which we now see scattered about 'boiled with great tumult, there also rose the rocks of vermilion color.'
Now, this was in the year Ce Tecpatl, One Flint, it was the day Nahui-Quiahuitl, Fourth rain. Now, in this day, in which men were lost and destroyed in a rain of fire, they were transformed into goslings; **the Sun itself was on fire**, and everything, together with the houses, was consumed."
Donnelly, I., "Ragnarok: The Age of Fire and Gravel" (N.Y.,D. Appleton, 1885)

"There was a tremendous hurricane that carried away trees, mounds, houses, and the largest edifices, notwithstanding which many men and women escaped, principally in caves, and places where the great hurricane could not reach them. A few days having passed, they set out to see what had become of the Earth, when they found it all populated with monkeys. **All this time they were in darkness, without**

seeing the light of the Sun, nor the Moon, that the wind had brought them."
Donnelly, I., "Ragnarok: The Age of Fire and Gravel"

The Algonquin Indians were located in Eastern Canada and the New England States:

"She is the evil minded woman who thus brings diseases upon men, who at the outset introduced pain and death in the world-our common mother, yet the cruel cause of our present woes . . . The ancient Algonquins believed brought death and disease to the race; 'it is she who kills men, otherwise they would never die; she eats their flesh and gnaws their vitals, till they fall away and miserably perish.' Who is this woman? **In the legend of the Muyscas it is Chia, the Moon, who was also goddess of water and flooded the Earth out of spite.**"
Brinton, D. C., "Myths of the New World" (Philadelphia, David McKay, 1896).

The Pawnee Indians of Nebraska are even more descriptive about what happened to the Sun when the cataclysm occurred. In their legend they say that when the Sun goes out, the world will come to an end. The legend:

"The old people told us that the Morning-Star **said that when the time came for the world to end the Moon would turn red; that if the Moon should turn black it would be a sign that some great chief was to die; that when the Moon should turn red the people would then know that the world was coming to an end. the Sun was also to shine bright and all at once that brightness would die out and the end would come....** The old people knew also that when the world was to come to an end there were to be many signs. Among the stars would be many signs. Meteors would fly through the sky. **the Moon would change its color once in a while. the Sun would also show different colors**, but the sign which was to be nearest to the people was that the rivers and the creeks were to rise."
Dorsey, G. A., "The Pawnee Mythology, Pt. 1 "

"The Pawnee legend mentions six of the points we mention in our theory: the flood, the high winds, the fire from the heavens, **the Moon changing color**, the waters sloshing around, and finally the Sun changing colors. We can forgive them for mixing up the order in which some of these occur; but after all, this legend had to have originated about 12,000 years ago, so we could expect some changes from the original legend. What is also interesting is that the Pawnee, like the other Indian tribes, **knew that God is the one who brought this cataclysm down upon man**."
Dorsey, G. A., "The Pawnee Mythology, Pt. 1"

"The Botocudos Indians of Brazil say that the destruction was caused by the Moon falling on the Earth from time to time." Brinton, D. C., "Myths of the New World" (Philadelphia, David McKay, 1896).

"Muyscas Indians of Columbia. At one time they associated the Moon with flooding the Earth out of spite." (same books as above-pl56)

The Menehune people, who migrated from New Zealand to the Hawaiian Islands thousands of years ago, said that one of their gods, Kahana, brought back the Sun after the Sun had vanished and the Earth was dark for a period of time. Now we come to the legends of Kana who restored the Sun. In this legend, the evil chief, Kahiki, is the one who takes away the Sun from the people. The legend goes as follows:

"Niheu treats roughly the messenger of Kahoalei (-Ii'i), ruling chief of Kahiki **and the chief in anger takes away the Sun, Moon, and stars from Hawaii**. Uli sends Kana with Niheu to bring them back. As Kana stretches to the sky to reach the light, Niheu dies of cold and is left behind, but Kana bends over to Kahiki and drops into the spring of two old relatives, who give him **fire** to guide him ahead and wind to bear him behind until he reaches the border of Kahoalei's land. He finds Uli's brother Manu-a guarding the pit (cave) down which the food is kept by the people below and handed up to those above. He puts down a plump black hand which his relatives recognize and fill, first with food, then with water, then with the birds called Kaiwea (fishhawk) which signal the day, then birds and the cock that crows for dawn, **finally stars, Moon, and Sun, all of which he places in the sky**. The chief himself next emerges and returns with Kana to tour the land, restoring Niheu to life on theway. When Ka-hoa-lei reaches Hawaii he finds that Kana and Niheu have both died and he rules there many years."
Beckwith, M., "Hawaiian Mythology" (Honolulu, University of Hawaii Press, 1970).

The philosopher, Wen-Tze, gives us the following description of what happened during the last cataclysm:

"When the sky, hostile to living beings, wishes **to destroy them, it burns them; the Sun and the Moon lose their form and are eclipsed; the five planets leave their paths**; the four seasons encroach one upon another; **daylight is obscured**; glowing mountains collapse; rivers are dried up; it thunders then in winter, hoarfrost falls in summer; the atmosphere is thick and human beings are choked; the state perishes; **the aspect and the order of the sky are altered**; the customs of the age are disturbed (thrown into disorder) all living beings harass one another."
Velikovsky, I., "Worlds in Collision"

"Another Taoist author, Hoei-Nan-Tze, says that **when the Earth and the Sun leave their path, a tremendous flood will cover the Earth**." Velikovsky, I., "Worlds in Collision"

Throughout Chinese recorded history, the emperors of China have insisted that astronomy be studied to its highest attainable level, for in their mythologies, it is recorded that certain catastrophic events occur on the Earth when particular celestial phenomena occur. One of the most important of these phenomena is a Solar Eclipse. A Solar Eclipse was considered a very evil omen. To quote from the lamentations from the Shih King, Ode 9, (Shik Yueh Kih Kiao):

"**The Sun and Moon announce evil, not keeping to their proper paths**. Throughout the kingdom there is no (proper) (order), because the good are not employed. **For the Moon to be eclipsed is but an ordinary matter. Now that the Sun has been eclipsed, how bad it is**.
Grandly flashes the lightning of the thunder. There is a want of rest, a want of good. The streams all bubble up and overflow. The crags on the hill-tops fall down. High banks become valleys; deep valleys

become hills. Alas for the men of this time! How does (God) not stop these things?"
Muller, F. M., (ed.), "The Sacred Books of the East-vol. 3

The Hindu book, *Vedanta-Sutras,* also tells us that the world is periodically devastated and changes both name and form. The surface of the Earth dissolves and is later produced a new. In the book, *Satapatha*Brahmana there is a good description of what the Sun does at the end of a Yuga (world age).

"**That one (the Sun) bakes everything here**, by means of the days and nights, the half-Moons, the months, the seasons, and the year; and this (Agni, the fire) (the Sun) bakes what is baked by that one: 'A baker of the baked (he is),' said BharadVaga of Agni; 'for he bakes what has been baked by that (Sun).' In the year these amounted to ten thousand and eight hundred: he stopped at the ten thousand and eight hundred. "
Muller, F. M., (ed.), "The Sacred Books of the East, vol. 43

"Hot rains came down from the heavens, scalding the flesh of the sinners. The entire time the deluge lasted: "**the Sun and the Moon shed no light**."
Ginzberg, L., "The Legends of the Jews"

."The Teotl Lixco Indians of Mexico which says that the Sun moved toward the East, "traveling toward the East and the disappearance in the East." After the Sun had stopped, they say the Earth was visited by upheavals and Earthquakes."
Velikovsky, I., "Words in Collision"

The Maya and Aztec performed human sacrifices to appease their gods; this was done in order to prevent the Sun from destroying them.

Quiches Maya of neighboring Guatemala have a legend that states that the Sun eventually did appear, but it was different from the Sun they had seen prior to the destruction:

"And the Sun, and the Moon, and the stars were now all established. Yet was not the Sun then in the beginning the same as now; his heat wanted force, and he was but as a reflection in a mirror; verily, say the histories, not at all the same Sun as that of to-day. Nevertheless he dried up and warmed the surface of the Earth, and answered many good ends."
Bancroft, H. H., "The Native Races, vol. 3-Myths and Languages"

"Now, **there had been no Sun in existence for many years**; so the gods, being assembled in a place called Teotihuacan, six leagues from Mexico, and gathered at the time round a great fire, told their devotees that he of them who should first cast himself into that fire should have the honor of being transformed into a Sun. So one of them, called Nanahuatzin . . . flung himself into the fire. Then the gods began to peer through the gloom in all directions for the expected light, and to make bets as to what part of heaven he should first appear in. And some said . Here, and some said There; but when the Sun rose they were all proved wrong, for not one of them had fixed upon the east. And in that same hour, though they knew it not, the decree went forth that they should all die by sacrifice.

The Sun had risen indeed, and with a glory of the cruel fire about him that not even the eyes of the gods could endure; but he moved not. There he lay on the horizon; and when the deities sent Tlotli, their messenger, to him, with orders that he should go on upon his way, his ominous answer was, that he would never leave that place till he had destroyed and put an end to them all. Then a great fear fell upon some, while others were moved only to anger; and among the latter was one Citli, who immediately strung his bow and advanced against the glittering enemy. By quickly lowering his head the Sun avoided the first arrow shot at him; but the second and third had attained his body in quick succession, when, filled with fury, he seized the last and launched it back upon his assailant. And the brave Citli laid shaft to string nevermore, for the arrow of the Sun pierced his forehead.
Then all was dismay in the assembly of the gods, **and despair filled their hearts for they saw that they could not prevail against the shining one; and they agreed to die**, and to cut themselves open through the breast. Xolotl was appointed minister, and he killed his companions one by one, and last of all he slew himself also.....Immediately on the death of the gods the Sun began his motion in the heavens; and a man called Tecuzistecatl, or Tezcociztecatl, who, when Nanahuatzin leaped into the fire, **had retired into a cave, now emerged from his concealment as the Moon**."
Bancroft, H. H., "The Native Races, vol. 3-Myths and Languages"

"Now began the struggle between the two brothers, Tezcatlipoca and Quetzalcoatl (the Sun), which was destined to destroy time after time the world, with all its inhabitants, and to plunge even the heavenly luminaries into a common ruin. **The half Sun created by Quetzalcoatl lighted the world but poorly**, and the four gods came together to consult about adding another half to it. Not waiting for their decision, **Tezcatlipoca transformed himself into a Sun, whereupon the other gods filled the world with great giants, who could tear up trees with their hands**. When an epoch of thirteen times fifty-two years had passed, Quetzalcoatl seized a great stick, a**nd with a blow of it knocked Tezcatlipoca from the sky into the waters, and himself became Sun**. For an epoch the Earth flourished under Quetzalcoatl as Sun, but Tezeatlipoca was merely biding his time, and the epoch ended, **he appeared as a tiger and gave Quetzalcoatl such a blow with his paw that it hurled him from the skies (the Sun disappears)**. The overthrown god revenged himself by sweeping the Earth with so violent a tornado that it destroyed all the inhabitants but a few, and these were changed into monkeys. **His victorious brother then placed in the heavens, as Sun, Tlaloc, the god of darkness, water and rains, but after half an epoch, Quetzalcoatl poured a flood of fire upon the Earth, drove Tlaloc from the sky, and placed in his stead, as Sun, the goddess Chalchiutlicue**, the Emerald Skirted, wife of Tlaloc. In her time the rains poured so upon the Earth that all human beings were drowned or changed into fishes, and at last the heavens themselves fell, and Sun and stars were alike quenched." (6-p74)

"It would be when Tezcatlipoca should **steal the Sun** from heaven for himself; in other words, when eternal night should close in upon the Universe."
Brinton, D.C., "American Hero-Myths" (Philadelphia, H. C. Watts & Co., 1882)

From a prayer to Tezcatlipoca, the Flying Serpent God:

"....is it already fixed in thy divine counsel that there is to be no mercy nor pity for us, until the arrows of thy fury are spent to our utter perdition and destruction? Is it possible that this lash and chastisement is not given for our correction and amendment, but only for our total destruction and obliteration; that **the Sun shall nevermore shine upon us, but that we must remain in perpetual darkness and silence**; that nevermore thou wilt look upon us with eyes of mercy, neither little nor much?"

Bancroft, H. H., "The Native Races, vol. 3-Myths and Languages"

In the Popol Vuh (16th century), the Maya Hero Twins are finally transformed into Sun and Moon.

Ix Chel, Maya goddess of the Moon, was sometimes depicted as the goddess of catastrophe (the goddess who stands by as the world floods). Many of her myths show her in a more benevolent light—as a goddess who refused to become a victim of oppression.

She was also depicted as an old woman wearing a skirt with crossed bones, and she had a serpent in her hand. She had an assistant sky serpent, whom they believed carried all of the waters of the heavens in its belly. She is often shown carrying a great jug filled with water, which she overturns to send floods and powerful rainstorms to Earth.

Aztec Coyolxauhqui was the Moon goddess according the Aztec mythology. Her name means "Golden Bells." She was the daughter of the Earth goddess, Coatlicue and the sister of the Sun god, Huitzilopochtli.

Coyolxauhqui encouraged her four hundred sisters and brothers to kill their dishonored mother. Coatlicue gave birth to Huitzilopochtli (Aztec Sun and war god) after a ball of feathers fell into the temple where she was sweeping (rift?) and touched her. Huitzilopochtli sprang out of his mother as an adult fully armed and saver her.

Coatlicue regretted such violence. Thus, Huitzilopochtli cut off Coyolxauhqui's head and threw it into the sky to form the Moon.

According to Maya mythology, before becoming the celestial objects that we see today, the Moon and the Sun were terrestrial creatures. The Moon was a young lady, while the Sun a brave hunter.
They fell in love. After some events, they fled away together.
The young lady's grandfather got mad at her and he had her killed. The dragonflies collected her body and her blood, and they put them in thirteen hollow stumps.
Meanwhile, the Sun had been looking for his lover for thirteen days. On the thirteenth day, he found the stumps. Twelve of them gave life to harmful insects and snakes. They started filling in the entire world with their offspring. From the thirteenth one, the Moon came out: she had come back to life.

***

The Ute Indians of California and Utah have a very detailed legend of the last destruction. They specifically mention the Sun as being the cause of the fire.

A Ute philosopher declares the Sun to be a living personage, and explains his passage across the heavens along an appointed way by giving an account of a fierce personal conflict between Ta-vi, the Sungod, and Tawats, one of the supreme gods of his mythology:

"In that long ago, the time to which all mythology refers, **the Sun roamed the Earth at will. When he came too near with his fierce heat the people were scorched, and when he hid away in his cave for a long time, too idle to come forth, the night was long and the Earth cold**. Once upon a time Ta-wats, the hare-god, was sitting with his family by the camp-fire in the solemn woods, anxiously waiting for the return of Ta-vi, the wayward Sun-god. Wearied with long watching, the hare-god fell asleep, **and the Sun-god came so near that he scorched the naked shoulder of Ta-wats**. Foreseeing the vengeance which would be thus provoked, he fled back to his cave beneath the Earth. Ta-wats awoke in great anger, and speedily determined to go and fight the Sun-god. After a long journey of many adventures the hare-god came to the brink of the Earth, and there watched long and patiently, till at last the Sun-god coming out he shot an arrow at his face, but the fierce heat consumed the arrow and it had finished its intended course; then another arrow was sped, but that also was consumed; and another, and still another, till only one remained in his quiver, but this was the magical arrow that had never failed its mark. Ta-wats, holding it in his hand, lifted the barb to his eye and baptized it in a divine tear; then the arrow was sped and struck the Sun-god full in the face, **and the Sun was shivered into a thousand fragments, which fell to the Earth,** causing a general conflagration. Then Ta-wats, the hare-god, fled before the destruction he had wrought, and as he fled the burning Earth consumed his feet, consumed his legs, consumed his body, consumed his hands and his arms-all were consumed but the head alone, which bowled across valleys and over mountains, **fleeing destruction from the burning Earth, until at last, swollen with heat, the eyes of the god burst and the tears gushed forth in a flood which spread over the Earth and extinguished the fire**. The Sun-god was now conquered, and he appeared before a council of the gods to await sentence. In that long council were established the days and the nights, the seasons and the years, with the length thereof, **and the Sun was condemned to travel across the firmament by the same trail day after day till the end of time."**

Donnelly, I., "Ragnarok: The Age of Fire and Gravel" (N.Y.,D. Appleton, 1885).

The Pawnee Indians of Nebraska are even more descriptive about what happened to the Sun when the cataclysm occurred. In their legend they say that when the Sun goes out, the world will come to an end. The legend:

".....that when **the Moon should turn red the people would then know that the world was coming to an end. the Sun was also to shine bright and all at once that brightness would die out and the end would come**.... The old people knew also that when the world was to come to an end there were to be many signs. Among the stars would be many signs. **Meteors would fly through the sky. the Moon would change its color once in a while. the Sun would also show different colors**, but the sign which was to be nearest to the people was that the rivers and the creeks were to rise."

Dorsey, G. A., "The Pawnee Mythology, Pt. 1 " 1906

The Ojibway Indians of the Great Lakes also mention in their legend that the Sun caused great havoc on the Earth. The legend is centered around a boy whose bird-skin coat is burned by the Sun.

"He swore that he would have vengence. He persuaded his sister to make him a noose of her own hair**. He fixed it just where the Sun would strike the land as it rose above the Earth's disk**; **and, sure enough, he caught the Sun, and held it fast, so that it did not rise**. The animals who ruled the Earth were immediately put into greatcommotion (Earthquakes). They had no light. They called a council to debate upon the matter, and to appoint some one to go and cut the cord, for this was a very hazardous enterprise, **as the rays of the Sun would burn up whoever came so near**. At last the dormouse undertook it for at this time the dormouse was the largest animal in the world (the mastodon?); when it stood up it looked like a mountain. **When it got to the place where the Sun was snared, its back began to smoke and burn with the intensity of the heat, and the top of its carcass was reduced to enormous heaps of ashes. It succeeded, however, in cutting the cord with its teeth and freeing the Sun, but it was reduced to very small size, and has remained so ever since**."
Donnelly, I., "Ragnarok: The Age of Fire and Gravel" 1885

In the Southeastern United States, we come to the Creek and Chewkee tribes. The Creeks have a creation legend similar to the one described in the Bible. They also say the Earth was entirely covered by water.

"After a long period of time, the Earth became populated with many; but something happened and "the Earth became angry and ate up a portion of her progeny; how the people started out on a journey toward the Sunrise; how they crossed a River of Slime, then a River of Blood, and came to the King of Mountains, whence a great fire blazed upward with a singing sound."

This legend seems to allude to the red clay that, that when mixed with the water, appeared to be like blood flowing in the streams. The Chewkee tribe, also on the Gulf of Mexico, say that the Sun was too close and too hot to the Earth, and it burned up many things.

The Mbocobi Indians of Paraguay are very specific in pointing out what caused the destruction.

"The destruction of the world was **due to the Sun**. **This orb once fell from the sky**, but a Mbocobi hastened to pick it up before it did any injury, and fastened it in its place with pegs. A second time it fell and burnt up the Earth. Two of the tribe, a man and his wife, climbed a tree and escaped destruction, but a flash of flame reached them and they fell to the ground, where they were changed into monkeys."
Brinton, D. C., "Myths of the New World" 1896

The Ainu culture in Japan does not have what we would call a very clear self-explanatory mythology of the last cataclysm. Instead the sequence of events that occurred is coded and concealed in some very fanciful mythologies. The main one being the legend of Izanagi. Izanagi we would consider their idea of God. The legend tells us his sister, Izanami (the old Sun), went to the underworld when she grew old and was about to die. The legend goes on:

"After Izanagi had slain his son, the fire-god, and brought into being new gods (a new Sun), including dragons, he was seized with longing to see Izanami once more. Accordingly he set out to find her in Yomi, the dark Hades of the Underworld ...
When Izanagi reached the gloomy dwelling of his sister, she raised the door, and he spoke to her, saying: 'Thine Augustness, my lovely young sister! The lands that I and you made are not yet finished; so come back.' " (20-p357)

This part of the legend is attempting to tell us that when the old Sun disappeared, surely into the rift (door, dwelling), there was a period of darkness that covered the Earth. Later on, the legend tells us that associated with the old Sun is fire, thunder, lightning, and Earthquakes. Eventually a new Sun reappears and the Earth is brought to life again.

In other writings, Markandeya goes into greater detail of what happens during the destruction:

"After a drought lasting for many years, seven blazing Suns will appear in the firmament; (the Sun novas) they will drink up all the waters. Then wind-driven fire will sweep over the Earth, consuming all things; penetrating to the nether world it will destroy what is there in a moment; it will burn up the Universe. **Afterwards many-coloured and brilliant clouds will collect in the sky**, looking like herds of elephants decked with wreaths of lightning. Suddenly they will burst aSunder, and rain will fall incessantly for twelve years until the whole world with its mountains and forests is covered with water. The clouds will vanish. Then the Self-created Lord, the First Cause of everything, will absorb the winds and go to sleep. The Universe will become one dread expanse of water."
Mackenzie, D. A., "Indian Myth and Legend" 1913

From Babylon myth:

"When that millennium has fully elapsed, which is the second of the religion of the Mazda-worshippers.... The **Sun conceals itself** . . .Then there will be 30 straight winters. Fifty-seven years after the Sun goes dark, a new Sun will appear in the sky."
Muller, F. M., (ed.), "The Sacred Books of the East, vol. 47, 1897

The idea of previous worlds and civilizations was well known to the ancient Jews. They believed that the previous civilizations had all been destroyed by cataclysms brought down upon them by God. They believed that some people perished by the deluge and that others were consumed by fire. It was believed that God summoned...

"....the Angel of the Face, (the Sun) ... to destroy the world. The angel opened his eyes wide, and scorching fires and thick clouds rolled forth from them." Ginzberg, L., "The Legends of the Jews" 1913

Isaiah 60:19 "**The Sun will no more be your light by day, nor will the brightness of the Moon shine on you**, for the LORD will be your everlasting light, and your God will be your glory."

The Nordics also have a legend describing the appearance of the new Sun which is called Dag.

In the beginning of the battle, Fenris and Iormungandr belched forth fire and smoke of "noxious, deathly vapours, which filled all heaven and Earth with their poisonous breath (the red dust from the Sun)." The roar of battle filled the universe. The first benevolent god to perish was Odin, the Sun. Eventually the other benevolent gods were also destroyed.

"The other gods who took part in the fray, and all the Einheriar having now perished, Surtr suddenly flung his fiery brands over heaven, Earth, and the nine kingdoms of Hel. The raging flames enveloped the massive stem of the world ash Yggdrasil, and reached the golden palaces of the gods, which were utterly consumed. The vegetation upon Earth was likewise destroyed, and the fervent heat made all the waters seethe and boil.

The great conflagration raged fiercely until everything was consumed, when the Earth, blackened and scarred, slowly sank beneath the boiling waves of the sea. Ragnarok (day of atonement) had indeed come; the world tragedy was over, the divine actors were slain, and chaos seemed to have resumed its former sway."

Guerber, H. A., "Myths of the Norsemen" 1908

The only survivors from this godly race were Lifthrasir and his wife, Lif. After the cataclysm had passed, they emerged from their **cave** to repopulate the Earth.

In Norse mythology, we can find a similar story to Rahu and Ketu, but with different names. Sköll is a warg (wolf) that chases the horses Árvakr and Alsviðr, that drag the chariot which contains the Sun (Sól) through the sky every day, trying to eat her. Sköll has a brother, Hati, who chases Máni, the Moon. At Ragnarök, both Sköll and Hati will succeed in their quests. (See Gylfaginning, chapters 38 and 51).

Note the above words, "At Ragnarök." In Norse mythology, Ragnarök is a series of future events which include a great battle. Therefore this event doesn't occur at every Sun and Moon eclipse, but just at Ragnarök.

Also in chapter 38, Sturluson says that there are many men in Valhalla, and many more who will arrive, yet they will "seem too few when the wolf comes." High foretells that as part of the events of Ragnarök, after Fenrir's son Sköll has swallowed the Sun and his other son Hati Hróðvitnisson has **swallowed the Moon, the stars will disappear from the sky**. The Earth will shake violently, trees will be uprooted, mountains will fall, and all binds will snap – Fenrisúlfr will be free. Fenrisúlfr will go forth with his mouth opened wide, his upper jaw touching the sky and his lower jaw the Earth, and flames will burn from his eyes and nostrils"

"Fenrir" is just another name for the destroyer, it is hidden in the Nordic legends behind many names, and the legend tells us that his children will swallow the Sun and Moon.

An Egyptian theory recorded by Herodotus said:

"That during eleven thousand three hundred and forty years of Egyptian history the Sun of four occasions altered his course, twice rising where he now sets and twice setting where he now rises."
Stuart, **J.** A., "The Myths of Plato" 1960

Also from Egyptian and Celtic scholars in The Kolbrin, we find:

"The Heavens will burn brightly and redly; there will be a copper hue over the face of the land, followed by **a day of darkness. A New Moon will appear and break up and fall**."
"The waters were purified, the sediment sank and life increased in abundance **... The Sun was not as it had been and a Moon had been taken away** ..."

"In a myth of the Luyia people of Kenya in East Africa, the Sun and Moon were brothers. **The Moon was older, bigger, and brighter**, and the jealous Sun picked a fight with him. **The two wrestled and the Moon fell into mud**, which dimmed his brightness. God finally made them stop fighting and kept them apart by ordering the Sun to shine by day and the mud-spattered Moon to shine by night to illuminate the world of witches and thieves."
hwww.mythencyclopedia.com/Mi-Ni/Moon.html

Rahu and Ketu Revisited

What recent mythology researchers know about Rahu and Ketu is false. They all try to put their own idea on an already *semi-fake* story. They mix the myth with the fake NASA ball Earth and ball planets cosmology, so the result is messy.

As we know, Rahu and Ketu are the head and tail of the serpent demon, and this is often referred to as the the Ouroboros serpent.

The Ouroboros (Greek for Tail Devourer) is found all around the world and on one of the many levels, symbolizes the cyclic nature of the universe: creation out of destruction, life out of death. The Ouroboros eats its own tail to sustain its life, in an eternal cycle of renewal. It is also related to Flat Earth; it grew so large that it could encircle the world and grasp its tail in its teeth. It is also symbolized as covering the Moon and Sun in a figure of eight, with the Head of a Lion and Tail of an Eagle.

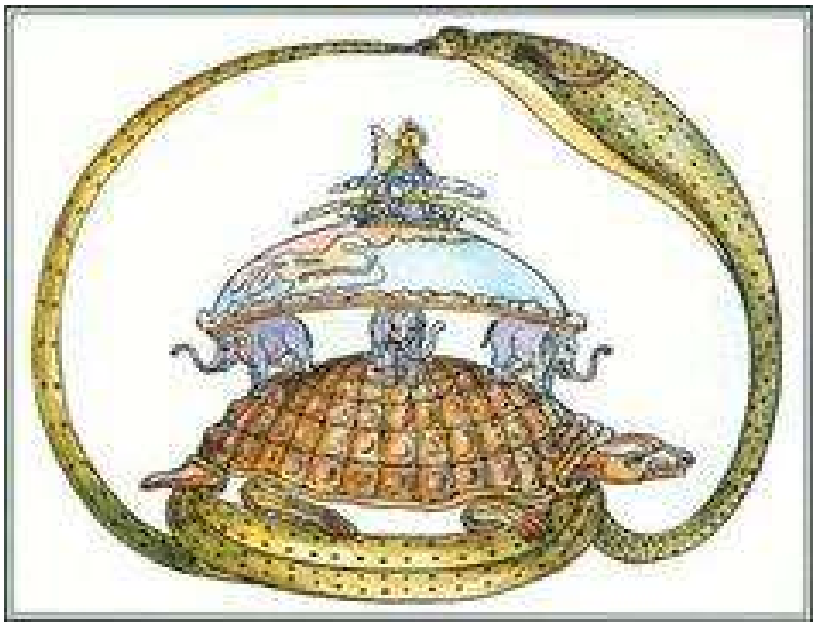

Where else do we see this lion and bird symbolism? Drumroll…..

The Ouroborus is a symbol of alchemical dissolution. He dissolves silver & gold which are symbolized by the Moon & Sun. The serpent is known all around the world as the milky-way, which we know as the dark rift. The dark rift as I´ve shown in other works (link1, link2) has many myths of a serpent coming out and devouring the Earth and sky.

Rahu is considered the serpent´s (or dragon's) head and Ketu is known as the serpent's tail.

Rahu and Ketu are originally symbols, deities, and archetypes of cataclysm – that devour the Sun and Moon.

Here is the myth most know:

"Just as Vishnu poured a drop of nectar into Svarbhanu's mouth, the Sun and Moon recognized that he was actually an asura! Vishnu chopped him in half, but it was too late. A drop of the immortal nectar was already in his mouth! Thus the two halves, Rahu and Ketu were born. The head of the asura who ingested the immortal nectar became Rahu and the lower half, the body of the asura, became Ketu. Rahu and Ketu are the head and tail of the demon, often referred to as the serpent. A Lunar Eclipse is connected to Rahu- this happens when he tries to eat the Moon, but it falls right out of him. Ketu is connected to the Solar eclipse. "

"It is believed that this immortal head from time to time swallows the Sun, causing eclipses. Then, the Sun passes through the opening at the neck, ending the eclipse. The body also turned into Ketu due to a boon, and it in turn swallows the Moon on timely basis to cause a Lunar Eclipse."

What the mythologists forget to mention is that this story was referring to the period of cataclysm, and not to the eclipses.

It´s the same story as in the Maya mythology and also Egyptian mythology, please read below:

"The Skyworld is a stage on which the players and events of cosmic ordering are made visible, while the Underworld stage leaves them hidden from view. In the Skyworld the Milky Way is on occasion seen as a serpent that swallows the souls of the recently dead. Also in the Skyworld the two-headed Celestial Monster of Classic Maya iconography could represent an "ecliptic snake" which takes the rising Sun into its eastern mouth, to discharge it through its western mouth at Sunset--the Sun traveling through its body in its path across the sky, in much the same way that the Sun was thought to travel through Nut's body in ancient Egypt."
http://pperov.angelfire.com/galactic.html#apocatastasis

This idea that Rahu and Ketu are planets came from people who tried to interpret the Vedic astronomy based on the ball Earth and ball planets belief.

SideNote; It seems that many of the Vedic scriptures were deliberately mistranslated, much has been distorted. One example is the that they also distorted periods of time, so instead of a few thousand years´ cycles, they exaggerated it so much and wrote stupid things like 1.000.000 years, Yuga x, Yuga y. Only the so called *scholars* knew how to translate the ancient Vedic writings, and they did a *great job* and for me, they intentionally distorted many ancient records. But rant over, let us press on.

But ancient Hindu records tell us other things about Rahu and Ketu.

Atharvaveda (13.16-24) employs Ketu to mean ray of light. These nine verses are taken from Rgveda (1.50.1-9) in the same order and more or less in the same form. They are also found "in one or more other Vedic texts" (Whitney 1905, Vol.2, p.722). More typically Ketu meant *combination of fire and smoke*. The Atharvaveda passage (19.9.10) refers to Dhumaketu as an epithet of mrtyu [death]. It either means a comet or literally as "smoke-bannered" to the smoke rising from a funeral pyre (Whitney 1905,Vol. 2, p. 914). Atharvaveda (11.10.1-2, 7) uses Ketu in the plural, asarunah ketavah [ruddy Ketus]. Here the reference seems to be to comets or meteors. Varahamihira's Brihatsamhita, composed in 6th century CE but containing much older material, quotes a still earlier astronomer Garga on a class of 77 comets, called Aruna, which are dark red in colour (Bhat 1981,Vol. 1, p.138).

The revered Ved Vyasa tells Dhrtarastra (Bhismaparva 3.46) as follows."Cows are giving birth to asses; and elephants to dogs. Sons are enjoying sexual pleasures with their mothers. Idols of gods are laughing, vomiting blood, feeling sad, and falling off their pedestals on their own. Animals are being born with three horns, four eyes, five feet, two urinary organs, and two tails. Women are giving simultaneous birth to four –five girls, who immediately start singing, dancing and laughing. Trees are floweringout of season. Lotus and water-lily are blossoming on tree tops. Even koel, peacock andparrot are making fearsome sounds. There is a downpour of blood and bones from the sky."

It is in this background that even the more-reasonable sounding descriptions of celestial phenomenon should be seen. A recurring theme is the reference at various places in the Mahabharata to Rahu, as if the occurrence of an eclipse was at par with holocaust on Earth. "Rahu has seized the Sun." (Bhismaparva 3.11).

"Rahu is approaching the Sun." (Bhismaparva 141.10).

Rahu swallowed the Sun most untimely." (Salyaparva 55.10).

"Rahu eclipsed the Sun and the Moon simultaneously." (Asvamedhaparva 76. 15, 16, 18).

Meteors (ulka) and Earthquakes are also similarly invoked. As part of the celestial foreboding it is stated that a very dangerous Dhumaketu has overcome the naksatra Pusya. This will bring destruction to both sides. (This ill-omen appears in the 4th century CE Buddhist text Sardulakarnavadana as well.

Continuing his listing of ill omens, Ved Vyasa tells Dhrtarastra that the sveta graha (white planet) has transgressed Citra, while the parusa graha (harsh planet) has established itself between Citra and Svati (Bhismaparva 3.11, 16). The translators have exercised their own discretion in rendering these terms. Sveta graha has been left untranslated (Sathe et al. 1985, p.39) or equated with Ketu (Ganguli 1884-1896, Book 6,p.12). Parusa graha has been identified with Rahu by one translator (Ganguli 1884-1896, Book 6, p.12) and with Ketu by a different translator (Sathe et al. 1985, p.39).

The arbitrariness is obvious. As we have argued it would be anachronistic to associate Rahu and Ketu with a planet in pre-Varahamihira times.

Greek astronomical elements made their documented appearance in India in 149 CE when a Greek astro-text was translated into Sanskrit by Yavanesvara. It was versified in 269CE by Sphujidhvaja under the title Yavanajataka (Pingree, p. 1959). The versification was a significant development, because it signifies assimilation of Greco-Babylonianelements into Indian tradition. And yet, Vedic astronomical tradition remained extant even after the introduction of Yavana texts, as can be seen from passages in Sardulakarnavadana, already referred to. "Irrespective of the naksatra, when the Sun or the Moon is seized by Rahu, the king along with his subjects comes to pain."Irrespective of the naksatra when Ketu enters the Moon, the neighboring enemy kinggets the upper hand." "When Dhumaketu establishes itself in the Pusya naksatra, then defeat in enemy's assault from all four directions is guaranteed" (Vaidya 1999, p. 374,couplets 462,463, 466). As we have already noted, Dhumaketu in Pusya as a bad omen is mentioned in the Mahabharata also. It is significant that Ketu and Dhumaketu are listed separately and along with Rahu under utpata.

In later Iranian (and Arabic) mythology, the ascending Node Rahu and the descending Node Ketu become the head and the tail of the dragon Al –Djawzahr, a serpent linked to cataclysm.

Chapter Conclusions

Not many people are looking into ancient cataclysmic myth with the Flat Earth viewpoint, or with even knowledge of chemtrails, so we are on new ground. But we can all see that an ancient story of our sky that exists with different archetypes *all around the world* just *has* to be based upon some larger truths.

My theory is that there´s a major cataclysm every so often that really shakes up all that is within the firmament (sky ceiling, sky dome). A new Sun or a Destroyer of some kind comes through the dark rift and wreaks havoc. The Sun dies in the carnage and becomes the Moon.

The celestial bodies *are* obstructed during cataclysm, some lose their power, and some of them fall. Also this obstruction of the Sun and Moon is probably made by the thick and black smoke that will cover the sky just as the cataclysm begins, possibly from the dying Sun or dark rift.

This is why the ancients feared the eclipses, they saw them as the cataclysm coming again, and they were scared because in ancient times there was a period without the Moon.

So then, this thesis presents that during ancient cataclysm, the sky mechanism was changed, and any sign of disturbance since, like an eclipse, and the people have become frightened that the cataclysm will start again.

What I would say currently is that climate change and chemtrails may be related to something coming soon, and for sure *the powers that be* would know this and be acting accordingly. It isn't something that will be put on Fox news that´s for sure.

I´m not trying to scare anyone, I´m just piecing together information.

Remember that everything you were taught is wrong….by design.

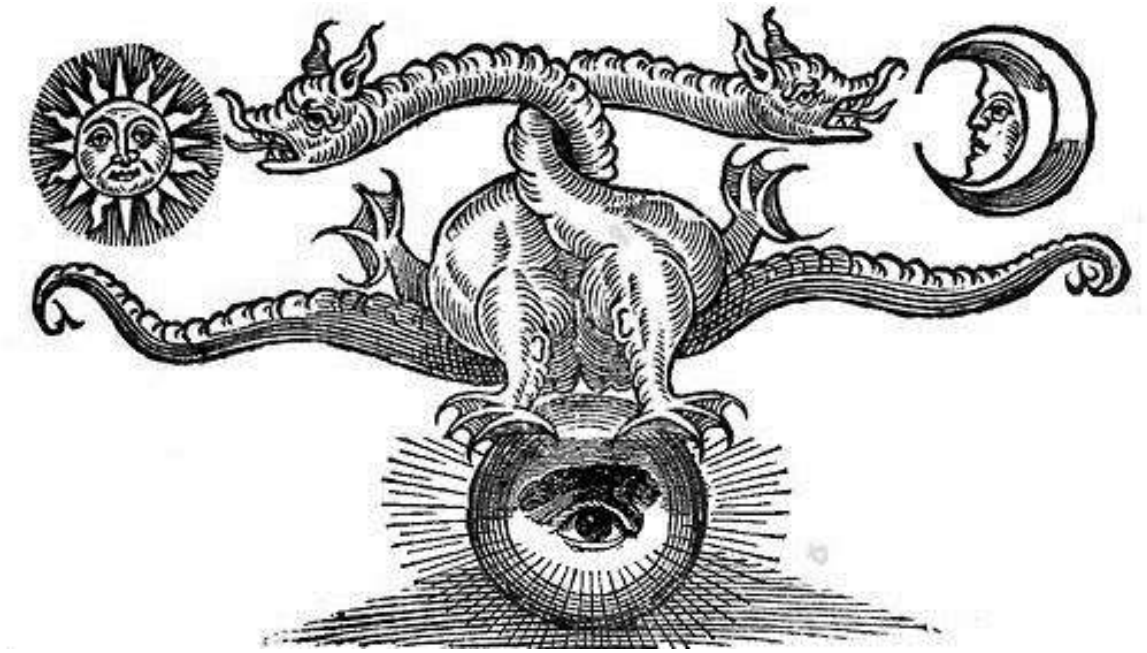

Afterword

The Moon is a mystical enigma, and cannot be solved with rigid science alone. Her secrets are there to be found, but we have to stroke her from many different angles.

We need to be resolute in our research, because there are government agencies employed to lie and send people the wrong way.

The exact science and mechanism behind the Sun is probably beyond our understanding, or it's that we just don't have access to better information sources.

The exact way the Sun and Moon function would reveal also a big part of the way the enclosed Earth mechanism functions, and that would be a huge find.

Not one person has the absolute right to claim their *complete* Flat Earth theory is *the* correct one and that all else is confusion. But by working together, individually studying, debating, being open, and creating tangible pieces of work, we can figure it out together. I prefer not to spend my time guarding and ring-fencing my work against the opinions of others. I prefer to collaborate and progress with more research, and then throw it into the pot. I am happy to be corrected - we were all wrong with the ball model, and this is a new paradigm with new information in its infancy.

This work is offered freely, but one is more than welcome to give support to my research and site via www.waykiwayki.com, where one can find hundreds of articles, videos, podcasts, and other literal offerings, spanning nearly a decade.

Come and join the facebook group: *Flat Earth Advanced – the Moon.*

You are welcome to email me at waykiwayki@hushmail.com, however I cannot promise a prompt reply, due to the increase of incoming traffic recently.

Much credit goes to other researchers, and also to those who have created the images I've included. I hope it´s ok that I´ve not yet sourced all of the creators, researchers, and artists. Special mention to Ashlee Ellegard Webster for her flattening of globe images. Super special credit goes to the man in the shadows who I think has read every book, pantheon, and myth ever known.

This work can be copied freely as I´m not a fan of copyright, which I see as old an old paradigm concept.

With love for truth,

Mx

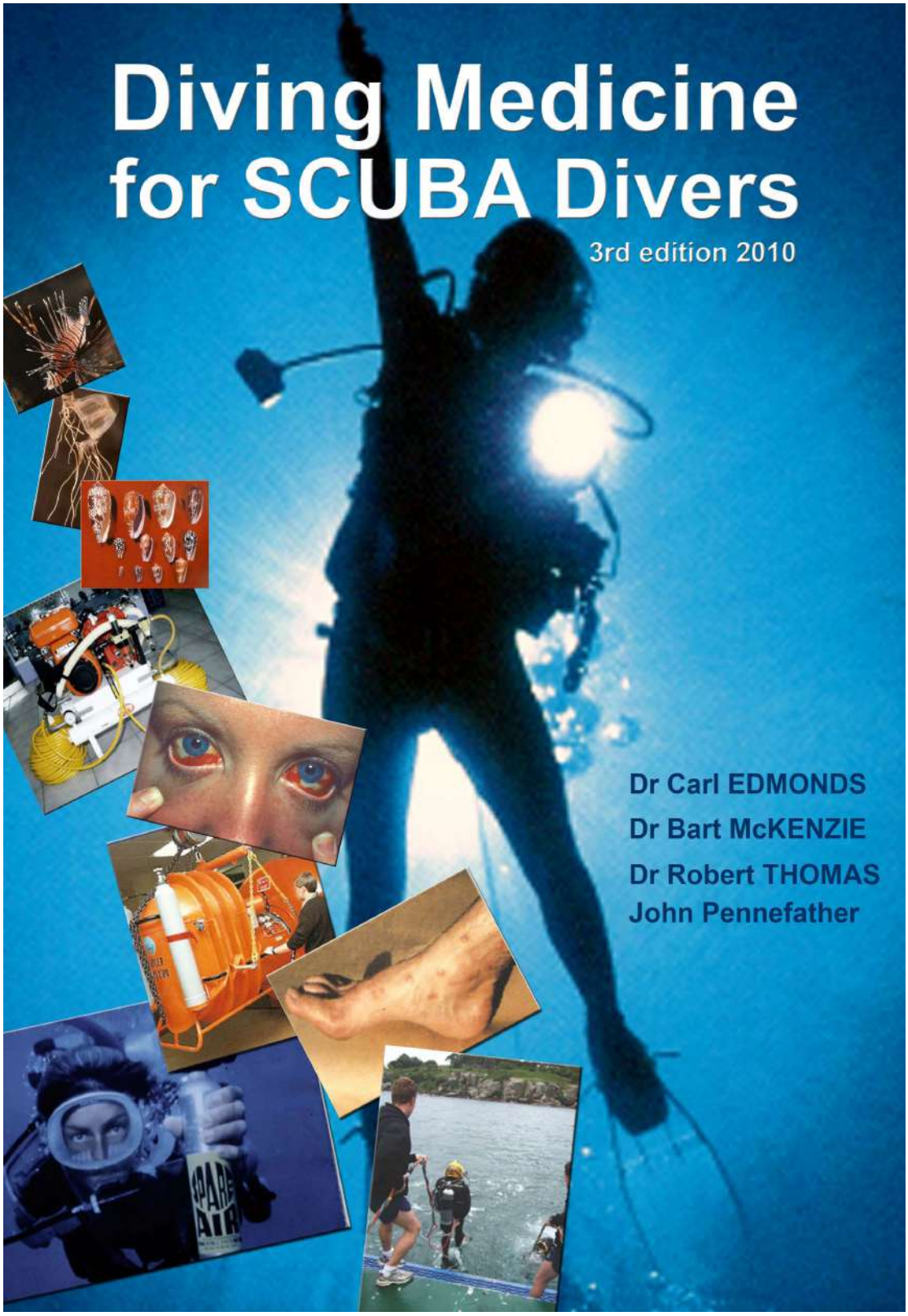

Diving Medicine for Scuba Divers 3rd Edition 2010

Published by Carl Edmonds
Ocean Royale, 11/69-74 North Steyne
Manly, NSW, 2095 Australia
puddle@bigpond.net.au

First edition, October 1992
Second edition, April 1997

National Library of Australia Catalogue

1. Submarine Medicine
2. Scuba Diving Injuries
3. Diving – physiological aspects

Copyright: Carl Edmonds

Title 1 of 1 - Diving Medicine for Scuba Divers
ISBN: [978-0-646-52726-0]

To download a free copy of this text, go to
www.divingmedicine.info

FOREWARD

"Diving and Subaquatic Medicine", in its third and highly respected edition, has provided a cornerstone of knowledge for the diving medical professional. Now, "Diving Medicine — for Scuba Divers" is a condensed, simplified and lighter publication for the general diving population. The authors — Drs Edmonds, McKenzie and Thomas, have done an excellent job of providing a comprehensive, useful and up to date resource base for the diver in the field.

The presentation of the material reflects the fact that the authors are experienced divers as well as specialists in diving medicine. Their thinly disguised sense of humour is reflected throughout the text in emphasising important issues and occasionally just lightening the academic loading on the reader. Their treatment of areas of controversy reflects their experience and background in treating diving emergencies. Some readers will find information which may be inconsistent with their teachings. It is strongly suggested that the reader pay attention to the advice that is presented by the authors. Their many years of cumulative experience is reflected in their advice.

This text represents the broadest coverage of diving medical issues that has been focussed upon the general diving public. It does an excellent job of bridging the gap between the "dive rescue" type materials put out by the various agencies, and the other excellent medical texts which focus primarily on physician education. This text should serve the reader well since its intent is clearly designed to stimulate the diving population at large to become better informed on current diving medical issues. Additional resource materials are presented for those who wish more detailed information.

Glen H. Egstrom Ph.D.
Professor Emeritus
Department of Physiological Sciences,
University of California, Los Angeles.

Addendum:

The success of this text, together with its Japanese and Korean editions, has inspired me to revise and update the material, even though the person I originally designed it for, my young son Mark, is now well into middle age.

The 2nd edition is well over a decade old, and it had run out of print. With the assistance of my esteemed co-worker, John Pennefather, and life long friend, Clarrie Lawler, the 3rd and final edition is now finished.

My intent is that it be made available freely (at least on Internet) to all divers, conscientious dive instructors, paramedics and doctors new to this fascinating and rewarding subject.

Carl Edmonds, 2010

ACKNOWLEDGMENTS

For the 1st and 2nd editions of this text

To my diving buddy and wife, Cindy, for her invaluable criticisms, suggestions and proof-reading. To my son, Mark, who inspired the predecessor of this text, "A Diving Manual for Mark", which I prepared for him. The motivation was to ensure that he had more factual information than that available from his open water scuba diving course. He is now a competent and perceptive diver, partly from his understanding of this text. Thanks also to Patracia Larke and Enid Page for their secretarial services.

Carl Edmonds.

I would like to thank my wife, Ann, for tolerating my on and off affair with a word processor, and also my daughter Deborah, Des Lund, Tammy Lye, Shirley Warner, and Joanne Wright who helped me in the gestation of this book.

Bart McKenzie

This author wishes to gratefully acknowledge the assistance given by my daughter, Natalia, to producing some of the diagrams used within the book. I also wish to thank my wife, Denise, for her tireless support and patience regarding my often frustrating use of the Macintosh computer and noisy printer in completing this text — something she never believed would really ever eventuate!

Bob Thomas

For the 3rd edition,

John Pennefather, who assisted Carl Edmonds in reviewing and updating the text

Clarrie Lawler, who supplied many excellent diagrams

Richard Chesher, for spectacular photographs

John Lippmann, CEO, DAN Asia-Pacific, for advice, corrections and preparation of both earlier and this edition

Richard Pyle for technical diving diagrams Cis-Lunar Development Laboratorie

DIVING MEDICINE FOR SCUBA DIVERS

BACKGROUND

1. History of Diving
2. Physics
3. Physiology
4. Breath-hold Diving
5. Equipment
6. Diving Environments
7. Stress Disorders — Panic and Fatigue
8. Female Divers

SPECIFIC DIVING DISEASES (PRESSURE RELATED DISEASES)

9. Ear Barotrauma
10. Sinus Barotrauma
11. Pulmonary Barotrauma
12. Other Barotraumas
13. Decompression Theory & Physiology
14. Decompression Procedures Tables
15. Decompression Sickness Clinical
16. Decompression Sickness Treatment
17. Dysbaric Osteonecrosis
18. Nitrogen Narcosis
19. HPNS
20. Hypoxia
21. Oxygen Toxicity
22. Carbon Dioxide Problems
23. Carbon Monoxide Toxicity
24. Breathing Gas Contamination

AQUATIC DISEASES

25. Drowning
26. Salt Water Aspiration Syndrome
27. Cold and Hypothermia
28. Infections
29. Dangerous Marine Animals

GENERAL DIVING RELATED MEDICAL PROBLEMS

30. Hearing Loss
31. Disorientation
32. Miscellaneous Disorders –
    - Pulmonary Oedema,
    - Contact Lenses
    - Muscular Cramps
    - Other Ear Problems
    - Headache
    - Sunburn
    - Seasickness
    - Jaw Joint Arthritis
    - Explosions
33. Unconsciousness in Divers
34. Why Divers Die
35. Sudden Death Syndrome
36. Psychological Disorders
37. Drugs and Diving

TREATMENT and PREVENTION

38. Medical Examinations for Divers
39. First Aid Kit
40. Oxygen Therapy Techniques
41. Training and Safety Checklist
42. Resuscitation Review

RECENT DEVELOPMENTS

43. Technical Diving

APPENDICES

A. Diving Medical Library
B. Emergency Contact Numbers
C. In-water O_2 Recompression Therapy
D. Divers Alert Network

Chapter 1

HISTORY of DIVING

Historians are unable to identify the first divers. Probably the techniques they used were similar to those of the native pearl and sponge divers. They may have used a stone weight to ensure rapid descent, but it is unlikely that they could dive deeper than 30 metres, or spend longer than 2 minutes underwater. Later, diving was employed for military purposes (such as destroying ships anchoring cables, boom defences, etc.) and for salvage work. Divers took part in great naval battles between 1800 BC and 400 BC. Alexander the Great was said to have descended in a diving bell (circa 330 BC) but the details are scarce and some of the stories of the descent are fanciful.

Commercial diving evolved through the 19th and 20th centuries and encompassed salvage and shell diving, extending into exploration, deep diving, off shore oil rigs, aquaculture, ecology and most importantly for you – recreational diving.

The history of diving evolved in two directions. The first is the development of diving equipment – described in this chapter. The second is the understanding of diving physiology and medicine – described in the rest of this text.

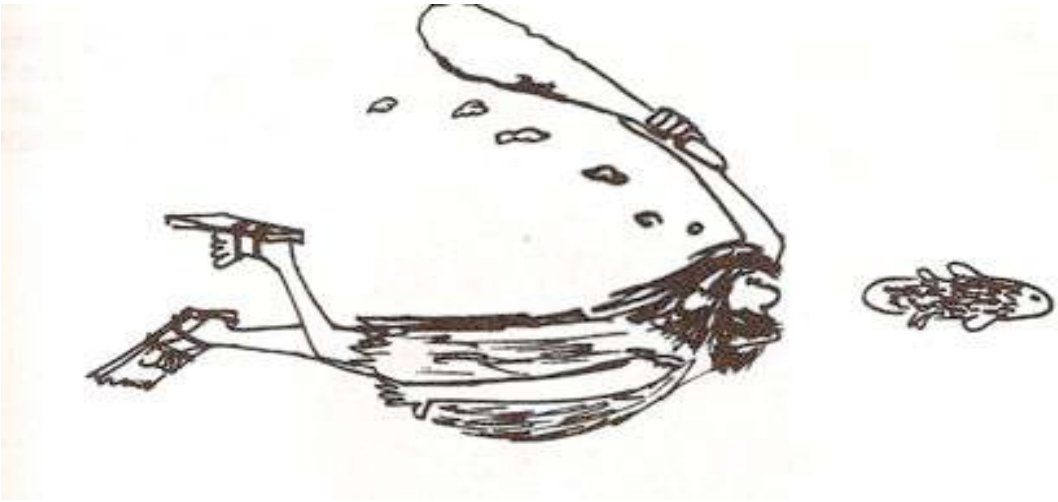

Fig. 1.1

A Roman historian, Pliny, recorded the earliest use of surface supplied breathing air by divers in AD77, when a breathing tube connected the diver to the surface. This possibly represents an

early "schnorkel". Its use was limited to very shallow dives, since man's respiratory muscles cannot draw air very far down from the surface – maximum half a metre. It was also depth limited due to the excessive volume of the breathing tube.

Leonardo da Vinci sketched several designs for diving equipment and submarines. Many diagrams of divers' hoods can be found in other historical texts from 1500 AD onwards, but much of this equipment would not have worked at depths greater than a few feet. They did, however, attest to man's desire to remain below the surface for extended periods. In 1680 Borelli, an Italian, designed a diving set which purported to be a self-contained diving apparatus. Although it was impracticable, the idea was revolutionary at that time. Despite the fact that much diving equipment was primitive and rarely functioned adequately, diving bells were used with success from the 17th century onwards.

Fig. 1.2

Diving with a helmet (the equivalent of an upturned bucket which enclosed the diver's head) gradually became an accepted method. It contained air that was pumped down from the surface following the development of efficient air pumps around 1800 AD. Bellows were used to force air down to the divers. This allowed longer and deeper dives and brought to light the many physiological problems caused by the undersea environment.

In 1837 Augustus Siebe marketed the first effective standard diving dress. This incorporated an air-supply line connecting a pump or compressor on the surface to a diving helmet. The helmet was attached by an airtight seal to a flexible suit that enclosed the diver and was filled with air.

Fig. 1.3
Siebe's helmet.

Fig 1.4
Rouquayrol and Denayrouze, Diving suit 1863

The development of **self-contained** air supplies was impeded by the lack of sufficiently powerful compressors and reservoirs. In 1863 the Frenchmen, Rouquayrol and Denayrouze, invented the first satisfactory demand regulator for self-contained underwater breathing apparatus (SCUBA), but due to lack of suitable high pressure air compressors and cylinders, it was limited to surface air supply lines.

In 1878 H. A. Fleuss made a workable self-contained (closed-circuit) oxygen breathing apparatus utilising caustic potash to remove exhaled carbon dioxide. "Closed" refers to the absence of an outlet for gas (i.e. no bubbles) and means that the exhaled gas is rebreathed. This was the forerunner of modern closed-circuit diving units.

Divers in the late 1800's were capable of reaching depths in excess of 50 metres, but the effects of decompression sickness (or bends) caused much concern and many injuries to divers. Paul Bert, a French scientist, was the first to explain that the disease was caused by the formation of nitrogen bubbles in the body and proposed the idea of a slow ascent to the surface. It was not until the early 20th century that Dr J. S. Haldane derived satisfactory mathematical decompression tables to overcome this physiological problem of deep diving. The first successful tables were based on the assumption that decompression sickness could be avoided by not exceeding a 2:1 pressure reduction between stops. It reflected a mathematical model of inert gas behaviour in a body and was to be the forerunner of current decompression tables. Later observations showed this principle to be incorrect in many cases, but these early tables and the later modified versions, prevented many divers from developing the bends.

Diving research this century has lead to a great improvement in all forms of diving equipment and since 1940 the use of such equipment has increased greatly. The design by Cousteau and Gagnan in 1943, of a proper demand-regulated air supply from compressed air cylinders worn on the back has developed into modern day **scuba**.

The scuba equipment used today, with the high-pressure regulator on the cylinder and a single hose to a demand valve in the mouth, was invented in Australia and marketed by an engineer named Ted Eldred in the early 1950s, under the *Porpoise* trademark.

Closed-circuit rebreathing apparatus using, oxygen or oxygen/nitrogen mixtures, has also been improved considerably since the early units used by Italian Naval divers in their attacks on shipping in Gibraltar in 1941. With the advent of deeper diving, gas economy has become a major problem and for this reason closed-circuit systems have achieved even greater importance.

Diving to depths in excess of 100 metres required not only the development of specialised closed or semi-closed circuit rebreathing apparatus, but also the use of other inert gas mixtures mixed with oxygen. Nitrogen, because of its narcotic effect at depth, has been replaced largely by other gases such as helium and hydrogen. These are not used without complications – as all gases cause specific physiological problems and no ideal mixture yet exists.

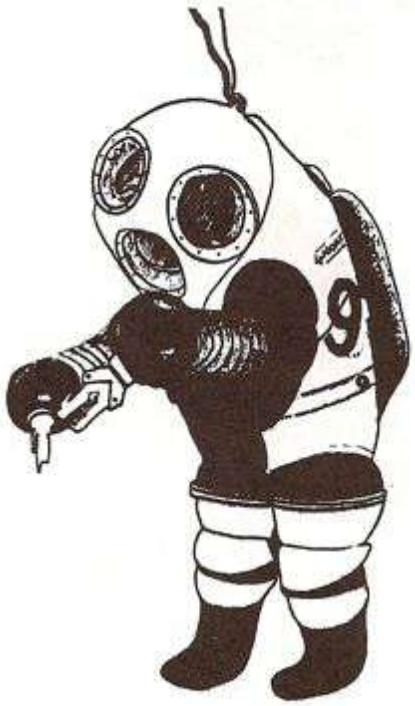

Fig. 1.5
"Jim" one-atmosphere diving suit.

The advent of **saturation diving** has completely revolutionised the ability to dive and work at great depth, and for lengthy durations, and this is economically rewarding. The system is based on saturation at any depth of all the diver's tissues by the inert breathing gas. Once this is achieved the body is incapable of absorbing further amounts of gas, no matter what the duration of exposure at this depth. Hence, further exposure does not lengthen decompression times. This practice is now adopted for most diving with extended bottom times at depths in excess of 100 metres.

In an attempt to reduce the risks in deep diving, **one-atmosphere diving suits** (ADS) have been developed out of strong lightweight alloys. These suits are fitted with articulated joints and use mechanical levers or claws for "hands". Some even have mobility and propulsion, but all require backup 'rescue' facilities. They are equipped with a self contained rebreathing apparatus and are often used at depths of 200–300 metres. Although somewhat bulky and requiring hoisting gear at the surface, divers can achieve a reasonable degree of movement at depth with the latest models. These suits are also useful for inspection-type work, although much of this is now done by non-manned **Remote Operated Vehicles** (ROV's) with video surveillance.

Fig 1.6 The Royal Australian Navy Clearance Diving Team, 1955, with oxygen breathing equipmenrt and wearing the famous “Clammy Death” suit.

Chapter 2

PHYSICS

To understand the physical and physiological problems which can confront a diver, it is helpful to recall a few basic physical laws of nature. Only a brief and simplified review of the physics of diving is given in this text. For more detailed explanations, refer to the diving manuals.

PRESSURE

Some of the major physical hazards are related to the effects of pressure. Pressure is defined as force per unit area. i.e.

$$\text{PRESSURE} = \frac{\text{FORCE}}{\text{AREA}}$$

If a force is spread over twice the area, the pressure is halved.

This explains why, for example, wide tyres are preferable for driving on beaches. The weight of the vehicle (force) when spread over a large area causes less pressure on the sand. This vehicle is less likely to sink into the sand than one with narrow tyres.

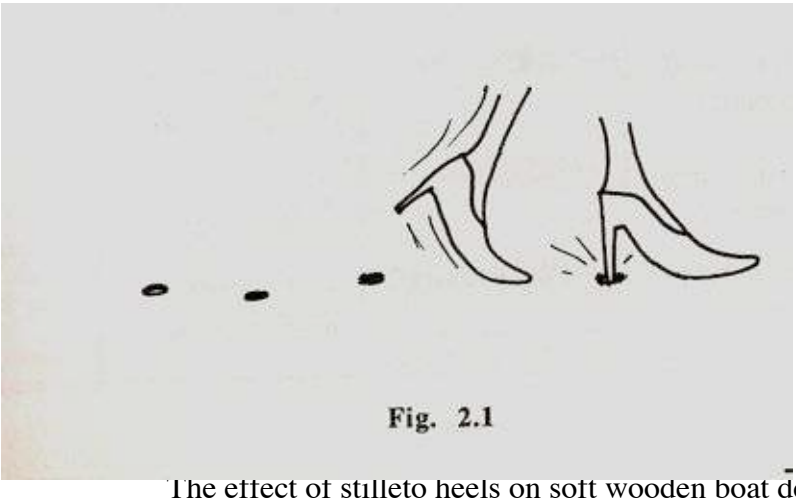

Fig. 2.1

The effect of stilleto heels on soft wooden boat decks

Gases exert pressure because they are made up of lots of fast moving molecules. The greater the number and the faster they move, the greater the pressure.

Pressure on a Submerged Diver

The pressure acting on a submerged diver has two components:

1. The atmosphere above the water, termed **atmospheric pressure,**
2. The weight of the water above the diver, termed **hydrostatic pressure.**

Divers' depth gauges are calibrated only to read the hydrostatic pressure (the depth of water) and so they read zero at sea level. They do not read the 1 atmosphere (1 ATA) above them. Thus the "gauge pressure" is always 1 atmosphere less than the true or "absolute" pressure. We will now elaborate.

Atmospheric Pressure

The atmosphere above the earth is some 150 km high. Although air is very light, this amount of air has significant weight and exerts substantial pressure on the earth's surface.

ABSOLUTE PRESSURE	GAUGE PRESSURE	DEPTH of SEAWATER
1 ATA	0 ATG	Surface
2 ATA	1 ATG	10 metres (33ft)
3 ATA	2 ATG	20 metres (66ft)
4 ATA	3 ATG	30 metres (99ft)

Table 2.1
Pressure at Depth

Atmospheric pressure at sea level is referred to as "one atmosphere" or "one bar". It is the same as 101.3 kPa, 1 kg/cm^2, 760mm Hg and 14.7 psi. At higher altitudes, atmospheric pressure is reduced, a factor which has a significant effect on diving in mountain lakes (see Chapter 6).

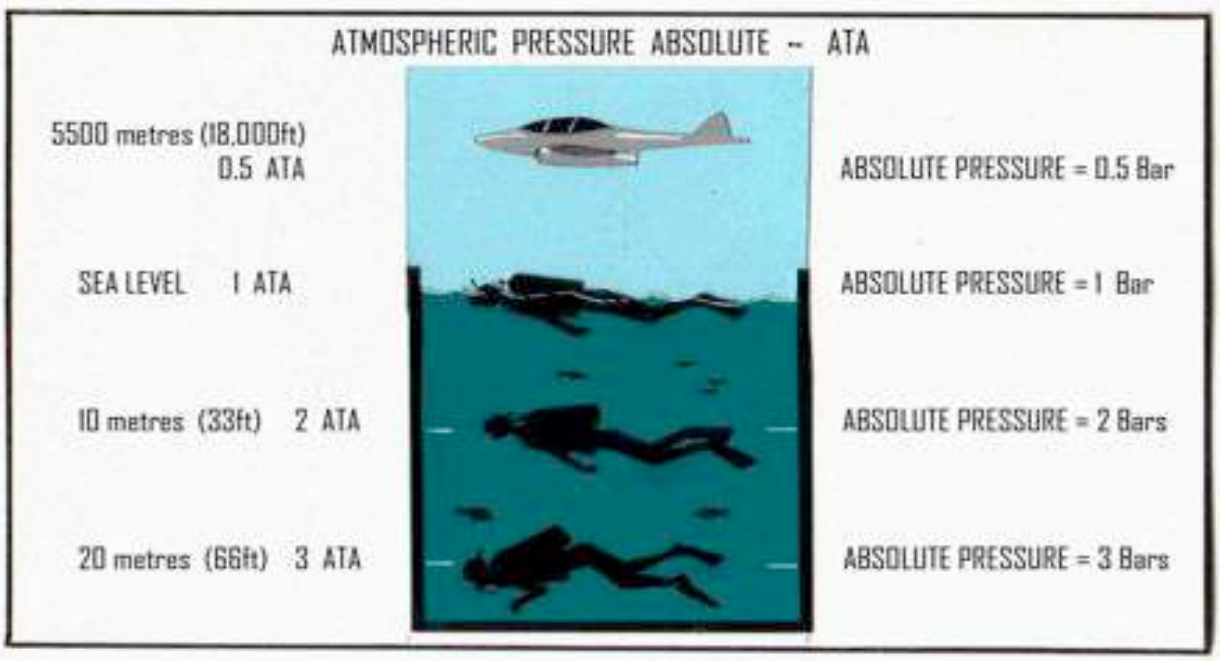

Fig. 2.2 Atmospheric and Hydrostatic Pressures (depth) added and thus converted to Absolute Pressure

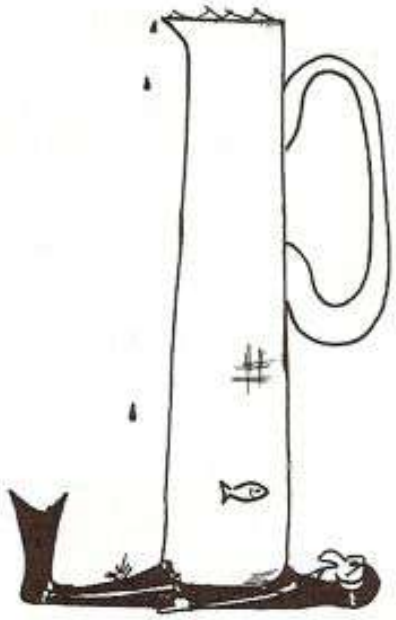

Fig. 2.3

Water is much denser than air and 10 metres (or 33 ft) of sea water exerts the same pressure (weight) as the whole 150 km of atmospheric air i.e. 1 ATA. For every additional 10 metres the diver descends, the water will exert a further pressure, equivalent to another atmosphere (1 ATA).

Common units of pressure (approximately):

1 ATMOSPHERE	=	10 metres sea water
	=	33 feet sea water
	=	34 feet fresh water
	=	1 kg/cm^2
	=	14.7 lbs/in^2, psi
	=	1 bar
	=	101.3 kilopascals, kPa
	=	760 millimetres mercury, mm Hg.

Absolute Pressure

The total pressure exerted on a diver at depth will be the pressure due to the atmosphere acting on the surface of the water (atmospheric pressure) plus the pressure due to the depth of the water itself (hydrostatic pressure).

The total pressure acting on the diver is termed the "absolute pressure". It is often expressed in terms of atmospheres and is called "atmospheres absolute" or "ATA".

To calculate the absolute pressure acting on a diver at a given depth in terms of atmospheres, divide the depth in metres by 10 (since every 10 m. sea water exerts 1 atmosphere pressure) and add 1 (the pressure of the atmosphere above the water).
e.g. the absolute pressure at 40 metres is [40 ÷ 10] + 1 = 5 ATA

(The depth in feet, divided by 33 + 1 also calculates absolute pressure, for those in the USA, e.g. the absolute pressure at 99 ft is 99/33 +1 = 4 ATA).

Gauge Pressure

As described above, hydrostatic pressure in diving is generally measured by a **pressure** or **depth gauge**. Such a gauge is normally set to register a pressure of zero at sea level and so it ignores the pressure due to the atmosphere (1ATA).

The pressure registered by a gauge at 10 metres sea water depth would thus be one atmosphere gauge (1ATG) or equivalent units. Gauge pressure is converted to absolute pressure by adding 1 atmosphere pressure.

Partial Pressure

With a mixture of gases, the proportion of the total pressure contributed by each of the gases is termed its partial pressure (its part of the pressure). The partial pressure contributed by each gas is proportional to its percentage of the mixture. Each gas contributes the same proportion to the total pressure of the mixture, as is its proportion in the composition of the mixture.

e.g. air at 1 ATA contains 21% oxygen, hence the partial pressure of oxygen is 0.21 ATA and air at 1 ATA contains 78% nitrogen, hence the partial pressure of nitrogen is 0.78 ATA.

GAS LAWS

Gases behave in nature and in diving according to several laws. Knowledge of these laws is important to the diver because they influence the duration of the air supply and affect the gas containing spaces in the body such as the ears, sinuses and lungs. They also cause other diving illnesses.

Boyle's Law

This defines the relationship between pressure and volume. It states that the **volume of a given mass of gas varies inversely with the absolute pressure (if the temperature remains constant).**

Stated simply, for a given amount of gas, if the pressure is increased, the volume is proportionally decreased and vice versa. This means that if the pressure is doubled, the volume is halved and vice versa.

Stated mathematically: V varies as $\frac{1}{P}$ (where V = volume and P = pressure)

It follows that for a given amount of gas, the volume multiplied by the pressure always has a constant value.

i.e. $P \times V$ is constant.

So if a sample of gas has an original volume of V_1 and an original pressure of P_1, and either the pressure or volume are changed, the new volume V_2 and the new pressure P_2 will multiply out to the same value.

i.e. $\mathbf{P_1 \times V_1 = P_2 \times V_2}$

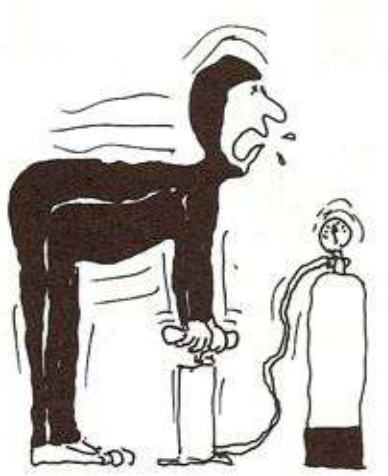

Fig. 2.4

This law can easily be demonstrated by a piston and cylinder such as a bicycle pump. If the piston is pushed into the cylinder half way, and the escape of gas prevented, the pressure in the cylinder will be found to have doubled. By this process, many litres of air can be crammed into a bicycle tyre but at the cost of an increase in pressure in the tyre (and hard work). Compressors work in this way, squeezing 2000 or more litres of air into a scuba cylinder – but at a high pressure.

Since water pressure increases with depth, the consequent reduction in gas volume becomes very important to the diver because his body has numerous air spaces.

Descent Problems: The air in the diver's middle ear and sinuses will contract in volume as the diver descends. If these volume changes are not compensated for by adding more air ("**equalisation**"), then pressure damage (**barotrauma**) to the tissues will result. For example:

If a 6 litre bag is filled at the surface (1 ATA) and taken to 20 metres depth (3ATA), the volume will be reduced by a factor of 3, to 2 litres.

$$P_1 \times V_1 = P_2 \times V_2$$
$$\therefore 1 \times 6 = 3 \times V_2$$
$$\text{i.e. } V_2 = 2 \text{ litres}$$

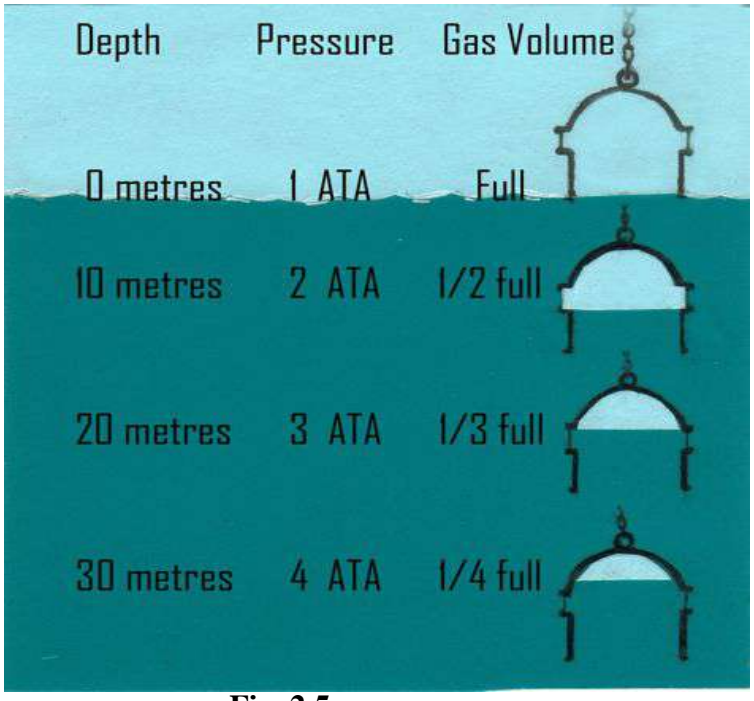

Fig. 2.5

In the same way, if a **breath-hold diver** takes a full breath at the surface and descends to 20 metres (3 ATA), the volume of air in his lungs may be reduced from 6 litres to 2 litres. The chest and lungs cope with compression better than distension. The limit for breath-hold diving is not known, but now has been shown to exceed 150 metres in certain individuals.

Ascent Problems. An average male diver's lungs may contain about 6 litres of gas. If a diver takes a full breath at 20 metres (3 ATA) from his scuba set and returns to the surface (1 ATA) without exhaling, the volume of gas in his lungs will increase from the 6 litre total lung capacity to 18 litres (6 × 3 litres).

This can be easily calculated this way:

$$P_1 \times V_1 = P_2 \times V_2$$

$$P_1 = 3 \text{ ATA}, \; V_1 = 6 \text{ litres}, \; P_2 = 1\text{ATA}, \; V_2 = ? \text{ litres}$$

$$V_2 = \frac{P_1 \times V_1}{P_2} = \frac{3 \times 6}{1} = 18 \text{ litres}$$

The lungs would have to expand to 18 litres to accommodate this volume – well beyond their rupturing point, causing **burst lung** (**pulmonary barotrauma of ascent**).

An important practical observation of Boyle's Law is that **the greatest volume changes take place near the surface. This means that the greatest danger from barotraumas is near the surface — and this applies with descent as well as ascent.**

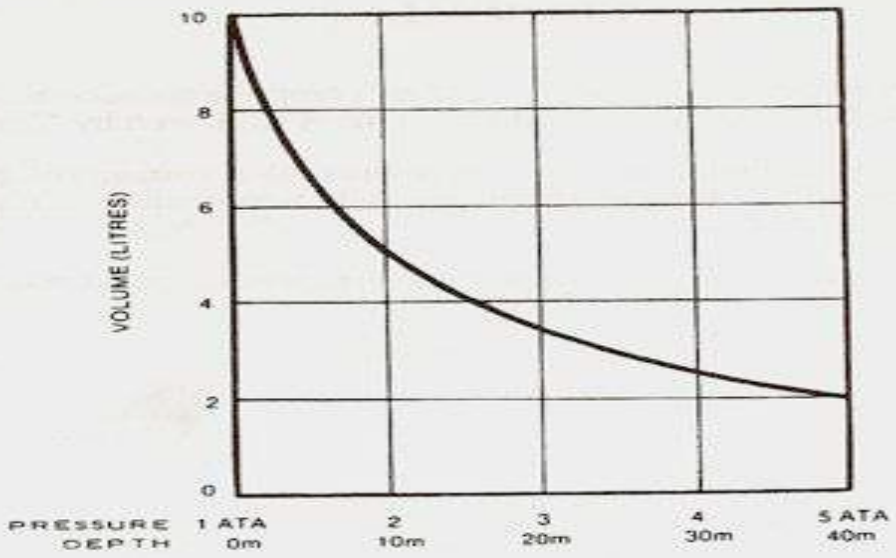

Fig. 2.6
This diagram shows changes in gas volume caused by pressure change at various depths. Note that maximal volume change is near the surface.

For example, if diver has a maximum of 4 litres of air in his lungs at 40 metres depth (5 ATA) and ascends 10 metres without exhaling (to 4 ATA), the volume in the lungs will increase to 5 litres:

$$P_1 \times V_1 = P_2 \times V_2$$
$$5 \times 4 = 4 \times V_2$$
$$\therefore V_2 = 5 \text{ litres}$$

Some people could possibly accommodate this expansion without lung damage.

If the same diver started at 10 metres depth (2ATA), and then ascended 10 metres to the surface (the same ascent distance as before), without exhaling, the pressure would change from 2ATA to 1ATA. The air in the lungs would expand from 4 to 8 litres. This would rupture his lungs.

Although the dives involved the same ascent distances, the volume change, and hence the danger, in response to Boyle's Law, is much greater near the surface.

Many divers are not aware of this and have a fallacious belief that if they confine their diving to shallow depths they will minimise the risk of barotrauma.

Buoyancy compensators are similarly affected by depth changes in response to Boyle's Law. **Wet suits** are also affected and lose their buoyancy and insulating properties with depth.

Charles' Law

Most divers will have noticed that bicycle pumps and air compressors become hot during use. As the volume of gas is compressed, heat is produced. This is explained by Charles' Law.

This Law states that **if the pressure remains constant, the volume of a given mass of gas varies directly with the absolute temperature** (absolute temperature is obtained by adding 273 to the temperature in degrees Celsius).

In other words, at a fixed pressure, if gas is heated it expands, and if gas is cooled its volume contracts.

Fig 2.7

Charles' and Boyle's laws can be combined into the **General Gas Law** : $\frac{PV}{T}$ is constant

For the non-mathematically minded this means that for a given amount of gas, the pressure multiplied by the volume, divided by the temperature, always comes to the same value – so if one of these factors is varied, it has an effect on the other two.

If a gas sample having $\frac{P_1V_1}{T_1}$ has one of these factors changed,

the new set of values $\frac{P_2V_2}{T_2}$ will multiply out to the same answer

i.e. $\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$

Fig. 2.8

Stated in another way; if a gas is compressed, its volume decreases and it gets hotter. If the gas is heated and the volume is prevented from expanding, the pressure rises.

The consequence of this law has lead to the demolition of several perfectly good automobiles (and divers!) following the storage of full scuba cylinders in the boot (trunk) in hot weather. Similarly, inflatable dive boats are often pressurised to the maximum and are then left in the sun. As the temperature rises, the pressure of the contained air progressively increases and then suddenly reduces – when the volume increases and when the boat explodes.

If gas is allowed to expand rapidly, it cools. Cooling from the expansion of previously compressed air, as it is breathed from a scuba cylinder, can lead to the regulator freezing up during cold water diving.

Problem: If the temperature of a scuba cylinder is 37˚C after being disconnected from the compressor. Its pressure gauge reads 199 ATG, what is the pressure after it has cooled to 17˚ C?

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$$

now because V_1 and V_2 are the same (the cylinder volume is unchanged), the equation can be written:

$$\frac{P_1}{T_1} = \frac{P_2}{T_2}$$

and this can be rearranged to:

$$P_2 = \frac{P_1T_2}{T_1}$$

Substituting the figures: (note that the cylinder pressure is in ATG and needs to have 1 atmosphere added to get ATA, also that the temperatures have to be converted to degrees absolute by adding 273 degrees)

$$\therefore\ P2 = \frac{(199+1) \times (273+17)}{(273+37)}$$

$$= 187 \text{ ATA}$$

Dalton's Law

With a mixture of gases, the total pressure exerted by the mixture, is the sum of the pressures that would be exerted by each of the gases if it alone occupied the total volume. That is, the total pressure is the sum of the partial pressures.

As the overall pressure increases (with descent underwater), so the partial pressure of each constituent gas increases.

e.g. if air contains approximately 21% oxygen (O_2) and 78% nitrogen (N_2), then in a sample of air at a given pressure, O_2 will contribute 21% of the total pressure and N_2 will contribute 78%.

AIR BREATHING

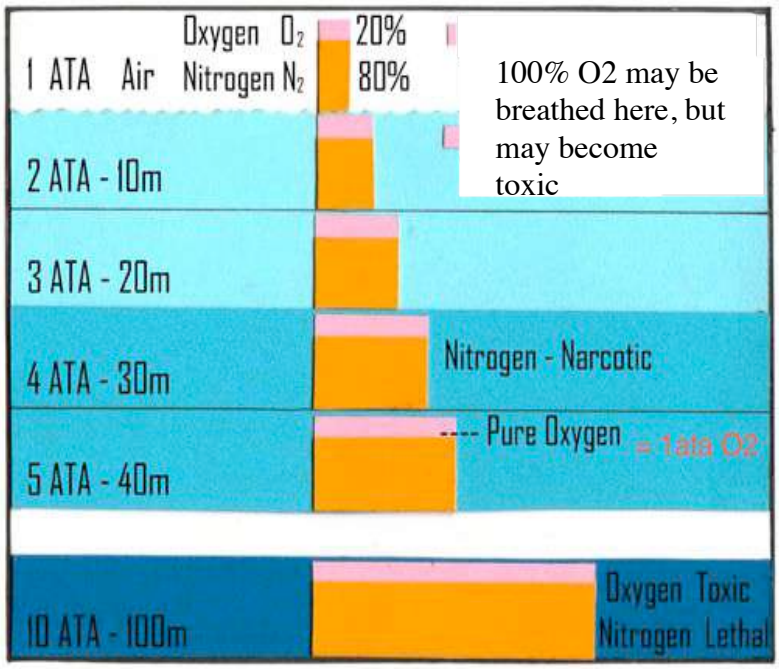

Fig. 2.9

At atmospheric pressure the partial pressure of O_2 in air is $^{21}/_{100}$ of 1 ATA. = 0.21 ATA while the partial pressure of N_2 is $^{78}/_{100}$ of 1ATA = 0.78ATA.

To calculate the partial pressure of a gas, multiply the percentage of gas by the absolute pressure.

This law is important when considering the toxic effect of gases at depth or the use of O_2 for treatment purposes.

Problem: Since O_2 can cause convulsions when breathed at greater than 1.8 ATA, would it be safe to breathe a mixture of 50% O_2 and 50% N_2 at 30 metres (4 ATA) ?

The partial pressure of O_2 = 50% × 4ATA = 2ATA

This oxygen / nitrogen mixture would be potentially toxic at this depth.

Henry's Law

This law describes the dissolving of gas in a liquid and states that the **quantity of gas which will dissolve in a liquid at a given temperature is proportional to the partial pressure of gas in contact with the liquid.** This means that if the pressure of gas exposed to a liquid increases, then more gas will dissolve in the liquid.

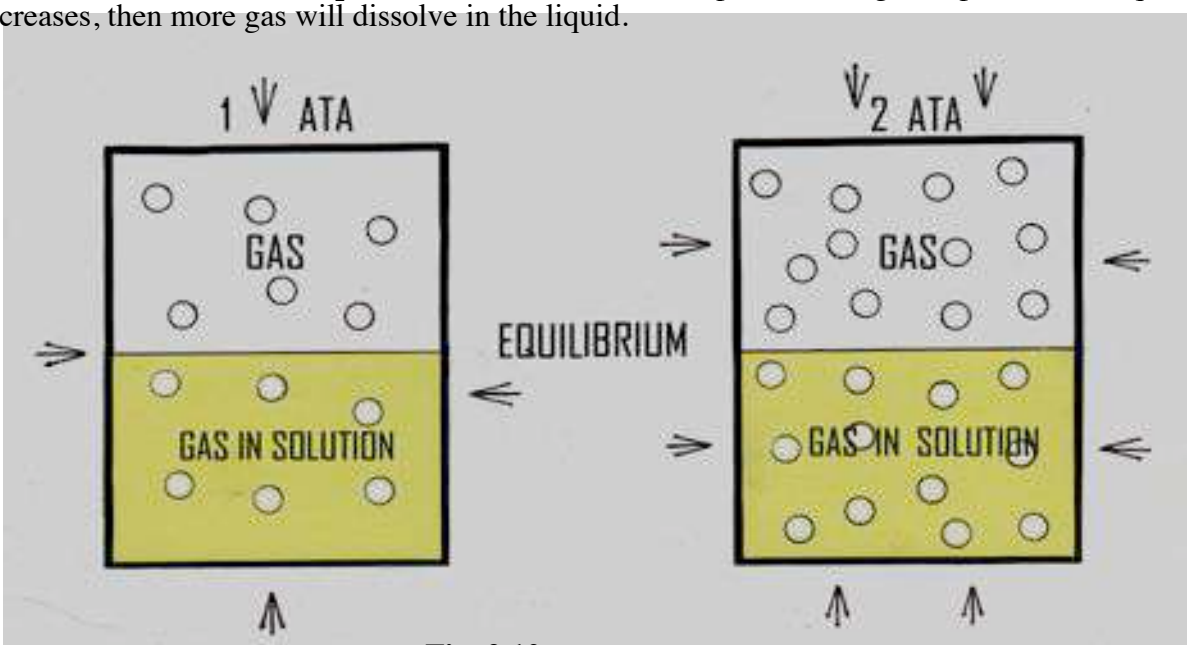

Fig. 2.10
This diagram shows how more gas molecules are dissolved in a liquid as the pressure of the gas exposed to the liquid is increased from 1 to 2 ATA.

An example of this law can be seen whenever a fizzy soft drink bottle is opened. During the manufacture of these drinks, carbon dioxide is dissolved in the liquid under pressure and the lid on the bottle maintains the pressure. When the bottle is opened and the pressure released, the liquid will not allow as much gas to be dissolved and so the excess gas is released from solution in the form of bubbles.

At sea level (1ATA) the human body contains approximately 1 litre of N_2 dissolved in the tissues. Whenever a diver breathes compressed air at depth, more N_2 will dissolve in the body because the partial pressure of N_2 in the air being breathed is increased. This is the cause of **nitrogen narcosis**.

Under certain circumstances, when the diver returns to the surface this N_2 can come out of solution in the form of bubbles. These bubbles cause tissue injury which is the basis of **decompression sickness** ("bends").

Fig. 2.11

Diffusion of Gases

If a diver were to pass wind in a confined room, all the occupants of the room would soon be aware of the fact but, fortunately, not necessarily the source.

This process of distribution of gas is termed diffusion. It is caused by the rapid random movement of gas molecules to all parts of a contained space. Gas molecules, being only single or small groups of atoms, are able to easily diffuse through watertight membranes such as blood capillaries or cell walls. This process allows O_2 and other gases to pass from the lungs to the blood and tissues, and then back.

GASES OF IMPORTANCE TO DIVERS

Air

Air consists of a mixture of O_2 + N_2 + a trace of carbon dioxide (CO_2), and minute amounts of rare gases. Rare gases such as Neon (Ne), Argon (Ar) and Xenon (Xe), and Hydrogen (H_2) exist in trace amounts only.

The approximate composition of air is:

Oxygen (O_2)	**—**	**21% by volume**
Nitrogen (N_2)	**—**	**78% by volume**
Carbon Dioxide (CO_2)	**--**	**0.04% by volume**
Others	**--**	**< 1 % by volume**

Some less reputable suppliers of air fills for scuba tanks provide free additives to the compressed air, such as dust, oil, hydrocarbons, rust, water vapor and carbon monoxide (CO).

Oxygen – O_2

This is a colourless, odourless, tasteless gas which is indistinguishable from air to breathe.

It is essential for metabolism and maintenance of life yet in quantities exceeding those in air it is **toxic** to man. Its proportion in air (21% or more specifically, a partial pressure of 0.21 ATA at sea level) is critical. A little more than this causes O_2 toxicity, a little less will not support human life. For this reason most gas mixtures breathed by deep divers contain an inert gas – usually either N_2 or helium (He), mixed with O_2 to ensure that the O_2 composition is maintained at a partial pressure close to 0.2 ATA (0.16 – 0.40 ATA).

O_2 supports combustion vigorously and can cause normally non-flammable substances (such as the occupants of a recompression chamber) to burn brilliantly if it is present at a sufficiently high partial pressure.

Divers should be aware of the potentially explosive and combustible properties of oxygen, as they may require to use it in first-aid, or be inadvisably enticed into diving with high oxygen mixtures.

Nitrogen – N_2

This gas, which is the major constituent of air, is also colourless, odourless and tasteless. N_2 dissolves well in body fluids and tissues, causing **narcosis** at depth and **decompression sickness** when it bubbles out of solution, after ascent.

It is termed an "**inert gas**" because it does not take part in human biochemical processes. The Creator appears to have included this gas in air to prevent us from developing O_2 toxicity, and to reduce the fire hazard.

Divers vary this N_2/O_2 ratio (in Nitrox, oxygen enriched air or mixed gas diving) in an attempt to improve on nature, extend diving durations, and reduce narcosis.

Carbon Dioxide – CO_2

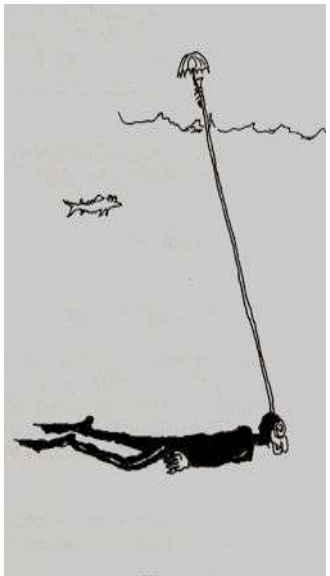

Fig. 2.12

This gas is also colourless, odourless and is said to be tasteless. However if a diver inhales a mouthful of CO_2 from a buoyancy vest inflated from a CO_2 cartridge it will be found to taste very nasty, due to its formation of carbonic acid in water.

CO_2 is a by-product of cellular metabolism and we exhale approximately 5% of CO_2 in our breath.

If a diver rebreathes some of his exhaled gas by using faulty breathing equipment or an excessively long snorkel the CO_2 will accumulate in the body leading to toxicity. These effects are discussed further in Chapter 22.

Carbon Monoxide – CO

This gas is colourless, odourless and tasteless. It cannot be detected by a diver and even in trace amounts can cause loss of consciousness or death.

It is usually produced as a product of incomplete combustion of carbon containing compounds and is a constituent of internal combustion engine exhausts and cigarette smoke.

Air contaminated by carbon monoxide, if supplied in scuba cylinders or by surface supply to divers, may have lethal results (see Chapter 23).

Helium – He

This is a colourless, odourless, tasteless gas, which is very light and very expensive. It is obtained from underground natural gas sources found in North America and elsewhere.

It is used to dilute O_2 in gas mixtures breathed at great depths because it has little tendency to produce narcosis (e.g. Heliox may be 90% He + 10% O_2, or any other proportion).

Due to its very low density it readily escapes through small leaks in pipes and valves making it difficult to retain. It is also a very effective conductor of heat, causing serious problems with **hypothermia**.

The low density of He alters the normal process of speech production causing "**Donald Duck**" like speech when a diver breathes this gas.

Hydrogen – H_2

This is a very lightweight gas that can replace N_2 to reduce narcosis at depth. Unfortunately it can combine explosively with O_2 and the resultant water (H_2O) is not sufficient to 'put out the diver'. It is sometimes used with very low O_2 percentages, at great depths, by skilled professional divers. It shares many problems with He.

Inert Gases:

Neon – Ne, Argon – Ar, Radon – Rn, and Xenon – Xe

These are more biologically inert gases which are present only in trace amounts in the atmosphere. They are of no importance to recreational divers.

Oil Gases

Because of lubrication needs in the compressor, oil vapors and hydrocarbons can be produced which may then contaminate the air supply. See Chapter 24.

BUOYANCY

It is important for divers to understand the factors affecting buoyancy. These are:

Density

Density is defined as **mass per unit volume** (density = mass ÷ volume).

For our purposes, mass can be considered to be the same as weight, so density is equivalent to weight per unit volume.

A substance is more dense than another if the same volume has more weight. Try lifting a bucket of water and then a bucket of lead, to illustrate this.

Specific Gravity

Specific gravity (S.G.) is the density of a substance compared to the density of fresh water which is given a value of one.

Lead has a specific gravity of 13.5 so it is 13.5 times as dense as water.
e.g.1 litre of water will weigh 1 kg., while the same volume of lead will weigh 13.5 kg.

The concept of specific gravity is important since the specific gravity of a substance determines whether it will float or sink in water.

Fig. 2.13

A substance with a specific gravity greater than 1 (i.e. denser than water) will sink. Lead, with a specific gravity of 13.5, does not float well, whereas oil, with a specific gravity of 0.8, floats easily — producing an oil slick.

The human body has a specific gravity of slightly greater than 1, depending on its content (fat has a specific gravity less than 1, and bones are greater than 1) but the air content of the lungs provides enough buoyancy to allow most people to float.

Archimedes Principle

The ancient Greek, Archimedes (apparently while reclining in his bath), discovered that when an object is immersed in a fluid, it appears to be lighter, and that the apparent loss of weight (or **buoyancy**) is equal to the weight of water displaced by the object.

Fig. 2.14

That is – the buoyant effect will be equivalent to the weight of fluid of equal volume to the immersed object.

Depending on whether the weight of fluid displaced is greater than, equal to or less than the weight of the object, an object immersed in the fluid will either float, remain suspended or sink. Even an object which sinks will still appear to be lighter than it would out of the fluid.

Sea water is denser than fresh water because of the salt content, so a greater weight of sea water will be displaced by an object. Hence objects in sea water are more buoyant than in fresh water.

Air (in the abdomen, buoyancy compensator and wet suit) contributes to buoyancy. Unfortunately air in these compartments varies in volume in response to the pressure changes with varying depth, making constant buoyancy adjustments necessary. This is usually accomplished by adding air to, or releasing it from, the diver's buoyancy compensator.

Divers go to considerable lengths to vary their buoyancy to help them submerge, to stay at a given depth, or to ascend or stay afloat in an emergency.

PHYSICAL EFFECTS OF THE ENVIRONMENT

Temperature

Body heat is a form of energy, the level of which can be estimated by measuring the body temperature.

Heat energy flows from areas of high temperature to areas of low temperature. The heat transfer which is important to the diver is **thermal conduction** (or transfer of heat by direct contact), and may cause **hypothermia** (low body temperature).

Since normal body temperature is 37°C and oceanic water temperature is commonly 12–20°C, the diver is almost always immersed in water at a lower temperature than his body. Usually

the water temperature decreases with depth, but there may be layers of water at different temperatures (thermoclines) – especially in still water.

Cold water creates a strong temperature gradient along which heat flows from the body, resulting in a continuous heat loss into the water. This process is assisted by water having a high capacity to conduct and absorb heat.

Since the maintenance of normal body temperature is essential for physiological functioning, the diver needs to take steps to minimise heat loss. This may be achieved by inserting a layer of air (which is a poor conductor of heat) between the diver and the water. It is conveniently contained in minute cells in a wet suit or under a rubber skin in a dry suit.

Light and Colour

Substances that transmit light have a tendency to slightly alter the path of the light rays which pass through them. This process is termed **refraction**. The degree to which they do this is termed the refractive index. Each time light passes through an interface between substances with different refractive indices, its path is bent.

When a diver views objects underwater, light must pass through the water, the face mask glass, and the air in the mask before it reaches his eyes. The light rays are refracted at each of these interfaces and the distortion makes objects appear larger and closer by a factor of about 25%.

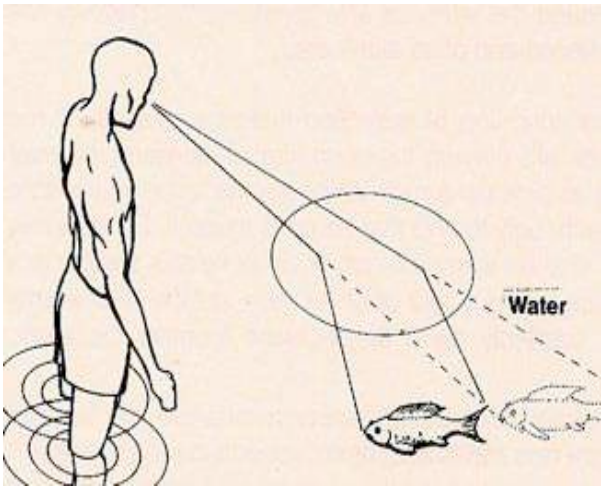

Fig 2.15 The fish appears closer, because the light rays are refracted at the air/water interface

Until the diver adapts to this distortion, it may be difficult to judge size and distances. This creates practical difficulties with simple tasks such as spear fishing.

Light rays are scattered by particles in the water making shadows less pronounced and reducing the ability to see clearly over large distances.

Clear focusing of the eye depends heavily on the refraction of light rays passing between the air in front of the eye and the cornea (the clear surface at the front of the eye). If the eyes are opened underwater without a face mask, the absence of this air/cornea interface results in very blurred vision.

Water absorbs colours to differing degrees. In clean oceanic water, red is absorbed in the first metre, orange in the first five metres, yellow in the first ten, and green and blue at greater depths. This explains why most things, regardless of their colour on the surface, appear to be coloured shades of blue or green at depths beyond about ten metres.

Inshore waters often contain yellowish products of vegetable decay which absorbs most colours except green. As a result, clean oceanic water appears blue, while inshore and estuarine water appears green from the surface.

Because deep water is lit mainly by blue and green light, coloured corals and fish at these depths look less brilliant unless illuminated by a torch or camera flash.

For safety reasons, it is advisable to wear conspicuously coloured diving equipment. However, the absorption of light underwater needs to be considered when choosing these colours. Red, for instance, which is easily visible on the surface, appears black at depth because of the significant absorption of red light by the water.

Fluorescent orange or **yellow** paint or fabric affords better visibility because the fluorescent dye actively emits light of its own colour and also provides a good contrast against natural aquatic backgrounds.

Sound

Sound waves in air are usually reflected at the air–water interface, and therefore shouting instructions to submerged divers is not of much value. Underwater, the sound wave travels much faster than in air, and this makes localisation of the source much more difficult. An example of this is the concern experienced when divers hear outboard engines, but cannot identify the distance or direction of the boat.

Altitude

If exposed to altitude (less than 1 ATA) a variety of effects may endanger the diver. Some equipment may be affected e.g. pressure gauges, and the diving profile needs to be modified to prevent pulmonary barotrauma and decompression sickness (see Chapter 6).

Chapter 3

PHYSIOLOGY*

A basic understanding of the bodies' processes is needed to grasp the physiological effects of diving and the application of diving medicine. The cardiovascular and respiratory systems are described here while the physiology of some other organs, such as the ear, are considered in specific chapters.

METABOLISM

The Need for Energy

Energy is a fundamental requirement for all life processes. It is needed for growth, repair, movement and all the active functions of the body. The fuel for this energy comes from **carbon compounds**, which are incorporated in complex molecules in the food we eat. This is biochemically dismantled in the digestive tract into simple chemical compounds which are absorbed and carried by the blood stream to the cells. Here they undergo further biochemical processing until ultimately the carbon is combined with oxygen (O_2), forming carbon dioxide (CO_2) and releasing energy.

This is similar to the energy formation which takes place in an automobile engine or a fire, where carbon in fuel or wood is combined with O_2 to produce energy. The body processes will only function under strict conditions of O_2 availability, temperature and acidity.

The body needs a means of transferring food products to the cells, together with delivery of O_2 and removal of CO_2. This is performed by the blood, in the vascular system. It comprises **arteries** which take blood to the tissues, a vast network of microscopic **capillaries** that bring the blood into contact with all the cells of the body, and **veins** which return blood to the heart.

The blood is circulated through the blood vessels by a muscular pump – the heart, and the whole system is called the **cardiovascular** system. It brings O_2 from the lungs to the cells and eliminates CO_2 through the respiratory system.

*** Reading this chapter can be delayed if time does not permit, and the reader is in a hurry to commence diving**

RESPIRATION

Anatomical Structure

The respiratory tract begins at the mouth and nose and ends in the microscopic air sacs called the **alveoli**, in the lungs.

The **nose**, apart from its decorative function, warms and humidifies the air that we breathe. It also filters large particles which might otherwise be inhaled. If the nose is bypassed by breathing through the mouth, a snorkel or scuba regulator, the lung then has to cope with drier, colder, unfiltered air.

After passing through the mouth or nose, the air then enters the throat where the **larynx** (or voice box) is situated. This is recognised as the "Adams Apple". The larynx produces the sounds of speech as well as helping to protect the lungs from inhalation of foreign material.

When sea-water from a flooded snorkel or scuba regulator enters the larynx, a trap-door like structure called the epiglottis closes over the opening and the vocal cords shut to prevent the foreign material from entering the lungs. If any material passes these structures, the cough reflex, activated by foreign material touching the inside of the air passages, may cause a **coughing** reaction which tends to expel whatever has been inhaled.

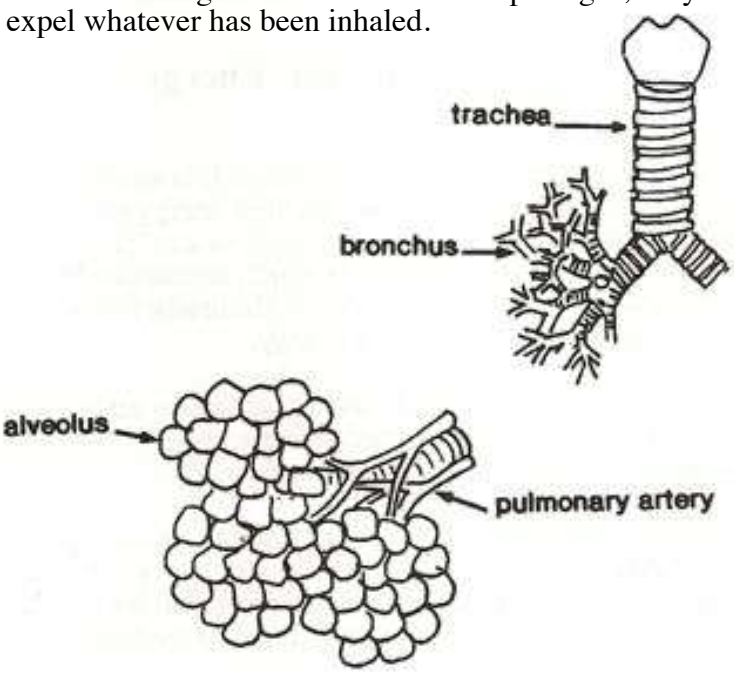

Fig. 3.1
Diagram showing the trachea and bronchial tree, terminating in the alveoli. The relationship between the bronchus, alveoli and branches of the pulmonary artery can be seen.

Below the larynx the air passes through a tube called the **trachea**. This is about as thick as the average snorkel and branches inside the chest into two tubes, the **bronchi**, which lead to the lungs. Those air passages are lined with cells covered with microscopic hairs (cilia) which move a sheet of secreted mucous slowly upwards towards the larynx. Small pieces of foreign material such as dust eventually find their way to the larynx, along with this mucous sheet. It is then either coughed-up or swallowed. The cilia may be damaged by smoking or infection, causing retention of mucous and inhaled material which may eventually obstruct the air passages.

The bronchi divide repeatedly into progressively smaller passages rather like the branches of a tree. These passages have encircling muscles in their walls which, by contraction or relaxation, can vary the diameter of the air passage.

In **asthma** the muscles of the small bronchi become oversensitive and overactive, causing excessive narrowing and obstruction of these air passages. This can occur in response to exercise, allergy, cold, infection, anxiety, smoking or other inhalants such as sea water. At the same time, the cells lining these passages produce excessive and thickened mucous. The combination of these factors causes airway narrowing which has serious repercussions for a diver.

The smallest branches of the bronchi end in bunches of microscopic air sacs called **alveoli**. The vast number of alveoli are packed together into the two sponge like organs, the **lungs**. There are about 300 million alveoli in the lungs and the combined surface area of all the alveoli in the lungs is equal to about half a tennis court. The alveoli are lined by a thin layer of fluid containing a detergent-like substance called **surfactant**. This acts as a wetting agent to prevent the alveoli from collapsing from surface tension.

The surfactant lining of the alveoli can be damaged in disease or by inhalation of water, leading to collapse of the lungs and serious respiratory difficulty.

Each alveolus is surrounded by a network of blood capillaries. These bring the blood into close contact with the air in the alveolus, with only the microscopically thin walls of the alveolus and capillary separating the two.

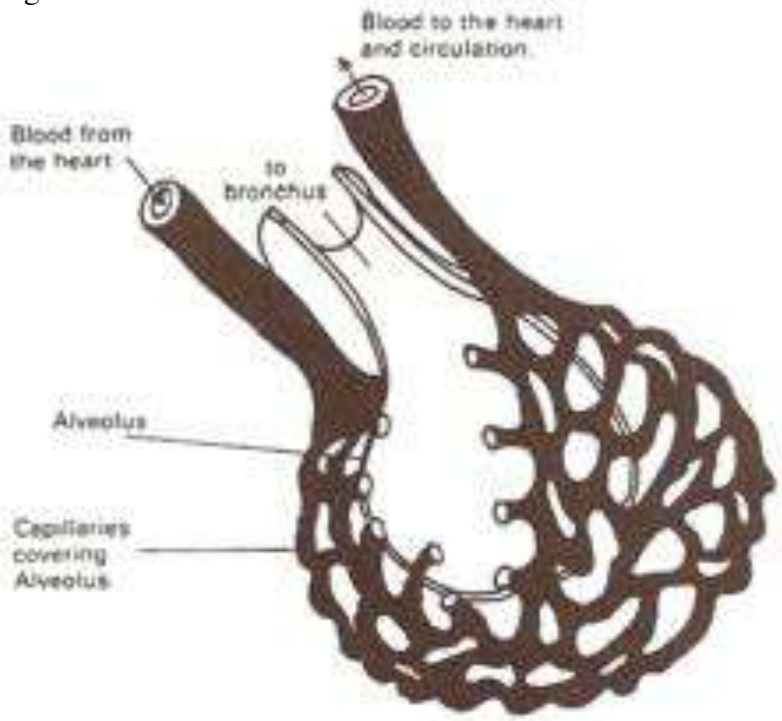

Fig. 3.2
This diagram illustrates an alveolus with its surrounding meshwork of capillaries.

If the wall of an alveolus is ruptured, as it may be in **pulmonary barotrauma** ("burst lung"), then air from the alveolus is able to enter the blood stream where it may cause blockage of distant vessels such as those in the brain. This is called an **air embolism**.

The lungs occupy a cavity about the size of a football on each side of the chest. The lung is covered by a thin membrane coating, called the **pleura,** and the inside of the chest wall is lined by a similar membrane. Between the two pleural layers is a narrow space which contains a small amount of lubricating fluid to minimise friction as the lungs expand and contract during breathing. If the outer surface of the lung tears, as it may in pulmonary barotrauma, then air can enter this pleural space causing the lung to collapse. This disorder is called **pneumothorax**.

The chest wall which encloses the lungs is made up of ribs with muscles between them - known as intercostal muscles. At the base of the chest cavity lies a large thin dome shaped muscle called the **diaphragm**. When the diaphragm contracts, it flattens and has a piston like effect, reducing the pressure in the chest cavity and increasing the volume of the lungs. The reduced pressure draws air into the lungs through the air passages.

Contraction of the diaphragm is the main method of **inhalation** in the resting state. It is assisted by contraction of the muscles between the ribs which rotate the rib cage upwards and outwards, enlarging the chest cavity and reducing the pressure in the chest. A group of neck muscles which are attached to the rib cage can also assist respiration when maximal breathing is required.

At the end of inhalation the elasticity of lungs and rib cage causes the lungs and chest wall to contract and **exhalation** takes place. With quiet breathing, this does nor require muscular effort. With heavy breathing, exhalation can be assisted by the abdominal and chest muscles.

Respiratory Function

During quiet respiration in adult males, about 500 ml of air is moved in and out of the respiratory tract with each breath. The volume per breath is termed "**tidal volume**". During extremely heavy exercise, the tidal volume can increase 10 fold, up to about 5 litres.

The total amount of air that can be held in the lungs (**total lung capacity** or **TLC**) in adult males is approximately 6 litres. Only about 10% of the air in the chest is exchanged with each breath during quiet respiration. The **vital capacity** (**VC**) is the maximum volume that can be exhaled in one breath, and the **forced expiratory volume** (**$FEV_{1.0}$**) is the maximum volume that can be exhaled in one second.

The **flow of air** through the respiratory passages varies at different stages of respiration. It reaches a peak about midway through inspiration — and during quiet breathing this peak flow rate is approximates 30 litres per minute. This value increases during exercise to 600–700 litres per minute.

Any breathing system (such as a snorkel or demand valve) which the diver is using, should be capable of handling these large air flows without significant resistance. If this does not occur, then the diver must exert extra effort during respiration in order to overcome this resistance. This problem is compounded when the diver is breathing compressed air at depth because the increased density of the gas will further increase the resistance to airflow in both the equipment and the lungs.

Gas Uptake and Loss

Air, which contains approximately 21% oxygen (O_2) and 78% nitrogen (N_2), is inhaled into the alveoli where it is brought into contact with the blood in the capillaries. This blood contains a lower partial pressure of O_2 than the air in the alveolus and a higher partial pressure of CO_2, since it has just returned from the body, which has been using O_2 and generating CO_2. Consequently, there is a pressure gradient causing O_2 to diffuse from the alveoli to the blood, and CO_2 to diffuse from the blood to the alveoli, where it is then exhaled. There is no net movement of N_2 since the N_2 in the alveoli and in the blood is in equilibrium, except when diving, altitude exposure or breathing different gases.

If the diver breathes air (78% N_2) or another inert gas such as helium, while descending or remaining underwater, this inert gas will pass from the alveoli to the blood because the partial pressure of the gas in the lungs is increasing as the diver goes deeper.

On ascent, the partial pressure of inert gas in the lungs will reduce, and this allows inert gas to move from the blood (returning from the tissues) to the alveoli, and be exhaled.

Respiratory Control

The partial pressures of CO_2 and O_2 in the blood are kept within very strict limits by a sensitive control system. There are sensors in the brain which detect small changes in the blood CO_2. If this increases, then the sensor causes stimulation of the **respiratory centre** within the brain, leading to faster and deeper respiration to eliminate more CO_2.

When a snorkel diver holds his breath, the CO_2 level in his blood increases. This produces respiratory stimulation which compels the diver to take a breath — hopefully after he has had time to return to the surface.

The sensors for blood O_2 pressure are in the carotid arteries which supply the brain. A reduction in the blood O_2 level also leads to respiratory stimulation, but this effect is not as powerful as that caused by CO_2 changes.

Smoking

The ingenious habit of rolling tobacco into a tube of paper, setting fire to it and inhaling the smoke, sabotages the complex respiratory and circulatory process at several points.

As well as predisposing to lung cancer and emphysema, noxious tars in the smoke precipitate out in the bronchi producing chronic irritation, narrowing of the bronchi and cause a persistent outpouring of mucous. This ultimately results in chronic bronchitis. The tar also poisons the cilia, which conduct the mucous up the airway to the larynx, resulting in retention of old mucous in the lungs (smell the breath!).

Various toxins in the smoke ultimately cause destruction of the alveolar walls producing cavities in the lungs and destruction of the lung architecture, resulting in the disease called **emphysema**. This, combined with obstruction of the air passages, makes the smoking diver less physically fit and more liable to air trapping in the lungs and pulmonary barotrauma (see Chapter 11).

The carbon monoxide content of the smoke reduces the capacity of the blood to carry O_2, thereby reducing oxygenation of the tissues.

Some of the chemical constituents of the smoke are absorbed into the blood stream producing changes in the walls of the blood vessels supplying the heart, brain and limbs. Ultimately these become obstructed. In later life this can cause heart attacks, strokes and peripheral vascular disease (gangrene).

CARDIOVASCULAR SYSTEM

Blood

Arteries take blood from the heart. **Veins** return blood to the heart. Arterial blood (which has absorbed O_2 as it passed through the lungs), is then pumped to the periphery by the heart and is brought close to all the cells in the body by the capillary system. Here the O_2 diffuses into the cells and the CO_2 diffuses out of the cells into the blood.

The blood transports O_2 and CO_2. The O_2 is mainly carried by an iron containing compound called **haemoglobin (Hb)** contained in the red cells. 100 ml of blood will transport approximately 20 ml of O_2. If the red blood cells are removed, blood plasma (the liquid part of blood) will transport only 0.3 ml of O_2 per 100 ml blood. A drop of blood contains approximately 300 million red cells.

In arterial blood, the haemoglobin is almost 100% oxygenated when the blood leaves the heart to go to the tissues. It is bright red in colour. If for any reason the arterial blood is not adequately oxygenated, it causes the blue colour of the skin and tongue (cyanosis) seen in **hypoxia** (see Chapter 20).

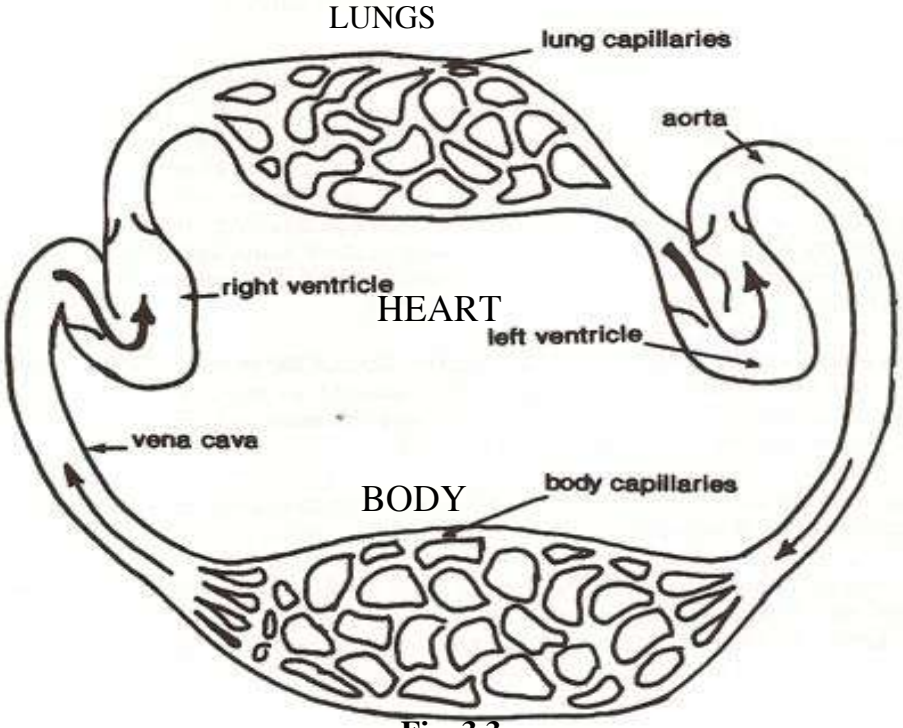

Fig. 3.3

A diagram showing the relationship between the circulations produced by the right and left sides (ventricles) of the heart.

In venous blood, the haemoglobin in red blood cells returns to the heart with 75% of its O_2 load still attached. It is then a more bluish colour.

The tissues need only 25% of the O_2 carried in arterial blood. This allows a reserve supply of O_2 which can be used during exercise or breath holding.

CO_2 is carried from the tissues by the blood in the veins, back to the lungs. Some of it is dissolved in blood plasma and some bound to the protein of the haemoglobin molecules. Although the CO_2 dissolved in the blood forms carbonic acid, the acidity of the blood is prevented from rising to excessive levels by a system of buffering compounds.

It is possible to increase the O_2 carrying capacity of blood by the use of **hyperbaric oxygen**. In recompression chambers, increased amounts of O_2 can be physically dissolved in the plasma, even though the haemoglobin is fully saturated with O_2.

Heart

The heart is a large muscular pump (about the size of a man's fist) located in the centre of the chest. See fig 3.4. It is composed of two functionally separate pumps which maintain two distinct circulations. The **right side** of the heart receives venous blood from the body and pumps this blood through the lungs where it picks up O_2 and eliminates CO_2. The **left side** of the heart receives this oxygenated arterial blood from the lungs and pumps it through the body.

Each side of the heart is essentially a two-stage pump which is not unlike a two-stage compressor. The **atrium** is the first or low pressure stage of the pump and it has a thin muscular wall. It receives blood from the veins at low pressure. When it contracts, it propels this blood into the second or high pressure stage – the more thickly walled and stronger ventricle.

The **ventricle** has two "one-way" valves, one valve preventing blood from flowing back into the atrium, and the other valve preventing blood flowing back into it from the arteries. When it contracts, it pumps blood into the arteries.

Occasionally there may be openings between left and right sides of the heart (patent foramen ovale, septal defects). In divers this allows bubbles to pass from the venous system to the arterial, causing serious manifestations of decompression sickness from dives that should otherwise be safe. People with significant heart abnormalities should not undertake scuba diving.

The heart, being a muscle, requires its own blood supply. This is provided by the **coronary arteries** which originate in the aorta, the main artery of the body. Any obstruction of these coronary arteries will cause damage to the heart muscle – a heart attack.

Partial obstruction of the coronary arteries may produce **angina** (which is pain or discomfort arising from insufficient O_2 in cardiac muscle), because it is receiving insufficient blood supply. Since a heart attack can take some of the fun out of a diving expedition, it is important for divers to have skilled medical examinations to exclude this problem or to help predict which divers will be susceptible to such heart conditions (coronary artery disease).

The resting output of the heart is about 5 litres of blood per minute. The heart has considerable reserve and if the tissues require it, can increase this output several fold by increasing its rate and strength of contraction.

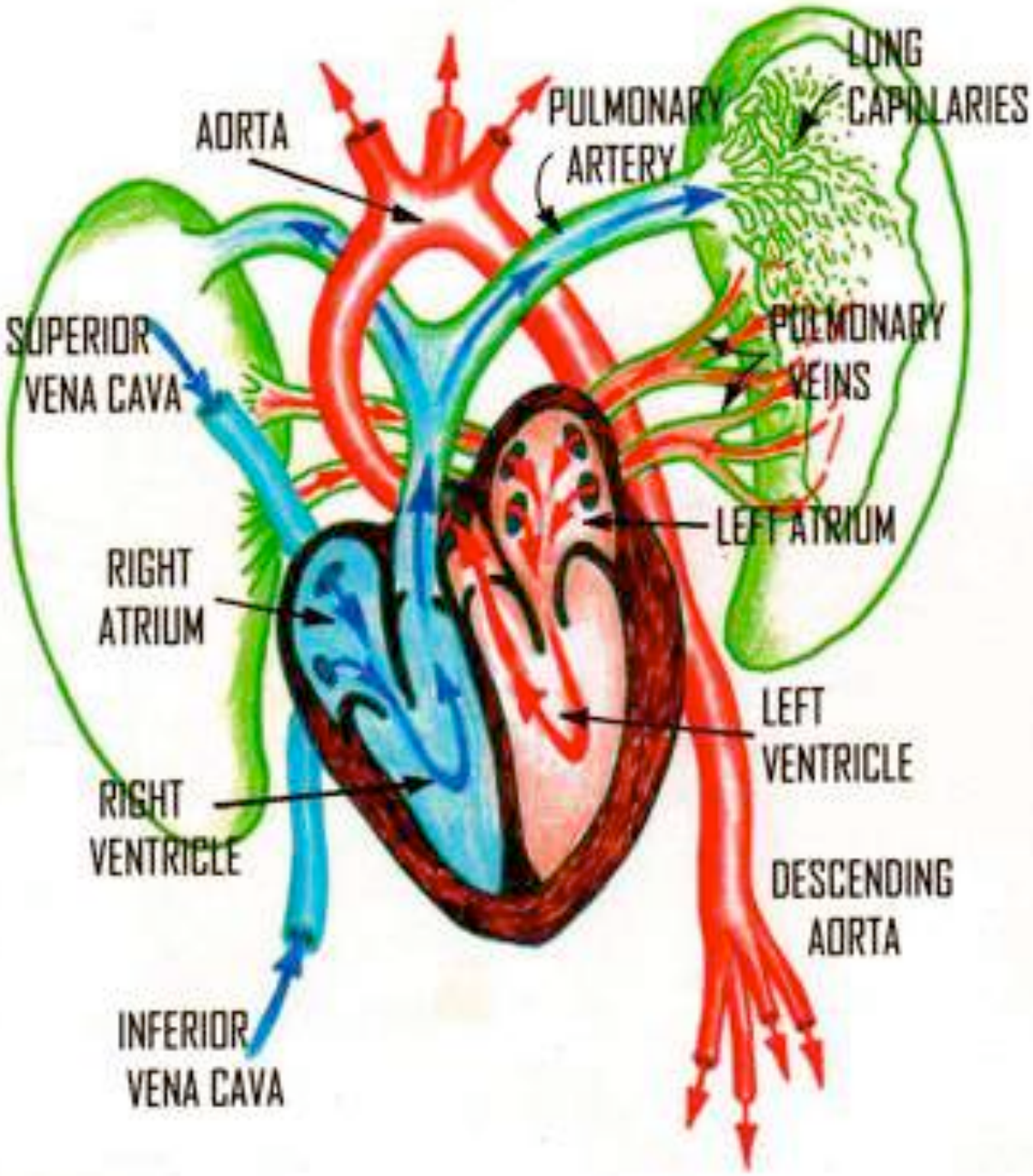

Fig. 3.4
This diagram shows a cutaway drawing of the heart to illustrate the flow of blood from the vena cava through the chambers of the heart and the lungs to the aorta.

Circulation

The blood flow from the heart is pulsatile and the **blood pressure** varies depending on the stage of heart contraction. The higher blood pressure during the heart's contraction is called the **systolic blood pressure** and has a normal value of around 100–140mm mercury (mm Hg). The pressure when the heart is not contracting is the **diastolic blood pressure** which has a normal value of around 60–90mm Hg. Blood pressure is normally recorded as Systolic / Diastolic – e.g. $^{130}/_{90}$.

Blood vessels can change their internal diameter under the control of the nervous system. This allows for some variation in blood flow to parts of the body depending on specific circumstances. For instance, during exercise the blood vessels dilate allowing more blood flow to the muscles, while under cold conditions the blood vessels to the skin constrict, reducing the blood flow to the skin (appearing pale) and so minimising heat loss.

The constriction or dilatation of the blood vessels also influences blood pressure. Excessively high blood pressure **(hypertension)** can ultimately cause damage to the blood vessels and an excessive strain on the heart. High blood pressure requires treatment, often with drugs which dilate the blood vessels but which may interfere with safe diving.

Blood pressure is constantly maintained by a sophisticated sensing and feedback mechanism. Variations in blood pressure caused by physical activity or standing from a reclining position are quickly compensated for by changes in the diameter of the blood vessel walls.

When a person is in a reclining position, blood pressure is maintained easily and the effect of gravity does not have to be opposed by the contraction of blood vessels. When standing up quickly from this position, blood pressure in the upper part of the body may fall. Occasionally, even in normal people, the heart and blood vessels cannot compensate rapidly enough and fainting or light-headedness can result. This is known as **syncope** or **postural hypotension**.

The cardiovascular system is able to compensate for changes in blood volume, such as those associated with severe bleeding (**haemorrhage**), by constricting the blood vessels and diverting blood from non-essential organs to essential organs such as the brain and heart.

In **pulmonary barotrauma,** air can gain access to the blood as it passes through the lungs. Air bubbles may be carried to vital organs such as the brain and heart, obstructing their blood flow and leading to serious consequences **(air embolism)**. In **decompression sickness,** gas bubbles may also be transported by the blood stream.

COMPRESSED-AIR DIVING

Scuba allows the diver considerable freedom but has its own limitations. It has all the potential problems of free diving, but adds special physiological problems of its own.

Resistance to Breathing

A major limitation to diving with scuba is resistance to breathing. During maximal exertion, a diver can consume over 70 litres of air per minute at the surface – but the peak flow rate during

inspiration is about three times this value. Some regulators may have difficulty delivering gas at this rate, adding considerable resistance to breathing.

This problem is magnified at depth because the greater pressure increases the density of the inhaled gas, especially at depths in excess of 30 metres when air is breathed. The same effect is seen at about 200 metres depth when helium /O_2 (heliox) mixture is breathed, because heliox is less dense than air. It is likely that resistance to breathing will ultimately limit the depth to which divers can reach.

An idea of the respiratory loads which the diver faces can be gained from the following table :

SCUBA SWIM	SPEED*	OXYGEN CONSUMPTION	RESPIRATORY MINUTE VOLUME
Slow scuba swim	0.5 knots	0.8 litres / minute	18 litres / minute
Average scuba swim	0.8 knots	1.5 litres / minute	28 litres / minute
Fast scuba swim	1.0 knots	1.8 litres / minute	40 litres / minute
Maximum scuba swim	1.3 knots	3–4 litres / minute	70–100 litres / minute

*. a knot is equal to 1 nautical mile per hour, or 1.85 km / hr

Table 3.1

Air Consumption

O_2 consumption is virtually the same for a given amount of exercise whether it is performed at the surface or deep under water. Because compressed air is being breathed at depth, more O_2 will be supplied than is needed by the diver. The actual volume of gas breathed at any depth will be the same as that which would be breathed at the surface. However, since the gas being breathed at depth is at greater pressure, the volume breathed, if converted to atmospheric (surface) pressure, will also be greater.

For example, during maximal effort a diver may consume 70 litres of air per minute at the surface. If he is performing an equivalent amount of effort at 20 metres depth (3ATA), he will still be breathing 70 litres per minute from his scuba regulator at 20 metres, but this will be equivalent to :

70 (litres) x 3 (atmospheres)
= 210 litres per minute at surface or atmospheric pressure.

So, the endurance of an air supply decreases with depth.

The regulator may not be able to meet the respiratory demands of a diver when certain conditions apply (see Chapter 5). Under these conditions, the diver may be aware of an inadequate air supply and either panic or take other dangerous action, such as a rapid ascent or omission of decompression requirements.

Skip Breathing

It is possible for a scuba diver to minimize his air consumption by deliberately slowing his breathing rate. This type of breathing pattern obviously limits the reserve of O_2 which will be stored in the divers lungs and haemoglobin, and may lead to retention of CO_2 and acidosis. It

reduces the safety margin in the event of air supply failure as well as increasing the likelihood of pulmonary barotrauma, and is recommended only for those with suicidal tendencies. Excess. CO_2 also acts as a narcotic, so the diver may make less sound decisions

Other Effects on a Scuba Diver

The physiological effects of scuba diving may parallel those of breath-hold diving (see Chapter 4). Hyperventilation, breath holding, the diving reflexes and the effects of immersion, may all be provoked. The dehydration effect of immersion is of importance in aggravating decompression sickness.

The many pathological problems developing from scuba diving are referred to later in this text. They include pulmonary barotraumas (see Chapter 11), respiratory decompression sickness (see Chapter 15), oxygen toxicity (see Chapter 21), breathing gas contamination (see Chapter 24), the drowning syndromes (see Chapters 25 and 26) and scuba divers' pulmonary oedema (see Chapter 32).

Gas Pressures

Because of the increased pressures on the diver, high nitrogen partial pressures cause nitrogen narcosis (or "narks"). Similarly, higher O_2 partial pressures may produce O_2 toxicity in very special circumstances. Gas coming out of solution in the divers body may cause decompression sickness and/or dysbaric osteonecrosis (or "bone rot").

Chapter 4

BREATH-HOLD DIVING

Free, breath-hold or snorkel diving is a prerequisite for successful scuba diving. The diver can thus become comfortable in the aquatic environment and gradually learn the swimming and snorkeling skills that may later be used in reaching dive sites and remaining safe even after the scuba set has been exhausted. One extension of snorkeling is free diving, used to spear fish, prolong underwater explorations, retrieve equipment, check anchors and many other activities .

It is not difficult for a diver to perform a breath-hold dive for a duration of one minute or more. This is possible because there is a reservoir of oxygen (O_2) stored in the lungs (about 1 litre O_2 when the lungs are full), in blood haemoglobin, and in myoglobin in the muscles.

With these reserves the diver is able to hold his breath for some time without the blood level of O_2 becoming dangerously low. Below a threshold blood O_2 partial pressure (about 30mm Hg – less than half the normal value), the brain ceases to function properly, causing loss of consciousness. At about this stage, the heart also becomes seriously starved of O_2 causing cardiac damage or disturbances of rhythm.

During a breath-hold dive, O_2 is consumed and carbon dioxide (CO_2) produced, decreasing the blood level of O_2 and elevating that of CO_2. Both effects may stimulate respiration but the CO_2 is the more dominant. Usually the diver develops an overpowering desire to breath (he reaches the **break point**) before the arterial O_2 level falls to a dangerous value. The urge to breath eventually becomes irresistible and the diver may even take a breath under water, if he is unable to reach the surface in time.

Breath-holding can be extended considerably, with experience and will-power but the break point is eventually reached. This is nature's safety mechanism to prevent people from losing consciousness from excessively prolonged breath-holding (see Case Histories 33.2 and 33.3).

ACCIDENTS and DEATHS

Breath-hold divers suffer from the same problems as scuba divers, except for those related to compressed gas inhalation. The common problems include environmental hazards, some equipment limitations and medical diseases such as the barotraumas, marine animal injuries,

infections, hypothermia, panic and fatigue, cardiac disorders and the drowning syndromes. These are discussed in the chapters dealing with scuba diving medicine.

In an Australian series of snorkeling deaths, the causes were dominated by drowning (45%), Cardiac disease (30%) and hypoxic blackout (20%). The first two were mainly in aquatically unfit, older tourists, and the last in younger, aquatically fit and experienced free divers. The inevitable predominance of time on the surface makes the breath-hold diver more susceptible to sunburn, boat injury and tidal currents.

The other problems more associated with breath-hold divers are discussed here.

Lung Squeeze (Pulmonary Barotrauma of Descent)

During a breath-hold dive the chest and lungs are compressed by the increasing pressure of water. As the air in the lungs is compressed, the volume is replaced to a limited degree by expansion and engorgement of the lung's blood vessels. Lung injury from this mechanism is known as lung squeeze, or pulmonary barotrauma of descent (see Boyle's Law, Chapter 2).

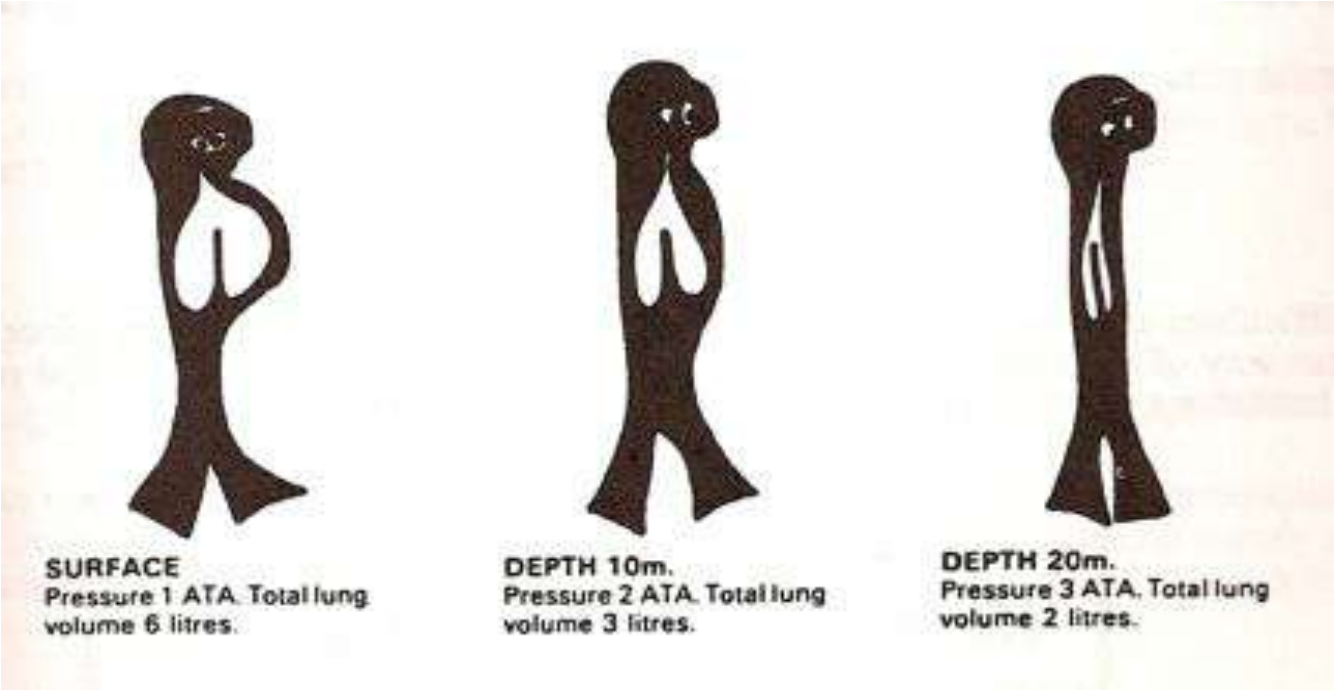

Fig. 4.1

Theoretically, the maximum safe depth for most divers should be about 30 metres (4ATA), but it probably varies between individuals, as much deeper breath-hold dives have now been performed - in excess of 200 metres.

Immersion

A neutrally buoyant diver is exempt from the main effects of gravity and this produces physiological changes in the body. The return of blood flow to the heart and lungs is increased. The body interprets this as an excess blood volume and compensates by **increasing urine production** (which may then lead to **dehydration**).

Cold water exposure produces many reflexes, including a desire to urinate. Temperature regulation is more difficult. The pressure variations may influence lung function with head out or vertical positions. Spatial orientation processes are disrupted. Trauma, in the form of physical injury from water movement, marine infections, dangerous marine animals, barotraumas, drowning etc. are dealt with in separate chapters.

Dive Reflex

Aquatic mammals display a reflex known as the "**dive reflex**". This is associated with profound slowing of the heart and redirection of the blood flow away from the muscles and non-essential organs to give a better blood supply to the heart and brain. It allows for longer submersions. This reflex is present to a less degree in humans and can be produced by immersing the face or head in cold water. The heart slowing component of the reflex has been used by physicians to treat certain cardiac disorders associated with a rapid heart rate. It can also result in heart arrhythmias.

Other potentially harmful reflexes can be induced by cold, Valsalvas, breath-holding *etc*.

Hypoxic Blackout

This loss of consciousness, due to an inadequate supply of $\mathbf{O_2}$ to the brain, usually develops without any warning. Underwater this leads to aspiration of water and drowning. It is a frequent cause of deaths amongst breath-hold divers. Sometimes the diver arrives on the surface still alive but in a state almost unconscious, and with some brain damage. This is called LMC (Loss of Motor Control) and causes unsteadiness and clumsiness. It also may occur after the diver has been rescued while unconscious or semi-conscious. At other times it may result in dementia, severe muscle impairment, visual damage or epileptic convulsions. There are two main types of hypoxic blackout, although they can occur together.

a. Hypoxic Blackout due to Hyperventilation and Breath-holding

There are some people who find the flaunting of safety mechanisms an overwhelming challenge. They may be trying to swim or stay underwater for as long as possible (such as swimming the length of a swimming pool, or to impress their peers or girl friend). The break point can be delayed by **hyperventilating** (taking a succession of rapid deep breaths) before a dive. This reduces lung and arterial CO_2 so that it takes longer for the blood level to reach the break point during a dive. During this delay, the blood O_2 level may fall below that necessary to maintain consciousness and the diver may become unconscious without any warning. This is one cause of **Hypoxic Blackout**. This can occur at any depth, such as in a one metre deep pool. Using this method some divers have been able to prolong their breath hold dives for extended periods — until the body is found!

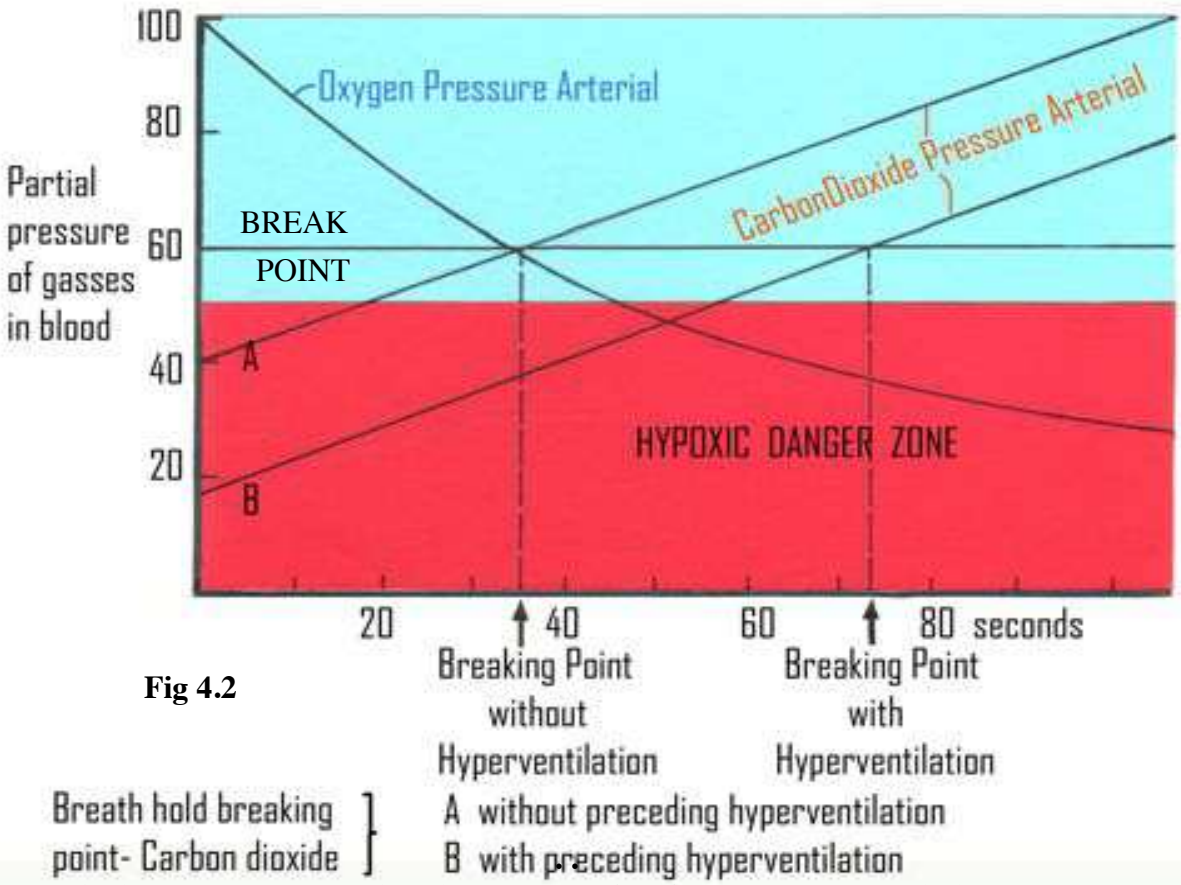

Fig 4.2

This diagram shows the relationship between the fall of oxygen and carbon dioxide levels in the blood with breath-holding. Normally, with breath-holding, (A) the breaking point is reached before the hypoxic zone is reached. After hyperventilation and breath-holding, (B) the breaking point is in the hypoxic zone.

b. Hypoxic Blackout due to Hypoxia of Ascent

Most divers will have noticed during a breath-hold dive that the desire to breathe often decreases with depth. This is probably due to the partial pressure of O_2 in the lungs increasing as they are compressed. There is a corresponding rise in the partial pressure of O_2 in the blood which will reduce the hypoxic stimulus to breathing. At depth the diver continues to exercise and use up his O_2 reserves. As the diver ascends however, the lungs will expand and the partial pressure of O_2 in them will correspondingly decrease. This produces an abrupt reduction in the O_2 partial pressure in the blood. It may fall below the threshold and cause unconsciousness during or immediately after ascent. This phenomenon is termed **hypoxia of ascent**. It may even occur after a diver has taken the first breath after surfacing, because of the delay between taking this breath and the replenished oxygen reaching the brain.

Some divers, especially those who are spear fishing, taking photographs or deep breath-hold diving for other reasons, will not only hyperventilate first, but then dive deep until they feel the necessity to breath (the break point). These are exposed to the dangers both of hyperventilation and breath-holding, and also to hypoxia of ascent. Death in such breath-hold divers is common, as unconsciousness occurs without warning.

Hypoxic Blackout - Prevention

1. Do not hyperventilate before breath-holding
2. Do not dive deep
3. Avoid excessive exercise
4. Avoid competitive behaviour resulting in the above activities
5. Ensure buoyancy near the surface
6. Dive with a buddy or have a safety observer/rescuer
7. Educate breath-hold divers of this hazard

Snorkel Diving

All the difficulties associated with breath-hold diving occur with snorkel diving – which is just breath-hold diving with a snorkel. Snorkel breathing is a convenient way of obtaining air whilst on the surface and with the head immersed, however it has several physiological and physical limitations due to the snorkel's structure (see Chapter 5).

MEDICAL CHECKLIST FOR SNORKELLERS

Have you ever had any of the following medical conditions :

1. Any cardiovascular disease? (Heart, blood pressure, blood, etc).	YES	NO
2. Any lung disease? (Asthma, wheezing, pneumothorax, TB, etc.)	YES	NO
3. Any fits, epilepsy, convulsions or blackouts?	YES	NO
4. Any serious disease? (Such as diabetes)	YES	NO
5. Serious ear, sinus or eye disease?	YES	NO
6. Any neurological or psychiatric disease?	YES	NO

Over the last month have you had any of the following :

7. Operations, illnesses, treatment?	YES	NO
8. Drugs or medications?	YES	NO
9. If female, are you pregnant?	YES	NO
10. Can you swim 500 metres without aids?	NO	YES
11. Can you swim 200 metres in 5 minutes or less?	NO	YES

NAME :

DoB :

ADDRESS :
(If under 16 years, guardian to sign.)

Note : If the candidate indicates an answer in the left hand column, then further investigation or action is required before snorkelling is considered as safe.

Chapter 5

DIVING EQUIPMENT

This chapter has been included to explain the operating principles and the limitations of some of the equipment currently used in free and scuba diving.

FREE DIVING EQUIPMENT

Mask

The reason we go diving is to view the spectacular underwater world. Unfortunately, our eyes are designed to see through air. In water the view is blurred and distorted – and somewhat magnified. The latter is the reason why divers exaggerate the size of the fish they see. Fishermen, who examine their catch on the surface, do not have such an excuse for their blatant lying.

The way we compensate for the water/air interface distortion, is to add an air space in front of our eyes. This can be achieved by air-filled contact lenses, swimming goggles or face masks. The last of these is used by divers.

The variety of face masks on the market suggests that the ideal mask has not yet been developed for the multitude of different face shapes.

The mask should cover the eyes and nose but not the mouth. Having the nose included allows the diver to exhale into the mask to compensate for the changes in water pressure, and so prevent **face mask squeeze** (see Chapter 12). The ability to exhale into the mask is also essential to clear a face mask flooded with water. The mask should be shaped so that the diver's fingers can reach and block his nostrils, to make ear equalising easier.

Ideally, the mask should have a small air volume so as to reduce the effort needed to equalise water pressure during breath-hold diving. The face plate should be close to the eyes, to maximise the field of vision. With a basic face mask this is limited by the nose. This problem could possibly be overcome by radical nasal surgery, but is best achieved by an indented rubber nose piece which allows the glass to be brought closer to the eyes. Clear plastic or glass side panels may also possibly help. Although this arrangement generally improves peripheral vision, the nose piece still restricts downward vision – obscuring the harness, weight belt and emergency gear.

Some face masks are fitted with a one-way exhaust or "purge" valve on the undersurface to aid elimination of water. This makes clearing the mask of surplus sea water much easier, but the valve can be an annoying source of water leakage into the mask, if it does not function correctly. Another source of leakage is a moustache or beard. These are still able to permit a seal if adequately wetted beforehand.

The body of the mask may be made of plastic, rubber, or silicone. The material needs sufficient rigidity to maintain the basic shape of the mask but a soft flanged edge (or skirt) is necessary to allow the mask to adapt to the contours of the face and provide a watertight seal. If the mask is excessively rigid it will not accommodate to water pressure changes, making face mask squeeze more likely. Masks made of silicone rubber are available for use – essential for divers with rubber or plastic allergy.

Fig. 5.1
Mask, snorkel and fins – basic free diving equipment

The viewing plate of the mask can be made of either glass or plastic. Masks, like the windscreens of cars, should be made of "tempered " or safety glass, thereby preventing shattering and damage to the face and eyes, in the occasional event of trauma. Then it will shatter into small cubes rather than sharp splinters when broken. Good quality plastic is less likely to break, but is prone to scratching

Fig. 5.2
Silicone masks

A mask can be chosen by fitting the mask to the face and gently inhaling through the nose. A mask which seals adequately will then adhere to the face without any air leak, and not fall off. The mask should seal properly without excessive tension from the mask strap. Some mask straps are broadened at the back of the head to distribute the tension over a greater area, as narrow bands are less secure and often cause local tenderness or headaches.

CAN YOU EASILY SEE?

Fig. 5.3

Diving masks with corrective lenses incorporated into or glued onto the faceplate.

For those divers who are myopic, they may wear soft or fenestrated contact lenses, or have negative corrected lenses built into the face plate. Lenses of differing refractive index are now available which slot directly into some masks, replacing the "blank" lenses. For the “oldies”, who need spectacles to read their gauges and see the miniscule aquatic animals and plants, they may stick a positive lens onto the face plate, low down and on one side only.

Snorkel

The snorkel allows the diver to breathe while floating, with the face submerged. Otherwise he would have to disrupt his view of the underwater world to turn his head to breath – like swimmers. It can also be used during periods of surface swimming before or following a scuba dive, to conserve compressed air or return to safety without relying on the scuba cylinder, which may be near empty.

Because of the limited strength of the respiratory muscles and the effect of water pressure, it is not possible to breath through a snorkel at a depth in excess of about 50 cm. The length of the snorkel should be sufficient to allow the diver to swim face down, to look around and to swim through choppy water without the snorkel flooding. It should not be excessively long as this increases breathing resistance and respiratory "dead space". The optimum length is about 30–35 cm.

Fig. 5.4
A range of snorkels. Some of these have "purge" valves for eliminating water.

The snorkel should be of maximum diameter to reduce breathing resistance but not wide enough to create excessive dead space. The optimum internal diameter is approximately 1.5–2 cm. It should have a minimum of angles and curves, and the interior should be smooth. Corrugated tubing or sharp angles increase breathing resistance. Mouthpieces are sometimes made to swivel and rotate in order to minimise drag on the mouth and permit a comfortable hold. The latter can be assisted by individual “bite moulding” of the mouthpiece.

The breathing resistance of a poorly designed snorkel is usually not noticed during quiet breathing, but may prevent the diver from exercising to his maximum capacity when needed. With moderately heavy breathing, as with anxiety-produced hyperventilation or swimming at >1 knot, snorkels restrict the breathing capability of divers engaged in surface swimming.

Several peculiar devices have been invented to prevent water entering the snorkel during a dive. These usually employ buoyant objects such as a table-tennis ball or cork which floats into, and obstructs the end of the snorkel when it is submerged. This requires an extra U shaped bend in the snorkel which increases resistance and dead space. These devices are unreliable and unnecessary, and often catch on other objects. Divers learn to expel water from the snorkel after returning to the surface by turning their head towards the side with the snorkel and blowing hard and fast.

Some snorkels are now fitted with a small purge valve near the mouth piece and which allows most of the water to drain from the snorkel automatically. It reduces the amount of water which needs to be expelled by the diver, and therefore the effort required. It is also a potential source of leaks.

If used excessively or gripped too tightly, sometimes the jaw becomes sore after a long dive. See Chapter 32.

Fins (Flippers)

The use of fins considerably improves the diver's swimming efficiency. There are several designs available.

Fins have two basic types of foot fitting — one has a shoe integrated with the fin (enclosed heel), and the other has a half shoe fitting and a heel strap (open heels) which allows the diver to wear neoprene boots. These boots can be used for protection when walking over reefs, and offer some thermal insulation to the feet.

The blades of the fins vary in size and rigidity and some types are fitted with vents (for a venturi effect). Studies of various fin types have shown that no one type is ideal for all divers. Fins with larger, more rigid blades, allow a more powerful action but require greater strength and are more difficult to manoeuvre. Muscular cramps can result from inappropriate powerful fins. In general, fins of medium size and medium flexibility are suitable for most recreational divers.

The way fins are used is important. Traditionally, a narrow, straight-leg kicking stroke has been taught. A less graceful wide-kicking stroke using bent knees is more efficient. This comes from directing the thrust of the fin along the direction of movement of the diver. Beginners may need coaching to avoid a bicycle peddling action, which is ineffective as a swimming stroke.

Fins with integrated shoes can often cause blistering and abrasions ("fin ulcers") on the foot or ankle, around the rim of the shoe fitting. This can be reduced by the use of socks until the diver becomes more accustomed to the particular fins. Correct fitting of the fins is necessary. If too loose, their loss will endanger the diver. An excessively tight fit may cause muscular cramps and fin ulcers.

New divers tend to rely on hand movements for propulsion. This is not effective underwater and they have to learn to employ fin propulsion

Fig. 5.5

Wet Suit

The wet suit provides protection, comfort and safety. It is made from rubber or neoprene which incorporates tiny air bubbles, to provide good thermal insulation. It also provides protection from scratching, abrasions and stinging animals.

On the surface and in shallow water, these suits give great buoyancy. To overcome this effect and therefore submerge, accessory weights are usually necessary. At depth however, the air bubbles in the wet suit are compressed (remember Boyles Law?), reducing its thickness, buoyancy and insulating properties. The variations in buoyancy at different depths may need to be offset by the use of a buoyancy compensator.

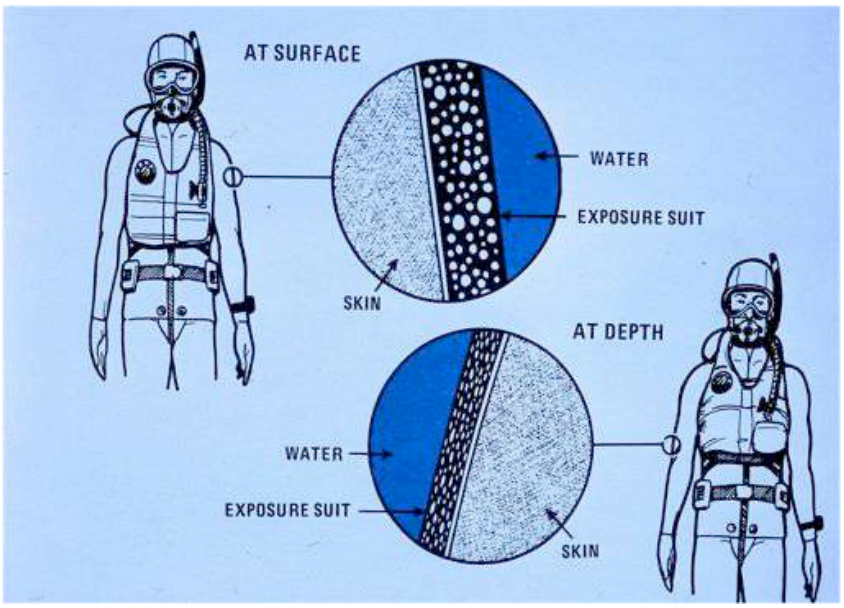

A poorly fitting wet suit can cause chafing, especially around the neck and arm-pit. A wet suit with an excessively tight neck can also compress the blood vessels to the brain leading to dizziness and fainting. Tightness around the chest may cause difficulty in breathing. Zippers allow for easier access and exit, but contribute to water leakage and reduced thermal reliability.

Fig.5.6

A wet suit variant containing inflatable gas compartments can partly overcome these problems. These suits can be inflated orally or directly from a scuba tank. Careful venting on ascent is necessary in order to prevent too rapid an ascent with these suits as the gases expand. They are a modern version of the traditional "**dry suit**". The latter is used in colder waters, when a layer of gas is injected between the rubberised suit and an insulating undergarment. A gas cylinder adds air into this space during descent, and the air is exhausted during ascent, to maintain neutral buoyancy. Because of the added buoyancy problems, special training is needed to use dry suits. See Chapter 12 for suit “blow up”. Urinating in dry suits is problemmatic, and the P-Valves that permit urinating into the ocean can cause rare but serious problems with retrograde flow of sea water, air and infections into the genito-urinary tract

There are **hybrid suits** that include characteristics of both wet and dry suits, and some include fluids and malleable solids that replace the gas spaces and avoid the variable buoyancy effects with depth.

Weight Belt

A weight belt is used to offset the buoyancy of the body, wet suit and other items of equipment. Ideally the diver should use enough weight to produce neutral buoyancy at the surface (without reliance on a buoyancy vest) or at a shallow depth, about 3-5 metres – where a safety stop is often indicated. The correct amount of weight is found by trial and error and this should be done in shallow water. As the diver descends, compression of the wet suit makes the diver less buoyant. This effect can be offset by the use of a buoyancy compensating vest (B.C.).

A diver without a wet suit will usually require less than 2 kg (5 lbs) weight and many divers will require no weight at all. A diver wearing a wet suit may require about 1 kg weight for each 1 mm wet suit thickness, with an extra 1 kg for neoprene booties or hood. Inexperienced divers tend to use more weights than experienced divers, and are therefore more at risk from buoyancy problems.

Fig. 5.7

The deliberate attachment of up to 10 kg of lead weight, or more, to an otherwise neutrally buoyant air-breathing creature in the water has obvious safety consequences. It aids in descent, but may impair the

ability to surface safely. "Lead poisoning" is a common contributor to recreational scuba diver deaths

Most weights are moulded lead shapes through which the belt is threaded. For comfort these are sometimes curved, and some newer belts incorporate zippered compartments filled with lead shot for better fit to the body. Weights are usually sold in 1, 2, or 3 kg. sizes.

Fig 5.8. Being over weighted is one of the most common faults of new divers.

Fig. 5.9
Sitting on the gunwale with a weight belt on, is living dangerously.

The weight belt should be fitted with a quick release buckle, preferably one which is separate from the scuba harness release. Exceptions to this requirement are found in saturation diving and in cave diving where a sudden ascent due to inadvertent release of the weight belt could have catastrophic consequences. The buckle should be easily identified by feel and therefore different from the harness buckle. The strap should not be too long or it will hinder quick release

The weight belt should fit firmly around the waist. If it does not, compression of the wet suit at depth may result in it becoming loose and rotating around the body, with the buckle becoming inaccessible. Some new belts are made from elastic material which conforms regardless of depth.

In a significant proportion of diving accidents, the diver fails to release the weight belt at the time of the emergency. Training of divers is required to ensure that release of the weight belt is routine in an emergency. When ditching the weight belt, the diver should release the buckle with one hand and hold the weight belt well clear of the body with the other, before dropping it — otherwise

entanglement with other equipment is possible. Unbuckelling the weight belt *per se* will not necessarily cause it to fall.

The attachment of weights to the diver using rope or an ordinary belt buckle which cannot be rapidly released, has sometimes proved more permanent than the diver would have wished.

The **weight belt should always be the last item of equipment put on before entering the water, and the first removed before leaving the water.** If this advice is followed, then an inadequately equipped diver who does fall back into the water, is more likely to float and not sink or drown.

Diving Knife

Fig. 5.10

Contrary to the popular Hollywood image, the diver's knife has limited usefulness in fighting marauding sharks. It is, however, an essential item of safety equipment which can be used to cut the diver free from entanglements such as rope, kelp, fishing lines and nets. Scissors may be more effective for this.

Although stainless steel blades resist rust, inferior quality steels do not hold a cutting edge well. The knife should be of robust construction and of a reasonable size. It should be strapped to the diver at a location where it will not cause snagging (e.g. the inner surface of the calf or arm), and easily accessible. It should not be attached to any item of equipment, such as the weight belt or scuba harness, which may be ditched in an emergency.

Spear Guns

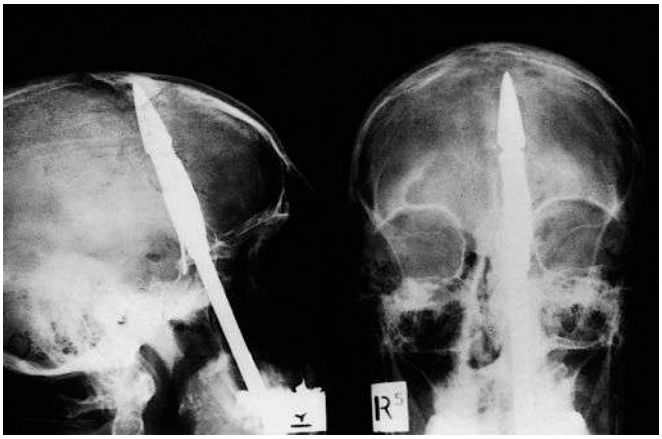

Fig 5.11

Although there is usually some value to specific bits of diving equipment, with inevitable problems accompanying them, the only piece of common diving equipment which can be universally condemned, is the spear gun. The senior author is so scared of divers carrying this equipment, that he departs the water as soon as one is observed. Sometimes the injury is to other divers, sometimes to the spear fisherman himself. The latter was the case in the fig 5.11. He managed to spear himself with the spear (seen as the white rod in the X-ray) penetrating his soft palate, the optic chiasm (nerves to the eye), the sinus and a lot of the frontal part of his brain (which he probably was not using, as he possessed this implement).

COMPRESSED GAS DIVING EQUIPMENT

The use of this equipment has given divers a high degree of freedom underwater and the capacity to go deep and stay there for long periods of time.

Strictly speaking, the term "**scuba**" refers to all **self contained underwater breathing apparatus** but these days it is generally restricted to open-circuit transportable air equipment only (initially called the "Aqualung"). With this equipment the diver breathes compressed air from a cylinder carried on his back, and then exhales into the water.

Other equipment used by divers includes **surface supply compressed air breathing apparatus** (Hookah or SSBA) and **rebreathing apparatus** (semi-closed or closed circuit). **Closed circuit** and **semi-closed circuit rebreathing apparatus** allows a diver to rebreathe some of his exhaled gas. It includes a chemical "scrubber" or absorber to remove exhaled CO_2. By re-using exhaled gas it makes economical use of the gas supply, as well as minimising bubble release into the water. They have obvious advantages for military, technical and commercial operations. See Chapter 43.

SCUBA

There are two basic forms of this system — the **twin-hose system** and the **single-hose** type. The twin-hose system is rarely used now.

The single-hose unit uses compressed air contained within a steel or aluminium cylinder ("**scuba tank**"). It is usually filled to a pressure of 150–200 Bar (2250–3000 psi). Some systems developed in Europe improve endurance by utilising cylinder pressures of approximately 300 Bar (4500 psi). New cylinders manufactured from alloy-mix materials permit greater pressures, and are smaller and lighter. In most countries, laws require that all cylinders are visually and hydrostatically tested every 1–2 years.

A **cylinder valve** fitted with a mechanical tap and connecting fitting is threaded into the neck of the cylinder. Standards require the fitting of a "**burst-disc**" to this valve so that this will burst before the cylinder in the event of overpressure.

A **"first-stage" pressure reducing regulator** attaches to the cylinder**,** usually by a universal screw-on or clamp fitting. This regulator reduces the pressure of the gas in the cylinder of 150–200 Bar to an intermediate pressure of 7-10 Bar (100-150 psi) above the local pressure and supplies air at this pressure to an **air hose** which passes over the diver's shoulder. The first stage regulator is thus designed to adjust the pressure in the air hose to the water pressure at the depth the diver is swimming (the environmental or "ambient" pressure). It automatically maintains this pressure differential as the diver changes depth.

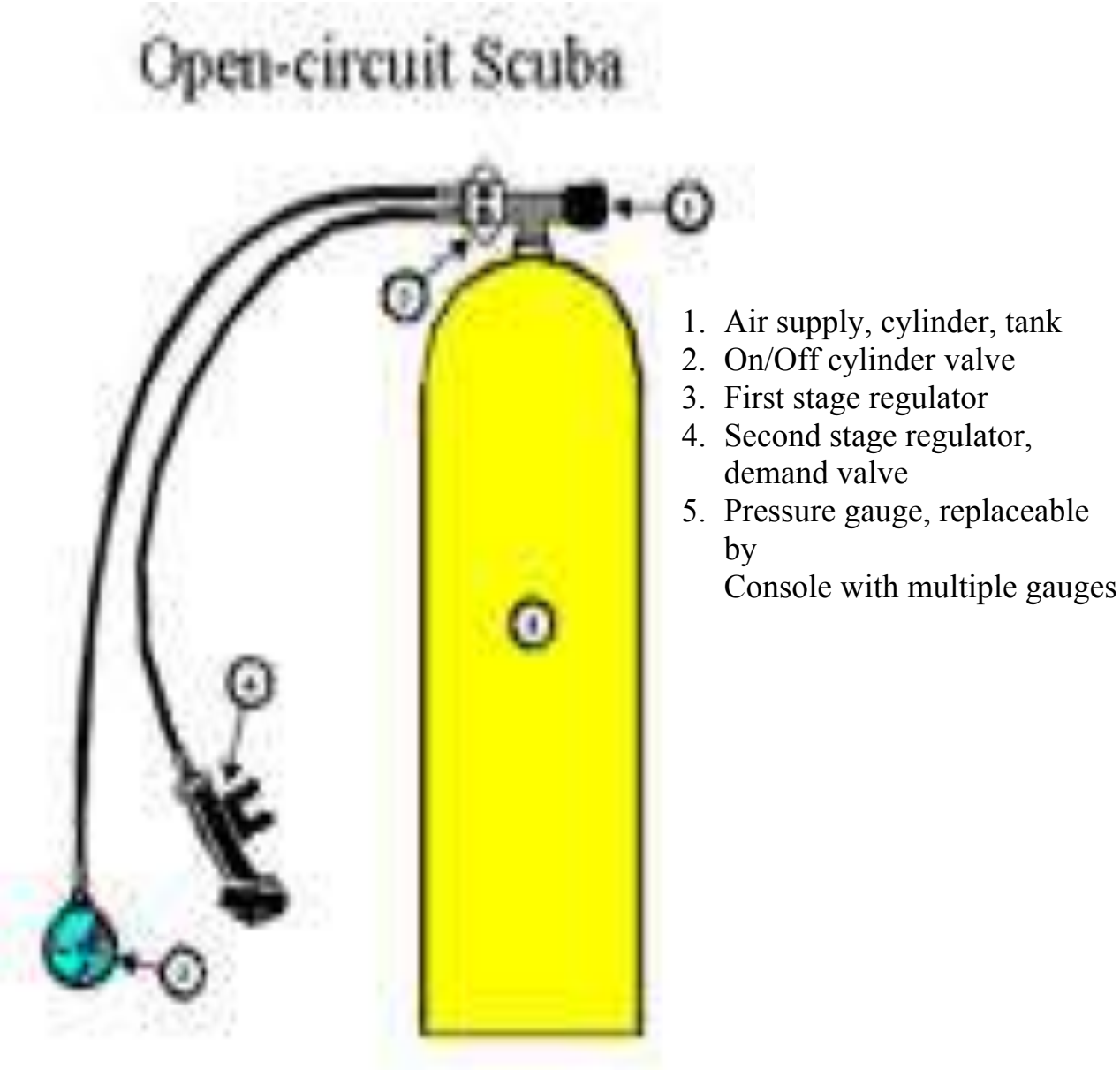

Fig. 5.12
A basic scuba set, without harness, console, octopus reg. or buoyancy compensator

The **air hose** is a small diameter flexible tube made of pressure resistant material which carries air from the first stage regulator to a **second stage regulator** (or **demand valve**) which in turn supplies air to the diver through the mouthpiece. With inhalation, a diaphragm moves to open a valve in the demand regulator, and air passes from the air hose to the diver, at the environmental pressure.

The diver exhales directly into the water through one or more one-way valves which should prevent water from entering the demand valve during inhalation.

It is important that the pressure of the air supply to a diver does not vary as the scuba cylinder empties, otherwise it will become progressively more difficult to breath as the tank pressure falls. Modern first stage regulators are much improved on the earlier models and have incorporated devices such as "balanced" valves to reduce this problem to some degree.

With most equipment, it is necessary for the diver to create a slight negative pressure in the mouthpiece during each inhalation in order to activate the demand valve mechanism. This negative pressure should be minimal or breathing becomes tiring. A regulator which is easy to breathe from at the surface, may not necessarily be able to deliver the large gas flows required during exertion at depth. When choosing a regulator, divers should refer to independent (e.g. U.S.Navy) testing.

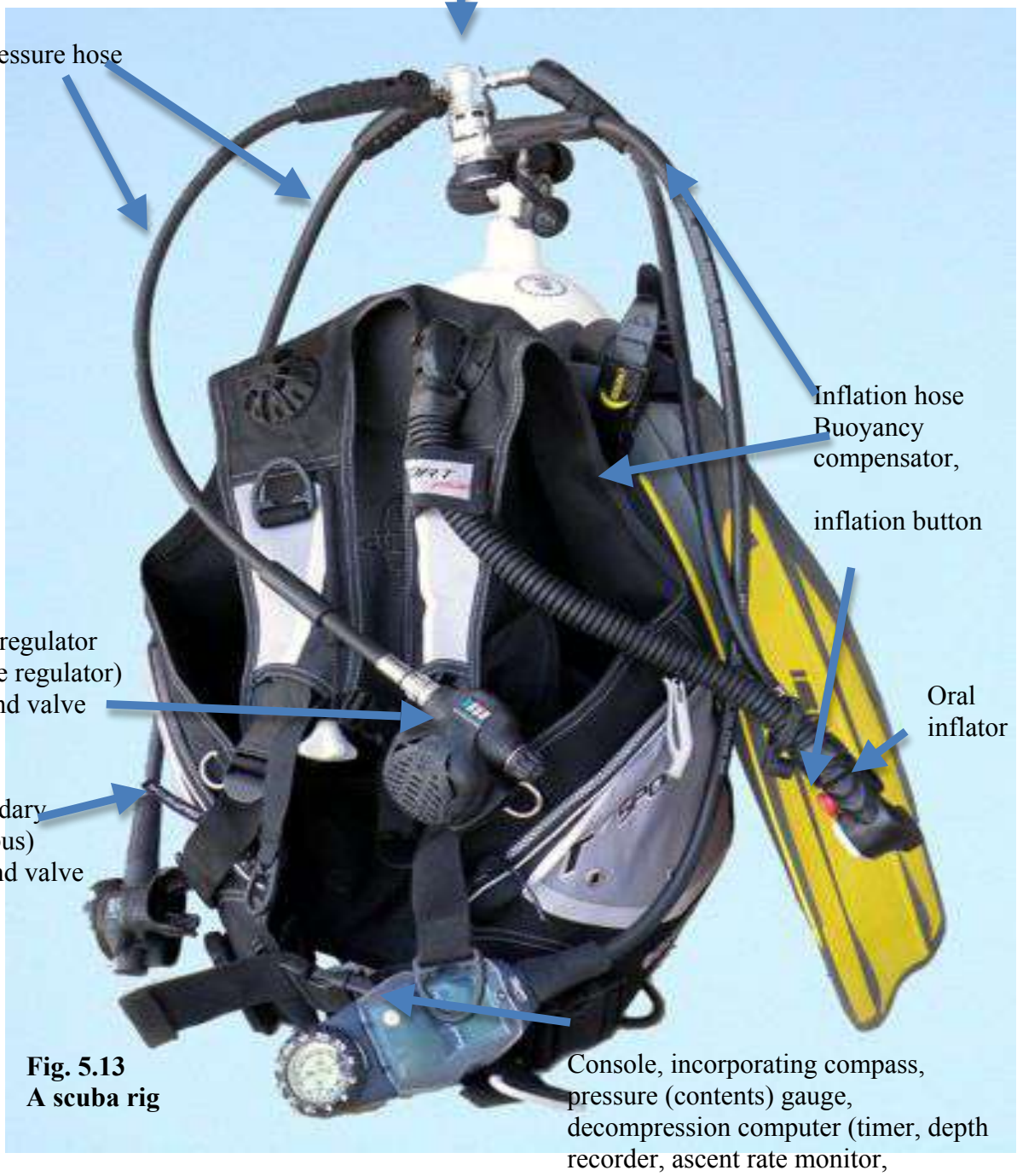

Fig. 5.13
A scuba rig

Difficulties are still encountered in obtaining adequate air supply with reasonable respiratory efforts under the following conditions :

- low cylinder pressures (observable on contents gauges), < 50 Bar
- cylinder valve not fully opened
- resistance in the first or second stage regulators (poor design or inadequate maintenance)
- increased respiration (exertion, hyperventilation, negative buoyancy etc.)
- at greater depths where the air breathed is more compressed (dense), > 30 metres
- with other demands on the air supply (inflating buoyancy compensator, octopus reg. etc.)

Some demand valves are bulky and quite heavy, requiring continual tension on the bite and the jaw to retain the mouthpiece. This can lead to painful spasm of the jaw muscles and a dysfunction of the jaw (temporo-mandibular) joint (see Chapter 32). Malleable plastic mouthpieces are available which attempt to spread the load evenly over the teeth. A soft silastic mouthpiece may be more valuable. Lugs attached to the mouthpiece are designed to keep the mouth open slightly, in a comfortable position and to locate and retain the demand valve correctly in the mouth. It should not be necessary to grip the lugs tightly.

Cylinder Valve

The gas outlet from the cylinder to the regulator is controlled by a high pressure valve or tap. High pressure "burst discs" are fitted by law to all scuba cylinder valves to minimise the risk of explosion if the tank is over-pressurised.

A common problem with divers is when they open the valve to check tank pressure, then close it to prevent accidental air loss *en route* to the dive site. The high pressure remains in the hoses and so the pressure gauge reads “full”. There is enough gas in the hose to permit a breath or two as the diver descends. Very soon he finds the supply of gas suddenly depleted, the pressure gauge then reads “zero”, or nearly so, and a rapid and embarrassing ascent is required.

Twin Hose Scuba

The twin or dual hose unit has both a first and second stage reducing valve combined in a single module attached to the cylinder yoke. Air is delivered by an intake hose to the diver's mouth at a pressure equal to the surrounding water. An outlet hose exhausts the exhaled air to the regulator for release to the water.

Since the diver's exhaust gas bubbles are released behind the head from the regulator, they tend not to interfere with vision. The twin hose apparatus has the disadvantage of requiring two bulky corrugated air hoses of around 2.5–3 cm diameter, and it is more difficult to purge the system of water. These units are rarely used today, except by photographers and in sites where regulators may freeze. The twin hoses were very prone to perishing and leakages.

HOOKAH and SSBA

Air can be supplied to the diver by a hose from the surface, either from a compressor (**hookah unit**) or from a cylinder or bank of cylinders (**surface supply breathing apparatus - SSBA**).

The air from SSBA is supplied directly to the demand valve at a pressure which is manually preset according to the depth at which the diver is operating. The first stage reducing valve (regulator) is located on the cylinder at the surface, and can be adjusted according to the diver's depth. This system can allow almost unlimited diving duration, which poses a risk of decompression sickness if the depth and time of the dive is not monitored.

Fig. 5.14
A hookah compressor and motor with capacity for two divers.

Fig 5.15
Surface supply

If the gas pressure in the hose from the surface fails due to a hose rupture, compressor failure or an empty cylinder, a pressure gradient can rapidly develop between the diver's respiratory tract and the failure site. Unless a non-return valve is incorporated in the gas supply line, near the diver, this pressure gradient can result in parts of the diver returning to the surface through his air hose. See Chapter 12.

Surface hookah units usually include a small pressurised reservoir as an emergency supply for breathing in case of compressor failure. Many divers carry small compressed air cylinders with them underwater ("pony bottles" or "bail-out bottles") which are able to be operated manually in the event of a main supply failure.

STANDARD DRESS or HARD HAT

This traditional piece of equipment uses compressed air delivered by a flexible hose to a rigid brass or copper helmet, usually connected to a heavy duty dry suit. The depth of the dive determines the pressure of the delivered air. A continuous air flow is supplied to the helmet at a rate sufficient to supply the diver's oxygen needs and to flush out exhaled gas. Originally hand powered compressors were used, later superseded by motorised compressors. A bank of compressed air cylinders can also be used, as with SSBA.

This system is bulky and requires heavy lead weights (usually boots and chest corsets) to offset the buoyancy of the helmet and the suit. Failure of the gas supply to keep up with the diver's rate of descent, or loss of the air supply (in the absence of a non-return valve), can lead to the diver being compressed into the helmet — causing head or body barotrauma. See Chapter 12.

Fig. 5.16
A "hard hat" or standard dress rig.

A modern variant of this system is used today in deep diving. It utilises a smaller, light-weight fibreglass or aluminium helmet or mask in conjunction with a dry or warmed wet suit, enabling the diver to swim and move more freely. The tethering line may go to the surface or a diving bell. The diver usually breathes gas mixtures which include helium, to prevent the development of nitrogen narcosis.

Fig. 5.17

A modern professional diving mask in rear with Standard Dress helmet in the foreground.

Closed and Semi-closed Circuit
REBREATHING APPARATUS

With this equipment some or all of the diver's exhaled gas is passed through a carbon dioxide absorber ("scrubber") and then rebreathed from a breathing bag ("counterlung"). This minimises gas usage, produces fewer bubbles and allows smaller cylinders to be used for an equivalent dive duration.

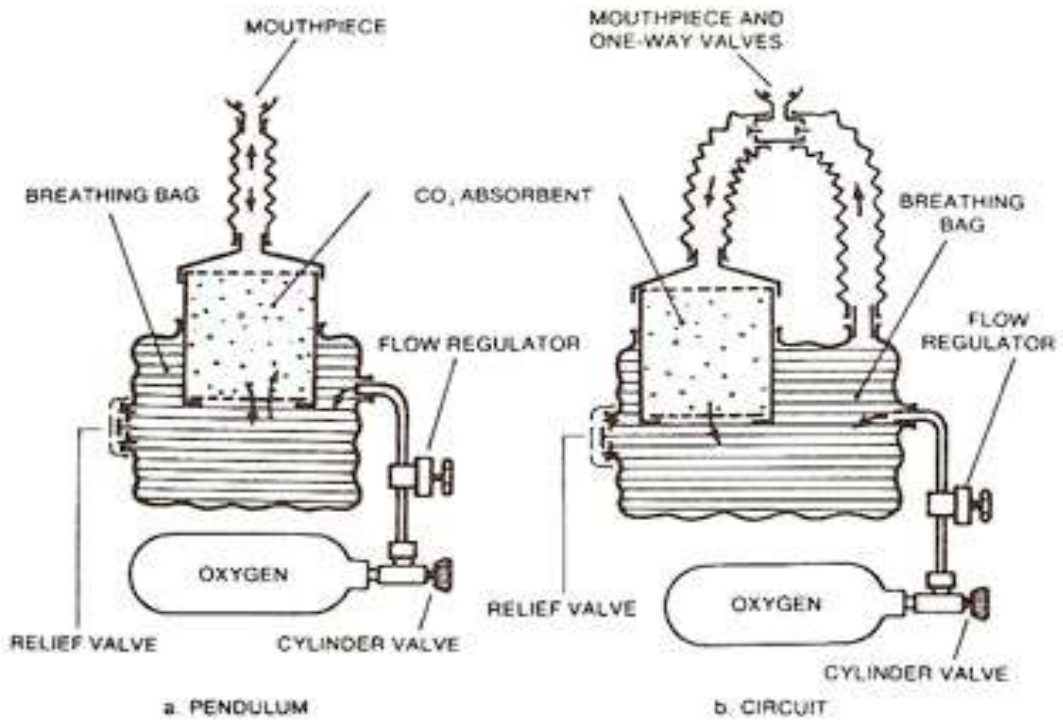

Fig. 5.18
Diagram of two types of closed circuit oxygen rebreathing sets.

A military system, using 100% oxygen in a closed circuit, is employed for clandestine operations (blowing up ships, examining potential landing sites, etc.). Because the diver rebreathes 100% oxygen, there is a risk of oxygen toxicity, so these sets have a practical depth limit of 9 metres.

Some of these sets have a demand-type system, where gas is supplied automatically when the volume in the counterlung is reduced. Others have a continual gas supply, with excess being exhausted.

Closed circuit mixed-gas rebreathing systems are used in technical and deep diving operations. These are further described in Chapter 43, but are not recommended for use by other than very experienced and meticulously trained divers.

Fig. 5.19 **Military diver wearing oxygen rebreathing sets.**

ANCILLARY DIVING EQUIPMENT

Buoyancy Compensator, Buoyancy Vest, B.C.

This device was originally devised as a modified life jacket to provide emergency flotation for the diver at the surface. Its value in compensating for changes of buoyancy due to wet suit compression with depth, was realised and it was modified to allow the gas content to be varied during the dive, depending on the buoyancy needs. It was also variously called a B.C.D. or B.C.V. (buoyancy compensating device or vest) or A.B.L.J. (adjustable buoyancy life-jacket)

Desirable features. When inflated the B.C. positive buoyancy should be sufficient to offset the negative buoyancy of the submerged weight of the diver and his equipment. It should support an unconscious diver so that his face is clear of the water. Ten kilograms (22 lbs) of buoyancy is more than adequate to achieve this. Most B.C.'s have excess capacity.

The B.C. should have a means of oral inflation, as well as a means of manually inflating with gas from a compressed air cylinder. With modern B.C.s the latter usually takes the form of an auxiliary "direct low pressure feed" line from the first stage or reducing valve. This direct-scuba-feed allows

the B.C. to be inflated using air from the scuba tank. This may provide insufficient or slow inflation with a low tank pressure at depth, especially if air is also needed for breathing.

Ideally there should be a separate emergency supply of inflating gas — this may be either a CO_2 cartridge or a small compressed air bottle. If a CO_2 cartridge is used it should have the ability to fully inflate the vest at depth, which usually requires at least a 20 gram capacity. The CO_2 cartridge triggering device is especially prone to corrosion and needs to be regularly maintained and inspected before each dive. The toggles which operate these cartridges can snag on passing obstructions, accidentally inflating the vest. This can have disastrous consequences in cave diving, saturation and decompression dives and other situations.

Some B.C.s are fitted with a small compressed air bottle for emergency inflation. This is activated by a rotating valve which will not open accidentally. The bottle can also serve as an emergency source for a few breaths of air if a modified demand valve is fitted to the vest. These bottles are usually charged from the main gas cylinder at the surface just prior to fitting the B.C.

The B.C. should have a pressure relief valve to prevent rupture from over-inflation on ascent. There also needs to be an easily accessible air-dumping valve to allow quick release of gas. The direct scuba feed line should also have an easily operated "quick-release" fitting at the B.C. end in case of a jammed inflator valve causing greater inflation than can be released by a dump valve.

The B.C. should be designed so that it will not ride up onto the throat when inflated. This was traditionally accomplished by fitting a crotch strap or attachment to the scuba harness. B.C.s are becoming increasingly more complex and expensive, and may contribute to diver errors and therefore injuries. Operating a B.C. requires repeated training.

Jacket B.C.s incorporating a scuba-tank backpack have become popular in recent years. These are comfortable and convenient to use, do not compress the chest and eliminate many of the straps associated with a traditional scuba-tank harness. With most of these units, however, it is difficult in an emergency to ditch an empty scuba tank on the surface without losing the B.C.

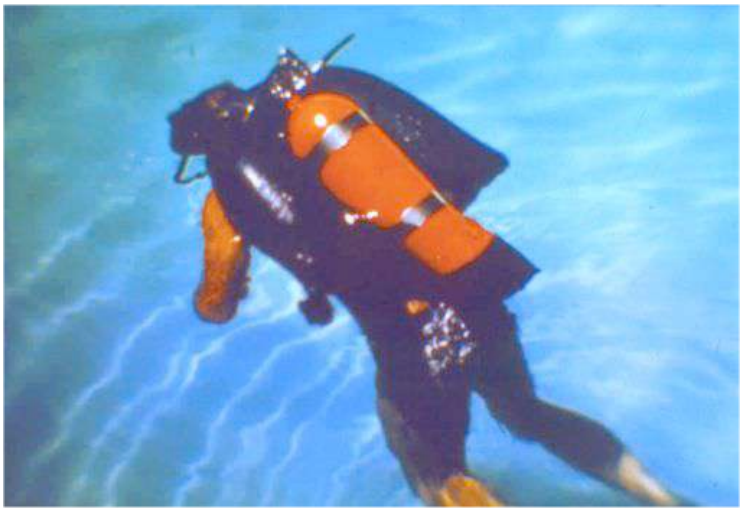

Fig. 5.20 This is the type of B.C. you do not want. One that tips the unconscious diver face down in the water

Contents Gauge

It is essential to monitor the air content of the scuba tank during a dive, to allow a sufficient air reserve for return to safety, emergency use and for decompression.

The pressure observed in the contents gauge overestimates the air available, because a substantial pressure is required just to drive air through the regulator. Thus something like 40 Bar should be deducted from the reading to calculate the remaining air available for the dive.

"Reserve" valves are not adequate substitutes for contents gauges since they may be inadvertently opened before or during the dive, and have been observed to leak or fail under operational conditions.

To gain maximum advantage from the contents gauge the diver should refer to it frequently, and should be aware of the values in respect to his own diving air consumption at that depth.

Fig. 5.21.
Depth and contents gauges (calibrated in feet of sea water and psig, respectively).

Alternate Air Source

The **octopus regulator** is a second-stage demand valve which can be used by the diver in the event of failure of the main demand valve, or which may be used by another diver who has an equipment failure or air exhaustion. The hose for the octopus or second reg. is longer than the primary reg so that it can be used easily by the OOA/LOA (out-of-air, low-on-air) diver. This facility eliminates the need for buddy breathing from a single demand valve, which can be difficult and dangerous to perform in high stress situations or between inexperienced divers.

Obviously, two divers using the same scuba system will halve the endurance of the tank. An alternative is to carry a complete separate emergency "**spare air**" unit with an adequate supply of air to reach the surface. At depth, and with a low tank pressure, insufficient air may be available for simultaneous use of the demand valve and the octopus regulator. Other alternative air sources include twin scuba cylinders (and independent regulators) and air breathing from a B.C. supply.

Fig. 5.22
A Spare Air unit

Diving Watch

A reliable, accurate, waterproof watch or dive timer is an essential piece of scuba diving equipment, in order that decompression requirements can be calculated.

The device should include some means of measuring elapsed time. A rotating bezel on the face of the watch is a simple and popular way of achieving this. It is not essential, but it is traditional, for divers to wear black-faced watches. Digital watches with elapsed time counters are also used.

Electronic dive timers, which are automatically triggered after a shallow descent, may not only record the dive duration but also the time between dives (surface interval).

Depth Gauge

It is necessary for the scuba diver to have an accurate knowledge of his depth exposure so that decompression requirements can be calculated. A depth gauge should be easily read under all visibility conditions. There are several types of depth gauge currently available. The simplest type uses an air-filled **capillary tube**. As the air in the tube is compressed during descent, water enters the capillary tube and the position of the water interface on a calibrated scale indicates the depth. This type of gauge is very accurate at depths down to about 10 metres but it is inappropriate in excess of 20 metres, due to the small scale deviations available on the gauge at these depths.

A **Bourdon tube gauge** incorporates a thin curved copper tube which straightens slightly as increased water pressure compresses the air within the tube. The movement of the tube is magnified by a gearing system which moves a needle across a scale. This type of gauge may become inaccurate due to salt obstructing the water-entry port, repetitive mechanical damage and altitude exposure.

Another type of gauge has a flexible **diaphragm** incorporated into the casing of the gauge. The diaphragm moves a needle through a magnifying gear system. This type of gauge has the advantage of relative simplicity and reliability.

Modern micro-processor technology has produced digital depth gauges which measure depth using a **pressure transducer**. This type of gauge is dependant on an adequately charged battery with water-tight integrity for reliable operation.

A device which records the **maximum depth** attained (maximum depth indicator or M.D.I.) is recommended as the diver may fail to note the greatest depth attained during a dive. This knowledge is necessary in calculating decompression requirements.

A depth gauge should be regularly recalibrated to ensure its accuracy. Some depth gauges incorporate a capillary depth gauge which will provide a cross check of calibration at shallow depth. Often depth gauges are contained in "consoles" which also contain cylinder contents gauges, timers and compasses.

Compass

Possibly one of the least appreciated pieces of equipment, until one needs it to navigate both underwater and on the surface.

Decompression Meters and Dive Computers (D.C.) (see Chapter 14)

A decompression meter or dive computer uses a mechanical or electronic model of the inert gas uptake and elimination by the diver. The dive computers (D.C. or D.C.M.) are based on decompression theories or algorithms (the principles on which the tables were developed) but often omit some of the safety factors incorporated in the formal tables. It is impossible for them to exactly duplicate the very complex gas uptake and elimination from a living diver and to allow for individual variation. They do however, accommodate the divers need to undertake both repetitive diving and multi-level diving in a much more manageable manner than the formal decompression tables.

Most current D.C.s also incorporate accurate devices for recording times, depths, ascent rates, cylinder contents and even water temperatures. Some provide "print-out" capabilities or connections to a computer. These enable accurate graphical representations of a diver's dive profile, and are useful to diving physicians treating cases of decompression sickness and to demonstrate where the diver went wrong. Unfortunately, sometimes it is the dive computer that goes wrong, not the diver.

Communication Systems

The safety of the buddy system of diving depends on the two divers being in constant communication. Divers who are not in constant communication are in reality only diving in the same ocean and may or may not be available to assist their buddy in an emergency. Even when they do, third party rescue is often needed. Buddy lines and buddy diving are discussed in Chapter 34.

Surface detection aids

The purposes of this class of personal equipment are to:

- allow the support boat to monitor and find divers on the surface during or after a dive
- prevent the diver being struck by boat traffic
- mark the diver's position when drift diving or while at the decompression stop
- help rescue services in lifeboats and helicopters to locate the diver

• Fig. 5.23
An inflatable safety sausage or "divers condom".

Surface detection aids include:

- Surface marker buoy, Decompression buoy, Delayed SMB, safety sausage or blob
- Red or yellow collapsible flag - high visibility, robust, bungeed to cylinder
- Glow stick - for night diving
- Whistle - cheap, will only be heard by people far from engine noise
- Torch/flashlight - if at sea after night fall
- Strobe light - needs long-lasting batteries
- High pressure whistle - expensive but effective
- Orange water dye - increases diver's visibility from search helicopters
- Mirror, such as a used compact disc, to reflect sunlight or searchlights
- Red or Orange Pyrotechnic flares -for helicopters and lifeboats
- Emergency Position-Indicating Rescue Beacon (EPIRB)

A **whistle** may be of value on the surface, in attracting support from the boat crew or other divers. Another system of drawing attention and demonstrating the divers position on the surface, where most accidents either commence or end up, is a depth-resistant **distress signal** (smoke for daytime, flare for night).

A 2 metre orange plastic tube, able to be inflated by scuba or mouth, is of value and is marketed as the **Safety Sausage**. If erect, it is easily seen from aboard boats. Aircraft can identify it more easily when it is laid flat on the water surface. It is also known as the "Diver's Condom".

Underwater a diver can be contacted by a variety of transmitting and homing devices. **Lights** are of real value at night, if the visibility is good.

It is a sad fact that most divers' bodies are retrieved only after a search — and usually death occurs without the buddy-diver's knowledge. Many deaths could possibly be prevented by the proper use of such simple and cheap systems of communication.

Most divers rely on diving close to each other, with visual communication only. Variations, such as one diver leading the other or diving with a group, results in an antithesis of the buddy system – as there is no clear and complete responsibility of each diver for the other.

A **buddy–line** keeps a pair of divers in close contact. It consists of a short length of cord (2–4 metres in length and preferably of floating line) which is attached to each diver's arm by a detachable strap. Any emergency affecting one diver will soon become apparent to the buddy even if he is not watching. Possibly the only instance where the buddy line should be discarded is when snagging is likely, or if a large shark takes a serious interest in one's buddy

Chapter 6

DIVING ENVIRONMENTS

When the aspiring diver is given his formal training, it is usually in a controlled and pleasant environment. He is trained to dive in that environment – not in other environments. He often has no concept of the demands and dangers of the other environments. He may subsequently encounter these dangers without an appreciation of the consequences of his actions, until it is too late. A perusal of this chapter may introduce him to some of the problems he could face and that need to be overcome. He needs specialised training and supervision when extending his diving profile to encompass these new environmental situations,

We will present only a brief overview. Reference should be made to diving manuals and texts for further details (see Appendix A). Divers are advised to obtain expert tuition from diving organisations specialising in these environments, before they contemplate venturing into them. Even then, they should be chaperoned by divers with specific experience of these environments

WATER MOVEMENTS

Tidal Currents

Currents of several knots* are commonly found in estuaries and ocean sites frequented by divers. These cannot be matched by the relatively puny swimming speeds achievable by a diver. For a short burst, a diver can manage about 1.5 knots, but a sustained speed of about 1.2 knots is the maximum which a fit diver can reach. For a relatively relaxed dive, a current of less than half a knot is acceptable. The problems posed by currents can be lessened by correct dive planning.

Firstly the diver needs to be aware that specific currents are a factor in any planned diving location. At other times they may be predicted by the tidal charts. This information is best obtained from local divers and/or maritime authorities. Currents can sometimes be identified by the behaviour of the dive boat at anchor, with the bows usually pointing into the current.

* 1 knot = 1 nautical mile per hour = 1.85 km per hour

The "Half Tank Rule". The best technique is to plan to swim into the current for the first half of the dive and use it to drift back to the boat during the second half. The dive is divided into halves on the basis of the air supply. After subtracting the air pressure needed to drive the regulator and as a reserve, allow half the remaining gas for the swim into the current and return using the second half of the supply. In tidal areas, it is necessary to anticipate any change in direction as the tide turns, or both halves of the dive may be into current. The best time to dive in tidal areas, for both ease and visibility, is usually during slack water - between tides.

As an example, if there is 200 ATA (or Bar) in the scuba cylinder, 40 ATA would be needed as a driving pressure for the regulator. That leaves 160 ATA for the dive. Eighty ATA could be used to swim into the current, and that leaves 80 ATA to drift back to the boat, allow for navigational errors and perform a safety stop. If one had allocated gas for emergencies, this would reduce the dive times accordingly.

The **anchor line** can be used to advantage. It is much easier to make headway against a current by pulling along a rope or chain, than by swimming. If a rope is attached loosely to the anchor line at the surface and run around the side of the boat, divers can enter the water holding onto this and use it to pull themselves to the anchor line. By using the anchor line, divers can pull themselves down to within 2 metres of the bottom, where the current is often less strong. Avoid swimming onto or dislodging the anchor, which can cast the boat adrift or lift and injure the diver. Swimming around the anchor allows the diver to check the anchor's security before continuing upstream.

Another rope (a **floating** or **Jesus line**) should drift with the current from the back of the boat, for 50 metres or more. This should be supported at regular intervals by buoys or floating plastic containers. This line has earned its name by "saving the sinners" who have missed the boat or surfaced down-current.

Another technique used in locations with strong currents is **drift diving**. Because of the fast currents, all equipment should be firmly attached and snagging on environmental hazards and other divers must be avoided. A **float** is towed to mark the diver's position and allow for signals to be sent to the surface craft. A rescue or pickup boat must drift with the divers and the current to another location, where the divers are hopefully recovered. Any such boat should have a **propeller guard** if it is to be used to rescue the divers.

Divers being left and lost at sea is unfortunately not a rare event, and it is often difficult for boatmen to find their divers if the sea surface is choppy (bubbles not detectable) and waves or swells block the diver from the low gunwaled boat. A divers float or a **"safety sausage"** (a long fluorescent inflatable plastic float) is a useful backup for a lost diver after an ocean or drift dive, and can be seen for a kilometre or so. A whistle can be used to attract attention, but is difficult to hear over engine and ocean noises. Other means of attracting attention of boatmen are low pressure horns (with >100dB sounds, for > 1km.), signal mirrors (if the sun is shining, for many km.), and waterproof **smoke flares** (lasting a minute or so, and visible for up to 10 km.). EPIRB and other electronic signalling devices may send emergency signals to commercial transports, including aircraft, over many km. See Chapter 5 for information on equipment.

Surge

In shallow water affected by waves, a **to-and-fro surge** which is too strong to swim against, may be encountered. This is best dealt with by gripping the bottom (with gloves) during the adverse surge, and moving with the favourable one. A diver contending with a powerful surge can become disoriented from the violent movement, injured from impact with rocks and can succumb to panic.

Surf

A surf entry without proper technique can be hazardous. The fully equipped diver presents a large vulnerable target to waves which can quickly divest him of essential equipment while engulfing him in unbreathable, non-buoyant foam.

The recommended technique of surf entry is to approach the water backwards after donning all equipment including fins before entry. The fins and mask must be firmly attached as they are easily lost and the regulator is attached to the vest by a clip in a readily accessible position. The diver watches oncoming waves over his shoulder while keeping an eye on his buddy, who is using the same procedure. Waves in shallow water should be met side-on to present the smallest surface area. The diver adopts a wide stance and leans into the wave. The diver should descend and swim while breathing through the regulator as soon as possible. Thus he avoids turbulence by diving under oncoming waves. After passing the waves break line, ascent allows reconnection with his buddy.

Floats are towed behind the diver on entry, and pushed in front when returning. Exit is achieved by the opposite process and by using incoming waves to help with progress towards the beach.

ENTRAPMENT

A variety of ropes, cords, fishing lines, nets, kelp and other material can easily snare the diver or his bristling array of equipment. Entrapment of this type can be safely dealt with by a calm appraisal of the situation and a sharp knife. Some divers prefer to use scissors (similar to wire cutters) instead of a knife, as they are more effective in cutting through particularly tough lines made of synthetic fibres. The limited field of view inherent with all face masks complicates these problems and makes the assistance of a buddy invaluable in tracing and untangling or cutting the causes of entrapment.

Kelp

This is a giant seaweed growing in forests from as deep as 30 metres and reaching the surface. It has a long trunk with branching fronds near the surface. It occupies cooler waters and provides a fascinating but potentially dangerous diving environment.

A diver can easily become entangled and drowned in kelp, especially near the surface where the fronds are thickest and special diving techniques are necessary for safe kelp diving.

Divers help minimize projections, which cause entanglement, by wearing the knife on the inside of the leg, use flush-fitting buckles and tape over protruding equipment. The scuba cylinder can be worn “upside down” to reduce regulator entanglement. The water is entered feet first and an attempt is made to push a hole in the kelp fronds, through which the diver passes. Divers should avoid twisting and turning in the kelp. A good kelp diver is a slow diver.The area near the bottom of the kelp causes the least likelihood of entanglement.

It is important to return to the surface with an ample reserve of air to ensure that the passage through the surface kelp is careful and unhurried. If entangled, be careful when cutting kelp stalks with a diameter similar to the regulator hose – you never know...

Enclosed Environments

Caves, wrecks, under-ice and even diving beneath large over-hangs are potentially hazardous environments which should not be entered without special training and planning. The following outline is by no means comprehensive. Specialised training and equipment are needed.

- **Caves.**

A diver in a cave usually cannot return directly to the surface in the event of an equipment malfunction or emergency. Even without these problems, it is easy to become lost and be unable to find the surface before the air supply is exhausted. The main problems are – panic, loss of visibility and navigational difficulties. The roof of a cave may collapse after air (expired from scuba) replaces the previously supporting water.

Caves are usually dark and lined with fine silt which is easily stirred into an opaque cloud by the use of fins. This is reduced with small fins, slow movements and avoiding the floors and roof. With silt, the natural or artificial illumination sources become valueless, reflecting the light back towards the diver.

Fig. 6.1

All essential equipment and lights are duplicated. A compass is mandatory. Cave divers carry a spare tank and regulator attached to a manifold, with the spare regulator on a long hose so it can be used by another diver following in a narrow passage, if necessary. Totally separate emergency air supplies are recommended.

Probably the safest diving equipment to use in caves, if not deep and the penetration distance not long, is a surface supply. Then return can follow the hose, which is being withdrawn by the surface tender.

Return to the entrance of the cave is marked by a line which is dispensed from a reel by the dive leader, who goes in first. The diver follows the leader into the cave.

This allows the way-out to be found by following the line, away from the leader. Vertical passage to the surface is marked by a heavier shot line which is less likely to entangle the diver ascending in haste.

- **Wrecks.**

Wreck diving shares many of the problems of cave diving (requiring similar preparations and precautions) as well as presenting some unique problems. In many areas the enduring wrecks are deep, adding the risk of decompression sickness and nitrogen narcosis to the general hazards.

Wrecks frequently contain physically or chemically unstable cargo, explosives and ordinance, toxic chemicals and unfriendly marine animals. Disturbed silt deep in a wreck and sharp jagged metal edges can make navigation through a labyrinth of ladders and passageways difficult. A compass may be of little help as the metal in the wrecks is often magnetised.

- **Ice diving.**

Diving under ice requires special equipment and know-how. It shares many of the hazards and precautions of cave diving but has the added complication of freezing conditions. Being trapped under ice can be an alarming experience for a diver with a frozen and therefore non-functioning regulator. Full reliance should not be placed in specialised "ice diving" regulators – in which the water is replaced by oil, alcohol or air. These can also freeze especially on the surface, using octopus regulators and with over-breathing. Attention has also to be paid to the exit procedure, as holes can "ice-over" rapidly. Protection may also be needed for surface tenders, as they may be exposed to wind and much colder temperatures than the diver, who is only at zero degrees C.

ENVIRONMENTAL VARIANTS

Cold Water

This can disrupt the performance of both the diver and his equipment. Diving in cold water requires the insulating qualities of a thick wet suit or dry suit, with gloves, boots and a hood. The wet suit, unfortunately, loses it efficiency when the insulating air layer is compressed with depth.

The cooling effect of compressed air expanding in the regulator, added to the low temperature of the water, makes freezing of the regulator a significant problem. Modified regulators which reduce these occurrences are available but cannot be fully relied upon.

Night Diving

This is not for everyone. The concept holds real fears for some divers who are perfectly comfortable diving in daylight. Because of the dangers of anxiety reactions and panic, night diving should be avoided by divers who are claustrophobic or feel excessively anxious at the prospect. The lack of visual cues can cause disorientation and imagination runs rife.

Lights well above the waterline, should be displayed on the boat and the shore exit. Torches should not be shone into a diver's face — it blinds him temporaribly (destroys night vision) — but they may be directed to display one's own hand signals.

The problems centre on impaired visibility. Vision is dependent on artificial light which is very restricted and can easily fail. It is important for the night diver to be able to find and use all items of equipment by touch alone.

Detecting and rescuing divers who develop problems and surface some distance away, may be difficult. An emergency flare, strobe light or chemical light stick (e.g."cyalume") attached to the diver's tank valve is worthwhile carrying for this eventuality, as is a whistle.

Deep Diving

Dives deeper than 30 metres have an increasing number of complications, possibly with inappropriate responses to these.

The endurance of the scuba air supply is severely limited at greater depths while the decompression requirements increases almost exponentially, adding a sense of urgency to the dive in the face of a diminishing reserve-air safety margin.

Decompression stops become obligatory for even short dives to depths in excess of 40 metres and requires the provision of extra air for this purpose. Unfortunately, the decompression tables, even if followed exactly, become less reliable as the depth increases, raising the possibility of serious decompression sickness even after a faultlessly executed dive.

Nitrogen narcosis can occur at less than 30 metres (100 feet) and progressively impairs judgement, attention, perception and an appropriate response to adversity as the depth increases. At depths in excess of 45 metres (150 ft.) mental stability, cognition and judgement are seriously impaired. See Chapter 18)

Equipment becomes more difficult to manage at these depths. Breathing through the regulator becomes harder. The buoyancy compensator takes much longer to inflate and uses more of the limited air supply. Wet suit compression reduces its insulating properties at the same time as the diver passes into colder deep water. This compression also progressively decreases buoyancy.

The environment beyond 30 metres is dark, colourless, cold, relatively devoid of marine life (although the fish and sharks are often larger), and replete with physiological hazards. In spite of this, some recreational divers feel compelled to experience it, albeit briefly because of the limited air endurance.

The authors recommend that, in view of the increased hazards and the limited diving satisfaction available in deep dives, recreational divers regard 30 metres (100 ft.) as the maximum recommended safe depth. Uneventful dives beyond this depth often impart a false sense of capability – which is then shattered when one or more things go wrong. It is then that the effects of narcosis are demonstrated.

Altitude Diving

Diving in waters located above sea level (e.g. a mountain lake or dam) introduces some potential hazards which are related to the cold temperatures at altitude and buoyancy problems with fresh water (see below). Other variations with altitude are much more important, but not immediately obvious.

Consider a dive in a mountain lake where the atmospheric pressure is half that at sea level (this would be at an unlikely altitude of about 6000 metres or 18,000 ft. elevation, but it makes the calculations easy). The pressure at the surface of the lake is that of the atmosphere, 0.5 ATA. Assume it is a salt water lake (fresh water is slightly less dense and so exerts slightly less pressure at a given depth).

The water in the lake will exert the same pressure at this altitude as it would at any other altitude.

That is, 10 metres of water will still exert a pressure of 1 ATA.

The pressure at 5 metres depth therefore will be 1 ATA, consisting of 0.5 ATA contributed by the atmospheric pressure, and 0.5 ATA contributed by the water.

The pressure at 10 metres will thus be 1.5 ATA.

One might think initially that this would give the diver a safety margin since the pressure at a given depth in a mountain lake is less than that in the ocean. The critical difference, however, is that the diver in the lake is returning to a lower surface pressure.

This can be illustrated by referring to one of Haldane's hypotheses (see Chapter 13). He indicated that a diver could spend an unlimited time at 10 metres (2 ATA) and return to the surface (1 ATA) without developing decompression sickness. In other words, a diver could return to a pressure of half the original pressure (i.e. a 2 : 1 ratio) without developing nitrogen bubbles in the tissues.

In the mountain lake, because the surface pressure is only half that at sea level (0.5 ATA), the diver need dive to only 5 metres (1 ATA) and return to the surface to encounter the same 2 : 1 “safe” ratio. A 10 metre dive exceeds the “safe” decompression ratio. This makes dive tables designed for sea level unreliable at altitude unless considerable corrections are made.

Decompression at altitude is further complicated by difficulties in estimating depth. Digital electronic gauges must be calibrated for altitude.

A mechanical depth gauge calibrated for sea level is likely to be unreliable at altitude. The gauge simply measures pressure and registers this as depth. Since the pressure at the surface of the lake is 0.5 ATA (half that of sea level), the gauge will be straining its mechanism and possibly bending the needle, trying to get its pointer past the zero stop to register what it interprets as negative depth. The gauge may only start to register a depth after it has returned to 1 ATA. This would not happen in the mountain lake until the water pressure and atmospheric pressure added up to 1 ATA – a depth of about 5 metres.

Even a capillary depth gauge, calibrated at sea level, will not really read accurately. At sea level, the air-to-water interface in the capillary will move half way along the capillary at 10 metres, since the pressure there is twice that at the surface. In the mountain lake with a surface pressure of 0.5 ATA, twice the surface pressure will be encountered at about 5 metres depth. So the capillary gauge will reach the "10 metre depth mark" (the half volume mark) at 5 metres.

The lower surface pressure also means that gas volume changes with depth are different. The gas in a diver's lungs will double in volume between 5 metres and the surface in the lake, instead of between 10 metres and the surface as would occur in the ocean. Ascent rates thus need to be reduced if the risk of barotrauma is not to increase. The gas expansion in a buoyancy vest will also be greater near the surface in the lake, which can lead to buoyancy changes unexpected by a diver used to ocean diving. Hyperventilation is more likely at altitude because the air is less dense, and over-breathing the regulator is more likely.

Flying after Diving

This creates some similar problems to altitude diving. The decompression tables were calculated on the assumption that the diver would be returning to a pressure of 1 ATA. If the diver then goes to altitude either in an aircraft of on a high mountainous road, with nitrogen still in his tissues, bubble formation is more likely because of the lower pressure experienced, and existing bubbles are liable to enlarge. Special "post diving flying rules" apply.

Diving in Freshwater and Dams

Buoyancy is less in freshwater than saltwater. Depth estimations and calculations are similarly disrupted (10 metres of seawater = 10.3 metres of freshwater). Freshwater is often still, and therefore develops dramatic thermoclines. Trees and other sources of entanglement tend to accumulate and not be destroyed as rapidly as in the sea. Some freshwater currents may cause difficulty. Chemical and sewerage pollution can be a major problem, and some specific freshwater organisms are very dangerous (e.g. *Naegleria* causing amoebic encephalomyelitis).

Dams have a specific problem with outflow below the surface. A diver may be unaware of the pressure gradient that can develop if part of the body covers an outflow orifice. Such a gradient will tether the diver underwater and may cause grotesque injuries as it forces the diver into and through the opening.

Conclusion

In this chapter, we have reviewed only some of the problems of diving in various environments and only some of the measures that can be taken to reduce the dangers. It is hoped that the reader will review this information whenever he is invited to join a diving activity different to one for which he has been trained. He may then be encouraged to undertake more specific training appropriate to his proposed dive program.

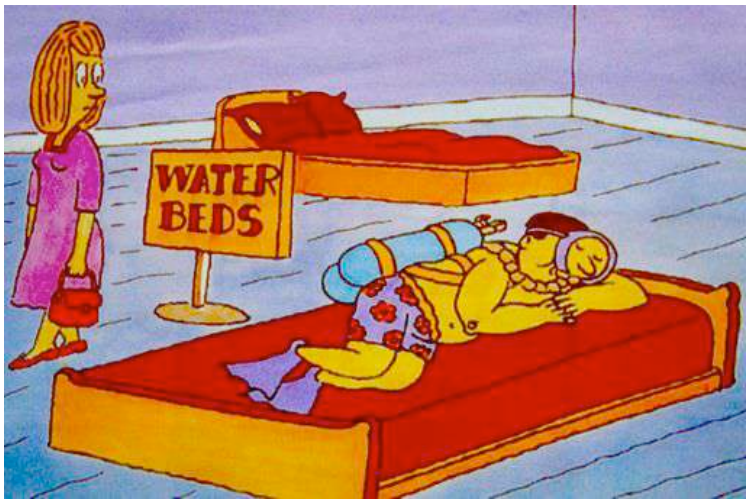

Fig 6.2 For some divers, there are few safe environments

.

Chapter 7

STRESS DISORDERS, PANIC & FATIGUE

INTRODUCTION

The diving equipment and facilities of the 1950's and 1960's were often spartan and the divers were to some extent influenced by the difficulties produced by this equipment and the environment. Diving was not easy, safe or comfortable and only the dedicated few were involved in the sport. The divers who survived tended to be capable and well adapted to the environment. They were “water people”.

The advent of more user-friendly equipment, together with the marketing and general popularity of diving, has seen the introduction into the sport of some divers who are less naturally suited to the environment. These divers may be more prone to **stress syndromes** when confronted with some of the threatening aspects of the marine world.

Some of the factors influencing the divers ability to cope with the diving equipment and environment will now be considered.

PERSONALITY FACTORS

Some personalities are better suited to scuba diving than others.

Military diving requires exacting physical and psychological standards and this is reflected in the high failure rate, generally about 50%. Many professional diving courses have a similar requirement and failure rate. This prompted researchers to look at the personality characteristics of successful trainees in an attempt to select out those who were not suitable.

In general, successful military divers were psychologically stable, not anxious about the dangers of diving, intelligent, practical, physically fit, self sufficient, good swimmers, capable and confident in the aquatic environment.

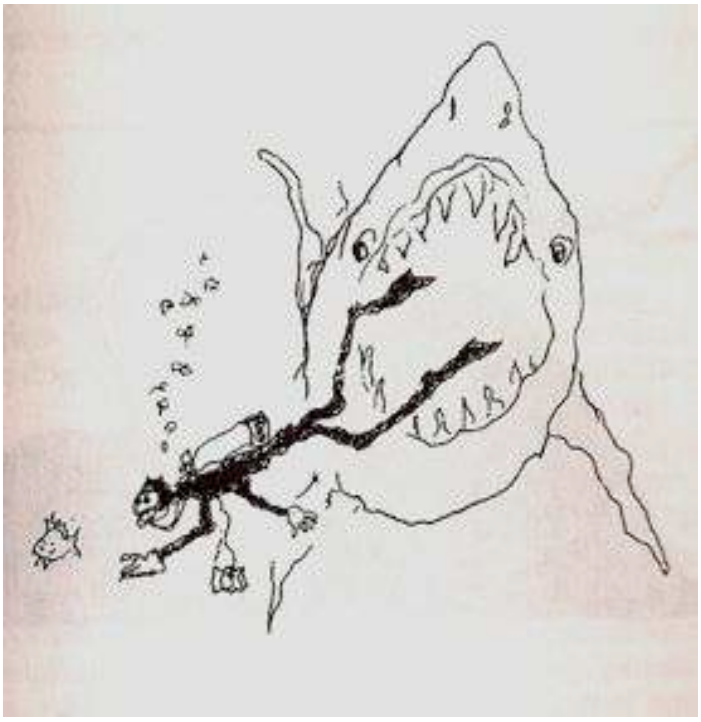

There is little data available for features which characterise successful and safe recreational divers. While the exacting requirements of a military diver probably are not as necessary, it would seem likely that similar characteristics would be shared by the most competent divers of both groups.

Fig. 7.1

Although there is a high failure rate with military and commercial diving courses, the failure rate in many recreational diving courses is close to zero. The standards set by some diving organisations is a source of concern, as it is possible that they may be overly influenced by commercial considerations. Between 5–10% of deaths in recreational divers occur while under training. Possibly this is the result of a combination of unsuitable people being trained in an unsafe manner. It is uncommon for a trainee to accept that he is not suited to diving, or for a diving organization to admit to less than ideal teaching methods.

STRESS RESPONSES

We all have an inbuilt automatic response to threats in the environment. This involves activation of the nervous system which prepares the body to confront the challenge or flee – the so called "**fight or flight**" response.

When this response is triggered, the sympathetic nervous system releases adrenalin into the body, stimulating the heart, increasing blood flow to the muscles, alerting the brain and increasing respiration. For example, a person suddenly confronted by a mugger is automatically primed to fight or run away. If the mugger is armed, the sensible victim usually considers the safest option is to quietly hand over money. This is an intellectual decision appropriate for survival and overrides the autonomic response. Logic can over-ride emotion.

Some divers may respond to certain levels of stress in ways inappropriate to survival. These potentially dangerous stress responses are :

- **Panic –** a psychological stress reaction characterised by excessive anxiety
- **Fatigue –** a physical stress response to exertion
- **Sudden Death Syndrome –** a lethal cardiac response to stress (see Chapter 35).

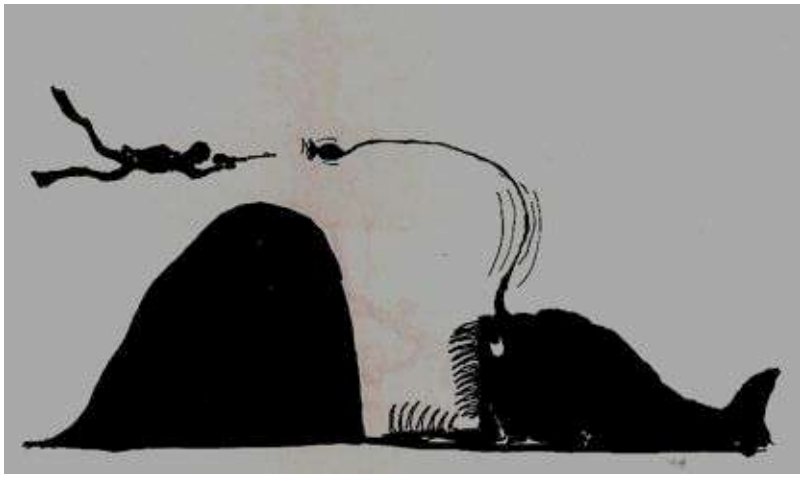

Fig. 7.2

PANIC

Panic is probably the most common contributor to death in recreational scuba diving. Studies have implicated panic as a contributor to between 40–80% of such diving deaths.

Fig 7.3

Panic is an extreme and inappropriate response to a real or imagined threat. Behavioural control becomes lost. Some readers will have experienced, or been near to, panic in some real life situations.

In general, the more naturally anxious a diver is, the more likely he is to panic.

As panic develops, the capacity to think rationally and solve the emerging problem deteriorates. The diver becomes more and more narrow minded and eventually may focus on only one goal e.g. reaching the surface – to the exclusion of other vital factors, such as exhaling during ascent.

Factors which upset a diver's emotional equilibrium can contribute to panic. Some of these **contributing factors are tabulated as follows**:

Personal Factors	**Equipment Problems**
Fatigue	Buoyancy
Physical unfitness or disability	Snorkel
Previous medical disorders	Face Mask
Seasickness and/or vomiting	Weight Belt
Alcohol or drugs	Wet Suit
Inexperience	Scuba Cylinder
Inadequate dive plan	Regulator
Dangerous techniques e.g. buddy breathing, free ascents	Low or Out-of-Air Situations
Psychological problems e.g. excessive general anxiety, phobias	Other Equipment faults
Sensory deprivation – night diving, blue orb syndrome, solo diving	Excessive reliance on equipment e.g. B.C.s
Vertigo and/or disorientation	Loss of equipment e.g. face mask or fins
	Excess weights
	Entrapment in lines, nets, harness etc.

Environmental Hazards

- Tidal currents
- Entry or exit problems
- White water e.g. surf
- Kelp
- Caves, wrecks
- Ice and cold water
- Deep diving – nitrogen narcosis, rapid air consumption, reduced buoyancy
- Dangerous marine animals
- Poor visibility
- Night diving
- Explosives
- Boat accidents

Consider the following scene, which has been gleaned from several diving fatalities, to illustrate some of the factors contributing (in ***italics***) to a panic-related death.

Case history.

Harry was a recently qualified diver who had ***borrowed*** equipment to undertake an open ocean dive in an ***unfamiliar*** area. His borrowed wetsuit was a little ***tight*** around his chest, restricting his breathing.

He decided to use two ***extra weights*** on his weight belt to help him descend in the ocean conditions, which were somewhat foreign to him. He was ***inexperienced*** at open ocean diving and the conditions were regarded as marginal so he felt a little ***uneasy*** about the dive.

His companions were more experienced than him and he was unsure of his ability to make his ***air supply*** last as long as his buddies. After all, he did not want to be the first to run out of air and force his buddies to shorten their dive.

During the dive he was sure he was using more air than the others but he had no way of checking this as his borrowed scuba set did not have a ***contents gauge*** .

He became a little more ***apprehensive***. They seemed to have swum a long way both from the dive boat and the shore. But he did not want to ***inconvenience*** his buddy or ***embarrass*** himself by ascending and checking his distance from shore or inquire about his buddies air supply. He had no idea how much air he had left but he felt that there probably wasn't much.

He became a little more ***anxious*** and his ***breathing rate*** increased. He noted some ***restriction*** to breathing. Was this just ***resistance in his regulator*** or was he now running out of air?

He activated his reserve valve. Perhaps this would improve the restricted gas flow. It didn't.

There was a tidal current running, which slowed their progress to the planned end of the dive – the safe exit point.

He was hoping that his companions were also running ***out of air***, as he appeared to be.

He was becoming more ***anxious***. His heart was pounding and his ***breathing rate*** was increasing. It was becoming harder to get sufficient air from his demand valve.

The difficulty in obtaining enough air settled the matter. He decided to get to the ***surface***, fast. In spite of his rapid ascent, he still did not seem to be getting more air from his demand valve. He must be out of gas.

He burst through the surface, gasping for breath. He wrenched off his ***face mask*** and ***demand valve*** and gasped air.

The water was choppy and waves washed over his face. He kicked hard with his ***fins*** to stay on the surface. One of the ill-fitting borrowed fins came off. A wave washed over his face and he ***inhaled water*** and started coughing. It was a real ***struggle*** to stay on the surface, he was becoming ***exhausted***. He wondered how long he could keep this up. He tried to keep his head well above the waves, but could not.

His buddy noticed he was missing and after a brief search, surfaced. Harry was no where to be seen. An organised search later found his body on the bottom, immediately below where he had surfaced.

His ***weight belt*** was still fastened, his ***buoyancy vest*** uninflated. There was ample air in his cylinder and testing of his demand valve revealed normal functioning, but demonstrating the usual resistance with high gas flows.

The autopsy report read "**drowning**". The **real cause was "death from panic".**

The above story illustrates some of the factors which combined to develop the anxiety which leads to panic and illustrates the irrational responses in a panicked diver. An appropriate logical response at any one of the steps that led to the disaster, would have prevented or relieved the situation.

Prevention.

If anxiety is an important precursor to panic, reducing anxiety is an effective counter measure. The most effective way to reduce anxiety is to have confidence in, and familiarity with, the task. This is achieved by **knowledge, training** and repeated **practice** of diving and safety procedures.

A good example is seen in the training of commercial airline pilots. They are required to fly a minimum number of hours per month and to practice and demonstrate emergency procedures at regular intervals. They spend many hours practicing emergency drills in a flight simulator. The usually cool and appropriate performance of these professionals in emergencies is a testimony to the success of this approach.

Another important preventive measure is for the diver to know his **limitations** and to dive within them. A diver may be comfortable, confident and competent in one diving situation but not in another. The first allows for safe diving, the second for a panic scenario. Panic is more likely when the diver is extending his dive parameters, especially if without competent training and supervision.

FATIGUE

Studies of recreational diver deaths show that **fatigue contributes in about 28% of cases.** This fatigue comes about from a combination of personal, equipment and environmental factors.

- ## Personal.

A high level of physical fitness is an important survival factor in diving. Even the calmest water dive can degenerate into a situation requiring maximal physical exertion due to unforeseen circumstances, such as currents, rescue requirements, etc.

During severe exertion, fatigue and its associated apathy will come sooner to physically unfit divers. Also, fatigue is experienced sooner by anxious or neurotic divers.

As a general rule, scuba divers should be able to swim 200 metres in < 5 minutes, without equipment. A fit diver will complete this in 4 minutes and a very unfit diver may take over 5 minutes.

- ## Equipment.

Much of the diver's equipment, the buoyancy compensator, tank, facemask, and wet-suit either increases resistance to swimming or restricts movement. Excessive weights make swimming more strenuous. Even the best regulators have appreciable resistance to airflow at high flow rates, significantly restricting breathing. All these factors accelerate fatigue.

- ## Environment.

Fully equipped, a diver cannot swim for long against a current of more than about 1 knot (see chapter 6). Rough water and cold exposure will make this even harder.

Chapter 8

THE FEMALE DIVER

Up to the 1960's, diving was almost exclusively a male domain with a certain associated macho image. Since then, an increasing proportion of female divers have come to enjoy this sport and have proven themselves equal to male divers in every regard.

Despite this, women are not the same as men and there are important consequences of this dissimilarity in diving activities. Unfortunately, much of the following information is based on very inadequate data.

History of Women in Diving

Probably the most famous of female diver groups are the **Ama** shell divers of Japan and Korea. These divers were originally men but the work was taken over by women, possibly because of their better tolerance to cold – the men only dived in summer while women were able to dive all year round. Another theory is that the men believed that diving impaired their virility.

The Ama underwent some remarkable physiological adaptations. During the winter months they increased their metabolic rate by 30%, which allowed them to generate more internal heat. Also, they reduced their skin blood flow by 30%. The fat content beneath their skin was increased. Both these changes improved their insulation.

There have been numerous famous women diving personalities. In the 1940's, Simone Cousteau dived alongside her husband Jacques Cousteau. In Australia, Valerie Taylor and Eva Cropp became well known because of their diving exploits. In America, Eugenie Clark was known as the 'shark lady' because of her brilliant work in this field. In 1969, Sylvia Earle led the first all woman team of aquanauts in the Tektite II habitat experiment.

Until recent years **diving instruction** was almost exclusively a male occupation. Many of these instructors basked in a 'superman' role and possessed more experience than knowledge.

In recent years women have become recreational diving instructors and have proven to be diligent and highly competent. In general they have been more keen to impress their students by knowledge and skills, rather than strength and bravado.

Women divers must be doing something right. Diving statistics show that females comprise 34% of trainees, but account for only 10-20% of fatalities (depending on the survey).

Scuba Training

In Western society, women are generally regarded as being less mechanically and mathematically adept than men. This prejudice is reflected in attitudes to diver training. In many cases, women are patronised by well-meaning male instructors and male companions.

Until recently, culturally acquired lack of assertiveness on the part of many women led them to refrain from asking what appeared to them to be naive questions of their instructors. On the other hand, prejudices by instructors led them to assume women would not be interested in, or understand, the intricacies of equipment functioning or decompression planning. This information tended to be directed towards the males in a training group. Women would often turn to a male friend or buddy rather than the instructor for answers to questions which arose during training. The information that they received was not always accurate.

A woman who has her equipment assembled and checked by a male companion, who has the equipment carried to the water and who is assisted into and out of the water, is overall less likely to become a competent and self sufficient diver.

The old stereotype of the woman in a dependent role can lead to problems in diving practice. Thus having men and women buddied together for basic training may be inappropriate. The concept, introduced by PADI some years ago, of an all-women class has much to commend it.

Fig. 8.1

Anatomical Differences

Womankind has been described as the "weaker sex." While it is generally true that on average women are less physically strong than men, there is not a vast difference in their performance in aquatic sports. For example, in the 1988 Olympics there was only a 10–12% difference in times between women and men for swimming events.

For the same physical size, men on average have greater physical strength than women. This is because men have a greater muscle mass per unit body weight. This minor difference in strength is much less significant in the weightless aquatic environment.

Being physically smaller, the woman has a lower requirement for oxygen at a given level of physical activity and will produce less carbon dioxide. With smaller lungs, women also take smaller breaths. Thus, women can often manage with less air than a male diving companion and so can use a smaller, lighter scuba cylinder. This can offset the apparent disadvantage of diminished body size and strength.

Because of differences in body shape, women have different equipment design requirements. There can be difficulty in obtaining appropriate sized or shaped wet suits, fins, boots, and gloves, in less developed countries.

Particular problems arise with male sized equipment, especially face masks which may not fit well, and large scuba cylinders which are unnecessarily bulky or heavy. Backpacks can be too long and so cover the weight belt, making the release of the belt in an emergency difficult. Over sized buoyancy compensators designed for males may give excessive buoyancy and drag with females.

Diving Activity

It is anecdotally believed that males tend to dive deeper, longer, more frequently and engage in more rapid ascents. Whether this is true, is unknown, but they certainly do engage in more risk-taking activities than females in all general epidemiological surveys, and so perhaps it is correct. They also suffer many more accidents and traumatic deaths than females – at all ages.

In one study, females dived mostly between 19-40m (61-68% of the activity) and much less between 41-60m (16-26%) or greater than 60m (1-2%). Young females dived at a more varied depth but also deeper than older females.

Thermal Variations

Women are better insulated than their male counterparts. They have a fat layer beneath the skin some 25% greater then men. They also have a better ability to constrict the blood flow to their limbs, reducing heat loss. These factors allow women to conserve their heat more effectively in a cold water environment while producing natural buoyancy, which improves their swimming and floating ability.

Unfortunately women tend to expose themselves less to demanding environmental temperatures, and so may not become as well adapted to cold exposure, and initially react more to this situation.

Menstruation

Some women perceive their ability to dive safely during the menstrual cycle may be impaired, and that the activity of scuba diving may alter the menstrual cycle. Statistically there is some justification for believing that diving related incidents are more frequent around the time of menstruation.

During menstruation, the average woman is likely to lose 50–150cc blood and cellular debris. There are some physical and physiological consequences of menstruation which will be discussed, but usually there is no reason why women should not dive during menstruation. For convenience, most women today prefer to wear internal protection such as tampons rather then

menstrual pads. In the early days, there was some concern that menstrual blood-loss may act as an attractant to sharks. In fact, females have a much lower incidence of shark attacks than males. This may be related to their different diving behaviour, or that haemolysed blood is a shark repellant.

Hormonal changes before and during menstruation tend to cause fluid retention and swelling. There is a theoretical possibility that this might encourage the development of DCS, and one study of altitude DCS suggested that DCS was more likely earlier in the menstrual cycle. This has not been validated for divers. There is minimal experimental data to support this association, but it may be wise for women, at least during this time, to add a safety margin to their decompression requirements.

Some women have significant psychological and physical problems around the time of menstruation, with abdominal pain, muscle cramps, headaches, nausea and vomiting. These may impair their diving ability. Women who suffer from severe problems of this nature, are advised to avoid diving at this time. The psychological disturbances associated with pre-menstrual tension and anxiety, may sometimes warrant the avoidance of diving during this time.

Some female migraine sufferers, have an increased likelihood of migraines around the time of menstruation. The problems associated with migraine are discussed further in Chapter 32 and the recommendations there should be followed.

Repeated exposure to high pressure environments seems to have no significant effect on hormone regulation, ovulation or menstruation.

Oral Contraceptives (the "Pill")

The physiological and psychological consequences of these hormonal tablets may have similar implications to those described above, under "menstruation". In theory, the increased coagulation effects from the pill could initiate or aggravate DCS. In practice this has not been observed. There is no relationship between the development of DCS and the taking of oral contraceptive tablets *per se*. It was considered prudent to cease oral contraceptives in the female team who undertook a long saturation dive during the Tektite No. 2 project. However, the absence of males probably made this decision an uncomplicated one.

Decompression Sickness (DCS)

Several studies have shown an increased incidence of DCS in women. Some studies of women divers showed a more than 3 times increased incidence of DCS compared with men who were exposed to the same dive profiles. Other studies have not shown this and perhaps the different observations are due to different cohorts with different dive profiles. For example, females in the space program seemed to be more susceptible to altitude DCS, and develop more serious DCS, than men.

Studies on female divers indicated that those who did “reverse profile” dives had more DCS symptoms than those who did their deeper dives first. In a large series of DCS cases it was shown that the men so affected dived deeper than the women (almost 3 metres) and it is possible that their diving exposures and ascent rates were greater.

The weight of evidence does tend to suggest that there might be an increased incidence of DCS among women. There are several possible explanations for this.

Women are frequently less physically fit than men and physical fitness is negatively related to DCS. Women usually have a higher proportion of subcutaneous body fat (+ 10%) for a given weight than men, and the body fat has a 4-5 times higher capacity for absorbing nitrogen. Fat tissue is slower to absorb and to eliminate the nitrogen. Logically, because of the different fat distribution between the sexes, studies that dealt with certain dive profiles (longer dives) could have more nitrogen absorbed and a greater incidence of DCS in females. Both reduced physical fitness and higher fat content probably increase the incidence of DCS .

Navy decompression tables were designed for and tested on physically fit, healthy young male divers. Strictly speaking, the tables should only apply to this population.

Because of this increased risk, it is wise for women divers to apply extra safety factors when using the dive tables: e.g. by reducing the allowable bottom time for any depth or by decompressing for a greater duration. Decompression computer programs should be on the more conservative settings.

A modern decompression problem has emerged with **breast implants**. Fortunately, gas filled implants are no longer used, as the barotrauma consequences of diving with these would be disturbing. However, even silicone filled implants do absorb nitrogen during a dive and a 4% expansion in the size of these implants has been recorded after dive profiles commonly used by women sports divers. This is not likely to cause a problem with the implants. However, if these women were involved in saturation diving there is the potential for significant volume changes which could lead to damage or rupture of the implant during or after ascent.

Other Diving Diseases

In a number of reports on **scuba divers pulmonary oedema**, a preponderance of females has been noted. Others have suggested an increased susceptibility to **dysbaric osteonecrosis**.

The relative risk of **oxygen toxicity** for women compared to men was 1.6 times for pulmonary toxicity and 2.9 times for neurological toxicity.

Panic is a common and potentially serious problem even amongst experienced divers. Although females reported this more frequently than males, the latter waited longer to react and so it was more likely to progress to a life threatening situation in them.

A study of dive masters and instructors found that female instructors and dive masters reported more diving-related **ear and sinus** symptoms than males. This relation was consistent even when controlling for potential confounders.

PREGNANCY

There has been considerable controversy over whether pregnant women should dive. This question arises because most women divers are in the child-bearing age group. The controversy hinges on the conflict between restricting the freedom of an individual and the risks (which have not been fully evaluated) to the unborn child. The potential problems of diving during pregnancy are as follows:

Maternal Effects

❑ Vomiting.

In the second and third months of pregnancy, many women are prone to vomiting – often manifest as "morning sickness". They are more prone to seasickness and to nausea and vomiting underwater during certain conditions. This predisposes to serious diving accidents.

❑ Barotrauma.

From the fourth month onwards, fluid retention and swelling of the lining of the respiratory tract, makes sinus and ear equalisation more difficult and predisposes to barotrauma.

❑ Respiratory function.

There is a decline in respiratory function as pregnancy progresses. There is an increase in resistance to air flow in the lungs. Later, the enlarging baby presses up into the chest, limiting breathing capacity. This combination impairs the pregnant woman's ability to cope with strenuous activity which may be required in an emergency, and may predispose to hypoxia or pulmonary barotrauma.

❑ Decompression sickness.

There are major alterations in blood volume and circulation during pregnancy. This may increase the uptake and distribution of nitrogen and may make the woman more prone to DCS.

❑ Infection.

It has been claimed that there is an increased risk of vaginal infection in pregnant women who dive. In the later stages of pregnancy some women develop minor leaks of the amniotic sac, which surrounds the baby with fluid. There is a possibility of infection of this fluid from organisms entering from the water before birth or directly into the womb after birth.

Effects on the Baby

The developing foetus is uniquely at risk from some of the physiological hazards associated with diving. The potential risk primarily consists of DCS, but hyperoxia, hypoxia, hypercapnia and increased nitrogen pressure may also be involved.

❑ Development of the foetus.

The foetus begins as a single cell organism and up until after the fourth month it is smaller than a mouse. A small bubble, such as develops during DCS, could have catastrophic effects.

The **circulation in the foetus** is unique and critical. In an adult diver, venous blood returning from the body passes through the lung capillaries, which filter out the bubbles frequently formed during or after ascent. In the foetus, the blood by-passes the lungs (since the foetus does not need to breathe) and passes directly to the left heart without passing first through this filtering network. Even one bubble forming in the tissues then veins of a foetus will be transported directly to the arterial circulation and will embolise to somewhere in its body.

In one study it was shown that diving >30 metres in the first trimester was associated with a 16% risk of foetal abnormality, but the numbers were small. Other studies have produced conflicting data.

❑ Hypoxia.

The outcome of many non-fatal diving incidents is hypoxia, most likely to be caused by salt water aspiration or near drowning. The pregnant diver will not only expose herself to hypoxia in this situation but will also expose the much more susceptible foetus to this.

❑ Hyperbaric oxygen.

Divers are likely to be exposed to hyperbaric oxygen in two situations. By simply breathing compressed air at depth they are inhaling elevated partial pressures of oxygen. In addition, if divers develop DCS or air embolism they will be given hyperbaric oxygen therapy for treatment of the condition.

Some foetal tissues are very sensitive to high partial pressures of oxygen. Great care is taken with newborn premature babies to avoid administration of high concentrations of oxygen because of the danger of retrolental fibroplasia, which causes blindness. The eye of the unborn baby is probably even more sensitive to high partial pressures of oxygen.

The circulation of the foetus contains a channel (the ductus arteriosus) which allows blood to by-pass the lungs. This channel closes after birth under the influence of a raised partial pressure of oxygen in the blood. There is a danger of premature closing of this and other shunts if the foetus is exposed to hyperbaric oxygen because of treatment given to the mother.

❑ Decompression sickness.

As mentioned earlier, women may have increased susceptibility to DCS and there are theoretical reasons to believe that pregnant women are even more susceptible. It is known from Doppler studies that showers of bubbles are regularly formed in the veins of divers ascending from many routine non-decompression dives. These bubbles do not usually cause any symptoms.

Some experiments in pregnant animals suggest that the foetus is more resistant to bubble formation than the mother but that bubbles do form after some dives, especially those deeper than 20 metres. Because of the unique circulation of the foetus even a few bubbles in the foetal circulation can have disastrous consequences.

Experiments with pregnant animals have produced conflicting results. One study on pregnant sheep (which have a placenta similar to a human) showed that the foetus developed bubbles in its circulation even after dives of less than 30 metres (100 ft.) and within the US Navy no-decompression limits. These results are disturbing when considering the vulnerability of the foetus to any bubble.

Other studies have shown an increased incidence of abortion, birth defects and still-births in pregnant animals after decompression.

Exposure to hyperbaric oxygen has also been shown in some studies to cause birth abnormalities and death.

Human Data

Japanese female divers, the Ama, often dived until late pregnancy, and had a 44% incidence of premature delivery with a high incidence of small babies when compared with non-diving women from the same area. Another survey on modern day Ama, who were not exposed to the same stresses, did not reflect this association.

Margaret Bolton from the University of Florida carried out a survey on 208 women who dived during pregnancy. She found an increased incidence of abortion, still-birth, low birth weight and death of the infant within the first month. Of the 24 women who reported diving to 30 metres (100 ft) or more, three had children with congenital defects. This contrasts with an incidence of 1 in 50 in the general population. One of the infants had an absent hand, a very rare abnormality.

An Australian case report, showing multiple grotesque abnormalities after diving, suggested that the effects and vulnerable time frame may be similar to taking congenital-abnormality producing drug, Thalidomide, during pregnancy.

It is difficult to draw firm conclusions from these studies, or others that have been conducted, because the numbers are too low for statistical validation, or are retrospective. They are however consistent with many of the animal studies.

The Bottom Line ---

There is considerable evidence suggesting that diving during pregnancy is harmful to the foetus. It is generally accepted that unnecessary drugs, alcohol and smoking should be avoided during pregnancy because of the risk to the foetus and we recommend a similar conservative approach to diving. The sacrifice of not diving during pregnancy may be easier to cope with than the guilt, valid or not, associated with abortion or giving birth to a malformed child.

Evidence in western societies suggest that female divers have accepted the above restriction and now do avoid diving when pregnant, or when attempting to become pregnant, making the collection of more data difficult, but good for the kids.

One interesting issue to consider by women who contemplate diving during pregnancy is that the foetus, who will have to live with any birth defect which may result, cannot be consulted when the decision to dive is made.

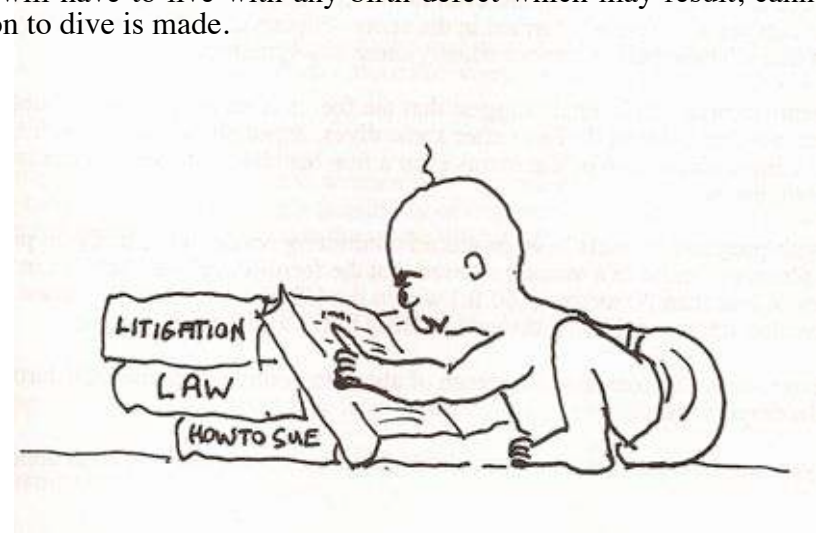

Fig. 8.2

Chapter 9

EAR BAROTRAUMA

ANATOMY OF THE EAR

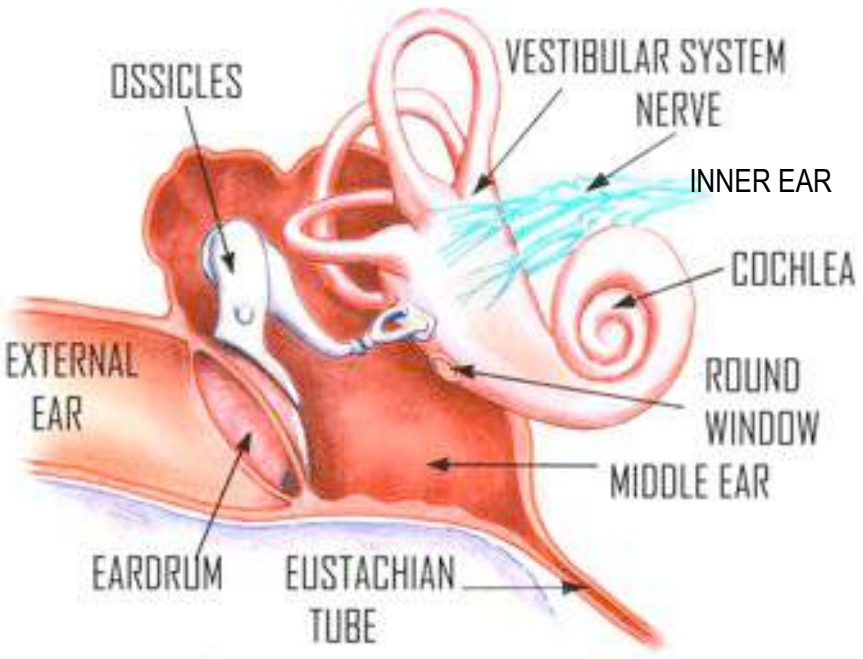

Fig. 9.1
Anatomy of the Ear

The ear is divided anatomically into the outer (external), the middle and the inner ear.

The **Outer Ear** comprises the visible part of the ear (the pinna) and the external ear canal. The pinna gathers sound waves and reflects them into the ear canal and onto the ear drum.

The **Middle Ear** is a pea sized cavity enclosed in a solid bony part of the skull. It is separated from the ear canal by the paper thin **ear drum**. There are several structures opening into the middle ear space.

• The **Eustachian Tube** joins the middle ear with the throat, allowing air to enter the middle ear cavity.

• The **Mastoid Sinus** (air pockets in the mastoid bone) also come off the middle ear.

There are two openings on the inner bony surface of the middle ear space called the **Round** and **Oval Windows,** because of their shape. These openings connect the middle to the inner ear. The oval window is a tough membrane attached to the end of one of the three interconnecting middle ear bones (ossicles), while the round window is closed by a thin delicate membrane.

The **Ear Drum (or Tympanic Membrane)** is connected by the three tiny ear bones or *ossicles* - the malleus, incus and stapes - to the oval window across the middle ear space. This bony chain, which is barely visible to the naked eye, transmits the sound vibrations from the ear drum to the inner ear.

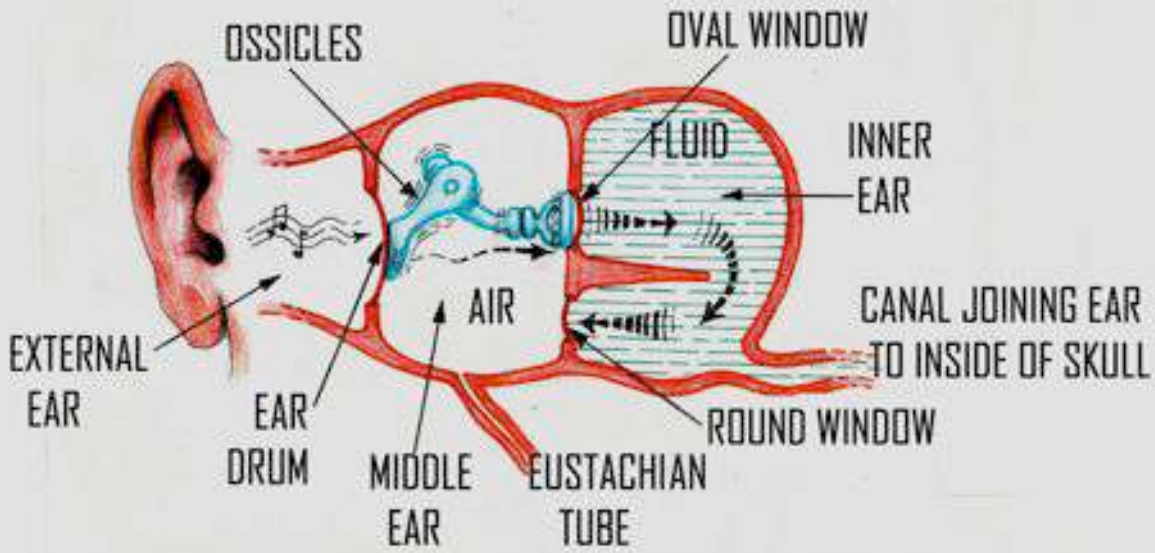

Fig. 9.2
Middle ear communicates with external and inner ear
Transmitting sound from the exterior to the inner ear

The **Inner Ear** contains the **Hearing** and **Balance organs**. It is entirely encased in bone and filled with fluid. The hearing organ (the **Cochlea**) is a spiral shaped structure containing fluid which surrounds nerve cells sensitive to sound vibrations.

A system of **3 semi–circular canals** is also filled with this fluid, and is the balance organ which is sensitive to position and movement. It is also called the **Vestibular System** (or **Vestibular Apparatus).**

THE MECHANISM OF HEARING

The hearing system works in an ingenious way. Sound vibrations, caught and reflected by the pinna, are directed down the ear canal causing the ear drum, at the end of the ear canal, to

vibrate. These vibrations are transmitted and magnified by the bony chain system of levers, to the oval window. The outer (air) and middle ears (ossicles) thus **conduct** sound waves to the inner ear.

Damage to the outer and middle ears interferes with the conduction of sound waves to the inner ear, and so causes **conductive deafness**. This affects the low frequency sounds (those commonly used with speech, 250-4000 Hz).

At the oval window, the sound waves are converted to pressure waves in the hydraulic fluid in the cochlea. As fluid is incompressible, the inward and outward vibrations of the oval window are compensated for by the round window bulging outward and inward, respectively.

The cochlea is tuned so that vibrations of various frequencies transmitted through it resonate in specific areas, allowing the ear to distinguish between differing frequencies of sound. This stimulates sensory nerve fibres within the cochlea and the impulses are perceived as sound when they reach the brain. Damage to this system causes **sensori-neural deafness.**

EXTERNAL EAR BAROTRAUMA

(EXTERNAL EAR SQUEEZE)

If the external ear canal is obstructed, the enclosed gas will be compressed and so reduce in volume during descent (Boyle's Law, again!). This will cause an outward bulging of the eardrum and swelling and bruising of the skin lining the ear canal.

Obstruction of the canal can be caused by a tight fitting **hood**, **wax** in the ears, **bony growths** (**exostoses**) in the ear or the wearing of **ear plugs**. As this condition can be encountered at depths as little as 2 metres, **ear plugs should not be worn during any type of diving.**

The symptoms include discomfort and pain on descent, bleeding from the external ear and the other pressure effects of barotrauma on the middle ear, including difficult equalisation.

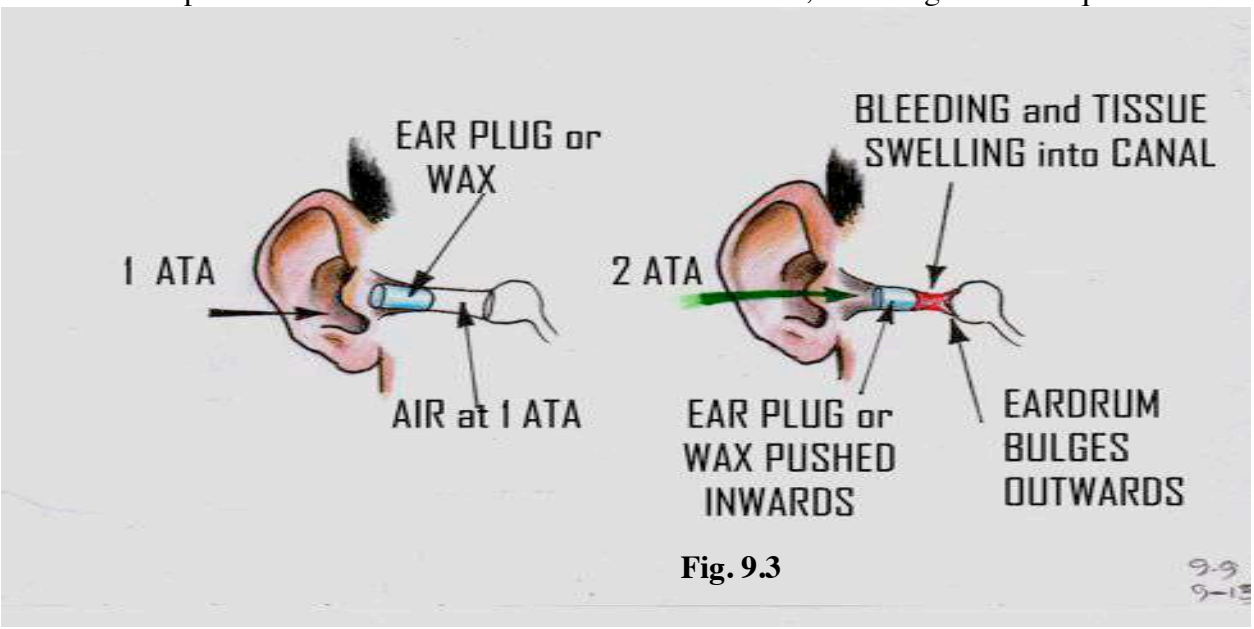

Fig. 9.3

MIDDLE EAR BAROTRAUMA OF DESCENT

(MIDDLE EAR SQUEEZE, AEROTITIS MEDIA)

The main risk of barotrauma to the ears is encountered on descent and the commonest site is the middle ear. About one quarter of diving trainees experience this, to a variable degree.

Water pressure around the diver increases as he descends. This pressure is transmitted to the body fluids and tissues surrounding the middle ear space causing compression of the gas space in the middle ear (Boyle's Law). The diver is aware of this sensation of pressure and voluntarily compensates for the reduction in middle ear gas volume by "**equalising the ear**" "auto-inflation" or "**clearing**". In this manoeuvre, air is blown up the Eustachian tube to replace the volume of gas compressed in the middle ear space. The ways of doing this are described later.

If the diver fails to equalise, water pressure will force the ear drum inwards, stretching it and escalating the sensation of pressure into one of pain. At the same time, reduced gas volume in the middle ear is compensated by blood and tissue fluid, swelling of the lining (mucosa) of the middle ear space. Ultimately, the blood vessels become over distended and rupture, bleeding into the ear drum and the middle ear space. This tissue damage takes days or weeks to resolve. Sometimes the ear drum itself will tear or rupture.

The depth at which this damage occurs depends on the size of the middle ear space and the flexibility of the ear drum. It is normally reached at 1–2 metres and if the diver does not equalise by the time he has reached this depth, barotrauma of the ear is likely.

The commonest diving problem!

If the diver can equalise his ears on the surface, the problem is due to incorrect diving technique

Fig. 9.4

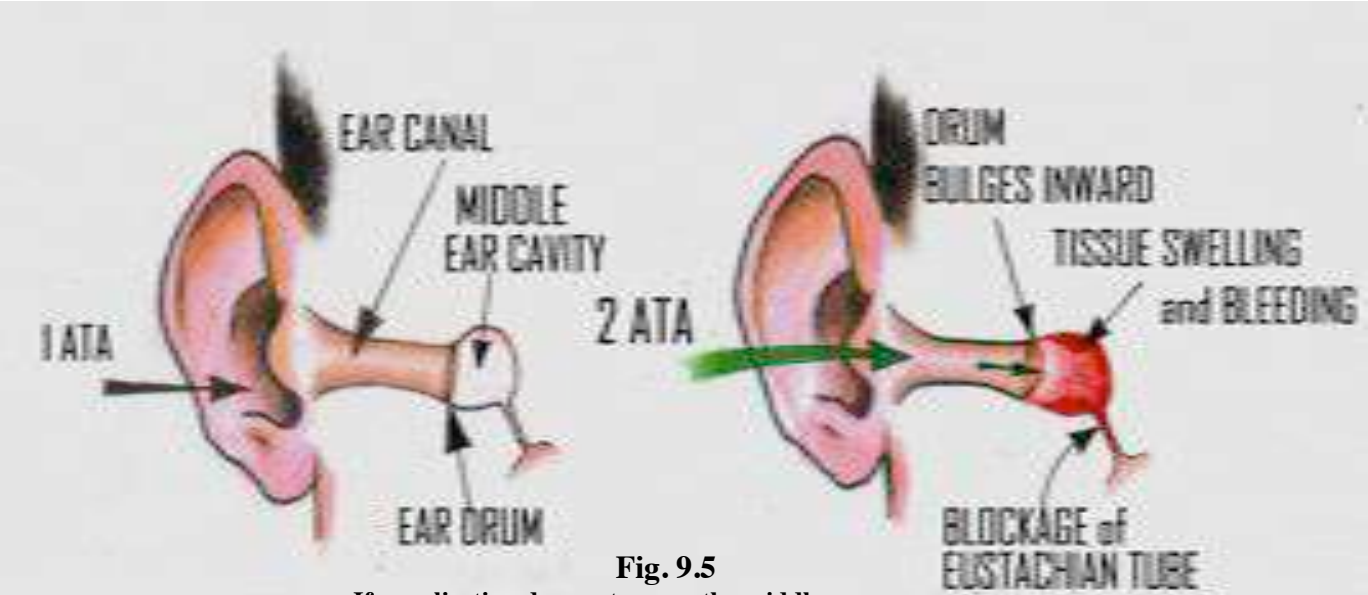

Fig. 9.5
If equalisation does not occur, the middle ear space is partly filled with blood, tissue fluid and swollen mucosa.

Clinical Features

A sensation of **pressure** is the first symptom of damage to the ear. This pressure sensation may develop into **pain**, which is usually severe, sharp and localised to the affected side. It increases as the diver descends, unless he equalises the middle ear spaces.

If the diver continues descending until the **ear drum ruptures**, he will experience relief of the pressure or pain, followed by a **cold feeling** in the ear. This is due to the sea-water which enters the middle ear space, cooling the bone and tissues near the balance organ. Thermal currents may be produced within the balance organ, causing stimulation and dizziness. Fluid may also be felt trickling down the throat, after running down the Eustachian tube from the middle ear space.

Fig. 9.6
Large perforation of the ear drum from a single dive to 8 metres without adequate equalisation.

Middle ear barotrauma, especially when associated with rupture of the ear drum, may be accompanied by **dizziness.** This sensation is termed **vertigo**. It may also be accompanied by **nausea** and **vomiting**. Vomiting underwater is a skill not frequently practiced and it can block the air supply and lead to aspiration of sea water (and vomitus) and drowning.

With lesser degrees of barotrauma, pain or discomfort in the ear may also be felt after the dive. There is often a **feeling of fullness** (or "water") in the ear and sounds may appear muffled. Crackling noises may also be heard (especially with chewing, swallowing or with jaw movements), caused by bubbles of air in the blood/body fluid mixtures within the middle ear.

Occasionally blood from the middle ear is forced down the Eustachian tube when the middle ear gas expands on ascent (Boyles Law). After surfacing the diver may notice **small amounts of blood** coming from the nose, on the same side as the barotrauma, or running down the throat.

Other symptoms due to the ear barotrauma include; a "**squeaking**" sound during equalisation (due to a Eustachian tube narrowed by mucosal swelling, or scarring from repeated episodes of barotrauma), or an **echo** sensation and/or a mild **ache** and **tenderness** over the ear/mastoid area following the dive.

Occasionally a diver has a naturally reduced pain appreciation, so that considerable barotrauma damage can be done despite the absence of much discomfort. These divers are vulnerable to permanent damage, as the usual warnings of middle ear barotrauma are absent. Others can suffer significant damage after exposure to even small pressures, e.g. in a swimming pool – especially if the submersion and pressure exposure is for many minutes duration.

The really serious problem from middle ear barotrauma in extension of the pathology to the inner ear with haemorrhage, or rupture of the round window and permanent hearing loss.

Treatment

A diver who has experienced middle ear barotrauma needs an examination by a diving **physician** to diagnose the condition, and check for complications such as a perforated eardrum or inner ear damage. Assessment of the cause, and advice on prevention of future difficulties is important (see later). **Audiograms** (hearing tests) are essential to test for damage to, and function of, the middle and inner ears.

Occasionally, the doctor may prescribe an oral **decongestant (or nasal spray)** to help open the Eustachian tube, while **antibiotics** may be prescribed if **infection** is present in the nose or throat area, or develops in the blood pooled in the middle ear cavity. This usually presents as a recurrence of pain hours or days after the barotrauma. **Ear drops** do not reach the middle ear and are of no value. They may be harmful if the ear drum has ruptured.

Once serious complications have been excluded, active treatment is usually unnecessary. In order to rest the ear and allow healing, diving, flying and middle ear equalisation should be avoided until the barotrauma has resolved. This commonly takes from 1–2 days up to 1–2 weeks.

The length of time away from diving depends on the severity of the barotrauma. The diver should not return to diving or flying until the physician has confirmed resolution of the barotrauma and the ability to equalise the ears. **Understanding why the barotrauma developed is necessary in order to prevent it happening again.**

If the ear drum was **perforated**, complete cure may take 1–3 months, even though it may appear to have sealed over within days. Early return to diving predisposes to recurrent perforation. Occasionally the drum fails to heal and requires surgical "patching" or "grafting" by a specialist. Later, diving may be permitted if repair is complete and easy voluntary equalisation of the ear is demonstrated.

It is necessary with the recurrence of symptoms, to perform repeat audiograms to confirm that no inner ear damage has occurred. About half of these divers may have ENT pathology, which may be treated by an otologist, and half will be recurring because of marginal Eustachian tube narrowing and failure to perform the equalisation techniques correctly. Both groups will gain from learning the correct “equalisation ahead of the dive” techniques.

PREVENTION OF BAROTRAUMA

Equalisation

Adequate equalisation of pressures in the middle ear space will prevent middle and inner ear barotrauma. This equalisation is necessary whenever increasing depth in the water. It should be performed frequently and before any ear discomfort is felt. It is necessary to equalise more frequently near the surface since the volume changes are greatest there (as explained by Boyle's Law). Equalisation should always be gentle to avoid damage. The technique of ear equalisation is a skill which improves with practice. Some divers can equalise without any apparent effort or action.

Upper respiratory tract infections (URTIs) cause congestion of the throat and Eustachian tube openings, making equalisation difficult or impossible. **Hay fever, allergies, snorting drugs** or **cigarette smoking** have a similar effect. Diving with these conditions is risky. A deviated nasal septum may also predispose to aural barotrauma as well as sinus barotrauma (see Chapter 10).

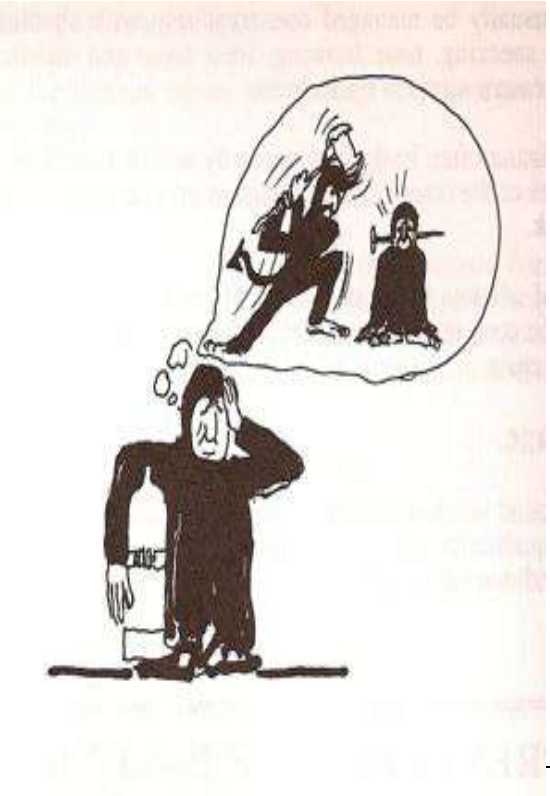

There are several ways of active and voluntary middle ear equalising before and during descent:

- **Valsalva manoeuvre.**

This technique is most frequently used because it is easy and effective. The diver **holds his nose, closes his mouth and blows gently against the closed nose and mouth.** This raises the pressure in the pharynx, forcing air up the Eustachian tubes into the middle ear. He hears the ear drums “pop”. If they produce a longer squeak, the tube is partially obstructed.

If there is infected material in the throat this can also be blown up the Eustachian tube into the middle ear, leading to **infection**. This is another reason why divers are advised against diving with an upper respiratory tract infection.

Fig.9.7

To supplement this manoeuvre, **opening of the Eustachian tubes** can be facilitated by wriggling the jaw from side to side or thrusting the lower jaw forward as the manoeuvre is performed (**Edmonds Number 1 technique**).

A drawback of the Valsalva technique is that if it employed too forcefully, it is theoretically possible that the inner ear may be damaged. Another drawback is that the nose must be held closed with the fingers, which is not always easy with some professional diving helmets or full-face masks.

- **The Toynbee manoeuvre.**

This involves **holding the nose and swallowing simultaneously**. This usually causes the Eustachian tubes to open momentarily, allowing air to enter or leave the middle ear.

The Eustachian tubes open only briefly with this manoeuvre and it causes a negative pressure in the pharynx, so only smaller amounts of air are able to pass into the middle ear space. Consequently, this manoeuvre is not as effective as the Valsalva manoeuvre, but it is used successfully by many divers.

- **Others.**

Voluntary Opening of the Eustachian tubes (BTV technique) can be performed at will by many experienced divers, by contracting certain muscles in their throat. This technique can be performed by holding the nose, closing the mouth, and then trying to lift up the larynx (Adams Apple), which can be viewed in a mirror. A clicking can be heard in both ears if the procedure is successful. This technique is difficult to describe but if it can be mastered it is convenient and effective, there is little force involved and the manoeuvre can be performed repeatedly.

If any difficulty is encountered, the **Lowry technique** ("swallow and then blow at the same time" – a Toynbee + Valsalva combination) or the **Edmonds Number 2 technique** ("sniff and blow" – suck the cheeks in with a sniff against the closed nostrils, immediately followed by a Valsalva), may be used.

Diving Technique

Anyone who has problems with middle ear equalisation should gently practice this procedure a few times each day, on land. Practice improves performance and makes perfect. Voluntary controlled equalisation becomes easier with repetition. Some may even need to use nasal decongestants to assist this, at first, but they should not be used when diving. Any diver who has difficulty with middle ear barotrauma should practice using the Valsalva, Lowry or Edmonds techniques, as they cause a positive pressure in the middle ears. They should not rely on swallowing, Toynbee or BTV techniques or any others that result in negative middle ear pressures, even though they may open the Eustachian tubes.

Many divers suffer hearing loss because they do not equalise their ears correctly.

Ideally on the day of a proposed dive, the diver should confirm that he is able to equalise easily before setting out on a diving expedition.

All divers should equalise on the surface immediately before descending. This confirms that equalisation is possible and the ear drums balloon slightly outwards, causing a slight positive pressure in the middle ear and allowing some margin for error if the diver becomes distracted and forgets to equalise during the first metre of descent.

The diver should then **equalise every metre or less as the descent proceeds**, so that no sensation of pressure is felt. This is called "**equalising ahead of the dive**" and is much safer than waiting until the pressure sensation (or actual pain) is felt. By that stage the middle ear mucosa is already swollen and obstructing the Eustachian tube, making equalisation more difficult.

If any difficulty is encountered, it is unnecessarily dangerous to descend further as equalisation will become more difficult - until the middle ear becomes congested with tissue fluid and blood. Instead the diver should either abort the dive or, if the dive is an important one, immediately ascend a little and repeat the equalisation manoeuvre. The diver should not persist with this "yo-yo" technique, or remain at a depth at which a "pressure sensation" is present. The middle ear is filling ("equalising") with blood. This is not a sensible situation. If the ears do not equalise easily, abort the dive. That way, he will be back diving much more quickly, having not damaged the middle ear.

If descent is continued, a '**locking effect' on the Eustachian tube** may develop**.** This is caused by the pressure difference between the middle ear and the throat. Equalisation is then impossible.

Descending 'feet first' makes equalisation considerably easier, and is best done on an anchor or shot line. This allows accurate control of the descent rate and depth while allowing the diver to concentrate on equalisation without the distractions of swimming and depth control. At least keep the head vertical when equalising, assisting the passage of air up the Eustachian tubes.

The novice diver and the diver who has difficulty with his ears should use a face mask which allows easy access to the nose to facilitate the various manoeuvres. If one ear causes more problems, then cock this ear to the surface when equalising (this brings that Eustachian tube even more vertical, and air travels upwards, under water.

Surgical correction of nasal septal deviations, cessation of smoking and adequate treatment of URTIs and allergies may be needed by those who have these predisposing causes.

Medication

Medication has been used to facilitate equalisation when there is some disorder of the ear, nose or throat. Topical nasal decongestant sprays and drops such as **phenylephrine and oxymetazoline** shrink the lining of the nose and Eustachian tube, reducing congestion and opening the air passages. Some oral medications such as **pseudoephedrine** have a similar effect.

While these drugs can make diving possible when it otherwise would not be, **they assist the diver to dive with conditions which should preclude diving**, such as upper respiratory tract infections. There is an added risk of the drug predisposing to barotrauma during the ascent - which is far more dangerous than ear barotrauma of descent. These drugs often relieve obstruction of the throat end of the Eustachian tube, improving equalisation during descent. Unfortunately they do not influence the middle ear end of the Eustachian tube, thus they do not assist release of gas from the middle ear during ascent. Decongestants may also permit diving despite minor descent barotrauma but can wear off during the dive, resulting in congestion of the pharyngeal end of the Eustachian tube and further obstruction to the outflow of gas from the middle ear (see below, "middle ear barotrauma of ascent").

Because these drugs mask the symptoms of conditions which would otherwise preclude diving, divers are advised that it is **better to avoid diving than to continue while taking these drugs.** See Chapter 37, last page, for more on the problems of drug treatment. They make other diving disorders more likely e.g. Sudden Death Syndrome.

•Middle Ear Equalisation (ME=) Client's Problem Check List

- **I descend a bit slower than my buddies.**
- **If there is any pressure, I halt my descent and wait a bit. Or,**
- **I may ascend until the ear clears. (Yo-Yo)**

Why do these procedures? If you are not ME= promptly or sufficiently, then these procedures merely allow the ME to fill with blood or tissue fluids, and thus allow further descent with less pain or discomfort. This is not a sensible way to ME=. It results in middle ear congestion, Eustachian tube obstruction and other pathology which may be temporary or permanent.

- **I am trying to use swallowing to ME=**

If you have any difficulty with ME=, employing techniques that result in a relatively negative ME pressure causes ME congestion and Eustachian tube blockage. Use the positive pressure Valsalva technique (or Lowry, or Edmonds) , prior to and during descents.

- **I dive down the shot line.**

This requires greater force to inflate the ME, as you are trying to force air down the Eustachian tube. Descend feet first and you can blow air up the Eustachian tube. Air travels easier up than down, in the water. Remember bubbles? They rise.

- **If there is any water in my ears (fullness, crackling) after the dive, I use alcohol ear drops to dry them out.**

It is likely that the "water" in your ear is really fluid in your ME!

- **I sometimes have a bit of blood from my nostril (or in my throat).**

Although it may be from your sinus, following expansion of air with ascent, it is more likely from the middle ear on that side. In either case, ME= correctly ("ahead of the dive") may well fix both. See previous pages.

- **When I dive and ME=, I hear a squeeking sound in my ear.**

This suggests a narrowed Eustachian Tube, possibly from inadequate ME=. The sound you should hear when you equalise, and the ear drum moves outward, is a click or pop. It takes a split second to achieve. Not a long drawn out sound.

- **I can often dive once, without problems, but cannot ME= on other dives.**

You have probably produced some middle ear congestion in the first dive, but continued the dive. By the second dive, you start off with significant middle ear congestion, and so ME= is more difficult.

- **One ear equalises before the other**

Not a problem. It is normal. You may wish to assist the slow ear by pointing it towards the surface as you ME=.

If reasonable, avoid the use of hoods which cover the external ear. If it is necessary to use these hoods, then it is preferable to have holes inserted over the ear, which will allow air and water movement and pressurisation. This is not, of course, necessary or possible with diving helmets or certain dry suits. Special ear plugs and masks used for prevention of ear barotrauma probably only work by delaying damage and therefore inducing fewer symptoms - not by preventing damage.

MIDDLE EAR BAROTRAUMA OF ASCENT

(ALTERNOBARIC VERTIGO, REVERSE SQUEEZE)

This condition is relatively uncommon by itself, but it is often a complication of a middle ear barotrauma of descent and/or the use of nasal decongestants. During ascent, air in the middle ear space expands and must escape. The air normally escapes down the Eustachian tube to the throat without any conscious effort by the diver. If very observant, he may actually hear or feel it escape from his ear.

Occasionally the Eustachian tube may obstruct this flow of air, with subsequent air distension and increased pressure sensation in the middle ear cavity during ascent. This causes bulging and possible rupture of the ear drum. There may also be damage to the inner ear, leading to hearing loss (see below, and Chapter 30). The increased pressure in the middle ear may also stimulate the nearby balance organ producing vertigo and its associated symptoms.

Clinical Features

Increasing pressure and pain is sometimes felt in the affected ear as the diver ascends. Often there may be **vertigo** as well as **nausea** and **vomiting**. After surfacing the diver may feel fullness or dullness in the ear. **Tinnitus** or **hearing loss** may indicate serious damage (inner ear barotrauma).

Vertigo may develop after only a metre or so of ascent (see Chapter 31, Case History 31.2). Many of these symptoms can be hazardous, especially as ascent may be prevented by the symptoms.

First–Aid

If a diver encounters ear pain or vertigo during ascent, he should **descend a little** to minimise the pressure imbalance and attempt to open the Eustachian tube by holding the nose and swallowing (Toynbee, or other equalisation manoeuvre). If successful, this equalises the middle ear by opening it up to the throat and relieves the distension in the affected middle ear.

Occluding the external ear by pressing in the tragus (the small fold of cartilage in front of the ear canal) and suddenly pressing the enclosed water inwards, may occasionally force open the Eustachian tube. If this fails then try any of the other techniques of equalisation described previously, and attempt a slow ascent.

Treatment

Uncomplicated cases resolve quickly but eardrum rupture or inner ear damage may need **specialised care. All cases need expert diving medical assessment, for diagnosis and advice.** Unless it can be understood and prevented, diving should be avoided.

INNER EAR BAROTRAUMA

A serious consequence of ear barotrauma is inner ear (hearing and balance organ) damage. The inner ear can be damaged in several ways.

- **Round window fistula (or "leak").**

If the diver fails to equalise the middle ear adequately, water pressure will bulge the eardrum inwards. Since the eardrum is connected to the oval window by the bony chain, this window is forced inwards and the round window bulges outwards. If these movements are excessive, the small end-bone can be pushed through the oval window or, more commonly, the round window may tear. After these injuries, the window may then leak the inner ear fluid (perilymph) into the middle ear.

Round window fistula may also be associated with an excessively **forceful middle ear equalisation manoeuvre**. Increased intravascular pressures in the head associated with this manoeuvre may be transmitted to the cochlea fluid, causing bulging and then rupture of the round window. Alternatively, the sudden displacement of the eardrum after an equalisation manoeuvre may set up a pressure wave in the inner ear fluid which tears the round window. The fluid which leaks out is crucial to the healthy function of the **cochlea** and its loss leads to damage to the hearing organ. Permanent severe **hearing loss** may follow if the fluid loss is not interrupted by healing of the round window or surgical repair,.

The same fluid also bathes the balance organ, and damage to this organ may also result in **dizziness (vertigo), nausea and vomiting.**

- **Other pathology.**

Permanent hearing loss or balance disturbance, unrelated to round window fistula, can be caused by direct cochlea damage from inner ear barotrauma. The cause may be haemorrhage (or bleeding), inner ear membrane rupture, or air entering the inner ear (from a stretched round window). This hearing loss may be temporary or permanent depending on the degree of damage and its management.

Clinical Features

The cardinal features of inner ear barotrauma are :

- **tinnitus** (ringing or buzzing noises in the ears)
- **hearing loss**
- **vertigo** (a feeling of being pulled, rotation, rocking or unsteadiness)
- nausea or vomiting
- rarely, **dysacusis** (painful hearing)

One or more of these must be present to make the diagnosis. Fluid may be noted in the middle ear.

Tinnitus is a ringing, buzzing or musical sound in the ear, usually high pitched, due to damage or irritation of the nerve cells of the cochlea.

Hearing loss is due to damage of the cochlea. It may improve, stay the same or deteriorate. Audiograms may differentiate this type of hearing loss (sensori-neural) from that due to middle ear barotrauma (conductive). See Chapter 30.

Vertigo is the spinning or pulling sensation due to balance organ damage (see Chapter 31).

Treatment

A diver presenting with any of these symptoms needs **immediate assessment** by a diving physician. In the interim, he should avoid any exertion, middle ear equalisation, altitude or diving exposure, sneezing or nose blowing.

The physician will examine the ear and perform **serial audiograms** to detect any hearing loss, which may not be obvious to the diver. **Tests** of **balance organ function (ENGs)** may be necessary.

Aspirin, nicotinic acid (vitamins), other vasodilators or anti-coagulants should not be taken.

An expert diving medical opinion concerning **future diving** should be sought if the diver has sustained permanent hearing loss, tinnitus or balance disturbance, as it is probable that further episodes of inner ear barotrauma will cause additional and possibly permanent disastrous effects. Training in correct middle ear equalisation techniques is essential if a return to diving is contemplated.

- ### Round window fistula.

This condition can usually be managed conservatively with **absolute bed rest in the sitting position.** Straining, sneezing, nose blowing, sexual activity, loud noise and middle ear equalising should be avoided, to prevent pressure waves in the inner ear.

The round window fistula often heals spontaneously within a week or two with this regimen but if hearing loss progresses or the other features persist, it may be necessary to resort to **surgery** to patch the round window leak.

Once an oval or round window fistula or cochlea injury has healed, the diver's future in this sport is bleak. Flying should be completely avoided for some months to allow complete healing of the injury or the surgical repair.

- ### Cochlea damage.

In the absence of a round window fistula, no specific treatment is available for this type of injury. Rest in a head elevated position, repeated pure tone audiograms, and avoidance of exertion, equalisation attempts and further exposures to barotrauma (flying or diving), is necessary until the condition has stabilised.

Chapter 10

SINUS BAROTRAUMA

ANATOMY OF THE SINUSES

The sinuses are air filled cavities contained within the bones of the base and front of the skull. Apart from causing inconvenience to divers, their exact function is unknown.

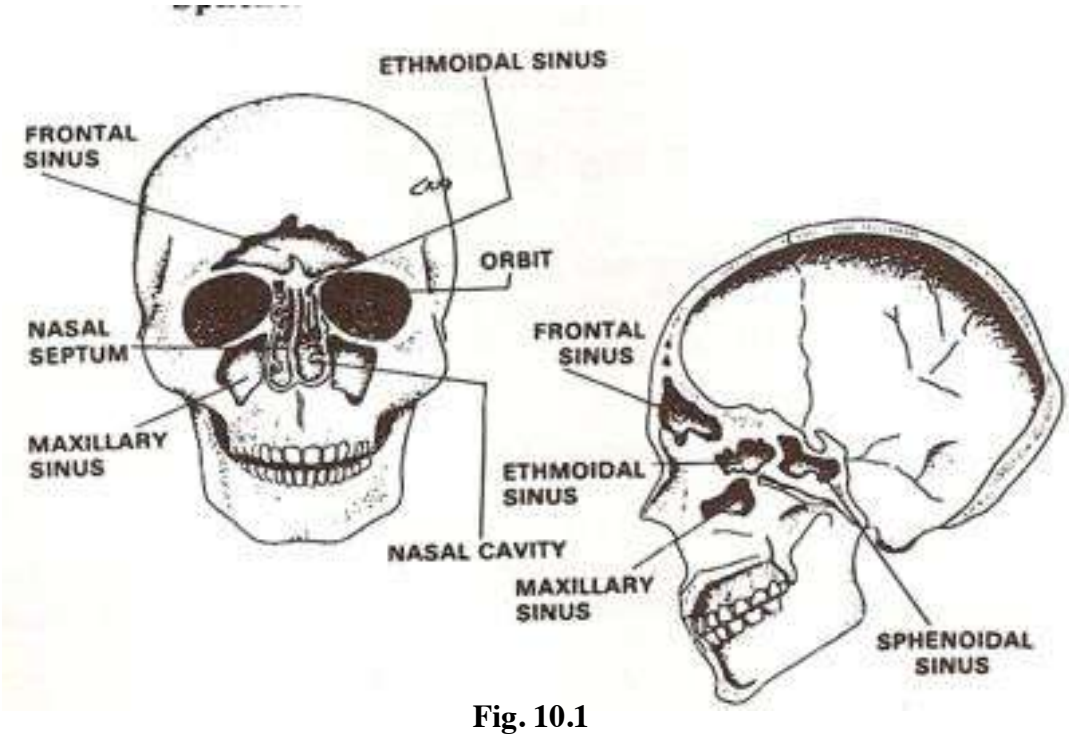

Fig. 10.1
Location of sinuses in the skull. They are connected by canals to the nose.

There are four main groups of sinuses, with openings into to the nose :

- Maxillary sinuses in the cheek bones
- Frontal sinuses in the skull above the eyes
- Ethmoid sinuses in the thin bone at the base of the nose
- Sphenoidal sinuses situated deep inside the central part of the skull.

All the sinuses are lined by a soft mucous-secreting tissue which is richly supplied with blood vessels. Each sinus communicates with the nose by its own narrow opening called the ostium, and through these, the sinuses are permanently open to the atmosphere.

The mastoid sinus or antrum is a similar structure that opens into the middle ear cavity. It more often reflects the pathology of the middle ear and reference should be made to Chapter 9 for this.

THE MECHANISM OF SINUS BAROTRAUMA

As the water pressure changes during a dive, the sinuses normally equalise automatically by free passage of gas into or out of their openings.

Problems are inevitable, however, if these openings become obstructed. Obstruction can be due to congestion associated with allergy, smoking, respiratory tract infection or the overuse of topical decongestants and other drugs. Other causes of ostia obstruction include chronic sinus inflammation (sinusitis), nasal inflammation (rhinitis), folds of tissue (polyps) and plugs of mucous.

When the sinus is blocked, during descent the gas in the sinus is compressed (according to Boyles' Law) causing sinus barotrauma of descent. The shrinking volume is replaced by swelling of the sinus lining, tissue fluid or bleeding – partly filling the sinus.

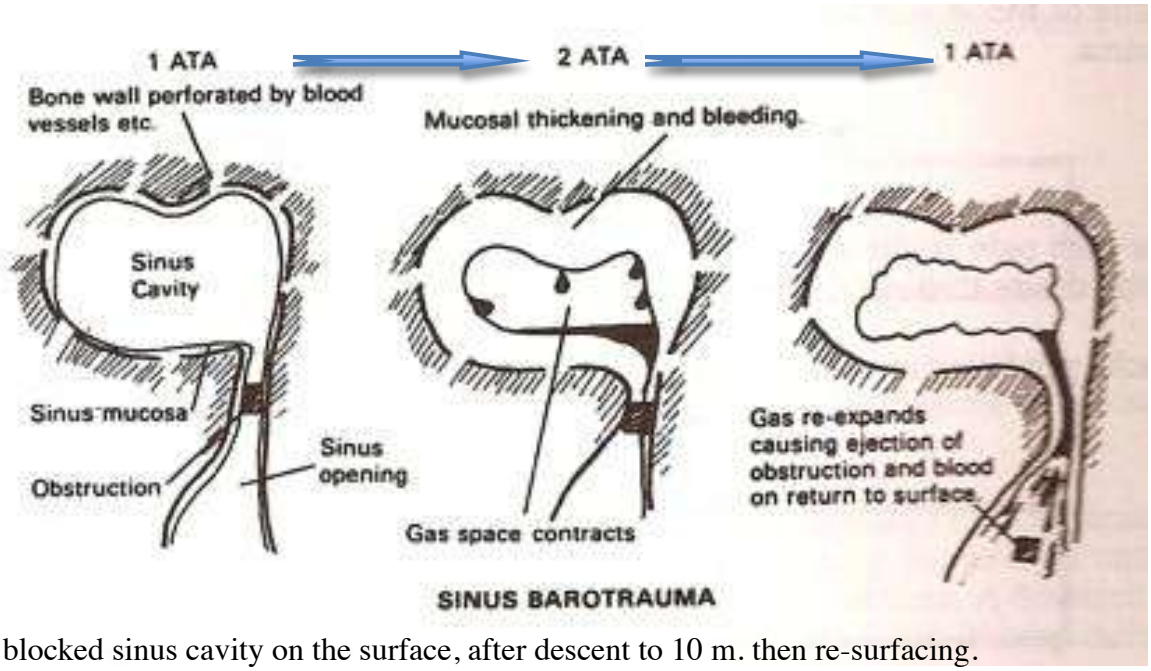

Fig. 10.2 A blocked sinus cavity on the surface, after descent to 10 m. then re-surfacing.

This tissue fluid and blood, which may take days or weeks to absorb, represents a rich nutrient medium for bacterial growth, promoting sinus infection (see Chapter 28).

During ascent, blood and tissue fluid from the sinus barotrauma of descent may be discharged into the nose or back of the throat by the gas expanding in the sinus, causing an apparent nose bleed from the same side as the injured sinus. Alternately, it may be spat out or swallowed.

If the sinus opening becomes obstructed during ascent, the expansion of gas flattens the sinus lining against its bony wall, causing pain and injury to this delicate tissue. This is called sinus barotrauma of ascent.

Sinus barotrauma of descent is more common than ascent, but they often coexist.

CLINICAL FEATURES

Sinus Barotrauma of Descent

This condition usually presents during descent with a sensation of pressure, developing into a pain in the region of the affected sinus. It is usually felt over the eye (frontal or ethmoidal), the cheek bone (maxillary), or deep in the skull (sphenoidal) depending on which sinus is involved.

Maxillary sinus barotrauma can also present as pain in the upper teeth.

The pain may settle during the dive, as blood and other fluids equalise the pressure differential, or it may recur as the damaged sinus causes a dull pain or headache afterwards.

A small amount of blood issuing from the nose during or after ascent is a frequent accompaniment of sinus barotrauma.

Occasionally the maxillary nerve may be involved, causing numbness over the cheek.

Sinus Barotrauma of Ascent

This presents with pain in the affected sinus during and after the diver's ascent. Bleeding from the sinus frequently drains through the nose or can be spat out.

Severe headache persisting or developing hours after the dive suggests either inflammation or developing infection (i.e. sinusitis) or sinus tissue damage.

Rarely, the bony walls of the sinus may rupture, with the expanding gas passing into the eye socket (orbital emphysema), the brain cavity (pneumocephalus) or tracking elsewhere. Any such severe manifestation must be treated as a medical emergency

TREATMENT

Any case of suspected sinus barotrauma accompanied by headache after a dive requires medical assessment, because decompression sickness and many other conditions can also present as headache. See Chapter 32.

Normally sinus barotrauma resolves without any treatment. Significant bleeding into the sinus may drain more rapidly if topical nasal and oral decongestants are used.

The diagnosis may be confirmed by X–Rays, CT scans or MRI (preferred) of the sinuses. The sooner this is done after the dive, the more likely it is to demonstrate the pathology.

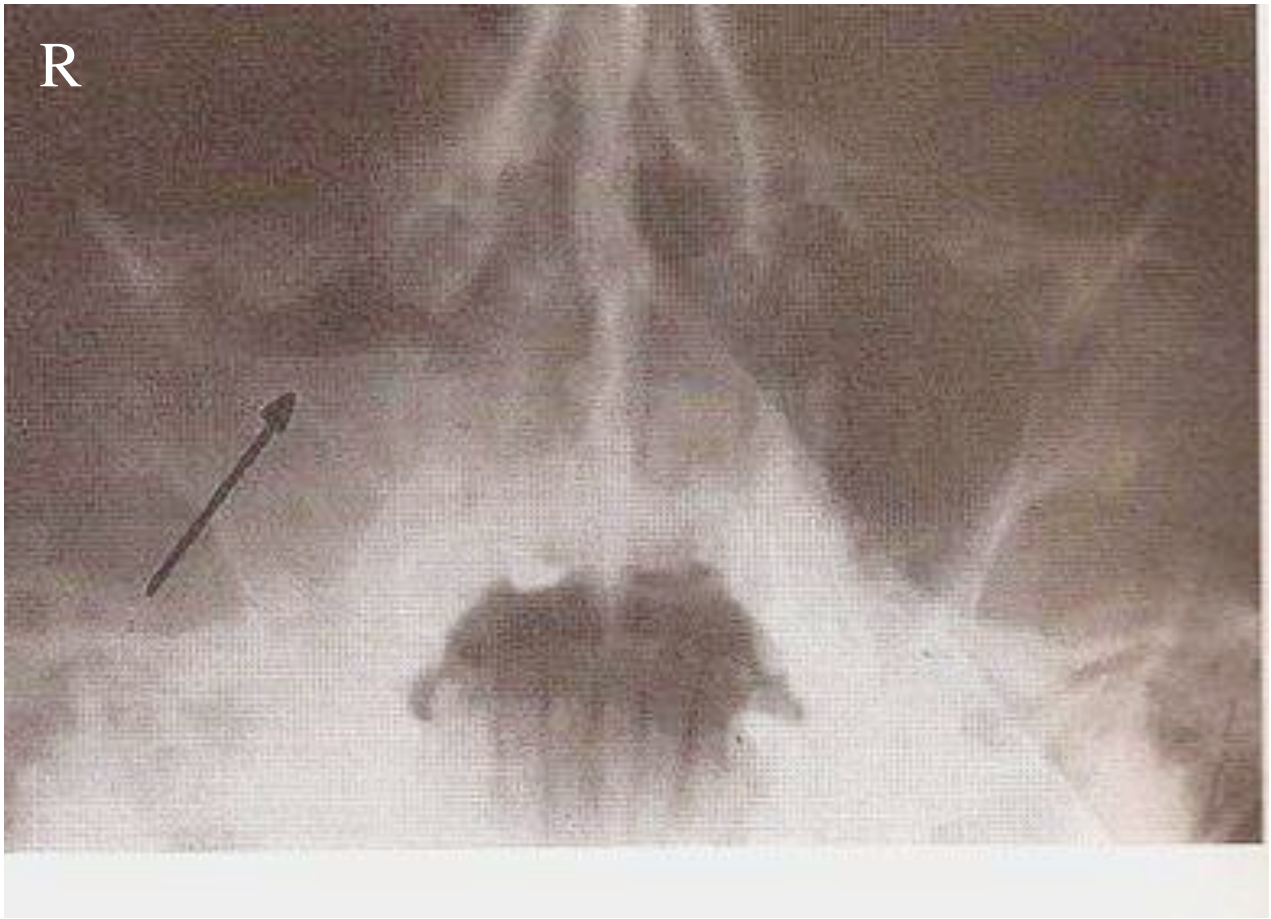

Fig. 10.3
Sinus x–ray showing fluid level (opaque area almost filling cavity below right orbit) in right maxillary sinus after barotrauma of descent. Left sinus cavity appears clear and filled only with air (black).

Increasing pain in the sinus, with fever or malaise developing after the dive suggests infection which is treated with decongestants and antibiotics.

Diving and flying should be suspended until the condition has resolved, usually from 2–10 days.

PREVENTION

Active and frequent middle ear equalisation, using positive pressure techniques such as the Valsalva, fortuitously assists by forcing air into the sinuses during descent and preventing barotrauma of descent. The "equalising ahead of the dive" technique is applicable (Chapter 8).

Diving should be avoided if the diver is suffering from any upper respiratory tract infection, to reduce both the risk of barotraumas and the infection complications. Smoking and allergic nasal congestion (hay fever) increases the risk of sinus barotrauma by obstructing the sinuses. A deviated nasal septum may also contribute to the development of sinus barotrauma, and if so, it can be surgically corrected.

Not all patients with chronic sinusitis need avoid diving. If the ostia is open and there are no polyps or other obstructions, the sinuses may get a good "washout" with air travelling into and out of the sinuses as the diver descends and ascends. This medical procedure is free to divers.

Nasal decongestants used at the time of diving tend to reduce the congestion of the sinus ostia (at least at the nasal end), but may not prevent sinus barotrauma of ascent. For this reason they should be avoided. It is better for the diver to be prevented from descending (sinus barotrauma of descent) than to be prevented from ascending (sinus barotrauma of ascent). See Chapter 37, last page, for a discussion on the effects of drugs used to prevent sinus barotrauma.

Repeated minor sinus barotraumas can result in progressive scarring of the ostia, causing obstruction and intractable sinus barotrauma. Then the diver has to choose between ceasing his diving career or problematic endoscopic surgery.

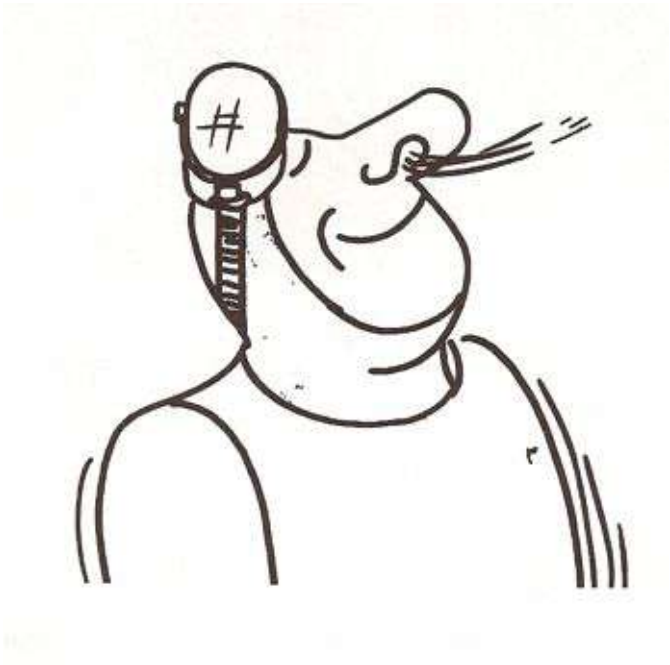

Fig. 10.4

Chapter 11

PULMONARY BAROTRAUMA

(LUNG BAROTRAUMA)

Pulmonary barotrauma is lung injury caused by pressure changes (Boyle's Law again). In divers it can occur on ascent or descent. Barotrauma of ascent is relevant to scuba diving, and barotrauma of descent to free diving (breath-hold).

PULMONARY BAROTRAUMA OF ASCENT

("BURST LUNG")

This is second only to drowning as a cause of death in young recreational scuba divers.

The lungs of a male diver normally contain about 6 litres of air, contained in the alveoli and air passages. If a diver takes a full breath at 20 metres (66 ft.) and returns to the surface, that 6 litre volume expands to 18 litres since the pressure at 20 metres is 3 ATA and at the surface, 1 ATA.

In this situation, to avoid over distension of his lungs, the diver must exhale 12 litres of air (measured on the surface) before or during his ascent. If he does not exhale this air, the expanding gas will distend his lungs, like a balloon, and even some normal lungs will rupture if they are distended more than 10%.

When near the surface, this 10% distension can be produced by an over-pressure of about 80mm Hg. - equivalent to the pressure difference between one metre depth (less than 4 ft.) and the surface, making pulmonary barotrauma a real possibility even for a scuba diver in a swimming pool. Divers have died from pulmonary barotrauma in shallow swimming pools

Even if the diver does exhale correctly, he can still encounter this problem if there is some obstruction to the venting of air from some part of the lung i.e. some form of lung pathology.

Like all other barotraumas, this is more a disease of the shallows, than of great depths. Gas volumes change more nearer the surface, for each alteration in depth.

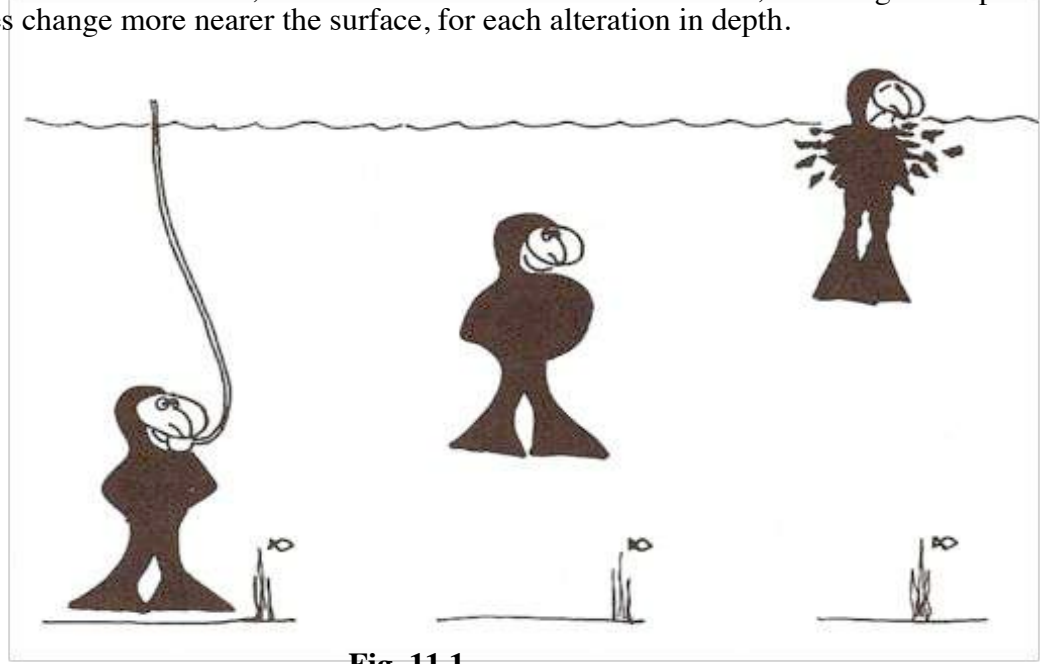

Fig. 11.1

Clinical Features of Pulmonary Barotrauma

If the lungs rupture due to excessive volume expansion, any or all of four consequences can follow:

- **Lung tissue damage**
- **Emphysema (gas in the tissues)**
- **Pneumothorax (gas in the chest cavity)**
- **Air embolism (gas bubbles in the blood)**

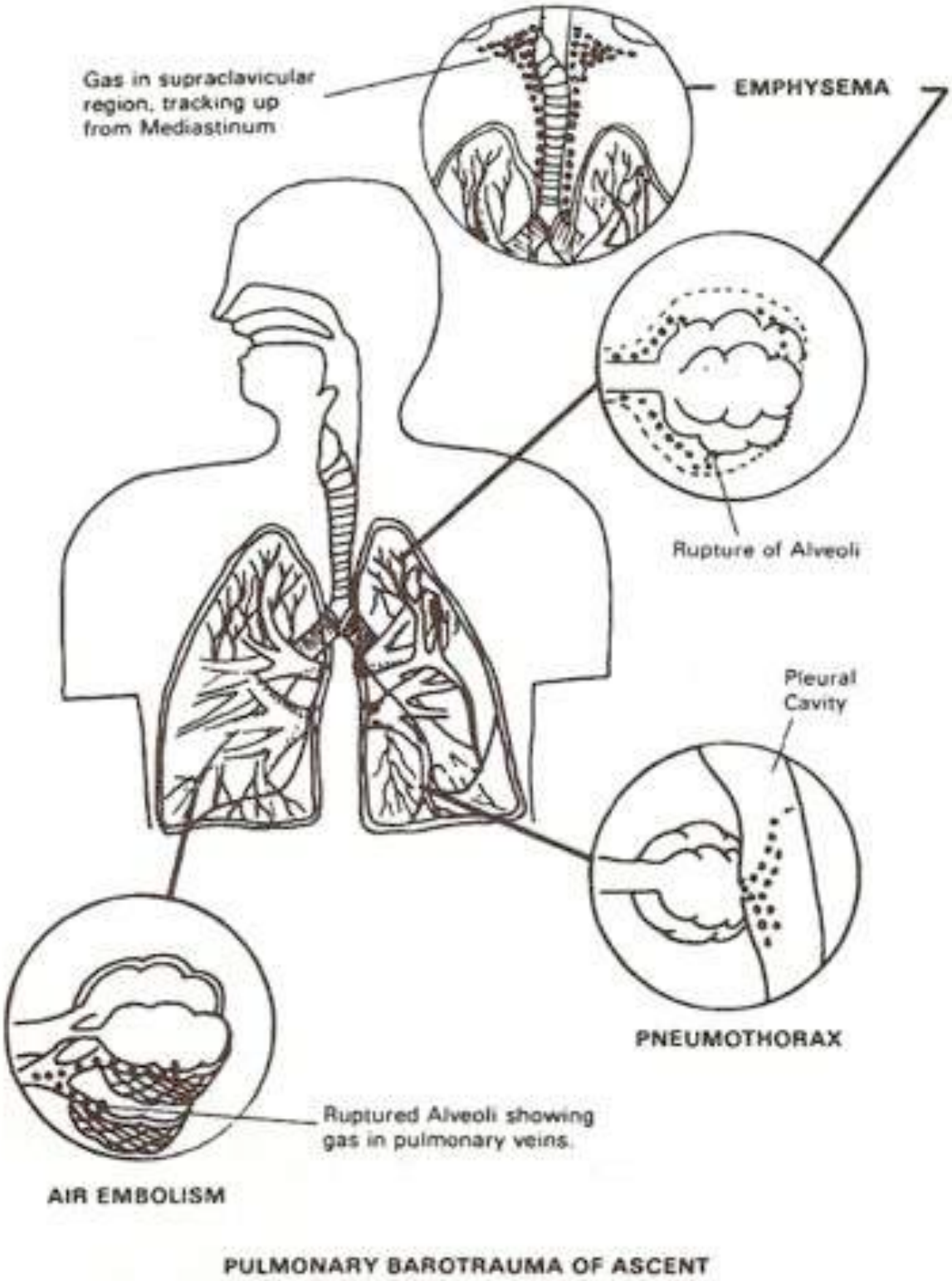

Fig. 11.2

The various clinical manifestations of a ruptured lung on ascent.

Lung Tissue Injury

If the lungs are over distended, generalised tearing of the lung tissues with severe diffuse damage to the lung structure is likely. Bleeding, bruising, and generalised destruction to the lungs causes severe breathing difficulties.

❑ Clinical features.

Shortness of breath, pain when breathing, coughing, coughing-up blood, and shock are the principal manifestations. Death may follow rapidly.

❑ Treatment.

The diver should be examined and treated for other manifestations of pulmonary barotrauma. Lung tissue damage alone has no specific first aid treatment apart from basic resuscitation measures (see Chapter 40 and 42). The patient should be given oxygen and taken immediately to hospital.

Surgical Emphysema

Tearing of the alveoli allows gas to escape into the tissues of the lung. Air tracks along the lung tissues to the **mediastinum** in the midline. From here it migrates into the **neck** or, in severe cases, tracking around the heart sac (pericardial sac) or even into the **abdominal cavity.**

If the diver has performed a long or deep dive and still has a nitrogen load in his tissues, nitrogen will continue to diffuse into these air spaces to expand them over the next few hours, with increasing symptoms.

The presence of the air in the tissues causes damage by compressing the blood vessels, nerves, larynx, or oesophagus. In severe cases air can compress the heart, causing malfunction.

❑ Clinical features.

It may take some time for symptoms to develop, as the air migrates slowly through the tissues. Air in the mediastinum and around the heart may cause **chest pain** and **shortness of breath.** Air in the throat leads to **voice changes** (the voice developing a "tinny" or "brassy" note), shortness of breath and/or **swallowing difficulties.** A **"crackling sensation"** may be felt under the skin around the neck – and especially just above the collar-bones (supra-clavicular space). It feels like "rice bubbles beneath the skin" or "cellophane paper", on pressing. The diver may complain of a sensation of **fullness in the throat.**

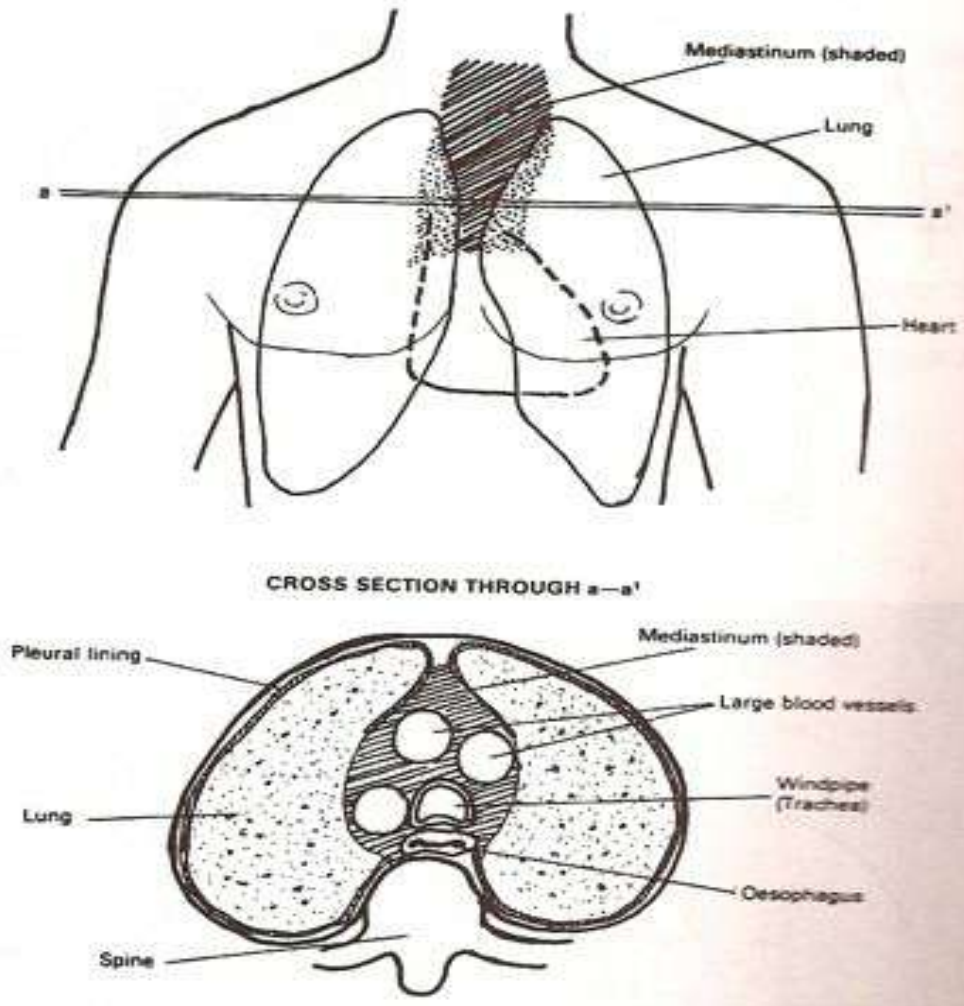

Fig 11.3

The location of the mediastinum can be seen from this frontal view and cross section. It is located deep in the chest between the lungs and above the heart, and its connection with the neck tissues can be noted.

❑ Treatment.

The diver should be examined and treated for the other manifestations of pulmonary barotrauma. Mild surgical emphysema alone responds to **100% oxygen.** (see Chapter 40). This causes a diffusion gradient for nitrogen (between the air space and the nitrogen-free blood) which eliminates the air bubbles. If not treated, the condition will slowly resolve, but it may last many days.

Severe surgical emphysema, especially if causing compression of the airway or blood vessels, will respond to **recompression** in a recompression chamber, compressing the emphysema. If air is breathed, more nitrogen may diffuse into the tissues, making the surgical emphysema even worse when the diver is decompressed. Breathing oxygen, especially under pressure in a chamber, produces a diffusion gradient of nitrogen out of the air spaces, with rapid resolution.

Pneumothorax

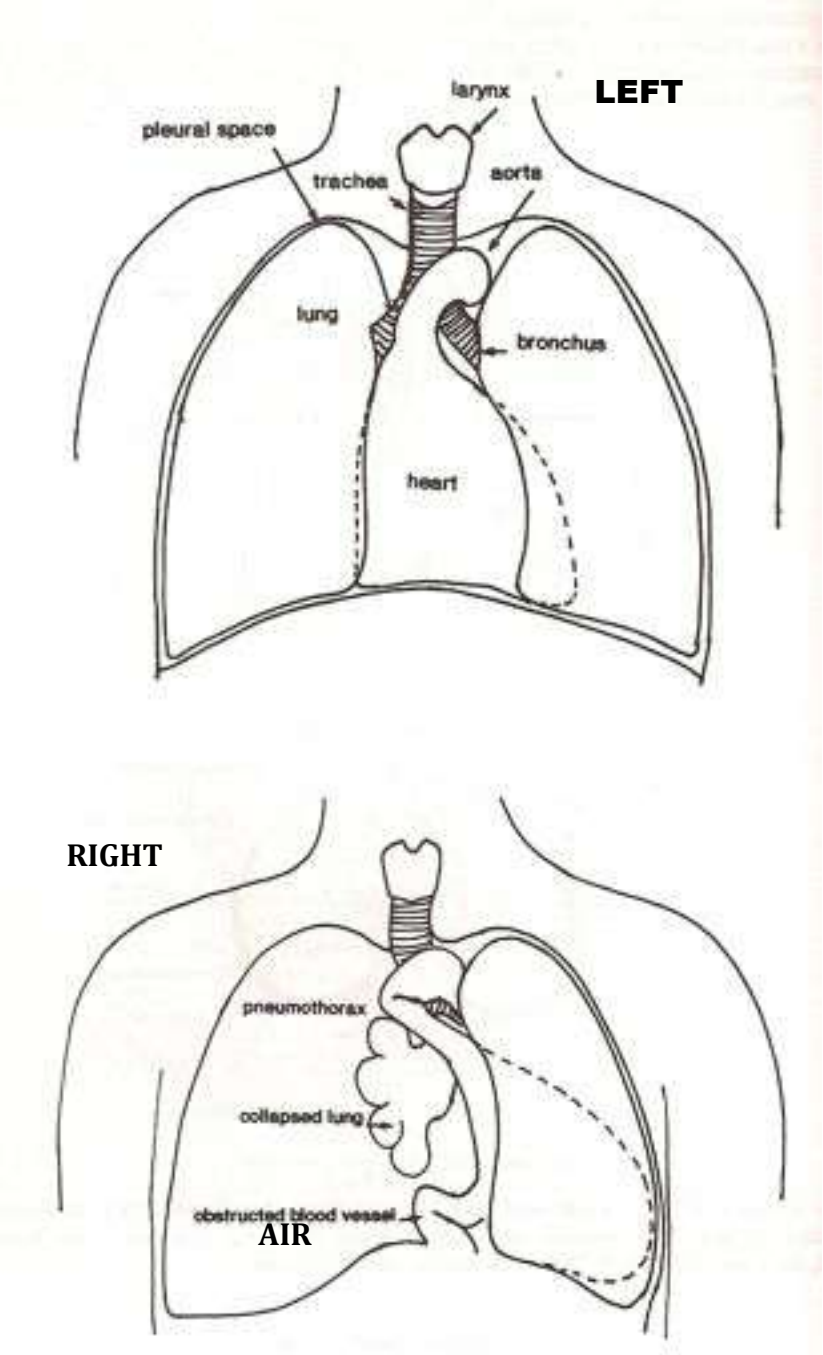

If the lung ruptures near its surface, air gains access to the pleural space, between the lung and the chest wall (pneumothorax). The elasticity of the lung causes it to collapse like a burst balloon and the lung tissue within the chest cavity is replaced by an air pocket.

The air pocket cannot escape.

Fig. 11.4

These schematic diagrams of a normal chest (as seen on x–ray) at the top, and a pneumothorax below, with collapse of the right lung. As air pocket in the right side expands with further ascent, it pushes the heart and midline structures towards the left side of the chest, causing a "tension" pneumothorax on the right.

Occasionally a valve effect allows air to pass from the air passages into the pneumothorax but prevents its return. As more and more air collects in the pneumothorax, the pressure in the thoracic cavity rises and forces the contents of the chest (including the heart and lungs) to the opposite side. This is called a **tension pneumothorax** and its effect on cardiac function is catastrophic and rapidly fatal if the air is not released.

If the lung rupture occurs at depth, the air in the chest cavity expands with ascent (Boyles Law) and this may also cause a tension pneumothorax

Bleeding may take place into the pneumothorax, leading to a **haemo-pneumothorax.**

❑ Clinical features.

A pneumothorax is usually heralded by **chest pain**, often made worse by breathing, and causes **shortness of breath**. Respiration becomes rapid and the heart rate increases.

With a **tension pneumothorax,** as the mediastinum is pushed to the opposite side, the trachea can be felt to be displaced to that side. The patient becomes increasingly short of breath and may become cyanotic (blue) and shocked. The pulse is difficult to feel as the blood pressure falls.

With severe cases of burst lung, a pneumothorax will be evident very soon after the diver reaches the surface, but in milder cases, the symptoms of pneumothorax may be delayed for many hours. Symptoms may be brought on by coughing or altitude exposure (e.g. mountain range, travel in aircraft, more diving).

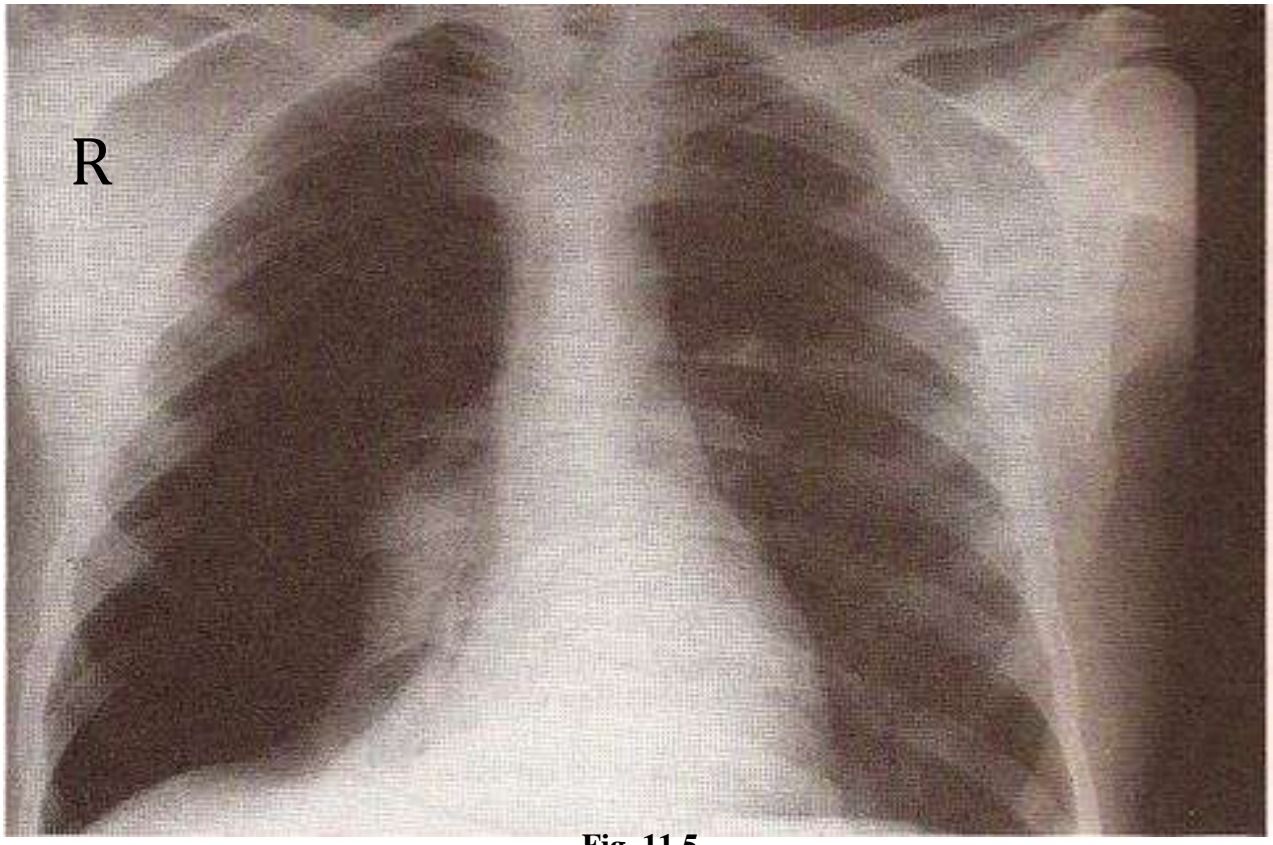

Fig. 11.5

X–ray of diver's chest after suffering pulmonary barotrauma of ascent with a right sided pneumothorax. The right chest cavity appears "black" due to its being filled with air and the collapsed lung (white) can be seen low, near the midline.

❑ **Treatment.**

A pneumothorax requires **urgent medical attention**. The extent of lung collapse is assessed clinically and confirmed by a chest x-ray. A large pneumothorax is treated by placing a **tube into the pleural air space** and connecting it to a one-way valve such as a Heimlich valve or an underwater drain. This allows air out of the pneumothorax but prevents its return. The placement of tubes in the chest is usually beyond the capability of untrained personnel as there are important structures, like the heart, which can be injured in the process. After a period of hours or days the tear in the lung usually heals and the lung slowly re-inflates.

A minor pneumothorax (less than 25% lung collapse) may be treated by the diver **breathing 100% oxygen** (see Chapter 40).

A **tension pneumothorax** is a medical **emergency**. The pressure in the pneumothorax must be relieved by the insertion of a needle or tube through the chest wall, into the pneumothorax.

If the diver is aware of the possibility of a pneumothorax, he may be able to alert a physician to the possible diagnosis if any of these clinical features are present. The physician may release the air by inserting a needle into the second intercostal space in the mid-clavicular line, or through the 5–6 intercostal space in the mid-axillary line. Both have potential complications.

Air Embolism

When the lungs rupture, tears in the alveoli walls (and contained blood capillaries) can allow air to enter the blood circulation. This air is conducted to the left side of the heart, from whence it is pumped through the arterial circulation.

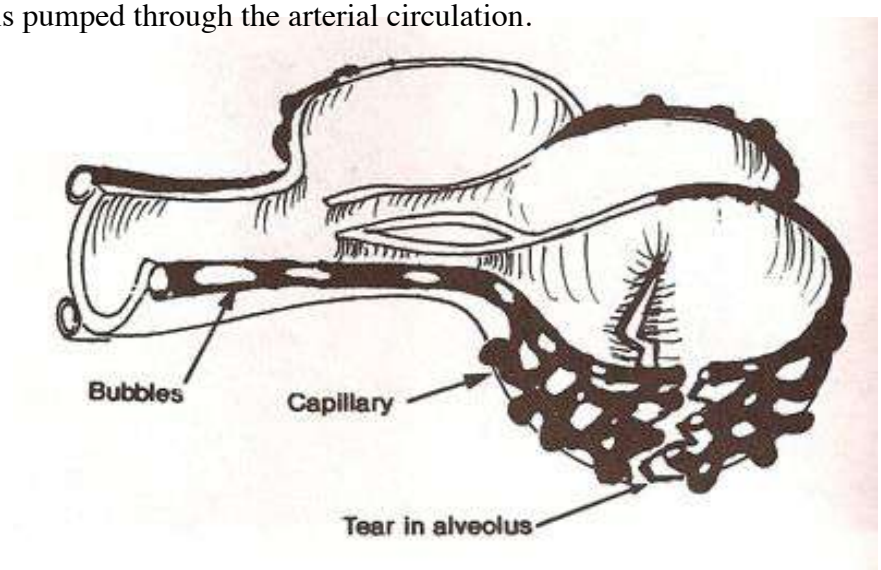

Fig. 11.6

Diagram of a ruptured alveolus and capillary vessel from pulmonary barotrauma of ascent. Air bubbles (emboli) are entering the veins carrying blood back to the left atrium of the heart.

The air bubbles obstruct or damage blood vessels in vital organs such as the heart and brain, leading to impairment of function, serious disability or death.

❑ Clinical features.

Symptoms present abruptly, usually immediately or within 10 minutes of the diver reaching the surface. Air bubbles lodging in the brain may cause **loss of consciousness**, **fits** or **confusion**, a pattern of symptoms similar to a "cerebral stroke". This is called **Cerebral Arterial Gas Embolism** or **CAGE**.

There is often **loss of function** of parts of the causing :

- **disturbances of sensation** such as **numbness** or **tingling**
- **disturbances of movement** including **paralysis** or **weakness**
- **disturbances** of **vision**
- **disturbances** of **speech**
- **disturbances** of **balance** or **co-ordination**
- **disturbances** of **intellectual function**

Air bubbles lodging in the coronary arteries which supply the heart with blood may lead to symptoms resembling a **heart attack** including **chest pain, shortness of breath** and **palpitations**. Air bubbles lodging in the circulation of the skin cause white or purplish patches (**marbling**).

If a diver surfaces after a deep dive and develops disturbances of brain function it is also possible that he might have **cerebral decompression sickness.** It may not be possible immediately to distinguish between air embolism and cerebral decompression sickness, now both called **acute decompression illness**. Fortunately the initial treatment is similar for both (see Case History 33.5).

❑ Treatment.

Air embolism causes hypoxic damage by obstructing important blood vessels with air bubbles. **Recompression therapy** in a chamber reduces the size of the bubbles and allows them to flow on, into smaller less-important blood vessels. The bubbles ultimately pass into the venous system, may be trapped in the lungs, and are eliminated as nitrogen diffuses out of them.

Divers who develop air embolism after free ascent or in submarine escape training, where a recompression chamber is available nearby, are immediately recompressed to reduce the air bubbles from their original size and so reduce the brain damage. The depth chosen will depend on the severity of the condition. High oxygen pressures are administered to reduce the bubble further. Otherwise, **transport** to a chamber must be arranged urgently.

There is some controversy about the best way to **position** the patient. Divers used to be taught to place the patient in the 30 degrees head down position, to keep the rising bubbles away from to the brain. This caused difficulty with resuscitation and transport however, and the increased venous pressure in the head worsened the cerebral oedema (brain swelling) which accompanies injury. It is no longer advised.

A more reasonable approach is to place the patient **horizontally**, **on their side** (the left side is theoretically preferable but this is probably not critical) without a pillow. This will place the head slightly lower than the heart. The patient is likely to be unconscious or drowsy and this position is also good for patency of the airway. This is often called the **coma position**.

The patient should **not be allowed to sit up or stand up** once this position is adopted as there are some case reports of divers rapidly deteriorating after sitting up, even some time after the barotrauma event. This may be due to bubbles rising to the brain.

The patient should be given **100% oxygen** to breathe. After an hour of breathing 100% oxygen, it is probably safe to allow him to assume a more comfortable position.

If the patient is unconscious the basic life support (**BSL**) priciples take precedence and should be followed (see Chapter 42).

The **other complications** of burst lung, such as pneumothorax or emphysema, must also be looked for and treated, if present.

The definitive treatment is **recompression therapy** in a well equipped chamber. Urgent transport is necessary in order to minimize brain or other essential organ injury. Unfortunately, even in the best facilities, full recovery is not always possible.

Predisposing Factors

❑ Breath-holding.

This may be due to failure to read "*Diving Medicine for Scuba Divers*", **panic, ignorance, forgetfulness or spasm of the larynx after inhalation of water.** The first rule that any aspiring scuba diver should learn is to exhale ("blow bubbles") during ascent. Breath-holding during ascent can lead to excessive distension of the lungs, and their rupture.

❑ Air trapping.

Anything preventing air from leaving all or part of the lungs can lead to pulmonary barotrauma.

Several factors may predispose to air trapping. Obstruction of the bronchi is frequent in; **asthma, acute and chronic bronchitis, respiratory tract infections.** This obstruction may allow air to enter the lungs but restrict exit of air – a ball valve effect. Other conditions which can cause this include; **tuberculosis (T.B.), tumours of the lung, calcified glands, cysts in the lung and emphysema. Heavy smoking** may cause **mucous plug obstructions.**

❑ Disorders of lung compliance.

Lung compliance is a measure of the stretching ability of the lungs. One published study investigated pulmonary barotrauma in Navy divers who had correctly exhaled during ascent and were previously medically fit. Studies of the lung compliance of these divers showed their lungs to be more "stiff" than normal, and therefore presumably more prone to tearing when slightly over-expanded.

Divers with **scars or fibrosis in the lungs** may have localised reduction in lung compliance which may cause shearing forces and tearing in these areas. Fibrosis of this type may be found after inflammatory lung disease such as **sarcoidosis, tuberculosis, lung abscess** or even some **severe pneumonias.**

Lung tearing has occasionally been described in **breath-hold divers.** These divers developed pneumothorax and mediastinal emphysema during breath-hold dives — the tearing of the lungs being caused by the diver taking very large breaths, with excessive respiratory pressures. On investigation, these divers were found to have relatively small lungs and relatively large chest cavities. Full expansion of the chest cavity in these divers led to over-expansion of the lungs and subsequent tearing.

❑ Rapid ascents

Any partially obstructed airway may restrict airflow. This can be overwhelmed by the massive volume changes which occur during rapid ascents. This risk can be reduced by adhering to the recommended slower ascent rate of 9 metres or 30 ft per minute, upon which most new decompression tables are based.

A slow ascent rate (9 metres or 30 feet/min, or slower), as in most Diving Manuals) is strongly recommended by the authors.

This rate may also help to reduce the risk of developing serious decompression sickness. The bottom time should be reduced accordingly

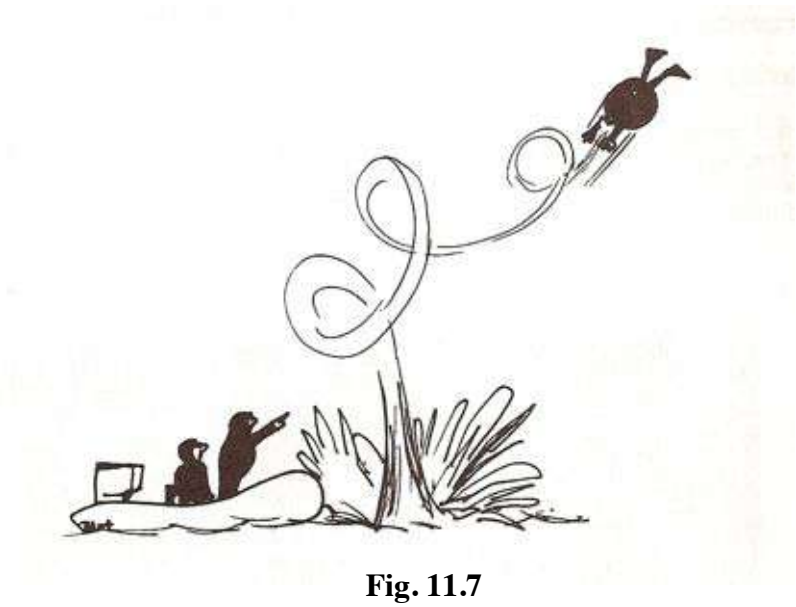

Fig. 11.7

❑ **Emergency ascents.**

The sudden failure of gas supply, especially at considerable depths, tends to alarm even the most sanguine diver. The subsequent emergency ascent is often undertaken with rapid ascent rates. Breath-holding due to anxiety together with a rapid gas expansion, greatly increases the likelihood of pulmonary barotrauma.

❑ **Free ascent training (or "Emergency swimming ascent training").**

A "free ascent" is a manoeuvre in which the diver breathes from compressed air equipment, takes a breath and then returns to the surface without breathing further from the equipment. Naturally, he must exhale to exhaust the expanding gas – but he may still encounter several problems during ascent. Most divers, aware of the dread consequences of breath-holding, tend to exhale excessively and may run out of breath before surfacing.

In many navies free ascent training, pulmonary barotrauma, salt water aspiration syndrome and near drowning were not uncommon accompaniments of these exercises. Thus, a **recompression chamber** and a **specialised diving physician** had to be available immediately adjacent to the practice ascent site. Perhaps because they are of less value, civilians do not impose the same standards.

Unless closely supervised, the rate of ascent is usually excessive, especially from greater depths since the diver knows that he only has one lung full of air to sustain him until the surface is reached. In fact he has the equivalent of 3 lungs full of air when compressed air is breathed at 20 metres depth. The excessive rate of ascent causes rapid gas expansion and damage if airways are partially obstructed.

This outmoded training technique was designed to prepare the divers to cope with an "out-of-air" situation in days before contents gauges, octopus rigs, and emergency air supplies were common.

Unfortunately deaths from this procedure made it a questionable practice**.** Also, the tendency of divers to over-inhale before commencing the ascent made the procedure more hazardous and not at all similar to the genuine out-of-air situation, where lack of air is usually detected after exhalation. Thus, the lungs in a real "out-of-air" situation are not fully inflated. As this 'real situation' usually happens without any inspiratory capability, it is presumably safer.

❑ **Submarine escape.**

Escape from a sunken submarine usually involves a rapid, buoyancy-assisted free ascent. This technique is practiced by most navies from depths of 20–30 metres in especially made submarine escape training facilities (SETF). Ascent rates are very fast and pulmonary barotrauma can occur in spite of good training and thorough preliminary medical examinations.

The emphasis on escape procedures demonstrates the optimistic outlook of submariners since the submarines often operate in water depths which exceed the crush depth of the submarine's hull.

Fig. 11.8

The Royal Australian Navy's S.E.T.F. in Western Australia.

❑ **Buddy breathing.**

Sharing a single regulator with a buddy, when his regulator fails, is not easy to achieve and unsuccessful attempts at buddy breathing, especially during ascent, are often followed by the diver abandoning the procedure and undertaking a free ascent to the surface. Either diver may tend to over-inhale prior to handing over the regulator and then breath-hold during ascent, while waiting for its return. These conditions are conducive to pulmonary barotrauma.

Employing an octopus regulator (a spare regulator with a longer hose) or another air supply system, is now almost universal, to facilitate safer buddy breathing. See Chapter 5.

Prevention of Pulmonary Barotrauma of Ascent

❑ **Medical fitness.**

Divers should be carefully screened to ensure there are no respiratory problems that predispose to pulmonary barotrauma (asthma, fibrosis, cysts, pneumothorax, infections, etc). Divers who do burst their lungs and survive are much more likely to have recurrences, which are then more severe – often with fatal consequences. Thus an episode of pulmonary barotrauma usually precludes further diving.

❑ **Diving techniques.**

Divers are advised to avoid situations which could lead to them having to perform an emergency free ascent. Such situations include greater **depth**, **reduced air supply**, **overweighting** and/or **excessive buoyancy**.

The use of well maintained **good quality equipment** (e.g. regulator), a **contents gauge**, an **octopus rig** or better still, and an **emergency air source** (to avoid the need for buddy breathing) are common sense measures which can be employed.

It is important for scuba divers to remember to **keep breathing normally** at all times, as a relatively small ascent in shallow water while the diver is holding his breath can lead to over-pressure (distension) of the lungs and pulmonary barotrauma. "**Skip**" (controlled or reduced) **breathing** is dangerous because it increases the time when a diver is not breathing.

Pulmonary barotrauma is a not uncommon accompaniment of "**free ascent training**" (also called "**emergency swimming ascent training**"). Included in this category is "**ditch and recovery**" drill, where the diver performs a free ascent as he returns to the surface after ditching his gear on the bottom. The greatest volume changes due to Boyle's Law take place near the surface, so that free ascents from even shallow depths are not safe.

The concept of training novice divers in "free ascent" technique is controversial. Obviously it is desirable for all divers to be familiar with safe "free ascent" principles in case of equipment failure. However, if a diver develops a serious air embolism after a free ascent, a fatal outcome is likely, unless a recompression chamber with an experienced diving physician is available at the site of the dive. This facility is rarely available in sport diver training. Even a few minutes delay in instituting recompression has a significant negative influence on the outcome of treatment.

Studies conducted on submarine escape trainees in Sweden showed an almost 4% incidence of EEG (electroencephalogram or "brain wave") changes in these divers indicating sub-clinical brain damage, presumably due to minor air embolism. Studies of free ascents by trainees in the U.S. Navy showed an incidence of pulmonary barotrauma of 1 in 3000.

❑ **Out of Air (OOA) and Low on Air (LOA) Situation.** (See Chapter 34.3)

All scuba divers should surface with at least 50 ATA air remaining. If not, the dive procedure needs reappraisal. Most of the OOA problems are due to failure to comply with this rule.

Without an air supply, hypoxia is inevitable. Whatever the excuse, running out of air underwater is hazardous. Alternately, failure to surface is uniformly fatal. Thus the diver should have a planned "bail out" procedure to be used if OOA, to reach the surface.

The more effort used in an OOA swimming ascent, the more the diver risks unconsciousness from hypoxia, panic and carbon dioxide build up.

From the <u>OOA diver's</u> point of view, a procedure that could be followed, is as follows (in priority):

1. At the first sign of any problem with the air supply, signal to a buddy.

Do not chase after him unless he is very close or is between you and the surface. Any unnecessary effort will further deplete your already limited air supply.

2. Commence a controlled ascent to the surface.

3. Unless in an enclosed space (a wreck, cave, etc.), ditch the weight belt – or unclasp it and hold it away from your body, so that in the event of more problems arising, it will be ditched automatically.

4. If the buddy responds to the signal by offering an air supply to you, in the form of a separate regulator, then remove your non-functioning regulator and accept his.

Unless you are well trained in buddy breathing and have practised this frequently with the rescuer, then it is usually not worth while to attempt an ascent while sharing a single regulator. Occasionally this may be necessary (such as when one is in an enclosed area).

5. If there is no secondary air supply available either from your own equipment or your buddies, then leave the regulator in your mouth, as some more air may become available due to the decreasing environmental pressure with ascent. It may also reduce the likelihood of salt water aspiration.

6. It may be necessary, depending on your equipment, to inflate the buoyancy compensator, but this is often neither prudent nor possible, because of the inadequate air supply. Inflating

the BC takes excessive time at depth and could also result in an accelerated ascent as the air in the BC expands, with ascent. More reliable and consistant buoyancy is obtained by ditching weights.

7. In the OOA situation there is little time available – but usually adequate if it is not wasted. Unnecessary and difficult underwater dialogue, especially followed by reassurance from narcotic rescuers-to-be, does not compensate for an air supply. Contacting everybody (no matter how important they may be) and evaluating the situation, is a topside indulgence. In some circumstances it is necessary to perform a rapid ascent, risking the possibility of decompression sickness and barotrauma, in order to avoid the inevitability of drowning .

If totally OOA, it is often preferable to ditch the weight belt and surface rapidly, exhaling if and as required.

From the **buddies** point of view it is prudent to :

1. Supply a secondary air source to the OOA diver. This is usually done via an octopus regulator or an alternative air source. In a panic situation, this may be done by giving him your own regulator, because this is what he grabs, because he can see it and knows it is functioning, and for you to use one of the secondary regulators.

2. Control the ascent of the OOA diver, assisting this by ditching his weight belt. It is preferable not to ditch your own weight belt as it may be necessary for you to descend later. As your buddy is already OOA, this requirement will not be needed by him. If necessary to obtain adequate buoyancy, it may sometimes be necessary to ditch your weight belt also.

3. Once you reach the surface, retain a secure hold on the OOA diver, as he could lose consciousness from CAGE, within the subsequent few minutes, and sink. Also, once you are on the surface, inflate his BC, either by the inflator button or orally. This will ensure his buoyancy and reduce his anxiety.

4. Attract attention and assistance by the use of many of the techniques already referred to in this text e.g. whistle, flare, safety sausage, etc.

PULMONARY BAROTRAUMA OF DESCENT

(LUNG SQUEEZE)

There is a slight risk of pulmonary barotrauma during descent as well as ascent, although from a different mechanism.

A diver descending during a breath-hold dive will have the air in his chest and lungs progressively compressed in accordance with Boyle's Law. Eventually a lung volume is reached when the compression of gas can no longer be accommodated by a further reduction in lung volume, and is instead compensated by the engorgement of blood vessels in the lungs. The lung blood vessels have only a limited ability to distend, and can be expected to rupture once this limit is exceeded, causing pulmonary haemorrhage.

A rapid descent when standard dress equipment is used, or failure of a surface-supply gas pressure in the absence of an effective non-return valve, are also possible causes of pulmonary barotrauma of descent. It is theoretically possible whenever a surface-supply of air is used e.g. standard dress, surface supply from a compressor or compressed air tanks, or pumping the air supply from the surface in commercial devices.

It is more likely in breath-hold diving, but case reports of this condition are infrequent and poorly documented. It is more likely in deep, record breaking attempts. The theoretical basis of the condition was severely tested when a world record descent to beyond 200 metres was made by a breath-hold diver several years ago.

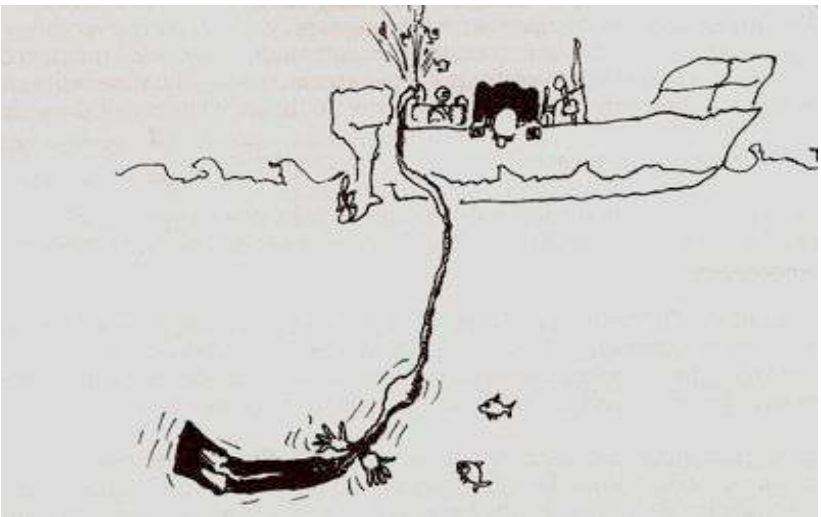

Fig 11.9

SSBA with no non-return valve, when the compressor stops

Chapter 12

OTHER BAROTRAUMAS

Barotrauma can develop wherever there is an enclosed gas space adjoining tissues. With descent, the space contracts, pulling tissue and blood into the space (implosions). With ascent, the space expands and disrupts tissue (explosions). Because gas changers are greater near the surface, barotrauma is more likely in the shallower depths.

The spaces may be within the body or between the body and the equipment. They include :

- **Facial (mask) squeeze**
- **Skin (suit) squeeze**
- **Body squeeze and "blow up"**
- **Gastro-intestinal barotrauma**
- **Dental barotrauma.**

Barotraumas dealt with earlier include lung, ear and sinus barotraumas.

FACIAL BAROTRAUMA OF DESCENT

(MASK SQUEEZE)

During descent the airspace inside a face mask is compressed and the contraction in volume of the gas space is accommodated by flattening of the mask against the face, and later by congestion of the facial skin and eyes.

It can lead to bleeding into the soft tissues under the skin and produce a characteristic **bruised** facial appearance under the mask area. The whites of the eyes may be grossly haemorrhagic ("**red eye**"). It may take 1–3 weeks to clear up.

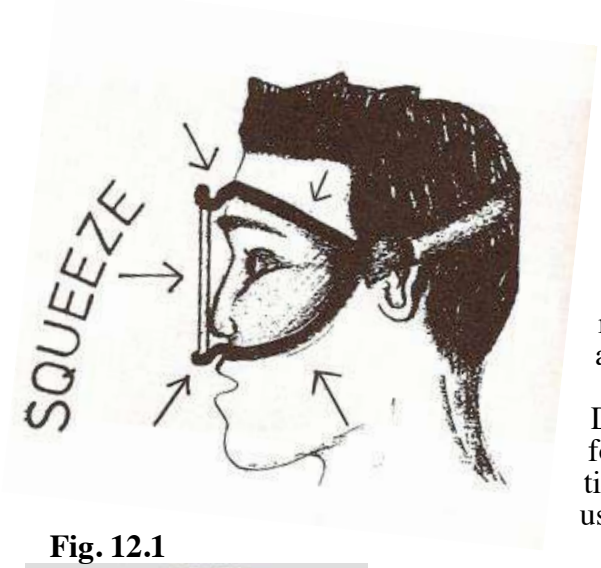

Fig. 12.1

This condition is easily prevented by exhaling into the face mask during descent, to equalise the mask with the water pressure. More cases have developed since rigid plastic face masks replaced the soft rubber ones. Expanding gas automatically vents around the edge of the face mask during ascent.

Divers using swim goggles run the risk of a similar form of barotrauma on a smaller scale involving the tissues around the eye. In the past, if goggles were used to dive, a method of equalising the space around the eye, such as that shown in diagram 12.2, was employed.

Fig. 12.2

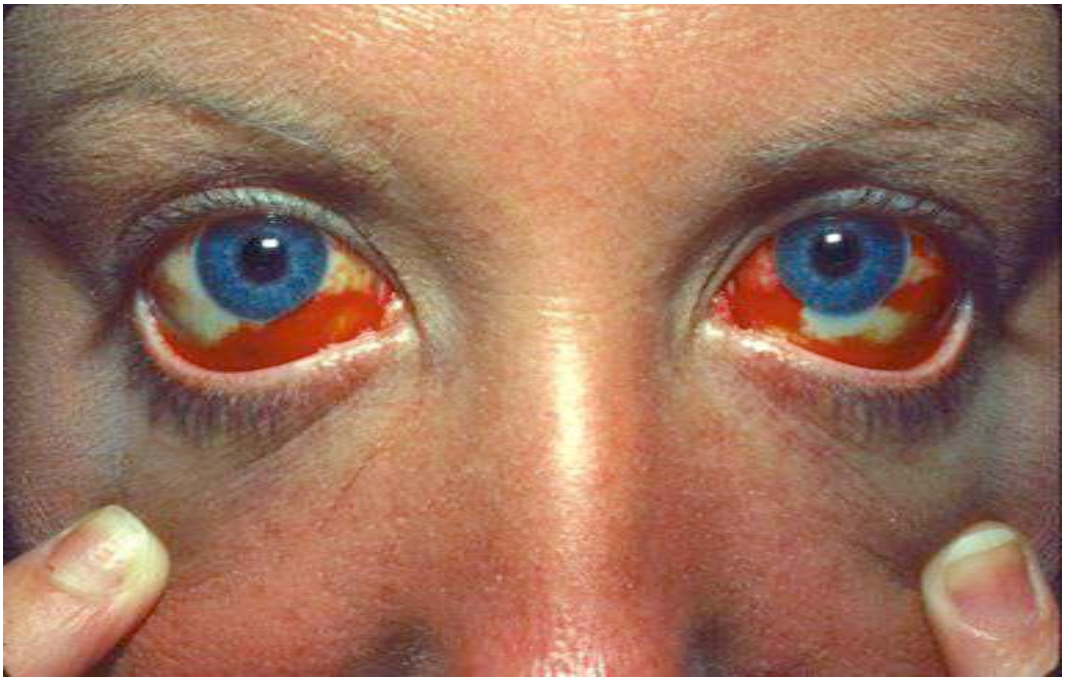

Fig. 12.3 Facial barotrauma from diving with goggles – note the bleeding into the sclera (white parts) of both eyes. Face masks can cause similar eye bleeds, but then the surrounding facial tissues are also bruised.

SKIN BAROTRAUMA OF DESCENT

(SUIT SQUEEZE)

Divers using a **dry suit** or a **loose wet suit** may experience this problem. During descent, pockets of gas can be trapped in folds under the suit. Where the suit has folds, the contraction of the gas space is accommodated by the skin being sucked into the space leading to strips or welts of bruising.

It may cause discomfort at the time. After surfacing the diver may notice bruising over the skin, corresponding to the folds.

BODY BAROTRAUMA OF DESCENT

(BODY SQUEEZE)

With the solid metal helmet used in standard diving there is a possibility of the diver descending and the pressure in the air hose not keeping up with the environmental pressure. If this occurs, the diver's body may be forced up into the helmet, and crushed.

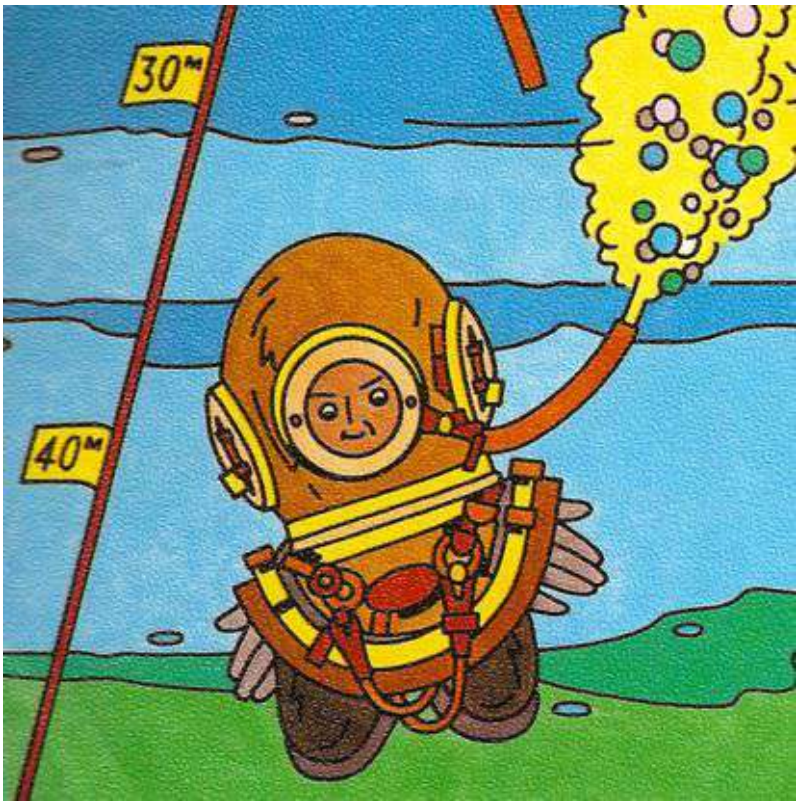

Fig. 12.4

This can also happen at a constant depth, when a non-return valve is not used (or is not functional) and the air supply fails. The only treatment for this bizarre injury is to wash out the helmet with a good antiseptic.

Even the modern plastic helmets, used in deep and helium diving, can cause minor variants of this condition.

SUIT BAROTRAUMA OF ASCENT

("BLOW UP")

With either dry suit or standard dress, the gas in the suit can expand with ascent, causing increasing buoyancy, more rapid ascent, etc. and a vicious circle develops where the diver may hurtle to the surface and be imprisoned in a balloon-like inflated suit. Special training and emergency procedures are needed for recreational divers who wear this equipment. As well as the physical injury that may result, other barotraumas and decompression sickness are likely.

Fig 12.5

GASTROINTESTINAL BAROTRAUMA

Gas is normally present in the gastro-intestinal tract. This finds its way into the atmosphere from time to time, as those who consume prunes, baked beans or cabbages will attest.

During a dive, gas may be swallowed when the diver equalises his ears, especially if in the inverted (head-down) position. It may accumulate in the stomach and gastrointestinal tract without initially causing any discomfort to the diver. During ascent however, this accumulated gas increases in volume, and can result in cramping colicky abdominal pain, belching and vomiting. Rare cases of stomach rupture have even occurred.

Divers are advised not to equalise their ears in the "head down" position.

Several cases of severe gastric discomfort have been reported during chamber dives when the divers drank carbonated beverages while under pressure. One amusing account relates the opening of a new hyperbaric facility which was toasted by champagne at 20 metres depth. The occupants were disappointed that the champagne appeared to be flat, but they drank with relish anyway because it tasted good. Their discomfort was exceeded only by their embarrassment during ascent as the gas in the champagne came out of solution and expanded in their stomaches.

DENTAL BAROTRAUMA

This uncommon form of barotrauma has on occasions been given sensational publicity, causing some divers to believe that they carry potential bombs set into their jaws.

Decayed teeth can occasionally contain a small air space which may lead to the tooth crushing inwards (imploding) during descent or fragmenting painfully (exploding) during ascent. The latter happens when there is an opening sufficient to allow air to enter during descent, but insufficient to allow it to escape during a fast ascent. As divers age, they develop small air spaces around their teeth, resulting in dental pain on descent. Because the air space is constant, the depth at which the barotrauma occurs also tends to be constant.

The explosive potential of this occurrence during descent or ascent has been overrated.

Diving within several days of a tooth extraction may occasionally allow air to enter the tissues through the tooth socket from the positive air pressure generated by breathing through a regulator. This results in air tracking into tissues around the face (**tissue emphysema**). This is rarely serious and is treated by the diver breathing 100% oxygen for several hours to eliminate the air.

Diving after any tooth extractions should be avoided until the tooth socket has healed — usually this takes about a week to ten days.

Chapter 13

DECOMPRESSION PHYSIOLOGY and SUSCEPTIBILITY

Decompression Sickness (DCS) is an illness caused by the effects of gas coming out of solution to form bubbles in the body after diving. It is due to the effect of Henry's Law (see Chapter 2) following diving exposures. Understanding decompression theories is difficult if not impossible, so the average diver may well bypass most of this chapter, if he is not technically inclined.

In sport divers the main gas formed in bubbles is nitrogen (N_2) because these divers almost always breathe air. However, the same principles apply to other inert gases, such as helium (He), which may be breathed by deep commercial and technical divers.

GAS UPTAKE

When a diver breaths air from scuba equipment at depth, N_2 is breathed at an increased partial pressure. Because gas diffuses from areas of high concentration (high partial pressure) to areas of lower concentration, N_2 is taken up from the lungs by the blood and transported around the body and into the tissues. The greater the depth, the greater the partial pressure of N_2, and therefore the amount of N_2 absorbed. Early in the 20th century, Haldane applied this concept to decompression.

The speed of N_2 distributing to the tissues depends on the their blood flow. Tissues with high metabolic needs such as the brain, heart, kidneys and liver receive most of the blood pumped from the heart. They will also receive most of the N_2 carried in the blood and will have a **rapid N_2 uptake.** Such tissues are termed "**fast tissues**" because of their fast N_2 uptake.

Because blood passing through the lungs immediately equilibrates with any change in inspired N_2 partial pressure, blood is the fastest tissue of all.

Other tissues such as ligaments, tendons and fat, with a relatively small blood flow, have a relatively slow N_2 uptake. These tissues are termed "**slow tissues**". Between the two are tissues of intermediate blood flow such as muscle. Some organs, such as the spinal cord, have both fast and slow tissue components. The rate of uptake of N_2 in a tissue is **exponential** i.e. it varies depending on the amount of gas already taken up by the tissue. As the tissue takes on gas, the uptake slows because the partial pressure gradient decreases.

The filling of a scuba cylinder is an example of an exponential process. When an empty cylinder is connected to a high-pressure source, the cylinder initially fills quickly, but the flow slows as the pressure in the cylinder increases and approaches that of the gas source.

The uptake of gas in any tissue is initially rapid but slows with time. Accordingly, it may take a long time for a tissue to become fully saturated with gas, but fast tissues become saturated sooner than slow tissues.

Since the exponential uptake takes a long time to reach completion, even if it starts rapidly, the concept of **tissue "half times"** is used to compare tissues. The half time is the time taken for a tissue to reach half its saturation level. A fast tissue may have a half time as little as a few minutes, while a slow tissue may have a half time of some hours.

GAS ELIMINATION

N_2 is eliminated in a reverse of the uptake process. As the diver ascends there is a reduction in the partial pressure of N_2 in the air he breathes, allowing blood to release N_2 into the lungs. The decrease in the blood level of N_2 causes N_2 to diffuse into the blood from the tissues. Fast tissues naturally unload N_2 quicker than slow tissues.

Theoretically, tissues should lose N_2 exponentially, and most decompression tables are calculated on this assumption. While large amounts of N_2 are lost initially, the process slows with time and it may take 24 hours or longer for all the N_2 taken up during a dive to be released. Diving again during the time of N_2 elimination will mean that the diver will start his second dive with a **N_2 retention** in some tissues. Adjustments are provided in the decompression schedule to allow for this and are incorporated as the **repetitive dive tables.**

If there is diminished circulation to a tissue during decompression, gas elimination will be reduced and thus bubble formation will be more likely.

In practice, even during routine conservative dives, bubbles of N_2 frequently form in the blood and tissues, interfering with N_2 elimination. It has been estimated that as much as 5% of N_2 taken up by the body after some dives is transformed into bubbles on decompression. These are often termed "**silent bubbles**" since they usually do not produce any symptoms. They do however have a profound and unpredictable influence on the decompression requirements for repetitive diving, because it takes much longer to eliminate gas bubbles in tissues than it does gas in solution.

SATURATION

When tissues are subjected to an increased partial pressure of inert gas during a dive, they take up dissolved gas in accordance with Henry's Law. However, there is a limit to the amount of gas which can be dissolved by a tissue exposed to any given partial pressure of gas (i.e. depth of dive). When this limit is reached the tissue is said to be **saturated**.

Our bodies are normally saturated with N_2 at atmospheric pressure and contain about one litre of dissolved N_2. If a diver were to descend to 20 metres (3 ATA) and remain there for a day or more, his body would take up the maximum amount of N_2 possible at that pressure and would then be saturated at that depth. His body would now have about 3 litres of N_2 dissolved in it.

Once the body is saturated with inert gas at a given depth, it will not take up more of that gas, no matter how long the diver spends at that depth. Consequently, once the diver is saturated the decompression requirement does not increase with time. This economy of time is exploited in **saturation diving**, when the diver is kept at depth for very long periods of time (days, weeks, months) but then needs only the same lengthy decompression.

BUBBLE FORMATION

The process of bubble formation can be demonstrated easily by opening a bottle of beer (or champagne, depending on taste and income). In a carbonated beverage CO_2 is dissolved in the liquid at a high pressure, which is then maintained by the lid. When the lid is opened, the pressure over the liquid becomes atmospheric and the partial pressure of CO_2 in solution exceeds the critical limit for bubble formation, causing bubbles to form. This could be avoided if the pressure was reduced slowly (decompressed).

During ascent, the pressure surrounding the diver (the environmental pressure) is reduced. Eventually, the pressure of N_2 dissolved in the tissues may become greater than the environmental pressure. The tissue is then said to be **supersaturated**.

The tissues are able to tolerate a certain degree of gas supersaturation. Nevertheless, Haldane explained that if the pressure of N_2 in the tissues exceeds the environmental pressure by a critical amount, then bubble formation is likely. The pressure differential needed to cause this varies between tissues but with most scuba diving it equals or exceeds 2 : 1 (i.e. the partial pressure of inert gas in the tissues should not be more than twice the environmental pressure). This explains why DCS under recreational diving conditions is unlikely after an isolated dive to less than 10 metres — the pressure at 10 metres is 2 ATA, while the pressure at the surface is 1ATA – a 2:1 ratio..

Gas bubbles in the tissue and blood are the cause of DCS. The exact mechanism of bubble formation is complex. It is likely that microscopic gas spaces (**bubble nuclei**) exist in all body fluids and that these form a nucleus for bubble formation during decompression.

Bubbles can form in any tissue in the body including blood. The pressure in each bubble will be the same as the environmental pressure (if it was not, the bubble would expand or contract until it was) and the bubble size is governed by Boyle's Law as the pressure changes.

At the onset of DCS, the pressure of N_2 in the tissues is supersaturated (greater than the environmental pressure) so there is an immediate diffusion (pressure) gradient of N_2 which then diffuses into any bubbles (or nuclei) present, causing them to expand.

A bubble of DCS contains mainly N_2 if the diver has been breathing air, but the other gases present in the tissues, such as carbon dioxide (CO_2), oxygen (O_2) and water vapour, also diffuse into it.

Once a bubble has formed its behaviour depends on several factors. Any increase in pressure such as diving or recompression will reduce its size while any decrease in pressure such as ascent in the water, over mountains or in aircraft, will expand it. The bubble will continue to grow in any tissue until the N_2 excess in that tissue has been eliminated. Once this has occurred (which may take hours or days) the bubble will begin to decrease in size but it may take hours, days or weeks to disappear. In the meantime the bubble can damage the tissues around it.

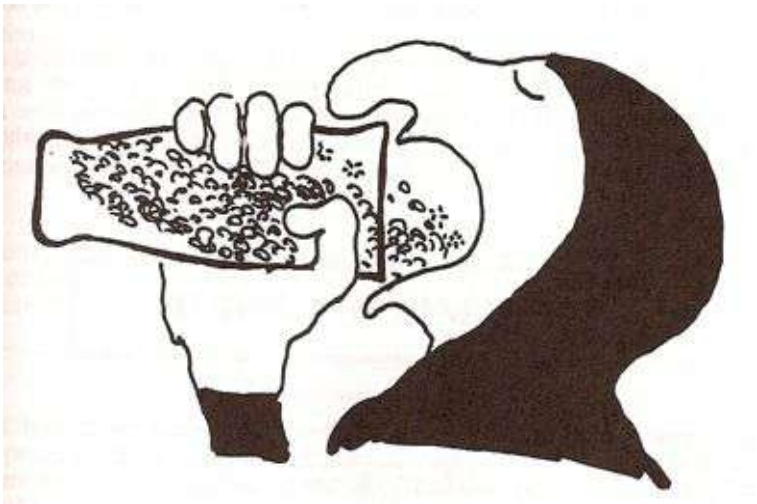

Fig. 13.1

There is good evidence that bubbles frequently form in tissues and blood of recreational divers after routine no-decompression dives, even when the tables have been faithfully followed. These bubbles do not usually cause symptoms but certainly cause doubt about the validity of both the decompression tables and dive computer algorithms.

Tissue damage by a bubble results from several factors. Bubbles in the blood obstruct blood vessels in vital organs such as the brain, while bubbles forming in the tissues may press on blood vessels and capillaries, obstructing their blood flow. Bubbles in the blood can also stimulate the clotting process causing the blood to clot in the blood vessels, obstructing blood flow to vital organs, and reducing the ability of the remainder of the blood to clot adequately. Many other biochemical and physiological changes with ill-defined sequelae occur in the tissues and blood vessels during both decompression and DCS. In the brain, spinal cord and other tissues, bubble pressure in or on nerves may interfere with nervous system functioning.

DIVE PROFILES

The type of dive has a significant bearing on where and when bubble formation takes place. **Short deep** dives (i.e. deeper than 30 metres) **tend to cause bubbles in the fast tissues** (blood, brain and spinal cord) while **long shallow dives tend to produce bubbles in the slow tissues** (like the joints). Long deep dives cause bubbles everywhere.

This distribution occurs because:

- in short dives, only the fast tissues take up enough N_2 to form bubbles on ascent and
- after shallow dives, fast tissues eliminate their relatively modest N_2 excess before a critical pressure differential develops.

It can thus be seen why it is important to ascend slowly from all deep dives. The slower the ascent rate, the longer the time for fast tissues to eliminate N_2 through the lungs, before a critical N_2 pressure-differential develops. Slow tissues are more affected by the total exposure (duration + pressure effects) and are more influenced by this and the duration of staging.

Diving folklore contains a myth that a diver using a single 2000 litres (72 cu. ft) tank cannot develop DCS. The air supply available was said to limit the diver to safe dive profiles. This is not true. For example, A single dive to 20 metres, may last 30-40 minutes, within the no-decompression time given by most tables, but it is not entirely safe. Remember, as mentioned previously, that even a single dive in excess of 10 metres can produce DCS.

The myth may become more apparent for deeper dives. For example, a single 2000 litre tank will give around 10 minutes duration for a 50 metre dive. According to most decompression tables, a 10 minute dive to 50 metres will require 10 minutes of decompression — but there may be no air remaining to complete these stops. Even if there was sufficient air, dives to this depth have a significant risk of DCS despite the tables being followed correctly.

FACTORS INFLUENCING DCS

DCS is unpredictable. In general, anything that increases blood flow to an organ will increase the rate of N_2 loading. Anything that interferes with blood flow from an organ will reduce the capacity to off-load N_2. These alterations may explain some of the possible predisposing factors that increase the likelihood of DCS.

❑ Depth/duration.

Any dive deeper than 10 metres can produce DCS although in general, the deeper the dive, the more gas absorbed, the greater the risk. The longer the dive at any one depth, the more gas absorbed (until saturation), the more the DCS risk.

❑ Individuals.

Some people are to be more susceptible to DCS than others. Even an individual may vary in susceptibility at different times, and DCS can develop after a dive profile which has been safely followed on many previous occasions. Others frequently get DCS despite conservative diving.

❑ Adaptation.

Repeated dives to similar depths over a period of time reduce the incidence of DCS. This may be due to the elimination of bubble nuclei. A diver returning to these dives after a 2 week break loses the benefits of this adaptation or acclimatisation.

❑ Age.

Older divers tend to be more predisposed to DCS (an old diver can be defined as anyone older than the senior author of this text). This age factor probably comes into play after the 3rd decade.

❑ Obesity.

This appears to be a predisposing factor probably due to increased N_2 solubility (4-5 : 1) in fat compared to water (obesity is defined as anyone heavier than the biggest author). This may be relevant for those with a BMI of > 25.

❑ Debilitation.

Factors causing the diver to be unwell such as **dehydration, hangover** or **exhaustion** tend to predispose to DCS. Fatigue (pre-dive) is observed in some cases.

❑ Injury.

DCS, particularly involving the musculo-skeletal system and joints, is more likely with recent bruises, strains or chronic injuries.

❑ DCS.

A previous episode of DCS, especially if it was unexpected from the dive profile, or if it damaged tissue (as in neurological DCS), makes the diver predisposed to similar subsequent episodes.

❑ Patent Foramen Ovale.

One reason for some people to have an increased susceptibility to DCS is that they have a small hole in their heart. All of us had a hole in our heart when we were a foetus. In about a third of the population some remnant of this hole remains, it is called a patent foramen ovale, or PFO. These people have an increased susceptibility to DCS, the likely reason is that bubbles that would normally be trapped in the lungs without causing symptoms pass through the hole, by-passing the lung filter, and on to other parts of the body, where they cause noticeable symptoms. However, the risk from a PFO is not great enough for it to be appropriate to test all divers for it, and repair of the hole is probably more dangerous than diving with it.

❑ Cold.

Diving in cold conditions makes DCS more likely, especially when the diver is inadequately insulated. More precisely, coldness during the dive inhibits inert gas uptake (because of restricted circulation) but allows more N_2 to dissolve in body fluids — whilst coldness during decompression inhibits inert gas release. Theoretically, it may be better to be cold during the dive and warm on decompression, unless bubble formation occurs. Warming will then reduce gas solubility and increase bubble growth and DCS.

The association between cold exposure and DCS is complex and contentious. During decompression and post-diving the cold environment may cause peripheral constriction of blood vessels and more bubble formation. Alternately, taking hot showers also tends to cause increased bubble formation and DCS.

❑ Alcohol and other drugs

It has been observed that divers who over-indulge in alcohol, or perhaps take other **drugs** or medications, may be especially susceptible to DCS. In the case of alcohol, especially taken the night before, the effects may be due to the associated dehydration or the vascular dilatation (remember the throbbing headache and "hangover"), increasing N_2 take-up.

❑ Exercise.

This also is complex and contentious. Some even claim that exercise 2-24 hours before diving, and even after diving, may reduce bubble formation. At least in rats.
Strenuous exercise during a dive is likely to increase the N_2 uptake by increasing blood flow to muscles, increasing gas uptake and favouring DCS development. Gentle exercise during decompression, by promoting circulation from the tissues probably aids in N_2 elimination. The effects may depend on whether bubbles or bubble nuclei have already formed. Strenuous exercise after the diver has returned to the surface makes the development of DCS, particularly in the musculo-skeletal system, more likely by promoting bubble formation. Strenuously activity, such as by shaking a beer can before opening it, aptly illustrates this phenomenon. During the first hour or two after a dive, particularly when there has been a large N_2 uptake, it is probably best to rest quietly as this is the period of maximal N_2 elimination.

Fig. 13.2

❑. Physical Fitness.

The less physically fit the diver, the more likelihood of DCS, probably because more energy is used and more blood flow is required for the same outcome – transporting more N_2.

❑ Gender.

There is some evidence that women have a higher incidence of DCS for certain dive profiles. There are subtle differences in physiology and body composition which could explain this. The decompression tables in current use only evolved after extensive testing on men alone (see Chapter 8).

❑ Dive profile.

Deep dives (greater than 18 metres), **long dives, decompression dives** and any dives exceeding the **limiting line** (in RN based tables) all have a higher incidence of DCS.

Square wave profiles (remaining at the maximum depth for all the dive) are probably more hazardous an equivalent N_2 load, than **multi-level diving**, if the levels are of diminishing depths ("forward dive profiles"). Otherwise they can be more dangerous.

❑ Reverse Dive Profiles

Divers are advised to dive from deep to shallow ("forward dive profiles"). The should dive their deep dive first in repetitive dives, and dive to progressively shallower depths when multi-level diving. If this order is not followed ("reverse dive profiles"), DCS is more likely.

❑ Rapid ascents.

These allow insufficient time for N_2 elimination from fast tissues, thus encouraging bubble formation.

❑ Multiple ascents.

Multiple ascents during a dive imply multiple decompressions and often involves rapid ascents. Bubbles in the blood (fast tissue bubbles) are likely to form during these ascents. The bubbles may not be adequately filtered by the lungs, passing along into the tissues, or may be reduced in size during the second or subsequent descent, allowing them to escape through the pulmonary filter into the tissues. DCS is then more likely.

❑ Repetitive dives.

Each repetitive dive begins with a N_2 load of some degree from the previous dive. Since bubble formation even after routine dives is common, a repetitive dive will often start with the diver carrying N_2 bubbles from the previous dive. N_2 elimination is less rapid from bubbles than it is from the same amount of gas in solution. These bubbles will be supplemented by N_2 taken up during subsequent dives, and make DCS more likely.

The algorithms used in dive computers are less accurate, and less validated, when used for repetitive diving.

Also with repetitive dives there may be the residual physiological effects of the previous dive, increasing the likelihood of decompression sickness. These physiological effects may include a lower body temperature, dehydration from immersion and recent exercise.

A very short surface interval may avoid some bubble formation in some dives, but if bubbles have developed, the longer the surface interval the safer the repetitive dive.

❑ Flying after diving.

The jet age often finds divers flying home after a dive holiday within hours of their last (sometimes literally) dive. International airliners are pressurised to an altitude of about 2000 metres (6500ft.) above sea level. This means a pressure reduction on the diver of about 25% with a corresponding increase in the degree of N_2 supersaturation as well as a corresponding increase in the size of any bubbles he may be carrying. The increase in size of critical bubbles may be sufficient to provoke symptoms or aggravate existing symptoms.

The DCIEM recommendation is "whenever possible it is inadvisable to fly above 600 metres in any aircraft within 48 hours of completing any dive. Travelling by vehicle over mountain ranges or hills can expose divers to the same dangers as flying and should be avoided in the same way for 24 hours. If flying after diving is considered essential, flying may be carried out after 24 hours but the increased risk of DCS must be borne in mind."

❑ Dive computers.

Using dive computers that are based on largely invalidated **theories** (as opposed to practical diving and decompression table experience) may result in a diver getting much more time underwater while diving — and in the recompression chamber during treatment. Both can be included in his log book if he survives.

They can be made safer by employing the advice given later (see Chapter 14).

❑ Multi-Factorial Effect.

Often there is more than one factor increasing the likelihood of DCS. Thus in one large Australian series over half the cases engaged in multiple dives, deep dives (greater than 30metres) and/or had ingested alcohol within 8 hours. Another 20% were precipitated by aviation exposure. Thus many of these divers would have had at least 2 factors increasing their likelihood of DCS.

Note: The senior (elder) author believes that the only explanation for most cases of DCS lies in the random application of Chaos Theory, which he also does not understand, or string theory which no-one understands.

Chapter 14

DECOMPRESSION TABLES & METERS

Since the work of J.S.**Haldane** (a British physiologist) early last century, decompression tables have been based on mathematical models of gas uptake and elimination in the body. He believed (wrongly) that the exponential rates of uptake and elimination of gases would be equal, and that body tissues could tolerate a supersaturation pressure gradient of air equal to 2:1, without causing bubble formation. He experimented on goats – the closest animal model that he could get to the diver – to determine this gradient. Thus a goat could tolerate 2 ATA of air pressure in its tissues, even when walking around at 1 ATA! Or, it could ascend from 6 to 3 ATA, 4 to 2 ATA and from 2 to 1 ATA, without bubble formation.

Haldane devised mathematical equations (later referred to as models or algorithms) that would represent a diver, and applied this to dive exposures. He postulated 5 hypothetical tissues (with half-times of 5, 10, 20, 40, 75 min) and decompressed the diver so that the supersaturation gradient in each of these 5 tissues would never exceed 2:1. Once any tissue reached that ratio, the diver would stop his ascent and "stage" there until the tissue "off-gassed" to allow him to ascend to the next stop. Thus bubble formation would be avoided by this decompression.

Although used by the Royal Navy for 50+ years, it was evident that some of Haldane's dive stops, or stagings, were too conservative and others too radical. Also, the assumptions on which the procedure was based were wrong.

The acclaimed 1957 **U.S. Navy Tables** are based on Haldane's theories. These include equal exponential uptake and elimination of inert gases and supersaturation gradients – as described in Chapters 1 and 13 – but have been modified by experimental trials and practical diving experience. In an attempt to overcome the flaws in Haldane's tables and to make decompression safer, they increased the number of hypothetical tissues to 6 (5, 10, 20, 40, 80 and 120 minutes), and calculated different maximum safe supersaturation ratios (now called M values) for the different tissues at different depths.

Following experiments and reviewing of the established decompression tables, between 1960 and 2000 a whole series of innovations and modifications were introduced. Dr Bruce Bassett (a USAF physiologist) concluded that the US Navy Tables resulted in an excessive incidence of about 6% DCS, when pushed to the no-decompression limits. Merrill Spencer in Seattle verified this observation and supported it with extensive Doppler monitoring, showing that bubbles developed in many routine dives – implying inadequate decompression. Many others observed similar inadequacies and in an attempt to cope with this new information, tables

based on modified Haldane principles were developed by Bassett, Huggins, NAUI, PADI, and many others.

The main alterations to improve safety were in;

- reducing the acceptable no-decompression times by 10-20%
- reducing ascent rates from 18 m/min to 9-10 m/min (at least in the top 30m)
- Adding a "safety stop" of 3-5 min at 3-5 m.

The Swiss decompression expert, Professor **Buhlmann** produced the Swiss model, which includes 16 theoretical tissue compartments with widely ranging half-times. The testing of these tables at altitude became more extensive than for most other tables, and he extended the concept into repetitive dives. Much later the US Navy E/L model assumed an exponential uptake of gas and a linear loss to eliminate gas

In the UK, Hempleman introduced a slab diffusion concept, which later evolved into the RNPL/BSAC tables. A cylinder of tissue was used by Hills in Australia. Other groups assumed the presence of bubbles or bubble forming nuclei, bypassing the traditional gas-in-solution supersaturation concepts, and decompressed in such a way as to keep the bubbles at a tolerable level. These are sometimes referred to as "**bubble models**"

Yount in Hawaii developed tables designed to keep gas nuclei from forming larger bubbles, and the allowable pressure gradients across these nuclei were less than Haldane's supersaturation ratios. Thus the no-decompression times were less, the initial decompression stops were deeper and the ascent rates were slower. Weinke developed these concepts further in the Reduced Gradient Bubble Model, and has collected some diving data based on these tables.

The Canadians (**DCIEM**) considered the transfer of gas between adjoining tissues as a major principle, and produced a very well researched table, popular amongst conservative recreational divers. Although based on decompression theory, it has been modified by extensive human testing in cold water and hard working conditions, with Doppler (ultrasound) monitoring. The single no-decompression times, and most repetitive dives, are more conservative than the US Navy tables and are often recommended for recreational diving.

Another approach involves the collecting of a vast number of well-documented dive profiles, some safe and some unsafe, and letting the computer devise the safe tables. Extending this concept, the US Navy, and now DAN, use such statistical analyses to determine the probable DCS disk with each dive profile and so "probabilistic" tables that accept a defined DCS incidence of 1%, 2.3% or 5% can be derived – in theory.

When the formal tables were pushed to the limits (decompression diving, deep diving and dives approaching non-decompression limits), there was an unacceptable 1–5% incidence of DCS. Even in divers who appeared unaffected, Doppler (ultrasound) studies often showed bubbles in the major veins during decompression. Nevertheless, in both recreational and navy diving there is only a 1/5,000 to 1/10,000 incidence of clinical DCS. This paradox is explained by the fact that most dives do not approach decompression or depth limits and so do not really test the tables. Also, divers include conservative modifications or "fudge factors", making their diving safer. Even the US Navy has modified its "gold-standard" table by reducing ascent rates and incorporating safety stops, making it marginally safer than it was.

It is not possible to make a perfect mathematical model of the decompressing human body. Even under normal conditions, the blood flow to some capillary beds shuts off from time to time, while flow to other areas increases. During decompression, nitrogen elimination will virtually cease from an area of shutdown capillaries. As decompression proceeds, this area will have a much higher N_2 concentration than predicted. No mathematical model can predict biological phenomena such as this.

When one considers the ill-defined complexities of gas transport through the range of different tissues, all with different solubilities, different supersaturation tolerances, different nuclei quantity and location - all at different depths, durations and ascent rates - the accurate representation of bubble development with mathematical equations, models or algorithms, is impossible at this time. Even if one could so define the evolution of bubbles, there are many other problematic biochemical and haematological sequelae, as well as individual susceptabilities, in the development of DCS manifestations, that reliable modeling is not yet likely

Any decompression table or computer algorithm that offers longer durations underwater, deeper or more dives, will result in a greater incidence of decompression illness, unless compensated for with slower ascent rates and longer decompression.

The imminent release of brilliant "new" tables is a permanent rumour. The difficulty in producing mathematical models which truly reflect human decompression physiology upon which to base the tables or computers, as well as the difficulty and expense of testing them, makes the development and validation of tables based on truly new models unlikely. However, tables and computers based on recycled mathematical models do frequently arise and often there is an amalgamation of decompression theories.

Towards the end of last century recreational diving involved more repetitive dives and multi-level diving. It reflected the interests of recreational divers, taking diving holidays, on live-aboard boats, exploring reefs and drop-offs. This was considerably different to the pre-planned single-depth/designated-duration dives of navies and commercial divers. With much recreational diving, relatively little time is spent at the maximum depth. Attempts to accommodate for these different diving profiles were attempted, with ingenious innovations by Graver, Huggins, the PADI RDP and Wheel, etc., with variable success. Others, such as the Buhlmann tables were more easily modified to incorporate the greater number of variables with multi-level and repetitive diving.

For any given nitrogen loading, multi-level diving from deep to shallow should produce less DCS than the single-depth dives documented in the formal dive tables. However, decompression from the infinitely variable multi-level diving is almost totally based on theory. There have been no adequate and comprehensive trials performed to show the tables reliability.

For both theoretical and practical reasons, most dive authorities advised divers to dive to progressively shallower depths, when repetitive or multi-level diving. This is referred to as a “Forward Dive Profile”. About a decade ago there was a misplaced academic campaign to accept “Reverse Dive Profiles” as having equivalent decompression obligations, as an acceptable diving practice, without incurring decompression penalties. Both experimental evidence and diving experience subsequently showed that the traditional “deep dive first” concept is correct, in both multi-level and repetitive diving. If reverse dive profiles are to be used, they will require new decompression schedules, often with longer decompression requiremts.

DECOMPRESSION METERS and COMPUTERS

With the preponderance of multi-level diving, repetitive diving and multi-day diving, the expedient introduction of decompression meters (DCM) and decompression computers (DC) resulted in a gradual reduction in the use of established tables and an acceptance of the new technology.

These meters are based on three different principles:

❑ Mechanical models of gas transfer. Decompression meters. (DCM)

The early meters in the 1960s relied on the movement of gas through small orifices to simulate uptake and elimination of nitrogen by compartments of the body. They are obviously a gross oversimplification of the gas transfer in divers but one, the **SOS meter**, was moderately safe for single shallow non-decompression dives to less than 24 metres.

❑ Electronic decompression models of existing tables.

These devices of the 1970s recorded the depth and time of a dive and related this to one of the existing decompression tables which was stored in its memory. One was even incorporated into a wristwatch. They saved the diver the difficulty of recording his depth and time, reading and remembering the tables.

These meters were not been very popular, because they did not offer any decrease in decompression obligation compared to the conventional tables and represented quite an expense merely to save the diver the trouble of applying a decompression table.

❑ Decompression computers (DC)

These are electronic models based on decompression theory. Microprocessors developed for computer use were incorporated into the diving world with the Edge (USA) and Decobrain (Switzerland), in the 1970s. Instead of following the tables, with their inevitable rounding up of depths and durations and the associated safety factors involved in overestimating pressure exposure, they were far more precise in their measurements and calculations. This meant that they could meticulously track the dive exposure and use algorithms to calculate in real time the nitrogen uptake and decompression required, according to any predetermined decompression theory. The safety of the DC then depends on the validity of the theory.

Fig. 14.1

DCs are of great value to divers, if they are used correctly. They have more accurate gauges (depth, duration) than previously available and this information is stored in a more reliable computer than the divers' narcotised memory. The incorporation of an ascent meter and alarm is invaluable in warning of rapid ascents.

These meters are programmed with one of the mathematical models (theories) on which the conventional tables have been based (usually the US Navy, Buhlmann or "bubble" tables) and this allows for multi-level and repetitive diving calculations. Because of this, and the avoidance of the rounding-up safety features inherent in the tables, the DCs usually permit longer dive durations, more flexibility in dive depths and repetitive dives, shorter surface intervals and less intellectual demand than "table diving". There are currently over a dozen different algorithms determining decompression in DCs. In general medicine, there is an axiom that the more treatments that there are available, the less likely it is that any work well. Same with diving medicine and DCs .

Flexibility can be incorporated and with the more advanced DCs, there are optional safety factors and restrictions. One can determine whether the DC behaves more conservatively than the theoretical algorithm normally calculates, allowing for increased age, weight, importance etc – but at the inevitable cost of a reduced dive duration. Many also indicate altitude (flying) restrictions post dive.

Most cases of DCS treated nowadays occur in divers who have correctly followed the instructions computed by their DCs. Probably over 90% of divers now use DCs instead of decompression tables, and considering the poor comprehension of the tables, this is probably a good thing. It certainly means that when an accident does occur, a comprehensive dive profile can be downloaded onto another computer for analysis, in retrospect – and sometimes to the embarrassment of the diver who pleads innocence of his misdemeanours.

❑ Disadvantages of decompression computers.

Sophisticated electronics, batteries and sea water are sometimes not compatible. Machines and technology sometimes fail. Back up plans are advisable.

The price to be paid for more dives and longer durations, without longer decompression, is increased DCS.

Most DCs employ algorithms that have not been adequately validated. When each of the conventional decompression tables was first formulated from a theoretical model, it had to be extensively modified based on human testing. Essentially, the DCs are offering the tables as they were, before they were modified with safety factors. With the removal of the tables rounding-up safety factors, the price divers pay for their extended underwater durations is that they can dive much closer to the DCS margin.

Most of the new DCs are (like the old SOS meters) conservative for short single, shallow dives. The dangers increase with longer, repetitive or deeper diving. They are also probably less safe with some multi-level profiles, if reverse profiles are used.

Nevertheless, DCs are popular and deservedly so. They are perceived as an alternative to personal responsibility in coping with a difficult problem (decompression table calculations). Unfortunately, to question the value of any specific computer can be interpreted as blasphemy, casting aspersions on the owners, bringing a deluge of unsolicited testimonials and threats of legal action by the manufacturers.

❑ **Safety suggestions** (the **DC Ten Commandments).**

If you rely on DCs, the following recommendations are made:

1. Do not dive in the 24 hours before using the DC.
2. The DC should be used on **all** dives, if it is used on **any**.
3. Ensure a back-up documentation of the dive profile and details.
4. In multi-level dives, the depths should be progressively shallower.
5. Repetitive dives (on the same day) should be progressively shallower.
6. Repetitive dives should have surface intervals of at least 2 – 4 hours (longer with greater depths and longer dives).
7. With multiple days diving, every fourth day should be non-diving.
8. Add an extra decompression safety stop for 5 minutes at 5 metres, on each dive in excess of 15 metres depth, if practicable.
9. Do not do dives that require decompression or go into the decompression mode. Stay as far away from those dives as possible.
10. Do not presume that the DC is accurate for diving at altitude or for altitude exposure (flying).

DCs are preferably used at depths of less than 30 metres and definitely depths of less than 40 metres. Pseudo-science accompanied the technology in the promotion of these DCs, and many claims of excellence were more applicable to the colour brochures and marketing than any research or development activities.

For those divers who are more important, or are more susceptible to DCS (age, gender, fitness, weight, medications, injuries, DCS history, etc.) or undertaking more hazardous dives (depth, duration, decompression obligation, temperature, currents, repetitive, multiple ascents, etc), the more conservative modes on the DC should be chosen.

PREVENTION of DCS

See the last page of Chapter 16

Chapter 15

DECOMPRESSION SICKNESS CLINICAL FEATURES

ACUTE DECOMPRESSION ILLNESS (ADI)

Although the term ADI can cover all types of DCS and barotraumas, it is commonly limited DCS and arterial gas embolism (AGE) from pulmonary barotrauma. Neurological DCS and AGE may be difficult to distinguish as they share similar symptoms - and the initial first aid treatment is similar. Rarely the diseases may even co-exist. Combining the two diseases under a common presentation is of value to those who find them difficult to distinguish from history, symtomatology and clinical signs.

Eventually, both in definitive treatment and in planning future advice to the diver, it is necessary to be more specific.

CLASSIFICATION OF SYMPTOMS

In the past it was traditional to describe decompression sickness (DCS) as Type 1 (minor – - musculo-skeletal or joint) or Type 2 (serious - cardio-pulmonary and neurological) DCS. Now the convention in is to classify the clinical features according to the organ or system involved (e.g. neurological, musculo-skeletal etc.). It is prudent to include with this diagnosis an indication of the evolution of the symptoms (as described later), as both influence the treatment required.

ONSET OF SYMPTOMS

The clinical features of DCS are seen during or after ascent. In the majority of cases, symptoms will be evident within six hours, and 50% within the first hour of the dive. Less commonly, a delay in onset of 24 hours or greater has been described. Divers often deny the reality of their symptoms, or will rationalise them. In one large series, only half the divers requested assistance within 12 hours of the first symptom developing.

The time of onset of symptoms depends to some degree on the type of dive. Deep dives (greater than 30 metres), especially those that require decompression or are close to the no-decompression limits or in which decompression has been omitted, are likely to present early. In extreme cases, symptoms may present during ascent or at the decompression stops. In general, the earlier the symptoms, the more potentially serious the DCS.

Symptoms may be initiated or aggravated by exposure to altitude (driving over mountains, air travel), exercise or breathing certain gases. Divers should be advised of the potentially serious complications of flying after diving.

JOINT PAIN (MUSCULO-SKELETAL DCS, BENDS)

Pain in or near one of the muscles or tendons around the joints is the **most common presenting feature of DCS from shallow diving.** The shoulder is most often affected while the elbows, wrists, hand, hips, knees, ankles are less frequent. It is not unusual for two joints to be affected, commonly adjacent ones e.g. the shoulder and elbow on the same side. It is less common for multiple joints to be affected in a symmetrical pattern.

Symptoms may begin with discomfort or an abnormal feeling in or near the joint. Over the next hour or two, pain and other symptoms may develop. The pain is generally of a constant aching quality (like a toothache), but occasionally may be throbbing. The diver may hold the joint in a bent position to reduce the pain. The stooping posture, that was adopted by Caisson (tunnel) workers affected by DCS near the hip, led to the term "bends".

The joint is usually not tender to touch but movement may aggravate the pain. Pressure, as from a blood pressure cuff (sphygmomanometer), may relieve the pain.

If not treated, pain usually continues for several uncomfortable days before slowly subsiding. In mild cases, minor and fleeting discomfort lasting only a few hours ("**niggles**") may be the only manifestation.

Symptoms are often found around a joint which has been subjected to unusual exertion or strain during or after the dive or which has been involved in a recent or chronic injury.

There may occasionally be difficulty distinguishing between DCS and other causes of a painful joint such as strain, injury or arthritis. In the latter conditions, the joint is usually tender to touch and may be red and swollen. Generalised arthritis is often bilateral and

symmetrical and involves smaller joints, and local pressure application produces no relief. These signs are uncommon in DCS.

In general, any pain in or near a joint after compressed air dives in excess of 10 metres (or shallower with repetitive or prolonged dives) must be assumed to be DCS until proven otherwise.

NEUROLOGICAL DCS

DCS can affect the **brain, spinal cord and/or peripheral nerves**. The clinical features are due to disturbance of activity in the nervous system, interfering with one or more of its five principal functions:

- **sensation**
- **movement** (including balance and co-ordination)
- **consciousness** and intellectual functions
- **autonomic** functions
- **reflexes** (e.g. knee jerk, cough reflex).

Of these, the first four are easier for the layman to assess.

❑ The Senses.

These include sight, hearing, smell, taste, pain and touch. Numbness and tingling (paraesthesiae) are frequent symptoms. Other abnormal signs include loss of sensation.

❑ Movement.

This includes the ability to move any muscle, the strength of the movement and the ability to co-ordinate it.

❑ Higher function of the brain.

The important intellectual functions are consciousness, orientation (awareness of time, person and place), thinking, speech and memory. Epileptic fits (convulsions) and confusion are possible. See the AMTS box at the end of this chapter for a rapid, simple method of assessing higher function brain involvement.

❑ Autonomic functions.

Interference with the control of breathing and heart function may cause shock and collapse. Bladder and bowel malfunction usually causes progressive abdominal discomfort and tenderness until the bladder or bowels are opened.

In **CEREBRAL (Brain) DCS**, the bubbles of DCS may be located in or near the blood vessels supplying the brain, causing obstruction of blood flow and direct pressure on the neurological tissues. The brain swells like any other tissue when injured, but because it is

confined within the solid bone of the skull, the pressure in the skull rises, further impairing blood supply to other parts of the brain. Swelling of the brain (cerebral oedema), as well as expansion of the nitrogen bubbles themselves, often leads to a steady worsening of this condition.

The onset of cerebral DCS is often heralded by **headache —** probably due to brain swelling. **Numbness** or **tingling (paraesthesiae)**, **weakness** or **paralysis** affecting a limb or one side of the body, difficulty with **speech**, **visual** disturbances, **confusion**, loss of **consciousness** or **convulsions** are all possible presenting symptoms of this serious disorder.

The part of the brain responsible for co-ordination (the cerebellum) may also be affected causing incoordination known as **"staggers".** The position sense and the balance organs can also be affected.

SPINAL DCS has a common association with DCS bubbles in the blood and lungs, commonly known as "**chokes".** It also may be preceded by "**girdle pains**" — or pain around the chest or abdomen. Disturbances in movement such as **weakness or paralysis** or disturbances in sensation such as **numbness or tingling** are also common. Interference with nerve supply to the bladder and intestines, may lead to **difficulty in passing urine, or opening the bowel**. **Paraplegia** or **quadriplegia** may develop.

INNER EAR DCS

The cochlea (hearing) or vestibular (balance) organs may be involved. This type of DCS is more commonly associated with **prolonged**,,**deep, repetitive, diving, breathing a helium-oxygen mixture**. **Hearing loss**, **ringing noises** in the ears (tinnitus), and/or **vertigo, nausea** and **vomiting** are the usual presenting features. This condition must be distinguished from the other major cause of these symptoms in divers – inner ear barotrauma (see Chapter 9).

LUNGS or PULMONARY DCS

Nitrogen bubbles are frequently found in the veins of divers ascending after deep dives, without necessarily the development of overt clinical DCS. When large numbers of these bubbles form, they may become trapped in the small vessels of the lungs, obstructing the blood flow. If excessive bubbles occur, this leads to a disturbance of lung function and a feeling of **breathlessness**, known as the "**chokes**".

Clinical features also include **a tight feeling in the chest, chest pain, difficult** and **rapid breathing**. Coughing is often observed, and may be precipitated by smoking, hyperventilating or exercise. Frequently the symptoms come on very soon after ascent, from relatively deep dives (over 30 metres) or after prolonged dives. Death may supervene.

HEART or CARDIAC DCS

The nitrogen bubbles that commonly form in the veins of divers after deep dives are usually filtered by the lung vessels.

There is a condition affecting the heart (**patent foramen ovale – PFO**) which is sometimes claimed to be present in about 30% of the population, and in which there is a potential communication between the right and left sides of the heart – between the right and left atria. It is a flap valve, normally kept closed by the naturally higher pressure on the left side (left atrium) of the heart. When large amounts of nitrogen bubbles obstruct the lungs, the back pressure in the right atrium can exceed the pressure in the left atrium. This flap may then open allowing gas bubbles to pass from the right to the left side of the heart, and then be pumped and distributed to any part of the body (similar to arterial gas embolism resulting from a burst lung – see Chapter 11).

Bubbles passing through a patent foramen ovale (PFO), or any other cardiac defects, tend to rise because of buoyancy. They can be easily carried into the blood vessels supplying the brain because it has a large blood supply and is located above the heart. This is one cause of cerebral DCS after apparently so called "safe" dives. It may be called cerebral arterial gas embolism or **CAGE,** but is due to DCS, and not pulmonary barotrauma. The term **acute decompression illness** covers both causes.

Bubbles can occasionally pass down the coronary arteries, which supply the heart, restricting the blood supply to the heart itself. In severe instances this can lead to a fatal destruction of heart muscle (myocardial infarction) just as for a "heart attack" in a non-diving person. In other cases, life threatening disturbances in cardiac pumping and rhythm may result.

Cardiac symptoms include chest pain, palpitations and shortness of breath.

GASTROINTESTINAL DCS

Obstruction of blood flow to the intestines by nitrogen bubbles can occasionally affect the gut. Clinical features are not common, but can include **vomiting** or **diarrhoea**, **cramping abdominal pain** and **haemorrhage** into the gut. Severe cases can show clinical shock, and can bleed to death.

SKIN MANIFESTATIONS of DCS

These are not common in scuba divers who wear wet suits.

Itching of the arms and legs, sometimes with a rash, is not uncommon after deep recompression chamber dives and with dry suits. This condition is probably due to gas passing from the surrounding high pressure atmosphere into the skin. The condition is not serious and requires no treatment.

In more severe DCS, nitrogen bubbles in the blood can obstruct blood supply to the skin, causing patchy white, blue and pink areas – "**marbling**". Obstruction of the lymphatic system (drainage of tissues) may produce **localised swelling of skin**.

GENERAL SYMPTOMS of DCS

Apathy, **tiredness**, **malaise,** and a **generalised weakened state** are common observations in many cases of DCS. In very severe cases there may be **generalised internal haemorrhages, shock**, and/or **death**.

DELAYED SYMPTOMS of DCS

Prolonged symptoms may be due to damaged nerve, spinal cord or brain (see above), body tissues or bone (see Chapter 17). In these cases there can be a persistence or recurrence of symptoms. Various psychological problems can also supervene on DCS. Even the stressful treatments and peer recriminations can exact an emotional toll on divers. Anxiety states and post-traumatic stress disorders are not uncommon sequelae

EVOLUTION OF SYMPTOMS

In general, the slower symptoms are to develop after ascent, the less serious the DCS, and *vice versa*. Also, the longer the delay between ascent and symptoms (say, 6-24 hours), the fewer and slower subsequent symptoms are to develop. The corollary of this is:

With symptoms developing soon after the ascent, the DCS is likely to be more serious.

In assessing DCS, the time of onset of symptoms should be related to the time of ascent. The clinical manifestations and their evolution should be described, together with any aggravating factors.

The manifestations may be progressive (getting worse), static or improving. They may also relapse.

Aggravating factors include not only those that predispose to DCS, but also those that precipitate it (see Chapter 13).

A major observation, supporting the diagnosis of DCS, is the favourable response of DCS to raised environmental pressure (re-immersion in the water, treatment in a recompression chamber) and, with a slower response, administration of 100% O_2.

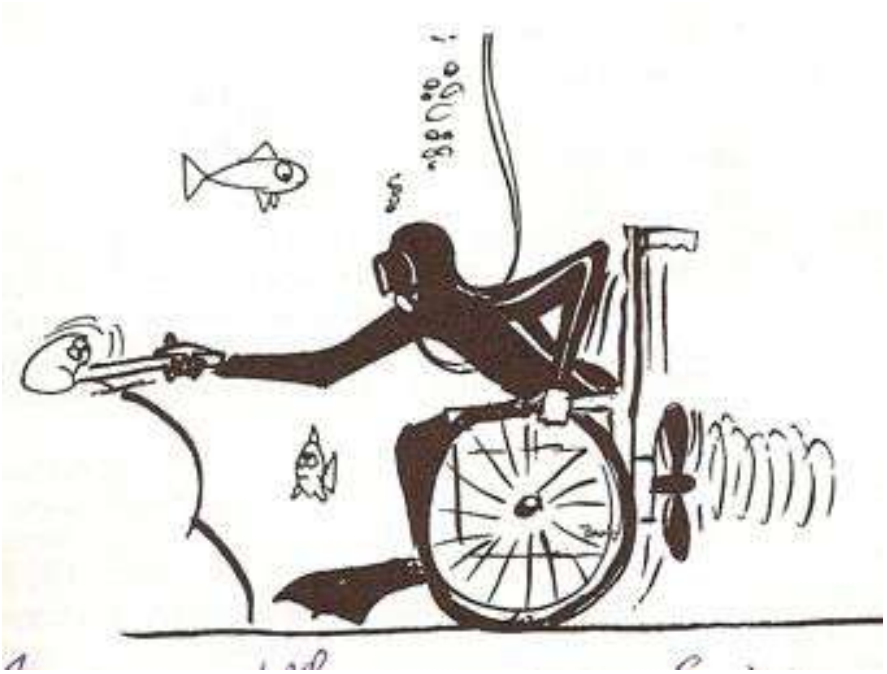

Fig. 15.1

The **ABBREVIATED MENTAL TEST SCORE (AMTS)** can be used to assess mild delirium, confusion and other cognitive impairment.

The following questions are put to the patient. Each question correctly answered scores one point. A score of 6 or less indicates impairment of brain function. A score above 6 does NOT mean that there is no cognitive impairment

Question
What is your age? (1 point)
What is the time to the nearest hour? (1 point)
Give the patient an address, and ask him or her to repeat it at the end of the test. (1 point) e.g. 36 Smithsonian Street
What is the year? (1 point)
What is the name of the hospital or number of the residence where the patient is situated? point)
Can the patient recognise two persons (the doctor, nurse, home help, etc.)? (1 point)
What is your date of birth? (day and month sufficient) (1 point)
In what year did World War 1 begin? (1 point) (other well known dates can be used, with a preference for dates some time in the past.)
Name the present monarch/prime minister/president. (1 point) (Alternatively, the question "When did you come to [this country]? " has been suggested)
Count backwards from 20 down to 1. (1 point)

As well as a "spot check", this test can be used to indicate improvement or deterioration, if repeated. It is not as comprehensive as others, which may be employed by paramedics and doctors e.g. the Mini-Mental State Examination, or the TYM.

Chapter 16

DECOMPRESSION SICKNESS TREATMENT

FIRST AID

❑ Expert advice.

Expert advice from a diving physician, a diving medical organisation (such as DAN) or a hyperbaric facility, should be sought as soon as possible. **Appendices B and D** contain a list of sources for such expert advice. In the interim, first aid treatment must be applied

The principles of first aid management of decompression sickness (DCS) are:

- **basic life support (BLS)** (see Chapter 42).
- **100% oxygen (O_2)** (see Chapter 40).
- **positioning and rest**
- **fluid replacement**

❑ Oxygen therapy

If the diver breathes 100% oxygen (O_2), nitrogen (N_2) is removed from the lungs. The breathing apparatus must supply to the lungs as close to 100% O_2 as possible. This means an **anaesthetic type mask** or an O_2 diving regulator as used in technical and commercial diving. It is not the simple plastic hospital oxygen mask that does not produce an air tight seal. 100% O_2 in the lungs results in a high diffusion (pressure gradient) of N_2 from the blood to the lungs, causing the increased elimination of N_2 from the blood and tissues – and also from any bubbles there.

Unfortunately 100% O_2 can be toxic to the lungs if given for 18–24 hours on the surface. This may complicate the hyperbaric O_2 which is later given to the diver in a recompression chamber (RCC). Ideally O_2 therapy should be given under the supervision of a diving physician however, if expert advice is unavailable, then **all suspected case of DCS should be given 100% O_2** from the outset, before and during movement of the patient and transport to the recompression chamber.

Unconscious divers, if not breathing of their own accord, will require assisted ventilation (intermittent positive pressure respiration or I.P.P.R.). Conscious patients can be treated with continuous flow or demand type masks. In such cases, treating personnel must always check the breathing mask on themselves first in order to ensure the oxygen system is working, and that resistance to breathing is not excessive.

Fig. 16.1
Photograph of diver with DCS breathing near 100% O_2 from a resuscitation bag. Note reservoir bag (lowermost) and high constant flow (>14 litres per minute) through an oxygen delivery tube.

❑ Position and rest.

An unconscious diver should be placed on his side in the coma position, to protect the airway at all times (see Chapter 42). If there is any likelihood of air embolism, the diver is best placed horizontal - preferably on the left side, but this is not essential. Some clinicians recommend that cases of cerebral DCS should be managed with the patient on the side, without a pillow, to prevent buoyant gas bubbles reaching the brain through the circulation.

The Trendelenburg (30 degree head down, feet up) position is no longer recommended. Having the head lowered raises the pressure in the brain – and this can aggravate the brain injury.

Due to the buoyancy of bubbles, sitting or standing may be dangerous in patients with air embolism or cerebral decompression sickness where the bubbles are still in the blood stream. As a general rule, 100% O_2 should be given for at least an hour before allowing the patient to sit or stand. After this, the diver can be allowed to adopt any comfortable position but should be kept relatively still and on 100% O_2. A diver with "chokes" will be more comfortable sitting up.

❑ Fluid replacement.

Severe DCS results in loss of blood and fluids into the tissues. It may be valuable to replace this fluid orally or intravenously. Oral fluids may be given if the diver has no abdominal pain, nausea or vomiting. Water or electrolyte fluids (eg."Gastrolyte") may be given – with the type and volume recorded. Acidic (orange juice) and glucose dinks should be avoided. If there is no bladder involvement (i.e. the patient can urinate), a litre of clear fluid every 2-4 hours should suffice. This fluid intake should be modified by the patient's thirst.

Some authorities have recommended large volumes of oral fluids while the diver is being transported to a RCC in an attempt to replace this deficit. One problem with this (as any party goer will attest) is that the fluid load will promote a vigorous urine flow, so the diver arrives

at the recompression facility with a stomach and bladder full of fluid. A patient with spinal DCS may be unable to empty the bladder and will therefore be in considerable pain.

The patient will usually be treated on a hyperbaric O_2 (HBO) treatment table. There is a very real risk of nausea, vomiting and convulsions as complications of this treatment. A full stomach can then possibly result in regurgitation of the stomach contents and aspiration into the lungs – further complicating treatment.

If the brain or spinal cord is involved and the patient has difficulty in voiding urine, an **in-dwelling urinary catheter** should be inserted whenever possible by a trained physician or nurse. If this is not feasible, care must be taken not to overload the patient with fluids.

Anyone who is trained to institute and monitor an intravenous infusion can be expected to be able to assess the state of hydration and determine the desirability and quantity of intravenous fluids, remembering that glucose fluids can aggravate cerebral oedema.

❑ Drugs.

Aspirin as a first aid measure has not been demonstrated to be of value in DCS. It may interfere with blood clotting and cause haemorrhage (bleeding) – especially in the stomach. Haemorrhage is already a major pathology in spinal cord and inner ear DCS.

The authors have seen one patient with severe DCS bleed to death from an internal haemorrhage just before he was to be given an "experimental last-ditch" anti-clotting agent. We are therefore reluctant to advocate the routine use of aspirin either for pain relief or to inhibit clotting in any DCS case.

Joint pains of DCS can be significantly eased without the risk of serious side effects by the administration of **paracetamol (acetaminophen) –** 1000 mg (or two tablets) 4 hourly. NSAID drugs may be requested by a diving physician, but are not routinely needed.

Other drugs such as steroids, diuretics and special intravenous fluids such as "Rheomacrodex" have been advocated but have not been proven to be beneficial. Anti-epileptics and other drugs such as diazepam ("Valium") may be needed to control fitting (convulsions), and for confusional states.

TRANSPORT OF PATIENT WITH DCS

With mild DCS, or if there has been a delay of 12-24 hours or more, and there is no progressive deterioration, local treatment with rest, monitoring and breathing 100% O_2 may be all that is necessary. This decision is best made by a diving physician.

With more severely affected divers, or those that are deteriorating or need medical attention, transport should be expedited. The diver should be transported with **minimum agitation** and as close as possible to sea level or at **1 ATA**. Mountainous roads should be avoided whenever an evacuation route by land is planned. 100% O_2. should be breathed before and during transport (see Chapter 40).

Transportation in aircraft presents problems. Apart from movement which aggravates DCS, environmental pressure decreases with altitude, causing DCS bubbles to expand (Boyle's Law) and more gas to pass from the tissues into any bubbles.

If the patient is evacuated by air, unpressurised aircraft should endeavor to fly at the lowest safe altitude. Even an altitude of 300 metres (1000 ft.) can make the symptoms of DCS worse. However, maintaining such an altitude can be alarming when flying over 297 metre terrain.

It should be remembered that most commercial "**pressurised**" aircraft normally maintain a cabin pressure of around 2000 metres (6000 ft.), which could seriously aggravate DCS.

Whenever possible the cabin altitude should be maintained at 1ATA. This is attainable by many modern commercial jet aircraft, executive aircraft such as the King Air and Lear Jet, and some military transport aircraft (Hercules C–130). This requirement is not popular with the commercial airlines since it necessitates the aircraft flying at lower than its most efficient altitude, resulting in excessive fuel consumption. This requirement may also limit the range of certain aircraft.

Breathing 100% Oxygen before and during the flight may be of value, especially from closed or semi-closed circuit equipment. There are risks to the aircraft and its inhabitants if open circuit O_2 is used as many airlines recirculate the cabin atmosphere and fire/explosion is possible.

Fig. 16.2

A portable patient treatment chamber connected to a second compartment for the exchange and transter of attendents and medics. The larger chamber also allows for transfer of patient to a larger recompression facility.

DEFINITIVE TREATMENT OF DCS

This is best controlled by diving physicians, hyperbaric facilities or specialised diving medical organisations, such as DAN.

❑ Therapeutic recompression.

This is the most effective treatment for DCS. Delay increases the likelihood of a poor final result. The diver is placed in a **recompression chamber (RCC)** and the pressure is increased according to a specified recompression treatment table.

The increase in pressure reduces the bubble size (Boyle's Law) and usually relieves the clinical features. It also increases the surface area to volume ratio of the bubble, which may collapse. The increased pressure in the bubble also enhances the diffusion gradient, encouraging nitrogen to leave the bubble.

The therapeutic recompression schedule determines the depth and duration of the treatment profile. The table selected may depend on factors such as the time elapsed since the onset of symptoms, the depth of the original dive and the type and severity of symptoms, as well as the capability of the treatment facility and the various breathing gases available. The treatment may be amended depending on the patient's response.

In recent years the **Oxygen Treatment Tables** are preferred because of their increased effectiveness. The injured diver is usually compressed to an equivalent depth of 18 metres (60 ft.) and is decompressed over 3–5 hours. He breathes oxygen from a mask while the rest of the chamber is filled with air. If the whole chamber is filled with oxygen, the **fire risk** increases dramatically.

Because the attendant breathes chamber air, great care must be taken to monitor his dive profile to avoid the embarrassing predicament of an attendant emerging from the RCC with DCS.

The diver breathes oxygen for the duration of the treatment except for 5 minute periods of air – or sometimes heliox ($He\text{-}O_2$) breathing – every 25 minutes, to minimise oxygen toxicity to the **lungs**. The regimen is not without hazard as there is a significant risk of **cerebral oxygen toxicity** causing convulsions in such persons.

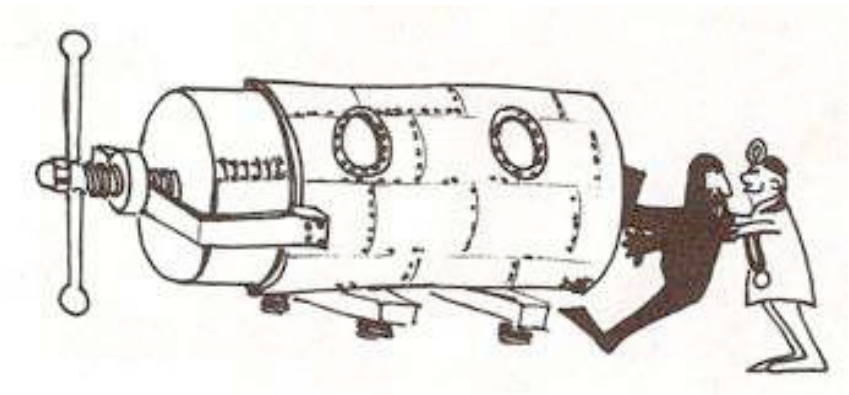

Fig. 16.3

Other tables are available which involve compressing the diver to a depth equivalent of 30–50 metres sea-water (100–165 ft.) in **air**, **heliox** and **nitrox** mixtures and then decompressing over periods ranging from several hours to several days depending on the severity of the symptoms.

Many other investigations and treatment modalities will be employed by experienced physicians in the RCC, including fluid balance, medications, etc, which need not concern the average diver.

A further emergency procedure, **Underwater Oxygen Treatment,** has been devised for use under expert supervision in remote localities. The diver is **recompressed in the water** to a maximum of 9 metres while breathing 100% O_2. Details of this procedure are outlined in Appendix C.

Treatment in water with the diver **breathing air** has been used in many parts of the world and water treatment tables are contained in some Navy diving manuals. While success has often been reported and delay in treatment can be avoided, this form of treatment has serious theoretical and practical difficulties which can result in worsening of the diver's condition.

The deep water air requirements (30 metres initially, and decompression for many hours)) renders the patient and accompanying divers prone to cold (hypothermia), narcosis, gas exhaustion, tide and other current changes. Attending divers may well develop DCS from extra exposures. This form of treatment is not generally recommended, unless other options (RCC, underwater oxygen treatment, etc.) are unavailable.

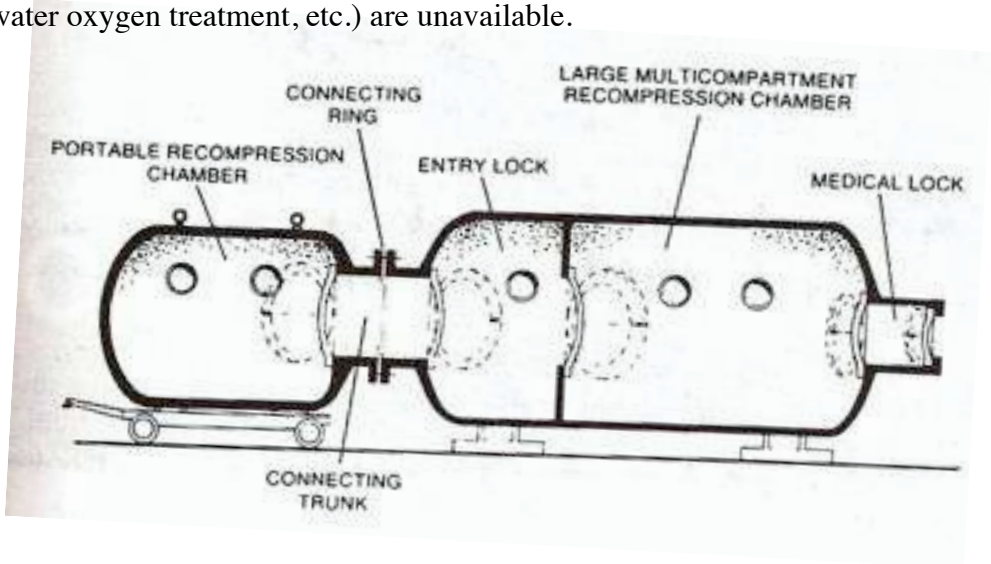

Fig. 16.4

Schematic outline of large static RCC treatment facility showing linking of a small portable chamber to the larger fixed unit. Both chambers are pressurised to identical pressures for a "transfer-under-pressure" (TUP).

❑ Hazards of therapeutic recompression.

While therapeutic recompression generally produces a dramatic relief of symptoms, it has several serious hazards. These include **oxygen toxicity** in the patient, the risk of **fire** and the risk of producing **DCS in the attendant.** It should only be used under the close supervision of a medical officer experienced in its use.

PREVENTION OF DCS

There are a number of factors which predispose to DCS. These are described in Chapter 13. Obviously these should be avoided wherever possible but some are unavoidable. Apart from falsifying a birth certificate there is little an individual can do about the predisposing factor of age.

Because of the unreliability of the currently available dive tables and DCs, and the unpredictability of the development of DCS, it is possible for even the most careful, well trained and conscientious diver to develop this condition. The following suggestions will help reduce the risks.

It is important that the diver never exceeds the no-decompression limits and ascent rates. In spite of this, all the dive tables, and especially the dive computers so far devised, have a significant failure rate.

❑ "Fudge factors".

The rate of development of DCS varies from less than 1% to as much as 5% depending on the table or computer algorithm used, the depth and the duration of the dive, and if these are pushed to their limits. The apparent safety of tables and algorithms are probably improved by intelligent divers incorporating "fudge factors" of their own. This is especially required for older, fatter, less fit divers. Also for more important people, such as your children, spouses, best friends etc. Fudge factors imply ascending slowly before it is theoretically necessary – reducing bottom times and depths.

Some fudge factors may be incorporated into the dive as follows:

- For decompression tables; decompress assuming that the dive was carried out at a greater depth and/or duration than was the case. Ascend earlier and at a slower rate than required.
- For dive computers; ascend well before required to do so according to the display. Select the most conservative mode permitted by the computer or chose an altitude based decompression whilst diving at sea level

❑ Accurate depth & time.

It is essential that the diver knows accurately the depths and durations of his diving. A depth gauge which indicates the maximum depth attained is useful, because it is common for divers to descend deeper than they realise. An underwater **watch** or, better, a **bottom timer** is valuable, as is a D.C.

❑ No-decompression diving.

Although the tables and computers are not totally reliable, they are less reliable for deep diving (greater than 30 metres). It is advisable to avoid pushing the dive to the limits when a no-decompression schedule is followed and to avoid dives requiring decompression.

❑ Slow ascent rates.

A slow ascent is prudent and the diver should certainly not ascend faster than the rate recommended by the tables. Preferably a slower ascent rate should be employed (8–10 metres or 25–33 ft. per minute is an acceptable safe rate) and the extra time taken is deleted from the bottom time. i.e. ascend earlier than permitted by the tables.

❑ Routine decompression stops.

Most authorities recommend a routine minimum safety ("decompression") stop at **3-5 metres for 3-5 minutes**, after a no-decompression dive greater than 15 metres, to allow partial nitrogen elimination and trapping of venous emboli in the lung vessels.

❑ Exercise

Avoid strong exertion as far as possible during the dive and decompression staging. Gentle exercise may assist in de-gassing during the ascent, staging and post dive, although some would advise no exercise at all post-dive. The latter would be applicable if bubble formation had occurred.

❑ Dive planning.

When repeated dives or multi-level dives are planned, the **deeper dives** should always be **performed first.** Recreational repetitive dives on the same day should have long surface intervals between dives, preferably 4 or more hours if possible, and a maximum of 3 dives per day..

With multi-day diving, a rest day is included after each 3 continuous diving days.

With **deep diving**, gradual build up (**acclimatisation**) is achieved by progressively deeper exposures.

❑ Post-dive restrictions.

It is advisable to **rest** for an hour or more after a deep or long dive to ensure elimination of nitrogen from the fast tissues. Surface intervals should be > 2-4 hours.

Flying and significant altitude exposures within 24 hours of diving is not recommended. The DCIEM recommendation is “whenever possible it is inadvisable to fly above 600 metres in any aircraft within 48 hours of completing any dive. Travelling by vehicle over mountain ranges or hills can expose divers to the same dangers as flying and should be avoided in the same way for 24 hours. If flying after diving is considered essential, flying may be carried out after 24 hours but the increased risk of DCS must be borne in mind.”

❑ **Dive computers** (See Chapter 14).

.A healthy scepticism towards reliance on any mechanical equipment, especially if promoted by a glossy brochure or a dive computer salesman, also has good survival value.In addition to the above precautions, the diver is advised to buy a good quality waterproof rabbit's foot.

Chapter 17

DYSBARIC OSTEONECROSIS

(DIVERS BONE DISEASE, AVASCULAR NECROSIS OF BONE, ASEPTIC BONE NECROSIS, BONE NECROSIS, BONE ROT, CAISSON DISEASE OF BONE)

This was first noticed in caisson (tunnel) workers in the 19th century, and was described as being an area of localised bone death, predominantly occurring in the long bones of the arms and thighs.

If this area of dead bone is located beneath the joint surface of the bones in the hip or shoulder joints, pain and symptoms of arthritis, along with a reduction in mobility of the joint is a common consequence – often occurring in mid or later life.

The exact cause of the disorder is probably a delayed effect of damage caused by gas bubbles produced during a dive (see Chapter 13). In this sense it is a delayed form of decompression sickness.

Cause

Bone is a living organ containing bone cells which constantly absorb and lay down new bone. It has a cleverly designed structure which resembles reinforced concrete or fibreglass and contains fibres of a sinew-like substance called collagen, embedded in a concrete-like calcium material. This is traversed by numerous vessels which supply the blood to the bone cells embedded in the bone. The bone cells permit the repair of fractures and allow the bone to change its structure to accommodate stresses which may vary during the person's life.

If the blood vessels supplying the bone cells are blocked by gas bubbles or any other cause, the bone cells die and the self-repairing ability of the bone stops. It becomes unable to fix the

repeated minor trauma which is common around joints and eventually the bone structure collapses causing permanent damage to the major load bearing joints, such as the hips or shoulders.

Fig. 17.1

The reported incidence of this condition varies from less than 1% in some Navy series, to 80% in Chinese commercial divers. This variance is probably due to factors such as different diagnostic criteria and differing ages, dive patterns, deco procedures and dive frequency.

Predisposing factors which are commonly associated with osteonecrosis include:

- **age greater than 30 years**
- **inadequate decompressions**
- **experimental dives**
- **deep dives**
- **decompression sickness**
- **long duration dives**

X-Ray changes have been seen as soon as 3 months after a dive and it has been reported following a single dive to 30 metres. When joint involvement does occur, the onset of symptoms is usually delayed for many years, reflecting the time required for joint destruction. Fortunately, in most cases the disease does not cause any serious damage to the joints and so produces no symptoms.

Occasionally bone pain may increase during recompression treatment, or may persist after treatment. Some of these progress to dysbaric osteonecrosis over the following months.

Classification of Bone Necrosis

The lesions are classified into two groups:

- **Type A lesions** – which are near the joint surface (juxta-articular).
- **Type B lesions** – which are remote from the joint surface (head, neck and shaft).

❑ Type A lesions.

With these, the joints may become involved as the under-lying bone is destroyed and the joint surface collapses. This may produce symptoms which are potentially crippling. Hips and shoulders are more frequently affected.

❑ Type B lesions.

These rarely cause symptoms and are generally of little clinical importance, except to suggest more conservative diving procedures. The most common areas affected are the long bones of the thigh, leg and upper arm. Occasional cases of bone cancer have developed in these lesions.

Clinical Features

When Type A lesions injure the joint, common symptoms are pain, which is usually aggravated by movement, in the affected joint and accompanied by a restriction of joint movement. As the condition progresses, severe osteoarthritis develops and the joint may eventually become frozen and incapacitating, due to pain.

Investigations

The lesions can also be identified in excellent detail, using MRI (Magnetic Resonance Imaging) scanning within days of the incident (but some of these may disappear). This is expensive but has no associated risks of irradiation. Early lesions can also be identified with radioactive Technetium ("bone scans"), which will bind to an osteonecrotic area and can be detected with a scanner within 2 weeks of the injury.

X–Rays have been the traditional investigative method but these will only reveal lesions once bone changes have developed. This may take months or years.

Treatment

The pain associated with movement can be reduced with an anti-inflammatory drug such as NSAIDS. Severe cases may require the fusion of a joint or its replacement with a synthetic joint made of either metal or plastic. While this procedure relieves the pain and increases mobility, a synthetic joint is never as robust as the "natural model" and its endurance is limited. Other techniques involve replacement from the divers healthy bone, or realignment of articular surfaces.

As the disease is regarded as an occupational hazard of diving, workers compensation claims may help off-set expensive medical costs.

Replacement hips may cause problems.

iously desirable. Most sensible y they should; avoid dives deeper han 40 metres, avoid dives equiring decompression, not pproach the no-decompression mits and ascend slowly. It is kely that the longer duration ives permitted with many dive omputers, together with the ncreased number of dives and the bility to dive close to the edge of ecompression commitment, now nakes this disorder more likely for ecreational divers.

f indicated, in divers who develop ecompression sickness a follow-p bone scan after 2–4 weeks hould detect areas of bone amage. MRI imaging can often e better used to determine the xtent of the lesion.

Divers with high risk factors who develop unexplained joint pain should be assessed to exclude this condition.

Occupational divers and other divers who are at increased risk because of their diving practices may require regular routine screening assessments. Since X-ray investigations can involve worrisome exposure to radiation, their frequency must be weighed against the risk of osteonecrosis development, and so MRIs are preferred. Divers who are likely to be at risk are required to have a baseline investigation performed before they are employed. For some susceptible occupational divers, follow-up assessments at 5 year intervals are recommended.

Replacement hips may cause problems.

Fig 17.3

Chapter 18

NITROGEN NARCOSIS

(COMPRESSED AIR INTOXICATION, RAPTURE OF THE DEEP, INERT GAS NARCOSIS, NARCS)

Intoxication in divers is not confined exclusively to beach barbeques and hotel bars. When breathed under pressure, nitrogen (which makes up 78% of air) has an intoxicating effect which, like alcohol, is variable and may lead to pleasure or disaster.

This phenomenon was regarded as an annoyance to the helmet diver who could be pulled to the surface by his attendant if he behaved irrationally, but the consequences to the scuba diver, who's safety is dependent on a buddy exposed to the same effect, can be more serious.

It will be present in all divers breathing air at a depth in excess of 30 metres, although some will notice it earlier. Others may not be aware of the effect, as judgment and perception are affected. The severity of symptoms and the exact depth of their onset varies between individuals. Because of narcosis, diving on air beyond 30 metres (100 feet) is not prudent, and 40 metres is considered unsafe for most recreational divers. A 50 metres depth is considered the maximum safe depth for experienced professional divers breathing air.

CAUSES OF NITROGEN NARCOSIS

The exact cause of this narcotic effect is uncertain. Nitrogen is classified as an inert gas because it does not participate in any chemical reactions within the human body. The influence of nitrogen on narcosis must therefore be due to some physical reaction.

When other inert gases such as neon, xenon and argon were investigated, it was found that their narcotic effect at depth correlated approximately with the relative weights of their individual molecules (i.e. their molecular weights). An increased molecular weight caused a greater narcotic effect. It was further shown that the inert gases which were more soluble in fat than water, tended to have a greater narcotic effect. There were unfortunately several inconsistencies in the behaviour of these gases, including hydrogen, which cast some doubt on these generalisations.

Other theories have been proposed implicating oxygen or carbon dioxide toxicity, lipid solubility and enzyme changes in the brain.

CLINICAL FEATURES

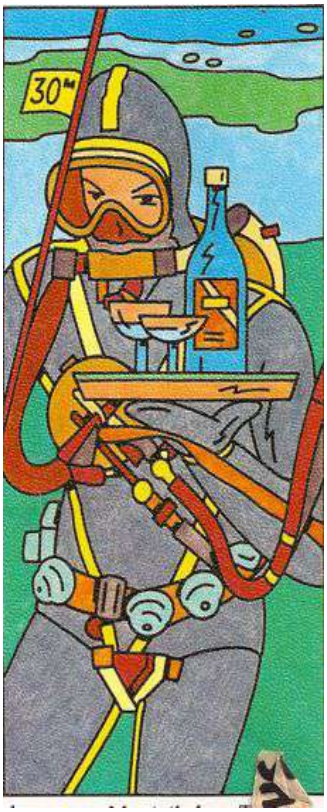

Fig. 18.1

The narcotic effect usually becomes effective within a few minutes of reaching a particular depth and does not worsen as exposure continues at this depth. Rapid descents may increase the effect, but with ascent it is dissipated.

The higher brain functions such as reasoning, judgment, memory, perception, concentration and attention tend to be the first affected by narcosis. This often leads to a feeling of well-being and stimulation in a diver secure in his surroundings. In a novice or an apprehensive diver, a panic reaction may follow. Some degree of tolerance develops at a given depth or with repetitive exposures.

The influence of narcosis may not be evident if the dive is uneventful, thus giving a false impression that the diver is in control of the situation. Memory and perception deficits may only be evidenced by a failure to follow instructions or the dive plan, or being inattentive to buoyancy, air supply or buddy signals. When a problem develops, the diver may be unaware of this – attention and perception being focused elsewhere (perceptual narrowing or "tunnel vision"). Thus emergency signals will go unheeded, emergency air supplies will not be offered, weight belts will not be released, rescue attempts will be crude and amateurish. Survival instincts and responses may be dampened. The safety of both the diver and his buddy are compromised.

Death may supervene due to errors provoked by impaired judgment or perception, and by over confidence. Loss of consciousness may happen without warning and be unnoticed by the diver's buddy. At great depths the diver may lose consciousness from the narcosis itself or the interaction between it and other factors such as sensory deprivation, carbon dioxide or oxygen toxicity.

Martini's Law - Table

20-30 metres	Mild impairment of performance on unpracticed tasks, mild euphoria
30-50 metres	Overconfidence and inadequate responses to danger. Perceptual narrowing, fixation on a particular function or exercise. Judgment impairment, affecting; air supply, buoyancy control. navigation, decompression obligations, ascent rates etc. Anxiety
50 metres	sedation , loss of judgment. Hallucinations possible.
50-70 metres	In a chamber, depending on conditions - talkative or terrified.
70 metres	Poor reasoning ability. Very poor response to signals or instructions.
70-90 metres	Poor concentration and mental confusion, stupefaction and loss of memory.
> 90 metres	Hallucinations and loss of consciousness.

Factors which are known to **increase** the effects of nitrogen narcosis include:

- **low intelligence**
- **fatigue or heavy work**
- **anxiety, inexperience or apprehension**
- **cold (hypothermia)**
- **recent alcohol intake or use of sedative drugs (includes seasickness medications), marijuana etc.**
- **poor visibility**

Factors which tend to **reduce** the effects of narcosis include:

- **strong motivation to perform a given task**
- **acclimatisation following prolonged or repeated exposures**
- **tolerance to heavy alcohol intake**

A diver who can "hold his liquor" is said to have a greater tolerance to nitrogen-narcosis. A plea of "acclimatising to narcosis" is generally not accepted by the courts however, as defence for an alcoholic intoxication charge.

The effect of nitrogen narcosis has been likened by some to that of drinking one martini on an empty stomach for every 10 metres depth (**Martini's Law**). The "olive" appears to be optional.

Case History Examples:

1. A group of divers descended into a deep clear freshwater cave in order to savour the pleasant intoxication of narcosis. Their bodies were found some weeks later in a deep confine of the cave. They were victims of over-confidence and impaired judgment induced by nitrogen narcosis.

2. Another diver became so elated during his dive that he removed his regulator and offered it to the other marine inhabitants.

3. A diver developed problems with his air supply but, possibly because of the 40 metre depth and narcosis, he did not attempt to ditch his weight belt. He triggered the dump valve of his BC instead of the inflation valve, and drowned with minimal struggling.

PREVENTION

Avoidance of compressed air diving to depths known to cause narcosis is a good policy. This implies a depth limit of 30 – 40 metres (100 – 130 feet) depending on the diver's experience, his tolerance to narcosis and the task performed. Safe diving beyond 30 metres requires an awareness of the ever increasing risk of this condition and its effects on human performance and judgment. Some experienced professional divers may be able to perform certain practised tasks at depths up to 60 metres with competency, but dives greater than 30 metres should be a source of concern for recreational divers and greater than 50 metres should be regarded as excessive even for professionals.

TREATMENT

A diver incapacitated by narcosis should be protected from injury and inappropriate behaviour, and bought to a shallow depth with a controlled ascent, bearing in mind decompression requirements. Symptoms clear rapidly as the nitrogen pressure is reduced. Any other symptoms present on surfacing (e.g. salt water aspiration and near drowning, decompression sickness etc.) are due to complications of experiencing narcosis at depth and not narcosis *per se*.

Chapter 19

HIGH PRESSURE NEUROLOGICAL SYNDROME

(HIGH PRESSURE NERVOUS SYNDROME, HPNS, HELIUM TREMORS)

This condition is a problem for deep commercial diving operations where helium/oxygen (Heliox) mixtures are breathed at depths in excess of 130 metres (430 feet). It causes a serious limitation to very deep diving and gets worse as the depth increases.

It is not a problem to recreational divers unless a grave (sic) miscalculation of depth or buoyancy has been made.

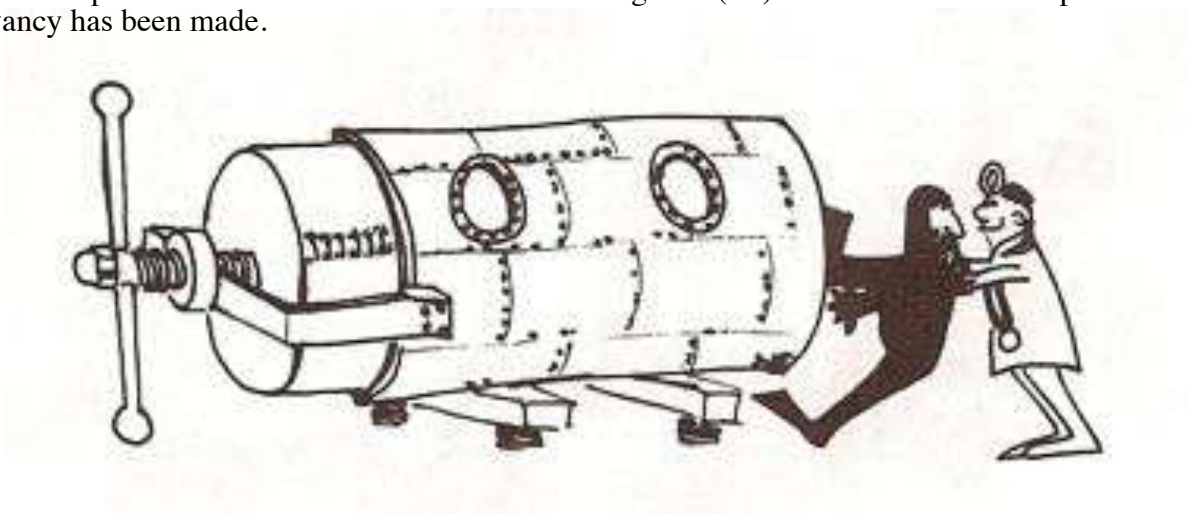

Fig. 19.1

CLINICAL FEATURES

The first sign is usually a mild uncontrollable tremor, with muscle twitching and difficulty in coordinating movements. If the diver then continues with the descent, confusion, drowsiness, disorientation and unconsciousness may follow. Respiration may also be affected by the neurologically induced incoordination of muscle activity. A tremor particularly affects the hands and arms and resembles the shivering due to cold (which helium breathing may also produce).

The condition is aggravated by rapid descent and thus a slow descent rate is a requirement in all deep diving operations using exotic gas mixtures.

The cause of HPNS is not fully understood. It is probably due to an excitation of a part of the brain by the direct mechanical effect of pressure. Evidence for this is that drug induced anaesthesia in animals can be reversed by simply compressing the animal to depths which provoke HPNS.

TREATMENT AND PREVENTION

Since the effects of HPNS resembled an excitation of the brain, early researchers reasoned that an agent which caused sedation might reverse the condition. Nitrogen narcosis is a common cause of sedation during diving, and so divers affected by HPNS were given small concentrations of nitrogen to breathe. This produced an effective reduction of some of the symptoms of HPNS. This helium-nitrogen-oxygen mixture (Tri-mix) is now used in most deep diving and by some technical divers. Small percentages of nitrogen have been included to reduce the HPNS – but not enough to cause significant nitrogen narcosis.

The added nitrogen in Tri-mix also permits better speech comprehension (because helium distorts sound production in the human larynx) and may slightly reduce the rapid heat loss with helium.

Chapter 20

HYPOXIA

(ANOXIA)

Hypoxia refers to an inadequate level of oxygen (O_2) within the cells. **Anoxia** implies there is no O_2 left at all in the cells, and is uncommon. Without O_2 most cells, especially those of the brain, die within a few minutes. This is the final outcome of many diving accidents and is often the ultimate cause of death.

Hypoxia is caused by an interruption in the chain of physiological processes (see Chapters 3 & 4) which bring O_2 from the outside air (or breathing gas) to the body's cells. There are **four links** in this chain where interruption can cause hypoxia, supplying a logical classification.

CLASSIFICATION OF HYPOXIA

HYPOXIC HYPOXIA

Inadequate Oxygen content in Arterial Blood

In diving, the most common form of hypoxia is **hypoxic hypoxia**. This usually follows an inadequate air supply and/or salt water aspiration or drowning. Either there is inadequate O_2 getting to the lungs because the diver has, for a variety of reasons, only water to breathe, or the lungs are unable to convey inhaled O_2 to the blood because of alveolar damage due to near drowning.

An inadequate air supply can also arise from an inadequate concentration of O_2 in the breathing gas (e.g. a gas mixture in which O_2 has been inadvertently omitted or an internally rusty scuba cylinder which extracts O_2). It may develop from equipment failure or obstruction somewhere in the respiratory tract between the nose or mouth and the alveoli, due to :

- upper airway obstruction due to unconsciousness
- tracheal obstruction from inhalation of vomit and
- alveolar damage from salt water inhalation (see Chapters 25 and 26).

Pulmonary barotrauma is another cause. A less common, but often catastrophic, cause is decompression sickness, when the gas bubbles are carried to the lungs in such quantities that the lungs cannot cope with them – resulting in "chokes".

STAGNANT HYPOXIA

O_2 Taken Up by the Blood Fails to Reach the Tissues

This is generally due to failure of the heart to pump the blood adequately to the tissues (e.g. from a heart attack or air embolism). Poor circulation to the extremities in cold conditions can cause localised hypoxia to these areas without generalised hypoxia.

ANAEMIC HYPOXIA

Inability of the Blood to Carry O_2 in the Presence of Adequate Circulation

This is generally due to inadequate amounts of circulating functional haemoglobin, usually from blood loss or carbon monoxide poisoning (see Chapter 23).

HISTOTOXIC HYPOXIA

Inability of the Cells to Use the O_2

This is caused by certain poisons including carbon monoxide (see Chapter 23).

HYPOXIA IN BREATHHOLD DIVING

Drowning Syndromes

See Chapters 25 and 26. The inhalation of water into the lungs is the commonest cause of hypoxia in all types of recreational diving.

Hyperventilation

As explained in Chapter 4, hyperventilation before a breath-hold dive reduces the urge to breathe during the dive and may cause the diver to lose consciousness from hypoxia while still underwater (see Case Histories 33.2, 33.3) and with little or no warning. Drowning frequently results from this.

Hypoxia of Ascent

The partial pressure of O_2 in the lungs falls as they expand during ascent from a breath-hold dive. In some circumstances, this can cause loss of consciousness from hypoxia during ascent. Details are explained fully in Chapter 4.

HYPOXIA IN COMPRESSED GAS DIVING

Scuba

Exhaustion of the air supply, equipment malfunction, regulator resistance or loss of the demand valve will leave the diver with nothing but water to breathe – inevitably resulting in hypoxia due to salt water aspiration or drowning. Panic and poor dive techniques are often precursors to these problems.

Diving induced asthma, pneumothorax (from pulmonary barotrauma) and decompression sickness (chokes) can also interfere with breathing sufficient to cause hypoxia.

Rebreathing Equipment

This type of equipment shares the same causes of hypoxia as scuba equipment, but has some additional hazards.

A hypoxic gas mixture can be breathed if the **wrong gas** or **wrong mixture** is used (i.e. a gas mixture containing insufficient O_2). A specific example of this is when a gas mixture intended for use at great depth (e.g. one containing 5% O_2) is breathed near the surface. With rebreathing equipment using a constant flow of gas, the **flow of O_2** must be sufficient for the energy needs of the diver, and this will increase with **exertion**. Thus, O_2 concentrations too low, inadequate flow rates or excessive O_2 consumption may all lead to hypoxia in rebreathers.

Dilutional hypoxia is a particular problem with rebreathing equipment. When the diver first begins to breathe from the diving set, a significant amount of nitrogen may be displaced from the lungs and body into the counterlung (breathing bag) of the equipment. If this is not vented after a few minutes breathing, the diver is likely to rebreathe almost pure nitrogen from the rebreathing bag. Oxygen is consumed by the diver and the carbon dioxide produced is

absorbed by chemicals used in the equipment. Because the counterlung will still contain gas (mostly nitrogen) the diver will be unaware of the danger. The O_2 % may be adequate at depth, but not near the surface.

CLINICAL FEATURES

In most cases of hypoxia, the diver is unaware that there is anything wrong and therefore can lose consciousness without warning.

Mild hypoxia starves the brain of O_2, causing confusion, impaired judgment and clouding of consciousness. The diver is frequently unaware that there is a problem and may even become over-confident. Sometimes he may notice a loss of colour vision, but this is infrequent. An observer should notice a deterioration of performance.

More **profound hypoxia** causes unconsciousness and in some cases, muscular jerking and spasms (especially of the jaw muscles) or epileptic type fits. **Severe hypoxia** results in rapid death.

Hypoxia makes the blood blue in colour. Hypoxic blood in the body capillaries gives the skin a blue appearance, and is termed **cyanosis**. It is not easily detected under the blue water, but is often obvious when the diver is surfaced

There is a form of localised cyanosis (stagnant hypoxia) associated with **cold** which does not denote generalised hypoxia. This is seen in the fingers and ears, due to (peripheral) blood vessel constriction causing inadequate circulation in these areas, in response to cold. It can be distinguished from the cyanosis of generalised hypoxia (hypoxic hypoxia) by looking at the colour of the tongue. The tongue is blue only in the generalised hypoxia.

TREATMENT

If hypoxia is due to insufficient O_2 in the cells, treatment should aim to reverse this. The basic resuscitation principles should be applied first (see Chapter 42).

A. Clear the airway.
B. Establish or maintain breathing.
C. Establish or maintain circulation.

Give the patient the highest possible O_2 concentration to breathe and use positive pressure if needed (see Chapter 40).

All diving operations should be undertaken with emergency O_2 equipment available, sufficient to last until the diver can access more formal medical facilities. A rule of thumb is to allow for 15 litres of O_2 per minute. There should always be at least one person, who is not diving, cognisant of and trained in the application of the O_2 equipment. This is usually the boatman or dive master.

PREVENTION

Most diving deaths are ultimately caused by the hypoxia associated with drowning, regardless of the initial problem. In many cases this can be avoided by the **buddy system**, which permits the diver to be rescued and employ positive **buoyancy** for the victim**.** A good buoyancy compensator should keep even an unconscious diver's face clear of the water.

Avoid prolonged or deep **breath-hold dives**. **Never hyperventilate** before a breath-hold dive and avoid exertion as much as possible.

With **scuba,** maintain equipment adequately and check it before a dive. **Monitor the air supply** continuously, using a contents gauge.

With **re-breathing equipment**, check the O_2 concentrations before diving, also the gas flows and pressures. Avoid excessive exertion and flush through with appropriate gas before ascending.

Chapter 21

OXYGEN TOXICITY

Oxygen (O_2) is toxic when breathed at a partial pressure in excess of 0.4 ATA (40% O_2 at atmospheric pressure) for sufficient time. The greater the concentration and pressure, the shorter the time. The two common forms of O_2 toxicity affect the lungs and the brain.

When O_2 is breathed at partial pressures between 0.4 and 1.6 ATA it is eventually toxic to the lungs. At partial pressures in excess of 1.6 ATA, it is toxic to the brain as well as the lungs. The effects are more pronounced and more rapid as the inspired partial pressure of O_2 increases.

MECHANISM

The exact cause of O_2 toxicity is unknown. It is generally considered that hyperbaric O_2 interferes with the activity of enzymes in the cells and that this disrupts the biochemical functions, particularly in the brain and lungs.

In the lungs, damage to the cells lining the alveoli causes a general thickening and stiffening of the lung tissues, accumulation of fluid and difficulty with breathing.

In the brain there is a reduction in the amount of certain nerve transmission chemicals as well as generalised damage to the nerve cells. If cerebral O_2 toxicity is allowed to develop, convulsions eventually follow.

The sensory organs are really neurological outposts. Thus vision, hearing and touch, may also be affected.

PREDICTION OF O_2 TOXICITY

To calculate the inspired partial pressure of O_2, multiply the percentage of inspired O_2 by the ambient pressure in atmospheres absolute and divide by 100.

e.g. the partial pressure of O_2 in room air is :

$$21\% \times 1\text{ATA} \div 100 = 0.21 \text{ ATA}.$$

The risks of O_2 toxicity increase with increasing partial pressure. In general it is usually possible to breath 100% O_2 (1ATA) for 12–24 hours without developing significant pulmonary O_2 toxicity. This duration reduces as the O_2 pressure increases. If therapeutic recompression is contemplated, a maximum period of only 6-12 hours breathing 100% O_2 may be acceptable since the subsequent therapeutic recompression will generally involve the use of hyperbaric O_2, and this will summate with existing O_2 toxicity. The amount of pre-treatment of diving casualties with O_2 will preferably be discussed with the diving physician responsible for the therapeutic recompression.

Oxygen toxicity results from a combination of O_2 pressure and duration of exposure. Both must be considered and tables are available to indicate the maximum values allowable for different pressures and durations, for both respiratory and cerebral O_2 toxicity.

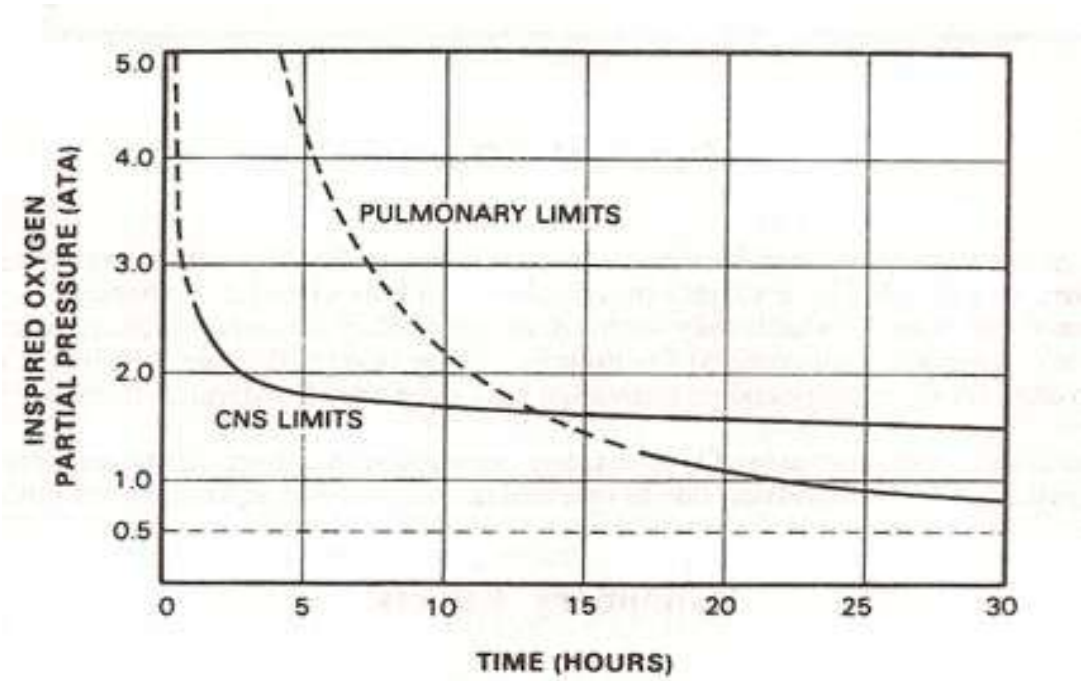

Fig. 21.1

This graph shows the predicted pulmonary and cerebral toxicity limits of exposure to varying partial pressures of oxygen. It can be noted that oxygen can be tolerated for much longer periods at lower partial pressures.

CAUSES OF O_2 TOXICITY

For **resuscitation, 100% O_2** should be used for hypoxic diving casualties without any fear of O_2 toxicity. As mentioned above, the treatment of decompression sickness and air embolism cases includes 100% O_2, even before consultation with the diving physician regarding any potential negative effects.

Oxygen re-breathing equipment should be restricted to military, commercial and trained technical divers use and diving with this should not be attempted by recreational divers. O_2 diving sets have an absolute depth limit of 9 metres for resting dives and 8 metres for working dives in order to reduce the risk of convulsions. Rebreathing and scuba sets employing nitrogen/O_2 (**nitrox**) mixtures are limited to depths which produce an inspired O_2 partial pressure of no more than 1.6 atmospheres, and often less.

In **deep diving** operations, gas mixtures of helium, nitrogen and O_2 should have the composition adjusted so that the inspired partial pressure of O_2 never reaches the toxic range.

Therapeutic recompression using O_2 tables often involves the compression of the diver to 2.8 atmospheres while breathing 100% O_2. There is a significant risk of both pulmonary and cerebral O_2 toxicity and these tables should only be employed on the advice of diving medical experts.

CLINICAL FEATURES

Cerebral Effects

In this case the earliest symptom may be a convulsion which can develop without any warning. It may sometimes be preceded by a variety of features such as facial pallor, visual or auditory disturbances, tunnel vision, faintness, or facial twitching – which are often not evident underwater. Nausea, retching and even vomiting are common with cerebral O_2 toxicity, as are anxiety and palpitations. There is considerable individual variation in susceptibility to cerebral O_2 toxicity and an individual may vary in his tolerance from day to day. It may be increased by anything that increases carbon dioxide levels, such as exercise, immersion, resistance from breathing equipment and nitrogen narcosis.

During therapeutic recompression using O_2 tables, any convulsion in a diver due to cerebral O_2 toxicity must be distinguished from a convulsion due to cerebral decompression sickness or air embolism. Sometimes the convulsions occur soon after the O_2 supply is removed (the "off effect").

Pulmonary Effects

The early symptom is an irritation deep in the central part of the chest, progressing to pain and a burning sensation which is aggravated by inspiration and accompanied by coughing. As the condition develops, shortness of breath ensues and a pneumonia type illness supervenes. Although the early symptoms are reversible, progressive serious symptoms may cause permanent lung damage or even death.

Fig 21.2

TREATMENT

Cerebral Effects

Whilst undergoing therapeutic recompression, if warning signs of cerebral toxicity develop, the patient should be encouraged to hyperventilate and then be given air to breath until the symptoms abate. Modification to the O_2 treatment table may then be necessary.

If the patient convulses he should be placed on his side to protect the airway from obstruction or aspiration of stomach contents. He should be protected from injuring himself on nearby solid objects. A padded mouth piece may be gently placed between the teeth to protect the tongue. After the convulsion has ceased the patient may be unconscious for a short time. His airway should be protected and he should be managed according to the principles outlined in Chapter 42. See Case Report 24.2.

Pulmonary Effects

These effects will usually resolve spontaneously if the supplementary O_2 administration is ceased as soon as symptoms develop. If it is essential to continue O_2 therapy however, a reduction in the partial pressure of O_2 given will slow the development of toxicity. Short periods of 'air breathing' (or Heliox), 5 minutes every half hour, are often used by experienced doctors to delay oxygen toxicity during O_2 therapy.

CONCLUSIONS

1. Recreational divers should not use O_2 enriched diving equipment. Technical divers should not expose themselves to O_2 pressures greater than 1.6 ATA or durations that could cause respiratory manifestations.

2. Resuscitation training with O_2 equipment is of great value to divers and dive boat operators. In diving accidents, the delayed risks of O_2 toxicity are outweighed by the benefits of treating the hypoxic diving casualty.

3. The use of O_2 in the first-aid treatment of decompression sickness and pulmonary barotrauma should always be undertaken whilst bearing in mind the prospect of eventual pulmonary oxygen toxicity. Breathing air for 5 minutes after 25 minutes of O_2 is one way of reducing the risk of pulmonary toxicity, but this should be discussed with the diving physician who will ultimately manage the case.

4. During therapeutic recompression using O_2, the use of short air or Heliox breaks during the treatment reduces cerebral and pulmonary O_2 toxicity.

5. There are other logistical problems with the use of oxygen, and some of these are discussed in Chapter 40.

Chapter 22

CARBON DIOXIDE PROBLEMS

Carbon dioxide (CO_2) is the gaseous by-product produced when the body consumes oxygen to fuel its metabolic processes. The body has an efficient way of disposing of CO_2, mainly through buffering systems in the blood and exhalation from the lungs.

CARBON DIOXIDE INSUFFICIENCY (HYPOCAPNOEA OR HYPOCAPNIA)

Hypocapnea refers to a blood carbon dioxide (CO_2) level below normal. The CO_2 partial pressure in the blood is normally maintained within narrow limits by a biological feedback mechanism. Voluntary or involuntary hyperventilation (overbreathing) will overcome this regulatory mechanism and lower the blood CO_2 level. The most common cause for this is the rapid sighing respiration associated with hysterical and anxiety states – the feeling one experiences on confronting a great white shark eye to eye.

A number of divers (fewer each year, due to natural selection) deliberately hyperventilate to lower their blood CO_2 level, before a breath-hold dive, in order to prolong the dive. They often succeed beyond their wildest dreams. The lethal consequences of this practice are explained in Chapter 4.

With most scuba equipment there is an increased breathing resistance, and this tends to diminish the likelihood of hyperventilation and hypocapnoea, Increasing depth also increases this resistance to breathing. Some more sophisticated equipment allows for assisted or pressure supplemented respiration, and this will increase the likelihood of hypocapnoea.

Clinical Features

A person hyperventilating from anxiety is not usually aware of an altered breathing pattern, although it may be evident to an observer. Hyperventilation causes increased resistance to breathing with scuba, and this causes more anxiety.

Symptoms include tingling or "pins and needles" (paraesthesiae) of the fingers, dizziness and light headedness, an altered conscious state or confusion. Muscular twitching or spasms can occur in extreme cases.

Treatment

The simplest treatment for hypocapnea is to reduce the breathing rate and depth. This restores the blood CO_2 level and cures the symptoms. On land, doctors often advise the patient to breathe in and out of a brown paper bag (rebreathing), but underwater most divers are not prepared to replace their regulator with a soggy paper bag.

Alternative Diagnoses

It is important to exclude other serious conditions such as decompression sickness, air embolism, carbon monoxide poisoning and salt water aspiration, whenever a diver presents with the symptoms of hypocapnoea. These illnesses can in themselves, also cause apparent hyperventilation and can mimic anxiety states.

CARBON DIOXIDE TOXICITY (HYPERCAPNOEA)

CO_2 toxicity is due to accumulation of CO_2 through excess production or inadequate ventilation (breathing).

The excess production is usually due to metabolism from increased exertion. Whereas only 0.5 litres/minute of CO_2 is produced at rest, this can rise to over 3 litres during maximal exercise.

Inadequate ventilation can be caused by breath-holding, breath control ("skip breathing"), rebreathers or extension of the respiratory passages ("dead space") with snorkels, etc.

The effect of depth on inspired partial pressure is important. While 3% inspired CO_2 may be tolerated at atmospheric pressure without significant symptoms, the same percentage at 20 metres (3 ATA) is the equivalent of 3×3 or 9% at the surface – a level which will cause serious toxicity.

Re-breathing exhaled CO_2 is the most common cause of CO_2 toxicity in divers. Hence, CO_2 toxicity is most commonly encountered with rebreathing equipment, but it can sometimes occur in diving helmets, compression chambers, saturation complexes (habitats) or possibly even scuba.

Causes of CO_2 Toxicity

❑ Rebreathing equipment.

Some types of military and technical diving equipment conserve gas and reduce exhaust bubble formation by allowing the diver to rebreathe his exhaled gas (exhaust bubbles can be detected by the enemy!). A canister of CO_2 absorbent (soda lime) is included in the circuit to remove the CO_2 which the diver exhales (see Chapters 5 & 43).

This mechanism can fail due to exhaustion of absorbent material, extended dive duration, salt water contamination, improper packing, excessive CO_2 production due to exertion, or improper assembly of the equipment.

❑ Diving helmet problems.

With a standard-dress helmet or with some helmets used in deep diving, the diver can partly rebreathe his exhaled gas if the fresh gas flow in the helmet is insufficient to flush out exhaled CO_2.

❑ Chambers and habitats.

CO_2 which is exhaled by chamber occupants must be removed by constant flushing of the chamber with fresh air or by the recirculation of the chamber gas through a CO_2 absorbent (scrubber). If either of these mechanisms is inadequate, the occupants can develop CO_2 toxicity by rebreathing their own exhaled CO_2.

❑ Scuba.

Since rebreathing is not possible with scuba equipment, CO_2 toxicity is not generally a problem for scuba divers unless there is excessive resistance to breathing (regulator resistance, increased gas density at depth) or a reduced respiratory response of the diver to CO_2 (possibly due to voluntary control or "skip breathing", adaptation, nitrogen narcosis, or high oxygen levels).

Clinical Features

These depend on the rate of onset and the actual partial pressure of the inspired CO_2.

A rapid accumulation of CO_2 may cause unconsciousness before any symptoms are experienced.

A slower build-up causes a variety of symptoms, including :

- **shortness of breath**, or air hunger.
- **flushing of the face** and **sweating** (sweating is not easy to detect underwater).
- **repetitive activity**, such as swimming, without awareness of this.
- **light headedness**, muscular **twitching**, jerks, tremors or **convulsions**.
- **impaired vision**.
- **unconsciousness**.
- **a splitting or throbbing headache**, usually at the front of the head. This may be severe and start after the CO_2 levels have been corrected. It often lasts for hours.
- **death**.

CO_2 toxicity may increase the likelihood of decompression sickness, oxygen toxicity, nitrogen narcosis and resistance to breathing (because of increased respiration). As with oxygen toxicity, there is sometimes an "off effect" whereby the symptoms of CO_2 toxicity are temporarily worsened when a diver suddenly resumes breathing normal gases after partially adapting to a high CO_2 pressure.

Treatment

Any diver, diving with rebreathing equipment, who experiences symptoms of CO_2 toxicity should immediately **cease exertion, inform his buddy, flush** the rebreathing system with fresh gas, then return to the surface by a **buoyant ascent** and **breathe air.**

Attendants of a surfaced diver suffering from CO_2 toxicity should isolate him from the source of CO_2 rebreathing, give **100% oxygen** by mask, and administer **basic life support** (see Chapter 42) including cardiopulmonary **resuscitation** if appropriate.

Other causes of headache and breathing difficulties such as pulmonary barotrauma, decompression sickness, carbon monoxide toxicity etc. should also be excluded (see Chapter 32).

The severe headache which follows CO_2 toxicity should be treated with a simple analgesic such as paracetamol (acetaminophen).

Chapter 23

CARBON MONOXIDE TOXICITY

Carbon monoxide (CO) is a gas produced by the incomplete combustion of carbon containing compounds. It is a component of the smoke from **engine exhausts**, slow combustion **stoves** and **cigarette** smoke. It can also be produced in divers' air compressors (see Chapter 24).

CO breathed in anything more than trace amounts can be lethal. It binds avidly to the oxygen (O_2) binding sites of haemoglobin (Hb) in the blood, preventing the haemoglobin from carrying O_2. CO bound to haemoglobin forms carboxyhaemoglobin (HbCO). If a sufficient number of the O_2 binding sites are occupied by CO, death from hypoxia ensues (see Chapter 20).

CO also binds with components of the energy-producing biochemical pathways in the cells, interfering with fundamental cellular function.

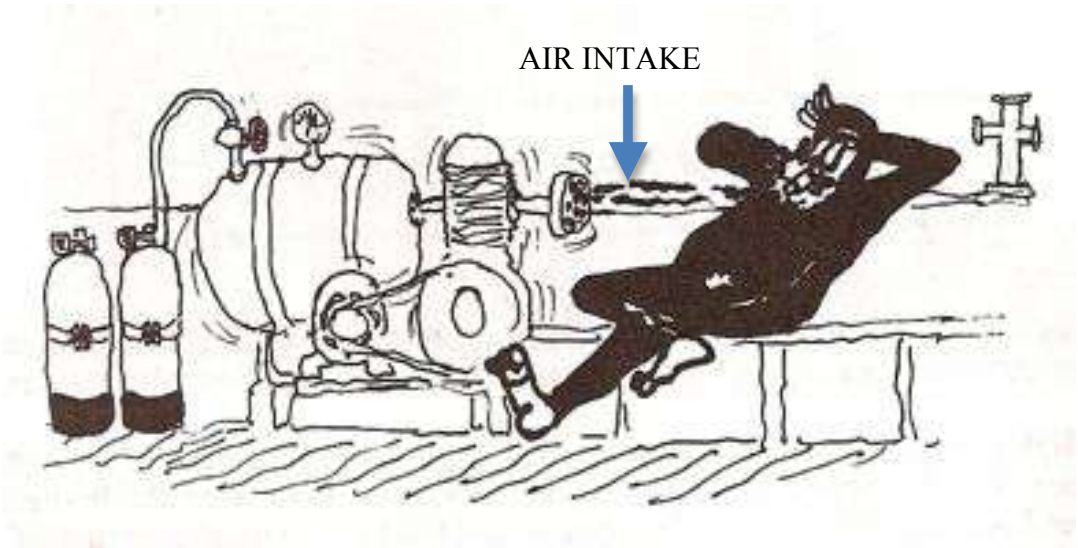

Fig. 23.1

CLINICAL FEATURES

Symptoms are those of progressive hypoxia due to the reduction in the oxygen transport by the blood. They vary with the **carboxyhaemoglobin** content of the blood as shown in the following table:

Concentration of CO in Breathing Gas	% Carboxy-haemoglobin	Effects on a Diver
400 parts per million (ppm)	7.2%	Nil or slight
800 ppm	14.4%	headaches dizziness, nausea breathlessness with exertion
1600 ppm	29.0%	confusion, vomiting, collapse.
3200 ppm	58.0%	paralysis, or loss of consciousness
4000 ppm	72.0%	coma
4500 ppm	87.0%	death

Table 23.1

The effects of CO are cumulative and are related to the concentration breathed and the duration of exposure. A concentration of 400 ppm may produce symptoms in an hour while 1200 ppm will need only 20 minutes. As the carboxyhaemoglobin (HbCO) level falls, following removal of the CO contamination, the clinical state may lag due to CO bound in the tissue, or from enzyme or protein damage. The classical "cherry pink" colour is only seen in the acute and early cases, before respiratory failure develops. Despite the above table, HbCO levels may not correlate directly with mortality or morbidity.

The effects of CO poisoning are greatly increased by increased pressure at **depth,** if the oxygen pressure is kept consistent. A 400 ppm contamination which would not produce clinical effects at atmospheric pressure will be equivalent to 4 × 400 ppm (or 1600 ppm) at 30 metres depth (4 ATA), a concentration sufficient to cause serious toxicity. Because the oxygen partial pressure reduces with ascent, the symptoms of mild CO poisoning may only become serious during or after ascent.

Serious **brain damage** is a frequent complication of significant CO toxicity due to prolonged hypoxia of the brain. (See Case History 24.1)

Because the contaminated compresed air may be given to other divers who employ the same supplier, it is possible that they may also be affected. This may have implications not only on the differential diagnosis of the victim, but for prevention of other casualties.

TREATMENT

The diver should be rapidly **isolated** from the contaminated gas and have **100% O_2** administered by mask. The administer **basic life support** (see Chapter 42) including CPR should be applied where appropriate.

Hyperbaric O_2 (HBO) is the treatment of choice. The high partial pressure of O_2, which occurs in a hyperbaric chamber, will dissolve enough O_2 in the blood plasma to meet the bodies needs without participation of the haemoglobin system. Oxygen is breathed at a partial pressure of 2 ATA or more to sustain life while the CO slowly detaches from the haemoglobin and is breathed out through the lungs, allowing the haemoglobin to resume its normal O_2 transport role.

If hyperbaric O_2 is to be of value it should be instituted as soon as possible, preferably within 6 hours of poisoning. Delay in treatment may result in irreversible and progressive brain damage.

PREVENTION

The major danger to any diver is from carbon monoxide contamination of the compressed air supply. Sources of contamination include:

❑ Direct contamination by CO from gasoline engine exhausts.

This may come either from the compressor motor itself, or from other nearby motors or gas exhaust outlets. The classic case occurs where the compressor air inlet hose is located downwind from the compressor motor exhaust.

❑ Contamination produced by the breakdown of unsuitable lubricants.

The incorrect use of hydrocarbon-based lubricants used to lubricate an air compressor is a common cause, however it may also result from overheating of the compressor. Both carbon and nitrogen oxides can be formed.

❑ The intake of polluted atmospheric air to fill air cylinders.

It is important for suppliers of compressed air to regularly check the quality of the air being compressed, to ensure that this and other pollutants are not included in divers' air supplies. Adequate filtration systems are necessary on all compressors, and these should always be properly maintained.

Chapter 24

BREATHING GAS CONTAMINATION

The supply of uncontaminated breathing gas (air) is of vital importance to the diver because of the magnifying effect on contamination by the partial pressure rise with increasing depth. For example, 5% contamination of gas at atmospheric pressure is equivalent to 20% at 30 metres depth (4 ATA).

Contamination usually arises either from impurities in the air taken into the compressor or from contaminants generated by the compressor itself.

PREPARATION OF COMPRESSED AIR

Atmospheric air is taken into the compressor and is compressed by one of two methods. Most dive shops use a piston and cylinder compressor which raises the pressure of the gas in several stages. A more advanced compressor uses a diaphragm pump similar in principle to that in a refrigerator.

Ideally the compressed air should be treated by passing it through several purifying cartridges (or filters) to remove contaminants. Silica gel is used to remove **water** vapour, activated charcoal removes **oil** and **hydrocarbons**, a molecular sieve removes **water droplets and dust particles** and a catalyst converts **carbon monoxide** to carbon dioxide, which can be absorbed. Less scrupulous air suppliers have been known to substitute women's sanitary pads, instead of filters.

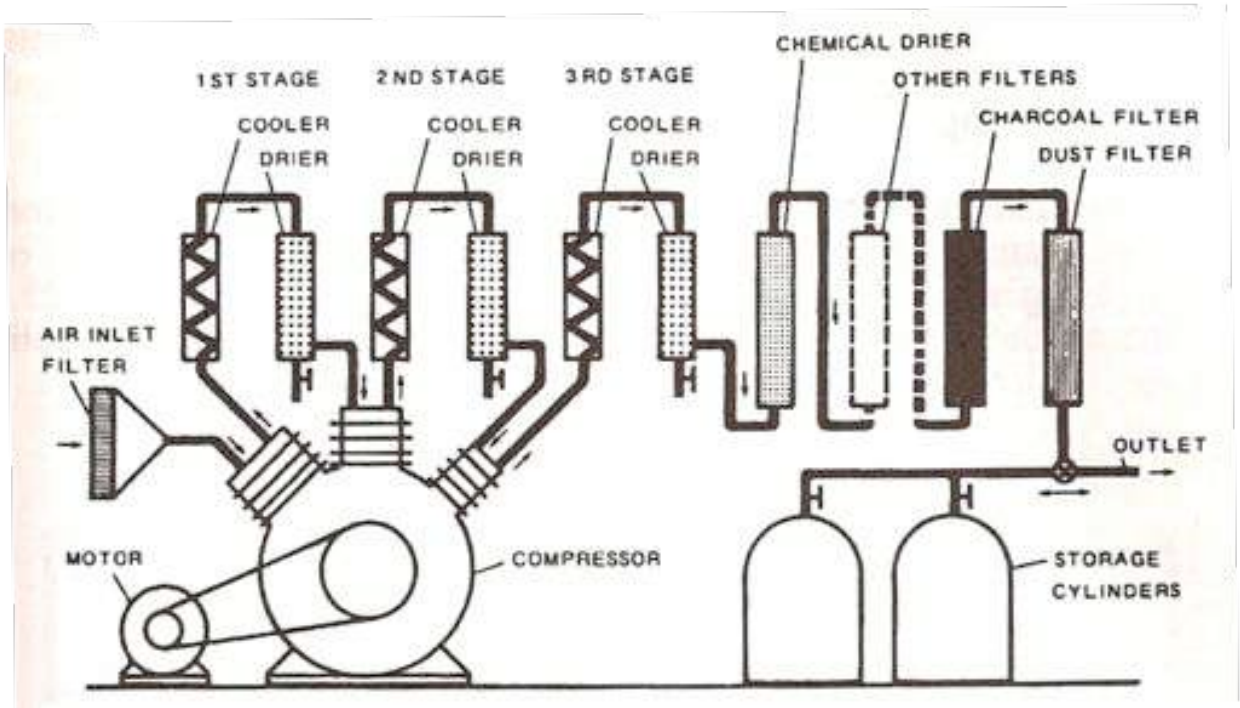

Fig. 24.1
Schematic diagram of a compressor system with filters.

GAS PURITY STANDARDS

Authorities such as the US Navy, NOAA and Standards Australia specify minimum standards of purity for breathing gas.

Case History 24.1 In an area subject to tidal currents an experienced diver planned to dive at slack water. He anchored his boat almost at low tide. The hookah compressor he used was correctly arranged with the air inlet upwind of the exhaust and the dive commenced. After 90 minutes at 10 metres the diver felt dizzy and lost consciousness but was fortunately pulled aboard by his attendant and resuscitated.
Diagnosis — Carbon monoxide poisoning.
Explanation — as the tide turned, so did the boat. This put the compressor **air inlet downwind of the motor exhaust** and carbon monoxide from the exhaust was breathed under pressure by the diver.

Compressors can also generate some lethal contaminants internally. The compressor piston requires lubrication and this is usually achieved by the use of special **oils**. In some circumstances, such as where there is excessive wear of the compressor, high temperatures can be generated and this may decompose the lubricating oil into toxic products such as **oxides of nitrogen** or **carbon monoxide**, which are then pumped into the diver's air tank. Poor maintenance of the compressor can also lead to an **oil** and **hydrocarbon** mist escaping into the air supply.

If the compressor is operated in an unclean environment, **dust** (and chemical vapours) can find its way into the diver's air causing abnormal wear on the moving parts of both the compressor and the regulator.

Fig. 24.2 A modern air compressor

Water vapour must be removed from the air delivered from the compressor or it can condense in the scuba cylinder causing rust, or allow the regulator to freeze up during diving in cold conditions.

Most compressors have a **filtration system** both on the inlet side to prevent the intake of dust, and on the outlet side to filter out oil and water vapour. Their efficiency depends on regular maintenance and the absence of over-loading.

Occasionally contamination comes from the destruction of the filters and lubrication systems. Non-hydrocarbon based lubricants with high "flash points' are preferable. The problems of oil lubrication can be overcome by using a compressor which is lubricated with water or dry Teflon materials. Unfortunately the expense of these is beyond the reach of many air suppliers. Diaphragm pumps also avoid the problem of oil lubrication but are also very expensive.

CLINICAL FEATURES

Contaminated air may have an unusual taste or smell, or alternately, it may appear quite normal.

As many of the divers may obtain their air supply from the same source, it is not uncommon for similar symptoms to be spread amongst the diving group. As only the most serious case may present, the other divers should be interrogated for similar but less severe symptoms.

- **Oxides of nitrogen** cause lung damage, which is likely to cause coughing, wheezing, shortness of breath and/or tightness in the chest.

- **Carbon monoxide** causes headache and unconsciousness – a detailed description can be found in Chapter 23.

- **Oil** can cause nausea, vomiting, chest pain, shortness of breath, coughing and pneumonia.

- **Trichloroethylene and other aliphatic halogens** can cause respiratory and gastrointestinal symptoms.

TREATMENT

If a diver is affected by breathing contaminated air he should be separated from the source, managed according to the **basic life support** principles outlined in Chapter 42, and gas from his scuba cylinder should be sent for analysis to a chemical or gas testing laboratory, as should that from others who used the same air compressors.

PREVENTION

The diver should breathe from his equipment before entering the water and should not use air which has an unusual taste or smell.

As the expertise of compressed air (and other breathing gases) suppliers vary, divers are well advised to obtain air fills only from a reputable supplier. Regular checks by local authorities on the quality of the air are advisable, and in many places are now mandatory. It can be tested by chemical detector tubes that determine the level of each specific contaminant. Drager (a gas and medical equipment company) supply these tubes in a Drager Gas Detection Kit. Many others are available.

Following any diving accident, suspect air can be tested by commercial gas suppliers and State Health authorities.

Case History 24.2

A diving club had for many years been filling their cylinders from an air bank made up of large cylinders, the source of which had been lost in the mists of time. It was decided to return the bank of cylinders to a major industrial gas supplier for testing. The cylinders had their original paint in good condition – black cylinders with a white collar. The gas company tested the cylinders, found them to be sound and refilled them according to the colour code on the cylinders. Unfortunately, this was the standard colour code for pure oxygen and that is what the company filled them with, having no idea that they would be ultimately used to fill scuba tanks.

A member of the dive club took delivery of the cylinders and reinstalled them in the bank. He did not know the significance of the colour coding and assumed that because he was using the cylinders to store air that the company would refill them accordingly. Because they were already full there was no need to fill them from the compressor and the bank was immediately used to fill several sets for a dive the following day. Two divers used tanks from this source on a dive to 20 metres. One abruptly convulsed 10 minutes into the dive and was fortunately rescued by his buddy before he too convulsed.

Some clever detective work performed by the rescuing diver, and the diving physician they consulted, established the cause of the problem as oxygen toxicity. Swift action by the police to round up all the contaminated scuba tanks before they could be used, averted a major disaster. In this case, breathing from the cylinder at the surface before the dive would not have disclosed any detectable difference from air.

Chapter 25

DROWNING SYNDROMES

It is vital for divers to understand the management of near-drowning because it is the final outcome of a large number of diving accidents.

GENERAL

Drowning is defined as the death of an air breathing animal due to immersion in fluid. When patients lose consciousness due to immersion and aspiration, but subsequently recover, the term 'near drowning' is used. When symptoms are not severe enough to classify as near drowned, another term the 'aspiration syndrome', is employed. There is a continuous progression between aspiration, near drowning and drowning cases.

Aspiration syndromes merge with near drownings - often the intensity of the symptoms and the degree of consciousness depending on various circumstances, the activity of the victim and the administration of oxygen.

"Near drowning" cases sometimes die hours or days later, thereby having to be re-classified as secondary or delayed drowning.

Some of the apparently "drowned" victims, because of enthusiastic CPR and capable intensive care, surprisingly recover without serious sequelae.

The incidence of death by drowning appears to be diminishing in the more civilised countries, from 7 per 100,000 in 1970s to around 2 per 100,000 now. Drowning is second only to motor vehicle trauma as a cause of accidental death in Australia and the USA, and is the major cause in some age groups (since the introduction of seat restraints in motor vehicles).

There is an over-representation of young males in most drowning series and there is a predictable age distribution for specific types of drowning. Most swimming pool deaths

occur in the very young, surf deaths mostly in teenagers and young adults, ocean deaths of sailors and fishermen throughout the whole adult range, and bathtub drownings are in either babies, the infirm or homicides.

When a fully conscious human accidentally falls in the water, he usually fights to survive, involving a panic reaction with violent struggling followed by automatic swimming movements. There may be a period of breath holding and swallowing of large gulps of water. Vomiting may occur, followed by gasping and aspiration of water and stomach contents. Blood stained froth develops in the airways and may be coughed up. Finally the patient convulses and then dies from cerebral hypoxia.

Drowning was traditionally associated with a "fight for survival" response but in other circumstances drowning may proceed in a quiet and apparently unemotional manner. In these cases loss of consciousness occurs without any obvious warning, and the underwater swimmer/diver then aspirates and drowns quietly. Examples of *quiet drownings* include:

1. Hyperventilation and hypoxia of ascent in breath-hold diving. This is a common cause of drowning in otherwise fit individuals who are good swimmers. Se Chapter 4.

2. Hypothermia and/or cardiac arrhythmia cases.

3. Drugs and alcohol effects. These increase the incidence of drowning by impairing judgment and reducing the struggle to survive. It is likely that nitrogen narcosis may have a similar effect in divers.

4. Diving equipment problems may produce hypoxia. These include the dilution and ascent hypoxias with rebreathers and carbon monoxide toxicity, interfering with oxygen metabolism. They are all likely to cause loss of consciousness without excess CO_2 accumulation, dyspnoea or distress. See Chapters 6, 20 & 43.

5. The salt water aspiration syndrome of divers. See Chapter 26.

6. Other causes of unconsciousness in divers, leading to drowning, have been described in Chapter 33, e.g. cerebral arterial gas embolism, some marine animal envenomations, and coincidental medical illnesses such as head trauma, epilepsy, insulin-induced hypoglycaemia in diabetics, cerebral haemorrhage etc.

Fresh or salt water entering the alveoli (air sacs of the lung) appears to wash out or damage the surfactant lining them, causing alveoli to collapse and become unavailable for gas exchange. Damage to the walls of the alveoli also causes the capillaries to leak blood and protein into the lungs. This interacts with air and water producing a foam which the victim may cough up in copious amounts. This is called pulmonary oedema.

The **sequence of events** in a near-drowning diving incident often goes as follows:

The degree of **panic** behaviour is variable, and may be reduced by such factors as personality, training, drug intake and nitrogen narcosis. If some air is still available from the regulator, the diver may persist with attempting to breathe from this (even at the cost of aspirating some water), and request assistance. Even if an alternative air supply is made available, **hypoxia** may still develop because of the water aspirated. Coughing and gasping may be voluntarily suppressed until the diver

reaches the surface. If the diver is totally deprived of his air supply for some reason, he initially breath-holds until the "break point" is reached and then takes an involuntary breath.

The resulting inhalation of a bolus of water usually provokes **coughing** and **closure of the larynx** producing involuntary breath-holding followed by unconsciousness. It is unusual for large amounts of water to enter the lungs after the victim loses consciousness as the tongue and loose tissues in the throat tend to close the airway. Instead, there is often swallowing of sea water, which makes the diver susceptible to vomiting and aspiration.

CLASSIFICATION of the Drowning Syndromes.

There are many ways of classifying the drowning syndromes.
The one used in this text relies on the dictionary definition of drowning as death due to submersion. The corollary of this is that near-drowning does not lead to death, but was close to it - and usually implies unconsciousness. Salt water aspiration is exactly what it implies and is less serious.

Others, based on the Alice in Wonderland principle that words mean whatever you want them to mean, have re-defined the classification and allow drowning victims to survive, near-drownings to die (delayed or secondary drownings) and others added a series of subdivisions (drowning without aspiration, drowning with aspiration, near-drowning with or without aspiration, the drowning process, etc.). Some classify them as to the aetiology – fresh or salt water drownings.

Clinicians have classified the drowning syndromes depending on their presentation at hospital emergency care units, and the ultimate prognosis is based on this essentially neurological classification, but with other factors being considered. Sensible and practical – for them.

Life-guards, first aid workers and other paramedics have an analogous, but different, classification based on what should be done at the scene of the recovery of the victim. Sensible and practical – for them.

CLINICAL FEATURES

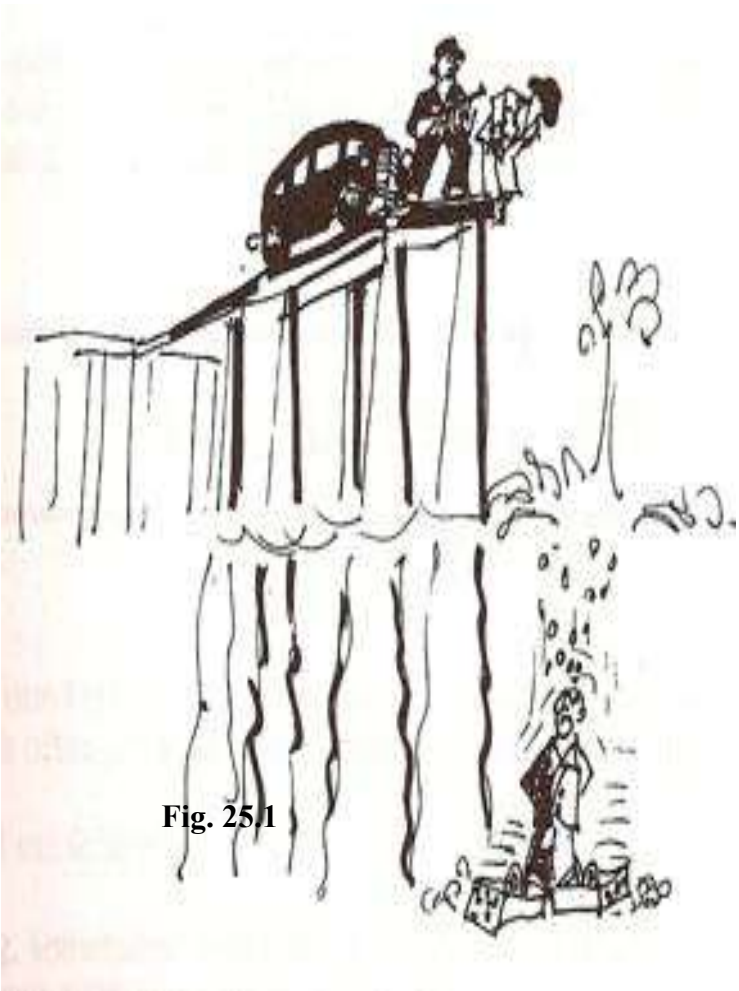

Fig. 25.1

The commonest ultimate cause of death in recreational scuba divers is drowning.
For example, any loss of consciousness or capability when engaging in terrestrial activities is unlikely to cause death. It would do so more frequently if the victim was diving under water.

When first rescued the condition of the near-drowned victim may vary from fully conscious to unconscious, with normal, laboured or absent respiration. The initial effects are on the respiratory system, but later damage is due to hypoxia on the brain.

If the victim is breathing, the stiff lungs cause laboured respiration and it is common for foam, often copious and blood stained, to be coughed up or to exude from the nose or mouth. Vomiting is also common, as is aspiration of stomach contents, either spontaneously or during resuscitation attempts. Cyanosis (bluish coloration of lips, tongue, ears) from hypoxia is frequent. The jaws may be clenched.

TREATMENT

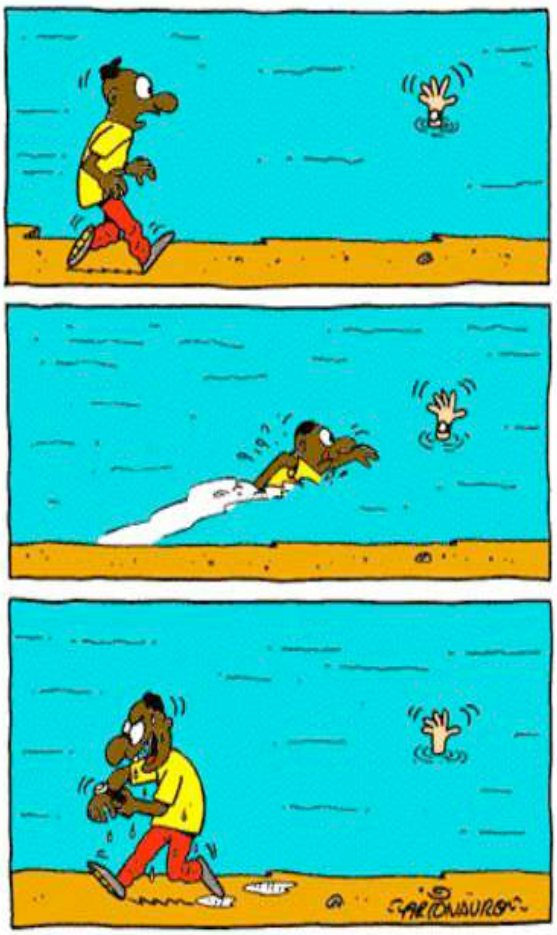

Fig. 25.2

Rescue and recovery from the water is the top priority and time should be spent in the diving course in this training. Success is reliant on buddy behaviour, acquiring positive buoyancy during ascent and on the surface (including ditching of weights) and attracting attention.

Treatment at the scene of an accident will often determine whether the victim lives or dies. The standard of first aid and resuscitation training of the rescuers therefore influences outcome.

The temperature of the water and thus the degree of hypothermia may also be a factor. Poorer results are achieved in warm bath water drownings. Other factors which influence outcomes include: the presence of chlorine and other chemicals and foreign bodies, the aspiration of stomach contents, the subsequent development of pneumonitis, respiratory infection and lung damage, haemolysis, renal failure and coagulopathies. These complications are for the intensive care physicians to cope with.

In exceptional circumstance, near drowned victims have fully recovered after periods of total immersion of over 15-45 minutes (especially so in children in cold waters), so it is worth attempting resuscitation even in apparently hopeless cases.

If the patient is unconscious the basic life support (**BSL**) priciples take precedence and should be followed (see Chapter 42). **Oxygen** in the highest concentration available should be given by mask to offset hypoxia. Masseter spasm ("jaw clenching") is a common feature of hypoxia, and may obstruct artificial respiration.

Near drowned cases are liable to deteriorate many hours after making an apparent recovery, so all near drowned victims should be taken to **hospital** and must remain there for at least 24 hours under observation.

The World Congress on Drowning, in 2002, made the following recommendations, which may also be applicable to divers, where the submersion time is unknown or is known to be less than 15 minutes.

Whenever a non-breathing victim is found in the water, the rescuer should bring the victim's face out of water and extend the neck to open the airway. In either shallow or deep water, if two or more rescuers are present, or a single rescuer is equipped with a floatation device, the victim should then be checked check for breathing. In the absence of spontaneous breathing, rescue breathing should be carried out for approximately one minute.

If breathing is restored, the victim's airway should be kept open during recovery and removal to dry land or boat. Only a brief stop should be made to monitor breathing and restart rescue breathing if necessary.

If breathing is not restored or there is no circulation, the rescuer should recover the victim to shore or boat without further attempts at rescue breathing. Rescuers should not check victim's pulse or start compressions while in the water. Cardiac compression in the water has been shown to be ineffective and pulse checks are both unreliable and slow the rescue process; even if CPR is necessary, this may place the victim in further danger of aspiration of water and needlessly tire the rescuer. After successful resuscitation the victim should be kept under observation for 5 to 10 minutes in case breathing stops. Even trained lifeguards cannot always accomplish in-water resuscitation technique effectively, especially in deep water.

Of course, resuscitation and then observation must be continued until professional medical assistance supervenes, as describer earlier.

In the diving situation it may sometimes be possible to improve on these measures, and it is strongly recommended to all competent divers that they should take advantage of some of the excellent Rescue and Resuscitation courses available through the diver training organizations and/or DAN.

PREVENTION

It is paradoxical that drowning, which causes more than 80 times the number of deaths in recreational divers than either decompression sickness or contaminated air, does not rate more than a paragraph or two in some diving medical texts.

A normally functioning diver, with adequate equipment in a congenial ocean environment, is protected from drowning as he carries his own personal life support system with him - his scuba. Drowning only occurs when there is;

- diver fault (pathology, psychology or technique),
- failure of the equipment to supply air, or
- hazardous environmental influences.

A survey was conducted of 100 recreational scuba deaths from drowning and compared these with near drownings, and it demonstrated that simple measures were available to avoid the fatal drowning cases. They were:

1. Diver health and fitness.
Ensure both medical and physical fitness, so that there is no increased likelihood of physical impairment or loss of consciousness, or difficulty in handling unexpected environmental stresses.

2. Experience.
Ensure adequate experience of the likely dive conditions (become trained and dive under the supervision of a more experience diver, when extending your dive profile).

3. Equipment.
Absence of appropriate equipment is a danger, but not as much as equipment failure and misuse. The latter includes the practice of overweighting the diver, and his over reliance upon the buoyancy compensator.

4. Environment.
Hazardous diving conditions should be avoided, using extreme caution with tidal currents, rough water, poor visibility, enclosed areas and excessive depths.

5. Neutral buoyancy (during the dive).
Ensure neutral buoyancy whilst diving. This implies not being overweighted and not being dependent on the buoyancy compensator.

6. Air supply.
An inadequate supply of air for unexpected demands and emergencies may convert a problematic situation into a fatal one. It also forces the diver to experience surface situations that are worrying and conducive to anxiety, fatigue and salt water aspiration. Equipment failure is not as common a cause of inadequate air supply as diver error -

failure to monitor the contents gauge and/or a reprehensible decision to breathe the tank down to near-reserve pressure.

7. Buddy diving.
Use traditional buddy diving practice - 2 divers swimming together. Solo diving, for the whole or part of the dive, is much more likely to result in an unsatisfactory outcome in the event of diving problems. It is the divers who are committed to the traditional buddy diving practices who are likely to survive the more serious drowning syndromes.

8. Positive buoyancy (post incident)
Positive buoyancy is frequently required if problems develop.
Failure to remove the weight belt during a diving incident continues to be a major omission, and must reflect on training standards. In most situations, unbuckling and then ditching (if necessary) the weight belt is the most reliable course of action once a problem becomes evident.
Buoyancy compensators cause problems in some emergency situations, and not infrequently will fail to provide the buoyancy required. They are of great value in many cases - but are not to be relied on.

9. Buddy communication.
If feasible, inform the buddy prior to ascent. If correct buddy diving practice is being observed, the buddy will automatically accompany the injured or vulnerable diver back to boat or shore.

10. Rescue
Employ the rescue, water retrievals, first aid facilities (including oxygen) and medevac systems which were planned before the dive. See Chapters 5 & 39.

These factors differentiate a drowning fatality from a successful rescue.

Chapter 26

SALT WATER ASPIRATION SYNDROME

This condition was first described in Royal Australian Navy divers in the late 1960's. Some divers were repeatedly presenting for treatment with a brief condition characterised by shortness of breath, sometimes a pale or bluish (cyanosis) skin colour, mild fever accompanied by shivering, malaise, anorexia, and generalised aches and pains. Chest X-rays sometimes showed an appearance similar to a patchy pneumonia and blood gases consistently verified hypoxia.

Close questioning of the divers revealed that nearly all the cases had aspirated a fine mist of seawater from a leaking or flooded demand valve. "Volunteer" experiments confirmed the association between aspiration of sea water and the development of the syndrome.

CLINICAL FEATURES

There is often a delay of half an hour or more between aspiration of the water and the major symptoms. The onset in mild cases is often provoked by exercise, movement or cold exposure. Others may develop after movement, such as arising from bed the next morning.

The diver has some or all the following symptoms:

- initial coughing, sometimes with expectoration, after surfacing
- fever with shivering (induced by cold exposure),
- malaise with anorexia, nausea or vomiting,
- shortness of breath, coughing, cyanosis
- headache and generalised aches and pains.

TREATMENT

The condition is self limiting and resolves without treatment within 2-24 hours. **Rest** and the administration of **100% oxygen** by mask for several hours until the symptoms have abated, is of considerable value. The oxygen not only relieves the hypoxia but produces dramatic resolution of the symptoms of this syndrome.

Because of the nature of the symptoms, it is necessary to distinguish the salt water aspiration syndrome from other serious conditions such as decompression sickness (chokes), pulmonary barotrauma (burst lung), severe infection and pneumonia – which can all present with some or all of the features of this condition.

PREVENTION

The condition can be prevented by avoiding situations which will result in the aspiration of seawater. Buddy breathing from a single regulator can be a fruitful source of the syndrome if the shared demand valve is not adequately cleared of water. Others include a towed search, poor regulator performance, exhaustion of air supply and free ascent practice. Proper maintenance of the demand regulator and its exhaust valves, is important.

Other cases develop on the surface, after divers remove the demand valve to talk – as they are wont to do. Similar symptoms are observed in surfers and victims rescued from the sea (especially by helicopters, which produce a strong down draft causing a sea water spray).

Some divers are especially vulnerable based on hyperactive airways, with a history of hay fever or asthma. In respiratory laboratories, aerosol inhalations of hypertonic saline (sea water) are used to provoke these breathing difficulties and demonstrate susceptibility to the syndrome.

Chapter 27

COLD & HYPOTHERMIA

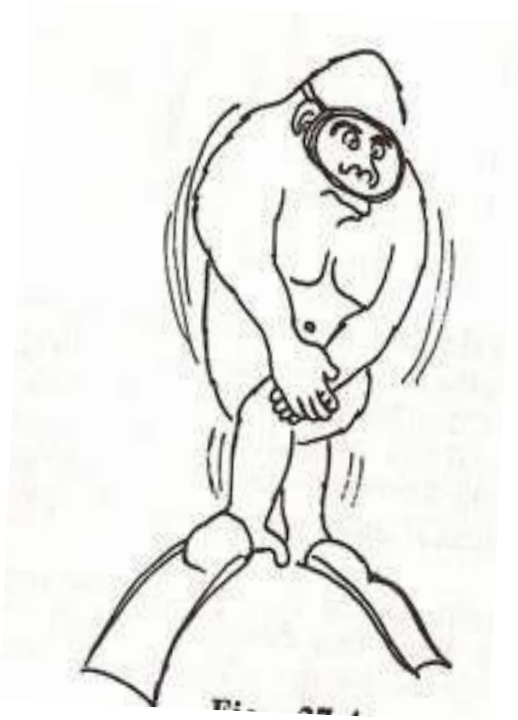

Fig 27.1

A diver is usually immersed in water which is considerably colder than the normal body temperature of 37°C. Unfortunately, water is particularly efficient at removing body heat, having a conduction capacity 25 times that of air and a specific heat (the amount of heat necessary to raise a given volume by a certain temperature) 1000 times that of air.

Without insulation, a diver will lose body heat much faster in water than in air at the same temperature. This can cause **hypothermia**, a harmful drop in body temperature to below 35°C.

The body can reduce temperature loss by generating heat through metabolism, exercise and shivering, and by restricting blood flow to the skin. The rate of heat loss also depends on factors such as the temperature of the water, the thickness of body fat, presence a wetsuit or other insulation, and the posture of the diver.

Recognition of the early clinical features of hypothermia may convince a diver to leave the water before a serious problem arises.

CLINICAL FEATURES

All divers will have experienced the early features of cold — numbness, blueness or pallor of the skin (especially in peripheral areas such as the fingers, toes and earlobes), clumsiness and shivering.

If the body temperature falls by about 2°C, loss of co-ordination and uncontrollable shivering may impair the ability to swim and render the performance of finely coordinated movements (like manipulating equipment and assisting buddies) impossible.

After a body temperature drop of 3–4°C, the diver may become weak, apathetic, confused and helpless. Drowning is a real possibility at this stage. A body temperature less than about 30°C results in unconsciousness. This may be confused with other causes of unconsciousness in divers. Often the diver appears to just lose consciousness without other obvious clinical manifestations.

A victim who is unconscious from severe hypothermia may have a very slow respiratory rate, and a barely detectable pulse, and may appear dead to the inexperienced observer. It is important to not assume the worst in this situation. He may even have fixed dilated pupils and still be resuscitated. Do not presume that he is dead, unless he is warm and dead.

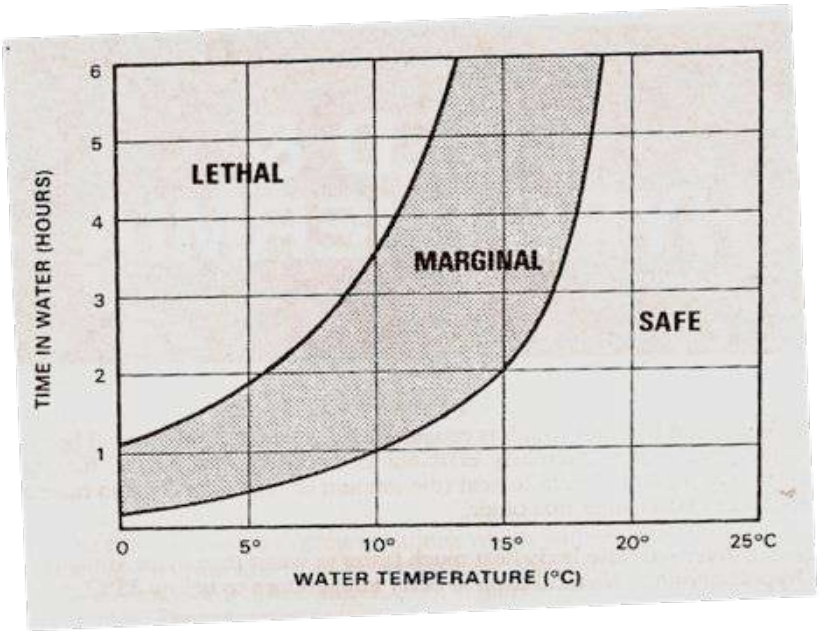

Fig. 27.2

This graph gives an indication of approximate survival times of an uninsulated human in water of various temperatures. These figures are overestimates – a diver would be severely incapacitated well before he reached the limits of survival. It is obvious that survival times of less than one hour can be expected without insulation in water temperatures found in many countries.

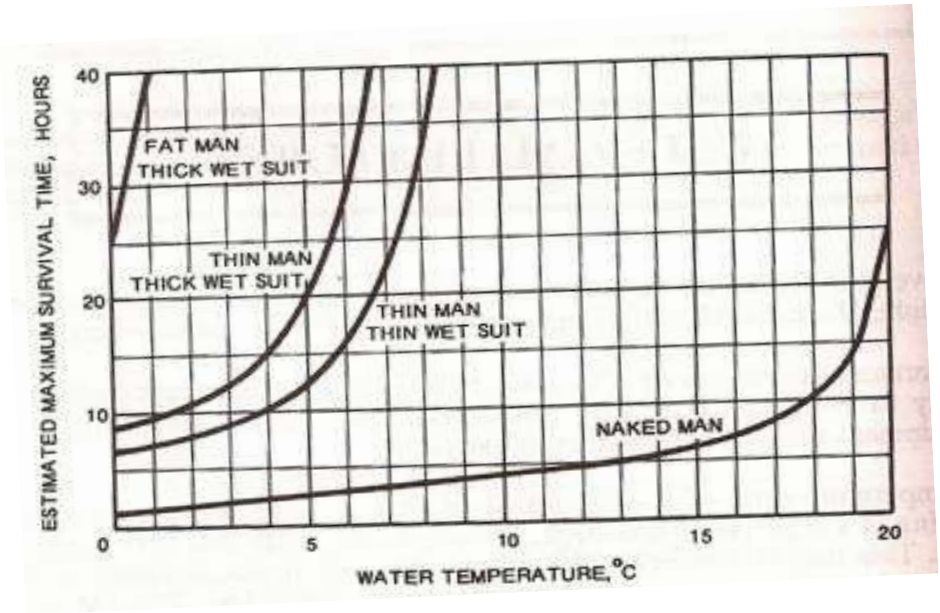

Fig. 27.3
Graph illustrating survival times in varying water temperatures for divers

FIRST AID TREATMENT

If required, the basic life support (**BLS**) first aid management principles take precedence (see Chapter 42). Removal from further danger is followed by assessment and treatment. It is recommended that expired air resuscitation (**EAR**) and external cardiac compression (**ECC**) be performed at half the normal rate in cases of hypothermia because body metabolism is slowed. However, unless the rescuer is confident that hypothermia is the sole cause of the victim's collapse, the usual resuscitation techniques and rates are probably indicated.

The diver must be handled gently. Both active and passive movements are to be avoided, as these tend to trigger serious or lethal cardiac arrhythmias. While the patient is hypothermic, ensure that he remains horizontal, as the vertical position can cause death. Always clear the airway, check for any evidence of heartbeat or respiration, and begin resuscitation as necessary.

The aim of management is to keep the victim alive, while returning the body temperature to normal. The usual methods of treating the diver include **wind-proofing**, **insulation** and **active warming**.

Fig. 27.4

If medical or hospital facilities are available, many other treatments are more effective than the first-aid and warm water immersion regimes, mentioned below. The immersion treatment is probably only indicated for those victims who have sudden or severe hypothermia.

Wind-proofing is essential. Unless the diver is protected from the wind and the wet, he will continue to lose heat. Usually it is best to dry the victim and clothe him, but under some exposed situations it may be necessary to leave his wet suit on and cover it with other materials, to supply insulation. If a wet suit has to be removed, it is preferable to cut it off.

Wet weather gear used alone, without a heat source, may help with insulation but may not be very effective as they do not generate heat and the victim's heat output is very slow. Wrapping in a **blankets**, **plastic** (garbage bag)**, tarpaulin** or even **newspaper**, may also help with insulation by reducing air flow over clothes, wet suit or skin. A reflective **survival blanket** over clothes and normal blankets may aid in wind-proofing and insulation.

Facilities to warm a diver are usually limited at a dive site and improvisation may be required. **Wrapping the victim in blankets with other divers** may be one way of transferring body heat to a mildly hypothermic diver. Warm diver buddies, especially of the opposite sex, may be sought by some unscrupulous divers who only pretend to be hypothermic.

The **engine room of larger vessels** is often warm enough to be of value in the management of hypothermia and engine cooling water may be a source of warm water in an emergency. Treatment can be suspended when the patient's body temperature reaches 37^{0}C, or he starts to sweat.

Although **alcohol** produces a warm inner glow, it actually worsens hypothermia by increasing blood flow to the skin, accelerating heat loss. It should not be given to hypothermic patients. Stimulating drinks such as tea and coffee should also be avoided. Warm water, glucose or electrolyte drinks may be given to fully conscious patients.

Re-warming is most simply achieved by **immersing the victim in a warm bath at a temperature of 37–38**^{0}**C.** A warm shower is a less efficient alternative and certainly not with the patient standing. A pleasantly warm bath or shower is approximately the right temperature. Warm packs or hot water bottles over the axilla, groin and abdomen may help – but avoid scalding.

It is possible that oxygen administration may be of value, especially if warmed or if used in a re-breathing system. The latter will reduce heat loss by re-breathing warm humidified gas.

Massage, alcohol or stimulant drinks (coffee), heat packs direct onto the skin and exposure to intense sources of heat (such as radiators), are all best avoided.

PREVENTION

Alcohol and other drugs may predispose to hypothermia by dilating peripheral blood vessels and losing heat by conduction into the water.

Diving in cold water is the commonest cause, but even in tropical waters loss of body temperature during a dive is likely if the diver is not effectively insulated.

The most popular and convenient insulator is the **wet suit** (see Chapter 5). Air bubbles enclosed in synthetic rubber provide an insulating barrier between the diver and the water without the need for the suit to be waterproof – hence the term "wet suit". They are available in various thicknesses depending on the expected water temperature. Wet suits have the disadvantage of compression of the air cells at depth, which reduces their insulation and causes inconvenient changes in buoyancy.

This problem is reduced in professional diving operations by the use of a "**dry suit**" which uses air as the insulating material. Other variations include electrical, chemical or hot water warming procedures, or even an inflatable air pocket enclosed in a wetsuit.

When immersed and in a survival situation, heat loss in an uninsulated person can be minimised by floating in the H.E.L.P position, a curled-up posture ("foetal" position) with the knees near the chest and the arms by the side, so covering the body areas which lose heat the most (axilla and groin). This can obviously be done only if the diver has a flotation aid. Huddling together with other survivors may be of value. Restriction of movement will also minimise heat loss.

To reduce heat loss, it is best not to swim more than a short distance, as although swimming generates some metabolic heat, this is more than offset by heat lost into the water during movement.

Divers should abort dives once they start feeling cold, and should ensure adequate time on the surface, in a protected and warm environment, before returning to dive. Hours are needed to regain the deep core body temperature. Sweating is a good sign that hypothermia no longer is a problem.

OTHER REACTIONS TO COLD

There are a variety of other problems which can eventuate from cold exposure. These include the following:

❑ Reflex Responses.

Exposure to cold can produce a number of adverse effects on the function of the heart and lungs. These reactions can kill a swimmer/diver on entering the water and can cause a problem before he develops hypothermia It can cause arrhythmias and interference with the coronary artery blood flow – with possible angina or myocardial infarction. It can also have reflex effects on the lung function, causing the diver to inhale excessively, causing aspiration of sea water, resistance to breathing, electrolyte changes, etc. Another problem is that a diver may respond to the cold stimulus with intense contraction of the blood vessels, causing a rise in blood pressure and a “stroke”.

Usually these changes are not noticed by the average diver and are of more interest to the diving physician investigating unusual diving accidents. (See Chapter 35)

❑ Muscular cramps

Cold exposure increases the likelihood of muscular cramps, especially in the legs (feet, calves), aggravated by finning.

❑ Cold Urticaria.

Some rare divers are particularly sensitive to the inhalation of cold air (due to the drop in pressure across the first and second stages of the regulator). This can cause difficulty in breathing as the cold air irritates the respiratory passages, producing an asthma-like syndrome. Cold water exposure on the skin can also produce an allergy-like effect, with a skin reaction similar to hives (urticaria). It can lead to generalised effects on blood vessels and blood pressure (shock), and it is important that the diver ceases to expose himself to cold, and obtains medical advice.

❑ Sinus and Ear Pain.

Exposure to cold seems to produce a reflex pain in susceptible divers, similar to the "ice cream headache". The site of pain may be related to an over-sensitive area of the skin being stimulated by the cold. Similar symptoms may develop during skiing or surface swimming in cold water.

If the external ear is affected, then ear plugs can be worn for surface swimming and fenestrated ear plugs for diving. Alternately a hood can be used with a small aperture over the ears. This will reduce the cold stimulus effect by retaining the water warmed to body temperature. Others instill oil or wax to reduce this effect.

Occasionally this disorder has been confused with barotrauma of descent, as it tends to occur within a few minutes of immersion, while the diver is descending.

Chapter 28

INFECTIONS

There are a variety of both exotic and mundane infections to which divers are exposed. Some are terrestrial and are the same as experienced by non-divers. Others are caused by specific marine organisms and require special methods of identification.

ABRASIONS AND INFECTED CUTS

Divers are frequently subjected to minor injuries including cuts and abrasions. These injuries are more prone to infection than those encountered in terrestrial pursuits because of the unusual bacteria encountered in the aquatic environment and because cuts and abrasions on divers frequently remain moist for long periods of time.

Cuts and abrasions which are not due to coral or other marine life do not require aggressive cleaning unless they are obviously contaminated. Bleach, antiseptic or antibiotic cream or powder should be applied as soon as possible. When out of the water they should be kept dry and loosely covered to prevent further contamination.

Coral Cuts

Coral often causes minor cuts and abrasions in unprotected divers in tropical waters. These cuts are particularly prone to infection, probably because of the large numbers of marine bacteria on coral and the retention of coral particles and slime in the wound.

They frequently become infected within hours. Even minor cuts or abrasions can become red, swollen, tender and painful. Later there may be a discharge of pus from the area.

A severe infection may spread to the lymphatics and blood stream, with fever, chills and tender swollen lymph glands in the groin or armpit, depending on the site of the injury.

❑ Treatment.

All coral cuts should be washed with bleach or soapy water as soon as possible and the surface of the cut or abrasion should be thoroughly cleaned. This removes foreign material that may be the source of inflammation. All cuts should then have local antibiotic powder, cream or ointment applied every 6 hours until healed. The senior (elderly) author, who has a tendency towards cowardice, relies more on the antibiotics than vigorous cleansing. Suitable topical antibiotics include neomycin or bacitracin.

Early attention to every coral cut in this way will usually prevent serious infections. If treatment is delayed, or if systemic effects occur, oral broad-spectrum antibiotics may be needed. The development of a chronic inflammation may cause severe itching over the next few weeks, but this usually responds to local steroid (cortisone) ointments.

❑ Prevention.

It is wise when diving on coral reefs to always wear protective clothing or a wet suit, gloves and booties. Modern lightweight "lycra suits" afford some protection and may be worn in very warm tropical waters. These provide no flotation or thermal insulation properties, and diving must be adjusted for this.

EAR INFECTIONS

Otitis Externa
(Swimmer's or Tropical Ear")

This outer ear infection is one of the most common and troublesome problems in divers. It is especially likely to occur in **hot humid conditions** (e.g. tropical climates, standard diving dress, compression chambers) or when **water is retained** within the ear after immersion, especially if with **contaminated water**. Small bony outgrowths (**exostoses**) are commonly found in the ears of swimmers and divers, and these may be large enough to cause retention of water, wax, debris and organisms. **Local injury** induced by scratching the ear canals (with a match or hair pin), or by clumsy attempts to remove **wax** (often using cotton buds) frequently precipitates the infection. Sometimes an underlying **skin disorder** is present such as eczema or dandruff. Many bacterial organisms have been incriminated, as well as fungi.

❑ Clinical features.

Mild infection causes **itching** of the ear which encourages the diver to scratch the ear canal, further breaking down the protective barrier and aggravating the infection. This has prompted the good advice that “nothing smaller than your elbow” should be inserted into the ear canal.

Serious infection may appear as a local boil in the ear canal, or as a diffuse inflammation with narrowing of the canal and an offensive smelling discharge. **Pain** with movement of the jaws or pulling on the ear is common. Occasionally a **mild hearing loss** or **dull feeling** in the ear may be noticed, and **dizziness** during diving is a possibility if one canal is completely blocked.

❑ Treatment.

The condition may be difficult to cure and treatment should be supervised by a doctor. Mild cases may only require careful cleansing of the ear canal followed by local (topical) **antibiotic + steroid** ear drops three times per day. More severe cases will need pain killing tablets such as paracetamol — two tablets four hourly as required, along with packing of the canal with special antibiotic + steroid ointments e.g. "Kenacomb otic". Oral antibiotics may be required in severe cases. Diving, along with further exposure to any water, should cease until the infection resolves.

❑ Prevention.

This can be achieved by the use of olive oil drops in the ears prior to diving, or the application of a few drops of a solution of 5% acetic acid in 85% isopropyl alcohol in each ear after a dive to ensure adequate drying. Commercial solutions include "AquaEar" and "Otic Domeboro". Scratching the ear canal with matches, hair pins, cotton buds and the like, although tempting, should be avoided.

Otitis Media
(Middle ear infection")

Middle ear infection is not very common in adult divers, but may occur after **middle ear barotrauma** or following upper respiratory tract infections (**URTIs)** or allergies. It may also follow an uneventful dive. Most infective organisms enter the middle ear cavity via the Eustachian tubes, which lead from the throat to the middle ear cavity, during middle ear equalisation. Occasionally a **perforation** in the ear drum will allow direct entry of contaminated water.

❑ Clinical features.

Clinically there may be a constant and/or stabbing **pain in the ear, fever**, **ringing** noises (tinnitus) and often a slight **hearing loss**. In this case the ear will not usually be tender to touch.

❑ Treatment.

Treatment is urgent and will include oral broad spectrum antibiotics, pain relieving tablets such as paracetamol and decongestants (such as pseudoephedrine). Ear drops are not advised. No diving or flying in aircraft should occur until resolution — usually 5-7 days.

GENERAL INFECTIONS

There are many diseases that can be caught while diving and swimming. Some are generalised and serious, and the organisms are especially adapted to water environments. Others, such as infectious hepatitis, poliomyelitis, typhoid, cholera, gastro-enteritis, etc., are common diseases and are found especially in contaminated water. Diving in contaminated water needs special clothing and protection as well as post-dive cleansing techniques.

❑ Vibrio and other Aquatic Organisms

Generalised manifestations such as fever, chills, septicaemia, shock etc. may be accompanied by respiratory symptoms, muscular pains, headaches, coma and death.

❑ Key West Scuba Divers Disease

This condition is not restricted to the city where it was described, in Key West. It is due to contamination of regulators (especially older ones with twin hoses) by bacteria and tends to occur in multiple divers– such as in diving courses where regulators are shared. It can also develop from breathing bags. It is similar to influenza with respiratory symptoms, but usually clears after 3 days.

❑ Naegleria (Amoebic Meningitis).

This lethal condition is encountered by divers or swimmers bathing in fresh water lakes, streams, hot springs, spas or hot tubs. It is caused by a microscopic amoeba which usually enters water by **faecal** contamination. It may survive in **warm fresh water** (not in sea water). The amoeba enters the body through the nose from where it burrows through the olfactory nerve to enter the brain. After an incubation period of about a week it causes meningitis and encephalitis, which is difficult to treat and is usually fatal.

The condition is usually manifest by a progressively worsening headache, fever, vomiting, discomfort on looking at bright light, neck stiffness, confusion and finally, coma. Death usually follows after 5-7 days.

There is very little that can be done to treat this dangerous condition apart from intensive nursing care in a major hospital and aggressive intravenous therapy with several antibiotics – none of which are very effective.

Because the organism enters through the nose, infection can be prevented by not immersing the head in fresh water, which is at risk of contamination. Such waters should be avoided if possible, however if diving is essential in these areas (police underwater searches, mining or drainage assessments etc.) then only diving equipment incorporating helmets which totally enclose the head and face should be used, and these rinsed off thoroughly prior to undressing after the dive. Heavy chlorination will kill the organism as will cold temperatures and seawater, eventually.

SINUSITIS

Because air passes into the sinuses during descent (see Chapters 2 and 10), if the diver has an **upper respiratory tract infection** and goes diving with this, then organisms will be transferred to the sinuses as he equalises pressures. Because of the overwhelming infection that is then produced, it is common to develop symptoms within hours or days of the dive exposure.

As a general rule, the more descents carried out, the greater the ammount of infective material which passes into the sinuses. Also, if there is any sign of **sinus barotrauma** (especially on descent) then there is blood and fluid in the sinus at body temperature, which makes an ideal medium for the growth of organisms.

❑ Clinical features.

With sinusitis there is not only a feeling of **fullness over the area of the sinus** (usually maxillary, frontal, ethmoid, sphenoid or mastoid), but there is **pain** which is likely to increase in severity. If there is any significant obstruction of the sinus ostium, then pressure develops within the sinus as infection flares. There may be severe systemic signs – similar to that of an abscess, thus the diver may be **feverish**, feel ill and may look sick.

❑ Treatment.

This usually involves oral broad spectrum antibiotics, pain relief (paracetamol) and decongestants (pseudoephedrine). Sometimes a fluid level can be seen on scans or X–Ray or MRI and rarely, surgical drainage is necessary.

Because infections tend to produce scarring, sinusitis must be avoided as much as possible by divers – otherwise the openings of the sinuses can become scarred and narrowed. This means that the diver is much more likely to develop sinus barotrauma in the future – thereby limiting his diving career.

❑ Prevention.

Avoid sinus barotrauma (see Chapter 10). The rapid and effective treatment of infections that do develop in the sinuses will be of some preventative value. Of more importance is the avoidance of diving during times in which there is any inflammatory disease of the upper respiratory tract (nose, throat), such as hay fever, rhinitis or upper respiratory tract infection.

SWIMMER'S ITCH

Swimmer's itch is a localised skin infection caused by a bird parasite (Schistosome cercaria) which can be encountered by persons swimming or wading in lakes or lagoons frequented by water birds. The parasite, which is present in the water, burrows through the skin and then dies, causing an inflammatory reaction under the skin. It causes multiple small, raised, red itchy lumps, which may last for a week or so.

The lesions usually resolve without treatment. Occasionally, more severe reactions may follow in individuals who are allergic to the parasites and may require medical attention by way of oral antihistamines and even topical or oral steroids (cortisone).

Fig 28.1

SWIMMING POOL GRANULOMA

Also called **Swimmer's Elbow**, this infection is due to an organism (marine *vibrios*) entering the skin via an underwater abrasion from a swimming pool, ship's hull etc. **Red swellings** covered with fine scales may develop 3–4 weeks after injury over bony prominences such as the elbows or knees. Thick pus may be found if the swelling is incised and spontaneous resolution may take up to a year or more. **Diagnosis** may only be confirmed by microscopic examination of a piece of the ulcer or lesion, and culture of the organism involved.

TINEA PEDIS – "TINEA" (or "ATHLETE'S FOOT")

This is a common fungal infection which affects the feet of divers and swimmers exposed to repeated wet and warm conditions, such as shower rooms. It causes itchy, scaly or raw areas between the toes and on the feet.

Many divers suffer from this infection, and are the source of cross-infection to others. The fungus can be found in many areas and makes the condition difficult to prevent because of repeated exposures.

Fortunately it responds readily to modern topical anti-fungal agents such as imidazole derivatives, (tolnaftate or undecylenic acid). The solution or cream should be applied twice daily and continued for two weeks after the condition appears to be cured. Attempts should be made to keep the feet as dry as possible, and drying with tissues between the toes after bathing or swimming is helpful. A light application of an anti-tinea powder (e.g. econazole dusting powder) daily may also be beneficial in preventing recurrence. In severe or resistant cases, oral anti-fungal medication such as ketoconazole or griseofulvin may even be necessary. Towels and footwear should not be shared.

PITYRIASIS VERSICOLOR (or "TINEA VERSICOLOR")

This mild fungus infection of the skin may either cause itching or no symptoms at all. With exposure to the sun however, a diver will notice coin-sized spotty areas which do not tan on the chest, back and arms. A fine scales on the skin will be seen on close inspection. It is best treated with topical anti-fungal lotions or creams such as clotrimazole or econazole applied twice daily. An alternative is 20% sodium thiosulphate (or photographer's 'Hypo' solution – but this stains clothing).

HIV/AIDS

More than 25 million people have died of AIDS, and there are 35 million people living with HIV/AIDS. Half are under the age of 25 years. The potential duration of the disease may be 1-2 decades, or longer. During this time, inevitably some will experiment with scuba diving. The risks to the diver or his companions are not known with certainty. The following is a general discussion based on the current information.

The virus must gain entry to the bloodstream to produce infection and is generally unable to enter through intact body surfaces such as skin or mouth lining. It can gain entry through small breaks in tissues such as cuts, abrasions or ulcers. Tissue fluid oozing from breaks in the skin or mouth lining can contain viruses in infected individuals. The virus is sometimes present in blood (viraemia) and saliva, making transmission a possibility (albeit probably only slight) in some situations encountered during diving. The viraemia is also the danger period to the patient, as this is the time when the virus can attack the central nervous system, especially the brain

❑ Sexual.

This is the most common mode of transmission. It can be eliminated by celibacy and reduced by using condoms ("safe sex" practices), avoidance of both promiscuity and anal intercourse. A wet suit should prove an effective barrier if worn at all times.

❑ Blood.

Blood from infected people can transmit the virus to others. This usually follows the sharing of needles or transfusions of infected blood, but there have been a few reports of transmission by infected blood splashing onto the skin — usually skin which has been broken. In diving, infection in this way could theoretically follow the first-aid treatment of an injured diver after trauma, cuts or marine animal injury. The sharing of mouthpieces and regulators, such as during buddy breathing, and which have not been adequately cleaned after each individual use, could pose a potential risk of infection through oral abrasions.

❑ Resuscitation.

Expired air resuscitation often requires mouth-to-mouth or mouth-to-nose contact. There is a theoretical risk of transmission of the virus during resuscitation, especially if either the victim or rescuer has ulcers or bleeding in the mouth. The risk from saliva is probably small but it would be wise, if a known AIDS carrier was involved, to use a mouth-to-mask or resuscitation tube technique. If the rescuer is trained in its use, an O_2 or air resuscitator bag would be even better. See Chapter 40.

❑ Sharing equipment.

It is common for equipment to be shared in diving schools. Since the virus is known to be present in the saliva of infected people, there is a theoretical risk of transmission of the disease from the sharing of demand valves (including buddy breathing practice) and snorkels. The risk is probably slight and there have not been any cases of such transmission yet reported.

Until the risk is excluded it would seem wise to disinfect shared equipment between uses by soaking in a solution known to be lethal to the virus. Check with the manufacturer beforehand to ensure that the chosen solution will not damage the equipment.

❑ The HIV positive or AIDS infected diver.

These individuals could pose a risk to their fellow divers in the situations outlined above. It would be considerate for them to take care to avoid situations which might bring their blood or other body fluids into contact with others. Breathing equipment should not be shared.

HIV positive cases (those with the virus infection but no obvious symptoms) have recently been shown to have neuropsychological abnormalities not obvious to themselves, but which could be detrimental to the normal intellectual functioning and judgment needed for scuba diving.

Infected divers may be exposing themselves to added risks by diving. Depression of the immune system makes them more susceptible to infection from coral cuts and abrasions, from exotic marine bacteria, and possibly to infections acquired from shared breathing equipment. It is shown by some researchers that hyperbaric environments and hyperbaric oxygen (as occurs with diving and diving treatment respectively) may reduce the integrity of the blood brain barrier and allow the extension of the virus into the brain, causing the dreaded neurological AIDS. The influence of some of the other physiological effects of increased pressure on AIDS infected divers is unknown.

❑ Prevention.

Both divers and non-divers should avoid contamination of the skin by other people's blood where possible, and use disposable plastic or latex gloves whenever possible. If it is unavoidable, wash the blood off as soon as practicable with soap or antiseptic solutions.

If medical attention is sought in **underdeveloped countries**, try to ensure that only single-dose vials, new or disposable syringes (i.e. totally unused) and needles are used, and that re-used instruments have been properly sterilised. Transfusions of blood and blood plasma in some of these countries carries a significant risk of HIV/AIDS or hepatitis infection. Artificial blood expanding solutions such as polygeline ("Haemaccel") should be used whenever available in such countries.

HEPATITIS

This is a highly contagious viral condition which infects and damages the liver. There are three or more variants of the virus but hepatitis-B probably poses the greatest potential threat to divers.

Hepatitis-B virus can cause a fatal infection in up to 10% of cases and fatal liver cancer or cirrhosis may develop after many years in apparent survivors. It is usually transmitted by infected blood.

There is potentially a small risk of infection in divers by sharing breathing equipment (as mentioned previously for HIV/AIDS), but the greatest risk comes from contamination of the skin by the blood of an infected person. In this regard it is similar to, but far more infectious than, HIV/AIDS. The virus is also far more "hardy" and can remain infectious for some time. Old dried blood in syringe needles found lying about can still infect anyone 'pricked' several weeks after being used by a carrier or actively infected person.

Prevention is along similar lines to HIV/AIDS.

INFECTIONS IN HYPERBARIC ENVIRONMENTS

Underwater habitats and compression chambers are humid areas which have a high concentration of oxygen.

This favours the growth of certain types of organisms, both in the chamber and on the skin of inhabitants. Outer ear infections (otitis externa), described earlier, are particularly common in underwater habitats and pressure chambers because of these environmental conditions, and divers occupying these environments are frequently given prophylactic ear drops to prevent these infections.

Should severe infections occur, treatment may prove difficult within such environments. These infections include sinusitis, bronchitis, pneumonias and skin infections. An acute attack of appendicitis occurring in a diver whilst decompressing from a saturation dive may even require surgery within the pressure chamber. The administration of general anaesthesia, along with the sterility necessary for such surgery, render this normally simple procedure much more difficult.

Chapter 29

DANGEROUS MARINE ANIMALS

There are many marine animals which are dangerous to eat, to be eaten by, or to touch. The diver who is content to observe or photograph the creatures of this undersea environment will rarely have his safety threatened by them. Of necessity, this chapter is an oversimplification, with many significant omissions. The photographs are copied from *Dangerous Marine Creatures*, by Carl Edmonds.

SHARKS

Although encounters with sharks are commonplace in diving, shark attacks on skin and scuba divers are not common. Many of the attacks recorded have been associated with spearfishing or shell harvesting, situations in which vibrations and chemicals given off by the wounded marine animal are likely to attract sharks.

In a large proportion of attacks on divers the victim was unaware of the presence of the shark until he was actually bitten. Several behaviour patterns preceding shark attacks have been documented. In some cases the **shark circles** the victim and occasionally bumps him (presumably to gain some sensory information about the nature of this unfamiliar but potential food source), before attacking.

In many tropical species, **sharks may exhibit a threat display (agonistic)**, apparently in response to a territorial invasion by the diver. This is characterised by the shark swimming with an irregular jerking motion, accompanied by an arched back, head up and pectoral fins

pointed downwards. This type of behaviour is the signal for the diver who wishes to experience old age, to depart the area.

The Great White shark has a "**bit and spit**" **technique** in which a single sudden powerful attack is made, with the shark then retreating until the victim (seal, dolphin, diver) haemorrhages in the water and loses consciousness. The shark can then feed without fear of damage from a counter attack.

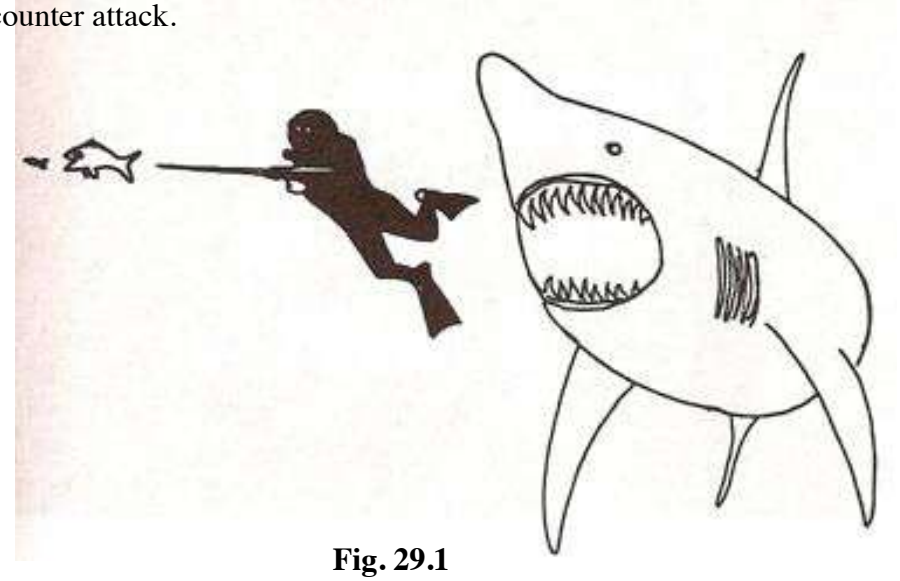

Fig. 29.1

Clinical Features

The seriousness of the injury depends on the size of the shark and the ferocity of the attack. Sharks larger than 2 metres in length have extremely powerful jaws equipped with razor sharp teeth which are easily capable of severing limbs or biting large pieces out of the torso. In spite of this, there have been many instances of divers surviving bites from sharks in excess of 4 metres in length. In some of these, the divers sustained severe lacerations from the puncture wounds of the teeth but no further injury. A shark of this size could easily bite a diver in two, so it appears that in some cases the shark will maul a victim and then not persevere, perhaps due to distaste for wet suit material or other items of the divers paraphernalia. Some divers may be as distasteful to sharks as they are to non-divers.

The blood loss from the massive lacerations accompanying shark attack is severe and immediate. Major blood vessels are frequently torn and generalised bleeding issues from the tissue laceration. Blood loss is often torrential and pulsates from severed arteries.

The victim will display clinical features of severe blood loss — pale clammy skin, a rapid weak pulse, low blood pressure and rapid respiration. Fatality occurs in 25% of cases.

Treatment

The principles of successful management of shark attack victims were first described by Australian and South African authorities following their combined experiences. They are:

❑ Stop the blood loss.

This must be done by rescuers **at the site** of the attack. Bleeding which is oozing or welling up from a wound can be stopped by applying a cloth pad (preferably but not necessarily clean) to the wound and pressing firmly with the hand or applying a tight bandage. Spurting arterial bleeders up to about 3 mm in size can also be stopped by a pressure bandage and pad. Larger arterial bleeders can be stopped by the application of pressure by a finger or thumb. Bleeding from major blood vessels (the size of a finger) can be stopped by pinching the end of the vessel between finger and thumb, or a tourniquet if a limb is involved. Tourniquets have to be released every 10–20 minutes to let blood return to normal tissues.

It is important that pads, used to stop bleeding, have pressure applied to them to force the blood vessels closed. It can be disastrous when rescuers merely cover bleeding areas with a dressing, without any pressure application. This soaks up and conceals the blood loss, without stopping it. Any clean material such as toweling, clothing or handkerchiefs are satisfactory in the first-aid situation.

❑ Resuscitate the victim at the site of the attack.

If the patient is unconscious the basic life support (**BSL**) priciples take precedence and should be followed (see Chapter 42).

I**mmobilisation** is advised. Once the victim is in a place of safety, (boat or shore) it is vital that he not be moved further. Bundling a victim into the back of vehicle for a bumpy ride to hospital has resulted in death of the victim on many occasions.

The victim should be kept lying horizontal at the rescue site and resuscitation equipment and personnel brought to him.

Resuscitation involves replacing the patient's blood loss by the **intravenous infusion** of blood or blood substitutes such as plasma, saline or other intravenous fluids. It is not safe to move the victim until a satisfactory circulating volume has been established. Evidence for this is a relatively normal pulse (rate less than 100) and blood pressure.

This management principle is sometimes difficult to accept by rescuers who understandably wish to dispatch the victim to hospital (anywhere!) as soon as possible. However, once the victim reaches there, exactly the same management as should have taken place at the shark attack site will be needed. i.e. arrest of the blood loss accompanied by the administration of intravenous fluids.

Major hospitals in shark attack prone areas have a shark attack protocol along the lines mentioned above. Equipment may be available for immediate transportation to a shark attack site. Shark attack is so rare, however, that practice at implementing this protocol is sometimes neglected.

In spite of the severity of the injuries, it is common for the patient not to experience significant pain for some time after the attack. This phenomenon is frequently seen in other forms of severe injury such as motor vehicle and war injuries. If the patient is suffering significant pain or shock, the rescuing medical team will administer morphine in an appropriate dose.

Nothing should be given by mouth to the victim, as an anaesthetic may be required.

Prevention

Since vibrations and chemicals given off by speared fish and other forms of marine life commonly attract sharks, the avoidance of fishing should lessen the risk to the diver. The carrying of speared fish or shell fish near the diver's body underwater invites a close inspection by an interested shark.

The well publicised practice of diving with a buddy should, on statistical grounds alone, reduce the likelihood of a shark attack on oneself by at least 50%.

Case Report 29.1 A young swimmer was attacked by a shark which amputated his leg above the knee. He was pulled from the water by his companions and the bleeding stump was wrapped in a blanket. He was noted to be pale and clammy with a weak thready pulse and was semi-conscious. He was placed in the back of a car and rushed to the nearest hospital which was over 20 kilometers away.

A subsequent newspaper report read: "shark attack victim died while being rushed to hospital".
It should have read: "shark attack victim died **because** he was rushed to hospital".

Swimmers are protected by swimming in enclosed or meshed areas. They should not swim where shark attacks have occurred, in estuaries and river mouths, or where fish or meat is ditched (fish markets, abattoirs etc.). It is safer to swim with groups of people and to avoid swimming at dusk (feeding time for sharks) or in areas of low visibility. Urine and blood are claimed to attract sharks and thus should not be released into surrounding water. Women who are menstruating, produce haemolysed blood which is not an attraction to sharks.

Divers are given the same advice, but also to avoid deep channels and drop-offs. If diving with sharks, carry something to fend them off (shark billy). Avoid sites where shark feeding is undertaken (a stupid act). A chain mail suit gives good protection, but it very heavy and thus dangerous for recreational divers. Ultrasonic, electrical, chemical and bubble deterrents are probably not effective against dangerous animals, but are enthusiastically marketed.

BOX JELLYFISH OR SEA WASP

This deadly stinging creature is found in the tropical waters of the Indian, Pacific and Atlantic oceans during certain seasons. The season for North Australia is October to March, but may be all year-long nearer the equator. They are rare in the temperate regions. The animal is an active swimmer which may be found even in very shallow water around beaches.

Its numerous tentacles may trail for up to 3 metres behind the body, which grows to 20 cm along each side of the cube. The tentacles cling to the victim's skin and contain many thousands of microscopic stinging cells (nematocysts) which can inject venom. The innumerable tiny doses of venom injected combine to form a large injection of toxin into the victim. The amount of venom injected depends on the length of tentacle in contact with the victim, and the area stung, as well as the thickness of skin.

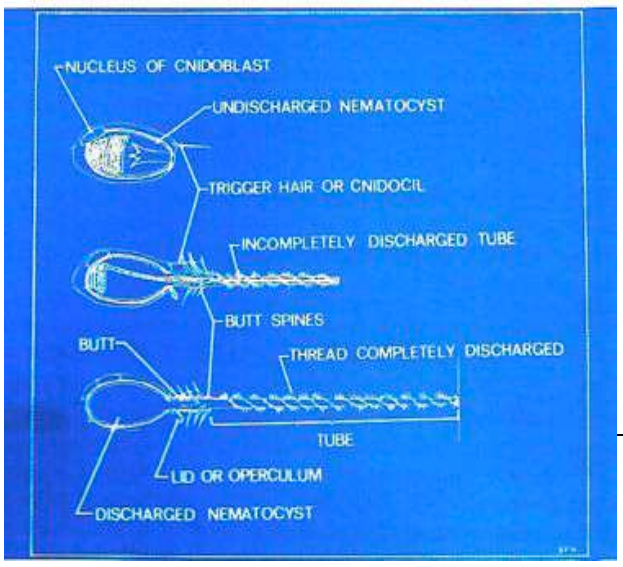

Fig. 29.2 Nematocysts from jellyfish

The venom has its most serious effects on the heart and the respiratory system. It paralyses the respiratory muscles leading to death. Weakening of cardiac contraction, as well as cardiac rhythm disturbances, compounds the problem. The venom exerts a local effect producing agonising pain with skin and muscle destruction.

Clinical Features

The victim experiences **immediate agonising pain** on contact with the tentacles. With a large sting, sudden **collapse**, **cessation of breathing, cyanosis**, **unconsciousness** and **death** may follow rapidly**.** These effects are particularly dangerous in small children or old frail swimmers.

If the victim recovers, severe pain still persists for many hours, and **scarring** is common in the stung areas due to local tissue destruction.

Treatment

Rescue the victim from the water and prevent drowning. This takes immediate precedence.

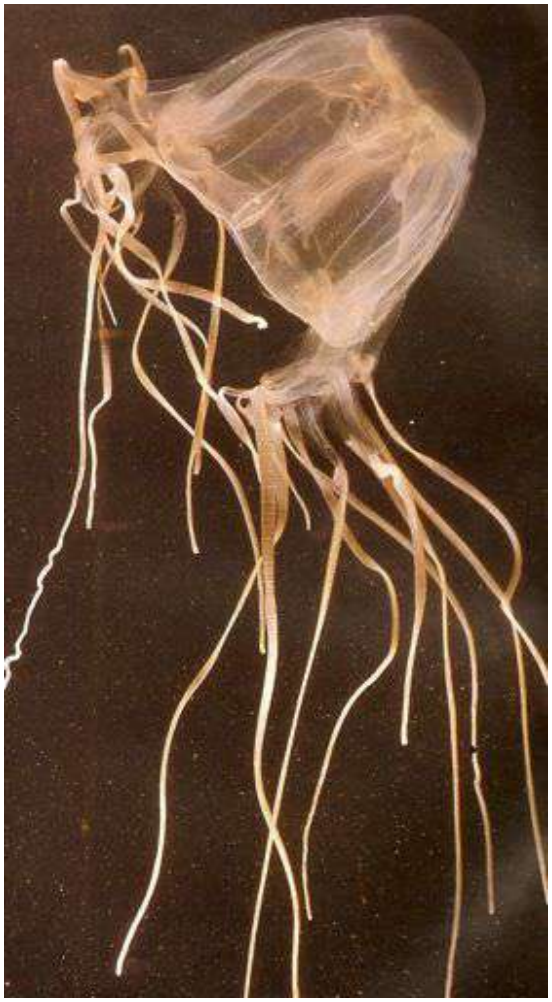

If the patient is unconscious the basic life support (**BSL**) priciples take precedence and should be followed (see Chapter 42) while enlisting medical assistance.

Apply copious amounts of ordinary **household vinegar** to the tentacles and gently remove the tentacles from the victim's skin. The tentacles cannot sting effectively through the thick skin of the palm of the hand and fingers so this may be safer than it sounds. It is important not to rub or damage the tentacles as this will encourage the injection of further venom into the victim.

Alcohol application is no longer advised, as there is some evidence that this may cause the discharge of further venom into the victim, as may pressure-bandage/immobilisation (see later). If the alcohol is of good quality, it may be more beneficial to the rescuer, once the victim has been taken safely to hospital.

The cause of death in box jellyfish sting is usually respiratory arrest. However, this may be transient if the victim is kept alive by expired air resuscitation or other **artificial ventilation** during this period. The victim should be transported to hospital urgently. Most survive, especially if still alive after the first few minutes.

Fig 29.3 Chironex – Box Jellyfish

An **antivenom** against the Chironex box jellyfish may neutralize some of the venom present in the victim's body. It has been developed by the Commonwealth Serum Laboratories (CSL Australia) and may be used in severe cases to prevent cardiovascular collapse, or where significant local scarring is threatened. It may not be as effective against other box jellyfish.

Prevention

The practice of covering as much exposed skin as possible by the wearing of a face mask, wet suit and hood, overalls or a Lycra suit, prevents the access of tentacles to the skin. This protection also reduces the risks of stings from other jellyfish and injuries from corals. Even water repellent skin preparations, such as sun-burn oils and creams, may reduce the danger.

OTHER JELLYFISH STINGS

Several other stinging jellyfish such as the **Portuguese Man-of-War, fire coral** and **stinging hydroids** can produce painful and sometimes incapacitating stings, although they are unlikely to be lethal.

The same technique of general management as described for box jellyfish should be followed (rescue, resuscitation etc. See Chapter 42). However, different local applications seem to work for different species. Vinegar or alcohol may cause further nematocyst discharge in some jellyfish stings. Some degree of pain relief can be afforded by the application of **local anaesthetic** (e.g. lignocaine ointment) to the stung area. Other preparations which have a variable effect, include "Stingose", "Stop-Itch", Tannic Acid Spray, etc. Any **anti-burn** preparation, including ice packs, may give some relief. More recently, application of heat (about 45^{0}C. as for fish stings – see later) has been shown to alleviate some jellyfish stings.

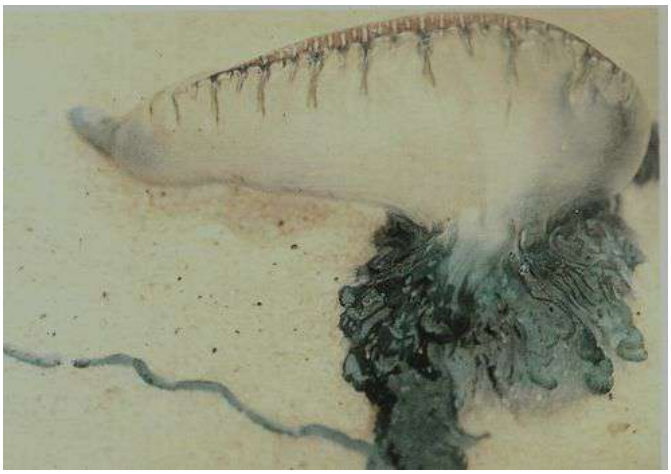

Fig 29.4 Physalia (Blue Bottle, Man-of-War)

Fig 29.5 Stinging Hydroid

IRUKANDJI SYNDROME

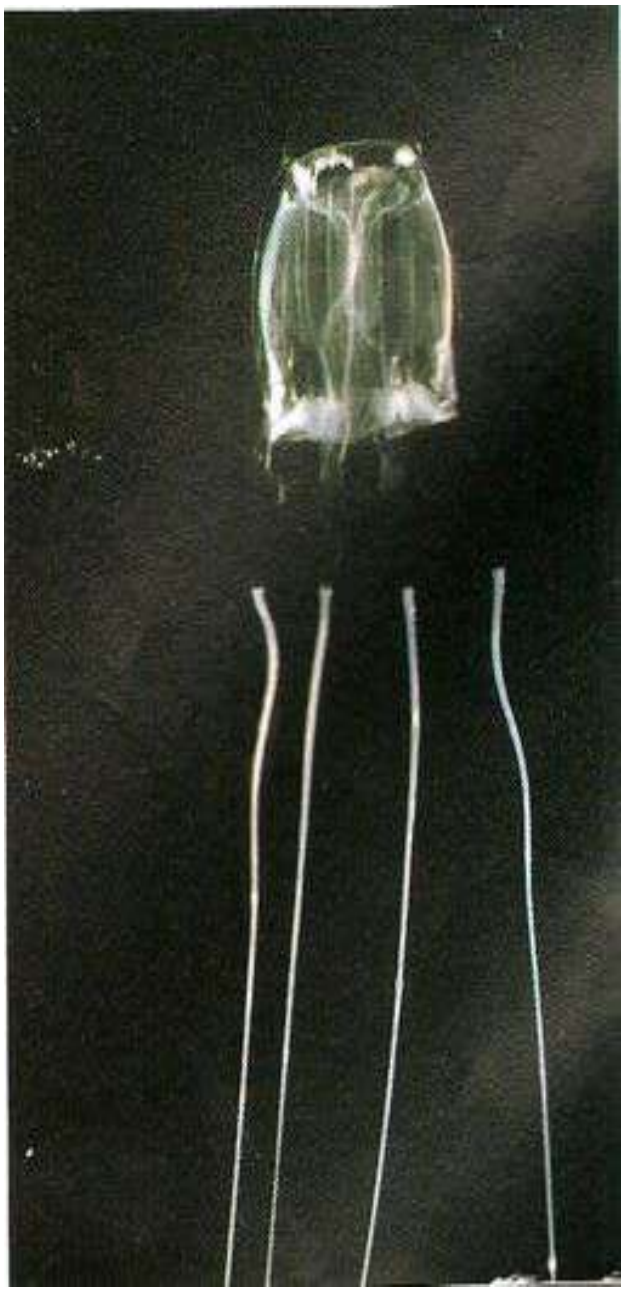

This disorder was first described as a result of an almost unnoticeable small box jellyfish (*Carukia barnesi*) sting with big after-effects. It is now known to be a possible complication of many other jellyfish stings. It has been the cause of severe cardiac illness and has been confused with both an acute abdomen and cardio-pulmonary diseases, because of its various symptoms.

The victim may, or may not, be aware of the sting. If not, there still may be a red patch visible where the sting occurred. After a latent period of between a few minutes and two hours, severe muscular cramps and pain develops (abdominal, spinal, limbs, chest). The diver becomes anxious, restless, sweating and may have gastro-intestinal and respiratory symptoms. There may be increased pulse rate and high blood pressure recorded. Cardiac damage and pulmonary oedema can develop.

Because of the latent period, the relationship to the jellyfish sting may not be appreciated, leading to many other medical diagnoses, including decompression sickness.

First aid treatment may involve copious vinegar application, as for *Chironex* (above) if identified early. Successful medical investigations and therapies are available, but deaths have been reported.

The injury can be prevented by wearing protective clothing (see above).

Fig. 29.6 ***Carukia barnesi***. This is the most frequent animal incriminated in the Irukandji syndrome. Its bell is only a couple of centimetres (one inch) long, but the tentacles, which are not always easily seen, are up to a metre long.

VENOMOUS CONE SHELLS

A small number of species of the cone shell family are capable of delivering a lethal venom. This is injected by a tiny dart shot from a tubular appendage which the animal can direct to any part of its shell. This apparatus is normally used by the animal to kill its prey (usually small fish), but it will use it as a weapon against a human who is careless enough to handle it.

Expert knowledge is required to differentiate venomous from harmless cone shells, and divers are advised to avoid handling them at all. Reef walkers, being less valuable than divers, may do as they wish.

Clinical Features

The initial sting may or may not be painful. It can penetrate cloth and skin. They should not be handled or put in pockets. The toxin affects the heart, skeletal and respiratory muscles. Muscle spasms develop. Death is usually from respiratory arrest.

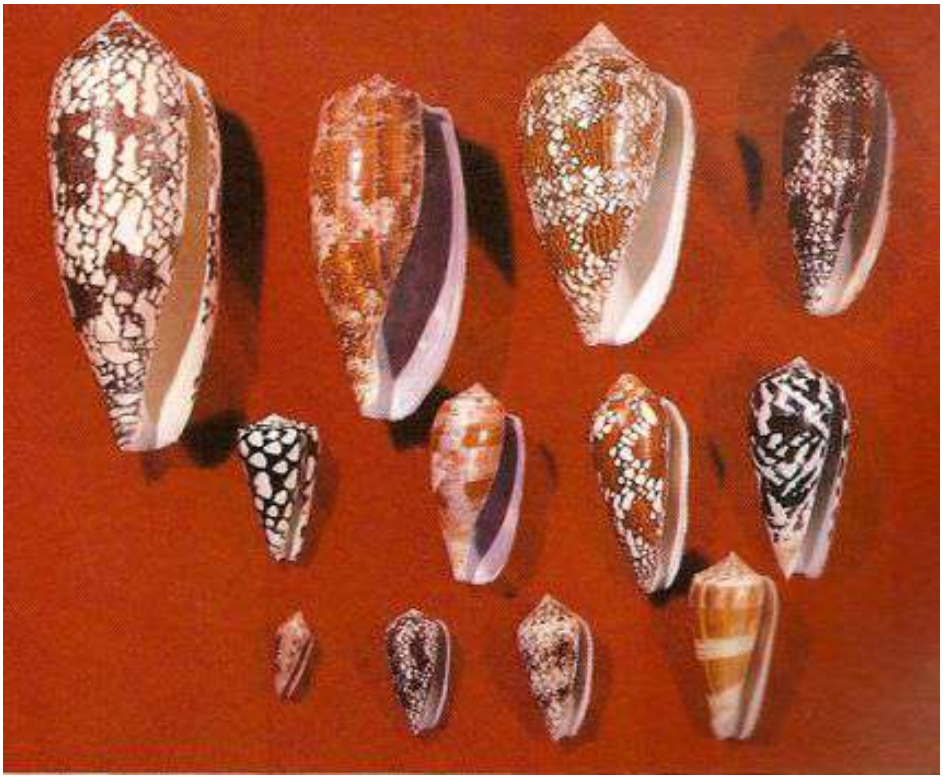

Fig. 29.7 **A collection of venomous cone shells**

Treatment

The prompt application of a **pressure bandage and immobilisation** (see later) should delay the spread of venom from the wound, although there have been no clinical case reports to verify this.

The first aid **basic life support** measures (see Chapter 42) may keep the patient alive until the respiratory paralysis has worn off. This may involve many hours of artificial respiration.

BLUE RINGED OCTOPUS

This attractive little animal is found in rock crevices along the water's edge of many islands in the Pacific and Indian oceans, as well as in deeper water. If annoyed it will display a colourful array of blue or purple rings on its skin. This may arouse the curiosity of a potential victim, especially a child.

Unfortunately it can inflict a small, relatively painless, bite and inject venom through a beak at the base of its tentacles. The bite may go unnoticed by the victim until the major effects of the venom develop.

The injected venom can produce general muscular paralysis within minutes, leading to cessation of breathing. The victim can then remain fully conscious- but unable to communicate with bystanders due to the paralysis. Death can then be due to respiratory failure, unless treatment is given.

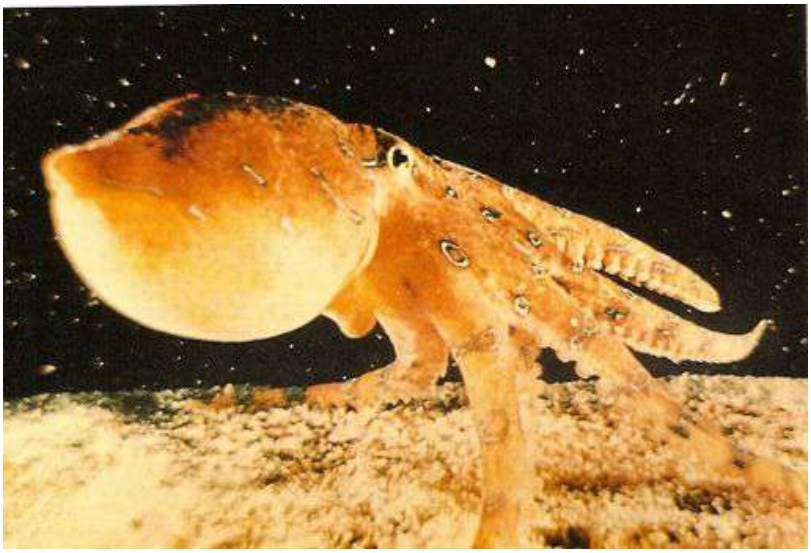

Fig. 29.8 Blue Ringed Octopus
This dangerous little animal should not be handled.

Treatment

Artificial respiration must be continued until recovery **(4 – 12 hours)**. This is necessary because of the respiratory muscle paralysis. Basic life support is needed (Chapter 42)

A **pressure bandage** and **immobilisation** (see later) should be applied promptly to delay spread of the venom, and maintained until full resuscitation measures are implemented.

Case Report 29.2. A diver found a small octopus with attractive iridescent blue rings - hiding in a shell. She placed it under her wet suit vest, intending to show it to her companion later. After the dive she complained of double vision and respiratory difficulty. When she showed the octopus to her buddy, the buddy correctly diagnosed the problem and kept the victim alive by mouth to mouth respiration until hospital was reached. The victim later pointed out that she was not encouraged by comments such as "it looks as though she is not going to make it " from bystanders who had not realised that she was fully conscious, in spite of being paralysed.

SEA SNAKE

Sea snake bites are not uncommon in the Indo-Pacific ocean waters. In certain areas, sea snakes will approach divers underwater. These advances may be inspired by curiosity, as it is rare for sea snakes to bite divers without provocation. They will retaliate if grabbed.

The venom of sea snakes is more potent than that of the cobra. Even when bites occur, the presence of short fangs at the back of the mouth deprives some sea snakes of an efficient way of delivering this venom into humans. Often venom is not injected, despite the biting.

Fig. 29.9 Yellow Bellied Sea Snake *(Pelamis platurus)*

Clinical Features

If envenomation occurs, **s**ymptoms may become evident within minutes to hours after the bite. Muscle weakness leading to paralysis, including **respiratory muscle paralysis** and **asphyxia**, and finally **cardiac failure** may follow the bite.

Occasionally the sea snake bite itself results in severe **lacerations** and **blood loss.**

Treatment

The prompt **pressure bandage + immobilisation** technique (see later) will delay the symptoms until medical assistance, resuscitation facilities and antivenom can be acquired.

The first-aid **basic life support** measures should be instituted where necessary (see Chapter 42). Mouth to mouth respiration is the major requirement. The victim should be taken to

hospital as soon as possible. Serious cases should be treated with sea snake **antivenom** (made by CSL – Australia).

STONEFISH

This is the most venomous fish known. It is extremely well camouflaged and may not move away when approached, as is implied by its name.

It is capable of inflicting severe stings by means of 13 poisonous spines along its back. The spines are able to penetrate rubber soled shoes or neoprene boots. At the base of each spine is a venom sac which empties its contents into the victim's wound.

Fig. 29.10 Stonefish

Clinical Features

Envenomation results in severe **agonising pain at the site** of puncture. Extreme **swelling** and **local paralysis** develops rapidly. The venom can lead to **respiratory distress**, **cardiac disturbances** and syncope (**fainting**) with a reduction in **blood pressure**. Death is uncommon except in children or the infirm.

Treatment

Immersion of the stung area in hot water about **45°C** (first tested by the attendant's hand, to ensure against scalding) often gives significant pain relief and should be employed as soon as possible as a first-aid measure. Elevating the wound may reduce swelling.

The severe pain of the sting can be relieved by the **injection of local anaesthetic** (with no added vasoconstrictor agent such as adrenalin) **into the puncture sites.** This treatment may need repeating several times before the pain stops recurring as the effects of the local anaesthetic injection wear off. A physician may prefer to block the nerve supply to the region with local anaesthetic as an alternative. Cleansing of the wound and antibacterial treatment is required.

The first-aid **basic life support** measures should be instituted where necessary (see Chapter 42). **Antivenom** from the Australian CSL Laboratories is available and its use may be necessary in severe cases.

OTHER SCORPION FISH

Other members of the scorpion fish family such as the **fortescue, lionfish** (or **butterfly cod)** and **bullrout**, produce painful stings similar to that of the stonefish, although both the local and generalised effects are usually not as severe. **Cat fish** have a similar effect.

Fig. 29.11 Butterfly Cod

Fig 29.12 Fortescue

Pain relief can be obtained by **immersing the area in hot water** at about **45°C** (previously tested by immersing an unaffected limb in the water) as for the Stonefish sting (above), while more sustained relief can again be obtained by **injecting the punctures with local anaesthetic (no adrenolin).** Cleansing of the wound and antibacterial treatment may be required, and the wound should be elevated.

STINGRAY

These flattened relatives of the shark have one or more long bony spines, which are intended for self defence, at the base of the tail.

The animals often bury themselves in the sand where they can inadvertently be stood upon, or otherwise disturbed, by an unsuspecting diver. The stingray defends itself by swinging its tail quickly over the top of its body, driving the spine into anything which happens to be above it.

The spine may produce a puncture and deposit venom. Its serrated edge can cause serious or even lethal lacerations. Parts of the spine, marine organisms and a toxic slime may be left in the wound to cause infections and local inflammation.

Clinical Features

Fig. 29.13 Typical manner in which a stingray injury occurs

Pain caused by the toxin is **immediate** and **very severe. Swelling** is rapid. Toxin may be absorbed into the body producing **generalised symptoms** of syncope (**fainting**), weakness, palpitations, low blood pressure and disturbances of cardiac rhythm. **Death** is rare – except in cases where a vital organ such as the heart have been pierced by the spine.

Despite initial improvement, there can be a deterioration in the clinical state some days later, if there is any foreign material or organisms left in the wound, or if damaged tissue becomes necrotic. For this reason, all cases should be referred for medical assessment.

.

Treatment

The **basic life support** resuscitation may be needed (Chapter 42). **Hot water** immersion treatment and/or **injected local anaesthetic**, as described for stonefish injury, are useful. The wound should be **cleaned** to remove any foreign body or venom. An X–ray, ultrasound, CT or MRI may demonstrate an embedded spine, which needs to be removed surgically. Local antibiotic cream, and often oral antibiotics (such as doxycycline), are indicated.

Prevention

Shuffling the feet while wading in areas frequented by stingrays will usually allow them to move away. Footwear may not be adequate to protect the feet or lower legs from these injuries. Diving into shallow waters where these animals inhabit could be dangerous. Divers should swim well above the sea bed.

OTHER MARINE ANIMALS

Many other marine animals may cause major or minor injuries, and require different first-aid treatments. These, together with more detailed descriptions of the potentially lethal animals and those poisonous to eat, are fully discussed in the companion text "Dangerous Marine Creatures" by Dr. Carl Edmonds (See appendix A).

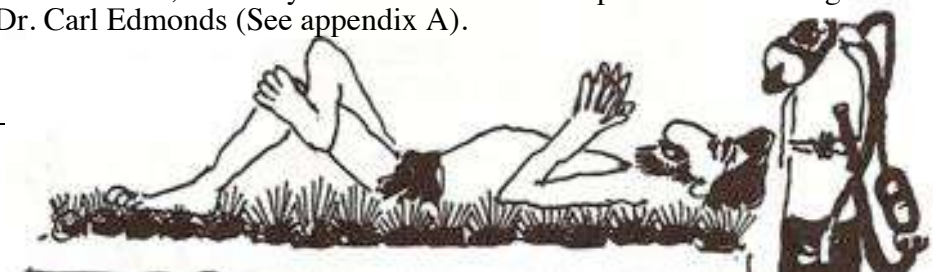

PRESSURE BANDAGE + IMMOBILISATION TECHNIQUE

This is used to delay the absorption of venom from a wound. A bandage (preferably stretchable) is applied over the bite and then wrapped around the limb (and extending up the limb) tight enough to block the drainage vessels (lymphatics). The pressure is approximately the same as that used to treat a sprained ankle.

Care must be taken not to put the bandage on so tight that it causes pain and cuts off circulation. For this reason the technique is not applicable to painful, swollen bites or stings that already have circulation impairment – such as fish stings

The limb should then be immobilised with a splint to prevent any local muscle movement (this spreads the venom despite the bandage).

The pressure bandage+immobilisation of a limb should be continued until the victim has knowledgeable medical personnel and facilities available to cope with the envenomation. This happens as the bandage is released and the venom moves into the bloodstream. The doctors may well administer antivenom (if available), before removing the bandage.

The technique is especially applicable to sea snake, blue ringed octopus and cone shell bites. A variant may be used if the bite is on the torso, with a pad and bandage to produce the pressure.

1

2

3

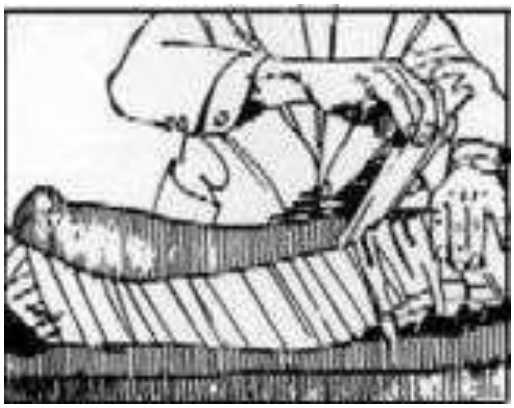

4

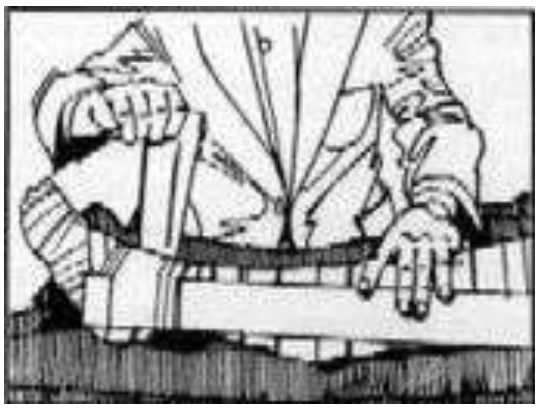

Chapter 30

HEARING LOSS

This chapter may be easier to understand if the structure and function of the ear, as outlined previously in Chapter 9, is reviewed. All cases of hearing loss should be assessed by a diving physician.

Divers frequently complain of a sensation of hearing loss which cannot be verified when hearing tests (pure tone audiometry) are performed. It is likely that currently available hearing tests, such as speech discrimination, are not sophisticated enough to detect such subtle alterations in the sensation of hearing.

The **causes** of demonstrable hearing loss fall into two categories:

- **Conductive hearing loss –** where there is some impediment to the conduction of sound vibrations (usually in the external and middle ear) *en route* to the hearing organ.

- **Sensorineural (nerve) hearing loss –** where sound vibrations reach the hearing organ (cochlea) in the inner ear, but the sound is not perceived due to damage of the cochlea or its nerve.

CONDUCTIVE HEARING LOSS

The likely causes of conductive hearing loss are in the external or internal ear.

External Ear Obstruction

Any obstruction to the outer ear such as wax accumulation, plugs or hoods, outer ear infections (see Chapter 28) or exostoses (see Chapter 32) can cause this.

Tympanic Membrane Damage

This membrane can be torn by:

• **Excessive stretching** during descent (middle ear barotrauma). See Chapter 9.
• A **shock wave** passing down the ear canal, such as an **underwater explosion** or a **pressure wave from a fin** passing close to the divers ear.
• An excessively forceful **Valsalva** manoeuvre has also been known to rupture the tympanic membrane from within.

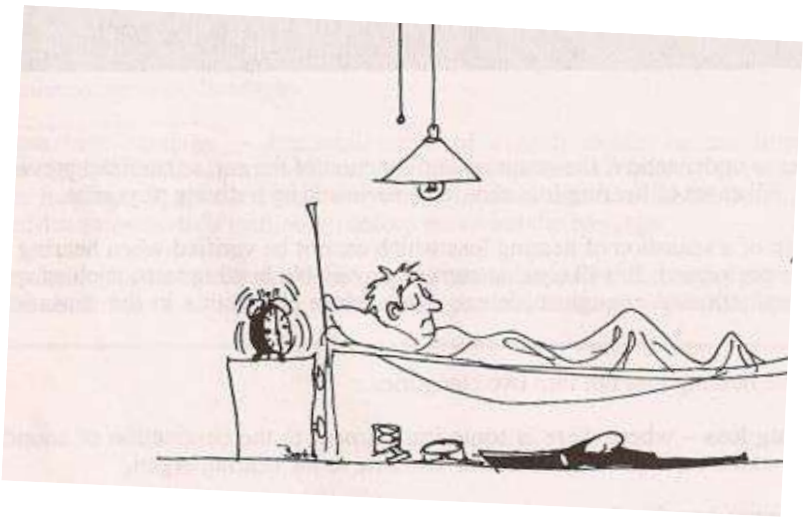

Fig. 30.1

Case History 30.1. A diver swimming closely behind his buddy suddenly felt pain in his left ear as his buddy's fin swept past his ear. Dizziness followed but soon settled. He surfaced and noticed a small amount of blood coming from his ear.

Diagnosis: Rupture of the ear drum caused by a pressure wave from a fin. The dizziness was due to cold water entering the middle ear through the ruptured ear drum. The blood was extruded by gas expanding in the middle ear, during ascent.

Case History 30.2. An old professional hard hat diver who smoked cigarettes and had suffered repeated tympanic membrane ruptures from barotrauma was in great demand at parties because of his ability to blow smoke from his ears. He claimed that during the latter part of his career he no longer needed to equalise.

Diagnosis A: Chronic perforation of the ear drums. His unusual talent was made possible by smoke passing from his throat to his ears through the Eustachian tubes, after he takes a drag from the cigarette and performs a Valsalva manoevre. His ears became self equalising later in his career because of permanent holes in his ear drums. Although in demand at parties, unfortunately he often misheard the directions and turned up at the wrong address.

Diagnosis B: Poor hearing due to recurrent barotrauma and chronic perforation of tympanic membranes.

Middle Ear Disorders

Disturbances of the middle ear impair conduction of sound vibrations from the ear drum, through the bony chain to the cochlea. Causes include:

• **Middle ear barotrauma** which produces bruising and swelling of the middle ear tissues and bleeding into the middle ear space. Both factors dampen sound transmission (see Chapter 9).

• **Middle ear infection (otitis media)** which causes swelling and inflammation. This fills the middle ear space with pus, which impairs sound conduction (see Chapter 28).

SENSORINEURAL HEARING LOSS

This is often accompanied by tinnitus (ringing in the ears) and sometimes by disorientation. Tinnitus can sometimes be more incapacitating than hearing loss. In recent hearing loss, early treatment increases the likelihood of improvement in hearing.

Noise Induced Deafness

Repeated exposure to loud noise may produce a progressive hearing loss which usually affects high frequency hearing first. This loss may be noticed by hi-fi enthusiasts who will complain that music has lost its sparkle. It is often insidious and may not be noticed for many years. Occasionally, a single exposure to loud noise can cause noticeable hearing loss immediately. Rock concerts and discos are also incriminated.

Noise induced hearing loss may be transient in the early stages but repeated exposure leads to permanent deafness, which worsens with more exposure. Industrial noise usually affects the ears symmetrically, but other noise such as gunfire, commonly affects only one ear (the one exposed to the noise or blast).

The diving environment is often a noisy one. Recompression chambers, compressors, boat engines, helmets and compressed air leaks are often loud enough to present a threat to the hearing of those in their vicinity. Divers should take care to protect their ears when necessary by the use of industrial protective ear muffs or ear plugs (but not when diving).

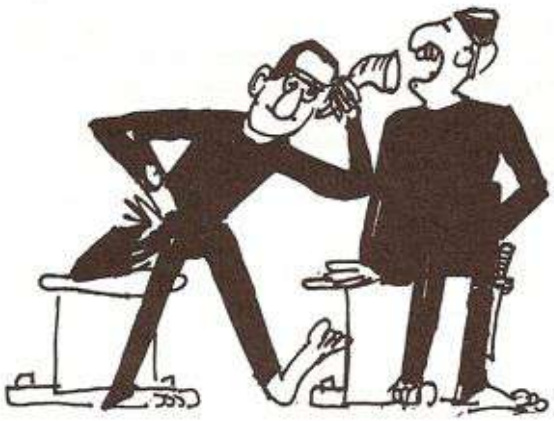

Fig. 30.2

High frequency hearing loss also renders consonants such as "S" or "CH" difficult to hear – hence the story of the yacht owner who was delighted when the curvaceous blonde diver with sensorineural deafness accepted his invitation to "crew" on his yacht.

Barotrauma

Inner ear barotrauma or associated round window fistula may lead to temporary or permanent hearing loss (see Chapter 9).

Decompression Sickness

Inner ear damage is an uncommon complication of decompression sickness (see Chapter 15) in shallow air breathing divers. It is more common in deep, helium or mixed gas divers.

OVERVIEW OF HEARING LOSS

• **All prospective divers must have their ears examined to exclude ear problems** likely to predispose to barotrauma.

• **All divers should have a baseline audiometry** performed, to enable the physician to detect early hearing loss, to make assessment of future hearing problems much easier and to allow early and more knowledgeable treatment to be administered in the (not uncommon) event of a diver presenting with hearing loss.

• **Any case of hearing loss in a diver should be assessed as soon as possible by a diving physician.** The doctor will take a history of the condition, examine the ears, test the hearing by pure tone audiometry at least, and possibly perform other specialised investigations such as bone conduction, speech discrimination, impedance audiometry, diving tympanogram, electro-nystagmograms and brain stem evoked auditory responses.

The cause is usually fairly obvious and **management** of the specific conditions is covered in other chapters.

• **Divers with pre-existing hearing loss** should realise that any deafness arising from barotrauma will be added to the loss they already have**.** It is also believed that people with hearing impairment are more susceptible to further damage than others. Divers who are aware of hearing loss should discuss the implications with a diving physician.

• **Occupational implications are raised.** Those who need excellent hearing for their livelihood, such as musicians, cardiologists, sonar operators and airline pilots, should consider whether the small but real risk of hearing damage associated with diving is worth taking.

• **Hearing loss is sometimes associated with abnormalities of the body's balance mechanism**, which might have safety implications with diving (see Chapter 31).

Chapter 31

DISORIENTATION

Accurate orientation whilst underwater is important for the diver so that he can find his way back to the surface. On land the diver uses a combination of vision, the feeling of gravity on his body, and the balance organs (vestibular system) to tell him which way is up.

When underwater, the diver becomes virtually weightless, depriving him of the sensation of gravity and making him reliant on vision and his balance organs for spatial orientation. With poor visibility, even the visual cues are lost, leaving the diver almost totally reliant on his balance organs for this orientation. A sensation of disorientation requires investigation by a diving physician.

The experienced diver can acquire some clues about his body position from:

- the way heavy objects such as the weight belt or other metal objects hang,
- the direction his bubbles are going,
- the direction of a life-line or hookah hose.

Inexperienced or panicking divers are often unable to use this subtle information. If the diver becomes disoriented he is likely to experience anxiety. Panic can easily ensue.

VERTIGO — OR "DIZZINESS"

This is a false sensation of spinning or moving. The diver may either have a sensation of himself spinning or the environment spinning about him. It happens because the balance organ (**vestibular system**) can be unreliable underwater – it was designed to work on land. Under certain circumstances it can supply the brain with misleading information which is falsely interpreted as movement.

The sensation of vertigo is bad enough, but it is often accompanied by **nausea** and **vomiting** which can threaten a diver's life. These symptoms may vary from mild to very severe. Vomiting into, and then breathing from, a demand valve is not easy underwater.

Function of the Vestibular System

The balance or vestibular system comprises two marble-sized structures located in the skull above and behind the middle ear space on either side of the head. Each vestibular apparatus has two parts. Abnormalities of either cause vertigo and disorientation. They are:

❑ A system of three interconnecting tubes (semi-circular canals).

These are aligned at right angles to each other and filled with fluid. They detect movement in all three planes. If the body rotates, the fluid in these three canals tends to lag behind, due to its inertia. The differential movement of the body and the fluid is detected by nerve endings – hair like projections into the fluid (hair cells) at the base of each canal.

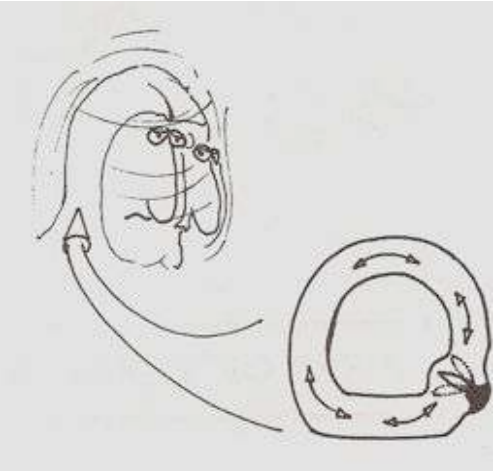

The semi-circular canals are located close to the ear canal. Cold water entering the ear canal can cool them slightly, causing convection currents in the fluid. The movement of the fluid is detected by the hair cells and, if this does not synchronise with stimuli from the other side, causes vertigo. This is termed **caloric** induced vertigo and is usually associated with the diver being in a near horizontal position**.**

Fig. 31.1

❑ The Otolith organ.

These other fluid-filled structures have a viscous base which contains minute calcium granules. Hair like projections of nerve cells penetrate this gel and detect any movement of the granules. Because of their weight, the granules tend to move in response to **gravity** and **acceleration**. The hair cells detect this and continuously inform the brain about which way is up and the direction of any acceleration.

Pressure changes can cause **barotrauma** induced vertigo, and it is possible that this could be due to stimulation of the otoliths or the semicircular canals and is usually associated with the diver being in a more vertical position.

Fig. 31.2

CAUSES OF VERTIGO

Problems arising from the vestibular system fall into two principle categories:

- unequal vestibular stimulation and
- unequal vestibular response.

Unequal Vestibular Stimulation

If both vestibular systems are equally sensitive but are stimulated unequally, then vertigo may result due to the unequal responses received by the brain.

Any condition causing more cold water to enter one external ear more than the other causes unequal **caloric** stimulation. Wax blocking one ear, an air bubble, otitis externa, exostoses, ear plugs or a ruptured ear drum will all have this effect.

With middle or inner **ear barotrauma** (see Chapter 9) affecting one side, or **decompression sickness** (see Chapter 15) on one side, unequal vestibular stimulation may result in vertigo.

Failure of the ears to equalise pressures to the same degree can stimulate the vestibular system unequally. This is not uncommon on ascent, as the pressure of the expanding gas in the middle ear spaces can become greater on one side than the other due to differences in patency of the Eustachian tubes releasing it. This is termed **Alternobaric Vertigo** or **middle ear barotrauma of ascent** (see Chapter 9). It is very common and is often noticed as the diver ascends a metre or so, from depth. He may even be aware of the sensation of one Eustachian tube opening before the other, or of the expansion of air in the other middle ear.

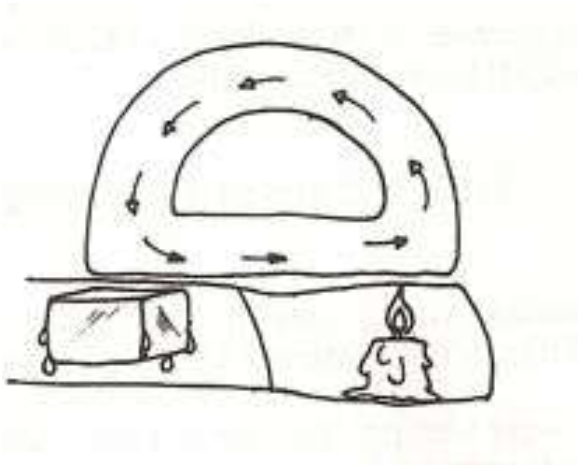

Fig. 31.3
Caloric stimulation producing "convection" current flows in inner ear fluids.

Case Report 31.1. A diver using hookah apparatus on a training dive at night lost contact with the bottom. He was unable to see his bubbles and had no idea which way was up, but by feeling the direction of his air hose, he was able to establish where the surface was.

Diagnosis: disorientation due to the reduced sensory input (decreased vision) of night diving.

Case Report 31.2. An inexperienced diver had difficulty equalising his middle ear during descent. He continued to descend in spite of this. The pain was abruptly relieved and he became aware of a hissing sound and cold sensation in his ear as his ear drum perforated. Seconds later he developed a severe sensation of spinning which was accompanied by nausea. He clung onto his shot line and was relieved when the vertigo gradually passed off after several minutes.

Diagnosis: Vertigo due to one sided vestibular stimulation (caloric) from cold water entering the middle ear when the ear drum ruptures. As the water warmed to body temperature, the cooling effect to the vestibular system subsided.

Unequal Vestibular Response

The two vestibular systems are normally equally sensitive to any stimulus such as movement. Some people may have unequal sensitivity. This can be due to a slight imbalance (either overactive or underactive function) which has been present from birth or to damage to one side from causes such as ear barotrauma or some medical conditions. In this situation, the same stimuli cause a greater response from one side than the other, and is experienced by the person as vertigo.

People with this problem unconsciously adapt by avoiding sudden movements of the head or body. They learn by experience to avoid gymnastics and roller coaster rides, but are usually not aware of the dangers posed to them by the extreme stimuli which are commonplace in diving.

Water entering the ear canals is a potent cause of vertigo in these people. As mentioned above, this water can cool the fluid in the semi-circular canals setting up convection currents (**caloric stimulation**). If the vestibular systems on each side are not equally sensitive, a stronger response will be produced from one side. The brain will interpret this information as indicating movement, and the diver will experience vertigo and feel disorientated.

During ascent or descent, equal **pressure changes** in the middle ear can also produce vertigo and disorientation in those with unequal vestibular response.

Other Causes of Vertigo

If the diver is deprived of normal visual cues in conditions of **poor visibility** or at night, it is possible for the resulting disorientation to culminate in vertigo, especially in inexperienced divers.

Nitrogen narcosis may aggravate vertigo. Vertigo has also been recorded as a symptom of other conditions not commonly encountered by recreational divers. These include **oxygen toxicity, carbon dioxide toxicity, carbon monoxide toxicity,** and **high pressure neurological syndrome**.

PREVENTION

A thorough diving medical examination before a prospective diver undertakes training can exclude some of the factors predisposing to vertigo, and is advised for all divers.

A diver experiencing vertigo under water should avoid unnecessary movement and hold onto a fixed object if one is available. Fortunately, in most cases, the vertigo is short lived. If the vertigo fails to abate and there are no obstacles above him, ditching of weights, possibly with cautious inflation of a buoyancy compensator, should return the diver to the safety of the surface.

Any diver who experiences vertigo under water and survives should abandon the dive, **consult a diving physician** to investigate, identify and correct the cause before diving again.

Occasionally the diver may be very astute and be aware of the cause of the vertigo, and may then correct the problem e.g. with middle ear barotrauma of ascent (see Chapter 9).

CONCLUSIONS

While **disorientation** under water can be unpleasant and dangerous, **vertigo** can be life threatening because of the risk of vomiting or panic.

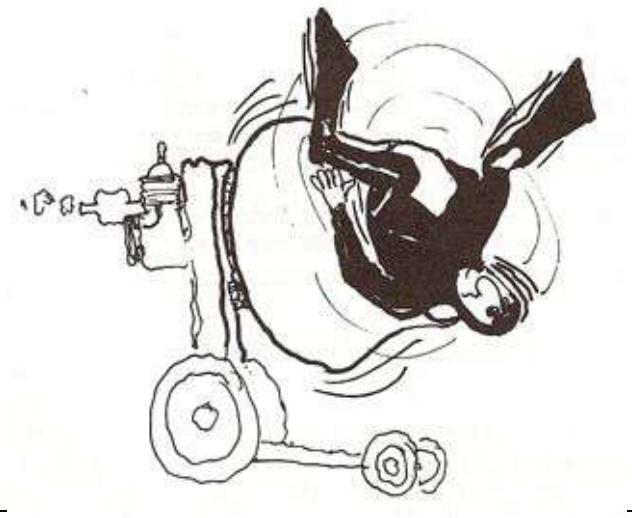

Fig. 31.4

Chapter 32

MISCELLANEOUS DISORDERS

PULMONARY OEDEMA

Scuba Divers Pulmonary Oedema is usually described as an uncommon disorder, often in apparently healthy individuals. In a survey of scuba divers about 1% described it. An individual predisposition for pulmonary oedema is a likely factor since a diver or swimmer with pulmonary oedema may have other episodes previously or subsequently.

It presents clinically with difficulty with breathing, with fast shallow respirations and a sensation of crackling sounds in the chest. It may be associated with fatigue, cough, sometimes blood-stained, and possibly a bluish tinge to the lips, tongue and face (cyanosis).

Symptoms usually resolve rapidly (some hours) after the immersion, but deaths have been reported. Treatment includes oxygen inhalation and medical assessment is required to verify the illness and exclude any predisposing features, such as undiagnosed heart disease.

Whether the disease is due to the individual diver, the dive profile, environmental conditions or the dive equipment, is unknown. It is more common in older divers, possibly aggravated by the effects of immersion, breathing against a inspiratory resistance from the regulator, cold exposure, hypertension, heart disease or using certain drugs (especially beta blockers).

Swimmers and surfers can also develop pulmonary oedema, sometimes for similar reasons as scuba divers, but often it is seen in very fit young athletes, produced because of the extreme exertion encountered by some excessive training regimes.

Other diseases that can produce pulmonary oedema and cause diagnostic confusion are the salt water aspiration syndrome, drowning, respiratory oxygen toxicity, gas contaminations, cold urticaria and asthma. Pulmonary decompression sickness, pulmonary barotrauma and the so-called 'deep diving dyspnoea' are diving disorders that may cause diagnostic confusion. Anxiety produced hyperventilation may also cause some diagnostic confusion, but this has none of the other respiratory manifestations.

CONTACT LENSES

Contact lenses are a convenient alternative to spectacles but can be a source of problems to the diver. The most common of these is **loss** of the expensive lens during removal of the face mask. The eyes should be shut while removing the mask underwater or on the surface.

In certain circumstances, especially during long or deep dives or in compression chambers, it is possible for **gas bubbles** to form behind the contact lens (particularly with hard, non gas-permeable lenses) causing pressure and damage to the cornea of the eye. If this happens, the diver may experience discomfort in the eye, blurred vision and the appearance of halos around bright lights. Long term effects could include scarring of the cornea.

Gas bubble damage can be overcome in the hard contact lens by an optician drilling a small hole in the centre of the lens (a **fenestrated** lens) which allows gas bubbles to escape. This has no effect on the visual performance of the lens. Soft contact lenses are usually not a problem because of their gas permeability and flexibility.

It is now relatively easy to have corrective lens ground into or attached to the diver's face mask, as an alternative to contact lenses (see Chapter 5).

MUSCULAR CRAMPS

Cramp is a painful spasm of a muscle group. It is common in divers and can cause a dangerous diversion or incapacity. The muscles most commonly affected are those in the sole of the foot, the calf and the thigh, but other muscles can also be involved.

Unusual exertion of the muscles, due to changes in fins or equipment, makes cramp more likely, especially if the diver is generally unfit. Cold water is another predisposing factor.

Cramp is managed by slowly stretching and maintaining tension on the muscle involved. Sometimes this may require the diver to actually stand and push down with his toes onto some firm underwater surface in order to stretch the muscle. Ditching of weights underwater or inflation of the buoyancy vest on the surface may be helpful in an emergency, avoiding the need to continue swimming.

This condition can be inconvenient or even dangerous if the diver is simultaneously coping with environmental problems such as white water, strong currents or tidal flows.

It is best **prevented** by maintaining a high level of physical fitness, using familiar and comfortable fins and having adequate insulation from cold water.

EAR PROBLEMS

Wax (Cerumen)

Ear wax (cerumen) is a protective substance which coats and waterproofs the external ear canal. Occasionally the ear produces excessive wax which accumulates and obstructs the ear canal, or contributes to water retention with subsequent otitis externa (see Chapter 28). It may produce curable hearing loss, or caloric induced vertigo if water is able to enter only one ear (see Chapter 31). Divers may try to remove this wax with cotton-tipped "buds", but unfortunately this often results in infection or the wax being compacted even tighter in the canal, precipitating total obstruction.

The excessive wax is easily removed by a diving physician using an instrument or syringe. This leaves the ear canal somewhat open to infection however (otitis externa – see Chapter 28), and therefore should not be done unless the wax totally occludes the canal. Ear drops are readily available (Cerumol, Waxsol, olive oil etc.) which help to soften wax so that the normal self cleaning function of the canal can proceed more easily. Diving itself aids in wax removal.

Exostoses

The inner part of the external ear canal passes through bone. People who swim or dive regularly, especially in cold water, sometimes develop outgrowths of this bone, known as exostoses, bulging into the ear canal. These can cause partial obstruction which may lead to the accumulation of wax and the retention of water causing infection or hearing loss.

Large troublesome exostoses can be removed surgically, however this is not usually necessary.

Others

Infections (otitis externa, otitis media) are discussed in Chapter 28, hearing loss in Chapter 30, vertigo and disorientation in Chapter 31, barotrauma in Chapter 9 and decompression sickness in Chapter 15.

HEADACHE

Headache during or after a dive is a frequent complaint and can be caused by conditions ranging from trivial to life threatening. It always requires careful assessment.

The most likely cause of the headache can usually be deduced from the past medical history, location of the pain, dive history, mode of onset and progression.

Details of clinical and diagnostic features can be found in the relevant chapters elsewhere. Although most headaches are not serious, the more serious causes will be discussed first.

Decompression Sickness and Pulmonary Barotrauma

Air emboli and bubble development in the brain can cause brain injury and swelling which often presents as headache. This may start within a short time after surfacing, or may be delayed for several hours. Headache followed by confusion or loss of consciousness is very suggestive of this dangerous disorder. The dive profile is helpful in diagnosing headaches of this type (see Chapters 11 and 15).

Sinus Barotrauma

This condition usually affects the various sinuses located around the eyes, or the maxillary sinuses in the cheek bones. Sharp pain in the affected sinus may be experienced during descent or ascent, or a more dull pain in the region of the sinus may be felt after the dive (see Chapter 10). A more serious and difficult-to-diagnose sinus headache can develop in the sphenoidal sinuses, a deep and central headache. The barotrauma headache is not usually long lasting.

Pain may be referred from the sinus to the upper teeth or behind the eyes. After minor barotrauma, an infection (**sinusitis**) can develop hours or days after the dive, causing a headache in similar sites to those mentioned (see Chapter 28).

Migraine

This condition can be a worrisome problem in divers. It is common in the general population.

Clinical features.

These may include an "aura" before the onset of the headache, with visual effects ranging from flashes of light, shimmering lines, partial loss of a visual field to mild blurring of vision. A severe headache aggravated by bright lights, usually accompanied by nausea and vomiting, and sometimes numbness, tingling, weakness or paralysis of the limbs, most often follows the visual aura.

Migraine headaches can be trivial or can be associated with vomiting, severe incapacity and neurological symptoms (visual disorders, numbness or 'tingling sensations' in arms or legs etc.). These more severe symptoms lead to diagnostic confusion with air embolism and decompression sickness and may result in an emergency evacuation and inappropriate treatment.

A severe migraine developing during a dive can incapacitate the diver or induce vomiting underwater with subsequent drowning.

For reasons which are not well understood, mild migraine sufferers can sometimes have very severe and unusual migraine attacks precipitated by diving. It may be that this is a response to bubbles within the cranial extravascular system. Migraine may also result from excessive exercise and carbon dioxide/oxygen pressure variations. Cold and exertion are also possible aggravating causes.

For these reasons migraine sufferers are not encouraged to dive. Nevertheless, some have no "neurological" features and are very infrequent and mild. Then if they do dive, they are usually **restricted to non-decompression dives** and to **less than 18 metres** and long surface intervals (i.e. dives that do not typically produce intra-arterial bubbles or cerebral decompression sickness).

Tension Headache

Diving and training for diving can be a stressful experience which can cause headaches in susceptible individuals from excessive muscular tension. These individuals will often recognise the headache as similar to those associated with other stressful experiences. Most are frontal or involve the neck and back of the head.

Mask Strap Tension

Inexperienced divers often tighten their mask strap excessively in the hope that the alarming prospect of loss of the mask underwater, can be avoided. Excessive tension of this strap interferes with the blood supply to muscles around the skull, causing a headache similar to tension headache. The pain is prevented by slackening the strap. As the diver gains confidence in his ability to deal with a flooded or displaced face mask, the need to keep the strap excessively tight disappears. Some headaches are related to the design of the strap (ie. wide single straps verses narrow split straps.) Trial and error may sort out this type of problem.

Carbon Dioxide Toxicity

This is a frequent problem with re-breathing equipment, but is sometimes observed with scuba. It can develop as a consequence of a breathing resistance from faulty regulators, possibly from excessive depth (when the air is more dense) or, more likely, from a voluntary inhibition of breathing (slowed or shallow breathing, “skip breathing”) in an attempt by the diver to reduce his air consumption. This popular explanation still awaits experimental or clinical proof.

The headache is often severe, throbbing and unresponsive to analgesics. It may last for an hour or more. See Chapter 22.

Other Types of Headache

Cold water entering an ear canal can cause headache (or earache) when it comes into contact with the ear drum (see Chapter 27). This is easily prevented in swimmers with normal ear plugs. These cannot be used safely by divers. Special ear plugs designed for divers, with perforations, may work The best prevention is a neoprene hood, which allows the trapped water to warm to body temperature. Some masks are designed to include the ears.

There are many **other causes** of headache, including neurological, barotraumatic, thermal, orthopaedic and vascular mechanisms, that are too complex to be assessed here. Any headache associated with diving deserves investigation, before its consequences during future diving become more serious than those occurring while on land. Exertional headaches cause particularly difficult diagnostic problems.

SUNBURN

Sunburn, especially in tropical areas, is a common problem for divers. It is caused by ultraviolet radiation from the sun. This radiation is scattered by the atmosphere and reflected from water so that even sheltering in shade does not provide complete protection.

The **clinical features** of sunburn have been experienced by almost all divers and do not require elaboration.

Treatment.

This is essentially symptomatic. Further exposure to sunlight (even indirectly) should be avoided. A soothing or cooling lotion is often of value in relieving the pain, and steroid (cortisone) creams may be beneficial in severe cases. Blisters should not be ruptured as this invites secondary infection.

Prevention.

Protection can be afforded by covering the skin by clothing, by wearing a hat and by the use of a broad spectrum UV screening cream or lotion. Snorkel divers are advised to wear one of the lightweight protective Lycra suits, which also give protection against marine stingers and coral cuts.

Ultraviolet screening agents are now coded by a SP number which gives an approximate indication of the degree of protection compared with

unprotected skin e.g. SP 10 cream will protect the skin from burning for at period 10 times longer than unprotected skin. Unprotected skin can begin to burn in 15 minutes in strong sunlight so that a sun screen with this level of protection can be expected to protect for 2.5 hours if an adequate thickness is maintained and the screen is not washed off. SP 15+ creams are even more effective and are advised.

Prolonged exposure to sunlight is associated with an increased incidence of skin cancer and premature skin ageing.

SEASICKNESS

This is a distressing and potentially hazardous problem for divers. It usually develops in susceptible individuals in the dive boat but can also develop underwater, during decompression on a shot line, in rough conditions or with underwater surge. On the boat, less attention is paid by the sea sick diver to dive planning and equipment preparation.

The associated vomiting causes dehydration on the boat and requires considerable skill to cope with underwater, if the diver is to continue breathing through his demand valve. It does have the advantage of attracting all sorts of fish homing in for a free feed.

Another potential problem relates to the sedating effect which is produced to some degree by most of the available anti-seasickness medications. This will affect judgment and aggravate nitrogen narcosis.

Prevention.

General measures to be taken include:

- remain in the centre line of the boat, but not near the bow (reduce spatial movements),
- positioning in the boat so that head movement is minimised, remain still (lie down),
- either keeping eyes closed or focusing on the distant horizon (avoid reading),
- if in an enclosed cabin, ensure air circulation with a fan if possible.

If mildly seasick, swimming or snorkeling around on the surface of a sheltered area for a short while will often settle symptoms. The diver can then reboard the boat to don gear and start the dive.

Short acting anti-seasickness tablets such as cyclizine are effective if taken 1 or 2 hours before boarding the boat. These last about 4 hours.

Another effective preventative measure is to take promethazine tabs (a well known oral antihistamine), 25 mg. at bed time the night before. It will cause sedation during the night but this corresponds to the normal sleep time. One dose at night will provide some resistance to sea sickness for the early part of the following day, with minimum sedation. The **depth** of diving should be **limited to less than 30 metres** (100 ft.), maximum, and preferably less than 18 metres (60 ft.) if drugs are used. A cup of coffee (caffeine) beforehand reduces seasickness and counters sedation in some.

In all cases, medication should have been tried previously (a "dry run") to ensure adverse side effects are not produced. It should not be taken if alcohol has been consumed because of additive effects.

Transdermal ("Scop") skin patches are not recommended for diving due to side effects and variable absorption, but may be effective for sailors.

Acupuncture (via acupressure pads) and ginger, although currently fashionable, are really only of psychological value.

TEMPORO-MANDIBULAR (JAW JOINT) ARTHRITIS

Novice divers tend to be apprehensive underwater, especially about the reliability of their air supply. They therefore clamp their jaws tightly on the mouth piece, causing excessive stress on the joint between the upper and lower jaw. This can cause minor injury to the joint, manifested by spasm of the jaw muscles, pain, tenderness over the joint (in front of the ear), and inability to fully open the jaw.

In recreational divers this condition is usually temporary and is reversible by correcting the cause. The diver is encouraged to grip the demand valve less tightly with the jaws. Some older demand valves are heavy and bulky, placing undue stress on the jaw, while other types may be positioned so that the air hose pulls the jaw to one side, causing uneven and excessive strain.

In some older divers, permanent arthritic changes to the joint can occur, from this cause. Individually mouldable lugs on the mouthpieces of snorkels and regulators may help minimize these effects in some cases.

EXPLOSIONS – UNDERWATER BLAST

Fig. 32.1

This topic is included only as a warning for recreational divers not to use explosives underwater. Military divers are particularly at risk from these hazards, even in training – because of the use of "scare charges" which are designed to discourage underwater saboteurs and are sometimes used in the vicinity of trainee divers to toughen them up.

When an underwater explosion is observed from the surface, a sudden explosive projection of water and foam into the air can be seen immediately after the explosion. This is the effect of the pressure wave emanating from the blast when it meets an air-water interface.

A similar effect is produced at air-tissue interfaces in the body as the shock wave travels through the diver. This can shred tissues such as lungs, intestines, sinus cavities and the middle ear spaces, which are in contact with air – all gas containing spaces within the body can be affected.

Clinical features.

The organs worst affected are the lungs and intestines. Rupture and bleeding of the tissues in the lungs and bowel cause:

- chest pain
- shortness of breath
- vomiting or coughing up of blood
- passage of bloody or black bowel motions.

Damage to the ears and sinuses causes features similar to barotrauma. Ruptured ear drums and deafness are particularly common.

If a diver is caught in the water where an explosion is inevitable, some protection can be afforded by attempting to float on his back, on the surface – this will remove some of the air containing tissues from contact with the water.

Chapter 33

UNCONSCIOUSNESS IN DIVERS

There are many causes of a diver losing consciousness in the water but the final outcome is very often the same – **drowning**. This chapter provides an overview of the underlying causes and basic treatment.

Unconsciousness on land rarely leads to death. Underwater, it frequently does. Because of the hazardous nature of diving in a state of impaired consciousness, great care must be applied to ensuring divers are medically fit and have no increased propensity to loss of consciousness. Also, once consciousness is lost, the adherence to a genuine "buddy system" is of demonstrable value.

When an unconscious diver is rescued and the first-aid measures necessary for all these cases are then instituted (see Chapter 39, 40 and 42), the remainder of the management depends on the cause of the unconsciousness. It is therefore important to be able to identify the likely causes.

They are best classified according to the type of diving being performed, and the equipment used. More information can be found on each topic elsewhere in this book.

CAUSES OF LOSS OF CONSCIOUSNESS

The causes **common to all types of diving** are:

- Hypoxia (from a diversity of causes)
- Salt water aspiration or near drowning
- Cold
- Marine animal injuries
- Vomiting and inhalation of vomit or sea-water
- Underwater explosions
- Miscellaneous medical conditions

In addition to these the causes associated with **scuba diving** are:

- Hypocapnea
- Decompression sickness
- Air embolism from pulmonary barotrauma
- Nitrogen narcosis
- Carbon monoxide toxicity
- Hypoxia due to faulty equipment or gas contamination

In addition to these, causes associated with **rebreathing** or **mixed gas diving equipment** (not commonly used by recreational divers) include :

- Hypercapnea
- Oxygen toxicity
- Hypoxia due to ascent, dilution or excessive consumption.

The more **common causes** of unconsciousness are as follows:

HYPOXIA

(SEE CHAPTER 20)

Hypoxia of the brain associated with near-drowning is the final event in many diving accidents and is the most common cause of unconsciousness in divers. It may follow events as diverse as breath-holding in free diving, inadequate air supply, salt water aspiration, equipment faults or misuse, inhalation of vomit, pulmonary barotrauma, gas contamination, etc. It is frequently associated with panic and physical exhaustion.

Case History 33.1 A diver breathing from a semi-closed breathing apparatus lost consciousness shortly after leaving the surface. He was brought back to the surface and revived with 100% oxygen. His slightly bluish face turned red later on when it was discovered that he had filled his cylinders with pure nitrogen.

Diagnosis: Hypoxia due to inadequate (i.e. nil) inspired oxygen.

HYPOXIA DURING BREATHHOLD DIVING

(SEE CHAPTER 4)

❑ Hyperventilation.

Hyperventilation before a breath-hold dive makes hypoxia more likely because the urge to breathe is suppressed. This technique was used by some divers to increase the duration of dives. The practice is gradually dying out, along with the divers who use it.

❑ Hypoxia of ascent.

A diver can lose consciousness from hypoxia of ascent during deep breath-hold dives. The partial pressure of O_2 in the lungs can fall dangerously low during ascent or immediately after surfacing.

Case Report 33.2. A young diver was attempting to beat a local swimming pool underwater distance record. He was seen to hyperventilate before the dive. Midway through the second lap, he ceased to swim and sank to the bottom. Luckily, he was quickly pulled from the water and revived by mouth to mouth respiration. He was discharged some weeks later, with permanent brain damage.

Case Report 33.3. An experienced diver taking part in a spear fishing competition was found dead on the bottom with a speared fish nearby. Post-mortem revealed no abnormality apart from drowning. He was known to practise hyperventilation, push himself to the limit and to dive deep.

NEAR DROWNING

(SEE CHAPTER 25)

This is the consequence of many diving accidents. The hypoxia associated with near drowning can render a diver unconscious, or the diver can become unconscious first, and then drown.

COLD OR HYPOTHERMIA

(SEE CHAPTERS 3, 27 AND 35)

Exposure to cold water can cause an progressive and excessive fall in body temperature which can make a diver initially confused (at a body temperature of around 34°C) and then unconscious (below 30°C). A diver suddenly entering cold water can sometimes develop a hypertensive spike or cardiac rhythm disturbances which can produce immediate unconsciousness from cardiac or cerebral accidents (heart attack or stroke).

MARINE ANIMAL INJURIES

(SEE CHAPTER 29)

Venomous animals can cause unconsciousness either from the direct effect of the venom on the brain, from hypoxia due to respiratory paralysis, or due to inadequate cerebral circulation from a lowering of blood pressure. Shock from blood loss after shark attack can also cause unconsciousness.

Case History 33.4. A group of divers on their first dive on a tropical reef eagerly took to the water. They returned later with speared fish, coral and shells among which were several varieties of venomous cone shells which they had handled, and in some cases, carried under their wet suits. A member of the boat crew recognised the cone shells and advised the divers of their narrow escape.

Diagnosis: potential loss of consciousness, or "accidents looking for somewhere to happen".

DECOMPRESSION SICKNESS

(SEE CHAPTERS 14–16)

Cerebral Decompression Sickness can lead to unconsciousness. It is more likely after deep dives and repetitive diving.

Case Report 33.5. A 26 year old diver undertook a 55 metre (180 foot) dive with a bottom time of 8 minutes. He returned rapidly to the surface when his contents gauge indicated an almost exhausted air supply. While climbing into the boat, he complained of numbness down one side, and developed slurred speech. He then had a convulsion and lost consciousness. He died en route to a recompression chamber 600 km. away.

Diagnosis: the probable diagnoses would include cerebral decompression sickness or air embolism (CAGE) from pulmonary barotrauma (burst lung). Autopsy verified the diagnosis.

AIR EMBOLISM FROM PULMONARY BAROTRAUMA OF ASCENT

(SEE CHAPTER 11)

Cerebral arterial gas (air) embolism (CAGE) can arise either during or soon after ascent, from any depth. It can cause abrupt loss of consciousness. It is sometimes associated with pneumothorax, which needs special management.

CARBON MONOXIDE TOXICITY

(SEE CHAPTER 23)

Divers breathing compressed air are vulnerable to this problem if the air source is contaminated, often from the exhaust of a nearby internal combustion engine.

OXYGEN TOXICITY

(SEE CHAPTER 21)

Military divers using oxygen equipment, technical or professional divers breathing mixed gases are at risk from convulsions due to oxygen toxicity under certain circumstances. A convulsion can be followed by unconsciousness and confusion.

GENERAL MEDICAL CONDITIONS

A variety of medical emergencies including hypoglycaemia (low blood sugar) in diabetics, heart attack, stroke, epileptic fit, drug overdose, head injury, severe infection and shock can cause unconsciousness. Many conditions with the potential to cause unconsciousness require exclusion in a diving medical examination and people with these conditions would normally be advised against diving.

Case Report 33.6. A 30 year old diver using scuba at 10 metres (33 foot) became unconscious ten minutes after the start of the dive while swimming strenuously on the bottom. He was brought rapidly to the surface by his buddy. He remained unconscious on the boat and was pale and sweaty with a rapid pulse. He was breathing adequately and 100% Oxygen breathing on a mask produced no improvement. Those present were at a loss for a diagnosis until the diver's wife informed them that he was a diabetic taking insulin. He was successfully treated in hospital by intravenous glucose.

Diagnosis: Hypoglycaemia (low blood sugar level) due to unexpected exertion, even though the diver took a reduced Insulin dosage. He was advised that diving and diabetes requiring medication was a suicidal combination and he agreed to take up a sport less dangerous to diabetics

RESCUE AND FIRST AID TREATMENT

(SEE CHAPTERS 39–42)

First **ditch the diver's weights.** The **diver must be brought to the surface and then removed from the water** as **rapidly** as possible. Emergency buoyant ascent takes precedence over concern for burst lung (pulmonary barotrauma) when the recreational diver is unconscious. These unconscious divers will usually exhale passively during ascent as the lung volume expands.

After surfacing the airway must be cleared and if there is no respiration, expired air resuscitation commenced. Details of the rescue and resuscitation technique are explained in Chapters 39 and 42. After securing the essential **basic life support** (DRABCD - rescuing the diver from the **dangerous** underwater environment, verifying the state of unconsciousness by the diver's **responses**, attending to the diver's **airway**, **breathing** and **circulation**) the cause must be sought and **specific treatment** started. Diving medical advice should obviously be sought as soon as possible.

Diagnosis of the cause is made by a logical process of elimination, taking into account the medical history of the diver, the equipment used, the type of diving, the dive profile, the events leading to the unconsciousness and the appearance of the diver. Assume the most serious and treatable diagnosis.

Contact medical authorities (see Appendices A and B) with full details (see Check List Chapter 39) and follow their **advice**. Transport may be required to a medical facility or formal **medevac** may be instituted. Divers may be made worse by transport in certain circumstances (shark attack, decompression sickness, pulmonary barotrauma, etc.) and so this decision is best left to experienced diving physician's cognisant with the prevailing circumstances. If aviation transport is employed, the type of aircraft and its ability to be pressurised, is also to be considered.

While waiting for assistance, administer 100% O_2, if needed, indicated or if in any doubt about the diagnosis.

Keep detailed records and ensure that these accompany the patient. Secure and retain the equipment for future assessment.

Chapter 34

WHY DIVERS DIE

INTRODUCTION

Experience of life suggests that anything which is fun tends to be either illegal, immoral, fattening, or dangerous. Recreational diving partly conforms to this universal law, ranking below hang gliding and parachuting but above most sports as regards the risk of a fatal accident.

Diving statistics from the USA, UK, Canada and Japans all show diving death rates of 15–30 per 100,000 divers per year, with the statistical chance of a fatality being about 2-3 per 100,000 dives.

These figures tend to contradict the misinformation issuing from some sections of the diving industry (fatalities of < 4 per 100,000 divers) which would have us believe that diving is a very safe recreation. It is not, but then we accept risks every day. Even driving an automobile to a dive site carries an appreciable (but much less) risk of death - a possibility which we generally regard with equanimity.

This chapter will show that many diving deaths should be preventable and that a diver ought to be able to minimise his chances of becoming a statistic by understanding and influencing the factors which are now known to be associated with diving deaths.

STATISTICAL EVIDENCE

The information presented here is mainly based on data gathered by valuable studies involving recreational diving fatalities. They have been conducted in different countries, but show strikingly similar results. The USA recreational diving deaths, originally compiled by John McAniff of the University of Rhode Island and then NUADC, are now collected and reported on by DAN, which recently analysed 947 open circuit scuba divers. The DAN survey also included technical divers, who dive deeper, longer and with gases other than

compressed air. The BSAC do a similar job in the UK and DAN-AP Diver Fatality Project is the Australian compiler. Unfortunately significant data is frequently not available and so relevant causal factors are often underestimated. Another Australian approach (the ANZ series of diving fatalities) was to select and analyse only the accidents in which sufficient data was available to make the analysis credible, and to determine what factors materially contributed to the fatality. Most of our statistics come from this source and are rounded up, for simplicity.

OVERVIEW

❑ Diving Fatality Data

- 90% died with their **weight belt on**.
- 86% were **alone** when they died.
- 50% did **not inflate their buoyancy** vest.
- 25% encountered their difficulty first on the **surface**,
  50% actually died on the surface.
- 10% were **under training** when they died.
- 10% were advised that they were **medically unfit** to dive.
- 5% were cave diving.
- 1% of "rescuers" became a victim.

❑ Age.

The recorded deaths range from children (pre-teens) to septuagenarians. Some decades ago the average age of the deceased was in the early 20s. Then there developed a small increase in the middle ages (45-60 years). This bimodal curve has now become distorted on the other side, and the average scuba death age is now 43 years. The reasons for this increasing age of death are:

- The "youngsters" from the 1970-80 scuba diving boom are now older
- Cardiac disease, the sudden death syndrome, affects the elderly and diving introduces more cardiac hazards than many other sporting activities
- Diving is becoming a life-style option for the increasingly active and affluent elderly, with more older people taking up this sport

❑ Gender.

In the 1990s 1 in 10 of the fatalities were women. The actual percentage of women in the overall diving population was about 1 in 3, suggesting that women are safer divers than men. Even now females account for only 20% of the deaths.

❑ Diving Experience.

In most series, 1/3 were inexperienced, 1/3 had moderate experience and 1/3 had considerable experience. The most dangerous dives were the first dive and the first open water dive. In half the cases the victim, based on witness statements and previously logged dives, was extending his diving experience (depth, duration, environment, equipment etc.) and thus did not have the experience to undertake the final dive. For this reason, any diver extending any of his dive parameters (depths, durations, environments, equipment) is advised to do this only with more experienced supervisors.

❑ Major Causes of Death identified at Autopsy.

According to death certificates, most divers ultimately drowned (over 80%), but a number of factors usually combined to incapacitate the diver before this terminal event. Drowning is really only the final act in a sequence of events that lead up to this. It is a reflection of the medium in which the accident happens, more than the accident itself. Often it obscures the real cause of death. Unless there are other factors, drowning should never happen to a scuba diver, as he carries his own personal air supply with him! Drowning develops because of preceding problems, such as cardiac disease, pulmonary barotrauma, the stress disorders, unconsciousness from any cause, salt water aspiration, trauma, equipment difficulties or environmental hazards, etc. These are referred to in the following sections and in other chapters.

CONTRIBUTING FACTORS

Deaths usually followed a combination of difficulties, which alone may have been survivable. The factors contributing to deaths are easier to understand when classified, and we have categorised them into the following groups:

- **Diving Techniques** (Inadequate air supply, buoyancy, buddy system)
- **Human Factors** (medical, physiological, psychological)
- **Equipment Factors** (misuse, faults)
- **Environmental Factors.**

DIVING TECHNIQUES

Inadequate Air Supply

In the ANZ survey in **half the deaths (56%), critical events developed when the diver was either running low or was out-of-air (LOA, OOA).** When equipment was tested following death, few victims had an ample air supply remaining. The DAN survey found 41% in this situation.

Most problems arose when the diver became aware of a low-on-air (LOA) situation. Some divers then died while trying to snorkel on the surface, attempting to conserve air (8%).

Concern about a shortage of air presumably impairs the diver's ability to cope with a second problem developing during the dive, or causes the diver to surface prematurely and in a stressed state of mind, where he is then unable to cope with surface conditions. In many cases the LOA diver faced these difficulties alone, as his buddy who had more air, continued the dive oblivious to the deteriorating situation (see later). LOA situations should be avoidable by adequate dive planning, using a cylinder with ample capacity for the planned dive, and frequent observation of the contents gauge.

A particularly dangerous technique was to intentionally use all the available air (breathing the tank dry). Then there is much less opportunity to cope with unexpected eventualities and greater likelihood of emergency ascent and salt water aspiration. The dive should always be completed with at least 50 ATA remaining.

A description of the methods of preventing and coping with an OOA and LOA situation is in Chapter 11.

Fig. 34.1

In some cases the diver was using a smaller cylinder than a 2000 litre (72 cu.ft) tank. A 1400 litre (50 cu.ft) cylinder has much less endurance than a conventional cylinder, and allows fewer breaths once a LOA situation develops at a significant depth. Also, a diver using a smaller cylinder will usually run out of air sooner, encouraging separation from his group.

Buoyancy

In the ANZ survey, half **the diving victims (52%) encountered buoyancy problems.** Most of these were due to inadequate buoyancy, but some (8%) had excessive buoyancy. The DAN survey buoyancy problems were the commonest adverse event leading to death.

The **buoyancy changes peculiar to wet suits** were a significant factor. The considerable buoyancy offered by a wet suit at the surface needs to be compensated by weights. An approximate formula for this is:

- 1 kg for each 1 mm thickness,
- 1 kg for "long john" extensions and a hood,
- 1 kg for an aluminium tank,
- ± 1–2 kg for individual body variations in buoyancy.

Based on the above formula, **40% of divers who perished were found to be grossly overweighted** at the surface. This factor would have been greater at depth. When weighted according to this formula, a diver should be neutrally buoyant at or near the surface. In this state, descent or ascent are equally easy.

During descent, the wet suit becomes compressed, making the diver negatively buoyant. This is where the **buoyancy compensator (B.C.)** comes in. It is inflated just sufficiently to restore neutral buoyancy. This is why it is called a buoyancy compensator.

Evidently, **some divers deliberately overweighted on the surface**, using this excess weight to descend more easily and were then using the B.C. to maintain depth and then later to return to the surface. This places excessive reliance on the B.C.. This dangerous practice is unfortunately promoted by some instructors. It has advantages from a commercial point of

view, as it expedites training. Groups of divers can be quickly taught to descend with minimum skill. The technique is less advantageous in terms of longevity of the diver.

In another fatality survey on buddied divers who ran into LOA/OOA situations, it was of interest that irrespective of who became OOA first, the overweighted diver was the one who died – at a 6:1 ratio. See Chapter 5, dealing with weights, buoyancy compensators, etc.

In spite of being heavily reliant on their B.C., many divers then misused them**.** Examples of this include accidental inflation or over-inflation causing rocket like ascents ("Polaris missile effect"), confusion between the inflation and dump valves, and inadequate or slow inflation due to being deep or LOA. The drag induced by the inflated B.C. (needed in many cases to offset the non-discarded weight belt) was a factor contributing to exhaustion in divers attempting to swim to safety on the surface.

There are other unpleasant consequences of buoyancy problems. The American Academy of Underwater Sciences, in a symposium in 1989, reported that **half the cases of decompression sickness** were **related to loss of buoyancy control.** After acquiring the initial open-water certificate, possibly the best course to undertake would be on buoyancy control.

Ditching of Weights

This was omitted by most victims (90%). This compelled them to swim towards safety carrying many kilos of unnecessary weight, and made staying on the surface very difficult in these cases. This critical and avoidable factor should be easily remedied by restoring the traditional weight belt ditching drills.

Earlier diving instructors taught that the weight belt was the last item put on, the first taken off. It was to be removed and held at arm's length in the event of a potential problem. The diver then had the option of voluntarily dropping the belt if the situation deteriorated, or replacing it if the problem resolved. When problems did develop, the belt was dropped automatically! Some current diving students now question the validity of dropping these lead (? dead) belts – perhaps the high cost of replacement is worth more than their lives. "Lead poisoning" is a frequent contribution to fatalities.

When ditched, the belt is held at arms length to avoid falling and fouling on other equipment. This entanglement occurred in some of the reported fatalities. In other cases, the belt could not be released because it was worn under other equipment (e.g. B.C., backpack harness, scuba cylinder etc.), or the release buckle was inaccessible because a weight had slid over it, or it had rotated to the back of the body. In some cases the belt strap was too long to slide through the release buckle. Other fatalities have occurred where release mechanisms have failed, due to the use of knotted belts (which could not be untied), or lead balls contained within a backpack.

In an emergency requiring either ascent or buoyancy, to keep the diver afloat on the surface, several kilograms of flotation are immediately available by simply discarding the weight belt. This action also results in a more consistent, controlled ascent than with an inflated B.C..

Buddy Diving System

The value and desirability of the buddy system is universally accepted in the recreational diving community. Two maxims have arisen in diving folklore from this concept:

- "Dive alone – die alone"
- "Buddies who are not in constant and direct communication are not buddies,

— merely diving in the same ocean".

In spite of this, **only 14% of divers who perished still had their buddy with them,** and in the Hawaiian series it was 19%. In 33% of the ANZ cases, the deceased diver either dived alone or voluntarily separated from his buddy beforehand, 25% left their buddy after a problem developed, and 20% became separated by the problem. Of those who started diving with a buddy in the DAN series, 57% were separated at the time of death.

A common cause of separation was one diver (the subsequent casualty) having inadequate air, OOA or LOA. In this case, the buddy often continued the dive alone, or accompanied the victim to the surface, before abandoning him and continuing the dive.

There were many misapplications of the buddy system. In some cases more than two divers 'buddied' together, leading to confusion as to who was responsible for whom. A particular variant of this is a training technique in which a group of inexperienced divers follows a dive leader. When one becomes LOA, he is paired with another (usually another inexperienced diver) in the same situation, and the two instructed to return to the surface together. Often the heaviest air consumers are the least experienced and are over-breathing through anxiety. Two such inexperienced, anxious divers, both critically low on air, are then abandoned underwater by the dive leader and left to fend for themselves!

In others, the buddy was leading the victim and therefore not immediately aware of the problem. Generally, the more experienced diver took the lead, affording him the luxury of constant observation by his buddy, while he gave intermittent attention in return. In this situation, unless a "**buddy line**" is used, the following diver (upon developing a problem such as LOA or OOA) has to expend precious time and energy and air, catching his buddy to inform him of the difficulty. Often this was impossible, and the first indication the leading diver had of the problem was the absence of his buddy, who by this time was unconscious on the sea bed or well on the way to the surface.

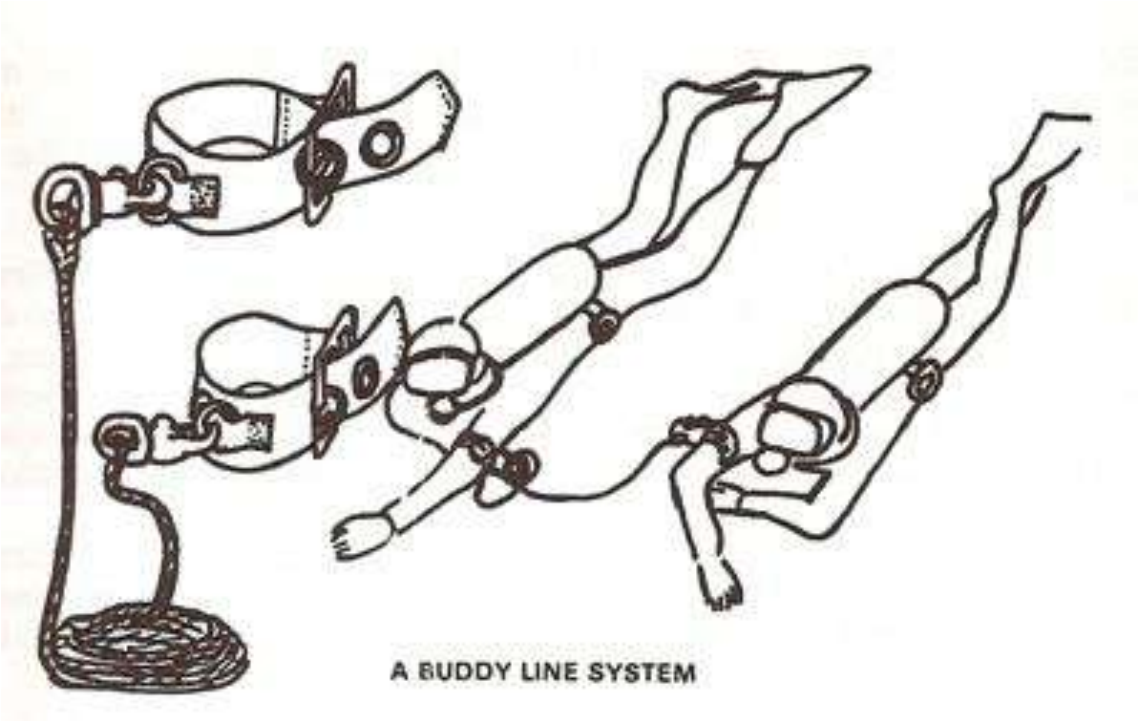

Fig. 34.2
A buddy line may be life saving

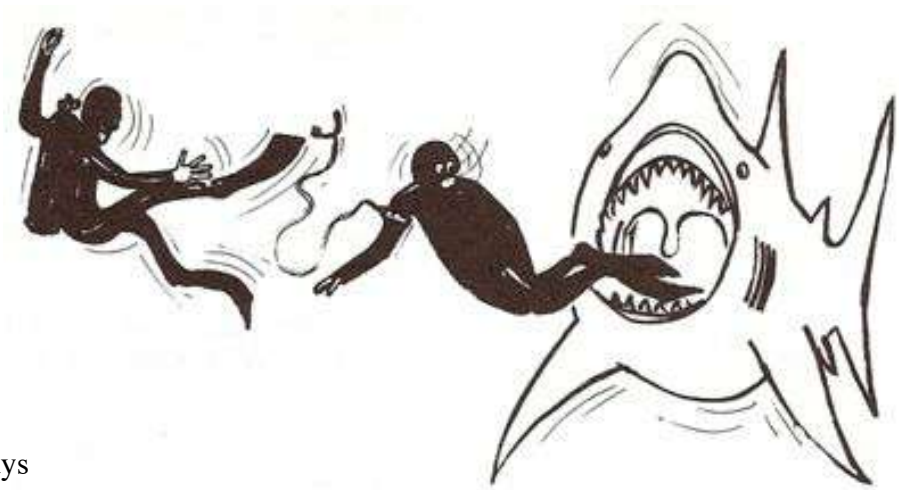

Fig 34.3
But not always

❑ Buddy rescue.

In only a minority of cases was the buddy present at the time of death. Most divers ultimately died alone, usually because of poor compliance with the principles of buddy diving. In only 1% of cases did the buddy die attempting rescue, indicating that adherence to the buddy principle is reasonably safe for the would-be rescuer.

❑ Buddy breathing.

4% of fatalities were associated with failed buddy breathing. In a study of failed buddy breathing conducted by NUADC, more than half were attempted at depths greater than 20 metres. In 29% the victim's mask was displaced and the catastrophe of air embolism occurred in 12.5% of cases.

One in 8 victims refused to return the demand valve, presumably to the righteous indignation of the donor. In one reported instance, knives were drawn to settle the dispute! Nevertheless, donating a regulator rarely results in the donor becoming the victim.

The use of an **octopus rig** or (more sensibly) a complete **separate emergency air supply** (e.g. "Spare Air") would appear to be a more satisfactory alternative, having the added advantage of providing a spare regulator for the owner in the (not so rare) event of a failure of the primary air supply.

HUMAN FACTORS MEDICAL, PSYCHOLOGICAL AND PHYSIOLOGICAL

In at least **25% of cases, the diver had a pre-existing disease which should have excluded him from diving** (compared to 8-10% in the potential diver trainee population). The diseases either killed the diver or predisposed him to the diving accident.

In assessing the cause of scuba fatalities, it is too easy to ignore the disorders which have no demonstrable pathology, such as panic and fatigue, but to do so results in less understanding of the incident. Drowning obscures many other pathologies and some, such as asthma or the sudden death syndrome, may not show up at autopsy.

Panic

39% of deaths were associated with panic. Panic is a psychological stress reaction of extreme anxiety, characterised by frenzied and irrational behaviour. It is an unhelpful response which reduces the chance of survival. This topic is covered in detail in Chapter 7.

Evidence of panic was derived from witness accounts of the diver's behaviour, in the Australasian series. Other studies suggest a 40–60% incidence of panic.

Panic was usually precipitated when the diver was confronted by unfamiliar or threatening circumstances such as LOA, OOA, poor visibility, turbulent water, unaccustomed depth, buoyancy problems (usually insufficient buoyancy), or separation from diving companions.

After panicking, the diver frequently behaved inappropriately by actions such as failure to ditch weights or inflate the B.C., rapid ascent, or abandoning essential equipment such as the mask, snorkel and regulator.

Fatigue

In 28% of cases fatigue was a factor. Fatigue is a consequence of excessive exertion, and limits the diver's capacity for survival. Physical unfitness aggravates it.

It commonly arose from a variety of circumstances including attempting to remain on the surface while overweighted, long swims in adverse sea conditions or swimming with excessive drag from an inflated B.C..

The fatigue factor was not restricted to unfit divers — under special circumstances any diver will become fatigued. In some cases the fatigue was associated with salt water aspiration syndrome, cardiac complications or asthma.

Salt Water Aspiration

This factor was present in 37% of cases. It refers to inhalation of small amounts of sea water by the conscious diver.

In many cases this was the result of; a leaking regulator, aspiration on the surface after removing the regulator, and buddy breathing. In most cases salt water aspiration was a pre-terminal event as the situation became critical. It frequently predisposed to the development of panic, fatigue, respiratory and other complications.

Pulmonary Barotrauma

13% of deaths had autopsy evidence of pulmonary barotrauma (burst lung)**.** In some cases it was a complicating factor rather than the initial cause. Factors promoting the barotrauma were diverse, including panic, rapid buoyant ascents, asthma and regulator failure. Half the cases had an identified cause for the illness. The other half were unexplained.

Cardiac (Sudden Death Syndrome)

In these cases there was either gross cardiac pathology or a clinical indication of cardiac disease (See Chapter 35). In the DAN series, 26% of deaths were due to this. Of the cardiac deaths, 60% complained of chest pain, dyspnoea or feeling unwell before or during the dive.

Victims tend to be older – cardiac causes explain 45% of the scuba deaths in those over 40 years. They tend to be more experienced divers, often with a history of known cardiac disease (arrhythmias or ischaemia) or high blood pressure - often under control with medication (especially beta blockers).

They usually die quietly and the pathophysiology is probably a cardiac arrhythmia (ventricular fibrillation). Resuscitation is difficult or impossible under these environmental conditions. The trigger factors producing this very rapid ineffective heart beat include the following; exercise, drugs, hypoxia from salt water aspiration, respiratory abnormalities from breathing under dysbaric conditions through a regulator and with restrictive clothing and harness, cardio-pulmonary reflexes and cold exposure.

Asthma

In at least 9% of deaths the diver was asthmatic in the ANZ survey, and in at least 8% of cases asthma contributed to the death. In some other surveys (especially those with less data on each fatality, or those that do not specifically check the previous medical history), this data is not so obvious.

Asthmatics should normally be excluded by a competent medical examination. Even so, surveys have shown that between 0.5 and 1% of divers are current asthmatics. When this figure is contrasted with the 9% of fatalities who have the condition, it implies that **asthma is a significant risk factor.**

There was often a series of adverse contributors to death in this group, including panic, fatigue and salt water aspiration. The ultimate pathology was usually drowning or pulmonary barotrauma.

The risk of pulmonary barotrauma is predictable, considering that asthma narrows and obstructs airways. Added to this is the possibility of an incapacitating asthmatic attack during the dive. A considerable number of divers in the survey died this way, some as they were returning to get their medication (aerosol inhalers). Others took it before the dive!

The diving environment can aggravate asthma in several ways:

❑ Salt water aspiration.

Respiratory physicians use nebulised salt water to provoke an asthmatic attack in cases of questionable asthma. Divers immerse themselves in such a solution and often breathe a fine mist of seawater through regulators.

❑ Cold dry air.

Breathing this air precipitates attacks in some asthmatics. Divers breathe this type of air continuously. It is carefully dried by the filling station before being used to fill scuba tanks, and cools as it expands in the regulator.

❑ Exertion.

This aggravates many attacks. Even the most routine dive can require unexpected and extreme exertion, due to adverse environmental factors such as rough water or currents.

❑ Hyperventilation.

The effects of anxiety cause hyperventilation and changes in respiratory gases. This will have little effect on normal lungs. It provokes asthma in those susceptible.

❑ Breathing against a resistance.

Many of the cases first notice problems at depth, where the air is more dense, or if there is increased resistance in the regulator – such as with a LOA or OOA situation.

A study from Denver showed that although normal divers did not show any change in respiratory function with exercise or breathing through scuba regulators, asthmatics had decreases of 15% and 27% respectively.

Vomiting

Apart from the cases that vomited during resuscitation – and there were many – in 10% vomiting initiated or contributed to the accident. It was often produced by sea sickness or salt water aspiration, but ear problems and alcohol over-ingestion also contributed.

Nitrogen Narcosis

This was an effect of depth, and contributed in 9%, but was never the sole cause of death in the ANZ series.

Respiratory Disease

A further 7% of casualties had chronic bronchitis, pleural adhesions, chest injury or other respiratory conditions. Because divers with these conditions are in a minority, they appear to be over represented in the deaths.

Drugs

Alcohol and cannabis (marijuana) are well known contributors to drowning. Cocaine is an established cause of sudden death in athletes. What surprised us was the apparent association between drugs taken for hypertension and the deaths from the sudden death syndrome. Anti-asthma drugs seemed to have the same association.

Decompression Sickness

The dread of DCS is prominent in the minds of most divers. Perhaps this is why there are no deaths due to this condition in the ANZ studies, and less than 1% in the NUADC. Hawaiians reached 4%, due to deep diving for black coral. The DAN survey has 2.5%, probably because of the inclusion of technical divers, who often dive deeper – the mean depth being 68 metres (226 ft) in that study.

While DCS is an important cause of serious disability (such as paraplegia) in all divers, it is not a frequent cause of mortality in recreational divers. This is not, however, true for professionals.

EQUIPMENT PROBLEMS

A significant proportion of deaths were associated with equipment malfunction (35%) or misuse (35%). There was some overlap in the equipment faults and the equipment misuse categories. In spite of the advanced technology available, modern equipment still frequently fails and divers need to be prepared for this possibility (see Chapter 5).

Regulator

In 14% of deaths there was a regulator fault, and in 1% it was misused. Subsequent testing of the regulators showed the majority of problems were due to leakage allowing inhalation of salt water, but in some cases there was excessive breathing resistance following a mechanical dysfunction. In a few cases, the regulator failed catastrophically, or the hose 'blew out'.

The difficulty of obtaining useable air from the regulator was often complicated by other factors such as panic or exhaustion.

Fins

13% of cases lost one or both fins. In some cases this was due to defective or ill fitting fins, but in the majority of instances the cause was not obvious.

A likely explanation is that the fin(s) was lost because of vigorous swimming efforts during attempts to stay afloat with inadequate buoyancy, or during an attempt to swim to safety. Once a fin is lost swimming efficiency is drastically impaired. Panic and fatigue probably had a significant role in these situations.

Buoyancy Compensator

In 8% of cases the B.C. malfunctioned. Usually this was due to failure of the inflation system, but some B.C.s did not remain inflated.

In 6% of deaths, the B.C. was misused. Some divers confused the inflation and dump valves, usually causing over-inflation of the B.C. and precipitating an uncontrolled ascent. Others pressed the wrong button and sank when they wanted to float.

Scuba Cylinder

12% of deaths had problems with the cylinder, usually from misuse. These included under-filling, using a cylinder too small for the dive, the cylinder being dislodged from its harness, and failure to turn on the cylinder valve.

Other Equipment Problems

In 5% or less of deaths, problems were experienced due to failure or misuse of:

- **weight belt** – usually inability to discard it (see Chapter 5)
- **harness** – design faults or covering the weight belt
- **mask** – loss, flooding, and broken straps
- **protective suit** – ill fitting, usually too tight
- **lines** – entanglement
- **gauges** – faulty readings, blow off.

ENVIRONMENTAL PROBLEMS

Environmental factors contributed to 62% of deaths (see Chapter 6).

Deaths near the Surface

25% of the accidents commenced on the surface, and 50% of the divers died at the surface. This may seem surprising as most divers would regard the surface as a safety zone. In many cases they were compelled to surface because of exhaustion of the air supply.

Turbulent (White) Water

Difficult water conditions caused problems in 36%. These included excessive **current, rough** water, **surf** and **surge** around rocks, **underwater surge** from wave movement, and impaired **visibility** caused by these conditions.

These unfavourable conditions often assailed the diver who was forced prematurely to the surface, OOA or LOA, and who was also frequently overweighted and hampered by the drag of his inflated B.C.. Exhaustion or panic then resulted in drowning.

Depth

Excessive depth was a factor in 12%. Often the fatal dive was the deepest ever for the victim. Deep water is a more gloomy and dangerous environment.

The dangers of excessive depth are predictable. They include; increased air consumption, impaired judgment from nitrogen narcosis, colder water, reduced visibility, slow or failed response to B.C. inflation, excessive air consumption, resistance to breathing, and a prolonged ascent in the event of problems.

Other Environmental Problems

Factors which contributed to less than 10% of fatalities included:

- **cave** dives – sometimes causing multiple deaths
- **marine animal injury** – including shark and other animal bites, marine stings (3–6%)
- **difficulties entering** and **exiting** the water
- **cold**
- **entanglements** with ropes, lines and kelp
- **entrapment** – under caves, ledges, or boats
- **night** diving.

DEATHS IN PROFESSIONAL DIVERS

Professional divers have a much higher death rate than recreational divers, especially when operating from deep sea oil rigs. Death rates up to 4.8 per thousand divers per year have been reported, however recent figures from the U.K. indicate that the current deaths in professional divers is now approaching that of recreational divers (30/100,000 per year). Unfortunately, technical divers, have taken over the high fatality rate previously claimed by professionals..

The causes of death differ from recreational divers. DCS and CAGE accounted for up to 28% of deaths. These divers not only frequently develop DCS, but sometimes die from the disease.

Because of the inhospitable environment in areas like the North Sea, cold and heavy seas were a significant factor in deaths, as was increased depth and duration in technical diving. Enclosed diving, such as in caves, salvage, wrecks, and under ice, were also hazards to which the recreational diver is not usually exposed.

Other important factors were equipment failure (saturation divers are highly dependent on equipment integrity for their survival), and the use of more complex gas mixtures and equipment.

Surprisingly, in spite of legislation requiring careful medical supervision, 6% of deaths had a contributing medical factor.

SUMMARY

Diving fatalities generally arise from a combination of factors, none of which alone would have caused disaster.

The contributing factors show an emerging pattern which needs to be addressed by diver education and training. For example, the majority of deaths were in divers who were medically unfit to dive or had a **LOA or OOA element.**

Competent and repeated **dive medical examinations** are essential. **Diver training** and re-training should result in proper planning, buoyancy control and air supply monitoring. Most of the deaths in recreational divers were preventable.

Case Report 34.1 A composite diving fatality might unfold as follows:

A young, inexperienced, slightly overconfident, indifferently trained, male diver undertakes a dive in open water under conditions with which he is relatively unfamiliar. He is healthy but does no regular exercise apart from occasional diving. He has a vague dive plan which he does not discuss with his equally casual buddy. He is mildly anxious because of the unfamiliar conditions. He follows his usual practice of using a generous number of weights, initially inflating and then deflating his B.C. on the surface, to allow his weights to help him descend. Fascination with the environment leads him and his buddy to descend to 40 metres, deeper than they originally intended. He checks his contents gauge and is alarmed to find he is close to his reserve. His anxiety is increased by the realisation that there may be a decompression requirement for this dive, but he may have insufficient air to complete even a safety stop. He is unsure of the decompression requirement, if any, and he did not bring any tables with him. He had not chosen any of the more conservative options on his decompression meter.

He activates the inflation valve on his B.C. but gets so little response that he swims for the surface. He heads for the surface alone with some urgency, unable to communicate with his buddy who is some distance away and preoccupied with other marine life. His air supply runs out during the ascent and he arrives at the surface in a state of panic.

He has extreme difficulty staying afloat but in his frenzied state, neglects to ditch his weight belt or orally inflate his B.C.. His predicament is aggravated by inhalation of sea water and the loss of one of his fins. He becomes exhausted trying to remain on the surface, because of his negative buoyancy and reduced propulsion.

A search team later found his body on the bottom – directly below where he surfaced. They have difficulty in surfacing the body, until they release the weight belt.

The most significant factors in recreational diving fatalities are:

- **diving with disqualifying medical conditions**
- **stress responses -panic and fatigue**
- **salt water aspiration**
- **environmental water movement**
- **buoyancy problems**
- **inadequate air supply - LOA or OOA**
- **adverse sea conditions**
- **failure to ditch the weight belt when in difficulty**
- **ignoring or misapplying the buddy system**
- **improper use of equipment**
- **failure of equipment.**

PREVENTION

Many of the factors associated with diving deaths are avoidable.

Contributing **medical** factors should be largely preventable by **adequate dive medical examinations** prior to commencing diving and periodically after that, especially after diving and non-diving illnesses. With increasing age, the examinations should be more frequent.
As a rule of thumb, routine medical examinations should be at least every 5 years when young, reducing to annual with old divers.

Some changes in the emphasis of diver instruction, aimed at better education concerning the high risk areas of diving, would be helpful. Divers who may be knowledgeable of decompression theory and practice, are running out of air and drowning in solitude, with their excessively laden weight belts still firmly attached.

Before diving is attempted, aquatic skills including unassisted swimming and snorkeling should be acquired. Scuba training should be from professional diving instructors, both for the initial open water training and subsequent courses on buoyancy control, rescue and advanced diving. Supervision is needed while extending diving activities involving different environments, equipment or dive parameters.

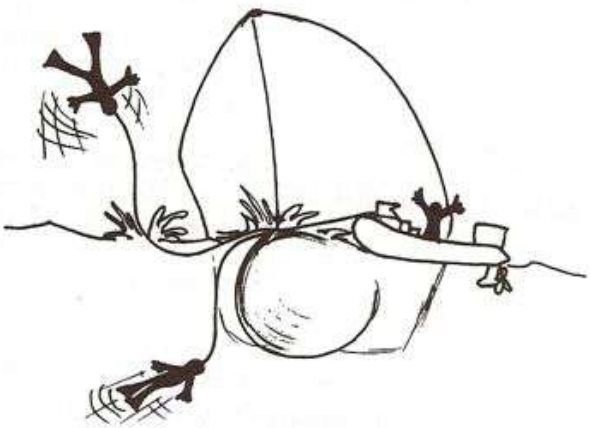

Fig 34.4

Chapter 35

SUDDEN DEATH SYNDROME (Cardiac Death)

PATHOLOGY

Sudden death in divers, especially middle aged divers, is not a rare event. The usual cause is cardiac — either a fatal disturbance of cardiac rhythm (arrhythmia), heart muscle death from a blockage of a diseased coronary artery (coronary occlusion causing ischaemia, myocardial infarction or "heart attack"), or a disease of the heart muscle itself (myocarditis, cardiomyopathy).

Statistical studies on deaths in diving show a disturbingly high incidence of death attributed to heart disease, and rising. In the DAN series, 26% of deaths were cardiac and they contributed to 45% of the scuba deaths in those over 40 years. They tend to be in more experienced divers, associated with known cardiac history, arrhythmias or ischaemia, hypertension or the use of cardio-active drugs.

Cardiac Arrhythmias

The heart normally beats in an orderly and regular way (see Chapter 3). The atria contract, first propelling blood into the ventricles which then in turn contract, ejecting blood into the major arteries.

If this rhythmic contraction is disturbed (an arrhythmia or "irregular heart beat"), the efficiency of cardiac function is impaired and the heart has to work harder, requiring more oxygen and blood flow of its own. Impaired efficiency may also cause lowered blood pressure, which can reduce blood flow to the brain, causing unconsciousness. The

arrhythmia which causes sudden death is called **ventricular fibrillation,** and this usually results in unconsciousness within a few seconds and death within a few minutes.

Severe arrhythmias tend to occur in divers who already have less severe arrhythmias, those with cardiac ischaemia, cardiomyopathies or who take cardio-active drugs (such as for hypertension, asthma, stimulants etc.)

Coronary Artery Disease

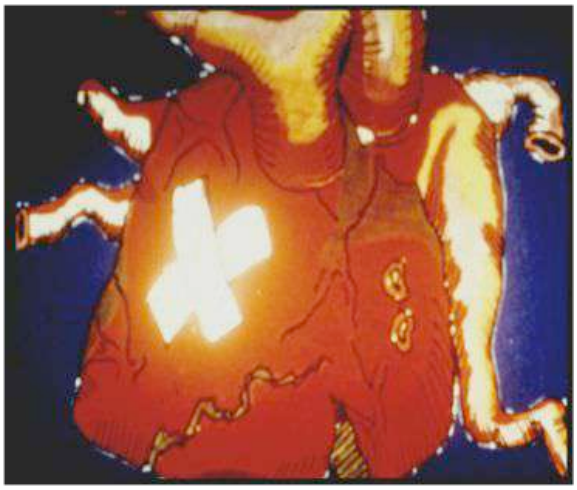

The heart receives its own blood supply from the coronary arteries. Its requirement for blood increases when it has to perform more work, for example during exercise. For a given level of exercise the heart has to work even harder if the blood pressure is elevated, or if the heart has to beat too rapidly, or if the resistance to blood flow is increased. Arrhythmias also increase the cardiac workload.

Fig.35.1

The heart is less able to cope with extra demands for work if the coronary arteries are obstructed, since the blood flow to the heart is reduced. When the coronary arteries do not supply sufficient blood and oxygen to the heart muscles, the latter becomes painful and produces central or left sided chest pain ("angina") or breathlessness (dyspnoea). This may be temporarily remedied by reducing the exercise and the demand for oxygen, by resting. If this deprivation of oxygen to the heart muscle is severe enough, heart muscle dies, and this is then called a myocardial infarction ("heart attack"). In divers, the first sign of this may be at autopsy.

Heart Muscle Disease

Some forms of heart muscle disease (cardiomyopathy – hereditary, alcoholic, or toxic) may affect its function and can occur at all ages. In non-divers who become aware of these diseases, heart transplants are often the only successful treatments. Divers may be unaware that they have these diseases. Viral infections sometimes involve the heart muscle (myocarditis), often without the patient being aware of this, and these infections predispose to cardiac deaths.

CONTRIBUTING CAUSES

There are a number of ways death or incapacity from cardiac diseases can come about and they are usually precipitated by one or more **trigger factors** encountered while diving. Some of these are:

Exercise

Severe exercise can cause sudden death by a number of mechanisms. Probably the most well known example was the death of the first marathon runner who dropped dead after running from Marathon to Athens to deliver the news of the Greek victory over the Persians. In reality, his death was probably due to heat stroke or heat exhaustion. Usually exercise will cause cardiac deaths only in those with some cardiac disease or malfunction.

A diver is at a disadvantage in some ways during exercise. During exertion on land, the cardiac output increases to meet the metabolic demands of the exercising muscles. In doing this, the work of the heart is made easier by blood vessel dilatation in the peripheral circulation, reducing the resistance to blood flow. In an exercising diver however, the skin blood vessels do not dilate because they are trying to conserve heat in response to the surrounding cold water. The diver's heart has to pump against an increased resistance and so work harder for a given amount of exercise, compared to a land athlete.

One of the limitations to exercise on land is the inability to disperse the metabolic heat of exercise. With the diver, much of this heat is conducted away by the water. As a result, it is possible to exercise in the water to a greater degree without the "hot and sweaty" discomfort.

It is therefore possible to exercise to a great degree in the water, with less discomfort but at a greater strain on the heart. In a trained athlete with a healthy heart this probably is only of academic interest. In a middle aged (i.e. over 40 years) diver with some degree of coronary artery disease ("narrowing of the arteries") the diver can overload the heart without realising it. This can result in sudden death.

Exercise, even in fit healthy divers, has been shown to cause significant arrhythmias with diving. It is much more likely to cause incapacitating or fatal arrhythmias in divers with cardiac disease.

Psychological and Personality Factors

Some personalities are more susceptible to cardiac disease than others. The so called **Type A** or **Type D personality** is believed to be most prone to cardiac disease. These individuals are intensely competitive, aggressive and as a result, by society standards, usually successful. They drive themselves hard and do not give up. They are twice as likely to develop coronary artery disease than others, and when they develop it they are likely to push their diseased heart beyond its limitations.

Traditionally this has been a **male personality trait**, but in a more competitive and equal society it is probable that a similar disease pattern will emerge in women competing in previously male dominated areas. Sudden death is not uncommon in Type A personalities.

Anxiety can have threatening cardiac consequences. Anxiety typically causes internal release of adrenalin, one of the stress hormones, which stimulates the heart to contract more forcefully, beat faster, and makes it more prone to arrhythmias. A fast beating heart has less time to replenish its own blood supply and becomes relatively starved of blood.

In the peripheral circulation, adrenalin causes constriction of blood vessels to the skin and internal organs, increasing the resistance to blood flow and the work of the heart. The stressed anxious individual thus has a fast beating heart with a poor blood supply which is more prone to arrhythmias and which has to work harder for a given exercise load.

A condition analogous to fainting (known as **vasovagal syncope**) is commonly seen in individually threatening situations such as a blood donation or the receiving of injections. A nervous response through central stimulation of the vagus nerve causes profound slowing of the heart. The end result is inadequate blood pressure and reduced cerebral circulation causing the diver to lose consciousness ("faint").

Cold

Sudden incapacity and death of divers soon after entering cold water has been frequently reported. The body has several immediate responses to cold water which could explain this.

During **cold water immersion** there is an increased sympathetic nervous system activity resulting in the release of adrenalin. This causes the potentially deleterious cardiac effects described above. A greater sympathetic response has been described in individuals who are **not adapted** to cold water exposure **or** who are **unfit.**

Sudden death from **vagal stimulation** associated with the **diving reflex** can occur after immersion of the face in cold water, although it can also be produced by immersion of the trunk in cold water. See below.

Sudden immersion in cold water is thought to be associated with a sudden death syndrome associated with reflex **coronary artery spasm, fatal arrhythmias** or **myocardial infarction.**

Divers will be familiar with the involuntary over breathing which can accompany sudden immersion in cold water or even a cold shower. In experimental animals, and also in man, the heart becomes more prone to arrhythmias caused by the reduction in blood carbon dioxide from this **involuntary hyperventilation**.

Hypothermia also makes the heart more prone to arrhythmias and may combine with some of the other problems mentioned above to cause sudden death.

Reflexes Associated with Diving

❑ The Diving Reflex.

Diving mammals such as whales are able to hold their breath for an hour and attain amazing depths. They are able to do this partly because of the evolution of the dive reflex. When the mammal leaves the surface there is a profound stimulus of the vagus nerve which slows the heart to about a fifth of its normal rate. At the same time, there is intense constriction of the blood supply to the skin and most organs with the exception of the heart, lungs and brain. This conserves oxygen reserves for use by the organs which need it most. The diving mammal maintains a normal blood pressure, but the output and work of the heart is dramatically reduced.

This reflex is present to some extent in humans. When a human is immersed in cold water there is vagal stimulation which slows the heart, as well as sympathetic nervous stimulation which constricts blood vessels to the skin and other organs. Because the reflex is only incompletely developed in man, there is often a rise in blood pressure but minimal or no fall in cardiac output. This increases, rather than reduces, the work of the heart.

The result of this process in man is increased work of the heart as well as the development of cardiac arrhythmias. Studies conducted on the traditional breath-hold Ama who showed an incidence of arrhythmias of 43% in summer, and an even higher incidence in winter.

❑ Carotid Sinus Syndrome.

The Carotid arteries, on each side of the neck, are the main arteries which supply the brain with blood, and these have a pressure sensing organ – the carotid sinus – in their walls at about the level of the larynx. External pressure on these carotid sinuses causes the cardiac control centre of the brain to mistakenly assume that the blood pressure has suddenly risen. This leads to a **reflex slowing of the heart** and **reduced blood pressure**. This can cause faintness or even loss of consciousness.

A similar effect is caused by pressure from the collar of a tight fitting wetsuit or dry suit neck seal, on the carotid sinus. The problem is especially likely in wet suits without a front zip fastener or having tight "crew necks".

In a series of 100 carefully documented diving deaths in Australia, only one case was thought to be due to this carotid sinus syndrome. In other cases however, distressed divers were seen to pull a constricting wetsuit away from the neck. This may have been a response to respiratory difficulty from cardiac causes, a tight neck opening in the suit, or the carotid sinus syndrome.

Hyperbaric Exposure

Studies in experimental subjects breathing air at pressures similar to those experienced by sports divers showed a significant incidence of arrhythmias caused by the hyperbaric exposure. This may be partly due to the elevated partial pressures of oxygen breathed at these depths.

Immersion

Simply immersing the body in water causes an increased return of blood to the heart, due to the change from a gravity influenced circulation to weightlessness. This rush of blood to the heart can rapidly double its workload until stability is returned. In redistributing blood from the periphery to the lungs, there may be a predisposition to pulmonary oedema (see Chapter 32)

With immersion, the effect of gravity on the blood vessels is removed. Sudden exposure to a gravity effect on the blood vessels, such as climbing up the ladder and out of the water, increases the possibility of hypotensive syncope. This is a good reason to not hover underneath divers on ladders.

Aspiration

Aspiration of seawater – always a possibility in diving – can cause immediate cardiac effects by a mechanism akin to the diving reflex (see above), It can also be followed by delayed effects due to hypoxia as the lungs are involved, as in the salt water aspiration syndrome (see Chapter 26).

Drug Effects

A large variety of drugs have arrhythmic and other effects on the heart, which may predispose to sudden death. Many can be purchased 'over-the-counter' in pharmacies or supermarkets. Some are contained in 'cold cures' and 'cough mixtures' and may be inadvertently used by divers. Some of these drugs include:

• Alcohol
• Nicotine – cigarette smoking
• Caffeine – coffee and tea, stimulant drugs to overcome sleepiness
• "Social" drugs such as cocaine, weight reducing and stimulant drugs such as amphetamines
• Blood pressure controlling drugs (e.g. calcium channel blockers, beta blockers)
• Drugs used to suppress arrhythmias (e.g. beta blockers)
• Drugs that change electrolyte concentrations in the blood – diuretics and electrolytes
• Sympathomimetic drugs (e.g. decongestants such as pseudoephedrine, anti-asthma medications such as salbutamol, and some anti-seasick drugs).
• Others that may cause arrhythmias –antidepressants, digoxin, some anti-malarials, local anaesthetics.

Cardiac Disease

❑ Coronary artery disease or CAD.

This heart disease (causing narrowing or obstruction of the coronary arteries), while considered to be a disease of middle and older age, is probably present to some degree even in some young adults.

It would appear from post-mortem studies done during the Korean and Vietnam wars that coronary artery disease begins in early adulthood, but usually only causes symptoms and death from heart attack after 40 years of age. The older the diver, the more significant this is likely to be. Divers with this disease are more prone to sudden death due to arrhythmias or myocardial infarction secondary to impaired blood supply to the heart muscle.

❑ Coronary artery bypass grafts.

Some blockages of the coronary arteries can be bypassed by blood vessel grafts – usually using arteries or veins. This reduces cardiac pain and improves cardiac performance but does not cure the underlying disease, which affects the many coronary arteries not improved with the by-pass. People with such grafts are still more prone to arrhythmias and cardiac dysfunction and should not dive, unless proven to be cured. A similar situation exits in those people who have already suffered a myocardial infarction or "heart attack" and those needing or possessing pacemakers

❑ Myocarditis.

Some viral infections which produce a flu-like illness can temporarily affect the heart muscle, impairing its performance and making it susceptible to arrhythmias. Sudden deaths from this often insidious condition (myocarditis) are occasionally reported in very fit athletes and in divers. It is unwise to dive or perform heavy exertion when suffering from a viral infection, for this reason.

Sometimes the heart is permanently and irreversibly damaged by such viruses. It is then similar to a cardiomyopathy. In terminal cases the only effective treatment is a heart transplant.

PREVENTION

All candidates should be carefully examined by an experienced diving physician before dive training. Those with known cardiac disease or a tendency to arrhythmias cannot dive safely. Middle aged divers, and those with high coronary risk factors need regular assessments.

Coronary risk factors include:

- a family history of heart disease at a similar age to the diver
- cigarette smoking
- hypertension
- cardio-active drugs (see earlier)
- obesity
- high cholesterol – hyperlipidaemia
- physical unfitness
- diseases such as diabetes and alcoholism.

Diving situations often require extreme physical exertion, which stresses the heart. A high standard of physical fitness brought about by regular exercise will improve the hearts ability to cope with this exertion.

While jumping into cold water gets the discomfort over quickly, it maximises the physiological stress. Enter cold water slowly to minimise these physiological stresses.

The combination of performance anxiety, transport stress, inadequate sleep, excessive alcohol, coffee and other drugs which often accompany a "high-living" diving holiday may be possible contributors to cardiac arrhythmias, and some deaths in divers.

Chapter 36

PSYCHOLOGICAL DISORDERS

Psychological Traits of Successful Divers

This topic has not been extensively researched, but the few studies which have been done on the psychological make-up of divers have shown the following — successful divers tend not to be anxious people; they are self sufficient, intelligent and emotionally stable. Their tolerance to stress often allows them to continue to function during difficulties which would incapacitate many non-divers. This may be helped by their tendency to use "denial", a mental mechanism by which they ignore the hazards which may confront them.

In spite of their overall stability, divers, like anyone else, can suffer from psychological disturbances (see Chapter 7 for a full description of stress and panic responses).

Anxiety States

It is quite normal for divers to feel concerned with the very real hazards of diving – and this book is full of them. Some people, however, develop an excessive and inappropriate anxiety to diving hazards which may become a **phobia** – an irrational fear. This may be the result of a previous traumatic event (such as near drowning during childhood) or may be an exaggerated reaction to some diving danger. Phobias may relate to diving in general or to a specific diving hazard (such as an excessive fear of sharks) or a situation (e.g. claustrophobia with the face mask, night diving, poor visibility).

Phobias can be treated by psychological de-conditioning therapy.

Most people who are anxious about diving are aware of this early in their training and quite sensibly desist. Why continue a recreational activity which causes apprehension? Unfortunately some continue because of peer pressure, ego challenge or other personal reasons. These divers tend to have a high baseline level of anxiety (**neuroticism** or "**high trait anxiety**") when diving and are more prone to panic when confronted with real or imagined hazards.

Flora Fisk Quiz for Recreational Divers:
Question: Is recreational diving supposed to be fun? Answer: Yes
Question: Are you having fun? If "yes" — continue diving.
If "no" — STOP DIVING.

Some divers experience a general anxiety reaction under special circumstances. It is an aquatic manifestation of a general medical (psychological) disorder produced by sensory deprivation - called the **Blue Orb Syndrome** by high altitude aviators. It usually happens to a lone diver in deep water, where there are no visual references. The diver develops an anxious feeling of being alone in the vastness of the ocean. This can lead to mounting anxiety and panic. The panicked diver may rush to the surface, omitting decompression or develop pulmonary barotrauma from failure to exhale adequately on ascent. The symptoms usually subside if the diver can establish visual contact with concrete objects such as the sea bed, a dive boat or even another diver, or by concentrating on diving instruments, such as a watch or depth gauge.

This syndrome can be avoided by diving with a buddy who provides reassuring company and a visual reference. Avoidance of deep water where there are no visual references, is also helpful.

Panic

This frenzied and irrational behaviour is the end result of a number of diving difficulties. It is more likely to occur in anxiety prone divers and frequently results in a diving accident or fatality. It is an important topic for divers to understand and is covered in detail in Chapter 7 and Case History 7.1

Psychological Disturbances due to Medical Causes

Brain function can be disturbed by physiological factors (such as **nitrogen narcosis**, **hypothermia**) and by other diving related illnesses.

Cerebral decompression sickness and **air embolism** can cause alteration of brain function during both the acute event and recovery. A diver may act irrationally because of these diseases, and not just his basic personality. Suspect this if he is acting “out of character”.

Near drowning, **hypoxia** and the **gas toxicity diseases** (**oxygen**, **carbon dioxide**, **carbon monoxide**, etc.) may also cause temporary or permanent brain damage.

Symptoms include confusion, irritability and irrational behaviour. This should always be borne in mind if a victim of a diving accident is unreasonably reluctant to undergo treatment. People who know the diver well will normally be the best judges of whether the behaviour is out of character.

Dementia

This is a deterioration of intellectual capacity and memory which is common in the elderly and has a variety of causes. Alzheimer's ("old timers") Disease is a severe form of dementia. Diving folklore holds that divers suffer an increased incidence of dementia. This belief has been supported by media reporting and anecdotal accounts of divers suffering from the condition. Some even believe that only "demented" persons would take up diving!

There are plausible theoretical reasons why divers could sustain brain damage sufficient to cause dementia from conditions such as repeated subclinical, or overt, cerebral **decompression sickness**, **air embolism**, **near drowning** or **carbon monoxide poisoning**, to name a few.

There are also some scientific studies which show evidence of at least transient brain damage in some divers. A study in Sweden showed 3.5% of free ascent trainees to have **EEG** ("electrical brain wave") abnormalities after free ascent training, and in another survey, EEG abnormalities were found in 43% of a group of Polish professional divers compared to 10% in a normal population.

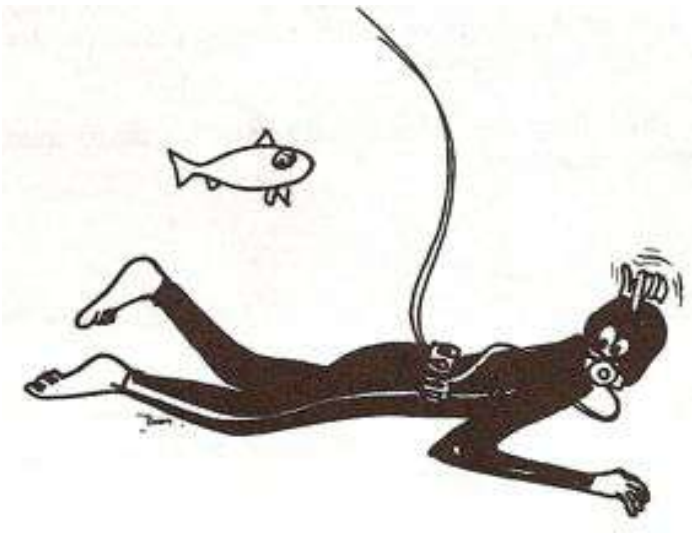

Fig. 36.1

In Australia, a group of divers, studied after treatment for **decompression sickness,** showed **neurological**, **psychological** and **EEG abnormalities** for some weeks after treatment, even in those who had no symptoms of neurological decompression sickness.

There have been several studies worldwide which appear to show deterioration of intellectual performance and psychological disturbances in divers suffering from neurological decompression sickness or "near miss" diving accidents. Unfortunately the methodology of these studies was grossly inadequate, making the conclusions unreliable.

To clear up some of the controversy a study of all 152 professional **abalone divers,** from a closed community, was undertaken in Australia by Edmonds and others in 1988. The divers in the study had diving exposures which would generally be regarded as extreme. On average they had been diving for 16 years and had been professional abalone divers for 12 years. They averaged 5 hours underwater per day on Hookah equipment for 105 days per year at an average depth of 15 metres

(50 ft) and admitted to being "bent" four times. Many other incidents of decompression sickness went unrecognised and untreated. Half of the divers used a dive profile which would, according to conventional dive tables, require decompression but which they omitted. Of the 69 cases of decompression sickness in this group which were actually treated, at least half were neurological.

It would seem that if there was any group of compressed air divers prone to brain injury after excessive diving exposure, it would be this one.

The divers were subjected to a wide battery of tests including **intelligence tests**, **psychometric** investigations to detect psychological abnormalities, **memory tests** and studies designed specifically to detect early dementia, **EEG studies** and **neurobehavioural tests**. The divers were compared with control groups taken from non-diving fishermen in the same localities.

The results showed the divers studied were within the normal range for the general population and displayed no evidence of brain damage or dementia. This implies that air divers, in general, have no greater risk of dementia or brain damage than non-divers. If brain damage does occur, it is either rare or so mild that it could not be detected by conventional testing.

Since the diving practices of this group were extreme, it seems reasonable to conclude that divers following more conservative practices, as well as other, more conventional, professional compressed air divers, have no greater risk of dementia or progressive brain damage than non-divers, unless they suffer a major diving accident affecting the brain (such as those mentioned above).

The signs of brain damage which have been described in studies performed soon after minor diving events are presumably temporary in nature.

Chapter 37

DRUGS AND DIVING

It is common for divers to enter the water under the influence of drugs. These may vary from paracetamol taken for a minor headache, to alcohol or marijuana from a beach party the night before, or a therapeutic drug for an illness such as high blood pressure.

Since some drugs are innocuous while others can have potentially lethal effects with diving, it is important to know something about them.

Problems can arise from effects of the drugs themselves, but commonly the condition for which the medication is taken poses a greater threat to the diver. For instance, most antibiotics have no harmful influences on divers, but a diver being treated for bronchitis with an antibiotic, has a significant risk of developing pulmonary barotrauma until the condition resolves.

We will consider commonly used drugs under four categories:

- Drugs taken for treatment of illnesses
- Drugs taken for prevention of illness (prophylaxis)
- Recreational or social drugs
- Drugs used for diving related illnesses.

TREATMENT DRUGS

In many cases the drug which is used to treat an illness can be a greater hazard to the diver than the illness itself. Some drugs may have frequent and predictable effects on diving activities and these are summarised here, but much more information can be sought by an Internet search.

The idiosyncratic effects of other therapeutic drugs on an individual are unpredictable. Thus, if drugs are to be used when diving, they should be introduced to the diver long before the diving

is undertaken. Then a drug reaction is less likely to complicate the interpretation and treatment of potential diving diseases.

Cardiac and Blood Pressure Medications

❑ Beta-blockers.

A variety of these drugs (e.g. atenolol, metoprolol) are used to treat high blood pressure or pain from coronary artery disease (angina). Their main action is to block the effect of the cardiac stimulant, adrenalin, on the heart. Adrenalin acts on specific drug receptors in the heart known as "beta receptors", thus giving rise to the term beta-blocker.

By inhibiting the action of adrenalin, beta-blockers reduce the force of contraction of the heart muscle. This diminishes the work it performs and so reduces angina symptoms, while the reduced output of blood lowers the blood pressure.

A diver taking beta-blockers has a significant limitation of the reserve pumping capacity of the heart. If a large blood supply is required by the muscles to extract the diver from an emergency e.g. heavy wave action or and adverse current, it may not be available.

Beta-blockers have been incriminated in the production of both arrhythmias and pulmonary oedema in divers.

Beta-blockers also act on the muscle lining of the bronchi and may unmask asthma in some individuals, exposing them to the dangers of both asthma and burst lung. Even those beta-blockers which are described as "cardio-selective" can still have these effects.

These drugs are often used in eye drops for the treatment of glaucoma. Significant amounts can sometimes be absorbed into the body causing generalised effects. Divers using these drops should seek medical advice to ensure that they are not affected in this way. They can avoid this complication by using a eye-drop technique called lacrimal compression.

The conditions for which the drugs are taken can cause difficulties as well. For example, a diver under treatment for high blood pressure is also at high risk of coronary artery disease, and may already have a sub-clinical form of this disease. The drugs may summate with, or potentiate, other causes of heart rate reduction associated with diving, and provoke arrhythmias and the sudden death syndrome (see Chapter 35).

❑ Other blood pressure drugs.

Apart from beta-blockers, blood pressure lowering drugs fall into two broad categories — blood vessel dilators and diuretics (stimulators of urine production).

- **Blood vessel dilators (vasodilators)** reduce blood pressure by widening peripheral blood vessels, where most of the resistance to blood flow occurs. These include prazosin and felodipine.

Some can inhibit the bodies ability to compensate for changes in posture, causing fainting on standing. This is an undesirable side effect in a diver attempting to ascend a ladder to leave the water, especially in adverse sea conditions or if he is preceding his buddy.

A newer drug, the ACE inhibitors, can produce a dry cough which can be troublesome in the diving environment. Others, (such as calcium channel blockers like verapamil) may affect the nerve conduction of the heart, making it more susceptible to the sudden death syndrome (see Chapter 35).

• **Diuretics** stimulate the production of urine and tend to dry the body out. This reduces blood volume and so tends to lower the blood pressure, but there are probably other mechanisms acting as well. The effects on diving are not clear but there are potential problems.

Reduction of the blood volume may affect blood viscosity and the dynamics of blood flow and so increase the dangers with bubble formation. In addition, changes to regional blood flow may alter the pattern of gas uptake and elimination (decompression).

Some diuretics tend to lower the blood potassium level, making the heart more prone to disturbances of rhythm. These arrhythmias may be unmasked by the effects of cold, the dive reflex, heavy exertion and the other causes of the sudden death syndrome (see Chapter 35).

One group of diuretics, carbonic anhydrase inhibitors, is also used as treatment for glaucoma. These cause increased paraesthesia especially in the cold-exposed hands of divers, and may lead to a mistaken diagnosis of decompression sickness.

Psychotropic Drugs

❑ Tranquillisers and sedatives.

This group of drugs includes **benzodiazepines**, of which diazepam ("Valium") is a common example, and sleep making drugs like barbiturates. A significant proportion of the population takes drugs to relieve anxiety. Excessive anxiety alone is a significant risk factor in diving, and the drugs taken to relieve it further complicate the problem.

Another class of tranquillisers, like **phenothiazines**, (e.g. chlorpromazine – "Largactil") and its modern equivalents, are used to treat serious psychiatric disorders such as schizophrenia. Apart from the side effects of these drugs, people suffering from this disorder often have a tenuous grip on reality and this can seriously impair their ability to make safe judgments related to diving.

Tranquillisers and sedatives cause; drowsiness, impaired judgment, slowing of thought processes and reduction in problem solving ability. These effects are intensified by nitrogen narcosis, but they are potentially dangerous at all depths.

❑ Antidepressants.

Depression is not an ideal state of mind for an active diver. Even when successfully treated with antidepressants, the diver is left with potentially harmful side effects from the drugs. Some of the antidepressants cause sedation, but the principal problem is the tendency of others to cause potentially lethal disturbances of heart rhythm and epilepsy. Some can react with certain foods and other drugs to affect blood pressure and consciousness.

❑ Anticonvulsants (anti-epileptic drugs).

These have similar sedative side effects to the tranquillisers as well as some other specific complications. Any form of epilepsy can have disastrous effects on cerebral activity, with loss of consciousness being common. The influence of nitrogen narcosis on these drugs is unknown. Some diving conditions (stress, glare and flickering light, high or low oxygen and carbon dioxide levels) can precipitate convulsions, despite these medications. Epilepsy and medications used for its control (e.g. phenytoin or carbamazepam), preclude safe diving.

Antihistamines

These are usually taken to treat allergic conditions. Pharmacologically, many are closely related to the psychiatric drugs and share a common side effect, sedation. They cause the same potential hazards to diving as other sedatives. In addition, if antihistamines are taken to treat hay fever, there is a strong possibility of the diver developing ear or sinus barotraumas. These drugs seldom completely cure the nose and throat congestion. Other recently developed drugs are less sedative, but may provoke cardiac arrhythmias or bronchospasm (asthma).

Antibiotics

These have a large number of side effects, but few of specific relevance to diving. Tetracyclines can occasionally cause photosensitivity, a condition resembling sunburn caused by enhancing sensitivity to sunlight. Many antibiotics increase susceptibility to **vomiting**.

The condition for which the antibiotic is taken is usually of more concern. This especially applies to respiratory tract infections which make the diver prone to barotraumas.

Analgesics

A diver suffering from pain which warrants the use of pain killers should generally not be diving. Apart from the adverse interactions some diseases can have on diving, the commonly used analgesics can have undesirable side effects. There is also the **diagnostic confusion** that may result between the painful condition aggravated by diving, and decompression sickness.

❑ Aspirin.

This commonly used analgesic causes an inhibition of the clotting ability of blood, with just one dose and lasts for many days.

This is usually not a problem in everyday use – in fact the blood effect is used to prevent heart attacks and strokes. However, it can have implications if the diver develops **inner ear barotrauma** or serious **decompression sickness**. The increased bleeding tendency can result in haemorrhage into injured tissues, such as the spinal cord, with greater consequences. Ulcer-like erosions can also occur in the stomach, with vomiting and occasionally bleeding from the gut

It may also cause bronchospasm, like **asthma**, in some divers.

❑ Paracetamol (acetaminophen).

If a diver needs to take analgesics for minor pain (hopefully after excluding diving related illnesses as a cause), it is better to use paracetamol which has fewer side effects than aspirin. Paracetamol does not effect blood coagulation and avoids the stomach upsets, common with aspirin.

❑ Strong analgesics.

Preparations containing codeine or dextropropoxyphene (both narcotic derivatives) and other strong analgesics are sometimes prescribed for severe pain.

These drugs have comparable sedative effects to the tranquillisers and can have similar adverse interactions with diving. People with pain of this intensity should not be diving.

Insulin and Anti–diabetic Agents

People taking these drugs are prone to sudden depression of the blood sugar level which produces **anxiety**, **confusion** and then **unconsciousness**. This is particularly likely during exercise. This complication in the water often has a fatal outcome. Because of this possibility and other potential physiological complications (e.g. acidosis and hyperventilation), diabetics are generally advised against diving.

Bronchodilators and other Asthma Medications

Asthma is an inflammatory condition of the air passages of the lungs. It causes swelling of the lining of the airways, spasm of the muscles in the airways (bronchospasm) and obstruction of airflow through them. The bronchospasm can be reduced by aerosol sprays containing drugs such as salbutamol ("Ventolin") or oral bronchodilators. These can disturb cardiac rhythm and precipitate the sudden death syndrome during diving, because of the multiple trigger factors (see Chapter 35).

While the use of these and other asthma medications will improve some of the airway flow and thus relieve symptoms, it does not cure the condition. Asthmatics have airways which are excessively sensitive to irritants, reacting with bronchospasm to stimuli such as cold dry air and sea water inhalation. There is usually a degree of obstruction in some of the airways most of the time. This makes them susceptible to pulmonary barotrauma or death from the diving sequelae of asthma - panic and drowning.

Some oral bronchodilators (theophylines) can cause pulmonary vasodilatation – which could potentially allow asymptomatic venous bubbles from normally safe dives to enter the arterial circulation as gas emboli, even without pulmonary barotrauma. **Asthma and these medications are incompatible with safe diving.**

Implanted Drug Delivery Systems

Implanted **reservoirs** are now being used to deliver drugs which cannot be taken orally and which need to be used over prolonged periods. Many of the conditions for which these reservoirs are used are incompatible with scuba diving.

Implants form a potential site of bubble formation during decompression. If bubbles form inside or around the reservoir, expansion in response to the gas laws may lead to excessive delivery of the drug. As experience with these devices in diving is limited, divers fitted with them are advised to seek expert medical advice concerning the possible complications.

PROPHYLACTIC (PREVENTION) DRUGS

Statins

These are frequently used to reduce the harmful effects of cholesterol, but may have other beneficial effects. Patients who need these drugs are likely to have an increased risk of cardiac disease, and so should be assessed more completely by their diving physician.

Some recipients will respond to statins with serious muscular and other disorders and so diving should be delayed for some weeks or months to ensure these problems are not experienced.

Oral Contraceptives — the "Pill"

These preparations can have serious side effects, even without the complication of diving. In older high dosage drugs, excessive blood clotting resulted in occasional deaths from pulmonary embolus or strokes.

Occasionally the pill produces serious psychological sequelae, migraine, nausea and vomiting, which may make diving more hazardous.

The newer low dose oral contraceptives have a lower incidence of these disorders. The concern with diving is the possibility of more congealable blood interacting with gas bubbles during decompression. There is no overwhelming evidence to either confirm or refute this theoretical risk, despite a number of surveys on female divers.

Anti-Sea Sickness Drugs

See Chapter 32.

Antimalarial Drugs

Tropical countries offer some spectacular diving locations but also frequently have endemic diseases, including malaria.

The chances of contracting this potentially lethal disease are reduced by the use of anti-malarial drugs such as chloroquine and pyrimethamine ("Maloprim"). Unfortunately many countries now have strains of malaria which are resistant to conventional anti-malarial drugs, making their use as a preventative measure not fully reliable. As well as being fallible, these drugs can have serious side effects including suppression of white blood cell production, anaemia, and eye damage.

One of the anti-malarials, mefloquine ("Lariam"), can cause coordination disturbances and vertigo which may have alarming implications and cause diagnostic problems for divers.

A diver intending to visit a malaria endemic area should seek expert medical advice regarding prophylaxis for malaria in that area, as well as other more exotic tropical diseases. A diving physician should also be asked about possible interactions of the prescribed drugs with diving.

RECREATIONAL (SOCIAL) DRUGS

Alcohol

Diving culture has traditionally included substantial use of alcohol. Like other drugs, it can have adverse interactions with diving.

There is no safe blood alcohol level for diving and few people in their right mind would consider diving while under its influence. Some may not be aware that the liver has a limited capacity to metabolise this drug, so it is possible to have an appreciable blood alcohol level on the morning after a night of heavy consumption. Traffic police are well aware of this. They frequently apprehend drivers going to work with an illegal blood alcohol level on the morning after.

The danger of alcohol consumption associated with aquatic recreations is well documented — 80% of **drownings** in adult males are associated with alcohol use, in Western countries. The hazards are predictable. Alcohol intoxication, as well as **impairing judgment** and **coordination**, causes **cardiac rhythm disturbances**, **impairs the pumping ability of the heart, reduces the blood volume** due to excessive urine production, and increases **heat loss** through the skin (**hypothermia**).

The disturbed physiology – otherwise known as a "hangover" – after excessive alcohol consumption is well known to the younger authors of this text. Used to excess, this drug is a toxin, damaging the liver, heart and brain. In divers, the vascular and metabolic dysfunction after heavy consumption is a probable risk factor for the development of **decompression sickness**. Increased susceptibility to both **sea sickness** and **vomiting** is frequently observed. Soporific effects may summate with those of **nitrogen narcosis**.

Tobacco

The gentle art of inhaling burnt tobacco leaf has some unfortunate side effects. The associated risks of lung cancer, heart and vascular disease are well known. There are more subtle effects.

A smoker inhales carbon monoxide, which binds to the haemoglobin and reduces the blood's ability to transport oxygen by up to 10%. This reduces the capacity for exertion and impairs the physical ability to respond to an emergency (e.g. fatigue from a surface swim). The nicotine in the tobacco also stimulates the heart, making it prone to **rhythm disturbances** (arrhythmias).

Airway narrowing caused by chronic smoke irritation impairs exercising ability and increases the risk of **pulmonary barotrauma**. A similar chronic irritation of the upper respiratory tract predisposes to **ear** and **sinus** b**arotrauma**.

Marijuana — Cannabis or "pot"

Chronic use of this drug causes many of the diving related respiratory problems attributable to cigarette smoke, and **chronic bronchitis** is especially common in heavy users. This predisposes to **pulmonary barotrauma**.

Marijuana causes altered perception, **impaired judgment**, and mood alterations, which are incompatible with diving safety. As with other drugs, these effects are compounded by the effects of nitrogen narcosis. It also is said to increase the likelihood of **hypothermia** by blocking blood vessel response to cold. The allegedly "beneficial effects" of marijuana are negated by pressure!

Cocaine ("coke") and Other Stimulants

These drugs have similar physiological effects to adrenalin, stimulating and irritating the heart, causing potentially lethal **cardiac rhythm disturbances**, and elevating the blood pressure. Sudden death in young people from the cardiac effects is common, especially in athletes who exercise after taking cocaine.

The mental stimulation and mood elevation **impair judgment** and encourage **risk taking**. Its use while diving, apart from being illegal, is very risky.

Caffeine

This drug is found in coffee, tea, cola, and many natural foods. Even chocolate drinks, periodically given to children at bedtime, contain it. It is one of the more innocuous drugs which is used almost universally.

When used to excess it can irritate the heart causing **rhythm disturbances** which are a potential problem in diving or other strenuous exercise. It also stimulates urine production which discourages some divers from lending their wet suit to known caffeine abusers.

Narcotics

The **sedative** and **judgment impairing** qualities of these drugs makes their use during diving even more dangerous and destructive than their use as a recreational drug.

Intravenous drug users have a significant risk of being infected with the **hepatitis** and **HIV (AIDS)** virus, which should be born in mind by their diving companions (see Chapter 28).

DRUGS FOR DIVING DISEASES

Anti Sea Sickness Drugs

See Chapter 32

Sinus and Ear Problems

Many inexperienced divers have difficulty equalising the ears and sinuses to pressure changes. Often this difficulty is associated with congestion of the lining of the nose, generally due to allergy (hay fever) or infection (URTI). Poor technique is a contributing factor.

Nasal congestion can be relieved to some extent by the use of tablets such as pseudoephedrine ("Sudafed"), or nasal decongestant sprays such as phenylephrine or ephedrine. They all have a **disruptive effect** on the **heart's conduction system** and may thus increase the likelihood of the sudden death syndrome.

These agents used in proper doses on land have few harmful generalised effects. However their activity on the nasal tissues can be unpredictable. Prolonged use causes localised tolerance to the drug, eventually aggravating the congestion which they are supposed to relieve. This applies particularly to nasal sprays. Their effect can wear off during the dive, allowing a trouble free descent, followed by sinus or ear barotrauma during ascent (see Chapters 9 and 10).

These drugs are sometimes used by divers to overcome the temporary nasal congestion of an upper respiratory tract infection (a cold or URTI). A safer approach is to avoid diving during the course of these infections.

If the decongestant is somewhat effective, it may result in the diver avoiding barotrauma of descent (as the beneficial effect is on the nasal lining) but have little or no effect on the "internal" opening of each of the air passages (Eustachian tube, sinus ostia, etc.). Thus the diver is now vulnerable to an internal blockage which manifests during ascent and produces **barotrauma of ascent**. This disorder is far more dangerous than the disease for which he is taking it, as it prevents ascent to safety. Barotrauma of descent merely stops him diving.

Self medication with these drugs is unwise and divers with “congestion” problems should seek the advice of a diving physician.

Medication for use in Decompression Sickness

In view of the relative unreliability of the decompression tables, researchers have experimented with drugs to inhibit the development of bubbles and speed gas elimination from the body. While some experimental drugs now allow laboratory animals to dive with a much greater margin of safety, no agents useful to human divers have yet been convincingly demonstrated.

Chapter 38

MEDICAL EXAMINATIONS FOR DIVERS

This chapter is not adequate to instruct a medical practitioner on the complexities of performing diving medical examinations. Special courses and qualifications are needed for this purpose.

Because of the unique physical and physiological conditions encountered in diving, medical standards for divers differ considerably from those of other sports. As a result it is sometimes necessary for a diving physician to advise a prospective diver against diving because of a disqualifying condition. Sometimes the recipient of this advice is supremely physically fit, and some have been of Olympic standard. These individuals understandably find it difficult to comprehend how a physically fit athlete is not necessarily fit to dive, medically.

To those with more knowledge of diving patho-physiology it becomes obvious that even the highest standard of physical fitness will not protect a diver from some of the complications from lung cysts or asthma, from a diving death.

The examining physician must consider many factors when conducting a diving medical examination. Almost 10% fail the medical and 10–15% incur specific diving limitations or advice, for safety reasons.

PSYCHOLOGY

The ideal diver is probably the cool James Bond like character we would all like to be - stable, calm under stress, able to endure physical and mental pressure, not prone to anxiety, able to conveniently ignore danger, slightly overweight and perhaps not surprisingly, a fluent liar.

Psychological stability is difficult to evaluate during the medical examination. Some clues may be gained from the history of sporting activities and occupation. Often the diving instructor is best able to evaluate the diver's psychological make-up during the course of instruction.

AGE

Fig. 38.1

Ideally the trainee diver should be aged between 18 and 35 years although exceptions can be made at both extremes of age. Divers over 45, if complying with the medical standards should be acceptable, but may require special tests such as a cardiac risk assessment and physical fitness checks.

Divers younger than 16 require very careful supervision during and after training because of their often smaller stature, limited strength and (most importantly) emotional immaturity. A buddy line to an experienced adult diver is recommended during the training of youthful divers. The mature and experienced buddy of an adolescent diver should take control of the dive and remember that his buddy may be an unreliable rescuer if difficulties arise. Most reputable medical authorities will not certify divers under the age of 15–16 years, without imposing serious limitations. This does not prevent younger divers being given a limited "diving experience" by qualified diving instructors under very strict and controlled conditions, and provided they are medically fit.

OCCUPATION

Pilots and aircrew are advised of the risks associated with flying after diving. Musicians, sonar operators, cardiologists, pilots and others reliant on excellent hearing for their livelihood are informed of the small but real risk to their hearing, or development of tinnitus, should they suffer ear barotrauma.

MEDICATION

Any illness requiring drug treatment needs careful consideration because either the illness or the drug may compromise diving safety. Sedatives, tranquillisers, antidepressants, antihistamines, anti-diabetic drugs, steroids, anti-hypertensives, anti-epilepsy drugs, alcohol and hallucinatory drugs such as marijuana and LSD all place the diver at risk. See Chapter 37 for more specific details.

Some antibiotics may have no direct adverse effect on diving, but the condition for which they were prescribed may have.

Experience and experiments indicate that many drugs which affect the brain have unpredictable effects on a diver exposed to the very high pressures encountered in deep diving.

HEART

Most heart diseases or abnormalities of heart rhythm are incompatible with safe diving and are disqualifying conditions. They can often be inferred from the personal or family history, clinical examination, biochemical tests or electrocardiograms (ECGs). The blood pressure should be normal for the age of the diver. See Chapter 35.

OBESITY

The overweight person is more prone to decompression sickness when air diving and is likely to have a reduced level of physical fitness. Most physically fit obese individuals may dive safely with appropriate reductions of the allowable durations of dives.

LUNGS

Lung disease is a disqualifying condition. The diver needs normal lung function to allow a reserve of respiratory function to cope with exertion and to permit easy air flow from the lungs to avoid pulmonary barotrauma. The lungs must be very elastic to enable them to stretch during sudden volume changes on ascent. A history of asthma, chronic bronchitis, bronchiectasis, fibrosis, cysts, spontaneous pneumothorax, chest injury or chest surgery are disqualifying conditions.

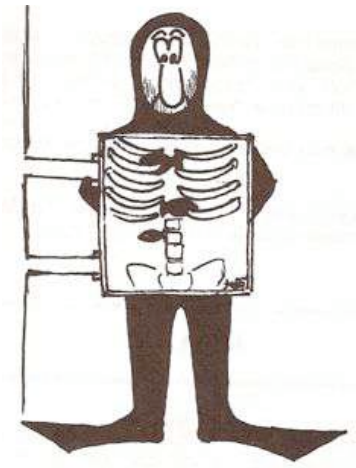

Fig 38.2

The doctor may be able to detect localised airway obstruction (which can lead to a burst lung) by listening to sounds made in the chest when the diver breathes deeply and rapidly. The history and **respiratory function tests** (expiratory spirometry) aid in the assessment. Occasionally radiological screening (Chest X-ray, CT scan etc.) may be necessary.

There was a dramatic drop in the incidence of burst lung in Australian Navy divers after the institution of these standards.

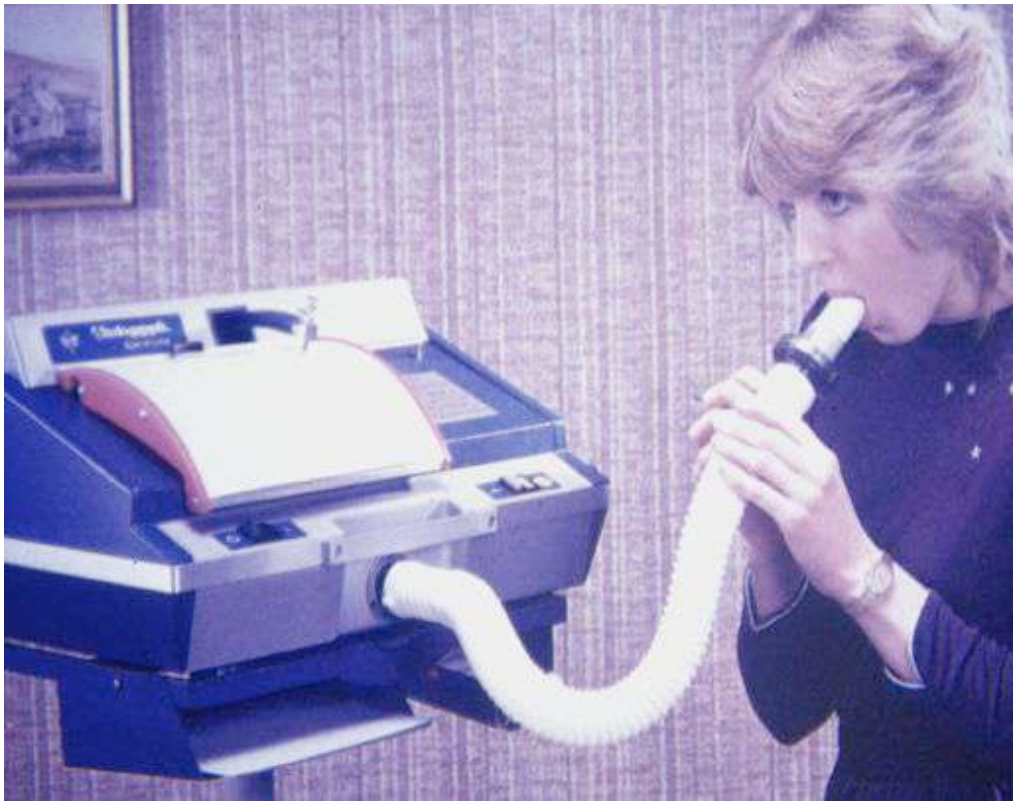

Fig. 38.3
Diving candidate blowing into a "Spirometer" to assess lung function. These devices have been largely replaced by digital expiratory spirometers, standardised for the specific population being tested.

EAR, NOSE AND THROAT

The ears, nose, throat and sinuses account for most diving induced illnesses. Any acute infection such as a cold will temporarily disqualify a candidate. A history of chronic or recurrent allergies, hay fever, sinusitis, tonsillitis, or tooth decay needs special assessment. Diving should be avoided while so affected. A deviated nasal septum (often appearing as a crooked nose) can cause obstruction of the sinus openings. All these factors can predispose to **sinus** or **ear barotrauma**.

The ears are carefully examined. The outer ear must be free from infection and not blocked with wax. The eardrum must be seen to be moved voluntarily during the **Valsalva**, or other equalising manoeuvre. An eardrum which has been scarred from previous perforation may be weakened. The examining physician, by viewing the ear drum while the diver attempts middle ear equalisation, can advise on correct techniques to be used when diving.

The hearing function test (**pure tone audiogram**) measuring hearing up to 8000 Hz is performed. Any significant hearing loss is regarded seriously since there is a risk of further hearing loss if **barotrauma** to the ears occurs during the diver's exposures.

Damage to the hearing organ may also be associated with disturbance of the balance organ. A special type of balance test is used to detect this, called the Sharpened Romberg, and further investigation is by an electronic measurement (electronystagmogram) if necessary. It is important to detect any balance organ dysfunction since it can lead to **vertigo** and **vomiting** underwater.

EYES

Good **vision** is essential for the diver to see his boat or buddy, if he surfaces some distance away. A diver who has impaired vision can have **corrective lenses** included into his face mask, but should always dive with a visually fit buddy in case the mask is lost or broken during the dive. See Chapter 5.

Contact lenses can pose problems and advice is needed about these. Hard lenses can trap bubbles between them and the cornea, causing pressure damage. Soft lenses are susceptible to loss – especially during mask removal. These divers are advised to keep the eyes closed when removing the mask, either underwater or on the surface. See Chapter 32.

The operation of **radial keratotomy**, used to surgically correct short sightedness, can cause problems. With this procedure, the cornea is cut radially in a sunburst pattern to change the curvature of the cornea. These cuts weaken the cornea which is prone to burst if the eye is bumped or subjected to external pressure reduction. If such a diver develops face mask squeeze (see Chapter 12), the eyeballs may actually rupture. Anyone who has undergone this operation should not dive. Most modern techniques, such as Laser resections for myopia, involve only minimal damage to the cornea, and are not a problem.

Colour vision is of lesser importance, apart from a few professional diving situations involving colour coded cylinders or wires (involving explosives).

BRAIN

Any disorder of the nervous system will complicate and confuse diagnosis and treatment of diving illnesses such as cerebral air embolism and decompression sickness.

Epileptics, even if controlled by drugs, should not dive as an epileptic fit underwater could prove fatal. The higher partial pressures of oxygen encountered during a scuba dive may render these persons more vulnerable to such attacks. Hypoxia, hyperventilation and sensory deprivation can aggravate fits. Many divers have had their first fit underwater.

Migraine is often made worse by diving (see Chapter 32). Severe migraine attacks leading to incapacity have occurred during dives in previously mild sufferers. It may also complicate recompression treatments. If certain precautions are observed some migraine sufferers can

engage in limited diving in reasonable safety. A patent foramen ovale in some divers may increase the frequency of migraine.

GENERAL CONDITIONS

Other diseases of the body such as **diabetes mellitus** (see Case History 33.6), severe **kidney** or **liver** disease also increase the risks of diving.

Muscle, bone and **joint** diseases or injuries can predispose to decompression sickness and make diagnosis and treatment of this disorder more difficult. Fatigue may be induced more easily.

Professional divers or those who frequently undertake decompression diving may require **long bone radiology or scanning** (see Chapter 17) to establish a baseline in the event of bone abnormalities developing, and for legal reasons. Because of the low risk of dysbaric osteonecrosis, the cost and the potential hazards posed by radiation exposure, these are not usually recommended for recreational divers.

A history of **motion sickness** is significant because it interferes with safe diving and it is difficult to vomit through a demand valve. Divers with a propensity to this condition need advice from the physician on remedies for seasickness which are compatible with safe diving (see Chapter 32).

Smoking diminishes physical fitness and can predispose to lung, sinus and ear barotrauma.

Pregnancy should preclude diving (see Chapter 8).

PHYSICAL FITNESS

This refers to the strength and speed, so necessary to athletes. It includes muscular, cardiac and respiratory capabilities. It is important to divers, as they are often called upon to exert themselves, to survive. One reasonable standard is to require an ability to swim, unaided, a distance of 200 metres in less than 5 minutes for recreational divers who do not subject themselves to difficult conditions. For professional or competent divers, this could be reduced to 4 minutes.

Medical fitness for diving refers to the freedom from illness likely to prejudice diving safety.
'Physical fitness' does not necessarily equate with 'diving medical fitness'.

It is not uncommon for physically fit young individuals to feel quite distressed when advised against scuba use by diving medical practitioners.

MEDICAL EXAMINATION FORMAT

There is little doubt amongst responsible diving instruction groups and diving medical associations, that mandatory full and comprehensive medical examinations should be performed on all divers before commencing scuba training. It is also needed before using scuba apparatus – even in such shallow and apparently safe locations as a swimming pool.

Fig 38.4

During a recent workshop on diving medical examinations, the following consensus was achieved with this advice for recreational divers:

• All diving candidates must be examined according to an established diving medical Standard. An example is the South Pacific Underwater Medicine Society (SPUMS) Medical Format (included in this Chapter) prior to commencing any use of scuba apparatus – even if only in a pool.

• The medical examiner must have been trained appropriately (at a recognised course) in diving medicine.

• Should any doubt exist as to the 'fitness' of an individual, then that person must be referred to a specialist diving medical practitioner (i.e. one with extensive training and experience in diving medicine).

This textbook is not aimed at instructing medical practitioners in Diving Medicine – although it will serve as a useful primer for those interested in this type of medicine. A list of recommended courses of instruction and reading texts is included in Appendix A.

A copy of a typical **Diving Medical Format** follows. It is suitable for candidates wishing to experience Scuba diving or to subject themselves for diver training. It must be performed and interpreted by a physician trained in diving medicine by an accredited body.

It comprises 3 sections:

(1) Medical history
(2) Diving [and diving medical] history
(3) Clinical examination and investigations.

Each is necessary and every item except for identification data, is of relevance to diver safety and diving limitations.

APPENDIX B
PRE-DIVE MEDICAL FORM FOR PROSPECTIVE ENTRY-LEVEL SCUBA DIVERS
The first two pages to be completed by candidate.

1	Surname Other Names	2	Date of Birth
		4	Sex: Male Female
3	Address	5	Telephone (Home)
6	Principal Occupation	7	Telephone (Work)
8	Intended Dive School		
9	Do you participate in any regular physical activity?	Yes	No
10	Description of activity		
11	Do you smoke?	Yes	No
12	Do you drink alcohol?	Yes	No
13	How many drinks a week?		
14	Are you taking any tablets, medicines or drugs? List:	Yes	No
15	Do you have any allergies? Details:	Yes	No
16	Have you had any reactions to drugs or medicines or foods? Details:	Yes	No

Have you ever had or do you now have any of the following? Tick Yes or No.

		Yes	No	Notes on History
17	Previous diving medical			
18	Prescription glasses			
19	Contact lenses			
20	Eye or visual problems			
21	Hay Fever			
22	Sinusitis			
23	Other nose or throat problem			
24	Dentures/Plates, etc.			
25	Recent dental procedures			
26	Deafness or ringing noises in ear(s)			
27	Discharging ears or other infections			
28	Operation on ears			
29	Giddiness or loss of balance			
30	Severe motion sickness			
31	Seasickness medication			
32	Problems when flying in aircraft			
33	Severe or frequent headaches			
34	Migraine			
35	Fainting or blackouts			
36	Convulsions, fits or epilepsy			
37	Unconsciousness			
38	Concussion or head injury			
39	Sleep walking			
40	Severe depression			
41	Claustrophobia			
42	Mental illness			
43	Heart disease			
44	Abnormal blood test			
45	ECG (Heart tracing)			
46	Consciousness of your heart beat			
47	High blood pressure			
48	Rheumatic fever			
49	Discomfort in your chest with exertion			
50	Short of breath on exertion			
51	Bronchitis or pneumonia			
52	Pleurisy or severe chest pain			
53	Coughing up phlegm or blood			

		Yes	No
54	Chronic or persistent cough		
55	TB		
56	Pneumothorax ("collapsed lung")		
57	Frequent chest colds		
58	Asthma or wheezing		
59	Use a puffer		
60	Other chest complaint		
61	Operation on chest, lungs, or heart		
62	Indigestion, peptic ulcer or acid reflux		
63	Vomiting blood or passing red or black motions		
64	Recurrent vomiting or diarrhoea		
65	Jaundice, hepatitis or liver disease		
66	Malaria or other tropical disease		
67	Severe loss of weight		
68	Hernia or rupture		
69	Major joint or back injury		
70	Limitation of movement		
71	Fractures (broken bones)		
72	Paralysis or muscle weakness		
73	Kidney or bladder disease (cystitis)		
74	In a high risk group for HIV or AIDS		
75	Syphilis		
76	Diabetes		
77	Blood disease or bleeding problem		
78	Skin disease		
79	Contagious disease		
80	Operations		
81	In hospital for any reason		
82	Life insurance rejected		
83	A job or licence refused on medical grounds		
84	Unable to work for medical reasons		
85	An invalid pension		
86	Other illness or injury or any other medical conditions		
Have any blood relations had			
87	Heart disease		
88	Asthma or chest disease		
89	TB		
Females Only			
90	Are you now pregnant or planning to be?		
91	Do you have any incapacity during periods?		

92 Date of most recent chest x-ray

Previous Diving Experience		Yes	No
93	Can you swim?		
94	Have you ever had any problem during or after swimming or diving?		
95	Have you ever had to be rescued?		
96	Do you snorkel dive regularly?		
97	Have you tried scuba diving before?		
98	Have you had previous formal scuba training?		
99	Year		
100	Approximate number of dives		
101	Maximum depth of any dive		
102	Longest duration of any dive		

I certify that the above information is true and complete to the best of my knowledge and I hereby authorise Dr. to give medical opinion as to my fitness, or temporary or permanent unfitness to dive to (Diveshop). I also authorise him or her to obtain or supply medical information regarding me to other doctors as may be necessary.

Signed: **Date:**

MEDICAL EXAMINATION: To Be Completed By An Approved Medical Practitioner.

1 Height cm	2 Weight kg	3 Visual Acuity R6/ Corrected 6/ L6/ Corrected 6/	4 Blood Pressure	5 Pulse
6 Urinalysis Albumen Glucose		7 Respiratory function test (Measured by equipment capable of reading to 7 litres) Vital capacity FEV_1 Percentage	8 Chest x-ray (if indicated) Date Place Result	

9 Audiometry (air conduction)						
Freqoency, Hz	500	1,000	2,000	4,000	6,000	8,000
Loss in DB(R):						
Loss in DB(L)						
If abnormal enter in diver's log book and on certificate						

Clinical Examination/Assessment

		Normal	Abnormal
10	Nose, septum, airway		
11	Mouth, throat, teeth, bite		
12	External auditory canal		
13	Tympanic membrane		
14	Middle ear auto-inflation		
15	Neurological Eye movements Pupillary reflexes Limb reflexes Finger-nose Sharpened Romberg		
16	Abdomen		
17	Chest hyperventilation		
18	Cardiac auscultation		
19	Other abnormalities		

Notes on Abnormalities

Fit to Dive Yes

Advice put on certificate

No Temporary Reason:

No Permanent Reason:

Printed Name: Date

Signed:

Detach the certificate below and hand to the candidate.

Medical Benefits Refund and/or Medical Rebate is not permissible, by law, for this examination. Issue of any Item Number which allows the candidate to claim such benefit will result in the physician being guilty of medifraud.

This is to certify that I have examined

Name:

Address:

in accordance with the requirements of the Australian Standard for the training and certification of recreational divers, and have found him/her to be:

Fit Permanently Unfit Temporarily Unfit (to be reviewed on) for diving and diving training to 18 m, undertaken using compressed air underwater. Audiogram normal/abnormal (see below).

Printed Name Signed

Address:

Date

Telephone

Advice

Chapter 39

FIRST – AID KIT

Certain drugs and equipment are of value in a diving accident and a diving team could reasonably be expected to acquire and carry these on diving expeditions. Training in the use of these, as well as in resuscitation, is of great importance.

FIRST-AID MATERIALS

For shark attack or trauma, large sized thick **cotton pads** (more than 20 cm square) with 10 cm **crepe bandages** (6 of each) are useful to make **pressure dressings** to stop bleeding and also for pressure bandages to reduce venom absorption. If obtainable, **shell** dressings of the type used by the military are ideal for this purpose. They can sometimes be obtained from army disposal stores.

A **rubber bandage** 10 cm wide ("esmarch" bandage obtainable from a medical equipment supplier) for use as a **tourniquet**. When wrapped tightly around the limb this is the best form of tourniquet. It covers a wide area, effectively stopping blood flow to the limb while minimising damage to tissues under the tourniquet.

Small adhesive skin **dressings** such as Elastoplast or Band-Aids.

Surgical instruments — scissors, artery forceps, fine forceps, disposable scalpel blade, disposable syringes and needles.

An aluminised **thermal blanket** such as a "Space blanket" to protect divers suffering from hypothermia.

Heat packs — of value in treating fish and jellyfish stings.

Cold packs — of value in treating jelly fish sting and general muscular strains

Eye irrigation solution.

Torch, pen and paper (for recording purposes).

RESUSCITATION EQUIPMENT

• **Airways (Guedel type)** in two adult sizes are useful if a victim loses consciousness and develops airway obstruction, or if artificial respiration is needed. A positive pressure air system (such as an AMBU Bag) is of value in combination with the airway for prolonged artificial respiration.

• **Oxygen Resuscitation Equipment**. A supply of oxygen and equipment to administer it can be lifesaving in some diving accidents. Devices as described in Chapter 40 should include a complete oxygen supply and delivery system in a robust portable container.

• A **large oxygen cylinder** with appropriate adaptors should be available if diving at a distance from diving medical facilities and recompression chambers.

An **underwater oxygen** system (appendix C) for recompression therapy by more sophisticated groups, in remote areas.

MEDICATIONS FOR DIVING PROBLEMS

• Household **vinegar,** preferably in a one litre container, to neutralise adherent stinging cells of box jelly fish and some other jelly fish. Household **bleach** is useful for sterilising coral cuts.

• **Local anaesthetic spray or ointment** (lignocaine) to relieve the pain from minor stings from animals such as Portuguese man-o-war and other jellyfish stings. **Solacaine** or other anti-burn preparations such as **Tannic acid** sprays may be efficacious for this purpose.

• Topical **antibiotic powder** to prevent infection from coral cuts and other minor injuries.

• **Skin antiseptic solution** such as chlorhexidine for cleaning wounds contaminated with dirt.

• Broad spectrum **antibiotic tablets** (e.g. erythromycine, doxycycline) to initiate treatment for serious infections, otitis externa, otitis media, sinusitis, and coral cuts etc.

• **Prophylactic ear drops** such as commercial preparations of Aqua Ear, Vosol or Otic Domoboro.

• **Therapeutic ear drops,** including antibiotic and steroid combination, for outer ear infections.

• **Local anaesthetic for injection** such as lignocaine 1% (without adrenalin) for wounds from stone fish and other fish stings. Up to 15 ml of this solution can be injected into the stung area in an adult and repeated every 2 hours if necessary.

• **Antivenoms** — depending on the geographical location.

GENERAL MEDICATIONS

• **Anti-diarrhoea** tablets such as diphenoxylate ("Lomotil") or loperamide ("Imodium").

• **Analgesics** (pain killers) such as paracetamol (acetaminophen). Aspirin, or drugs containing this substance, may be unpredictable and hazardous and are best avoided.

• **Ultra-violet** blocking sunscreen (SP15+ or greater). A **1% hydrocortisone cream** is useful to treat sunburn, allergic dermatitis or itching.

• **Anti-Seasickness** tablets (see Chapter 32).

• **Decongestants —** pseudoephedrine tablets, and topical nasal sprays.

• **Topical antibacterial and antifungal** preparations, such as Cicatrin or Neosporin.

TRAINING

A diving team venturing to a remote locality should have at least one member (preferably two in case that one becomes the victim of an accident) trained in first aid relevant to divers. Resuscitation and oxygen administration requires expert training and supervision. Training in the use of injections is an advantage, both for the administration of local anaesthetics, antivenoms and other drugs under the direction and advice of a distant medical specialist.

MEDICAL INFORMATION

Perhaps the most valuable addition to any first aid box is a source of information. This should include diving **medical texts** (see appendix A) and general **contact numbers** (see appendix B & D) for both medical assistance and recompression chamber availability. This should be supplemented by local contacts and phone numbers of knowledgeable divers and diving physicians.

A copy of **this book** should remain with the First-aid kit.

Also in the kit should be a list of its **contents**, including purchase and expiry dates of the drugs.

INFORMATION NEEDED
about a DIVING ACCIDENT.
<u>CHECK LIST</u>

Name of informant

Name and age of victim

TELEPHONE NUMBER to return call, or if disconnected

Geographical location (+ local medical facilities or RCCs)

Case history
- Initial symptoms, including time of onset
- Description of clinical symptoms + progress
- Signs of illness
- Negative findings (symptoms NOT present e.g. micturition, dyspnoea, skin wounds)
- First aid given (including oxygen, amount and method)

Dive details: Profile, gases, deco,(+ recent dives)

Personal history
- Diving
- Illnesses – diving and general
- Medications and surgery

Any other persons injured

First aid facilities and/or medical personnel available on site

Follow up arrangements

Medevac possibilities

Chapter 40

OXYGEN THERAPY TECHNIQUES

100% Oxygen (O_2) therapy is an essential technique in first-aid treatment of many diving emergencies. Unfortunately, the correct use of O_2 equipment is often poorly understood by divers. See Chapter 16 for details of the use of O_2 in decompression sickness.

OXYGEN BREATHING EQUIPMENT

The apparatus for administering O_2 is not unlike a scuba system. It comprises a **cylinder** which holds the oxygen at a pressure comparable to scuba tank pressure, and a pressure **reducing valve (regulator)** connected either to a **demand valve** or to a **constant flow system**. There is a need to avoid the use of flammable lubricants (such as oils and silicone grease) which can cause explosions in the presence of O_2.

Oxygen can accelerate burning, cause spontaneous combustion, and can make ordinarily non-combustible materials burn furiously, so its use requires a strict fire prevention attitude. Oxygen should not be administered near heat sources, and smoking near O_2 is even more hazardous to your health than usual. Oxygen should not be used in poorly ventilated areas where high concentrations can build up. All equipment must be kept clean and the cylinder valve should be turned on slowly and the system purged before O_2 is given to the patient. If the patient is unconscious, always check that the apparatus is working by you or an attendant breathing on it first.

The O_2 can be delivered in high or low concentrations depending on the apparatus used. With a constant flow system, the O_2 can be delivered to the patient through either a cheap disposable (usually loose fitting) plastic **oxygen mask, nasal prongs,** a **nasal catheter** or a **bag-valve-mask resuscitator system.**

Where the administration of 100% O_2 is required (e.g. in first aid for decompression illness) this means that nothing but O_2 is inhaled by the diver. Many of the O_2 masks commonly used

in medicine have 100% O_2 delivered to the mask, but, the patient inhales only about 25-50% O_2 because the design of the mask allows this O_2 to mix with air, which dilutes the final breathing mixture. The disposable plastic O_2 masks in common use in ambulances and hospital casualty rooms are of this type.

These are **NOT** adequate for treating serious diving injuries.

Fig. 40.1
These oro-nasal plastic masks, catheters and nasal prongs are not adequate for treating divers with 100% O_2.

As a general rule, **decompression sickness**, **gas embolism** (and other manifestations of **pulmonary barotrauma**), carbon monoxide poisoning and **near drowning** cases should be given 100% O_2 from the outset.

Other diving accidents which produce shock or hypoxia can sometimes be helped with lower concentrations of O_2. In general, if the patient is cyanosed (blue) they need O_2 in a sufficient concentration to make them pink again. Usually, 100% O_2 is needed for treatment of diving accidents.

CONSTANT FLOW SYSTEMS

Such devices deliver a constant flow of O_2 to a mask or an alternative delivery system.

There are several types of plastic **disposable oxygen masks** available, the **Hudson** mask being typical. Each is normally accompanied by instructions which specify the correct O_2

flow to use, most masks using a flow of about 4–6 litres per minute. These masks allow the O_2 flow to be diluted by inspired air, the flow of which varies during inspiration and peaks at about 30 litres per minute. Because of this, the concentration of O_2 which the patient actually inhales is reduced from 100% to about 40 or 50%.

To increase the inspired O_2 percentage, increase the O_2 flow. Unless this is increased beyond 30 litres per minute an inspired percentage of 100% is not attainable. Such high flow rates rapidly deplete most divers' O_2 supply.

A device known as a **nasal prong** (see above photo) is available which delivers O_2 by small tubes directly into the patient's nostrils. An elastic head strap holds the prongs in place. This system has a similar efficiency to the common loose fitting O_2 mask but is more comfortable and more effective for long term use because it is less likely to dislodge when the patient sleeps.

These systems are acceptable when O_2 supplementation in low concentrations is all that is required. Cases in this category would include typical general medical conditions found in hospitals such as heart attack, and mild recovering cases of near drowning, salt water aspiration syndrome and shock associated with serious trauma or shark attack.

HIGH CONCENTRATION OXYGEN SYSTEMS

When near 100% O_2 delivery is required in cases such as decompression sickness, pulmonary barotrauma or near drowning, a more efficient O_2 delivery system is necessary.

This can be achieved by a **demand valve,** an O_2 circuit incorporating a **rebreathing bag,** or a bag–valve–mask device with an O_2 inlet and a reservoir bag with a very high O_2 flow rate.

Demand Valves

The simplest and most effective way to deliver 100% O_2 is via a **demand valve -** like a second stage regulator.

Some demand valves have been specially designed for O_2 administration. Some, such as the Oxiden and LSP demand valves, are designated to provide O_2 only to spontaneously breathing patients. Others, such as the Robert Shaw (which is used on the Oxy Viva, marketed in Australia) and the Elder demand valves, can produce 100% O_2 to a spontaneously breathing patient as well as provide O_2 resuscitation to a non-breathing casualty via a manual trigger. DAN also supplies an oxygen kit with this type of valve.

These demand valves are usually used with a tight fitting anaesthetic type mask. However, some can be fitted with a scuba mouthpiece to provide O_2 to a breathing diver (in this case the diver's nose should be sealed with a nose clip or by some other means).

In an emergency in remote locations, certain demand valves can be adapted to deliver O_2, with an adaptor connected to an O_2 reducing valve, or alternatively by connecting the diver's first stage regulator to an O_2 cylinder, using the specialised adaptor. It is essential if this system is used that all components in the breathing system (including the lubricants) are O_2

compatible. Otherwise the diving illness may be complicated by fire, explosion, or shrapnel injuries. When using this type of system the regulator should be tested for safety by purging it before anyone breathes from the system.

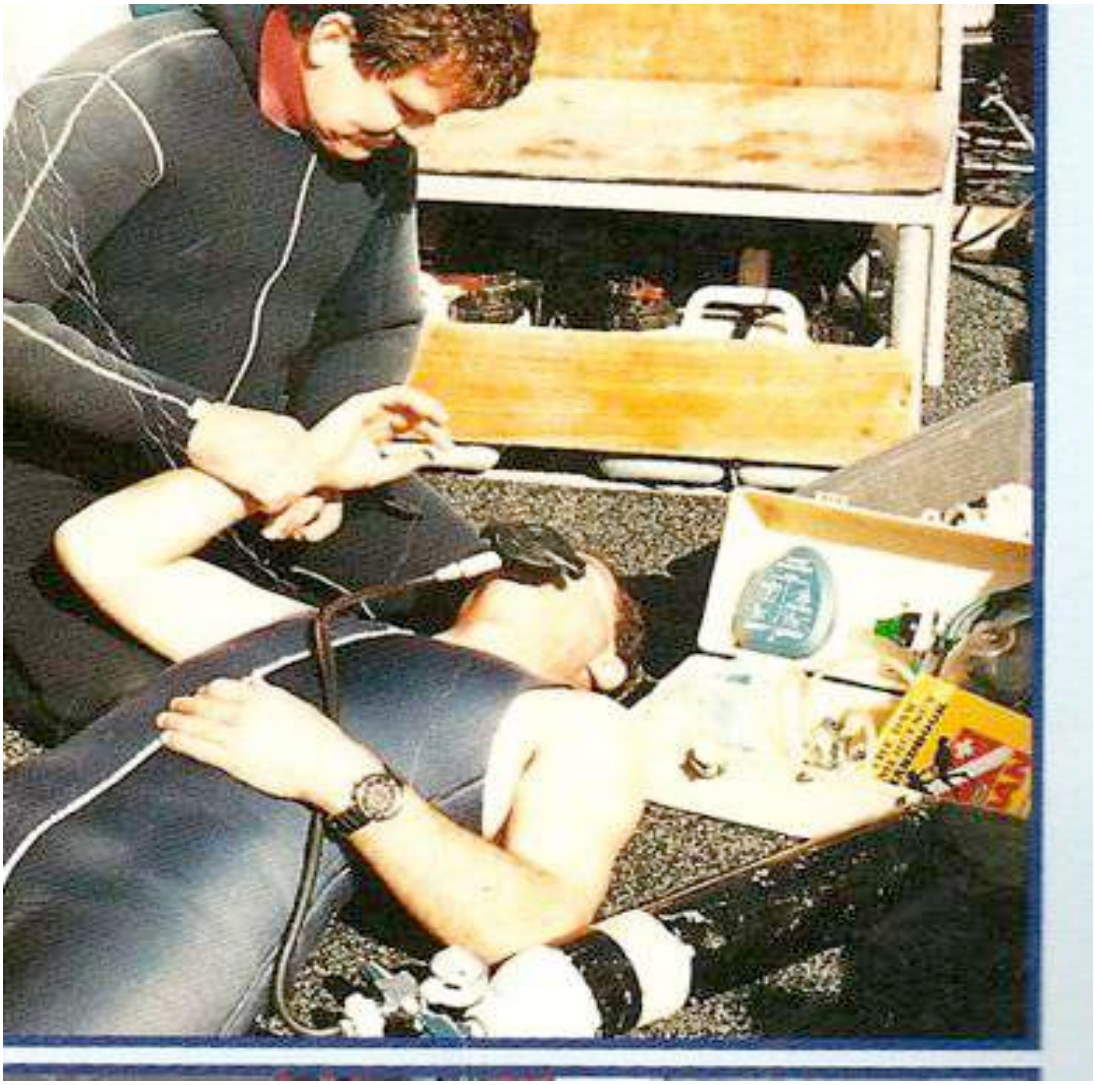

Fig. 40.2

A system designed by DAN (USA) specific for divers and incorporating a demand valve for 100% O_2 inhalation. The diver must be breathing spontaneously.

Rebreathing System

Oxygen use can be reduced if the diver breathes from a system which permits him to rebreathe some of his exhaled gas. This requires the nitrogen to be flushed from the breathing circuit after a few minutes rebreathing, and the exhaled CO_2 to be absorbed in a soda line (or similar chemical) canister. There are commercially available units using this system. Technical divers frequently use diving sets that can be used for 100% O_2.

PSA Oxygen Concentrator

These devices use a molecular sieve to filter nitrogen from air and deliver high concentrations of oxygen. They are available in a portable size and can be battery powered.

Bag–Valve–Mask Respirator

These devices comprise a self inflating rubber or plastic bag and a valve system which allows the patient to be ventilated with air through a mask by squeezing the bag. Oxygen can be piped into the circuit to increase the inspired O_2 concentration. If a reservoir bag is attached to the inflation bag, the inspired O_2 concentration can be increased to almost 100% by using a large O_2 flow (approx. 12–14 litres per minute), providing a good seal is achieved and the patient's respiratory volume is not excessive. The valve system also allows the patient to breath spontaneously from the device. Details of the operation of these devices vary between manufacturers and are outlined in the instruction manual supplied.

Fig. 40.3
A resuscitator showing anaesthetic-type mask, hand-squeezed breathing bag, oxygen reservoir bag and head strap to secure mask to diver's face.

Multiple Systems

Most countries have a variety of multi-purpose near 100% O_2 delivery systems available. Aga, AG, CIG, Ambu, LSP, Drager and Laerdal and DAN manufacture such units. Self-contained O_2 therapy units and resuscitation devices are now available with a demand valve allowing the patient to either breathe spontaneously out of the device, be given positive pressure ventilation by manual pressure, or use a breathing bag. A flow meter can be fitted to another part of the system, to allow O_2 to be delivered to a mask at a high continuous flow rate. The details vary between resuscitators and are explained in instructions supplied with them. The entire device should be contained in a sturdy rustproof metal or heavy duty plastic box which is compact, water resistant and easy to carry. It should have an adaptor and be able to be supplemented by connection to a larger O_2 cylinder, if necessary.

GENERAL INFORMATION

Oxygen Toxicity

(See Chapter 21). 100% O_2 will cause some reversible damage to lungs after 18–24 hours. This has to be balanced against the benefit of the condition being treated and this will usually be physician's decision. If medical advice cannot be obtained, in decompression sickness and gas embolism cases, or with near-drownings or carbon monoxide toxicity, O_2 toxicity is generally the lesser of two evils, and it is usually best to continue giving 100% O_2 until expert advice to the contrary is given.

Oxygen toxicity is generally not a consideration with low concentration devices as they do not delivery much more than 40% O_2, which is probably below the threshold for toxicity for the estimated duration of exposure.

Contraindications to O_2 Therapy

In addition to the problems of O_2 toxicity, there are some problems associated with oxygen usage in the general community.

There is a theoretical risk to premature babies (eye damage) and to sufferers of emphysema (respiratory depression). Neither of these groups are numerous in the diving population.

In recent years a problem of sensitivity to O_2 has emerged in cancer sufferers who have been treated with the drug **Bleomycin,** and similar substances. These people can suffer from severe, permanent lung damage if given O_2 in concentrations greater than 21% (air). This problem is not likely to be frequently encountered, as these people should be excluded from scuba diving.

Practicalities of O_2 Administration

The main disadvantage of resuscitation apparatus is the limited O_2 supply available from the contained O_2 cylinder, because sufficient O_2 must be supplied to allow for transport of the diver from the accident site to an appropriate medical facility. This can be overcome by carrying additional cylinders or by an adaptor which allows a connection to a larger O_2 cylinder. Estimate the rate of consumption of O_2, the supply available, and plan accordingly. During transport, and whenever O_2 is given, attention must be paid to adequate ventilation to prevent O_2 build up to dangerous levels.

All divers are encouraged to undergo additional training in resuscitation and O_2 administration. This should be mandatory for dive masters and dive instructors.

Cases such as gas embolism and decompression sickness generally require **100% inspired O_2** from the outset. A system capable of delivering 100% O_2 must be chosen. Even then, an ill fitting mask may allow air to be breathed around the seal, diluting the O_2. Attention to the mask fit and attachment is necessary to prevent this. It is especially likely if the patient sleeps. If a demand valve is used, make sure the patient does not breathe air through the nose at the same time. This can be prevented by using a nose clip (improvise if necessary) or a diver's face mask.

Whenever O_2 is administered there is a serious **fire hazard** since increased concentrations of O_2 accelerate burning and can make ordinarily non-combustible materials burn furiously. The area where the O_2 is administered should be well ventilated and sources of ignition and combustible materials (including cigarettes) should be avoided. The system should be turned on slowly and should be tried and running before it is applied to the patient's face.

Further information may be obtained from the text "Oxygen First Aid for Divers"— by John Lippmann, J.L. Publications, Australia.

Fig 40.4

A portable oxygen resuscitator. This system allows positive pressure ventilation with 100% O_2 for a non-breathing diver, or passive demand supply for a breathing diver. A long supply hose can be fitted to a larger cylinder containing 100%O_2 for prolonged use. The airways and suction appliance are included

Chapter 41

TRAINING & SAFETY CHECKLIST

SAFETY INFORMATION AND BACKUP

- Diving manuals and library (see Appendix A).
- Diving medical and safety manuals and texts (see Appendix A).
- Diving medical organisations (see Appendix B & D).
- Medical Insurance — DAN has organised this for North Americans, world wide.

DAN.AP covers the area from the eastern Indian ocean to the western Pacific ocean
BSAC arranges insurance from the UK/Europe.
Most diving shops and organisations advise on the relevent local contacts.

DIVER TRAINING

- Physical fitness – 200 metres unassisted swim in less than 5 minutes.
- Medical examination fitness for diving by a qualified diving physician. Includes pure tone audiogram and expiratory spirometry.
- Entry level diving certificate. Qualified by a diving instructor from a reputable training organisation.
- Advanced level training, specialty courses such as buoyancy training.
- Diving rescue and resuscitation training.
- Specialised courses for the specific diving environments (cave, deep, etc.).

DIVE PLAN

- Boat operator's safety and equipment check.
- Dive plan — includes terrain information, navigation and emergency plan, maximum depth, air supply, ascent rate, safety and decompression stops.
- Basic diving equipment check and maintenance.
- Dive responsibility, documentation and buddy system.
- Diving rescue equipment — includes BC, emergency air supplies, alarm signals, etc
- Lost diver strategy, including EPIRB in open ocean.
- Diving rescue plan - includes boat / buoy / diver backups.
- First-Aid kit and resuscitation equipment. (See Chapters 39 and 40).
- Emergency contact numbers (local and elsewhere).

FOR DIVE OPERATORS

- Register trip plan with local maritime or port authority
- Rescue and resuscitation equipment and training
- Onboard observer and recorder (diver in/out times)
- Communication system. Satellite phone, radio etc
- A good lawyer!

CHECK LIST FOR DIVING ACCIDENTS

INFORMATION NEEDED if reporting a DIVING ACCIDENT.
for both reporter and recipient

Check List

- NAME of informant
- NAME AND AGE of victim
- TELEPHONE NUMBER to return call, or if disconnected
- Geographical LOCATION (+ any local RCCs)
- CASE HISTORY
    - Initial SYMPTOMS, including time of onset
    - Description of clinical symptoms + progress
    - Signs of illness
- NEGATIVE FINDINGS (symptoms that are NOT present)
- FIRST AID given (including oxygen, amount, method and time)
- DIVE DETAILS Profile, gases, deco, (+ recent dives)
- PERSONAL HISTORY
    - Diving
    - Illnesses – diving and general
    - Medications and surgery
- Any other persons injured
- First aid FACILITIES and/or medical personnel available on site
- Follow up ADVICE and ARRANGEMENTS
- Possibility of medevac

n.b You may think that you have read this before, in Chapter 39. You have. In emergencies, either chapter may be accessed.

Chapter 42

RESUSCITATION REVIEW

It is not possible to learn the techniques of resuscitation from a book. To acquire these skills, the authors recommend that all divers undertake a resuscitation course from one of the many organisations worldwide which teach these techniques. Once learnt, the skills need to be practised regularly, just as do diving emergency procedures.

The protocol used here is meant as a reminder to divers who have already been trained in resuscitation. It is based on the basic life support (**BLS**) recommended by the Australian Resuscitation Council <http://www.resus.org.au/>. Organisations in other countries may have slightly different protocols, but if from reputable organisations they should be equally effective.

WHAT IS RESUSCITATION?

Resuscitation is the restoration or preservation of life using **basic life support (BLS**). This includes the A–B–C — Airway, Breathing, and Circulation — to preserve oxygenation to vital tissues. The most important tissue to protect from hypoxia is the brain.

Expired Air Resuscitation (EAR) is usually the best method of initially ventilating the lungs. If the rescuer has the equipment and skill to ventilate the victim with an Oxygen Resuscitator, then that is preferable for diving accidents.

ASSESSMENT OF THE DIVING CASUALTY

D and R - DANGER and RESPONSE

DANGER

It is important to protect yourself, others and the victim from further injury. It includes retrieval of the victim from a drowning or hypothermic situation, protection from marine animal injuries (shark attack, chironex tentacles etc.) and avoidance of physical trauma from boats, surf etc.

RESPONSE. Is the Victim Conscious?

Most problems arise in an unconscious victim. If the victim appears unconscious, confirm this by shouting at him and squeezing the shoulder. If the victim does do not respond, he is probably unconscious. If the victim is **conscious** he will normally take care of his own airway and breathing.

Exceptions to this are the sea snake, blue ringed octopus or box jellyfish (Chironex) envenomation, where the victim may be conscious but paralysed. In these cases the victim will not respond to shouting, so the management for an unconscious victim (which is appropriate) will be undertaken.

If the victim is **unconscious** he will be in danger of hypoxic hypoxia from obstruction of the airway or inhibition of breathing, or of stagnant hypoxia from lack of circulation (see Chapter 20).

With an unconscious victim, take care of the following systems:

- **AIRWAY**
- **BREATHING**
- **CIRCULATION**

This is easily remembered by the mnemonic — A-B-C.

Maintenance of airway, breathing and circulation takes precedence over other forms of care. Without these functions, the victim is certain to die.

A — AIRWAY

An unconscious victim loses muscle control. Loss of control of the muscles of the throat and tongue can cause the airway to become obstructed. This is particularly likely when the victim is lying on his back, mainly due to the tongue falling backwards into the throat, due to gravity.

The airway can be further obstructed by vomit, saliva or foreign material. This would normally be swallowed or spat out by a conscious person, while any material which entered the larynx or trachea would elicit the reflexes of coughing and laryngeal closure. These reflexes may be lost in the unconscious patient.

Clear the airway

Cases of **near-drowning**, and other unconscious victims rescued from the ocean, are probably best managed initially with the victim on his side.

To prevent airway obstruction, turn the victim on his side (left side down for gas embolism).
Sweep any foreign material from the mouth with the fingers.
Position the mouth slightly downward to allow any fluid to drain out. (In these cases, there is frequently fluid, vomit or other material in the airway).
The side (lateral) position with the mouth turned slightly downward will help any fluid to drain from the airway.

In **other cases**, and after clearing the airway of obstructions, secure the airway with the victim on their back. (See the chapter on pulmonary barotrauma for special considerations in this condition).

Tilt the head backwards, to open the airway, if necessary apply the chin lift maneuver (taught in resuscitation courses) if the airway remains obstructed. If there is a possibility of spinal injury, further injury by unnecessary movement of the spine should be avoided when positioning the patient, but clearing the airway takes precedence.

B — BREATHING

Even with a clear airway the victim may not breathe because of respiratory muscle paralysis, cerebral injury, hypoxia, cardiac arrest or other reasons.

Check for breathing – look, listen and feel.

Look and feel for respiratory movements of the chest and abdomen. Listen and feel for air moving from the nose and mouth. Occasional gasping respirations should be treated as not breathing.
If the victim is making respiratory movements ensure that the airway is clear.

Recognition of airway obstruction
If the airway has been cleared of foreign material, it can still be obstructed.
Signs of **obstructed airway** are snoring sounds, crowing sounds or movements of the chest and abdomen with no air moving from the nose and mouth.
If there is airway obstruction, look to improving the head tilt and chin lift maneuver.

If the victim is breathing, put him or leave him on his side (coma position) and keep the airway clear.

If the victim is not breathing:

Turn onto the back, commence expired air resuscitation (EAR).

The theory and practice of this are covered in a resuscitation course. The steps are:

- clear the airway
- tilt the head back, apply chin lift
- use mouth to mouth or mouth to nose expired air resuscitation
- look for the chest rising with each breath (if it is not, clear and open the airway)
- look, listen and feel for exhalation

If the rescuer has the equipment and skills, help secure the airway using an airway device (eg guedel airway), and ventilate using mouth to mask, a self inflating bag (eg Ambubag) with or without oxygen, or an oxygen inflating device (eg Oxyviva).

Give two full inflations of one second each then:

Check the circulation

C — CIRCULATION

If the victim is unconscious, unresponsive and not breathing it can be assumed that there is no circulation and external cardiac compression (ECC) should be started (Australian Resuscitation Council recommendation).

Note: Previous protocols recommended feeling for a pulse at this stage. This has now been omitted as it is often difficult to feel a pulse in these circumstances, even by medical personnel, and feeling for a pulse and then deciding there is none delays the start of life saving ECC.

Advanced resuscitators such as doctors or paramedics may check for circulation by feeling for a carotid pulse before commencing ECC. But this may be unreliable even in experienced hands and no more than 10 seconds should be used checking for a pulse. If there is no pulse, or any doubt, commence ECC.

When there is no circulation, the sooner ECC is commenced, the better the outcome.

Commence external cardiac compression (ECC) and Cardiopulmonary Resuscitation (CPR)

External cardiac compression. Circulation sufficient to keep the victim alive can be restored with this technique. The chest is compressed by pressure on the sternum, forcing blood into the major arteries and producing some circulation to the vital organs. It takes numerous compressions to establish a blood pressure adequate to provide circulation. If compressions are stopped, the blood pressure immediately falls and the process has to be restarted. Thus the number of times compressions are stopped should be minimized.

The combination of artificial respiration and external cardiac compression is called CPR. This is performed as follows.

CARDIOPULMONARY RESUSCITATION (CPR)

When the heart stops, so does the breathing (except for occasional gasping respirations in some cases). It is necessary to maintain respiration and cardiac function simultaneously. CPR is EAR and ECC combined.

Give two EAR (expired air resuscitation) breaths of one second per inspiration each, then

Start external cardiac compression (ECC)

Practical details of this are covered in a resuscitation course. Remember, compress:

– vertically
– over the lower half of the sternum (the centre of the chest)
– with the heel of the hand, both hands locked together
– with no pressure on the ribs
– to a depth of 4–5 centimetres (1.5–2 inches or 1/3 the depth of the chest) in adults
– a rate of 100 per minute (almost 2 per second)
- time of compression equal to relaxation – allow complete relaxation

CPR with one rescuer

2 respiratory inflations to 30 cardiac compressions.

CPR with two rescuers

2 respiratory inflations to 30 cardiac compressions

Co-ordinate the inflation with the relaxation phase of compressions. Pause the compressions to allow the inflations.

Continue ventilation and chest compressions at a 30 to 2 ratio until the patient recovers.

Do not stop resuscitation to check for signs of life. If the victim recovers it should be obvious.

If multiple rescuers are present, rotate chest compression duty every 2 minutes as it is very tiring.

If rescuers are unable or unwilling to give expired air resuscitation, chest compressions alone are better than nothing.

Duration of CPR

Continue CPR until:

signs of life return
more qualified help arrives
rescuers are exhausted.
an authorized person pronounces the patient dead.

Recovery checks

Frequent recovery checks (stopping cardiac compressions to feel the pulse) are no longer considered appropriate. Interruptions to cardiac compressions result in a poor outcome. Lay people often have difficulty feeling a pulse. Continue CPR until there are signs of life.

D – AUTOMATIC DEFIBRILLATION

If an AED (Automated External Defibrillator) is available attach it and follow the prompts.

Some causes of cardiac arrest (e.g. ventricular fibrillation) can be fixed by giving an electrical shock through the chest.

An AED can analyse the cardiac electrical activity and automatically give a shock if appropriate.

If required, the AED should be used as soon as possible after initial CPR.

CPR should be continued until the AED is attached.

If the device detects a shockable rhythm and delivers a shock, CPR should be continued for a further two minutes - the heart takes a few minutes to recover its strength. The machine should then be used to recheck the rhythm.

AUSTRALIAN RESUSCITATION COUNCIL <http://www.resus.org.au/>

BASIC LIFE SUPPORT (**BLS)**

Using the mnemonic, **DRABCD** stands for:

D Danger – always check the danger to you, any bystanders and then the injured or ill person. Make sure you do not put yourself in danger when going to the assistance of another person.

R Response – is the person conscious? Do they respond when you talk to them, touch their hands or squeeze their shoulder?

A Airway – can the person breathe? Is their airway clear?
If the person is responding, they are conscious and their airway is clear. Assess how you can help them with any injury.
If the person in not responding, they are unconscious. You need to check their airway by opening their mouth and having a look inside. If the mouth is clear, tilt their head gently back (by lifting their chin) and check for breathing. If the mouth is not clear, place the person on their side, open their mouth and clear the contents. Then tilt the head back and check for breathing.

B Breathing – check for breathing by looking for chest movements (up and down). Listen by putting your ear near their mouth and nose. Feel for breathing by putting your hand on the lower part of their chest or against your cheek.

C CPR (cardiopulmonary resuscitation) – if the person is not breathing, place them flat on their back. Tilt their head back gently by lifting their chin. Pinch their nostril closed, place your open mouth firmly over their open mouth and give them 2 breaths – one after the other. Stop. Check for signs of life. If there are none, commence compressions on the chest. Place the heel of one hand in the centre of their chest and your other hand on top. Press down firmly and smoothly (compressing to one-third of their chest depth) 30 times. Give 2 breaths. Keep going until medical assistance arrives.

D Defibrillator – use a portable defibrillator (if you have one). This is a machine that applies electrical therapy to overcome irregular heart beat (arrhythmia), allowing the heart to re-establish an effective rhythm. Be sure to follow the instructions and picture on the machine and on the package of the pads as well as the voice prompts. If the person responds to defibrillation, follow the DRABCD Action Plan.

Chapter 43

TECHNICAL DIVING

Most recreational SCUBA divers use compressed air in a single cylinder, with a single-hose regulator, to depths up to 30-40 metres, and avoid any decompression staging obligation (although routinely a safety stop is included for dives exceeding 12 metres).

Technical diving is a term used to describe extended diving or where the gas and/or the equipment is different from the original "Aqualung" concept, which only used open circuit, compressed air. The purpose of this technical diving is to extend durations or depths. The simplest modification is the use of a different gas in the scuba cylinder. Then comes the use of several different gases, in multiple cylinders. To reduce the increasing bulk of this diving apparatus, as well as to reduce gas loss, complex rebreathing equipment can be added.

The increasingly complex equipment bestows some advantages, but with increased risks. Some industry representatives and other enthusiastic advocates have promoted this technical diving to the recreational diving community as a safer alternative, enticing others who may not be as adept. It is a more complex and more risky activity than recreational diving and requires expensive equipment and extensive training – two reasons why it appeals to the diving industry.

Most of the diving accidents and deaths which occur in recreational scuba diving, are not due to decompression sickness. Indeed the major causes include the ocean environment (Chapter 6), the stress responses on the individual (Chapter 7), equipment failure or misuse (Chapter 5 & 34) and some diving practices which are especially hazardous – exhaustion of the air supply, buoyancy problems, and failure to follow buddy diving practices (Chapter 34). Technical diving techniques do not reduce and often increase these risks.

The complexity and novelty of technical diving has attracted many established divers, mainly men. Possibly the element of danger and the *avant guarde* nature of the activity combine to offer an enticing challenge, extending diving experiences and excitement.

The technical diver, often studious and attracted to risk taking behaviour, operates on a reduced safety margin but usually with a quiet confidence in his skills and equipment.

He may have commercial interests, being involved in wreck salvage, equipment manufacture, marketing and sales, diver training, or other related activities. An appreciable number of high-profile experts in technical diving have died undertaking this activity, and their deaths have hopefully served to caution many younger and less experienced divers.

DEFINITION

1. USING GASES OTHER than COMPRESSED AIR
   e.g. OXYGEN
   NITROX, (Oxygen Enriched Air),
   HELIOX, (Helium & Oxygen)
   TRIMIX, (Helium, Nitrogen & Oxygen)
2. DECOMPRESSION DIVING
3. DEEP DIVING (> 40 metres)
4. REBREATHINGEQUIPMENT

Technical diving refers to diving in excess of the usual range for recreational scuba divers. This may involve an extension of duration at any depth, the depth itself (in excess of 30–40 metres), changing the gas mixtures to be used, or using different types of diving equipment. All these fall into the realm of technical diving.

It is important, when discussing technical diving, to specify which type, as the relative risk varies from little or no additional risk (compared with recreational diving) to a high one, such as with rebreathing equipment.

Decompression dives and deep diving using only compressed air have added risks and have already been dealt with in previous chapters.

The other risks increase as the gas mixture deviates from normal air and with increased complexity of the equipment.

Diving on 32% oxygen, 68% nitrogen instead of air in a scuba cylinder, to a maximum of 40 metres on a no-decompression conventional air profile, could possibly incur slightly less risk than a recreational scuba air dive.

THE TECHNICAL DIVE

1. DIVE PROTOCOLS, PROFILES and GAS MIXTURES
2. EQUIPMENT COMPLEXITY
3. PHYSIOLOGICAL ASSUMPTION
   — EAD / 0_2 / $C0_2$ / INERT GAS TRANSFER
4. ENVIRONMENTS
5. ACCIDENT & RESCUE IMPLICATIONS

1. DIVE PROTOCOLS, PROFILES and GAS MIXTURES

The diver attempts to select the theoretically ideal gas mixture for the ascent and descent (travel mixes), the bottom (bottom mix) and the decompression staging (usually oxygen or a high oxygen mixture).

The simplest form of technical diving has the diver breathing a mixture of 32% or 40% oxygen (O2) in nitrogen (N2). With the increased O2, there is proportionately less N2. That means less decompression obligation (and less N2 narcosis). For the same decompression risk, the dive can therefore be prolonged and this is highly desirable in some dive trips. The additional risk of oxygen toxicity must be appreciated.

Using a single O2 enriched gas mixture limits the technical diver to shallower dives than with compressed air. A series of gas mixtures in separate cylinders, with diminishing O2 percentages, allows the technical diver to reach greater depths. Substituting helium (He) for N2, either in Heliox (He/O2) or Trimix (N2/He/O2) allows the technical diver to descend further, while avoiding or reducing N2 narcosis. During ascent the changes of gas mixtures is reversed until nearing the surface, when higher O2 percentages may be breathed to expedite the elimination of inert gas (He or N2).

When there are various gas mixtures being breathed, the safe profile of the dive may be very complex and errors may be made in the choice of gas breathed. Nevertheless, using open circuit equipment and several gas mixes, dives to over 100 metres have been safely performed.

The use of rebreathing equipment enormously increases the potential hazards (see later), while attemting to control and monitor the gases breathed and thedecompression required..

2. EQUIPMENT COMPLEXITY

Technical diving involves more complex equipment for producing, supplying and delivering the various breathing gases, other than air. With an increase in the complexity of the equipment there is an associated increase in the likelihood of human error at all these 3 stages.

The handling of mixtures with higher than normal oxygen percentages implies greater risk of fire and explosions. Gas mixtures may not be as compatible as the "normal" oxygen/nitrogen mix in air, and the heat generated during compression must be appreciated. Although not common, explosions associated with high oxygen percentages are very destructive.

Problems and mistakes develop from the use of multiple gases and complex equipment:

- Mixing, labelling and transport of gas;
- Handling it at the dive site;
- Analysing the gases and confirming that they are the ones appropriate for the dive to be performed;
- Selection of appropriate gases during the dive.
- Different gases require different cylinders together with the various attachments; manifolds, O rings, contents gauges, high pressure hoses, and often, separate regulators.

Because of the added complexity of the equipment, the use of multiple gas mixtures and the increased support facilities, there are substantial initial capital outlays, operating and maintenance costs.

3. PHYSIOLOGICAL ASSUMPTIONS

There is considerable doubt regarding some of the physiological assumptions on which technical diving is based. It is claimed that the equivalent air depth (EAD) calculation can be used to determine the different influence of the gas mixture on the diver, and this has been applied to both nitrogen narcosis and decompression sickness (DCS). There is, in fact, no really good evidence that this EAD is a strictly accurate concept. Experience in highly controlled navy diving has been reassuring, and the implication is that the EAD concept is a valid approximate assumption.

Divers using O2/N2 mixtures decompress using tables of EAD. These calculate of the actual partial pressure of N2 for the dive and from this calculate the depth of an air dive that has the same N2 pressure. The diver then decompresses as if he had done an air dive to the calculated depth, the EAD.

Thus a diver breathing 40% O2 at 30 metres (60% of 4 ATA = 2.4 ATA of N2) has an EAD of 20 metres (80% of 3ATA = 2.4 ATA of N2). So, after this dive to 30 metres, our diver decompresses as if he had dived to 20 metres on air.

For the same depth/duration dive, the O2 enriched diver may have less DCS risk and less N2 narcosis, but he will have more risk of O2 toxicity, than the air breathing diver. There are the physiological implications of breathing oxygen at varying partial pressures, as well as increased carbon dioxide retention with both increased oxygen diving and deep diving.

Inadequate factual information is available regarding the physiological interactions of multiple gases. The rate of inert gas transfer between the breathing gases in the lungs, the body tissues and any gas spaces (including decompression bubbles) varies both with the gas and the depth. Thus the selection of different gas mixtures is likely to influence the transfer of inert gases in many ways, far more complex than can be deduced from a simplistic formula. Anyone who doubts this should peruse one of the more sophisticated texts on such topics as nitrogen narcosis and the counter diffusion of gases.

Decompression procedures and algorithms are often unproven. Even with the vast data available on air diving to 40 metres, there are many inexplicable decompression accidents. Adding the vagaries of extended depths and durations, multiple gas mixtures and computer modelling, makes for greater uncertainty in technical diving. Technical divers should question the origin and validity of the decompression schedules they are encouraged to use. Some have had to be altered to reduce their incidence of DCS. The lack of controlled trials have caused some to compare the promoters of these decompression protocols to a pharmaceutical company marketing a drug without testing it and then expecting the consumers to determine the correct dosages.

4. ENVIRONMENTS

The main purpose of technical diving is to extend the environments into which diving is performed. This usually results in an increase in the hazards associated with such environments. The exception is a reduction of the N2 narcosis of deep diving, by the substitution of helium. Most of the other problems with deep diving are aggravated. Not only can the depth or duration of the dive be extended, but so can the actual diving terrain. This is the reason why many wreck divers and cave divers have embraced this activity.

5. ACCIDENT & RESCUE IMPLICATIONS

For the above reasons, the mixed-gas diver often wears a large amount of equipment, complex and bewildering (especially when problems develop during the dive). The likelihood of superimposed equipment problems is thus compounded. Difficulties include those of buoyancy and entrapment. Depth control requires greater discipline and skill as the margin between the "safe" depth and the oxygen toxicity depth, is much reduced. Sometimes a full facemask is indicated so that drowning becomes less likely and rescue more possible.

Because of the different equipment and gases, and the extension of the environments, the procedures for accident management and rescue may have to be altered to take into account the specific problems – such as difficulty in removal of heavy and cumbersome gear. With each variation from the conventional scuba system, there is a price to pay, and a possible modification of the first aid and treatment procedures.

OXYGEN PRESSURE

1. CNS AND RESP. TOXICITY
2. EFFECT ON RECOMPRESSION THERAPY
3. ? NITROGEN NARCOSIS & DCS
4. MIXING & HANDLING DANGER
5. HYPOXIC MIXTURES
6. CO2 BUILD UP
7. EQUIPMENT CHANGES

There is little concern about oxygen toxicity when diving with compressed air within recreational diving limits. Neurological and respiratory oxygen toxicity are virtually impossible. Also, the amount of oxygen exposure is unlikely to significantly influence any recompression treatments that may be needed for decompression accidents. Neither statement can be applied to technical diving.

It had been assumed that oxygen, by virtue of its replacement of nitrogen, would to some degree reduce the severity of nitrogen narcosis and decompression sickness. Although this is possibly so in theory, the scant experimental evidence that there is available, would suggest that oxygen may actually contribute to nitrogen narcosis. The possibility that O2 could contribute to DCS has been proposed, but this is unproven.

Inadequate mixing can result in oxygen pressures being higher or lower than intended. This has implications regarding the safety of the dive profile.

Higher oxygen levels are also likely to interfere with carbon dioxide transport in the blood. This has implications as regards carbon dioxide and oxygen toxicity, nitrogen narcosis and possibly decompression sickness.

The handling of gas mixtures, where oxygen is used or is added to air or other gases, can produce some hazards. Oxygen increases the risk of fire and explosion.

Some divers have observed that O2 aggravates the deterioration of soft materials, such as O-rings and other materials that comprise the diving equipment. It can also accelerate corrosion in cylinders.

See The Technical Dive, above.

OXYGEN ENRICHED AIR
or
NITROX (EANx)

CONFUSION of TERMINOLOGY and JARGON ——— SPECIFY or DIE

40/60 = 40% O2 (EUROPE) or 60% O2 (USA)

NOAA 36% + 32% O2	–	max. 1.6 ATA O2 pressure	
NURC, NC	–	max. 1.45 ATA	"
SWEDEN	–	max. 1.4 ATA	"
DAN (R. VANN)	–	max. 1.2 ATA	"

BUT O2 TOX = ?CNS, ?RESP, ?CO2. Specify which.

"EAD" = EQUIVALENT N_2 PRESSURE

Much of the technical diving now performed involves the use of nitrogen/oxygen mixtures in which the oxygen concentration is greater than that of compressed air. Under these conditions it is very important to specify exactly how much oxygen is being used. Such phrases as 40–60 or 60–40 are not only confusing but often misleading. In Europe 40–60 could imply 40% oxygen, whereas in the USA it is more likely to imply 40% nitrogen.

The actual percentages used in technical diving do vary with different countries and establishments but NOAA in the USA have chosen 32% oxygen and 36% oxygen as their two major mixes. These should not be referred to as Nitrox 1 or Nitrox 2, as this could also be misleading.

The EANx refers to enriched air (nitrox) with the x = oxygen percentage. Thus EAN 32 should mean 32% oxygen and not 32% nitrogen! Do not rely on jargon. Specify the exact mixture, in full.

Any EANx diving has a safe depth range less than air, due to oxygen toxicity.

The oxygen pressures that are considered acceptable vary with different authorities, and in many cases there is confusion between the neurological oxygen toxicity (which can result in nausea, vomiting, seizures, etc.) and respiratory oxygen toxicity, which tends to only occur with prolonged exposure. Also, many of the pressures being quoted in the literature refer to the oxygen pressures observed with rebreathinge quipment, when the carbon dioxide levels have not been measured – complicating considerably the actual cause of symptoms. Most of the work carried out during World War 2 and soon after, failed to measure the carbon dioxide levels and therefore their conclusions regarding safe oxygen limits, are questionable.

NOAA states that the maximum oxygen pressure acceptable is 1.6 ATA. The National Undersea Research Centre in North Carolina recommends 1.45 ATA. The Swedish authorities have recommended 1.4 ATA and Dr Richard Vann of the Divers Alert Network has suggested 1.2 ATA. The US Navy gives a much greater range, and relates it to the duration of the exposures.

The claimed advantages of EANx diving include a probable reduction in decompression sickness incidence, and a possibility of reduced nitrogen narcosis.

On a theoretical basis, presuming nitrogen pressure as the sole cause of nitrogen narcosis, a 20% oxygen mixture (air) at 23 metres could be replaced with 36% oxygen at a depth of 30 metres. to give an equivalent "narcotic effect". Experimental verification for belief in this theory has been sought, but it was unable to be verified.

A common claim is made that there is less post-dive fatigue with EANx than there is with air. This has not yet been verified.

LOW RISK NITROX DIVING

NITROX (EANx) REPLACES AIR. SAME EQUIPMENT
(Same Profile as AIR DIVE). RANGE 15 – 40 metres depth.

ADVANTAGES

1. LESS DCS
2. ? LESS NITROGEN NARCOSIS
3. ? LESS POST DIVE FATIGUE

DISADVANTAGES

1. GAS MIXING PROBLEMS
2. LESS MAX DEPTH (O2 TOXICITY)
3. ? DETERIORATION OF DIVE EQUIPMENT
4. ? MORE CO2 RETENTION

It is possible to use EANx to obtain possible advantages, with relatively few disadvantages, under certain conditions.

In this type of technical diving, the nitrox mixture, usually 32% or 36% oxygen, replaces air, but the same equipment is used and the same decompression profiles permitted, within the 15 – 40 metre range. Others use 28–40% oxygen, the latter with appropriate depth reduction.

It has been claimed that there is deterioration in the dive equipment by using high oxygen mixtures but this has not been verified. It is believed that halofluorocarbon O-rings (e.g. Viton) are less

likely to oxidise and have a higher ignition point – and are thus frequently preferred by technical divers.

It is likely, because of the higher oxygen levels inhaled that there will be a concomitant degree of carbon dioxide retention, based on the common and competitive pathways for the transfer and transport of these gases.

HIGHER RISK NITROX DIVING

NITROX (EANx) REPLACES AIR — PROFILE AS FOR E.A.D.

ADVANTAGES RANGE 15 – 40 metres

1. INCREASED DURATION of NO-DECO DIVE
   or LESS DECO STOPS
   or GREATER DURATION/DEPTH of DIVE for SAME DECO
2. DECO VALUE — IF AIR STOPS FOLLOWED (LESS N_2)

DISADVANTAGES

1. GAS MIXING, HANDLING & CORRECT USAGE
2. MAX DEPTH LIMITED (O2 TOXICITY)
3. ?ALTERATION OF DCS & RECOMPRESSION THERAPY
4. ? DYSBARIC OSTEONECROSIS (SLOW TISSUES AFFECTED BY LONGER DIVES)

In this type of diving (EANx) the profile of the dive is altered to make allowance for the high oxygen, lower nitrogen levels, based on the EAD or similar calculations. Thus the diver is likely to increase the duration of his no-decompression dive, reduce the decompression stops required or increase the duration or depth of the dive for the same decompression time commitment. Whether this calculation is justifiable under all conditions, has yet to be demonstrated.

The probable only genuine safety advantage of this kind of diving occurs if "air stop" times are followed during decompression, whilst using EANx.

There is a possibility of an increased risk of decompression sickness, due to the effects of oxygen contributing to this disorder, or because of the use of untested algorithms used in commercial nitrox decompression profiles. The "bent" diver is also more likely to have had a higher oxygen dose, contributing to respiratory damage during the recompression therapy, than his air breathing colleague.

There may well be an alteration in the type of decompression sickness sustained with this form of diving because of the increased duration that it frequently entails. Thus the slower tissues are more

likely to be affected, and this should be considered during the subsequent recompression therapies, and of a possible increased susceptibility to dysbaric osteonecrosis.

HIGH RISK, HELIUM DIVING

LESS DENSE, LESS SOLUBLE, FASTER DIFFUSION, HEAT CONDUCTIVITY

ADVANTAGES

1. LESS NARCOSIS — GREATER DEPTH
2. LESS BREATHING RESISTANCE — GREATER DEPTH
3. REDUCED CO2 RETENTION
4. LESS DECO (for LONGER DIVES)

DISADVANTAGES

1. DEEPER DIVING
2. MORE DECO (for SHORT DIVES)
3. HEAT LOSS (ENVIRONMENT, ? RESPIRATORY)
4. VOICE DISTORTION 5. MIXING 6. HPNS

There are significant differences in the way the body handles helium and nitrogen. Both are inert gases, but helium is much less dense and is also less soluble in some tissues than nitrogen. It does, however, have a much greater speed of diffusion and also conducts heat more rapidly.

The real advantage compared to nitrogen is that it does decrease the incidence of nitrogen narcosis. For dives in excess of 30 – 40 metres, the risks of nitrogen narcosis can be proportionately decreased as helium replaces nitrogen. It thus tends to be used for dives of greater depths. An additional factor is the reduction in breathing resistance due to its decreased density and other factors, also allowing dives to greater depths.

The effects on decompression likelihood are more complicated. It is probably likely to produce less decompression requirement for the longer dives, but may well require more decompression for shorter dives. Many of the helium and Trimix decompression tables are less well validated than the air tables, and herein lies a major difficulty with helium diving.

The main aggravating problem is that the divers are diving deeper with helium and Trimix than with compressed air, and therefore are exposed to all the associated problems of depth (other than nitrogen narcosis and breathing resistance). Barotrauma and DCS risks are aggravated. The environmental difficulties associated with depth include poor visibility, buoyancy implications, excess gas consumption, stress factors and the increased problems with first aid, rescue and resuscitation.

There is also a greater conductive heat loss from helium, even though there is some question regarding the respiratory heat loss. Heliox feels colder to breathe, and in a helium environment the heat is lost more rapidly. Increased depth also aggravates heat loss.

Voice distortion can produce communication problems. At greater depths the high pressure neurological syndrome (HPNS) also becomes relevant.

The difficulties with mixing gases, referred to above, are also present with helium and are complicated by the different compressibility of helium, as well as the risk of ascending with low oxygen pressures – which are commonly used with deep helium diving.

Comparison with the commercial deep divers is noteworthy. These experts usually require a surface supply of gas, full facemasks, communication systems, a standby diver, a wet bell and a recompression chamber on site. Experience has demonstrated the need for these. The less trained amateurs appear to have no such requirements.

VERY HIGH RISK. RE-BREATHERS or CIRCUIT SETS

Rebreathing equipment has been in use for more than a century, causing many deaths and cases of unconsciousness. Despite the recent electronic mechanisms, the essential problems of rebreathing equipment remain. It is very much a high risk strategy to employ for specific reasons, by professionals.

The value of rebreathing equipment is that it produces fewer bubbles, and is therefore quieter. This is of use both in clandestine military operations and for marine photography. It is more economical on gas, as the gas is recycled through the diving equipment, in a "circuit". It can also be constructed with low magnetic materials, which are useful if one is working around magnetic mines.

The disadvantage that is inherent in all types of rebreathers is the failure of the carbon dioxide absorbent system to work effectively under all diving conditions. This may occur for various reasons. One is an inappropriate canister design. The early absorbent canisters were inadequate for maximal exertion. It is surprising how few improvements the manufacturers have included in some of the carbon dioxide absorbent canisters in the sets now being promoted. Also, the absorbent itself is not always reliable. It frequently varies in efficiency, and each absorbent batch needs to be tested. This is not feasible for the individual diver. The handling and storing of absorbent may result in deterioration in efficiency, as will the degree and type of wetting that may occur.

When diving in sea water, hypertonic saline can enter the system, causing a great reduction in efficiency. The absorbent itself, when combined with carbon dioxide, produces water as a by-product, which can also influence the efficiency. Water traps are incorporated in some sets.

REBREATHERS

ADVANTAGES

SILENT, ECONOMICAL, +/- MAGNETIC

DISADVANTAGES

1. CO2 TOXICITY
2. DILUTION HYPOXIA, HYPEROXIA
3. CAUSTIC COCKTAIL
4. INITIAL and MAINTENANCE EXPENSE

OXYGEN RE-BREATHERS :
DEPTH LIMIT 8 – 9metres

CONSTANT FLOW. O2 = FLOW vs. ENERGY

OXYGEN MONITORS = FAILURE. DCS?

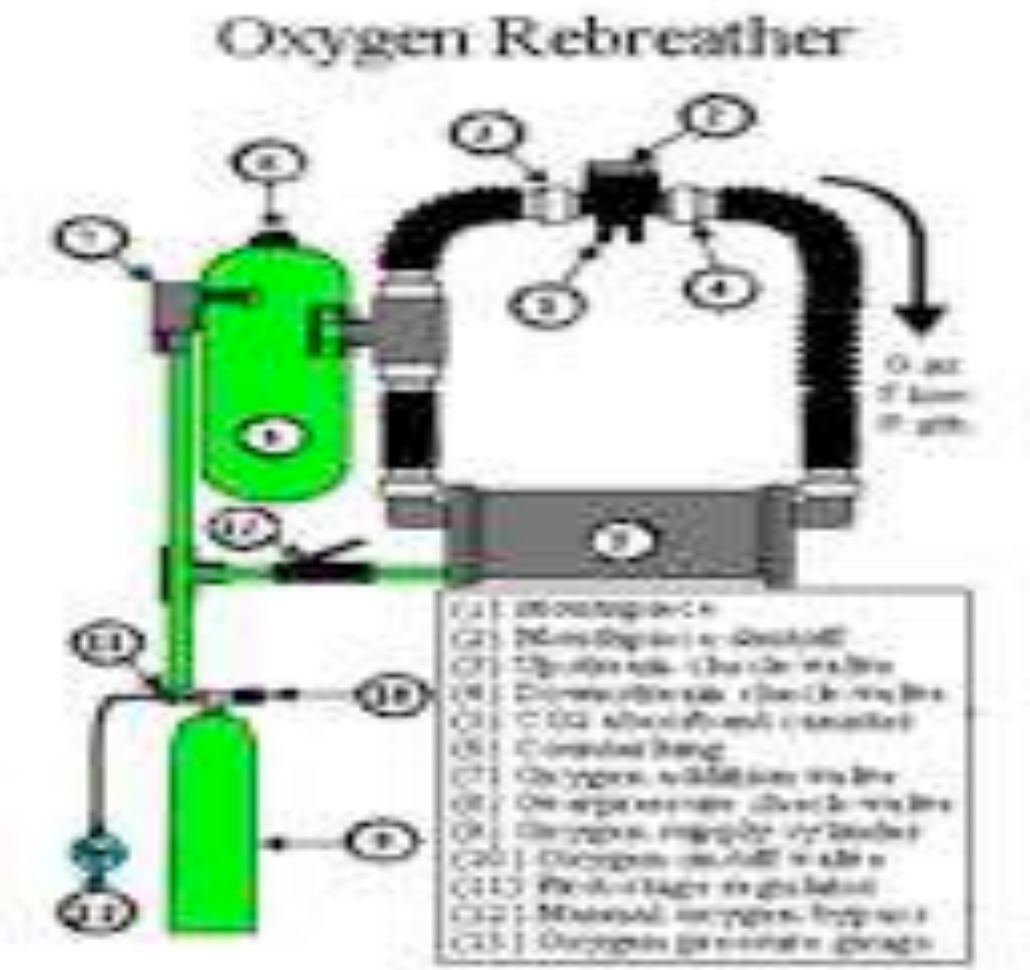

Fig. 43.1

The carbon dioxide absorbent must be packed correctly into the canister. This is an acquired skill and requires training. The density of packing influences the efficiency. Lower temperatures also reduce the efficiency of the absorbent. Some absorbents can be pre-pachaged.

Often absorbent canisters will work very well at a moderate work load, but when exertion is required, the absorbent canister will eventually fail – especially if it has been in use for a considerable time.

The manufacturers' claims regarding the safe duration of carbon dioxide absorption in their diving equipment are usually very optimistic, and do not apply to emergency situations where the diver is exerting himself maximally (such as when swimming against a current, or trying to rescue and tow a companion – even on the surface).

When water gets into the rebreathing set, it may collect some of the alkali from the absorbent and enter the divers mouth and lungs, which can be very unpleasant. This is called a "caustic cocktail".

A rebreathing set can cause dilution hypoxia, usually by incorrect technique and failure to "clear the set" (and the lungs) of the inert gas. This is more likely when the supply gas is on demand, compared to the old fashioned constant flow sets. It can also occur if there is a small amount of inert gas in the gas cylinder, and especially so when there is a considerable amount of nitrogen or helium, such as with nitrox, heliox or trimix diving. It may be induced by an incorrect mix, a leak from or obstruction to the inflow, or low cylinder pressure. It can even occur in 100% O2 sets, especially those that supply gas on demand.

Sometimes the hypoxia will only be noticed during ascent. A lower oxygen percentage at depth may translate to a dangerously low oxygen partial pressure nearer the surface.

Re-breathers require specialised diving protocols, when rescue and resuscitation are needed. It is not just a matter of removing a mouthpiece and replacing it with another. Companion diver drill needs to be tailored for each type of re-breather.

The problems of gas mixing and handling, as described above, also relate to this equipment.

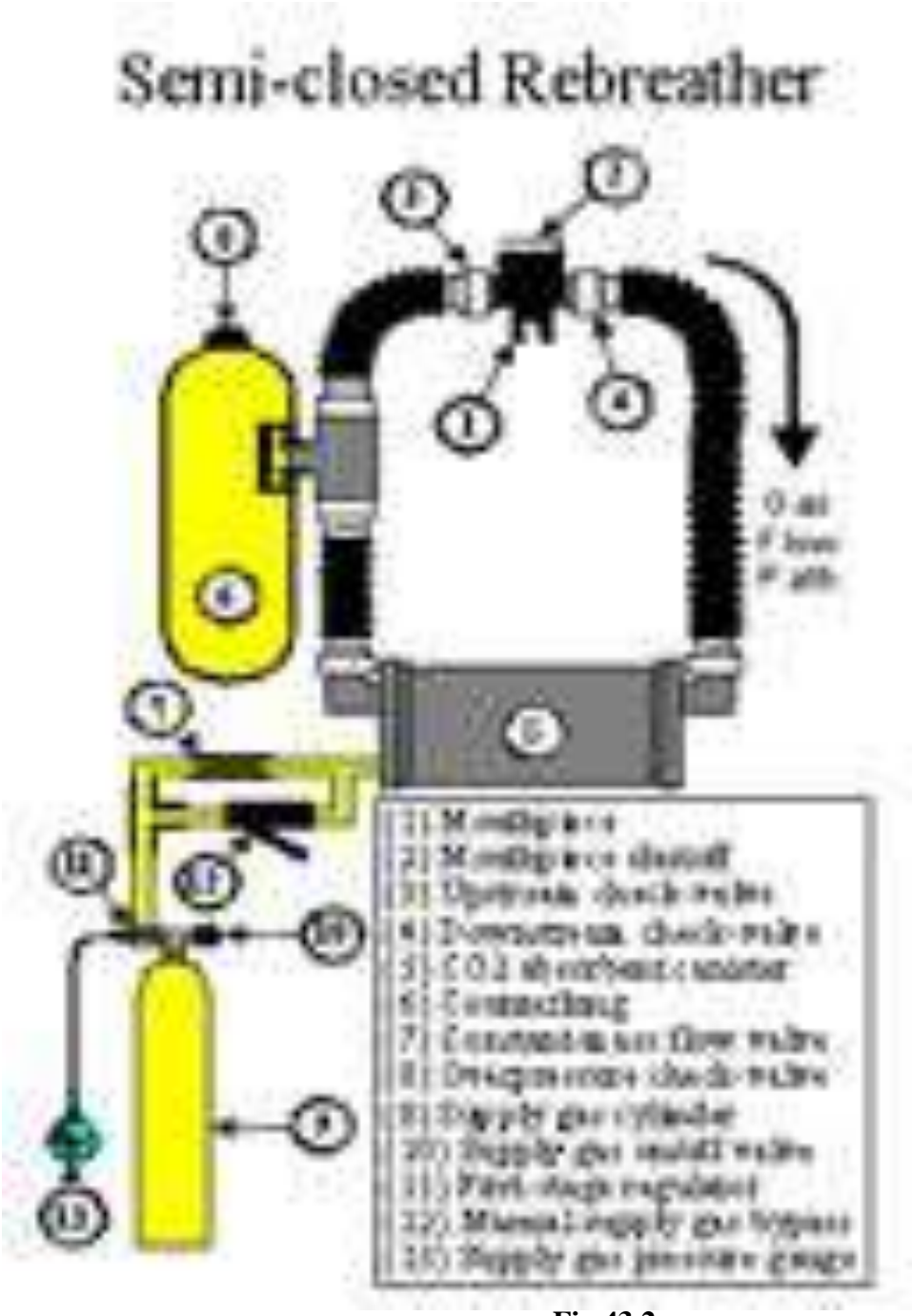

Fig 43.2

• The oxygen rebreathers are closed circuit sets, used to a maximum depth of about 8-9m, are usually restricted to specialised navy divers. They have resulted in many cases of unconsciousness and death.

Occasionally photographers use this equipment, but they would be considered unwise to do so. The companion rescue drill is often required and marine photographers are not gregarious beasts.

• Some rebreather sets have a constant flow of nitrox, heliox or trimix gas. They are usually semi-closed circuit sets. With these, the oxygen level in the breathing bag or inspiratory tube may vary according to two major factors. The first is the flow of oxygen into the set, and the second is the amount lost from the set. The inspiratory oxygen range can be a variable quantity, and should be designated pre-dive. The relevant factors determining inspiratory oxygen include:

– the volume and mixture of the incoming gas
– the energy utilised in metabolism (oxygen consumption)
– the volume and mixture of gas released as bubbles (e.g. with ascents).

The interaction between the input and output of oxygen will result in a variable oxygen percentage and ascent or descent will then determine the oxygen pressure. These sets are especially likely to cause dilution hypoxia and hypoxia of ascent.

As hypoxia usually produces no warning prior to it causing unconsciousness, the use of constant flow rebreathing sets would be considered unwise. Close attention to the cylinder pressure, ensuring an adequate inflow of gas, and a replacement with fresh gas prior to ascent (a "flush-thru") is essential

• The more expensive closed circuit mixed gas rebreathing sets use analysers and solenoids to measure and control the oxygen pressures during the dive and a feed back system adds oxygen or a diluent gas (nitrogen, helium, mixtures) as required – to ensure that the oxygen partial pressure remains within a certain range. This equipment is extremely expensive, often not reliable and should only be used by those with faith in this technology.

N.B. Anyone who uses a rebreather should be aware of the much greater risk of unconsciousness. Without a full facemask, this problem usually converts to subsequent drowning and death.

Closed-Circuit Rebreather

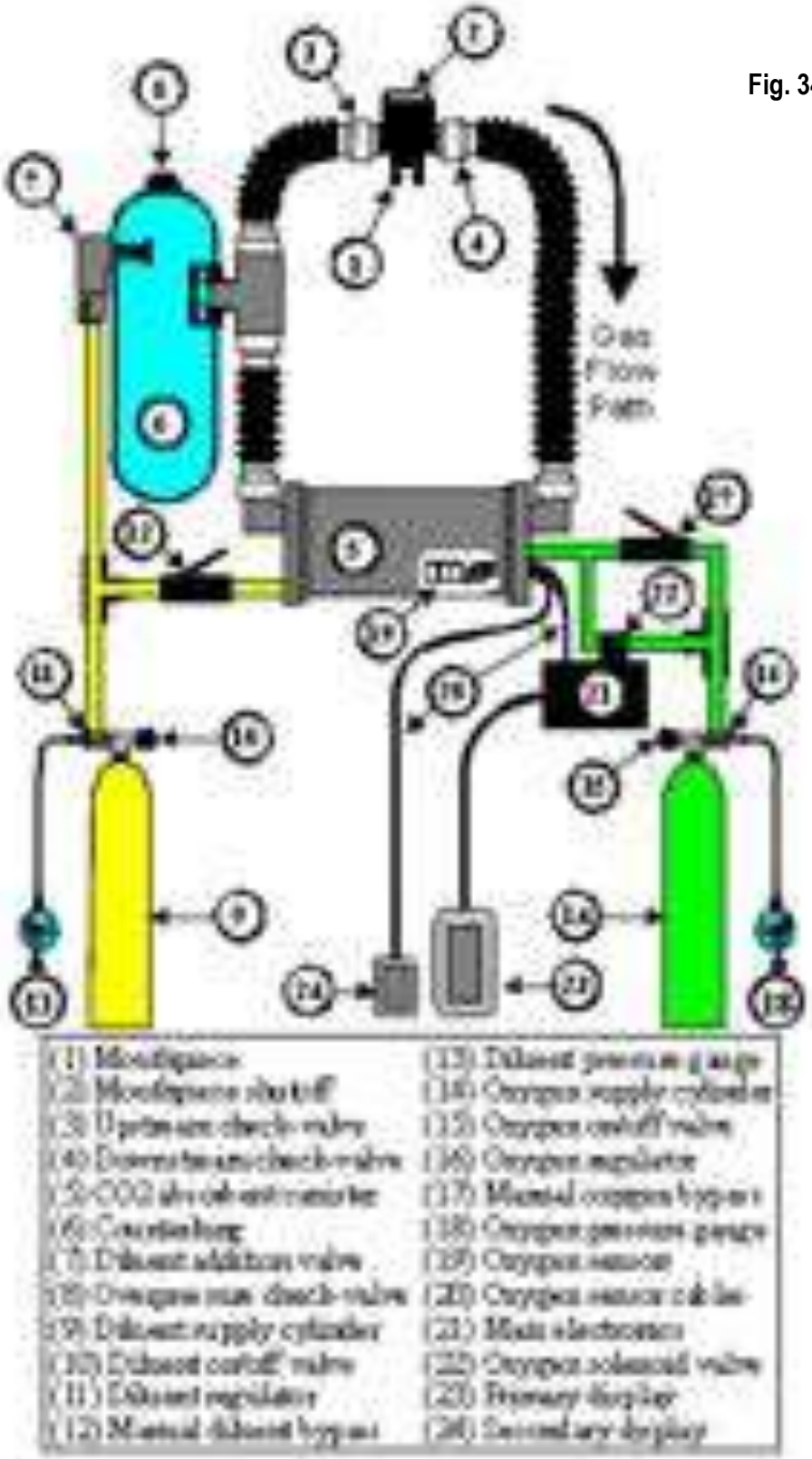

Fig. 34.3

CONCLUSION

There are few problems in understanding the general concepts of technical diving, using different gas mixtures, different equipment and extending limits beyond those of recreational diving. The principles are relatively simple. The devil is in the application.

Perhaps the most important thing about technical diving is to realise that the majority of the diving deaths that occur in recreational divers occur for reasons which will be aggravated by the use of more complex equipment, in more hazardous environments. Technical diving is therefore, by its very nature, likely to have greater risks than normal recreational diving, other factors being constant.

The margin for error in this type of diving is appreciably less, and therefore it should only be employed by divers with enormous experience, detailed training and meticulous attention to equipment selection, maintenance and use. The advocates of technical diving tend to lay great stress on certain aspects of safety - which are relatively unimportant compared to the others referred to in Chapter 34.

To overcome some potential equipment related problems, technical divers may stress the need for redundancy, redundancy, redundancy. Complexity is an unintended accompaniment.

They will stress the importance of decompression sickness, and the physiological advantages of oxygen, but may ignore the more frequent causes of diving deaths, such as exhaustion of gas supply, buoyancy problems, stress responses, etc. They will also tend to ignore the areas in which the "technical advances" have been meagre e.g. the efficiency of carbon dioxide absorbents, to focus in preference on high-tech oxygen sensors and theoretical decompression algorithms.

The leaders in this field will be experienced, highly skilled, very fit, entrepreneurial divers often with a high public profile. It is not for occasional, pleasure seeking divers who anticipate a relaxed, hassle free, unencumbered dive and an automatic expectation of survival at the end of the day.

Appendix A

DIVING MEDICAL LIBRARY

DIVING TEXTS

- **Physics, Physiology and Decompression Theory** for the Technical and Commercial Diver – Bruce Wienke
- **Deeper Into Diving** – John Lippman and Simon Mitchell
- **High Altitude Diving** - Bruce Wienke
- **Scuba Equipment Care and Maintenance** - Farley and Royer
- **Oxygen and the Scuba Diver** - Hendrick and Thomson
- **Oxygen First Aid** – John Lippmann
- **The Application of Enriched Air Mixtures** - Betts
- **Sport Diving in Depth** - Griffiths
- **The Physics and Engineering of Diving** - Dickens
- **The Underwater Handbook,** a Guide to Physiology and Performance for the Engineer (1976) — by C. W. Shilling, M. F. Werts & N. R. Schandolmeier.
- **Mastering Rebreathers -** Jeff Bozanic
- **Nitrox Manual -** Dick Rutkowski
- **Recreational Nitrox Diving -** Robert N. Rossier
- **Technical Diving In Depth -** Bruce R. Wienke
- **Snorkeling and Free Diving -** Tom Scott, Jim Flagg
- **Cold Water Diving -** John N. Heine
- **Scuba Diving -** Dennis K. Graver

DIVING MANUALS

• U S Navy Diving Manual, Revision 6, (Vol 1 Air Diving) - Naval Sea Systems Command, Supervisor Of Diving. Free download on internet
• British Subaqua Club (BSAC) Diving Manual http://www.bsac.com/
• National Oceanic and Atmospheric Administration (NOAA) Diving Manual
• Professional Association of Diving Instructors (PADI) – Open Water Diving Manual.
• National Association of Underwater Instructors (NAUI) – Diving Manual.

NEWSLETTERS

• **Undercurrent** – this doubles both as a consumer guide to holiday dive sites, as well as a forum for the presentation of recent technical data, diving accidents, etc. :
Address : PO BOX 1658 Sausalito, California 94965, U.S.A.

• **Alert Diver** – a quarterly newsletter of the Divers Alert Network. An excellent informative newsletter regarding diving medical safety and accident information, produced by DAN :
Address : BOX 3823 Duke University Medical Centre, Durham, North Carolina 27710, U.S.A.

• **Pressure** – a bi-monthly newsletter of the Undersea and Hyperbaric Medical Society, with information of a social, academic and educational nature, relevant to the Society and its members – mainly physicians and paramedics:
Address : Undersea and Hyperbaric Medical Society 9650 Rockville Pike, Bethesda Marylands 20814, U.S.A.

• **Diving and Hyperbaric Medicine Journal** – a bi-monthly journal/newsletter of the South Pacific Underwater Medical Society and the European Undersea Biomedical Society. This contains both original diving medical articles and reviews, and also summaries of other diving medical research conducted throughout the world:
Address**: SPUMS, c/- ANZCA, 630 St Kilda Rd, Melbourne Vic, 3004, Australia**
email spumsadm@bigpond.net.au

- **Undercurrent.** A USA newsletter with excellent independent information on equipment and dive travel sites**.**
  http://www.undercurrent.org/

- **X-Ray Mag. A free internet magazine.** More into ecology, marine animals, photography, etc. URL: http://www.xray-mag.com/

- **Advanced Diver.** More into high tech diving and equipment.
  http://www.advanceddivermagazine.com/

NB. Many other countries have their own variations of the above newsletters, from local institutions.

DIVING MEDICAL TEXTS

- ***Diving Medicine – for Scuba Divers —** (2010) by Carl Edmonds, Bart McKenzie, John Pennefather and BobThomas .
- ***The DAN Emergency Handbook** — by John Lippmann & Stan Bugg.
- ***Dangerous Marine Creatures *** — by Carl Edmonds,
- ***** **Oxygen First-Aid for Divers** — by John Lippman, J.L. Publications,

• **Diving And Subaquatic Medicine (4th Edition)** Carl Edmonds, Christopher Lowry, John Pennefather, And Robyn Walker. Arnolds Publ.

• **Bove and Davis' Diving Medicine**— A.A. Bove by, published by Saunders, Philadelphia.

- **Bennett and Elliott's Physiology & Medicine Of Diving-5th Edition -** Alf Brubakk, Tom Neuman

• **The Physicians Guide to Diving Medicine** (1984) — by Shilling, Carlston and Mathias. Published by Plenum Press, New York.

• **Diving Accident Management** (1990) — the 41st Undersea and Hyperbaric Medical Workshop, edited by Bennett and Moon, and published by the Undersea and Hyperbaric Medical Society,

- **Assessment Of Diving Medical Fitness For Scuba Divers And Instructors -** Peter B. Bennett, Frans J. Cronje, Ernest S. Campbell
- **Encyclopedia Of Underwater Investigations -** Robert G. Teather

*** = Suitable for recreational divers**

NOTE Most of the above can be acquired over the internet, and most texts can be obtained from:

Amazon.com

Best Publishing

Divers Alert Network.

Some can be accessed from the internet, free. (such as the US Navy Diving Manual) and

Diving Medicine – for Scuba Divers — (2010) by Carl Edmonds, Bart McKenzie, John Pennefather and BobThomas . www.divingmedicine.info

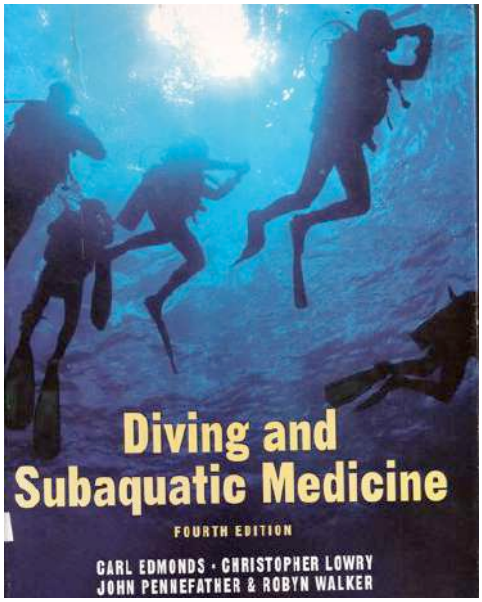

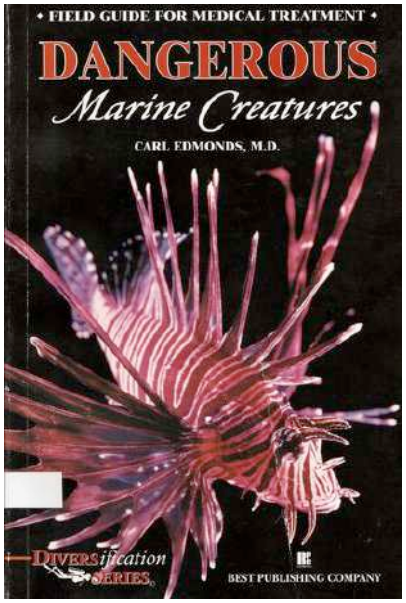

DIVING MEDICAL REFERENCES

Apart from the Diving Medical Societies (see Appendix B) and in the medical texts referred to above, the best source of diving medical references is:

http://rubicon-foundation.org

In the past there have been excellent sources of references and reports. They include

A Bibliographical Sourcebook of Compressed Air, Diving and Submarine Medicine. Vols 1, 2 and 3. U.S. Govt publication.

Key Documents of the Biomedical Aspects of Deep-Sea Diving, Vols 1-5. Undersea Medical Society publication.

Underwater Medicine and Related Sciences. A Guide to the Literature. Vols 1-6. Undersea Medical Society publication.

It is possible to conduct research without access to these publications, but I do not see how. CE

Appendix B

INFORMATION CONTACTS

SOCIETIES

Apart from the instructor organisations (NAUI, PADI, SSI, YMCA, BSAC, CMAS, IANTD, etc.) there are certain societies which would be very useful for any diving paramedic to belong to.

These include the following :

• **Divers Alert Network (DAN)** — BOX 3823 Duke University Medical Centre, Durham, NC 27710 (USA and the American continent, offshore and USA islands and trust territories). www.diversalertnetwork.org

• **Divers Alert Network, Asia Pacific(DAN.AP). P.O. Box 384, Ashburton, Vic, Australia.** www.danasiapacific.org Email: info@danasiapacific.org

• **Undersea and Hyperbaric Medical Society** —Mailly American. www.uhms.org

• **South Pacific Underwater Medicine Society (SPUMS)** — C/- The Australian and New Zealand College of Anaesthetists. Mainly the Indo-Pacific region, centred in Australia). www.spums.org.au

- **European Underwater and Biomedical Society.** www.eubs.org

• **British Subaqua Club (BSAC)** — (mainly around the U.K. and Europe). www.bsac.org

• CMAS Confédération Mondiale des Activités Subaquatiques. Mainly European. http://www.cmas.org

- IANTDAn organisation that publishes information for technical divers. http://www.iantd.com/iantd2.html

See **Appendix D** for all **DAN** regional offices and emergency contacts

EMERGENCY TELEPHONE NUMBERS

Your Area : _____
_____
_____

Telephone :

_____
_____
_____

Telephone :

_____
_____
_____

Telephone :

Others : _____
_____
_____

Telephone :

Others : _____
_____
_____

Telephone :

Others : _____
_____
_____

Telephone :

Others : _____
_____
_____

Telephone :

Others : _____
_____
_____

Telephone :

Appendix C

IN–WATER O_2 RECOMPRESSION THERAPY

AUSTRALIAN UNDERWATER OXYGEN TABLE

Notes :

1. It is recommended that the application of this therapy technique usually be limited to trained and experienced diving medical practitioners and paramedics.

2. This technique may be useful in treating cases of decompression sickness in localities remote from recompression facilities. It may also be of use while suitable transport to such a centre is being arranged.

3. In planning, it should be realised that the therapy may take up to 3 hours. The risks of cold, immersion and other environmental factors should be balanced against the beneficial effects.

4. The diver must be accompanied by an attendant at all times.

❑ Equipment.

The following equipment is essential before attempting this form of treatment :

1. Full face mask with demand valve and surface supply *or* helmet with free flow.
2. Adequate supply of 100% oxygen for patient, and air for attendant.
3. Wet suit for thermal protection.
4. Shot with at least 10 metres of rope (a seat or harness may be rigged to the shot).
5. Some form of communication system between patient, attendant and surface.

❑ Method.

1. The patient is lowered on the shot rope to 9 metres, breathing 100% oxygen.
2. Ascent is commenced after 30 minutes in mild cases, or 60 minutes in severe cases, if improvement has occurred. These times may be extended to 60 minutes and 90 minutes respectively if there is no improvement.
3. Ascent is at the rate of 1 metre every 12 minutes.
4. If symptoms recur, remain at depth a further 30 minutes before continuing ascent.
5. If oxygen supply is exhausted, return to the surface, rather than breathe air underwater.
6. After surfacing, the patient should breathe 1 hour on 100% oxygen, one hour off, for a further 12 hours.

Fig. 1 App C. Breathing intermittent O2 after treatment

Fig. 2 App C. An Underwater O2 unit

Fig. 3 App C. A cheap full face mask. *Cressi-Sub*

Appendix D

DIVERS ALERT NETWORK

(DAN)

The Divers Alert Network (DAN) is a group of not-for-profit organisations dedicated to improving diving safety for all divers. Regional DAN organisations have been established throughout the world to provide safety services to divers. These include DAN America, DAN Europe, DAN Asia-Pacific (AP), DAN Europe, DAN Southern Africa and DAN Japan.

As part of its mission, DAN provides or funds 24-hour diving emergency hotlines throughout the world; offers emergency evacuation cover and dive injury insurance to members; provides diving medical information services; offers quality training in accident management including first aid and oxygen provision; collects data and reports on dive accidents; funds research into dive medicine and dive safety; and provides support for recompression chambers in remote locations.

DAN Asia-Pacific
P.O. Box 384 (49A Karnak Rd)
Ashburton, Victoria 3147
AUSTRALIA
Tel +61-3-9886 9166
Fax +61-3-9886 9155
info@danasiapacific.org
www.danasiapacific.org
Diving emergencies:
DES Australia
within Australia 1800-088200
outside Australia +61-8-8212 9242
DES New Zealand 0800-4DES 111

DAN AP Korean Hotline
010-4500 9113
DAN AP Chinese Hotline
+852-3611 7326

DAN America
6 West Colony Place Durham, N.C. 27705 USA
Tel +1-919-684-2948
Fax+1-919-490-6630
dan@DiversAlertNetwork.org
www.diversalertnetwork.org
Diving emergencies:
DAN America
+1-919-684 8111
DAN America-Mexico
+52-5-629 9800 code 9912935

DAN Europe
P.O. Box DAN
Roseto 64026 ITALY
Tel +39-085-893-0333
Fax +39-085-893-0050
mail@daneurope.org
www.daneurope.org
Diving emergencies:
+39-06-4211 8685

DAN Japan
Japan Marine Recreation Association Kowa-Ota-Machi Bldg, 2F,
47 Ota-Machi 4-Chome, Nakaku, Yokohama City, Kagawa 231-0011 JAP
Tel +81-45-228-3066
Fax+81-45-228-3063
E-mail dan@danjapan.gr.jp
www.danjapan.gr.jp
Diving emergencies:
+81-3-3812 4999

DAN Southern Africa
Private Bag X 197 Halfway House 1685 SOUTH AFRICA
Tel +27-11-254 1991
Fax: +27-11-254 1993
E-mail mail@dansa.org
www.dansa.org
Diving emergencies:
inside South Africa:
0800-020111
outside South Africa
+27-10-209 8112

Fifth Edition

SCUBA DIVING

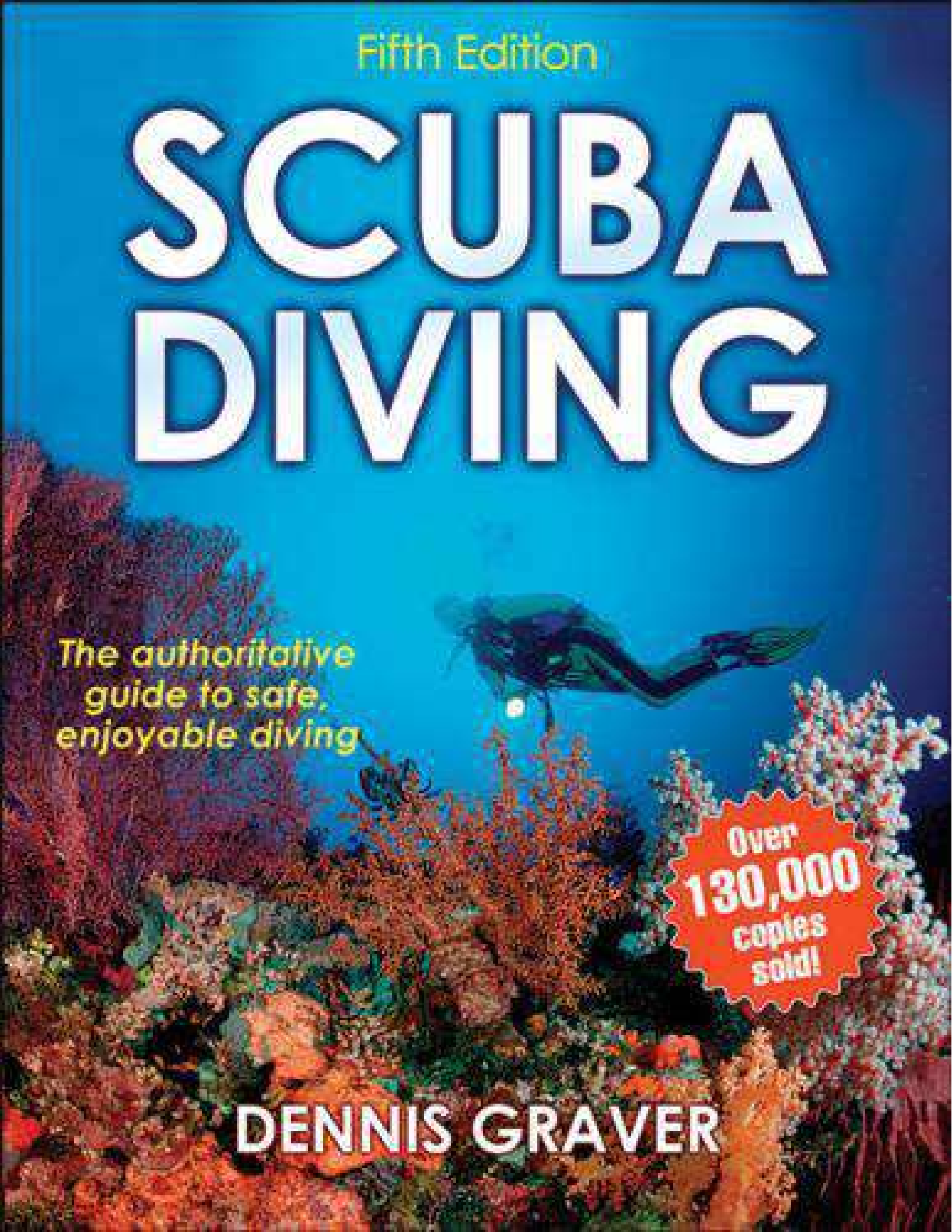

The authoritative guide to safe, enjoyable diving

Over 130,000 copies sold!

DENNIS GRAVER

HUMAN KINETICS

Scuba Diving

FIFTH EDITION

Dennis K. Graver, EMT, SEI

HUMAN KINETICS

Library of Congress Cataloging-in-Publication Data

Names: Graver, Dennis, author.
Title: Scuba diving / Dennis K. Graver, EMT, SEI.
Description: Fifth Edition. | Champaign, IL : Human Kinetics, [2016] |
Includes bibliographical references and index.
Identifiers: LCCN 2016018110 (print) | LCCN 2016031644 (ebook) | ISBN
9781492525769 (print) | ISBN 9781492542032 (ebook)
Subjects: LCSH: Scuba diving.
Classification: LCC GV838.672 .G74 2016 (print) | LCC GV838.672 (ebook) | DDC
797.2/34--dc23
LC record available at https://lccn.loc.gov/2016018110

ISBN: 978-1-4925-2576-9 (print)

Copyright © 2017, 2010, 2003, 1999, 1993 by Dennis K. Graver

All rights reserved. Except for use in a review, the reproduction or utilization of this work in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including xerography, photocopying, and recording, and in any information storage and retrieval system, is forbidden without the written permission of the publisher.

This publication is written and published to provide accurate and authoritative information relevant to the subject matter presented. It is published and sold with the understanding that the author and publisher are not engaged in rendering legal, medical, or other professional services by reason of their authorship or publication of this work. If medical or other expert assistance is required, the services of a competent professional person should be sought.

The web addresses cited in this text were current as of June 2016, unless otherwise noted.

Acquisitions Editor: Tom Heine; **Developmental Editor:** Anne Hall; **Managing Editor:** Nicole Moore; **Copyeditor:** Jan Feeney; **Permissions Manager:** Martha Gullo; **Senior Graphic Designer:** Nancy Rasmus; **Cover Designer:** Keith Blomberg; **Photograph (cover):** Dirscherl Reinhard/Prisma/age fotostock; **Photographs (interior):** Dennis Graver, unless otherwise noted; **Photo Asset Manager:** Laura Fitch; **Visual Production Assistant:** Joyce Brumfield; **Photo Production Manager:** Jason Allen; **Art Manager:** Kelly Hendren; **Associate Art Manager:** Heidi Richter; **Illustrations:** © Human Kinetics; **Printer:** Walsworth

Human Kinetics books are available at special discounts for bulk purchase. Special editions or book excerpts can also be created to specification. For details, contact the Special Sales Manager at Human Kinetics.

Printed in the United States of America 10 9 8 7 6 5 4 3 2 1

The paper in this book is certified under a sustainable forestry program.

Human Kinetics
Website: www.HumanKinetics.com

United States: Human Kinetics
P.O. Box 5076
Champaign, IL 61825-5076
800-747-4457
e-mail: info@hkusa.com

Canada: Human Kinetics
475 Devonshire Road Unit 100
Windsor, ON N8Y 2L5
800-465-7301 (in Canada only)
e-mail: info@hkcanada.com

Europe: Human Kinetics
107 Bradford Road
Stanningley
Leeds LS28 6AT, United Kingdom
+44 (0) 113 255 5665
e-mail: hk@hkeurope.com

Australia: Human Kinetics
57A Price Avenue
Lower Mitcham, South Australia 5062
08 8372 0999
e-mail: info@hkaustralia.com

New Zealand: Human Kinetics
P.O. Box 80
Mitcham Shopping Centre, South Australia 5062
0800 222 062
e-mail: info@hknewzealand.com

E6733

Scuba Diving

FIFTH EDITION

Contents

Foreword vii
Acknowledgments ix

1 Diving Into Scuba 1

The Joys of Diving 2
How Diving Evolved 3
The Diving Community 3
Diver Training 5
Proficiency Testing 5
Diving Prerequisites 5
Diving Risks 6
Selecting a Dive Course 8
Diving Responsibilities 9

2 Diving Science 11

Anatomy for the Diver 12
Respiration and Circulation 15
Contrasts in Air and Water Environments 19
Gas Laws 23
Air Consumption 28
Heat, Humidity, Light, and Sound 28

3 Diving Adjustments 37

Thermal Adjustments 38
Buoyancy Adjustments 41
Pressure Equalization 43
Breathing Adjustments 46
Heart and Lungs 50
Equilibrium Adjustments and Seasickness 52
Visual Adjustments 53
Ingassing and Outgassing 53
Dehydration Prevention 58
Mobility Adjustments 59

4 Diving Equipment 63

Basic Equipment for All Divers 65
Exposure Suits 71
Weighting Systems 78
Buoyancy Compensators 80
Scuba Cylinders 83
Cylinder Valves 87
Scuba Regulators 89
Alternate Air Sources 94
Instrumentation 96
Dive Knives and Accessories 100
Specialized Equipment 104

5 Diving Environment 109

Aquatic Biology 110
Potentially Dangerous Aquatic Animals 114
Conservation and Preservation 117
Aquatic Conditions 121
Boat Diving 132

6 Diving Skills 137

Preparing to Skin Dive 138
Skin Diving Skills 141
Preparing to Scuba Dive 148
Scuba Diving Skills 152
Managing Physiological Problems 169
Managing Physical Problems 173
Managing Emergencies 178

7 Dive Planning 183

Dive-Planning Factors 184
Advance Planning and Preparation 185
Short-Term Planning and Preparation 187
On-Site Planning and Preparation 187
Area Orientations 189
Dive Profile Planning 190
Dive Tables 191
Dive Calculators and Dive Computers 204
Postdive Review 206

8 Diving Opportunities 209

Continuing Education 210
Local Opportunities 212
Dive Travel 213
Dive Destinations 216
Career Opportunities 218
Leadership Opportunities 219

Appendix A: Associations and Periodicals 221
Appendix B: AOK Answers 223
Glossary of Scuba Diving Terms 231
Bibliography 239
Index 241
About the Author 245

This page intentionally left blank.

Foreword

Recreational scuba diving has progressed from its beginnings in the 1950s. Dennis Graver's fifth edition of *Scuba Diving* is uniquely designed toward a level of education that provides new divers with an exceptional level of knowledge while promoting safety and enjoyment. With our training agencies, Scuba Educators International (the former YMCA scuba program) and PDIC International, our goal is to provide a full education for new divers. This text accomplishes that goal and perhaps can also be used for advanced programs as well.

At a time when instant gratification leads many to seek recreational activities that can be accomplished with little effort in a brief time, scuba takes new divers into an environment that remains challenging. Therefore, the need for a full education has not changed. Perhaps the delivery of information will change in the future; however, the need to understand the complexities of the undersea world will not.

In addition to bringing this fifth edition up to current levels of information, Dennis has provided end-of-chapter questions that result in additional learning. This edition is among the best of texts in the scuba industry for new divers.

Dave Barry wrote, "There is nothing wrong with looking at the surface of the ocean itself, except that when you finally see what goes on underwater, you realize that you've been missing the whole point." I couldn't agree more.

Tom Leaird, CEO
Scuba Educators International
PDIC International

This page intentionally left blank.

Acknowledgments

My appreciation includes the review of the Application of Knowledge (AOK) questions by Dr. Terry Beattie, former diving medical services officer for the Washington, DC diving unit. His physiology expertise was also helpful.

Wall diving in Haiti

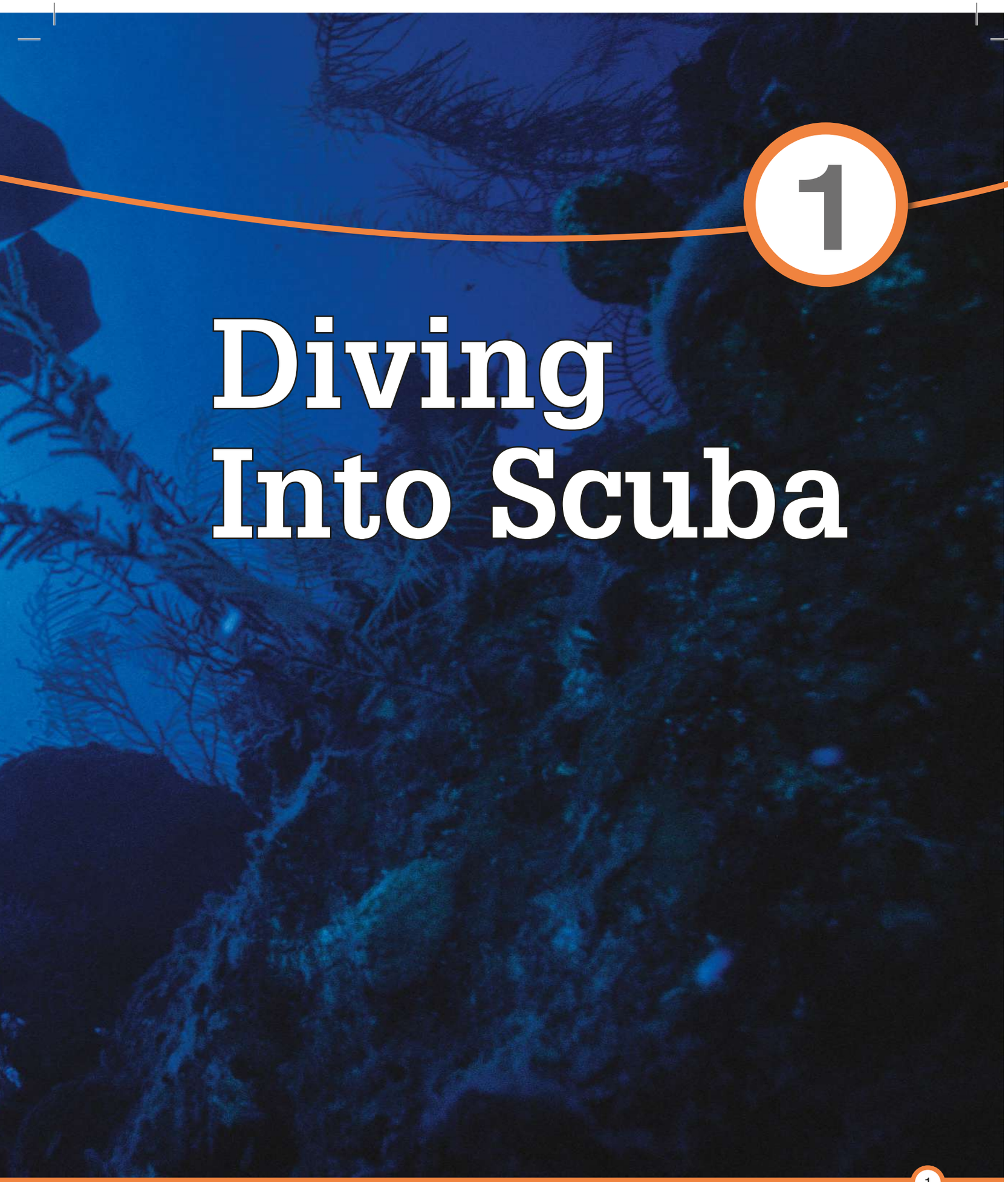

1

Diving Into Scuba

Dive In and Discover

By the end of this chapter, you will be able to do the following:

- Describe the joys of diving.
- Define the terms *closed circuit*, *open circuit*, *scuba*, *semi-closed circuit*, and *C-card*.
- Explain the training requirements for participation in scuba diving activities.
- List two medical conditions that disqualify an individual from scuba diving.
- Describe three risks associated with scuba diving.
- List several questions to ask when selecting an entry-level scuba diver training course.
- Explain several responsibilities that you assume when you become a certified scuba diver.

The Joys of Diving

As a diver, you are weightless and can move in all directions. Your freedom approaches that of a bird as you move in three dimensions in a fluid environment. Diving weightless in clear water in a forest of underwater plants with sunlight streaming down is only one of many unforgettable experiences awaiting you.

Just as there are mountains, plains, and various environments above water, there are various environments for you to experience underwater. Coral reefs, kelp forests, incredible rock formations, and other natural wonders await divers in various geographical regions. In addition, divers can explore piers, jetties, quarries, mysterious shipwrecks, and other artificial structures. The variety of underwater sights is limitless. There is more to view underwater than anyone could possibly see in an entire lifetime. A range of activities—such as photography, hunting, and collecting—make diving challenging and rewarding. Scuba diving includes an activity to interest everyone.

Divers are friendly and easy to get to know; their camaraderie is well known. Diving is a sharing activity, and there is much to share. If you enjoy traveling, you will probably love diving. Dive travel is the number one business of recreational diving. Reasonably priced dive vacations to exotic islands abound. Most divers plan one or more diving vacations each year.

The sensations of diving are fantastic but are difficult to explain. Words cannot describe the peaceful solitude of inner space. Diving contributes to good health, can help reduce work-related stress, can increase self-esteem, and can make you feel great. You need to experience the emotions and sensations for yourself. You will then begin to know the joys of diving.

A giant grouper

When you descend beneath the surface of the water, you enter an entirely new and beautiful world. You have opportunities to see incredible life-forms that only a few people ever see. Imagine swimming in a giant aquarium, and you'll get a glimpse of what you can expect to experience in the underwater world.

How Diving Evolved

Interest in the underwater world has always existed. Driven by curiosity and the need for food, people have ventured beneath the surface of Earth's seas for thousands of years. Records exist of sponge divers, oyster pearl divers, military divers, and even salvage regulations dating back to 3000 BC. In ancient times, the two methods employed for diving were breath holding and the use of an inverted, air-filled bell (which was large and heavy). John Lethbridge, an Englishman, developed an oak cylinder diving engine in 1715 to pump air into a bell. The hand-operated air compressor, invented in 1770, allowed fresh air to be pumped into a submerged diving bell. In 1772, Sieur Freminet of France invented a helmet-hose system that permitted divers to work without a bell. The greatest early advancement in underwater equipment occurred in 1837, when Augustus Siebe (a German living in England) invented the first closed-dress diving suit, which became known as the hard-hat system. Siebe's system was so effective that it dominated underwater work for 100 years, and it is still in use today.

The origin of recreational diving can be traced to 1825. That year, William H. James, an Englishman, invented the first **open-circuit** self-contained underwater breathing apparatus **(scuba)** system. Although this system was not very practical, it did solve the problems of the air hose tether and the depth restrictions related to oxygen poisoning. Benoit Rouquayrol (a French mining engineer) and Auguste Denayrouze (a French naval officer) invented the aerophore in 1865. Their creation is considered the source of modern scuba equipment.

Later, two Englishmen, Henry Fleuss and Robert Davis, developed a **closed-circuit** oxygen rebreather system in 1878. This system eliminated the air hose that had previously tethered divers and restricted their movement. The system had great military application because bubbles were not exhausted into the water. Oxygen rebreathers were used extensively by Italian and British frogmen during World War II. However, the use of oxygen rebreather systems is limited to a depth of 20 feet (about 7.6 m) because breathing pure oxygen at greater pressures causes convulsions, which can be fatal during submersion.

Yves Le Prieur of France developed a manually controlled open-circuit scuba system in the early 1920's (year disputed); however, modern scuba diving was launched with the development of a "demand" scuba system that was perfected by Frenchmen Emile Gagnan and Jacques Cousteau in 1943. With this system, a diver could demand compressed air from a steel cylinder by simply inhaling from a pressure regulator held in the mouth. Scuba was introduced in the United States in 1950 and has been popular ever since.

The military has continued to improve underwater breathing systems that have found their way into use by recreational divers. The latest scuba systems are **semi-closed-circuit** rebreathers. These systems use sophisticated electronics to control the oxygen mixture that a diver breathes, periodically releasing only a small amount of bubbles. High levels of training, frequent use, and dedicated maintenance are required for the safe use of these expensive rebreathers. Table 1.1 outlines the basic features of each type of breathing system in more detail.

The Diving Community

The recreational diving community consists of equipment manufacturers, diving retailers, diving educators, diver training organizations, dive resorts, diving

Table 1.1 Scuba System Comparisons

Open-circuit scuba	Semi-closed-circuit scuba	Closed-circuit scuba
User inhales compressed air	User inhales oxygen and inert gas	User inhales pure oxygen
Exhaled air exhausted into water	Exhaled gases enter closed system	Exhaled gases enter closed system
Air supply duration varies with depth	Carbon dioxide absorbed within system	Carbon dioxide absorbed within system
Components simple	Oxygen added as needed	Oxygen added as needed
Unit easily maintained	Air supply duration unaffected by depth	Depth limit of 25 feet
Affordable	Components sophisticated and complex	Convulsions and death possible
Basic training required for use	Unit requires high maintenance	Components simple
	Expensive	High maintenance
	Highly technical specialty training required	Not available for recreational use
		A predecessor to semi-closed-circuit scuba

supervisors, dive guides, dive clubs and associations, publishing companies, and certified divers. Commercial, scientific, and professional diving are not considered recreational pursuits; separate communities are involved in these types of diving. This book addresses only recreational skin diving and scuba diving. A scuba diver breathes compressed air underwater, while a skin diver holds his or her breath while submerged.

Few laws pertain to recreational scuba diving. The laws that exist do not govern who may dive. The diving industry is self-regulating. The diving community realizes that it is dangerous when people who have not completed a sanctioned course of instruction attempt scuba diving. Dive businesses require proof of completion of training before they will allow you to have your scuba tanks filled or allow you to participate in diving activities. Many dive operations also require proof of recent experience documented in a diving logbook. If you have not been diving for a year or more, you may be required to complete at least one dive under the supervision of a diving professional. The supervised dive requirement helps increase the safety of divers whose skills may need to be refreshed.

When you complete your training requirements as a scuba diver, you receive a certification card called a **C-card**. Most C-cards do not require renewal, but the recreational diving community universally recommends the completion of refresher training after periods of inactivity in excess of six months.

Certified divers may dive without supervision or may employ the services of a diving guide. Just because a divemaster or diving supervisor is aboard a dive boat, you should not assume that this person is a guide who will dive with you. Guide services are not necessarily included with diving trips. If you want a guide to lead you about underwater and show you the sights, you should arrange for guide services in advance.

You will learn more about the community as your diving experience increases. Many opportunities for adventure and enjoyment can be found within this community. Get actively involved in the community when you complete your training and officially qualify as a scuba diver.

Diver Training

A national diver training organization must sanction your training. The training organization establishes standards of training that you must meet before the organization will issue a certification card. The Appendix A includes a list of national diver training organizations. Your instructor should have credentials that identify her as a qualified instructor. The instructor's membership in the training organization must be current in order for the person to be qualified to teach and certify divers. Be sure to confirm your instructor's qualifications.

An entry-level training course usually consists of a series of academic sessions, pool or confined-water (pool-like conditions in open water) sessions, and open-water training (in actual diving locations). You will learn theory in the classroom, learn skills in controlled conditions, and then apply your skills in an actual diving environment. This logical progression is common for all approved diver training courses.

The minimum requirements for your training are as follows: You should have four or more academic sessions, four or more pool sessions, and at least four scuba dives in open water. A skin dive in open water may also be part of your training.

Your initial training should involve a total of 30 to 40 hours of instruction. The instruction should occur over a period of several weeks instead of a few days. The time between class sessions allows you to reflect on your training and helps you absorb and retain the knowledge and skills better than a concentrated training schedule would.

Proficiency Testing

After you have learned and practiced the skills of skin and scuba diving, you must demonstrate your competence at a level established by the agency sponsoring your training. Proficiency testing may include diving exercises that are challenging and fun. Examples include mask recovery and clearing for skin diving, simulated boat exit for scuba diving, alternate air source breathing, buddy breathing, equipment handling, and a sequence of surface entry and equipment donning known as a bailout.

Diving Prerequisites

Scuba diving can be undertaken by anyone over 12 years of age who is in normal health and has a reasonable degree of physical fitness. People younger than this should not participate in scuba diving (even when supervised by adults) because they do not have the mental and emotional maturity to deal with the problems that might arise. Skin diving is a good activity for youngsters if they are well supervised.

For scuba diving, you need to have swimming ability, but you do not need to be a competitive swimmer. At the beginning of your training, you should be able to swim 200 yards (183 m) nonstop at the surface using any combination

of strokes. There is no time requirement for the swim. Being comfortable in the water is more important than being able to swim fast. You also need to be able to swim 25 feet (7.6 m) underwater with no push-off. By the end of the course, you should be able to swim 300 yards (274 m) nonstop at the surface using any combination of strokes; you should also be able to swim 50 feet (15.2 m) underwater with no push-off. The goal is to increase your aquatic proficiency during the course.

Good health means your heart, lungs, and circulation are functional and that you do not have any serious diseases. Any medical conditions—even if controllable under normal conditions—that might incapacitate you in the water could cause you to drown while scuba diving. Some individuals with asthma or diabetes may be able to dive if they have obtained special medical approval. People with physical disabilities may also dive if they have medical approval from a physician. The air spaces in your body—sinuses, ears, and lungs—must be normal because changes in pressure affect them. Other medical conditions, such as seizure disorders, absolutely preclude a person's involvement in diving; a seizure while diving can be fatal. Women who are pregnant should not scuba dive. Increased pressure can adversely affect an unborn child. Pregnant women may choose to participate in snorkeling as an alternative to scuba diving. Many women ask whether they may dive during menstruation. Menstruation does not preclude a woman from diving if her health permits participation in other sports during that time.

You need to be emotionally fit as well as physically fit for diving. If you are terrified of water or of feeling confined, diving is probably an activity you should avoid. Normal concerns are to be expected, but stark terror is unacceptable.

You should have a physical examination before you begin your training, especially if it has been more than a year since your last exam. Ask your instructor to recommend a diving physician. Physicians who do not understand the physiology of scuba diving sometimes inappropriately grant approval to people who have medical conditions that place them at great risk in and under the water. Your instructor can likely recommend a physician who understands medical issues related to scuba.

Diving Risks

All activities present some risk. There is risk involved in walking across the street or driving a car. To avoid injury while participating in an activity, people take precautions for their safety. Precautions must be taken for scuba diving just as for any other pursuit. The level of risk in diving is similar to that of flying in an

SCUBA WISE

I nearly drowned when I was 4 years old, and I became terrified of water. When I was 8, I spent a summer with my uncle in Ohio. He would take me to Lake Erie and give me pennies if I would wade into water deep enough to cover my navel. I learned to swim as a Boy Scout at age 11. Although I completed a lifesaving class at age 16, I was still apprehensive about water. When I learned skin and scuba diving and discovered that I could actually relax in water, the water became my ally. For the first time in my life, I enjoyed water and was able to rid myself of my childhood fears. Just because someone feels anxiety about water does not mean that the person can't enjoy scuba diving. If you can swim 200 yards, you can learn to dive and to love being in and under the water.

airplane. Both are low-risk activities when done with well-maintained equipment according to established rules and in good environmental conditions. Unfortunately, both activities are unforgiving if you ignore the rules and recommendations designed to minimize the risks.

The following information (and the information throughout this book) will make you aware of injuries that scuba divers can incur. This information alerts you to potential hazards and, more important, helps you learn to avoid injury. If you do what you are taught to do as a diver, your risk will be minimal, and all of your diving experiences will likely be pleasant ones.

Pressure changes with depth. Changes in pressure can severely injure bodily air spaces if you are not in good health or if you fail to equalize the pressure in the bodily air spaces with the surrounding pressure. You will learn equalizing techniques as part of your training. Gases are normally dissolved in the fluids and tissues of your body. Increased pressure increases the amount of gas dissolved in your body. If you ascend too rapidly from a dive, the gases in your system can form bubbles and produce a serious illness known as **decompression illness**. By regulating your depth, the duration of your dive, and your rate of ascent, you can avoid decompression illness. Failure to heed depth and time schedules and ascent rates can result in serious, permanent injuries.

Diving can be strenuous at times. You need sufficient physical fitness and stamina to handle long swims, currents, and other situations that may arise. If you become winded from climbing a flight of stairs, you may need to improve your level of fitness before learning to scuba dive. Exhaustion in and under the water is hazardous. A good exercise to improve fitness for diving is swimming with fins while breathing through a snorkel.

Diving takes place in water, an alien environment. You use life-support equipment to dive, but you cannot depend entirely on the equipment for your well-being. Aquatic skills are essential in and around water. People with very weak aquatic ability can drown when minor equipment problems occur—problems that could be handled easily by a person with good water skills. To be a scuba diver, you must be comfortable in the water.

You should not be overly concerned with the potential risks of diving because the possible

Red Sea coral cave

injuries are preventable. Learning how to dive as safely as possible is the purpose of your training. You will learn to minimize the risk of injury and maximize your enjoyment of the underwater world.

Selecting a Dive Course

There are many diver training organizations and thousands of professional diving educators. Your phone book might list diving businesses that offer sanctioned courses. Many universities, community colleges, and recreational departments also offer scuba courses. (See Appendix A for a list of diver training organizations.) Ask about the qualifications, experience, and reputation of several diving instructors in your area to select the course that can provide you with the best possible training. Here are some questions you should ask:

- Is this training sanctioned by a diver training agency?
- How long has the instructor been teaching scuba diving?
- Which levels of training is the instructor qualified to teach?
- May I speak with the graduates of a recent class?
- Why is this course better than others in the area?
- Are assisting and rescue techniques taught in the course?
- How many instructor-supervised open-water dives are included?

The tuition for diving instruction is usually between $200 and $300. The lowest-priced course may not necessarily be a bargain. Find out what is included with the course fee and, more important, what the total cost will be for you to become certified as a scuba diver. You do not have to purchase all the equipment needed to scuba dive, but you need to have a mask, snorkel, fins, and usually boots and gloves for your training (see figure 1.1). Use of the additional required equipment is typically part of the course tuition.

You should find out whether the price of the course includes the costs of educational materials and certification. There may be additional costs for travel, lodging, parking, boat fees, and equipment rental for open-water training. Determine the complete cost before enrolling in a course.

When you have selected the best program for you and have enrolled in a course, you should receive a reading assignment for your first session. If you are not given an assignment, speak with the instructor; your learning will be enhanced if you read in advance about the topics to be presented in class. Good diving instructors provide a handout with reading assignments.

Figure 1.1 Required scuba diving training equipment.

Diving Responsibilities

When you qualify as a scuba diver, you assume many responsibilities. You are responsible for your safety, for the safety of those you dive with, for the image of scuba divers, and for the preservation of the diving environment. The diving community encourages divers to accept responsibility for their actions. To be part of the diving community, you need to be a responsible diver. Learn what you should do, then do what you learn.

Summary

Diving can be a source of great joy. Many exciting experiences await the trained diver. You need dive credentials to participate in diving activities. You must complete diver training to obtain your C-card and logbook. But diving is not for everyone. You must have normal health, good swimming skills, and reasonable physical fitness. Diving poses risks that a well-trained, wise, and fit diver can minimize. Compare training programs and choose the best education, which may not be the quickest or the least expensive. Remember that you accept a great deal of responsibility when you become a diver. Do not assume that you can transfer the responsibility for a dive accident to someone else. Ultimately, you control your actions underwater. Become a competent, self-reliant diver who adheres to recommended safety practices, and you will discover the joy of diving.

Application-of-Knowledge (AOK) Questions

The following questions will enhance your understanding of what you have learned in this and every chapter. Take time to consider each question before you look at the answer, which is at the back of the book in Appendix B. When you apply the basic knowledge you have learned and correctly respond to the following questions, you demonstrate understanding, which is a higher level of learning than mere knowledge. Do not be concerned if you do not have a correct response. Your analysis of the question and the revelation of the answer will increase your understanding.

1. What is your primary reason for learning to scuba dive? What do you think you will do while scuba diving after completion of your initial training?
2. Why do beginning recreational divers use open-circuit breathing systems instead of closed-circuit systems?
3. Why should you document your dives in a logbook?
4. What are the advantages of completing a 40-hour scuba diving course rather than a course that is much shorter in duration?
5. Why is a medical examination recommended for those who want to learn scuba diving? What type of doctor is the best choice for a scuba diving medical exam?
6. List three actions you can take to minimize your risk of injury when scuba diving.
7. What are some of the factors that you should consider when selecting a scuba diving course?
8. List three actions you can take to demonstrate that you are a responsible diver.

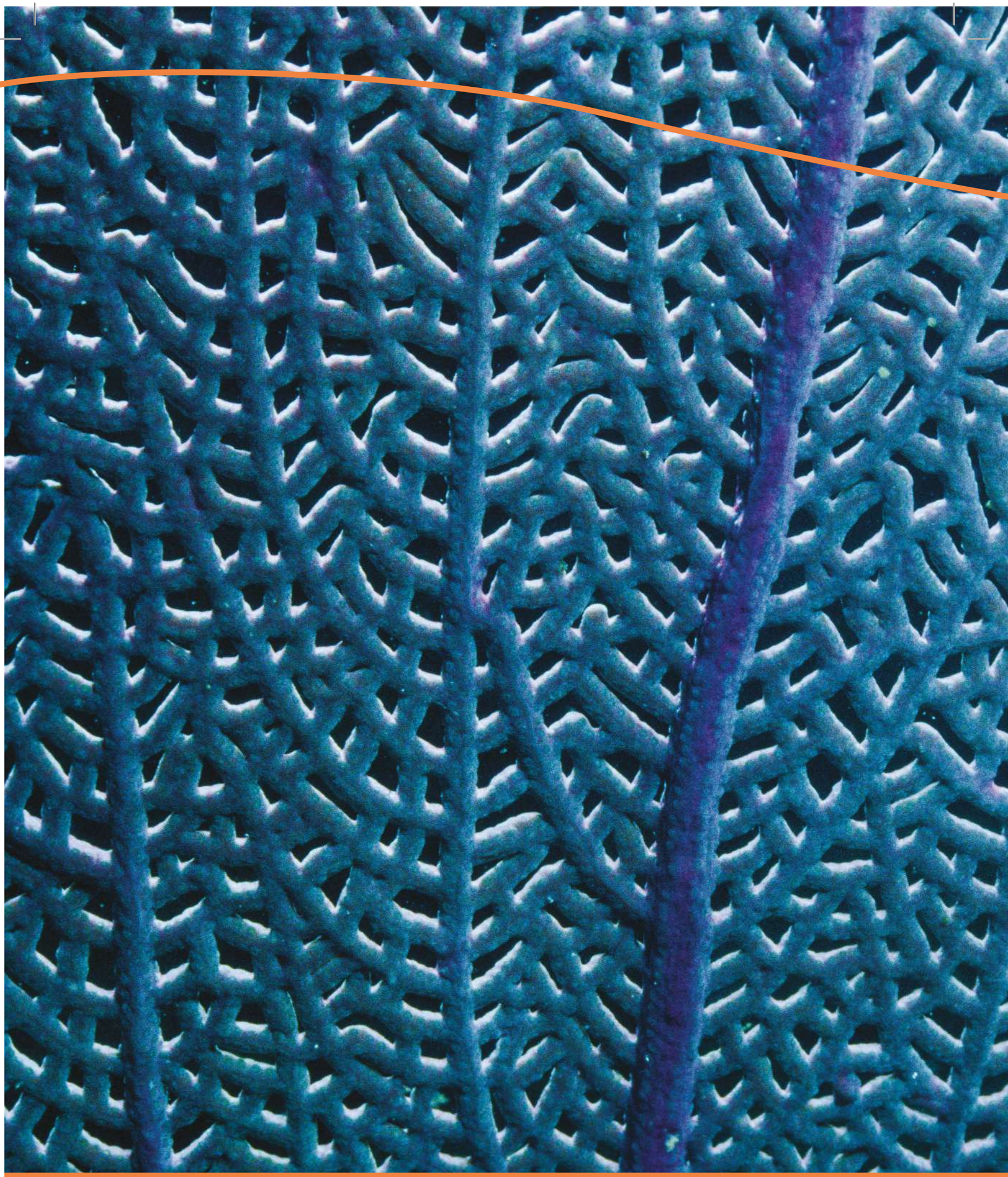

A sea fan

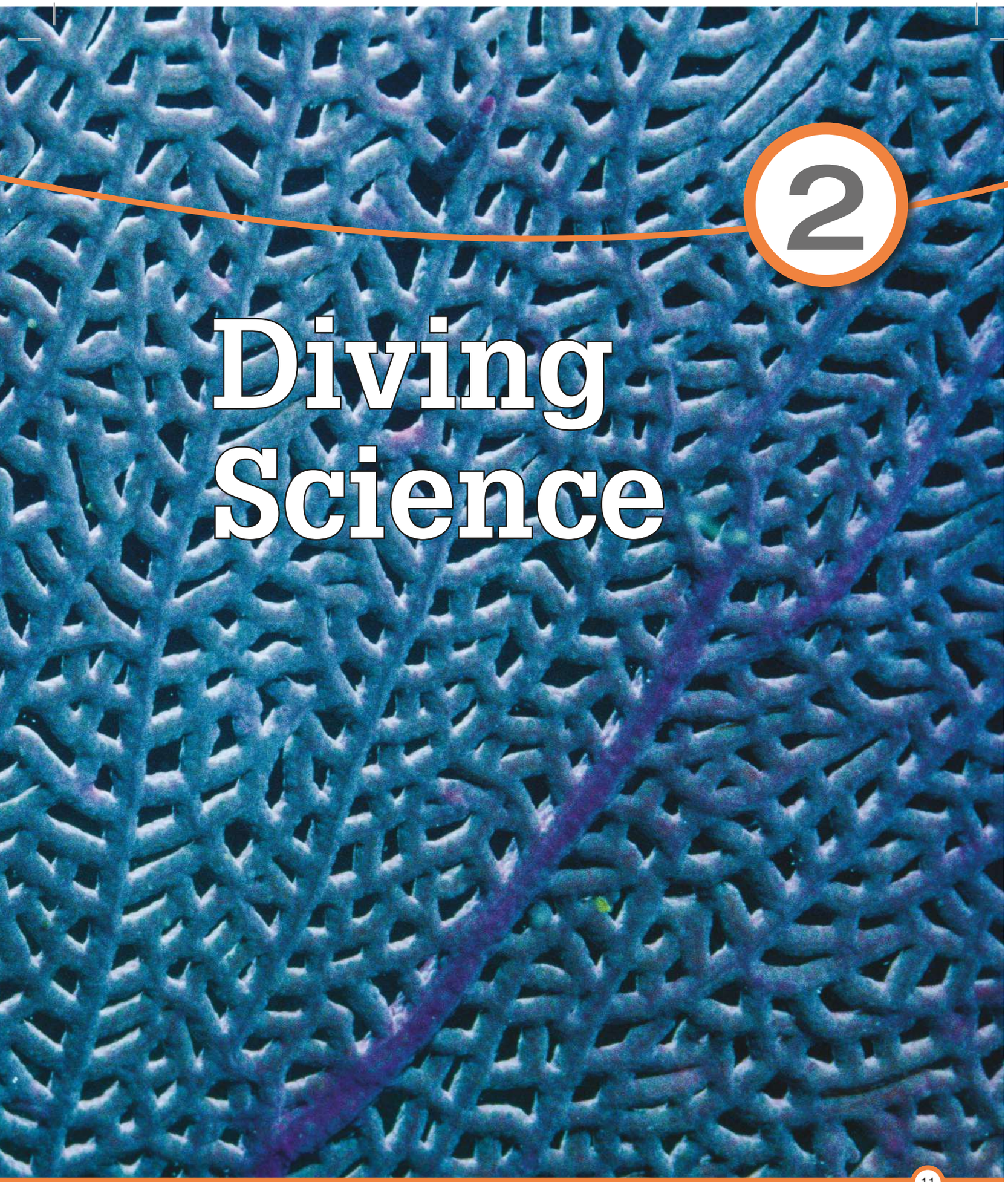

2 Diving Science

Dive In and Discover

By the end of this chapter, you will be able to do the following:

- List three body air spaces of concern to divers.
- Describe the process of hearing in air.
- Describe the cardiorespiratory process.
- Explain the effect of carbon dioxide on breathing.
- Define the terms *eardrum*, *middle ear*, *eustachian tube*, *nitrogen narcosis*, *hyperventilation*, *hypoventilation*, *density*, *buoyancy*, *pressure*, *gauge pressure*, *absolute pressure*, *Boyle's law*, *Gay-Lussac's law*, *Dalton's law*, *squeeze*, *reverse block*, *ingassing*, *outgassing*, and *partial pressure*.
- Explain the principle of buoyancy and the key to controlling buoyancy.
- Explain the effects of pressure and temperature on a volume of air in a flexible container.
- Explain why it is important to vent your lungs when you ascend in water.
- Convert any temperature to absolute temperature.
- Explain the process of ingassing and outgassing.
- State the two primary factors that affect the air consumption of a scuba diver.
- List four methods of heat loss.
- Describe three potential problems for scuba divers that may be caused by humidity.
- Describe the effects of water on vision and hearing.

Anatomy for the Diver

Inside your body are air-filled spaces that are affected by changes in pressure. The three body air spaces of primary concern to you as a diver are the lungs, the ears, and the sinuses. Understanding the structure and function of your throat will also help you as a diver. Figure 2.1 illustrates the structure and functions of the sinuses, throat, and lungs.

Sinuses

The sinuses warm and humidify inspired air. They secrete mucus to help protect the body by trapping airborne germs. The small airways that connect the sinuses to the nasal passages are normally open. Congested sinuses pose problems for divers. In the next chapter, you will learn more about sinus problems and how to prevent them.

Throat

In addition to being the organ of voice, the throat and larynx help prevent foreign matter from entering the lungs. If something foreign, such as food or water, comes into contact with the larynx, a reflex action causes a spasm of the larynx. Coughing expels the foreign substance. You have experienced this sensation when something has "gone down the wrong pipe." Review the throat section of figure 2.1. During your scuba diving training, you will learn how to keep water out of your larynx to avoid coughing and choking in and under the water.

Lungs

Healthy lungs are essential for scuba diving. The lungs are large organs that contain millions of microscopic air sacs. Your lungs have a maximum capacity and a minimum capacity. When you exhale completely, your lungs are not empty. They contain about 2 pints (1L) of air. The air remaining in your lungs after you have exhaled completely is your residual volume. The amount of air you move in and out of your lungs is your tidal volume. When you are at rest, your tidal volume is small. When you exert yourself, your tidal volume increases until you reach both your maximum lung volume and your residual volume with each breath. Your vital capacity is the difference between the volume of air for a maximum inhalation and the volume of air for a maximum exhalation—typically about 6 to 8 pints (2.8 to 3.8L). In the next chapter, you will learn several reasons why your lungs are the most critical air spaces when diving.

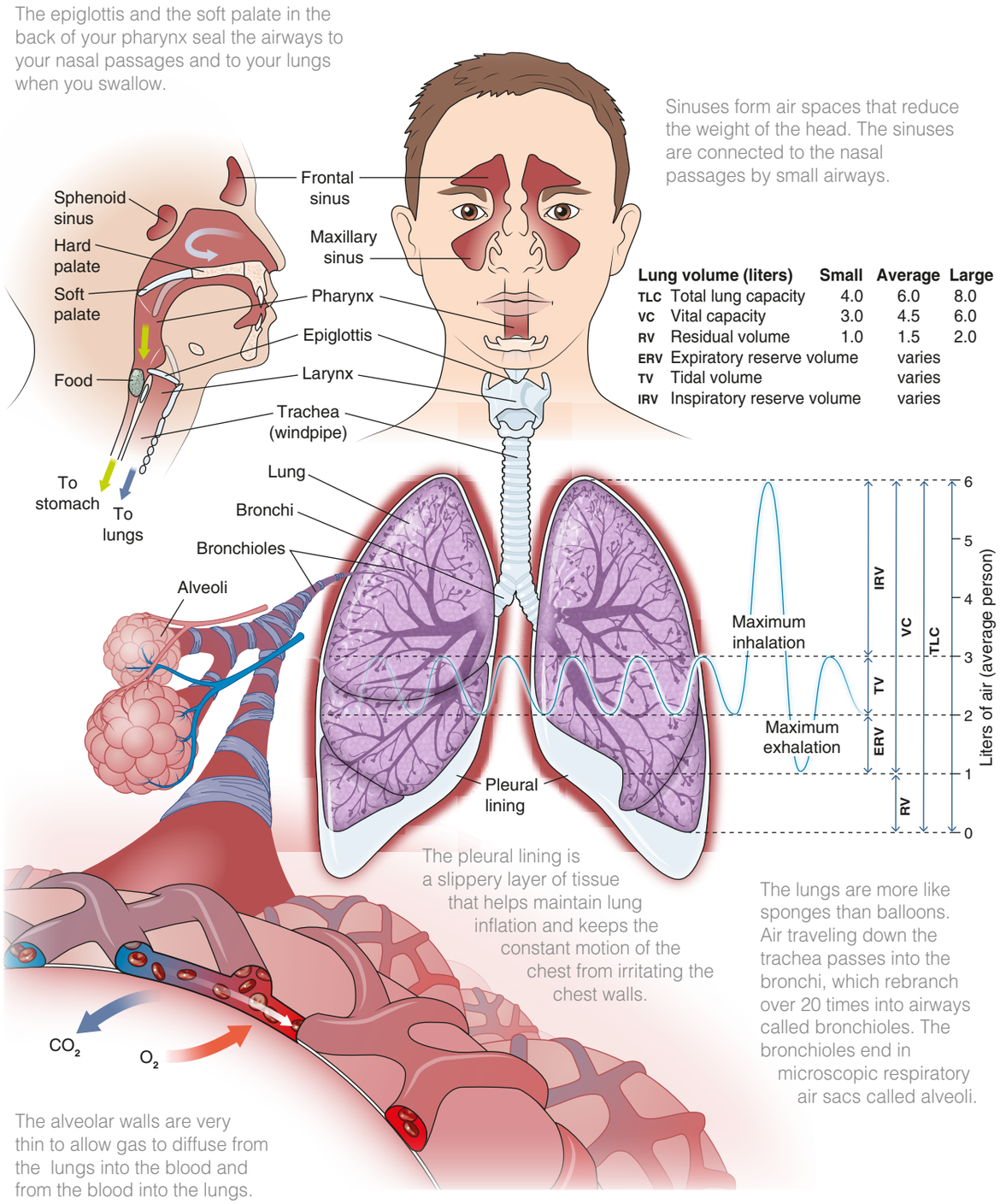

Lung volume (liters)		Small	Average	Large
TLC	Total lung capacity	4.0	6.0	8.0
VC	Vital capacity	3.0	4.5	6.0
RV	Residual volume	1.0	1.5	2.0
ERV	Expiratory reserve volume		varies	
TV	Tidal volume		varies	
IRV	Inspiratory reserve volume		varies	

Figure 2.1 Functions of the sinuses, throat, and lungs.

Colorful fish in a Florida coral cave

Your body is a marvelous machine. It performs many complex functions automatically. Your body is well adapted to an air environment, but it can also adjust in many ways to the aquatic environment. In this chapter, you will become familiar with some of the structures and functions of your body that are important for scuba diving. You will learn the differences between the air and water environments and how changes in pressure affect your body. As a diver, you face many challenges in the underwater environment, but these challenges can be managed.

Ears

Behind your **eardrum** is an air space called the **middle ear**, which is illustrated in figure 2.2. The pressure in the middle ear must equal the pressure in the outer ear; otherwise, the eardrum cannot move freely. The next chapter explains how to keep pressure inside your ears equal to external pressure. Your **eustachian tube** allows the equalization of pressure in the middle ear. The liquid-filled cochlea contains hairlike projections called cilia, which convert mechanical movement to electrical signals for the brain. The movement of the oval window by the tiny bones of the middle ear causes the liquid and the cilia in the cochlea to move back and forth. The oval window movement could not take place without a second window in the hearing organ—the round window. When the oval window moves inward, the round window moves outward, and vice versa.

If the motion sensed by your semicircular canals and the visual cues received by your eyes are not in harmony, motion sickness can result. Sudden changes in temperature or pressure in the middle ear can affect your semicircular canals

Abbreviations

✓ **ATA**	atmospheres absolute	✓ **FSW**	feet of seawater
✓ **atm**	atmospheres	✓ **ft**	feet
✓ **CO**	carbon monoxide	✓ **m**	meters
✓ **CO_2**	carbon dioxide	✓ **O_2**	oxygen
✓ **°C**	degrees Celsius	✓ **psia**	pounds per square inch absolute
✓ **°F**	degrees Fahrenheit	✓ **psig**	pounds per square inch gauge
✓ **FFW**	feet of freshwater		

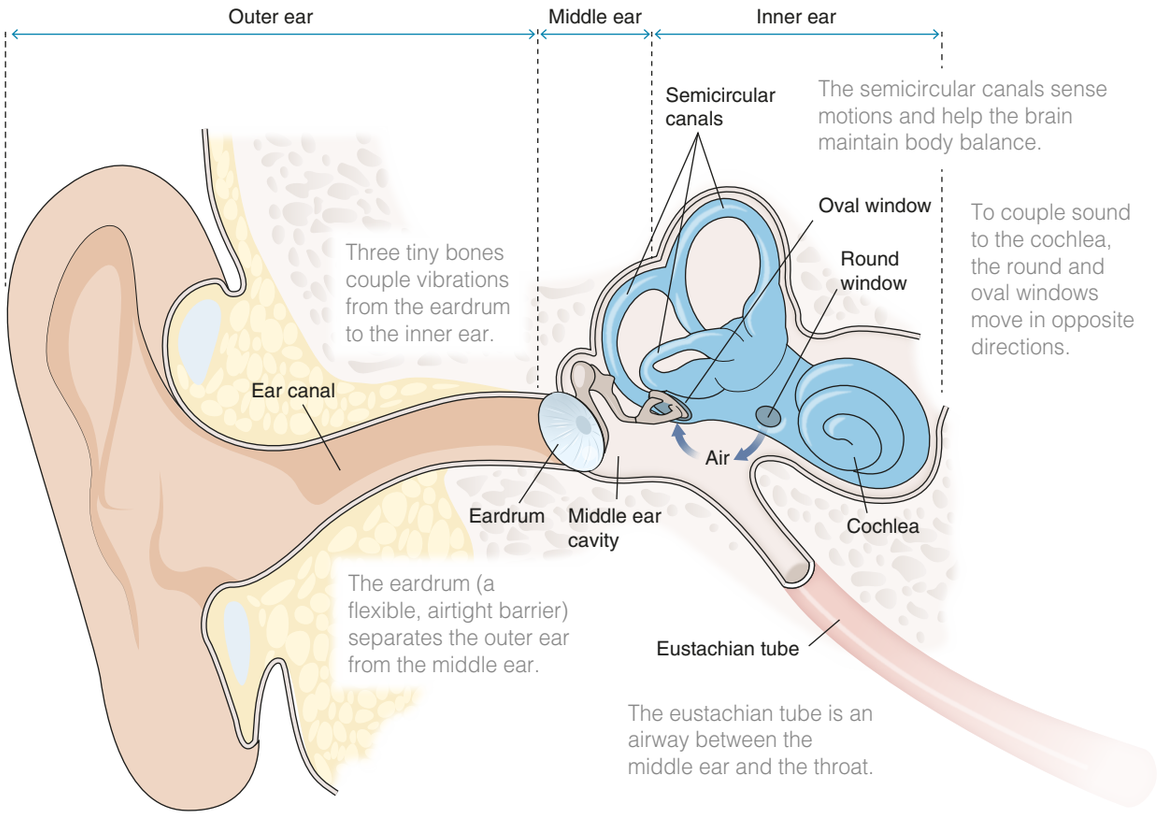

Figure 2.2 Process of hearing in air.

and cause temporary disorientation. (The next chapter presents potential ear problems for divers, how to avoid them, and how to handle them if they happen.)

Teeth

You may be surprised to learn that there are dental concerns for divers. Pressure can affect air pockets in improperly filled teeth and can cause tooth pain. If a tooth hurts only under pressure or only after a dive, see your dentist and tell him what you suspect. The roots of some upper molars extend into the sinus cavities. You should postpone diving for several weeks after you have had a tooth extracted.

Your mouth and jaws are designed for an even bite. If you bite hard on a mouthpiece with only your front teeth for prolonged periods, your jaws will become sore. Special mouthpieces designed for a proper bite can help reduce the problem. You should not have to bite hard on a mouthpiece to hold it in place. If you find biting necessary, get lighter equipment. Prolonged, improper biting that irritates your jaws can lead to serious inflammation of your jaws and ears.

Respiration and Circulation

One of the fascinating processes within the human body is your ability to breathe in air and circulate oxygen to the tissues with no conscious effort. As your level

of exertion increases, your heart and lungs automatically adjust to meet the increased demands for oxygen and nourishment. An understanding of the gases involved in respiration and the basics of respiration and circulation can help you understand the effects and the demands of diving on your lungs and heart.

Gases We Breathe

Several gases affect recreational divers. You need to know about their effects on your body. About 80 percent of air is nitrogen (N^2). At sea-level pressures, nitrogen has no effect on your body. At a depth of about 100 feet (30 m), the increased pressure of the gas has a detrimental effect, which is called **nitrogen narcosis**. Excessive nitrogen in your body at the end of a dive can produce a serious illness known as decompression sickness. You will learn more about nitrogen narcosis and decompression sickness in the next chapter.

Oxygen (O_2) is the gas that supports human life. Any other gas mixed with oxygen serves only as a vehicle for oxygen to be inspired. Approximately 21 percent of air is oxygen (see figure 2.3). You need to breathe at least 10 percent oxygen to remain conscious. However, oxygen breathed under high pressure is poisonous and causes convulsions because oxygen at increased pressure affects your nervous system. You usually have compressed air—not pure oxygen—in your scuba tanks. A specialty form of diving uses a nitrogen and oxygen mixture with a higher percentage of oxygen than is found in air. The mixture, which reduces the effects of nitrogen at depth, is called **nitrox**. The use of special mixed gases, including nitrox, requires special training, equipment, and procedures.

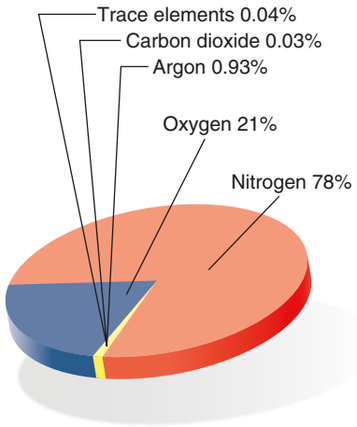

Figure 2.3 Composition of air.

As your tissues use oxygen, they produce carbon dioxide (CO_2). Carbon dioxide is the primary stimulus for respiration. The greater the level of carbon dioxide in your body, the greater your urge to breathe will become. If the level of carbon dioxide in your body becomes too great, unconsciousness will result.

Carbon monoxide (CO) is a poisonous gas produced by the incomplete combustion of gas or oil. The exhaust from an internal combustion engine contains carbon monoxide. An oil-lubricated air compressor that overheats can produce carbon monoxide. Even a minute amount of carbon monoxide in your scuba tank can poison you and lead to unconsciousness or death. Air filling stations must take care to avoid contamination of air with carbon monoxide.

Breathing and Circulation Mechanics

When you need to breathe, sensors at the base of your brain send a signal that stimulates your diaphragm to contract and your chest to expand. This draws air into your lungs in the same way that an old-fashioned bellows draws in air when you expand it. Your diaphragm contracts and increases the volume of the chest cavity, while the muscles of your chest expand your chest cavity to inspire air. Figure 2.4 illustrates how the heart, lungs, and circulatory system work in the process of respiration.

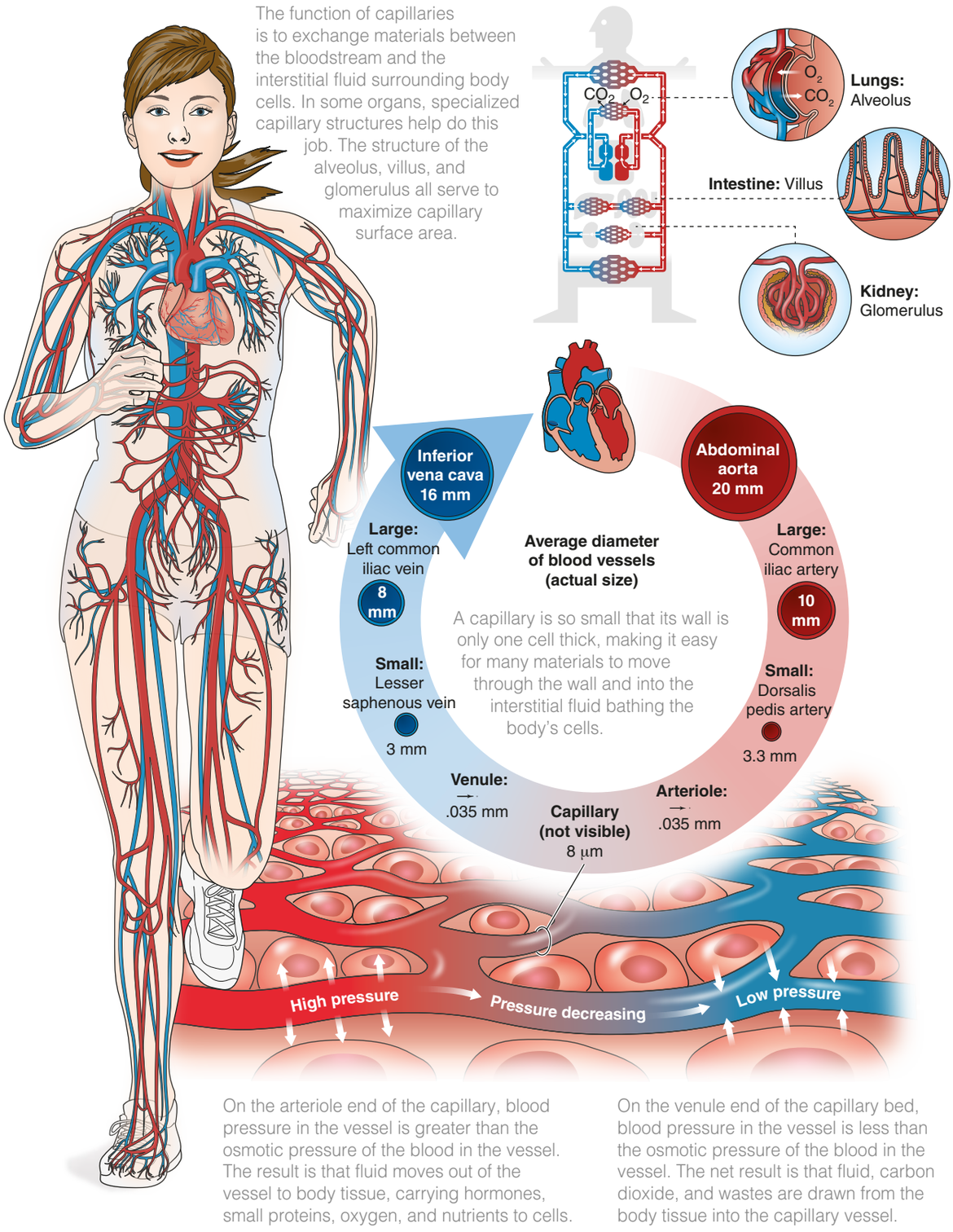

Figure 2.4 The cardiorespiratory system.

Blood consists of plasma (which is a colorless liquid) and a variety of cells. Hemoglobin, a blood component, is the primary oxygen-carrying mechanism in the blood. Approximately 45 percent of the blood is hemoglobin. Hemoglobin releases oxygen when it reaches tissues that need oxygen.

When the tissues use oxygen, they produce carbon dioxide. The carbon dioxide diffuses into the venous system and into the lungs in exchange for oxygen. This completes a circulatory cycle, which takes about 30 seconds.

Carotid sinuses on each side of the neck sense blood pressure within the circulatory system. Excessive pressure on the carotid sinuses during exercise causes the heart to slow when it should be working hard to meet the oxygen demands of the body. Decreased output from the heart can lead to insufficient oxygen for the brain, which can cause unconsciousness. A blackout caused by pressure on the carotid sinuses (see figure 2.5) is a carotid sinus reflex. Therefore, beware of diving equipment that is tight around your neck.

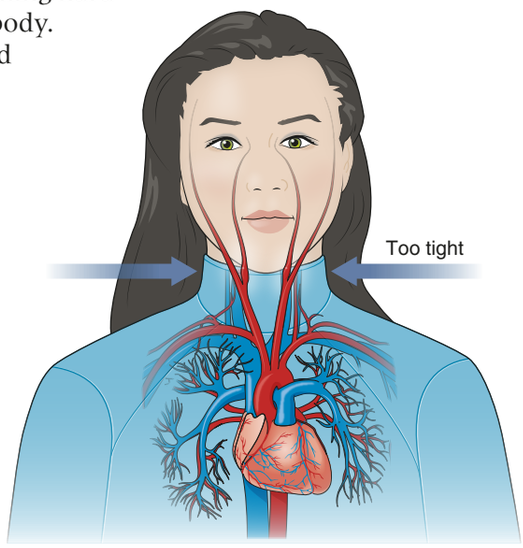

Figure 2.5 Excessive pressure on the carotid sinuses can lead to unconsciousness.

Exhalation is usually a passive process. To exhale carbon-dioxide-laden air from the lungs, the diaphragm relaxes, and the elasticity of the chest cavity forces air from the lungs. You ventilate your lungs approximately 12 to 20 times per minute when at rest. Respiration functions automatically. The key to respiration is the level of carbon dioxide in your circulatory system. When the carbon dioxide in your body reaches a certain level, your brain stimulates respiration. When you voluntarily hold your breath, the buildup of carbon dioxide within your body urges you to breathe. Many people believe that the amount of oxygen in the body controls respiration, but it is primarily the level of carbon dioxide that regulates breathing.

Hyperventilation is rapid, deep breathing in excess of the body's needs. Limited hyperventilation—three or four breaths—enhances breath holding (see figure 2.6*a*). But if you hold your breath after excessive hyperventilation, you may lose consciousness without warning before being stimulated to breathe (see figure 2.6*b*). A breath-holding diver who loses consciousness from lack of oxygen usually blacks out near the surface during ascent. The sudden loss of consciousness near the surface is called **shallow-water blackout**. Loss of consciousness while in the water can cause drowning. You should avoid excessive hyperventilation.

If you breathe rapidly and shallowly, carbon dioxide continues to build in your system, but you do not expel it from your lungs. Inadequate breathing is **hypoventilation**. Shallow breathing is dangerous, especially when you exert yourself, because you can lose consciousness from lack of oxygen. You need to breathe sufficiently to exchange the air in your lungs.

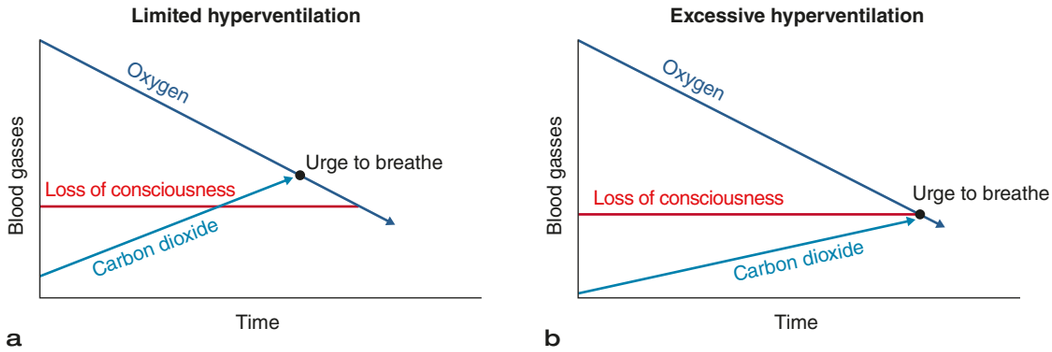

Figure 2.6 Charts showing (*a*) limited hyperventilation and (*b*) excessive hyperventilation.

Contrasts in Air and Water Environments

We live immersed in air, which is a fluid. Air has weight and takes up space. We don't pay much attention to our immersion in air because we are adapted to this environment. We have lived in it all of our lives, and we cannot see the air. The weight of the atmosphere does affect us, however.

Air weighs about 0.08 pound per cubic foot (1.28 mg per cubic cm) at sea level. As altitude increases, air becomes thinner, so its weight per volume is less in the mountains than it is at the seashore (see figure 2.7). The change in the

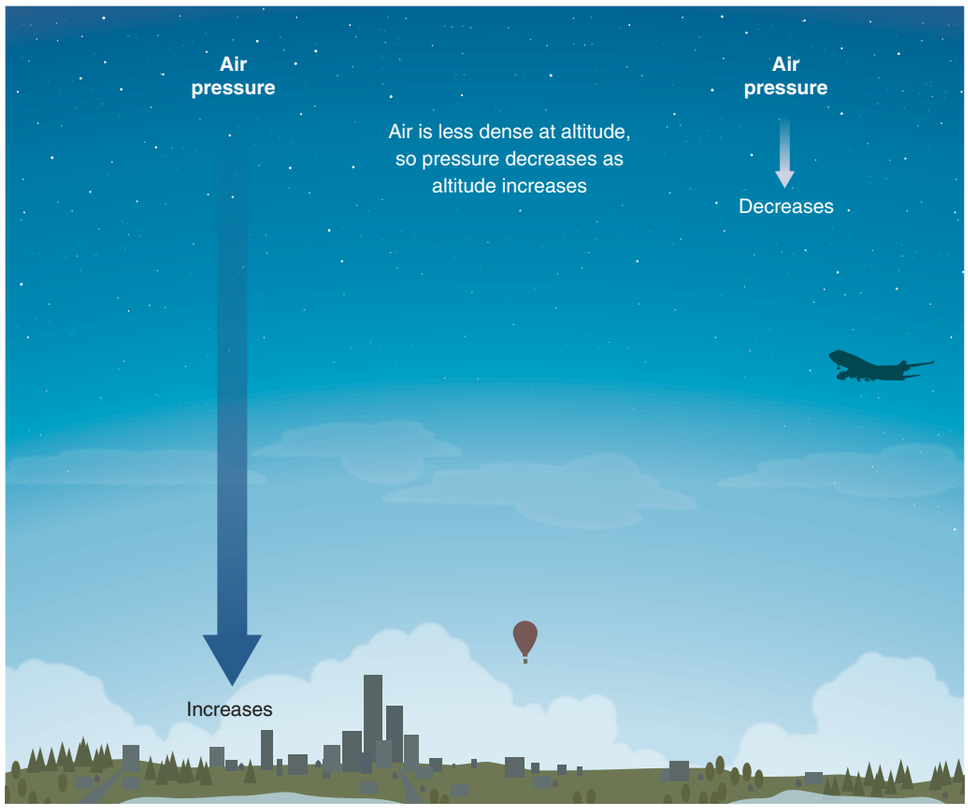

Figure 2.7 Air density and pressure are affected by altitude.

weight of air affects the air spaces in our ears when we fly or when we drive in the mountains.

Density

Density is weight per unit volume (for example, pounds per cubic foot). Water is a fluid, but it is much heavier than air. Seawater weighs about 64 pounds per cubic foot (1.025 g per cubic cm), which makes it about 800 times denser than air. Freshwater, because it does not contain salt, weighs a little less than seawater: 62.4 pounds per cubic foot (1 g per cubic cm). Temperature affects the density of water, air, and all fluids; cold water is slightly denser than warm water.

Air can be compressed, but water is essentially incompressible. Air becomes thinner as altitude increases, whereas water density remains constant throughout the water column.

Drag

Drag is a force that slows movement. Resistance to movement is much greater in water than in air. Factors affecting drag include the viscosity of the fluid, the speed of motion, and the size and shape of the object moving through the fluid. Drag is greater when the fluid is denser, the motion is faster, the object is larger, or the surface of the object is more irregular (see figure 3.9 in chapter 3).

Because of water's higher density, the molecules of water are much closer together than those of air. The tightly packed molecules affect the transmission of light, sound, and heat (see table 2.1). Light travels about 27 percent slower in water than in air. Sound travels about 4 times faster in water than in air. Still water conducts heat nearly 25 times faster than air does and moving water can

Table 2.1 Water Compared With Air

Property	Air	Water	Comparison	Effects
Density	0.08 lb/ft^3 (1.3 kg/m^3)	62.4 to 64 lb/ft^3 (1,000 kg/m^3)	Water is 800 times denser than air.	Resistance to movement
Compressibility	Yes	No	Air density varies; water density is constant (at dive pressures).	Affects body & attached air spaces
Speed of light	186,000 mi/s (300,000 km/s)	140,000 mi/s (225,400 km/s)	Light travels 27% more slowly in water.	Affects vision
Light absorption	Low	High	Water absorbs color quickly.	Light & color loss
Speed of sound	1,125 ft/s (340 m/s)	4,900 ft/s (1,400 m/s)	Sound travels 4 times faster in water.	Unable to determine source
Conductivity	0.17	3.86 to 4.12	Heat loss is 22 to 24 times faster in water than in air.	Rapid loss of body heat
Heat capacity	0.24	0.94 to 1.0	The heat capacity of water is 4 times greater than air.	Absorbs heat quickly

conduct heat hundreds of times faster than air. Water has an enormous capacity for absorbing heat with little change in its temperature. The higher density of water affects you in many ways when you dive. The next chapter explains how to deal with the effects of water density.

Buoyancy

An object's ability to float in a liquid depends on the density of the object compared with the density of the fluid in which the object is immersed. Water exerts pressure equally in all directions, even upward. You can feel the upward force (**buoyancy**) of water when you try to push something under the water. Buoyancy results from the difference in pressures on the upper and lower surfaces of an object. The weight of an object plus the weight of the fluid (air, water, or both) above the object exert a downward force. Fluid pressure pushes upward from below. The difference between these two forces is the buoyancy of the object.

Archimedes, an ancient Greek scientist, discovered that the force of buoyancy acting on a submerged object equals the weight of the water displaced (this is known as **Archimedes' principle**). A hot-air balloon floats in air because the hot air inside the balloon weighs less than the volume of cooler air the balloon occupies. A diver is buoyed upward with a force equal to the weight of the water that the diver displaces (see figure 2.8). If you and your equipment weigh less than the weight of the water you displace, you will float, or have positive buoyancy. If you and your equipment weigh more than the water being displaced, you will sink. An object that sinks has negative buoyancy. If you and your equipment weigh exactly the same as the water displaced, you neither float nor sink. Instead, you remain at the depth where you are; you have **neutral buoyancy**.

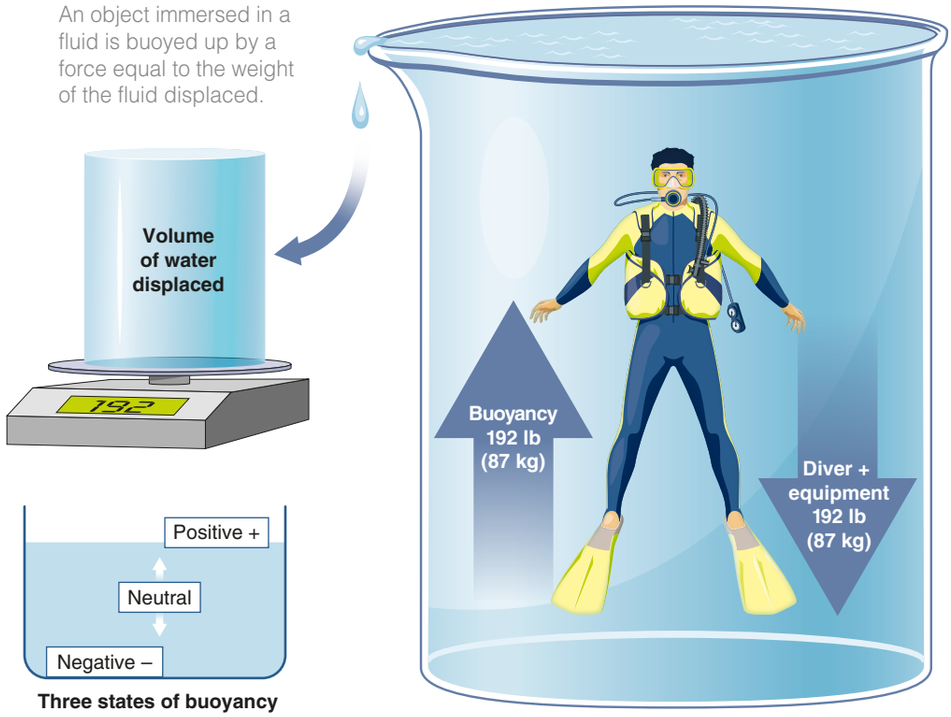

Figure 2.8 Principles of buoyancy.

As a diver, you can float at the surface, sink to the bottom, or hang suspended between the bottom and the surface. If the volume of an object increases with very little change in its weight, buoyancy increases. This happens when you add air to an inflatable jacket or vest. The next chapter addresses the factors affecting buoyancy and the principles of buoyancy control.

Pressure Measurement

Force (often weight) per unit area—such as pounds per square inch or grams per square centimeter—is **pressure**. The envelope of air surrounding the earth is the atmosphere. The weight of one square inch of the atmosphere at sea level is 14.7 pounds (1.03 kg per square cm), or 1 atmosphere (atm) of pressure. As you descend in water, the weight of the fluid—the pressure—exerted on each square inch of your body increases. One square inch of saltwater that is 33 feet (10.1 m) in height weighs 14.7 pounds, 1 atm, or 1.01 bar. One square inch of freshwater 34 feet (10.36 m) in height also exerts a pressure equivalent to 1 atm. Because 1 bar is almost equal to 1 atm, we'll consider them identical in this book. Because water does not compress (at the pressures involved with recreational diving), it follows that water pressure increases by 1 atm for every 33 feet of saltwater (33 FSW) and for every 34 feet of freshwater (34 FFW). Figure 2.9 shows how **atmospheric pressure** and water pressure are measured.

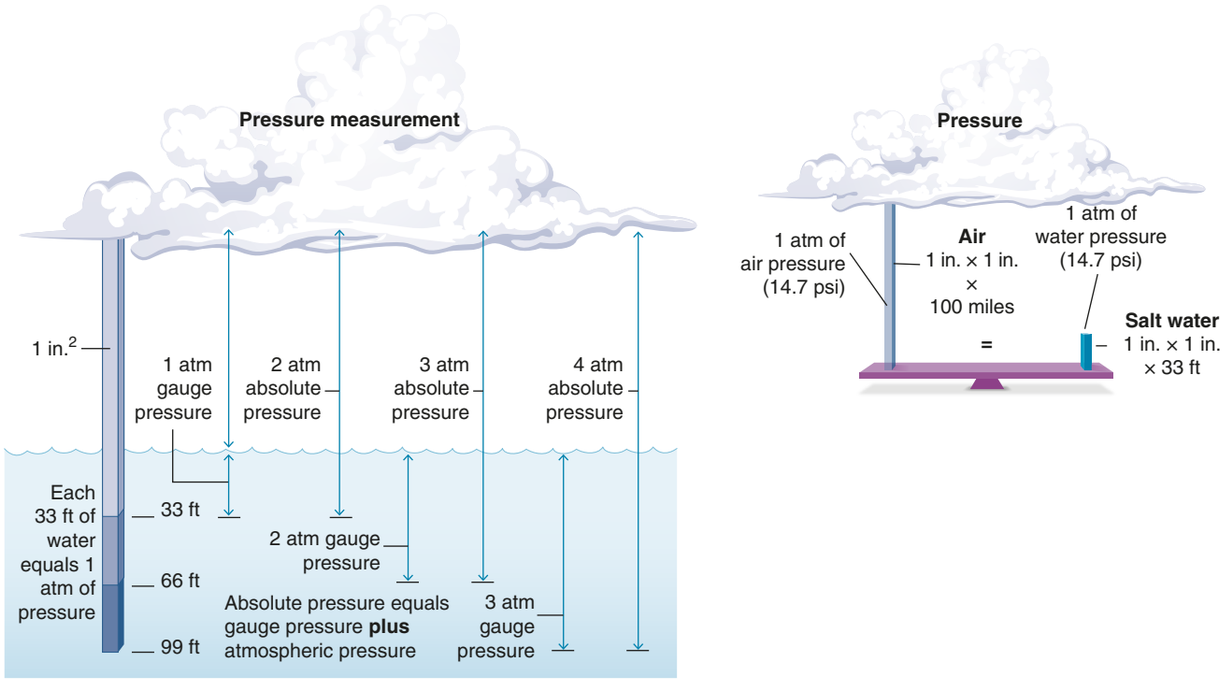

Depth		Gauge pressure			Absolute pressure		
FSW	FFW	In atm		In psig	In ATA		In psia
0	0	0	=	0.0	1 ATA	=	14.7
33	34	1	=	14.7	2 ATA	=	29.4
66	68	2	=	29.4	3 ATA	=	44.1
99	102	3	=	44.1	4 ATA	=	58.8
132	136	4	=	58.8	5 ATA	=	73.5

Gauge pressure compared to absolute pressure in fresh and salt water.

Figure 2.9 Principles of pressure measurement.

The reference for pressure is either the atmospheric pressure at sea level or zero pressure (outer space). A pressure gauge that reads zero at sea level displays only the pressure in excess of one atmosphere. Tire gauges and depth gauges are good examples of instruments that indicate **gauge pressure**, which may be measured in pounds per square inch gauge (psig), feet or meters of seawater, or bars of pressure.

The total pressure exerted is what matters to divers. The pressures of both the atmosphere and the water apply to diving. The reference for the total pressure is zero, as in a vacuum. The total pressure is called **absolute pressure**, which is measured in pounds per square inch absolute (psia). When people express absolute pressure increments in atmospheres, they use atmospheres absolute (ATA).

You obtain absolute pressure by adding atmospheric pressure to gauge pressure. Be sure you understand the concept because we'll use absolute pressure when dealing with the effects of pressure in this and later chapters.

Gas Laws

When you compress a quantity of gas, you reduce its volume and increase its density and temperature. Allowing a quantity of gas to expand increases its volume, decreases its density, and decreases its temperature. Because divers use gases, you need to understand the relationships between these properties (pressure, volume, density, and temperature) of gases.

Boyle's Law

Boyle's law states that for any gas at a constant temperature, the volume varies inversely with the absolute pressure, while the density varies directly with the absolute pressure. If you compress a closed, flexible air space (such as a balloon), you reduce its volume in proportion to the increase in pressure. When you double the pressure, a closed, flexible air space occupies only half the volume that it did originally. No air is lost. The molecules compress into a smaller area. The density of the air is twice as great as it was originally. When you return the compressed air space to its original pressure, the air inside expands until the object reaches its original volume. You compress your lungs during a breath-hold descent, and they return to normal volume when you return to the surface, provided that you do not expel air underwater.

Scuba equipment provides air to you at the exact pressure of the surrounding water. This allows you to expand your lungs to their normal volume regardless of the depth. The density of the air inside the lungs increases in proportion to the water pressure. If the water pressure doubles, the density of the air in your lungs also doubles. Figure 2.10 shows the relationships between pressure, volume, and density for a gas-filled, flexible-walled container.

Boyle's law also applies when you reduce the surrounding pressure. As outside pressure decreases, compressed air in a closed, gas-filled, flexible container expands in proportion to the reduction in pressure; for instance, if the pressure halves, the volume doubles. If a container filled with compressed air at depth is vented correctly during ascent, expanding air escapes through the vent, and the container remains full throughout the ascent. If the container is not vented, pressure inside increases when the container reaches its maximum volume. If the container is weak, the increase in pressure will rupture the container. This concept is important to scuba divers, who have many air spaces filled with

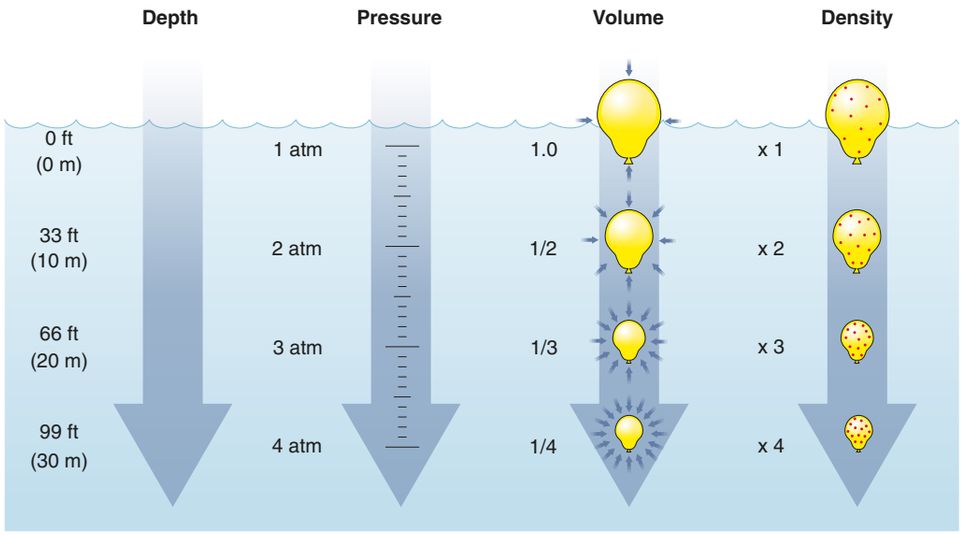

Figure 2.10 Pressure, volume, and density relationships.

compressed air. Vented air spaces do not pose a hazard. But if your lungs are not vented during ascent, life-threatening injuries will result. If you do not vent air from a flotation jacket during ascent, control of buoyancy will be lost as air expands and the jacket volume increases.

Figure 2.10 shows an interesting point about the rate of change of pressure (and volume) in water. The pressure doubles from 1 atm to 2 atm in 33 feet (10 m) of seawater. Doubling the pressure again requires a depth of 99 FSW (30 m). Note that you must ascend from 99 feet to a depth of 33 feet—a distance of 66 feet (20 m)—to experience the same rate of change of pressure that you experience when you ascend from 33 feet to the surface. In other words, the closer you get to the surface, the greater the rate of change of pressure (and of the volume of an air space). You must be more attentive to compressed air in air spaces the nearer you are to the surface.

The change in pressure that you experience during descents and ascents in water is perhaps the most significant challenge of diving. Changes in pressure have direct, mechanical effects on your body. Pressure imbalance in your body air spaces can cause discomfort. In the air environment, you feel pressure changes as a result of changes in altitude, but pressure changes in water occur at a much greater rate than in air. You can sustain serious injury unless you keep the pressure in air spaces inside and attached to you equalized with the surrounding water pressure. Boyle's law causes **squeezes** and **reverse blocks**, which figure 2.11 illustrates. Knowledgeable and experienced divers routinely equalize pressures to avoid squeezes and blocks. Keeping pressure in air spaces equalized with the surrounding pressure is one of the main subjects of the next chapter.

Gay-Lussac's Law

Boyle's law addresses gas at a constant temperature because the temperature affects the pressure and volume of a gas. Jacques Charles, a French chemist, discovered that the volume of gas at a constant pressure changes with temperature. Joseph-Louis Gay-Lussac, a French scientist, defined the effect of

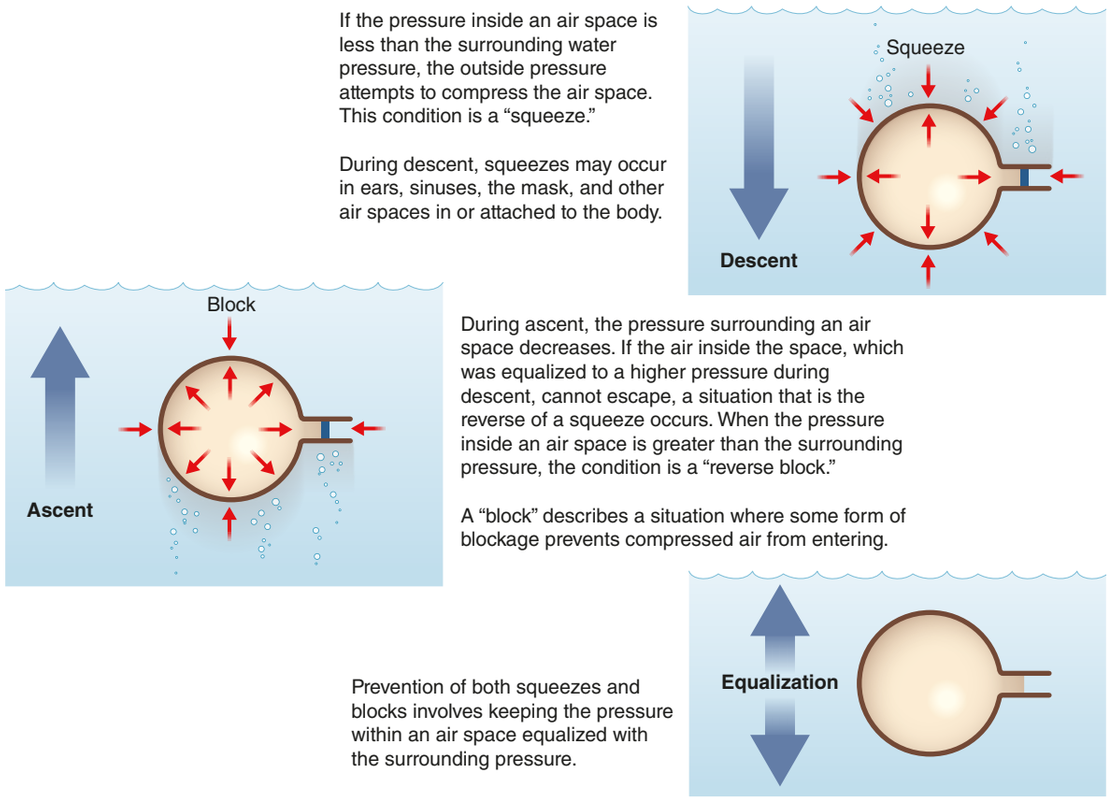

Figure 2.11 An explanation of squeezes and reverse blocks when equalizing pressure.

temperature: For any gas at a constant volume, the pressure of the gas varies directly with the absolute temperature. Just as absolute pressure must be used for pressure calculations, absolute temperature must be used for temperature computations. The absolute temperature scale for Fahrenheit temperatures is Rankine. To convert a Fahrenheit temperature to Rankine, add 460 degrees. The absolute temperature scale for a Celsius temperature is Kelvin. To convert a Celsius temperature to Kelvin, add 273 degrees.

You can observe the effect of **Gay-Lussac's law** with a scuba tank, which has a constant volume. Decreasing the temperature of the air in a tank causes the pressure to decrease. Increasing the temperature of the air in a tank causes the pressure to increase. A scuba tank taken from the trunk of a hot car and cooled in water experiences a drop in pressure, although no air leaves the cylinder. Pressure increases or decreases in a standard 80-cubic-foot (2,265 L) aluminum scuba cylinder at a rate of about 6 psig per degree Fahrenheit temperature change. Pressure increases or decreases in a standard 71.2-cubic-foot (2,016 L) steel scuba cylinder at a rate of about 5 psig per degree Fahrenheit temperature change.

Dalton's and Henry's Laws

A gas can diffuse into or out of a substance. When a gas comes into contact with a liquid, the gas dissolves into the liquid. The amount of gas that diffuses into the liquid depends on the density and temperature of the liquid, the pressure of the gas in contact with the liquid, and the length of time that the gas is

in contact with the liquid. Another term for the process of gas diffusion into a liquid is **ingassing**. Because the human body is primarily liquid, the gases you breathe diffuse into your body tissues.

In a mixture of gases, such as air, the percentage of the total pressure exerted by each gas is the **partial pressure** of that gas. **Dalton's law** states that the total pressure exerted by a mixture of gases is the sum of the pressures that would be exerted by each gas if it alone were present and occupied the total volume. The partial pressure of a gas determines the amount of that gas that dissolves into a liquid. **Henry's law** expresses gas absorption: The amount of a gas that dissolves in a liquid at a given temperature is directly proportional to the partial pressure of that gas. Table 2.2 shows the partial pressures of gases in the air at 1 ATA.

The partial pressure of a gas is the percentage of the gas in the mixture times the absolute pressure of the mixture. The partial pressure of a gas at depth has the same effect as a higher percentage of that gas at the surface. If a mixture of gas contains 2 percent carbon dioxide (CO_2) at sea level (14.7 psia, or 1.03 kg per square cm), the partial pressure of the CO_2 is 0.294 psia (0.021 kg per square cm). If the absolute pressure of the same mixture of gases increases to the pressure found at a depth of 99 feet (30 m) in the ocean (58.8 psia, or 4.12 kg per square cm), the partial pressure of the CO_2 is 1.176 psia (0.082 kg per square cm). The amount of CO_2 sensed by the body at 99 feet is four times greater than the amount that is sensed at the surface. Breathing 2 percent CO_2 at a depth of 99 feet is the same as breathing 8 percent CO_2 at the surface! A high level of CO_2 has a profound effect on respiration. The surface-equivalent effect of partial pressures makes minute amounts of contaminants in breathing gases unsafe at depth. Table 2.3 shows the surface-equivalent effect of partial pressures at various depths.

Table 2.2 Partial Pressures of Gases in Air at 1 ATA

	Percentage	Partial pressure at 1 ATA	Partial pressure of gas at 1 ATA (metric)
Nitrogen	78%	11.466 psia	0.803 kg/cm^2
Oxygen	21%	3.087 psia	0.2163 kg/cm^2
Argon	0.93%	0.137 psia	0.0095 kg/cm^2
Trace gases	0.04%	0.006 psia	0.0004 kg/cm^2
Carbon dioxide	0.03%	0.004 psia	0.0003 kg/cm^2
Totals	100%	14.7 psia	1.03 kg/cm^2

Table 2.3 Surface-Equivalent Effect of Partial Pressures

Depth	Pressure	O_2*a	CO*b	CO_2*c
0 ft (0 m)	1 ATA	20%	20 ppm	2%
33 ft (10 m)	2 ATA	40%	40 ppm	4%
66 ft (20 m)	3 ATA	60%	60 ppm	6%
99 ft (30 m)	4 ATA	80%	80 ppm	8%
132 ft (40 m)	5 ATA	100%	100 ppm	10%

*Breathing 20% oxygen at 132 ft (40 m) has the same effect as breathing 100% oxygen at the surface.

*Breathing a mixture containing 40 ppm CO at 66 ft (20 m) is the same as breathing 120 ppm CO at the surface!

*Breathing a mixture containing 2% CO_2 at 99 ft (30 m) is the same as breathing 8% CO_2 at the surface and causes shortness of breath, rapid breathing, and headache.

Additional Information About Gas Laws

The following gas formulas can be used to make precise mathematical calculations of pressure, volume, and temperature.

Boyle's law: $P_1V_1 = P_2V_2$

P_1 = Initial pressure (psia or ATA)

P_2 = Final pressure (psia or ATA)

V_1 = Initial volume

V_2 = Final volume

Example: A balloon with 2 pints of air floats from 2 ATA to the surface (1 ATA). What is the volume of the balloon at the surface?

P_1 = 2 ATA

P_2 = 1 ATA

V_1 = 2 pints

V_2 = Unknown

Rearranging the formula to solve for V_2, we find that:

$$V_2 = \frac{P_1V_1}{P_2} = \frac{(2 \times 2)}{1} = 4 \text{ pints}$$

Partial pressure = absolute pressure 3 percentage of gas

Example: What is the partial pressure of oxygen if the gas constitutes 20 percent of a gas mixture that has an absolute pressure of 58.8 psi?

$$PP = 58.8 \times 0.2 = 11.76 \text{ psia}$$

Gay-Lussac's law: $\frac{P_1}{T_1} = \frac{P_2}{T_2}$

P_1 = Initial pressure (psia or ATA)

P_2 = Final pressure (psia or ATA)

T_1 = Initial temperature (°R or °K)

T_2 = Final temperature (°R or °K)

Example: A scuba tank with a pressure of 2,250 psig and a temperature of 70 °F is heated to a temperature of 150 °F. What is the pressure of the scuba tank at the higher temperature? First, convert readings for pressure and temperature to absolute measures.

P_1 = 2,250 psig + 14.7 psia = 2,265 psia

P_2 = Unknown

T_1 = 70 °F + 460 = 530 °R

T_2 = 150 °F + 460 = 610 °R

Rearranging the formula to solve for P_2, we find that

$$P_2 = \frac{P_1T_2}{T_1} = \frac{2,265 \times 610}{530} = 2,607 \text{ psia}$$

2,607 psia – 14.7 psia = 2,592 psig

When a liquid has absorbed all of a gas that it can hold, the liquid is saturated. When you reduce the partial pressure of the gas in contact with the liquid, gas diffuses out of the liquid. This process is **outgassing**. Ingassing and outgassing provide a foundation for the dive computers and tables used to prevent decompression sickness; these tables are presented in chapter 7.

Air Consumption

The volume of air you breathe per minute during exertion is much more than the volume you breathe at rest—up to 17 times more on land and about 14 times more in the water. In the water, pressure on the torso allows only 85 percent of normal inhalation.

Because the density of the air breathed increases with depth, depth significantly affects the rate at which you consume air. For a given level of exertion, a supply of air lasts only half as long at a pressure of 2 ATA as it does at a pressure of 1 ATA. With heavy exertion at a pressure of 4 ATA (at a depth of 99 feet, or 30 m), you exhaust an air supply over 40 times faster than you would when at rest at the surface! The rapid depletion of your air supply is one reason that you must avoid heavy exertion while diving.

The rate of air consumption is expressed in cubic feet per minute (or liters per minute) or psig (or atmospheres or bars) per minute. By knowing your consumption rate for various levels of activity, you can plan your dives. When you know your consumption rate and the amount of air available, you can calculate air supply duration for future dives.

Heat, Humidity, Light, and Sound

You experience many changes when you enter water. You lose body heat faster, and you lose body moisture when you use scuba equipment. In addition, what you see is deceiving, and what you hear can cause confusion. When you understand what happens to you in water (and why it happens), you can better manage the differences between the water and air environments.

Heat Transfer

The net effect of the various forms of heat transfer is that you can chill quickly while diving. Radiation, convection, and conduction transfer heat from one medium to another (see figure 2.12). Heat waves radiate from exposed surfaces, heat travels upward through fluids by convection, and heat is transferred directly via conduction between substances in contact with each other. Metals are good conductors. Water is a poor conductor compared with metal, but water conducts heat about 25 times faster than air (depending on density). Conduction and convection are the primary means by which heat is transferred from a diver to the surrounding water. Heat rises from the skin, and water carries the heat away. You also lose body heat through the process of evaporation. Moisture evaporates from your lungs when you are breathing underwater and from the surface of your skin when you perspire above water. Scuba equipment expands high-pressure air and cools it. Your body heat warms the air you breathe, and you lose the heat with each exhalation.

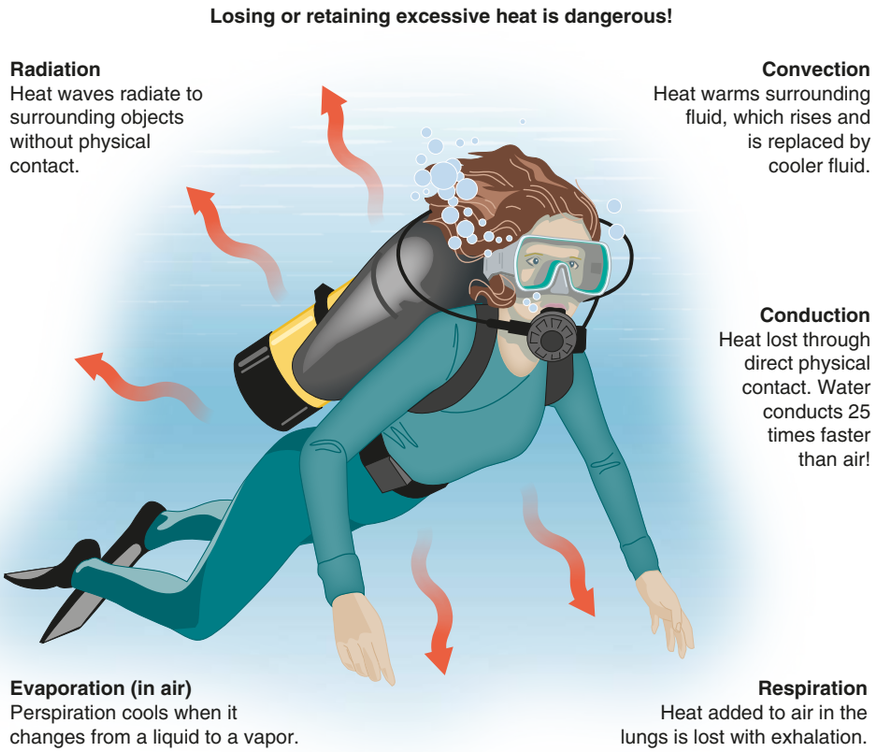

Figure 2.12 Heat transfer and loss.

You can slow the transfer of heat by insulating yourself with a material that is a poor conductor of heat. Exposure suits help insulate you from the environment, but insulation does not help reduce heat lost through respiration. The next chapter presents ways to manage the problems of heat loss.

Humidity

Scuba divers must be aware of the effects of humidity, or the amount of water vapor present in a gas. The temperature of the gas determines the amount of water vapor a gas can absorb and retain. The warmer the gas, the more humidity the gas can contain.

You humidify inspired air. The process of compressing the air that is put into scuba tanks dehumidifies the air in the cylinder. You lose moisture from body tissues when you breathe dry scuba air. The resulting fluid loss can cause partial dehydration, an undesirable condition, especially for a scuba diver. In the next chapter, you will learn how to avoid the problems of dehydration.

Diving poses other humidity challenges that you must manage. Moisture in the air inside your mask condenses on the faceplate as the air cools. Unless you thoroughly clean your mask lens in advance so that the condensation runs off in a thin sheet, foggy beads of condensation will form and blur your vision. Chapter 6 presents the process for cleaning, or defogging, your mask.

Additional Information About Air Consumption

The following air consumption formulas can be used to calculate consumption rates and air supply duration. To calculate the rate of air consumption, you need three items of information: the depth at which you have remained for a period of time, the length of time you have remained at that depth, and the amount of air you have used during that time.

1. Determine your depth air consumption rate (DACR). This is simply the amount of air used divided by the time at depth. For example, the DACR for a diver who uses 1,000 psi (68 atm) in 10 minutes is 100 psi (6.8 atm) per minute.

$$\text{DACR} = \frac{\text{Air used}}{\text{Time at depth}}$$

2. Convert the DACR to the surface air consumption rate (SACR). You need to express the rate in terms of volume rather than pressure. After you do this, you can apply the air consumption rate to any depth and to a cylinder of a different size from the one used initially to calculate the air consumption rate. Obtain the surface consumption rate by multiplying the DACR by the ratio of the pressure at the surface to the pressure at depth. Because you can express pressure in terms of depth, you can use the following formula:

$$\text{SACR} = \text{DACR} \times \frac{33\text{ ft (or 10 m)}}{\text{Diving depth} + 33\text{ ft (or 10 m)}}$$

   If, for example, your depth consumption rate for a depth of 33 feet (10 m) is 30 psi (2 atm) per minute, your SACR is 30 (33/66) = 15 psi per minute, or 2 (10/20) = 1.0 atm per minute.

3. Convert the rate to volume, establish a ratio of the tank volume and pressure (when the tank is full) to the breathing rate volume and pressure, then solve for the breathing rate volume (BRV) as follows:

$$\frac{V_1}{P_1} = \frac{V_2}{P_2}$$

   where

$$V_1 = \text{Full tank volume}$$

$$V_2 = \text{Breathing rate volume (BRV)}$$

$$P_1 = \text{Full tank pressure}$$

$$P_2 = \text{Breathing rate pressure}$$

   so

$$\text{BRV} = \frac{V_1 \times P_2}{P_1}$$

   For example, the breathing rate volume (BRV) for a diver with an 80-cubic-foot (2,265 L), 3,000 psi (204 atm) tank and an SACR of 30 psig (2.04 atm) per minute is

$$\text{BRV} = 80\text{ ft}^3\ 3\ 30\text{ psi/min}/3{,}000\text{ psi} = 0.8\text{ ft}^3/\text{min}$$

$$\text{Metric BRV} = 2{,}265\text{ L } 3\ 2.04\text{ atm/min}/204\text{ atm} = 22.65\text{ L/min}$$

4. For the same level of activity, you can calculate the approximate duration (in minutes) of any amount of air from a tank of any size used at any depth. Here is an example: How long will 1,750 psi (119 atm) of air from a 71.2-cubic-foot (2,016 L), 2,475 psi (168 atm) tank last at a depth of 70 feet (21.3 m) for a diver with a breathing rate volume (BRV) of 0.8 cubic feet (22.7 L) per minute?

First, determine the volume of air in the tank at a pressure of 1,750 psi (119 atm). The formula for determining the volume of air in the tank is

$$V_2 = \frac{V_1 \times P_2}{P_1}$$

where

V_1 = Full tank volume

V_2 = Partially filled tank volume

P_1 = Full tank pressure

P_2 = Partially filled tank pressure

The air supply volume for the partially filled tank is therefore

$$V_2 = \frac{71.2 \times 1{,}750 \text{ psi}}{2{,}475 \text{ psi}} = 50.3 \text{ ft}^3\text{, or}$$

$$V_2 = \frac{2{,}016 \text{ L} \times 119 \text{ atm}}{168 \text{ atm}} = 1{,}428 \text{ L}$$

The formula for air supply duration (ASD) is

$$\text{ASD} = \frac{\text{Air supply volume}}{\text{BRV}} \div \frac{\text{Diving depth} + 33 \text{ ft (or 10 m)}}{33 \text{ ft (or 10 m)}}$$

The air supply duration (ASD) for the question posed earlier is

$$\text{ASD} = \frac{50.3 \text{ ft}^3}{0.8 \text{ ft}^3 / \text{min}} \div \frac{70 \text{ ft} + 33 \text{ ft}}{33 \text{ ft}} = 20 \text{ min}$$

$$\text{Metric ASD} = \frac{1{,}428 \text{ L}}{22.7 \text{ L} / \text{min}} \div \frac{21.3 \text{ m} + 10 \text{ m}}{10 \text{ m}} = 20 \text{ min}$$

These calculations may seem complicated at first, but the ideas are simple. The calculations become easy with practice. Let's review the four steps of air consumption calculations: (1) Determine your depth air consumption rate (DACR); (2) determine your surface air consumption rate (SACR); (3) determine your breathing rate volume (BRV); and (4) determine the air supply duration (ASD) for a quantity of air. The abbreviated formulas for the calculations are as follows:

$$\text{DACR} = \frac{\text{Air used}}{\text{Time at depth}}$$

$$\text{SACR} = \text{DACR} = \frac{33 \text{ ft (or 10 m)}}{\text{Diving depth} + 33 \text{ ft (or 10 m)}}$$

$$\text{BRV} = \frac{V_1 \times P_2}{P_1}$$

$$\text{ASD} = \frac{\text{Air supply volume}}{\text{BRV}} \div \frac{\text{Diving depth} + 33 \text{ ft (or 10 m)}}{33 \text{ ft (or 10 m)}}$$

In freezing temperatures, moisture from your exhaled breath can cause a scuba regulator to freeze. Water in other items of diving equipment can also freeze. If you intend to dive in a cold environment, you should complete special training and should know how to prepare and use your equipment in those conditions.

Light and Vision

The density of water makes it challenging for divers to interpret what they see and hear. Light travels faster in air than in water. When rays of light traveling in water pass through the lens of your mask, they accelerate and bend (refract). The effect is that what you see underwater is magnified. Objects appear to be at three-fourths of their actual distance (25 percent closer) and four-thirds of their actual size (33 percent larger). The visual distortion requires adjustments, and you will learn to make these with experience. An object that is 12 feet (3.7 m) away appears to be only 9 feet (2.7 m) away. A fish that appears to be about 2 feet (0.6 m) long is actually only 1.5 feet (0.5 m) in length. Many new divers discover that items they bring back from diving are much smaller than they perceived them to be underwater. Figure 2.13 illustrates how light is perceived differently in water.

One difficulty caused by the refraction of light is that distant objects appear closer than they are. This can create a hazard in clear water when you look downward from a drop-off. You may be tempted to go to a point that appears to be at a safe distance but might actually take you beyond your planned maximum

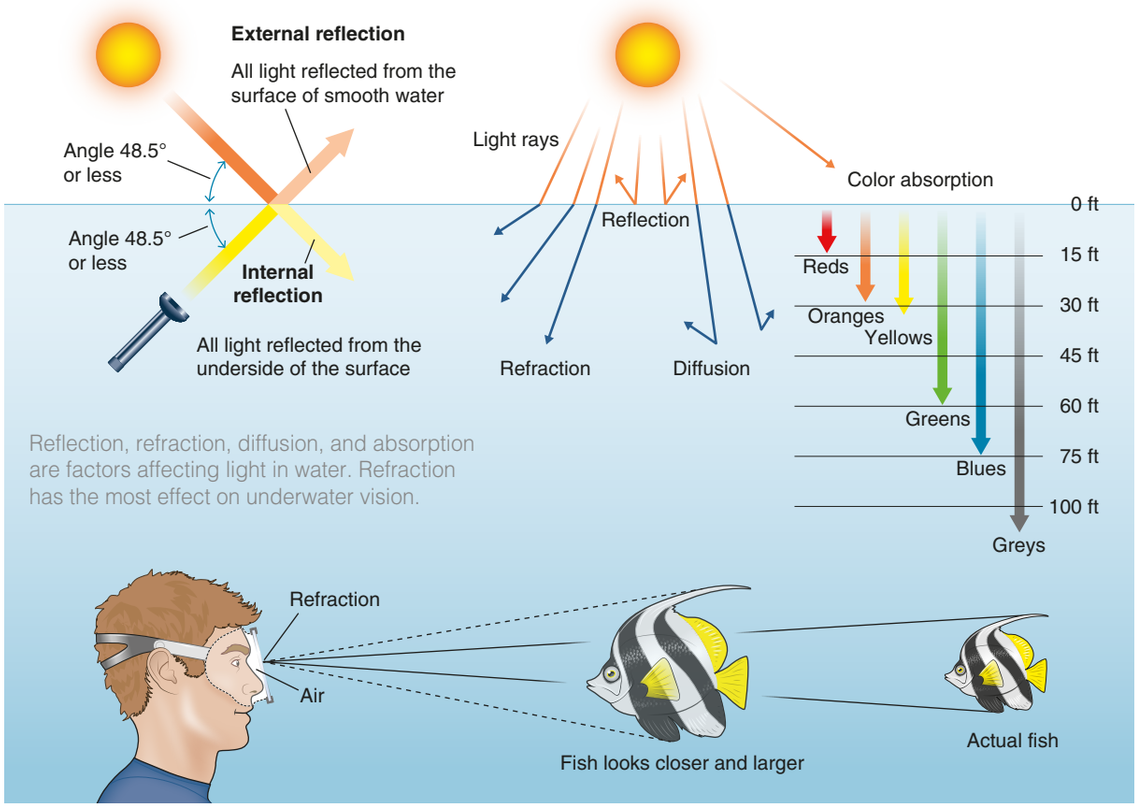

Figure 2.13 Visual perception is distorted underwater.

SCUBA WISE

There is an old joke about a young woman who returned home from her first session of scuba diving training and was unenthusiastic when asked whether she enjoyed the class. Her reply was "First the instructor told us all the ways that we could die, then he worked us in the water until I thought that I was going to die." Fortunately, instructors do not teach classes that way today. The reason for learning about potential injuries is to understand how to prevent them. Armed with knowledge that you can convert to wisdom through proper application, you can avoid all of the potential problems identified in this chapter. The use of scuba diving instruction for fitness training is also outdated. Reasonable physical fitness is a prerequisite for scuba diving. Any healthy person who can swim 200 yards (61 m) continuously is fit enough to dive. The objective of modern scuba training is to help you learn to relax in and under the water, not to whip you into shape.

depth. You must realize that distance perception is inaccurate, and you must rely on your depth gauge instead of your vision.

You have two types of vision: day vision and night vision. You use different parts of your eyes for each type. When you move from a brightly lit area into a dimly lit area, your vision needs 15 to 30 minutes to adapt to the lower level of light. Even after the adaptation, your ability to see fine details is much less than your ability with day vision. In addition, particles in water diffuse, scatter, and attenuate light. The deeper you go, the less light there is. In **turbid** water, the amount of light decreases very quickly with depth. A dive in turbid water involves a change from day vision to night vision.

Water affects light in many ways. The surface of water reflects light. When light strikes the surface of calm water at an angle less than 48 degrees, the water reflects all the light. Early morning and late afternoon sunlight do not penetrate calm water. Light rays going toward the surface from underwater also reflect from the interface, making the underside of the surface appear like a mirror when viewed at the proper angle.

Objects you view underwater often lack their natural colors. White light, such as sunlight, comprises all the colors of the spectrum. Increasing depth absorbs various colors. The water absorbs warm colors, such as red and orange, with only 30 feet (9 m) of depth. Cooler colors, such as blue, penetrate deepest. This is why deep, clear water is blue. The underwater scene appears drab at depths below 100 feet (30 m). Fortunately, you can restore all the colors of the spectrum underwater by using an artificial light at close range.

Another visual challenge is disorientation. When you are weightless in water and do not have a visual reference, **vertigo** can result. Vertigo is the feeling of movement perceived as a spinning sensation. This problem can occur when the water is turbid and visibility is poor, when the water is crystal clear but there are no visual clues for orientation, and when waters of different densities mix together. You will learn how to manage the challenges of disorientation in the next chapter.

A final visual problem is poor visibility in water compared to air. In air environments, visibility is measured in miles, but in water it is measured only in feet. Particles suspended in water cause turbidity, which hampers vision. Reduced visibility makes it difficult to locate objects underwater, to maintain contact with your buddy, and to find your way. You must develop skills to cope with these challenges, including search techniques, buddy system techniques, and underwater navigation skills.

Sound and Hearing

Sound travels well in water. You can hear many sounds while diving. On land you can determine the direction of a sound by the difference in the time it takes the sound to reach one ear and then the other. This interval is brief but sufficient for your brain to discern it. Underwater, you hear sound conducted through water and the bones of your head to your inner ears. Sound travels about four times faster in water than in air. The time delay between a sound reaching one ear and then the other is so small that directional discernment is difficult underwater.

Sound does not transfer well from air to water. Only 0.01 percent of sound can travel directly between air and water. You must use special devices to make voice communications effective underwater.

Summary

The aquatic environment affects your body in several ways. Pressure affects your air spaces: your ears, sinuses, and lungs. The rate of change of pressure in water is many times greater than in air and increases as you approach the surface during ascent. Changing pressures also affect the diffusion of gas into and out of liquids. The level of carbon dioxide in your body controls your respiration, and the partial pressure of the carbon dioxide is affected by depth. Temperature also affects the pressure of a constant volume of gas. Depth and activity have the greatest effect on air consumption. Water absorbs heat from the body, so divers need insulation to help prevent excessive heat loss. Humidity can cause several problems for scuba divers. The difference in density between air and water affects your buoyancy, mobility, heat loss, vision, and hearing. Now that you are aware of the effects of the aquatic environment on your body, you are prepared to learn to adjust to the underwater environment.

Application-of-Knowledge (AOK) Questions

1. If you are exerting underwater and find yourself breathing rapidly and deeply and suddenly feel that you are not receiving enough air, what should you do?
2. A sudden change of conditions while you are diving increases your anxiety, and you notice that your breathing is rapid and shallow. Your anxiety level increases because you feel unable to manage the situation in which you find yourself. What action will help alleviate your anxiety and abnormal breathing?
3. While diving with an experienced diver, you compare tank pressures and discover that you are using twice as much air as your buddy, who is the same size as you. You are using equipment that is identical to your buddy's. What could be causing you to have the high rate of air consumption?
4. Maintained scuba equipment is extremely reliable, but if your regulator started leaking water when you inhaled and you were at a depth of 60 feet (18.3 m), what should you do?
5. While swimming in a pool, you notice a swimmer breathing rapidly and deeply before swimming underwater. You keep an eye on the swimmer because he may lose consciousness without warning while submerged. What is the cause of this problem, and what could you tell the swimmer to help him understand how to prevent an accident?
6. You are diving in water that has a temperature of 60 °F. (15.6 °C) and begin shivering. You decide to increase your activity level to produce more body heat. Is this a good course of action?

Cup corals in the Red Sea

3

Diving Adjustments

Dive In and Discover

By the end of this chapter, you will be able to do the following:

- Define the terms *hypothermia*, *hyperthermia*, *vasoconstriction*, *heat exhaustion*, *heatstroke*, *equalization*, *Toynbee maneuver*, *Valsalva maneuver*, *trapdoor effect*, *skip breathing*, *barotrauma*, *pulmonary barotrauma*, *arterial gas embolism*, *mediastinal emphysema*, *subcutaneous emphysema*, *pneumothorax*, *vertigo*, *perfusion*, *ambient pressure*, *half-time*, *compartment*, *controlling compartment*, *decompression illness*, *dehydration*, and *nitrogen narcosis*.
- State the cause, effect, signs and symptoms, first aid treatment, and prevention of hyperthermia, hypothermia, squeezes, reverse blocks, respiratory distress, pulmonary barotrauma, vertigo, seasickness, decompression illness, and nitrogen narcosis.
- Explain three ways to control buoyancy while scuba diving.
- Explain when to equalize pressure in your ears during descent and what to do if the pressure does not equalize.
- Explain how to minimize resistance to movement for underwater swimming.

To function effectively underwater, you need specific attitudes, equipment, knowledge, and skills. This chapter provides the fundamental knowledge and begins shaping the attitudes required to minimize the risk of injury. You will apply the basic aspects of anatomy, physiology, and physics you have learned to a typical person descending into the depths and ascending back to the surface.

Thermal Adjustments

Maintaining your body's core temperature within a few degrees of normal is challenging in water. When you are immersed in water, you lose body heat rapidly. A water temperature of 50 °F (10 °C) can incapacitate an unprotected diver within 15 minutes. Even water at a temperature of 80 °F (27 °C), which feels relatively warm, can chill a diver within an hour. Wearing only a bathing suit in 80 °F (27 °C) water is the same as being without any clothing in air that is 42 °F (6 °C).

Your brain regulates your body functions to maintain your body temperature. If your core temperature is less than 95 °F (35 °C), you will suffer from **hypothermia**. You need to guard against mild and severe hypothermia, both of which can be dangerous. If your core temperature is higher than normal, you experience the effects of **hyperthermia**. You need to understand the effects of two types of hyperthermia—heat exhaustion and heatstroke—and you must try to prevent them. They both can be dangerous.

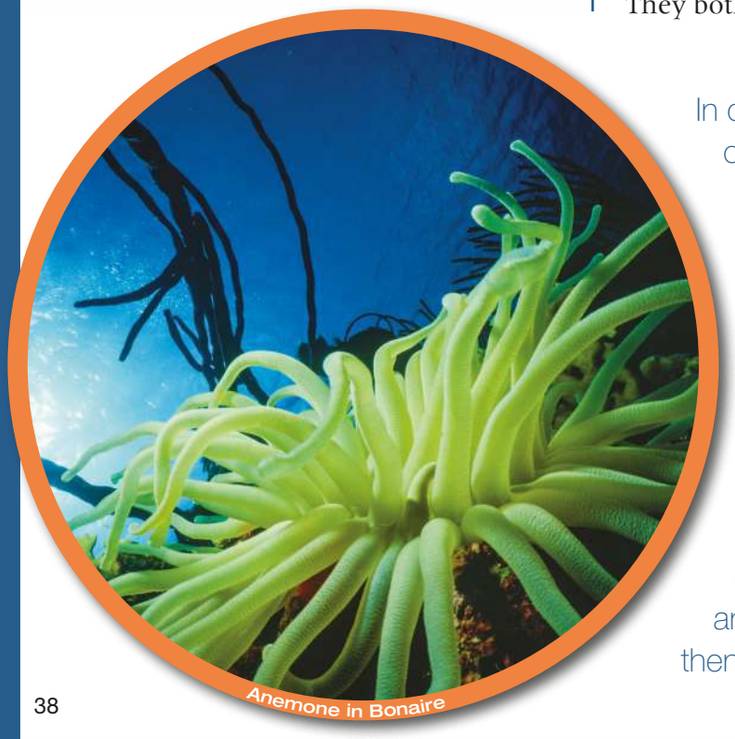

Anemone in Bonaire

In chapter 2, you learned about the effects of water on your body. In this chapter, you will learn how to deal with those effects. Most people with average intelligence and normal health can make adjustments to the aquatic environment, such as buoyancy control and the equalization of pressure. You make many adjustments automatically, but some you must make consciously. Learning to make the transition from an air environment to an underwater environment requires professional instruction and guidance. You need to learn what to do, then do what you have learned.

Heat Loss

Your body has a variety of physiological responses to the loss of heat. Respiration increases automatically when you get chilled. This is undesirable because you heat and moisturize inspired air, and you lose the heat and moisture with each exhalation. The more you breathe, the more heat and moisture you lose to the environment (and the shorter your dive time). Water depth compounds the problem because the greater the surrounding pressure, the greater the density of the air that you breathe. Denser air absorbs more heat than air that is less dense. The deeper you dive, the quicker you get cold.

Anything that affects the function of your body—excitement, fear, seasickness, and other forms of illness—can increase heat loss. This is why good health and a confident state of mind are safety recommendations.

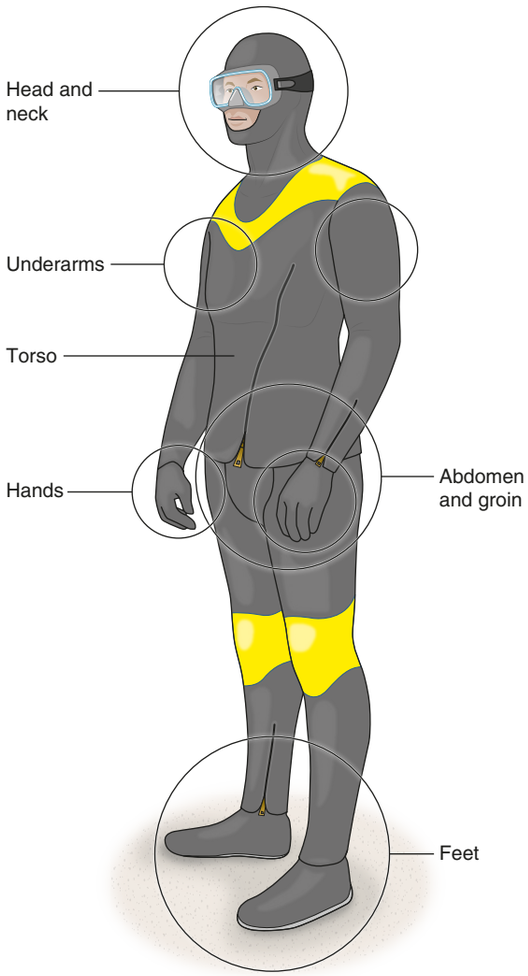

Figure 3.1 Areas prone to high heat loss.

One way your body responds to cold is to shunt blood from the extremities through **vasoconstriction**. The circulatory shunting reduces heat loss because it keeps warm blood from passing through areas of your body that have little insulation. Your head, neck, underarms and sides, groin, hands, and feet are the areas of your body most prone to heat loss underwater (see figure 3.1). Fortunately, you can insulate these areas easily. In cold water, you can lose considerable heat from your head and neck because the head receives a large supply of blood and lacks natural insulation. Your body does not shunt blood from your head as it does from other body extremities. In water at a temperature of 70 °F (21 °C) or less, you must be sure to insulate your head and neck.

Hands have large surface areas relative to their volume. To prevent excessive heat loss from your hands when you get cold, your body shunts blood from them until they reach a temperature of 50 °F (10 °C). At this temperature, your body restores circulation to your hands to partially rewarm them. Hands quickly lose their warmth to the water. If you dive without hand protection in cold water, you lose body heat through your hands. Because your hands may become numb and lack strength when they get cold, you should insulate them to keep them functioning and to conserve body heat.

Small people chill more quickly in water than large people do. Small people have less muscle mass to generate and store heat. Insulation is important for all divers, but protection against heat loss is more critical for those of smaller stature.

Failure to wear adequate insulation leads to hypothermia. Hypothermia also results from repeated or prolonged exposure. Slow chilling of your body is undesirable. You lose muscle strength and feeling, and your muscles may cramp. Severe heat loss also affects your ability to reason. Another body response to heat loss is shivering, which restores heat through muscular activity. Shivering, which generates about five times as much heat as your body produces at rest, is helpful on land, but is not beneficial in water. Water conducts away the heat you produce by shivering, and you get colder. Uncontrollable shivering indicates that you have lost too much heat from the core of your body and that you cannot rewarm yourself without getting out of the water. When you are shivering, you should terminate the dive. Rewarm yourself thoroughly before diving again. To help your body temperature return to normal, you should put on clothes that are warm and dry, stay in warm surroundings, and consume warm nonalcoholic drinks. You can become so cold that shivering ceases. This occurs because severe hypothermia has incapacitated the body's ability to produce heat. A cold person who is not shivering requires medical care.

You must understand that warming the surface of your body is different from warming the core of your body. You may feel warm, but your deep core temperature may remain below normal. If you return to the water in this condition, you will quickly become chilled. The only way to be sure that you are thoroughly warmed is to keep warming yourself until you begin to perspire. Perspiration occurs when the core temperature begins to rise above normal. To warm yourself, get dry, wrap everything except your face in blankets, and stay that way until you perspire. Most people will shed the blanket when their skin is warm, not being aware that the core of their body is still below normal temperature.

Overheating

You can prevent excessive loss of body heat by insulating your body with an exposure suit, but insulation can cause another problem. When you insulate your body to reduce heat loss in water, you reduce your body's ability to rid itself of excess heat above water. The evaporation of perspiration helps cool your body, but if you cover your body, perspiration cannot evaporate. You may become overheated in warm climates when you are preparing to dive. Maintaining your body temperature within acceptable limits before, during, and after dives can be a challenge. Figure 3.2 describes some of the causes and effects of overheating.

When a person is unable to stop the rise of core temperature, **heat exhaustion** occurs. This condition is serious. A person affected by heat exhaustion becomes weak and may collapse. The victim looks pale and feels sweaty. Place a person with this condition in a cool place, remove the exposure suit, and take steps to lower the person's body temperature. If the victim is alert and not nauseous, have him or her drink a diluted electrolyte fluid or water to replace lost body fluids.

A more serious form of hyperthermia is **heatstroke**, which occurs when the body temperature becomes so high that the body's temperature-regulating ability shuts down. A victim of heatstroke looks flushed and has hot, dry skin. The mental status of a heat stroke victim is not normal. The victim may be delirious, or unresponsive, or have seizures. Heat stroke is extremely serious. Cool the patient's body immediately and summon medical assistance.

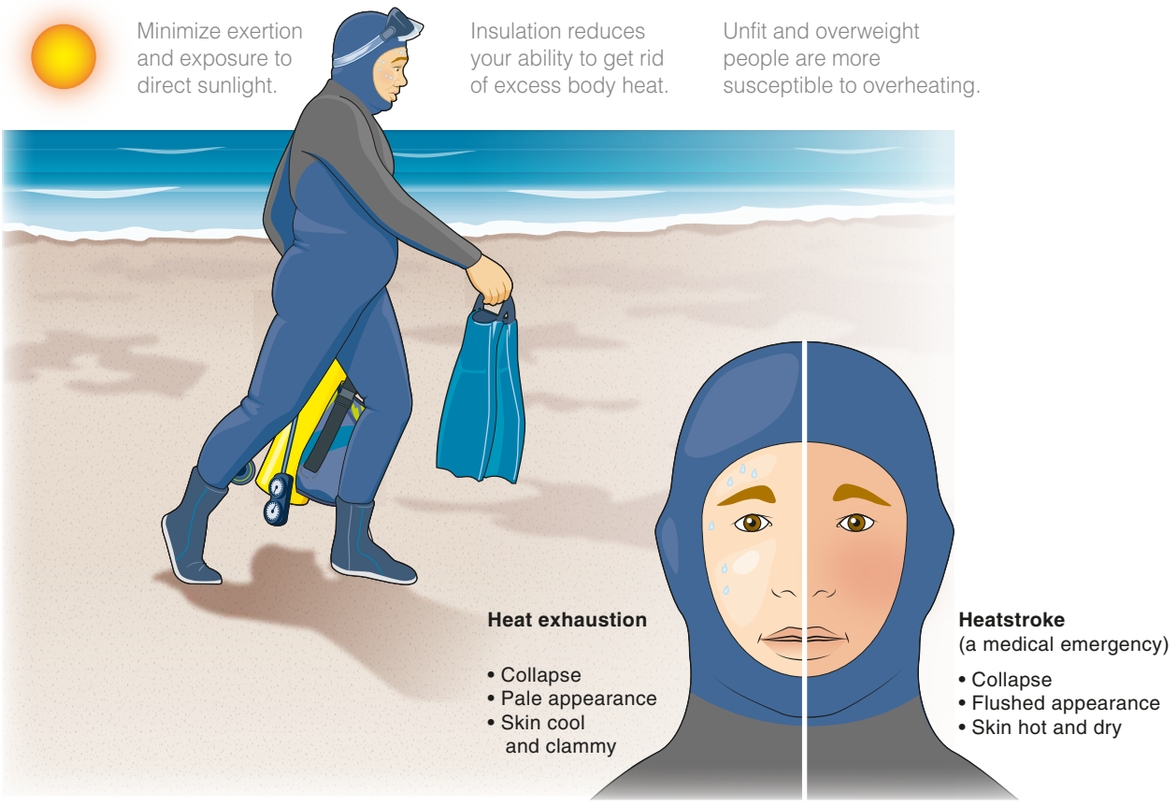

Figure 3.2 Take precautions against overheating by learning the symptoms of heat stroke and exhaustion.

Preventing hyperthermia is much better than treating it after it occurs. You should avoid prolonged exposure to warm temperatures when wearing insulation. If the air temperature is warm, douse yourself with water after donning your exposure suit and before donning the remainder of your diving equipment. Stay out of direct sunlight if possible. All thermal considerations for diving are especially important for divers whose physical fitness is marginal.

Buoyancy Adjustments

Exposure suits and other equipment affect your buoyancy. You must adjust the amount of weight you wear to control your depth in the aquatic environment. When you are too buoyant, you must fight to remain submerged; when you carry excess weight, you must work hard to keep from sinking or to stay off the bottom. You need to maintain neutral buoyancy underwater and positive buoyancy at the surface. In chapter 2, you learned the principle of buoyancy and the three states of buoyancy: positive, negative, and neutral. Now you will learn some practical applications of buoyancy.

Your body—which is composed of solids, liquids, and air spaces—has an average density nearly the same as water. A typical human body immersed and relaxed in water has a positive buoyancy of a few pounds when the lungs are filled with air; it has a negative buoyancy of a couple of pounds when the lungs contain

the minimum amount of air. Fortunately, the density of water and resistance to movement prevent you from moving up and down rapidly when you breathe.

Factors Affecting Buoyancy

People usually wear some type of exposure suit when diving. Most exposure suits increase buoyancy, so divers wear weights to offset the buoyancy of the suit and achieve neutral buoyancy. These weights are made of lead, which is about 12 times denser than water.

Your initial state of buoyancy in water will depend on the volume of water you and your equipment displace. You can vary your volume—with a negligible increase or decrease in your weight—by adding air to or venting air from your buoyancy compensator (BC). Increasing BC volume increases buoyancy, while decreasing BC volume decreases buoyancy.

Your buoyancy is affected by your physical size and weight, your lung capacity, the equipment you wear, and the items you carry (see table 3.1). Exposure suits use air or small bubbles of gas for insulation. When you descend while wearing an exposure suit, pressure compresses your suit and reduces its volume, so you become less buoyant. You must add air to your BC to compensate for buoyancy lost from suit compression. On the other hand, buoyancy increases as you consume air from your scuba cylinder. Air weighs 0.08 pounds per cubic foot (1.3 g per L). A typical scuba tank contains 80 cubic feet (2,265 L) of air. A full tank weighs 6 pounds (2.7 kg) more than an empty tank. As you consume air from your scuba tank, you must vent some of the air that you added to your BC to compensate for suit compression. The trade-off helps you keep buoyancy constant during a dive.

Inflating your lungs increases your buoyancy, while deflating them reduces buoyancy. A high average lung volume makes you float; a low average volume makes you sink. When you become excited or begin moving quickly, your respiration increases, which affects your buoyancy. For optimal control of buoyancy, you should maintain a calm, relaxed state.

The density of water also affects buoyancy. Saltwater is denser than freshwater, so you are more buoyant in the ocean than in a lake. This means that if you

Table 3.1 Factors Affecting Buoyancy

Factor	Effect
Size and weight of diver	Obese divers are more buoyant.
Type and amount of equipment	Larger equipment is more buoyant (because of increased volume).
Amount of weight worn	Weights decrease buoyancy.
Amount of air in BC	Increasing volume increases buoyancy.
Amount of air in tank	Buoyancy increases as air decreases.
Amount of air in lungs	Exertion or excitement increases volume and buoyancy.
Suit compression	Pressure decreases volume and buoyancy.
Items carried	Added weight decreases buoyancy.
Type of water (salt or fresh)	Denser water increases buoyancy.

SCUBA WISE

At some point, many divers discover a special feeling of becoming part of the underwater world. I remember vividly an overwhelming feeling—a strange combination of peace and exhilaration—that I experienced during a dive in the Red Sea. The goal of training is to help you adjust to a new environment. When you adjust, you can relax, and when you can relax, you can focus more on your exciting new surroundings than on yourself. Meeting the demands of diving is challenging, but when you succeed, the exuberance is worth every adjustment you have to make. People can adjust well to new situations. Training helps people adjust more quickly, easily, and safely than they can by trial and error. Study the remaining chapters carefully to learn how to adjust to the subaquatic environment. I want you to feel what I felt in the Red Sea and have felt many times since while diving. The feeling of being one with the sea is powerful, moving, and unforgettable.

are weighted for neutral buoyancy in the ocean, you must remove some weight to achieve neutral buoyancy for freshwater diving. The amount of weight you remove is about 3 percent of the combined dry weight of you and your equipment. For example, if a neutrally weighted 160-pound (73 kg) diver with 60 pounds (27 kg) of equipment, including 16 pounds (7 kg) of weights, wants to dive in freshwater instead of seawater, the diver must remove about 7 pounds (3 kg) of weight in order to be weighted correctly.

Ways to Control Buoyancy

As a diver, you can control buoyancy in three ways: by the amount of weight you wear and carry, by the amount of air in your BC, and by the amount of air in your lungs. These means of control are coarse, medium, and fine adjustments, respectively. The skills you need to learn in order to adjust to the aquatic environment include determining the correct amount of weight to be worn, regulating the amount of air in your BC, and varying your breathing for minor buoyancy adjustments. Chapter 6 describes these skills.

Pressure Equalization

One of the most important adjustments you must learn is how to handle the effects of pressure changes in water. Pressure changes rapidly as you descend and ascend. To avoid discomfort and injury, you must keep the pressure in the air spaces inside and on your body equalized with the outside pressure. This section identifies the procedures for pressure equalization.

Equalizing Pressure in the Sinuses

Your sinuses equalize pressure automatically as long as they are healthy. But when you have a cold or respiratory illness, the membranes lining your sinuses become swollen. The swelling can close the narrow air passages leading to the sinuses. If you descend with swollen sinus membranes, a sinus squeeze will result. Because the sinuses are formed in bone, they do not compress as a flexible container does. When the pressure inside the sinuses is less than the surrounding pressure, the outside pressure causes the linings of the sinuses to swell and ooze fluids into the cavities. The swelling and fluid reduce the volume and compress the air that is there. Do not attempt this painful method of equalization

because your problems are not over after you descend. When you ascend in this condition, the compressed air in the sinuses expands to its original volume. The expansion can force the fluid in your sinuses out through the openings, which are swollen shut. Avoid this painful process; don't dive unless your sinuses are clear and normal. Figure 3.3 shows what happens to both healthy and congested sinuses under pressure.

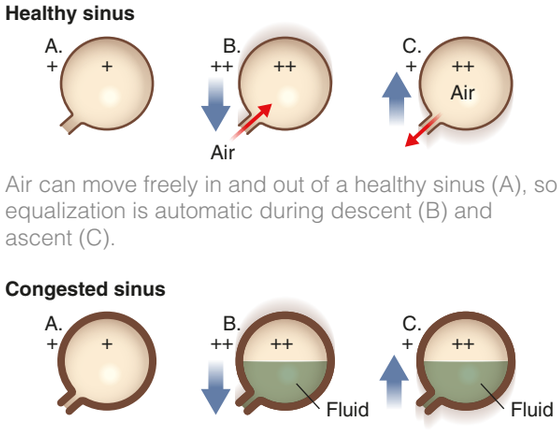

Air can move freely in and out of a healthy sinus (A), so equalization is automatic during descent (B) and ascent (C).

Air cannot get into or out of a congested sinus (A). The body forces fluids into the sinus at depth to equalize the pressure (B). The air in the sinus tries to expand to its original volume during ascent (C).

Figure 3.3 The mechanisms of equalizing pressure in the sinuses.

Do not use medications to relieve stuffiness and congestion caused by an illness before you go diving. Increased pressure can reduce the medication's effects and the duration of its effectiveness. Decongestants do not cure an illness; they simply mask its symptoms. When medication taken to open swollen airways wears off, a rebound effect occurs. The airways become more swollen than they were before you took the medication. If the rebound occurs while you are diving, you can trap high-pressure air in your sinuses. If you are not well enough to dive without medications, you should not dive.

Equalizing Pressure in the Ears

You must equalize the pressure in your ears more often than any other air space. Whereas the air passages to your sinuses are normally open, the **eustachian tubes** from your throat to your ears are normally closed. You must learn how to open the tubes at will to allow air to pass through them and equalize pressure in the air spaces of your middle ears.

You can make several movements to open the ends of the eustachian tubes where they connect with the throat. Swallowing, yawning, lifting the base of your tongue, and jutting the jaw forward (either individually or in combination) should produce a cracking sound in your ears. The opening of the tubes causes the sound. Some divers are fortunate because they can use simple movements to equalize pressure in their ears during descent. Most people require a more forceful means of equalization. Many divers use a technique known as the **Toynbee maneuver**, which you do by blocking your nostrils, closing your mouth, and swallowing. The **Valsalva maneuver** is even more forceful and is done by blocking your nostrils, closing your mouth, and gently attempting to exhale.

You must avoid excessive force when you use the Valsalva method because you could damage your ears permanently by rupturing the round window. When pressure in the outer ear increases, the eardrum bulges inward. The bones of hearing in the middle ear transmit the movement to the oval window in the inner ear. When you attempt to exhale against closed airways, you create an internal pressure on your inner ears. The attempted exhalation pressure in conjunction with the external pressure exerted on the oval window can cause the

round window to rupture; this is a serious injury that can result in a permanent high-frequency hearing loss and constant ringing in the ear. Because you control how hard you attempt to exhale in a Valsalva maneuver, you can prevent this injury. Be careful!

You must equalize pressure in the air spaces of your middle ear early (before descent) and frequently. If you delay equalization during descent, increasing pressure will hold your eustachian tubes closed—and if you attempt to force air through them, this will only close them tighter. This is the **trapdoor effect**, a difficulty you can avoid by keeping the pressure in your middle ears and your throat equal. You can then open your eustachian tubes and allow air to pass through them. Figure 3.4 shows the trapdoor effect.

You should equalize pressure in your ears before descending and about every 2 feet (0.6 m) for the first 15 feet (4.6 m) of descent, about every 3 feet (0.9 m) from 15 to 30 feet (4.6 to 9.1 m), and as needed thereafter. You can feel and hear air enter your ears when you eliminate any difference in pressure. If you attempt to equalize pressure but cannot get air into your ears, you should ascend several feet to reduce the pressure on your eustachian tubes and try again. If that works, continue your descent. If that does not work, ascend a few more feet and try again. Your initial descents may be somewhat jerky until you become accustomed to equalization techniques.

Failure to equalize pressure in your middle ears is as bad as trying to equalize forcefully. In an unequalized middle ear, pressure bows the eardrum inward. If the squeeze continues, pressure forces blood and fluid into the middle ear. This

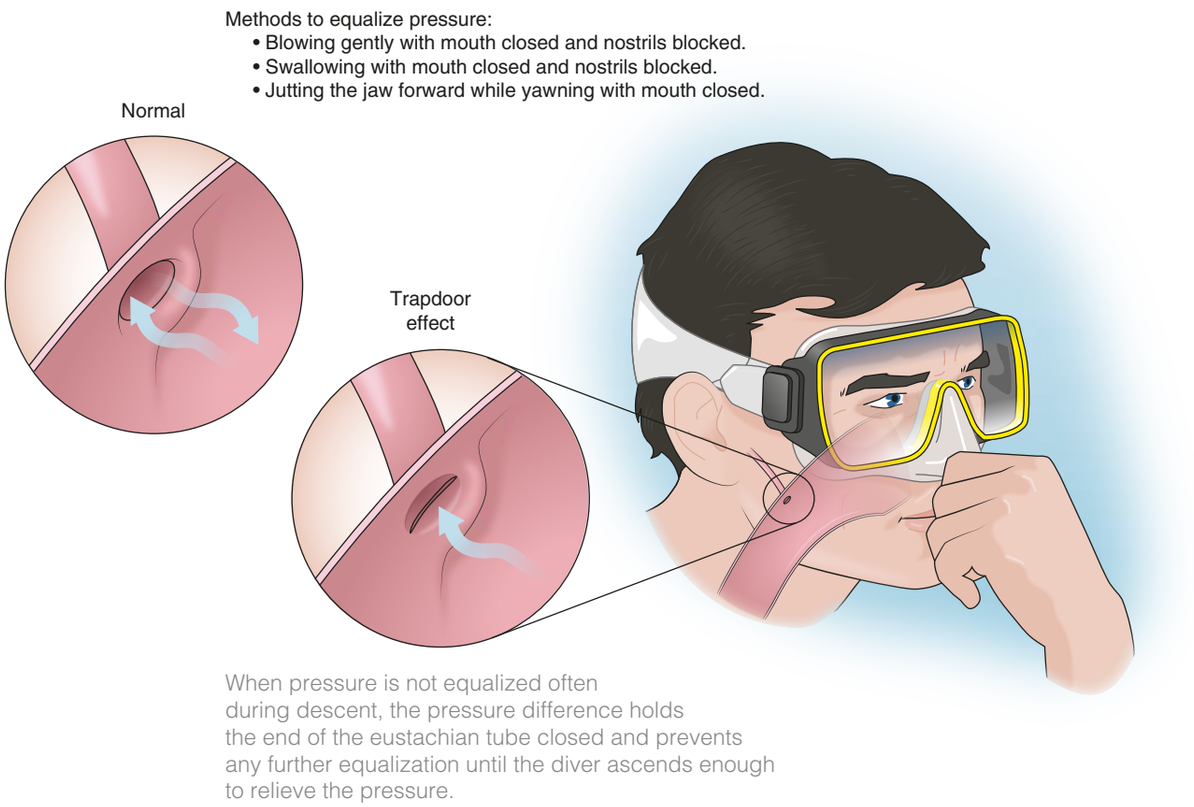

Figure 3.4 Helpful techniques for pressure equalization.

process takes time, but it is usually painful. It also damages the ear. If you feel discomfort or pressure in your ears during descent, you should ascend until the pain is gone, and then ascend another couple of feet before trying to equalize.

If you ignore an ear squeeze and continue your descent, the pressure differential can rupture your eardrum. The rupture instantly equalizes the pressure in your middle ear but damages your ear in the process. A rupture of your eardrum causes a temporary loss of hearing and a feeling of fullness in your ear. Cold water rushing into the middle ear can also cause temporary disorientation (see the section on Equilibrium Adjustments and Seasickness later in this chapter). If you suspect an ear injury, see a physician; prompt treatment can minimize the risk of permanent injury.

You can cause an outward rupture of an eardrum if you block the ear canal with a plug or if you cover your ears with a watertight covering. When you obstruct an ear canal, the air in your outer ear remains at surface pressure while the pressure in your middle ear increases with equalization during descent. The difference in pressure between the middle and outer ear pushes the eardrum outward until it breaks. You can prevent such an injury. Do not wear earplugs while diving, and avoid waterproof seals over your ears.

Equalizing is easier when you descend in an upright position as opposed to a head-down position. Membranes line the airways in your head, and gravity affects blood in the vessels within the membranes. When you are upside down in water, the membranes of your air passages swell and narrow.

Difficulties with ear equalization during ascent are uncommon. Air expanding inside the middle ear escapes through the eustachian tube. Air passes out through the tube much more easily than it goes in; you do not have to do maneuvers to open your tubes so the air can escape. But if a plug of mucus happens to block a tube, pressure can build up in the middle ear and cause an uncomfortable reverse block. If you feel pain or pressure in an ear during ascent, stop the ascent. Excess pressure inside the ear usually works its way out if given time. If you are forced to surface, the pain will increase, and an injury can result. Have the ear examined by a physician, especially if you have recurring reverse blocks.

Equalizing Pressure in the Mask

The air space inside the dive mask is affected by changes in pressure. If you descend without increasing the amount of air inside your mask, pressure pulls your face and eyes into the mask slightly. The pulling sensation, if ignored, ruptures capillaries in your eyes and on your face and causes a mask squeeze. After a mask squeeze, the whites of your eyes will be red, and your face will be red and puffy. There is no excuse for a mask squeeze. You can prevent the problem by exhaling through your nose as needed during descent to keep the pressure inside your mask equal to the surrounding pressure.

Breathing Adjustments

Breathing while in water differs from breathing while on land in several ways. You cannot breathe as freely in water as you can on land. Pressure on the lungs from immersion in water prevents you from breathing as fully as you can in air. Also, it is not as easy to get large quantities of air from a scuba system as it is to breathe deeply above water.

When you ascend, compressed air in your lungs expands, which can cause your lungs to rupture unless you allow the excess air to escape. A normal breathing pattern allows expanding air to escape; breath holding does not. A primary rule of scuba diving is to always breathe continuously.

There is a chance that you will inhale some water. This may cause you to choke and gasp when the water strikes your vocal cords. You need to learn the correct methods of breathing while in and under the water.

Lung Overexpansion

The most important aquatic breathing adjustment is overcoming the instinct to hold your breath underwater. When you breathe compressed air at depth, the density of the air in your lungs is greater than it is at the surface. When you ascend, the air in your lungs expands as the surrounding pressure decreases. If you hold your breath, your lungs expand until they reach their maximum volume, then rupture with an ascent of as little as 4 feet (1.2 m). You must avoid breath holding when breathing compressed air underwater. You should breathe continuously or exhale a small amount of air continuously, and you should never hold your breath during ascent after breathing compressed air. Figure 3.5 illustrates why you must breathe continuously, and the section following the figure describes the possible consequences of breath holding while ascending with compressed air in your lungs.

Skip Breathing

Some divers attempt to extend their air supplies by holding each breath for several seconds. This dangerous practice is called **skip breathing**. When you hold your breath, you increase the amount of carbon dioxide in your circulatory system, which increases your urge to breathe and reduces your breath-holding ability. A high level of carbon dioxide in your body will reduce your ability to cope with difficulties that may arise. And if you skip breathe, you may forget to exhale while ascending, or you may have a headache after the dive. You must be sure to breathe continuously when breathing compressed air.

Breathing Problems

Divers can encounter respiratory difficulties if they overexert themselves, breathe contaminated air, inhale water, or run out of air. Fortunately, you can prevent all of these problems.

Overexertion

Scuba equipment allows you to breathe comfortably while underwater. A regulator delivers air with little respiratory effort, but breathing underwater requires more effort than breathing above water. If you do not maintain your scuba equipment properly, the effort required to inhale and exhale can be excessive and can cause respiratory distress.

Scuba regulators have a limited capacity to supply air. Regulators are not designed for activities involving heavy exertion. Commercial divers use helmets with air hoses to supply large amounts of air to meet their requirements while working underwater. You must avoid strenuous activities underwater because

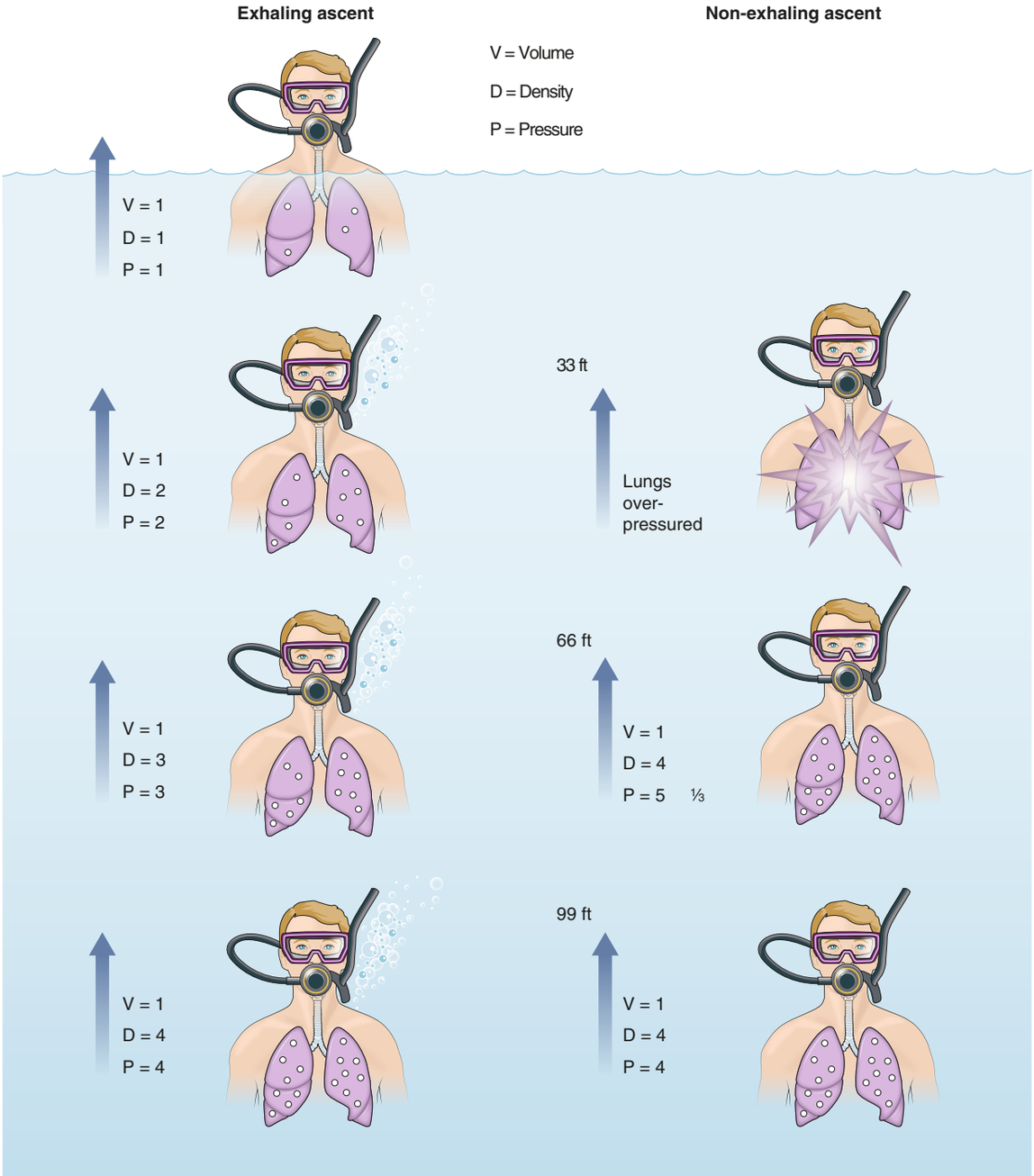

Figure 3.5 A breath-holding ascent after breathing compressed air can cause lung rupture. Breathe continuously while ascending.

you can overbreathe your equipment and experience air starvation, a suffocating feeling of not being able to get enough air.

While diving, you should make your breaths longer and slower than your breaths on land. Pace your activity to keep respiration at a slow, controlled rate. If breathing becomes rapid or labored, cease all activity immediately and focus on exhaling deeply until your respiration returns to a controlled rate.

Contaminated Air

If the compressed air in your tank comes from an air compressor that is not operated or maintained properly, your tank can contain contaminated air. The contamination is likely to be carbon monoxide (CO), a gas produced by incomplete combustion. CO in the body impedes the blood's ability to transport oxygen. Blood hemoglobin's affinity for CO is 210 times greater than its affinity for oxygen. Hemoglobin normally exchanges oxygen for CO_2 about every half minute, but when CO attaches to hemoglobin, the contaminant remains bonded to the hemoglobin for hours. Breathing contaminated air under pressure can make you ill and render you unconscious. Avoid contaminated air by having your tank filled only with pure air. CO is an odorless, tasteless gas, but it is usually accompanied by other gases that have a foul taste and odor. If the air from a scuba tank smells or tastes foul, do not use it. Report the situation immediately to the facility that filled the cylinder. Good air stations have their air tested regularly to ensure purity.

Water Inspiration

If you inhale water, your larynx spasms as a reflex action to keep the liquid out of your lungs. Avoid coughing and choking in water; not only are these reflexes unpleasant, but they can cause you to inspire more water and lose buoyancy. You need to breathe differently underwater than you do on land. Avoid quick inhalations. Begin a breath with a light, slow inhalation to ensure that you are inhaling air and not water. Once air begins flowing, the remainder of your inhalation can be normal.

You can also prevent droplets of water from going down your throat by placing the tip of your tongue on the roof of your mouth behind your upper teeth to act as a splashboard. If, in spite of all you do to avoid it, you inhale some water and start coughing, try swallowing hard three times in rapid succession. The swallowing helps you overcome your coughing reflex.

If water does enter your lungs, it interferes with respiration. If an extensive area of the lungs is irritated by water, you can drown. This is a risk you face in any aquatic pastime. The purpose of your training is to allow you to enjoy the aquatic environment with the minimum risk of injury, especially of drowning. When you follow the rules and practices that you are taught, the risk of injury while diving is negligible.

An automatic response to the inhalation of water is to swallow it in an effort to keep your airways dry. When you swallow saltwater, you can experience illness, nausea, or diarrhea. If you swallow several mouthfuls of saltwater while diving, you should terminate the dive.

Running Out of Air

Some divers do not monitor their air supplies and run out of air at depth. Running out of air underwater is no more excusable than running out of gas on the freeway when you have a working gauge. There are ways to manage an out-of-air situation (which you will learn in chapter 6), but it is much better to avoid the problem by monitoring your instruments and keeping them in working order.

Potential Lung Injuries

Pulmonary Barotrauma

Barotrauma is trauma or injury caused by pressure. **Pulmonary barotrauma** is any lung injury caused by pressure. Failure to allow expanding air to escape from the lungs during ascent can cause several forms of pulmonary barotrauma, either singularly or in combination.

Arterial Gas Embolism

An **embolism** is a blockage of circulation. An embolism resulting from an air bubble blocking the arterial circulation is an **arterial gas embolism** (AGE). This occurs when air expanding in the lungs forces bubbles of air into the circulation. Air bubbles enter the capillary beds of the lungs and pass through the heart, which pumps the bubbles into arteries supplying blood to the body. The diameter of an artery decreases as the distance from the heart increases. At some point, a bubble lodges in an artery and becomes an embolus (plug). It is common for an arterial gas embolism to occur in an artery leading to the brain. The embolus has the effect of a stroke, causes unconsciousness, and is an extremely serious injury. Anytime a diver loses consciousness after a dive, you should suspect AGE. The temporary obstruction of an airway, such as that caused by a cold, increases the risk of AGE. Healthy lungs are a prerequisite for diving.

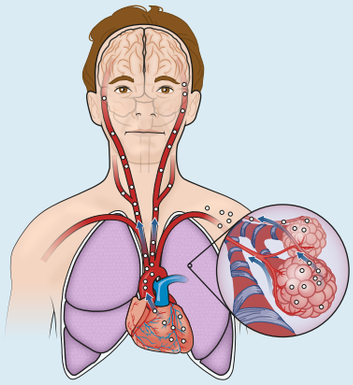

Air embolism

Mediastinal Emphysema

If a rupture of the lung does not force air into the circulation, the air may travel along the bronchi and enter the middle area of the chest (called the mediastinum). This results in a **mediastinal emphysema**, which means that air is in the tissues in the middle of the chest. The injury causes a dull ache or tightness that worsens with coughing, swallowing, or taking a deep breath. Expanding air can interfere with the circulation of the heart and cause loss of consciousness.

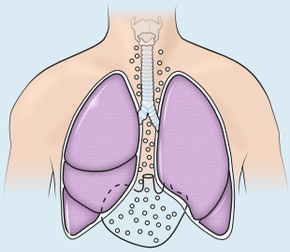

Mediastinal emphysema

Heart and Lungs

Immersion changes your cardiac workload. When you are upright in water, there is greater pressure on your lower limbs than on the upper part of your body. The pressure differential shifts more blood than normal into the upper part of your body. Your heart pumps a greater volume with each beat and works from 1.3 to over 1.5 times harder in this situation than it does on land.

Water temperature alters your heart's rate and rhythm. In addition, physical and emotional stress make your heart work harder. The combined effects of water pressure, exercise, cold, and stress can cause serious problems if your heart is not healthy. If you have heart problems and wish to dive, you should first obtain medical approval from a diving physician. Even minor heart problems can cause you to suddenly lose consciousness in the water, and you could drown. Fitness for diving is an important safety issue.

Subcutaneous Emphysema

Expanding air in the mediastinum can migrate upward along the breastbone. The air then swells the tissues around the neck, producing an injury known as **subcutaneous emphysema**, which means that air is in the tissues under the skin. The injury can cause changes in the voice, crackling of the skin, and a feeling of fullness in the neck.

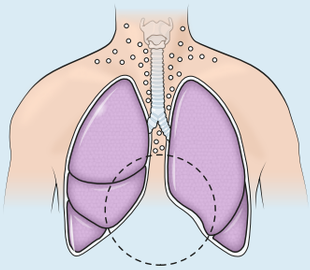

Subcutaneous emphysema

Pneumothorax

If a lung rupture forces air into the space between the lungs and the lining of the chest wall, a condition known as **pneumothorax** occurs. This means that air is trapped in the chest cavity. As the air trapped in the pleural space expands during ascent, it collapses the lung and can affect the action of the heart. Symptoms include severe pain and breathing difficulty.

Lung injuries are serious and can be life threatening. Life support may be required. This is one reason that you should complete a course in cardiopulmonary (heart and lung) resuscitation (CPR). You may have to administer first aid until professional medical treatment (which all lung injuries require) is available.

Failing to allow excess air to escape causes nearly all lung overexpansion injuries. You can prevent injuries by breathing continuously. If you remove the scuba regulator from your mouth for any reason, exhale lightly and continuously to avoid breath holding.

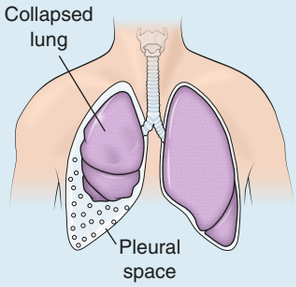

Pneumothorax

Cold water increases the effort required by your heart because cold causes the body to reduce circulation to the extremities in order to conserve heat. With more blood in the core of the body and an increased workload from pressure, your heart may not be able to pump the increased blood volume. The result can be excess blood in the small blood vessels of the lungs. These small blood vessels will transfer fluid to the air spaces of the lungs to reduce pressure. The accumulation of fluid in the lungs—known as pulmonary edema—causes severe breathing distress and can lead to panic or even loss of consciousness. You should test your circulatory ability gradually. Avoid strenuous dives in cold water until you have gradually adjusted to being underwater.

Your ability to exert effort in water is not as good as it is on land. When under pressure, your heart and lungs do not function as well as they normally do. A situation in which you are trying to work hard underwater can quickly result in a frightening sensation of suffocation when your equipment is unable to deliver

the amount of air that your body requires. To stay within your cardiopulmonary limitations, you must learn to limit your activities and to pace yourself.

Equilibrium Adjustments and Seasickness

If you rupture an eardrum, water that is colder than body temperature can come into contact with the semicircular canals of your inner ear. The canals provide information for your brain to maintain equilibrium and are sensitive to temperature and pressure changes. Cold water cooling the semicircular canals can cause vertigo, a subjective feeling of movement perceived as a spinning sensation. This disorientation will pass when the water in the middle ear warms to body temperature. Obviously, it is better to prevent vertigo than to deal with it.

When a sudden change in pressure occurs in the air space of the middle ear of one ear (a change that affects the semicircular canals), this can also produce a disorienting sensation called alternobaric vertigo. The disorientation from alternobaric vertigo passes quickly when the brain recovers from the sudden pressure change.

Vertigo has many causes. Fortunately, instances of severe disorientation are rare in diving. If you experience disorientation while underwater, try to grasp a solid object for a point of reference until the feeling passes. If suspended in the water, close your eyes and hug yourself to reduce the effects of the vertigo. Avoid panic by telling yourself that the sensation will not last long.

Motions detected in your inner ears, visual references, and pressures on the joints of your limbs all affect your equilibrium. When your brain receives mixed signals from your inner ears, eyes, and body, you can experience motion sickness. You must avoid seasickness—motion sickness experienced aboard a vessel—because vomiting in or under the water is hazardous.

Medications can help reduce the tendency to be seasick. The medicine dulls the senses of the organs of balance in your ears. Unfortunately, the medications can have other undesirable effects. If you are prone to seasickness, consult a diving physician about the type of medication you should try. Take some of the medicine several days before you plan to dive, and note the effects, if any. If the medication produces drowsiness or blurred vision, do not use it while diving. Seek an alternative that does not produce side effects. Medications that cause dizziness, drowsiness, changes in heart rhythm, or blurred vision can cause you to lose consciousness under pressure. Many divers do successfully use medications to prevent seasickness; you need to find a type that works for you. Take medication for motion sickness at least 30 minutes before you are exposed to motion.

If you do not want to use medication, you may use other techniques to reduce the likelihood of seasickness. Eat a good, non-spicy meal before diving. An empty, acid-filled stomach becomes upset more easily than a full one. People whose breakfast before diving consists of coffee and orange juice are good candidates for motion sickness. When aboard a vessel, you should position yourself as near the center of the boat as possible. Avoid the front end of the boat. You should also avoid breathing engine fumes and avoid reading. You can reduce your reaction to the boat's motion by lying down for a while with your eyes closed. Being still allows your inner ears to adjust to the motion without visual signals confusing your brain.

You can become disoriented when you are weightless in a dimly lit environment. Under some conditions, it can be difficult to tell which way is up if you rely only on your sense of balance. To prevent disorientation, you must learn to recognize clues about your orientation in the water. Water in your mask settles to the lowest point, bubbles ascend, and heavy objects you hold (such as your weight belt) orient themselves up and down.

Visual Adjustments

Experience will help you adjust to magnified vision underwater. You can adjust so quickly to distance corrections in water that you have to readjust when you surface from a dive. At the end of a dive, the distance to a boat or to the shore can look much greater than it is. You may be surprised to find that you require less time than you think to swim to a destination.

Your vision also compensates for color differences. If you know an object's true color, the object will look more like that color. The use of artificial light at close range makes it easier to view the rich and magnificent colors in the underwater world.

Your vision adjusts to low light levels, but the process takes time. Short, deep dives in turbid water do not allow complete visual adjustment, and details will not be clear. You cannot see well without artificial light. You can improve your ability to see while diving by avoiding bright light and glare before a dive. Wear good, dark sunglasses above water during the day. When you complete training for night diving, you will learn other techniques to help your eyes adjust for diving at night.

Humidity can cause condensation to form on the lens of your mask and obscure your vision. A clean glass surface will not fog, so make sure you clean the lens or lenses of your mask thoroughly before diving.

Ingassing and Outgassing

Your body absorbs nitrogen underwater, and there are limits to the amount of nitrogen that you can safely eliminate at the end of a dive. If you exceed these limits, you can be injured. In this section, you will learn the theory of decompression, which will help you understand how dive tables and dive computers function to help you avoid injuries.

Decompression Theory

Gases diffuse by moving from areas of greater concentration to areas of lesser concentration. When external pressure increases, gases diffuse from your lungs into your blood, and then from your blood into your tissues. When the **ambient (surrounding) pressure** decreases, diffusion occurs in the reverse sequence.

Two factors that affect diffusion of gases in your body are time and **perfusion** (the circulation in a tissue). The greater the circulation in a tissue, the sooner the pressures of the gases in that tissue come into balance with the pressures of the gases you breathe. Reaching this state of equilibrium takes time. The amount of time it takes a tissue to accumulate half of the gas it can hold at a

given pressure is a **half-time**. A tissue is saturated (holds all the gas it can at a given pressure) after six half-times. If perfusion permits 50 percent of a gas to diffuse into a tissue in 5 minutes, the tissue saturates in 30 minutes. Outgassing also occurs in six half-times.

Air is primarily nitrogen and oxygen. The oxygen in the air you breathe is of no concern during decompression within the limits of recreational diving (130 feet or 39.6 m maximum) because you use the oxygen. Nitrogen, the primary component of air, is inert. Your body cannot use the nitrogen, so when you ascend, you must eliminate the excess nitrogen that you absorbed at depth. Because nitrogen diffuses from your body, there is no problem unless the reduction in pressure is so great that you cannot eliminate the nitrogen fast enough. When you reduce the pressure on a liquid rapidly and there is sufficient gas dissolved in the liquid, the gas forms bubbles. An excellent example of this is CO2 dissolved in carbonated beverages (see figure 3.6). The gas remains dissolved in a sealed, pressurized container. When you reduce the pressure suddenly by opening the container, the gas forms bubbles because it cannot diffuse out of solution slowly. If a beverage container has a tiny leak, however, the carbon dioxide comes out of solution slowly without bubbling. There is no bubbling when you open a container that has a slow leak because the gas has already diffused out of solution.

Figure 3.6 Rapid pressure reduction causes dissolved gases to form bubbles within the decompressed liquid.

You must be concerned about the amount of nitrogen in solution in your body and the rate at which you eliminate it. If you absorb too much nitrogen while diving and do not ascend in a manner that allows the excess nitrogen to be eliminated without forming bubbles, the bubbles are likely to cause decompression sickness, a serious diving illness.

Mathematical models provide estimates of the amount of nitrogen in different parts of your body. Because circulation varies in different parts of your body, tissues absorb nitrogen at differing rates. Decompression experts use mathematical models, called **compartments**, to estimate gas absorption and elimination by various areas of the body. A compartment is identified by its half-time. A 5-minute compartment has a half-time of 5 minutes. Experts use compartments ranging from 5 minutes to as long as 960 minutes when calculating gas absorption and elimination.

A compartment that has absorbed gas can withstand some lowering of pressure before bubbling occurs. Originally, scientists believed that a reduction in pressure by a ratio greater than two to one would cause bubbling to occur in divers whose tissues were saturated at a particular depth. This ratio was considered a surfacing limit that could not be exceeded. Then scientists discovered that the surfacing ratios are different for various compartments because of differences in perfusion. For example, brain tissue absorbs and eliminates inert gas faster than

bone tissue. Body areas that eliminate gas quickly have a higher surfacing ratio than compartments that eliminate gas slowly. The difference in ratios poses an interesting and complex problem for divers because one tissue controls how long a diver can remain at one depth, while another tissue controls how long the diver can remain at another depth. The **controlling compartment** was used to establish our modern time limits for diving. Do not remain at any depth longer than the time it takes the controlling compartment to exceed its surfacing ratio. If you do, you must prevent bubble formation by stopping during the ascent to allow time to eliminate the excess gas. The time limits for various depths have been conveniently arranged into tables and included in dive computer software. These tools help divers keep the amount of nitrogen in solution in the various tissues of the body within the surfacing ratio of each compartment.

It takes time to eliminate excess gas from tissues. The rate at which you reduce pressure (ascend) is important. The dive tables presented in chapter 7 use an ascent rate, or pressure reduction rate, no greater than 30 feet (9 m) per minute. The mathematical models for some dive computers use rates two or three times slower than the dive tables. You must ascend slowly to prevent the formation of bubbles in your body tissues.

Partial ascents with stops before continuing the ascent can help with outgassing and the prevention of bubble formation. These delays are called safety stops or rest stops, and they are recommended when ascending from any depth in excess of 30 feet. The deeper you dive, the more safety stops you should take when ascending. The recommended practice is to stop for five minutes at half of the deepest depth to which you have been diving and to stop again at a depth of 15 to 20 feet (4.5 to 6.1 m) for five minutes before surfacing.

Decompression Illness

Commonly known as the bends, decompression illness (DCI) is the result of a reduction in pressure (decompression) that is too rapid for the amount of gas in solution in body tissues. The gas forms bubbles in the tissues or in the blood before it can diffuse into the lungs and be eliminated. Scientists do not fully understand what occurs when DCI strikes. The symptoms may appear immediately after surfacing from a dive or days after diving. About half of all DCI cases occur within one hour after diving. The symptoms vary depending on the amount and location of the bubbles. Severe cases of decompression illness can have severe neurological effects and produce permanent paralysis. Common symptoms include the following:

- A mottled skin rash
- Joint pain
- Numbness
- Tingling
- Weakness
- Paralysis

The following factors can increase the chances of DCI:

- Lack of sleep
- Tight-fitting apparel

- Alcohol and its aftereffects
- Exercise during and after diving
- Dehydration
- Cramped position
- Illness
- Altitude after diving
- Age
- Scar tissue or injuries
- Cold water
- Medical conditions

Some experts believe that you should avoid postdive activities that stimulate circulation because increased perfusion may contribute to DCI. Specific activities to avoid include physical exercise, drinking alcoholic beverages, and taking hot showers or baths. Reduced pressure at altitude, such as flying in a plane or driving into the mountains too soon after diving, also poses an additional risk factor that can cause DCI. Avoid all activities that increase the likelihood of decompression illness.

A person suffering from DCI requires prompt first aid and medical treatment because it is an extremely serious condition. The illness worsens with time. The best first aid measure is to administer oxygen in the highest concentration possible. Breathing oxygen eliminates nitrogen in inspired air and enhances the diffusion of nitrogen from the body. The patient should remain still and should sip water. Make arrangements to have the patient transported to the nearest medical facility. After initial medical treatment, the medical staff will arrange to transfer the patient to a facility that has a hyperbaric (high-pressure) chamber. The patient is placed inside a large vessel called a recompression chamber (see figure 3.7). Hyperbaric treatment consists of increasing the pressure in the chamber to reduce or eliminate the symptoms, administering medications, and then slowly decompressing the patient. Recompression must be done in a chamber.

Figure 3.7 Recompression chambers are the only safe method to restore pressure after becoming sick.

Decompression illness is extremely serious because it can cause permanent injury. DCI can cause joint pain and neurological deficits that cannot be cured. To reduce the likelihood of DCI, you should remain well within the established time and depth limits for diving. Surface at a rate no faster than that specified by the dive computer you use. Stop at least twice during ascent (as previously recommended) and eliminate excess nitrogen before surfacing (see figure 3.8). Ascending at the proper rate and stopping are forms of decompression that reduce the likelihood of DCI. (The details about planning your ascent and time delays are provided in chapter 7.) An additional precaution that can help you avoid DCI is to delay excursions to altitude after diving.

Another point to keep in mind is that it is often difficult to distinguish between the symptoms of an arterial gas embolism and DCI. Any diver who has neurological symptoms after a dive should seek medical attention immediately.

Figure 3.8 Divers completing a decompression stop on their ascent to the surface.

Nitrogen Narcosis

The increased partial pressure of nitrogen can cause a condition known as nitrogen narcosis, or rapture of the deep, at depths of about 100 feet (30 m) and deeper. Scientists do not know the exact mechanism of nitrogen narcosis, but its effects are similar to those of anesthetic gases. The feelings associated with narcosis range from euphoria to overconfidence to terror, although some divers do not experience any unusual feelings. Narcosis affects you whether or not you experience abnormal feelings. And it impairs thinking, judgment, reasoning, memory, and the ability to do physical tasks. Narcosis is hazardous; it reduces your awareness and your ability to respond to an emergency. Susceptibility to narcosis varies from person to person and within an individual from day to day.

Narcosis begins suddenly at depth—usually around 100 feet (30m). You can relieve its symptoms rapidly by ascending to a shallower depth to reduce the narcotic effect.

The following factors predispose a person to narcosis:

- A high level of systemic CO2 caused by exertion
- Alcohol or its aftereffects
- Anxiety
- Cold
- Medications
- Social drugs

Experience, frequent diving, and concentration will reduce susceptibility to narcosis. Preventing narcosis is always better than attempting to cope with it.

Specialty training can qualify you to dive with special gas mixtures that reduce the effects of nitrogen. Diving with mixed gases without completing the required training is dangerous and should never be attempted.

Dehydration Prevention

You need to preserve body fluids to prevent **dehydration**. Getting cold makes you produce more urine than normal. Each breath you take from your scuba cylinder must be humidified as explained in the previous chapter. Breathing dehumidified air causes dehydration. You receive air underwater by creating an inhalation pressure that opens valves in your air delivery system. The inhalation pressure is slight, but it is greater than normal. Inhaling harder than normal is negative-pressure breathing, which also has the physiological effect of increasing urine production. Breathing underwater compounds the problem of dehydration.

Diuretic beverages (such as coffee and alcohol) and some types of medications cause increased urine production. Avoid ingesting anything that makes you urinate more than normal. You must prevent excessive dehydration because the condition predisposes you to diving injuries. To prevent dehydration, you should do the following:

- Insulate yourself to stay as warm as possible.
- Keep your regulator well maintained so you can breathe as easily as possible.
- Avoid diuretic drinks and medications.

- Replenish body fluids frequently; drink fluids before and between dives.
  Note: If you wait until you are thirsty to drink when diving, you are already dehydrated. Be proactive and drink at least 1 pint (500 ml) of appropriate fluid after every dive. A good after-dive beverage is a sports drink diluted with an equal amount of water. Avoid alcoholic beverages because they increase dehydration.

Mobility Adjustments

The equipment that you wear for diving limits your mobility. It reduces your range of motion and makes it difficult to walk. The colder the water, the thicker your exposure suit, so the less your range of movement.

Diving equipment is fairly heavy, and it is challenging to lift and move. Improper lifting techniques can cause back injuries. Squat down and lift with your legs to pick up a tank or weight belt instead of bending over. The weight of the equipment changes your center of gravity and affects your balance. Keeping your balance during entries and exits can be challenging on a rocking boat or an uneven surface for your feet. Move carefully and hold on to something or someone for support when you move around out of the water.

You wear fins so that you can use the large muscles of your legs for propulsion. When diving, you need to learn to use your legs for swimming and your body angle to control direction. These techniques free your hands for other uses underwater. Minimize the use of your hands for propulsion. Fins make it difficult for you to walk when you're out of the water. Shuffle your feet while moving backward or sideways, keep your knees bent, and be careful not to fall.

Drag slows your movement in the water. One factor affecting drag is the speed of motion. The greater the speed, the greater the resistance to movement. The average diver can sustain a speed of a little more than 1 mile an hour (0.9 knot, or 1.6 km per hour). A knot equals 1.15 miles per hour (1.85 km per hour) and is the method of measuring speed in water. Doubling the speed increases the energy requirement fourfold. Trying to move in water as if it were air causes exhaustion quickly. Use a slow, steady pace and slow, deliberate motions to reduce the effects of drag.

Another factor affecting drag is the size of the object in motion. The larger the surface area exposed, the greater the resistance. A swimming diver is under the influence of four forces: Weight (negative buoyancy) pulls the diver downward, lift (positive buoyancy) pulls the diver upward, thrust moves the diver forward, and drag slows forward progress (see figure 3.9). Divers usually wear weight around the waist, which pulls the lower half of the body downward. Buoyancy from the BC lifts the upper half of the body. The effect of these forces increases the surface area of the diver and therefore the drag. You should adjust the amount of weight you wear so that your body is as horizontal as possible in the water. Correct weighting minimizes drag and the effort required to swim. Also consider placing a small amount of weight near the top of your body to adjust your trim—that is, your position in the water.

Water flows smoothly across a smooth and rounded surface but flows turbulently across an irregular surface. When the surface is irregular, the turbulent flow increases drag. Just as vehicles designed to travel through fluids are streamlined to reduce the drag caused by turbulent flow, you can choose and configure your equipment so that it presents the smoothest surface possible to the flow of water.

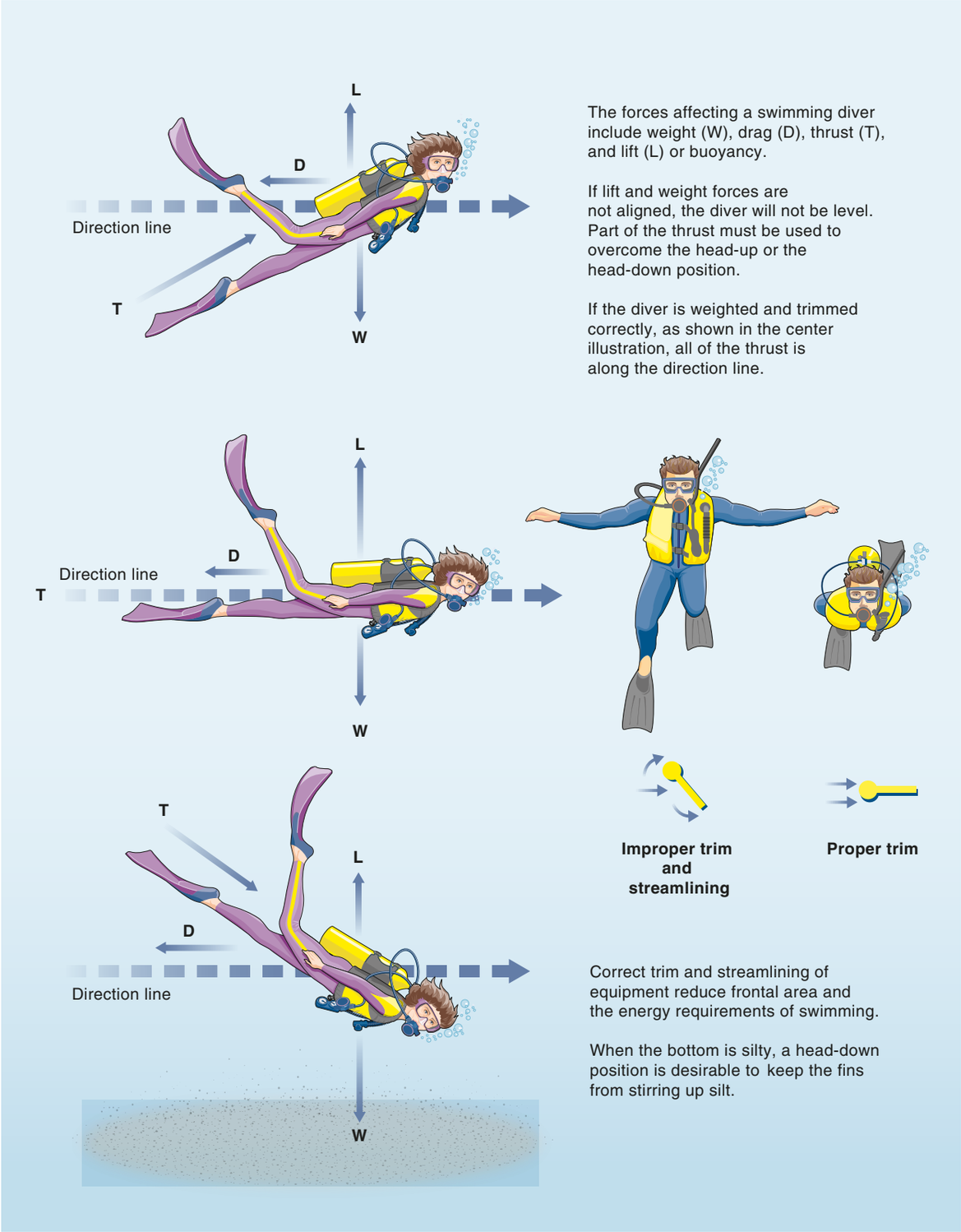

Figure 3.9 Weight, drag, thrust, and lift (buoyancy) affect divers as they move through the water.

Summary

You have to make many changes in your normal behavior to adjust to the underwater environment. The way you breathe is the most important adaptation. You must breathe continuously and avoid breath holding when you use scuba equipment. You must also limit your activity and pace yourself to avoid overexertion. Managing the mechanical and physiological effects of pressure requires major adjustments. You must keep pressure equalized in your air spaces to avoid pressure injuries, and you must limit your depth and time at depth to avoid nitrogen narcosis and decompression illness. Being in the underwater environment seems strange at first, but you can conform to weightlessness and other strange feelings. The sensations of diving become exhilarating as you gain experience in the new world beneath the surface.

Application-of-Knowledge (AOK) Questions

Remember that this is an exercise in thinking, not a quiz. If you carefully consider each situation and how you would respond to it, your time will be well invested even if your thoughts are different from the solution provided. Imagine each situation as vividly as possible.

1. You are wearing a wetsuit and swimming against a current (which is a good idea under most circumstances) in cold water. You are becoming colder with each passing minute. What action, other than ending the dive, can you take to reduce your heat loss?
2. After a dive during which you shivered, you exit the water on a cool, cloudy, windy day. What actions can you take to get warm that are not likely to cause physiological problems? Hint: Rapid rewarming can affect your heart rhythm and may also cause decompression illness.
3. You are wearing 17 pounds of lead weights while diving at a depth of 40 feet (12 m) when your weight belt suddenly slips off. How rapidly do you think you will float upward, and what actions can you take to recover from this situation?
4. You are wearing a wetsuit, a BC, and a weight belt while scuba diving at a depth of 33 feet (10 m). Your buoyancy is adjusted to be perfectly neutral at this depth. You have to ascend 10 feet (3.3 m) to clear a reef. What actions will you need to take to control your buoyancy as you swim over the reef?
5. You are pulling yourself hand over hand down an anchor line to descend to a dive site and are having difficulty equalizing the pressure in your ears. Your normal equalization methods are not working. What should you do to equalize the pressure in your ears?
6. Why do you need to equalize the pressure in your body air spaces and your mask more frequently at the beginning of the descent than when you are approaching your diving depth?
7. You are having trouble keeping up with your buddy, who is swimming ahead. You are breathing heavily and beginning to feel starved for air. What actions should you take?
8. When you switched tanks between dives, you accidently selected one that is almost empty. You were also distracted and failed to check the pressure before you started the dive. You are at a depth of 50 feet (15 m) when it becomes hard to breathe. You look at your pressure gauge and find a reading of zero. Based on what you know at this point in your training, what action do you think is appropriate?

Large, beautiful angelfish at a coral reef in Cozumel, Mexico

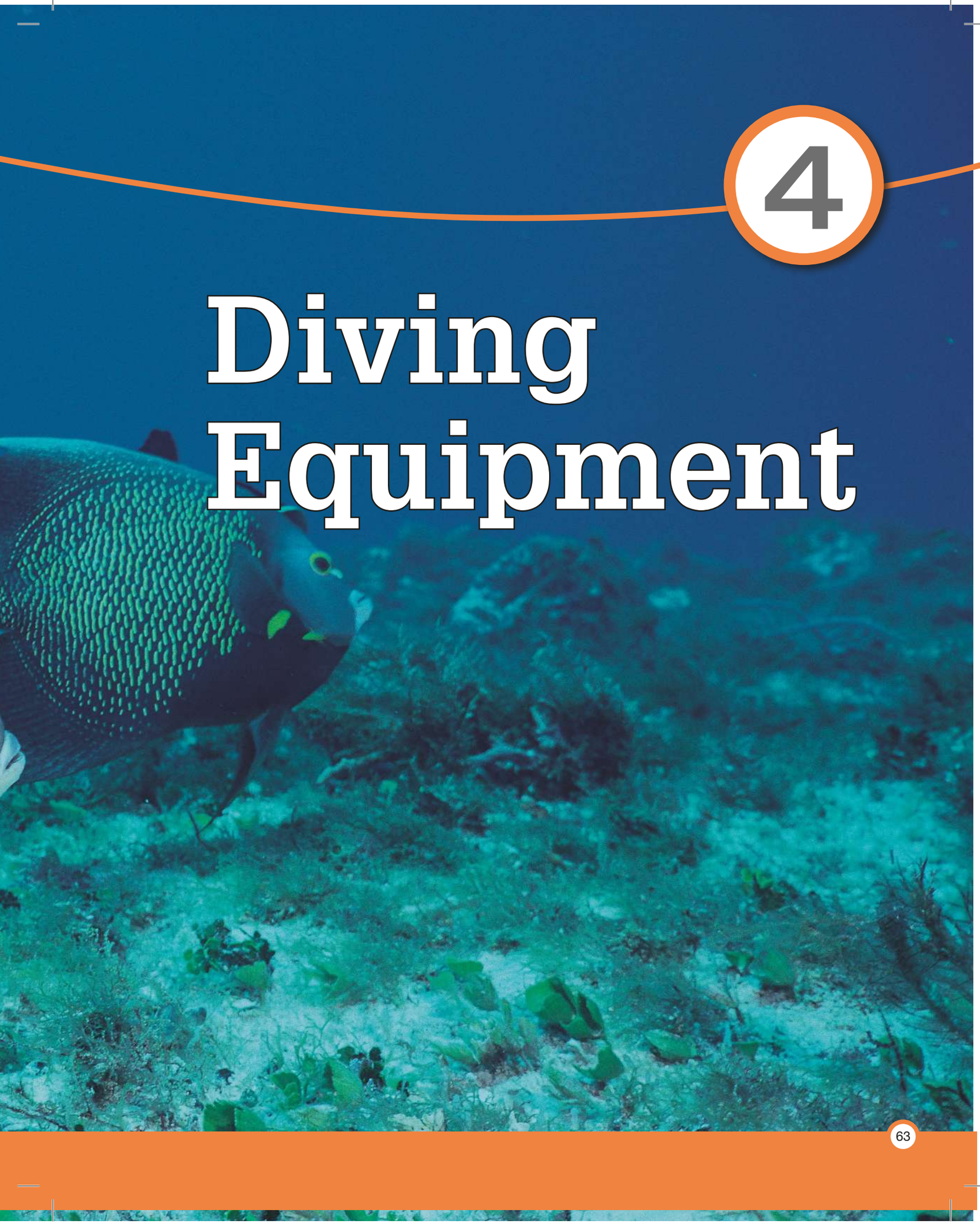

4

Diving Equipment

Dive In and Discover

By the end of this chapter, you will be able to do the following:

- Describe the purpose, types, features, selection criteria, and care and maintenance of skin and scuba diving equipment.
- Contrast steel and aluminum scuba cylinders, wet and dry suits, integrated and nonintegrated weighting systems, and reserve and nonreserve valves.
- Define the terms *buoyancy*, *compensator*, *defogging*, *shorty*, *jumpsuit*, *Farmer Johns*, *K-valve*, *J-valve*, *DIN valve*, *valve seat*, *burst disk*, *O-ring*, *port*, *stage*, *octopus*, *capillary gauge*, *Bourdon tube*, *Boyle's law*, *outgassing*, *bezel*, *submersible pressure gauge (SPG)*, *blowout plug*, *alternate air source (AAS)*, *pony tank*, *Spare Air unit*, *lubber line*, *console*, and *ceiling*.

You will become familiar with the following items in this chapter:

- Masks
- Snorkels
- Fins
- Skin diving vests
- Exposure suits
- Weighting systems
- Buoyancy compensators
- Scuba cylinders
- Cylinder valves
- Scuba regulators
- Alternate air sources
- Instrumentation
- Dive knives and other accessories
- Specialized equipment

You need to be equipped properly to dive in open water. A snorkeler (a diver who remains at the surface) should wear a mask, a snorkel, fins, and a skin diving vest. A skin diver (a breath-holding diver who dives beneath the surface) uses snorkeling equipment and may wear an exposure suit and a weight belt (if the suit requires weights). Scuba equipment includes, at a minimum, a mask, a snorkel, fins, an exposure suit, a weighting system (if needed), a **buoyancy compensator** (BC), a scuba unit (cylinder, valve, regulator, **alternate air source**), instrumentation, and a dive knife. For cold water, a diver also needs a hood, boots, and gloves.

Figure 4.1 shows fully equipped divers.

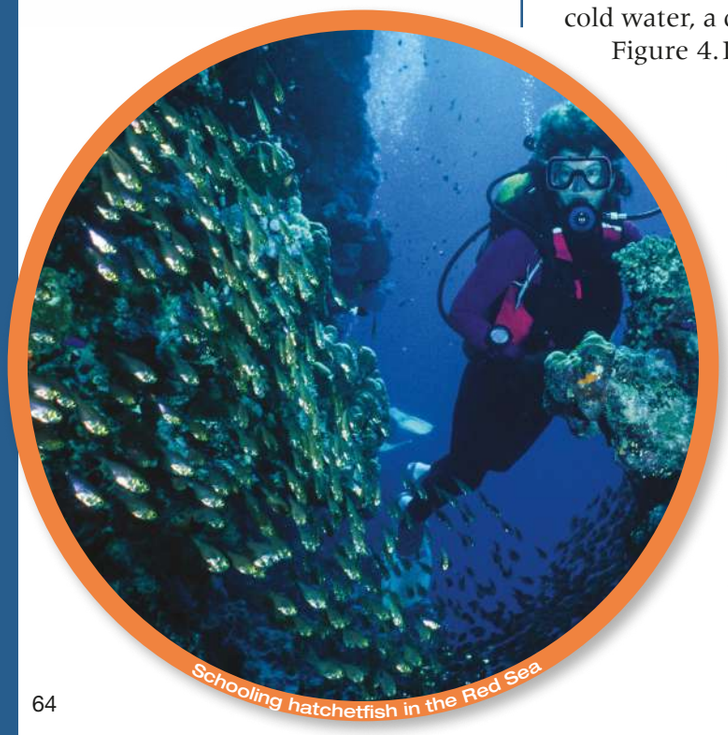

Schooling hatchetfish in the Red Sea

Equipment helps people adapt to the underwater environment. It allows them to see, breathe, move, monitor their position, remain comfortable, have fun, and rest. Diving is an equipment-intensive activity. In this chapter, you will learn what equipment you need, how to select the best equipment for your needs, and how to care for your equipment.

Figure 4.1 Fully equipped cold-water (left) and warm-water (right) divers.

Basic Equipment for All Divers

The mask, snorkel, fins, and some flotation device are basic equipment for all types of recreational diving (snorkeling, skin diving, and scuba diving).

Masks

The eyes require an air space in front of them to focus sharply. Your mask provides an air space and a window to another world. Many styles of masks are available, but only two basic types are used for recreational diving: purge and nonpurge. A purge is a one-way valve through which you can expel water that enters the mask. As you will learn in chapter 6, you can remove water from a mask without a valve, so many masks do not feature a purge valve. A third type of mask, a full-face mask, is for commercial and specialty applications only. Figure 4.2 shows the most common types of masks.

Figure 4.2 Test dive masks before you buy to ensure a snug and comfortable fit.

The type of mask you choose is not nearly as important as the fit of the mask on your face. The mask must fit your facial contours perfectly so it will remain comfortable and watertight throughout a dive. Fit and comfort are the most important features to consider when you select a mask. To test the fit of a mask, remove the strap or position it on the front of the mask. Tilt your head back and lay the mask (do not push it) on your face. Make sure your hair is not under the sealing edge of the mask; then inhale gently. If the mask pulls onto your face snugly from the partial vacuum created by your inhalation, the mask fits. If you have to push the mask to get it to seal on your face, the mask will probably leak when you use it underwater.

Factors that affect the fit and comfort of a mask include the style, the type of material, and the type of sealing edge of the mask. The best masks are made from silicone, which is soft, pliable, and nonallergenic and resists deterioration better than rubber compounds (see figure 4.3). The buoyancy of smaller masks poses no problems, but the tug of buoyancy of a larger-volume mask can affect the seal. Low-volume masks are easier to clear of water and provide excellent visibility. A wide, double-edged seal does a better job of keeping out water than a single-edge seal.

A film of oil from production is on the surface of the glass lenses of new masks. You must remove the film completely—a process called **defogging**—or the mask will fog continuously underwater. Clean your mask thoroughly with scouring powder. The glass (which is tempered) is too hard to be scratched by the abrasive, so do not be timid when cleaning your mask. Commercial defogging solutions help keep your mask clear while you dive. If a clean mask fogs slightly during use, you can allow a small amount of water into the mask and wash it across the fogged area to resolve the problem.

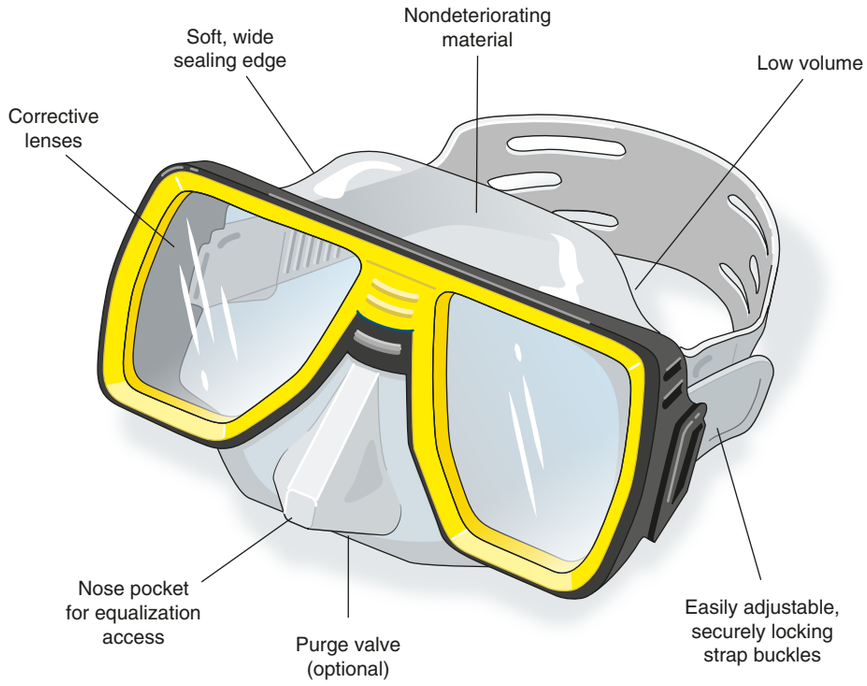

Figure 4.3 Mask features.

Several options are available for visual correction if you need it. Hard contact lenses cannot be used, because gases that are released from the body during ascent can be trapped between the lens and the surface of the eye. This can injure the eye. If you wear soft contact lenses, you can use them for diving after you complete your training; however, you should not use them during training because you may lose them when you are learning how to clear water from your mask. Some contact lens wearers prefer a mask with a small purge valve because they can expel water from the mask with little risk of losing a contact lens. Several companies prepare and bond corrective lenses into any diving mask. Interchangeable corrective lenses are available for some masks. If you require only a simple correction, you may be able to obtain corrective lenses for your mask when you purchase it.

Store your mask in a mask box when you are not using it. The box helps keep the mask from getting damaged and helps prevent discoloration of the silicone mask skirt.

Snorkels

A human head weighs about as much as a bowling ball. If you had to swim while holding a bowling ball out of water, you would become exhausted quickly. If you swim while holding your head above water, you will tire rapidly. But if you allow buoyancy to support your head in the water, you can relax and swim for hours. A snorkel allows you to breathe while water supports your head—allowing you to conserve energy and enjoy continuous underwater viewing.

In its simplest form, a snorkel is nothing more than a breathing tube that extends from a diver's mouth to a point above the waterline. A basic diving snorkel is a J-shaped tube with a mouthpiece on one end. Just as with masks, there are two types of snorkels: purge and nonpurge (see figure 4.4). A purge snorkel has a one-way valve through which you can expel water that enters the tube. As you will learn in chapter 6, you can clear water from a snorkel that lacks a purge valve, so some snorkels do not have a valve. Some types of purge snorkels are self-draining; gravity drains water from the tube automatically when you are at the surface of the water.

Additional snorkel options include a swivel mouthpiece and a flexible hose for the lower half of the tube. With the flexible hose, the lower part of the tube—which is usually curved—hangs straight down when you are not using your snorkel (see figure 4.5). The flex-hose snorkel reduces interference between the snorkel and your scuba regulator. Special mouthpieces can be used to maximize comfort, and devices are available to prevent water from entering the top end of the snorkel, although water exclusion devices are not essential.

Figure 4.4 Three common types of snorkel are purge (two at left and far right) and nonpurge (center right).

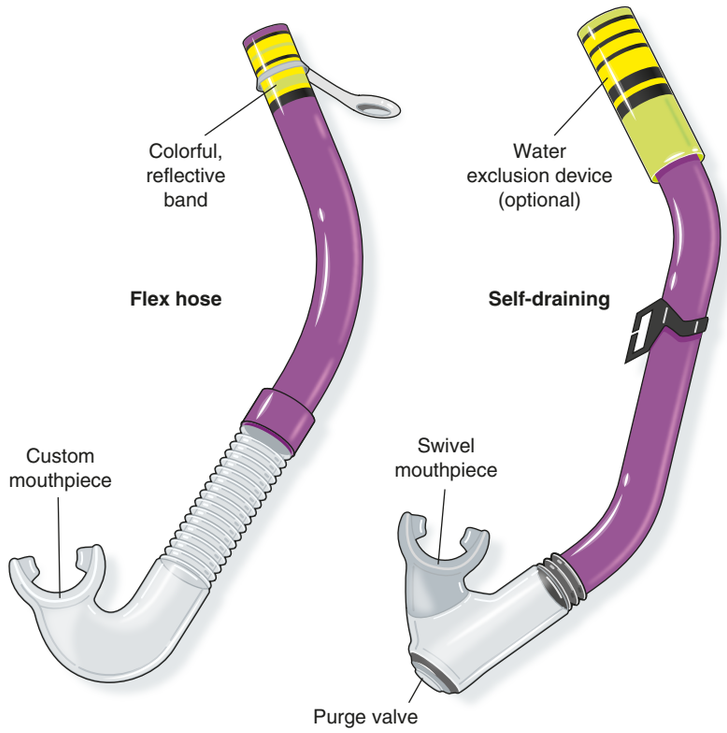

Figure 4.5 Features of a flex hose and self-draining snorkel.

The fit of your snorkel is the most important consideration, much more important than the type. The mouthpiece must not irritate your mouth, gums, or jaw when in place for extended periods of time. The angle of the mouthpiece in your mouth should not require you to bite hard to hold the mouthpiece in place. A snorkel that fits poorly may cause sore gums or jaws. A snorkel tube should have an inside diameter of approximately 3/4 inch (1.9 cm) so that resistance to airflow through the tube does not make breathing difficult.

Attach the snorkel to your mask strap on the left side using a snorkel keeper because your regulator hose goes over your right shoulder. Several types of snorkel keepers are available. Popular keepers are the simple ones depicted in figure 4.6. The adjustment is correct if the snorkel mouthpiece remains in your mouth when you open your mouth wide.

Figure 4.6 Two types of snorkel keepers.

Fins

Diving would be much less enjoyable without fins. Fins increase your ability to move in the water and free your hands for other activities. You move in water by pushing against the water. Fins increase your ability to move by presenting a larger surface area than your hands or feet alone. The muscles of your legs are much larger and stronger than your arm muscles. Your arms would tire quickly if fins were attached to your hands, but your legs are strong enough to handle the load. Fins also help stabilize you in water by providing a large surface to resist movement. The resistance provides leverage for countermovement and directional control.

The two basic types of fins are shoe (full-foot) fins and open-heel fins. Shoe fins slip on to bare feet, so they are good snorkeling fins for tropical climates. You wear open-heel fins with foot coverings called boots. Open-heel fins, generally used for scuba diving, are usually larger and stiffer than shoe fins (see figure 4.7). Small, flexible snorkeling fins may be inadequate for the harder work of scuba diving.

Special materials and designs abound. The fundamental features of a fin are the size and stiffness of the blade (see figure 4.8). The larger and stiffer the blade, the greater the propulsion but also the greater the physical demand when you move the fin through the water. A blade that is too stiff can cause you to cramp and become fatigued. You should begin with fins of a moderate size and stiffness. When you can use those fins for extended periods without difficulty, you may consider fins that can provide greater propulsion.

The most important criteria for the selection of any item of diving equipment are fit and comfort, and this is especially true of fins. To help ensure proper fit, sit down and try on a fin. Wear a boot if you are trying on an open-heel fin. Hold your foot in the air and wiggle it up and down and from side to side. The fin and your foot should move as a single unit. If your foot moves inside the foot pocket, the fin is too large. Diving equipment should fit snugly but not tightly.

Figure 4.7 Three styles of open-heel fins.

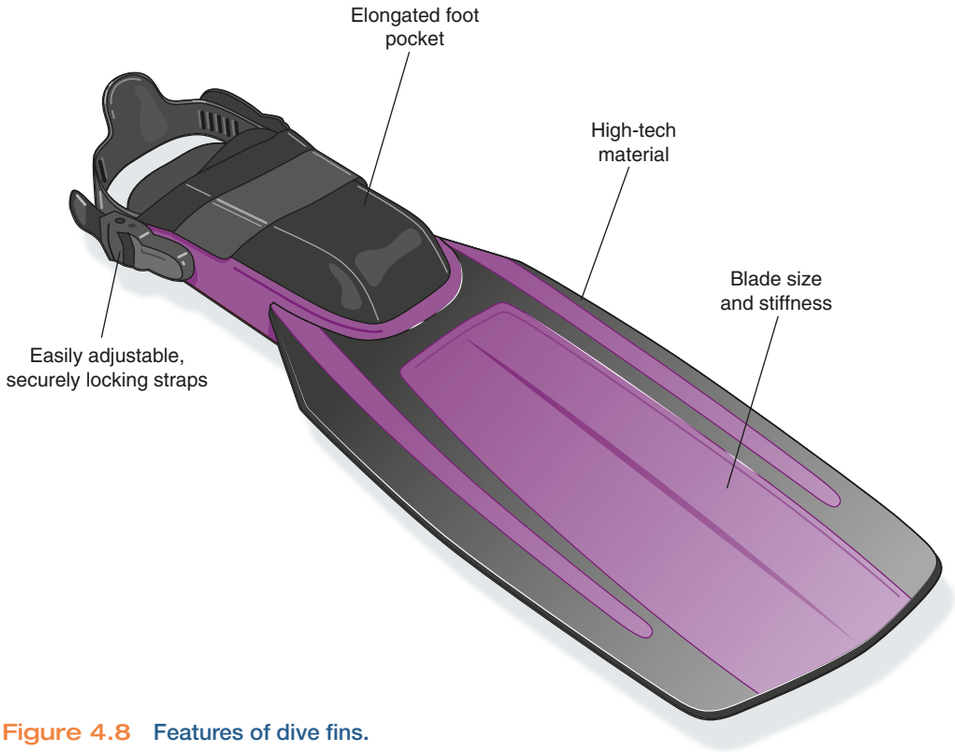

Figure 4.8 Features of dive fins.

A fin that is too tight can cause your foot to cramp. The foot pocket of the fin should fit snugly but should not exert pressure on your foot. Select your fins using fit, comfort, and blade size and stiffness as your primary criteria. Features, fashion, and cost should be secondary considerations.

Skin Diving Vests

When you blow air into a skin diving vest (which is smaller and less complex than a buoyancy compensator), you can achieve positive buoyancy (floating). Positive buoyancy is invaluable when you want to rest at the surface or carry items you have collected while diving. Wear a buoyancy device whenever you dive. You can achieve negative buoyancy (sinking) by venting air from the skin diving vest.

Divers use one basic type of skin diving vest; this vest fits around the neck and secures at the waist (see figure 4.9). A standard feature of a skin diving vest is an oral inflation tube, which comes in various sizes. It is easier to use larger tubes with mouthpieces than smaller tubes without mouthpieces.

Figure 4.9 Skin diving vest.

Care and Maintenance of Skin Diving Equipment

Your skin diving equipment will provide years of service if you take care of it properly. Rinse the equipment with clean, freshwater after use; dry it in the shade; and store it in a cool, dark, smog-free location. Prolonged sunlight, smog, salt crystals, and swimming pool chlorine will harm your equipment. Fill a skin diving vest partially with freshwater after use, and swish the water inside. Drain the vest completely, and then inflate it for drying and for storage. Periodically inspect the straps on your mask and fins for drying and cracking, and replace straps when they begin to deteriorate. After your mask has dried, rub a drop of defogging solution on the lens to help keep the lens clean during storage.

Exposure Suits

You should wear some type of exposure protection when diving regardless of the water temperature. An exposure suit provides protection against scrapes and stings in addition to providing insulation.

The two basic types of exposure suits are wet suits and dry suits. Generally speaking, the colder the water, the thicker your suit should be and the more of your body should be covered by the suit. Suits come in many configurations, but most are one piece or a combination of a jacket and pants. Vests are accessories that help conserve heat in vital organs. Hoods, boots, and gloves help reduce heat loss from the head, feet, and hands, respectively.

Wet Suits

A wet suit is a good exposure suit for water temperatures ranging from 60 to 85 °F (15.6 to 29.4 °C). A wet suit allows water inside, but any water that enters can also exit and carry heat with it. The better the fit of the suit, the less water inside and the warmer the suit. Table 4.1 provides a comparison of the various types of wet suits. The following descriptions provide some additional information.

Table 4.1 Wet Suits

Type	Warmth	Features	Temperature range
Spandex	45% more than bare skin	Light, compact; useful as undergarments for thicker suits	78 °F (25.6 °C) +
Thermoplastic	30% warmer than spandex	Neutrally buoyant; no weights required; wicks perspiration; windproof	75 °F (23.9 °C) +
Plush-lined thermoplastic	10% warmer than unlined	Same as thermoplastic	72 °F (22.2 °C) +
Foam neoprene 1/8 in. (2–3 mm) 3/16 in. (4 mm) 1/4 in. (5–6.5 mm)	20-100% warmer than plush-lined thermoplastic	Buoyant; weights required; long drying time; evaporation chills wearer; minor repairs easy to do	Down to 60 °F (15.6 °C)
Hoods, vests, boots, and gloves or mitts	16-66% more warmth	Reduces water circulation; layering allows flexibility for various temperatures	Down to 60 °F (15.6 °C)

Spandex suits are thin, stretchy, attractive, full-body garments that are popular for diving in tropical waters. Spandex garments provide 45 percent more insulation than bare skin and provide protection against stings, scrapes, and sunshine. Spandex suits are light and compact for travel. They are also useful as undergarments for thicker neoprene wet suits, adding an extra layer of insulation and allowing you to slip into the heavier suit more easily. If you live in a temperate climate, you can use a spandex body suit beneath a neoprene wet suit for local diving and use the spandex suit by itself for vacation diving in the tropics (see figure 4.10).

Thermoplastic—sandwiched between two layers of spandex—is another type of wet suit material. A suit made from this three-ply material is about 30 percent warmer than a spandex suit. Thermoplastic suits are thin (1.2 to 1.4 mm) and designed for use in tropical waters (75 to 85 °F [24 to 29 °C]). Some suits feature soft, plush lining for extra warmth. The plush lining increases the warmth of the suit by about 10 percent. You can wear a thermoplastic suit beneath a neoprene wet suit for extra warmth in colder water.

Thermoplastic material offers several advantages: The balance between the weight and volume of this unique material makes it neutrally buoyant, so you may not need a weight belt. The stretchy fabric allows you to move freely. The material draws perspiration away from your body but is windproof.

A third type of wet suit material is foam neoprene, as in the suits shown in figure 4.11. The foam contains tiny bubbles of inert gas, which provide insulation. The thicker the wet suit material, the greater the insulation quality of the suit. A 1/8-inch (about 3 mm) neoprene wet suit is about 20 percent warmer than a plush thermoplastic suit.

The thickness of foam neoprene in wet suits ranges from 1/8 inch for warm-water diving to 3/8 inch (about 9 mm) for extreme cold-water diving. The most common wet suit thicknesses are 1/8 inch, 3/16 inch (about 4 mm), and 1/4 inch (about 6 mm). You can layer wet suit material on critical areas of your body to

Figure 4.10 Common styles of warm-water wet suits: foam neoprene jumpsuit (left), shorty (center), and nylon (right).

Figure 4.11 Common styles of cold-water wet suits: Jackets at left and center (worn with or without an attached hood) are worn over Farmer Johns at right to form layers.

reduce heat loss, but the thicker the insulation you wear, the more difficult it is for you to control buoyancy. Choose the thickness of wet suit material used by experienced divers in the area where you plan to dive.

Nylon usually covers both sides of the neoprene used for wet suits. The nylon increases the strength and durability of the suit, which is glued and sewed together. You can make minor repairs with wet suit cement, but you should have a wet suit manufacturer do extensive repairs.

Neoprene wet suits require weights to achieve neutral buoyancy. The suit provides immediate buoyancy when you release the weights, but the buoyancy of the suit can be either a benefit or a hazard. Neoprene is not windproof, and the evaporation of water from the suit between dives can chill you. You can wear a wet suit overgarment to retain warmth between dives in colder climates. Neoprene suits take longer to dry than other types of exposure suits. Mobility is good with thin neoprene wet suits but decreases as the thickness of the material increases.

Numerous wet suit designs are available, including the **shorty**, the one-piece **jumpsuit**, and **Farmer Johns** (see figures 4.10 and 4.11). You must consider several features when selecting a wet suit design. The more zippers a suit has, the more water circulates inside the suit and the greater the loss of heat. You can get zipperless suits for cold-water diving. Good wet suits feature a spine pad to minimize water circulation along the channel formed by your spine, and some suits for cold-water diving have attached hoods to minimize water circulation at the neck. You may spend time kneeling around and in the water, so knee pads are a desirable feature.

Heat packs are available for wet suits. These packs contain a nontoxic, reusable chemical that heats to about 130 °F (54 °C) for half an hour or more, depending on conditions. The packs fit into special pockets that are an optional suit feature.

You can buy wet suits in standard sizes, or you can have a suit tailored for a custom fit. The fit of a wet suit is its most important feature. The suit must fit snugly all over, but it must not fit so tightly that it hampers your breathing and circulation. A suit that fits well may feel slightly restrictive out of the water. The true test of the fit of a suit is to dive with it. You may be able to rent a wet suit identical to one you would like to purchase.

Wet Suit Accessories

Scuba divers wear footwear for warmth and foot protection. Several types of footwear—called boots or booties—are available for various needs. Boots, usually made from neoprene, may cover only the foot or both the foot and the ankle (see figure 4.12). Boots range from inexpensive neoprene socks to sturdy footwear with durable, molded soles. They may or may not have zippers. Zippered boots are easier to don and remove, but boots without zippers are warmer.

Figure 4.12 Boots are available with or without zippers.

You should wear hand coverings when the water temperature is less than 70 °F (21 °C). Some divers wear gloves for protection, such as when catching lobsters. Types of hand coverings include gloves, mitts, and gauntlets (see figure

4.13). Wear mitts in cold water because they are thicker and have less surface area for heat dissipation than gloves have. Wear gauntlets (neoprene mitts with long cuffs) when the water is extremely cold. Thin neoprene gloves provide sufficient insulation in temperate water. Do not wear gloves in tropical areas. You are more likely to touch things when wearing gloves, and grabbing delicate coral reefs and marine animals while wearing gloves may harm the animals.

Figure 4.13 Gloves and mitts protect divers' hands from cold temperatures and injuries.

A hood is an important warmth accessory that can reduce your heat loss from 20 to 50 percent, depending on the temperature of the water. Two basic types of hoods—attached hoods and separate hoods—are available for wet suits (see figure 4.14). Cold-water divers like attached hoods because they restrict water circulation in a suit more than separate hoods do. Some separate hoods have skirts that end at the base of the neck; other cold-water hoods have large bibs that cover the neck and shoulder area. Thin hoods are used for protection in warm water, and thicker hoods are used for insulation and protection in cold water.

Figure 4.14 Hoods conserve warmth in cold waters.

Figure 4.15 The addition of a vest can increase the warmth of your wet suit by as much as 16 percent.

You can increase the warmth of your wet suit by as much as 16 percent by wearing a vest. Wet-suit vests come in all types of material and provide an additional layer of insulation (see figure 4.15). Layering is an effective technique that reduces water circulation while increasing insulation. Some vests have attached hoods.

Wet Suit Care and Maintenance

With proper care, wet suits can provide years of service. Soak your wet suit in clean, warm freshwater after each use. If you can't soak your suit, rinse it. Hang your suit on a wet suit hanger, which is wider than a typical clothes hanger; dry

the suit in the shade; and store it on the hanger in a cool, dark, smog-free location. A garage is a poor environment for storing equipment because of emissions from automobile exhaust. Do not fold your wet suit for storage because folds can form permanent creases, which insulate poorly. Inspect your suit regularly for seam integrity and tears. Make repairs or have them made as needed.

Dry Suits

When the water temperature is less than 60 °F (15.6 °C), consider using a dry suit, which is an exposure suit that keeps out water. Dry suits are much warmer than wet suits for three reasons:

1. Air is in contact with your skin instead of water. Because air conducts less heat than water, you lose less heat via conduction.
2. You wear undergarments beneath a dry suit. The undergarments trap a layer of air, which is a good insulator.
3. Suit compression affects two of the three types of dry suits only slightly, so the insulating ability of those types of dry suits remains nearly constant regardless of depth. Constant volume in a dry suit provides an advantage over a wet suit, which compresses with depth and expands during ascent.

Dry suits are identified by the material from which they are made. Table 4.2 compares the three most common types of dry suits. The foam neoprene dry suit is made of the same material as a wet suit, but the dry suit has seals at the wrists and neck, attached dry boots, and a waterproof zipper (see figure 4.16).

A dry suit can cost two to five times as much as a wet suit, but with proper care and maintenance, a good dry suit will last many times longer than a wet suit. If you dive primarily in cold water, a dry suit is a good investment.

A dry suit creates a closed air space that is subject to squeezing. A low-pressure inflator valve, which is a standard feature, allows you to add air to the suit during descent to prevent suit squeeze. Because air expands during ascent, an exhaust valve is another standard feature. Get a dry suit that has a constant-volume exhaust valve that automatically maintains a constant state of buoyancy during ascent.

Diving is a diuretic activity—that is, it increases the amount of urine you produce. Urination in a wet suit is a regular practice by divers in open water, but the problem is more complex when you wear a dry suit. An optional feature for a dry suit is a relief zipper, which is useful only when you are out of the water.

Table 4.2 Dry Suits

Type	Advantages	Disadvantages
Foam neoprene	Form fitting and streamlined	Long drying time; hard to locate and repair leaks; buoyancy control difficult
Crushed neoprene	Durable; easy to repair; long lasting; less buoyant than foam neoprene	Less insulation than foam neoprene; expensive; somewhat bulky
Shell (two kinds) • Coated nylon • Rubberized fabric	Fast drying; easy to repair; nylon suits inexpensive; rubberized suits long lasting	Easily punctured; bulky; nylon suits do not last long; rubberized suits expensive

Divers wearing a dry suit have to exit the water to urinate, and those without a relief zipper must disrobe.

Several types of undergarments for dry suits are available. Some are inexpensive but compress with depth and lose insulating ability when wet. More expensive undergarments are highly resistant to compression and retain most of their insulating quality even if wet. Many divers who wear dry suits prefer to wear two layers of undergarments: a thin garment against the skin to carry perspiration away from the body and a thicker overgarment to provide the bulk of the insulation. Moisture conducts heat, so you are warmer when perspiration is wicked away from your body.

Dry suits have several drawbacks. Controlling buoyancy is more difficult with a dry suit than with a wet suit. Rapid, uncontrolled ascents can occur unless you control the suit. Training is essential before attempting to dive in a dry suit. Dry suits are bulkier than wet suits. It is easier to don and remove a dry suit than a wet suit, but the bulkiness of a dry suit makes surface swimming difficult. A dry suit also restricts your mobility more than a wet suit does. The inconveniences of a dry suit are of little consequence, however, when warmth is your primary concern.

Figure 4.16 Dry suits can be made from foam neoprene, crushed neoprene, or nylon.

Berenika Lychak/iStock/Getty Images

Dry Suit Accessories

Most dry suits have attached boots, a desirable feature. Thin latex booties are found on less expensive suits; more expensive suits have hard-soled boots. You should wear heavy socks for insulation with latex booties, and you must wear wet suit boots over the latex booties to protect them. The end result is that divers who wear dry suits usually need fins with large foot pockets.

You can wear neoprene gloves or mitts with a dry suit. If the water is extremely cold, you can get dry gloves with insulating liners that attach to some dry suits. An attached latex dry hood is an option for some models, but most divers use a separate neoprene dry suit hood. Figure 4.17 shows examples of dry suit accessories.

Dry Suit Care and Maintenance

Dry suits require more care than wet suits. The zipper and the control valves are expensive to replace, so do not allow salt crystals to form in the zipper or in the valves. Soak and rinse the zipper and valves in clean freshwater as soon as possible after a saltwater dive. Wash the neck and wrist seals with soapy water, and then rinse them. Coat latex seals with pure talc after they dry. The talc helps protect the rubber from the elements. Fold dry suits in half over a wide hanger for drying. Lubricate the suit zipper according to the manufacturer's instructions, then store the suit with the zipper open. If the suit needs to be repaired, have an authorized dealer make the repairs. As with wet suits, store your dry

Figure 4.17 Dry-suit boots and gloves help keep divers warm.

suit inside your home. The rubber seals on dry suits are adversely affected by exhaust emissions.

Exposure Suit Selection

You should consider many factors when selecting an exposure suit and accessories: your physical characteristics, where you intend to dive, how you intend to dive, how much diving you intend to do, and what you intend to do while diving. The amount of money you want to invest is also a factor, but keep in mind that buying an inexpensive suit may be false economy. If the suit does not meet your needs, you will have to spend more money for another suit.

The amount of diving you intend to do is an important factor. If you plan to make only one dive per day, your insulation requirements are not as great as they are if you plan to dive several times per day. The more time you plan to spend in the water, the warmer your suit needs to be.

If you are a thin person who gets chilled easily, you need more insulation than the average person for a given water temperature. People with more-than-average body fat may not require as much insulation as those with average body fat; fat is a good, natural insulator. Thermal comfort is essential for diving safety and enjoyment.

If you intend to do most of your diving in one area, the most popular type of suit in the area is probably the best type for you. If the local waters are cold, you have to choose between a wet suit and a dry suit. If you choose to get a wet suit as your first exposure suit for cold-water diving, a layered design with Farmer John pants, a vest, and a step-in jacket with an attached hood retains more warmth than high-waisted pants, a regular jacket, and a separate hood. A custom fit retains more warmth than a suit of a standard size.

Dry suits are not desirable for long surface swims because you can overheat, and the drag caused by the bulkiness of some dry suits may cause you to tire or cramp. On the other hand, dry suits retain much more warmth at depth than wet suits. If most of your diving will be deeper than 40 feet (12 m), if the water is 60 °F (15.6 °C) or colder, and if you can avoid long surface swims, a dry suit is a good choice—as long as you obtain training before using the suit.

If you plan to dive in a variety of climates, a spandex or thermoplastic suit combined with a neoprene wet suit may be a good option. You can wear different parts of the suits to meet different warmth requirements. Your diving activity affects your needs. An underwater hunter looking for game generates more body heat than an underwater photographer whose movements are minimal. The less active you are while diving, the more insulation you need. Your desires, needs, and budget determine the features you select for your exposure suit. When choosing features for your suit, you should consider the features that are popular with local experienced divers and diving leaders.

The accessories you select for your suit depend on the type of suit you choose, the temperature of the water, the activity you intend to pursue, and your budget. A hood may be thin and short for warmer water, thick with a long skirt for colder water, or a dry suit type. Foot coverings may be low cut, ankle high, or attached to the suit. The soles may be soft or hard. Hand coverings range from nothing to thick mitts, gauntlets, or dry gloves.

Local diving professionals can help with your selection of an exposure suit and accessories. No matter what type of exposure suit you choose, keep in mind that it is an investment in your enjoyment of diving. Diving is not fun if you get cold.

Weighting Systems

Exposure suits increase your buoyancy. You need weights for ballast in order to achieve neutral buoyancy. One type of weighting system is a weight belt (see figure 4.18); another type integrates the weights into the scuba unit.

Weights can be attached to or inserted into a belt that you wear around your waist. The belt is heavy nylon webbing two inches (five cm) wide. You can thread the belt through lead weights, wrap pouches of lead shot around the belt, or put weights or pouches of lead shot into pockets on pocket-type belts. A hollow fabric belt that you can fill with lead shot is more comfortable on your hips than hard weights. Lead shot, which comes in different sizes, causes less damage than hard lead weights if you drop the belt accidentally. Smaller-diameter shot weighs more per volume than larger shot, so the more weight you need, the smaller the shot you should use.

Because exposure suits compress with depth, a weight belt around your waist loosens unless it has a means of compensating for the suit compression. An

Figure 4.18 Weight belt with various sizes of weights.

elastic compensator is a desirable feature for a weight belt; you can select from a variety of designs.

Secure the weights on your weight belt so that they will not shift position. Pocket-type belts are good in this respect. When you thread separate weights onto a belt, you should secure the first and last weights with retainers, which are available as accessories.

The most important feature of any weighting system is the quick release. In the event of an emergency, you must discard weights quickly to establish positive buoyancy. No matter what type of system you choose, a reliable, easily located, and easily operated quick release is essential.

Several types of weights are available. Large, curved hip weights help offset the buoyancy of cold-water exposure suits. Smaller, rectangular weights find widespread use because they are economical. Coated weights are attractive and practical. Manufacturers mold weights into various shapes and offer a heavy vinyl coating as an option. The coating reduces pollution from lead, improves appearance, makes the weights easier to see in the water, and reduces suit abrasion. Fabric mesh packets filled with lead shot are popular. You can get bulk lead shot for hollow fabric belts. Coated lead shot, although slightly more expensive, is better than uncoated shot. Shot-filled tubular ankle weights are also available, although some diving experts believe ankle weights are unnecessary.

An integrated weighting system eliminates the need for a weight belt by holding lead weight in the **backpack** that holds the scuba cylinder or in the buoyancy compensator. An integrated weighting system makes the scuba unit heavier, but it also allows better distribution of weight than a separate system can provide. Some divers consider an integrated weighting system advantageous because weights are less likely to shift, a suit compression compensator is not needed, and having the weight above the waist improves the diver's trim in the water and reduces strain on the lower back. However, when you raise your center of gravity, you are more likely to lose your balance and fall when you are out of the water. You need strength, good balance, and caution to use an integrated weighting system. Integrated systems usually use lead shot, either in bulk or in pouches.

Weighting System Care and Maintenance

Weighting systems do not require as much care and maintenance as other items of diving equipment. Do not soak or rinse raw lead after use because lead in the runoff pollutes the environment. If gray water seeps from your weight system, replace the lead with new, coated lead. (You can recycle old lead, so do not throw it away.) Pocket-type weighting systems allow you to remove the weights and rinse the remainder of the system.

Regularly inspect the functional aspects, such as the quick release and the compensator. If you use a belt, inspect the weight retainers to make sure they are not broken, and make sure the end of the belt is clean and neat for easy insertion into the buckle. If the end of your belt is frayed, trim and singe it. When you fix a frayed end, be careful that you do not cut or burn yourself.

Weighting System Selection

When selecting a weighting system, you should consider your physical characteristics, the amount of weight you need, and how frequently you will need to change the amount of weight you use.

If you have a large build, your waist is larger than your hips, and your exposure suit is thick, a weight belt may not work well—especially if you need more than 30 pounds (13.6 kg) of weight. In chapter 6, you will learn how to test your buoyancy to determine the amount of weight that you need. Consult with divers whose physical characteristics are similar to yours for advice about weighting systems. If you require more than 30 pounds of weight, you may need weights on a weight belt plus an integrated system. If you require only a small amount of weight, nearly any type of weight system is acceptable.

If the type of diving you do varies, the amount of weight you need also varies. If you dive in freshwater and in saltwater, you need to adjust your weights. When you vary your exposure suit configuration, you also need to vary the amount of weight you wear. The more you need to change weights, the more you need a weighting system that allows changes to be made easily.

Buoyancy Compensators

A buoyancy compensator (BC) helps you control buoyancy. You can inflate your BC at the surface to increase buoyancy, deflate it to reduce buoyancy for descent, and add air to it to achieve neutral buoyancy underwater. Most BCs also contain a backpack to hold your scuba cylinder.

The three types of BCs are jacket style, back mounted, and front mounted (see table 4.3 for style comparison). Most BCs used in diving today have a wraparound jacket design, which provides front and rear buoyancy. There are two basic jacket designs: One style has inflation tubes over the shoulders; a newer style has straps over the shoulders. The straps have convenient, adjustable releases (see figure 4.19).

Figure 4.19 Jacket-style buoyancy compensator with shoulder straps and quick-release clasps.

Table 4.3 Buoyancy Compensators

Type	Location of buoyancy chamber	Advantages	Disadvantages
Jacket style	Front and rear	Even lift; diver can remain upright	Not suitable for skin diving
Back mounted	Rear	Does not interfere with valve operation for dry suits	Pushes diver forward; difficult for diver to remain upright
Front mounted	Front	Suitable for skin diving; allows diver to remain upright	Need separate backpack for cylinders; requires disconnection of inflator hose before removal; does not provide as much lift as jacket

Back-mounted BCs place buoyancy chambers behind you. These systems are useful for underwater modeling and for specialty diving activities. Models look better without bulky BCs covering them, and back-mounted units do not interfere with valve operation on dry suits as much as other types of BCs do.

Front-mounted BCs fit around your neck and cover your chest area. This was the first type of BC, but few divers use front-mounted BCs today. Most divers today believe that BC jackets are superior to front-mounted ones. You can use a front-mounted BC for both skin diving and scuba diving. Jacket-style and back-mounted BCs are not suitable for skin diving.

In its simplest form, a BC is a bladder with attachments (see figure 4.20). The bladder may be coated material that constitutes the BC itself, or a separate bladder may be inside a fabric shell. The seams of a BC bladder are glued or electronically welded, depending on the type of material.

Buoyancy Compensator Selection

When selecting a BC, you should consider your physical characteristics, where you plan to dive, and the type of diving you plan to do. Some BCs provide more buoyancy than others. A cold-water diver wearing a wet suit needs more lift than a warm-water diver wearing a thin exposure suit, although a large amount of lift is not necessarily desirable.

If you are a large, strong individual, the size and bulk of a BC may not be of much concern. But if you are small, you may be wise to choose a small, compact BC. The length of the BC is important. If the BC extends downward too far, this makes donning and removing your weight belt difficult. If you are shortwaisted, consider an integrated weighting system. The fit is important because a BC needs

Figure 4.20 Desirable BC features.

to support you in the water. The BC should fit snugly and not ride up on your body. Models that allow adjustment for a custom fit are desirable.

Another desirable feature is a single-bladder construction, which is less expensive and requires less maintenance than a separate bladder. Renting and using different types of BCs will aid you in making a selection. Talk to experienced divers and diving professionals, and observe BC preferences in the area where you dive.

Buoyancy Compensator Care and Maintenance

Your investment in a BC is large. But, like most diving equipment, your BC will provide years of service if you take care of it. You need to rinse your BC inside and out after use, especially after use in a swimming pool or in the ocean. Chlorine in pool water and salt crystals from seawater are harmful to your BC. Drain the water from your BC after use, fill it about a third full with freshwater, swish the water around, and then drain the bladder. Rinse the inflator assembly thoroughly, and leave the BC fully inflated until it dries. Inflation will test the airtight integrity of the bladder and valves. If the BC does not remain firm for

Diving Equipment Checklist

- ✓ Mask, snorkel, and snorkel keeper
- ✓ Fins and boots
- ✓ Scuba tank (filled)
- ✓ Buoyancy compensator
- ✓ Exposure suit, hood, and gloves
- ✓ Weight system
- ✓ Regulator with pressure gauge
- ✓ Alternate air source
- ✓ Instruments to monitor depth, time, and direction (separate or integrated)
- ✓ Signaling devices (whistle, mirror, safety tube)
- ✓ Dive knife
- ✓ Float, dive flag, and anchor
- ✓ Dive tables
- ✓ Dive light
- ✓ Slate and pencil
- ✓ Marker buoy
- ✓ Collecting bag
- ✓ Gear bag

Spare equipment

- ✓ Scuba tank(s)
- ✓ Weights
- ✓ Straps
- ✓ O-rings
- ✓ Snorkel keeper

Secondary equipment

- ✓ First aid kit
- ✓ Emergency phone numbers and radio frequencies
- ✓ Logbook
- ✓ Swimsuit
- ✓ Towel
- ✓ Jacket
- ✓ Hat or visor
- ✓ Sunglasses
- ✓ Dive kit
- ✓ Save-a-dive kit
- ✓ Drinking water

at least an hour, take it to a professional repair facility. Anytime your BC fails to function properly, have it professionally repaired. It is hazardous to attempt BC repairs without special training, tools, and parts.

Scuba Cylinders

A scuba cylinder stores compressed air at high pressure. The container must be strong and free of corrosion. Scuba cylinders (also called tanks) are made of either aluminum or steel. Each type has advantages and disadvantages. Figure 4.21 shows examples of various steel and aluminum tanks.

Figure 4.21 Scuba cylinders come in many sizes and are made of steel or aluminum.

Steel Cylinders

Steel scuba cylinders come in various sizes with various pressure ratings. Common sizes are 63, 71.2, 80, and 100 cubic feet (1,784, 2,016, 2,265, and 2,832 L). The pressure to which the cylinders can be filled, called the working pressure, ranges from 1,800 to more than 4,000 pounds per square inch (psi), or 122 to 272 standard atmospheres (atm).

Some compressors pump air only to 2,500 psi (170 atm). When you cannot get air at a pressure higher than 2,500 psi, a steel tank may be more desirable than an aluminum tank. A steel 71.2-cubic-foot scuba tank filled to 2,250 psi (153 atm) contains about 5 cubic feet (142 L) more air than an 80-cubic-foot aluminum tank filled to the same pressure. The aluminum tank must be filled to 3,000 psi (204 atm) to obtain 80 cubic feet.

The main disadvantage of a steel cylinder is that it can rust, which can render a tank unsafe and unusable. Do not allow water inside a scuba cylinder. The high-pressure atmosphere has a large amount of oxygen to fuel corrosion. You can keep the inside of a steel scuba tank dry, but the outside is exposed to moisture. Galvanizing inhibits rust on the outside of steel tanks, but the inside may not be galvanized because galvanization affects air purity. Painting a galvanized surface may improve the appearance of a cylinder, but paint alone is an inadequate finish because cracks or chips in the paint allow moisture to reach the steel. The tank will begin to rust unless there is a galvanized coating beneath the paint.

Another disadvantage of a steel scuba tank is that it has a rounded bottom because of the manufacturing process. The tank cannot stand by itself unless you place a rubber or plastic boot, called a tank boot, on the end of the cylinder. The boot makes the base of the tank flat so it will stand. Some boots have flat sides to help keep a tank from rolling when you lay the cylinder on its side. Moisture and salt trapped between the tank boot and the cylinder can cause corrosion. Boots with internal ridges—the preferred type of boot—are self-draining.

Aluminum Cylinders

Aluminum alloy cylinders also come in various sizes with various pressure ratings. Common sizes are 63, 80, and 100 cubic feet (1,784, 2,265, and 2,832 L). The working pressure for aluminum cylinders is 3,300 psi (224 atm).

Aluminum corrodes, but the oxide that forms arrests the corrosive process—this is a significant advantage over the corrosive process of rust in steel tanks. Rust is an accelerating process, but corrosion in an aluminum cylinder is a self-arresting process.

The bottom of an aluminum cylinder is flat. You do not need a tank boot on the cylinder to allow it to stand by itself, but many divers put boots on aluminum tanks to protect the tank and any objects struck by the bottoms of the cylinders.

Aluminum cylinders also have drawbacks. Aluminum is softer than steel, so aluminum tanks can be dented and gouged more easily than steel tanks. On aluminum tanks, brass cylinder valves control the flow of air. Electrolytic action between the dissimilar metals of the cylinder and the valve can cause the valve to seize in the aluminum cylinder threads unless the valve is removed periodically and coated with a special compound. Valve seizing seldom is a problem with steel cylinders.

Aluminum cylinders do not need to be galvanized. You can paint them to improve appearance, but do not bake the paint finish. Temperatures hotter than 180 °F (82.2 °C) reduce the strength of an aluminum cylinder and can cause it to explode when filled. If you would like your cylinder painted, have it done by a professional tank-painting service.

Cylinder Markings

Several rows of markings on the neck of a scuba cylinder provide useful information about the cylinder. You should be able to determine the meaning of several of the marks. Figure 4.22 shows an example of cylinder neck markings.

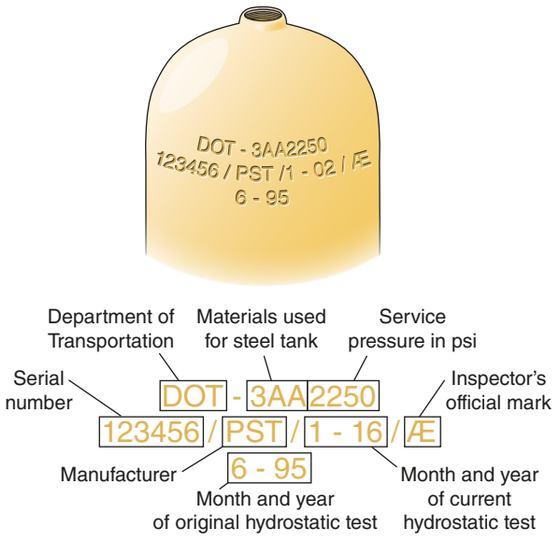

Figure 4.22 Cylinder neck markings.

The first row of marks on tanks manufactured in the United States discloses the government agency that sanctioned the manufacture of the tank, the type of metal from which the tank was made, from, and the working pressure. The first letters in the row identify the government agency, such as DOT (U.S. Department of Transportation), CTC/DOT (Canadian Transportation Commission and the U.S. Department of Transportation), or ICC (the former U.S. Interstate Commerce Commission). The next characters in the first row identify the type of metal that the tank was made from. The marks 3A and 3AA are designations for steel cylinders. The marks 3AL, E 6498, and SP6498 are designations for aluminum cylinders. The final figures in the first row—the important ones for you to remember—are numbers indicating the working pressure of the cylinder in pounds per square inch.

The second row of markings includes the serial number of the tank (which you should record for identification purposes) and letters or numbers identifying the manufacturer of the cylinder. The month and year of the tank's first pres-

sure test are located somewhere below the second row. U.S. regulations require compressed gas cylinders to be pressure-tested before being put into service and every five years thereafter. Subsequent test dates can be stamped anywhere on the neck of the tank. A registered symbol between the month and year of a pressure test identifies the facility that did the testing.

Tanks are filled at an air (or "fill") station (see figure 4.23).When you take your cylinder to an air station for filling, the facility personnel will examine the tank markings to determine whether the cylinder test date is current and to identify the pressure to which the tank can be filled.

Cylinder Selection

The main criteria for the selection of diving equipment—fit and comfort—apply to scuba tanks as well. A small person should use a small tank. Larger divers may require larger tanks because they have larger lungs and use more air.

The material from which a tank is made affects its capacity, size, and working pressure. Although steel is heavier than aluminum, aluminum tanks are larger and heavier than steel tanks with similar capacities. Because aluminum is not as strong as steel, the walls of an aluminum tank are thicker than the walls of a steel tank of approximately the same capacity. The higher the pressure rating, the greater the capacity of a tank of a given size. Modern high-pressure steel scuba tanks are high-capacity cylinders, but they are quite heavy. Good sizes for small divers to consider for an initial scuba tank are 50 and 63 cubic feet (1,416 and 1,784 L). Popular sizes for divers of average size are 71.2 and 80 cubic feet (2,016 and 2,265 L).

Another important factor to consider when selecting a scuba tank is the buoyancy of the cylinder, which is determined by the volume and weight of the tank. Aluminum tanks are more buoyant than steel tanks. High-capacity tanks have a greater change in buoyancy than smaller cylinders as you use air from them. The change in buoyancy between a full and an empty tank can vary by more than 8 pounds (3.6 kg; see table 4.4). Some tanks are negatively buoyant

Figure 4.23 Scuba cylinder filling station.

Table 4.4 Cylinder Size, Working Pressure, and Buoyancy

Capacity (ft^3/L)	Working pressure (psi)	Buoyancy (lb) from full to empty
Aluminum 50/1,416	3,000	–2.7 to +1.3
Aluminum 63/1,784	3,000	–2.3 to +2.7
Steel 71.2/2,016	2,250	–2.0 to +3.6
Aluminum 80/2,265	3,000	–2.0 to +4.4
Steel 76/2,152	2,400	–6.5 to –0.1
Steel 80/2,265	3,500	–7.4 to –1.0
Steel 102/2,888	3,500	–7.6 to +0.5

whether full or empty, but most tanks sink when full and float when empty. Buoyancy for tanks varies so much that you should select the tank most often used for diving in your area or try diving with several different tanks to determine which is easiest to manage.

You need to decide whether you are going to dive using compressed air, enriched-air nitrox, or mixed gases. If your breathing gas is anything other than compressed air, your cylinder, cylinder valve, and regulator must be dedicated to the specialized gas. It is unsafe to use a compressed air cylinder and regulator with mixed gases. Regulators, valves, and cylinders must be "oxygen clean" and "oxygen safe" when used with mixed gases.

Multiple-tank scuba units are used for specialty applications. As a beginning diver, you do not need double or triple scuba tanks. A single tank is adequate for most diving activities.

Cylinder Accessories

Fabric or plastic sleeves can help protect the exterior of your scuba cylinder. Some sleeves have places to attach items within easy reach. Tank bands, which secure a cylinder to a buoyancy compensator, vary. Some allow you to attach a small, backup scuba cylinder to your main cylinder. Additional accessories are available to help you carry or transport your scuba cylinders. Tank boots are a desirable cylinder accessory.

Cylinder Care and Maintenance

Scuba tanks are high-pressure vessels. They are strong, but you should handle them with care. Exterior damage can render your scuba tank useless. Avoid throwing scuba tanks or allowing them to roll about on the deck of a boat or in the trunk of your car. Secure cylinders for transportation or storage. Unless you are holding a scuba cylinder, do not leave it standing in an upright position, especially at a dive site. If the cylinder falls, it can injure someone or damage the tank valve or regulator. However, you should store your scuba tanks in an upright position so that any moisture inside will settle to the bottom where inspectors can detect it with relative ease. Rinse the outside of your cylinder with clean, fresh water after use, and pay special attention to the boot area of steel tanks.

Corrosion can ruin a cylinder rapidly, and pieces of corrosion can damage a tank valve or scuba regulator. One way to prevent moisture from entering a tank is to keep air in the tank. Water can get into an empty scuba cylinder while you

are diving, so avoid using all the air in your tank. Moisture can enter an empty tank if you store it with the valve open. Store your scuba cylinder with a few hundred pounds (about 20 atm) of pressure inside. A low pressure keeps moisture out but provides little oxygen to aid corrosion if there is moisture in the tank.

The filling process can force water into the tank. A water trap in an air compressor is supposed to remove moisture from air, but if the moisture-removal system does not function properly, water can be pumped into your tank along with air. The filling hose attachments for scuba tanks can get wet, and water inside the filling attachments can be forced into your cylinder. Using a high-quality air station is important.

The diving industry requires an annual visual inspection of scuba cylinders. The examination consists of an external inspection, the removal of the valve from a cylinder, an internal inspection using a special light, special electrical testing for aluminum cylinders, the replacement of the valve, and the attachment of a decal indicating the inspection date. Most air stations require a current inspection sticker on a tank before they will fill it. When you handle your scuba tank, listen for sounds of anything moving inside the cylinder. If you hear anything, have the tank visually inspected.

U.S. government regulations specify that compressed gas cylinders must be pressure-tested every five years. Some countries require pressure testing every year or two. The test is hydrostatic because it takes place in water. An inspector fills a scuba tank with water and submerges it in a closed container that is completely filled with water. The inspector applies pressure to the scuba tank hydraulically, and the tank expands slightly from the pressure. The expansion displaces water from the container holding the scuba tank. The inspector measures the expansion and then releases the pressure. The tank must return to within 10 percent of its original volume within a specified period of time. If the tank is too brittle to expand and contract correctly, the inspector condemns it.

You may transport a scuba tank on an airplane only if the tank is completely empty and the valve is open—a situation that is not good for scuba tanks. Do not transport your cylinder by air. Diving destinations have tanks readily available, so there is no need to take a tank on a dive trip.

Cylinder Valves

Cylinder valves control the flow of a liquid or gas. Four types of valves are available for scuba tanks: the simple valve, the reserve valve, the high-pressure valve, and the multiple-cylinder valve. Because multiple tanks are for advanced specialty diving activities, this section addresses only the simple and high-pressure valves.

Simple Valve

A simple valve is an on-off valve that operates like a faucet. You turn the valve handle counterclockwise to open it and clockwise to close it. The first catalog of diving equipment listed this type of valve as item K, and the valve has been identified as a **K-valve** ever since (see figure 4.24).

The **valve seat** is a soft-sealing surface. It is the portion of the valve that closes and stops the flow of air. You can damage the seat with excessive closing pressure.

Valves for scuba tanks have several features, one of which is a snorkel tube that extends from the bottom of the valve into the scuba cylinder. The valve

snorkel prevents moisture or particles from entering the valve when you invert the tank. Another standard feature of tank valves is a thin, metal disk called a **burst disk**. If a tank is overfilled or the heat from a fire causes the tank pressure to increase to a hazardous level, the disk will burst and vent the tank to prevent an explosion. The disks corrode over time, and occasionally a burst disk ruptures. The failure makes a loud noise, and the tank hisses loudly, but the situation is not dangerous (even if it occurs while you are diving, which is unlikely). If the burst disk in your tank valve ruptures, you need to have the valve serviced professionally. Manufacturers rate burst disks for various pressures, and the correct disk must be used. Keep the pressure rating of the burst disk in mind if you want to change a valve from one scuba tank to another. A valve with a low-pressure burst disk will rupture if you use it on a tank with a higher pressure rating.

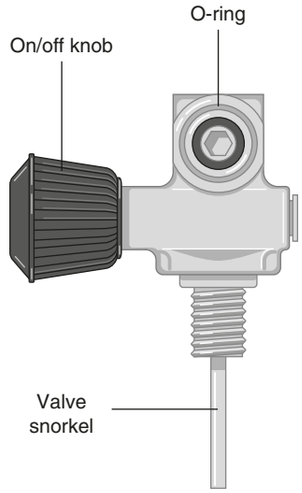

Figure 4.24 K-valve.

There are two types of outlets for scuba tank valves. The traditional outlet is nearly flush with the surface of the valve and surrounded by a soft, circular ring called an **O-ring**. The ring forms the high-pressure seal between the valve and the scuba regulator, so the ring must be clean and free of nicks or cuts. A newer type of threaded outlet with a recessed O-ring seal is a **DIN valve**, which withstands higher pressures than a traditional O-ring valve (see figure 4.25). Tank pressures in excess of 3,000 psi (204 atm) require a DIN fitting.

Figure 4.25 Close-up of a DIN valve.

Reserve Valve

A **J-valve** (the valve was listed as item J in the first equipment catalog) was designed to maintain a reserve of air to permit a normal ascent. The introduction of **submersible pressure gauges** (SPGs) for scuba tanks has rendered the J-valve obsolete.

The reserve lever on a J-valve created problems. If the lever was in the incorrect position (down), the valve would not maintain a reserve. If the diver failed to put the reserve lever in the up position before a dive or if the lever was bumped during a dive, the diver relied on a reserve of air that was not available. Safety concerns also contributed to the demise of the J-valve.

Valve Care and Maintenance

Cylinder valves are made of soft metal and have thin areas, so physical abuse can ruin them. A protective cap is a good accessory for a tank valve. The cap helps

prevent loss of the O-ring, helps keep dirt out of the valve, and helps prevent physical damage to the high-pressure sealing surface around the valve opening. Protection of the valve is one reason that you should secure scuba tanks when you store them and should not leave them standing unattended. If a tank rolls about or falls over, this can damage the valve and render it inoperable.

When you open a tank valve, you should turn it slowly, open it all the way, and then close it one half turn. If something strikes the handle, a valve seat that is fully open will be damaged more than if it were closed slightly. When you close the valve, avoid excessive force, which shortens the life of the valve seat.

Rinsing your tank valve after use is a good idea, but soaking the valve by inverting your scuba tank in a container of warm water is better than rinsing. Water remains in the valve opening after rinsing or soaking. Open the valve momentarily to blow the water from the opening. If you leave the water in the opening and have the tank filled before the water evaporates, moisture will be forced into the tank.

Have your tank valve serviced annually by a professional. Also have your valve professionally serviced anytime it fails to operate easily or when the burst disk needs to be replaced. Valves receive partial servicing (lubrication) during the annual visual inspection of your scuba tank. The large O-ring that seals the valve to the scuba tank can be replaced at the time of the visual inspection. But partial servicing of your valve during a tank inspection is not the annual servicing of the valve. In a complete valve servicing, a qualified repair technician completely disassembles the valve, cleans the parts, replaces various parts, reassembles the valve, and then tests the unit.

Scuba Regulators

The function of a regulator is to reduce high-pressure air to a breathable level. Most scuba regulators use two **stages** of pressure reduction. The first stage of a regulator attaches to the valve of a scuba cylinder and reduces the high pressure to an intermediate pressure of about 140 psi (9.5 atm). The first stage of the regulator connects via a hose to the second stage, which contains the mouthpiece. The second stage reduces the pressure from the intermediate level to the surrounding pressure. A scuba regulator is a demand system; it delivers air only when you demand it by inhaling, unlike a constant-delivery, free-flow system, which commercial divers use. Scuba regulators are highly reliable and have a fail-safe design that turns the demand system into a free-flow system in the event of a component failure. Figure 4.26 provides more information about scuba regulator terms.

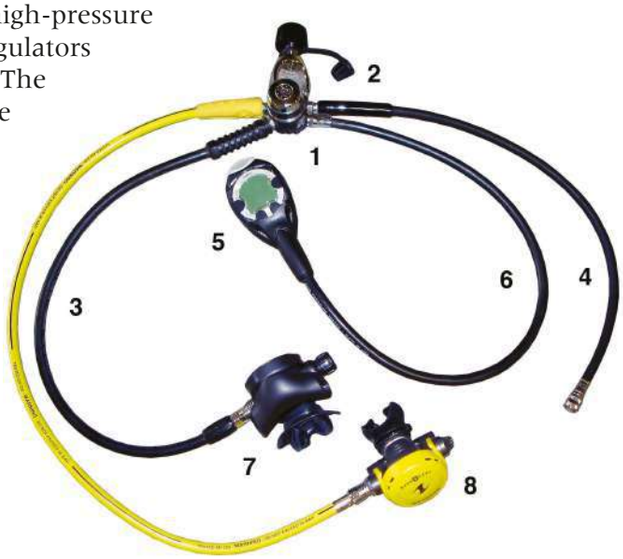

Figure 4.26 Scuba regulator nomenclature: (1) first stage, (2) dust cover, (3) low-pressure hose, (4) low-pressure inflator hose, (5) console, (6) high-pressure hose, (7) primary second stage, (8) extra second stage.

First Stages

The first stages on scuba regulators are either balanced or unbalanced. Changes in tank pressure affect the performance of a balanced first stage only slightly. With an unbalanced first stage, the performance of the regulator changes as tank pressure changes, so a balanced first stage is desirable.

The two main types of valves for the first stages of regulators are diaphragm and piston (see figure 4.27). A diaphragm first stage has a diaphragm that keeps water and dirt from the working parts inside. A bias spring combined with water pressure pushes the valve open. Tank pressure closes the valve when the first-stage pressure equals intermediate pressure plus water pressure. A diaphragm valve has more parts than a piston first stage, so it is more expensive to manufacture and service. The diaphragm valve's exclusion of water and dirt from the mechanics allows high performance for longer periods of time than a piston valve.

Piston first stages have an open, simple design with few moving parts. Water pressure in direct contact with the piston combines with the force of a bias spring to open the piston valve. Tank pressure causes the piston to move and the valve to close when the first-stage pressure equals intermediate pressure plus water pressure. Piston first stages are easier and less expensive to service than diaphragm first stages. But dirt, salt crystals, and mineral deposits that accumulate inside a piston regulator can affect its performance. Each type of first stage has advantages and disadvantages, so either type is acceptable. A diaphragm-piston regulator combines the diaphragm and piston concepts. The diaphragm excludes water and dirt and transfers pressure to a piston.

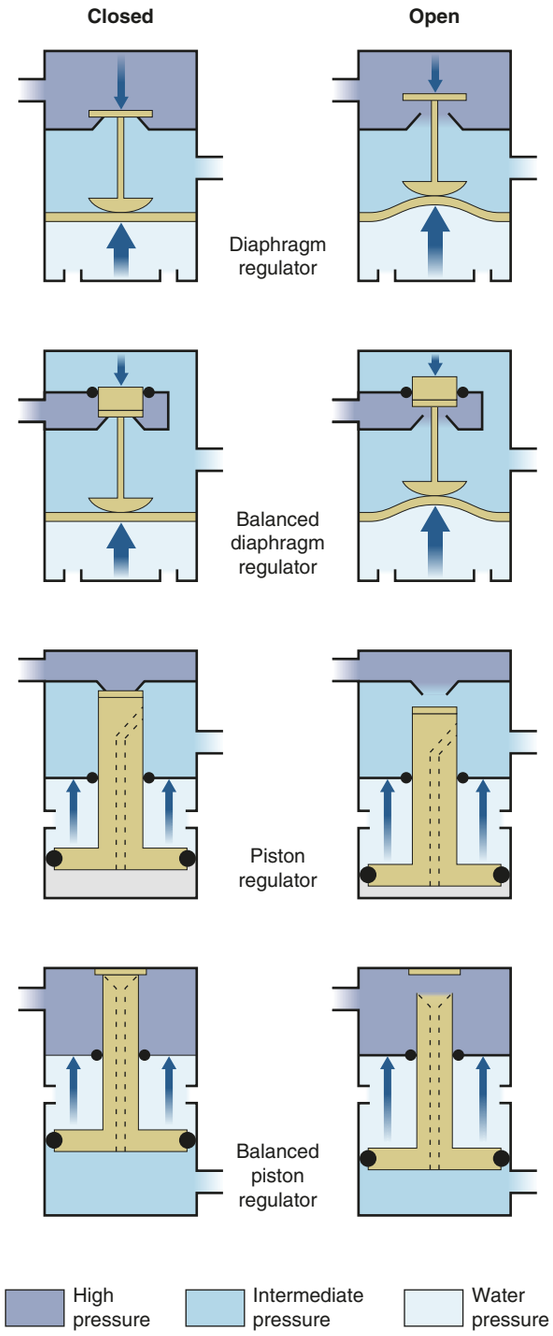

Figure 4.27 Typical regulator first stages. Red arrows represent spring pressure. Black dots represent O-rings.

A first stage must have a means of attachment to the tank valve. A typical regulator has a yoke that surrounds the valve and mates the regulator to the tank valve. The regulator inlet, which has an inlet filter, secures to a tank with a yoke screw. Scuba regulators that operate at above-average pressures use a DIN fitting instead of a yoke screw. A DIN fitting screws directly into a DIN valve and does not have a yoke.

First stages include multiple openings called **ports**. One of the ports is for high-pressure air measurement with an SPG. The remaining ports are for low-pressure air. A regulator should have several low-pressure ports to supply air to the primary second stage, an alternate second stage, a BC inflator, and possibly a dry suit inflator. The sizes of ports vary. The high-pressure port is usually larger than the low-pressure openings. This helps prevent connection errors; a low-pressure hose inadvertently connected to the high-pressure port would rupture.

Some regulators feature environmental shielding by sealing special fluid inside a flexible chamber attached to the first stage. The sealed, flexible chamber transmits water pressure to the regulator, but no water, salt, or dirt can enter the first stage. Extremely cold water can cause an unshielded regulator to freeze, but the fluid in an environmental chamber does not freeze.

Second Stages

The most common second stage is shaped like a cup lying on one side (see figure 4.28). Imagine a pliable diaphragm across the top of the cup, a mouthpiece attached to the bottom of the cup, and an exhaust valve attached to the lower side. A lever that activates a valve inside the container is in contact with a static diaphragm. An inhalation through the mouthpiece creates a partial vacuum inside the second stage. The pressure reduction pulls in the diaphragm, which moves the lever and opens the valve, allowing air to flow from the first stage

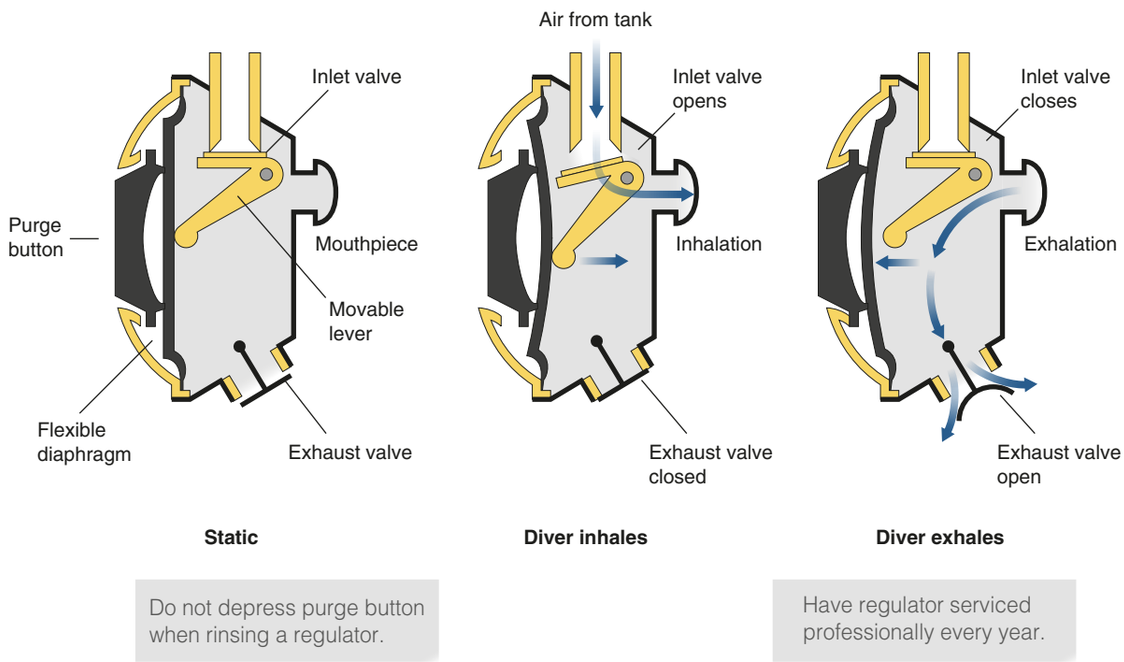

Figure 4.28 Regulator operation.

of the regulator into the second stage. When you stop inhaling, the buildup of pressure returns the diaphragm and lever to their normal positions, closing the valve and stopping the flow of air. When you begin to exhale, increased pressure inside the second stage opens the exhaust valve, allowing air to escape.

The two types of second-stage valves for scuba regulators are the downstream valve and the pilot valve. In a downstream-valve regulator, a small bias spring holds a valve closed. Inhalation moves the diaphragm, which moves a lever. The lever movement overcomes the resistance of the bias spring and opens a valve, allowing intermediate-pressure air to enter the second stage. After inhalation, air flows until the diaphragm moves outward, allowing the bias spring to close the valve. Downstream second-stage valves are simple and inexpensive, and they tolerate more sand and dirt than pilot valves do.

In a pilot-valve regulator, the movement of the diaphragm in the second stage opens a small valve that in turn opens a larger valve. When you stop inhaling from the regulator, the diaphragm returns to its normal position, and the valves close. A pilot valve delivers air up to four times more easily than a downstream valve. But pilot valves are more expensive to manufacture and service than downstream valves, and in shallow water, you may experience the effect of shuddering air movement with a pilot-valve regulator.

All regulators have a purge button or area on the regulator that you depress to manually open the second-stage valve. Use the purge to test the regulator, to expel water and debris from inside the second stage, and to relieve the pressure in the regulator after you close the tank valve.

The location of the exhaust valve varies. The exhaust valve may be at the bottom, the side, or the front of the second stage. The position of the exhaust affects the bubble pattern and the clearing of the regulator when it has water inside and the mouthpiece is in your mouth. Some regulators direct exhaust bubbles by means of an exhaust tee. You will learn more about regulator positioning for clearing in chapter 6.

A regulator is either right-handed, left-handed, or bidirectional, referring to the direction from which the regulator hose must come when the regulator is in your mouth. For example, the hose must come from the right side when you are using a right-handed regulator. The hose may come from either side when you are using a bidirectional regulator. The directional configuration is important only for knowing how to orient the regulator when you place it in your mouth. Figure 4.29 shows the possible directional configurations.

Some second-stage casings feature strong, light, durable materials that do not bend or corrode as metal does. Several types and styles of mouthpieces are available. Use a soft, comfortable mouthpiece that does not cause jaw fatigue. A repair technician can replace mouthpieces quickly and easily.

Figure 4.29 Scuba regulators have different configurations.

Regulator Accessories

A regulator hose is flexible but has rigid metal connectors crimped onto the ends. The points where the hose and the metal meet are stress points because the hose fibers strain against an unyielding surface. To prevent the breakdown of the hose fibers at the stress points, you should use hose protector sleeves. At a minimum, equip all regulator hoses with hose protectors at the first-stage end of each hose.

Padded regulator bags help protect your regulator during transportation or storage. The bag should be large enough to accommodate the regulator and all hoses without bending the hoses sharply.

Adapters allow you to use DIN-fitting regulators on standard cylinder valves. Use protective covers for the threads on DIN fittings.

Purge depressors are a built-in feature on some regulators and an accessory for regulators that do not have them. A purge depressor partially depresses the purge button to remove bias spring pressure from the second-stage valve when you are not using the regulator, thus extending the life of the valve seat.

Colored second-stage covers allow you to color-coordinate your regulator with your other equipment. Other accessories include various mouthpieces and hose adapters (as previously mentioned).

Regulator Selection

It takes effort to inhale and exhale through a scuba regulator. Effort is needed to overcome resistance; good regulators have minimal breathing resistance. Compare the performance data for various regulators, and select one that lets you breathe easily over a wide range of tank pressures, which implies a balanced first stage.

Choose a widely used and easily serviced regulator. You should get a regulator that can be serviced by facilities anywhere you happen to be and that uses readily available parts.

The type of diving you do should affect your selection. If you plan to do most of your diving from shore, you should avoid a pilot-valve second stage, which is adversely affected by sand and dirt. Diaphragm first stages are a better choice if most of your diving is from shore.

If you do not have a scuba tank, consider purchasing the tank and regulator at the same time so that you can match the fittings of the regulator and the cylinder valve. If you will be diving in water that is near freezing, you should select an environmentally shielded regulator. Remember that enriched-air nitrox and mixed gases require specially cleaned and dedicated regulators.

Regulator Care and Maintenance

Your regulator is a precision instrument; it requires care and maintenance to ensure the best possible performance. Keep sand and dirt out of your regulator. Do not allow salt crystals to form inside. Soak your regulator in clean, fresh, warm freshwater as soon as possible after diving in the ocean. You need to remove salt before it dries. If you cannot soak your regulator, you should at least rinse it. A combination of soaking and rinsing is best. Follow these rules when rinsing or soaking a regulator:

- *Keep the inside of the first stage dry.* The purpose of the dust cover on the first stage is to keep out water and dust. Develop the habit of replacing the dust

cover and securing it in place with the yoke screw anytime you do not have the regulator attached to a tank. Make certain the dust cover is in place before you rinse or soak a regulator.

- *Allow low-pressure water to flow gently through the second stage and also into the openings on the first stage.* High-pressure water can force dirt or grit into crevices, where it causes damage. Gentle pressure washes the dirt away.
- *Do not press the purge button when you rinse the second stage unless the regulator is pressurized.* If you do not pressurize the regulator and you press the purge button while water is inside the second stage, you open the second-stage valve and allow water to flow through the hose into the first stage.

When the regulator has dried thoroughly, lay it flat for storage. For prolonged storage, place the regulator in a plastic bag to help protect it against the harmful effects of smog. Do not bend the hoses sharply because bending damages hose fibers. Replace hoses that are cut, bulging, or leaking.

You can avoid most problems with your regulator by having it serviced annually. Have your regulator serviced even if it seems to be functioning properly. Failure to invest in regular service can affect your safety and shorten the life of your regulator.

Alternate Air Sources

Several equipment options can help if you run out of air underwater (although running out of air is due only to sheer negligence). Your best option is an alternate air source (AAS), which is a source of compressed air other than your primary scuba regulator. An AAS is valuable if your primary source of air begins free-flowing or leaking during a dive; you can switch to the AAS and make a normal ascent. The two primary types of alternate air sources are extra second stages and backup scuba units. An extra second stage allows two divers to share air without passing a single mouthpiece back and forth. Backup scuba units are fully redundant scuba systems that provide an independent source of air in an emergency. You are not dependent on a buddy to provide air when you are equipped with a backup system. Extra second stages are less expensive than backup scuba units but do not provide the benefits of an independent scuba system.

Extra Second Stages

Two types of extra second stages are available. The first is an extra second stage for your regulator; the second is a BC low-pressure inflator that has an integrated regulator second stage (see figure 4.30). An extra second stage, or **octopus**, should meet the following criteria:

- The first stage of the regulator should be capable of meeting the airflow demands of two second stages.

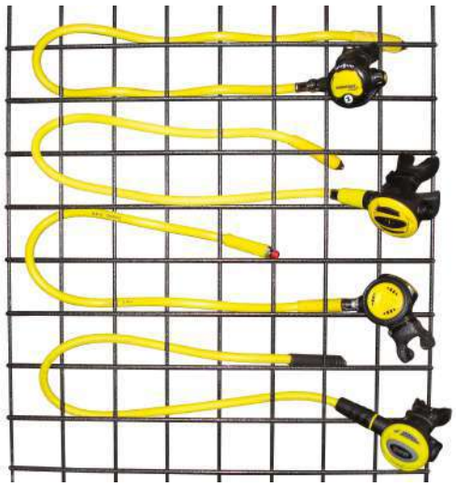

Figure 4.30 Extra second stages.

- The hose on the extra second stage should be several inches longer than the hose on the primary second stage.
- The extra second stage should attach to your chest area in such a way that your buddy can remove it quickly and easily. Do not allow the extra second stage to dangle.
- The attachment device should cover the mouthpiece opening of the extra second stage to prevent the regulator from free-flowing and to keep dirt and debris from getting inside.
- The extra second stage should be brightly colored for easy identification.

A regulator second stage may be integrated into a BC low-pressure inflator in one of two ways. The BC inflator can have a built-in regulator second stage, or quick-release hose fittings on the regulator second stage can allow it to be connected in series in the low-pressure hose leading to the inflator assembly. Both types of extra second stages have advantages and disadvantages. An integrated second stage requires one less hose on your regulator because a single hose provides air for both the extra second stage and the BC inflator. When you must share air and you have an integrated second stage, you must breathe from the integrated second stage because the hose is too short for your buddy to use. With an extra second stage, you or your buddy can use either air source. Extra second stages integrated into the BC inflator may leak air. To stop a leak, you must disconnect the low-pressure air, thereby losing the functions of the low-pressure inflator and the extra second stage.

Backup Scuba Units

Two types of backup scuba units are available. A **pony tank** is a small (about 13- to 20-cubic-foot, or 368 to 566 L) scuba cylinder with a separate, standard regulator. You clamp a pony tank to the side of your main scuba cylinder. A **Spare Air unit** is a smaller (2 to 4 cubic foot, or 57 to 113 L) scuba cylinder with a special regulator integrated directly into the valve. A pony tank provides an adequate supply of air for many situations, but a Spare Air unit provides only enough air to permit an ascent from shallow depths. On the other hand, a Spare Air unit is small and light, whereas a pony tank is bulky and heavy (see figure 4.31).

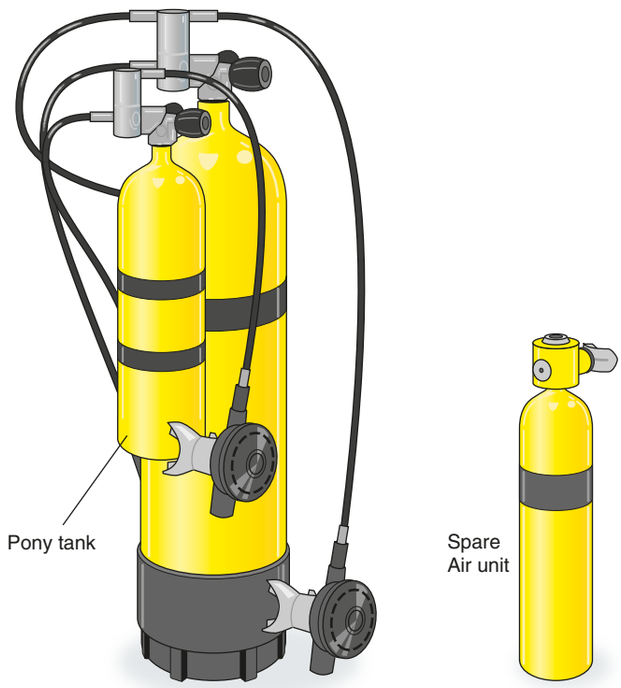

Figure 4.31 A pony tank and a Spare Air unit.

Alternate Air Source Care and Maintenance

AAS equipment should receive the same care and maintenance as your primary scuba equipment. Some divers attempt to save money by having only their primary scuba equipment serviced annually. This is false economy. AAS equipment needs servicing just as much as primary scuba equipment does.

Instrumentation

You move in three dimensions in water the way a pilot does in air. Instruments are important safety requirements for flying. A pilot needs information about altitude, direction, time, and amount of fuel remaining. Similarly, you need information about depth, direction, time, and amount of air remaining. The minimum instrumentation you need is a depth gauge, an underwater timer, a submersible pressure gauge (SPG), and a compass. A dive computer is highly recommended. All required instrumentation can be provided by a single integrated device.

Depth Gauges

To tell how deep you have descended, you can use one of four types of depth gauges: a **capillary gauge**, a **Bourdon tube**, a diaphragm gauge, or an electronic gauge. Manufacturers calibrate depth gauges in seawater. Most mechanical gauges do not indicate the correct depth in freshwater, but they indicate the equivalent seawater depth if you use them starting at sea level in freshwater or if they adjust for altitude pressure changes. The reading of gauges in freshwater is acceptable because tables for time limits at various depths are based on sea-level depths. Table 4.5 compares the four types of depth gauges.

A capillary depth gauge is a simple instrument. It is a hollow, air-filled, transparent plastic tube sealed at one end and placed around a circular dial. The open end of the tube aligns with zero on the gauge dial. A capillary gauge uses the principle of **Boyle's law**. Water pressure compresses the air inside the tube during descent. The position of the air–water interface inside the tube relative

Table 4.5 Depth Gauges

Type	Accuracy	Advantages	Disadvantages
Capillary	Accurate only to depth of about 40 ft (12 m)	Rugged, inexpensive	Can clog with debris or air bubbles
Bourdon tube	Reading accurate within 1% to 2%	Accurate	Can be damaged by reduced pressure at altitude
Diaphragm	Very accurate	Can adjust zero setting for pressure changes at altitude	Expensive
Electronic	Reading accurate within 6 in. (15 cm)	Gauge may zero itself to compensate for changes in atmospheric pressure	Must have sufficient battery power; expensive

to markings on the dial indicates the depth. At two atmospheres of pressure, the air column in the tube compresses to one half its original length.

A Bourdon tube is a thin metal tube formed into a spiral. The tube may be open to the water, or it may be closed and placed inside a housing filled with oil. Oil-filled Bourdon gauges are more popular than open-tube gauges. Pressure on an open-tube gauge begins to straighten the tube and increases the coil diameter. The straightening produces a spiral movement of the tube, which is linked mechanically to a needle to indicate the depth. Water pressure causes the coil of a closed-tube gauge to decrease in diameter. The movement of the coil, linked mechanically to a needle, indicates the amount of pressure exerted on the gauge.

The accurate but expensive diaphragm gauge uses elaborate mechanics to connect a thin, movable diaphragm to an indicating needle. Electronic depth gauges, also accurate and expensive, use a pressure sensor (transducer), electrical circuitry, a display, and a battery to indicate depth.

A maximum-depth indicator is a desirable feature for depth gauges. As you will learn in chapter 7, you must know the depth of a dive for planning purposes. A digital depth gauge retains the maximum depth you attain. The instrument displays the information until the next dive or for 12 or more hours after a dive; it then resets automatically. Many modern depth gauges with needle displays have a thin indicating wire that the gauge needle pushes along the dial face. When the needle retreats, the wire remains at the highest point reached on the dial. You can reset the indicating wire by turning a screw on the dial face. When you use this type of gauge, you must remember to reset the maximum-depth indicator before each dive.

Underwater Timers

You can use either an automatic or a manual underwater timer to keep track of time during a dive. Either type may indicate time with hands on a dial or with a digital display. Pressure activates automatic timers, which start timing at a depth of about 3 to 5 feet (1 to 1.5 m) and stop timing when the depth is less than that. Automatic timers are better than watches because you do not have to remember to start or stop the timing of your dive, although you have to reset some watch-type automatic timers before a dive. Waterproof watches that you can use as underwater timers usually feature a rotating **bezel**, a movable ring that you can set to indicate elapsed time. Digital watches are accurate, but their small buttons can make them difficult to operate. The best timers are electronic automatic timers, which can keep track of how long you dive, how long you are at the surface between dives, and how many dives you make. You do not have to remember to reset or activate anything when you use an automatic electronic timer. Dive computers track underwater time precisely. Some computers can display a timed profile of the dives that you make.

Dive Compasses

Rarely can you see more than 100 feet (30 m) underwater, so a navigational aid can be valuable. If you dive without a directional reference, you can end a dive a long distance from your planned exit point. A dive compass can help you avoid long surface swims or swims through thick surface canopies of water plants. You can use a compass to navigate beneath the canopies, where there

are passages through the plants. A compass is also useful for relocating a precise area underwater and as a surface navigation device if fog arises.

Three types of compasses are card types, needle types, and electronic. Card and needle compasses are mechanical. Magnetic deposits in the earth near the North Pole attract either a magnetized disk or a magnetized needle to provide a directional reference. The compass card or needle can deviate from its correct alignment if it is in close proximity to ferrous metal or a source of magnetic forces such as a magnet or electrical motor.

The earth's magnetic field varies in different locations, and the variations affect the balance of the compass card or needle. Some manufacturers offer compasses balanced for different zones of the earth. A good compass has a good tilt potential of plus or minus 30 degrees. Be sure to ask about balancing when selecting a compass.

Diving compasses have liquid inside to dampen the swinging of the needle or disk. To be useful for diving, a compass needs a reference line, called a **lubber line**, to indicate the direction of travel relative to the needle reference. Another desirable feature is a rotating bezel with bracketing index marks that allows you to mark the needle position for a specific direction.

You view some compasses from the top and some from the side. Side-reading compasses display the course in a window on the side of the instrument. You look across a top-reading compass. You will learn how to read and use a compass in chapter 6.

Digital compasses are usually an integrated component of a dive computer (see figure 4.32). The direction of travel is displayed both graphically and numerically. A heading (desired direction) can be set and the device will indicate any correction needed to remain on course. Bearings can be stored in a memory device.

Figure 4.32 Console with computer and SPG (left) and electronic dive compass (right).

Submersible Pressure Gauges

An SPG is analogous to the gas gauge of a car and is just as essential. The SPG measures scuba cylinder pressure. You can use either a mechanical or an electronic SPG. A mechanical SPG is a high-pressure Bourdon tube. High-pressure air from the cylinder passes through the regulator first stage, through a high-pressure hose, and into a Bourdon tube inside a housing at the end of the high-pressure hose. The pressure tries to straighten the spiral tube, which moves a needle on a dial to indicate the tank pressure. Physical shock can damage a mechanical SPG.

An electronic SPG has a pressure sensor (transducer), circuitry, a battery, and a display. It is a form of a high-pressure depth gauge. The display may be digital or graphic. Either a symbol or numbers (in pounds per square inch, atmospheres, or bars) indicate the amount of air in your tank. If the electronics get wet or if the battery dies, an electronic SPG will not function. Some electronic SPGs are

hoseless—they do not require an air hose between the cylinder and the gauge, which usually serves multiple functions (dive computer, compass, and SPG). You should retain at least 300 to 500 psi (20 to 34 atm) in your tank at the end of a dive. Mechanical SPGs typically have a red area on the dial for the last 500 psi. When you dive, you should monitor your air supply and make sure that you surface before the needle gets into the red area. Electronic depth gauges usually warn of a low supply of air by blinking the display.

An SPG has a **blowout plug** to relieve pressure in the housing in the event of a high-pressure leak. Identify the blowout plug on your SPG, and do not place anything over the plug that will prevent it from functioning. If the plug cannot come out to release high pressure inside the SPG housing, the face of the instrument can explode.

Instrument Consoles

You can purchase diving instruments individually or in combination. It is convenient to combine several gauges into a display unit called a **console**. An instrument console attaches to the high-pressure hose coming from your regulator first stage. When your instruments are in a console, your arms are free of gauges, and dive preparations are quicker.

The two types of instrument consoles are mechanical and electronic. A mechanical console contains an SPG and a depth gauge. Some also contain an underwater timer, a compass, and a thermometer. The instruments usually feature luminous displays, which are easy to read in low light.

With an electronic console, all instrument information is in a single display (see figure 4.32). If one gauge fails in a mechanical console, the remainder can still function; but when an electronic console fails, all the information provided by the unit is lost. Electronic displays are difficult to read in the dark unless they feature some type of illumination.

Dive Computers

A dive computer is an electronic instrument with a pressure sensor, electronic circuitry, a battery, and a display. A programmed computer inside the instrument uses pressure and time information to continuously calculate the uptake of nitrogen by various compartments that have different half-times. A half-time is the length of time required for a mathematical model (compartment) to increase or decrease its gas absorption or elimination by one-half. A compartment—which resembles, but does not duplicate body tissues—is completely saturated or desaturated in six half-times. When the absorption by any one of a computer's compartments reaches a predetermined level, the device indicates that you are approaching the time limit after which a direct ascent to the surface will no longer be possible. On reaching that time limit, the computer indicates a minimum depth—a **ceiling**—that you cannot exceed during ascent. You risk decompression sickness unless you wait until the computer indicates that sufficient outgassing has occurred to allow you to continue your ascent. A dive computer provides extremely accurate time and depth information. Other common features are a low-battery warning, a rapid-ascent warning, a dive log mode, a dive-planning mode, and information about flying after diving. Additional information about dive computers is provided in chapter 7.

Instrument Care and Maintenance

Physical abuse can damage instruments, so protect your instruments from shock. Secure your console instead of allowing it to swing freely.

Heat or prolonged exposure to hot sunlight can cause oil in a liquid-filled gauge to expand and break the seal on the housing encasing the instrument. If you break the seal, you must have the gauge repaired. Hot water in a whirlpool or a shower can cause an underwater timer to expand, break a seal, and allow water inside. Do not subject diving instruments to high temperatures.

Have air leaks in SPGs repaired at the first opportunity. Have your depth gauge tested for accuracy from time to time by a professional repair facility, or compare your gauge with an extremely accurate instrument, such as a digital depth gauge. Follow all the manufacturer's recommendations. Reduced pressure at elevations above sea level can damage some instruments. Unless an instrument is designed for use at altitude, pack it in an airtight container for flying. Soak and rinse instruments with clean, fresh water after use. Give special attention to pressure-sensing areas to prevent them from becoming clogged with dirt or salt crystals.

Electronic gauges require batteries. Some gauges must be returned to the manufacturer for battery replacement. Gauges that permit the consumer to replace batteries will flood if the batteries are not replaced properly.

Dive Knives and Accessories

This section addresses some additional required and optional equipment. Figure 4.33 shows examples of a dive knife and other dive accessories.

Figure 4.33 Dive knives and accessories.

Dive Knives

A dive knife is a required equipment item. Lines and cords in water can cause entanglements, so you must have a knife to cut yourself or your buddy free if necessary.

The many designs of dive knives include large knives and small knives. A small knife positioned where you can reach it easily is better than a large knife that you cannot reach. The most important feature of a dive knife is an effective cutting edge. Good blades are corrosion resistant and hold a sharp edge; serrated blades cut lines more effectively than straight blades.

Some knives are multipurpose tools designed for prying, digging, pounding, and measuring in addition to cutting. If you have a multipurpose knife as a diving tool, you should also have a small, separate dive knife.

A dive knife comes with a sheath, which has some means to lock the knife in place. Make sure the sheath lock is reliable to prevent losing a knife. You can attach a small knife to your leg, your arm, your console, or your BC. Wear

a large knife on the inside of your leg and secure it with straps that stretch to compensate for exposure suit compression.

To prevent corrosion, rinse your dive knife after diving. Inspect the edge for sharpness, remove any rust, and coat the blade with grease.

Accessories

You are likely to need several small, but important, accessory items. These items include a gear bag, a dive flag and float, dive lights, dive kits, signaling devices, dive slates, and a diver's first aid kit.

Gear Bags

You need so much equipment for diving that you also need a means to carry it. Gear bags can be simple or complex. They can feature multiple compartments, padding, novel ways for carrying or moving, sealed fabric edges, and various fabrics. The best type for you depends on your needs and how much you want to invest. Be sure to get a bag with webbing handles that completely surround the bag to provide full, durable support. No matter what your budget, you need a gear bag for your equipment.

Dive Flags and Floats

In many areas, local law requires the use of a dive flag while diving. In the United States, the traditional dive flag is a red flag with a white diagonal stripe (see figure 4.34). The flag is usually vinyl, mounted on a fiberglass staff, and stiffened with

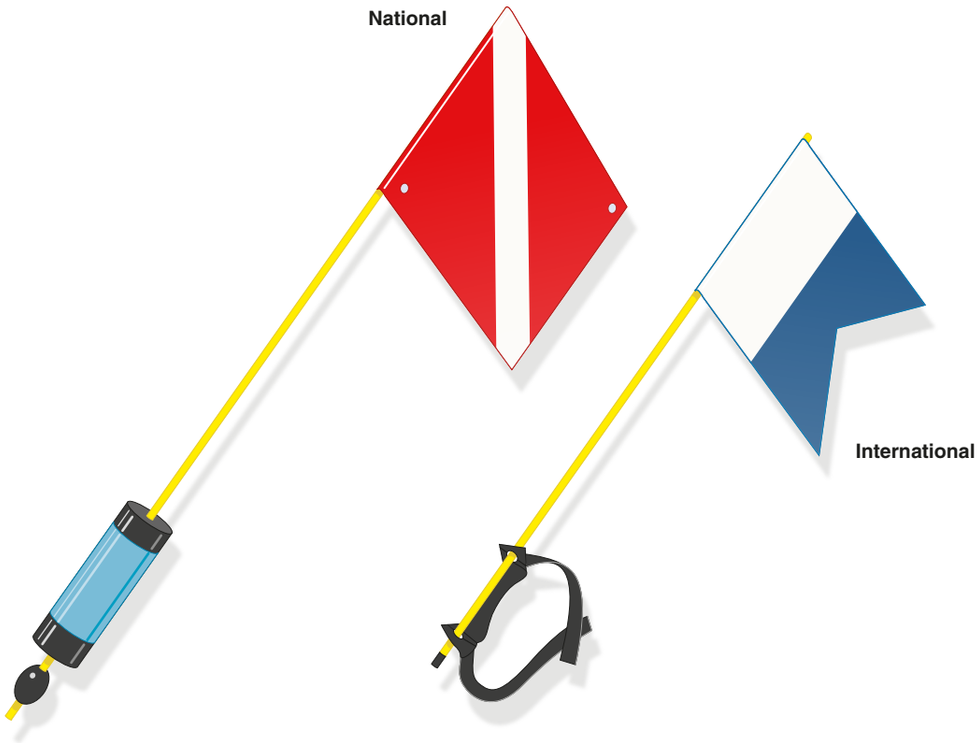

Figure 4.34 Dive flags.

a wire so that it stands out from the staff at all times. In addition to the red and white flag, you should use the international Alpha flag—a swallow-tailed blue and white flag—when diving from a vessel. The Alpha flag is a general dive flag in countries other than the United States.

Unless you are diving from a boat, you need a float to support your dive flag. Some flagstaffs have a float attached. You can also get attachments to secure a dive flag to an inner tube. Another option is a flag-holding canvas covering for an inner tube.

Dive Lights

A dive light can increase your diving enjoyment. Light levels are low underwater, and a light restores color to objects at depth and allows you to peer under ledges and into holes. You can see and enjoy much more when you use a dive light.

The many types of dive lights available include large, powerful, rechargeable lights and small lights that use disposable batteries. Consider a small dive light initially. A large light is for night diving, which is an advanced activity. A small light is easy to carry and is useful both for day dives and as a backup light for night diving. Many of today's small dive lights are bright and compact (see figure 4.35). Maintain your dive light according to the manufacturer's instructions.

Figure 4.35 Dive lights are useful during the day or night to restore color visibility and illuminate dark places.

Kits

Two equipment kits are recommended: a dive kit and a save-a-dive kit. The dive kit contains items you use frequently for diving. The save-a-dive kit contains items you may need to salvage a dive. See the Equipment Kits sidebar for items that you might include in each kit. Because the items in your dive kit are small and get wet, you may want to keep them in a container by themselves so you can find them easily.

Once you get yourself and your equipment to a dive site, you don't want a minor equipment problem to keep you from diving. A broken strap, a missing O-ring, and a torn mouthpiece are examples of problems that can stop you from diving unless you have spare parts. Keep save-a-dive items in a waterproof box. Do not mix wet items from your dive kit with dry items in your save-a-dive kit.

Dive Slates

You need to record and refer to information around and in the water when you dive. Plastic slates are better than paper because water does not affect the slates. You can write on dive slates with a standard lead pencil or a grease pen.

Equipment Kits

Dive Kit

- Defogging solution for your mask
- Lip balm
- Wet suit shampoo
- Seasickness medication
- Sunscreen

Save-a-Dive Kit

- Mask strap
- Regulator mouthpiece
- Fin strap
- Crescent wrench
- Snorkel keeper
- Screwdrivers
- Tank valve O-rings
- Cable ties
- Quick-release buckle
- Twine
- Weight keepers
- Duct tape
- Superglue
- Extra batteries
- Heavy nylon thread and heavy-duty needles
- Neoprene cement
- Fiberglass cast tape for temporary repairs

Dive slates include checklist slates, reference slates, logbook information transfer slates, and writing slates. All types have value. You will probably have several slates when you become an experienced diver. Initially, you should have an equipment checklist, a dive-planning slate, and an underwater writing slate.

Signaling Devices

Long-range signaling devices are invaluable if you become adrift, especially in areas where currents are strong. You should have a whistle readily available. A whistle is more effective than shouting to gain attention. The sound from a whistle does not require much energy to produce, and it travels well over water.

A diver's air horn, powered by low-pressure air, works with as little as 50 psi (3.4 atm) of tank pressure and produces a sound that can be heard up to a mile away. Although this device is small, it is so loud that you must point the horn away from yourself to avoid hearing damage.

You can get long, bright, thin, inflatable safety tubes that fit easily into your BC pocket. Inflate a safety tube at the surface to make yourself easier to spot in the water.

You can use a signal mirror to flash sunlight long distances over water. Other forms of signaling devices for divers are flashlights, strobe lights, and waterproof flares. The greater the chance of being caught in a current, the greater your need for signaling devices.

Diver's First Aid Kit

Diving takes place in remote areas, and because it is a physical activity, someone may be injured. You need to be prepared for an emergency. You should have a first aid kit at the dive site. See the First Aid Kit sidebar for items that should be included in a standard first aid kit. Pack the items in a waterproof container. Chapter 6 addresses the use of the first aid items.

First Aid Kit

- Rescue breathing mask
- Tweezers
- Seasickness medication
- Bandage scissors
- Isopropyl alcohol
- Penlight
- Hydrogen peroxide
- Coins for emergency phone calls
- White vinegar
- Emergency contact information for diving accidents
- Baking soda
- Diving first aid book
- Analgesic and antiseptic ointment
- Space blanket
- Hot packs
- Pen and small notebook
- Eyewash
- Rubber band tourniquet
- Triangular bandages

Specialized Equipment

Other specialized equipment for scuba diving includes nitrox (oxygen-enriched air) equipment, mixed-gas equipment, and rebreathers. This equipment permits divers with specialty training to obtain longer dive times. The risks associated with the use of specialized equipment make training and adherence to the rules pertaining to its use absolutely essential.

Enriched-Air Nitrox (EANx)

Breathing gas that contains levels of oxygen greater than the 21 percent found in air provides benefits for divers. Higher oxygen levels reduce the amount of nitrogen absorbed, thereby reducing the problems posed by decompression illness. However, divers must know the technical requirements and dangers associated with the use of EANx. Dedicated equipment and specialty training are essential.

Filling standard cylinders with EANx is not permitted. Nitrox equipment is clearly marked to differentiate it from standard scuba equipment (see figure 4.36). Nitrox blending stations are available at many modern dive facilities. Standard blends of nitrox are EAN32 and EAN36, which contain 32 and 36 percent oxygen respectively (also called Nitrox32 and Nitrox36). Nitrogen and oxygen are blended together during the filling process, but the final mixture may not be exact. Divers need to test the oxygen percentage of the gas in their cylinder before diving. The divers use a handheld oxygen analyzer to ensure that the oxygen level is acceptable.

Figure 4.36 Nitrox tanks have explicit markings to differentiate them from standard cylinders.

If you breathe oxygen at a partial pressure that is too high,

convulsions can occur. Partial pressures of oxygen must be kept within safe limits (1.4 to 1.6 atm or bars). For this reason, there are maximum operating depths (MODs) for nitrox diving. The MODs for EAN32 and EAN36 are 110 feet (33.5 m) and 95 feet (29 m) respectively. These are absolute depths and must not be exceeded. Additionally, divers must determine equivalent air depths (EADs) for decompression. Modern dive computers allow divers to select the type of breathing gas being used. The computers automatically calculate EADs and are recommended for mixed-gas diving.

Dive operators and gas blending facilities require proof of specialty training for the use of EANx. Training is widely available.

Technical Diving

This form of diving is even more specialized than the use of enriched-air nitrox because it involves the use of mixed gases and highly technical equipment. Breathing gases include helium, neon, nitrogen, and oxygen; these gases are mixed in various quantities depending on the depth. Obviously, divers need substantial training before they attempt this method of diving. Some gas mixtures will not sustain life until the partial pressure of the oxygen in the mix increases with depth. In this instance, technical divers must use travel tanks to reach a safe depth before switching to the low-oxygen mixture in the primary unit. The principal goal of technical diving is to allow divers to exceed recreational depth and time limits. Technical diving has greater risks than recreational diving, but many consider the rewards worth the risk.

Some technical divers use multiple large-capacity cylinders for open-circuit scuba diving. Others use complex and expensive closed-circuit rebreathers (CCRs) to obtain greatly extended dive times. Rebreathers, which use a mixture of oxygen and inert gas, allow exhaled oxygen to be breathed repeatedly until it is metabolized. Additional oxygen is added as needed to the inert gas in the system to maintain required levels. Exhaled carbon dioxide is absorbed by a canister called a scrubber. Oxygen replacement may be controlled automatically by a central processing unit or may be controlled manually by a valve. The result, which is not affected by depth, allows a diver to remain submerged for hours at a time or to make multiple dives with a single fill of breathing gas. Figure 4.37 illustrates a modern CCR.

Dive preparation and predive procedures vary greatly from open-circuit scuba diving. For example, a diver needs to prebreathe a CCR to activate the absorption process of the carbon dioxide scrubber. This procedure can take several minutes and is affected by the temperature of the absorbent. The diver also needs to check the unit's calibration at a depth of 20 feet (6 m) during descent. Changes in the gas mixture that are made during ascent must be managed by a central processing unit or managed manually by the diver. This truly is technical diving.

Figure 4.37 CCR 2000 rebreather.

Courtesy of Dan Wible.

SCUBA WISE

If you want to avoid diving accidents, you should use familiar, correctly fitting, well-maintained dive equipment the way that it is intended. As I look back over several decades of dive experience, I can recall many instances when divers got into trouble because they did not follow the essential equipment safety procedures summarized in the previous sentence. Diving allows us to explore an alien environment using life-support equipment. Modern scuba equipment is manufactured well, but you must take care of it and have it serviced periodically to keep it functioning properly. Would you even consider going into outer space with life-support equipment that was not carefully maintained and serviced? We go into inner space when we dive. Our life-support equipment is just as vital to us as an astronaut's.

Equipment problems while diving do arise, but equipment failure is rarely the cause of an accident. Problems cause anxiety and stress. A diver experiencing stress from environmental factors and fear may panic from the added stress that a minor equipment problem might pose. Learning to manage yourself and your equipment helps you reduce stress, avoid panic, and prevent an accident. One of the most important rules of accident-free diving is to have good, familiar equipment that is in good condition and to use it properly.

Oxygen levels are managed by employing multiple sensors, which are accurate devices when maintained and calibrated. Current CCRs lack reliable carbon dioxide sensing devices. Carbon dioxide toxicity can be fatal, and a diver's symptoms are not a reliable indicator. This problem requires the usage time of the carbon dioxide absorbent to be monitored and the absorbent must be discarded when time limits (based on statistical tests) have been attained.

Because CCRs are expensive, must be used frequently, and require extensive maintenance, they are used by relatively few divers. But the advantage of greatly extended dive times appeals to many divers, so the popularity of rebreathers is increasing.

Summary

Diving involves a great deal of equipment. You are beginning to understand the equipment you need to have, how to select it, and how to care for it. You will learn more about equipment from your instructor, from retailers, from magazines, and from other divers. Get good equipment, and give it the best care possible. Diving is not enjoyable if you have constant equipment problems.

Application-of-Knowledge (AOK) Questions

1. Diving equipment is expensive. What are the best procedures to follow that will increase the life of your equipment?
2. What two concerns are most important when selecting diving equipment?
3. You were unaware that your regulator was free-flowing behind you at the surface at the end of a dive, and your tank lost all of its air. When you move your tank after the dive, you feel water sloshing inside. How was the water able to get into your cylinder?
4. You have weighted yourself to be weightless in water and are able to sink with your BC deflated just by exhaling. At the end of the dive you are unable to stop your ascent as you near the surface, even when you exhale completely. What is causing this undesirable change in buoyancy?
5. You are diving in a dry suit. You have added air to the suit to maintain a constant volume as you descend. You find a small anchor on the bottom and decide to salvage it. You pick up the anchor and carry it with you. To offset the weight of the anchor, you add more air to your dry suit. Why is this action a dangerous practice?
6. You and your buddy are practicing compass navigation by swimming a compass course. You place your compass next to your buddy's to compare headings and proceed along your planned course, but you missed your mark and had to surface. You find that you swam in the wrong direction. What caused this error?
7. What is the best way to remove salt, debris, or chlorine from you diving equipment?
8. A scuba regulator can have four or five hoses attached to the first stage. Can you think of the purposes of these hoses?
9. What are the best reasons for having good dive equipment and maintaining it properly?

School of barracuda in the Florida Keys

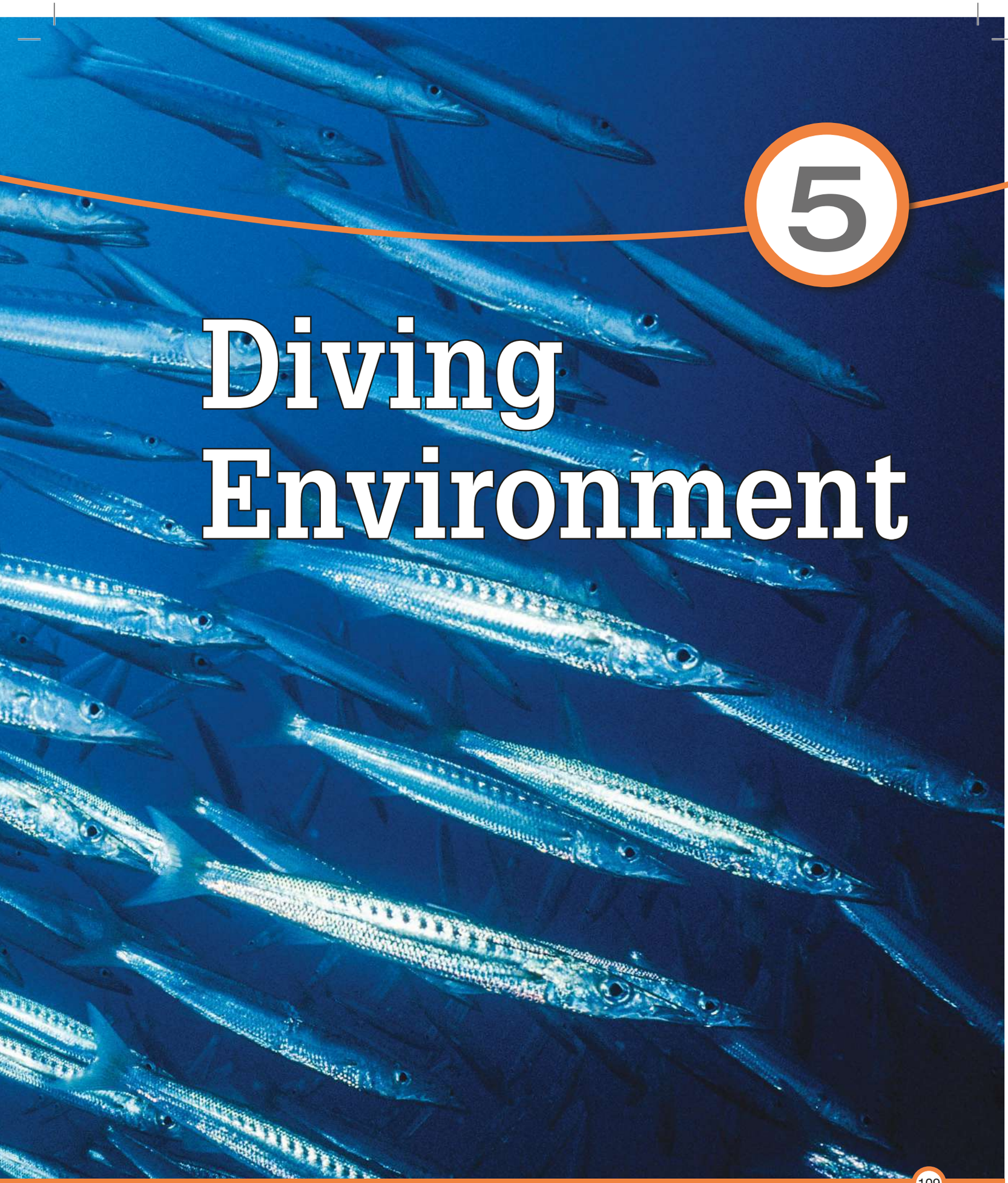

5 Diving Environment

Dive In and Discover

By the end of this chapter, you will be able to do the following:

- Describe the aquatic food cycle and explain how the process contributes to red tide.
- List five types of aquatic life injuries and three ways to avoid such injuries.
- List four types of pollutants and describe the effects they have on the underwater environment.
- List three actions that divers can take to help preserve aquatic life.
- List five ways to keep from damaging the environment while you are underwater.
- Contrast the general diving conditions for freshwater diving and saltwater diving.
- Describe the hazards of cavern diving, cave diving, and ice diving.
- Explain the cause of tides and their effects on diving activities.
- Explain the cause of waves and surf and their effects on diving activities.
- Explain the causes of currents and their effects on diving activities.
- Describe a rip current and explain how to escape from one.
- Define the terms *plankton*, *bloom*, *red tide*, *kelp*, *stipes*, *fronds*, *ciguatera*, *scombroid*, *tetrodotoxin*, *thermocline*, *overturn*, *reverse thermocline*, *upwelling*, *eddies*, *cavern*, *cave*, *sink*, *siphon*, *turbid*, *continental shelf*, *tsunami*, *stand*, *spring tide*, *neap tide*, *flood*, *ebb*, *slack water*, *fetch*, *swell*, *crest*, *trough*, *wave height*, *wavelength*, *wave period*, *wave train*, *surge*, *surf*, *gyres*, *downwelling*, *rip current*, *drift*, *set*, *drift dive*, and *trail line*.

Aquatic Biology

The plants and animals of the underwater world are wondrous and diverse. The millions of animals in the aquatic realm range from those that are microscopic to those weighing tons. To appreciate, respect, and enjoy aquatic life, you need to learn a bit about biology. This section familiarizes you with the flora and fauna of the aquatic realm. Aquatic life fits into three categories: life-forms that drift with the currents, those that swim freely and are able to move against the currents, and those that dwell on the bottom.

Drifters

The drifters are called **plankton**. Animals that drift are called zooplankton. Drifting plants are called phytoplankton. Plankton begins the food cycle in the waters (see figure 5.1). Small animals eat plankton, and larger animals eat the smaller animals. When the large animals die, their remains sink to the bottom and decompose. The decomposed material rises to the surface, where it becomes food for the plankton. Warm water and nutrients cause some plankton to multiply. Overpopulations of plankton, called **blooms**, can color the water, destroy underwater visibility, and form toxins in animals that feed by filtering water. Toxin from blooms makes clams and mussels unsafe for consumption during summer months in some areas. One type of red phytoplankton often creates blooms, known as **red tide**, in the seas. Diving conditions are poor in areas affected by plankton blooms.

Another type of aquatic plant, called algae, is an important part of the aquatic world. Plants use light to produce their own food and become food for animals. They convert water and carbon dioxide to oxygen and carbohydrates through the process of photosynthesis. Various types of algae are found in underwater areas where light is available. Most algae grow in shallow water where light is most abundant. Thick moss drapes objects in some freshwater areas. Turtle grass is a grasslike, green, tropical saltwater plant that provides a habitat for many forms of life. Some types of long, flowing algae in cold and shallow water, such as surf grass or eel grass, can cause you to trip if you try to walk in them. Slippery plants cover rocks and can cause you to slip and fall unless you move cautiously.

Giant algae, called **kelp**, produce long strands, called **stipes**, in which you can become entangled. You can learn how to avoid and how to deal with this problem. A rootlike structure, called a holdfast, anchors kelp to the bottom; numerous gas bladders, called floats, lift

Learning to dive gives you an opportunity to become more familiar with the aquatic environment, which covers over 70 percent of the surface of the earth. The underwater world is fascinating. This chapter introduces you to biological and physical conditions of the diving environment. You will learn the effect of people on the environment and the effect of the environment on people. You affect the underwater world more than you might imagine.

A juvenile octopus in California

the top parts toward the surface. What appear to be the leaves of kelp are **fronds**. Large areas of kelp are known as kelp beds, which have thick canopies that blanket the surface of the water (see figure 5.2). It is difficult to swim through a kelp canopy at the surface, but it is easy to swim between the clumps of stipes beneath the surface. Underwater navigational skills are important when diving in areas where kelp is dense. Kelp forests are popular diving areas because they contain great quantities of life.

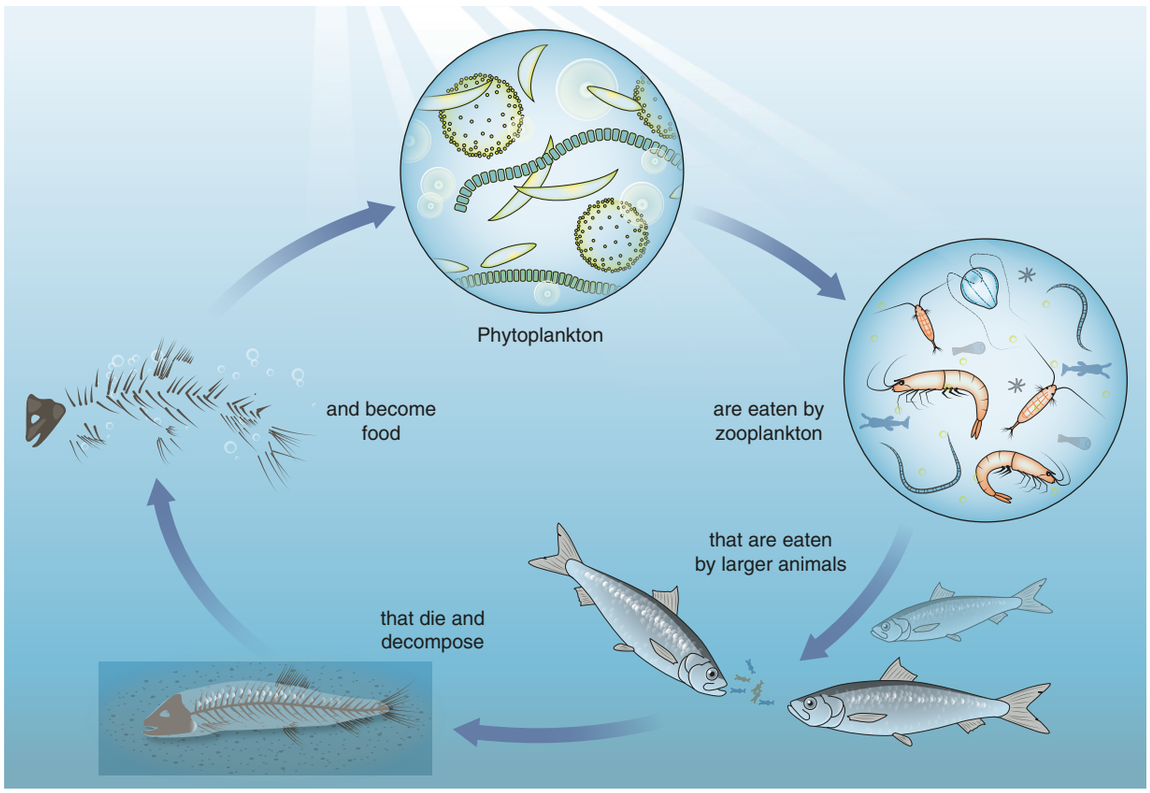

Figure 5.1 The underwater food cycle.

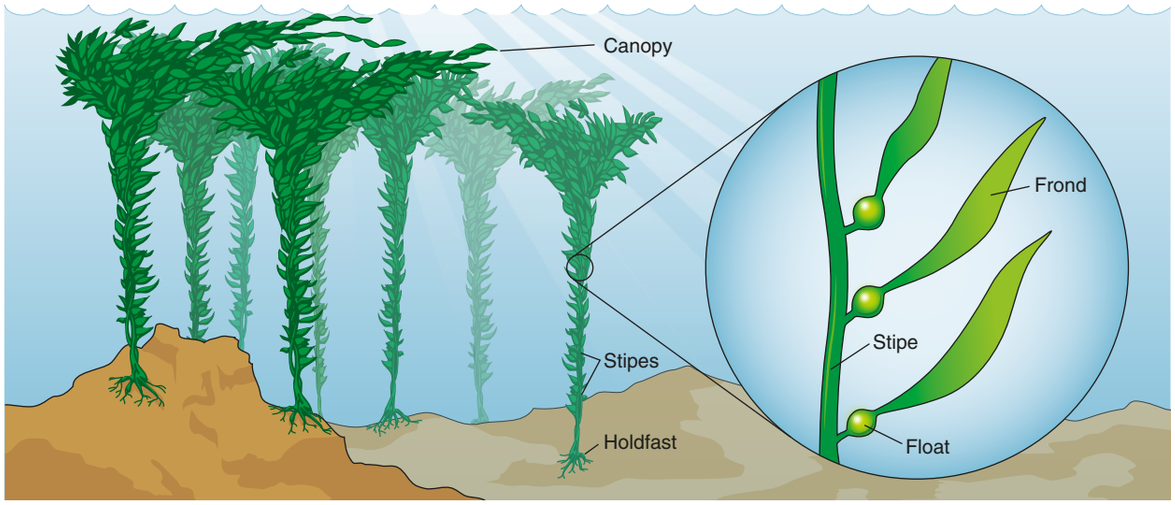

Figure 5.2 A typical kelp bed.

Swimmers

One of the rewards of diving is seeing fish. Fish can be found in nearly all the waters of the world. You cannot outswim even the slowest fish, so do not chase them. If you want to observe fish closely, you must blend into the environment. Fish will get closer to you than you could ever get to them.

Collecting for aquariums requires specialized knowledge and procedures. Most fish have an internal air bladder for buoyancy control. If a fish is taken to the surface too quickly, the air bladder expands and kills the animal. Avoid handling fish because the experience can be traumatic as well as physically damaging to the animal. And the fish's spine could injure you.

Eating some types of fish can harm you, as some fish are poisonous. Types of fish poisoning include **ciguatera**, **scombroid**, and **tetrodotoxin**. Ciguatera results from eating fish that consume a certain species of algae. Ciguatera poisoning causes gastrointestinal problems within 6 to 12 hours. Scombroid poisoning, which produces nausea and vomiting within an hour, can result if you eat fish that have not been kept chilled. Tetrodotoxin, the most serious fish poisoning, results from eating exotic fish such as blowfish or puffer fish (figure 5.3). Tetrodotoxin poisoning can cause death within minutes. Avoid eating large and unusual-looking fish. Check with local fishermen to determine which fish are safe for consumption.

Many large mammals—sea lions, seals, dolphins, whales, and manatees—inhabit the water. They are graceful, beautiful, and sometimes awesome in

Figure 5.3 Eating improperly prepared puffer fish can result in potentially fatal poisoning.

appearance. Viewing them is an exciting experience. Some are curious and may approach you. Aquatic mammals will usually not harm you in the water if you leave them alone. But sea lions and seals are defensive on land and might bite if you get too close.

Bottom Dwellers

Bottom dwellers include animals that are stationary, such as coral and sea fans, as well as animals that move about, such as crabs and lobsters. Living, stationary bottom dwellers are usually not included in fish and game regulations, and you should not take them. Do not take coral, sea fans, and animals that appear stationary, such as starfish. If you hunt crabs and lobsters for food, you should know how to determine the gender of the animals, how to catch them without harming them, and how to measure them for minimum size. Do not take females, particularly those bearing eggs. Some divers take only one claw from a crab to conserve the species. A crab can feed and defend itself with one claw and can regenerate a new claw to replace the one taken.

Hydroids

If you dive in saltwater, you should know about hydroids, a category of animal that includes bottom dwellers, such as coral, and swimmers, such as jellyfish (see figure 5.4). Some hydroids, such as the beautiful sea anemone, have a round, columnlike body with a mouth surrounded by tentacles. This is a polyp form of hydroid. Another type, called the colonial form, can assume many different

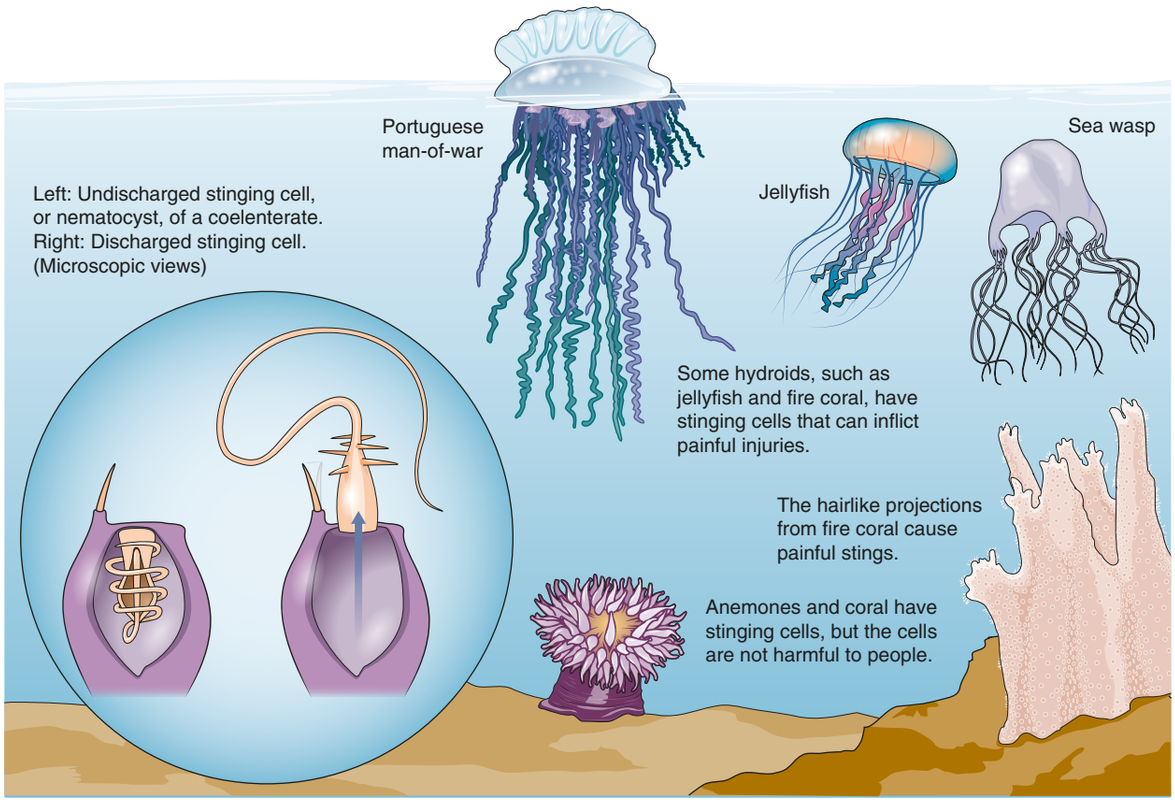

Figure 5.4 Several types of hydroids have stinging cells.

shapes. Colonies of hydroids can encrust a surface, and groups of colonies form jellyfish. Learn to recognize and avoid hydroids that can injure you.

Potentially Dangerous Aquatic Animals

Aquatic animals use a variety of mechanisms to obtain food and to defend themselves from attack. You can minimize your chances of injury by being familiar with the ways in which animals can inflict injury. Aquatic animals are rarely aggressive toward humans; they flee, hide, or stand still as you approach. If you do not touch, threaten, or provoke an animal, the animal is unlikely to injure you intentionally. But remember, aquatic animals are wild animals (figure 5.5). If you feed them, and they bite you in the process, you should not blame the animals. Table 5.1 lists types of injuries that you could sustain from aquatic animals and recommended first aid.

Animals That Cause Abrasions or Cuts

Many animals, such as some types of coral and barnacles, have sharp, hard structures that easily cut flesh. Such cuts can be painful, are slow to heal, and can get infected. Avoid contact with reefs and rocks covered with sharp animals. Some fish have knifelike protrusions on their gill plates or at the base of their tails. They defend themselves by flailing rapidly back and forth and slashing anything near them.

Animals That Puncture or Lacerate

Sea urchins are the porcupines of the seas. Cold-water urchins have short, thick spines, while warm-water urchins have spines that are long and thin. The spines of all urchins can penetrate an exposure suit if you bump into them. The spines break off in your flesh and produce a painful, red, swollen wound. The spines can be difficult to remove, so if you are unable to avoid a sea urchin injury, see a physician to have the spines removed.

Some fish have a row of long, sharp spines along their back. The scorpionfish has hollow dorsal spines that have venom-filled sacs at the base. If you are punctured by the spines and compress the venom sacs, you inject yourself with a toxin. The toxin of some species, such as the stonefish, can cause serious symp-

Figure 5.5 Potentially dangerous aquatic animals: (*a*) White tip shark; (*b*) cone snail; (*c*) sea snake; and (*d*) jellyfish with jacks.
Scott Frier: www.wildlifephotoworkshops.com

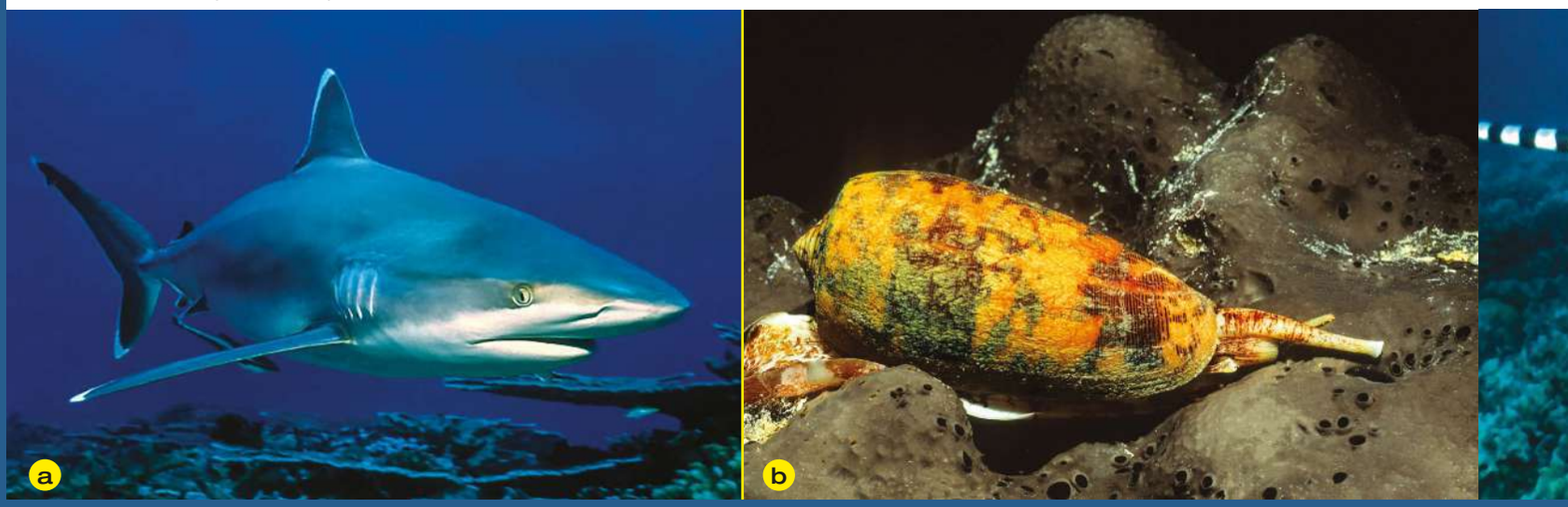

toms. Some fish, such as the lionfish (also called the turkey fish or zebra fish), have spines in other fins. The toxin of a lionfish produces serious symptoms in humans. Freshwater catfish also have venomous spines.

A shellfish called the cone shell uses venom to kill animals for food. The venom is highly toxic, and these animals can inject their venom into humans. Do not handle conical shells in tropical waters.

Rays are bottom dwellers that are round and flat. Some rays have a sharp, serrated barb at the base of the tail. The rays blend into the bottom. When someone disturbs a ray, it defends itself by arching its back and thrusting its barb into the attacker. A sheath, which often remains in the laceration wound, covers the barb and contains a toxin. Stingray wounds in the ankles of divers, waders, and swimmers are common in some areas. Avoid stingray wounds by shuffling your feet along the bottom as you walk. The torpedo ray, an animal found on the West Coast of the United States, can generate electricity. This ray can stun a diver, so you need to be able to recognize it. Avoid contact if you encounter one.

Animals That Bite

Use discretion to avoid being bitten by aquatic animals. One large freshwater fish that can bite is the gar. Turtles can inflict serious wounds. Muskrats may attack in defense, and moray eels in the ocean can deliver a serious bite. Alligators also have the potential to inflict an injury but have not been known to hurt divers.

Sharks can bite, but attacks on scuba divers are nearly nonexistent. Hollywood has exaggerated the danger of sharks; only a few species of sharks are aggressive, and it is rare to encounter these types in the waters visited by divers. Most divers would be delighted to see a shark because these animals usually retreat from an area frequented by scuba divers.

Retreat from snakes if you encounter them. Do not handle snakes; they can be extremely venomous. A sea snake's bite can produce a life-threatening emergency. Some freshwater snakes, such as the cottonmouth, are also venomous. The blue-ringed octopus, while only the size of a golf ball, is one of the most toxic creatures on earth. Don't allow curiosity to endanger your life.

Animals That Sting

Many aquatic animals can sting. Learn to identify and avoid contact with jellyfish, featherlike or whiplike hydroids, certain worms, and even some sponges.

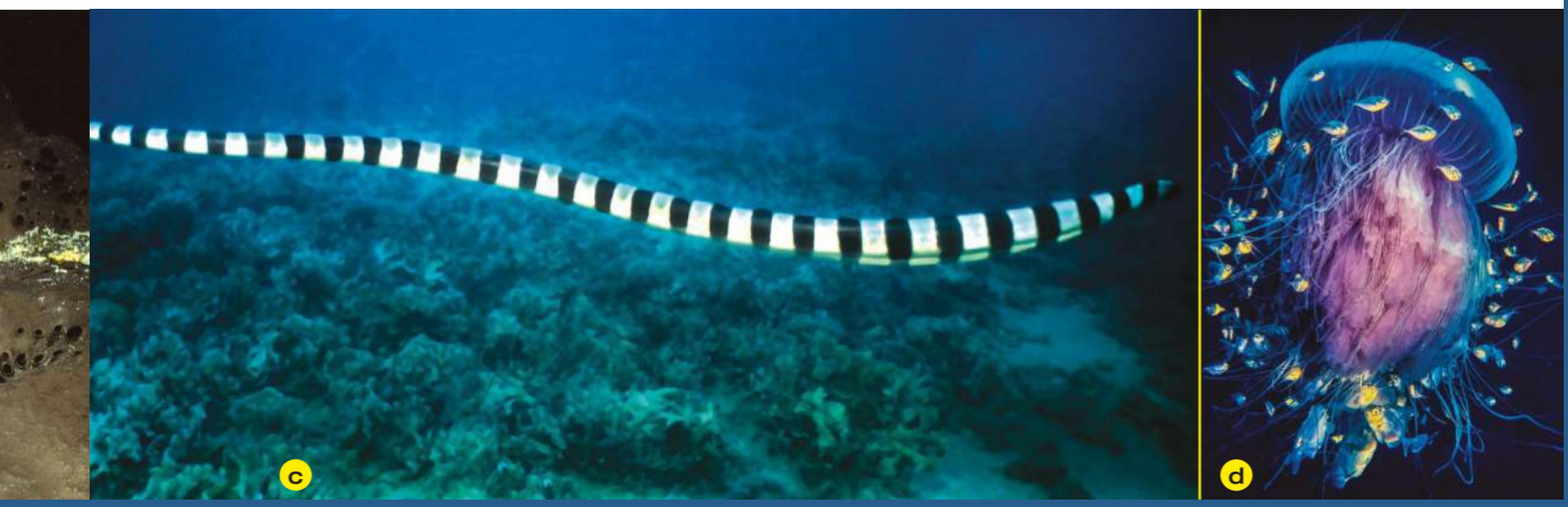

Table 5.1 Potentially Dangerous Aquatic Animals

Injury	First aid
Abrasions or cuts	
Barnacle Coral	Scrub and disinfect the wound.
Punctures or Lacerations	
Cone shell Sea urchin Venomous fish (e.g., scorpionfish, lionfish) Stingray	Soak the injured area in hot water. For a cone shell sting, apply a venous tourniquet and immobilize the injured area. It is advisable to seek medical attention for injuries caused by the spines of the scorpionfish family.
Bites	
Barracuda Blue-ringed octopus Moray eel Sea snake Shark Turtle, muskrat, alligator	Stop the bleeding; clean and disinfect the wound. For a blue-ringed octopus bite, apply pressure to the wound and seek immediate medical attention—it may be necessary to administer rescue breathing. For a sea snake bite, apply pressure, immobilize the area, and get prompt medical attention.
Stings	
Bristle worm Fire coral Fire sponge Jellyfish	Soak the injured area in vinegar.

Seek medical attention as needed for any injury.

Hydroids have tiny stinging cells, called nematocysts, that they use to kill food. The sting of some animals, such as encrusting fire coral, is merely annoying; but the sting of others, such as the Portuguese man-of-war or the box jellyfish, can cause a medical emergency. A tropical marine worm, called a bristle worm, has tufted, silky bristles along the sides of its body. The bristles, which are fine and brittle, penetrate the skin easily, are difficult to remove, and cause a burning sensation. Do not touch or handle these worms. You should wear an exposure suit at all times while diving in the ocean. The suit provides protection from stings, but be careful when you remove it. The stinging cells of marine animals remain active even if they are not on the animal. Parts of jellyfish and other animals that are on your equipment can sting if they come into contact with your skin when you remove the equipment. Avoid contact with any unusual debris clinging to your equipment after diving in salt water because it could be part of a jellyfish. Some jellyfish come to the surface at night in tropical waters. Dangerous stinging animals are seasonal in some regions. Check with local divers to find out what to avoid and when to be on the alert.

Avoiding Danger

Many aquatic animals are potentially dangerous. Tropical waters have the most dangerous animals. Serious injuries to divers are not common because divers avoid animals that can hurt them. You should respect animals, but do not be overly concerned about them. Do not panic or flee just because you see a potentially dangerous animal; simply avoid contact with it. Learn to recognize the dangerous animals in an area, know where to look for them, be alert for them, and keep clear of them. Move slowly and look carefully. Shuffle your feet when wading. Don't touch anything unless you know it is safe to touch it. Neutral buoyancy—the ability to hover—is an excellent defense against aquatic injuries.

The biology of the underwater world is so interesting that many divers study it as a hobby. It is fun to learn about animals and plants and then observe them in their natural habitats. Some divers, fascinated with aquatic life, pursue biology as a career.

Conservation and Preservation

Life in the waters is beautiful and precious, but pollution from people jeopardizes life in the lakes and seas. Divers are physically ruining beautiful reefs. Unless we take action immediately, there is a real danger that many areas of the underwater world will become barren and lifeless. We must preserve and conserve the resources of our waters.

Because the waters are so vast, people often take them for granted. Lakes and seas seem too big to be harmed, but that is not the case. The underwater environment is fragile; the balance of nature can be upset more easily than many people realize. We think of the seas as being great and powerful because we see big waves and pounding surf. People who do not see beneath the surface may not realize the delicate nature of the animals that inhabit the waters. Some living things in the seas grow slowly, only a fraction of an inch (or centimeter) per year.

As a diver, you can help reduce damage to the underwater world. Every diver should take two actions: (1) Do everything possible personally to preserve the diving environment, and (2) do everything possible to educate others and help them learn to preserve the aquatic environment. You will get closer to the aquatic environment than most of your friends. You will see firsthand the effects of pollution, litter, and exploitation. You will also see the beauty and abundance that are possible when the environment is clean and unmolested. Your influence in society can make a difference. If you do not become part of the solution to problems in the aquatic environment, you are part of the problem.

Pollution

Stemming pollution is one of the greatest challenges facing the world. Humans are incredible polluters of the environment; we have polluted the air, the land, and the waters. People dump billions of gallons and pounds of waste into water every day. People seem to think that when something is out of sight, it is no longer a problem. This is not true of pollutants. Sewage, industrial waste, garbage, and sediment have killed and are continuing to kill many underwater environments.

If we stopped all pollution today, the aquatic environment would continue to suffer for decades from the waste materials that are already in the water.

Runoff from land that enters rivers and streams and flows into the ocean also causes pollution. Chemicals used in agriculture, on lawns, and in gardens cause death and destruction in the aquatic environment. Sediment from construction and drilling finds its way into water. The sediment blocks out life-giving sunlight and smothers bottom-dwelling creatures.

Lakes and oceans have been viewed as bottomless toilets for waste disposal. There are two problems with this narrow view: (1) Animals and plants live in the water and are killed by the pollution; (2) the water cannot be replaced because large bodies of water cannot be flushed.

Diver Impact

As a diver, you can harm the environment in several ways. You can remove living things from the environment, you can smash and kill living things while moving about underwater, and you can stir up clouds of silt on the bottom. The silt can choke and kill some organisms. Good intentions can also cause problems; handling and feeding animals can kill them.

You can be an effective predator underwater. Many animals are available for you to hunt and take. A few callous individuals kill things for sport, but responsible people take only what they will eat. Although the impact of divers is of little significance compared with commercial fishing, you should keep in mind that you do have an effect. If you spearfish on a reef, the fish in the area will soon be unapproachable by divers. If you want to be a predator, you must do so in a responsible manner. Conservation and preservation of aquatic resources are important topics that are discussed later in this section.

It is difficult to resist touching animals underwater, but you should refrain until you know what animals may be touched and how to touch them without harming them. Many animals are delicate; rough handling will kill them. Predators will eat animals that a diver removes from a protected area for viewing. Some animals, including fish, have a protective coating of mucus. If the coating is removed by handling, the animal can develop an infection and die. The stress of being handled by a gigantic, bubble-blowing monster may be more than some aquatic animals can take. Sea turtles may abort their eggs after being harassed by a playful diver. Do not be guilty of killing things for your amusement. After

SCUBA WISE

There are similarities between the environments above and below water. Temperature, weather, plants, and animals change when you travel from region to region on land. We enjoy the variety of mountain, plain, forest, and desert regions. Similarly, the world beneath the water varies from region to region. The enjoyment that the various underwater environments provide is a major reason why dive travel is so popular.

When you go to places on land that are rugged, remote, and hazardous, you need special training and equipment. There are dangers everywhere you go. Careful planning and preparation help minimize risks and are essential. A guide makes exploration safer and more enjoyable. When you go diving in places that are rugged, remote, and hazardous, the same requirements apply. Enjoy the wonderful differences beneath the waters—coral reefs, vertical walls, wrecks, and so forth—but always be properly prepared and equipped and always obtain an orientation to a new area.

you learn how, you can get extremely close to animals underwater. You can interact with them and enjoy them without handling them.

Feeding animals underwater was a popular activity until environmentalists showed that this is harmful (not to mention dangerous). Several potential problems are associated with feeding aquatic animals. Unnatural food given to animals can interfere with their digestion. The animals may become dependent on the food fed to them by divers and may not be able to forage if the food supply is discontinued. Food can make animals overcome their natural fear of divers. When an animal that would normally take cover at the sight of a diver becomes accustomed to divers providing food, the animal will readily approach a diver who is a hunter.

A diver underwater has the potential to be a bull in a china shop. Overweighted divers plow along the bottom with their fins pointed downward, stirring up great clouds of silt. Buoyancy difficulties cause divers to hold on to and crash into reefs and other living things. Divers who rest on the bottom crush plants and animals without realizing the damage. Divers who swim too close to a reef often kick animals to death.

Prevention of Diver Impact

One of the most important reasons for developing good buoyancy control skills is to prevent damage to the underwater environment. An environmentally responsible diver is properly weighted and in control of buoyancy at all times. Make your diving no-contact diving. You should be able to hover above a reef, move your mask within inches of an animal, and view it without touching anything but water and without stirring up silt. Learn to use your hands to scull into position while your fins remain still. Sculling is positioning achieved with short movements of the hands (not the arms). Buoyancy control and sculling are excellent skills for reducing diver impact. Figure 5.6 shows a diver sculling.

Another way to protect the environment is to keep your equipment secured close to you so that it does not dangle and drag. Equipment that drags along the bottom for an entire dive can do a lot of damage. Streamline your equipment. Moving slowly underwater conserves energy and air and makes you less likely to come into contact with animals. This helps you avoid injuring both yourself and animals.

If you must hold on to something underwater or push yourself away from something, you should look before you touch, and you should avoid touching anything that is living. If you must settle onto the bottom for some reason, select an area where there is no visible life. If you are weightless, keep in mind that one finger may be all you need to provide the leverage to move. Leave no evidence that you ever visited an underwater area.

Figure 5.6 Sculling helps minimize the diver's impact on the aquatic environment.

Conservation

Many animals that were once plentiful are nearly extinct because of a lack of conservation. Buffalo and passenger pigeons are good examples of this happening on land. In some areas, this is happening to aquatic animals. Bait fish that once swarmed in enormous schools in some areas no longer exist. With no food to eat, larger predatory fish no longer frequent the areas.

Fish and game regulations are designed to conserve natural resources. Rules regarding sizes, seasons, limits, and the means by which game may be taken have been established to help ensure an ongoing supply of a resource. Obey fish and game regulations, and encourage others to abide by them as well. In most areas you need a license to harvest fish and seafood. There are written regulations you need to be familiar with. Size, sex, limit, and season laws can impose a stiff fine if you fail to adhere to the regulations. Check with your local Department of Fish and Game. The rules can benefit everyone in the long run.

If you take living things from the water, do it in a responsible manner. Avoid taking animals from areas that are popular dive sites. Limit hunting and collecting to remote areas where the impact of divers is much less. Take only what you need, not what you can get or what you are allowed. The two types of hunters most harmful to the environment are the quantity hunter and the trophy hunter. The quantity hunter seeks to take as much and as many of everything as possible to build an image as a mighty hunter. The trophy hunter seeks the largest animals, which destroys the fittest of the breeding stock. If you kill an animal, you have a duty to know how large one must be to breed and what the maximum size of the animal is. You should take animals that have had the opportunity to reproduce, but do not take the largest ones. It is difficult at times to be selective, but you should attempt to conserve life.

Nature left undisturbed maintains a balance of life. Animals are both predator and prey. An animal eats another animal and is in turn eaten by others. If there are too many predators, their population diminishes because of an inadequate food supply. If there is a temporary overpopulation of prey, the number of predators increases. People upset nature's balance. We are the most reckless predators of all. We disturb the food cycle with pollution, hunting, fishing, boating, and every other way that we affect the aquatic environment. It takes nature longer to recover from our impact than from any natural disaster. People must lessen their effect on the environment to reduce interference with the natural management of the life within the waters.

Preservation

Preservation is everyone's business, but it is more your business now than it was before you became a diver. You probably know more than your friends and neighbors about the aquatic environment. You must be an ambassador for preservation. You must educate and motivate people to help preserve the diving environment.

What people do above water affects life beneath the surface. What people put down their drains and toilets winds up in the aquatic environment, as do the chemicals they use on their lawns and gardens. Litter in and around the water kills animals, birds, fish, and shellfish Think twice about the products you buy, how you use what you purchase, and how you dispose of your wastes. Where will the toxic chemicals and waste you use end up? Be environmentally conscien-

tious; then teach others to be so, too. A simple act, such as the use of detergents without phosphates, can make a difference. Phosphates are powerful nutrients that upset the balance of nature when dumped into the aquatic environment.

Be involved in your community. Be concerned about issues such as sewage treatment, toxic wastes, and construction, which can be extremely harmful. Waste products from manufacturing are also harmful. Help others understand the seriousness of pollution.

Become more informed and keep informed. Join groups that are working to preserve the environment. Organizations such as the Ocean Conservancy, the Reef Environmental Education Foundation (REEF), and the Oceanic Society merit your membership. These groups provide up-to-date information and details on how you can help. Various organizations sponsor underwater cleanups from time to time; these are enjoyable and worthwhile. (For a list of environmental organizations you can join, see Appendix A.)

There are actions you can and should take when you go diving. Dispose of trash properly, and encourage others to do the same. If you have a boat for diving, anchor it away from reefs to avoid reef damage caused by the anchor and chain. Retrieve trash you find in the environment, especially plastic. Collect plastic, monofilament, lead, and stainless steel leaders. Not all trash is bad. Bottles and cans provide homes for animals. Report unlawful dumping and lost or discarded fishing nets and traps, which continue to catch and kill after abandonment.

Aquatic Conditions

The particular state of the environment and the physical situations in which divers find themselves are referred to as the aquatic conditions. The conditions of concern include the temperature, visibility, and the degree of movement of the water. You need to be familiar with aquatic conditions in general and local aquatic conditions specifically.

The Big Picture

Many factors affect the seas. The sun bears down on the earth at the equator more directly than on other parts of the earth. The climate and waters near the equator are warm; the greater the distance from the equator, the colder the water. The difference in water temperatures in different parts of the world causes the movement of air and water by convection currents. Winds move from areas of high pressure to areas of low pressure. Winds and the turning of the earth move water. On a global scale, water currents move in a clockwise direction in the northern hemisphere of the earth and in a counterclockwise direction in the southern hemisphere. Weather generally moves from west to east. The gravitational attraction between the earth and other planetary bodies produces changes in the water level called tides. Storms at sea produce energy in the form of waves that travel thousands of miles before giving up their energy in the form of surf.

Seasonal changes, winds, and storms affect inland bodies of water. Water from snow and rain in the mountains and hills flows into streams, rivers, and lakes. Water seeping into the ground resurfaces in quarries and springs.

Weather, seasons, geography, and other factors affect both surface and underwater diving conditions. The water can be rough at the surface but calm at depth. There may be a current at the surface but none on the bottom. The visibility may

be good at the surface but poor on the bottom, or vice-versa. The temperature is usually warmer at the surface than it is at depth. You should become familiar with both surface and underwater conditions and their effect on your diving.

Perhaps the most important fact to remember about diving conditions is that they vary. Thus, an environmental orientation is important when you dive in a new region. The diving conditions usually dictate the way you dive. What works well in one area may be totally ineffective for another. An understanding of the effects of diving conditions in an area is important. You should know what conditions to expect, how the conditions affect your approach to diving, and how to manage the effects of the conditions. For example, in one area, it may be safe to wade into the water and put on your fins; however, in another area, you may need to put your fins on before you wade in.

General Freshwater Diving Conditions

Freshwater is 2.5 percent less dense than saltwater, so you are less buoyant in freshwater than in saltwater. The density of water varies with temperature. Freshwater is most dense at a temperature of 39.2 °F (4 °C)—the usual temperature and density of the water at depths greater than 60 feet (18 m) in freshwater lakes and quarries. The water becomes lighter when it is either warmer or cooler than this. Freshwater often forms layers: a layer of warmer, lighter water on top of a layer of colder, denser water. The change from the warm water to the colder water, called a **thermocline**, is abrupt. When the water is calm, the thermocline appears to have wisps of smoke on top of it when viewed from above. The refraction of light is different when the density of the water is different, so there is a slight visual blurring at the interface of a thermocline. When you dive in freshwater, you must insulate yourself for the water temperature below the thermocline. Although the surface of a lake may be warm and sunny, the water at the bottom may be close to freezing.

The best season for diving in a freshwater lake or quarry depends on the body of water. Spring and fall are often good times because the water temperature is the same from surface to bottom, there is oxygen for fish at all depths, populations of plankton are low, and visibility is usually good.

In later spring, the sun warms the surface water, and the lack of wind keeps the surface water from mixing with the colder water at depth. The water stratifies, and a thermocline forms that remains until fall. The layer of water below the thermocline stagnates in most lakes during the summer. Decaying matter depletes oxygen and creates toxins. Fish seek refuge above the thermocline. Sunlight and warm water often lead to plankton blooms, which are aggravated by pollution.

In the fall, the surface water in lakes and quarries cools until it equals the temperature of the water at depth. Colder water and reduced sunlight stifle plankton growth, and visibility improves. Winds cause water circulation. The movement of water from the top to a depth of about 60 feet, called an **overturn**, carries oxygenated water to the bottom and leads to what is called spring diving conditions. Visibility tends to be poor during an overturn.

In the winter, the surface water in a lake or quarry is colder than the water at the bottom. A **reverse thermocline** exists and remains until the temperature of the surface water warms to that of the water at the bottom. This allows the spring overturn to begin. Water becomes 10 percent lighter when it freezes. If ice sank when it froze, bodies of water could become solid ice from top to bottom.

Ice that forms over water insulates the water beneath it. Figure 5.7 shows the annual cycle of freshwater lakes and the concept of thermoclines.

When a strong wind blows along a shore for a sustained period over a body of water, the surface water is pushed away from shore. This water is replaced by colder water flowing up from the depths. This is called an **upwelling**, which can also occur in the ocean. If the wind conditions persist, the water temperature can become constant throughout the water column, even in the summer. An upwelling carries nutrients from the depths into shallow water and brings animals into the area. After an upwelling, a bloom can occur because of the increase in nutrients in warm, shallow water.

Specific Freshwater Diving Conditions

People can have good diving experiences in many kinds of freshwater environments. Springs, low-altitude lakes, and quarries are good sites for divers who do

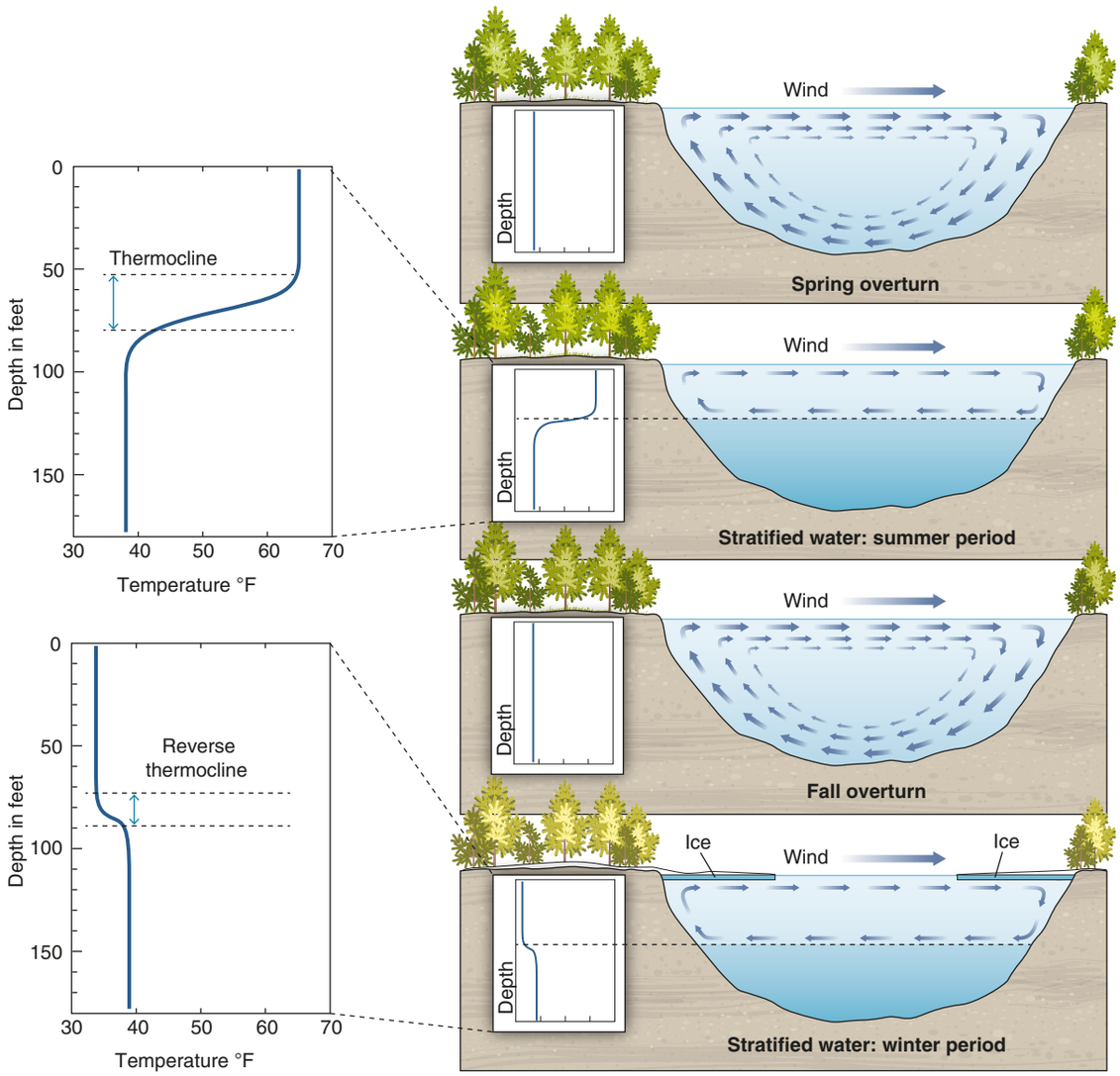

Figure 5.7 Thermocline, reverse thermocline, and the annual cycle of overturn in lakes.

not live near a coast. Other freshwater environments—such as rivers, caverns and caves, high-altitude and frozen lakes, and areas with submerged wreckage—can be dangerous. You should complete specialty training courses before diving in any hazardous area.

You may encounter strong currents in rivers (see figure 5.8). The currents are strongest at the surface and on the outside of bends. Countercurrents and swirling currents, called **eddies**, are common, and entanglements are likely in many rivers. Water flowing through a porous obstruction, such as a pile of logs, can pin a diver against the debris. Currents can undercut the bank of a river, so direct access to the surface can be impossible at times. Rivers are especially prone to seasonal changes and are often unpredictable.

Freshwater springs can provide beautiful diving environments. The flow of clean, calm water, which is usually at a moderate temperature (65 to 78 °F, or 18 to 26 °C), provides excellent visibility. The water often flows through underground limestone cave systems that extend for thousands of feet. Diving in open-water basins in a spring is appropriate for certified divers, but a diver should not enter any overhead area where a direct, vertical ascent to the surface is not possible. A large, roomlike opening where light from the surface can be seen is called a **cavern**. Areas that extend farther than a cavern where no surface light can be seen are called **caves** (see figure 5.9). Divers must complete specialty courses and must meet several requirements to dive in caverns and caves. Entering such an environment—even for a short distance—without the required training and equipment can be fatal. No amount of diving experience without cave diving training qualifies a diver to enter a cave. It is easy to

Figure 5.8 Because many rivers have strong currents and pose other unique diving challenges, specialty training is advised.
Berenika/iStock/Getty Images

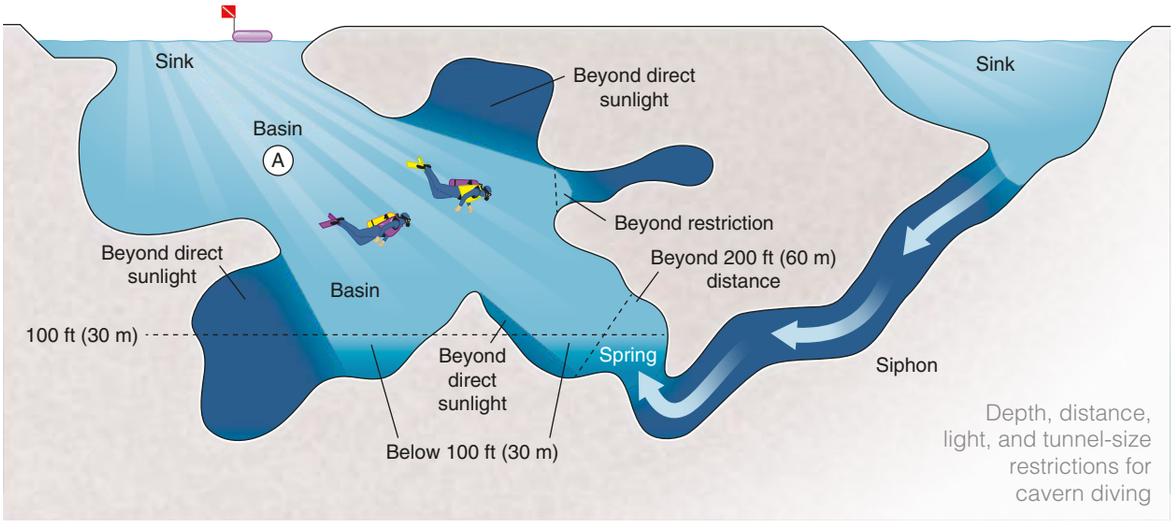

Figure 5.9 Dark blue areas are caves; light blue areas are caverns.

become disoriented, to stir up thick clouds of silt that instantly reduce visibility to nothing, and to panic and drown. You may dive in spring basins, but stay out of caverns and caves unless you meet the requirements to dive in them. Appendix A contains contact information for organizations that offer training in cavern and cave diving.

Sometimes the earth collapses into an underground cave system and forms a **sink**. The water flowing into the sink forms a basin. An opening called a **siphon** channels water from a sink back into the system. Diving in siphons can be dangerous. The amount of water moving through a system depends on the amount of rainfall in the area. Under extreme conditions, the normal flow of water can reverse.

People dig pits to excavate sand, gravel, and stone. At some depth, they encounter the water table, and the pit floods, forming a quarry or sand pit. At some of these sites, you can see submerged construction equipment that was abandoned when the pit flooded. Sand pits and quarries tend to have fair to good visibility, although the disturbance of silt that has settled to the bottom can ruin the visibility quickly. Gravel quarries usually have more silt than other types of quarries, so the water may be turbid.

Lakes can be excellent dive sites. There are freshwater lakes at altitudes above 10,000 feet (3,048 m). Diving at altitudes above 1,000 feet (305 m) is a specialty called altitude diving. This type of diving presents many hazards because the rate of pressure change is greater when you descend into water at an atmospheric pressure that is less than the pressure at sea level. You must follow special procedures to avoid decompression sickness. Problems can also result from the thinner air, which provides less oxygen with each breath taken above water.

When ice forms over water in the winter, diving becomes hazardous. Perils of ice diving include hypothermia, the freezing of regulator and buoyancy control valves, and getting lost beneath the ice. Ice diving can be beautiful and adventurous, but it is dangerous when attempted without the proper training, equipment, and procedures.

Potential hazards when diving in freshwater include body heat loss, submerged trees, wire, fishing line, hooks and lures, log jams, debris, currents, rapids, whirlpools, poor visibility, silt, and overhead restrictions. You can minimize these hazards with training, experience, equipment, correct diving techniques, area orientations, and good judgment.

General Saltwater Diving Conditions

The earth contains five oceans. Seas, gulfs, and bays are smaller sections of the oceans. The land beneath the oceans is not flat. The **continental shelf** is an underwater area that extends from land and slopes gradually to a depth of about 600 feet (183 m). Beyond the continental shelf are great underwater canyons and mountains, as well as hills, valleys, and great plains. In some areas, the tips of mountains and volcanoes extend above the surface of the ocean from great depths, forming islands. In other areas, deep canyons sever the continental slope, creating deep-water conditions close to shore.

The saltwater environment varies greatly from one part of the earth to another. The clear waters of the Caribbean, with temperatures that can exceed 85 °F (29 °C), feature coral reefs covered with beautiful, lush animals that look like plants. Schools of colorful tropical fish abound. Temperate waters, with temperatures ranging from 55 to 70 °F (13 to 21 °C), may have great forests of kelp, which contain more life per volume than a tropical rain forest. Cold ocean waters in northern latitudes are nutrient rich and teeming with life. Beautiful sights and wonders can be experienced in all the oceans of our planet.

Oceans are always in motion. Tides, winds, and currents cause water movement. An underwater earthquake can move water by creating gigantic seismic waves known as a **tsunami**. Although these giant waves are also called tidal waves, the tide has nothing to do with them. Tsunamis can cause great destruction, but they are rare and can be forecasted.

The energy source for movement of ocean water may be local or may originate thousands of miles away. You need to understand what causes water to move, how it moves, and how to dive in ocean water that is in motion.

Tides

The gravitational attraction between the sun, the moon and the earth pulls water toward these heavenly bodies. The resulting increase in water depth is a high tide. Water pulled away from the sides of the earth in the process produces a decrease in water depth called low tide. A high tide forms on the side of the earth opposite the moon because the attraction of the moon is least at that point and because of the centrifugal effect created by the rotation of the earth. At any given time, two areas on the earth are experiencing high tides and two areas are experiencing low tides. During a brief period of time called a **stand**, the tide neither rises nor falls; four tidal stands can occur per day—two high and two low. Geographical formations interfere with or enhance the rising and falling of the water in some areas.

The sun also affects the tides, but only about half as much as the moon because the moon is much closer to the earth. When the moon and the sun are aligned with the earth, which occurs twice monthly (during the new and full moons), tides are highest and are called **spring tides** (see figure 5.10). When the moon and sun are at right angles to each other in relation to the earth, tides are lowest and are called **neap tides**.

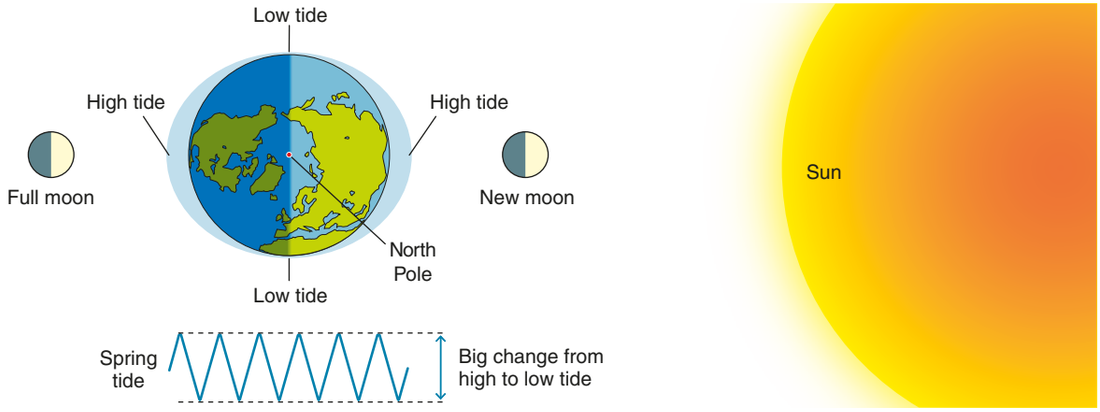

Figure 5.10 The effect of the sun and moon on tidal patterns.

The moon rotates around the earth in the same direction that the earth spins; therefore, the duration of a lunar day (moonrise to moonrise) is nearly 25 hours. This causes tides to occur at different times each day and explains why the heights of subsequent tides vary.

During a tidal change, water flows from one area to another. The water movements caused by the tides are called tidal currents, and they vary in velocity between tides. The flow of water into an area because of high tide is called a **flood**. The flow of water away from an area because of low tide is called an **ebb**. Between flood and ebb, a period of time occurs when the water has minimal movement. The term for this period is **slack water**. Because of geographical features of the earth, water cannot move instantly in many areas. There is usually a delay between the predicted time of a tide and the time when the water is slack.

Tidal changes affect diving activities and operations. Large differences in the water level between tides affect visibility and other conditions. Water height also affects entries and exits, shoal areas when boating, the mooring of boats, and the loading and unloading of vessels. Divers should know how high the tide will be. Tidal changes are small (less than 2 feet, or 0.6 m) in some areas or on some days, but they are large (more than 6 feet, or 1.8 m) in other areas or on other days. The greater the difference between high and low tide, the greater the effect of the tides.

Because the earth, moon, and sun follow set courses, the tides are predictable. Tides vary because the distance between planetary bodies and their positions relative to one another change constantly. Despite all the variables, scientists can predict with accuracy when the tides will occur and the heights of the tides. The U.S. government publishes tide tables, and national weather channels broadcast tidal information continuously. Local correction tables and current tables can provide more precise information. Tide tables refer to the height of the tide in relation to an average of the low tides in an area (see figure 5.11). Wind and barometric pressure affect the height of tides, aiding or opposing the movement of tidal waters.

November 2015

	Low tide				High tide			
	a.m.	ht.	p.m.	ht.	a.m.	ht.	p.m.	ht.
	Sunrise 6:19		-PST-		Sunset 5:08			
1 Su	1:35	1.0	2:35	0.5	7:49	6.4	8:46	4.7
2 M	2:06	1.3	3:17	0.9	8:21	6.7	9:36	4.4
3 Tu	2:38	1.7	4:06	1.0	8:56	8.8	10:35	4.0
4 W	3:14	2.1	5:02	0.9	9:33	6.7	11:44	3.7
5 Th	3:50	2.6	6:03	0.7	10:19	6.5	—	—
	Sunrise 6:24		-PST-		Sunset 5:03			
6 F	4:42	3.0	7:17	0.4	1:14	3.5	(11:15	6.0)
7 Sa	6:00	3.3	8:36	0.2	2:57	3.8	12:28	5.4
8 Su	8:23	3.3	9:43	0.1	4:15	4.0	2:03	
9 M	10:12	2.8	10:42	0.1	5:01	4.4	3:39	
10 Tu	11:22	2.1	11:27	0.3	5:37	4.9	4:55	
	Sunrise 6:28		-PST-		Sunset 5:00			
11 W	—	—	12:14	1.4	6:09	5.3	5:	
12 Th	12:06	0.6	12:59	0.8	6:34	5.6	6	
13 F	12:38	0.9	1:38	0.3	6:59	5.9		
14 Sa	1:07	1.3	2:13	0.1	7:24	8.1		
15 Su	1:31	1.6	2:45	0.3	7:47	6.2		
	Sunrise 6:33		-PST-		Sunset			
16 M	1:56	2.0	3:21	0.4	8:12			
17 Tu	2:21	2.2	3:53	0.3				
18 W	2:40	2.5	4:33	0.				
19 Th	3:03	2.7	5:17					
20 F	3:21	3.0						

Figure 5.11 Tide table.

Generally, it is best to dive at high tide, but you may need to time your diving to coincide with slack water in areas with strong tidal currents. The timing may not be important in areas where tidal changes are small. Consider local knowledge about the effects of tides when planning your dives.

Waves and Surf

Ripples form when wind blows across water. You can see this in a puddle on a windy day. When ripples form on a large body of water and the wind continues to blow, the sides of the ripples form a surface against which the wind can push to set the water in motion. The longer and harder the wind blows in a constant direction, the larger the waves that form. As waves move away from the area in which they were created—an area called the **fetch**—the tops of the waves become rounded, undulating forms called **swells**. The water within a swell moves in a circular motion but has little forward motion. The effect is similar to the transmission of a waveform along a rope that is tossed up and down at one end. Waves of energy travel along the rope, but the rope itself does not move forward.

Swells are energy forms that can travel thousands of miles through water and still contain a great deal of energy. The top of a wave is called the **crest**. The bottom is called the **trough**. The distance from the crest to the trough is the **wave height**. The distance between waves is a **wavelength**. The time it takes two waves to pass a given point is a **wave period**. A long series of waves is a **wave train**. The greater the height of a wave and the greater the wavelength of a train of waves, the greater the energy contained in the waves. Two trains of waves can have phases during which the waves of each train reinforce one another and produce larger waves, called wave sets. The formation of larger and smaller waves can occur in phases. By timing the rhythm of the surf, you can make entries and exits easier by moving through the surf zone when the wave sets are small. Figure 5.12 illustrates fundamental concepts of waves and surf.

The water within a passing wave moves in a circular motion (this motion is called an orb). The diameter of the motion equals the height of the wave at the surface and diminishes with depth. Motion from a passing wave can be felt to a depth equal to half the wavelength for a series of waves. If there is 100 feet (30 m) between waves, you can feel the effect of a passing wave to a depth of 50 feet (15 m).

When waves enter shallow water, contact with the bottom interrupts the circular motion of the water within the waves. The circular movements flatten and eventually become a back-and-forth subsurface motion called **surge**. Wave contact with the bottom causes waves to slow and to become steeper. The wave height increases as the depth of the water decreases. When the depth of the water is about the same as the wave height, the wave becomes unstable and tumbles forward. At this point, the water within the wave moves forward and gives up its energy in breaking waves known as **surf**. In areas with offshore reefs, sand bars, and underwater obstructions, waves can break in shallow water, pass over the obstruction, reform, and break again in shallow water near the shore. Waves that break offshore indicate the presence of shallow water in the area where the waves are breaking.

Waves can break and release their energy all at once, or they can spill forward and expend their energy over a wide area. Waves that build quickly and break suddenly form plunging breakers. Waves that spill forward over some distance to shore are spilling breakers. Plunging breakers occur on steeply sloping shores,

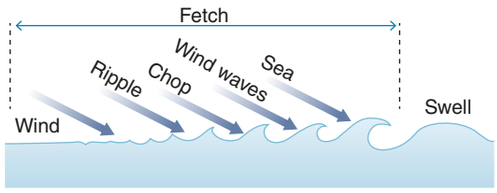

Wind over water causes waves—undulating forms of energy that can travel thousands of miles.

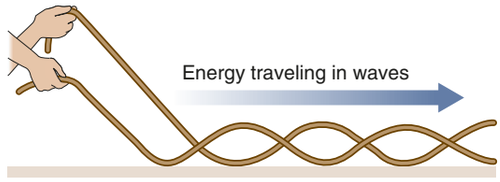

Wave movement is a flow of energy similar to that caused by vertically shaking one end of a rope.

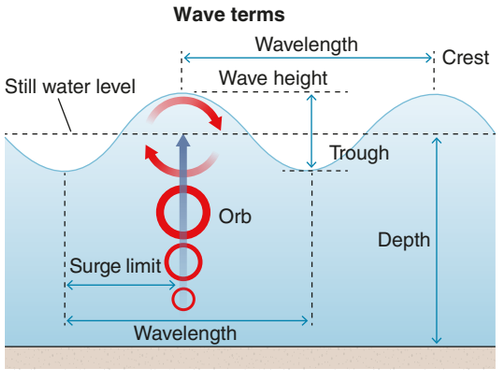

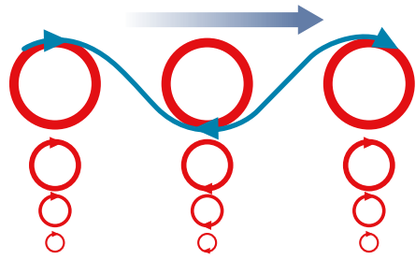

As a wave passes, water particles (represented by the circles) do not move with the wave; rather, they complete their orbit by returning to their starting point.

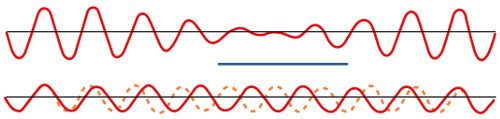

Wave sets resulting from two wave trains

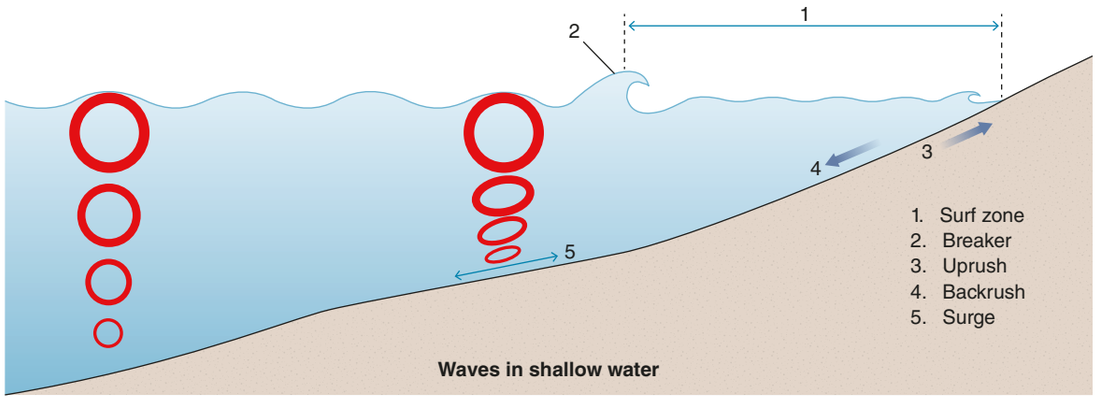

Figure 5.12 Wind speed, the absence or presence of obstacles, and the shape and depth of the ocean floor influence the formation of waves and surf.

while spilling breakers occur on bottoms with shallow slopes. Spilling breakers tend to reduce visibility more than plunging breakers.

The crashing water of surf contains air. The white water and foam in the surf zone do not provide as much buoyancy as water that does not contain air. It can be difficult to remain above water in the foam of surf, but you should not

attempt to rise above the breakers because you can be picked up and tossed forward by the moving water. You should remain low in the wave and breathe from your scuba regulator.

Surf rushes up the face of a beach and then flows back again to the still water level. The return flow of water is called the backrush. This countercurrent, sometimes called an undertow, does not extend beyond a depth of three feet (about a meter). Although backrush can be strong when surf is large on a steep beach, the belief that an undertow can carry a swimmer out to sea is a misconception.

The surf's crashing onto the shore moves sand. Gentle summer waves carry sand onto the beach. Large, rough winter waves carry the sand offshore, where it forms sand bars. This action explains why some beaches are rocky in the winter and smooth in the summer.

Currents

Currents are to water what wind is to air—fluid in motion. Because water is 800 times denser than air, water poses much more resistance to movement than air. You need to know how and why water moves and how to move with it. The force of water in motion is too great to resist. You must learn to use the flow of water to your advantage.

Wind, gravity, tides, and convection cause currents; the most common currents are wind-generated surface currents. Because of the rotation of the earth, currents flow at an angle to the wind that generates them. When the wind pushes water away from an area, a compensating current replaces the water displaced. Overall, the effect on the earth is to produce large circulating currents, called **gyres**, that move clockwise in the northern hemisphere and counterclockwise in the southern hemisphere.

When wind blows along a coast, a vertical compensating current—an upwelling—occurs. Another vertical current, the opposite of an upwelling, is a **downwelling**. A strong downwelling current is rare, but it can occur under the right conditions where there is a steep drop-off near shore. It is frightening to be pulled downward in water. You should find out if and where a downwelling occurs in an area, and you should avoid that location. If you get caught in a downwelling, swim horizontally until you are clear of the current. Trying to swim up against the current can lead to exhaustion and panic.

Water flows in layers (see figure 5.13). The water at the surface moves quickly with the wind, whereas the water a few feet beneath the surface moves slower. The greater the depth, the less the effect of a surface current. After 12 hours of a sustained wind in a given direction, the speed of a surface current is approximately 2 percent of the wind that drives it. Because water flows in layers, it is possible for a surface current to flow in one direction and for a subsurface current to flow in the opposite direction only a few feet beneath the surface.

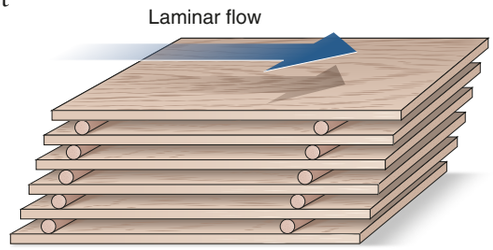

Figure 5.13 The laminar flow of water is similar to a stack of boards with rollers between them. When the stack is tipped, the top board moves much farther than the bottom board. Surface water moves more than water at depth.

Surface water at the equator expands as it is warmed and flows slowly toward the poles of the earth. At the same time, water cooled near the poles increases in density and sinks. Slow-moving convection currents result. These currents do not affect diving procedures, but they greatly affect the oceans by moving nutrients and pollution.

Water can move swiftly, and water movement intensifies whenever an amount of water passes through a restriction. When water moves past irregular formations, turbulence produces dangerous eddies and whirlpools. Swift-water diving is a specialty that is considered hazardous even with experience and training.

Incoming waves can pass over an underwater obstruction, such as a sand bar, and trap water on a shore. If the backrush of water flows back to sea through a narrow opening in the obstruction, a **rip current** exists. The current is narrow and strong and moves away from shore. The current dissipates shortly after passing through the restricted area. You can identify rips by a fan-shaped area of water on the shore, by muddy and foamy water in the rip area, and by a section of waves that breaks before the other waves (see figure 5.14). Rip currents can be stationary or moving. You must learn to recognize and avoid rip currents. If you are unable to swim toward shore, you are probably in a rip current. Swim parallel to the shore for about 60 feet (18 m) to get clear of the rip; then turn and proceed toward shore.

Waves breaking on a shore at a slight angle move water along the shore. The current is called a longshore current, and it affects diving. On steep beaches, the current can cut a shallow trough in the bottom close to shore. The trough, called a longshore trench, is a sudden drop-off and can cause an unsuspecting wading diver to fall. Items dropped into the surf zone move along the shoreline in the direction of the longshore current. The movement of sand on beaches occurs because of wave action combined with longshore currents.

Two types of currents result from all the forces of nature: standing currents, which are constant, and transitory currents, which occur briefly. The Gulf Stream that flows around Florida and up the eastern coast of the United States and the Antilles Current that flows through the Caribbean are examples of standing

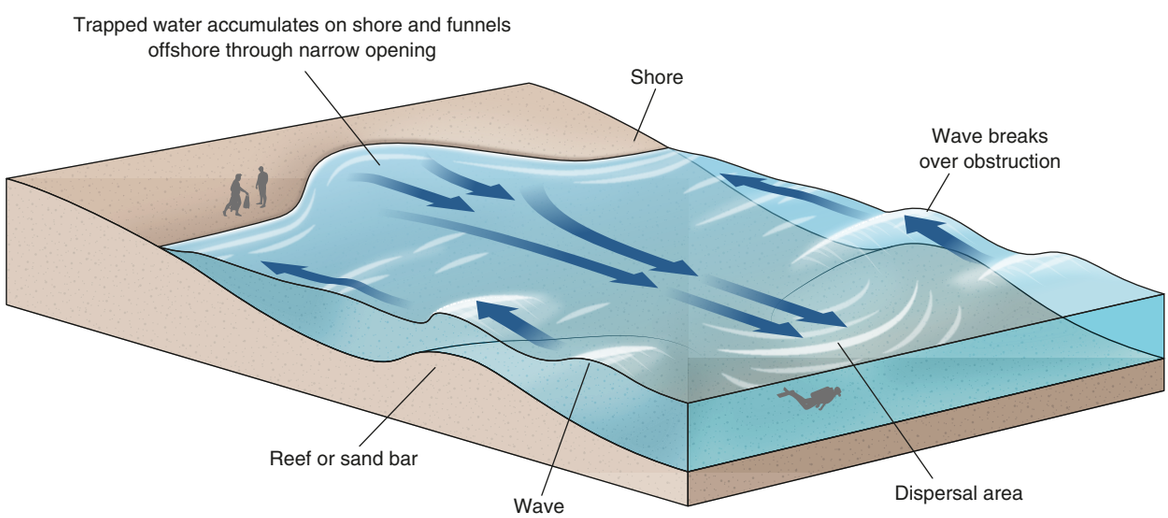

Figure 5.14 A classic rip current.

currents. A longshore current is a transitory current. A rip current can be either type. You should be familiar with the currents you will encounter or may encounter in an area.

Currents can flow at a rate of hundreds of feet per minute. The speed of a current is its **drift**, and the direction assumed by a current is its **set**. A fully equipped diver can swim at a rate of 60 to 100 feet (18 to 30 m) per minute (0.6 to 1.0 knots). Attempting to fight even a mild current is a futile waste of energy. Strong water movement can dislodge equipment, cause regulators to free-flow, and cause rapid heat loss. You must learn to recognize and estimate currents. If the currents are strong, you must avoid them. Even if they are mild, you should avoid struggling against them. The primary rule for diving in currents—when you begin and end your dive at the same point—is to dive in a direction opposite the flow at the surface; then use the surface current to aid you when you return at the end of the dive. (See chapter 7 for techniques to estimate currents.)

You also need to know some other general rules for diving in currents. Water movement is minimal at the bottom. Whenever possible, descend along a line from the surface to the bottom. If the current is still strong when you reach the bottom, ascend on the line and abort the dive. When you are unsure of your location during a dive, it may help to surface, determine the direction to the exit point, redescend, and move along the bottom toward your destination.

When there is a surface current, you should minimize the time spent at the surface. Whenever possible, hold on to a fixed object at the surface to keep from drifting downcurrent at a dive site. When a surface current is mild, swim against it to maintain your position at the surface. Develop the habit of noting your surface position in relation to a fixed reference. If you are caught in a sudden, strong current, swim across it to escape from it.

A dive in which the diver moves with the flow of a current is a **drift dive**. Simple drift dives along a coast are acceptable when planned. You enter the water at one point, move along the shore with the current, and exit at another point downstream. Another type of drift diving, which is a specialty activity, involves diving in currents from a boat. This type of diving is hazardous because the boat must be operated while divers are in the water. Do not attempt drift diving from boats unless you have completed specialty training and have a trained, experienced captain for the vessel.

Potential saltwater diving hazards include body heat loss, currents, surf, marine life, poor visibility, and fishing nets and equipment. You can minimize the hazards of ocean diving by following the recommendations that were described in this section.

Boat Diving

All divers eventually dive from boats, either small private boats or large charter vessels. Space is limited aboard boats, so you should organize your equipment compactly and pack it in the order in which you will need it. You may need to carry equipment for some distance to a boat, so plan to wear your scuba tank and weight belt and carry a bag in each hand. One bag should contain your dive gear, and the other bag should hold your lunch, camera, and any items that you want to keep dry.

Charter Boat Diving

You need to be well rested, so you should avoid partying the night before boat diving. Eat a meal that is free of acid and fat before you depart. Arrive at least 30 minutes before the scheduled departure time. Check in with a crew member. Usually, you must sign a roster and complete some forms. Ask where to stow your gear, and pick out a suitable space to store your dry items. If you are prone to seasickness, take preventive medication before the boat gets underway.

Tour the vessel. Locate the head (bathroom) and the emergency equipment. Stay out of the engine room and the wheelhouse (see example in figure 5.15). If you did not come with a buddy, you should find one before reaching the dive site. Listen carefully to the orientation provided by the captain or crewmember.

If you begin to feel queasy after the boat gets underway, you should go outside to the center of the boat and concentrate on the horizon. Sleep may help you acclimate to the boat's motion. If you are nauseated or vomiting when it is time to dive, avoid scuba diving. Vomiting underwater is hazardous. Mildly seasick divers may recover by snorkeling or by swimming to a nearby shore and waiting until they feel better.

Diving operations begin only when indicated by the captain. When you are ready to enter the water, check out with the divemaster. Typically, you will enter the water by stepping off the side of the boat, and you will exit the water by boarding a swim step at the stern (rear) of the vessel. Using the anchor line for your descent is a good practice.

Begin your dive against any current so that the water movement can assist you back to the boat at the end of the dive. Crew members often deploy a **trail line**—a long line with a float attached—to the stern for you to use in pulling yourself to the boat against the current (see figure 5.16). You should dive upcurrent; however, if you end up downcurrent from the boat and are unable to reach the trail line, you should make yourself buoyant, signal the boat, remain calm, and wait to be picked up after all the other divers have boarded the vessel.

If you hear a yelping, sirenlike sound while underwater, you should surface. The sound is from an electronic diver recall system. Remain where you are, and

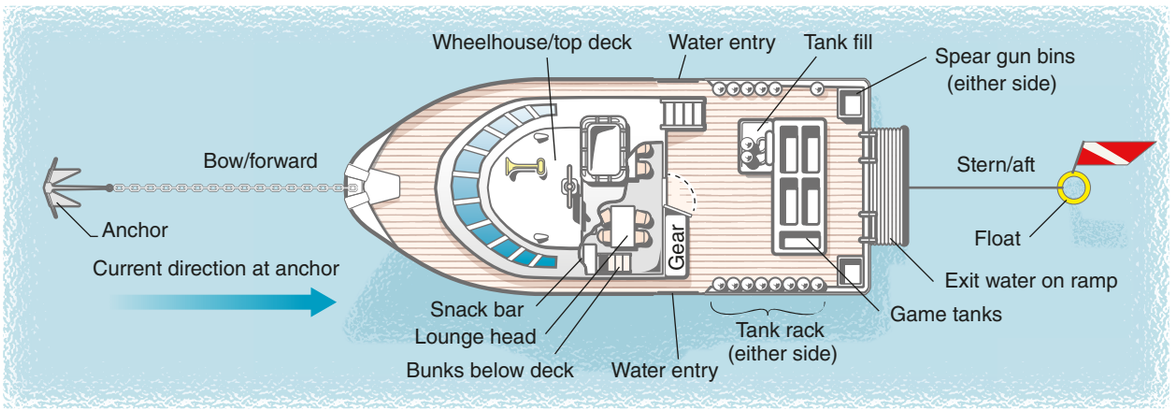

Figure 5.15 Typical charter boat configuration.

Figure 5.16 Divers can use a trail line to pull themselves to the boat if they surface downcurrent.

look to the boat for instructions. Learn the vessel's recall procedures in advance.

At the end of the dive, return to the boat by swimming along the bottom or at the surface. You should avoid skimming (swimming just beneath the surface), because you cannot be seen and you could be struck by a passing boat. Also avoid crowding the exit area. Divers should exit one or two at a time and remain clear of divers climbing the boarding ladder. Retain your equipment until immediately before exiting the water. If you do not know what equipment to remove, ask for instructions. Immediately after exiting, pick up any equipment that you handed to someone during your exit, and clear the exit area to make room for others.

After your last dive, pack your equipment before the boat gets underway. Be available for a visual roll call, which is required. Watch and learn from experienced divers.

Charter boat dives provide memorable diving experiences. Learn the proper techniques for diving from charter boats and then do what you have learned.

Small-Boat Diving

When you plan to dive from a small boat, make sure the boat is equipped for safety and has all the equipment required by the Coast Guard. The boat should also have a first aid kit. Leave a float plan—information about where you are going and when you plan to return—with someone who can summon assistance if you are delayed. Have a radio or cell phone aboard for emergency communications. Check the weather and water conditions in advance, and reschedule the dive if conditions will be unfavorable or are marginal on the planned dive day. Diving is not good when the weather is bad.

Small boats pitch, roll, and pound, so stow your equipment securely and expect it to get wet. If the trip to the dive site is short, you should suit up on shore. Doing so is easier and safer. If you must suit up in

a small boat, remain seated and help one another. Before you enter the water, make sure that there is a means to get back into the boat. More than one diver has been stranded alongside a boat, unable to exit the water. Display the dive flag, but only while diving. Deploy a trail line and float.

The usual entry from a small boat is a seated back roll made by two divers at the same time. This method enables the divers to avoid rocking the boat and causing one diver to lose balance and fall in when unprepared. The descent should be down the anchor line to ensure that the anchor is set and clear. If possible, you should leave an operator in the boat. A sudden weather change, a slipping anchor, or a change of water movement could pose serious problems at the end of a dive if the boat is unattended.

At the end of the dive, remove your weights and scuba unit before exiting. If you cannot hand them to someone in the boat, clip them to gear lines that have been hung over the side. Exit the water and pull in your equipment.

Boat diving is good diving. Be prepared, conserve space, plan ahead, learn the rules, and follow those rules. Learn more about boating to make boat diving more enjoyable. Complete a course in seamanship and small-boat handling.

Summary

The aquatic environment is vast and wonderful. We depend on it for our existence. We need to learn about it, care for it, and respect it. We must protect the environment and encourage everyone we know to assist in that effort. Because the environment varies greatly from place to place, you need an area orientation before diving. There are hazards in the underwater world, just as there are hazards on land. You can minimize aquatic hazards with training, experience, planning, proper procedures, and common sense.

Application-of-Knowledge (AOK) Questions

This chapter provides knowledge but does not explain how to apply it. In the remaining chapters you will learn the skills for diving in various environments. Right now you need to digest what you know about the diving environment and consider how it might affect your diving.

1. In what ways will the temperature of the water affect your diving?
2. In what ways will the visibility under water be affected by the environment and by you?
3. What ways can you recall that can help you manage currents while diving?
4. When you encounter strange animals while diving, how should you respond?
5. How much training and experience are required until you are able to dive anywhere at any time?
6. When you are an experienced scuba diver, where is the safest and most comfortable place to be when diving?
7. If you were equipped for diving in the ocean and wanted to dive in fresh water, what equipment changes would you need to make?
8. What do you think you need to know about diving conditions when you intend to go diving?

Beautiful coral reefs in the Red Sea

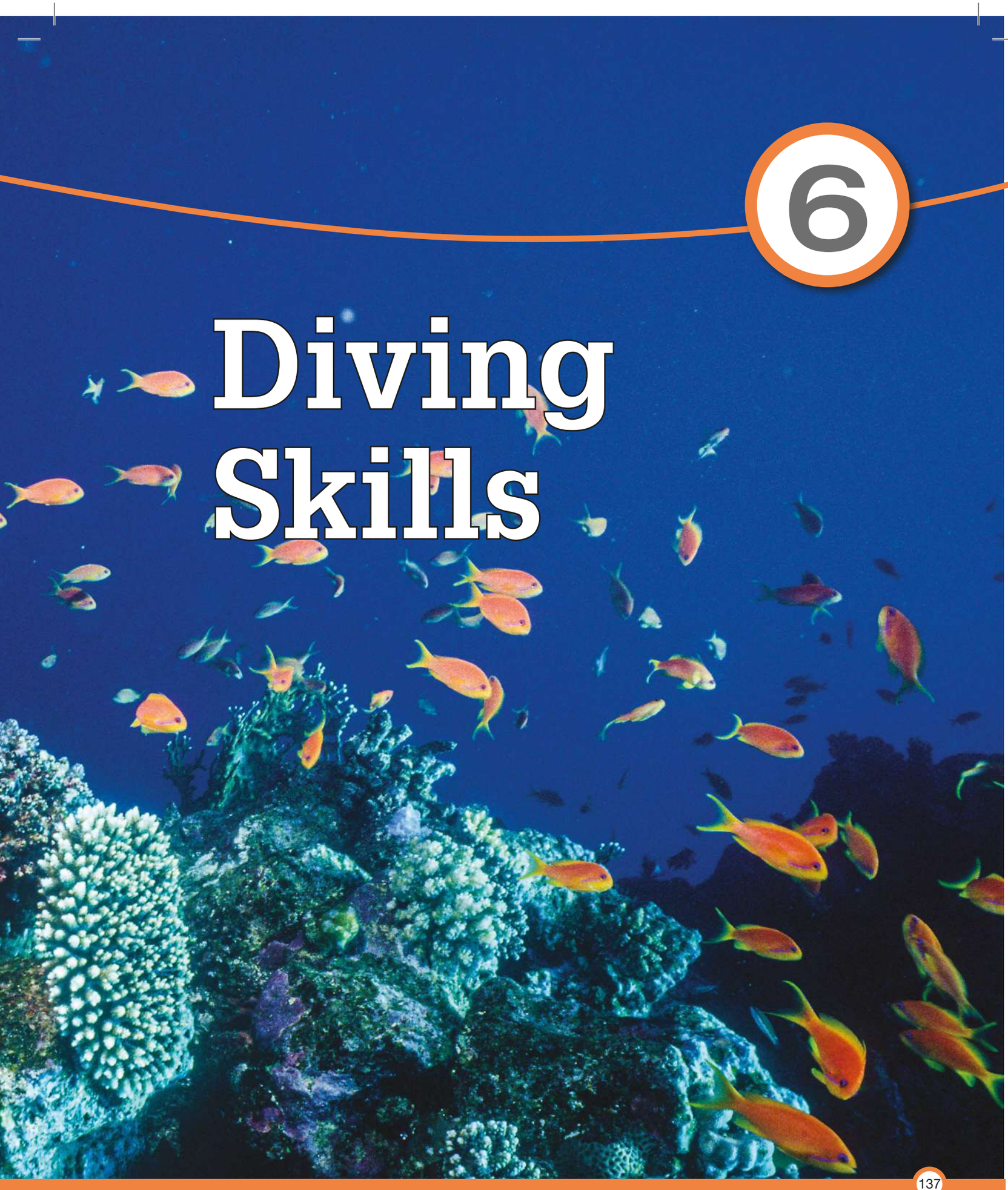

6 Diving Skills

Dive In and Discover

By the end of this chapter, you will be able to do the following:

- Prepare, assemble, don, adjust, inspect, and disassemble your equipment for skin and scuba diving.
- Demonstrate 16 skin and scuba diving hand signals.
- Explain how to test and control buoyancy.
- Explain how to clear water from a mask, a snorkel, and a scuba regulator.
- Explain how to descend, swim with fins, and ascend when skin and scuba diving.
- Explain how to handle, remove, and replace skin and scuba diving equipment.
- Describe various entry and exit techniques for skin and scuba diving.
- Explain how to use dive flags.
- Explain the buddy system and lost buddy procedures.
- Explain how to use a compass for underwater navigation.
- Explain how to prevent and manage seasickness, dizziness, stress and panic, overexertion, coughing, cramps, entanglement, loss of buoyancy control, loss of air supply, and dive emergencies.
- Explain how to rescue a diver who is not breathing and the first aid procedures for an injured scuba diver.
- Define the terms *buddy system*, *flutter kick*, *scissors kick*, *modified frog kick*, *dolphin kick*, *surf*, *neutral buoyancy*, *diver's push-ups*, *open-valve ascent*, *reference descent*, *nonreference descent*, *buddy line*, *compass heading*, *compass course*, *reciprocal compass course*, *flaring*, and *buddy breathing*.

You will learn most of the diving skills in controlled conditions; when you have developed the basic skills, you will apply them in open water. (Open water is any body of water representative of dive sites in the local area.) When you can demonstrate the skills of diving in open water, you are ready to receive your diving certification.

Preparing to Skin Dive

This section covers the skills of skin diving, many of which apply to scuba diving. When you swim at the surface and breathe through a snorkel, you are snorkeling. Scuba divers snorkel to conserve the air in their cylinder for underwater activities. When you breathe through a snorkel at the surface, hold your breath, and dive beneath the surface, you are skin diving. As a diver, you need to master the skills of snorkeling, skin diving, and scuba diving in order to enjoy all that the underwater world has to offer.

Preparing Equipment

Be ready to start your lesson when you come to your first water session. Prepare your mask, snorkel, and fins in advance. Clean your mask lens thoroughly so it will not fog (a description of the cleaning procedure is provided in chapter 4). Adjust the mask strap so it is snug but not tight. Attach the snorkel to the mask strap on the left side, and adjust the snorkel so it is comfortable in your mouth. If your fins have heel straps, adjust them so they are snug but not tight. Complete all adjustments before your first water session.

Inspect your skin diving vest or buoyancy compensator by inflating it, making sure it does not leak, and then deflating it. Put it on, reinflate it, and adjust the strap or straps so it will stay in position in the water. You may need to use a strap that runs between your legs and attaches to the front and back of the flotation device to keep it from riding up when you are in the water.

If you wear an exposure suit, your instructor will suggest an initial amount of weight for your weight belt. Adjust the length of the belt at the buckle end so the excess strap at the opposite end does not exceed 6 inches (15 cm). Allow 2 inches (5 cm) for each two-slot weight that you thread onto the belt. Distribute the weight on both sides of the belt so that you will be balanced in the water. Lock the weights in place with weight retainers.

Mark your personal equipment with your initials so that you can identify it. You can use special paint or markers (available at dive stores) or colored tape. Pack

Diving requires many skills, and you must repeat these skills correctly until you can execute them automatically. When you have mastered the skills of diving, your enjoyment can begin because you will be able to devote more attention to things of interest. The basic skills of skin diving and scuba diving are introduced in this chapter.

Anemones and clownfish in the Red Sea

your equipment—with the exception of your weight belt—in a gear bag. Place the items you will don last, such as your fins, on the bottom. Place the items you will don first, such as your exposure suit, on the top.

Donning and Inspecting Skin Diving Equipment

Don the pants or legs of your exposure suit first, then your boots, and then the top part of your suit. Donning a snug-fitting wet suit is easier if you wet the inside of your suit with water (to which you have added some mild shampoo) or if you wear a spandex suit as an undergarment. Place the ends of the legs of your wet suit over the tops of your boots so that water can drain from your suit when you exit from a dive. If the boots are outside the legs of your wet suit, water from the suit will balloon your boots when you get out of the water. If you become warm while donning the exposure suit, cool yourself before proceeding with your preparations. For open-water diving, don a cold-water hood before donning your wet suit jacket so that the skirt of the hood is underneath your jacket.

After you put on your exposure suit, you should clean and defog your dive mask and then set it aside until you are ready to don it (see the description of the cleaning procedure in chapter 4). Leave about 1/2 inch (1.27 cm) of water inside the mask so it will not dry before you put it on, and place the mask where it will not get damaged.

Don your skin diving vest or BC next. Place the skin diving vest over your head, and then secure the straps. Inflate the vest fully to make sure it will not be too tight. The straps should be as snug as possible without being uncomfortable or interfering with breathing.

Don your weight belt after the skin diving vest so the belt will be clear of the vest straps. Grasp the free end of the weight belt (the end without the buckle) in your right hand, and grasp the buckle end in your left hand. Pick up the belt, step through it, pull it into position across your back, and bend forward so that gravity supports the weight of the belt. Then tighten the belt and secure the quick release. Always wear the weight belt with a right-hand release, even if

you are left-handed, so that a rescuer will know how to release your belt in an emergency. If you always hold the free end of the belt in your right hand when donning it, you will always have a right-hand quick release.

Don the remainder of your skin diving equipment—mask, snorkel, and fins—at the water's edge or, if it is calm, in the water. Place the mask on your forehead and pull the mask strap over the back of your head using both hands to position the strap. Position the mask on your face, clearing any hair from beneath the sealing edge of the mask. Reposition the mask strap so it lies flat. If the mask has a split strap, position the split above and below the crown of your head. Don your snorkel when you don your mask. Check the adjustment of the snorkel when your mask is in place.

Stabilize yourself when putting on fins so that you don't lose your balance and fall. Hold on to your buddy (see "The Buddy Diving" section later in this chapter) or an object for support, or sit at the water's edge. Hold one side of a fin, bend one leg into a figure-four position, push the fin onto your foot, and pull the strap or heel pocket into place (see figure 6.1). Repeat the process for the other fin. Avoid walking with fins. If you must walk a few steps while wearing fins, shuffle your feet and walk backward. If you try to walk forward, you can lose your balance or damage your fins.

After you have donned your equipment, you should inspect it for completeness, correct positioning, and adjustment. When you are satisfied with your

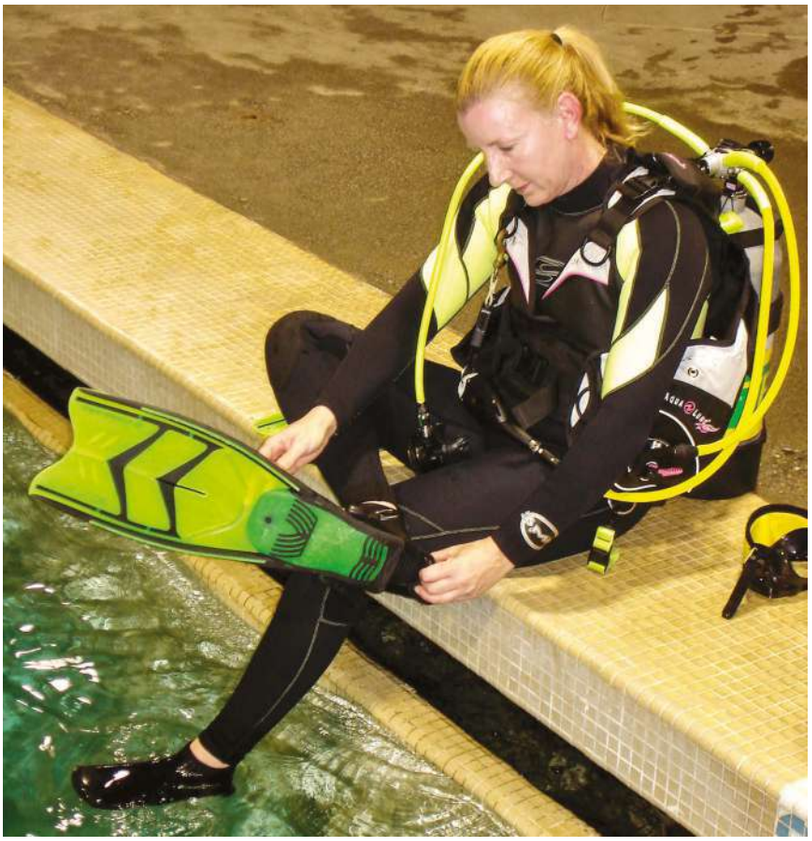

Figure 6.1 Donning fins.

equipment, inspect your buddy's equipment while he inspects yours. Inspect from head to toe. You may be able to see something that your buddy could not see, or vice versa. Develop the habit of inspecting each other's equipment before every dive.

Skin Diving Skills

Diving skills can be divided into three categories: skin diving skills, scuba diving skills, and problem management skills. This section introduces skin diving skills.

Using Skin Diving Hand Signals

You cannot talk with a snorkel in your mouth and your face in the water, so you must use hand signals as a primary means of communication. You need to learn and use the standard hand signals shown in figure 6.2. You will learn additional hand signals for scuba diving. Display hand signals clearly and deliberately when you send them, and acknowledge all hand signals you receive.

Figure 6.2 Skin diving hand signals.

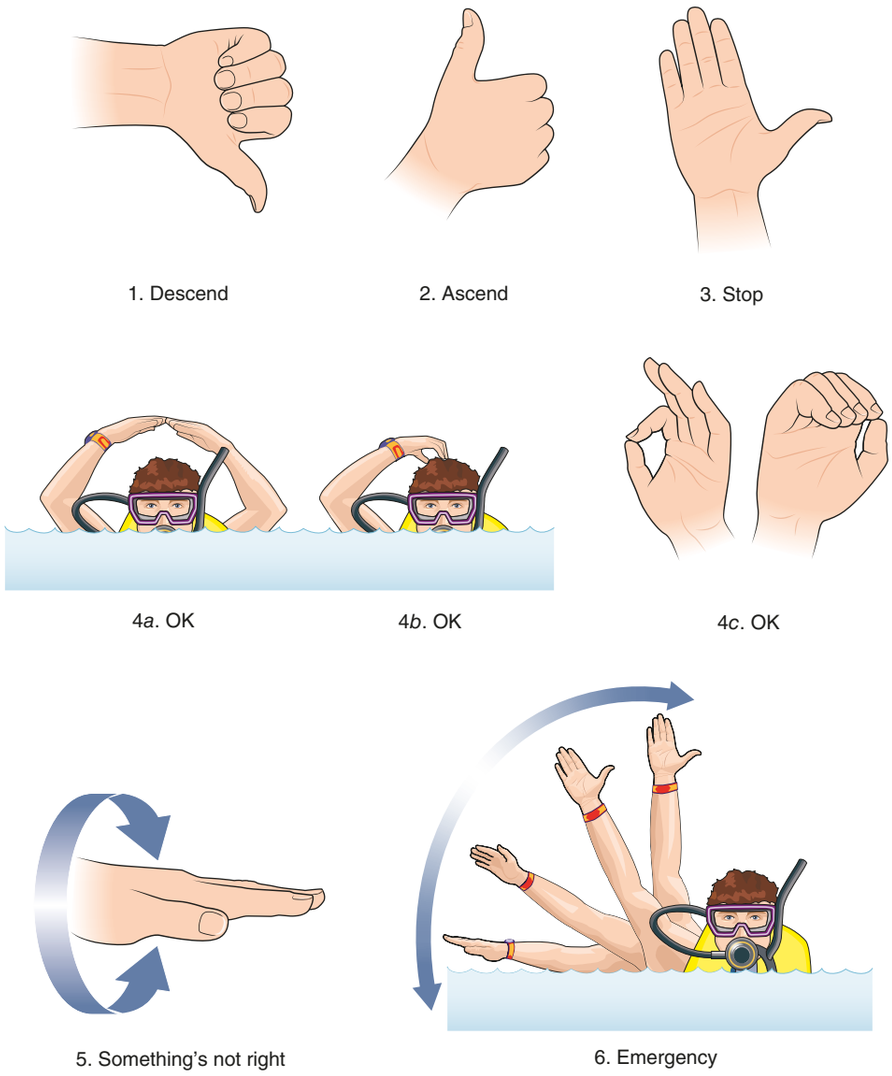

Using a Skin Diving Vest

You inflated and deflated your skin diving vest as part of your diving preparations. You also need to learn how to deflate and inflate the vest in the water. To deflate the vest, position your body so the vest's exhaust port is the highest point. Hold the deflation valve open while you sink lower in the water. The water pressure helps force the air from the vest. You may have to hold the collar of the vest down with one hand to remove the air from that area. When you have vented all of the air from the vest, close the valve to prevent water from entering. Get your buddy to confirm that all the air is out of your vest.

Inflating the vest in the water is easier with your head underwater instead of above water. Take a breath of air, duck your face beneath the surface, insert the oral inflator into your mouth, open the valve, exhale into the vest, and close the valve. Repeat this process until you have the desired buoyancy.

Testing Buoyancy

Your buoyancy should be neutral at the surface. If you are too buoyant, you will have to struggle to descend and remain underwater. If you are not buoyant enough, you will tire quickly at the surface and sink while swimming underwater. You must be able to adjust your buoyancy correctly.

If you are wearing an exposure suit, you need weights to offset the buoyancy of the suit. With all equipment in place, position yourself in chest-deep water. Exhaust all air from your skin diving vest. Take a full breath, hold it, lift your feet from the bottom, and remain motionless while you slowly count to 10. If you sink, remove some weight and try again. If you exhale half of the air in your fully inflated lungs and still can't submerge, you need to add weight until you can. When your weighting is correct, you remain at the surface while holding a full breath and sink when exhaling. Adjust your weighting until you achieve neutral buoyancy, which means that you neither sink nor float when holding an average breath. Have your buddy watch while you make the final test. With your lungs about half full of air, the top of your head should remain even with the surface of the water when you are motionless in a vertical position.

Clearing the Mask

When your head is above the surface, water inside a mask will run out if you pull the bottom of the mask away from your face. Water inside a mask underwater also flows out the bottom of the mask if you displace the water with air. Putting air inside the mask is easy: Just exhale lightly through your nose. A long, light exhalation is better than a short, forceful one because a strong exhalation blows air past the seal of the mask and does not displace water effectively.

To clear a mask that has a purge valve, seal the mask against your face, tilt your head downward to make the purge the lowest point in the mask, and exhale through your nose until the mask is clear of water. To clear a mask without a purge, hold the top of the mask against your forehead, take a breath, and start exhaling slowly (see figure 6.3). When the level of the water is below your eyes, tilt your head back while continuing to exhale, and the remainder of the water will flow out the bottom of the mask. You must be exhaling when you tilt your head back or water will run up your nose. Bubbles from the bottom of the mask indicate when you have cleared all the water from your mask. It may

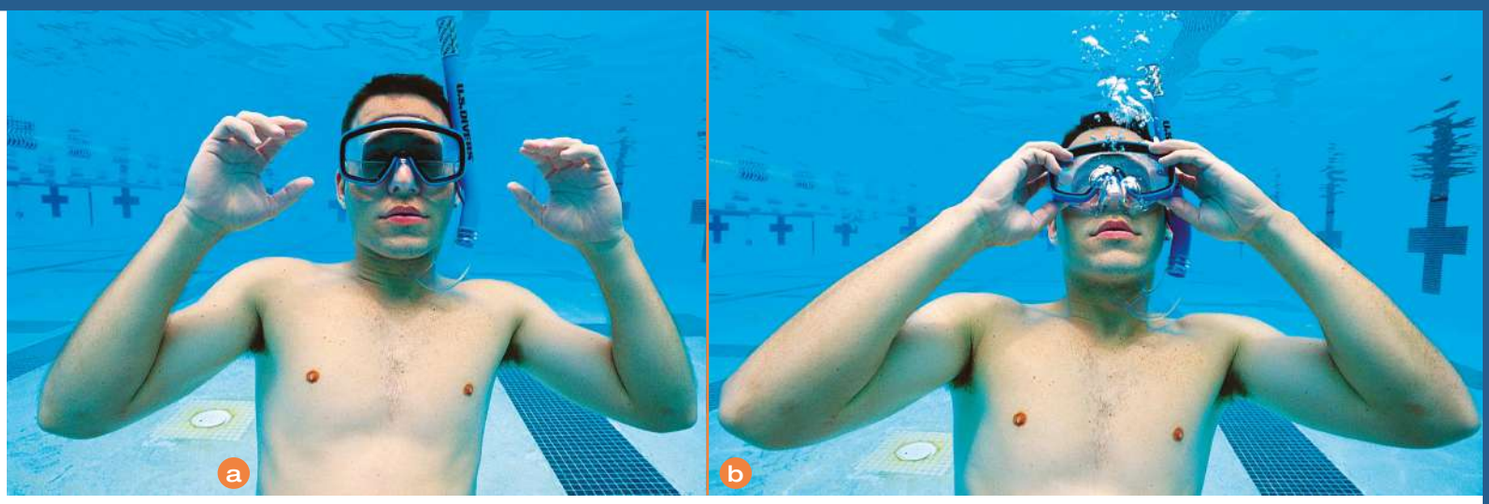

Figure 6.3 Clearing a mask.

sound as though it takes a long time and a lot of air to clear a mask, but with practice you will be able to flood and clear your mask several times after a single inhalation. Clearing requires only a few seconds.

Using the Snorkel

Stand in the water and lean forward with the snorkel in your mouth. Place your face in the water and inhale gently. If you inhale forcefully at first, you might inhale water. When you are sure the tube is clear, you may breathe more forcefully through your snorkel. When you descend beneath the surface, the snorkel tube can fill with water. Self-draining snorkels drain nearly all the water from the tube when you surface.

If your snorkel is not the self-draining or water exclusion type, you have to blast the water from the tube with a strong, sharp exhalation. To do this, blow hard into the snorkel to force the water out. Inhale cautiously after exhaling. If a little water remains in the tube, you can breathe past it if you inhale gently. After you fill your lungs again, expel the remaining water with another forceful exhalation.

If you have a simple snorkel without valves, you can clear the tube while ascending by using the displacement method. Begin by looking up to invert the snorkel; then exhale a small amount of air into the tube. As you ascend, the air expands according to Boyle's law and displaces the water in the tube, which will be clear of water when you reach the surface. The tube is inverted at this point, however, so when you turn it to the upright position, water will flow into the tube unless you exhale while rolling your head forward at the surface. Continuous exhalation prevents water from entering the tube. When you use the displacement method, which is easier than the blast method, you can take a breath the moment you reach the surface. Figure 6.4 shows the displacement method.

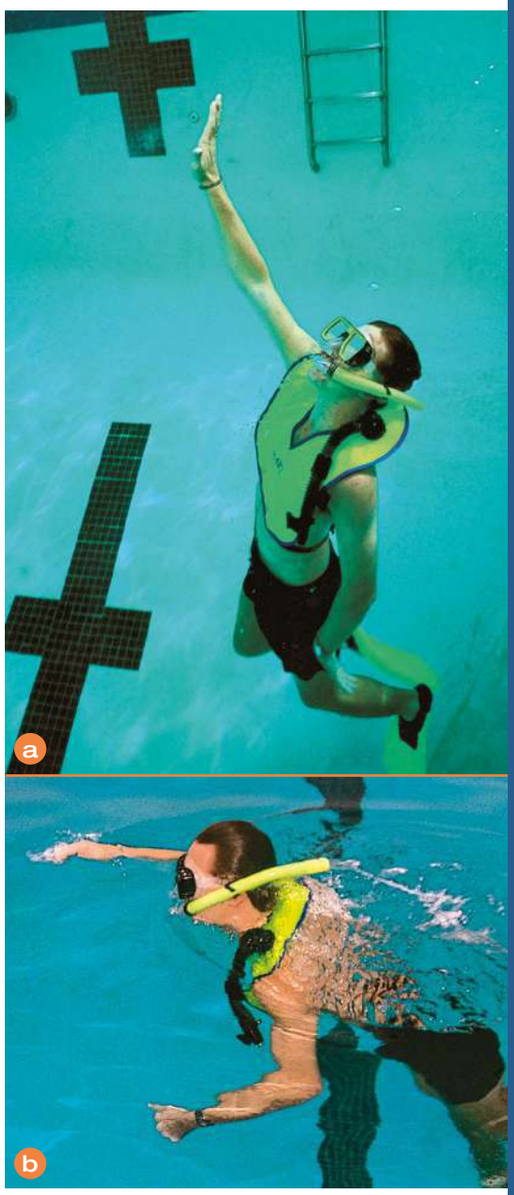

Figure 6.4 Displacement method of clearing a snorkel.

Learn to keep the snorkel in your mouth when you are in the water. Avoid the temptation to remove the mouthpiece and shake the water from your snorkel. You do not have to remove the mouthpiece when you can clear your snorkel proficiently. When your hands are occupied, you will not be able to remove and replace the snorkel to clear it.

Using Fins

Fins provide large surfaces that you can push—using the strong muscles of your legs—against the water for propulsion. The blades of the fins need to push the water the way a broom sweeps a floor. A broom cannot sweep if you move it up and down, and your fins will not provide propulsion if you move them lengthwise in the water. You must kick them back and forth in wide, sweeping kicks. Small fins allow short, quick kicks, but larger fins require wider, slower kicks. Moving efficiently, not speedily, is your objective when wearing fins.

The most common fin kick is the **flutter kick**, which is an up-and-down kick that you can do facing down, up, or to the side. Your fins need to be underwater to provide propulsion. It is easier to keep your fins submerged at the surface when you are on your back or on your side instead of your belly. Move your legs up and down from the hip while bending your knees only slightly. Extend your legs almost fully throughout each stroke, and use wide, slow kicks. Keep your hands at your sides or extend them in front. Do not use your hands and arms to swim. Figure 6.5 shows the flutter kick.

The **scissors kick**, shown in figure 6.6, is a resting kick. The scissors kick is similar to a flutter kick, but it requires less energy. While lying on your side, slowly extend one leg backward while you extend the other leg forward, and then pull your legs together quickly. Hold a streamlined position with your toes pointed while you glide through the water. When you come to a stop, repeat the stroke. Extend your legs almost fully throughout the stroke.

A **modified frog kick** is useful because it uses different muscles than the flutter and scissors kicks. Changing kicks is helpful if you become tired or develop leg cramps. The modified frog kick is different from a swimmer's frog kick. Assume a facedown position with your fins perpendicular to the bottom. Rotate your ankles so that the tips of the fins point outward, slide the fins tip first in an outward direction, then pull the bottoms of the fins together quickly in a wide, sweeping arc. Extend your legs almost fully throughout the stroke, and then hold the final position with your feet together and your toes pointed while you glide (see figure 6.7). When you come to a stop, repeat the stroke.

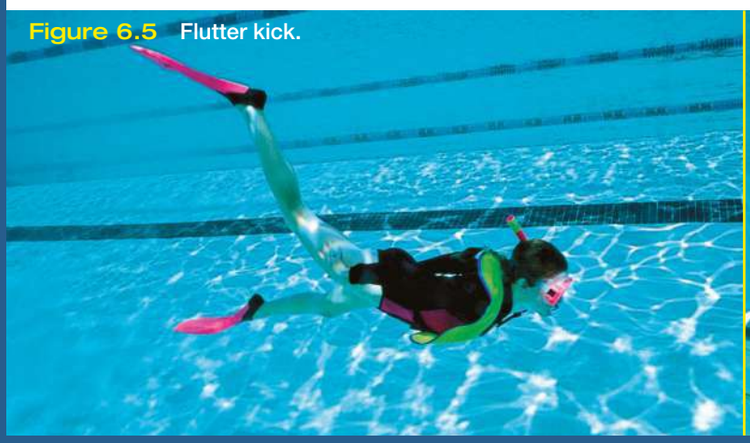

Figure 6.5 Flutter kick.

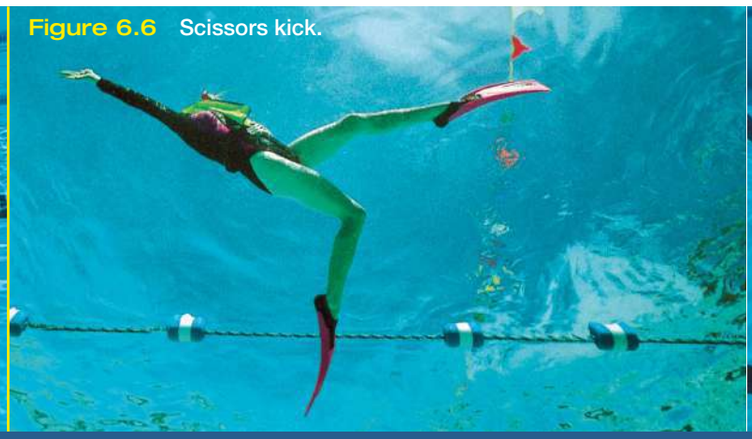

Figure 6.6 Scissors kick.

The **dolphin kick** is also useful for a change of pace or if you lose a fin. Hold both feet together continuously. Exert force against the water with a wavelike up-and-down motion of your body. To begin, straighten your body, extend your feet together, bend your shoulders forward, and pull your feet slightly backward. Next, lift your shoulders, thrust forward forcefully with your hips, and then pull down with your fins. When you see this kick demonstrated, you will be able to learn it by imitation. If you lose a fin, you can use the dolphin kick to propel yourself. Cross your legs and put the leg without a fin behind the one that has a fin. Figure 6.8 shows the dolphin kick.

Performing Surface Dives

You need to be able to dive down to see the beautiful world beneath the surface. The initial dive beneath the surface of the water is a surface dive. The principle of a surface dive is simple: You raise part of your body above the water and point yourself straight down; the part of your weight above the water drives you downward. Once you are submerged, you start swimming to continue your descent.

You should learn three types of surface dives: the pike, the tuck, and the feet-first dive (see figure 6.9). You do the pike dive while moving forward at the surface. Bend forward at the waist to make your trunk vertical in the water. Next, quickly lift your legs out of the water to a vertical position. The more of your legs you can lift out of the water, the farther your surface dive will push you under the water. For shallow dives, you can do the pike dive by lifting only one leg.

When you have learned the pike dive, you are ready for the tuck dive, which is similar. Do the tuck dive from a stationary position. Begin in an upright position in the water, pull your knees to your chest, and sweep backward with your arms to roll yourself forward in the water. When you are inverted, extend yourself fully into a vertical position. The procedure is a coordinated movement that you must do quickly to get your legs above water. After you extend your legs, the remainder of the dive is the same as the pike dive.

Use a feet-first dive in areas where surface plant growth is dense because you are less likely to get entangled. Begin in a stationary, upright position, and use a strong scissors kick to propel yourself up and out of the water as far as possible. Pull your arms downward to your sides for added lift. When you reach the highest point, point your toes downward, and hold your arms to your sides. The weight of your body will push you below the surface. Raise your arms upward when you are below the surface to push yourself even lower. When your downward momentum ends, do a tuck dive to invert yourself, then continue your descent.

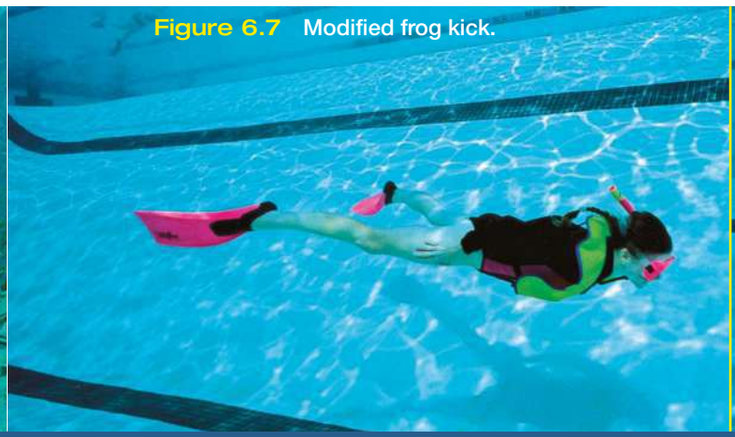

Figure 6.7 Modified frog kick.

Figure 6.8 Dolphin kick.

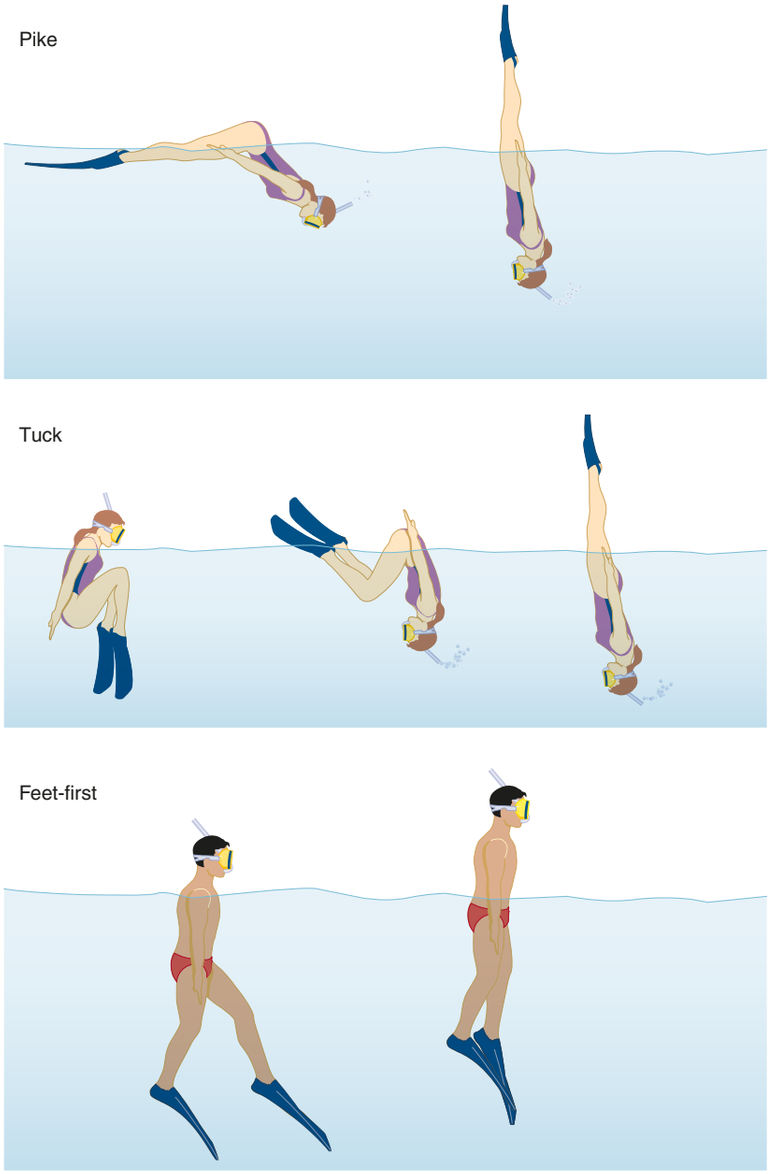

Figure 6.9 Surface dives.

Performing Descents, Underwater Swimming, and Ascents

To prepare to descend, vent all air from your flotation device. Use an ear-clearing maneuver (described in chapter 3) at the surface so that it will be easy to equalize pressure in your ears while you descend. Hyperventilate three times, inhale as fully as possible, and don't exhale until you return to the surface.

Initiate your descent with a good surface dive. Equalize the pressure in your ears and mask every couple of feet during your descent. Use wide, slow, powerful kicks to propel yourself to your desired depth, and then relax as much as possible.

Swimming underwater sounds simple, but some people have trouble controlling direction. Your head is your rudder. Pull your head backward while swimming facedown to go up. Bend your head forward to go down. Bend your trunk to the left or right to turn. With practice, you can go in any direction you choose without using your hands.

Follow these four rules for ascents:

1. Reach up; extend one arm over your head for protection (see figure 6.10).
2. Look up to avoid obstructions and to determine when the surface is near.
3. Come up slowly. A slow ascent is better than a rapid ascent because swimming rapidly uses more oxygen.
4. Revolve once or twice during your ascent to check your surroundings.

Alternate breath-hold dives with your buddy. One diver should remain at the surface, watching, while the other diver is underwater. Make sure your buddy knows where you are at all times.

Figure 6.10 Extend your arm for protection during a skin diving ascent.

Handling Equipment

Occasionally you may need to remove, adjust, and replace skin diving equipment while you are in the water. You should be able to handle your equipment easily with training and practice.

To remove, adjust, and replace your mask, begin by inflating your skin diving vest. Because you will make the adjustment above water, you should use buoyancy to reduce the effort. After you make the adjustment, replace your mask using the procedure for donning the mask described earlier in this chapter.

You may need to adjust a fin strap or remove sand or gravel from your fins while diving. The fins are easier to remove, adjust, and replace than the mask is. You do not need as much buoyancy to work with your fins because you can look into the water and make the adjustment underwater. In fact, too much air in your skin diving vest can be a nuisance. Keep your face in the water, breathe through your snorkel, and work with one fin at a time.

You may need to tighten, adjust, or replace your weight belt. The weights may slip, or the belt may fall off. To tighten or replace your weight belt, you need to get the belt across the small of your back while you are in a facedown position breathing through your snorkel. If you try to replace or tighten the belt in an upright position, you will have a constant fight with gravity. When you are in the correct, facedown position, gravity becomes your ally while you secure the buckle. To get the belt into position across your back, begin in an upright position, holding the free end of the weight belt against the outside of your right thigh. Lean back into a horizontal, faceup position momentarily, and then roll to your left to a facedown position while continuing to hold the free end of the belt against your right thigh. At this point, the belt will be draped across the backs

of your thighs. Clear your snorkel so that you can breathe. While holding the free end of the belt in your right hand, reach down with your left hand, grab the buckle end, pull the belt into position across your back, and secure the buckle.

Removing Equipment

Your equipment helps you adapt to the underwater environment. Develop the habit of keeping your equipment in place while you are in the water. You may need to defog your mask or make an adjustment, but other than that you should wear your equipment continuously. Avoid the tendency to prop your mask on your forehead when you are at the surface. Mask propping is a sign of distress and a good way to lose your mask and snorkel. If you must remove your mask while in the water, pull it down around your neck where it will be secure.

Bottom conditions at the dive site determine whether you should remove your fins when leaving the water. In some areas, you can remove your fins in waist-deep water and wade out of the water. In other areas, you need to wear your fins until you are clear of the water. Fins can be removed in several ways. When you prepare to climb a boat ladder, hold the ladder continuously while crossing your legs in a figure-four position to remove each fin. If the boat has a platform at the rear, swim onto the swim step and remove your fins in a kneeling position. When you remove fins in waist-deep water or on land, you should use your buddy for support. On steep beaches, you may choose to crawl from the water. Your buddy can remove your fins while you are on your hands and knees. Your buddy then crawls ahead of you, and you remove your buddy's fins. When you stand up, you exchange fins.

Weight belt removal also varies. When you can walk or climb out of the water, keep your belt on while doing so. When you remove a weight belt, lower it gently instead of dropping it. You can cause damage or injury if you develop the habit of dropping your weight belt. If you must pull yourself onto a dock or into a small boat, remove your belt first and hand it up.

Keep your skin diving vest in place until you are clear of the water. Remove your skin diving vest, and then your wet suit. Remove your wet suit by turning it inside out. Be sure to remove your boots before your wet suit pants. You will need assistance from your buddy to remove a pullover jacket or a jumpsuit.

Preparing to Scuba Dive

Nearly all of the skills of skin diving are used in scuba diving, but you need to learn many additional skills for scuba diving. This section introduces the preparatory skills of scuba diving; the next section covers basic and postdive skills. The more familiar you become with the skills by studying them, the better you will be able to perform them. You must learn the skills correctly from the outset in order to develop good habits for diving safety.

Packing Equipment

Nothing is more inconvenient than getting ready to dive and discovering that something is missing or not right. You should take steps to prevent equipment inconveniences. Begin by using a checklist when you pack your diving equipment to ensure that you have everything you need (see the diving equipment checklist). After taking inventory, inspect your equipment as you pack it. If you

have not used your equipment for a while, assemble it and test it first. Make sure everything works properly.

Pack your equipment in your gear bag—except for your weight belt, tank, and BC. When you have to carry your equipment, put your BC on your tank, and carry the tank on your back. Carry your gear bag in one hand, and carry your weight belt in the other hand. Equipment-carrying devices with roller wheels can reduce the work of moving your diving equipment (see figure 6.11).

Figure 6.11 A gear bag with wheels can make transporting equipment easier.

Assembling Equipment

Today's backpacks are part of the BC and have one or more bands to secure the scuba cylinder. Many of the bands are made of webbing, which stretches when wet. Soak a fabric tank band in water for a couple of minutes before securing it around a scuba tank. The soaking softens the webbing and allows it to stretch when you tighten the band. A dry fabric belt may allow the tank to slip when the band gets wet and stretches. A tank that slips from a pack can be a hazard.

Orient your backpack with the opening of the tank valve facing the backpack. Stand the tank up with the valve handle facing to your right. Slide the tank band over the tank with the tank between you and the pack. Tighten the band so the valve opening points directly toward the pack (see figure 6.12).

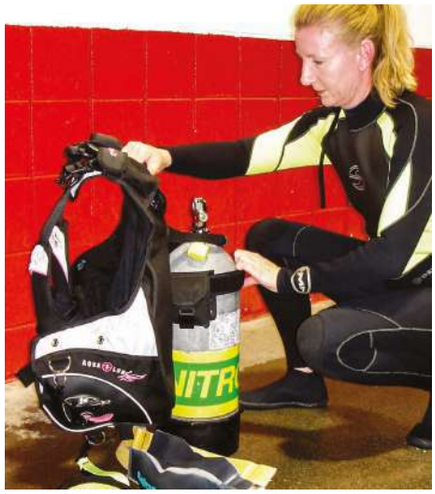

Figure 6.12 Strapping the backpack to the tank.

The height of the tank band on the tank varies with the type of backpack. Generally, the top of the pack is even with the base of the tank valve. After you attach the pack and before you attach the regulator, don the scuba tank and check the height adjustment by slowly tilting your head backward. If your head hits the tank valve, the tank is too high in the pack. Reach over your shoulder for the tank valve. If you cannot touch it, the tank is too low in the pack. Adjust the height as needed. When you become familiar with the correct height adjustment, you will not need to test it before completing the assembly of the scuba unit, but you should check it before entering the water.

Tighten the band as much as possible before securing it, and test the tightness by grasping the tank valve with one hand and the top of the pack with the other hand. Try to move the pack up and down on the tank. If there is movement, the band needs to be tighter.

Attach the regulator assembly to the scuba tank. The tank valve should have a valve protector or a piece of tape over the valve opening. Remove the cover

or tape (but do not litter—put the tape in a trash container or in your gear bag). Make sure that the O-ring remains in place and is free of debris. Loosen the regulator yoke screw, and remove the dust cover. If your regulator second stages have purge depressors, release the purges. Expel any water or dirt from the valve by opening the tank valve slightly and momentarily before you attach the regulator to the tank.

Your regulator has several hoses. Orient the hose to your primary second stage toward the same side of the tank as the handle of the tank valve so that the hose will come over your right shoulder. When you orient the hose this way, the other hoses are properly oriented automatically as long as they are untangled and can hang freely.

Carefully seat the inlet opening of the regulator on the tank valve outlet, and then turn the yoke screw or DIN fitting until it is snug but not tight. Tighten the fittings using only your fingers and thumb. See figure 6.13 for an example of a regulator attachment.

Attach the various regulator hoses to the scuba unit. It is easier to attach the low-pressure hose to the BC before you pressurize the hose. Attach your instrument console to the BC, but do not attach your extra second stage yet.

Turn on the air. Hold the SPG with the front facing away from you and others so it will not cause an injury if it fails when pressurized. The SPG blowout plug should prevent an explosion from occurring, but holding the gauge is a good precaution. Open the tank valve slowly in a counterclockwise direction. Open the valve all the way; then close it one-quarter turn. You should feel the hoses stiffen under pressure. Listen for leaks in the system. If the regulator free-flows, cover the mouthpiece opening with your thumb to stop the flow. If air leaks

Figure 6.13 Your regulator and hoses should look like this if attached correctly to the cylinder.

from the tank valve seal, turn the air off, remove the regulator, and inspect the seal. You may have to replace the O-ring. Solve all air leakage problems before you use the scuba unit.

Testing the unit is the final step. Use your SPG to make sure the tank is full. Reset instruments in your console as needed. While looking at the SPG, depress the regulator purges momentarily to test the second stages and clear them of any debris; then breathe deeply through each second stage several times. The tank pressure reading should remain constant while you breathe from the regulator. If the pressure reading drops when you inhale, you have not opened the tank valve sufficiently. Do not open the valve partially to check the pressure of the tank. When you have finished testing the regulators, attach the extra second stage to the BC. Depress the low-pressure inflation valve on the BC for one or two seconds to make sure the valve functions correctly. When you have completed the assembly and testing of the unit, lay it on its side with the regulator and instruments on top.

Equipment assembly procedures are the same for both new and experienced divers. Always assemble and test your scuba equipment before you suit up. Practice will help you recall the assembly procedures.

Donning and Inspecting Scuba Equipment

With skin diving equipment, you don the weight belt after the skin diving vest, but with a jacket-type BC, you should don the weight belt before the scuba unit.

A good way to don a scuba unit out of the water is for your buddy to hold the system and help you into it. Once your arms are through the shoulder straps, you should bend forward and balance the unit on your back while you secure the waistband. Place the regulator hoses over your shoulders while you fasten the waistband so the hoses are not trapped beneath the waistband. Some dive boats have vertical tank racks on seats to allow you to sit while donning the system. Do not sit on the deck or ground to don a scuba tank, however, because the tanks of other scuba divers can strike you in the head.

Some divers prefer to don the scuba unit in the water—this is a good practice if you dive from a small boat or have a minor back problem. Put the scuba unit into the water first. Make sure that it does not float away! To don a jacket BC unit in the water the way you would don a coat, inflate it fully, sit on it, put your arms through the arm holes, and slide off and into the unit. Or you may put your scuba unit on over your head. Maneuver the tank in front of you with the valve facing you and the backpack facing up. Place the regulator in your mouth and keep the hose of the primary second stage between your arms. (If you place your right arm inside the loop of the hose, the hose will wrap around your arm when you try to lift the tank over your head.) Place your forearms completely through the arm holes just past your elbows, and begin to lift the scuba unit over your head. Rather than lift the unit, you should push yourself down in the water and duck beneath the scuba unit. Lower the tank into position on your back. Use your left hand to pull your snorkel clear of your BC as the tank lowers into position. When the tank is in place, lean forward and secure the waistband.

When you have donned your scuba equipment and the remainder of your diving equipment, inspect all your equipment to make sure it is positioned, adjusted, and functioning correctly. Your buddy should do likewise. Then inspect each other's equipment.

Do a head-to-toe buddy check. Specific checks for each item of equipment are listed in the Head-to-Toe Buddy Check sidebar.

Equipment inspections are important. Solving problems is easier before you enter the water than afterward. It does not take much time to inspect each other's equipment, but the practice can save a lot of time later. Make predive inspections a habit.

Scuba Diving Skills

To be a scuba diver and enjoy diving with minimum risk of injury, you need to learn the proper procedures for entering and exiting the water, controlling your buoyancy, descending and ascending, monitoring your instruments, and coordinating with a buddy. This section provides a helpful introduction to these skills, but you must learn them by doing. You will learn the skills in pool-like conditions and then apply them in open water.

Using Entry Techniques

The four basic types of water entries for divers are wading, seated, feetfirst, and back roll. You need to learn when to use each type and the procedures for each. The objective of any entry is to get into the water as easily as possible without injuring yourself or losing any equipment. After you enter the water and are under control, switch from your regulator to your snorkel if you are going to swim or remain at the surface.

You can make open-water entries from the shore or from a boat. You can wade in from the shore, or you can enter from an artificial structure such as a dock, pier, or jetty. The conditions vary greatly, and so do the entry techniques. Surf may be a factor at some locations. Bottom conditions can range from smooth and soft to rough and firm. The bottom may slope gradually or steeply, and there

Head-to-Toe Buddy Check

Mask	In place, flush on face, no hair under skirt, not fogged
Snorkel	Attached to left side of mask
Scuba cylinder	Valve open all but a quarter turn, valve facing user's right side, height OK, snug in pack
SPG	Full pressure, not damaged or leaking
Scuba regulator	Primary and alternate air sources function OK, both buddies know the placement of the alternate air systems, regulator hoses oriented properly and free for use
Buoyancy compensator	Inflates and deflates properly
Weighting system	In place, proper amount of weight, right-hand quick release accessible, weights clear to drop, both buddies familiar with weight release system
Exposure suit	Properly positioned and zipped
Dive knife	Attached properly, easy to access, secured
Fins	Securely strapped or worn
Instruments	Watch, dive timer, or dive computer ready for use

may be holes and drop-offs. Plants, animals, and rocks may be present in the entry area. A good entry technique for one location may be inappropriate for another. It takes experience and knowledge of the area to determine an effective entry procedure.

You need to know some general techniques when making wading entries at an open-water site. A wading entry sounds simple, but keep in mind that diving equipment affects your center of gravity, your mobility, and your peripheral vision. You must walk backward or sideways when wearing fins.

If there is little or no surf, you can wade in, don your fins, and begin your dive. Breathe through your regulator and have your BC inflated partially. Shuffle your feet to detect holes and rocks and to chase away bottom-dwelling creatures (see figure 6.14). When the water reaches your thighs, lie down in the water and start swimming. In some areas with muddy bottoms, you should not wade because you can sink deeply into the mud and lose a fin when you try to extract your foot. When the bottom is firm and the water is calm, you can wade into the water without your fins and don them in the water. It is helpful to have information about the bottom conditions of a dive site.

In most areas where there is surf, you should don your fins before you enter the water, and you should not remove them until you are clear of the water after the dive. Time your entry to coincide with small waves (see the discussion of waves and surf in chapter 5). Keep all equipment in place, and breathe through your regulator. Deflate your BC because you need to duck beneath breaking waves when they are higher than your waist. If you inflate your BC, you will be unable to duck beneath the waves, and a large wave could lift and toss you. Hold your mask with one hand at all times; spread your fingers and curl them over the top of the mask so you can see. Keep your knees bent and shuffle sideways into the waves to minimize your profile to the moving water. Stop moving just before a wave hits you, allow the wave to pass, and then resume your shuffling until the water is deep enough for you to swim. Allow incoming waves to pass over you, and move through the surf zone quickly.

From a commercial dive boat, you can enter the water from the side or from a water-level platform at the back of the boat. Have all equipment in place, breathe

Figure 6.14 When entering surf, shuffle sideways (keeping knees bent), hold your mask, and lean into waves.

Courtesy of Fred Humphrey.

from your regulator, and hold your mask securely. Have any specialty items, such as a camera, handed to you after you are in the water; another option is to retrieve these items from an equipment line. When entering from the side of a boat, note the direction of boat movement. Wind will cause an anchored boat to swing from side to side. If you enter the water on the side of the boat when the boat is moving in that direction, the boat may pass over you after you enter.

You can do a controlled, seated entry from a dock, from a swimming platform on a boat, or from any surface where you can sit close to the water, whether the water is only a few feet deep or too deep for standing. With all equipment in place, turn and place both hands on one side of your body on the surface that you are seated on. Then lift yourself slightly, move your body out over the water, and lower yourself into the water. A controlled, seated entry is a simple and easy entry (see figure 6.15).

Use feet-first entries when the distance to the water is too high for a seated entry, such as when entering from a commercial charter boat. Two types of feet-first entries can be used: giant stride and feet together.

Use the giant-stride entry, shown in figure 6.16, when the distance to the water is about 3 to 5 feet (1 to 1.5 m) and you want to remain at the surface during the entry. Stand at the entry point with your equipment in place and your BC partially inflated. Observe the point of entry and make sure the area is clear. Hold your mask firmly with one hand, spreading your fingers so you can see. Look straight ahead while stepping out with one leg. The entry is a step, not a hop or a jump. If you step out as far as you can, your trailing leg follows automatically. Keep one leg extended forward and the other leg extended backward until you hit the water; then pull your legs together quickly to stop your downward momentum. As soon as you stabilize at the surface, turn and signal that you are OK, and then move away from the entry point so that the next diver can enter.

Figure 6.15 Entering the water from a seated position.

Figure 6.16 Giant-stride entry.

Use the feet-together entry to avoid discomfort from the impact when the distance to the water is too high for a giant-stride entry. The procedures are the same as for the giant-stride entry, except that after you step out from the entry point, you bring your legs together before you hit the water. A feet-together entry submerges you. After you bob back to the surface and stabilize, signal the next diver and clear the entry area.

A back-roll entry can be done from either a seated or squatting position. Use the seated back-roll entry from a low, unstable platform, such as a small boat. You can also use a back-roll entry when the distance to the water is too high for a regular seated entry and the platform is too unstable for you to stand. To do a seated back roll, sit with all equipment in place and your back to the water. Move your bottom to the very edge of the surface on which you are sitting. Have someone make sure the entry area is clear. Hold your mask with one hand and your mask strap with the other hand. If you do not hold the mask strap, the force of the water may wash it up over your head, and your mask may fall off. Lean backward to begin the entry. Hold your knees to your chest as you roll backward; if you don't, you may clip your heels on the edge of the surface you were seated on (see figure 6.17). You are likely to do a backward, disorienting somersault in the water with this entry. You can reorient yourself when you bob back to the surface.

Figure 6.17 Entering with legs extended risks clipping the heels on the boat. Tuck the legs into the chest instead.

Christian Wheatley/Getty Images

A squatting back roll is used when there is no suitable surface on which to sit. The thin side of a small, rocking boat is a good example of an unsuitable seat. For a squatting back-roll entry, prepare yourself while sitting on an adequate seat immediately adjacent to the entry area. Make sure the entry area is clear, then stand partially, turn your back to the entry area, and literally sit down into the water (see figure 6.18). Pull your knees to your chest as you enter so that you do not catch your heels.

Figure 6.18 Squatting back-roll entry.

Surface Snorkeling With Scuba Equipment

You may use your snorkel for surface swimming in order to conserve the air in your scuba tank. This procedure is used in areas where divers need to swim from shore to reach the dive area. If you use air from your cylinder for surface swimming, you will reduce the amount of time you can remain underwater. When snorkeling while wearing scuba equipment, you should inflate your BC partially. Excessive air in your BC combined with the weight of your scuba tank

Three Important Scuba Diving Rules

1. Equalize pressure in your ears early and often.
2. Always breathe; avoid breath holding.
3. If the scuba regulator is out of your mouth, exhale continuously.

can cause you to roll sideways when swimming facedown. It is better to swim on your back or side; these positions allow wider kicks than when swimming facedown, and the tank weight is not a factor. If you swim on your back, you can breathe without the snorkel. Swim beside—not behind or in front of—your buddy. Move slowly and steadily, pacing yourself. If you feel short of breath after snorkeling, rest before beginning your descent.

Recovering and Clearing the Regulator

When you are in the water wearing scuba equipment and have removed your regulator second stage from your mouth—such as when snorkeling or talking at the surface—you need to recover it from behind your right shoulder (where it falls naturally). The second stage will have water inside, and you must clear the water before you breathe. If you start with the regulator in your mouth, you need to place the regulator in the water. The task is more difficult than it sounds because if you place the second stage in the water with the mouthpiece facing up, the regulator free-flows. Place the second stage in the water with the mouthpiece facing downward to prevent free-flow.

Two methods can be used to recover the regulator second stage from behind your shoulder. The most popular technique is the sweep method. Lean to the right side so gravity swings the second stage away from you. Reach back with your right hand until you touch the bottom of your scuba tank, and then extend your arm and sweep it forward in a large arc. The hose will lie across your arm, where you can retrieve the second stage easily.

The second method of recovering the second stage is the over-the-shoulder reach. Reach back toward your regulator first stage with your right hand while lifting the bottom of your scuba tank with your left hand. Grasp the second-stage hose where it attaches to the first stage and follow it down to the second-stage end. Some divers find this recovery method difficult or impossible.

Clearing the regulator can be as simple as exhaling into it. As long as the exhaust valve is at the lowest point, the water inside the second stage will be displaced. If the exhaust valve is not at the lowest point when you exhale, only part of the water may be exhausted, and you may inspire some water when you inhale. To avoid choking on inhaled water, find out where the exhaust valve is on your regulator, be sure to make it the lowest point when you clear your regulator, and inhale cautiously after clearing the regulator.

Another way to clear a regulator second stage—the purge method—involves using low-pressure air to clear the water from the chamber. If you insert the mouthpiece into your mouth and depress the purge, you may blow water down your throat. You can prevent this in two ways. One way is to depress the purge lightly as you place the bubbling regulator in your mouth, and to use back pressure from your lungs to keep water out of your mouth and throat. The other method is to place the regulator into your mouth, block the opening with your tongue, and depress the purge momentarily to clear the chamber. Either method

is acceptable. If you are purging the regulator while placing it in your mouth, release the purge the moment the regulator is in place to avoid overinflating your lungs. Develop the habit of blowing a continuous stream of tiny bubbles any time the regulator is out of your mouth to avoid holding your breath while breathing compressed air.

Performing Snorkel–Regulator Exchanges

You must be able to switch from breathing from your snorkel to breathing from your regulator and vice versa. When you prepare to descend for diving, you exchange your snorkel for your regulator; when you surface after a dive, you exchange your regulator for your snorkel. Do both exchanges with your face in the water. Take a breath, exchange one mouthpiece for the other, and clear the new breathing device.

Inflating and Deflating the BC

You should be familiar with two ways to inflate a BC and two ways to deflate it. The easiest and most commonly used means of inflation is the low-pressure inflator, which is used to add air to the BC in short bursts. Inflating the BC for several seconds can cause serious buoyancy control problems if the inflator valve sticks. By adding air a little at a time, you have better control of your buoyancy.

If your low-pressure inflator or your integrated second stage develops a problem while you are diving, you may need to disconnect the low-pressure hose. When you disconnect the hose, the low-pressure inflator no longer functions, so you have to control buoyancy by orally inflating the BC. The mouthpiece of a BC is usually more elaborate than the mouthpiece of a skin diving vest. To help keep water out of a BC, the mouthpiece has a purge so that the water inside can be cleared before you open the valve to the BC. To clear the BC mouthpiece and orally inflate the BC, follow this procedure:

1. Insert the mouthpiece into your mouth.
2. Exhale a small amount of air into the mouthpiece to clear it.
3. Keep the mouthpiece in your mouth.
4. Depress the manual inflator-deflator valve.
5. Exhale into the BC.
6. Release the manual inflator-deflator valve.
7. Repeat the procedure until you achieve the desired amount of buoyancy.

You can use the same bobbing technique at the surface that you learned for the skin diving vest (see the previous explanation in this chapter), but oral inflation procedures are different underwater. To inflate a BC orally underwater, follow these steps:

1. Grasp the inflation valve with your left hand, and grasp your regulator second stage with your right hand.
2. Take a breath.
3. Insert the BC mouthpiece into your mouth.
4. Clear the mouthpiece.
5. Exhale most of the air in your lungs into the BC.

6. Save enough air to clear the regulator, which fills with water when you remove it from your mouth.
7. Repeat the procedure until you achieve the desired amount of buoyancy.

You can deflate your BC with the manual inflator-deflator valve or with a dump valve. If you do not use a dump valve, you must open the deflator valve at the lower end of the BC hose and hold it higher than the highest point of the BC. Using a dump valve is more convenient than using the deflator valve. Note that you can deflate your BC only when the exhaust port is the highest point. Air cannot escape if you try to deflate a BC in a horizontal or inverted position, so you need to be in an upright position to deflate your BC.

Testing Buoyancy

Similarities exist between buoyancy testing for skin diving and buoyancy testing for scuba diving, but there are also some important differences. Buoyancy varies more when you scuba dive than when you skin dive. The volume of air in your lungs varies more in scuba diving. You dive deeper, so suit compression affects your buoyancy more. As you use air from your scuba tank, your buoyancy changes. You need to sense buoyancy changes quickly to maintain control of your buoyancy.

Your initial buoyancy test for scuba diving can be the same as your buoyancy test for skin diving. Begin by testing your BC at the surface. Inflate it fully and then deflate it. Make sure the low-pressure inflator and all deflation valves function correctly. With your regulator in your mouth and your BC completely deflated, relax and breathe slowly. When your lung volume is high, you should remain at the surface with your eyes just below the surface–water interface. If you sink with your lungs full of air, you need to remove some weight. When you exhale completely, you should sink. If you cannot sink after a complete exhalation, you need additional weight. Test your buoyancy while you are close to your point of entry, and correct any buoyancy problems before you attempt to dive.

As you descend in open water, you may need to add air to your BC in order to maintain neutral buoyancy. Strive to maintain neutral buoyancy continuously. Add air to your BC in small amounts, and test your buoyancy by stopping all motion and observing what happens. If you are sinking, inhale and add air to your BC. If you are rising in the water, exhale and release air from your BC. With experience you will know when to add air to your BC, when to vent air from your BC, and how much air to add or release.

You can learn buoyancy control in a swimming pool by assuming a rigid, face-down position with your arms at your sides. If your buoyancy is correct, a slow, full inhalation raises your shoulders while your fin tips remain on the bottom; a slow, complete exhalation causes your shoulders to sink. Some people call this buoyancy evaluation **diver's push-ups** (although you do not use your hands). If you do not rise with a full inhalation, you should add a small amount of air to your BC and try again. Figure 6.19 illustrates a diver doing diver's push-ups.

As a scuba diver, you need enough weight to get down at the beginning of a dive and enough weight to allow you to remain in control when you ascend at the end of the dive. The precise weight for scuba diving is the amount you need to hover at a depth of 15 feet (4.6 m) with 300 psi (20 atm) of air in your tank and no air in your BC. This amount may overweight you slightly at the beginning of a dive. Test your buoyancy at a depth of 15 feet at the end of a dive to see if you need to make an adjustment for your next dive.

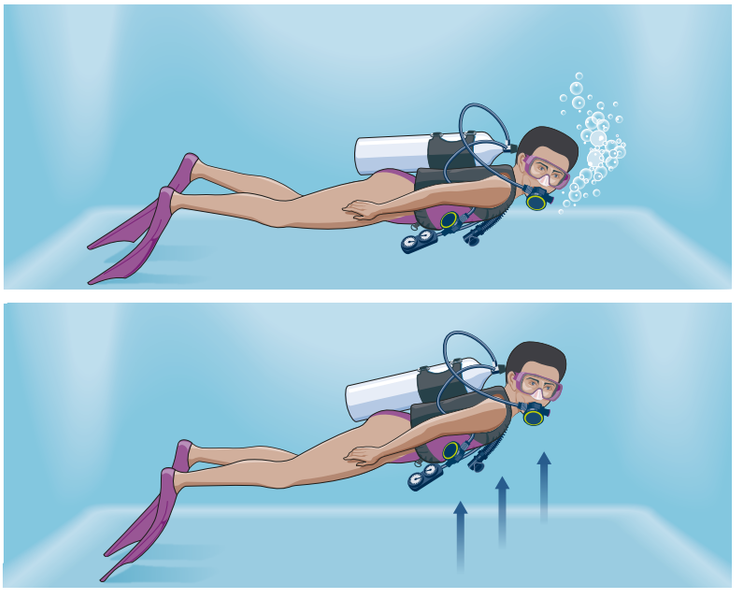

Figure 6.19 Executing proper diver's push-ups ensures good buoyancy control.

Controlling Buoyancy

No skill identifies a scuba diver's ability as much as buoyancy control. The ability to finely control buoyancy is important for safety, for enjoyment, and for the welfare of the environment. When your buoyancy is out of control, a hazard exists for both you and the environment around you.

With your buoyancy adjusted to the point where you can pivot slightly up and down on your fins on the bottom (in a pool) while inhaling and exhaling, push yourself about 2 feet (0.6 m) off the bottom and remain motionless. You may or may not remain in a horizontal position, but that is not important. Maintain your depth by controlling your average lung volume, but remember to breathe continuously. If you are sinking, keep more air in your lungs. If you begin rising, reduce the amount of air in your lungs. With practice, you will be able to hover motionless just off the bottom.

To demonstrate mastery of hovering, assume a vertical position in the water. Cross your legs at the ankles, grasp one wrist with the opposite hand, and remain motionless. Find an eye-level reference and develop the ability to hover motionless while upright in the water. Pay attention to the buoyancy effects of breathing. Once you master hovering both horizontally and vertically, you will realize several benefits. Your air will last longer, swimming will not tire you as much, you will need fewer buoyancy adjustments, and you will do less damage to the diving environment. Your buoyancy control continues to improve with practice until you become a highly experienced diver.

A useful skill for maintaining neutral buoyancy during ascent is a BC venting technique called an **open-valve ascent**. When you ascend, air in your BC expands and affects your buoyancy. You must release the expanding air to control buoyancy. If you vent the air from time to time, your buoyancy is in a constant state of change: You may not release enough air, or you may release too much. The open-valve ascent is a better alternative. Hold the BC inflator-deflator valve just below the level of your shoulder, point the mouthpiece of

the inflator-deflator valve downward, and open the deflator valve. Air cannot escape because the opening is lower than the exhaust port on the BC. While you hold the mouthpiece downward and the deflation valve open, raise the inflator-deflator slowly until air just begins to bubble from the mouthpiece. Bubbling occurs when the mouthpiece and the exhaust port on the BC are at the same level. If you hold the inflator-deflator valve at the bubbling level while ascending, the expanding air inside the BC bubbles out through the open valve, and your buoyancy remains constant. If you need to release additional air to control buoyancy, raise the inflator-deflator slightly; if you need to decrease the amount of air you release, lower the assembly slightly.

Using the Dive Flag

You should display the appropriate dive flag or flags when diving. Authorities in some areas require you to use a flag, and flag use is encouraged in all areas. Follow these conventions when you use a dive flag: Display the flag only when divers are in the water, and surface within 100 feet (30 m) of your flag—the closer the better. The flag does not guarantee your safety from boaters, who are supposed to give the flag a berth of at least 100 feet; however, the flag does serve as a signal to many boaters who recognize that the flag means there are divers in the vicinity. Figure 6.20 shows a boat anchored with a dive flag indicating there are divers in the area.

If you support your flag with a surface float, you need to tow your float to the area where you wish to dive. In areas lacking underwater plants, you can tow the float while you dive. In areas with plants, you need to anchor the float to keep it from drifting away, and you must navigate back to the flag at the end of your dive. When entering through surf, you should tow the surface float behind you. Push the float ahead of you when exiting through surf.

Figure 6.20 Surface as close as possible to the dive flag to keep you safe from other boaters.

Descending

Scuba descents are different from skin diving descents. You usually descend headfirst as a skin diver, but scuba divers descend feet-first. A feet-first descent allows better control of buoyancy, provides better orientation, allows buddies to maintain contact during descent, helps prevent the swallowing of air, and allows easier pressure equalization of air spaces.

Descending involves several procedures, which seem complex at first but become routine with practice. The process of descending includes preparing to descend, initiating the descent, and completing the descent.

1. Preparing to descend
   - Check your instruments. Orient yourself at the surface and set a reference on your compass. Be sure your underwater timer and depth gauge are zeroed.
   - Confirm that your buddy is ready to begin.
   - Exchange your snorkel for your regulator.
   - Pressurize your ears slightly to begin the equalization process (see chapter 3).
   - Hold your BC deflator valve in your left hand.
   - Give or acknowledge the signal to descend.
2. Initiating the descent
   - Begin the descent by venting your BC. It is better to do this with the dump valve than by holding the deflator valve above your head. Hold the inflator-deflator valve in your left hand throughout the descent so you can add or release air from your BC at any time. Exhale fully to help get started downward.
   - Breathe shallowly for the first 10 feet (3 m).
   - Equalize pressure in your ears about every 2 feet (0.6 m) for the first 15 feet (4.6 m). If you experience equalizing problems, ascend a few feet to reduce the pressure, equalize again, and descend again. Exhale some air into your mask to prevent a mask squeeze. (See chapter 3 for a description of pressure equalization.)
   - Keep your fins still while you descend so that you do not stir up silt on the bottom.
   - Control your rate of descent by the average amount of air you keep in your lungs. When you begin sinking while your lungs are full, add a short burst of air to your BC to regain neutral buoyancy.
3. Completing the descent
   - Remain with your buddy throughout the descent.
   - Avoid contact with the bottom.
   - Hover above the bottom, level off to a swimming position, agree on a direction with your buddy, and begin your exploration.

Two types of descents are performed in open water: **reference descents** and **nonreference descents**. A reference descent is one that you control by following a line or the slope of the bottom. You should do a reference descent whenever possible. A descent made vertically in water without any external reference is a nonreference descent, which is more difficult to control than a reference descent.

Clearing the Mask

To clear a mask while scuba diving, you need to develop the skill of breathing through your mouth with your nose exposed to water. With concentration and practice, you can master this skill quickly. First, try inhaling through your mouth and exhaling through your nose. Next, try inhaling and exhaling through your mouth. If you feel any water going up your nose, exhale immediately to keep the water out.

To practice clearing a mask, you need to flood it with water. This is not difficult, but a few tips make it easier. If you exhale lightly while tilting the mask forward on your face to break the seal at the top, the mask releases from your face easily, and the air escapes at the highest point. When you reseat your mask to begin clearing it, be sure to hold back any strands of hair with one hand while you reseat the mask with the other. Hair under the mask causes leakage. If you are wearing a hood, make sure that the hood is clear of the mask before you attempt to clear water from the mask.

Buddy Diving

You should always dive with a companion, using the **buddy system**. A buddy provides reminders and assistance and sees things that you might not see. Buddies inspect each other's equipment, provide feedback based on observations, and work as a team. Dive buddies should remain close enough to each other that one can immediately assist the other in the event of an emergency. The more turbid the water and the greater the depth, the closer buddies should remain to each other. During your training, you should strive to remain within touching distance of your buddy at all times; learn to keep track of your buddy. Maintaining contact with a dive buddy in open water is not difficult when you follow a few standard procedures.

Agree on a position relative to one another and maintain that position as much as possible. That way, your buddy will know where to look for you, and you will know where to look for your buddy. The best dive team configuration is side by side; the least desirable is for one diver to be above and behind the other. Looking up and back is difficult when you are hampered by your equipment.

Agree on a direction of movement. Both parties should maintain that direction until both agree to proceed in a different direction. When you follow this practice, there are fewer directions to consider if you and your buddy become separated.

Confirm your buddy's position every few seconds. If you scan the areas ahead from side to side while swimming, you should get a glimpse of your buddy each time you turn your head in your buddy's direction.

When visibility is poor, physical contact can keep you and your buddy together. Holding hands is appropriate. Or you can use a short line—a **buddy line**—to keep in contact with each other. If you become separated from your buddy underwater, look for your buddy for up to one minute. If you are unable to locate your buddy, ascend slightly and turn in a circle while looking for bubbles. The visibility is often better a few feet above the bottom than it is at the bottom. If you do not see your buddy's bubbles, ascend to the surface and wait for your buddy, who should also duplicate this procedure. When you have reunited at the surface, you can descend again and continue your dive. Obviously, it is better to remain together underwater than to surface to reunite.

The Buddy System

Teaming up with a qualified partner is appropriate for activities that involve risk, such as swimming, rock climbing, and skin and scuba diving. Safety is the primary purpose of the buddy system, but having a dive buddy also makes diving more enjoyable. The following list describes the duties and responsibilities of a good dive buddy:

- Helps to plan the dive in advance and after arrival at the dive site
- Reviews signals and emergency procedures
- Assists with the donning of the scuba unit
- Inspects equipment
- Maintains contact while diving and follows procedures for reuniting if separation occurs
- Provides reminders concerning depth, direction, time, air pressure, and ascent rate
- Points out items of interest
- Identifies problems that you are unaware of, such as an air leak
- Provides reassurance and assistance as needed
- Summons additional assistance as needed
- Provides first aid as needed
- Assists with exits and the removal of the scuba unit

If you are unable to relocate your buddy underwater and your buddy does not surface promptly, look at your surroundings to mark your position so that you will know the approximate location where your buddy was last seen. If someone is overseeing the diving operations, notify this person that your buddy is missing so that a search can be initiated. If you are alone, try to locate bubbles that could indicate your buddy's position.

Underwater Swimming

When you swim along the bottom, your fins can raise a cloud of silt that harms the environment and reduces visibility. Silting is more of a problem when you are overweighted because the excess weight angles your fins toward the bottom when you swim (see figure 6.21). The first step in reducing silt is to weight yourself properly. In areas where the bottom silt is thick, add air to your BC to make yourself slightly buoyant underwater. The buoyancy forces you to swim at a slight downward angle and keeps the thrust of your fins directed upward. Another way to reduce silt is to remain far enough from the bottom to keep from disturbing it.

Figure 6.21 Overweighting can cause divers to stir up a cloud of silt.

Finally, consider changing your kick if you boil up silt with your kick strokes. You should also consider redistributing your weight by placing a small portion of it higher on your body to achieve level trim as described in the "Mobility Adjustments" section of chapter 3.

When you kick something while swimming, you must overcome the tendency to want to get away from whatever you kick. Your kicks are strong and can damage the environment or another diver. As soon as you feel something with a fin, stop kicking, look back to see what you have hit, and maneuver yourself clear before proceeding.

Navigating

To find your way underwater, you can use natural navigation and compass navigation. You can best determine your relative position with a combination of both types of navigation.

With natural navigation, you use your natural surroundings to determine where you are. Light, shadows, plants, formations, water movement, depth, and other indicators can help you navigate. As you move, note your surroundings. Observe which way you are going relative to the movement of the water, the sand ripples on the bottom, the depth contour, and the angle of the sun. By noting natural aids to navigation, you can find your way underwater.

A dive compass increases the accuracy of navigation. Figure 6.22 shows two types of dive compass. You need to be able to set a direction (called a **compass heading**) and to determine which way you are going relative to the directional reference you have set. Your compass should have a reference line, called a lubber line, that you point in the direction of travel. The north-seeking needle or card of the compass establishes a position relative to the lubber line as long as you hold the compass in a level position. Many dive compasses have index marks on a movable bezel. You set the index marks to indicate the heading. Electronic compasses allow you to set a heading and use arrows to indicate the direction that you need to turn to remain on course.

Figure 6.22 Traditional dive compass (left) and electronic dive compass (right).

To go in the direction set on the compass, you must hold the compass so the lubber line is directly in line with the centerline of your body. If the lubber line points to one side, you will not be on course even though the north reference is at the correct point on the dial. Figure 6.23 shows a compass lock position.

A **compass course** is a series of headings that leads to a destination. Many types of compass courses may be used. One frequently used course is the **square compass course**. To navigate a square course, set your initial heading and proceed in that direction for a given distance, which can be measured by time, tank pressure, or fin kicks. Stop and turn 90 degrees to the right while continuing to keep the lubber line aligned with the centerline of your body. Note the

relative position of north on the compass, and proceed in the new direction for the same distance as you covered on the first leg of the course. Stop again, and turn another 90 degrees to the right. Note the position of north on the compass, which should be opposite your initial heading. Proceed along the third leg of the course for the same distance as before. Stop once more, turn again 90 degrees to the right, note the relative position of north, and follow the new heading back to your starting point.

Figure 6.23 The compass-lock position.

Divers also frequently follow a **reciprocal compass course**—an out-and-back course. Set the initial heading on the compass. At the midpoint of the dive, turn 180 degrees until north on the compass is directly opposite the original heading. Then follow the reciprocal heading back to your starting point. Advanced scuba diving courses provide additional compass navigation training.

If you do not know precisely where you are when the end of a dive approaches, you may need to surface, find a reference for your exit, and set a compass heading that leads directly to the end-of-dive location. Be especially careful if you surface more than 100 feet (30 m) from your dive flag.

A compass provides correct directional reference information when it is not affected by nearby objects. Metal objects, other compasses, and electrical fields within a couple of feet of a compass can cause it to deviate from its correct reading. Keep metal, magnets, dive lights, and other compasses away from your compass to help ensure accuracy.

Monitoring Instruments

Most diving instruments are passive; that is, they do not provide information unless you look at them. Some instruments emit an audible beep, but most require observation to provide information. Develop the habit of checking your instrumentation frequently while diving so that you can control your depth, dive time, and direction. This will also help you avoid running out of air. You should be able to accurately estimate your tank pressure at any time during training. If you cannot estimate the pressure within 300 psi (20 atm) at any time, you need to monitor your SPG more frequently.

When you are planning to dive in open water, you should look at your instruments when you assemble your equipment. Look at your instruments again when you inspect your equipment. Look at your instruments again before you descend and again while you are descending. Refer to your compass for directional reference before you begin moving underwater. Monitor your gauges every few minutes while diving, and compare your air pressure with your buddy's several times during a dive. At any given time during a dive, you should be able to accurately estimate your depth, your dive time, your direction, your tank

pressure, and your buddy's tank pressure. If you cannot do this, you need to improve your instrument-monitoring skills.

Using Scuba Diving Hand Signals

Scuba divers use several hand signals that are not used for skin diving. The signals specific to scuba diving relate to air supply. Learn and use the standard hand signals shown in figure 6.24. Remember to display hand signals clearly and deliberately and also to acknowledge the hand signals you receive.

Ascending

Scuba ascents are different from skin diving ascents, although there are a few similarities. The procedures you need to learn become automatic with practice and experience.

To initiate an ascent, one member of a buddy team gives the ascent signal, which the other acknowledges. Always obey the ascent signal. Prepare to ascend by noting your time, depth, and remaining air. Locate your BC inflator-deflator assembly, and hold it in your left hand. Begin the procedures for the open-valve ascent previously described (see "Controlling Buoyancy").

Begin ascending slowly with your buddy while breathing continuously. Monitor your depth gauge and keep tabs on your buddy. The maximum rate of ascent is 0.5 feet (0.15 m) per second, which is quite slow. Some instruments warn you when your rate of ascent is too rapid. You need training, practice, and awareness to avoid exceeding the maximum rate of ascent.

Stop and decompress (outgas) when indicated by your dive computer. Whether or not you make a "deep stop," you should stop for one to three minutes at a depth of 15 feet (4.6 m) to help prevent decompression sickness (DCS). The procedures for decompression are described in chapter 7. As you ascend, look up and around. Extend one hand above your head for protection against overhead obstacles. Make one full rotation to view the surrounding area as you near the surface. When you reach the surface, make another rotation to view the area, and then inflate your BC to establish buoyancy. Exchange your regulator mouthpiece for your snorkel.

Handling Equipment

Sometimes you will need to remove, adjust, and replace scuba equipment while you are in the water. You may need to remove equipment to exit the water onto a boat, to make an adjustment, or to free the equipment from an entanglement. With training and practice, you should be able to handle your equipment easily.

Removal of the scuba unit is easy because it is similar to removing a coat. Open the releases, slip your left arm free, swing the scuba tank forward under your right arm, hold the scuba unit with your left hand, and pull your right arm free. Freeing your left arm is easier if you insert your hand and wrist through the armhole of the BC first and remove it hand first (instead of pulling your arm through first). If you are at the surface in water too deep for you to stand when you want to remove your scuba unit, you should remove your weight belt first and place it on a surface float or support station. Don the scuba unit in the water according to the in-water donning procedures previously presented in this chapter.

Figure 6.24 Scuba diving hand signals.

Using Exit Techniques

The technique you use to exit the water depends on the situation. To exit from shallow water in a swimming pool, begin by removing your weight belt, tank, and fins, in that order. Carefully place the equipment on the side of the pool; then climb out by using the ladder or lifting yourself up onto the edge of the pool.

To exit from the deep end of a swimming pool (as shown in figure 6.25), begin with your BC inflated partially. If you exit on a ladder, grasp the ladder with one hand and remove your fins with the other; maintain contact with the ladder at all times. Place your fins on the edge of the pool, or slide the heel straps over your wrists; then climb the ladder to exit the water. Clear the exit area at once, and take your fins with you. Remove your scuba system and weight belt when you are well clear of the exit area.

Figure 6.25 Exiting from the deep end using a ladder.

To exit from deep water without a ladder, begin by removing your weight belt and carefully placing it out of the water. Remove the scuba unit and use one hand to trap the regulator hose against the surface onto which you will exit. Use your other hand to remove your fins, and place them out of the water. Place both hands on the exit edge. With the regulator hose trapped under one hand, lower yourself to about chin level in the water while you extend one leg forward and one leg backward. Pull yourself upward with your arms, and pull your legs together forcefully in a strong scissors kick to provide upward momentum. Pull with your arms until you are far enough out of the water to push downward and lift yourself from the water (see figure 6.26). Immediately after your exit, turn around and pull your scuba unit from the water carefully.

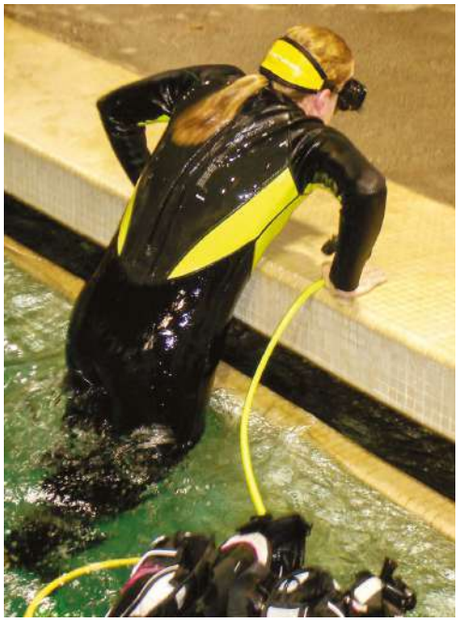

Figure 6.26 Exiting from the deep end without a ladder.

If you are boarding a boat that has a ladder, you should keep your tank and mask on and keep your regulator in your mouth in case you fall back into the water (see figure 6.27). Use the ladder exit technique described for exiting the deep end of the pool. Maintain contact with the ladder at all times when you

are in the water. If the boat has a platform at the rear, you usually swim onto the platform, remove your fins, and then stand on the platform and board the vessel.

Techniques for wading exits in open water vary with the environment. You should usually wear all your equipment until you are clear of the water. Shuffle your feet along the bottom while moving backward. Surf exits require training and practice. Stop outside the breaking waves and evaluate the surf. Approach the surf zone with your regulator in your mouth and your BC deflated. Hold your mask continually in the surf zone. Follow a breaking wave, and allow additional waves to pass over you until the water is only a couple of feet deep. If the surf is mild, you can stand at that point and back out of the water. If the surf is strong, swim until you can crawl; then crawl clear of the water (see figure 6.28).

When you are clear of the water, work with your buddy to remove your fins. The buddy system is in effect even when you are out of the water.

Figure 6.27 Boarding a boat with a ladder.

Figure 6.28 Using a crawling exit in strong surf.
Courtesy of Fred Humphrey.

Disassembling Equipment

The first step in disassembling your scuba equipment is to turn off the air by turning the valve in a clockwise direction. Release the pressure in the hoses by depressing the purge on the regulator second stage. Keep the purge depressed until you bleed all the air from the system. Next, disconnect all hoses that are connected to the scuba unit: the low-pressure inflator, the extra second stage, and the SPG. Loosen the yoke screw and remove the regulator from the tank. Dry the first-stage dust cover thoroughly and replace it. Loosen the tank band and remove the BC from the scuba tank. Be sure to rinse and care for your equipment as soon as possible (following the procedures described in chapter 4).

Managing Physiological Problems

If you do everything you are trained to do as a scuba diver, you can avoid problems. But it is not a perfect world. If you fail to pay attention or forget to do something, a problem can occur. Good divers can deal with nearly any problem. This section introduces you to proven ways of dealing with potential diving difficulties. Do not be overly concerned about the problems presented. You can prevent them, but knowing how to deal with them helps reduce your apprehension.

Difficulties affecting your physiology include seasickness, dizziness, stress and panic, overexertion, coughing, and cramping. When you have one of these problems, your body sends messages that something is wrong. You need to know the messages and the actions you can take to overcome physiological difficulties.

Seasickness

The best strategy is to try to prevent seasickness (see "Equilibrium Adjustments and Seasickness" in chapter 3) because taking medication after you are seasick is usually ineffective. If you do get seasick, you are likely to vomit. Vomiting underwater can be dangerous because of involuntary gasping that can cause you to choke. Only you can determine your degree of nausea. If you throw up or feel as if you are on the verge of doing so, do not dive. If you feel queasy, getting into the water may help you overcome the feeling. Some divers who feel slightly nauseated find that they feel better if they get into the water quickly and dive. After that first dive, they are fine for the remainder of the day.

You should surface if you feel nauseated while diving. If you must vomit underwater, do not vomit through your regulator. Hold the second stage against one corner of your mouth, and depress the purge fully while you vomit. You should get air instead of water if you gasp. When you have finished throwing up, place the second stage in your mouth, clear it, and resume breathing. The purge method should be a last resort. Vomiting through the regulator is not recommended because this method sometimes results in clogging.

If you are seasick and have to throw up while aboard a boat, do it over the rail on the side of the vessel opposite the wind. Do not use the restroom (or head, as it is called on a boat). The best remedy is to get to land, rest until you feel better, and take seasickness medication before returning to the vessel. If you are ill from motion sickness—and if the dive boat has a dinghy and there is land nearby—you should request to be taken to shore for a while.

Dizziness

The absence of visual clues in a weightless environment can cause temporary dizziness. Visual references can help you prevent disorientation (see chapter 3). Injury, temperature changes, and pressure changes affecting the inner ear can also cause a whirling feeling called vertigo, which may be more difficult to overcome than dizziness caused by disorientation.

To cope with either dizziness or vertigo, you should first seek a fixed visual reference. If possible, make physical contact with something solid for a point of reference. If there is nothing to see or to grasp, close your eyes and hug yourself. In most cases, dizziness passes in a minute or two. If you then move slowly and keep your head still, you should be able to surface. A good buddy can recognize your difficulty and assist you.

Stress and Panic

Stress is the perception of a substantial imbalance between environmental demand and a person's response capability, and it occurs under conditions where failure to meet demand is perceived as having serious consequences. Stressors are conditions or attitudes that cause stress. Dive stressors, which can be internal or external, include cold, illness, exhaustion, injury, fear, equipment problems,

loss of air supply, buddy separation, depth, darkness, currents, and disorientation. Stress is not always bad. Moderate stress can cause a feeling of exhilaration, improve performance, and lead to a positive condition called eustress. Excessive stress causes anxiety, decreases performance, and leads to a negative condition called distress.

Knowledge and appraisal of a situation affect your reaction to stress. Training, experience, and your predisposition toward a situation affect your knowledge and appraisal. Thus, your perception of the circumstances is determined by what you know and are able to do, by what you have done, and by any inherent fears about the circumstances that you might have.

The problem with stress in diving is that it can lead to panic—a sudden, uncontrolled, irrational reaction to a perceived danger. Divers who panic often perish. You must manage stress in order to prevent panic. Dr. Tom Griffiths, who has researched stress, said the following: "The most critical factor in the progression of panic after stress increases is whether or not a problem arises." Divers who can recognize and manage stress can overcome problems and are far safer than those who cannot cope with the effects of stress.

Stress has both physiological and psychological components. Anxiety causes involuntary physiological changes. When stress leads to anxiety, your breathing rate and heart rate increase, your nervous system becomes more active, and your awareness decreases. These factors decrease your performance and increase your anxiety. Then heightened anxiety begins the cycle again. Unless you interrupt the cycle, anxiety escalates into panic. Psychological difficulty is every bit as critical as the physical problems. Frank Pia, a chief lifeguard, says, "Much of the distress that a person experiences when difficulty arises stems from what the person tells himself about the situation." You can be exhilarated by a situation or allow it to cause anxiety and distress. Pia continues, "The difference between panic and a heightened physiological state is the thought process."

Michael J. Asken, who has written books about stress, encourages the use of task-relevant instructional self-talk (TRIST). You determine your emotional state when in distress by what you tell yourself. Asken says that in a stressful situation, "Success is not achieved by focusing on the outcome, hoping for a good outcome, or even telling yourself that the outcome will be OK. The most effective self-talk involves imagining that your instructor is sitting on your shoulder guiding you through your response."

With the previous concepts in mind, you can recognize and manage stress. The first steps are to sense your breathing rate and identify your self-talk when a problem occurs. If your breathing rate is fast and you are having negative thoughts about the situation, it is time to break the stress cycle. Stop all physical activity, establish buoyancy, and breathe deeply. Divers in distress usually have an adequate supply of air! As you gain control of respiration, you can think more clearly. Take control of your thoughts before they take control of you. Imagine that you are telling someone else how to deal with this situation. Assess your options, determine the best course of action, and then take deliberate action. As you begin to overcome the difficulty, your confidence increases and your physiological condition begins to return to normal. Figure 6.29 shows how to break the stress cycle that leads to panic.

Now you can understand why some divers are thrilled by the same situation that causes another diver to panic. Remember that stress is simply a matter of perception and that you can change your perception with training and experience.

You can also learn to manage difficult situations by imagining them vividly in your mind. Your body cannot distinguish the difference between an actual event and one that you imagine. As you learn to solve problems in your mind, you can learn to recognize and manage stress.

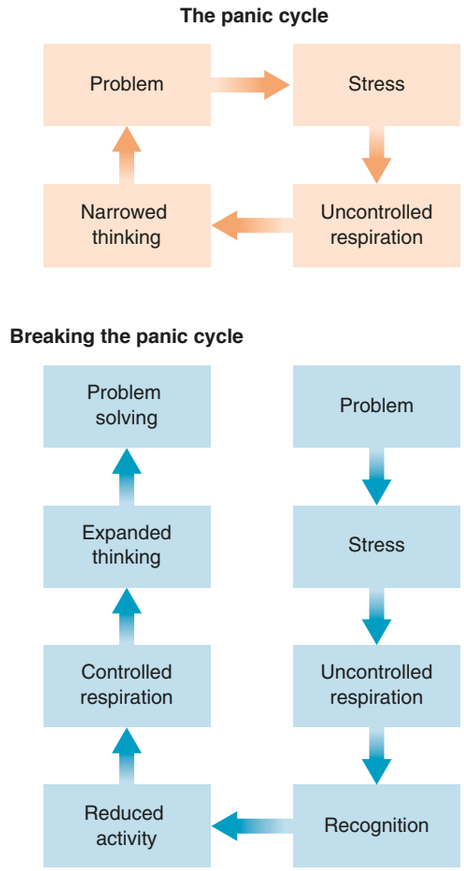

Figure 6.29 The panic cycle and how to break it.

Overexertion

If you work too hard while scuba diving, the equipment may not be able to supply enough air to meet your respiratory needs. You experience a sudden feeling of suffocation, and you may suspect that your equipment has malfunctioned. It is unlikely that scuba equipment will cease to function suddenly in a way that restricts airflow. More likely than not, overexertion is causing the feeling that you cannot get enough air. You can manage overexertion similarly to the way you manage stress. Stop all physical activity and breathe deeply; you should overcome your respiratory problem within a minute or two.

Coughing

When water strikes your larynx, a reflex coughing action tries to clear the foreign matter from your airway. When you cough while in water, you may inhale additional water, which can complicate the situation. You need to overcome the reflex as quickly as possible. The best way is to swallow hard three times in rapid succession. If you must cough, try to do so through a regulator so that you inhale air rather than water if you gasp. You can lose buoyancy as you expel air when coughing, so you may need to establish positive buoyancy if you cough excessively.

Cramps

When your muscles get cold or when circulation to your muscles is inadequate, cramping can occur. A cramp is a sudden, strong, involuntary, persistent, painful muscle contraction. Divers tend to get cramps in the lower legs and the feet. To alleviate a cramp, stretch the affected muscle and rub it to increase circulation. Pounding a cramped muscle is ineffective and causes tissue damage. If you experience a cramp in your leg or foot in the water, you may be able to get rid of it by grasping the tip of your fin and pulling it toward you. Another technique to help release a leg cramp is to contract the muscles of the opposite leg. This action can inhibit nerve impulses to the cramped muscles. Buddies should assist one another with getting rid of cramps. Figure 6.30 shows a diver relieving a cramp.

Figure 6.30 Your diving buddy can help you relieve a cramp by rubbing and stretching the cramped muscle.

Managing Physical Problems

Potential physical difficulties in diving include entanglement; loss of buoyancy control; loss of air supply; and a distressed, injured, or incapacitated buddy. As with physiological problems, you can overcome these problems, but it is better to avoid them.

Entanglement

You will probably encounter fishing line, nets, wire, string, and rope in the water—and these items can entangle you. Underwater plants can also entangle divers in some areas. Streamlining your equipment to minimize places where things can get caught helps reduce the chance of entanglement. Being aware also helps. When you encounter something that can entangle you, swim around it or push it beneath you and swim over it. Avoid swimming beneath things that can cause entanglement. The area of your tank valve and regulator first stage is one of the easiest places to become entangled and also one of the most difficult areas to free from entanglement.

If you become entangled underwater, you should stop. Then try to examine the problem without turning because turning often compounds the problem. If you can reach the entanglement, you should free yourself. If not, signal your buddy to help you get free. If you can see what you are doing, you can use your dive knife to cut yourself free. Trying to cut yourself free from something in the area of your tank valve is an obvious mistake. If your tank is entangled and you can't reach the entanglement—and your buddy is not nearby to lend assistance—you should remove your scuba unit, free it from the entanglement, and put it back on.

Loss of Buoyancy Control

You can lose control of buoyancy underwater if you lose weights or if a low-pressure inflator on your BC or dry suit sticks. You can take steps to prevent loss of buoyancy control. Check your weight system from time to time while you are

diving to make sure it is secure. Inspect your low-pressure inflators before each use, and have them serviced at the first sign of unusual operation.

If your weights fall off while you are underwater, immediately maneuver yourself into an inverted position, swim down forcefully, and retrieve the weights. If you are successful, you can avoid an uncontrolled ascent; if you are unsuccessful, you will float to the surface. Your rate of ascent depends on how buoyant you are without your weights, the type of exposure suit you are wearing, the amount of air in your BC at the time, your depth, and the amount of surface area you can expose to the direction of motion. You learned in chapters 2 and 3 that the greater the cross-sectional area of an object moving in a given direction, the greater the resistance to movement in that direction. If you lose control of buoyancy and your ascent is uncontrolled, you can slow your rate of ascent by **flaring**—arching your back, extending your arms and legs, and positioning your fins parallel to the surface. Flaring (shown in figure 6.31) is the method recommended to slow a buoyant ascent.

Unless you care for your BC inflator carefully and have it serviced annually, the inflator valve will eventually stick. If your low-pressure inflator sticks in the open position, you should first hold your BC deflator valve in the open position. Modern BCs vent air faster than the low-pressure inflator admits air. If the inflator valve remains stuck, disconnect the low-pressure hose. If an uncontrolled ascent results, flare to slow your rate of ascent.

Loss of Air Supply

You are unlikely to have air supply difficulties if you have your regulator serviced annually, if you maintain your regulator properly, and if you monitor your SPG. Potential problems include a regulator that free-flows, low air pressure, and no air to breathe. Divers have ways to deal with each of these difficulties.

Figure 6.31 Flaring is an excellent way to slow a rapid ascent.

Sand, dirt, vomit, and freezing can cause regulator free-flow. If free-flow occurs, your best course of action is to switch to your extra second stage. You could also use your buddy's extra second stage. If there is no source of air except the regulator that is free-flowing, you can breathe from it by pressing your lips lightly against the mouthpiece, taking the air you need, and allowing the excess air to escape. Look down while you breathe from the regulator to keep the escaping air from causing your mask to leak.

You are supposed to end a dive with at least 500 psi (34 atm) of air in your tank. If you are inattentive and breathe nearly all the air from your cylinder while diving, it will become difficult to get air from your regulator. Divers often refer to this situation as being "out of air," but, in reality, they are out of air only at the depth at which breathing is difficult. As you ascend, the lower ambient pressure allows you to obtain additional air from your tank. When breathing becomes difficult and your tank is nearly empty, you should use your buddy's alternate air source (AAS) or ascend while continuing to breathe shallowly through your regulator.

In the rare event that you should completely lose your primary source of air while underwater, you have five ascent options. Figure 6.32 shows the order of preference of these options.

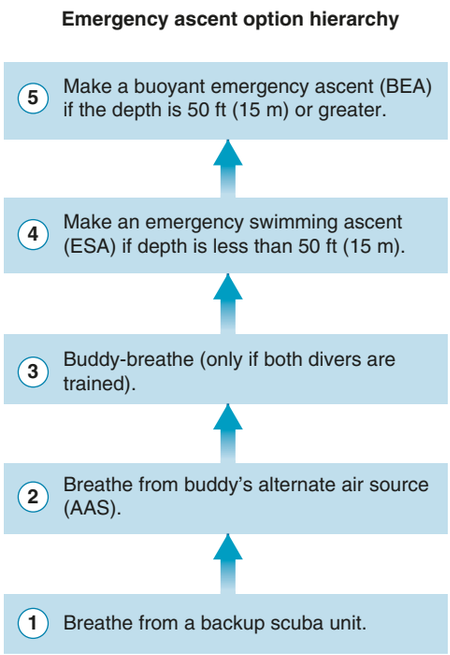

Figure 6.32 Loss-of-air options.

An extra-second-stage-assisted ascent closely approximates a normal ascent. You ascend while breathing from your buddy's extra second stage. When you require air underwater, get your buddy's attention and give the signals for "out of air" and "give me air," if possible. If your buddy's extra second stage and primary second stage are similar, your buddy will hand you the extra second stage, or you may take it. If your buddy's extra second stage is integrated into the BC low-pressure inflator, your buddy will hand you the primary second stage and breathe from the integrated second stage. If you cannot get your buddy's attention, take the extra second stage, begin breathing, and then signal your buddy that you have no air. After you establish a breathing rhythm, grasp each other's right forearm or BC and ascend normally (see figure 6.33).

An emergency swimming ascent (ESA) is a scuba ascent you do using only the air in your lungs. The ascent rate of an ESA is faster than a normal ascent, but it is not rapid. Retain your regulator in your mouth, and try to breathe from it from time to time. Do not hold your breath or you risk a lung overexpansion injury. If you exhale too much air, you will have a strong urge to inhale. The key to a successful ESA is to exhale enough air that your lungs remain at a comfortable volume. When you do an ESA correctly, you can ascend 50 feet (15 m) easily without an overwhelming desire for air. If the depth is over 50 feet, discard your weights to initiate a buoyant emergency ascent (BEA). Swim

Figure 6.33 Air sharing involves cooperation and coordination with your buddy.

for the first portion of the ascent, but allow yourself to drift up when buoyancy can replace swimming. Flare during the last 15 feet (4.6 m) of a BEA, and keep your lungs at a comfortable—but not maximum—volume.

Buddy breathing—the sharing of a single regulator second stage by two divers—is not a desirable loss-of-air option because it jeopardizes the safety of two people. You and your buddy should practice buddy breathing at the surface before beginning a dive if buddy breathing is a loss-of-air option for the dive. Not all divers are trained to buddy-breathe, and buddy breathing attempted by two divers who are not proficient with the skill can result in disaster. However, those who are proficient buddy breathers can make a loss-of-air situation a mere nuisance.

Initiate buddy breathing with the "out-of-air" signal followed by the signal for "give me air." Your buddy holds the regulator second stage in the right hand and grasps your shoulder strap with the left hand. Your buddy extends the second stage toward you and holds it in such a way that you have access to the purge. You grasp your buddy's wrist (not the regulator) with your left hand and grasp your buddy's shoulder strap with your right hand. Guide the second stage to your mouth and push your lips against the mouthpiece to make a seal instead of inserting the mouthpiece into your mouth. By not putting the mouthpiece into your mouth to breathe, you can exchange the regulator quickly and reduce mask leakage caused by facial movements. Take several quick breaths initially; then pass the regulator back to your buddy. Exhale a small amount of air continuously when you are not breathing from the regulator. The exhalation helps prevent a lung overexpansion injury during ascent. After the initial contact, you and your buddy each take two breaths before passing the regulator. Do not inhale fully when buddy breathing because full breaths can cause buoyancy problems or lung injuries. A medium inhalation is adequate because you receive air every few seconds. You should exhale fully after each first breath that you take. The complete exhalation helps expel carbon dioxide and reduces your urge to breathe.

As soon as you and your buddy establish a breathing rhythm, you should swim to the surface, holding on to each other throughout the ascent. Blow bubbles continuously when the regulator is not in your mouth. Remember to control your buoyancy.

Skills for loss-of-air situations require proficiency, so they must be learned well and renewed periodically. Discuss the procedures for a loss-of-air situation with your buddy, and agree on the options you will use. You should both be familiar with the signals, positions, and techniques.

Assisting Your Buddy

You have read about many ways to assist your buddy and how your buddy can assist you, so you should realize the importance of the buddy system while diving.

In addition to helping your buddy handle entanglements, cramps, equipment problems, and loss-of-air situations, you may need to provide assistance if your buddy is incapacitated from exhaustion, illness, or injury.

A buddy who becomes agitated at the surface needs assistance in regaining control. Help such a buddy establish buoyancy, calm down, and breathe slowly and deeply. When the situation is under control, you may be able to help your buddy resolve the difficulty that caused the agitation.

If your buddy becomes exhausted, provide assistance at the surface with a biceps push, a fin push, or a do-si-do (arm-under-arm) push. The three types of pushes are illustrated in figure 6.34. Use the biceps push when your buddy can help, and use the fin push when your buddy is too exhausted to help at all. Monitor your buddy and offer encouragement while you provide assistance. (Note: Scuba rescues require techniques different from surface assistance.)

Figure 6.34 Providing assistance to an exhausted buddy using the (*a*) biceps push, (*b*) fin push, and (*c*) do-si-do push.

Managing Emergencies

Diving accidents occur when divers do not exercise good judgment or when they fail to follow recommended practices. If you do what you are supposed to do, the chances of a diving accident are extremely small. But you may have to render aid to someone else who violates safety rules. This section identifies the aid you should be capable of providing.

Training and Preparation

Three types of emergency preparedness training are recommended for all divers: first aid, cardiopulmonary resuscitation (CPR), and diving rescue techniques. You can get first aid and CPR training from various public service organizations. You'll learn some diving rescue techniques in your entry-level course, but you should also complete a rescue specialty course.

Emergency preparedness includes having emergency equipment and information available. The emergency equipment available at a dive site should include these items:

- Diving first aid kit (see the checklist of kit items in chapter 4)
- Oxygen delivery system
- Blanket (if appropriate)
- Drinking water

You may not have all of the emergency equipment yourself, but you can determine whether it is available aboard a boat or as part of an organized dive.

Have a means of communication—a telephone, cellular phone, CB radio, or marine radio—to summon assistance. Have contact information for local emergency medical assistance and for the emergency treatment of divers at the dive site. Have phone numbers and radio frequencies for local emergency support services. Examples of numbers to have include the Coast Guard, paramedics, hospital, ambulance, police or sheriff's office, recompression facility, and the Divers Alert Network (DAN). The Divers Alert Network has a 24-hour emergency number 919-684-8111 to assist with the coordination of responses to diving accidents.

Scuba Lifesaving

The principal elements of scuba lifesaving are cognizance, assessment, rescue, and evacuation (CARE). Cognizance is the ability to detect signs that identify or predict an accident. Trained and experienced divers can recognize situations that can lead to accidents before divers enter the water, when they enter the water, at the surface, during descents, underwater, during ascents, and during exits. As your experience increases, you will become more cognizant of potential difficulties.

Assessment is an ongoing evaluation of various factors that can affect a lifesaving situation. Rescuers need to consider their own ability to rescue the victim, the victim's condition, the environmental conditions, the equipment available, and additional resources.

Rescue techniques include approaches, extensions, throws, assists, establishing buoyancy, defenses, towing, deep-water rescue breathing, equipment removal, and removing the victim from the water. The ability to rescue both conscious

and unconscious divers at the surface and underwater is an important skill to develop; you should take a scuba lifesaving or rescue specialty course as soon as possible after completing basic scuba training.

An unconscious diver underwater can drown unless rescued immediately. Illness, drugs, and blows to the head can cause loss of consciousness. If a diver loses consciousness underwater, make the diver buoyant and get her to the surface immediately. You do not need to concern yourself about expanding air if the diver is not breathing because an unconscious person exhales automatically regardless of head position. Do not concern yourself about the diver's decompression status; she can be treated for DCS but will die after about four minutes without air. Do not jeopardize your own safety when attempting to rescue another diver.

A diver who is not breathing needs air quickly and must get it at the surface. This is where your CPR skills and diving rescue techniques are invaluable. Open the victim's airway. Often this is all an unconscious person needs to be able to breathe. To open the airway, tilt the head and lift the chin. Turn the person's head to the side to drain water from the mouth and throat. Vomiting is common, so be prepared for it. Clear vomit from the victim's mouth and throat at once or the person may inhale it and choke.

A person who is not breathing after you have opened the airway requires rescue breathing and medical assistance. Shout for help. If you can remove the victim from the water quickly, you should do so. If removal will be delayed by even a couple of minutes, begin rescue breathing while you remain in the water. Keep the victim's airway open and lightly pinch the victim's nostrils to seal them. Seal your mouth over the other person's mouth and fill the person's lungs with air until the chest rises gently. Give the victim one breath every 5 seconds or two breaths every 10 seconds while swimming to safety. A gurgling sound from the victim indicates water or vomit in the airway. Roll the person to the side and drain the fluid before continuing rescue breathing. The preferred method of in-water artificial respiration involves using a rescue-breathing mask, which you can carry in the pocket of your BC. Figure 6.35 shows a person doing rescue breathing using such a mask.

You will probably not be able to detect a pulse in the water, so don't bother trying. You cannot administer CPR in the water; you must remove the victim from the water and position her on a firm surface.

Evacuation is the final element of scuba lifesaving. Evacuation procedures include accident preparedness, communications, and transportation options. Advance planning is essential to evacuate an injured diver to a medical facility.

Figure 6.35 Rescue breathing with a breathing mask.

First Aid

Some aspects of diving first aid are not taught in standard first

aid courses. This section touches on those aspects, but you need the additional study and training that you can get in rescue specialty and oxygen administration courses.

Nothing is more important than attending to basic life support: airway, breathing, and circulation. The next priority for any serious diving injury is treatment for shock. Lay an injured diver who is breathing on his side, keep the person warm (but avoid overheating), and administer sips of water if he is conscious. If you suspect an air embolism, DCS, or near drowning, have the person breathe oxygen in the highest possible concentration. Keep any diver who has lost consciousness or who has symptoms of DCS lying down until he can be evaluated at a medical facility. Monitor the victim continuously.

You should be able to recognize signs and symptoms that indicate a serious diving illness. The following signs and symptoms indicate an injury that requires the prompt administration of oxygen and medical treatment:

- Confusion
- Seizure
- Loss of consciousness
- Nausea or vomiting
- Shortness of breath
- Sudden, extreme weakness
- Numbness or a pins-and-needles sensation
- Inability to do simple motor skills
- Paralysis
- Unequal pupils

Venomous marine animals may be encountered in some parts of the world. A few venomous animals can inflict life-threatening wounds (see chapter 5). The wounds can cause pain, weakness, nausea, shock, mental confusion, paralysis, convulsions, depression, arrest of breathing, and even cardiac arrest. Fortunately, such wounds are rare.

Venomous injuries are either punctures or stings. First aid for venomous puncture wounds involves removing all foreign matter from the wound, applying hot packs to the injured area for half an hour, and keeping the injured area below the level of the heart. The injured person should obtain medical attention.

First aid for a venomous sting includes killing any stinging cells that are in contact with the skin, removing any residue, cleansing the area, applying an analgesic ointment for pain relief, and seeking medical attention. Vinegar is a good solution to apply to all stings to neutralize stinging cells initially.

You may not recall the appropriate first aid procedures in the event of an accident; therefore, you should bring along a diving first aid book to help you identify an injury and administer the appropriate first aid. A wilderness first aid book is also strongly recommended because diving often occurs in remote areas.

Managing Accidents

If a serious diving accident occurs, and no supervisory personnel are available to take charge, you have to manage the situation to the best of your ability. Summon help, but do not leave a seriously injured diver unattended. Enlist the aid of others. Try to locate the injured diver's identification and medical information. Write down what happened, as well as the person's dive profile,

SCUBA WISE

People are creatures of habit, and habits are the result of repetition. If you repeat an action correctly enough times, you form a good habit. On the other hand, if you repeat an action incorrectly enough times, you create a bad habit. Divers need good habits to avoid accidents, but sometimes they fail to take the time to develop them. I often watch divers on charter boats and at dive sites as they prepare to dive. I notice that some fail to inspect their equipment adequately, some fail to plan their dive properly, and some fail to follow the practices outlined in this chapter. They usually know what to do, but because they have bypassed many steps every time they go diving, they have a habit of skipping important items. How safe is a pilot if he ignores the preflight checklist? Taking the correct action requires concentration initially, but when repeated until the action becomes a habit, the process becomes automatic. People can make complex processes simple by repeating them. When a skill is executed properly every time, it not only becomes habit but also is easier to recall when you really need it, such as in an emergency. I strongly encourage you to take the time to form good diving habits. If you do, I assure you that the odds of being injured are extremely small.

symptoms, dive times, and so forth. Pin the information in a conspicuous place, and send it with the injured diver to the medical facility. Accompany the diver to the medical facility, if possible.

Summary

The skills of diving range from simple skin diving procedures to complex scuba skills to problem management. You need to learn the skills correctly the first time, practice them until you can do them easily, and renew them frequently to stay proficient. You also need to be trained and prepared to handle a diving emergency. As mentioned previously, visualization of skills can help you develop skills. This is especially important for problem-management skills. Some problems are not common, and you may never encounter them; but you need to be prepared to manage them. If you visualize a problem vividly in your mind, you will be able to remain relatively calm when that problem occurs because you will know what actions to take. Knowing what to do helps you remain relatively calm, which allows you to think and to better manage your situation.

Application-of-Knowledge (AOK) Questions

1. What are ways in which you can control buoyancy?
2. What do you need to remember to do while ascending from a scuba dive?
3. What can you do to minimize the chances of becoming separated from your buddy while scuba diving?
4. The inflator valve on your BC sticks when you open it to add air. What actions can you take to prevent a rapid ascent?
5. What should you do if your regulator begins free flowing while you are underwater and cannot be stopped?
6. When preparing to exit the water into a small boat, what is the correct order for the removal of your equipment?
7. While swimming against a current at the bottom in 40 feet (12.2 m) of water, you begin to experience air starvation. What is the proper way to manage this problem?
8. If you become entangled underwater, what can you do to get free of the entanglement?

Sunset at Grand Cayman

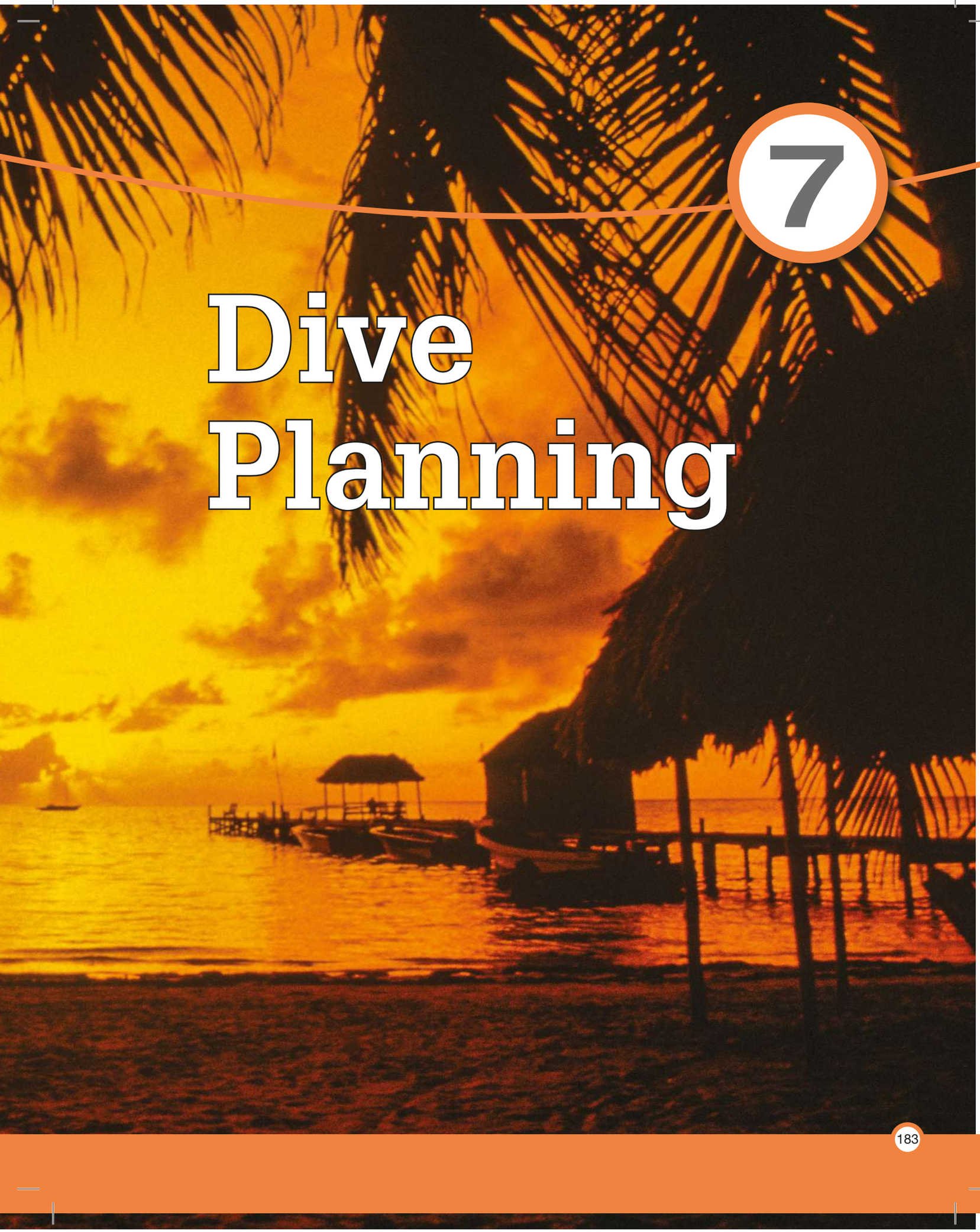

7 Dive Planning

Dive In and Discover

By the end of this chapter, you will be able to do the following:

- List at least nine factors to consider in dive planning.
- Explain the why, who, where, when, how, and what of advance planning.
- Explain the on-site planning procedures for a dive.
- List at least five methods of obtaining area orientations for scuba diving.
- Use dive tables to plan repetitive dives that do not require decompression.
- Explain the planning procedures for cold or strenuous dives, variations in ascent rate, multilevel dives, omitted decompression, diving after required decompression, going to altitude after diving, exceeding maximum time when doing precautionary decompression, and a repetitive-dive residual nitrogen time that exceeds the absolute bottom time of the previous dive.
- Define the terms *residual nitrogen*, *repetitive dive*, *surface interval time*, *maximum bottom time*, *decompression stop*, *repetitive group*, *safety stop*, *residual nitrogen time*, *actual bottom time*, *total bottom time*, *emergency decompression*, *no-decompression-stop limit*, *dive profile*, *step dive*, *spike dive*, *multilevel dive profile*, *sawtooth dive profile*, *scrolling*, and *contingency plan*.
- Explain contingency planning for scuba diving.
- Compare the advantages and disadvantages of dive computers.

In this chapter, you will learn about all phases of dive planning: advance planning, short-term planning, on-site planning, and postdive planning. You will also learn about area orientations and how to do dive profile planning. An essential part of dive planning is scheduling your time and depth to avoid DCS (decompression sickness).

A well-planned dive increases enjoyment and satisfaction and decreases the risk of injury. A poorly planned dive can result in disappointment, embarrassment, and discomfort. By the end of this chapter, you will understand the significance of the expression "Plan your dive, and then dive your plan."

Dive-Planning Factors

Many factors affect your plans for a dive or a diving trip. Keep the following considerations in mind when you are looking ahead to a dive outing:

- Health and fitness are important. Illnesses, required medications, and recent operations probably disqualify you for diving. If your health is not normal, consult a diving physician. If there is any doubt about your physical condition, refrain from diving until you are in good health. If you are prone to motion sickness, take steps to try to prevent it.
- Climate is a big factor affecting dive planning. If you dive close to where you live, dive planning is easier than if you intend to dive thousands of miles away. A difference in climate usually means a big difference in diving conditions, which means a difference in your equipment requirements.
- The distance you travel to a diving destination affects your planning. If you travel far to reach the destination, allow a day to rest and recover from travel before you dive. After even one day of repetitive diving, wait one full day before flying home.
- Weather affects diving conditions significantly. Storms and sudden changes in the weather can make diving dangerous. Know the weather forecast, and reschedule your dive if poor weather is predicted. Know the expected wind speed, air temperature, and water conditions.
- Seasonal changes affect water movement, water visibility, air and water temperatures, entry and exit areas, and the presence of certain types of animals. You

First you learn the theory of diving, then the skills of diving, and then you apply what you have learned. Your training objective is to qualify to dive without supervision, which involves planning your underwater excursions.

Fish in a Red Sea coral cave

should know what to expect at a dive site at different times of the year. It helps to know the visibility, water temperatures, tides, surf, surge, currents, bottom composition, silt conditions, plants, and animals.

You need to be physically and mentally fit for diving. Fitness for diving implies that you are well rested, are well nourished, have the physical strength and stamina to meet the requirements of the environment and the activity, are qualified for the activity, are not apprehensive about your plans, are not goaded into doing something you are not prepared to do, and do not allow pride to affect good judgment.

Your objective for the dive affects your planning. Different diving activities require different plans and different equipment. The planning of an underwater photography dive is not the same as the planning of a dive where you intend to hunt for game.

You must know and observe laws, regulations, and customs. Some areas have laws that require the use of a dive flag. Obey fish and game regulations. Some diving professionals discourage the taking of any living thing in an area. You need to know the behavior expected of you. Knowing the expectations in advance can help you avoid being embarrassed at the dive site.

Etiquette is important. Will early-morning diving activities be offensive to residents near the dive site? Will the parking of vehicles at a site irritate people? Be considerate of others who may be in the area where you intend to dive, including those who are fishing nearby. Consider the impact of noise, changing clothes, and dive site access. Then make your plans using good etiquette.

Advance Planning and Preparation

The first phase of dive planning is the determination of why, who, where, when, how, and what.

Why? Determine the objective of the dive. What do you want to do? Take photos? Explore? Look for artifacts?

Who? Determine with whom you want to dive. Select a buddy who is interested in your dive objective.

Where? Determine a primary and an alternative site. If conditions at the primary site are unfavorable, go to the alternate site.

When? Determine the best time to dive. The water at most areas is usually calmer in the morning than it is in the afternoon. Tidal currents and height may affect the best time to dive.

How? Decide how to reach the dive site. Who will drive? What are the directions?

What? Determine what equipment is needed for the dive. Who will bring the float and flag? How many tanks do you need? Are there any special needs for the intended activity?

Advance preparation can include

- making reservations,
- paying deposits,
- buying or renting equipment,
- having equipment serviced or repaired,
- getting tanks filled,
- obtaining a fishing license or permit,
- preparing equipment for photography, and
- obtaining emergency contact information.

Fresh water diving at Lake Mead, Nevada

Your preparations will usually include a trip to your local dive facility. Inspect your equipment before you go. You may discover a needed repair that requires some time to complete. Identify your equipment needs early.

Short-Term Planning and Preparation

The day and evening before you intend to go diving, you need to take three actions. First, you should find out the weather forecast and current water conditions so you can determine if conditions will be acceptable for your diving activities. Call your dive buddy to discuss and confirm your plans. Last-minute revisions, such as going to the alternative site, may be necessary. If you anticipate poor diving or weather conditions, reschedule the dive.

The second step of short-term preparation is packing your diving equipment and your personal items. (See chapter 4 for an equipment checklist.)

The third step of short-term planning is to write down your dive plans and schedule. Leave this information with a friend. Instruct your friend to notify the authorities if you fail to return by a certain time.

On-Site Planning and Preparation

When you and your buddy arrive at the dive site, you must determine whether the conditions are acceptable for diving. If not, go to an alternative site. If the conditions at the alternative site are also unacceptable, abort the dive.

An important step in the assessment of a dive site is the estimation of the current. Look for telltale signs such as kelp bent over from water movement, a wake around the anchor line or behind an anchored boat, or objects drifting on the surface. Determine the velocity of moving water by measuring how long it takes a floating object to move a known distance, such as the length of your boat. When an object moves 100 feet (30 m) in one minute, the speed of the object is approximately 1 knot (1.15 miles or 1.85 km per hour). When a current exceeds about 1/3 knot (0.4 miles or 0.6 km per hour), you must heed it because you can swim at a speed of only about 3/4 knot (0.86 miles or 1.4 km per hour). Plan the dive so that the current assists you in swimming to your exit point at the end of the dive. Figure 7.1 includes a table that can help you estimate current velocity.

If the diving conditions are favorable, you should then determine the diving area. Select the entry and exit areas for the dive, and discuss the entry and exit procedures. Agree on the course to be followed during the dive. Agree on time, minimum air pressure, and landmarks for changes in direction. You and your buddy should know in advance approximately where you will be at any time during the dive.

An important part of your planning is discussing and agreeing on the procedures for the buddy system. Decide who is in charge of the team, where you will position yourselves relative to each other, how you will move (steadily or start and stop), and what reunion procedures you will follow in the event of separation. Remember that communication is much easier on land than it is underwater, so take advantage of the opportunity you have to communicate and coordinate while preparing for a dive.

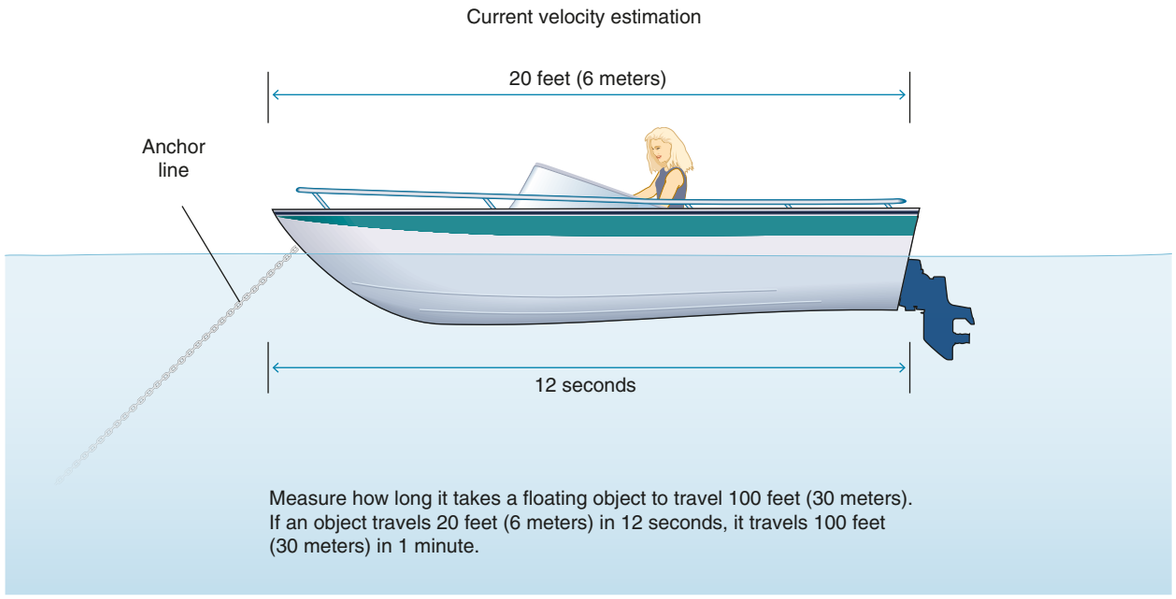

Current velocity table
(time to travel 100 feet or 30 meters)

Time (seconds)	Speed (knots)	Time (seconds)	Speed (knots)
5	12.0	95	0.62
10	6.0	100	0.59
20	3.0	110	0.54
30	2.0	120	0.49
40	1.5	130	0.46
50	1.2	140	0.42
60	1.0	150	0.39
70	0.84	160	0.37
80	0.74	170	0.35
90	0.66	180	0.33

Figure 7.1 You can estimate the velocity of a current if you time how long it takes a floating object to move the length of your boat.

Always review your plans for emergencies. Agree on air-sharing procedures. Discuss what to do in the event of a serious diving emergency. Know where, how, and whom to call for help. Make sure you both have access to a first aid kit and other emergency equipment. A few minutes spent coordinating procedures before an emergency can save precious seconds if an accident occurs.

Scuba diving requires dive profile planning. You and your buddy need to agree on the maximum time and depth for your dive. You must limit time and depth to avoid decompression illness. Dive profile planning is discussed later in this chapter.

Area Orientations

You should recognize the importance of learning about a dive site before diving there. Because orientation is vital and because you want to be a responsible diver, you need to learn how to obtain orientations. Area orientations can be formal or informal. A formal orientation is provided as a service by a diving professional. The professional will tell you what to look for and what to look out for in the area and will lead you on a dive. A professional dive guide provides suggestions and points out items of interest and potential hazards. When you have completed a formal orientation, ask the professional to sign and stamp your logbook. A continuing education dive course is another excellent form of formal orientation to a new area.

Formal orientations are ideal, but if you cannot arrange to have a professional introduce you to a new region, consider some or all of the following options for an informal orientation:

- Read books, articles, and brochures about diving in the area. Learn as much about an area as you can before you go there.
- Write to dive stores in the region where you intend to dive. Ask if you can participate in a dive class session for your orientation to the area.
- Write to dive clubs in a region where you intend to dive. Ask if you can participate in a club-sponsored dive when you are in the area. Ask for contact information for several club members who dive regularly and may be willing to allow you to go diving with them.
- When you arrive in a new area, find local dive sites and visit them when divers are likely to be there. Ask the divers about the sites while they are preparing to dive or after they exit from a dive. If you have your equipment ready, you may be able to accompany them on a dive. Make sure they have experience in diving at the site.

SCUBA WISE

Having reviewed dive accident reports for many years, I have found that failure to adequately plan or to carry out the plan for a dive is a common cause of accidents and injuries. I have been able to avoid serious injury during decades of diving. However, I have had some bad experiences. These experiences occurred when I failed to plan adequately, when I attempted activities without first completing training for the activity, or when I did not have adequate knowledge of a new dive site.

I dived from shore in California on a beach with a steep incline without taking the time to research the dive. I was used to donning my fins, wading into the water, and swimming beneath the oncoming waves. When I attempted this procedure on this type of beach, I found that there was a trough created by plunging waves, and that the trough was a dropoff. The trough is also where the waves break suddenly and violently. As I waded into the water, I lost my footing at the trough at the same time that a large breaker suddenly formed and pounded me into the trough. I managed to kick past the trough before the next wave hit, but I was humbled by the experience. A key part of dive planning is to understand the terrain and learn the local procedures. After the dive, I watched a local diver literally run into the water after a wave broke, jump over the trough, turn onto his back, pull on one fin, and kick beyond the area where the waves broke. I have used that technique on steep beaches ever since that day when the plunging surf taught me a lesson. Avoid surprises that could be dangerous by learning about any new location before diving there.

- Purchase a space on a diving charter boat. When you board the vessel, tell the crew you are new to the area. Ask for advice about diving procedures, and ask to be introduced to an experienced diver who can provide additional information.

When you dive with local divers, allow them to go first. Do as they say and do as they do. Procedures vary from region to region. A procedure you use in your area may be inappropriate in a different area. For example, in your normal diving environment, you may be able to enter the water from a boat in a current without holding on to a line. But if you tried to do so in a different area, you might be swept away at once. Be humble, listen to others, and follow their example to avoid embarrassment.

Dive Profile Planning

There are limits on how long you may remain at various depths. The amount of nitrogen absorbed by your body determines the limits. You need to be aware of the effects of **residual nitrogen**, which is nitrogen remaining in your system from a dive made within the past 12 hours. Whether you do one dive or a series of dives, it takes time to eliminate nitrogen from your body. If you dive again before the excess nitrogen has had time to outgas, you add to the nitrogen already in your body, and you reach critical nitrogen levels faster than if you had not already absorbed excess nitrogen. A **repetitive dive** is any dive made within 6 to 24 hours (depending on the dive-planning device) of a previous dive. Figure 7.2 shows how the amount of residual nitrogen in the body builds from repetitive dives. A **rest stop**—a precautionary decompression stop during ascent—reduces the risk of decompression illness.

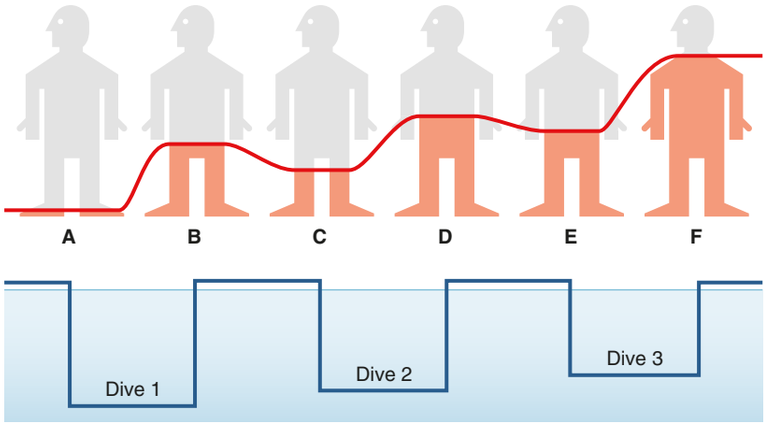

A—Normal amount of nitrogen in body
B—Amount of nitrogen in body after first dive
C—Amount of nitrogen in body after first surface interval
D—Amount of nitrogen in body after second dive
E—Amount of nitrogen in body after second surface interval
F—Amount of nitrogen after third dive

Figure 7.2 If you make a repetitive dive before allowing the residual nitrogen to leave your body, the nitrogen continues to accumulate.

Decompression experts use complex mathematical calculations and field testing to establish time limits for various depths for single and repetitive dives. The time limits are published in tables and programmed into calculators and dive computers. You need to know how to use these dive-planning devices to plan your dive profile so that you can minimize the risk of decompression illness.

No dive-planning device can guarantee that you will not develop decompression illness. Dive tables, calculators, and computers provide information based on statistics acquired through testing. If a diver adheres to the profile limits of a device, the statistical probability that the diver will develop decompression illness is small. The devices assume that you are in good health, that you do not get cold during the dive, that you do not exert strenuously, and that you ascend at the correct rate.

If you dive to the maximum time limits indicated by any dive-planning device, you increase the likelihood of decompression illness. Reducing your exposure to pressure reduces the likelihood of decompression illness.

Dive Tables

The U.S. Navy (USN) developed tables for dive planning. The recreational diving community adopted the military tables in 1980 and has used modified versions successfully. Although the tables were designed for military diving rather than recreational diving, many decompression experts still consider the USN dive tables—with reduced maximum dive times—appropriate tables for recreational use. New Doppler-tested dive tables developed by the USN are the basis for the dive tables presented in this chapter.

The USN tables use compartments with half-times of 5 minutes, 10 minutes, 20 minutes, 40 minutes, 80 minutes, and 120 minutes. The amount of nitrogen remaining in the 120-minute (2-hour) compartment determines the **repetitive group designations** of the tables, which are represented by letters. Because outgassing of a compartment takes six half-times, you can see why the USN tables define a repetitive dive as any dive within 12 hours (6 × 2 hours) of a previous dive.

The tables include four sets of dive tables:

A. Total bottom timetable

B. Surface interval timetable

C. Residual nitrogen timetable

D. Decompression timetable

Dive Table Modifications

Diving organizations and manufacturers have modified the dive tables to make them more appropriate for recreational diving. Here are some typical differences between standard dive tables and the modified dive tables presented in this chapter:

- Reduced time limits—Dive-planning devices today include maximum time limits that are less than the no-decompression-stop limits of the USN dive tables.
- Reduced depth limits—The USN dive tables provide dive-planning information for depths to 190 feet (58 m). The recommended maximum depth limit for recreational diving is 100 feet (30.5 m). New divers should limit diving to

AIR DIVING TABLES

SCUBA EDUCATORS INTERNATIONAL® PDIC INTERNATIONAL®

READ BEFORE PLANNING A DIVE

SEI & PDIC International Air Diving Tables are designed to enhance diver safety and assist in dive planning. The Air Diving Tables should be used only by persons properly trained in their use. SEI & PDIC International Air Diving Tables provide a conservative safety margin to help avoid decompression sickness. Use of these Air Diving Tables is not guaranteed to eliminate the possibility of decompression sickness. To gain the greatest conservative margin from these tables, divers must carefully plan their dives in accordance with the No-Decompression Stop Limits and ascend no faster than 30 feet per minute with appropriate safety stops. Do not exceed the No-Decompression Stop Limits. Do not dive deeper than 100 feet. Do not use these tables for high altitude diving without calculating an adjustment for high altitude. Susceptibility to decompression sickness varies in accordance with a diver's physical wellness and readiness, environmental conditions, ascent rates, and many other situations and conditions.

© 2009 Scuba Educators International Revised 10-2013

Table A

TOTAL BOTTOM TIME (TBT) IN MINUTES

No-Decompression Stop Limit

START HERE

DEPTH Meters	DEPTH Feet	No-Decompression Stop Limit	A	B	C	D	E	F	G	H	I	J	K	L	M
3	10		57	101	158	245	426								
4.6	15		36	60	88	121	163	217	297	449					
6.1	20		26	43	61	82	106	133	165	205	256	330	461		
7.6	25	354	20	33	47	62	78	97	117	140	166	198	236	285	354
9.1	30	223	17	27	38	50	62	76	91	107	125	145	167	193	223
10.7	35	168	14	23	32	42	52	63	74	87	100	115	131	148	168
12.2	40	135	12	20	27	36	44	53	63	73	84	95	108	121	135
13.7	45	102	11	17	24	31	39	46	55	63	72	82	92	102	
15.2	50	80	9	15	21	28	34	41	48	56	63	71	80		
16.7	55	63	8	14	19	25	31	37	43	50	56	63			
18.3	60	51	7	12	17	22	28	33	39	45	51				
21.3	70	42	6	10	14	19	23	28	32	37	42				
24.4	80	32	5	9	12	16	20	24	28	32					
27.4	90	24	4	7	11	14	17	21	24						
30.5	100	18	4	6	9	12	15	18							
33.5	110	16	3	6	8	11	14	16							
36.6	120	10	3	5	7	10									
39.6	130	6	2	4	6										

Repetitive Group Designation

Table C

RESIDUAL NITROGEN TIME (RNT) IN MINUTES

Repetitive Dive Depth in Feet/Meters

35	40	45	50	55	60	70	80	90	100	110	120	130
11	12	14	15	17	18	21	24	27	30	33	36	39
15 / 153	13 / 122	12 / 90	11 / 69	10 / 53	9 / 42	8 / 34	7 / 25	6 / 18	5 / 13	5 / 11	5 / 5	4 / 2
24 / 144	21 / 114	18 / 84	17 / 63	15 / 48	14 / 37	12 / 30	10 / 22	9 / 15	8 / 10	8 / 8	7 / 3	6
33 / 135	29 / 106	25 / 77	23 / 57	20 / 43	19 / 32	16 / 26	14 / 18	12 / 12	11 / 7	10 / 6	9 / 1	9
43 / 125	37 / 98	32 / 70	29 / 51	26 / 37	24 / 27	20 / 22	18 / 14	16 / 8	14 / 4	13 / 3	12	11
53 / 115	45 / 90	40 / 62	35 / 45	32 / 31	29 / 22	25 / 17	22 / 10	19 / 5	17 / 1	16	14	13
64 / 104	55 / 80	48 / 54	42 / 38	38 / 25	35 / 16	29 / 13	25 / 7	22 / 2	20	18	17	15
75 / 93	64 / 71	56 / 46	49 / 31	44 / 19	40 / 11	34 / 8	29 / 3	26	23	21	19	18
88 / 80	74 / 61	64 / 38	57 / 23	51 / 12	46 / 5	39 / 3	33	29	26	24	22	20
101 / 67	85 / 50	73 / 29	65 / 15	58 / 5	52	44	38	33	30	27	24	22
116 / 52	97 / 38	83 / 19	73 / 7	65	58	49	42	37	33	30	27	25
132 / 36	109 / 26	93 / 9	81	72	65	54	46	41	36	33	30	27
149 / 19	122 / 13	104	90	80	72	59	51	44	40	36	32	30
169	136	115	99	88	79	65	55	48	43	39	35	32

Note: For repetitive dives less than 35 feet use the RNT values for 35 feet

Table B

SURFACE INTERVAL TIME (SIT) IN HOURS AND MINUTES

New Repetitive Group	A	B	C	D	E	F	G	H	I	J	K	L	M
A	2:20 / 0:10	3:36 / 1:17	4:31 / 2:12	5:23 / 3:04	6:15 / 3:56	7:08 / 4:49	8:00 / 5:41	8:52 / 6:33	9:44 / 7:25	10:36 / 8:17	11:29 / 9:10	12:21 / 10:02	13:13 / 10:54
B		1:16 / 0:10	2:11 / 0:56	3:03 / 1:48	3:55 / 2:40	4:48 / 3:32	5:40 / 4:24	6:32 / 5:17	7:24 / 6:09	8:16 / 7:01	9:09 / 7:53	10:01 / 8:45	10:53 / 9:38
C			0:55 / 0:10	1:47 / 0:53	2:39 / 1:45	3:31 / 2:38	4:23 / 3:30	5:16 / 4:22	6:08 / 5:14	7:00 / 6:07	7:52 / 6:59	8:44 / 7:51	9:37 / 8:43
D				0:52 / 0:10	1:44 / 0:53	2:37 / 1:45	3:29 / 2:38	4:21 / 3:30	5:13 / 4:22	6:06 / 5:14	6:58 / 6:07	7:50 / 6:59	8:42 / 7:51
E					0:52 / 0:10	1:44 / 0:53	2:37 / 1:45	3:29 / 2:38	4:21 / 3:30	5:13 / 4:22	6:06 / 5:14	6:58 / 6:07	7:50 / 6:59
F						0:52 / 0:10	1:44 / 0:53	2:37 / 1:45	3:29 / 2:38	4:21 / 3:30	5:13 / 4:22	6:06 / 5:14	6:58 / 6:07
G							0:52 / 0:10	1:44 / 0:53	2:37 / 1:45	3:29 / 2:38	4:21 / 3:30	5:13 / 4:22	6:06 / 5:14
H								0:52 / 0:10	1:44 / 0:53	2:37 / 1:45	3:29 / 2:38	4:21 / 3:30	5:13 / 4:22
I									0:52 / 0:10	1:44 / 0:53	2:37 / 1:45	3:29 / 2:38	4:21 / 3:30
J										0:52 / 0:10	1:44 / 0:53	2:37 / 1:45	3:29 / 2:38
K											0:52 / 0:10	1:44 / 0:53	2:37 / 1:45
L												0:52 / 0:10	1:44 / 0:53
M													0:52 / 0:10

37 — Minutes of Residual Nitrogen (RNT) Added to Actual Bottom Time to Compute TOTAL BOTTOM Time

98 — Calculated Maximum ACTUAL BOTTOM TIME (ABT) in minutes. If This Number is Exceeded, a Decompression Stop is Required

SEI & PDIC INTERNATIONAL ACCEPTS NO RESPONSIBILITY FOR ACCIDENTS ARISING FROM THE USE OF THESE TABLES

Maximum Time For This Interval — 0:52

Minimum Time For This Interval — 0:10

Courtesy of Scuba Educators International.

REPETITIVE DIVE PROFILES

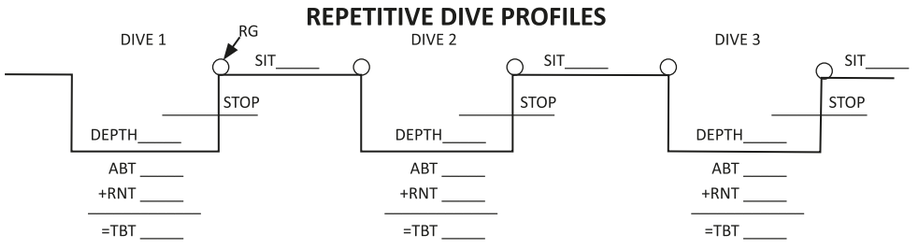

DEFINITIONS

BOTTOM TIME (BT). The elapsed time in minutes, starts when the diver leaves the surface in descent and stops when the diver begins safety stop or surfaces (round up to next whole minute). On repetitive dives this is Actual Bottom Time (ABT).

DEPTH. The depth of a dive; use the maximum depth attained during the dive in feet or meters of seawater. When depth is not indicated on the table use the next greater depth available.

NO-DECOMPRESSION STOP LIMIT. The longest amount of time a diver can spend at depth without requiring a decompression stop. Table A lists these for all depths. For repetitive dives the Residual Nitrogen Time (RNT) must be added to the Bottom Time (BT) and the Total Bottom Time (TBT) must not exceed the No-Decompression Stop Limit.

REPETITIVE DIVE. Any dive conducted within the time period specified by the surface interval table (Table B) of a previous dive with a minimum surface interval of 10 minutes. For repetitive dives of less than 35 feet, use the RNT (Table C) values for 35 feet to calculate Total Bottom Time (TBT).

REPETITIVE GROUP DESIGNATION (RG). Letters that relate to the amount of residual nitrogen in a diver's tissues for the time specified in the Suface Interval Table.

RESIDUAL NITROGEN TIME (RNT). Located in Table C, an amount of time, in minutes, which must be added to the Bottom Time (BT) of a repetitive dive to calculate Total Bottom Time (TBT). Represents residual nitrogen still in a diver's tissues from previous dives.

SAFETY STOP. A 3 minute minimum safety stop at 15-25 feet is recommended at the end of all dives.

SURFACE INTERVAL TIME (SIT). Located in Table B, the elapsed time from when a diver surfaces following a dive to the time (next whole minute) that diver starts the descent of the next dive. SIT must be a minimum of 10 minutes.

TOTAL BOTTOM TIME (TBT). Calculated by adding the Bottom Time (BT) and the Residual Nitrogen Time (RNT). For times not indicated at a particular depth, use the next longer time available for that depth.

RULES

ASCENT RATE. Divers should ascend at 30 feet per minute or slower.

DIVES LESS THAN 25 FEET. Dives less than 25 feet do not have a specific No-Decompression Stop Limit. However, if conducted as repetitive dives, they must be calculated as a 35 foot dive for Residual Nitrogen Times (RNT), Table C.

EXCEEDING THE NO-DECOMPRESSION STOP LIMITS. If divers err and stay longer than the times allowed on the No-Decompression Stop Limit, Table A, the divers have exposed themselves to an increased decompression requirement and must refer to the Decompression Stops Table, Table D, for required decompression stops to complete the dive. Once on the surface these divers should not dive again for at least 12 hours or fly for at least 24 hours.

FLYING AFTER DIVING. The longer the surface interval before flying the less likely a diver will experience decompression sickness. A minimum surface interval of 12 hours is required before ascent to altitude in a pressurized aircraft (or a ground altitude of 1000 feet above sea level). Divers who plan to make multiple dives for several days or computer assisted dives should take extra precaution and wait an extended surface interval of greater than 12 hours before flying or ascending to altitude. Divers who have made dives requiring decompression stops should wait a minimum of 24 hours before flying or ascending to altitude.

HIGH ALTITUDE DIVING. These Air Diving Tables are not to be used for diving at altitudes greater than 1000 feet above sea level without calculating altitude adjustments.

ORDER OF DIVES. All repetitive dives should be conducted with each successive dive equal in depth or shallower than the previous dive.

SPORT DIVING MAXIMUM DEPTH. Sport divers should not exceed 100 feet of depth.

SPORT DIVING TIME LIMIT. On any dive the Total Bottom Time (TBT) should not exceed the No-Decompression Stop Limit. Dives requiring mandatory decompression stops should not be planned.

Table D

DECOMPRESSION STOPS TABLE

Depth Feet Meters	Bottom Time (min.)	Time to First Stop (min:sec)	Decompression Stops (min) @ 20ft.	Total Assent (min:sec)	Repetitive Group
40 12.2	135		0		M
	150	0.40	6	7:20	O
	160	0.40	14	15:20	Z
	170	0.40	21	22:20	Z
	180	0.40	27	28:20	Z
50 15.2	80		0		K
	90	1:00	2	3:40	M
	95	1:00	4	5:40	N
	100	1:00	8	9:40	O
	110	1:00	21	22:40	O
60 18.3	51		0		I
	60	1:20	2	4:00	L
	65	1:20	7	9:00	L
	70	1:20	14	16:00	N
	80	1:20	23	25:00	O
70 21.3	42		0		I
	50	1:40	9	11:20	L
	55	1:40	14	16:20	M
	60	1:40	24	26:20	N
	70	1:40	44	46:20	O
80 24.4	32		0		H
	35	2:00	1	3:40	J
	40	2:00	10	12:40	K
	45	2:00	17	19:40	M
	50	2:00	24	26:40	M
90 27.4	24		0		G
	30	2:20	4	7:00	J
	35	2:20	14	17:00	L
	40	2:20	23	26:00	M
	45	2:20	31	34:00	N
100 30.5	18		0		F
	25	2:40	3	6:20	J
	30	2:40	15	18:20	L
	35	2:40	26	29:20	M
	40	2:40	36	39:20	N
110 33.5	16		0		F
	20	3:00	3	6:40	I
	25	3:00	14	17:40	K
	30	3:00	27	30:40	M
	35	3:00	39	42:40	N
120 36.6	10		0		D
	15	3:20	2	6:00	H
	20	3:20	8	12:00	J
	25	3:20	24	28:00	L
	30	3:20	38	42:00	N
130 39.6	6		0		C
	10	3:40	1	5:20	G
	15	3:40	4	8:20	I
	20	3:40	17	21:20	K
	25	3:40	34	38:20	M

Courtesy of Scuba Educators International.

Dive Table Terms

Actual Bottom Time (ABT)—The elapsed time in minutes (rounded up to the next whole minute) from when a diver leaves the surface in descent until the diver begins a rest stop or surfaces.

No Decompression Stop Limit (NDSL)—The maximum time that a diver may stay at a specified depth (indicated in table A). For repetitive dives, the residual nitrogen time (RNT) from table C must be added to the actual bottom time (ABT) to obtain the total bottom time (TBT), which must not exceed the NDSL specified in table A. Table C also indicates the maximum ABT for various depth and repetitive group combinations that will prevent the TBT from exceeding the NDSL.

Repetitive Dive—Any dive conducted within 12 hours of a previous dive with a minimum 10-minute surface interval. For repetitive dives of less than 40 feet, use the 40-foot (12 m) values in table C (RNT table) to determine the TBT.

Repetitive Group Designation (RGD or RG)—A letter that indicates the amount of residual nitrogen in a diver's tissues for 12 hours after diving.

Residual Nitrogen Time (RNT)—An amount of time (in minutes) that must be added to the actual bottom time (ABT) of a repetitive dive to determine the total bottom time (TBT). RNT compensates for residual nitrogen remaining in a diver's tissues from previous dives.

Surface Interval Time (SIT)—The elapsed time (rounded up to the next whole minute) from when a diver surfaces after a dive to the time the diver begins the descent of a repetitive dive. The minimum SIT is 10 minutes.

Total Bottom Time (TBT)—The sum of ABT and RNT. For times not indicated at a particular depth, use the next longest time available for that depth.

depths of about 60 feet (18 m). Advanced divers qualify to dive to approximately 100 feet. Divers who complete a deep-diving specialty course qualify to dive to 130 feet (39 m). Technical divers qualify to dive at even greater depths. Depth in water is similar to speed on land. When you qualify for a driver's license, you may drive at speeds up to the legal limit. Driving at much higher speeds is hazardous and requires specialized training, special equipment, and controlled situations to minimize the risk of serious injury or death. Deep diving (beyond 100 feet) is similar to race car driving in many respects. Deep diving is a professional endeavor that unqualified recreational divers should not attempt.

- Revised **surface interval times**—In 1983, a USN study reported a number of errors in the surface interval timetable. The errors usually do not affect the type of diving that recreational divers do, but some modified versions of the tables include the corrections.

- Combined tables—Modified versions of the USN dive tables often combine information to make the tables easier to use. Tables may include the **total bottom time** (TBT) for repetitive dives along with the **residual nitrogen time** (RNT). Tables may include required decompression information to eliminate a separate decompression timetable.

Dive Table Use

In this section we'll take a closer look at some of the individual elements comprising the combined dive tables.

Table A provides the NDSL, without required decompression stops for various depths. This table also provides a letter designation for various dive profiles. The letter designation indicates the amount of excess nitrogen remaining in your body after a dive.

Table A

TOTAL BOTTOM TIME (TBT) IN MINUTES

No-Decompression Stop Limit

START HERE

DEPTH Meters	DEPTH Feet	No-Decompression Stop Limit	A	B	C	D	E	F	G	H	I	J	K	L	M
3	10		57	101	158	245	426								
4.6	15		36	60	88	121	163	217	297	449					
6.1	20		26	43	61	82	106	133	165	205	256	330	461		
7.6	25	354	20	33	47	62	78	97	117	140	166	198	236	285	354
9.1	30	223	17	27	38	50	62	76	91	107	125	145	167	193	223
10.7	35	168	14	23	32	42	52	63	74	87	100	115	131	148	168
12.2	40	135	12	20	27	36	44	53	63	73	84	95	108	121	135
13.7	45	102	11	17	24	31	39	46	55	63	72	82	92	102	
15.2	50	80	9	15	21	28	34	41	48	56	63	71	80		
16.7	55	63	8	14	19	25	31	37	43	50	56	63			
18.3	60	51	7	12	17	22	28	33	39	45	51				
21.3	70	42	6	10	14	19	23	28	32	37	42				
24.4	80	32	5	9	12	16	20	24	28	32					
27.4	90	24	4	7	11	14	17	21	24						
30.5	100	18	4	6	9	12	15	18							
33.5	110	16	3	6	8	11	14	16							
36.6	120	10	3	5	7	10									
39.6	130	6	2	4	6										

Repetitive Group Designation: A B C D E F G H I J K L M

Dive table courtesy of Scuba Educators International.

Table B, the surface interval timetable, provides letter group information based on your surface interval time (SIT). You begin a surface interval with one letter group designation; as you outgas nitrogen, you acquire lower letter group designations.

Table B

SURFACE INTERVAL TIME (SIT) IN HOURS AND MINUTES

Repetitive Group Designation

Repetitive Dive Depth in Feet/Meters

Repetitive Dive Depth in Feet/Meters	A	B	C	D	E	F	G	H	I	J	K	L	M
130 39													
4 2	2:20 0:10	3:36 1:17	4:31 2:12	5:23 3:04	6:15 3:56	7:08 4:49	8:00 5:41	8:52 6:33	9:44 7:25	10:36 8:17	11:29 9:10	12:21 10:02	13:13 10:54
6		1:16 0:10	2:11 0:56	3:03 1:48	3:55 2:40	4:48 3:32	5:40 4:24	6:32 5:17	7:24 6:09	8:16 7:01	9:09 7:53	10:01 8:45	10:53 9:38
9			0:55 0:10	1:47 0:53	2:39 1:45	3:31 2:38	4:23 3:30	5:16 4:22	6:08 5:14	7:00 6:07	7:52 6:59	8:44 7:51	9:37 8:43
11				0:52 0:10	1:44 0:53	2:37 1:45	3:29 2:38	4:21 3:30	5:13 4:22	6:06 5:14	6:58 6:07	7:50 6:59	8:42 7:51
13					0:52 0:10	1:44 0:53	2:37 1:45	3:29 2:38	4:21 3:30	5:13 4:22	6:06 5:14	6:58 6:07	7:50 6:59
15						0:52 0:10	1:44 0:53	2:37 1:45	3:29 2:38	4:21 3:30	5:13 4:22	6:06 5:14	6:58 6:07
18							0:52 0:10	1:44 0:53	2:37 1:45	3:29 2:38	4:21 3:30	5:13 4:22	6:06 5:14
20								0:52 0:10	1:44 0:53	2:37 1:45	3:29 2:38	4:21 3:30	5:13 4:22
22									0:52 0:10	1:44 0:53	2:37 1:45	3:29 2:38	4:21 3:30
25										0:52 0:10	1:44 0:53	2:37 1:45	3:29 2:38
27											0:52 0:10	1:44 0:53	2:37 1:45
30												0:52 0:10	1:44 0:53
32													0:52 0:10

New Repetitive Group

Dive table courtesy of Scuba Educators International.

Table C, the residual nitrogen timetable, provides adjusted (reduced) no-stop time limits based on the amount of nitrogen in your body from a previous dive or dives. The table also converts your residual nitrogen to an amount of time for each depth. This residual nitrogen time (RNT) must be added to your actual dive time to obtain the total bottom time (TBT) for the repetitive dive.

Table C

RESIDUAL NITROGEN TIME (RNT) IN MINUTES

Repetitive Dive Depth in Feet/Meters	35 11	40 12	45 14	50 15	55 17	60 18	70 21	80 24	90 27	100 30	110 33	120 36	130 39
A	15 153	13 122	12 90	11 69	10 53	9 42	8 34	7 25	6 18	5 13	5 11	5 5	4 2
B	24 144	21 114	18 84	17 63	15 48	14 37	12 30	10 22	9 15	8 10	8 8	7 3	6
C	33 135	29 106	25 77	23 57	20 43	19 32	16 26	14 18	12 12	11 7	10 6	9 1	9
D	43 125	37 98	32 70	29 51	26 37	24 27	20 22	18 14	16 8	14 4	13 3	12	11
E	53 115	45 90	40 62	35 45	32 31	29 22	25 17	22 10	19 5	17 1	16	14	13
F	64 104	55 80	48 54	42 38	38 25	35 16	29 13	25 7	22 2	20	18	17	15
G	75 93	64 71	56 46	49 31	44 19	40 11	34 8	29 3	26	23	21	19	18
H	88 80	74 61	64 38	57 23	51 12	46 5	39 3	33	29	26	24	22	20
I	101 67	85 50	73 29	65 15	58 5	52	44	38	33	30	27	24	22
J	116 52	97 38	83 19	73 7	65	58	49	42	37	33	30	27	25
K	132 36	109 26	93 9	81	72	65	54	46	41	36	33	30	27
L	149 19	122 13	104	90	80	72	59	51	44	40	36	32	30
M	169	136	115	99	88	79	65	55	48	43	39	35	32

Note: For repetitive dives less than 35 feet use the RNT values for 35 feet

New Repetitive Group

Dive table courtesy of Scuba Educators International.

Table D, the decompression stops table, provides the amount of time that you must decompress if your total bottom time (actual dive time plus residual nitrogen time) exceeds the maximum no-stop time (NDSL) limit.

Now you are ready to learn how to plan dives using the air diving tables. Use table A to determine the maximum no-stop time (NDSL) for a dive. This is the maximum time you can spend underwater for the deepest depth you reach during a dive. If your total bottom time exceeds the NDSL time, you must complete mandatory decompression using table D (this will be discussed later). Your goal is to avoid exceeding the NDSL.

Rows are horizontal, and columns are vertical. Enter table A on the row corresponding to the deepest depth you plan to reach during a dive, and then move across to the end of the row to find your maximum NDSL. After your dive, reenter the row and find the first number (in minutes) that does not exceed your actual dive time. Proceed downward on the column containing your dive time and locate the repetitive group designation letter (from A to M) for your dive. For example, a dive of 50 feet (15 m) for 30 minutes assigns you to group E. Remember that whenever the depth or duration of your dive exceeds a listed number in the table, you must use the next larger number.

Table D

DECOMPRESSION STOPS TABLE

Depth Feet Meters	Bottom Time (min.)	Time to First Stop (min:sec)	Decompression Stops (min) @ 20ft.	Total Assent (min:sec)	Repetitive Group
40 12.2	135		0		M
	150	0.40	6	7:20	O
	160	0.40	14	15:20	Z
	170	0.40	21	22:20	Z
	180	0.40	27	28:20	Z
50 15.2	80		0		K
	90	1:00	2	3:40	M
	95	1:00	4	5:40	N
	100	1:00	8	9:40	O
	110	1:00	21	22:40	O
60 18.3	51		0		I
	60	1:20	2	4:00	L
	65	1:20	7	9:00	L
	70	1:20	14	16:00	N
	80	1:20	23	25:00	O
70 21.3	42		0		I
	50	1:40	9	11:20	L
	55	1:40	14	16:20	M
	60	1:40	24	26:20	N
	70	1:40	44	46:20	O
80 24.4	32		0		H
	35	2:00	1	3:40	J
	40	2:00	10	12:40	K
	45	2:00	17	19:40	M
	50	2:00	24	26:40	M
90 27.4	24		0		G
	30	2:20	4	7:00	J
	35	2:20	14	17:00	L
	40	2:20	23	26:00	M
	45	2:20	31	34:00	N
100 30.5	18		0		F
	25	2:40	3	6:20	J
	30	2:40	15	18:20	L
	35	2:40	26	29:20	M
	40	2:40	36	39:20	N
110 33.5	16		0		F
	20	3:00	3	6:40	I
	25	3:00	14	17:40	K
	30	3:00	27	30:40	M
	35	3:00	39	42:40	N
120 36.6	10		0		D
	15	3:20	2	6:00	H
	20	3:20	8	12:00	J
	25	3:20	24	28:00	L
	30	3:20	38	42:00	N
130 39.6	6		0		C
	10	3:40	1	5:20	G
	15	3:40	4	8:20	I
	20	3:40	17	21:20	K
	25	3:40	34	38:20	M

Dive table courtesy of Scuba Educators International.

Next, you use table B, the surface interval table, to determine your letter group for a repetitive dive. Your group designation depends on the amount of time that you remain at the surface until you dive again. The longer your surface interval, the closer to the beginning of the alphabet your letter group will be.

Enter table B using the group designation you obtained from table A. Move downward along the column until you find the time range (expressed as hours: minutes; for example, 1:26 is 1 hour and 26 minutes) that your surface interval time (SIT) falls into. Remember that when you exceed a number on the dive tables, you must use the next larger number. Follow the row to the left to find your ending letter group designation. For example, if your letter group was E at the beginning of a 2:00 surface interval, you would be in group C at the end of the surface interval. You would be in group C after a surface interval ranging from 1:45 to 2:39.

Use table C, the residual nitrogen time (RNT) table, to determine both your adjusted maximum no-stop time for a repetitive dive and your residual nitrogen time (which must be added to your actual dive time) for your planned depth. For example, if your repetitive group letter from a previous dive is C and your planned depth for your next dive is 50 feet, your actual bottom time (the smaller, lower number for the depth column coordinate) must not exceed 57 minutes. Additionally, the residual nitrogen time (RNT) for the depth (the larger, upper number for the depth column coordinate) must be added to your actual bottom time. If you dived to 50 feet for 40 minutes as a group C diver, you would add 23 minutes of RNT to your actual bottom time to obtain a total bottom time (TBT) of 63 minutes.

After a repetitive dive, you return to table A and use your total bottom time (TBT) to obtain a new repetitive group designation. For example, your dive to 50 feet for 40 minutes as a group C diver yielded a TBT of 63 minutes. According to table A, 63 minutes of TBT at a maximum depth of 50 feet would assign you to repetitive group I at the beginning of your next surface interval.

Table D is for emergency use only when you exceed the maximum no-stop time (NDSL) for a depth. Avoid dives that require planned decompression. If you unintentionally allowed your total bottom time (TBT) to exceed the NDSL, you would refer to table D to determine the required decompression. For example, if your TBT for a depth of 60 feet (18 m) was more than 51 minutes (the NDSL), but less than 61 minutes, you would need to decompress at a depth of 20 feet (6 m) for 2 minutes. Your total ascent time, including the decompression, would require 4 minutes. Your repetitive group after the decompression dive would be L. This example is for emergency purposes only. You should never plan a decompression dive unless you complete specialty training and meet every requirement for safety.

Special-Equipment Dive Planning

You must use special dive tables if you dive using oxygen-enriched air (nitrox); you learn how to use these tables when you take a nitrox specialty course. Do not attempt specialized diving without proper training. Exceeding the 130-foot (39 m) depth limit by a few feet when breathing compressed air is not particularly dangerous, but exceeding the maximum depth limit (which varies with the gas mixture) when breathing mixed gases can cause seizures and drowning. Gas mixtures other than compressed air are only for divers who have the prerequisite training and equipment.

Dive Profile Terms and Rules

Divers use several types of **dive profiles**, which are created by plotting the time and depth of a dive. Figure 7.3 shows a diagram of a standard profile. A dive to a constant depth for a given period of time is typically depicted as a profile with square corners. A **multilevel dive profile** is a dive that progresses from deep to shallow during a given period of time. When you make a dive in a series of steps, you can refer to it as a **step dive**. A **sawtooth dive profile** is a dive that progresses from deep to shallow and back to deep. Avoid this type of dive. Another profile is the bounce profile. This represents a dive with a short ABT,

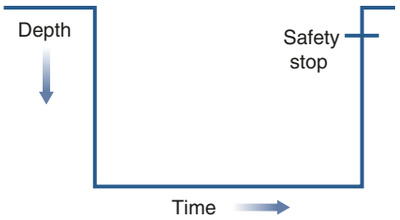

Figure 7.3 Standard dive profile.

such as a dive to free a fouled anchor. You should also avoid this profile, which is depicted graphically as a spike. Bounce profiles are also called **spike dives**. Figure 7.4 graphically depicts these four profiles.

Dive Profile Diagramming

Diagram your dive profiles when planning and recording your dives. Include planned and actual depths, rest decompression stops, bottom times, repetitive group letter designations, and surface interval times. For repetitive dives, include RNTs and TBTs. A simple method for diagramming dive tables is to use a worksheet like the one shown in figure 7.5. Some dive computers automatically record a profile of your dive that may be displayed for review.

Try the following exercise, which combines diagramming with the dive table procedures you have learned. Use the blank diagramming worksheet in figure 7.5 as a template. If you have any difficulties, refer to the preceding section about the use of the dive tables. Assume that all dive times include three-minute rest

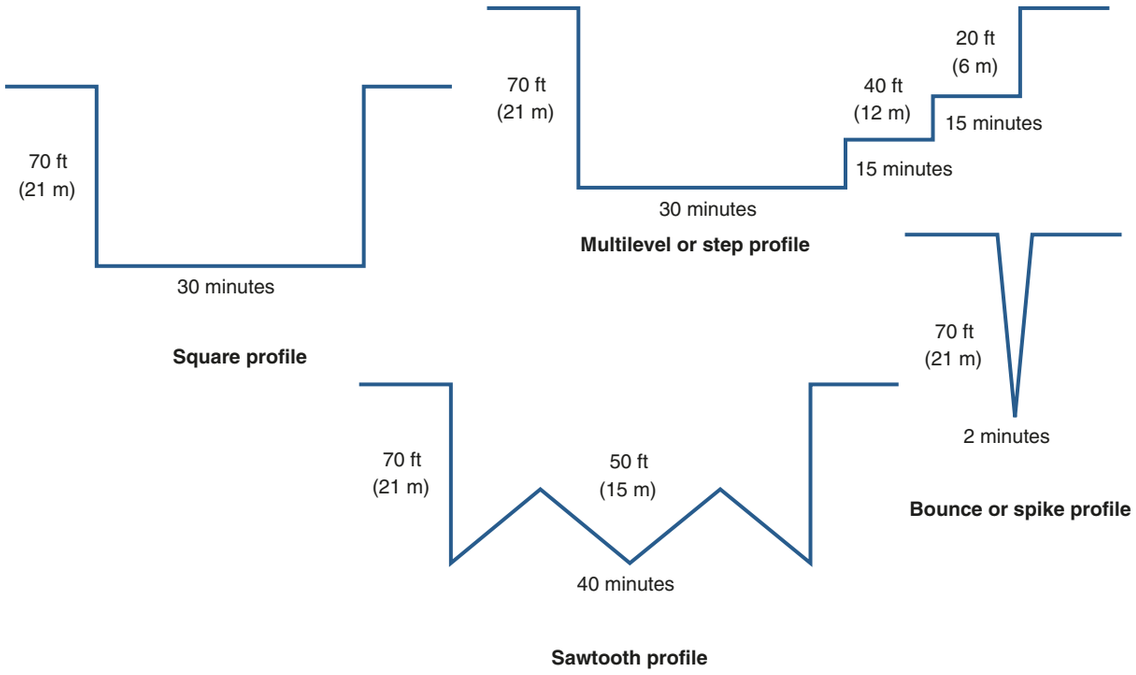

Figure 7.4 Types of dive profiles.

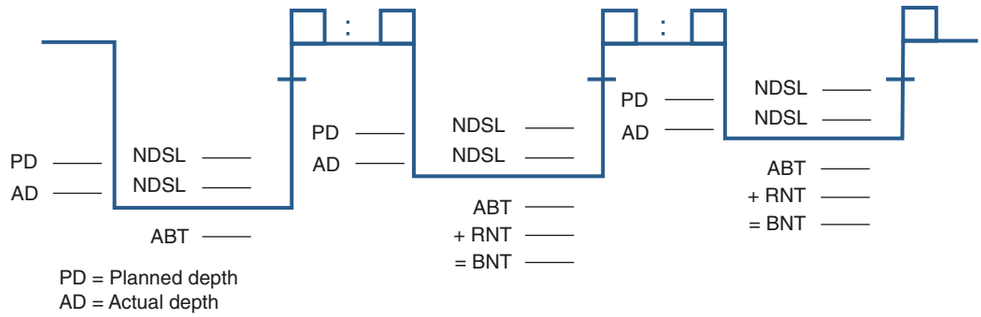

Figure 7.5 Use a worksheet like this to plan your dives.

stops. Calculate and diagram the following series of dives (the answers to the diagramming problem are in the next paragraph and in figure 7.6):

- The first dive is to 78 feet (24 m) with an ABT of 20 minutes, followed by a surface interval of 1.5 hours.
- The second dive is to 55 feet (16.8 m) with an ABT of 25 minutes, followed by a surface interval of 2 hours.
- The third dive is to 40 feet (12 m) with an ABT of 25 minutes.

Solution: The repetitive group after the first dive is E. After the surface interval, the repetitive group changes to D. The RNT for a D diver at 55 feet (16.8 m) is 26 minutes. The TBT (ABT + RNT) for the second dive is 51 minutes (25 + 26). The repetitive group after the second dive is I. After the second surface interval, the group changes to G. The RNT for a group G diver at 40 feet is 64 minutes. The TBT for the third dive is 89 minutes (25 + 64). The repetitive group after the third dive is J.

The following tips help make dive profile diagramming easier:

- Enter the repetitive group letter at every upper corner of a profile except the first corner.
- Add ABT and RNT to obtain TBT for every repetitive dive. Use the recall word ART to help remember that you add ABT and RNT to obtain TBT.

Now combine the dive table procedures to plan a series of dives. Use the blank diagram shown in figure 7.5. Assume that the depth of the first dive—the deepest dive—is 60 feet. The NDSL, according to table A, is 51 minutes. Plan an ABT

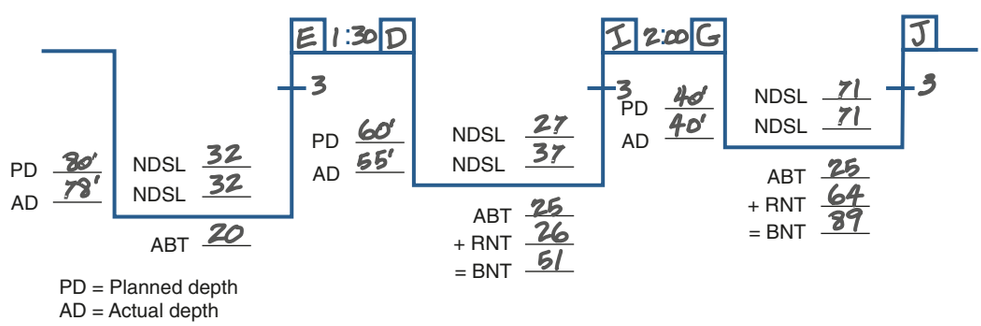

Figure 7.6 Completed diagramming worksheet.

Rules

Ascent Rate. Divers should ascend at 30 feet per minute or slower.

Dives Less Than 25 Feet. Dives less than 25 feet do not have a specific No Decompression Stop Limit. However, if conducted as repetitive dives, they must be calculated as a 35 foot dive for Residual Nitrogen Times (RNT), Table C.

Exceeding the No Decompression Stop Limits. If divers err and stay longer than the times allowed on the No Decompression Stop Limit, Table A, the divers have exposed themselves to an increased decompression requirement and must refer to the Decompression Stops Table, Table D, for required decompression stops to complete the dive. Once on the surface, these divers should not dive again for at least 12 hours or fly for at least 24 hours.

Flying After Diving. The longer the surface interval before flying, the less likely a diver will experience decompression sickness. A minimum surface interval of 12 hours is required before ascent to altitude in a pressurized aircraft (or a ground altitude of 1000 feet above sea level). Divers who plan to make multiple dives for several days or computer-assisted dives should take extra precaution and wait an extended surface interval of greater than 12 hours before flying or ascending to altitude. Divers who have made dives requiring decompression stops should wait a minimum of 24 hours before flying or ascending to altitude.

High Altitude Diving. These Air Diving Tables are not to be used for diving at altitudes greater than 1000 feet above sea level without calculating altitude adjustments.

Order of Dives. All repetitive dives should be conducted with each successive dive equal in depth or shallower than the previous dive.

Sport Diving Maximum Depth. Sport divers should not exceed 100 feet of depth.

Sport Diving Time Limit. On any dive, the Total Bottom Time (TBT) should not exceed the **No Decompression Stop Limit.** Dives requiring mandatory decompression stops should not be planned.

Definitions

Bottom Time (BT). The elapsed time in minutes starts when the diver leaves the surface in descent and stops when the diver begins safety stop or surfaces (round up to next whole minute). On repetitive dives, this is Actual Bottom Time (ABT).

Depth. The depth of a dive; use the maximum depth attained during the dive in feet or meters of seawater. When depth is not indicated on the table, use the next greater depth available.

No Decompression Stop Limit. The longest amount of time a diver can spend at depth without requiring a decompression stop. Table A lists these for all depths. For repetitive dives, the Residual Nitrogen Time (RNT) must be added to the Bottom Time (BT) and the total bottom time (TBT) must not exceed the No Decompression Stop Limit.

Repetitive Dive. Any dive conducted within the time period specified by the surface interval table (Table B) of a previous dive with a minimum surface interval of 10 minutes. For repetitive dives of less than 35 feet, use the RNT (Table C) values for 35 feet to calculate Total Bottom Time (TBT).

Repetitive Group Designation (RG). Letters that relate to the amount of residual nitrogen in a diver's tissues for the time specified in the Surface Interval Table.

Residual Nitrogen Time (RNT). Located in Table C, an amount of time, in minutes, which must be added to the Bottom Time (BT) of a repetitive dive to calculate Total Bottom Time (TBT). Represents residual nitrogen still in a diver's tissues from previous dives.

Safety Stop. A 3-minute minimum safety stop at 15-25 feet is recommended at the end of all dives.

Surface Interval Time (SIT). Located in Table B, the elapsed time from when a diver surfaces following a dive to the time (next whole minute) that diver starts the descent of the next dive. SIT must be a minimum of 10 minutes.

Total Bottom Time (TBT). Calculated by adding the Bottom Time (BT) and the Residual Nitrogen Time (RNT). For times not indicated at a particular depth, use the next longer time available for that depth.

of 28 minutes (including rest stop) for the dive. Your repetitive group designation after the first dive is E. Now plan a repetitive dive to the same depth. Refer to table C to plan the dive because table C provides the maximum allowable times for various letter group designations. If your surface interval is less than 52 minutes, you remain in the E group, your RNT is 29 minutes, and your ABT must not exceed 22 minutes. Your surface interval should be at least 1 hour, as recommended. If you wait an hour between the first and second dive, your repetitive group changes to D, your RNT is 24 minutes, and your ABT cannot exceed 27 minutes. If you want to dive for 28 minutes, you need to extend your surface interval to at least 1 hour and 45 minutes to obtain repetitive group C.

Assume that you repeat the first dive as a group C diver. Your RNT is 19 minutes and your ABT is 28 minutes, so your TBT is 47 minutes. Your repetitive group after the second dive is I.

Assume that the dive site is so good that you want to make a third dive to 60 feet. No ABT minutes are allowed for a diver with a group I designation. Proceed up the column of the table. If you want to dive to 60 feet again for 28 minutes, you need to attain repetitive group C, which allows 32 minutes of diving without required decompression. When you know the group that you need to attain as well as your starting group, you can plan your surface interval. Table B tells you that you must wait at least 5 hours and 14 minutes to move from group I to group C. It would be wise to consider a shallower dive at this point.

Use the tables to plan dives that enable you to avoid emergency decompression. You have three options for planning repetitive dives that do not require mandatory decompression. If your RNT prevents you from making a desired dive, you can (1) reduce the duration of the dive, (2) reduce the depth of the dive, or (3) increase the surface interval preceding the dive.

Special Procedures

Unusual circumstances may arise that require special procedures. You should know the procedures for dive profile planning for each of the following situations:

- A cold or strenuous dive—When a dive is particularly cold or strenuous, use the next greater time for the dive. If the dive is both cold and strenuous, use the next greater time and depth.
- Variations in the rate of ascent—If you ascend faster than 30 feet (9 m) per minute, extend your rest stop by at least two minutes. The faster you ascend, the more you should increase the stop time.
- Multilevel dives—Treat dives to multiple levels as square-profile dives, as if all the time of the dive was at the deepest depth of the dive. Do not attempt to extrapolate the dive tables.
- Omitted decompression—If you need to decompress but fail to do so, use the following procedure for omitted decompression. If you have no symptoms of decompression illness after the dive, remain out of the water, breathe oxygen in the highest concentration possible, rest, drink water, and be alert for symptoms of decompression illness. Wait 24 hours before diving again. If you suspect decompression illness, proceed at once to a hyperbaric facility for a medical examination. The USN has a procedure for in-water decompression, but diving medical experts agree that this procedure is inappropriate unless there is no alternative.

- A rest stop that causes your ABT or TBT to exceed NDSL—If a rest stop causes ABT or TBT to exceed the maximum time limits, determine your repetitive group letter designation using the NDSL for the profile.
- Diving after mandatory decompression—After a dive that requires decompression, wait at least 24 hours before diving again.
- A repetitive dive with an RNT that exceeds the ABT of the previous dive—If the RNT for a repetitive dive exceeds the ABT of the previous dive, use the RNT for planning the repetitive dive.
- Altitude after diving—Ascending to altitude after diving increases the likelihood of decompression illness because of further reduction in pressure. Driving in the mountains or flying after diving can cause decompression illness that would not occur if you remained at sea level until you eliminated the excess nitrogen.

The Divers Alert Network (DAN) recommends that divers wait a minimum of 12 hours before ascent to altitude in a commercial jet airliner (higher than 8,000 feet, or 2,438 m) after a single no-decompression dive. If you make multiple dives for several days, wait at least 18 hours before flight. If your dive requires decompression, a preflight interval substantially longer than 18 hours is required. Lengthening your preflight time interval reduces the chance that you will experience decompression illness.

DAN does not have a recommendation for flying or driving at lower altitudes. The most extensively tested tables for altitudes are the Swiss dive tables. The Swiss tables use a compartment with a much longer half-time than the USN tables. A reasonable approach to altitude delays after diving makes the surface intervals of the USN dive tables equivalent to those of the Swiss tables. Converted USN minimum surface intervals in table 7.1 specify the minimum time required to attain permissible nitrogen levels for various altitudes.

Altitude is any elevation above 1,000 feet (305 m). The altitude delay timetable provides recommended time delays for altitudes up to 10,000 feet (3,048 m). To use the altitude delay timetable, enter the table horizontally on the top line and

Table 7.1 Altitude Delay Timetable

	Starting repetitive group										
Altitude (ft/m)	ABC	D	E	F	G	H	I	J	K	L	Group*
2,000/610	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	2:26	K
3,000/914	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	2:37	4:08	J
4,000/1,219	0:00	0:00	0:00	0:00	0:00	0:00	0:00	2:53	4:30	5:51	I
5,000/1,524	0:00	0:00	0:00	0:00	0:00	0:00	3:04	4:57	6:29	7:44	H
6,000/1,829	0:00	0:00	0:00	0:00	0:00	3:20	5:24	7:12	8:38	9:54	G
7,000/2,134	0:00	0:00	0:00	0:00	3:41	6:02	8:06	9:43	11:10	12:36	F
8,000/2,438	0:00	0:00	0:00	4:08	6:50	9:11	11:04	12:41	14:19	15:40	E
9,000/2,743	0:00	0:00	4:57	8:06	10:48	12:58	14:51	16:39	18:11	23:09	D
10,000/3,048	0:00	6:18	10:37	13:25	15:56	18:05	20:10	21:18	23:24	24:50	C

Note: Times represent the minimum recommended time delay before ascending to the listed altitude and are USN surface interval times with a delay factor of 5.4. Times are in hours:minutes (for example, 5:24 is 5 hours 24 minutes).

*Recommended minimum repetitive groups for indicated elevations.

find your starting repetitive group. Next, find the altitude to which you wish to ascend. If you exceed a number, use the next greater one. The time at the coordinates of the desired altitude and your starting repetitive group indicates the minimum time delay recommended before ascending to the altitude selected.

Profile Contingency Planning

When you plan a dive profile, you should also plan for contingencies. You should know what to do in case you unintentionally exceed your planned depth or time or both. A simple matrix, such as that depicted in figure 7.7, is helpful. When diving using dive tables, you should prepare a contingency matrix in advance of your dives and carry it with you while diving. Dive computers automatically provide contingency information.

Not all dive-planning information displayed by a computer can be accepted at face value. You must learn to apply the dive computer guidelines included in this section.

After you complete your plans, you need to implement them. The first rule of dive planning is to plan your dive, and then dive your plan. You and your buddy should make every effort to do as you agreed to do before the dive. When circumstances force changes in your plans, it helps to have **contingency plans**.

Times in boxes are emergency decompression times.

Sample

	Planned time 30 minutes	Next greater time 40 minutes	Next greater time 50 minutes
Planned depth 60 ft	0	0	0
Next greater depth 70 ft	0	0	5
Next greater depth 80 ft	0	5	10

	Planned time ____	Next greater time ____	Next greater time ____
Planned depth ____			
Next greater depth ____			
Next greater depth ____			

Figure 7.7 A contingency matrix helps you plan ahead for unexpected events.

These plans need to address many possible contingencies for a dive. For example, you should know what to do if you

- surface downcurrent from a boat when you had planned to surface in front of it,
- are unable to reach or use the exit point you had selected on shore,
- end the dive a long distance from your exit location,
- exceed the maximum allowable bottom time (ABT or TBT) for a dive,
- experience a failure of your dive computer while diving, or
- ascend directly to the surface without precautionary decompression.

Dive Calculators and Dive Computers

The benefits of a dive computer outweigh the potential problems associated with the use of these devices. Because dive computers simplify dive planning, you should obtain a dive computer as soon as you can. However, make sure you learn how to use dive tables and dive calculators so that you will know how to plan a dive if you do not have a dive computer. Although dive computers have minimized the use of dive tables, divers must be able to use dive tables as a backup in case the dive computer fails to work.

Dive calculators are circular planning devices that eliminate the mathematics required with the dive tables. Dive calculators provide precalculated numbers, and they eliminate the need to add and subtract times. Guides on the calculators help eliminate line-jumping errors frequently made when using dive tables.

Dive tables have 5- or 10-foot (1.5 or 3 m) increments and require all the time of a dive to be counted at maximum depth of the dive. If the first part of your dive is deeper than the remainder of the dive, you suffer a penalty because the tables treat the entire dive as if it took place at the deepest depth (see figure 7.8). At the end of the dive, you receive a repetitive group designation that is higher than you deserve. Dive tables are advance-planning devices, while dive computers provide constant information on decompression status.

Dive computers use 1-foot (0.3 m) increments for profile planning, and they calculate nitrogen absorption continuously. As you vary depth during a multilevel dive, you are charged only for the nitrogen you absorb. You do not incur the maximum-depth penalty (see figure 7.8) of the dive tables, so your residual nitrogen time after a multilevel dive is less than when calculated using dive tables. The penalty avoidance is the primary advantage of a dive computer as a planning device. Dive computers also provide advance-planning information by **scrolling** a sequential display of time limits for various depths. You

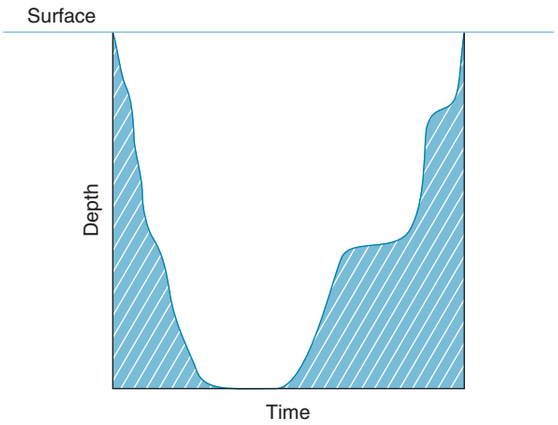

Figure 7.8 The striped area represents penalty time imposed by dive table procedures.

must carry your dive computer with you when you dive because it provides information about your decompression status continuously and keeps track of your residual nitrogen.

Regardless of the type of dive computer you choose, you need to understand some basic principles. First, be sure to read the instruction manual that comes with your computer. Wait 24 hours after diving with dive tables before using a dive computer to ensure that you eliminate all residual nitrogen from your body. Do not exceed the ascent rate specified by the manufacturer, and do a five-minute rest stop at a depth of 15 to 20 feet (4.6 to 6.1 m) at the end of every dive. If you exceed the rate of ascent specified for your computer, extend the duration of your precautionary stop by at least the amount of time it should have taken you to ascend to the stop depth. Do not make repetitive dives deeper than 80 feet (24 m). Keep your dive computer activated until its outgassing is complete. If your computer fails at any time while you are diving at a depth in excess of 30 feet (9 m), terminate the dive immediately with a rest stop. If your computer fails or if you switch it off accidentally, discontinue diving for 24 hours.

If you exceed the maximum time limit for a dive, decompression is mandatory. The computer displays a ceiling—the minimum depth to which you may ascend. As you decompress, the ceiling becomes shallower until the computer indicates that you may surface. After a minimum interval at the surface, the computer begins scrolling—displaying the time limits for various depths sequentially. Wait 24 hours before diving again after any dive for which a ceiling was displayed.

The concept of backup planning with dive tables is good for square profiles; however, for multilevel dives planned with a computer, the use of the dive tables for backup planning is not feasible.

Modern dive computers offer many features. Here are some examples of the features that a computer may include:

- Operating modes for various breathing gases
- Decompression stop data
- A wireless transmitter that provides both cylinder pressure and remaining air data
- Dive profile lifetime history memories

Dive Computers

Advantages

- The maximum-depth penalty of the dive tables is avoided.
- Dive profile information is accurate.
- Dive computers store a record of the dive profile.
- You can eliminate common errors made when using manual dive planners.
- Dive computers offer additional features, such as an ascent rate indicator.

Disadvantages

- Dive computers are electronic instruments that can fail.
- Dive computers are expensive to purchase and service.
- Each diver must have a separate computer.
- The mathematical model varies from one type of computer to another, so some confusion results when each diver of a buddy team uses a different type of computer.

- A multiple-step ascent rate indicator
- Altitude adjustment
- A dive-planning mode
- Backlighted displays
- Audible alarms

An integrated, wireless-transmitting, digital-compass dive computer provides dive-planning information before, during, and after your dives. This type of computer is a highly desirable device. Central processing units (CPUs) in closed-circuit rebreathers (CCRs) are also sophisticated dive computers that provide highly technical dive-planning information.

The advantages of dive computers greatly outweigh the disadvantages, and computers have become an almost essential item for scuba divers.

Postdive Review

After a dive, you and your buddy should reflect on your experience. How closely did the actual dive match the dive that you planned? If there were deviations from the plan, what caused them? Could you have prevented the deviations with a different plan or approach? What changes can you make to improve the next dive? Some problems require research, or you may need to ask the advice of a diving professional. The experience of each dive should affect your plans for future dives. Your dives with your buddy should progress more smoothly each time you dive together. Each time you visit a dive site, your dive procedures should improve. A review of each dive with your buddy and a discussion about future diving are valuable parts of dive planning. Even if you have a new dive buddy for a dive, you should plan your dives together and discuss the experience afterward. It is a good idea to keep notes about your dives on a waterproof slate and transfer unusual experiences to your dive log along with the usual data. Reflecting on past experiences from time to time will help you become a better diver.

Summary

The saying "If you fail to plan, you plan to fail" is true for scuba divers. All phases of dive planning are important and help ensure enjoyable and successful underwater experiences. Follow the recommended steps for planning, get area orientations when appropriate, have contingency plans, and discuss your dives with your buddy.

Dive profiles are a large portion of dive planning. Plan to dive conservatively. No dive-planning device can guarantee that you will not develop decompression illness after diving. The deeper, longer, and more frequently you dive, the greater your risk of decompression illness. Limit multiple-day and multilevel dives because repetitive multiple-depth profiles make you more susceptible to decompression illness. Make a rest stop at the end of every dive, avoid improper profiles, and have surface intervals of an hour or more. After three consecutive days of repetitive diving, wait a full day before diving again. Exercise good judgment and common sense.

Application-of-Knowledge (AOK) Questions

1. What is the most difficult task of dive planning?
2. How can you make dive planning easier and be more likely to accomplish it?
3. You have a dive planned, the weather and water conditions are good, and you catch a cold virus the day before your dive. Cold medicine minimizes your congestion. What should you do?
4. A friend who has more diving experience than you invites you to dive on a shipwreck that is located at a depth of 120 feet (36.6 m). How should you respond to his invitation?
5. Would you rather plan a dive by using a dive computer or by using dive tables? What is the advantage of each approach?
6. You have been diving at a resort for a week. None of your dives have required decompression other than safety stops. Your plane leaves the day after tomorrow at 10 a.m. When should you stop diving?
7. If you have to go over a mountain pass with an altitude of 5,000 feet (1,525 m) to return home after diving, what action do you need to take?

Sergeant major fish in Cozumel

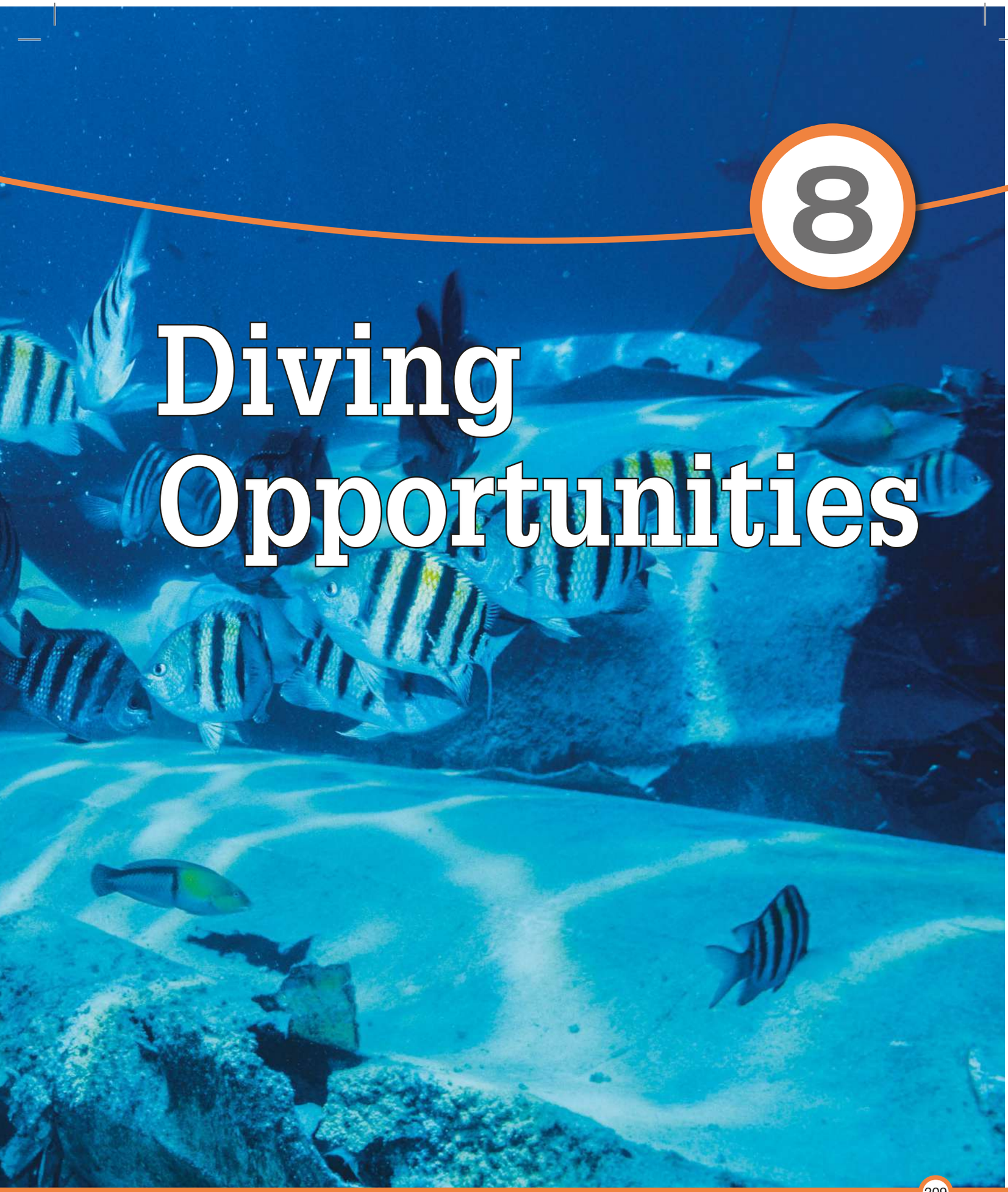

8

Diving Opportunities

Dive In and Discover

By the end of this chapter, you will be able to do the following:

- List 11 special-interest areas for novice divers and 6 special-interest areas for advanced divers.
- List at least five ways to continue your scuba diving education.
- List at least three ways to get involved with the local diving community.
- List several actions to take when preparing for a dive travel trip.

Continuing Education

A good diver never stops learning. You cannot learn everything you need to know about diving during a single course of instruction. After you complete your entry-level course, you can enroll in an intermediate or advanced scuba course immediately. These courses allow you to gain additional experience under supervision, help you develop important skills (such as navigation), and introduce you to special interest areas of diving. When you have identified a diving specialty you would like to pursue, you should complete a specialty course for that topic. A specialty course helps you begin enjoying the special area from the outset and helps you avoid mistakes and injury. Specialty areas that are of interest to new divers include the following:

- Underwater photography
- Underwater hunting and collecting
- Underwater environment
- Scuba lifesaving and rescue
- Night diving
- Boat diving
- Dry suit diving
- Drift diving
- Cavern diving
- Altitude diving
- Nitrox diving (enriched air)

Chestnut cowry by a starfish in California

As a qualified scuba diver, you can do many things. You can learn more, dive in different areas, be recognized for expertise, help others, or earn money, just to name a few. Opportunities abound.

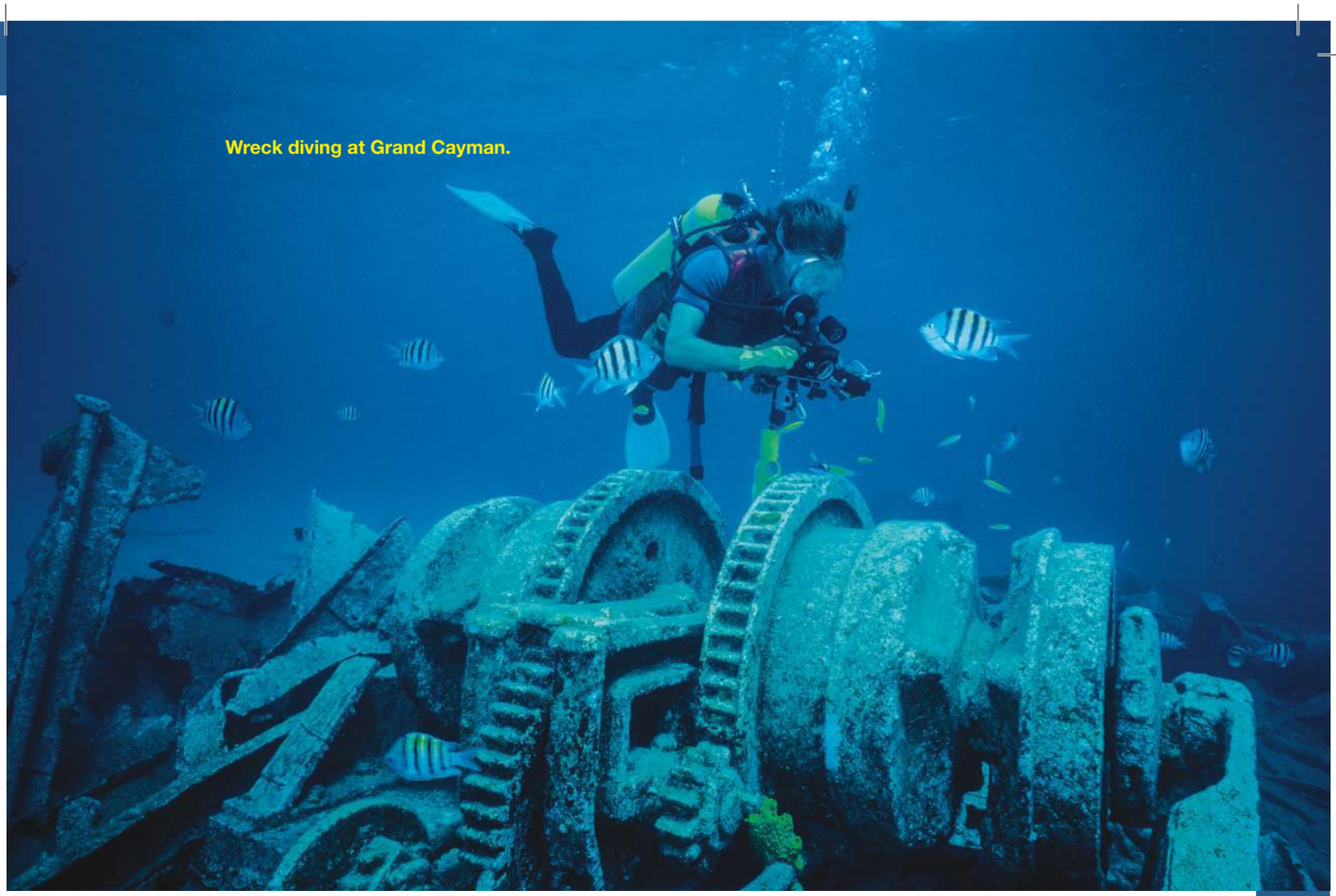

Wreck diving at Grand Cayman.

The following specialty courses may be of interest to experienced scuba divers with advanced training:

- Wreck diving (penetration)
- Ice diving
- Cave diving
- Deep and wall diving
- Light salvage
- River diving
- Search and recovery
- Rebreather diving
- Mixed-gas diving

Courses are only one way to learn about diving. Many diving seminars, workshops, and conferences are also available. Check the calendar section of diving publications to learn of events scheduled in your area. Diving changes constantly, so you need to keep updating your knowledge of diving medicine, diving equipment, and diving procedures. Continuing education programs give you the opportunity to learn from professionals who are continually involved in diving.

You can learn more about scuba diving by reading books, magazines, and newspapers. Subscribe to a scuba periodical and read all you can about diving. The appendix contains a list of scuba periodicals. You can also visit websites by and for scuba divers.

Another good way to continue your diving education is to join a dive club. Dive clubs offer many benefits to divers, including education. A talk on an interesting subject by a good speaker is usually part of the monthly meeting of a local club.

Continuing your education is important for your safety and enjoyment. When you have increased your knowledge about diving, you may become interested in helping others to learn.

Local Opportunities

You do not need to live in a coastal area to dive regularly or to get involved in diving. You can dive nearly anywhere you can find water, and there are dive stores, dive clubs, and diving events in noncoastal areas. Get involved in diving in your area right away. Find all the diving-related businesses and groups in your area: dive clubs, dive stores, dive boats, diving events, and diving publications. Join a dive club, attend the meetings, and participate in club dives. Find a local diving website to make contact with other divers in the area. Complete continuing education courses. Attend local diving seminars, workshops, con-

Underwater photography at Cozumel.

ferences, and shows. Talk to other divers when you attend diving events. Seize every opportunity to learn and every opportunity to dive with those who are more experienced than you.

Dive Travel

Many divers enjoy taking dive trips to exotic destinations, and there are thousands of beautiful and exciting diving locations to choose from around the world. If you plan to travel somewhere to dive, you need to know how to arrange dive travel, how to prepare to travel, and how to enjoy your trip.

Travel agencies that specialize in dive trips provide valuable information, so it is a good idea to book your trip through one of them. You might want to consider a dive package that includes accommodations and diving. Dive resorts advertise these packages in diving publications and often offer packages for bargain prices at dive shows. Many resorts have toll-free numbers you can call to obtain information about dive packages. To ensure the best value for your money, compare several packages before you book a trip. If a brochure from a dive resort is out of date, check with the resort for current offers.

When planning a dive travel trip, you need to decide whether you want to dive from a live-aboard dive boat or from a shore-based dive resort. A live-aboard trip allows you to dive in a variety of locations. The vessel moves from place to place throughout the day. You dive in remote areas that only a few divers visit. You may find a live-aboard dive trip to be a wonderful experience if diving is the only objective of your trip and you are not affected by motion sickness.

Consider a shore-based dive trip if you do not care for confinement and if you want to participate in activities other than diving. Many beautiful islands and resorts offer a variety of things to do in addition to diving. Nondiving members of your travel party can usually find many sources of enjoyment at a land-based diving destination.

Many diving publications and diving websites are available to help you decide where to dive, although making a choice from all the fantastic destinations available can be difficult. Recommendations from other divers, trips sponsored by your dive club or dive store, and travel presentations can help you decide.

When you have decided on the region for your dive trip, get as much information as possible about diving in the area. Obtain and review brochures, books, articles, videos, and any other materials about the region. The Internet is another excellent resource for information on dive locations. Talk to divers who have been to the location where you plan to go. The more you know in advance, the more enjoyable you can make your experience.

After you select your diving destination, you should make your reservations well in advance, confirm all arrangements in writing, and get clarification concerning cancellations, refunds, and so forth.

Research and plan your dive trip by learning about the water temperature for the time of year when you will be diving. Then obtain the appropriate exposure protection. Remember that too much insulation is better than too little. Also make sure that all your equipment is in good working order.

Plan to avoid sunburn. Sunlight in tropical areas is more intense than in other climates. Use a sunscreen, and unless you are well tanned, keep yourself covered at all times, even while in the water. Use lip balm with a sunscreen ingredient.

Underwater salvage operations.

You cannot develop a tan in a few days, so it is silly to try if you do not already have one. Sunburn can ruin an expensive vacation, and donning an exposure suit is painful when you are burned. It takes only a few minutes for an untanned person to get burned in the tropics, so be careful!

Obtain all required documentation well in advance. Obtaining passports and visas can take months. Be sure to find out whether the country to which you are traveling requires immunizations.

Learn what to expect when you arrive at your destination. Know the frequency and voltage of the electricity; take converters if you are bringing electrical devices that operate on a different voltage. Know the monetary exchange rate. Consider taking a small pocket calculator to help compute the conversion of money.

If you have an expensive camera, video equipment, or jewelry, take your property to a customs office before you leave the country and have it documented to avoid being charged duty on the items when you return home.

Do not procrastinate in making your preparations. Procrastination can cause a cancellation of your trip, a forfeiture of your deposit, and a great deal of frustration. Consider travel insurance.

Limit what you pack for your trip. You do not need much clothing for a dive trip unless you plan to attend formal events. A few pairs of shorts and T-shirts, some swimsuits, and a couple of sets of nice clothes for dinner should suffice. Experienced dive travelers travel light when it comes to clothing and personal effects.

Diving destinations provide tanks and weights, so you do not need to take those items with you. When you pack for your trip, keep in mind that there may be weight limitations for your baggage. Excess baggage costs can be very high. To guard against theft, avoid flaunting expensive diving, photography, and video equipment. Use inconspicuous containers to ship your equipment, and it will be less likely to disappear. Insure luggage that contains expensive equipment. Items I always take with me are mask, snorkel, fins, exposure suit, boots, gloves, regulator, buoyancy compensator, and dive computer. I rent tanks and weights. My gear can be packed in a dive bag, but I prefer to use a suitcase for better protection of the equipment. I pack a dive gear bag in the suitcase because I don't want to use a suitcase at the dive site. There are also dive gear bags that have wheels, which are useful when you can roll the bag.

Be prepared for travel illnesses. Obtaining medications for nausea, diarrhea, and colds is easier before a trip than it is if you become ill in a foreign land.

Flying is the most common method of transportation for dive travel. Air travel usually produces jet lag and dehydration. A viral illness a couple of days after reaching your destination is common because many virus germs are concentrated inside the cabins of airliners. Here are some tips and suggestions concerning air travel:

Schedule your flight to arrive the day before you start diving. It is unwise to plan to dive the same day you arrive.

Drink a full glass of water or fruit juice every hour during your flight to avoid dehydration. Avoid alcohol, milk, and drinks containing sugar. The humidity in an aircraft is about 8 percent, so you will become dehydrated unless you drink plenty of fluids.

Avoid eating heavy meals and salty foods when flying. Light foods cause less dehydration than fatty foods. Consider ordering a special meal of salad and fruit, or bring your own food if meals aren't served on the flight.

Wash your hands frequently when traveling. Germs from your hands can get into your eyes and nose and lead to viral infections.

Remember that baggage compartments on airliners are unpressurized and that the reduced atmospheric pressure can damage your gauges. Therefore, you should either bring your gauges inside the cabin in your carry-on luggage or store them in airtight containers.

As soon as possible after reaching your destination, get some exercise and drink plenty of liquids. Limit your consumption of alcohol, which causes dehydration. Getting intoxicated at the outset of a diving vacation is one of the worst things you can do because you increase your chances of getting DCS.

Confirm your return air reservations as soon as you arrive, especially if you travel to another country. If you fail to do this, you are likely to lose the reservations. You could end up arriving at the airport to go home and discovering that your tickets are not valid.

Your first dive at your destination should be an orientation dive. In addition to learning about diving in the area, you should use the dive to check your buoyancy and make any needed adjustments. Work out any equipment problems during a dive in shallow, still, calm water. Avoid deep dives or moving water until you have acclimated to the area.

After three days of repetitive diving, refrain from diving for a day so that your body can outgas. Go shopping or take a land tour for a refreshing change of pace. Allow one day between your final scuba dive and your scheduled flight home.

SCUBA WISE

Scuba equipment is a means to an end—a mechanism that enables you to go to areas that are otherwise unreachable by the average person. I think of scuba diving as a means of transporting myself to a destination where I can explore or participate in a special activity. Scuba diving without an objective loses its thrill after only a few dives. You need to develop an interest area, which can be anything from bottle collecting to wreck diving. When you have an objective for your dives, diving can be extremely rewarding. Over the years, I have participated in many different diving activities. My favorite underwater pursuit is photography. I can collect priceless memories without removing anything from the environment. In my opinion, underwater photography is the most challenging and the most satisfying dive activity. Select a diving activity that appeals to you, complete training for the activity, and then pursue it. You may choose several different activities. No matter what you choose to do while diving, I hope that your experiences will be as wonderful as mine have been.

The nondiving day before flying helps prevent DCS caused by altitude, and it gives you the opportunity to rinse, dry, and pack your diving equipment. Snorkeling is a good last-day activity, and your mask, snorkel, and fins dry quickly for last-minute packing.

Dive travel can be fun, exciting, and memorable when you plan it well. You can avoid unpleasant, frustrating, and disappointing experiences by researching your destination and preparing properly. Good diving to you!

Dive Destinations

One of the great pleasures of scuba diving is visiting different regions of the country and the world. Fantastic dive sites can be found nationally and internationally. The following areas are popular dive destinations that you might want to visit. The pictures throughout this book were taken at these destinations.

California

Dense kelp beds that form surface canopies make diving in California an unforgettable experience. The kelp provides shelter for many forms of life. Rock formations provide anchorage for the buoyant kelp and support additional life-forms, such as abalone, lobster, and moray eels. An expedition on a live-aboard dive boat to one or more of California's offshore islands during the summer months is an extremely worthwhile scuba experience. Shore diving in California can be difficult because of surf and low visibility. An area orientation is essential before attempting to dive through surf.

Washington

Puget Sound, which was carved into the landscape by glaciers, forms a gigantic body of ocean water that is much more protected than the coastal waters. The Sound contains many shipwrecks and artificial reefs, which are havens for many life-forms. Puget Sound is the home of the world's largest octopus as well as giant lingcod, monkey-faced eels, and beautiful orange and white anemones called metridians. The water is cold, but the visibility is 15 to 20 feet (4.5 to 6.1 m) most of the time. Dry suits are recommended.

Florida

Florida features some of the best and most diverse diving in the United States. From crystal clear springs and rivers to the Keys, there are many spectacular dive sites. Cavern and cave diving are popular, but training for these specialties is mandatory. The Florida coast has abundant shipwrecks. The Keys boast coral reefs and colorful fish. All scuba divers should definitely have a dive excursion to Florida included in their dive log.

New England

The cold waters of the New England area offer a great variety of diving opportunities. Wrecks, reefs, kelp, and creatures captivate your attention. Lobsters with claws and other interesting life-forms are great attractions. If you are able to schedule some time to dive when you travel to this part of the country, you should try to include the adventure in your itinerary.

Great Lakes

These great bodies of freshwater boast the most well-preserved shipwrecks in the United States. Many of these wrecks are in deep water, so you must have specialty training to explore them. Divers who are unfamiliar with the Great Lakes should use the services of a divemaster. If you like wreck diving, this is one of the places to include on your must-see list.

Hawaii

Hawaii offers exciting diving that includes good visibility, spectacular underwater formations, beautiful reef fish, large turtles, and giant marine mammals. Hawaii is a popular diving destination that you should visit if you get the chance. Aloha!

Caribbean

The Caribbean is one of the world's most popular places to dive, and for good reason. The area includes reefs, wrecks, drop-offs, and life-forms that are unique and breathtaking. Visibility is usually excellent. Many dive resorts in the area offer trips to phenomenal locations. Nearly every experienced diver talks about Caribbean diving experiences.

South Pacific

The islands in the South Pacific offer some of the most beautiful diving to be found anywhere on Earth. Divers can see fantastic soft corals, sea fans, and incredible fish. Visibility is outstanding. Many of the island waters feature wrecks and war artifacts. The beauty found above and beneath the water in this part of the world makes trips to the South Pacific highly desirable for scuba divers.

Underwater photography in Southern California

Middle East

The diving destination in this region is the Red Sea, which is unparalleled for scuba diving. The shear walls, the abundant and colorful life-forms, and the crystal clear waters make this one of the most desirable of all diving destinations. If you can afford a trip on a live-aboard charter boat to offshore islands, you will be treated to unforgettable sights.

To learn details of the world's most popular diving destinations, you can do a search of the topic on the Internet. Many websites provide excellent information. Traveling to exotic dive destinations is one of the most enjoyable aspects of scuba diving.

Career Opportunities

Some people love diving so much that they want it to be their vocation. Career opportunities exist for those who are willing to put forth the effort to achieve their desires. Jobs within the recreational dive industry include resort dive guide, instructor, journalist, dive travel coordinator, retail salesperson, and sales representative for equipment manufacturers. People can also work in nonrecreational scuba jobs, such as scientific research, archaeology, engineering studies, hull cleaning, salvage, and underwater repairs. Two of the same rules that apply to recreational pleasure diving also apply to underwater work: (1) Complete training for an activity before you attempt it, and (2) use complete, proper equipment for the activity. Career training opportunities abound. If you are interested in diving as a vocation, you can find information about training in dive periodicals.

Night diving preparations at Catalina Island, California.

Leadership Opportunities

After you have become an experienced diver, you may want to help others learn and enjoy scuba diving. You can take courses to become qualified as a divemaster, assistant instructor, and instructor. Many people who love diving work part-time in the industry in leadership positions. When you are ready for leadership training, contact your scuba training agency for information.

Summary

Always remember that a good diver never stops learning. You should complete additional training, make sure you are trained for any special activity before you attempt the specialty, read as much as possible, join a dive club, and attend educational events. Take advantage of every opportunity to learn more about diving.

Increase your diving abilities and experience by applying what you learn. Participate in dive trips and charters. Research and plan your excursions thoroughly to maximize your enjoyment and minimize your disappointment. Many opportunities are available for diving adventures throughout the world.

Another opportunity for you as a diver is to make a positive contribution to the diving community. You can do this by conducting yourself in a responsible manner at all times and encouraging other divers to act responsibly. Help establish a good image of diving, get involved in issues that pertain to the diving environment, and help educate people who do not dive. Whether your contribution is small and personal or vast and international, you can make a difference. Decide now that your involvement in the wonderful pursuit of scuba diving will be a positive one.

Application-of-Knowledge (AOK) Questions

1. Why are you learning to dive? Do you just want to investigate a new activity, or are you interested in an activity that requires scuba diving?
2. What do you visualize as your progression of training as a scuba diver?
3. How often do you think you will dive? Will most of your diving be local or related to dive travel?
4. If a nondiving friend asked you what can be done to reduce the risk of injury while scuba diving, what advice would you provide?
5. What causes you the greatest concern about scuba diving? What can you do to minimize your concerns?
6. How do you see yourself as a scuba diver in 5 years? In 10 years?

This page intentionally left blank.

Appendix A

Associations and Periodicals

Diver Training Organizations

International Diving Educators Association (IDEA): www.ideascubausa.com/home_15.html

Multinational Diving Educators Association (MDEA): http://neptuneshop.wix.com/mdea#!

National Association of Scuba Educators (NASE): http://naseworldwide.org

National Association of Underwater Instructors (NAUI): www.naui.org

Professional Association of Diving Instructors (PADI): www.padi.com/Scuba-Diving

Scuba Diving International: www.tdisdi.com

Scuba Educators International (SEI): www.seidiving.org

Scuba Schools International (SSI): www.divessi.com

Environmental Organizations

Blue Frontier Campaign: http://bluefront.org

Cousteau Society: www.cousteau.org

Earthecho International: http://earthecho.org

Living Oceans Society: www.livingoceans.org

Marine Conservation Society: www.mcsuk.org

Oceana: http://oceana.org

Ocean Defense International: www.oceandefense.org

Oceanic Preservation Society: www.opsociety.org

Reef Check: www.reefcheck.org

Reef Relief: www.reefrelief.org/mission

Save Our Seas: http://saveourseas.com

Sea Shepherd Conservation Society: www.seashepherd.org

Coral Reef Alliance: http://coral.org

Ocean Conservancy: www.oceanconservancy.org

The Ocean Project: http://theoceanproject.org

NOAA National Marine Sanctuary Project: http://sanctuaries.noaa.gov

Scuba Diving Magazines

Advanced Diver: www.advanceddivermagazine.com
Alert Diver: www.alertdiver.com
Dive Training: https://dtmag.com
Diver: http://divermag.com
Scuba Diving: www.scubadiving.com
Undercurrent: www.undercurrent.org

Appendix B
AOK Answers

Chapter 1

1. Some people learn to scuba dive because they seek a new experience or adventure. Scuba diving training opens the gate to a beautiful underwater world, which can provide many pursuits. Scuba diving should be considered a mode of transportation that takes you to places where you can pursue underwater interests such as exploring, photography, and wreck diving. After completion of your basic scuba diver training, continue your education and obtain an advanced scuba diver rating. With the additional skills acquired with advanced training, you will be ready to learn specialty activities and seek the rewards of your chosen endeavors.
2. An open-circuit scuba system allows a diver to breathe compressed air from a pressurized cylinder and exhaust the used gases into the water. It is a simple, reliable, easily maintained air delivery system, which makes it the safest and most popular scuba system. A closed-circuit scuba system allows a diver to rebreathe exhaled air by removing carbon dioxide and adding the correct amount of oxygen without allowing gases to escape from the system. It is a complex air delivery system that requires specialty training and extremely careful maintenance. Closed-circuit scuba systems are very expensive compared to open-circuit systems.
3. A scuba certification card (C-card) indicates that you have completed initial training, but it does not reflect how recently you have been diving. Certification agencies recommend that you complete simple refresher training if you have not dived for a year or longer. This training is a review of your basic skills under the supervision of an instructor during an open-water dive. The confidence you gain from knowing that your skills are refreshed will make your diving safer. A good reason to log your dives is to document your diving activity. Some dive resorts and charter boats will want to review your log book to determine your diving experience, especially your most recent dives. Logged dives also can help you recall many facts that you would otherwise forget.
4. It is logical that the more time you can use your knowledge and skills of scuba diving under the supervision of a certified diving instructor, the better your knowledge and skills will become. Extended training also allows you to ponder your learning between training sessions. You have time to think of questions that will clarify concepts and increase your understanding of principles. Short scuba training courses do not provide these benefits. The goal of learning is to transfer knowledge and skills acquired in a controlled environment to the actual environment in which they will be used. Longer scuba courses develop a greater ability to transfer your

knowledge and skills, which increases your safety for your independent diving after training.

5. A medical exam for scuba diving should be performed by a scuba diving physician who understands medical issues that could be problematic for those considering scuba training. Knowledge of how pressure affects medical conditions can help prevent pain, suffering, and even death.
6. Actions you can take to minimize your risk of injury while scuba diving are to dive only when you are in good health, use well-maintained equipment, dive in good environmental conditions, and adhere to the safety rules of scuba diving.
7. When you are in the process of selecting an instructional course for scuba diving, you need to know the sanctioning national agency, the instructor's experience, the total cost of the course, the course curriculum and duration, and how much open-water (real environment) training is included.
8. You can show that you are a responsible diver if you do only what you are qualified to do, dive within your personal limitations, adhere to safety rules, and remain aware of how your conduct will be viewed by others—divers and nondivers—when you go diving.

Chapter 2

1. Your exertion has used much oxygen, which has produced much carbon dioxide (CO_2), which stimulates you to breathe more to expel the CO_2. The feeling of a lack of air is caused by the need to reduce the amount of CO_2 dissolved in your blood. You need to minimize your activity immediately and breathe deeply until your breathing returns to normal.
2. Rapid, shallow breathing (hypoventilation) moves only a small amount of air in and out of your body. The structures from your mouth to your lungs do not allow gas exchange, so your body is not getting oxygen or ridding itself of CO_2. Gaining control of your breathing by stopping or reducing activity and breathing deeply until you feel better is the correct course of action and can help prevent panic. When your breathing is under control, you can think clearly and manage the difficulty that caused your anxiety.
3. Many factors affect air consumption, but the primary cause of high air consumption is exertion. Relax and move slowly. If you are too buoyant and have to exert to maintain your position in the water, your air supply will be depleted rapidly. If you are wearing too much weight, your attitude in the water will increase your frontal surface area and drag will increase air consumption. Adjust your weight and trim so you can hang suspended in a level position, and then relax and enjoy diving.
4. The first step is to tell yourself that you have time to manage this problem. The next step is to think of the options available to you. The quickest and easiest solution is to switch to your extra second stage or alternate air source. Your buddy's extra second stage is another good choice. The worst way to manage this problem is to spit out your regulator and swim to the surface.
5. Rapid and deep breathing before holding your breath is hyperventilation, which flushes carbon dioxide from your body. The buildup of carbon

dioxide when you hold your breath is what stimulates you to breathe. If you eliminate too much carbon dioxide from your body, you may lose consciousness while breath holding before you are stimulated to breathe. Loss of consciousness underwater can cause drowning. Hyperventilation should be limited to a few breaths before underwater breath holding to avoid what is called shallow-water blackout.

6. Your body loses heat rapidly in water. When you increase your heat production by exercising in cold water, you lose more heat than you would if you had not increased exertion. If you are wearing a wet suit, ascending to shallower water will increase your insulation as the water pressure decreases. Avoid moving directly into a current, which will push cold water through your suit. Your best action is to obtain better insulation for your entire body for future dives.

Chapter 3

1. The current is pushing cold water into and through your wet suit. If you swim at an angle to the current, you will reduce the flow of cold water through your suit and not get as cold. You can also look for places where you can be shielded from the current. You do not need to swim constantly when diving. Most of the life in the ocean is less than 1 inch (2.45 cm), so it is a good idea to stop and look closely at any underwater formations.
2. Some of the best and safest ways to warm yourself after diving are to remove your wet suit, get dry, get out of the wind and into a warmer environment, wrap yourself in a blanket or sleeping bag, eat high-energy food, drink warm drinks, and avoid stimulants. The principle is to warm yourself from the inside out instead of using external heat, which is dangerous. Remember that your core temperature will not be normal until you begin to perspire.
3. Did you think about the density of water and how much resistance to movement it causes? People usually think they will ascend rapidly if they lose their weights, but if you respond quickly, this situation is not difficult to manage. Quickly deflate any air from your BC. Then invert yourself to get your fins above your head, swim downward forcefully for a few kicks, grasp your weight belt, and replace it. (You will learn how to do this skill in chapter 6.)
4. When you ascend 10 feet (3.3 m) to swim over the reef, your buoyancy will increase because your wet suit and the air in your BC will expand. If the upward excursion is brief, you may be able to control your buoyancy just by keeping the amount of air in your lungs minimal. If the excursion is long or you are unable to control your buoyancy by breathing, vent some air from your BC to maintain neutral buoyancy until you descend again.
5. Remember that it is easier to equalize pressure in your ears if you descend feet first. It is good to use an anchor line for reference and control during a descent, but remain upright as you make your way down the line.
6. You should recall that the rate of change of pressure is greatest near the surface. Therefore, you need to equalize pressure in your ears and mask more frequently near the surface than will be required as the depth of

descent increases. If you can keep your air spaces equalized for the first 33 feet (10 m), you are not likely to have problems beyond that depth.

7. If you don't stop swimming, you will experience air starvation. And if you do stop swimming, your buddy will disappear and you may become separated. Prevention is best. Get your buddy to agree that you will swim side by side at all times and that when one of you stops, you both stop. If your buddy is ahead of you and you want or need to stop, use a rock or the butt of your dive knife to bang on your tank to get your buddy's attention. If your efforts fail, stop and get your breathing under control and then surface to reunite with your buddy if he does not return to you.
8. You are out of air, but only at a depth of 50 feet (15 m). As amazing as it sounds, if you start swimming slowly to the surface, there will be additional air to breathe from your tank as the surrounding pressure decreases with ascent. This is not the best option, however. In chapter 6 you will learn other ways to deal with situations involving loss of air supply.

Chapter 4

1. Minimize sun exposure, soak and rinse after use, store in a dark and dry location that is free of smog, inspect your gear regularly, and have your equipment serviced regularly or as needed.
2. The most important concerns for most diving equipment are fit and comfort. Features, colors, and price do not matter if the equipment does not fit well or is uncomfortable to use. If possible, rent equipment before you make a purchase so you will know if an item is right for you.
3. If all of the air escapes from your tank and the valve in your second stage is open (or if you happen to depress the purge button and open the valve), water can enter your tank. You will learn how to minimize the chances that your regulator will free-flow. If you detect water inside you scuba cylinder, take it to a facility for maintenance as soon as possible.
4. Air has weight. The air in a typical scuba tank weighs five to six pounds (~2.25-2.7 kg) when the tank is full. As you deplete the air, your buoyancy increases. The loss of weight may not be apparent at depth if you wear an exposure suit because pressure compresses the suit. You simply add less air to your BC than you would if you were not wearing an exposure suit. Problems arise as you near the surface with a scuba tank that is nearly empty because your suit expands and you are unable to overcome the positive buoyancy caused by the loss of weight from your cylinder. To avoid this problem, you need to adjust your weight to be neutrally buoyant just below the surface while using a tank with approximately 500 psi of air remaining in the tank.
5. You need just enough air in a dry suit to keep it from squeezing you. Any excess air forms a bubble inside the suit, and the bubble rises to the highest point. If you tip upside down to do something, the bubble can rise to your feet, may loosen your fins, and can cause you to float upward in an inverted position. This problem is just one example of why you need special training in the use of a dry suit.

6. A compass is a magnet, and magnets attract one another. Placing your compass near your buddy's compass causes the needles to shift position and the readings to be inaccurate. You need to keep your compass away from any items that can cause the needle to deviate from its natural direction.
7. The best way to clean your diving equipment is to soak it in warm water, rinse it gently with flowing water, and dry it thoroughly before storing it. Some equipment parts, such as dry suit zippers, require scrubbing. Follow the manufacturer's instructions.
8. The attachments to a scuba regulator first stage may include a cylinder pressure gauge, a primary second stage regulator, an alternate air source second stage, a BC inflator hose, and possibly a dry suit inflator hose.
9. The best reasons for having good dive equipment and maintaining it properly are safety and comfort. You will know when your equipment is right when you are not concerned about it while you are diving.

Chapter 5

1. The water temperature affects the equipment you use, your body temperature, your diving procedures, the visibility underwater, and the underwater life, among other things. Water temperature has a profound impact on your diving adventures. Knowing what to expect and how to be prepared for the temperature of the water is essential.
2. Water visibility is affected by depth, pollution, water movement, weather conditions, algae, temperature, your movements, and other factors. Check present diving conditions with your local dive facility and postpone diving when visibility is poor.
3. A good way to manage currents is to know in advance what to expect. Obtaining training for diving in various currents is important. Being aware that the current is weaker at the bottom than at the surface is helpful. Swimming at right angles to a current may help you get out of some currents. Diving against the current is a common practice, but in some situations it is best to drift with the flow of the current.
4. When you encounter a strange animal underwater, respond by stopping movements, respecting the animal, observing it from a distance, and refraining from provoking or feeding the animal. Nearly all marine animals are defensive, not aggressive. Think and respond. The worst action to take when you see a strange or large animal is to panic and ascend rapidly to the surface.
5. No matter how much training or experience you obtain, you will always need to acquire local knowledge about diving in a new area. Watching and mimicking experienced local divers are a good way to avoid embarrassment, but learning in advance about why the divers use the techniques they use is even better.
6. While it may be difficult to accept until you gain experience, the safest and most comfortable place to be while diving is near the bottom. It is much more uncomfortable and dangerous at the surface of the water.

7. The transition from saltwater diving to freshwater diving requires changes that include the amount of weight you wear and perhaps the type and thickness of your exposure suit.
8. When you plan to dive in a new area, you need information about the area. Weather conditions, seasons, forecasts of diving conditions, area hazards, equipment requirements, and local diving procedures are important to know in advance. Chapter 8 provides resources to help you obtain the information that you need.

Chapter 6

1. Buoyancy can be controlled by the amount of weight worn, the amount of air in your BC, and the amount of air in your lungs. Think of these methods as large, medium, and small ways of control, respectively.
2. To ascend properly from a scuba dive, alert your buddy, hold your BC inflator-deflator with the mouthpiece facing down and the valve open, reach up, swim up, ascend slowly, monitor your rate of ascent and your buddy, turn slowly, stop at a depth of 10 feet (3.05 m) for precautionary decompression, ascend to the surface, and make a complete turn to assess your surroundings. Ascending is a complex skill. Learn it well.
3. Successful buddy contact requires predive planning and communications, agreeing on a direction and maintaining that direction until you both agree on a new direction, maintaining the same position (side by side) relative to one another, holding hands or using a buddy line when visibility is poor, and committing to reunite using standard procedures if you become separated.
4. This situation can be frightening unless you respond quickly. All you have to do is raise and open the deflator valve with one hand and disconnect the low-pressure hose with the other hand. You do not need to terminate your dive because you should have the skills needed for manual inflation of your BC.
5. When a regulator free-flows and will not stop, simply switch to your extra second stage and ascend. Remove your tank at the surface and turn off the air supply. If you allow the tank to empty and the second-stage valve is stuck in the open position, water can get inside your cylinder.
6. To exit the water into a small boat, remove your weight belt and place it in the boat, inflate your BC about half full, and remove your scuba unit. Next, grab the rail of the boat and, while holding the regulator hose from your tank under one hand, kick yourself up and into the boat. Remove remaining equipment after you are in the boat.
7. Swimming against a current can be tiring. If there is a nearby obstruction, such as a rock formation, that you can get behind to minimize the current and rest, seek shelter. If there is an outcropping you can grasp that will allow you to rest and catch your breath, use it. If the bottom is sandy, plunge your knife into the bottom to maintain your position while you recover. The most important action is to reduce your exertion and get your breathing under control. If the situation is unmanageable, surface and drift back to your exit point.

8. The assistance of your buddy is the best way to get free of an entanglement, but you can also remove your scuba unit, clear the entanglement, and replace your equipment. Be sure you learn the skills of equipment removal and replacement well so you can manage minor problems that may occur while scuba diving.

Chapter 7

1. The most difficult task of dive planning is getting started! Most people agree that planning is a good idea, but they procrastinate. You can be successful with almost any task if you push yourself to get started. If you keep a list of things to do, consider adding dive planning to that list.
2. To make dive planning easier, find a buddy who would like to dive with you and agree on which items each of you will plan. Set a date to talk again about your plan. By taking responsibility and setting a date, you will have less work to do and will be more likely to meet your goal of developing a good dive plan.
3. If your health is not normal, do not dive. Congestion from a cold can make it difficult to equalize pressure in your ears and sinuses. Pressure can cause the medicine to terminate sooner than it normally would. When decongestants stop working, air passages can swell and close rapidly and trap air. Notify your dive buddy that you need to postpone your dive trip until you are well again.
4. Assuming you are a novice scuba diver, it would be unwise to attempt a dive on a wreck in deep water. Shipwrecks often occur where conditions, such as currents, can be treacherous. Tell your friend that you are extremely interested but that you need to complete advanced and specialty training so you will be adequately prepared for the dive.
5. When you dive using a dive computer, you avoid the dive table penalty rule of having to assume that all of your dive time was spent at the maximum depth. The computer also stores valuable information about your dive, helps you control your rate of ascent, and indicates when you need to stop for decompression. The disadvantage of a dive computer is that it is an electronic instrument that can fail. Dive tables do not need batteries. The best approach to planning time and depth limits is to develop a contingency plan using dive tables, write the plan on an underwater slate, and take it with you when you dive with your computer. If your computer stops working or you inadvertently turn it off, your contingency plan can help keep you safe.
6. You need to plan. After multiple dives, you should wait 18 hours before flying, so your diving needs to end today.
7. When driving decreases pressure with altitude after you have been diving, you need to consult the altitude delay timetable. If your driving altitude reaches 5,000 feet and your starting repetitive-dive group is 1, you need to wait 3 hours 4 minutes before driving over that pass. This requires planning. If your altitude delay timetable is at home in your book, you won't know how long you need to delay your journey.

Chapter 8

1. Some people learn to dive just to experience the activity. They complete training, purchase equipment, make a few dives, and stop diving. Other people continue training, get involved in the diving community, pursue an underwater hobby, and dive for many years. If you think of scuba diving as a means to an end, it can provide wonderful experiences that will keep you involved for many years.
2. You might see yourself completing an advanced diver course and several specialty courses. It helps to gain some dive experience in addition to your training dives. Just be sure that your excursions are limited to the conditions and depths to which you are qualified.
3. If you are not certain that you will dive regularly or that you will dive only locally, consider renting the major equipment (cylinder, regulator, exposure suit, buoyancy compensator, and weight belt). If you find yourself diving frequently, buy your gear. Renting equipment will help you make a good selection. If you plan to dive primarily at dive destinations, rent the major equipment at the destination. When you become an avid scuba diver, you are likely to have full sets of equipment for both cold- and warm-water diving.
4. The best ways to reduce the risk of injury from scuba diving are to be trained for what you do while diving, have the proper equipment for your dive, and gain experience gradually. Making a dive that greatly exceeds your training and limitations without first gaining several experiences that gradually increase your training and limitations is unwise.
5. Phobias are natural, but if something terrifies you and it is related to diving, learn how to deal with the fear before you are exposed to it. For example, many people fear sharks because of what they have seen in media. What you should know is that dangerous shark encounters are rare and most likely occur at the surface of the water. If you see a shark, remain calm, stay on the bottom, and enjoy the exciting experience.
6. When you look to the future, do you envision yourself diving regularly? Are you just going to wait and see what happens, or do you intend to continue your training, get involved, and become a frequent diver with many friends who also enjoy diving? You can dive for a lifetime and not even begin to see everything the underwater world has to offer. May all your diving experiences be enjoyable and memorable!

Glossary of Scuba Diving Terms

absolute pressure—The total pressure exerted on a diver, which is measured by adding atmospheric pressure to gauge pressure.

actual bottom time—The elapsed time in minutes from when a diver leaves the surface in descent until the diver begins a rest stop or surfaces.

alternate air source—A source of compressed air other than the diver's primary scuba regulator. The two primary types of alternate air sources are extra second stages and backup scuba units.

ambient pressure—The surrounding pressure.

Archimedes' principle—This principle states that the force of buoyancy acting on a submerged object equals the weight of the water displaced.

arterial gas embolism (AGE)—An embolism resulting from an air bubble blocking the arterial circulation.

atmospheric pressure—The pressure exerted by the atmosphere.

backpack—The equipment that holds the diver's scuba cylinder.

barotrauma—Trauma or injury caused by pressure.

bends—See decompression illness.

bezel—A movable ring on a waterproof watch (being used as an underwater timer) or compass that you can set to indicate elapsed time or direction.

blooms—Overpopulations of plankton that can color the water, destroy underwater visibility, and form toxins in animals that feed by filtering water.

blowout plug—A plug included in a submersible pressure gauge to relieve pressure in the housing in the event of a high-pressure leak.

Bourdon tube—A type of depth gauge that uses a thin metal tube formed into a spiral. The movement of the coil, linked mechanically to a needle, indicates the amount of pressure exerted on the gauge.

Boyle's law—This law states that for any gas at a constant temperature, the volume varies inversely with the absolute pressure, while the density varies directly with the absolute pressure.

buddy breathing—The sharing of a single regulator second stage by two divers.

buddy line—A short line used by two divers to keep in contact with each other when visibility is poor.

buddy system—The practice of teaming up with a qualified partner for activities that involve risk, such as skin and scuba diving. A buddy provides reminders and assistance and sees things that you might not see.

buoyancy—The upward force of water.

buoyancy compensator—An item of equipment that helps a diver control buoyancy. You can inflate your buoyancy compensator at the surface to increase

buoyancy, deflate it to reduce buoyancy for descent, and add air to it to achieve neutral buoyancy underwater.

burst disk—A thin metal disk that is a standard feature of the valves for scuba tanks. If a tank is overfilled or the heat from a fire causes the tank pressure to increase to a hazardous level, the disk will burst and vent the tank to prevent an explosion.

capillary gauge—A type of depth gauge that is made up of a hollow, air-filled, transparent plastic tube sealed at one end and placed around a circular dial. Water pressure compresses the air inside the tube during descent (based on Boyle's law), and the position of the air–water interface inside the tube relative to markings on the dial indicates the depth.

cavern—A large, roomlike opening in a natural formation where light from the surface can be seen.

caves—Openings in a natural formation that extend farther than a cavern where no surface light can be seen.

C-card—The certification card you receive when you complete your training requirements as a scuba diver.

ceiling—The minimum depth that a diver cannot rise above without the risk of DCS.

ciguatera—Fish poisoning that results from eating fish that consume a certain species of algae.

closed-circuit—Refers to a scuba system that eliminates carbon dioxide from exhaled breath and resupplies oxygen without allowing gases to escape from the system.

compartments—Mathematical models used to estimate gas absorption and elimination by various areas of the body.

compass course—A series of headings on a compass that leads to a destination.

compass heading—A direction that is set on a compass.

console—A display unit into which several instruments can be combined. An instrument console attaches to the high-pressure hose coming from the regulator first stage.

continental shelf—An underwater area that extends from land and slopes gradually to a depth of about 600 feet (183 meters).

contingency plans—Plans that address possible circumstances that could force changes in the original plans for a dive.

controlling compartment—The area of the body that determines how long a diver can remain at a given depth. The determination is made by how quickly a gas diffuses from that compartment.

crest—The top of a wave.

Dalton's law—A law stating that the total pressure exerted by a mixture of gases is the sum of the pressures that would be exerted by each gas if it alone were present and occupied the total volume.

decompression illness (DCI)—A serious medical condition in which a diver has neurological symptoms on surfacing. Also known as the bends.

decompression stops—Precautionary stops made during ascent in order to reduce the risk of decompression illness.

defogging—The process of removing the film of oil from the surface of the glass lenses on a scuba mask in order to prevent the mask from fogging underwater. Commercial defogging solutions can help keep your mask clear while you dive.

dehydration—A physical condition resulting from the excess loss of body fluids.

density—Weight per unit volume.

DIN valve—A newer type of threaded outlet for a scuba tank valve that has a recessed O-ring seal and withstands higher pressures than a traditional O-ring valve. Tank pressures in excess of 3,000 psi (204 atmospheres) require a DIN fitting.

dive profiles—Diagrams used to plan a dive by plotting the time and depth of the dive.

diver's push-ups—A buoyancy evaluation used to learn buoyancy control. When the diver is weighted properly, a full inhalation will raise the shoulders while the fin tips remain on the bottom, and an exhalation will cause the shoulders to sink.

dolphin kick—A type of fin kick in which a diver holds both feet together continuously and exerts force against the water using a wavelike up-and-down motion of the body.

downwelling—A compensating downward current that occurs when wind blows along a coast where there is a steep drop-off near shore.

drag—A force that slows movement.

drift—The speed of a current.

drift dive—A dive in which the divers move with the flow of the current.

eardrum—A thin membrane that separates the ear canal from the middle ear and transmits vibrations to the inner ear via a series of small bones that are attached to the membrane.

ebb—The flow of water away from an area because of low tide.

eddies—Swirling currents that result from non-linear water flow

embolism—A blockage of circulation.

emergency decompression—A required delay in ascent that a diver must take if the actual or total bottom time (ABT or TBT) exceeds the no-decompression-stop limit for a dive.

equalization—The process of keeping the pressure inside an air space in or attached to the body equal to the ambient pressure.

eustachian tube—The part of the ear that allows the equalization of pressure in the middle ear.

Farmer Johns—A type of wet suit design where the leg material extends upward to cover the chest and is held in place with shoulder straps

fetch—The area in which waves were created.

fetch length—The distance wind travels unobstructed. The length of water over which a given wind has blown.

flaring—A method of slowing an uncontrolled ascent by arching the back, extending the arms and legs, and positioning the fins so they are parallel to the surface.

flood—The flow of water into an area because of high tide.

flutter kick—The most common type of fin kick. The flutter kick is an up-and-down kick that you can do facing down, up, or to the side.

fronds—The portions of kelp that appear to be the leaves.

gauge pressure—The pressure indicated by a pressure gauge that reads zero at sea level (this type of gauge displays only the pressure in excess of one atmosphere).

Gay-Lussac's law—This law states that for any gas at a constant volume, the pressure of the gas varies directly with the absolute temperature.

gyres—Large circulating currents that move clockwise in the northern hemisphere and counterclockwise in the southern hemisphere.

half-time—The amount of time it takes a tissue to accumulate half of the gas it can hold at a given pressure.

heat exhaustion—A medical condition that occurs when the core temperature of the body is above normal and the victim has become dehydrated.

heatstroke—An emergency medical condition that occurs when the body temperature becomes so high that the body's temperature-regulating ability shuts down.

Henry's law—This law states that the amount of a gas that dissolves in a liquid at a given temperature is directly proportional to the partial pressure of that gas.

hyperthermia—Higher-than-normal body core temperature.

hyperventilation—Rapid, deep breathing in excess of the body's needs.

hypothermia—Excessive loss of heat from the core of the body.

hypoventilation—Rapid and shallow breathing that does not allow you to expel carbon dioxide from your lungs.

ingassing—The process of gas diffusion into a liquid.

jumpsuit—A one-piece dive suit.

J-valve—A cylinder valve that was designed to maintain a reserve of air to permit a normal ascent. The introduction of submersible pressure gauges for scuba tanks has rendered the J-valve obsolete.

kelp—Giant algae that produce long strands that a diver can become entangled in.

Kelvin—The absolute temperature scale for Celsius temperatures. To convert a Celsius temperature to Kelvin, add 273 degrees.

K-valve—A cylinder valve that is a simple on-off valve that operates like a faucet. You turn the valve handle counterclockwise to open it and clockwise to close it.

lubber line—A reference line on a compass that indicates the direction of travel.

mediastinal emphysema—A lung injury in which air is present in the tissues in the middle of the chest.

middle ear—The air space behind the eardrum.

modified frog kick—A fin kick in which the diver rotates the ankles so that the tips of the fins point outward, slides the fins tip first in an outward direction, then pulls the bottoms of the fins together quickly in a wide, sweeping arc.

multilevel dive profile—A dive that progresses from deep to shallow during a given period of time.

neap tide—The twice-monthly period of lowest tides, which occurs when the moon and sun are at right angles to each other in relation to the earth.

neutral buoyancy—A state of buoyancy that a diver achieves when the diver neither sinks nor floats when holding an average breath.

nitrogen narcosis—A detrimental effect caused by the increased pressure of nitrogen that occurs at a depth of about 100 feet (30 meters).

nitrox—A nitrogen and oxygen mixture with a higher percentage of oxygen than is found in air. The mixture reduces the effects of nitrogen at depth.

no decompression stop limits—The maximum time that a diver may stay at a specified depth.

nonreference descent—A descent made vertically in water without any external reference.

octopus—A term used to describe an extra second stage.

open-circuit—Refers to a scuba system that permits a diver to breathe compressed air from a pressurized cylinder and exhausts the used gases into the water. Open-circuit systems are the safest and most popular form of scuba system.

open-valve ascent—A method for maintaining neutral buoyancy during an ascent by keeping the BC inflator-deflator valve open while holding the valve in a special way.

O-ring—A soft, circular ring that surrounds the outlet for scuba tank valves.

outgassing—The process of gas diffusing out of a liquid. For diving, it pertains to the diffusing of nitrogen absorbed under pressure.

overturn—The movement of water from the top of a lake to a depth of about 60 feet as a result of water circulation caused by winds. This water movement carries oxygenated water to the bottom and leads to what is called spring diving conditions.

partial pressure—The percentage of the total pressure exerted by each gas in a mixture of gases.

perfusion—The circulation in a tissue.

plankton—Aquatic life-forms that drift with the currents.

pneumothorax—A lung injury in which air is trapped in the chest cavity.

pony tank—A backup scuba unit consisting of a small scuba cylinder with a separate, standard regulator.

ports—The openings in the first stage of a regulator. One of the ports is for high-pressure air measurement with an SPG. The remaining ports are for low-pressure air.

pressure—Force (often weight) per unit area.

pulmonary barotrauma—Any lung injury caused by pressure.

Rankine—The absolute temperature scale for Fahrenheit temperatures. To convert a Fahrenheit temperature to Rankine, add 460 degrees.

reciprocal compass course—A course where the diver moves in a given heading for a given distance and returns to the origination point by reversing the direction of travel.

red tide—Blooms created by a type of red phytoplankton.

reference descent—A descent that the diver controls by following a line or the slope of the bottom.

repetitive dive—Any dive made within 6 to 24 hours (depending on the dive-planning device) of a previous dive.

repetitive group or repetitive group designation—The letter on a dive table that indicates the amount of nitrogen in a diver's body.

residual nitrogen—Nitrogen remaining in your system from a dive made within the past 12 hours.

residual nitrogen time—An amount of time that must be added to the actual bottom time (ABT) of a repetitive dive to determine the total bottom time (TBT). RNT compensates for residual nitrogen remaining in a diver's tissues from previous dives.

rest stop—A precautionary decompression stop during ascent.

reverse block—Discomfort that occurs when the pressure in the middle ear is above normal during an ascent due to blockage of the Eustachian tube.

reverse thermocline—The point of transition from a layer of cold water toward the surface of a lake to a layer of warmer water toward the bottom. This state exists in winter.

rip current—The narrow, strong current that moves away from shore as a result of water flowing back to sea through a narrow opening in an underwater obstruction.

sawtooth dive profile—A dive that progresses from deep to shallow and back to deep.

scissors kick—A fin kick that can be used as a resting kick. While lying on the side, the diver slowly extends one leg backward while extending the other leg forward, and then pulls the legs together quickly.

scombroid—A type of fish poisoning that can result if you eat fish that have not been kept chilled. It produces nausea and vomiting within an hour.

scrolling—A method used by dive computers to provide advance-planning information through a sequential display of time limits for various depths.

scuba—Self-contained underwater breathing apparatus.

semi-closed-circuit—Refers to a scuba system that combines the benefits of closed-circuit and open-circuit systems. Small amounts of used gases are released periodically.

set—The direction assumed by a current.

shallow-water blackout—The sudden loss of consciousness near the surface of the water.

shorty—A one-piece dive suit with short arms and legs used for diving in warm water.

sink—The area formed when the ground collapses into an underground cave system.

siphon—An opening where water is channeled from a sink back into the system.

skip breathing—An attempt to extend one's air supply by holding each breath for several seconds.

slack water—The period of time between flow and ebb when the water has minimal movement.

Spare Air unit—A backup scuba unit consisting of a small scuba cylinder with a special regulator integrated directly into the valve. A Spare Air unit is smaller than a pony tank and provides only enough air to permit an ascent from shallow depths.

spike dives—A dive with a short actual bottom time (ABT), such as a dive to free a fouled anchor.

spring tides—The twice-monthly period of highest tides, which occurs when the moon and sun are aligned with the earth.

square compass course—A series of headings that allows the diver to travel in a square pattern so that the dive ends where it began.

squeeze—A condition that occurs when the pressure on the outside of an air space in or on the body is greater than the pressure inside the air space. This condition may cause pain and injury and may be avoided by keeping the pressures equal (see equalization).

stages—The two parts of a scuba regulator that perform the two steps in the pressure reduction process.

stand—A brief period of time during which the tide neither rises nor falls.

step dive—A dive that is made in a series of steps.

stipes—Long strands of kelp.

subcutaneous emphysema—A lung injury in which air is present in the tissues under the skin. The air swells the tissues around the neck.

submersible pressure gauges—An instrument that measures scuba cylinder pressure.

surf—Breaking waves that occur when the water within the wave moves forward and gives up its energy. This occurs when the depth of the water is about the same as the wave height.

surface interval time—The elapsed time from when a diver surfaces after a dive to the time the diver begins the descent of a repetitive dive.

surge—The back-and-forth subsurface motion of the water within waves when the waves enter shallow water.

swells—The rounded, undulating forms that waves take as they move away from the area in which they were created.

tetrodotoxin—A type of fish poisoning that results from eating exotic fish such as puffer fish or blowfish. Tetrodotoxin is the most serious form of fish poisoning and can cause death within minutes.

thermocline—An abrupt transition from a warmer layer of water to a colder layer of water.

total bottom time—The sum of actual bottom time and residual nitrogen time.

Toynbee maneuver—A method of opening the eustachian tubes by blocking the nostrils, closing the mouth, and swallowing.

trail line—A long line with a float attached that is deployed from the stern of a boat. Divers use the trail line to pull themselves to the boat against the current.

trapdoor effect—A condition that exists when pressure prevents the Eustachian tube from opening. This condition may be overcome by ascending to reduce pressure.

trough—The bottom of a wave.

tsunami—Gigantic seismic waves caused by an underwater earthquake.

turbid—Water that is unclear or murky because of stirred-up sediment.

upwelling—The replacement of warm water at the surface with colder water flowing up from the depths as a result of a strong wind blowing along the shore for a sustained time.

Valsalva maneuver—A method of opening the eustachian tubes by blocking the nostrils, closing the mouth, and gently trying to exhale.

valve seat—The portion of the cylinder valve that closes and stops the flow of air.

vasoconstriction—Narrowing of the blood vessels.

vertigo—A subjective feeling of movement perceived as a spinning sensation.

wave height—The distance from a wave crest to the trough.

wave period—The time it takes two waves to pass a given point.

wave train—A long series of waves.

wavelength—The distance between waves.

Bibliography

Auerbach, P. (1987). *A medical guide to hazardous marine life.* Jacksonville, FL: Progressive Printing.

Barnhart, R., & Steinmetz, S. (1986). *Dictionary of science.* Maplewood, NJ: Hammond.

Bascom, W. (1964). *Waves and beaches.* Garden City, NY: Anchor Books.

Bove, A., & Davis, J. (1990). *Diving medicine.* Philadelphia: W.B. Saunders.

Divers Alert Network. (1989). *Medical requirements for scuba divers.* Durham, NC: Author.

Edmonds, C., Lowry, C., & Pennefather, J. (1981). *Diving and subaquatic medicine.* Mosman, NSW, Australia: Diving Medical Centre.

Foley, B. (1989). *Physics made simple.* New York: Doubleday.

Graver, D. (2004). *Aquatic Rescue and Safety.* Champaign, IL: Human Kinetics.

Lee, P., Lidov, M., & Tyberg, T. (1986). *The sourcebook of medical science.* New York: Torstar.

Lehrman, R. (1990). *Physics the easy way.* Hauppauge, NY: Barron's Educational Series.

Lippmann, J. (1992). *The essentials of deeper sport diving.* Locust Valley, NY: Aqua Quest.

Maloney, E. (1983). *Chapman piloting seamanship and small boat handling.* New York: Hearst Marine Books.

McGraw-Hill Professional Publishers (1998). Concise encyclopedia of science. Camden, Maine: McGraw-Hill.

Miller, J. (Ed.). (1979). *NOAA diving manual* (2nd ed.). Washington, DC: U.S. Government Printing Office, U.S. Department of Commerce.

National Association of Underwater Instructors (NAUI). (1991). *Advanced diving technology and techniques.* Montclair, CA: Author.

Professional Association of Diving Instructors (PADI). (1998). *The encyclopedia of recreational diving.* Santa Ana, CA: Author.

Sebel, P., Stoddart, D., Waldhorn, R., Waldmann, C., & Whitfield, P. (1985). *Respiration: The breath of life.* New York: Torstar.

Taylor, E. (Ed.). (1985). *Dorland's illustrated medical dictionary.* Philadelphia: W.B. Saunders.

Whitfield, P., & Stoddart, D. (1984). *Hearing, taste, and smell: Pathways of perception.* New York: Torstar.

This page intentionally left blank.

Index

Note: The italicized *f* and *t* following page numbers refer to figures and tables, respectively.

A

AAS. *See* alternate air source
abrasions 114, 116*t*
absolute pressure 23, 26-27
ABT. *See* actual bottom time
accessories 101-104, 101*f*-102*f*
accidents 180-181, 189
actual bottom time (ABT) 194, 199-202
adjustments. *See* diving adjustments
advance preparation 185-187
AGE (arterial gas embolism) 50, 57
air
  composition 16, 16*f*
  consumption 28, 30-31, 34
  environment 19-23, 19*f*, 20*t*, 21*f*-22*f*, 118
  supply, loss of 49, 174-177, 175*f*-176*f*
alternate air source (AAS) 64, 94-96, 94*f*-95*f*
alternobaric vertigo 52
altitude 200, 202-203, 202*t*
aluminum cylinders 83-85
ambient (surrounding) pressure 53
anatomy 12-15, 13*f*, 15*f*
animals, dangerous 114-117, 114*f*-115*f*, 116*t*, 180
aquatic biology 110-117, 111*f*-115*f*, 116*t*
aquatic conditions 121-132, 123*f*-125*f*, 127*f*, 129*f*-131*f*
Archimedes' principle 21
area orientations 189-190
arterial gas embolism (AGE) 50, 57
ascents
  emergency 175-176, 175*f*
  scuba diving 159-160, 166, 174-176, 174*f*-175*f*, 200-201
  skin diving 146-147, 147*f*
atmospheric pressure 22-23, 22*f*

B

backpack 79, 149, 149*f*
back-roll entry 155, 155*f*
backup scuba units 94-95, 95*f*
barotrauma 50
BC. *See* buoyancy compensator
BEA. *See* buoyant emergency ascent
the bends. *See* decompression illness
bezel 97
biology, aquatic 110-117, 111*f*-115*f*, 116*t*
bites, animal 115, 116*t*
blooms 110
blowout plug 99
boat diving 132-135, 133*f*-134*f*, 153-154
boots 73, 73*f*, 76, 77*f*, 78
bottom dwellers 113
bottom time (BT) 200
bounce dive 198, 198*f*
Bourdon tube 96-97, 96*t*
Boyle's law 23-24, 24*f*, 27, 96
breathing 15-18, 16*f*-17*f*, 19*f*, 28, 47
  adjustments 46-51, 48*f*
  buddy 175-177, 175*f*-176*f*
BT (bottom time) 200
buddy breathing 175-177, 175*f*-176*f*
buddy check 152
buddy line 162
buddy system 152, 162-163, 172, 173*f*, 175-177, 176*f*-177*f*
buoyancy 21-22, 21*f*, 59, 60*f*
  adjustments 41-43, 42*t*
  control 119, 159-160, 173-174
  equipment and 66, 70, 73, 75-76, 78-83, 80*f*-81*f*, 80*t*, 85-86, 86*t*
  neutral 21, 142, 158-159, 161
  testing 142, 158
buoyancy compensator (BC)
  equipment 64, 80-83, 80*f*-81*f*, 80*t*, 138-139, 150-151, 153, 157-160, 174
  inflating and deflating 152-153, 157-160, 174
buoyant emergency ascent (BEA) 175-176
burst disk 88

C

calculators, dive 204-206, 204*t*
California 216
capillary gauge 96-97, 96*t*
cardiorespiratory system 16, 17*f*
CARE. *See* cognizance, assessment, rescue, and evacuation
career opportunities 218
Caribbean 217
carotid sinuses 18, 18*f*
cavern 124, 125*f*
caves 124, 125*f*
C-card 4
CCR. *See* closed-circuit rebreather
ceiling 99
certification 4
charter boat diving 133-134, 133*f*-134*f*, 190
ciguatera 112
circulation 15-18, 17*f*-18*f*
closed-circuit rebreather (CCR) 3, 4*t*, 105-106, 105*f*
cognizance, assessment, rescue, and evacuation (CARE) 178-179
cold 32, 51, 201
collapsed lung 51
communication 178
community, diving 3-5
compartments 54-55
compass course 164-165
compasses, dive 97-98, 98*f*, 164-165, 164*f*-165*f*
compass heading 164
computers, dive 99, 204-206, 204*t*
congested sinuses 44, 44*f*
conservation 117-121, 119*f*
consoles 98*f*, 99-100
continental shelf 126
contingency plans 203-204, 203*f*
continuing education 210-212, 219
controlling compartment 55
coughing 172
cramps 172, 173*f*
crest 128
currents 124, 124*f*, 130-132, 130*f*-131*f*, 187, 188*f*
cuts 114, 116*t*
cylinders 83-87, 83*f*-85*f*, 86*t*, 149-150, 149*f*-150*f*
cylinder valves 87-89, *88f*

D

Dalton's law 25-26, 26*t*
dangerous aquatic animals 114-117, 114*f*-115*f*, 116*t*, 180
decompression illness (DCI) 7, 55-57, 56*f*-57*f*, 190, 200
decompression stops 55, 57*f*, 190, 196, 196*t*, 200-201
decompression theory 53-55, 54*f*
defogging 29, 66, 139
dehydration 29, 58-59
density 19*f*, 20, 20*t*, 23-24, 24*f*
depth 200
depth gauges 96-97, 96*t*, 100
descents 146-147, 161
destinations 216-218
diaphragm gauge 96-97, 96*t*
DIN valve 88, 88*f*

dive calculators 204-206, 204*t*
dive compasses 97-98, 98*f*, 164-165, 164*f*-165*f*
dive computers 99, 204-206, 204*t*
dive courses 8, 8*f*
dive flags 101-102, 101*f*, 160, 160*f*
dive floats 102
dive kits 102-104
dive knives 100-101, 100*f*, 152
dive lights 102, 102*f*
dive planning
  advance preparation 185-187
  area orientations 189-190
  calculators for 204-206, 204*t*
  computers for 99, 204-206, 204*t*
  contingency 203-204, 203*f*
  dive profile 188, 190-191, 190*f*, 197-204, 198*f*-199*f*, 202*t*, 203*f*
  dive tables for 191-197, 192*t*-193*t*, 195*t*-196*t*, 204, 204*f*
  factors 184-185
  on-site 187-188, 188*f*
  postdive review 206
  short-term 187
  special-equipment 197
  special procedures 201-203, 202*t*
dive profile 188, 190-191, 190*f*, 197-206, 198*f*-199*f*, 202*t*, 203*f*
diver impact 118-119, 119*f*
diver's push-ups 158, 159*f*
dive slates 102-103
dive tables 191-197, 192*t*-193*t*, 195*t*-196*t*, 204, 204*f*
diving. *See also* equipment; scuba diving; skin diving
  altitude and 200, 202-203, 202*t*
  boat 132-135, 133*f*-134*f*, 153-154
  buddy 162-163, 172, 173*f*, 175-177, 176*f*-177*f*
  community 3-5
  conditions 121-132, 123*f*-125*f*, 127*f*, 129*f*-131*f*
  drift 132
  evolution of 3
  guides 4
  joys of 2, 9
  periodicals 222
  prerequisites 5-6
  risks 6-8
  surface 145, 146*f*
  training for 5-8, 8*f*, 221
  travel 213-218
diving adjustments
  breathing 46-51, 48*f*
  buoyancy 41-43, 42*t*
  dehydration prevention 58-59
  equilibrium 52-53
  heart and lungs 50-52
  importance of 38, 61
  ingassing and outgassing 53-58, 54*f*, 56*f*-57*f*
  mobility 59, 60*f*
  thermal 38-41, 39*f*, 41*f*
  visual 53
diving environment
  aquatic biology 110-117, 111*f*-115*f*, 116*t*
  aquatic conditions 121-132, 123*f*-125*f*, 127*f*, 129*f*-131*f*
  for boat diving 132-135, 133*f*-134*f*
  conservation and preservation 117-121, 119*f*
diving opportunities
  career 218
  continuing education 210-212, 219
  leadership 219
  local 212-213
diving skills
  automatic execution of 139, 191
  for emergencies 178-181, 179*f*
  physical problem management 173-177, 174*f*-177*f*
  physiological problem management 169-172, 172*f*-173*f*
  scuba diving 152-169, 153*f*-155*f*, 159*f*-160*f*, 163*f*-165*f*, 167*f*-169*f*
  skin diving 141-148, 141*f*, 143*f*-147*f*
dizziness 170
dolphin kick 145, 145*f*
downwelling 130
drag 20-21, 20*t*, 59, 60*f*
drift 132
drift dive 132
drifters 110-111, 111*f*-112*f*
dry suits 75-77, 75*t*, 76*f*-77*f*

E

EANx. *See* enriched-air nitrox
eardrum 14, 15*f*, 52
ears 14-15, 15*f*, 44-46, 45*f*, 52
ebb 127
eddies 124
education, continuing 210-212, 219
electronic gauge 96-97, 96*t*
embolism 50
emergencies 178-181, 179*f*, 188
emergency decompression 201, 203*t*
emergency swimming ascent (ESA) 175-176, 175*f*
emphysema 50-51
enriched-air nitrox (EANx) 104-105, 104*f*
entanglement 173
entry techniques 152-155, 153*f*-155*f*
environment. *See also* diving environment
  air 19-23, 19*f*, 20*t*, 21*f*-22*f*, 118
  water 19-23, 19*f*, 20*t*, 21*f*-22*f*, 28-29, 29*f*, 32-34, 32*f*
environmental organizations 121, 221
equalization 7, 14, 24, 25*f*, 43-46, 44*f*-45*f*
equilibrium adjustments 52-53
equipment. *See also* exposure suits; fins; mask
  AAS 64, 94-96, 94*f*-95*f*
  accessories 101-104, 101*f*-102*f*
  assembling 149-151, 149*f*-150*f*
  BC 64, 80-83, 80*f*-81*f*, 80*t*, 138-139, 150-151, 153, 157-160, 174
  buoyancy and 66, 70, 73, 75-76, 78-83, 80*f*-81*f*, 80*t*, 85-86, 86*t*
  checklist 82
  cylinder valves 87-89, 88*f*
  disassembling 169
  dive knives 100-101, 100*f*
  diver impact and 119
  donning 139-141, 140*f*, 151-152
  handling 147-148, 166
  inspecting 139-140, 151-152
  instrumentation 96-100, 96*f*, 98*f*, 152
  introduction to 64, 65*f*
  nitrox 104-105, 104*f*, 197
  packing 148-149, 149*f*
  preparing 138-141, 140*f*
  removing 148
  scuba cylinders 83-87, 83*f*-85*f*, 86*t*, 149-150, 149*f*-150*f*
  scuba regulators 89-94, 89*f*-92*f*, 149-152, 150*f*, 156-157
  skin diving 70-71, 70*f*, 138-141, 140*f*, 147-148
  snorkel 67-68, 67*f*-68*f*, 138, 140, 143-144, 143*f*, 152, 155-157
  specialized 104-106, 104*f*-105*f*, 197
  surface snorkeling with 155-156
  for technical diving 105-106, 105*f*
  training 8, 8*f*
  weighting systems 78-80, 78*f*, 138-140, 147-148, 152
  weight of 59, 60*f*
ESA. *See* emergency swimming ascent
eustachian tubes 14, 15*f*, 44-46, 45*f*
exit techniques 168-169, 168*f*-169*f*
exposure suits
  dry suits 75-77, 75*t*, 76*f*-77*f*
  preparing 138-139, 152
  wet suits 71-75, 71*t*, 72*f*-74*f*, 77-78
extra second stages 94-95, 94*f*

F

Farmer Johns 72*f*, 73
feet-together entry 155
fetch 128
fins 69-70, 69*f*-70*f*, 152
  donning 138, 140, 140*f*
  removing 148
  using 144-145, 144*f*-145*f*, 147
first aid 103-104, 116*t*, 178-180
flags 101-102, 101*f*, 160, 160*f*
flaring 174, 174*f*, 176
floats 102

flood 127
Florida 217
flutter kick 144, 144*f*
flying 200, 202, 202*t*, 215
food cycle, underwater 110-111, 111*f*
freshwater diving conditions 122-126, 123*f*-125*f*
fronds 111

G
gases 16, 16*f*, 18, 19*f*, 26-27, 26*t*
gas laws 23-28, 24*f*-25*f*, 26*t*, 96
Gay-Lussac's law 24-25, 27
gear bags 101
giant-stride entry 154, 154*f*
Great Lakes 217
guides 4
gyres 130

H
habits 191
half-time 54, 99
hand coverings 73-74, 74*f*, 76, 77*f*, 78
hand signals
  scuba diving 166, 167*f*
  skin diving 141, 141*f*
Hawaii 217
head-to-toe buddy check 152
hearing 14, 15*f*, 34
heart 50-52
heat
  exhaustion 40, 41*f*
  loss 28-29, 29*f*, 39-40, 39*f*
heatstroke 40, 41*f*
Henry's law 25-26, 26*t*
high altitude diving 200, 202, 202*t*
hoods 74, 74*f*, 76, 78, 139
humidity 29
hydroids 113-114, 113*f*, 116
hyperthermia 38
hyperventilation 18, 19*f*
hypothermia 38, 40
hypoventilation 18

I
ingassing 26, 53-58, 54*f*, 56*f*-57*f*
injuries 50, 114-117, 116*t*, 179-181
instrumentation
  equipment 96-100, 96*f*, 98*f*, 152
  monitoring 165-166

J
job opportunities 218
jumpsuit 72*f*, 73
J-valve 88

K
kelp 110-111, 112*f*, 216
kicks 144-145, 144*f*-145*f*, 164
kits 102-104
knives 100-101, 100*f*, 152
K-valve 87, 88*f*

L
lacerations 114-115, 116*t*
laws 4
leadership opportunities 219
lifesaving scuba 178-179, 179*f*
lift 59, 60*f*
light 32-33, 32*f*
lights, dive 102, 102*f*
local diving opportunities 212-213
lubber line 98
lungs 12, 13*f*, 47, 48*f*, 50-52, 176

M
magazines 222
mask 46, 65-67, 65*f*-66*f*, 138-140, 147-148, 152
  clearing 142-143, 143*f*, 162
  defogging 29, 66, 139
maximum-depth indicator 97
mediastinal emphysema 50
menstruation 6
middle ear 14, 15*f*, 52
Middle East 218
mobility adjustments 59, 60*f*
modified frog kick 144, 145*f*
multilevel dive profile 197, 198*f*

N
navigation 97-98, 98*f*, 164-165, 164*f*-165*f*
NDSL. *See* no decompression stop limit
neap tides 126
neutral buoyancy 21, 142, 158-159, 161
New England 217
nitrogen narcosis 16, 58
nitrox equipment 104-105, 104*f*, 197
no decompression stop limit (NDSL) 191, 194-197, 195*t*, 199-200, 202
nonreference descents 161

O
octopus 94
on-site planning and preparation 187-188, 188*f*
open-circuit scuba 3, 4*t*
open-valve ascent 159-160, 166
organizations 121, 221
O-ring 88-89, 150-151
outgassing 26, 28, 53-58, 54*f*, 56*f*-57*f*, 166
overexertion 47-48, 172
overheating 40-41, 41*f*
overturn 122, 123*f*
overweighting 163, 163*f*

P
packing 148-149, 149*f*
panic 170-172, 172*f*
partial pressure 26-27, 26*t*
periodicals 222
physical problems, managing 173-177, 174*f*-177*f*
physiological problems, managing 169-172, 172*f*-173*f*
plankton 110, 111*f*
planning. *See* dive planning
pneumothorax 51
pollution 117-118
pony tank 95, 95*f*
ports 91
postdive review 206
preparation
  scuba diving 148-152, 149*f*-150*f*, 185-188, 188*f*
  skin diving 138-141, 140*f*
prerequisites, diving 5-6
preservation 117-121, 119*f*
pressure
  absolute 23, 26-27
  ambient 53
  atmospheric 22-23, 22*f*
  equalization 7, 14, 24, 25*f*, 43-46, 44*f*-45*f*
  gas laws and 23-24, 24*f*
  injury from 50
  measurement 22-23, 22*f*
  partial 26-27, 26*t*
  reduction, stages of 89-95, 90*f*-92*f*, 94*f*
proficiency testing 5
Puget Sound 216
pulmonary barotrauma 50
punctures 114-115, 116*t*

R
Rankine 25
reciprocal compass course 165
recompression chamber 56, 56*f*
recreational diving 3-4
Red Sea 218
red tide 110
reference descents 161
regulators, scuba 89-94, 89*f*-92*f*, 149-152, 150*f*, 156-157
repetitive dive 190-191, 190*f*, 194, 200
repetitive group designation (RGD or RG) 191, 194, 200
reserve valve 88
residual nitrogen 190, 190*f*, 194-197, 195*t*
residual nitrogen time (RNT) 194-202, 195*t*
respiration 15-18, 16*f*-17*f*, 19*f*
responsibilities 9
rest stops 55, 57*f*, 190, 196, 196*t*, 200-201
reverse blocks 24, 25*f*
reverse thermocline 122, 123*f*
RG. *See* repetitive group designation
RGD. *See* repetitive group designation
rip current 131, 131*f*

risks 6-8
RNT. *See* residual nitrogen time

S

safety stops 55, 57*f*, 190, 196, 196*t*, 200-201
saltwater diving conditions 126-132, 127*f*, 129*f*-131*f*
save-a-dive kit 102-103
sawtooth dive profile 197, 198*f*
scissors kick 144, 144*f*
scombroid 112
scrolling 204-205
scuba (self-contained underwater breathing apparatus)
  backup 94-95, 95*f*
  cylinders 83-87, 83*f*-85*f*, 86*t*, 149-150, 149*f*-150*f*
  lifesaving 178-179, 179*f*
  regulators 89-94, 89*f*-92*f*, 149-152, 150*f*, 156-157
  systems 3, 4*t*
scuba diving
  ascents 159-160, 166, 174-176, 174*f*-175*f*, 200-201
  descents 161
  entry techniques 152-155, 153*f*-155*f*
  exit techniques 168-169, 168*f*-169*f*
  hand signals 166, 167*f*
  at high altitude 200, 202, 202*t*
  laws 4
  navigation 97-98, 98*f*, 164-165, 164*f*-165*f*
  organizations 221
  preparation for 148-152, 149*f*-150*f*, 185-188, 188*f*
  prerequisites 5-6
  recreational 3-4
  responsibilities 9
  rules 156, 200
  skills 152-169, 153*f*-155*f*, 159*f*-160*f*, 163*f*-165*f*, 167*f*-169*f*
  technical 105-106, 105*f*
  underwater swimming in 163-164, 163*f*
  wreck diving 217
sculling 119, 119*f*
seasickness 52-53, 170
seated entry 154, 154*f*
self-contained underwater breathing apparatus. *See* scuba
semi-closed-circuit scuba 3, 4*t*
set, of current 132
shallow-water blackout 18
short-term planning and preparation 187
shorty 72*f*, 73
signaling devices 102-103
signals. *See* hand signals
silting 163, 163*f*
simple valve 87-88, 88*f*
sink 125
sinuses 12, 13*f*, 43-44, 44*f*
siphon 125
SIT. *See* surface interval time
skills. *See* diving skills
skin diving
  ascents 146-147, 147*f*
  equipment 70-71, 70*f*, 138-141, 140*f*, 147-148
  hand signals 141, 141*f*
  preparation 138-141, 140*f*
  prerequisites 5-6
  recreational 4
  skills 141-148, 141*f*, 143*f*-147*f*
skip breathing 47
slack water 127
slates 102-103
small-boat diving 134-135
snorkel 67-68, 67*f*-68*f*, 138, 140, 143-144, 143*f*, 152, 155-157
snorkeling 155-156
snorkel–regulator exchanges 157
sound 34
South Pacific 217
Spare Air unit 95, 95*f*
specialized equipment 104-106, 104*f*-105*f*, 197
SPGs. *See* submersible pressure gauges
spike dive 198, 198*f*
spring tides 126, 127*f*
square compass course 164-165
squatting back-roll entry 155, 155*f*
squeezes 24, 25*f*
stages, of pressure reduction 89-95, 90*f*-92*f*, 94*f*
stand 126
steel cylinders 83-85
step dive 197, 198*f*
stings, animal 115-116, 116*t*
stipes 110
stress 170-172, 172*f*
subcutaneous emphysema 51
submersible pressure gauges (SPGs) 88, 98-100, 98*f*, 150, 152, 165
surf 128-130, 129*f*, 153, 153*f*, 169, 169*f*
surface dives 145, 146*f*
surface interval time (SIT) 194-196, 195*t*, 200
surface snorkeling 155-156
surge 128
surrounding pressure. *See* ambient pressure
swells 128
swimmers 112-113, 112*f*
swimming
  ESA 175-176, 175*f*
  kicks 144-145, 144*f*-145*f*, 164
  surface dives 145, 146*f*
  underwater 146-147, 147*f*, 163-164, 163*f*

T

tanks. *See* cylinders
task-relevant instructional self-talk (TRIST) 171
TBT. *See* total bottom time
technical diving 105-106, 105*f*
teeth 15
temperature 20, 23-27, 32, 34. *See also* thermal adjustments
tetrodotoxin 112
thermal adjustments 38-41, 39*f*, 41*f*
thermocline 122, 123*f*
throat 12, 13*f*
thrust 59, 60*f*
tides 126-128, 127*f*
total bottom time (TBT) 194-202
Toynbee maneuver 44
trail line 133, 134*f*
training 5-8, 8*f*, 210-212, 221
trapdoor effect 45, 45*f*
travel 213-218
TRIST (task-relevant instructional self-talk) 171
trough 128
tsunami 126
turbid water 33

U

underwater food cycle 110-111, 111*f*
underwater swimming 146-147, 147*f*, 163-164, 163*f*
underwater timers 97
upwelling 123

V

Valsalva maneuver 44
valves, cylinder 87-89, 88*f*
valve seat 87
vasoconstriction 39
venomous injuries 114-115, 116*t*, 180
vertigo 33, 52
vests 74, 74*f*, 138-139, 142, 148
visibility 33, 162
vision 32-33, 32*f*, 67
visual adjustments 53
volume 23-24, 24*f*
vomit 52, 133, 170

W

wading entry 153, 153*f*
Washington 216
water
  environment 19-23, 19*f*, 20*t*, 21*f*-22*f*, 28-29, 29*f*, 32-34, 32*f*
  slack 127
wave height 128
wavelength 128
wave period 128
waves 128-130, 129*f*
wave train 128
weighting systems 78-80, 78*f*, 138-140, 147-148, 152
wet suits 71-75, 71*t*, 72*f*-74*f*, 77-78, 139, 148
wreck diving 217

About the Author

Dennis Graver has more than 30 years of experience as a scuba instructor and instructor trainer. He has authored 27 books and manuals on scuba diving, including three previous editions of *Scuba Diving* from Human Kinetics. In his position as director of training for the Professional Association of Diving Instructors (PADI), he designed the PADI modular scuba course and wrote the *PADI Dive Manual*, which revolutionized scuba instruction. During his tenure as director of education for the National Association of Underwater Instructors (NAUI), Graver wrote several diving texts, including the *NAUI Openwater I Scuba Diver Course Instructor Guide*. He has also contributed hundreds of articles on diving to such magazines as *Skin Diver*, *Sources*, and *Undercurrents* as well as several NAUI technical publications.

Since 1977 Graver has been photographing many wonders under the water, from the Red Sea off the coast of Egypt to the barrier reefs of Australia. He has won numerous awards from the Underwater Photographic Society, and his photos have graced the covers of numerous magazines and illustrated several diving texts and audiovisual educational programs.

Graver and his wife, Barbara, live in Camano Island, Washington.

You'll find other outstanding
outdoor and adventure resources at

www.HumanKinetics.com/outdoorandadventure

In the U.S. call 1-800-747-4457

Australia 08 8372 0999 • Canada 1-800-465-7301
Europe +44 (0) 113 255 5665 • New Zealand 0064 9 448 1207

HUMAN KINETICS
The Premier Publisher for Sports & Fitness
P.O. Box 5076 • Champaign, IL 61825-5076 USA

eBook
available at
HumanKinetics.com

Equipment

Introduction

The wide array of equipment available is often bewildering to a novice diver. Some items are essential in all circumstances, while others are used only in specific diving environments. The more research you do and advice you seek, the more likely you are to buy safe, suitable gear.

BUYING AND RENTING GEAR

You could begin dive training with as little as a swimsuit, but it may be advisable to at least obtain a good-quality mask, snorkel, and fins. This means that you always have three key pieces of gear that fit perfectly, and with which you are comfortable.

Many divers own no more than this basic gear, preferring to rent the rest. There are pros and cons to this approach: it frees you from carrying heavy gear around, but it can also mean using ill-fitting equipment that has been worn by many people, and which may not have been well cared for. The alternative is to buy and transport all your own gear. This is recommended if you are likely to be diving regularly in your local area (although a local club or shop may have gear you can rent). It is not so practical if you plan to dive any farther afield—taking your own gear overseas can be expensive, although most dive tour operators are fairly generous with their excess baggage allowances. You can rent a tank and weights from any dive center, so you need never travel with these items.

The best places to buy equipment and obtain advice are dive shows or local dive shops. Second-hand gear is worth investigating, but inspect it for signs of wear and corrosion before making a purchase, and make sure that it is serviced before use.

DIVERS TOURING A REEF
Knowledge and care of gear is vital, since your equipment is your life-support system for exploring the underwater world.

CORE EQUIPMENT

There are several items of gear that are essential, in addition to your mask, fins, and snorkel. A buoyancy compensator (BC) is an inflatable jacket that keeps you afloat at the surface, helps you adjust your buoyancy underwater, and supports other vital pieces of equipment. Weights, and a belt to mount them on, are another necessity, although weights can also be carried on some BCs (*see p.59*). You will also need some kind of dive suit (*see pp.50–55*) depending on the water temperature—and boots. A regulator (the device that supplies air to the diver) is also necessary, preferably one that incorporates an alternative air source (an octopus) and an instrument console with a compass. Dive computers are a useful aid to safe diving, and you should also carry a knife to cut yourself free if you become snagged underwater.

DIVING ACCESSORIES

There is a range of equipment and accessory items designed for specialized diving. Try to borrow or rent specialty gear that you may only use infrequently.

Slates are useful accessories, since they allow you to write messages to your buddy, or to make notes from your dive—for example, on species that you do not recognize—or to sketch a site.

Flashlights are essential for diving at night, in murky water, and in low-light environments, such as caverns and wrecks.

If you will be doing a lot of open-water diving, especially in currents, you should purchase a marker buoy, either a surface marker buoy (SMB) or an inflatable surface marker buoy. These indicate your position to observers on the surface, and also aid your ascent.

MASK
The glass lenses should be tempered and offer a good field of view.

EQUIPMENT FOR COLD WATER
Cold seas make extra demands on diving equipment, and additional peripheral items may be needed to ensure safe diving. Thermal protection is a priority.

GAUGES
Gauges with large, clear dials are easier to read in low-light environments.

FLASHLIGHT
Temperate waters tend to be murky, so a flashlight may be required.

BUOY AND REEL
A marker buoy and reel are useful for guiding an ascent to the surface.

DEMANDING ENVIRONMENT
Diving in cold water requires extra equipment, and appropriate training in how to use it. Safety devices are especially important in such waters.

Lastly, you may wish to record your dive experiences with a camera. Digital cameras in housings are now small enough to fit into pockets in your kit.

DIVING IN COLD WATER

If you plan to dive frequently in cold water, you might want to invest in a drysuit (*see pp.54–55*). A good thermal underlayer is essential with a drysuit, since loss of body heat can result in poor decisions and lead to exhaustion, speeding the onset of hypothermia. Dive clothing manufacturers make special thermal undersuits with wicking properties to draw moisture away from the skin. If you are planning on doing any extreme cold water diving or ice diving (*see pp.188–89*) you will need to use an environmentally sealed regulator, because these are more resistant to freezing.

Also necessary in cold water are a hood and gloves, the thickness of which will depend on the water temperature. The thicker the neoprene, the more insulation it offers, but the more it will restrict movement. Thick gloves or mittens reduce dexterity considerably, which is worth bearing in mind if you intend to use smaller items of gear.

MAKING THE RIGHT CHOICE

Don't rush into buying the first bit of new gear you see—always look at a range of products and compare prices. Take advice from fellow divers and shop staff, but remember that everyone has their favorite brands. Reviews in magazines and on the Internet can give you more information, so that you can weigh the options before you buy.

Masks

A well-fitting mask is arguably your most important piece of diving gear, so take time to choose one that is right for you. There are many styles and designs to choose from, but above all you need one that fits correctly, and suits your personal diving requirements.

SEEING UNDERWATER

Human eyes cannot focus in water, which is why you have to wear a mask when diving. The way masks work is to trap a pocket of air in front of the eyes, allowing them to focus normally. A standard mask consists of a single or double panel of tempered glass inside a frame, with a silicone skirt forming an airtight seal against the face. A strap and buckles hold it all in place.

If you take care of your mask and keep it in a sturdy case to avoid damage, it should last for many years. The lenses of new masks have a protective coating from the manufacturers that should be removed before use. This can be wiped off using a nonabrasive cloth and non-gel toothpaste (or a fine abrasive cleaner) to clean the coating off without scratching the glass. If in doubt, consult the manufacturer's instructions. Make sure you rinse your mask in fresh water after use and store it away from direct sunlight in a dry place.

TESTING FOR FIT

There are lots of shapes and sizes of mask, so take time to find one you like that fits. A simple test is to hold it up to your face and inhale gently through your nose. If you then lean forward, it should remain in place.

1 **Hold the mask** in front of your face with the strap hanging loose.

2 **Place it on** your face. Inhale through your nose—the mask should stay on without support.

MASK DESIGNS

CLASSIC This is a slightly old-fashioned design, but it is still popular with photographers because it gives a clear view of the diver's face. It offers a good field of view, but its high internal volume of air makes clearing the mask a little harder.

EXTENDED VISION Many new designs of mask are constructed to give the diver an extended field of view, especially downward. This model is a multi-lens design, which also allows corrective lenses of different strengths to be fitted.

MATTE FINISH Some divers prefer to use a mask with a black silicone skirt to avoid confusing reflections on the inner surface of the glass. This is especially useful to marine photographers. However, this feature does reduce peripheral vision.

SIDE VIEW Some designs offer improved peripheral vision through glass panels on either side of the main lens, but are still low-volume, so are easy to clear. When choosing, make sure that additional lenses do not distort perspective.

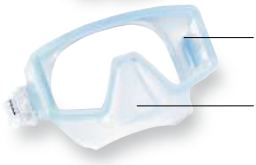

FULL FACE Mostly used in commercial diving, full-face masks are advantageous in very cold or polluted water. Some recreational models have intercom radio units so divers can communicate.

EYE TO EYE
Evolution has equipped fish with eyes that focus underwater, but divers must rely on masks for clear vision.

CORRECTIVE LENSES

Some masks can be fitted with prescription lenses, so if you wear glasses you will still be able to see underwater. Contact lens wearers can use normal masks—they provide an adequate degree of protection from water, unless accidentally removed.

Snorkels and fins

Along with a mask, a snorkel and a pair of fins are probably the first pieces of gear you will buy. Snorkels vary very little, but there are plenty of styles of fins to choose from, depending on your budget. As always, look for a good fit and features that suit your diving needs.

CHOOSING A SNORKEL

A snorkel is simply a tube that allows you to breathe easily at the surface without needing to raise your head from the water. When diving, you do not need a snorkel underwater; however, you may use one before and after submerging. A good snorkel should be easy to breathe through, and should not be too long—ideally no more than 17 in (43 cm)—so it is easy to clear. Attach your snorkel to the left-hand side of your mask, leaving the right-hand side free for your regulator hose (*see pp.60–61*). Always rinse your snorkel after use and store out of direct sunlight.

DESIGN FEATURES

Basic snorkels are little more than a curved tube, but some models have a splash-guard at the top of the tube to prevent ingress of water, and a drainage valve at the lower end. Collapsible types are easy to stow away when not in use.

FITTING FINS

Sometimes incorrectly referred to as flippers, fins are used to aid propulsion when diving. A standard fin consists of a shoe made of rubber and a blade made of a stiffer material, often with ridges and vents to enhance performance. The shoe can either fit the whole foot like a slipper or leave the heel exposed. The latter is held on with a strap, and is designed to be worn over neoprene boots. When buying fins of this type, it is advisable to take your boots along with you, since their thickness will need to be taken into account when testing for fit.

SELECTING THE RIGHT PAIR

Advances in fin technology mean that there is a wide range for the diver to choose from, depending on their specific preferences. For example, split fins offer a high degree of agility underwater, but many divers believe traditional models deliver more outright thrust. Before deciding which pair to buy, consider the kind of diving you will be doing, and seek advice from store staff or other divers.

FIN POWER
A diver employing the "frog kick" with conventional fins.

TYPES OF FINS

CONVENTIONAL A basic fin with a single, fairly rigid blade, the conventional design is a good starter fin, but you may want to try other types to see if they offer features that suit you better.

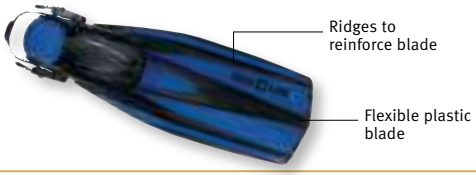

FORCE FINS Made exclusively by a California manufacturer, this range of fins is designed to minimize the water turbulence produced, and reduce the diver's kicking effort.

PIVOTING A design that has been adopted by two manufacturers, this fin features a flex-point in its structure and a V-shaped area of softer plastic to scoop water for better traction and speed.

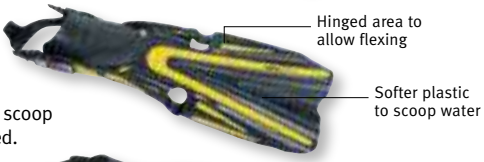

SPLIT A popular style, split fins offer better agility than standard fins. However, some divers believe they are less effective than conventional fins in strong currents.

SNORKELING Fins designed for snorkeling are lighter than scuba fins, and of a foot-pocket design, so they can be used with bare feet. Some also have shortened blades.

Foot-pocket fin Heel-strap fin

FIT AND FASTENING
Fins are held on the foot in two basic ways: in a form-fitting foot pocket, or by a strap over the heel.

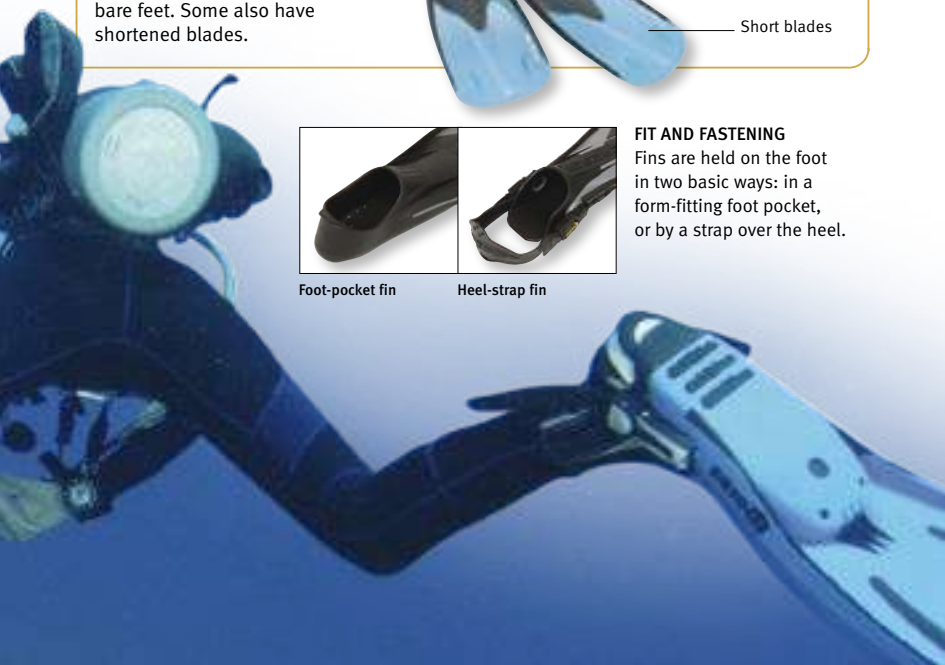

Diving suits

In all but the very warmest waters, you will need some kind of thermal protection during your dive. Good diving suits (which are technically known as "exposure suits") are available in every price range, but always look for comfort, fit, and quality of construction.

BEATING THE CHILL

Water conducts heat away from the body much more efficiently than air. Unless the water you are in is the same temperature as the surface of your skin—about 90°F (32°C), which is hotter than most corals can withstand—your body will lose heat, and you will get cold. This means that in almost all diving environments, thermal protection is required. If you are diving in warm waters, a normal wetsuit will be sufficient. Only in the warmest waters should you consider a non-insulating or "skin" suit. Never be tempted to dive in just a swimming suit; even if you don't need insulation, a thin layer will prevent stings and abrasions. If you are diving in cooler waters, you should consider the heavyweight options: a semi-dry suit or a drysuit.

SECOND SKIN
Modern wetsuits are made of materials that offer highly efficient heat insulation.

WHICH SUIT?

Personal warmth requirements vary. If possible, take various suit options—you can layer these to give you the best protection. It is rare to be too hot underwater, but in tropical climates you may overheat in your suit before you enter the water. This can be prevented by wetting the suit before putting it on.

Buying a suit for every temperature range is too expensive for most divers, but most suits will suffice for a variety of waters. Although in the tropics you would be too warm in a drysuit, you could opt for a full-length wetsuit, as its extra warmth would be welcome in the cooler depths.

In addition to a suit, you may need accessories such as gloves, boots, a hood, and talcum powder, which is useful for easing entry into suits.

EXPOSURE PROTECTION

There are guidelines for the type and thickness of suit needed in different conditions. For the tropics, where the waters rarely drop below 68°F (20°C), a 2–3-mm wetsuit is fine. In the Mediterranean in summer, a 5–6-mm full wetsuit is more appropriate, rising to 7–8 mm during winter months. A semi-dry suit of 10–15-mm thickness, with a hood and thick gloves, will cope with most conditions in colder temperate waters, but in waters below 50°F (10°C), a drysuit is necessary.

SUIT LAYERS

Wetsuits and drysuits keep the diver warm in different ways. In wetsuits, a thin film of water is trapped next to the diver's skin and acts as an insulator. Drysuits seal a layer of air inside the suit, and this permits the use of thermal undergarments for extra insulation.

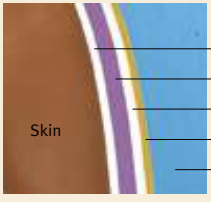

SKINSUITS AND SHORTIES

In warm, tropical waters, you will probably not require any thermal insulation, but a thin layer over your body will prevent your equipment from rubbing on your skin, as well as giving protection from the Sun before and after dives and providing a barrier to stings and scrapes.

The thinnest skinsuits are made of Lycra, but some manufacturers offer suits made from 0.5-mm neoprene. Lycra suits can also be used as a base layer under a heavier suit, so they are a good add-on option if you are not sure how warm or cold your destination will be.

So-called "shortie" wetsuits offer slightly more thermal protection than skinsuits, typically being constructed of 3-mm neoprene. However, they do leave the arms and legs exposed and unprotected from stinging plankton and jellyfish or abrasions on rocks or coral.

SHORTIE WETSUIT
This type of wetsuit is a good choice in warm waters, and does not cost as much as a full suit.

Light gloves

Skinsuit vest

TROPICAL DESCENT
Even in warm waters, a full-length wetsuit may prove invaluable, because water temperatures drop as the diver ventures deeper.

FULL-LENGTH WETSUITS

The next level of thermal protection after shorties and skinsuits is a full-length wetsuit. These offer increased thermal protection for the arms and legs, and are typically made from 2–8-mm neoprene. But thickness is not the only factor in thermal efficiency—a close-fitting suit provides better insulation. A correctly fitting wetsuit should not only be easy to put on, but should follow the contours of your body as closely as possible. Wetsuits rely on trapping a thin film of water next to the body, which gradually becomes warm. If the suit fits badly, water flows more freely between the body and the suit, and chills—rather than insulates—the diver.

CARING FOR YOUR WETSUIT

A simple maintenance regimen is vital to prolong the life of your wetsuit. After every dive, wash your suit to remove salt and debris, which can cause it to rot. It is best to use special wetsuit shampoo, which helps to maintain the suppleness of the neoprene and prevent mildew. Once the suit is fully dry, place it on a hanger and store it in a cool, dry place.

FULL-LENGTH WETSUIT
Scuba wetsuits are available in a range of styles, categorized by the thickness of neoprene they are made from.

SEMI-DRY SUITS

Though intended for colder temperate waters, where they are worn with a hood and gloves, semi-dry suits are a versatile style that can be used even in warm Mediterranean seas. Like standard wetsuits, semi-dry suits are made of neoprene and work by trapping an insulating layer of water between the body and the suit. However, they have more efficient seals at the wrists and ankles and around the neck or hood, which stops water from flushing through the suit and allows the trapped water to warm up. They also tend to be made of thicker neoprene and are often designed to be used in layers (with a jacket fitting over a coverall-style base layer).

The thickness of the neoprene in some semi-dry suits has a tendency to restrict movement on land (although not in water). However, suits made of super-flexible material are available that offer an equivalent degree of thermal protection, but are much easier to put on and move around in. Although a semi-dry suit offers protection similar to a drysuit (*see p.54*), some divers prefer a drysuit, because a damp body can be uncomfortable after the dive, especially in cold weather.

SEMI-DRY SUIT WITH EQUIPMENT
Good-quality semi-dry suits are suitable for a range of dive environments and conditions, and allow you to withstand the chill of greater depths.

DIVE COMPUTER
Often no bulkier than a wrist watch, a dive computer may need a strap extension when worn with thicker suits such as a semi-dry suit.

FLASHLIGHT
In the colder and deeper waters accessible with a semi-dry suit, visibility can be impaired, so a light should always be carried.

DIVING IN DRYSUITS

Drysuits are the most expensive type of diving suit but offer excellent thermal protection. They use trapped air, not water, as an insulating barrier. Drysuits are available off the rack, but it is worth spending a little extra to get a made-to-measure garment, because a well-fitting suit will be more comfortable and more likely to avoid buoyancy problems caused by unequal air distribution.

Drysuits are made from a variety of materials, including crushed neoprene and membrane materials. Crushed neoprene suits are tough and offer a good degree of thermal insulation. Membrane drysuits are made from strong, thin, waterproof fabrics, which offer less inherent thermal insulation than crushed neoprene, and are designed to be used with a thermal undersuit. They are more supple than neoprene types, and so easier to move around in. However, they can be looser-fitting, and allow excess air to move around inside, which can shift buoyancy to the legs, and tip the diver upside down.

KEEPING WARM
Even in conditions as extreme as those encountered diving under ice, drysuits provide good thermal protection, ensuring the diver stays comfortable.

INFLATOR VALVE
This is connected to the regulator first stage (*see p.61*) and allows air to be introduced into the suit.

CRUSHED NEOPRENE SUIT
The inherent buoyancy of a suit made of neoprene that has been "crushed"—treated with heat and pressure to flatten its internal air cells—remains the same at any depth.

USING A DRYSUIT

Drysuits are more challenging to use than wetsuits, as they have variable buoyancy, due to their air-inflation mechanism. If you plan to use one for the first time, take an orientation course in a pool before diving in open water. A drysuit should be serviced once a year (or after long periods of storage) by an approved repair technician.

FASTENINGS AND SEALS

Drysuits generally have an "across the shoulder" zipper to allow entry. You will need help from your buddy to do this up. These zippers need regular coating with wax or "zip slip" to keep them lubricated and protected from corrosion.

To prevent water from entering the suit, drysuits have watertight seals—made of thin neoprene or latex—at the wrist and neck. Latex seals are generally less likely to leak than neoprene versions, but do not offer any insulation and are more fragile, so there is the risk that they may tear. Neoprene seals, by contrast, are tough, easy to repair, and warm, but it can be difficult to get a good fit.

INFLATING WITH AIR

Drysuits have an inflator valve on the chest that injects air to reduce the "shrink wrap" sensation caused by water pressure squeezing the suit against the body. This air expands on ascent and must be dumped via a valve, normally located on the shoulder, which can be manual or automatic. Cuff dumps are also available that work by raising your arm to release excess air.

TOP COVER
A neoprene hood is necessary in cold waters to insulate the head.

THERMAL UNDERSUIT
Specialty garments, undersuits can be worn under drysuits for extra insulation and comfort.

Heavy wetsuit gloves

MEMBRANE SUIT
The tough, lightweight fabric of a membrane suit is more flexible than neoprene and so allows greater ease of movement.

Buoyancy compensators

Unlike fish, humans have no natural way to adjust their buoyancy, or floating ability, when underwater. Divers wear a buoyancy compensator (BC)—a jacket that can be filled and emptied of air—to control their buoyancy underwater and at the surface.

HOW A BC WORKS

The BC (also known as BCD, or buoyancy compensation device) offers flotation at the water's surface and fine adjustment of buoyancy underwater, and acts as a harness for equipment items. At the surface, a fully inflated jacket will keep your head well out of the water. Underwater, buoyancy can be increased by inflating the BC with small amounts of air.

All BCs are connected via an air hose to the regulator first stage (*see p.61*), which supplies pressurized air from the tank for inflating the jacket. BCs can also be inflated orally, via a corrugated hose that hangs down over the left shoulder. Controls for automatic inflation and deflation are usually located at the end of this hose.

Of course, BCs must be able to reduce buoyancy as well as increase it. This is achieved by venting air from the jacket via dump valves at various points. Air can be released through the oral inflation hose by depressing a button, and can also be vented by using a pull-cord dump valve, which acts more quickly. These are often located on the opposite shoulder from the oral inflation hose, and at the bottom edge of the jacket. They should be used cautiously, since they alter buoyancy rapidly. Many BCs also vent air when the oral inflation hose is tugged sharply.

STOWING KIT
Buoyancy compensators (*right*) are fitted with handy stowage points for peripheral gear.

SURFACING IN SAFETY
When fully inflated, BCs allow the diver to float effortlessly at the surface. Here, two divers signal OK after a dive.

FIXINGS AND FASTENINGS

The tank (*see p.68*) is also attached to the BC. A strap attached to a rigid backplate wraps around the tank and holds it in place with a cam-action buckle. It is important that your tank is fastened securely and does not move during the dive. A well-designed BC will have various adjustable straps around the waist and at the shoulders, to ensure that it fits snugly. These can be easily re-adjusted underwater. Steel or plastic D-rings (*see below*) can be used to attach smaller items of gear.

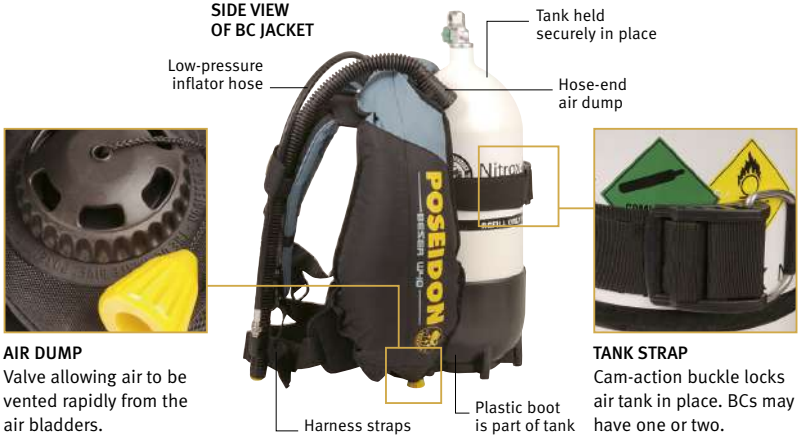

SPECIALTY JACKETS
High-volume "wing" style BCs (*top*) are used in advanced diving (*left*). The large air capacity of these jackets provides extra lift for heavily laden divers.

CHOOSING A JACKET

A basic BC is like a padded vest, and these are often only adjustable at the waist. More advanced models have adjustable straps to make sure the jacket fits as snugly as possible underwater. Ideally, you should buy one that fits like a glove and does not roll around with a tank attached. You should also find it easy to take on and off. If possible, you should try it on with your wetsuit (or drysuit) to check the fit, and to ensure that there is room for more adjustment on every strap in case of changes in your suit or your body shape. Beyond the basic models of BCs there are the more specialized features available on some jackets, such as weight integration, or a "wing" style design (*see opposite page*). Whatever the features of the jacket you decide to buy, make sure they suit the type of diving you will be doing the most. Try out several different jackets before deciding.

SERVICING YOUR BC

Careful maintenance of your BC will keep it in good working order for many years. It should be serviced at least once a year, or after it has been stored unused for more than six months. Take it to a recommended dive shop. Only service it yourself if you have been trained to do so. Do not take risks by diving with old or unserviced gear.

CARING FOR YOUR BC

As with all items of equipment, your BC should be rinsed with fresh water after every dive. The inside should also be rinsed by flooding it through the corrugated hose. The jacket should then be emptied again by turning it upside-down and draining it through the same hose. After this, you should reinflate it to help locate and remove any remaining water, which will rattle around inside when the BC is shaken. The jacket should be stored in a cool, dry place in a partially inflated state.

BUOYANCY COMPENSATOR DESIGNS

INTEGRATED WEIGHT BC
Some BCs feature integral weight systems. Many people find these quick-release weight pockets more comfortable to use than a traditional weight belt. The weights should be removed before putting the BC on and taking it off.

Inflator mechanism incorporates a backup regulator

INTEGRATED BREATHING BC
The Mares HUB (Human Underwater Breathing) system was the first to integrate the regulator into the design of the BC. A great advantage of this top-of-the-line product is its simplified profile, which reduces drag in the water.

Regulator integrated with jacket design

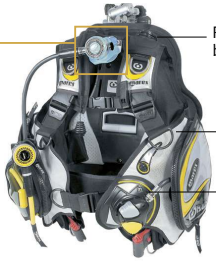

WING / SEMI-WING BC
Favored by technical divers (*see p.194*), these offer high lift capacity. The air bladder is rear-mounted: this aids horizontal posture under the water, but can push the diver face-forward at the surface. A semi-wing offers better surface support.

Semi-wing

Wing

WOMEN'S BC Women divers are increasingly well catered to with special jacket designs to give a comfortable fit without the usual tension points across the chest. Such jackets offer other convenient options like weight integration.

Specially shaped harness

TRAVEL BC If you travel often, it is a good idea to consider a lightweight BC without a rigid back, which can be rolled up. These designs have a lower lift capacity, as they are mostly used in warm-water sites, where minimal weight is carried by the diver.

Toggle for jettisoning weight pouches

Regulators

One of the most important items of diving gear is a regulator. This two-stage gas-supply system provides you with air from your tank at the correct pressure for you to breathe underwater. Since it is a key life-support device, most divers buy their own regulator.

BREATHING UNDERWATER

Your tank can store large amounts of air in a small space because the air it contains has been highly compressed (*see pp.68–69*). Before you can breathe this stored air, it must be reduced to a safe pressure, which varies according to how deep you are underwater. The deeper you go, the greater the weight of water pressing on your chest cavity from outside (*see pp.94–95*). This is called ambient pressure. Air for breathing must be supplied at an equivalent pressure, otherwise your lungs will not be able to inflate properly.

STORING REGULATORS
Regulators are tough, but they should be well cared for. Hang them up in a cool, dry place away from sunlight.

HOW REGULATORS WORK

Supplying air at the right pressure is the job of the regulator, which is made up of the first stage (which fits on to the tank) and the second stage (the piece you put in your mouth), which is also known as the demand valve. The first stage reduces the air passing out of the tank to a lower pressure of around 116–145 psi (8–10 bar) above ambient. A valve in the first stage allows some of this low-pressure air to pass into a hose, and then closes again. When the air in the hose has been inhaled by the

UNDERSEA EXPLORER
The regulator reduces the pressure of the highly compressed air in the tank to a level that is safe to breathe.

diver, the first stage detects a decrease in pressure in the hose, and allows more air to flow in, to replace that which was used. It automatically alters air-pressure in the hose when depth changes occur. The first stage also supplies air for BC and drysuit inflation (via separate hoses).

The second stage reduces the low-pressure air in the hose to the same pressure as the water you are swimming through, allowing normal breathing. The second stage houses a diaphragm that is subject to equal forces of water pressure on one side, and air pressure from your breathing passages on the other. When you inhale, the diaphragm flexes inward, causing an inlet valve to open, and air flows into your mouth. Exhaling causes the diaphragm to flex outward and triggers an exhaust valve to open, allowing used air to be expelled. The unit can be cleared of water by using the purge button, which is located on its front surface.

A-CLAMP FIRST STAGE
A common and convenient clamp-style first stage, suitable for most recreational diving requirements. Less secure than DIN.

DIN-TYPE FIRST STAGE
A screw-in first stage that can handle high tank pressures of more than 3,000 psi (200 bar). Highly secure.

SECOND STAGE
Designed to supply air safely at all depths, your second stage should fit comfortably in your mouth.

INSTRUMENT CONSOLE
A high-pressure hose from the first stage feeds directly into a pressure gauge, indicating remaining air supplies.

OCTOPUS SECOND STAGE
The octopus is for use in the event that your main second stage fails, or for donation to other divers.

BC JACKET
A low-pressure hose from the first stage supplies air to your BC to inflate the jacket underwater.

THE OCTOPUS SECOND STAGE

Most recreational regulators are set up in an "octopus" configuration (so called because of the multiple hoses emerging from the first stage). This incorporates a backup second stage, which is intended for use if your main one fails, or if your buddy needs a donated air supply. Over time, this backup air source has become known as the octopus second stage, or simply the octopus.

The octopus and the hose to which it is attached are usually finished in a bright yellow color for easy location in an emergency, and to distinguish them from your main second stage (which ordinarily has a black hose). The octopus hangs unused from your BC most of the time, and for this reason it is often detuned (adjusted) to prevent "free-flow"—an accidental discharge of air that occurs more easily when the mouthpiece of the regulator is left continually flooded. Some types of octopus are attached to the hose via a swiveling joint, so that they can be easily and safely donated to your buddy regardless of whether they are beside you, or face-to-face.

The hose supplying an octopus should be longer than the one feeding the main second stage. This will allow it to be offered to a buddy easily and for the buddy pair to remain face-to-face during a shared-air ascent (*see p.140*). Practice air-sharing procedures regularly.

STOWING THE OCTOPUS

Consider where to stow your octopus so that it can be easily located in an emergency by either you or your buddy. You can buy specialized clips of various designs to attach the octopus to your BC, which allow it to be freed easily when required, and replaced again afterward. Many divers hang their octopus near their midriff, but others mount it higher up on the chest, so it remains constantly within sight and close at hand. You can also buy regulator retainers made from rings of surgical tubing. These go around your neck, meaning the octopus is always right under your chin, though some find this arrangement constricting.

SAFE STOWAGE
Attach your octopus to your BC at a point where you can easily access it, and where it will be clearly visible at all times.

EMERGENCY ACCESS
Special quick-release clips are available for fastening your octopus to your gear.

TYPES OF REGULATORS

BASIC REGULATOR Simple, inexpensive regulators at the budget end of the market can be very reliable and durable. However, some may deliver air a little less smoothly than high-performance models.

OCTOPUS SECOND STAGE Octopus second stages are often just standard second stages with high-visibility color schemes. However, some designs can be swiveled to any angle for easy access.

HIGH-PERFORMANCE REGULATOR At the top end of the market are regulators that combine high build quality with low inhalation effort and toughness. Many are environmentally sealed for cold-water use.

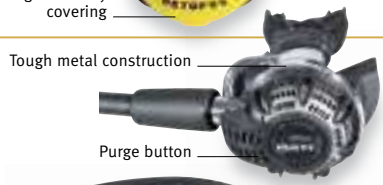

OXYGEN-COMPATIBLE REGULATOR Divers who wish to use nitrox mixes (*see pp.194–95*) that are very rich in oxygen will need a regulator designed to handle high oxygen concentrations without risk of combustion.

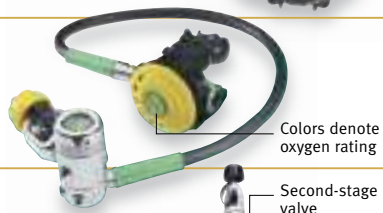

VINTAGE-STYLE REGULATOR In twin-hose regulators, the valve of the second stage is separated from its mouthpiece, and sits behind the diver's head, where it releases bubbles out of the diver's field of vision.

CHOOSING A REGULATOR

As with all items of dive equipment, prices vary enormously, but the regulator is one piece of gear worth investing decent money in. Basic regulators use a piston valve in the first stage and can be very reliable but often feature rather chunky second stages. The more expensive models use a diaphragm in the first stage and tend to have more streamlined second stages. Top-of-the-line models may be designed specifically for cold water diving, or be optimized for efficient air delivery on deep dives. They may also have dials to adjust the inhalation effort required to open the valve and deliver air. These can be used either to tune the unit to free-flow less easily, or maximize air delivery during a normal dive. Always rinse your regulator in fresh water after use, and have it serviced annually by a qualified dive-equipment technician.

DEEP BREATH
Regulators are designed to stringent standards, and are very reliable if well-maintained.

Gauges and consoles

To stay within safe time and depth limits and avoid running out of air, you must carry instruments that monitor your air consumption, time underwater, and depth. There are various devices on the market to help with this, from simple gauges to complex digital consoles.

BASIC INSTRUMENTS

You should never dive without some means to measure the time you spend underwater, the amount of air you have left in your tank, and your depth. A waterproof watch with a bezel—a rotatable outer ring—allows you to mark significant times, and thus keep track of how much time has elapsed. It also serves as a backup in case of failure of digital equipment used to monitor time, such as a dive computer.

Depth gauges allow you to monitor your depth even if you cannot see the water's surface. They are usually analog (needle-and-dial) devices, although digital versions are also available. Both work in the same way: they measure the pressure of the water around you, and convert this into an accurate reading of your depth. Make sure you buy one that records the maximum depth of your dive, and remember to reset it after every dive. Equally important is an air-pressure gauge, also known as a submersible pressure gauge (SPG) or contents gauge. This is connected to your regulator first stage via a high-pressure hose and measures the pressure of the air in the tank. You should check your air-pressure gauge every few minutes during the dive to carefully monitor your air consumption.

INDICATING REMAINING AIR SUPPLY
The air-pressure gauge allows you to keep track of the amount of air in your tank; let your buddy know (*see p.145*) when your supply is running low.

CHOOSING A CONSOLE

Your console is your dive information center, so you must be happy with the way it handles. Different consoles need to be held at different angles to read the dials; experiment before you buy to find out which is easiest for you to use. Choose one with large, easy-to-read numbers. Some are equipped with other instruments, such as a compass; these can be either on top of the console or on the back.

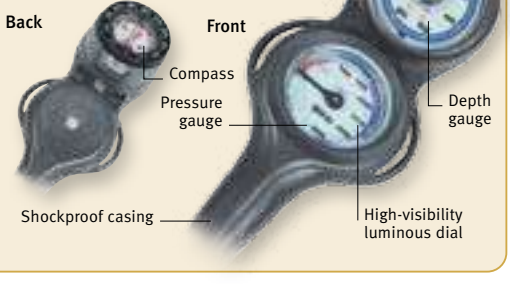

GAUGE CONSOLES

Depth and air-pressure gauges are often combined in a single integrated console that attaches to your regulator first stage and hangs by your side during a dive. Mounted together like this, the gauges are easy to locate and can be

TYPES OF GAUGES AND CONSOLES

STAND-ALONE DEPTH GAUGE A depth gauge is usually oil-filled and reacts to ambient pressure changes. Mostly analog, these gauges are accurate and durable.

STAND-ALONE AIR-PRESSURE GAUGE Typically analog, this type of gauge contains a tube that expands or contracts according to the air pressure inside your tank.

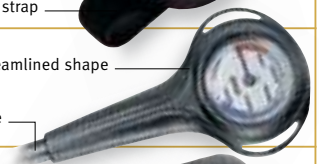

DEPTH AND AIR-PRESSURE CONSOLE This option allows at-a-glance checking of both depth and tank pressure, and also registers your maximum dive depth.

DIGITAL CONSOLE This console has an integrated computer and displays a range of information, such as air consumption, temperature, depth, time, and ascent rate.

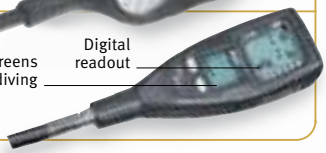

read simultaneously, like a "diver's dashboard." They can sometimes also include other instruments, such as a compass or a thermometer.

Consoles should be rinsed with fresh water after use and stored away from direct sunlight, and kept in a padded or protective case to avoid damage.

ADVANCED CONSOLES

Some manufacturers offer instrument consoles with digital rather than analog gauges. Consoles are also available that perform the same advanced functions as a dive computer (*see pp.66–67*). As well as indicating dive time, depth, and decompression schedules like a normal computer, these also incorporate a digital air-pressure gauge. For divers who use a wrist-mounted computer to monitor dive information, console instruments can provide an essential backup.

ANALOG VERSUS DIGITAL
Although less easy to read, analog gauges sometimes give slightly more accurate readings than digital ones, particularly at shallow depths.

Dive computers

To avoid decompression sickness, it is critical to monitor your depth and time underwater on every dive. In the past, this would have been done using a watch and a depth gauge, but these functions and many others can now be performed by a dive computer.

USING A DIVE COMPUTER

The main functions of a dive computer are to monitor your dive time and depth and calculate safe dive schedules. A basic model uses this data to assess the level of nitrogen absorption in your body while you are under the water, and also displays information about your ascent rate, including any stops you may need to make during an ascent to avoid decompression sickness (*see pp.100–01*). More advanced models also inform you about the temperature of the water and your air consumption (if linked via a transmitter to your regulator first stage), both for normal air and technical gas mixes (*see pp.194–95*).

Computers are more flexible than dive tables (*see pp.100–01*). Dive tables show how long you can spend at different depths before needing decompression stops. Most tables assume that you spend your entire dive at the greatest depth; computers can make more complicated calculations, including the time spent at all the different depths reached on a dive. Some computers will also schedule deep decompression stops. These reduce the risk of nitrogen microbubbles forming in your body, meaning that less time is spent decompressing in shallow water. When making repeat dives, especially in places where you are not limited by times of tides and slack water, using dive

TYPES OF DIVE COMPUTERS

MULTI-GAS This type of computer can calculate dive schedules for several different mixtures of breathing gas, from normal air to various oxygen-enriched blends of nitrox.

CONSOLE-INTEGRATED This computer is built into the diver's instrument console (*see p.65*), which is attached to the high-pressure hose from the regulator. As such, it also displays remaining air supplies.

WRIST-MOUNTED This instrument is typical of a basic level, watch-style dive computer. It provides data on depth, dive time, ascent rate, and safe dive schedules, and is for use with normal breathing air.

RADIO-UPDATED This wrist-mounted computer also displays how much air remains in the diver's tank by means of radio updates from a module screwed into the regulator's first stage (*see pp.60–61*).

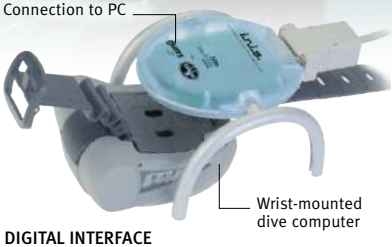

DIGITAL INTERFACE
Peripheral devices are available that link dive computers to PCs, enabling you to download dive profiles and create dive logs.

tables instead of a computer may not give you the maximum number of dives you can safely make each day.

Computers are also able to monitor your ascent rate accurately. In your dive training, you will be taught to ascend no faster than the smallest exhaled bubble, but it is more accurate to follow a digital readout that tells you the safe ascent rate.

BUYING A DIVE COMPUTER

A key consideration when buying a computer is the readability of the display. Too many figures on a small screen are difficult to read—the simplest models are often the best. Choose one with digits that are large enough to read with ease, without any nonessential data on the primary screen. It is also advisable to opt for a brand with a proven track record.

SAFETY CHECKS
As you ascend, keep a close watch on your dive computer to make sure that you are not exceeding the maximum ascent rate it recommends.

Computers are the most precise way of monitoring your dive, but they are not a foolproof means of avoiding DCS. It is impossible for data from a computer to exactly reflect your body's requirements. All models have built-in safety margins, but to be sure of avoiding problems, you should always dive conservatively, and not push your depths and times to the limit.

MALFUNCTION PRECAUTIONS

Never rely solely on a computer. As a precaution against malfunction, it is always worth practicing how to calculate dive times using tables. Although your computer will give early warning of a low battery, make sure you carry back up devices—either a watch and depth gauge, or a second computer, which can be mounted on your console.

Tanks

Also referred to as "cylinders" and "bottles," tanks contain pressurized air for breathing underwater and have a valve that delivers this air to the regulator. Requiring regular testing and inspection, they can almost always be rented from dive centers.

SIZE AND MATERIALS

Different sizes of tanks are available, but most commonly they have a volume, or capacity, of 72, 80, or 95 cubic feet. The maximum pressure to which these can be filled varies—it can be as much as 4,350 psi (300 bar), but is more commonly 3,000 psi (206 bar). These figures are stamped on the neck of the tank, along with the dates of manufacture and most recent testing. Visual inspection at a certified dive shop must be carried out annually, and hydrostatic testing (for metal fatigue) is required at different intervals (from one to five years) in different countries. Your test stamps must be in date for your tank to be filled by any reputable dive center.

SECURED AGAINST MOVEMENT Tanks stored upright must be secured. Always store them out of direct sunlight.

Many recreational divers use an 80 cubic foot tank, but if you find you usually run low on air before your buddies do, you should consider using a larger size. On deeper, or more lengthy dives, two tanks (a "twin-set") may be needed. Compact "pony bottles" are often used as a backup air source; mounted next to the main tank, these are equipped with an independent regulator.

TYPES OF TANKS

Tanks may be made of aluminum or steel. Steel is heavy and corrodes in salt water, but if well maintained it is durable and long-lasting. Aluminum is a lighter metal, but because tank walls must therefore be thicker, this adds bulk, and largely negates any weight advantage; they are also more prone to damage.

Suit inflation tank

3-liter pony bottle

Nitrox tank

A steel tank

Tanks are most commonly filled with compressed air. For technical diving (*see pp.194–95*), however, they can be filled with special gas mixes, such as nitrox and trimix. Drysuit inflation tanks, used mainly by technical divers, are filled with the inert gas argon, which, being denser than air, retains body heat for longer.

Because the air inside the tank is used up during the course of a dive, it gradually becomes lighter. Most people forget that air weighs a certain amount too, and compression packs a great deal of it into one tank. Aluminum tanks may even become positively buoyant as air is used up.

THE TANK VALVE

A valve screwed into the neck of the tank provides an attachment for the regulator first stage, for either an A-clamp or a DIN (*see pp.60–63*), and a means of turning the gas supply on and off. Always check that there is a sound O-ring in the valve before fitting your regulator. Without one, or with a damaged one, the fitting won't seal.

MAINTENANCE AND STORAGE

Fill your tank at reputable dive shops or centers and make sure it is "in date" and regularly serviced. If you are storing your tank on a boat, do not leave it standing up (especially if gear is attached) unless secured in a rack. Take care to lay it down and secure it so that it does not roll around. Tanks are full of highly-pressurized gas, so handle with care. Store tanks out of direct sunlight, especially when full, as this can cause a pressure buildup inside. If you have not used your tank for more than six months, have it refilled to get rid of any stale air.

DOUBLE CAPACITY
A twin-set is ideal for deep dives that require long, scheduled decompression stops.

Weight systems

When diving, your suit, BC, and even your own body all add to your buoyancy. There are many systems for carrying the lead ballast used to offset this, and they can be tailored to the environment you will be exploring. Comfort and ease of use should be a consideration.

STAYING DOWN

When wearing a drysuit or thick wetsuit (*see pp.50–55*), you will need to wear a considerable amount of weight to counteract the buoyancy of the suit. The most inexpensive way of wearing ballast is on a belt carrying hard lead tablets. Other systems use lead shot as the ballast, which is more comfortable to wear around your waist, as it molds to your contours. This is of most relevance to women, who often find that a standard weight belt is uncomfortable on the hips.

PASSING A WEIGHT BELT
When handing over a belt, always hold it with the buckle down, so that the weights cannot slide off.

Many BCs now feature integrated weight pockets, offering a more adjustable approach to weight distribution and trim. The lead carried by such BCs is mostly stored in quick-release pockets, but there are usually secondary stowage areas that are inaccessible during a dive. It is important to make sure the amount of releasable lead in the BC will give you sufficient positive buoyancy to guarantee an unassisted return to the surface if you have to jettison it.

WEIGHT SYSTEMS

STANDARD WEIGHT BELT A basic nylon belt carries lead tablets that have two slits through which the belt is threaded. The entire belt is dumped in an emergency, via a rapid-release cam buckle.

POUCH-STYLE WEIGHT BELT Some belts have pouches that hold bags filled with lead shot, which are easy to pull out and discard. A variation on this is a belt with a single compartment to which you add loose shot.

WEIGHT HARNESS Taking weight off your waist and suspending it from both shoulders makes carrying the load more comfortable. Quick-release catches allow dumping in an emergency.

WEIGHT-INTEGRATED BC This is perhaps the most comfortable way of carrying weight. The weight can be distributed around your body, aiding correct posture in the water (*see also* BCs, *pp.56–59*).

SECURING WEIGHTS TO A BASIC BELT

To prevent the lead tablets from sliding along your belt and altering your stability during a dive, you can secure them in place with plastic keepers. Alternatively, you can twist the belt to hold them in place. Insert the belt through the first slit in the weight, make a half turn, and then run the belt through the second slit.

Securing weights by twisting the belt

1 **To use a keeper** on a basic weight belt, thread the belt through the weight tablet, and locate in the desired position.

2 **Take the belt** up through the keeper, bend and thread back through the keeper, then through the weight.

3 **Pull the belt** taut on each side of the weight. Be careful to avoid kinking the webbing as you do so. Repeat as necessary.

REFINING YOUR WEIGHT SYSTEM

Weights are something you will not want to travel far with—and you do not need to, because they can be rented at dive centers. You will find that when you dive in different environments and in different suits, your weight requirements change. Record the configurations that work best for you somewhere handy, such as at the back of your log book. Whichever way you decide to wear your weights, it is vital that you can release them quickly and safely in an emergency.

LEAVING GRAVITY BEHIND
Heavy weights, though awkward on land, will not be a burden underwater.

ANKLE WEIGHTS

When a diver wears a drysuit (*see pp.54–55*), the air within the suit can become trapped around the legs and feet, inverting the diver. Ankle weights will help to counteract this undesirable effect.

Weights mounted in ankle strap

Dive flashlights

A flashlight is an essential item of dive gear. Although primarily used when diving at night and in conditions of reduced visibility, such as in caves or in murky water, flashlights also have other useful roles, including illuminating subjects for underwater photography.

LIGHTING UNDERWATER

Lights vary enormously in design and cost, but the basic function is to provide reliable illumination when submerged. There are good designs in nearly every price range, but flashlights, especially hand-held ones, are easy to lose and damage, so it is worth weighing up carefully how much you want to spend. It is also wise to obtain an equally reliable backup flashlight, such as a smaller pencil-style model, to stow in a pocket, for use if your primary light fails during a dive.

BURN TIME AND BATTERIES

One of the most important factors to consider when buying a flashlight is burn time—you do not want to run out of battery charge halfway through a dive,

USING A DIVE TORCH

Flashlights are easy to lose, especially during dives in strong currents. Wrist straps offer a degree of security, but it is safer to clip them on with a flexible leash. Stow them close to your body to minimize the risk of being hit on rocks. Also, try not to switch the light on and off too often during the dive, because this may cause the bulb to fail.

LIGHT IN THE DARKNESS
A cavern diver surveys the inside of a sea cave. Lights are an essential piece of diving safety equipment.

TYPES OF FLASHLIGHTS

HIGH-INTENSITY DISCHARGE This type of light burns brighter than a halogen bulb and, because it requires less power, has a long burn time. Such lights can be heavy and expensive, however.

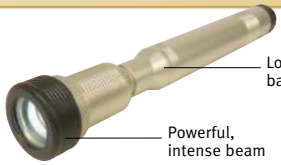

BACKUP A range of flashlights have evolved that are intended for use if the diver's primary light fails. As such, they are small and lightweight, but only generate enough light for emergency use.

PISTOL-GRIP This is probably the most widely used style of light among recreational divers. They are fairly inexpensive and usually use conventional non-halogen filament flashlight bulbs.

UMBILICAL An umbilical light has the battery housed in a separate case. This model uses high-intensity light-emitting diodes (LEDs) to generate a beam. Umbilicals have long burn times.

HIGH-TECH RECHARGEABLE This model can be charged via a special "wet" contact that means the case does not have to be opened, minimizing the risk of flooding through careless reassembly.

Rear of light

which could compromise safety. There is usually a trade-off between output and burn time. Some lights are very bright and offer an excellent field of illumination for a short duration. But many lower-output models can outlast much brighter competitors, so consider carefully which would be more useful for you before making a purchase.

Rechargeable flashlights are more cost-effective than those requiring battery replacements, and tend to offer the greatest light output but the shortest burn time. Also consider the recharge time, especially if you are diving from a liveaboard boat and the vessel's generator is run only for short periods of time.

BUOYANCY AND BEAM

Another factor to consider is buoyancy. Some flashlights are positively buoyant, because they have a large airspace within the reflector. This makes them unbalanced in the water and difficult to handle. A positively buoyant light cannot be laid down, for example, if you need to light a subject for photography.

Lamps also vary in the size of the hotspot (central area of beam) and peripheral halo. The halo can be useful because it lights up a wider area than just the main hotspot. The hotspot is the brightest part of the beam, and is best for highlighting specific objects or for illuminating murky surroundings.

Surface marker buoys

Marker buoys may be used either during a whole dive or just before you ascend, to let those at the surface or on the shore know where you are, and also to aid your ascent. These devices offer a sense of security for both diver and surface cover, and aid safe recovery.

SMB OR DSMB?

A surface marker buoy (SMB) is an inflatable float linked to the diver by a line. It allows an observer at the surface to monitor the diver's position underwater. SMBs should be in a high-visibility color—most are orange.

The line to which the float or buoy is attached is usually stowed on a special reel. The reel should be held in one hand, with not too much slack on the line. You should use an SMB when your boat cover needs to be aware of your position at all times, such as during drift dives (where undersea currents propel the diver along), or where there is a likelihood of heavy boat traffic overhead.

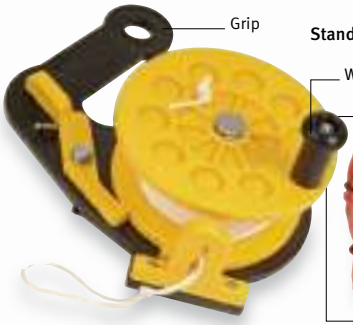

STANDARD REEL AND LINE
Reels are available for both right- and left-handers for ease of winding.

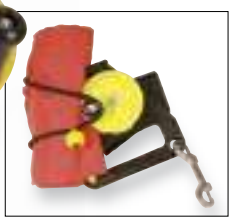

REEL AND DSMB
Line and buoy can be wound and clipped together for compact storage.

Another type of buoy is deployed only at the end of a dive, or in an emergency. The delayed SMB (or DSMB) is carried deflated during the dive, then inflated underwater (*see pp.130–31*) before ascent, carrying a line from the reel up to the surface. When it pops up,

DIVER WITH DSMB INFLATED
By the time you surface, your boat cover should be en route to pick you up.

DIVER WITH SURFACE MARKER BUOY
Only one SMB, inflated before the dive, is needed per buddy pair, provided you remain together during the dive.

it shows where you are about to surface. Its line also acts as a guide for making a slow, safe ascent and for decompression stops (*see p.101*). Always make sure you are some distance from other divers using buoys before deploying, or you risk fouling each other's lines.

USING A REEL

Both SMB and DSMB lines should be tied with a bowline knot to the reel. A reel should be compact enough to fit into a pocket or clip onto your BC, but not so small that it is difficult to operate, especially if gloves are worn. Some reels have a ratchet action to stop the line from unwinding on its own, and this helps to avoid tangles. The length of line you use should correspond to the depth you are expecting to dive to; this may be 130 ft (40 m) or even more. If the line is too short, then when it runs out, the DSMB will yank you upward at a dangerously fast rate until the buoy reaches the surface. This can also happen if the reel jams, which is why you must never deploy a DSMB while the reel is still clipped to your BCD: you must hold the reel in your hand so that in an emergency you can simply let it go.

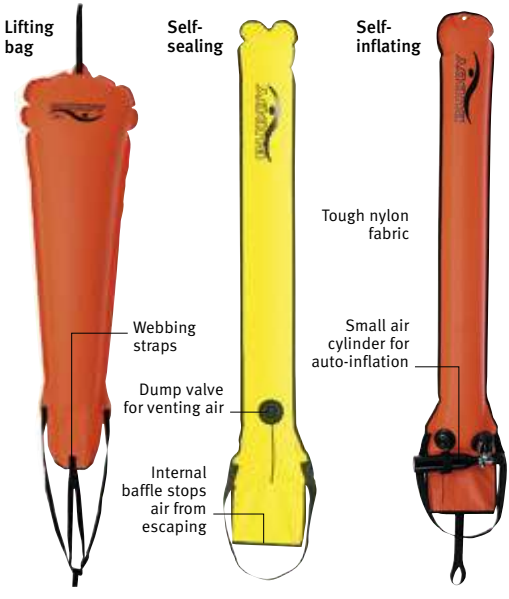

TYPES OF DSMBS
The cheapest lifting bags have an open base from which air can escape after inflation. Self-sealing types are more reliable.

Safety and signaling devices

It is rare to be lost at the surface, but you must be equipped for such an eventuality. To guide rescuers to your position, you should carry high-visibility surface location aids. The boat crew must know what color equipment you are wearing, and which aids you are likely to be using.

SIGNALING OPTIONS

Even on a relatively straightforward dive, the current can take you a long way away from your boat cover, and heavy swell can make you difficult to locate. If this happens, stay close to your buddy, and, if possible, secure yourself to them. Signaling to the boat will speed up recovery, and there are various gadgets available for this. The devices that are most commonly used are inflatable surface marker buoys (*see pp.74–75*). It is best to use the long "safety sausage" type, which stands up in the water, and remains inflated once deployed, rather than an open-ended one. If you opt for the latter, make sure you also carry an alternative device, such as a collapsible flag. Surface horns and power whistles are also used. They are of limited range, but can help to attract

SURFACE OBSERVATION SIGNAL This tall, inflatable signal is easily seen by boats or aircraft. It is inflated using compressed air from your cylinder.

SIGNALING DEVICES

FLAG Collapsible plastic flags in high-visibility or fluorescent colors are excellent for signaling at the surface, especially in a swell. They can be rolled and stowed alongside your cylinder.

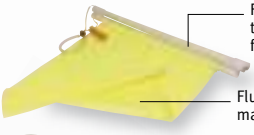

STROBE LIGHT Strobes are particularly effective when attached to the top of a folding flag. If attached to your BC, the flash may only be intermittently visible as waves lap over and around it.

DISTRESS FLARES Aerial flares do not pinpoint your position, so they are best used in conjunction with other devices. They can be difficult to operate when wearing gloves, or with cold hands.

PERSONAL LOCATOR BEACON (PLB) Devices that broadcast radio distress signals can be used to alert surface and air searchers to your presence. Some are also equipped with bright lights for close-range visual detection.

attention if the boat is not too far away. You should always carry a flashlight, since the beam can be easily spotted at night, especially when moved slowly in a scanning motion, both vertically and horizontally (*see pp.184–85*). Make sure you also carry a backup light. Strobe lights are even more effective, since their flashing light travels in all directions, eliminating the need for scanning. Flashlights and strobes must be robust and reliable; regular inspection and maintenance is vital, with special attention to sealing surfaces and O-rings.

Reflective signals are useful in bright weather. It is worth carrying an old CD or mirror to use as a reflective device to attract attention. Reflective patches on your hood and upper body may be visible to aircraft searching overhead. Dye markers stain the water and can help a search from the air, but should only be used once the aircraft is overhead, since the dye rapidly dissipates.

If you find yourself on the surface with no signaling devices whatsoever, waving your arms or even your fins and shouting may help to alert attention if your boat cover is nearby. Keep calm and persist with signaling.

INTEGRATED SAFETY DEVICES

Some safety devices can be integrated into existing gear. Air-horns, for example, attach to the low-pressure hose on your BC. These are far louder than whistles, and are useful if you are out of breath or under stress. Safety sausages can also be inflated with your BC.

Standard BC

Air horn attachments

RESCUE FLAGS
A brightly colored flag is a cheap, simple-to-carry, easy-to-deploy, and highly visible location device.

Knives

A dive knife is essential for freeing yourself if you get snagged on fishing lines and other obstacles. The simplest dive knife—little more than a basic blade in a plastic mount—will get you out of most tangles, but a good one will prove an invaluable multipurpose tool.

SELECTING A KNIFE

Knives must be rustproof, but the better the rust-proofing of stainless steel, the harder it is to maintain a sharp edge. Titanium knives, while expensive, are light, rust-resistant, and stay sharp. Knives come in various sizes. Small ones can be mounted on your BC or belt, where they are easy to reach and out of the way.

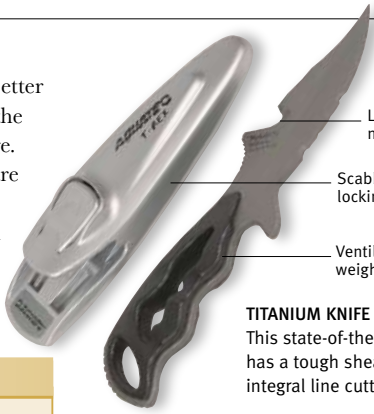

TITANIUM KNIFE
This state-of-the-art blade has a tough sheath and an integral line cutter.

SHEARS AND MULTI-TOOLS

Multi-tools incorporate both types of knife blades, for slicing and sawing—useful for rope and monofilament line. Underwater shears are only really needed if diving in very hazardous conditions or for use as tools in salvage.

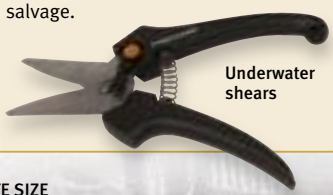

Standard knives with 4½ in (11.5 cm) blades are most popular. Blades may be serrated for sawing rope or smooth for slicing monofilament line, and may have a line-cutter notch. Blunt-ended blades are ideal for tasks that need leverage, such as prying open containers. It is worth choosing a knife with a metal pommel on the end, which can be used as a hammer.

KNIFE SIZE
Large knives are popular, but a smaller blade can be just as useful on a dive.

Bags

Whenever you move to and from a dive location, you will need a bag to carry all your gear. Ideally, your chosen dive bag should be capacious and light, but it must also be durable enough to survive the rigors of transporting heavy equipment.

BUYING A BAG

When choosing a bag, check that it has strong seams, sturdy handles, and no parts that are likely to rust. Also consider security, especially when travel will involve your being separated from your baggage. Thin bags can be cut open, so a padlock will not guarantee that the contents stay safe. You may want to invest in a rigid or semi-rigid bag (*see p.238*).

Many divers use a large duffel bag for travel, and a smaller day bag (or even two) for dive days. Duffel bags range in size from 6,000 to 10,000 cu in (100 to 180 liters), day bags from 1,800 to 5,000 cu in (30 to 80 liters). There are countless "extra" bags, such as fin bags, mask bags, tool pouches, and pony-bottle pouches. While useful, most of these are not essential if you pack your gear carefully into one large duffel bag.

FLIGHTS TO FAR-AWAY LOCATIONS
Dive baggage has to be hard-wearing and tough, especially as several transfers are often needed to reach a dive location.

BAGS FOR DIVERS

MESH BAG Perforated bags can be used to rinse all your gear at once in fresh water, or to carry small items aboard a boat without fear of water ruining the bag.

DUFFEL BAG These can accommodate most of your gear, and also have various compartments where you can stow small items, such as slates and carabiners.

REGULATOR BAG Your regulator is valuable, and needs its own padded bag. It can be put in your main bag, but when flying, most divers stow it safely in their carry-on.

DRY BAG When you reach the dive site, a watertight bag or case is essential to protect clothing and electrical equipment, such as cameras, from moisture and salt.

TRAVEL CASE A secure, rigid-sided case gives extra protection to your gear while in transit. Wheels and a telescoping handle make transportation easier.

Gadgets and accessories

There are a host of diving gadgets on the market, including fun items such as underwater MP3 players. Other accessories, however, have more practical uses. Clips and retractors are very useful for securing equipment, and may save you the loss of a camera or flashlight.

DIVER PROPULSION DEVICES

Underwater scooters, usually referred to as DPVs (diver propulsion vehicles), make great toys, but they can be practical and useful, too—for example, when exploring a large wreck site. There are two basic types: those that you sit astride, and those you hold on to, which tow you along. It is safest to use DPVs only to travel on a horizontal plane. Descending or ascending under power can be very dangerous, due to rapid pressure changes. DPVs are not cheap, but there are less expensive models suitable for general recreational use—that is, where failure of the device would hardly constitute an emergency. More expensive but tougher, more reliable machines are generally not aimed at the leisure diver but intended for expeditions, such as cave explorations, where failure of the unit mid-dive would be a more serious development.

Jetboots, made by a California diving company, also propel the diver along effortlessly: they comprise a pair of propellers that attach to your legs, controlled by a switch at your waist. Jetboots can propel you as fast as some DPVs, and leave your hands free.

DIVER USING SCOOTER
A diver propulsion vehicle, or DPV, takes the effort out of exploring large dive sites.

BANDS, CLIPS, AND FASTENINGS

Cam bands—webbing straps that hold your cylinder to your BC—are available with a number of different buckle mechanisms that ensure that the tank is secure and held close to your body. Retractors are very useful for attaching items, such as a flashlight or slate, to your BC because they pull the item close to your body when not in use. Carabiners and bolt-snap clips (as used on dog leashes) are useful for attaching equipment securely to your BC during the course of a dive.

Retractor

Spring-loaded snap clips

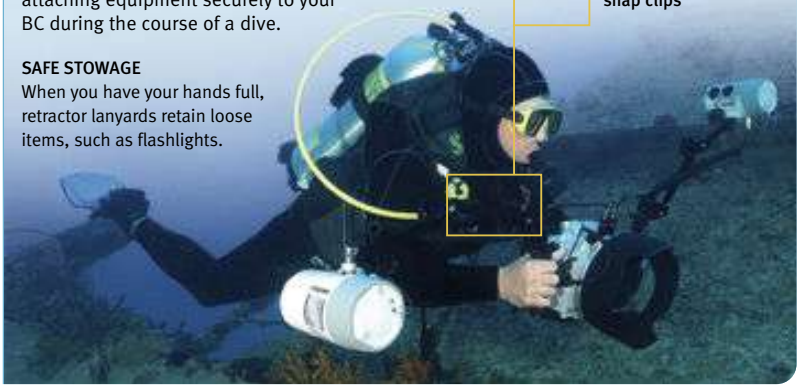

SAFE STOWAGE
When you have your hands full, retractor lanyards retain loose items, such as flashlights.

ATTRACTING ATTENTION

There is a range of gadgets that you can use to attract your buddy's attention underwater. Some BCs have a buzzer that uses a small amount of compressed air to generate a sound. Slightly more low-tech are tank pingers—a piece of elastic with a hard plastic ball attached. This is slipped around the cylinder and can be stretched and released to "ping" against it, creating a metallic sound that travels well underwater. Shakers are tubes filled with metal balls. When agitated, these create a clearly audible rattle.

WRITING UNDERWATER

Slates are handy for mapping dive sites or making notes on any unusual species encountered on your dive. They are made of white plastic, with a matte surface that can be written on with a pencil. Flat slates can be attached to a BC or slipped into a pocket; wrist-mounted slates are contoured to fit around your arm. The latter are especially useful when making technical dives with various decompression stops: you can read your computer and the list of stops from the same arm. A compass can be mounted on a flat slate, enabling you to look at navigational notes while checking your bearings.

AIR-LOCK TANK CAM BAND
Some dive gadgets are labor-saving devices. Tank band buckles that can be tightened using compressed air are a useful example.

New technology

In the 1950s, development of the scuba unit revolutionized diving, and its technology continues to develop apace. Modern equipment is sleek, lightweight, and colorful, and advances in the efficiency of rebreather systems are transforming diving once again.

MASKS AND VEHICLES

One of the most notable recent advances in modern dive equipment is the use of head-up displays in masks, which are similar to those used by military pilots. Essential dive information is sent via radio frequency from the tank to the mask, where it is presented on an LCD. Masks with an integral radio communication system, which allows divers to talk directly to one another, are also available.

Advanced synthetic materials have made dive equipment stronger and lighter, and have been used in the development of diver propulsion vehicles (DPVs; *see p.80*) that are smaller and easier to handle.

MASK TECHNOLOGY
New mask developments include head-up displays of critical information (*top*) and masks that correct optical distortion underwater (*left*).

FROM SCUBA TO REBREATHER

In the last decade, rebreather technology has become available to the recreational diver, representing the most dramatic leap forward in diving equipment since the development of scuba itself.

In contrast to scuba—an "open circuit" system that discharges most of a diver's air supply into the water with the exhaled breath—rebreathers recycle the diver's exhaled breath, using chemicals to remove waste (mainly carbon dioxide) and allowing the diver to inhale the gas again. Each breath uses about 4 percent of the oxygen in the inhaled gas mix, so one of the main tasks of a rebreather is to maintain oxygen levels in the breathing loop, generally by adding extra oxygen.

UNDERWATER SCOOTER DPV
The breathing observation bubble (BOB) is powered by an electric motor and has a top speed of 2.5 knots.

Rebreathers using pure oxygen were used by frogmen in World War II, for whom the absence of bubbles was vital to hide their presence. Dive depths were limited to 20 ft (6 m), at which point pure oxygen becomes toxic. A revolution in rebreather technology occurred in the mid-1980s, allowing the use of nitrogen–oxygen gas mixes rather than pure oxygen. Modern rebreathers electronically monitor the concentration of oxygen in the mix and adjust it to safe levels for a given depth.

Rebreathers offer longer dive times and fewer decompression stops. They remain costly, and need careful maintenance and specialized training.

REBREATHER TECHNOLOGY

Modern rebreathers are divided into two categories—semi-closed rebreathers (SCRs) and closed-circuit rebreathers (CCRs). SCRs expel a small quantity of exhaled gas into the water, whereas CCRs only dump gas into the water on ascent. Both types clean exhaled gas in a scrubbing unit, which removes CO_2. They also ensure that the oxygen content of the inhaled gas is at a suitable level for the diver, and monitor toxicity, which the diver can check manually on a gauge, or on a computer display. Computer control has dramatically increased the efficiency and safety of rebreathing devices.

HOW A REBREATHER WORKS

A chemical in the scrubbing unit absorbs carbon dioxide from exhaled gas, while pure oxygen is introduced to the breathing loop to replace that consumed by the diver.

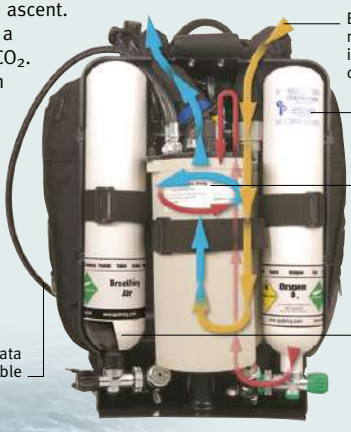

Mouthpiece

Oxygen flow

Inhaled gas flow

Exhaled gas flow

Computer display

REBREATHER BENEFITS

Because gas is recycled in a rebreather, you use it up at a slower rate than with scuba gear, and so can spend longer underwater. The consistently high levels of O_2 help reduce post-dive fatigue.

Equipment assembly

The gear assembly process is a key part of any dive. During your training, you will be taught the basics of assembling dive gear. With experience and a few simple tips, you can perfect this process and make it efficient, rapid, and—most importantly—consistently safe.

THE IMPORTANCE OF PREPARATION

With clear, blue water beckoning, the Sun shining, and the dive site awaiting you, you may find that there is a great temptation to rush the gear assembly process. However, the equipment used for diving is nothing less than life-support gear, and should be treated with a corresponding level of respect. It is reasonable to say that putting dive gear together safely before jumping into the sea is as important as packing a parachute correctly before jumping out of a plane.

CHECKING FOR DAMAGE

The first stage in assembling your gear is to carry out a visual inspection. This is particularly important if the equipment is unfamiliar or rented. It is essential to examine the test dates (*see p.68*) for the

1 **Lay out equipment** on a mat or other clean, grit-free surface where you can assemble the gear easily.

2 **Wet the tank** band on your BC, if possible, since this helps improve its grip on the tank's metal surface.

3 **Position the tank** with the pillar valve O-ring facing the backplate of the BC, then fasten the band around it.

4 **Attach the regulator** first stage to the tank. Do not overtighten the clamp (or screw thread, on a DIN fitting).

INFLATOR VALVE
The spring-loaded collar on the end of the hose fits on to the inflator valve.

5 **Pass the inflator hose** from your regulator's first stage (*see pp.60–61*) through the jacket's guides and straps before connecting it to the BC's inflator valve. Make sure the valve has engaged securely with the hose coupling.

tank and also the integrity of the O-ring (*see p.88*). Any nicks or cuts in the O-ring may lead to serious problems during the dive, yet replacing a damaged O-ring before the dive takes only seconds. Also look for wear and tear on the regulator hoses, particularly where they join the ports of the first stage. This is a common point of strain for regulator hoses, resulting in cracks and splits in the hose with sustained use. If the integrity of a hose has degraded, it may rupture when the air supply is turned on.

Check for signs of damage to the BC. With regular diving, there will be inevitable signs of wear and tear, but be sure to check the integrity of the buckles and load-bearing straps. These must be completely sound.

ASSEMBLING YOUR GEAR

When you are satisfied that your equipment is in good order, follow the steps below to make it ready. Take particular care to ensure that the band securing the tank to the BC is neither too high (resulting in the tank possibly coming free), or too low (resulting in the pillar valve restricting movement of your head during the dive, or actually striking it on entry).

6 **Turn on the air** supply, opening the tank's pillar valve all the way. Then turn the knob a quarter-turn back.

7 **Test air supply** by breathing from both your primary and secondary air sources. Check air quality and delivery.

8 **Secure the gear**, if you are not diving immediately, by tucking any loose parts inside the BC for protection. Most breakages occur in transit.

Disassembly

The end of a dive is an exciting time, and neglecting your dive gear at this stage is an easy trap to fall into. Take a little time, however, to clean and store your equipment properly: not only will it last much longer, but everything will be in order at the start of your next dive.

DEVELOPING A ROUTINE

As with so many things in diving, routine is the key. Many experienced divers attend to their gear the moment the dive is finished, and regard the dismantling, rinsing, and storage of their gear as an integral part of the dive itself. This is particularly true for diving from a boat, where a shortage of space frequently leads to cluttered decks and broken gear.

Diving equipment can cost a considerable amount of money, so it is worth taking a few moments to make sure it is properly dismantled and safely stowed.

The first stage of disassembly is to turn the tank off and purge any remaining air from the hoses using the purge button on the second stage (*see also pp.60–63*). The regulator first stage may then be removed from the pillar

1 **Shut off the gas** supply, turning the knob on the tank's pillar valve all the way off before disassembly.

2 **Purge gas from** the system using the purge button on the regulator to vent any air remaining in the hoses.

3 **Disconnect the inflator** hose from your BC's inflator valve, and the hose from the drysuit inflator, if worn.

4 **Disengage the regulator first stage** from the pillar valve of the tank, and free the hoses from the BC by undoing all the clips and straps.

A-CLAMP FIRST STAGE
Remove this type of first stage by loosening the clamp first.

DIN-TYPE FIRST STAGE
Turn the knurled collar to unscrew the regulator first stage from the pillar valve.

valve of the tank. It should take no more than finger pressure to turn the clamp or collar; any more resistance indicates that compressed air remains in the system. You will need to check that the tank is completely turned off, then purge the hoses again.

After removing the regulator first stage from the tank, use a blast of air from the tank to dry the dust cap and filter. Fit the dust cap securely, then rinse the regulator thoroughly in fresh water. Never forget to put the dust cap on, otherwise water will enter the delicate mechanism of the first stage: this is to be avoided at all costs. Hang the regulator up to dry, ensuring that the hanging method does not put strain on the hoses at any of the points of attachment to the first stage. You can then detach your BC from the cylinder and drain it of any saltwater. Rinse the BC inside and out with fresh water, and inflate it orally to allow quick drying and prevent the formation of salt crystals inside.

Dive gear must be dried thoroughly between dives, ideally by hanging it up out of direct sunlight. Do not store it where it is compressed or where moisture may accumulate, such as in a boat deck locker or a diving bag.

5 **Remove the BC** from the cylinder by undoing the cam buckle on the strap, and then sliding the BC upward until it is clear.

6 **Drain the BC** of saltwater by depressing the deflator button at the end of the corrugated hose. This allows the water to escape through the oral inflation point.

Equipment maintenance

Diving equipment is your life-support system underwater, yet many divers take a worryingly relaxed approach to maintaining their gear. A few simple steps will increase the life span of your dive gear, avoid the expense of constant replacements, and increase your safety.

TOOL KITS

Most divers carry a set of Allen keys, screwdrivers, or adjustable wrenches. Scuba multi-tools are all-in-one alternatives that incorporate virtually every basic tool that a diver might need on a dive boat or at a dive site. These include a range of hexagonal Allen keys designed to fit dive equipment and standard wrenches for regulator adjustment, as well as screwdrivers and a marlin spike.

Diving shops also sell "save-a-dive" kits, which consist of a set of lubricants, parts, and tools that can be used in the majority of equipment-related incidents on site. O-ring replacement and adjustments to hose configuration, for example, are simple, commonly required tasks that can be undertaken by anyone with such a kit.

SCUBA MULTI-TOOL
A useful gift for any diver, this contains tools for most small maintenance jobs.

Experienced divers also carry spare parts with them. This is sensible diving practice, and while the range of items differs from diver to diver, it usually includes items such as spare O-rings, silicone grease, tape, adhesive, an O-ring extractor, spare plugs for the regulator first stage, and any spare buckles or straps that might be necessary.

O-RING MAINTENANCE

The most disappointing reason for having to cancel a dive is the lack of a spare O-ring for a cylinder. O-ring malfunction may be indicated by air escaping once the regulator is fitted, but even in the absence of this you should make regular inspections to check that the O-ring has not degraded. In many cases, simply cleaning the O-ring and the valve groove with a lint-free towel will solve the problem. Pull the O-ring through your fingers—if you can feel cracks, it should be replaced. When removing an O-ring, never use a sharp tool, as this is likely to damage the ring.

1 **To remove the O-ring** from a cylinder's pillar valve (*see pp.68–69*), use a specially designed extractor tool. Alternatively, use the rounded end of a tie-wrap or even a credit card.

2 **Inspect the O-ring closely** and replace it if any cracks or tears are visible. Clean the valve groove with a lint-free towel, then press the new O-ring in carefully with your fingertips.

ADDING A REGULATOR HOSE

To add a hose to your regulator or change an existing hose is a simple process. You must avoid attaching a low-pressure hose to the high-pressure port, as this will rupture the hose. Most regulators now make this impossible by having different- sized ports, although older models may not have this feature. The high-pressure port is for the pressure gauge, while other ports are for a second stage or the inflator hoses. It is worth taking time to ensure that the hose configuration is practical and comfortable, to avoid strain on the base of the hoses and reduce drag.

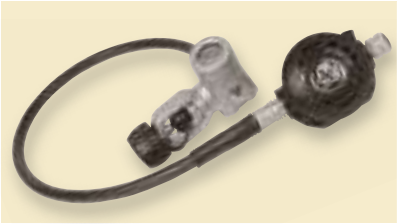

1 **When adding a hose**, be careful to select the correct port (high/low pressure). If the ports are not clearly differentiated, or if you are in any doubt, consult the manufacturer's manual.

2 **Remove the plug** from the regulator port using the correct-sized Allen key. Some regulators still have raised bolts—these can be loosened with a wrench.

3 **Attach the hose** by simply screwing it into place. Tighten the nut with a wrench of the appropriate size—hose nuts are soft and can easily be damaged by pliers.

4 **Exactly the same** procedure can be used to attach further hoses to the other ports, and to replace damaged hoses. Always check hoses for cracks, fraying, and abrasion.

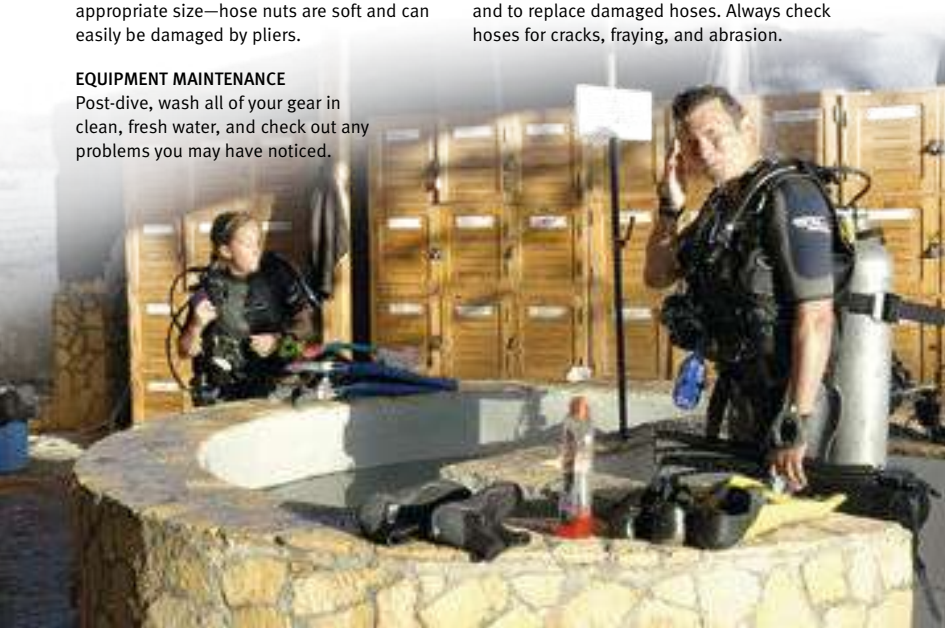

EQUIPMENT MAINTENANCE
Post-dive, wash all of your gear in clean, fresh water, and check out any problems you may have noticed.

The diving environment

Introduction

When you go diving, you are introducing your body to an alien environment. Evolution has not equipped our bodies for the underwater world, and you will encounter unusual sensory phenomena when you descend into the depths for the first time.

LIGHT AND VISION

Our eyes are designed to focus in air; underwater, our vision blurs unless we wear a mask, which puts an air space in front of our eyes. However, light travels more slowly in water than it does in air, and as it passes through the mask from water to air it bends, causing refraction—an optical effect that makes objects appear bigger and closer than they are.

Colors also look different to us as depth increases (*see p.183*). Objects near the surface look as colorful as they would above water, but as you go deeper, first reds disappear, then oranges, yellows, greens, and blues, until all light is lost. Colors can be revived by illuminating a submerged object with a flashlight.

SOUND AND HEARING

Above the surface, sound travels at approx 1,150 ft (350 m) per second, but underwater it travels four times faster. This makes it hard to identify the direction of a sound source: if your buddy signals to you by tapping his tank, you will hear the sound easily but may not be able to pinpoint where it is coming from. The range of audible sounds is also reduced, but the underwater environment is not as quiet as is popularly supposed: you can hear the scratching and snapping sounds made by fish and crustaceans, along with the sound of your breathing and any

ANATOMY OF A DIVER
The land-based development of the human frame means that your body is not adapted to many aspects of the underwater environment.

boat noise close by. However, you will find it difficult to communicate verbally underwater, even if you remove your regulator and shout, which is why divers use hand signals (*see pp.144–45*).

Sounds made above water do not easily penetrate below the surface, so any noises made to alert divers must be made underwater. A typical recall signal from a dive boat is to bang on the metal ladder on the craft's stern, which, as its lower part is in the water, transmits sound to divers beneath.

HEAT AND PRESSURE

Water conducts heat 25 times more efficiently than air, so the human body loses heat quickly underwater. In extreme cases, a diver may lose heat faster than it can be generated by their body, leading to hypothermia. When diving, it is important to always wear thermal protection appropriate to the conditions.

The airspaces in the body—not just the lungs, but the inner ear, sinuses, and gut—will be increasingly affected by pressure as you make your descent. This can lead to uncomfortable sensations (*see pp.98–99*) that must be relieved by equalizing internal and external pressure manually using the Valsalva maneuver.

SEA CHANGE
Divers patrol a reef wall in the Red Sea. As depth increases, physiological effects become more profound and numerous.

Depth and pressure

To understand the principles of diving, it helps to have a grasp of basic physical laws. When you dive, water exerts pressure on your body, and this pressure increases with depth. This simple fact influences all aspects of diving, from buoyancy control to safety considerations.

UNDERSTANDING PRESSURE

On land, your body is constantly subjected to external pressure from the weight of all the gases in the atmosphere. This is called atmospheric pressure, and it is measured in pounds per square inch (psi) or bar. Atmospheric pressure at sea level is 14.7 psi (1 bar)—that is, 14½ lb pressing on every square inch (1 kg per square centimeter) of your body's surface. This pressure is not fixed; it varies slightly with weather conditions and decreases with altitude. At the summit of Mount Everest (29,028 ft/8,848 m), atmospheric pressure is only 4.5 psi (0.31 bar).

When you dive underwater, pressure is also exerted on your body by the weight of the water above and around you. Because water is denser than air, the pressure generated by just 33 ft (10 m) of seawater is equal to that generated by all the gases in the atmosphere above sea level. Seawater is denser than fresh water, because it contains dissolved salts and minerals. This increased density means that pressure increases slightly faster with depth in seawater than in fresh water.

When you dive, you are subject to the cumulative weight of both atmospheric pressure and water pressure. The combined total is known as absolute pressure. Your depth gauge translates pressure into an accurate depth reading, but it is calibrated to ignore atmospheric pressure, so it indicates zero depth when you are at the surface. This is known as gauge pressure. Always use absolute pressure to calculate the total pressure at a given depth.

HOW PRESSURE RELATES TO DEPTH

Because water is much denser than air, pressure increases quickly with depth underwater. Water cannot be compressed and transmits pressure freely, so pressure increases cumulatively, and at a constant rate, as you dive deeper. At a depth of 33 ft (10 m) of seawater, absolute pressure is 29.4 psi (2 bar). This figure is the sum of the pressure generated by the atmosphere (14.7 psi) and by 33 ft of water (14.7 psi). At 66 ft (20 m) absolute pressure is 44.1 psi (3 bar), and so on, rising by 14.7 psi for every additional 33 ft of water depth.

If diving at a high altitude site, like a mountain lake, atmospheric pressure is reduced; consequently, absolute pressure is lower, too. You must use special dive tables at such sites.

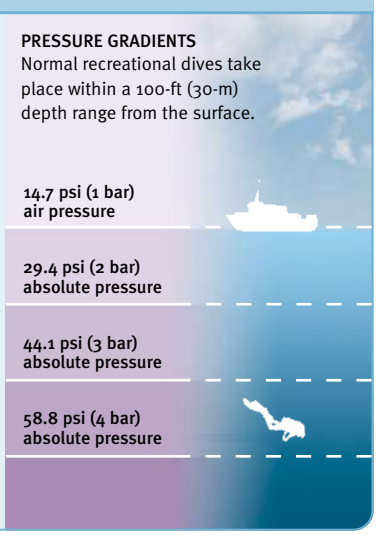

HOW PRESSURE AFFECTS GASES

In the 17th century, scientist Robert Boyle made an important discovery. He found that at a constant temperature, the volume of a gas is inversely proportional to the pressure exerted on it—meaning that as pressure increases, the volume that a gas occupies decreases. This relationship is known as Boyle's law.

Boyle's discovery has important implications for divers. Buoyancy devices such as BCs contain bladders filled with air. These provide upward lift to help you float weightlessly underwater. As you descend, however, increasing water pressure compresses the air inside the BC. This means it displaces a smaller volume of water, and so generates less buoyancy (*see p.97*), causing you to sink. You must then inject more air into the BC to boost volume to its former level. On ascent, this must be released again, because as pressure decreases, the extra air inside the BC expands, causing you to become very buoyant and rise too quickly. For the same reason, you should never hold your breath underwater. Compressed air inhaled at depth expands as you ascend, and if it is not exhaled, it can burst delicate vessels in your lungs.

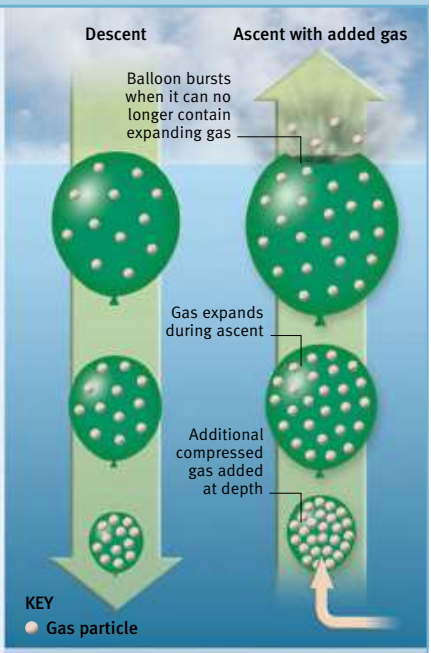

If you took a volume of gas, like that contained within a balloon, and submerged it, the space the gas occupied would get smaller with increasing depth.

If you added extra gas to the same balloon at depth, it would expand slightly. During ascent, the gas inside would expand to more than the balloon's original volume.

IN DEEP
Diving exposes you to pressure effects not encountered on land.

Breathing compressed gas

On land, we seldom need to worry about the availability of air, its composition, or the pressure it is supplied at. But when you venture underwater, things are very different. Being aware of how the diving environment affects your respiration cycle is of prime importance.

BREATHING UNDERWATER

Scuba divers draw breath from a tank of highly pressurized air, via a regulator. The regulator (*see pp.60–61*) reduces the pressure of the air in the tank to that of the water around you (ambient pressure), allowing it to be inhaled safely. As depth increases, so does ambient pressure, meaning that the regulator must adjust the pressure of the air it delivers to match.

Air, like any gas, is compressible, and becomes denser when pressurized. This means that at depth you will be inhaling more air per breath than you would at the surface (because the gas molecules are packed in more tightly). Because air is mainly composed of oxygen and nitrogen, this means you will be breathing higher concentrations of both gases than you would be at the surface, though the percentage of each in the mixture does not change. Both of these gases can have toxic effects on the body when pressurized. Nitrogen enters your body's tissues, and can have a narcotic effect (*see pp.100–01*). Oxygen is vital for supporting life, but becomes toxic if inhaled in sufficient concentrations. As the pressure of the air you are breathing rises, there are more and more molecules of oxygen packed into each breath you take. Beyond a depth of 200 ft (60 m), there is a risk that the concentration of oxygen in pressurized air could have a

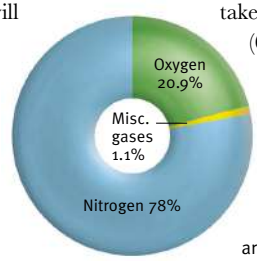

WHAT AIR IS MADE OF
Air is mainly composed of nitrogen, with oxygen forming nearly 21 percent. Inert gases like argon and freon form the remainder.

DEPTH AND AIR CONSUMPTION

Most divers consume approximately 1 cubic foot of air per minute at the surface, though this rate varies between individuals, and is also dependent on how much you exert yourself, and what your mental state is. An anxious or excited diver will consume more air than a calm one.

Because your regulator supplies you with denser air as you go deeper, your air consumption will also increase as you descend. In seawater, your Surface Consumption Rate (SCR) will have doubled by the time you reach a depth of 33 ft (10 m), tripled by 66 ft (20 m), and so on. There are various methods of obtaining an accurate calculation of your SCR, which you will find in your diving manual. Once you know this you will be able to establish how long the air in different sizes of tank (filled to maximum operating pressure) will last at different depths, and plan your dive accordingly.

	Air consumption (cu ft per minute)
Surface	1
33 ft (10 m)	2
66 ft (20 m)	3
100 ft (30 m)	4

toxic effect on the body, leading to convulsions, loss of consciousness, and even death, in extreme cases.

To get around these problems, "technical" divers going beyond the limits of normal recreational diving (*see pp.194–95*) use special gas mixes. These contain altered percentages of oxygen and nitrogen, and also include helium. These mixes are used during dives in excess of 165 ft (50 m).

WASTE GASES

One by-product of the respiration cycle is carbon dioxide. The more you exert yourself, the more carbon dioxide is produced. Since your body's breathing reflex is triggered by a build-up of carbon dioxide in the blood, not by falling oxygen levels, overexertion underwater will make you feel short of breath. On the surface, your body would respond to this by breathing faster, thus flushing out carbon dioxide. Your regulator, while efficient, cannot process the same volume of air as you can breathe on land. Always rest for a moment if you begin to feel breathless.

BREATHING AND BUOYANCY

An object immersed in water is subject to an upward thrust equal to the weight of the water it has displaced. You will only float, therefore, if your weight is less than that of the water you displace. If it is greater, you will sink; if the two are equal, you will do neither—you will have neutral buoyancy. The compressed gas you breathe underwater alters your buoyancy. When you inhale, the volume of your chest gets larger—you displace more water without becoming any heavier, and so become more buoyant.

A diver displays buoyancy control

DEEP BREATH
Slow, even breathing patterns promote low air consumption and help prevent hyperventilation.

Pressure and the body

When you dive, your body is subjected to pressure generated by the weight of the water around you. As you descend, you will feel the increase in pressure in some of your body's internal air spaces, chiefly your ears. If you do not act to correct this, you may be injured.

EQUALIZING PRESSURE

The body contains four major air spaces that are affected by pressure underwater: the ears, the sinuses, the airways and lungs, and the stomach and gut. Air spaces in our respiratory and digestive systems are compressible, in that they are contained by soft tissue that can expand or contract to accommodate pressure changes. However, your regulator (*see pp.60–63*) ensures that the air pressure in your lungs and airway is always in equilibrium with external water pressure, so pressure effects are never felt here. You should barely notice any effect on your stomach and gut, although you may break wind, especially on ascent.

The air spaces in your head, however, such as the ears and sinuses, are rigid and cannot be compressed. When the pressure outside them is greater than that within, a "squeeze" is created, which can cause pain and damage, especially to the eardrum. You must equalize the pressure in these cavities with the ambient pressure of the water around you as you descend. Equalizing is simple, using a procedure that clears the ears called the Valsalva

ROUTINE MANEUVER
Equalize pressure in your ears and sinuses every 3 ft (meter) or so, during your descent to the bottom. Don't wait until your ears begin to hurt.

THE IMPORTANCE OF EQUALIZATION

The ear is divided into the outer, middle, and inner ear. The eardrum separates the middle and inner ear from the outside world, and it is this membrane that is most vulnerable to pressure changes. If you do not equalize as you descend to raise the pressure behind your eardrum—by sending air to the ear from your throat via your eustachian tubes—then the eardrum will eventually rupture and let water into your middle ear, leading to possible infection and deafness. Your sinuses are pairs of cavities within the bones of your face. Under normal circumstances they are open, via your nasal passages, but if you have a cold, or congestion caused by allergies, they fill with mucus and become blocked. Failure to equalize will cause blood vessels in the lining of the sinuses to rupture and bleed into the sinuses, which can be quite painful.

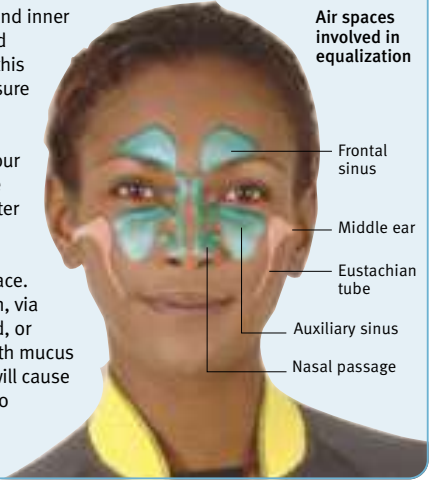

maneuver. You just have to hold your nose and breathe out against your nostrils. This will send air from your throat through the eustachian tubes into the middle ear so that the pressure there becomes equal to that of the water outside, thus easing the discomfort. This maneuver should also equalize the pressure in your sinuses, and in addition resolve any problems of dizziness or vertigo you may be experiencing (as balance is also controlled within your ears). Some people get the same result by simply swallowing repeatedly.

You need to equalize regularly on descent, but on ascent equalization should occur automatically. If you have problems equalizing on the way down, ascend a little to reduce the pressure slightly, and try again. This equalization process is vital for comfort when diving; never dive with ear plugs or with anything tight over your ears. Check that your hood does not fit too tightly, as this can hinder efficient equalization. If you have any problems with your sinuses or ears, do not attempt to dive, since this may make them worse. A cold, or any nasal congestion, may cause blockages in your airways and you will not be able to clear your ears and sinuses. If you notice a slight nosebleed after a dive, this is a common sign that you have a mild sinus blockage and you should not dive again until it has cleared up. Decongestants will not necessarily help once you are at depth, and you should not dive while you are taking them.

SAFE DESCENT
To adapt safely to the underwater world, you must understand how depth affects the human frame, especially internal cavities.

Nitrogen and decompression

When you dive using scuba gear, nitrogen gradually accumulates in your body, and is released when you ascend. In order for this to happen safely, you must plan your dive carefully and employ diving procedures that minimize the risk of decompression sickness (DCS).

GAS ABSORPTION

Gases are soluble in liquids, and dissolve more easily when they are pressurized. When you breathe compressed air underwater, the two main gases it is composed of (nitrogen and oxygen) enter and dissolve in your bloodstream. This "gas exchange" takes place in the lungs, via tiny saclike vessels called alveoli.

Nitrogen is particularly easily absorbed into the blood and body tissues. Your regulator supplies denser and denser air as depth increases (*see also p.60*), which means that every breath you take is loaded with more nitrogen molecules than an equivalent breath at the surface. This surplus of nitrogen rapidly accumulates in the tissues of your body. If you ascend too quickly, the nitrogen in your tissues can reenter the blood in the form of harmful bubbles. It is therefore very important for you to manage your dive time and depth,

BOTTOM TIME
The longer you spend at depth, the more nitrogen builds up in your body.

NITROGEN AND BODY TISSUES

When you breathe compressed air underwater, its nitrogen content enters the blood and begins to penetrate the body's tissues. When you ascend, ambient pressure drops and nitrogen starts to come out of the tissues again. If your ascent is slow and controlled, the nitrogen reenters the blood and exits the body harmlessly, via your lungs. If your ascent is too fast, the escaping nitrogen can form harmful bubbles, leading to decompression sickness (DCS; *see pp.142–43*). DCS can cause minor discomfort if the bubbles are lodged under the skin or in the joints, but if they are in the brain or nervous system, it can lead to paralysis or even death.

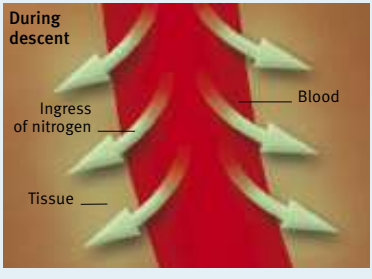

NITROGEN ENTERS BODY TISSUES
Pressurized air dissolves in blood and its nitrogen content enters tissues, especially fatty areas, where it is temporarily stored.

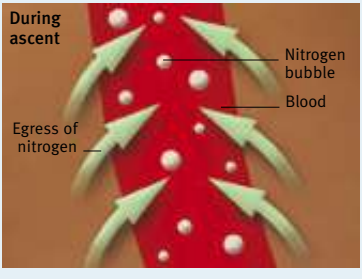

FORMATION OF NITROGEN BUBBLES
If pressure drops too quickly during ascent, nitrogen forms bubbles in the bloodstream, which can lead to decompression sickness.

PERFECT PERCH
Dive boats sometimes lower a "trapeze" for divers to hold on to while making decompression stops, with a spare air tank attached for safety.

using either dive tables or a dive computer, to ensure that nitrogen does not build up to levels that cannot be safely dissipated during ascent.

Dive tables help you to calculate how long you can safely dive at a given depth, and what the safe limits are for any following dives that day (excess nitrogen takes 24 hours to fully leave the body). The tables will also tell you if you need to perform a decompression stop. This is a scheduled pause in your ascent that allows nitrogen to dissipate safely. It is a good idea to stop at 16 ft (5 m) for three minutes during ascent, even if your dive tables do not require it, just as a precautionary measure. This is called a "safety stop." Dive computers also generate safe diving schedules, by making continuous calculations during the dive (*see pp.66–67*).

NITROGEN NARCOSIS

The nitrogen in compressed air can also have a narcotic effect at depth, ranging from a feeling of euphoria to anxiety. Nitrogen narcosis rarely causes death directly but at depths below 30m (100ft) it causes divers to make increasingly bad judgments. If you feel the effects of narcosis during a dive, ascend to a shallower depth and the narcosis should rapidly disappear. Keep an eye on your buddy for symptoms of narcosis, such as sluggish responses and irrational behaviour.

USING DIVE TABLES
An understanding of the use of dive tables is a key element in all dive training programs.

Tides and currents

The tides have a huge impact on the marine life and conditions of many dive sites, and localized currents can create unseen offshore hazards. Knowing how tides and currents work and how they affect diving procedures is essential for safe diving at any location.

TIDAL MOVEMENT

Water levels at the shoreline can vary considerably over the tidal cycle, in extreme cases by as much as 50 ft (15 m). The region between the highest and lowest tide marks is called the intertidal zone.

Make a thorough check of local tidal conditions and coastal topography when planning a dive, otherwise you may encounter unexpected hazards, like areas of strong tidal flow. Water clarity varies from site to site, but it is often best at high tide, while incoming tides tend to bring marine species closer in to shore with them. The most important tidal event for divers is "slack water," a short interval when tidal flow halts at the changeover of every tide. Tidal currents mean that some sites are only safe for diving during the slack water period.

INTERTIDAL ZONE
Sometimes submerged and sometimes exposed, this is a harsh habitat for marine life.

HOW TIDES ARE FORMED

The Earth's seas are subject to the gravitational pull of both the Sun and the Moon. Although the Moon is smaller than the Sun, it is much closer to the Earth, so its influence is stronger. The gravitational pull of the Moon draws water into a bulge on the side of the planet closest to it. A counter-bulge forms on the opposite side of the Earth because the water there is pulled less toward the Moon than the planet is itself. As the Earth rotates, the two opposing bulges of water move across the face of the planet, creating two high tides and two low tides in just over 24 hours.

NEAP TIDES: FIRST AND THIRD QUARTER MOONS
With the Sun and Moon at right angles, high tides are at their lowest, as they are subject to the gravitational pull of the Moon alone, but low tides are at their highest, pulled by the Sun.

SPRING TIDES: NEW AND FULL MOONS
Twice a month, the Sun and Moon are in line with the Earth. Their combined gravitational pull creates the highest high tides, and low tides are at their lowest.

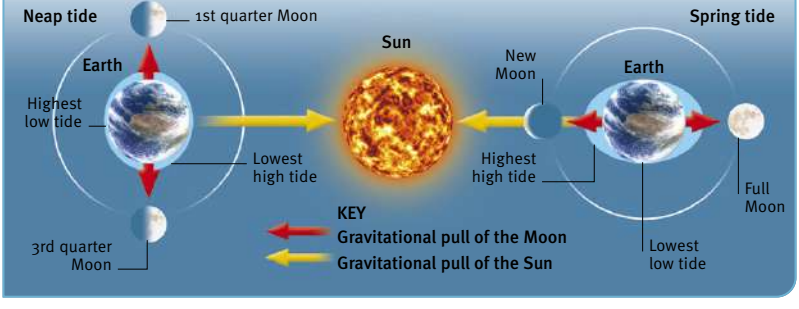

TYPES OF CURRENTS

Currents are created by winds, tides, the rotation of the Earth, and the confluence of bodies of water of differing salinity and temperature. Tidal currents occur when local topography, such as a narrow channel, creates an area of unusually strong flow when the tide is coming in (flooding) or going out (ebbing). These currents can be powerful enough to carry away even strong swimmers.

DIVERS EXITING SEA
Plan your entrance and exit so that you avoid swimming against the tide. The best time to dive is usually at "slack tide"—at high or low tide, when the tidal pull is weakest.

Standing currents are permanent currents independent of tides. Longshore currents, which flow in one direction along the coast, are localized standing currents. There are also standing currents in the open ocean, such as the Gulf Stream, which flows northward from the Gulf of Mexico, driven by the prevailing southwest winds, to the mid-North Atlantic, where it divides into two lesser currents.

Rip currents occur when incoming waves force the backwash from previous waves sideways. The backwash streams along the shore until it finds a route back to the open sea, such as a gap in a sandbar, and surges out in a strong, narrow current, perpendicular to the shore. To escape a rip current, swim across the current, parallel to the shore. You may find that trying to swim out of a rip current leads you into a longshore current, which pushes you back into the rip current. If this happens, swim out of the rip current in the opposite direction.

WAVE SURGE
Underwater topography and weather conditions can combine to create dangerously strong water movement where surf breaks on the shore.

Diving skills

Introduction

Good basic skills are key to being a competent diver. Some techniques, such as buoyancy control, may seem difficult to master at first, but with practice they will become second nature, allowing you to relax and enjoy your dives. Skilled divers are also safer divers.

SUPPLEMENTING YOUR TRAINING

This chapter of the book discusses diving skills, but is not intended to be, or in any way replace, a training manual. While it covers the skills and techniques that are taught in the basic dive training offered by approved agencies, it must not be used in place of any element of that training. There is no substitute for comprehensive dive training with a certified instructor.

In this chapter, therefore, we outline the skills you need in order to be a safe and competent diver. Many people only dive once a year on vacation, or even less frequently, so it is very important to keep up to date with your diving skills. If you have been away from diving for any length of time, take a refresher course or an easy trial dive.

STEPPING OUT
Knowing how to enter the water properly will help you get your dive off to a good start.

USING A COMPASS
Don't forget to practice navigation skills regularly. Getting lost can ruin a dive, and may hinder your return to a safe exit point.

You will find that most of your basic skills are still familiar to you after a break, or will come back in due course. However, if you are put in a stressful situation, you may not be able to react quickly and effectively if your skills are rusty. Even experienced divers can benefit from regular practice of basic skills, and you should always be thinking about how to improve your equipment configuration and technique.

KEEPING SKILLS CURRENT

A great way of keeping your skills up to date is to practice in a pool (*see pp.146–49*). This might not be as much fun as getting out into the open water, but it will certainly provide you with a safe environment in which to perfect key skills such as buoyancy. Regular training in a pool or sheltered water can help you to meet the challenges of open-water diving more confidently. If your local waters are cold and stormy in winter, use the off-season to brush up drills and procedures. Joining a diving club may give you access to a nearby pool.

DEPLOYING A BUOY
Diving requires the use of many items of special equipment that you must learn how to use safely.

Pre-dive skills

Good organizational skills will help ensure a trouble-free diving trip. A successful dive begins before you even leave your home, with checks on weather conditions and equipment. Even if someone else is organizing the dive, make sure you are well prepared and informed.

GETTING OFF TO A GOOD START

On the morning of the dive, get a detailed, up-to-date weather forecast specific to the site you are diving, and double-check tide times if appropriate. If you don't feel well for any reason, make an honest assessment of whether you are fit to dive that day. It is better to sit out a dive than to feel ill underwater and have to abort. Illnesses affecting lung or circulatory efficiency also increase the risk of decompression sickness (*see p.142*)—a very good reason for skipping the dive.

BRIEFING FROM DIVE LEADER TO GROUP
While a group leader will take overall charge of the dive plan, you must take responsibility for your own fitness and equipment readiness.

Always check your gear before you set off on a diving trip. Before you leave home, you must be sure that you have everything you need. It is good practice to make a checklist of all the items you will need, and keep this in your bag. When you arrive, be sure to transfer all your gear from the initial mustering point to the actual dive site, or onto the boat you are diving from. All too often, divers leave a vital piece of

DIVING WITH A NEW BUDDY

Diving abroad often means that you will dive with strangers. Make sure you give the dive organizer an honest assessment of your level of ability and interests, so they can pair you with a suitable buddy. Even if you don't know your buddy, give them all the respect and assistance you would offer a familiar dive partner.

gear on the dock, such as a weight belt or mask, and only realize their mistake when the dive boat reaches its destination—by which time it is too late.

DIVING FROM A BOAT

The boat should have been booked in advance, but even so, it is polite to call the skipper again before you arrive. He or she will be able to confirm that weather and conditions are favorable for diving. When you board, check that the boat has adequate first aid and safety signaling devices on board, and appears to be in a seaworthy condition. Before you leave shore, notify the coastguard or relevant authority—and do not forget to inform them of your safe return when you get back. Ensure that the skipper reviews safety procedures, including the location of lifejackets, flares, and life rafts, and demonstrates the underwater recall signal.

COMFORT AFLOAT
Remember to bring seasickness tablets with you on boat trips. Test them 24 hours beforehand to make sure they have no adverse effect on you.

GETTING READY TO DIVE
If you have organized your day efficiently, you will be able to relax and concentrate on getting yourself ready to dive.

DIVING FROM THE SHORE

If you are diving from the shore, always check for the easiest possible access to your dive site on arrival (or before) to avoid long walks down to the beach with heavy gear. You may need permission to drive to the water's edge to unload.

WATCHING THE WEATHER

Obtain a forecast that includes the wind speed and direction, sea state, and visibility (fog, mist, etc). Do not take any risks with diving in bad weather—it is better to miss the dive than dive in surging seas or in visibility so poor your boat cover will not be able to locate you. You should also check tide tables to establish the timings of slack water. This is critical when diving areas affected by strong tidal currents.

Rough seas

Final checks

Diving with a partner (or buddy) means that there is someone to help if you encounter problems underwater. For the system to work properly, however, you and your buddy need to conduct your own briefing before a dive and check that all equipment is functioning properly.

MAKING BUDDY CHECKS

Just before the start of the dive, get together with your buddy and assemble your gear (*see pp.84–85*) and put it on. There is no particular order in which to do this, but most divers find that to avoid overlooking anything, it is useful to develop a routine. When you are both suited up and before either of you enter the water, you should carry out a buddy check—sit or stand next to your buddy and carefully check each other's gear, following the sequence shown below. Do not be tempted to rush these vital checks—you may regret it later. They ensure that you know how each other's gear is assembled, how it works, and that it is functioning. They also serve as a double-check that neither of you has overlooked anything before you dive.

1 **Check BCs** so that you and your buddy know where each other's inflation and deflation points are, and ensure that they are working. Do the same for drysuits, if worn.

2 **Check that weights** are present and securely fastened. You and your buddy must be especially aware of how each other's weights are released, in case either diver is incapacitated.

3 **Check harness is secure** and note where, on your buddy's kit, key fastening points and harness release clips are located, and how they are operated.

4 **Check air contents** gauges and breathe from your regulators to check they are working. Test each other's octopus second stage.

BUDDY BRIEFING

Suiting up provides a good opportunity to talk over a dive plan with your buddy. Ensure first that you are both agreed, as a pair, on the aim and course of the dive, your entry and exit points, and your predicted maximum depth and time for the dive. Check that you both have the required amount of air for your plan (including a reserve for emergencies). Agree on who will lead the dive and on whether you will dive to the left or right of your buddy. Decide on all communication signals, including how and when you will signal for the end of the dive, and on what you will do if you become separated. Once you have agreed a plan, stick to it unless it becomes impossible to do so. If circumstances change during the dive, use hand signals to discuss how, as a pair, you are going to modify the dive.

5 **Ready to dive?** Give each other a last once-over to establish who is carrying any miscellaneous pieces of gear, such as reels and slates, and where they are fastened. When you are ready to dive, make a final OK signal.

PREVENTING FOGGING

Mask fogging is a very common inconvenience, and is caused by oils on the mask's lens allowing moisture to bead. The traditional way to prevent fogging is to rub saliva onto the inside of the lens glass, then lightly rinse clear, before putting on the mask for a dive. Anti-fogging sprays are also available.

Using an anti-fogging spray

Entry

Every site will require a different method of entry, and the easiest method possible at each site is the best one to use. In all cases, it is safer to enter without large, heavy equipment, such as cameras, so if conditions allow, have them passed to you once you are in the water.

SHORE ENTRY

When entering the water from a beach, make sure your gear is assembled and on your back, then carry your mask and fins to the water's edge. There, put on your fins and mask with your buddy's help and walk backward into the water (walking forward with your fins on is difficult). As soon as you are deep enough, lower your center of gravity by bending down, then start to swim, or "fin," out into open water. If there is surf, make sure your BC is partially inflated and wait until a big wave approaches; lean into it, and let the return surge take you away from shore.

ENTERING FROM THE SHORE
If entering through surf, be careful not to lose your fins as you put them on. Breathe from your regulator throughout entry.

CONTROLLED SEATED ENTRY

This is the easiest method of entry into a pool or from a platform just above water level, and an especially useful method if you have difficulty with mobility on land or with keeping your balance when standing on a floating platform. It also enables you to put on your gear while seated.

1 **Turn from the waist** to place both hands on the hard surface to one side of your body, ready to take your weight.

2 **Pivot your body around,** using your hands to support you, so that you turn to face the side while easing yourself into the water.

GIANT STRIDE ENTRY

Stepping into the water is a good entry method to use from a large diving boat or off a dock or platform, provided there is sufficient depth of water. Partially inflate your BC, place your regulator in your mouth and hold it and your mask in place with one hand, and take a large stride out into the water.

1 **Take a giant stride,** holding your mask and regulator in place. Keep looking forward, not down.

2 **On splashdown** the splayed-scissor position of your legs will cushion the impact and keep you at the surface.

BACKWARD ROLL

Use this entry from small boats with a low freeboard (distance from deck to waterline). With your regulator in place, check that your fins are free from obstruction, and that the water behind you is clear. Ease yourself over the edge and roll over backwards, aiming to hit the water first with your upper back.

ENTRY POSITION

Hold your mask and regulator in place with one hand and ease yourself backward until your tank is hanging over the water.

BUDDY PROCEDURE AND SIGNALS

After entering the water, meet on the surface and make an OK signal (hand on head or fingertips together) to shore or boat cover. One exception to this is when making a negatively buoyant entry through strong surface currents. In this case, descend as quickly as possible, and rejoin your buddy on the seabed.

Standard "OK" signals

Descent

Descending to the seabed is always an exciting experience, but while you may want to get there quickly, you must always descend in a controlled manner. The method you use will depend partly on the site and prevailing conditions, and partly on your own preferences.

MAKING YOUR DESCENT

Having met up with your dive buddy in the water, follow the preparatory steps below. If you find that you do not sink, check that no air remains in your BC (and drysuit, if applicable). If the problem persists, you probably do not have sufficient weight on your belt (*see pp.70–71*), and should return to the boat or the shore to get some more. It is a good idea to note how much weight you need each time you try a different equipment configuration—record the details in your logbook for reference.

At a depth of 10 ft (3 m) you should carry out a bubble check. This involves briefly stopping so that you and your buddy can check each other's equipment for signs of air leaks—for example, from an incorrectly fitted hose. This will allow you to return to the surface to resolve the issue before you resume the dive. By fixing minor problems now, you may avoid bigger problems during the dive that could lead to an emergency.

As you descend, you will feel pressure in your ears. Release this regularly by swallowing or holding your nose and

1 **Meet up with your buddy** on the surface, well clear of any boat cover. If using a snorkel, remove it, and switch to breathing from primary regulators.

2 **Give an OK signal** to each other when ready to begin the dive. If using a "buddy line" to link yourselves together (useful when one buddy is a novice), ensure now that you are both attached.

3 **The "down" signal** confirms your intention to descend immediately. If your buddy is a novice or nervous, reassure them by holding their hand and helping them to descend.

4 **Both you and your buddy** should deflate your BCs and exhale together, so that you become negatively buoyant and start to sink simultaneously. If you have a problem sinking, address it now.

blowing against the closed nostrils—a process called "equalization" or "ear-clearing" (*see also pp.98–99*). If you feel you are descending too fast, allow a little air into your BC. Your mask will also start to press on your face; relieve this by exhaling gently through your nose. If you are wearing a drysuit, you will feel the water pressing it against your body. Ease this by letting air into the suit, but not so much that it alters your buoyancy too greatly—and remember to release it again on ascent.

EQUALIZATION
Pinch your nose and blow at intervals to equalize pressure in the skull with that of the water.

BUBBLE CHECK
Buddy pairs check each other's gear for telltale air leaks before descending further.

DESCENT OPTIONS

Many divers prefer to descend feet-first so that they can remain face to face with their buddy and maintain eye contact. This is good practice, unless you need to descend quickly through a current and therefore have to use your fins.

A free descent is where no line is used and you descend freely (head- or feet-first) to the seabed. If you are using an SMB (*see pp.74–75*), ensure before descent that the buoy is fully inflated and the line is free. Your buddy should stay on the same side throughout the descent to avoid confusion and tangling in the line. On group dives there may already be a weighted guideline in place, attached to a buoy and leading to a wreck or fixed point on the seabed, which will give you and your buddy a handy point of reference. Do not use the line to pull yourself down, since it may not be anchored securely at the base.

FINAL CHECK
A group of divers meeting at the bottom of a guideline give a last "OK" signal to each other before exploring the dive area.

Breathing underwater

Swimming freely underwater for the very first time can be exhilarating, but also potentially unnerving and even frightening. Learning and perfecting good breathing techniques allows the novice diver to become calm and relaxed in this new environment.

LEARNING TO RELAX

The sheer thrill of your very first dive will cause your heart rate to increase and make you breathe faster than normal. You may be alarmed at how quickly your air supply runs out, but after a few dives, you will learn to be calmer and more relaxed, and your air consumption will decrease. When you feel comfortable underwater, take a few moments to concentrate on your breathing and remind yourself of the dos and don'ts of underwater breathing.

CONSERVING AIR
The more relaxed you are, the less air you will use and the longer your dive will last.

First, and most important, never hold your breath underwater, because you will run the risk of bursting a lung. Inhale and exhale more deeply than you would normally do on land. Try to maintain neutral buoyancy (*see pp.118–21*), since this will decrease your need for leg and arm movements to maintain a constant position, and thus reduce air consumption. Furthermore, avoid

BREATHING OUT
If you remove your regulator, exhale slowly to minimize the risk of lung expansion injury.

CLEARING THE REGULATOR
A regulator can be cleared of water by simply exhaling, or by pressing the button on the front to purge the water.

FREQUENT CHECKS
Buddies should regularly check each other's air reserves, to maintain a good awareness of the remaining levels of both members of the team.

swimming unnecessarily against currents—it is better to go with the flow than to waste your air by overexertion.

Always be aware of where your buddy is and what he or she is doing; many novice divers lose sight of their buddy and rapidly become anxious. If there is a problem, stop, breathe, and think before taking action. Acting too quickly and without thought could lead to panic and further problems.

Make sure that you are familiar with your equipment and fully understand how it works. You must know exactly where each item is stored to avoid panic if things go wrong during a dive—for example, if you lose your regulator or your mask begins to fill up. Regularly practice emergency procedure drills (*see pp.136–43*), either in the pool or in open water, to ensure confidence in a real crisis.

PLANNING YOUR AIR CONSUMPTION

On an open-water, fairly shallow dive, the rule of air consumption is to plan your dive so that you can carry out your descent, exploration, and ascent, and still surface with a reserve of 725 psi (50 bar). If you have to come back to a given point to exit, such as a guideline or anchor line, start your return when you have 1,500 psi (100 bar) left. If there is a current, swim out against it and swim back with it, so that the return is never longer than the outward trip. It is also wise to swim back to your point of entry or shotline at a shallower depth than on the outward journey, since this reduces air consumption. In difficult environments, follow the "Rule of Thirds" (*below*).

THE RULE OF THIRDS

When planning a dive in an environment that is covered overhead—such as a wreck, cave, or under ice—or when making deeper dives, always apply the Rule of Thirds. The rule states that you should use a third of your air for descent and exploration, and a third returning to your guideline or point of entry and for ascent. The remaining third is an emergency reserve.

Buoyancy control

Without doubt, the most important skill for a diver to master is that of buoyancy control. Until you achieve this, you will be constantly expending effort in order to maintain your position in the water, wasting air resources and distracting you from your dive.

SUITS AND WEIGHTS

Wearing any type of exposure suit will make you more buoyant in the water, and thus make it more difficult to dive. Drysuits contain an airspace between the suit fabric and the diver's body. Wetsuits have air cells within their neoprene fabric, and so the thicker the neoprene, the more buoyant you will be. The function of weights is to counteract the buoyancy your suit gives you so that, with your lungs full of air but your BC uninflated, you float at the correct level at the water's surface.

Correct assessment of how much weight you need to carry for a given dive is very important. Spend time conducting weight tests for different suit and equipment combinations, and note your findings for future use. Remember, saltwater is denser than fresh water, so you will require more weight when you dive in the sea than in fresh water.

USING YOUR BC

At the start of a dive, you will usually enter the water with your BC inflated, as having positive buoyancy allows you

Too high

Too low

CONDUCTING A WEIGHT TEST
Empty your BC, and take a deep breath and hold it. Float in an upright position. If you are wearing the right amount of weight, you will float at eye-level to the waterline. Add or discard weights as necessary until this is achieved.

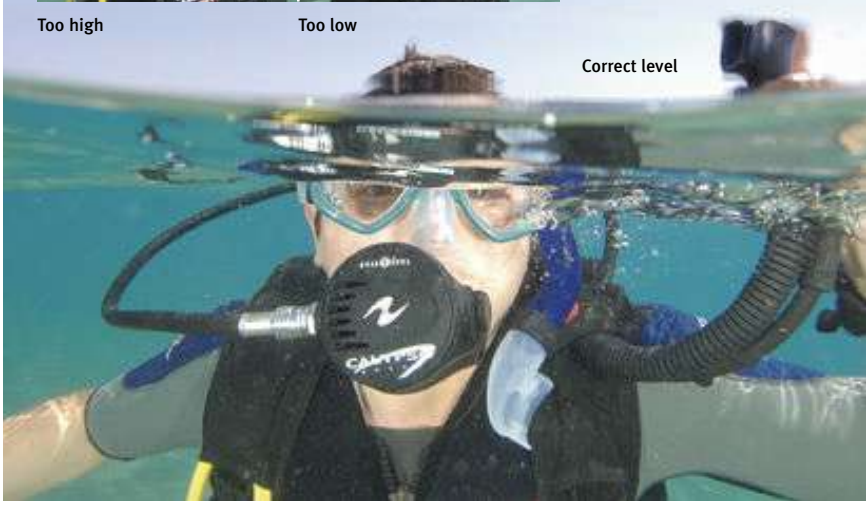

Correct level

to float at the surface effortlessly while you meet up with your buddy—the only exception to this is at sites where there are strong surface currents, when it is important to get to the bottom quickly.

In order to descend, you must be negatively buoyant, which you achieve by deflating your BC and exhaling (*see pp.114–15*). However, as you sink, you will find that your descent rate accelerates as your suit—and the remaining air in your BC—becomes compressed. To counteract this, you must add air to your BC. If this halts you before you have reached the desired level, let air out. Repeat the process, letting small amounts of air in and out of your BC until you become stationary at the desired depth.

When this happens, you will have achieved neutral buoyancy underwater—a state in which you neither float nor sink. Achieving this perfect equilibrium of weight ballast and lifting forces means you do not waste effort and air using your fins to maintain a certain depth.

LEARNING THE BASICS
An instructor looks on as a student diver practices fin pivots (*see pp.146–49*), an exercise designed to improve basic buoyancy control.

When you want to ascend, the same control principles apply, but in reverse. As you go up, the air in your suit and BC expands as external water pressure decreases, making you increasingly buoyant, and causing you to rise rapidly. As it is important to ascend at a slow,

HOVERING OVER A REEF
Buoyancy control is critical when diving over delicate coral that can be damaged by physical contact.

steady rate (*see pp.128–31*), you'll need to vent air accordingly as you rise. To descend and ascend as smoothly as possible, you must become familiar with how your BC responds to inputs of air, and how quickly it vents air.

Drysuits have independent inflation and deflation controls (*see pp.54–55*). If you are diving in a drysuit, the air space between your suit and your body will become compressed, and you will need to inject a little air into the suit to avoid discomfort from being "shrink-wrapped" by external water pressure. This will make you more buoyant, so you will need to vent air from your BC to counteract it. On ascent, you will need to dump the excess air from your drysuit.

FINE CONTROL OF BUOYANCY

Small movements up and down—to look at different parts of a reef, for example—are best achieved through breath control, rather than letting air in and out of your BC. Good buoyancy control only comes with practice, and is worth perfecting. Take time at the start of each dive to practice fine control of buoyancy before you set off exploring. From a neutrally buoyant position on the seabed, slowly inhale—you should rise up and hover just above the seabed. Breathe out a little—you should sink. Practice breathing in and out and become familiar with how it influences your gentle rise and fall. You should be able to hover over the seabed and control your buoyancy by breathing alone, without constantly filling and

JETTISONING YOUR WEIGHTS

In extreme circumstances, such as being caught in a powerful downcurrent, it may be necessary—as a last resort only—to dump your weight belt. You will instantly start to rise up rapidly through the water, so be ready to vent air from your BC to slow your ascent. Breathe out on the way up, to minimize the risk of lung expansion injury (*see p.95*).

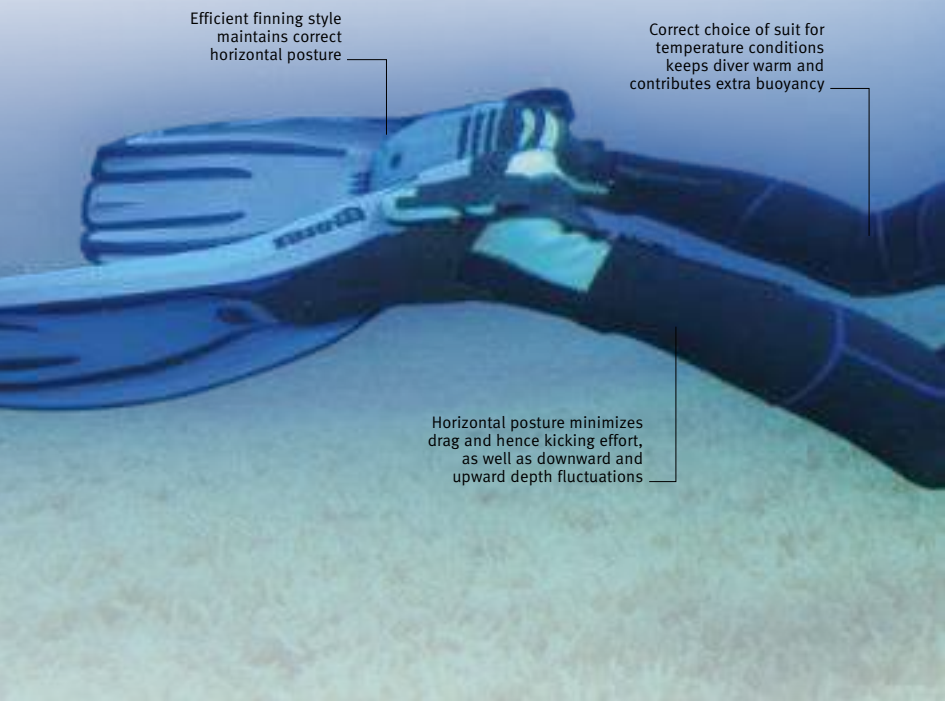

emptying your BC. Once you have mastered this, you will be able to float at any depth without effort.

TRIM AND WEIGHT DISTRIBUTION

The position of your weights and equipment has an important part to play in maintaining good "trim" (correct posture and balance) underwater. Distribution of weight is key in achieving a good horizontal attitude, which makes buoyancy control easier and helps you to propel yourself along with the minimum of drag. A streamlined profile also helps posture and ease of movement, which helps to reduce air consumption. Experiment with different weight systems and stowage points until you find a comfortable combination that helps you achieve the best position in the water. Some designs of BC allow you to stow weight at different points, not just on a belt. For example, some have special quick-release pouches that are designed to take moderate weight loads (*see p.59*). Always check that you are able to jettison the bulk of your weight load in case of emergency.

ACHIEVING PERFECT BUOYANCY
Thinking about buoyancy as a system of influences to be harmonized, rather than just weights and BC, will help you perfect the art of moving underwater.

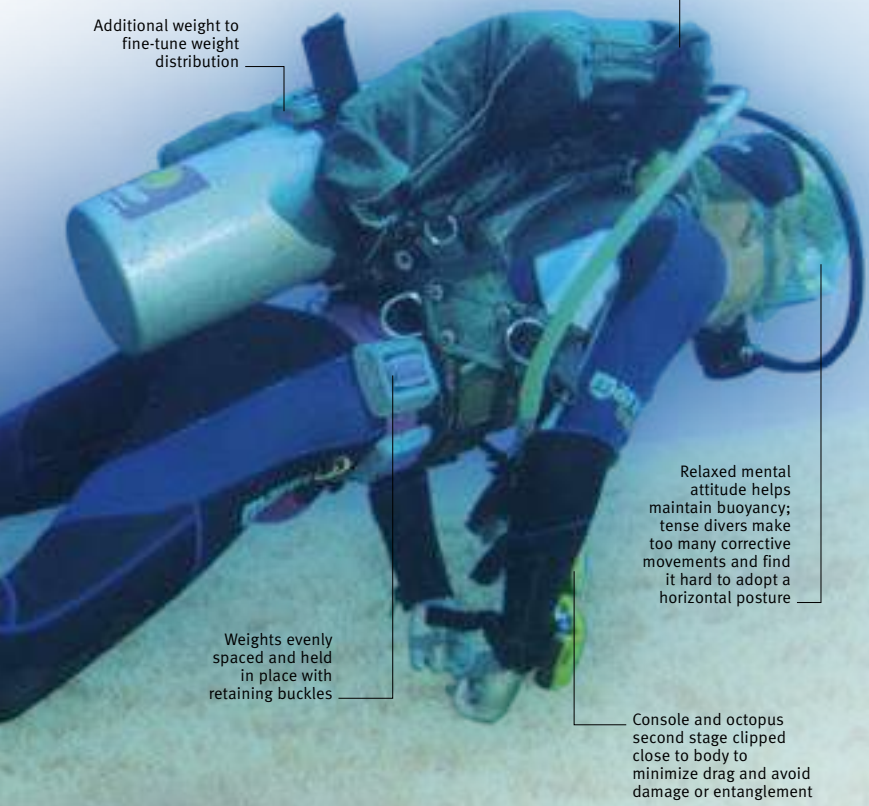

Moving underwater

The duration of your dive is limited by your rate of air consumption, which in turn depends on how much energy you expend. Good finning technique, buoyancy control, and a streamlined profile will all make your movement underwater more energy-efficient.

EFFICIENT PROPULSION

Your movement through the water should be driven and controlled by your legs. If you watch experienced divers, you will notice that they use their arms very little, and move effortlessly through the water, saving energy, minimizing air consumption, and avoiding damage to reefs and the marine environment.

Swimming with fins should be an extremely efficient way of propelling yourself through the water, but many divers are let down by a poor finning technique. There are several different strokes you can use, but all should be carried out with strong, rhythmic movements, fully extending the leg (or legs) with each kick. Uncontrolled thrashing and incomplete kicks will simply waste energy. It is worth asking your buddy to give you feedback on your technique—even to photograph or film you. You may be surprised at the results.

Scuba diving is not a race, and the faster you go, the less you see. Learning to pace yourself and move at a leisurely speed will help you to relax, and a relaxed diver is an aware, efficient, and safe one.

MINIMIZING ARM USE
To practice with your legs only, cross your arms in front of you or let them hang by your side.

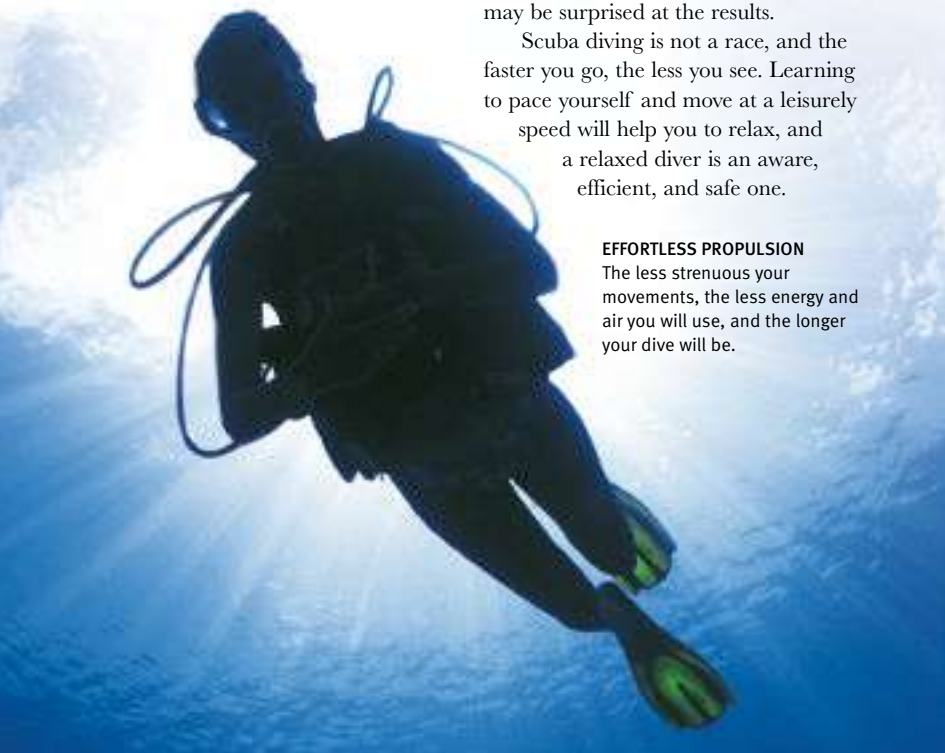

EFFORTLESS PROPULSION
The less strenuous your movements, the less energy and air you will use, and the longer your dive will be.

PROFILE AND BUOYANCY

To achieve the most streamlined (and therefore most energy-efficient) swimming position, you should be almost horizontal in the water. Your hips will be slightly lower than your upper body due to that area being weighted. During a dive, you should look forward and downward most of the time—raising your head too much will lift your upper body and make your profile less streamlined.

POOR PROFILE
Swimming with the wrong profile may mean that your fins kick up silt, impairing visibility.

Carrying the correct amount of weight is important. Being overweighted will drag your hips and lower body downward. Swimming will then require greater effort due to your unnatural angle in the water, which will produce increased drag. You can improve your profile by repositioning the weights in your BC, or refining your weight requirements (*see pp.118–21*).

Achieving neutral buoyancy (*see pp.118–21*) is the key to maintaining a good dive profile with ease, and also increases your efficiency. If you are positively buoyant, you will be constantly swimming downward; if negatively buoyant, you will have to swim upward all the time or crawl along the bottom—all of which require more effort than horizontal, neutrally buoyant, swimming.

TYPES OF KICKING STROKES

With the flutter kick (*below*), all fin movement takes place above the body's centerline, reducing the amount of water directed downwards. This can be useful when you wish to avoid stirring up sediment on the seafloor. The frog kick is also ideal for use where fine silt can ruin visibility if disturbed, because it directs thrust horizontally behind you. The swimmer kick is good for power, but uses up energy quickly. Most divers tend to drift slightly when using the flutter kick or the swimmer's kick, because one leg is slightly stronger than the other.

FROG KICK
Spread your legs and bring up your knees. Rotate your fins so that the bottoms of the blades face behind you. Bring the fins together in an arc, thrusting out behind.

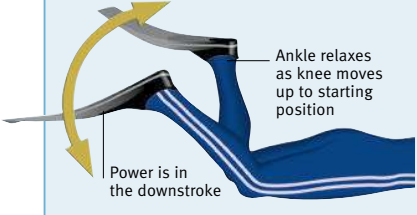

FLUTTER KICK
Bend the knees slightly, shorten the stroke of the kick, and keep the hips fairly static. The stroke ends when the knee is straight.

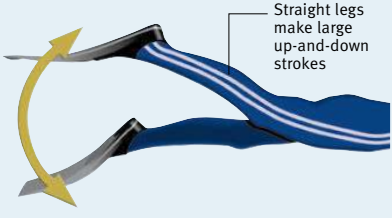

SWIMMER KICK
Lock your legs straight and kick from the hips with long strokes. This technique exploits the power in your thigh and hip muscles.

Exploring and navigating

Learning how to use instruments and natural features to find your way around underwater is important. Accurate navigation will help you to reach your planned exit point at the end of a dive, and will also help you find and explore features of interest, such as wrecks.

MAPPING THE SITE

It is easy to become disoriented in poor visibility, but good navigation skills will help you find your way, even when conditions are against you. Developing such skills will make you less dependent on other divers for your own safety—if you always rely on your buddy to lead the way, you could find yourself in trouble if you become separated.

On first entering the water, make mental notes of the surrounding features, such as unusual rocks, coral, or your position on a wreck site. As you progress through the dive, try to build up a map of the site in your mind. At every turning point, note nearby features and which direction you turned in relation to them.

FOLLOWING A REEF WALL
In clear conditions, a reef wall makes an easy navigational reference point. Try to remember distinctive features as you go along.

MENTALLY MAPPING THE DIVE SITE
Reef, rock, and seabed features, sunlight, and even shoals of fish can all help you to orient yourself at the dive site.

Look at the seabed geology, too: rocks often have features or show strata that make memorable visual markers. It may be helpful to jot notes down on a slate.

NATURAL NAVIGATION AIDS

A compass (*see pp.126–27*) gives the most accurate directional information, but there are also clues in nature. If the water is clear and shallow enough, or if there are any shadows cast by rocks, you can use the Sun as a directional reference. On a morning dive, for example, the Sun will be in the east. Take note of which way the current is running. (If you are diving close to slack water (*see pp.102–03*), at high or low tide, this may change by as much as 180° as the tide begins to turn.) Even if you cannot feel the current, you can find visual clues to its direction. Plankton, for example, drift along with the current, so shoals of fish station themselves facing into the current so that they can feed on them. Kelp fronds flow in the direction of the current. Exhaust bubbles will rise up toward the surface (unless you are in a down current).

READING SAND RIPPLES

Sand ripples run parallel to the shore, at 90 degrees to the path of incoming and outgoing tidal currents. Sand ripples get closer together nearer to shore, providing a useful reference point for underwater navigation. If the spacing of sand ripples underneath you appears to be getting narrower, you are swimming toward shallow water and a possible exit point.

Sand ripples on the seabed

FOLLOWING A REEF

On a basic reef dive, swim out with the reef wall on one side of your body and then return with the wall on your other side. When diving around a wreck, try to maintain either a clockwise or counter-clockwise motion, especially if the wreckage is fragmented. Keep the main part of the wreck to one side of your body during the dive, so that you return to your starting point after one circuit.

USING A HANDHELD COMPASS
In case you are separated for any reason, both you and your buddy should carry and be confident at using a magnetic compass.

USING A COMPASS

A compass has a circular body, often divided into 360-degree increments, and a needle that swivels to point to magnetic north. Many also have a "bezel"—an outer ring, also marked with increments and the points of the compass, that can be rotated independently to the body. This can be used in conjunction with the "direction of travel" arrow (known as the "lubber line"), which is marked on the compass body, to establish the direction you want to travel in relation to magnetic north. Point the lubber line in the direction in which you wish to travel, then turn the bezel so that its "north" mark matches the direction in which the needle is pointing. If you ensure that the needle always points to the north mark, as you move in the direction indicated by the lubber line you will maintain your intended direction. Digital compasses work in the same way, but display the information on an LCD.

A compass is a fairly basic device, but it is easy to use it incorrectly. If it is a wrist-mounted model, ensure that your

MEASURING DISTANCE
When navigating underwater, you can judge relative distances by counting how many fin strokes you have made in a given direction, or by timing your progress along a course.

wrist is held square to the direction of travel, and not at an angle. It is simpler to use a compass mounted on a console or a retractable lanyard, which can be held out straight in front of you. Use both compass readings and environmental observations, so that you can correct any inaccuracies in either method. Mark readings down on a slate as you go, and try to integrate these into your mental map of the dive site. In slack water, timing how long it takes to get from one point to the next gives you an idea of your position if you return along the same compass course. Combining this information with depth readings taken along the way will help you navigate with more accuracy.

LAYING A DISTANCE LINE

In low-visibility environments—such as murky temperate waters—it may be necessary to lay lines to aid navigation back to a convenient point. Lines can also be used to search an area: if you know the approximate position of a specific site, natural object, or artifact, you can use a line, in arcs of different lengths, to scan a large area of seabed efficiently.

1 **Find a solid anchor point** to attach the line to. Make a couple of turns around the object for extra security. Do not secure a line to jagged edges that will fray or cut through it.

2 **Secure the line** to the anchor point. If you have a reel and line you use regularly, fit a spring clip to the end of the line so that it can wrap around the anchor point and clip back on to itself.

3 **Unspool the line** as you explore, making sure that your buddy keeps clear to one side. The line should be tied off again wherever you change heading.

Ascent

All good things come to an end, and at some point your dive will be over. Ascending at the end of a dive should be carried out with great care. This is the part of the dive where decompression sickness (*see pp.142–43*) can develop, so your ascent must not be hurried.

WHEN TO ASCEND

Dives often end when you reach a prearranged time to surface, either to meet the boat cover or to account for changing tides. However, you may decide to ascend before this if the site proves uninteresting, or if you or your buddy are cold or tired. You may also be forced to ascend if you have only 725 psi (50 bar) left in your tank, if you are making a decompression dive (*see pp.130–31*) and reach the scheduled minimum volume of air in your tank, or if problems cause you to abort the dive.

SIGNALING TO ASCEND
To ask your buddy if they are ready to ascend, or to signal your intent, raise your thumb.

Before ascent, you and your buddy should signal the "thumbs up" to each other that you are both able and ready to ascend to the surface.

If possible, avoid making a "free ascent," with no visual reference to guide you. If you can follow a reef wall or a permanent mooring line, for example, this will give you a natural sense of your ascent rate, in addition to your computer readings, and can help avoid disorientation. Sometimes, though, it is not possible to ascend up a line or along a natural feature, and it can be helpful to deploy an delayed surface

MONITORING A DIVE COMPUTER
Dive computers (*see pp.66–67*) help the diver judge ascent rate. Some devices give audible warnings if safe limits are exceeded.

BUDDY PAIR ASCENDING A GUIDELINE
A buoyed line moored to a wreck or seabed feature offers one of the safest guides for an ascent, especially in a strong current.

marker buoy (DSMB; *see pp.130–31*)—this will give you a vertical reference and also signals your position at the surface. Ideally, both members of the buddy pair should carry a DSMB. If you are diving on a wreck or a site with a guideline, aim to get back to the line for your ascent—not only will boat cover know where to pick you up at the surface, but the line will assist you in making decompression stops and, in a current, can provide something to hold on to. If you do not manage to get back to the guideline, send up a DSMB.

MAKING YOUR ASCENT

To start your ascent, slowly breathe in or let a little air into your BC—this should cause you to rise up. Rise up next to, and ideally face-to-face with, your buddy so that you can check each other's rate of ascent and condition. As ascent continues, carefully monitor the amount of air in your BC (and drysuit if relevant)—the air will expand as the pressure decreases during your ascent, causing your rate of ascent to increase rapidly. Keep one hand on the BC exhaust hose in order to vent air from

FOLLOWING A SLOPE
Ascending next to a natural feature, such as a reef, offers a reassuring reference point.

it as you ascend. You may also need to dump air from your drysuit. You can fin to assist your ascent, especially to gain momentum to begin with, but this will not be necessary if you are correctly weighted and positively buoyant. Keep breathing normally and do not hold your breath.

A conservative estimate of a safe rate of ascent is 50 ft (15 m) per minute (or slower) up to 16 ft (5 m) in depth. At this point many divers make a 3-minute safety stop. Whether you do or not, the last 16 ft (5 m) to the surface should take another full minute. If making a decompression dive, make all your scheduled safety stops and check with your buddy as you do so.

MAKING A SAFETY STOP
Divers make a precautionary safety stop at 16 ft (5 m) to allow nitrogen to exit the body's tissues at a safe rate.

DEPLOYING AN INFLATABLE BUOY

Since ascent must be conducted in a controlled way, it is helpful to carry your own portable guideline—an inflatable buoy, or safety sausage. These highly visible tubes are launched using air from your second stage, and allow you to ascend up an attached line. Using them is a real skill, so seek training before your first attempt. Remember to reel the line in as you ascend, since slack line can get wrapped around your gear.

1 **Unfurl the buoy**, which should be attached to a reel. Unclip the reel from yourself, taking care not to drop it.

2 **Partially inflate** with a small burst of purged air from your octopus (*see p.61*) to stand the buoy upright.

3 **Hold the reel** in one hand, and with the other, hold your octopus valve under the buoy's inflation point.

4 **Launch the** buoy by purging a long blast of air into it with your octopus. Release the spool, holding tight to the reel, and the buoy will shoot upward.

5 **When the** spool stops spinning, start to ascend, reeling in the line as it slackens. Ascend at a safe pace: do not tug on the line to pull yourself up. During ascent, keep one hand ready to vent air from your BC if required.

6 **Await recovery** at the surface from your boat cover, which will have sighted the buoy at the start of your ascent. The buoy should be left in its inflated state until you have been recovered.

Surface skills and recovery

Once you and your buddy have completed your dive and are at the surface, you need to head back to shore or wait to be picked up by your boat cover. Always remember that the dive isn't over until you and your buddy are back on dry land or on your dive vessel!

EXITING ON TO THE SHORE

If completing a shore dive, you and your buddy should swim to the shore and find a safe, appropriate method of getting back on to dry land. If you are diving from a sandy beach, it is usually best to keep your fins on for as long as possible and walk backward up the beach until you are well out of the way of any swell. Removing your fins too early can lead to problems if you are swept off your feet or back out to sea, since you will have no propulsion.

If diving from a rocky shore, try to find an easy route back on to the rocks. Again, aim to keep your fins on for as long as possible. As you take off your gear, place it clear of any swell that could carry it back into the water.

EXITING ON TO A BEACH
Only remove your fins and walk forward when exiting the shallows of a calm sea, or once you are out of the way of any swell.

SIGNALING TO A BOAT
Your dive plan should ensure that your boat cover knows roughly when you will surface. If you need help quickly, shout and wave a signaling flag.

SIGNALING TO BOAT COVER

If you are diving from a boat, remain close to your buddy with your regulator in place and signal to your boat that you are OK and awaiting recovery. If there is a problem or emergency (*see p.141*) and you or your buddy require immediate recovery, wave to the boat, or draw extra attention to yourselves by blowing a whistle, shouting, or waving

a flag. As the boat approaches, move to it with caution, keeping your distance until you have a signal from the boat that it is in neutral and safe to approach.

DIVER GIVES "OK" SIGNAL
A standard signal for "I'm OK" at the surface is to put your hand on top of your head, forming a closed arc.

BOARDING A BOAT
It is usually easiest to hand up any small items first, then hoist yourself up the ladder while still suited up.

EXITING ON TO A BOAT OR RIB

To get into a rigid inflatable boat (RIB), you should first get a hold on the boat, either by a rope or by a handle on the side. Make sure your BC is inflated, hand your weight belt in first, and then slip out of your BC by unclipping one of its shoulder straps. Keep hold of it until someone from the boat can pull it in, along with your tank. You should then be able to kick with your fins and pull yourself up on to the RIB.

If returning to a hard boat, head for the ladder and hold on to the bottom. Take great care in a swell—the ladder will move in the water. The skipper or dive master will probably tell you whether to de-kit in the water or to enter the boat wearing your gear. You may be able to hand in your weight belt, BC and tank, but often it is easy enough to remove just your fins at the base of the ladder, hand them up to someone in the boat or hold on to them, and simply climb the ladder fully kitted up. If you feel you are not strong enough to do this, let someone on board know (preferably before you dive).

Once on the boat, be aware that others may be waiting in the water to get on board. Stand back to let others on, stow your gear quickly and neatly, and then offer assistance if it is needed.

Post-dive and debriefing

Your dive is over, and there is equipment to be cleaned and stowed, and records to write up while your dive is still fresh in your mind. Problems may have occurred during the dive, and questions may have arisen that need to be addressed. Now is the best time to do this.

PRACTICAL MATTERS

The first post-dive task is to wash your gear. You need to ensure that it is clean and free from mud or sand. If you have been diving in saltwater, then this also needs to be thoroughly rinsed off before the salt starts to corrode metal parts. Large dive boats and liveaboards usually have one facility for washing electronic equipment, such as computers and cameras, and another for your BC, regulator, and other gear. Rinse all items thoroughly to remove excess salt and sand, which can easily work itself into O-rings and regulators. There may also be facilities on board to dry your gear before stowing it away in a bag or box for your next dive. Make sure you follow instructions and do not leave gear lying around—this can lead to loss or damage, which may impede your next dive. If you are diving from a rigid inflatable boat (RIB) or small dive boat, you will need to find a water source back on land to rinse your gear, and then leave it to dry or stow it for transport.

ANALYZING THE DIVE

There will normally be some sort of debriefing with your dive leader after a dive. This gives you an opportunity to ask questions about the dive (for example, to identify fish species that you did not recognize, or discuss problems with gear, navigation, or currents). This is a time when you can learn about the site for a future dive and share with others ways to improve skills.

DIVER CLEANING GEAR
Rinsing, drying, checking, and stowing your gear should become a post-dive routine.

You should always fill in a log book after every dive, recording the depth, time, and conditions (including visibility, water temperature, and currents), as well as key features of the dive—details of a wreck, species encountered, equipment configuration, the type of suit you were wearing and the weights you used, and any problems you experienced. The log will be a valuable reference for future dives.

Finally, ensure that after every dive you are well rehydrated. You will have expended a high level of energy and will have lost fluids. Choose non-fizzy soft drinks or, ideally, water, to ensure that you rehydrate your body properly. Keep your energy levels up, especially if you are diving again soon, by having a good meal or eating a snack with a high energy content. If you are cold after a dive, get warm and dry—have a hot drink and get out of your wetsuit, if possible.

REACHING THE SHORE
Eager as you may be to talk about what you have seen, attend to your gear first.

LOOKING FORWARD

If you are planning to make more than one dive in the same day, consider beforehand what your surface interval will be between dives. Aim to do your deepest dive first, and shallower dives subsequently. Make sure you are in good shape to dive again—if you have become cold or fatigued on the first dive, it is better to spend time recovering than to dive again too soon.

DIVERS DEBRIEFING
Talking to other divers in the post-dive get-together will help broaden your knowledge of diving, and you may pick up some handy tips.

Problem solving

Diving is a safe sport if you follow the rules, avoid taking unnecessary risks, and always dive with a buddy. But that does not mean you will never encounter difficulties during a dive. Learning how to anticipate and solve problems is part of becoming an accomplished diver.

KNOW YOUR LIMITS

One of the most exciting things about diving is that no two dives are ever the same. Even consecutive dives on the same dive site, with the same buddy, will be different due to changes in currents, visibility, and your own attitude. Some days you will feel great, and everything will work out perfectly, and other days you might be tired, have a slight headache, perhaps, and there may be minor problems underwater that you have to deal with. In this unpredictable sport, you must be mentally and physically prepared for every eventuality.

REGULAR SIGNALING
Remain in close communication with your buddy at all times, and signal to each other regularly that you are OK.

To minimize the likelihood of problems and incidents occurring during a dive, do not dive without preparation in conditions with which you are unfamiliar. For example, if you have never dived in a cave, wreck, or under ice, make sure that you undergo the relevant training before attempting to do so. Diving in conditions for which you are unprepared will lead to anxiety, and sometimes to panic—and that is when problems often arise.

The buddy system is important for your safety, but as a diver you must ultimately take responsibility for your own welfare. If you feel uneasy during a dive, tell your buddy, and either change or abort the dive. Never let peer pressure push you beyond your limits and training, even if this means that your buddy's dive may be unsatisfactory. Buddy compatibility is important, since accidents can occur when two egos clash.

LOOKING OUT FOR YOUR BUDDY
Your buddy is your underwater lifeguard, but it is a two-way relationship—you also have to support your buddy when problems arise.

MENTAL AND PHYSICAL FITNESS
Confidence, a positive mental attitude, good health, and good levels of fitness are essential ingredients for a safe, successful dive.

ANTICIPATING EQUIPMENT PROBLEMS

It is rare that equipment fails during a dive, but you should nevertheless take great care when assembling, storing, and servicing your gear. After all, this is what keeps you alive underwater, so it is worth spending time and money on keeping it in perfect working order.

Consider "what if" situations ahead of a dive, and think through how you will deal with the failure of any element of your gear. Do you have a spare if your buddy is not close by? Your regulator should always have an octopus second stage (*see p.61*), to use as a backup in case your main second stage fails.

HANDLING INCIDENTS CALMLY

Many things can go wrong on a dive, but most are minor issues that can be solved without aborting the dive. Even serious problems can be dealt with if you remain focused and positive. Be persistent, and say "no" to failure. When you become aware of a problem, stop, think, breathe, and then act. Whatever happens, do not panic—remember that your buddy is there to help. Make sure you are there for your buddy, too, and be quick to reassure or assist them if they appear disoriented or anxious.

THE INCIDENT PIT

Minor problems can escalate into major ones if you do not stay calm and deal with them methodically. Imagine that you lose your regulator while trying to ease cramp in your foot, and then fumble as you try to retrieve it. If you panic, you may inhale water, choke, and become disorientated. This type of scenario is known as the "incident pit"—stay out of it by thinking clearly and acting methodically.

TROUBLE-SHOOTING

The equipment problems you will encounter most commonly will involve your regulator, mask, and inflator hose.

If your regulator comes away from your mouth and you cannot retrieve it, reach for your octopus second stage (*see p.61*) and breathe from it. If your octopus is not accessible for any reason, signal to your buddy, who should offer you their own octopus to use. Take a few breaths from this to calm down and, with your buddy's help, locate your main regulator and replace it.

Should air start to flow uncontrollably from your regulator, take the regulator out of your mouth and push the purge button, which should clear any grit that may be jamming the valve open. If this does not work, switch to your octopus second stage and signal to your buddy that your air supply is running down more quickly than normal. Monitor your air supply carefully, and ascend immediately with your buddy. Note that in very cold fresh water, freezing up of the regulator will probably be the cause of any free-flow, in which case purging will not help. In this scenario, you may need to breathe from your buddy's octopus while performing a shared-air ascent together (*see p.140*).

RETRIEVING LOST REGULATOR
A lost regulator is usually easy to retrieve, but if your buddy is finding this difficult, go to their aid and, if necessary, offer them your octopus.

CALMING YOUR BUDDY
Reassuring physical contact can help to calm a disoriented buddy and prevent panic from setting in.

SECURING A BUDDY'S LOOSE TANK
Some problems are difficult to solve alone, and you should not hesitate to alert your buddy if you need assistance.

REMOVING YOUR WEIGHT BELT
You must be competent at removing your weight belt underwater, in case you need to adjust your weights or if your belt becomes snagged.

SOLVING MASK AND HOSE ISSUES

If your mask is dislodged or starts to fill with water, simply clear it in the normal way. Lean your head back, press in the upper edge of the mask, and breathe out through your nose. The air bubbles you exhale will clear any excess water in your mask. At depth, this can seem daunting, so be prepared to practice this procedure until you are comfortable with it and can execute it confidently.

On occasion, you may find that the inflator hose for your BC or drysuit fails to work. This is usually caused by grit becoming lodged in the valve. If necessary, BCs can be inflated orally, both underwater or at the surface.

LOCATING A LOST BUDDY

During a dive, you may find you lose sight of your buddy, especially in low visibility conditions. The correct procedure is to look around for a minute or two, and then surface independently. If your buddy is using a surface marker buoy (*see p.74*), it will indicate their position. Having located it, you can then follow the line down to find your buddy again.

If neither buddy nor buoy is visible, signal to any boat cover to recover you; alert the coastguard if your buddy has not appeared after a few minutes. In all probability, your buddy will surface shortly after you do. You can then decide whether or not to resume the dive.

FREE AND BUOYANT ASCENTS

If you run out of air, locate your buddy and follow the procedure below. If you cannot find your buddy, you will need to make a rapid "free ascent" by finning to the surface. This can be aided by ditching your weights, but be ready for a sudden increase in buoyancy. Breathe out slowly during ascent to prevent lung expansion injury. If, however, you have a little air left, you can make a more controlled "buoyant ascent." Let some air into your BC to kick-start your ascent, and tilt your head back to watch for the surface. Again, breathe out during the ascent. Control your ascent rate by venting air from the BC, but not so much that you lose buoyancy. Do not rise any faster than

1 **Signal to your buddy** that you are out of air by making a chopping motion at your neck with the flat of your hand.

2 **Your buddy will offer** you their octopus second stage, or, if they have one, a separate regulator attached to a pony bottle.

3 **When your buddy is** sure that you are OK and your breathing has settled, you should lock arms and ascend to the surface.

OUT-OF-AIR PROCEDURE WITH BUDDY

If you have not kept track of your air consumption and find that you are out of air, locate your buddy and ascend using their octopus or pony bottle. If no alternate air source is available, then buddy breathe. This involves both divers using the same regulator, with each diver taking two breaths before passing it to their buddy. This is a complex procedure, however, that needs to be practiced if it is to succeed in a real-life crisis.

4 **Hold on to one another** as you rise to the surface. Make any decompression stops as you would do during a normal ascent.

your exhaled bubbles. At the surface, signal to boat cover immediately. If there is no boat, swim to the shore. Your buddy will need to be found and, as you have ascended without safety stops, you will need to be monitored for DCS (*see p.142*).

MAKING EMERGENCY LIFTS

If your buddy is unconscious or injured, keep their regulator in their mouth and perform a buoyant lift to get them to the surface. This means holding on to your buddy's harness as you ascend, using their BC to adjust buoyancy for both of you. Alert any boat cover once at the surface. If your buddy is not breathing, artificial ventilation (AV) may be needed until medical help arrives. If there is no cover, or if you are close to shore, you may decide to tow your buddy to safety. Both towing and AV require special training; if your dive training did not cover them, a lifesaving course is recommended.

OVERCOMING PANIC

If you feel panic coming on, alert your buddy, stop moving, and steady your breathing. If your buddy panics, reassure them with hand signals but observe them from a safe distance, since a flailing arm can knock your regulator from your mouth or injure you. When they have calmed down, hold their hand or arm, make sure their regulator stays in place, and remain close until normal breathing resumes.

SIGNALING FOR HELP
Supporting your buddy at the surface, wave one arm in a wide arc and shout to alert boat cover.

Decompression sickness

If a diver ascends too quickly, the rapid pressure change causes bubbles of nitrogen to form in tissues and blood vessels *(see pp.100–01)*, resulting in decompression sickness (DCS). This condition must be watched closely and treated quickly, since in extreme cases it is fatal.

DCS SYMPTOMS

Severe DCS may cause impaired vision and balance, convulsions, and even unconsciousness. Nitrogen bubbles can form under the skin, or become trapped in capillaries and restrict the blood supply to the lungs, slowing breathing, lowering blood pressure, and causing shock. If a bubble enters the brain or spinal cord, it can lead to paralysis or even death. Symptoms usually appear between one and six hours after surfacing, but occasionally do not manifest themselves until 24 hours later. They start with minor aches in joints, rashes, and itching. These will disappear after a few hours in mild DCS, but with severe DCS they will be precursors of serious complications. At the first sign of symptoms, request emergency medical attention (from the coast guard, if you are at sea) and give basic first aid: supply the patient with 100 percent oxygen; keep them warm and still (lying down if possible); provide nonalcoholic fluids (ideally sports or rehydrating drinks); and apply pressure to joints to alleviate pain. The patient should be referred for medical attention and may need to be treated in a recompression chamber.

EMERGENCY OXYGEN
It is vital that the dive boats carry an oxygen administration kit in case of emergencies.

If so, provide those in charge of the chamber with the diver's symptoms and details of the dive. In the chamber, the pressure is raised until it causes gas bubbles in the diver's tissues and blood to shrink. The diver is decompressed under controlled conditions, allowing any residual nitrogen to dissipate safely.

LOOKING FOR SYMPTOMS
Skin rashes and joint pains may be precursors of more serious problems.

DCS AT SEA
Record the symptoms, give first aid, and get to shore if feasible—or wait for the coast guard to arrive.

Hand signals

Good communication with your dive buddy is important, not just in an emergency, but also to monitor each other's progress, share underwater experiences, and agree when to end the dive. The most effective method of communication is to use hand signals.

USING HAND SIGNALS

Verbal communication is not an option underwater (*see p.92*) unless you are wearing a full-face mask, but most messages can be conveyed using a combination of hand signals and drawings on a slate. Using hand signals will allow you to exchange information with your dive buddy and even ask each other questions underwater. They can also be reassuring.

Signal to your buddy regularly during a dive, as this will ensure that you remain in close contact and can easily notify each other of any problems that may develop. Repeat any sign that is not clear to your buddy, and acknowledge every signal that he or she makes with an "OK" to show you understand. If your buddy is slow to respond to a signal, check that they are OK. If you are diving at depth, sluggish responses may indicate the onset of nitrogen narcosis (*see pp.100–01*), which impairs reaction times and mental acuity.

Signals are not just limited to showing each other how much air you have left or highlighting problems. They can also be used to inform your buddy of interesting things you have seen, or to alert them to something you are both looking for, such as a particular marine organism. There are numerous signals for marine animals, and you can develop your own within your buddy pair. A common example is the use of a straight, upright hand (representing a fin) on top of your head, meaning "I've seen a shark." Signals for animals like turtles and rays can be readily improvised.

QUALIFYING THE SIGNAL
To make it clear that a signal refers specifically to yourself—for example, "watch me"—point to yourself first.

COMMON HAND SIGNALS

There are a number of standard hand signals that are essential to all divers, and you can also improvise new signals with friends. Always review signals at the start of a dive, especially if you are diving with a new buddy. Give signals to your buddy slowly and clearly, to ensure that they are understood. Use combinations of signals to convey messages more precisely (for example, something is wrong/I can't clear my ears). Below are some of the most important and common diving signals.

OK/Are you OK?
Form an O with thumb and index finger. Point the other fingers up.

Something is wrong
Tilt the flat of your hand from side to side, palm down.

Up/Let's ascend
Point thumb straight upward with fingers clenched in a fist.

Down/Let's go down
Point thumb downward with fingers clenched in a fist.

1,500 psi (100 bar) left
Form T-shape with hands to indicate 1,500 psi remaining.

750 psi (50 bar) left
Make a fist, with palm outward, to indicate 750 psi remaining.

I am out of air
Make a horizontal chopping motion across base of neck.

Stop
Present flat of hand, palm outward, to halt buddy or other divers.

Slow down
Move palms slowly downward together (pivoting at elbows).

Stay/Move together
Move index fingers together, until touching side by side.

Stay at this depth
Hold both hands horizontally and move one over the other.

Watch/Look
Point at your eyes, then at subject of interest, or self, or another diver.

I am cold
Hug yourself with both arms crossed to indicate chill.

I can't clear my ears
Point at ear to indicate you are having difficulty equalizing.

Feeling breathless
If you are feeling out of breath and need to rest for a moment, you should use the flat of your hand to mimic your chest rising and falling.

Developing your skills

With any skilled pastime, it is important to keep your technique sharp. Many divers only dive once a year (or even less frequently), and it is easy to forget even basic skills during your time off from diving. Honing your skills in sheltered water is a good way to stay proficient.

PERFECTING BUOYANCY

Beginners, occasional divers, and even those who dive regularly will benefit from practicing key skills in a pool or sheltered water. Buoyancy is the most fundamental diving skill of all, so it is a good place to start when reviewing your technique. Refamiliarize yourself with buoyancy control with "fin pivots," using inhalations and exhalations to make the upper part of your body rise and fall. Keep your fins on the bottom and your legs apart. They will act as the "hinge" of your upward and downward movements. The value of this exercise is that it trains you to use breathing to adjust your buoyancy, rather than controlling it only with your BC. Also, try swimming through a hoop with your arms folded—this will test your buoyancy control and give you an idea of the space you can squeeze through with your equipment on.

1 **To practice fin pivots**, breathe in gently and feel your upper body rise, keeping your fin tips on the floor. Stop at a 45° angle to the bottom.

2 **Slowly release the air** in your lungs until your body begins falling. Then, before you hit the bottom, begin inhaling again.

CONFIDENCE-BUILDING EXERCISES

Mask clearing is one of the most basic, yet important, skills in diving. Building coordination and confidence by removing and replacing your mask, and even performing tasks without it, will improve your diving. It will also prepare you for minor upsets like your mask being accidentally dislodged. Begin building up your confidence slowly by removing and replacing your mask in shallow water, before repeating the exercise in deeper water. Repeat while swimming, and then try it with your regulator removed. You can also try swapping masks with your buddy. This is simply a variation on the solo exercises but also helps you to develop useful cooperative skills.

Giving yourself other handicaps can also help improve your confidence underwater. Try swimming with just one fin, for example. Losing a fin can happen on an open-water dive, but if you are already prepared, you will be able to cope with the situation calmly and effectively. Practice all of these exercises in a pool or sheltered water.

REDUCED VISIBILITY
Most divers feel uncomfortable swimming with their mask removed, but with practice, you will become confident performing tasks without it.

1 **To exchange masks**, kneel on the bottom facing each other. Give each other the "OK" signal, then remove your masks, while continuing to breathe normally.

2 **Swap masks** with your partner. It is much easier to keep your eyes open throughout the exchange, but perfecting this by touch, with your eyes closed, is a very useful exercise.

3 **Don each other's masks**, and blow them clear (*see p.139*). Ignore any urge you feel to hurry when putting the mask on. In particular, make sure the strap is not twisted.

4 **Adjust the mask** for comfort, then open your eyes and signal OK to indicate a successful swap. The masks can then be swapped back again for further practice.

VARYING THE ROUTINE

Skills exercises don't have to be simple drills. Introducing an element of fun or setting an objective can motivate you to train for longer. Games for the pool or sheltered water include "hide the mask," which forces you to navigate and search without your mask; or you can practice breath control by performing simple exercises without breathing apparatus, such as retrieving objects from the bottom, or swimming through a series of hoops, extending the course as your stamina improves.

1 **"Hide the mask"** begins with an "OK" signal from the seeker, once the diver who is going to hide the mask has removed it.

2 **The seeker closes his eyes** while the hider finds a corner of the pool to put the mask in, and returns to signal the seeker to start searching.

3 **The seeker opens his eyes** and starts to methodically search the pool bottom until the mask is found and replaced.

REMOVING AND REPLACING GEAR

Familiarity with your basic gear is a vital skill. One exercise that helps to foster confidence in this area is to remove your scuba unit—that is, your BC and breathing apparatus, all fully connected—and put it back on underwater, keeping your regulator in your mouth throughout. This exercise should be carried out at the bottom of a pool or in sheltered water, and if you are a beginner or haven't tried this before, it may be useful to have an instructor present who can guide you through the process. Start by undoing all fastening clips on the front of your BC and pull your left arm out of it first (even if you are left-handed), using your right arm to pull it around to your right. The BC jacket should now be in front of you, between you and the tank. Keep the hose for your regulator second stage between your arms (otherwise it may get caught under your shoulder strap when you put the unit back on). Then put the scuba unit back on by reversing the operation.

For an extra challenge, once you have removed your BC, try removing your regulator from your mouth and swimming away to a distance of about 30 ft (10 m), before returning and putting your gear back on again.

Alternatively, try entering the water without your gear, then putting it on while treading water, keeping your head above the surface. This is a much more difficult exercise, and should only be carried out under the guidance of an instructor.

1 **To take off** your scuba unit underwater, undo all clips and take your left arm out of the BC, gripping the jacket with your right hand. Breathe normally through your regulator.

2 **Pull the whole unit around** in front of you. Keep a firm grip on your BC to avoid losing hold of the scuba unit.

3 **To put gear back on**, reverse the operation, making sure all clips and fastenings are reconnected.

Dive fitness

As a diver, you are reliant on your body as well as your gear for a safe and enjoyable dive—so make sure both are well maintained. Diving regularly helps keep you in shape, but it is important to maintain your fitness levels if you go for long periods without diving.

HOW FITNESS AFFECTS DIVING

Staying in shape reduces some of the risks associated with diving, and makes it more enjoyable. Having healthy lungs and an efficient circulatory system means you will use less air, so dives last longer. A good level of cardiovascular fitness will also help to prevent panic attacks, which can be triggered by a buildup of carbon dioxide in the blood. A fit body will expel this by-product of respiration more efficiently, reducing that risk. General strength is required for lifting equipment and other incidental tasks.

Maintaining a healthy weight for your size is also advisable. When you dive, nitrogen builds up in the tissues of your body with each intake of breath, and fat retains nitrogen longer than all other tissue types. Therefore, if you are overweight and breathe rapidly because you are out of shape, you expose yourself to an increased risk of DCS (*see pp.142–43*).

Weight loss and gain can also alter your buoyancy. Fat is inherently buoyant, so if you have put on or lost a lot of weight since your last dive, you must adjust how much lead weight you carry as ballast (*see pp.118–21*).

TALKING THINGS THROUGH
If you have any injuries or physically weak areas, let those diving with you know before you dive. Don't be afraid to ask for assistance if this will help prevent injury.

DEVELOPING A FITNESS REGIMEN

There is no quick fix for getting and staying fit. Aim to exercise regularly, tailoring your routine to strengthen the muscles you use most during diving: those of your back, legs, shoulders, and arms. If you do not already exercise regularly, seek medical advice before embarking on a fitness routine. Start gently with easy exercises, and build up slowly. A routine based on the exercises overleaf could start, for example, with ten repetitions of each, building up to 50, three times a week. You should also aim to undertake regular cardiovascular exercise (such as walking or cycling), for 30 minutes two or three times a week to improve your general fitness and respiratory health.

Whatever exercise routine you decide to adopt, try to incorporate it into your normal daily routine wherever possible. A walk to work every day does more good than a monthly gym session. Never exercise after diving. This is a time when it is important not to do any strenuous activities—there is still residual nitrogen in your body, and this may form bubbles if you exercise too hard. Do some gentle stretches and take it easy for the rest of the day. Avoid hot baths or showers after diving, since these can also raise your vulnerability to DCS.

GOING AGAINST THE FLOW
Finning into strong currents places demands on your leg muscles and lungs.

PROTECT YOUR BACK

Be very careful when carrying and loading heavy equipment. A poor lifting technique can result in pulled muscles, and possibly long-term back problems. A back injury can spoil a trip, but diving with a weak back can also be a safety risk: surface maneuvers, climbing onto boats, and exiting surf, for example, are all compromised by weakness in this area. If your lower back is vulnerable, consider transferring your weights from a belt to a harness (*see p.70*).

Loading dive equipment

WARMING UP

Always warm up and stretch—especially your thighs, calves, and upper and lower back—before exercising, and spend a few minutes gently stretching and cooling down afterward. Many sporting injuries are caused by failure to warm up properly, so treat it as an important part of your fitness routine.

ARM AND LEG EXERCISES

Basic squats build up strength in your legs. Standing upright with your feet shoulder-width apart and your knees soft, place your hands on your hips and shift the weight back into your heels. Keeping your head and chest pointing forward and your back straight, bend at the hip and knees, as if you were sitting down onto a chair. Keep your heels on the floor; you should be able to move your toes, since all your weight is on your heels. Straighten and repeat.

Lunges are a more advanced exercise for legs. Stand upright with one foot a good stride in front of the other, with your hips parallel and your toes forward. Your back heel should be off the ground and your front foot flat on the floor. Both knees should be bent and your back straight. Drop your body weight toward the floor until your front knee is bent at a 90° angle. You may have to tilt your pelvis forward to get this position right. Rise up and repeat.

Biceps curls strengthen the biceps—muscles in the upper arms that are used whenever you lift anything. Using hand weights or a weighted bar, stand upright,

BASIC PUSH-UP
The push-up is a great all-around exercise for the upper body and arms. Beginners can start with their knees resting on the ground.

HEALTHY LUNGS
Cardiovascular training builds up lung capacity, making snorkeling feel even more effortless.

LUNGE EXERCISE
Practicing lunges in front of a mirror will help you to get the position of your front leg right: knee directly over foot, bent to a right angle.

ABDOMINAL CRUNCH
Always keep your lower back flat on the floor and never arched. For a harder workout, aim your right elbow to your left knee, and vice versa.

with your back straight and knees slightly soft, and hold the weights in your hands with your palms facing upward. Without rocking or moving your shoulders, bend the arms toward the chest and then lower, taking care not to lock the elbows.

EXERCISING YOUR TORSO

Push-ups strengthen your shoulders, upper body, and core muscles. Lying on your front, bend your knees, curl your toes under your feet, and place your hands shoulder-width apart below your shoulders. Press your body weight up and down from this position. Keep your back straight at all times and your elbows pointing out to the sides.

Crunches strengthen abdominal muscles, which will help protect your lower back. Lying on your back, bend your knees slightly and place the soles of your feet on the floor. Place your fingertips behind your ears and lift your upper torso and shoulders off the floor. Hold this for three to five seconds while looking forward, then lower and repeat.

HYDRATION AND NUTRITION

During a dive, you can burn as much energy as you would playing an aerobic sport, such as squash or badminton. The precise amount will vary according to your fitness, the temperature of the water, and any prevailing currents. But while it is tempting to stoke up on a big fried breakfast in preparation for a day's diving, better food choices before a dive are complex carbohydrates, which will keep you feeling fuller for longer and provide a steady supply of blood sugar to maximize energy levels while you are diving. Some foods may make you feel uncomfortable on a dive—commonly, greasy or hot and spicy dishes—so these are best avoided.

After a dive, you need to replace the energy you have burned with foods that release energy into your blood relatively quickly, such as refined carbohydrates. Have something to eat within an hour of surfacing, if possible. Energy-replacement drinks can be a quick fix, but avoid those with a high caffeine content.

GOOD DIVING FOODS
Whole-wheat bread, legumes, and unrefined cereals and pasta release energy slowly during a dive. Dried fruit and nuts make good snacks.

Above all, the most important thing is to remain well hydrated before and after a dive. Your physical performance will be reduced if you become even mildly dehydrated, and this is a big issue if you are diving in warm waters where the outside temperature is also high. Aim to drink at least 2 quarts (liters) of water during the course of the day (more if it is hot). This will help to reduce muscle cramps and fatigue, and also helps decrease the risk of DCS.

AVOIDING PROBLEMS

It is important to get plenty of sleep before and after diving, to give your body a chance to recharge. Diving is physically demanding, and if you are run down, you are more susceptible to catching colds. If you do have a cough or a cold, you should not dive, because some symptoms of respiratory illness can prevent you from being able to equalize and may increase your vulnerability to DCS (*see pp.100–01*). Avoid taking decongestants, since these can impair your performance underwater.

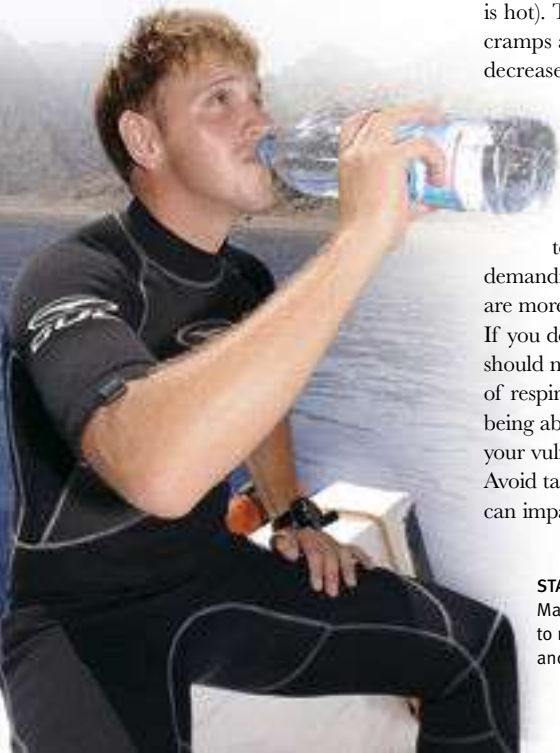

STAYING HYDRATED
Make sure you drink plenty of still water to maintain good hydration. Tea, coffee, and carbonated drinks are no substitute.

FOODS TO EAT WHEN DIVING

The following list gives examples of the kinds of foods it is good to include in meals and snacks before and after a dive:

BEFORE A DIVE Multigrain, whole-wheat, or rye bread; whole-wheat pasta; sweet potatoes; brown rice; oatmeal; bran-based cereals; legumes and nuts; beans; leafy vegetables; and plain yogurt.

AFTER A DIVE Pasta and rice (ideally in a salad or low-fat sauce); potatoes; fresh fruit; white bread or bagels.

So can alcoholic drinks, which are best avoided the night before a dive because they can also contribute to dehydration. Divers sometimes experience minor ear infections, especially after days of continuous diving or if diving in slightly polluted waters. There are various ways of avoiding ear problems, including rinsing the ears with fresh water after every dive, or using an alcohol-based fluid designed for swimmer's ear.

Lastly, you should have a regular dive medical to check that you are in good enough shape to dive. If, between medicals, there have been any changes in your health since you last dived, then it is always best to have a checkup.

AFTER A DIVE
It's tempting to head for the bar after a day's diving, but have something to eat, too. Even standing still in cold water burns calories.

Going diving

Introduction

Becoming a qualified diver is a tremendously exciting experience, but the range of choices facing you post-qualification can be daunting. The joy of diving is that within one sport lies a host of different specialties, all requiring different skills, knowledge, and planning techniques.

RECREATIONAL DIVING

Upon qualification, the world of diving opens up to you. Whether you become a casual resort diver or an advanced technical diver with a liking for the more extreme elements of the sport, only time and experience will tell.

The broad spectrum of recreational diving encompasses many different environments and experiences, including reefs, caves and caverns, wreck diving, ice diving, freshwater diving, and night diving. There are also various specialty activities in which you can participate, such as filming and photography, wildlife and conservation diving, underwater archaeology, and free diving (diving without an air supply). The majority of divers find that they derive most enjoyment from sampling a range of different environments and activities, rather than limiting themselves to one specific aspect of diving.

BEACH BRIEFING
Becoming confident enough to conduct your own dives without a guide or instructor gives you greater freedom. Always plan and brief dives thoroughly before entering the water.

Many new divers aim to acquire further specializations, pins, and qualifications from the moment they pass their initial training course. This urge, while understandable, should be tempered by the need to gain general diving experience. There is no substitute for time spent in the water—what was once intimidating becomes familiar, helpful routines start to emerge, skills and safety consciousness improve, and knowledge increases. You will also feel more confident and comfortable in the water as you learn to tailor your equipment to your own needs.

Only when you have served your "apprenticeship" by gaining experience of different environments and diving skills should you make any decisions about future specializations. Without putting in the hours underwater, you will find further training and courses more difficult, no matter how well structured and delivered they are. In addition, the more experience you have, the more aware you will be of your own likes, dislikes, and aptitudes, allowing you to make more informed choices for the future direction of your diving.

Fortunately, gaining general diving experience is relatively easy, with dive groups and clubs springing up in abundance to meet the sport's growing appeal. Furthermore, dive travel is so well established that almost any region of the world is now accessible, and opportunities to dive exist in virtually every coastal resort around the world. A whole range of diving encounters are available—from a casual beach dive through to full-blown expeditions—all easy to arrange via the Internet or through dedicated dive travel agencies.

PLOTTING YOUR FUTURE COURSE
The more experience you gain, the better informed you will be when deciding whether or not to pursue a particular aspect of diving.

PROFESSIONAL DIVING

The enjoyment of diving leads some people to consider a full-time career in which diving plays an essential part. Diving professionals range from recreational instructors and dive masters, whose main role is to introduce others to diving, through to military experts, commercial specialists, scientific divers, fishermen, and underwater cameramen and photographers.

Further skills are always needed to succeed in your chosen career—an underwater filmmaker, for example, must be an accomplished camera operator as well as a superb diver, and an instructor needs good communication and teaching skills, in addition to first-class abilities in the water.

Professional diving can offer excellent rewards, both personal and financial, but they tend to be well earned. Professionals must dive in all conditions and at all times, and the work is often arduous and sometimes dangerous.

PREPARING FOR AN ICE DIVE
Diving under ice is thrilling, but it is not for beginners. The hazards of enclosed surroundings and extreme cold mean that special training is required.

Identifying your aims

Time spent researching a dive is time well spent. Not only does it help you to unearth new attractions and clarify the aims of the dive, but it is also likely to heighten anticipation, enhance safety, and maximize the pleasure and interest you derive from the trip.

ASSESSING AIMS

The nature of dive research will vary from trip to trip, but certain core issues should always be taken into consideration.

Identifying the divers is obviously one of the first items to be addressed. Diving is not, as a rule, an individual sport, so of necessity others will be involved in the dive itself. The experience and qualifications of divers within a group will vary, and what may be a comfortable dive for one person may be extreme for another. Time spent discussing qualifications and experience levels may well prove invaluable in establishing the ideal dive location and activity for the group as a whole.

BEGINNING RESEARCH
Use magazines, books, and maps to highlight basic issues that you can then investigate in depth on the Internet.

Along with making sure that you have the right mix of divers, you will need to ensure that their aims are not likely to conflict. Imagine the scenario, for example, of an avid photographer preparing to take an

TROPICAL CONSIDERATIONS
When planning tropical dives, consider the effect of local weather patterns. Runoff from tropical islands in the wet season, for example, may impair visibility.

DIVING WEBSITE
The Internet is an invaluable research resource, allowing you to access the opinions and advice of authoritative dive organizations.

award-winning image of a sunken ship, when an impatient wreck enthusiast swims in front of the camera lens and ruins the shot. Establishing the main aim of the dive at the outset, or even allocating buddy teams with the same interests, should prevent such problems.

When diving with a local operator, try to check out their credentials. In many parts of the world there are associations set up by groups of respectable operators to govern the local dive industry. Failing that, many tourist boards and even government agencies require dive operators to register—it is always worth getting an independent view of your potential host from these organizations.

LOCATION AND TIME

Conditions can vary greatly along even the most bland section of coastline or within the most innocuous-looking body of water. Choosing an exact site in advance has the benefit of allowing very specific research into local issues. Key questions to be answered include: what are the hazards involved? Where are the established entry and exit points? What are the divers likely to encounter during the dive? What is the tidal range? What are the prevailing currents?

The timing of the trip is also important. Tides, for example, may affect not only visibility, but also the marine life present, as well as entry and exit considerations. Marine life will also be influenced by seasonal variations, as will water temperatures. Furthermore, a different range of marine animals may well be present at night than during the day, while boat traffic and use of the area for other watersports will vary throughout the day and the week. Diving at the wrong time can make the event a disappointing experience.

Safety is the most important area of your research. You will need to assess the conditions you are likely to encounter—waves, currents, depth, temperature—and the facilities available to you should an accident occur. You should not show up at a dive site unless you know the exact location of the nearest hospital, recompression chamber, and casualty evacuation facilities. Knowing how to contact these agencies in an emergency is also crucial—researching phone numbers in the comfort of your own home is infinitely better than trying to do it on a rocking boat while caring for an injured diver.

UNDER THE ICE
The possibilities for thrilling diving trips around the world are almost endless, limited only by time, budget, and imagination.

Planning a trip

Armed with the fruits of your research, you will have a clear idea of where you want to dive, when you want to go, who you will be diving with, and what you will see when you get there. The final task before your departure is to plan the trip in as much detail as possible.

REALISTIC TIMETABLING

Most trips revolve around a key moment in time—the instant when the divers enter the water. This is often determined by a certain phase in the tide when the diving is most favorable, or by the desire to dive at a specific time of the day or night, such as when the Sun is high for photography, or in the gathering dusk when the reef is at its most active. A realistic, workable timetable is thus vital to achieve such time-specific goals.

Devising a timetable, such as the chart below, by working backward from the moment you aim to enter the water is the best way to plan your diving day. This method is used by many divers to ensure that their diving is as stress-free and as enjoyable as possible.

PACKING YOUR EQUIPMENT
Checklists and routines will help to ensure that all your gear makes it to the dive site.

PRESSURE TEST

Before you start packing your equipment for a dive, carry out a pressure test. This involves rigging the gear to a tank, and ensuring that the seals and regulators are all functioning as they ought to be. This is also a useful means of ensuring that your tank is full prior to departing for the dive site. There is no more sickening sensation than journeying to a dive site, looking out over perfect conditions, assembling your dive gear, and turning on the air to be greeted by a cacophony of hissing regulators and seals—or, indeed, the tank itself being empty.

REVERSE TIMETABLING

A diver knows they must enter the water at 3:00 pm to dive during slack water between the tides. They work backward from this to calculate a realistic departure hour, building in a contingency to ensure arrival at the dive site in good time.

Activity	Time Taken	Actual time
Entering the water		3:00 pm
Suiting up and buddy checks	25 min	2:35 pm
Site inspection and briefing	40 min	1:55 pm
Arrival and parking	10 min	1:45 pm
Driving to site	1 hour	12:45 pm
Loading car	30 min	12:15 pm
Checking and packing dive gear	30 min	11:45 am
Contingency time	30 min	11:15 am

CAVE DIVING
For more complicated trips, such as cave dives, special equipment may be necessary, and this must be factored into your planning.

PACKING WITH CARE

The greatest cause of the postponement or cancellation of dives is forgetting pieces of gear. A fool-proof method of guaranteeing that you pack all of your equipment is to load your gear into your dive bag at home in the order that you will be putting it on at the dive site. Visualizing your arrival at the site and suiting up will lead you to load your dive bag in the following sequence: suit, boots, knife, weight belt, tank, BC, regulator, computer, hood, mask, snorkel, fins, gloves, and finally any ancillary items (safety sausage or DSMB, flashlight, slate, and so on).

An alternative packing technique is to run through a checklist from head to toe, and again pack your gear in that order. The list then runs: hood, mask, snorkel, suit, regulator, BC, tank, computer, gloves, weight belt, knife, boots, fins.

There are innumerable ways of trying to ensure that no item of equipment is forgotten, but the most effective are those that are convenient, quickly become a habit, and take little time.

DIVERS ENTERING THE WATER
A realistic timetable will ensure that suiting up and briefing on site is not rushed, and that you do not miss your window of opportunity with the tide.

Site-specific planning

Pre-dive research is important, but there is no substitute for a thorough assessment of the site on the day of the dive. This may highlight issues not raised by your research, and allow you to make adjustments to accommodate changes in conditions before the dive starts.

FINALIZING YOUR PLAN

Once you know when and where you are going to dive, you must devise a practical plan for each of the dives you mean to conduct on the day. Whether you are planning for just yourself and a friend, or for a larger group, always take local conditions into account before settling on a final dive plan.

You must address a number of key points in your dive plan. First, it should broadly define the aim of the dive, the dive's entrance and exit points, and its general route, as well as describing an alternative exit in case the main choice is inaccessible. It should also define the dive's maximum depth and time, as well as decompression restrictions, and should specify the minimum reserve of air that divers must surface with at the conclusion of the dive.

Depending on the aim of the dive and the experience of the group, special signals may be required. It is imperative to select diver recall systems—the signals or sounds that will be used to indicate that the dive should be aborted. Your plans for the dive should also identify any potential hazards that may be encountered. If an incident occurs, what procedures should be followed? Define what action to take if divers become separated—in murky waters, this is one of the most common reasons for dives being aborted. Make sure that the divers you intend to pair together as buddies share similar interests or levels of experience, so that there are no conflicts of interest during the dive.

Assess local conditions and examine weather reports; worsening weather conditions might dictate a move to a more sheltered site nearby, or rule out the use of a boat. Finally, ensure that the designated shore marshal or dive supervisor understands their duties, including recording tank pressures and the times that divers enter the water, and responding to incidents. Provide them with contact details of local agencies to be informed in the event of an emergency.

UNLOADING THE GEAR
Even if you feel sure that every item of gear was packed before the trip, double-check on site that you have everything you need.

INSPECTING THE DIVE SITE
Resist the temptation to enter the water as quickly as possible. Assess the conditions and inspect the site, noting the dive's entry and exit points.

OUTLINING THE PLAN
If it helps, you can always draw simple sand maps to clarify the route to the other divers.

THE DIVE BRIEFING

The final task is to brief the entire dive crew. The briefing, which should ideally take place overlooking the dive site, must cover all aspects of the dive plan, and include not only divers, but also shore marshals, deck hands, and anyone else associated with the dive. At the end of the briefing, ensure that all the divers clearly understand the constraints of the dive plan, and are fully prepared for the dive. After any final questions from the group, the dive can begin.

The simple maxim "plan the dive, and dive the plan" is an old one, but its worth has been proven over time. Following it means that all of the divers will get the most they can from the dive, and, importantly, will enjoy it in safety.

ENTERING THE WATER
By the time you enter the water, everyone involved with the dive should be fully aware of the dive parameters.

Wreck diving

Exploring wrecks is—for many people—the main reason to dive. This activity is particularly popular in countries without coral reefs to dive on. However, it also has a general appeal, because wrecks give a historical context for a dive and demand mastery of special skills.

DIVING ON WRECKS

Sunken ships have an obvious attraction as sites of archaeological and historical interest. A wreck is a time capsule from a given period, and the waters around most countries contain thousands of them.

Although most known wrecks are metal-hulled vessels, unusual conditions will sometimes preserve wooden-hulled ships dating back hundreds of years. These provide an excellent opportunity to examine past shipbuilding techniques, and sometimes hold interesting and unusual artifacts. To explore significant wrecks, you should be properly trained and, preferably, work within an archaeological organization.

Information on recent wrecks, such as warships sunk during the two world wars, is usually readily accessible. It is often possible to research the specifications and history of the vessel, along with details of its crew and mission. Historical records may give details of the events surrounding the sinking of the ship.

DESCENDING A GUIDELINE
Popular wrecks often have a buoy moored to their structure to make location and descent easier.

Wrecks also make for good diving because they act as artificial reefs, attracting all sorts of marine life. Over time, the structure becomes encrusted with marine growth, especially filter-feeders, if it lies in a tidal stream. Fish exploit the safety and shelter that wrecks offer: smaller fish tend to shoal around wrecks; larger fish, such as moray or conger eels, can hide safely in nooks and crannies; and predators, such as sharks, may be found in the wreck's surroundings, because they feed off the smaller fish.

PREPARATION AND SKILLS

Wreck diving can be a fascinating activity, but it is not without risks. Under no circumstances go inside a wreck if you lack appropriate training and equipment; the danger of entrapment is very real, and specialist knowledge and skills are required to enter wrecks safely. Exploring the outside of wrecks is safer, but always treat such sites with respect.

TANK IN TRUK LAGOON, SOUTH PACIFIC
Wreck diving is not just about ships; crashed aircraft and military vehicles carried by sunken vessels provide equally enthralling alternatives.

SUNKEN LEVIATHAN
The sight of a major wreck, such as the *Giannis D* in the Red Sea, rarely fails to send shivers down the spine.

DIVING IN CONFINED SPACES
Wreck penetration often involves exploring in dark, confined spaces. It is not recommended for those with claustrophobic tendencies.

FINNING TECHNIQUE

Wrecks often contain a lot of static silt and mud, which is easily disturbed by changes in water currents caused by divers finning overhead. Good visibility can suddenly and dramatically change to almost zero, so using a frog kick (*see p.123*), which generates relatively little turbulence, is recommended to minimize disruption of the silt. The best time for wreck diving is at slack water, but a small current can help sweep away any disturbed silt.

USING GUIDELINES

When you arrive at the dive site, ensure that your boat cover is moored to the wreck's marker buoy. If it doesn't have one, an experienced diver needs to swim down and attach a guideline directly to the wreck, to act as a guide for other divers to follow. If you intend to return to the surface via the same guideline, it can be useful to use a distance line (*see p.127*) to help you find your way back to the bottom of the guideline, especially in poor visibility. If you plan to ascend from another part of the wreck, deploy an inflatable marker buoy (*see pp.130–31*).

PENETRATING WRECKS

When wreck diving, you should carry backups of certain pieces of gear, such as your reel and your flashlight, as well as an independent air source, such as a pony bottle (*see p.68*), in case of emergencies. You must also ensure that you follow the "Rule of Thirds" (*see p.117*) with your air consumption—a third of your air is used to get to the site and explore, a third

LEVEL POSITIONING
The frog kick is a gentle swimming stoke that mainly causes turbulence in the water behind and not below you, and so does not kick up silt.

to return to the point of entry, and a third is left in reserve as a contingency. Allow extra time at the end of the dive to find the guideline, and do not take any unnecessary risks with your dive time, depth, or penetration.

EXPLORATION TIPS

It is a good idea to investigate the area around a wreck, since artifacts may have spread over time and pieces of wreckage may have been knocked off that might be of interest, but take great care. Always be aware that the structure (both inside and outside a wreck) may be much degraded and potentially dangerous. Wear protective gloves, even in warm waters, as corroded metal can be sharp. Wrecks are often heavily fished, so carry a knife in case you become entangled in fishing line. You also need a powerful flashlight (and a backup) to illuminate your way in poor visibility, and to signal to your buddy. In some very enclosed wrecks, it is advisable to wear a helmet and head lamp.

THE GOLDEN RULE

Never remove items from a wreck. As with any artifact, marine finds are more useful to archaeologists when examined *in situ*, and their context provides more information about the wreck than if they are handed in without details of where they were located. Worse still is if they are lost forever in a diver's personal artifact collection.

US troop ship artifacts

RESPONSIBILITIES WHEN DIVING

Wrecks are historical sites, and just as you would not remove artifacts from a historical site on land, such as a castle or monument, you should also respect a wreck and not damage it in any way. Be considerate of the enjoyment of others who may dive the wreck after you. Dive responsibly and with care, and do not remove anything without being given permission to do so. When you find an artifact, such as a ship's bell, after a long,

DIVER USING WRECK KIT
Carrying extra air and nitrox is a failsafe, and can help to extend dive times and shorten decompression stops. Good light sources and dive plan information are also essential.

hard search, you may feel that you deserve to be able to take it away as a trophy, but all such discoveries should remain where they were found, or be raised by trained archaeologists for preservation in a museum.

WOODEN-HULLED VESSELS

Most of the wrecks that you are likely to explore will be metal-hulled, but wooden ships are also preserved under some conditions. Avoid touching the wooden structure, as it is likely to be extremely fragile. Historic wooden vessels may have been lying protected under sand or silt for many hundreds of years. If you want to expose parts of the wreckage for any reason, you should do so by wafting your hand over it to remove the silt, creating a current to take the silt away.

LIVE AMMUNITION

Wrecks from the two world wars can provide fascinating dive experiences for military history buffs, but remember at all times that they may also contain live ammunition that could potentially be very dangerous. Never touch anything that resembles a shell or mine, and research what you are likely to find on the wreck before diving.

ARTIFICIAL REEFS

At some popular dive sites, wrecks have been sunk deliberately to provide a safe, interesting dive at a reasonable depth for recreational divers. Such wrecks quickly become colonized by marine life and provide a fantastic opportunity for diving and marine research. They are generally a good place to start wreck diving, since the structure will be intact and any hazards will have been removed. At some wreck sites there are special diving trails with underwater display boards to

WRECKED BY DESIGN
The former US Coastguard cutter *Keith Tibbetts* was sunk deliberately at Cayman Brac to provide an artificial reef for divers to explore.

RESPECTING HISTORY
Many wrecks sunk during conflict are official war graves and may not be dived.

explain the layout. There may also be laminated booklets that you can take with you on the dive, which enable you to read about the most interesting features of the wreck while on site.

VIRGIN WRECKS

Every wreck diver's dream is to find a previously undiscovered wreck—and there are certainly still plenty to be found. Before you embark on such a search, you should familiarize yourself with the law regarding wrecks, and any wreck you find should be reported to the authorities. Generally speaking, you can dive any wreck around the world apart from those that are designated as historic wrecks and war graves, which are restricted and require special permission to dive.

Locating a virgin wreck can be a difficult, lengthy task, and is likely to be expensive. But a modern hardboat with an experienced skipper and high-tech equipment such as an echo sounder, 3-D imaging of the seabed, and a GPS system, combined with charts and some good background research on your part, could help you hit the jackpot. Many undiscovered wrecks are found in very deep water, so make sure you are fully qualified and appropriately trained before undertaking deep wreck dives.

FINDING OUT MORE

If your interest in a wreck goes beyond merely diving it and you wish to enhance your dive by finding out more about the wreck's history, there are a number of available sources of information. Good starting points include books on wrecks in the area that you are diving, and using the Internet to do some background research. There are numerous websites and forums where you can discover more about wrecks, and also make contact with other divers interested in finding new wrecks. Furthermore, local museums may have displays of artifacts recovered from wrecks in the area, and local divers and boat skippers may be able to reveal information about the history of a wreck.

EXPLORING VIRGIN WRECKS

If you are searching for a particular undiscovered wreck, it is important to start your research in the archive. Studying the ship's specifications and plans will help you identify any remains that you find, and determine whether they belong to the subject of your search. It is a good idea to take video footage of anything you find on a dive, as this is a nonintrusive method of recording the remains, and the footage will be critical in identifying the wreck and for showing to experts.

Navigational chart

Nature and conservation diving

Recreational divers represent a huge pool of manpower that can potentially contribute to conservation work. Many divers take up the sport due to an inherent fascination with nature, and more and more are now applying their skills to organized conservation projects.

DIVING RESPONSIBLY

Most divers have a certain level of knowledge and enthusiasm for marine life and the ocean environment, and realize that along with the thrill of exploration comes responsibility. Divers are in a unique position to observe the deterioration of busy coastal margins around the world and, with the development of the modern ecotourism market, are now also able to actively contribute to conservation projects in the same regions.

At the most basic level, conservation work simply means responsible diving by anyone taking up the sport in the first place. Always adhere to a "no-touch" rule when diving delicate reef systems; you will find this vigorously enforced by most dive operations around the world. Good buoyancy control, streamlined equipment, and a respect for the varied inhabitants of the sea are now widely taught as the only way to dive, and as a result even heavily dived areas experience less habitat disturbance than they used to.

GOOD PREPARATION
A group of divers at a conservation site discuss the work they will be conducting.

USING IDENTIFICATION SLATES
To familiarize themselves with marine life during a dive, divers use species identification slates—waterproof cards with pictures of the species likely to be found at the dive site.

MAKING A DIFFERENCE

Voluntary work, though often conducted at the most basic level, can nonetheless be extremely significant at a local and even international level. The efforts of one voluntary organization alone—Coral Cay Conservation—contributed greatly to the establishment of the Belize Barrier Reef (*see pp.316–17*) as a World Heritage Site, and have seen eight Marine Protected Areas (MPAs) established around the world in the last 20 years.

Formally observing and recording marine life is not just the preserve of the scientific expedition. Significant work has been done by enthusiastic amateurs. Organized conservation and monitoring projects range from casual one-day beach cleanups run by local diving clubs, to multi-national operations setting up nature

reserves throughout the world. Diving conservation is now big business, with hundreds of organizations around the world claiming (with varying levels of accuracy) to run projects that contribute to scientific knowledge or the establishment of marine reserves and protected areas. Most organizations put volunteers through an intense training period before allowing them to take part in even relatively simple sampling or monitoring programs.

REEF EXPOSED
A healthy reef at low tide. Rising global temperatures are harming coral, and sights like this may become less common in future years.

FIRST PRINCIPLES

There are two principal considerations that must be borne in mind by anyone working as a diver on a marine conservation project. The first is safety. You need good diving skills, and enough experience to be able to keep a watchful eye on the ever-present factors of depth, time, and potential hazards while simultaneously making meticulous observations. Enthusiasm for the work being undertaken must not distract you

PRESERVING HABITATS
Carrying out conservation work not only allows you to give something back to the marine environment, but may give you the opportunity to visit exotic destinations.

from the normal rules of safe diving. The second factor is consideration for marine life. Unless you are working under the direction of a well-recognized conservation group, you must only gather data without touching or harming the marine environment. Both marine algae and animals should be recorded only *in situ*.

WORKING TOWARD OBJECTIVES

Good science starts with a clear means to identify the subject of the research. In the case of most amateur enthusiasts, this means reading up in accurate reference works and guides about the dive environment in question. A further, essential part of any observation program is the need to record data effectively, requiring not only a means of noting down results, but a framework of headings and categories that remain the same for the entire observation program. Only by combining these two important factors—accurate identification and consistent record-keeping—can you accomplish truly effective research and monitoring. Repeated observation of species populations within a given area, for example, can provide invaluable data for analysis by research groups. Your recordings may help local initiatives, or possibly go forward to form part of a much larger-scale project, in which

FINDING A PROJECT

Conservation, to paraphrase an old adage, begins at home. There are marine ecology programs in most countries, and interested divers can usually find groups in their area that contribute to protecting the local marine environment. Search the Internet and specialty diving magazines to find suitable organizations local to you.

CORAL MONITORING PROGRAM
Advanced or potentially invasive techniques should only be used under the direction of a recognized conservation group.

DIVING AT A MARINE RESERVE
Even a little research into how and why an area is protected will add food for thought to your enjoyment of the dive.

complex statistical techniques may be used to draw conclusions with far greater significance than you perhaps imagined when gathering your data.

EQUIPPED FOR THE JOB

Although this is the age of the computer and of instant global communications, most scientific expeditions still use some fairly basic equipment to record data in the field. Observation strategies can begin with nothing more than a sharp pair of eyes, an underwater slate, and a good guidebook. An additional item that can be extremely useful is a magnifying glass. This simple tool greatly enhances enjoyment and has become essential for many divers around the world when diving complex reefs.

Underwater photography and videography, if correctly documented, can provide potentially valuable information. In an organized program, you may receive training in advanced techniques, such as the use of catch bags, nets, and quadrats, or in ways to ensure that the target region is sampled effectively, such as GPS systems and datum lines. Effective sampling and observation does not necessarily require high-tech gadgets, but an effective strategy to ensure scientific integrity in the results you achieve is essential.

A CAREER IN CONSERVATION

Should you wish to develop your interest in conservation techniques, there are many courses and reference works available. Formal qualifications may be sought at undergraduate and postgraduate level, although demand for places is extremely high. Jobs in this area are also relatively difficult to come by. However, the rewards for persisting in either scientific or conservation work can be immense.

The loggerhead turtle, an endangered species

Photography and videography

Since the dawn of diving, there has been interest in capturing images of the remarkable ecosystems and animals of the sea. Although underwater photography was initially a complex process, advances in technology have seen it become an accessible, absorbing pastime.

EARLY PIONEERS

The development of underwater filming and photography has mirrored the evolution of diving itself. Early pioneers made their own waterproof housings for cameras, but the invention of the O-ring (enabling watertight hydraulic seals) in 1937 and the Aqua Lung in 1943 convinced camera companies that they should make and sell their own designs.

What prompted this was the huge appetite for images of the underwater world. The rapid growth of recreational diving, and growing interest in underwater films among movie audiences, spurred the development of underwater film and stills photography equipment. Although the first underwater film was made in 1939 by diving pioneer Hans Hass—an instructional documentary for spear-fishermen called *Stalking Underwater*—the first truly commercial underwater movie was made by Jacques Cousteau in 1956. His feature film *The Silent World* was an Oscar-winning sensation that established him as the world's premier underwater filmmaker and explorer.

MINUTE SUBJECTS
One of the most interesting aspects of underwater photography is macro (close-up) work. Tiny creatures can make dramatic subjects.

GROWING MOVEMENT

Until the early 1980s, underwater stills photography and filming was seen as the preserve of either the dedicated amateur or full-time professional diver. Underwater photography was brought to the masses with the development of the Nikonos series of amphibious cameras—culminating in the Nikonos V, still viewed by many as the best underwater camera ever made. Underwater filming did not become a popular pastime until the mid-1990s, when digital video cameras came onto the market, making moving pictures cheaper and technically easier for amateurs to produce. Digital imaging also revolutionized stills photography underwater. Small, easy-to-use digital cameras with

STALKING A SUBJECT
Photographing marine species underwater requires guile and a good understanding of animal behavior and habitats.

many automated functions have made it simpler for divers to record high-quality images of their underwater experiences.

Despite this, the jump to producing professional-standard material remains as great as ever. Light behaves very differently underwater than in air (*see p.183*), and there are myriad extra factors to take into consideration when aiming for truly professional results. The work that you see on the page or the screen may well have involved lengthy planning, complex equipment, and the specialized skills of a professional post-production studio.

However, the same principles that guide professionals can enhance amateur work. You must have knowledge of the underwater environment and the behavior of target animals, and possess impeccable diving skills. Add to this mix some tips on technique when photographing and filming underwater, and you will be well placed to get the best from the dazzling range of camera and video gear available to divers.

CAMERA CONTROL
Though cameras, especially digital models, have become smaller and easier to use, it still takes great skill to wield one effectively underwater, especially for video footage.

USING LIGHT CREATIVELY
An understanding of how light behaves underwater is essential to capture a mood effectively.

STILLS PHOTOGRAPHY

The advent of the digital age has seen an explosion in the popularity of underwater photography, with more and more divers creating images of their dives for posterity. However, producing good results consistently is a craft, and this is what separates the professional from the amateur photographer.

CAMERA SYSTEMS

Underwater camera systems for stills photography can be divided into two categories. The first consists of truly amphibious cameras that do not require a housing to be used underwater, and can also operate on dry land. The pioneering Nikonos V camera is a famous example of the type. Amphibious cameras often have interchangeable lenses, and are available in both film and digital formats. These highly specialized cameras feature oversized controls for ease of operation underwater.

The second category, housed systems, are conventional cameras in waterproof cases. The cameras themselves can be divided into two sub-categories: compact devices, and SLR (single lens reflex) models.

Compacts come in film and digital formats, and are light and convenient to carry and use. Most models offer fully automated focus and exposure control. You cannot change the lens on a compact, but some have variable focal length (zoom) lenses. After-market add-on lenses that attach to the outside of camera housings are also available.

ANIMAL MAGIC
To capture the vibrant colors of marine creatures and environments, you will need a purpose-designed strobe lighting system.

SLR cameras, which are available in both digital and film versions, allow the use of a range of lenses and offer complete creative control over key functions such as aperture, shutter speed, and focus. For this reason they are invariably the choice of professional underwater photographers, although SLR-compatible ports, housings, and lighting systems can be very expensive.

CAMERA HOUSINGS

Both compacts and SLRs require a housing to protect them from water damage. These are normally made of plastic, though aluminum housings are also available for SLRs. Aluminum housings are strong and durable but expensive and bulky, and hide the camera within, making it harder to operate. Plastic housings are cheaper and allow you to see the camera, but can be less durable. They are especially vulnerable to scratches and abrasion.

NIKONOS CAMERA
Amphibious cameras like the Nikonos V have an integral water-resistant housing.

DIGITAL COMPACT
Fully automated functions make digital compacts easy to use, but limit creative control.

SLR CAMERA
Available in film and digital formats, SLRs are bulky but offer superior image control.

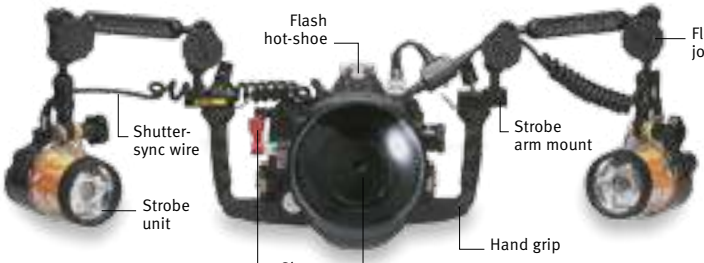

SLR HOUSING (FRONT VIEW)
Serious investments in their own right, large SLR housings offer excellent access to key controls.

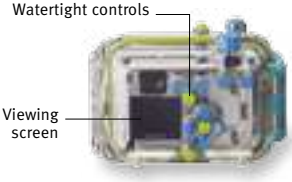

COMPACT IN PLEXIGLAS HOUSING
The housing features push-button mechanical controls for taking, reviewing, and even editing images underwater.

COMPACT WITH STROBE
A diffuser lens placed over the strobe softens the light it casts.

Bare strobe unit

UNDERWATER LIGHTING SYSTEMS

Lighting is one area where compact digital cameras fall short. Their built-in flashes are too close to the lens to avoid backscatter (*see p.183*). You can distance the light source from the lens by using an auxiliary strobe light, bolted to the housing. SLR-based and amphibious systems often use strobes in pairs, mounted on flexible arms to allow the subject to be lit from different angles. Knowing how to position strobes to best advantage and manipulate power settings is essential, though many cameras will automatically trigger the strobe to fire for the correct duration for the exposure.

HOUSING MAINTENANCE

Maintaining a housing is relatively straightforward. O-ring integrity should always be checked and the ring lightly lubricated with silicone grease. Make a final visual check before diving. Housings should be rinsed in fresh water as soon as possible after a dive. Operate the mechanical controls during rinsing to prevent a build-up of salt crystals. If water penetrates the housing, it will ruin the camera, so a few moments taken to prepare and rinse gear can prevent expensive damage.

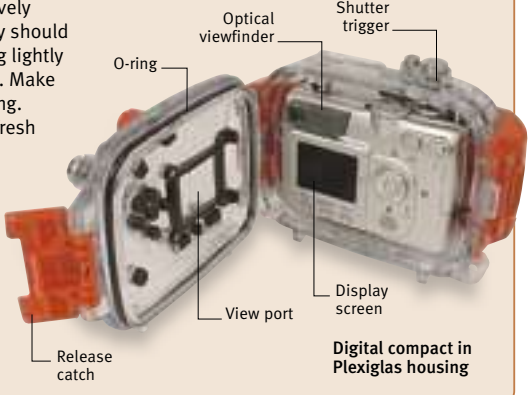

Digital compact in Plexiglas housing

MAKING FILMS UNDERWATER

Like underwater stills photography, videography has been revolutionized by the advent of digital cameras, which are small and easy to use, compared to film or analog video. Basic point-and-shoot underwater videography is quite easy with consumer-grade digital cameras. To advance beyond the novice stage, however, you need special skills, more advanced equipment, and a flair for using video editing software.

PRECISE POSITIONING
To be a good underwater film-maker, you need to perfect your buoyancy control and finning.

A robust housing will be necessary to protect the camera from water damage, while a basic underwater lighting system is essential for capturing good color footage, since water filters out various colors from natural light as depth increases (*see p.183*).

CHOOSING A VIDEO SYSTEM

Choice of camera is the first decision for novice videographers. Though analog video cameras are still available, digital is now the dominant technology. Digital cameras are available in digital tape, flash media, HDD (hard disk drive), and DVD recording formats. Look for cameras with three-chip processors (called 3CCD cameras). These divide red, green, and blue color information into separate processing paths, offering richer colors than cheaper one-chip models. At the top end of the market are cameras with HD (high definition) imaging, which produce video of astonishing clarity, due to the high rate at which they scan visual information.

FILMING SKILLS

Creating high-quality underwater video requires you to do more than merely jump in with everything set to automatic, press record, and film throughout the dive. Many automatic systems perform poorly underwater. For example, autofocus becomes inaccurate in low-contrast environments, resulting in hazy images. Careful manual control of focus and manipulation of white balance (which sets the baseline for color values in different light conditions) are key skills.

Your diving skills also must be excellent; you need to be a steady platform for the camera—camera shake is a common cause of ruined footage. The camera and housing should also be neutrally buoyant (they should neither sink nor rise underwater) for ease of

BASIC VIDEO CAMERA
Affordable cameras for home use usually feature automatic focus and exposure control, but may also have some manual options for skilled users.

ADVANCED DIGITAL CAMERA
Top-end video cameras have advanced features like high definition, multi-chip processing, plus image stabilization systems for sharp images.

Housing with lights

Basic light

PRESSURE HOUSING AND LIGHTS
Housings fall into two categories—electronic or manually controlled. The latter use plungers, rods, pins, and cogs to allow the camera controls to be manipulated from outside the housing; they require more skill, but have the advantage of avoiding electronic systems that can go wrong or short out in the event that the housing is flooded.

Bare housing

handling, and you should familiarize yourself completely with how your camera is operated, so its use becomes second nature underwater.

PRODUCTION AND EDITING

Thorough planning is the secret to good filmmaking, and video should be shot with an audience or editor in mind. Building a story around a given dive is not a complex process, but creates a target shoot list for the camera operator to work from, and gives the finished film an engaging narrative. You can place scenes in order and polish the results to a professional standard using video editing software that is commercially available for use on home computers.

KEEPING CLEAR
While filming, be careful to avoid causing physical damage to the organisms you have come to film. Avoid touching, or kneeling on, reefs.

KEY PRINCIPLES

Modern cameras and housings allow even the novice to achieve consistently good photographs and video footage. One key development of the last ten years has been the ability to immediately view the results—even during the course of the dive—and then make adjustments accordingly before the next shot. But throughout the history of underwater photography and videography, certain principles have applied as much to hardened professionals as they have to happy snappers. Following these simple rules can immeasurably improve even your first forays into photography and filmmaking underwater.

LINING UP FOR A SHOT
Good positioning relative to your subject can help make subjects stand out from the background.

GETTING CLOSER

The first principle when shooting underwater is to get as close to the subject as is feasibly possible. Even the most crystal-clear water contains a mass of suspended particles, and even if your eyes cannot pick them out, you can be sure your lighting system or lens will. Getting close to the subject means greater clarity of the resulting image, and this is one of the reasons why "close-up wide-angle" is such a popular approach. This uses a wide-angle lens at close range, allowing the whole subject to be viewed with the lens only a short distance away. For the same reason, "macro" (close-up) photography is often a good place to start if you are a beginner, since clear, well-lit macro images are relatively easy to achieve.

For the novice photographer or videographer, another excellent principle is to remember to take shots with an upward orientation. There is nothing worse than viewing an endless series of dull blue or green shots where the subject is lost in the background of reef or seabed. Shooting upward isolates the

SUNLIGHT THROUGH WATER
The shifting light patterns encountered underwater can make for dramatic photography, but are challenging to expose correctly.

WAITING FOR A SHOT
Many fish are inquisitive, but easily alarmed; good buoyancy control will allow you to hold still until a subject presents itself.

subject against the background of the open water, and creates a more striking contrast. The composition of such shots can still include reef or coral, wall or wreck; however, the subject itself should be consistently easier to spot.

USING LIGHT SOURCES

Effectively lighting a subject is one area where most amateurs struggle. It is also one of the areas where modern compact digital cameras built-in flashes will let a picture down. If the light source is too close to the lens, particles in the water will reflect the flash directly back into the lens. This creates a phenomenon known as "backscatter," a snowstorm effect of white dots on the photograph. One of the most basic prerequisites of taking good underwater images in anything other than perfect underwater conditions is an ability to position a strobe away from the lens. This allows the flash to come in at an obtuse angle, and helps to prevent backscatter from spoiling a shot.

These simple principles do not, unfortunately, guarantee good images every time, but they are useful for anyone starting out in photography or filming. Trial and error will help to refine your skills over time, and the best advice is to jump in, remember the basics, snap or film away, and enjoy!

DEPTH AND COLOR LOSS

Water absorbs different wavelengths of light as depth increases, causing images taken without artificial illumination to seem washed-out and dominated by blue tones. Reds fade from visibility at a depth of 10–15 ft (3–5 m), oranges at 35 ft (10 m), and yellows around 50 ft (15 m); by 80 ft (25 m) greens have almost disappeared, and by 100 ft (30 m) everything appears deep blue or black to the naked eye.

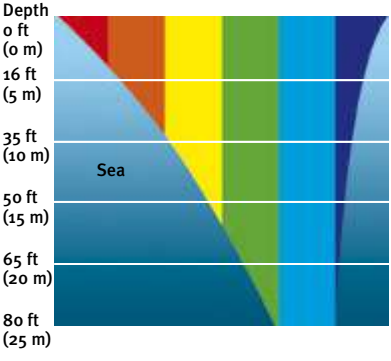

Night diving

To many, the idea of diving at night may seem a little crazy, but there is far more to it than just the thrill of exploring in the dark. Night diving reveals a whole host of underwater organisms that lie hidden during the day and only become active at dusk.

LIGHT AND COMMUNICATION

Night diving is obviously less simple than diving during the day, but when properly organized, it is relatively straightforward. A powerful flashlight will easily light your way and the creatures around you, revealing marine life in its true colors. However, if you cover your flashlight (never switch it off; the bulb may blow when you turn it back on), you will be surprised at how much light there is underwater. Many creatures use phosphorescence at night, and as you move through the water, you will cause plankton to release tiny pulses of light, leaving beautiful glowing wakes trailing behind you.

NIGHT ON THE OCEAN
When diving on a clear, moonlit night such as this, you may be surprised at just how effectively the Moon is able to illuminate the underwater environment.

The main restrictions compared to daytime diving relate to communication, not only between divers, but also—and perhaps more importantly—between divers and boat cover. You will be able to monitor your buddy's location from the position of their flashlight beam. Signaling to your buddy can be done in the normal way, but while shining your light on to your hand as you make the sign. Attract your buddy's attention by waving your light, but take great care not to shine it at their face, since it can take as long as 15 minutes for their eyes to fully readjust to the near-darkness and restore night vision.

At the surface, avoid flashing your light unless you need to alert the surface cover in an emergency. If you are diving from the shore, leave a light source there—for example, a bright boat light, flashlight, or beacon—so you can easily find your way back to land after surfacing. Exit points should be marked with colored lights, to distinguish them from any other light sources.

You will find that most gauges and computers are easily readable by flashlight, and just shining the beam directly on to your gauge will make it luminous for a while.

LIGHTING TIPS

It is vital that all divers carry a flashlight, and at least one other as a backup. Take spare batteries; if using rechargeables, make sure they are fully charged. Use hand-held lights, not head-mounted ones: one look at your buddy with a head-mounted light will ruin their night vision. A glow-stick attached to the back of your suit, while not essential, will increase your own visibility.

Hand-held flashlight

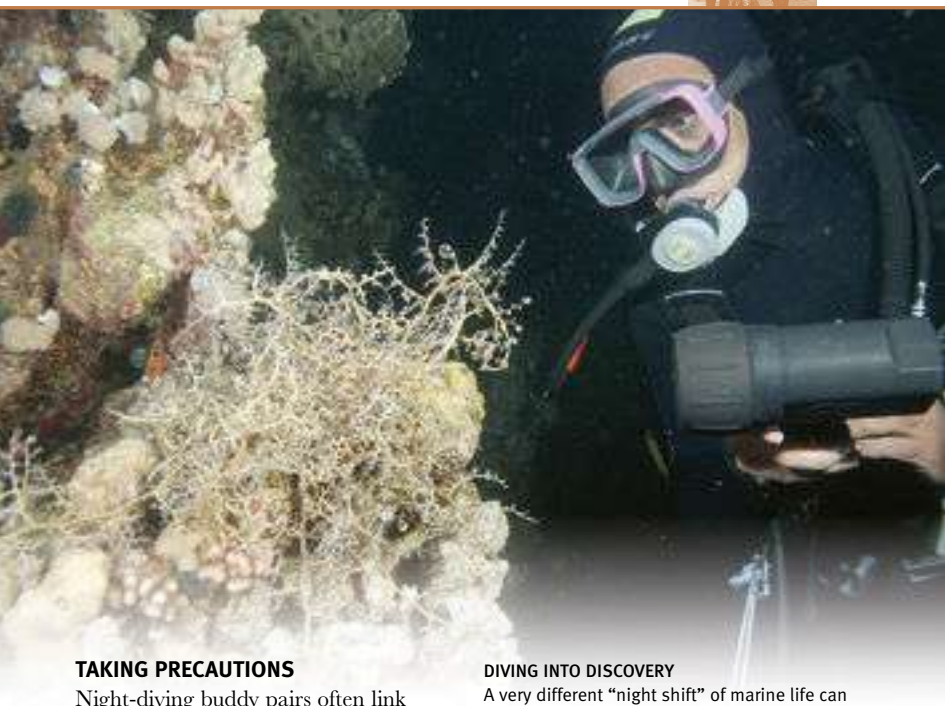

DIVING INTO DISCOVERY
A very different "night shift" of marine life can be seen during nocturnal dives, like this outlandish basket star.

TAKING PRECAUTIONS

Night-diving buddy pairs often link themselves together with a short line. While not essential, this is recommended for novice night divers and in poor visibility. If you are diving on a wreck to which a guideline and buoy are attached, secure a strobing marker light to the base of the line so it can be easily located in the dark. When choosing a dive site, opt for something simple, such as a shallow wreck or reef with plenty of distinctive features. An interesting daytime dive may be worth revisiting to investigate marine life that emerges at night. To minimize the risk of incidents, do not dive too deep, since any problems will be much harder to cope with in the gloom. If diving a more open site, use a surface marker buoy with marker light attached to identify it to any boat cover. Be careful not to lose contact with surface cover: getting lost at night could mean you are not recovered until morning.

FLASHLIGHT SIGNALS

During a dive, your torch beam will alert your buddy to your position. You can also make simple signals with your flashlight beam on the seabed to communicate with your buddy. At the surface, wave your torch vigorously from side to side if you require immediate assistance.

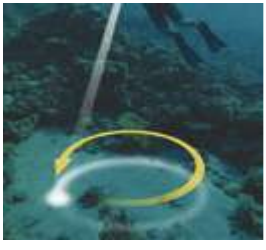

SIGNALING "OK"
Tracing a large circle on the seabed communicates to your buddy that you are OK.

SIGNALING FOR ATTENTION
A rapid side-to-side movement means that you want to show or tell your buddy something.

Cavern diving

Rich with opportunities for dramatic dives, caverns offer a bridge between ordinary open-water diving and the far more specialized activity of cave diving. Nonetheless, diving in caverns demands careful preparation and respect for potential risks.

WHAT IS CAVERN DIVING?

A cavern is a large, enclosed area of rock or coral that is of sufficient size to allow a diver to enter and move around. These fascinating environments can be extremely dramatic, are often eerily lit, and may play host to an array of animals not encountered on the open reef.

Unlike the more extreme sport of cave diving, in which divers penetrate beyond the reach of natural light, in cavern diving the exit is always clearly within view. However, cavern diving still carries inherent risks because the diver does not have direct access to the surface in the event of an emergency. If simple safety drills are not followed at all times, there is the potential for very serious problems to develop when diving even in simple caverns.

CAVERN DIVING EQUIPMENT

A flashlight is an essential item for cavern diving, but it is also sensible to bring along a backup light. This need not be a large, bulky model but must be able to provide enough light if your main flashlight malfunctions. Be sure to keep batteries freshly charged. An alternative air source, such as a pony bottle (*see p.68*), is also an excellent idea in caverns, since the diver cannot ascend directly to the surface if the primary air source suffers an unexpected failure.

A backup flashlight

SAFETY AND SKILLS

The first rule of cavern diving is to prevent your natural curiosity from turning a cavern dive into a cave dive! The exit should be clearly visible at all times—if not directly in your line of sight, then at least the natural light of the cavern entrance should remain visible. For many divers, the temptation to explore deep into cave systems is strong; however, the moment the exit is lost from view and natural light fades, the dive becomes a cave exploration, which should not be attempted without appropriate equipment and special safety and skills training.

Certain skills are essential when diving larger caverns. Buoyancy control is important, particularly when there is a silty substrate on the cavern floor.

ON THE THRESHOLD
A diver hovers at the entrance of a freshwater cavern in Australia. Following basic safety rules is essential at such sites.

INNER SPACE
A diver shines light on the interior of a sea cave. Caverns have a mysterious ambience quite different from that of open diving environments.

It is also advisable to use a finning technique known as the frog kick (*see p.123*). This is a gentle, circular fin-stroke that helps you avoid stirring up sediment. One misplaced kick from a fin can turn a cavern into an impenetrable soup of silt and lead to disorientation and panic. For this reason, and as a matter of plain common sense, it may also be appropriate to tie off a line at the entrance of larger caverns (*see p.127*) to provide a direct route to the exit in case of reduced visibility. It is also strongly advisable not to enter a cavern without local knowledge or a guide; there may be specific issues about a cavern system and the hazards it contains—such as seasonal changes in visibility—that do not appear in any standard dive guidebooks.

Ice diving

Diving under ice might not sound like an attractive proposition, but it is a uniquely exciting experience. Visibility is typically excellent beneath ice, with soft light filtering through from the world above. Ice diving can also take you to breathtakingly beautiful parts of the world.

UNIQUE ENVIRONMENT

Ice diving can be done in the sea or in frozen lakes. The aquatic life you encounter will depend on where you are in the world and whether you are in a marine or freshwater environment. Visibility is often stunning; the ice provides a protective layer over the water, so it is not churned up by the wind, and low light levels under the ice also help to inhibit algal growth.

COLD-WATER GEAR
A diver waits to enter the water wearing special equipment designed to cope with conditions of extreme cold.

KNOWING THE ROPES

If you are planning to go ice diving, you will need to undergo special training. This will teach you not only the safety aspects of diving while tethered on a rope under ice, but also how to cut the ice hole through which you enter the water, and how to tend safety ropes for other divers while you remain on the surface in a support role.

Ice diving requires a minimum of four to six people per dive. There will normally be two divers operating as a buddy pair (*see pp.110–11*), and two people to tend their lines. It is advisable to have at least one extra person on standby. It is standard practice for each diver to be tethered separately, and it is not advisable to tether two divers on one line, because if anything happens to this rope, then both divers can be lost. Lines should always be tied using a bowline knot, which is very secure.

SIGNALS AND SAFETY

Signaling to your tender is an important part of ice diving. Since you are diving in extreme conditions, it is vital that he or she knows you are OK throughout the dive. You will learn about giving rope signals during training.

Apart from the obvious dangers of diving in an enclosed environment, you must also be careful to monitor your core body temperature. Repeat dives

SPECIAL PROCEDURES
Ice divers rely on a surface support team (*above*) for safe diving under the ice.

should only be undertaken when you are fully warmed up. You should also ensure that you eat high-energy, hot food to replace the calories you will burn during the dive and in keeping warm afterward.

FROZEN DEPTHS
A diver films the fascinating world beneath the surface. Many unique animals live in cold seas.

SPECIAL GEAR

Ice diving requires special equipment. You will need an environmentally sealed regulator set (*see p.63*), which is designed to be less susceptible to freezing. You will be taught to submerge this in the water before diving, to acclimatize it to the temperature. You must only breathe from it when submerged, to reduce the risk of free-flow (an uncontrolled release of air due to a frozen mechanism). It is also recommended that your primary air source and octopus are mounted on separate first stages (*see p.61*) to ensure that one will work if the other freezes. Your cylinder will need a special pillar valve to accommodate this.

You will need to wear warm undergarments beneath your drysuit, thick neoprene gloves or a pair of waterproof dry gloves (*see p.54*), and two thick hoods for extra warmth.

Freediving

Once simply a fishing technique, freediving—diving using only the air carried in the lungs—has become increasingly popular as a sport. Extreme athletes now constantly push the perceived limits of human performance, diving to remarkable depths on a single breath of air.

PURE SPORT

Freediving is not exclusively the preserve of the extreme athlete: anyone who has snorkeled and duck-dived to explore the seabed could describe themselves as a freediver, making this the most accessible of all diving disciplines. Many would argue that it is also the purest form of diving, requiring only the minimum amount of equipment and relying entirely on discipline, courage, and physical conditioning.

The roots of freediving stretch back to when the human race took its first tentative steps into the sea, diving into the water to harvest food and explore the seabed. Modern freediving, by contrast, is a disciplined sport, with record attempts highly regulated by the two governing bodies—A.I.D.A. (Association Internationale pour le Développement de l'Apnée) and F.R.E.E. (Freediving Regulations and Education Entity). Divers compete in a number of categories, two of which usually take place in a swimming pool: static breath-hold—a simple endurance test with a timed immersion in a shallow pool—and dynamic breath-hold, a test of the distance divers can swim using only a single breath. In open water, competitors are judged on the depths they can reach and return from safely.

Freediving naturally carries inherent risks—the most advanced and extreme disciplines demand that the body's supply of oxygen is completely used on a dive, and losing consciousness is always a danger. Exploration of freediving limits should never be conducted alone, and only after attending one of the many introductory courses available to those interested in taking up the sport.

EQUIPMENT

Freediving does not require much equipment, but specialized items, such as streamlined suits and fins of variable flexibility, are used to maximize the speed of movement through the water. Masks with low internal volume are also useful, as they allow divers to conserve air that would otherwise be used to exhale into a standard mask to counteract mask squeeze (*see p.115*).

Specialist mask

STATIC BREATH-HOLD COMPETITION
Divers compete for the length of time they can hold their breath. World-class competitors can achieve times of over seven minutes.

TOUCHING THE DEEP
A freediving competitor reaches for the tag that proves that the designated target depth has been reached—this must be presented at the surface for inspection by judges.

DIVING DEPTHS

Competitive depth records are judged in strict categories, defined by precise parameters such as the use of weights or fins. In "No Limits"—the category in which the highest-profile records are set—the diver is dragged down by a weighted sled, and rockets to the surface using an airbag.

Depth obtainable by an average club freediver.

100 ft (30 m)

Competent amateur freediver

Depth obtainable by a competitor in the "Constant Weight" category, which allows the use of fins and ballast. Top divers can reach 330 ft (100 m)—roughly the height of the Statue of Liberty in New York or Big Ben in London.

330 ft+ (100 m+)

Constant Weight

Depth obtainable by the top "No Limit" class of freedivers.

650 ft+ (200 m+)

Underwater archaeology

While marine archaeology is mainly confined to the academic and scientific world, it could be said that every diver who explores a wreck has an archaeological interest and responsibility. Both the seabed and many freshwater sites hold compelling historical secrets.

RULES FOR THE AMATEUR

Although every dive has the potential to reveal historic artifacts, targeted methodical surveying is an altogether more serious business, as is the correct means of excavating and recording any finds. Archaeological divers can therefore be placed in one of two distinct categories—the serious professional diving archaeologist, and the enthusiastic amateur.

Although this is a field in which amateurs can make a very valuable contribution, it is essential that they do so only through involvement in an organized project, conducted by a recognized and authorized group. Far too many sites have been destroyed, and artifacts removed, by what amounts to vandalism and theft—the equivalent on land might be digging a historic site or burial ground without permission, which is unthinkable, of course.

There are many organizations that welcome help from amateur volunteers, and will supervise and guide you. The thrill of discovery really does reach a new level when it is shared with individuals who understand the full significance of a find. You may find that more scope for joining an organized project exists if you first undertake some training, perhaps by volunteering on digs on land.

Learning professional marine excavation and recovery techniques can be fascinating for the amateur. Archaeologists go to great lengths to avoid damaging a surveyed location, and even when they locate an item of interest, try to avoid intrusive collecting techniques or digging.

MORE THAN YOU BARGAINED FOR

By taking an artifact from a dive site, you destroy its integrity and hinder efforts to learn about past peoples. And that may not be the end of the matter. In many parts of the world, sites of marine archaeological interest are strictly regulated. The recovery of artifacts is—quite rightly—taken very seriously, and penalties for illegal activity and for removing items can be severe.

Amphorae, used to store oil and wine, recovered from a Turkish site

RELIC OF CONFLICT
Divers inspect the remains of a Japanese fighter aircraft shot down in the Pacific during World War II.

WHAT TO EXPECT

Marine archaeological expeditions are usually initiated because research or local knowledge has pointed to a region that has archaeological potential. Modern GPS (Global Positioning System) equipment has made pinpointing such sites much easier. On reaching the site, a detailed survey is generally first carried out. This may involve using a datum line (a fixed reference line on the seabed from which measurements are taken), or pegging out a grid. Precise recording of the location and state of the site before digging is vital, and is usually conducted through photography or extensive sketching. Small test digs may take place before a full-scale excavation occurs, with the exact site of each find carefully logged, and the finds themselves sketched or photographed. Preservation and restoration techniques are extremely specialized for such finds. Marine archaeology is a long way from the derring-do of Indiana Jones, requiring organization, attention to detail, fine diving skills, and persistence.

CAREFUL SURVEY
Meticulous recording of survey data is one way in which amateurs can help on a professional dig.

Technical diving

The desire for ever-deeper dives and longer periods underwater has led to the development of special gas mixtures. The use of these in technical diving offers an exciting range of new skills and marine environments for qualified technical divers to explore.

PUSHING BACK BOUNDARIES

Scuba divers are limited, by simple physical laws, to a certain amount of time under water (dependent on the depth of the dive) and an overall maximum depth. Training agencies recommend that recreational divers using compressed air go no deeper than 130–165 ft (40–50 m). Beyond this, nitrogen can build up dangerously quickly in the body, greatly increasing the risk of decompression sickness (DCS; *see pp.142–43*). In addition, the increased levels of nitrogen in the bloodstream below 165 ft (50 m) can lead to severe nitrogen narcosis (*see pp.100–01*). The oxygen content of air, too, becomes increasingly toxic to the nervous system the deeper you dive (*see pp.96–97*).

To extend dive times and allow exploration to greater depths, special mixtures of gas have been developed in which the balance of oxygen and nitrogen has been altered, or that contain new gases to dilute the impact of either oxygen or nitrogen. The use of these mixes is known as technical diving, and involves three types of gas mixes: nitrox, which allows longer dive times, and trimix and heliox, to reach greater depths.

EXTENDING DIVE TIMES

Normal air is composed of around 21 percent oxygen and 79 percent nitrogen. Nitrox is a gas mixture that contains a higher percentage of oxygen and a lower nitrogen content than normal, allowing longer dive times before the diver's body stores up too much nitrogen, and also reducing diver fatigue after the dive. Nitrox can be mixed in various proportions, from

UNDERGROUND ODYSSEY
A diver emerges from the water wearing the complex equipment required for cave diving, including tanks filled with special gas mixes, and sophisticated dive computers.

THE COMPLETE ENCYCLOPEDIA OF MINERALS

Description of over 600 Minerals from around the world

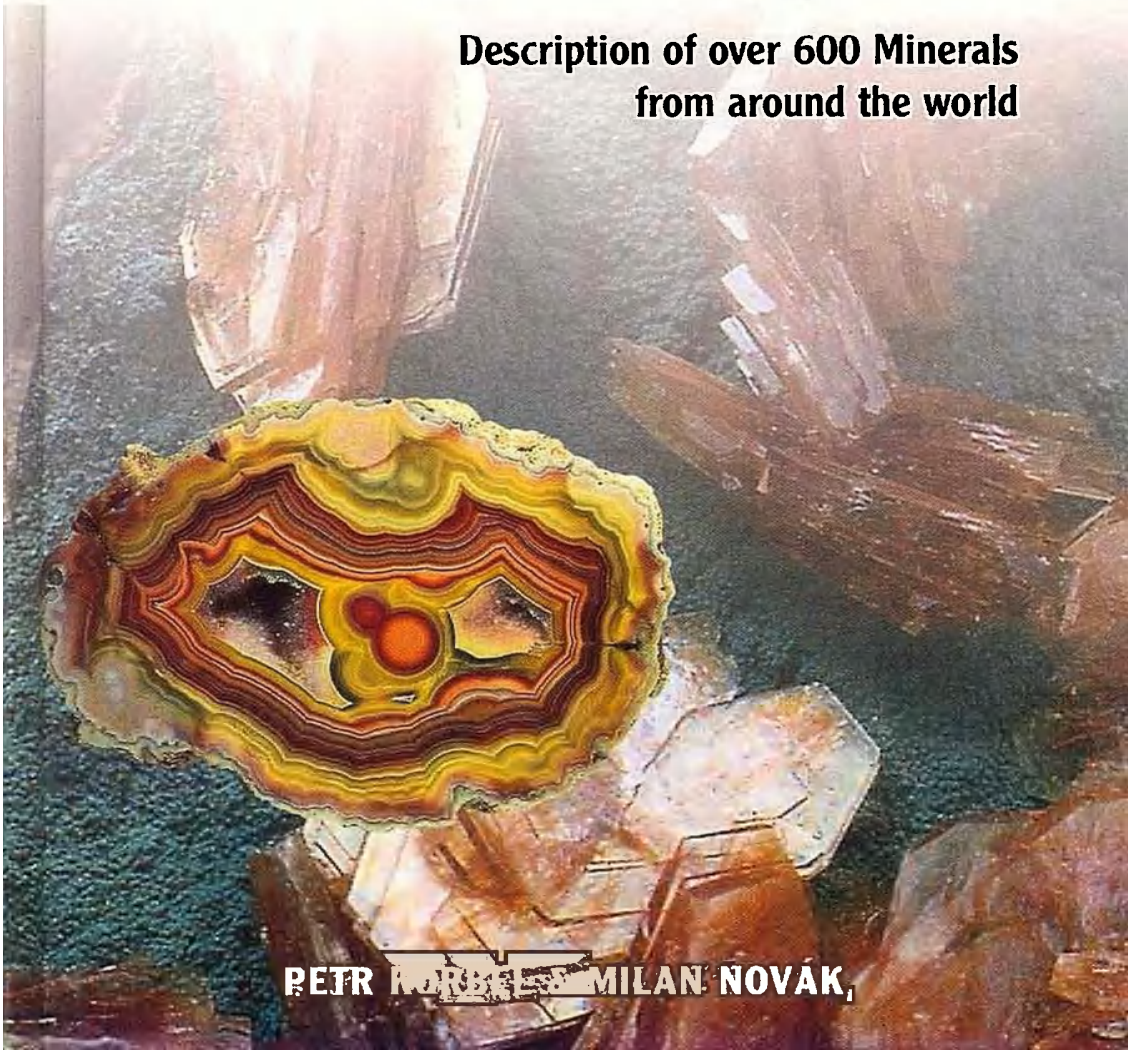

PETR KORBEL & MILAN NOVÁK

Explanation of the abbreviations used in the book

C – color
S – streak
L – luster
D – diaphaneity
DE – density
H – hardness
CL – cleavage
F – fracture
M – morphology
LU – luminescence
R – remark
X – crystal
XX – crystals

In the figure captions, the bigger dimension is always mentioned

Scale of the frequency of the occurrence

● - very rare
●● - rare
●●● - uncommon
●●●● - abundant
●●●●● - common

© 1999 Rebo International b.v., Lisse, the Netherlands
© 2001 Grange Books PLC, United Kingdom

Text: Petr Korbel, Milan Novak
Fotografie: Fotobanka Granit (100), Petr Korbel (220), Geza Kulcsar (16), Jeffry A. Scovil (331), Sandor Szakall (34)

Production and layout: Granit Ltd., Praha

ISBN 1 84013 404 6

All rights reserved.

Copying of the entire book as well as of its parts in any way is forbidden by law without prior written consent of the publisher.

Contents

	Introduction	7
1.	Elements	9
2.	Sulphides	19
3.	Halides	67
4.	Oxides	75
5.	Carbonates	113
6.	Borates	131
7.	Sulfates	137
8.	Phosphates	155
9.	Silicates	193
10.	Organic compounds	281
11.	Rocks, meteorites and tektites	285
	Recommended literature	291
	Index	293

Introduction

The ever increasing number of publications about minerals reflects a growing interest in nature. Most of those publications though only deal with a few dozen of the most common minerals or gemstones. This book fills the gap by also featuring less common and rare minerals. The authors describe over 600 mineral species and varieties, illustrated with about 750 color photographs. In choosing illustrations of particular minerals, aesthetic criteria such as size of crystal and color played a role in addition to their importance and distribution in nature. This book includes some rare minerals, known only from one locality, because they form very attractive crystals or aggregates. There are minerals known to humankind since prehistoric times such as quartz and gold, but also minerals first described quite recently like rossmanite. The photographs show well-formed and colorful crystals but many aggregates, which are more common in nature, are also included. The minerals in the book are listed according to the mineralogical system of Hugo Strunz, in his book *Mineralogische Tabellen* in 1978. The chemical formulae of individual minerals follow the form of Glossary of *Mineral Species* 1995 by M. Fleischer and J.A. Mandarino. The information is complemented in both cases with the latest knowledge from scientific literature, such as new nomenclature of amphiboles, micas and zeolites.

The mineral descriptions cover the basic physical and chemical data, including chemical formula and crystal system. The data provided correspond mainly to the end-members. The less common valence of the chemical elements is marked in the chemical formula (Fe^{3+}, Mn^{3+}, As^{3+}, Mn^{4+}, Pb^{4+}). Where an element features in both valences in the mineral, they are both marked (e.g. ilvaite, braunite).

The origin of individual minerals is described in detail. We chose for a relatively simplified scheme because the normal complexity cannot be described here in detail. Minerals can be distinguished as either primary (resulting directly from a solidifying of magma, crystallizing of an aqueous solution or metamorphism – re-crystallization in a solid state) and secondary (resulting from alteration of the original mineral, e.g. during its oxidation or reduction under low temperature and pressure close to the surface of the Earth). Primary minerals are divided into following groups: 1. magmatic, when a mineral crystallizes directly from a melt (it includes magmatic and effusive igneous rocks, including granitic and alkaline syenite pegmatites and meteorites): 2. sedimentary, when a mineral crystallizes during a process of diagenesis or from hydrous solutions under normal temperature (clastic, organic and chemical sedimentary rocks); 3. metamorphic, when a mineral crystallizes during metamorphic processes in a solid state at a wide range of temperature and pressure (it includes regionally and contact metamorphosed rocks and skarns); 4. hydrothermal, when minerals crystallize from aqueous solutions and fluids under high to low temperatures (it includes ore and the Alpine-type veins, cavities in volcanic rocks, minerals and rocks, hydrothermally altered under high temperature, e.g. greisens).

Secondary minerals are divided into following groups: 1. oxidation, when minerals result from the oxidation (weathering) of the primary minerals in the oxidation zone of ore deposits and other rocks (it includes the origin of malachite and azurite during the chalcopyrite oxidation, also the origin of secondary phosphates in granitic pegmatites during the oxidation of primary phosphates); 2. cementation, when minerals result from the reduction of the primary minerals (the origin of native copper and native silver under the reduction conditions in the cementation zone of ore deposits). This classification is very much simplified of course, because in many cases we cannot readily determine a specific origin of a particular mineral. This relates to minerals that crystallize under conditions which approximately represent a transition between separate phases of the origin, such as the magmatic or hydrothermal origins of elbaite in the pegmatite cavities; the metamorphic or hydrothermal origins of grossular in skarns; the magmatic or metamorphic origins of cordierite in migmatites; and the hydrothermal or secondary origins of some phosphates in granitic pegmatites etc.

With the localities for individual minerals we have tried to list the most important worldwide localities regardless of their recent production but we have also included recent discoveries since these may produce important mineral specimens. Where a mineral has an important use this is listed at the end of mineral description. We would like to acknowledge all who contributed in any way to the production of this book, particularly those private collectors and institutions which loaned minerals for photography. We hope those fascinated in the world of minerals and of nature will find this book a fascinating source of information.

This book is dedicated to the memory of Dr. Jaroslav Venek, who was of extraordinary influence to several generations of Czech and Slovak mineralogists and mineral collectors with his enthusiasm for mineralogy and attitude to life.

Marcasite, 100 mm, Misburg, Germany

1. Elements

Copper
Cu

CUBIC ● ● ● ● ●

Properties: C – light pink to copper-red, it darkens and covers green to black in air; S – red; L – metallic; D – opaque; DE – 8.9; H – 2.5 – 3; CL – none; F – hackly; M – cubic crystals and its combinations, dendritic aggregates, sheets, slabs, massive.
Origin and occurrence: Primary hydrothermal copper is mainly related to basic igneous rocks; it is also common as a product of supergene cementation. It is associated with cuprite, malachite, azurite, silver, chalkocite, bornite and other minerals. The largest accumulations of primary copper are in the Keweenaw Peninsular, Lake Superior, USA, the largest being 15 x 7 x 3m (approx. 50 x 23 x 10ft)and weighing 420 tons. Fine crystals up to 50 mm(approx. 2in) also occur there, as do calcite crystalwith copper inclusions. Superb supergene coppercrystals come from many localities like Tsumeb, Namibia and Chessy, France. Crystals up to 140mm (5½ in) long occurred in the Ray mine and in Bisbee, Arizona, USA. Very fine spinel-law twins up to 5 cm in size and dendritic aggregates come from Mednorudnyansk, Ural mountains, Russia; crystals up to 30 mm ($^{15}/_{16}$ in) were found in Dzhezkazgan, Kazakhstan. Fine specimens of copper, associated with cuprite, azurite and malachite occurred in Rudabanya, Hungary. *Application:* electronics, electrical engineering, ingredient in gold alloys.

Copper, 42 mm, Cornwall, UK

Silver, 52 mm, Freiberg, Germany
Copper, 120 mm, Keweenaw Peninsula, U.S.A.

Silver

Ag

CUBIC ● ● ●

Properties: C – silver-white, tarnishes gray to black; S – silver-white; L – metallic; D – opaque; DE – 10.5; H – 2.5 – 3; CL – none; F – hackly; M – cubic crystals, dendritic aggregates, wires, leaves, massive.

Origin and occurrence: Hydrothermal in ore veins and also of secondary cementation origin in association with acanthite, stephanite, proustite, pyrargyrite, copper and many other minerals. The best specimens of crystallized and wire silver come from Kongsberg, Norway, where wires up to 400mm (16 in) long and crystals up to 40 mm (1 9/16 in) in size have been found. Beautiful specimens of wire silver with wires over 100 mm (4 in) long are known from Freiberg, Schneeberg and St. Andreasberg, Germany. Wires several cm long were also found in Pribram and Jachymov, Czech Republic. Dendritic aggregates from Batopilas, Chihuahua, Mexico, reached up to 150 mm (6 in). Crystals and aggregates of silver, grown together with copper are genetically unique in the basalt cavities in the Keweenaw Peninsula near Lake Superior, Michigan, USA. Wires, up to 100 mm (4 in) long, come also from the San Genaro Mine in Huancavelica and Uchucchaqua, Peru. New finds of silver wires, up to 150 mm (6 in) long, have been made in Dzhezkazgan, Kazakhstan.

Application: photographic industry, jewelry, electronics.

Silver, 55 mm, Schwarzwald, Germany

Gold, 48 mm, El Dorado, California, U.S.A.

Gold, 68 mm, Eagle's Nest Mine, California, U.S.A.

Gold

Au

CUBIC ● ●

Properties: C – gold-yellow; S – yellow; L – metallic; D – opaque; DE – 19.3; H – 2.5-3; CL – none; F – hackly; M – octahedral and cubic crystals, skeletal and dendritic aggregates, leaves, nuggets.

Origin and occurrence: Primary hydrothermal in ore veins, also in contact metamorphic deposits and pegmatites. Placer deposits are secondary. It occurs with pyrite, arsenopyrite, quartz, sylvanite, calaverite, krennerite and other minerals. Beautiful leaves and crystals of gold found in many localities in California, USA (Colorado Quartz mine, Nigger Hill and others). Fine leaf gold comes from Rosia Montana, Romania. The best crystals, skeletal octahedra, up to 50 mm (2 in) have been found in alluvial sediments near Gran Sabana, Roraima Shield, Venezuela. Gold wires up to 110 mm (4 5/16 in) long were very rare in Ground Hog mine, Gilman, Colorado, USA. The largest known sheets of crystallized gold occurred in the Jamestown mine, California, USA, where a cavity, which yielded 49 kg (108 lb) of golden leaves, was discovered on 26.12.1992. The largest measures about 300 mm (11 13/16 in) and has about 25.79 kg (56 lb 13 oz) of gold on it. Typical aggregates of fine gold wires come from Farncomb Hill near Breckenridge, Colorado, USA. Fine crystals were also found in Berezovsk, Ural mountains and in the Lena River basin, Siberia, Russia. Fine scales and larger nuggets from placer deposits were found in Klondike, Alaska; Tuolumne County, California, USA and in Ballarat, Victoria, Australia. Fine dendritic aggregates occurred in the Hope's Nose, Devon, UK. A unique find of leaves up to 100 mm (4 in) was made in Krepice near Vodnany, Czech Republic.

Application: practically the only source of gold as a metal; used in jewelry, electronics and medicine.

Mercury

Hg

TRIGONAL ● ●

Properties: C – tin white; L – metallic to adamantine; D – opaque; DE – 13.6; M – liquid at temperatures above -39°C (-38.2°F); R – very poisonous fumes.

Origin and occurrence: Hydrothermal in low-temperature ore deposits, also connected with hot springs. It is associated with cinnabar, calomel and other Hg minerals. It occurred in Almaden, Spain;

Mercury, 30 mm, Socrates Mine, California, U.S.A.

Idria and Avala, Serbia; New Almaden and New Idria, California, Terlingua, Texas, USA; Dedova hora, Czech Republic and Rudnany, Slovakia as droplets and liquid cavity fillings.
Application: chemical industry, measuring instruments, metallurgy.

Moschellandsbergite
Ag_2Hg_3

CUBIC ● ●

Properties: C – silver-white; S – silver-white; L – metallic; D – opaque; DE – 13.5; H – 3.5; CL – good; F – conchoidal; M – dodecahedral crystals and their combinations, granular, massive.
Origin and occurrence: Hydrothermal in low-temperature deposits, associated with cinnabar, tetrahedrite, pyrite and other minerals. Crystals, several mm long, were found in Moschellandsberg, Germany. They are also known from Sala, Sweden; Les Chalanches, France and Brezina, Czech Republic.

Moschellandsbergite, 6 mm grain, Moschellandsberg

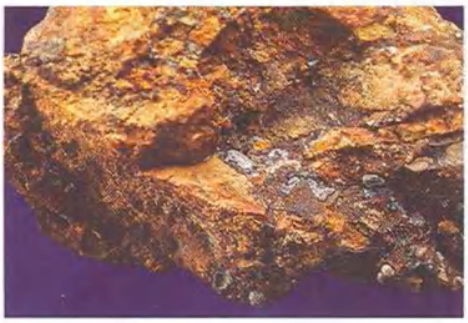

Lead
Pb

CUBIC ● ●

Properties: C – gray-white, tarnishes to lead-gray and gets dull; S – lead-gray; L – metallic; D – opaque; DE – 11.3; H – 1.5; CL – none; M – octahedral and cubic crystals, massive.
Origin and occurrence: Hydrothermal, also sedimentary (authigenic), associated with willemite and other minerals. The best specimens with crystals up to 40 mm (1 9/16 in) in size, come from LDngban, crystallized also from Pajsberg, Sweden. Octahedra, up to 10 mm (2/5 in), are described from El Dorado, Gran Sabana, Venezuela. It also occurs in Franklin, New Jersey, USA and Jalpa, Zacatecas, Mexico.

Iron
Fe

CUBIC ● ●

Properties: C – steel-gray to black; S – gray; L – metallic; D – opaque; DE – 7.9; H – 4; CL – perfect; F – hackly; M – crystals, granular, massive.
Origin and occurrence: Terrestric iron occurs mainly in basic rocks, but it is also known from carbonate sediments and the petrified wood. The most famous locality is Blaafjeld near Uivfaq on Disko Island, Greenland, where masses up to 20 tons were found.

Lead, 53 mm, Langban, Sweden

Iron, 80 mm, Bühl, Germany

Chunks, weighing over 10 kg (22 lb) come from Bühl near Kassel, Germany. Impregnations of iron in dolerite occur in the Khuntukun massif and masses up to 80 kg (176 lb) are known from Ozernaya Mt., Siberia, Russia.

Platinum

Pt

CUBIC ● ●

Properties: C – steel-gray to dark gray; S – steel-gray to silver-white; L – metallic; D – opaque; DE – 21.5; H – 4-4.5; CL – none; F – hackly; M – cubic crystals, nuggets, grains and scales.

Origin and occurrence: Platinum occurs in magmatic segregations, together with chromite, olivine and magnetite in ultrabasic rocks; secondary in placers. The best crystals up to 15 mm (19/32 in) come from Konder in Khabarovsk Region, small crystals, but mainly nuggets, weighing up to 11.5 kg (25 lb 5 oz) found in the Tura River basin near Turiinsk, Ural Mountains. Primary platinum is known from deposits in the vicinity of Nizhniy Tagil, Ural Mountains, Russia; from the Onverwacht mine, Bushveld, South Africa and Sudbury, Ontario, Canada. Fine smaller nuggets, weighing up to 75 g (165 lb) were found in the Trinity River sediments in California, USA and the Choco River sediments in Columbia.

Application: chemical industry, catalytic convertors, rocket industry.

Platinum, 8 mm, Konder, Russia

Arsenic, 45 mm, Alden Island, Canada

Arsenic
As

TRIGONAL ● ● ●

Properties: C – tin-white, tarnishes quickly to black; SB tin-white; L – metallic; D – opaque; DE – 5.8; H – 3.5; CL – perfect; F – uneven; M – rhombohedral crystals, botryoidal aggregates, granular, massive.
Origin and occurrence: Mainly hydrothermal, together with other As minerals. It forms massive veins, up to 200 mm (7⅞ in) thick, with botryoidal surface in Jachymov, Czech Republic and in Freiberg, Germany. In Akatani, Japan, spherical aggregates consisting of small crystals were found. It is also known from Sacarimb, Romania. Botryoidal aggregates of arsenic with leaf gold were found in the Royal Oak mine, Coromandel, New Zealand. Crystals of metamorphic origin come from Sterling Hill, New Jersey, USA.

Stibarsen, 58 mm, Atlin, Canada

Antimony, 37 mm, New Brunswick, Canada

Stibarsen
SbAs

TRIGONAL ● ● ●

Properties: C – tin-white to gray, tarnishes black; S – gray; L – metallic, sometimes dull; D – opaque; DE – 6.3; H – 3-4; CL – perfect; M – indistinct crystals, botryoidal aggregates.
Origin and occurrence: In pegmatites with antimony, stibiotantalite and microlite; hydrothermal in ore veins withpyrargyrite, proustite, pyrostilpnite and dyscrasite. Beautiful botryoidal aggregates of stibarsen (previously labeled as allemontite), up to 100 mm (approx. 4 in) in size, come from Pribram and Trebsko, Czech Republic. Botryoidal aggregates up to 80 mm (3⅛ in) and imperfect crystals were found in quartz veins in Atlin, British Columbia, Canada. Fine specimens occurred in a Li-bearing pegmatite near Varutrask, Sweden.

Antimony
Sb

TRIGONAL ● ● ●

Properties: C – tin-white; S – gray; L – metallic; D – opaque; DE – 6.7; H – 3-3.5; CL – perfect; F – uneven; M – rhombohedral crystals, botryoidal aggregates, massive.
Origin and occurrence: Hydrothermal in ore veins with silver, stibnite, stibarsen, sphalerite and other minerals; also in pegmatites. As veinlets in a pegmatite in Varutrask, Sweden. Cleavable plates, up to 50 mm (2 in) known from Tornäva, Finland.

Massive aggregates up to 200 mm (7⅞ in) come from Pribram, Czech Republic. Rhombohedral crystals up to 10 mm (⅜ in) across and accumulations up to 300 mm (11 13/16 in) in size described from Lake George, New Brunswick, Canada.

Bismuth

Bi

TRIGONAL ● ● ●

Properties: C – silver-white, tarnishes pink; S – silver-white; L – metallic; D – opaque; DE – 9.8; H – 2-2.5; CL – perfect; M – rhombohedral crystals, granular, massive.
Origin and occurrence: It is found in pegmatites, greisens and hydrothermal in ore veins together with chalkopyrite, arsenopyrite, lollingite, nickeline, breithauptite and many other minerals. Common in pegmatites in Anjanabonoina, Madagascar. Very fine crystals, up to 20 mm (25/32 in) known from Schlema and Hartenstein, Germany. Skeletal aggregates, overgrown with other arsenides, occurred in Jachymov, Czech Republic. Masses weighing several kg found in Bolivia (Tasna, Velaque) and Australia (Kingsgate, New South Wales). Cleavable masses up to 12 cm (4 11/16 in) described from Cobalt and Gowganda, Ontario, Canada.
Application: Bi ore.

Arsenolamprite

As

ORTHORHOMBIC ●

Properties: C – gray-white, it covers with a black coating; S – black, L – metallic to adamantine; D – opaque; DE – 5.6; H – 2; CL – perfect; M – acicular crystals, tabular and fan-shaped aggregates, massive.
Origin and occurrence: Hydrothermal in ore veins associated with arsenic, bismuth, silver and other minerals. Its crystals and veinlets were found with Cu arsenides in Cerny dul, Czech Republic. Occurs also in Jachymoc, Czech Republic and Marienberg Germany. Found recently in Cavnic, Romania.

Bismuth, 20 mm, Cinovec, Czech Republic

Arsenolamprite, 70 mm, Jachymov, Czech Republic

Graphite

C

HEXAGONAL ● ● ● ● ●

Properties: C – black to steel-gray; S – black to steel-gray; L – metallic, dull, earthy; D – opaque; DE – 2.3; H – 1-2; CL – perfect; M – hexagonal tabular crystals, massive.
Origin and occurrence: Metamorphic, from metamorphism of a sedimentarymaterial with C contents: also primary magmatic.Associated with many materials, stabile underconditions of the graphite origin. Crystals several cm in size known from Nordre Stromfjord, Greenland. Crystals were also found in Sterling Hill, New Jersey and Crestmore, California, USA Foliated aggregates are found in Sri Lanka (Radegara, Galle region). Accumulations in Buckingham and Grenville, Quebec, Canada are industrially important. Also common in Shunga deposit in Karelia, Russia; in Cesky Krumlov, Netolice and Blizna, Czech Republic.
Application: metallurgy, nuclear industry, production of lubricants.

Graphite, 130 mm, Krichim, Bulgaria

Diamond

C

CUBIC ● ●

Varieties: bort (opaque technical diamonds), balas (dark colored, spherical radial aggregates), carbonado (brown-black to black massive aggregates, up to egg sized)

Properties: C – colorless, yellow, brown, white, pink, black, red, blue, green; S – white; L – adamantine; D – transparent to translucent; DE – 3.5; H – 10; CL – perfect; F – conchoidal; M – octahedral and cubic crystals; LU – sometimes fluorescent, sometimes phosphorescent.

Origin and occurrence: Primary magmatic occurrences are limited to kimberlite pipes, secondary occurrences to placers. Large primary deposits are known from South Africa (Premier mine, Kimberley) and Yakutia, Russia (pipes in the vicinity of Mirnyi). Primary and secondary occurrences of diamonds are located in lamproites and placers near Argyle, Western Australia. Australia. Most historical diamonds from India (Golconda). Brazil (Diamantina, Minas Gerais), Congo, Angola and Namibia were found in placer deposits. Diamonds of industrial grade always prevail over the gem quality stones. The largest gem-grade diamond ever found, the Cullinan, weighing 3106 carats, comes from the Premier mine in Kimberley, South Africa. It yielded gem rough for 104 faceted stones, the heaviest of which weighs 531 carats. The largest faceted diamond known, called Golden Jubilee, was found in the same place in 1986. It weighed 755 carats before cutting and as a finished stone it weighs 545.65 carats. The dark blue Hope (44 carats) and the green Dresden (76 carats) probably came from India. Absolutely unique red diamond, weighing 5 carats, which is at the Smithsonian Institution, Washington, DC, USA, is of unknown origin.

Application: the most popular gemstone, bort and carbonado varieties are used as abrasives.

Sulfur

S

ORTHORHOMBIC ● ● ● ●

Properties: C – sulfur-yellow, yellow-brown, greenish, reddish to yellowish-gray; S – white; L – resinous to greasy; D – transparent to translucent; DE – 2.1; H – 1.5-2.5; CL – imperfect; F – conchoidal to uneven; M – dipyramidal, disphenoidic and thick tabular crystals, botryoidal and stalactitic aggregates.

Origin and occurrence: Hydrothermal product of fumaroles, product of an activity of microorganisms, disintegration of sulfides and acidic chemical reactions; associated with gypsum, anhydrite, aragonite, calcite,

Diamond, 15 mm x, Mirnyi, Russia

Sulphur, 66 mm, Sicily, Italy

celestite and halite. The world's best crystals come from many localities near Girgenti, Sicily, Italy (Caltanisseta, Cianciana) where they reached up to 12 cm (4 11/16 in) in size. Fine crystals are also known from Tarnobrzeg, Poland and Yavokskoye near Lvov, Ukraine. As a product of solfataras it occurs in many volcanically active places, like Solfatara near Pozzuoli, Italy or in sulfur lava near Shiretoku, Japan. Sulfur layers, up to 30 m (100 ft) thick, associated with salt diapirs, are located near Charles Lake, Louisiana, USA. It originates during intensive oxidation/reduction reactions of pyrite in Rio Tinto, Spain and in Kostajnik, Serbia.

Application: chemical, paper-making, rubber and leather-making industries, agriculture.

Selenium, 110 mm, Kladno, Czech Republic

Selenium

Se

TRIGONAL ● ●

Properties: C – gray to red-gray; S – red; L – metallic, D – opaque to translucent; DE – 4.8; H – 2; CL – good; M – acicular crystals, droplets of vitreous surface, felt-like aggregates.

Origin and occurrence: Secondary, resulting from alteration, fumaroles and from burning coal dumps, with sulfur, sal ammoniac and other sulfates. Also from oxidation of organic compounds in U- and V-bearing deposits of the Colorado Plateau type, associated with pyrite, zippeite and other minerals. Red needles up to 20 mm (25/32 in) long come from the United Verde mine, Jerome, Arizona, USA. Black selenium needles found in burning coal dumps in Kladno and Radvanice, Czech Republic. Occurs with ores of U and V along sandstone fissures in the Peanut Mine, Bull Canyon, Colorado, USA. Occurred through volcanic activity in Vulcano, Lipari Islands, Italy.

Tellurium

Te

TRIGONAL ● ●

Properties: C – tin-white; S – gray; L – metallic; D – opaque; DE – 6.2; H – 2-2.5; CL – perfect; M – prismatic and acicular crystals, granular, massive.

Origin and occurrence: Primary hydrothermal in low-temperature ore deposits; it originates also as secondary through the oxidation/reduction reactions of tellurides. It is associated with gold, sylvanite, altaite, pyrite and other minerals. Crystals up to 30 mm (1 3/16 in) long are known from Balya, Turkey. Crystals up to 20 mm (25/32 in) long occurred in the Au deposits Cripple Creek and Colorado City, USA. Crystals up to 10 mm (3/8 in) across come from Kawazu and Suzuki, Japan. Rich cleavage masses and crystals up to 70 mm (2¾ in) long were found in Uzbekistan. Crystallized tellurium is also known from Fata Baii and Baia de Aries, Romania.

Tellurium, 1 mm xx, Zlatna, Romania

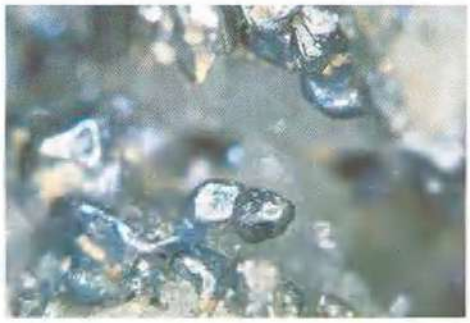

2. Sulfides

Algodonit
Cu_6As

HEXAGONAL ● ●

Properties: C – steel-gray to silver-white, it quickly covers with a brown coating on air; S – gray; L – metallic; D – opaque; DE – 8,7; H – 4; CL – none; F – conchoidal; M – crystals, granular, massive.
Origin and occurrence: Hydrothermal, mainly intimately inter-grown with other Cu arsenides. Its largest accumulations are known from Cu deposits in melaphyres in the Keweenaw Peninsula, Lake Superior, Michigan, USA. It is also known from Chile (Algodones mine near Coquimbo, Atacama). Other localities are Talmessi, Iran and Långban, Sweden.

Domeykite
Cu_3As

CUBIC ● ● ●

Properties: C – tin-white to steel-gray, it tarnishes yellow and covers with a brown coating; S – gray; L – metallic; D – opaque; DE – 7,9; H – 3-3,5; CL – none; F – uneven; M – botryoidal aggregates, massive.
Origin and occurrence: Hydrothermal with copper, cuprite, algodonite and silver. Common in masses, weighing several kg, together with algodonite in the Keweenaw Peninsula, Michigan, USA. The largest accumulations are known from Talmessi and Anarak, Iran. It occurs near Copiap and Chanarcillo, Chile. Massive aggregates in cuprite up to 50 mm (2 in) come from Biloves, Czech Republic.

Sphalerite, 56 mm, Picos de Europa, Spain
Algodonite, 60 mm, Keweenaw Peninsula, U.S.A.

Domeykite, 3 mm x, Rudabanya, Hungary

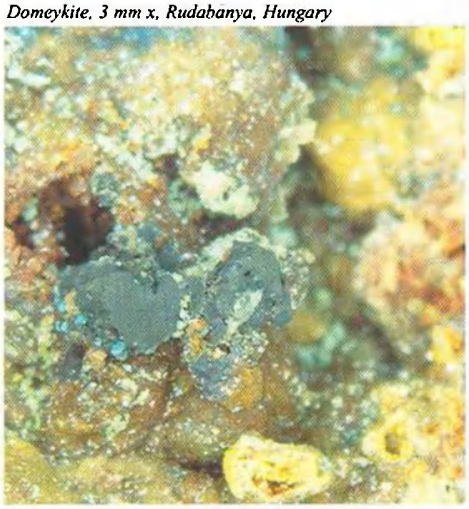

Allargentum
$Ag_{1-x}Sb_x$

HEXAGONAL ●

Properties: C – silver-white; S – gray; L – metallic; D – opaque; DE – 10,1; H – not determined; CL – none; M – small crystals, granular.
Origin and occurrence: Hydrothermal in the silver-bearing ore veins, associated with silver, breithauptite and dyscrasite. Crystals up to 1 mm (1/32 in) known from Hartenstein, Germany. Its intergrowths with silver ores were found in Cobalt, Ontario, Canada. It is also known from Broken Hill, New South Wales, Australia and microscopic in Rejska vein in Kutna Hora, Czech Republic.

Allargentum, 50 mm, Schlema, Germany

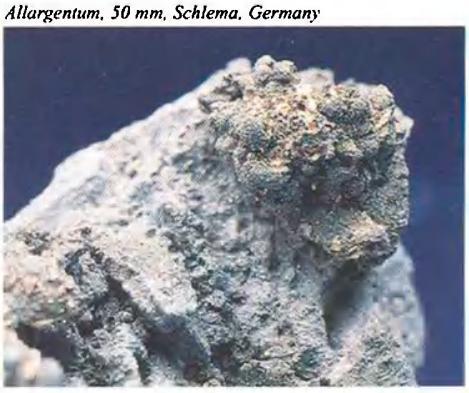

Dyscrasite
Ag_3Sb

ORTHORHOMBIC ● ●

Properties: C – silver-white, it tarnishes yellow to black; S – silver-white; L – metallic; D – opaque; DE – 9,7; H – 3,5-4; CL – good; F – uneven; M – pyramidal and prismatic crystals, granular, massive.
Origin and occurrence: Hydrothermal in the ore veins, associated with silver, stibarsen, pyrargyrite, calcite and other minerals. The best specimens come from the silver-bearing veins, cross-cutting the U deposit Haje near Pøibram, Czech Republic, where prismatic crystals up to 50 mm (approx. 2 in) long and striated tabular twins were found. They are mostly embedded in stibarsen; all the specimens, appearing in the mineral shows, are etched out of matrix. Deformed crystals of completely different habit are known from St. Andreasberg, Germany. Crystals occurred also in the Consols mine in Broken Hill, New South Wales, Australia.

Chalcocite
Cu_2S

MONOCLINIC ● ● ● ●

Properties: C – lead-gray to black; S – lead-gray to black; L – metallic; D – opaque; DE – 5,8; H – 2,5-3; CL – imperfect; F – conchoidal; M – prismatic to tabular crystals, granular, massive.
Origin and occurrence: Hydrothermal, also sedimentary and metamorphic, mostly secondary, in the oxidation and cementation zones of ore deposits. It occurs together with pyrite, chalcopyrite, covellite, bornite and other minerals. Crystals up to 25 cm (9⅞ in) found in the M'Sesa mine, Zaire. Beautiful crystals several cm across come from Redruth and St Just, Cornwall, UK. Crystals over 20 mm (25/32 in) across occurred in Bristol, Connecticut and in Butte, Montana. Crystals up to 50 mm (2 in) in size known from the Flambeau mine near Ladysmith, Wisconsin, USA. Shiny cyclic twins of crystals up to 20 mm (25/32 in) found in Dzhezkazgan, Kazakhstan. Massive aggregates are important Cu ore in Rio Tinto, Spain; Bor, Serbia; Bisbee, Arizona, USA and Tsumeb, Namibia.
Application: important Cu ore.

Djurleite
$Cu_{31}S_{16}$

MONOCLINIC ● ● ●

Properties: C – lead-gray to black; S – lead-gray; L – metallic; D – opaque; DE – 5,8; H – 2,5-3; CL – none; F – conchoidal; M – short prismatic to tabular crystals, granular, massive.
Origin and occurrence: Secondary as a product of the cementation zone in ore deposits. Crystals up to 10 mm (⅜ in) across known from the Botallack mine

Dyscrasite, 75 mm, P?ibram, Czech Republic

Chalcocite, 17 mm x, Cornwall, UK

near St. Just, Cornwall, UK. Aggregates of thick acicular crystals up to 30 mm (1 1/16 in) long found in Dzhezkazgan, Kazakhstan.

It occurs in massive form in many porphyry copper deposits (Butte, Montana; Bisbee, Arizona, USA), also in Tsumeb, Namibia, together with chalcopyrite, pyrite and other minerals.

Djurleite, 3 mm xx, Dzhezkazgan, Kazakhstan

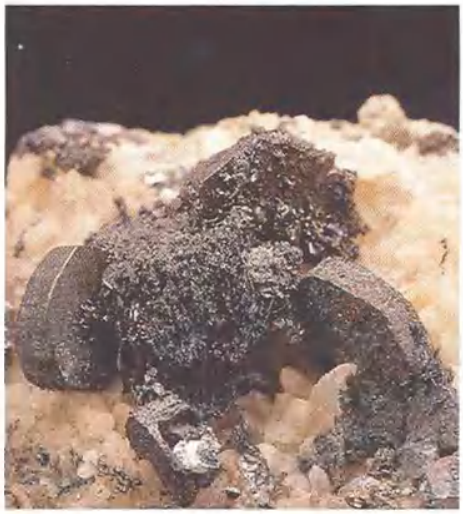

Berzelianite

Cu_2Se

CUBIC ● ● ●

Properties: C – silver-white, tarnishes black; S – silver shiny; L – metallic; D – opaque; DE – 7,3; H – 2,5; CL – none; F – uneven; M – granular, massive.

Origin and occurrence: Hydrothermal, together with other selenides in U, Fe and Au deposits. It is the main mineral in the selenide mineralization in Tilkerode, Germany. Grains, up to several tens of cm across, greenish tarnished, occurred together with other selenides in Bukov, Habøi, Petrovice and Pr edborice , Czech Republic. Similar occurrence is known from near Pinky Fault near Athabasca Lake, Saskatchewan, Canada.

Berzelianite, 60 mm, Bukov, Czech Republic

Bornite, 15 mm xx, Dzhezkazgan, Kazakhstan

Bornite

Cu_5FeS_4

ORTHORHOMBIC ● ● ● ●

Properties: C – copper-red, tarnishes iridescent; S – gray-black; L – metallic; D – opaque; DE – 5,1; H – 3-3,5; CL – imperfect; F – uneven to conchoidal; M – pseudo-cubic octahedral crystals, massive.

Origin and occurrence: Magmatic, hydrothermal, sedimentary, in skarns and pegmatites together with chalcocite, chalcopyrite, pyrite, quartz and other minerals. Fine crystals up to 10 mm (3/8 in) across are known from Carn Brea, Cornwall, England, UK. Crystals up to 30 mm (13/16 in) come from Likasi, Shaba, Zaire. Beautiful crystals up to 40 mm (1 9/16 in) across were found recently together with chalcocite in Dzhezkazgan, Kazakhstan. Massive aggregates are common and used as Cu ore in Kipushi, Shaba, Zaire. Fine-grained, sedimentary bornite occurs in Cu-bearing shales in Mansfeld, Germany, where it forms the main ore layer. Crystals, up to 20 mm (25/32 in) in size, occurred in the Cole shaft and masses, weighing several thousands of tons, were mined in the Campbell shaft, Bisbee, Arizona, USA. *Application:* Cu ore.

Bornite, 3 mm xx, Dzhezkazgan, Kazakhstan

Umangite

Cu_3Se_2

TETRAGONAL ● ● ●

Properties: C – blue-black with reddish tint, tarnishes purple; S – black; L – metallic; D – opaque; DE – 6,6; H – 3; CL – imperfect; F – uneven to conchoidal; M – granular, massive.

Origin and occurrence: Hydrothermal in ore veins together with other selenides (clausthalite, berzelianite). It is common, associated with berzelianite in Tilkerode, Germany; Sierra de Umango, Argentina and Slavkovice, Czech Republic. Larger accumulations occur in the Martin Lake mine near Athabasca Lake, Canada.

Acanthite
Ag_2S

MONOCLINIC ● ● ●

Properties: C – black; S – black; L – metallic; D – opaque; DE – 7,2; H – 2-2,5; CL – none; F – uneven; M – pseudo-cubic crystals, massive. It mainly occurs as paramorphs after argentite (high-temperature phase of the same composition).
Origin and occurrence: Hydrothermal in ore veins. Beautiful crystals over 50 mm (2 in) long occurred in the Himmelsfürst mine in Freiberg, in Annaberg and Schneeberg, Germany. Acicular crystals are known from Jachymov, Czech Republic. It is common in association with silver, proustite, pyrargyrite, polybasite, stephanite, galena and other minerals in Mexico. Probably the best paramorphs after argentite up to 70 mm (2¾ in) across come from the Rayas mine, Guanajuato. Fine crystals occur in the Las Chispas mine, Arizpe, Sonora and many localities in Zacatecas, Chihuahua.
Application: important Ag ore.

Argentite
Ag_2S

CUBIC ● ● ●

Properties: C – black-gray, tarnishes black; S – black; L – metallic; D – opaque; DE – 7,1; H – 2-2,5; CL – imperfect; F – uneven to conchoidal; M – octahedral and cubic crystals, dendritic aggregates, massive. Stabile at temperatures over 179EC (354.2°F), below this temperature there are paramorphs of acanthite after argentite.

Acanthite, 23 mm, Arizpe, Mexico

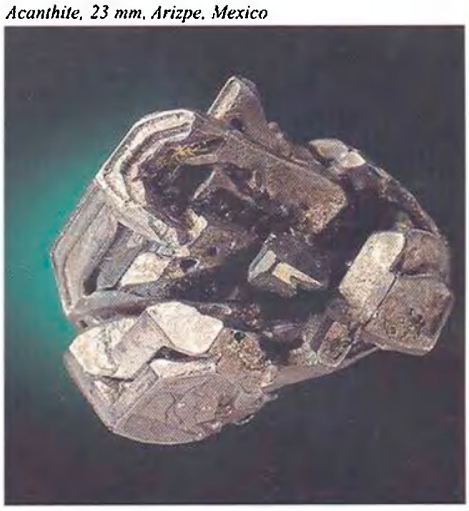

Origin and occurrence: Hydrothermal in low-temperature ore deposits, associated with silver, galena and Ag sufosalts. Occurs between the oxidation and cementation zone with stromeyerite, silver, jalpaite, iodargyrite and other minerals. Fine crystals up to 40 mm (1 9/16 in) across, are known from Freiberg and Schneeberg, Germany. Similar crystals found in Jachymov and Midinec, Czech Republic. Crystals up to 30 mm (1 3/16 in) occur in Sarrabus, Sardinia, Italy. Maybe the best argentite crystals occurred in Mexico (Arizpe, Sonora; Zacatecas; Guanajuato), where crystals reached up to 40 mm (1 9/16 in). Fine crystals up to 20 mm (25/32 in) across reported from Chañarcillo in Chile.
Application: important Ag ore.

Umangite, 40 mm, Beaverlodge Lake, Canada

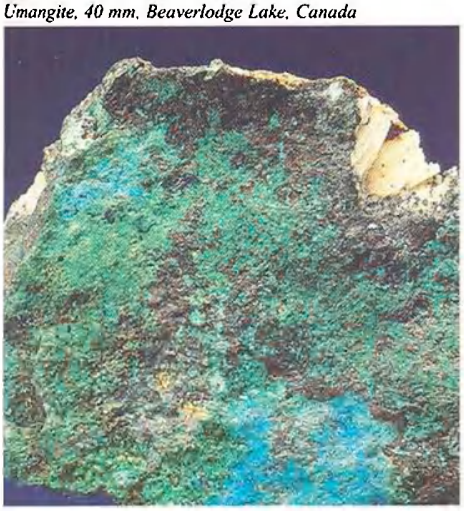

Argentite, 29 mm, Zacatecas, Mexico

Aguilarite, 4 mm xx, Guanajuato, Mexico

Hessite

Ag_2Te

MONOCLINIC ● ● ●

Properties: C – lead to steel-gray, tarnishes black; S – light gray; L – metallic; D – opaque; DE – 8,4; H – 2-3; CL – imperfect; F – even; M – pseudo-cubic crystal combinations, granular, massive.

Origin and occurrence: Hydrothermal in medium- and low-temperature ore veins together with calaverite, sylvanite, altaite, gold, tellurium and other sulfides. The best specimens with crystals up to 20 mm ($^{25}/_{32}$ in) across ($^{25}/_{32}$ in) and aggregates up to 100 mm (4 in) come from the Bote mine, Romania. Small crystals occur in the Jamestown mine, California, USA. Massive aggregates were found in Gold Hill, Colorado, USA and Moctezuma, Mexico. Aggregates, up to tens of cm² (several sq in) in size, were known in the Zavodinskii mine, Altai, Kazakhstan.

Aguilarite

Ag_4SeS

ORTHORHOMBIC ●

Properties: C – lead-gray, tarnishes black; S – gray-black; L – metallic; D – opaque; DE – 7,7; H – 2,5; CL – none; F – hackly; M – skeletal crystals, massive.

Origin and occurrence: Hydrothermal in ore veins, together with silver, stephanite, proustite, pearceite, calcite and quartz. The best crystallized specimens with crystals up to 30 mm ($1^{3}/_{16}$ in) across come from the San Carlos mine, Guanajuato and Chontalpan, Taxco, Guerrero, Mexico. It is also known in inter-

Hessite, 31 mm, Botes, Romania

growths with acanthite and naumannite from the Comstock Lode, Virginia City, Nevada, USA.

Argyrodite
Ag_8GeS_6

ORTHORHOMBIC ●

Properties: C – steel-gray, tarnishes black; S – gray-black; L – metallic; D – opaque; DE – 6,3; H – 2,5-3; CL – none; F – uneven to conchoidal; M – combinations of cubic crystals, botryoidal aggregates, massive.
Origin and occurrence: Hydrothermal in low-temperature base metal deposits, associated with Ag sulfosalts. It occurred in crystals in the Himmelsfhrst mine, Freiberg, Germany. Crystals were also found in several localities in Bolivia (Atoche, Colquechaca, Potosí). Crystal measuring 60 mm (24 in) across is reported from Porco, Bolivia.

Stromeyerite
AgCuS

ORTHORHOMBIC ● ● ●

Properties: C – dark steel gray, tarnishes blue; S – steel-gray; L – metallic; D – opaque; DE – 6,3; H – 2.5-3; CL – none; F – conchoidal; M – pseudo-hexagonal tabular crystals, massive.
Origin and occurrence: Mostly secondary in the cementation zone of ore veins, associated with freibergite, bornite, chalcocite, galena and other minerals. Fine tabular crystals up to 10 mm (7/8 in) across found in Dzhezkazgan, Kazakhstan, where pseudo-morphs of stromeyerite after silver wires also occur. Skeletal prismatic pseudo-morphs after chalcocite crystals come from Vraneice, Czech Republic. Massive aggregates are common in many deposits in Colorado, USA (Aspen; Red Mountain), Chile (Copiapo), Bolivia (Potosi) and Canada (Cobalt, Ontario).

Stromeyerite, 6 mm xx, Dzhezkazgan, Kazakhstan

Argyrodite, 34 mm, Mexico

Jalpaite
Ag_3CuS_2

TETRAGONAL ● ●

Properties: C – light gray, tarnishes dark gray to iridescent; S – black; L – metallic; D – opaque; DE – 6,8; H – 2-2,5; CL – good; F – conchoidal; M – crystals, granular.
Origin and occurrence: Hydrothermal in low-temperature ore veins. Crystals up to 25 mm (1 in) across known from Jalpa, Queretaro, Mexico. Crystals up to 30 mm (13/16 in) across come from the Caribou mine, Colorado, USA. Massive aggregates, associated with galena, sphalerite, pyrite, stromeyerite, polybasite and other minerals, occurred in Pøibram, Czech Republic.

Jalpaite, 30 mm veinlet, P?ibram, Czech Republic

Pentlandite, 90 mm, Sudbury, Canada

Pentlandite

$(Fe,Ni)_9S_8$

CUBIC ● ● ● ●

Properties: C – light bronze to red-brown; S – light bronze; L – metallic; D – opaque; DE – 5,0; H – 3,5-4; CL – none; F – conchoidal; M – crystals, massive.
Origin and occurrence: Typical magmatic liquid mineral, associated with pyrrhotite and chalcopyrite. It is common in Sudbury, Ontario, Canada, as mostly microscopic inclusions in chalcopyrite, but also as imperfect crystals. Large accumulations are known from Talnakh near Norilsk, Siberia, Russia, where it occurs together with Cu, Pt and Pd sulfides. It is also important ore in the deposit near Rustenburg, South Africa.
Application: the most important Ni ore.

Sphalerite, 70 mm, Olkusz, Poland

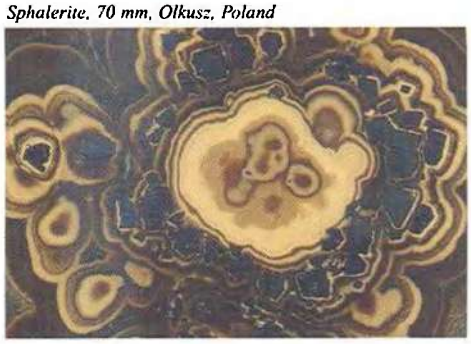

Sphalerite

ZnS

CUBIC ● ● ● ●

Varieties: cleiophane (green, yellow, orange), marmatite (black)

Properties: C – colorless, yellow, orange, green, brown, black; S – brownish, light yellow, white; L – resinous to adamantine; D – transparent, translucent, opaque; DE – 4,1; H – 3,5-4; CL – perfect; F – conchoidal; M – tetrahedral and dodecahedral crystals, botryoidal, fibrous and stalactitic aggregates, massive; LU – sometimes orange.
Origin and occurrence: Magmatic (liquid, in pegmatites); hydrothermal in low- to high-temperature deposits, skarns, hydrothermal sedimentary deposits;

Sphalerite, 31 mm, Elmwood, U.S.A.

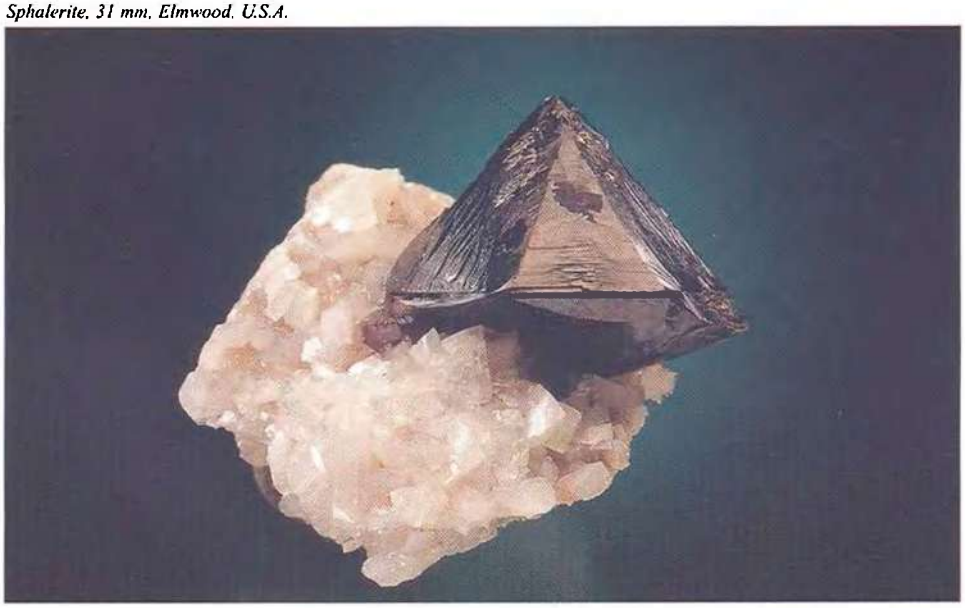

rare sedimentary and metamorphic. It occurs together with galena, pyrite, chalcopyrite, marcasite, fluorite, barite, quartz and other minerals. Beautiful crystals up to 100 mm (4 in) across come from Trepea, Serbia. Green and red crystals up to 100 mm (4 in) known from Cananea, Sonora, Mexico. Fine yellow crystals up to 30 mm (1 3/16 in) were common in Banska "tiavnica, Slovakia; similar crystals occur in Madan, Bulgaria. The most beautiful yellow, orange and red crystals up to 150 mm (6 in) across found in Picos de Europa, Santander, Spain. They are sometimes faceted. Brown crystals up to 50 mm (2 in) come from Joplin, Missouri; stalactitic aggregates up to 150 mm (6 in) long come from Galena, Illinois, USA. Perfect black, shiny crystals and twins up to 50 mm (2 in) are famous from Dalnegorsk, Russia; yellow crystals, up to 30 mm (1 3/16 in), occurred in Dzhezkazgan, Kazakhstan. Transparent crystals up to 30 mm (1 3/16 in) also found in Franklin, New Jersey, USA. Green crystals up to 100 mm (approx. 4 in) across occurred in the Big Four mine, Colorado, USA. Crystals up to 50 mm (2 in) known from the Oppu mine, Aomori, Japan.
Application: principal Zn ore.

Coloradoite
HgTe

CUBIC ● ●

Properties: C – black-gray; S – black-gray; L – metallic; D – opaque; DE – 8,1; H – 2,5; CL – none; F – uneven to conchoidal; M – granular, massive.
Origin and occurrence: Hydrothermal, associated with altaite, calaverite, krennerite, gold, pyrite and other minerals in Au-bearing veins. It was common in Cripple Creek and in the Smuggler mine, Colorado; in the Norwegian mine, California, USA. Grains, reaching up to several mm, come from Jilove, Czech Republic.

Coloradoite, 70 mm, Kalgoorlie, Australia

Chalcopyrite, 10 mm xx, Cavnic, Romania

Chalcopyrite
$CuFeS_2$

TETRAGONAL ● ● ● ● ●

Properties: C – brass-yellow, tarnishes iridescent; S – green-black; L – metallic; D – opaque; DE – 4,3; H – 3,5-4; CL – imperfect; F – uneven; M – tetrahedral crystals, botryoidal aggregates, massive.
Origin and occurrence: Magmatic, hydrothermal and sedimentary, in association with sphalerite, galena, tetrahedrite, pyrite and many other sulfides. Fine crystals up to 30 mm (1 3/16 in) across are known from Banska "tiavnica, Slovakia and from Cavnic, Romania.
Crystals up to 120 mm (4 11/16 in) across, associated with other sulfides, come from the Nikolai mine in Dalnegorsk, Russia. Beautiful crystals up to 120 mm (4 11/16 in) found in Japan (Arawaka, Osarizawa). Fine crystals reaching up to several cm occur in Peru (Huanzala, Huaron). Massive aggregates are important Cu ore in Sudbury, Ontario, Canada; Bingham, Utah; Bisbee, Arizona, USA and Rio Tinto, Spain.
Application: important Cu ore.

Chalcopyrite, 15 mm x, Ground Hog Mine, U.S.A.

Luzonite, 60 mm, Recsk, Hungary

Luzonite
Cu_3AsS_4

TETRAGONAL ● ●

Properties: C – dark pink-brown; S – black; L – metallic; D – opaque; DE – 4,5; H – 3,5; CL – good; F – uneven to conchoidal; M – crystals, granular, massive.
Origin and occurrence: Hydrothermal in low- to medium-temperature veins, associated with enargite, tetrahedrite, sphalerite, bismuthinite, Ag sulfosalts and other minerals. Crystals up to 40 mm (1 9/16 in) across come from Quiruvilca; it is common in Cerro de Pasco, Peru. It is also common in the Teine mine, Hokkaido, Japan; crystals also occur in Kinkwaseli, Taiwan. It is known from Bor, Serbia and Recsk, Hungary.

Stannite
Cu_2FeSnS_4

TETRAGONAL ● ● ●

Properties: : C – steel-gray to black, tarnishes blue;

Stannite, 10 mm xx, Potosí, Bolivia

Germanite, 40 mm, Tsumeb, Namibia

S – black; L – metallic; D – opaque; DE – 4,5; H – 4; CL – imperfect; F – uneven; M – pseudo-octahedral crystals, granular, massive.
Origin and occurrence: Hydrothermal in high-temperature Sn deposits. The best specimens come from Bolivia; crystals up to 50 mm (2 in) across known from Llallagua; crystals up to 30 mm (1 3/16 in) from Chocaya and cross-like inter-growths from the San Jose and Itos mines near Oruro. It occurs as massive vein fillings in Carn Brea, Cornwall, UK and Cínovec, Czech Republic. It was also found in amblygonite pegmatites in Caceres, Spain and in quartz-amblygonite veins near Verneøov, Czech Republic.
Application: Sn ore.

Germanite
$Cu_{26}Fe_4Ge_4S_{32}$

CUBIC ● ●

Properties: C– pink to purple-grey, S – black, L - metallic, D - opaque, DE – 4,5, H - 4, CL – none, F – uneven, M – tetrahedral crystals, granular, massive.
Origin and occurrence: Hydrothermal, often intergrown with tennantite, bornite and other minerals. It was important only in one particular place in Tsumeb, Namibia, where larger accumulations were found. Smaller occurrences are known from the Shikanai mine, Japan; small cubic crystals come from the Humboldt mine, Colorado, USA.

Tennantite
$Cu_{12}As_4S_{13}$

CUBIC ● ● ● ●

Properties: C – steel-gray; S – black, brown to dark

red; L – metallic; D – opaque; DE – 4.6; H – 3-4.5; CL – none; F – conchoidal to uneven; M – tetrahedral crystals, granular, massive.
Origin and occurrence: Hydrothermal in ore veins and greisens with pyrite, calcite, dolomite, quartz and other sulfides and Cu-Pb-Zn-Ag sulfosalts. Fine crystals are known from Cornwall, UK (Wheal Jewel, Gwennap; Carn Brea). Crystals up to 150 mm (6 in) across, come from Tsumeb, Namibia. Crystals of binnite in Lengenbach, Binntal, Switzerland, up to 30 mm (1 3/16 in). Crystals up to 20 mm (25/32 in) across were found recently in Dzhezkazgan, Kazakhstan. Large masses occurred in Kipushi, Zaire. Crystals up to 25 mm (1 in) known from El Cobre, Zacatecas, Mexico. *Application:* Cu ore.

Tennantite, 26 mm, El Cobre, Mexico

Tetrahedrite

$Cu_{12}Sb_4S_{13}$

CUBIC ● ● ● ●

Properties: C – steel-gray to black; S – black, brown; L – metallic; D – opaque; DE – 5.0; H – 3-4.5; CL – none; F – conchoidal; M – tetrahedral crystals, granular, massive.
Origin and occurrence: Hydrothermal in low- to medium temperature veins; in contact metamorphic deposits together with chalcopyrite, galena, sphalerite, pyrite, bornite, calcite, quartz and other minerals. The largest known crystals up to 25 cm (9 7/8 in) across come from Anzen and Irazein in Pyrennees, France. Common crystals several cm in size occur in Cavnic, Romania. Fine specimens with crystals up to 70 mm (2¾ in) across found in the Mercedes mine in Huallanca, Peru. Other Peruvian localities like Casapalca and Morococha yielded fine crystallized specimens. Fine crystals up to 20 mm (25/32 in) in size known from Poibram, Czech Republic.
Application: ruda Cu.

Tetrahedrite, 10 mm xx, Peru

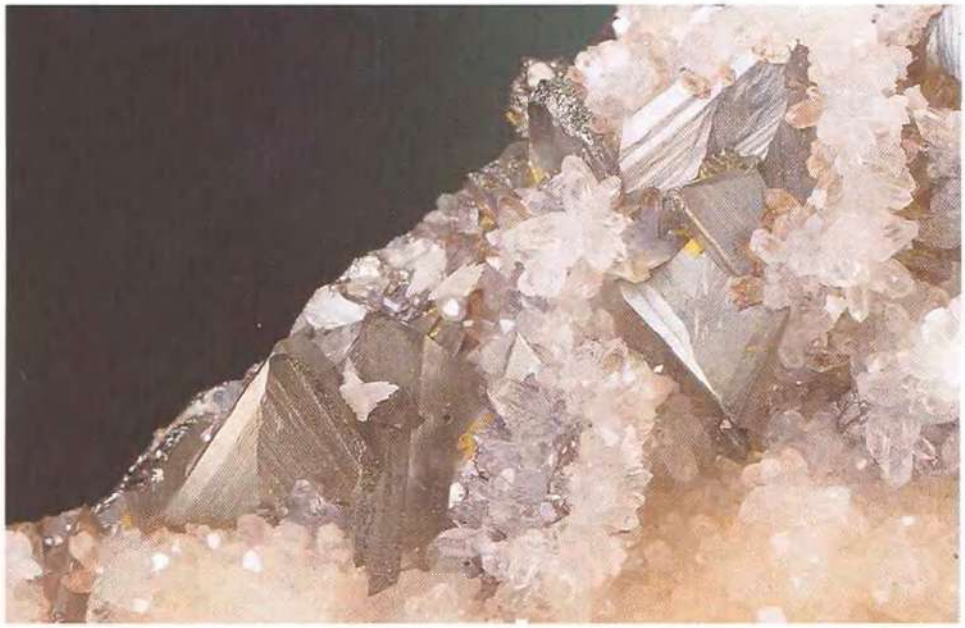

Freibergite, 15 mm xx, Obecnice, Czech Republic

Freibergite

$(Ag,Cu,Fe)_{12}(Sb,As)_4S_{13}$

CUBIC ● ● ●

Properties: C – gray to black; S – black, brown to dark red; L – metallic; D – opaque; DE – 5,4; H – 3-4,5; CL – none; F – uneven to conchoidal; M – tetrahedral crystals, massive.
Origin and occurrence: Hydrothermal in ore deposits, associated with many sulfides and sulfosalts. It is mainly massive in the Himmelsfurst mine in Freiberg, Germany; Cobalt, Ontario, Canada and in Mount Isa, Queensland, Australia. It was a principal Ag ore in Kutná Hora, Czech Republic.
Application: Ag ore.

Wurtzite

ZnS

HEXAGONAL ● ● ●

Properties: C – dark red-brown, dark brown to brown-black; S – brown; L – resinous to submetallic; D – translucent to opaque; DE – 4,1; H – 3,5-4; CL – good; M – pyramidal, prismatic to thick tabular, striated crystals, concentric banded and radial aggregates.
Origin and occurrence: Hydrothermal in ore veins with sphalerite, marcasite, pyrite and other minerals. Also of low-temperature origin along the cracks of clay concretions. The best wurtzite crystals come from Bolivia; crystals up to 40 mm ($1^{9}/_{16}$ in) across from Animas and crystals up to 20 mm ($^{25}/_{32}$ in) across from Llallagua and Potosi. Fine crystals up to 30 mm ($1^{3}/_{16}$ in) across found in Talnakh near Norilsk, Siberia, Russia. Interesting radial aggregates, up to several cm in diameter, occurred in Pøibram, Czech Republic.

Wurtzite, 30 mm xx, Animas, Bolivia

Greenockite

CdS

HEXAGONAL ● ● ●

Properties: C – yellow to orange, dark red; S – orange-yellow to brick-red; L – adamantine to resinous; D – opaque to translucent, DE – 4,8; H – 3-3,5; CL – good; F – conchoidal; M – trillings, pulverulent coatings.
Origin and occurrence: Mainly secondary as pulverulent coatings on sphalerites. It also occurs in cavities of volcanic rocks together with prehnite, zeolites and calcite. Crystals are known from ore veins in Llallagua, Bolivia. Pulverulent coatings are described from Pøibram, Czech Republic; Bleiberg, Austria and from the deposits in the Tri State region, Missouri, USA. Crystals up to 10 mm ($^{3}/_{8}$ in) occurred in the cavities of volcanic rocks near Renfrew, Scotland, UK.

Enargite

Cu_3AsS_4

ORTHORHOMBIC ● ● ● ●

Properties: C – gray-black to black; S – gray-black;

Grenockite, 20 mm, Ocna de Fier, Romania

L – metallic; D – opaque; DE – 4,4; H – 3; CL – perfect; F – uneven; M – tabular and prismatic, striated crystals, massive.
Origin and occurrence: Hydrothermal in medium-temperature, sometimes in low-temperature deposits, associated with quartz, pyrite, sphalerite, galena, bornite and other minerals. Beautiful crystals, up to 100 mm (4 in) across, come from the Luz Angelica mine in Quiruvilca; crystals up to 150 mm (6 in) across from Morococha and Cerro de Pasco, Peru. Fine crystals were found in Butte, Montana, USA and in Mancayano, Luzon, Philippines. It occurs as a principal Cu ore in several deposits (Bor, Serbia; Huaron, Peru).
Application: Cu ore.

Enargite, 41 mm, Butte, U.S.A.

Cubanite
Cu_2FeS_3

ORTHORHOMBIC ● ● ●

Properties: C – brass-yellow to bronze; S – black; L – metallic; D – opaque; DE – 4,1; H – 3,5; CL – none; F – conchoidal; M – tabular crystals, massive.
Origin and occurrence: Magmatic in liquid deposits as inclusions in chalcopyrite, as hydrothermal in high-temperature deposits, associated with chalcopyrite, pyrite, pyrrhotite and sphalerite. The best crystals, twins, up to 40 mm (1 9/16 in) across, come from the Henderson No.2 mine, Chibougamau, Quebec, Canada. It was also described in crystals from Sudbury, Ontario, Canada and Morro Velho, Brazil. Its large accumulations are important Cu ore in Sudbury, Ontario, Canada; in Mooihoek, South Africa and in Prince William Sound, Alaska, USA.
Application: important Cu ore.

Cubanite, 33 mm x, Chibougamau, Canada

Sternbergite, 20 mm, St.Andreasberg, Germany

Sternbergite
Ag_2FeS_3

ORTHORHOMBIC ● ●

Properties: C – golden-brown; S – black; L – metallic to adamantine; D – opaque; DE – 4,3; H – 1-1,5; CL – perfect; M – thin tabular pseudo-hexagonal crystals, often in rosettes and fan-shaped aggregates.
Origin and occurrence: Hydrothermal in Ag-bearing veins, associated with stephanite, acanthite, proustite, argentopyrite and other minerals. Tabular crystals up to several mm across known from Jachymov and Midinec, Czech Republic; from St. Andreasberg, Johanngeorgenstadt, Schneeberg and Freiberg, Germany.

Argentopyrite, 60 mm, M?d?nec, Czech Republic

Argentopyrite
Ag_2FeS_3

ORTHORHOMBIC ● ●

Properties: C – gray-white, tarnishes iridescent; S – gray; L – metallic; D – opaque; DE – 4,3; H – 3,5 – 4; CL – none; F – uneven; M – thick tabular pseudo-hexagonal crystals, granular.
Origin and occurrence: Hydrothermal in ore veins, associated with proustite, pyrargyrite, stephanite, sternbergite, dolomite, quartz and other minerals. Crystals up to 5 mm (3/16 in) across come from Jachymov and Midinec, Czech Republic; it is also known from Schlema, Freiberg and Schneeberg, Germany and from Colquechaca, Bolivia.

Pyrrhotite, 68 mm, Dalnegorsk, Russia

Pyrrhotite
$Fe_{1-x}S$ (x = 0-0,17)

MONOCLINIC ● ● ● ●

Properties: C – bronze-yellow to brown, tarnishes quickly; S – dark gray-black; L – metallic; D – opaque; DE – 4,7; H – 3,5-4,5; CL – none; F – uneven to conchoidal; M – tabular, pyramidal and prismatic crystals, massive.
Origin and occurrence: Magmatic liquid in basic rocks, together with pyrite and pentlandite; in pegmatites; hydrothermal in high-temperature and metasomatic deposits; sedimentary and metamorphic. Tabular crystals up to 300 mm (11 13/16 in) across come from Trepea, Serbia and Dalnegorsk, Russia; prismatic crystals up to 150 mm (6 in) long found in Santa Eulalia, Chihuahua, Mexico and Chiuzbaia, Romania. Tabular crystals up to 110 mm (4 5/16 in) occurred in Cavnic, Romania. Large imperfect crystals, coated with wavellite, are known from Llallagua, Bolivia. Huge masses of industrial importance occur in Sudbury, Ontario, Canada; Talnakh near Norilsk, Siberia, Russia and elsewhere. *Application:* sometimes as Fe ore.

Nickeline, 138 mm, Pohla, Germany

Nickeline
NiAs

HEXAGONAL ● ● ● ●

Properties: C – light copper-red, tarnishes gray to black; S – light brown-black; L – metallic; D – opaque; DE – 7,8; H – 5-5,5; CL – none; F – conchoidal; M – striated crystals, botryoidal and dendritic aggregates, granular, massive.
Origin and occurrence: Hydrothermal in high-temperature ore veins with skutterudite, nickel-skutterudite, safflorite, rammelsbergite and other minerals. Also magmatic in norites and peridotites. Crystals up to 15 mm (19/32 in) across come from Pohla; small crystals were found in Richelsdorf, Germany. Fine massive aggregates and small crystals occurred in Bou Azzer, Morocco. Huge accumulations are known in Cobalt and Gowganda, Ontario, Canada.
Application: Ni ore.

Breithauptite
NiSb

HEXAGONAL ● ●

Properties: C – light copper-red, purplish; S – red-brown; L – metallic; D – opaque; DE – 8,6; H – 5,5; CL – none; F – conchoidal to uneven; M – thin tabular crystals, dendritic aggregates, massive.
Origin and occurrence: Magmatic, hydrothermal and metamorphic, associated with silver, nickeline, cobaltite and other sulfides. It occurs inter-grown with pyrrhotite and pentlandite in magmatic liquid deposit Vlakfontein, South Africa. It is common in hydrothermal veins in Cobalt, Ontario, Canada; also known from Sarrabus, Sardinia, Italy and St. Andreasberg, Germany.

Breithauptite, 80 mm, Cobalt, Canada

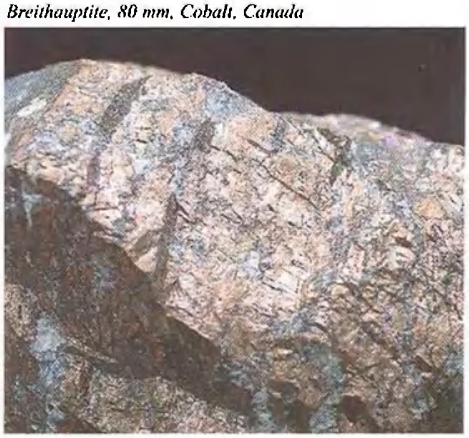

Millerite, 26 mm, Antwerp, U.S.A.

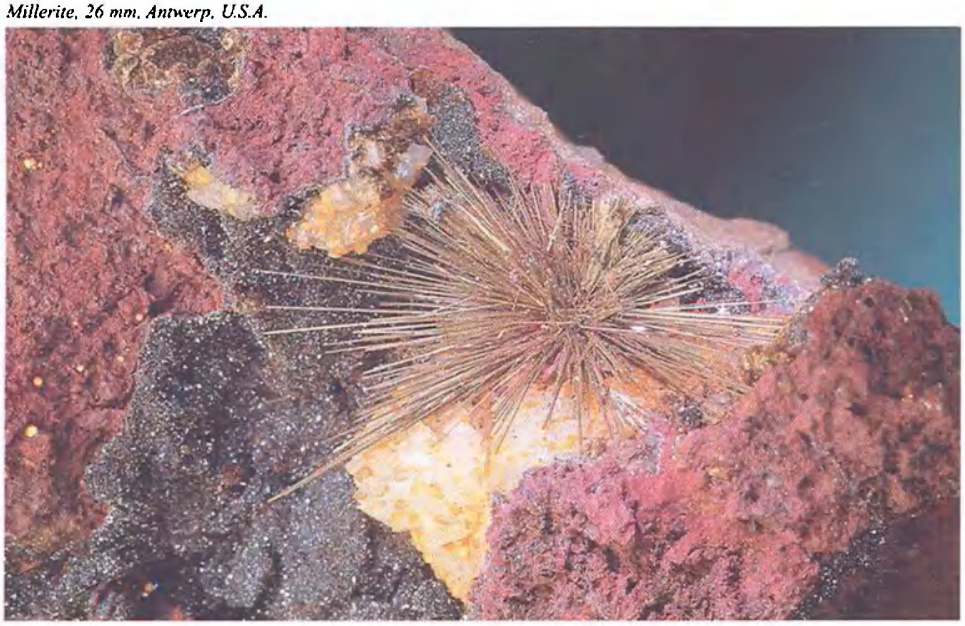

Millerite
NiS

TRIGONAL ● ● ●

Properties: C – light brass-yellow to bronze, tarnishes iridescent; S – greenish-black; L – metallic; D – opaque; DE – 5,4; H – 3-3,5; CL – perfect; F – uneven; M – acicular crystals, cleavable masses, aggregates of parallel inter-grown crystals with velvety surface.
Origin and occurrence: Hydrothermal in low-temperature ore deposits, also as a product of decomposition of Ni sulfides. It is associated with pyrrhotite, ankerite, whewellite, barite and other minerals. Cleavable masses occurred near Temagami near Sudbury, Ontario, Canada and in Kambalda, Western Australia, Australia. Acicular crystals come from ore veins in Jachymov and Příbram, Czech Republic; needles in cavities of siderite concretions in the vicinity of Kladno, Czech Republic, reached up to 70 mm (2¾ in). Acicular crystals up to 80 mm (3⅛ in) long known from the limestone cavities near Dortmund, Germany. Beautiful specimens with crystals up to 50 mm (2 in) long in the cavities in hematite occurred in the Sterling mine, Antwerp, New York, USA. Velvety fibrous aggregates were found in the Thompson mine, Manitoba, Canada. *Application:* Ni ore.

Alabandite, 40 mm, Sacarimb, Romania

Alabandine
MnS

CUBIC ● ●

Properties: C – black; S – green; L – submetallic; D – opaque; DE – 4,1; H – 3,5-4; CL – perfect; F – uneven; M – cubic and octahedral crystals, granular, massive.
Origin and occurrence: Hydrothermal in ore veins, associated with rhodochrosite, calcite, galena, sphalerite, pyrite and other minerals. Crystals up to 20 mm (25/32 in) across together with granular aggregates are relatively common in Romania (Sacarîmb, Baia de Arieş , Roşia Montana). Crystals are also known from the Queen of the West mine, Colorado and from the Lucky Cuss mine, Tombstone, Arizona, USA.

Galena, 56 mm, Neudorf, Germany

Galena, 66 mm, Picher, U.S.A.

Galena

PbS

CUBIC ● ● ● ● ●

Varieties: steinmannite

Properties: C – lead-gray; S – lead-gray; L – metallic; D – opaque; DE – 7,6; H – 2,5; CL – perfect; F – conchoidal; M – cubic crystals and their complex combinations, tabular crystals, skeletal aggregates, massive.
Origin and occurrence: Magmatic, hydrothermal, metamorphic, very rare sedimentary, associated with sphalerite, chalcopyrite, pyrite, quartz and other minerals. Large crystals up to several tens of cm in size come from many localities in the USA. (Joplin, Missouri; Galena, Kansas; Picher, Oklahoma; Sweetwater mine, Missouri). Beautiful, often skeletal crystals or spinel-law twins up to 200 mm (7⅞ in) across known from the Nikolai mine in Dalnegorsk, Russia. Fine crystals occur also in Naica, Chihuahua, Mexico. Famous complicated combinations of crystals were found in Neudorf, Germany. Octahedral crystals up to 10 mm (⅜ in) (steinmannite variety) were common in Pribram, Czech Republic. Beautiful specimens with cubes up to several cm across come from Madan, Bulgaria; spinel-law twins occurred in Herja, Romania.
Application: the most important Pb ore.

Clausthalite, 20 mm veinlet, Tilkerode, Germany

Clausthalite

PbSe

CUBIC ● ● ●

Properties: C – lead-gray, bluish; S – gray-black; L – metallic; D – opaque; DE – 8,3; H – 2,5 – 3; CL – good; F – granular; M – granular, massive..
Origin and occurrence: Hydrothermal in ore veins with a low S content, together with berzelianite, umangite, uraninite and other minerals. It occurs as massive aggregates in calcite veins with other selenides in Clausthal and Tilkerode, Germany. It is similar in Skrikerum, Sweden; common in the U deposits in Predborice, Bukov and Zlatkov, Czech Republic.

Altaite. 65 mm, Stania, Romania

Altaite
PbTe

CUBIC ● ● ●

Properties: C – tin-white to yellowish, tarnishing to bronze; S – black; L – metallic; D – opaque; DE – 8,3; H – 3; CL – perfect; F – conchoidal; M – cubic and octahedral crystals, massive.

Miargyrite. 20 mm. Kutna Hora, Czech Republic

Origin and occurrence: Hydrothermal in vein Au deposits, associated with other tellurides, galena and other minerals. Crystals up to 20 mm (25/32 in) across are known from the Revenge mine, Colorado, USA. It is massive with aguilarite in Kalgoorlie, Western Australia, Australia. It is relatively common with other tellurides in Sacarîmb , Romania; in the Zavodinskii mine, Altai, Russia and in Zod near Sevan Lake, Armenia.

Miargyrite
$AgSbS_2$

MONOCLINIC ● ● ●

Properties: : C – black to steel-gray; S – cherry-red; L – metallic, adamantine; D – opaque; DE – 5,3; H – 2,5; CL – imperfect; F – conchoidal; M – thick tabular, striated crystals, massive.
Origin and occurrence: Hydrothermal in low-temperature ore veins, together with proustite, pyrargyrite, polybasite, silver, quartz and other minerals. Fine crystals, up to 10 mm (3/8 in) across, occurred in Príbram and Kutna Hora, Czech Republic; in Braunsdorf and Freiberg, Germany. It is also known from many localities in Bolivia (Tatasi, Oruro – 10 mm (3/8 in) crystals, Potosí) and Mexico (Sombrerete, Catorce). It was also found in Hiendelaencina, Guadalajara, Spain.
Application: Ag ore.

Franckeite
$(Pb,Sn)_6FeSn_2Sb_2S_{14}$

TRICLINIC ● ●

Properties: C – gray-black; S – gray-black; L – metallic; D – opaque; DE – 5,9; H – 2,5-3; CL –

Franckeite. 30 mm, Oruro, Bolivia

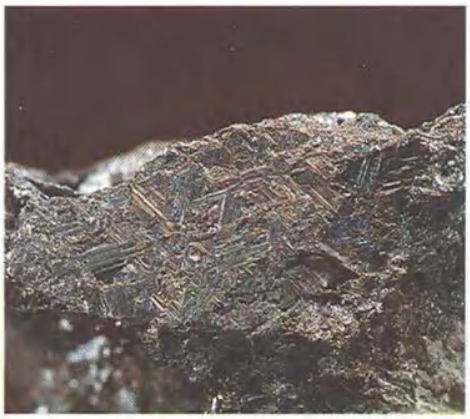

Cylindrite, 50 mm, Poopo, Bolivia

perfect; M – thin tabular, curved crystals, often rosette-like aggregates of crystals.
Origin and occurrence: Hydrothermal in Ag-Sn deposits, associated with cylindrite, zinkenite, cassiterite, wurtzite and other minerals; in contact metamorphic limestones. The best crystals up to 60 mm (24 in) across come from various localities in Bolivia (Mollecagua near Poopo; Huanuni; Chocoya near Potosi; Colquechaca and Oruro). It is known in contact metamorphic limestone from the Kalkar quarry, California, USA.
Application: Sn ore.

Cylindrite
$Pb_3Sn_4FeSb_2S_{14}$

TRICLINIC ● ●

Properties: C – lead-gray to black; S – black; L – metallic; D – opaque; DE – 5,4; H – 2,5; CL – perfect; M – cylindrical, conical and spherical crystal aggregates, massive.
Origin and occurrence: Hydrothermal in Sn-bearing veins with franckeite, stannite, cassiterite, galena and other minerals. World famous specimens with crystals up to 50 mm (2 in) long come from the Trinacria and Santa Cruz mines near Poopo, Bolivia. It was also reported from the Smirnovsk deposit, Transbaikalia, Russia.

Cinnabar
HgS

TRIGONAL ● ● ● ●

Properties: C – red to brownish-red; S – crimson; L – adamantine to metallic, also dull; D – opaque; DE – 8,2; H – 2-2,5; CL – perfect; F – conchoidal to uneven; M – rhombohedral, thick tabular and prismatic crystals, massive.
Origin and occurrence: Low-temperature hydrothermal mineral, associated with realgar, mercury, pyrite, marcasite and other minerals. The world's best specimens with crystals up to 70 mm (2¾ in) across are known from many localities in China (Hunan and Guizhou provinces). Fine shiny crystals up to 10 mm (⅜ in) across come from Nikitovka, Ukraine and in Khaidarkan, Kyrgyzstan. Crystals up to 30 mm (1 3/16 in) also found in Monte Amiata and Rippa near Seravezza, Italy. Its massive aggregates are an important Hg ore in Allchar, Macedonia; Almaden, Spain and elsewhere.
Application: the most important Hg ore.

Cinnabar, 30 mm x, Hunan, China

Covellite, 65 mm, Butte, U.S.A.

Covellite
CuS

HEXAGONAL ● ● ●

Properties: C – indigo-blue, tarnishing iridescent; S – lead-gray; L – submetallic to resinous; D – opaque; DE – 4,6; H – 1,5-2; CL – perfect; F – uneven; M – hexagonal tabular crystals, massive.
Origin and occurrence: Rare hydrothermal; mainly secondary in the oxidation zone of ore deposits, associated with chalcopyrite, chalcocite, djurleite, bornite and other minerals. Tabular crystals up to 30 mm ($1^{1}/_{16}$ in) across are known from Butte, Montana and Summitville, Colorado, USA; also from Sarrabus, Sardinia, Italy. Massive aggregates are common in Bor, Serbia; Bisbee, Arizona, USA and elsewhere.
Application: Cu ore.

Linneite
$Co^{2+}Co^{3+}{}_2S_4$

CUBIC ● ● ●

Properties: C – light gray to steel-gray; S – black-gray; L – metallic; D – opaque; DE – 4,9; H – 4,5-5,5; CL – imperfect; F – uneven to conchoidal; M – octahedral crystals, granular, massive.
Origin and occurrence: Hydrothermal in ore veins and metamorphic deposits, together with chalcopyrite, pyrrhotite, millerite, bismuthinite and sphalerite. The best crystals up to 30 mm ($1^{1}/_{16}$ in) across come from the Kilembe mine, Uganda. Crystals are also known from Musonoi, Shaba, Zaire; Mhsen, Germany; siderite concretions in Kladno, Czech Republic and from the Bastnas mine near Riddarhyttan, Sweden.

Carrolite
$Cu(Co,Ni)_2S_4$

CUBIC ● ● ●

Properties: C – light to steel-gray, tarnishing red-purple; S – gray; L – metallic; D – opaque; DE – 4,8; H – 4,5-5,5; CL – imperfect; F – conchoidal to uneven; M – octahedral crystals, granular, massive.
Origin and occurrence: Hydrothermal in ore veins, associated with linneite, chalcopyrite and other minerals. It is a principal Co ore in deposits in Zaire. Beautiful crystals up to 20 mm ($^{25}/_{32}$ in) across are known from the M'Sesa mine near Kambove, Kolwezi and from Kamoto, Shaba.
Application: Co ore.

Linneite, 60 mm, Ruwenzori, Uganda

Carrollite, 8 mm x, Kambove, Zair

Stibnite

Sb_2S_3

ORTHORHOMBIC ● ● ● ●

Properties: C – steel-gray, tarnishing iridescent or black; S – lead-gray; L – metallic; D – opaque; DE – 4,6; H – 2; CL – perfect; F – conchoidal to uneven; M – thick to thin prismatic crystals, thin needles, massive.

Origin and occurrence: Hydrothermal in medium- and low-temperature ore veins with quartz and gold, the other associated minerals are rare (arsenopyrite, berthierite, gudmundite, antimony). The largest known stibnite crystals occurred in the in the Ichinokawa mine, Shikoku, Japan, where they were up to 60 cm (24 in) long. Similar sized crystals were found recently in several localities in China (the Xikuangshan mine, Hunan). Crystals up to 200 mm (7⅞ in) across come from Manhattan, Nevada, USA. Beautiful crystals are known from Romania; long prismatic ones with barite crystals are prevalent in Baia Sprie, clusters of thin needles come from Herja and thick prismatic crystals from Baiuţ. Perfect druses of crystals up to 150 mm (6 in) long, associated with purple fluorite crystals, barite and calcite, occurred in Kadamdzhai, Kyrgyzstan. Beautiful druses of stibnite crystals with quartz come also from Kremnica, Slovakia. Also Kostajnik, Serbia and La Lucette, France yielded fine crystals in the past.

Application: Sb ore.

Stibnite, 187 mm, Ichinokawa, Japan

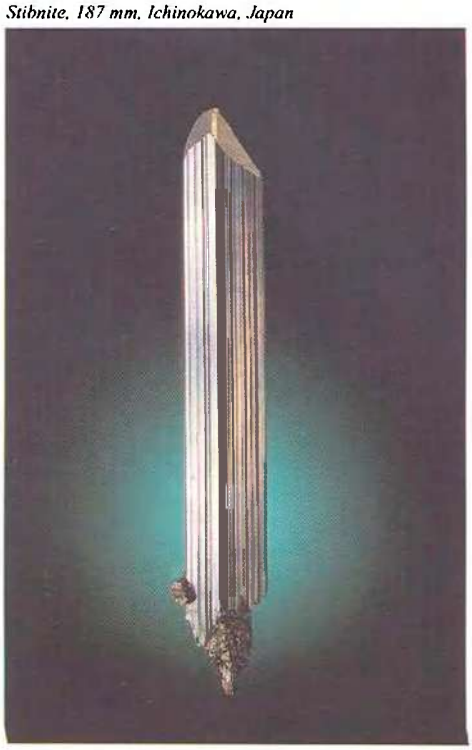

Stibnite, 120 mm, Baia Sprie, Romania

Bismuthinite, 160 mm, Tasna, Bolivia

Bismuthinite

Bi_2S_3

ORTHORHOMBIC ● ● ● ●

Properties: C – lead-gray to tin-white, tarnishing yellow and iridescent; S – lead-gray; L – metallic; D – opaque; DE – 6,8; H – 2-2,5; CL – perfect; F – uneven; M – thick prismatic to acicular, striated crystals, fibrous aggregates, massive.

Origin and occurrence: Hydrothermal in ore deposits; also in recent volcanic exhalation deposits, associated with bismuth, arsenopyrite, stannite, galena and other minerals. The world's best crystals over 50 mm (2 in) long, come from Bolivia (Tasna; Huanuni; Llallagua). Fine crystals were also found in Redruth, Cornwall, UK. Rich finds were made in Biggenden, Queensland, Australia. Interesting crystals are known from Spind, Norway.

Application: Bi ore.

Kermesite, 20 mm xx, Freiberg, Germany

Tetradymite, 80 mm, Ciclova, Romania

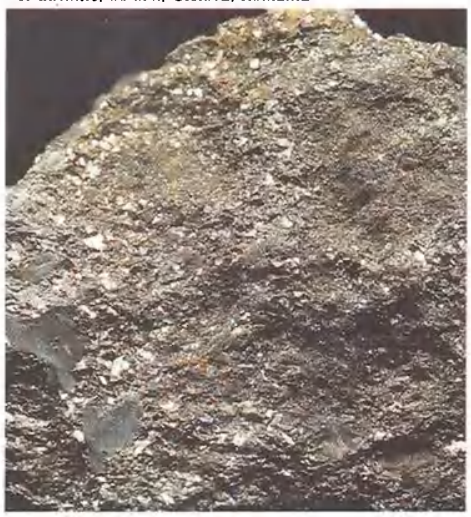

Kermesite

Sb_2S_2O

TRICLINIC ● ● ●

Properties: C – cherry-red; S – brown-red; L – adamantine to submetallic; D – translucent to opaque; DE – 4,7; H – 1-1,5; CL – perfect; M – acicular crystals, radial aggregates.

Origin and occurrence: Secondary, as a result of a stibnite oxidation in Sb deposits, associated with stibnite, antimony, senarmontite, valentinite and stibiconite. Famous specimens with needles up to 100 mm (4 in) long, in radial aggregates, come from Pezinok and Pernek, Slovakia. Crystals up to 50 mm (2 in) long found in the Globe and Phoenix mines, Zimbabwe. It is also known from Bolivia (San Francisco mine, Poopo; Oruro) and Braunsdorf, Germany.

Nagyagite, 100 mm, Sacarimb, Romania

Tetradymite
Bi_2Te_2S

TRIGONAL ● ● ●

Properties: C – light steel-gray, tarnishing dark to iridescent; S – light steel-gray; L – metallic; D – opaque; DE – 7,3; H – 1,5-2; CL – perfect; M – pyramidal crystals, curved bladed aggregates, massive.
Origin and occurrence: Hydrothermal in medium- and high-temperature Au deposits; also in contact metamorphosed deposits, together with gold, hessite, calaverite, pyrite and other minerals. Crystals up to 10 mm (3/8 in) across are known from Župkov , Slovakia. It is common in Colorado (Red Cloud) and in California (Carson Hill), USA. It occurs with gold in skarns in Baiţa Bihorului, Romania.

Nagyagite
$Pb_5Au(Te,Sb)_4S_{5-8}$

TETRAGONAL ● ●

Properties: C – black-gray; S – black-gray; L – metallic; D – opaque; DE – 7,5; H – 1,5; CL – perfect; F – uneven; M – thin tabular to foliated crystals, granular, massive.
Origin and occurrence: Hydrothermal in low-temperature veins, associated with altaite, arsenic, gold, rodochrosite and other minerals. Foliated crystals up to 40 mm (1 9/16 in) across come from Sacarîmb and Baia de Arieş , Romania. Fine specimens were also found in Tavua, Viti Levu, Fiji and in the Sylvia mine, Tararu Creek, New Zealand. It also occurred in Gold Hill and Cripple Creek, Colorado, USA and in Kalgoorlie, Western Australia, Australia.

Sylvanite
$(Au,Ag)_2Te_4$

MONOCLINIC ● ● ●

Properties: C – steel-gray to silver-white, tarnishing to yellow; S – steel-gray to silver- white; L – metallic; D – opaque; DE – 8,2; H – 1,5-2; CL – perfect; F – uneven; M – short prismatic to thick tabular crystals, skeletal aggregates, granular.
Origin and occurrence: Hydrothermal in low-temperature ore veins, also in medium- and high-temperature deposits as one of the latest minerals, associated with gold, calaverite, hessite, krennerite and other minerals. The best specimens come from Sacarîmb and Baia de Arieş , Romania, where it occurs as skeletal crystals and aggregates. Crystals were also found in Cripple Creek, Colorado, USA; crystals up to 10 mm (3/8 in) across come from the Emperor mine, Viti Levu, Fiji.

Sylvanite, 20 mm, Cripple Creek, U.S.A.

Krennerite, 3 mm xx, Sacarimb, Romania

Krennerite

$AuTe_2$

ORTHORHOMBIC ● ●

Properties: C – silver-white to brass-yellow; S – silver-white; L – metallic; D – opaque; DE – 8,9; H – 2-3; CL – perfect; F – conchoidal to uneven; M – short prismatic, striated crystals, massive.
Origin and occurrence: Hydrothermal in gold-bearing veins, associated with gold, tellurium, pyrite, quartz and other tellurides. Crystals up to 20 mm ($^{25}/_{32}$ in) across found in Tavua, Viti Levu, Fiji. Small crystals occurred in Sacarîmb and Baia de Arieş , Romania and in Cripple Creek, Colorado, USA.
Application: Au ore.

Calaverite

$AuTe_2$

MONOCLINIC ● ● ●

Properties: C – brass-yellow to silver-white; S – greenish; L – metallic; D – opaque; DE – 9,3; H – 2,5-3; CL – none; F – uneven to conchoidal; M – bladed and short prismatic, striated crystals, granular, massive.
Origin and occurrence: Hydrothermal in low-temperature Au-bearing veins, sometimes in medium- and high-temperature deposits, associated with coloradoite, altaite, krennerite and other tellurides. It is common in the Mother Lode in California (Carson Hill). It is important, together with hessite in Cripple Creek, crystals up to 10 mm ($^{3}/_{8}$ in) across come from the Cresson mine, Colorado, USA.. It is also known from several mines near Kirkland Lake, Ontario, Canada.
Application: Au ore.

Pyrite

FeS_2

CUBIC ● ● ● ● ●

Properties: C – light brass-yellow, tarnishing iridescent and darkens; S – green-black to brown-black; L – metallic; D – opaque; DE – 5,0; H – 6-6,5; CL – imperfect; F – conchoidal to uneven; M – combinations of cubic crystals, striated, stalactitic and spherical aggregates, massive.
Origin and occurrence: Magmatic segregations in basic rocks with pyrrhotite and pentlandite, in pegmatites and skarns; hydrothermal in porphyry and vein deposits together with other sulfides; hydrothermal sedimentary, sedimentary and metamorphic. Magmatic segregations are known from Sudbury, Ontario, Canada and Merensky Reef, Transvaal, South Africa. Large crystals up to 200 mm ($7^{7}/_{8}$ in) across are known from Rio Marina, Elba, Italy. Fine octahedra come from Llallagua, Bolivia. Crystals, up to 120 mm ($4^{15}/_{16}$in) across, were found in Bingham and Park City, Utah, USA. Crystals up to 150 mm (6 in) are known from Huanzala and Quiruvilca, Peru. The largest pyrite deposit is Rio Tinto, Spain, where fine-grained pyrite formed accumulations about 1 billion tons. Beautiful cubes up to 80 mm ($3^{1}/_{8}$ in) come from Navajun, Spain. Large deformed crystals

Calaverite, 21 mm, Nagyag, Romania

Pyrite, 74 mm, Navajun, Spain

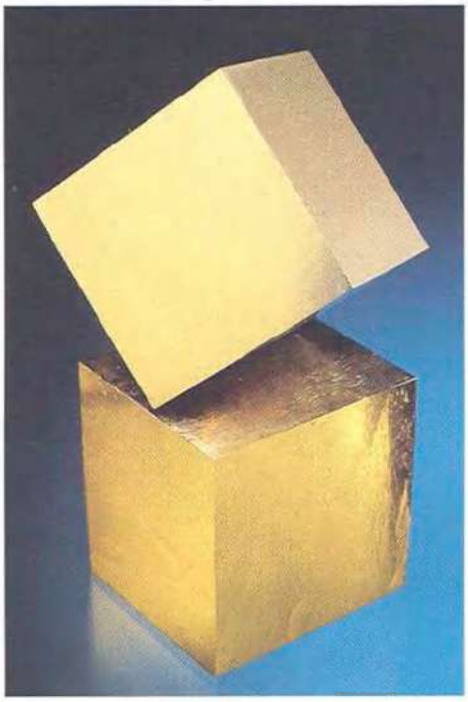

Pyrite, 135 mm, Washington, U.S.A.

up to 200 mm (7⅞ in) across occurred in the Samo mine near Hnust'a , Slovakia. Interesting, complex combinations of crystals found in Nanisivik, on an edge from Climax, Colorado, USA.
Application: production of sulfuric acid.

Hauerite

MnS_2

CUBIC ● ● ●

Properties: C – red-brown to brown-black; S – brown-red; L – metallic to adamantine; D – opaque; DE – 3,44; H – 4; CL – perfect; F – uneven to conchoidal; M – octahedral crystals and their combinations, spherical aggregates..
Origin and occurrence: Low-temperature sedimentary mineral, limited to clays with high S contents. The best crystals up to 50 mm (2 in) across come from the Destricella mine near Raddusa, Sicily, Italy. Crystals up to 25 mm (1 in) and their aggregates are known from Víglasska Huta (former Kalinka) near Zvolen, Slovakia. Crystals up to 15 mm (19/32 in) across occur in Tarnobrzeg, Poland. Also found with sulfur, gypsum, realgar and calcite in the salt domes near High Island, Texas, USA.

Hauerite, 39 mm, Raddusa, Italy

Sperrylite, 55 mm, Talnakh, Russia

Sperrylite
$PtAs_2$

CUBIC ● ●

Properties: C – tin-white; S – black; L – metallic; D – opaque; DE – 10,8; H – 6-7; CL – imperfect; F – conchoidal; M – complex combinations of cubic crystals, massive.
Origin and occurrence: It is mainly of magmatic liquid origin, associated with pyrrhotite, pentlandite, cubanite and other minerals. The best crystals come from Talnakh near Norilsk, Siberia, Russia, where inter-grown crystals reached up to 50 mm (2 in). Crystals up to 40 mm (1 9/16 in) across known from Tweefontein, Potgietersrust, Bushveld, South Africa. It also occurred as disseminated aggregates in Sudbury, Ontario, Canada.

Aurostibite
$AuSb_2$

CUBIC ● ●

Properties: C – gray, tarnishing iridescent; S – gray; L – metallic; D – opaque; DE – 9,9; H – 3; CL Bnone; M – granular.
Origin and occurrence: Hydrothermal in quartz veins, associated with gold and other Sb minerals. Grains up to 5 mm (3/16 in) across come from Krasna Hora, Czech Republic. It also occurs in the Giant Yellowknife mine, Northwest Territories and Hemlo, Thunder Bay, Ontario, Canada and is also reported from Bestyube, Kazakhstan.

Krut'aite
$CuSe_2$

CUBIC ●

Properties: C – gray; S – gray; L – metallic; D – opaque; DE – 6,5; H – 4; M – microscopic crystals, massive.
Origin and occurrence: Hydrothermal, associated with clausthalite, uraninite and other minerals. The richest accumulations of massive aggregates were found in the El Dragon mine, Potosí, Bolivia, where crystals up to 1 mm (1/32 in) occurred as well. It was originally described as microscopic from Petrovice, Czech Republic.

Cobaltite
CoAsS

ORTHORHOMBIC ● ● ●

Properties: C – silver-white; S – gray-black; L – metallic, adamantine to dull; D – opaque;DE – 6,3; H – 5,5; CL – perfect; F – uneven; M – pseudo-cubic crystals, granular, massive.
Origin and occurrence: Hydrothermal in high-temperature ore deposits and metamorphic, together

Aurostibite, 4 mm grain, Krasná Hora, Czech Republic

Krut'aite, 40 mm, El Dragon, Bolivia

Cobaltite, 5 mm x, Hakansboda, Sweden

with magnetite, sphalerite, chalcopyrite, other sulfides and arsenides of Co and Ni. The best dodecahedra, several cm across, come from metamorphic sulfide deposits in Tunaberg, Sweden. Other Swedish localities, like Håkansboda and Ramsberg, yielded crystals up to 60 mm (24 in) across. Fine crystals about 10 mm (3/8 in) occurred in the magnesite deposit Mutnik near Hnust'a , Slovakia. Cubic crystals up to 30 mm (1 3/16 in) across found in Espanola, Ontario, Canada. Beautiful crystals are known from the skarn in Bimbowrie, South Australia, Australia.
Application: Co ore.

Gersdorfite

NiAsS

CUBIC ● ● ●

Properties: C – silver-white to steel-gray, tarnishing gray-black; S – gray-black; L – metallic; D – opaque; DE – 6,0; H – 5,5; CL – perfect; F – uneven; M – octahedral striated crystals and their combinations, massive.
Origin and occurrence: Hydrothermal in medium-temperature ore deposits in association with nickeline, nickel-skutterudite, ullmanite, siderite and other minerals. Crystals up to 100 mm (4 in) across come from the Snowbird mine, Montana, USA. It is important in several mines near Sudbury and Cobalt, Ontario, Canada. Large cleavable masses are known from Dobsina, Rudnany and Nizna Slana, Slovakia. Crystals from the Aït Ahmane mine near Bou Azzer, Morocco, reached up to 40 mm (1 9/16 in) . *Application:* Ni ore.

Gersdorffite, 10 mm xx, Bou Azzer, Morocco

Ullmanite, 60 mm, Bou Cricha, Morocco

Ullmanite

NiSbS

CUBIC ● ●

Properties: C – steel-gray to silver-white; S – gray-black; L – metallic; D – opaque; DE – 6,8; H – 5-5,5; CL – perfect; F – uneven; M – combinations of cubic crystals, massive.
Origin and occurrence: Hydrothermal in ore veins together with other Ni minerals. Fine crystals, up to 20 mm (25/32 in) across, come from Monte Narba near Sarrabus, Sardinia, Italy. Twins are known from Lolling, Austria. Crystals up to 10 mm (3/8 in) were found in Ksice near Støibro, Czech Republic. It is common in Broken Hill, New South Wales, Australia and in Cochabamba, Bolivia.
Application: Ni ore.

Marcasite, 42 mm, Vintirov, Czech Republic

Marcasite

FeS_2

ORTHORHOMBIC ● ● ● ●

Properties: C – tin-white, bronze-yellow, tarnishing iridescent; S – grayish to brownish- black; L – metallic; D – opaque; DE – 4,9; H – 6-6,5; CL – good; F – uneven; M – tabular, pyramidal and prismatic crystals, often twinned into the form of cockscomb-like aggregates, stalactitic, botryoidal and massive.
Origin and occurrence: It originates at low temperatures in very acidic environment, either in sedimentary, or in hydrothermal deposits, associated with pyrite, pyrrhotite, galena, sphalerite, fluorite, dolomite and calcite. Hydrothermal crystals and pseudo-morphs after pyrrhotite are known from Freiberg, Germany; Llallagua, Bolivia and Chiuzbaia, Romania. Crystals from Wiesloch, Germany and Reocin, Santander, Spain are of similar origin. Large crystals occurred in Joplin, Missouri and in Galena, Illinois, USA. The best crystals of sedimentary marcasite come from coal basins. Fine crystals from black coal are known from Essen, Germany. Cockscomb-like aggregates up to 150 mm (6 in) across, come from the brown coal basin in Vintiøov, Czech Republic. Spherical, radial concretions with pyrite are known from Sparta, Illinois, USA and from Champagne, France.
Application: production of sulfuric acid.

Marcasite, 70 mm, Sparta, U.S.A.

Lollingite, 7 mm xx, Cobalt, Canada

Lollingite
$FeAs_2$

ORTHORHOMBIC ● ● ●

Properties: C – steel-gray to silver-white; S – gray-black; L – metallic; D – opaque; DE – 7,5; H – 5-5,5; CL – sometimes good; F – uneven; M – prismatic crystals, massive.
Origin and occurrence: Magmatic in pegmatites; hydrothermal in greisens and Sn-W veins, rare in the other types of ore veins, together with skutterudite, bismuth, nickeline, siderite, calcite and other minerals. Fine crystals are known from syenite in Langensundsfjord, Norway. Crystals up to 50 cm (20 in) across come from a pegmatite in Kaatiala, Finland. Masses occur in the Kobokobo pegmatite, Kivu, Zaire. Massive aggregates with schorl were found in Dolni Bory, aggregates with cassiterite in Pøebuz, Czech Republic.

Safflorite, 70 mm, Bou Azzer, Morocco

Safflorite
$(Co,Fe)As_2$

ORTHORHOMBIC ● ● ●

Properties: C – tin-white, tarnishing to dark gray; S – gray-black; L – metallic; D – opaque; DE – 7,5; H – 4,5-5; CL – good; F – uneven to conchoidal; M – prismatic crystals, radial aggregates, massive.
Origin and occurrence: Hydrothermal in medium-temperature with skutterudite, rammelsbergite, nickeline, silver, bismuth and lollingite. It is common in Schneeberg, Germany; Cobalt, Ontario, Canada; Batopilas, Chihuahua, Mexico and in Bou Azzer, Morocco.

Rammelsbergite
$NiAs_2$

ORTHORHOMBIC ● ● ●

Properties: C – tin-white, pinkish; S – gray-black; L – metallic; D – opaque; DE – 7,1; H – 5,5-6; CL – good; F – uneven; M – prismatic crystals, radial and fibrous aggregates, massive.
Origin and occurrence: Hydrothermal in medium-temperature veins associated with other Ni and Co minerals. Botryoidal aggregates come from Schneeberg, Germany. It is common in Sarrabus, Sardinia, Italy; in the Eldorado mine near Great Bear Lake; in Cobalt, Ontario, Canada and Bou Azzer, Morocco.

Rammelsbergite, 60 mm, Schneeberg, Germany

Arsenopyrite, 26 mm x, Portal, Mexico

Arsenopyrite
FeAsS

MONOCLINIC ● ● ● ●

Properties: C – silver-white to steel-gray; S – black; L – metallic; D – opaque; DE – 6,2; H – 5,5 – 6; CL – good; F – uneven; M – thick tabular to prismatic striated crystals, granular, massive.
Origin and occurrence: It occurs in pegmatites; hydrothermal in high-temperature vein deposits and greisens; metamorphic in contact metamorphic skarns, gneisses and mica schists. Long prismatic crystals up to 30 cm (12 in) long are known from the Obira mine, Japan. It is very common in greisen Sn and W deposits, fine crystals are known from Horni Slavkov, Czech Republic and Ehrenfriedersdorf, Germany. Beautiful crystals up to 50 mm (2 in) across come from Panasqueira, Portugal, where they occurred associated with fluorapatite, wolframite and siderite. Historically important were large crystals from Tavistock, Devon , UK. Crystals up to 40 mm (1 9/16 in) found in Llallagua, Bolivia. Crystals up to 50 mm (2 in) across were found recently in the Nikolai mine in Dalnegorsk, Russia.

Gudmundite, 10 mm xx, Polar Urals, Russia

Shiny crystals up to 30 mm (1 3/16 in) across, come from Hunan province, China. *Application:* As ore.

Gudmundite
FeSbS

MONOCLINIC ● ● ●

Properties: C – silver-white to steel-gray; S – black; L – metallic; D – opaque; DE – 7,0; H – 5-6; CL – none; F – uneven; M – prismatic twinned crystals, massive.
Origin and occurrence: Late hydrothermal mineral of ore deposits, also in metamorphic deposits and skarns. It is common in metamorphic sulfide deposits in Sweden (Boliden, Gudmundstorp). Massive aggregates are known from Kutna Hora and Vlastijovice, Czech Republic. It was common in Broken Hill, New South Wales, Australia.

Molybdenite
MoS_2

HEXAGONAL ● ● ● ●

Properties: C – lead-gray; S – blue-gray; L – metallic; D – opaque; DE – 4,0; H – 1-1,5; CL – perfect; M – tabular and prismatic crystals, scaly aggregates.
Origin and occurrence: Magmatic in pegmatites, granites and aplites, hydrothermal in high-temperature veins, also in porphyry ore deposits and in contact metamorphic deposits; associated with chalcopyrite, quartz and other minerals. Large crystals come from pegmatites in Blue Hill Bay, Maine, USA and in Mutue-Fides-Stavoren, Transvaal, South Africa, where they reach several tens of cm in size. Crystals up to 150 x70 mm (6 x 2¾ in) across come from the transitional type between pegmatites and quartz veins near Arendal and Moss, Norway; large crystals also occur in the Temiskaming district, Quebec, Canada; tabular crystals up to 120 mm (29/32 in) across found in quartz-molybdenite breccia pipes in veins in Australia (Queensland, New South Wales). Fine crystals are also known from Kladnica near Vito_a, Bulgaria; Horni Slavkov, Czech Republic and Ehrenfriedersdorf, Germany. As a fine grained disseminated ore was mined in Bingham, Utah and Climax, Colorado, USA.
Application: Mo ore.

Skutterudite
$CoAs_{2-3}$

CUBIC ● ● ●

Properties: C – tin-white to silver-gray, tarnishing to

gray and iridescent; S – black; L – metallic; D – opaque; DE – 6.8; H – 5,5 – 6; CL – good; F – conchoidal to uneven; M – combinations of cubic crystals, skeletal aggregates, granular, massive.
Origin and occurrence: Hydrothermal in medium- to high-temperature ore veins, associated with other Ni and Co minerals. Crystals up to 50 mm (2 in) across come from Bou Azzer, Morocco. Crystals up to several cm in size were found in Schneeberg and Annaberg, Germany. Large massive accumulations occur in Cobalt and Gowganda, Ontario, Canada. *Application:* Co ore.

Skutterudite, 25 mm, Schneeberg, Germany

Nickel-skutterudite

$NiAs_3$

CUBIC ● ● ● ●

Properties: C – tin-white to silver, tarnishing to gray and iridescent; S – black; L – metallic; D – opaque; DE – 6,5; H – 5,5 – 6; CL – good; F – conchoidal to uneven; M – combinations of cubic crystals, skeletal aggregates, granular.
Origin and occurrence: Hydrothermal in medium-temperature veins with arsenopyrite, arsenic, bismuth, calcite and siderite. Known in crystals from Chatham, Connecticut and Chester, Massachusetts, USA; also Val d'Anniviers, Wallis, Switzerland. Massive aggregates come from Dobsina, Slovakia; Les Chalanches, France; Mohawk mine, Michigan, USA; Schneeberg, Germany. *Application:* ruda Co a Ni.

Nickel-skutterudite, 36 mm, Saxony, Germany

Proustite, 25 mm, Chanarcillo, Chile

Proustite

Ag_3AsS_3

TRIGONAL ● ●

Properties: C – crimson, darkens upon exposure to light; S – crimson; L – adamantine; D – translucent to opaque; DE – 5,6; H – 2-2,5; CL – good; F – conchoidal to uneven; M – prismatic, rhombohedral and scalenohedral crystals, massive.

Origin and occurrence: Low-temperature hydrothermal mineral, also in the oxidation and cementation zone together with stephanite, silver, xanthoconite, acanthite and other minerals. The best specimens with crystals up to 100 mm (4 in) long come from the Dolores mine, Chanarcillo, Chile. Crystals up to 80 mm ($3^1/_8$ in) long found in the Himmelsfürst mine in Freiberg, Niederschlema and Schneeberg, Germany. Large druses with crystals up to 40 x 20 mm ($1^9/_{16}$ x $^{25}/_{32}$ in) across occurred in Jachymov; crystals up to 20 mm ($^{25}/_{32}$ in) across known from Pøibram and Stara Vozice, Czech Republic. Fine crystals come from Batopilas, Chihuahua and Sombrerete, Zacatecas, Mexico. Crystalline masses of proustite, weighing over 250 kg (550 lb), were found in 1865 in the Poorman mine, Silver City, Idaho, USA. *Application:* Ag ore.

Pyrargyrite, 61 mm, Zacatecas, Mexico

Pyrargyrite
Ag_3SbS_3

TRIGONAL ● ●

Properties: C – dark red, darkens upon exposure to light; S – crimson; L – adamantine; D – translucent to opaque; DE – 5,9; H – 2,5; CL – good; F – conchoidal to uneven; M – prismatic, rhombohedral and scalenohedral crystals, granular, massive.
Origin and occurrence: Low-temperature hydrothermal mineral, also secondary in the oxidation and cementation zone, together with silver, acanthite, other Ag sulfosalts, calcite and quartz. Crystals in Colquechaca, Bolivia and Chañarcillo, Chile reached several cm in size. Crystals up to 70 mm (2¾ in) long occurred in the Santo Nino vein in Fresnillo, Zacatecas, Mexico. San Genaro mine in Huancavelica, Peru yielded crystals up to 50 mm (2 in) across. Crystals up to 40 mm (1 9/16 in) across found in Freiberg; smaller crystals only are known from St. Andreasberg, Germany. Crystals up to 20 mm (25/32 in) across come from Pribram and Stara Vozice, Czech Republic. Crystals up to 50 mm (2 in) across found in the San Carlos mine, Hiendelaencina, Spain.
Application: Ag ore.

Xanthoconite
Ag_3AsS_3

MONOCLINIC ● ●

Properties: C – dark crimson, orange-yellow to yellow-brown; S – orange-yellow; L – adamantine; D – translucent; DE – 5,5; H – 2-3; CL – good; F – conchoidal; M – tabular and lath-like crystals, botryoidal and radial aggregates.

Xanthoconite, 2 mm xx, Marienberg, Germany

Pyrostilpnite, 60 mm, Potosí, Bolivia

Origin and occurrence: Hydrothermal in ore veins together with proustite, pyrargyrite, acanthite, arsenic and calcite. Botryoidal masses with yellow crystals up to 7mm (9/32 in) long come from Freiberg; other important localities are St Andreasberg, Germany; Ste-Marie-aux- Mines, France; Cobalt, Ontario, Canada; Pribram, Trebsko and Jachymov, Czech Republic.

Pyrostilpnite
Ag_3SbS_3

MONOCLINIC ● ●

Properties: C – hyacinth- to orange-red; S – orange-yellow; L – adamantine; D – translucent; DE – 6,0; H – 2; CLB perfect; F – conchoidal; M – tabular to lath-like crystals, radial aggregates.
Origin and occurrence: Hydrothermal in low-temperature veins, associated with pyrargyrite, stephanite, acanthite and other Ag minerals. The best crystals come from St. Andreasberg, Germany. Crystals up to 10 mm (3/8 in) long were found in Pribram, Trebsko and Jachymov, Czech Republic. It is also described from Colquechaca, Bolivia and Chañarcillo, Chile.

Samsonite
$Ag_4MnSb_2S_6$

MONOCLINIC ●

Properties: C – steel-gray; S – dark red; L – metallic; D – opaque; DE – 5,5; H – 2,5; CL – none; F – conchoidal; M – prismatic striated crystals.
Origin and occurrence: The only locality, where it occurred in relatively larger amount, was the Samson mine in St. Andreasberg, Germany, where crystals up to 10 mm (3/8 in) across were found.

Samsonite, 12 mm x, St.Andreasberg, Germany

Chalcostibite
$CuSbS_2$

ORTHORHOMBIC ● ●

Properties: C – lead-gray, tarnishing to blue and green; S – lead-gray; L – metallic; D – opaque; DE – 5,0; H – 3-4; CL – perfect; F – conchoidal; M – long prismatic, striated crystals, granular, massive.
Origin and occurrence: Hydrothermal in ore veins, associated with jamesonite, chalcopyrite, tetrahedrite, stibnite, andorite and other minerals. Partly altered crystals up to 100 mm (4 in) long are known from Rar-el-Auz near Casablanca, Morocco. It is often in deposits in Bolivia (Huanchaca, Oruro, Colquechaca), where its crystals reach 10 mm (3/8 in)

Chalcostibite, 29 mm, Saint Pons, France

Emplectite, 60 mm, Schwarzenberg, Germany

in size. The world's best crystals up to 16 cm (6 5/8 in) long were discovered recently near St Pons, France.

Emplectite

$CuBiS_2$

ORTHORHOMBIC ● ● ●

Properties: C – gray to tin-white; S – gray; L – metallic; D – opaque; DE – 6,4; H – 2; CL – perfect; F – conchoidal to uneven; M – thin prismatic to acicular striated crystals.

Origin and occurrence: Hydrothermal in high-temperature veins associated with chalcopyrite,

Berthierite, 98 mm, Herja, Romania

Wittichenite, 50 mm, Wittichen, Germany

molybdenite, quartz, tetrahedrite and other minerals. Fine acicular crystals up to 30 mm (1 3/16 in) long come from Krupka, Czech Republic. Crystallized specimens were also found in Wittichen and Johanngeorgenstadt, Germany; Colquijirca, Peru and in the Akenobe mine, Japan.

Wittichenite

Cu_3BiS_3

ORTHORHOMBIC ● ●

Properties: C – steel-gray to tin-white, tarnishing yellow to steel-gray; S – black; L – metallic; D – opaque; DE – 6,2; H – 2-3; CL – none; F – conchoidal; M – prismatic crystals, massive.

Origin and occurrence: Hydrothermal in ore veins, associated with other Bi minerals, Cu-Fe sulfides, selenides and secondary U minerals. It occurs in Wittichen, Germany; Baiţa Bihorului, Romania; Tsumeb, Namibia and Cerro de Pasco, Peru.

Berthierite

$FeSb_2S_4$

ORTHORHOMBIC ● ● ●

Properties: C – dark steel-gray, tarnishing iridescent to brown; S – dark brown-gray; L – metallic; D – opaque; DE – 4,7; H – 2-3; CL – imperfect; M – long prismatic, striated crystals, fibrous, felt-like and radial aggregates.

Origin and occurrence: Hydrothermal in low-temperature Sb deposits. Acicular crystals up to 10 mm (3/8 in) long are known from the St. Antoni de Padua gallery in Kutna Hora, Czech Republic; thick prismatic crystals come from Poproc, Slovakia. Iridescent columnar aggregates up to 200 mm (7 7/8 in) long occur in Herja, Romania. Fine specimens are known also from Oruro, Bolivia.

Stephanite, 8 mm xx, Zacatecas, Mexico

Stephanite
Ag_5SbS_4

ORTHORHOMBIC ● ●

Properties: C – black; S – black; L – metallic; D – opaque; DE – 6,3; H – 2-2,5; CL – imperfect; F – conchoidal; M – short prismatic to tabular striated crystals, massive.
Origin and occurrence: Late hydrothermal mineral in Ag deposits, associated with proustite, acanthite, silver, tetrahedrite, galena, sphalerite and pyrite. Crystals up to 40 mm (1 9/16 in) long come from Pribram and Jachymov, Czech Republic and from St. Andreasberg, Germany. Crystals up to 50 mm (2 in) across found in the Las Chispas mine, Arizpe, Sonora, Mexico and Hiendelaencina, Spain. Smaller crystals occurred in Freiberg, Schneeberg and Annaberg, Germany. *Application:* Ag ore.

Pearceite
$Ag_{16}As_2S_{11}$

MONOCLINIC ● ●

Properties: C – black; S – black; L – metallic; D – opaque; DE – 6,1; H – 3; CL – none; F – conchoidal to irregular; M – short prismatic to tabular crystals and rosette-like aggregates, massive.
Origin and occurrence: Hydrothermal in low- to medium-temperature deposits with acanthite, silver, proustite, quartz, barite and calcite. Crystals several mm across are known from Jachymov, Moldava and Midinec, Czech Republic; from Arqueros, Chile and from the Veta Rica mine, Coahuila, Mexico. Crystals up to 12 mm (15/32 in) across were found in the Caribou mine, Colorado, USA; also in Dzhezkazgan, Kazakhstan.
Huge accumulations of almost pure pearceite occurred in the Mollie Gibson mine near Aspen, Colorado, USA.
Application: Ag ore.

Polybasite
$(Ag,Cu)_{16}Sb_2S_{11}$

MONOCLINIC ● ●

Properties: C – black; S – black; L – metallic; D Bopaque; DE – 6,4; H – 2-3; CL – imperfect; F – uneven; M – pseudo-hexagonal tabular crystals, massive.
Origin and occurrence: Hydrothermal in low- to medium-temperature ore veins, associated with

Pearceite, 15 mm xx, Guanajuato, Mexico

Polybasite, 39 mm, Zacatecas, Mexico

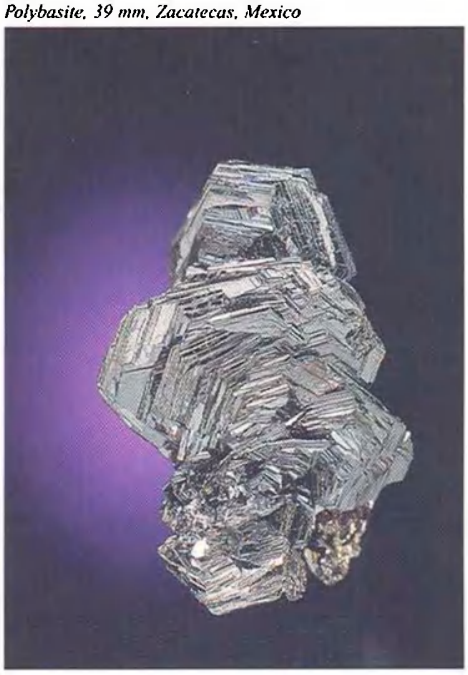

pyrargyrite, tetrahedrite, stephanite, acanthite, quartz and other minerals. Crystals, several cm across, are known from Wolfach, St. Andreasberg, Freiberg and Schneeberg, Germany; also from Guanajuato, Mexico. The best specimens with tabular crystals up to 90 mm (3 9/16 in) across come from the Las Chispas mine, Arizpe, Sonora, Mexico.
Application: Ag ore.

Lorandite

$TlAsS_2$

MONOCLINIC ● ●

Properties: C – crimson, lead-gray, it covers with a yellow coating; S – cherry-red; L – metallic to adamantine; D – translucent to transparent; DE – 5,5; H – 2-2,5; CL – perfect; M – short prismatic to tabular crystals, granular, massive.
Origin and occurrence: Hydrothermal, associated with stibnite, realgar, orpiment, pyrite and other minerals. Crystals up to 50 mm (2 in) across were found in Allchar, Macedonia. It is also known from Djijikrut, Tajikistan and from the cavities in dolomite from Lengenbach, Binntal, Switzerland.

Livingstonite

$HgSb_4S_8$

MONOCLINIC ● ●

Properties: C – black-gray; S – red; L – metallic to adamantine; D – opaque; DE – 5,0; H – 2; CL – perfect; M – acicular crystals, columnar and fibrous aggregates, massive.
Origin and occurrence: Hydrothermal in low-temperature veins, associated with cinnabar, stibnite, getchellite and other minerals. The best specimens with prismatic crystals up to 50 mm (2 in) long are known from Khaidarkan, smaller crystals only found from Kadamdzhai, Kyrgyzstan. It is also described from the La Cruz mine, Huitzuco, Guerrero, Mexico and from the Matsuo mine, Japan.

Lorandite, 10 mm x, Allchar, Macedonia

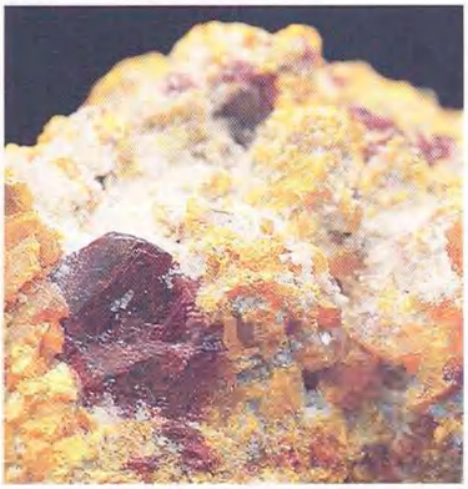

Livingstonite, 40 mm, Khaidarkan, Kyrgystan

Bournonite. 62 mm. Cornwall. UK

Bournonite
$PbCuSbS_3$

ORTHORHOMBIC ● ● ● ●

Properties: C – steel-gray to black; S – steel-gray to black; L – adamantine to dull; D – opaque; DE – 5,8; H – 2,5-3; CL – imperfect; F – conchoidal to uneven; M – short prismatic to tabular crystals, often striated, granular, massive.
Origin and occurrence: Hydrothermal in medium-temperature ore deposits, together with galena, tetrahedrite, pyrite, siderite and other minerals. The finest crystals, complex twins called cogswheel ore, over 50 mm (2 in) across come from the Herodsfoot mine near Liskeard, Cornwall, UK. Tabular crystals up to 40 mm (1 9/16 in) across found in siderite cavities in Příbram, Czech Republic; large prismatic and tabular crystals known from Neudorf, Germany. Smaller crystals common in Cavnic and Baia Sprie, Romania. Crystals, up to 100 mm (4 in) across, occurred in Machacamarca, Bolivia. Crystals up to 40 mm (1 9/16 in) come from Huancavelica, Peru. Crystals up to 100 mm (4 in) across reported from the Les Malines mine, France. Crystals up to 20 mm (25/32 in) across found at Chenzhou, Hunan province, China. *Application:* Pb, Cu and Sb ore.

Aikinite
$PbCuBiS_3$

ORTHORHOMBIC ● ●

Properties: C – black-gray, tarnishing brown; S – gray-black; L – metallic; D – opaque; DE – 7,3; H –

Aikinite. 2 mm xx. Rud?any. Slovakia

Betekhtinite, 10 mm xx, Dzhezkazgan, Kazakhstan

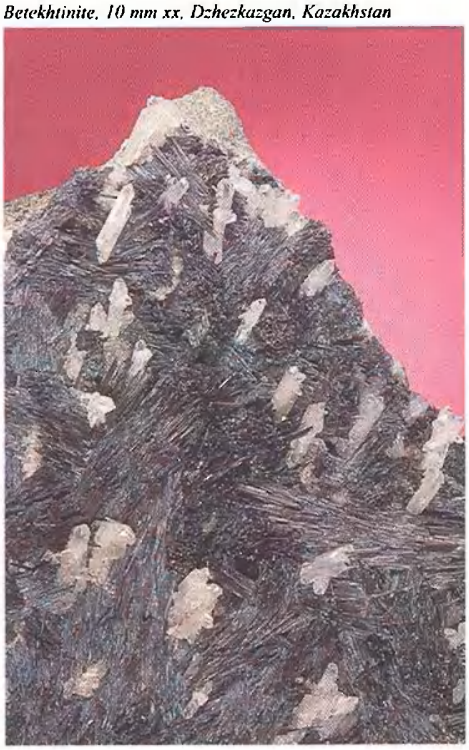

2-2,5; CL – imperfect; F – uneven; M – prismatic to acicular, striated crystals, massive.
Origin and occurrence: Hydrothermal in ore veins, associated with gold, pyrite, galena, tennantite and other minerals. It is common in quartz veins with gold in Berezovsk, Ural Mountains, Russia, where crystals up to 30 mm (1 3/16 in) long were found; fine crystals come also from La Gardette, Bourg d'Oisans, France. Grains up to 50 mm (2 in) across occurred in the Outlaw mine, Nevada, USA. It is also known in metamorphic veins in Val d'Anniviers, Switzerland.

Betekhtinite

$Cu_{10}(Fe,Pb)S_6$

ORTHORHOMBIC ● ●

Properties: C – black; S – black; L – metallic; D – opaque; DE – 6,1; H – 3-3,5; CL – good; M – acicular crystals, granular.
Origin and occurrence: Hydrothermal in ore deposits. The best specimens come from Dzhezkazgan as clusters of acicular crystals up to 70 mm (2¾ in) long, associated with bornite, chalcocite, djurleite and other minerals. Rich specimens found in Kipushi Kipushi, Shaba, Zaire. Granular aggregates fairly common in calcite veinlets, cross-cutting Cu-bearing shales near Eisleben, Mansfeld, Germany.

Andorite

$PbAgSb_3S_6$

ORTHORHOMBIC ● ●

Properties: C – dark steel-gray, tarnishing to yellow and iridescent; S – black; L – metallic; D – opaque; DE – 5,4; H – 3-3,5; CL – none; F – conchoidal; M – prismatic and tabular striated crystals, massive.
Origin and occurrence: Hydrothermal in ore deposits, together with cassiterite, jamesonite, stannite and other minerals. The world's best specimens come from the Itos and San Jose mines in Oruro and the Tatasi mine in Potosí, Bolivia, where it forms crystals up to 30 mm (1 3/16 in) across. Thin tabular crystals are known from Baia Sprie, Romania and from the Keyser mine, Nevada, USA. Needles up to 10 mm (⅜ in) long occurred in Trebsko, Czech Republic.
Application: ruda Ag.

Andorite, 45 mm, Oruro, Bolivia

Freieslebenite. 2 mm xx, Baia Sprie, Romania

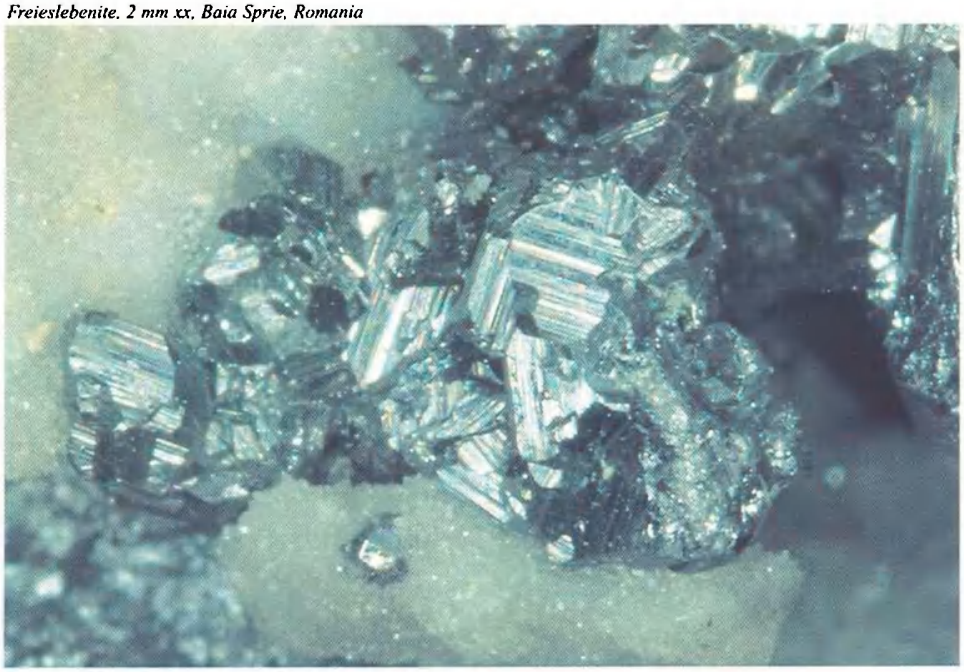

Freieslebenite
$AgPbSbS_3$

MONOCLINIC ● ●

Properties: C – light steel-gray, lead-gray to silver-white; S – light steel-gray, lead-gray to silver-white; L – metallic; D – opaque; DE – 6,2; H – 2-2,5; CL – imperfect; F – conchoidal to uneven; M – prismatic, striated crystals.
Origin and occurrence: Hydrothermal in ore deposits, associated with acanthite, pyrargyrite, silver, galena, siderite and andorite.
Crystals up to 20 mm ($^{25}/_{32}$ in) across come from the Santa Cecilia, Guadalajara, Spain; Freiberg, Germany; Oruro, Bolivia and from the Treasury Lode, Colorado, USA.

Diaphorite
$Pb_2Ag_3Sb_3S_3$

MONOCLINIC ● ●

Properties: C – steel-gray; S – steel-gray; L – metallic; D – opaque; DE – 6,0; H – 2,5-3; CL – none; F – conchoidal to uneven; M – prismatic striated crystals.
Origin and occurrence: Hydrothermal in medium-temperature ore veins, associated with galena,

Diaphorite, 5 mm xx, Pribram, Czech Republic

Sartorite, 4 mm x, Binntal, Switzerland

sphalerite, pyrargyrite, pyrite and other minerals. Beautiful striated crystals up to 10 mm (⅜ in) come from the cavities of quartz veins in Poíbram; rare small crystals, several mm across found in the cavities of quartz veins in the St. Antoni de Padua gallery in Kutna Hora, Czech Republic. Complicated combinations of crystals were described from Braunsdorf, Germany. Crystals up to 80 mm (3⅛ in) long occurred in Hiendelaencina, Guadalajara, Spain. It is also reported from Catorce, San Luis Potosí, Mexico.

Sartorite

$PbAs_2S_4$

MONOCLINIC ● ●

Properties: C – dark lead-gray; S – chocolate-brown; L – metallic; D – opaque; DE – 5,1; H – 3; CL – good; F – conchoidal; M – prismatic striated crystals.
Origin and occurrence: Hydrothermal in dolomite, associated with tennantite, dufrenoysite, pyrite and realgar. The best crystals up to 100 mm (4 in) long come from Lengenbach, Binntal, Switzerland. It was also found in the Zuni mine, Colorado, USA.

Baumhauerite

$Pb_3As_4S_9$

TRICLINIC ●

Properties: C – lead- to steel-gray, tarnishing iridescent; S – chocolate-brown; L – metallic; D – opaque; DE – 5,4; H – 3; CL – perfect; F – conchoidal; M – tabular to short prismatic striated crystals, granular.
Origin and occurrence: Hydrothermal, associated with realgar and other sulfosalts. The best crystals up to 25 mm (1 in) across come from Lengenbach, Binntal, Switzerland. Massive aggregates were found in Hemlo, Thunder Bay, Ontario, Canada and Sterling Hill, New Jersey, USA.

Baumhauerite, 50 mm, Binntal, Switzerland

Rathite, 60 mm, Binntal, Switzerland

Rathite

$(Pb,Tl)_3As_5S_{10}$

MONOCLINIC ●

Properties: C – lead-gray, tarnishing iridescent; S – chocolate-brown; L – metallic; D – opaque; DE – 5,3; H – 3; CL – perfect; F – conchoidal; M – prismatic striated crystals.
Origin and occurrence: Hydrothermal, associated with other Pb-Tl-As-S minerals. Crystals up to 10 mm (⅜ in) across come from dolomite in Lengenbach, Binntal, Switzerland.

Dufrenoysite

$Pb_2As_2S_5$

MONOCLINIC ● ●

Properties: : C – lead- to steel-gray; S – red-brown to chocolate-brown; L – metallic; D – opaque; DE – 5,6; H – 3; CL – perfect; F – conchoidal; M – elongated striated tabular crystals.
Origin and occurrence: Hydrothermal low-temperature mineral, associated with rathite, sartorite, baumhauerite and realgar. Crystals up to 25 mm (1 in) across come from dolomite in Lengenbach, Binntal, Switzerland. Also found in Batopilas, Chihuahua, Mexico.

Dufrenoysite, 10 mm x, Binntal, Switzerland

Jordanite, 12 mm x, Binntal, Switzerland

Jordanite

$Pb_{14}(As,Sb)_6S_{23}$

MONOCLINIC ● ●

Properties: C – lead-gray, tarnishing iridescent; S – black; L – metallic; D – opaque; DE – 6,4; H – 3; CL – perfect; F – conchoidal; M – tabular crystals, botryoidal aggregates.

Geocronite, 70 mm, Pribram, Czech Republic

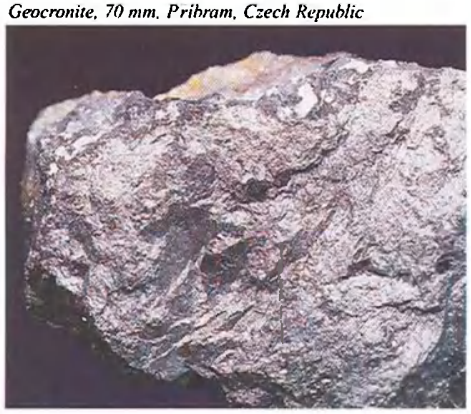

Origin and occurrence: Hydrothermal in low-temperature ore veins, also in metamorphic dolomites, together with tennantite, sphalerite, galena, dolomite and other minerals. The most famous crystals up to 50 mm (2 in) across come from cavities in dolomite in Lengenbach, Binntal, Switzerland. Tabular crystals occurred also in Sacarîmb , Romania. Botryoidal aggregates, growing on barite crystals, were described from the Yunosawa mine, Japan.

Geocronite

$Pb_{14}(Sb,As)_6S_{23}$

MONOCLINIC ● ●

Properties: C – light lead-gray; S – light lead-gray to gray-blue; L – metallic; D – opaque; DE – 6,4; H – 2,5; CL – good; F – uneven; M – tabular crystals, massive.
Origin and occurrence: Hydrothermal in ore veins, associated with galena, pyrite, tetrahedrite, barite, fluorite and quartz. Crystals, up to 80 mm (3 1/8 in) across, come from Pietrasanta, Italy. Crystals up to 90 mm (3 9/16 in) found in Virgem da Lapa, Brazil. Tabular crystals up to 40 mm (1 9/16 in) across occurred in the Kilbricken mine, Ireland. Massive aggregates are known from Příbram, Czech Republic.

Zinkenite
$Pb_9Sb_{22}S_{42}$

HEXAGONAL ● ● ●

Properties: C – steel-gray, tarnishing iridescent; S – steel-gray; L – metallic; D – opaque; DE – 5,3; H – 3-3,5; CL – imperfect; F – uneven; M – thin prismatic striated crystals, radial to felt-like aggregates, massive.
Origin and occurrence: Hydrothermal in ore veins, associated with stibnite, jamesonite, boulangerite, bournonite, stannite and other minerals. Crystals up to 50 mm (2 in) across are known from the Itos and San Jose mines in Oruro, Bolivia. It was also found in Wolfsberg, Germany; St. Pons, France; Sacarîmb and Baia Sprie, Romania.

Jamesonite
$Pb_4FeSb_6S_{14}$

MONOCLINIC ● ● ● ●

Properties: C – gray-black, tarnishing iridescent; S – gray-black; L – metallic; D – opaque; DE – 5,8; H – 2,5; CL – good; F – uneven; M – acicular crystals, fibrous and felt-like aggregates, massive.

Zinkenite, 39 mm, Oruro, Bolivia

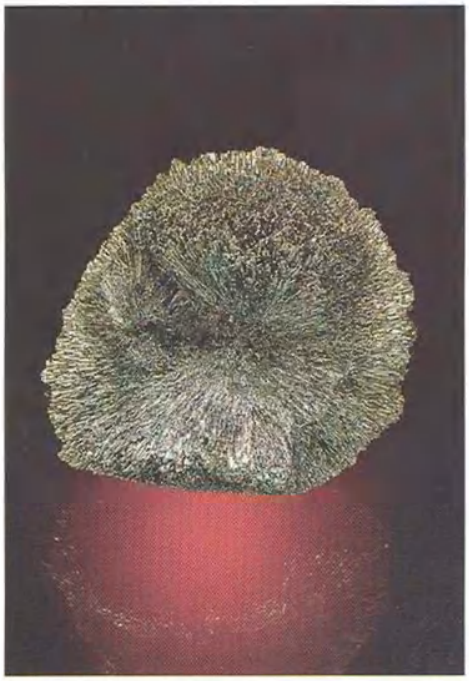

Jamesonite, 30 mm xx, Sombrerete, Mexico

Origin and occurrence: Hydrothermal in medium- and low-temperature base metal ore veins, associated with other Pb-sulfosalts, pyrite, sphalerite, galena, tetrahedrite, quartz and other minerals. Needles up to 80 mm (3 1/8 in) long occur in many localities in Bolivia (Tasna; Bolivia mine, Poopo; San Jose and Itos mines, Oruro). It also comes from Wolfsberg and Freiberg, Germany; Nizna Slana , Slovakia and Sombrerete, Zacatecas, Mexico.
Application: Pb and Sb ore.

Semseyite
$Pb_9Sb_8S_{21}$

MONOCLINIC ● ●

Properties: C – gray to black; S – black; L – metallic; D – opaque; DE – 6,1; H – 2,5; CL – perfect; M – tabular and prismatic crystals and their rosette-like aggregates.
Origin and occurrence: Hydrothermal in medium-temperature veins, associated with jamesonite, bournonite, zinkenite, sphalerite and other minerals. Fine rosette-like aggregates of crystals over 10 mm (3/8 in) across come from Baia Sprie, Herja and Rodna, Romania. It was also found in the San Jose mine, Oruro, Bolivia; Wolfsberg, Germany and Huancavelica, Peru.

Boulangerite
$Pb_5Sb_4S_{11}$

MONOCLINIC ● ● ● ●

Properties: C – lead-gray; S – brownish; L – metallic to silky; D – opaque; DE – 6,2; H – 2,5-3; CL – good; M – acicular striated crystals, fibrous and felt-like aggregates.
Origin and occurrence: Hydrothermal in low- and medium-temperature ore veins, together with other Pb sulfosalts, galena, sphalerite and other minerals. Fine needles over 100 mm (4 in) long come from cavities in quartz in Příbram, Czech Republic. It is common in the Coeur d'Alene district, Idaho, USA. Acicular crystals up to 30 cm (12 in) long were found in Trepea, Serbia and Leadville, Colorado, USA. It is also known from Wolfsberg, Germany and Bolivia (Colquechaca, Huanuni, Isca-Isca).

Cosalite
$Pb_2Bi_2S_5$

ORTHORHOMBIC ● ● ●

Properties: C – lead- to steel-gray, silver-white; S – black; L – metallic; D – opaque; DE – 7,1; H – 2,5-3; CL – none; F – uneven; M – prismatic to acicular crystals, radial and fibrous aggregates.
Origin and occurrence: Magmatic in pegmatites, hydrothermal in medium-temperature deposits; also metamorphic, associated with sphalerite, chalco-

Semseyite, 40 mm, Cavnic, Romania

Boulangerite, 40 mm, Zacatecas, Mexico

Cosalite, 3 mm xx, Ocna de Fier, Romania

pyrite, pyrite, cobaltite and other minerals. Elongated crystals are known from Crodo, Italy. Fine needles up to 40 mm (1 9/16 in) long, included in quartz crystals, were found in Kara-Oba, Kazakhstan. It occurs also in Au deposits (Homestake mine, South Dakota, USA) or in skarns (Baiţa Bihorului, Romania).

Kobellite

$Pb_{22}Cu_4(Bi,Sb)_{30}S_{69}$

ORTHORHOMBIC ● ● ●

Properties: C – black-gray to steel-gray; S – black; L – metallic; D – opaque; DE - 6,5; H - 2,5-3; CL – good; M – fibrous aggregates, granular, massive.
Origin and occurrence: Hydrothermal in high-temperature veins and pegmatites, together with cobaltite, arsenopyrite, chalcopyrite and other minerals. It is known from a sulfide rich pegmatite in the Superior Stone quarry, North Carolina, USA. Massive aggregates are common in Jedl'ovec , Slovakia. It was originally described from the Hvena mine near Åskersund, Sweden.

Kobellite, 20 mm xx, Rozňava, Slovakia

Realgar, 64 mm, Shimen, China

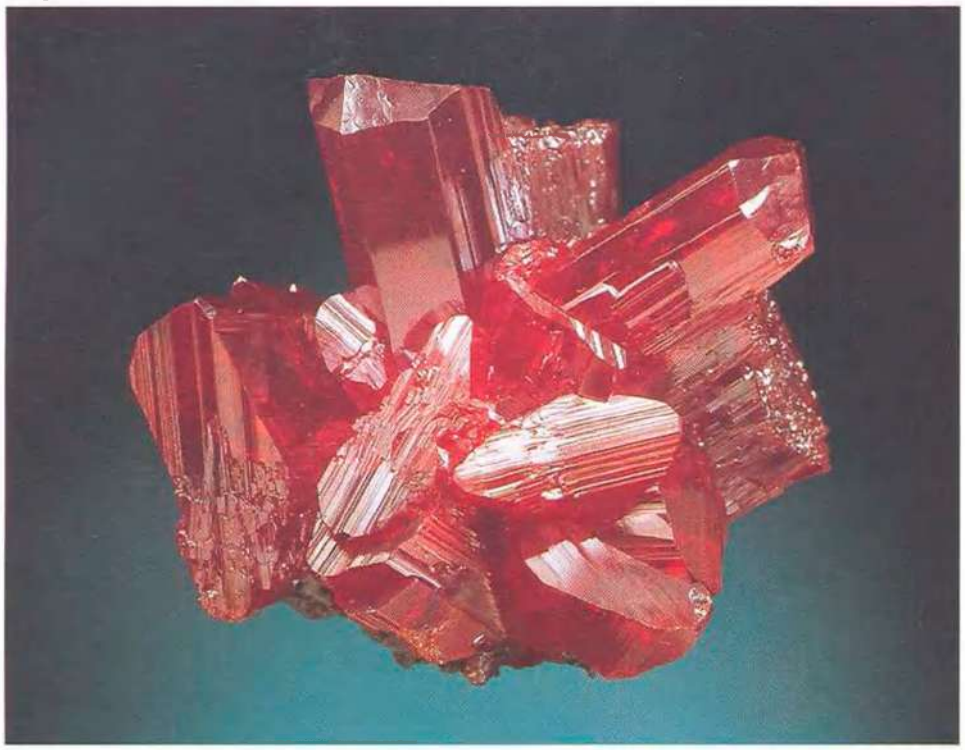

Realgar
AsS

MONOCLINIC ● ● ● ●

Properties: C – red to orange-yellow; S – orange-red to red; L – resinous to greasy; D – transparent to translucent; DE – 3,6; H – 1,5-2; CL – good; F – conchoidal; M – prismatic striated crystals, granular, massive.

Orpiment, 40 mm, Baia Sprie, Romania

Origin and occurrence: Hydrothermal in low-temperature veins, associated with other As-Sb minerals; also as a sublimation product of volcanic gasses, in hot springs and sediments. The most beautiful crystals over 100 mm (4 in) long come from Shimen, Hunan, China. Crystals up to 70 mm (2¾ in) long occurred in the Getchell mine, Nevada, USA. Crystals up to 80 mm (3⅛ in) long were found in Lengenbach, Binntal, Switzerland. Crystals are also known from Baia Sprie, Romania. Massive aggregates are common in Allchar, Macedonia.

Orpiment
As_2S_3

MONOCLINIC ● ● ● ●

Properties: C – lemon-yellow to bronze-yellow; S – light lemon-yellow; L – resinous to pearly; D – transparent to translucent; DE – 3,5; H – 1,5-2; CL – perfect; M – prismatic crystals, foliated and fibrous aggregates.
Origin and occurrence: Hydrothermal in low-temperature veins, together with realgar, stibnite, calcite etc., also from hot springs and fumaroles. It is also a common product of realgar oxidation. The best crystals up

to 100 mm (4 in) long come from Shimen, Hunan, China. Fine cleavable lamellae occur in Lukhumi, Georgia and Men-Kyule, Yakutia, Russia. Crystals up to 50 mm (2 in) across found in the La Libertad mine, Quiruvilca and Huayllapon, Ancash, Peru. Crystals up to 80 mm (3¼ in) long described from the Getchell mine, Nevada, USA.. Fine specimens are also known from Allchar, Macedonia and Khaidarkan, Kyrgyzstan. *Application:* As ore, pigment.

Getchellite

$AsSbS_3$

MONOCLINIC ● ●

Properties: C – dark red, tarnishing green and iridescent; S – orange-red; L – pearly to glassy, resinous; D – transparent; DE – 4,0; H – 1,5-2; CL – perfect; F – splintery; M – imperfect curved crystals, massive.
Origin and occurrence: Hydrothermal in low-temperature ore deposits, associated with orpiment, realgar, stibnite, cinnabar and other minerals. It was described from the Getchell mine, Nevada, USA. Beautiful specimens with grains up to several cm across come from Khaidarkan, Kyrgyzstan. It is also known from Zarehshuran, Kurdistan, Iran.

Getchellite, 60 mm, Khaidarkan, Kyrgystan

Orpiment, 35 mm, Huayllapon, Peru

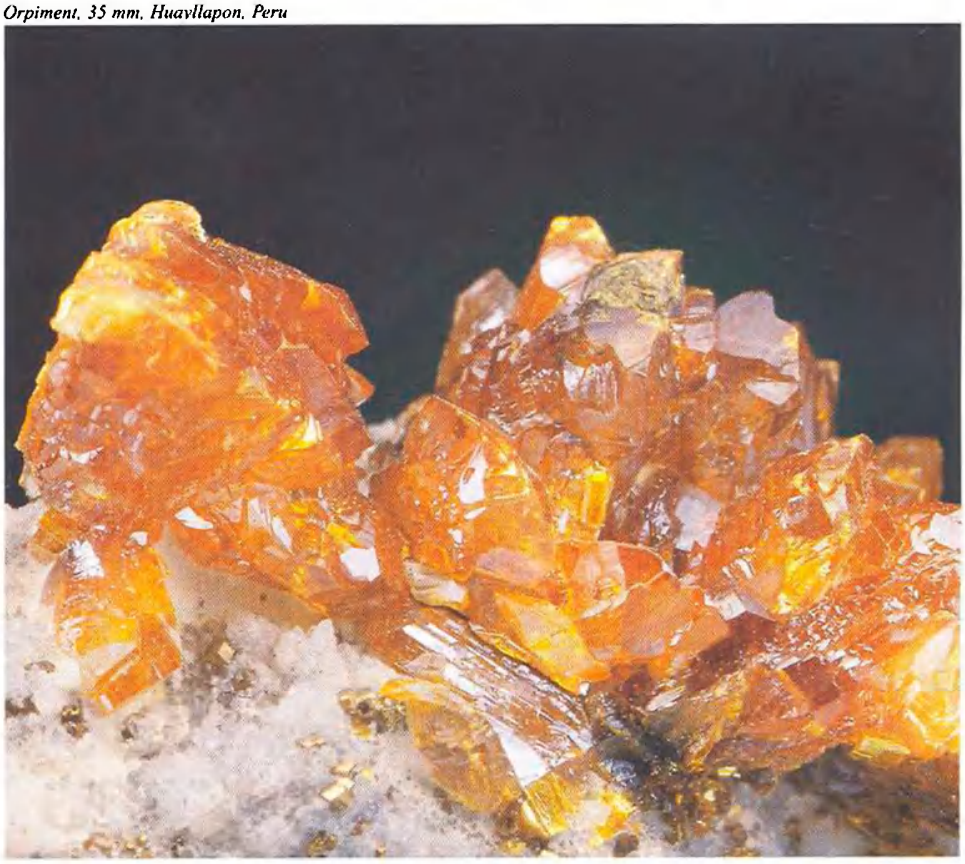

3. Halides

Fluellite

$Al_2(PO_4)F_2(OH) . 7 H_2O$

ORTHORHOMBIC ● ●

Properties: C – colorless, white, yellow; S – white; L – vitreous; D – transparent; DE – 2.2; H – 3; CL – imperfect; M – dipyramidal crystals.
Origin and occurrence: Hydrothermal in greisens, also secondary as a result of triplite alteration. Crystal druses over 10 mm (3/8 in) in size come from Horní Slavkov, Czech Republic. Very similar specimens were found in Stenna Gwyn near St. Austell, Cornwall, UK. Also found in pegmatites in Kynzvart, Czech Republic and in Hagendorf, Germany as a product of triplite alteration.

Fluellite, 10 mm x, Horní Slavkov, Czech Republic

Cryolite

Na_3AlF_6

MONOCLINIC ● ● ●

Properties: C – colorless, white, purple, brownish; S – white; L – greasy to pearly; D – transparent to translucent; DE – 3; H – 2.5; CL – none; F – uneven; M – pseudo-cubic crystals, massive.

Origin and occurrence: Characteristic mineral of the cryolite pegmatites. Crystals up to 30 mm (1 3/16 in) in size were found in Ivigtut, Greenland, where it was mined as the Al ore for more than 100 years. It was associated with other aluminofluorides, sphalerite, cassiterite, ferrocolumbite and other minerals. It is also known from the Francon quarry in Montreal, Quebec, Canada in crystals up to 10 mm (3/8 in) across. Massive cryolite occurs in Miass, Ural Mountains, Russia and in St. Peter's Dome, Pikes Peak batholith, Colorado, USA.
Application: it was an important Al ore.

Fluorite, 67 mm, Berbes, Spain
Cryolite, 35 mm, Ivigtut, Greenland

Creedite. 70 mm. Santa Eulalia. Mexico

Creedite

$Ca_3Al_2(SO_4)(F,OH)_{10} \cdot 6\ H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, purple; S – white; L – vitreous; D – transparent; DE – 2.7; H – 4; CL – perfect; F – conchoidal; M – short prismatic to acicular crystals, granular, massive.
Origin and occurrence: Hydrothermal, associated with fluorite and barite. Purple crystals, several cm long, come from Wagon Wheel Gap near Creede, Colorado, USA. Nice druses were found in Santa Eulalia, Chihuahua, Mexico. The best creedite specimens with purple crystals up to 30 mm (1 3/16 in) long were recently found in Akcha-tau, Kazakhstan.

Carnallite

$KMgCl_3 \cdot 6\ H_2O$

ORTHORHOMBIC ● ● ● ●

Properties: C – colorless, white, yellowish, red, blue; S – white; L – vitreous to greasy; DE – 1.6; H – 1-2; CL – none; F – conchoidal; M – pseudo-hexagonal pyramidal and tabular crystals, granular; LU – strong; R – decomposes under wet conditions.
Origin and occurrence: Sedimentary, one of the last products of evaporation of salty solutions; also supergene as a product of a reaction of older salts with solutions, rich in potassium, associated with halite, sylvite and other minerals. Crystals up to 40 mm (1 9/16 in) across are known from the vicinity of Carlsbad, New Mexico, USA. Nice crystals come also from Stassfurt and Alexanderhall, Germany. Massive aggregates are common in many salt deposits, like in Saskatchewan, Canada; in Kalush, Ukraine and elsewhere.
Application: : the most important potassium salt, used as fertilizer and for production of metal Mg.

Atacamite

$Cu_2Cl(OH)_3$

ORTHORHOMBIC ● ● ●

Properties: C – emerald-green, black-green; S – green; L – vitreous; D – translucent; DE – 3.8; H – 3-3.5; CL – perfect; F – conchoidal; M – prismatic crystals, columnar, radial and lamellar aggregates, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of Cu deposits in the arid climate, associated with other Cu minerals. Crystals up to 10 mm (3/8 in) long were described from Burra district, Southern Australia, Australia. Crystals up to 10 mm (3/8 in) across come from Tsumeb, Namibia; also from Bisbee, Arizona, USA. Rich aggregates of acicular crystals occur in many localities in Atacama province, Chile (Copiapo, Remolinos).

Carnallite. 80 mm. Merkers. Germany

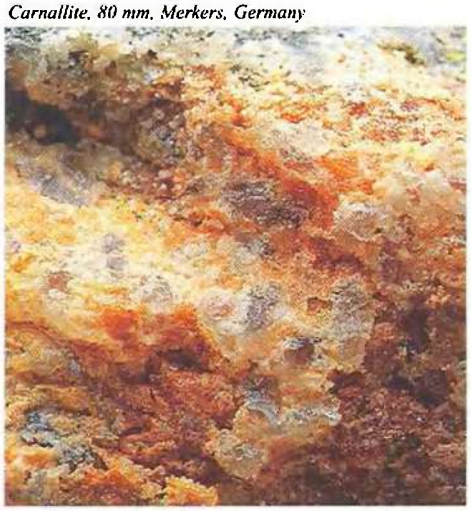

Atacamite, 78 mm, Atacama, Chile

Boleite

$Pb_{26}Ag_9Cu_{24}Cl_{62}(OH)_{48}$

CUBIC ●

Properties: C – blue; S – blue; L – pearly; D – translucent; DE – 5.1; H – 3-3.5; CL – perfect; M – cubic crystals; R – soluble in water.

Origin and occurrence: Secondary, originated in the oxidation zone of Cu deposits in the arid climate. By far the best specimens were found in Boleo, Baja California, Mexico, where cubes up to 25 mm (1 in) in size were found. It is also known from Phillipsburg, Montana, USA and Challacollo, Chile.

Boleite, 65 mm, Santa Rosalia, Mexico

Cumengite, 21 mm, Boleo, Mexico

Cumengite

$Pb_{21}Cu_{20}Cl_{42}(OH)_{40}$

TETRAGONAL ●

Properties: C – indigo-blue; S – blue; L – vitreous; D – translucent; DE – 4.7; H – 2.5; CL – good; M – tetragonal pyramidal crystals, also epitaxially overgrown on boleite cubes. *Origin and occurrence:* Secondary in the oxidation zone of Cu deposits in the arid climate, associated with boleite. The largest crystals up to 35 mm (1⅜ in) are known from Boleo, Baja California, Mexico. It is also reported from Newport Beach near Falmouth, Cornwall, UK.

Iodargyrite

β -AgI

HEXAGONAL ● ● ●

Properties: C – colorless, tarnishes to yellow; S – yellow; L – adamantine; D – transparent to translucent; DE – 5.7; H – 1-1.5; CL – perfect; F – conchoidal; M – prismatic to tabular crystals, granular, massive.
Origin and occurrence: Secondary, product by oxidation of Ag ores, with other Ag minerals. Common greenish crystals over 10 mm (⅜ in) in size occur in Broken Hill, New South Wales, Australia. Also found in Vrancice, Czech Republic; Tonopah, Nevada, USA; Chanarcillo and Copiapo, Chile.

Villiaumite

NaF

CUBIC ● ● ●

Properties: C – dark red; S – white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 2-2.5; CL – perfect; M – small crystals, granular, massive; R – soluble in water.
Origin and occurrence: Late mineral in cavities in alkaline igneous rocks (nepheline syenites). Crystals, several cm long, are known from the Rasvumchorr Mountain, Khibiny massif, Kola Peninsula, Russia; only slightly smaller crystals come from Mont St.-Hilaire, Quebec, Canada and Illimaussaq, Greenland.

Iodargyrite, 2 mm xx, Rudabánya, Hungary

Villiaumite, 10 mm xx, Khibiny Massif, Kola, Russia

Halite, 55 mm, Trona, U.S.A.

Halite

NaCl

CUBIC ● ● ● ●

Properties: C – colorless, gray, white, red, blue; S – white; L – vitreous; DE – 2.2; H – 2; CL – perfect; F – conchoidal; M – cubic crystals, granular, massive; R – soluble in water.

Origin and occurrence: Product of high-temperature fumaroles (Etna, Mt. Vesuvius; Italy); mainly sedimentary, as a result of evaporation of sea water, associated with sylvite, carnallite and other minerals. Very fine cubes over 10 mm (¾ in) are known from Weliczka and Bochnia, Poland. Blue cleavable aggregates are found in Bernburg, Germany. Salt deposits in Austria (Hallstatt, Hallein) are also important. Huge halite deposits, associated with potassium salts are mined in the vicinity of Stassfurt, Germany. Fine skeletal crystals are known from many localities in California, USA.

Application: food and chemical industries.

Halite, 18 mm xx, Sonora, Mexico

Sylvite, 77 mm. California, U.S.A.

Sylvite
KCl

CUBIC ● ● ●

Properties: C – white, gray, blue, red; S – white; L – vitreous; D – transparent; DE – 2; H – 2; CL – perfect; F – uneven; M – cubic crystals and their combinations; granular, massive; R – soluble in water.
Origin and occurrence: Sedimentary as a result of evaporation of sea water, together with halite, carnallite and other minerals. Nice cubes up to 50 mm (2 in) come from Stassfurt; it forms stalactites in Wathlingen, Germany. Crystals are also known from Kalush, Ukraine and from Salton Sea, California, USA, where it occurs as octahedra on halite crystals.
Application: chemical industry.

Sal ammoniac, 10 mm x, Batonyterenye, Hungary

Sal ammoniac
NH_4Cl

CUBIC ● ● ●

Properties: C – colorless, white, gray, yellow, brown; S – white; L – vitreous; D – transparent; DE – 1.5; H – 1-2; CL – imperfect; F – conchoidal; M – combinations of cubic crystals, dendritic and skeletal aggregates; earthy.
Origin and occurrence: Typical mineral for fumaroles and burning coal dumps, associated with sulfur and other minerals. Complicated crystals are known from Mt. Vesuvius, Etna and Vulcano, Italy. Crystals over 10 mm (3/8 in) in size occurred on burning coal dumps near Kladno, Czech Republic; similar from localities in Eastern Pennsylvania, USA and near Ste-Etienne, France.

Calomel, 30 mm, Terlingua, U.S.A.

Calomel
HgCl

TETRAGONAL ● ●

Properties: C – colorless, white, gray, brown, it darkens on air; S – white; L – adamantine; D – transparent to translucent; DE – 7.2; H – 1.5; CL – good; F – conchoidal; M – tabular to pyramidal crystals, coatings, earthy; LU – dark red.
Origin and occurrence: Secondary as a result of alteration of Hg minerals, associated with cinnabar, mercury and other minerals. Crystals were found in Moschellandsberg, Germany; Avala, Serbia;

Khaidarkan, Kyrgyzstan and Terlingua, New Mexico, USA.

Fluorite

CaF_2

CUBIC ● ● ● ● ●

Properties: C – colorless, white, yellow, red, green, blue, purple, brown, black; S – colorless; L – vitreous; D – transparent to translucent, opaque; DE – 3.2; H – 4; CL – perfect; F – conchoidal to splintery; M – combinations of cubic crystals, granular, massive; LU – blue, blue-green, also phosphorescent.

Origin and occurrence: Rare magmatic, mainly hydrothermal and metasomatic. Associations are very diverse, depending on a type of the deposit, in which it occurs. Beautiful crystals are known from many localities all over the world. Pink octahedra, several cm in size, are known from pegmatites in Nagar, Pakistan. Nice crystals were also found in greisens in Cornwall, UK (Wheal Mary mine) and from Horni Slavkov, Czech Republic. Beautiful green cubes up to 20 cm across and colorless cubes up to 10 mm (3/8 in) across from Dalnegorsk, Russia are of hydrothermal origin. Famous green and purple crystals come from Alston Moor and Weardale, England, UK. Nice pink octahedra up to 30 mm (1 3/16 in) occurred in Huanzala, Peru. Beautiful yellow cubes up to 50 mm (2 in) associated with barite, are known from Halsbrücke and Annaberg, Germany. Purple complex combinations of crystals come from La Collada, Spain. Mainly purple cubes up to 10 mm (3/8 in) occurred in Rosiclare, Illinois; in association with honey-yellow calcite crystals are known from the Elmwood mine, Tennessee; similar occurrences are also in several localities in Kentucky, USA. The most valuable fluorite specimens are pink octahedra up to 150 mm (6 in) from Goschenen, Switzerland; Mont Blanc massif, France and other Alpine localities.

Application: metallurgy, chemical industry, special optics.optika.

Fluorite, 50 mm, Argentina

Fluorite, 23 mm x, Chamonix, France

4. Oxides

Cuprite
Cu_2O

CUBIC ● ● ●

Varieties: chalcotrichite (acicular to hair-like crystals)
Properties: C – red; S – red; L – adamantine to submetallic; D – transparent to translucent; DE – 5.8-6.2; H – 3.5-4; CL – imperfect; F – conchoidal to uneven; M – combinations of cubic crystals, hair-like aggregates, granular, massive.
Origin and occurrence: Secondary, as a result of the oxidation of Cu sulfides. Crystals up to 150 mm (6 in) in size, covered with malachite, occur in Onganja, Namibia. Shiny octahedra, up to 40 mm (1 9/16 in) in size, come from the Mashamba West mine, Zaire. Acicular and fibrous crystals of the chalcotrichite variety were found in Bisbee, Arizona, USA. Combinations of cubic crystals, covered with malachite, are known from Chessy near Lyon, France. Crystals up to 30 mm (1 3/16 in) across are reported from Tsumeb, Namibia.
Application: important Cu ore.

Zincite
ZnO

HEXAGONAL ● ●

Properties: C – yellow, orange, red; S – orange-yellow; L – adamantine; D – transparent to translucent; DE – 5.7; H – 4.5-5; CL – perfect; F – conchoidal; M – pyramidal crystals, granular, massive.
Origin and occurrence: Metamorphic, associated with willemite and franklinite. It forms very rare crystals up to 40 mm (1 9/16 in) in size in the metamorphosed Zn deposits in Franklin and Sterling Hill, New Jersey, USA, it is mostly granular and massive. Zincite crystals and aggregates of vitreous luster from Poland, which are offered at the mineral shows, are not of a natural origin, there are smelter products.

Zincite, 11 mm x, Franklin, U.S.A.

Amethyst, 122 mm, Guerrero, Mexico
Cuprite, 18 mm x, Mashamba West, Zair

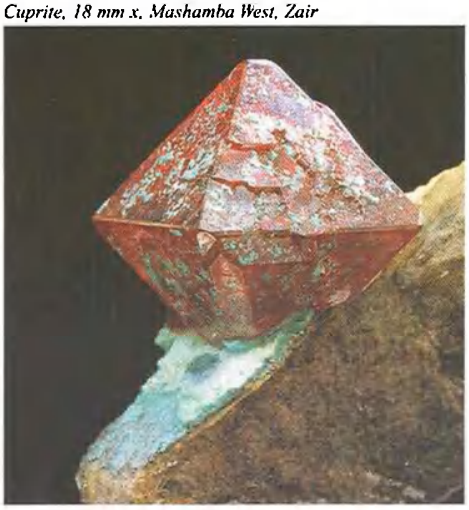

Chalcotrichite, 50 mm, Bisbee, U.S.A.

Tenorite
CuO

MONOCLINIC ● ● ●

Properties: C – steel-gray to black; S – gray; L – metallic; D – opaque; DE – 6.5; H – 3.5; CL – imperfect; F – uneven to conchoidal; M – thin tabular to scaly crystals, earthy, massive.
Origin and occurrence: Secondary, in the oxidation zone of Cu deposits, together with other Cu supergene minerals. It was common in Cu deposits in the Keweenaw Peninsula, Michigan, USA. It was mined as Cu ore in Bisbee, Globe and Morenci, Arizona, USA. Thin tabular crystals are known from Tsumeb, Namibia.

Spinel
$MgAl_2O_4$

CUBIC ● ● ●

Varieties: pleonast (black)
Properties: C – pink, red, green, blue, brown, black; S – white; L – vitreous to dull; D – transparent to opaque; DE – 3.6; H – 7.5-8; CL – imperfect; F – conchoidal to uneven; M – octahedral crystals, granular, massive.
Origin and occurrence: Magmatic, metamorphic, also in placers, associated with corundum, sillimanite and other minerals. Large pleonast crystals reaching up to 150 mm (6 in) were found in the Aldan massif, Yakutia, Russia. Fine crystals weighing up to 14 kg (30 lb 12 oz) come from Amity, New York. Crystals up to 120 mm (4 7/16 in) across are known from Sterling Hill, New Jersey. Blue crystals of spinel come from Bolton, Massachusetts, USA and South Burgess, Ontario, Canada. Gemmy pink and red crystals up to 20 mm (25/32 in) reported near Ratnapura, Sri Lanka and in Mogok, Burma. Fine pink crystals up to 50 mm (2 in) in size occur in Kukh-i-lal, Tajikistan.
Application: gemstone.

Tenorite, 3 mm xx, Bisbee, U.S.A

Gahnite
$ZnAl_2O_4$

CUBIC ● ● ●

Properties: C – black-green, black; S – gray; L – vitreous to greasy; D – translucent to opaque; DE –

Spinel, 10 mm x, Mogok, Burma

Pleonast, 15 mm xx, Vietnam

Gahnite, 90 mm, Broken Hill, Australia

4.4-4.6; H – 7.5-8; CL – imperfect; F – conchoidal to uneven; M – octahedral crystals, granular.

Origin and occurrence: Magmatic and metamorphic, associated with wolframite, chalcopyrite and other minerals. Crystals up to 120 mm (4 7/16 in) in size come from Franklin and Sterling Hill, New Jersey, USA. Crystals from Broken Hill, New South Wales, Australia reached up to 30 mm (1 3/16 in). Blue-green crystals were found in Rowe, Massachusetts, USA. Cuttable blue crystals occur near Gidan Wayo, Nigeria.

Magnetite

$Fe^{2+}Fe^{3+}{}_2O_4$

CUBIC ● ● ●

Properties: C – black; S – black; L – metallic; D – opaque; DE – 5.2; H – 5.5-6.5; CL – none; F – uneven to conchoidal; M – octahedral crystals, granular, massive.

Origin and occurrence: Magmatic, hydrothermal and metamorphic, rare sedimentary. Parageneses differ according to the origin. Fine crystals up to 170 mm (6 11/16 in) in size, come from Traversella, Italy. A crystal 25 cm (9 15/16 in) in size was found in Vastanfors, Sweden. Fine crystals up to 40 mm (1 9/16 in) occur in Dashkesan, Azerbaijan, where it is associated with andradite, epidote and apatite. Beautiful shiny octahedra up to 40 mm (1 9/16 in) are known from Alpa Lercheltini, Binntal, Switzerland. Rare cubes up to 20 mm (25/32 in) on edge come from the ZCA No.4 mine, Balmat, New York, USA.. Magnetite crystals reaching up to 10 cm were found in pegmatites in Jaguaraçu, Minas Gerais, Brazil. Crystals up to 20 cm (7 7/8 in) in size reported from the Gardiner complex, Greenland. Crystals up to 50 mm (2 in) were lately found in Kovdor, Kola Peninsula, Russia.

Application: Fe ore.

Magnetite, 20 mm x, Chester, U.S.A.

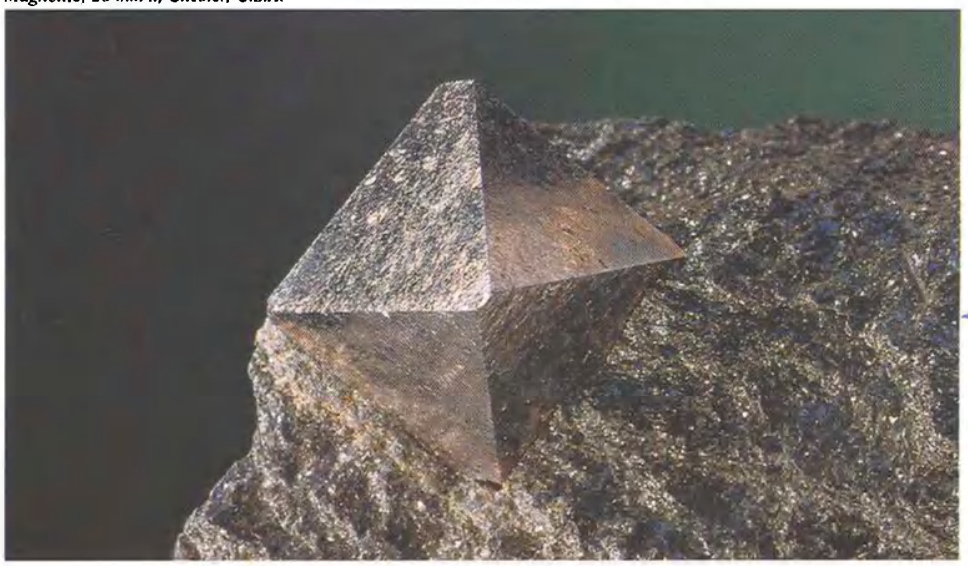

Franklinite, 30 mm, Sterling Hill, U.S.A.

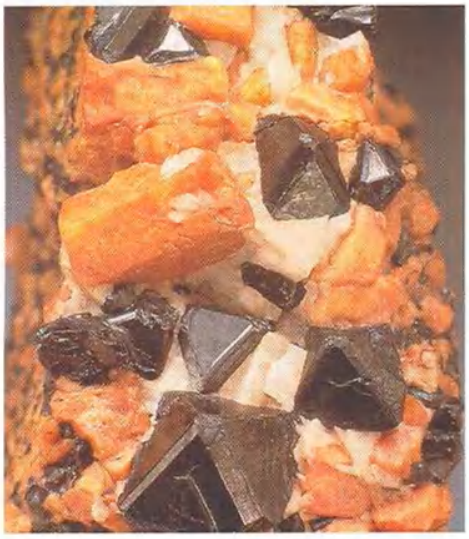

Franklinite
$(Zn,Mn^{2+},Fe^{2+})(Fe^{3+},Mn^{2+})_2O_4$

CUBIC ● ●

Properties: C – black; S – dark brown; L – metallic; D – opaque; DE – 5.1-5.5; H – 6; CL – none; F – uneven to conchoidal; M – octahedral crystals, granular, massive.
Origin and occurrence: Metamorphic, associated with willemite, zincite and other minerals. The only localities, where it is common and occurs in very large accumulations, are Franklin and Sterling Hill, New Jersey, USA. Crystals up to 170 mm (6 11/16 in) across are known from there. It is rare in Langban, Sweden and Ocna de Fier, Romania.
Application: Zn ore.

Chromite
$FeCr_2O_4$

CUBIC ● ● ●

Properties: C – black; S – brown; L – metallic; D – opaque; DE – 4.5-4.8; H – 5; CL – none; F – uneven; M – octahedral crystals, granular, massive.
Origin and occurrence: Magmatic, together with magnetite, uvarovite and other minerals. Rare crystals reaching up to 10 mm (3/8 in) are known from Uzun Damar, Turkey. It occurs mostly massive, like in deposits in Bushveld, South Africa; in Sarany, Ural Mountains, Russia and in Guleman, Turkey.
Application: Cr ore.

Hausmannite
$Mn^{2+}Mn^{3+}{}_2O_4$

TETRAGONAL ● ● ●

Properties: C – black; S – brown; L – submetallic; D – opaque; DE – 4.8; H – 5.5; CL – perfect; F – uneven; M – pseudo-octahedral crystals, granular, massive.
Origin and occurrence: Hydrothermal in high-temperature Mn deposits, also as a product of the contact metamorphism. The best specimens with crystals up to 30 mm (1 3/16 in) in size come from the N'Chwaning mine, Kuruman, South Africa. Smaller crystals were found in Ilfeld and Ilmenau, Germany. It also occurred as fine crystals in Langban and Jakobsberg, Sweden.

Chromite, 87 mm, Finnland

Minium, 60 mm, Broken Hill, Australia

Minium

$Pb^{2+}{}_2Pb^{4+}O_4$

TETRAGONAL ● ● ●

Properties: C – red, S – orange-yellow; L – dull to greasy; D – opaque; DE – 8.9; H – 2.5; CL – perfect; M – earthy and pulverulent aggregates, massive.

Origin and occurrence: Secondary mineral, as a result of the galena oxidation. It occurs in Langban, Sweden; in Anarak, Iran; in Leadhills, Scotland, and in Broken Hill, New South Wales, Australia.

Hausmannite, 42 mm, Kuruman, South Africa

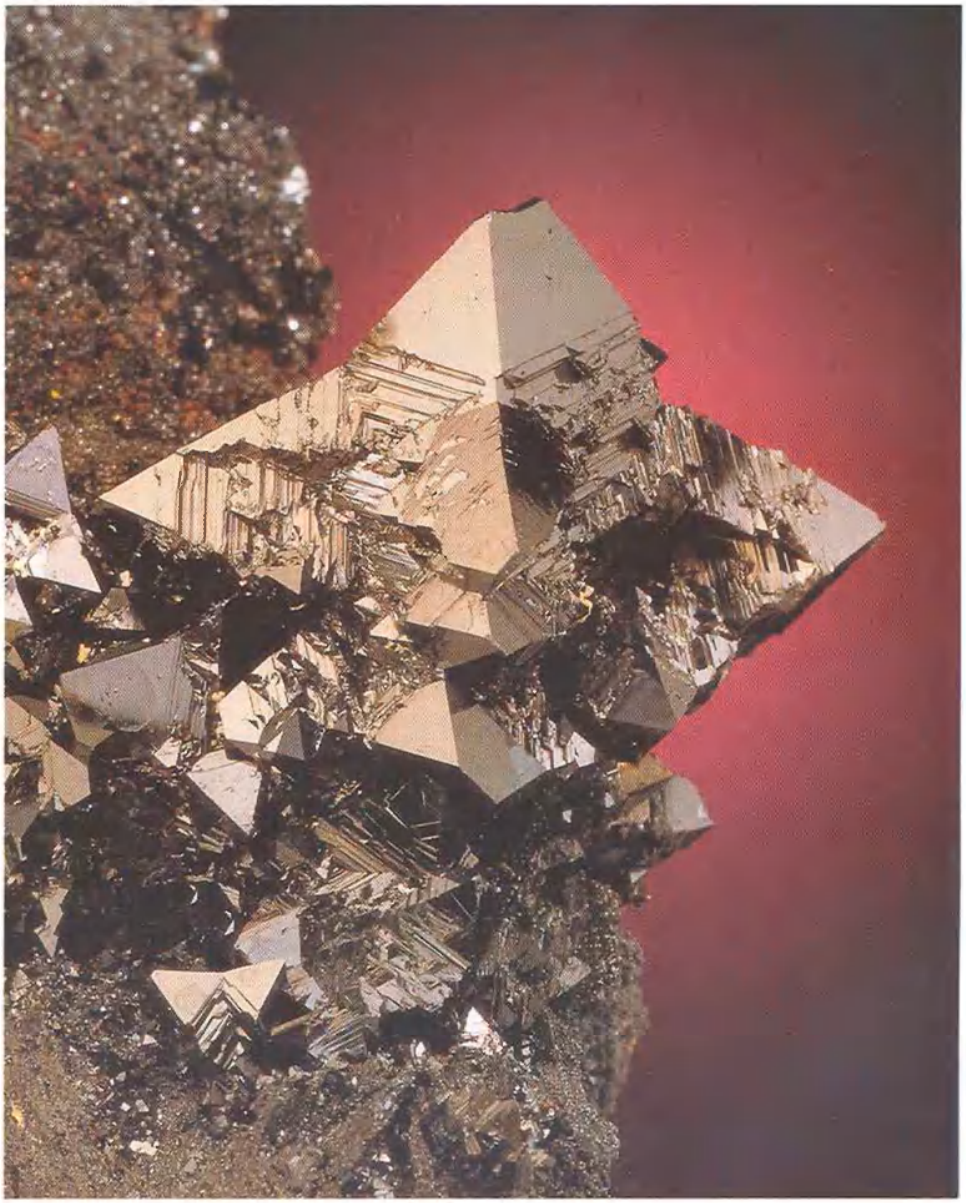

Chrysoberyl, 35 mm, Espírito Santo, Brazil

Chryzoberyl

$BeAl_2O_4$

ORTHORHOMBIC ● ● ●

Varieties: alexandrite

Properties: C – yellow-green, yellow, blue-green, alexandrite is green in daylight, purple in artificial light; S – white; L – vitreous; D – transparent to translucent; DE – 3.8; H – 8.5; CL – good; F – conchoidal to uneven; M – thin to thick tabular crystals common cyclic twins.

Origin and occurrence: Magmatic in pegmatites, prevailing as metamorphic, in association with schorl, phenakite and other minerals. Twins up to 22 cm ($8^{11}/_{16}$ in) occurred near Pancas, Espirito Santo, Brazil. Complicated twins up to 100 mm (4 in) in

Alexandrite, 75 mm, Malyshevo, Russia

size come from Ambatondrazaka and other localities in Madagascar. Tabular crystals several cm in size, embedded in the sillimanitic rock, were found in Mar_íkov, Czech Republic. Fine alexandrite crystals up to 80 mm (3 1/8 in) are known from Malyshevo, Ural mountains, Russia, together with emerald and phenakite. Alexandrite crystals reach up to 30 mm (1 7/16 in) in Nyanda, Zimbabwe. Gemmy chrysoberyls, commonly with a cat's eye effect, come from the vicinity of Ratnapura, Sri Lanka.
Application: gemstone.

Valentinite
Sb_2O_3

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, brownish; S – white; L – adamantine; D – transparent to translucent; DE – 5.7-5.8; H – 2.5-3; CL – perfect; M – prismatic to tabular crystals, radial aggregates, massive.
Origin and occurrence: Secondary mineral, originated in the oxidation of stibnite. The best specimens with crystals up to 30 mm (1 7/16 in) were found in Pribram, Czech Republic. Fine crystals come also from Braunsdorf, Germany. Crystals up to 20 mm (25/32 in) long occur in Oruro, Bolivia. Beautiful radial aggregates up to 40 mm (1 9/16 in) in diameter, associated with kermesite, are known from Pezinok and Pernek, Slovakia. Pseudo-morphs after stibnite crystals up to 35 cm (13 5/8 in)long, are reported from the Xikuangshan Mine, Lengshuijiang, China.

Arsenolite
As_2O_3

CUBIC ● ●

Properties: : C – white; S – white; L – vitreous; D – transparent to translucent: DE – 3.9; H – 1.5; CL – good; F – conchoidal; M – octahedral crystals, crusts, coatings; R – soluble in water.

Arsenolite, 3 mm xx, Recsk, Hungary

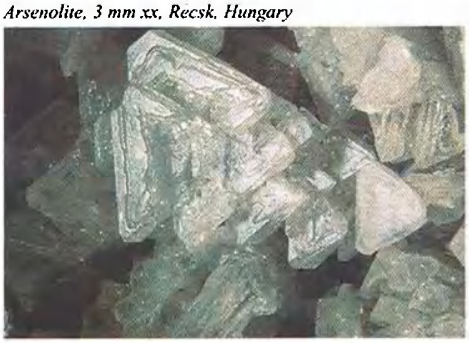

Valentinite, 40 mm, Nicolet, Canada

Origin and occurrence: Secondary mineral, resulting from the oxidation of As ores. Poorly developed crystals several mm long, occur in Jachymov, Czech Republic; in Johanngeorgenstadt and St. Andreasberg, Germany. Crystals up to 20 mm (25/32 in) long originated during a mine fire in the White Caps mine, Nevada, USA.

Senarmontite
Sb_2O_3

CUBIC ● ● ●

Properties: C – white, light gray; S – white; L – greasy, vitreous to adamantine; D – transparent to opaque; DE – 5.2-5.8; H – 2-2.5; CL – imperfect; F – uneven; M – octahedral crystals, granular, massive.
Origin and occurrence: Secondary, produced by stibnite oxidation, with valentinite and cerussite. The finest crystals up to 30 mm (1 7/16 in) in size come from Djebel Hammimate, Algeria. Also occurred in Cetine, Italy and in Dubrava, Slovakia.

Senarmontite, 3 mm xx, Pernek, Slovakia

Bixbyite, 11 mm x, Thomas Range, U.S.A.

Bixbyite
$(Mn^{3+},Fe^{3+})_2O_3$

CUBIC ● ●

Properties: C – black; S – black; L – metallic to sub-metallic; D – opaque; DE – 5; H – 6-6.5; CLB imperfect; F – conchoidal to uneven; M – cubic crystals, also twins.
Origin and occurrence: Hydrothermal in rhyolite cavities and metamorphic. Cubes up to 12 mm (15/32 in) are found together with topaz in Thomas Range, Utah, USA. Crystals up to 25 mm (1 in) occurred in the Postmasburg mine, South Africa. Crystals up to 80 mm (3 1/8 in) come from Ultevis, Sweden.

Leucosapphire, 30 mm, Sri Lanka

Corundum
Al_2O_3

TRIGONAL ● ● ● ●

Varieties: ruby, sapphire, leucosapphire, emery

Properties: C – colorless (leucosapphire), yellow, pink, red (ruby), blue (sapphire), purple, green,

Corundum, 10 mm xx, Miass, Russia

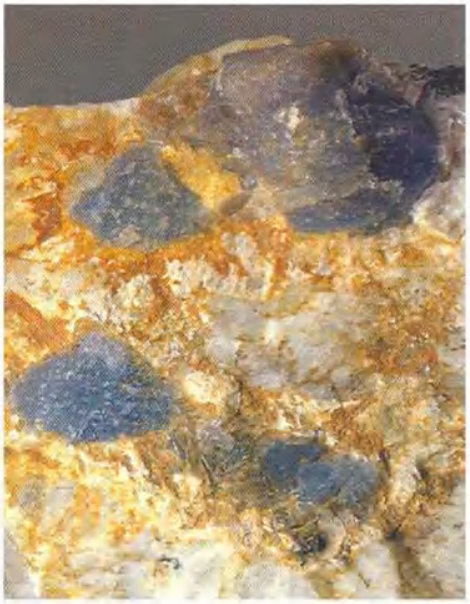

Ruby, 15 mm xx, Kola, Russia

Corundum, 4 mm, Montana, U.S.A.

Ruby, 16 mm x, Jegdalek, Afghanistan

gray; S – white; L – vitreous to adamantine; D – transparent to opaque; DE – 4.0-4.1; H – 9; CL – none; F – conchoidal to uneven; M – long prismatic to barrel-like crystals, pebbles; LU – rare dark red.

Origin and occurrence: Magmatic in andesites, pegmatites and basalts, metamorphic and in placers, in association with andalusite, topaz, spinel and other minerals. Crystals of common corundum weighed up to 30 kg (66 lb) near Bancroft, Ontario, Canada. A crystal, weighing 151 kg (333 lb 3 oz) was also found in the Letaba district, South Africa. Sapphire crystals weighing up to 20 kg (44 lb), come from the vicinity of Ratnapura and Rakwana, Sri Lanka. Fine sapphire crystals are also known from Kashmir, India. Rough gem sapphire is mined from the Yogo Gulch sediments in Montana, USA and from Anikia, Queensland, Australia. Fine blue crystals up to 50 mm (2 in) long, occur near Miass, Ural mountains, Russia. Ruby is even much rarer variety of corundum. Its beautiful crystals up to 50 mm (2 in) long come from Jegdalek, Afghanistan; Mogok, Burma and from Luc Yen, Vietnam. Prismatic crystals of opaque ruby, up to 40 mm (1 9/16 in) in size were found in the Khit Island near Kola Peninsula, Russia. Ruby crystals up to 30 cm (12 in) in size embedded in green zoisite from the vicinity of Arusha, Tanzania are very decorative.

Application: emery as abrasive material, sapphire and ruby as gemstones.

Sapphire, 41 mm, Kashmir, India

Hematite, 50 mm, La Fibbia, Switzerland

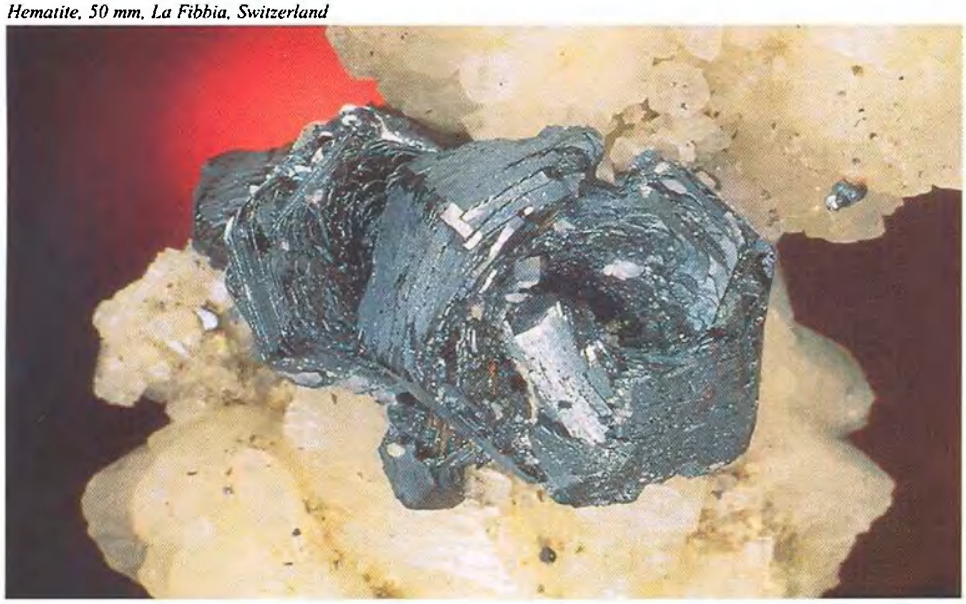

Hematite

Fe_2O_3

TRIGONAL ● ● ● ● ●

Properties: C – red, gray, black; S – red; L – metallic, dull; D – opaque; DE – 5.3; H – 6-6.5, earthy to 1; CL – none; F – uneven to conchoidal; M – thick to thin tabular crystals, massive, earthy.
Origin and occurrence: Magmatic, hydrothermal, sedimentary, also metamorphic, parageneses vary according to the origin. Beautiful crystals up to 100 mm (4 in) in size come from Brumado, BahRa, Brazil. Crystals up to 30 cm (12 in), were found in the Wessels mine, Kuruman, South Africa. So called iron roses reached up to 100 mm (4 in) near St. Gotthard, Switzerland. Fine crystals several cm in size, occurred in Rio Marina, Elba, Italy. Very fine tabular crystals reached up to 70 mm (2¾ in) in Nador, Morocco. New finds of fine crystals up to 40 mm (1 9/16 in) in size were made in the Korshunovskoye deposit, Russia. Fine botryoidal aggregates come from Hradiste and Horni Blatna, Czech Republic and from Botallack, Cornwall, UK. Sedimentary banded iron ores form huge deposits near Krivoy Rog, Ukraine or in the vicinity of Lake Superior (Mesabi Range, Minnesota; Marquette, Michigan, USA).
Application: important Fe ore.

Loparite-(Ce), 10 mm xx, Khibiny Massif, Kola, Russia

Ilmenite

$FeTiO_3$

TRIGONAL ● ● ● ●

Properties: C – black; S – black; L – metallic to dull; D – opaque; DE – 4.5-5; H – 5-6; CL – none; F – conchoidal to uneven; M – thick tabular crystals, granular, massive.
Origin and occurrence: Magmatic, metamorphic and in placers, associated with pyrrhotite, rutile, magnetite and other minerals. Crystals weighing up

Ilmenite, 11 mm x, Binntal, Switzerland

to 30 kg (66 lb) were described from the Faraday mine near Bancroft, Ontario, Canada. Crystals up to 150 mm (6 in) in size occurred near Girardville, Quebec, Canada. Crystals also reached up to 100 mm (4 in) near Miass, Ural mountains, Russia. Crystals up to 120 mm 120 mm (4 7/16 in), were found in Arendal and Kragero, Norway. Crystal rosettes up to 10 mm (3/8 in) in size come from Maderanertal, Switzerland. It is also common in placers (Kamituga, Kivu, Zaire; Sri Lanka; Travancore, India; Madagascar etc.).

Perovskite
$CaTiO_3$

ORTHORHOMBIC ● ● ●

Properties: C – dark brown to black; S – colorless to gray; L – metallic to adamantine; D – opaque; DE – 4.0-4.3; H – 5.5-6; CL – imperfect; F – conchoidal to uneven; M – pseudo-cubic crystals, granular.
Origin and occurrence: Magmatic in basic and ultrabasic rocks, metamorphic, together with magnetite, zircon and other minerals. Fine pseudo-cubic crystals up to 40 mm (1 9/16 in) in size come from Zlatoust and Akhmatovsk, Ural mountains, Russia. It occurs as crystals up to 80 mm (3 1/8 in), associated with magnetite crystals in the Gardiner complex, Greenland. Crystals up to 40 mm (1 9/16 in) were found in Jacupiranga, Sao Paulo, Brazil. Crystals from Val Malenco, Italy, reached up to 20 mm (25/32 in). Crystals up to 20 mm (25/32 in) were lately found in Afrikanda, Kola Peninsula, Russia.

Loparite-(Ce)
$(Ce,Na,Ca)TiO_3$

ORTHORHOMBIC ●

Properties: C – black, S – dark red-brown, L – metallic; D – opaque; DE – 4.6-4.9; H – 5.5-6; M – pseudo-cubic crystals, granular; R – metamict.
Origin and occurrence: Magmatic in alkaline rocks, with lorenzenite, eudialyte and aegirine. Fine interpenetration twins up to 20 mm (25/32 in) in size come from Mount Nyorkpakhk, Kola Peninsula, Russia.

Perovskite, 60 mm, Zlatoust, Russia

Stibiconite, 230 mm, Catorce, Mexico

Stibiconite
$Sb^{3+}Sb^{5+}{}_2O_6(OH)$

CUBIC ● ● ●

Properties: C – white, creamy, light yellow, brown; S – white; L – vitreous, greasy to dull; D – opaque; DE – 4.1-5.8; H – 3-6; M – pseudo-morphs after stibnite crystals, earthy, massive.
Origin and occurrence: Secondary, as a result of the stibnite oxidation, associated with valentinite and other minerals. Fine pseudo-morphs after stibnite crystals up to 30 cm (12 in) long come from Catorce, San Luis Potosí, Mexico. Similar pseudo-morphs were also found in Kostainik, Serbia; in the Ichinokawa mine, Japan and in Pereta, Italy. Pseudo-morphs after stibnite up to 20 cm (7⅞ in) long occur also in Çukuroren, Turkey.

Bindheimite
$Pb_2Sb_2O_6(O,OH)$

CUBIC ● ● ●

Properties: C – yellow, brown, gray; S – yellow, L – resinous, dull to earthy; D – translucent to opaque; DE – 4.6-5.6; H – 4-4.5; F – conchoidal to earthy; M – botryoidal, nodular and earthy crusts.
Origin and occurrence: Secondary in the oxidation zone of Pb-Sb deposits. Needles up to 10 mm (⅜ in) long come from Rudník, Czech Republic. It is common in Broken Hill, New South Wales, Australia; in Bisbee, Arizona, USA and in Sidi-Amor-ben-Salem, Tunisia. Lamellar pseudo-morphs up to several cm in size are known from Tsumeb, Namibia.

Pyrochlore
$(Na,Ca)_2Nb_2O_6(OH,F)$

CUBIC ● ● ●

Properties: C – yellow-brown, brown, black; S – brown; L – vitreous to greasy; D – translucent to opaque, DE – 4.5; H – 5-5.5; CL – locally good; F – conchoidal to uneven; M – octahedral crystals, granular; R – radioactive (admixtures of U, Th).
Origin and occurrence: Magmatic in alkaline rocks, together with zircon, astrophyllite and other minerals. Fine brown shiny crystals up to 20 mm (25/32 in) in size come from the vicinity of Vishnevogorsk, Ural mountains, Russia. Crystals reaching 10 mm (⅜ in) occur in the Panda Hill deposit, Tanzania. Crystals are also known from Oka, Quebec, Canada. Single octahedra measuring 5 mm (3/16 in) were found in Luesha, Kivu, Zaire. *Application:* Nb,U and Th ore.

Bindheimite, 4 mm, Rudabánya, Hungary

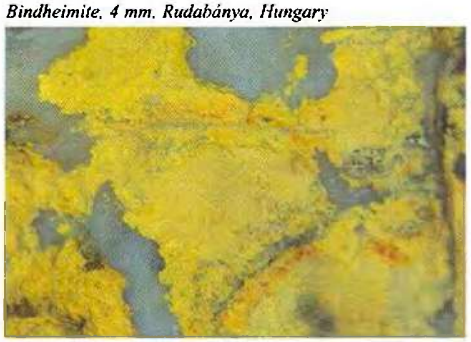

Microlite, 7 mm x, Gillette Quarry, U.S.A.

Pyrochlore, 58 mm, Vishnevogorsk, Ural Mts., Russia

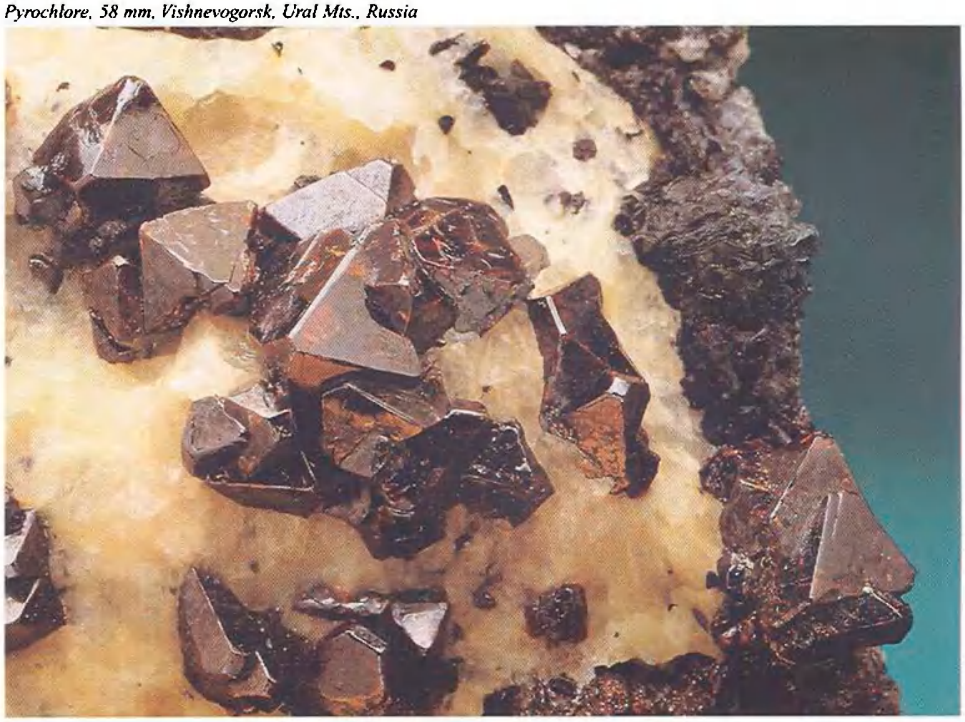

Betafite

$(Ca,Na,U)_2(Ti,Nb,Ta)_2O_6(OH)$

CUBIC ● ●

Properties: C – black, brown, yellow-brown; SB red-brown; L – resinous to greasy; D – translucent to opaque; DE – 4.2; H – 3-5.5; CL – none; F – conchoidal to uneven; M – octahedral crystals; R – radioactive, metamict.
Origin and occurrence: Magmatic in granitic pegmatites, rich in U, Th and rare earth elements, associated with beryl, euxenite-(Y) and other minerals.

Betafite, 20 mm, Silver Crater, Canada

The world's best specimens come from many localities in Madagascar (Betafo, Ambatofotsikely etc.), where crystals up to 6 kg (13 lb 3 oz) were found. Beautiful specimens with crystals up to 100 mm (4 in) in size occur in the Silver Crater mine near Bancroft, Ontario, Canada. It is also known from Evje, Norway.

Mikrolite

$(Na,Ca)_2Ta_2O_6(O,OH,F)$

CUBIC ● ● ●

Properties: C – brown, yellow, green, reddish; S – white; L – vitreous to greasy, locally adamantine; D – translucent to opaque; DE – 5-6.4; H – 6-6.5; CL – locally good; F – conchoidal to uneven; M – octahedral crystals, granular, massive.
Origin and occurrence: Magmatic, typical for granitic pegmatites, together with manganocolumbite, manganotantalite and other minerals. Octahedra up to 65 mm ($2^{9}/_{16}$ in) in size occur in Ankola, Uganda. Crystals up to 30 mm ($1^{3}/_{16}$ in) in size come from Virgem da Lapa, Minas Gerais, Brazil. Crystals up to 75 mm (3 in) are reported from the Harding pegmatite, New Mexico, USA. It occurs in important accumulations near Wodgina, Western Australia.

Quartz, 60 mm, Herkimer, U.S.A.

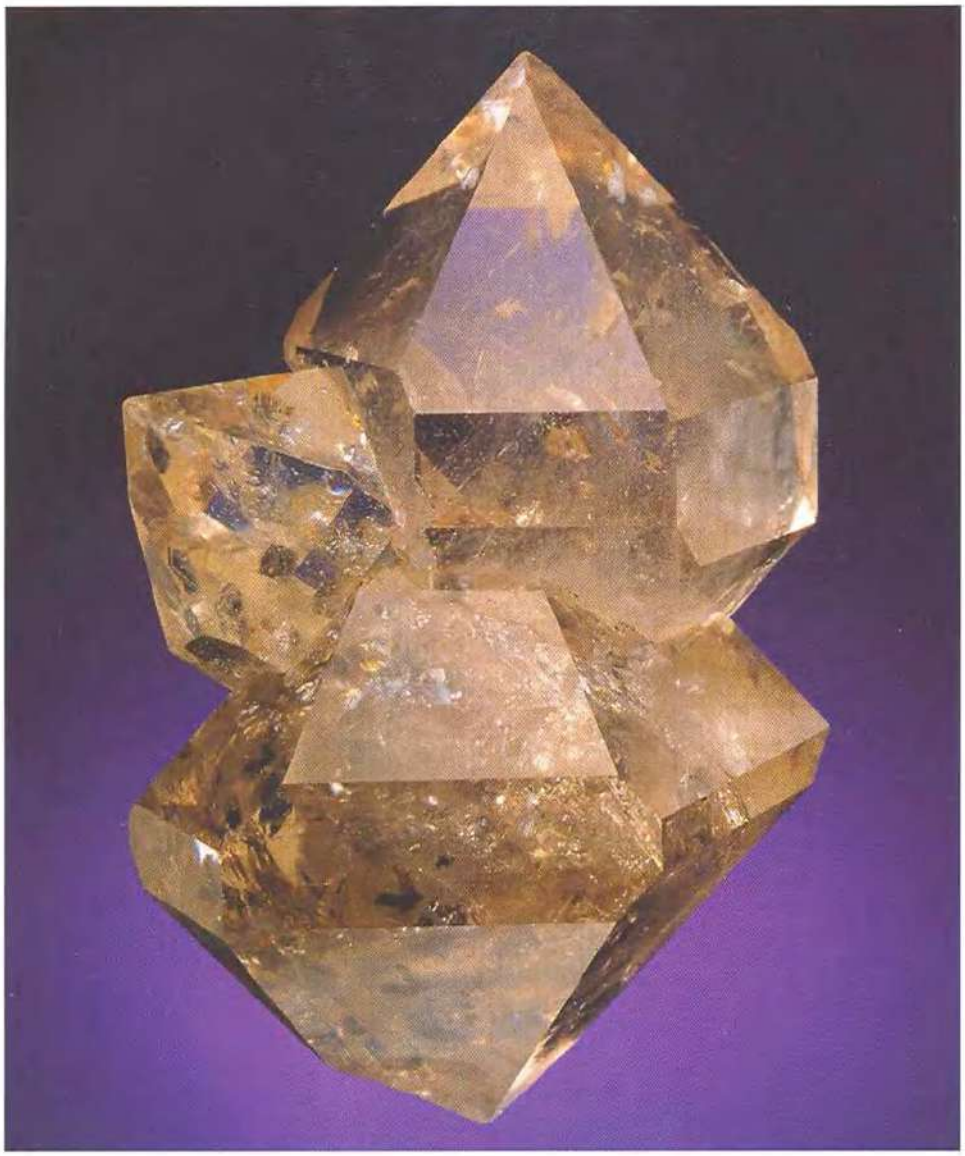

Quartz
SiO_2

TRIGONAL ● ● ● ● ●

Varieties: rock crystal, citrine, smoky citrine, morion, amethyst, rose quartz, chrysoprase, jasper, chalcedony, agate, onyx, sardonyx, aventurine, heliotrope, tiger's eye, falcon s eye.

Properties: C – colorless (rock crystal), white, yellow (citrine), brown (smoky citrine), black (morion), purple (amethyst), pink (rose quartz), green (chrysoprase); D – these varieties are mostly transparent, often translucent; C – other varieties are mainly multicolored, separate colors have different hues and the color is commonly caused by microscopic admixtures of other minerals; varieties: red, green, brown, yellow (jasper), banded with different colors (agate); white

Rock crystal, 95 mm, La Gardette, France

and black bands (onyx), white and red-brown bands (sardonyx), green to red-brown with mica or hematite inclusions (aventurine), dark green with red spots (heliotrope), yellow-brown to black-brown, fibrous

Smoky quartz, 70 mm, Middle Moat Mt., U.S.A.

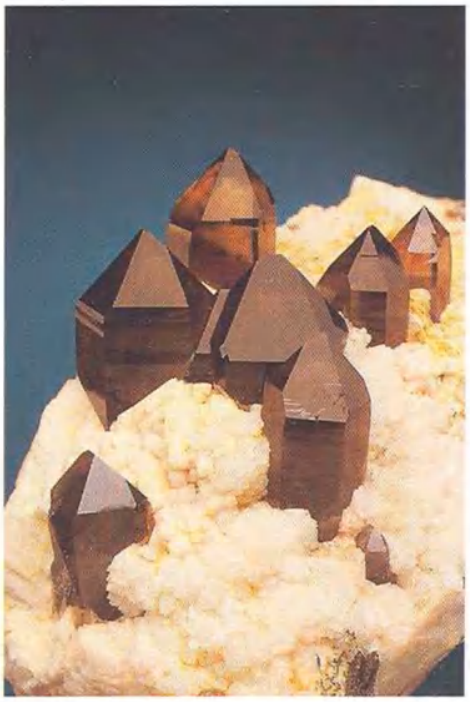

Citrine, 32 mm, Charcas, Mexico

with silky luster (tiger's eye), blue-gray to yellow-brown, fibrous with silky luster (falcon's eye); S – white; L – vitreous, silky, dull; D – transparent to translucent, opaque; DE – 2.6; H – 7; CL – none; F –

Smoky quartz, 81 mm, Switzerland

Morion, 100 mm, Agadir, Kazakhstan

Amethyst, 100 mm, Bochovice, Czech Republic

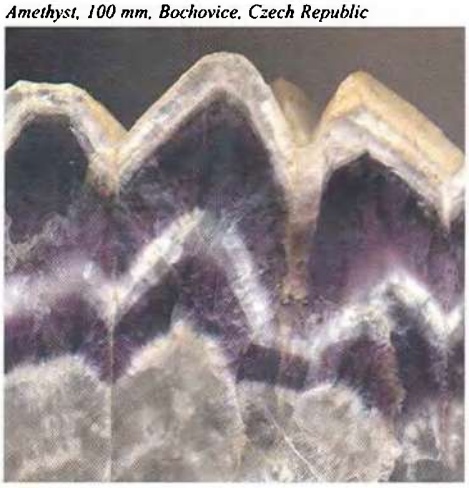

conchoidal; M – long to short prismatic, acicular, dipyramidal to tabular crystals, fibrous, botryoidal and stalactitic aggregates and coatings, concretions, geodes, granular, massive.

Origin and occurrence: Magmatic in different types of rocks, mainly in granites, granitic pegmatites and volcanic rocks; metamorphic in different types of rocks, mainly in quartzites and mica schists; hydrothermal in different types of ore and Alpine-

Rose quartz, 145 mm, Minas Gerais, Brazil

Amethyst, 200 mm, Guerrero, Mexico

type veins; secondary in the oxidation zone of ore deposits; also in different types of sedimentary rocks and in organic remains, also in placers. Probably the most common mineral in the Earth's crust and the most important rock-forming mineral, as well.

Large crystals of rock quartz up to 7 m (23 ft) long come from pegmatites in the Betafo region in Madagascar and from the Alpine-type veins, like Uri, Grimsel and Furka, Switzerland; perfect crystals are known from the cracks in marbles near Carrara, Italy; it also occurs in the quartz veins in Herkimer, New York and Hot Springs, Arkansas, USA. Citrine occurs mainly in granitic pegmatites and large crystals come from Goias, Brazil; from Suky and Netín, Czech Republic; from Murzinka, Ural mountains, Russia. Smoky quartz originates mostly in granitic pegmatites, it also occurs in the

Chalcedony, 68 mm, High Atlas, Morocco

Carnelian, 40 mm, Nova Paka, Czech Republic

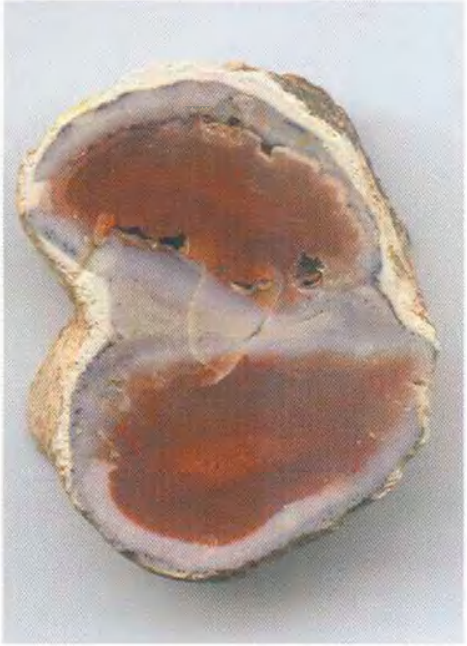

Chrysoprase, 60 mm, Szklary, Poland

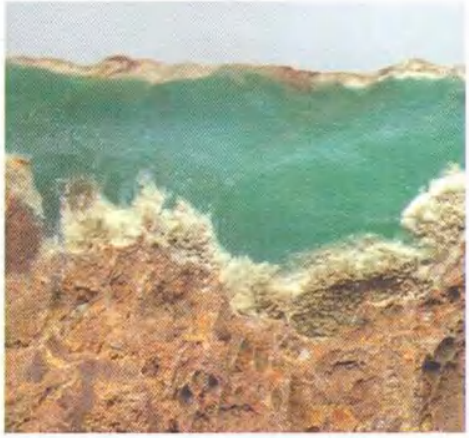

Alpine-type quartz veins and in cavities of volcanic rocks. Perfect crystals up to several meters long, come from many places, the largest crystal, weighing 77 tons , was found in Kazakhstan; perfect crystals occur in pegmatites in many places in Brazil; also in Korostenskiy massif, Ukraine; in the Pikes Peak batholith, Colorado, USA; in the Alpine-type veins in Maderanertal and in Grimsel, Switzerland. Morion crystals, commonly associated with smoky citrine, were found in quartz veins and in pegmatites.

Its crystals are known from St. Gotthard, Switzerland. Amethyst comes from quartz and ore veins, cavities in volcanic rocks, rare in the Alpine-type veins. Famous localities in volcanic rocks are in the states of Rio Grande do Sul and Minas Gerais, Brazil, doubly terminated crystal, weighing 5.5 tons, come from Diamantina. In Serra do Mar, Rio Grande do Sul, a cavity covered with amethyst

Jasper, 50 mm, Ural Mts., Russia

Jasper, 60 mm, Oregon, U.S.A.

Petrified wood, 70 mm, Podkrkonosi, Czech Republic

crystals measuring 10 x 2 x 1 m (33 x 6 x 3 ft 3in) was found; rich druses occur also in the ore veins in Porcura, Romania and Julimes, Mexico. Rose quartz, forming masses up to several meters in granitic pegmatites in the Rose Quartz pit, Quadeville, Ontario, Canada; in Ambositra, Madagascar; crystals up to 10 mm (2/5 in) long growing on quartz crystals, come from Sapucaia, Minas Gerais, Brazil. Dark green chrysoprase veins up to 50 mm (2 in) thick are known from serpentinites in Szklary,

Iron stained quartz, 95 mm, Horovice, Czech Republic

Agate, 50 mm, Železnice, Czech Republic

Onyx, 65 mm, Brazil

Poland. Jasper is known from volcanic rocks and their contacts with sediments, locally as a result of petrification of organic matter, mainly plants, it is also known from quartz veins. Rich aggregates occur in Idar-Oberstein, Germany; in Podkrkonosi region and in Krusne hory mountains, Czech Republic; in the Petrified Forest National Park, Holbrook, Arizona, USA; in Ural mountains, Russia; in Kabamby, Madagascar. Chalcedony is mostly found in quartz veins and geodes in volcanic rocks, also in sediments.

Rich aggregates come from Idar-Oberstein, Germany; Julimes, Mexico: many localities in Uruguay and Brazil; in Huttenberg, Austria. Agates are known from cavities in volcanic rocks, rare in hydrothermal veins and in sediments. The most important localities are located in the southern part of Brazil in the state of Rio Grande do Sul; in Uruguay; also in Yemen; India; Mongolia; in several

Moss agate, 40 mm, Krdjali, Bulgaria

Agate, 140 mm, Horni Halze, Czech Republic

Agate, 60 mm. Brazil

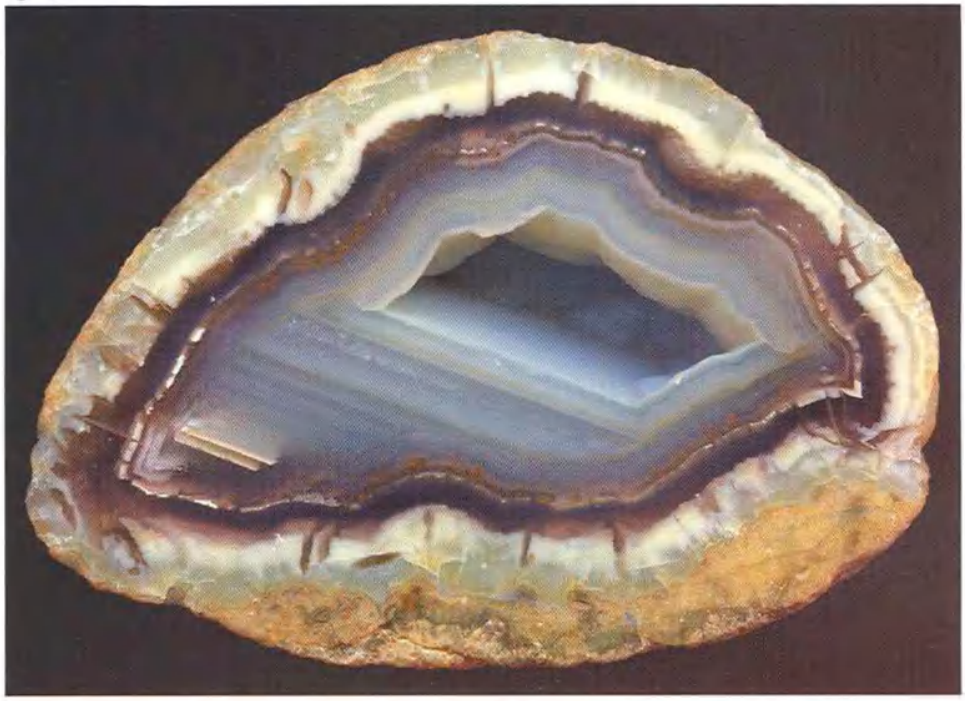

localities in the USA; in Idar-Oberstein, Germany; in Podkrkonosi, Czech Republic. The most famous localities of onyx and sardonyx are in Brazil and Uruguay.

Rich aggregates of aventurine come from Miass, Ural mountains, Russia; Mariazell, Austria; Belany, India. Heliotrope occurs in Idar-Oberstein, Germany; Kozakov, Czech Republic and in Brazil.

Application: important raw material in glass industry, many colored varieties, like amethyst, smoky citrine, citrine, onyx, sardonyx, and heliotrope are cut as gemstones.

Aventurine, 50 mm, India

Tiger's eye, 50 mm, Griqualand, South Africa

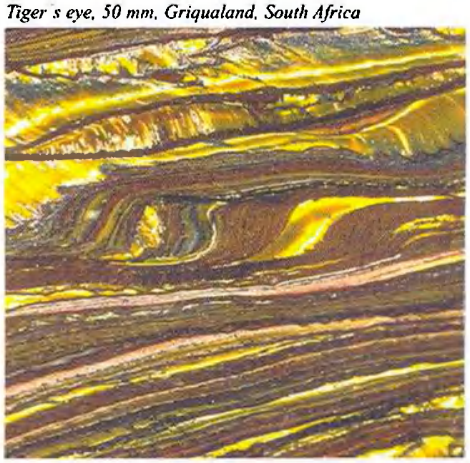

Tridymite, 9 mm, Big Luc Mt., U.S.A.

Tridymite

SiO_2

ORTHORHOMBIC ● ● ●

Properties: C – colorless to white; S – white; L – vitreous; D – transparent; DE – 2.3; H – 6.5-7; CL – none; F – conchoidal; M – pseudo-hexagonal tabular crystals.

Cristobalite, 110 mm, Vechec, Slovakia

Origin and occurrence: Magmatic in cavities of young felsic volcanic rocks in association with cristobalite, chalcedony and other minerals. Pseudo-hexagonal tabular crystals up to 10 mm (3/8 in) in size, come from Vechec, Slovakia. Similar crystals found in Ichigayama, Japan. Crystals up to 10 mm (3/8 in) occur with topaz and other minerals in the Thomas Range, Utah, USA.

Cristobalite

SiO_2

TETRAGONAL ● ● ●

Varieties: lussatite (fibrous)

Properties: C – colorless to white; S – white; L – vitreous; D – translucent; DE – 2.3; H – 6.5; CL – none; M – pseudo-octahedral crystals, spherical and botryoidal aggregates.

Origin and occurrence: Magmatic in cavities of young felsic volcanic rocks, associated with tridymite. Crystals up to 4 mm (5/32 in) are known from Cerro San Cristobal, Hidalgo, Mexico. Crystals up to 2 mm (1/16 in) long occur in Vechec, Slovakia. Gray spherical aggregates come from Coso Hot Springs, California, USA.

Common opal, 60 mm, Křemže, Czech Republic

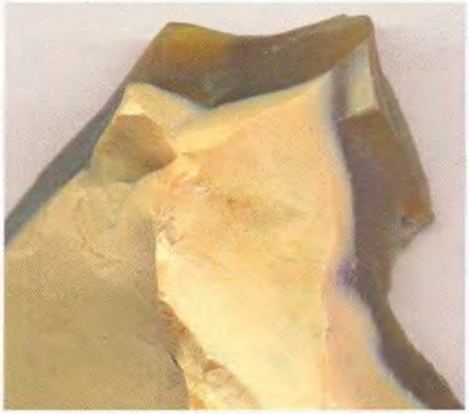

Opal, 50 mm, Herlany, Slovakia

Opal

$SiO_2 \cdot n\ H_2O$

AMORPHOUS ● ● ● ● ●

Varieties: hyalite, milky opal, fire opal, precious opal, wooden opal, geysirite.

Properties: C – colorless (hyalite), white (milky opal), red (fire opal), iridescence (precious opal), brown, red-brown, yellow, green, gray, blue; S – white; L – vitreous, dull, earthy, waxy; D – transparent to translucent, opaque; DE – 2.1; H – 5.5-6.5; CL – none; F – conchoidal; M – botryoidal and stalactitic aggregates, coatings, concretions, geodes, massive; LU – white, yellow-green, green.

Origin and occurrence: Hydrothermal in volcanic rocks and tuffs, also in various types of volcanic rocks and tuffites, in different types of sedimentary rocks, in organic remnants and hot springs, rare in

Hyalite, 45 mm, Valec, Czech Republic

Opal, 47 mm, Opal Butte, U.S.A.

Wooden opal, 40 mm, Lubietová, Slovakia

hydrothermal veins; secondary in the weathering zone of different types of rocks. It is often associated with chalcedony. Coatings and stalactitic aggregates of hyalite up to 50 mm (2 in) thick known from Cerritos, Mexico; Valec, Czech Republic; Klamath Falls, Oregon, USA. Milky opal occurs in Dubník, Slovakia; Smrcek, Czech Republic and many other localities. The most famous locality of fire opal is Zimapan, Hidalgo, Mexico. Precious opal comes from many localities in Australia, e.g. Baracoo River, Queensland; Coober Pedy, Southern Australia and White Cliffs, New South Wales, where it forms rich aggregates and veinlets in sandstones; classic locality is Dubník, Slovakia, where it occurs in

Dendritic opal, 56 mm, Křemže, Czech Republic

Fire opal, 20 mm, Mexico

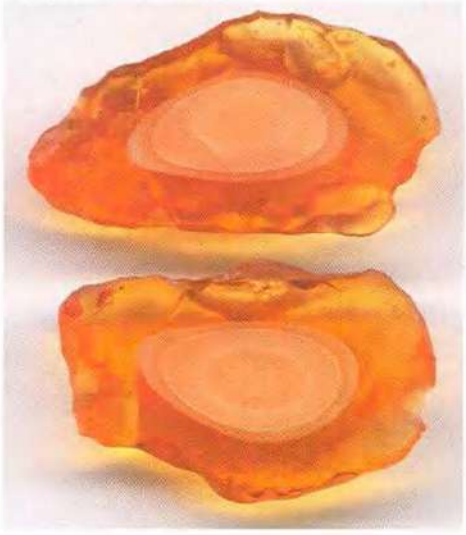

andesites and was probably mined already by ancient Romans. Beautiful precious opals come also from the Virgin Valley, Nevada, USA. Petrified trees, known as wooden opal, reach lengths of several meters in the Petrified Forest National Park, Holbrook, Arizona, USA; in L'ubietova and Povraznik, Slovakia. White geysirite is known mainly from the hot springs in Iceland; Yellowstone National Park, Wyoming, USA; New Zealand.

Application: colored opal varieties, primarily precious opal and fire opal are cut as gemstones, diatomite in chemical industry.

Precious opal, 55 mm, Opal Butte, U.S.A.

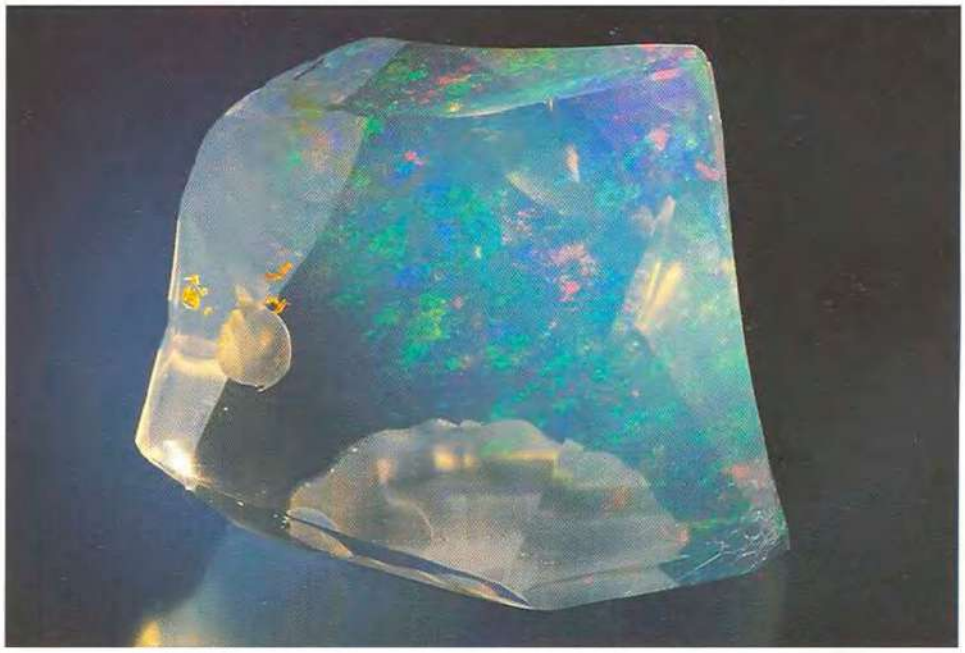

Melanophlogite, 80 mm, Fortullino, Italy

Melanophlogite

SiO_2

CUBIC ●

Properties: C – colorless, white; S – colorless; L – vitreous; D – transparent; DE – 2; H – 6.5; CL – none; M – pseudo-cubic crystals, spherical aggregates.
Origin and occurrence: Hydrothermal, associated with sulfur and other minerals. It was originally described from the sulfur deposit in Racalmuto, Sicily, Italy as crystals up to 4 mm (5/32 in) in size. It is also known from Chvaletice, Czech Republic, where it forms crystals up to 2 mm (1/16 in) in size, in Alpine-type veins.

Rutile

TiO_2

TETRAGONAL ● ● ● ●

Properties: C – red-brown, red, brown, yellowish, black; S – light brown; L – metallic to adamantine; D – transparent to translucent; DE – 4.2; H – 6-6.5; CL – good; F – conchoidal to uneven; M – short prismatic, striated crystals, common twins, acicular crystals, granular, massive.
Origin and occurrence: Magmatic and metamorphic, also in placers, together with monazite-(Ce), topaz, beryl, quartz and other minerals. The largest crystals up to 150 mm (6 in) in size come from the Mount Graves, Georgia, USA. Beautiful epitaxial intergrowths with hematite occur in Cavradischlucht, St. Gotthard, Switzerland and in Ibitiara, Bahía, Brazil. It is common as inclusions in smoky citrine (quartz) crystals from Ibitiara, Bahia and Itabira, Minas Gerais, Brazil. Knee-like crystal twins up to 70 (2¾ in) cm in size were found in the vicinity of Golcuv Jenikov and Sobeslav, Czech Republic.
Application: Ti ore.

Rutile, 21 mm, Ibitiara, Brazil

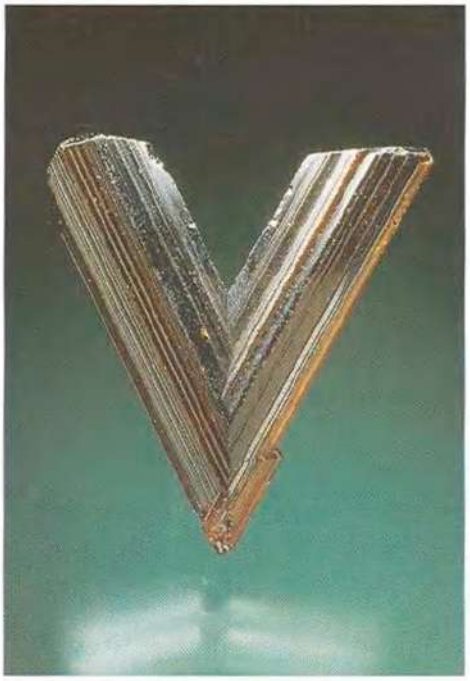

Rutile, 48 mm, Bahía, Brazil

Cassiterite, 110 mm, Yunnan, China

Cassiterite

SnO_2

TETRAGONAL ● ● ● ●

Properties: C – colorless, brown, black; S – white, grayish, brown; L – metallic to adamantine, dull; D – transparent to opaque; DE – 6.3-7.2; H – 6-7; CL – imperfect; F – conchoidal to uneven; M – dipyramidal and short prismatic crystals, multiple twins, granular, massive.
Origin and occurrence: Magmatic in pegmatites, hydrothermal in high-temperature deposits, metamorphic and in placers, together with wolframite, topaz and other minerals. Crystal twins up to 150 mm (6 in) in size come from Horni Slavkov, Czech Republic. Crystals of similar size were also found in Panasqueira, Portugal. Fine twins up to 80 mm (3 1/8 in) were found in Rossarden, Tasmania, Australia. Crystals up to 70 mm (2¾ in) in size were found in Llallagua, colorless and transparent crystals up to 50 mm (2 in) in size in Viloco, Bolivia. Crystals up to 110 mm (4 5/16 in) occurred lately in Tenkergin, Chukotka, Russia. Crystals up to 130 mm (5 7/16 in) in size are known from pegmatites in Minas Gerais, Brazil (Fazenda do Funil). New finds of shiny crystals up to 100 mm (4 in) long were made in Hunan and Yunnan provinces, China.
Application: Sn ore.

Plattnerite

PbO_2

TETRAGONAL ● ●

Properties: C – black; S – brown; L – metallic to adamantine; D – opaque; DE – 9.6; H – 5.5; CL – none; M – acicular crystals, botryoidal aggregates, massive.
Origin and occurrence: Secondary, as a result of the oxidation of other Pb minerals, together with pyromorphite, hemimorphite and other minerals. Fine crystals come from Mina Ojuela, Mapimi, Durango, Mexico and from the Blanchard mine, New Mexico, USA. Botryoidal aggregates, weighing up to 100 kg (220 lb), were found in the Morning mine, Mullan, Idaho, USA.

Plattnerite, 150 mm, Mapimi, Mexico

Pyrolusite, 88 mm, Baraga, U.S.A.

Pyrolusite
MnO_2

TETRAGONAL ● ● ● ●

Properties: C – black, steel-gray; S – black; L – metallic to dull; D – opaque; DE – 5.1; H – 6-6.5; CL – perfect; F – uneven; M – prismatic to acicular striated crystals, stalactitic and botryoidal aggregates, granular, massive.
Origin and occurrence: Secondary, as a result of alteration of manganite and other primary Mn minerals, also hydrothermal. Crystals up to 20 mm (25/32 in) long come from Horní Blatna, Czech Republic. Radial shiny aggregates were found in Ohrenstock, Germany. It occurred in Ilfeld, Germany, too. Large sedimentary Mn deposits, where pyrolusite is the main constituent, are known near Chiaturi, Georgia or near Nikopol, Ukraine. Crystals are reported also from Tsumeb, Namibia and Hotazel, South Africa.
Application: important Mn ore.

Hollandite
$Ba(Mn^{4+},Mn^{2+})_8O_{16}$

MONOCLINIC ● ● ●

Properties: C – gray-black; S – black; L – submetallic; D – opaque; DE – 5; H – 6; CL – good; M – short prismatic crystals, racemous and columnar aggregates.
Origin and occurrence: Metamorphic in Mn deposits with braunite, scheelite and other minerals, also secondary. Crystals up to 5 mm (3/16 in) long come from the Bradshaw mountains, Arizona, USA. It is common in the metamorphosed Mn deposits in Ultevis, Sweden; in Nagpur and Balaghat, India.

Coronadite
$Pb(Mn^{4+},Mn^{2+})_8O_{16}$

TETRAGONAL ● ● ●

Properties: C – black, black-gray; S – brown-black; L – submetallic to dull; D – opaque; DE – 5.4; H – 4.5-5.5; M – botryoidal crusts with fibrous structure, massive.
Origin and occurrence: Secondary in the oxidation zone of Mn deposits. Spherical aggregates up to 100 mm (4 in) in diameter come from the Bou Tazoult mine, Imini, Morocco. Small crystals were found in the Beltana mine, Southern Australia, Australia and in the Silver Bill mine, Arizona, USA.

Todorokite
$(Mn^{2+},Ca,Mg)Mn^{4+}{}_3O_7 \cdot H_2O$

MONOCLINIC ● ● ●

Properties: C – gray, brown-black, black; S – brown;

Todorokite, 60 mm, Kamogun, Japan

L – metallic to dull; D – opaque; DE – 3.5-3.8; H – 1; CL – perfect; M – platy crystals, stalactitic and nodular aggregates.
Origin and occurrence: Secondary mineral, as a result of the oxidation of Mn minerals. Crystals are reported from several mines in the Kalahari region, South Africa (Hotazel, Smart). It was originally described from the Todoroki mine, Hokkaido, Japan. It is an important constituent of occanic Mn concretions.

Ferrotapiolite
$FeTa_2O_6$

TETRAGONAL ● ● ●

Properties: C – black, brown; V – red-brown; L – submetallic, adamantine, resinous; D – opaque; DE – 7-7.8; H – 6-6.5; CL – none; F – uneven to conchoidal; M – dipyramidal and short prismatic crystals, massive.
Origin and occurrence: Magmatic in pegmatites, together with manganotantalite, microlite, cassiterite and other minerals. Crystals up to 40 mm (1 9/16 in) in size are known from the vicinity of Governador Valadares, Minas Gerais, Brazil. A crystal 120 mm (4 9/16 in) long has been described from Angarf, Morocco. Short prismatic crystals come from pegmatites near Topsham and Paris, Maine, USA.

Ferrotapiolite, 60 mm, Maršíkov, Czech Republic

Ilmenorutile
$(Ti,Nb,Fe)O_2$

TETRAGONAL ● ● ●

Properties: C – black; S – brown; L – submetallic; D – opaque; DE – 4.2; H – 6-6.5; CL – good; F – conchoidal to uneven; M – prismatic crystals, granular.
Origin and occurrence: Magmatic in pegmatites, together with schorl, zircon, fluorapatite and other minerals. Prisms several cm long come from Udraz near Pisek, Czech Republic. It is also known from the vicinity of Miass, Ural mountains, Russia and Evje, Norway.

Ilmenorutile, 10mm x, Pisek, Czech Republic

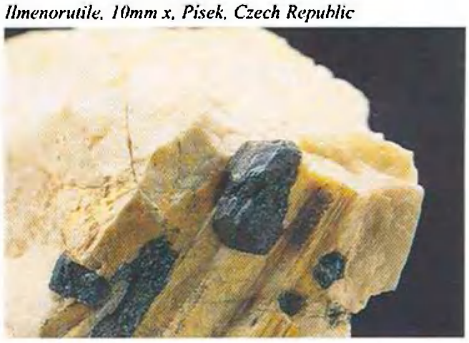

Anatase
TiO_2

TETRAGONAL ● ● ●

Properties: C – black-gray, brown, red-brown, blue, rare colorless; S – white; L – submetallic to adamantine; D – transparent to opaque; DE – 3.8-4; H – 5.5-6; CL – perfect; F – conchoidal; M – dipyramidal and tabular crystals.
Origin and occurrence: Hydrothermal in the Alpine-type veins, associated with brookite and quartz, also sedimentary and metamorphic. Beautiful crystals up to 50 mm (2 in) long were found in Alpa Lercheltini, Binntal, Switzerland. Famous black-blue crystals reaching up to 30 mm (1 3/16 in) come from Matskorhae, Hardangervidda, Norway. Crystals up to 15 mm (1 9/32 in) were recently found in Dodo, Polar Ural, Russia. Crystals up to 30 mm (1 3/16 in) in size occurred in the Old Lot and Vulcan mines, Colorado, USA.

Tellurite
TeO_2

ORTHORHOMBIC ● ●

Properties: C – yellow, yellow-orange; S – yellowish; L – adamantine; D – transparent; DE – 5.8; H – 2; CL – perfect; M – acicular and thin tabular crystals, radial aggregates, pulverulent.
Origin and occurrence: Secondary, resulting from the oxidation of AuBTe ores. Crystals up to 10 mm (3/8 in) long occurred in the Kawazu and Susaki mines, Japan. Beautiful specimens were found in Moctezuma, Sonora, Mexico; it also comes from Cripple Creek, Colorado, USA.

Brookite
TiO_2

ORTHORHOMBIC ● ● ●

Properties: C – light to dark brown, yellow-brown, black; S – white to gray; L – submetallic to adamantine; D – transparent to translucent, opaque; DE – 4.1; H – 5.5-6; CL – imperfect; F – conchoidal to uneven; M – tabular, dipyramidal, long and short prismatic crystals.
Origin and occurrence: Hydrothermal along the fissures of the Alpine-type veins and in granitic and alkaline pegmatites; it occurs as pseudo-morphs after titanite and ilmenite; also in sedimentary rocks. Perfect tabular crystals up to 50 mm (2 in) in size found in Rieder Tobel, Switzerland; Magnet Cove, Arkansas, USA; Passo di Viza, Italy. New finds of crystals up to 50 mm in size made in Dodo, Polar Ural, Russia.

Anatase, 16 mm, Hardangervidda, Norway

Brookite, 50 mm, Puiva, Polar Urals, Russia

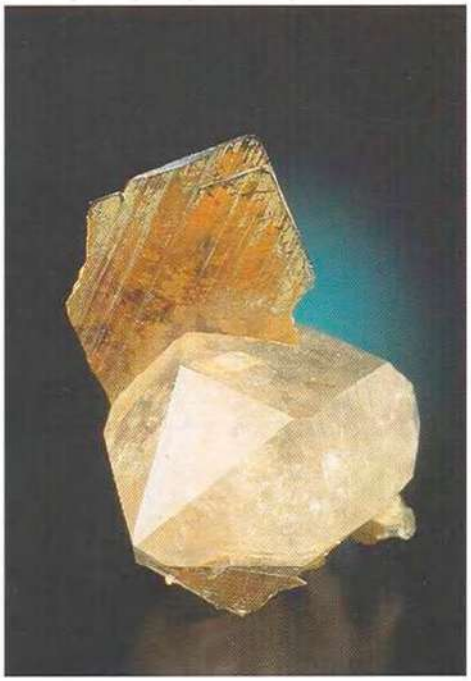

Tellurite, 1 mm x, Fata Baii, Romania

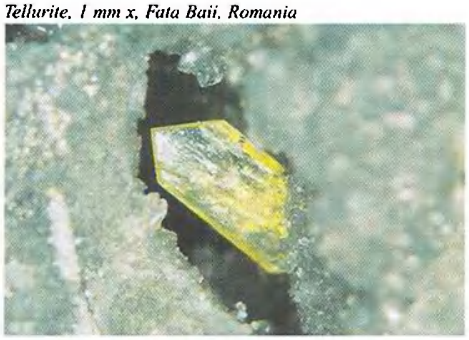

Ferberite
$FeWO_4$

MONOCLINIC ● ● ●

Properties: C – black; S – brown-black to black; L – submetallic; D – opaque; DE – 7.5; H – 4-4.5; CL – perfect; F – uneven; M – short prismatic to tabular crystals, granular, massive.
Origin and occurrence: Hydrothermal in high- to medium-temperature ore veins, in greisens and skarns; rare magmatic in granitic pegmatites and granites; it also occurs in placers. It is usually associated with cassiterite, scheelite, sulfides and quartz. Perfect tabular crystals up to 120 mm ($4^{7}/_{16}$ in) in size found in the Quartz Creek mine, Colorado, USA; also in Cínovec, Czech Republic; Panasqueira, Portugal; Ehrenfriedersdorf, Germany and Potosí, Bolivia.
Application: W ore.

Ferberite, 68 mm, Mundo Nuovo, Peru

Hübnerite
$MnWO_4$

MONOCLINIC ● ● ●

Properties: C – yellow-brown, red-brown, black; S – yellow-brown to black-gray; L – submetallic; D – translucent to opaque; DE – 7.2; H – 4-4.5; CL – perfect; F – uneven; M – short prismatic to tabular crystals, granular, massive.
Origin and occurrence: Hydrothermal in high- to medium-temperature ore veins, in greisens; rarely magmatic in granitic pegmatites; also in placers. Perfect tabular to short prismatic crystals up to 25 cm ($9^{13}/_{16}$ in) in size, come from the Huayllapon mine, Pasto Bueno, Peru; also from Baia Sprie, Romania; Kara-Oba, Kazakhstan; the Sweet Home mine, Alma and Silverton, Colorado, USA.
Application: W ore.

Hubnerite, 39 mm, Silverton, U.S.A.

Titanowodginite, 20 mm, Tanco, Canada

Titanowodginite

$MnTiTa_2O_8$

ORTHORHOMBIC ●

Properties: C – dark brown, black; S – dark brown; L – submetallic; D – translucent to opaque; DE – 6.9; H – 5.5; CL – imperfect; F – uneven; M – dipyramidal crystals, granular.
Origin and occurrence: Magmatic in granitic pegmatites. Dipyramidal crystals up to 10 mm (3/8 in) long occur in the Tanco mine, Bernic Lake, Manitoba, Canada.
Application: Ta ore.

Ferrocolumbite, 90 mm, Middletown, U.S.A.

Ferrocolumbite

$FeNb_2O_6$

ORTHORHOMBIC ● ● ●

Properties: C – black, red-brown; S – red-brown to black; L – submetallic; D – translucent to opaque; DE – 5.2; H – 6; CL – good; F – uneven to conchoidal; M – long to short prismatic and tabular crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites and granites; rare hydrothermal in high-temperature ore veins and in greisens; also in placers. Tabular crystals up to 1 m (39 3/8 in) in size occur in granitic pegmatites near Custer and Keystone, South Dakota, USA; in Malakialina, Madagascar; Ichikawa, Japan; masses, weighing up to 270 kg (594 lb) come from the Meyers quarry, Colorado, USA.
Application: Nb ore.

Manganotantalite

$MnTa_2O_6$

ORTHORHOMBIC ● ● ●

Properties: C – red, red-brown, black-brown, black; S – dark red to black; L – submetallic; D – translucent to opaque; DE – 8.0; H – 6-6.5; CL – good; F – uneven

Manganotantalite, 19 mm, Nuristan, Afghanistan

Euxenite-(Y), 40 mm, Ambatofotsy, Madagascar

to conchoidal; M – prismatic to tabular crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites and granites; also in placers. Crystals up to 100 mm (4 in) in size come from Li-bearing pegmatites in the Tanco Mine, Bernic Lake, Manitoba, Canada; also from Sao Jose da Safira, Minas Gerais, Brazil.
Application: Ta ore.

Euxenite-(Y)
$(Y,Ce,U,Th)(Nb,Ta,Ti)_2O_6$

ORTHORHOMBIC ● ●

Properties: C – black with brownish and green hues; S – gray, yellowish, brownish; L – submetallic, resinous; D – translucent to opaque; DE – 4.6; H – 6; CL – none; F – conchoidal; M – tabular crystals, granular, massive; R – locally weakly radioactive, commonly metamict.
Origin and occurrence: Magmatic in granitic and alkaline pegmatites; also in placers. Typically associated with monazite-(Ce), zircon, ilmenite and other oxides of rare earth elements. Crystals up to 150 mm (6 in) in size are known from Kragero and Hitterö, Norway; from Ankazobe, Madagascar.

Aeschynite-(Ce)
$(Ce,Ca)(Ti,Nb)_2O_6$

ORTHORHOMBIC ● ●

Properties: C – black, red-brown, yellow; S – red-yellow; L – vitreous, resinous, adamantine; D – translucent to opaque; DE – 5.0; H – 5.5; CL – none; F – conchoidal; M – prismatic and tabular crystals, granular, massive; R – locally weakly radioactive, commonly metamict.
Origin and occurrence: Magmatic in alkaline and granitic pegmatites and carbonatites. It is associated

Aeschynite-(Ce), 45 mm, Hittero, Norway

with zircon and oxides of rare earth elements. Crystals up to 190 mm (7 3/16 in) long occur in Quadeville, Ontario, Canada; other localities are Kragero, Norway; Trout Creek Pass, Colorado, USA.

Stibiotantalite
$SbTaO_4$

ORTHORHOMBIC ● ●

Properties: C – yellow, yellow-brown, red-brown, yellow-green; S – yellow-brown; L – submetallic, vitreous, resinous; D – transparent to translucent; DE – 7.5; H – 5-5.5; CL – good; F – uneven to conchoidal; M – prismatic, tabular and dipyramidal crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites; also in placers. Crystals up to 120 mm (4 7/16 in) in size occur in Muiane, Alto Ligonha, Mozambique; also found in the Little Three mine, Ramona, and the Himalaya mine, Mesa Grande, California, USA. Also known from Greenbushes, Western Australia, Australia.
Application: Ta ore.

Stibiotantalite, 10 mm grain, Dobrá Voda, Czech Republic

Uraninite
UO_2

CUBIC ● ● ● ●

Properties: C – black, black-brown, black-gray; S – black, black-brown to greenish; L – submetallic, greasy, earthy; D – opaque; DE – 7.5-10,6; H – 5-6. earthy aggregates 3; CL – imperfect; F – uneven to conchoidal; M – cubic crystals, botryoidal aggregates, granular. massive; R – strong radioactive.
Origin and occurrence: Mainly hydrothermal in ore veins, skarns; magmatic in granitic pegmatites; in sedimentary rocks; also in placers.

Usually associated with other U minerals, e.g. coffinite and secondary alteration products, mainly U micas. Perfect crystals up to 100 mm (4 in) in size and weighing up to 2.5 kg (5 lb 8 oz) come from the Fissure mine, Wilberforce, Ontario, Canada, where they occur in simple pegmatites, cross-cutting marbles.
Crystals are also known from Dieresis, Spain and Shinkolobwe, Zair. Rich botryoidal aggregates were found in Jachymov and Slavkovice, Czech Republic; in Bois-Noirs and Margnac, France.
Application: U ore.

Uraninite, 8 mm x, Portland, U.S.A.

Gibbsite, 80 mm, Gamba, Brazil

Gibbsite
$Al(OH)_3$

MONOCLINIC ● ● ●

Properties: C – colorless, gray, white, greenish; S – white; L – vitreous, pearly; D – transparent to translucent; DE – 2.4; H – 2.5-3.5; CL – perfect; F – uneven; M – tabular crystals, lamellar and earthy aggregates, coatings and stalactitic films, granular, massive.
Origin and occurrence: Hydrothermal as a product of alteration Al-rich rocks; secondary in the oxidation associated with goethite; metamorphic in weakly metamorphosed Al-rich rocks, typically with diaspore; a constituent of bauxites. Tabular crystals up to 100 mm (4 in) in size were found in Zlatoust, Ural mountains, Russia; also in Villa Rica, Minas Gerais, Brazil.

Brucite
$Mg(OH)_2$

TRIGONAL ● ● ●

Properties: C – colorless, gray, white, bluish, blue, yellow, brown; S – white; L – vitreous, pearly; D – transparent to translucent; DE – 2.4; H – 2.5; CL – perfect; F – uneven; M – tabular crystals, foliated, acicular and earthy aggregates, granular, massive.
Origin and occurrence: Hydrothermal in veins in serpentinites or dolomitic marbles, a product of periclase alteration; rare metamorphic in skarns and marbles. Perfect crystals up to 18 cm in size, come from the Low's mine, Pennsylvania and the Tiliy Foster mine, New York, USA; also known from Asbestos, Quebec, Canada; Predazzo, the Alps, Italy; blue crystals up to 50 mm (2 in) in size were found in the Bazhenovskoye deposit, Azbest, Ural mountains, Russia.

Brucite, 30 mm, Azbest, Russia

Diaspore
AlO(OH)

ORTHORHOMBIC ● ● ●

Properties: C – colorless, gray, white, greenish, yellowish, pink, purplish; S – white; L – vitreous, pearly; D – transparent to translucent; DE – 3.4; H – 6.5-7; CL – perfect; F – conchoidal; M – tabular crystals, foliated aggregates, stalactitic films, granular, massive.
Origin and occurrence: Hydrothermal as a product of alteration Al-rich minerals, e.g. andalusite, typically with pyrophyllite and corundum; metamorphic in Al-rich rocks; a constituent of bauxites. Tabular crystals up to 120 mm ($4^7/_{16}$ in) in size come from Menderess, Turkey; also from Naxos, Greece; Chester, Massachusetts, USA.

Diaspore, 35 mm, Milas, Turkey

Goethite, 77 mm, Santa Eulalia, Mexico

Goethite, 60 mm, Pribram, Czech Republic

Goethite
$Fe^{3+}O(OH)$

ORTHORHOMBIC ● ● ● ● ●

Varieties: velvet ore

Properties: C – black-brown, yellow-brown, brown; S – yellow-brown; L – submetallic, metallic, silky, earthy; D – translucent to opaque; DE – 4.3; H – 5-5.5; CL – perfect; F – uneven to conchoidal; M – acicular to prismatic crystals, botryoidal aggregates, commonly with radial structure, coatings and stalactitic films, earthy, granular, massive.
Origin and occurrence: Secondary as one of the most common minerals of the oxidation zone of ore deposits, it forms a significant part of limonite; hydrothermal in ore veins, in cavities in pegmatites and volcanic rocks. It forms pseudo-morphs after pyrite and other Fe sulfides. Rich botryoidal aggregates of velvet ore with a velvety surface come from Příbram, Czech Republic; acicular crystals up to 50 mm (2 in) long are known from Bottalack and Redruth, Cornwall, UK; it also occurs in Siegen and Horhausen, Germany; in Florissant, Colorado, USA.
Application: Fe ore.

Lepidocrocite, 30 mm, Rudabanya, Hungary

Manganite
$Mn^{3+}O(OH)$

MONOCLINIC ● ● ● ●

Properties: C – black to black-gray; S – red-brown to black; L – submetallic to dull; D – opaque; DE – 4.3; H – 4; CL – perfect; F – uneven to conchoidal; M – long to short prismatic crystals, acicular and earthy aggregates, concretions, granular, massive.
Origin and occurrence: Hydrothermal in low-temperature ore veins, together with quartz; secondary in the oxidation zone of ore deposits; sedimentary and rare metamorphic in Mn-rich rocks. Druses of black crystals up to 40 mm (1 9/16 in) long come from the classic locality Ilfeld, Germany; it also occurs in Ohrenstock and Ilmenau, Germany; in Nikopol, Ukraine; in Sterling Hill, New Jersey, USA; in the N'Chwaning No. 2 mine, Kuruman, South Africa.
Application: Mn ore.

Lepidocrocite
$Fe^{3+}O(OH)$

ORTHORHOMBIC ● ● ●

Properties: C – dark red to red-brown; S – orange to brick-red; L – submetallic, adamantine to silky; D – transparent to opaque; DE – 4.0; H – 5; CL – perfect; F – uneven to conchoidal; M – tabular to short prismatic crystals, acicular, bladed and earthy aggregates, concretions, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of ore deposits, overgrown on botryoidal goethite. It occurs together with goethite as a constituent of limonite, its tabular crystals and their aggregates are known from Herdorf, Germany and Rancie, France.

Manganite, 85 mm, Ilfeld, Germany

Lithiophorite

$(Al,Li)Mn^{4+}O_2(OH)_2$

MONOCLINIC ● ● ●

Properties: C – black, commonly with bluish tint; S – black-gray to black; L – metallic to dull; D – opaque; DE – 3.3; H – 3; CL – perfect; F – uneven; M – scaly crystals, botryoidal and earthy aggregates, coatings, granular, massive.

Origin and occurrence: Secondary in the oxidation zone of ore deposits and along the cracks in sedimentary rocks. Botryoidal aggregates occur in Schneeberg, Germany; Jivina and Zajecov, Czech Republic; Miyazaki, Japan.

Lithiphorite, 60 mm, Rangersdorf, Germany

Curite

$Pb_2U_5O_{17} \cdot 4\,H_2O$

ORTHORHOMBIC ● ●

Properties: C – dark orange to red-orange; S – light orange; L – adamantine to earthy; D – transparent to translucent; DE – 7.4; H – 4-5; CL – good; F – uneven; M – acicular crystals, earthy aggregates, coatings, massive; R – strong radioactive.

Origin and occurrence: Secondary in the oxidation zone of U deposits, associated with other secondary U minerals, e.g. torbernite, kasolite and uranophane. Rich aggregates were found in Shinkolobwe, Zaire; also known from La Crouzille, France; South Alligator, Northern Territory, Australia.

Curite, 30 mm, Shinkolobwe, Zair

5. Carbonates

Magnesite
$MgCO_3$

TRIGONAL ● ● ● ●

Properties: C – colorless, white, yellowish, brownish, black; S – white; L – vitreous to dull; D – transparent; DE – 3.1; H – 4; CL – perfect; F – conchoidal; M – rhombohedral and prismatic crystals, massive cleavable aggregates, earthy; LU – occasionally blue or green.
Origin and occurrence: Rarely magmatic, mainly hydrothermal metasomatic and metamorphic. The largest crystals are known from Brumado, Bahía, Brazil, reaching up to 1 m (39 3/8 in) in size, embedded in metamorphosed dolomites. Crystals in cavities in the same locality are up to 50 mm (2 in) in size. Also crystals up to 50 mm (2 in) found in the Eugui quarries, Spain. Crystals up to 10 mm (3/8 in) across come also from Val Malenco, Italy. It prevails as massive aggregates, forming huge deposits, like Veitsch, Austria; Liao-Tung, China. Many deposits are located in Slovakia (Jelsavska Dubrava, Hnusta).
Application: heat-resistant material.

Calcite, 40 mm, Houghton Co., U.S.A.
Magnesite, 148 mm, Brumado, Brazil

Smithsonite, 32 mm, Tsumeb, Namibia

Smithsonite
$ZnCO_3$

TRIGONAL ● ● ●

Properties: C – white, gray, green, pink, blue; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 4.4; H – 4-5; CL – perfect; F – conchoidal to uneven; M – rhombohedral crystals, botryoidal and stalactitic aggregates, massive; LU – sometimes green or blue.
Origin and occurrence: Supergene, as a result of oxidation of the primary Zn ores, associated with other supergene Pb minerals. The largest yellow scalenohedra crystals up to 40 mm (1 9/16 in) in size come from Broken Hill, New South Wales, Australia. Pink crystals, up to 30 mm (1 3/16 in) long occurred in Tsumeb, Namibia. World famous blue-green botryoidal crusts up to 100 mm (4 in) thick found in the Kelly Mine, Magdalena, New Mexico, USA. Nice aggregates and banded stalactites described from Monte Poni, Sardinia, Italy.

Smithsonite, 40 mm, New Mexico, U.S.A.

Siderite, 26 mm, Governador Valadares, Brazil

Siderite

$FeCO_3$

TRIGONAL ● ● ● ●

Properties: C – yellow, brown, black; S – yellowish-white; L – vitreous; D – translucent; DE – 4; H-4; CL – perfect; F – uneven to conchoidal; M – rhombohedral crystals, granular, massive.

Origin and occurrence: Hydrothermal in medium- and low-temperature deposits, sedimentary. Crystals up to 40 cm long found in Mont St.-Hilaire, Quebec, Canada. Crystals up to 100 mm (4 in) in size come from Panasqueira, Portugal. Crystals up to 30 mm (1 3/16 in) in size occurred in Neudorf, Germany. Rhombs up to 20 mm (25/32 in) also found in Pribram, Czech Republic. Pseudo-morphs of goethite after siderite up to 70 mm (2¾ in) across described from Pikes Peak, Colorado, USA. Deposits of massive siderite in Erzberg and Hüttenberg, Austria yielded crystals up to 50 mm (2 in) long. Fine crystals reported from Tavistock, Devon and Redruth, Cornwall, UK.

Application: important Fe ore.

Sfaerocobaltite

$CoCO_3$

TRIGONAL ● ● ●

Properties: C – pink, gray, brown; S – red; L – vitreous; D – transparent to translucent; DE – 4.1; H – 4; CL – perfect; M – scalenohedral and rhombohedral crystals, radial aggregates, massive.

Origin and occurrence: Secondary, as a product of the oxidation of primary Co minerals. The best specimens, with crystals up to 30 mm (1 3/16 in) long, come from Zaire (Musonoi; Kakanda). Crystals up to 10 mm (4 in) long known from Bou Azzer, Morocco.

Sphaerocobaltite, 16 mm, Bou Azzer, Morocco

Rhodochrosite, 117 mm, Sweet Home Mine, U.S.A.

Rhodochrosite

$MnCO_3$

TRIGONAL ● ● ● ●

Properties: C – white, pink, red, brown, locally black coatings on crystals; S – white; L – vitreous; D – transparent to translucent; DE – 3.6; H – 3.5-4; CL – perfect; F – conchoidal to uneven; M – rhombohedral and scalenohedral crystals, hemispherical and botryoidal aggregates, granular, massive.

Origin and occurrence: Only rare in pegmatites, hydrothermal in medium- and low-temperature veins, sedimentary and metamorphic. The most beautiful crystals come from the Sweet Home mine, Alma, Colorado, USA, where rhombs up to 150 mm (6 in) in size occur, associated with purple fluorite, hubnerite, tetrahedrite and other minerals. Beautiful dark red scalehohedra, up to 100 mm (4 in) found in the N'Chwaning No.1 and 2 mines, Kuruman, South Africa. Pink rhombs up to 80 mm (3⅛ in) come from Silverton, Colorado, USA. Nice pink hemispheres and botryoidal aggregates are known from Cavnic and Baia Sprie, Romania. Similar specimens occurred in Huaron, Peru. Pink banded crusts and stalactites were found in the Mina Capillitas, Catamarca, Argentina. Pink rhombs, associated with bertrandite, were lately found in Kounrad, Kazakhstan.

Rhodochrosite, 70 mm, Mina Capillitas, Argentina

Calcite, 150 mm, Elmwood Mine, U.S.A.

Calcite

$CaCO_3$

TRIGONAL ● ● ● ● ●

Properties: C – colorless, white, gray, yellow, brown, pink, red, blue, green, black; S – white; L – vitreous to pearly; D – transparent to opaque; DE – 2.7; H – 3; CL – perfect; F – conchoidal; M – crystals of various habit, concretions, stalactites, oolitic aggregates, granular, massive; LU – sometimes red to orange.

Calcite, 30 mm, Tunguzka, Russia

Origin and occurrence: One of the most common minerals, resulting from a wide range of conditions, it is magmatic, hydrothermal, sedimentary, metamorphic and secondary, it occurs in various parageneses. Large crystals found in many localities throughout the world. Pinkish and yellow crystals over 500 mm (20 in) long come from Joplin, Missouri and from the Elmwood mine, Tennessee, USA. Beautiful calcite crystals, crystallographically of very complex habits, found in Dalnegorsk, Russia. Nice calcite specimens occurred also in Mexico (Naica, Chihuahua; Charcas, San Luis Potosí). European localities, like Příbram, Czech Republic; St. Andreasberg, Germany; Kongsberg, Norway are famous by their calcites, too. Wine yellow, complicated crystals are known from the Sarbayskoye deposit in Rudnyi, Kazakhstan. Clear cleavable aggregates of the birefringent calcite (so called Iceland spar) were found in basalt cavities in Helgustadir, Iceland. The largest of them reached up to 6 x 2 m (20 x 6 ft 6 in) in size. Very nice scalenohedra up to 80 mm (3 1/8 in) long with copper inclusions occur in Keweenaw Peninsula, Michigan, USA. Perfect scalenohedra and their twins up to 100 mm (4 in) long are known from Egremont and Frizington, UK. Beautiful butterfly twins up to 80 mm (3 1/8 in) recently reported from Guiyang, Hunan, China. Crystals of calcite with sand inclusions up to 100 mm (4 in) in size come from the vicinity of Fontainebleau, France.

Application: building industry, optical industry.

Dolomite, 33 mm x, Navarro, Spain

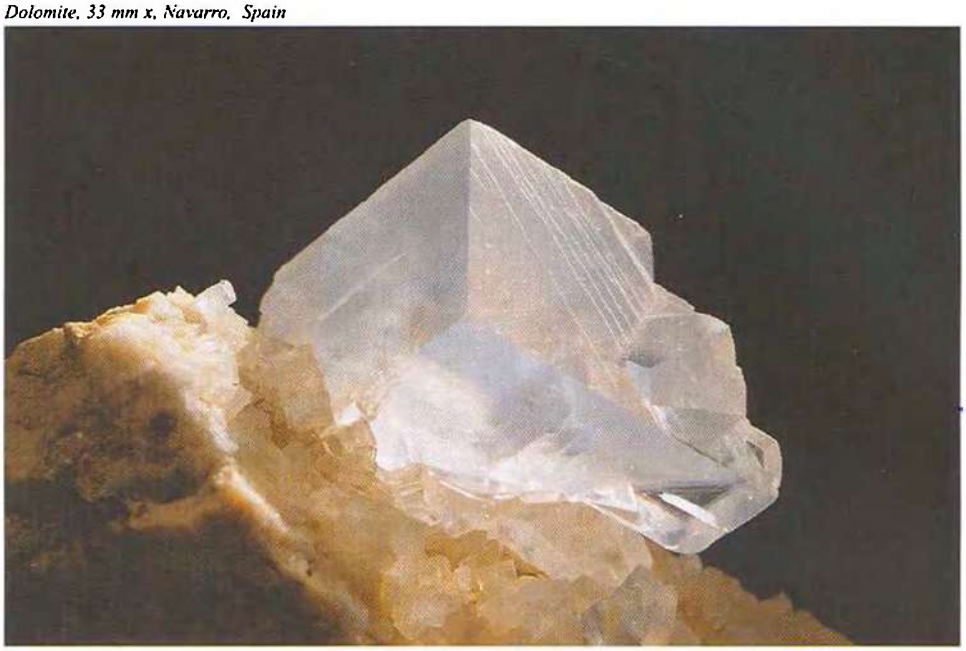

Dolomite
$CaMg(CO_3)_2$

TRIGONAL ● ● ● ● ●

Properties: C – gray-white, pink, red, green, brown, black; S – white; L – vitreous to pearly; D – transparent to translucent, DE – 2.9; H – 3.5; CL – perfect; F – conchoidal; M – rhombohedral crystals, massive.
Origin and occurrence: Magmatic in pegmatites, hydrothermal, metasomatic, sedimentary and metamorphic, together with siderite, magnesite, calcite and other minerals. Crystals up to 100 mm (4 in) long found in Brumado, Bahia, Brazil. Fine crystals also occurred in Banska Stiavnica, Slovakia and in Cavnic, Romania. Crystals up to 200 mm (7⅞ in) in size come from Eugui, Spain. Crystals up to 150 mm (6 in) long found in cavities in dolomitic rocks in Lengenbach, Binntal, Switzerland. Crystals several cm long known from Jachymov, Czech Republic. Large accumulations of massive dolomite are common in magnesite deposits.
Application: metallurgy, heat-resistant material.

Ankerite, 20 mm xx, Roudny, Czech Republic

Ankerite
$CaFe(CO_3)_2$

TRIGONAL ● ● ● ●

Properties: C – white, yellowish, brown-yellow; S – white; L – vitreous to pearly; D – translucent; DE – 3; H – 3.5-4; CL – perfect; F – conchoidal; M – rhombohedral crystals, granular.
Origin and occurrence: Hydrothermal in medium- and low-temperature veins, also sedimentary and metamorphic, together with siderite and other minerals. Crystals up to 50 mm (2 in) occurred in the Tui mine, New Zealand. Brown rhombs up to 40 mm (1⁹⁄₁₆ in) in size known from Gilman, Colorado, USA. Yellowish crystals up to 10 mm (⅜ in) come from concretions near Kladno, Czech Republic. Massive aggregates are common in metasomatic deposits of siderite (e.g. Nizna Slana, Slovakia).

Kutnohorite, 60 mm, Kutna Hora, Czech Republic

Huntite, 110 mm, Koksin, Czech Republic

Kutnohorite

$CaMn(CO_3)_2$

TRIGONAL ● ● ●

Properties: C – white, gray, pink, yellowish; S – white; L – vitreous; D – translucent; DE – 3.1; H – 3.5-4; CL – perfect; F – conchoidal; M – poorly developed crystals, granular, massive.
Origin and occurrence: Hydrothermal and metamorphic, associated with ankerite, quartz and other minerals. Poorly developed crystals are known from Kutna Hora, Czech Republic. Small crystals several mm in size are described from Mont St.-Hilaire, Quebec, Canada. Large crystals occurred in Moncure, North Carolina, USA. Small gray-white crystals were found in Broken Hill, New South Wales, Australia.

Huntite

$CaMg(CO_3)_2$

TRIGONAL ● ●

Properties: C – white; S – white; L – earthy; D – opaque; DE – 2.7; H – 1.5; CL – none; F – conchoidal; M – fibrous aggregates, earthy.
Origin and occurrence: Secondary mineral resulting from the oxidation of dolomite, associated with magnesite and dolomite. Fine fibrous aggregates come

Aragonite, 49 mm, Tazouta, Morocco

Flos ferri, 110 mm, Erzberg, Austria

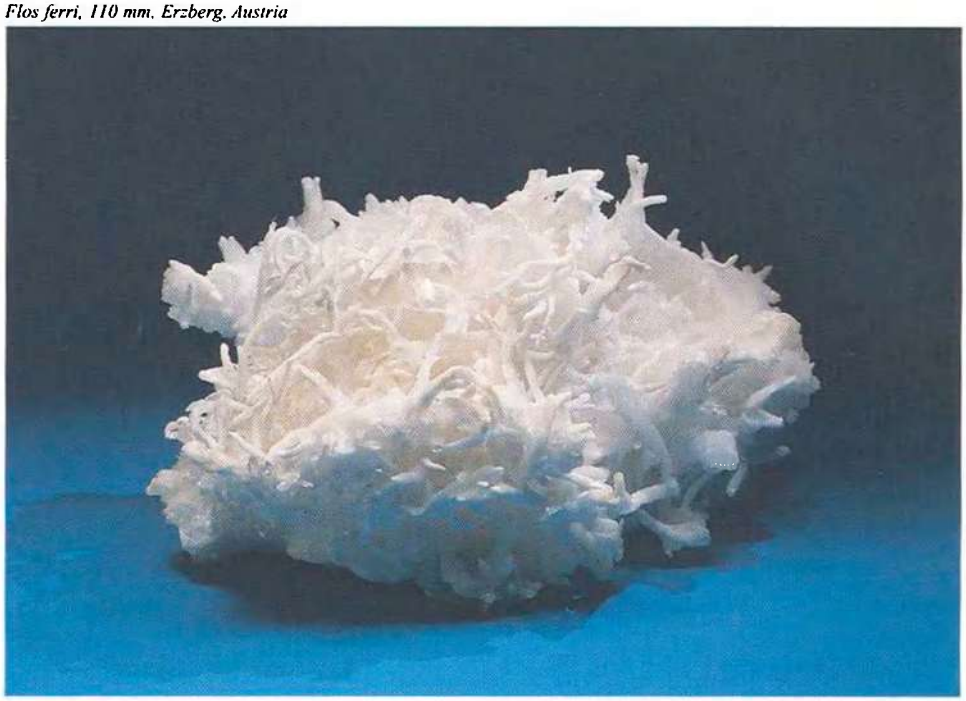

from Koksin, Czech Republic. It also occurs in the Ala-Mar deposit, Nevada, USA. Massive aggregates are known from the Boquira mine, Bahia, Brazil.

Aragonite

$CaCO_3$

ORTHORHOMBIC ● ● ● ●

Varieties: flos ferri, hot-spring tufa, peastone, tarnowitzite.

Properties: C – colorless, white, yellow, reddish, greenish, purplish, bluish, gray; S – white; L – vitreous; D – transparent to opaque; DE – 3; H – 3.5-4.5; CL – imperfect; F – conchoidal; M – prismatic crystals, oolitic, banded, columnar and dendritic aggregates, massive; LU – locally weak cream yellow.

Peastone, 35 mm, Karlovy Vary, Czech Republic

Origin and occurrence: Primary as a late hydrothermal mineral of high-temperature deposits more commonly secondary as a product of the oxidation of siderite and pyrite. It also results from precipitation of thermal springs, it is sedimentary and metamorphic. Fine white prismatic crystals up to 70 mm (2¾ in) long found together with blue celestite crystals in _pania Dolina, Slovakia. Similar crystals come from sulfur deposits in Cianciano, Italy and Tarnobrzeg, Poland. Maybe the best aragonite crystals in the world are known from the gossan of the magnesite deposit in Podrecany, Slovakia, where druses of crystals up to 200 mm (7⅞ in) long were found. Wine yellow crystals up to 100 mm (¾ in) long occurred in cavities of volcanic rocks in Horenec near Bilina, Czech Republic. Those crystals were the only gem rough, suitable for facetting, in the world. Very interesting copper pseudo-morphs after aragonite come from Corocoro, Bolivia. Banded and oolitic aggregates found in Karlovy Vary, Czech Republic. Dendritic aggregates from Erzberg, Austria and elsewhere are known as flos ferri.

Application: decorative stone.

Strontianite, 34 mm, Hardin Co., U.S.A.

Strontianite
$SrCO_3$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, gray, yellowish, greenish, reddish, brown; S – white; L – vitreous to resinous; D – transparent to translucent; DE – 3.8; H – 3.5; CL – perfect; F – conchoidal to uneven; M – prismatic and acicular crystals, columnar and fibrous aggregates, massive, earthy.
Origin and occurrence: Hydrothermal in low-temperature veins, in cavities of volcanic rocks, mainly sedimentary, together with calcite and zeolites. Crystals up to 80 mm (3¼ in) long occur near Bleiberg, Austria. Smaller crystals known from Strontian, Scotland. Crystals up to 50 mm long found in marls in Ahlen near Münster, Germany.

Witherite
$BaCO_3$

ORTHORHOMBIC ● ●

Properties: C – white, gray, yellowish; S – white; L – vitreous to greasy, D – translucent; DE – 4.3; H – 3.5; CL – good; F – uneven; M – pseudo-hexagonal dipyramidal crystals, fibrous and botryoidal aggregates, massive.
Origin and occurrence: Hydrothermal in low-temperature deposits with fluorite, barite, calcite and other minerals. Beautiful yellowish crystals up to 120 mm (4 11/16 in) long come from the Minerva No. 1 mine, Cave-in-Rock, Illinois, USA. Crystals, up to 70 mm (2¾ in) in size occurred in Hexham and Alston Moor, England, UK. Its pseudo-hexagonal crystals or botryoidal aggregates were very rare in Pribram, Czech Republic.

Cerussite
$PbCO_3$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, whitish, yellow, black; S – white; L – greasy to adamantine; D – translucent; DE – 6.6; H – 3-3.5; CL – good; F – conchoidal; M –

Witherite, 20 mm x, Alston Moor, UK

Cerussite, 92 mm, Tsumeb, Namibia

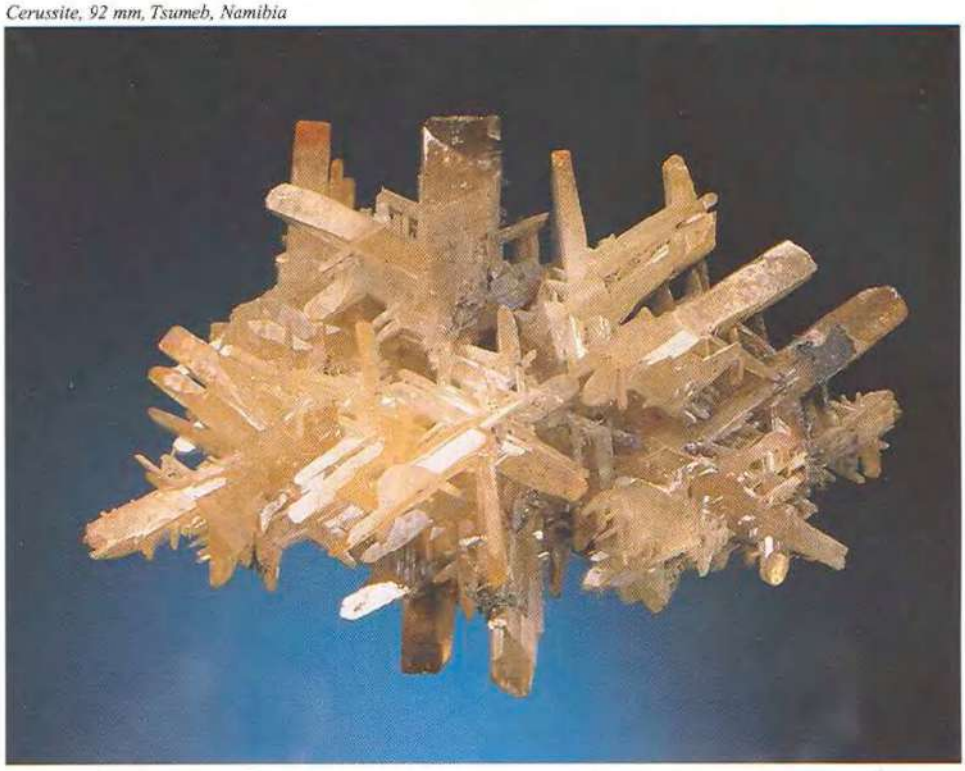

prismatic and pyramidal crystals, common trillings and twins, granular, massive; LU – sometimes yellowish.

Origin and occurrence: Secondary mineral, resulting from the oxidation of galena and other Pb minerals, together with pyromorphite, vanadinite, barite and other minerals. The best specimens come from Tsumeb, Namibia, where trillings up to 200 mm (7⅞ in) in diameter occurred. Large twins also known from Broken Hill, New South Wales, Australia. Beautiful crystals up to 50 mm (2 in) in size reported from Mibladen, Morocco. Fine crystals up to 50 mm (2 in) long found in Stribro, Czech Republic. Typical white acicular crystals, up to 60 mm (2⅜ in) long come from the Flux mine, Arizona, USA.

Barytocalcite

$BaCa(CO_3)_2$

MONOCLINIC ● ●

Properties: C – colorless, white, gray, yellowish; S – colorless; L – vitreous to resinous; D – transparent to translucent; DE – 3.7; H – 4; CL – perfect; F – conchoidal to uneven; M – prismatic, often striated crystals, massive; LU – light yellow.

Origin and occurrence: Hydrothermal in low-temperature veins, together with calcite, barite and other minerals. Crystals several cm in size and cleavable masses are known from Alston Moor, Cumbria, UK. Imperfect crystals about 10 mm (⅜ in) in size occurred in Stříbro, Czech Republic. It is also described from Freiberg, Germany and from Langban, Sweden.

Barytocalcite, 52 mm, Alston Moor, UK

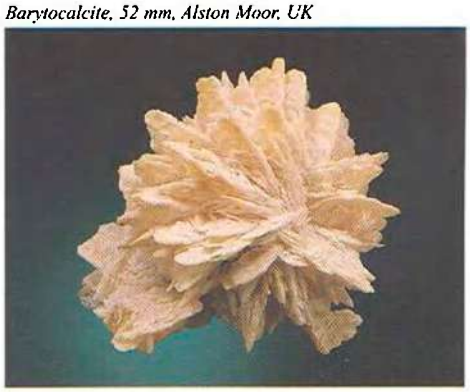

Azurite, 48 mm, Touissit, Morocco

Malachite, 63 mm, Mashamba West, Zair

Azurite

$Cu_3(CO_3)_2(OH)_2$

MONOCLINIC ● ● ● ●

Properties: C – blue; S – blue; L – vitreous; D – transparent to opaque; DE – 3.8; H – 3.5; CL – perfect; F – conchoidal; M – tabular and prismatic crystals, pulverulent.

Origin and occurrence: Secondary, resulting from the oxidation of Cu sulfides, mainly associated with malachite. Crystals up to 200 mm (7⅞ in) long come from Tsumeb, Namibia. Crystals from Touissit, Morocco reach up to 70 mm (2¾ in). Tabular crystals up to 50 mm (2 in) in size found in Chessy near Lyon, France. Famous crystals up to 50 mm (2 in) across occurred in the Copper Queen mine in Bisbee, Arizona, USA. Fine azurite concretions with crystals on the surface come from La Sal, Utah, USA. Crystal rosettes, reaching up to 130 mm (5 1/16 in) in size were found in the Yang Chweng Mine, Guang Dong, China.

Application: Cu ore.

Malachite

$Cu_2(CO_3)(OH)_2$

MONOCLINIC ● ● ● ●

Properties: C – green; S – light green; L – vitreous, dull, earthy; D – opaque; DE – 4.1; H – 3.5-4; CL – perfect; F – conchoidal to uneven; M – acicular and prismatic crystals, botryoidal aggregates, stalactites, massive.

Origin and occurrence: The most common supergene mineral of Cu, associated with azurite, cuprite and other minerals. Crystals up to 30 mm (1 3/16 in) in size occur in Kambove, Zaire. Crystals up to 20 mm (25/32 in) also found in Rudab<nya, Hungary. Pseudomorphs after azurite crystals up to 100 mm (4 in) in size known from Tsumeb; pseudo-morphs after cuprite crystals up to 50 mm in size come from Onganja, Namibia. The blocks of banded malachite weighing up to 250 tons , occurred in Mednorudnjansk, Ural mountains, Russia. Similar material in huge quantities come from many deposits in Shaba

Malachite, 70 mm, Shaba, Zair

Rosasite, 270 mm, Gleeson, U.S.A.

province, Zaire, where stalactites up to 500 mm (20 in) long were also found.
Application: Cu ore, production of decorative objects and jewelry.

Rosasite
$(Cu,Zn)_2(CO_3)(OH)_2$

MONOCLINIC ● ●

Properties: C – green, blue; S – greenish; D – opaque; DE – 4.0-4.2; H – 4.5; CL – good; M – small acicular crystals, fibrous and botryoidal crusts.
Origin and occurrence: Secondary, forming in the oxidation zone of Cu and Zn deposits, together with other secondary minerals of Cu. Spherical aggregates of acicular crystals up to 10 mm (⅜ in) found in Mina Ojuela, Mapimi, Durango, Mexico. Similar finds were made in Bisbee, Arizona, USA.

Hydrozincite, 65 mm, Tiger, U.S.A.

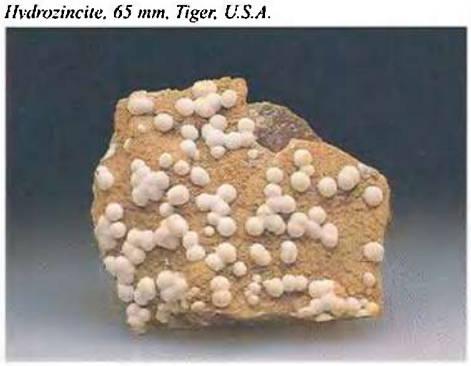

Hydrozincite
$Zn_5(CO_3)_2(OH)_6$

MONOCLINIC ● ● ●

Properties: C – white, yellowish; S – white; L – pearly to dull; D – opaque; DE – 4; H – 2-2.5; CL – perfect; F – conchoidal; M – tabular crystals, massive, earthy; LU – locally blue.
Origin and occurrence: Secondary, resulting from the oxidation of sphalerite, together with cerussite, smithsonite and hemimorphite. Small crystals are known from Mapimi, Durango, Mexico. Spherical aggregates, several mm in diameter, were found in Sterling Hill, New Jersey, USA. Stalactites and thick crusts occurred in Long-Kieng, Burma. Crusts and stalactites were also described from Bleiberg, Austria; from Mezica, Slovenia and from Raibl, Italy.

Aurichalcite
$(Zn,Cu)_5(CO_3)_2(OH)_6$

ORTHORHOMBIC ● ● ●

Properties: C – light green, blue-green, blue; S – blue-green; L – silky to pearly; D – translucent; DE – 4; H – 1-2; CL – perfect; M – acicular crystals, crusts.
Origin and occurrence: Secondary in the oxidation zone of Cu and Zn deposits in the arid climate, associated with linarite and other minerals. Large prismatic crystals come from Mina Ojuela, Mapimi, Durango, Mexico. Nice rosettes of acicular crystals are known from Bisbee and from the 79 mine, Banner district, Arizona, USA. It also occurred in Monteponi, Sardinia, Italy.

Aurichalcite, 80 mm, Arizona, U.S.A.

Dawsonite, 2 mm aggregate, Recsk, Hungary

Dawsonite

$NaAl(CO_3)(OH)_2$

ORTHORHOMBIC ● ● ●

Properties: C – colorless to white; S – white; L – vitreous; D – transparent; DE – 2.4; H – 3; CL – perfect; M – acicular to blade-like crystals, radial aggregates.
Origin and occurrence: Hydrothermal in low-temperature deposits, associated with calcite, dolomite and other minerals. Fine acicular crystals up to 35 mm (1⅜ in) long come from Mont St.-Hilaire, Quebec, Canada. Radial aggregates along the rock cracks were found in Dubnik and Zlata Bana, Slovakia.

Bastnäsite-(Ce)

$(Ce,La)(CO_3)F$

HEXAGONAL ● ● ●

Properties: C – yellow to brown; S – yellow-brown; L – vitreous to greasy; D – translucent; DE – 4.8-5.2;

Parisite-(Ce), 71 mm, Boyaca, Colombia

Bastnasite-(Ce). 6 mm x, Ariege, France

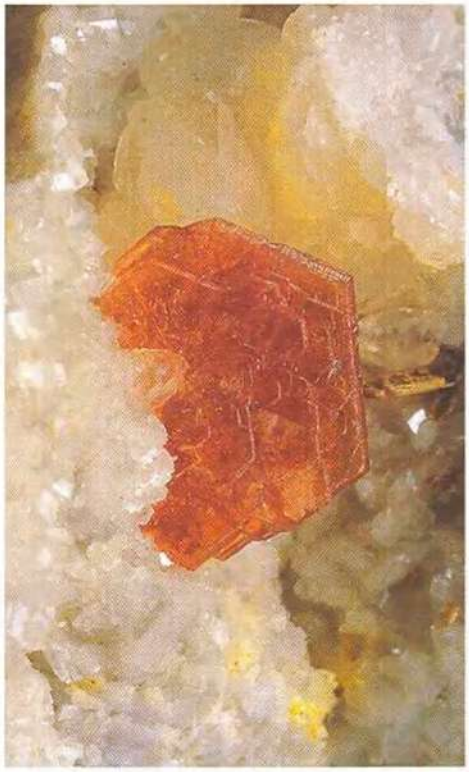

H – 4-4.5; CL – good; F – uneven; M – tabular crystals, granular, massive.
Origin and occurrence: Magmatic in pegmatites, also metamorphic, together with allanite-(Ce) and other rare earth elements minerals. Crystals up to 200 x 150 mm (7⅞ x 6 in) in size occur in Andekatany, Madagascar. Crystals up to 100 mm (4 in) long are known from Karonge, Burundi. Transparent crystals up to 25 mm (1 in) long were found recently in the Trimouns quarry, France. It forms important local deposits, as those of Mountain Pass, California, USA.*Application:* ore of rare earth elements.

Parisite-(Ce)
$CaCe_2(CO_3)_3F_2$

TRIGONAL ● ● ●

Properties: C – brown, yellow-brown, gray-yellow; S – brownish; L – vitreous to resinous; D – transparent to translucent; DE – 4.4; H – 4.5; CL – perfect; F – conchoidal to splintery; M – dipyramidal striated crystals.
Origin and occurrence: Magmatic in pegmatites, also hydrothermal and rarely metamorphic, together with bastnasite-(Ce) and other minerals of rare earth elements. Crystals several cm long come from Quincy, Massachusetts, USA. Crystals up to 80 mm (3⅛ in) long occurred in pegmatites near Hundholmen, Norway. Crystals up to 23 cm (9 7/16 in) long found in the Snowbird mine, Montana, USA. Transparent crystals up to 15 mm (1 9/32 in) were lately reported in the Trimouns quarry, France. Very unusual association have been described from Muzo, Columbia, where crystals up to 50 mm occur together with emeralds.

Phosgenite
$Pb_2(CO_3)Cl_2$

TETRAGONAL ● ●

Properties: C – colorless, white, yellow-white, gray, brown; S – white; L – adamantine; D – transparent to translucent; DE – 6.1; H – 2-3; CL – good; F – conchoidal; M – short to long prismatic and tabular crystals, granular, massive; LU – sometimes yellow.
Origin and occurrence: Secondary, resulting from the oxidation of galena, associated with cerussite and other secondary Pb minerals. Crystals up to 150 x 100 mm (6 x 4 in) across known from Monteponi, Sardinia, Italy. Crystals from Tsumeb, Namibia reached up to 100 mm. Crystals up to 30 mm (1 3/16 in) in size found in Matlock, Derby, UK. Crystals up to 35 mm (1⅜ in) long occurred in the Mammoth mine, Tiger, Arizona, USA.

Phosgenite. 28 mm, Monteponi, Italy

Bismutite. 3 mm xx. Brdarka, Slovakia

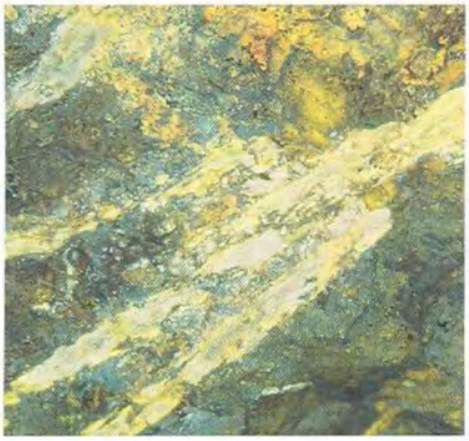

Bismutite

$Bi_2O_2(CO_3)$

TETRAGONAL ● ● ●

Properties: C – yellow, brown, gray, blue, black; S – white; L – vitreous, pearly to dull; D – translucent to opaque; DE – 6.1-7.7; H – 3.5; CL – good; M – spherical, radial aggregates, massive, pulverulent.
Origin and occurrence: Secondary, originated from the oxidation of Bi minerals, associated with bismuthinite and other minerals. It is common from pegmatites in Madagascar (Ampangabe) and in Mozambique. Large pebbles found in Kivu province, Zaire. It also occurred in Tasna, Bolivia.

Weloganite. 30 mm xx. Francon Quarry, Canada

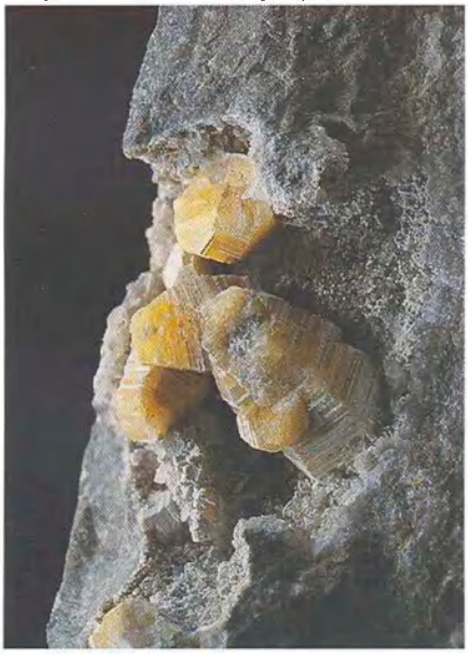

Gaylussite. 30 mm. Natron Lake, Tanzania

Gaylussite

$Na_2Ca(CO_3)_2 \cdot 5\,H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white, gray, yellowish; S – colorless; L – vitreous; D – transparent to translucent; DE – 2; H – 2.5-3; CL – perfect; F – conchoidal; M – lenticular to prismatic crystals.
Origin and occurrence: Sedimentary, typical a constituent of salt sediments. Crystals up to 80 mm (3 1/8 in) long come from Searles Lake, California, USA. It also occurs in Borax Lake and Mono Lake, California, USA. Large crystals are known from Amboseli Lake, Kenya.

Weloganite

$Na_2Sr_3Zr(CO_3)_6 \cdot 3\,H_2O$

TRICLINIC ●

Properties: C – white, yellow; S – white; L – vitreous; D – transparent to opaque; DE – 3.2; H – 3.5; CL – perfect; F – conchoidal; M – pseudo-hexagonal striated crystals.
Origin and occurrence: Hydrothermal in the alkaline rocks, together with zircon, dresserite and other minerals. Crystals up to 50 mm (2 in) come from cavities in the Francon quarry, Montreal; crystals up to 30 mm (1 3/16 in) are known from St. Michel, Quebec, Canada.

Artinite, 10 mm, Fethiya, Turkey

Artinite

$Mg_2(CO_3)(OH)_2 \cdot 3\,H_2O$

MONOCLINIC ● ● ●

Properties: C – white; S – white; L – silky; D – transparent; DE – 2; H – 2.5; CL – perfect; M – spherical, radial aggregates, veinlets, crusts.
Origin and occurrence: Hydrothermal, originating at low temperatures in serpentinites, associated with magnesite, aragonite and other minerals. Radial aggregates up to 20 mm (²⁷/₃₂ in) in size occur in the Gem mine, San Benito County, California, USA. Needles up to 20 mm (²⁵/₃₂ in) long come from Val Malenco, Italy. Clusters of acicular crystals were found on Staten Island, New York, USA.

Zaratite, 70 mm, Lancaster Co., U.S.A.

Hydrotalcite, 1 mm xx, Dunabogdany, Hungary

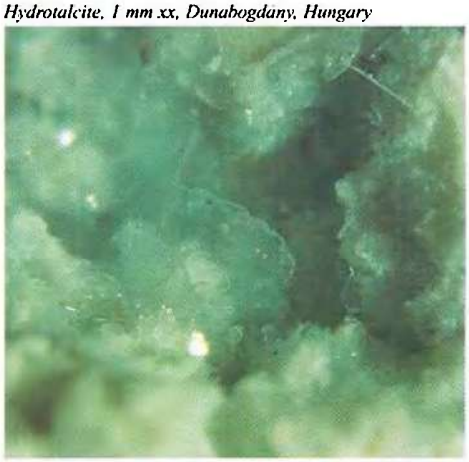

Zaratite

$Ni_3(CO_3)(OH)_4 \cdot 4\,H_2O$

CUBIC ● ● ●

Properties: C – emerald-green; S – light green; L – vitreous to greasy; D – transparent to translucent; DE – 2.6-2.7; H – 3.5; CL – none; F – conchoidal; M – crystalline crusts, stalactites, coatings.
Origin and occurrence: Secondary, originating from the oxidation of Ni minerals, associated with millerite, brucite and other minerals. It occurs as a product of the oxidation of Ni minerals in ultrabasic rocks in Kraubat and Stubachtal, Austria. Large green coatings found in Heazlewood, Tasmania, Australia. It covers millerite needles in the vicinity of Kladno, Czech Republic.

Hydrotalcite

$Mg_6Al_2(CO_3)(OH)_{16} \cdot 4\,H_2O$

TRIGONAL ● ●

Properties: C – white; S – white; L – pearly to waxy; D – transparent; DE – 2.1; H – 2; CL – perfect; M – fibrous and layered aggregates, massive.
Origin and occurrence: : Hydrothermal in ultrabasic and metamorphic rocks. It occurred with serpentine in Nordmark, Norway and also found at Franklin and Sterling Hill, New Jersey, USA.

Stichtite, 35 mm, Rouchovany, Czech Republic

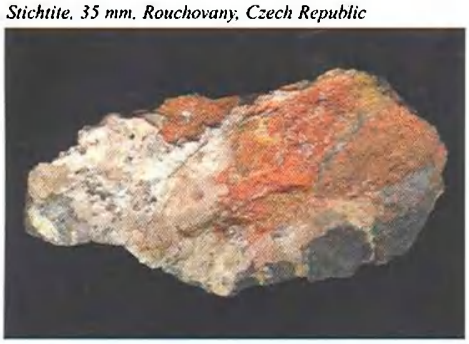

Stichtite

$Mg_6Cr_2(CO_3)(OH)_{16} \cdot 4\ H_2O$

TRIGONAL ● ●

Properties: C – pink to purple; S – white to light purple; L – pearly, waxy to greasy; D – translucent; DE – 2.2; H – 1.5-2; CL – perfect; M – lamellar and fibrous aggregates, massive.
Origin and occurrence: It occurs in serpentinites as scaly aggregates and veinlets in Bou Azzer, Morocco; in Dundas, Tasmania, Australia and in Barberton, South Africa.

Alumohydrocalcite

$CaAl_2(CO_3)_2(OH)_4 \cdot 3\ H_2O$

TRICLINIC ● ●

Properties: C – white, gray; S – white; L – earthy; D – opaque; DE – 2.2; H – 2.5; CL – perfect; M – chalky aggregates, consisting of acicular crystals.
Origin and occurrence: Secondary, associated with alophane and other minerals. It was described from the Khakasy deposit, Siberia, Russia, where it originates from the oxidation of alophane. Nice white radial aggregates occurred along the cracks in shales in L'adomírovo, Slovakia.

Ancylite-(Ce)

$Sr_3(Ce,La)_4(CO_3)_7(OH)_4 \cdot 3\ H_2O$

ORTHORHOMBIC ● ●

Properties: C – colorless, yellow, yellow- brown, light purple, brown; S – white; L – vitreous to greasy; D – transparent to opaque; DE – 4; H – 4-4.5; CL – none; F – splintery; M – short to long prismatic and pseudo-octahedral crystals.
Origin and occurrence: Hydrothermal in alkaline rocks, associated with nepheline and other minerals. Crystals up to 6 mm (¼ in) long are known from

Alumohydrocalcite, 70 mm, Ladomirovo, Slovakia

6. Borates

Ludwigite
$Mg_2Fe^{3+}BO_5$

ORTHORHOMBIC ● ● ●

Properties: C – dark green, black-green, black; S – blue-green; L B silky to dull; D – opaque; DE – 3.9; H – 5; CL – perfect; F B uneven; M – prismatic crystals, fibrous aggregates, granular, massive.
Origin and occurrence: Metamorphic in skarns and dolomitic marbles, locally associated with magnetite and other borates. Rich aggregates are known from Ocna de Fier, Romania; Kamineichi, Japan and the Hol Kol mine, Suan, North Korea.
Application: chemical industry, B ore.

Gaudefroyite
$Ca_4Mn^{3+}{}_3(BO_3)_3CO_3(O,OH)_3$

HEXAGONAL ●

Properties: C – black; S – black; L – adamantine to dull; D – opaque; DE – 3.4; H – 6; CL – good; F – uneven; M – prismatic crystals, fibrous and acicular aggregates.
Origin and occurrence: Hydrothermal in calcite veins. Prismatic crystals up to 50 mm (2 in) long were found on the mine dumps near Tachgagalt, Morocco and in the N'Chwaning No. 2 mine and the Wessels mine, Kuruman, South Africa.

Gaudefroyite, 40 mm, Kuruman, South Africa
Ludwigite, 120 mm, Ocna de Fier, Romania

Inderite, 230 mm, Boron, U.S.A.

Inderite
$MgB_3O_3(OH)_5 \cdot 5\,H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white to pink in aggregates; S – white; L – vitreous; D – transparent to translucent; DE – 1.8; H – 2.5; CL – good; F – conchoidal to uneven; M – prismatic crystals, acicular and fibrous aggregates, nodules, massive.
Origin and occurrence: Sedimentary in boron deposits, commonly associated with colemanite and other borates. Prismatic crystals up to 100 mm (4 in) long occur in Boron, California, USA. It is also known from Inder, Kazakhstan.
Application: chemical industry, B ore.

Invoite, 90 mm, Turkey

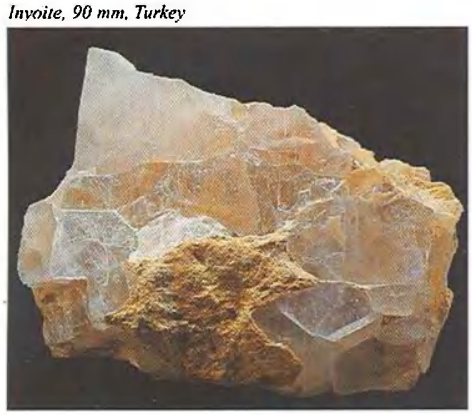

Borax, 60 mm x, Searles Lake, U.S.A.

Inyoite

$CaB_3O_3(OH)_5 \cdot 4\,H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white; S – white; L – vitreous; D – transparent to translucent; DE B 1.9; H – 2; CL – good; F – uneven; M – short prismatic to tabular crystals, columnar aggregates, massive.
Origin and occurrence: Sedimentary in boron deposits, commonly associated with colemanite and other borates. Clear tabular crystals up to 100 mm (4 in) long found in Kirka and Emet, Turkey. It also comes from Inder, Kazakhstan and the Corkscrew mine, California, USA.
Application: chemical industry, B ore.

Ulexite, 20 mm aggregate, Death Valley, California, U.S.A.

Borax

$Na_2B_4O_5(OH)_4 \cdot 8\ H_2O$

MONOCLINIC ● ● ●

Properties: C – white, colorless, yellowish, gray, greenish; S – white; L B vitreous to dull; D – transparent, translucent to opaque; DE – 1.7; H – 2-2.5; CL – perfect; F – conchoidal; M – short prismatic to tabular crystals, columnar to earthy aggregates, crusts, coatings, granular, massive; R – soluble in water.
Origin and occurrence: Sedimentary in boron deposits, associated with other borates and halite. Prismatic crystals up to 150 mm long are known from Borax Lake, also Searles Lake and Boron, California, USA. It also occurs in Kirka, Turkey.
Application: chemical industry, B ore.

Ulexite

$NaCaB_5O_6(OH)_6 \cdot 5\ H_2O$

TRICLINIC ● ● ●

Properties: C – white, colorless, light gray; S – white; L B vitreous to silky; D – transparent to translucent; DE – 2.0; H – 2.5; CL – perfect; F B uneven; M – elongated prismatic to acicular crystals, fibrous aggregates, crusts, nodules, granular, massive.
Origin and occurrence: Sedimentary in boron deposits, associated with other borates. Slabs up to 100 mm (4 in) thick consisting of fibrous aggregates occur in Boron, California; also found in Esmeralda, Nevada, USA; Emet, Turkey and Inder, Kazakhstan.
Application: chemical industry, B ore.

Colemanite

$Ca_2B_6O_{11} \cdot 5\ H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white, yellowish, light gray; S – white; L – vitreous; D – transparent to translucent; DE – 2.4; H – 4.5; CL – perfect; F B uneven; M – short prismatic and isometric crystals, granular, massive.
Origin and occurrence: Sedimentary in boron deposits, associated with other borates. Perfect crystals up to 200 mm (7⅞ in), come from Emet and Kirka, Turkey. Crystals up to 70 mm (2¾ in) long occur in Gower Gulch, Inyo, California; also known from Esmeralda, Nevada, USA and Inder, Kazakhstan.
Application: chemical industry, B ore.

Kernite, 100 mm, Boron, U.S.A.

Kernite

$Na_2B_4O_6(OH)_2 . 3 H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white, light gray; S – white; L – vitreous to dull, silky; D – transparent, translucent to opaque; DE – 1.9; H – 2.5-3; CL – perfect; F – uneven; M – isometric crystals, fibrous aggregates, granular, massive; R – soluble in water.

Rhodizite, 23 mm, Antsirabe, Madagascar

Hambergite, 64 mm, Pamir, Tadzhikistan

Origin and occurrence: Sedimentary in boron deposits, associated with borax and other borates. Platy aggregates and crystals up to 2.5 x 1 m (8 ft x 39 3/8 in) found in Boron, California, USA; also known from the Tincalayu mine, Salta province, Argentina.

Hambergite

$Be_2BO_3(OH,F)$

ORTHORHOMBIC ● ●

Properties: C – colorless, white, light gray; S – white; L – vitreous; D – transparent to translucent; DE – 2.4; H – 7.5; CL B perfect; F B uneven; M – tabular to prismatic crystals, granular.

Origin and occurrence: Magmatic in granitic and rarely also in alkaline pegmatites; hydrothermal in cavities within the pegmatites, associated with tourmaline, danburite and beryl. Tabular crystals up to 200 mm (7 7/8 in) from the Little Three mine, Ramona, California, USA. Crystals up to 110 mm (4 5/16 in) also known from several localities in Madagascar, e.g. Imalo and Anjanabonoina; also found in Hyakule, Nepal and Ctidruzice, Czech Republic .

Rhodizite

$(K,Cs)Al_4Be_4(B,Be)_{12}O_{28}$

CUBIC ● ●

Properties: C – colorless, white, yellow, light gray; S – white; L – vitreous to adamantine; D – transparent to translucent; DE – 3.5; H – 8.5; CL – imperfect; F – conchoidal to uneven; M – isometric crystals, granular.

Origin and occurrence: Magmatic in granitic pegmatites, associated with tourmaline, danburite and beryl. Cubic to tetrahedral crystals, reaching up to 30 mm (1 3/16 in) occur in several localities in

Madagascar, e.g. Sahatany and Antandrokomby; also known from the Animikie Red Ace pegmatite, Florence, Wisconsin, USA.

Boracite

$Mg_3B_7O_{13}Cl$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, yellowish, light gray, light to dark green; S – white; L – vitreous; D – transparent to translucent; DE – 3.0; H – 7-7.5; CL – none; F – conchoidal to uneven; M – isometric crystals, fibrous aggregates, granular, massive.

Origin and occurrence: Sedimentary in evaporitic deposits, together with halite, gypsum and anhydrite; metamorphic in metamorphosed evaporates. Crystals up to 15 mm (19/32 in) come from Alto Chapare, Cochabamba, Bolivia; crystals about 5 mm (3/16 in) in size, occur in Stassfurt, Hanover and Kahlberg, Germany. Crystals are also known from Choctaw, Louisiana, USA and the Boulby mine, North Yorkshire, England, UK.

Boracite, 12 mm x, Alto Chapare, Bolivia

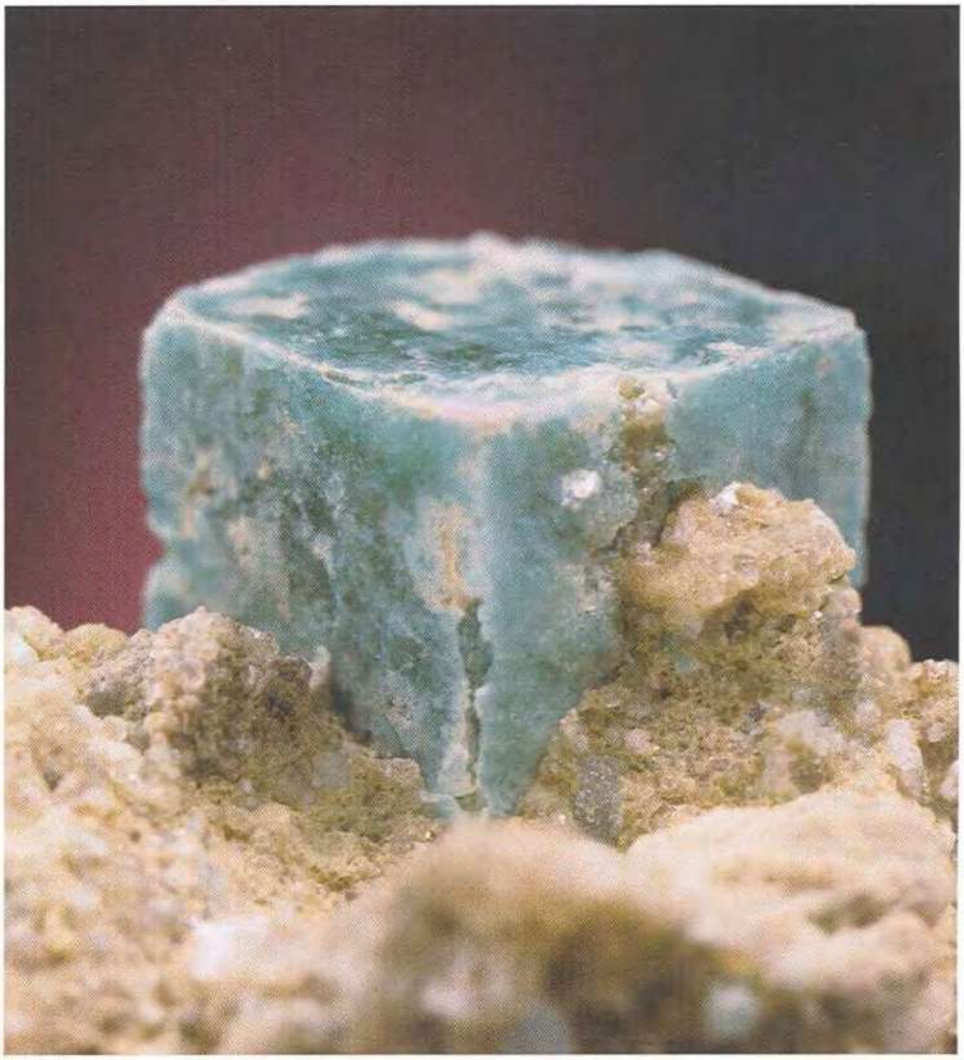

7. Sulfates

Anglesite
$PbSO_4$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, yellowish, gray, greenish; S – white; L – adamantine to greasy; D – transparent to translucent; DE – 6.4; H – 2.5 – 3; CL – dobra; F – conchoidal; M – thick tabular crystals, massive, LU – sometimes yellowish.
Origin and occurrence: Secondary, as a result of the galena oxidation, together with cerussite and other minerals. Beautiful yellowish crystals up to 100 mm (4 in) in size come from Touissit, Morocco. Crystals from Tsumeb, Namibia, reached up to 40 mm (1 9/16 in). Large crystals are also reported from Phoenixville, Pennsylvania, USA. Crystals up to 20 mm (25/32 in) occur in cavities in weathered galena in Sardinia, Italy. Crystals, reaching up to 40 mm (1 9/16 in) also found in Mezica . Slovenia. Prismatic crystals up to 75 mm (3 in) long known from the Bunker Hill mine, Idaho, USA.

Barite, 17 mm xx, Pöhla, Germany
Anhydrite, 55 mm, Simplon Tunnel, Switzerland

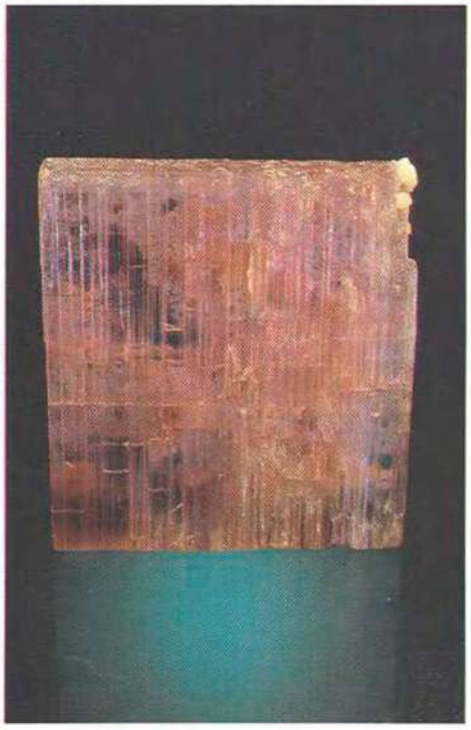

Anglesite, 81 mm, Touissit, Morocco

Anhydrite
$CaSO_4$

ORTHORHOMBIC ● ● ● ●

Properties: C – colorless, white, bluish, purplish, red, brown; S – white; L – pearly to vitreous; PS – transparent to translucent; DE – 2.8-3; H – 3-3.5; CL – good; F – splintery to uneven; M – isometric and prismatic crystals, granular, massive.
Origin and occurrence: Mostly sedimentary, as a result of the evaporation of sea water, associated with gypsum, calcite and other minerals; rare in pegmatites and hydrothermal. It is very common in form of massive aggregates in salt deposits in Stassfurt and Wathlingen, Germany, where small crystals also occur. Folded layers in clays known from Wieliczka, Poland. Purplish crystals up to 20 mm (25/32 in) found in cracks in metamorphic rocks near Simplon and St. Gotthard, Switzerland. Fine druses of bluish crystals up to 200 mm (7 7/8 in) long found in Naica, Chihuahua, Mexico.

Celestine, 270 mm, Madagascar

Celestite

$SrSO_4$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, yellowish, blue; S – white; L – vitreous; D – transparent to translucent; DE – 4; H – 3-3.5; CL – perfect; F – uneven; M – prismatic and tabular crystals, columnar aggregates, concretions.

Origin and occurrence: Rare hydrothermal, mainly sedimentary, together with halite, anhydrite and gypsum. Beautiful blue crystals several cm long occur in cavities of concretions in the Sakoany mine, Madagascar. Prismatic crystals up to 100 mm (4 in) long in sulfur deposits in Poland (Tarnobrzeg) and Italy (Caltanissetta). Fine blue tabular crystals associated with aragonite crystals found in _pania Dolina, Slovakia. Fine crystals are known from marls in Tunisia and Libya.

Barite

$BaSO_4$

ORTHORHOMBIC ● ● ● ●

Properties: C – white, yellow, blue, red, brown, black; S – white; L – vitreous; D – transparent to translucent; DE – 4.3-4.7; H – 3.5; CL – perfect; F – conchoidal to splintery; M – tabular to prismatic crystals, massive; LU – sometimes blue.

Origin and occurrence: Hydrothermal, it originates under medium and low temperatures, also sedimentary, together with fluorite, calcite, cinnabar and other minerals. Beautiful barite crystals up to 50 mm (2 in) found in hydrothermal veins in Pribram, Czech Republic. Blue tabular crystals up to 100 mm (4 in) from Dedova hora, Czech Republic. Beautiful tabular crystals up to 100 mm (4 in) known from Banska Stiavnica, Slovakia; Cavnic and Baia Spric, Romania. Very fine crystals of up to 200 mm (7⅞ in) occurred in Alston Moor, Frizington and Mowbray, Cumbria, UK. Beautiful druses of honey-brown crystals up to 70 mm (2¾ in) long found in Pohla, Germany. Fine druses of prismatic crystals known from Elk Creek, South Dakota, USA; so called

Antlerite, 60 mm, Chuquicamata, Chile

Barite, 100 mm xx, Stoneham Co., U.S.A.

desert roses (crystals with inclusion of sand grains) originating in desert climate occur in the vicinity of Norman, Oklahoma, USA. Shiny crystals come from Freiberg and Halsbrucke, Germany.
Application: an ingredient of drilling fluids, chemical, glass, paper and rubber industries.

Antlerite
$Cu_5(SO_4)_2(OH)_6 \cdot 3\,H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – green; S – light green; L – vitreous; D – translucent; DE – 3.9; H – 3.5; CL – perfect; F – conchoidal to uneven; M – short prismatic and acicular crystals, crusts and earthy aggregates.
Origin and occurrence: Secondary, resulting from weathering of Cu ores, associated with other secondary Cu minerals. It is a principal mineral of the oxidation zone in Chuquicamata, Chile, where it forms crystals up to 5 mm (3/16 in). Acicular crystals, up to 20 mm (25/32 in) long, occurred in Bisbee, Arizona, USA. Crystalline crusts are known from Spania Dolina and Picsky, Slovakia.

Brochantite, 14 mm aggregate, Bingham, U.S.A.

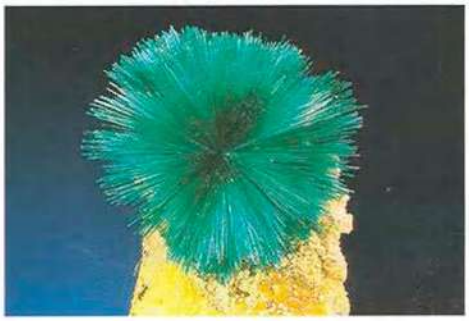

Brochantite
$Cu_4(SO_4)(OH)_6$

MONOCLINIC ● ● ●

Properties: C – green; S – light green; L – vitreous, in cleavage planes pearly; D – translucent to transparent; DE – 4; H – 3.5-4; CL – good; F – conchoidal to uneven; M – long prismatic to acicular crystals, granular.
Origin and occurrence: Secondary, as a result of the oxidation of Cu ores usually in the arid climate, together with other secondary Cu minerals. Crystals up to 70 mm (2¾ in) long found in Bisbee, together with pseudo-morphs of malachite after brochantite. Crystals are also known from Tsumeb, Namibia. Prismatic crystals up to 50 mm (2 in) long from Cerro Verde, Peru. Emerald-green crusts occur in L'ubietova, Slovakia.

Linarite, 23 mm xx, Graham Co., U.S.A.

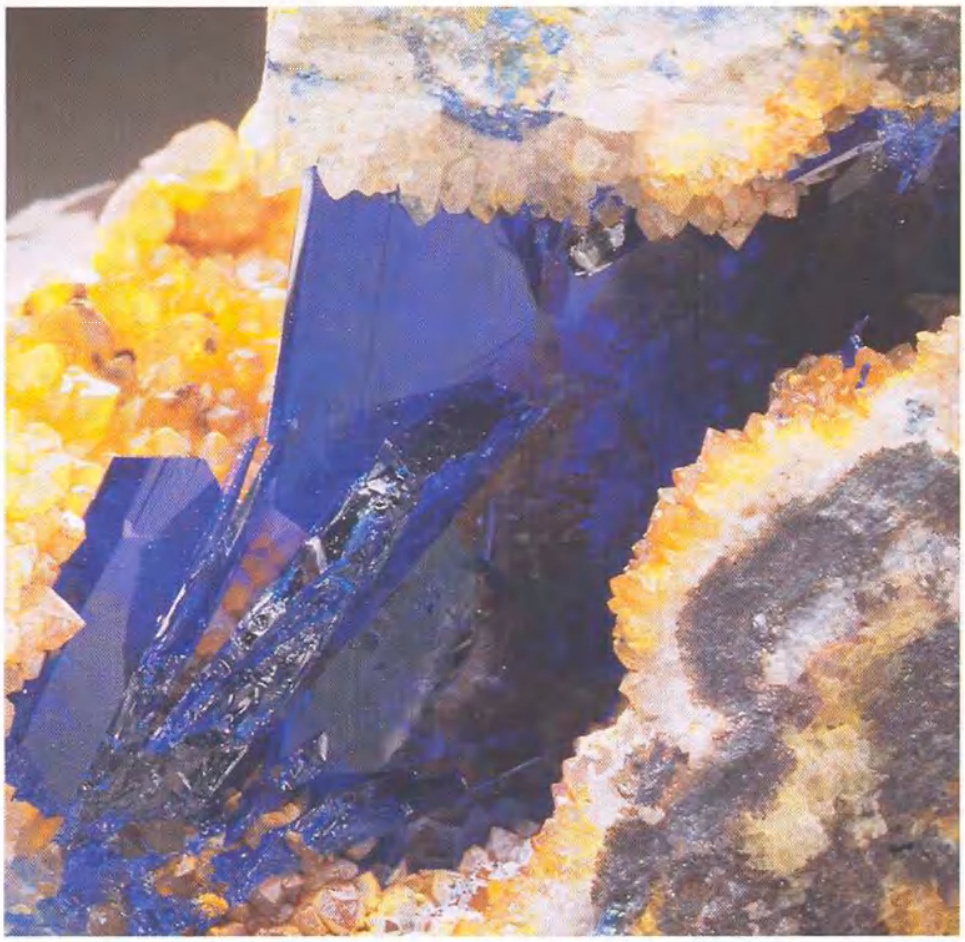

Linarite

$PbCu(SO_4)(OH)_2$

MONOCLINIC ● ●

Properties: C – azure-blue; S – light blue; L – vitreous to adamantine; D – translucent; DE – 5.3-5.5; H – 2.5; CL – good; F – conchoidal; M – tabular and prismatic crystals, coatings.
Origin and occurrence: Secondary, as a result of the oxidation of Cu and Pb sulfides at low pH. The largest crystals, reaching up to 80 mm (3 1/8 in), come from the Mammoth mine, Tiger, Arizona, USA. Crystals from the Grand Reef mine in Arizona up to 50 mm (2 in) long. Very fine crystals, up to 30 mm (1 3/16 in) were found in Keswick, Cumbria, England, UK. Large crystals were also described from Tsumeb, Namibia and from Broken Hill, New South Wales, Australia.

Alunite

$KAl_3(SO_4)_2(OH)_6$

TRIGONAL ● ● ●

Properties: C – white, yellowish, gray; S – white; L – vitreous to pearly; D – transparent to opaque; DE – 2.8; H – 3.5-4; CL – perfect; F – conchoidal; M – rhombohedral crystals, porous aggregates, granular.
Origin and occurrence: Secondary, as a result of reactions of sulfuric acid with Al-rich rocks, associated with gypsum. Small crystals are known in Berehovo, Ukraine. Massive aggregates occur together with turquoise in Cu deposits in Arizona, USA. (New Cornelia mine and others). Huge alunite deposits, like Zaglik, Azerbaijan, are mined for Al.
Application: Al ore.

Alunite, 1 mm xx, Nagyegyháza, Hungary

Natrojarosite, 50 mm, Laurion, Greece

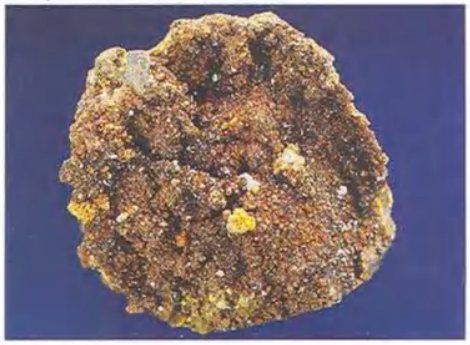

Natrojarosite
$NaFe^{3+}{}_3(SO_4)_2(OH)_6$

TRIGONAL ● ●

Properties: C – yellow, brown; S – light yellow; L – vitreous; D – transparent to translucent; DE – 3.2; H – 3; CL – good; F – conchoidal; M – tabular to rhombohedral crystals, earthy aggregates..
Origin and occurrence: Secondary, as a result of weathering Fe sulfides, associated with fibroferrite, alunite and other minerals. It occurs in Modum, Norway; Soda Springs Valley, Nevada, USA; Chuquicamata, Chile and elsewhere.

Jarosite
$KFe^{3+}{}_3(SO_4)_2(OH)_6$

TRIGONAL ● ● ●

Properties: C – yellow; S – light yellow; L – adamantine to dull; D – translucent; DE – 2.9-3.3; H – 2.5-3.5; CL – good; F – conchoidal to uneven; M – rhombohedral to tabular crystals, granular, massive.
Origin and occurrence: : Secondary, as a result of weathering Fe sulfides, usually associated with natrojarosite and other sulfates. The world's best specimens come from Pena Blanca Uranium mine, near Aldama, Chihuahua, Mexico, where crystals, up to 20 mm (25/32 in) were found. Crystals, up to 10 mm (3/8 in), were found in Tombstone, Arizona, USA. Crystals, together with pseudo-morphs of jarosite after alunite, occurred in Chuquicamata, Chile. Tabular crystals also were described from Horni Slavkov, Czech Republic.

Jarosite, 2 mm xx, Gyongyösoroszi, Hungary

Beudantite
$PbFe^{3+}{}_3(AsO_4)(SO_4)(OH)_6$

TRIGONAL ● ●

Properties: C – dark green, brown, red-brown; S – greenish, gray-yellow; L – vitreous to resinous; D – transparent to translucent; DE – 4; H – 3.5-4.5; CL – good; M – rhombohedral crystals, crusts.
Origin and occurrence: Secondary, occurring in the oxidation zone of Pb deposits, together with scorodite and other minerals. Crystals, several mm in size, found in Tsumeb, Namibia, are the best in the world. Small crystals are known from Bisbee, Arizona, USA; Ashburton Downs, Western Australia, Australia and Kamarcza near Laurion, Greece.

Beudantite, 70 mm, Moldava, Czech Republic

Hanksite
$Na_{22}K(SO_4)_9(CO_3)_2Cl$

HEXAGONAL ● ●

Properties: : C – colorless, yellowish; S – colorless; L – vitreous to dull; D – transparent to translucent; DE – 2.6; H – 3.5; CL – good; F – uneven; M – tabular to short prismatic crystals.
Origin and occurrence: Sedimentary in salt lake sediments, together with halite, borax and other minerals. Crystals up to 200 mm (7⅞ in) found in Searles Lake, California, USA. It is also known from borax deposits in the Death Valley, California, USA.

Caledonite
$Cu_2Pb_5(SO_4)_3(CO_3)(OH)_6$

ORTHORHOMBIC ● ●

Properties: C – green to blue-green; S – light green; L – resinous; D – transparent to translucent; DE – 5.6; H – 2.5-3; CL – perfect; F – uneven; M – prismatic crystals, radial aggregates.

Origin and occurrence: Secondary, occurring in the oxidation zone of Cu and Pb deposits, associated with linarite and other minerals. Crystals up to 20 mm (25/32 in) come from Leadhills, Scotland, UK. Crystals, up to 15 mm (19/32 in), were found in the Mammoth mine, Tiger, Arizona, USA. Rich druses are known from Anarak, Iran. Crystals reaching up to 20 mm (25/32 in) occurred in the Blue Bell mine, California, USA.

Hanksite, 50 mm, Searles Lake, U.S.A.

Leadhillite
$Pb_4(SO_4)(CO_3)_2(OH)_2$

MONOCLINIC ● ●

Properties: C – colorless, yellowish, gray; S – colorless; L – resinous to adamantine; D – transparent to translucent; DE – 6.5; H – 2.5-3; CL – perfect; F – conchoidal; M – pseudo-hexagonal tabular crystals.
Origin and occurrence: Secondary mineral from the oxidation zone of Pb deposits, associated with cerussite, anglesite, linarite and other minerals. The largest crystals, measuring up to 15 cm, come from Tsumeb, Namibia. Crystals up to 25 mm (1 in) found in the Mammoth mine, Tiger, Arizona, USA. Crystals in Leadhills, Scotland, UK were of similar size.

Chalcantite
$CuSO_4 . 5 H_2O$

TRICLINIC ● ● ●

Properties: C – deep blue; S – white; L – vitreous to resinous; D – transparent to translucent; DE – 2.3; H – 2.5; CL – imperfect; F – conchoidal; M – short prismatic to tabular crystals, stalactites, films; R – soluble in water.

Caledonite, 5 mm x, Graham Co., U.S.A.

Leadhillite, 35 mm, Leadhill, UK

Chalcanthite, 30 mm, Pima Co., U.S.A.

Origin and occurrence: Secondary, as a product of the oxidation of Cu sulfides, stable only in arid climate. Crystals up to 40 mm (1 3/16 in) in size, and stalactites, up to 1 m (39 3/8 in) long are known from Bisbee, Arizona, USA. Crystals were also found in the oxidation zone in Chuquicamata, Chile. Beautiful stalactites come from Rio Tinto, Spain.
Application: rarely mined as Cu ore.

Melanterite, 80 mm, Harz Mts., Germany

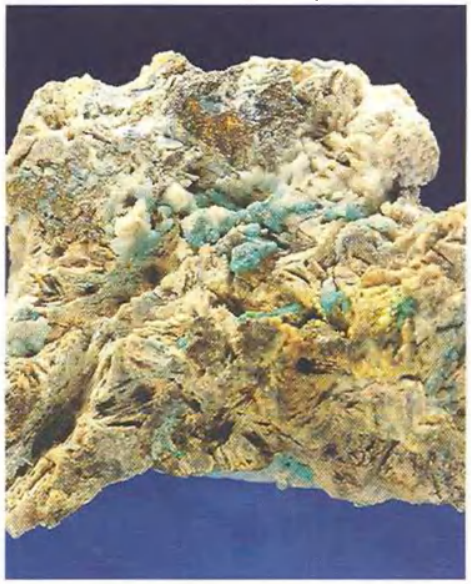

Melanterite

$FeSO_4 . 7 H_2O$

MONOCLINIC ● ● ● ●

Varieties: pisanite (with Cu contents), kirovite (with Mg contents)

Properties: C – light green; S – colorless; L – vitreous; D – translucent; DE – 1.9; H – 2; CL – perfect; F – conchoidal; M – granular crusts, botryoidal and stalactitic aggregates, films; R – soluble in water.
Origin and occurrence: Secondary, as a result of the oxidation of Fe sulfides, together with other sulfates. It is unstable under atmospheric conditions. Crystals up to 20 mm (25/32 in) long are known from Bisbee, Arizona and from the Boyd mine, Tennessee, USA. Stalactites, up to 50 cm (20 in) long, occurred in Chvalctice, Czech Republic. It is also common in Rio Tinto, Spain and Banska Stiavnica, Slovakia.

Epsomite
$MgSO_4 . 7 H_2O$

ORTHORHOMBIC ● ● ● ●

Properties: C – colorless to white; S – colorless; L – vitreous to silky; D – transparent to translucent; DE – 1.7; H – 2; CL – perfect; F – conchoidal; M – small crystals, granular, stalactitic aggregates; R – soluble in water, decomposing under atmospheric conditions.
Origin and occurrence: Secondary, as a result of the oxidation of Fe sulfides, also from crystallization of the salt lake water, associated with halite and other minerals. Crystals up to 1 m (39⅜ in) long were found on Mount Kruger, Washington; prismatic crystals up to 50 mm (2 in) long also occur in Bisbee, Arizona and needles of a similar size come from the White Caps mine, Nevada, USA. It is also known as a product of the activity of fumaroles in Mount Vesuvius, Italy.

Morenosite
$NiSO_4 . 7 H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – green, green-white; S – greenish; L – vitreous; D – transparent to translucent; DE – 2; H – 2-2.5; CL – good; F – conchoidal; M – stalactitic crusts and efflorescences; R – soluble in water.
Origin and occurrence: Secondary, as a product of the oxidation of Ni minerals, stable mainly in the arid regions. It is very common in Sudbury, Ontario, Canada; also in Richelsdorf, Germany and Potucky, Czech Republic.

Coquimbite
$Fe^{3+}_2(SO_4)_3 . 9 H_2O$

TRIGONAL ● ● ●

Properties: C – purple, yellow, green, colorless; S – colorless; L – vitreous; D – transparent; DE – 2.1; H – 2; CL – imperfect; F – conchoidal to uneven; M – prismatic to tabular crystals, granular, massive; R – soluble in water.
Origin and occurrence: Secondary, associated with other sulfates. Crystals, several cm long, were found in the Dexter No.7 mine, Calf Mesa, Utah, USA. Small prismatic crystals are known from Zelezník, tabular crystals occurred in Banska Stiavnica, Slovakia. It is very common in Chilean deposits, like Chuquicamata and Tierra Amarilla.

Alunogen
$Al_2(SO_4)_3 . 17 H_2O$

TRICLINIC ● ● ●

Properties: C – colorless, white, gray-yellow; S – colorless; L – silky; PS – transparent; DE – 1.8; H – 1.5; CL – perfect; M – pseudo-hexagonal crystals, granular; R – soluble in water.
Origin and occurrence: Secondary, as a result of the pyrite oxidation, also a product of sublimation on volcanoes and burning dumps, associated with other sulfates. Large crystals are known from the Dexter No.7 mine, Calf Mesa, Utah, crusts, over 1 m (39⅜ in) thick were found on Mount Alum, New Mexico,

Epsomite, 10 mm xx, Stassfurt, Germany

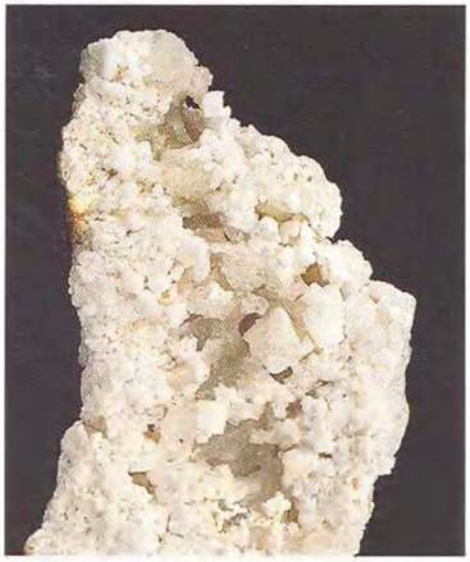

Morenosite, 60 mm, Potucky, Czech Republic

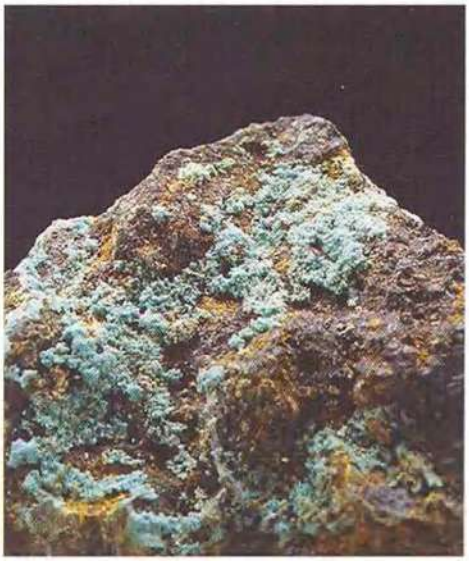

Coquimbite, 110 mm, Cerritos Bajos, Chile

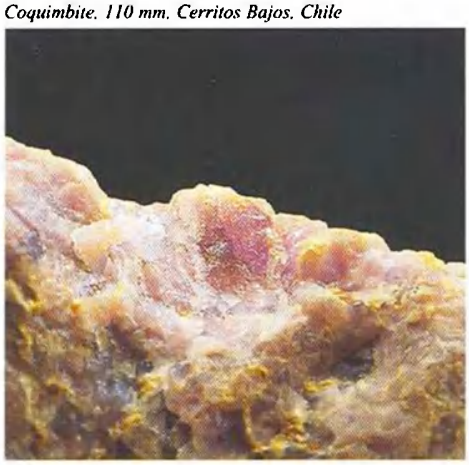

Alunogen, 80 mm, Dubnik, Slovakia

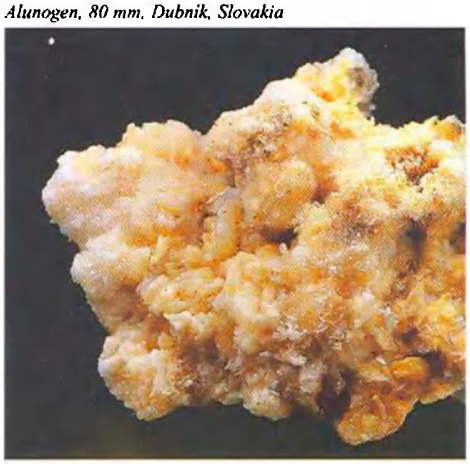

USA. Fine aggregates occur on the walls of underground workings in the old opal mines in Dubník, Slovakia.

Halotrichite

$FeAl_2(SO_4)_4 \cdot 22\ H_2O$

MONOCLINIC ● ● ●

Properties: C – white, greenish; S – colorless; L – vitreous; PS – transparent to translucent; DE – 1.9-2.1; H – 1.5-2; CL – imperfect; F – conchoidal; M – acicular crystals, fibrous aggregates, efflorescences; R – soluble in water.
Origin and occurrence: Secondary, as a result of the pyrite oxidation, also a product of the activity of hot springs ans solfataras, associated with other sulfates. Common fibrous crusts occur in Dubnik, Slovakia. It looks similar in Rio Marina, Elba, Italy and in Chuquicamata, Chile. It is known from solfataras in Pozzuoli, Italy. It is a product of hot springs activity in the Lassen Peak National Park, California, USA.

Halotrichite, 90 mm, Recsk, Hungary

Apjohnite

$MnAl_2(SO_4)_4 \cdot 22\ H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white, pink, greenish, yellow; S – colorless; L – silky; D – transparent; DE – 1.9; H – 1.5; M – fibrous aggregates, coatings, massive; R – soluble in water.
Origin and occurrence: Secondary, occurring together with other sulfates. Large accumulations are known from Little Pigeon Creek, Alum Cave, Tennessee, USA. It also occurs in Delagona Bay, Mozambique.

Apjohnite, 60 mm, Smolnik, Slovakia

Bilinite, 60 mm, Bilina, Czech Republic

Bilinite

$Fe^{2+}Fe^{3+}{}_2(SO_4)_4 \cdot 22\ H_2O$

MONOCLINIC ● ●

Properties: C – white to wellowish; S – white; L – silky; D – opaque; DE – 1.9; H – 2; M – fibrous aggregates; R – soluble in water.
Origin and occurrence: Secondary, as a result of the pyrite oxidation, associated with other sulfates. Fibrous aggregates are known from Svitec near Bilina, Czech Republic. Also described from Bisbee, Arizona, USA.

Tschermigite

$(NH_4)Al(SO_4)_2 \cdot 12\ H_2O$

CUBIC ● ● ●

Properties: : C – colorless; S – colorless; L – vitreous; D – transparent; DE – 1.6; H – 1.5; F – conchoidal; M – octahedral crystals, fibrous and columnar aggregates; R – soluble in water.

Tschermigite, 50 mm, Mogyororosbanya, Hungary

Origin and occurrence: Secondary in the brown coal, also on burning dumps and as a product of solfataras. Skeletal crystals up to 10 mm (3/8 in) in size are known from the brown coal basin near Most and Bilina, Czech Republic. Original locality Eermniky, Czech Republic yielded fibrous crusts, up to 50 mm (2 in) thick. Fine octahedra, up to 10 mm (3/8 in) come from burning dumps near Zastavka near Brno, Czech Republic. Octahedra occur also in Geysers, California, USA. Small crystals on Etna, Sicily, Italy are of volcanic origin.

Polyhalite

$KCa_2Mg(SO_4)_4 \cdot 2\ H_2O$

TRICLINIC ● ● ●

Properties: C – colorless, brown, red-brown; S – colorless; L – vitreous; D – transparent to translucent; DE – 2.8; H – 3-3.5; CL – perfect; M – acicular to prismatic crystals, columnar, scaly and fibrous aggregates.
Origin and occurrence: Sedimentary, as a constituent of salt deposits, also as a product of volcanic activity. Tabular and prismatic crystals are found rarely together with massive polyhalite aggregates in Stassfurt, Germany. Fibrous aggregates come from Halstatt, Austria. Large deposits of polyhalite are located near Carlsbad, New Mexico, USA. Coatings occur on Mount Vesuvius, Italy.

Görgeyite

$K_2Ca_5(SO_4)_6 \cdot H_2O$

MONOCLINIC ● ●

Properties: C – colorless, yellowish; S – colorless; L – vitreous; D – translucent; DE – 3; H – 3.5;

Polyhalite, 54 mm, Carlsbad, U.S.A.

Gypsum, 62 mm, Carneville, Utah, U.S.A.

CL – good; F – conchoidal; M – tabular crystals.
Origin and occurrence: Sedimentary in salt deposits, associated with halite and other minerals. Fine crystals up to 80 mm (3⅛ in) long found near Inder Lake, Kazakhstan; also known from Bad Ischl, Austria.

Gypsum

$CaSO_4 . 2 H_2O$

MONOCLINIC ● ● ● ● ●

Varieties: Maria-glass, alabaster, satin spar, selenite.
Properties: C – colorless, white, gray, yellowish; S – colorless; L – vitreous to pearly; D – transparent to translucent; DE – 2.3; H – 1.5-2; CL – perfect; F – conchoidal; M – typical monoclinic crystals, often twins, fibrous and platy aggregates, granular, massive; LU – crystals with inclusions sometimes bluish to yellowish.

Origin and occurrence: Rare primary hydrothermal, mostly sedimentary and as a weathering product, associated with anhydrite, halite and other minerals. Crystals up to 50 mm from Cavnic, Romania are probably hydrothermal. Huge crystals up to 1.5 m (5 ft) long found in karst cavities (Cave of Swords) in gossan in the PbBZn deposit Naica, Chihuahua, Mexico. Crystals up to 9 m (29 ft 6 in) long from Santa Eulalia, Chihuahua, Mexico, where they were found combined with interesting aggregates, called 'ram's horns' from their shape. Crystals up to 4 m (13 ft) long occurred in Tarnobrzeg, Poland. Common crystals are also known from Gorguel, Spain. The 'desert roses' or crystal druses, from Sahara desert in Tunisia and Algeria, with sand grain inclusions are mineralogically interesting. Large slabs of the transparent variety, called Maria-glass, found in Friedrichsrode, Germany. Fine-grained variety alabaster occurs e.g. in Italy. Fibrous variety selenite (sometimes also called satin spar) comes from the Sylva river basin, Perm, Russia.
Application: building, chemical and medical industries.

Langite

$Cu_4(SO_4)(OH)_6 . 2 H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – blue, blue-green; S – light blue; L – vitreous to silky; D – translucent; DE – 3.3; H – 2.5-3; CL – perfect; M – small isometric crystals, fine-grained crusts, earthy aggregates..
Origin and occurrence: Secondary, as a product of the oxidation of Cu ores, associated with gypsum and other minerals. Crystals known from St. Just, Cornwall, UK. Crystals occurred in Tsumeb, Namibia. Fine specimens come from Spania Dolina and L'ubietova, Slovakia and from Borovec, Czech Republic. It is also common in Ely, Nevada, USA and in El Cobre, Chile.

Gorgevite, 80 mm, Inder Lake, Kazakhstan

Langite, 49 mm, Allihies, Cork, Ireland

Fibroferrite, 70 mm, Dubník, Slovakia

Fibroferrite
$Fe^{3+}(SO_4)(OH) \cdot 5\,H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – yellowish, greenish, gray; S – white; L – silky; D – opaque; DE – 1.9; H – 2.5; CL – perfect; M – small crystals, fibrous and botryoidal crusts; R – soluble in water.
Origin and occurrence: Secondary as a product of the pyrite oxidation, together with other sulfates. Crystals in cavities in melanterite found in the Dexter No.7 mine, Calf Mesa, Utah, USA. Vein fillings up to 3 m (10 ft) thick come from the Santa Elena mine, La Alcaparrosa, Argentina. Fine fibrous aggregates occur in Dubnik, Slovakia. It is common in many localities in Chile (Tierra Amarilla, Chuquicamata).

Aluminite
$Al_2(SO_4)(OH)_4 \cdot 7\,H_2O$

MONOCLINIC ● ● ●

Properties: C – white; S – white; L – dull; D – opaque; DE – 1.7-1.8; H – 1-2; CL – none; M – earthy nodules, consisting of microscopic needles.
Origin and occurrence: Secondary, resulting from the reaction of sulfuric acid with Al in Al-rich rocks. Large nodule, 30 cm (12 in) in diameter, known from Newhaven, Sussex, UK. Fine white nodules described from Mala Chuchle, Prague, Czech Republic. It also occurs near Halle, Germany and covers limestone in the vicinity of Joplin, Missouri, USA.

Aluminite, 60 mm, Bicske-Csordakut, Hungary

Botryogen
$MgFe^{3+}(SO_4)_2(OH) \cdot 7\,H_2O$

MONOCLINIC ● ● ●

Properties: C – orange-red; S – yellow; L – vitreous; D – transparent to translucent; DE – 2.1; H – 2-2.5; CL – perfect; F – conchoidal; M – prismatic striated crystals, racemose and spherical aggregates with radial structure.
Origin and occurrence: Secondary, as a result of the pyrite oxidation in arid regions, together with other sulfates. Crystals up to 35 mm ($1^5/_{16}$ in) long from the Libiola mine near Genoa, Italy. It is common in Chuquicamata and Quetena in Chile; also known from Rammelsberg, Germany.

Botryogen, 3 mm xx, Knoxville, U.S.A.

Devilline, 10 mm xx, Spania Dolina, Slovakia

Copiapite

$(Fe^{2+},Mg)Fe^{3+}{}_4(SO_4)_6(OH)_2 \cdot 20\ H_2O$

TRICLINIC ● ● ●

Properties: C – yellow, green-yellow, orange; S – yellow; L – pearly; D – transparent to translucent; DE – 2.1-2.2; H – 2.5; CL – perfect; M – thin tabular crystals, scaly and pulverulent aggregates, earthy.
Origin and occurrence: Secondary, originated from the pyrite oxidation, together with other sulfates.
Tabular crystals come from the Dexter No.7 mine, Calf Mesa, Utah, USA and also from Zeleznik , Slovakia. Fine crystals were found in many localities in Chile (Tierra Amarilla, Chuquicamata). It also occurs in Rammelsberg, Germany.

Copiapite, 96 mm, Copiapó, Chile

Serpierite, 80 mm, Pribram, Czech Republic

Devilline

$CaCu_4(SO_4)_2(OH)_6 \cdot 3\ H_2O$

MONOCLINIC ● ●

Properties: C – emerald-green; S – white to light green; L – vitreous to pearly; D – transparent to translucent; DE – 3.1; H – 2.5; CL – perfect; M – thin tabular pseudo-hexagonal crystals and coatings.
Origin and occurrence: Secondary, as a result of the oxidation of Cu sulfides, associated with other secondary Cu minerals. The world's best specimens come from Spania Dolina, Slovakia, where crystal rosettes up to 10 mm (⅜ in) in diameter were found in the past century; also known from Botallack, Cornwall, UK and Tsumeb, Namibia.

Serpierite

$Ca(Cu,Zn)_4(SO_4)_2(OH)_6 \cdot 3\ H_2O$

MONOCLINIC ● ●

Properties: C – blue; S – white; L – vitreous; D – transparent to translucent; DE – 3.1; H – 2.5; CL – perfect; M – tabular crystals, coatings.
Origin and occurrence: Secondary, originated in the oxidation zone of Cu-Zn deposits, together with smithsonite and other minerals. Its small crystals and aggregates occurred in Laurion, Greece. Recently confirmed at Pribram, Czech Republic.

Ettringite, 45 mm, Kuruman, South Africa

Sturmanite, 35 mm xx, Kuruman, South Africa

Ettringite

$Ca_6Al_2(SO_4)_3(OH)_{12} \cdot 26\ H_2O$

HEXAGONAL ● ●

Properties: C – colorless, yellow; S – white; L – vitreous to silky; D – transparent; DE – 1.8; H – 2.5; CL – perfect; M – prismatic crystals, fibrous aggregates.
Origin and occurrence: Metamorphic, associated with sturmanite. The world's best crystals reaching up to 100 mm (4 in) come from the N'Chwaning mine, Kuruman, South Africa. Crystals up to 4 mm (5/32 in) occurred in Franklin, New Jersey, USA. Also known from the contact metamorphic conditions in Crestmore, California, USA.

Sturmanite

$Ca_6Fe^{3+}{}_2(SO_4)_2[B(OH)_4].(OH)_{12} \cdot 25\ H_2O$

TRIGONAL ●

Properties: C – yellow, yellow-green; S – white; L – vitreous; D – transparent to translucent; DE – 1.9; H – 2.5; CL – perfect; M – flat dipyramidal crystals.
Origin and occurrence: : Probably metamorphic, associated with barite, hematite and ettringite. Crystals up to 140 mm (5½ in) long come from the N'Chwaning mine, Kuruman, South Africa.

Johannite

$Cu(UO_2)_2(SO_4)_2(OH)_2 \cdot 8\ H_2O$

TRICLINIC ● ●

Properties: C – green; S – light green; L – vitreous; D – transparent to translucent; DE – 3.3; H – 2-2.5; CL – good; M – prismatic to thick tabular crystals, scaly aggregates, coatings; R – radioactive.
Origin and occurrence: Secondary, as a result of the uraninite oxidation, together with other secondary U minerals. Crystals are known from Jachymov, Czech Republic and Johanngeorgenstadt, Germany. It was common with zippeite in Central City, Colorado, USA and also reported from Mounana, Gabon.

Zippeite

$K_4(UO_2)_6(SO_4)_3(OH)_{10} \cdot 4\ H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – orange-yellow; S – yellow; L – dull to earthy; D – opaque; DE – 3.7; H – not determined; CL – perfect; M – acicular and tabular crystals, pulverulent and acicular aggregates, coatings; LU – green; R – radioactive.
Origin and occurrence: Secondary, forming during the uraninite oxidation, associated with other secondary U minerals. Small tabular crystals come from Drmoul, Czech Republic. Coatings and

Johannite, 70 mm, Jachymov, Czech Republic

Zippeite, 40 mm, Lodeve, France

acicular aggregates are known from Jachymov, Czech Republic and Central City, Colorado, USA; also occurs in Shinkolobwe, Zaire.

Scheelite

$CaWO_4$

TETRAGONAL ● ● ● ●

Properties: C – colorless, gray-white, yellow-brown, orange, red, greenish; S – white; L – greasy to adamantine; D – translucent; DE – 6.1; H – 4.5-5; CL – good; F – conchoidal to uneven; M – pseudo-octahedral crystals, granular, massive; LU – blue-white.
Origin and occurrence: Magmatic in pegmatites, hydrothermal in greisens and metamorphic; parageneses vary significantly according to the origin. Beautiful brownish crystals up to 100 mm (4 in) in size come from Taewha and Tongwha in Korea. Similar crystals were recently found in China. Orange crystals up to 40 mm (1 9/16 in) in size are associated with cassiterite crystals on quartz crystals from Iultin and Tenkergin, Russia. Clear crystals, weighing up to 50 kg (110 lb), occurred in pegmatites near Natas, Namibia. Crystals up to 70 mm (2¾ in) reported from several mines near Traversella, Italy. Beautiful red crystals up to 20 mm (29/32 in) were very rare in Pøibram, Czech Republic. Yellow crystals up to 40 mm (1 9/16 in) are known from quartz veins with pumpellyite in Oboi Dsl, Czech Republic.
Application: W ore.

Scheelite, 77 mm, Sichuan, China

Stolzite, 20 mm, Arizona, U.S.A.

Stolzite
$PbWO_4$

TETRAGONAL ● ●

Properties: C – gray-brown, orange-yellow, red, green; S – colorless; L – adamantine to resinous; D – transparent to translucent; DE – 7.9-8.3; H – 2.5-3; CL – imperfect; F – conchoidal to uneven; M – dipyramidal and thick tabular, striated crystals.
Origin and occurrence: Secondary, as a product of the oxidation of primary W minerals. Crystals up to 60 mm (2 3/8 in) from St Leger-de-Peyre, France. Prisms and needles up to 25 mm (1 in) long found in Broken Hill, New South Wales, Australia. Crystals up to 25 mm (1 in) also known from Tsumeb, Namibia. Crystals up to 20 mm (25/32 in) reported from Cínovec, Czech Republic and from the Black Pine mine, Montana, USA.

Crocoite, 44 mm, Dundas, Australia

Crocoite
$PbCrO_4$

MONOCLINIC ● ●

Properties: C – orange, red; S – orange-yellow; L – adamantine to greasy; D – translucent; DE – 6; H – 3; CL – good; F – conchoidal to uneven; M – long prismatic to acicular crystals, crusts..
Origin and occurrence: Secondary, as a result of the galena oxidation in basic rocks. The world' best specimens come from the Dundas district, Tasmania, Australia where crystals up to 100 mm (4 in) long were found in several mines. Fine crystals up to 40 mm (1 9/16 in) long are known from Berezovsk, Ural mountains, Russia. Crystals up to 20 mm (25/32 in) were recently found in Callenberg, Germany.

Ferrimolybdite, 60 mm, Hurky, Czech Republic

Betpakdalite, 70 mm, Bayan Tsogdo, Mongolia

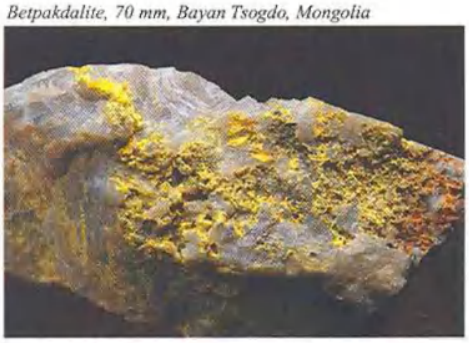

Ferrimolybdite
$Fe_2(MoO_4)_3 \cdot 8\ H_2O$

ORTHORHOMBIC ● ●

Properties: C – yellow, whitish; S – light yellow; L – adamantine, silky, earthy; D – opaque; DE – 4.4; H – 1-2; M – small acicular crystals, fibrous and radial aggregates, earthy.
Origin and occurrence: Secondary, as a product of the molybdenite oxidation. Microscopic crystals were found in Glen Innes, New South Wales, Australia. It occurred as yellow coatings in Climax and Telluride, Colorado, in the Getchell mine, Nevada, USA; also in Hsrky near Eista, Czech Republic.

Betpakdalite
$CaFe^{3+}H_8(MoO_4)_5(AsO_4)_2 \cdot 8\ H_2O$

MONOCLINIC ●

Properties: C – lemon-yellow; S – yellow; L – vitreous, waxy, dull; D – opaque; DE – 3; H – 3; CL – good; M – short prismatic microscopic crystals, pulverulent.
Origin and occurrence: Secondary mineral. Originally described from Kara-Oba, Kazakhstan. It also occurred in Tsumeb, Namibia and in Krupka, Czech Republic.

Wulfenite
$PbMoO_4$

TETRAGONAL ● ● ●

Properties: C – yellow, orange, brownish, red, greenish; S – white; L – greasy to adamantine; D – transparent to translucent; DE – 6.3-7; H – 3; CL – good; F – uneven to conchoidal; M – thin tabular and dipyramidal crystals, granular, massive.
Origin and occurrence: Secondary, as a result of the galena oxidation, together with cerussite, vanadinite and other minerals. The best specimens are known from the Red Cloud mine near Yuma, Arizona, where red tabular crystals found up to 50 mm (2 in) diameter. Yellow-brown crystals up to 100 mm (4 in) come from the Glove mine, Arizona, USA. Thick tabular orange crystals up to 20 mm ($^{25}/_{32}$ in) in diameter found in the Erupcion mine, Villa Ahumada, Los Lamentos, Chihuahua. Beautiful yellow plates up to 60 mm ($2^3/_8$ in) with orange mimetite spheres occurred in the San Francisco mine, Magdalena, Sonora, Mexico. Rare tabular crystals up to 70 mm (2¾ in) across come from Tsumeb, Namibia. Fine orange-yellow tabular crystals up to 20 mm ($^{25}/_{32}$ in) across, and pyramidal crystals are known from Bleiberg, Austria and Mezica , Slovenia. Crystals up to 100 mm (4 in) were found recently in Touissit, Morocco.

Wulfenite, 38 mm, Los Lamentos, Mexico

8. Phosphates

Lithiophosphate
Li_3PO_4

ORTHORHOMBIC ●

Properties: C – colorless, white, pinkish; S – white; L – vitreous; D – transparent to translucent; DE – 2.5; H – 4; CL – perfect; F – uneven; M – prismatic crystals, granular.
Origin and occurrence: Hydrothermal in granitic pegmatites, where it forms by replacement of spodumene and montebrasite. Cleavable aggregates up to 100 mm (4in) across occur in the Tanco Mine, Bernic Lake, Manitoba, Canada; crystals up to 25 mm (1 in) across in the Foote mine, Kings Mountain, North Carolina, USA.

Beryllonit
$NaBePO_4$

MONOCLINIC ● ●

Properties: C – colorless, white, yellowish; S – white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 5.5-6; CL – perfect; F – uneven; M – short prismatic to tabular crystals, granular.
Origin and occurrence: Hydrothermal in cavities in granitic pegmatites where it is associated with herderite, albite and tourmaline. Crystals and their twins, up to 150 mm (6 in) across come from Stoneham and Newry, Maine, USA; also from Viitaniemi, Finland and Paprok, Afghanistan.

Adamite, 52 mm, Mapimi, Mexico
Lithiophosphate, 30 mm, Tanco, Canada

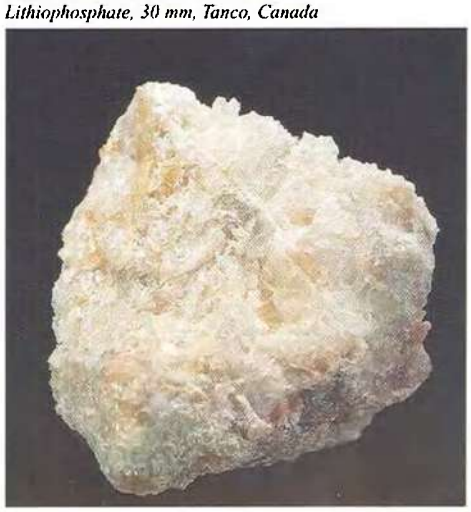

Beryllonite, 31 mm, Paprok, Afghanistan

Triphyllite
$LiFePO_4$

ORTHORHOMBIC ● ● ●

Properties: C – gray-green, gray-blue, gray, brown; S – gray-white; L – greasy to vitreous; D – transparent to translucent; DE – 3.4; H – 4-5; CL – good; F – uneven; M B short prismatic crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites, associated with graftonite, sarcopside and many secondary phosphates. Large triphyllite crystals up to 1.5 m (5 ft) across are known from Hagendorf, Germany; from the Tip Top mine, Custer, South Dakota and Palermo No. 1 mine, North Groton, New Hampshire, USA; also from Hühnerkobel, Germany.

Triphylite, 70 mm, Hagendorf, Germany

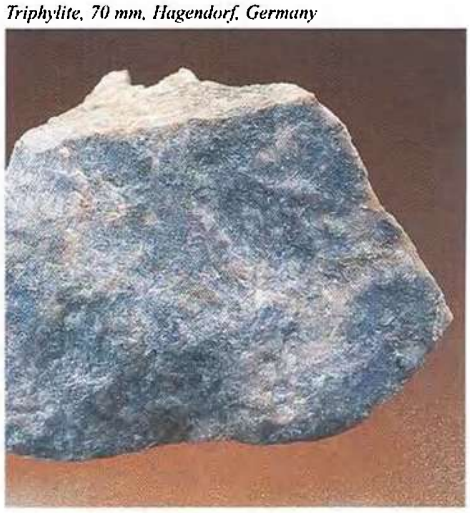

Lithiophilite, 120 mm, Viitaniemi, Finnland

Lithiofilit
$LiMnPO_4$

ORTHORHOMBIC ● ●

Properties: C – pink, red-brown, brown; S – white; L – vitreous; D – transparent to translucent; DE – 3.3; H – 4-5; CL – good; F – uneven; M – short prismatic crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites. It is sometimes associated with triplite and triphyllite, typically replaced by many secondary phosphates. Large masses of lithiophillite, reaching up to 1 m (39⅜ in), occur in Karibib, Namibia; Kitumbe, Rwanda. Other localities are Mangualde, Portugal; Tanco mine, Bernic Lake, Manitoba, Canada; Stewart Lithia Mine, Pala, California, USA.

Purpurite, 40 mm, Usakos, Namibia

Berzeliite, 120 mm, Langban, Sweden

Purpurite
$Mn^{3+}PO_4$

ORTHORHOMBIC ● ●

Properties: C – pink, purple, dark brown; S B red-purple; L - dull to velvet; D – translucent to opaque; DE – 3.7; H – 4-4.5; CL – good; F – uneven; M B granular, massive.
Origin and occurrence: Hydrothermal, as a product of lithiophillite replacement in granitic pegmatites. It is usually associated with many secondary phosphates. It is known from Kitumbe, Rwanda; Usakos and Sandamab, Namibia; the Tip Top and Bull Moose mines, Custer, South Dakota and Branchville, Connecticut, USA.

Berzeliite
$(Ca,Na)_3\ (Mg,Mn)_2\ (AsO_4)_3$

CUBIC ● ●

Properties: C – yellow, orange; S – red-purple; L – resinous; D – transparent to translucent; DE – 4.1; H – 4-4.5; CL – none; F – conchoidal to uneven; M – isometric crystals, granular, massive.
Origin and occurrence: Metamorphic, together with haussmanite, rhodonite and tephroite. It occurs as massive in Langban and Nordmark, Sweden.

Whitlockite
$Ca_9(Mg,Fe)H(PO_4)_7$

TRIGONAL ● ●

Properties: C – colorless, white, yellowish, pinkish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.1; H – 5; CL – none; F – conchoidal to uneven; M – rhombohedral crystals, granular, massive.

Origin and occurrence: Hydrothermal, as a product of replacement of primary phosphates in granitic pegmatites, rare in sedimentary rocks – phosphorites: very rare magmatic in meteorites. Mainly associated with apatite and carbonates. Imperfect crystals around 10 mm (3/8 in) across, occur in the Palermo No. 1 mine, North Groton, New Hampshire; also in the Tip Top mine, Custer, South Dakota, USA.

Whitlockite, 10 mm xx, New Hampshire, U.S.A.

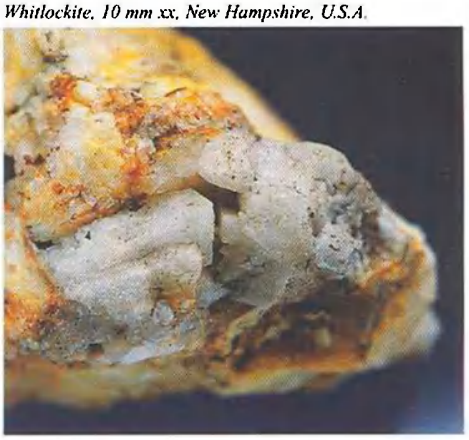

Xenotime-(Y)

YPO_4

TETRAGONAL ● ● ●

Properties: C – brown, yellow, gray, greenish; S – white; L – vitreous to resinous; D – transparent, translucent to opaque; DE – 4.5; H – 4.5; CL – good; F – conchoidal to uneven; M – long prismatic to tabular crystals, granular; R – sometimes weakly radioactive and metamict.

Origin and occurrence: Magmatic in granitic and alkaline pegmatites, granites and syenites; hydrothermal in the Alpine-type veins; metamorphic in gneisses; common in placers. It is associated with monazite-(Ce) and zircon. Perfect prismatic crystals up to 100 mm (4in) across occur mainly in pegmatites in Kragero and Hitterö, Norway; Ytterby, Sweden; in several places in Madagascar; in Ichikawa, Japan. Crystals about 20 mm (25/32 in) are known from the Alpine-type veins in Binntal, Switzerland.

Application: ore of rare earth elements.

Xenotime-(Y), 14 mm x, Washington Co., U.S.A.

Monazite-(Ce)
$CePO_4$

MONOCLINIC ● ● ●

Properties: C – yellow, brown, red-brown, orange, gray-green; S – white; L – vitreous to resinous; D – transparent, translucent to opaque; DE – 4.6; H – 5-5.5; CL – good; F – conchoidal to uneven; M – long prismatic to tabular crystals, granular; R – sometimes weakly radioactive and metamict.
Origin and occurrence: Magmatic in granitic and alkaline pegmatites, granites, syenites, carbonatites; hydrothermal in the Alpine-type veins and greisens; metamorphic in gneisses; common in placers.
It is associated with apatite, xenotime-(Y) and zircon. Perfect prismatic crystals up to 200 mm (7⅞ in) come from Mars Hill, North Carolina, Trout Creek Pass, Colorado, USA; also from Arendal, Norway; Ambatofotsikely and Ampangabe, Madagascar, where masses weighing several kg are common; crystals, up to 200 mm (7⅞ in) across found in Minas Gerais, Jaguaraçu, Brazil.
Application: ore of rare earth elements.

Monazite-(Ce), 31 mm, Felicio, Brazil

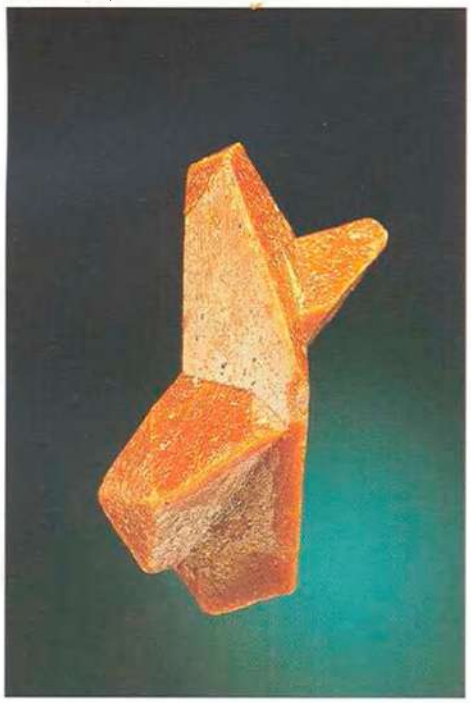

Hydroxylherderite
$CaBe(PO_4)(OH,F)$

MONOCLINIC ● ●

Properties: C – colorless, white, yellowish, greenish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.0; H – 5-5.5; CL – good; F – conchoidal to uneven; M – prismatic to tabular crystals, radial aggregates, granular.
Origin and occurrence: Hydrothermal in cavities in granitic pegmatites and in greisens. Perfect crystals, up to 120 mm (4 11/16 in) long come from Marilaca and together with colored tourmalines from Virgem da Lapa, Minas Gerais, Brazil; also known from Topsham and Stoneham, Maine, USA.

Amblygonite
$LiAl(PO_4)(F,OH)$

TRICLINIC ● ● ●

Properties: C – colorless, white, yellowish, bluish, gray; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.1; H – 5.5-6; CL – good; F – conchoidal to uneven; M – short prismatic crystals, granular, massive.

Hydroxylherderite, 22 mm, Anza, California, U.S.A.

Montebrasite, 80 mm, White Picacho, Arizona, U.S.A.

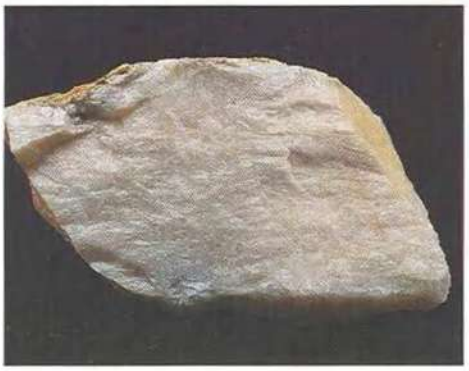

Origin and occurrence: Magmatic in granitic pegmatites and some granites; rare hydrothermal in greisens and in ore veins. Large amblygonite masses, several meters across, come from pegmatites, like the Beecher, Custer and Hugo mines, Keystone, South Dakota, USA, where its blocks weighed up to 200 tons; also known from Viitaniemi, Finland; Uto, Sweden. Typical representative of the quartz-amblygonite veins is Vernerov, Czech Republic.
Application: Li ore and raw material for ceramics.

Montebrasite

$LiAl(PO_4)(OH,F)$

TRICLINIC ● ● ●

Properties: C – colorless, white, yellowish, yellow, bluish, gray; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.0; H – 5.5-6; CL – good; F – conchoidal to uneven; M – short prismatic to tabular crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites and granites; hydrothermal in granitic pegmatites. It occurs very frequently with amblygonite, which it replaces. Large masses are known from Montebras, France; also from the Tin Mountain mine, Custer, South Dakota, USA.. Perfect tabular crystals up to 150 mm (6 in) across found in Taquaral, Minas Gerais, Brazil.

Wagnerite

$Mg_2(PO_4)(F,OH)$

MONOCLINIC ● ● ●

Properties: C – light yellow, yellow-green, yellow-brown, green; S – white; L – vitreous to greasy; D – transparent, translucent to opaque; DE – 3.2; H – 5-5.5; CL – imperfect; F – conchoidal to uneven; M – short prismatic crystals, granular, massive.

Wagnerite, 18 mm x, Werfen, Austria

Origin and occurrence: Magmatic in granitic pegmatites; metamorphic in gneisses and eclogites; hydrothermal in quartz veins and in salt deposits. The most famous finds come from the quartz veins in Hollgraben and Radelgraben, the Alps, Austria, where wagnerite occur together with lazulite and forms crystals up to 30 mm ($1^3/_{16}$ in) across; also known from Mangualde, Portugal and Bodenmais, Germany.

Amblygonite, 55 mm, Minas Gerais, Brazil

Zwieselite, 30 mm grain, Dolni Borv, Czech Republic

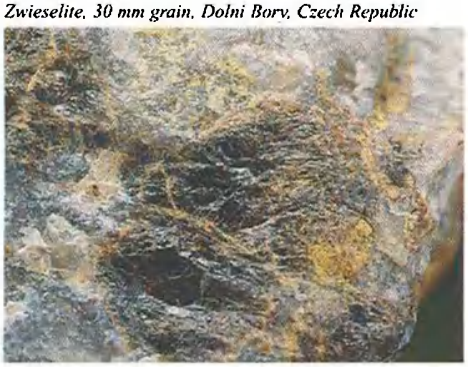

Zwieselite
$Fe_2(PO_4)(F,OH)$

MONOCLINIC ● ● ●

Properties: C – dark brown to black-brown; S – light brown; L – vitreous to greasy; D – translucent to opaque; DE – 4.0; H – 5-5.5; CL – imperfect; F – conchoidal to uneven; M – short prismatic crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites: rare hydrothermal in greisens. It is associated with apatite, triplite and secondary phosphates. Imperfect crystals and granular aggregates are known from pegmatites near Zwiesel, Germany: Dolni Bory, Czech Republic.

Libethenite, 80 mm, L'ubietova, Slovakia

Triplite, 70 mm, Viitaniemi, Finnland

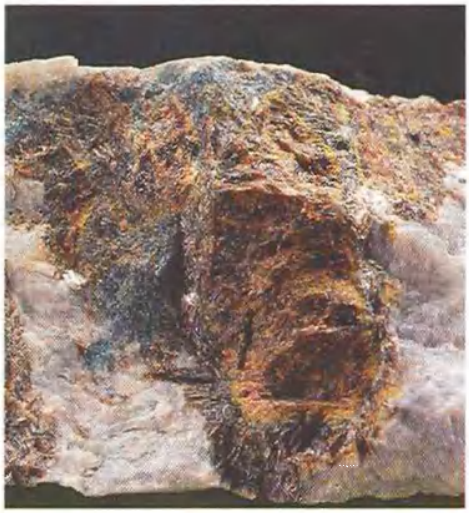

Triplite
$Mn_2(PO_4)(F,OH)$

MONOCLINIC ● ● ●

Properties: C – pink, light brown; S – light brown; L – vitreous to greasy; D – translucent to opaque; DE – 3.8; H – 5-5.5; CL – imperfect; F – conchoidal to uneven; M – short prismatic crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites and granites; hydrothermal in greisens and in quartz veins. It is usually hydrothermally altered and replaced by secondary phosphates. Huge masses several meters in size come from pegmatites in Karibib, Namibia; it also occurs in Mangualde, Portugal and in Sukula, Finland.

Libethenit
$Cu_2(PO_4)(OH)$

ORTHORHOMBIC ● ● ●

Properties: C – black-green to light green; S – olive-green; L – greasy; D – translucent; DE – 3.9; H – 4; CL – imperfect; F – conchoidal to uneven; M – short prismatic and dipyramidal crystals, botryoidal aggregates, granular, massive.
Origin and occurrence: Secondary in Cu deposits, where it occurs together with malachite, pseudomalachite and brochantite. Perfect crystals, up to 30 mm (1 3/16 in) across, come from the Rokana mine, Zambia; crystals, up to 10 mm (3/8 in), are known from Kambove, Zaire; L'ubietova, Slovakia; Nizhniy Tagil, Ural Mts., Russia.

Olivenite, 80 mm, Tsumeb, Namibia

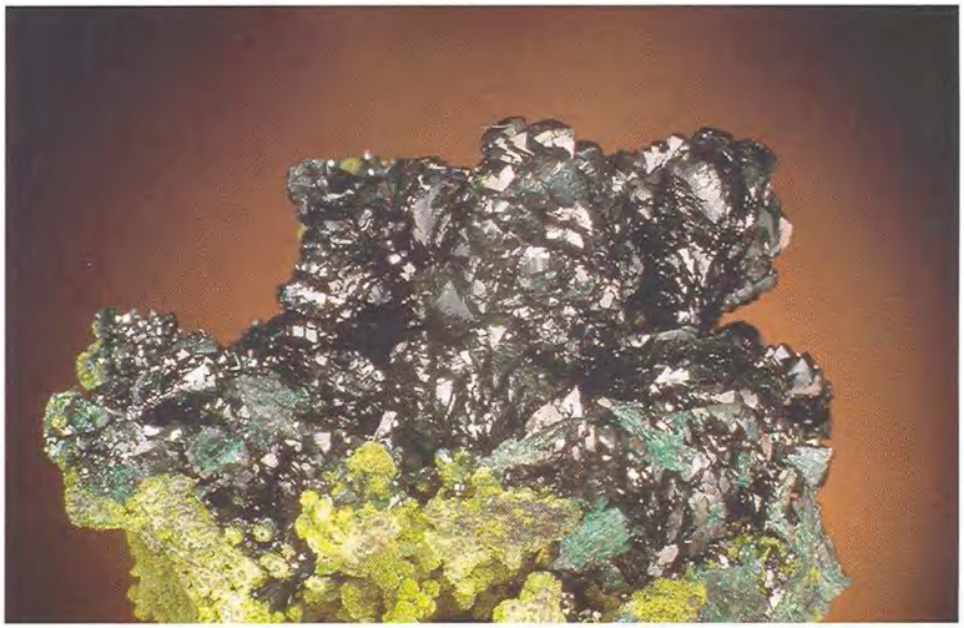

Olivenite
$Cu_2(AsO_4)(OH)$

ORTHORHOMBIC ● ● ●

Properties: C – olive-green, green-brown, gray-green to gray; S – light green; L – greasy; D – translucent to opaque; DE – 4.4; H – 3; CL – imperfect; F – conchoidal to uneven; M – long to short prismatic and dipyramidal crystals, acicular and radial aggregates, massive.
Origin and occurrence: Secondary in Cu deposits. It is associated with other secondary Cu minerals, malachite, azurite and scorodite. Perfect crystals up to 10 mm (3/8 in) across occurred in Wheal Gorland, Cornwall, and Tavistock, Devon UK; also in Tsumeb, Namibia and Ashburton Downs, Western Australia, Australia.

Adamine
$Zn_2(AsO_4)(OH)$

ORTHORHOMBIC ● ● ●

Properties: C – yellow-green, yellow, green, colorless, purple; S – white; L – vitreous; D – transparent to translucent; DE – 4.4; H – 3.5; CL – imperfect; F – conchoidal to uneven; M – long to short prismatic and dipyramidal crystals, acicular and radial aggregates, massive; LU – yellow-green.

Origin and occurrence: Secondary in Zn deposits, associated with hemimorphite, goethitee and smithsonite. Rich druses of green and rare purple crystals up to 70 mm (2¾ in) across come from Mina Ojuela, Mapimi, Durango, Mexico; also known from Tsumeb, Namibia; Laurion, Greece and Cap Garonne, France.

Adamite, 34 mm, Mapimi, Mexico

Lazulite, 11 mm x, Lincoln Co., U.S.A.

Lazulite
$MgAl_2(PO_4)_2(OH)_2$

MONOCLINIC ● ● ●

Properties: C – dark to light blue, blue-green; S – white; L – vitreous; D – transparent to translucent; DE – 3.1; H – 5.5-6; CL – imperfect; F – conchoidal to uneven; M – short prismatic and dipyramidal crystals, granular, massive.
Origin and occurrence: Hydrothermal in quartz veins and granitic pegmatites, where it is formed by decomposition of primary phosphates; metamorphic in quartzites.
Imperfect crystals up to 100 mm (4 in) across come from Horrsjoberg, Sweden; perfect crystals about 50 mm found in Ashudi, Pakistan; also known from Big Fish River, Yukon, Canada and near Werfen, Austria.

Scorzalite
$FeAl_2(PO_4)_2(OH)_2$

MONOCLINIC ● ● ●

Properties: C – dark blue, blue-green; S – white; L – vitreous; D – transparent to translucent; DE – 3.3; H – 6; CL – imperfect; F – conchoidal to uneven; M – prismatic crystals, granular, massive.
Origin and occurrence: Hydrothermal in granitic pegmatites, as a replacement product of primary phosphates, rare in quartz veins. Dark blue, granular aggregates up to 100 mm (4in) across occur in the Palermo No. 1 and No. 2 mines, North Groton, New Hampshire and the Victory mine, Custer, South Dakota, USA.

Rockbridgeite
$Fe^{2+}Fe^{3+}{}_4(PO_4)_3(OH)_5$

ORTHORHOMBIC ● ● ●

Properties: C – dark and light green, black-green; S – green; L – dull; D – translucent to opaque; DE – 3.4; H – 4.5; CL – good; F – uneven; M – acicular crystals, radial aggregates and crusts, granular, massive.
Origin and occurrence: Secondary in granitic pegmatites and in Fe deposits. It mostly originates from the hydrothermal alteration of primary phosphates, mainly triphyllite, and it is associated with other secondary phosphates. Rich radial aggregates up to 50 mm (2 in) across come from Hagendorf, Germany; the Tip Top mine, Custer, South Dakota and the Fletcher mine, Groton, New Hampshire, USA.

Scorzalite, 80 mm, Palermo No.2 Mine, U.S.A.

Rockbridgeite, 70 mm, Hagendorf, Germany

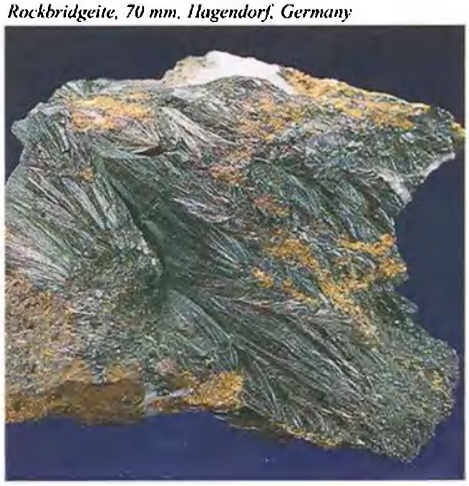

Frondelite, 15 mm xx, Custer, U.S.A.

Frondelite

$MnFe^{3+}{}_4(PO_4)_3(OH)_5$

ORTHORHOMBIC ● ● ●

Properties: C – light olive-green, brown, black-green; S – green; L – dull; D – translucent to opaque; DE – 3.5; H – 4.5; CL – good; F – uneven; M – acicular crystals, radial aggregates and crusts, granular, massive.

Origin and occurrence: Secondary in granitic pegmatites, where it forms as a result of the hydrothermal alteration of primary phosphates, mainly lithiophillite. Radial aggregates occur in the Fletcher mine, Groton, New Hampshire, USA; also in Sapucaia, Minas Gerais, Brazil.

Dufrenite

$Fe^{2+}Fe^{3+}{}_4(PO_4)_3(OH)_5 \cdot 2\,H_2O$

MONOCLINIC ● ● ●

Properties: C – dark green, black-green, black; S – green; L – vitreous to dull; D – translucent to opaque; DE – 3.4; H – 3.5-4.5; CL – good; F – uneven; M – radial aggregates and crusts, granular, massive.

Origin and occurrence: Secondary in granitic pegmatites and in the oxidation zone of Fe deposits; it forms in pegmatites as a result of the hydrothermal alteration of primary phosphates. Radial aggregates occur in pegmatites in Hagendorf and Hühnerkobel, Germany; also in gossan in the Wheal Phoenix mine, Cornwall, UK.

Dufrenite, 9 mm, Lancaster, U.S.A.

Pseudomalachite

$Cu_5(PO_4)_2(OH)_4$

MONOCLINIC ● ● ●

Properties: C – green, black-green; S – green; L – vitreous to dull; D – translucent to opaque; DE – 4.3; H – 4.5-5; CL – good; F – uneven; M – short prismatic crystals, botryoidal aggregates with radial structure, massive.

Origin and occurrence: Secondary in Cu deposits, associated with malachite, chrysocolla, libethenite and goethite. Beautiful botryoidal aggregates and large masses come from Nizhniy Tagil, Ural mountains, Russia; also from Ehl and Virneberg, Germany; from L'ubietova , Slovakia; and also from many localities in Shaba province, Zaire.

Pseudomalachite, 80 mm, L'ubietova, Slovakia

Cornubite, 30 mm, Farbiste, Slovakia

Cornubite
$Cu_5(AsO_4)_2(OH)_4$

TRICLINIC ●

Properties: C – light to dark green; S – light green; L – vitreous to dull; D – translucent; DE – 4.8; H – not determined; CL – not determined; F – uneven; M – botryoidal aggregates, massive.
Origin and occurrence: Secondary in Cu deposits. It was found together with olivenite and clinoclase in the Bedford United quarry, Cornwall, UK; also in Ashburton Downs, Western Australia, Australia and Farbiste, Slovakia.

Cornetite, 10 mm aggregate, Lumumbashi, Zair

Augelite, 34 mm, Yukon Territory, Canada

Augelite
$Al_2(PO_4)(OH)_3$

MONOCLINIC ● ● ●

Properties: C – colorless, white, yellowish, pinkish; S- white; L – vitreous to dull; D – transparent to translucent; DE – 2.7; H- 4.5-5; CL – good; F – uneven; M – thick tabular to prismatic crystals, granular, massive.
Origin and occurrence: Hydrothermal in granitic pegmatites as a product of primary phosphates replacement; rare metamorphic in quartzites. Massive aggregates several decimeters across occur in pegmatites in Burango, Rwanda; also in the Hugo mine, Custer, South Dakota and Mount White, California, USA. Crystals up to 20 mm (25/32 in) across come from Rapid Creek, Yukon, Canada and from the Champion mine, California; rare small crystals found in the Palermo No. 1 mine, North Groton, New Hampshire, USA.

Cornetite
$Cu_3(PO_4)(OH)_3$

ORTHORHOMBIC ●

Properties: C – dark blue to blue-green; S – light blue; L – vitreous; D – transparent to translucent; DE – 4.1; H – 4.5; CL – none; F – conchoidal to uneven; M – short prismatic crystals, coatings.
Origin and occurrence: Secondary in Cu deposits. It is rare in the Etoile mine near Lumumbashi and in Kalagi, Shaba province, Zaire; also occurs in Yerington, Nevada and in Saginaw Hill, Arizona, USA.

Clinoclase, 40 mm, Cornwall, UK

Clinoclase
$Cu_3(AsO_4)(OH)_3$

MONOCLINIC ● ●

Properties: C – dark green-blue to black-green; S – blue-green; L – vitreous; D – transparent to translucent; DE – 4.4; H – 2.5-3; CL – perfect; F – uneven; M – prismatic and tabular crystals, botryoidal aggregates.
Origin and occurrence: Secondary in Cu deposits, often associated with malachite, azurite and other secondary Cu minerals. Spherical aggregates up to 10 mm (⅜ in) known from the Majuba Hill mine, Nevada, USA; occurs rarely near Tavistock, Devon, UK; also in Novoveska Huta, Slovakia.

Conichalcite
$CuCa(AsO_4)(OH)$

ORTHORHOMBIC ● ● ●

Properties: C – yellow-green to emerald-green; S – light green; L – vitreous to greasy; D – transparent to translucent; DE – 4.3; H – 4.5; CL – none; F – uneven; M – short prismatic crystals, botryoidal aggregates with radial structure, massive.
Origin and occurrence: Secondary in Cu deposits. It occurs as rich botryoidal aggregates in Otavi, Namibia; Tintic, Utah, in the Higgins mine, Bisbee, Arizona and Yerington, Nevada, USA.

Conichalcite, 70 mm, Tintic, U.S.A.

Duftite
$PbCu(AsO_4)(OH)$

ORTHORHOMBIC ● ●

Properties: C – yellow-green, olive-green to gray-green; S – light green; L – vitreous to greasy; D – translucent; DE – 6.5; H – 3; CL – not determined; F – uneven; M – small crystals, botryoidal aggregates and coatings.
Origin and occurrence: Secondary in base metals deposits. It is associated with malachite and azurite in Tsumeb, Namibia; in Mina Ojuela, Mapimi, Durango, Mexico; in Moldava, Czech Republic.

Duftite, 90 mm, Tsumeb, Namibia

Descloizite, 38 mm, Grootfontein, Namibia

Descloizite

$PbZn(VO_4)(OH)$

ORTHORHOMBIC ● ● ●

Properties: C – red-orange, red-brown to brown-black, gray-green; S – light yellow-brown; L – greasy; D – transparent, translucent to opaque; DE – 6.1; H – 3-3.5; CL – none; F – conchoidal to uneven; M – crystals of different habits, mostly pyramidal or prismatic, botryoidal and skeletal aggregates, massive.

Origin and occurrence: Secondary in base metals deposits. It is mainly associated with pyromorphite, mimetite, vanadinite and other secondary Pb minerals. Occurs as crystals up to 30 mm (1 1/16 in) long in Tsumeb and Berg Aukas, Namibia; also Broken Hill, Zambia; and the Mammoth mine, Tiger, Arizona, USA.

Arsendescloizite

$PbZn(AsO_4)(OH)$

ORTHORHOMBIC ●

Properties: C – light yellow; S – white; L – adamantine to greasy; D – transparent to translucent; DE – 6.1; H – 4; CL – none; F – conchoidal to uneven; M – tabular crystals, rosette-like aggregates.

Arsendescloizite, 95 mm, Mapimi, Mexico

Origin and occurrence: Secondary in base metals deposits, associated with mimetite and goethite. Found rarely as crystals of 1 mm (1/32 in) in size in Tsumeb, Namibia; also Mina Ojuela, Mapimi, Durango, Mexico.

Mottramite

$PbCu(AsO_4)(OH)$

ORTHORHOMBIC ● ●

Properties: C – grass-green to black-green; S – light green; L – vitreous to dull; D – transparent to opaque; DE – 5.9; H – 3-3.5; CL – none; F – conchoidal to uneven; M – crystals of different habits, botryoidal and dendritic aggregates, crusts and coatings, massive.
Origin and occurrence: Secondary in base metals deposits, associated with mimetite, descloizite and vanadinite. It occurs in Mottram, Cheshire, UK; in Tsumeb, Namibia; Mammoth mine, Tiger, Arizona, USA.

Brazilianite

$NaAl_3(PO_4)_2(OH)_4$

MONOCLINIC ● ● ●

Properties: C – colorless, white, yellowish, yellow-green; S – white; L – vitreous; D – transparent to translucent; DE – 3.0; H – 5.5; CL – good; F – uneven; M – short prismatic to isometric crystals, radial aggregates, granular.
Origin and occurrence: Hydrothermal in cavities in granitic pegmatites, where it is associated with fluorapatite, albite and tourmaline. It occasionally originates as a product of amblygonite replacement. Perfect yellow-green crystals up to 150 mm (6 in) across found in cavities in pegmatites in Conselheira Pena and Corrego Frio, Linopolis, Minas Gerais and from Pietras Lavradas, Paraiba, Brazil.

Mottramite, 60 mm, Tiger, Arizona, U.S.A.

Cafarsite, 9 mm x, Binntal, Switzerland

Cafarsite

$Ca_8(Ti,Fe^{2+},Fe^{3+},Mn)(As^{3+}O_3)_{12} \cdot 4\,H_2O$

CUBIC ●

Properties: C – dark brown; S – yellow-brown; L – submetallic; D – translucent; DE – 3.9; H – 5.5-6; CL – none; F – conchoidal; M – isometric crystals.
Origin and occurrence: Hydrothermal along cracks in Alpine-type veins. Cubic crystals up to 30 mm (1 1/16 in) across come from Binntal, Switzerland and Pizzo Cervandone, Italy.

Brazilianite, 47 mm, Conselheira Pena, Brazil

Carminite
$PbFe^{3+}{}_2(AsO_4)_2(OH)_2$

ORTHORHOMBIC ● ●

Properties: C – crimson-red, red-brown; S – red-yellow; L – adamantine to pearly; D – translucent; DE – 5.5; H – 3.5; CL – good; F – conchoidal to uneven; M – prismatic crystals, acicular, radial, felt-like to porous aggregates, coatings and crusts.
Origin and occurrence: Secondary in base metals deposits, associated with mimetite and scorodite. Crystals up to 10 mm (3/8 in) across occur in Tsumeb, Namibia; in Mina Ojuela, Mapimi, Durango, Mexico; in Calstock, Cornwall, UK.

Carminite, 2 mm xx, Nadap, Hungary

Bayldonite, 34 mm, Tsumeb, Namibia

Bayldonite
$PbCu_3(AsO_4)_2(OH)_2 \cdot H_2O$

MONOCLINIC ● ●

Properties: C – yellow-green, olive-green; S – light green; L – resinous; D – translucent; DE – 5.5; H – 4.5; CL – not determined; F – uneven; M – earthy aggregates, coatings, massive.
Origin and occurrence: Secondary in the oxidation zone in hydrothermal Cu deposits and in greisens. It occurs as common yellow-green coatings and crystals up to 10 mm (3/8 in) across, associated with large azurite and mimetite crystals in Tsumeb, Namibia; also as coatings in St. Day, Cornwall, UK; in Moldava, Czech Republic.

Vesignieite
$BaCu_3(VO_4)_2(OH)_2$

MONOCLINIC ● ●

Properties: C – yellow-green, olive-green; S – light yellow-green; L – vitreous to dull; D – translucent;

Vesigneite, 80 mm, Vrancice, Czech Republic

DE – 4.1; H – 3-4; CL – good; F – uneven; M – earthy and pulverulent aggregates and coatings, massive.

Origin and occurrence: Secondary in base metals deposits and in sediments, containing Cu sulfides. It occurs as yellow-green coatings in sediments in Horni Kalna, Czech Republic; also in the Mashamba West mine, Zaire; together with barite and psilomelane in Friedrichsroda, Germany.

Arsentsumebite

$Pb_2Cu(AsO_4)(SO_4)(OH)$

MONOCLINIC ●

Properties: C – bluish-green to light green; S – light green; L – dull; D – translucent; DE – 6.5; H – 3; CL – good; F – uneven; M – earthy aggregates, coatings, massive.

Origin and occurrence: Secondary in the oxidation zone, where it forms as a result of mimetite replacement. Found rarely in Moldava, Czech Republic; in Tsumeb, Namibia, where it forms pseudo-morphs after azurite crystals.

Arsentsumebite, 40 mm, Moldava, Czech Republic

Fluorapatite, 50 mm, Dusso, Pakistan

Fluorapatite

$Ca_5(PO_4)_3F$

HEXAGONAL ● ● ● ● ●

Properties: C – colorless, white, yellowish, pinkish, blue, purple, green, brown with various hues; S – white; L – vitreous to dull; D – transparent to translucent, sometimes opaque; DE – 3.2; H – 5; CL – imperfect; F – conchoidal to uneven; M – long prismatic to tabular crystals, botryoidal, earthy and fibrous aggregates, massive; LU – yellow.

Origin and occurrence: Magmatic in granites, syenites, diorites, gabbros and various types of pegmatites, also in volcanic rocks; hydrothermal in quartz veins, ore veins, greisens and Alpine-type veins; metamorphic in different types of gneisses, migmatites, mica schists, skarns and amphibolites; in different types of sedimentary rocks. Perfect, short prismatic, purple transparent crystals up to 40 x 40 mm (1 9/16 x 1 9/16 in) are renowned from pegmatite in the Pulsifer quarry, Mount Apatite, Auburn, Maine, USA; pink crystals from Dusso, Pakistan. Perfect crystals about 100 mm (4in) across also from Alpine-type veins, e.g. in Fiesch, Switzerland; also known from the quartz veins in greisens, associated with wolframite, in Panasqueira, Portugal; in Horni Slavkov, Czech Republic; Ehrenfriedersdorf, Germany. Also found in skarns at Cerro de Mercado, Durango, Mexico. Large deposits of mas-

Chlorapatite, 20 mm x, Bob's Lake, Ontario, Canada

sive apatite located in Kola Peninsula, Russia; crystals weighing up to 300 kg (660 lb) from the vicinity of Clear Lake, Ontario, Canada.
Application: main source of P, chemical industry, fertilizer.

Chlorapatite
$Ca_5(PO_4)_3Cl$

HEXAGONAL ● ● ●

Properties: C – white, various hues of yellow; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.2; H – 5; CL – imperfect; F – conchoidal to uneven; M – long prismatic to tabular crystals, granular.
Origin and occurrence: Magmatic in nepheline syenites and their pegmatites, some gabbros and volcanic rocks, also in meteorites; metamorphic in skarns. It is usually associated with scapolite, amphibole, titanite and magnetite. Perfect prismatic crystals up to 35 cm ($13^{12}/_{16}$ in) long, come from pegmatites in Bamle, Norway; also from Kurokura, Japan.

Hydroxylapatite
$Ca_5(PO_4)_3(OH)$

HEXAGONAL ● ● ●

Properties: C – white, yellow, various hues of gray; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.2; H – 5; CL – imperfect; F – conchoidal to uneven; M – short prismatic to tabular crystals, acicular aggregates, granular.
Origin and occurrence: Metamorphic in talc schists and serpentinites; hydrothermal in granitic pegmatites; sedimentary in organic remnants. Crystals up to 30 mm ($1^3/_{16}$ in) across known from Snarum, Norway; Hospental, Switzerland; Eagle, Colorado, USA.

Carbonate-fluorapatite
$Ca_5(PO_4, CO_3)_3F$

HEXAGONAL ● ● ●

Properties: C – white, gray; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.2; H – 5; CL – imperfect; F – conchoidal to uneven; M B spherical and botryoidal aggregates, massive.
Origin and occurrence: Hydrothermal in ore veins and along the cracks in volcanic rocks. Rich botryoidal aggregates are known together with hyalite opal from Valec, Czech Republic; the Wheal Franco mine, Tavistock, Devon, UK.

Hydroxylapatite, 2 mm xx, Tornaszentandrás, Hungary

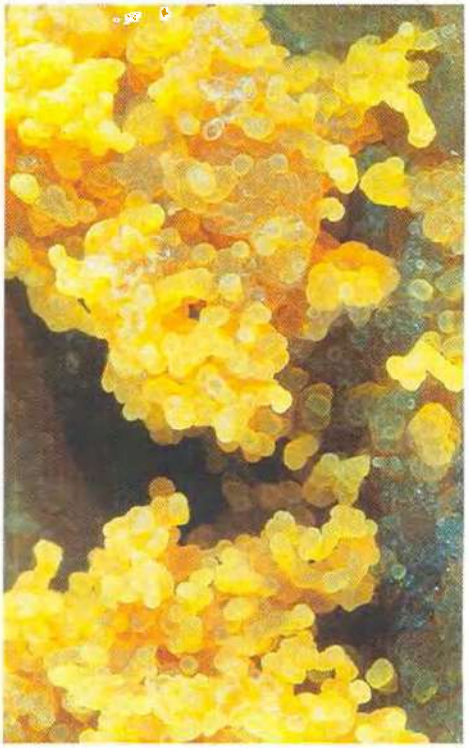

Carbonate-fluorapatite, 80 mm, Staffel, Germany

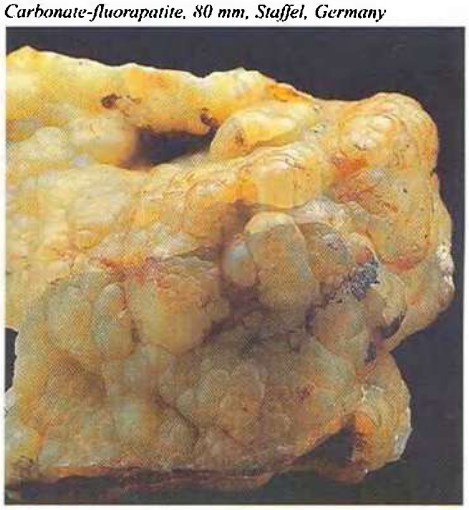

Pyromorphite, 41 mm, Les Farges, France

Pyromorfite

$Pb_5(PO_4)_3Cl$

HEXAGONAL ● ● ● ●

Properties: C – green, brown, yellow, orange, white, gray; S -white; L – adamantine to greasy; D – transparent to translucent; DE – 7.1; H – 3.5-4; CL – imperfect; F – conchoidal to uneven; M – long prismatic to tabular and pyramidal crystals, botryoidal aggregates with radial structure, acicular and earthy aggregates, massive.

Origin and occurrence: Secondary in the oxidation zone of Pb deposit, associated with cerussite, goethitee and other secondary minerals. Perfect crystals up to 40 mm (1 9/16 in) long come from many localities, e.g. green and brown ones from Bad Ems and Zschopau, Germany; yellow-brown ones from Les Farges, France; green, brown and yellow ones from Mina Ojuela, Mapimi, Durango, Mexico. Green, orange and brown crystals, up to 60 mm (2 3/8 in) long, are known from the Bunker Hill mine, Idaho, USA.

Pyromorphite, 62 mm, Les Farges, France

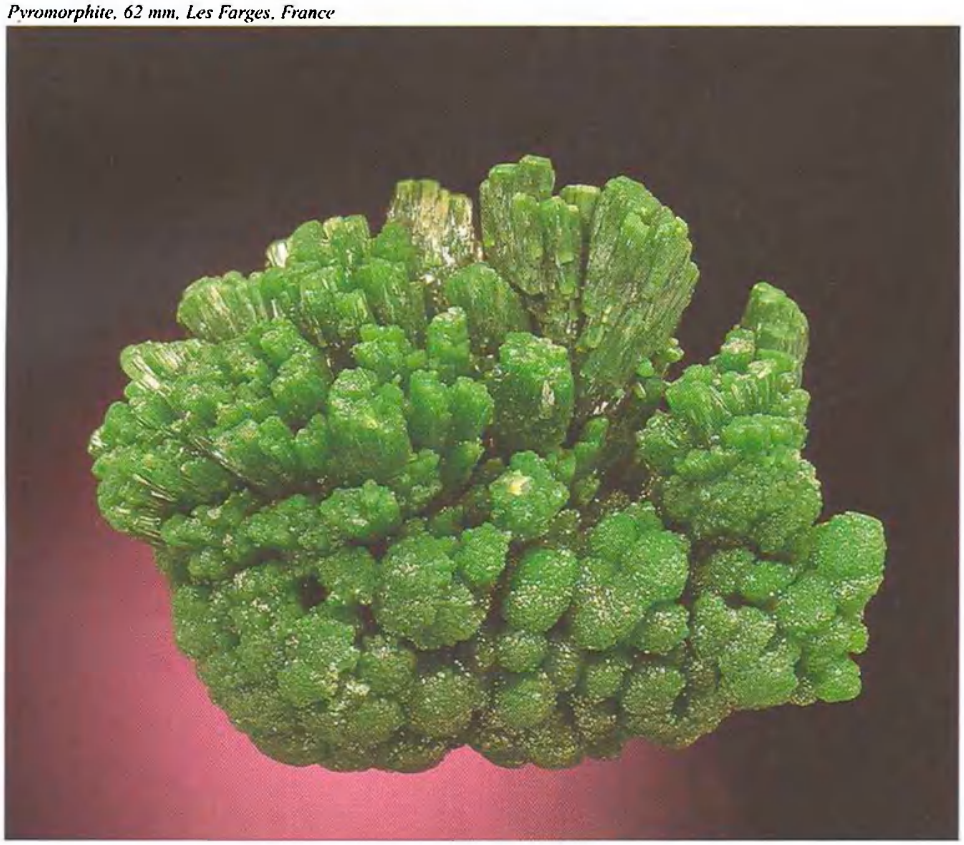

Mimetite, 14 mm x, Johanngeorgenstadt, Germany

Mimetite

$Pb_5(AsO_4)_3Cl$

HEXAGONAL ● ● ●

Varieties: campylite

Properties: C – yellow, orange, brown, yellow-brown, white, gray; S – white; L – adamantine to greasy; D – transparent to translucent; DE – 7.3; H – 3.5-4; CL – imperfect; F – conchoidal to uneven; M – long to short prismatic and pyramidal crystals, botryoidal aggregates with radial structure, acicular and earthy aggregates, granular.
Origin and occurrence: Secondary in Pb deposits, associated with pyromorphite and goethite. Perfect crystals up to 20 mm ($^{25}/_{32}$ in) long come from Johanngeorgenstadt, Germany; campylite occurs in Drygill, UK; crystals up to 50 mm (2 in) across found in Tsumeb, Namibia; Santa Eulalia and San Pedro, Chihuahua, Mexico. Beautiful yellow crystals up to 30 mm ($1^{1}/_{16}$ in) found recently in Hat Yai province, Thailand.

Vanadinite

$Pb_5(VO_4)_3Cl$

HEXAGONAL ● ● ●

Properties: C – yellow, orange, red, brown, yellow-brown; S – white; L – adamantine to greasy; D – transparent to translucent; DE – 6.9; H – 2.5-3; CL – none; F – conchoidal to uneven; M – long to short prismatic and pyramidal crystals, botryoidal aggregates with radial structure, acicular and earthy aggregates, granular.
Origin and occurrence: Secondary in Pb deposits, associate with pyromorphite, wulfenite and goethite. Perfect crystals up to 130 mm ($5^{2}/_{16}$ in) long come from Djebel Mahseur and Mibladen, Morocco. Prismatic crystals were also found in Tsumeb, Namibia; in the Old Yuma, Red Cloud, Apache and Mammoth mines, Arizona, USA.

Vanadinite, 64 mm, Mibladen, Morocco

Atelestite, 70 mm, Schneeberg, Germany

Atelestite
$Bi_2(AsO_4)O(OH)$

MONOCLINIC ●

Properties: C – yellow, yellow-green; S – white; L – adamantine to greasy; D – transparent to translucent; DE – 7.0; H – 4.5-5; CL – imperfect; F – conchoidal to uneven; M – small tabular crystals, spherical aggregates.
Origin and occurrence: Secondary in the oxidation zone of Bi deposits. It was found in Schneeberg, Germany.

Hureaulite
$Mn_5(PO_4)_2[PO_3(OH)]_2 \cdot 4\,H_2O$

MONOCLINIC ● ● ●

Properties: C – orange-red, pink, purplish, white, gray; S – white; L – vitreous; D – transparent to translucent; DE – 3.2; H – 3.5; CL – good; F – uneven; M – prismatic to tabular crystals, coatings, granular, massive.

Hureaulite, 53 mm, Shingus, Pakistan

Variscite, 30 mm, Fairfield, U.S.A.

Origin and occurrence: Secondary in granitic pegmatites, as a result of alteration of the primary phosphates, mostly lithiophillite, triphyllite and graftonite. Purplish crystals up to 50 mm (2 in) across come from Shingus, Pakistan. Similar crystals found in Hagendorf, Germany. Also known from Mangualde, Portugal; Sao Jose da Safira, Minas Gerais, Brazil; from the Tip Top mine, Custer, South Dakota, USA.

Variscite
$Al(PO_4) \cdot 2\,H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, greenish, blue-green; S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 3.5-4.5; CL – good; F – uneven; M – isometric crystals, botryoidal aggregates, nodules, coatings, massive.
Origin and occurrence: : Hydrothermal in cracks in sedimentary rocks, rich in Al and P, also in phosphates deposits. It is associated with apatite, wavellite and other phosphates. Renowned greenish nodules up to 30 cm (12 in) in diameter come from Clay Canyon, Fairfield, Utah, USA; also known from Ronneburg, Germany and Jivina near Beroun, Czech Republic.

Strengite
$Fe^{3+}(PO_4) \cdot 2\,H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, pinkish, red-purple; S – white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 3.5-4.5; CL – good; F – conchoidal; M – isometric, tabular to short prismatic crystals, botryoidal aggregates with radial structure, coatings.
Origin and occurrence: Secondary in granitic

Strengite, 34 mm, Svappavaara, Sweden

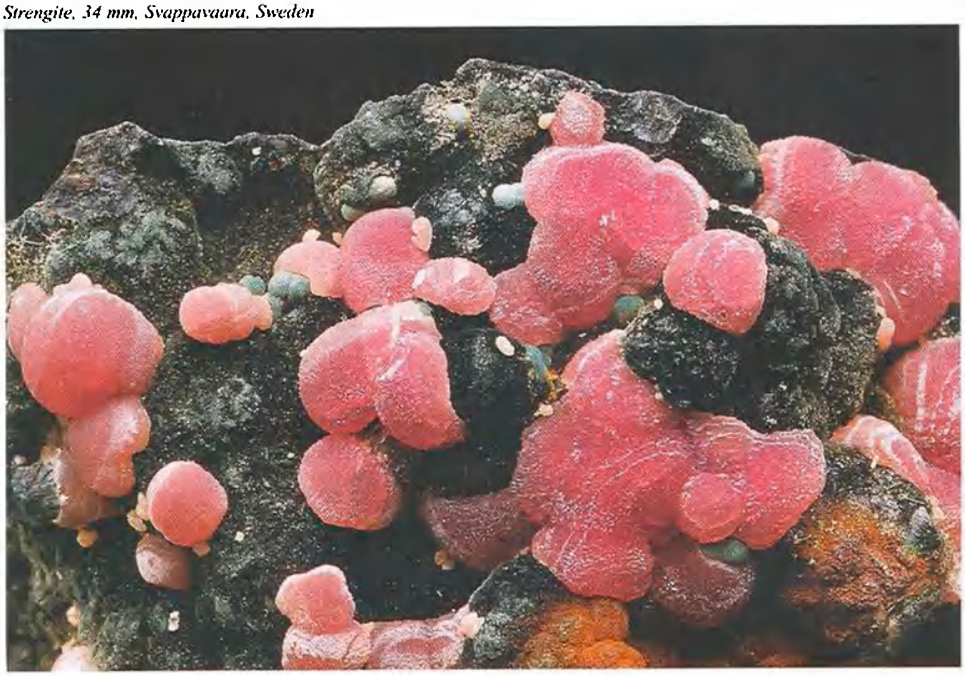

Scorodite, 10 mm x, Zacatecas, Mexico

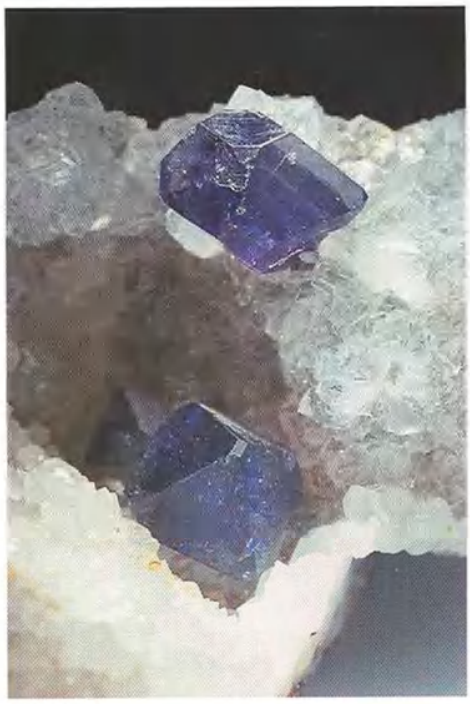

pegmatites where it forms as a result of hydrothermal replacement of primary phosphates; in the oxidation zone of Fe deposit together with goethite. Purple crystals up to 5 mm (3/16 in) across, come from the Bull Moose mine, Custer, South Dakota, USA; also known from Pleystein, Germany and from Teskov, Czech Republic.

Scorodite

$Fe^{3+}(AsO_4) \cdot 2\,H_2O$

ORTHORHOMBIC ● ● ●

Properties: B – light green, gray-green, olive-green, colorless, blue, yellow-brown; S – light green; L – vitreous to resinous; D – transparent to translucent; DE – 3.3; H – 3.5-4; CL – imperfect; F – conchoidal; M – dipyramidal to short prismatic crystals, botryoidal and earthy aggregates, coatings, granular, massive.

Origin and occurrence: Secondary in the oxidation zone of ore deposits, associated with arsenopyrite, lollingite and other arsenides; in granitic pegmatites; hydrothermal in the hot springs.

Perfect crystals up to 50 mm (2 in) across come from Tsumeb, Namibia; the Kiura mine, Oita, Japan; Mina Ojuela, Mapimi, Durango, Mexico; green crusts, several cm thick, were found in Djebel Debar, Algeria.

Phosphophyllite, 18 mm x, Potosí, Bolivia

Phosphophyllite
$Zn_2Fe(PO_4)_2 \cdot 4\,H_2O$

MONOCLINIC ● ●

Properties: C – colorless, blue-green, blue; S – white; L – vitreous; D – transparent to translucent; DE – 3.1; H – 3-3.5; CL – good; F – uneven; M – long to short prismatic and thick tabular crystals, granular.
Origin and occurrence: Secondary in the oxidation zone of ore deposits; in granitic pegmatites, where it replaces primary phosphates. Perfect crystals, up to 140 mm (5½ in) across come from sulfide cavities in the Unificada mine, Potosi, Bolivia. Also occurs at Hagendorf, Germany.

Ludlamite
$Fe_3(PO_4)_2 \cdot 4\,H_2O$

MONOCLINIC ● ●

Properties: C – light green, green; S – white; L – vitreous; D – transparent to translucent; DE – 3.2; H – 3.5; CL – perfect; F – uneven; M – thin to thick tabular crystals, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of ore deposits; in granitic pegmatites, where it replaces primary phosphates. Perfect crystals up to 90 mm (3 9/16 in) across known from the San Antonio mine, Santa Eulalia, Chihuahua, Mexico; also from Morococala, Bolivia and the Blackbird district, Idaho, USA.

Ludlamite, 36 mm, Morococala, Bolivia

Anapaite
$Ca_2Fe(PO_4)_2 \cdot 4\,H_2O$

TRICLINIC ● ●

Properties: C – light to dark green, colorless; S – white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 3.5; CL – perfect; F – uneven; M – thin to thick tabular crystals, rosette-like aggregates, granular.
Origin and occurrence: Secondary in the oxidation zone of Fe deposits and in sedimentary rocks, rich in P. Tabular crystals several mm across and spherical aggregates, up to 30 mm (1 3/16 in) across, come from the cracks within oolitic ores near Anapa and Kerch, Crimea, Ukraine; also known from Bellaver de Cerdena, Spain.

Vivianite
$Fe_3(PO_4)_2 \cdot 8\,H_2O$

MONOCLINIC ● ● ● ●

Properties: C – colorless when fresh, quickly oxidizes to blue, green, purple, black-blue; S – white to bluish; L – vitreous; D – transparent, translucent to opaque; DE – 2.7; H – 1.5-2; CL – perfect; F – uneven; M – long prismatic to acicular crystals, fibrous, earthy to pulverulent aggregates, coatings, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of Fe deposits; in granitic pegmatites, where it forms by the replacement of primary phosphates and in sedimentary rocks in proximity of organic

Anapaite, 18 mm aggregates, Anapa, Ukraine

material; hydrothermal in the ore deposits. Crystals up to 1.5 m (5 ft) long found in clay sediments in Anloua, Cameroon; crystals up to 200 mm (7⅞ in) across known from Morococala, Bolivia. Smaller crystals come from Trepca, Serbia; Leadville, Colorado, Bingham Canyon, Utah, USA; also from Kerch and Anapa, Crimea, Ukraine.

Erythrine

$Co_3(AsO_4)_2 \cdot 8\,H_2O$

MONOCLINIC ● ● ●

Properties: C – dark purple, pink, colorless; S – light pink to white; L – vitreous; D – transparent to translucent; DE – 3.2; H – 1.5-2.5; CL – perfect; F – uneven; M – long prismatic, acicular to tabular crystals, earthy aggregates, coatings, granular, massive.

Origin and occurrence: Secondary in the oxidation zone of Co, Ni and U deposits. Tabular crystals up to 60 mm (2⅜ in) long come from Bou Azzer, Morocco. Other important localities are Schneeberg, Germany; Talmessi, Iran; Cobalt, Ontario, Canada; Mount Cobalt, Queensland, Australia.

Erythrite, 29 mm, Mexico

Vivianite, 48 mm, Morococala, Bolivia

Phosphophyllite, 18 mm x, Potosí, Bolivia

Phosphophyllite
$Zn_2Fe(PO_4)_2 \cdot 4\,H_2O$

MONOCLINIC ● ●

Properties: C – colorless, blue-green, blue; S – white; L – vitreous; D – transparent to translucent; DE – 3.1; H – 3-3.5; CL – good; F – uneven; M – long to short prismatic and thick tabular crystals, granular.
Origin and occurrence: Secondary in the oxidation zone of ore deposits; in granitic pegmatites, where it replaces primary phosphates. Perfect crystals, up to 140 mm (5½ in) across come from sulfide cavities in the Unificada mine, Potosí, Bolivia. Also occurs at Hagendorf, Germany.

Ludlamite
$Fe_3(PO_4)_2 \cdot 4\,H_2O$

MONOCLINIC ● ●

Properties: C – light green, green; S – white; L – vitreous; D – transparent to translucent; DE – 3.2; H – 3.5; CL – perfect; F – uneven; M – thin to thick tabular crystals, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of ore deposits; in granitic pegmatites, where it replaces primary phosphates. Perfect crystals up to 90 mm (3⁹/₁₆ in) across known from the San Antonio mine, Santa Eulalia, Chihuahua, Mexico; also from Morococala, Bolivia and the Blackbird district, Idaho, USA.

Ludlamite, 36 mm, Morococala, Bolivia

Anapaite
$Ca_2Fe(PO_4)_2 \cdot 4\,H_2O$

TRICLINIC ● ●

Properties: C – light to dark green, colorless; S – white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 3.5; CL – perfect; F – uneven; M – thin to thick tabular crystals, rosette-like aggregates, granular.
Origin and occurrence: Secondary in the oxidation zone of Fe deposits and in sedimentary rocks, rich in P. Tabular crystals several mm across and spherical aggregates, up to 30 mm (1³/₁₆ in) across, come from the cracks within oolitic ores near Anapa and Kerch, Crimea, Ukraine; also known from Bellaver de Cerdena, Spain.

Vivianite
$Fe_3(PO_4)_2 \cdot 8\,H_2O$

MONOCLINIC ● ● ● ●

Properties: C – colorless when fresh, quickly oxidizes to blue, green, purple, black-blue; S – white to bluish; L – vitreous; D – transparent, translucent to opaque; DE – 2.7; H – 1.5-2; CL – perfect; F – uneven; M – long prismatic to acicular crystals, fibrous, earthy to pulverulent aggregates, coatings, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of Fe deposits; in granitic pegmatites, where it forms by the replacement of primary phosphates and in sedimentary rocks in proximity of organic

Anapaite, 18 mm aggregates, Anapa, Ukraine

material; hydrothermal in the ore deposits. Crystals up to 1.5 m (5 ft) long found in clay sediments in Anloua, Cameroon; crystals up to 200 mm (7⅞ in) across known from Morococala, Bolivia. Smaller crystals come from Trepca, Serbia; Leadville, Colorado, Bingham Canyon, Utah, USA; also from Kerch and Anapa, Crimea, Ukraine.

Erythrine

$Co_3(AsO_4)_2 \cdot 8\,H_2O$

MONOCLINIC ● ● ●

Properties: C – dark purple, pink, colorless; S – light pink to white; L – vitreous; D – transparent to translucent; DE – 3.2; H – 1.5-2.5; CL – perfect; F – uneven; M – long prismatic, acicular to tabular crystals, earthy aggregates, coatings, granular, massive.

Origin and occurrence: Secondary in the oxidation zone of Co, Ni and U deposits. Tabular crystals up to 60 mm (2⅜ in) long come from Bou Azzer, Morocco. Other important localities are Schneeberg, Germany; Talmessi, Iran; Cobalt, Ontario, Canada; Mount Cobalt, Queensland, Australia.

Erythrite, 29 mm, Mexico

Vivianite, 48 mm, Morococala, Bolivia

Annabergite, 55 mm, Laurion, Greece

Annabergite

$Ni_3(AsO_4)_2 . 8 H_2O$

MONOCLINIC ● ● ●

Varieties: cabrerite (Mg contents)

Properties: C – light to dark green, white; S – white; L – vitreous; D – transparent to translucent; DE – 3.2; H – 1.5-2.5; CL – perfect; F – uneven; M – long prismatic to acicular and tabular crystals, acicular and earthy aggregates, coatings, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of Ni deposits, associated with erythrite. Crystals up to 5 mm (1/16 in) across occur in Gukuroren, Turkey; Sierra Cabrera, Spain; Laurion, Greece. Nodules up to 20 mm (25/32 in) come from the Snowbird mine, Montana, USA.

Symplesite

$Fe_3(AsO_4)_2 . 8 H_2O$

TRICLINIC ● ●

Properties: C – bluish, dark blue, light green, black-green; S – white; L – vitreous; D – transparent to translucent; DE – 3.0 H – 2.5; CL – perfect; F – uneven; M – acicular to tabular crystals, spherical aggregates with radial structure, earthy aggregates and coatings, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of ore deposits and in granitic pegmatites with

Symplesite, 12 mm xx, Baia Sprie, Romania

Picropharmacolite, 128 mm, France

As minerals, mainly arsenopyrite and lollingite. It occurs in Schneeberg, Germany; Baia Sprie, Romania and Trebsko, Czech Republic.

Picropharmacolite

$Ca_4MgH_2(AsO_4)_4 \cdot 12\ H_2O$

TRICLINIC ● ●

Properties: C – colorless, white; S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 2.5; CL – perfect; F – uneven; M – acicular crystals, spherical aggregates with radial structure, coatings.
Origin and occurrence: Secondary in the oxidation zone of ore deposits with As minerals, mostly arsenopyrite and lollingite. It is also known from Ste-Marie-aux-Mines and Salsigne, France; Freiberg, Germany and Jachymov, Czech Republic.

Brushite

$CaH(PO_4) \cdot 2\ H_2O$

MONOCLINIC ● ●

Properties: C – colorless, white; S – white; L – vitreous, pearly on cleavage planes; D – transparent to translucent; DE – 2.3; H – 2.5; CL – perfect; F – uneven; M – prismatic, acicular to tabular crystals, earthy aggregates and coatings, massive.
Origin and occurrence: Secondary on bat and bird

Brushite, 4 mm xx, Domica, Slovakia

Legrandite, 25 mm xx, Mapimi, Mexico

excrements and bones, it impregnates bones, along the cracks of phosphorites. Tabular crystals up to 20 mm ($^{25}/_{32}$ in) across come from Quercy, France also occurs near Oran, Algeria and Pig Hole, Virginia, USA.

Legrandite

$Zn_2(AsO_4)(OH) \cdot H_2O$

MONOCLINIC ● ●

Properties: C – colorless, yellow, purple; S – white; L – vitreous; D – transparent to translucent; DE – 4.0 H – 4.5; CL – imperfect; F – uneven; M – long prismatic crystals and their inter-growths, radial aggregates.
Origin and occurrence: Secondary in the oxidation zone of Zn deposits and in pegmatites. Prismatic crystals up to 250 mm ($9^{13}/_{16}$ in) long found in Mina Ojuela, Mapimi, Durango, Mexico. It is also known from Galileia, Minas Gerais, Brazil and Tsumeb, Namibia.

Euchroite, 10 mm xx, L'ubietova, Slovakia

Euchroite

$Cu_2(AsO_4)(OH) . 3 H_2O$

ORTHORHOMBIC ● ●

Properties: C – emerald-green; S – white; L – vitreous; D – transparent to translucent; DE – 3.5; H – 3.5-4; CL – imperfect; F – conchoidal to uneven; M – short prismatic to thick tabular crystals, massive.

Origin and occurrence: Secondary in the oxidation zone of Cu deposits, associated with olivenite and malachite. Thick tabular crystals up to 30 mm (1 3/16 in) come from L'ubietova, Slovakia. It also occurs in Zapacica, Bulgaria and Chessy, France.

Vauxite, 20 mm xx, Llallagua, Bolivia

Strunzite, 10 mm xx, Hagendorf, Germany

Vauxite

$FeAl_2(PO_4)_2(OH)_2 . 6 H_2O$

TRICLINIC ● ●

Properties: C – light to dark blue; S – white; L – vitreous; D – transparent to translucent; DE – 2.4; H – 3.5; CL – none; F – uneven; M – tabular crystals, radial aggregates, massive.

Origin and occurrence: Secondary in the oxidation zone of Sn deposits, associated with wavellite. Crystals occur in the Siglo XX Mine, Llallagua, Bolivia.

Strunzite

$MnFe^{3+}{}_2(PO_4)_2(OH)_2 . 6 H_2O$

MONOCLINIC ● ● ●

Properties: C – light yellow, yellow-brown; S – yellowish; L – vitreous; D – transparent to translucent; DE – 2.5; H – not determined; CL – not determined; F – uneven; M – acicular crystals, acicular and fibrous aggregates.

Origin and occurrence: Secondary in granitic pegmatites, as a result of weathering of primary phosphates, mostly triphyllite; rarely hydrothermal in the cracks of Fe-rich sedimentary rocks. Acicular crystals up to 20 mm (25/32 in) across known from Hagendorf, Germany; the Palermo No. 1 and No. 2 mines, North Groton and the Fletcher mine, Groton, New Hampshire, USA.

Cacoxenite, 1 mm xx, Hellertown, Pennsylvania, U.S.A.

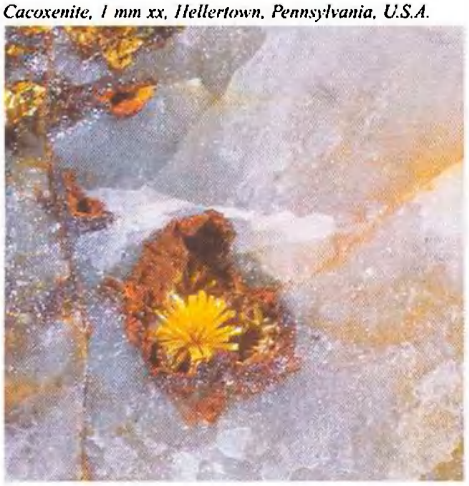

Cacoxenite

$(Fe^{3+},Al)_{25}(PO_4)_{17}O_6(OH)_{12} \cdot 75\ H_2O$

HEXAGONAL ● ● ●

Properties: C – light yellow, yellow-brown, orange; S – yellow; L – silky; D – transparent to translucent; DE – 2.3; H – 3-4; CL – not determined; F – uneven; M – acicular and fibrous aggregates, often with radial structure, botryoidal crusts and coatings.
Origin and occurrence: Hydrothermal on the cracks of sedimentary Fe ores, associated with wavellite; rare as secondary in granitic pegmatites, as a product of weathering of primary phosphates. Crystals up to 10 mm (⅜ in) across come from the Horcajo mine, Ciudad Real, Spain; golden-yellow acicular aggregates occur in Hrbek near Svata Dobrotiva and Trenice, Czech Republic and in Amberg, Germany.

Beraunite

$Fe^{2+}Fe^{3+}{}_5(PO_4)_4(OH)_5 \cdot 4\ H_2O$

MONOCLINIC ● ●

Properties: C – red-brown, red, gray-green; S – yellow to green-brown; L – vitreous to dull; D – translucent; DE – 3.0; H – 3.5-4; CL – good; F – uneven; M – acicular aggregates, often with radial structure, botryoidal crusts and coatings.
Origin and occurrence: Hydrothermal on the cracks of sedimentary Fe ores, typically together with wavellite; secondary in granitic pegmatites, as a product of weathering of primary phosphates. Acicular aggregates up to 10 mm (⅜ in) across occur in Mount Indian, Alabama, USA; also known from Hrbek near Svata Dobrotiva, Czech Republic and Amberg, Germany.

Beraunite, 3 mm aggregates, Nekeszeny, Hungary

Diadochite

$Fe^{3+}{}_2(PO_4)(SO_4)(OH) \cdot 5\ H_2O$

TRICLINIC ● ● ●

Properties: C – yellow-brown, brown, red-brown, yellow-green, gray-green; S – yellow to light brown; L – dull, waxy; D – translucent to opaque; DE – 2.0-2.4; H – 3; CL – not determined; F – uneven, conchoidal, earthy; M – nodules, coatings and crusts, massive.
Origin and occurrence: Secondary in the oxidation zone of Fe deposits. Diadochite caves in abandoned mines are known from Saalfeld, Germany; nodules found in New Idria, California and Eureka, Nevada, USA.

Diadochite, 40 mm, Recsk, Hungary

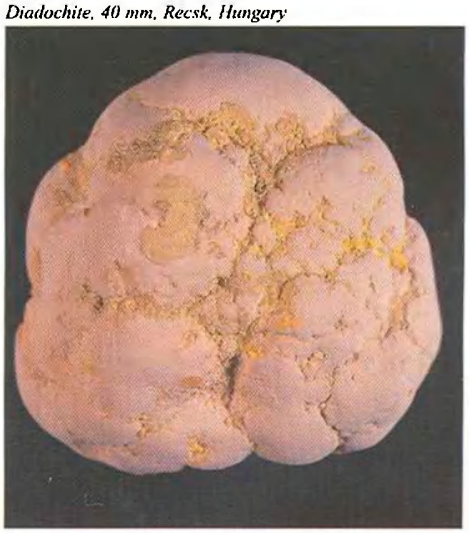

Wavellite, 40 mm, Trenice, Czech Republic

Eosphorite, 4 mm xx, Lavra da Ilha, Brazil

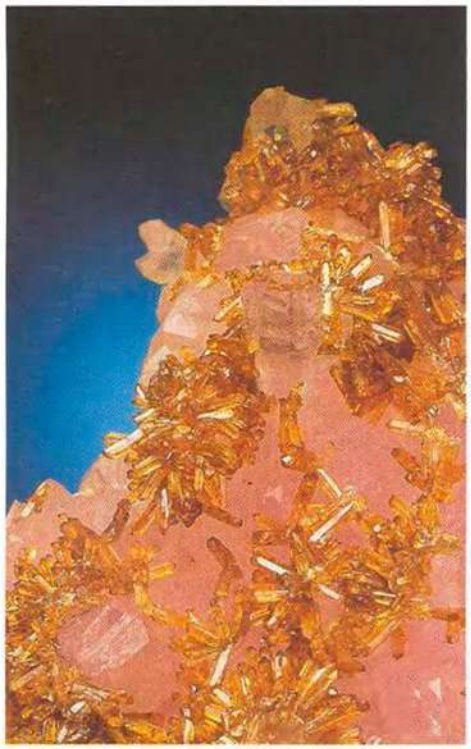

Wavellite

$Al_3(PO_4)_2(OH,F)_3 \cdot 5\,H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, greenish, light blue-green, green, yellowish; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 2.4; H – 3.5-4; CL – perfect; F – uneven; M – isometric crystals, hemispherical aggregates with radial structure, botryoidal aggregates, nodules, coatings, massive.
Origin and occurrence: Hydrothermal in the cracks of Al and P-rich sediments, also in phosphate deposits, ore veins and pegmatites. Beautiful hemispherical aggregates up to 40 mm (1 9/16 in) in diameter occur in Pencil, Garland and Magnet Cove, Arkansas, USA; also known from Trenice and Milina, Czech Republic and Ronneburg, Germany.

Eosphorite

$MnAl(PO_4)(OH)_2 \cdot H_2O$

ORTHORHOMBIC ● ●

Properties: C – pinkish, colorless, white, brownish, red-brown; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 3.1; H – 5; CL – imperfect; F – uneven to conchoidal; M – long to short prismatic crystals, radial aggregates, granular.
Origin and occurrence: Secondary in granitic pegmatites, as a product of hydrothermal replacement of primary phosphates. Crystals up to 100 mm (4in) long found in the Joao Modesto dos Santos mine, Minas Gerais, Brazil. It occurs in Rapid Creek, Yukon, Canada, too.

Turquoise

$CuAl_6(PO_4)_4(OH)_8 \cdot 4\,H_2O$

TRICLINIC ● ● ●

Properties: C – blue, blue-green, green; S – white; L – waxy; D – transparent, translucent to opaque; DE – 2.9; H – 5-6; CL – good; F – conchoidal to uneven; M – short prismatic crystals, botryoidal aggregates, coatings, massive.
Origin and occurrence: Secondary in the surface parts of rocks with elevated contents of P and Cu, e.g. in the oxidation zone of some Cu deposits. Small crystals occurred near Lynch Station, Virginia, USA. Massive blue and blue-green concretions come from Mount Ali Mirsai near Maden, Iran. Other localities are Cortez, Nevada, Los Cerillos and Eureka, New Mexico and Bisbee, Arizona, USA.
Application: popular gemstone.

Turquoise, 3 mm aggregates, Humboldt Co., U.S.A.

Chalcosiderite

$CuFe^{3+}{}_6(PO_4)_4(OH)_8 \cdot 4\,H_2O$

TRICLINIC ● ●

Properties: C – dark green; S – white; L – vitreous; D – transparent to translucent; DE – 3.3; H – 4.5; CL – good; F – conchoidal to uneven; M – short prismatic crystals, coatings.

Origin and occurrence: Secondary in the oxidation zone of some Cu deposits, together with goethite, dufrenite and pharmacosiderite. It occurs in Bisbee, Arizona, USA; in the Wheal Phoenix mine, Cornwall, UK; Schneckenstein, Germany; Horni Slavkov, Czech Republic.

Turquoise, 50 mm, Kazakhstan

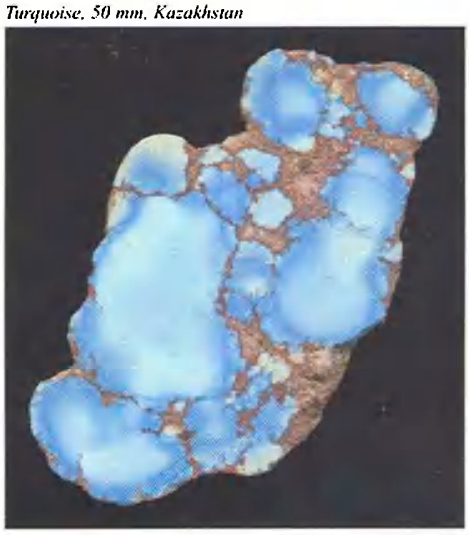

Chalcosiderite, 90 mm, Cornwall, UK

Chenevixite, 100 mm, Chuquicamata, Chile

Chenevixite

$Cu_2Fe^{3+}{}_2(AsO_4)_2(OH)_4 \cdot H_2O$

MONOCLINIC ●

Properties: C – dark green, olive-green to yellow-green; S – yellow-green; L – greasy; D – translucent; DE – 3.9; H – 3.5-4.5; CL – not determined; F – conchoidal to uneven; M – earthy aggregates, coatings, massive.
Origin and occurrence: Secondary in the oxidation zone of Cu deposits, associated with malachite, tyrolite, azurite and other minerals. Massive aggregates occur in the Mammoth mine, Tintic, Utah, USA; also in Klein Spitzkopje, Namibia.

Tyrolite

$CaCu_5(AsO_4)_2(CO_3)(OH)_4 \cdot 6\ H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – apple-green, green-blue to blue; S – light green to blue-green; L – vitreous to pearly; D – transparent to translucent; DE – 3.3; H – 2; CL – perfect; F – uneven; M B scaly and fan-shaped aggregates, coatings and crusts.
Origin and occurrence: Secondary in the oxidation zone of Cu deposits, frequently associated with chalcophyllite. Rich aggregates occur in the Majuba Hill mine, Nevada, also in Tintic, Utah, USA. It is also known from Brixlegg, Austria; Saalfeld and Schneeberg, Germany; Novoveska Huta, Slovakia.

Tyrolite, 90 mm, Novoveska Huta, Slovakia

Delvauxite

$CaFe^{3+}{}_4(PO_4,SO_4)(OH)_8 \cdot 4\text{-}6\ H_2O$

AMORPHOUS ● ● ●

Properties: C – yellow-brown, brown, red-brown, black-brown; S – yellow; L – greasy, waxy; D – translucent to opaque; DE – 1.8-2.0; H – 2.5; CL – not determined; F – conchoidal, earthy; M – nodules, stalactites, coatings and crusts, massive.
Origin and occurrence: Secondary in oxidation zone of Fe deposits. Nodules of 50 cm (20 in) across in Czech Republic. Also known in Berneau and Richelle, Belgium; Zelezník, Slovakia; Kerch, Crimea, Ukraine.

Delvauxite, 80 mm, Nucice, Czech Republic

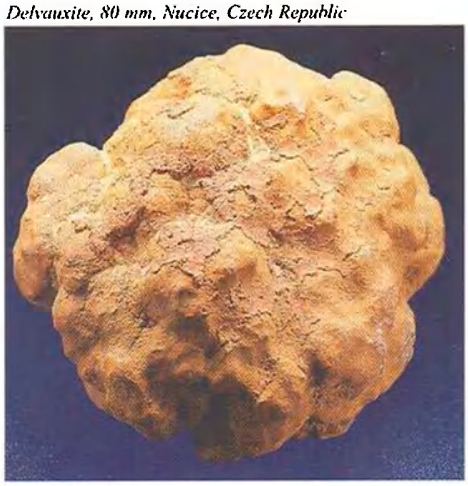

Bukovskyite, 40 mm, Kaňk, Czech Republic

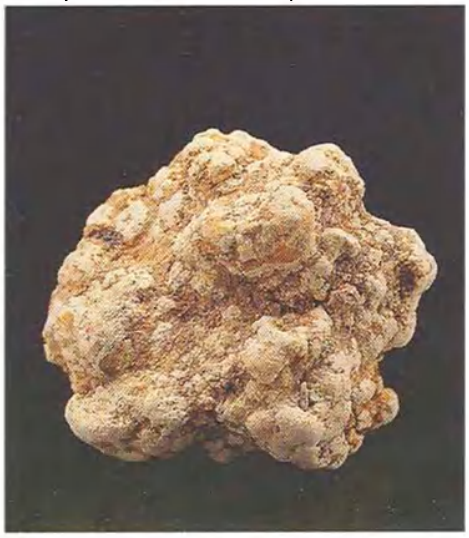

Bukovskyite

$Fe^{3+}{}_2(AsO_4)(SO_4)(OH) \cdot 7\ H_2O$

TRICLINIC ●

Properties: C – yellow-green, gray-green; S – yellowish white; L – dull to earthy; D – translucent to opaque; DE – 2.3; H – not determined; CL – not determined; F – earthy; M B botryoidal aggregates and nodules.
Origin and occurrence: Secondary on the old mine dumps where it forms as a product of arsenopyrite weathering. Nodules up to 60 cm (24 in), occur on medieval dumps in Kank near Kutna Hora, Czech Republic.

Chalcophyllite, 36 mm, Chile

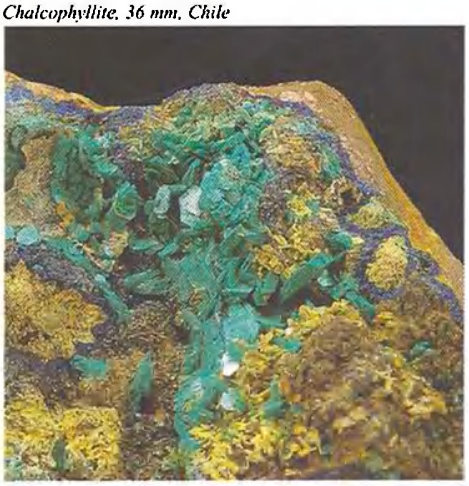

Veszelyite

$(Cu,Zn)_3(PO_4)(OH)_3 \cdot 2\ H_2O$

MONOCLINIC ● ●

Properties: C – green, blue-green, dark blue; S – green; L – vitreous; D – translucent; DE – 3.4; H – 3.5-4; CL – good; F – uneven; M – short prismatic to tabular crystals, granular.
Origin and occurrence: Secondary in the oxidation zone of Cu-Zn deposits. Crystals up to 50 mm (2 in) across found in the Black Pine mine, Philipsburg, Montana, USA; also from Moravita, Romania; Arakawa, Japan; Wanlockhead, Scotland, UK.

Chalcophylite

$Cu_{18}Al_2(AsO_4)_3(SO_4)_3(OH)_{27} \cdot 36\ H_2O$

TRIGONAL ● ● ●

Properties: C – emerald-green, blue-green; S – light green; L – vitreous, pearly; D – transparent to translucent; DE – 2.6; H – 2; CL – perfect; F – uneven; M – tabular crystals, scaly, fan-shaped aggregates, coatings, massive.
Origin and occurrence: Secondary in the oxidation zone of Cu deposits, usually associated with tyrolite. Rich aggregates occur in the Majuba Hill, Nevada and in the Tintic district, Utah, USA. Nice specimens also come from Novoveska Huta and Piesky, Slovakia; Nizhniy Tagil, Ural mountains, Russia; Cap Garrone, France.

Veszelyite, 13 mm, Philipsburg, U.S.A.

Liroconite, 86 mm, Cornwall, UK

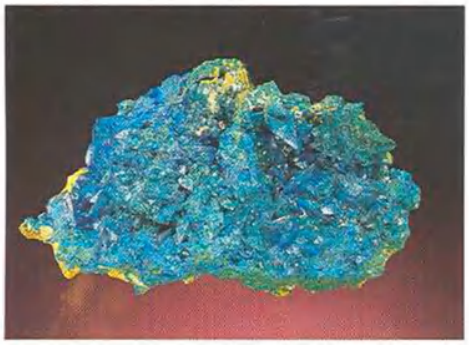

Lirokonit

$Cu_2Al(AsO_4)(OH)_4 \cdot 4\,H_2O$

MONOCLINIC ● ●

Properties: C – blue, green; S – light blue; L – vitreous to resinous; D – transparent to translucent; DE – 3.0; H – 2-2.5; CL – imperfect; F – conchoidal to uneven; M – lenticular, dipyramidal crystals, massive.
Origin and occurrence: Secondary in the oxidation zone of Cu deposits, together with olivenite, malachite and azurite. Perfect crystals up to 30 mm (1 3/16 in) across come from Redruth and St. Day, Cornwall, UK. It is also known from Cerro Gordo, California, USA.

Evansite

$Al_3(PO_4)(OH)_6 \cdot 6\,H_2O$

AMORPHOUS 3

Properties: C – colorless, white, greenish, light blue-green, yellowish; S – white; L – vitreous, resinous, waxy; D – transparent to translucent; DE – 1.8-2.2; H – 3-4; CL – none; F – conchoidal; M – botryoidal, stalactitic and hemispherical aggregates and coatings.
Origin and occurrence: Secondary in the oxidation zone of Fe deposits, rich in P, associated with alophane and goethite. Rich stalactitic aggregates come from Zelezník, Slovakia and from Épernay, France.

Whiteite-(CaFeMg)

$CaFeMg_2Al_2(PO_4)_4(OH)_2 \cdot 8\,H_2O$

MONOCLINIC ● ●

Properties: C – brown; S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 4; CL – good; F – uneven; M – short prismatic crystals.
Origin and occurrence: Hydrothermal in the cracks of P-rich sediments. Perfect crystals up to 20 mm (25/32 in) across come from Big Fish River, Yukon, Canada. Crystals up to 5 mm (3/16 in), are known from Lavra da Ilha de Taquaral, Minas Gerais, Brazil.

Evansite, 5 mm aggregates, Sirk – Zeleznik, Slovakia

Whiteite-(CaFeMg), 65 mm, Yukon Territory, Canada

Jahnsite-(CaMnMg), 60 mm, Custer, U.S.A.

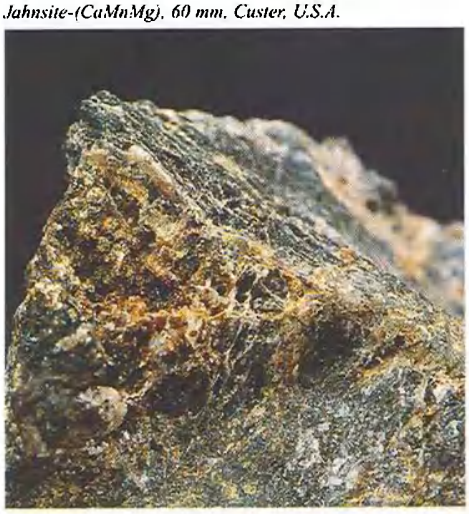

Jahnsite-(CaMnMg)

$CaMnMg_2Fe^{3+}{}_2(PO_4)_4(OH)_4 \cdot 8\,H_2O$

MONOCLINIC ● ●

Properties: C – yellow, light to dark brown; S – light yellow; L – vitreous; D – transparent to translucent; DE – 2.6; H – not determined; CL – good; F – uneven; M – short to long prismatic crystals, granular.
Origin and occurrence: Secondary in granitic pegmatites, where it forms as a result of replacement of primary phosphates. Perfect crystals up to 10 mm (⅜ in) across occur in the Tip Top mine, Custer, South Dakota, USA. It is also known from Hagendorf, Germany.

Wardite

$NaAl_3(PO_4)_2(OH)_4 \cdot 2\,H_2O$

TETRAGONAL ● ● ●

Properties: C – colorless, white, greenish, light blue-green; S -white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 5; CL – perfect; F – uneven; M – dipyramidal crystals, radial and hemispherical aggregates, coatings, crusts, granular.
Origin and occurrence: Secondary in granitic pegmatites, where it forms as a product of primary phosphate replacement; hydrothermal in the cracks in P-rich sediments. Perfect crystals up to 30 mm (1¼ in) across come from Rapid Creek, Yukon, Canada. It is also known from Piedras Lavradas, Paraiba, Brazil.

Wardite, 48 mm, Yukon Territory, Canada

Cyrilovite

$NaFe^{3+}{}_3(PO_4)_2(OH)_4 \cdot 2\,H_2O$

TETRAGONAL ● ●

Properties: C – yellow, orange, brown-yellow; S – yellow; L – vitreous; D – transparent to translucent; DE – 3.1; H – 4; CL – good; F – conchoidal; M – tabular and dipyramidal crystals, coatings and crusts.
Origin and occurrence: Secondary in granitic pegmatites, where it forms as a result of replacement of primary phosphates. Small crystals occur in Cyrilov, Czech Republic; Hagendorf, Germany and Sapucaia, Minas Gerais, Brazil.

Cyrilovite, 60 mm, Iron Monarch, Australia

Liroconite, 86 mm, Cornwall, UK

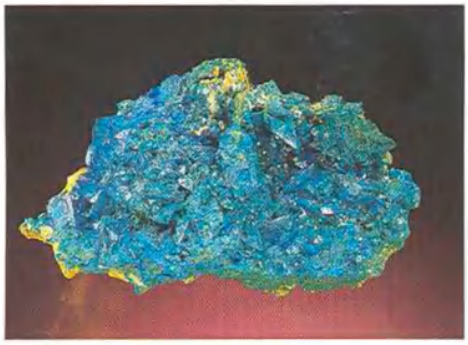

Lirokonit

$Cu_2Al(AsO_4)(OH)_4 \cdot 4\,H_2O$

MONOCLINIC ● ●

Properties: C – blue, green; S – light blue; L – vitreous to resinous; D – transparent to translucent; DE – 3.0, H – 2-2.5; CL – imperfect; F – conchoidal to uneven; M – lenticular, dipyramidal crystals, massive.
Origin and occurrence: Secondary in the oxidation zone of Cu deposits, together with olivenite, malachite and azurite. Perfect crystals up to 30 mm (1 3/16 in) across come from Redruth and St. Day, Cornwall, UK. It is also known from Cerro Gordo, California, USA.

Evansite

$Al_3(PO_4)(OH)_6 \cdot 6\,H_2O$

AMORPHOUS 3

Properties: C – colorless, white, greenish, light blue-green, yellowish; S – white; L – vitreous, resinous, waxy; D – transparent to translucent; DE – 1.8-2.2; H – 3-4; CL – none; F – conchoidal; M – botryoidal, stalactitic and hemispherical aggregates and coatings.
Origin and occurrence: Secondary in the oxidation zone of Fe deposits, rich in P, associated with alophane and goethite. Rich stalactitic aggregates come from Zeleznik, Slovakia and from Epernay, France.

Whiteite-(CaFeMg)

$CaFeMg_2Al_2(PO_4)_4(OH)_2 \cdot 8\,H_2O$

MONOCLINIC ● ●

Properties: C – brown; S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 4; CL – good; F – uneven; M – short prismatic crystals.
Origin and occurrence: Hydrothermal in the cracks of P-rich sediments. Perfect crystals up to 20 mm (25/32 in) across come from Big Fish River, Yukon, Canada. Crystals up to 5 mm (3/16 in), are known from Lavra da Ilha de Taquaral, Minas Gerais, Brazil.

Evansite, 5 mm aggregates, Sirk – Zeleznik, Slovakia

Whiteite-(CaFeMg), 65 mm, Yukon Territory, Canada

Jahnsite-(CaMnMg), 60 mm, Custer, U.S.A.

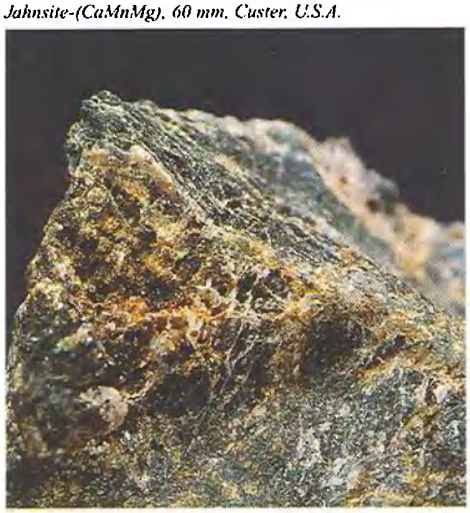

Jahnsite-(CaMnMg)

$CaMnMg_2Fe^{3+}{}_2(PO_4)_4(OH)_4 \cdot 8\ H_2O$

MONOCLINIC ● ●

Properties: C – yellow, light to dark brown; S – light yellow; L – vitreous; D – transparent to translucent; DE – 2.6; H – not determined; CL – good; F – uneven; M – short to long prismatic crystals, granular.
Origin and occurrence: Secondary in granitic pegmatites, where it forms as a result of replacement of primary phosphates. Perfect crystals up to 10 mm (3/8 in) across occur in the Tip Top mine, Custer, South Dakota, USA. It is also known from Hagendorf, Germany.

Wardite

$NaAl_3(PO_4)_2(OH)_4 \cdot 2\ H_2O$

TETRAGONAL ● ● ●

Properties: C – colorless, white, greenish, light blue-green; S -white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 5; CL – perfect; F – uneven; M – dipyramidal crystals, radial and hemispherical aggregates, coatings, crusts, granular.
Origin and occurrence: Secondary in granitic pegmatites, where it forms as a product of primary phosphate replacement; hydrothermal in the cracks in P-rich sediments. Perfect crystals up to 30 mm (1 3/16 in) across come from Rapid Creek, Yukon, Canada. It is also known from Piedras Lavradas, Paraiba, Brazil.

Wardite, 48 mm, Yukon Territory, Canada

Cyrilovite

$NaFe^{3+}{}_3(PO_4)_2(OH)_4 \cdot 2\ H_2O$

TETRAGONAL ● ●

Properties: C – yellow, orange, brown-yellow; S – yellow; L – vitreous; D – transparent to translucent; DE – 3.1; H – 4; CL – good; F – conchoidal; M – tabular and dipyramidal crystals, coatings and crusts.
Origin and occurrence: Secondary in granitic pegmatites, where it forms as a result of replacement of primary phosphates. Small crystals occur in Cyrilov, Czech Republic; Hagendorf, Germany and Sapucaia, Minas Gerais, Brazil.

Cyrilovite, 60 mm, Iron Monarch, Australia

Pharmacosiderite, 35 mm, Cornwall, UK

Arseniosiderite, 3 mm, Nagyborzsony, Hungary

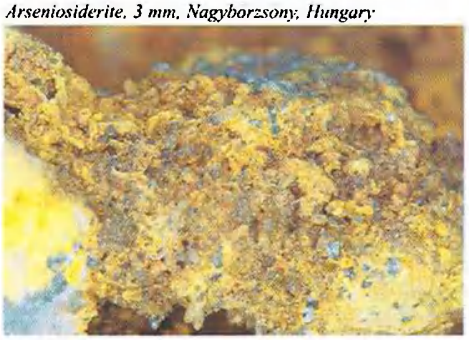

Lavendulan, 36 mm, San Juan, Chile

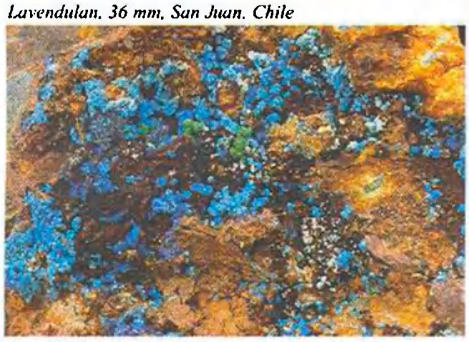

Mixite, 70 mm, Cínovec, Czech Republic

Pharmacosiderite

$KFe^{3+}_4(AsO_4)_3(OH)_4 \cdot 6\ H_2O$

CUBIC ● ● ●

Properties: C – green, yellow-brown, brown; S – white; L – adamantine to greasy; D – transparent to translucent; DE – 2.8; H – 2.5; CL – imperfect; F – uneven; M – cubic crystals, coatings, crusts, granular, massive.
Origin and occurrence: Secondary in granitic pegmatites and in the oxidation zone of ore deposits, where it forms as a product of arsenopyrite and lollingite replacement. Cubic crystals up to 10 mm (3/8 in) across occur in St. Day, Liskeard and Redruth, Cornwall, UK. It also comes from the Majuba Hill mine, Nevada, USA; Horhausen, Germany and Cap Garrone, France.

Arseniosiderite

$CaFe^{3+}_3(AsO_4)_3O_2 \cdot 3\ H_2O$

MONOCLINIC ● ●

Properties: C – yellow, light to dark brown; S – yellow; L – submetallic to silky; D – opaque; DE – 3.6; H – 1.5; CL – good; F – uneven; M – fibrous and earthy aggregates, coatings, crusts, massive.
Origin and occurrence: Secondary in the oxidation zone of ore deposits and in granitic pegmatites, where it forms as a product of arsenopyrite and lollingite replacement. Rich aggregates occur in Romanche, France; also from the Eureka mine, Tintic, Utah, USA and Wittichen, Germany.

Lavendulane

$NaCaCu_5(AsO_4)_4Cl \cdot 5\ H_2O$

ORTHORHOMBIC ● ●

Properties: C – light blue-purple, blue; S – white; L – vitreous to waxy; D – translucent; DE – 3.5; H – 2.5; CL – good; F – uneven; M – acicular crystals, acicular and earthy aggregates, coatings.
Origin and occurrence: Secondary in the oxidation zone of Co, Cu and Ni deposits, as a result of arsenide weathering, associated with erythrite. Rich acicular aggregates come from Talmessi and Anarak, Iran. Small crystals found in the Blanca mine, San Juan, Chile and in Annaberg, Germany. Crystals up to 4 mm (5/32 in) across occur in the Gold Hill mine, Utah, USA.

Kovdorskite, 15 mm x, Kovdor, Kola, Russia

Mixite

$BiCu_6(AsO_4)_3(OH)_6 \cdot 3\ H_2O$

HEXAGONAL ● ●

Properties: C – emerald-green, blue-green, light blue, light green, whitish; S – light green, light blue; L – adamantine to dull; D – translucent; DE – 3.8; H – 3-4; CL – not determined; F – uneven; M – acicular crystals, acicular and earthy aggregates, coatings, massive.
Origin and occurrence: Secondary in the oxidation zone of Bi and Cu deposits, together with bismuthinite. It occurred in Jachymov, Czech Republic; in Tintic, Utah, USA; Schneeberg and Wittichen, Germany.

Kovdorskite

$Mg_5(PO_4)_2(CO_3)(OH)_2 \cdot 4{,}5\ H_2O$

MONOCLINIC ●

Properties: C – light pink-brown, white, blue; S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 4; CL – none; F – conchoidal to uneven; M – tabular crystals, massive.
Origin and occurrence: Hydrothermal, associated with magnesite, magnetite and other minerals. Blue and pink-brown crystals up to 25 mm (1 in) across come from the Zhleznyi mine, Kovdor, Kola Peninsula, Russia.

Torbernite

$Cu(UO_2)_2(PO_4)_2 \cdot 8\text{-}12\ H_2O$

TETRAGONAL ● ● ●

Properties: C – emerald-green to grass-green; S – light green; L – vitreous to dull, pearly on the cleavage planes; D – transparent to translucent; DE – 3.3; H – 2-2.5; CL – perfect; F – uneven; M – tabular to pyramidal crystals, earthy aggregates, coatings, granular, massive; R – strongly radioactive.
Origin and occurrence: Secondary in the oxidation zone of U deposits, also in pegmatites and sedimentary rocks, resulting from the hydrothermal alteration of uraninite and other U minerals. Emerald-green tabular crystals, several cm across, come from Sabugal, Portugal; Jachymov, Czech Republic; Shinkolobwe, Zaire; Bois-Noirs, France; Moctezuma, Mexico and many localities in the Colorado Plateau, Utah, USA. Beautiful druses of crystals up to 20 mm ($^{25}/_{32}$ in) across, found in the Margabal mine, Aveyron, France.
Application: U ore.

Torbernite, 120 mm, Katanga, Zair

Autunite, 15 mm xx, Troncasa, Portugal

Novacekite, 70 mm, Brumado, Brazil

Autunite

$Ca(UO_2)_2(PO_4)_2 \cdot 10\text{-}12\ H_2O$

TETRAGONAL ● ● ●

Properties: C – light to dark yellow, yellow-green to green; S – light yellow; L – vitreous to dull; D – transparent to translucent; DE – 3.1; H – 2-2.5; CL – perfect; F – uneven; M – tabular crystals, foliated, scaly, earthy and pulverulent aggregates, coatings, massive; LU – yellow-green; R – strongly radioactive.

Origin and occurrence: Secondary in the oxidation zone of U deposits, in pegmatites and in some U-rich sedimentary rocks, as a result of hydrothermal alteration of uraninite and other U minerals. It is frequently associated with torbernite and other U secondary minerals. Tabular crystals up to 30 mm (1 3/16 in) across come from Schneeberg and Johanngeorgenstadt, Germany and Autun, France. It is also known from Rum Jungle, Northern Territory, Australia; St. Austel, Cornwall, UK; Mount Spokane, Washington, USA and Jachymov, Czech Republic.

Application: U ore.

Uranocircite

$Ba(UO_2)_2(PO_4)_2 \cdot 10\ H_2O$

TETRAGONAL ● ●

Properties: C – light to dark yellow, light yellow-green; S – light yellow; L – vitreous to dull, pearly on the cleavage planes; D – transparent to translu-

Uranocircite, 78 mm, Minas Gerais, Brazil

Zeunerite, 5 mm xx, Cínovec, Czech Republic

cent; DE – 3.5; H – 2-2.5; CL – perfect; F – uneven; M – tabular crystals, foliated and earthy aggregates, pulverulent coatings, massive; LU – green; R – strongly radioactive.
Origin and occurrence: Secondary in the oxidation zone of U deposits. Yellow tabular crystals up to 10 mm (3/8 in) across occurred in Dametice, Czech Republic; Bergen and Wolsendorf, Germany and in the Sao Pedro mine, Suaçui, Minas Gerais, Brazil.

Novacekite
$Mg(UO_2)_2(AsO_4)_2 \cdot 10\ H_2O$

TETRAGONAL ● ●

Properties: C – straw-yellow, light yellow; S – light yellow; L – vitreous to dull; D – transparent to translucent; DE – 3.7; H – 2.5; CL – perfect; F – uneven; M – tabular crystals, lamellar, earthy and pulverulent aggregates, massive; LU – dark green; R – strongly radioactive.
Origin and occurrence: Secondary in the oxidation zone of U deposits. Tabular crystals up to 50 mm (2 in) across come from the Pedra Preta Mine, Brumado, Bahia, Brazil. Lamellar aggregates are known from Zalesi, Czech Republic; Aldama, Chihuahua, Mexico; Wittichen, Germany.

Zeunerite
$Cu(UO_2)_2(AsO_4)_2 \cdot 10\text{-}16\ H_2O$

TETRAGONAL ● ●

Properties: C – emerald-green, yellow-green; S – light green; L – vitreous to dull; D – transparent to translucent; DE – 3.4; H – 2.5; CL – perfect; F – uneven; M – tabular crystals, foliated aggregates, massive; R – strongly radioactive.
Origin and occurrence: Secondary in the oxidation zone of U deposits. Tabular crystals up to 30 mm (1 3/16 in) across, come from the Pedra Preta Mine, Brumado, BahRa, Brazil. It is also known from Zalesí, Czech Republic and Schneeberg, Germany.

Carnotite
$K_2(UO_2)_2(VO_4)_2 \cdot 3\ H_2O$

MONOCLINIC ● ● ●

Properties: C – light to dark yellow, yellow-green; S – light yellow; L – dull; D – transparent to translucent, opaque; DE – 4.9; H – not determined; CL – perfect; F – uneven; M – earthy aggregates, massive; R – strongly radioactive.
Origin and occurrence: Secondary in the oxidation zone of sedimentary U deposits, typically associated

Carnotite, 55 mm, Utah, U.S.A.

with tyuyamunite. Platy crystals, up to 2 mm (1/16 in) across found in the Mashamba West mine, Zaire. Earthy and pulverulent aggregates occur in many localities in the Colorado Plateau, e.g. Paradox Valley, Colorado and La Sal, Utah, USA. Also known from Tyuya Muyun, Uzbekistan and Radium Hill, Southern Australia, Australia.
Application: U and V ore..

Tyuyamunite
$Ca(UO_2)_2(VO_4)_2 \cdot 5\text{-}8\ H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – yellow-green, canary yellow; S – light yellow; L – silky to adamantine; D – translucent to opaque; DE – 3.6; H – 2; CL – perfect; F – uneven; M – earthy aggregates, massive; LU – weak yellow-green; R – strongly radioactive.
Origin and occurrence: Secondary in the oxidation zone of sedimentary U deposits, together with carnotite. Common earthy and pulverulent aggregates occur in many localities in the Colorado Plateau, e.g. Paradox Valley, Colorado and Red Creek, Utah, USA. It was described from Tyuya Muyun, Uzbekistan.
Application: U and V ore.

Tyuyamunite, 80 mm, Fergana, Uzbekistan

9. Silicates

Phenakite

Be_2SiO_4

TRIGONAL ● ● ●

Properties: C – colorless, white, yellowish; S – white; L – strong vitreous; D – transparent to translucent; DE – 3.0; H – 8; CL – imperfect; F – conchoidal; M – long prismatic to tabular crystals, radial aggregates, granular.
Origin and occurrence: Magmatic in granitic pegmatites; hydrothermal in greisens; metamorphic in mica schists, associated with beryl, chrysoberyl, apatite and quartz. Prismatic crystals up to 250 mm (9 13/16 in) long occurred in Kragero, Norway. It is also known from Sao Miguel de Piracicaba, Minas Gerais, Brazil in crystals, up to 100 mm (4 in) long. The other localities are Habachtal, Austria; Malyshevo, Russia; Anjanabonoina, Madagascar.
Application: sporadically cut as a gemstone.

Phenakite, 200 mmx, Kragero, Norway

Willemite

Zn_2SiO_4

TRIGONAL ● ● ●

Properties: C – white, yellowish, gray, green; S – white; L – vitreous; D – translucent; DE – 4.0; H – 5.5; CL – good; F B conchoidal to uneven; M – prismatic to tabular crystals, radial aggregates, granular; LU – distinct light green.
Origin and occurrence: Metamorphic in marbles; secondary in the oxidation zone of ore deposits, associated with zincite, franklinite, hemimorphite and smithsonite. Crystals up to 100 mm (4 in) across come from Franklin, New Jersey, USA, Mont St.-Hilaire, Quebec, Canada.
Application: as Zn ore.

Topaz, 60 mm, Thomas Range, U.S.A.
Willemite, 76 mm, Tsumeb, Namibia

Forsterite, 40 mm, Suppat, Pakistan

Forsterite

OLIVINE GROUP

Mg_2SiO_4

ORTHORHOMBIC ● ● ● ●

Properties: C – yellowish, greenish, colorless; S – white; L – vitreous; D – transparent to opaque; DE – 3.3; H – 6.5-7; CL – good; F – conchoidal to uneven; M – tabular to prismatic crystals, granular.

Origin and occurrence: Metamorphic in regionally and contact metamorphosed dolomites. Typical rock-forming mineral, associated with enstatite, spinel, phlogopite and chlorite. Green gemmy crystals up to 80 mm (3 1/8 in) long come from Suppat, Pakistan. It also occurred in Crestmore, California, USA, Mount Timobly, British Columbia, Canada and Monte Somma, Italy.

Olivine

OLIVINE GROUP

$(Mg,Fe)_2SiO_4$

ORTHORHOMBIC ● ● ● ● ●

Varieties: chrysolite

Properties: C – green, yellow-green (chrysolite), brown-green to black-green; S – white; L – vitreous; D – translucent to opaque; DE – 3.3-3.6; H – 6.5-7; CL – good; F B conchoidal to uneven; M – imperfect crystals, granular.

Origin and occurrence: Magmatic in some ultra-basic rocks, e.g. dunites, lherzolites, peridotites, gabbros and in meteorites. Typical rock-forming mineral, usually associated with diopside, magnetite and pyrope. Classic locality of chrysolite is Zebirget Island in the Red Sea, Egypt, where tabular crystals

Olivine, 60 mm, Smrci, Czech Republic

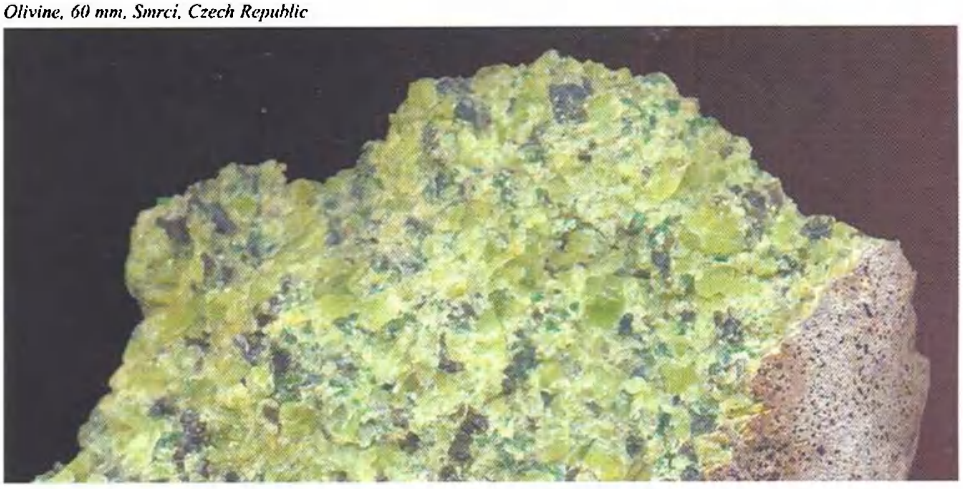

up to 100 mm (4 in) across occurred. It is also known from the San Carlos Indian Reservation, Arizona, USA. Olivine was found in many basaltic rocks in Laacher See, Germany; Rockport, Massachusetts, USA and elsewhere.
Application: chrysolite is cut as a gemstone.

Fayalite
OLIVINE GROUP
Fe_2SiO_4

ORTHORHOMBIC ● ● ●

Properties: C – black-green to black; S – gray; L – dull to vitreous; D – opaque; DE – 4.2; H – 6.5-7; CL – good; F B conchoidal to uneven; M – imperfect prismatic crystals, granular.
Origin and occurrence: Magmatic in granitic pegmatites, granites and syenites, associated with orthoclase, gadolinite-(Y) and epidote; rare metamorphic. Poorly developed crystals up to 150 mm (6 in) long found in pegmatites near Baveno, the Alps, Italy.

Fayalite, 80 mm, Rockport, U.S.A.

It is also known from Strzegom, Poland and the Sawtooth Batholith, Idaho, USA.

Tephroite
OLIVINE GROUP
Mn_2SiO_4

ORTHORHOMBIC ● ●

Properties: C – gray, olive-green, red-brown; S – white; L – dull to vitreous; D – translucent to transparent; DE – 4.2; H – 6; CL – good; F B conchoidal to uneven; M – prismatic crystals, granular.
Origin and occurrence: Metamorphic in skarns and Mn-rich metamorphosed sediments, together with rhodonite, franklinite and spessartine. Granular aggregates and perfect crystals up to 50 mm (2 in) across known from Franklin, New Jersey, USA; Langban, Sweden. It comes also from Tarnobrzeg, Poland, in crystals, up to 80 mm (3 1/8 in) across.

Tephroite, 60 mm, Harstigen, Sweden

Pyrope, 3 mm grains, Trebivlice, Czech Republic

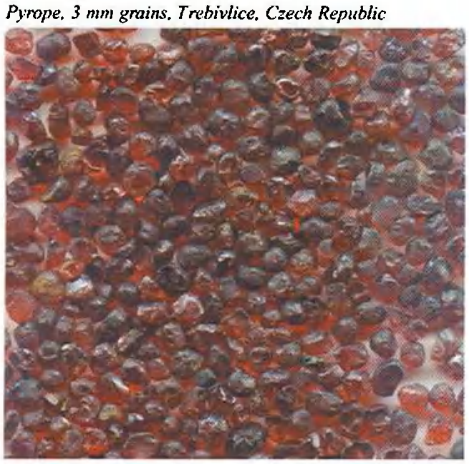

Pyrope

GARNET GROUP

$Mg_3Al_2Si_3O_{12}$

CUBIC ● ● ● ●

Properties: C – red to purple-red, light purple, black-brown; S – white; L – vitreous; D – transparent to translucent; DE – 3.5; H – 7-7.5; CL – none; F B conchoidal to uneven; M – isometric crystals, granular.

Origin and occurrence: Magmatic in some ultrabasic rocks, e.g. lherzolites, peridotites, kimberlites, eclogites and serpentinites; metamorphic in quartzites; also known from placers. It is associated with diopside, magnetite and diamond. It comes from many localities in ultrabasic rocks, like Trebenice and Mirunice, Czech Republic; Zoblitz, Germany; Madras, India; Kimberley, South Africa. Crystals up to 250 mm ($9^{13}/_{16}$ in) across were found in Dora Maria, the Alps, Italy.

Application: cut as a gemstone.

Almandine

GARNET GROUP

$Fe_3Al_2Si_3O_{12}$

CUBIC ● ● ● ●

Properties: C – red to purple-red, black-brown; S – white; L – vitreous; D – transparent to translucent; DE – 4.3; H – 7; CL – none; F B conchoidal to uneven; M B well-formed crystals, granular.

Origin and occurrence: Metamorphic in regionally metamorphosed rocks, as chlorite schists, gneisses,

Almandine, 38 mm x, Ötztal, Austria

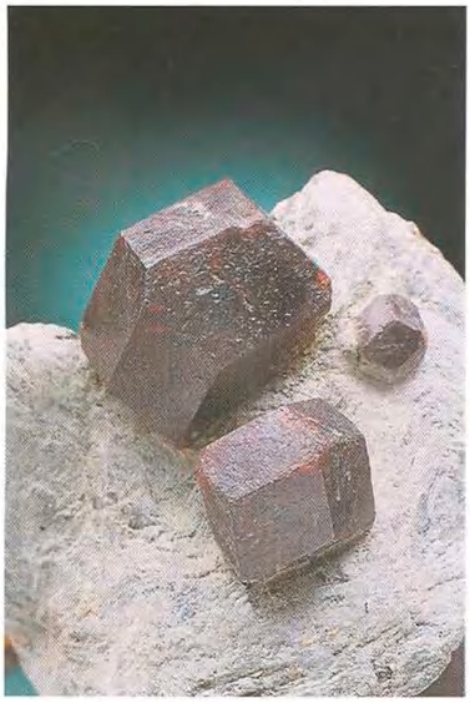

Spessartine, 96 mm, Gilgit, Pakistan

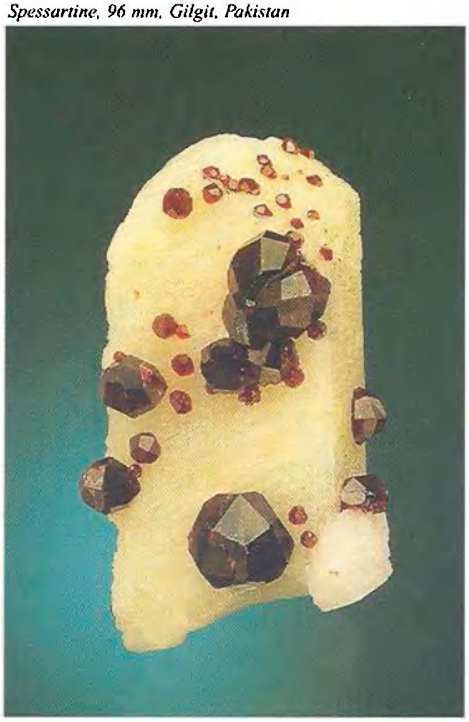

Grossular, 17 mm x, Asbestos, Canada

mica schists and migmatites; magmatic in some granites and pegmatites; also in placers. Well-developed crystals up to 150 mm (6 in) across are known from Ishikawa pegmatites, Japan and Shingus, Pakistan. It comes from many mica schists and gneisses as crystals, up to about 50 mm (2 in), like Fort Wrangel, Alaska, USA; Ötztal, Austria; Bodo, Norway. It occurs in placers near Ratnapura, Sri Lanka.
Application: cut as a gemstone, abrasive material.

Spessartine
GARNET GROUP
$Mn_3Al_2Si_3O_{12}$

CUBIC ● ● ● ●

Properties: C – red, orange, light brown to yellowish; S – white; L – vitreous; D – transparent to translucent; DE – 4.3; H – 7-7.5; CL – none; F B conchoidal to uneven; M – perfect crystals, granular.
Origin and occurrence: Magmatic in granitic pegmatites and some granites; hydrothermal in cavities in rhyolites; metamorphic in some skarns and Mn-rich metamorphic rocks. Perfect crystals up to 30 mm (1 3/16 in) across were found in granitic pegmatites in the Hercules mine, Ramona, California, USA; near Marienfluss river, Namibia and in rhyolite cavities in Nathrop, Colorado, USA.
Gemmy crystal fragments up to 50 mm (2 in) across were recently found in an undisclosed locality in Minas Gerais, Brazil.
Application: cut as a gemstone.

Grossular
GARNET GROUP
$Ca_3Al_2Si_3O_{12}$

CUBIC ● ● ● ●

Varieties: hessonite, tsavorite

Properties: C – red, green (tsavorite), orange, red-brown (hessonite) to colorless; S – white; L – vitreous; D – transparent to tranlucent; DE – 3.4; H – 6.5-7; CL – none; F B conchoidal to uneven; M – perfect crystals, granular.
Origin and occurrence: Metamorphic in Ca-rich, contact metamorphic rocks, skarns, rodingites; hydrothermal along the cracks in these rocks, associated with diopside, vesuvianite, wollastonite, scapolite and epidote. Perfect crystals about 30 mm (1 3/16 in) across occur in the Jeffrey quarry, Asbestos, Quebec, Canada and Sierra de las Cruces, Coahuila, Mexico.
Tsavorite crystals up to 50 mm (2 in) across come from the Tsavo National Park, Kenya and Merelani Hills, Arusha, Tanzania. The other well-known localities are Ala, Italy and Ciclova, Romania. Grossular crystals up to 100 mm (4 in) in size in Vapenna, Czech Republic; Xalostoc, Mexico and Sandare, Mali.
Application: cut as a gemstone.

Andradite, 30 mm xx, Graham Co., U.S.A.

Andradite

GARNET GROUP

$Ca_3Fe^{3+}{}_2Si_3O_{12}$

CUBIC ● ● ● ●

Varieties: demantoid, melanite

Properties: C – dark red, black-brown, brown, green to yellow-green (demantoid), black-brown to black (melanite); S – white; L – vitreous; D – transparent to translucent; DE – 3.9; H – 6.5-7; CL – none; F B conchoidal to uneven; M B well-developed crystals, granular.

Origin and occurrence: Metamorphic in Ca and Fe rich contact metamorphosed rocks, skarns, rodingites; hydrothermal along the cracks of these rocks; magmatic in some alkaline igneous rocks, usually in the same localities as grossular. Crystals up to 40 mm (1 9/16 in) across come from Sinerechenskoye, Russia. Fine crystals were also found in the Namgar mine, Usakos, Namibia and Ciclova, Romania. Fine melanite crystals occur in Magnet Cove, Arkansas, USA. Demantoid crystals up to 30 mm (1 3/16 in) across known from Val Malenco, Italy and in the Bobrovka river basin, Ural mountains, Russia.

Application: demantoid is cut as a gemstone.

Demantoid, 3 mm xx, Bobrovka, Russia

Melanite, 5 mm xx, Rudnyi, Kazakhstan

Uvarovite, 2 mm xx, Sarany, Russia

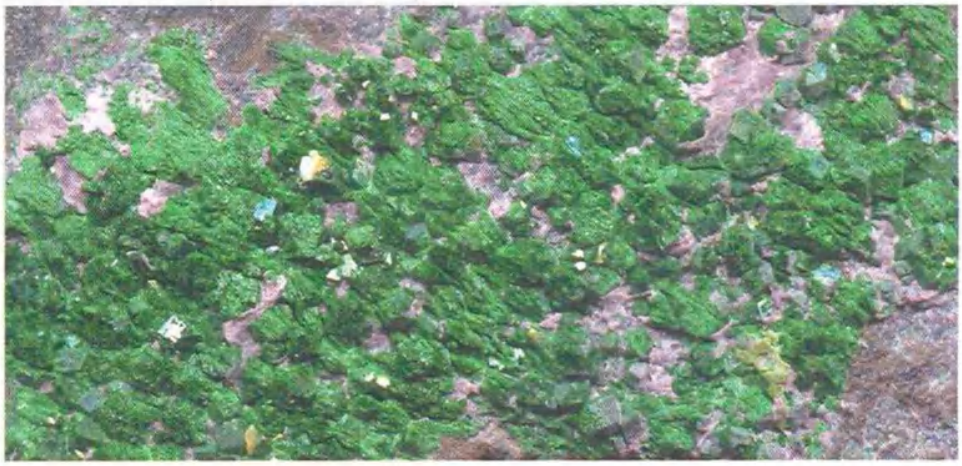

Uvarovite

GARNET GROUP

$Ca_3Cr_2Si_3O_{12}$

CUBIC ● ●

Properties: C – dark emerald-green; S – white; L – vitreous; D – transparent to translucent; DE – 3.9; H – 6.5-7; CL – none; F B conchoidal to uneven; M – perfect crystals, granular.

Origin and occurrence: Metamorphic and also hydrothermal are almost only limited to rocks with increased Cr content, ultrabasic rocks with chromite, serpentinites and skarns. It occurs as crystals up to 8 mm (5/16 in) across along cracks in chromite in Sarany, Ural mountains, Russia. Crystals up to 20 mm (25/32 in) across, come from Outokumpu, Finland. It is also known from Orford, Quebec, Canada.

Zircon

$ZrSiO_4$

TETRAGONAL ● ● ●

Varieties: jargon, hyacinth.

Properties: C – yellow (jargon), brown, yellow-brown, red-orange (hyacinth), red to colorless; S – white; L – vitreous, greasy to adamantine; D – transparent to translucent; DE – 4.7; H – 7.5; CL – imperfect; F B conchoidal to uneven; M – perfect long and short prismatic crystals, granular; LU – yellow; R B sometimes radioactive and usually metamict.

Origin and occurrence: Magmatic and metamorphic as an accessory mineral in different rock types; rarely hydrothermal in the Alpine-type and quartz veins; also in sediments and placers. Prismatic crystals up to 30 cm (12 in) across, occur in syenite pegmatites in Renfrew and Bancroft, Ontario, Canada. It is also known from Miass, Ural mountains and from Mount Vavnbed, Lovozero massif, Kola Peninsula, Russia. Classic occurrences in granitic pegmatites are Alto Ligonha, Mozambique; Arendal, Norway; Ytterby, Sweden and Jaguaraçu, Minas Gerais, Brazil. Gemmy zircons of different colors up to 80 mm (3 1/8 in) in size come from placers near Ratnapura, Sri Lanka and elsewhere.

Application: hyacinth and jargon are cut as gemstones, Zr ore.

Zircon, 20 mm x, Vishnevogorsk, Ural Mts., Russia

Eulytite, 2 mm xx, Schneeberg, Germany

Sillimanite, 10 mm grains, Havlíčkův Brod, Czech Republic

Eulytine

$Bi_4Si_3O_{12}$

CUBIC ● ●

Properties: C – brown, yellow, gray, colorless; S – white; L – greasy; D – translucent; DE – 6.6; H – 4.5; CL – good; F B conchoidal to uneven; M – small dipyramidal crystals, radial aggregates, granular.
Origin and occurrence: Secondary in Bi deposits, typically associated with bismuth.
Its small crystals come from Horní Slavkov, Czech Republic; Schneeberg and Johanngeorgenstadt, Germany.

Euclase

$BeAlSiO_4(OH)$

MONOCLINIC ● ●

Properties: C – colorless, white, greenish, blue; S – white; L – vitreous; D – transparent to translucent; DE – 3.1; H – 7.5; CL – good; F B conchoidal to

Euclase, 42 mm, Minas Novas, Brazil

uneven; M – long prismatic crystals, radial aggregates, granular.
Origin and occurrence: Hydrothermal occurs in pegmatites, greisens and in quartz and Alpine-type veins; rare in placers. Well-formed crystals up to 80 mm (3 1/8 in) in size known from Santa do Encoberto, Minas Gerais, Brazil. Blue crystals up to 50 mm (2 in) across occurred in the Last Hope Mine, Karoi, Zimbabwe. It also come from the sediments of Sanarka River, Ural mountains, Russia. Dark blue crystals up to 150 mm (6 in) across were found recently in the Chivor Mine, Colombia.
Application: locally cut as a gemstone.

Sillimanite
Al_2SiO_5

ORTHORHOMBIC ● ● ● ● ●

Properties: C – white, gray, greenish, yellowish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.3; H – 6.5-7.5; CL – good; F – uneven; M – long prismatic crystals, fibrous aggregates, granular.
Origin and occurrence: Almost exclusively metamorphic in gneisses and migmatites; only rare magmatic in pegmatites and granites; also in placers. Typical rock-forming mineral, very commonly associated with andalusite. Typical localities are Bodenmais, Germany and Mar_íkov, Czech Republic. Gemmy crystals up to 20 mm (25/32 in) long are known from Rakwana-Deniyaya, Sri Lanka.

Andalusite
Al_2SiO_5

ORTHORHOMBIC ● ● ● ●

Varieties: chiastolite, viridine

Properties: C – pink, red-brown, red, gray, whitish, green (viridine); S- white; L – vitreous to dull; D – translucent to transparent; DE – 3.2; H – 6.5-7.5; CL – good; F – uneven; M – prismatic crystals, fibrous aggregates, granular, massive.
Origin and occurrence: Metamorphic in regionally and contact metamorphosed rocks; magmatic in pegmatites and granites, hydrothermal in quartzites. Typical rock-forming mineral, commonly associated with sillimanite, corundum and cordierite. Renowned localities are Lisens, the Alps, Austria; Bimbowrie, South Australia, Australia. Green gemmy crystals come from Morro do Chapeu, Bahia, Brazil. Viridine is known from Darmstadt, Germany.

Andalusite, 28 mm x, Aracuay, Minas Gerais, Brazil

Chiastolith, 20 mm, Bimbowrie, Australia

Kyanite
Al_2SiO_5

TRICLINIC ● ● ● ●

Properties: C – blue, gray, white, green, dark gray, colorless; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.6; H – 4.5-7.5; CL – good; F – uneven; M – prismatic to tabular crystals, fibrous aggregates, granular, massive.
Origin and occurrence: Almost only metamorphic in regionally metamorphosed rocks, mica schists, gneisses, granulites and eclogites; less frequently magmatic in pegmatites and granites; rarely hydrothermal in quartz veins. Typical rock-forming mineral, associated with andalusite and sillimanite. Blue columnar aggregates and crystals up to 150 mm (6 in) long occur in Barra do Salinas, Minas Gerais, Brazil. Other renowned localities are Pizzo Forno, Switzerland; Prilep, Macedonia; Keivy, Kola Peninsula, Russia.

Kyanite, 87 mm, Minas Gerais, Brazil

Topaz, 18 mm xx, Ghundao Hill, Pakistan

Topaz
$Al_2SiO_4(F,OH)$

ORTHORHOMBIC ● ● ●

Varieties: pycnite

Properties: C – colorless, blue, yellow, gray, white, greenish, pinkish, red; S – white; L – vitreous; D – transparent to translucent; DE – 3.6; H- 8; CL – good; F – uneven; M – perfect prismatic to tabular crystals, radial and columnar aggregates, granular.
Origin and occurrence: Magmatic in pegmatites and granites; hydrothermal in greisens, in rhyolite cavities, in quartz veins, also in placers. Topaz crystals in pegmatites are occasionally very large, like the crystal measuring 80 x 60 x 60 cm (31½ x 24 x 24 in) across from Fazenda do Funil; crystals

Topaz, 49 mm, Ramona, U.S.A.

up to 30 cm (12 in) from Virgem da Lapa, both Minas Gerais, Brazil. Blue, brownish and bicolored crystals up to 40 cm (15¾ in) long come from Volodarsk Volynskii, Ukraine. Other topaz localities are Iveland, Norway; Murzinka, Ural mountains, Russia; Little Three mine, Ramona, California; Pikes Peak, Colorado, USA; Gilgit, Pakistan; Spitzkopje, Namibia. Orange to red topaz (imperial topaz) comes from quartz veins near Ouro Preto, Minas Gerais, Brazil. Pink topaz crystals up to 70 mm (2¾ in) found at Mount Ghundao, Mardan, Pakistan. Important topaz specimens were also found in Schneckenstein, Germany; Thomas Range, Utah, USA; Nerchinsk, Siberia, Russia. Columnar aggregates of pycnite come from Cínovec, Czech Republic and Altenberg, Germany. *Application:* cut as a gemstone.

Pycnite, 60 mm, Cínovec, Czech Republic

Staurolite, 80 mm, Keivy, Kola, Russia

Staurolite

$Fe_2Al_9Si_4O_{22}(OH)_2$

ORTHORHOMBIC ● ● ● ●

Properties: C – dark to light brown, yellow; S – white; L – dull to vitreous; D – transparent to almost opaque; DE – 3.7; H – 7-7.5; CL – good; F – uneven; M – prismatic to tabular crystals and their combinations, granular.

Origin and occurrence: Metamorphic in regionally metamorphosed rocks, gneisses and mica schists; rare magmatic in granites; also in placers, commonly associated with almandine, andalusite and kyanite. Crystals up to 50 mm (2 in) across come from Pizzo Forno, Switzerland. It is also known from Rio Arriba, New Mexico, USA. Its cross-like twins, up to 200 mm (7⁷/₈ in) across, occur in Keivy, Kola Peninsula, Russia and Morbihan, France.

Sapphirine, 26 mm, Vohimena, Madagascar

Sapphirine

$Mg_2Al_4SiO_{10}$

MONOCLINIC ● ●

Properties: C – dark to light blue, green; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.5; H – 7.5; CL – imperfect; F – uneven; M – tabular crystals, granular.

Origin and occurrence: Metamorphic in regionally metamorphosed rocks, rich in Al and Mg and poor in Si, associated with spinel and corundum. Crystals up to 40 mm (1⁹/₁₆ in) across come from Fiskenaesset, Greenland; Betroka and Androy, Madagascar. Other localities include Val Codera, Italy and Enderby Land, Antarctica.

Chondrodite, 15 mm x, Brewster, U.S.A.

Chondrodite

$Mg_5Si_2O_8(OH,F)_2$

MONOCLINIC ● ● ●

Properties: C – yellow, greenish, brown; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.2; H – 6-6.5; CL – imperfect; F – uneven; M – crystals of different habits, granular.
Origin and occurrence: Metamorphic in contact and regionally metamorphosed carbonate rocks; rare magmatic in carbonatites, typically associated with spinel, chlorite and phlogopite. Perfect crystals up to 50 mm (2 in) long come from the Tilly Foster mine, Brewster, New York, USA. Other localities are Pargas, Finland; Monte Somma, Italy; Riverside, California, USA.

Clinohumite

$Mg_9Si_4O_{16}(OH,F)_2$

MONOCLINIC ● ● ●

Properties: C – yellow, red, brown, white; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.3; H – 6; CL – imperfect; F – uneven; M – crystals of different habits, granular.
Origin and occurrence: Metamorphic in contact and regionally metamorphosed carbonate rocks, serpentinites and talc schists, associated with spinel, chlorite, forsterite, serpentine and phlogopite. Gemmy yellow crystals up to 30 mm ($1^3/_{16}$ in) across come from Kukh-i-Lal, Pamir, Tajikistan. Other localities are Pargas, Finland; Monte Somma, Italy; Jensen quarry, California, USA.

Clinohumite, 19 mm, Kukh-i-lal, Tadzhikistan

Chapmanite

$SbFe^{3+}{}_2Si_2O_8(OH)$

MONOCLINIC ● ●

Properties: C – olive-green to dark yellow; S – yellowish to yellow; L -dull; D – translucent; DE – 3.7; H – 2.5; CL – imperfect; F – uneven; M – elongated crystals, massive.
Origin and occurrence: Hydrothermal and secondary in the cracks of rocks, sometimes with stibnite. Its massive aggregates occur in Smilkov near Votice, Czech Republic; in the Keely mine, Cobalt, Ontario, Canada.

Chapmanite, 4 mm, Pernek, Slovakia

Braunite, 60 mm, Langban, Sweden

Braunite

$Mn^{2+}Mn^{3+}{}_6SiO_{12}$

TETRAGONAL ● ● ●

Properties: C – black, black-gray, black-brown; S – gray; L – submetallic to metallic; D – opaque; DE – 4.7; H – 6-6.5; CL – perfect; F – uneven to conchoidal; M – dipyramidal crystals, granular.
Origin and occurrence: Metamorphic in regionally and contact metamorphosed rocks; hydrothermal in sedimentary rocks rich in Mn and in hydrothermal veins, associated with hausmanite, pyrolusite and other Mn minerals. Its perfect crystals up to 70 mm (2¾ in) long, come from Kacharhavee and Tirodi, India. It is also known from Langban, Sweden; Ilfeld and Ilmenau, Germany; St. Marcel, Italy; Tizi Bashkun, Morocco.
Application: Mn ore.

Thaumasite

$Ca_6Si_2(CO_3)_2(SO_4)_2(OH)_{12} \cdot 24\ H_2O$

HEXAGONAL ● ●

Properties: C – white to colorless; S – white; L – vitreous to dull; D – transparent to translucent; DE – 1.9; H – 3.5; CL – imperfect; F – uneven; M B acicular aggregates, granular, massive; LU – white.
Origin and occurrence: Hydrothermal or metamorphic in contact metamorphosed carbonate rocks, usually associated with other Ca silicates and carbonates, like ettringite and prehnite. Typical localities are Crestmore, California; West Paterson, New Jersey, USA; Langban, Sweden; N'Chwaning mine No. 2, Kuruman, South Africa.

Titanite

$CaTiOSiO_4$

MONOCLINIC ● ● ● ●

Properties: C – colorless, yellow, brown, green, gray to black; S – white; L – vitreous to dull; D – transparent, translucent to opaque; DE – 3.5; H – 5-5.5; CL – good; F – uneven to conchoidal; M – tabular crystals and their combinations, granular, massive.
Origin and occurrence: Metamorphic and magmatic as a common accessory mineral in many igneous and

Thaumasite, 3 mm aggregate, Maglovec, Slovakia

Fersmanite, 10 mm x, Khibiny Massif, Kola, Russia

Titanite, 20 mm xx, Stubachtal, Austria

metamorphic rocks and pegmatites; hydrothermal in the Alpine-type veins; also in placers.
The most beautiful crystals up to 180 mm (7 1/16 in) long occur in Alpine-type veins in Tavetsch and Binntal, Switzerland; Zillertal and Felbertal, Austria; Dodo, Polar Ural, Russia. Large, poorly developed crystals weighing up to 40 kg (88 lb), come from Eganville, Ontario, Canada and Rossie, New York, USA.

Chloritoid, 30 mm, Ile de Croix, France

Fersmanite

$(Na,Ca)_4(Ti,Nb)_2Si_2O_{11}(OH,F)_2$

TRICLINIC ● ●

Properties: C – light to dark brown; S -white; L – vitreous to dull; D – translucent; DE – 3.5; H – 5-5.5; CL – none; F – uneven to conchoidal; M – tabular crystals.
Origin and occurrence: Magmatic to hydrothermal in alkaline pegmatites together with pectolite, apatite and sulfides. Crystals up to 30 mm (1 3/16 in) across known from Mount Eveslogchorr, Khibiny massif, Kola Peninsula, Russia.

Chloritoid

$Fe_2Al_4Si_2O_{10}(OH)_2$

MONOCLINIC, TRICLINIC ● ● ●

Properties: C – dark gray, gray-green to black-green; S – gray; L – vitreous to dull; D – translucent; DE – 3.6; H – 6.5; CL – perfect; F – uneven to conchoidal; M – tabular crystals, foliated and scaly aggregates, granular.
Origin and occurrence: Metamorphic in some mica schists and phyllites; hydrothermal alteration product in lavas. Typical localities are Zermatt, Switzerland; Ottrez, Belgium; Pregraten, Austria.

Datolite, 30 mm xx, Dalnegorsk, Russia

Datolite

$CaBSiO_4(OH)$

MONOCLINIC ● ● ●

Properties: C – white to colorless, yellowish, greenish, gray; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.1; H – 5-5.5; CL – none; F – uneven to conchoidal; M – prismatic to tabular crystals, nodules, granular.

Origin and occurrence: Metamorphic and hydrothermal in contact metamorphosed rocks, in cavities in volcanic rocks, in ore veins and pegmatites. It is commonly associated with zeolites, prehnite, calcite and also with tourmaline. Prismatic and tabular crystals up to 100 mm (4 in) long occur in Dalnegorsk, Russia. Other typical localities are West Paterson, New Jersey; Keweenaw Peninsula, Michigan, USA; Haslach, Germany; Charcas, San Luis Potosi, Mexico.

Gadolinite-(Y)

$Y_2FeBe_2Si_2O_{10}$

MONOCLINIC ● ●

Properties: C – black, dark red, brown, greenish; S – gray-green; L – vitreous to greasy, D – transparent to translucent; DE – 4.4; H – 6.5-7; CL – none; F –

Gadolinite-(Y), 60 mm, Hitterö, Norway

Howlite, 30 mm, California, U.S.A.

Dumortierite, 60 mm, Dehesa, U.S.A.

conchoidal; M – prismatic crystals, granular; R – locally metamict.
Origin and occurrence: Magmatic in granites and pegmatites, locally associated with fluorite and allanite; rare hydrothermal in the Alpine-type veins. Aggregates, up to 40 cm (15¾ in) across come from Hittero, Norway. Barringer Hill and Clear Creek, Texas, USA yielded aggregates, weighing up to 90 kg (198 lb). Other famous localities are Blatchford Lake, Northwest Territories, Canada; Baveno, the Alps, Italy; Iveland, Evje, Norway.

Kornerupine
$Mg_4Al_6(Si,Al,B)_5O_{21}(OH)$

ORTHORHOMBIC ● ●

Properties: C – colorless, white, greenish, gray, brown; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.3; H – 6.5-7; CL – none; F – conchoidal; M – prismatic crystals, radial aggregates, granular.
Origin and occurrence: Metamorphic in strongly metamorphosed rocks, as granulites. Crystals up to 230 mm (9 1/16 in) across known from Fiskenaesset, Greenland. It also occurs in Lac Ste-Marie, Quebec, Canada; Waldheim, Germany; Itrongay, Madagascar.

Dumortierite
$(Al,Ti,Mg)Al_6BSi_3O_{16}(O,OH)_2$

ORTHORHOMBIC ● ● ●

Properties: C – purple, pink, blue, brown; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.4; H – 8.5; CL – good; F – uneven; M – prismatic crystals, radial and acicular aggregates, massive.
Origin and occurrence: Metamorphic in some Al-rich metamorphic rocks, e.g. in migmatites and gneisses; magmatic in granites and pegmatites; hydrothermal in altered rocks. It comes from Dehesa, California; Rochester, Nevada, USA; Beaunan, France; crystals are known from the vicinity of Kutna Hora, Czech Republic; Soavina, Madagascar.

Howlite
$Ca_2B_5SiO_9(OH)_5$

MONOCLINIC ● ●

Properties: C – white; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.5; H – 3.5; CL – good; F – uneven; M – tabular crystals, nodules, massive.
Origin and occurrence: Hydrothermal in borate deposits, associated with ulexite and colemanite. Crystals, several cm across, come from the vicinity of Bras d'Or Lake, Nova Scotia, Canada. It occurs also in Lang and Daggett, California, USA.

Kornerupine, 90 mm, Lac Ste.-Marie, Canada

Cuprosklodowskite, 82 mm, Musonoi, Zair

Cuprosklodowskite

$Cu(UO_2)_2Si_2O_7 \cdot 6\,H_2O$

MONOCLINIC ● ●

Properties: C – various hues of green; S – greenish; L – vitreous to dull; D – transparent to translucent; DE – 3.8; H – 4; CL – good; F – uneven; M – radial, acicular aggregates, thin coatings, granular, massive; R – strong radioactive.

Origin and occurrence: Secondary in the oxidation zone of U deposits, associated with autunite, torbernite, uranophane and other U secondary minerals. Crystals, up to several cm long, come from

Uranophane, 125 mm, Brewster, Texas, U.S.A.

Kasolite, 30 mm, Musonoi, Zair

the Mashamba West mine, Musonoi, Zaire. It also occurred in Jachymov, Czech Republic.

Uranophane
$Ca(UO_2)_2Si_2O_7 \cdot 6\,H_2O$

MONOCLINIC ● ● ●

Properties: C – various hues of yellow to brown; S – yellowish; L – vitreous to dull; D – transparent to translucent; DE – 3.9; H – 2.5; CL – good; F – uneven; M – radial, acicular aggregates, thin coatings, granular, massive; LU – yellow-green; R – strong radioactive.
Origin and occurrence: Secondary in the oxidation zone of U deposits as a product of uraninite alteration, associated with autunite, torbernite and other secondary U minerals. Needles up to 10 mm (3/8 in) long, come from Musonoi, Shinkolobwe, Zaire.
It is also known from the Faraday mine, Bancroft, Ontario, Canada. It occurs in Wolsendorf, Germany, as well as in Jachymov, Czech Republic.

Kasolite
$Pb(UO_2)SiO_4 \cdot H_2O$

MONOCLINIC ● ●

Properties: C – various hues of yellow, green to brown; S – yellowish; L – vitreous to dull; D – translucent to opaque; DE – 6.2; H – 4.5; CL – good; F – uneven; M – prismatic crystals, radial and acicular aggregates, thin coatings, granular, massive; R – strong radioactive.
Origin and occurrence: Secondary mineral in the oxidation zone of U deposits. It is a product of uraninite alteration and is associated with uranophane, torbernite and U hydroxides. Crystals up to 10 mm (3/8 in) long occur in Shinkolobwe, Kasolo, Zaire and Mounana, Gabon.

Akermanite
$Ca_2MgSi_2O_7$

TETRAGONAL ● ● ●

Properties: C – white, gray, green, brown; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.9; H – 5.5; CL – good; F – uneven to conchoidal; M – short prismatic crystals, granular.
Origin and occurrence: Magmatic in volcanic basic rocks; metamorphic in contact metamorphosed marbles. A typical representative of localities in marbles is Crestmore, California, USA. It occurs also in Ca-rich volcanic rocks in Velardena, Mexico.

Akermanite, 5 mm xx, Mt. Vesuvius, Italy

Gehlenite
$Ca_2Al_2SiO_7$

TETRAGONAL ● ● ●

Properties: C – white, gray, yellowish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.0; H – 5-6; CL – good; F – uneven to conchoidal; M – short prismatic crystals, granular.
Origin and occurrence: Magmatic in volcanic basic rocks; metamorphic in contact metamorphic marbles.
It occurs in Crestmore, California, USA; Monzoni, Italy; Oraviza, Romania and elsewhere.

Gehlenite, 80 mm, Vata de Sus, Romania

Ilvaite, 20 mm xx, Dalnegorsk, Russia

Bertrandite, 36 mm, Mt.Antero, U.S.A.

Ilvaite

$CaFe^{2+}{}_2Fe^{3+}Si_2O_8(OH)$

ORTHORHOMBIC ● ● ●

Properties: C – black to black-gray; S – black; L – submetallic to dull; D – opaque; DE – 4.1; H – 5.5-6; CL – good; F – uneven; M – prismatic crystals, granular.
Origin and occurrence: Metamorphic and hydrothermal in contact metamorphosed Fe, Zn and Cu deposits. Prismatic crystals up to 100 mm (4 in) long come from Rio Marina, Elba, Italy. Crystals up to 30 cm (12 in) long are known from Serifos Island, Greece. Crystals, several cm long occur also in Dalnegorsk, Russia and in the Laxey mine, Idaho, USA.

Bertrandite

$Be_4Si_2O_7(OH)_2$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, yellowish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.6; H – 6-7; CL – perfect; F – uneven; M – tabular crystals and their combinations, radial aggregates, granular, massive.
Origin and occurrence: Hydrothermal in granitic pegmatites, greisens and in hydrothermal veins, together with beryl, also as pseudo-morphs after beryl. Tabular crystals up to 50 mm (2 in) across occur in Conselheira Pena, Minas Gerais, Brazil. It is known from Kounrad and Kara-Oba, Kazakhstan in crystals up to 30 mm ($1^3/_{16}$ in) across. It also comes from Stoneham, Maine, USA; Písek, Czech Republic and Iveland, Norway.
Application: the most important Be ore.

Hemimorphite

$Zn_4Si_2O_7(OH)_2 \cdot H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, yellowish, greenish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.4; H – 4.5-5; CL – perfect; F – uneven; M – tabular crystals and their combinations, botryoidal and radial aggregates, granular, massive.
Origin and occurrence: Secondary in the oxidation zone of Zn deposits, associated with sphalerite, smithsonite and cerussite. Crystals up to 100 mm (4 in) long come from Bisbee, Arizona, USA.. It is also from the El Potosi Mine, Santa Eulalia, Chihuahua, Mexico; Bleiberg, Austria; Cho-Dien, Vietnam.
Application: Zn ore.

Hemimorphite, 30 mm xx, Mapimi, Mexico

Lamprophylite

$Na_2Sr_2Ti_3Si_4O_{16}(OH,F)_2$

MONOCLINIC ● ●

Properties: C – brown to dark brown; S – white; L – vitreous to submetallic; D – translucent; DE – 3.5; H – 2-3; CL – perfect; F – uneven; M – tabular crystals, radial aggregates.

Origin and occurrence: Magmatic in alkaline syenites and their pegmatites, associated with nepheline, aegirine and eudialyte. The best crystals up to 150 mm (6 in) long come from Mount Flora, Lovozero massif, Kola Peninsula, Russia. It is also known from Langesundsfjord, Norway and Mont St.-Hilaire, Quebec, Canada.

Hemimorphite, 26 mm, Arizona, U.S.A.

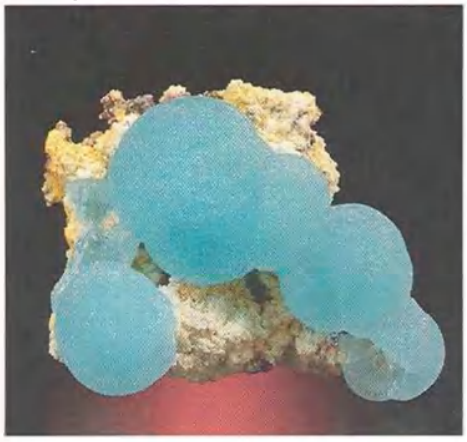

Lamprophyllite, 50 mm xx, Lovozero Massif, Kola, Russia

Clinozoisite, 36 mm, Eden Mills, U.S.A.

Clinozoisite
$Ca_2Al_3Si_3O_{12}(OH)$

MONOCLINIC ● ● ● ●

Properties: C – colorless, yellowish, green, pink; S – white; L – vitreous to dull; PS – transparent to translucent; DE – 3.4; H – 6.5; CL – good; F – uneven; M – prismatic crystals, columnar, radial aggregates, granular, massive.
Origin and occurrence: Metamorphic and hydrothermal in contacts of marbles and in the Alpine-type veins. Prismatic crystals up to 60 mm (2³/₄ in) long come from Radoy Island, Norway. It comes also from Eden Mills, Vermont; Allens Park, Colorado, USA; Pinos Altos, Baja California, Mexico.

Epidote
$Ca_2(Al,Fe^{3+})_3Si_3O_{12}(OH)$

MONOCLINIC ● ● ● ● ●

Properties: C – green, brown, greenish, yellow-green; S – white; L – vitreous to dull; D – transparent to translucent, locally opaque; DE – 3.4; H – 6-7; CL – good; F – uneven; M – prismatic crystals, columnar and radial aggregates, granular, massive.
Origin and occurrence: Metamorphic in marble/granite contacts; hydrothermal in the Alpine-type veins and in hydrothermally altered rocks, typically associated with albite, prehnite and amphibole. Its green crystals up to 140 mm (5½ in) long, were found in Sobotín, Czech Republic. Perfect crystals up to 100 mm (4 in) across come from Knappenwand, Austria. Thick tabular crystals occur in Prince of Wales Island, Alaska, USA. Fine crystals, resembling Austrian crystals, were found recently in Alchuri, Shigar, Pakistan. Fine columnar aggregates of crystals are known from Pampa Blanca, Peru and also reported from Arendal, Norway.

Piemontite
$Ca_2(Al,Mn^{3+})_3Si_3O_{12}(OH)$

MONOCLINIC ● ●

Properties: C – red-brown to black, crimson, red-yellow; S – white; L – vitreous to dull; D – translucent, locally opaque; DE – 3.5; H – 6; CL – good; F – uneven; M – prismatic crystals, radial aggregates, granular.
Origin and occurrence: Metamorphic in shales and Mn-rich metamorphic rocks; rare magmatic in rhyolites and pegmatites. Needles up to 30 mm (1¹/₁₆

Epidote, Dashkesan, Azerbaidzhan

Epidote, 50 mm, Knappenwand, Austria

in) long come from St. Marcel, Piedmont, Italy and Otakiyama, Japan.

Allanit-(Ce)

$(Ca,Ce,Y)_2(Al,Fe^{3+})_3Si_3O_{12}(OH)$

MONOCLINIC ● ● ●

Properties: C – black to dark brown; S – light gray; L – greasy to submetallic; D – translucent to opaque; DE – 3.9; H – 5.5-6; CL – none; F – conchoidal to uneven; M – tabular crystals, granular; R – usually metamict.

Origin and occurrence: Magmatic in pegmatites and granites; metamorphic in various types of metamorphic rocks, e.g. migmatites, amphibolites and gneisses. Grains, up to 70 cm (27$^{3}/_{16}$ in) across found in pegmatites near Bancroft, Ontario, Canada. Also occurs in Barringer Hill, Colorado; Amelia district, Virginia, USA; Arendal and Hittero, Norway; Ytterby and Riddarhyttan, Sweden; Yates mine, Quebec, Canada.

Piemontite, 45 mm, Todyryact, Morocco

Allanite-(Ce), 40 mm, Vizcaya, Spain

Tanzanite, 37 mm, Arusha, Tanzania

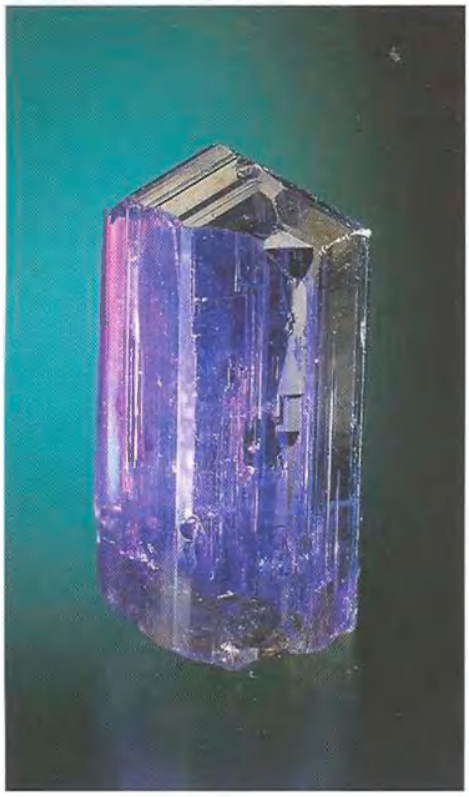

Zoisite, 60 mm, Weisenstein, Germany

Zoisite

$Ca_2Al_3Si_3O_{12}(OH)$

ORTHORHOMBIC ● ● ●

Varieties: thulite, tanzanite

Properties: C – colorless, yellowish, green, pink, red (thulite), blue (tanzanite); S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.4; H – 6.5; CL – good; F – uneven; M – prismatic crystals, radial aggregates, granular.
Origin and occurrence: Metamorphic in regionally metamorphosed Ca-rich rocks, mainly in pyroxenic gneisses, amphibolites and in marble contacts. It is known from many localities like Saualpe, Austria; Zermatt, Switzerland; Lexviken, Norway (thulite); Traversella, Italy; Alchuri, Pakistan; Merelani Hills, Arusha, Tanzania (tanzanite), where crystals up to 70 mm (2¾ in) long were found.
Application: tanzanite and thulite are cut as gemstones.

Thulite, 40 mm, Lexviken, Norway

Vesuvianite

$Ca_{19}(Al,Mg,Fe)_{13}Si_{18}O_{68}(OH,F,O)_{10}$

TETRAGONAL ● ● ●

Properties: C – brown, yellowish, green, blue, purple, colorless; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.3; H – 6-7; CL – imperfect; F – uneven to conchoidal; M – prismatic to tabular crystals, columnar aggregates with radial structure, granular, massive.

Origin and occurrence: Metamorphic and hydrothermal in contact metamorphosed Ca-rich rocks, mainly in skarns, in marble contacts, also in rodingites; rarely magmatic in alkaline rocks. It is usually associated with grossular, wollastonite and diopside. Perfect green crystals up to 180 mm (7 1/16 in) long and purple crystals up to 70 mm (2¾ in) long come from the Jeffrey quarry, Asbestos, Quebec, Canada. It is also known from Hazlov, Czech Republic; Crestmore, California, Franklin, New Jersey, USA; Monzoni, Italy.

Vesuvianite, 110 mm, Asbestos, Canada

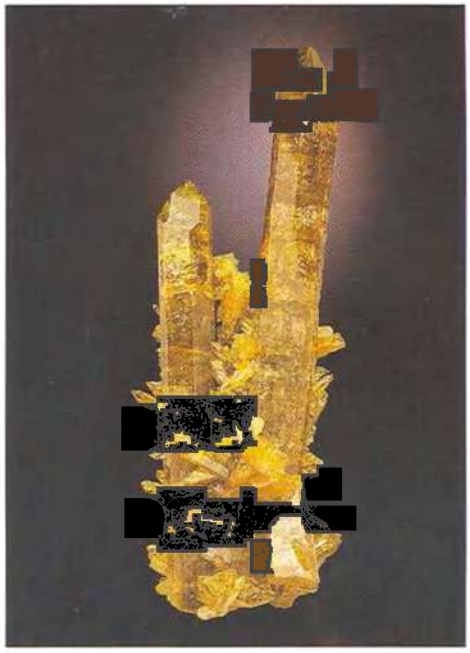

Viluite

$Ca_{19}(Al,Mg)_{13}B_5Si_{18}O_{68}(O,OH)_{10}$

TETRAGONAL ●

Properties: C – dark green, gray-brown; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.4; H – 6; CL – imperfect; F – uneven to conchoidal; M – prismatic crystals.

Origin and occurrence: Metamorphic in serpentinized skarn, associated with grossular. Its perfect prismatic crystals up to 50 mm (2 in) long are only known from the Vilui River basin, Yakutia, Russia.

Viluite, 30 mm xx, Vilui River, Russia

Benitoite
$BaTiSi_3O_9$

TRIGONAL ● ●

Properties: C – blue, pink, white, colorless; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.6; H – 6-6.5; CL – imperfect; F – uneven to conchoidal; M – prismatic to tabular crystals, mainly with trigonal cross-section; LU – bluish.
Origin and occurrence: Hydrothermal in veins, cross-cutting serpentinites, always associated with neptunite and natrolite. Its classic locality is the Benitoite Gem mine, San Benito Co., California, USA, where it forms tabular crystals up to 40 mm (1 7/16 in) across.

Catapleiite
$Na_2ZrSi_3O_9 \cdot 2\,H_2O$

HEXAGONAL ● ●

Properties: C – light yellow, yellow-brown, pink, brown, blue; S – white; L – vitreous to dull; D – transparent to opaque; DE – 2.8; H – 5-6; CL – good; F – uneven; M – thin tabular crystals, lamellar aggregates.
Origin and occurrence: Magmatic in nepheline syenites and their pegmatites, together with acgirine, titanite, nepheline and microcline. It occurs in Mont St.-Hilaire, Quebec, Canada, as tabular crystals up to 150 mm (6 in) across. Crystals up to 30 mm (1 3/16 in) across come from Mount Yukspor, Khibiny massif, Kola Peninsula, Russia. It is also known from Langesundsfjord, Norway; Magnet Cove, Arkansas, USA.

Eudialyte
$Na_4(Ca,Fe,Ce,Mn)_2ZrSi_6O_{18}(OH,Cl)$

TRIGONAL ● ●

Properties: C – red, pink to brown; S – white; L – vitrous to dull; D – transparent to translucent; DE – 2.8; H – 5-5.5; CL – imperfect; F – uneven; M – prismatic and tabular crystals, granular.
Origin and occurrence: Magmatic in nepheline syenites and their pegmatites, associated with acgirine, nepheline and microcline. Crystals up to 80 mm (3 1/8 in) across come from Mount Kukisvumchorr, Khibiny massif, Kola Peninsula, Russia. Crystals up to 50 mm (2 in) across are also known from Mont St.-Hilaire, Quebec, Canada. It also occurs in Langesundsfjord, Norway and in Los Island, Guinea.

Ferroaxinite
$Ca_2FeAl_2BSi_4O_{15}(OH)$

TRICLINIC ● ● ●

Properties: C – brown to purple- brown, light purple; S – white; L – vitreous to dull; D – trans-

Benitoite, 28 mm x, San Benito Co., U.S.A.

Catapleiite, 37 mm, Mont St.-Hilaire, Canada

Eudialyte, 20 mm grain, Khibiny Massif, Kola, Russia

Ferroaxinite, 62 mm, Khapalu, Pakistan

parent to translucent; DE – 3.3; H – 6. 5-7; CL – good; F – uneven to conchoidal; M – tabular crystals, platy aggregates, granular, massive.
Origin and occurrence: Metamorphic and hydrothermal in contacts of marbles and granites, associated with clinozoisite, prehnite, calcite and actinolite, also in the Alpine-type veins and pegmatites. Perfect tabular crystals up to 150 mm (6 in) across come from Puiva, Polar Ural, Russia. Other renowned localities are Obira, Japan; Bourg d'Oisans, France; Monte Scopi, Switzerland.

Tinzenite

$Ca(Mn,Fe)_2Al_2BSi_4O_{15}(OH)$

TRICLINIC ● ●

Properties: C – yellow, orange to red; S – white; L – vitrous to dull; D – transparent to translucent; DE – 3.3; H – 6.5-7; CL – good; F – uneven to conchoidal; M – tabular crystals, platy and fibrous aggregates, granular, massive.
Origin and occurrence: Hydrothermal in the Alpine-type veins, cross-cutting a rock, rich in braunite. It comes from Tinzen, Val d'Err, Switzerland and in the Cassagna mine, Genova, Italy.

Tinzenite, 10 mm xx, Val Graveglia, Italy

Beryl

$Be_3Al_2Si_6O_{18}$

HEXAGONAL ● ● ● ●

Varieties: emerald, aquamarine, heliodor, morganite, goshenite, red beryl (bixbite).

Properties: C – variable in different varieties; common beryl – mostly yellow, yellow-green, light green to white, rare blue; varieties: emerald – dark emerald-green; aquamarine – light to dark blue-green: morganite – pink; heliodor – light yellow to yellow-green; goshenite – white to colorless; red beryl B red; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.6; H – 7.5-8; CL – imperfect; F – uneven to conchoidal; M – long prismatic to tabular crystals, columnar and radial aggregates, granular, massive.

Origin and occurrence: Magmatic in pegmatites and granites; hydrothermal in greisens, in cavities in rhyolite, in quartz veins; metamorphic in mica schists. Perfect prismatic crystals of common beryl up to 9 m (29 ft 6 in) long found in the Etta mine, Keystone, South Dakota, USA. Crystals weighing up to 177 tons, come from Namivo, Alto Ligonha, Mozambique. Other localities are Pici, Brazil; Iveland, Norway; Antsirabe, Madagascar. Emerald crystals occur in mica-schists, marbles and ultrabasic rocks, associated with other Be minerals, phenakite and

Emerald, 30 mm x, Boyaca, Colombia

Emerald, 30 mm, Ural Mts., Russia

chrysoberyl. Beautiful dark green transparent crystals are known from Malyshevo, Ural Mts., Russia; Muzo and Coscuez, Colombia; Habachtal, Austria,

Emerald, 90 mm, Muzo, Colombia

Aquamarine, 97 mm, Koronel Hurta, Brazil

where the largest crystals reach up to 120 mm ($4^{11}/_{16}$ in) in size. Aquamarine is mainly known from pegmatites and hydrothermal veins, commonly

Aquamarine, 158 mm, Shingus, Pakistan

Heliodor, 87 mm, Volodarsk Volynskii, Ukraine

associated with tourmaline, quartz and albite. Intense blue crystals are known from many pegmatites in Minas Gerais, Brazil, where in the Marambaia mine, crystals, up to 70 cm (27$^{9}/_{16}$ in) long and weighing up to 110 kg (242 lb) found. Very beautiful gemmy crystals over 30 cm (12 in) long recently occurred in the Medina Mine. It also come from Murzinka, Ural mountains and Adun Chilon, Siberia, Russia. It also comes from Spitzkopje, Namibia; Gilgit, Pakistan and elsewhere. Morganite is a typical mineral of granitic pegmatites, where it occurs in mainly in cavities, usually associated with color varieties of tourmaline, quartz and albite. Its tabular crystals up to 100 mm (4 in) across found in the White Queen mine, Pala, California, USA. Smaller crystals come from San Piero in Campo, Elba, Italy. Crystals up to 50 cm (20 in) across reported from several localities in Minas Gerais, Brazil. Heliodor also occurs in pegmatite cavities, hydrothermal veins and in gneisses, commonly associated with quartz and albite. Prismatic crystals up to 200 mm (7$^{7}/_{8}$ in) long, come from Volodarsk Volynskii, Ukraine. It is also known from Nerchinsk, Siberia, Russia and several mines in Minas Gerais, Brazil. Goshenite occurs only in pegmatites. Its prismatic crystals come from Goshen, Massachusetts,

Heliodor, 26 mm x, Pamir, Tadzhikistan

Goshenite, 39 mm, Apalygun, Pakistan

Red beryl, 17 mm x, Wah Wah Mts., U.S.A.

USA; San Piero in Campo, Elba, Italy. Red beryl occurs in Violet Claims, Wah Wah mountains, Utah, USA, where crystals up to 50 mm (2 in) were found. *Application:* Be ore, color varieties are cut as gemstones.

Morganite, 43 mm x, San Diego Co., U.S.A.

Bazzite, 70 mm, Tordal, Norway

Bazzite

$Be_3(Sc,Al)_2Si_6O_{18}$

HEXAGONAL ●

Properties: C – light to intense blue; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.8; H – 6.5; CL – imperfect; F – uneven to conchoidal; M – long prismatic crystals, columnar and radial aggregates.

Origin and occurrence: Hydrothermal in the Alpine-type veins and pegmatites. Crystals up to 20 mm (25/32 in) long come from Tørdal, Norway. It is also known from Lago Maggiore, Italy and St. Gotthard, Switzerland.

Cordierite

$Mg_2Al_4Si_5O_{18}$

ORTHORHOMBIC ● ● ● ●

Varieties: iolite (gemmy blue)

Sekaninaite, 40 mm x, Dolni Bory, Czech Republic

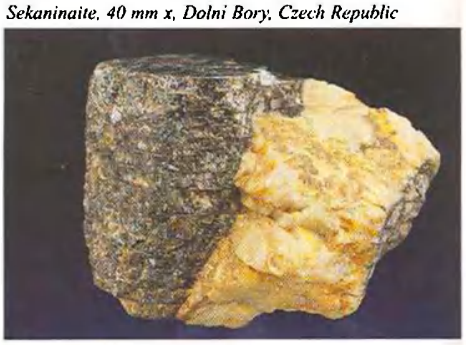

Properties: C – blue, purple, gray, gray-green, gray-brown, trichroic; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.5; H – 7-7.5; CL – imperfect; F – uneven to conchoidal; M – short prismatic crystals, granular.

Origin and occurrence: Metamorphic in migmatites and gneisses and in contact cherts; magmatic in granites and granitic pegmatites, usually associated with andalusite and sillimanite, also known from placers. It is a typical rock-forming mineral. Its prismatic crystals are very rare. Poorly developed transparent crystals up to 200 mm (7⅞ in) long come from Naverberg, Sweden. It also occurs in Orijarvi, Finland; Kragero, Norway; Bodenmais, Germany. Gemmy pebbles are known from the vicinity of Ratnapura, Sri Lanka.

Sekaninaite

$Fe_2Al_4Si_5O_{18}$

ORTHORHOMBIC ● ● ●

Properties: C – blue, purple, strongly pleochroic; S

Cordierite, 50 mm, Fishtail Lake, Canada

Dravite, 27 mm, Gujarkkot, Nepal

Buergerite, 12 mm x, San Luis Potosí, Mexico

– white; L – vitreous to dull; D – transparent to translucent; DE – 2.8; H – 7-7.5; CL – imperfect; F – uneven to conchoidal; M – short prismatic crystals, granular.
Origin and occurrence: Magmatic in granitic pegmatites and some granites, associated with andalusite and tourmaline; metamorphic in gneisses and migmatites. Conical, imperfect crystals, up to 70 cm (27 9/16 in) long typically come from Dolni Bory, Czech Republic; also known from San Piero in Campo, Elba, Italy.

Dravite

TOURMALINE GROUP

$NaMg_3Al_6(BO_3)_3Si_6O_{18}(OH)_4$

TRIGONAL ● ● ● ●

Properties: C – light to dark black-brown, blue, colorless, commonly pleochroic; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.0; H – 7-7.5; CL – none; F – uneven to conchoidal; M – long to short prismatic crystals, columnar to acicular aggregates, granular.
Origin and occurrence: Metamorphic in migmatites, gneisses, mica schists, marbles and in contact metasomatic rocks; magmatic in some granitic pegmatites; hydrothermal in quartz veins and ore veins, also known from placers. Brown to dark brown perfect crystals up to 200 mm (7 7/8 in) long come from mica schists near Dravograd, Slovenia. It is also known from marbles in Gouverneur, New York, USA. Crystals up to 150 mm (6 in) across found in Yinnietharra, Western Australia, Australia. Crystals up to 50 mm (2 in) long were recently found in Gujarkot, Nepal.
Application: transparent crystals are cut as gemstones.

Buergerite

TOURMALINE GROUP

$NaFe^{3+}{}_3Al_6(BO_3)_3Si_6O_{21}F$

TRIGONAL ●

Properties: C – black, strongly pleochroic; S – yellow-brown; L – vitreous to dull; D – translucent to opaque; DE – 3.3; H – 7; CL – none; F – uneven to conchoidal; M B prismatic crystals, granular.
Origin and occurrence: Hydrothermal in rhyolites. Its black crystals, up to 40 mm (19/16 in) long, are known from Mexquitic, San Luis Potosi, Mexico.

Schorl, 40 mm, Gilgit, Pakistan

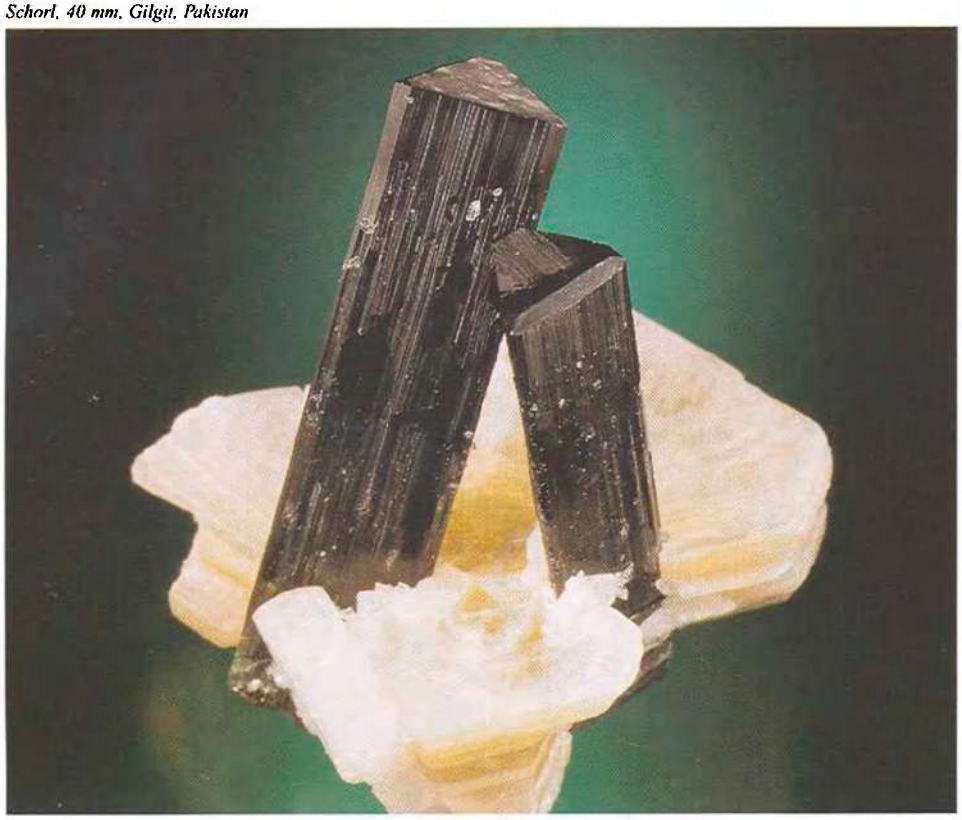

Schorl

TOURMALINE GROUP

$NaFe_3Al_6(BO_3)_3Si_6O_{18}(OH)_4$

TRIGONAL ● ● ● ●

Properties: C – black, black-brown, blue-black, strongly pleochroic; S – white; L – vitreous to dull; D – translucent to opaque; DE – 3.3; H – 7-7.5; CL – none; F – uneven to conchoidal; M – long to short prismatic crystals, columnar to acicular aggregates, granular, massive.

Origin and occurrence: Magmatic in granites and granitic pegmatites; hydrothermal in greisens, in quartz and ore veins; metamorphic in migmatites, gneisses, mica schists and tourmalinites; also known from placers. It is usually associated with muscovite, quartz and albite. Perfect black crystals come from many pegmatite localities.

Its long prismatic crystals, up to 5 m long, come from Arendal, Norway. Very good crystals are also known from Kaatiala, Finland; Dolni Bory, Czech Republic; Conselheira Pena and Galileia, Minas Gerais, Brazil.

Povondraite

TOURMALINE GROUP

$NaFe^{3+}{}_3Mg_2Fe^{3+}{}_4(BO_3)_3Si_6O_{18}(OH)_4$

TRIGONAL ●

Properties: C – black; S – gray; L – vitreous to dull; D – translucent to opaque; DE – 3.3; H – 7; CL –

Povondraite, 60 mm, Alto Chapare, Bolivia

Elbaite, 115 mm, Tourmaline Queen Mine, U.S.A.

Elbaite, 153 mm, Afghanistan

none; F – uneven to conchoidal; M – short prismatic crystals, granular.
Origin and occurrence: Hydrothermal along the cracks in metamorphosed evaporites. Its black crystals up to 10 mm (3/8 in) long typically come from Alto Chapare, Cochabamba, Bolivia.

Elbaite

TOURMALINE GROUP

$Na(Li_{1,5}Al_{1,5})Al_6(BO_3)_3Si_6O_{18}(OH)_3F$

TRIGONAL ● ● ●

Varieties: rubellite, verdelite, indicolite, achroite

Properties: C – varies in different varieties, rubellite – pink to red; verdelite – various hues of green; indicolite – blue; achroite – colorless, other colors include yellow, brown and black; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.0; H – 7; CL – none; F – uneven to conchoidal; M – prismatic crystals, columnar aggregates, granular, massive.
Origin and occurrence: Almost only magmatic and hydrothermal in granitic pegmatites, typically associated with lepidolite and albite; also in placers. Its crystals are mostly known from pegmatite cavi-

Rubellite, 23 mm, Pala, U.S.A.

ties in many localities. Rubellite crystal called 'The Rocket' (109 cm/43 5/16 in long) was found in the Jonas mine, Minas Gerais, Brazil. Rubellite crystals up to 40 cm (15¾ in) long come from Alto Ligonha district, Mozambique. Rubellite crystals up to 250 mm (9 13/16 in) long occurred in the Stewart Lithia, Tourmaline King and Tourmaline Queen mines, Pala, California, USA. Verdelite crystal 270 mm (10⅝ in) long, is known from the Dunton mine, Newry, Maine, USA. Other important localities include San Piero in Campo, Elba, Italy; Malkhan, Transbaikalia, Russia; Rozna , Czech Republic; Uto, Sweden; Gilgit, Pakistan; Paprok, Afghanistan.
Application: commonly used as a gemstone.

Uvite

TOURMALINE GROUP

$CaMg_4Al_5(BO_3)_3Si_6O_{18}(OH,F)_4$

TRIGONAL ● ● ●

Properties: C – gray, black, brown, green, red, pleochroic; S – white; L – vitreous to dull; D – translucent to opaque; DE – 3.3; H – 7-7.5; CL – none; F – uneven to conchoidal; M – long to short prismatic crystals, columnar aggregates, granular.
Origin and occurrence: Metamorphic in Ca-rich rocks, marbles, skarns; magmatic in some pegmatites; hydrothermal in ore veins.
Perfect green and red crystals, up to 30 mm (1 3/16 in) across, occur in Brumado, Bahía, Brazil. It is also known from Gouverneur and Pierrepont, New York, USA.

Uvite, 41 mm, Brumado, Brazil

Liddicoatite

TOURMALINE GROUP

$Ca(Li_2Al)Al_6(BO_3)_3Si_6O_{18}(OH)_3F$

TRIGONAL ● ● ●

Varieties: rubellite, verdelite

Properties: C – varies in different varieties, mainly pink, green, green-brown to yellow-brown; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.1; H – 7; CL – none; F – uneven to conchoidal; M – prismatic crystals, columnar aggregates, granular.
Origin and occurrence: Magmatic in granitic peg-

Liddicoatite, 60 mm, Anjanobonoina, Madagascar

Verdelite, 24 mm, Gillette Quarry, U.S.A.

Indigolite, 32 mm, Esmeralda Mine, Mesa Grande, U.S.A.

matites in association with lepidolite, spodumene and albite. Its perfect red and green crystals up to 250 mm (9 13/16 in) long come from pegmatite cavities in many localities in Madagascar, e.g. Sahatany and Anjanabonoina. It is also known from Blizná, Czech Republic.

Application: cut as a gemstone.

Liddicoatite, 41 mm, Madagascar

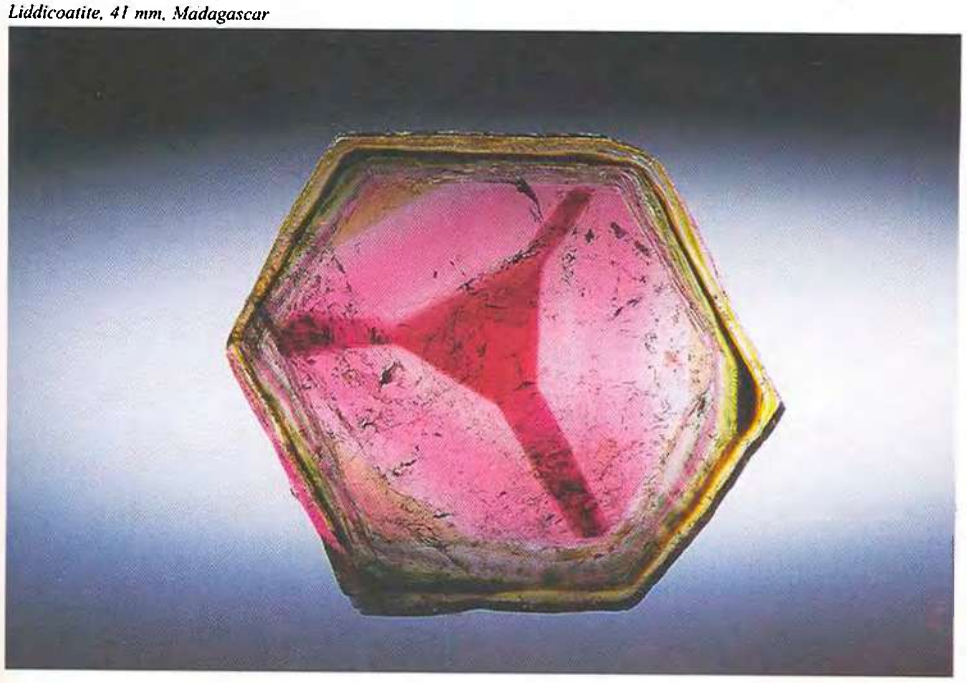

Foitite

TOURMALINE GROUP

□ $Fe_2AlAl_6(BO_3)_3Si_6O_{18}(OH)_4$

TRIGONAL ● ● ●

Properties: C – black to black-purple; S – white; L – vitreous to dull; D – translucent to opaque; DE – 3.3; H – 7; CL – none; F – uneven to conchoidal; M – prismatic crystals, acicular aggregates.
Origin and occurrence: Magmatic in granitic pegmatites and granites; hydrothermal in pegmatite cavities. Foitite typically forms black tips of elbaite crystals (moors heads) in Dobra Voda, Czech Republic; San Piero in Campo, Elba, Italy; White Queen mine, Pala, California, USA. It occurs together with schorl in some granitic pegmatites, like Rozna, Czech Republic.

Rossmanite

TOURMALINE GROUP

□ $LiAl_2Al_6(BO_3)_3Si_6O_{18}(OH)_4$

TRIGONAL ●

Properties: C – pink; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.1; H – 7; CL – none; F – uneven to conchoidal; M – prismatic crystals, columnar aggregates.
Origin and occurrence: Magmatic in granitic pegmatites. Prismatic crystals, up to 20 mm (5/32 in) long, were found in massive lepidolite in Rozna, also in Lastovicky, Czech Republic.

Foitite, 20 mm x, Dobra Voda, Czech Republic

Dioptase

$Cu_6Si_6O_{18} \cdot 6\,H_2O$

TRIGONAL ● ●

Properties: C – emerald-green to blue-green; S – light blue-green; L – vitreous to dull; D – transparent to translucent; DE – 3.3; H – 5; CL – good; F – uneven to conchoidal; M – long to short prismatic crystals, granular.
Origin and occurrence: Secondary in the oxidation zone of Cu deposits, associated with other secondary Cu minerals. Beautiful crystals up to 50 mm (2 in) across occur in Tsumeb, Namibia; Altyn Tyube, Kazakhstan; Renneville, Congo. It also comes from the Mammoth mine, Tiger, Arizona, USA.

Milarite

$KCa_2AlBe_2Si_{12}O_{30} \cdot H_2O$

HEXAGONAL ● ●

Properties: C – colorless, white, gray, light green; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.5; H – 5-6; CL – none; F – uneven to conchoidal; M – long prismatic to acicular crystals, granular.
Origin and occurrence: Hydrothermal in the Alpine-type veins, in pegmatites and hydrothermal veins, associated with adularia. Crystals up to 40 mm (1 5/16 in) long come from Jaguaraçu, Minas Gerais, Brazil. It was also found in St. Gotthard, Switzerland; Klein Spitzkopje, Namibia; Valencia mine, Guanajuato, Mexico.

Rossmanite, 10 mm xx, Lastovicky, Czech Republic

Dioptase, 40 mm, Mindouli, Zair

Sugilite

$KNa_2(Fe,Mn,Al)_2Li_3Si_{12}O_{30}$

HEXAGONAL ●

Properties: C – purple; S – white; L – vitreous to dull; D – transparent to translucent; DF – 2.7; H – 6-6.5; CL – none; F – uneven to conchoidal; M – long prismatic to acicular crystals, granular.

Origin and occurrence: Hydrothermal in alkaline syenites, associated with pectolite, albite and aegirine. It occurs in Iwagi Island, Japan and Hotazel, South Africa. *Application:* cut and polished as a gemstone and decorative stone.

Milarite, 22 mm, Jaguaracu, Brazil

Sugilite, 40 mm, Hotazel, South Africa

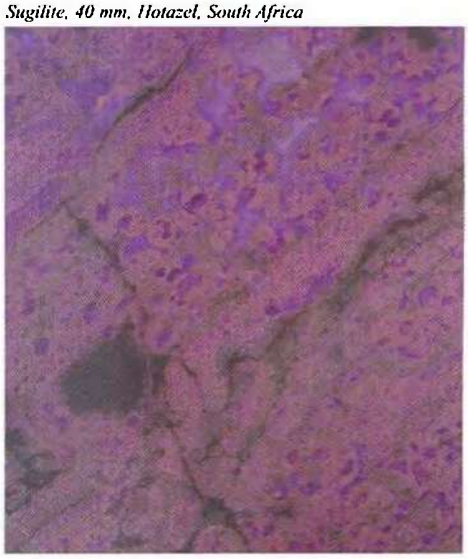

Bronzite, 15 mm, Minas Gerais, Brazil

Enstatite

PYROXENE GROUP

$Mg_2Si_2O_6$

ORTHORHOMBIC ● ● ● ●

Varieties: bronzite

Properties: C – colorless, gray, yellowish, greenish, brown (bronzite); S – white; L – vitreous to dull; D – transparent to opaque; DE – 3.2; H – 5-6; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular.

Origin and occurrence: Magmatic in ultrabasic rocks, gabbros, peridotites and in meteorites; metamorphic in marbles, associated mainly with olivine and pyrope. Known from Bamle, Norway in crystals up to 50 cm (20 in) long. It was originally described from Ruda nad Moravou, Czech Republic. Gemmy crystals up to 20 mm ($^{25}/_{32}$ in) long occur in Brumado, Bahia, Brazil.

Diopside

PYROXENE GROUP

$CaMgSi_2O_6$

MONOCLINIC ● ● ● ● ●

Varieties: chrome diopside, fassaite, jeffersonite

Properties: C – light green, dark green (chrome diopside), colorless, white, gray, brown, rare blue; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.3; H – 5.5-6.5; CL – good; F – uneven to conchoidal; M – long to short prismatic crystals, granular.

Chrome diopside, 5 mm grains, Inagli, Russia

Origin and occurrence: Metamorphic in Ca-rich rocks, skarns, pyroxene gneisses, marbles; magmatic in basic igneous rocks, pegmatites and meteorites; hydrothermal in the Alpine-type veins. Fassaite occurs in skarns, chrome diopside in metamorphic deposits of Cu and Cr; jeffersonite in metamorphic deposits of Mn and Zn. Diopside is a typical rock-forming mineral.

Diopside, 22 mm x, Ala, Italy

associated with plagioclase, grossular and epidote. Well formed crystals are relatively rare. Jeffersonite forms prismatic crystals up to 250 mm (9 13/16 in) long in Franklin, New Jersey, USA. Crystals in Corrego Setuba, Minas Gerais, Brazil reach up to 30 cm (12 in). Other typical localities are Zillertal, Austria; Nordmarken, Sweden; Orford mine, Quebec, Canada. Fassaite is known from Val di Fassa, Italy. Chrome diopside crystals reach up to 100 mm (4 in) in Outokumpu, Finland and gem rough recently found in Inagli, Yakutia, Russia.

Hedenbergite

PYROXENE GROUP

$CaFeSi_2O_6$

MONOCLINIC ● ● ● ●

Properties: C – dark green, brown-green, brown to black; S – white to gray; L – vitreous to dull; D – transparent to translucent; DE – 3.6; H – 6; CL – good; F – uneven to conchoidal; M – long to short prismatic crystals, granular.

Origin and occurrence: Metamorphic in Ca-rich rocks, like Fe-skarns and pyroxene gneisses; magmatic in some granites and syenites, associated with magnetite, grossular and epidote. Crystals are relatively rare, reaching up to 50 mm (2 in) from Dalnegorsk, Russia and Franklin, New Jersey, USA. Also occurs in Nordmarken, Sweden. Large lamellar aggregates come from skarns in Rio Marina, Elba, Italy.

Hedenbergite, 40 mm, Dalnegorsk, Russia

Augite, 17 mm, Firmerich, Germany

Augite

PYROXENE GROUP

$(Ca,Mg,Fe,Al)(Si,Al)_2O_6$

MONOCLINIC ● ● ● ● ●

Properties: C – dark brown to black; S – gray-green; L – vitreous to dull; D – translucent to opaque; DE – 3.6; H – 6; CL – good; F – uneven to conchoidal; M – short prismatic crystals, granular.

Origin and occurrence: Magmatic in basic rocks (basalts, gabbros, diabases and their tuffs); rare metamorphic in skarns. It is a typical rock-forming mineral, known from many localities. Well formed crystals up to 150 mm (6 in) across occur mainly in volcanic rocks, such as near Lake Clear, Ontario, Canada. Also known from Laacher See, Germany; Lukov and Pa_kapole, Czech Republic, in crystals, up to 50 mm (2 in) in size.

Omphacite, 90 mm, St.Leonhard, Austria

Omphacite

PYROXENE GROUP

$(Ca,Na)(Mg,Fe^{2+},Fe^{3+},Al)(Si,Al)_2O_6$

MONOCLINIC ● ● ● ●

Properties: C – green to dark green; S – gray-white; L – vitreous to dull; D – translucent to opaque; DE – 3.3; H – 5-6; CL – good; F – uneven to conchoidal; M – short prismatic crystals, acicular aggregates, granular. *Origin and occurrence:* Metamorphic in ultrabasic and basic rocks, originated under high-temperature, usually associated with pyrope, diopside and kyanite. It occurs in eclogites and granulites in Rubinberg, Germany; Headsburg, California, USA and elsewhere.

Jadeite, 85 mm, Guerro Negro, Baja California, Mexico

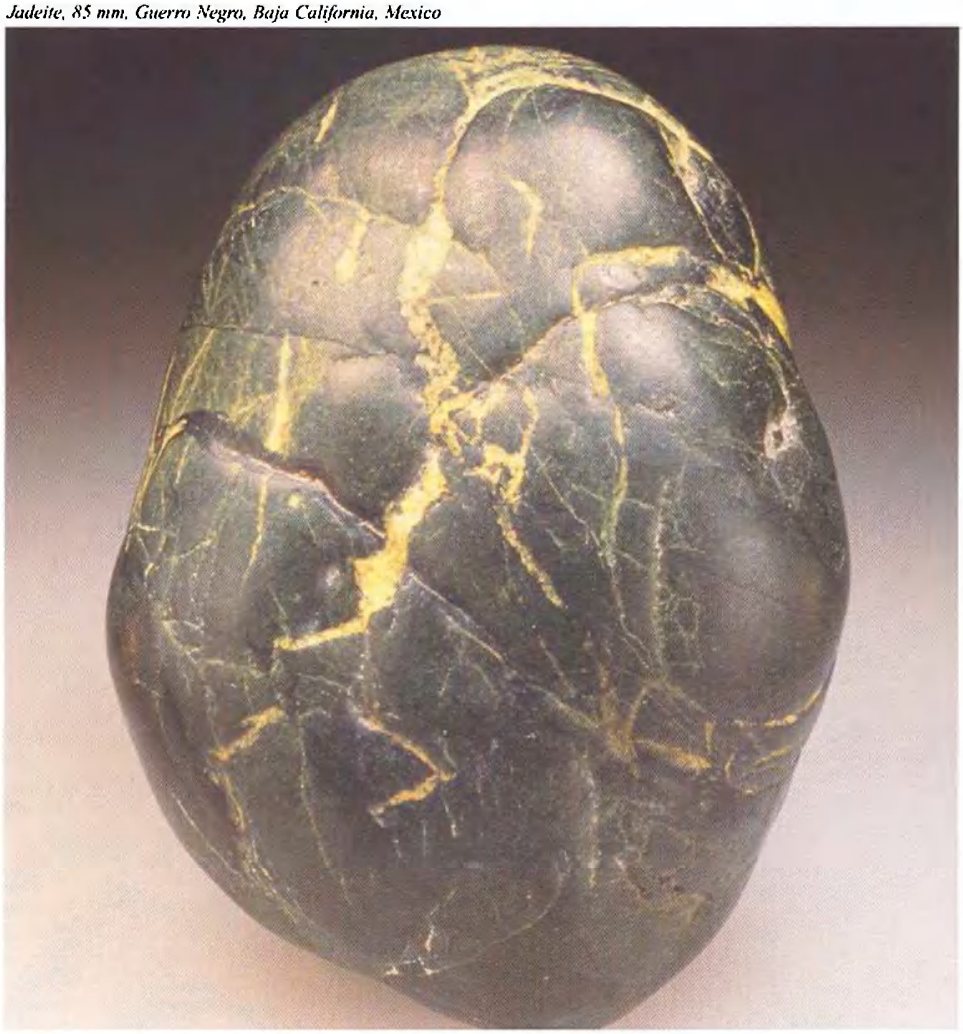

Nephrite, 60 mm, Jordanow, Poland

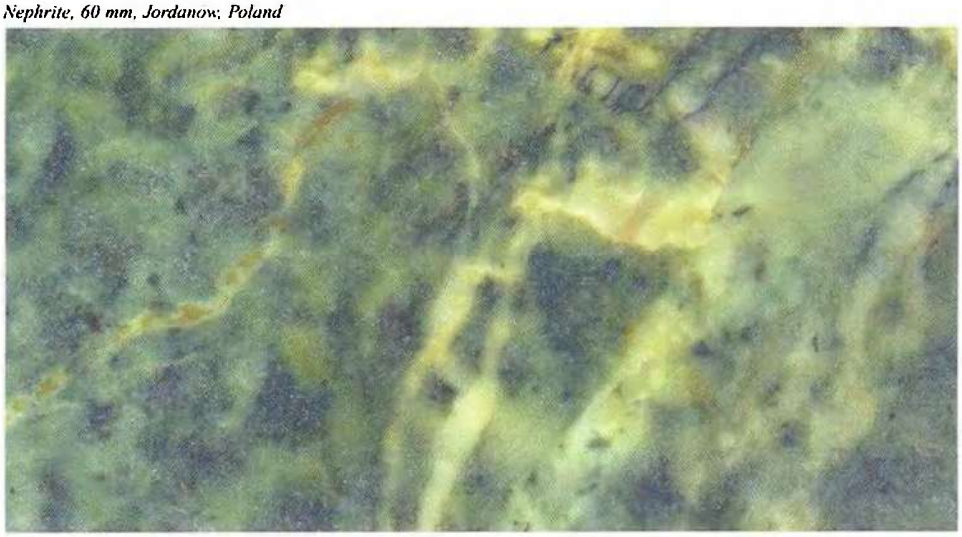

Jadeite

PYROXENE GROUP

$NaAlSi_2O_6$

MONOCLINIC ● ●

Varieties: nephrite

Properties: C – white, lavender to gray, in aggregates also light green (nephrite); S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.2; H – 6; CL – good; F – uneven to conchoidal; M – prismatic crystals, acicular aggregates, granular.
Origin and occurrence: Exclusively metamorphic in strongly metamorphosed rocks; also in placers. Blocks, weighing several tons, are known from Ben Sur, California, USA. It also come from Tawmaw, Burma; New Zealand; Tibet and elsewhere.
Application: as a material for carvings and decorative purposes.

Aegirine

PYROXENE GROUP

$NaFe^{3+}Si_2O_6$

MONOCLINIC ● ● ●

Properties: C – dark green to black-green; S – light yellow-gray; L – vitreous to dull; D – translucent to opaque; DE – 3.6; H – 6; CL – good; F – uneven to conchoidal; M – long prismatic to acicular crystals, acicular aggregates.
Origin and occurrence: Magmatic, mainly in alkaline magmatic rocks (syenites, carbonatites, alkaline granites and their pegmatites); rare hydrothermal in sediments. Well formed prismatic crystals up to 150 mm (6 in) long come from Mount Malosa, Malawi. Crystals up to 30 cm (12 in) occur in Langensundsfjord, Norway. Other localities are Mount Karnasurt, Lovozero massif, Kola Peninsula, Russia and Mont St.-Hilaire, Quebec, Canada.

Aegirine, 77 mm, Mt.Malosa, Malawi

Spodumene, 52 mm x, Kunar, Afghanistan

Hiddenite, 35 mm, Adams Property, U.S.A.

Kunzite, 118 mm, Nuristan, Afghanistan

Spodumene

PYROXENE GROUP

$LiAlSi_2O_6$

MONOCLINIC ● ● ●

Varieties: kunzite, hiddenite, triphane

Properties: C – white, gray, yellowish (triphane), green (hiddenite), pink to purple (kunzite); S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.2; H – 6.5-7.5; CL – good; F – uneven to conchoidal; M – long to short prismatic crystals, cleavable aggregates, granular; LU – orange.

Origin and occurrence: Magmatic in granitic pegmatites; hydrothermal in pegmatite cavities. It is known from many localities. Its poorly developed crystals up to 6 m (20 ft) long occur in the Etta mine, Keystone, South Dakota and in Kings Mountain, North Carolina, USA. Transparent kunzite crystals up to 40 cm (15¾ in) long come from Mawi, Laghman, Afghanistan. Kunzite crystals up to 280 mm (11 in) long found in the Pala Chief mine, Pala, California, USA. Cracked kunzite crystals, up to 1 m (39⅜ in) long known from Araçuai, Minas Gerais, Brazil. Hiddenite occurs in crystals up to 250 mm (9 13/16 in) long in Resplendor, Minas Gerais, Brazil. It also comes from the Adams property, North Carolina, USA.

Application: raw material for ceramics, kunzite and hiddenite are cut as gemstones.

Carpholite, 70 mm, Horni Slavkov, Czech Republic

Carpholite

$MnAl_2Si_2O_6(OH)_4$

ORTHORHOMBIC ● ●

Properties: C – various hues of yellow; S – white; L – vitreous to dull; D – translucent; DE – 3.O; H – 5-5.5; CL – good; F – uneven to conchoidal; M – acicular crystals and their aggregates.
Origin and occurrence: Hydrothermal in veins, cross-cutting greisens, associated with fluorite, quartz and cassiterite; also in metamorphic rocks. It was described from Horni Slavkov, Czech Republic, where it forms acicular crystals up to 10 mm (3/8 in) long and radial aggregates. It is also known from Meuville, Belgium and Wippra, Germany.

Lorenzenite, 20 mm xx, Flora Mt., Kola, Russia

Lorenzenite

$Na_2Ti_2Si_2O_9$

ORTHORHOMBIC ● ●

Properties: C – brown to black; S – yellowish; L – vitreous to dull; D – translucent to opaque; DE – 3.4; H – 6; CL – good; F – uneven to conchoidal; M – prismatic crystals, columnar aggregates, granular.
Origin and occurrence: Magmatic in alkaline syenites and their pegmatites, associated with astrophyllite, nepheline and aegirine. Crystals up to 80 mm (3 1/8 in) long come from Mount Flora, Lovozero massif, Kola Peninsula, Russia. Other localities are Narssarssuk, Greenland and Mont St.-Hilaire, Quebec, Canada.

Ajoite

$(K,Na)Cu_7AlSi_9O_{24}(OH)_6 \cdot 3\ H_2O$

TRICLINIC ●

Properties: C – blue-green; S – light green; L – vitreous to dull; D – translucent to opaque; DE – 3.0; H – not determined; CL – not determined; F – uneven to conchoidal; M – prismatic crystals, massive.
Origin and occurrence: Secondary in Cu deposits, associated with shattuckite. It occurs rarely as small crystals in the New Cornelia mine, Ajo, Arizona, USA. It is also known as inclusions in quartz crystals in the Messina mine, Transvaal, South Africa.

Ajoite, 60 mm, Ajo, U.S.A.

Holmquistite, 60 mm, Lac Malartic, Canada

Antophyllite

AMPHIBOLE GROUP

$Mg_7Si_8O_{22}(OH,F)_2$

ORTHORHOMBIC ● ● ● ●

Properties: C – white, light brown, yellow, light green; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.4; H – 5.5-6; CL – good; F – uneven to conchoidal; M – long prismatic crystals, columnar and radial aggregates, granular.
Origin and occurrence: Metamorphic in gneisses, serpentinites and at the contact of serpentinites and pegmatites, usually with actinolite or tremolite and cordierite. Crystals up to 150 mm (6 in) long come from the Marbridge No.1 mine, Quebec, Canada. It is also known from Bodenmais, Germany; Snarum, Sweden and Hermanov, Czech Republic.

Anthophyllite, 60 mm, North Carolina, U.S.A.

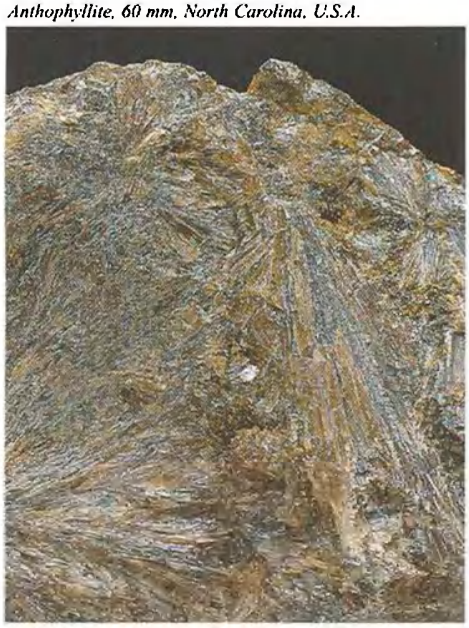

Holmquistite

AMPHIBOLE GROUP

$Li_2Mg_3Al_2Si_8O_{22}(OH,F)_2$

ORTHORHOMBIC ● ● ●

Properties: C – blue, purple, gray, black; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.0; H – 5.5-6; CL – good; F – uneven to conchoidal; M – acicular and prismatic crystals, fibrous aggregates, granular.
Origin and occurrence: Hydrothermal at the contacts between complex Li-bearing pegmatites and amphibolites, sometimes associated with biotite. It occurs in Greenbushes, Western Australia, Australia where its fibers reach up to 180 mm (7 1/16 in) in length. It is also known from Uto, Sweden; Manono, Zaire and Brandbrucken, Austria.

Manganogedrite

AMPHIBOLE GROUP

$Mn_2Fe_5Si_8O_{20}(OH)_2$

MONOCLINIC ● ●

Properties: C – gray, dark green, brown, greenish; S – white; L – vitreous to dull; D – translucent to opaque; DE – 3.5; H – 5-6; CL- good; F – uneven to conchoidal; M – columnar, acicular and radial aggregates, granular.
Origin and occurrence: Metamorphic in contact and regionally metamorphosed Mn rich rocks, associated with chlorite and magnetite. It comes from Dannemora, Sweden and elsewhere.

Manganogedrite, 70 mm, Franz Joseph Iceberg, New Zealand

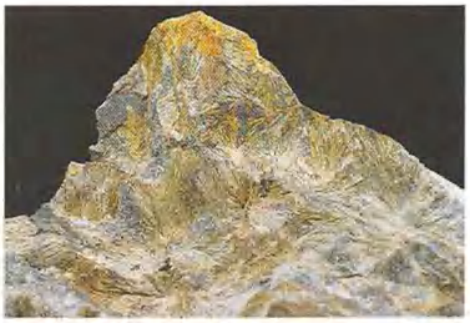

Tremolite, 37 mm, Maricopa Co., Arizona, U.S.A.

Tremolite

AMPHIBOLE GROUP

$Ca_2Mg_5Si_8O_{20}(OH)_2$

MONOCLINIC ● ● ● ● ●

Properties: C – white, gray, greenish, green, brown, pink; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.9; H – 5-6; CL – good; F – uneven to conchoidal; M – columnar, acicular and radial aggregates, granular.

Origin and occurrence: Metamorphic in contact and regionally metamorphosed rocks, dolomites and ultrabasic rocks; hydrothermal in the Alpine-type veins, typically associated with diopside, talc, dolomite and calcite. It is a typical rock-forming mineral, known from many localities. Prismatic crystals up to 40 cm (15¾ in) long come from Brumado, Bahia, Brazil. It occurs also in Campolungo, Switzerland; Zillertal, Austria; Gouverneur, New York, USA and elsewhere.

Actinolite

AMPHIBOLE GROUP

$Ca_2(Mg,Fe)_5Si_8O_{20}(OH)_2$

MONOCLINIC ● ● ● ● ●

Properties: C – light green to almost black; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.2; H – 5-6; CL – good; F – uneven to conchoidal; M – columnar, acicular and radial aggregates, granular.

Origin and occurrence: Metamorphic in contact and regionally metamorphosed rocks, also in dolomites and some basic rocks (amphibolites, shales). It is a typical rock-forming mineral, associated with anthophylite, chlorite, talc, dolomite and calcite. It is known from many localities where it forms columnar aggregates up to 250 mm (9 13/16 in) long, as in Knappenwand and Zillertal, Austria; Val Malenco, Italy; Brumado, BahRa, Brazil and Sobotín, Czech Republic.

Actinolite, 30 mm xx, Austris

Edenite

AMPHIBOLE GROUP

$NaCa_2(Mg,Fe)_5Si_7AlO_{22}(OH)_2$

MONOCLINIC ● ● ● ●

Properties: C – white, gray to dark green; S – white; L – vitreous to dull; D – transparent, translucent to opaque; DE – 3.1; H – 5-6; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular.

Origin and occurrence: Metamorphic in regionally metamorphosed basic rocks and marbles; magmatic in diorites and gabbros. Typical rock-forming mineral, its crystals up to 40 mm (1 9/16 in) long come from Edenville, New York, USA and Bancroft, Ontario, Canada.

Edenite, 48 mm, Wilberforce, Canada

Pargasite

AMPHIBOLE GROUP

$NaCa_2(Mg,Fe)_4AlSi_6Al_2O_{22}(OH)_2$

MONOCLINIC ● ● ● ●

Properties: C – light brown, green-blue, brown to black-brown; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.1; H – 5-6; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular, massive.
Origin and occurrence: Metamorphic in regionally metamorphosed basic rocks, marbles and skarns; magmatic in diorites and gabbros; typical rock-forming mineral. Crystals up to 80 mm (3½ in) across come from the Jensen quarry, Riverside, California, USA; Pargas, Finland; Hunza, Pakistan.

Common Amphibole

AMPHIBOLE GROUP

$(Ca,Na)_2(Mg,Fe,Al)_5(Si,Al)_8O_{22}(OH,F)_2$

MONOCLINIC ● ● ● ● ●

Properties: C – dark green, brown to black; S – gray; L – vitreous to dull; D – translucent to opaque; DE – 3.4; H – 5-6; CL – good; F – uneven to conchoidal; M – short prismatic crystals, columnar aggregates, granular.
Origin and occurrence: Metamorphic in basic rocks, as amphibolites and shales; magmatic in some igneous rocks, as syenites, diorites, andesites and basalts. Typical rock-forming mineral, known from many localities. Large crystals up to 1 m (39½ in) long come from Silver Crater, Ontario, Canada. It is also known from Zillertal, Austria; Lukov, Czech Republic and elsewhere.

Common Amphibole, 50 mm x, Lukov, Czech Republic

Richterite

AMPHIBOLE GROUP

$Na_2Ca(Mg,Fe,Al)_5(Si,Al)_8O_{22}(OH,F)_2$

MONOCLINIC ● ● ●

Properties: C – brown, yellow, dark red, dark green; S – gray; L – vitreous to dull; D – transparent to translucent; DE – 3.1; H – 5-6; CL – good; F – uneven to conchoidal; M – long prismatic crystals, columnar aggregates, granular.
Origin and occurrence: Magmatic in alkaline rocks, as trachytes and alkaline granites; rare metamorphic in some metamorphosed rocks, as skarns and marbles. Crystals up to 100 mm (4 in) long come from Wilberforce, Ontario, Canada and Langban, Sweden.

Pargasite, 70 mm, Limberg, Parainen, Finnland

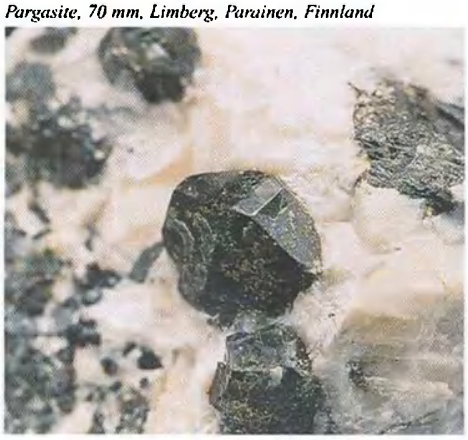

Richterite, 40 mm, Wilberforce, Canada

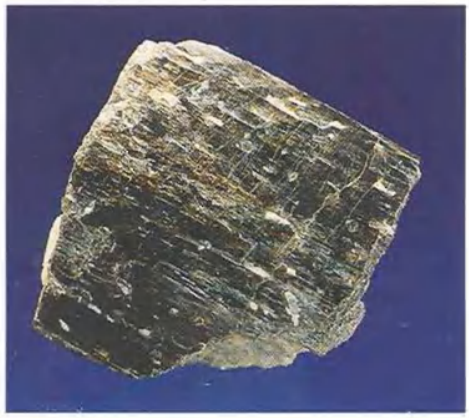

Riebeckite, 45 mm, Griqualand, South Africa

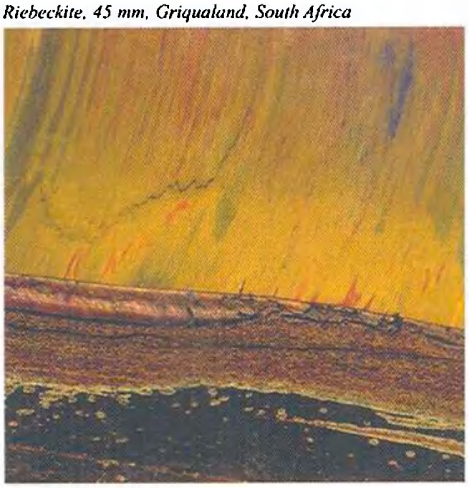

Riebeckite

AMPHIBOLE GROUP

$Na(Fe^{2+},Mg)_3Fe^{3+}{}_2Si_8O_{22}(OH)_2$

MONOCLINIC ● ● ● ●

Varieties: crocidolite (tiger's eye)

Properties: C – dark blue to black, crocydolite – gray-blue; S – white; L – vitreous to dull; D – translucent to opaque; DE – 3.4; H – 5; CL – perfect; F – uneven to conchoidal; M – long prismatic crystals, acicular to fibrous aggregates.
Origin and occurrence: Magmatic in alkaline rocks, as granites, syenites and rhyolites; metamorphic in regionally metamorphosed Fe rich shales. Large crystals up to 150 mm (6 in) long come from Khangay, Mongolia. Crocidolite is famous from Griqualand, South Africa.

Wollastonite, 44 mm, Crestmore, U.S.A.

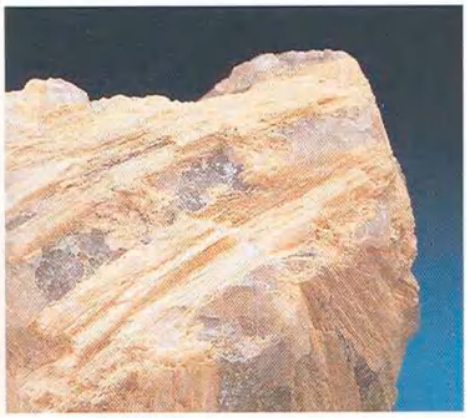

Aenigmatite, 40 mm xx, Eveslogchorr, Kola, Russia

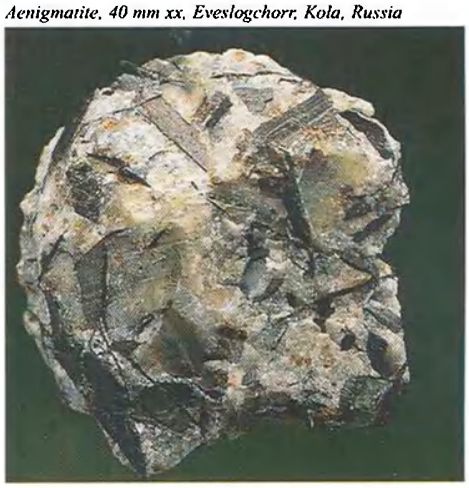

Aenigmatite

$Na_2Fe_5TiSi_6O_{20}$

TRICLINIC ● ●

Properties: C – black; S – red-brown; L – submetallic to dull; D -almost opaque; DE – 3.8; H – 5.5; CL – perfect; F – uneven to conchoidal; M – long prismatic crystals, columnar aggregates, granular.
Origin and occurrence: Magmatic in alkaline syenites and volcanic rocks, associated with aegirine and arfvedsonite. It occurs in Julianehab, Greenland; Lipari Island, Italy and elsewhere.

Wollastonite

$Ca_3Si_3O_9$

TRICLINIC ● ● ● ●

Properties: C - white, gray, light greenish, pink; S – white; L – vitrous to dull; D – transparent to translucent; DE – 3.0; H – 4.5-5; CL – good; F – uneven to conchoidal; M – tabular to short prismatic crystals, columnar and radial aggregates, granular; LU - orange.
Origin and occurrence: Metamorphic mainly in contact metamorphosed rocks, as marbles, skarns, less common in pyroxene gneisses and quartzites, associated with grossular, diopside and vesuvianite. Columnar and acicular aggregates up to 180 mm ($7^1/_{16}$ in) long occur in the Strickland quarry, Connecticut, USA. It also come from Crestmore, California, USA; Ciclova, Romania and Zulova , Czech Republic. Short prismatic crystals up to 30 cm (12 in) across are known from the Santa Fe mine, Chiapas, Mexico.

Bustamite, 20 mm xx, Broken Hill, Australia

Bustamite
$(Mn,Ca)_3Si_3O_9$

TRICLINIC ● ● ●

Properties: C – pink to red-brown; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.4; H – 5.5-6.5; CL – good; F – uneven to conchoidal; M – tabular crystals, acicular aggregates, massive.
Origin and occurrence: Metamorphic or hydrothermal in Mn rich metamorphic rocks or skarns. It is associated with rhodonite and other Mn silicates in Franklin, New Jersey, USA; Langban, Sweden; Broken Hill, New South Wales, Australia.

Pectolite, 70 mm, Zelechovske udolí, Czech Republic

Pectolite
$Ca_2NaSi_3O_8(OH)$

TRICLINIC ● ● ●

Properties: C – white, pinkish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.9; H – 4.5-5; CL – good; F – uneven to conchoidal; M B acicular and radial aggregates; LU – locally yellow to orange.
Origin and occurrence: Hydrothermal in basalt cavities, less frequently along the cracks in marbles, associated with zeolites. Classic localities are West Paterson, New Jersey, USA, where it forms acicular spherical aggregates up to 180 mm (7 1/16 in) across; also Monte Baldo, Italy; Zelechovske udolí, Czech Republic. Prismatic crystals up to 50 mm (2 in) long come from Mont St.-Hilaire, Quebec, Canada.

Serandite
$Mn_2NaSi_3O_8(OH)$

TRICLINIC ●

Properties: C – pink, red; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.9; H – 4.5-5; CL – good; F – uneven to conchoidal; M – tabular and prismatic crystals, granular.
Origin and occurrence: Hydrothermal in cavities of igneous rocks (alkaline syenite, carbonatite) together with aegirine and analcime. Perfect pink crystals up to 200 mm (7 7/8 in) long come from Mont St.-

Serandite, 25 mm, Mont St.-Hilaire, Canada

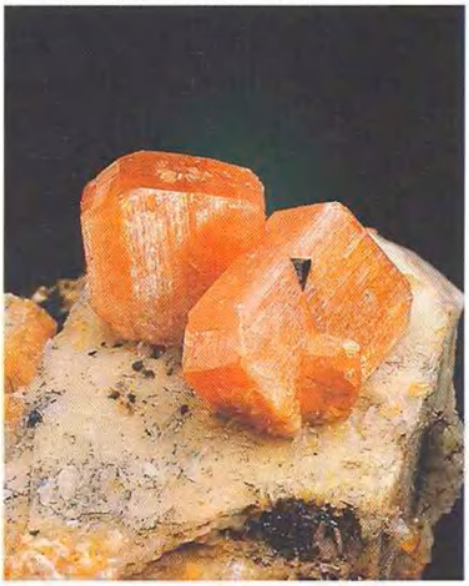

Hilaire, Quebec, Canada; crystals up to 60 mm (2⅜ in) long found in the Yubileinaya dike, Mount Karnasurt, Lovozero massif, Kola Peninsula, Russia.

Charoite

$(K,Sr,Ba)(Ca,Na)_2(Si,Al)_4O_{10}(OH,F)$

TRICLINIC ● ●

Properties: C – purple; S – white; L – vitreous to dull; D – translucent to opaque; DE – 2.6; H – 5-6; CL – good; F – uneven; M – fibrous aggregates, massive.
Origin and occurrence: Hydrothermal in alkaline igneous rocks, associated with aegirine and nepheline. Rich aggregates occur in Sirenevyi Kamen, Chara river basin, Murun massif, Yakutia, Russia.
Application: cut and polished as decorative stone.

Agrellite

$NaCa_2Si_4O_{10}F$

TRICLINIC ●

Properties: C – gray-white to greenish; S – white; L – vitreous, dull to pearly; D – transparent to translucent; DE – 2.9; H – 5.5; CL – good; F – uneven; M – long prismatic crystals, granular.
Origin and occurrence: Metamorphic in alkaline gneisses. It occurs as prismatic crystals up to 100 mm (4 in) long in the Kipawa river basin, Quebec, Canada.

Okenite

$CaSi_2O_4(OH)_2 \cdot H_2O$

TRICLINIC ● ●

Properties: C – white, yellowish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.3; H – 4.5-5; CL – good; F – uneven to conchoidal; M B blade-shaped crystals, acicular crystals and their spherical aggregates.

Charoite, 50 mm, Sirenevyi Kamen, Russia

Origin and occurrence: Hydrothermal in basalt cavities together with zeolites. Rich acicular spherical aggregates up to 80 mm (3⅛ in) in diameter come from basalt cavities in the vicinity of Poona, India; also known from Faeroe Islands.

Agrellite, 120 mm, Kipawa, Canada

Okenite, 125 mm, Mumbai, India

Xonotlite, 20 mm vein, Stare Ransko, Czech Republic

Xonotlite

$Ca_6Si_6O_{17}(OH)_2$

MONOCLINIC ● ●

Properties: C – white, pinkish, gray; S – white; L – vitreous, dull to pearly; D – transparent to translucent; DE – 2.7; H – 6.5; CL – good; F – uneven to conchoidal; M – acicular to fibrous aggregates, massive.
Origin and occurrence: Hydrothermal in cavities of basic and ultrabasic rocks and in the Alpine-type veins, associated with calcite and zeolites; rare metamorphic in skarns and in contact metamorphosed marbles. It comes from Tetela de Xonotla, Puebla, Mexico; Crestmore, California, USA; Mihara mine, Japan. It is was also found in Stare Ransko, Czech Republic.

Elpidite

$Na_2ZrSi_6O_{15} \cdot 3\,H_2O$

ORTHORHOMBIC ● ●

Properties: B – colorless, yellowish, red; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.6; H – 5.5-6.5; CL – good; F – uneven to conchoidal; M – long prismatic crystals, fibrous aggregates.
Origin and occurrence: Hydrothermal in albitized parts of alkaline pegmatites, associated with albite and aegirine. Perfect crystals up to 30 cm (12 in) long come from Tarbagatay, Kazakhstan. Crystals up to 200 mm (7⅞ in) long found at Mont St.-Hilaire, Quebec, Canada. Also known from Khan Bogdo, Gobi desert, Mongolia; Umbozero mine on Mount Alluaiv, Lovozero massif, Kola Peninsula, Russia.

Elpidite, 128 mm, Mont St.-Hilaire, Canada

Rhodonite, 35 mm x, Broken Hill, Australia

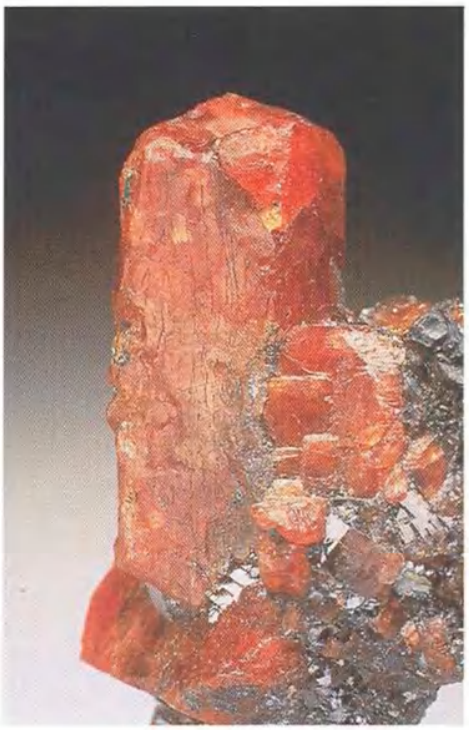

Rhodonite, 70 mm, Maloye Sedelnikovo, Ural Mts., Russia

Rodonite

$CaMn_4Si_3O_9$

TRICLINIC ● ● ●

Properties: C – pink, red, red-brown, gray; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.7; H – 5.5-6.5; CL – good; F – uneven to conchoidal; M – tabular crystals, granular, massive.
Origin and occurrence: Metamorphic in Mn-rich rocks, hydrothermal in ore veins, associated with spessartine and Mn oxides. Perfect crystals up to 200 mm (7⅞ in) long come from Franklin, New Jersey, USA; also from Broken Hill, New South Wales, Australia.
Granular and massive aggregates occur in Langban, Sweden and in Maloye Sidelnikovo, Ural mountains, Russia.
Application: cut and polished as a decorative stone.

Babingtonite

$Ca_2Fe^{2+}Fe^{3+}Si_5O_{14}(OH)$

TRICLINIC ● ●

Properties: C – green-black to black-brown; S – green-gray; L – vitreous to dull; D – translucent to opaque; DE – 3.3; H – 5. 5-6; CL – good; F – uneven to conchoidal; M – short prismatic crystals, granular.
Origin and occurrence: Hydrothermal along the cracks in rocks, associated with prehnite, adularia and epidote.
Well-formed crystals about 20 mm (25/32 in) across are known from Arendal, Norway. Other localities are Westfield, Massachusetts, USA; Baveno, Italy.

Babingtonite, 5 mm x, Mumbai, India

Kinoite
$Ca_2Cu_2Si_3O_8(OH)_4$

MONOCLINIC ●

Properties: C – blue; S – light blue; L – vitreous to dull; D – transparent to translucent; DE – 3.2; H – 5; CL B perfect; F – uneven to conchoidal; M – prismatic crystals, granular.
Origin and occurrence: Hydrothermal along cracks in rocks, associated with apophyllite and copper. Small crystals were found in the Christmas mine, Santa Rita mountains, Arizona, also in the Kearsarge vein, Keweenaw Peninsula, Michigan, USA.

Inesite
$Ca_2Mn_7Si_{10}O_{28}(OH)_2 \cdot 5\ H_2O$

TRICLINIC ● ●

Properties: C – pink, brown; S – white; L – vitreous to dull; D – translucent; DE – 3.0; H – 5.5; CL – good; F – uneven to conchoidal; M – short prismatic and tabular crystals, columnar and radial aggregates, granular.
Origin and occurrence: Hydrothermal in ore veins, associated with rhodochrosite, rhodonite and calcite. Rich aggregates and well-formed crystals about 10 mm (⅜ in) across found in Langban, Sweden; also in Hale Creek, California, USA; Broken Hill, New South Wales, Australia. Its spherical aggregates up to 30 mm (1⅛ in) in diameter are known from the Wessels mine, Kuruman, South Africa.

Kinoite, 36 mm, Christmas Mine, U.S.A.

Neptunite
$KNa_2LiFe_2Ti_2Si_9O_{22}$

MONOCLINIC ● ●

Properties: C – dark brown to black; S – red-brown; L – vitreous to submetallic; D – translucent to opaque; DE – 3. 2; H – 5-6; CL – good; F – uneven to conchoidal; M – long prismatic crystals, granular.
Origin and occurrence: Hydrothermal in veins, associated with natrolite, benitoite and eudialyte. Well-formed prismatic crystals up to 80 mm (3⅛ in) long come from the Benitoite Gem mine, San Benito Co., California, USA. It is also known from Mont St.-Hilaire, Quebec, Canada and Narssarssuk, Greenland.

Epididymite
$NaBe_2Si_3O_7(OH)$

ORTHORHOMBIC ● ●

Properties: C – white, gray, yellowish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.6; H – 6-7; CL – good; F – uneven to conchoidal;

Inesite, 50 mm, Kuruman, South Africa

Neptunite, 28 mm, San Benito Co., U.S.A.

M – tabular crystals, lamellar aggregates, granular. *Origin and occurrence:* Hydrothermal in cavities of alkaline pegmatites, associated with zeolites. Tabular crystals up to 60 mm (2³/₄ in) across come from Mount Malosa, Malawi. It is also known from Langesundsfjord, Norway; Mont St.-Hilaire, Quebec, Canada; Narssarssuk, Greenland.

Bavenite, 40 mm, Yermolayevskove, Russia

Bavenite

$Ca_4Al_2Be_2Si_9O_{26}(OH)_2$

ORTHORHOMBIC ● ● ●

Properties: C – white, yellowish, pinkish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.8; H – 5.5; CL – good; F – uneven to conchoidal; M – prismatic to tabular crystals, radial and lamellar aggregates. *Origin and occurrence:* Hydrothermal in pegmatic cavities with albite: typically produced by beryl replacement in pegmatites or helvite replacement in skarns. Found in Baveno, Italy with crystals up to 20 mm (²⁵/₃₂ in) across. Also known from Strzegom, Poland. Crystals up to 4 mm (⁵/₃₂ in) across found in the Hewitt quarry, USA.

Epididymite, 70 mm, Apatity, Kola, Russia

Prehnite, 28 mm, Brandberg, Namibia

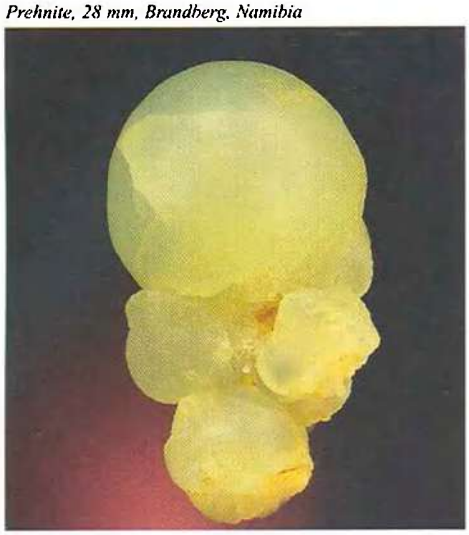

Prehnite
$Ca_2Al_2Si_3O_{10}(OH)_2$

ORTHORHOMBIC ● ● ● ●

Properties: C – white, greenish to green, yellow, gray; S – white; L – vitreous; PS – transparent to translucent; DE – 2.9; H – 6-6.5; CL – good; F – uneven to conchoidal; M – tabular to prismatic crystals, usually forming botryoidal or spherical aggregates, granular.
Origin and occurrence: Hydrothermal in basalt cavities and in the Alpine-type veins, commonly associated with calcite, albite, epidote and zeolites. Rich aggregates are known from West Paterson, New Jersey, USA; Val di Fassa, Italy; Poona, India.

Astrophyllite, 30 mm xx, Khibiny Massif, Kola, Russia

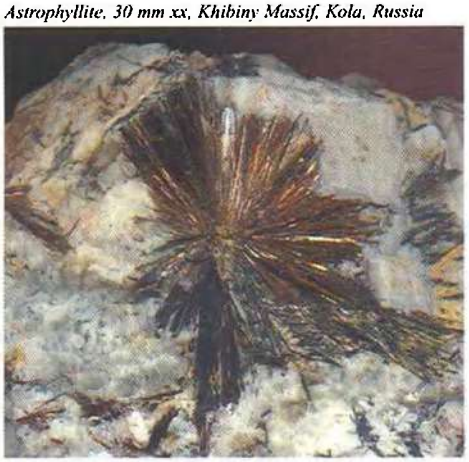

Fluorapophyllite, 47 mm, Poona, India

Crystals up to 45 mm (1¾ in) across found in Copper Valley, Brandberg, Namibia, Asbestos, Quebec, Canada and Talnakh, Siberia, Russia.

Astrophyllite
$(K,Na)_3(Fe,Mn)_7Ti_2Si_8O_{24}(O,OH)_7$

TRICLINIC ● ●

Properties: C – various hues of yellow; S – white; L – vitreous to dull; D – transparent to translucent; DE – 3.3; H – 3; CL – good; F – uneven to conchoidal; M – long prismatic crystals, bladed aggregates.
Origin and occurrence: Magmatic in alkaline pegmatites. Crystals and their radial aggregates reaching up to 100 mm (4 in) come from Mount Eveslogchorr, Khibiny massif, Kola Peninsula., Russia. It is also known from Mont St.-Hilaire, Quebec, Canada.

Fluorapophyllite
$KCa_4Si_8O_{20}(F,OH) . 8 H_2O$

TETRAGONAL ● ● ●

Properties: C – white, greenish yellowish, pinkish; S – white; L – greasy to pearly; D – transparent to translucent; DE – 2.4; H – 4.5-5; CL – perfect; F – uneven; M – prismatic and tabular crystals, lamellar aggregates, massive.
Origin and occurrence: Hydrothermal in cavities in volcanic rocks, along the cracks in the Alpine-type veins, in ore veins and pegmatites. Well-formed crystals up to 100 mm (4 in) long are known from Jalgaon and Poona, India; Centreville and Fairfax,

Pyrophyllite, 37 mm, California, U.S.A.

Agalmatolite, 70 mm, India

Virginia, USA. Other localities are St. Andreasberg, Germany; Bento Gonçalves, Rio Grande do Sul, Brazil.

Pyrophyllite

$Al_2Si_4O_{10}(OH)_2$

MONOCLINIC ● ● ● ●

Varieties: agalmatolite

Properties: C – white, greenish, yellow, gray; S – white; L – greasy to pearly; D – translucent to transparent; DE – 2.8; H – 1-2; CL – perfect; F – uneven; M B foliated and lamellar aggregates, massive (agalmatolite), radial to acicular aggregates.
Origin and occurrence: Metamorphic in Al-rich rocks, together with andalusite and kyanite; hydrothermal in veins with quartz and micas. Rich foliated and radial aggregates and poorly-formed crystals up to 10 mm (3/8 in) across are known from Zermatt, Switzerland; Berezovsk, Ural mountains, Russia and the Champion mine, California, USA.
Application: heat-resistant material, agalmatolite as decorative stone.

Talc

$Mg_3Si_4O_{10}(OH)_2$

MONOCLINIC ● ● ● ● ●

Properties: C – white, greenish, yellow, pinkish, gray; S – white; L – greasy to pearly; D – transparent to translucent; DE – 2.8; H – 1; CL – perfect; F – uneven; M – tabular crystals, foliated aggregates, massive.
Origin and occurrence: Metamorphic in metamorphosed basic rocks and dolomites, together with actinolite, dolomite and chlorite; hydrothermal in veins. Tabular crystals up to 20 mm (25/32 in) across come from Brumado, Bahia, Brazil and Chester, Massachusetts, USA. Rich foliated aggregates are known from Zillertal, Austria. Pseudo-morphs after quartz crystals come from St. Gotthard, Switzerland. Large deposits of massive talc are mined in China.
Application: as a filling material in textile, paper and chemical industries, heat-resistant material.

Talc, 61 mm, Rochester, U.S.A.

Muscovite
MICA GROUP
$KAl_3Si_3O_{10}(OH,F)_2$

MONOCLINIC ● ● ● ● ●

Varieties: fuchsite
Properties: C – white, greenish, yellowish, pinkish, green (fuchsite), gray; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 2.8; H – 2.5-3; CL – perfect; F – uneven; M – tabular crystals, lamellar and scaly aggregates, massive.
Origin and occurrence: Metamorphic in various rock types, as mica schists, gneisses; magmatic in granites and pegmatites; hydrothermal in veins or next to ore veins (fuchsite). Important rock-forming mineral, usually associated with quartz, K-feldspar, albite and biotite. Its sheets, reached up to 5 x 3 m in size and weighed up to 85 tons in the Inikurti mine, Nellore, India. Large sheets are also known from Custer, South Dakota, USA. Well-formed crystals up to 100 mm (4 in) across come from Alabashka near Murzinka, Ural mountains, Russia. Other famous localities include Mamsk, Ural mountains, Russia and Cruzeiro mine, Minas Gerais, Brazil, where gemmy crystals were found.
Application: insulation material in electrical applications and construction.

Muscovite, 35 mm, Plumos, California, U.S.A.

Celadonite
MICA GROUP
$KFe^{3+}(Mg,Fe^{2+},Al)Si_4O_{10}(OH)_2$

MONOCLINIC ● ● ● ●

Properties: C – light green to blue-green; S – white; L – dull; D – translucent to almost opaque; DE – 3.0; H – 2; CL – perfect; F – uneven; M – earthy, sometime small scaly aggregates, massive.
Origin and occurrence: Hydrothermal as a product of mafic mineral replacement in volcanic rocks, associated with zeolites, prehnite and calcite. Famous localities are Val di Fassa and Monte Balda, Italy; also known from Bisbee, Arizona, USA.

Boromuscovite
MICA GROUP
$KAl_2BSi_3O_{10}(OH)_2$

MONOCLINIC ●

Properties: C – yellowish; S – white; L – dull; D – transparent to translucent; DE – 2.8; H – 2.5; CL – perfect; F – uneven; M – small scaly aggregates, massive.
Origin and occurrence: Hydrothermal in cavities in

Fuchsite, 4 mm xx, Pamir, Tadzhikistan

Celadonite, 60 mm, Unhošťany, Czech Republic

complex pegmatites, associated with albite, elbaite and lepidolite. Fine-grained scaly aggregates were described from the Little Three mine, Ramona, California, USA; also in Řečice, Czech Republic.

Paragonite

MICA GROUP

$NaAl_3Si_3O_{10}(OH,F)_2$

MONOCLINIC ● ● ●

Properties: C – white, greenish, yellowish, pinkish; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 2.8; H – 2.5; CL B perfect; F – uneven; M – tabular crystals, scaly aggregates, massive.

Origin and occurrence: Metamorphic in various rock types, as mica schists and gneisses. Fine-grained scaly aggregates are known from Pizzo Forno, Switzerland and elsewhere.

Boromuscovite, 40 mm, Ramona, U.S.A.

Paragonite, 80 mm, St.Gotthard, Switzerland

Glauconite

MICA GROUP

$K_{0.8}(Al,Fe^{2+},Fe^{3+},Mg)(Si,Al)_4O_{10}(OH)_2$

MONOCLINIC ● ● ● ● ●

Properties: C – light green, yellow-green to blue-green; S – light green; L – dull; PS – translucent to opaque; DE – 2.9; H – 2; CL B perfect; F – uneven; M B earthy and platy aggregates, massive.

Origin and occurrence: Hydrothermal in sedimentary and volcanic sedimentary rocks, also as a product of replacement of mafic minerals in volcanic rocks. Common in sandstones and limestones, locally associated also with phosphorites. It occurs in many localities in the Karpathians, Poland; in Polabi region, Czech Republic. Massive aggregates are known from the N'Chwaning No. 2 mine, Kuruman, South Africa.

Glauconite, 80 mm, Malomerice, Czech Republic

Annite, 120 mm. Budec, Czech Republic

Annite

MICA GROUP

$KFe_3AlSi_3O_{10}(OH,F)_2$

MONOCLINIC ● ● ●

Properties: C – black, locally with reddish or greenish tint; S – colorless; L – vitreous to pearly; D – transparent to translucent, locally opaque; DE – 2.8; H – 2.5-3; CL – perfect; F – uneven; M – tabular crystals, lamellar aggregates, massive.
Origin and occurrence: Metamorphic in skarns; magmatic in some pegmatites and granites. Its lamellar aggregates are known from Langban, Sweden and Cape Ann, Massachusetts, USA. Crystals up to 150 mm (6 in) across come from Mont St.-Hilaire, Quebec, Canada.

Phlogopite

MICA GROUP

$KMg_3AlSi_3O_{10}(OH,F)_2$

MONOCLINIC ● ● ● ● ●

Properties: C – light brown, greenish, yellowish, colorless, gray; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 2.8; H – 2.5-3; CL – perfect; F – uneven; M – tabular crystals, platy and scaly aggregates, massive.
Origin and occurrence: Metamorphic in various rock types, as marbles and some ultrabasic rocks, at the contacts of pegmatites and serpentinites, magmatic in some ultrabasic rocks and pegmatites, associated with dolomite, diopside, anthophyllite. Rich scaly aggregates come from many localities, as Pargas, Finland and elsewhere. Crystals up to 5 m (16 ft) across found in Sludyanka, Siberia, Russia. Crystals up to 50 cm (20 in) across occurred in the Gardiner complex, Greenland. The largest crystals, 10 x 5 m (33 x 16 ft) across, weighing up to 90 tons, come from the Lacy mine, Ontario, Canada.
Application: as insulation material in electrical applications.

Phlogopite, 45 mm, Sludyanka, Russia

Biotite

MICA GROUP

$K(Fe,Mg)_3AlSi_3O_{10}(OH,F)_2$

MONOCLINIC ● ● ● ● ●

Properties: C – brown to black, commonly with red or green tint; S – colorless; L – vitreous to pearly; D – transparent to translucent, locally opaque; DE – 2.8; H – 2.5-3; CL – perfect; F – uneven; M – tabular crystals, scaly aggregates, massive.
Origin and occurrence: Metamorphic in various rock types, as mica schists, gneisses, migmatites and different types of metamorphosed shales; magmatic in pegmatites, granites, syenites and diorites, rare in basalts and ultrabasic rocks. Typical rock-forming mineral, usually associated with quartz, feldspars and muscovite. Rich scaly aggregates are known from a pegmatite in Evje, Norway, where its crystals reach up to several meters across. Other famous localities are Bessnes, France; Uluguru mountains, Tanzania; Silver Crater mine, Ontario, Canada and Laacher See, Germany.

Polylithionite

MICA GROUP

$KLi_2AlSi_4O_{10}(F,OH)_2$

MONOCLINIC ● ● ●

Properties: C – gray, colorless, yellowish, purple; S – colorless; L – vitreous to pearly; D – transparent to translucent; DE – 2.8; H – 2.5-3; CL – perfect; F – uneven; M B tabular crystals, scaly aggregates, massive.
Origin and occurrence: Magmatic and locally also hydrothermal in alkaline pegmatites, granites and carbonatites, rare in Li-bearing granitic pegmatites. Well-formed crystals up to 40 mm (1 5/16 in) across are known from cavities in alkaline pegmatites from Mont St.-Hilaire, Quebec, Canada. It comes also from Illimaussaq, Greenland and from cavities in granitic pegmatites in Rožice, Czech Republic and

Biotite, 30 mm x, Ontario, Canada

elsewhere. Large industrial deposits known in Blatchford Lake, Northwest Territories, Canada.
Application: Li and Cs ore.

Trilithionite

MICA GROUP

$KLi_{1,5}Al_{2,5}Si_3O_{10}(F,OH)_2$

MONOCLINIC ● ● ● ●

Properties: C – purple, pink, blue, colorless, green; S – colorless; L – vitreous to pearly; D – transparent to translucent; DE – 2.8; H – 2.5-3; CL – perfect; F – uneven; M – tabular crystals, scaly aggregates, massive.
Origin and occurrence: Magmatic in Li-bearing pegmatites and granites, commonly associated with elbaite, spodumene, petalite, quartz and albite; sometimes hydrothermal in quartz veins and pegmatites. Well-formed crystals from pegmatite cavities are known from Virgem da Lapa, Minas Gerais, Brazil. Massive aggregates come from the Stewart Lithia mine, Pala, California and the Brown Derby No. 1 mine, Colorado, USA; Varutrask, Sweden; Rozna , Czech Republic; Meldon Quarry, Devon, UK.
Application: Li and Cs ore.

Polylithionite, 30 mm, Lovozero Massif, Kola, Russia

Zinnwaldite

MICA GROUP

$KLiFeAl_2Si_3O_{10}(F,OH)_2$

MONOCLINIC ● ● ●

Properties: C – gray, colorless, brown; S – colorless; L – vitreous to pearly; D – transparent to translucent; DE – 3.7; H – 2.5-3; CL – perfect; F – uneven; M – tabular crystals, scaly aggregates, massive.
Origin and occurrence: Hydrothermal in quartz veins and greisens; magmatic in pegmatites and granites, associated with fluorite, cassiterite and wolframite. Rich aggregates with scales up to 100 mm (4 in) across are known from quartz veins in Cínovec, Czech Republic. Crystals up to 150 mm (6 in) across come from pegmatite cavities in Virgem da Lapa, Minas Gerais, Brazil. It also occurs in the Pikes Peak batholith, Colorado, USA; Baveno, Italy.
Application: Li ore.

Trilithionite, 70 mm, Tanco, Canada

Zinnwaldite, 10 mm xx, Cinovec, Czech Republic

Margarite, 73 mm, Chester, U.S.A.

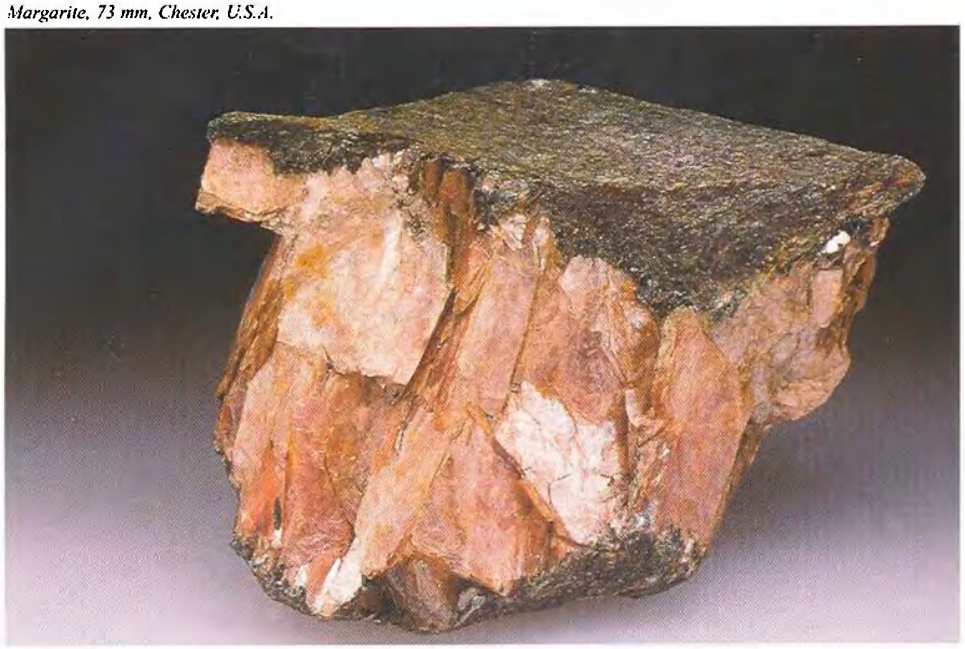

Margarite

$CaAl_4Si_2O_{10}(OH)_2$

MONOCLINIC ● ● ●

Properties: C – pinkish, colorless, yellowish; S – colorless; L – vitreous to pearly; D – transparent to translucent; DE – 3.1; H – 3.5-4.5; CL B perfect; F – uneven; M – tabular crystals, platy and scaly aggregates, massive.

Origin and occurrence: Metamorphic, associated with corundum, diaspore, tourmaline and staurolite in various types of metamorphosed shales. Coarse platy aggregates come from Chester, Massachusetts and Sterling Hill, New Jersey, USA.

Clintonite

$Ca(Mg,Al)_3(Al_3Si)O_{10}(OH)_2$

MONOCLINIC ● ●

Properties: C – colorless, yellowish, pinkish, greenish; S – colorless; L – vitreous to pearly; D – transparent to translucent; DE – 3.1; H – 3.5; CL – perfect; F – uneven; M – tabular crystals, lamellar aggregates, massive.

Origin and occurrence: Metamorphic in contact metamorphosed marbles, associated with vesuvianite, grossular, diopside and spinel. Rich aggregates with lamellae up to 20 mm ($^{25}/_{32}$ in) across occur in Green Monster mountain, Alaska and Crestmore, California, USA and Monzoni, Italy.

Clintonite, 3 mm xx, Pomaz, Hungary

Stilpnomelane, 100 mm, Horni Udoli, Czech Republic

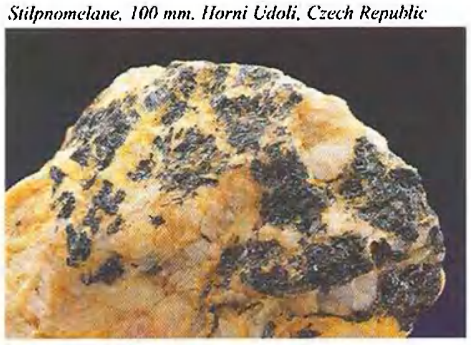

Stilpnomelane

$K(Fe^{2+},Mg,Fe^{3+})_8(Si,Al)_{12}(O,OH)_{27}$

MONOCLINIC ● ● ● ●

Properties: C – black, black-brown, black-green, yellow-brown; S – colorless; L – vitreous to dull; D – translucent to opaque; DE – 2.8; H – 3; CL – perfect; F – uneven; M – tabular crystals, foliated, lath-like and acicular aggregates..
Origin and occurrence: Metamorphic in Fe-rich shales, usually associated with chlorite, magnetite and albite. Crystals up to 20 mm (25/32 in) across come from Jim Pond township, Maine, USA. Foliated aggregates are known from Horní Udolí near Zlate Hory, Czech Republic; Mesabi Range, Minnesota, USA.

Montmorillonite

$(Ca,Na)_{0,33}(Al,Mg)_2Si_4O_{10}(OH)_2 \cdot n\ H_2O$

MONOCLINIC ● ● ● ●

Properties: C – white, yellowish, greenish, bluish; S – white; L – greasy; D – translucent to opaque; DE – 2.3; H – 1-2; CL – perfect, wet massive aggregates are plastic; F – uneven; M – earthy aggregates, massive.
Origin and occurrence: Hydrothermal as a product of replacement of other minerals in volcanic rocks, granitic pegmatites and sediments. It is abundant in many localities. Montmorillonite deposits are known from Antrim, Northern Ireland, UK; in Hungary; Slovakia and many places in the USA.
Application: ceramics and chemical industry.

Saponite

$(Ca,Na)_{0.3}(Fe,Mg)_3(Si,Al)_4O_{10}(OH)_2 \cdot 4\ H_2O$

MONOCLINIC ● ● ●

Properties: C – white, yellowish, gray-green, bluish; S – white; L – greasy; D – translucent to opaque; DE – 2.1; H – 1-2; CL – perfect, massive aggregates are plastic under wet conditions; F – uneven; M – earthy, nodular, massive.
Origin and occurrence: Hydrothermal as a product of replacement of other minerals in volcanic rocks and serpentinites. It occurs together with copper in many places in Keweenaw Peninsula, Michigan, USA. Scales up to 10 mm (3/8 in) across described from Mont St.-Hilaire, Quebec, Canada.

Vermiculite

$(Fe,Mg,Al)_3(Si,Al)_4O_{10}(OH)_2 \cdot 4\ H_2O$

MONOCLINIC ● ● ●

Properties: C – yellow-brown, gray-green, green-brown; S – white; L – greasy; D – translucent to opaque; DE – 2.5; H – 1.5; CL B perfect; F – uneven; M – scaly aggregates, massive; R – it expands when heated.
Origin and occurrence: Hydrothermal as a product of replacement of phlogopite and other mafic micas in various rock types. It occurs in Palbora, South Africa; Milbury, Massachusetts, USA; Kovdor massif, Kola Peninsula, Russia.
Application: electrical applications and paper industry.

Saponite, 2 mm crust, Erdobenye, Hungary

Montmorillonite, Kisthaas, Syria

Vermiculite, 50 mm, Drahonin, Czech Republic

Cookeite, 60 mm, Dobra Voda, Czech Republic

Cookeite

CHLORITE GROUP

$LiAl_4Si_3O_{10}(OH)_8$

MONOCLINIC ● ●

Properties: C - white, yellowish, pinkish, brown; S – colorless; L – vitreous to pearly; D – transparent to translucent; DE – 2.6; H – 2.5-3.5; CL – perfect; F – uneven; M B scaly, locally radial aggregates, massive.

Origin and occurrence: Hydrothermal in cavities in Li-bearing pegmatites, associated with elbaite, lepidolite, fluorapatite and albite, it also replaces spodumene; metamorphic in Al-rich shales, together with diaspore and pyrophyllite. Massive aggregates, replacing spodumene, occur in the Tanco mine, Bernic Lake, Manitoba, Canada. It also comes from cavities in pegmatites in the Little Three mine, Ramona and the Himalaya mine, Mesa Grande, California. It is also known from the Pulsifer quarry, Maine, USA and Muiane, Mozambique.

Clinochlore, 5 mm xx, Sarany, Russia

Clinochlore

CHLORITE GROUP

$(Mg,Fe)_5Al_2Si_3O_{10}(OH)_8$

MONOCLINIC ● ● ● ● ●

Varieties: kammererite

Properties: C – greenish, gray, white, yellowish, brown, red-purple (kammererite); S – colorless; L – vitreous to pearly; D – transparent to translucent; DE – 2.7; H – 2-2.5; CL – perfect; F – uneven; M B poorly-developed tabular crystals, foliated and radial aggregates, massive.

Kammererite, 11 mm xx, Kop Daglari, Turkey

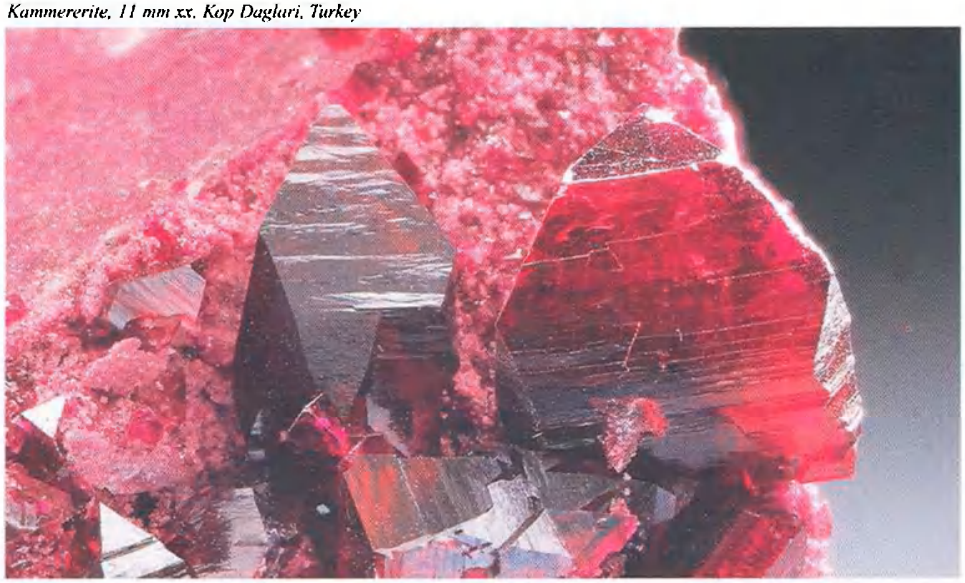

Chamosite, 70 mm, Hajan, Hungary

Origin and occurrence: Metamorphic in various types of shales and marbles; hydrothermal in quartz veins and Alpine-type veins, occasionally replaces certain minerals as biotite. Crystals up to 50 mm (2 in) across, associated with chondrodite and magnetite, are known from the Tilly Foster mine, Brewster, New York and Chester, Pennsylvania, USA. It also occurs in Val d'Ala, Italy and Zillertal, Austria. Scales up to 40 cm (15¾ in) across come from Beramy, Madagascar. Kammererite crystals up to 20 mm (25/32 in) across found in Kop Daolari, Erzerum, Turkey.

Chamosite

CHLORITE GROUP

$(Fe,Mg)_5Al_2Si_3O_{10}(OH)_8$

MONOCLINIC ● ● ● ●

Properties: C – gray, gray-green, brown; S – gray-green; L – vitreous to dull; D – translucent to opaque; DE – 3.2; H – 2-2.5; CL – perfect; F – uneven; M B scaly and oolitic aggregates, massive.
Origin and occurrence: Metamorphic to hydrothermal in various types Fe-rich sediments, typically associated with siderite and magnetite.
It is common in Chamoson, Switzerland; Nucice, Czech Republic. Spherical aggregates up to 15 mm (19/32 in) in diameter found in Mont St.-Hilaire, Quebec, Canada.

Kaolinite

$Al_2Si_2O_5(OH)_4$

TRICLINIC ● ● ● ● ●

Properties: C – white, yellowish, greenish, gray; C – colorless; L – earthy; D – translucent to opaque; DE – 2.6; H – 1; CL – perfect; F – uneven; M – earthy and exceptionally scaly aggregates, massive.
Origin and occurrence: Hydrothermal, as a result of feldspar replacement in various rock types, as granites and arcoses. It forms large deposits in China, France, UK and Czech Republic.
Application: raw material for ceramics.

Dickite, 100 mm, Nowa Ruda, Poland

Dickite

$Al_2Si_2O_5(OH)_4$

MONOCLINIC ● ● ●

Properties: C – white, yellowish; S – colorless; L – earthy to pearly; D – translucent to opaque; DE – 2.6; H – 1; CL – perfect; F – uneven; M – earthy and platy aggregates, massive.
Origin and occurrence: Hydrothermal in cavities in hydrothermal veins, together with quartz, carbonates and sulfides. It occurs in Essen, Germany; Anglesey, Wales , UK; Kladno, Czech Republic. Microscopic crystals come from Mas d'Alary, France.

Kaolinite, 50 mm, St.Austell, UK

Nakrite, 70 mm, Rochlitz, Germany

Nacrite
$Al_2Si_2O_5(OH)_4$

MONOCLINIC ● ● ●

Properties: C – white, yellowish, gray-white; S – colorless; L – earthy to pearly; D – translucent to opaque; DE – 2.6; H – 1; CL – perfect; F – uneven; M – earthy and scaly aggregates, massive.
Origin and occurrence: Hydrothermal in cavities of hydrothermal veins, associated with quartz, fluorite and topaz. It occurs in Horni Slavkov, Czech Republic; Freiberg, Germany. Pseudo-morphs after topaz come from Ouro Preto, Minas Gerais, Brazil.

Antigorite
$Mg_3Si_2O_5(OH)_4$

MONOCLINIC ● ● ● ● ●

Properties: C – white, yellowish, greenish, gray; S – colorless to gray; L – vitreous, pearly, silky, earthy; D – translucent to opaque; DE – 2.6; H – 2.5-3.5; CL – perfect; F – uneven; M B foliated, lamellar and fibrous aggregates, massive.
Origin and occurrence: Metamorphic in serpentinites, marbles and other Mg-rich rocks; hydrothermal in veins cross-secting these rocks. It is important rock-forming mineral. It occurs in Kraubat, Austria; Hrubsice, Czech Republic; Antigorio, Italy and elsewhere.

Antigorite, 110 mm, Felsocsatar, Hungary

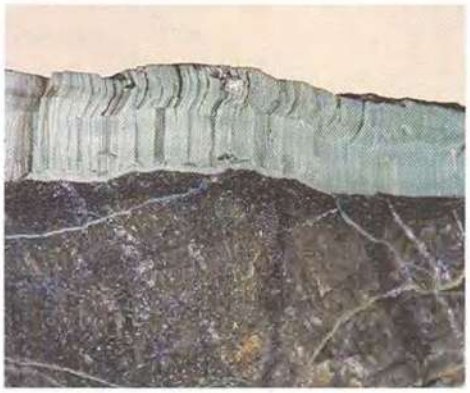

Amesite
$Mg_2Al(SiAl)O_5(OH)_4$

TRICLINIC ● ●

Properties: C – greenish, gray; S – colorless; L – vitreous to pearly; D – translucent; DE – 2.8; H – 2.5-3; CL – perfect; F – uneven; M B foliated to radial aggregates, massive.
Origin and occurrence: Hydrothermal in Mg-rich rocks, associated with diaspore, magnetite and chromite. It comes from Chester, Massachusetts, USA. Crystals up to 40 mm (1 5/16 in) across found in Sarany, Ural mountains, Russia.

Cronstedtite
$Fe^{2+}{}_2Fe^{3+}{}_2SiO_5(OH)_4$

MONOCLINIC ● ●

Properties: C – black, black-brown, black-green; S – dark olive-green; L – vitreous to submetallic; D – translucent to opaque; DE – 3.6; H – 3.5; CL – perfect; F – uneven; M – prismatic crystals, columnar aggregates.
Origin and occurrence: Hydrothermal in ore veins. Fan-shaped aggregates of crystals come from Pribram and Kutna Hora, Czech Republic; Ouro Preto, Minas Gerais, Brazil; Salsigue mine, Auge, France.

Amesite, 40 mm, Sarany, Ural Mts., Russia

Cronstedtite, 42 mm, Hunan, Bolivia

Chrysocolla, 50 mm, Arizona, U.S.A.

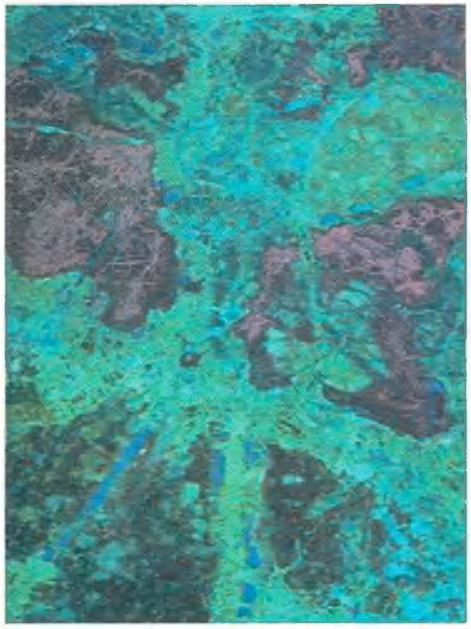

Chrysocolla

$(Cu,Al)_2H_2Si_2O_5(OH)_4 \cdot n\ H_2O$

MONOCLINIC ● ● ● ●

Properties: C – different hues of blue-green; S – white; L – vitreous, greasy to earthy; D – translucent to almost opaque; DE – 2.0-2.4; H – 2-4; CL – none; F – conchoidal to uneven; M – microscopic acicular crystals, botryoidal, stalactitic and earthy aggregates.
Origin and occurrence: Secondary in Cu deposits, associated with malachite and other secondary Cu minerals.
Rich aggregates come from many places in Arizona, USA, like Bisbee and Morenci. It is also known from Mednorudnyansk, Ural mountains, Russia and Broken Hill, New South Wales, Australia.

Allophane

approximately $Al_2SiO_5 \cdot n\ H_2O$

AMORPHOUS ● ● ●

Properties: C – white, gray, bluish, greenish, brown; S -white; L – vitreous, greasy and earthy; D – transparent, translucent to opaque; DE – 1.9; H – 2-3; CL – none; F – conchoidal to uneven; M – botryoidal, stalactitic and earthy aggregates.
Origin and occurrence: Hydrothermal product of alteration along the cracks of sedimentary rocks in coal deposits and in the oxidation zone of ore deposits, associated with other secondary minerals. Rich botryoidal aggregates occur in Dehr, Germany; Moldova Nuova, Romania; New Cornelia mine, Arizona, USA; El Dragon mine, Potosi, Bolivia.

Allophane, 90 mm, Zlate Hory, Czech Republic

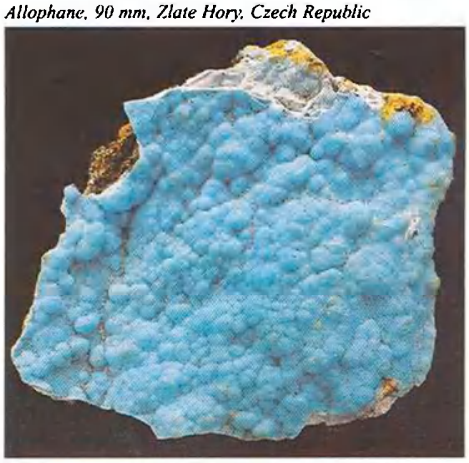

Palygorskite, 120 mm, Burda, Hungary

Palygorskite

$(Mg,Al)_2Si_4O_{10}(OH) \cdot 4\ H_2O$

MONOCLINIC ● ● ● ●

Properties: C – white, gray; S – white; L – dull; D - translucent to almost opaque; DE – 2.2; H – 1; CL – good; F – uneven, aggregates plastic under wet conditions; M – acicular crystals, fibrous aggregates, massive.
Origin and occurrence: Hydrothermal as a product of alteration of different rock types, such as serpentinites, granites, marbles and graywackes, also in ore veins and the Alpine-type veins. Rich aggregates resembling leather occur along the cracks in marbles in Hejná near Horazd'ovice, Czech Republic; in the Mammoth mine, Tiger, Arizona, USA and in Palygorskava, Russia.

Sepiolite, 70 mm, Eskisehir, Turkey

Sepiolite

$Mg_4Si_6O_{15}(OH)_2 \cdot 6\ H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – white, yellowish, gray; S – white; L – dull; D – opaque; DE – 2.0; H – 2-2.5; CL – not determined; F – uneven; M – earthy aggregates, massive.
Origin and occurrence: Hydrothermal as a product of serpentinite alteration, typically associated with magnesite. Classic locality is Eskişehir, Turkey; also known from Biskoupky, Czech Republic.

Zeophylite

$Ca_4Si_3O_8(OH,F)_4 \cdot 2\ H_2O$

TRICLINIC ● ●

Properties: C – white; S – white; L – pearly; D – transparent to translucent; DE – 2.6; H – 3; CL – perfect; F – uneven; M – platy crystals, commonly with radial structure.
Origin and occurrence: Hydrothermal in cavities of volcanic rocks, associated with zeolites. Its most important locality is Radejeín, Czech Republic, where it forms spherical aggregates up to 10 mm (3/8 in) in diameter. Also known from Schellkopf, Germany and Monte Somma, Italy.

Cavansite

$CaVOSi_4O_{10} \cdot 4\ H_2O$

ORTHORHOMBIC ● ●

Properties: C – green-blue, blue; S – light blue; L – vitreous; D – transparent; DE – 2.2; H – 3-4; CL – good; F – uneven; M – long prismatic to acicular crystals, radial aggregates.
Origin and occurrence: Hydrothermal in cavities of volcanic rocks, associated with calcite, apophyllite and zeolites. Rich radial aggregates up to 30 mm

Zeophyllite, 60 mm, Radejcin, Czech Republic

Cavansite, 120 mm, Poona, India

(1 1/16 in) in diameter come from the Wagholi Quarry, Poona, India.

Nepheline
$(K,Na)AlSiO_4$

HEXAGONAL ● ● ● ●

Properties: C – colorless, white, greenish, yellowish, gray, green, brown; S – white; L – vitreous to greasy; D – transparent to translucent; DE – 2.7; H – 5.5-6; CL – none; F – uneven to conchoidal; M – prismatic crystals, granular, massive.
Origin and occurrence: Magmatic in many alkaline rocks, as alkaline syenites and their pegmatites, also in some basalts; rare metamorphic in gneisses. It is a typical rock-forming mineral, associated with leucite, augite and apatite. Well-formed crystals found in Mount Vesuvius, Italy. Rich aggregates are known from many localities in Khibiny massif, Kola Peninsula, Russia. Crystals up to 70 cm (27 9/16 in) long occur in Davis Hill, Bancroft, Ontario. Perfect crystals up to 35 mm (1 3/8 in) long, come from Mont St.-Hilaire, Quebec, Canada.
Application: locally as Al ore and in ceramic industry.

Nepheline, 43 mm, Aouli, Morocco

Petalite, 23 mm, Paprok, Afghanistan

Petalite

$LiAlSi_4O_{10}$

MONOCLINIC ● ● ●

Properties: C – colorless, white, greenish, yellowish; S – white; L – vitreous; D – transparent to translucent; DE – 2.5; H – 6-6.5; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular.
Origin and occurrence: Magmatic in Li-bearing pegmatites, associated with lepidolite, elbaite, amblygonite and locally replaced by a mixture of quartz and spodumene. Gigantic crystals several meters long, come from Bikita, Zimbabwe; and Varutrask, Sweden. Clear gemmy crystals up to 230 mm (9 1/16 in) across found in pegmatite cavities in Paprok, Afghanistanand also in San Piero in Campo, Elba, Italy. Crystals up to 100 mm (4 in) across occur in Araçuai, Minas Gerais, Brazil.
Application: ceramic industry.

Chkalovite, 70 mm, Umbozero, Kola, Russia

Chkalovite

$Na_2BeSi_2O_6$

ORTHORHOMBIC ● ●

Properties: C – colorless, white; S – white; L – vitreous; D – transparent to translucent; DE – 2.7; H – 6; CL – imperfect; F – conchoidal; M – prismatic crystals, granular, massive.
Origin and occurrence: Hydrothermal in alkaline pegmatites, usually associated with natrolite, sodalite and eudialyte. Crystals up to 200 mm (7 1/8 in) across come from the Umbozero mine, Mount Alluaiv, Lovozero massif, Kola Peninsula, Russia; also at Julianhab, Greenland and Mont St.-Hilaire, Quebec, Canada.

Sanidine, 30 mm x, Drachenfels, Germany

Sanidine

FELDSPAR GROUP

$KAlSi_3O_8$

MONOCLINIC ● ● ● ●

Properties: C – colorless, white, yellowish, pinkish, gray; S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 6-6.5; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular, massive.

Origin and occurrence: Magmatic in acid volcanic rocks, as rhyolites and trachytes, a typical rock-forming mineral. Well-formed crystals and their combinations up to 100 mm (4 in) in size occur in rhyolites and trachytes in Drachenfels and Laacher Sec, Germany. It also comes from Roc de Courlande, France; Beaverdell, British Columbia, Canada; Kyustendil, Bulgaria.

Orthoclase

FELDSPAR GROUP

$KAlSi_3O_8$

MONOCLINIC ● ● ● ● ●

Varieties: adularia, moonstone (gemmy variety with chatoyancy)

Properties: C – colorless, white, yellowish, pinkish, gray, brown, yellow; S – white; L – vitreous to pearly (adularia); D – transparent to translucent; DE – 2.5; H – 6-6.5; CL – good; F – uneven to conchoidal; M – prismatic and tabular crystals and their combinations, granular, massive.

Origin and occurrence: Magmatic in rhyolites, trachytes, granites, syenites and pegmatites; metamorphic in various rock types, as orthogneisses and migmatites; hydrothermal in the Alpine-type veins, ore veins and some sediments, also in placers, a typical rock-forming mineral. Well-formed crystals and twins up to 200 mm (7⅞ in) across come from granites in Twentynine Palms, California, USA; Marina di Campo, Elba, Italy; Loket and Karlovy Vary, Czech Republic; in pegmatite cavities in Strzegom, Poland; San Piero in Campo, Elba, Italy. Yellow gemmy crystals up to 70 mm (2¾ in) across found in Itrongay, Madagascar. Gigantic feldspar crystals up to tens of meters long known from several pegmatite localities in Black Hills, South Dakota, USA; also Hagendorf, Germany. Adularia is known from Alpine-type veins in St. Gotthard, Switzerland; pebbles of moonstone occur in gem-bearing gravels in Ratnapura, Sri Lanka.

Application: ceramic and glass industry, moonstone as a gemstone.

Orthoclase, 32 mm x, Oro Grande, U.S.A.

Adularia, 60 mm, Alps, Switzerland

Microcline, 110 mm, Lake George, Teller Co., U.S.A.

Microcline

FELDSPAR GROUP

$KAlSi_3O_8$

TRICLINIC ● ● ● ● ●

Varieties: amazonite

Properties: C – colorless, white, yellowish, pinkish, gray, light to dark green (amazonite); S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 6-6.5; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular, massive.

Origin and occurrence: Magmatic in granites, syenites and pegmatites; metamorphic in various rock types, as orthogneisses and migmatites; hydrothermal in the Alpine-type veins and ore veins; a typical rock-forming mineral. Well-formed amazonite crystals up to 40 cm (15¾ in) across occur in pegmatite cavities in Crystal Peak, Colorado, USA; Keivy, Kola Peninsula, Russia; Morefield mine, Virginia, USA. Gigantic microcline crystals up to 12 m (39 ft) across come from pegmatites in Black Hills, South Dakota, USA; Kaatiala, Finland.

Application: ceramic industry, amazonite as a decorative stone.

Hyalophane

FELDSPAR GROUP

$(K,Ba)Al(Si,Al)_3O_8$

MONOCLINIC ● ● ●

Properties: C – colorless, white, yellowish; S – white; L – vitreous; D – transparent to translucent; DE – 2.9; H – 6-6.5; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular, massive.

Amazonite, 60 mm, Keivy, Kola, Russia

Hyalophane, 50 mm, Busovaca, Bosnia

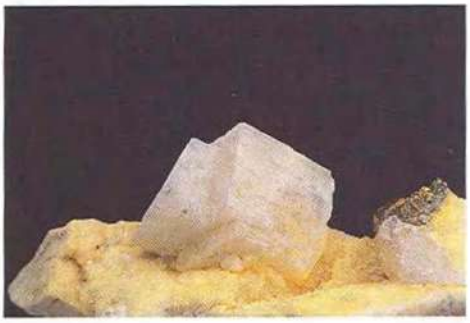

Origin and occurrence: Magmatic in phonolites; metamorphic in various rock types, as marbles, gneisses and Mn-rich rocks; hydrothermal in the Alpine-type and ore veins. Crystals up to 100 mm (4 in) across found in Alpine-type veins near Busovaca, Bosnia-Hercegovina.

Albite

FELDSPAR GROUP

$NaAlSi_3O_8$

TRICLINIC ● ● ● ● ●

Varieties: pericline, cleavelandite

Properties: C – colorless, white, yellowish, pinkish, gray, greenish, bluish; S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 6-6.5; CL – good; F – uneven to conchoidal; M – tabular crystals and their twins, platy aggregates, granular, massive.

Albite, 85 mm, Governador Valadares, Brazil

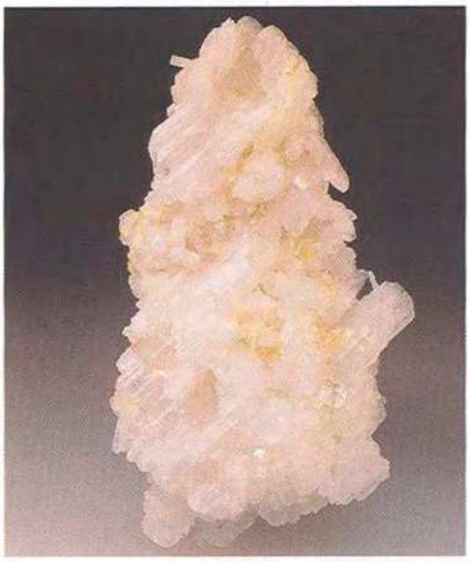

Origin and occurrence: Magmatic in granites, syenites and their pegmatites; metamorphic in various rock types, as orthogneisses, migmatites, phyllites and metamorphosed shales; hydrothermal in the Alpine-type veins and ore veins, a typical rock-forming mineral. Well-formed tabular to platy cleavelandite crystals up to 150 mm (6 in) across occur in pegmatite cavities in the Amelia district, Virginia and the Pala district, California, USA. Albite crystals also come from Strzegom, Poland; San Piero in Campo, Elba, Italy; Murzinka, Ural mountains, Russia; many localities in Minas Gerais, Brazil. Pericline crystals are known from the cracks along Alpine-type veins in Grossgreiner, Austria; St. Gotthard, Switzerland; also from ore veins in Roznava, Slovakia.

Oligoclase, 40 mm, Ural Mts., Russia

Oligoclase

FELDSPAR GROUP

$(Na_{0,9-0,7}Ca_{0,1-0,3})Al_{1,1-1,3}Si_{2,9-2,7}O_8$

TRICLINIC ● ● ● ● ●

Varieties: peristerite, sunstone

Properties: C – white, yellowish, pinkish, greenish, iridescent (peristerite); S – white; L – vitreous; D – transparent to translucent; DE – 2.6; H – 6-6.5; CL – good; F – uneven to conchoidal; M – tabular crystals, platy aggregates, granular, massive.

Origin and occurrence: Magmatic in granites, syenites, andesites and pegmatites; metamorphic in gneisses and migmatites; hydrothermal in the Alpine-type veins; typical rock-forming mineral. Large crystals, several dm long, are known mainly from pegmatites. Peristerite comes from Arendal, Norway; Miass, Ural mountains, Russia; Quadeville, Ontario, Canada; Tvedestrand, Norway. It also occurs in Vezna , Czech Republic. Sunstone is mined in the Ponderosa mine near Lakeview, Oregon, USA.

Application: peristerite and sunstone are used as gemstones.

Andesine, 50 mm, Bodenmais, Germany

Andesine

FELDSPAR GROUP

$(Na_{0,7-0,5}Ca_{0,3-0,5}Al_{1,3-1,5}Si_{2,7-2,5}O_8$

TRICLINIC ● ● ● ●

Properties: C – yellowish, pinkish, gray, light green; S – white; L – vitreous; D – transparent to translucent; DE – 2.7; H – 6-6.5; CL – good; F – uneven to conchoidal; M – platy aggregates, granular, massive.

Labradorite, 50 mm, Labrador, Canada

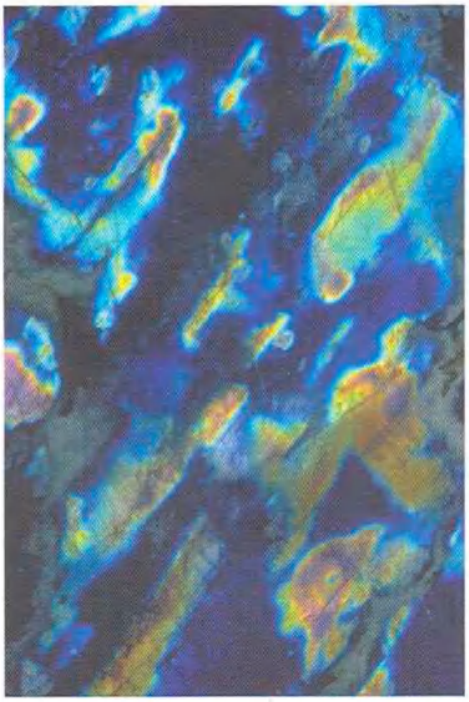

Anorthite, 10 mm xx, Pasmaada, Finnland

Origin and occurrence: C – white to dark gray, greenish; S – white; L – vitreous, pearly; D – transparent to translucent; DE – 2.7; H – 6-6.5; CL – good; F – uneven to conchoidal; M B cleavable aggregates, granular, massive.

Labradorite

FELDSPAR GROUP

$Ca_{0,7-0,5}Na_{0,3-0,5}Al_{1,5-1,7}Si_{2,5-2,3}O_8$

TRICLINIC ● ● ● ●

Properties: Magmatic in andesites, diorites; metamorphic in gneisses and migmatites; typical rock-forming mineral. Large grains are known from migmatites in Bodenmais, Germany and Adamello, Italy.
Origin and occurrence: Magmatic in gabbros, basalts, anorthosites; metamorphic in amphibolites. Large iridescent aggregates up to 1 m (39⅜ in) across, come from Nain, Labrador, Quebec, Canada and also occur in Korostenskiy massif, Ukraine and Ylamaa, Finland.
Application: as a decorative stone.

Anortite

FELDSPAR GROUP

$CaAl_2Si_2O_8$

TRICLINIC ● ● ●

Properties: C – gray, greenish, pinkish; S – white; L – vitreous, dull; D – transparent to translucent; DE – 2.8; H – 6-6.5; CL – good; F – uneven to conchoidal; M B granular, massive.
Origin and occurrence: Magmatic in gabbros, basalts, anorthosites; metamorphic in contact metamorphosed rocks. Pinkish grains and poorly-developed crystals come from Val di Fassa, Italy. Crystals up to 50 mm (2 in) across are known from Miyake-Jima Island, Japan. It was also found in Mount Erebus, Antarctica and Monte Somma, Italy.

Danburite
$CaB_2Si_2O_8$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, gray, greenish, pinkish, yellow, brown, red-brown; S – white; L – vitreous, dull; D – transparent to translucent; DE – 3.0; H – 7; CL – imperfect; F – uneven to conchoidal; M – prismatic crystals, columnar aggregates, granular, massive.
Origin and occurrence: Hydrothermal in pegmatite cavities, ore veins and the Alpine-type veins; metamorphic in skarns and contact metamorphosed rocks.
Well-formed prismatic crystals up to 250 mm ($9^{13}/_{16}$ in) long come from Russell, New York, USA; Toroku, Japan; Dalnegorsk, Russia; and Charcas, San Luis Potosi, Mexico.

Danburite, 109 mm, Charcas, Mexico

Cancrinite
$Na_6CaAl_6Si_6O_{24}(CO_3) \cdot 2\,H_2O$

HEXAGONAL ● ● ●

Properties: C – colorless, white, yellow, orange, bluish; S – white; L – vitreous; D – transparent to translucent; DE – 2.4; H – 5-6; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular, massive.
Origin and occurrence: Magmatic in alkaline syenites; hydrothermal as a product of replacement of volcanic rocks, associated with nepheline and sodalite.
Crystals up to 20 mm ($^{25}/_{32}$ in) across come from Mont St.-Hilaire, Quebec, Canada. It was also found in Litchfield, Maine, USA; Cancrinite Hill, Bancroft, Ontario, Canada and elsewhere.

Leifite
$Na_2(Si,Al,Be)_7\,(O,OH,F)_{14}$

TRIGONAL ●

Properties: C – colorless; S – white; L – vitreous; D – translucent; DE – 2.5; H – 6; CL – good; F – uneven; M – acicular crystals.
Origin and occurrence: Hydrothermal in cavities of alkaline pegmatites, associated with microcline and zinnwaldite. Crystals up to 20 mm ($^{25}/_{32}$ in) across occur in Narssarssuk, Greenland. Fine crystals are also known from Mont St.-Hilaire, Quebec, Canada, associated with serandite.

Cancrinite, 70 mm, Ditrau, Romania

Leifite, 100 mm, Mont St.-Hilaire, Canada

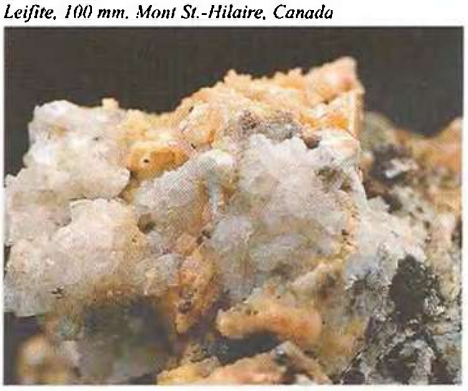

Sodalite, 50 mm, Bancroft, Canada

Sodalite

$Na_8Al_6Si_6O_{24}Cl_2$

CUBIC ● ● ●

Varieties: hackmanite

Properties: C – colorless, white, blue, yellow, pink (hackmanite); S – white; L – vitreous; D – transparent to translucent; DE – 2.3; H – 5.5-6; CL – imperfect; F – uneven to conchoidal; M – isometric crystals, granular, massive; LU – orange-red.
Origin and occurrence: Magmatic in alkaline syenites and phonolites, associated with nepheline, zircon and titanite; hydrothermal in marbles. Granular aggregates are known from the Princess Sodalite mine, Bancroft, Ontario, Canada and Ditrau, Romania. Crystals up to 100 mm (4 in) across come from Kangerdluarssuk, Greenland.
Application: cut as a gemstone.

Noseane

$Na_8Al_6Si_6O_{24}(SO_4)$

CUBIC ● ●

Properties: C – colorless, white, blue, gray, black; S – white; L – vitreous; D – transparent to translucent; DE – 2.3; H – 5-6; CL – good; F – uneven to conchoidal; M – isometric crystals, granular, massive.
Origin and occurrence: Magmatic in alkaline basalts and similar effusive rocks, associated with nepheline and haüyne. Granular aggregates are known from Laacher See, Germany and Monte Somma, Italy.

Lazurite

$(Na,Ca)_8Al_6Si_6O_{24}(S,SO_4)$

CUBIC ● ● ●

Properties: C – dark blue; S -light blue; L – vitreous; D – transparent to translucent; DE – 2.4; H – 5-5.5; CL – imperfect; F – uneven to conchoidal; M – isometric crystals, granular, massive.
Origin and occurrence: Metamorphic in contact metamorphosed marbles, associated with pyrite. Crystals up to 50 mm (2 in) across come from Sar-e-Sang, Badakhshan province, Afghanistan. Granular aggregates are known from many localities, such as Malobystrinskoye deposit near Lake Baikal, Russia or Monte Somma, Italy.
Application: decorative stone and gemstone.

Nosean, 6 mm grains, Laacher See, Germany

Lazurite, 60 mm, Afghanistan

Lazurite, 21 mm x, Nuristan, Afghanistan

Tugtupite

$Na_8Be_2Al_2Si_8O_{24}Cl_2$

TETRAGONAL ● ●

Properties: C – white, pink, red, bluish; S – white; L – vitreous; D – transparent to translucent; DE – 2.3; H – 5; CL – good; F – uneven to conchoidal; M – small crystals, granular, massive.

Origin and occurrence: Hydrothermal in alkaline syenites and their pegmatites, also as a product of chkalovite alteration. Aggregates up to 60 mm (2¼ in) across come from the Umbozero mine, Mount Alluaiv, Lovozero massif, Kola Peninsula, Russia

Tugtupite, 90 mm, Illimaussaq, Greenland

Lazurite, 60 mm, Chile

and IIRmaussaq, Greenland. Small crystals occur in the Poudrette quarry, Quebec, Canada.

Danalite

$Fe_4Be_3Si_3O_{12}S$

CUBIC ● ●

Properties: C – gray,yellow,pink, red, brown; S – white; L – vitreous; D – transparent to translucent; DE – 3.4; H – 5.5-6; CL – imperfect; F – uneven to conchoidal; M – isometric crystals, granular, massive.

Origin and occurrence: Magmatic in granitic pegmatites; hydrothermal in greisens, skarns and ore veins. It occurs in Cape Ann, Massachusetts, USA; Hortekollen, Norway and Coolgardie, Western Australia, Australia.

Danalite, 70 mm, Yxjoberg, Sweden

Helvite, 2 mm xx, Cavnic, Romania

Helvite

$Mn_4Be_3Si_3O_{12}S$

CUBIC ● ● ●

Properties: C – brown, gray, yellow, yellow-green; S – white; L – vitreous; D – transparent to translucent; DE – 3.4; H – 6; CL – good; F – uneven to conchoidal; M – isometric crystals, granular, massive.
Origin and occurrence: Magmatic in granitic pegmatites and alkaline syenites; hydrothermal in greisens, skarns and ore deposits. Cubic crystals up to 25 mm (1 in) across known from the Sawtooth Batholith, Idaho, USA. Crystals also come from Schwarzenberg, Germany; Cavnic, Romania; and Oslofjord, Norway.
Application: Be ore.

Scapolite, 44 mm x, Leslie Lake, Quebec, Canada

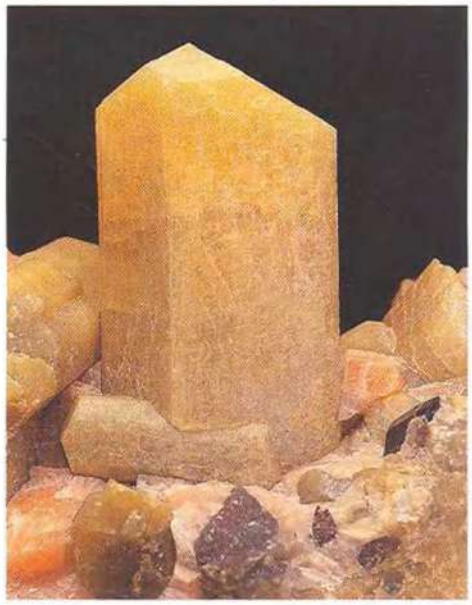

Marialite

SCAPOLITE GROUP

$Na_8(AlSi_3O_8)_6(Cl_2,SO_4)$

TETRAGONAL ● ● ●

Properties: : C – colorless, white, gray, purplish, yellow; S – white; L – vitreous, locally pearly; D – transparent to translucent; DE – 2.5; H – 5-6; CL – good; F – uneven to conchoidal; M – prismatic crystals, columnar aggregates, granular, massive; LU – yellow to orange.
Origin and occurrence: Hydrothermal in veins, cross-cutting alkaline metamorphic rocks and in pegmatites, cross-cutting ultrabasic rocks; metamorphic in marbles and metaevaporites. It comes from Ankazobe, Madagascar; Umba, Tanzania and Crestmore, California, USA.

Meionite

SCAPOLITE GROUP

$Ca_8(Al_2Si_2O_8)_6(CO_3,SO_4)$

TETRAGONAL ● ● ● ●

Properties: C – colorless, white, gray, purplish, green, blue; S – white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 5-6; CL – good; F – uneven to conchoidal; M – long prismatic crystals, columnar aggregates, granular, massive; LU – yellow to orange.

Leucite, 30 mm, Mt.Vesuvius, Italy

Analcime, 45 mm, Tunguzka, Russia

Origin and occurrence: Metamorphic in skarns, marbles, granulites and in contacts of volcanic rocks; hydrothermal in veins, cross-cutting Ca-rich rocks. Well formed crystals up to 40 cm (15¾ in) across come from Lake Clear and Eganville, Ontario, Canada. Crystals are also known from Monte Somma, Italy; Sludyanka, Siberia, Russia and Pargas, Finland.

Leucite

ZEOLITE GROUP

$KAlSi_2O_6$

TETRAGONAL ● ● ● ●

Properties: C – colorless, white, gray; S – white; L – vitreous; D – transparent to translucent; DE – 2.5; H – 5.5-6; CL- imperfect; F – uneven to conchoidal; M – isometric crystals, granular, massive.
Origin and occurrence: Magmatic in effusive K-rich basalts, associated with nepheline and sanidine. Well-formed crystals, several cm in size, come from Mount Vesuvius, Italy and Laacher See, Germany.
Application: ceramic industry.

Analcime

ZEOLITE GROUP

$NaAlSi_2O_6 \cdot H_2O$

CUBIC ● ● ● ●

Properties: C – colorless, white, pinkish, yellowish; S – white; L – vitreous; D – transparent to translucent; DE – 2.3; H – 5-5.5; CL – imperfect; F – uneven to conchoidal; M – isometric crystals, granular, massive.
Origin and occurrence: Mainly hydrothermal as a product of nepheline or sodalite replacement; rare magmatic in effusive rocks; also in sediments, associated with calcite and zeolites. Well-formed crystals up to 30 cm (12 in) across come from Nidym, Siberia, Russia; Lago Maggiore, Italy; Mont St.-Hilaire, Quebec, Canada; West Paterson, New Jersey, USA and elsewhere.

Pollucite

ZEOLITE GROUP

$(Cs,Na)AlSi_2O_6 \cdot H_2O$

CUBIC ● ● ●

Properties: C – colorless, white, gray; S – white; L – vitreous; D – transparent to translucent; DE – 2.9; H – 6.5-7; CL – none; F – uneven to conchoidal; M – isometric crystals, granular, massive.
Origin and occurrence: Magmatic and rarely also hydrothermal in Li-bearing pegmatites, associated with lepidolite, albite, quartz and petalite. Almost monomineral layer of pollucite several meters thick occurs in the Tanco mine, Bernic Lake, Manitoba, Canada. White and colorless crystals from pegmatite cavities up to 60 cm (24 in) across come from Paprok, Afghanistan. Also known from San Piero in Campo, Elba, Italy and Gilgit, Pakistan.
Application: ceramic industry, Cs ore.

Pollucite, 22 mm x, Gilgit, Pakistan

Natrolite, 47 mm, Maricopa Co., Arizona, U.S.A.

Scolecite, 165 mm, Nasik, India

Natrolite

ZEOLITE GROUP

$Na_2Al_2Si_3O_{10} \cdot 2\,H_2O$

ORTHORHOMBIC ● ● ● ●

Properties: C – colorless, white, yellowish, pinkish; S – white; L – vitreous to silky; D – transparent to translucent; DE – 2.2; H – 5-5.5; CL – perfect; F – uneven to conchoidal; M – long prismatic to acicular crystals, fibrous and radial aggregates, granular, massive.

Origin and occurrence: Hydrothermal in cavities in volcanic rocks, alkaline pegmatites, along the Alpine-type fissures, also as a product of plagioclase replacement, commonly associated with calcite and zeolites. Colorless and white acicular crystals up to 30 cm (12 in) long come from cavities of alkaline pegmatites in Mount Putelichorr, Khibiny massif, Kola Peninsula, Russia. Crystals were also found in Narssarssuk, Greensland and Mont St.-Hilaire, Quebec, Canada. Crystals also occur in cavities of basaltic rocks in Teigarhorn, Iceland; Zálezly and Soutisky, Czech Republic and in Faeroe Islands.

Skolecite

ZEOLITE GROUP

$CaAl_2Si_3O_{10} \cdot 3\ H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white; S – white; L – vitreous to silky; D – transparent to translucent; DE – 2.3; H – 5; CL – good; F – uneven; M – long prismatic to acicular crystals, fibrous and radial aggregates.

Origin and occurrence: Hydrothermal in cavities in volcanic rocks and along the Alpine-type fissures, associated with calcite and zeolites. Clear prismatic crystals up to 200 mm (7⅞ in) long come from basalt cavities near Nasik, India. Crystals come also from Teigarhorn, Iceland and Suderoy, Faeroe Islands.

Mesolite, 150 mm, Mahrashitra, India

Mesolite

ZEOLITE GROUP

$Na_{16}Ca_{16}Al_{48}Si_{72}O_{240} \cdot 64\ H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white; S – white; L – vitreous to silky; D – transparent to translucent; DE – 2.3; H – 5; CL – good; F – uneven; M – acicular crystals, fibrous and radial aggregates, granular, massive.

Origin and occurrence: Hydrothermal in cavities in volcanic rocks, associated with other zeolites. Colorless needles up to 150 mm (6 in) long come from the basalt cavities in Berufjord and Teigarhorn, Iceland. Radial aggregates up to 200 mm (7⅞ in) in diameter known from the vicinity of Poona, India. Prismatic crystals up to 100 mm (4 in) long found in Skookumchuck Dam, Washington, USA.

Thomsonite

ZEOLITE GROUP

$Ca_2NaAl_4(Al,Si)_2Si_4O_{20} \cdot 6\ H_2O$

ORTHORHOMBIC ● ● ●

Properties: C – colorless, white, yellowish, brown-red; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 2.3; H – 5-5.5; CL – good; F – uneven to conchoidal; M – prismatic crystals in clusters with radial structure, botryoidal aggregates, massive.

Origin and occurrence: Hydrothermal in cavities in volcanic rocks, also as a product of hydrothermal feldspar replacement. Colorless and white acicular crystals, forming radial aggregates up to 50 mm (2 in) in diameter come from Old Kilpatrick, Scotland, UK and West Paterson, New Jersey, USA. Hemispherical aggregates up to 30 mm (1³⁄₁₆ in) across known from Vinaicka hora near Kladno, Czech Republic.

Thomsonite, 50 mm, Bazsi, Hungary

Gonnardite, 10 mm xx, Zalahalap, Hungary

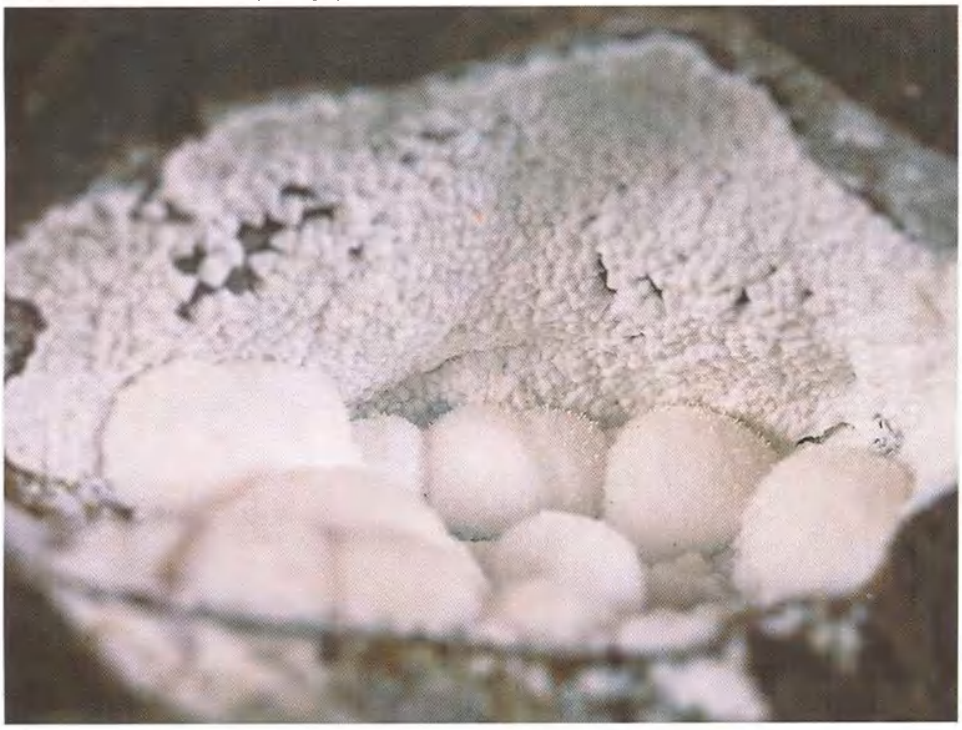

Gonnardite

ZEOLITE GROUP

$(Na,Ca)_{6-8}(Al,Si)_{20}O_{40} \cdot 12\ H_2O$

TETRAGONAL ● ● ●

Properties: C – colorless, white, yellowish; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 2.3; H – 4.5-5; CL – good; F – uneven to conchoidal; M – prismatic crystals, fibrous and radial aggregates.

Origin and occurrence: Hydrothermal in cavities in volcanic rocks, associated with zeolites, also in the contact zone of marbles, associated with wollastonite. Acicular crystals and fibrous aggregates occur in Aci Castello, Sicily, Italy; Weilberg, Germany; Bundoora, Victoria, Australia.

Edingtonite

ZEOLITE GROUP

$BaAl_2Si_3O_{10} \cdot 3\ H_2O$

TETRAGONAL ● ● ●

Properties: C – white, gray, pinkish; S – white; L – vitreous; D – transparent to translucent; DE – 2.8; H – 4; CL – good; F – uneven to conchoidal; M – prismatic crystals, granular, massive.

Origin and occurrence: Hydrothermal in cavities and along the cracks in volcanic rocks, associated with calcite and harmotome. Colorless and white prismatic crystals up to 40 mm (1 9/16 in) across come from Old Kilpatrick, Scotland, UK; Stare Ransko,

Edingtonite, 100 mm, Stare Ransko, Czech Republic

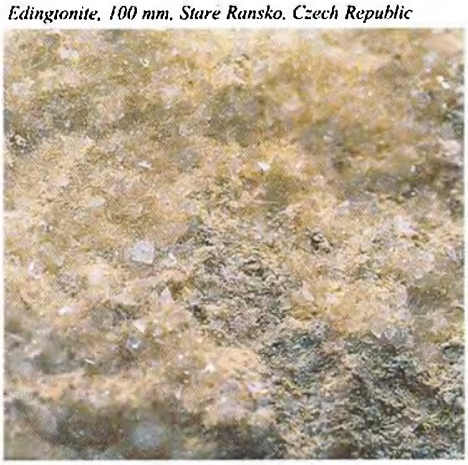

Czech Republic and Mont St.-Hilaire, Quebec, Canada.

Dachiardite-Ca

ZEOLITE GROUP

$(Ca_{0.5},Na,K)_5Al_5Si_{19}O_{48} \cdot 13\ H_2O$

MONOCLINIC ● ●

Properties: C – colorless, white, yellowish; S – white; L – vitreous to pearly; D – transparent to translucent; DE – 2.1; H – 4-4.5; CL – good; F – uneven to conchoidal; M – prismatic crystals and complex interpenetration twins.
Origin and occurrence: Hydrothermal in cavities in granitic pegmatites, associated with albite, petalite and elbaite, also as pseudo-morphs after petalite. It occurs in San Piero in Campo, Elba, Italy and in the Opal Hill quarry, Riverside, California, USA.

Ferrierite-Ca

ZEOLITE GROUP

$(Ca,Na,K,Mg)_6Al_6Si_{30}O_{72} \cdot 18\ H_2O$

ORTHORHOMBIC ● ●

Properties: C – colorless, white, greenish, pink, brownish; S – white; L – vitreous; D – transparent to translucent; DE – 2.1; H – 3-3.5; CL – good; F – uneven to conchoidal; M – tabular crystals, platy and columnar aggregates, massive.
Origin and occurrence: Hydrothermal in cavities in volcanic rocks, along the cracks of the Alpine-type veins and in volcanic tuffs, associated with calcite and zeolites. Platy aggregates occur in Svojanov, Czech Republic; Albero Bosso and Monastir, Sardinia, Italy; Kamploops Lake, British Columbia, Canada.

Ferrierite-Ca, 2 mm aggregates, Regec, Hungary

Dachiardite-Ca, 2 mm xx, Regec, Hungary

Laumontite

ZEOLITE GROUP

$Ca_4Al_8Si_{16}O_{48} \cdot 18\ H_2O$

MONOCLINIC ● ● ● ●

Properties: C – colorless, white, yellowish, pinkish; S – white; L – vitreous; D – transparent to translucent, weathered almost opaque; DE – 2.4; H – 3-4; CL – good; F – uneven to conchoidal; M – prismatic crystals, fibrous and radial aggregates.
Origin and occurrence: Hydrothermal in cavities in volcanic rocks, along the cracks in the Alpine-type veins, in ore veins and sediments, associated with calcite and other zeolites. Colorless and white acicular crystals up to 38 cm (14 15/16 in) long come from the Pandulena Hill Quarries, India. Crystals up to 150 mm (6 in) long known from Pine Creek, Bishop, California, USA. It was also found in Dillenburg, Germany.

Laumontite, 120 mm, Markovice, Czech Republic

Heulandite-Ca, 8 mm xx, Parana, Brazil

Heulandite-Ca

ZEOLITE GROUP

$(Ca_{0.5},Sr_{0.5},Na,K)_{4.5}Al_9Si_{27}O_{72} \cdot 24\ H_2O$

MONOCLINIC ● ● ● ●

Properties: C – colorless, white, yellowish, pinkish, red, brown; S – white; L B vitreous to pearly; D – transparent to translucent; DE – 2.1; H – 3.5-4; CL – perfect; F – uneven to conchoidal; M – prismatic to tabular crystals, granular, massive.

Origin and occurrence: Hydrothermal in cavities in volcanic rocks, in the Alpine-type fissures, ore veins and sedimentary rocks, usually associated with calcite and other zeolites. Colorless and white tabular crystals up to 100 mm (4 in) across come from Nasik, India; West Paterson, New Jersey, USA; Teigarhorn, Iceland. Red heulandite found in Val di Fassa, Italy.

Clinoptilolite-Ca, 2 mm xx, Hon?ova Hurka, Czech Republic

Clinoptilolite-Ca

ZEOLITE GROUP

$(Ca_{0.5},Na,K)_{4.5}Al_9Si_{27}O_{72} \cdot 24\ H_2O$

MONOCLINIC ● ● ● ●

Properties: C – colorless, white, yellowish, pinkish, red, greenish; S – white; L – vitreous; D – transparent to translucent; DE – 2.2; H – 3.5-4; CL – perfect; F – uneven to conchoidal; M – platy crystals, massive.

Origin and occurrence: Hydrothermal in volcanic-sedimentary rocks. Large industrial deposits occur in New Zealand, Japan and Australia. Crystals come from Agate Beach, Oregon, USA.

Application: construction, chemical industry and agriculture.

Stilbite-Ca

ZEOLITE GROUP

$(Ca_{0.5},Na,K)_9Al_9Si_{27}O_{72} \cdot 28\ H_2O$

MONOCLINIC ● ● ● ●

Properties: C – colorless, white, yellowish, pinkish, brown; S – white; L – vitreous, locally pearly; D – transparent to translucent; DE – 2.2; H – 3.5-4; CL – good; F – uneven to conchoidal; M – prismatic and tabular crystals, commonly complicated interpenetration twins and sheaf-like aggregates, granular, massive.

Origin and occurrence: Hydrothermal in cavities of volcanic rocks, along Alpine-type fissures, in ore veins, also in sedimentary rocks and hot springs, typically associated with calcite and other zeolites. Colorless and white tabular crystals and their combinations up to 200 mm (7⅞ in) across come from the Pandulena Hill Quarries, India. Crystals are also known from Teigarhorn, Iceland; Faeroe Islands; West Paterson, New Jersey, USA and elsewhere.

Epistilbite, 15 mm x, Nasik, India

Epistilbite

ZEOLITE GROUP

$(Ca,Na_2)Al_2Si_4O_{12} \cdot 4\ H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white, yellowish, pinkish, light brown; S – white; L – vitreous; D – transparent to translucent; DE – 2.2; H – 4; CL – good; F – uneven to conchoidal; M – prismatic crystals, radial aggregates.

Origin and occurrence: Hydrothermal in cavities in volcanic rocks; associated with other zeolites. Clear and white tabular crystals and their twins up to 30 mm (1³/₁₆ in) across come from Jalgaon, India; also in Teigarhorn, Iceland and Faeroe Islands.

Stilbite-Ca, 83 mm, Nasik, India

Gismondine, 2 mm xx, Tapolca-Diszel, Hungary

Gismondine

ZEOLITE GROUP

$CaAl_2Si_2O_8 \cdot 4{,}5\ H_2O$

MONOCLINIC ● ●

Properties: C – colorless, white, bluish, pinkish; S – white; L – vitreous; D – transparent to translucent; DE – 2.3; H – 4.5; CL – good; F – uneven; M – complex twins of crystals, platy aggregates.
Origin and occurrence: Hydrothermal in cavities in volcanic rocks, in hydrothermally altered rocks. It comes from Capo di Bove, Italy; Schiffenberg, Germany and Dobrna, Czech Republic.

Harmotome, 79 mm, Strontian, UK

Phillipsite-K, 3 mm xx, Limberg, Germany

Phillipsite-K

ZEOLITE GROUP

$(K,Na,Ca_{0.5})Al_5Si_{11}O_{32} \cdot 12\ H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white, reddish; S – white; L – vitreous; D – transparent to translucent; DE – 2.2; H – 4-4.5; CL – good; F – uneven; M – prismatic crystals and complex interpenetration twins, granular, massive.
Origin and occurrence: Hydrothermal in cavities in volcanic rocks, sedimentary rocks and hot springs; associated with calcite and other zeolites. Colorless and white interpenetration twins up to 20 mm ($^{25}/_{32}$ in) across come from Capo di Bove, Italy; Doughboys, Tasmania, Australia; Sovinec near Litomioce, Czech Republic and elsewhere.

Harmotome

ZEOLITE GROUP

$(Ba_{0.5},Ca_{0.5},K,Na)_5Al_5Si_{11}O_{32} \cdot 12\ H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white, gray, reddish, yellow, brown; S – white; L – vitreous; D – transparent to translucent; DE – 2.4; H – 4.5; CL – good; F – uneven; M – prismatic crystals, commonly complex interpenetration twins, granular massive.
Origin and occurrence: Hydrothermal in cavities in volcanic rocks, in the Alpine-type veins, ore veins and in pegmatites. Prismatic crystals and their twins up to 20 mm ($^{25}/_{32}$ in) across are known from Strontian, Scotland, UK. Crystals also come from St. Andreasberg and Oberstein, Germany; Kozakov, Czech Republic and Kongsberg, Norway.

Goosecreekite

ZEOLITE GROUP

$CaAl_2Si_3O_{10} \cdot 5\ H_2O$

MONOCLINIC ●

Properties: C – white, colorless; S – white; L – vitreous; D – transparent to translucent; DE – 2.2; H – 4.5; CL B perfect; F – uneven to conchoidal; M – prismatic crystals, granular.
Origin and occurrence: Hydrothermal in cavities in volcanic rocks. It is known from Goose Creek, Virginia, USA. Crystals up to 30 mm (1 3/16 in) across found in the Pandulena Hill quarries, India.

Goosecreekite, 25 mm xx, Maharashtra, India

Chabasite-Ca

ZEOLITE GROUP

$(Ca_{0.5},Na,K)Al_4Si_8O_{24} \cdot 12\ H_2O$

TRIGONAL ● ● ●

Properties: C – colorless, white, yellowish, pinkish, greenish; S – white; L – vitreous to dull; D – transparent to translucent; DE – 2.2; H – 4-4.5; CL – imperfect; F – uneven to conchoidal; M – rhombohedral crystals and their twins, granular, massive.
Origin and occurrence: Hydrothermal in cavities in volcanic rocks and pegmatites, along Alpine-type fissures and in hot springs, usually associated with calcite and other zeolites. Colorless and white interpenetration twins up to 60 mm (24 in) across, come from Faeroe Islands. Crystals are also known from Repcice, Czech Republic; Berufjord, Iceland; Panvil, India; Maglovec, Slovakia and elsewhere.

Chabazite-Ca, Passboro, Nova Scotia, Canada

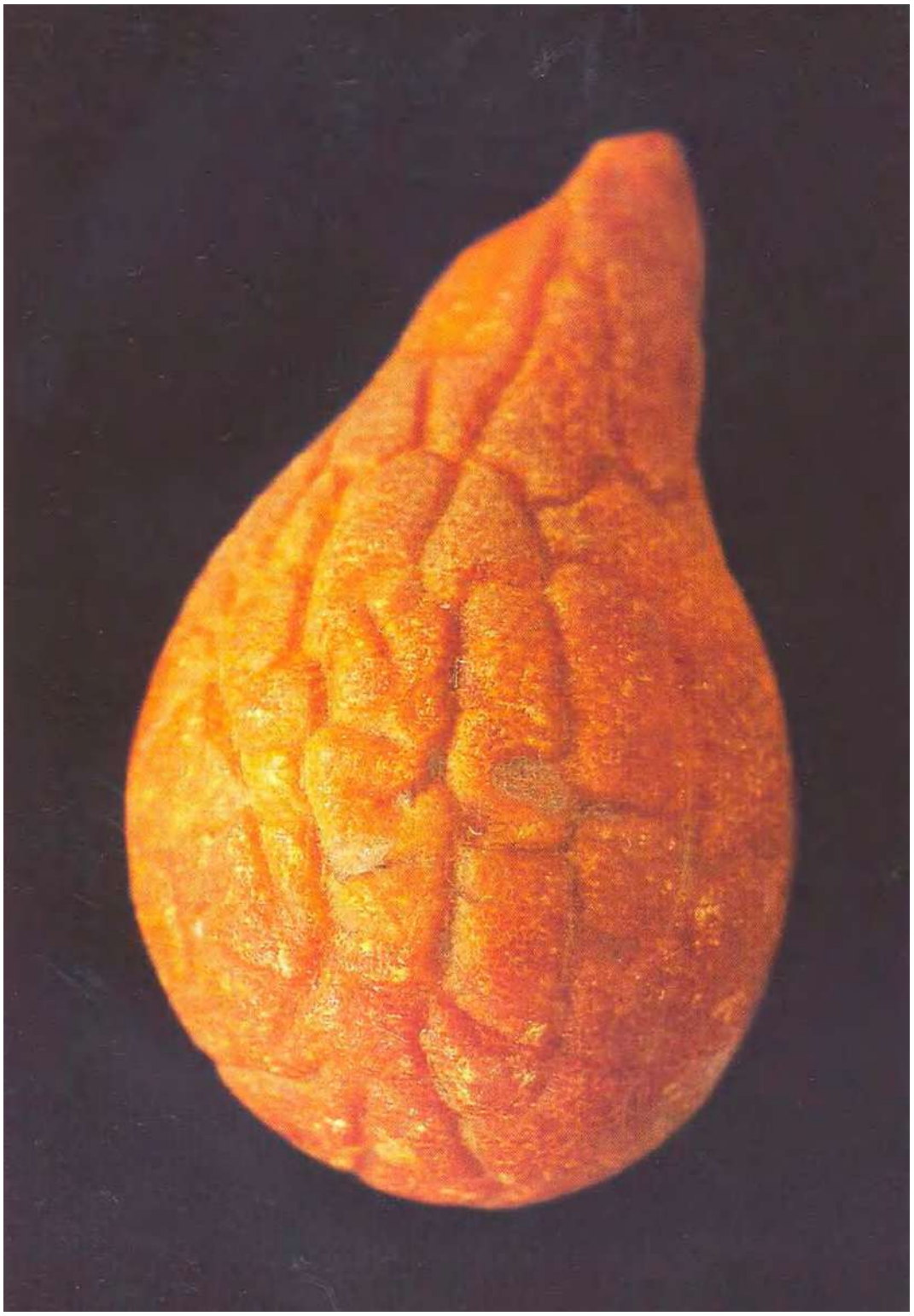

10. Organic compounds

Whewellite

$CaC_2O_4 \cdot H_2O$

MONOCLINIC ● ● ●

Properties: C – colorless, white, grayish; S – colorless; L – vitreous; D – transparent to translucent; DE – 2,2; H – 2,5; CL – good; F – conchoidal; M – prismatic crystals, commonly twinned.
Origin and occurrence: Rare hydrothermal, mainly sedimentary in coal basins, associated with barite, ankerite and other minerals. Hydrothermal crystals up to 70 mm (2¾ in) long, come from U-bearing veins in PrRbram, Czech Republic. Similar crystals, up to 70 mm (2¾ in) long, occurred in Cavnic, Romania. Heart-shaped and butterfly twins, up to 100 mm (4 in) found in concretions near Kladno, Czech Republic. Similar specimens are known from Burgk near Dresden, Germany. Interesting flat radial aggregates found along cracks in clays in the vicinity of Most, Czech Republic.

Amber, 40 mm, Baltic Sea, Latvia
Whewellite, 27 mm x, Burgk, Germany

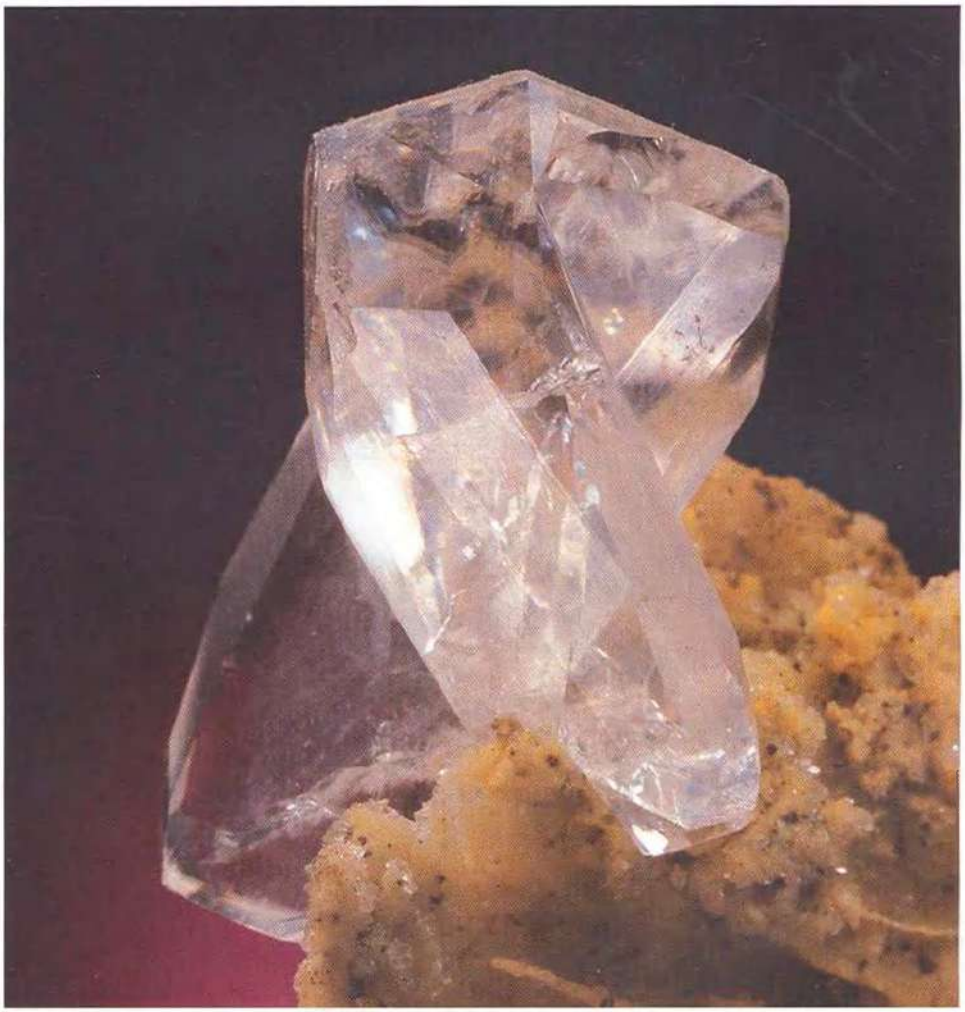

Mellite, 170 mm, Bicske-Csordakút, Hungary

Mellite

$Al_2[C_6(COO)_6] \cdot 18\ H_2O$

TETRAGONAL ● ●

Properties: C – honey-yellow; S – white; L – resinous to vitreous; D – transparent to translucent; DE – 1,7; H – 2; CL – imperfect; F – conchoidal; M – dipyramidal crystals, granular; LU – blue.
Origin and occurrence: Secondary in the cracks in brown coal and lignite. The best specimens with crystals up to 40 mm (1 5/16 in) in size, come from Csordakut near Tatabanya, Hungary. Crystals, up to 10 mm (3/8 in) across, were found in Artern, Germany. Granular aggregates are known from Valchov, Czech Republic.

Evenkite

$C_{24}H_{50}$

MONOCLINIC ●

Properties: C – colorless to light yellow; S – colorless; L – waxy; D – translucent; DE – 0,9; H – 1; CL – good; M – pseudo-hexagonal tabular crystals; R – melts at 50EC (122°F).
Origin and occurrence: Probably secondary. It occurs within geodes near Evenki, Siberia, Russia. Also known from the cracks in altered andesite in Dubnik, Slovakia.

Fichtelite

$C_{19}H_{34}$

MONOCLINIC ● ● ●

Properties: C – colorless to yellowish; S – colorless; L – vitreous; PS – transparent to translucent; DE – 1; H – 1; M – thin tabular crystals, scales, crystalline crusts.

Evenkite, 50 mm, Dubnik, Slovakia

Origin and occurrence: Secondary, typical mineral of peat-bogs. Its crystals are known from Borkovice near Sobeslav, Czech Republic; also known from Marktredwitz, Germany.

Amber

a mixture of hydrocarbons

AMORPHOUS ● ● ●

Properties: C – honey-yellow, yellow-brown, brown, red-brown, blue, green, black; S – white; L – resinous, dull; D – transparent to translucent, rare opaque; DE – 1,0-1,1; H – 2-2,5; CL – none; F – conchoidal; M – massive irregular or drop-shaped aggregates, nodules and fragments; LU – light blue, yellow.
Origin and occurrence: Amber is a petrified resin from Tertiary and Mesozoic conifers, occurring rarely in sediments. Plant or insect remnants are sometimes found trapped in amber. The most famous localities are located along the southern coast of the Baltic Sea in Poland, Germany, Lithuania, Estonia, Latvia and Russia. The largest masses found weighed up to 10 kg (22 lb). Blue amber is known from the Dominican Republic. Other amber localities are found in Syria, Lebanon, Thailand, Vietnam, Canada and the USA.
Application: as a gemstone.

Fichtelite, 80 mm, Mazice, Czech Republic

Amber, 95 mm, Madagascar

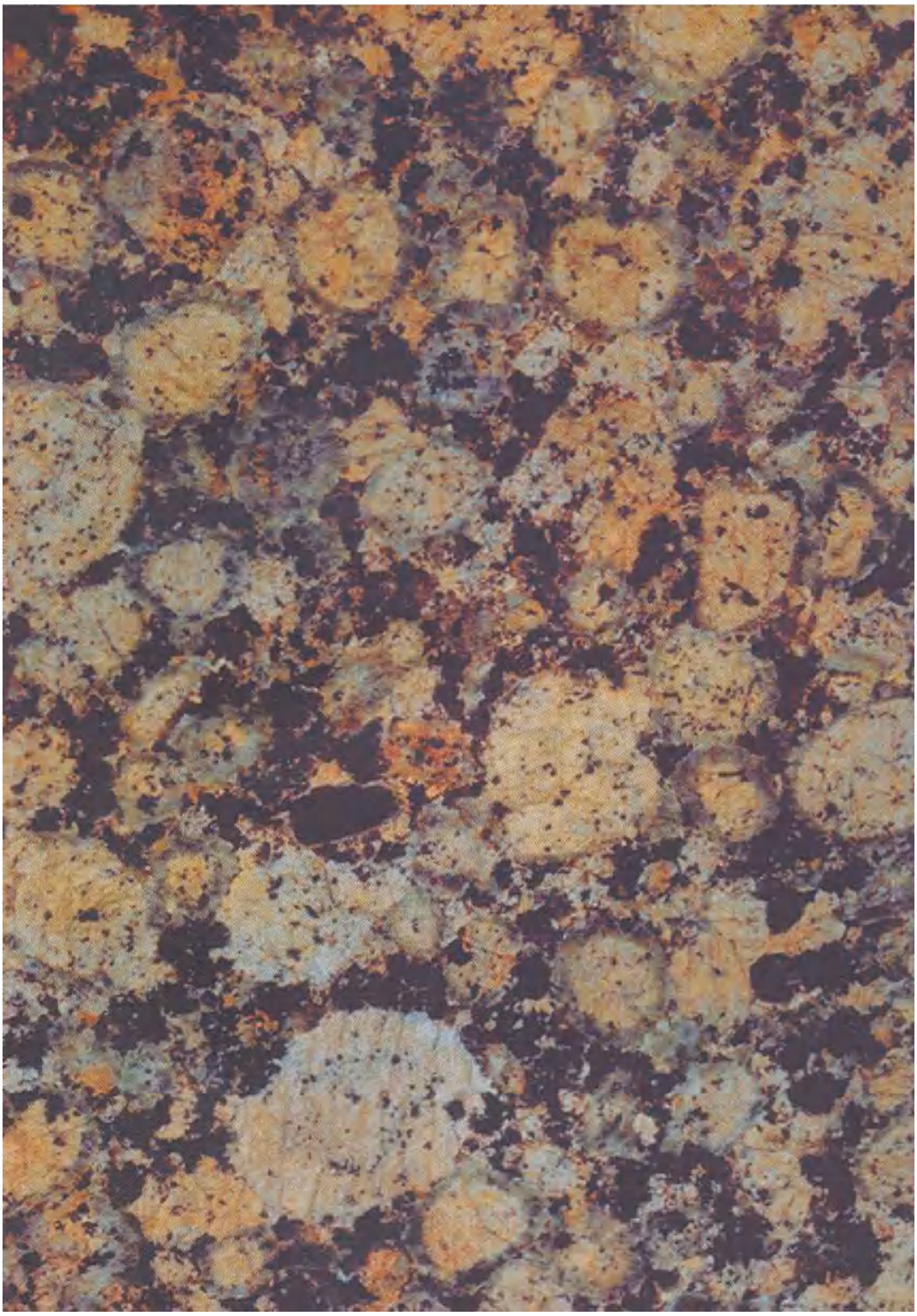

11. Rocks, meteorites and tektites

Rocks consist of minerals and often of fragments of other rocks and organic matter. Since rocks are mixtures, we cannot determine certain physical / chemical data (chemical formula, crystal system, hardness, etc.). Unlike minerals, some rocks form huge bodies which may cover several thousand km² so no single localities are shown in this book. Rocks are mainly used for building, in agriculture, and as raw materials for chemicals, ceramics and metals.
Rocks are generally divided into three principal groups according to their origin:
1. **igneous rocks**
2. **sedimentary rocks**
3. **metamorphic rocks**.

1. Igneous rocks

Properties: C – light gray, gray, white, brown, black, gray-green, pinkish, red-brown; D – opaque, rare transparent to translucent; L – dull, rare vitreous, DE – varies from 2.6 up to about 3.8; F – uneven, rare conchoidal; M – coarse to fine grained, commonly massive aggregates, consisting of microscopic or larger grains or crystals of various minerals, up to several decameters across, glass is also present rarely. Minerals in rocks usually have characteristic texture features, like graphic granite and others.
Origin and classification: Igneous rocks form at high temperature and commonly at high pressure in solidification of mainly silicate magma of variable composition. The mineral composition reflects a chemical composition of magma. Typical rock-forming minerals are quartz, orthoclase, microcline, plagioclase, biotite, muscovite, amphiboles, pyroxenes, olivine and nepheline. Rocks may solidify at various depths, according to which they are either intrusive, dyke or effusive. They can be grouped according to chemical composition (SiO_2 content):
felsic – granite, syenite, pegmatite, rhyolite, obsidian
intermediate – diorite, andesit **basic** – gabbro, basalt
utrabasic – peridotite, varies from 1.0 up to about 2.8; F – uneven, sometimes conchoidal; M – coarse to fine grained and massive aggregates, consisting of grains or crystals, from microscopic to several dm in size, also of rock fragments and organic matter, banded textures are typical and some rocks are fossiliferous.

Granite, 65 mm, Norway

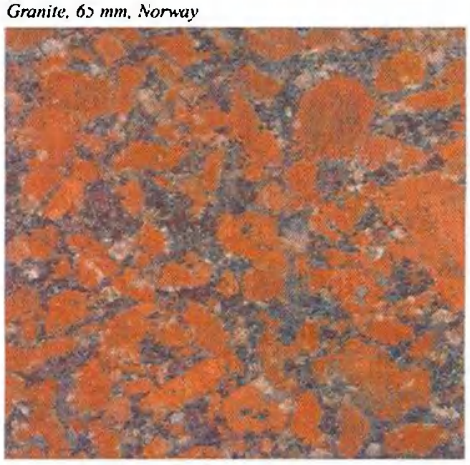

Orbicular granite, 150 mm, Sweden
Pegmatite, 65 mm, Dolni Bory, Czech Republic

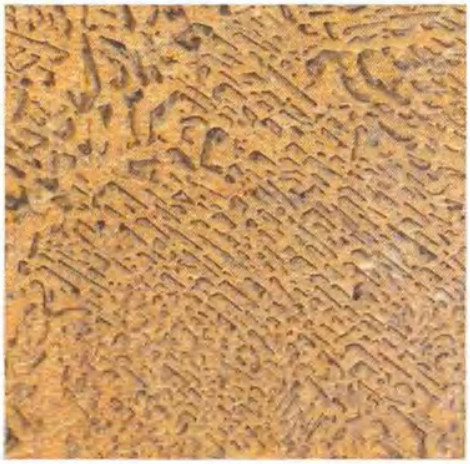

Diabase, 65 mm, Rožany, Czech Republic

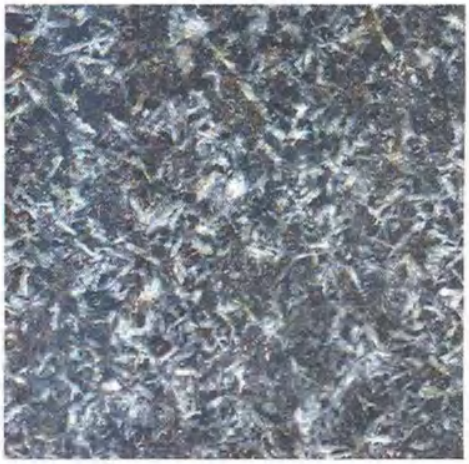

Sandstone, 65 mm, Buffalo Gap, South Dakota, U.S.A.

Conglomerate, 65 mm, St.Alban, UK

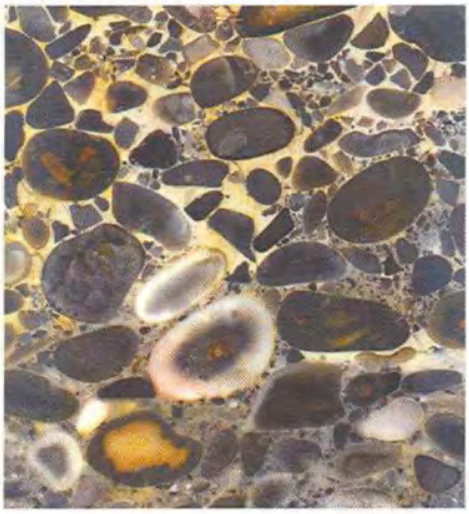

2. Sedimentary rocks

Properties: C – grey, brown, red-brown, black, green-grey, pinkish, white, yellow, often variable even within one rock, banding is typical; D – opaque, rare translucent; L – dull, sometimes vitreous, greasy or earthy; DE -varies from 1.0 up to about 2.8; F - uneven, sometimes conchoidal; M – coarse to fine grained and massive aggregates, consisting of grains or crystals, ranging from microscopic to several dm in size, also of rock fragments and organic matter, banded textures are typical and some rocks are fossiliferous.

Origin and classification: Sedimentary rocks originate under surface temperatures and pressures, as a result of sedimentation of mineral and rock fragments and organic matter of different size through the water and wind activity, or by precipitation from water solutions. Typical rock-forming minerals are quartz, calcite, dolomite, halite, clay minerals and others. According to their origins we can distinguish several groups of sedimentary rocks:

clastic (consisting of rock fragments) – sandstone, conglomerate, quartzite, siltstone.

organic (consisting mainly of organic matter) – limestone, coal.

chemical (originating by precipitation from water solutions) – evaporites, travertine.

Limestone, 65 mm, Italy

Evaporite, 40 mm, Walkenried, Germany

Serpentinite, 65 mm, Hrubšice, Czech Republic

Gneiss, 65 mm, Doubravcany, Czech Republic

3. Metamorphic rocks

Properties: C – light gray, gray, brown, green, red-brown, black, green-gray, pinkish, white, sometimes variable within one rock, banding is relatively common; D – opaque; L – dull, rare vitreous; DE – varies from 2.5 to 4.8; F – uneven, rare conchoidal; M – coarse to fine-grained, platy, acicular and sometimes massive aggregates, consisting of grains and crystals, ranging from microscopic to several decimeters across. Typical are planar textures and foliation of some minerals.

Origin and classification: Metamorphic rocks originate under higher temperature and pressure during metamorphism of originally igneous or sedimentary rocks. The source of a thermal energy could be magma, then this type is called contact metamorphism, or the thermal source lies in the depth of the earth's crust and effects large areas, then this type is called regional metamorphism. During a process of metamorphism new minerals originate. Typical rock-forming metamorphic minerals are quartz, orthoclase, plagioclases, biotite, muscovite, amphiboles, pyroxenes, calcite, dolomite, sillimanite, kyanite, almandine, staurolite and serpentine.

Regionally metamorphosed rocks: serpentinite, mica schist, gneiss, marble.

Contact metamorphosed rocks: contact chert (porcelanite), skarn.

Marble, 100 mm, Greece

Skarn, 65 mm, Lisna, Czech Republic

Iron meteorite, 50 mm, Sikhote-Alin, Russia

Iron meteorite, 40 mm detail, Gibeon, Namibia

Meteorites

Properties: C - light gray, gray, gray-green, black; D – opaque; L – dull, metallic, rare vitreous; DE – varies from 3.0 to 7.3; F – uneven; M – coarse to fine-grained, sometimes massive aggregates, consisting of irregular grains of different minerals, ranging from microscopic to several cm across.

Origin and classification: Meteorites are igneousrocks formed in space. Most originate in the asteroidbelt between Mars and Jupiter. They consist of various minerals and their chemical composition differs greatly. Typical rock-forming minerals in meteorites are olivine, pyroxenes, plagioclases, Fe and Ni alloys and sulfides, rarely also organic matter. Meteorites fall into four main groups, accordingto metallic iron and silicate component:

iron meteorite;
siderolite;
chondrite;
achondrite;

Occurrence: Meteorite falls are known throughout the world. The largest known iron meteorite, weighing approximately 60 tons , is located near the Hoba farm, Namibia. The largest known chondrite, weighing about 1 ton , fell in 1948 in Norton County, Nebraska, USA. Most meteorites that have been found recently come from large glaciers such as those of Antarctica and from desert in Namibia.

Moldavite, 35 mm, Southern Bohemia, Czech Republic

Tectite, 55 mm, Dalat, Vietnam

Tektites

Properties: C – light to dark green, yellow-green, brown-green, brown, green-gray, black; D – transparent, translucent to opaque; L – vitreous; DE – varies from 2.3 to 2.6; H – 6-7; F – conchoidal to uneven; M – massive irregular, drop-shaped or disc-shaped aggregates, irregular fragments, sometimes with typical sculptured surface.

Origin, classification and occurrence: Natural glasses, rich in SiO2, which formed as a result of rapid melting of surface rocks during impacts of large meteorites or comets.

They are classified according to their age and occurrence (following sequence from the oldest to the youngest):

Bediasites and georgianites – the USA, Mexico, Barbados, Cuba;

Urengoites – Novyi Urengoi, Russia;

Moldavites – southern Bohemia and western Moravia, Czech Republic;

Ivorites – Ivory Coast;

Irghisites – Zhamanshin, Russia;

Indochinites, phillipinites, javanites, bilitonites – southeastern Asia;

Australites – Australia.

Application: some tektites, mainly moldavites, as gemstones.

Moldavites, 40 mm, Southern Bohemia, Czech Republic

Recommended literature

Anthony J.W. et al. (1990, 1995, 1997):
Handbook of Mineralogy I, II, III,
Mineral Data Publishing, Tucson.

Bernard J.H., Rost R. a kol. (1992):
Encyklopedický prehled minerálů,
Academia, Prague, in Czech.

Clark A.M. (1993):
Hey's Mineral Index, 3rd print,
Chapman & Hall, London.

Duda R., Rejl L. (1986):
Minerals of the World,
Hamlyn, Twickenham.

Fleischer M., Mandarino J.A. (1995):
Glossary of Mineral Species,
Mineralogical Record, Tucson.

Roberts W.L. et al. (1990):
Encyclopedia of Minerals,
Van Nostrand Reinhold, New York.

Strunz H. (1977):
Mineralogische Tabellen, 6th print,
Geest & Portig, Leipzig.

Porcelanite, 85 mm, Komňa, Czech Republic
Agate, 50 mm, Zeleznice, Czech Republic

THE COMPLETE
ENCYCLOPEDIA OF

MINERALS

The Complete Encyclopedia of Minerals is a book for all the people, who are interested in the beautiful world of minerals. They can find here hundreds of photographs of beautiful crystals and crystal aggregates, together with a list of classic and new occurrences for each mineral species. Minerals are arranged into chapters according to the mineral system:

- Elements • Sulfides • Halides • Oxides
- Carbonates • Borates • Sulphates • Phosphates
- Silicates • Organic compounds
- Rocks, meteorites and tektites

ISBN 1-84013-404-6
9 781840 134049

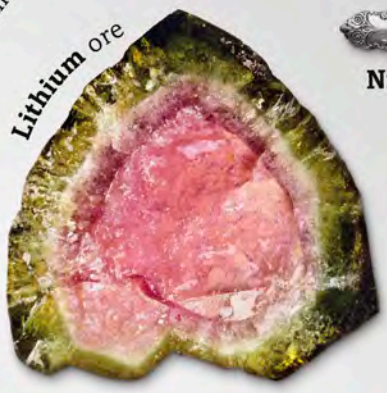

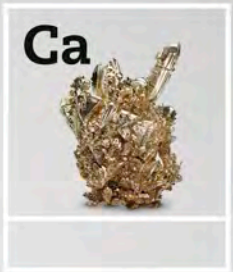

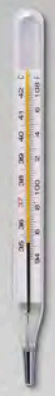

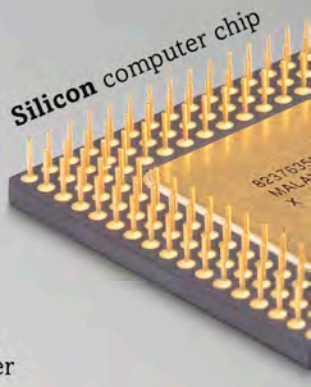

THE PERIODIC TABLE BOOK

A VISUAL ENCYCLOPEDIA OF THE ELEMENTS

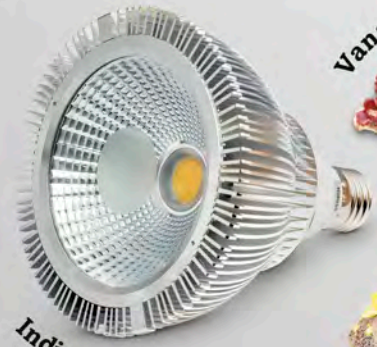

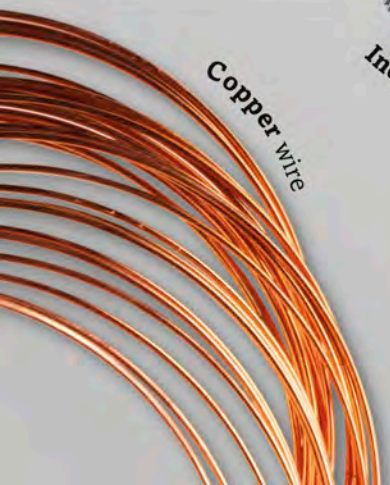

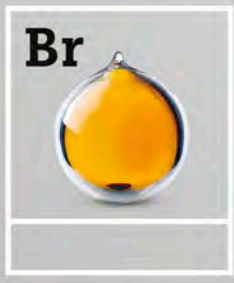

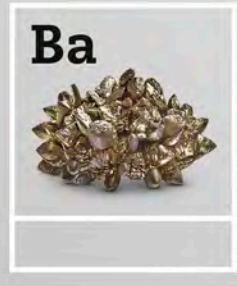

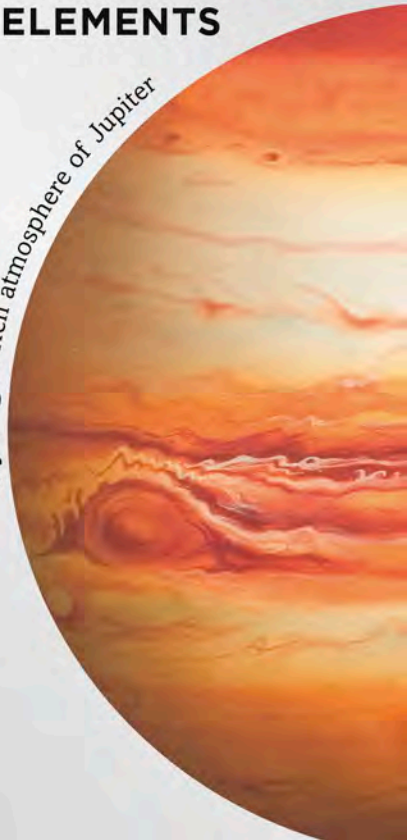

Penguin Random House

DK India

Senior Editor Bharti Bedi
Project Art Editor Amit Verma
Editorial Team Neha Ruth Samuel, Charvi Arora, Deeksha Saikia
Art Editors Mansi Agrawal, Amisha Gupta, Ravi Indiver
Assistant Art Editors Neetika Malik Jhingan, Nidhi Rastogi
Jacket Designer Suhita Dharamjit
Jackets Editorial Coordinator Priyanka Sharma
Senior DTP Designer Harish Aggarwal
DTP Designers Sachin Gupta, Syed Md Farhan, Nityanand Kumar, Mohammad Rizwan
Picture Researcher Nishwan Rasool
Managing Jackets Editors Saloni Singh, Sreshtha Bhattacharya
Picture Research Manager Taiyaba Khatoon
Pre-production Manager Balwant Singh
Production Manager Pankaj Sharma
Managing Editor Kingshuk Ghoshal
Managing Art Editor Govind Mittal

DK UK

Project Editor Ashwin Khurana
Senior Art Editor Smiljka Surla
Jacket Editor Claire Gell
Senior Jacket Designer Mark Cavanagh
Jacket Design Development Manager Sophia MTT
Managing Editor Dr Lisa Gillespie
Managing Art Editor Owen Peyton Jones
Producers, Pre-production Dragana Puvacic, Catherine Williams
Producer Anna Vallarino
Publisher Andrew Macintyre
Art Director Karen Self
Associate Publishing Director Liz Wheeler
Design Director Phil Ormerod
Publishing Director Jonathan Metcalf

Photographer Ruth Jenkinson
Photography Assistant Julie Stewart

Element samples prepared and supplied by RGB Research Ltd
www.periodictable.co.uk

First published in Great Britain in 2017 by
Dorling Kindersley Limited
80 Strand, London WC2R 0RL

Copyright © 2017 Dorling Kindersley Limited
A Penguin Random House Company
10 9 8 7 6 5 4 3 2 1
001–289022–April/2017

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the copyright owner.

A CIP catalogue record for this book is available from the British Library

ISBN: 978-0-2412-4043-4

Printed in China

A WORLD OF IDEAS:
SEE ALL THERE IS TO KNOW

www.dk.com

CONTENTS

Foreword 6
Elemental building blocks 8
Chemical discoveries 10
Inside an atom 12
Periodic table of elements 14
Reactions and uses 16

Hydrogen 18
Hydrogen 20

Alkali Metals 22
Lithium 24
Sodium 26
Salt flats 28
Potassium 30
Rubidium 32
Caesium, Francium 34

Alkaline Earth Metals 36
Beryllium 38
Magnesium 40
Calcium 42
Fly Geyser 44
Strontium 46
Barium 48
Radium 50

Transition Metals 52

Scandium, Titanium 54
Vanadium, Chromium 56
Manganese 58
Iron 60
Steelmaking 62
Cobalt 64
Nickel 66
Copper 68
Copper wires 70
Zinc 72
Yttrium 74
Zirconium, Niobium 76
Molybdenum, Technetium 78
Ruthenium, Rhodium 80
Palladium 82
Silver 84
Cadmium, Hafnium 86
Tantalum, Tungsten 88
Rhenium, Osmium 90
Iridium 92
Platinum 94
Gold 96
Golden Buddha 98
Mercury 100
Rutherfordium, Dubnium, Seaborgium 102
Bohrium, Hassium, Meitnerium 104
Darmstadtium, Roentgenium, Copernicium 106

Lanthanides 108

Lanthanum, Cerium, Praseodymium 110
Neodymium, Promethium, Samarium, Europium 112
Gadolinium, Terbium, Dysprosium, Holmium 114
Erbium, Thulium, Ytterbium, Lutetium 116

Actinides 118

Actinium, Thorium, Protactinium 120
Uranium, Neptunium, Plutonium, Americium 122
Curium, Berkelium , Californium, Einsteinium 124
Fermium, Mendelevium, Nobelium, Lawrencium 126

The Boron Group 128

Boron 130
Aluminium 132
Jet turbine 134
Gallium, Indium 136
Thallium, Nihonium 138

The Carbon Group 140

Carbon 142
Pink diamond 144
Silicon 146
Germanium, Tin 148
Lead, Flerovium 150

The Nitrogen Group 152

Nitrogen 154
Drag racing 156
Phosphorus 158
Arsenic, Antimony 160
Bismuth, Moscovium 162

The Oxygen Group 164

Oxygen 166
Sulfur 168
Danakil Depression 170
Selenium, Tellurium 172
Polonium, Livermorium 174

The Halogen Group 176

Fluorine 178
Chlorine 180
Ocean clean up 182
Bromine 184
Iodine, Astatine, Tennessine 186

Noble Gases 188

Helium 190
Nebula 192
Neon, Argon 194
Krypton, Xenon 196
Radon, Oganesson 198

Glossary 200
Index 204
Acknowledgements 208

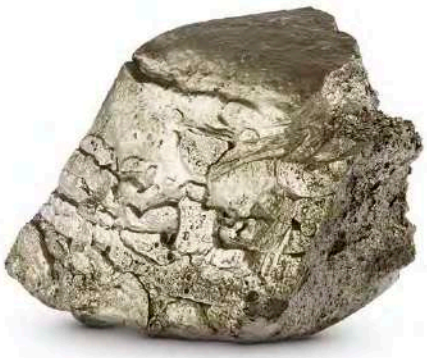

Chunk of yttrium

Chunk of silver

Zirconium crystal bar

Foreword

Everything in nature, from the mountains and the oceans to the air we breathe and food we eat are made up of simple substances called elements. You may have already heard of several of them, including gold, iron, oxygen, and helium, but these are just four out of a total of 118. Many have unique – and sometimes surprising – chemical and physical properties. Gallium, for example, is a solid but melts in your hand. A compound of sulfur gives off a nasty smell of rotten eggs. Fluorine is a gas that can burn a hole straight through concrete!

The elements are rarely found in their pure form. Mostly, they are combined with each other to make compounds, which make up substances around us. For example, hydrogen and oxygen make water, sodium and chlorine form salt, and carbon is found in millions of compounds, many of which – including proteins and sugars – make our bodies work.

To find out more about the elements, we need to take a good look at the periodic table. This is used by scientists around the world to list and detail the elements. It shows the key information

Nickel balls

Cube of melting gallium

Iodine in a glass sphere

Barium crystals

Chunk of grey selenium

Magnesium crystals

Osmium pellet

for each element, grouping them into similar types. With this information, we can use the elements to make many things we need: a fluorine compound in toothpastes toughens our teeth and silicon crystals engineered into microchips operate our gadgets and phones.

Every element has its own story of where it comes from, what it can do, and how we use it. Let's begin a tour of every element one by one. It's going to be a fascinating journey.

Tom Jackson

Throughout this book you will find boxes with the following symbols. This is what each of them mean.

	This shows the structure of an atom of an element, with the nucleus (made of protons and neutrons) at the centre and electrons surrounding it in their shells.
	Electron
	Proton
	Neutron
State	The state of the element at a temperature of 20°C (28°F). It can be a liquid, solid, or gas.
Discovery	This details the year in which the element was discovered.

Chunk of uranium

Gold crystals

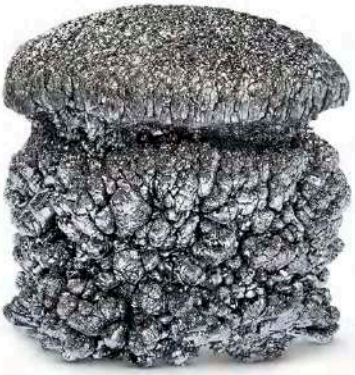

Thulium crystals

Calcium crystals

Elemental building blocks

Elements are everywhere: some you can see, like gold, others are almost invisible, like oxygen gas. An element is a substance that cannot be broken up into simpler ingredients. Each one is made up of tiny building blocks called atoms, which are unique for every element. Most elements are joined with other elements to make compounds, which are made by combining two or more elements. This includes water, which is a compound of hydrogen and oxygen.

Elements in our world

There are 118 elements in the periodic table; 92 of them are found in nature, while the others are made by humans. Every element is unique. Most of the elements are solids, like the metals. At room temperature, 11 elements are gases, while bromine and mercury are the only two liquids.

Bromine liquid with bromine gas

Bismuth crystals

Earth

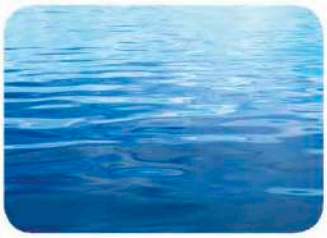

Water

Air

Fire

Ancient ideas

The idea of elements is very old, dating back about 2,600 years to ancient Greece. However, Greek thinkers believed that the world was made of just four elements: earth, water, fire, and air. Empedocles, an influential scholar, was the first to propose that these elements made up all structures. Only much later did scientists learn that none of these are actually elements. For thousands of years, everybody from ancient Egyptian priests to medieval European alchemists, speculated about the definiton and classification of an element.

Elements in and around us

About 99 per cent of the human body is made from just six elements, though they are combined together to form thousands of different compounds. On the other hand, Earth's atmosphere is a mixture of gases, most of which are pure elements. About 99 per cent of the air is made from nitrogen and oxygen.

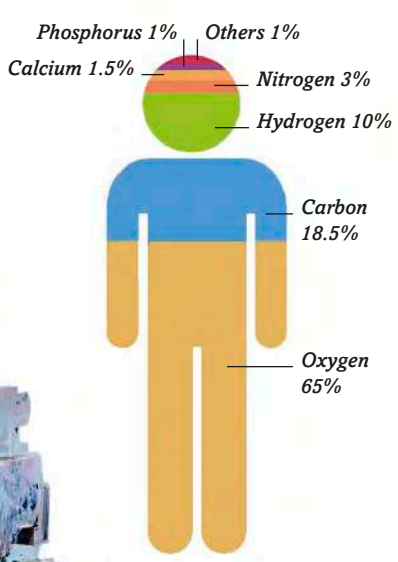

Human body

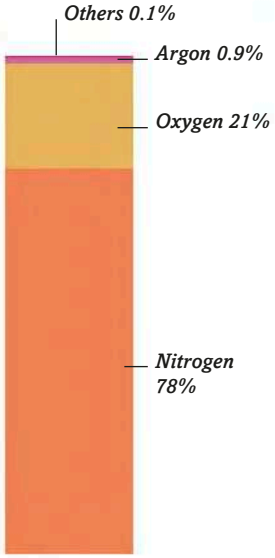

Earth's atmosphere

Gold crystals

Iranian alchemists in their workshop

Alchemy and mysticism

Chemists are scientists who study elements and compounds. However, before they existed, the alchemists were medieval researchers. Believing in a mixture of science and magic, alchemists tried to change ordinary metals (such as lead) into gold. They failed because elements cannot be changed from one type to another. But, in the process, they discovered many new elements and developed several processes that chemists still use today.

ROBERT BOYLE

The first person to use science to understand the elements was the Irish scientist and inventor Robert Boyle. He pursued science through reason, and in the 1660s he performed the first chemistry experiments to show that much of what the alchemists believed was wrong.

Chemical discoveries

The ancient concept of four elements – earth, water, fire, and air – expanded to a belief that every substance on Earth was made from a mixture of these elements. However, many substances including mercury, sulfur, and gold did not fit this idea. Over the last 300 years, chemists have followed a long series of clues to reveal the true nature of elements, their atoms, and what happens to them during chemical reactions.

Humphry Davy
In the early 19th century, the English scientist Humphry Davy discovered several new metals. He used a revolutionary process called electrolysis, in which electric currents split chemical compounds into their elements. Davy discovered a total of nine new elements, including magnesium, potassium, and calcium.

Magnesium crystals

Pioneering chemists

Many of the first breakthroughs in chemistry came in the 1700s, from investigations into the composition of air. Chemists such as Joseph Black, Henry Cavendish, and Joseph Priestly discovered several different "airs", which we now call gases. They also found that the gases could react with solid substances, which they called "earths". These discoveries began a journey that revealed that there were dozens of elements, not just four. Today, scientists have identified 118 elements, but more may be discovered in time.

Antoine Lavoisier
In 1777, the French scientist Antoine Lavoisier proved that sulfur was an element. This yellow substance was familiar for thousands of years, but Lavoisier performed experiments to show that it was a simple substance that could not be divided up any further. In the same year, he also found out that water was not an element, but a compound of hydrogen and oxygen.

Granule of pure sulfur

JOHN DALTON

Like many scientists of his day, the English scientist John Dalton already believed that matter must be made of tiny particles. In 1803, he began to think about how these particles might join together. He came to realize that there are different particles for every element, and that the particles of one element all have the same mass. He also realized that the particles of different elements combine in simple proportions to make compounds. So, for example, the particles of the elements carbon and oxygen can combine to make carbon monoxide. He suggested that during a chemical reaction, the particles rearrange to make compounds. He formulated the first modern theory of atoms.

Dalton's table of elements

Jacob Berzelius
In the early 1800s, the Swedish doctor Jacob Berzelius investigated chemicals in rocks and minerals. He found two minerals that contained new elements. He named these elements cerium (after Ceres, the dwarf planet) and thorium (after Thor, the Viking god of thunder). Berzelius also invented a system of using symbols and numbers that chemists still use to identify elements and compounds today.

Chunk of pure cerium

Pure caesium inside a sealed container

States of matter

Elements can exist in three states of matter: solid, liquid, and gas. At room temperature, most elements are solids, 11 are gases, and only two are liquids. However, elements can change from one state into another. These changes don't alter the atoms of these elements, but arrange them in different ways.

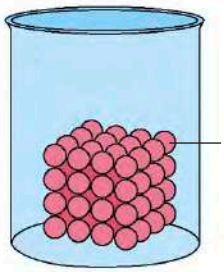

A solid keeps its shape and has a fixed volume.

In a solid, all the atoms are attracted to each other and locked in position.

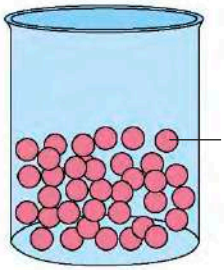

A liquid takes the shape of its container, but its volume remains fixed.

In a liquid, the atoms begin to move around as the attraction between them weakens.

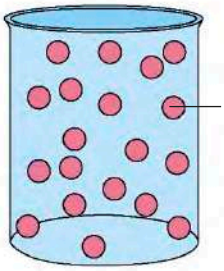

A gas will fill any container, no matter how large or small.

In a gas, the atoms are weakly attracted to each other, so they all move in different directions.

Robert Bunsen
The German chemist Robert Bunsen is best known for inventing a gas burner that is often used in laboratories. In the 1850s, Bunsen used such a burner – which produced a hot, clean flame – to study the unique flame colours produced by different elements. When an unknown substance made bright blue flames, he named it caesium, meaning "sky blue".

Inside an atom

An atom is the smallest unit of an element. Atoms are too small to see (even with the most powerful microscopes) but they are everywhere. They consist of smaller particles called protons, neutrons, and electrons. Every element has a unique number of protons.

What's the atomic number?

The number of protons in an atom of an element is called the atomic number. The atomic number of an atom identifies the element it belongs to. Every atom also has an equal number of electrons. For elements found naturally on Earth, hydrogen has the smallest atomic number (1), while uranium atoms have the highest atomic number (92).

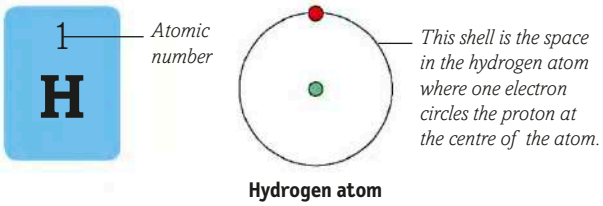

Hydrogen atom

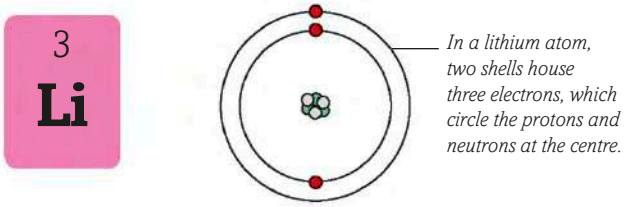

Lithium atom

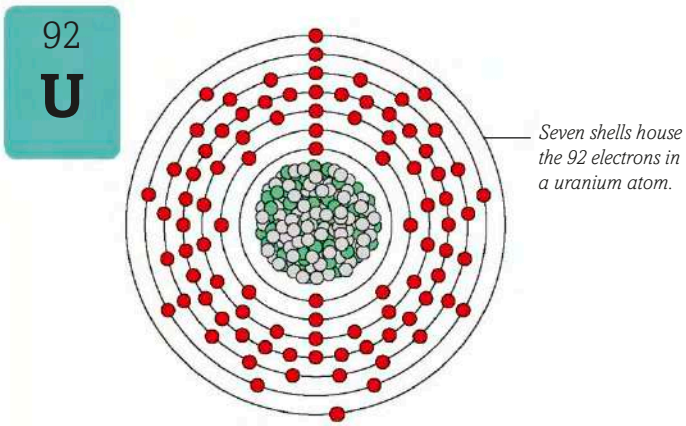

Uranium atom

Electron ❯ The tiny, negatively charged particles in an atom are called electrons. They are involved in the way the atoms of an element react and form bonds with the atoms of other elements.

Shell ❯ The electrons in an atom move around the nucleus. They are arranged in layers called shells. When reacting with each other, atoms tend to fill up their outer shells to become more stable.

Neutron ❯ As its name suggests, neutrons are neutral particles, which means they do not have an electric charge. A neutron weighs the same as a proton, and much more than a electron.

Proton ❯ Protons have a positive electric charge. This charge attracts the negatively charged electrons, holding them in place around the nucleus. Because each proton's charge is cancelled out by the equal charge of an electron, the atom has no overall charge, and is therefore neutral.

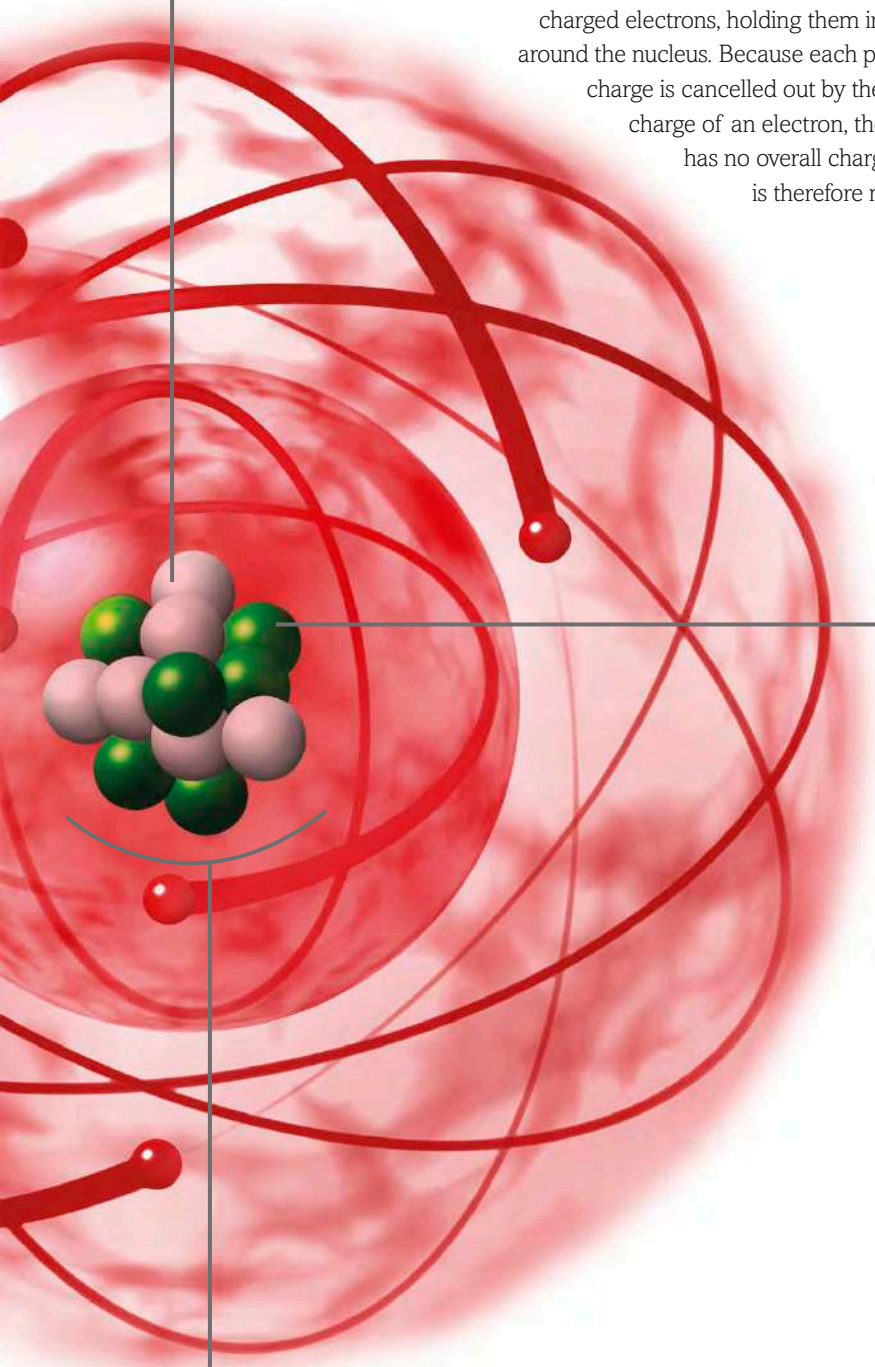

Nucleus ❯ The central core, or nucleus, of an atom is made up of protons and neutrons. Nearly all the mass of the atom is packed into the nucleus, and this gives every element a unique atomic mass.

Atomic facts

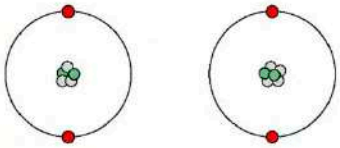

Isotopes

While every element has a unique number of electrons and protons in its atoms, the number of neutrons can vary. These different forms are called isotopes. For example, helium has two isotopes: one contains three neutrons (He-3), the other has four (He-4).

Electromagnet attracts metal pieces

Electromagnetism

Atoms work like tiny magnets. A force called electromagnetism holds them together. It makes particles with opposite charges, such as protons and electrons, attract each other. Those with similar charges repel each other. A magnet is an object in which the magnetic forces of the atoms attract and repel other objects. An electromagnet develops magnetism when an electric current runs through it.

ATOMIC PIONEERS

During his atomic research in the early 20th century, Sir Ernest Rutherford, a New Zealand scientist, expanded our understanding of the structure of atoms. He discovered protons and proved that they were located in an atom's nucleus.

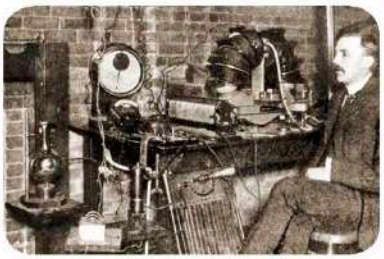

Sir Ernest Rutherford

Periodic table of elements

The periodic table is a useful way of organizing the elements. It arranges the elements in order of their atomic number, which is the number of protons in the nucleus of an atom, and is unique to every element. The table also divides the elements into rows, called "periods", and columns, called "groups". Dmitri Mendeleev, the chemist who devised the table, arranged the elements based on the similarity of certain physical and chemical properties.

1 H 1.0079											
3 Li 6.941	4 Be 9.0122										
11 Na 22.990	12 Mg 24.305										
19 K 39.098	20 Ca 40.078	21 Sc 44.956	22 Ti 47.867	23 V 50.942	24 Cr 51.996	25 Mn 54.938	26 Fe 55.845	27 Co 58.933	28 Ni 58.693	29 Cu 63.546	30 Zn 65.39
37 Rb 85.468	38 Sr 87.62	39 Y 88.906	40 Zr 91.224	41 Nb 92.906	42 Mo 95.94	43 Tc (96)	44 Ru 101.07	45 Rh 102.91	46 Pd 106.42	47 Ag 107.87	48 Cd 112.41
55 Cs 132.91	56 Ba 137.33	57-71 La-Lu	72 Hf 178.49	73 Ta 180.95	74 W 183.84	75 Re 186.21	76 Os 190.23	77 Ir 192.22	78 Pt 195.08	79 Au 196.97	80 Hg 200.59
87 Fr (223)	88 Ra (226)	89-103 Ac-Lr	104 Rf (261)	105 Db (262)	106 Sg (266)	107 Bh (264)	108 Hs (277)	109 Mt (268)	110 Ds (281)	111 Rg (272)	112 Cn 285
			57 La 138.91	58 Ce 140.12	59 Pr 140.91	60 Nd 144.24	61 Pm (145)	62 Sm (150.36)	63 Eu 151.96	64 Gd 157.25	65 Tb 158.93
			89 Ac (227)	90 Th 232.04	91 Pa 231.04	92 U 238.03	93 Np (237)	94 Pu (244)	95 Am (243)	96 Cm (247)	97 Bk (247)

The actinides and the lanthanides *are placed between the alkaline earth metals and the transition metals, but have been moved below to give them more space.*

KEY

- Hydrogen
- Alkali Metals
- Alkaline Earth Metals
- Transition Metals
- Lanthanides
- Actinides
- The Boron Group
- The Carbon Group
- The Nitrogen Group
- The Oxygen Group
- The Halogen Group
- Noble Gases

Elements of this group *are semi-metals (elements with the properties of metals and non-metals): they are shiny like metals but crumble easily like non-metals.*

This group *contains the noble gases, which never form bonds with other elements, and are unreactive.*

					2 He 4.0026
5 B 10.811	6 C 12.011	7 N 14.007	8 O 15.999	9 F 18.998	10 Ne 20.180
13 Al 26.982	14 Si 28.086	15 P 30.974	16 S 32.065	17 Cl 35.453	18 Ar 39.948
31 Ga 69.723	32 Ge 72.64	33 As 74.922	34 Se 78.96	35 Br 79.904	36 Kr 83.80
49 In 114.82	50 Sn 118.71	51 Sb 121.76	52 Te 127.60	53 I 126.90	54 Xe 131.29
81 Tl 204.38	82 Pb 207.2	83 Bi 208.96	84 Po (209)	85 At (210)	86 Rn (222)
113 Nh 284	114 Fl 289	115 Mc 288	116 Lv 293	117 Ts 294	118 Og 294
66 Dy 162.50	67 Ho 164.93	68 Er 167.26	69 Tm 168.93	70 Yb 173.04	71 Lu 174.97
98 Cf (251)	99 Es (252)	100 Fm (257)	101 Md (258)	102 No (259)	103 Lr (262)

Reading the table

Element symbol

Every element has a unique symbol of one or two letters. These symbols ensure that scientists who speak different languages do not get confused while describing the same element.

3
Li
6.941

The atomic number is the number of protons in the nucleus of this element's atoms.

The first letter of a symbol is always a capital, but the second is lower case.

The atomic mass number is the average of all the atoms of the element. It is not a whole number because there are different isotopes (forms) of each element, each with a different number of neutrons.

Periods

Elements in the same period, or row, have the same number of electron shells in their atoms. So elements in period one have one electron shell, while those in period six have six electron shells.

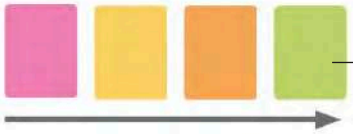

Periods run from left to right.

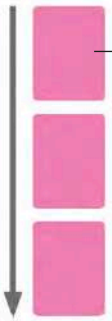

Groups run from top to bottom.

Groups

Members of a group, or column, all have the same number of electrons in their outermost shell. For example, group one elements have one outer electron, while group eight elements have eight outer electrons.

DMITRI MENDELEEV

The periodic table was developed by the Russian chemist Dmitri Mendeleev in 1869. Others had tried before, but his table was periodic, or repeating, because the characteristics of elements follow a pattern. The table was incomplete as some elements had not yet been discovered. However, Mendeleev predicted the positions of the missing elements, and was proved right when they were finally isolated many years later.

Reactions and uses

Explosive reaction
In this chemical reaction, pure lithium reacts with air to make the compound lithium oxide. It takes energy to break the links between the lithium atoms and then make bonds with oxygen in the air. Reactions need energy to begin, but they often produce energy as heat and light.

The elements can combine in different ways to make 10 million compounds, possibly more. As well as learning about the physical and chemical properties of elements, chemists also want to find out how and why certain elements react with each other to form compounds. Chemical reactions are happening all the time. During a reaction, substances change into new substances. The bonds that hold them are broken and then remade in a different combination.

1. This piece of pure lithium is placed on a surface and is exposed to the air.

2. A gas torch is used to heat the lithium, and in just a few seconds it turns red, which is a typical colour for this metal when it becomes hot.

3. Very quickly, the lithium catches fire. The white areas forming here are the compound lithium oxide, which is a combintion of lithium and oxygen.

Mixtures

A mixture is a combination of substances that can be separated by physical means, such as filtering. It is not the same as a compound, where the ingredients are connected by bonds and can only be separated using a chemical reaction. Mixtures can be classified as solutions, colloids, and suspensions.

Solution
In this mixture, a substance is completely and evenly mixed, or dissolved, into another substance. Seawater is a solution.

Colloid
This mixture contains unevenly spread particles and clusters that are too small to see. Milk is a colloid.

Suspension
This type of mixture consists of large particles of one substance floating in another substance. Muddy water is a suspension.

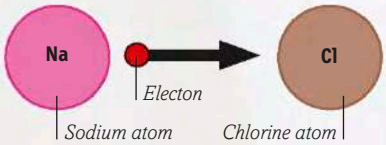

1. A sodium atom donates one electron to a chlorine atom. This gives both atoms full outer electron shells.

2. These are now charged atoms known as ions. The sodium ion has a positive charge and the chlorine ion has a negative charge.

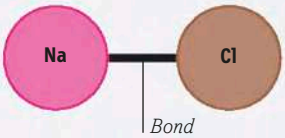

3. Sodium is attracted to – and forms a bond with – chlorine, forming a molecule of the compound sodium chloride.

Forming compounds

There are two kinds of bonds formed between elements during a chemical reaction. In an ionic bond, such as in sodium chloride (above), one atom gives away its electron(s) and another accepts them. This results in each having full outer electron shells. The other type is called covalent bonding. In this, atoms sit together and share their electrons so they both have full outer shells.

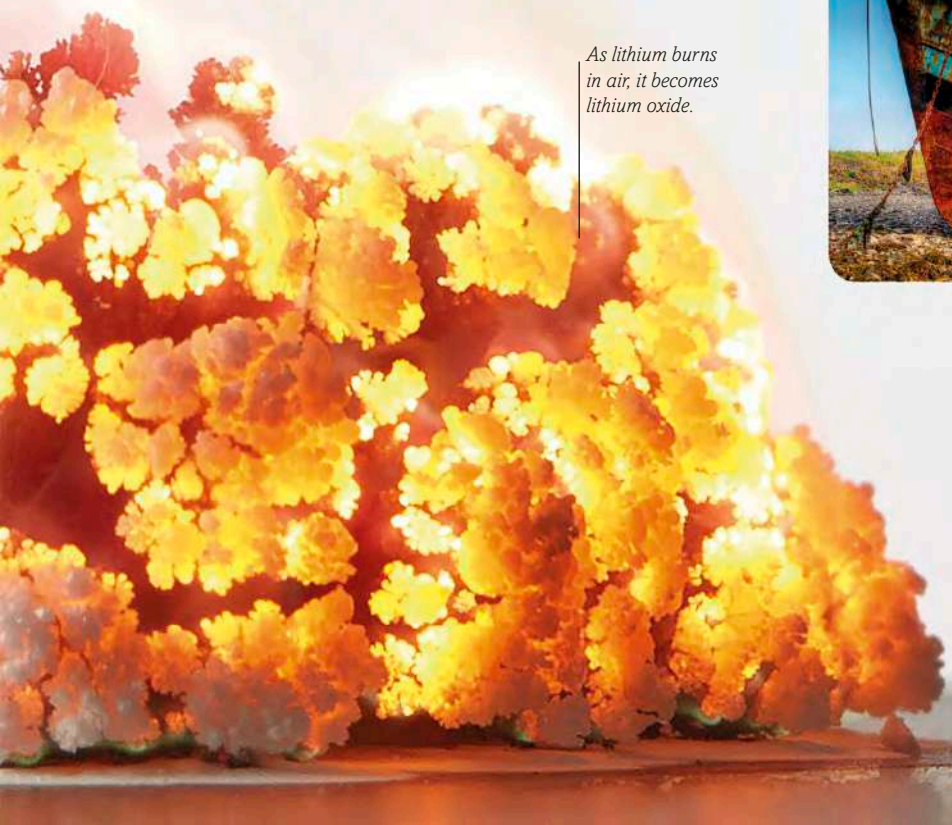

As lithium burns in air, it becomes lithium oxide.

Reactions in the real world

Chemical reactions happen all around us. There are reactions when we cook, take medication, or breathe. The image above shows a rusty iron ship. Over time, the element iron develops this red, flaky layer when it reacts with oxygen present in water or air to form the compound iron oxide – more commonly known as rust.

Pure hydrogen (H) fills this glass sphere, and produces a purple glow when electrified.

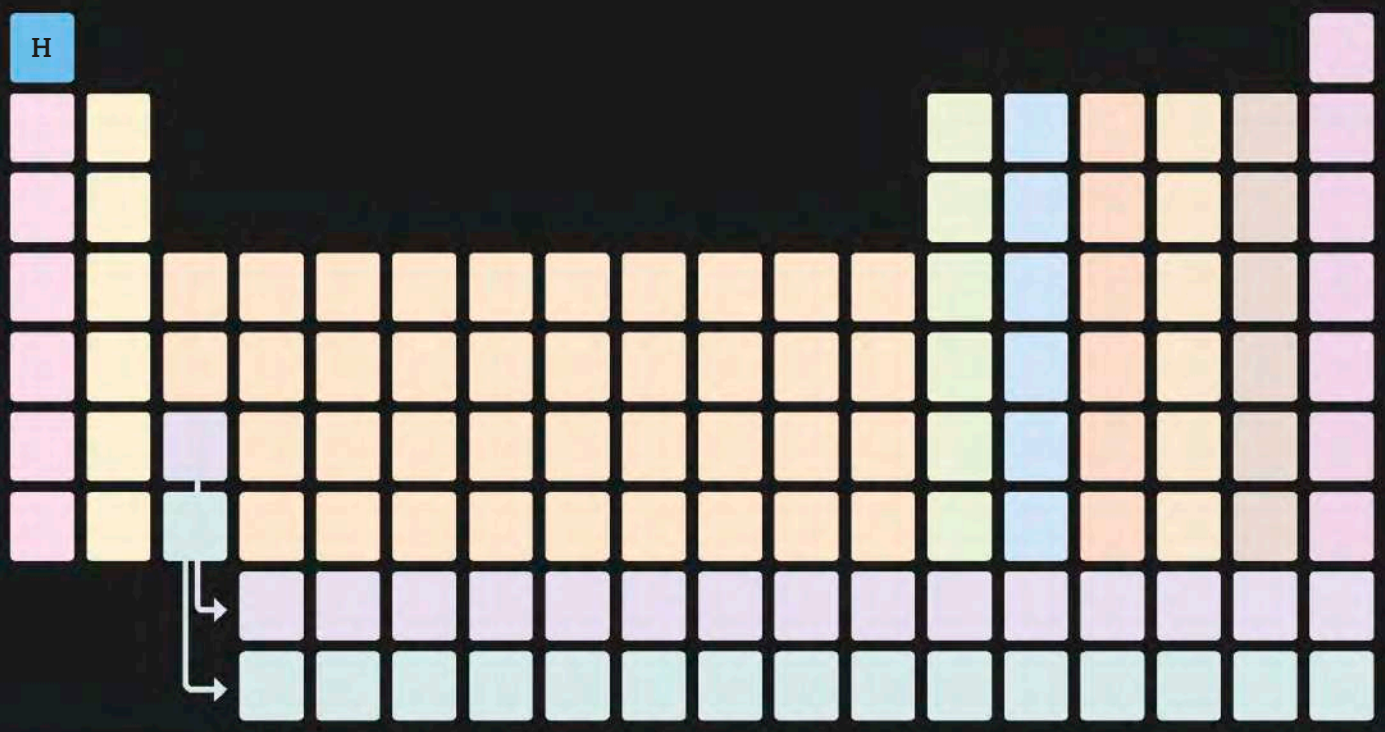

Hydrogen

The first element, hydrogen (H), is located above the alkali metals in the first column of the periodic table. However, because it is so different to the elements below it, hydrogen is not included in their group. This gas has the simplest atoms of any element with one electron and one proton. It is highly reactive and forms compounds with all kinds of other elements.

Atomic structure
A hydrogen (H) atom has one electron moving around a nucleus consisting of a single proton.

Physical properties
Hydrogen gas is the lightest material in the Universe. Pure hydrogen is rare on Earth, as it escapes quickly from the atmosphere into space.

Chemical properties
Hydrogen is highly flammable. It forms compounds with both metals and non-metals.

Compounds
The most common hydrogen compound is water. Acids are compounds that contain hydrogen.

1 H Hydrogen

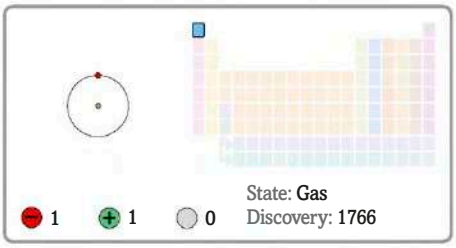

Forms

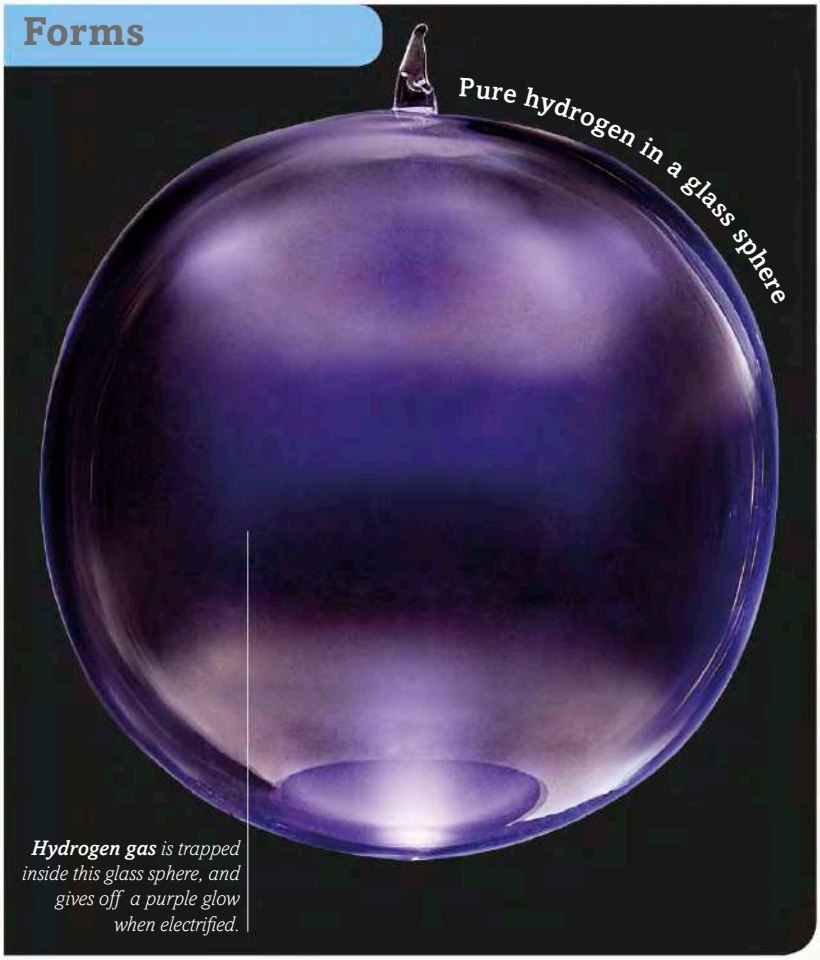

Hydrogen gas *is trapped inside this glass sphere, and gives off a purple glow when electrified.*

The Orion Nebula

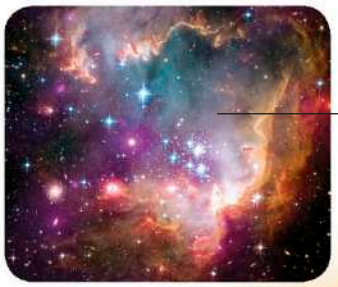

This gaseous stellar nursery *is giving birth to thousands of stars.*

The Sun

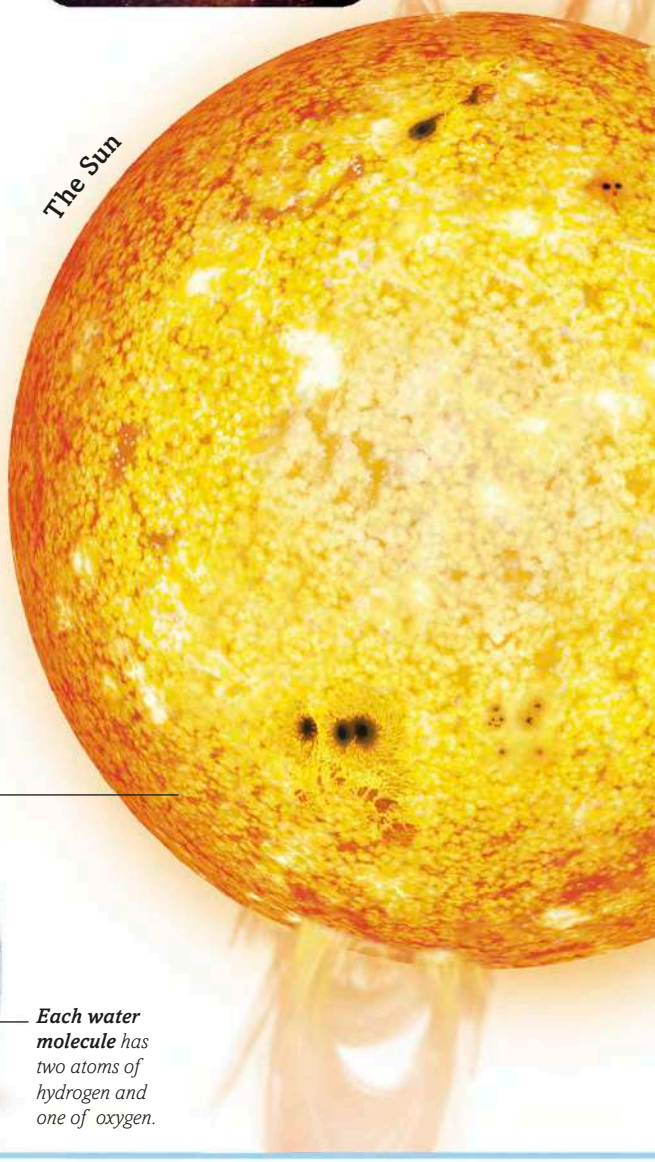

The Sun *is four-fifths hydrogen.*

Jupiter

Three quarters of this planet *is made up of layers of gaseous and liquid hydrogen.*

Water

Each water molecule *has two atoms of hydrogen and one of oxygen.*

Hydrogen is the first member of the periodic table because it has the simplest atoms of all elements: they contain just one proton and one electron. **Pure hydrogen** is a transparent gas. The biggest planets, such as **Jupiter**, are vast balls of hydrogen mixed with other gases, such as helium and methane. On Earth, hydrogen is commonly found in **water**. Although it is rare in Earth's atmosphere, hydrogen is the most common element in the Universe. Stars, such as the **Sun**, contain large amounts of hydrogen. At the centre of a star, atoms of

Uses

Hydrogen-filled balloon

This balloon *can rise high into the atmosphere where sensors gather information about atmospheric pressure, temperature, and wind speed.*

HOW ROCKET FUEL WORKS

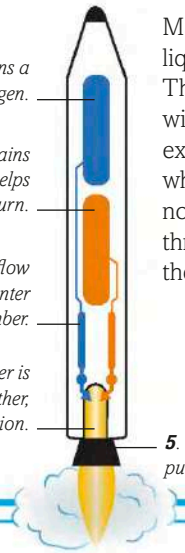

1*. This chamber contains a fuel called liquid hydrogen.*

2*. This chamber contains liquid oxygen, which helps the hydrogen burn.*

3*. Pumps control the flow of the liquids as they enter the combustion chamber.*

4*. The combustion chamber is where the liquids mix together, creating an explosion.*

5*. The nozzle emits hot vapour, pushing the rocket upwards.*

Many space rockets use liquid hydrogen as a fuel. The hydrogen reacts with oxygen to form extremely hot steam, which blasts out of the nozzle. This creates thrust, which pushes the rocket upwards.

Delta IV rocket

This powerful rocket *uses 45,460 litres (12,000 gal) of liquid hydrogen as fuel.*

Margarine

Margarine *is made of vegetable oils thickened by adding hydrogen.*

Hydrogen peroxide

This liquid *is used as a cleaner.*

The only **waste product** of hydrogen fuel is **steam**.

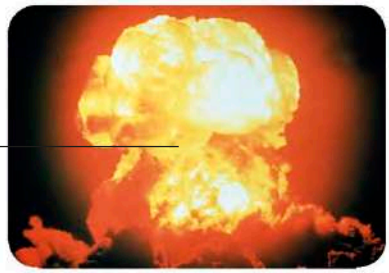

Hydrogen bomb explosion

This powerful explosion was *created by fusing hydrogen atoms.*

This energy-efficient bus *runs on a fuel cell fed by hydrogen.*

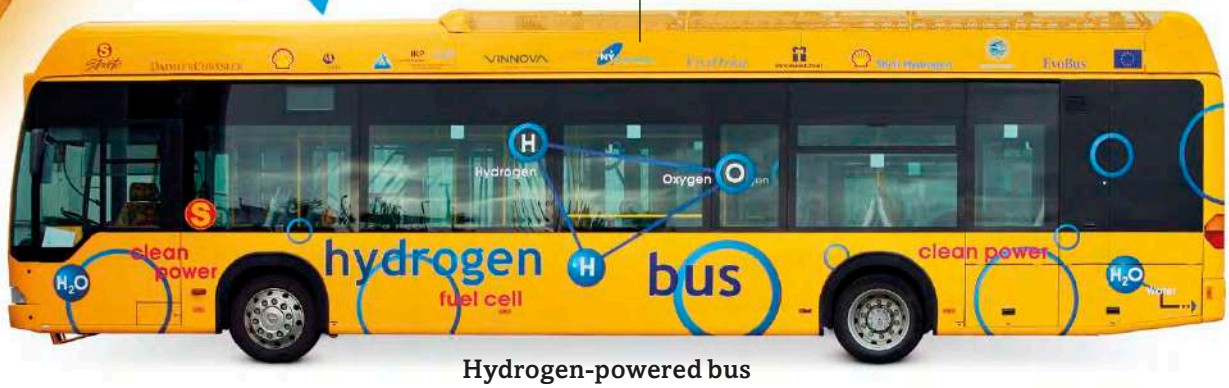

Hydrogen-powered bus

this element are fused together, releasing heat and light. New stars form inside **nebulae** – such as the **Orion Nebula**. They are clouds of hydrogen gas that slowly collapse in on themselves. Hydrogen gas is the lightest element of all, and much lighter than air. This is why **hydrogen-filled balloons** can fly higher than air-filled ones. Supercold liquid hydrogen is used as **rocket** fuel. Atoms of hydrogen fuse together to produce a lot of energy in **hydrogen bomb** explosions. Pure hydrogen is also a clean energy source used to power some **buses** and cars.

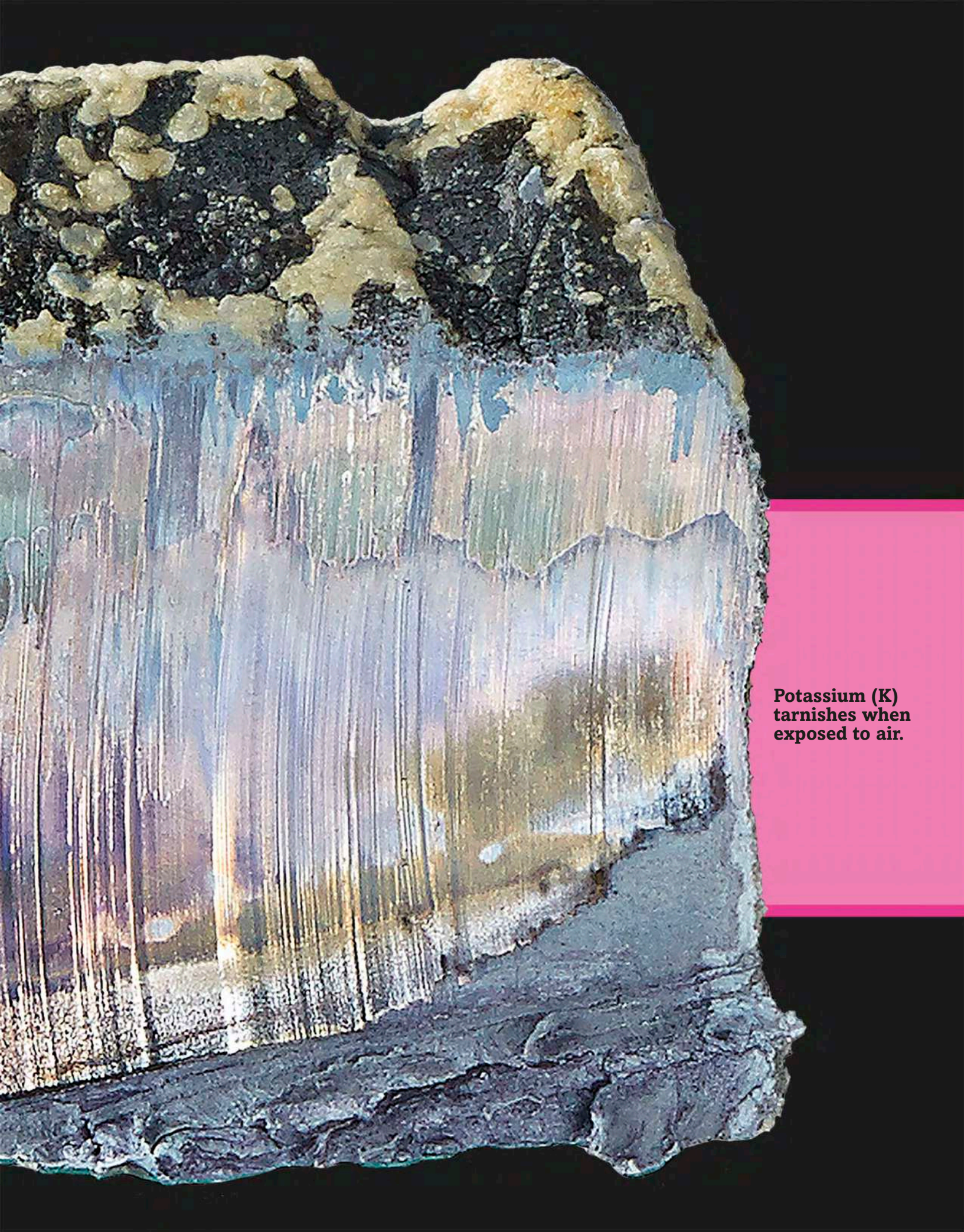

Potassium (K) tarnishes when exposed to air.

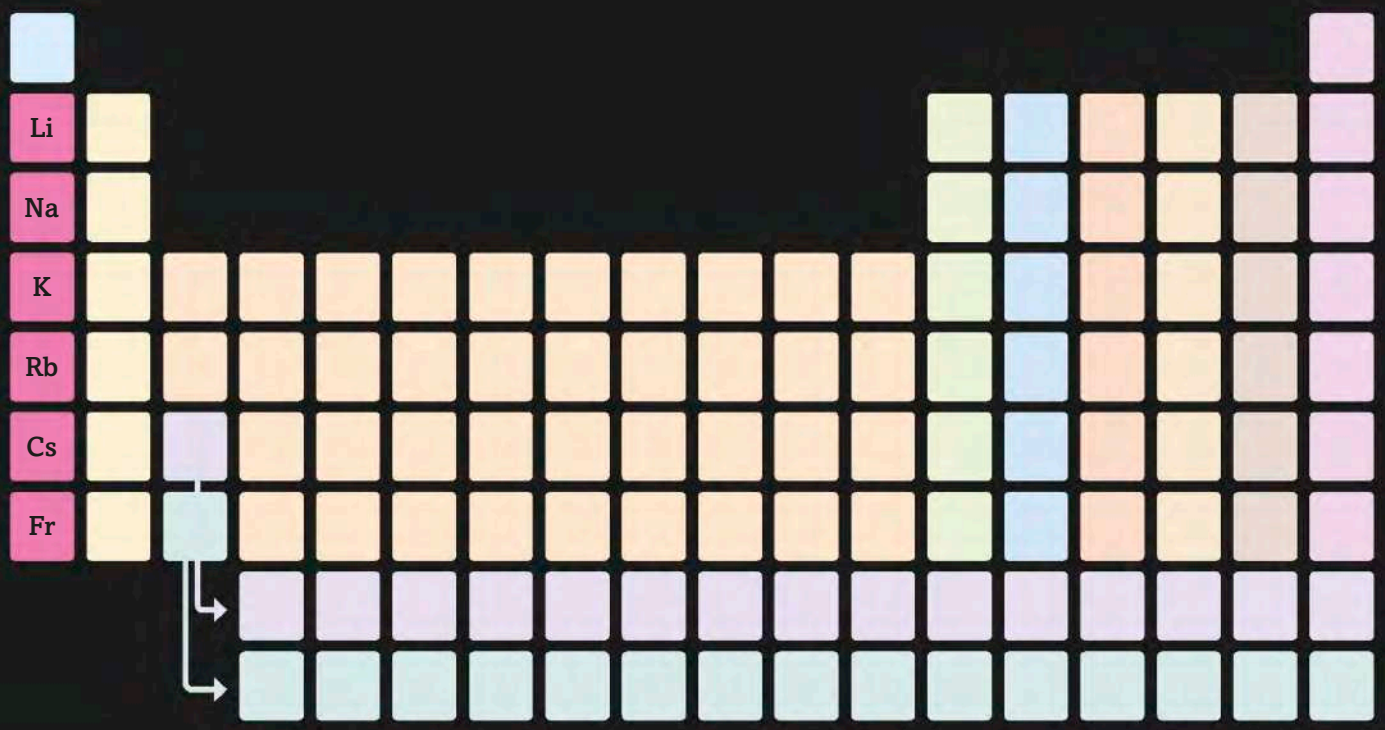

Alkali Metals

After hydrogen (H) – which is in a group of its own – the first column of the periodic table contains the alkali metals. This group gets its name from the way the elements react with water. These vigorous reactions always produce acid-attacking compounds called alkalis. None of the alkali metals are ever found in a pure form in nature. The first three metals are common in many minerals, while the last three are rarer.

Atomic structure
The atoms of all alkali metals have just one electron in their outer shell. Alkali metal atoms are among the biggest of all atoms.

Physical properties
These metals are soft enough to be cut with a knife. They are all silvery and very shiny when clean.

Chemical properties
Alkali metals are highly reactive. They form bonds with other elements, giving away their single outer electron.

Compounds
These metals react with water to form compounds called hydroxides. They react easily with halogens to form salts, such as sodium chloride.

3 Li Lithium

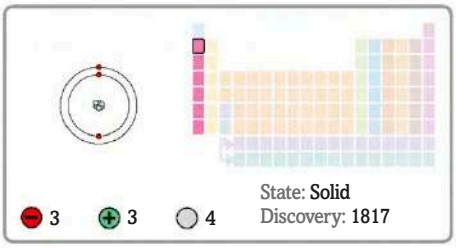

Forms

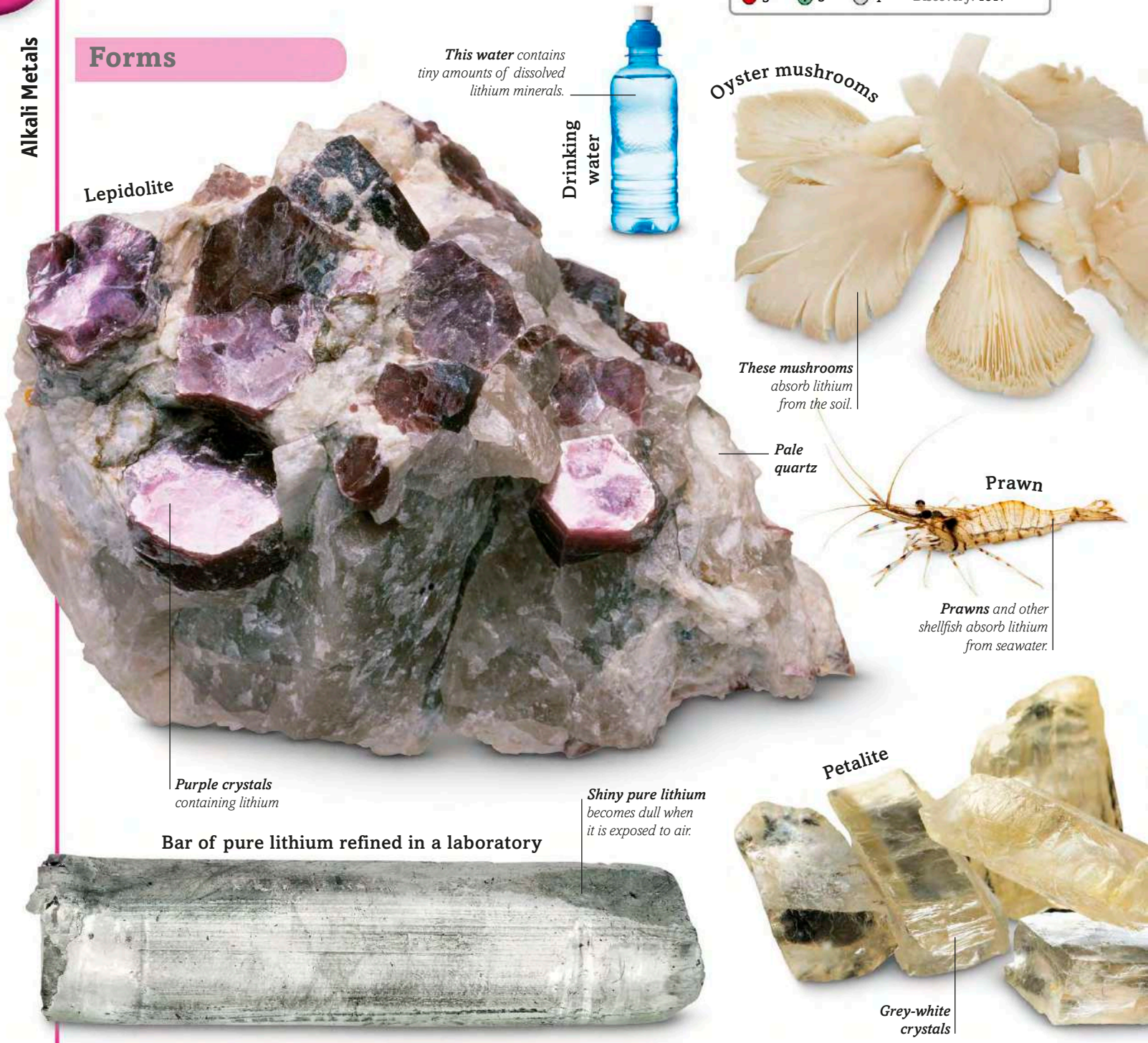

Lithium is the the lightest of all metals: in fact, it can easily float on water. **Pure lithium** is very reactive and exists in nature only in minerals, such as **lepidolite** and **petalite**. Many lithium minerals dissolve well in **water**, and the world's seawater contains millions of tonnes of dissolved lithium. Lithium is found in many foods, such as **mushrooms**, **prawns**, nuts, and seeds. It also has many everyday applications. Glass composed of lithium is resistant to heat and is used in scientific equipment, such as **mirrors inside**

Uses

Smartphones *run on rechargeable batteries that use lithium to store electricity.*

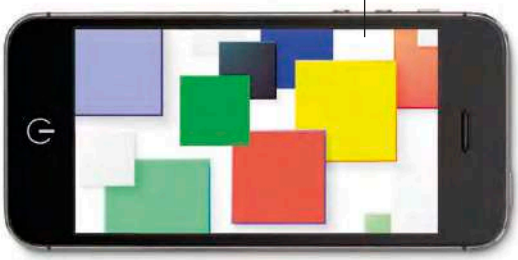

Smartphone

Hale telescope mirror

Lithium added to the glass in this mirror *stops the disc warping at extreme temperatures.*

LITHIUM-ION BATTERY

Lithium-ion batteries are widely used in digital devices. They store electrical energy to power gadgets and are rechargeable. This diagram shows a device's battery in use; when it is charging, this process is reversed.

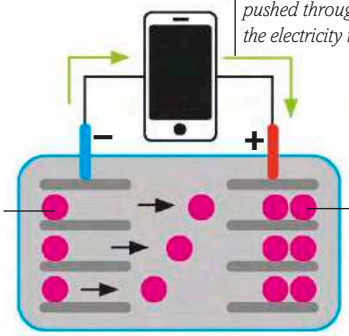

3*. As ions move inside the battery, negatively charged electrons are pushed through the phone, providing the electricity to make it work.*

1*. Inside the battery, positively charged lithium ions move from the negative electrode (-) to the positive electrode (+).*

2*. The positive electrode receives lithium ions as the battery loses charge.*

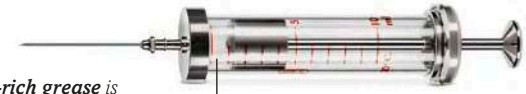

Syringe

Lithium coating on the inside of some syringes *delays the clotting of the blood sample.*

Lithium-rich grease *is used to keep mechanical parts of engines running smoothly, even when hot.*

Grease

Some artificial teeth *contain lithium discilate, which makes them strong.*

Artificial teeth

This air scrubber *used lithium hydroxide to purify the air inside the Apollo 13 spacecraft.*

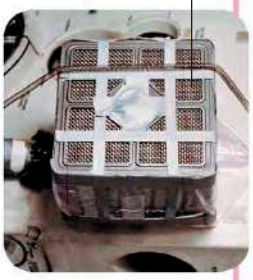

Air scrubber

This charging point *can recharge an electric car in one hour.*

This car runs for at least **64 km** (40 miles) on one charge of its lithium-ion battery.

Electric car

telescopes. The main use for lithium is in rechargeable batteries. Lithium-ion batteries are small but powerful, so they are ideal for **smartphones** and tablet computers. Larger lithium batteries can power **electric cars**, which are less polluting than petrol-powered ones. A soapy compound called lithium stearate is used to make **grease**, which helps automobile engines run smoothly. This element also forms hard ceramics that are used to produce strong **artificial teeth**. Lithium compounds are used in some medicines as well.

11 Na Sodium

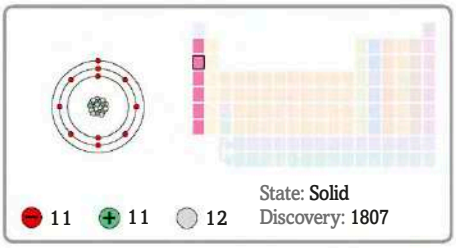

Forms

Clinoptilolite

This sodium-rich mineral *is an example of a zeolite, a mineral with tiny holes that can trap gases.*

Pure halite crystals

Cube-shaped transparent crystals

Sodalite cabochon

Polished gemstone *made of the mineral sodalite*

Soft, shiny metal

Laboratory sample of pure sodium in an airless vial

This glass case *holding pure sodium has no air in it, to prevent the metal from reacting with oxygen in air.*

The thick, white crust *covering this salt flat contains sodium chloride and other salts.*

Salar de Uyuni, Bolivia

Everyday salt contains lots of sodium. Although abundant on Earth, sodium is never found in its pure form naturally: it forms compounds with other elements. Sodium chloride, which also contains chlorine, is the most common sodium compound. It is also known as the mineral **halite**, and it is what makes seawater salty. Other sodium minerals include **sodalite**, a soft blue stone that can be shaped and polished. **Pure sodium** is soft enough to be cut with a knife. It reacts with oxygen in the air, forming a compound called sodium oxide, and bursts

Uses

Common salt

Edible salt *is made by refining the mineral halite.*

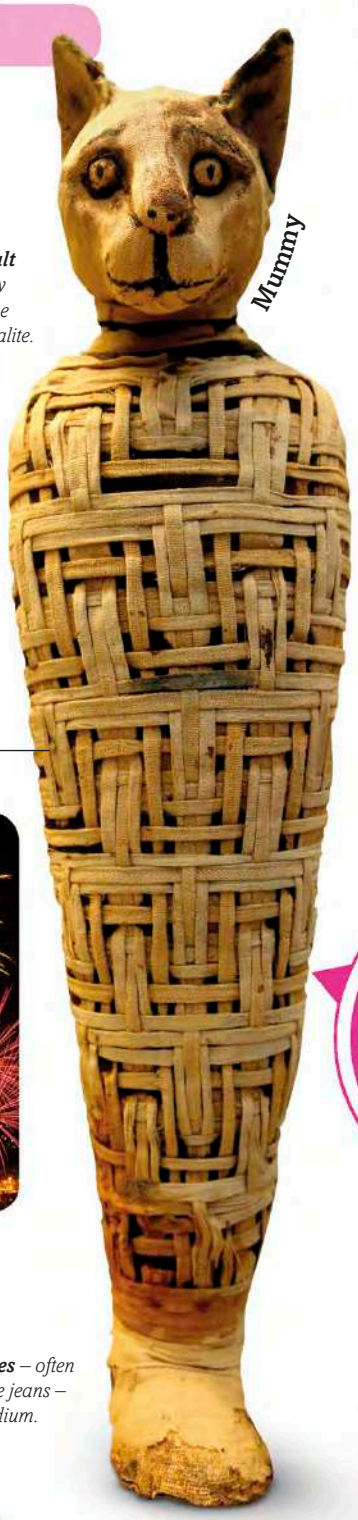

Mummy

This mummified body*, or mummy, was preserved using sodium compounds.*

Sodium fireworks

Bright yellow lights in fireworks *get their colour from burning sodium compounds.*

Indigo dyes *– often used in blue jeans – contain sodium.*

Indigo dye powder

Cats were sacred in ancient Egypt, so their bodies were mummified.

MUMMIFICATION

Ancient Egyptians believed in life after death and so preserved the bodies of their dead. A dead body was washed and the organs removed, then crystals of sodium compounds were used to dry it out. Finally, the body was wrapped, which completed the process of mummification.

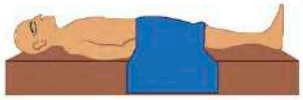

1*. Organs, such as the stomach and lungs, were removed from the dead body.*

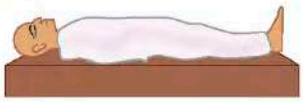

2*. Sodium compounds were spread over the body to dry it.*

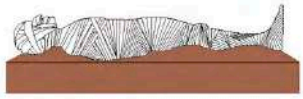

3*. The body was wrapped in cloth to mummify it.*

This tube *glows bright yellow-orange when sodium gas is electrified.*

Sodium gas lamp

Bar of soap

Some soaps *contain sodium hydroxide.*

Baking soda

Odourless white powder

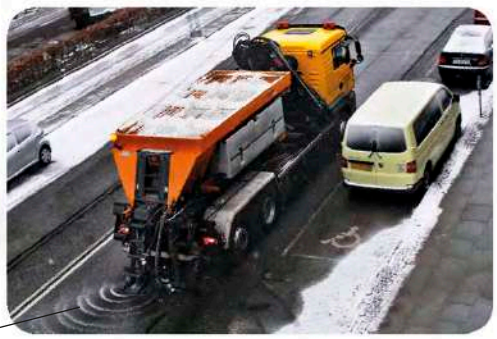

De-icing

Spraying salt *keeps roads free from ice and frost.*

into flames when in contact with water. Sodium compounds in **fireworks** burn with a yellow-orange colour. In ancient Egypt, crystals of sodium compounds were used to preserve dead bodies as **mummies**. Another useful compound is sodium bicarbonate, or **baking soda**, which makes dough rise by releasing bubbles of carbon dioxide. When refined, sodium chloride, or **common salt**, has several uses. It makes ice melt so it is used in salty grit added to slippery, frozen roads. This helps **de-ice** them to make them safer. It is also an important seasoning for meals.

SALT FLATS Hundreds of artificial ponds dot the hillside near the small town of Maras, high in the Andes of Peru. The ponds fill with water from a stream that runs down from the nearby mountains. In the sunshine, the water evaporates, leaving behind a thick salt crust that can be collected. The people of Maras have been gathering salt in this way for at least 500 years.

The salt forms part of rocks deep underground before it is dissolved by the stream and flows into the pools. Evaporation can also be used to collect salt from seawater or other salty water sources (known as brines). Today, however, most of the world's salt comes from underground mines containing thick layers of salt that are a result of ancient seas drying out. Over millions of years, that dry salt has become buried under dense layers of rocks. This so-called "rock salt" is sometimes unearthed using excavators. At other mines, it is washed out by piping in warm water, which dissolves the salt. The brine is then pumped up to the surface for evaporation.

19 K Potassium

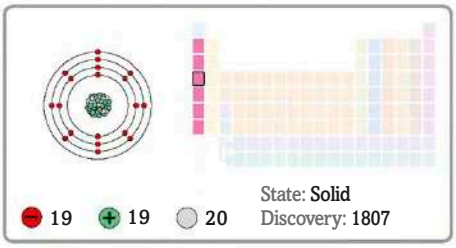

Forms

Potash

This mineral is rich in potassium chloride.

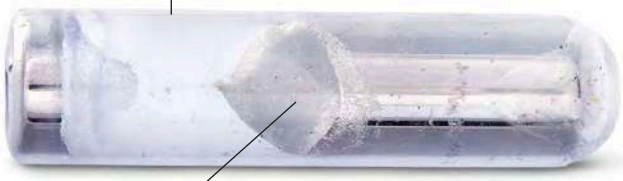

Laboratory sample of pure potassium in an airless vial

This glass case holding pure potassium has no air in it, preventing the metal from reacting with oxygen in air.

Soft and shiny solid

The yellow and green colour comes from impurities.

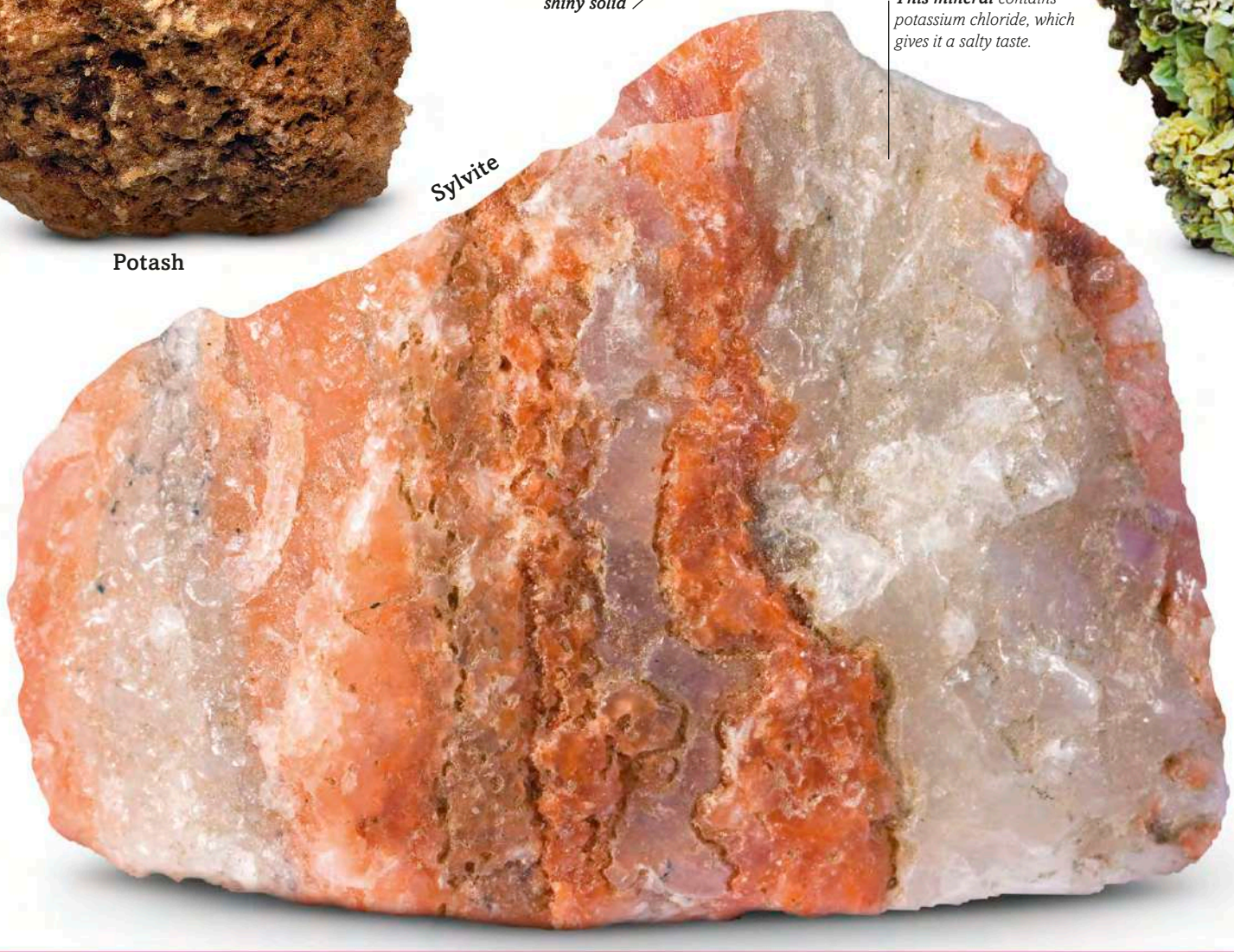

Sylvite

This mineral contains potassium chloride, which gives it a salty taste.

Potassium was first found in the dust of burnt plants. It was discovered by Sir Humphry Davy when he experimented with **potash** – a mixture of substances made from the ash of burnt plants soaked in water. The name potassium comes from potash but the element's chemical symbol, K, is taken from *kalium*, a Latin word for "ash". Potassium is never found pure in nature, but is present in minerals such as **aphthitalite** and **sylvite**. Potassium is vital for the human body, helping muscles and nerves work properly. For this, we rely on

Uses

Soda water *contains potassium compounds for added flavour.*

This salt *contains potassium chloride, which helps lower blood pressure.*

REBREATHER

A rebreather is a machine used by expert divers so they can stay underwater for long periods.

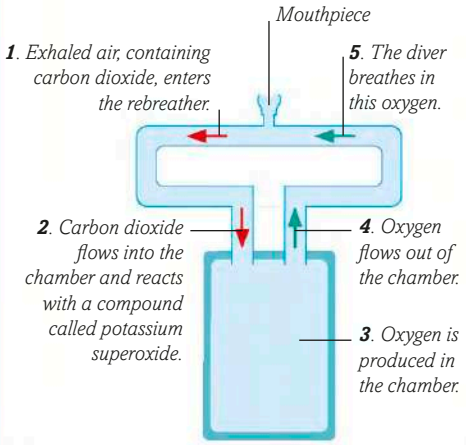

This cylinder *contains a compound called potassium superoxide.*

Potassium solutions *are used to hydrate patients.*

Saline drip

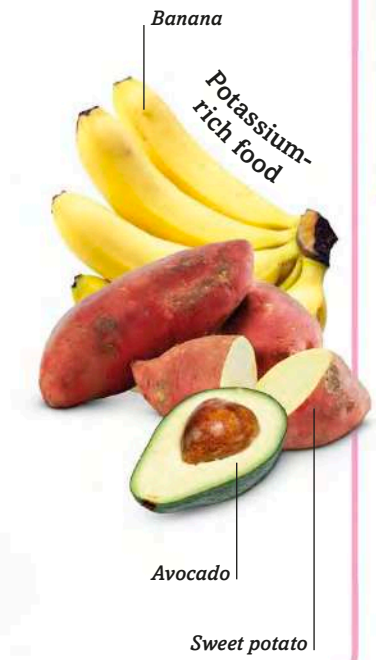

Potassium-rich fertilizer *is easily absorbed by the soil and boosts plant growth.*

This soap *contains potassium hydroxide, which is a cleaning agent.*

This explosive mixture *contains powdered potassium nitrate.*

This strengthened glass sheet *contains potassium nitrate.*

potassium-rich food, such as bananas, root vegetables, and avocados, which contain potassium chloride. In tiny amounts, this compound can enhance flavours, as it does in **soda water**. It is also a healthy alternative to sodium chloride, or common salt, and an important ingredient in **saline drips** for rehydrating patients who are seriously ill. Potassium nitrate is a compound of potassium, oxygen, and nitrogen, and is found in **gunpowder** and **toughened glass screens** for mobile phones.

37 Rb Rubidium

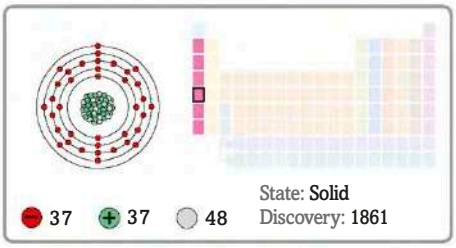

Forms

Laboratory sample of pure rubidium in an airless vial

This glass case contains ***pure rubidium****, preventing it from coming into contact with air and catching fire.*

Leucite

Pale, waxy mineral

Lepidolite

This soft mineral contains up to 3.5% rubidium.

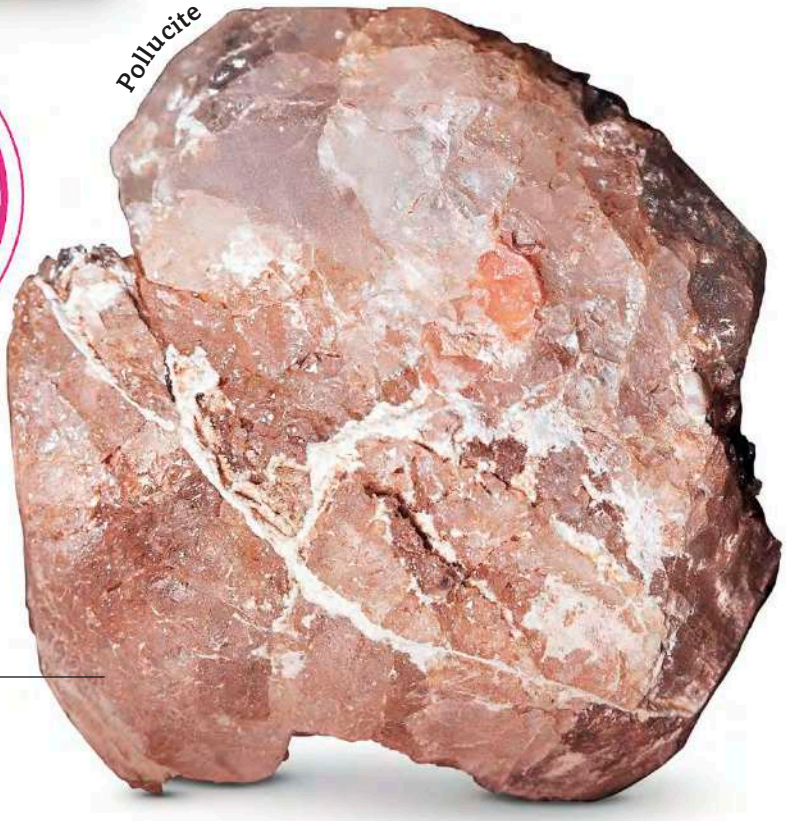

Pollucite

This ore contains caesium and rubidium.

Rubidium makes up only about **1 per cent** of this mineral.

Rubidium was named after the Latin word *rubidius*, meaning "deepest red". This refers to the red-coloured flame it produces when burned. This highly reactive element ignites on contact with air. On contact with water, it reacts vigorously, producing hydrogen gas and a lot of heat. Rubidium is not often concentrated in particular minerals, but instead is spread in small amounts through a wide range of minerals, such as **leucite** and **pollucite**. The pure metal is sourced mainly from the mineral **lepidolite**. Another mineral called rubicline has even more

Uses

These lenses *contain rubidium, which aids night vision.*

Night-vision goggles

RUBIDIUM-STRONTIUM DATING

About a quarter of all rubidium atoms are radioactive. Slowly over time, they break down into strontium atoms. Comparing the amounts of these elements in a rock shows when that mineral was formed. Older rocks have less rubidium and more strontium in them.

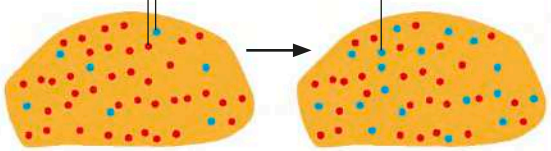

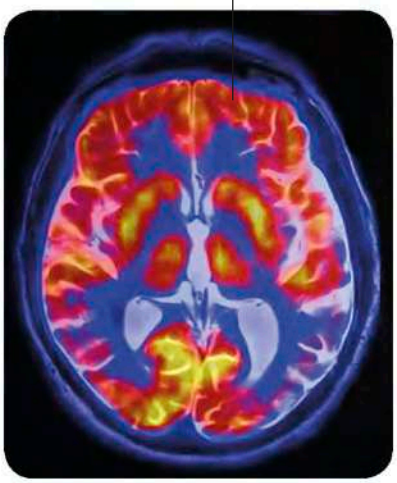

The structure of the brain *can be seen clearly because of the use of radioactive rubidium.*

PET scan

Photomultiplier

This sensitive device *detects light by using a rubidium compound.*

Fireworks

This purple colour *comes from burning a nitrogen rubidium compound.*

Electricity cables *are hung from these rubidium-rich insulators.*

Ceramic insulator

Magnetometer

This device *from the early 20th century used rubidium to measure the strength of magnetic fields.*

rubidium in it but is very rare. Rubidium atoms are sensitive to light and can be used in photoelectric cells (devices that convert light energy into electricity) and **night-vision equipment**. This element has radioactive forms, which can be used to measure the age of rocks. When injected into a patient's body, rubidium targets tumours, which show up clearly on **PET (positron emission tomography) scans**. Rubidium is also used by light-sensitive electronics called **photomultipliers**, and in making **insulators** for high-voltage cables and some special types of glass.

55 Cs Caesium

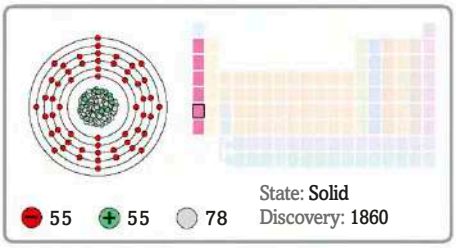

Forms

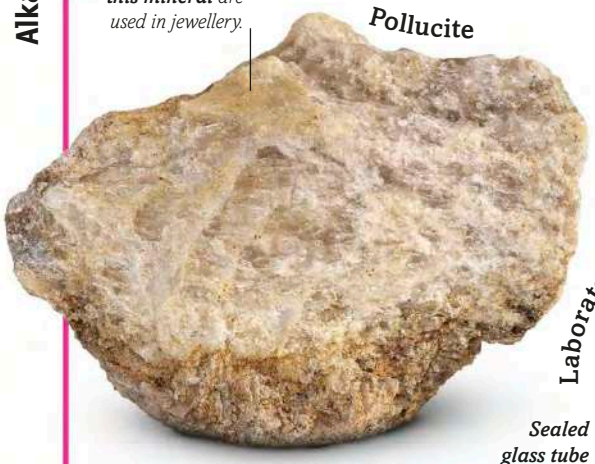

The crystals of this mineral are used in jewellery.

Pollucite

Shiny, silver-gold metal

Sealed glass tube

Laboratory sample of pure caesium in an airless vial

KIRCHHOFF AND BUNSEN

Caesium was discovered in 1860 by German scientists Robert Bunsen and Gustav Kirchhoff. They burned a sample of mineral water on a burner, which split the flame's light into individual colours. One of them was a distinctive light blue, which came from caesium.

Gustav Kirchhoff (left) and Robert Bunsen (right)

Uses

This highly accurate clock is also called a caesium clock.

Atomic clock

High-density caesium compounds in this fluid stop toxic gases rising to the surface.

Drilling fluid

As the most reactive metal on Earth, caesium explodes into flames if in contact with air or water. Therefore, **pure caesium**, is stored in a sealed glass tube from which all the air has been sucked out. This element is rare, and most of it is extracted from the mineral **pollucite**. Its name means "sky blue" and refers to the colour of caesium's flame when burning. Caesium is used in **atomic clocks**, which measure time down to a billionth of a second. These clocks are so accurate that they would gain or lose no more than one second every 300 years.

87
Fr

Francium

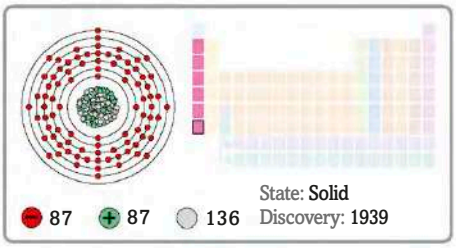

This mineral *was discovered in 1828 in Norway.*

MARGUERITE PEREY

The French chemist Marguerite Perey discovered francium in 1939 while studying the way a pure sample of another radioactive metal – actinium – decayed. She found that actinium broke down to form thorium and a previously unknown element. She named this element francium after her home country.

The dark crust *is a uranium mineral that holds tiny amounts of francium.*

Earth's rocks have **one** francium **atom** for every million trillion uranium atoms.

Francium is the rarest natural element on Earth. Scientists think there may be just 30 g (1.1 oz) of francium in Earth's rocks. Francium atoms are created when radioactive elements break down. Francium can be extracted from radioactive ores such as **thorite** and **uraninite**, both of which contain tiny amounts of this element. Even so, to date the largest sample of the metal made contained only 300,000 atoms, and lasted only a few days. Francium has no known uses outside of research.

Barium (Ba) crystals turn black in air.

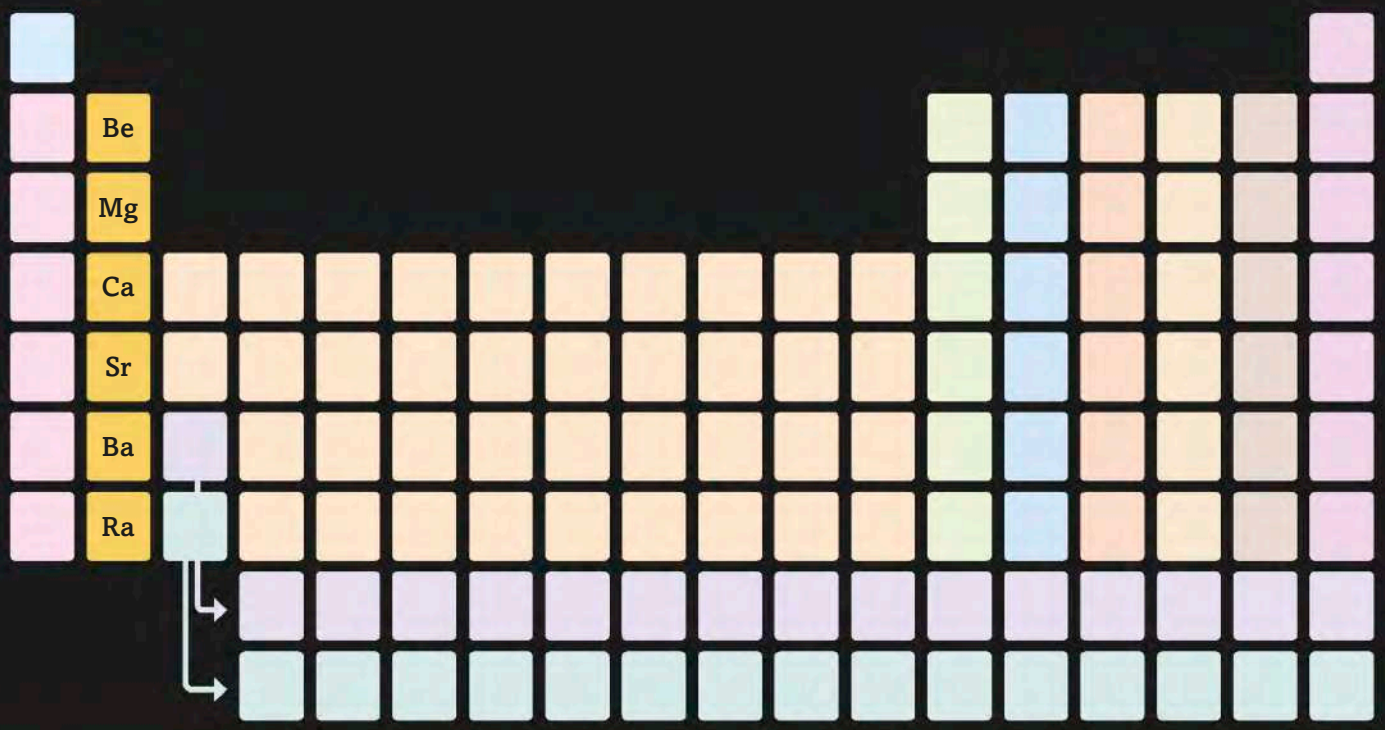

Alkaline Earth Metals

This group is a collection of reactive metals that were discovered as compounds inside common minerals in Earth's crust. Most of these minerals – known in the past as "earths" – are alkaline (alkali-producing), and this is how the group got its name. All alkaline earth metals were first purified in the 19th century.

Atomic structure
The alkaline earth metals have two electrons in their outermost electron shell. Radium (Ra) is the most radioactive member.

Physical properties
All members of this group are soft and shiny when pure. They are solid at room temperature.

Chemical properties
These metals are similar to the alkali metals, but not as reactive. Except for beryllium (Be), all alkaline earth metals react with hot water or steam.

Compounds
These elements form compounds with non-metals by losing their outermost electrons. Several compounds are found in teeth and bones.

4 Be Beryllium

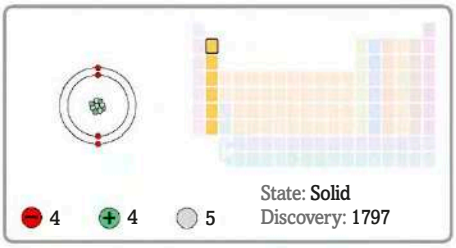

Forms

This mineral can also be brown, green, or orange.

Chrysoberyl

Aquamarine

These crystals have this pale blue colour due to iron impurities.

Laboratory sample of pure beryllium

Lightweight metal

Beryllium is found in more than **30 kinds of mineral**.

This widely used element gets its name from the Greek word *beryllos*, after which the mineral beryl is also named. Beryllium is the lightest of the alkaline earth metals, but it does not share many of the group's properties. For example, it does not react with water and is much harder than the other metals in its group. Two common beryllium minerals are **chrysoberyl** and beryl. Beryl has different forms, such as **aquamarine** and emerald. Beryllium is useful in many ways. For example, some military **helicopters** use windows made

Uses

This seal *made of a beryllium-nickel alloy is strong enough to stop high-pressure water supply leaking through.*

These beryllium mirror segments *will not contract in the cold of space.*

James Webb Space Telescope

This beryllium pipe *delivers beams of protons into this device.*

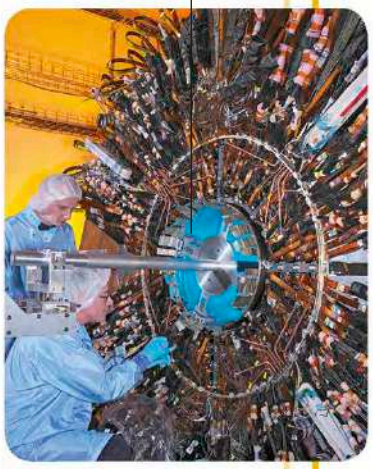

ATLAS, a device at the Large Hadron Collider, CERN, Switzerland

This airbag *is triggered by a sensor that contains beryllium.*

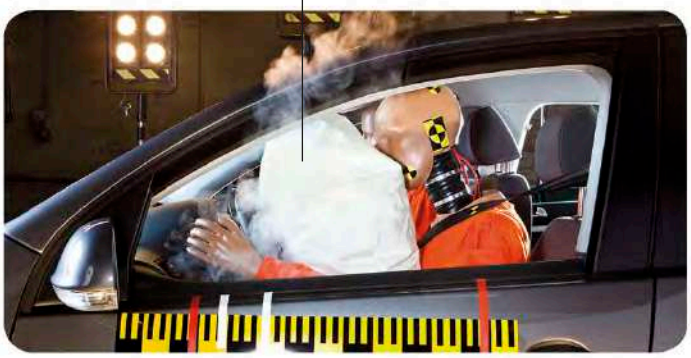

Airbag

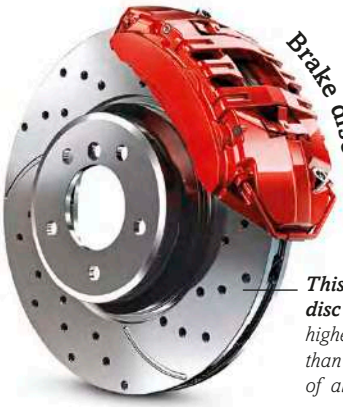

Brake disc

This beryllium disc *can withstand higher temperatures than a disc made of aluminium.*

LOUIS NICOLAS VAUQUELIN

Beryllium was discovered in 1798 by the French chemist Louis Nicolas Vauquelin. He extracted the pure metal from emerald, which is a valuable green form of the mineral beryl. He had already discovered the element chromium, which is also found in emerald and gives it its green colour.

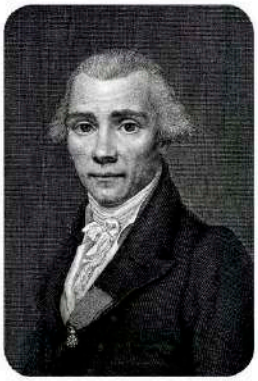

of beryllium-rich glass to shield optical sensors to aid pilots flying at night or through fog. Objects made of this metal keep their shape well and hardly expand or contract if the temperature changes. This makes beryllium useful in valves for **fire sprinklers** and car sensors that trigger **airbags**. NASA's **James Webb Space Telescope** will use a large beryllium mirror that is light and strong. Beryllium is also used to make **brake discs** for racing cars. Alloys of beryllium and copper are used in springs as well.

12 Mg Magnesium

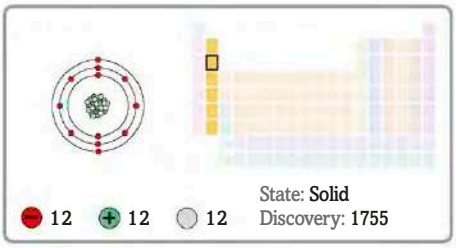

Forms

This green, magnesium-rich ***mineral*** *forms deep underground.*

Feather-like appearance

Shiny, grey crystallized form

Magnesium has **22 known isotopes**.

This ore *is a natural form of magnesium carbonate.*

Magnesium was named after Magnesia in Greece. This element largely exists deep inside Earth's mantle, but it can also be found in seawater and many minerals in our planet's crust, including **serpentine**. Another mineral, **dolomite**, is also a source of **pure magnesium**.

Magnesium has many important applications. Alloys of magnesium are not only strong, but also lightweight, so are used in a range of objects, from **car wheels** to **cameras**. For centuries, many naturally occurring magnesium minerals have been used in traditional medicines.

Uses

Alloy wheel

Magnesium alloy *makes this wheel strong and shiny.*

MAGNESIUM IN CHLOROPHYLL

Chlorophyll is an important molecule in plants and is what makes them green. At its centre sits a magnesium atom, which helps plants convert sunlight into energy in a process called photosynthesis.

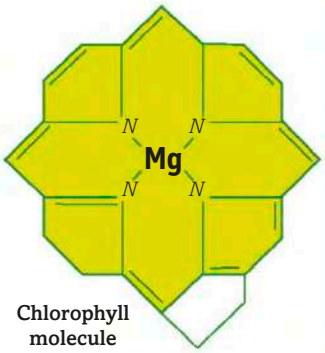

Chlorophyll molecule

Digital camera

The magnesium alloy body *of this camera is lightweight and will not rust.*

Epsom salts

Crystals containing magnesium sulfate *are added to warm water for a soothing bath.*

This powder *makes skin smooth and soft.*

Talcum powder

Milk of magnesia

This indigestion medicine *is a mixture of water and magnesium carbonate.*

White lights *from burning magnesium compounds*

Magnesium fireworks

Portland cement

This widely used cement *contains powdered magnesium oxide.*

The magnesium alloy case *of this laptop is strong but lightweight.*

Laptop

Magnesium carbonate, or **magnesia**, reacts with acid in the stomach to settle indigestion. Heating magnesia produces magnesium oxide, which is one of the ingredients in **cement**. Magnesium compounds are also used in **fireworks**, and they burn hot with a white flame. Salts composed of magnesium, called **Epsom salts**, named after the place in England where they were first mined, work as a muscle relaxant. Magnesium silicate, known as **talc**, is a soft mineral used in body powders.

20 Ca Calcium

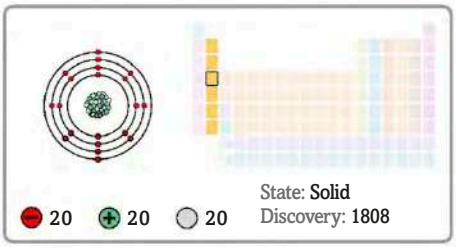

Forms

This crystal has a shiny surface.

Calcite

Large, fang-like crystals

These column-shaped crystals contain calcium carbonate.

Aragonite

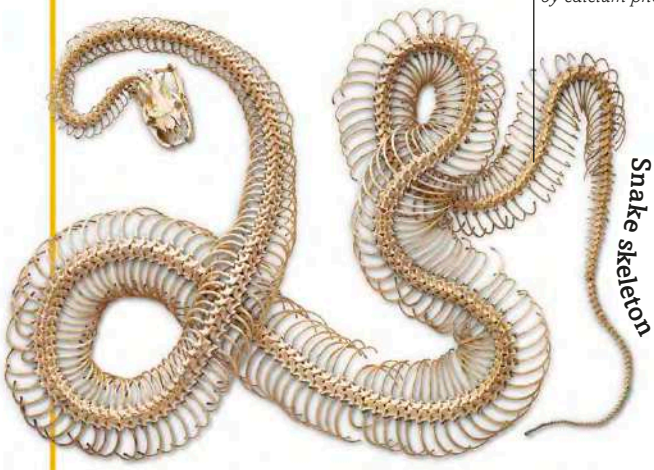

Bones are hardened by calcium phosphate.

Snake skeleton

This pure metal is soft enough to cut with a knife.

Crystals of pure calcium refined in a laboratory

The most abundant metal in the human body, calcium is also the fifth most common element on Earth. It appears in many minerals: **calcite** and **aragonite** are made of a compound of calcium and carbon called calcium carbonate. Bones in animal **skeletons** contain the compound calcium phosphate. The hard, outer layers of many other animals, such as the **shells** of sea snails, are made of calcium carbonate. Calcium is very important in our diet. We get calcium by eating **calcium-rich food**, including dairy products, green

Uses

Shells of sea snails *are hardened by calcium carbonate absorbed from sea water.*

This tablet *contains calcium carbonate, which is an alkali – a substance that balances out acids.*

An adult **human** contains about **1 kg (2 lb) of calcium** in the body.

CALCIUM CAVES

As running water flows into caves, it deposits calcium carbonate. These deposits build up to form structures called stalactites and stalagmites.

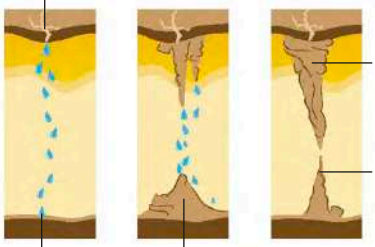

This plaster of Paris cast *hardens when dry, supporting broken bones.*

These chalks *contain calcium sulfate.*

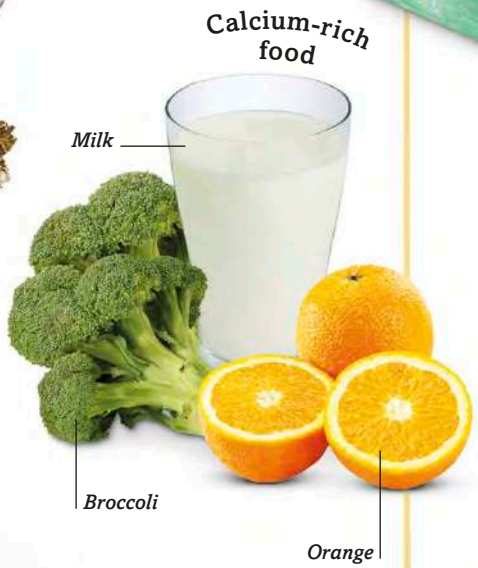

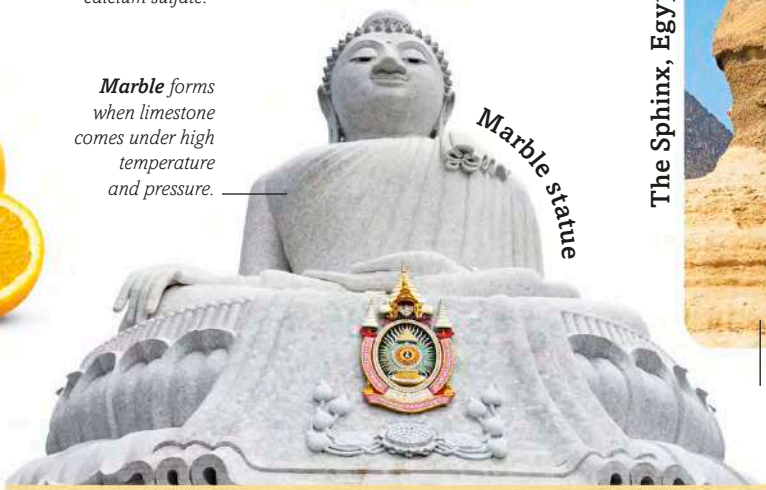

Marble *forms when limestone comes under high temperature and pressure.*

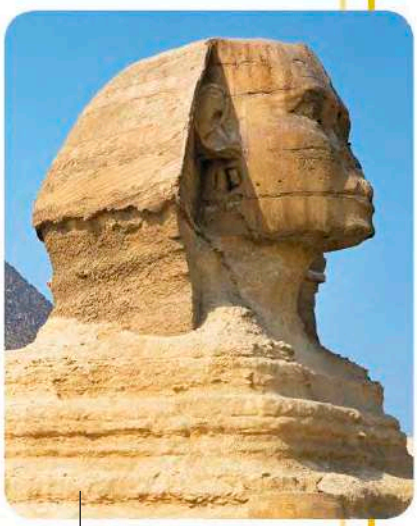

This statue *is made of limestone, a natural rock containing calcium carbonate.*

Oranges are also a good source of calcium, and most orange juices have extra calcium added to them. **Antacid tablets**, used to settle indigestion, contain calcium carbonate. This compound reacts with acid in the stomach. Calcium compounds are also common in construction materials. Plasterboard, which is used to make walls smooth, **writing chalk**, and **Plaster of Paris** are all made from the mineral gypsum. Calcium oxide is an important ingredient in cement and helps turn it into hard concrete.

FLY GEYSER The multicoloured Fly Geyser in the Black Rock Desert of Nevada, USA, is made from a mound of calcium carbonate rock. Such mounds and pools are made naturally in many other places where springs gush out warm, calcium-rich waters. The amazing colours of the rocks are caused by algae and bacteria that live in this water.

Fly Geyser is not a natural wonder. It was made by accident in 1964 when engineers were drilling a well to find a source of hot water. They did find a small reservoir of water that is heated by volcanic activity deep beneath the surface, but they chose to cap the well and look elsewhere. Eventually, the hot water broke through, creating a natural fountain, or geyser. Over the decades, the calcium deposits have slowly built up. The central mound is now 1.5 m (5 ft) tall and nearly 4 m (13 ft) wide. The scalding water that gushes out can reach a height of 1.5 m (5ft).

38
Sr

Strontium

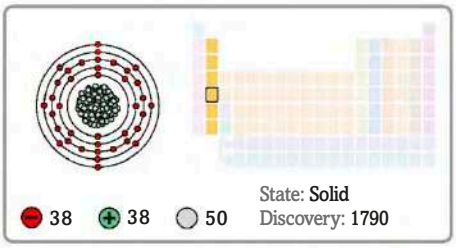

Forms

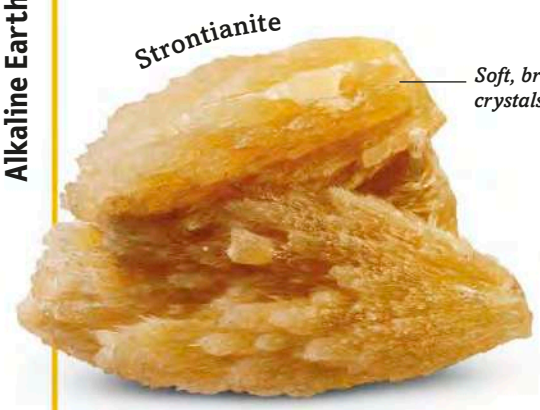

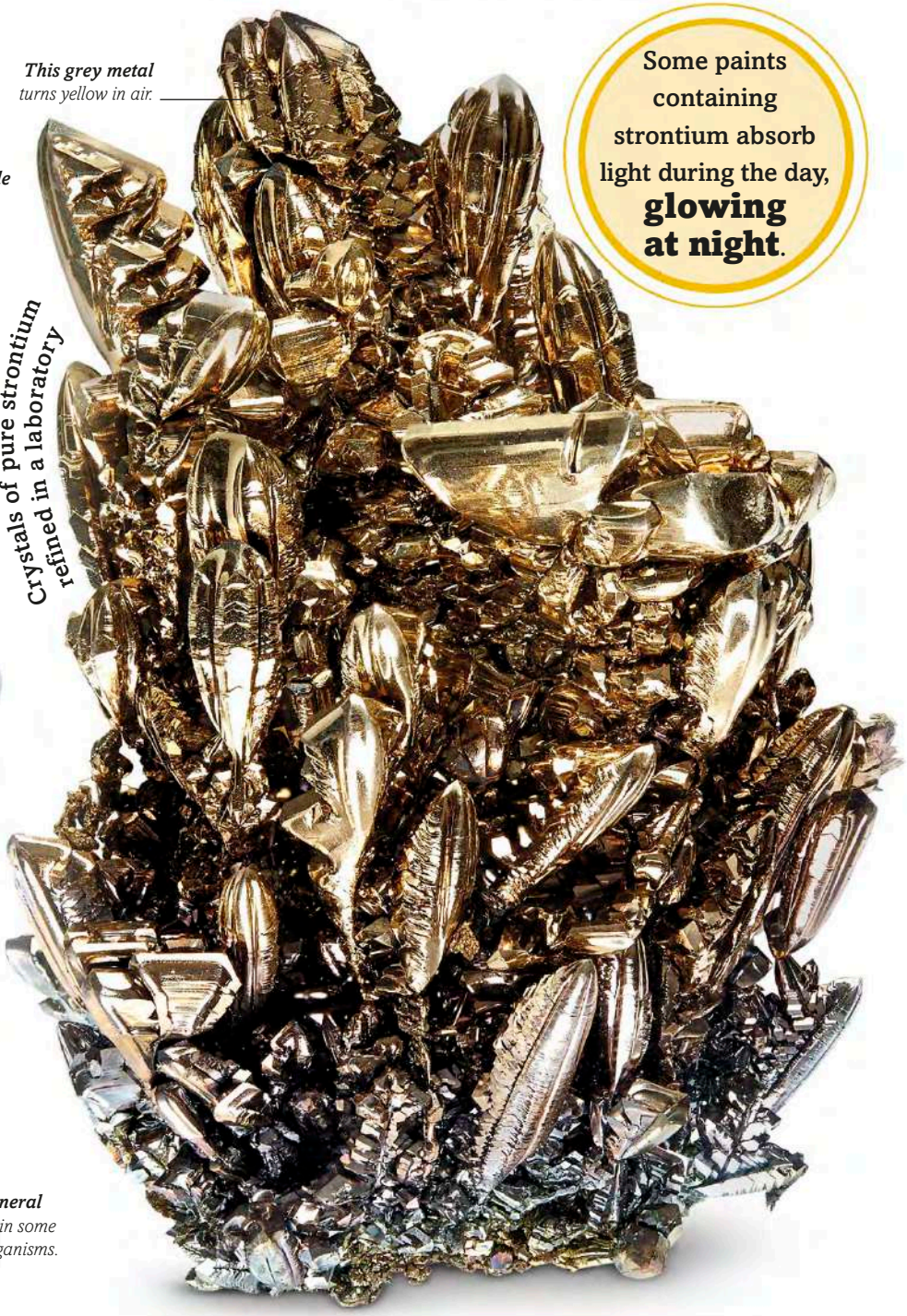

Some paints containing strontium absorb light during the day, **glowing at night**.

Strontium was discovered in 1791 in a mineral found near the Scottish village of Strontian. The mineral burned with a bright crimson flame, and Scottish chemist Thomas Charles Hope studied it and found that it contained a new element. This mineral was called **strontianite**, and it is the main ore of strontium. **Pure strontium** was first extracted by British chemist Humphry Davy in 1808, who conducted an experiment using electricity to obtain the element from the mineral. Strontium was once used in television screens, but today

Uses

Glazed ceramic

The bowl's smooth finish *is due to strontium oxide.*

Lights in unmanned buoys *can be powered by radioactive strontium.*

Navigation buoy

Loudspeaker

Strontium burns *in air with a bright red colour.*

Flare

Magnets inside this loudspeaker *contain strontium.*

Strontium compounds in some toothpastes *provide relief from pain.*

Toothpaste for sensitive teeth

Unmanned radar stations *run on electricity produced using a form of strontium called strontium-90.*

Weather radar station

GENERATING ELECTRICITY

A radioactive form of strontium, called an isotope, can be used to produce electricity. A radioisotopic thermoelectrical generator (RTG) converts heat from the element into electricity for use in spacecraft.

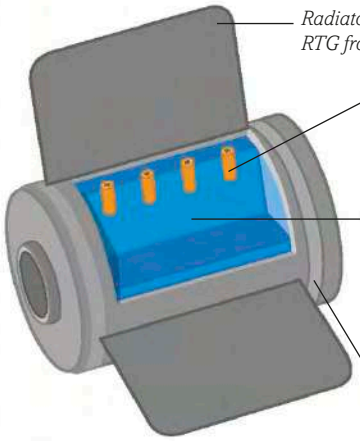

there are fewer uses for it. Strontium oxide in pottery and **ceramic** glazes creates distinctive colours, while strontium carbonate produces a red colour in **flares** and fireworks. Magnets that contain iron oxide can be made stronger by adding strontium to them. These strong magnets are used in **loudspeakers** and microwave ovens. Strontium chloride is added to some kinds of **toothpaste**, while radioactive strontium is a source of electricity for **radar stations** in remote places where there are no power lines or fuel supplies.

56 Ba Barium

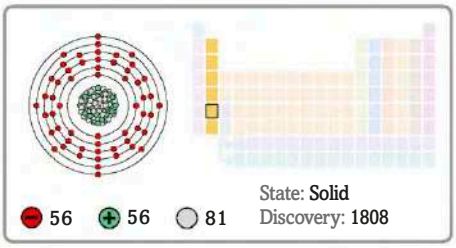

Forms

Farmers used witherite as **rat poison** until the 18th century.

This mineral *is used to make ceramic glazes.*

This soft metal *has a light, golden shine.*

A grey layer of tarnish *forms when the metal comes into contact with air.*

Petal-like shapes *form in the desert when sand mixes with barite or with gypsum.*

These glassy, blue crystals *contain barium and titanium.*

Barium is named after the Greek word *barys*, which means "heavy", because barium and its minerals are dense. The **pure form** of this element was first discovered in 1808 by the English chemist Humphry Davy, who extracted it from an oxide of barium. This does not exist in nature. Davy obtained it by heating the mineral **witherite**. Today, the main source of barium is barite, a mineral of sulfur that forms in deserts and in rock deposits that come into contact with hot water. A rarer mineral called **benitoite** also contains barium. The

Uses

Laboratory sample of pure barium crystals

Spark plug

This plug *contains an alloy of barium and nickel.*

Glass-making

This glass *can be made shinier by adding barium oxide and barium carbonate.*

Jasperware pot

This pot *is made from clay that is rich in barium.*

Vacuum tube

The barium in the metallic strip *absorbs gases in the tube, maintaining a vacuum.*

X-ray scan

The intestine *is filled with a barium solution.*

BARIUM SOLUTION

Barium is used in a medical test for checking a patient's digestive tract for problems. In this test, a patient swallows a barium liquid solution, which fills the organs in the digestive tract.

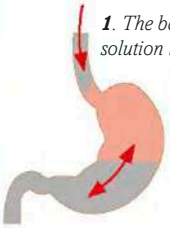

1*. The barium solution is ingested.*

2*. The solution enters the stomach and begins to fill it.*

3*. Under an X-ray scan, the barium-filled stomach shows up clearly.*

element is used in **spark plugs** to make them produce more powerful sparks and is added to **glass** to increase its shine. Barium compounds are added to some types of clay used for making **pots** and vases. In oil wells, barium compounds are added to drilling fluids to increase their density. Doctors make use of barium's density by giving patients a solution of barium compound to swallow, before taking **X-rays** of their digestive system. The barium makes the soft digestive organs denser, allowing them to be seen clearly with an X-ray machine.

88
Ra

Radium

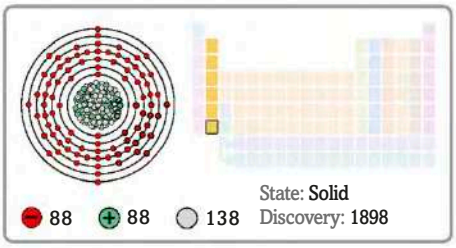

Forms

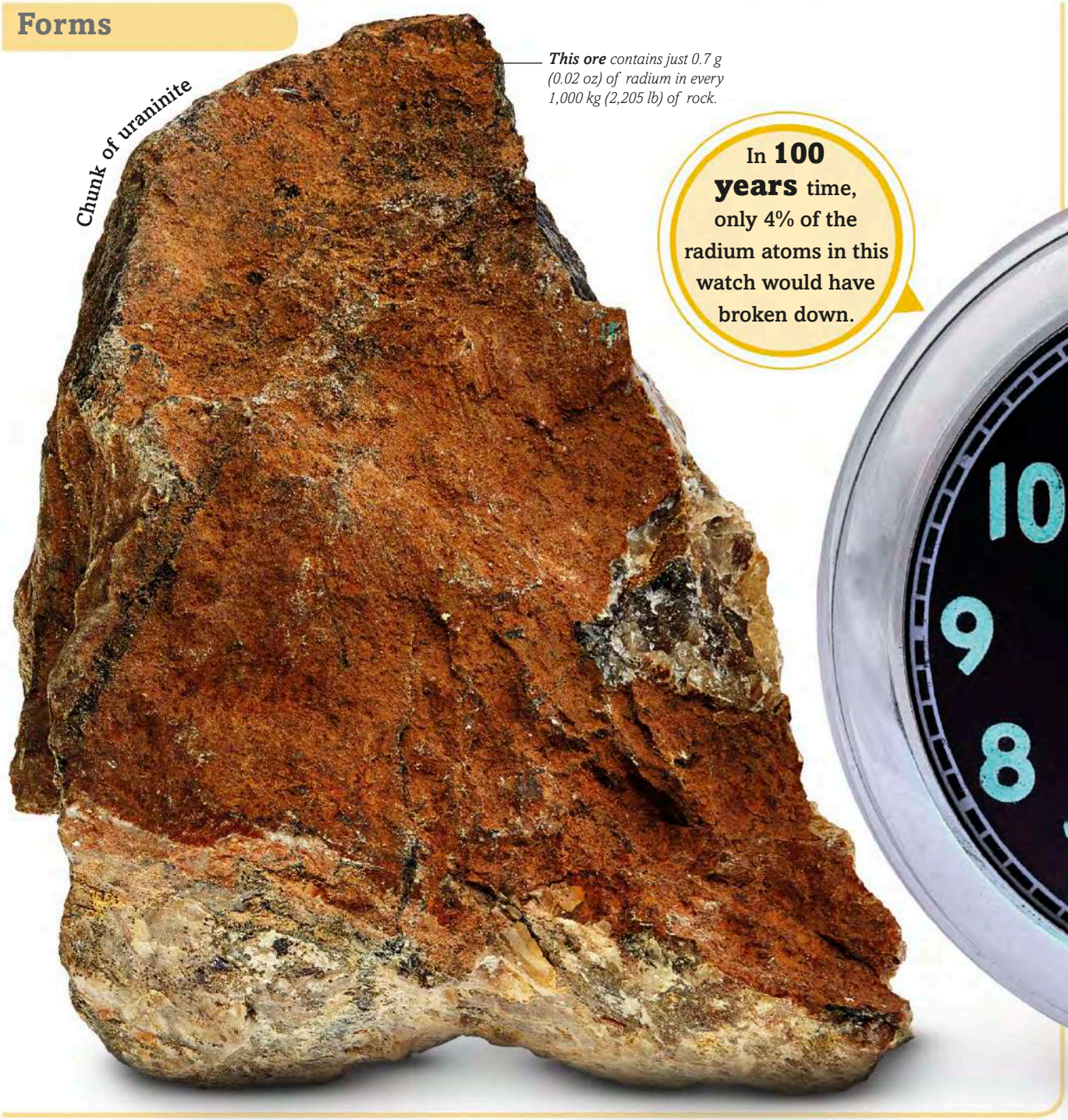

Radium is the only radioactive member of the alkaline earth metals. It is also the rarest element in this group, and forms in small amounts when the atoms of more common metals – such as uranium and thorium – break down. Radium atoms do not survive for long, with most of them quickly decaying into radon, a radioactive noble gas. This element is highly dangerous and is rarely used today. However, in the early 20th century, radium compounds were in common use. Luminous paints, like those used to make **watch dials**

Uses

PIERRE AND MARIE CURIE

Radium was discovered in 1898 by Marie and Pierre Curie. They found that uranium ores produced more radioactivity than expected from samples of uranium. They realized another radioactive metal was present and named it radium.

The radium paint *in this clock makes the numbers glow green-blue in the dark.*

Pocket watch with a luminous dial

Vials for radium treatment

This vial *contains a liquid called radium chloride.*

Radium emanator

This machine *from the early 20th century mixed radium into water, which was thought to make it healthier to drink.*

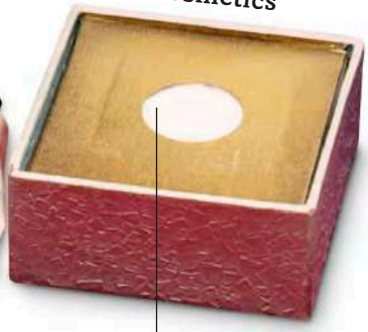

Cosmetics

Skin lotions *containing radium were common in the 1920s.*

Radium face powder *was once thought to be good for the skin.*

glow in the dark, were created using radium. People working with this paint often became ill, especially with cancer, because the radiation produced by radium damages DNA. Nevertheless, until the 1940s, many people thought radium's radioactivity made them stronger, not weaker. They injected themselves with **vials containing a radium compound**, believing it gave them an energy boost. They also thought that creams and **cosmetics** with radium in them made the skin healthier, even though they did exactly the opposite.

Oddly shaped piece of pure cobalt (Co).

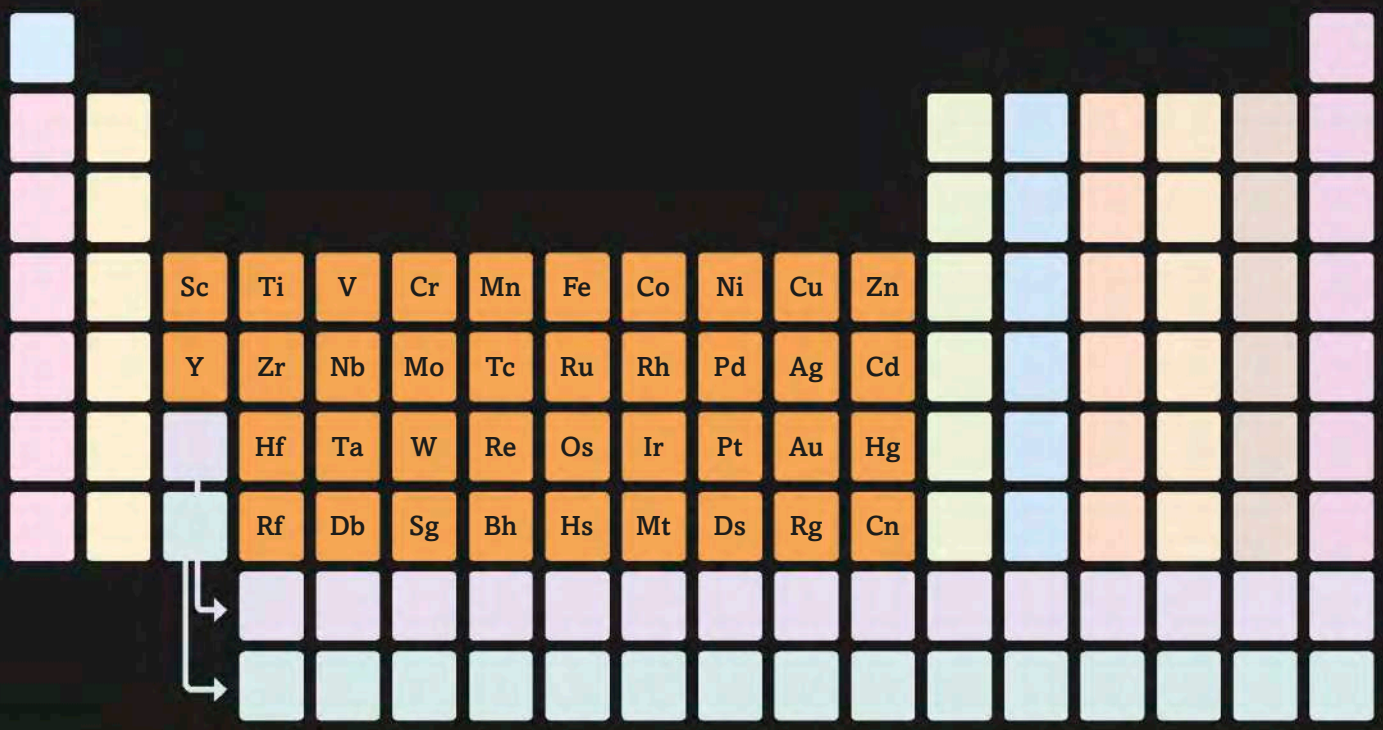

Transition Metals

This is the largest set of elements in the periodic table. This block of metals contains useful elements, such as gold (Au), iron (Fe), and copper (Cu). Many of these metals are easy to shape. The fourth period of the block – from rutherfordium (Rf) to copernicium (Cn) – are artificial and do not occur in nature. They were created by scientists in laboratories.

Atomic structure
Most transition metals have two outer electrons, but a few, such as copper (Cu), have just one.

Physical properties
These elements are generally hard and dense metals. Mercury (Hg), the only element that is liquid at room temperature, also belongs to this group.

Chemical properties
Transition metals are not as reactive as alkali and alkaline earth metals. However, they form many varied and colourful compounds.

Compounds
Many compounds of transition metals are brightly coloured. These metals are often used in alloys, such as brass and steel.

21 Sc Scandium

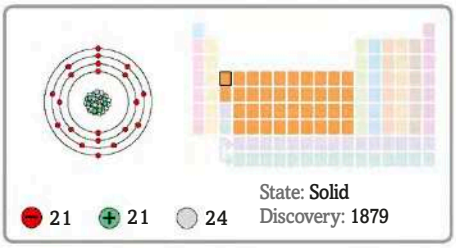

Forms

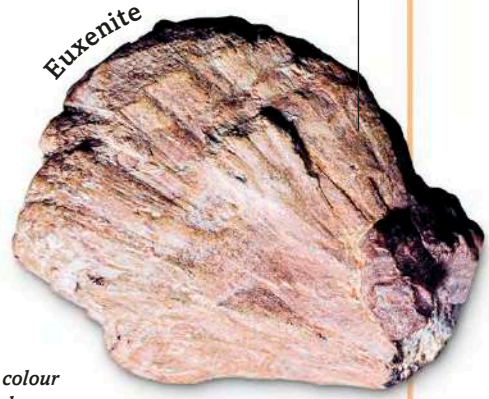

Uses

A soft and lightweight metal, scandium is similar to aluminium. It is spread so thinly in Earth's rocks that it is very difficult to collect a large amount of this element. Scandium is only used for specialist applications. Its main ores are the minerals **gadolinite** and **euxenite**, which also contain small amounts of many other rare metals, such as cerium and yttrium. Scandium mixed with aluminium makes a strong alloy, which is used in lightweight equipment for sports, such as **lacrosse**, and to make high-speed jets, such as the **MiG-29**.

22 Ti Titanium

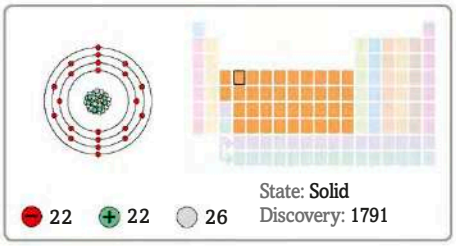

Forms

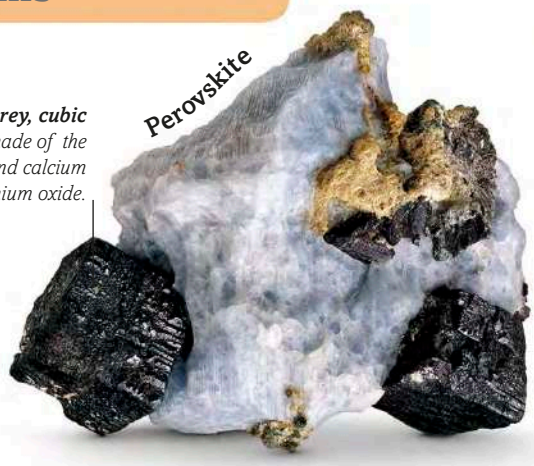

This grey, cubic crystal *is made of the compound calcium titanium oxide.*

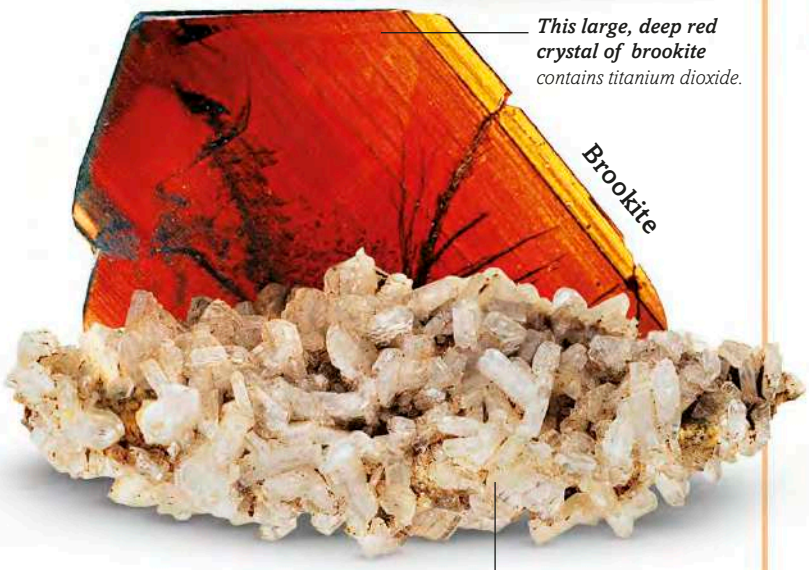

This large, deep red crystal of brookite *contains titanium dioxide.*

These are crystals *of the mineral albite.*

This metal's shine fades to grey *when exposed to air.*

Laboratory sample of pure titanium

Uses

The titanium plate *placed inside acts as as shield.*

Body armour

This titanium joint *can replace damaged bone in body.*

The titanium dioxide *in this sunscreen blocks damaging ultraviolet (UV) radiation in sunlight.*

Drill bit

This drill bit *is hardened with a coating of titanium nitride.*

This watch casing *is made of a titanium alloy.*

This titanium frame *is lightweight but strong.*

Named after the Titans, a race of mythic Greek gods, titanium is a silvery metal. It is as strong as steel but much lighter, and it is not corroded by water or chemicals. This strong metal also makes excellent protective shields in **body armour**. Titanium is commonly used to prepare titanium dioxide, a compound of titanium and oxygen that is used in paints and **sunscreen**. Titanium is not toxic so it can be used to make medical implants, such as artificial **hip joints**. **Wristwatches** made with titanium alloys are light and strong.

23 V Vanadium

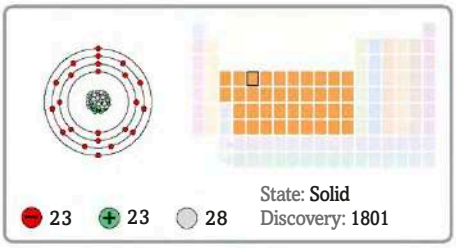

Forms

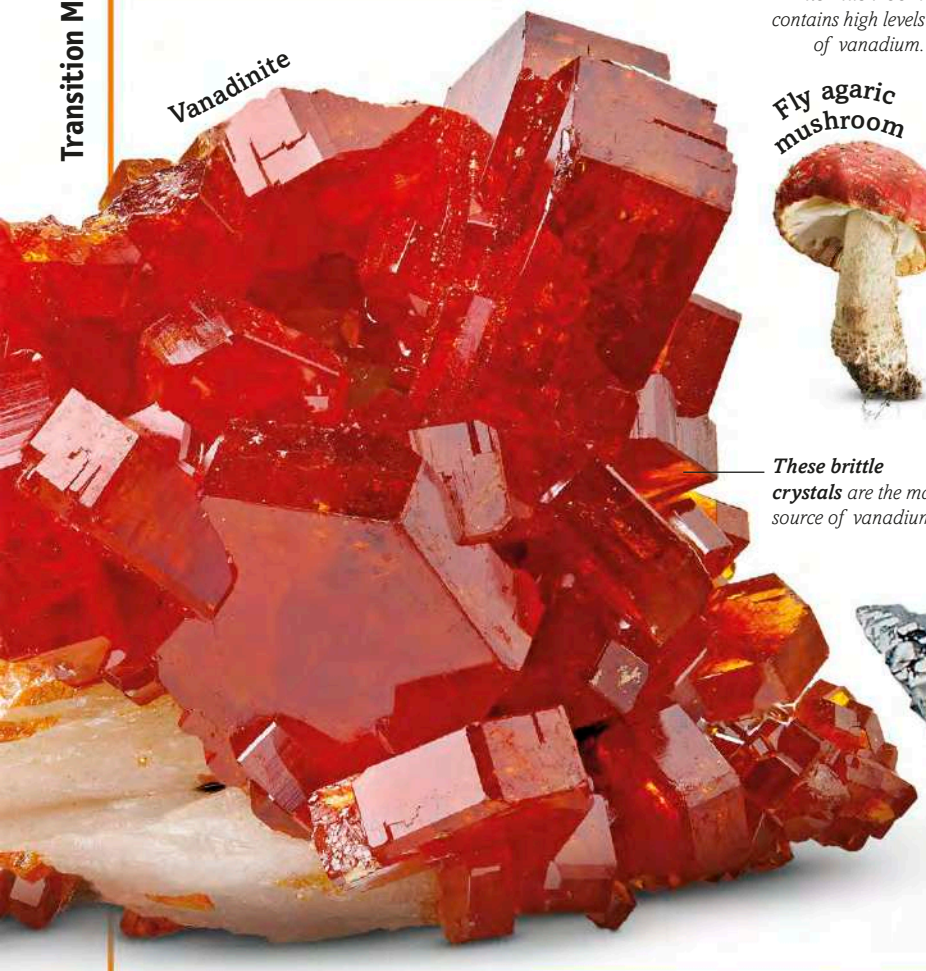

This mushroom *contains high levels of vanadium.*

These brittle crystals *are the main source of vanadium.*

This powdery yellow crust *contains tiny amounts of vanadium.*

Silvery surface

Uses

Tools *made with alloys of vanadium and steel are durable.*

About **85% of all vanadium** is used to toughen steel.

This knife *has been strengthened by the addition of vanadium.*

Vanadium can be hammered and stretched without breaking. This hard, strong metal is easy to shape. **Vanadium** was first **purified** in 1869 by the British chemist Henry Roscoe. Today, it is commonly extracted from its ore **vanadinite**. Ancient metalworkers used tiny amounts of vanadium compounds to make a very tough substance called **Damascus Steel**. This was named after the capital city of Syria, where ironworkers made the world's sharpest swords. Vanadium is still used to toughen tools, such as **spanners** and knives.

24 Cr Chromium

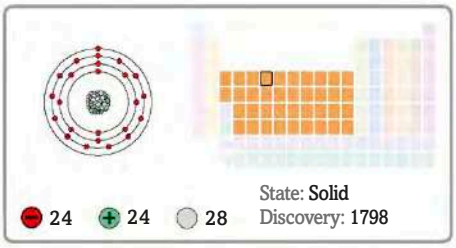

Forms

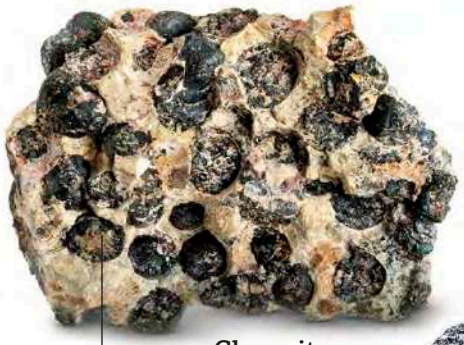

Chromite

***Chromite** is dark grey to brown in colour.*

Crocoite

***These large red crystals** contain chromium and lead.*

Laboratory sample of pure chromium

***This metal stays shiny** even when exposed to air and water.*

Uses

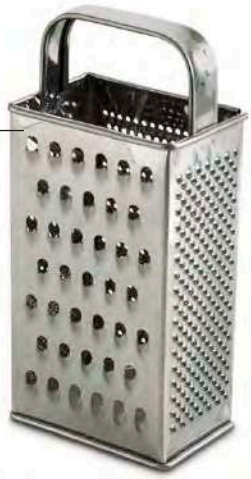

Stainless steel kitchenware

***This grater** can resist corrosion because it contains chromium.*

Ruby

***This red colour** is due to the presence of tiny amounts of chromium oxide in the crystal.*

Motorcycle

***The chromium plating** protects this motorbike from rusting.*

Chromium is named after *chroma*, the Greek word for "colour". Many minerals of chromium, including **chromite** and **crocoite**, are brightly coloured. An artificial form of crocoite, known as "chrome yellow", was once used in paints, but it was banned when scientists discovered it to be poisonous. **Pure chromium** doesn't corrode easily, so it is combined with iron and carbon to produce **stainless steel**. Chromium also gives gemstones, such as **rubies**, their deep-red colour. Some **motorcycles** have chromium-plated bodywork, giving them a shiny finish.

25 Mn Manganese

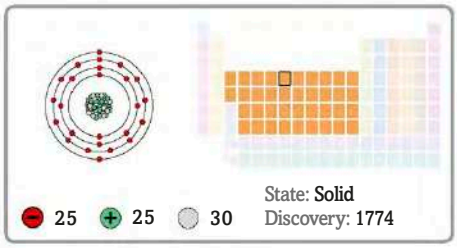

Forms

Manganese was **purified** from **pyrolusite** for the first time in 1774.

Like magnesium, this element gets its name from the Greek region of Magnesia. There are many manganese minerals, including the colourful mineral **rhodochrosite**. The pure form of the metal is obtained mainly from the ore **pyrolusite**. **Pure manganese** is dense, hard, and brittle. This element is present in seawater as the compounds manganese hydroxide and manganese oxide, which have built up in layers over millions of years to form masses on the sea bed. The human body needs a tiny amount of manganese, which we can get

Manganese-rich food

Laboratory sample of pure manganese

Uses

This US coin *from World War II was made using manganese and silver because nickel was in short supply.*

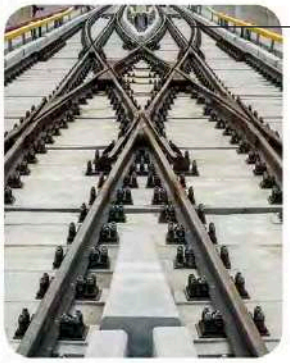

Railway tracks

These steel tracks *have manganese added to them to make them stronger.*

Dry cell battery

This battery *contains manganese dioxide.*

Unleaded petrol

This petrol *contains a manganese compound, which is less toxic than lead.*

Lascaux cave paintings, France

The black colour *comes from manganese dioxide.*

Purple glass bottle

This glass is coloured *by adding a manganese compound called permanganate.*

JOHAN GOTTLIEB GAHN

In 1774, Swedish chemist Johan Gottlieb Gahn discovered manganese by reacting manganese dioxide with charcoal – which contains carbon – under a lot of heat. The carbon took the oxygen away from the compound, leaving behind pure manganese.

from mussels, nuts, oats, and pineapples. The applications of manganese include its use in strengthening steel, which is used in making **railway tracks** and tank armour. Certain **dry cell batteries** carry a mixture containing manganese oxide. Manganese compounds are also added to **petrol** and used to clean impurities from **glass** to make it clear or to give it a purple colour. In prehistoric times, the compound manganese dioxide was crushed to make the dark colours used in **cave paintings**.

26 Fe Iron

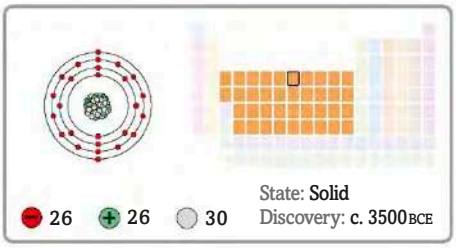

Forms

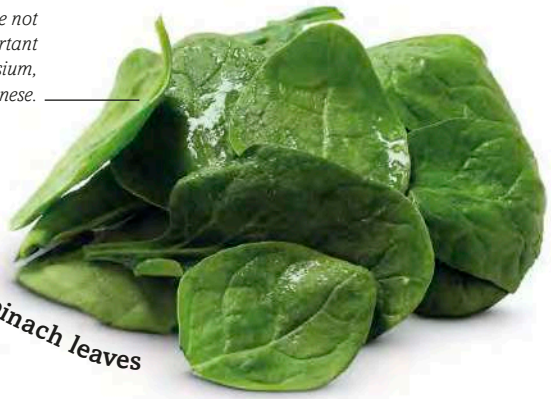

Iron is the **most common metal** on our planet.

Most of the iron on our planet is locked away in Earth's hot, molten core. This element is widely found in rocks worldwide, and almost 2.5 billion tonnes of iron is purified every year. Mineral ores rich in iron include **pyrite**. Other ores, including haematite, are used to extract **pure iron** in a process called smelting. **Iron-rich meteorites** – chunks of rock from outer space that crash to Earth – are one of very few sources of naturally pure iron. The human body uses iron to make haemoglobin, a substance in blood that carries oxygen around our body (oxygen helps

Uses

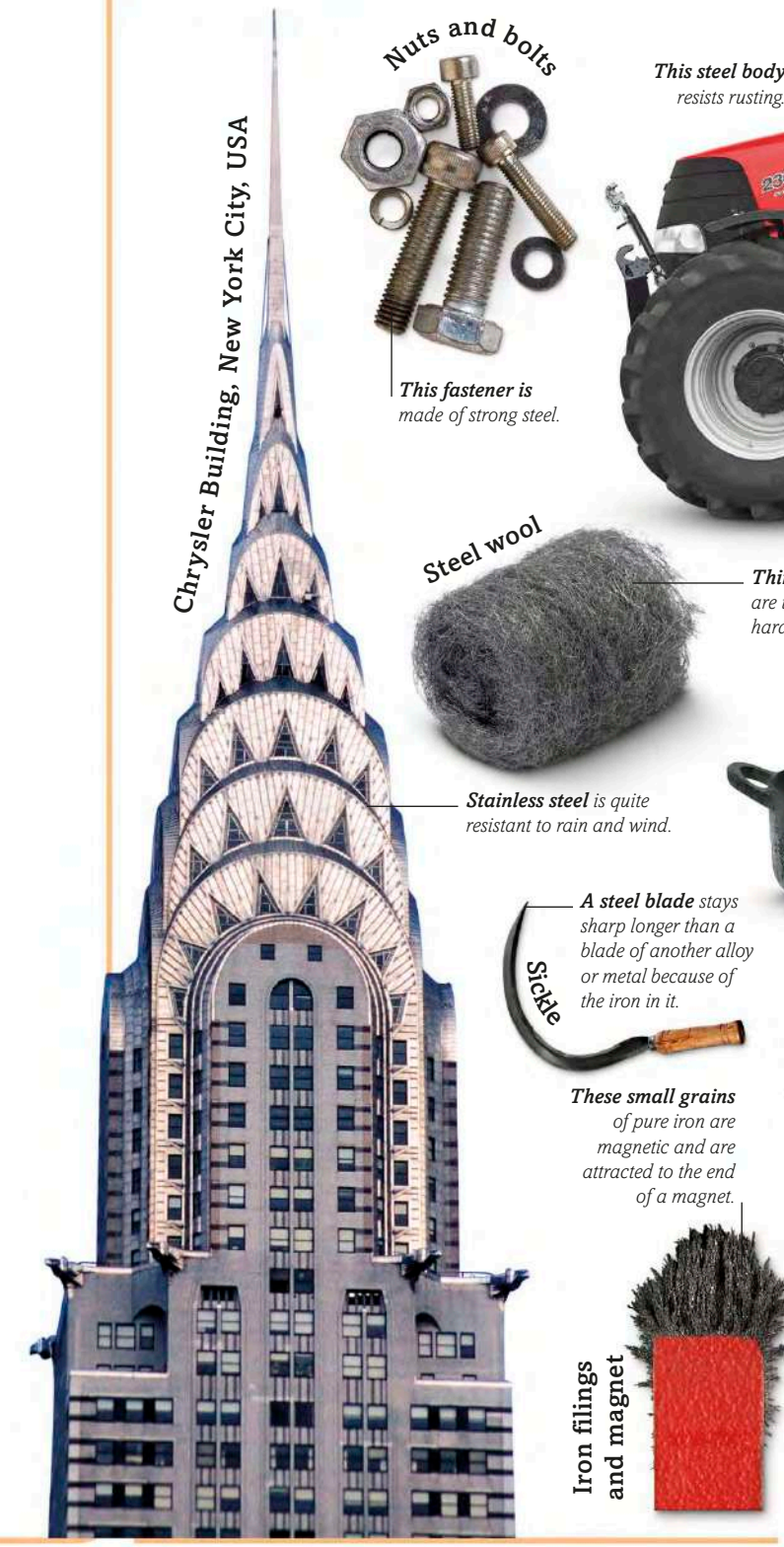

Transmission towers

SMELTING

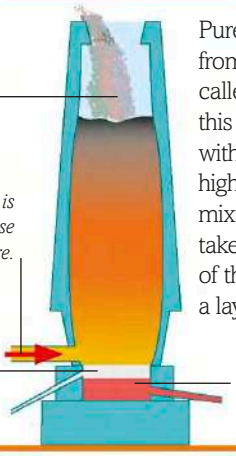

Pure iron is separated from its ores in a process called smelting. During this process, iron reacts with carbon in coal at a high temperature. As the mixture burns, the carbon takes the impurities out of the ore, leaving behind a layer of pure iron.

our cells produce energy for the body to work). Foods containing iron include meats and green vegetables, such as **spinach**. When pure iron comes into contact with air and water, it develops a flaky, reddish-brown coating called rust, which weakens the metal. In order to make iron tougher, tiny amounts of carbon and other metals, such as nickel and titanium, are added to it. This forms an alloy called steel, which is used to make **bolts** and strong **tractor** bodies, among other applications. Adding the element chromium to steel creates a stronger alloy called stainless steel.

STEELMAKING A stream of red-hot, liquid metal pours from a furnace at a steelmaking workshop. This is the end of a long process in which iron ore is transformed into steel, a tough alloy that is strong enough to make girders for supporting skyscrapers and bridges. The steel may even be moulded into car bodies, woven into superstrong cables for elevators, or turned into powerful magnets that can levitate maglev trains.

Steel is an alloy of iron that contains about two per cent carbon and some other elements. The carbon locks all the atoms together and prevents the metal from cracking. This makes steel harder than iron: it bends before it breaks and doesn't shatter easily. To make steel, iron ore is smelted in a blast furnace to remove its impurities, such as nitrogen, sulfur, or phosphorus. Other elements can be added to create different varieties of steel. For example, chromium in steel stops it from rusting, while manganese makes it harder. Adding silicon to steel can make the alloy more magnetic, while nickel makes it less brittle at extremely low temperatures.

27 Co Cobalt

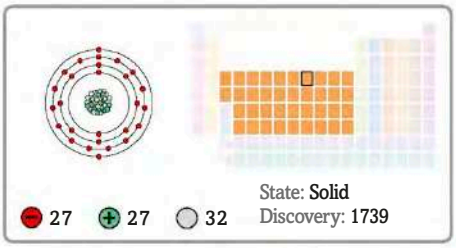

Forms

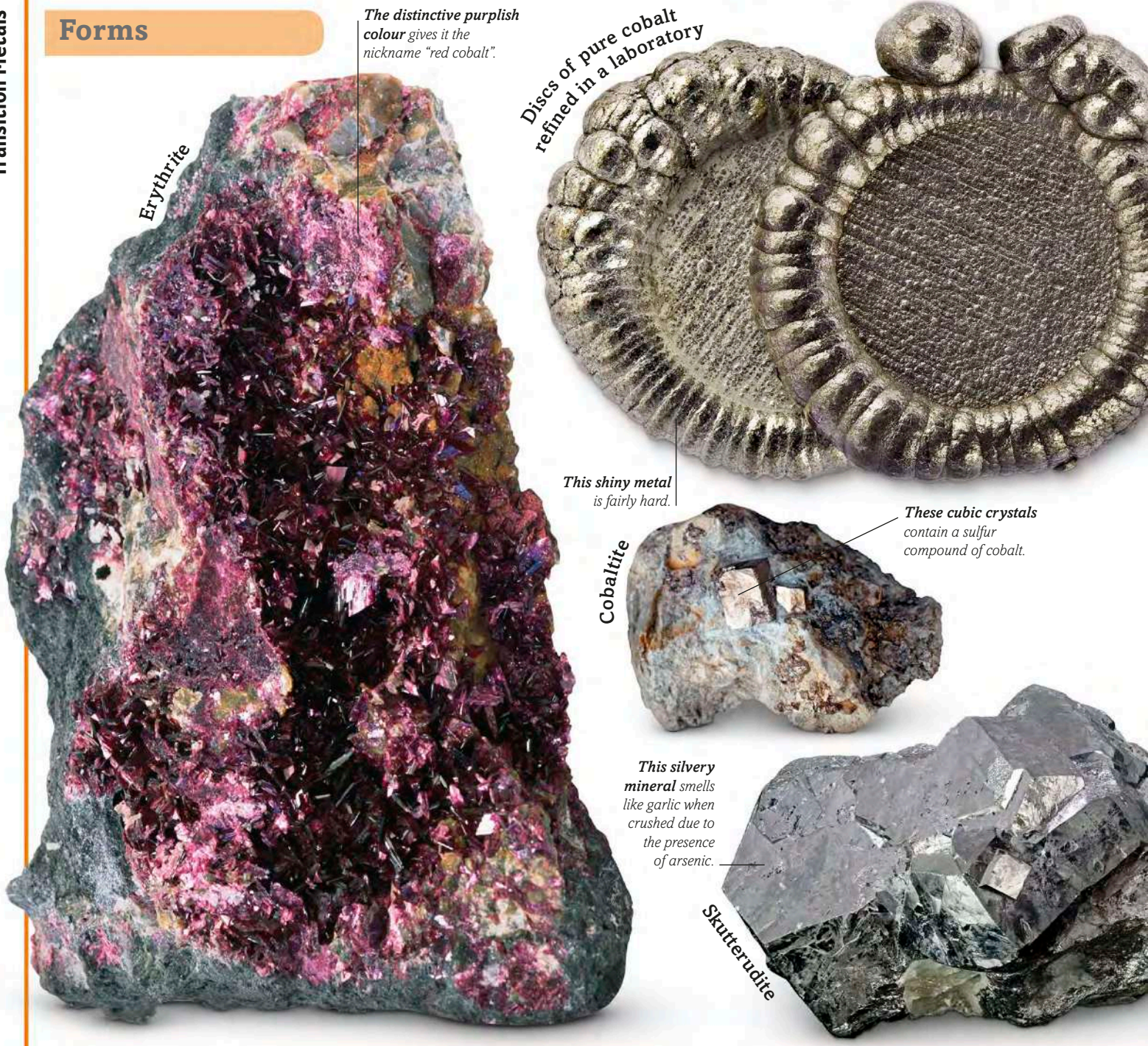

Medieval German miners often mistook ores of cobalt for precious metals. When they tried to purify these, the arsenic gas released made them sick. This unwanted side-effect led to the name *kobold*, which is German for "goblin", a mischievous spirit.

Pure cobalt is hard and shiny, and is added to steel and other alloys to make them stronger. Alloys containing cobalt are used in the blades of **jet engines** and in **artificial joints**, such as hip and knee joints. Cobalt is one of the few elements

Uses

Tough and lightweight artificial joints *can be made from an alloy of cobalt and chromium.*

This part of the joint *is screwed into the hip bone.*

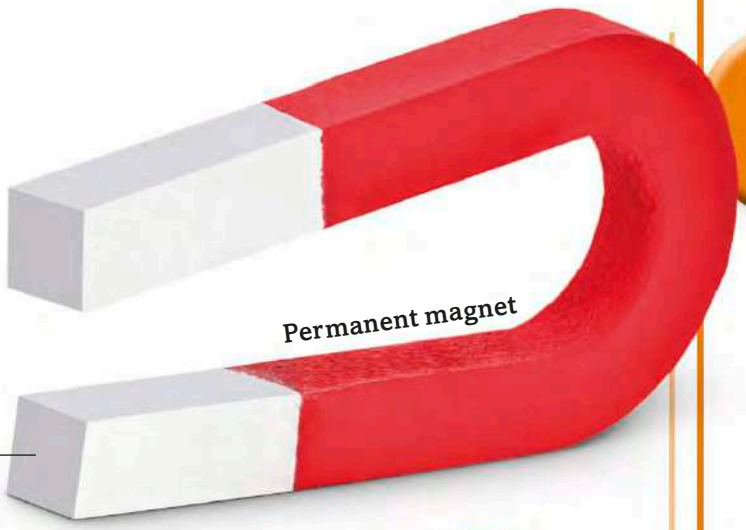

This magnet *can work at temperatures as high as 800°C (1,400°F).*

These blades *made of a cobalt alloy stay hard even when very hot.*

Jet engine turbine

This blue glass *is manufactured using cobalt compounds.*

Cobalt blue paints have been in use **since 3000 BCE**.

This intense colour *does not fade easily over time or upon exposure to light.*

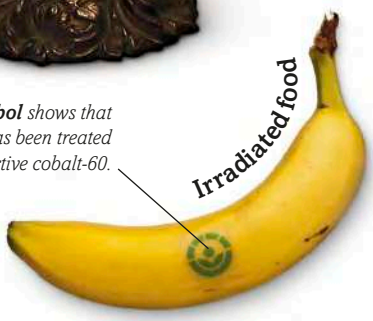

This symbol *shows that this fruit has been treated with radioactive cobalt-60.*

FORMING AN ISOTOPE

Cobalt-60 is an isotope, or form, of this element. It is created artificially in nuclear reactors. Because of its radioactivity, it is useful in some cancer treatments.

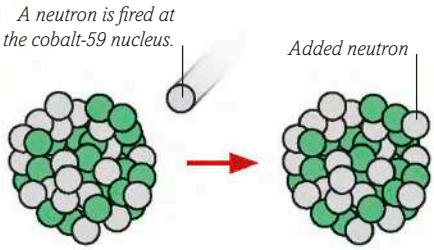

Cobalt-59 is a stable atom with 59 neutrons.

Cobalt-60 is a radioactive atom with 60 neutrons.

that can be used to make a permanent magnet. Large **permanent magnets** are made from a tough alloy of cobalt, nickel, and aluminium, called alnico. A radioactive form of cobalt, called cobalt-60, is produced in nuclear reactors. This form is widely employed to **irradiate food**, a process by which food is exposed to a tiny dose of radiation to kill harmful germs. Cobalt can also produce a deep shade of blue: **cobalt blue paints** and dyes are formed by reacting aluminium with cobalt oxide.

28 Ni Nickel

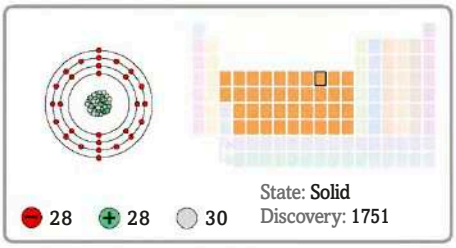

Forms

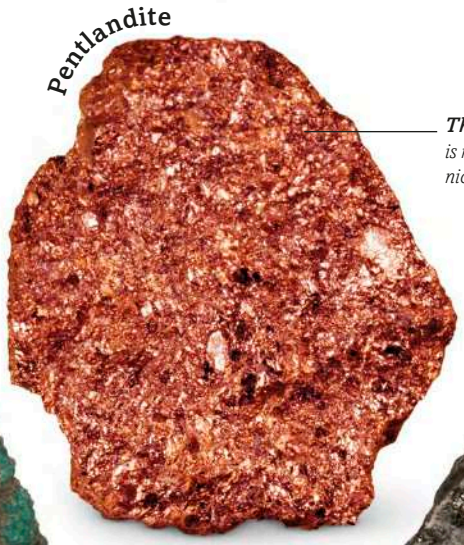

Nickel is named after Old Nick, a demonic spirit from Christian lore that was believed to live underground. In the 18th century, German miners mistook a poisonous nickel mineral, now known as **nickeline**, for a copper ore. When this mineral failed to yield copper, they named it *Kupfernickel*, meaning "Old Nick's copper". Nickel is also found in other ores, such as **garnierite** and **pentlandite**. This element is one of the most useful metals, with a number of applications. Because **pure nickel** does not rust, it is used to coat objects to make

Uses

Nickel goblet drum

This nickel-plated instrument has a shiny finish.

Rust-resistant handle

Nickel-plated dagger

Nickel-copper alloy plating makes propellers strong and durable.

Ship propeller

Nickel-based coin

The US five cent coin is made of an alloy of 75% copper and 25% nickel.

This **coin** from the **USA**, made of a nickel-copper alloy, is called a **nickel**.

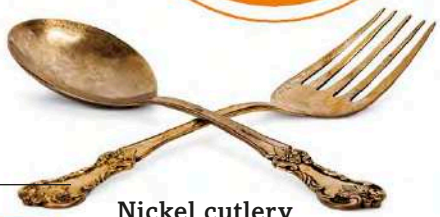

This silver-coated fork is made of an alloy of nickel, copper, and zinc.

Nickel cutlery

Electric guitar

These nickel-plated guitar strings help create a clear tone.

PERMANENT MAGNETS

A temporary magnet works when it is put in a magnetic field, but a permanent magnet retains its charge even when it is taken out of that field. Nickel is one of only a few elements that can be used to make permanent magnets.

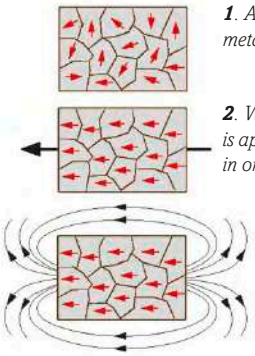

***1**. Atoms in a piece of nickel metal are arranged randomly.*

***2**. When a magnetic field is applied, the atoms align in one direction.*

***3**. The atoms then stay aligned and produce a magnetic field of their own, even after the external field is removed.*

Toaster

Nickel alloy wires heat up to make toast.

them look like silver – a trick still used to make inexpensive ornamental objects. Nickel is also mixed with copper to make an alloy called cupronickel. This is used as plating on **propellers** and other metallic parts of ships, as the alloy does not corrode in seawater. The same alloy is used in most of the world's silver-coloured **coins**. Nickel is used in the strings of **electric guitars**. This element is added to chromium to make an alloy called nichrome. Wires made of this alloy conduct heat very well, so are used in **toasters**.

29 Cu Copper

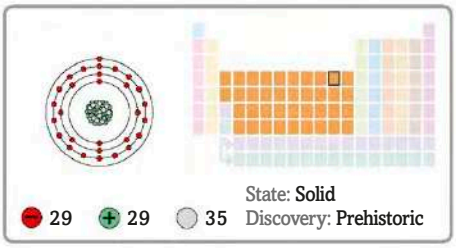

Forms

Copper growth on limonite

Branchlike crystals of copper

Malachite

In caves, feather-like crystals are often formed.

Pellets of pure copper refined in a laboratory

Unique reddish orange colour

Chalcopyrite

These golden yellow crystals contain copper sulfide.

Bornite

Colourful tarnish develops when the mineral reacts with air.

Crustacean blood

Crustacean blood is blue because it contains copper.

Copper is a soft, bendy metal that is an excellent conductor of electricity and heat. Although it is one of the few elements found pure in nature, most of it exists in ores such as **chalcopyrite**. Other copper minerals, such as **malachite** and azurite, are brightly coloured.

Copper is the only metal that has a reddish colour in its pure form. **Pure copper** is mainly used as **wires** in electrical equipment. Copper wire wrapped around an iron core and then electrified helps create an **electromagnet**. Because they can be switched on or off, electromagnets can

Uses

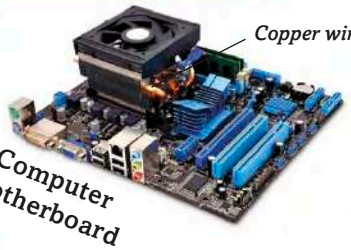

Copper coating on steel *slows down corrosion.*

Massive crane-mounted electromagnets *contain huge coils of copper wire.*

ELECTROPLATING

A thin layer of copper can be added to metal, usually iron, to stop it from corroding easily. This process is called electroplating.

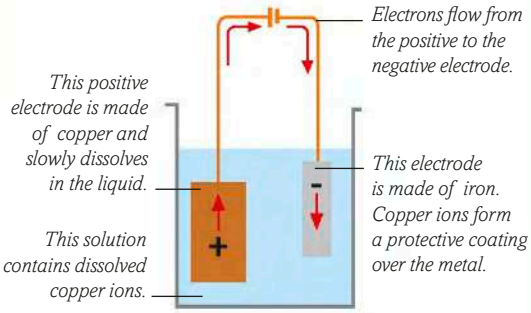

A verdigris layer *protects pure copper from additional weathering.*

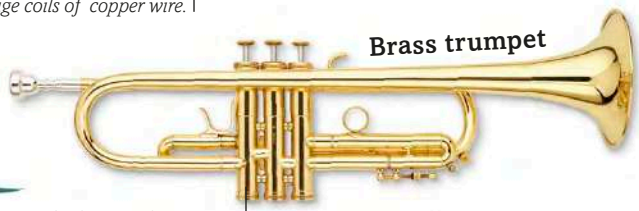

The brass tube *contains air that vibrates to produce musical sounds.*

This tough alloy *does not weaken over time.*

A copper glaze *on the vase gives it a metallic shine.*

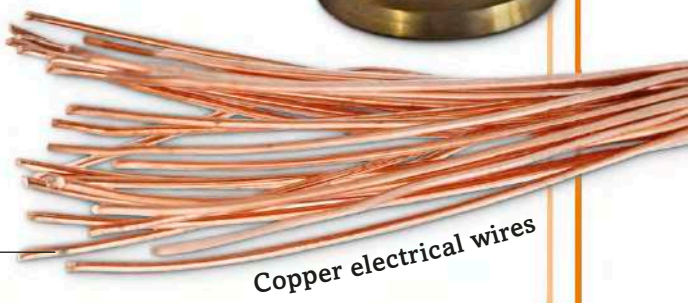

Pure copper *can be stretched to form long wires.*

be magnetic as and when they are needed. They can be much more powerful than normal magnets and can lift heavy objects. Pure copper does not rust, but it reacts with air over time to form a layer of grey-green copper carbonate called verdigris. This can be seen on copper statues, such as the **Statue of Liberty**. Copper is often mixed with other metals to produce tougher alloys. Bronze, a copper-tin alloy, is more durable than pure copper and has been used since ancient times. Brass, a copper-zinc alloy, is used in musical instruments, such as **trumpets**.

COPPER WIRES Not much thicker than a human hair, these copper wires are twisted together and woven into a tight bundle. One of the main uses for these wires is to shield a thicker copper wire that transmits a signal to a television. As the signal carries pictures and sounds in the form of electrical currents, the wires wrapped around it prevent interference from other electrical sources nearby.

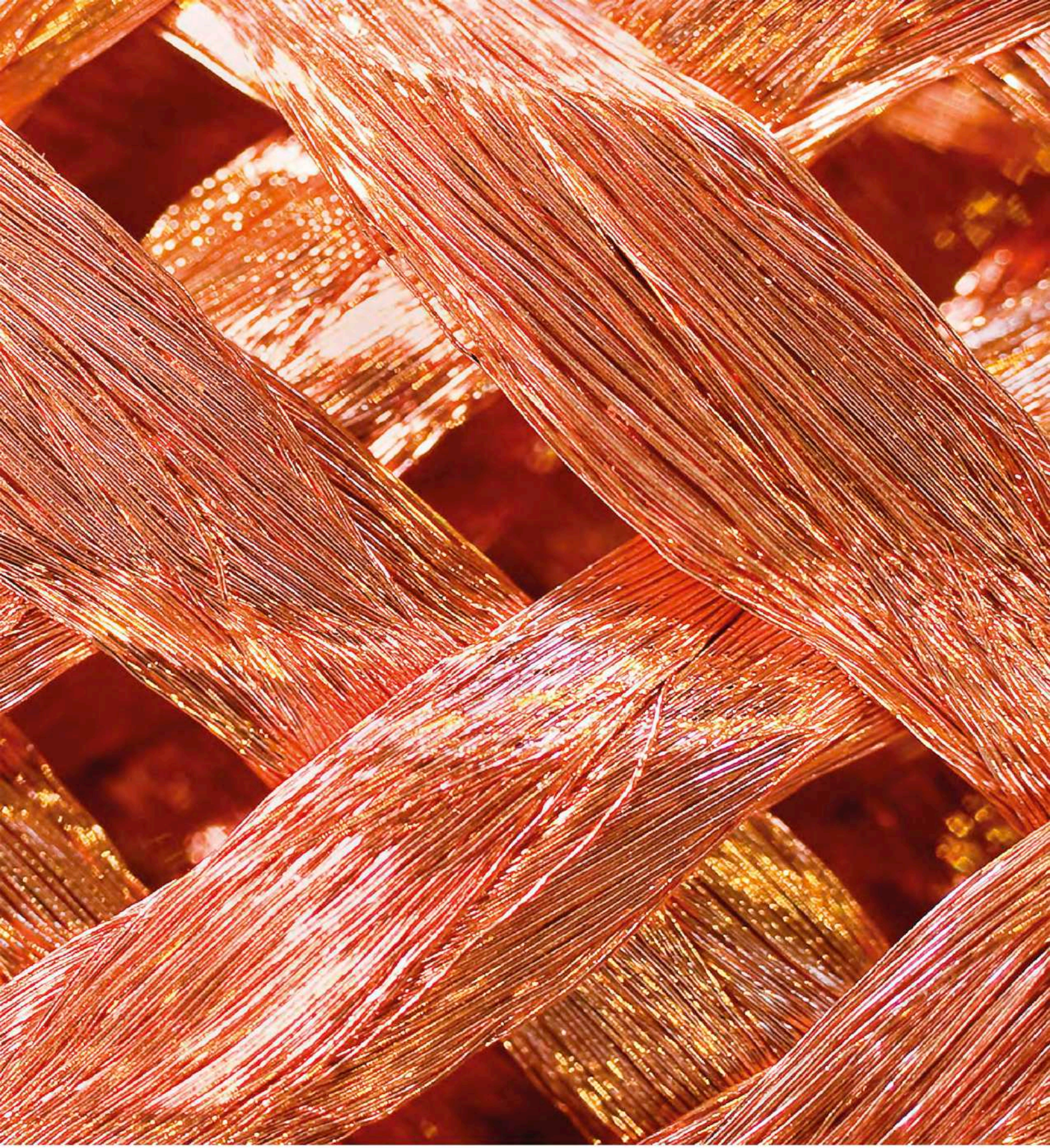

Copper is a very good conductor of electricity, but not the best; silver is better. However, copper is more widely used because it is much cheaper to find and purify. Each year, about 15 million tonnes of pure copper is produced, and more than half of it is used to make electrical components, such as this mesh. Today, more than a billion kilometres of copper wiring is running unseen in power supplies, buildings, and electronics. Copper is now the most common electrical metal, but it has a long history. It was the first element to be refined from ores in large amounts about 7,000 years ago in the region that is now Iraq. Today, Bingham Canyon in Utah, USA, is the world's largest copper mine.

30 Zn Zinc

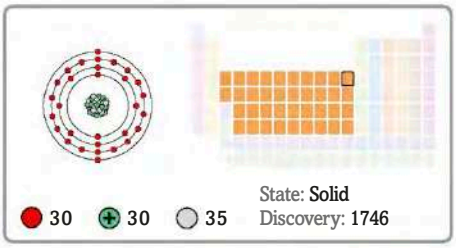

Forms

***This ore** is the main source of zinc.*

***This zinc mineral** forms rough nodules inside cracks.*

Hard, shiny metal

Zinc was used in India and China hundreds of years before the German chemist Andreas Marggraf identified it as a new element in the 18th century. This element is a rare transition metal that is never pure in nature, but is found in many minerals. The mineral **sphalerite**, containing zinc sulfide, is the major source of **pure zinc**. Another principal mineral, **hemimorphite**, contains zinc and silicon. Zinc is essential in our diet. We consume it from food such as cheese and sunflower seeds. Zinc compounds have a wide

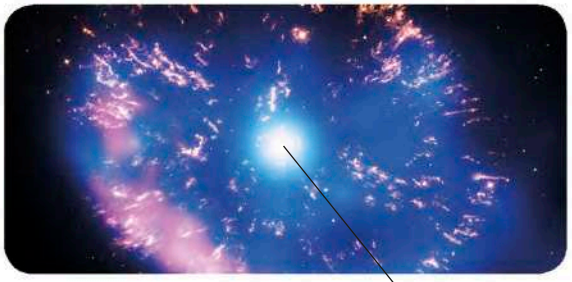

Supernova

Along with many other elements, zinc atoms *are formed inside supernovae (exploding giant stars).*

Hemimorphite

This mineral *contains zinc carbonate.*

Hemimorphite was discovered by James Smithson, the founder of the **Smithsonian** Institution.

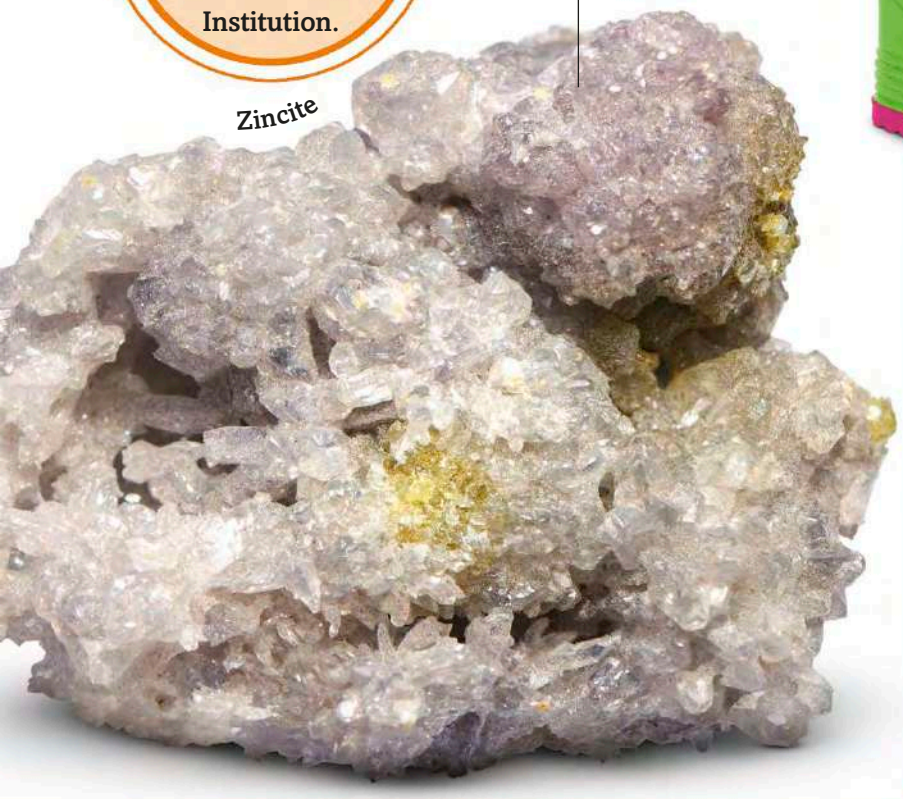

Zincite

Zinc oxide crystals *are generally colourless.*

Uses

First-aid tape

Medical tapes *that contain zinc oxide stop wounds from getting infected with microbes.*

The zinc coating *on this steel bridge protects it from rust.*

Akashi Kaikyo Bridge, Kobe, Japan

American penny

This zinc coin *is coated with copper.*

Rubber boots

This flexible rubber *is made stronger by adding zinc oxide.*

Calamine lotion

This soothing skin lotion *contains a mixture of zinc compounds.*

GALVANIZED STEEL

Steel is protected from corrosion by coating it with zinc. Alloys of iron and zinc form in layers between the steel and pure zinc. This process is called galvanization.

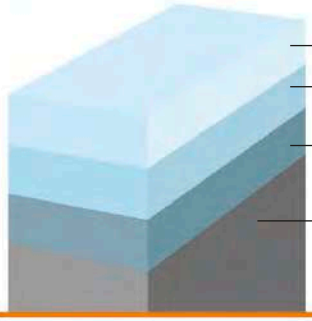

range of applications. For example, a compound of zinc and oxygen called zinc oxide is used in **medical tape** and sunscreen. Zinc oxide can also be used to toughen the rubber used in **boots** and tyres. A compound of zinc and sulfur called zinc sulfide is used to make some paints that glow in the dark. When pure zinc is exposed to air, the metal reacts with oxygen to form a protective layer of an oxide. This coating can prevent objects covered in zinc, such as **bridges**, from corroding easily.

39 Y Yttrium

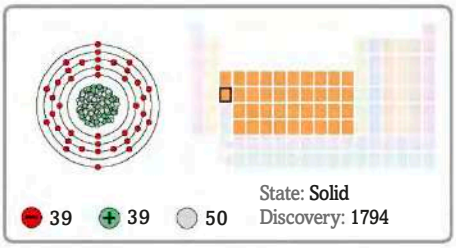

Forms

Yttrium is **400 times more common** in Earth's crust than silver.

Xenotime

This mineral contains trace amounts of uranium, which is radioactive.

Moon rock

This rock sample was brought to Earth by NASA's Apollo 16 mission.

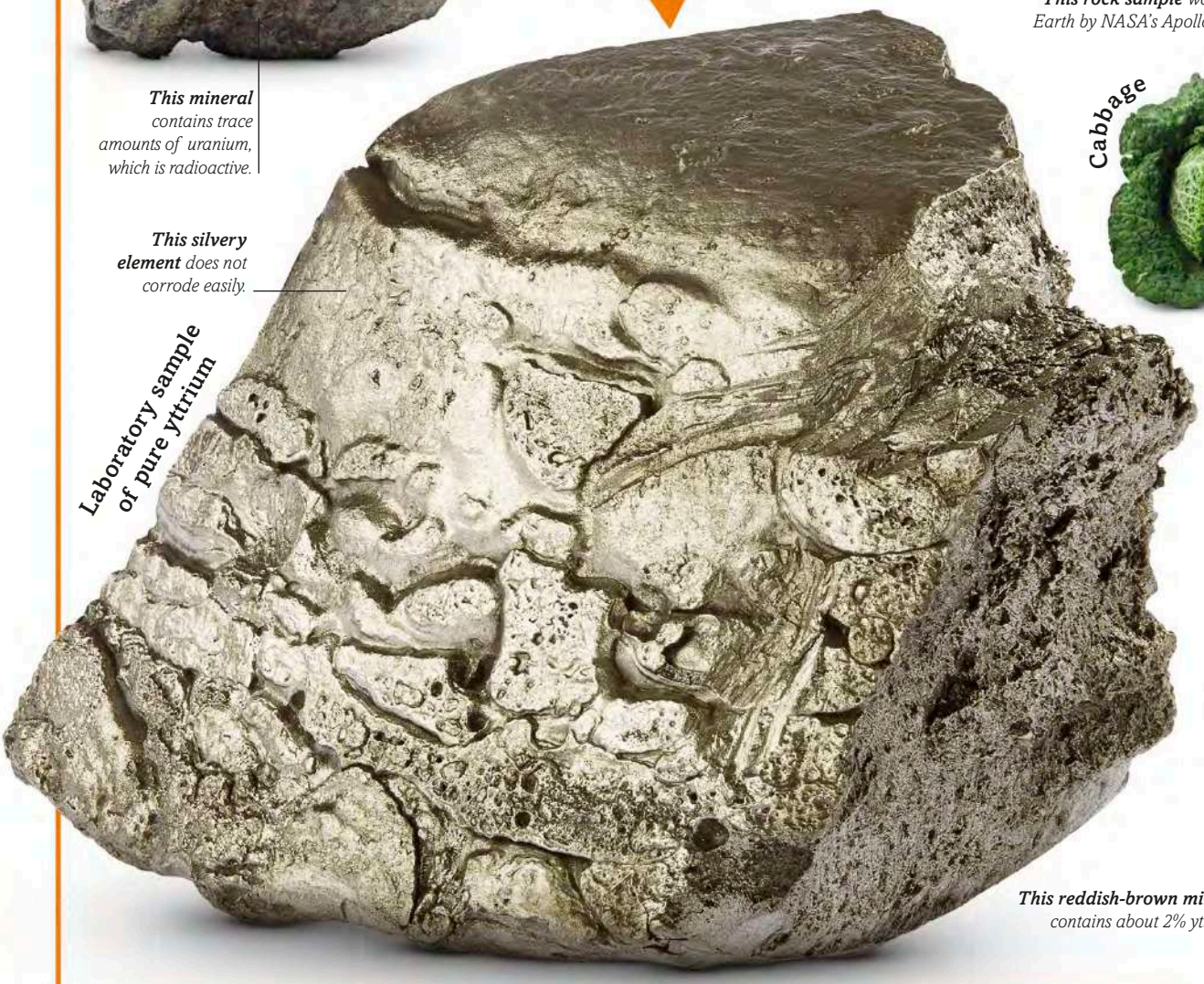

Laboratory sample of pure yttrium

This silvery element does not corrode easily.

Cabbage

This vegetable contains yttrium.

Monazite

This reddish-brown mineral contains about 2% yttrium.

The samples of rock brought back from the Moon by astronauts in NASA's Apollo missions contained higher levels of yttrium than rocks on Earth. This element is never found in pure form in nature, but small traces of it are present in many minerals, including **xenotime** and **monazite**. Yttrium was discovered in a compound in 1794 by the Finnish chemist Johan Gadolin, but it wasn't isolated until 1828. Other yttrium compounds have since been found in vegetables, including **cabbage**, and in seeds of woody plants. In

Uses

LED lamp

This bulb *contains yttrium compounds that enable it to produce a warmer, more yellow glow.*

Laser

This laser *is powered using crystals composed of yttrium and silicon, and can cut through metal.*

Yttrium-90

This radioactive form of yttrium *is used to treat cancers in the body.*

Yttrium gas mantle

This fabric mantle *holds the hot flame inside.*

Digital camera lens

Shock-proof lenses *are made from yttrium-infused glass to make them tough.*

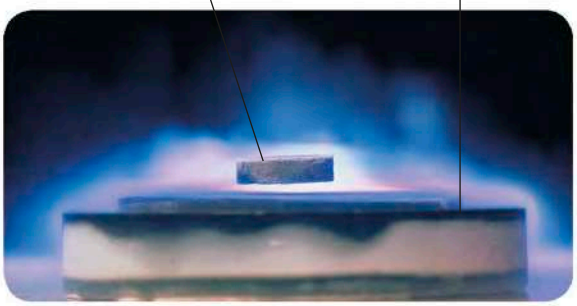

Yttrium superconductor

A small magnet *floats over the superconductor.*

This superconductor *produces a magnetic field that repels the magnet above it.*

NASA spacecraft use yttrium lasers to map the surface of asteroids in space.

FRIEDRICH WÖHLER

In 1828, the German chemist Friedrich Wöhler became the first person to purify yttrium. He did so by extracting it from the compound yttrium chloride. He was also the first person to extract the metals beryllium and titanium from their ores.

LED lamps, yttrium converts blue light to other colours. Many **lasers** use an artificial mixture of yttrium and aluminium inside a silicon-rich crystal called garnet. Powerful yttrium lasers are used for treating some skin infections, as well as by dentists during tooth surgery. A radioactive form of this element has medical applications. Yttrium is added to the glass in a **camera lens** to make it tough. Yttrium compounds are also used in **superconductors** – materials that conduct electricity easily when cooled to very low temperatures.

40 Zr Zirconium

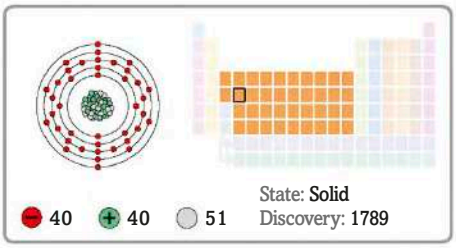

Forms

Zircon crystals

This dark brown colour *is caused by iron impurities.*

Bar of pure zirconium refined in a laboratory

This greyish-white pure element *is easy to shape.*

Uses

Dental crowns

These tough dental crowns *are made of a zirconium-rich ceramic.*

Ceramic knife

This hard, non-metallic blade *does not require frequent sharpening.*

Camera flash from the 1960s

This zirconium-filled bulb *produces a bright light.*

Zirconia crystal ring

This ring *contains cubic zirconia crystals.*

This element is named after the mineral zircon, which means "golden" in Persian, a reference to the golden-brown colour of its crystals. The Swedish chemist Jacob Berzelius was the first person to isolate **pure zirconium**, in 1824. Today, however, the element is mostly used in the form of the compound zirconium dioxide, or zirconia. Powdered zirconia is heated to produce a hard glass-like ceramic, which is used to create **dental crowns** and sharp **ceramic knives**. Powdered zirconia also forms sparkling **zirconia crystals** that look like diamonds.

41 Nb Niobium

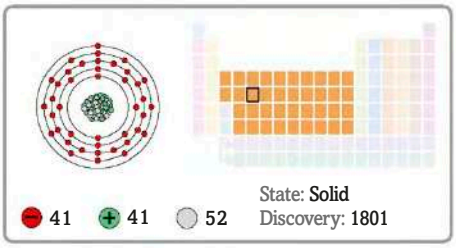

Forms

This dark, dense ore *has a light metallic shine.*

This greyish pure metal *is soft.*

Columbite is named after the country of **Columbia**.

Uses

The niobium case *houses a battery that regulates the beating of a human heart.*

This nozzle *made of a niobium alloy kept its shape even when very hot.*

This pair of spectacles *has thin, powerful lenses made of a niobium compound.*

Niobium is so similar to the metal tantalum that the two were wrongly thought to be the same element for almost 40 years. The mineral **columbite** is the main source of this shiny metal. Niobium is not found naturally in its **pure form**. When extracted, it has many uses. As the element does not react adversely in the human body, it is used in implants, such as **pacemakers**. Niobium also does not expand when hot, so it is used to make parts of rockets, such as the one on the **Command Module** from NASA's Apollo 15 spacecraft that went to the Moon in 1971.

42 Mo Molybdenum

Forms

Molybdenite

This mineral *feels greasy to the touch.*

Chunk of pure molybdenum refined in a laboratory

This metal's pure form *is silver-grey and has a very high melting point: 2,623°C (4,753°F).*

Uses

Lubricant

This slippery lubricant*, which contains finely powdered molybdenite mixed with oil, protects fast-moving mechanical parts in engines.*

Chrome-Moly steel bike

This lightweight but stiff frame *is made from a steel containing molybdenum and chromium.*

Ratchet set

These closely fitting parts *are hard so they do not get damaged easily.*

Molybdenum gets its unusual name from the Greek word *molybdos*, which means "lead". Miners once mistook **molybdenite**, a dark mineral containing this metal, for an ore of lead. This element is much harder than lead, so it is easy to distinguish between these two elements when they are pure. Molybdenite is soft and slippery, and it is the main molybdenum ore. **Pure molybdenum** is mainly used to make alloys that are resistant to corrosion. These are lightweight so are ideal for constructing **bike frames**, but

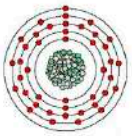

42 42 54

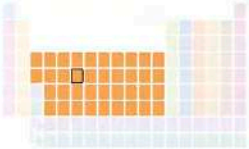

State: Solid
Discovery: 1781

This experimental sports car is built with a rust-resistant alloy that contains molybdenum.

are hard enough for making sturdy tools, too. Molybdenum alloys are used in the latest designs of supercars, such as the **Vencer Sarthe**.

43
Tc

Technetium

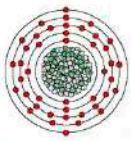

43 43 55

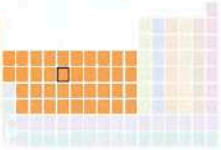

State: Solid
Discovery: 1937

This pure form of the metal is produced inside nuclear reactors.

Foil of pure technetium produced in a reactor

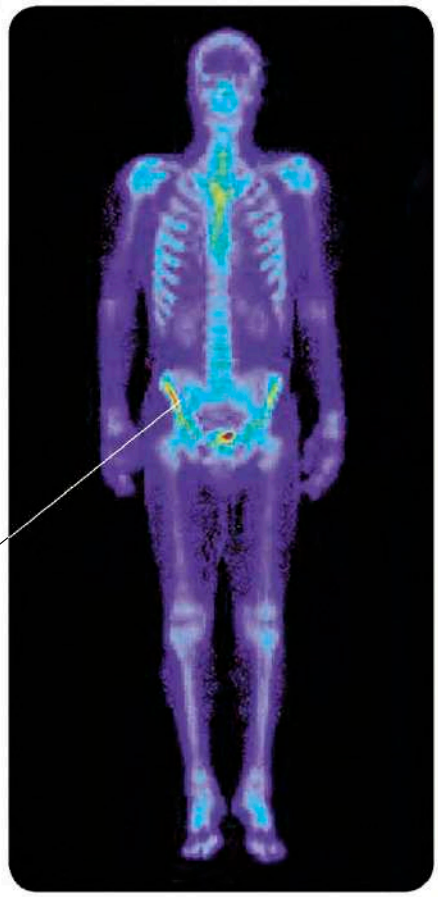

This body scan was created using the radioactive effects of technetium.

Technetium-based imaging

This box contains radioactive molybdenum, which breaks down into technetium.

Generating technetium

Technetium was the first element to be produced artificially by researchers. It is named after the Greek word for artificial, *tekhnetos*. Technetium does not exist in nature: any of its atoms that once existed on Earth broke down millions of years ago. Tiny amounts of this element were discovered in the waste produced by early nuclear reactors. Technetium is the lightest radioactive element. It is used extensively in **medical imaging**. It is injected into a patient's body, where it emits radiation for a short while. Some machines use this radiation to show bones clearly.

44
Ru

Ruthenium

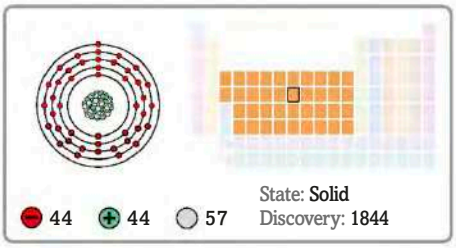

Forms

Pentlandite

This yellow-brown ***mineral*** *is often found deep underground.*

Pure ruthenium refined in a laboratory

These crystals have a bright, silver colour.

Uses

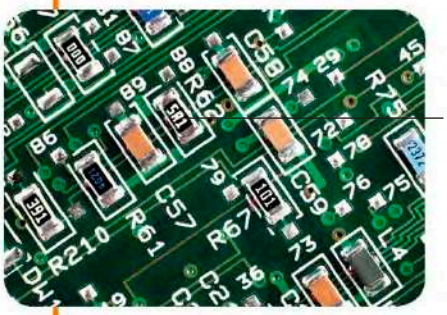

Electronic circuit board

This ***component*** *contains ruthenium dioxide.*

Toggle switch

The metal alloy in the ***switch*** *is toughened by adding ruthenium.*

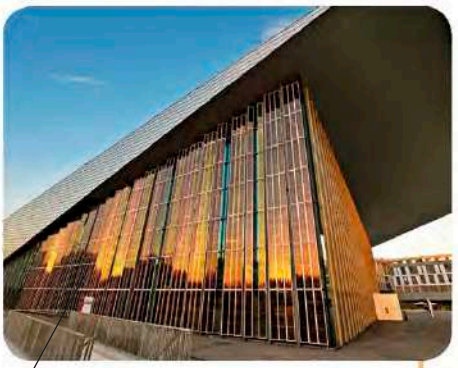

SwissTech Convention Centre, Switzerland

These low-cost solar panels are made using ruthenium.

Ruthenium is named after *Ruthenia*, an old Latin name for Russia. This rare metal is found in the mineral **pentlandite**, and its **pure form** is commonly extracted from this ore. A compound called ruthenium dioxide is used in several components in **electronic circuits**, including resistors and microchips for computers and other digital devices. Adding a small amount of ruthenium makes softer metals, such as platinum and palladium, much tougher. Moving parts in devices such as **switches** benefit from this property.

45 Rh Rhodium

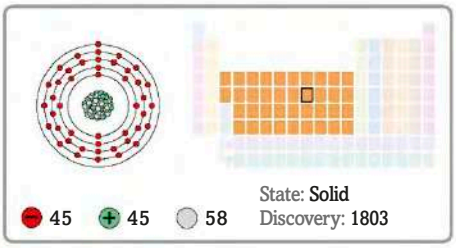

Forms

Pure rhodium pellet refined in a laboratory

This metal *is shiny and silvery when pure.*

Millerite

This golden mineral *is defined by its needle-shaped crystals.*

Uses

Rhodium-plated jewellery

The rhodium plating *prevents jewellery from losing its shine.*

Headlight reflector

This rhodium-alloy reflector *provides a bright light.*

Rhodium-plated microscope

Parts of this microscope *are coated with rhodium and can resist corrosion.*

Fibreglass production

These fibreglass strands *are made by passing molten glass through rhodium-enriched trays.*

The rosy red colour of one of its compounds inspired the name rhodium. The Greek word *rhodon* means "rose-coloured". Rhodium is unreactive and does not form compounds easily. It is a rare metal. Most of the **pure form** is extracted when platinum is mined. Pure rhodium is hard and is used to toughen precious **jewellery**, mirrors, and optical devices, such as **microscopes**. It is mainly used in the production of catalytic convertors for cars. **Fibreglass**, which is often found in protective gear – like helmets – also contains rhodium.

46 Pd Palladium

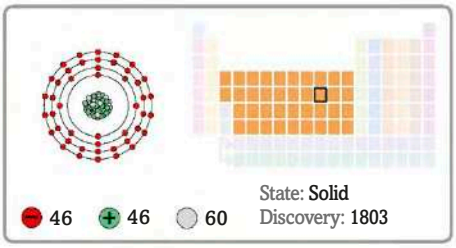

Forms

The mines in this area have a high concentration of palladium.

Blue Ridge mines, South Africa

Palladium absorbs **hydrogen**, like a sponge soaking up water.

Pure form can be produced by separating it from the ores of other metals, such as copper and nickel.

Pure palladium pellet refined in a laboratory

Uses

The converter gets hotter as more pollutants enter the exhaust.

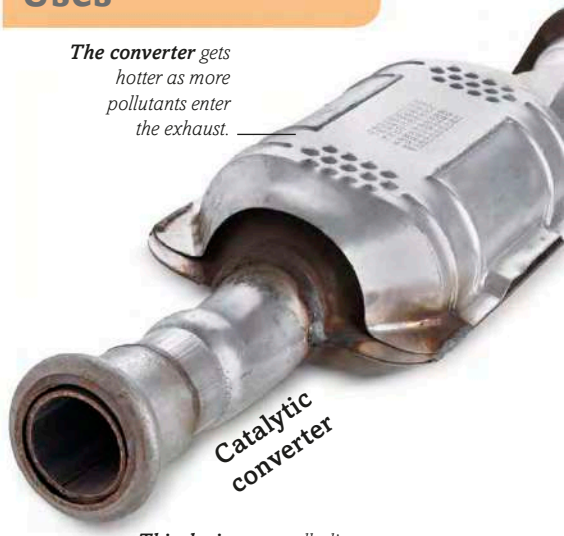

Catalytic converter

This device uses palladium compounds that change colour when poisonous carbon monoxide is detected, which triggers an alarm.

Carbon monoxide detector

This commemorative coin is made from the palladium produced by the Stillwater Mining company in Montana, USA.

Palladium coin

Palladium is a rare, precious metal: it is 10 times rarer than silver and twice as rare as gold. Like these metals, palladium has a shiny surface and does not corrode easily. Palladium is found **pure** in nature, but it also has a few rare minerals, such as braggite. Of its many applications, the element's main use is in **catalytic converters**, which are devices used in vehicles to convert poisonous exhaust gases into less harmful ones. A compound called palladium chloride is used in **carbon monoxide detectors**. Because the element is

CATALYTIC CONVERTER

Many cars contain engines that are attached to catalytic convertors. These important devices convert toxic exhaust gases into less harmful pollutants. Palladium plays a key role in the process.

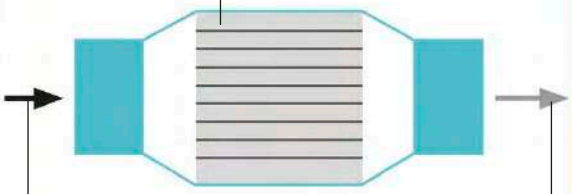

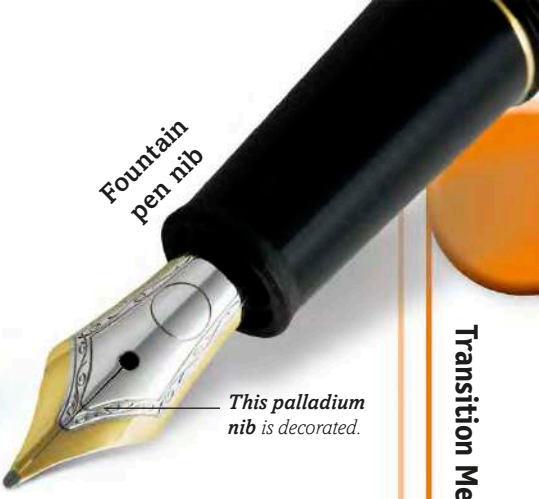

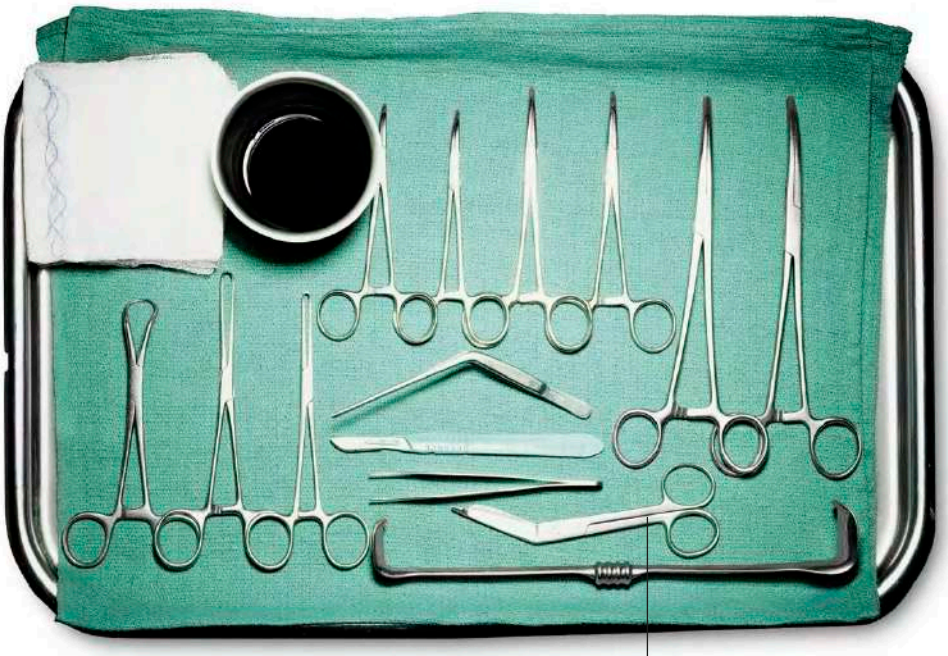

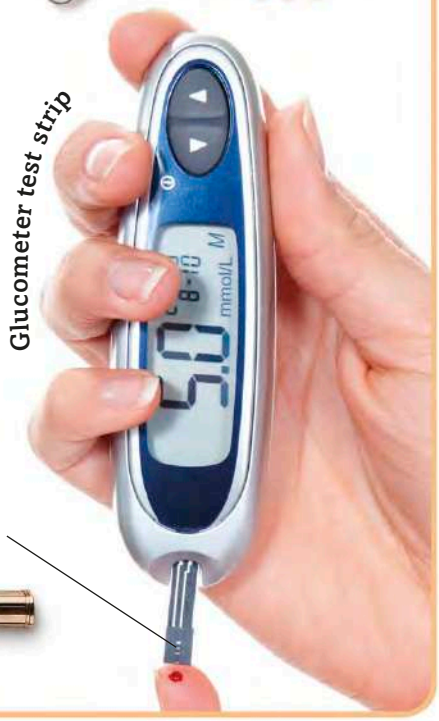

Orchestral flute

precious, it is used to make **commemorative coins** in some countries. Palladium is alloyed with steel to make it more resistant to corrosion. These alloys are used to make **surgical tools** and expensive musical instruments, such as some **flutes**. Palladium is often mixed with gold to form an alloy called white gold, which is used in jewellery. Some fountain pens have **nibs** decorated with palladium. The element is also used in **glucometer test strips** so that patients can check the level of glucose in their blood.

47 Ag Silver

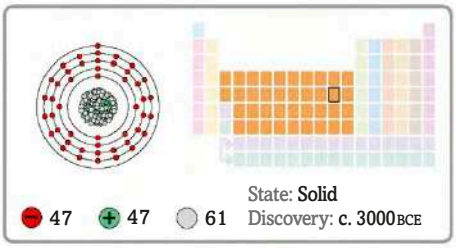

Forms

This mineral's ***colour*** *changes to purple when it is exposed to bright light.*

The bright surface tarnishes after reacting with air.

These large ***opaque crystals*** *have a glistening sheen.*

Black silver sulfide forms twisted crystals.

1 g (0.03 oz) of silver can be drawn into a **2-km- (1.2-mile-) long** wire.

Silver gets its symbol "Ag" from its Latin name *argentum*, which means "shiny white". It is considered a precious metal because its pure form has a grey shine that does not corrode quickly, and it stays untarnished if cleaned regularly. Silver can be found pure in nature, but mostly it is mined from ores, such as **pyrargyrite** and **acanthite**. Because this element is valuable and can be moulded easily, pure silver was used historically to make **coins**. This metal is also ideal for making **bracelets** and settings for gems. Some people even use

Uses

MAKING CLOUDS

Rain is crucial to our Earth, especially for growing healthy crops. Where there are no clouds, scientists can form tiny water droplets that cling to silver iodide powder, forming artificial rain clouds.

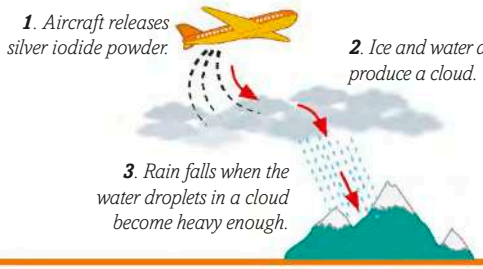

Soft silver *is easily pressed into coins.*

Silver coating *is used on some circuit board parts.*

Circuit board

Silver bracelet

Pure silver *moulded and cut into varying shapes*

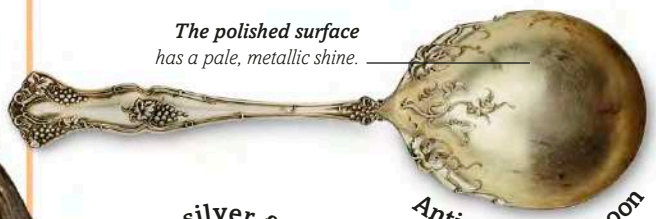

The polished surface *has a pale, metallic shine.*

Antique silver spoon

These thin sheets of silver *called "vark" are edible.*

Silver nitrate *is mixed with water to clean cuts and scrapes.*

Silver nitrate

Glass infused with silver chloride *turns brownish when exposed to sunlight.*

Photography plate

An image forms *when silver bromide darkens quickly on exposure to light.*

flattened **silver foil** to decorate food. **Silver spoons** and forks were the only pieces of cutlery that did not create a nasty metallic flavour in the mouth in the days before the invention of stainless steel. Silver conducts electricity better than copper, and is used in some **circuit boards**. **Silver nitrate** (a compound of silver, nitrogen, and oxygen) is a mild disinfectant used in some anti-bacterial soaps. Silver forms light-sensitive compounds with chlorine (used in **sunglasses**) and bromine (used in old **photography plates**).

48 Cd Cadmium

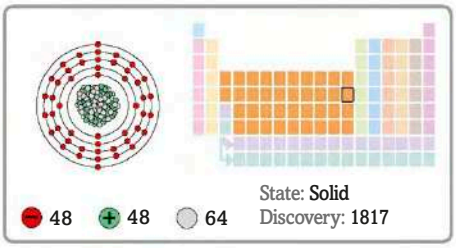

Forms

This mineral *contains a rare form of cadmium sulfide, a compound of cadmium and sulfur.*

Greenockite

This soft metal *has a bluish tinge.*

Pellet of pure cadmium refined in a laboratory

The yellow colour *of this zinc mineral is due to cadmium impurities.*

Smithsonite

Uses

Nickel-cadmium battery

Cadmium and nickel layers *create electricity in this rechargeable battery.*

This deep red pigment *contains powdered cadmium oxide.*

Red paint containing cadmium

This electronic component *used in circuits contains a compound of cadmium and sulfur.*

Light-sensitive resistor

Cadmium-covered screws *do not rust.*

Cadmium-plated screws

This research sample *is being viewed under ultraviolet (UV) light produced by a cadmium laser.*

Fluorescent microscope

Cadmium is a highly toxic metal, and is known to cause cancer. This rare element is found in the ore **greenockite**, but it is mostly obtained as a by-product of zinc extraction. **Cadmium** was discovered in 1817 from a mineral called calamine. Today, this metal is mainly used in conjunction with nickel in **rechargeable batteries**. The compound cadmium oxide was once used in preparing **red paints**, but not anymore because of its toxicity. Cadmium is also used to create lasers for use in powerful **microscopes**.

72 Hf Hafnium

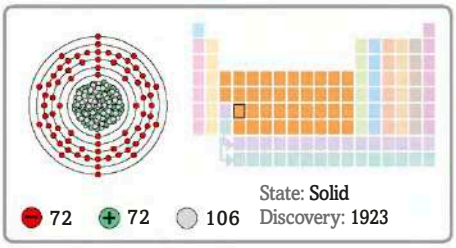

Forms

Zircon crystals

***Hafnium** uses up 4% of the mass of this zircon crystal.*

A zircon crystal can be up to **4 billion years** old.

Laboratory sample of pure hafnium

***This element's pure form** is resistant to corrosion in air.*

Uses

***Sections of this cutter** are made of hafnium.*

Metal cutter

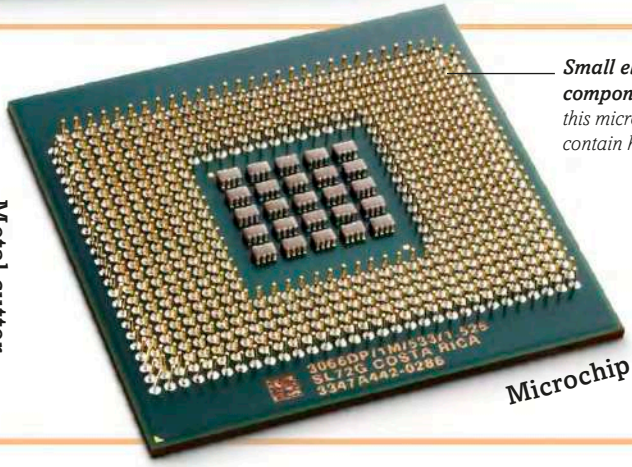

***Small electronic components** in this microchip contain hafnium.*

Microchip

Hafnium is named after *Hafnia*, the Latin word for the city of Copenhagen in Denmark. It took a long time to distinguish hafnium from zirconium because the two elements are present together in crystals of the mineral **zircon** and their atoms are similar sizes. Hafnium is used in powerful **cutters** that pierce metallic objects with a hot stream of sparks. It is also used to make ultra-small electronics – only a few millionths of a millimetre wide – in **microchips**.

73 Ta Tantalum

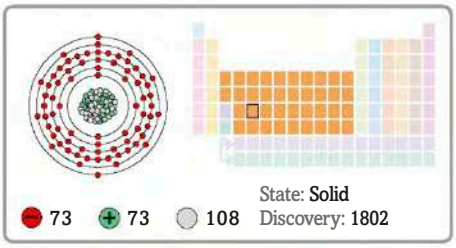

Forms

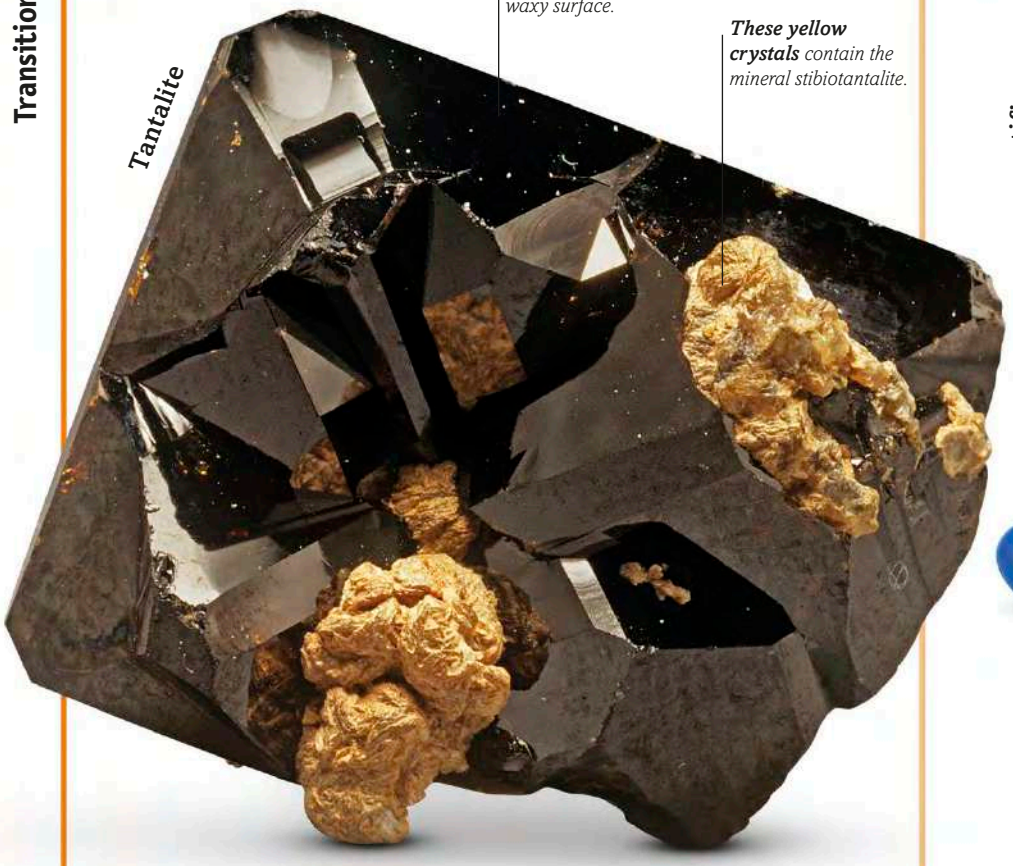

Rods of pure tantalum refined in a laboratory

Uses

Tantalum is a hard metal named after Tantalus, a man from Greek mythology who was punished by the gods. It is extracted from a rare mineral called **tantalite**. This tough metal is not harmful to the human body, so it is used to make **artificial joints** and other body implants. Tantalum powder is used in **capacitors** – devices used in electronic circuits to store electricity. This strong metal toughens **watches** made of softer, precious metals. Tantalum is also used to create strong turbine blades that do not corrode.

74 W Tungsten

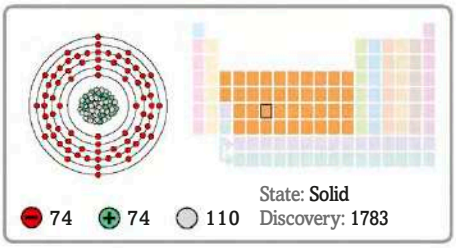

Forms

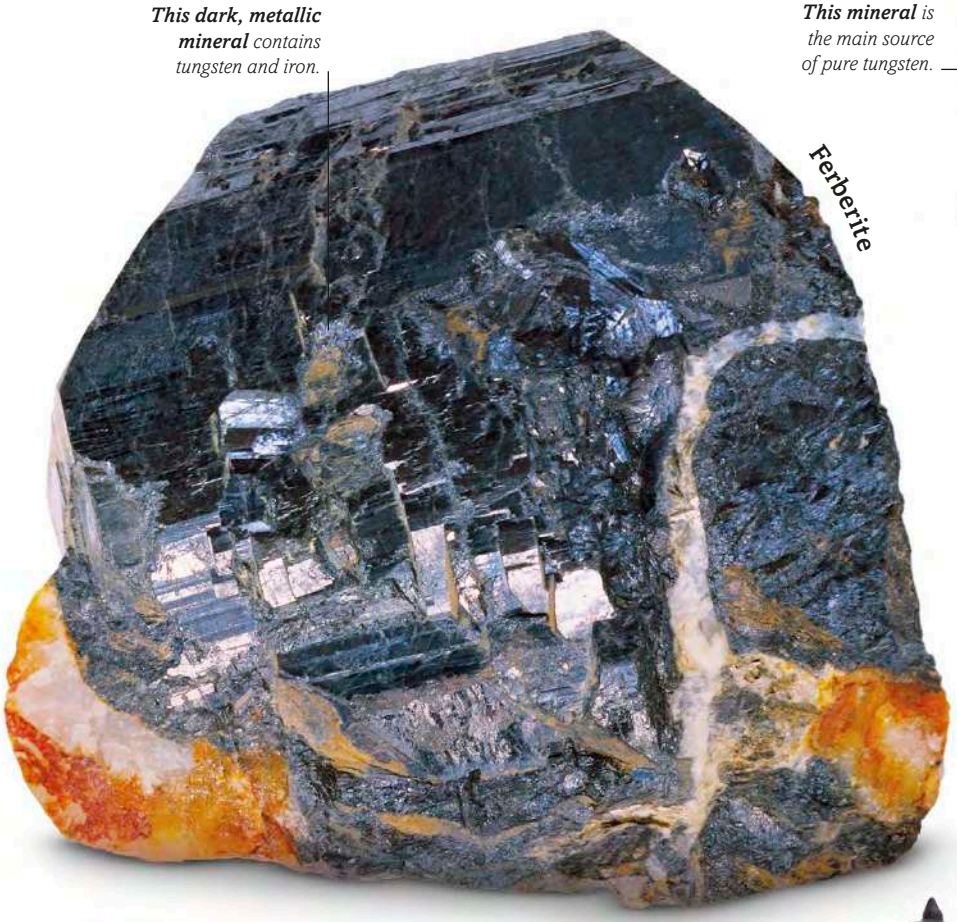

This dark, metallic mineral *contains tungsten and iron.*

Ferberite

This mineral *is the main source of pure tungsten.*

Wolframite

Pure tungsten *is a hard, grey metal.*

Cylinder of pure tungsten refined in a laboratory

Uses

Drill bit

This drill bit *has a coating of tungsten carbide, which protects it from damage.*

Tungsten pigments were in use in Chinese porcelain 350 years ago.

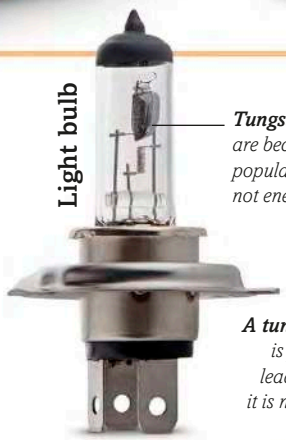

Light bulb

Tungsten filaments *are becoming less popular as they are not energy-efficient.*

Fishing sinker

A tungsten sinker *is preferred to a lead one because it is not poisonous.*

Tungsten has the highest melting point of any metal: it turns to liquid at a searing 3,414°C (6,177.2°F). It is a very dense metal, and its name comes from the Swedish phrase for "heavy stone". This metal is usually obtained from the mineral **wolframite**. A compound called tungsten carbide is used to harden objects such as **drill bits**. Tungsten's high melting point allows it to be used in the filaments of **light bulbs**. This element is also useful in producing weights, such as **sinkers** used with fishing lures.

75 Re Rhenium

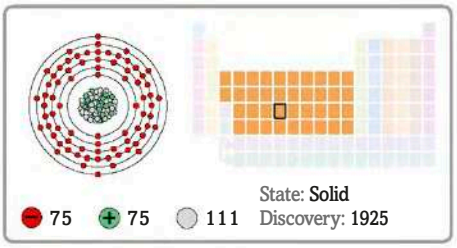

Forms

Molybdenite

This ore *contains molybdenum and small amounts of rhenium.*

Pure rhenium pellet

Pure rhenium *is denser than gold.*

Uses

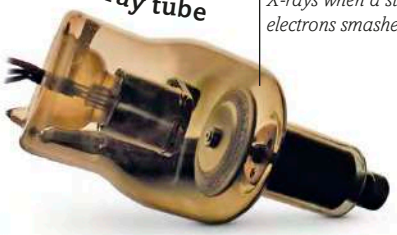

X-ray tube

These tubes *contain a rhenium alloy that produces X-rays when a stream of electrons smashes into it.*

Rhenium has the **highest boiling point** of any element.

F-22 Raptor fighter plane

This plane *has jet engines containing a heat-proof rhenium alloy.*

Rhenium is very rare in nature: only one atom out of every billion in Earth's crust is a rhenium atom. Discovered in Germany in 1925, and named after the Rhine river – it was the last stable, non-radioactive element to be found. Rhenium has a very high melting point, and can stay solid at extreme temperatures. This allows alloys made of this element to be used in very hot conditions, such as those inside the tubes of **X-ray machines**, as well those in the exhaust nozzles of rockets and the jet engines of **fighter planes**.

76
Os

Osmium

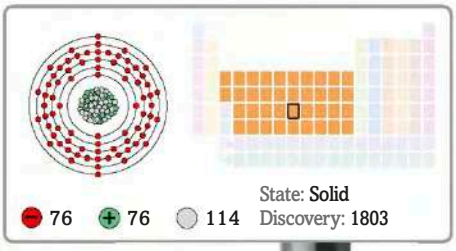

Forms

This is a natural alloy *of osmium and iridium.*

The pure form *is hard but shatters easily.*

Uses

Transmission electron microscope (TEM) image

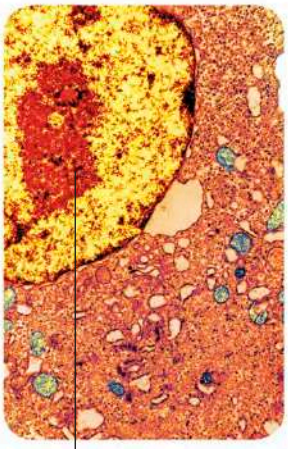

Osmium oxide *is used to highlight objects inside a cell.*

Fingerprint powder in use

Black osmium oxide powder *clings to oily fingerprints.*

The needle *of this old record player is made of osmium.*

Record player

Fountain pen

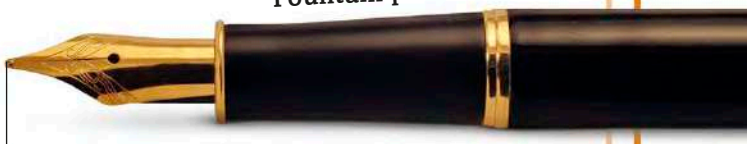

The nib of this pen *moves smoothly because of its hardy osmium alloy.*

Osmium is the densest of all naturally occurring elements: 250 ml (8.5 fl oz) of this metal (in its liquid form) weighs 5.5 kg (12 lb). This rare element is found in the ore **osmiridium**. **Pure osmium** reacts with oxygen in the air to form a poisonous oxide, so the metal is used safely by combining it with other elements or alloys. A red osmium oxide stains cells so they can be seen clearly under a powerful **microscope**, while a black oxide powder allows **fingerprints** to be revealed in crime investigations. A hard osmium alloy is used in **fountain pen** nibs.

77 Ir Iridium

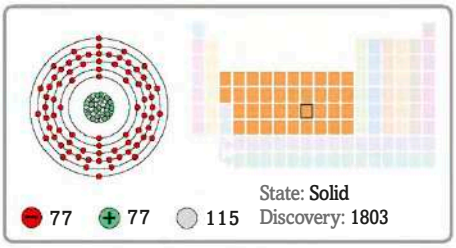

Forms

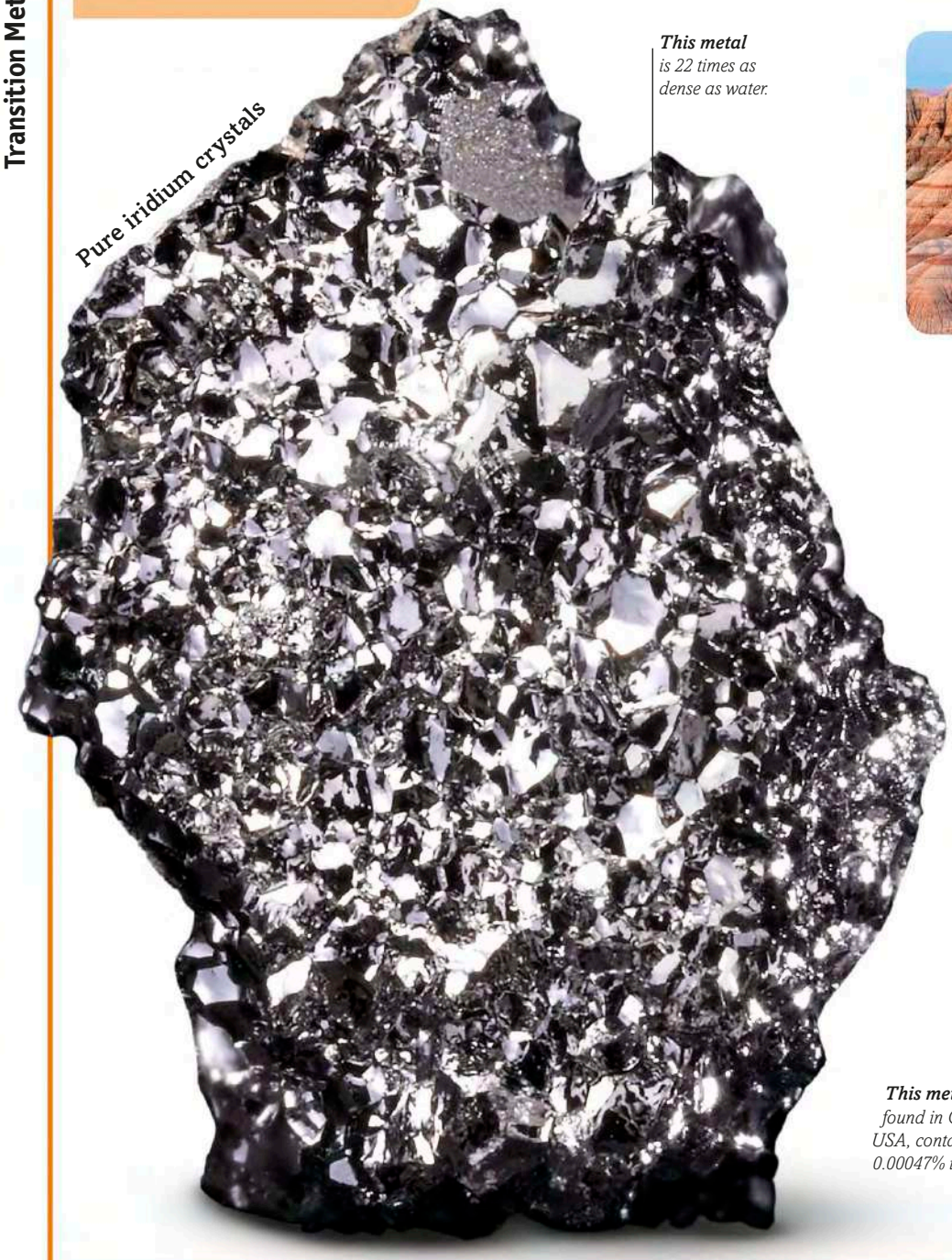

Pure iridium crystals

This metal *is 22 times as dense as water.*

These rocks *contain a layer of iridium-rich clay.*

Badlands National Parks, South Dakota, USA

Willamette meteorite

This meteorite*, found in Oregon, USA, contains just 0.00047% iridium.*

Iridium is the rarest natural element on Earth: there is one iridium atom out of every billion atoms in Earth's rocks. This dense metal can be found in its **pure** form in nature as well as in other common ores that contain nickel and copper. Iridium is present in **meteorites** and other space rocks. A layer of iridium-rich clay is found in Earth's crust all over the world, especially in the **Badlands of South Dakota, USA**. Scientists believe this small quantity of iridium in our planet's crust was deposited by the dust from an

Uses

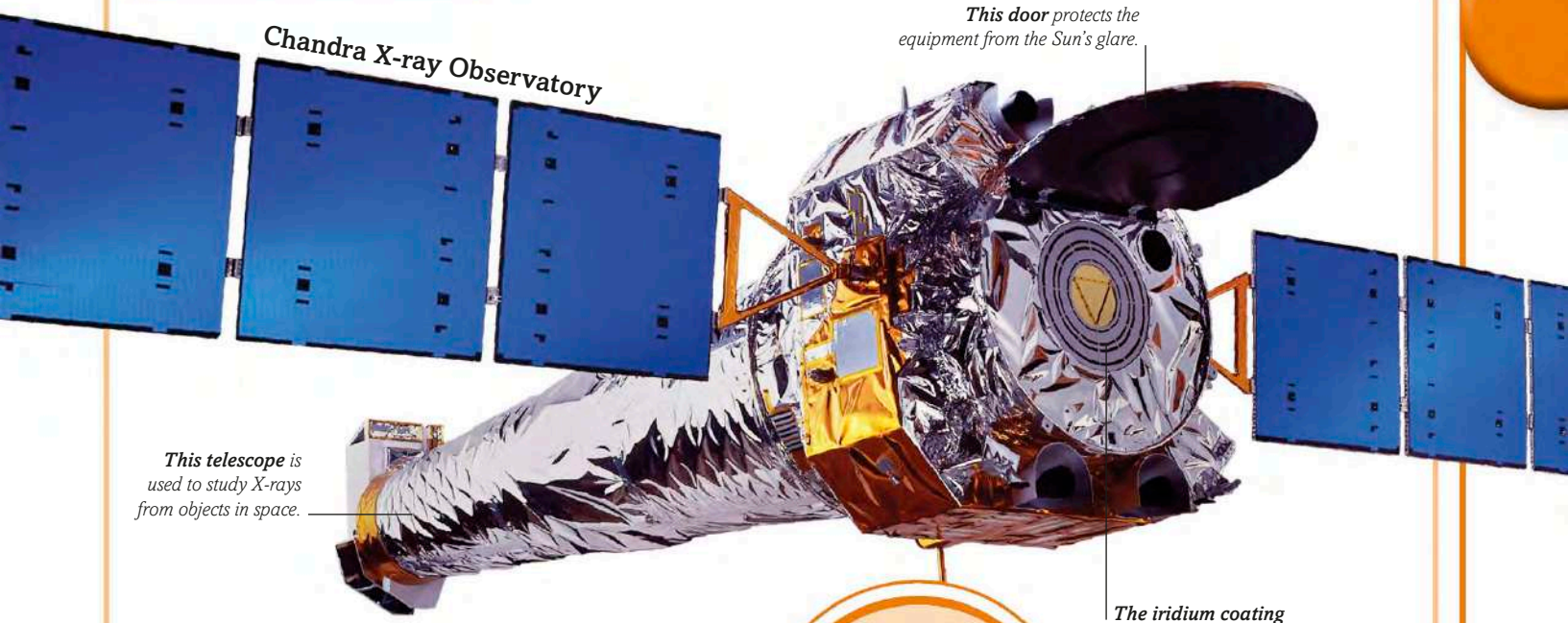

Iridium is the **second densest metal**, after osmium.

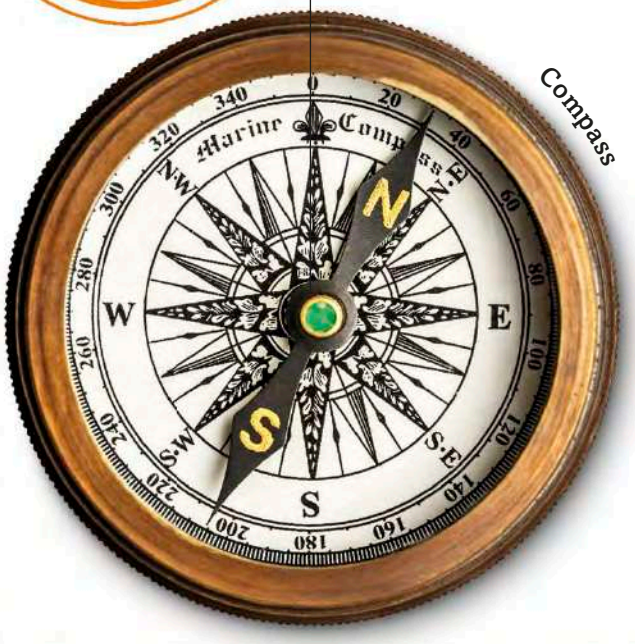

LUIS WALTER ALVAREZ

In 1980, the American physicist Luis Walter Alvarez and his son Walter discovered a layer of iridium-rich clay in rocks all over the world. They suggested that this was the result of a meteorite impact about 66 million years ago, which led to the extinction of the dinosaurs.

explosion 66 million years ago when a large meteorite hit our planet. The applications of this element include coating the mirror of NASA's **Chandra X-ray Observatory**, an Earth-orbiting telescope that studies X-rays from distant stars. Iridium is more durable than platinum and copper, and is therefore preferred over these metals for use in spark plugs. Iridium is also mixed with osmium to make an alloy called osmiridium, which is used in **compasses** and put in nibs for some fountain pens to make them hardy.

78 Pt Platinum

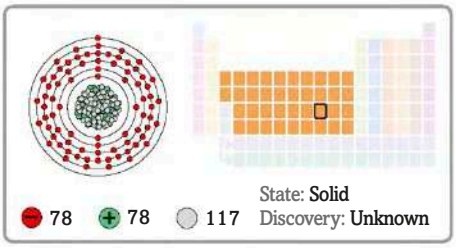

Forms

This dense, shiny mineral *consisting of platinum and arsenic is the most common ore of platinum.*

Platinum melts at the high temperature of 1,768°C (3,214°F).

Sperrylite

Large nuggets *of pure platinum are rare.*

Nugget of pure platinum refined in a laboratory

Spanish explorers first found platinum in the mines of South America in the 1700s. They obtained a whitish substance that the locals living near there called *platina*, meaning "little silver". This precious metal has a silvery white shine. Platinum rarely reacts with other elements, even at high temperatures. This makes it difficult to extract from its ores, such as **sperrylite**. **Pure platinum** does not corrode or tarnish. It is, however, not easy to shape or mould, so use of platinum was limited to the making of simple jewellery and **watches**. By the 20th century,

Uses

Platinum resistance thermometer

This thermometer records temperature by measuring the electric current flowing through a fine platinum wire.

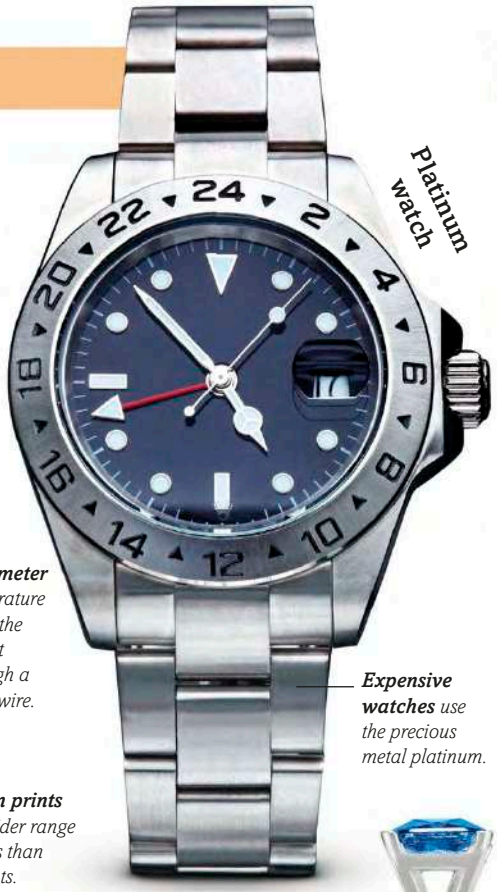

Platinum watch

Expensive watches use the precious metal platinum.

Platinum prints have a wider range of shades than silver prints.

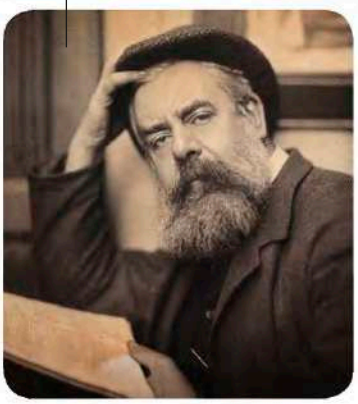

Black and white photographic print

Jewellery made of platinum does not lose its shine.

Platinum ring

Dental fillings once contained platinum and mercury.

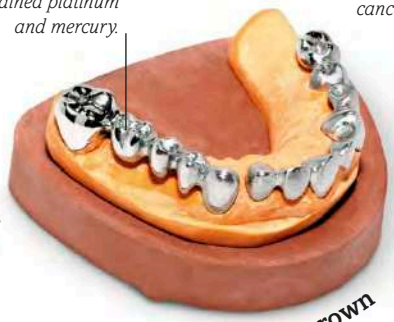

Dental crown

This drug contains platinum and kills cancer cells in the body.

Cancer drug

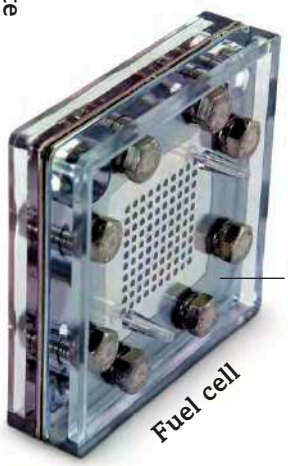

Fuel cell

This fuel cell contains platinum, which speeds up the reaction between hydrogen and oxygen.

This stent made of platinum is not harmful to the body and anchors a damaged blood vessel as it heals.

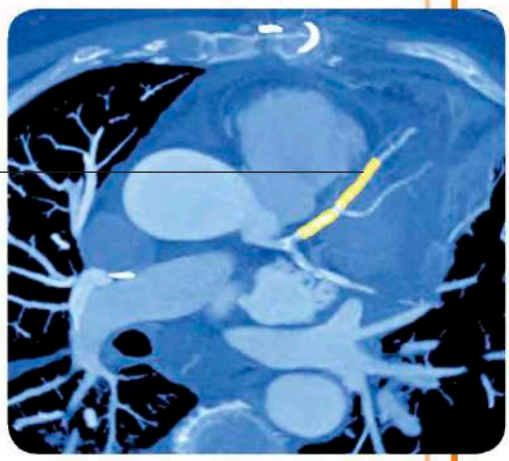

Medical stent

Platinum was found in an Egyptian casket from the **7th century BCE**.

ANTONIO DE ULLOA

Although platinum had been in use in jewellery on the west coast of South America for more than 2,000 years, it was Spanish naval officer Antonio de Ulloa who made the first major study of it. In 1735, while on a South American expedition, he found grains of the metal in river sands. He brought them back to Spain to examine them.

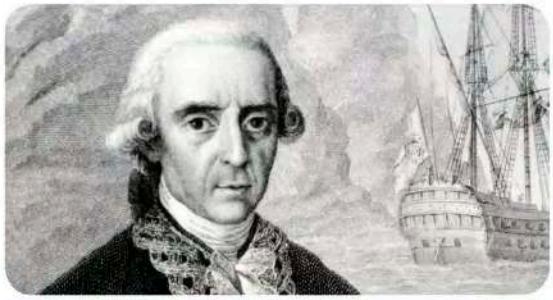

more applications were discovered. Platinum can be used in place of silver to generate **photographic prints**, and in place of gold for making **dental fillings**. Today, platinum plays an important role in various technologies. For example, it is used in **fuel cells** – devices that generate electricity by combining hydrogen and oxygen. These cells do not need to be recharged like other batteries. Powerful **drugs for treating cancer** contain this element, while **stents** made of pure platinum help heal damaged blood vessels.

79 Au Gold

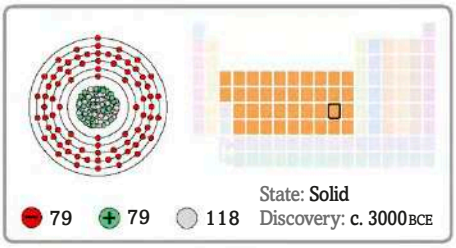

Forms

Gold's chemical symbol **Au**, comes from its **latin name**, *aurum*.

Pure gold *has a dark yellow colour unlike any other metal.*

These crystals *may be elongated, rectangular, or cubic.*

Crystals of pure gold refined in a laboratory

Calaverite

This mineral ore *contains gold and tellurium.*

Natural gold in quartz

These flakes of pure gold *are locked into quartz crystals.*

People were making gold ornaments more than 6,000 years ago. This was many centuries before they learned how to purify copper, iron, and other metals. Gold is believed by many to be the first metal element to be identified. It is a dense, unreactive metal with a distinctive deep yellow colour. Gold is naturally **pure** and seldom makes compounds in nature; the compound in the mineral ore **calaverite** is an exception. **Pure gold** found in nature may form nuggets but mostly is found as tiny specks embedded in rocks. Gold miners crush up these rocks and wash out the

Uses

This very thin layer of gold *protects the astronaut from the Sun's heat.*

Astronaut's visor

This mask *was placed over the pharaoh's mummified face.*

Tutankhamun's death mask

Royal Crown Derby plate

This glass plate *contains specks of gold.*

Gold bars

Gold bars *stored in banks are a sign of wealth.*

Edible gold flakes

The flakes *decorating this expensive chocolate are edible.*

Gold teeth

These replacement teeth *are made of gold and mercury.*

Wat Phrathat Doi Suthep temple, Thailand

A thin layer of gold *covers this entire temple.*

Gold foil *keeps this car engine at a stable temperature.*

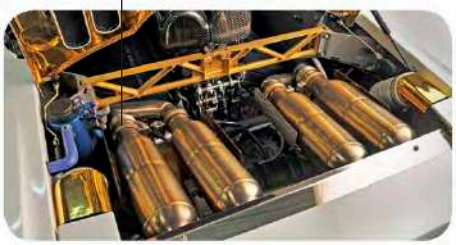

McLaren F1 car engine

Ancient gold jewellery

This neck ornament *is made from cast gold.*

THE HOLTERMANN NUGGET

The largest piece of natural gold was found on 19 October, 1872, near the small town of Hill End in Australia. Named after its discoverer, Bernhardt Holtermann, the piece contained moe than 90 kg (198 lb) of pure gold.

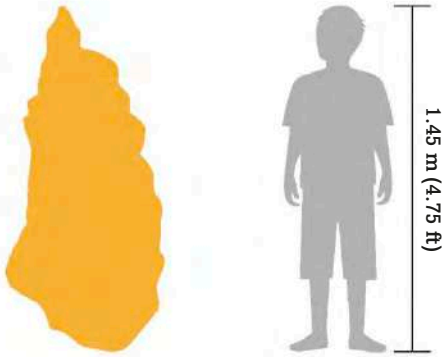

Holtermann Nugget Child aged 10 years old

gold dust with water or strong acids. The applications for gold include heat shields in **astronaut's visors**. This metal has always been seen as valuable and many ancient artefacts, such as the **3,300-year-old death mask** of Egyptian pharaoh Tutankhamun, were forged from it. Some of the earliest coins, found in Turkey, were made of it. Gold is used to cover important buildings, such as Thailand's **Wat Phrathat Doi Suthep** temple. This precious metal is most commonly used today in **jewellery** or decorations.

GOLDEN BUDDHA A precious statue of Buddha with one thousand eyes and one thousand hands stands in Long Son Pagoda, a temple in Nha Trang, Vietnam. The Buddha is depicted as holding a range of sacred objects, including scrolls and white lotus flowers. This statue is completely covered in a layer of pure gold, and it draws in hundreds of devotees from across the world.

Although humans have discovered many strong metals and useful elements, gold has remained one of the most valuable. Before people knew what it was, they saw glittering gold dust in river beds or dug large gold nuggets out from rocks. They found that gold has many valuable qualities: it is soft enough to hammer into any shape and can be melted down for moulding into ornaments. Best of all, its gleaming golden colour never fades away. Ancient cultures prized items made of gold: in ancient Egypt, gold was used to make coins as well as to cap the tops of pyramids. Gold is, however, so rare that if all the world's mined gold were forged into a cube, it would fit inside the penalty area of a soccer pitch.

80 Hg Mercury

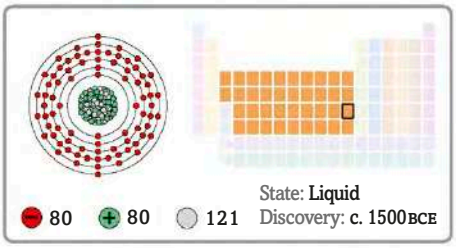

Forms

Solid mercury is soft enough to be **cut with a knife**.

Mercury is the only metal that is liquid at room temperature. Along with water, it is one of the few liquids found naturally on Earth's surface. **Pure mercury** forms around volcanoes where the heat separates it from its minerals, such as **cinnabar**. This red mineral has been used for many centuries: ancient Romans roasted cinnabar to release a liquid they called *hydrargyrum*, meaning "silver water". This was the element mercury. It was later known as quicksilver because of how fast it flowed as a stream of liquid. This metal is very poisonous:

Uses

This thermometer *contains mercury, which expands as it gets warmer and contracts as it cools.*

Mercury thermometer

Pills containing mercury *were commonly used to treat constipation and toothache.*

Mercury pills

This bulb *glows when the mercury vapour inside it is electrified.*

CFL bulb

Mercury has been in use for more than **4,000 years**.

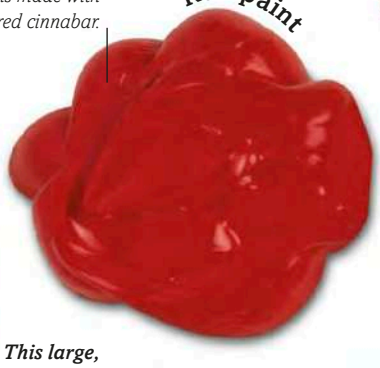

This bright red paint *is made with powdered cinnabar.*

Red paint

This large, low-cost mirror *made of a pool of pure mercury is used in an astronomical telescope.*

Liquid mirror inside telescope

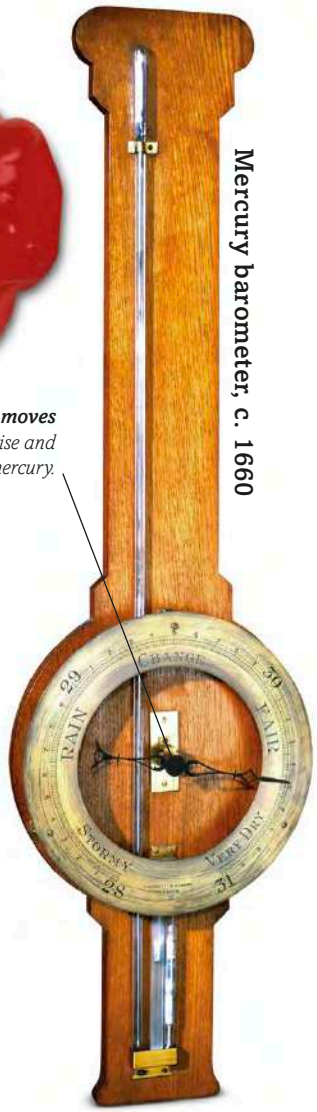

The needle moves *with the rise and fall of mercury.*

Mercury barometer, c. 1660

HOW DOES A BAROMETER WORK?

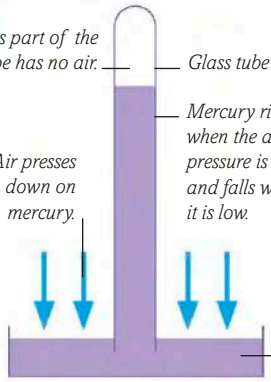

A barometer is an instrument used to measure air pressure to forecast the weather. The simplest – and earliest – designs used a column of mercury inside a glass tube.

it can damage organs and nerves if inhaled or swallowed. As a result, the use of this metal is carefully monitored today. Mercury is used in some batteries, some **thermometers**, and in low-energy, **compact fluorescent light (CFL) bulbs**. Its compounds are used to prepare strong, **red paints**. Until the early 18th century, mercury was used in **pills** for treating some common ailments. It gradually fell out of use when it was found to be toxic. The first accurate **barometers** also contained this liquid, but such devices are rarely seen today.

104 Rf Rutherfordium

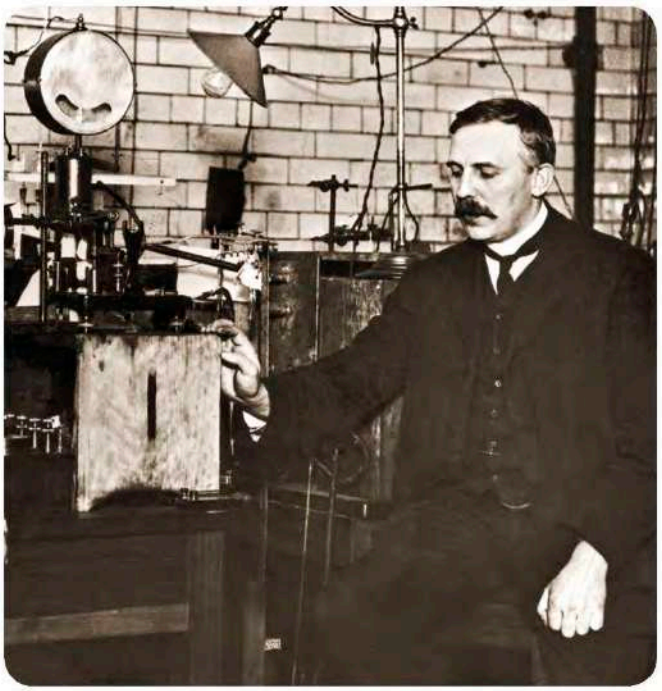

Ernest Rutherford

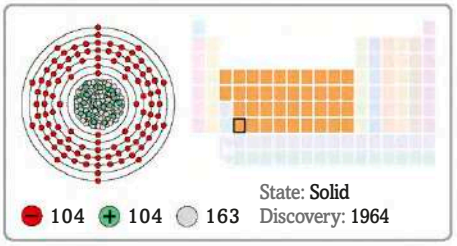

Rutherfordium was the first superheavy element to be discovered. In this type of element, each atom has 104 or more protons in its nucleus. It is named after the New Zealand scientist **Ernest Rutherford**, who, in 1913, suggested that every atom has a nucleus, or core. Pure rutherfordium is synthesized by researchers in a laboratory.

105 Db Dubnium

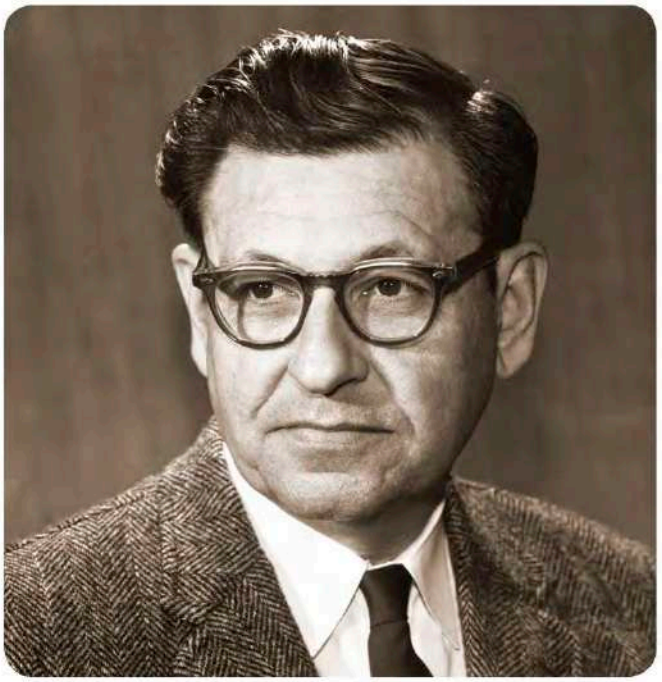

Albert Ghiorso

The US scientist Albert Ghiorso discovered **12 elements** in the 20th century.

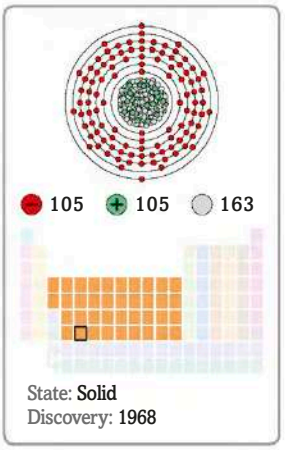

It took scientists nearly 30 years to agree on a name for this element. Dubnium was finally named after the Russian city of Dubna, where the first atoms of this artificial, radioactive element were created, in 1968. However, a team of American scientists led by **Albert Ghiorso** also produced samples of the element at the same time. This radioactive element has 12 isotopes, or forms, with different numbers of neutrons.

106 Sg Seaborgium

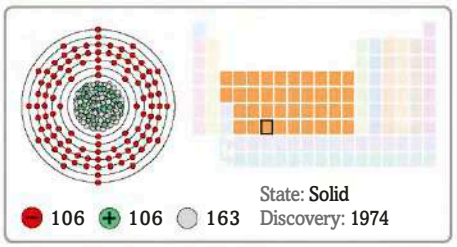

Atoms of seaborgium break apart in about three minutes, so little is known about it. Scientists think it may be a metal. The element was isolated in 1974 in a machine called the **Super Heavy Ion Linear Accelerator** at the Lawrence Berkeley National Laboratory. It was named after the US scientist Glenn T Seaborg.

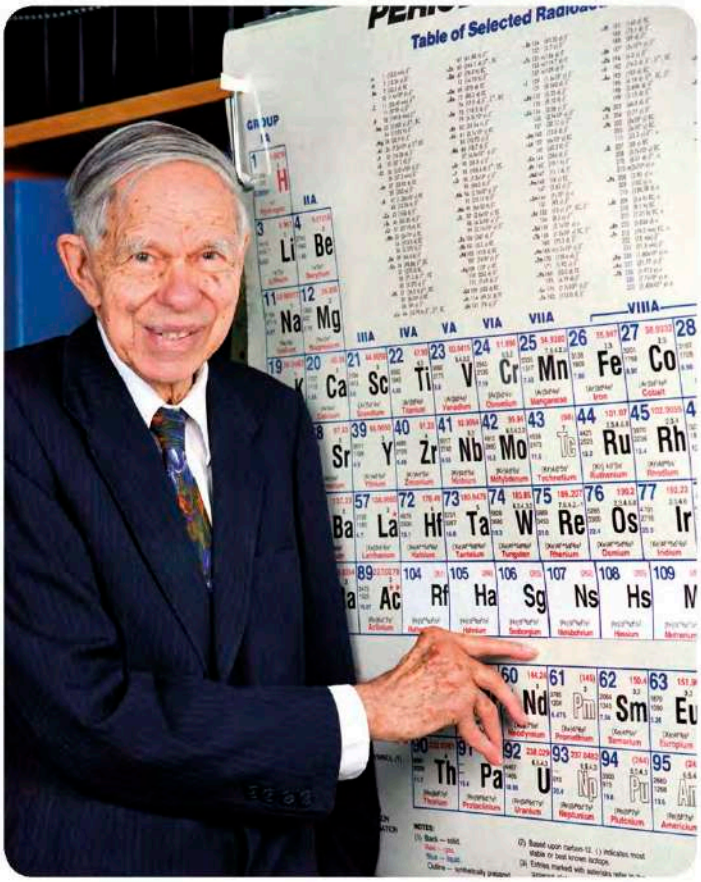

Glenn T Seaborg

This huge machine was used to discover **five new elements**.

This giant tube *forms part of the Super Heavy Ion Linear Accelerator, which is a type of particle accelerator – a machine in which atoms are smashed together.*

Super Heavy Ion Linear Accelerator, Lawrence Berkeley National Laboratory, California, USA

NOBEL PRIZE IN CHEMISTRY

Glenn T Seaborg and his fellow US researcher Edwin McMillan were awarded the Nobel Prize for Chemistry in 1951 for their work in creating neptunium. This was the first element to be isolated that was heavier than uranium – the heaviest natural element.

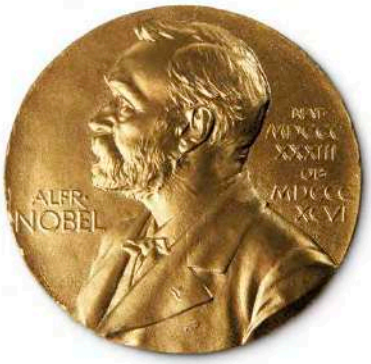

Nobel Prize medal

107 Bh Bohrium

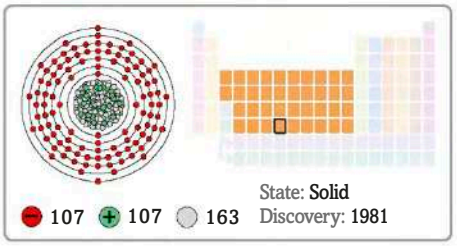

Niels Bohr

Bohrium is an artificial element named after the Danish scientist Niels Bohr. This was to honour his model of the structure of atoms' electron shells. Bohrium was first produced by firing chromium atoms at bismuth atoms in a particle accelerator (a machine in which atoms are smashed together). Atoms of this metal are unstable: half of any sample of bohrium atoms breaks apart in 61 seconds. As a result, it is not very well understood.

108 Hs Hassium

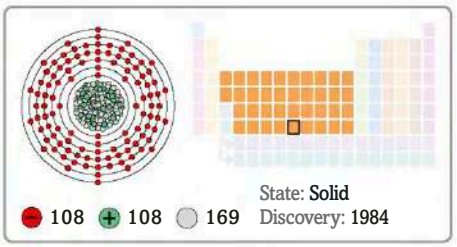

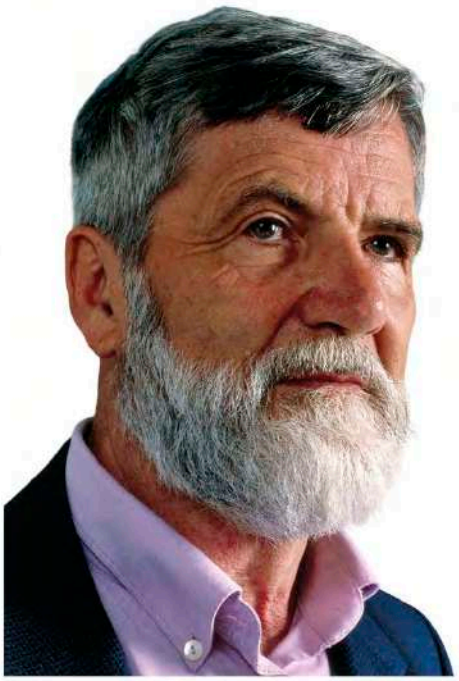

Peter Armbruster

Hassium was produced inside this chamber.

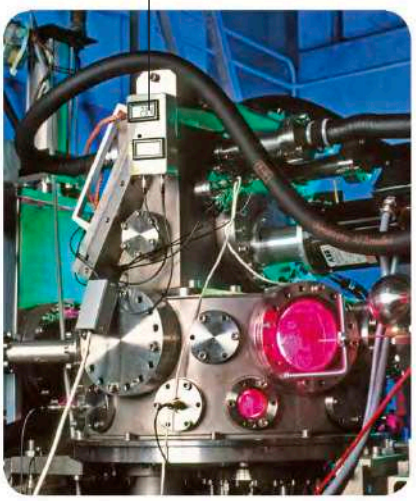

A chamber at Centre for Heavy Ion Research, Darmstadt, Germany

Scientists think hassium is a metal, but they have not been able to produce enough of its atoms to study it in any detail. Hassium is very radioactive, and most of its atoms break apart within a few seconds. This element is named after the German state of Hesse, the location of the **Centre for Heavy Ion Research**, where hassium was first created artificially by a team led by the German physicist **Peter Armbruster**.

109
Mt

Meitnerium

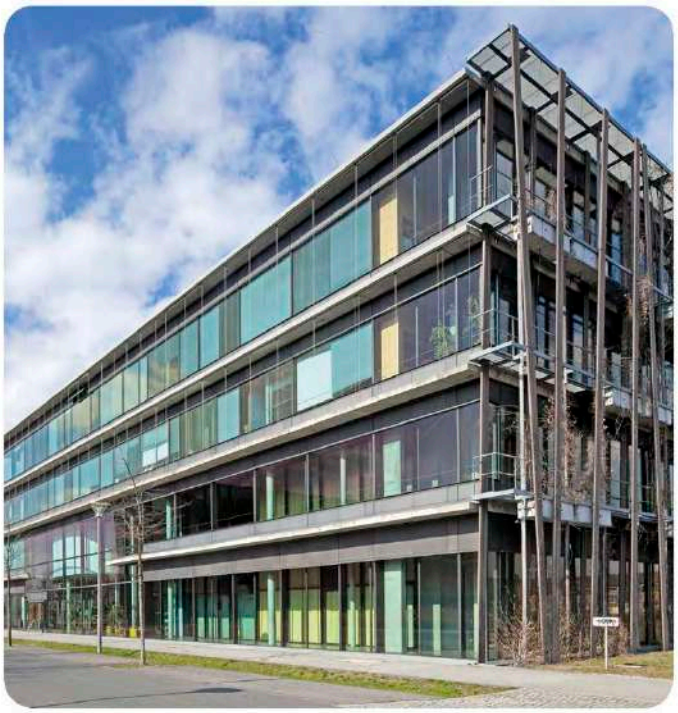

Meitner Haus building,
Humboldt University, Germany

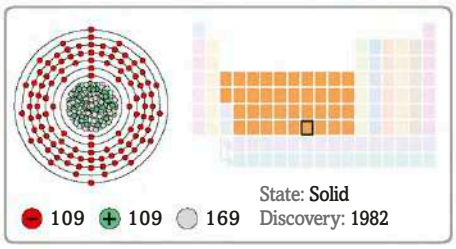

Researchers think meitnerium might be the densest of all elements. It is very unstable, and even the atoms of its most stable isotope, or form, break apart in a matter of seconds. Meitnerium is named after the Austrian physicist **Lise Meitner**, to honour her achievements in physics. Several universities, such as **Humboldt University** in Berlin, Germany, also have buildings in her name.

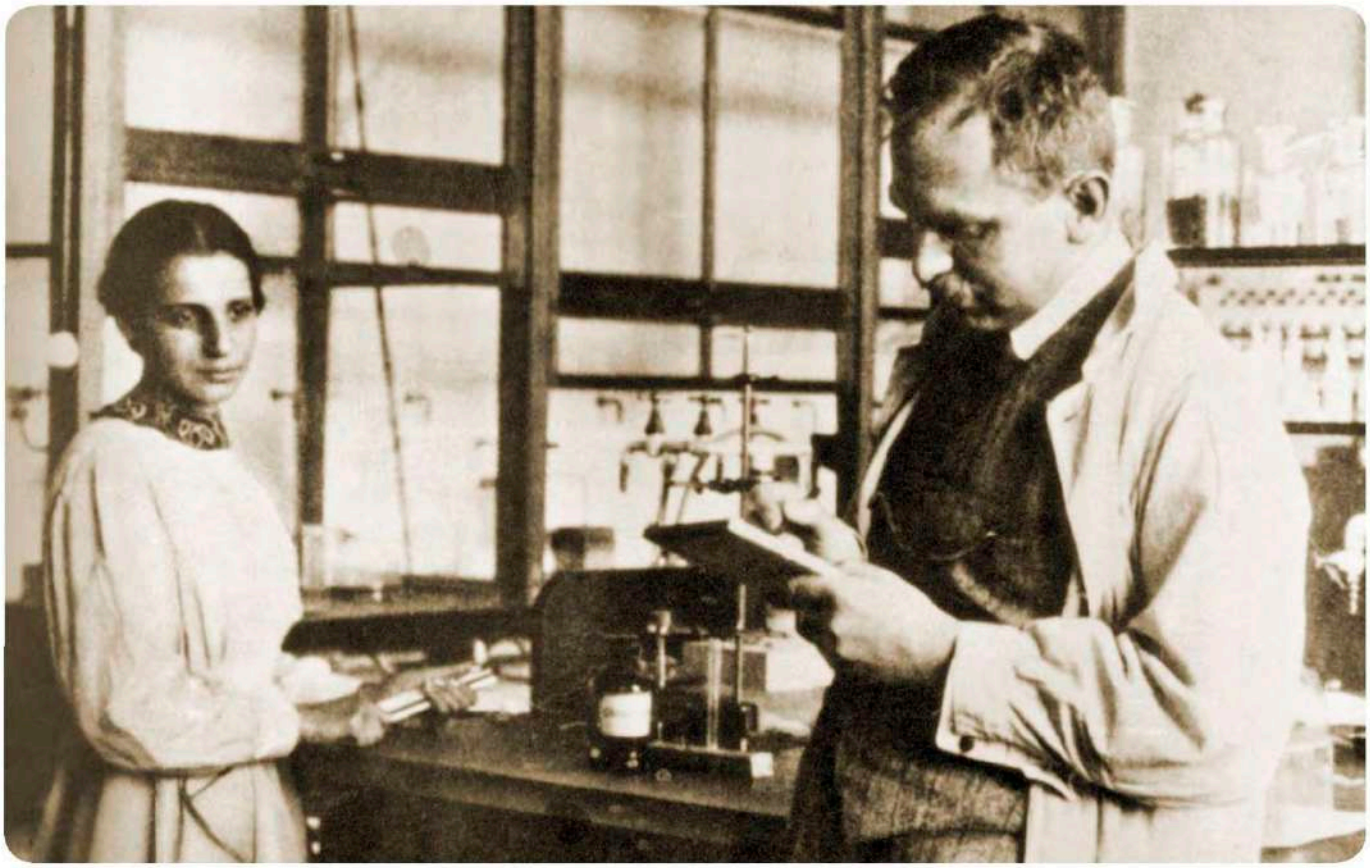

Lise Meitner (left) works with the German chemist Otto Hahn

110 Ds Darmstadtium

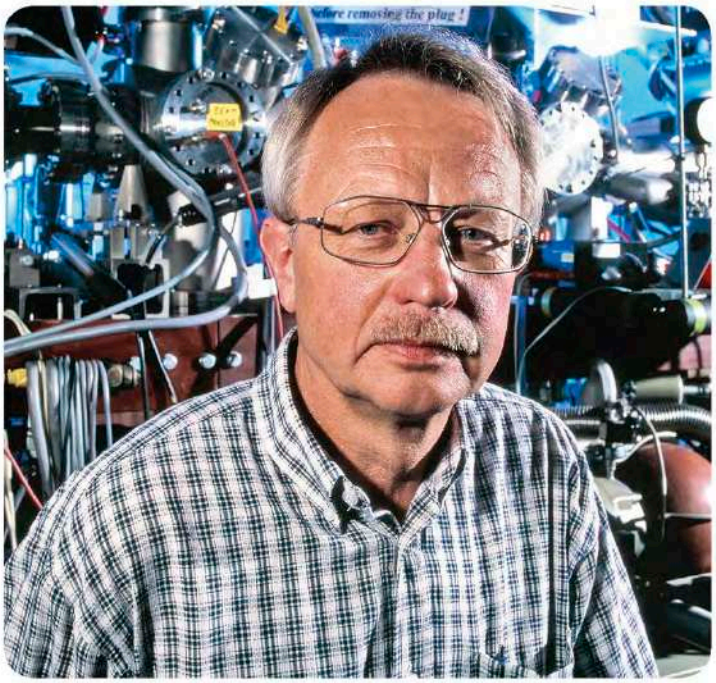

Sigurd Hofmann

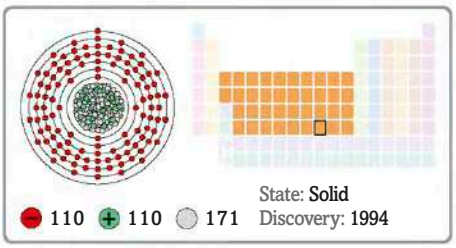

This artificial element is named after the German city of Darmstadt – the home of the Institute for Heavy Ion Research where this element was first produced. A team led by the German physicist **Sigurd Hofmann** created darmstadtium by smashing nickel atoms into lead atoms in a particle accelerator (a machine in which atoms are smashed together).

111 Rg Roentgenium

Wilhem Röntgen

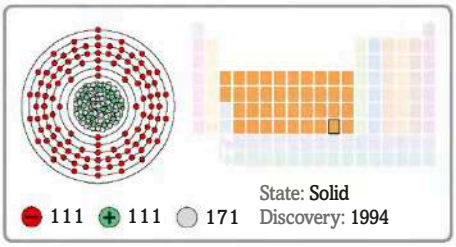

Scientists believe that this metal shares many characteristics with precious metals, such as gold and silver. However, its atoms break apart within seconds, so this has not yet been confirmed. Roentgenium was created in Darmstadt, Germany. It was named after **Wilhelm Röntgen**, the German scientist who discovered X-rays in 1895.

112
Cn

Copernicium

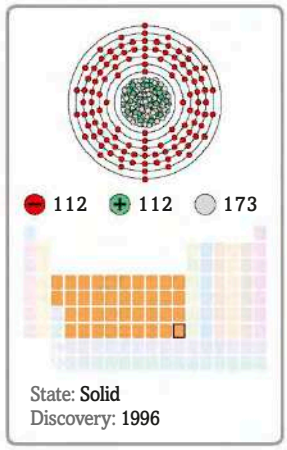

Some scientists think copernicium could be the **only gaseous metal**.

This statue *stands in front of the Polish castle in which Copernicus lived.*

Statue of Nicolaus Copernicus

This German research institute *is where copernicium was discovered.*

Heavy Ion Research Centre, Germany

The atoms of this radioactive element survive only for a few minutes, before breaking down. Copernicium is created in particle accelerators by smashing together atoms of lead and zinc. Only a few atoms of this artificial element have ever been produced. Copernicium is named after **Nicolaus Copernicus**, the Polish astronomer who theorized that our planet orbits the Sun.

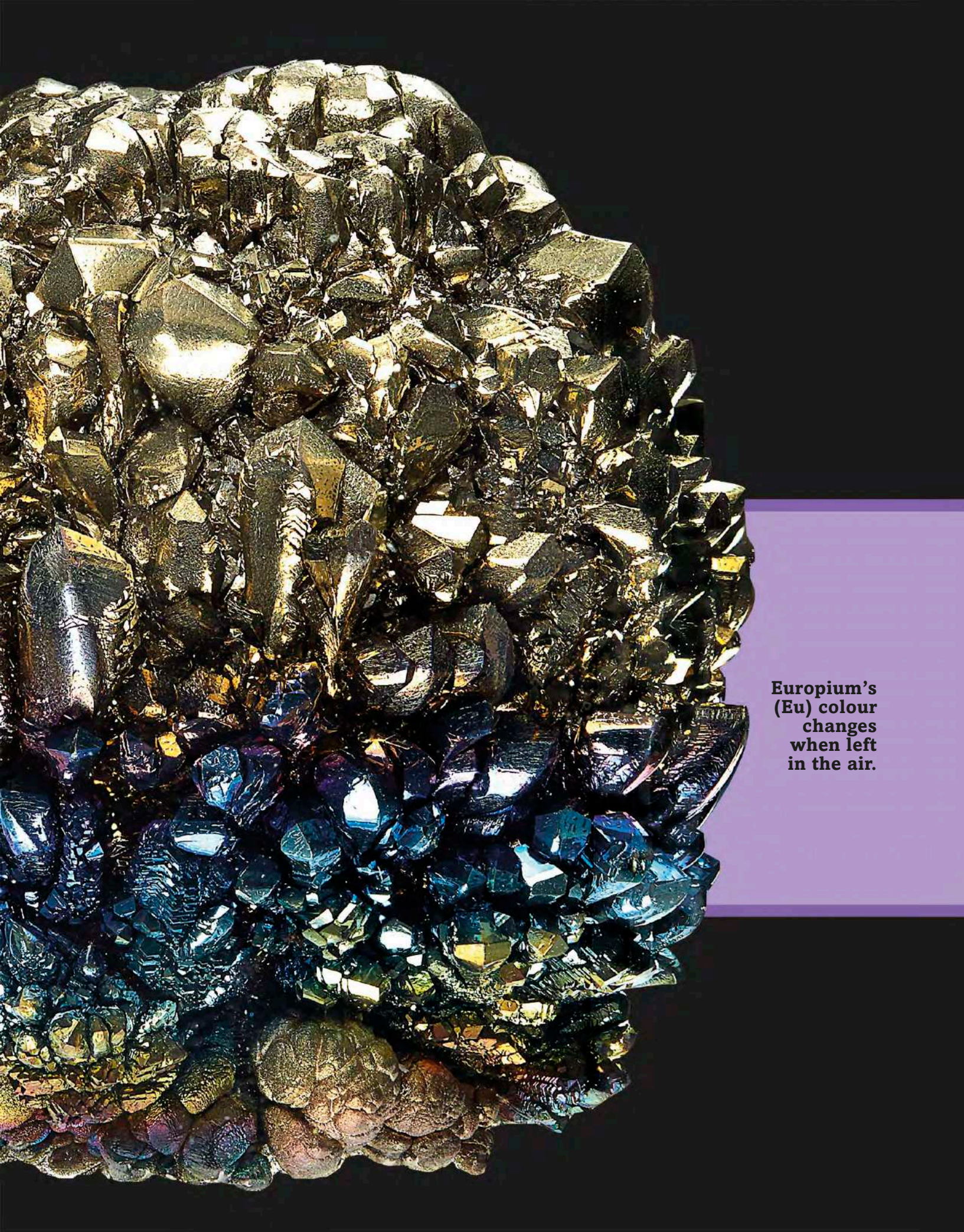

Europium's (Eu) colour changes when left in the air.

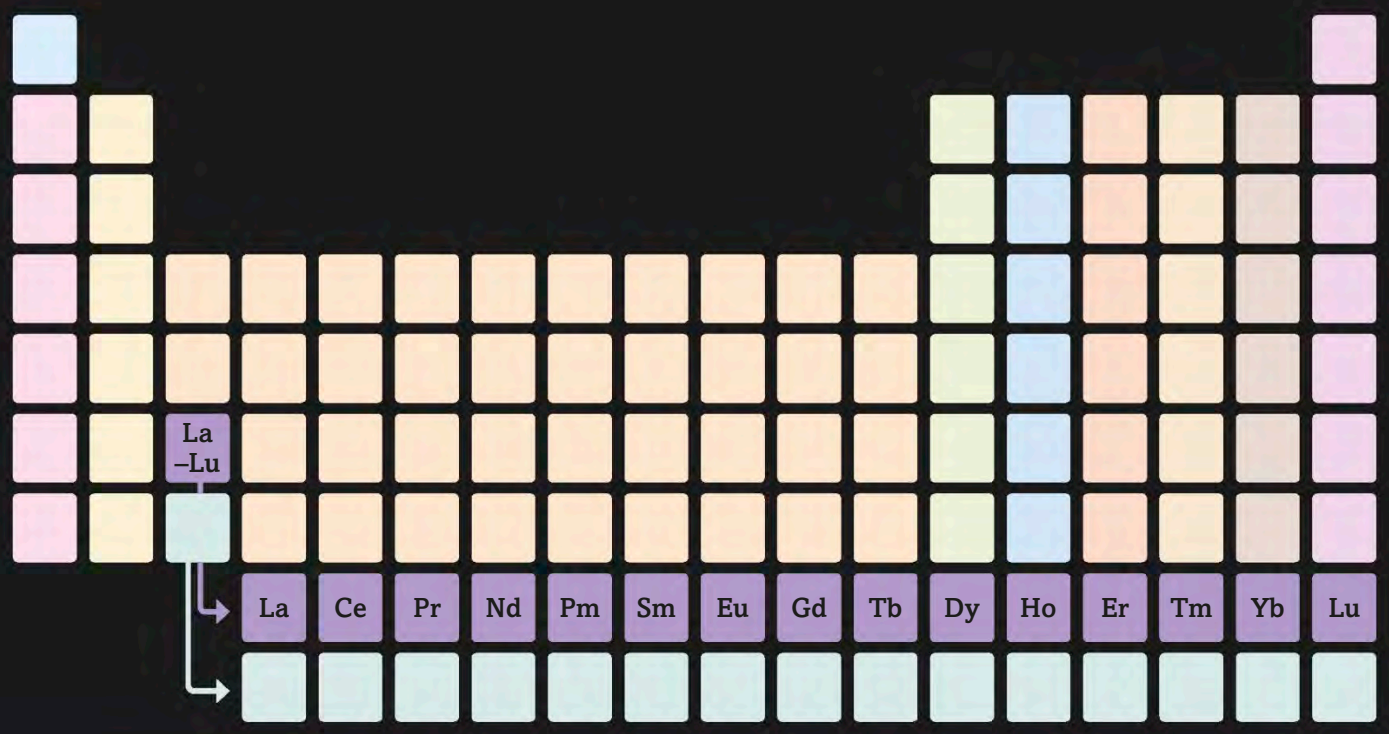

Lanthanides

This set is named after lanthanum, the first element in this series. The name "rare earth metals" is also given to these elements because they were discovered mixed together in complex minerals in Earth's crust, and were thought to be uncommon. However, they are actually not rare but abundant. These metals – between barium (Ba) and hafnium (Hf) – should fit between the alkaline earth metals and the transition metals, but they are normally shown underneath the main table to save on space.

Atomic structure
Atoms of every element in this group have two outer electrons. The lanthanides have large atoms, all with six electron shells.

Physical properties
The lanthanides are dense, shiny metals, which tarnish easily when exposed to air. They do not conduct electricity very well.

Chemical properties
These elements react slowly with oxygen (O) at room temperature, but the reactions speed up when heated.

Compounds
Many lanthanides form compounds with oxygen called oxides. These are often used in lasers and magnets.

57 La Lanthanum

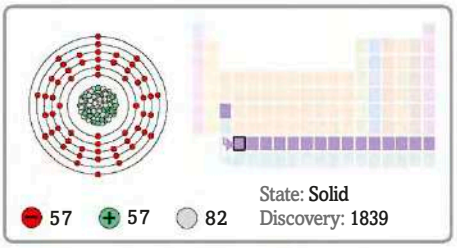

Forms

Lanthanum carbonate is used to treat patients with **kidney disease**.

Bastnasite

This reddish-brown mineral *is also found in other colours, including white, tan, and grey.*

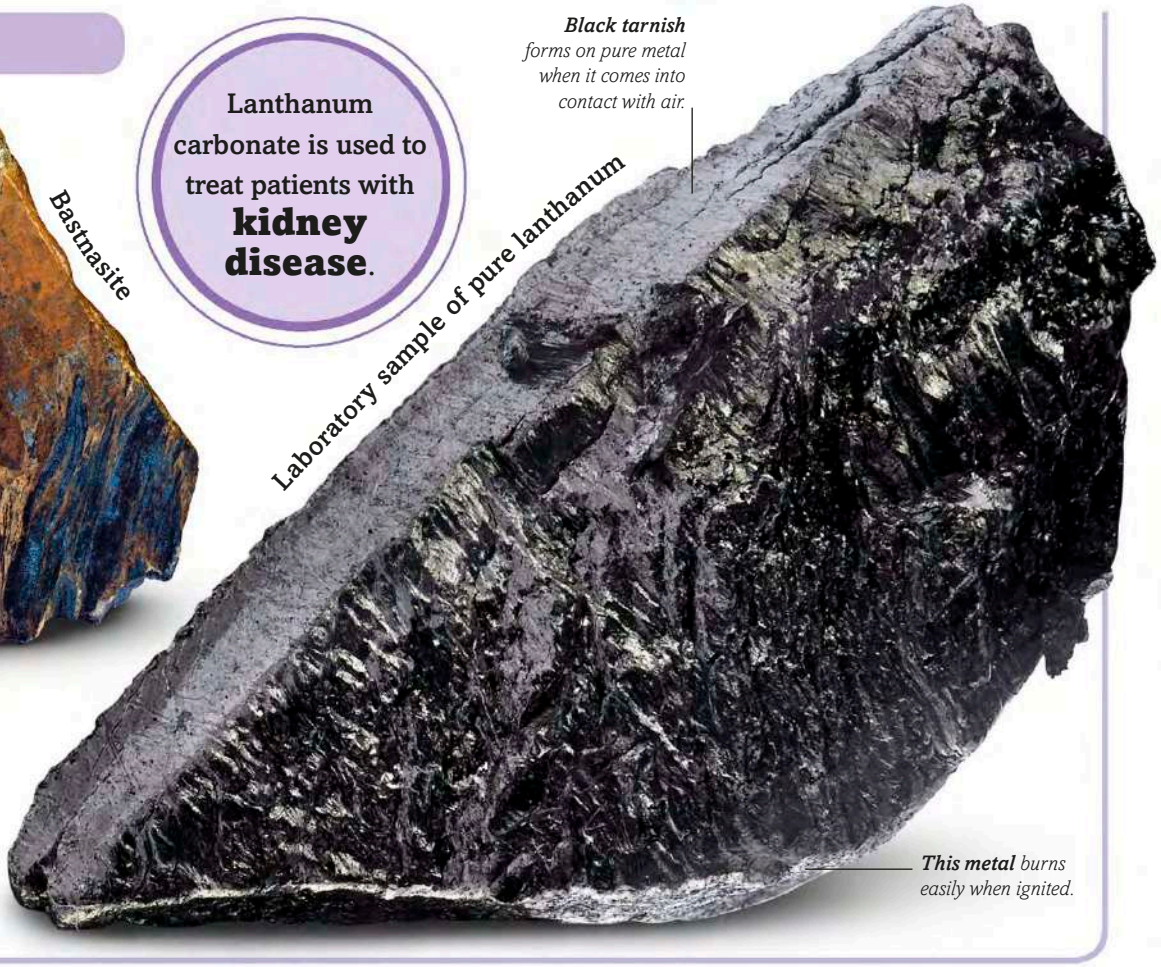

Laboratory sample of pure lanthanum

Black tarnish *forms on pure metal when it comes into contact with air.*

This metal *burns easily when ignited.*

Uses

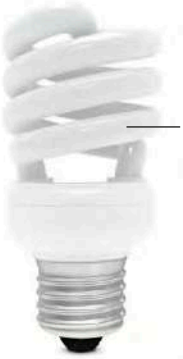

Fluorescent lamp

This lamp *uses lanthanum to reduce the yellow colour in its light.*

Molten lanthanum

In its molten state*, lanthanum is used to smooth rough diamonds.*

This lens *can better focus light on an object due to the presence of lanthanum oxide in the glass.*

Camera lens

Although the word "lanthanum" means "to lie hidden", it is more abundant than most metals. For example, it is three times more common than lead. This element was discovered in the mineral cerite in 1839. However, it took chemists almost another 100 years to find a way to purify the metal. Today, the mineral **bastnasite** is a source of **pure lanthanum**. The element's applications range from its use in film studio lights and **lens-making** to refining petroleum.

58 Ce Cerium

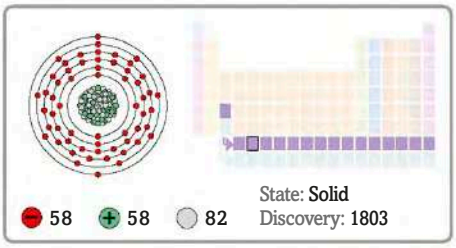

Cerium was the first of the lanthanides to be discovered. It is named after the dwarf planet Ceres, which was discovered two years before the element was isolated. Cerium is highly toxic when **pure**, but safer cerium compounds have some uses. The main use of cerium is in making phosphors, which are chemicals that produce lights of different colours. Phosphors are present in **flatscreen TVs** and bulbs.

Television

The inside of this screen *is coated with cerium-containing phosphors, which emit red, green, and blue light.*

Laboratory sample of pure cerium

The pure form *of the metal tarnishes on contact with air.*

Kitchen spatula

This red colour *comes from a compound called cerium sulfide.*

59 Pr Praseodymium

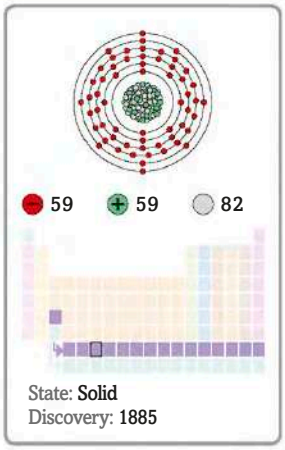

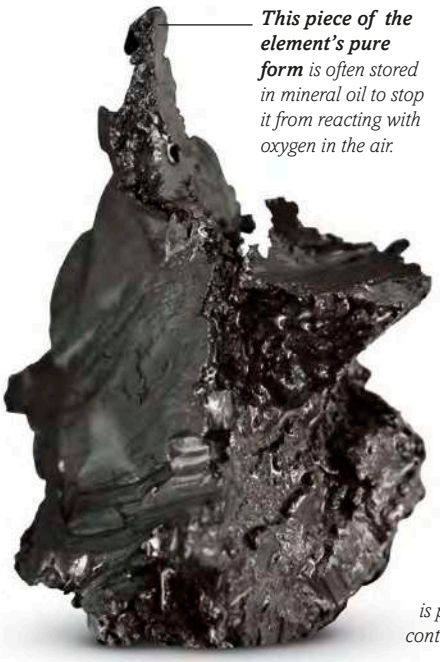

This piece of the element's pure form *is often stored in mineral oil to stop it from reacting with oxygen in the air.*

Laboratory sample of pure praseodymium

Part of this element's name comes from *prasinos*, the Greek word for "green". Normally a grey colour when pure, the element reacts slowly on contact with air to form a green coating. Praseodymium compounds give a yellow colour to glass and heat-resistant **ceramics**, and provide a green colour to some **artificial jewels**. This element also boosts the strength of magnets that contain it.

Yellow ceramic pot

This yellow colour *is produced by a solution containing praseodymium.*

Green cubic zirconia

This artificial gem *gets its green colour from tiny amounts of a compound of praseodymium and oxygen.*

60 Nd Neodymium

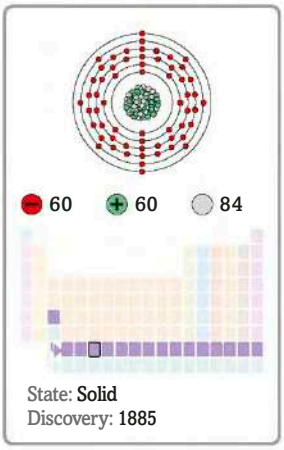

Pink glass

This glass *gets its colour from very tiny amounts of neodymium.*

Strong magnets made of neodymium can be used to lift thousands of times their own mass. This element was discovered in 1885 by the Austrian chemist Carl Auer von Welsbach, and it was originally used to colour glass. Small amounts of neodymium turn **glass** pinkish purple. Today, this element is also employed in lasers used in eye surgery.

Laboratory sample of pure neodymium

The pure element *turns black when it reacts with air.*

61 Pm Promethium

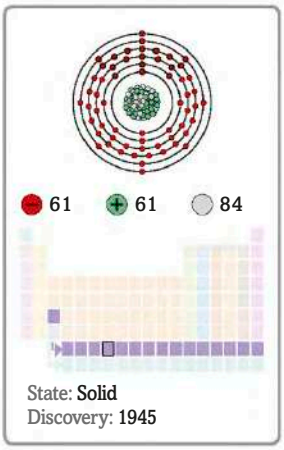

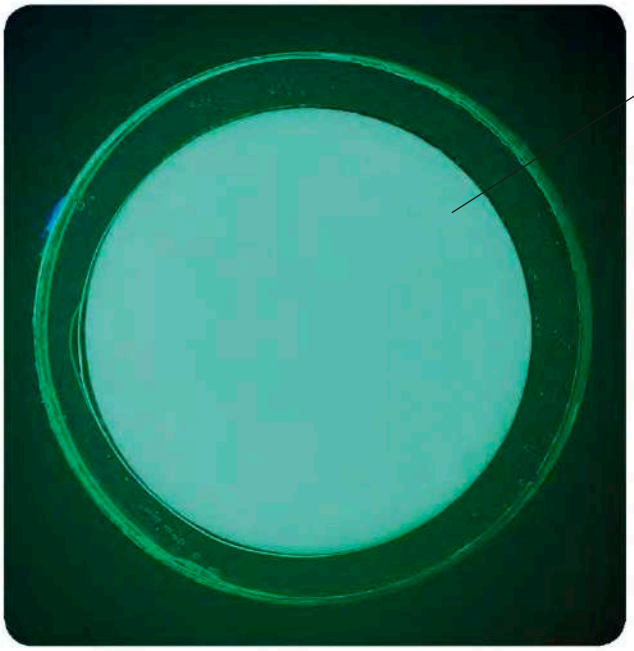

This paint *glows as a result of radioactive promethium.*

Promethium-rich paint in a tin seen from above

This missile *uses radioactive promethium for electrical power.*

Missile

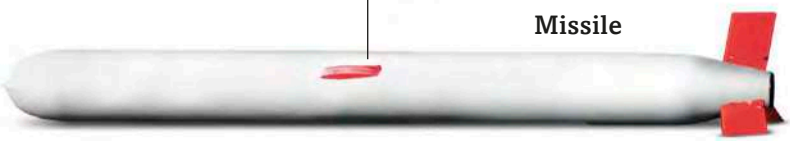

Promethium is the rarest lanthanide element. Any promethium that was in Earth's rocks decayed billions of years ago. Promethium is therefore produced artificially in nuclear reactors. Being very radioactive, it is used in some **missiles**, because it converts this radioactivity into electrical power. The addition of promethium also makes some **paints** glow in the dark.

62 Sm Samarium

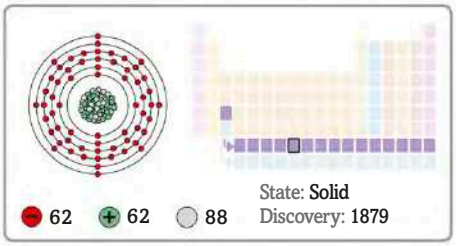

This element is named after the mineral samarskite from which it was first purified. However, another lanthanide-rich mineral called monazite is the main source of this element today. Samarium is mixed with cobalt to make permanent magnets that are often used in **electric guitars**.

These pickups (components that sense vibrations produced by guitar strings) are made of samarium-cobalt magnets.

This silvery white metal darkens on contact with air.

63 Eu Europium

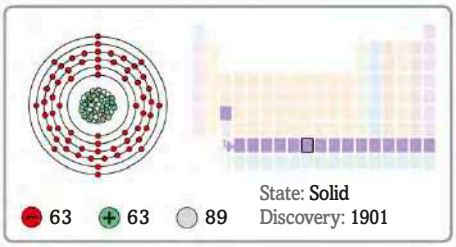

The crystals of this yellowish metal often have patches of dark oxides.

Europium was named after the continent of Europe. However, most of the world's supply of the element comes from the USA and China, where the mineral bastnasite is mined for the extraction of **pure europium**. A compound called europium oxide is used in euro and **British bank notes**. When placed under ultraviolet (UV) light, the compound gives off a red glow.

This red glow proves this note is real.

Section of British note under UV light

64 Gd Gadolinium

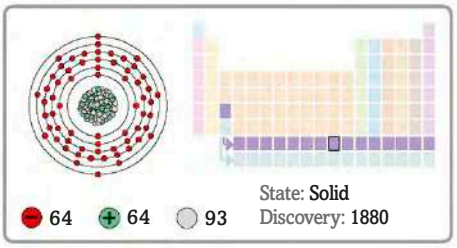

Gadolinium, and its mineral ore gadolinite, are named after the Finnish chemist Johan Gadolin, who discovered the element. Gadolinium compounds are used to obtain clear **MRI scans**. It is also used in electronics, and to make rust-resistant steel.

Laboratory sample of pure gadolinium

This soft, silvery metal *darkens when exposed to air.*

Gadolinite

This mineral *contains tiny amounts of gadolinium.*

This image of the brain *is clear because a gadolinium compound was injected into the patient's blood.*

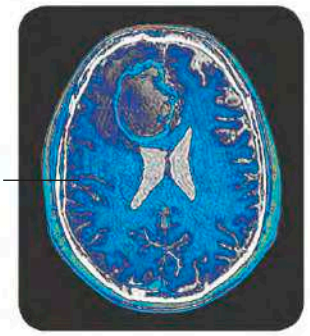

MRI scan of the human brain

65 Tb Terbium

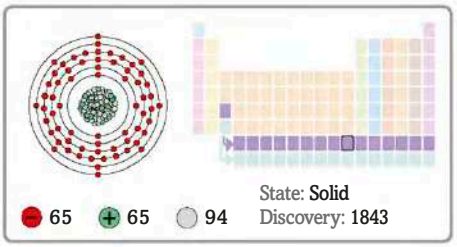

Terbium is named after the village of Ytterby in Sweden. It is a silvery metal that can be obtained from the ore monazite. This element has only a few uses. **Pure terbium** is added to other metals to make powerful magnets used in sound-producing devices, such as the **SoundBug™**. Its compounds are used to line **mercury lamps**.

SoundBug™ device

This device *uses magnets to turn any flat surface, like a window, into a loudspeaker.*

The mercury vapour in this lamp *produces ultraviolet light when electrified, and this is turned into a bright yellow glow by terbium.*

Mercury lamp

The pure metal *is soft enough to be cut with a knife.*

Laboratory sample of pure terbium

66 Dy Dysprosium

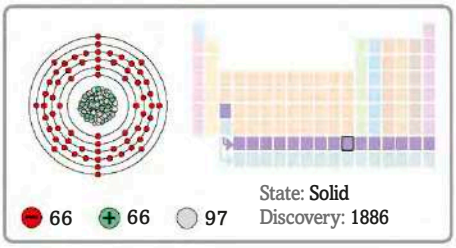

This mineral contains tiny amounts of dysprosium.

Dysprosium reacts more easily with air and water than most other lanthanide metals. Although it was discovered in 1886, it took until the 1950s to purify it. This metal is often used with neodymium to produce magnets that are used in **car batteries**, wind turbines, and generators.

This pure metal remains shiny at room temperature.

Laboratory sample of pure dysprosium

Some hybrid car batteries contain dysprosium.

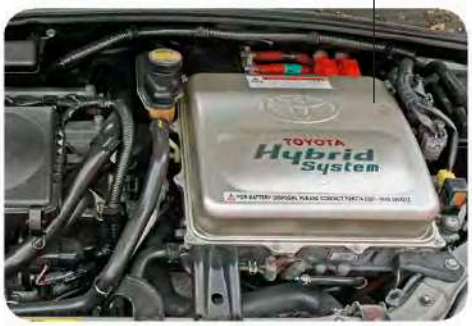

Hybrid car battery

67 Ho Holmium

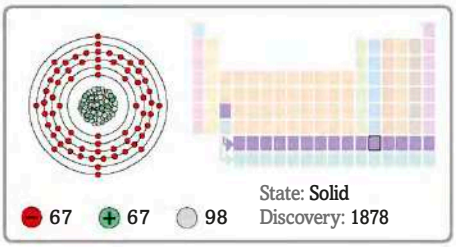

The Swedish chemist Per Teodor Cleve named holmium after the Swedish city of Stockholm. **Pure holmium** can produce a strong magnetic field and is therefore used in magnets. Its compounds are used to make lasers, and to colour glass and artificial jewels, such as **cubic zirconia**.

This artificial gemstone is coloured red by small amounts of holmium.

Red zirconia gemstone

Bright, silver shine

Laboratory sample of pure holmium

68 Er Erbium

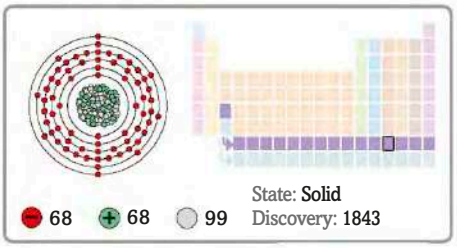

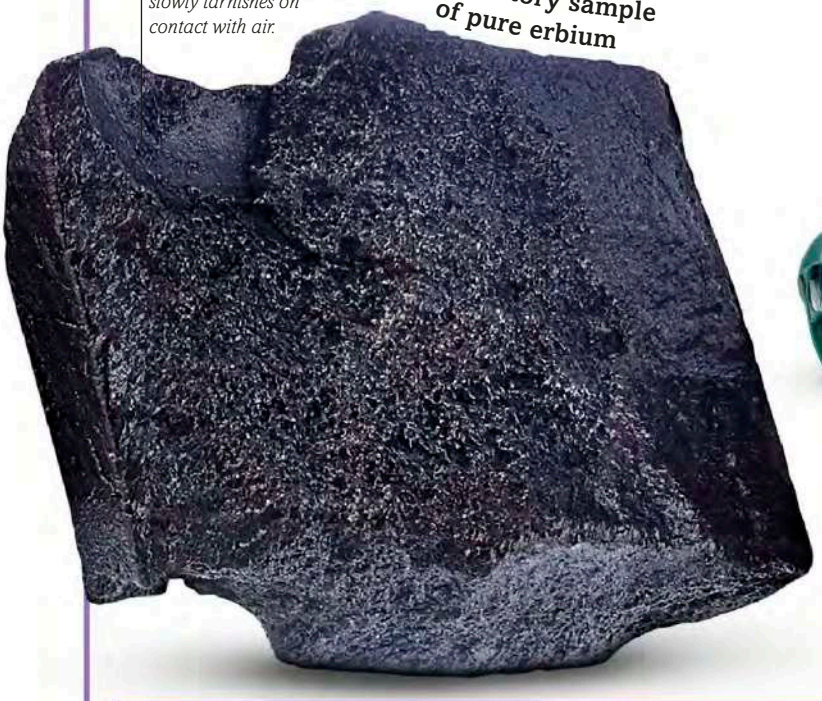

Laboratory sample of pure erbium

This silvery element *slowly tarnishes on contact with air.*

This glass *contains erbium, which protects a welder's eyes from heat and bright light.*

Welding goggles

The rose pink finish *of this vase is from an erbium chloride glaze.*

Pink pottery

Like terbium and ytterbium, erbium is also named after the Swedish village of Ytterby, near which it was discovered. This element does not occur in its **pure form** in nature, but it can be obtained from the mineral monazite. Many erbium compounds are pink in colour and are used to colour **pottery** and glass.

69 Tm Thulium

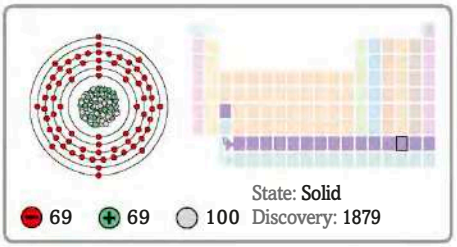

Laboratory sample of pure thulium

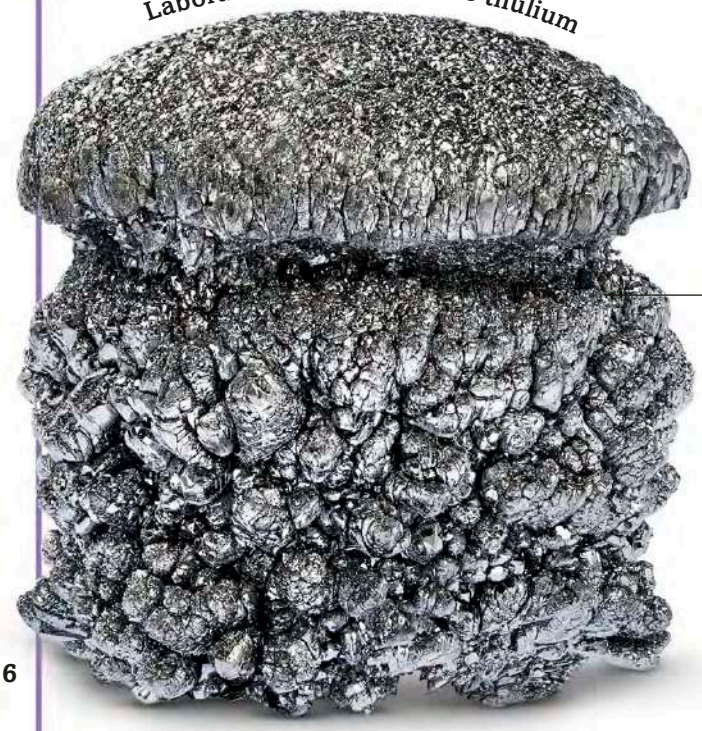

This soft metal *glows blue under ultraviolet (UV) light.*

This machine *emits X-rays using a very small amount of thulium.*

Portable X-ray machine

Thulium is the least abundant of all the lanthanide metals. It is used to create lasers that surgeons use to cut away damaged body tissue. Thulium also has a radioactive form that can produce X-rays: **portable X-ray machines** make use of this form.

70 Yb Ytterbium

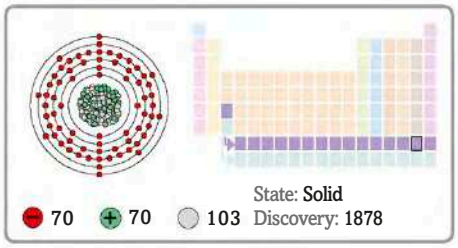

Ytterbium tends to be more reactive than other lanthanide metals. It is stored in sealed containers to stop the metal from reacting with oxygen. The **pure metal** has only a few uses. A small amount of ytterbium is used in making steel, while its compounds are used in some **lasers**.

Laser cutting

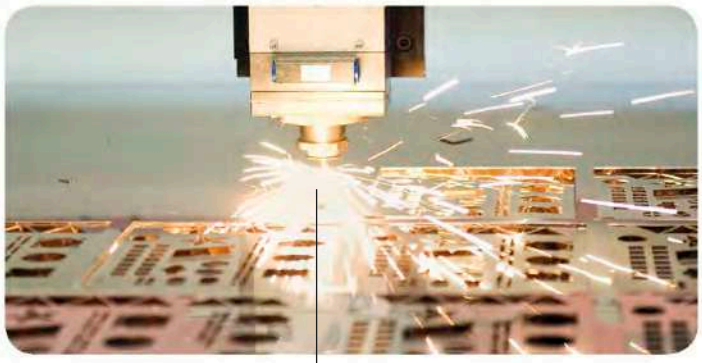

An ytterbium laser *can cut through metals and plastics.*

This bright, shiny metal *can be hammered into thin sheets.*

Laboratory sample of pure ytterbium

71 Lu Lutetium

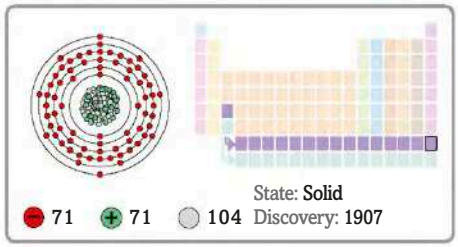

Laboratory sample of pure lutetium

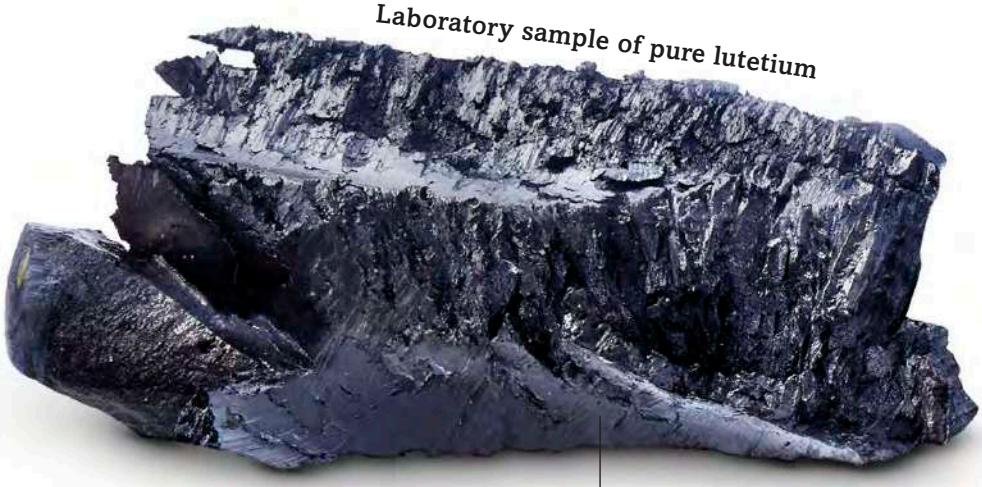

This element *is the hardest and densest lanthanide metal.*

Some oil refineries *use lutetium to break down crude oil to make fuels, such as petrol and diesel.*

Oil refinery

Lutetium was the last of the rare earth metals to be discovered. It is also the final member of the lanthanides. In its **pure form**, lutetium is very reactive and catches fire easily. It is rare and has few uses, mainly as a substance mixed with crude **oil**.

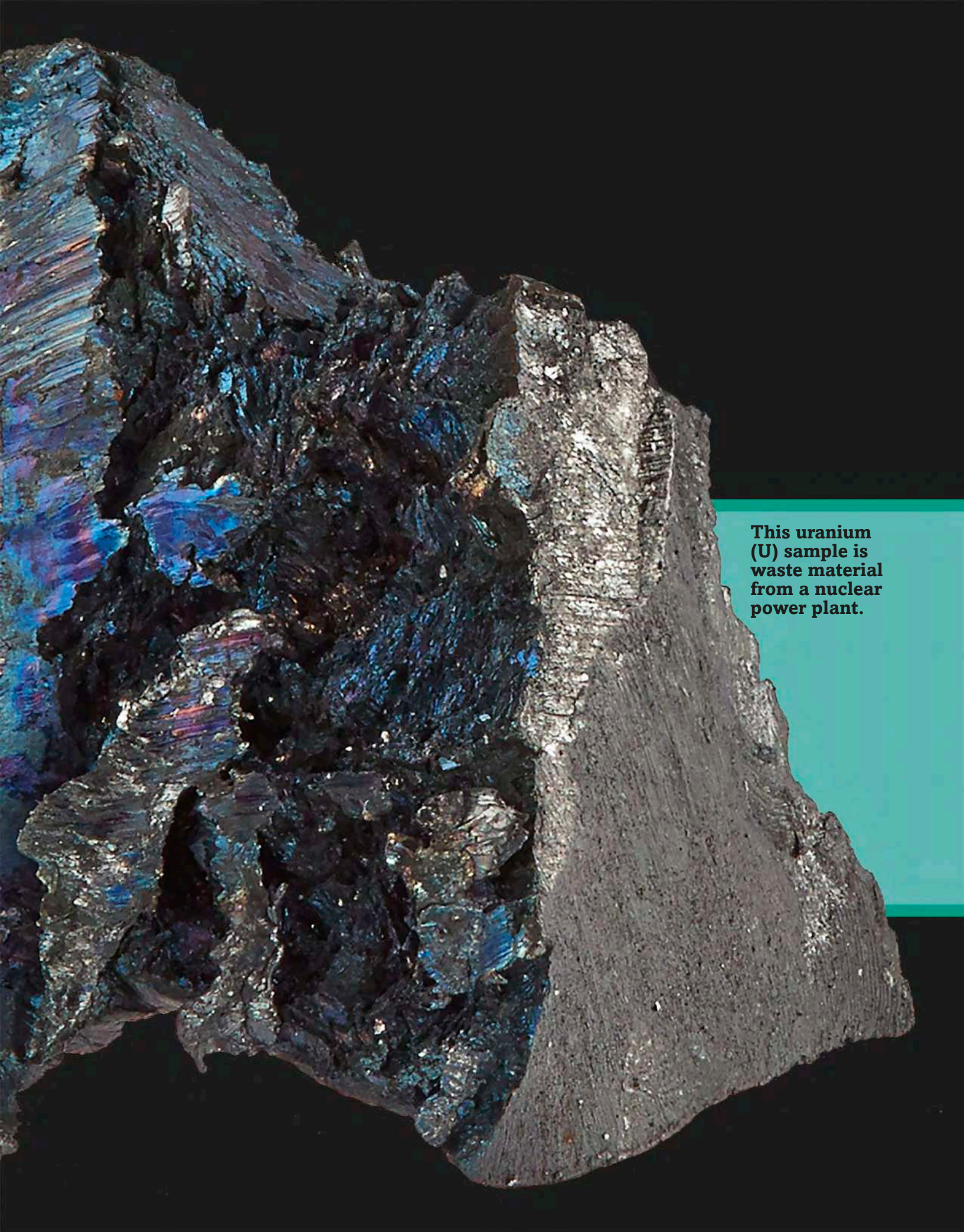

This uranium (U) sample is waste material from a nuclear power plant.

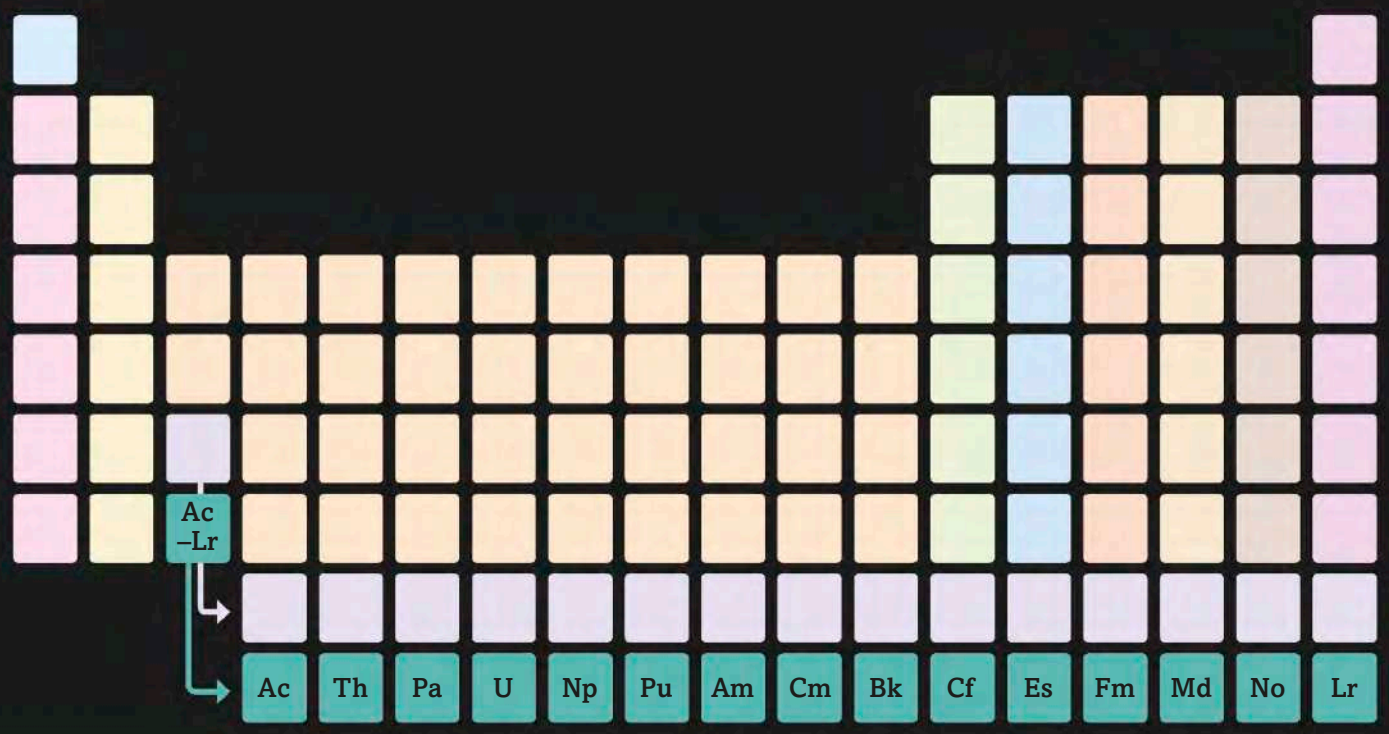

Actinides

These metals are named after actinium (Ac), the first member of the group. Although this group is often shown as the bottom row in the periodic table, to save space, they actually sit between radium (Ra), an alkali earth metal, and Rutherfodium (Rf), a transition metal. All the elements in this group are radioactive, and the final nine members are artificially produced in laboratories.

Atomic structure
All the elements in this group have two electrons in their outer shell. Their atoms all have seven electron shells.

Physical properties
Natural actinides are dense metals with high melting points. The physical properties of most of the artificial ones are unknown.

Chemical properties
The actinides are reactive metals and are never found in pure form in nature. They react easily with air, the halogens, and sulfur (S).

Compounds
Actinides form colourful compounds with halogens. Most actinide ores also contain compounds of oxygen (0) called oxides.

89 Ac Actinium

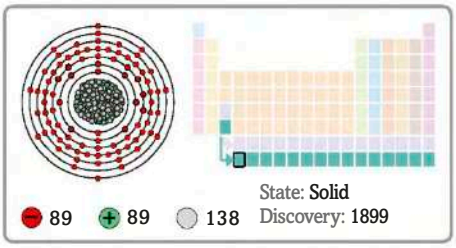

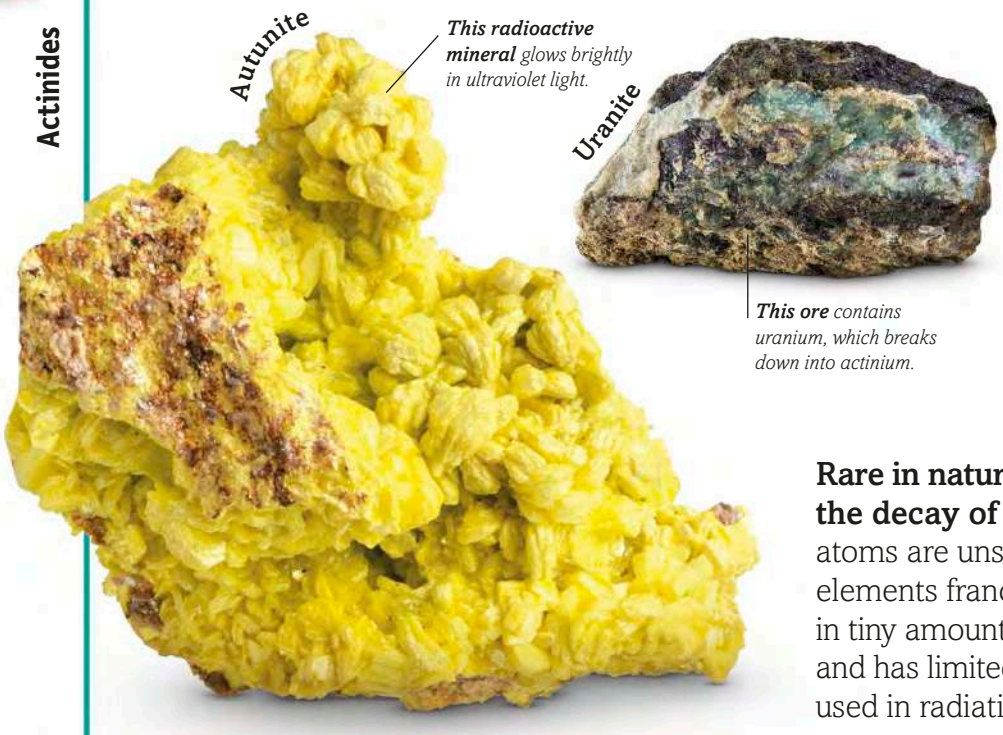

Neutron probe

Rare in nature, actinium is a metal formed by the decay of other radioactive elements. Its atoms are unstable and break down to make the elements francium and radon. Actinium is found in tiny amounts in uranium ores, such as **uranite**, and has limited applications. Its isotopes are used in radiation therapy to treat cancer.

90 Th Thorium

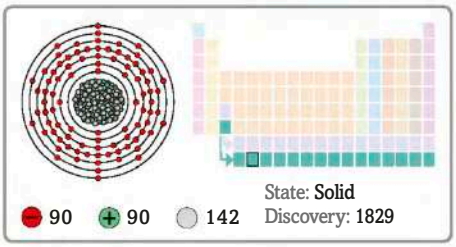

The most common natural radioactive metal, thorium is used inside vacuum tubes to allow an electric current to flow. It can also undergo nuclear fission, a process in which atoms split in two and release energy. Scientists are exploring ways of making thorium-powered nuclear reactors that produce electricity.

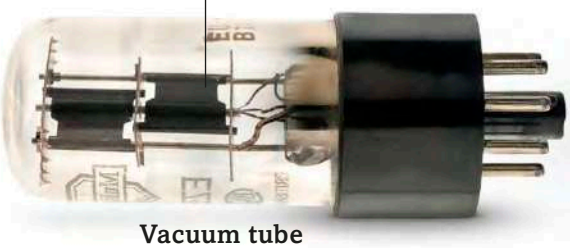

Vacuum tube

91
Pa

Protactinium

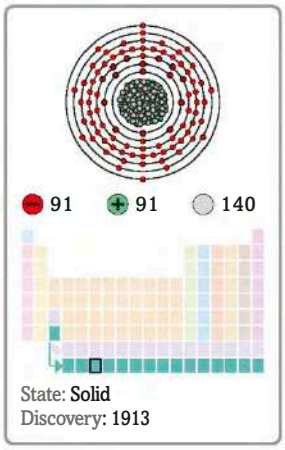

This vibrant green radioactive mineral *contains tiny amounts of protactinium.*

This brittle, shiny ore *feels waxy.*

Torbernite

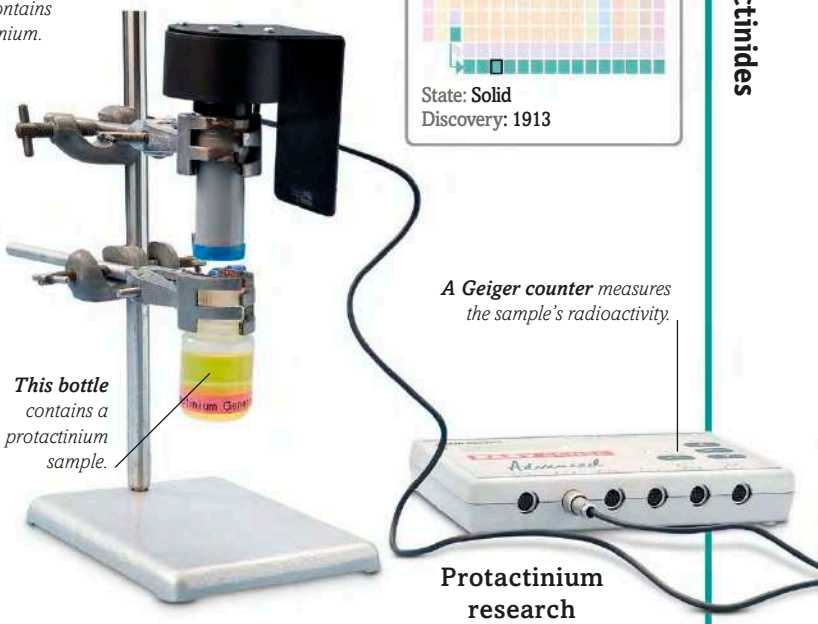

This bottle *contains a protactinium sample.*

A Geiger counter *measures the sample's radioactivity.*

Protactinium research

The name protactinium means "before actinium". This is because a uranium atom decays to form a protactinium atom, which then quickly breaks down into an actinium atom. Small quantities of protactinium are found in ancient sands and mud. Geologists use Geiger counters to carry out **research** to calculate how old the sands are.

These used nuclear fuel rods *contain protactinium.*

Nuclear waste

92 U Uranium

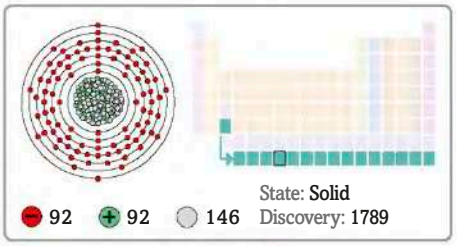

Named after the planet Uranus, uranium was the first known radioactive element. In the early 20th century, some manufacturers used uranium in **glass bowl** glazes, only to realize later that it was a harmful metal. An unstable form, called uranium-235, is used as fuel in nuclear reactors and in atomic bombs.

This sample of pure uranium *is waste from a nuclear plant.*

These black sections *contain uranium dioxide, which is the main source of uranium.*

Uraninite

Uranium mixed into glass *makes this bowl glow bright green under ultraviolet (UV) lamps.*

Glass bowl

93 Np Neptunium

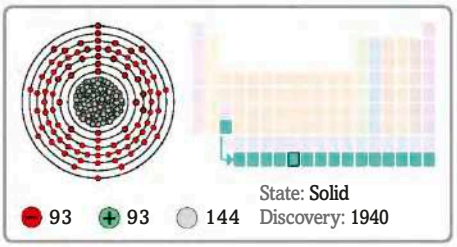

The radioactive elements *in this mineral decay to form neptunium.*

Uraninite

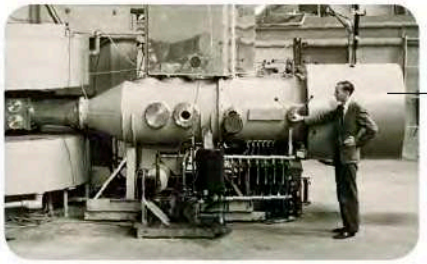

This cyclotron*, built in 1938, was used to discover neptunium.*

Cyclotron at the University of California, Berkeley, USA

Sitting next to uranium in the periodic table, neptunium was named after the planet Neptune. It exists in small amounts in radioactive ores, such as aeschynite. It forms during nuclear explosions and was first identified inside a machine called a **cyclotron**. There are no known uses for neptunium.

94 Pu Plutonium

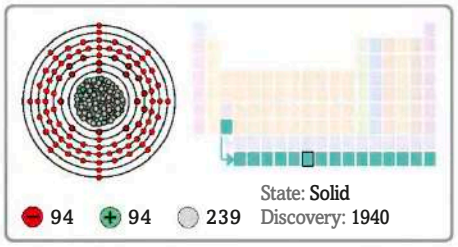

Hardly any plutonium exists in nature: most of it has decayed into other elements over time. It was discovered during the development of nuclear bombs in World War II. Today, plutonium is used mostly as a nuclear fuel.

This ore *contains trace amounts of plutonium.*

This plutonium battery *was used in early pacemakers.*

1970's pacemaker battery

This Martian rover *uses the heat given off by a supply of plutonium to generate electrical power.*

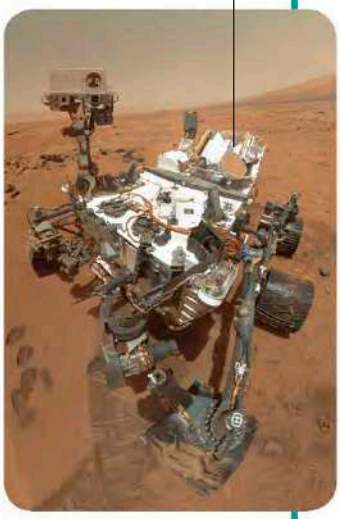

Curiosity Rover

95 Am Americium

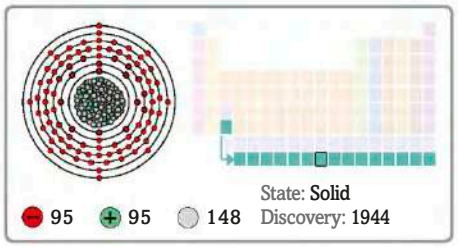

This metallic element is not found in nature. Instead, it is produced inside nuclear reactors when uranium or plutonium atoms are bombarded with neutrons. Remarkably, americium is the most common radioactive element used in the home. Radioactivity emitted by americium atoms causes the air inside **smoke detectors** to conduct electricity. When smoke disrupts the electric current, an alarm goes off.

This smoke detector *contains tiny, harmless quantities of americium.*

96 Cm Curium

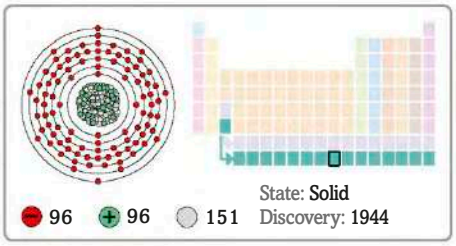

Marie Curie working in her laboratory

This lander studied the surface composition of the comet 67P.

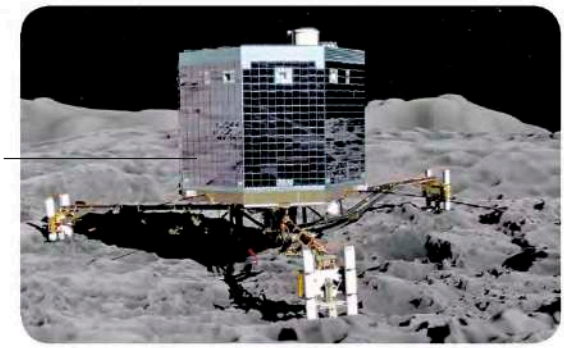

Philae lander

Curium is a silvery, radioactive metal that glows reddish purple in the dark. This element was discovered by the US scientist Glenn T Seaborg at the University of California. It was named after **Marie Curie**, the scientist who discovered the element polonium. Several space probes, such as the **Philae comet lander**, use X-ray devices containing curium to study their environment.

97 Bk Berkelium

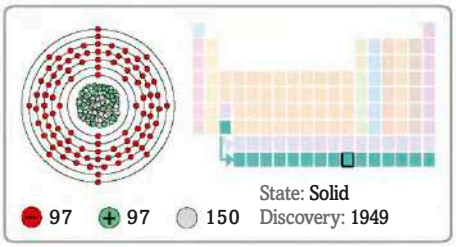

University of California, Berkeley campus, USA

This element was named after the city of Berkeley – home to the University of California – where this artificial element was discovered. It was first synthesized by **Glenn T Seaborg**. Berkelium has no uses other than the creation of heavier elements, such as tennessine.

Seaborg helped develop the atom bomb, but **opposed using it** in World War II.

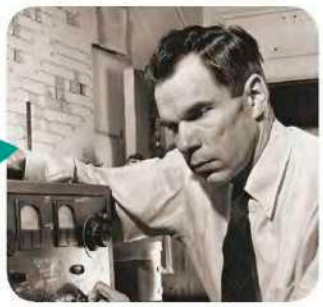

Glenn T Seaborg

98 Cf Californium

Water detector

This machine uses californium to find water underground.

This isotope, or form, of californium produces a lot of neutrons.

Pellets of radioactive californium

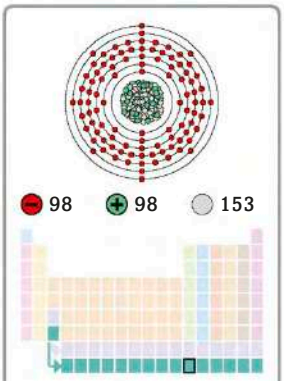

State: Solid
Discovery: 1950

Californium is named after the US state of California. This soft, silvery metal does not exist in nature and is made by smashing berkelium atoms with neutrons in a particle accelerator (a machine in which atoms are smashed together). This **radioactive element** is used in the treatment of cancer.

99 Es Einsteinium

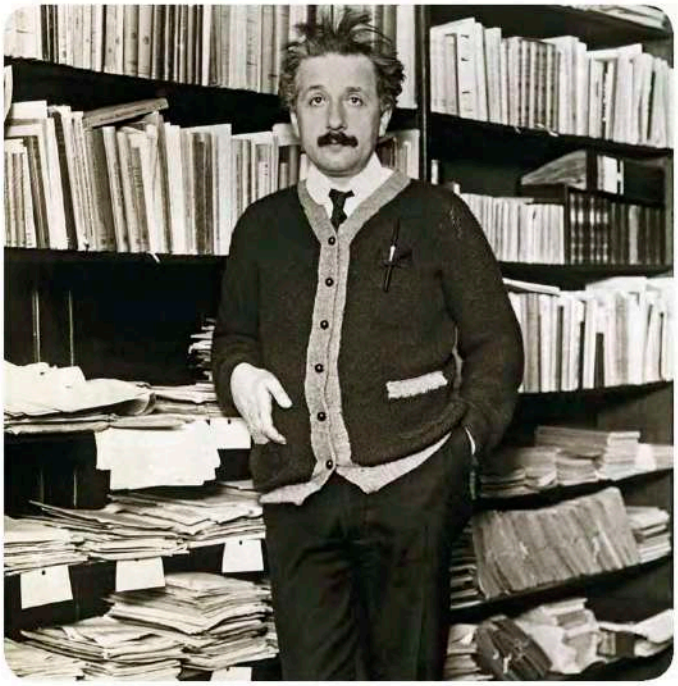

Albert Einstein in his study

Only a **few milligrams** of einsteinium are made **every year**.

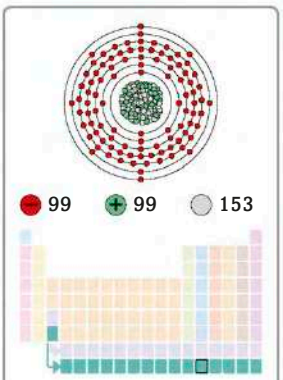

State: Solid
Discovery: 1952

Einsteinium was discovered in the chemicals left over after the first hydrogen bomb test in 1952. The huge explosion fused smaller atoms together to make larger ones, including einsteinium. This element was named after the great German-born scientist **Albert Einstein**, and was found to be a silvery, radioactive metal that glows blue in the dark. It is only used for making heavier elements, such as mendelevium.

100 Fm Fermium

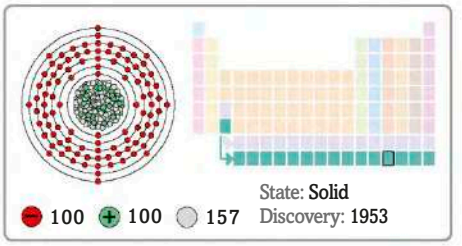

This artificial element was named after the Italian scientist Enrico Fermi. He built the first nuclear reactor in 1942, starting the American effort to build nuclear weapons during World War II. Fermium was first identified in the debris of an atom bomb test in 1953. This unstable element has no known uses beyond research.

Some scientists call Enrico Fermi the "father of the **atomic age**".

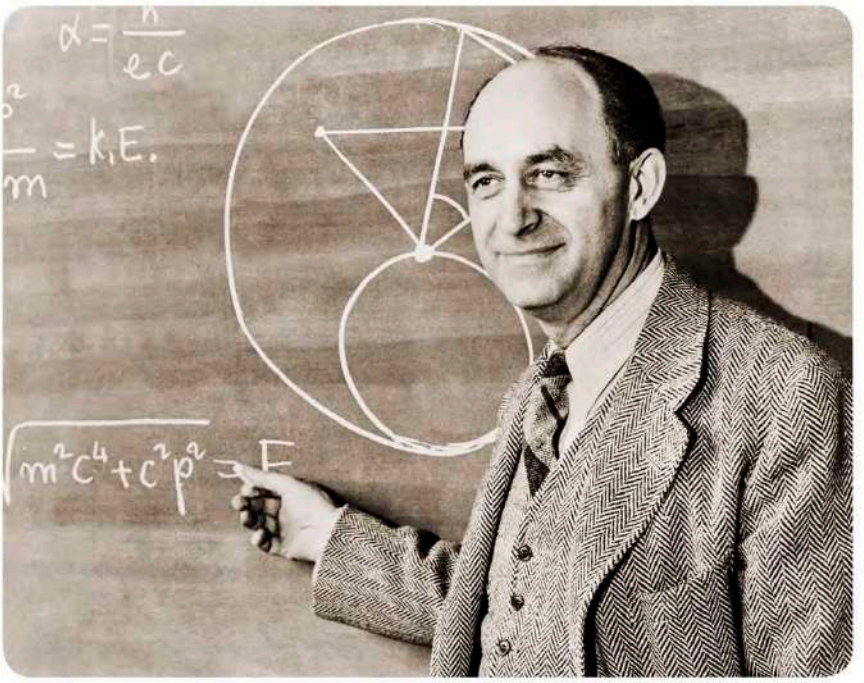

Enrico Fermi

101 Md Mendelevium

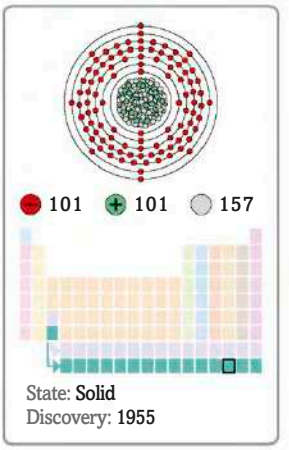

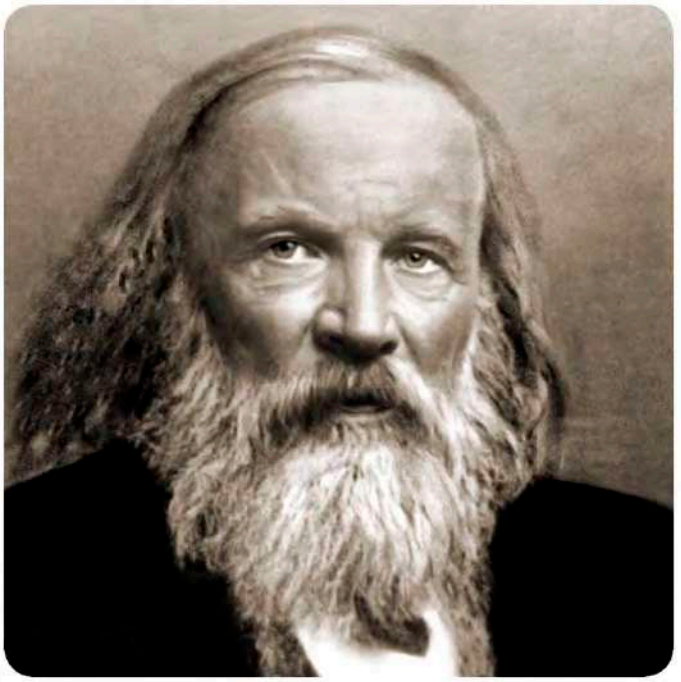

Dmitri Mendeleev

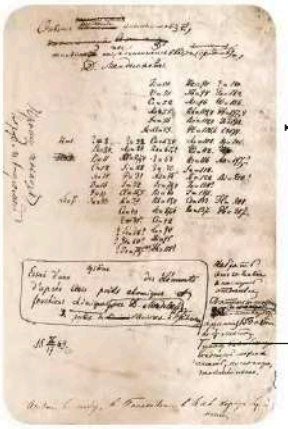

Mendeleev's periodic table

Mendeleev's notes *from 1869 show his method of arranging elements in columns and rows.*

Mendelevium is named after the Russian chemist Dmitri Mendeleev, who invented the periodic table. Mendelevium is produced in very small amounts by firing parts of helium atoms at einsteinium atoms in a particle accelerator (a machine in which atoms are smashed together).

102 No Nobelium

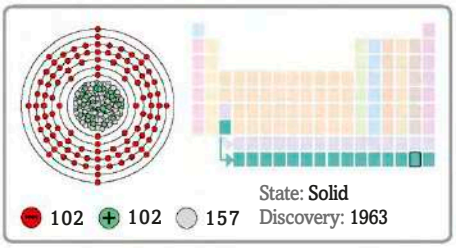

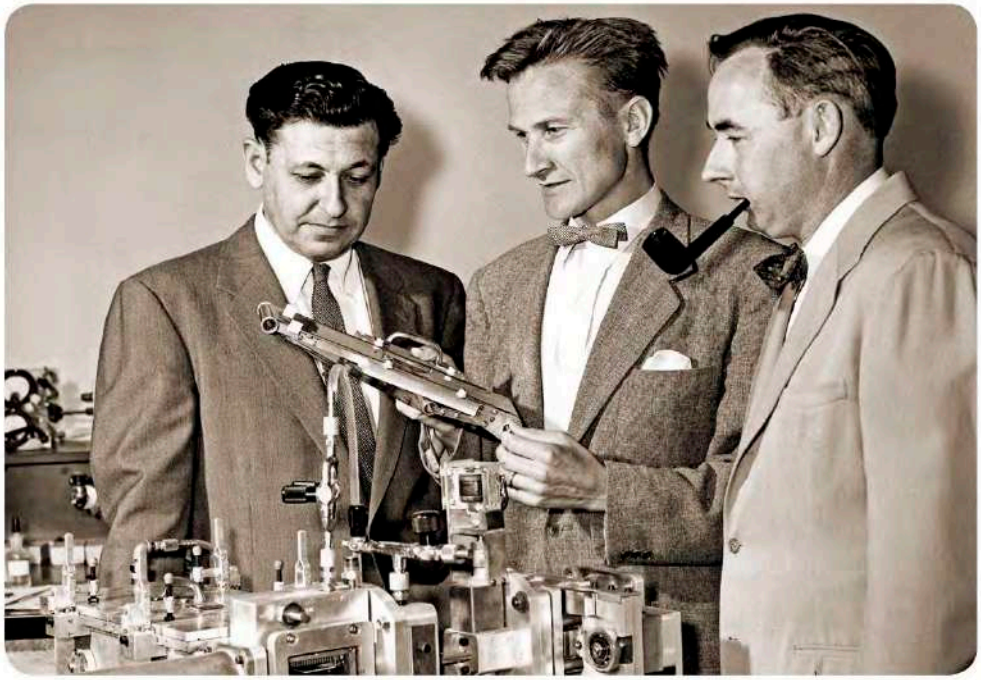

Albert Ghiorso, Torbjørn Sikkeland, and John R Walton

This artificial metal is named after the Swedish chemist Alfred Nobel, who started the Nobel Prize. It was discovered in 1963 by a team of scientists working in California, USA. This team included **Albert Ghiorso**, **Torbjørn Sikkeland**, and **John R Walton**. They used a particle accelerator to fire carbon atoms at curium atoms, creating nobelium atoms, which broke apart within minutes.

103 Lr Lawrencium

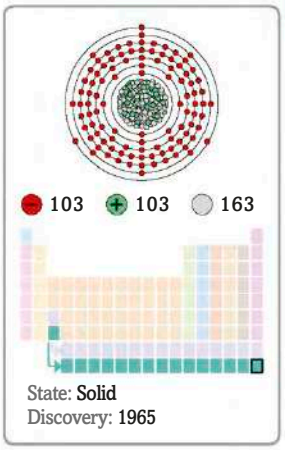

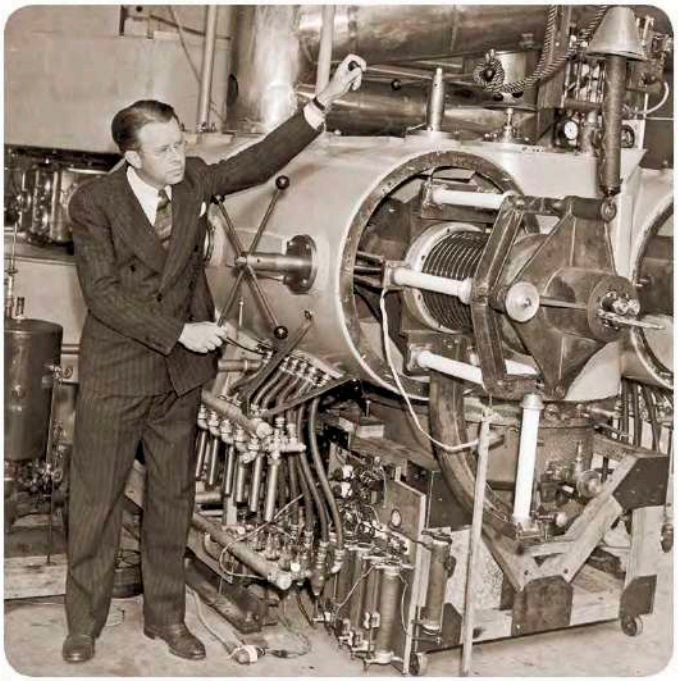

An early cyclotron

Lawrencium was produced at the **Berkeley lab** set up by Ernest Lawrence.

Lawrencium is named after the US scientist Ernest Lawrence, who developed the first cyclotron particle accelerator. This is a machine in which parts of atoms are smashed together by making them spin round in circles. Lawrencium atoms were produced in a similar machine by firing boron atoms at californium atoms.

Pure gallium (Ga) becomes liquid at 29°C (84.2°F).

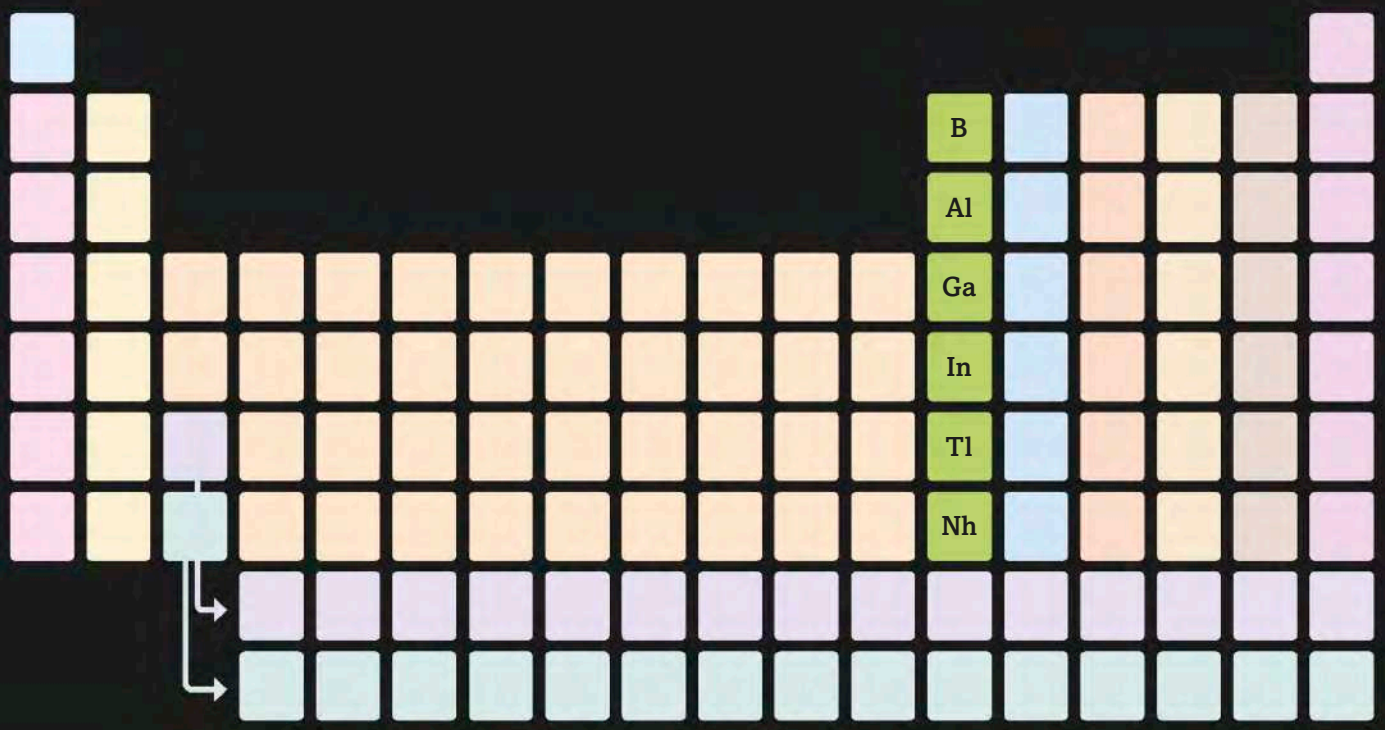

The Boron Group

This group contains five natural elements and one artificial element called nihonium (Nh). Although these elements are not very reactive, none of them are found in a pure form in nature. Boron (B), the first member, is a semi-metal (an element that has properties of both metals and non-metals), while the rest are metals. The second member, aluminium (Al), is the most common metal in Earth's rocks.

Atomic structure
Members of this group have three electrons in the outer shell of every atom. Some elements have unstable isotopes.

Physical properties
All elements, except boron, are shiny solids. Every member of this group is soft, except for boron (B), which is one of the hardest elements.

Chemical properties
Most of the elements don't react with water. Aluminium (Al) forms an oxide layer in water, and can react with it when this layer is corroded.

Compounds
They form compounds by losing electrons to other elements. All of them react with oxygen (O) by bonding to three oxygen atoms.

5 B Boron

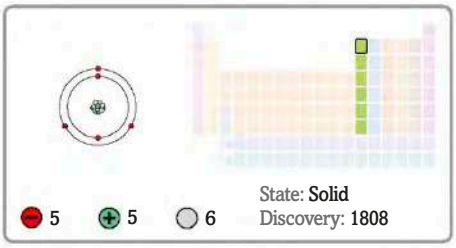

Forms

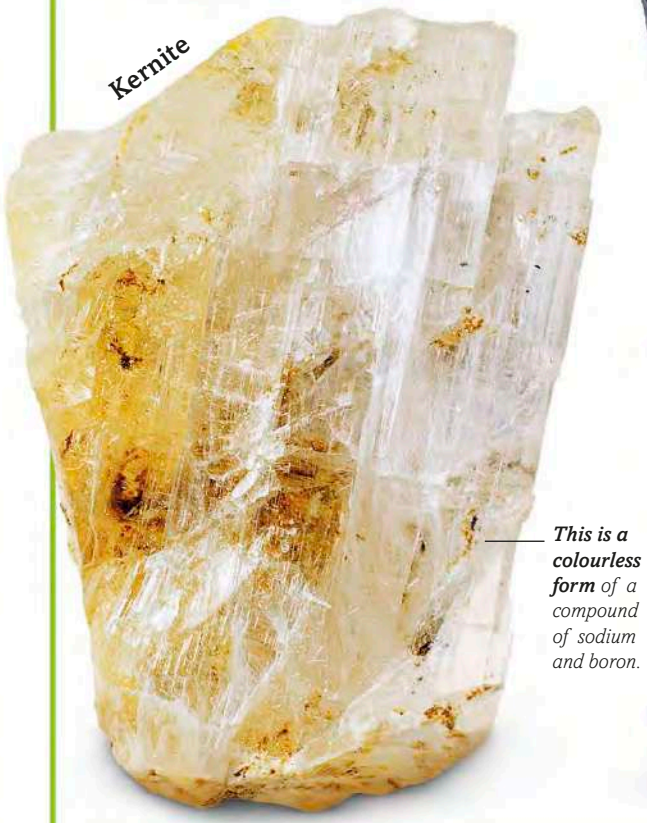

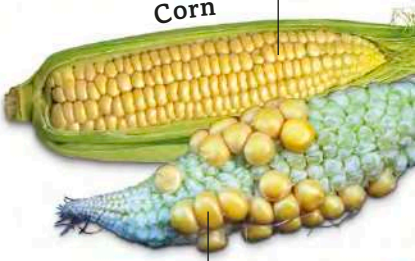

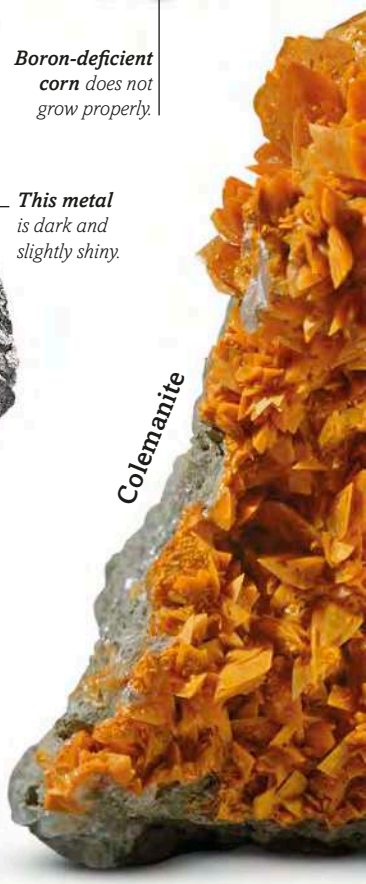

Some boron compounds are among the toughest artificial substances on Earth, with only diamond being harder. This element is a very hard material and becomes even harder when made to react with carbon or nitrogen. **Pure boron** can be extracted from various minerals, including **ulexite** and **kernite**. The demand for this element was once so high that people moved to live in the extreme heat of **Death Valley, USA**, to work in boron mines there. Compounds of boron in soil are essential for plants to grow healthily. We use boron in our homes every day.

This searing desert is one of the main places on Earth where boron is found.

Death Valley, USA

Needle- and leaf-shaped crystals

Uses

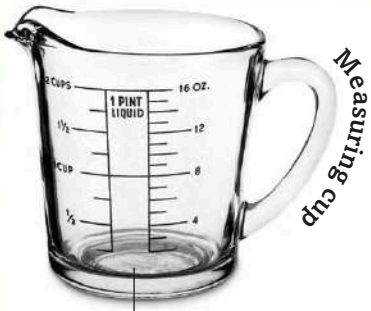

Measuring cup

This tough glass contains boron oxide.

Boric acid

These white crystals are obtained from sodium borate.

THÉNARD AND GAY-LUSSAC

The salt sodium borate, also called borax, was already in use 1,000 years ago. In 1808, the Frenchmen Joseph Louis Gay-Lussac and Louis Jacques Thénard isolated pure boron by heating borax with potassium.

Louis Jacques Thénard
Born into a poor family, Thénard excelled as a scientist. He also discovered a compound called hydrogen peroxide.

Joseph Louis Gay-Lussac
This French chemist is also remembered for discovering that the pressure of a gas goes up with its temperature.

LCD screen

This screen is composed of boron-rich glass, which makes it scratch-resistant.

This clay is bouncy but also firm because it contains boron.

Modelling clay

Boron carbide is one of the **hardest** materials in use today.

The protective body of this tank contains boron carbide, a compound of boron and carbon.

Military tank

Tough, heat-resistant glassware, such as **measuring cups**, are strengthened with boron. **Boric acid** is a natural antiseptic and can be used to treat minor cuts and scrapes. A flexible layer of boron-based glass fibres is used to toughen thin **LCD screens** for televisions and laptops. Even some kinds of **modelling clay** and bouncy silly putty contain boron compounds. Boron is named after a crumbly white salt called borax, which is used in detergents. The element is also present in a diverse range of objects, from insecticides to armour for **military tanks**.

13
Al

Aluminium

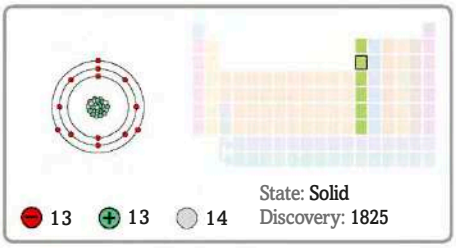

Forms

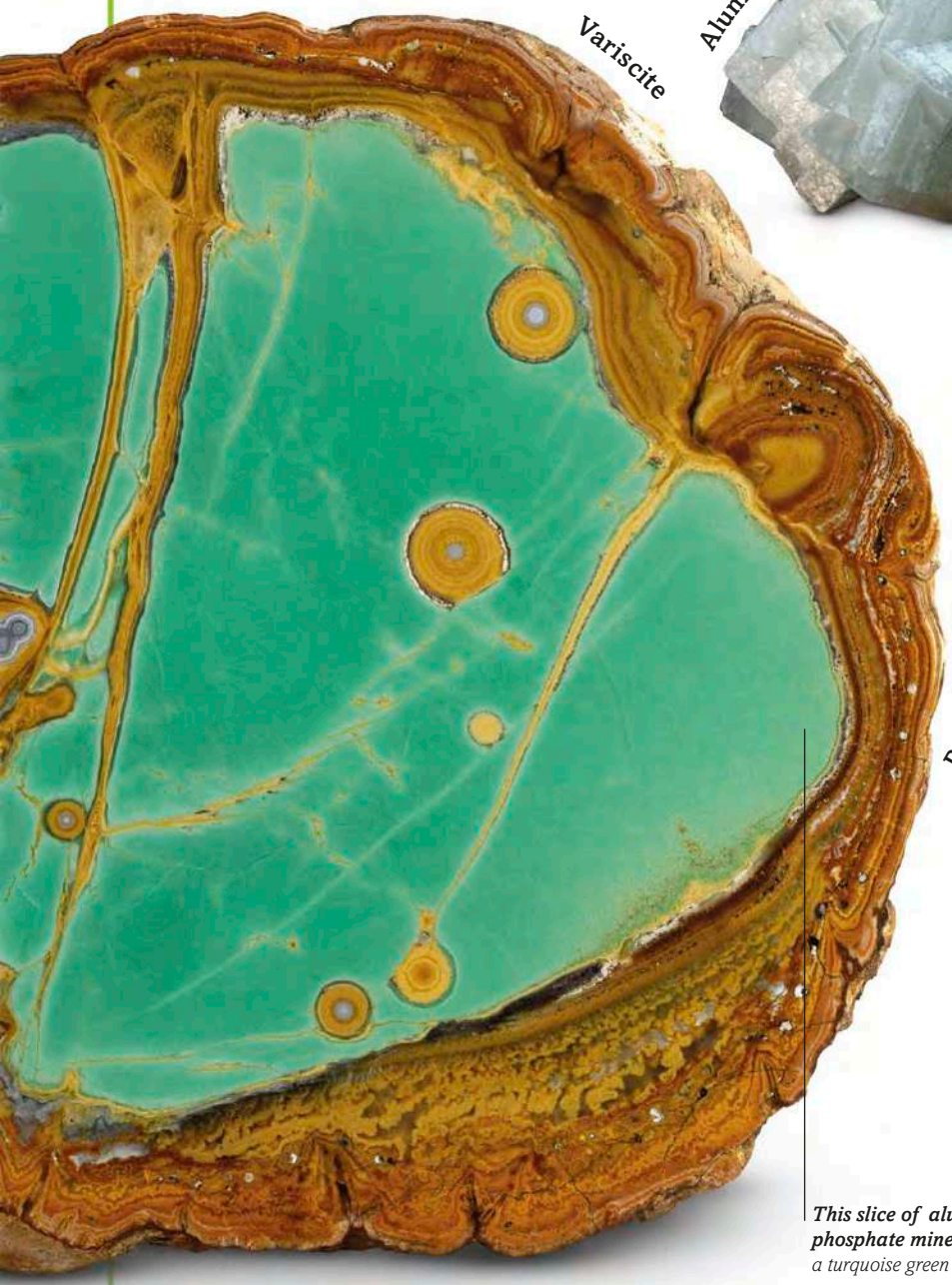

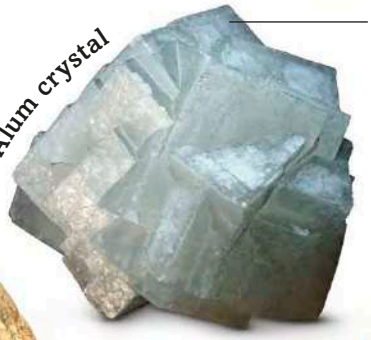

This crystal *contains aluminium and sulfur compounds.*

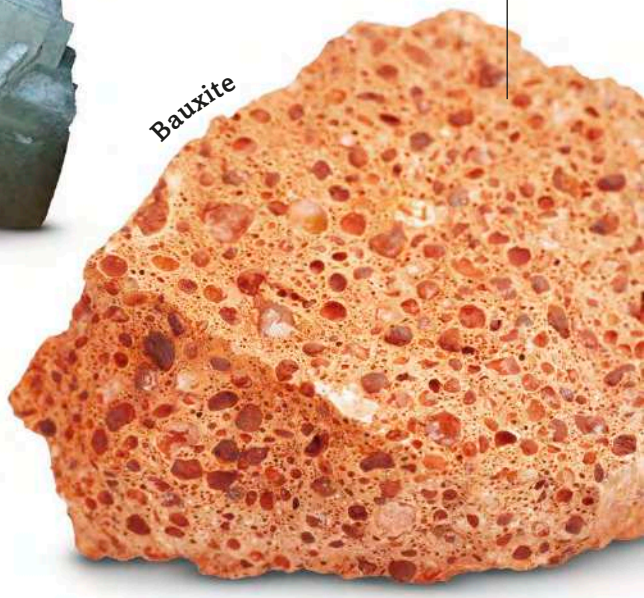

This aluminium ore *is the world's main source of the element.*

Reflective surface

This slice of aluminium phosphate mineral *has a turquoise green colour.*

Although aluminium is the most common metal in Earth's rocks, scientists did not discover it until the early 1800s. Even then, it took a further 80 years for scientists to work out how to use the ore **bauxite** to extract large amounts of **pure aluminium**. It can also be found in other minerals, including **variscite**. Today, aluminium is often recycled because producing it anew requires 15 times more energy. The metal makes a strong, shiny **foil** when rolled flat, and is useful for storing foods. A **fire protection suit** made from

Uses

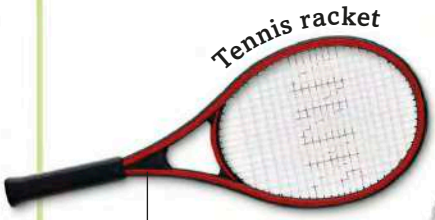

This aluminium frame *makes the racket light.*

This foil *does not break even as it is bent and twisted.*

This suit *protects against temperatures up to 1,000°C (1,800°F).*

This can *is produced from recycled aluminium.*

Recycling
one aluminium can saves enough energy to run a TV for three hours.

Aluminium cables *are lightweight.*

The aluminium casing *protects the touchscreen.*

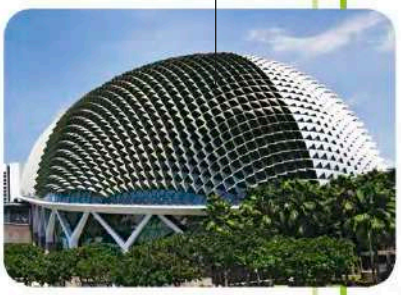

Parts of this dome *are made of aluminium.*

This plane's fuselage *is built from sheets of aluminium stretched around a frame.*

RECYCLING ALUMINIUM

Aluminium is expensive to purify, so it is often recycled instead. Drinks cans are almost 100 per cent pure aluminium and can be shredded, melted down, and made into new cans.

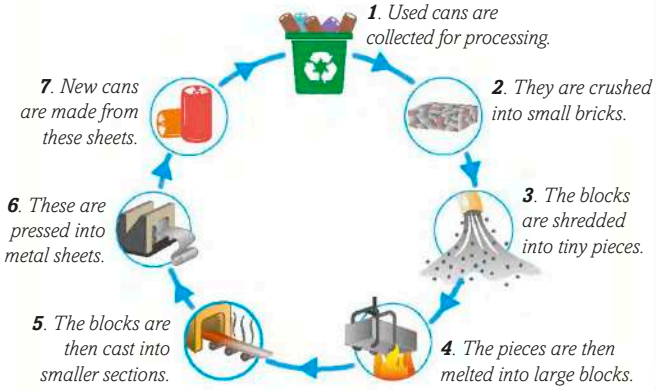

this foil reflects away heat. Aluminium is the most widely used metal after iron. It is very lightweight compared to iron's alloy steel and almost as strong. A dome made from aluminium, such as the one in the **Esplanade Theatre** in Singapore, can be much larger than a steel-based one, which would collapse under its own weight. Aluminium is also a good electrical conductor and so is used in **overhead cables**. Tough aluminium alloys are used to produce parts of some aircraft, including the **Boeing 737**.

JET TURBINE The curved blades of this jet engine are shaped very precisely to catch the air, and they are also strong enough to stay stiff when working at high temperatures. There are several tough metals that fit these requirements, but most are very dense, making them too heavy for an engine powering an aeroplane into the air. That leaves only one metal for the job: aluminium.

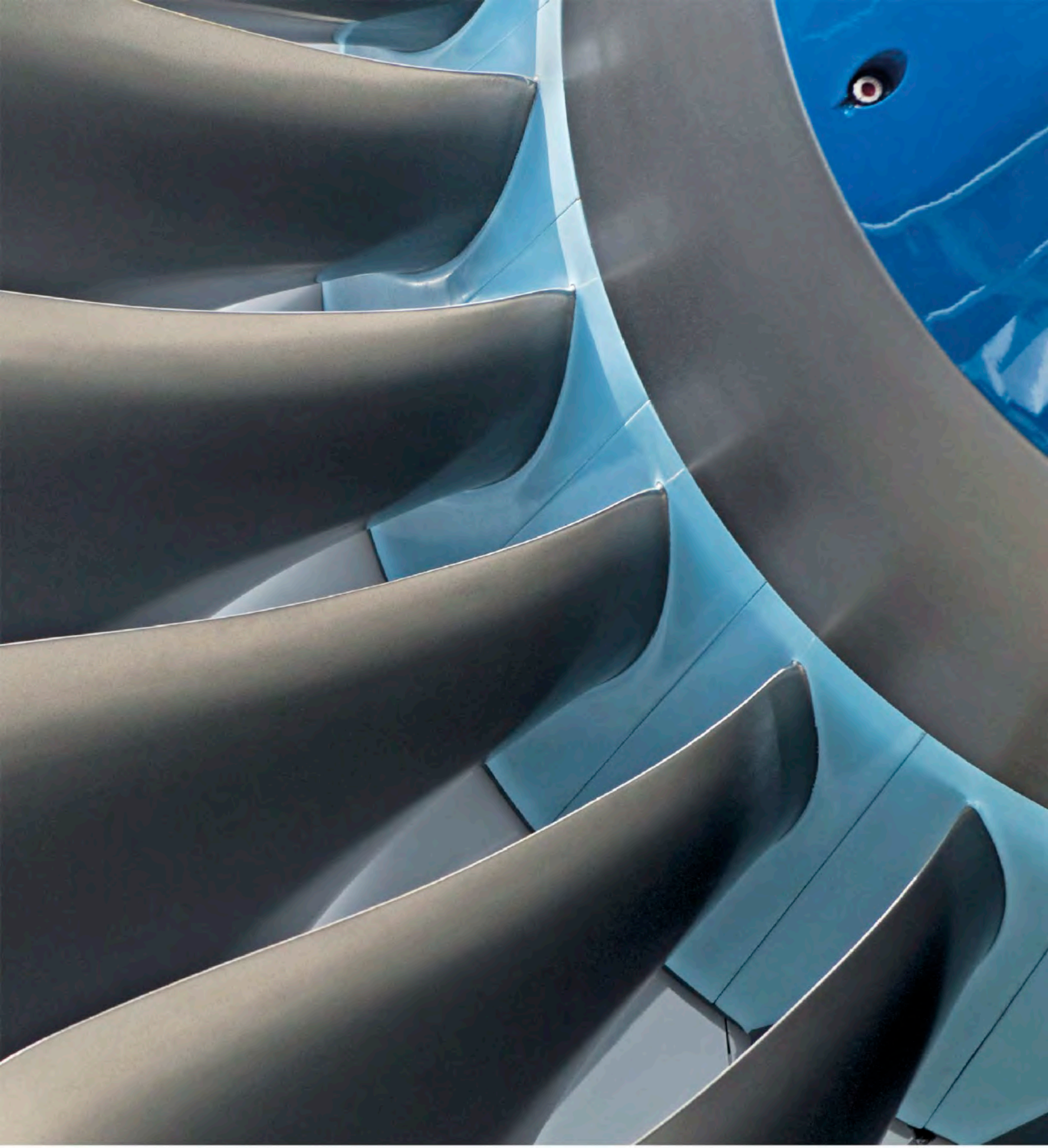

Aluminium is what makes high-speed, long-range air travel possible. Easily moulded, it is one-quarter the weight of steel, and it never rusts. Steel is stronger, but a plane made from it would be too heavy to fly. Instead aluminium is mixed with titanium and steel to produce tough yet lightweight alloys, which are used in the engines and bodies of jet aircraft. There is almost twice as much aluminium in Earth's rock as there is iron. However, purifying aluminium takes a lot of energy. Once pure, though, it can be recycled over and over again. So, one day these engine blades might transform into a fizzy drink can.

31 Ga Gallium

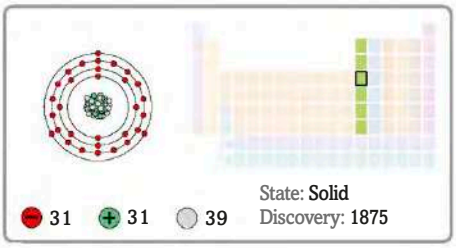

Forms

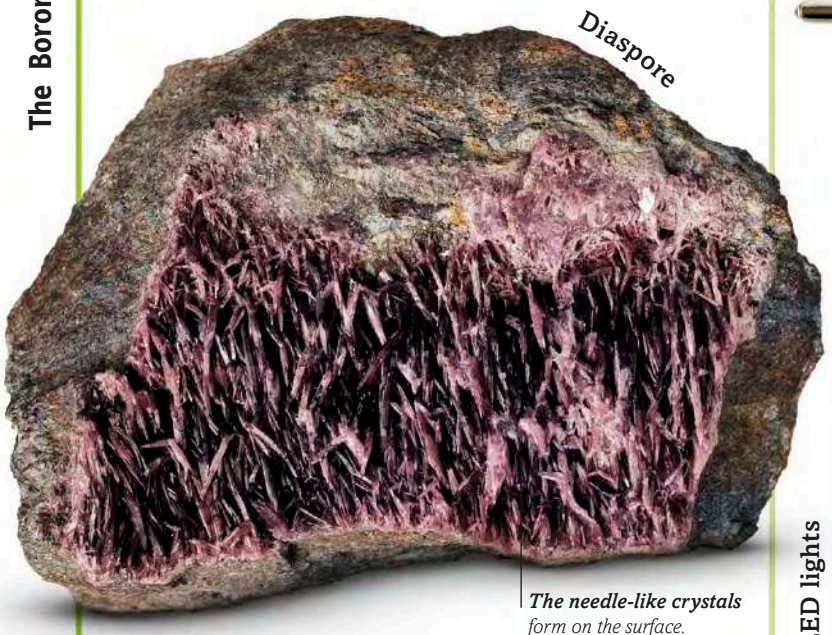

The needle-like crystals form on the surface.

Pure gallium has a very low melting point.

Uses

Thermometer

This medical thermometer uses a gallium alloy instead of mercury.

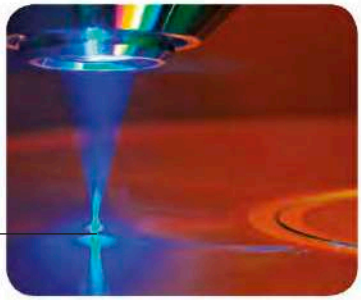

Blu-ray disc

A gallium laser is used to read Blu-ray discs.

Red LED lights

The red LED gets its colour from gallium compounds.

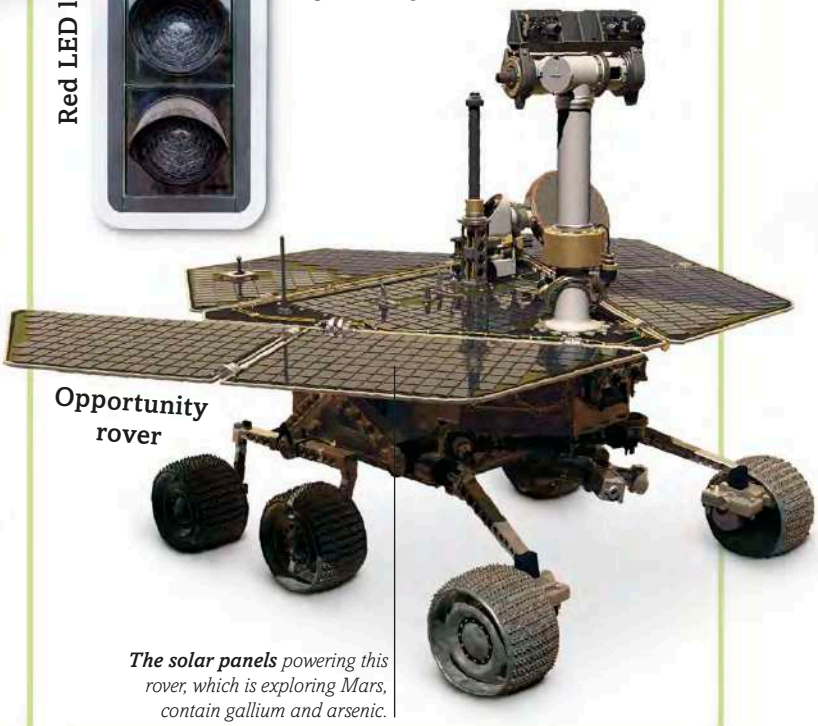

Opportunity rover

The solar panels powering this rover, which is exploring Mars, contain gallium and arsenic.

Gallium melts at just 29°C (84.2°F), which means it soon becomes liquid when held in the hand. This element is found in small amounts in ores of zinc and aluminium, such as **diaspore**. **Pure gallium** is isolated when the other elements from this ore are extracted.

Gallium has a number of uses. It is mixed with indium and tin to form a liquid alloy called galinstan, which can be used in **thermometers**. Gallium is also found in **Blu-ray lasers**, **LEDs**, and some solar panels, such as those on NASA's Mars **rovers**.

49 In Indium

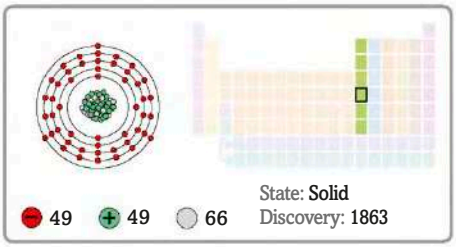

Forms

When bent, indium produces a **"tin cry"** – a sound similar to a **scream**.

Sphalerite

This zinc mineral *is the main source of indium.*

Pure indium mould cast in a laboratory

Pure indium *is soft enough to draw lines on paper.*

Uses

This touchscreen *has a grid of very thin, transparent wires made of indium tin oxide.*

Touchscreen tablet

Welder's goggles

Protective goggles *coated with indium stop the heat damaging the welder's eyes.*

This glass coated with indium oxide *is shiny but still lets light through.*

Transistor

The tiny electronic switches *inside this transistor contain indium.*

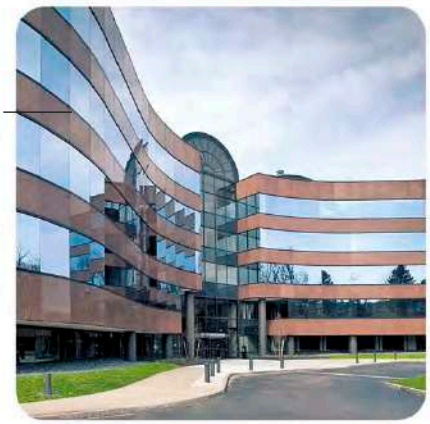

Windows in a building

Indium is named after indigo, which is the colour of the light its atoms release when electrified. Its minerals are rare, and most of the metal is obtained from lead and zinc ores, such as **sphalerite**. **Pure indium** is very soft, and the element is mostly used in compound form. For example, a compound called indium tin oxide used on a **touchscreen** allows the computer to detect when a finger makes contact with the screen. Indium is also required in microchips, and to produce **welder's goggles** and **windows** that are heat- and glare-proof.

81 Tl Thallium

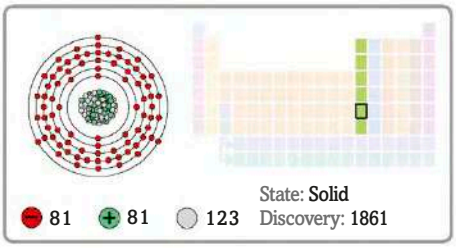

Forms

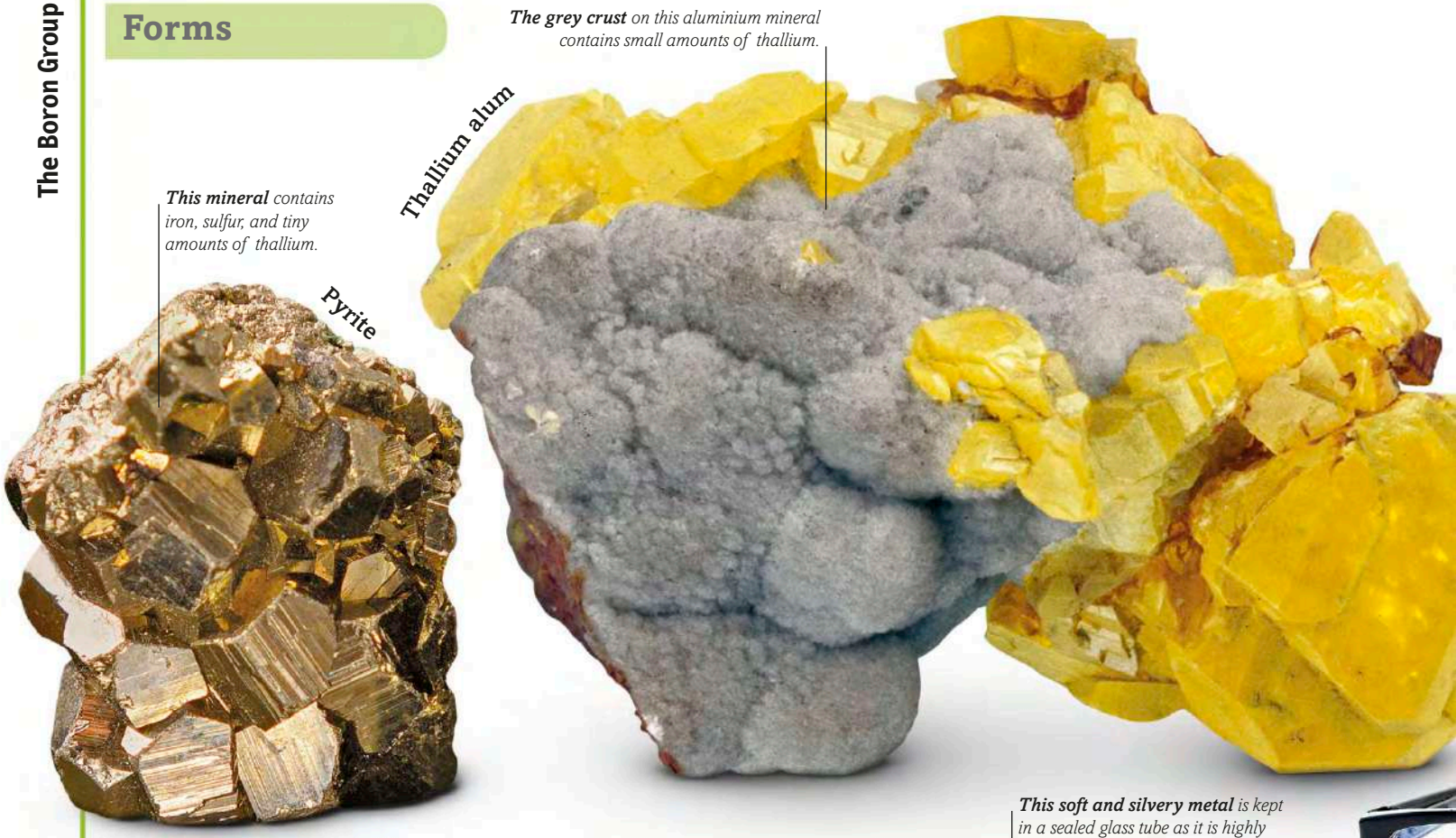

The grey crust *on this aluminium mineral contains small amounts of thallium.*

This mineral *contains iron, sulfur, and tiny amounts of thallium.*

This soft and silvery metal *is kept in a sealed glass tube as it is highly toxic and reacts easily with air.*

Laboratory sample of pure thallium in an airless vial

Thallium was named after the Greek word *thallos*, which means "green shoot": it was first identified from the colours in its flame, which includes a bright green light. Thallium was discovered in 1861 by William Crookes and Claude-Auguste Lamy. Although both chemists worked separately, they found the element in the same way – as a residue while making strong acids using the mineral **pyrite**. Thallium was later found to exist in larger amounts in other minerals, including **thallium alum**. **Pure thallium** is toxic and has to be handled

Uses

Heart function scan

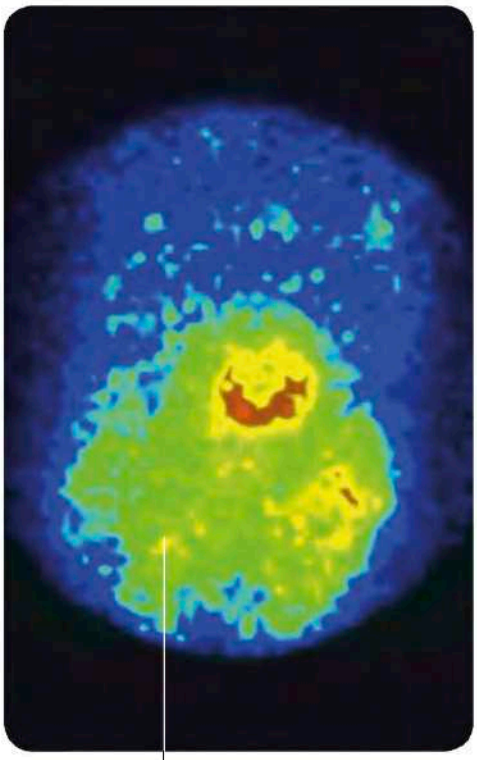

Blood injected with a thallium compound *shows up on a patient's heart scan.*

Spectacles

These thin lenses *contain strong, thallium-infused glass.*

Until the 1970s, thallium salts were commonly used as ant poison.

toxic and has to be handled with care when used. A chlorine compound of thallium is used in **scans** to study a patient's blood circulation. Thallium oxide also helps make glass stronger for use in **spectacles** and cameras.

113
Nh

Nihonium

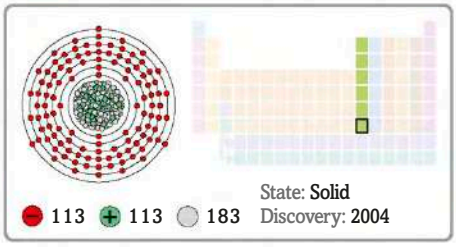

Kozuka Morita (left), with a visiting official at the RIKEN Nuclear Research Centre, Wako, Japan

Nihonium was named after the Japanese word *nihon*, which means Japan. A metallic element, nihonium was first detected in 2003 by teams studying the artificial element moscovium, which has the atomic number of 115. They noticed that atoms of moscovium broke apart after only a few seconds into atoms of an element with an atomic number of 113. In 2004, Kozuka Morita and a team of scientists at the **RIKEN Nuclear Research Centre** in Japan isolated this element in a different way: they fused bismuth and zinc atoms together.

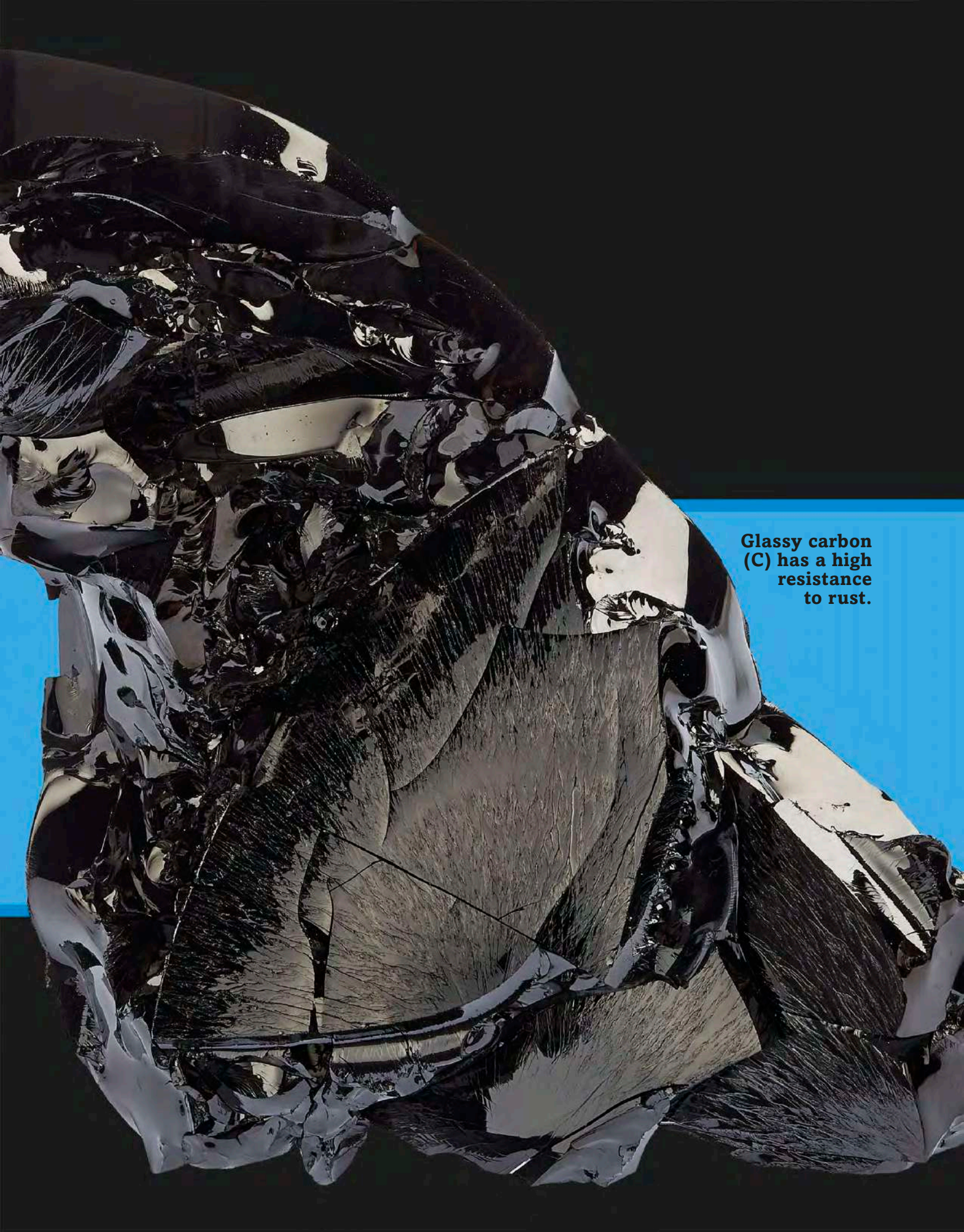

Glassy carbon (C) has a high resistance to rust.

The Carbon Group

This group contains one non-metal, two semi-metals, and three metals. The non-metal carbon (C) is the main element in all living things. The semi-metals – silicon (Si) and germanium (Ge) – are elements that have the properties of both metals and non-metals, and they are essential in electronics. Two of the metals – tin (Sn) and lead (Pb) – have been used by humans for centuries. Flerovium (Fl), an artificial element, has no known uses.

Atomic structure
Members of this group have four electrons in the outer shell of each atom. These atoms can bond with up to four other atoms.

Physical properties
At room temperature, all natural elements in this set are solid. Flerovium (Fl) is an artificial element, and scientists think it may be a solid.

Chemical properties
All natural elements in this group can react with hydrogen (H). Carbon (C) and silicon (Si) can react with both metals and non-metals.

Compounds
These elements react with hydrogen to form compounds called hydrides. Each element can lose up to four electrons when forming compounds.

6 C Carbon

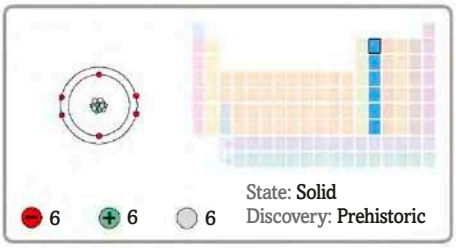

Forms

Carbon has the largest number of compounds of any element – with more than nine million known. Carbon is the fourth most common element in the Universe. Each carbon atom can bond to four others, allowing them to form chains and rings. Pure carbon exists in three forms on Earth – **graphite**, **diamond**, and buckminsterfullerene (a structure based on 60 interlinked carbon atoms). Diamond is the hardest substance in nature. It is often used in jewellery. The blades of some **saws** are coated with diamonds, and

Uses

This sharp blade is coated with tiny diamonds.

This carbon filter absorbs harmful chemicals in water.

This battery has a graphite core that carries electric current.

Sturdy and lightweight frame

This **strong** but **light** frame is made by fusing together carbon fibres.

This wheel is composed of carbon fibre and some other materials, which make it rigid.

Soft graphite leaves a mark when pressed against paper.

Pencil "lead" containing graphite, not lead

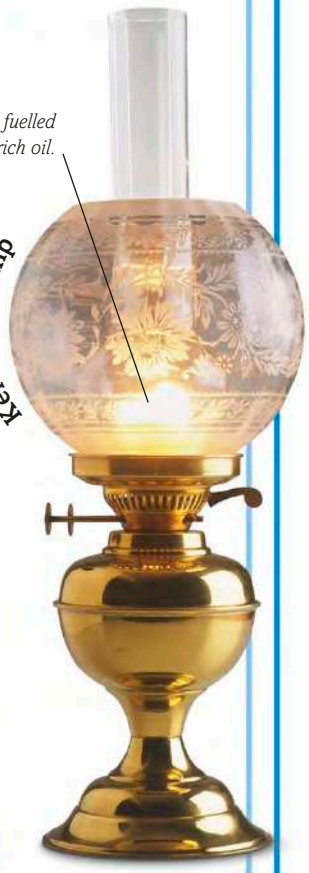

This flame is fuelled by carbon-rich oil.

This flexible plastic material is composed of chains of carbon and hydrogen atoms.

HARD AND SOFT CARBON

Diamond and graphite have different properties because of the arrangement of their atoms – tetrahedrons in diamond, and sheets of hexagons in graphite.

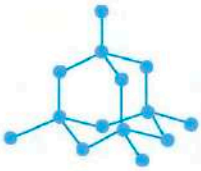

Hard as diamond
The tetrahedron, or pyramid structure, of atoms creates a rigid shape that is equally strong in all directions.

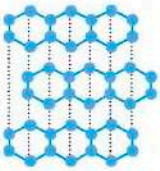

Soft as graphite
Graphite contains layers of carbon atoms that slide over each other easily, as there are only weak forces between them.

can cut into anything. Only a diamond can cut another diamond. Graphite is much softer, which is why it is used in **pencil "lead"**. It is also used in some batteries. Coal is currently the largest source of fuel for the generation of electricity, but its fumes are also known to have harmful environmental and health effects. Crude oil, natural gas, and **coal** are hydrocarbons (compounds containing only hydrogen and carbon) that occur in nature. They can be used as fuels and as plastics for objects such as **polythene bags**.

PINK DIAMOND With a mass of only just over 3 g (0.1 oz), this jewel – known as The Sweet Josephine – is one of the largest pink diamonds ever sold. Diamonds are normally colourless forms of pure carbon, and if there is any colour, it comes from tiny amounts of another substance. For example, boron makes the gem appear blue. Strangely, pink diamonds have no impurities, and no-one knows why they are pink.

The Sweet Josephine was cut from raw diamond that is more than 1.5 billion years old. This formed 150 km (93 miles) beneath Earth's surface and was then pushed up by a volcanic eruption, before eventually being dug out at a mine in Australia. Diamonds form when carbon is squeezed and heated to more than 1,000°C (1,832°F). This process rearranges the carbon atoms into a rigid crystal that makes diamond the world's hardest substance. The process also gives diamond the ability to bend light, a property that gives these jewels their glorious sparkle. With the right cut and polish, a diamond can be made into a beautiful gem that is prized throughout the world.

14 Si Silicon

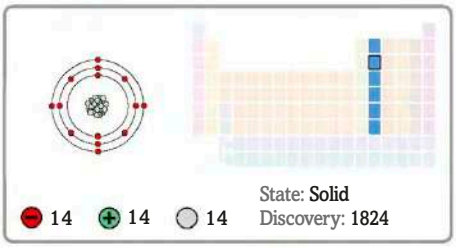

Forms

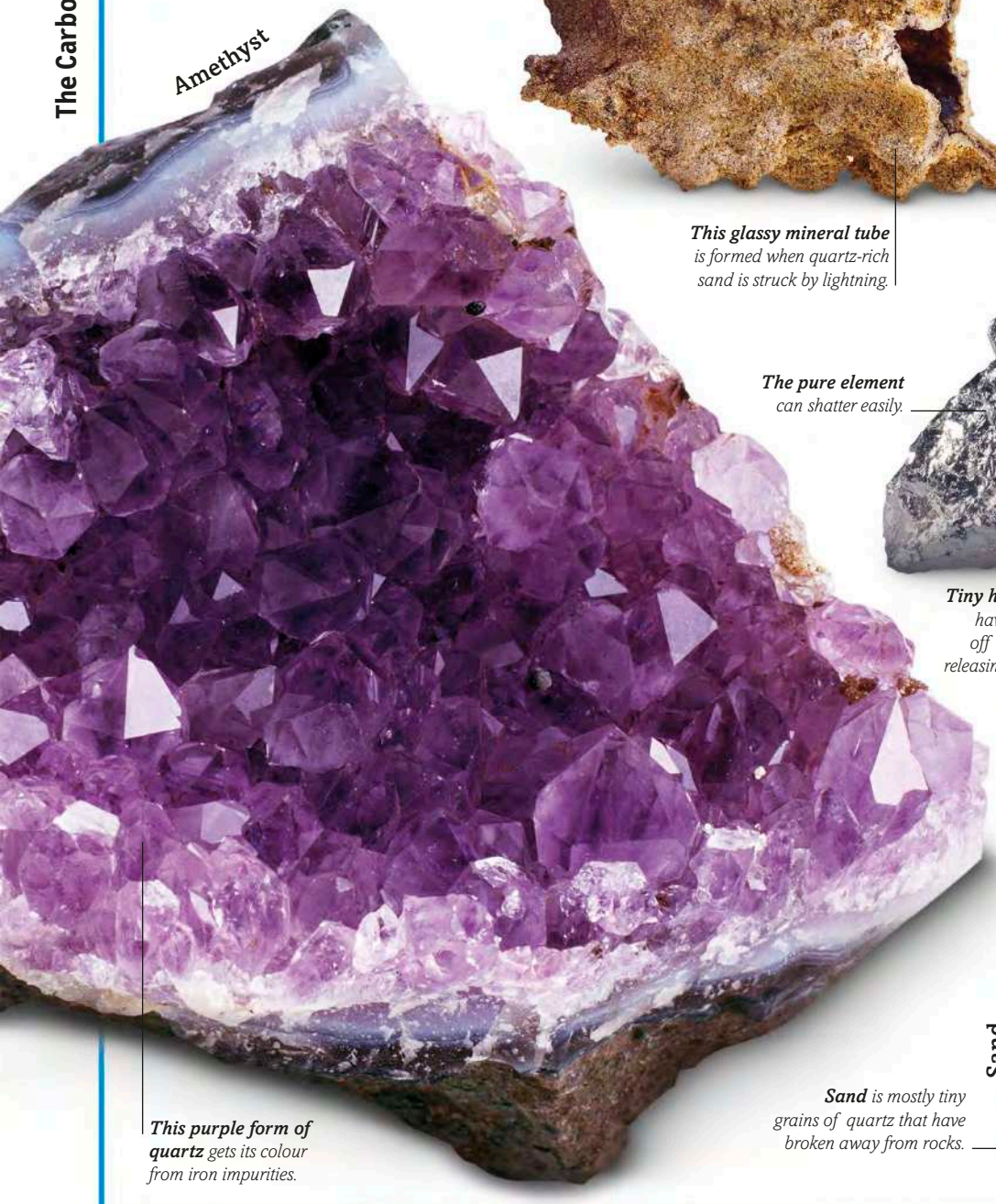

About 90 per cent of the minerals that make up Earth's rocks contain silicon, a common element in our planet's crust. Nearly all silicon minerals are compounds of silicon and oxygen, known as silicates. The most common silicate is quartz, the mineral form of silicon dioxide, or silica. It is also the most common substance in **sand**. **Amethyst** is a type of quartz. Quartz deposits are widely found in rocks such as granite and sandstone. A valuable type of silica is **opal**, which is used as a gemstone. The clays used to make pottery and **ceramics** are also silicates.

Uses

Opal ring

Opal is a jewel *containing both silicon oxide and water molecules.*

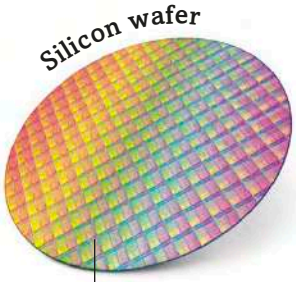

Silicon wafer

Thin slice of silicon

Screens of silicon-based smoke *can be used in battles.*

Smokescreen

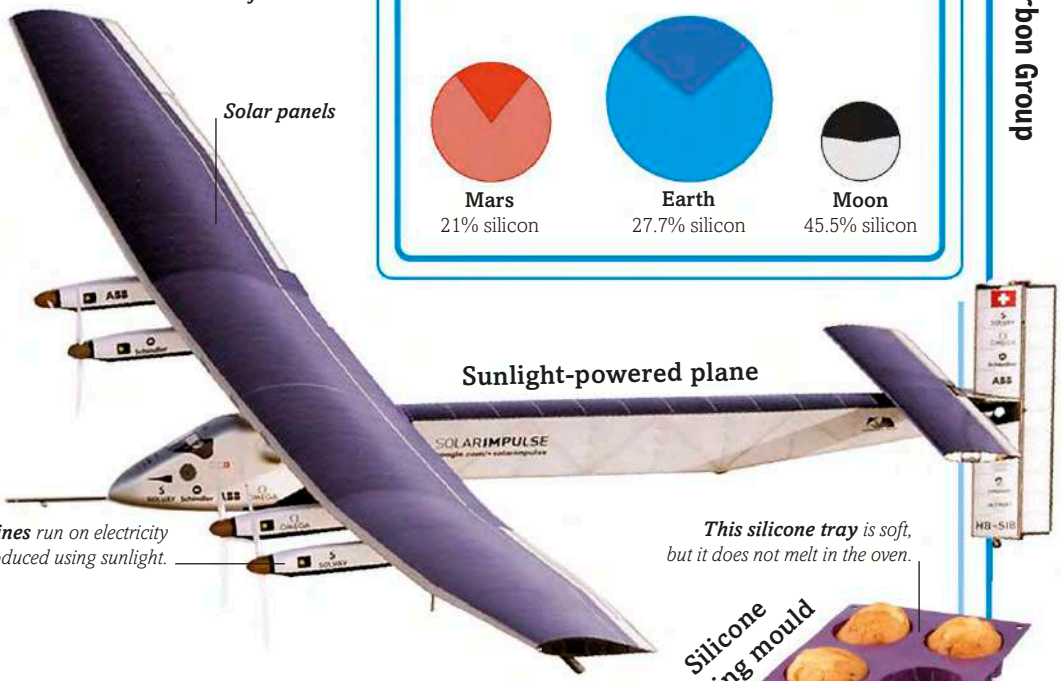

Solar panels

The engines *run on electricity produced using sunlight.*

Sunlight-powered plane

Ceramic vase

Ceramic *is made of a silicon-based clay that has been shaped and permanently hardened by heat.*

Aerogel *conducts the heat from the flame poorly, preventing it from passing to the flower.*

Silicate aerogel in an experiment

This silicone tray *is soft, but it does not melt in the oven.*

Silicone baking mould

Silicone watch

This silicone band *is flexible and strong.*

SILICON IN SPACE

Earth and Mars contain similar amounts of silicon. This element forms on the outer layer, or crust, of both planets. In contrast, the Moon is almost half silicon. Astronomers think this tells us that the Moon was formed from Earth's surface after an asteroid smashed into our planet about 4.4 billion years ago.

Mars 21% silicon

Earth 27.7% silicon

Moon 45.5% silicon

One of the most important uses of silicon is in electronics. Thin slices called **silicon wafers** drive electronic circuits. This versatile element is also used to turn sunlight into electricity in solar panels. Artificial silica is used to create **aerogel**, a lightweight but tough substance that does not conduct heat well. It is used in fire-fighting suits, and prevents flames reaching a firefighter. Another silicon compound is silicone, which can be moulded into any shape, and is used in a wide range of products from **baking moulds** to **watches**.

32 Ge Germanium

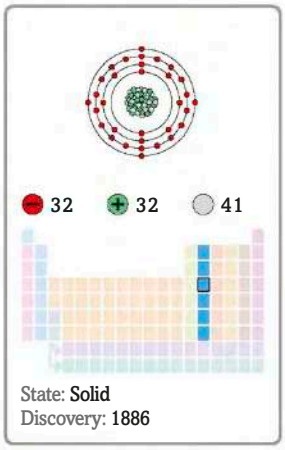

Forms

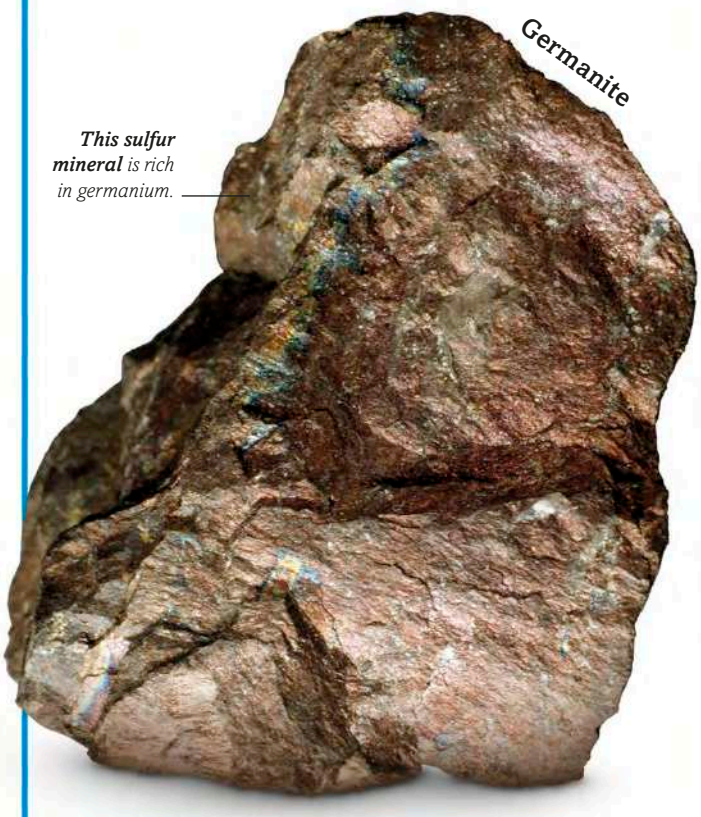

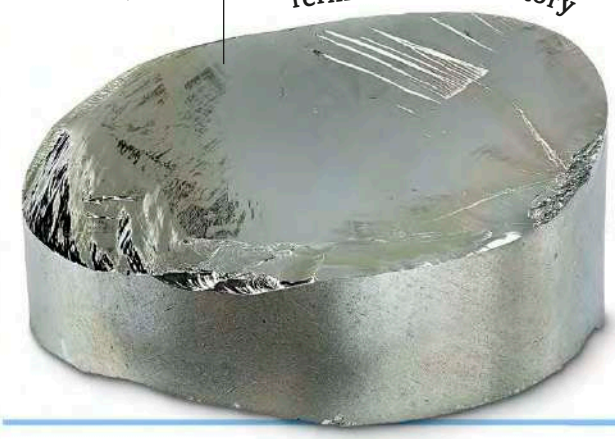

Uses

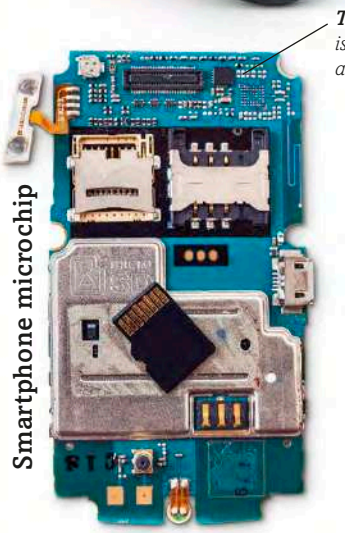

Germanium is found in the **atmosphere** of **Jupiter**.

This semi-metal is named after the country Germany. It was discovered there in 1886 by chemist Clemens A Winkler, nearly 20 years after Russian chemist Dmitri Mendeleev predicted its existence and properties. **Germanite** is a mineral rich in germanium, but this element is mainly extracted from the ores of silver, copper, and lead. One of its compounds, germanium oxide, is used in wide-angle **camera lenses**. It is also used in some **microchips** and in a number of **car sensors** that aid in navigation.

50 Sn Tin

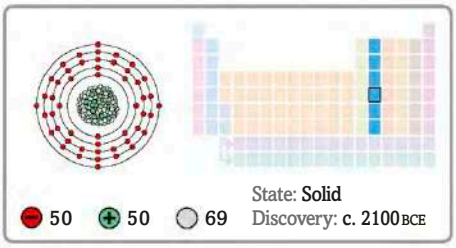

Forms

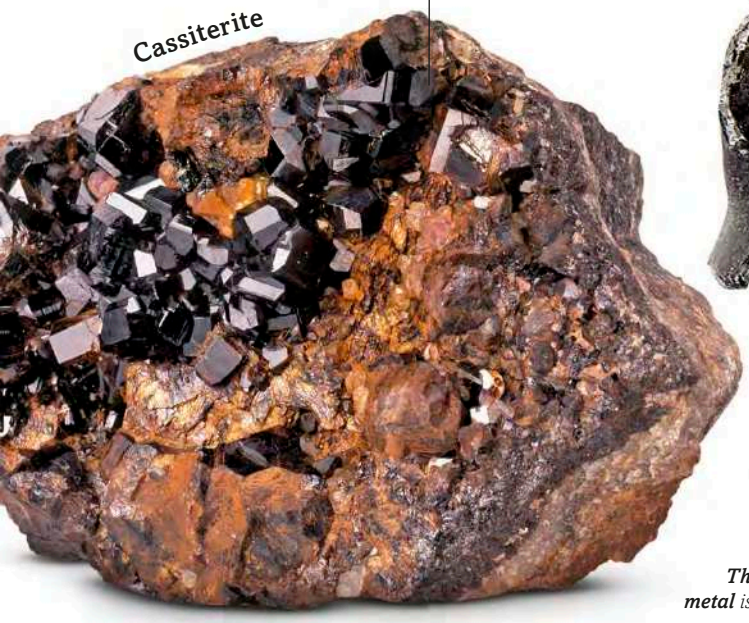

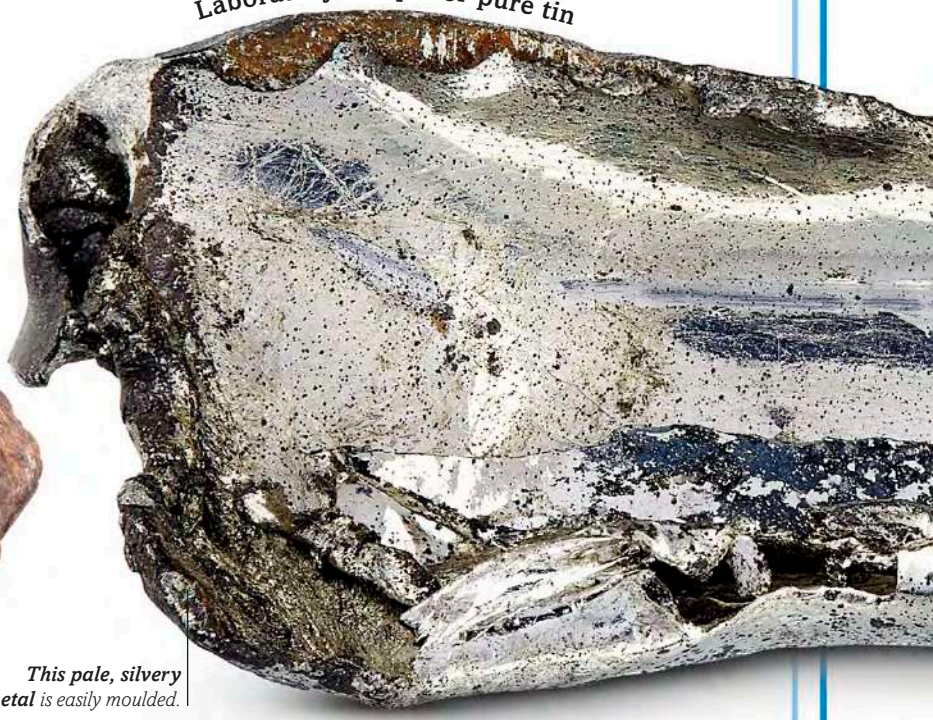

Uses

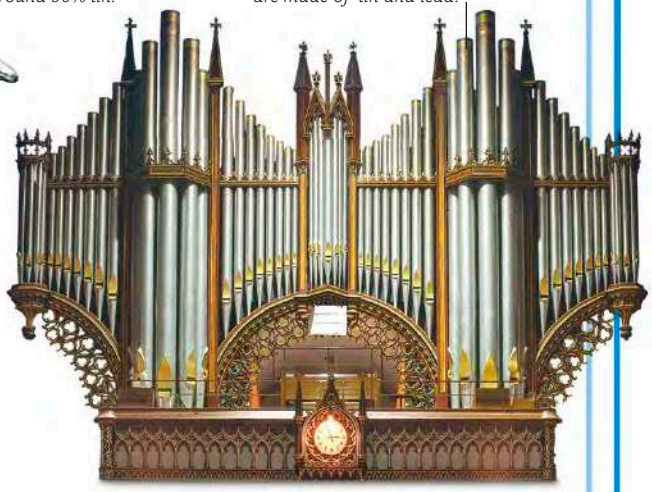

Tin was one of the first metals used by humans. As long as 5,000 years ago, tin was mixed with copper to make bronze, an alloy that was stronger than either pure metal. The ore **cassiterite** is the main source of **pure tin**. Uses for tin are many, including plating steel objects, such as **cans**, to stop them corroding. A compound called tin chloride is used for dyeing silks. This metal continues to be used in a variety of tough alloys, including **pewter**, soft solder, and bronze.

82 Pb Lead

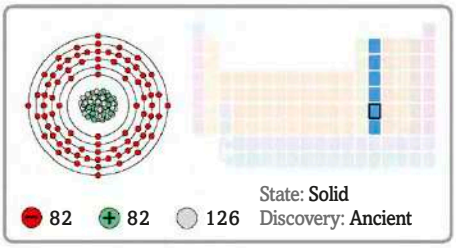

Forms

Brown crocoite

This soft, crumbly mineral *is made of lead chromate.*

Galena

This mineral *has a bright, silvery shine.*

The prism-like crystals *of this mineral contain a lead-sulfur compound.*

Anglesite

Uses

Lead pipe

Rust-resistant pipe

This crystal glass *sparkles more than regular glass because it contains lead oxide.*

Lead crystal glass mug

For many years, lead and tin were thought to be different forms of **the same metal**.

The chemical symbol for lead, Pb, comes from the Latin word *plumbum*. This is where the word "plumber" comes from: in ancient Roman times, water pipes used in plumbing were made from this soft metal. Lead compounds are found in the minerals **crocoite**, **anglesite**, and **galena** – the main source of **pure lead**. Lead was used far more commonly in the past as an important ingredient in paints, hair-dyes, and insecticides. A common historical use was in **glassware**. It has limited applications today. Lead can

Pure strip of lead refined in a laboratory

***This pure form of the metal** has a dull grey colour.*

***These pale crystals** on the galena are a calcium mineral.*

***Flashings**, or "lead sheets", cover exposed corners on roofs to make them waterproof.*

Flashings

because it absorbs radiation. It is also used in weights for diving, car batteries, and bendy **"flashings"** for sealing roofs. Lead fell out of favour because it turned out to be poisonous.

114 Fl Flerovium

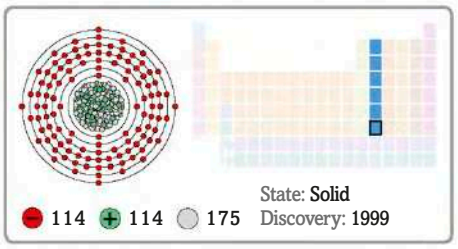

This machine produces flerovium by **smashing together** atoms of calcium and plutonium.

Particle accelerator at Joint Institute for Nuclear Research, Dubna, Russia

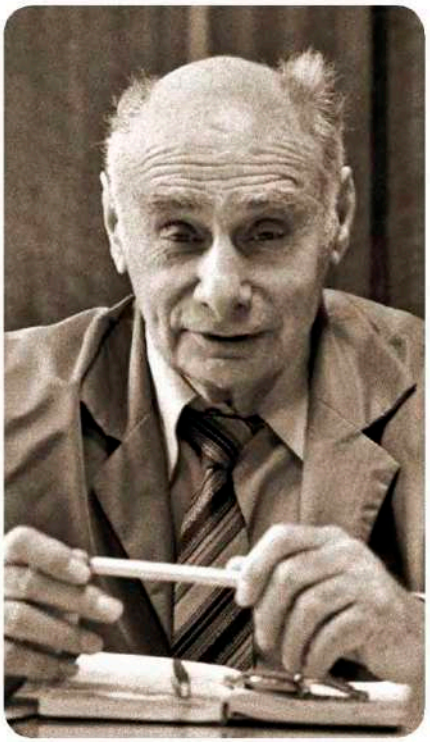

Georgy Flerov

Flerovium takes its name from the Russian scientist Georgy Flerov. He founded the **Joint Institute for Nuclear Research** in Dubna, Russia, where this element was first produced in a particle accelerator (a machine in which atoms are smashed together). Flerovium is highly radioactive and its atoms last for only a few seconds before breaking apart.

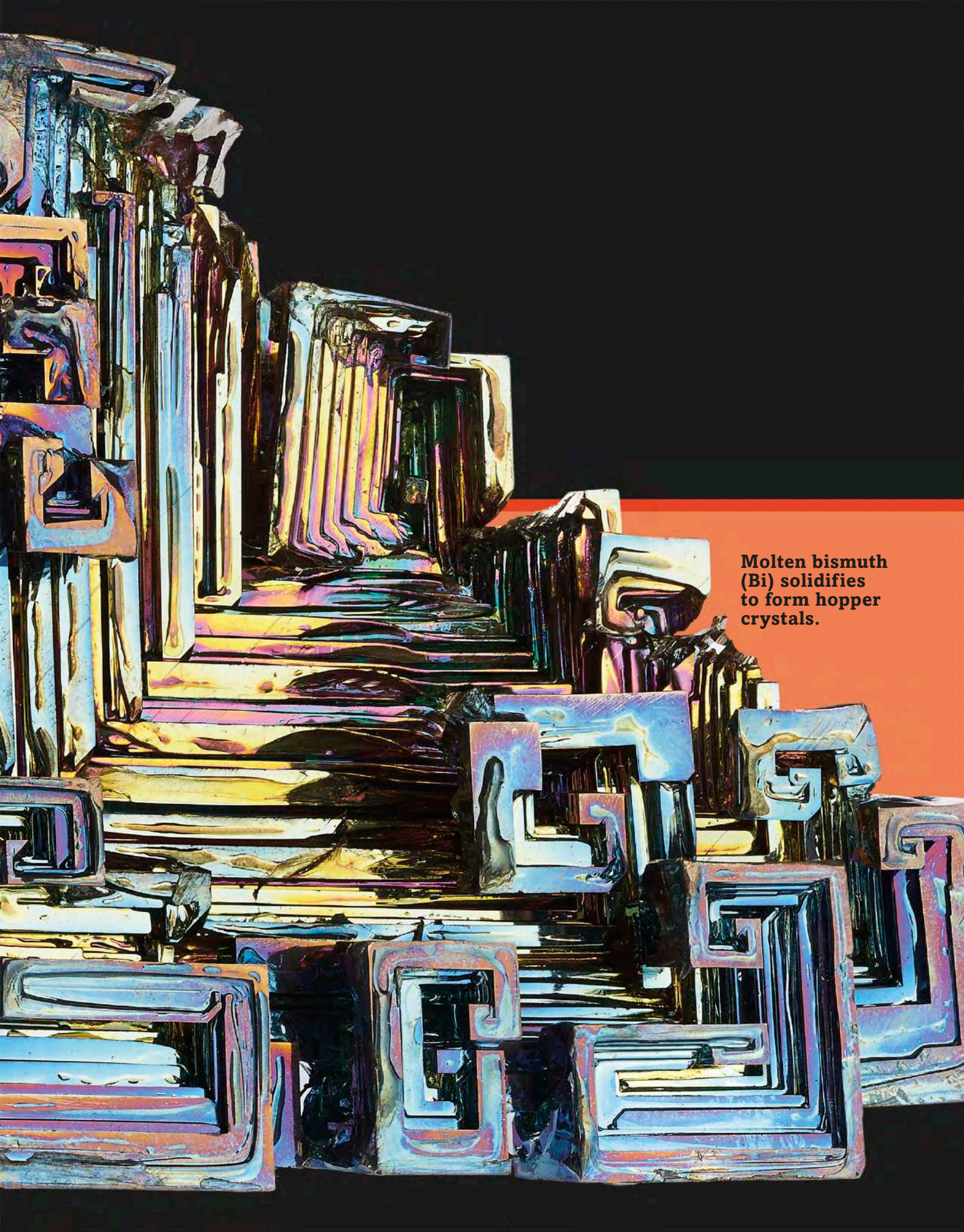

Molten bismuth (Bi) solidifies to form hopper crystals.

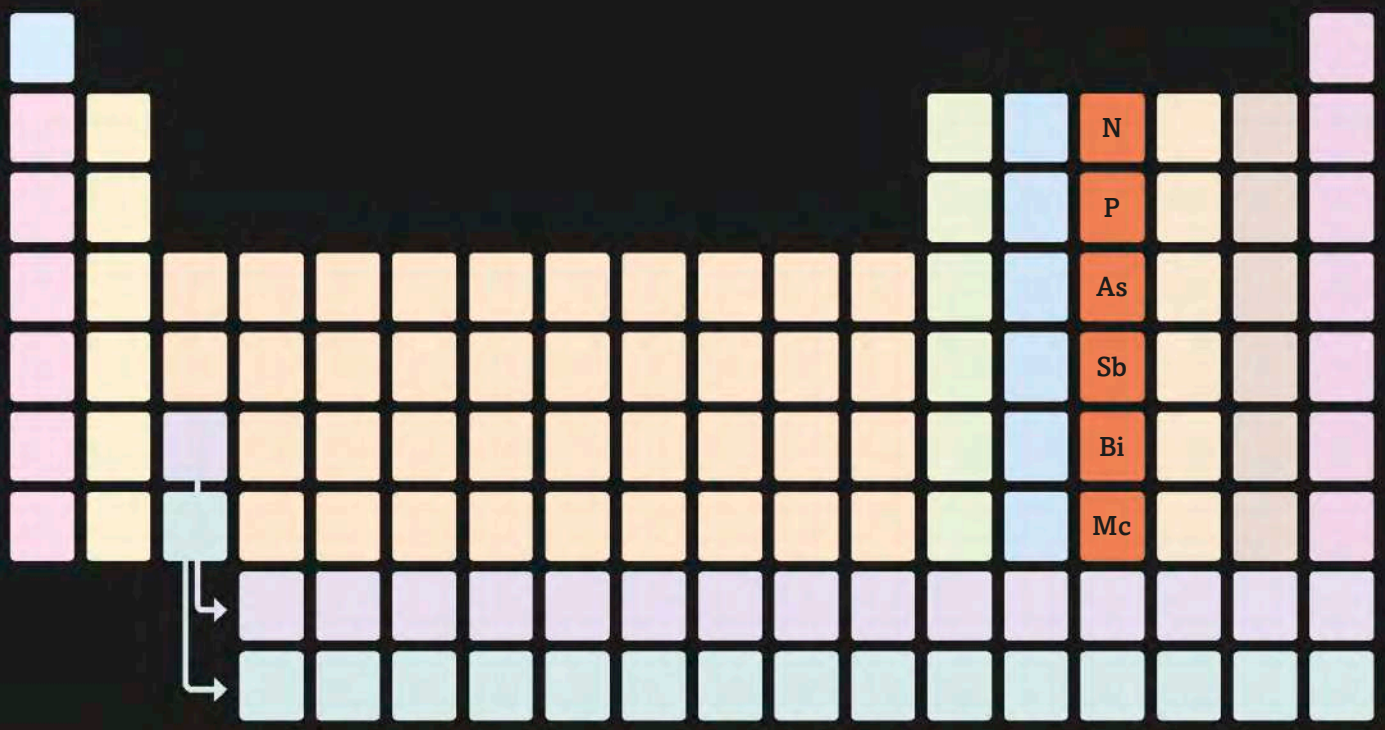

The Nitrogen Group

This group includes different types of natural element – non-metals, semi-metals, and dense metals – as well as moscovium (Mc), an artificial element. The group is also known as "pnictogens". This derives from the Greek word *pnígein*, which means "to choke" and refers to the potential toxicity of nitrogen (N) in certain forms.

Atomic structure
Members of this group have atoms with five electrons in the outer shell. These atoms can form up to three bonds at the same time.

Physical properties
All the members are solids, except nitrogen (N). The density of elements increases down the group: bismuth (Bi) is 8,000 times denser than nitrogen.

Chemical properties
Phosphorus (P), which exists in two main forms, is reactive, but the others in this group are fairly stable.

Compounds
When reacting with three hydrogen (H) atoms, all members of the group form reactive gaseous compounds called hydrides.

7 N Nitrogen

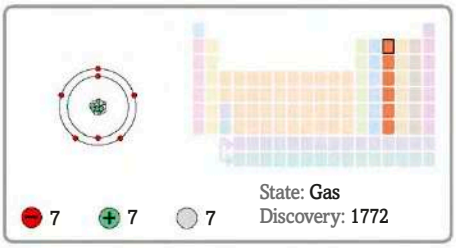

Forms

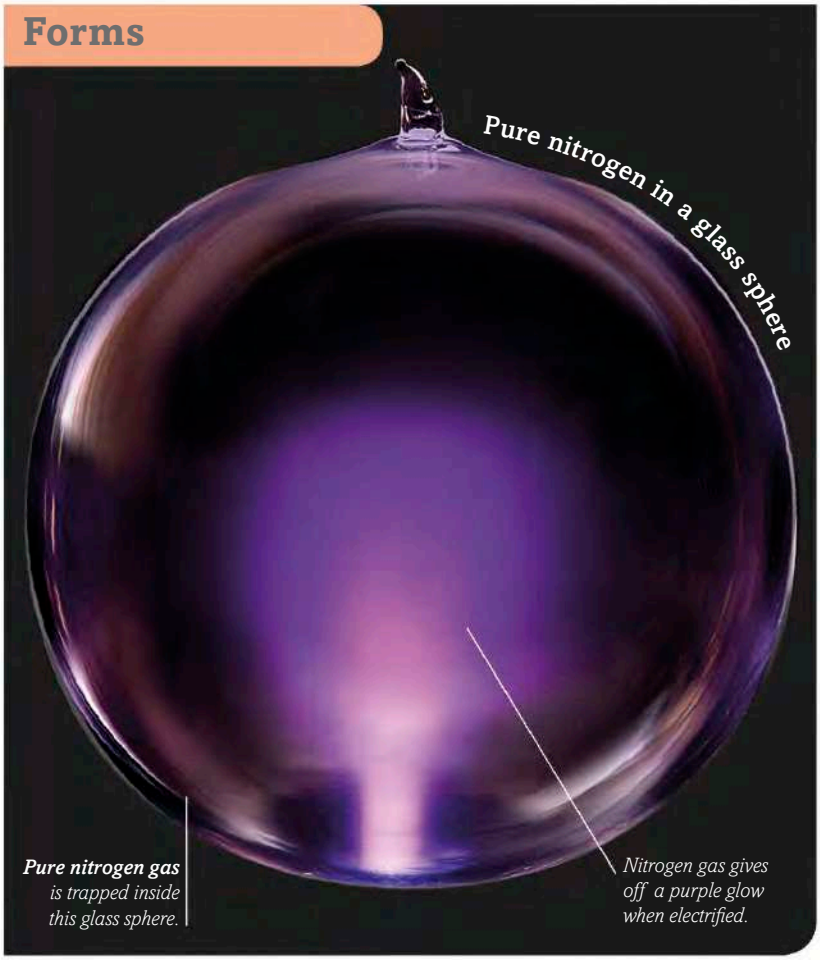

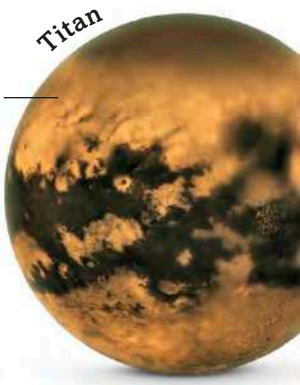

Saturn's largest moon *has an atmosphere containing 48% nitrogen.*

Liquid nitrogen

This clear and colourless liquid *forms when nitrogen is cooled to below –195°C (–319°F).*

Microscopic image of root nodule

Bacteria living inside plant roots *can take nitrogen from the air for use by the plant.*

This is a naturally *occurring form of sodium nitrate.*

Nitrogen surrounds us all the time because it is the transparent gas that makes up nearly three-quarters of Earth's atmosphere. Since **pure nitrogen** does not react easily, its **liquid form** can be used to freeze and preserve items such as blood and tissue samples.

Nitratine is one of the few minerals rich in nitrogen. Some useful nitrogen compounds can be made by industrial processes. A group of nitrogen compounds is used in explosives, including **TNT** and **nitroglycerine**. When ignited, they explode because the bonds

Uses

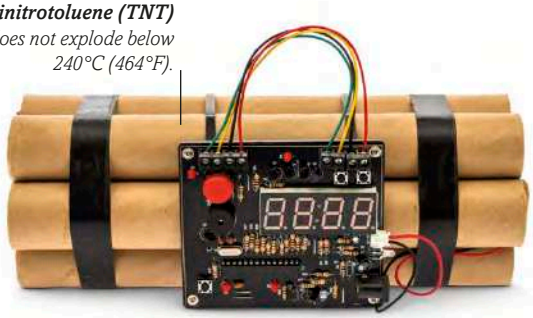

TNT

Trinitrotoluene (TNT) *does not explode below 240°C (464°F).*

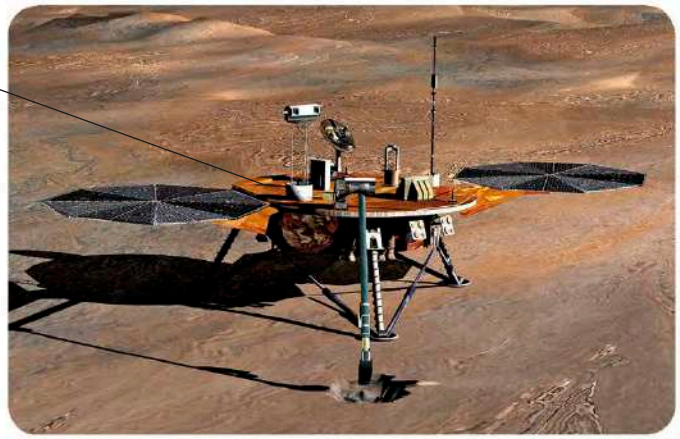

Phoenix Mars Lander

This lander *used 12 hydrazine-powered thrusters to land on Mars.*

Nitroglycerine spray

Nitroglycerine *can be used to treat heart conditions.*

Drag bike

Motorcycles *use nitromethane to fuel their powerful engines.*

Textile dye

These nitrogen-based azo dyes *are often used to colour textiles.*

Super glue

Superstrong glues *contain small nitrogen compounds that link together.*

Nitrogen fertilizer

These fertilizers *contain compounds called ammonium nitrates that promote growth in plants.*

NITROGEN CYCLE

Nitrogen is essential for life. The nitrogen cycle is a process that constantly recycles nitrogen between Earth's atmopshere and all living things.

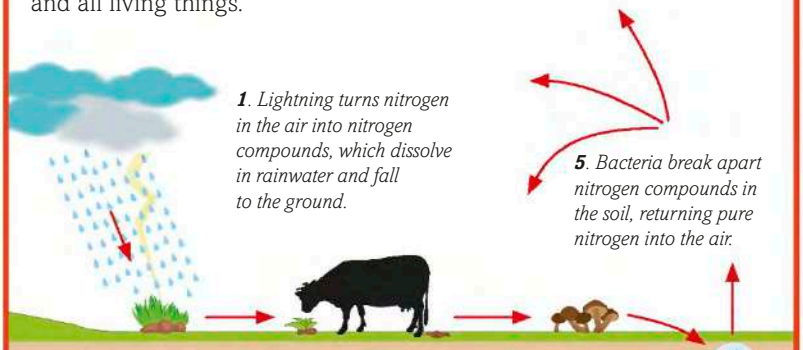

1. *Lightning turns nitrogen in the air into nitrogen compounds, which dissolve in rainwater and fall to the ground.*

2. *Bacteria in the soil and plant roots produce nitrogen compounds from pure nitrogen in the air.*

3. *Animals take in nitrogen compounds when they eat, and release it in their dung.*

4. *Fungi, such as mushrooms, break down dead plants and animals, which release their nitrogen compounds back into the soil.*

5. *Bacteria break apart nitrogen compounds in the soil, returning pure nitrogen into the air.*

between nitrogen atoms detach very quickly. Nitrogen fuels, such as nitromethane, are used in **drag bikes**, providing a lot more power than carbon-and-hydrogen only fuels, such as petrol. The compound hydrazine is used in thrusters on spacecraft, such as the **Phoenix Mars Lander**.

Some nitrogen compounds are put in **dyes** and **glues**. An industrial technique called the Haber process turns nitrogen and hydrogen gas into ammonia, a liquid commonly used to make **nitrogen fertilizers**. When mixed with soil, these fertilizers boost plant growth.

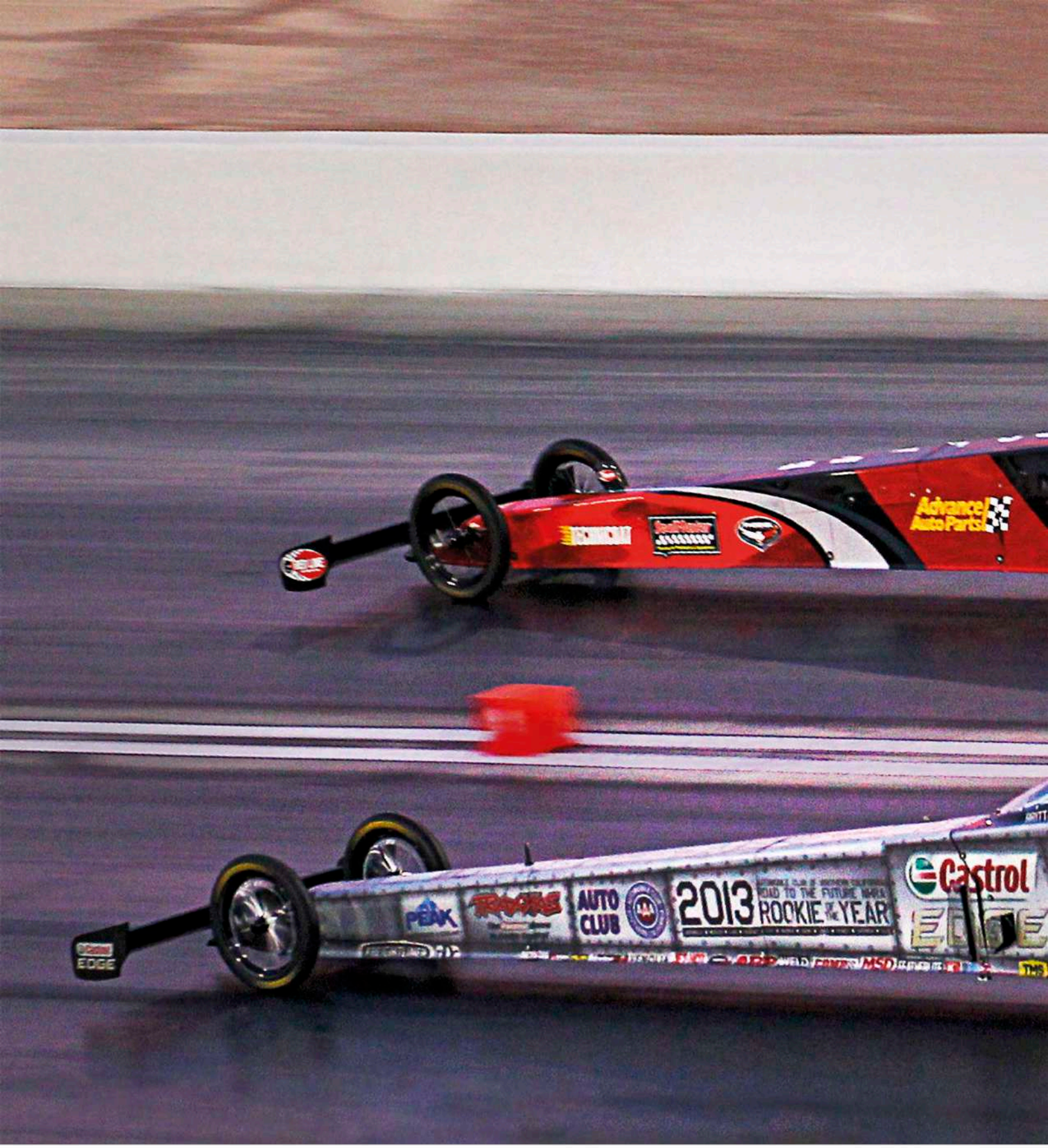

DRAG RACING Zooming along a dead straight track, these all-powerful dragsters accelerate all the way to the finish line. They contain massive engines filled with an extra-powerful fuel called nitromethane, which is often shortened to “nitro”. Burning eight times faster than regular petrol used in most cars, this super fuel can push dragsters to speeds in excess of 480 km/h (300 mph).

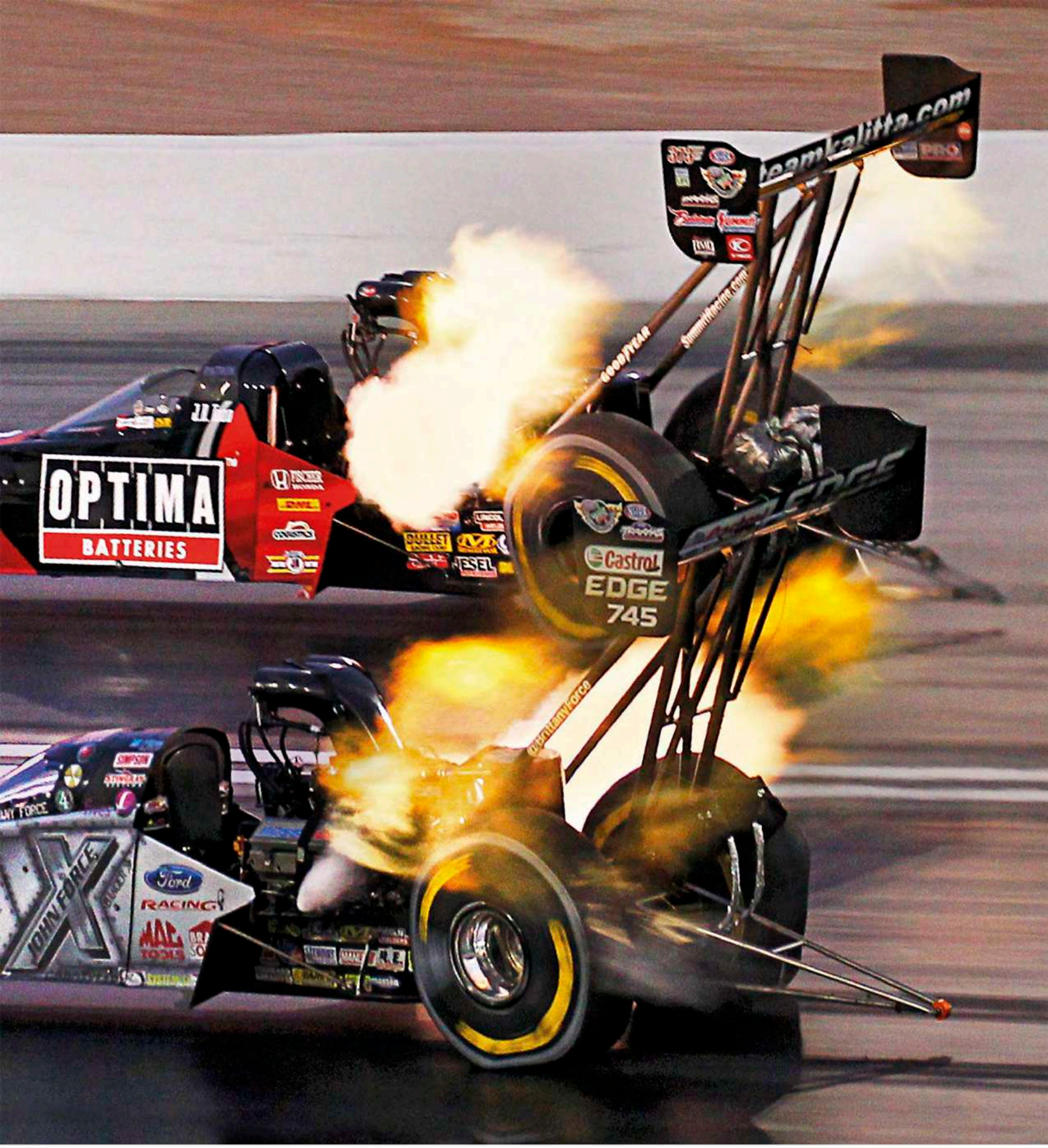

Nitromethane contains carbon, hydrogen, and nitrogen, but it is the latter element that really gives this fuel its immense power. During the process of combustion – when oxygen is mixed with the fuel in the race car's mighty engine – nitromethane burns so violently that nitrogen escapes from the fuel and returns to its pure form. This chain of events leads to a release of energy, which propels the dragsters to breakneck speeds. Although these races are an incredible spectacle to behold, driving using nitromethane can be dangerous due to the explosive nature of nitrogen when used in this way: drag racers are taking a risk to win.

15
P

Phosphorus

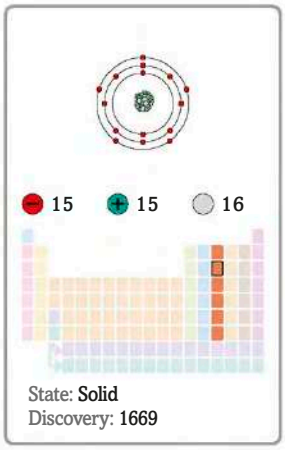

Forms

Phosphate mining in Nauru

Deep phosphate mines *make up 80% of this tiny Pacific island.*

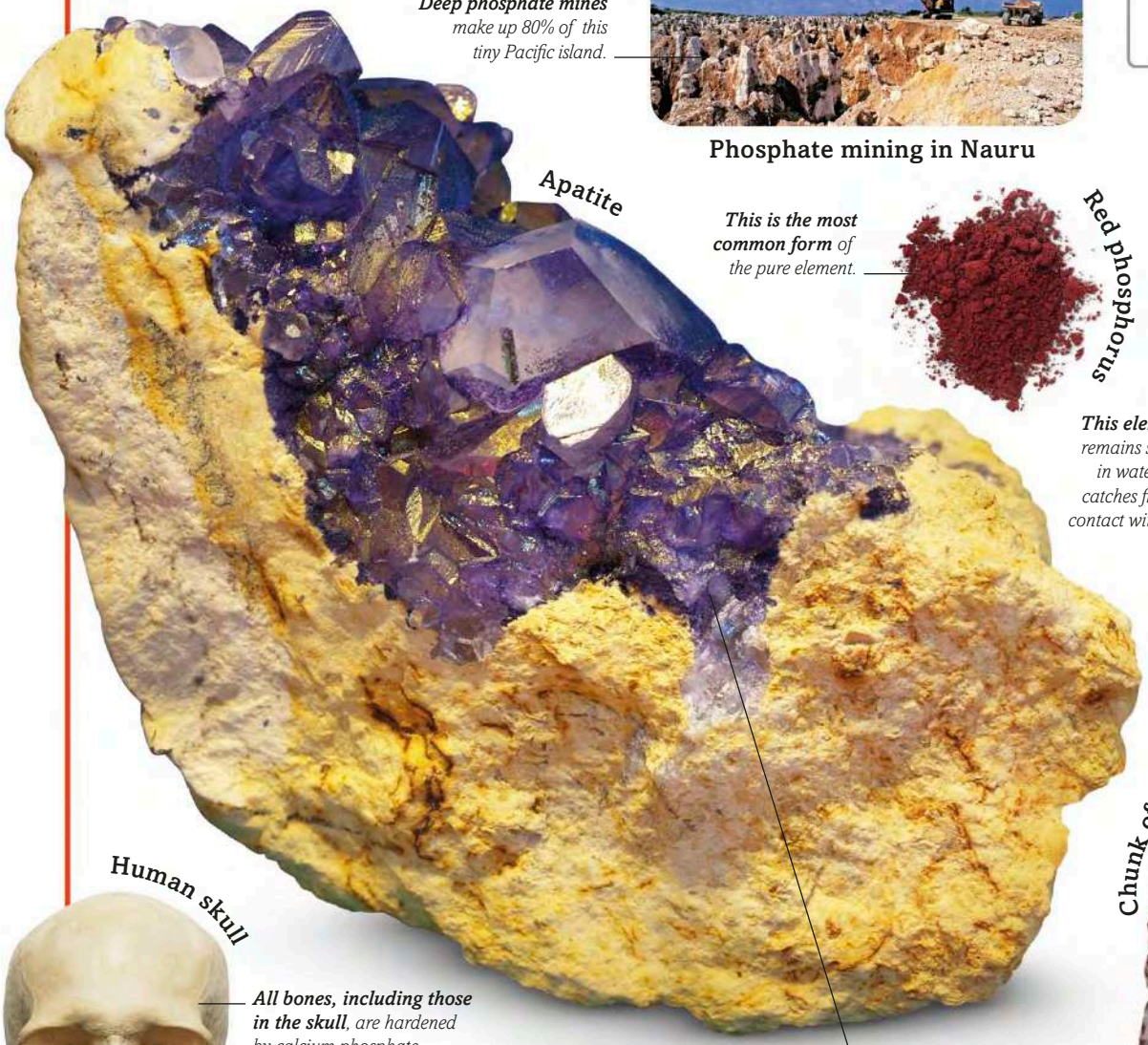

Apatite

This purple colour *of the crystals is due to metal impurities.*

Red phosphorus

This is the most common form *of the pure element.*

White phosphorus

This element *remains stable in water, but catches fire on contact with air.*

Chunk of violet phosphorus

This form of phosphorus *is made in a laboratory by heating red phosphorus.*

Human skull

All bones, including those in the skull*, are hardened by calcium phosphate.*

Tuna

This fish *is rich in phosphorus.*

Phosphorus was accidently discovered by German alchemist Hennig Brand. In 1669, in his quest for the mythical Philosopher's Stone (a material some believed could turn any metal into gold), he boiled a large pot full of urine for days. This produced a mysterious glowing substance, which he called phosphorus, meaning "giver of light". Phosphorus is the first element to have a discoverer with a recorded name. It is never pure in nature, and occurs in different minerals. Phosphorus has several flammable,

Uses

This light, strong china *contains calcium phosphate.*

China tea set

LIFE'S BUILDING BLOCKS

DNA – short for deoxyribonucleic acid – is like a mini database packed with instructions telling your body how to work properly. It consists of a chain of molecules and looks like a twisted ladder called a double helix. The edges are sugar molecules linked together by units containing phosphorus.

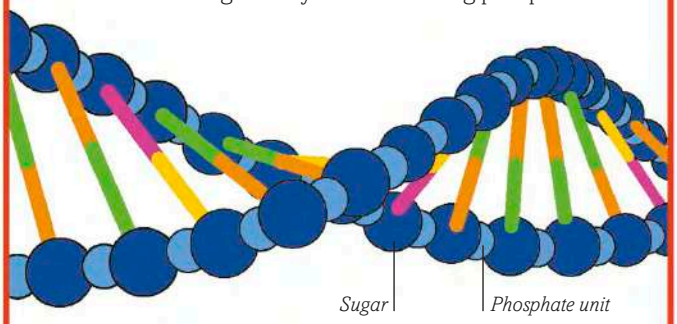

Safety match box

The striking surface on the sides of the box *contains phosphorus, which ignites the match upon contact.*

These flexible fibres *are composed of phosphate-rich glass.*

Fibre optics

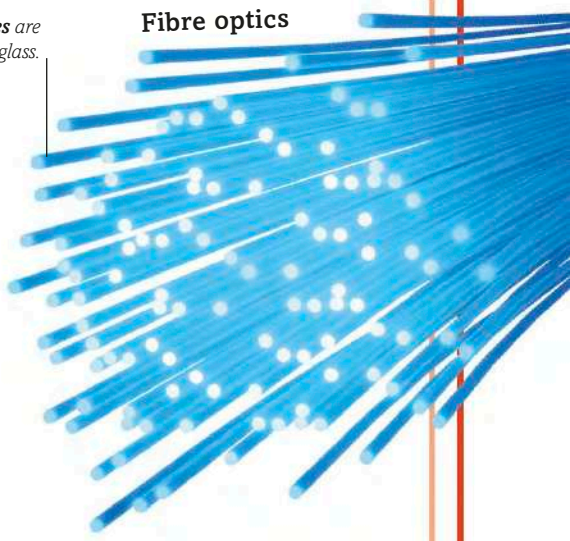

Spraying ammonium phosphate *over burning material cuts off its supply of oxygen. This extinguishes the fire.*

Fire extinguisher

Spraying crops with phosphates *kills pests, such as insects.*

Pesticides

Fertilizer

This fertilizer *contains ammonium phosphate to boost plant growth.*

solid forms, including **red**, **white**, black, and **violet**. The glow seen by Brand was caused by white phosphorus reacting with oxygen. Phosphorus is mainly found in **phosphate** minerals (in which phosphorus links to oxygen), such as **apatite**, its main ore.

Phosphates are present in **fine china**, and are an important ingredient in **fertilizers**. The strips on the sides of **safety match boxes** contain pure phosphorus. More complex phosphorus compounds used in **pesticides** are poisonous.

33 As Arsenic

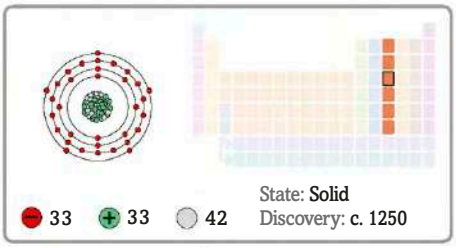

Forms

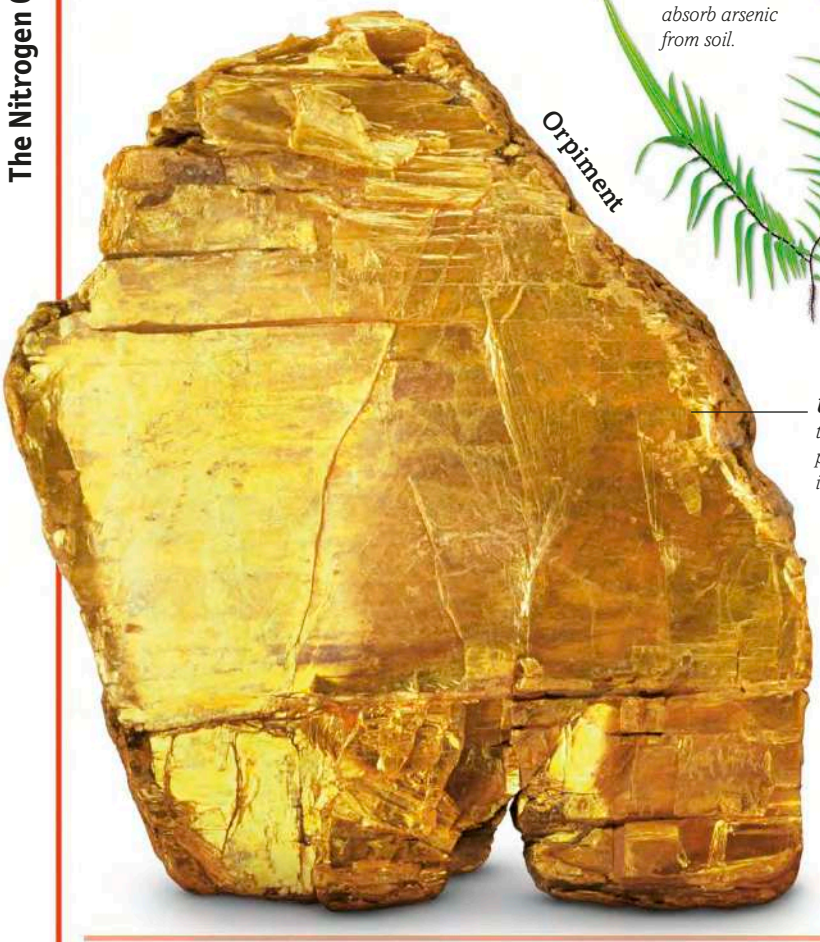

Orpiment

Chinese brake fern

These leaves *absorb arsenic from soil.*

Until the 1800s, *this mineral was powdered for use in paints.*

Pure arsenic refined in a laboratory

Metallic shine

This ore *is found in hot volcanic springs.*

Realgar

Uses

Bullets

This tip *is an alloy of arsenic and lead.*

Rat poison

This poisonous arsenic compound *kills rats.*

When heated, arsenic **doesn't melt**, instead it **turns into a gas**.

Car battery

The electrodes *inside this battery contain arsenic.*

Arsenic is often called the "king of poisons". Every form of arsenic – either pure or in a compound – is poisonous to animals. In fact, arsenic poisons have been used for centuries. This semi-metal is found in several minerals with striking colours, including **orpiment**. Naturally occurring **pure arsenic** has a shiny, grey colour. Arsenic compounds are used in making some **rat poisons**. The leading use of arsenic today is for strengthening lead. This is done by mixing arsenic with lead to create a tough alloy that is often used in **car batteries**.

51 Sb Antimony

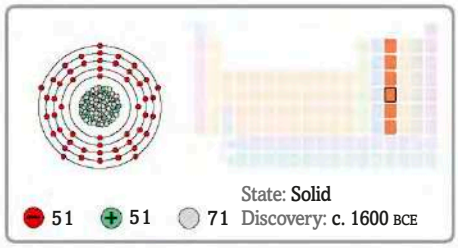

Forms

These needle-like crystals *contain antimony, lead, and iron.*

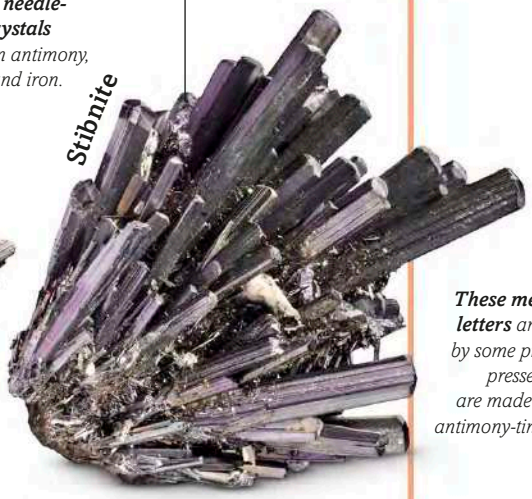

This mineral *tarnishes on contact with air.*

This silvery semi-metal *is hard but brittle.*

Pure antimony crystals refined in a laboratory

Uses

These metallic letters *are used by some printing presses, and are made of an antimony-tin alloy.*

Printing press metal type

These matches *with antimony in the tip burn brighter than the ones without it.*

Safety matches

Kohl *is a dark eyeshadow.*

Egyptian kohl

Kohl makes it **easier to see** in strong sunlight.

Antimony gets its name from the Greek word *anti-monos*, meaning "not alone". This may refer to the fact that the element is never found pure in nature, but is always found combined with heavier metals, such as lead. The element's symbol, Sb, comes from *stibium*, the Latin word for kohl, a form of eye make-up. The mineral ore **stibnite** is the largest source of **pure antimony**. Its pure form is mostly used to make hard alloys, such as that in the **metal type** used by some printers. Ancient **Egyptian kohl** was made from powdered stibnite.

83 Bi Bismuth

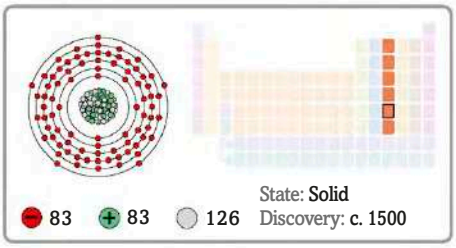

Forms

Bismuthinite

This mineral is the main source of pure bismuth.

Hopper crystals refined in a laboratory

These rainbow colours form when the metal on the surface begins to react with oxygen.

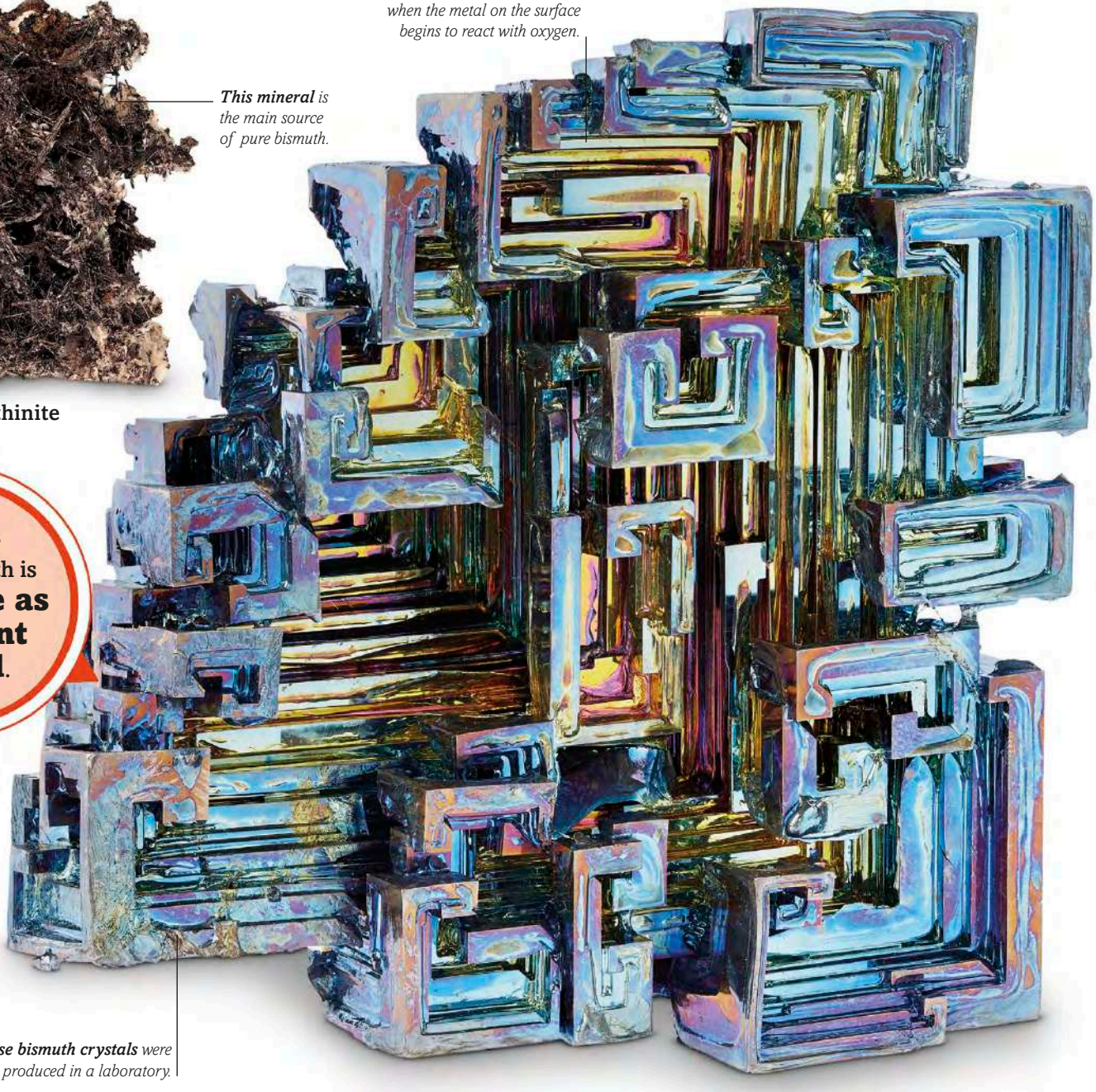

These bismuth crystals were produced in a laboratory.

In Earth's crust, bismuth is about **twice as abundant as gold**.

Bismuth is a radioactive element but its atoms are relatively stable and last for millions of years. People have known about bismuth for centuries. The Incas of South America added it to weapons made of the alloy bronze to harden them, while ancient Egyptians used a bismuth mineral to make their cosmetics glittery. Pure bismuth forms an oxide in air that is seen as colourful crystals called **hopper crystals**. This element is very brittle and has few uses when not in a compound form. Yellow bismuth

Uses

This cold box *uses a compound called bismuth telluride, which becomes cold when electrified, and keeps items cool.*

Portable refrigerator

This nail polish *gives a pearly effect because of bismuth compounds.*

Yellow cosmetics

This medicine *contains a bismuth compound that helps settle an upset stomach.*

Digestive medicine

Unlike most elements, bismuth's liquid state is heavier than its solid.

pigments are used in paints and **cosmetics**, while several bismuth compounds are also in **medicines**. An alloy of bismuth and tin is an ingredient in fire sprinklers.

115 Mc Moscovium

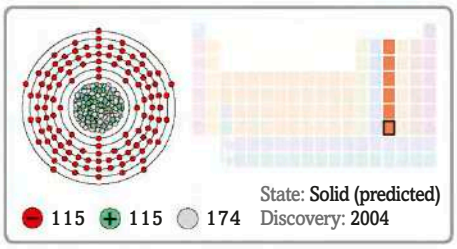

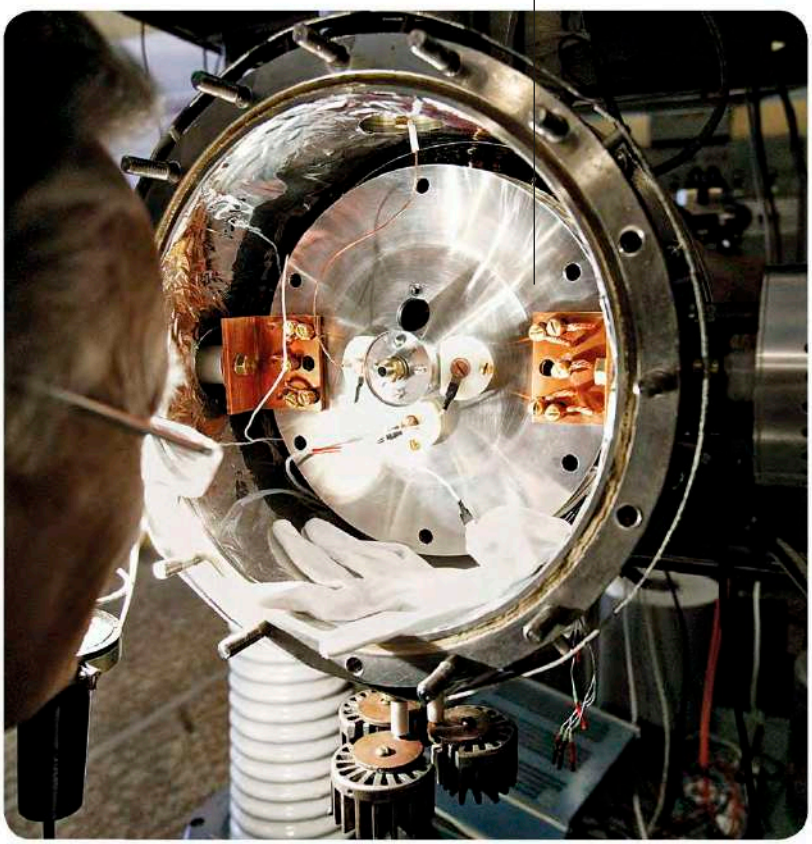

This is one of the machines *in this research centre.*

Joint Institute for Nuclear Research, Dubna, Russia

Only about a hundred or so atoms of this heavy, artificial element have been made. Moscovium was first created at the **Joint Institute for Nuclear Research** in Dubna, Russia. A team of Russian scientists, led by Yuri Oganessian, created this element by smashing americium atoms with parts of calcium atoms. It is named after the Russian capital city of Moscow. This element is extremely radioactive, and its atoms break up within a fraction of a second. Scientists think that moscovium would be a dense, metallic solid but with such small samples, they can only measure how big the atoms are before they break up.

Pure tellurium (Te) can form silvery crystals.

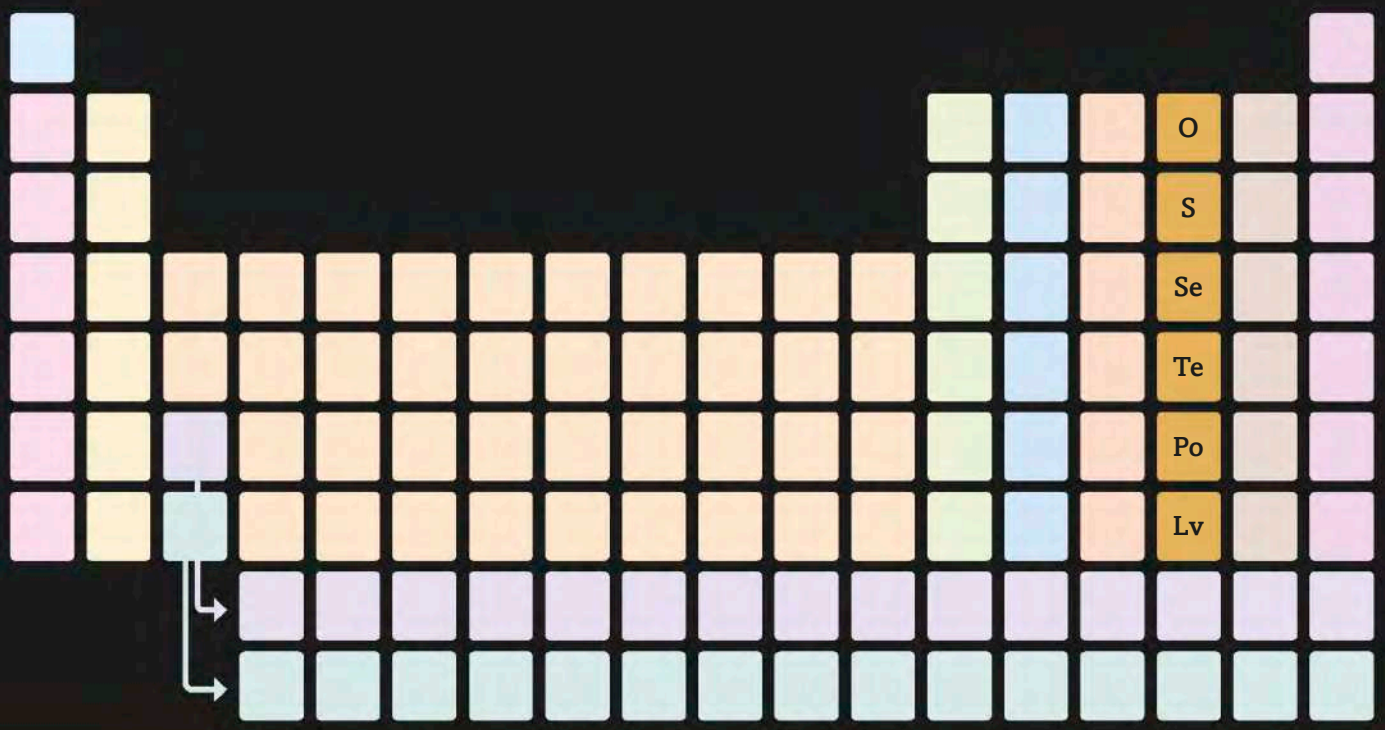

The Oxygen Group

This group does not include any natural metals. The first two members, oxygen (O) and sulfur (S), are non-metals widespread in nature. The remaining three natural elements are semi-metals. Only the artificial member, livermorium (Lv), is thought to be a metal, but chemists don't really know for sure.

Atomic structure
All members have six electrons in the outer shell of each atom. This electron structure makes these elements highly reactive.

Physical properties
The members of this group are solids, except oxygen (O), which is a gas at room temperature. The density of the elements increases down the group.

Chemical properties
The reactivity of these elements decreases down the group. Oxygen is always involved in the process of combustion.

Compounds
These elements can form compounds with each other. They all react with carbon (C) to form compounds, some with strong smells.

8 O Oxygen

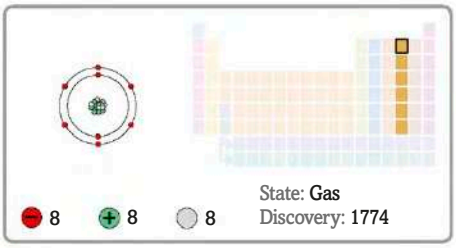

Forms

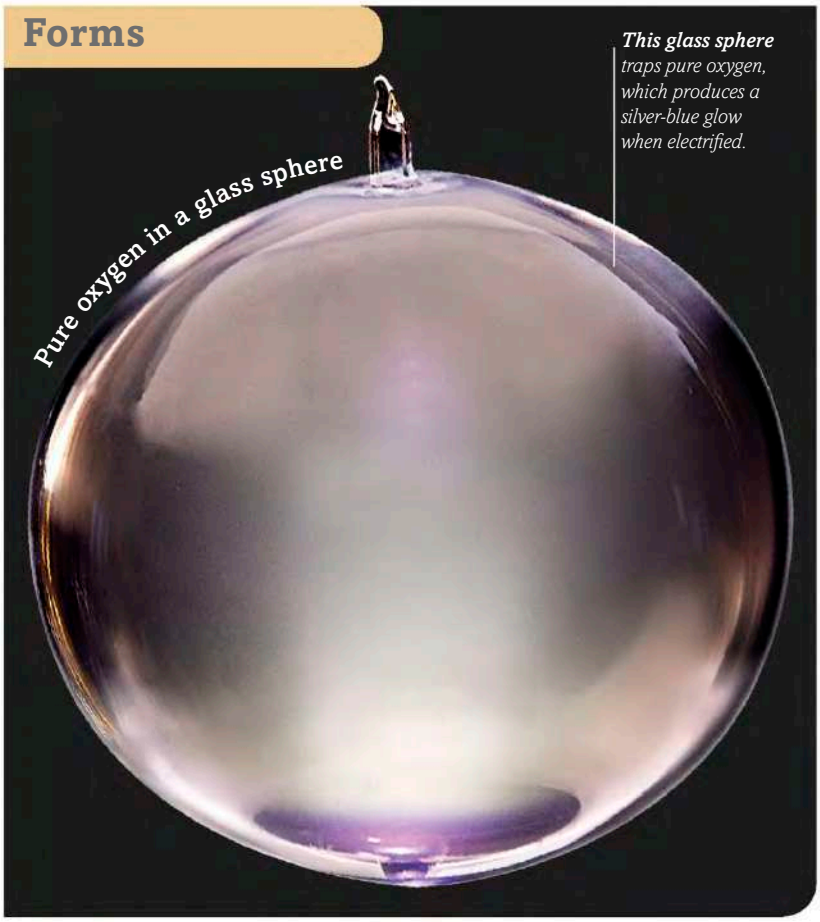

This glass sphere *traps pure oxygen, which produces a silver-blue glow when electrified.*

Streaks of light *adorn the sky as atoms of oxygen in the air are hit by a stream of particles blasting from the Sun.*

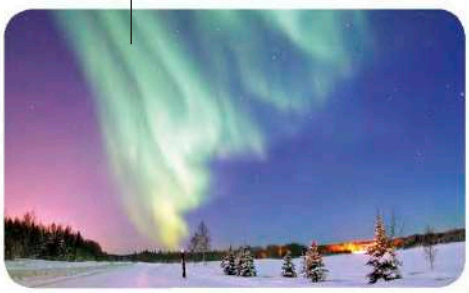

Northern lights

Plants *release oxygen in sunlight.*

Flames *are produced when oxygen reacts with fuel, such as wood.*

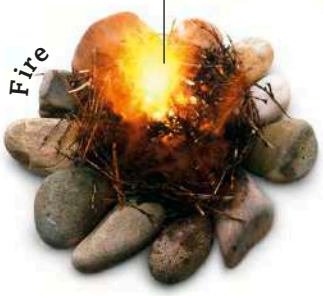

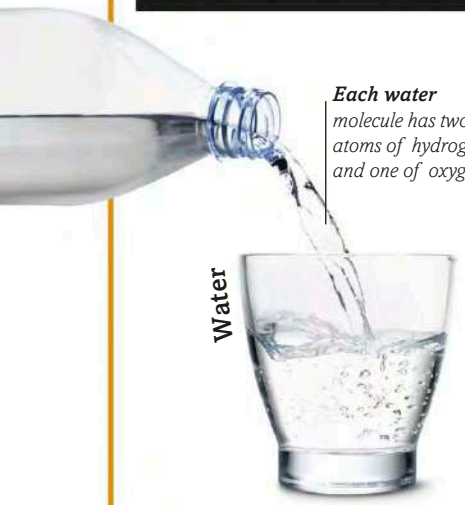

Each water *molecule has two atoms of hydrogen and one of oxygen.*

WHAT IS COMBUSTION?

Combustion is a chemical reaction that produces heat and light. Oxygen is essential for combustion to take place.

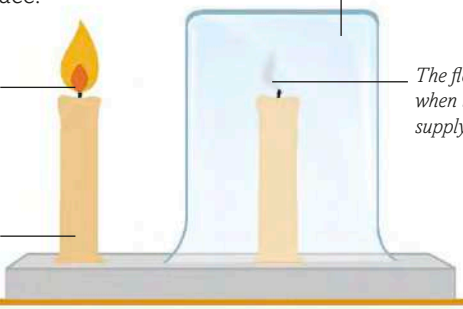

The flame is steady in the presence of oxygen.

Candle wax is made of a flammable carbon compound, which acts as a fuel.

A beaker is placed over a candle, starving the flame of oxygen.

The flame goes out when the oxygen supply is cut off.

Oxygen is the most common element in Earth's crust. Oxygen and its compounds make up half of all rocks and minerals on our planet. In the atmosphere, **pure oxygen** makes up around one-fifth of the air. This element is a transparent gas. Life on Earth depends on oxygen for survival. Animals breathe in air to collect the oxygen in it. Our bodies' cells then use that oxygen to break apart sugars to release energy, which powers our bodies. Another process that involves oxygen is the burning reaction called combustion, in which oxygen

Uses

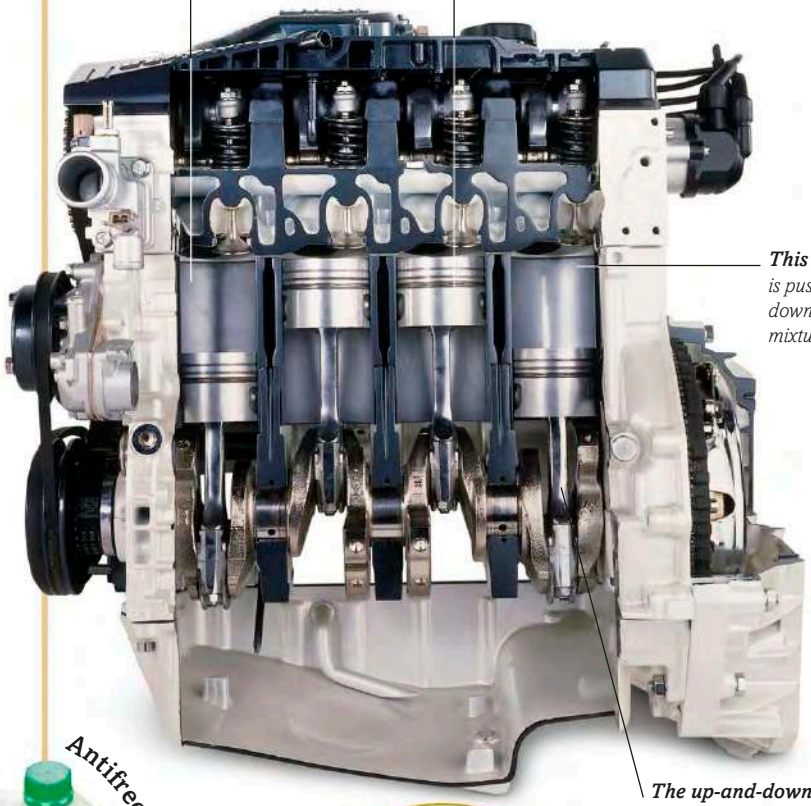

This cylinder *takes in oxygen along with the fuel.*

The piston *then squeezes the mixture of oxygen and fuel.*

This piston *is pushed down when the mixture explodes.*

The up-and-down motion of the pistons *makes the wheels of a car spin.*

Internal combustion engine

Hot steel *may have impurities that are burned away using pure oxygen.*

Steelmaking

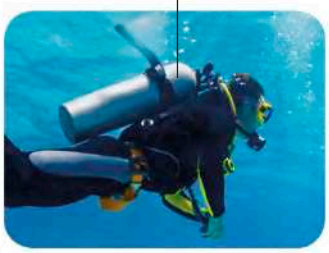

The oxygen in this cylinder *can last for up to 30 minutes.*

Scuba diver

Antifreeze

This liquid *contains an oxygen compound that prevents water from freezing in an engine.*

Oxygen is **colourless** as a gas, but looks **pale blue** as a liquid.

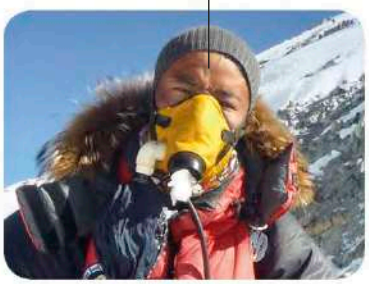

Mountaineers *have to adjust slowly to changing levels of oxygen as they climb higher.*

Mountaineer with oxygen supply

Medical oxygen cylinder

The oxygen in this cylinder *is given to patients in controlled amounts.*

Welding torch

This flammable gas-oxygen mixture *burns hot enough to melt metal.*

This rocket *carries liquid oxygen that mixes with the fuel to produce hot gases, which help the rocket lift off.*

Atlas V rocket

reacts with a fuel and produces **fire**. Oxygen is also used up when it reacts with other elements to form compounds called oxides. However, it is replenished by **plants** through a process called photosynthesis, which releases fresh oxygen. Car **engines** are powered by the combustion of petrol or other fuels. Oxygen is also useful in the **making of steel**. Tanks of oxygen let **mountaineers** breathe easily in environments that have low levels of this gas. Rockets, such as the **Atlas V**, carry liquid oxygen to burn fuel in the absence of air in space.

16 S Sulfur

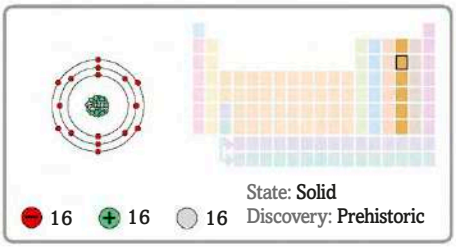

Forms

These yellow sulfur crystals *often attach to volcanic mud.*

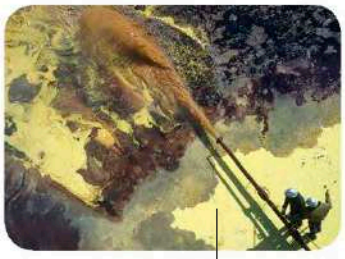

Hot liquid sulfur *is pumped up from an underground mine.*

This grey-blue crystal *contains a sulfur compound called strontium sulfate.*

Soft, brittle granules

Skunks *release a foul spray that contains three kinds of sulfur compounds.*

Cut onions *release sulfur compounds that make our eyes water.*

Hydrogen sulfide bubbles *form in volcanic mud.*

Known since ancient times, sulfur is one of the few non-metals that can be found pure in nature. This yellow, crystalline element is found in large amounts near volcanic craters. Another name for sulfur is "brimstone", which refers to the way its crystals burn, melting into a blood-red liquid. In some religions, brimstone is thought to be the fuel that burns in the underworld. **Pure sulfur** is extracted from underground deposits using hot water. The hot **liquid sulfur** is then pumped to the surface. This element is a common ingredient in many

Uses

This plant *emits a rotting smell to attract meat-eating insects.*

Titan arum

Vulcanized rubber *– made by heating sulfur with natural rubber – is weatherproof.*

Vulcanized tyre

Preserved dried fruits

Some dried fruits *are preserved using powder containing a sulfur compound.*

When burned, sulfur in this candle *drives away pests.*

Sulfur candle

ACID RAIN

Sulfur dioxide gas produced by burning fuel dissolves in rain water, making a sulfur acid. This falls to the ground as acid rain.

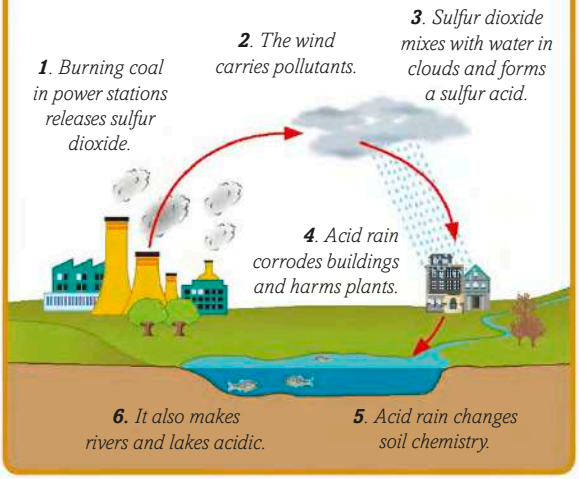

Creams that contain sulfur compounds *can disinfect skin.*

Skin cream

Lead acid battery

This battery *contains strong sulfuric acid.*

Penicillin pills

Some antibiotics *contain sulfur compounds that kill harmful bacteria.*

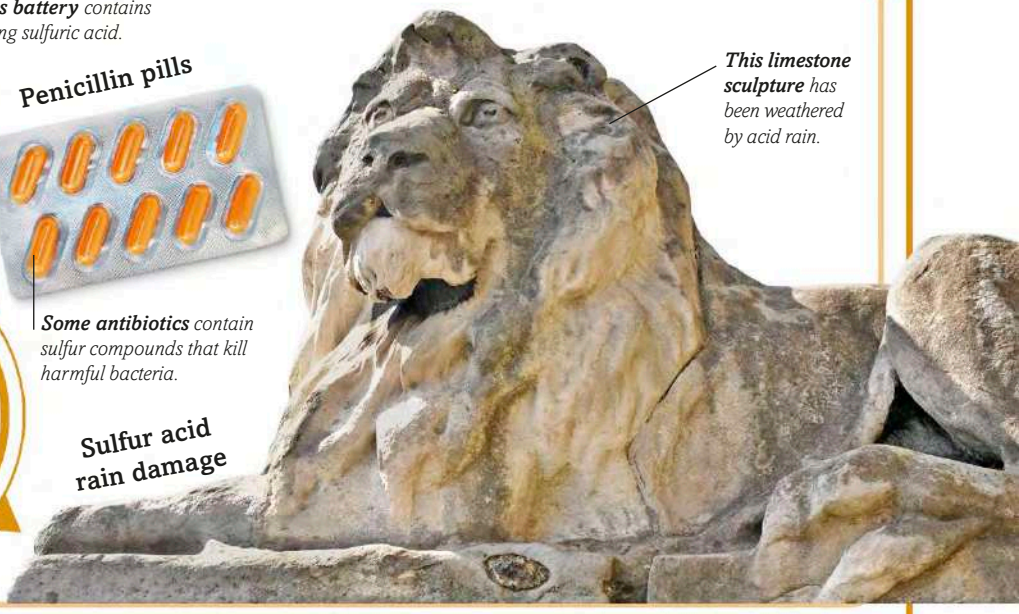

This limestone sculpture *has been weathered by acid rain.*

Sulfur acid rain damage

Acid rain **destroys forests** by removing nutrients in the soil and in the leaves.

minerals, such as **celestine**. Many sulfur compounds smell bad. For example, the rotten-egg smell of volcanic pools is due to hydrogen sulfide gas. Other examples include **skunk** spray, the gaseous substance emitted by chopped **onions**, and the odour of the **titan arum** flower. There are many uses for this non-metal. Its compounds can harden natural rubber for use in **tyres**, preserve **dried fruits**, and make strong **battery** acids. The element has anti-bacterial properties and is used in antibiotic medicine, such as **penicillin**.

DANAKIL DEPRESSION This hot spring in Africa's Danakil Depression is surrounded by a yellow crust of pure sulfur. The sunken region between Ethiopia and Eritrea in East Africa is a wild volcanic area, packed with erupting craters, arid deserts, boiling mud, and pools with unusual colours caused by the presence of sulfur and many mineral salts.

At more than 100 m (330 ft) below sea level, the Danakil Depression is one of the lowest points on Earth's surface. This area receives little or no rainfall and the weather is hot and dry, with temperatures soaring above 50°C (120°F). The scalding green water of the springs inside the depression contains pure sulfur as well as a toxic sulfur compound called sulfuric acid. As the water evaporates, sulfur deposits build up around the edges of the pools, making beautiful shapes across the vast landscape. Tourists visit to marvel at the remarkable sights at Danakil, even though the inhospitable conditions in the area give it the title of the "cruellest place on Earth".

34 Se Selenium

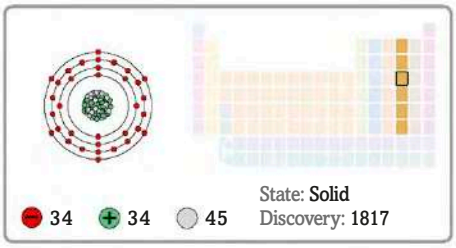

Forms

This form of the element *has a metallic sheen on its surface.*

Chunk of pure grey selenium refined in a laboratory

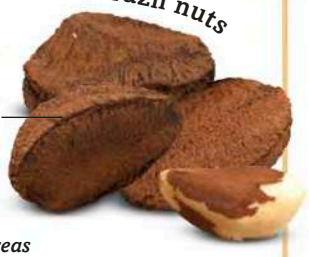

Brazil nuts

These nuts *are the richest food source of selenium.*

Berzelianite

These dark areas *contain selenium and copper.*

Uses

This calculator *is powered by a solar cell made of selenium and nickel.*

Calculator

Anti-dandruff shampoo

A selenium compound *in this shampoo treats dandruff.*

This bright colour *comes from the red selenium in the glaze.*

Ceramic vase

Photocopier

Many office photocopiers *use selenium in powdered form.*

Selenium is named after Selene, the Greek goddess of the Moon. This element is a semi-metal and so has the properties of both metals and non-metals. Selenium has two main pure forms: **grey selenium**, which is a hard substance, and red selenium, which is a soft powder. The most common use of selenium is as an ingredient that provides colour in glass and **ceramics**. Selenium is sensitive to light, so it is used in solar cells that convert sunlight into electricity. It is also utilized in **photocopy machines**.

52
Te

Tellurium

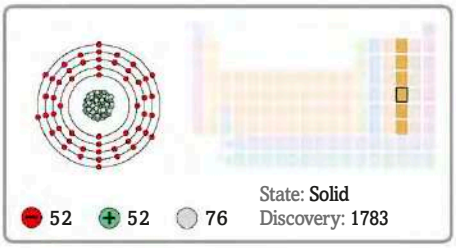

Forms

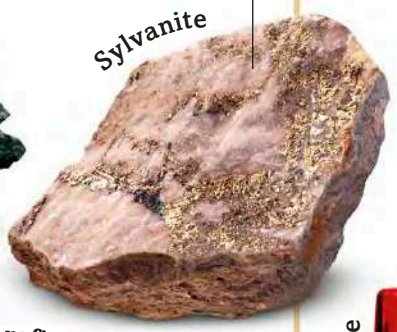

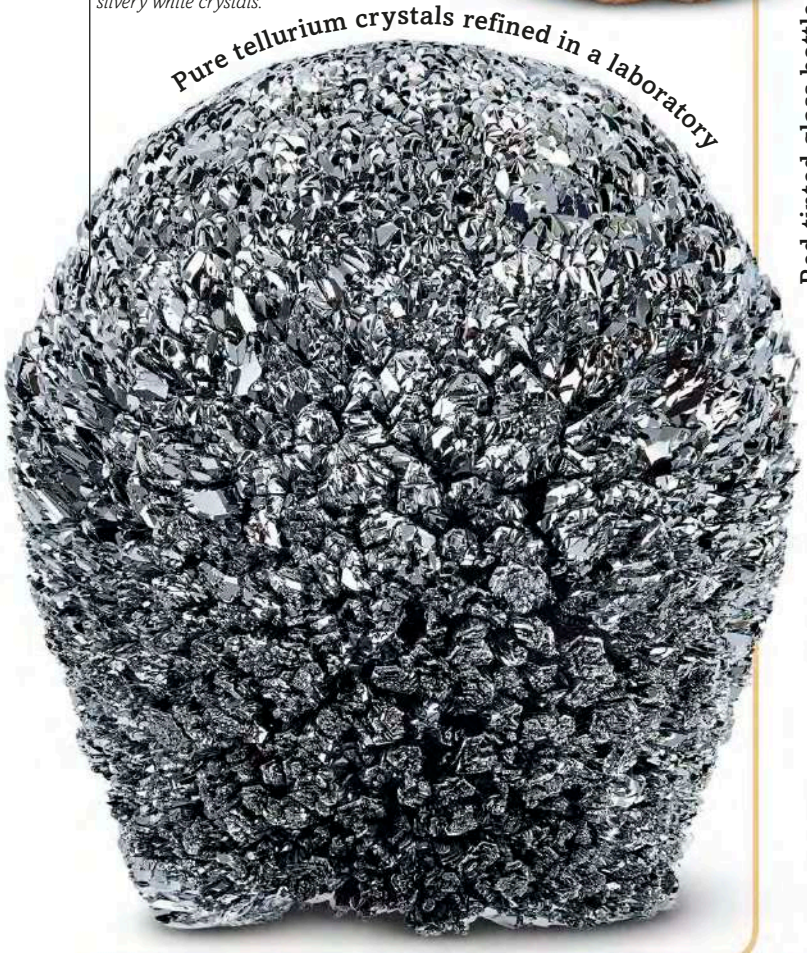

Uses

Tellurium **protects** bronze **from corroding** easily in air.

Tellurium is one of the 10 rarest elements on this planet. It gets its name from the Latin word *tellus*, which means Earth. This element is often found as a compound with another element, such as the metal nickel, as in the case of the ore **melonite**. Tellurium is also produced as a waste product when lead and copper are refined. **Pure tellurium** can take on two forms: a shiny, metallic solid or a brown powder. This element is mainly used in the glass of **fibre optics**, which carry high volumes of information much faster than copper cables.

84 Po Polonium

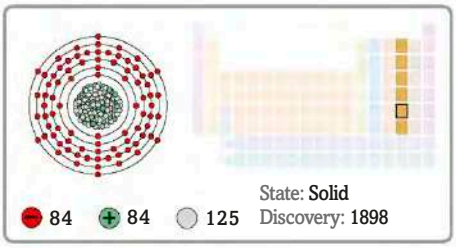

Forms

This uranium ore contains **0.0000001 per cent** polonium.

This mineral *contains uranium atoms that break apart to form atoms of polonium.*

Uses

This brush *is used to remove static electric charge from camera lenses and music records.*

This bomb *can be detonated when the polonium inside it ignites.*

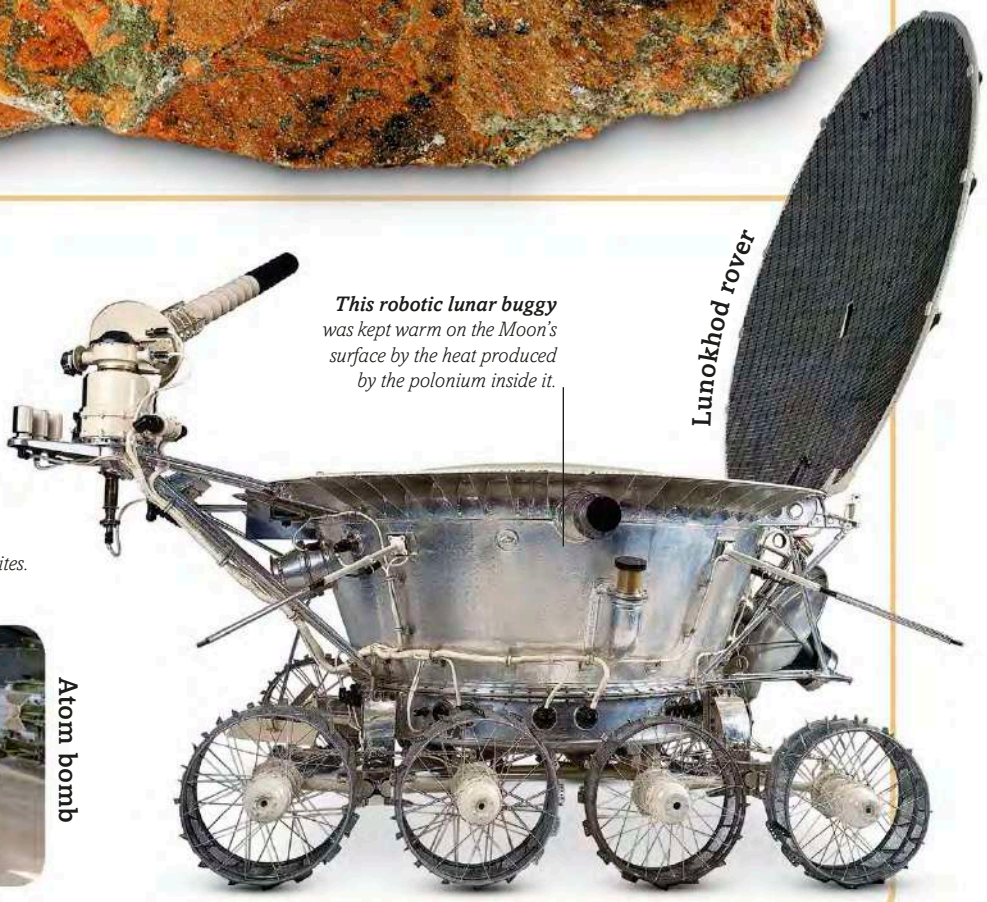

This robotic lunar buggy *was kept warm on the Moon's surface by the heat produced by the polonium inside it.*

Polonium is very radioactive: 1 g (0.03 oz) of this metal quickly heats up to 500°C (932°F) because of the radiation it emits. This element was discovered by Marie and Pierre Curie in 1898. Marie named it after Poland, her homeland. It is rare in nature, and is normally produced in nuclear reactors. Despite its radioactivity, this element is used in a few ways. It can trigger the explosion of **atom bombs**. It heats and powers spacecraft, such as the Russian **Lunokhod rovers**, which landed on the Moon in the 1970s.

116
Lv

Livermorium

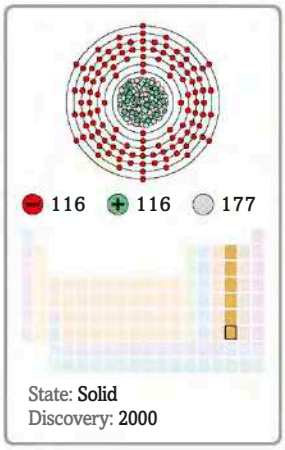

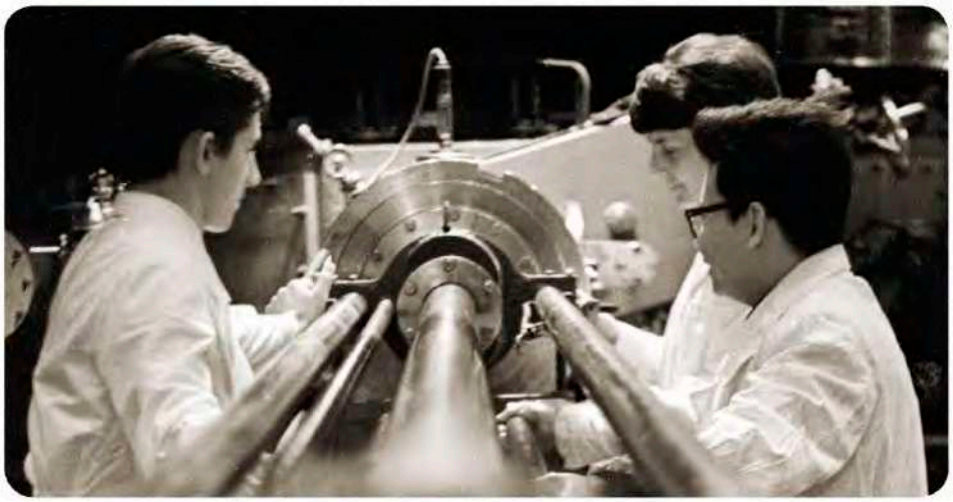

Heavy ion cyclotron,
Joint Institute of Nuclear Research, Dubna, Russia

Livermorium has been **named after** this **laboratory**.

Lawrence Livermore National Laboratory, California, USA

When livermorium atoms were first produced in 2000, they broke apart in a fraction of a second. The first successful attempt to create atoms of this element was made at the **Joint Institute of Nuclear Research** at Dubna, Russia. The team worked with material provided by the **Lawrence Livermore National Laboratory** in California, USA. This highly radioactive element was produced by firing calcium atoms at curium atoms in a particle accelerator (a machine in which atoms are smashed together).

This glass sphere contains pure iodine (I).

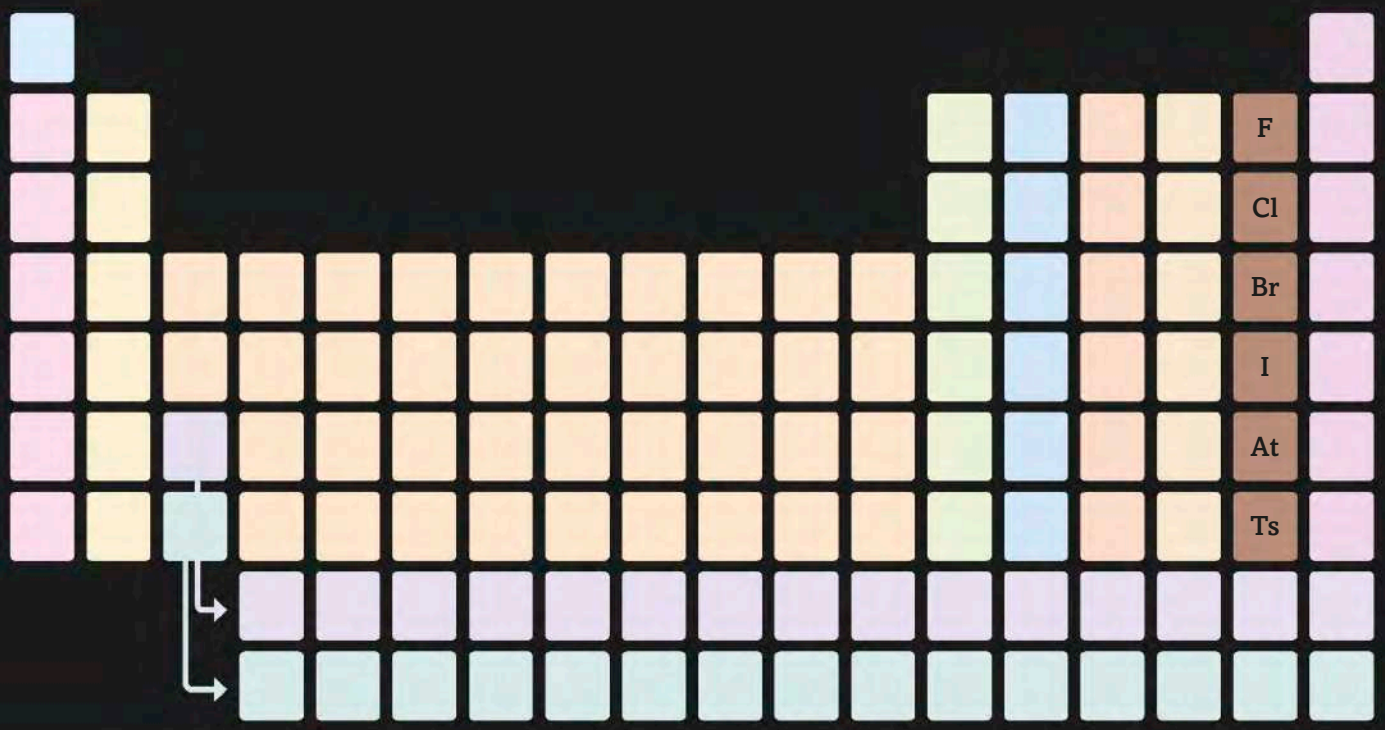

The Halogen Group

One of the most reactive groups in the periodic table, this set includes non-metals. The name "halogen" means "salt former", which refers to the way that elements in this group react with metals to form salts, such as sodium chloride, widely known as common salt. Scientists don't know much about tennessine (Ts), an artificial halogen.

Atomic structure
All members have seven electrons in the outer shell of each atom. There is space for one more electron in each outer shell.

Physical properties
Bromine (Br) is the only halogen that is liquid. Fluorine (F) and chlorine (Cl) are gases, while iodine and astatine (At) are solids.

Chemical properties
Every halogen atom receives one electron from other atoms to form a compound. Reactivity decreases down the group.

Compounds
The halogens react with hydrogen (H) to form acidic compounds. Halogen compounds are used in products such as bleach.

9 F Fluorine

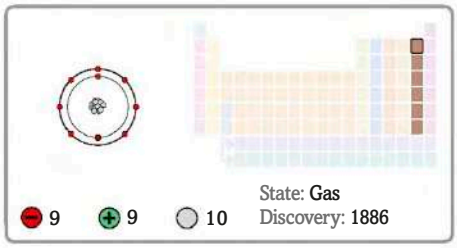

Forms

This soft mineral *is brittle, which means it breaks easily into lots of pieces.*

Cryolite

Laboratory sample

This sealed case *contains a mixture of fluorine and helium.*

Topaz **means "fire"** in the ancient Indian language of Sanskrit.

These cubic crystals *are green due to impurities.*

Fluorite

Topaz

This precious gemstone *contains 20.7% fluorine.*

This highly reactive element is incredibly dangerous when pure: just a tiny amount added to the air can kill a person. A pale yellow gas, fluorine reacts with brick, glass, and steel, burning a hole straight through them. Because it is so dangerous, pure fluorine is often stored in nickel containers that can resist its attack. Minerals such as **cryolite** and **fluorite** contain this element. This gas and its less harmful compounds have a wide variety of uses. Hydrofluoric acid is a toxic liquid used to etch patterns on glass, as seen in some **glass vases**.

Uses

Etched glass vase

These patterns *form after the surface of the glass is burned by an acidic fluorine compound.*

These circuit breakers*, containing a compound of fluorine and sulfur, cut electricity supply in an emergency.*

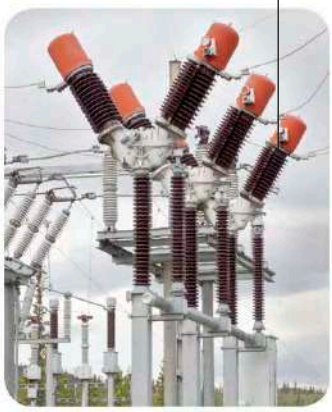

Circuit breakers

Ceramic pot

The shiny coating *on this ceramic dish is from a fluorine-containing glaze.*

Fluorine-rich plastic clothing *repels water.*

Waterproof clothing

Upon injection, this fluorine-rich liquid *heals damaged tissues by carrying oxygen to them.*

Oxycyte

PTFE was used by NASA to make **spacesuits** because it protected against heat.

This pan *has a PTFE coating that resists heat.*

Non-stick pan

Some toothpastes *contain fluorides that harden tooth enamel.*

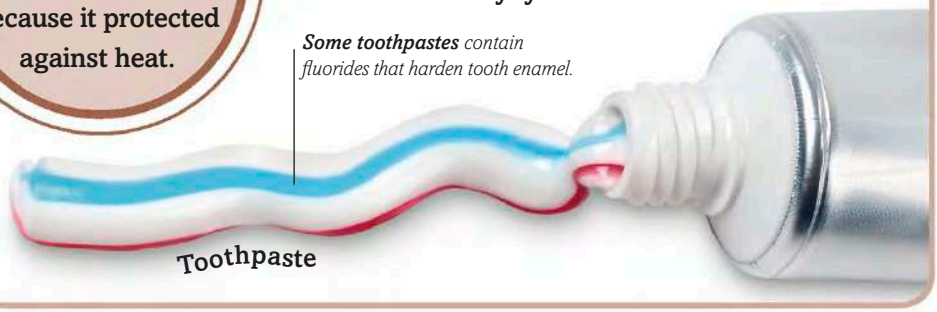

Toothpaste

HENRI MOISSAN

In the early 1800s, chemists in Europe realized that minerals such as fluorite contained an unknown element. However, it took another 70 years before the French chemist Henri Moissan extracted pure fluorine following a series of dangerous experiments that even poisoned him several times.

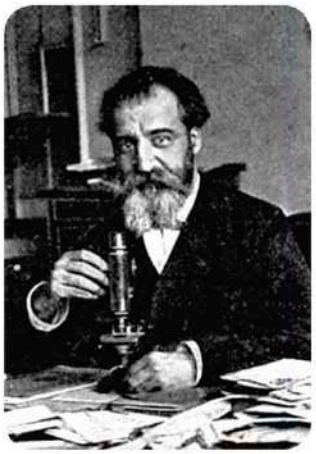

Some glazes used to coat ceramics contain fluorine minerals. When heated, these glazes release fluorine, which hardens the ceramic underneath. Another compound called polytetrafluoroethylene (PTFE) is commonly used to make **non-stick pans**: this material is slippery and prevents food that has burned while cooking from sticking to the pan. Thin fibres made of PTFE are also used to make lightweight, **waterproof clothing**. One of the most common uses of fluorine compounds is in **toothpaste**: they toughen teeth against decay.

17 Cl Chlorine

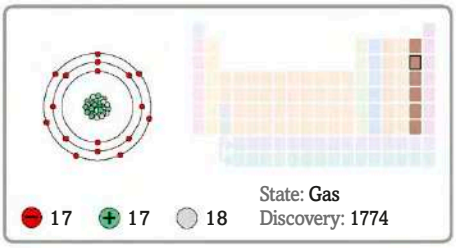

Forms

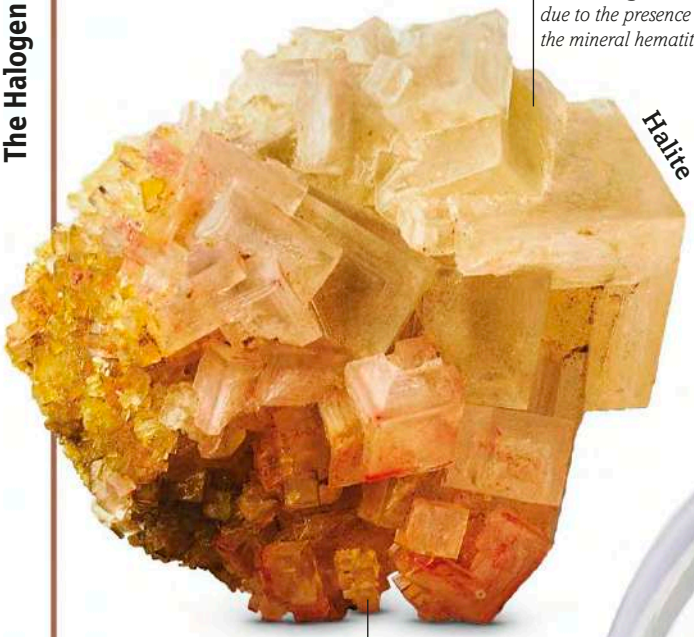

Halite

This orange colour *is due to the presence of the mineral hematite.*

Cube-shaped crystals

Red-eyed tree frog

The skin *of this tree frog contains a chlorine compound.*

Pure chlorine in a glass sphere

This glass sphere *traps pure chlorine so as to prevent any reactions with air.*

Pure chlorine *is heavier than air.*

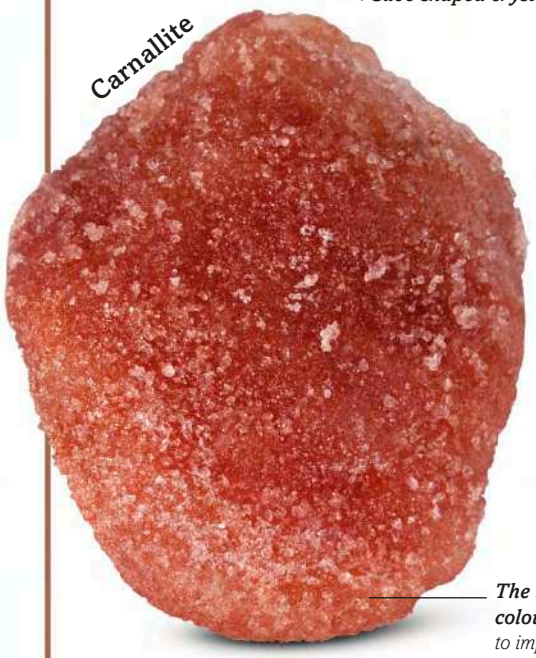

Carnallite

The intense red colour *is due to impurities.*

Chlorine is named after the Greek word *chlóros*, which means "pale green", a reference to the colour of this gaseous element. Chlorine is a highly reactive gas that forms a number of compounds, and does not exist pure in nature. The most common chlorine compound is sodium chloride, found in nature as the mineral **halite**. Chlorine compounds are important for the body and are used by muscles and nerves. They are also present in sweat. As it is poisonous in its pure form, chlorine gas was used as a weapon during World War I: soldiers had to wear masks for protection

Uses

The soles of some running shoes *contain chlorine compounds.*

Sodium chloride *is used to season meals.*

Common salt

Breathing in this chlorine-infused liquid *makes people fall unconscious.*

The chlorine levels in this pool *need to be regulated carefully to maintain the quality of water.*

Swimming pool

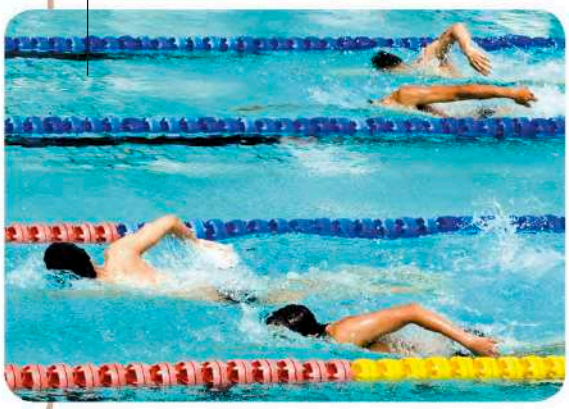

This chlorine-rich plastic *is tough.*

This bleach cleaner *contains a compound called sodium hypochlorite.*

This case *is tough but flexible.*

PVC pipes

Strong water pipes *are composed of thick PVC.*

CHLORINATION

The process of cleaning dirty water with chlorine is called chlorination. This method involves passing the water through a filter.

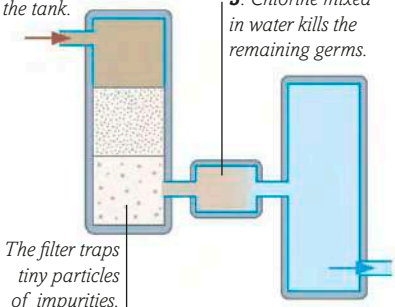

against this weapon. Today, chlorine is used in many ways. Its compounds are present in everything from **running shoes** to **choloroform**. It reacts with hydrogen to make hydrochloric acid, an industrial cleaner. This corrosive liquid eats away at most metals, releasing hydrogen gas. A weaker chlorine acid is used to clean water in **swimming pools**, while **bleach** and other cleaners use chlorine compounds to kill germs. One of the most widely used plastics, **polyvinyl chloride (PVC)**, contains chlorine. It is a tough plastic, used to make many rigid objects.

OCEAN CLEAN UP Chlorine is a common ingredient in cleaning products, and can be used for scrubbing everything from bathroom tiles to ocean floors. These divers are trying to remove harmful seaweed in the Mediterranean Sea using the power of chlorine. This green weed grows quickly and can potentially kill other sea plants by depriving them of their essential nutrients. Some fish are also poisoned if they eat this toxic weed.

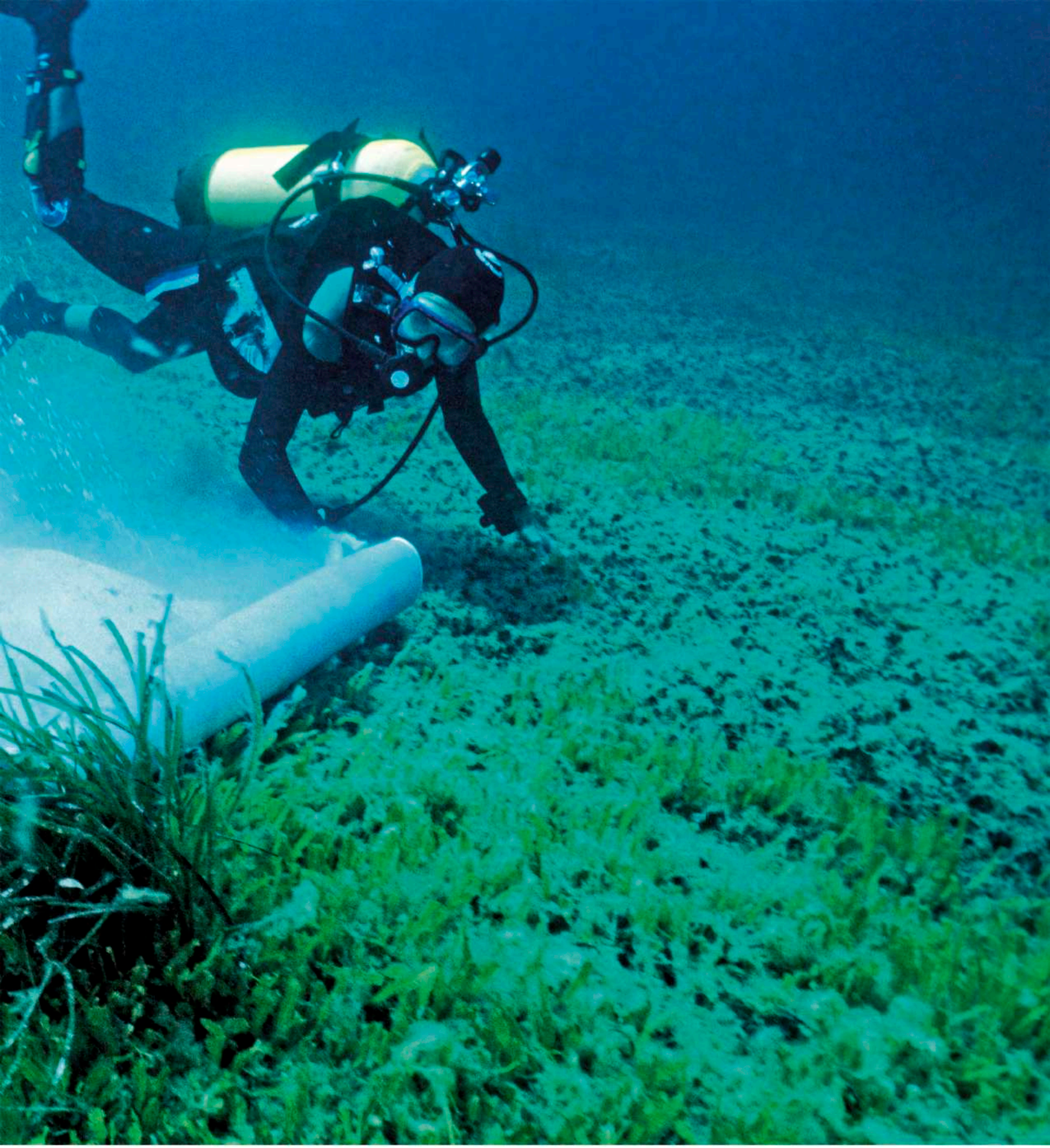

The two divers use chlorine twice in the cleaning process. First, they cover the thick seaweed with a sheet made of PVC, a tough plastic containing chlorine. Next, they pump a compound of sodium and chlorine called sodium hypochlorite under the sheet. This powerful liquid bleach kills the unwanted seaweed. Several weeks later, the divers return to remove the PVC sheets. The invading seaweed will not regrow, and the plants on the seabed will gradually return. Although chlorine is highly reactive and can damage skin and other body parts, divers are well-protected by their rubber wetsuits.

35 Br Bromine

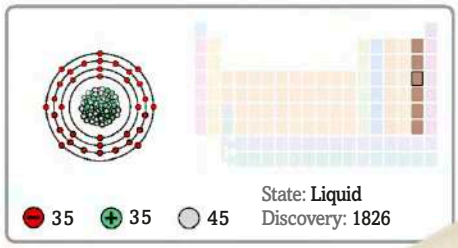

Forms

Bromine is named after the Greek word for **"stench"** because of its strong smell.

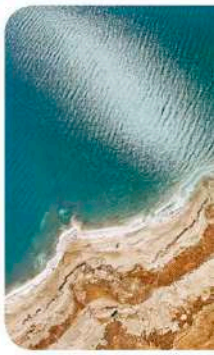

Bromine is the only non-metal that is a liquid at room temperature. A thick vapour given off by this liquid is dangerous if breathed in. **Pure bromine** is never found in nature. Its compounds can easily mix in water, and are found dissolved in seawater and extremely salty lakes, such as the **Dead Sea** in the Middle East. Solid bromine salts, including **potassium bromide**, collect as the water evaporates away, leaving behind crusts of white crystals. Bromine can then be extracted from the solid salts. A common

Uses

This bromine salt *was used in the late 19th century to help patients sleep.*

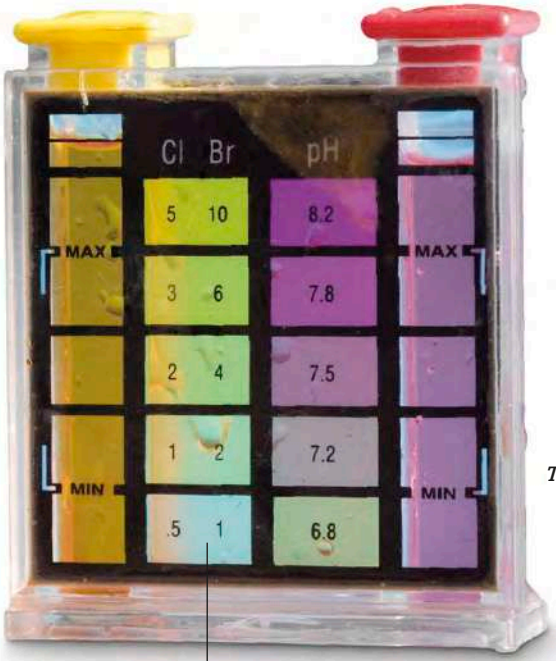

Bromine chemical test kit

The colours on this test kit *show how much bromine is in the water.*

Fire extinguisher

This fire extinguisher *uses a non-flammable, bromine-rich gas to extinguish fires.*

This image *was created when light reacted with silver bromide.*

Photograph negative

Bromine was used as a weapon in World War I.

Crusts of bromine salts *form along the Israeli shoreline.*

The Dead Sea Coast

ANTOINE-JÉRÔME BALARD

In 1826, French chemist Antoine-Jérôme Balard discovered bromine. He heated seawater from salt marshes, and after most of the water had evaporated, he passed chlorine gas through what was left. The remaining liquid turned orange-red: this was bromine.

The fabric *used to make this suit contains bromine compounds that make it fireproof.*

Fireproof suit

use of this element is as a disinfectant to clean water. It works better than chlorine in hot tubs as chlorine escapes into the air easily from the warm water. The concentration of bromine in swimming pools can be regulated using **chemical test kits**. Bromine compounds can be used in film photography in which images are printed using chemicals on **negatives**. Today, bromine is mainly used in **fireproof material**, such as firefighter suits or furnishings, because it doesn't catch fire easily.

53 I Iodine

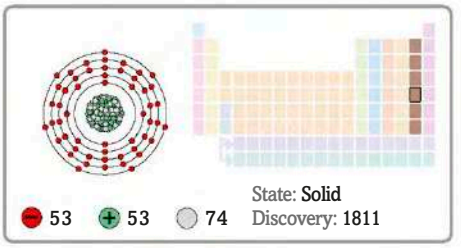

Forms

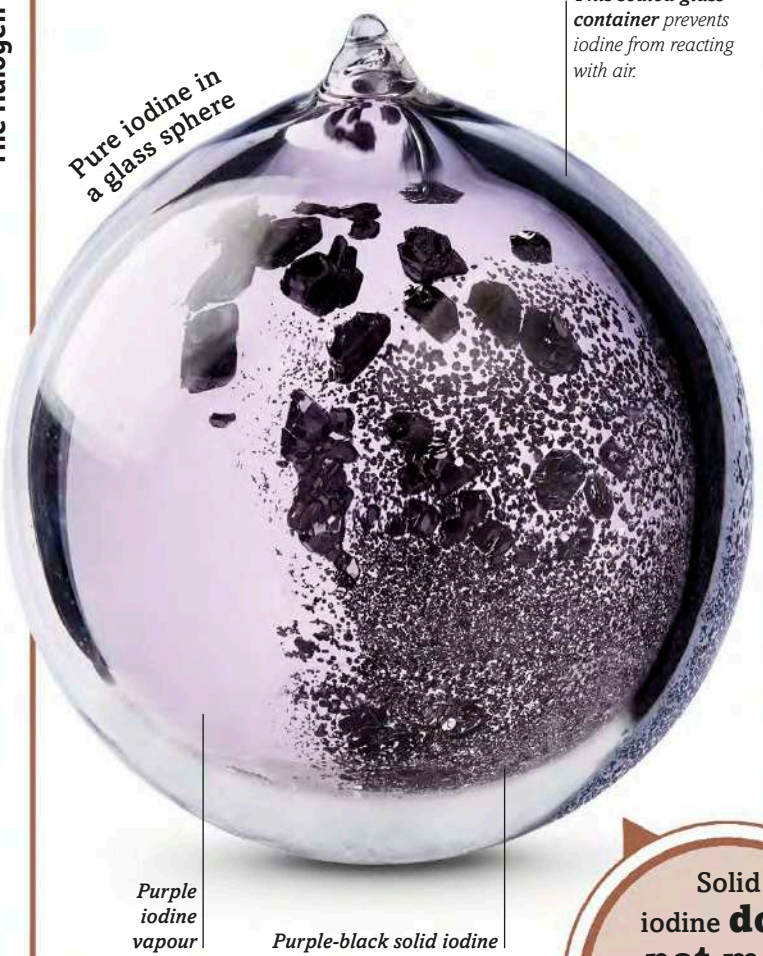

Solid iodine **does not melt** when heated, but turns into a vapour.

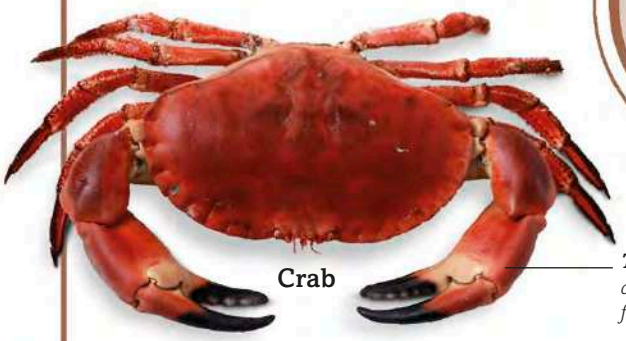

Uses

Iodine is the only halogen that is solid at room temperature. The element forms a purple gas when heated, and is named after the Greek word *iodes*, which means "violet". Iodine was first discovered in seaweed, and many plants and animals in the sea have high levels of iodine.

Seafood, including **crabs** and fish, provide the element in our diet. The human body needs small amounts of iodine to make an important substance called thyroxine, which helps us grow. Iodine is also used to make **printing ink**, red and brown food dyes, and disinfectants.

85
At

Astatine

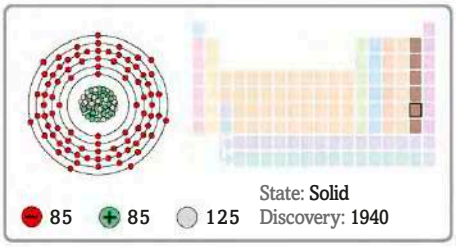

Atoms of astatine are unstable, and typically break down after just a few hours, into atoms of lighter elements, such as bismuth. This radioactive element itself forms in a similar way when atoms of a heavier element called francium break apart. Tiny amounts of this rare element are found in uranium ores, such as **uraninite**. The Italian physicist Emilio Segrè was one of the first scientists to isolate a sample of pure astatine. He was able to do so by using a particle accelerator: this is a machine that smashes together atoms and then studies the results.

***Inside this mineral**, unstable atoms of the element francium are breaking apart, forming astatine atoms.*

117
Ts

Tennessine

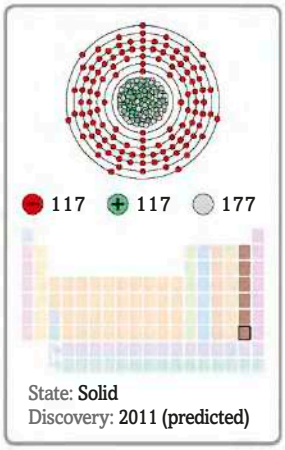

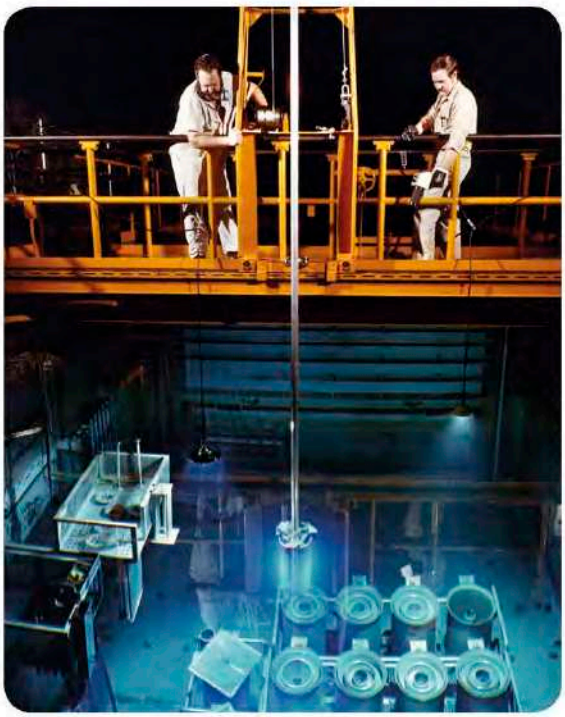

Nuclear Reactor, Oak Ridge National Laboratory, Tennessee, USA

Atoms of tennessine **existed for a few seconds** after they were formed.

Tennessine is the youngest element in the periodic table. It was produced in 2011, in the Russian city of Dubna. The element was named after the US state of Tennessee, home to the **Oak Ridge National Laboratory**, which houses one of the first, large-scale nuclear reactors ever built. Only a few atoms of this halogen element have ever been made. Even so, scientists have predicted it to be a semi-metal, not a non-metal like all the other halogens.

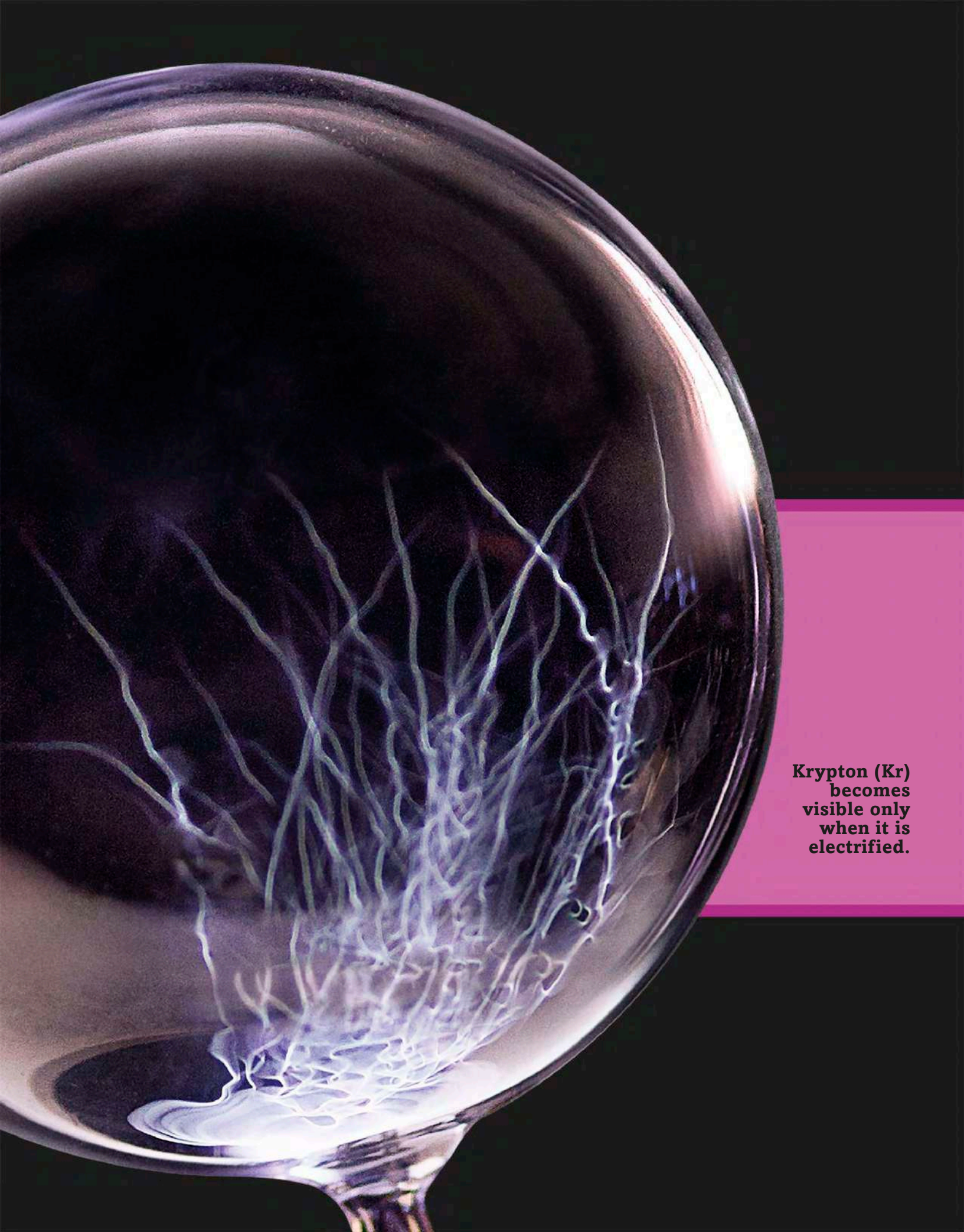

Krypton (Kr) becomes visible only when it is electrified.

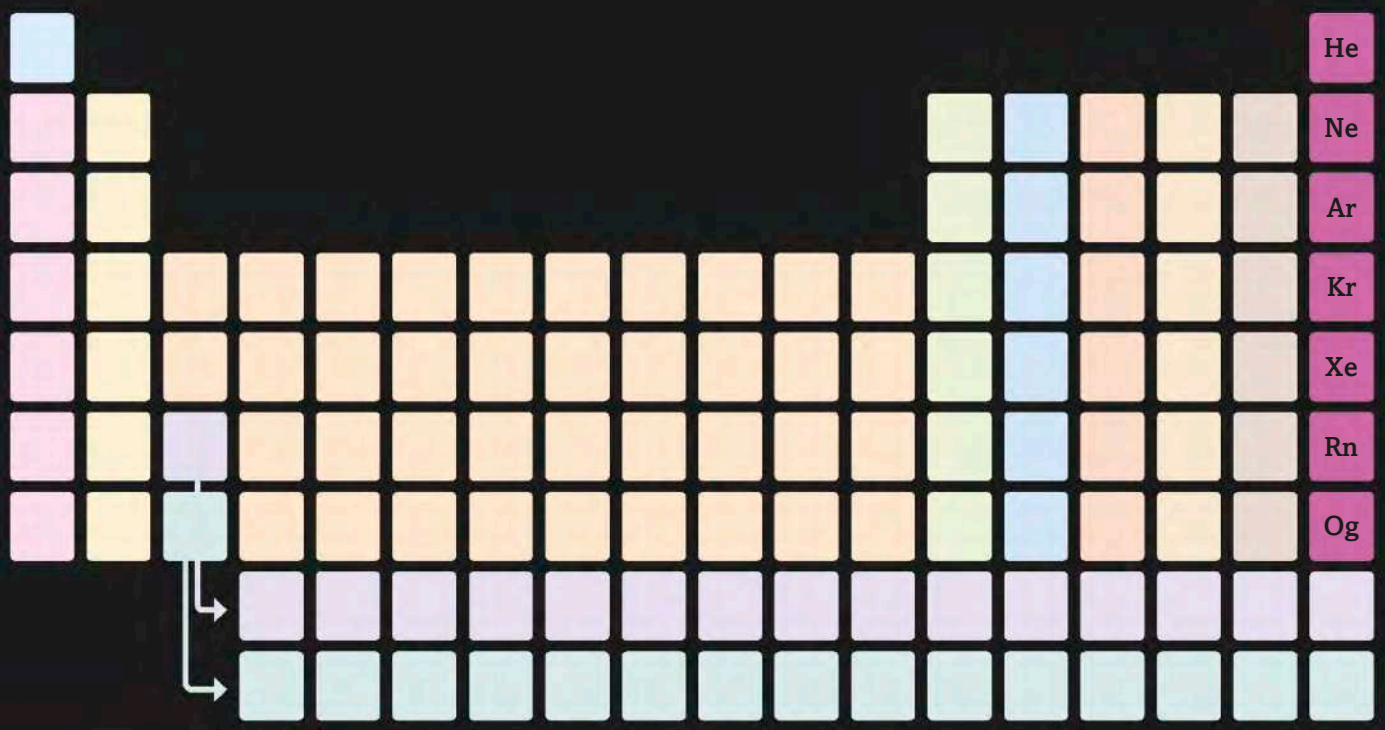

Noble Gases

The group on the far right of the periodic table belongs to the noble gases. These elements are described as "noble" because they do not react with the other "common" elements, such as oxygen (O). Their atoms never form bonds in nature, not even with atoms of their own kind, and so they are always gases at room temperature.

Atomic structure
Apart from a helium (He) atom that has two electrons in its outer shell, all other elements in this group have atoms with eight electrons.

Physical properties
All the members of this group are colourless gases. Going down the group, the density increases – radon (Rn) is 54 times denser than helium.

Chemical properties
Noble gases never react in nature. In the laboratory, heavier noble gases can be forced to form compounds with fluorine (F).

Compounds
These gases form no natural compounds. However, xenon (Xe), krypton (Kr), and argon (Ar) can be made to form compounds.

2 He Helium

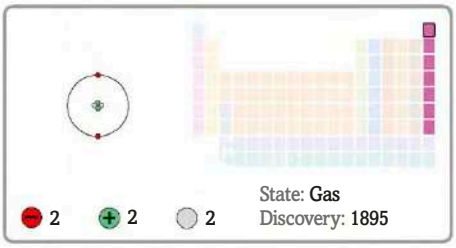

Forms

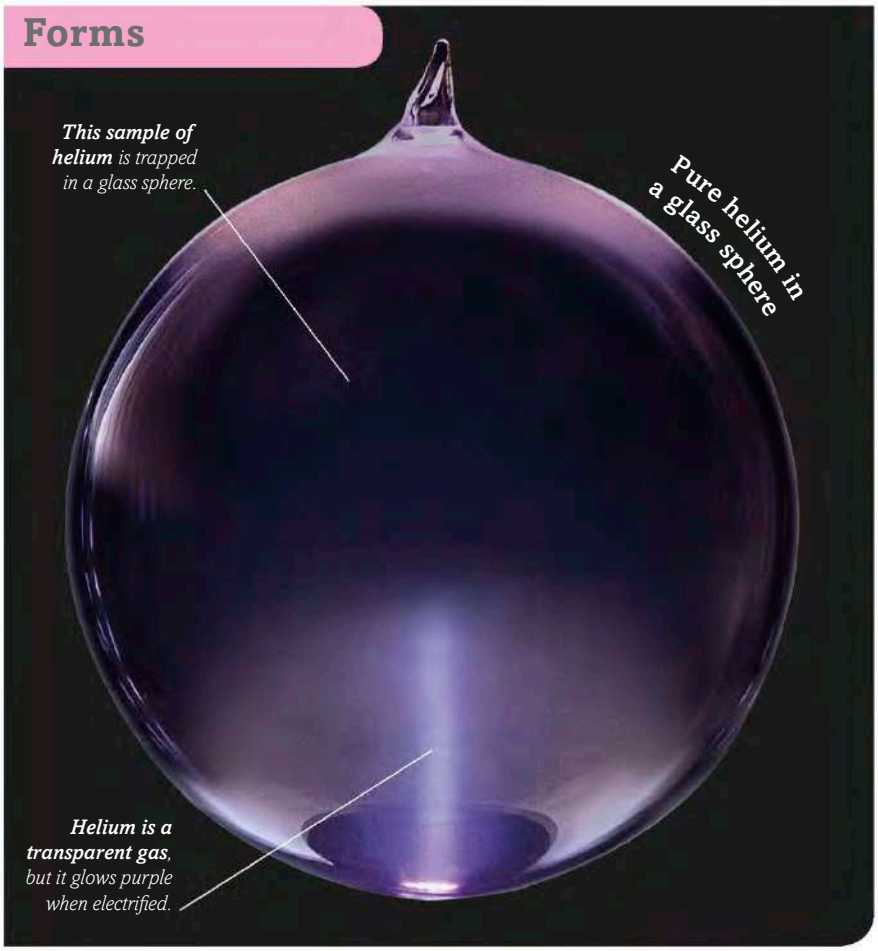

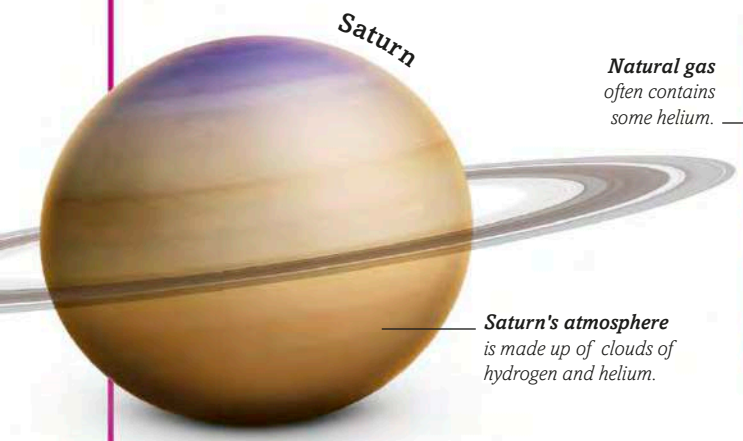

Gas flare rig

Uses

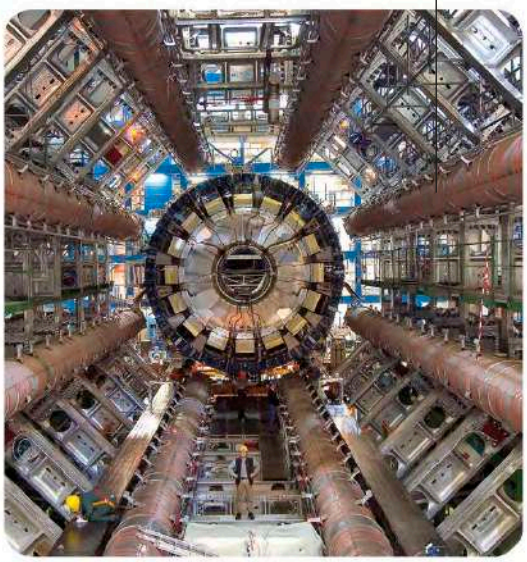

Large Hadron Collider, CERN, Switzerland

Helium is the second lightest element after hydrogen. This transparent gas was first discovered in 1868 by Sir William Ramsay, a Scottish chemist. Today, we know that a quarter of all the atoms in the Universe are helium. It is one of the main gases in the atmospheres of giant gas planets, such as **Saturn**. Being so light, however, helium is very rare on Earth: it escapes from our atmosphere into space. It was not until 1895 that chemists managed to collect a sample of helium gas coming from uraninite, a radioactive uranium mineral. Today, helium is

Airship

This airship *contains large quantities of helium to make itself lighter than air.*

Party balloon

This balloon *contains a mixture of helium and air.*

Helium-ion microscope

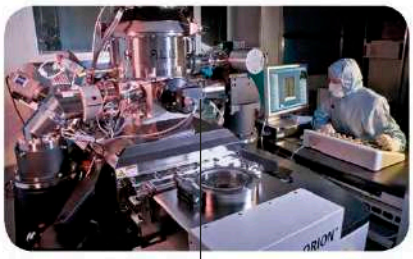

This powerful microscope *can zoom in to view much smaller details than most other microscopes.*

SUN GAS

In 1868, during a total solar eclipse (when the Moon passes directly in front of the Sun), helium was discovered in the cloud of gas seen around the Sun. The yellow colour of this cloud showed it contained an unknown gas, which was named after Helios, the Greek god of the Sun.

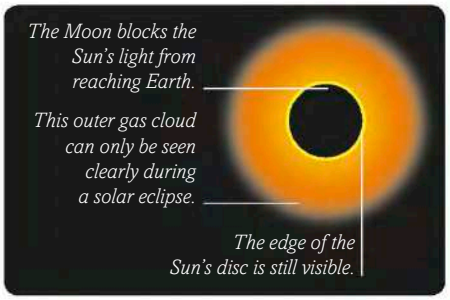

This high-speed train *uses a pair of magnets: one to move forward and one to float over the track.*

The track is lined with a magnet *that repels the one on the train, making it float.*

Helium-cooled maglev train

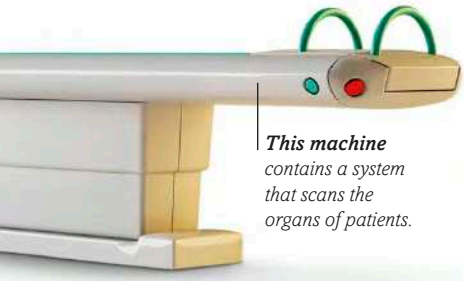

This machine *contains a system that scans the organs of patients.*

Helium has the **lowest melting point** of any element.

Rocket helium tank

Helium in this container *will fill up a rocket's fuel tanks as they empty out during liftoff.*

collected from underground reservoirs or is found mixed in natural **gas** and oil. Unlike hydrogen, which is very reactive, helium is a noble gas and does not react at all. This property makes it safe to use in objects such as **party balloons** and **airships**. To turn helium into a liquid, it must be cooled to an extreme temperature of –269°C (–452°F). Liquid helium is used to make things very cold, including the powerful magnets used to make **maglev trains** float along special tracks. **MRI scanners** also use liquid helium for cooling.

NEBULA This glowing nebula (cloud of gas and dust) is the Crescent Nebula. It is so vast that our entire Solar System would fit inside it seven times over. The nebula's light comes from a super-heated star at its centre. Known as WR 136, this star is 15 times heavier than our Sun and 250,000 times brighter. Its immense power comes from its fuel – helium.

Helium makes WR 136 hot and bright. The star once burned using hydrogen, like our Sun. Hydrogen atoms smashed together in the star's core until they became helium atoms, releasing energy in the process. However, the star ran out of hydrogen about 200,000 years ago. It began smashing together helium atoms instead, and ballooned into a gigantic red star, sending out a cloud of gas that spread around it. The star is producing a wind of electrified gases that hurtles out at 1,700 km (1,056 miles) every second. This wind continues to crash into the gas cloud, making it glow into the nebula we see. Eventually, WR 136 will run out of helium and its other fuels, and explode into an enormous fireball called a supernova.

10 Ne Neon

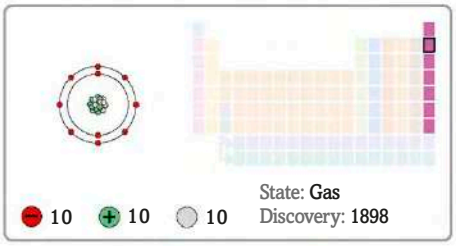

Forms

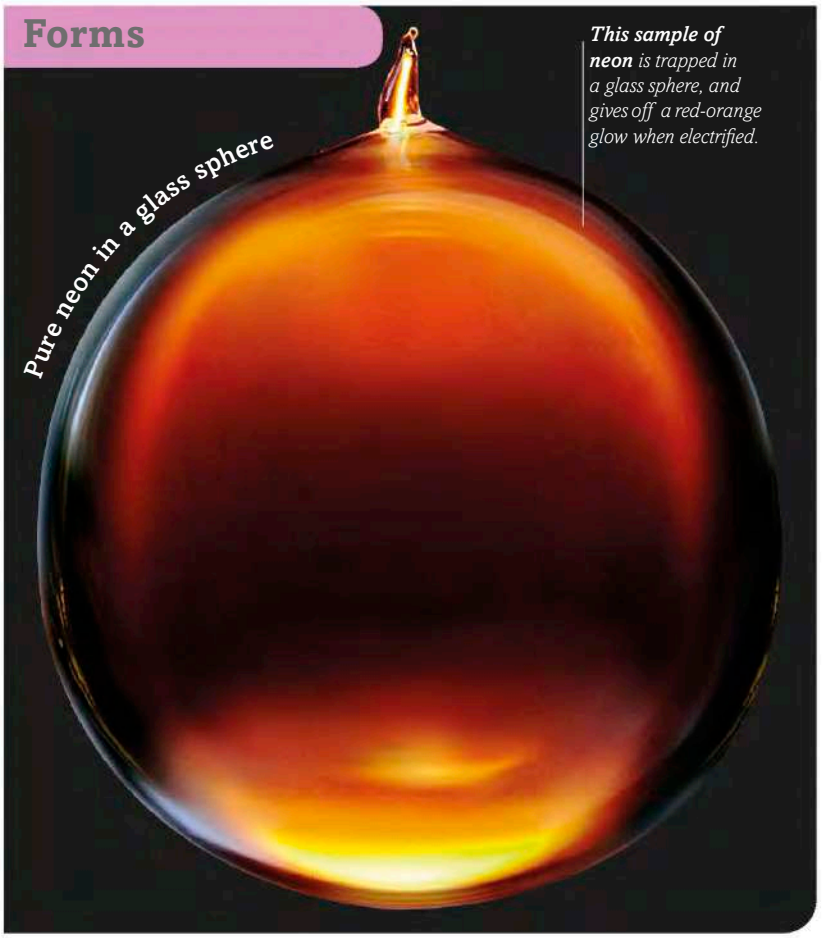

Pure neon in a glass sphere

***This sample of neon** is trapped in a glass sphere, and gives off a red-orange glow when electrified.*

Volcanic eruption

***Volcanic eruptions** release neon gas into the atmosphere.*

Neon lights may refer to lighting produced using any noble gas.

Uses

***A deep red beam** is emitted by this laser.*

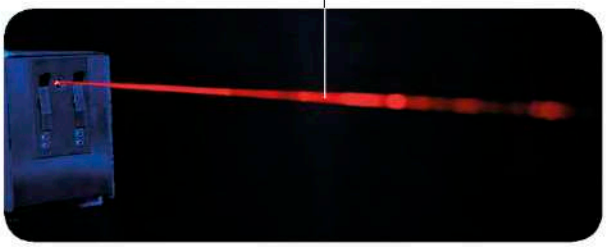

Helium-neon laser

***Glass tubes filled with neon** are used as brightly lit signs.*

Neon sign

Neon is a rare element: it makes up just 0.001 per cent of our atmosphere. Some of it was locked in Earth's rocks when the planet formed, and this is released into the air by **volcanic eruptions**. **Pure neon**, a transparent gas, is extracted by cooling air to a temperature of −189.34°C (−308.81°F), at which point the neon gas in the air turns to liquid. Neon can be mixed with helium to create research **lasers**. However, it is most commonly used in lighting, such as in **illuminated signs** or as bright warning beacons in the path of aircraft at airports.

18 Ar Argon

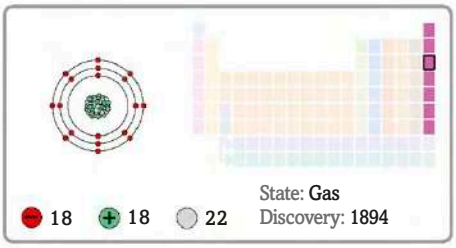

Forms

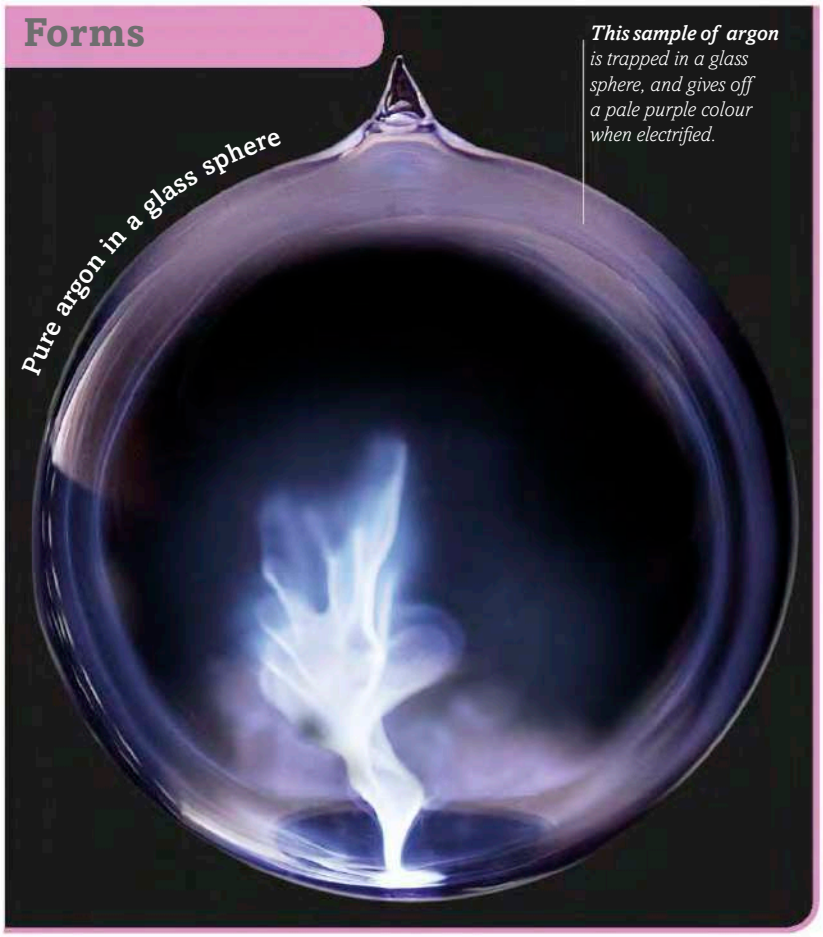

This sample of argon is trapped in a glass sphere, and gives off a pale purple colour when electrified.

Uses

Some diving suits are inflated with argon to retain heat in cold water.

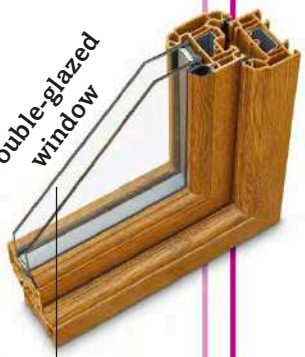

The space between the glass panes is filled with argon to slow the loss of heat.

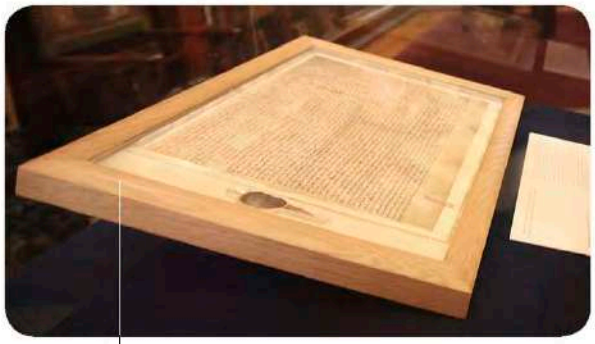

The Magna Carta, a historical document, is stored in argon, which forces out oxygen and water vapour that would damage the parchment.

Argon in this flame prevents metals from reacting with oxygen

Argon is the third most abundant gas in the atmosphere, after nitrogen and oxygen. It undergoes no reactions with any other element, and was named after the Greek word *argos*, meaning "idle". Argon does not conduct heat well so it is put in **double-glazed windows**, and in **diving suits** during cold, deep dives. Its lack of reactivity is useful. Argon is used in **museum displays** to protect delicate exhibits. It also stops metals reacting during hot **welding**. This element can also be useful in the production of titanium.

36 Kr Krypton

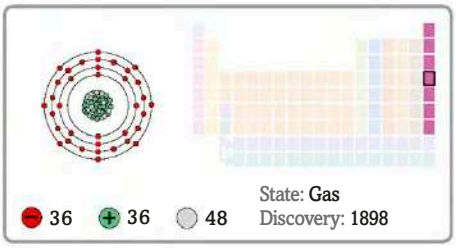

Forms

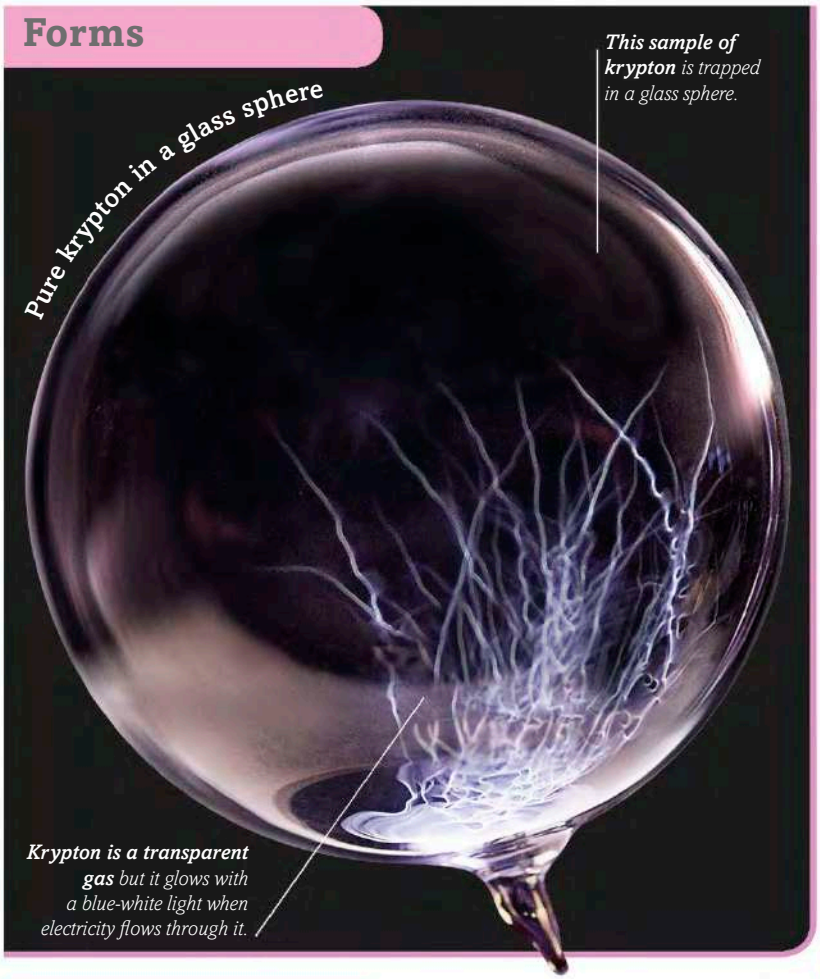

Digital camera

Sir William Ramsey won a **Nobel Prize** in Chemistry for discovering the noble gases.

Laser lighting

Uses

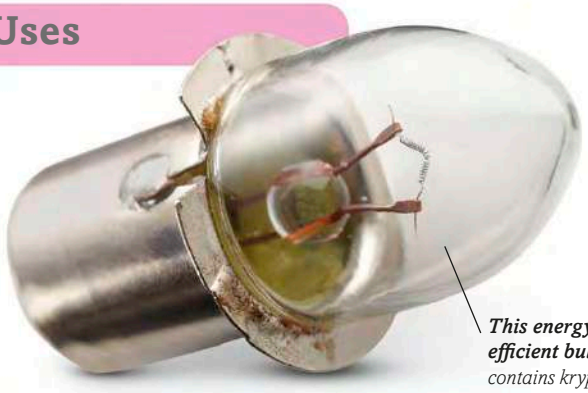

Incandescent bulb

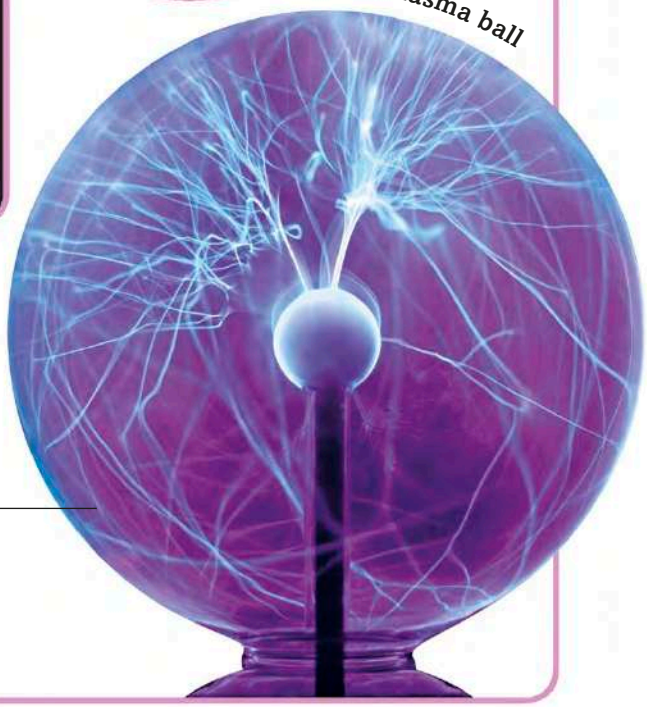

Plasma ball

The word krypton means the "hidden one" in Greek. This element exists as an inert gas in nature, which means that it does not react with almost any other element. Krypton is not found in any minerals and only tiny amounts of it can be found in the air. **Pure krypton** produces a very bright white light when electrified with a current, which makes it ideal for use in **flash bulbs**. Krypton can react with the element fluorine to form the compound krypton fluoride, which is used to power some kinds of **laser**.

54 Xe Xenon

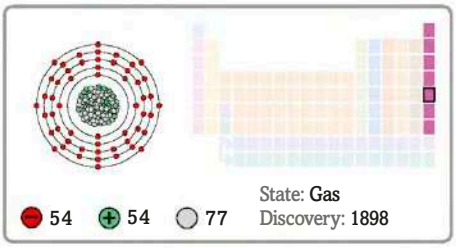

Forms

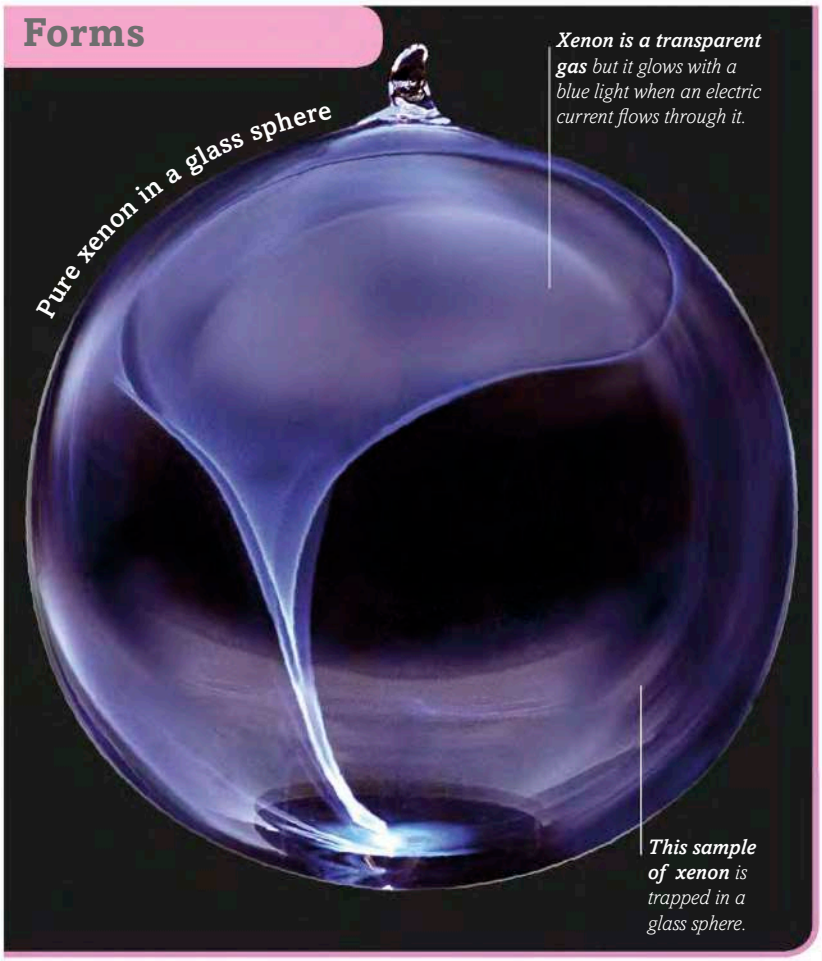

Xenon is a transparent gas *but it glows with a blue light when an electric current flows through it.*

This sample of xenon *is trapped in a glass sphere.*

The xenon produced by this device *makes a patient unconscious before an operation.*

Xenon is used in powerful **lasers** that can **kill bacteria**.

Xenon headlights *glow brighter than the more common halogen headlights.*

Xenon glow *disinfects the air in a room.*

Uses

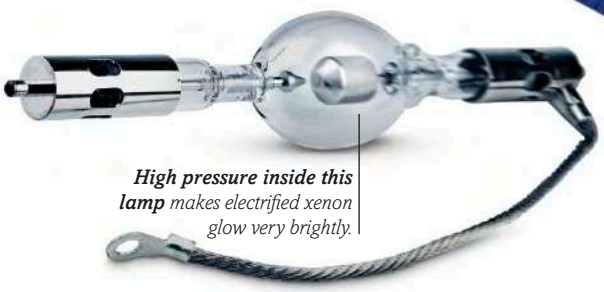

High pressure inside this lamp *makes electrified xenon glow very brightly.*

Modern film projector lamp

Electrified xenon *blasts from the exhaust, pushing the probe forwards.*

Xenon is so rare that there is only one atom of this gaseous element for every 10 million atoms in the air. Like the other noble gases, xenon is colourless and odourless. It glows brightly when electrified, making it useful in very powerful lamps, such as those used in **film projectors** and **car headlights**. The gas is harmless when breathed in and can be used as an **anaesthetic**. When preparing food, **xenon lamps** can purify the air. To propel spacecraft, xenon is used in some rocket engines that produce streams of fast-moving, electrified atoms.

86 Rn Radon

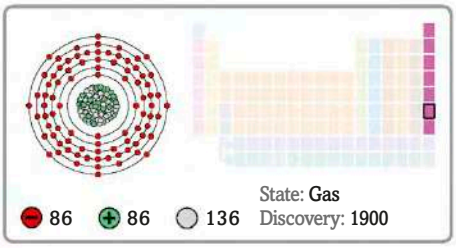

This uranium mineral *releases radon gas, as radioactive metals in it break apart.*

These yellow crystals *belong to another uranium mineral called uranophane.*

It takes only **3.8 days** for half of radon's atoms to split into atoms of other elements.

As it decays*, a compound called thorium dioxide emits radon.*

Radon is the only natural radioactive noble gas. This element is produced by the breakdown of uranium and other radioactive metals. Being a gas, radon escapes from minerals, such as **uraninite**, into the air. Radon is very radioactive and breathing it in can cause illness, such as lung cancer. In most places, the amount of radon in the air is incredibly tiny. However, its levels are high around **volcanic springs** and mud, where it bubbles out with other hot gases. Radon is also present in the water at **geothermal power plants**, which use heat energy from deep, volcanic rocks to make electricity. Radon is also more common in areas rich in granite rock. In these places, people use **test kits** to monitor their homes' radon levels.

118 Og Oganesson

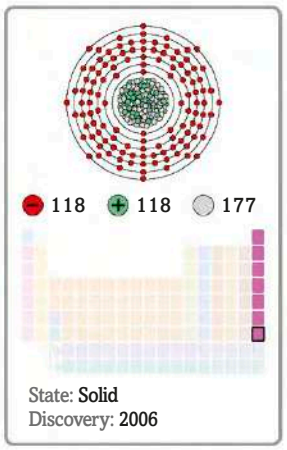

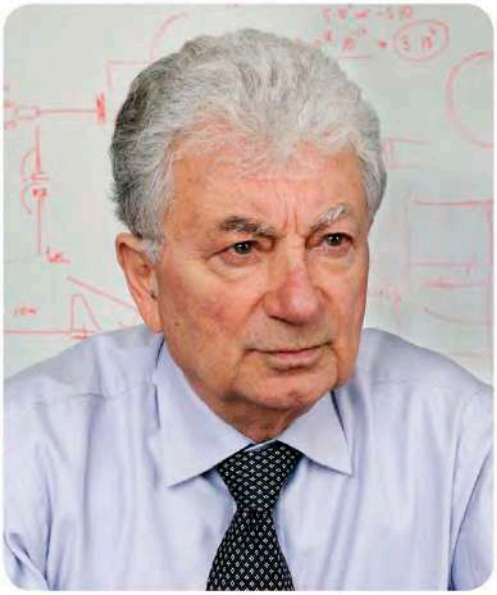

Yuri Oganessian

The heaviest element yet made is oganesson. Scientists think it would be solid at room temperature, but it may really be an unreactive, noble gas. However, only a few atoms have been created so far, so its properties are not well understood. Oganesson was first produced by a team of Russian and American scientists who smashed californium and calcium atoms together at the **Joint Institute for Nuclear Research** in Dubna, Russia. The element was named after **Yuri Oganessian**, the leader of the team.

Joint Institute For Nuclear Research, Russia

The muddy water *from volcanic springs contains radon.*

Hot springs in Lisvori, Greece

These pipes *draw water containing radon from deep under the ground, and this is then used to power the plant.*

Geothermal power plant

This kit *collects radon from the air so the amount of the gas in the area can be measured.*

Radon home test kit

DICTIONARY OF SELF-AFFIRMING POWERFUL QUOTATIONS

By Christian H. Godefroy
with the help of
Didier Penissard

Published by Christian H. Godefroy (2001 Christian H. Godefroy. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, recording or otherwise, without the prior written permission of the author.

Manufactured in the United States of America.

Table of contents

Contents

Introduction 3
A 10
B 36
C 47
D 78
E 103
F 125
G 152
H 160
I 181
J 204
K 209
L 211
M 229
N 254
O 261
P 272
Q 306
R 307
S 331
T 387
U 408
V 412
W 418
Y 437

Table of contents

Introduction

What power do words have?

You surely remember fairy tales from your childhood where the wicked queen cast a spell on the hero. The spell usually consisted of some gesture, accompanied by a "Magic Formula" with powerful effects. Sometimes the spell was a curse or malediction (literally: words which convey a wish that harm come to someone).

During your childhood people around you cast all kinds of spells on you. Very often you were not aware of it, but Magic Formulas - transformed into so-called "killer sentences" - had their effect on your mind, and continued harming you in subtle ways.

Let's look at a few examples. Has anyone ever said something to you like:

- You're really good for nothing.
- You'll never make it.
- Blockhead!!!
- You're just a chip off the old block!
- You're even worse than your mother.
- You're not going to start crying again, are you!?!
- Stop trying to be so special...
- It never fails : you're incapable of being on time.
- What an idiot you are!
- You have absolutely no sense of humor.
- You're always sick.

- You just don't have any talent.
- You should never...

Every time someone used the words "be" or "have" or "always" or "never" when referring to you, chances are they were casting a harmful spell on you.

How harmful spells affect your life

Because these kinds of sentences were stored in your subconscious at an age when you were not able to question their validity, they continued to affect you in subtle ways right into adulthood.

Of course they may be more specific:

- Beware of women.
- Beware of men.
- You'll never understand anything about math.
- Mechanics is a man's job.
- Never trust a Frenchman.
- Making love is dirty.

These kinds of thoughts may not prevent you from getting married, making love or repairing your car, but the causes for your lack of confidence, for the problems and failures you encounter in life, can very often be traced back to these negative seeds.

You are also casting spells on yourself

Do you sometimes talk to yourself? Of course you do, just like everyone else. And do you sometimes blame yourself for the way

you behave? Are you ever negative about yourself? Of course you are.

- What an idiot I am!
- I'll never make it.
- I'm always so clumsy when I'm with other people.
- I'm just not lucky.
- I'm going to get shafted (taken for a ride) again.
- I'm not good enough.
- I'm so much less talented than...

You've taken up where the others have left off, and are casting spells on yourself! Ideally you would identify these kinds of negative reproaches and replace them with stimulating, positive thoughts which liberate you, make life easier, and affirm your worth instead of inhibiting and prohibiting you.

How to weed out these negative thoughts?

Using the Dictionary of Self-Affirming Powerful Quotations is an excellent way to discover where these seeds of failure, anxiety, health or relationship problems, etc. are hidden. It's as if, by consulting the dictionary, you are trying to defend yourself against them. Whenever a positive thought makes you uncomfortable, or bothers you without your knowing exactly why, or just seems stupid, take it as a sign. Continue working with that thought until it reveals its secret meaning to you.

You can also make a list of the difficulties you have to face, the problems you encounter, and then use a corresponding positive thought as an antidote.

Try to remember what was said to you just before you started feeling "down on yourself" or depressed.

How to neutralize the power of negative spells

Many people before you have had to deal with these kinds of negative thoughts. That's why therapists, doctors, and specialists in the art of suggestion have developed antidotes in the form of positive affirmations. The Dictionary of Self-Affirming Powerful Quotations offers you a selection of the best of these.

Better still, they've all been transcribed into the first person (I) so that you can use them to replace the negative spells you are casting on yourself. Here's what to do:

1. Using the alphabetical headings and index, find a positive thought that can serve as an antidote. If you like you can copy the structure of existing affirmations to create your own new ones.

2. Read it carefully, out loud if possible. Then repeat it silently to yourself. Copy it on a pad or in a notebook that you carry around with you. Make it a theme of meditation: while driving, or riding on the train or bus, or lying the beach - any time your mind starts wandering - repeat the positive thought.

3. Use altered states of consciousness resulting from relaxation, meditation or self hypnosis to impregnate your subconscious more deeply. You can tape an hypnosis-inducing text on cassette, followed by your positive thought(s), and then a wake-up message.

4. Discuss your situation with a friend or close relative. Exchange points of view with persons who do not have the same problem as you do, in order to understand how their minds work, and eventually to use them as models for your own behavior. Also try to read personal accounts of how others have successfully dealt with the same type of problem.

5. Allow at least 21 days for your inner transformation to take place. If you are trying to make some very drastic changes, persevere for at least 3 months.

A marvelous tool for personal development

The Dictionary of Self-Affirming Powerful Quotations is a constant companion, always ready to inspire you and guide you along the paths of life. It will help you learn to know yourself better, and exploit the formidable power of your subconscious to help you attain your goals.

You can also use it as an instrument of divination: just ask yourself a question, and then open the book at random. You'll be sure to find an answer, or at least its essential elements.

Copy the thoughts you find most inspiring. The simple act of writing out a thought helps make it your own.

If you're living through a difficult period in your life or feel that you're under a lot of stress, immerse yourself in the Dictionary of Self-Affirming Powerful Quotations and let it wash away all the negative influences and emotions you experience during the course of a day.

And if you happen to be a therapist or teach some form of relaxation or meditation, or if you want to use positive suggestions to help your children, the dictionary will provide you with a constant source of counsel and guidance.

A valuable tool in your interpersonal relations

Does your child have a tummy ache? Does he or she have trouble concentrating? Is your sister in the process of rebuilding after a difficult divorce? Is your partner having business problems?

Open the Dictionary of Self-Affirming Powerful Quotations to the appropriate page. Select one or a number of sentences. A single positive thought can change your life. It can be the branch that saves a drowning person, the talisman, the blessing that breaks the evil spell.

Offer these thoughts to others as you would offer a bouquet of flowers. The Dictionary is like a jewel box, the positive thoughts it contains are like precious gems. Offer them as gifts to your friends and loved ones. And like jewels they will reflect the light of love and dispel the shadows of despair.

As you'll soon discover, this dictionary is an amazing collection of positive thoughts with almost magical powers. It's almost like an Aladdin's lamp, ready to grant all your wishes. All you have to do is keep reading it. In no time at all your life will become simpler and easier, free of all the obstacles which have been holding you back in the past.

Life gives you what you expect of it, on condition that you cultivate positive and creative thoughts. And the Dictionary of Self-Af-

firming Powerful Quotations will help you change your way of thinking. You'll see opportunities where others see only obstacles, solutions where others see only problems, lessons to be learned where others see failure, and beauty where others see nothing at all.

Consider this dictionary as a great prayer of enthusiasm and confidence in life and in humanity, full of gratitude.

Christian H. Godefroy, mailto:webmaster@hypnotic-words.com

ABUNDANCE

This great power, the power to choose, gives me the necessary hope and the necessary state of mind and the necessary courage to get everything I really want out of life.

J. Martin Kohe, Your Greatest Power, published by A Different Life Ltd., 1979

I believe in abundance, I desire abundance, I receive abundance.

Shakti Gawain, Reflections In The Light, New World Library, 1978

I allow this infinite abundance to pass freely through me and through all those who listen to me, just as I freely give up the money I possess, allowing it to flow through my hands, giving where it can do the most good.

Robert Collier, Riches At Your Doorstep, S.I.P. 1962

I am grateful, and I rejoice in a life full of abundance.

Dr. Joseph Murphy, The Miraculous Power To Attract Infinite Wealth, Godefroy Publications, 1979.

I lead a rich and abundant life, physically, spiritually, and on other levels.

Shad Helmstetter, The Power Of Inner Motivation, 1987.

The universe is a horn of plenty which everyone can draw from. My very nature is abundance, and I accept it fully. Abundance is what I really am. And I accept it fully and joyfully. The universe is pure abundance!

Shakti Gawain, Techniques Of Creative Visualization, Editions Soleil, 1978.

Table of contents

ACCOMPLISHMENT (Also see Success)

By using my imagination I move towards accomplishment. I remove the debris caused by the wounds I have suffered, which deprive me of the inner security I deserve.

Dr. Maxwell Malts, Psychocybernetics and Self Accomplishment, published by A Different World Ltd. 1970.

I was conceived to accomplish, built for success, and gifted with the seeds of greatness.

Zig Ziglar, Rendezvous At The Summit, published by A Different World Ltd., 1975

I will do things that will make me live longer... live better.

J.V. Cerney, Stay Younger - Live Longer through the Magic of Mental Self-Conditioning, Parker Publishing Company, Inc. 1968.

ACNE (See Urticaria, Eczema)

ACTION (Also see Opportunity)

From now on I will savor every one of my actions!

Shakti Gawain, Techniques Of Creative Visualization, Editions Soleil, 1978.

I will act immediately.

Og Mandino, The Greatest Secret In The World, published by A Different World, 1979.

I act constructively. I do what my conscience dictates.

W. Clement Stone, Napoleon Hill, Success Through Constructive Thinking, Godefroy Publications, 1978.

I flow with my energy - all the time.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Table of contents

I act in a positive way.

Subli Power - Subli Control, Edit Inter, 1989.

I act.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I cease all actions which work against me. Then the right actions are revealed to me.

Vernon Howard, Psycho- Pictography, S.I.P., 1965

I don't want to remain passive in the face of life any longer, nor allow myself to be carried along by circumstances. I want to act. I want to attain my goals.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I now know that to overcome fear I have to act without hesitation and my doubts will disappear. I now know that action transforms fear into serenity. I will live this day as if it were my last. Action is the bread and the wine that nourishes my success. Action is the spark that gives life to my dreams, my plans, my goals, and transforms them into reality.

Og Mandino, The Greatest Secret In The World, published by A Different World, 1979.

I am capable of action.

Jean Chartier, How To Acquire A Superior Personality, Published by Dangles, 1971.

I am in the process of changing my life through my actions. This is a choice that belongs to me.

Richard Corriere & Joseph Hart, Staying In Psychological Shape, Les Editions de l'Homme, 1979.

I am completely ready.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I simplify my life. I get rid of disorder and actions which are not fruitful. I continue asking myself questions, at least once a week:

Table of contents

"What is it, above and beyond my daily routine, that constitutes the essence of my life, and what are the real activities to which I should be devoting my time?"

Denis Waitley, Seeds Of Greatness, Editions Pocket Books, 1984.

ADVERSITY

I face up to things courageously. I stay calm.

Alan Houel, How To Deal With Difficult People, Godefroy Publications, 1990.

I am happy to encounter a certain form of opposition. It offers me an opportunity to test my strength and my ideas.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

Adversity always contains the seed of a new opportunity. I can do it, if I believe I can!

W. Clement Stone & Napoleon Hill, Success Through Constructive Thinking, Godefroy Publications, 1978.

Once I start playing a game I always do my best to hit the ball back and defeat my opponent.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

AEROPHAGY (Flatulence)

Although my stomach is swollen, I know the swelling will gradually disappear. Little by little my stomach will regain the strength and elasticity it lost, and as it regains this strength and elasticity it will return to its normal size, and carry out the functions of digesting food and sending it to the intestines more and more easily, until my stomach and intestines are in perfect form.

Emile CouÈ, Complete Works, Astra Publications, 1976.

Table of contents

ADVICE

God will tell me what I should do.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

My subconscious becomes a more and more important part of myself... I can ask it questions... it answers... it advises me and guides me.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

AFFECTION

I am affectionate.

Dr. Wayne W. Dyer, Your Erogenous Zones, Tchou, 1976.

The positive emotions of love, affection and patience appear when I am in touch with my true self.

Vernon Howard, Psycho-Pictography, S.I.P., 1965

AFFIRMATION

I have the right to be myself. Affirmation is a measure of my success. I affirm myself more and more, in all areas of my existence.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

AGING

I avoid being inactive and doing nothing... I make sure not to let my movements slow down too much... I cultivate vivacity in order to keep myself alert.

Table of contents

I am sure I will stay forever young... my mind will remain lucid... and I will not suffer the pains of aging until very late in life... my life is rich and productive.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I adapt to the ups and downs of life by maintaining my sense of humor. I keep the child alive and awake within me.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I may be aging, but I am also improving with age.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

AGITATION

At the least sign of becoming over-excited I react and remain impassive.

Paul-Clement Jagot, How To Develop Your Personal Magnetism, published by Dangles, 1982.

AGREEMENT

I try to reach explicit agreements in the form of contracts.

Pierre Nicolas, Time Is Money - and Pleasure, Inter Editions, 1981.

AGGRESSION

The spirit in me is my shield. The spirit in me is my saviour. I have nothing to fear - the power of the spirit protects me.

Marcelle Auclair, The Book Of Happiness, Editions de Seuil, 1959.

The spirit in me is my refuge and my fortress.

Marcelle Auclair, The Book Of Happiness, Editions de Seuil, 1959.

Table of contents

AIDS

My immune system is back to full strength again.

Louis Proto, Self Healing, Piatkus Publications, 1990.

ALCOHOLISM (Also see Sobriety)

Drinking is no longer important to me.
Drinking is no longer a necessity for me.
I forgive myself for past mistakes which led me to drink.
I am becoming a more relaxed person.
My self confidence is growing day by day.
I like people, and I feel they like me.
I think of myself as a positive person.
I like to think of myself as a person who doesn't drink.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec/Amerique Publications, 1874.

I like drinking water.

Subli Power, Sublihealth, Edi Inter, 1989.

I free myself of the habit of alcohol. I love myself, and I love life more every day. I am proud of myself.

Karl O. Stoeber, Psycho-training, Godefroy Publications, 1984.

I don't have to have a drink whenever I am upset or frustrated. I can survive life's frustrations without alcohol. And because I have proven to myself over and over again that I have never been able to handle alcohol intelligently, I am going to abstain altogether.

Frank S. Caprio, M.D. and Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc. 1974.

I have no trouble controlling my drinking habit. I now drink alcohol in moderation. I can easily go several days without having a drink. The less I drink the healthier I get. I am mastering my habits,

and my general health is improving day by day. I like my body, and I want to keep it healthy for as long as possible. So I take care of it. I drink very moderately. There are many days I don't drink at all, and I'm perfectly happy.

I am giving up alcohol completely. I have no trouble not drinking at all, and I feel a lot better. My body and mind are completely free of the alcohol habit. I have more and more control over my emotions. I can stop drinking completely, starting right now. From now on I'll be sober and happy.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I couldn't care less about drinking, any place, any time, under any circumstances. Others can drink, but I'm really not interested.

Clement Auger, Transform Your Life Through Auto-suggestion, Heritage Press, 1979.

I feel as if all my problems were fading away.. I've solved them all... nothing can change my mind... I feel absolutely no desire for alcohol any more... drinking doesn't interest me... it turns me off... more and more... even the thought of alcohol turns me off... just looking at alcohol makes it even more repulsive.

From now on I won't drink any alcohol... nothing can make me change my mind... I'm happy to be improving my health, and I feel better every day... I'll never drink alcohol again. I find it more and more repulsive. Even the though of alcohol disgusts me.

If I drink a single glass, I'll get sick. My stomach can't take any alcohol... even a glass will make me very sick... I will try not to succumb to temptation.

I'll never drink alcohol again.

Just the sight of alcohol makes my disgust grow... I'm so happy to be free of it... nothing can make me change my mind. I can't drink any alcohol... and I won't!

Professor Kurt Tepperwein, Secret Techniques of Hypnosis, Godefroy Publications, 1981.

Table of contents

I am capable of resolving my problems, and thus of freeing myself from the burden of alcoholism.

Professor Kurt Tepperwein, Think and Heal, Godefroy Publications, 1984.

I'm going to stop drinking not because others want me to, but because I really want to.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I'm a nice person when I don't drink. I look better when I don't drink. I work better when I don't drink. I lose weight when I don't drink. I'm healthier when I don't drink.

Editorial staff of Practical Health Guides: How To Activate Your Body's Natural Defences, Godefroy Publications, 1989.

ALLERGIES

I have more and more fun with cats. Cats are gentle, affectionate animals. My body reacts completely normally when I'm with them.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I'm a healthy normal person. My body functions normally, under any conditions. I function perfectly normally in the presence of animals. My body functions normally, whatever the surrounding odors are. I am free of past negative influences which have been causing my allergic reactions. Soon I'll be able to eat any food I like. Each day that goes by strengthens my conviction that I am in the process of resolving all past problems. I perceive of myself as a person who is not allergic to anything.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

ALLERGIES (Food)

I can now eat —— (name the food) easily and find it delicious. I

Table of contents

am free of past negative influences. My body now reacts normally to all the foods I eat.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

ALTRUISM

To me loving means listening unconditionally and without prejudices, to the needs of another person, and fully appreciating his or her worth.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I have the right to be open with other people.

Christian. H. Godefroy & John Clark, How To Have More Time, Godefroy Publications, 1989.

I stop thinking only about myself, and think more about others. Every day I try to do something kind, something that brings a smile to a suffering face.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I always take other people's feelings into consideration. I am aware of their feelings, sensitive to their attitudes and opinions. I know how to listen. I show a sincere interest in the convictions, thoughts and ideas of others. And since my interest is sincere, other people recognize that and start counting on me.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

AMBITION
(Also see Vitality, Dynamism, Enthusiasm)

I maintain a victorious attitude towards life. I know there is no task a person cannot accomplish once he or she has made up his mind. I want to accomplish... (name the task). My ambition is completely

Table of contents

legitimate. I want to make progress in my work, as well as in my intellectual and moral development. I will leave nothing to chance. Attaining the success I aspire to depends solely on me.

Professor Robert Tocquet, The powers Of The Will, Godefroy Publications, 1989.¥

I reconstruct myself with a brighter future.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am not content with being less than I can become. I'm as capable as any other human being who has ever lived. I am not afraid of my own greatness.

I always remember that I can become anything I choose to be, and that greatness is there for the taking if I really want it.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

I am no longer satisfied with what I can get, but only with what I want to get.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I know I can obtain everything... gain access to the highest circles... I only have to want it... and have faith in my success, and I will succeed.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I am aware, conscious, and ready! I am overflowing with vitality and ambition. I exploit every opportunity with dynamism and enthusiasm.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am a CEO (or a lawyer, or any profession of your choice). Every day brings me closer to my goal. I know I am already there. It's fantastic. I accept my own amazing future.

Charles Baudouin, Psychology And Practice of Auto Suggestion, Idegraph Publications, 1990.

I am full of ambition, and there is nothing wrong with that. I want

to succeed in my work, as well as in my personal development.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

ANAEMIA

My blood is getting richer and richer, more and more red, more and more nourishing. It is getting to be more and more like healthy blood. My anaemia is gradually disappearing, carrying away with it the host of problems that accompany being anaemic.

Emile Coue, Complete Works, Astra Publications, 1976.

My blood is rich, and is replenished daily.

Louis Proto, Self Healing, Piatkus Publications, 1990.

ANAESTHESIA (Also see Pain, Wounds)

With each pinch the anesthesia spreads. At first I pinch lightly, adding pressure each time. The fourth time I pinch, the anesthesia will be complete.

Leslie M. Le Cron, Liberate The Forces In You Through Self Hypnosis, Tchou/Ariston, 1964.

All right, now I'm going to count to three, and on the count of three my arm will become rigid, and the more I try to bend it, the more rigid it will get. During this time my right hand will grow more and more numb, more and more numb... more and more insensitive. Now even my left hand is becoming rigid, so the anesthesia in my right hand is getting stronger.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

ANAESTHESIA - DENTAL

My cheek is numb, I can prick it with a needle, and not feel a thing... My lower gum is numb, I can prick it with a needle, and not feel a thing... My whole mouth including gums and teeth are totally numb, like a block of wood.

Giovanni Sciuto, Real Hypnosis, Dangles Press, 1975.

ANGER (Also see Migraines)

What other people think or say about me doesn't bother me. If my inner Guide is with me, if He (or She) approves of my thoughts and actions, then everything is all right. I am calm and strong!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

It's normal to get angry from time to time.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I take responsibility for my emotions. When I begin to get angry I acknowledge the fact that I own my emotions. I remove myself from any potentially hostile situations. When I talk about my anger or dissatisfaction I say: "I feel angry when I see that happening," instead of saying, "You make me angry when you do that." Only I can make myself angry.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I master my emotions. I never get angry for no reason, nor do I react violently to any situation whatsoever.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

I am very careful not to get angry. I have a good character, I demonstrate patience, and I always think before I act. I avoid complaining.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

When I get angry about not achieving the results I want in a situation, I move quickly from "stewing to doing.' I ask, "What can I do to turn this situation around? Is there anything I can salvage from this situation? What can I learn from this situation? How can I prevent it from happening again?"

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

When I get angry I remind myself that everyone has the right to be what s/he chooses to be, and demanding that others be different will only result in prolonging my anger. I make sure to let other people make their own choices. After all, don't I claim that right for myself?

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

When I feel myself getting angry I quickly give myself the suggestion that I can switch my thoughts to something else. If need be I can always walk away from the person who is provoking the anger reaction, or expend my energy in doing something that will divert my mind - finding some other outlet for my pent-up emotions. I am going to remind myself that people can make me angry only if I allow them to do so, and that getting angry is actually flattering the other person.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

ANGINA (Chest)

The attack is passing and my heart is beating normally.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

Table of contents

ANSWERS

My Real Self knows all the answers I need.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

ANXIETY

From now on, whenever I feel anxious I will breathe deeply, relax, and feel a surge of energy rising through me.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I express the anxieties I have about myself, and then let go of them.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am completely calm... completely calm... A marvelous feeling of peace and harmony spreads through my body... I am happy... my circulation is regular and my digestion is perfect.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

I feel a hot current flowing through my solar plexus.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

In the past I would look anxiously outside, waiting nervously for something bad to happen. From now on I look inwards, full of confidence. I look to my Inner Guide.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I am learning to interpret my anxiety as a state of excitement instead of nervousness.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

Table of contents

I stop anxiously fighting all the little complications of life. I simplify my affairs.

I have to work towards freeing myself of all forms of fear and anxiety.

I feel less tense and anxious.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am in the process of becoming a much more self confident person.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I see my life and everything that happens from the calm, clear perspective of my conscience, and profound understanding.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

My tension can be an ally because if I feel any anxiety, I take it as a sign to call on the means at my disposal to deal with the situation.

I take a deep breath and relax. Everything's all right. I am in control of the situation, and of myself.

It's all over, I've succeeded.

It wasn't as bad as I thought it would be.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

Everything is fine.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

ANXIETY - SOCIAL

It does no good to worry. It's much better thinking about what I can do rather than becoming anxious. I have plenty of time to take a deep breath and relax before dealing with the situation.

I don't have to prove anything to anyone. If others accept me, all the better. If they don't, too bad. I'm not trying to make everyone

love me.

I concentrate on the present moment. What do I have to do now?

I think mostly about the things I have to do, and about the positive results these actions will bring.

If I deal with the situation step by step, I won't have any trouble.

I can't expect people to always think as I do, and do what I want them to.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

APLOMB

In all situations where I used to be too impressionable, I am now protected by a suit of armor through which nothing can penetrate.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

I develop my aplomb and my confidence.

SubliPower, Subli Confidence, Edi Inter, 1989.

I feel an imperturbable calm in the face of anything that happens.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I don't allow anyone to influence me.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

APPEARANCE (PHYSICAL)

I know that my appearance, my weight and my well being depend solely on me. I do everything necessary, every day and every instant, to create the "me" I really want to be.

I am in control of myself. Only I decide what my physical appearance is like, how much I weigh, and how I feel.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

APPETITE (Also see Digestion, Stool)

I'm hungry at mealtimes... I am attracted by natural, healthy foods.

Sublipower, Subli Health, Edi Inter, 1989.

APPREHENSION

I am calm, courageous and free! My inner Guide fills me with confidence, self assurance, love and harmony. I feel free, joyous and strong. I am free, and secure under Your wing. I am free!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

Negative suggestions, phobias, apprehensions, dark or sad thoughts all gradually disappear of their own accord, without my having to fight them. Even dark thoughts and feelings which are based on reality are gradually receding.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

APPROVAL

I feel good because I want to, and my feeling good has nothing to do with you.

No one is unhappier than people who need (the approval of) other people. But those who want to love others become happy themselves.

I thank people who tell me things that can help me, even if what they say is difficult to accept. My feeling of gratitude replaces my need for approval.

I stop looking to my spouse, or to anyone else, for approval of what I say, by cutting out things like, "Isn't that true, dear?" or... "Don't you agree?" or... "Ask so and so, she'll tell you."

Table of contents

I reprimand myself out loud every time I find myself seeking other people's approval. By doing so I become aware of this kind of behavior, so that I can stop it.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

It is completely normal that people forget to express or even feel gratitude; therefore, if I expect other people to reward me through a show of gratitude, I am only exposing myself to a flood of bitter disappointments.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

I'm sure the sales manager will notice my effort and my results.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

ARCHERY

I am perfectly relaxed when I practice archery.

My reflexes are getting faster and more efficient. I can hit the target more and more easily, shot after shot. I am improving all aspects of my archery. My vision and reflexes are working marvelously well. Archery is easy for me.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

ARTHRITIS

I relax the muscles around my joints. I feel a cascade of clear water flowing over my joints, washing them, purifying them, making them strong. The water carries away the pain. I see the cool clear water flowing over my fingers, my knees, my shoulders. It calms me, washing away the inflammation. I can already feel the water sooth-

ing my joints, relaxing them, so that I can move with no pain at all.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

My joints are supple again, and feel smooth and comfortable.

Louis Proto, Self Healing, Piatkus Publications, 1990.

My energy circulates more and more freely through my whole body, which is becoming light and luminous. My pains completely disappear because the amazing power of my subconscious mind is healing me completely.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

ARRHYTHMIA

My heart beats slowly and calmly.
I am my heart.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

ASKING

I have the courage to ASK.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

ASSIDUITY (Also see Perseverance)

I'm interested in material things, and I do my work conscientiously.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

Table of contents

Work that you're interested in is never difficult, and sure to succeed.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

ASSIMILATION

I'm getting stronger day by day. I assimilate information more and more rapidly and precisely, and my willpower is getting stronger.

Paul Clement Jagot, The Power Of The will, Dangles Press, 1950.

ASSISTANCE

The universe is always there to help me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

ASSISTANCE - SPIRITUAL

I send you strength and courage. United with the forces of your inner Guide, I consciously direct our plenitude towards you, so that we can offer you assistance and stimulate the right conditions which will help your development.

I love you and send you energy, faith and courage. May a clear vision awaken in you of the right path to follow. Everything good you wish to accomplish will succeed.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

Right now divine love is doing its perfect work for the benefit of all.

From now on may divine love and light work through me.

Table of contents

Divine love goes before me and prepares the way.
From now on God will show me the way.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

ASSOCIATES

My time and energy are precious. I choose to surround myself with positive associates who are focused on possibilities, learning, improving and taking responsibility for results. I guard my mental life against negative influences.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

ASSURANCE (Also see Self Confidence)

Every day I feel more and more sure of myself, and it's great. A beautiful, exciting life is awaiting me; it's wonderful to feel this sense of assurance growing in me, this confidence in myself that I feel getting stronger and stronger.

Charles Baudouin, Psychology And Practice of Auto Suggestion, Idegraph Publications, 1990.

I'm ready for action, I look more assured, more vibrant, more convincing. I speak with conviction and tread firmly.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

I feel strong and calm no matter what happens.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I always feel calm and self assured.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I can be whoever I want to be with total support. I can go fast or slow - it's my choice. I am protected.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

Table of contents

I am calm and confident. I possess the tranquil assurance that my life will be a success.

Shad Helmsletter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am happy, self assured and strong.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I am confident.

Matthew McKay & Patrick Fanning, Self Esteem, New Harbinger Publications, 1990.

ASTHMA

Every day I feel more and more free and relaxed. I react normally to all situations. I fill my lungs fully when I breathe. And I can easily overcome any tension or anxiety I feel. I am in perfect control of my body in all situations.

My bronchial tubes and lungs are strong and relaxed. I breathe calmly and deeply in all situations. I always stay calm, and control my body's reactions. I feel more and more confident and relaxed all the time.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I send energy from my brain into my nostrils, which will remain open, sending air to my longs, for as long as I want.

My lungs are relaxing, air is entering, the attack is subsiding.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

I am completely relaxed, breathing deeply and freely.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am convinced that my asthma problem is getting better every day.

I free myself of the causes of my asthma attacks.

Table of contents

I am convinced that I don't have to have these attacks.

I'm becoming a relaxed person.

I feel stronger and healthier every day.

I see myself as a healthy, normal person.

I always breathe easily and normally.

I don't have to use an asthma attack to get other people's attention.

I can now share my feelings with others more easily.

I have strong willpower, and I can easily control my physical reactions.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

My breathing is relaxed and effortless.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

My breathing is slow and regular; with each breath the muscles in my bronchial tubes and lungs relax more and more deeply; air enters my rib cage more and more freely, and my breathing becomes easier and easier.

After each session I see myself gradually improving. I breathe much more easily, the wheezing in my chest is disappearing, I am becoming much calmer, much more relaxed, and much less nervous.

As soon as I place my left hand on my chest, the muscles in my lungs and bronchial tubes relax; I feel them relaxing, I feel the air flowing more and more freely into my lungs, my breathing becomes easier and easier; with each breath my chest muscles relax more and more, and my breathing becomes calm, regular, calm, regular and easy.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

My breathing is completely normal, deep, regular and calm.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

Table of contents

ASTHMA ATTACK

It's stopping, I'm already breathing easier.

I can feel myself breathing more easily. One more breath and it'll be over.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

AUDACITY (Also see Courage)

I will go where failures fear to tread.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I dare to do what I want. I act! I'm no longer afraid, and I will succeed.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I dare take the first step towards my rightful castle.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I can't always be reasonable and please everybody if I want to feel good about myself; I have to take risks, risks which develop a feeling of direction in me.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

I Can, I Want, I Dare, I Act.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I want to dare.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

To succeed I risk failure, and that is the proof of audacity.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Table of contents

If there's nothing to lose and a lot to gain, I'll go for it.

W. Clement Stone, Passport To Success, Godefroy Publications, 1962.

AUTOSUGGESTION (Also see Hypnosis)

I am going to follow the suggestions which I implant in my subconscious mind.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

AVAILABILITY

I am always available for people whom I consider important, and for things which I consider important.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

AWARENESS (See Self Awareness)

Table of contents

BAD CHARACTER

I am acquiring a pleasant and joyous character.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I believe that my bad temper can be controlled, and I am not going to rationalize by thinking that I cannot help myself.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

BAD HABITS (See Habits - Negative)

BACKACHE

All the muscles in my back are relaxing. The tension in my dorsal region is dissolving completely. My whole back feels better and as this happens my confidence grows. My spine is relaxing. My back is getting stronger every day, and all tension is disappearing.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

BALANCE (Also see Peace, Strength, Wisdom, Health, Harmony)

Table of contents

In the past I used to suffer from fears that almost made me physically sick, but I am gaining strength as I become more and more balanced, and the strong and healthy person I know myself to be.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, 1978.

I eat and sleep well.

I take care of my physical balance, since it plays such an important role in my ability to work well, and remain emotionally stable.

I know that the best thing is to maintain a constant and balanced relationship with my body, instead of getting preoccupied with it from time to time.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am well structured and balanced.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I am a balanced person.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am perfectly balanced.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

A sign of the presence of God in me is my feeling of peace and balance.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

My body is balanced, in perfect harmony with the universe. I am always deeply relaxed and balanced.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

BALDNESS

Now that I am no longer using my energy to work, I will channel

Table of contents

it to my scalp; the energy flows more normally, replenishing my hair as it did in the past so that it grows quickly.

Like crabs who grow new legs I can regrow my hair.

I want my hair to look like... (add description).

My hair is growing... growing.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

BEAUTY (Also see Health)

Every day my beauty shows a little more clearly. Every day I move closer to this promise of beauty. I feel better and better.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

I give thanks for ever increasing health, beauty and vitality.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

My body is pure and beautiful. I love nature and count my blessings. I see beauty in others, and experience it in myself.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

BEHAVIOR

The totality of my actions and behavior determines how others decide they should treat me.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

BEING

I am a strong, affectionate and creative being.

Table of contents

I am a radiant expression of God. From now on my body and mind will express divine perfection.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

BELIEF

I allow myself to hold something sacred in life.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

BELONGING

I move forward with confidence and chances are I will experience a sense of belonging.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

BENEVOLENCE (Also see Sincerity)

I look at myself with benevolence and I can thus fend off my solitude indefinitely.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am more and more benevolent towards myself and towards others.

Sublipower, Subliconfidence, Edi Inter, 1989.

I am always serene and benevolent.

Alan Houel, How To Deal With Difficult People, Godefroy Publications, 1990.

Everyone wants the best for me.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Editions, 1982.

Table of contents

[I treat everyone with kindness. I treat everyone as a precious and passing gift.]

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

BEST

I accept and deserve the best, which is flowing to me now that I have become an open channel.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

BIRTHING

I prepare for giving birth every day, and make sure to practice the techniques I'm learning in my prenatal classes to make the birthing easier. I am a woman who is in full control of myself and my reactions.

Marcel Rouet, Mastering Your Subconscious, published by Dangles, 1982.

I feel very good, and I look forward to the birth impatiently... all this is engraved in my subconscious... As soon as I look at my right hand, I experience a state of deep relaxing sleepiness... pain does not exist... All I have to do is look at the palm of my hand and all the pain disappears... I feel completely fine.

Dr. Kurt Tepperwein, Secret Techniques of Hypnosis, Godefroy Publications, 1981.

BLAME

I don't expect anything from anybody, and I have no one and nothing to blame.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

Table of contents

BLESSING

May God bless all my undertakings.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I accept the good that comes to me, here and now.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I love so that I see only what is good, give only what is good, and receive only what is good.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I count my blessings.

Og Mandino, The Greatest Secret In The World, A Different World Press, 1979.

I thank God for the gift of life.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

From this moment on I am a complete expression of love and gratitude for God. I realize that there is so much good in my life.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I am ready to receive all the blessings of this abundant universe.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am grateful for the blessings I receive daily.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am thankful for everything that happens to me.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

BLISS

I feel the happiness and bliss of being alive.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

BLUSHING

My blood is descending into my feet. My feet are getting hot. My feet are burning!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

BODY

I have a body, but I am not my body. My body may be tired or rested, healthy or sick, but that has nothing to do with the real me. My body is a precious instrument which I use to experience and act on the outer world. But it is only an instrument. I treat it well, I try to keep it in good health, but it is not me. I have a body, but I am not my body.

Dr. Robert Assagioli, Psychosynthesis, Epi, 1965.

BOREDOM

Every day I try to overcome my boredom by working on achieving worthwhile goals.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I overcome boredom because I concentrate on the details and nuances of each new situation. Because I'm always interested in learning, I stay interested in my work.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

Boredom disappears for good when I observe the vastness of life.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

No matter what happens I will not allow boredom or obstacles on my path to deprive me of my peace of mind.

Orison Swett Marden, The Joy Of Living, J.H. Jeheber Publications.

Table of contents

BOWLING

My bowling style is getting better every day. My approach is supple and relaxed. I release the ball with more and more precision. I am able to control the ball and put the effect I want on it. I can hit strikes with ease. My score is getting better every time I play. Bowling is easy for me.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

BRAGGING

I learn to control myself when I am about to brag or try to take advantage of someone.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

BRAVERY (Also see Strength, Courage)

I am in the process of becoming a brave person.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

BREAKING UP

I chose to love you. That was then, but now I've changed my mind.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

BREASTS

My breasts are developing... growing... at the same time they are

Table of contents

becoming firmer and firmer... my breasts are becoming beautiful.

My breasts are becoming more beautiful every day... more and more beautiful... firmer and firmer... my breasts are becoming fantastic... I feel them swelling... pushing outward... growing... they grow larger every week.

I see myself with beautiful breasts... pushing against my sweater... attracting men's looks.

My chest will develop more and more... and I'll be proud of my breasts... which are becoming more and more beautiful... more and more attractive.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

BREATHING

My breathing is full, regular and deep; I am developing the habit of correct breathing even when I don't think about it, and I know that it is helping to keep my blood healthy.

Charles Baudouin, Psychology And Practise of Autosuggestion, Idegraph Publications, 1990.

My breathing is regular and calm.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

My breathing is completely calm and regular.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I am breathing a little more slowly.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

Every day I do deep breathing exercises... to maintain the elasticity of my lungs... increase their capacity... to purify my blood... and give my organism even more vitality.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

Table of contents

BRIDGE

I defend myself very well at bridge.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

BUDGET

I allow myself a budget for personal expenses.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

BUSINESS (Also see Money)

All my money comes from God. He has sincerely given me enough to meet all the needs of my business, or if He hasn't given it to me yet, it won't be long in coming. If I need more, He will provide more. Therefore, I don't even want to think about easy ways of making money by taking advantage of others. God gives abundantly, and I'm going to manage this business as if He were always here, close to me.

Robert Collier, Riches At Your Doorstep, S.I.P., 1962

I am becoming a better businessman / woman.

Vernon Howard, Psycho-Pictography, S.I.P., 1965

My business affairs are the affairs of God - they are owned, operated and expanded by divine guidance.

R. Charles Barker, The Science Of Success, Dangles Publications, 1983.

Table of contents

BUYING

I get the coat (house, car, etc.) that I want for a great price. Wearing it (or living in the house, or driving the car, etc.) gives me an enormous amount of pleasure.

Raymond Hull , Wanting Is Power, Les Editions de l'Homme, 1969.

Table of contents

CALM (Also see Self Control)

I feel calm and in control of myself under all circumstances.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I use my resolute calm to combat the threatening attitudes of vain, excitable and impulsive people.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I cultivate calm.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I am becoming a calmer, more relaxed person.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I feel calm, rested and radiant.
I am becoming calm and balanced.
And as I become calm, I feel peaceful.
I am completely calm and reposed.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I stay calm, no matter what happens!

Alan Houel, How To Deal With Difficult People, Godefroy Publications, 1990.

I am calm and am comfortable being who I am.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am calm... calm... calm... calm... calm... calm...

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

Table of contents

I am calm, tranquil, imperturbable.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

I am calm, very calm.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Editions, 1982.

I am calmer and calmer in all situations.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am imperturbable, calm. I am relaxed and peaceful.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

I am completely calm.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

My breathing is calm. My heart is beating peacefully. I am as tranquil as the depths of the sea.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I am cool, calm and collected.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

I control myself with the confidence of a master psychologist, and I stay cool under the most difficult circumstances.

Franck Rudolph Young, Cyclomancia, SIP, 1966.

I am cool, calm and lucid no matter how important the person I'm speaking to is.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

CANCER

I now imagine my tissues getting rid of this waste, becoming healthy again, clean and normal. My treatment is a success. I'm in

good health. I'm getting stronger and stronger. I feel better and better. I smile a lot. I see myself as a smiling, happy, active, healthy person. Everything's fine. I am healed.

Now I imagine my cancer being defeated. My cancerous cells retreat, they are beaten, they disappear over the horizon, far, very far away, out of sight. My cancer disappears, melts into the distance, very far away, so far no one can ever find it.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

For me, the word 'incurable' simply means that the illness cannot be cured from the outside, but only from the inside ... Dissolving mental patterns dissolves disease.

Anything can be healed. So, I am realistic, and plan for a miracle!

The doctors are amazed at my total remission.

My body is healing itself totally.

Louis Proto, Self Healing, Piatkus Publications, 1990.

CARDIAC TROUBLE

My heart is beating calmly and regularly.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

CARE

The more I take care of myself, the more others take care of me.

Louis Proto, Self Healing, Piatkus Publications, 1990.

CAREER

God leads me forward in a career filled with success.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

Table of contents

I know that I can earn a living in different ways, that I don't have to stick to the same job or even the same career just because I already have a lot of time invested in what I'm doing. I give myself permission to be whoever I want, and get into new areas which may have a bearing on my professional decisions.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

I find the ideal job that perfectly fulfils all my needs and desires.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I see every job change or career change as an opportunity to attain my objectives. I have a positive opinion of myself, and I feel good about myself. I am clear, calm, confident, and in control.

I never feel committed to work or a career that isn't totally beneficial.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

CHALLENGE

I look at every day as a challenge; and I try to give it the best shot I have. I start the day by making myself receptive to the outside world, and setting objectives which I try to define very clearly. I then decide what I can do to attain those objectives, and get right to work. I develop the habit of exploiting the opportunities that arise every day in order to attain my goals.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Each victory strengthens me for the next challenge.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

At this moment, even as I express these truths about myself, I know that I can succeed and that I will succeed. If I think about the

challenges awaiting me, whatever they are, I know I can win.

Just look what I can accomplish today! I am incredible... and today is the ideal day to prove it!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am eager to face new challenges!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Challenges are stimulating. I meet each challenge with determination and dynamism, and with the absolute certitude that I will get the best possible results.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

CHANGE (Also see Habit)

It is easy for me to reprogram the computer of my mind. All of life is change, and my mind is ever new.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I have the courage and strength to make the changes that need to be made.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I have the power to change myself the way I want.

Sublipower, Sublicontrol, Edi Inter, 1989.

I stop thinking that I have to change my situation before I can change myself. I change myself, and my situation changes naturally.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I start this day in a new and better way.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I will really begin a new life.

Table of contents

The world will start to become better as soon as I choose to make it better. I won't wait for others to improve the world. I won't wait for my neighbor to improve him or herself. I'll be the one to start.

I can change my own small world, and that's what counts. I choose to believe that I can change it.

If I use my power to choose it is amazing how well things work out for me, even beyond my wildest expectations.

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

I can change if I make the necessary effort.

I am going to be different.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I want to change this.

I want to change.

I want something different.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

CHARM (Also see Slimness, Health, Sincerity)

I am a radiant being, full of love and light.

From now on my perfect partner will find me divinely irresistible.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am a warm, sincere, caring and affectionate person. And I see these qualities in others as well.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Other people appreciate me because I try to make them feel good. I'm attractive in many ways. I attract other people, their interest, their enthusiasm, their friendship and their confidence. I bring the best

out in everyone, and that makes me attractive.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

My relations with men and women are more and more harmonious. I'm becoming more charming every day. I am capable of loving and of being loved.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

CHILL

I feel fine, I'm not cold, I'm not going to catch cold.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

CHOICE OF PROFESSION

Before making a choice that will determine the course of my entire existence I will spend weeks - months if necessary - gathering as much information and data as I can find concerning my future profession.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I completely abandon myself to You, my inner Guide. Lead me along the right path, and tell me what the right decision is. Enlighten me as to my vocation so that I can choose the profession that is best suited for me.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986

CIRCULATION (Also see Anemia)

My heart is beating normally and regularly; my circulation is regu-

lar; I know how my heart and circulation are affected by emotions, thoughts and suggestions. Thus I can achieve perfect circulation; if an organ has a tendency to become congested and swollen, it will receive less blood and less nourishment from now on, so that it gradually shrinks back to its normal size. If an organ tends to become anemic it will receive more blood. This, plus the fact that the blood it receives will be of a better quality due to my improved digestion, will make it grow stronger.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

My blood circulates freely, my heart beats calmly and strongly.

Sublipower, Sublihealth, Edi Inter, 1989.

CIRCUMSTANCE

I create my own future circumstances. I change myself today, and tomorrow's circumstances will be as I wish them to be.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

CLARITY

By cleaning and tidying up my physical environment, I clean and enlighten my whole life.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

CLIENTS

The people I call need my products and services.

I can't lose what I don't have. Any prospect I gain is one more than I have now.

Table of contents

I'm widening the base of my contacts.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

COLD (Common)

It's getting better now. It's getting better.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I am getting rid of my tendency to catch colds.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I am breathing more and more easily now. My nose and sinuses are clearing up completely, and I feel myself breathing well. I feel wonderful. All my respiratory passages are opening.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

My sinuses are perfectly healthy. My head is completely clear, and my eyes are dry.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

COLIC (Biliary or Hepatic)

I am relaxing my biliary passages; the channels are dilating; the stone is moving down into my intestine.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

COMFORT (Feeling At Ease)

I feel completely comfortable when I'm with other people.

Shakti Gawain, Techniques of Creative Visualization, Editions Soleil, 1978.

Table of contents

I feel at ease, sure of myself, and others appreciate me.

Claude Arpel, "I Have Confidence In Myself" - A Subliminal Method, Procauvi Publications, 1986

Every day I appreciate other people more, and find them fascinating. I feel more and more comfortable in the presence of strangers.

Paul-Clement Jagot, Education of the Word, Dangles Press, 1975.

COMMITMENT

I learn to say NO as if it means YES, I am already committed. One of the best ways to relieve stress is to schedule my time so that I can, comfortably, keep my commitments.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I don't commit myself halfway, but all the way in order to attain my objectives.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I honor my commitments.

Sublipower, Sublisuccess, Edi Inter, 1989.

Everything I do involves a price to pay.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

COMMUNICATION
(Also see Confidence, Tension)

It's easier and easier for me to see other people's points of view.

Christian H. Godefroy, Human Relations and Communication: The Subliminal Method, Godefroy Publications, 1987.

I have the right to be open with other people.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

Table of contents

I communicate with ease and with joy.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I always communicate clearly and effectively.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

The first thing I do is show the respect I have for the other person.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I speak with ease, anywhere and to anyone.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I share my emotions.
I enjoy speaking in public.

Christian H. Godefroy, Human Relations and Communication: The Subliminal Method, Godefroy Publications, 1987.

I always think about saying what I feel.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

I am the best judge of my actions, thoughts and feelings, and I am the one who is responsible for their consequences.

In order to develop, learn and evolve I communicate with others about my limitations.

In my search for happiness, it is in my best interests to communicate with others about conditions that promote our mutual well being.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

I am in charge of what I say and how I say it. Realizing that nothing is learned while I talk, much of my communication is listening, observing and qualifying. When I communicate I am prepared to deliver a message that offers value and leads to mutual understanding on the part of the receiver.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Table of contents

There's no mystery about communicating. I know how to listen, and I always listen with a clear and receptive mind.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

People feel at ease with me. People feel good in the presence of someone who radiates relaxation.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

COMPANY

My company is there to serve me, and not vice versa.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

COMPASSION

I will not forget that my compassion can make me free.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I can't offer compassion to others unless I am compassionate with myself. I am not unfair with myself.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I know that others are equally worthy, equally imperfect. I have compassion for them because they are engaged in the same struggle for survival that I am. Since I know that everyone is doing his or her best, it's easy for me to feel sympathy and compassion for other people.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

COMPETENCE (Also see Money, Skill)

I have confidence in my competence.

Table of contents

I am competent.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I'm always learning something new. I'm happy with the skills I already possess, and try to become even more competent in all my activities.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am becoming the best in my field, and money flows to me as if by magic.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I'm competent, hard-working, and really good at what I do.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I am open to learn new things every day, and that gives me the optimism, energy and inspiration to create a future full of success.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I am strong and competent.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

I want to develop my skills to their highest level, and acquire superior competence and ability in my fields of expertise.

Paul Clement Jagot, Power Of The Will, Dangles Press, 1950.

My confidence is growing, and as my confidence grows so does my competence.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

COMPETITION

I have confidence in myself and in my subconscious mind. I can

Table of contents

develop all the skills necessary for me to get ahead of the competition.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

I never compare myself to others in order to compete with them, because I can never become someone else, and therefore I cannot attain exactly what someone else attains. But what I can do that no one else can do is to be myself, fulfill my life's tasks, and reach the very summit of my own field of endeavour.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

COMPLEX

I embrace life openly... with no complexes... so that I can partake of the quintessential joy and pleasure of really being alive.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

COMPLEX - INFERIORITY

I have no complexes at all towards people I judge superior to myself.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I am no longer preoccupied with what other people think about me, since I know that all they're doing is wondering what I think about them.

I am totally indifferent to the opinions of others.

I thank my enemies for using their sharp eyes and tongues to point out the weaknesses in my life, and myself.

I thank everyone who has deceived me for making me more vigilant, and for having prevented me from becoming dishonest myself.

I know they probably all enjoyed doing this for me, but that is small payment for the service they provided.

Clement Auger, Transform Your Life Through Autosuggestion, Heritage Publications, 1979.

I know that in certain respects I am superior to many people I am acquainted with.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

The best way not to feel inferior of others is to do something about my inferiority complex. Confidence follows accomplishment. Feeling sorry for myself will never cure me of my shyness. Working hard, making myself useful, doing something for someone else will give me the kind of inner satisfaction that brings with it a feeling of self confidence.

I must not use my sense of inferiority as an alibi for being too lazy to improve myself.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

When I am with another person I never consider myself inferior to that person, not even in my thoughts. I always see myself as an equal. Because my personality is unique, just like his or hers, and I am the only one who can do what I must do.

I don't think about another person's professional or financial standing, but only about the person in whom, like myself, there lives an inner Guide whose role it is to harmonize personal good with the general good, so that I have no reason to fear, but rather sufficient reason to love and see in the other person an accomplice to my evolution.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

COMPLEX - SEXUAL

I let Nature do its work; I let myself get carried away in (sensual)

Table of contents

improvisations which arise spontaneously in my partner or myself, because I know that everything is normal and good about the ways men and women use to express their love for one another.

Clement Auger, Transform Your Life Through Autosuggestion, Heritage Publications, 1979.

I am free to express my sexuality as I choose.

Louis Proto, Self Healing, Piatkus Publications, 1990.

COMPLIMENTS

I always say "Thank you" when I am paid any compliment, by any one, for any reason. Neither do I try to play down or play up value that is bestowed. The ability to accept is the universal mark of an individual with solid self esteem.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

COMPOSURE

I cultivate composure... and stay calm under all circumstances.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I want to acquire constant composure, imperturbable self assurance and confidence that is worthy of who I am.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

COMPREHENSION

I make sure I understand the person I'm talking to correctly by reformulating his or her point of view, and asking if I've summed it up well.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Table of contents

CONCENTRATION (Also see Delegating)

I am learning to concentrate.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I concentrate all my energies on the challenge at hand, and my actions help me forget about everything else.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I'm in control of my thoughts and my imagination.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I believe I can concentrate. I am going to make concentrating a habit. I am going to reduce distractions to a minimum whenever I am reading or studying. I am going to visualize myself as a person having a good mind, the capacity to listen well, to become interested in everything, to concentrate better and achieve success in life.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I develop and improve my powers of concentration.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I master my powers of concentration perfectly. I can direct my attention and control my thoughts. My mental faculties are getting sharper every day, and I'm always looking for opportunities to make them even better, and to develop and improve my concentration. I practice concentrating my attention and my thoughts, so that I am the master of my own mind.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I concentrate on what I desire and prevent myself from thinking about what I don't want.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Table of contents

I can control my thoughts.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I can concentrate.
I can concentrate better and better.
I have a powerful ability to concentrate.
Nothing can distract me when I concentrate.
I do concentration exercises every day, I do them with pleasure, and they don't tire me out at all.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

I know I'm moving in the right direction; I look ahead, never backwards. Since I am able to concentrate on one thing at a time, I concentrate on the work at hand and finish it!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am calm. I feel good. I feel very good. Nothing can disturb my mind.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

My powers of concentration are improving day by day. I can therefore accomplish all my tasks more rapidly and effectively.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

The power of concentration has no limits, and I use it every time I speak or act.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

My mind does not wander.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

My mind can focus effectively on one thing at a time. The best use of my time and energy is to concentrate on the details of the task at hand and take things one step at a time.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

Table of contents

Not now, I'm concentrating on something.

Professor Kurt Tepperwein, How To Learn The Painless Way, Godefroy Publications, 1983.

CONFIDENCE

The answer is Yes!

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I trust in the process of life. I am safe.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I have confidence in — (name the person).

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

([I have a right to be here. I am a person too. I am okay.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

There are no weaknesses in my outlook... I always see the future serenely... with tranquil certitude and confidence.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

I leave yesterday behind and search for goals that will liberate my creative abilities.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I will not let this insignificant impression (or event) affect me.

Paul Clement Jagot, Power Of The Will, Dangles Press, 1950.

I can do everything I want.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I can handle it.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

Table of contents

I am capable.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I relax and trust the environment to support me. I place my trust in the Universe which helps me whenever I follow the dictates of my heart.

Louis Proto, Self Healing, Piatkus Publications, 1990.

People like me.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

The Real Me is completely confident, in any situation.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

No person and no thing has power over me.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I do nothing I cannot talk about, and I'm ready to talk about everything I do.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

If I confide in others, others will confide in me.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

CONFIDENCE IN THE FUTURE

My future is bright, healthy, loving, prosperous and SAFE. Being healthy, happy, loved and with no problems feels totally SAFE. My life gets better and better with every day that passes. With every day that passes I feel stronger and more alive. I look forward to the future with confidence and trust.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Table of contents

CONFIDENCE - SELF (See Self Confidence)

I think about my conflict situations. I try to find solutions. But I also now live joyously, as if my problems were already resolved. I live with this confidence, which never leaves me.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

If I'm in conflict with someone, I imagine that I am both myself and that other person.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

If I have to deal with a delicate or a conflict situation I make sure to create a positive atmosphere of respect for the persons opposing me at the moment.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

From this moment on all the difficulties between —— (name) and myself are erased and our relationship is marvelous.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

CONFORMING

I am infinitely more important than all these rules and social norms which are imposed on us for appearance's sake... talking in a certain way, never leaving the beaten path... that kind of repressive nonsense.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

CONGRATULATIONS

Tonight when I go to bed I will review my accomplishments of the day and reinforce my self image. I will think about them calmly and with pleasure; chances are I will have a very good sleep.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

I did my job perfectly today. I'm becoming more efficient every day.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I'm pleased with myself. My work was impeccable.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

CONSENSUS

I always look out for an agreement or some kind of consensus, even if certain things cannot be formulated officially.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

CONSIDERATION (Also see Respect)

I show others my consideration by making myself really available when I'm with someone, and through concrete actions that can help, rather than just empty words and wasted time.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I give myself the consideration and respect I deserve.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I treat myself as well as I would like others to treat me.

Sublipower, Sublicontrol, Edi Inter, 1989.

People see me as a good worthwhile person. Those who fail to see my worth have something on their screen that keeps them from seeing what a good person I am. When people really see me, they see me as a good person.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

Table of contents

CONSTIPATION

From this moment on I will feel the need to go to the bathroom every morning after breakfast... every morning without fail I will feel the need to empty my bowels...

I will feel the need to go to the toilet... at night before going to bed, I will go to the toilet... where all my tensions will be released... my stool is normal... evacuation is easy... all my tension disappears... my digestion is perfect... I'm so happy not to have digestion problems any more.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

Using my own nervous energy I can provoke a bowel movement... I am trying.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

Every morning when I get up, or more precisely twenty minutes after breakfast, I have a bowel movement without having to take any medication, or depend on any artificial substances whatsoever.

Emile Coue, Complete Works, Astra Publications, 1976.

CONSTRAINT

I am now ready to give up control. I no longer hurt myself. I am strong, wonderful and free.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

CONTACT

Every person is a link in the golden chain of my own, as well as the universal well being.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

Table of contents

I can choose the people I want to associate with.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I touch others more often.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

CONTRACT

I make sure the contract is mutually profitable, i.e. that the sum of the gains for both parties present is as large as possible.

I also make sure the contract is equitable, i.e. that the sacrifices or losses (and not the gains) are equally spread among the parties present.

I then establish a contract with the other parties which precisely defines "who does what, how, and with what means."

And I always include a clause that allows for renegotiation.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

CONVERSATION

Whenever I'm involved in important or delicate discussions I take careful notes and eventually I offer a summary of the conversation to the other party and ask if he or she agrees with my interpretation.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

CONVICTION

I speak with conviction to everyone, and this influences them.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

Table of contents

COOPERATION (Also see Help)

It is easier and easier for me to receive and give aid and cooperation.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

I commit myself to cooperate with others on constructive projects.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I meet my ideal business partner.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

COUGH

I don't cough anymore.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

COURAGE (Also see Dynamism, Resistance, Energy, Enthusiasm, Strength, Bravery)

Today is a magnificent day for courage.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Everything important in life deserves fighting for.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I have the courage to aim higher.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I summon my courage.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Table of contents

I feel courageous, full of energy and enthusiasm.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I feel happy, joyous and courageous.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I try to enjoy everything I do. I feel courageous, energetic and enthusiastic; the things I have to do seem easy. Work distracts me, interests me and attracts me.

Professor Robert Tocquet, The powers Of The Will, Godefroy Publications, 1989.

I take hold of my courage with both hands; it's so easy to abandon everything, and so hard to keep your head held high.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I am brave, I am strong, I am courageous.

George Barbarin, How To Overcome Fears And Anxiety, Dangles Press, 1974.

When I work I am supple, relaxed and happy.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Work distracts me, interests me and attracts me.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

The things I have to do seem easy.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

COURTESY

I am honest and sincere with everyone. I treat everyone I meet with courtesy and respect.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

COWARDICE (See Lying, Laziness)

CREATION (Also see Money)

From now on I have the firm intention of creating — (specify what) here and now!

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

CREATIVITY (Also see Sensitivity, Love)

In my own unique way, I am a genius.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

My aim is to understand an idea with maximum clarity.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I recognize and approach problems creatively. I view all problems as "situations needing improvement" or as "temporary inconveniences" and "opportunities to grow."

I take time to ride my bike, build sand castles, fly a kite, smell a rose, walk in the woods or walk barefoot in the sand. I explore the wonderful creative world of children (and I'm going to do just that this weekend).

I know that ideas are expendable, and there's always a new and better one. I challenge myself to make them work and get them into practical use.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I am becoming more and more creative every day. I realize that the only limits on my creativity are the ones I imposed on myself in the past. I now choose not to limit my creativity any longer.

Table of contents

I'm not afraid of exploring new ground or looking for the best solutions to problems concerning every aspect of my life.

The more I make use of my creativity, the more creative I become. The more creative I am the more successful I am in my undertakings.

I use my creativity to explore the source (of universal energy) from which all ideas spring forth. I ask myself a question, and the answer appears in my mind. I define the problem, and my creative mind indicates a solution.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I never experience "writer's block" when I write a letter or memo.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I step beyond the limits of conventional ideas.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I will always be a part of the living spirit. The creative power of God is in me and acts through me.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I am more and more creative and inspired in all areas of my life. I find plenty of ideas and solutions to my problems because I am able to draw from the inexhaustible wealth of my subconscious mind.

I am more and more creative.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I feel more and more inspired every day. I feel more creative in dealing with life situations, full of new and original ideas.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

From this moment on I am an open channel for creative energy.

I am the architect of my life.

From now on I will create my life exactly as I want.

I am an inexhaustible source of creative energy.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

I am the creative power in my world.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I am wise. I am intelligent and creative. I accept that a positive mental attitude is an important part of creating my reality.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

My subconscious provides me with answers for everything. I can always look to my subconscious for solutions to problems that arise. Now I feel creative, I feel inspired to get the best out of life, always.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

If I try to get both hot and cold water out of the same tap at the same time, I only end up with water that is lukewarm. If I try to criticize and create at the same time, I can neither criticize objectively, nor create ideas with enthusiasm. I am therefore going to concentrate on creating new ideas, and leave the criticizing for later.

Virginia Godefroy and Thierry de Foichaings, How I Got Rid Of My Money Problems Once And For All, Godefroy Publications, 1985.

If the problem appears to be a new one requiring new solutions, I rely on the creative approach.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I do what is necessary to make each day a creative day.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

CREDIT

I am going to get my loan.
My bank manager will approve the loan.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

Table of contents

CRISIS

It is entirely possible not to suffer the burdens of one crisis after another.

My Real Me can understand and resolve each crisis.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I never fail to take five minutes alone to get hold of myself (during a time of crisis).

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

CRITICISM

I want to judiciously examine all the ideas that come to me spontaneously, that I read about or hear about, before accepting them as valid.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

I do my best. Then I open my old umbrella and walk tranquilly through the storm of criticism, which no longer gets under my skin.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I laugh at critics... at bad reviews... which cannot disturb my serenity... nor destroy my confidence.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

I never forget that unjust criticism is often a compliment in disguise. I tell myself that criticism is often just a form of jealousy.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I never talk badly about anybody. I avoid criticizing, and even teasing people. I am only hard on myself. I try to notice other people's qualities in order to imitate them, and the bright side of situations in order to enjoy life.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

Table of contents

I keep a scrupulous account of all my blunders and write my own criticism of each one. Since no one can hope to be perfect, I try to do what a little soap salesman once told me: "I seek out frank, useful and constructive criticism."

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

Outside criticism doesn't bother me as long as I stay calm and don't react negatively. I have criticized others, it's human nature, but I want to get rid of the habit. In future I won't judge anyone on appearances. I'll let my inner Guide be the judge, which means I can accept everything and everyone with benevolence instead of criticism. And this will immunize me against hostile remarks. I'll take them with a smile, learn what I can and then continue on my way without being affected. My success is the best proof that the path I have chosen is the right one.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I thank my critics for spurring me on to ever greater achievements by offering criticism, which I may have misconstrued as a desire for vengeance, in the past.

Clement Auger, Transform Your Life Through Autosuggestion, Heritage Publications, 1979.

If I had to read all the criticism that is directed at me, never mind answer it all, I'd be better off closing down shop. I do my best with the capabilities I have, and try to see things through to the end. If it turns out that I'm right, then what was said against me is of no importance. And if I'm wrong, well ten angels could swear I'm right and it wouldn't make any difference.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

CURIOSITY

I experience the world with excitement and curiosity so that every day is a great day for me.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

DANGER

In case of danger I don't panic... I summon all my energy.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

There is a force in me that always makes me stronger than whatever is threatening me. I dominate the situation and come out a winner.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

DEAFNESS

I hear the truth. I know and love the truth. My ears are in tune with God's perfect plan. That is why my hearing is good. Divine harmony is revealed to me through the perfection of its instruments.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

DEBT (Also see Money, Projects)

I obtain the sum of money I need to finance my project and repay my debts.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

My business is not encumbered with debts, and my cash flow is good.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

Table of contents

DECISIONS

I forget about past mistakes; I stop torturing myself about them. Today is a day that will never return; I have to do everything I can to seize good opportunities when they occur.

I will not run around in circles any longer; I will set a direction for myself that will help me attain objectives which are really worthwhile.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I will always choose the correct course of action and accomplish it, because my Guide influences my thinking, inspires my words and directs my actions. With my Guide's help I can solve any problem and make positive decisions. I know what I want. I will succeed. I am succeeding!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I create a hierarchy of my priorities in order to develop clear criteria upon which to base my decisions.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I will not allow anyone to influence my decisions... to make me do what I don't want to do.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I will never make a hasty decision.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

I make judicious decisions.

When I learn to think with an open mind, the decisions I have to make on a daily basis become easy and profitable.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I make my decisions at the most opportune time.

I transform my decisions into actions that result in both profit

Table of contents

and satisfaction.

I never go around in circles.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I will be able to define what would please me the most, and thus make decisions. I will feel that my decisions are the right ones. At the same time I will try to improve my life, by continuing with my studies, for example.

Charles de Liguori, Hypnotism, de Vecchi Publications, 1975.

I am certain that I've based my decisions on a solid foundation, one that would benefit both myself and those close to me.

I have no trouble making decisions.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

My subconscious mind guides me towards making the right decisions in all my undertakings.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

When I have an important decision to make I not only try to optimize the quality / price ratio, but also the quality / time-spent ratio.

I try to involve the people close to me in the decisions I make, and try to explain my way of thinking to them.

If I can't decide between two options, I look for new ways to approach the problem.

Before committing myself to a course of action I evaluate the impact my decision will have on the people concerned (and on my eventual public).

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I approve of myself and my decisions are always perfect for me.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

Table of contents

DEFEAT

I never talk about defeat.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

DEFEATISM

There is only goodness, and the manifestations of goodness.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I never accept "no" as a satisfactory answer. I say "yes" to victory, I say "yes" to life!

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I never, never accept defeat.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

DEFENCE

I have the courage and strength to defend myself.

Louis Proto, Self Healing, Piatkus Publications, 1990.

DEFIANCE

I become more self confident, and this confidence gives me the certitude that I am capable not only of doing well, but of doing very well, of accomplishing anything I desire, within reason of course, and also of fulfilling my duty.

Emile Coue, Complete Works, Astra Publications, 1976.

Table of contents

I am regaining confidence in myself.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

DELEGATING

I concentrate on what is essential, on where my efforts are profitable, and delegate the rest.

I only do one thing at a time, with variations in what I do.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am learning to organize, to delegate a portion of my powers, to direct and to survey.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I delegate as much as possible.

I know how to make other people work.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I delegate a portion of my reading to subordinates and then tell them to sum up the text orally, or in a memo.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

DEPENDENCE

I am convinced that my whole life depends on my relation with the Creator... I don't rely on people, or on the conditions around me... I rely on the Infinite... on the Eternal.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

Table of contents

DEPRESSION
(Also see Conditioning, Confidence, Happiness, Health, Abundance, Optimism)

Day by day I feel better and better... I feel more and more happy... more and more joyful... I feel completely comfortable with my life.

My self confidence is growing day by day... I feel more and more happy... more and more joyful... I feel absolutely wonderful.

My mind is clear, free of all dark thoughts. I feel joyful and happy.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

I try to consider depression as a passing feeling. I don't consider it a permanent part of myself.

I banish my depression.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I start and end each day with a smile. All day long I make a conscious effort to smile and to feel good about myself.

I only allow myself to get carried away by thoughts that can help me attain my most important goals. My thoughts are always clear, constructive and powerful.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I believe that life with all its complications and hardships is worth living. I am going to stop brooding because I'm depressed. I am going to remind myself that depressed spells quite often vanish - that they are transient and that everyone at times experiences the 'blues' or low moods.

If my depression is chronic or too deep-seated, I will contact my physician, have him check me physically, and will let him decide whether I should consult some specialist.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Table of contents

I am becoming a stronger person.
I am free of the negative conditioning of the past.
I like myself.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I allow my hidden feelings to surface. I study them and sort through them. I keep those I like and reject the others. It's normal to feel depressed from time to time. But then I free myself of these feelings of melancholy. I don't allow myself to become a prisoner. I am free because I am able to reject feelings that make me sad or depressed.

I feel strong, healthy and full of energy. Everything is working out in my life. I am liked by other people. I like what I do. My life is happy and serene. I know what peace is. I feel good about myself. I have confidence in the future.

I don't know what real depression is.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

Although I feel depressed for the moment, I know it won't last. And soon, when I've practiced more, these kinds of feelings won't affect me at all.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1982.

I will not allow myself to miss out on any more of the good things in life. I have no intention of falling back into a stupid feeling of depression which I know only too well.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am abundance.

I am becoming a stronger and stronger person. I have confidence in life and I'm recovering my sense of joy. May this feeling of joy stay with me. I feel more and more happy. The current of life flows through me, bringing me an abundance of love and happiness.

I congratulate myself for handling the affair so well. And in fu-

ture I will always handle myself just as well. I'm in great shape, and I'm completely positive. I'm in a very good mood. I congratulate myself.

Every day, and in all aspects, I am becoming stronger, more confident, happier and healthier.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am relaxed, calm and happy. Everything is fine. I feel good. I await the future with confidence. Everything is fine, everything is calm.

Even if the weather is bad at the moment I know it won't last, and besides the weather isn't important.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I am happy, confident and strong.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I want my perceptions of things to be just as positive, and my feelings just as pleasant as they are negative and demoralizing when I'm depressed.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

Everything is going well, really well.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

DERMATOSIS

My blood is pure. My skin is beautiful.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

DESIRE

Table of contents

I have desires, but I am not my desires. My desires are transitory and sometimes contradictory. Attraction alternates with repulsion. There are desires in me, but they are not me.

Dr. Robert Assagioli, Psychosynthesis, Epi, 1965.

I now have enough time, energy, wisdom and money to realize all my desires.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I know that by applying the principles of autosuggestion, any desire that I choose to go after will soon materialize, first through external signs, and then as the attainment of the goal itself.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

The power of Infinite Understanding is at work in my consciousness and will realize all the desires in my heart.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

Now I very much want what I desire. Now I have the strength, energy and courage to be myself.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

If I really desire something I can obtain it.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

DESIRE - SEXUAL (See Sexual Desire)

DESTINY

Thus, day by day I write my own destiny; for inexorably I become what I do.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Table of contents

I have the power to create the life I wish for.

Sublipower, Sublicontrol, Edi Inter, 1989.

I can create my life with my thoughts.
I am responsible for everything that happens to me.
Change my thoughts, and I change my life.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I can orient my thoughts accordingly and, this done, and by using the greatest power I have - the power of choice - I am indirectly in a position to control circumstances.

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

I am the Master of my Destiny, I am the Captain of my Soul.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I am the master of circumstances.

Sublipower, Subliconfidence, Edi Inter, 1989.

Only I can decide what my destiny will be. I must do what I want to do, despite the obstacles.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

Above all I want, I want with all my strength, all the power lying dormant within me. I want to build a beautiful destiny.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

DETERMINATION (Also see Energy)

It is only by being relentless that I will succeed. I will therefore not be stingy with my efforts...

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Table of contents

I keep my head high. I look like the winner I am, and I act and think like a winner. Every time a problem starts wearing me down, I act immediately. I straighten up and attack my problems, and I solve them. Frustration and failure only make me stronger, more positive, better organized and more determined than ever!

I can imagine myself attaining my goals. I'm becoming more positive every day, and I'm more committed than ever!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I will never fail if my determination to succeed is strong enough.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I never let myself stray from the path I have chosen, nor allow anything or anyone to influence me in any way.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I can do what I please if I have enough determination and energy.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I have great determination. I know that nothing can stop me except myself. That's why I never stop moving forward.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I know that determination is the ultimate ingredient of all success, and I have a lot of it.

I am resolute, able to make decisions, constant, persevering and committed.

Each day I see how determined I am to attain all the goals I have established for myself, concerning both work and my personal life.

And because I accept and respect myself, I know that I really do deserve to be a winner in life. And that gives me the determination I need to be a winner, every day, in the most positive and beneficial sense of the word.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I want to do the right thing at the right time, despite any obstacles or warnings which may try to prevent me.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

If I work assiduously enough and for long enough to attain what I believe in, I will make the dream a reality.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I don't act with precipitation, but with determination.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I want to decide my own course of action and adhere inflexibly to the resolutions and decision I have made after proper deliberation.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

DIALOG

I always let other people express themselves freely and completely. I am attentive and reassuring when other people talk to me.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

DIET (Also see Hygiene, Negative Habits)

At each meal I eat half as much as usual.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

From now on I will no longer eat between meals.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

I like to eat the foods that are part of my diet. They are quite enough for me.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

Table of contents

Now I no longer want to eat what I crave at any given moment.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

My diet is doing me a world of good. I'm losing a pound and a half every week.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

DIFFERENCE

Please forgive me. I don't ask you to approve of what I did, or agree with me, but I do ask you to forgive me. I want to put our differences in the past, wipe the slate clean, and start fresh.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

DIFFICULT TASKS

This job looks so big that I'm avoiding doing anything. I'm going to break it down into smaller, manageable tasks and then get started on the first step.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

This task was destined for me. I can do it.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

DIFFICULTY (Also see Worry)

This is a tough situation. I'll have to dig deeper. Now is the time to really use my professional skills. I welcome the chance to test myself.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

Each difficulty is a challenge which I accept.

Christian H. Godefroy, Positive Thinking: The Subliminal Method, Godefroy Publications, 1987.

I succeed in dominating all difficulties that arise.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1982.

The Spirit in me knows, the Spirit in me can.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

When a difficulty arises I don't ask, "Who's to blame?" Instead I ask, "What's to be done?"

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Whatever happens I will react as calmly and intelligently as possible.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

DIGESTION (Also see Appetite, Stool)

I have a good appetite, I eat calmly, and I digest my food perfectly normally.

Professor Kurt Tepperwein, The Powers Of The Will, Godefroy Publications, 1989.

I feel no more pain, I have no weakness, my digestion is good, my circulation is very regular, all my bodily functions are normal.

Henry Durville, Course In Hypnosis and Suggestion, Durville Publications, 1938.

I transform the sunlight warming my skin into energy which I send to my lazy stomach to help it digest properly.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

My digestion and metabolism are functioning well. Every day at mealtime I tell myself that I am feeling very hungry; I eat with pleasure, but also with moderation; I get into the habit of chewing my

food well so that digesting is easier. I also know how sensitive my stomach and intestines are to emotions and to cerebral activity, and therefore to suggestion. Thus I can make my stomach and intestines function normally through the power of suggestion. If I digest properly I will assimilate my food properly as well. This means my organism will be nourished as it should be. And finally, my bowel movements are completely normal.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

My digestion is fine.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I digest food freely and easily.

Sublipower, Sublirelax, Edi Inter, 1989.

My stomach is ready to do its work harmoniously. I'm happy to sit down at the table since the meals I eat will be perfectly assimilated by my body.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

My stomach is functioning normally again. It has all the energy it needs to do its work and assimilate all the foods I eat. I am now digesting more and more easily. I assimilate foods with ease. My digestion takes place normally and easily, and I am relaxed and in good spirits.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Since my digestion is good I will also assimilate foods well; my body will benefit from all the foods I give it, using the nourishment to create blood, muscles, strength, energy - in short, life - so that each day I get stronger and more vigorous.

Emile Coue, Complete Works, Astra Publications, 1976.

Since I digest well, my intestinal functions are also normal. Every

morning when I get up I feel the need to have a bowel movement, and without having to use any kind of medication or artificial stimulus whatsoever I perform the function normally and to my complete satisfaction.

Emile Coue, Complete Works, Astra Publications, 1976.

DIGNITY

I stop debasing myself.
I never neglect my human dignity.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I will by worthy of respect, confidence and esteem. I want to set a noble example. I will therefore overcome any faults I have, in order to live up to this esteem.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I am worthy.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

DIPLOMA

I will make sure to finish my program of studies since I want to obtain my —— (specify which) diploma.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I am a successful person. There is no way I won't get my doctoral degree (or any other diploma - specify which). I am certain to succeed, because I know my success is already a part of the order of things. A new and fascinating life is opening up for me.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

Table of contents

DIRECTING OTHERS

I know how to be encouraging and understanding. I incite others to do their best and to fulfil themselves. I always seek out and find the best in the people who work for me, and instruct them in a positive way. Since I expect the best from others, I always get it. I assume complete responsibility for my person. And when I exercise my directorial functions, I let others assume their responsibilities as well.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

DISAGREEMENT (Also see Feud)

Tomorrow ——— (add name) will announce that s/he has changed his / her mind.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I don't feel insulted by people who disagree with me.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

If, by chance, our opinions happen to coincide, all the better. If they don't, we can't do anything about it.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

My willpower is strong; I will calmly refrain from arguing.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1982.

DISAPPOINTMENT

I try to laugh at my little failures and disappointments.

I try to see the funny side of things. I tell myself that it's not the end of the world, and that things will work out better next time. I

overcome my disappointments by not making mountains out of molehills.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

I'm disappointed with the results I'm getting so far. I'd better find out as much as I can about what is going wrong in this situation so I can improve it or at least get feedback to help me perform better in the future.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

DISAPPROVAL

When faced with someone's disapproval I react automatically by using sentences beginning with "you." I let anyone who is trying to make me feel guilty know that I am quite capable of surviving their disapproval.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

DISCIPLINE (Also see Laziness, Disorder)

I am teaching myself to be disciplined, and to listen with deference and goodwill to the people in charge of my education.

Professor Robert Tocquet, The powers Of The Will, Godefroy Publications, 1989.¥

DISCOMFORT

It won't be long before I feel better.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

Table of contents

DISCOURAGEMENT

I will not let the hesitations and errors caused by my inexperience discourage me. I can always learn something very important.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

DISCUSSION

At the end of a discussion I make sure everyone involved has understood each other.

At the end of a discussion I specify the conclusions I have drawn, and ask the other parties to do the same.

I define, along with the other persons, the goals of the discussion and the points to be covered.

I make sure I am completely available for the entire duration of the discussions.

I try to terminate a discussion within the allotted time, even when I am talking to long-winded persons.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

When a difficult moment in a discussion arises I try to concentrate on what I like or appreciate about the other person(s).

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

DISOBEDIENCE

I am becoming more obedient and reasonable.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Table of contents

DISORDER (Also see Discipline, Laziness)

All areas of my life are becoming more and more orderly. I am becoming a disciplined and orderly person. I am more and more enthusiastic and I enjoy my work. I take pleasure in accomplishing all my day to day activities. I have all the energy I need to do what I have to do. Every morning I wake up overflowing with energy, and I attack the day with determination and discipline.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am learning to organize my mind and my time more effectively, and to relax and free myself of worries about all kinds of things that are out of my control.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, 1978.

DISSATISFACTION

When I am dissatisfied I ask myself what I would like to be doing at that moment, and then I try to do it.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

DISTRACTION

I am going to keep myself busy and utilize my leisure time to advantage. I will always find something to do that will give me a lift, like seeing a good movie, reading an interesting book, listening to music or anything that will help me relax and make me forget my troubles.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Table of contents

DISTRESS

My inner Guide always advises me correctly, and shows me the right path to take. My Guide will rescue me from distress. S/He has the answers to all the questions that are tormenting me. S/He dissolves all my doubts. I have complete confidence in my Guide, and I know S/He is advising and assisting me at this very moment.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

DOUBT (Also see Self Confidence)

Here we are, with no doubts whatsoever.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

DREAMS (Also see Fear of the Unknown)

I let myself dream. I tell myself that I can have everything I want. With no taboos or prohibitions. That I have all the money I need to live out my fantasies for a period of fifteen days. I see that almost all my fantasies are realizable, that I don't desire the moon but simply things that I can obtain on condition that I get rid of my fear of the unknown and make the effort to stretch out my hand.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I want to remember one of my dreams, I can remember one of my dreams, and tomorrow morning I will have a clear recollection of that dream.

Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

Tonight, as on all nights, I am going to live a new adventure, and tomorrow I will remember all of it!

Pierre Fluchaire, The Dream Revolution, Dangles Press, 1985.

Table of contents

My dreams are becoming concrete reality more easily every day.
Sublipower, Sublisuccess, Edi Inter, 1989.

I have peaceful dreams, and my mind is at rest.
R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I can modify my dreams when I want to.
Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

I can (I want to) remain conscious when I dream.
Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

When I have a clear and lucid dream I feel perfectly rested.
Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

DRIVING

I always stay calm when I'm driving. All my tension and nervousness disappears. I think before I react.
Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

I am a competent driver. I pass my test easily and effortlessly. Driving gives me pleasure. I find it easy and fun.
Vera Pfeiffer, Positive Thinking, Element Books, 1989.

DRUG ADDICTION

I feel more and more detached from my real life... I am free of my negative, annoying thoughts... Nothing is important... I abandon myself to this feeling of deliverance.

Everything around me becomes clear, clean and more and more beautiful... Colors and music blend into pleasant images... I abandon

myself to the sounds and colors... nothing can disturb me.

I am indifferent to the noises from outside... I concentrate on myself... I am in a state which is much higher than any drug can take me... I can enter this state whenever I like... and be free of all worries, problems and negative influences.

No drug can do this for me... I don't need —— (specify the drug) which is ruining my health... and my health is improving every day.

I have confidence in my destiny, which I control through my desires... but nothing will make me go back to taking drugs... I don't see the point anymore.

Drugs disgust me... nothing on earth could make me start taking drugs again... if I feel the need to shut out reality, I will do some self hypnosis or meditation... things which regenerate my entire organism.

I do a few minutes of meditation every day... I concentrate on my future... I am overflowing with strength and energy... the solution to my problems becomes clearer every day... and I'm doing more and more to attain my goals.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

I'm proud of myself and my success.

People like me when I'm not on drugs.

I'm healthy when I'm not on drugs.

I work much more efficiently when I'm not on drugs.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

DRUGS

I will not forget that tobacco, coffee and alcohol are drugs. i.e. means of stimulating my body. They can be positive, but there is also a risk of becoming dependent.

Table of contents

I avoid using one drug to compensate for the effects of another (for example drinking coffee when I've had too much alcohol).

I change my habits from time to time, especially those concerning drugs, to determine if I'm addicted or not.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

DUTY

I will do today's duty today.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

DYNAMISM (Mental - Also see Ambition, Enthusiasm, Vitality)

When I'm in the alpha state I condition my mind to become more open to the suggestions and directives I give it. I feel better every time I practice my alphagenic exercises.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

From this moment on I will always be ready and alert, strong and conscious of my goals. Nothing can distract me from attaining my objectives. I am untiring, opinionated and persevering. I take more and more pleasure in the joy of creation.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

From this moment on I am positive and I am going to keep moving forward!

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

I am a super dynamic person.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

Table of contents

DYSFUNCTION (Organ)

Generally speaking, if I have one or a number of organs which are defective to a greater or lesser degree, I tell myself that these organs are starting to function better and better, so that they gradually return to normal and I am completely healed.

Emile Coue, Complete Works, Astra Publications, 1976.

My mind is in command. All my organs obey my mind.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I make sure all my organs are functioning well; my heart beats normally and my blood circulates as it should; my lungs are working well; my stomach, intestines, liver, bladder and kidneys all perform their functions normally. If one of these organs happens to be malfunctioning at the moment, the problem will diminish day by day, so that in the very near future it will disappear completely, and the organ will resume its normal function.

Emile Coue, Complete Works, Astra Publications, 1976.

DYSPEPSIA

I send my stomach the order to contract.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

Table of contents

EASE

It's easier for me to do what is good for me.

Sublipower, Sublicontrol, Edi Inter, 1989.

I master X technique (specify) easily and pleasantly.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

ECONOMIZING

I never spend my money stupidly. I know the value of wise spending. All my financial decisions are based on the principles of economy and moderation.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

ECZEMA (Also see Acne, Urticaria)

From now on I will stop thinking about my eczema, and think about my intestines instead; I have to get rid of the toxins that are being eliminated through my skin through the mucous membrane of my large intestine instead. As soon as I start doing this my eczema will disappear.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

Table of contents

I am a healthy person, pure and perfect; my skin is healthy, and I am healed.

My skin is perfect.

Dr. Joseph Murphy, The Miracles Of Your Mind, Dangles Press, 1984.

My skin is becoming normal, my eczema is disappearing and my skin is becoming smooth, beautiful to the touch, and perfectly normal.

Clement Auger, Transform Your Life Through Autosuggestion, Heritage Publications, 1979.

EDUCATION

As a father of six children I know from experience that the greatest gifts a parent can give their children (and that managers can give their employees) are roots and wings. Roots of responsibility and wings of independence.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

It is my responsibility to teach my child, by example, what real love is and how to work out his problems intelligently, instead of emotionally, how he can live with the greatest amount of happiness in this society of ours.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I never resort to exaggerated threats or warnings which are forgotten in any case. I stick to my word.

I tend to give large or small rewards on all kinds of occasions.

For me each day is an opportunity to show, by setting an example, what the best way to live is.

I teach the values I believe in by setting a good example.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I understand the difference between punishment, discipline and teaching, and I always try to consider these elements from their proper perspective.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

My children are my greatest joy. Their noisy games are a sign of their minds' vitality and vivacity. My spouse brings life into the home, and merits all my love. We smile at each other a lot and understand each other better each day.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

First, I am able to recognize my children's unique abilities and talents - to reinforce them, nurture them and help them recognize what is special about themselves.

Second, I am able to understand their behavior in the context of who they are - I don't misinterpret natural shyness as being unfriendly, or a need for privacy as rejection. Seen in context, even negative behavior is more understandable and predictable.

Third, seeing my children accurately helps me focus on changing only the behavior that is important to change - behavior that's harmful to them, behavior that isolates them socially, or behavior that is disruptive to the family.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

EFFICIENCY

Every day I am improving, I'm improving more and more in all ways.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I give the best of me in everything.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Table of contents

I am efficient in everything I do.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I use my time well. I do everything I have to do quickly and passionately.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I concentrate my time and energies on the 20% of my activities, contacts and concepts that have proven most productive to me in the past.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I do my best.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I do my work as well as possible, and I am no more incompetent than anyone else.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I always do my best according to my awareness at the moment.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I remind myself to only do one thing at a time.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am an organized and efficient person. I am mastering the art of doing more in less time.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I'm doing my job better. My performance is improving.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

Table of contents

EFFORT

Master of my body, master of my thoughts, I want my effort to also be others' salvation.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

Every effort also constitutes an investment.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

EGOISM

Attaining an objective implies sharing it with other people, and thus I move from isolation to happiness.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

By developing a love of humanity in myself I eliminate all hate, envy, jealousy, all egoism and all cynicism from my heart, because I know that harboring a negative attitude towards others can only result in disappointment.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

ELIMINATION (Bowel Movement)

I will assure better elimination of my body waste.

J.V. Cerney, Stay Younger, Live Longer Through The Magic Of Mental Self-Conditioning, Parker Publishing Company Inc., 1968.

ELOCUTION

Every one of my words is spoken deliberately, fluently, and pro-

nounced clearly.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I have all the tools I need for perfect elocution.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I dare speak out.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I pronounce my words strongly, clearly and fluidly.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I am determined to make my words become influential and dominating.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I'm going to think carefully before saying anything - being mindful of the effect my words may have on others. I am going to avoid being sarcastic, learn to become more diplomatic and develop the art of friendly and interesting conversation. I am going to be less argumentative. If I learn to control the impulse to say the wrong thing in haste, I will become a better person, one who will be loved and respected.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I make sure I have the language, with enough vocabulary and sufficiently varied examples, to best adapt what I'm saying to the person or people listening to me.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I'm becoming much more articulate and effective in my use of language.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

My elocution is becoming clear, harmonious and supple.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Table of contents

EMOTION

Today I will be the master of my emotions.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I have emotions, but I am not my emotions. These emotions are innumerable, changing, often conflicting, yet I always remain myself, through hope and despair, joy and pain, irritation and calm. I can observe, understand and judge my emotions; I can master them and channel them and use them in better and better ways. I have emotions, but I am not my emotions.

Dr. Robert Assagioli, Psychosynthesis, Epi, 1965.

I dominate my emotions.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I am able to control my emotions.

Claude Arpel, Self Confidence: The Subliminal Method, Procauvi Publications, 1986.

When a problem arises I avoid becoming emotional. I Stay Calm.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

EMOTION - NEGATIVE
(Also see Fear, Anxiety, Nervousness)

I cease being the victim of negative emotions. I realize that they have absolutely no power over me.

I know I am on the path towards dissolving all my negative emotions when I begin to consider what they really are.

I become my best friend by trying to understand my negative emotions.

The power of self awareness allows me to reject negative emo-

Table of contents

tions whenever they try to influence me.

I destroy unhealthy thoughts and harmful emotions as soon as I start becoming aware of them.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am able to control my emotions perfectly in all situations. I am gaining complete mastery of my reactions. I am more and more positive and fulfilled in all situations.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am going to keep my negative emotions under control.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

EMOTIONAL SCARS

I erase the scars of my negative feelings.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

EMPHYSEMA (Also see Asthma Attacks)

I am calm, I am breathing, my lungs are relaxing, my kidneys are opening.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

ENCOURAGEMENT

I play a role in the lives of all the people I get close to. By treating someone like the person they should be, I offer a powerful encouragement for them to become that person.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

Table of contents

ENDURANCE (Also see Dynamism, Courage)

Since I take good care of myself I have a lot of energy, endurance and vitality.

I now have more energy and endurance than I did before.

I love life, and I'm happy to be alive.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I have more and more endurance... more and more strength... I get stronger every day... I can make progress... and become stronger.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

I know that success isn't due to chance. Real and durable success requires a sum of intellectual and moral qualities. The principal physical quality is endurance. I must be healthy. Therefore I want to live in a healthy way.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I possess dynamism, courage and endurance. I have a very positive attitude towards myself and towards everything I do. I am practical and realistic, but I also believe in getting the best possible results from any situation.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I have more and more endurance, more and more strength.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

I have more endurance than X (specify).

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

Table of contents

ENERGY (Also see Courage, Determination, Strength, Vitality, Health)

Each day I accumulate new energy in my solar plexus region.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

I have energy.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I inhale and I am inspired. I exhale and energy flows through me.

Christian H. Godefroy, The Seven Eternal Laws of Success, Godefroy Publications, 1980.

I am overflowing with more and more energy and strength.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I will destroy the negative images my enemies have tried to inject into my mind, and replace them with joyful images that encourage me and increase my energy.

Orison Swett Marden, The Joy Of Living, J.H. Jeheber Publications.

I feel very good today! I have more energy than ever.

I can do great things. And every task I accomplish gives me even more energy!

I have a lot of energy, which I nourish with my curiosity and enthusiasm.

I have more and more energy every day. I am full of life, in great shape, and in an exceptionally good mood.

When I do something that requires a surplus of energy I call on my reserves and find that I always have more than I need.

Every day I am becoming more and more aware of my energy and vitality. I create an unlimited reserve of energy in myself.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I do not waste my emotional energy.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

Table of contents

I feel the life force flowing through my body.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I am in excellent health, overflowing with energy.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I am bursting with energy.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

Energy flows freely through my body.

Sublipower, Sublirelax, Edi Inter, 1989.

My breathing is getting deeper and energy is circulating freely through my body.

Sublipower, Sublihealth, Edi Inter, 1989.

My most powerful energy for creating my life is present in me at this moment.

My energy is stimulated by my breathing, by love, and by life-giving thoughts and words.

I am bursting with energy and light.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

To be energetic I act energetically!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Every day and from all points of view I am more and more energetic.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I feel great! I'm full of energy! I'm so strong, I'm always going to feel healthy!

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I balance my energy.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Table of contents

I'm in good physical condition.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Editions, 1982.

I'm in great shape.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

ENERGY - CREATIVE

I am overflowing with divine light and creative energy.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

ENERGY - VITAL

I absorb the energy scattered throughout the atmosphere, and store it in me. I exhale air, but I retain the energy which is carried into my body by the air I breathe, and which I make my own.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1982.

ENJOYMENT

I relish each golden day.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

All the obstacles to my total expression in life and my complete enjoyment of existence are now dissolved.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

ENLIGHTENED MIND

My mind is clear, alert and ready to meet any challenge.

Table of contents

I enthusiastically wait for opportunities to put my mind to the test and improve it. Instead of considering problems as difficulties to avoid, I see them as opportunities to sharpen my mental faculties.

I am awake and fully conscious. My senses are sharp, my mind is clear. I am aware of everything around me, and I am in control of my life.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Divine light and divine love are flowing through me.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

ENTERITIS

If I am suffering from enteritis I imagine the enteritis is clearing up; the intestinal inflammation is gradually disappearing, and the phlegm and other secretions are becoming less and less abundant until they disappear and I am completely cured.

Emile Coue, Complete Works, Astra Publications, 1976.

ENTERPRISING SPIRIT

The search for the best in myself is my greatest enterprise. It is a creative adventure, because it leads me towards self acceptance.

With an enterprising spirit I can combat all negative feelings which belittle me and destroy my true identity.

With an enterprising spirit I increase the power of my personality as I find new horizons, and fully accept what I have already discovered.

With an enterprising spirit I progress towards a world of pure dynamic energy.

My enterprising spirit is the creative effort I make to resolve the

enigma which is in me, to try and understand my complexities and problems in order to overcome them.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

ENTHUSIASM (Also see Ambition, Dynamism, Vitality, Energy, Courage)

Today is the first day of the rest of my life.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

Thunder and lightning! Life electrifies me!

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I am naturally enthusiastic.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am enthusiastic about my goals; I develop this quality which is so essential to my success.

I apply myself with gusto to creative endeavors. I understand that life can be enjoyable and fun.

Every day I see myself in top form.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I have an iron will... nothing can discourage me... on the contrary, problems only give me more energy... the energy I need to overcome them... nothing can stifle my enthusiasm... absolutely nothing.

I cultivate my enthusiasm... my passion for everything I do.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I give myself a chance to become spontaneously enthusiastic about what I'm doing.

I am enthusiastic about everything I choose to undertake.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

Table of contents

I stimulate my enthusiasm by maintaining a high level of interest. The more my activities interest me, the more enthusiastic I am. The more enthusiastic I am, the more energy I have.

I regularly program my mind with positive instructions designed to keep me in shape, full of youthful enthusiasm.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am enthusiastic about life and filled with energy and enthusiasm.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

Life interests me.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

My day starts when I wake up in the morning and shake off the numbness of sleep, overcome my negative impulses and promise to make this day exciting.

I will not spend my day doing nothing; I have exciting goals and I'm full of enthusiasm.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am enthusiastic about my life.

Christian H. Godefroy, Positive Thinking: The Subliminal Method, Godefroy Publications, 1987.

Oh, this is a great day for getting out there and conquering the world!

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

To be enthusiastic I act enthusiastically!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Every day, in all ways, I am becoming more and more enthusiastic.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

Table of contents

ENURESIS (Bedwetting)

From now on I will stop wetting my bed, and I won't get angry when I have to go to bed. My bed won't be wet when I wake up in the morning.

I like going to bed at night.

If I have to go to the bathroom at night I get up and use the toilet. As soon as I get back into bed I fall asleep and my bed stays dry.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

From now on I will have a very strong desire to overcome this shameful habit, which is really a feeling of inferiority. A moment will come when my desire to control my bedwetting will be stronger than anything else.

I can now control myself much better than before.

Every night before falling asleep I order myself to wake up as soon as I feel the need to go to the bathroom. If It happens I will wake up.

I can control myself. I will not wet my bed. I am capable of controlling my bladder no matter what. Even if bedwetting has some hidden meaning for me, even if I want to be obstinate, I can still overcome my enuresis.

And even if I have a relapse, it doesn't mean anything. I will look for the cause of the relapse. I no longer suffer from enuresis. I can control it, and I will succeed in controlling it.

Charles de Liguori, Hypnotism, de Vecchi Publications, 1975.

I always call my mummy, and my bed is always clean.

Emile Coue, Complete Works, Astra Publications, 1976.

I always ask my mummy when I have to go to the toilet, and I always stay clean when I sleep. It's easy for me to stay dry and clean the whole night long.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Table of contents

I can restrain myself all night long and my bed is always clean and dry.

Emile Coue, Complete Works, Astra Publications, 1976.

I am a pretty little girl and people like me a lot. Everyone likes me because I'm so likable, so happy and gay. People like me because I only sleep in a clean, dry bed. I always sleep in a warm, dry bed. If I have to go to the bathroom, I always wake up in time.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

My bed is clean and dry, and I always wake up when I have to.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

If I concentrate on what I feel as I urinate during the day, I will feel the same sensation at night when I'm asleep, and after I've gone to the bathroom I'll get back in bed and fall asleep instantly.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

ENVY

This is so wonderful! I rejoice in that person's prosperity, and wish him/her more and more good fortune.

Dr. Joseph Murphy, The Miracles Of Your Mind, Dangles Press, 1984.

EQUILIBRIUM (See Balance)

EQUANIMITY

I maintain my tranquillity and equanimity.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Table of contents

ERRORS (See Mistakes)

ESTEEM (Also see Self Esteem)

I accord myself this elementary form of respect. I do not deny who I am. I'm worth more than that. I am in no way obliged to lead a monotonous existence.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I esteem myself. I know I am a good person.

Claude Arpel, Self Confidence: The Subliminal Method, Procauvi Publications, 1986.

I never belittle myself to myself.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

EVOLUTION

I am a directly linked representative of the ever-evolving, ever-expanding consciousness of the Creator. I am spirit expressing through a physical structure. I am patient with myself.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

EVOLVING

I have the right to develop, to learn and to evolve.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

EXAMS

(Say this to yourself on the night before an exam): I will not waste

Table of contents

a single minute more on this stupid activity, and I really don't care if I pass or not.

Dr. Maxwell Maltz, Psychocybernetics, Godefroy Publications, 1978.

From now on I always concentrate only on what I'm doing, and any noises only serve to reinforce my concentration.

I prepare very carefully for all my exams.

I am calm and relaxed. I can always remember what I learned when I need to, and therefore all my answers are clear and precise.

I channel all my attention towards my goal.

My marks are improving daily.

I am in complete control in all situations, and I attain the goals I set for myself within their prescribed time limits.

Professor Kurt Tepperwein, How To Learn The Painless Way, Godefroy Publications, 1983.

I study hard. I am certain to be ready on exam day.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I will pass my exam.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I am going to pass the exam. I am calm and my mind is clear.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

I am secure in the knowledge that I have prepared well for my exam. My thorough revision is now paying off. Everything I have revised is ready in my mind to be recalled easily and effortlessly during the exam.

Vera Pfeiffer, Positive Thinking, Element Books, 1989.

EXCHANGE

I like to give and I like to receive.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

I give and I receive.

Sublipower, Sublisuccess, Edi Inter, 1989.

EXCUSES

Once and for all, and from this moment on, I will get rid of these excuses.

I will not excuse myself all the time, especially when I'm really not sorry for what I said or did.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I can decide not to start acting in such and such a way again, I can realize that I'm wrong to adopt such and such a behavior, but excusing myself all the time is a sickness which makes me feel responsible for the emotions of others.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

EXERCISE (Also see Obesity, Nutrition)

I have to exercise every day so that I won't get sick... and stay in top shape.

By exercising every day... never missing a single day... even if only for a few minutes if I'm pressed for time... I will live at least ten healthy years longer.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

Every time I exercise I have a better opinion of myself. I'm really proud of the extraordinary work I'm accomplishing.

I always get the exercise I need. I don't stop until I have attained my objective, every time I exercise.

I never force myself too much or hurt myself physically or other-

wise. I do exactly the amount of exercise I need.

I couldn't live without exercise. It helps keep my healthy. It makes me alert, makes me feel better, and allows me to give the best of myself!

I set objectives for myself and then I achieve them. I'm on a regular exercise program, I eat healthy food and I get enough sleep. I spend a part of each day staying in shape.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I like doing exercise.

I feel like moving, walking...

Sublipower, Sublihealth, Edi Inter, 1989.

I can't wait to do some exercise today.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I will do a moderate amount of exercise every day, and I'll have fun doing it.

J.V. Cerney, Stay Younger, Live Longer Through The Magic Of Mental Self-Conditioning, Parker Publishing Company Inc., 1968.

I am going to engage in some form of moderate daily exercise.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Any amount of exercise (running, jumping rope, aerobic dancing, bicycle riding, etc.) gets my mind and body working much better. Exercise makes joy!

Exercising brings me joy.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

EXPERIENCE

I have learned from my experiences, even if sometimes they were

Table of contents

painful.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I listen for the secret messages hidden in my experiences. I let these help me develop positive and useful habits.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

To get into a specialty I first have to learn the basics.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

EXPRESSING ONESELF (See Self Expression)

Table of contents

FAILURE

From this moment on I will very calmly devote all my energy to finding ways likely to reduce the impact of the consequences which I have already mentally accepted.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

This is NOT a catastrophe! This is not horrible, humiliating, or the end of the world! It's inconvenient. I'm disappointed. Now I'm going to learn something and start again. So let's get going!

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I am learning that failure, no matter how serious, can be a trampoline to success and not a chopping block.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

At least once a week I review my goals and find out what I truly want now. I am willing to risk giving up all I have been for what I can become.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I'll do better next time.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I transform failure into success.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

Next time I will succeed.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

Table of contents

FAILURE (In School)

I can, therefore I must. I could have avoided this mistake, I should have been able to avoid it. I could have learned the lesson better and more quickly, so I should have done it. I could have done better, so I should have.

Professor Robert Tocquet, The powers Of The Will, Godefroy Publications, 1989.

FAITH (Also see Fear)

Belief helps me follow the way.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I have faith, faith in my ideas, in my destiny, and faith in God.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I am free of all fear, I am firm in my faith.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

Faith is invincible.
I believe, I really believe.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Well directed faith adds amazing power to all of my thoughts.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

FALSE BELIEFS

I never let my convictions trap me into passivity. Becoming attached to a belief based on past experience is like trying to escape reality. Only the present exists, and what is true in the present may

Table of contents

not necessarily be true of the past. I do not base my behavior on what I believe but on what is, what I experience in the present. By opening myself to the present experience instead of clouding reality with my beliefs, I realize that the unknown is a vast and fascinating domain.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I am ready to clear up any beliefs that hold me back from the good I want in life.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

FAMILY LIFE

My professional life and my family life are mutually enriching.

My family life is rich and creative.

I am there for my loved ones, giving them both time and my full attention.

My family understands and appreciates my professional activities.

I don't feel guilty about being absent from home when I work.

I make sure that my family life plays an important role in my balance as a human being and my personal fulfillment.

I always make a note of any meetings with my spouse or children in my agenda, in the same way that I keep track of meetings with my boss, or clients.

I ask myself if there is anything positive in my family life that can also be applied to my professional life (and vice versa).

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

FASCINATION

Others find me fascinating and terrific.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

Table of contents

FATIGUE

I like my work, but I also make sure to take time out to have fun: in this way I rarely feel tired.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I am not in the least tired.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

I am never lazy... physically or mentally.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

I'm in great shape.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

The sensations of fatigue and weakness that I sometimes feel will gradually disappear and be replaced by sensations of strength and vigor which will become stronger every day.

Emile Coue, Complete Works, Astra Publications, 1976.

FAULTS

Every day I am going to improve and eventually I will succeed in eliminating the faults I have discovered in myself.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

I accept the helpful criticism of my friends, and I use self hypnosis to purge myself of whatever personality shortcomings I possess.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I accept my weaknesses and my faults so that I feel all right even when I fail.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

I use my faults as inspiration to motivate me to improve and save more time.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

FEAR (Also see Emotional Paralysis, Anxiety, Nervousness, Guilt, Rejection)

I confront my fears with a positive state of mind and positive behavior.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I do what frightens me.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I have nothing to fear, the Spirit in me is watching, I have everything to hope for.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I fear nothing since my faith in the Mystery is great.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I know that the love of goodness fills my soul and dissolves all fears. I now live in the joyous expectation of goodness.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

Simply because I am aware of it, and because I abhor it, because I absolutely don't want this to happen - it will not happen!

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

Nothing can harm me since I am united with my inner Guide!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

The Real Me knows no fear.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

Table of contents

No matter what threatens me I will face up to it and meet the challenge!

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

If I find myself in a situation that frightens me, for example, I think about something that will calm me down.

Franck Rudolph Young, Cyclomancia, SIP, 1966.

What cannot defeat me only makes me stronger.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

Every day I overcome my fears, worries and any other destructive thoughts.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

[To overcome fear I use my mind.]

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I confront my fears and I accept them.

Sublipower, Sublicontrol, Edi Inter, 1989.

I free myself of my fear of what other people think.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

Only I can free myself from my fears. I start by becoming aware of what I fear, and stop myself whenever I find that I'm thinking or talking about it. I become aware of the 'fear-talk' around me, and stop listening to it.

I specify my fears. As long as they remain unclear they continue to have power over me.

I ask myself what the worst that could happen is. Not only is the worst hardly likely to occur, the answers often make me laugh!

I don't suppress fear. I feel it, talk about it objectively.

I am as safe as I allow myself to be. Only good comes to me. I

have nothing to fear.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I allow myself to be, without the need to prove myself to anyone. I live my power without fear of defeat or rejection. I control my dreams, my life.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I am not afraid of being afraid.
Knowing myself helps free me from fear.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I reject useless fears.
I understand my fears and I can explain them rationally.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I smile and my fears evaporate. I am strong. I am powerful. I have confidence in myself.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I am calm and relaxed.
Fears may come and go, but it doesn't matter; I can control them as long as I stay relaxed.
When I can control my thoughts, I can control my fears.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

I am inviolable, I am invulnerable.

George Barbarin, How To Overcome Fears And Anxiety, Dangles Press, 1974.

The Eternal is my refuge, and I make the Highest of the high my home.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

The Real Me cannot be afraid of anyone.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

Table of contents

FEAR OF ANIMALS

From now on I am no longer afraid of —— (name the animal). I find them tranquil, calm animals. When I see them I identify as much as possible with them.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I love all animals.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

FEAR OF DEATH

I meditate on death every day in order to give my life more meaning.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

FEAR OF ELEVATORS

I have no trouble taking the elevator. I now feel perfectly relaxed whenever I enter an elevator. I am able to control my reactions and emotions completely. I am able to master the situation completely, remaining calm and relaxed.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

FEAR OF EXAMS (Also see Stagefright)

I know I can pass my exam. I know the material perfectly. I'm ready. I have no reason to worry.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

FEAR OF FAILURE (Also see Disapproval)

I'm making a common mistake. I'm thinking of failure as the enemy of success. But it isn't at all. Failure is a teacher - a harsh one perhaps, but the best. I can be discouraged by failure - or I can learn from it.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I remind myself that fear of failure is really often a fear of disapproval or ridicule, and I tell myself that everyone has a right to their own opinion, but that other people's opinions don't concern me.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I do not fear failure. I am responsible for my actions. I acquire the necessary knowledge and skills to succeed, and I move forward.

I am never afraid to try.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

FEAR OF FLYING

I like airplanes, I am fascinated by them, they are a wonderful way to travel, and I am fortunate to be living in an age where this mode of travel is available to me. I am sharing the benefits of modern progress with many thousands of other people.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

FEAR OF REALITY

I remind myself that nothing human is foreign to me. I can be what I choose to be. I engrave this sentence in my mind, and I recall this aphorism every time I fall into the familiar habit of trying to es-

cape reality.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

FEAR OF REJECTION

It's better to ask for what I wish to obtain and get "no" for an answer than never to ask for anything, and never get anything.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

FEAR OF THE DARK

From now on I am no longer afraid of the dark. Nothing changes when it gets dark. There is no reason to be afraid. On the contrary, I know that darkness is my friend because it helps me rest.

The darkness is my friend, it helps me get a good night's rest.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

FEAR OF THE FUTURE

Starting today I am no longer afraid of the future. I feel strong. I am in control of my destiny. I have absolutely no reason to be afraid. The future belongs to me. I am happy, loved and respected.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

(From now on only beautiful things will happen in my life.

I rejoice at the thought of the beautiful moments I am going to experience.)

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I am very happy... I feel good about who I am... I look confidently

into the future... and I know: my destiny is full of wonderful events.

Day by day, in all ways, things are getting better.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

Life is for me.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

Guide me, celestial light, guide me on the right path; I don't ask to see what the distant future holds in store for me. Just the next step is enough.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

FEAR OF THE OPPOSITE SEX

I feel better and better in my relations with the opposite sex. I feel more and more comfortable with men / women. My relationships are harmonious. I am completely at ease, and men / women find me attractive. I am more and more confident and sure of myself with men / women.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

FEAR OF THE UNKNOWN

If I really believe in myself, everything is possible. The totality of human experience is at my disposal from the moment I decide to adventure into unknown territory, without asking for guarantees of any kind. Just because it's different, and I'll have fun doing it.

If I see myself shrinking back from the unknown, I ask myself: "What is the most terrible thing that can happen?"

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Table of contents

FEAR OF VEGETABLES

From now on I am no longer afraid of — (name the vegetable). I know I am stronger. I see them as friends. They are beautiful plants, basking in the same sunshine as we do. We share the planet with them.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

FEAR OF WATER

From now on I am no longer afraid of water. I know there is nothing to fear because water is an element of nature. I am confident, I am happy, I feel good in the water.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses , Godefroy Publications, 1989.

FEELINGS

To attain emotional maturity each of us must learn to develop two critical capacities: the ability to live with uncertainty and the ability to delay immediate gratification in favor of long range goals.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I accept all my feelings as forming a part of myself.

From now on I will try to live all my feelings.

It's good to express my feelings, and from now on I will allow myself to do so.

I like myself when I express my feelings.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I have the right to express my feelings.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

Table of contents

I am in control of my feelings.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I can assume responsibility for my own feelings.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I can control my feelings.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I know how to communicate my feelings to other people. I understand them and they understand me. I am not afraid of approaching people and telling them what I feel.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

If there's no feeling in it I don't want it any more.

Louis Proto, Self Healing, Piatkus Publications, 1990.

FEELINGS - NEGATIVE

I refuse to allow images of disaster and of a terrible future to take root in my mind and dominate my personality with feelings of failure.

I refuse to allow distorted, rambling thoughts to occupy any space in my mind.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Trying to rise above my negative feelings means that I am making progress.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

FEELINGS - POSITIVE

I am working to continue my past successes and create entirely

Table of contents

new ones. As a constructive observer of the mind, I make sure my feelings of success are constantly growing in my imagination.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

FERMENTATION (Intestinal)

Under my hand my intestine is contracting and pushing the gas upwards.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

FEUDING

I want my problems with —— (name the person) to stop.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

To cut the thread of my negative thoughts towards someone I cut the knot attached to my own mind.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

Your inner Guide is allied with my own. Together they can bring us peace. I understand you, and you understand me. Our inner Guide is our only guide.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

FEVER

My temperature is going down.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

Table of contents

FIBROMA

The contractions of my uterine muscles strangle the blood vessels which nourish my fibroma; lacking blood, it atrophies and disappears.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

FIGHTING SPIRIT

I know that I have the fighting spirit to attain my objectives; that means I try, and when I try I succeed. If I fail I try again.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I develop the habit of reacting positively and with a fighting spirit to threats and problems. I do this by adopting a positive, fighting attitude towards all day to day problems, as well as the problems in my mind.

Dr. Maxwell Maltz, Psychocybernetics, Godefroy Publications, 1978.

FILING

I equip myself with the necessary material (office furniture) to set up an efficient filing system.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

FINANCE

I find it easy to organize my finances.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

[I attain my financial objectives.]

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

My finances are prospering more every day.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

FIRMNESS

I can achieve goals in this situation with firmness. I don't need to be this angry. I can assert myself and be firm without being so angry or demanding.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I have the right to say no to someone without losing that person's love or respect.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am learning to say no.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I never cave in to pressures or demands imposed by others... except if I have personal reasons for doing so.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I know how to say 'No' without being aggressive or timid.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I want to maintain a firm and dignified attitude, without becoming arrogant or cynical, without violence and without fear.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

Other people can say 'No' to me without my being affected.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

Table of contents

FLAB

I am aware of the weight I've put on around the waist and stomach... my waist is going to get a little smaller every day... and my kidneys will purge my body of fat.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

FORCE OF NATURE

No force of nature is bad - it all depends on how it is used.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

FOREIGN LANGUAGES

Studying this language gives me more pleasure every day... I await each class with impatience. I concentrate exclusively on learning this new language... nothing can distract me. Every new word I learn is engraved forever in my subconscious... I don't forget anything that I've learned.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

I learn the —— (specify) language easily. My memory absorbs new words more and more easily. I understand the language better and better, and I'm also speaking much more fluently. My memory for this new language is improving day by day. Learning —— is easy for me.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Table of contents

FORGIVENESS (Also see Resentment)

It is over, I can forgive myself.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

Every night I forgive all my enemies for everything.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I forgive my father with a sincere heart. He acted as he thought best. I accept what he did. I wish him peace, joy and harmony. I am sincere. I say what I think.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I forgive all those who have offended me.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I forgive others, they are only human. I forgive myself, I'm human also. If I harbor resentment I may be able to hurt other people, but I am sure to hurt myself with the poison my resentment builds up in me. A person who can't forgive is not really alive.

When I forgive others I take a giant step towards maximum relaxation; and when I forgive myself, I attain it.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I forgive —— (name the persons who offended you). I will radiate love, peace, benevolence and all good things under heaven towards them. They are free and I am free. I know I am free of them because I can think about them without any pain, but rather with a gentle feeling of peace and a blessing that springs from my heart.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

I am willing to forgive anyone who ever hurt me, and wish them well.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Table of contents

I fully and freely forgive you. I loose you and let you go. I do not wish to hurt you. I am free and you are free and all is well again between us. I feel love toward you and I glory in your success, prosperity and complete good.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I now forgive you and liberate you all.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I may not be able to overcome my human weakness and actually love my enemies, but for the sake of my own health and happiness I forgive them and forget about them.

Please open your heart to me. Understand, accept and forgive.

I wish this hadn't happened, but I was trying to meet my needs.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

FORESIGHT

I never undertake something without having a contingency solution, just in case.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I consider what I will do in case of success as well as in case of momentary failure.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

FORTUNE

My happiness? My fortune? They are within me.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

FOUNTAIN OF YOUTH

I feel and think as if I had a magic charm, as if I were much younger than I am. I NEVER consider myself older than twenty-eight to thirty-five years old.

Franck Rudolph Young, Cyclomancia, SIP, 1966.

FRANKNESS (Candor)

I am always frank about everything I do. I never lie out of self interest, negligence or even for amusement. Everything my conscience finds reprehensible I refrain from doing.

Professor Robert Tocquet, The powers Of The Will, Godefroy Publications, 1989.

I force myself to play the game frankly and openly.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

If I realize I've made a mistake I admit it; I try to learn from the experience in order to change my line of action.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

FRATERNITY

My activities and my life, in fact my entire being finds meaning in the spirit of human generosity and fraternity with all people.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

FREEDOM

I am not resistant to new truths. They can free me. I am not bound

Table of contents

to a sad past except by my own mental chains. I break them apart. I can be totally free of negative feelings linked to the past. And this freedom makes everything in the present seem fresh and new.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I plan ahead and free my energy in order to benefit others. I breath in goodness, which liberates me.

I breath the desire to give into my heart, instead of just taking all the time, and I feel free and full of power.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I know that there are infinite forces at work within me. I want to free them, and make them grow constantly.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

(I am limitless. I know no boundaries as I live unconditional love and creativity. I recognize my true identity.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I use my power to choose wisely.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I choose what happens in my life instead of leaving things to chance.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am now at peace with my past. It's finished, and I am free!

From now on I will forgive everyone whom I come in contact with. We are all happy and free. I don't have to try and please others. I am worthy of being loved and appreciated, whatever I do!

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am free to act as I wish.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Editions, 1982.

I am free to live in any way I choose.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I want to be free; free of myself as well as of others.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

When my mind is really free I won't be imprisoned by my emotions. A man who is freed from prison doesn't have to think about freedom or imprisonment any longer.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

In harmony with the whole universe I find my individual freedom.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

FREE TIME

I never miss an opportunity to do nothing, and give myself the time to do so regularly.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

FRIENDSHIP

I need friends, I need to be surrounded with people who will help me as I help them.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I like others and I like myself. I find it easy to make new friends.

Vera Pfeiffer, Positive Thinking, Element Books, 1989.

I create close links of friendship around me, in which the principle attractions are not problems or needs, but values and goals.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Table of contents

I cultivate friendships which are important to me. I allow myself to have confidential relationships with certain people.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

I keep making new friends.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

I make friends easily, and I have a lot of respect for them. My friends can count on me.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I only have friends.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

I never neglect my friendship with others.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am funny and adorable, and I have some very, very good friends.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I am surrounded by friends.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Publications, 1982.

I appeal to You, my Inner Guide, to give me the power to radiate love and sympathy all around me, so that I can attract others to me. Help me find those with whom I can forge real and lasting friendships. Let these people cross my path, and help us recognize immediately that we are on the same wavelength, that we belong to each other. I thank You for your help and guidance.

K.O. Schmidt, Success At Your Doorstep, Astra Publications, 1986.

FRIGIDITY

I have gained a better understanding of my problem. I know that

Table of contents

I am not really frigid, that I am a warm person and that the sex act is an expression of love.

I am not frigid and I can achieve an orgasm.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I love my body and I have the right to be sexually satisfied. I find it easier and easier to have an orgasm. I love orgasms, and abandon myself completely, with complete confidence. I am becoming a sexually fulfilled woman, vibrant with sensuality. I find it easy to attain orgasm, and to satisfy my partner. I love my body, and my partner's body, and I am becoming a more and more desirable woman.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I like my body and I like myself. I reach orgasm with ease. I awaken my partner's desire and I know how to satisfy him. I am a sensual woman and I can abandon myself to my desire.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I am a normal woman who functions normally.

I can be perfectly relaxed and enjoy a sexual encounter.

I know that having an orgasm is completely normal.

I am a healthy woman; I relax completely and open myself to my partner.

I accept my orgasm without forcing it or wanting it too much.

I am convinced that making love is a beautiful and pleasant experience.

I see myself as a woman who can always have an orgasm.

My body and mind are free of past sexual failures.

I am a woman, a feminine being, and I love to make love with my companion.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

Table of contents

FRUGALITY

Whether I'm out or at home I enjoy eating small portions.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I found health by living in moderation and eating frugally, simply in order to assure my survival.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I like leaving the table still feeling a little hungry.

Sublipower, Sublihealth, Edi Inter, 1989.

I will only eat healthy foods, and never to excess.

J.V. Cerney, Stay Younger, Live Longer Through The Magic Of Mental Self-Conditioning, Parker Publishing Company Inc., 1968.

When I sit down to eat I always tell myself: "I live better when I eat less."

I like eating less. Although I am always well nourished, I really like eating small portions and staying a bit hungry.

I never eat more than I should. I know how much food I need, and I'm proud of myself because I eat well, I am reasonable, and I've never looked better!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

FRUSTRATION

I stop brooding about the best way to do things and simply start doing them.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I remind myself that frustration is a waste of time.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

I can always reap an advantage of some kind from a negative experience. I no longer react to frustration as I did at one time, for I know that I can profit from whatever happens, good or bad, in one way or another.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

FULFILLMENT

I am burning to discover the best part of myself, that confident, companionable and responsible person I sometimes am. I avidly search for occasions to fulfill myself and refuse to get frustrated and boxed into a dead end.

I never neglect the responsibility I have to myself to open the doors of my real personality and live my life to the fullest.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

[I am excited to be alive. I am glad to be here. I am a bright shiny star that glows every day and night.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

In future everything I say will tend to magnify my life. I want to live a noble and full life, spreading good to those around me, in an atmosphere of clarity and serene joy.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I want to live fully.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

I feel fulfilled.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

Table of contents

FUTURE

I try to be as conscious and as positive as possible when making decisions that affect my life and my future.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

This is a great opportunity. I am constructive and positive, and I am starting to work on my future.

Vera Pfeiffer, Positive Thinking, Element Books, 1989.

Table of contents

GAMBLING (Games Of Chance)

I am incredibly smart. I know which (card, horse, number, etc.) is going to win.

Win, win, win.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

I never play.

NEVER.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

GASTRITIS (Inflammatory)

Blood flows into my stomach and intestines, repairing my mucous membranes, which are red and inflamed; my white blood cells close up the sores, and my mucous membrane is healed.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

GENEROSITY

[I am gentle and generous with the people I like.]

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am a generous and loving person.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

Table of contents

The more I share my possessions the more is given to me in return.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

GIFTS (Talents)

Every day I develop the gifts that have been bestowed upon me. I cultivate them assiduously, and with more and more noticeable success. Once again I let my inner Guide act through and for me. His will is my own.

I am united with my inner Guide. It activates my inner aptitudes, especially my gift for.... (specify). Every day this gift gets stronger.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

GOALS (Aims)

Every time I want to do something or change something about my life I write it down, I develop a plan for attaining my goal, and set myself a deadline. In this way all my goals are transformed into action.

Whenever I want to change something about myself, or get something in life, I set myself a goal, I write it down, I read it over every day, and I take the necessary steps to attain my objectives.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I find my goals and decide now to go after them (intense desire makes things happen!).

I write my goals down. (Writing goals has magic. I have a plan, rather than a wish or daydream. Writing is a commitment).

I read my goals every day (I reinforce my commitment, and in-

spire action).

I visualize myself reaching my goal (just after awakening and just before falling asleep).

I make an affirmation for my goal and affirm myself. (Affirmations are the fuel for motion. I affirm my objective(s) many times each day. I keep building my confidence. I repeat each day's words aloud 10 times, 3 times daily).

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I will not waste my strength on idle pleasure or useless work. All my actions will lead towards my goal.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I don't get all worried about results. I take pleasure in the journey towards my goal. I know exactly what I want.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I always set new and higher goals once the old ones have been attained.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I know what I want.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I know I am capable of attaining the goal I set for myself; therefore I demand continued and indefatigable action from myself, directed towards my goal.

I have a clear written description of the precise goals I set for myself, and I will not stop trying to attain them.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

I am strong. I am courageous. I pursue my choices with focus and confidence.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

Table of contents

Once I set goals for myself, I find it easy to attain them.

Sublipower, Subli Success, Edi Inter, 1989.

My desires lead to positive goals which direct my life.

Sublipower, Subli Success, Edi Inter, 1989.

My prayers are powerful, and I know they will help me attain my goal.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

One of the keys to my success is knowing precisely what I want to do.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

GOLF

My golf swing is improving day by day. I hit the ball with power and precision, no matter what the situation. My swing is becoming elegant and effective. I hit my irons as well as my woods. And I'm doing very well with my putter too. Golf is becoming easy for me. My score is improving all the time.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

GOOD HUMOR

At the center of every human being there lives an inner Guide. May my Guide awaken and transform the sullen side of my personality into a friendly, joyous and harmonious disposition.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I try to maintain my good humor and enthusiasm.
Every morning I try to wake up in a good mood, ready for action!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

GOODNESS

I try to see the good in others, but first I look for what is good in myself.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I put my best self out there.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I am good.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

If I want to find happiness I have to stop thinking about the gratitude or ingratitude of others, and give solely for the joy of giving.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

If man is good, the world is good.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

GRATITUDE

I now give thanks for all the good things I have and all that is still to come.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

If I want my children to be grateful I have to show them how myself, by being grateful to others.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I give thanks for having been allowed to discover the divine aspect of my mind and body, and for the help I received in settling my business affairs and personal relations.

Now I give thanks for this life full of health, wealth, happiness,

and perfect self expression.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

GREED

I have more material wealth than I need, because I am not greedy and desire only what I can use constructively in my life. It took a lot of effort to get where I am, and I have always observed the laws of nature by developing positive habits to which I adhere.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

GRIEF (Also see Forgiveness)

I free myself of my grief, and the incessant flow of negative thoughts.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

GRIEVANCES

I agree to let bygones be bygones, and bury my grievances in the past.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

GROWTH

I'm open to new ideas. The best in me surges forward, never stopping, never ending, filling me with joy and freedom. That is the law of growth working within me, leading me towards what is good.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

Table of contents

My world is opening and expanding.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

GROWTH - PHYSICAL

I have a good appetite, I am digesting my food better, my lungs are getting stronger, I'm becoming robust, and developing normally.

Emile Coue, Complete Works, Astra Publications, 1976.

GUILT (Also see Fear, Rejection)

Feeling guilty won't change the past and won't make me a better person.

I tell myself that the past can never be changed, no matter how unpleasant it may be. What's done is done. I can feel as guilty as I like, but that won't change anything.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I have the right not to feel responsible for other people's feelings.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I stop feeling needlessly guilty. I try to find the reason for a mistake, without blaming myself.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I give myself the right to make mistakes.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I pardon and absolve myself of all guilt, here, now and forever!

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I forgive all those who tried to make me feel guilty about sex and

my body.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I thank God every day for what life brings me. I an thankful for my luck and for my joys, and also for the obstacles and deceptions. I ask forgiveness for my prejudices against others, and thus free myself of guilt.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

Everybody makes mistakes, it's only human.

I accept the consequences of my actions without embarrassing myself by blaming myself for past errors.

However bad a decision I made, it doesn't matter. I accept myself as a person who, like everyone else, is just trying to survive.

I can learn from my mistakes without guilt or worry.

I am not responsible for others.

Everyone's awareness is different, so comparisons are worthless.

Matthew McKay & Patrick Fanning, Self Esteem, New Harbinger Publications, 1990.

I shake off all my erroneous thoughts, as well as those of the people around me. My mind is perfectly clear and free.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

HABIT

I reject my negative habits, those which don't do me any good; I cultivate my positive habits, those which work to help me.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I know that habits can be changed. I have already altered old habits and acquired new ones, whether in sales, bowling, or thinking. I have done it (1) when I became aware of the habit; (2) when I became convinced it was blocking me from reaching my goals; (3) when I worked at substituting a new, more useful habit.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I am completely free of this habit. Harmony and peace of mind reign over me.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

HABITS - NEGATIVE
(Also see Diet, Habit, Hygiene)

No negative habit can dominate or influence me. I am in control of myself and my actions. I always do what is best for me and for my future.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I have decided not to have any unhealthy habits.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

I do not identify with undesirable habits. They are only acquired traits, and are not part of my real Self.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am free of this habit, I have peace of mind.

Dr. Joseph Murphy, The Miracles Of Your Mind, Dangles Press, 1984.

My bad habits disappear easily and are replaced by positive ones, especially concerning my diet and hygiene. And this naturally contributes to improving my general state of physical health.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

It is easy for me to replace one habit with another.

Sublipower, Sublicontrol, Edi Inter, 1989.

HABITS - POSITIVE

I will develop good habits and become their slave.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I have good working and living habits.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am going to develop habits that are going to make me think better and feel better.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

HAIR

I can improve the condition of my hair... help it grow back... through mental action... I feel my hair revitalizing... getting stronger and more vigorous every day.

I'm not afraid my hair will fall out... losing my hair is no longer a

Table of contents

problem... because I know I can keep my hair.

I think about caring for my hair every day... helping to give it strength and vitality.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

HAIR LOSS

I'm not at all worried about going bald one day... I know I will keep my hair... I don't think about losing it any more... on the contrary I'm certain I have the power to keep my hair... I am free of the fear of losing my hair... I actually feel it getting stronger and thicker.... growing more and more...

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

HALITOSIS

My breath smells fresh.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

HANDICAP

I can be greater than anything that happens to me.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

HAPPINESS
(Also see Love of Self, Confidence, Intimacy)

None of my past sufferings, nothing unpleasant or tragic that

happened to me will enter my mind. I will only allow positive thoughts, thoughts of joy and of the friends with whom I shared happy times, thoughts of peace and success, to enter my mind today.

I will not allow anything to disturb my happiness: I have the right not only to exist today, but to really live.

I don't want to be unhappy today, whatever happens. I want to live and rejoice to the fullest on this day. I won't let anything make me angry.

Orison Swett Marden, The Joy Of Living, J.H. Jeheber Publications.

Today is the best day ever.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

God wants me to be happy.

The happier I am, the happier the people around me are.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I assert my permanent good health, my happiness, the perfect creative expression of myself.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I have the right to be happy. I work on getting better at being happy.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

I have the right to try to be as happy as possible as long as I respect others.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

I always have what I need to be happy here and now.

I am rich, I feel good about myself, I am happy.

I do, however, also accept responsibility for trying to make the world an oasis of happiness and abundance for everyone.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

I believe that happiness comes from the ability to enjoy life and an unselfish desire to comfort and help others.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I make an effort to be happy.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I deserve to be happy, healthy and wealthy. My mind's natural state is happiness.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I am immersed in happiness.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I think of happiness as a state of inner freedom.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I know how to savor each instant of happiness. I live intensely, in the present moment.

Christian Godefroy, 1987.

I will be happy.

Og Mandino, The Greatest Secret In The World, Different World Press, 1979.

I am happy to be alive and to be building my success.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

The spirit in me is the artisan of my happiness.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

My mind and my soul are a haven of happiness, joy and confidence.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

To be happy I make others happy!
I have found happiness by helping others find it.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Table of contents

For me, true happiness never comes alone...

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

HARMONY

I accept the divine balance between male and female. I experience this male / female relationship with love and understanding. I allow this relationship to expand and strengthen me every day.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I am in perfect harmony with the universe.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

HARMONY - CONJUGAL

Each day I shall remind myself to be less sensitive. Love frustrations are often the result of over sensitivity. We all know couples who truly love each other, yet constantly bicker over trivialities, each oversensitive and on the defensive. Their love is like a duet, played on a discordant note.

I must give my wife companionship.

I must not expect my wife to conform to my way of thinking about everything.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I like being married and I take advantage of the numerous benefits it offers.

My marriage is a success because I do my part. I don't take my success for granted, and I make a conscious effort to make my marriage a success, and help it grow.

My marriage is proof of the effectiveness of teamwork. I am never

alone in life, because I am part of a loving, attentive and successful team.

I respect my marriage and my partner, and I always try to resolve our differences in as positive a way as possible.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I begin and end each day with a declaration of love for my spouse, and during the day, if possible, I devote three minutes of telephone time just to chat with him / her and express my love.

I'm a good listener. I care about my spouse's life, about the thousand and one details and little incidents that occur every day. I remind myself that things done out of duty may be done well, but things done out of love are done with grace. I see how something begun as a duty can be transformed into a work of love. It's actually amazing how fascinating the 'little things' in life can be.

I do not force my spouse to compete with the children for my attention. I make sure to spend time alone with my spouse.

When we disagree about something I think that it is possible to have different opinions without arguing. However, we never go to sleep before resolving our differences.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I know and believe that there is a man out there who will love and take care of me. I also know that I can contribute to his happiness and peace. He will accept me the way I am. He won't want to re-educate me, and I don't want to be re-educated. It won't be necessary, because our relationship will be based on love, freedom and mutual respect.

I can picture very clearly the qualities I'd like to see in my husband. I think about these qualities, so that they also become my own.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I am you and you are me. We are one, nothing can separate us.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

Table of contents

Every day our relationship is becoming more and more harmonious. We are happy to live together, we compliment each other and our love for each other is growing.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

We always solve our differences with understanding and consideration.

I respect the unique character of my spouse. I admire the numerous qualities in her / him which makes this relationship unique.

I am honest and sincere, and I express my thoughts and opinions openly, while taking what my partner has to say into account.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Our relationship is going very well.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

HARMONY - FAMILIAL

(I invest in the well being of my children. And I must make my wife feel like a woman and not a mother substitute.)

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

HARVEST

I ask God to bless each furrow of ploughed earth. Every seed that is put into the seeder is blessed, and I have a clear image in my mind of the abundant harvest that will come later, God willing.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

Table of contents

HATE (Also see Love)

Whenever I feel an impulse of hate or anger arising in me I replace it with a loving thought and act as if I feel love instead of aversion.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I celebrate a general amnesty, freeing all those who have hurt me. I wish them all peace, health and happiness, and send them my blessings.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

HEALING

My healing powers are developing and getting stronger.

Sublipower, Sublihealth, Edi Inter, 1989.

What the mind causes the mind can cure.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

My thoughts are becoming more positive every day, healing me, assuring my success and bringing me happiness.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

I love my body. I have the power to heal myself.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I believe and know that divine love is in the process of healing me.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I direct all the energy in me towards God and His powers of healing. And the seed of this attitude sprouts and grows in the fertile soil

of my subconscious mind.

I know that the infinite power of healing is healing me right now. It can heal me despite any logical proofs to the contrary. I feel it, I believe it, and I'm ready to attain my objective, which is my complete recovery.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I will recover.

Emile Coue, Complete Works, Astra Publications, 1976.

I rejoice in this new body and give thanks.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I am getting better.

I want to get better.

Professor Robert Tocquet, The powers Of The Will, Godefroy Publications, 1989.¥

The love of God saturates my entire being. The peace that comes from God fills my soul, and I give thanks for the marvelous healing that is taking place in me at this very moment.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

The power of my subconscious heals my body and makes it perfectly healthy.

I have the power to heal myself.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

The doctors are amazed at my progress.

Louis Proto, Self Healing, Piatkus Publications, 1990.

My state if health is improving.

Dr. Joseph Murphy, The Miracles Of Your Mind, Dangles Press, 1984.

If one of my organs is functioning badly the natural power of healing will work on curing the anomaly day by day.

Table of contents

My healing powers are strong and protect me against all disease.
Sublipower, Sublihealth, Edi Inter, 1989.

The power of suggestion accompanied by effective treatment will assure my recovery.
R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Every day, and in all ways, I am feeling better and better.
Emile Coue, Complete Works, Astra Publications, 1976.

Everything is fine.
Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

HEALTH - GENERAL (Also see Beauty, Vitality, Confidence, Happiness, Looking Good, Slimness, Charm, Balance, Harmony, Self Expression, Peace, Success, Prosperity, Love of Life, Energy)

As my essential functions of digestion, circulation, breathing and sleep get better and better, so does my general health.
Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

From today on I will develop my body and mind. I will remain in good health. I am solid and strong because I have decided to be so.
Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

Calm and relaxed, I regain my strength and health.
R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Every cell of my body breathes harmony and health. My whole body radiates health. I feel wonderful. I'm in great shape. The way I

Table of contents

look and move expresses health and harmony. I radiate well being.

Every day my positive thoughts keep me healthy. My thoughts create harmony in my body and in my relations with people. These harmonious thoughts give me strength and health.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

Every day I get more beautiful and healthy!

Everything I do promotes my health and beauty.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

The prototype of my body in all its perfection can be found in my subconscious mind.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I can feel my body more healthy now. I visualize, think and speak well of my health. I use positive self-talk on a regular basis.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I am acquiring good health. I make sure to keep all the parts of my body healthy.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I have an increasingly positive mental attitude in a perfectly healthy body.

I am constantly creating positive images and thoughts which keep me healthy and in top shape.

Christian Godefroy, Mental Dynamics Seminar, 1978.

I listen to my inner voice which always tells me the truth about my physical condition. I always listen to its warnings and advice.

My mind takes good care of me. By giving it clear inner instructions, I consciously condition it to keep my body bursting with health, energy and vitality. My mind always obeys my commands.

I am very healthy, but I don't take my health for granted. I am

Table of contents

proud of the efforts I make to stay healthy.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I use the power of my mind to control my thoughts, and keep my body and mind healthy.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I will consult my dentist and physician for periodic checkups. I am going to devote a portion of each day to improving my health.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

(I feel good, I look good, I am good.)

Louis Proto, Self Healing, Piatkus Publications, 1990.

I feel very healthy! I feel happy! I am in great shape!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

(I am a friend to my body. I always let my body guide me in perfect equilibrium. I have confidence in my body.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I feel good.
I am in perfect health.

George Barbarin, How To Overcome Fears And Anxiety, Dangles Press, 1974.

I feel fine... I'm healthy... I have confidence in my subconscious... which watches over me and keeps my sympathetic nervous system in balance... my sympathetic nervous system which regulates all my bodily functions... and which keeps me in perfect health...

Nothing can affect me... nothing can harm me... my health can withstand anything... my vital forces are inexhaustible... I am in perfect health... in great shape.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

Table of contents

I am now in perfect health.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I am very healthy. God is my health.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

I will always live with a positive frame of mind, and that will keep my body healthy.

Day by day my health is improving.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

My body's natural state is exuberant health! I see myself perfectly healthy, radiant and strong. I focus on health.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

The idea of perfect health now fills my mind.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

Every cell, every tissue of my body is regaining its health and harmony.

My health is improving day by day, I am glowing with health. My strength is growing and my health is improving daily.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

The life of God penetrates every fiber of my being. I AM well and whole in every part of my body. I am composed of billions of cells all containing intelligent life, and this intelligence guides me towards health, happiness and prosperity.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

My body is my friend.

The power of life is in me and regenerates me.

Sublipower, Sublihealth, Edi Inter, 1989.

My body is perfect.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

Table of contents

Everything about me is an affirmation of balance and health.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

HEART

I feel each heart beat. My heart is strong. It is doing its work perfectly. I can count on it. It is strong, healthy and vigorous. It is beating regularly, calmly. It is strong.

My heart is vigorous. The blood it circulates is pure.

My cardiovascular system is purified. My arteries are free of all obstructions. My arteries are healthy. I feel good. I feel fine.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

My heartbeat is slow and regular.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

My heart is beating slowly and regularly.

My heart is beating regularly, calmly, and strongly. It is working completely normally.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

My heart is strong again, and functioning normally.

Louis Proto, Self Healing, Piatkus Publications, 1990.

My heart is getting stronger and stronger, more and more calm, more and more rested. It is beating slowly and regularly, easily sending fresh blood full of oxygen to all parts of my body. My heart is getting stronger day by day, becoming more and more calm and powerful.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

My heart feels fine.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

Table of contents

I'm feeling better and better.
My heart is beating normally and regularly.
My heart is perfectly calm.
My heart is perfectly healthy.
My heart is doing fine.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

HEART ATTACK (See Tachycardia)

HELP (Also see Cooperation)

God is with me, God helps me, He cannot abandon me.

Marcel Baugin, Hypnotism, Suggestion and Telepsychic Phenomena, Scientific Distribution, 1956.

I accept the help others give me.

Subli Power, Sublicontrol, Edi Inter, 1989.

I have the courage to ask for help.

Shakti Gawain, Reflections In The Light, New world Library, 1978.

I make use of everything that can or could help me.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I am supported by life.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

May God give me the patience to accept what I cannot change, the courage to change what I can, and the wisdom to know the difference.

W. Clement Stone & Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978

Everybody loves me and supports me wherever I go.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Table of contents

HESITATION

As soon as I decide something... I do it, after having considered all the advantages... I do it right away... enthusiastically... and I resolutely eliminate all doubt and hesitation from my mind.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

HINDRANCES

I now get rid of all the feelings of guilt, fear, resentment, disappointment and hate I may have accumulated. I am free of all hindrances!

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I free myself of the hindrances that are no longer useful.

SubliPower, SubliControl, Edi Inter, 1989.

HOMEWORK (Also see Studies)

When I'm at home I concentrate on my homework. I finish all of it carefully. I don't let myself get distracted. I'm not going to be tempted to do anything else until my homework is completed and my courses for tomorrow are prepared.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

HOMOSEXUALITY

I am a man.
A boy's penis smells bad. It's horrible, disgusting.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

Table of contents

HOME

I find my dream home. It's exactly what I've been looking for, and corresponds to all my needs. My subconscious is an infallible guide.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

HONESTY (Also see Frankness, Sincerity)

Honesty demands that I rid myself of the need to evaluate my own worth in terms of exterior values, instead of looking at myself as objectively as possible, with an eye towards an even better life than the one I am living today - in other words, instead of being more or less faithful to 'the way it's always been.'

Being honest with myself requires that I realistically evaluate my strong points and my weak points so that I can identify and separate all the defence mechanisms I've built up around me from my valid inner messages, and eliminate them.

I can decide to continue listening to false inner signals, trying to hide my real self from others, but even if I do, from now on I will be totally honest with myself.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

HONOR

I would be ashamed not to live up to my responsibilities and delude myself; therefore I must live up to my responsibilities to others, and not try to delude them.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

Table of contents

HOPE

I build my hope, and never destroy it.
I choose to believe that something good can happen.

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

Everything is tranquil around me. I feel good. I am confident. My life unfolds under the sign of hope. I am proud of what I have accomplished up to now. And I will continue being proud. My life is calm, pleasant and full of hope.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

HUMILIATION

No one can humiliate or upset me. I won't allow it.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

Humor

I am going to convince myself that a sense of humor is an antidote to worry. It will help me develop a balanced personality, and is as essential to intelligent living as food is for survival.

I need to laugh for the maintenance of good health, for the relaxation of my mind and body. I am going to remind myself not to take either myself or my troubles too seriously. Laughter will help me make life more bearable. Laughter is a habit. I am capable of cultivating it by simply repeating each day that I will find something to laugh about.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I have a kindly sense of humor about myself.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

Table of contents

HURT

Only the False Me can hurt me.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

No one can hurt me unless I authorize them to do so.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I eliminate the hurt of a negative emotion by refusing to identify with it; this means that I don't see it as a permanent part of myself, but rather as a feeling that is just passing through me.

I try to put a definitive end to my inner hurt instead of seeking temporary solutions.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

HYGIENE (Also see Negative Habits, Dieting)

I give my body the best.

I love my body and I admire my beauty.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I will keep my body and mind clean because I want to live longer and stay young.

J.V. Cerney, Stay Younger, Live Longer Through The Magic Of Mental Self-Conditioning, Parker Publishing Company Inc., 1968.

I am good to my body and my body is good to me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am going to develop habits that are conducive to making me think and feel better. I am going to cultivate healthy living habits.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Table of contents

HYPERTENSION

My blood pressure is always normal, and my heart beats regularly.

Christian Godefroy, Mental Dynamics Seminar, Godefroy Publications, 1978.

HYPNOSIS

I am certain I can put anyone under, at any time.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

I am going to use the technique of self hypnosis to develop a will to improve, a desire to study, learn and make progress in every aspect of my life.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

My hypnotic powers are developing day by day.

Pr. Lawrence Powers, The Ten Secrets Of Hypnotic Power, Godefroy Publications, 1982.

I am influencing your mind. I am going to put you to sleep. You cannot resist me.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

My hypnotic power is developing day by day.

Dr. Lawrence Powers, Ten Secrets Of Hypnotic Power, Godefroy Publications, 1982.

Table of contents

IDEALS

I intend to become.... (specify).

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I will become great.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I can be one of the cogs in a great cause. I strive to attain my highest ideals, which always include excellent physical and mental health.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

My role is to try to acquire the intellectual and moral qualities of the person I wish to imitate. I therefore observe my ideal model closely. I am aware of the value of each word I speak.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

My ideal is my guide. It is easier and easier for me to attain the goals I set for myself.

Christian H. Godefroy, The Seven Eternal Laws of Success, Godefroy Publications, 1980.

IDEAS

I don't forget the ideas that enter my mind.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Table of contents

I make a conscious effort to think of new and better ideas, which are the subconscious keys to the vault of my mind.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I try to put an idea to practical use in my business or job, at home, and in my social relations.

I often review the immensely beneficial ideas I have concerning my Real Self and my False Self.

I take all these ideas and store them in my mental filing cabinet.

If an idea is really effective, it will prove itself sooner or later. I want to give my ideas a chance to prove their worth.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

The idea that will enable me to succeed is already in my mind. I just have to think to get it out.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I am part of the vital flow of divine ideas. I gather them and rejoice in them.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

The exact idea I need today will appear in my mind. Coming from the Divine Intelligence, it also contains all the subsidiary ideas I need. I allow this right idea to govern my thoughts.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

IDEAS - LUCRATIVE

I find the idea I need to double my revenue in one year.

My subconscious comes up with ideas for rapidly earning $50,000 (or the amount you need).

My subconscious comes up with ideas for rapidly earning

Table of contents

$100,000.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

My mind is always creating new and lucrative ideas.

Christian H. Godefroy, Prosperity: The Subliminal Method, Godefroy Publications, 1987.

IDEAS - NEGATIVE

I let go of any mental pattern which is creating negative effects in my life.

Louis Proto, Self Healing, Piatkus Publications, 1990.

IDEAS - PRECONCEIVED

I have a conversation with someone whom I've avoided up to now. I soon notice that my preconceived ideas, when re-evaluated, paralyze me and nourish my indifference.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

IDENTITY

I master the principles of perception. I learn to know myself. I discover my real identity.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I am —— (specify what kind) of person.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I am happy being me right now.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Table of contents

I am ME.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences the same way as I give , Godefroy Publications, 1989.

IGNORANCE (Also see Sickness)

As I give myself the right to make mistakes, I also give myself the right to be ignorant.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Divine force is manifesting itself in me, I am positive, positive, positive. I have the power to destroy sickness and ignorance.

Richard Ingalese, The Power Of The Mind, Dangles Publications, 1981.

ILLNESS (Also see Pain, Ignorance)

In the final analysis I must love in order not to fall ill.

I have learned the lesson(s) I needed to learn and no longer need to be ill.

I intend to enjoy myself in spite of my illness, rather than brooding over it and bringing myself down with worry.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I have now transcended all preconceived ideas about illness. I am free and healthy!

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I forget about illness.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I feel better.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

Table of contents

I know I can resist all diseases... whatever they are... I cannot be sick... I feel full of vitality... no discomfort whatsoever... no illness can catch my subconscious off guard.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

Illness begins in the mind. I will never allow my mind to fall prey to thoughts which are likely to result in disease.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

You couldn't get the better of me, you will never get the better of me again.

Emile Coue, Complete Works, Astra Publications, 1976.

I am in good health, and my body no longer has any trace of disease. I look in the mirror and I feel happy. My skin is clear, my eyes shine with health. I feel good. I have energy. Life interests me. Everything is fine. My body is healthy, perfectly healthy.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

ILLUSION

If I accept myself I choose reality and its accompanying roles: happiness, confidence and success. I don't have to live in a world of illusion any longer, because I'm very happy to be who I am.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I get rid of outdated attitudes and acquired convictions. My life will be free of illusions.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

IMAGINATION

I use my imagination in a constructive way.

My imagination is a friend, responsible for my past successes. I construct my present based on these positive images of the past, and make progress towards building a good and pleasant life for myself.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

The power of my imagination is my greatest strength.

Even if I haven't had a lot of success, what I have accomplished remains in my imagination, where it really counts.

I use my imagination to help establish goals for myself; it provides me with realistic objectives, unclouded by fantasies, which will make me a happier person.

I allow my imagination to work creatively for me and for others, by planning sensible and attainable goals.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

My imagination is a powerful tool for creation.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

My imagination always helps me find new ways to succeed... new and enriching paths to follow.

I have confidence in my imagination, which shows me the best ways to overcome obstacles.

My imagination is becoming more and more fertile. I am constantly discovering things I had no idea about... things that enrich and diversify my mind... so that I have a definite advantage over those whose horizons are limited.

I know I can count on my imagination.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

All the strokes of luck I have benefited from in life already ex-

isted in my imagination.

Imagination is my mind's workshop. It is where I convert my mental energy into action and wealth.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

IMMUNITY

(I am convinced that unconditional love is the most powerful stimulant for the immune system.)

Louis Proto, Self Healing, Piatkus Publications, 1990.

IMPASSE

When faced with an insurmountable obstacle I try to see things on a higher level.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

IMPATIENCE

I will no longer feel impatient or angry; on the contrary I will always be patient, always in control of myself. And from now on the things that used to bother or irritate me won't affect me any more.

Emile Coue, Complete Works, Astra Publications, 1976.

IMPOTENCE (Also see Premature Ejaculation)

Tonight when I am alone with my partner I will feel exactly as I do now. I will be completely confident, sure of myself, I'll feel vigor-

Table of contents

ous and healthy, exactly as I do now.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

It's easy for me to get a strong and long lasting erection. And I have no trouble satisfying my sexual partner. I only ejaculate when I want to - I have perfect control. I like making love more and more, and I'm more and more skillful and loving with my partner. I am becoming more and more virile, and I satisfy my partner completely.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I have perfect control of my erections. I love to make love, and I know how to satisfy my partner. Day by day I am becoming more and more virile and loving.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I am a normal man, with normal functions.

I feel confident with the woman I have chosen as my sexual partner.

Just as I am able to stay erect when I masturbate, I can do the same when I am making love.

I am a normal man, with normal sexual habits.

I only think about my successful sexual encounters.

I know that my sexual functions are normal.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

IMPULSE - NEGATIVE

I want to control my physical and emotional impulses and refuse to act on those which run counter to goals, principles and which disturb my vital balance. Next time I'm going to think before I act. I am absolutely determined to control my impulses.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

Table of contents

I am always calm... and I control all my impulses. I never act before deliberating... but once I've made a decision I see it through.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

INCOME

I now have a satisfactory income of $—— per month.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I'm doing what I really like. I therefore excel at what I do and my income is constantly increasing.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

My abilities are improving and helping to increase my income.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

INDECISION

I am no longer undecided, I make no more mistakes or errors of judgment. My conscience is a perfect center of Intelligence of the universal Understanding which is God.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

INDEPENDENCE

As an independent and positive observer of the mind, I do my best to be a complete and independent person. I refuse to let others dominate and control my thoughts.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

I don't need to be accompanied by a friend to do what I want to do.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

I have the right to live my own life, as long as I don't harm myself or others.

I live my own life. It belongs to me.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I will never allow others to think for me. I must remain the master of my thoughts and actions. I am capable of steering my boat the way I want.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I am the sole creator of my universe and day and night I watch over my thoughts and feelings in order to guarantee my independence from others.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

INDEPENDENCE - FINANCIAL

I make a good living! I am heading for even greater success and financial independence.

I appreciate the financial independence which allows me to attain my financial objectives. The more precise and detailed my financial plan is, the more rapidly I will attain my objectives, and the more financially independent I will be.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I aim for financial independence and I make sure I don't owe anyone anything. If I have to beg for money I become its slave. In cases where I can't borrow money I find a creative way to earn it.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Table of contents

INDIGESTION

My digestive system is calming down.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

INDIVIDUALITY

I am a unique creature.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

INFARCTUS

After each session I feel the pain in my chest diminishing. I will not feel my heart laboring. It now beats regularly and much more strongly.

The fear I feel about my heart is disappearing, and I will stop thinking about it all the time. My heart is now healthy and strong enough, and I am recovering.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

INFERIORITY (See Complex)

INFLUENCE

I have a positive influence on other people.

Sublipower, Subliconfidence, Edi Inter, 1989.

I influence the people around me at will.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Table of contents

I will tell X that ——— (specify) and my words will have a powerful effect.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I am sure of myself, no one can resist my influence.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

I intimidate you, and you can do nothing to resist me.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

The fact of understanding another person gives me power and influence over that person.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

The acts and words of others have no power or authority over me.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

X! During our meeting on — (specify) your opinions will change when I tell you that —— (specify).

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

INFORMATION

I try to promote the exchange of a maximum of information, considering that other people are, in a sense, really strangers, and that I am also a stranger to myself.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I make sure I'm well informed about new developments in my field.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Other people (superiors, subordinates and colleagues) know what

I'm doing, as well as why and how.

I am well informed about everything that has to do with my work and with my business.

I make sure there are no misunderstandings about the information I circulate.

I consider that a part of my work consists of circulating information.

If I'm in doubt about whether or not to divulge a piece of information, I go ahead and do it.

I make sure to always reformulate the information I receive in order to verify that I've understood it correctly.

I do not hesitate to be redundant, i.e. to repeat the same information two or three times in slightly different forms, to make sure that I have been understood correctly.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

INNER DIALOG

I stop focusing my thoughts on the negative. I just let them go! Being realistic doesn't demand focusing on a problem continuously.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I turn all negative 'self-talk' into positive 'self-talk.'

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

INNER GUIDE

I learn to take charge of myself.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I learn to take orders from my inner headquarters - the Real Me.

Table of contents

It is always a sure guide.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I listen to my inner guide for everything that is good, honest and positive.

Sublipower, Sublicontrol, Edi Inter, 1989.

The law of love guides me.

The inner force which guides me shows me how I can best serve humanity. In my dealings with friends and clients I am conscious of the fact that God resides in them, and I work with them for the triumph of happiness, prosperity and peace.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

My true Self guides me in all my actions.

Divine love guides me and always takes care of me.

It is now my inner wisdom that is my guide.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

INNER PEACE

I often practice mental relaxation. When I want to relax completely, even if only for a moment, I use my imagination to project myself to a place where I know I will immediately feel inner peace.

I possess inner peace. I live in a positive and interesting way. I have a reason for being here. I like what I am, and that gives me even greater assurance and inner peace.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

The secret place where I find refuge is found deep in my own mind.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

Table of contents

INNER STRENGTH (See Strength - Inner)

INSOMNIA

Every night from the moment I want to go to sleep to the moment I want to wake up the next morning, I will fall into a deep, calm, tranquil sleep during which I will have no nightmares. When I awaken I will be completely refreshed and feel great.

Emile Coue, Complete Works, Astra Publications, 1976.

As soon as I go to bed I fall asleep.

As soon as I go to bed I fall into a deep sleep. I sleep all night long, and wake up refreshed and full of energy.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

Being physically and mentally relaxed will enable me to sleep well again... a deep, restorative sleep... This sleep will make me feel completely rested when I wake up.

I will never be nervous again... always relaxed... I know how to relax and fall asleep when I want to.

I will sleep deeply all night long... nothing can bother me. And in the morning when I wake up I'll feel happy, lucid, completely rested... and full of joyous and well channeled energy.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I always get enough sleep; my body makes sure to get all the rest that it needs!

I love sleeping. I get tired, and sleep comes. I am going to have a deep sleep.

My body craves sleep. I will sleep well right through the night, and I'll wake up happy and full of new energy!

Table of contents

I feel the peaceful calm of the night making me drowsy; I am completely relaxed; I am in a state of repose, peace and harmony. I am calm, I feel I am a part of the rhythm of the night... I am weary and relaxed... I feel free and in tune with the natural rhythm of nature, which sleeps at night... my thoughts are getting fuzzy... I'm so tired... I'm falling asleep... sleep.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I sleep deeply and calmly.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I sleep better and better.

Emile Coue, Complete Works, Astra Publications, 1976.

I fall asleep easily, I will sleep deeply all night long and wake up refreshed and full of energy and good humor.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I fall asleep easily. I savor my sleep. Every night when I go to bed I forget my worries, fears and anxieties. I am happy, relaxed, calm, and ready for a great night's sleep.

I don't wake up during the night. Everything is fine. Everything is calm. I sleep deeply right through to the morning. I am calm and I don't think about anything. I sleep deeply.

Everything around me is calm. I do not wake up. I sleep deeply. I am relaxed and happy. Everything is fine.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I couldn't care less about whether I'm going to sleep or not. Even if I stay up till tomorrow morning it wouldn't bother me.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I am calm... my muscles are loose... my nerves relaxed. I am calm...

and I am falling asleep.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

I am a normal person and I can fall asleep naturally.

The noise in my room helps me sleep.

The noise outside my room helps me sleep.

I can sleep the whole night through, without waking up, as I have done many times.

When I go to bed I leave my worries and preoccupations behind.

I am relaxed, free of all the negative influences that troubled my sleep in the past.

I don't have to try to sleep any more - I just let myself fall asleep.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I am going to use self hypnosis and practice relaxing my body and my mind. I am going to sleep soundly and awaken in the morning feeling refreshed and invigorated.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

A wonderful feeling of quiet and harmony spreads through my body... I feel happy... Every night when I go to bed this feeling overcomes me.

I fall asleep immediately... my sleep is healthy and natural... when I wake up I feel rested and refreshed... I can feel the peace and harmony spreading through me... I'll be happy when I wake up... day by day things are getting better and better.

Every night when I go to bed I eliminate all negative or annoying thoughts... I fall asleep right away... and sleep all night long until the next morning... When I awaken this marvelous feeling of peace and harmony is still there... I feel happy... day by day my condition is improving.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

INSOMNIA - CHILDREN

I love to go beddy-by because it feels so good.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

INSPIRATION (Also see Creativity, Optimism)

Every day ideas and creative inspiration come to me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Something good is waiting for me. And God sends me love. God is the light of my life, the source of all my knowledge and inspiration. God, who is at the heart of my being, knows all. He provides the nourishment for my thoughts, ideas which serve me well, and clear divine intelligence.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I am now guided by God. His love fills my soul. God inspires me and enlightens my life. I am full of goodwill and love for all. Every instant of my life I submit to the will of God.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I am optimistic and inspired.

Sublipower, Subliconfidence, Edi Inter, 1989.

The inspiration I need will come to me while I sleep. Tomorrow I will resolve all my problems easily.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

My subconscious helps me discover the solution I am looking for. The answer I seek comes to me with ease. What is the best way to —— (specify)?

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Table of contents

If I am stymied or at a dead end with a problem, I change the scenery and my mood. I try relaxing and reflecting by getting away for a day at the ocean or in the country. I remember that the right brain creative problem solver in my mind is always available for review when my left brain logic slows down.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

INTELLECT

I have an intellect, but I am not my intellect. It is fairly developed and active; it lacks discipline but it can be controlled and channeled; it is an instrument for gaining knowledge, both of the exterior world and the inner world, but it is not me. I have an intellect, but I am not my intellect.

Dr. Robert Assagioli, Psychosynthesis, Epi, 1965.

INTELLIGENCE (Also see Self Confidence)

The self image I choose depends only on me. I can choose to consider myself intelligent by applying my own criteria. For me, the happier I am the more intelligent I am.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I am intelligent.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

I am wise. I am intelligent and creative. I accept that a positive mental attitude is an important part of creating my reality.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I am an intelligent person.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

Table of contents

INTIMACY

I have the right to be intimate with someone.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I am capable of being intimate. I like having someone close to me. I am not afraid of intimacy. I'm ready to commit myself to an enriching relationship. I am happy to love someone.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I am completely happy. My intimate relations are deeply satisfying and enriching.

I look at my being and my intimacy with enthusiasm and hope. I believe that I deserve the best. I am worthy of the very best, and I obtain it.

I am aware of my most intimate feelings and desires. I possess a fabulous treasure - my magnificent intimacy with another human being.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

INTRANSIGENCE

I will never convince anyone by using force. The feelings this would create are very negative and the results are short lived.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

INTRODUCTIONS

I always greet the people I meet with a smile. When introducing myself in any new association I take the initiative to volunteer my own name first, clearly; I always extend my hand first, looking the

person in the eyes when I speak.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

INTUITION

By listening to my intuition I can hear the voice of God.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

I am intuitive... my intuition is the voice of my inner Guide. S/He leads me and I obey.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

My intuition is getting better and better... and this intuition comes from my subconscious, which protects me... which knows what I should do... to stay healthy... to be loved... to succeed in my affairs... to earn money.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I develop my intuition and heed its advice.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I recognize my own intuitive ability.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I know what I need to know each moment. My consciousness is forever expanding and deepening.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I am receptive to the counsel of my inner self.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

The Universal Understanding which is in me knows things that I don't. I let this great force of understanding and the power of action do what has to be done. All the wisdom, all the spirit and all the

intelligence of God are now working through me to accomplish my goals.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

The spirit in me knows.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

Intuition is not merely a hunch, but is an irresistible light which must be followed.

I never ignore my inner messages.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

My intuition is perfect, my perception of things is also perfect.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

INVESTMENT

I resolve to get my money working instead of me.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

INVOLVEMENT

I insist on my right to be a constructive member of the great human family.

No one can force me to isolate myself.

I live in the present, even in moments of tension, pain or grief.

I voluntarily strive to attain new and inspiring objectives.

I make sure I am open towards people.

I get involved in constructive projects.

Today is another day during which I can become wholeheartedly involved with other people.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

IRRITATION

I am starting to free myself from anger, resentment and irritation as I recognize that the causes of my anger are mostly in my mind, and disappear as soon as I modify my mental attitude.

If I get into a bad mood I sit or lie down, I close my eyes, I rest, breathing deeply a number of times and concentrating on an image of tranquility I repeat, 'I am at peace. I am at peace. I am calm, I am strong."

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I don't get irritated every time a problem arises.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

ISOLATION

I reject whatever can separate me from others.

If I neglect my self esteem I will find myself isolated.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am overcoming my feeling of isolation from others.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

JEALOUSY

I have complete confidence in my husband / wife.

I am no longer afraid that s/he will leave me.

I won't try to eavesdrop on his / her telephone conversations any more.

I know s/he loves me as much as I love him / her.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I love him / her as s/he loves me - unselfishly.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I am ready to allow Love to flow through me, working for the benefit of all. I free myself of all hate, all unpleasant thoughts, and all jealousy.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

May my confidence shine through all my actions and attitudes. I show my love for my spouse by doing special little things for him / her, and I do everything I can to fulfil his / her needs and desires. I don't ask for anything s/he is not willing to give voluntarily; I firmly believe in our love, and I see how the magic power of love and mutual confidence changes both of us for the better.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

Table of contents

JOBS (Also see Remuneration)

From now on I have a perfectly satisfying and well paying job.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I now have a good job.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I will find the ideal job.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I am keeping my job.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I easily find a well paid and pleasant job such as ——— (specify).

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

The ideal job for me does exist. But I have to start by believing it if I want to find it.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

Infinite Spirit, give me the possibility of having a job (or a house, or abundance, or anything else you desire) which pleases me, a job that keeps me busy, a job where I am needed, and which no one else can do as well as I. May I meet the right people so that I find a job where I can offer good service and get well paid for it. Lead me to where I can make the right contacts. I leave it up to You, and I know that all will be well.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I accept the inevitable.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

Table of contents

JOIE DE VIVRE (See Love Of Life)

JOY

I smile a lot. I feel happy inside, and I show it.

I'm a happy person by nature. I try to look at life with warmth and joy, and I believe that life is worth living.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Today is a great day. And I have everything I need. So I choose to live wonderfully, in joy and love!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Each day I prepare myself to be receptive to joy.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Every day my mind is filled with joy, love and peace.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I now choose harmonious thoughts and let the joy flow freely through me.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I let go of the past and now live totally in the present.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I think and act joyously, and so I am joyous.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I smile because I want to share my joy.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

I am as joyous as possible.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

(I am excited to be alive.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I am always joyous, active and energetic. Eternal joy is my strength. I am always in a good mood, free and full of happiness.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

Because my morale is getting stronger and stronger, I feel more and more in control of all situations, however difficult they are. Instead of all those negative and depressing thoughts, I will concentrate my mind on thoughts of health, courage, confidence in life and in myself, certitude and joy.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

JUDGMENT (Also see Action, Memory)

Moderation, common sense, good judgment and balance are all important elements of the way I live.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I channel my thoughts and take the time to exercise judgment concerning all my activities.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I never act contrary to my own better judgment.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

My thoughts are ordered and clear.
My judgment is rapid and sure, my memory is accurate.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

In order to clarify my judgment I try to understand other people's opinions about my behavior, my ideas and my emotions.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

Table of contents

JUSTICE

To be just, I act justly.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

JUSTIFICATION

I refuse to justify everything I do. If someone asks why I do such and such a thing, I reply that I have no reasonable explanation. I do what I do simply because I feel like it.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Table of contents

KLEPTOMANIA (Also see Honesty)

I no longer feel the need to steal. I am in perfect control of my reactions and my impulses. In stores I always pay for the merchandise I take. I can easily master my impulses. I take pleasure in spending money for objects I desire. I am a perfectly honest person, and happy about it.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

KNOW-HOW

I trust myself to acquire the knowledge and skills I need if this plan or approach does not lead to the results I want.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

KNOWLEDGE

I seek knowledge in order to make wiser choices.

My daily job is to broaden my knowledge.

I can do whatever I want, but what I want is determined by what I know.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I raise my consciousness, my level of understanding.

Table of contents

I raise my level of knowledge.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I make harmonious and confident use of the knowledge I acquire in my life.

Professor Kurt Tepperwein, How To Learn The Painless Way, Godefroy Publications, 1983.

I collect all the information I can.

I replace the effort of thinking with a tranquil understanding of everything that is going on.

I transform my acquired knowledge into personal knowledge. And then I experience the miracle of peace, every day.

The supreme secret for breaking out of the vicious circle of problems is knowledge.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I invest in my own knowledge. I enroll in an extension or seminar class in some aspect of personal or professional development. I make the bookstores and fitness centers my new haunts for 'happy hour.'

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

(Each person's knowledge is different, therefore no comparison is of any value.)

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

Table of contents

LACK

People give to me freely: I do not have to beg or fight.

Louis Proto, Self Healing, Piatkus Publications, 1990.

LAUGHTER

From now on I will cultivate laughter. I will smile, and my digestion will improve; I will rejoice and my burden will grow lighter; I will laugh, and my life will be prolonged, because that is the secret of long life, and it is now mine.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

My scale of values changes as my inner self develops. So now I laugh about things that caused me to suffer in the past.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

LAZINESS (Also see Cowardice, Lying, Disorder, Discipline)

Tomorrow I will wake up at — (specify) o'clock. I will feel rested and eager; I will not waste a minute lying around in bed; I will jump out of bed and joyously greet the day that awaits me. I will maintain my enthusiasm throughout the day, and I will work with joy and

presence of mind, certain that my efforts will lead to the desired results.

Starting today I will not allow myself to waste time on insignificant things, in order to prevent the bacteria of laziness from taking over my mind and my body. I walk straight and stand tall, not bent over like a crescent moon.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I am becoming a good worker, committed, courageous, and more and more frank and sincere.

Emile Coue, Complete Works, Astra Publications, 1976.

I will react against laziness and I will work with courage and regularity (dedication).

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

I'm starting to like what I do again.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I can now get up on time every morning, feeling great.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am rested and full of drive. I rejoice in the things I have to do. Today is a lucky day for me. I am calm and serene, and I can handle any situation!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

LEARNING (Also see Knowledge)

Learning is like depositing gold in the bank of your mind. I invest in myself every day. I can learn whatever I want.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

I love to learn. I find the time to learn new things, and constantly develop my mind.

Christian H. Godefroy, The Seven Eternal Laws of Success, Godefroy Publications, 1980.

I am constantly learning!

W. Clement Stone, Passport To Success, Godefroy Publications, 1962.

After you have gained general knowledge in a field or subject, concentrate on learning one aspect of it well. Succeed in specializing, before you diversify. Doing one thing well, until you have mastered it, brings confidence and a reputation for excellence.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I take great joy in learning.

I am completely relaxed, and assimilate the text perfectly.

I can easily remember everything I learn when I wish, and I am satisfied with my progress.

I will stick to the program I set up for myself right to the end, and I will not let myself get distracted.

I am very happy to learn so quickly and easily in this way, and I feel great.

I listen enthusiastically to the information I want to acquire, and assimilate it as if it were the easiest thing in the world to do.

I always concentrate on what I am doing, at the moment I'm doing it.

Learning brings me more joy every day.

I feel free of all obstacles, and learn with enthusiasm.

Each time I make progress I get more pleasure from my work.

Each day brings me a greater and greater understanding of the deeper meaning of what I learn.

Professor Kurt Tepperwein, How To Learn The Painless Way, Godefroy Publications, 1983.

I believe I can learn anything.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

Table of contents

LEARNING WHILE ASLEEP

Soon I will be deeply asleep. And yet I will easily be able to hear what I taped, and memorize it without waking up. I will learn the text on the tape easily, while I am completely calm and relaxed.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

In a few minutes, when I fall asleep, and even if I sleep very deeply, I will be able to hear the tape recorder; it will be easy to learn and memorize everything I hear.

Valery Sanfo, Practical Guide To Self Hypnosis, de Vecchi Publications, 1987.

LEISURE

I engage in inspirational recreation. I get out the kites, grab the Frisbees, dust off the picnic basket, share a project with the children, attend little theatre productions, musicals, and concentrate most of my television and movie viewing on shows that warm my heart.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

My leisure time is full.

I am aware that I can take an active part in managing my leisure time, just as I manage my professional life, or anything else I want to do.

I don't consider leisure time as wasted, or just time to kill. On the contrary it is an opportunity to develop myself and to relax.

Every now and then I review how I make use of my leisure time, and plan various ways to use it better (theatre, vacations, etc.).

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Table of contents

LESIONS

If there are lesions in one or another of my organs, these will heal up day by day, and will rapidly disappear.

Emile Coue, Complete Works, Astra Publications, 1976.

LEVEL HEADEDNESS

I am always level headed, calm and serene. God's peace permeates my mind and my soul, my entire being. I am full of goodwill, and sincerely wish all people peace and happiness.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

LIFE

Life is an amusing game which I want to play.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I now dedicate my entire mind to the Divine Plan of creative life. I refuse to be a mediocre person achieving mediocre results in my life. God has given me the genius of the Spirit.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

Now is the best time to be alive.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

LIKABILITY

I am likable and affectionate, and I have a lot to share with others.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

I am likable.

Dr. Wayne W. Dyer, Your Erogenous Zones, Tchou Press, 1976.

I am always relaxed, never irritable, and therefore always smiling and likable.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

LIMITATION

I never allow the obtuse thoughts of others to limit me; instead I open myself to limitless possibilities.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

There is no limit to how good I can feel.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am limitless. I know no boundaries as I live unconditional love and creativity. I recognize my true identity.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I am free, nothing frightens me, nothing can limit me. The realm of the mind has no boundaries.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

(I am a limitless being.)

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

All thoughts of limitation and poverty are now erased from my mind.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

No task is impossible if I know and understand my limits.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

My only limit is my mental image.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

LISTENING

I listen to what other people are saying instead of preparing my answers, and I make sure they have finished expressing their point of view before responding.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I listen to everything the people I'm talking to have to say, not only a part of it. To me listening doesn't mean waiting for my turn to speak. When I listen, I really listen.

I am an attentive listener. I don't listen just for the sake of listening, but to understand.

I listen to other people's thoughts, but I also listen to my inner voice, the voice of my best friend and advisor. I listen to the advice I receive.

Every day I become more and more skilled in the art of positive listening. I like to use my listening skills every time I have an opportunity to do so, no matter who it is I'm talking to, what the occasion is, or where I am.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

LOGIC

I am a rational person.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

My thoughts are coherent and organized in an orderly way.

Sublipower, Sublimemory, Edi Inter, 1989.

Table of contents

LOOKING GOOD
(Also see Weight, Losing Weight, Health)

I am gradually reaching my ideal weight. Every day I am closer to having the figure I dream about. I'm looking better every day. My body is firm, and I feel light. Life is nourished by life: I feel like eating living nourishing food, rich in vitamins. I love the life-giving taste of fresh raw fruits and vegetables, which enhance my beauty and energy.

Each day I feel my body becoming more beautiful and I rejoice in the life flowing through me.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

When I take an honest look at myself in the mirror I like what I see. I'm proud of the efforts I make to look and feel good!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

(I know I have to look good as good as I can to please my husband / wife.)

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I vibrate with immense happiness, in a magnificent body sculpted by the immense power of my subconscious.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

LOSING WEIGHT
(Also see Strength, Body Fitness)

I'll soon reach my ideal weight. I feel happy, and my body is looking better and better every day. I don't have to deprive myself of anything. I like my body, I like myself. My subconscious mind is my most

Table of contents

powerful ally. Through the power of my subconscious I can completely master my eating habits. I am free and happy. As I lose weight I feel more fulfilled, I am more active, more considerate of others, and prouder of my image.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

Every time I start eating I rapidly lose my hunger. I will not eat any more foods that make me gain weight, and I won't eat between meals. My body will function on a strict minimum of food and, by perspiring, will eliminate all toxic substances.

Valery Sanfo, Practical Guide To Self Hypnosis, de Vecchi Publications, 1987.

I lose a couple of pounds every week, and I feel fine.

Professor Kurt Tepperwein, Secret Techniques of Hypnosis, Godefroy Publications, 1981.

From now on I will be slim, strong and in perfect condition, whatever I do.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I find it pleasant and easy to lose X number of pounds.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I love this new beautiful body my subconscious is sculpting day after day. I'm losing weight. I'm looking better. I feel good. Every breath I take is filled with health and energy. Every time I exhale I get rid of toxins and excess fat.

My fat is melting. I'm looking better and better every day. I have less and less desire for useless foods that do nothing for my health. I am free of all that. I feel light, so light. I'm in great shape. And I feel great.

I feel so much better being myself in a slim body, free of all those false cravings and superficial pleasures.

Charles Baudouin, Psychology And Practice of Auto Suggestion, Idegraph Publications, 1990.

Table of contents

I am gradually reaching my ideal weight, and can maintain it easily. My health is improving. My appearance is also getting better and better, and people like to be with me more. I'm looking better every day. I only eat healthy foods, which help me lose weight without feeling hungry. I am becoming more and more attractive and healthy, thanks to the power of my subconscious mind. I can easily reach my ideal weight, and stick to it with no effort. I like my body, and the new self image my subconscious is helping me create. I am in perfect control of my eating habits. I love my body, and I have the power to lose weight.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I have no trouble maintaining my ideal weight.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am always slim and healthy.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

Whenever I sit down to eat I never let anyone else influence me, tempt me, or discourage me in a negative way. I am learning to say no to food and yes to success.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Losing weight the self hypnosis way will definitely improve my appearance.

Losing weight the self hypnosis way will change my personality, giving me greater self confidence, a sense of pride, a feeling of accomplishment.

Losing weight the self hypnosis way will make me less tired and sluggish.

Losing weight the self hypnosis way will improve my sex life by making me more attractive.

Losing weight the self hypnosis way will prolong my life.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

LOST OBJECTS

You know all things; therefore You know where my ring is and You will reveal where it is to me.

Dr. Joseph Murphy, The Miracles Of Your Mind, Dangles Press, 1984.

LOVE (Also see Success, Happiness)

To love is to look beyond other people's faults and failings. I concentrate on the good in other people.

I love, and I never hesitate to go that extra mile, or spend that extra time to show someone my love.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Every time my partner is affectionate, I forget all my worries, and feel full of love and tenderness.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

My life is an experience of love and warmth.

Christian H. Godefroy, Positive Thinking: The Subliminal Method, Godefroy Publications, 1987.

Giving and accepting love makes me strong. I always treat love as the most prized gift. The natural state of my heart is love.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I accept appreciation for my loving actions. As I give freely, I also receive freely. I appreciate myself as a male and female.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I greet this day with my heart full of love.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I love to love and be loved.

Table of contents

The more I love, the more love I have to give to others.

From now on I will love and receive love without any reservations.

All the love I seek is already in my heart.

The more I love myself, the more I love ———.

I love ——— and ——— loves me.

I now invoke the quality of love.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I love life, and life loves me in return.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I love.

Jean Chartier, How To Acquire A Superior Personality , Dangles Press, 1971.

I learn to think in terms of love in my mind, and I feel love in my heart.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I give and receive love freely.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I don't hate anyone, I don't envy anyone, but I love and respect the whole of humanity.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I could stop loving you, but for the time being I prefer to continue.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I radiate love through my thoughts, words and actions. Love brings harmony and happiness to my life, and to my relations with others.

Dr. Joseph Murphy, Triumph Over Yourself And Others, Le Jour Publications, 1988.

Table of contents

I am always in love with the universe, with life, and with myself.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I am a person worthy of love and affection.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I live in a cosmic universe, and I can hate no person, nation or race.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I live according to the law of love.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I love you.

Marcelle Auclair, The Book Of Happiness, Seuil Publications, 1959.

The emotions of love exists in me when I can be with other people and not want anything from them.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

The Spirit in me loves the Spirit in you.

Marcelle Auclair, The Book Of Happiness, Seuil Publications, 1959.

Harmony surrounds me. I listen with love to the pleasant and the good. I am a centre for love.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I fill myself with love and always expect the best.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

LOVE - FILIAL (Also see Creativity, Sensitivity, Optimism, Peace, Joy, Tenderness)

I love and respect my father because he is good for me.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

Table of contents

I love my mother, but I'm not responsible for her happiness.

Louis Proto, Self Healing, Piatkus Publications, 1990.

LOVE LIFE

I respect myself and assume the responsibility of being myself in my love life.

I am loving and caring both towards my partner and myself.

I don't consider marriage an institution where one partner dominates the other. I have faith in love, in the support and opinions of my partner, but I am the only one who is in control of my life.

I like the way my partner looks at me, and I know that this look is a reflection of my approval and appreciation for him / her.

I feel sincere admiration for my partner, and I show it. I often express my admiration and approval, both in public and at home.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

LOVE OF LIFE (Also see Prosperity, Happiness, Self Confidence, Courage, Peace, Love)

I am going to regain my zest for life.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Life is fantastic. That's right, fantastic.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

[I am going to make a positive effort to enjoy life, based on the principle that happiness is an attitude that can be cultivated.]

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I am excited to be alive. I'm glad to be here. I am a bright, shiny

star that glows every day and night.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I rejoice in life, and I am happy to be alive at this point in history. I contemplate the bounty and richness of life all around me.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I profit from life.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

I am full of life. I love life and I'm happy to be alive. I am a very special person, living an extraordinary life.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Life is fun and pleasurable.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

LOVE OF SELF
(Also see Self Confidence, Self Esteem)

(The more responsibilities I accept, the more energy and self love I have.)

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I love the body I was given. It is a part of me, and I love who I am.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I love and appreciate myself as I am.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I love myself without condition or judgment. I am an energy pattern like no other before or after me. I am comfortable in my individuality.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

Table of contents

I love myself the way I am.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I respect and love myself.

Sublipower, Sublisuccess, Edi Inter, 1989.

I feel warm and loving towards myself.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

LOYALTY

I invest my sincerity and loyalty in others. This brings me closer to people.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

LUCIDITY (Also see Self Control)

By acquiring a lucid and constructive mind I become my own judge; not a judge who condemns with impunity, but a compassionate judge who does everything to plead my cause.

The lucidity of my mind is my laboratory.

My lucidity helps me develop positive habits.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I think clearly and judiciously.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I want to be perfectly lucid under all circumstances, remain in control of myself whatever happens, and feel perfectly at ease with everyone I meet.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

Table of contents

My mind is always clear and lucid, and the results I obtain are the proof.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

LUCK (Also see Money)

I am lucky.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I am getting luckier and luckier.

Christian H. Godefroy, Positive Thinking: The Subliminal Method, Godefroy Publications, 1987.

I have the right to be lucky and rich.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I go through life smiling.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

I am a magnet for everything good.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

Love, that divine magnet, attracts everything that is good.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

Everything comes to me easily and effortlessly.

From now on everything will contribute to make my life beautiful.

Everything good comes to me easily and effortlessly.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

LUNG DISEASE

My lungs are getting stronger.

Sublipower, Sublihealth, Edi Inter, 1989.

Table of contents

My lungs are clean and strong. I can breathe deeply.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

LYING (Also see Cowardice, Laziness)

I always tell the truth.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Table of contents

MAGNETISM (Also see Influence)

I am accumulating great magnetic power.

I accumulate all ambient magnetic energy in my solar plexus.

My magnetism is becoming more radiant every day.

The magnetic power I emit makes my suggestions irresistible.

I can easily concentrate the magnetism in me, and project it where and how I want.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

I am projecting my magnetic force. I am saturating this person with the flow of my magnetic energy.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

The people who enter my life are attracted by my own personal level of magnetism.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I have magnetism because I am in control of my mind and body... this gives me great inner strength... this force... each day... I feel it developing in me.

I control and channel my thoughts at will... and this cohesion of my personality has a positive influence on those around me.

Such thoughts attract similar forces... of love and benevolence... which reinforce my personal magnetism.

I feel myself radiating a powerful magnetic force... which helps me succeed in life... which makes me attractive... to both men and

Table of contents

women... I always try to be friendly, hoping to benefit them in some way... with my magnetic force.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I now exert an irresistible attraction on people.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I project my personal magnetism.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

I feel attractive, bursting with all the power in my soul.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I feel more and more attractive, more and more powerful and magnetic. I have a positive influence on everyone around me. More and more people are attracted to me. Yes, day by day I feel my magnetism getting stronger.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

Other people love to have me around.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

My personality is becoming more and more magnetic. I can influence the people around me at will.

People are attracted to me more and more. I'm liked by everyone I meet.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I'm working on my facial expression: I want it to be frank, full of light and strength, ready to spread my power wherever necessary.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

No one can resist the power of my eyes.

Professor Lawrence Powers, The Ten Secrets Of Hypnotic Power, Godefroy Publications, 1982.

Table of contents

Nothing can resist my gaze.

Marcel Rouet, Techniques And Practise Of Hypnotism, Paris Productions N.O.E., 1974.

MAIL

I answer my mail, or throw it away, or delegate it every day, so that it doesn't pile up.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

The mail I send projects a positive image of my service and of myself.

Taking care of my correspondence doesn't take a lot of time.

I use empty blocks of time to take care of my mail.

I write creatively.

I am learning (through practice) to write clean copy on the first draft.

I am learning (through practice) to dictate a clean copy to my secretary on the first draft.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

MANAGEMENT

I build other winners too.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

MANAGEMENT - MONEY

I enjoy managing money carefully and successfully.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

Table of contents

I know how to manage my money. I have a healthy respect for what I earn and what I know, and I put them to the best possible use.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I know where I stand financially.

My spending is in line with my resources, with a sufficient margin for maneuvering.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

MANAGEMENT - TIME

I dedicate about twenty minutes a day to managing my time: ten minutes in the morning going over the list of things I have to do, and one minute per hour throughout the day.

I regularly ask myself what I can do to save time.

I allow myself sufficient time alone to think about how I'm organizing my time.

I know how to say 'no' and avoid time wasting activities.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I make a list of my daily tasks, in order of importance.

I know how important it is to remain in control of my time and my priorities.

Since I establish my own priorities and since I am aware of the importance of each of my actions, I give them exactly the amount of time and degree of attention they merit.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

(I manage my time well.)

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I take time for sunsets and flowers.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Table of contents

MARRIAGE

I am wanted by someone. I am happily married to a good and loving man / woman who is interested in things of the mind. I am secure.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

MATURITY

I am becoming more mature every day.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

MEDICAL TREATMENT

This treatment is totally successful and feels very comfortable. My pills are working wonders - and with no side effects.

Louis Proto, Self Healing, Piatkus Publications, 1990.

MEDITATION

I bow before the power that is in me.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I remain awake in a passive way.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

MEETING PEOPLE

I like to meet people and I focus on their good qualities. My goal

Table of contents

is to make each person glad they talked with me.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I take great pleasure in meeting other people, and I always see their good side. I am in no way unconsciously fearful of people, I know that the Spirit in all people is one of warmth and loving friendship. I give thanks for all the people I know, and for those I will meet in the future.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I approach people with heartfelt friendship.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I can't select all the people I'd like to work with and be with; but I control whom I spend most of my time with, and I can meet new people. I can change my environment and seek out successful role models to learn from and share with.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Now I meet exactly the type of people I want to know.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

The more I meet different people, the more opportunity I have to realize that my fears were all a waste of time. With this in mind the unknown becomes something that I want to explore instead of something I want to shy away from.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

You are you and I am me.
I am not in this world to meet your expectations.
You are not in this world to meet mine.
I am here to do my thing.
You are here to do your thing.
Sometimes we will meet - and that's beautiful.

Table of contents

Sometimes we won't. Too bad.

Louis Proto, Self Healing, Piatkus Publications, 1990.

MEETINGS

I always get to my meetings slightly ahead of time, and I make sure I never miss a plane or train.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Each time I am called upon to make a presentation to my department I prepare carefully and arrive at the meeting feeling completely sure that I know my material and that I will be in control of the meeting.

Each time I meet with our financial advisors I feel very much a part of the family. Being young only emphasizes my capabilities. Sitting with my colleagues I feel calm and serene, one member of a team all collaborating to achieve the same objectives. Because my mind is concerned with the decisions that we will soon have to make, I am oblivious to my body, and I am calm, concentrated and completely without tension.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, The 1978.

I prepare for my meetings.

I make sure that the person chairing the meeting does his or her job properly.

Only the people who are required to make the meeting more effective are allowed to participate.

The meetings I chair always result in explicit decisions being made and understood by all participants.

When I'm not chairing a meeting that I find is going badly I try to exert a positive influence by getting myself appointed secretary, or by suggesting that we take time out to evaluate the proceedings.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Table of contents

Whenever I am asked a point-blank question in a group meeting I am usually able to respond because I always pay attention to the discussion. I feel calm. I don't worry about coming up with the "right" answer because after all, I am the expert.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, 1978.

MELANCHOLY

I am never melancholy.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

When I am in one of my melancholy moods I will make sure not to project my disturbed feelings of gloomy disposition onto others.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

MEMORY
(Also see Concentration, Judgment, Willpower)

Starting today all aspects of my memory are improving. It is getting better and better. I retain everything I need to know more and more easily. I have confidence in my memory. It is more and more accurate. It is becoming more and more effective.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

What I record in my memory is written in indelible ink, and stays there forever.

Professor Kurt Tepperwein, How To Learn The Painless Way, Godefroy Publications, 1983.

Whenever I repeat the word X three times in a row my memory becomes perfectly active. It is ready to assimilate data, and I learn everything I have to with ease. Whenever I read or listen to some-

thing my mind will be more open and will store as much information as possible. My memory is improving, and it is much easier to concentrate on the subject at hand.

Valery Sanfo, Practical Guide To Self Hypnosis, de Vecchi Publications, 1987.

I have a good memory... a faultless memory... I can retain whatever I want.

My memory is powerful... and improving every day... as I organize my thoughts.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I like going to classes more and more as my memory improves.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

My memory is improving daily. My memory is getting better day by day, so that I can remember everything I need for my work.

All aspects of my memory are improving. From now on I can easily remember everything I have to, under any circumstances and in any place. My memory is more and more precise and faithful. I absorb all the information I need in my day to day life more and more easily. My memory is improving daily, and will soon be completely effective. I control my memory under any circumstances.

I congratulate myself. I really have a fantastic memory. I should be proud to have such a good memory. I can easily remember all the dates, events, statistics, thoughts, words, reading and work-related information I need to know. I really have an effective memory. Once again, bravo! I congratulate myself!

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I train my memory in order to make it more powerful, and I use it to advance in my chosen profession.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I imagine my memory as a fruit (any fruit). It is healthy, perfectly

smooth, free of imperfections. It is perfect. That is my memory.

I am relaxed. Everything is going well, everything is calm. My memory is improving day by day. I have a good memory, a faithful memory. I can count on my memory because it is becoming more and more accurate.

My memory is faithful. I can count on it. It is becoming more faithful every day. I am happy about the progress it is making.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I always remember my conversations with others clearly. I can repeat important parts word for word. This helps in my work, and in all other areas of my life.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am going to devote a part of each day to improving my memory.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I want to make my memory fast and accurate, my assimilation (of information) clear and correct, and my willpower faultless.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

My mind stores information with ease.

All I have to do is give myself a mental command to retrieve any information I want from my memory.

Every day I work on developing my concentration and memory.

My memory is getting better every day.

I have confidence in my memory, as it becomes more and more accurate.

My memory is getting stronger every day.

My thoughts transform into images, which are engraved on my mind.

I can easily recall any information I need, when I need it.

I have the power to remember everything I see, hear or read.

My memory is my friend, and I can count on it.

If I want to forget an event, a thought, a person or an object, my memory obeys.

The more I use my memory, the more accurate and faithful it becomes.

Sublipower, Sublimemory, Edi Inter, 1989.

I now have complete confidence in my memory. It is now fully effective. I can control it under any circumstances. It absorbs everything I need to know, and provides me with the information I need, when I need it.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

MENOPAUSE

I am not at all worried about my life as a woman... because I can stay young... stay young for a very long time.

I feel good... I am relaxed... I am not worried... I am resolutely optimistic... at this age where women reach their full potential... A great life lies ahead... full of promise... and I feel good, wonderfully good.

I am going to manage my menopause and make it another, even more enriching stage in my life as a woman.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

MENSTRUATION

I know that the menstrual cycle should occur regularly every twenty-eight days, and not every thirty days as many people believe. It will last for four days, no more, no less, and the flow will be neither

too abundant or too sparse, both before and during the period; I should not suffer from kidney or lower back pains, headaches or any other kind of pain, since this natural function should take its course naturally, without my having to suffer in any way at all.

Emile Coue, Complete Works, Astra Publications, 1976.

I am balanced and peaceful in all changes of cycles, and I bless my body with love.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

My uterus is tightening, emptying itself.
My uterus is becoming empty; I want it to be empty.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

MENTAL CONTROL

I take charge of my thoughts. I can do what I will with them.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I am responsible for the thoughts in my head.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

MENTAL FACULTIES (Also see Concentration)

My faculties are developing day by day.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

MENTAL FATIGUE

My thoughts are clear and precise, my mind is fresh and alert.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

Table of contents

I do not hesitate to take a break from my work if I'm tired... to relax my muscles... to breathe deeply and restore my energy... or to relax my brain by emptying my mind.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

MENTAL PROJECTION

I develop my faculties of concentration and mental projection. I learn to project my mind and all my senses in the desired direction, to any place and any level in the universe.

Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

I am learning to develop my five senses, my concentration, and to project my mind to any place and any level in the universe.

Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

MENTAL RETREAT

In my place of retreat I am free, and I am in control of my subconscious.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

MERIT

I merit the best, and from now on only the best will come to me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I was just talking about luck, but luck had nothing to do with it. I got that good grade because I deserved it.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Table of contents

I want to imitate the examples of those who succeed, of those who fear no obstacles. I want to merit and obtain success, joy, power and love.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

MIGRAINE (Also see Emotion)

I can relax at any time. My head feels cool... my head feels light... my neck and shoulders are supple and relaxed.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

In a few moments my mind will become clearer. The excess blood congesting my head is flowing back down to other parts of my body. Soon my headache will be gone.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I feel my head getting calm.
I feel my head relaxing, getting lighter.
I feel free of all mental stress.
I feel my head is normal, clear.
I am in control of my emotions.
I feel free of the past events which caused my migraines.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I will feel good today, my head will be clear and lucid.

Leslie M. Le Cron, Liberate The Forces In You Through Self Hypnosis, Tchou/Ariston, 1964.

I am going to free myself of migraines.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

My head is opening, becoming calm and light. Energy is once again circulating normally though my head, eliminating all fatigue and all tension. My blood is irrigating every part of my brain fully,

bathing it in a soft light. My head is completely relaxed.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

My forehead feels pleasantly cool.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

My headache has disappeared.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

Go away! Go away!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

MIND (Also see Memory, Concentration)

I use the vast power of my mind.
I concentrate my mind on one thing at a time.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I want to be mindful of everything because whatever I am aware of can inspire new ideas and opportunities.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

All the functions of my mind, like memory, attention, etc. are improving.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

Like an eagle my mind was created to rise to great heights.
I rely on my aptitude for developing maximum mental power.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I affirm the existence of the Real Me.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I will never use the power of my mind to do harm to others.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

Table of contents

I rejoice in this Life which exists for me. I am an incarnation of the Mind of all life.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I AM everything of substance that is.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

[I am going to think with my mind and not my emotions. I am learning to think love in my mind, so that I will feel love in my heart.]

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

My mind is a divine center of operations.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

Everything is the Mind. When the Mind is well, everything is well.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

MIND PROJECTION

I want X (specify who) to be forced to reveal —— (specify what).

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

MIRACLES

The light in me is accomplishing miracles in my life, here and now.

The light in me is accomplishing miracles in my body, in my mind, and in my affairs, here and now.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am the greatest miracle in nature.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

Table of contents

Divine love is working in me here and now in order to create something.

The Christ in me is accomplishing miracles in my life, here and now.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

My own words create miracles in my life.

I expect a miracle.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

MISOGYNY

I will make an attempt to understand and appreciate feminine psychology so as to better comprehend the underlying motivations for my wife's behavior.

I cannot expect love and respect from my wife if I belittle or abuse her.

I must also remind myself that little courtesies, tender attentions and acts of kindness are as important to her as material gifts.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

MISTAKES

I collect my mistakes with pleasure. I consider them as tools for learning.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I have the right to make a mistake.

To err is human. I am a human being.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

Table of contents

I nourish my conviction to rise above my mistakes.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I learn from my mistakes.

Sublipower, Sublisuccess, Edi Inter, 1989.

(My father thought that it is important not to make mistakes. But I have to learn how to do this work. I can only learn by trying. If I worry about mistakes I will stagnate, I won't learn anything.)

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I know that I am free to make mistakes, it is fine to make mistakes. I do the best I can. I always do the best I can within the limits of my awareness. Within the limits of what I know at any given time I do the best I can. I love myself, mistakes and all. I am a good person. I learn from my mistakes, and I am free to make them.

I am free to make mistakes. I can learn from my mistakes without guilt or worry. The past is over and there is nothing I can do about it.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

When I make a mistake or get ridiculed or rejected, I look at mistakes as learning experiences, and ridicule as ignorance. After a rejection, I take a look at my BAG - Blessings, Accomplishments, Goals. I look at rejection as part of one performance, not as a turndown of the performer.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Here's a mistake I promise myself I'll never make again, with the help of God!

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

Table of contents

MISTRUST

I believe everything —— (name the person) tells me about his / her activities.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

MISUNDERSTANDING

I make sure not to mistake a misunderstanding for a real disagreement.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

MODERATION (Also see Discernment)

I am happy to possess this or that object. I promise to use it wisely and in moderation, so that I get as much pleasure as I can from it.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

When I reprimand someone or express my unhappiness I try to do so after the urge to fight or become upset has subsided. The best way to get my feelings across is when I can speak in a normal voice, without all of the warlike body language. When I am upset I try a substitute physical exercise such as running, tennis, racquetball, or handball, in which there is impact involved to release the built-up adrenaline in my system. I do speak my mind, but I criticize the behavior without attacking the other person.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Everything has its just measure, and I respect it.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Table of contents

I am going to observe moderation in everything I do.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

MONEY (Also see Business, Luck, Competence, Creation, Projects, Debts, Vocations)

From now on I have more money at my disposal than I need.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I earn more money than I spend.

Christian Godefroy Collection, Magic Formulas For Being Happier And More Successful, Godefroy Publications.

I manage my affairs wisely and confidently. I earn enough money to satisfy my needs.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I don't dislike money... it gives me security... and I know that I can make money if I want to.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

I can do anything I want to extremely well, including making money.

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

I am receiving plenty of money for doing exactly what I want to do.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I am mentally and spiritually connected to the money I gave ——— (name) and it will come back to me multiplied and magnified in the Divine Scheme.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

Table of contents

I bless you so that you can bless others in turn. May you enrich everyone you touch. I thank God on your behalf, but I thank Him even more for the thousands of others like you. I bless this infinite abundance. I thank God for it, and receive it with an ever more open mind.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I find the money I require to buy the car I need.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

Money is energy, and I control its flow.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

Money makes me, and the people close to me, happy.
Money allows me to help others.
Money comes to me without effort.

Christian H. Godefroy, Prosperity: The Subliminal Method, Godefroy Publications, 1987.

Money comes to me freely.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Life places people on my path who can help me progress financially.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

On the first of —— (month), 19— (five years from now) I will have $50,000, accumulated gradually over a 5 year period. In exchange, I will do my best to provide excellent service as a (describe the service or merchandise) salesman.

I firmly believe that I will have that money. My conviction is so strong that I can already see it in front of me. I can touch it... it's there waiting for me to take it. I instruct my subconscious mind to come up with a plan that will get me that money, and as soon as my subconscious mind forms a plan I will put it into action.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

Table of contents

I become more financially comfortable every day.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

MOODS

From now on I will be in a good mood more and more often. Joy, happiness and serenity will become my normal states of mind.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I am going to devote time to studying the reasons for my mood swings.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

MOROSITY

I greet this day with a heart full of love.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I no longer think there is no point to life. I use the principles of truth to discover goals in life.

Life is worth living when I understand the value of the richness of life.

My life becomes richer and fuller when I can free myself from my negative self and the errors that part of me commits.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

Oh, what a wonderful day for going out and enjoying what the world has to offer!

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

This is the day God has made; I will rejoice and be happy on this day. I thank God for the abundance of life, I thank God for lasting love. I thank God for joy. I thank God for magnificent health, I thank

God for infinite abundance.

I have awakened to a new day. I feel at one with the birds and with all living creatures, celebrating this new day with a burst of song and thanksgiving. God, I thank You for the abundance that is mine.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

MOTIVATION

I like ——— (specify an activity which you like but which you haven't had the courage to undertake.)

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I feel rich! I feel happy! I feel great!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I enjoy my professional activities.

I am enjoying the new project I'm working on.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I show my self respect through this inner motivation which keeps me on the path of achieving my goals, on the path of success, even if I fail now and then.

I am my goal.

I am success, even before I attain it, since I am already on the path of full realization.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

If I want to achieve permanent, sustaining success, the motivation that will drive me toward that goal must come from within. It must be personal, deep-rooted and a part of my innermost thoughts. All other motivation, the excitement of a crowd, the stimulation of a pep-talk, the exhilaration of a passing circumstance is external and

temporary. It will not last.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

MOTIVATION - SCHOOL

After I finish this lesson I can go and watch the football game.
I'll do three more chapters, and then I can watch the movie on TV.

Professor Kurt Tepperwein, How To Learn The Painless Way, Godefroy Publications, 1983.

MOVING

I'm soon moving to a really nice apartment that will make my life more stimulating.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

MUSCLES

Every muscle in my body is controlled by my Muscle Coordination Center.

Franck Rudolph Young, Cyclomancia, SIP, 1966.

My muscles are relaxed and warm.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

My muscles are strong and supple.
My muscles are supple and healthy.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

MUTUAL AID

Every day I find occasions to help my peers and accomplish constructive things.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

MYOPIA (See Nearsightedness)

Table of contents

NAIL BITING

Every time I bring my nails to my mouth to bite them, as soon as I put my nails in my mouth, I will experience an extremely bitter taste, and the taste will get more and more bitter, more and more unbearable to the point where I will feel sick and want to vomit.

Gradually, as I grow up, I will become more and more attractive. So I don't want to harm my appearance by having ugly hands; beautiful hands with well trimmed nails will make me even more attractive, and for this reason I will make a determined effort to stop biting my nails.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

I no longer feel any need to bite my nails. I find it easy to break the habit, and my hands are becoming beautiful and elegant.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

NATURAL

I am natural and confident.
I act as naturally as I feel inside.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

NEARSIGHTEDNESS

The lens of my eye, too rounded (too convex) ... is relaxing.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

Table of contents

NEED

I have legitimate needs and wants. I can choose what I need and want without having to justify it to anybody. I make choices, and I take responsibility for them.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

It's all right to meet my needs as I see fit.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I express my needs honestly, listen to the needs of others, and look for the deepest truth.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I first understand my own needs, and then the needs of others.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I can invent new ways to satisfy a need and wisely choose the best option.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I am secure enough to grow. I ask for what I need. All my needs will be fulfilled.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

The universe always provides for my needs.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

What I say, I get.
And I always have everything I need.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

Table of contents

NEGOTIATING

I know how to break off negotiations if I have to.

My negotiations always result in clear, explicit and concrete agreements.

Before negotiating I clearly define my objective(s) in order to know exactly what I want and what I don't want to happen.

Before negotiating I clearly define my bottom line (without necessarily telling the other party) in order to use dissuasive strategies without being forced to walk away from the table.

When I negotiate I try to remember that any business relationship is comprised of one part cooperation and one part competition, each of which may predominate according to the time and circumstances.

I am careful not to get boxed into an undesirable choice: friend or enemy, trust or mistrust, cooperation or competition.

I make clear pronouncements of the points of agreement as they occur.

I know how to break down my negotiations, so that the discussion about one clause does not systematically bring into question all the clauses already agreed to.

Breaking down the negotiating agenda allows me to remain firmly opposed to the other party on certain points.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

If I negotiate without having the power of decision, I make sure the scope of my delegated authority is clear in my mind, and in the minds of everyone involved.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Table of contents

NERVOUS BREAKDOWN (Also see Nervousness)

From now on I will no longer be taken by surprise as I was in the past; every time I feel a nervous attack coming on I will be warned by certain symptoms; but these symptoms will no longer frighten me at all, because at the same time as I feel them I will hear a voice inside me saying, as quick as lightning, "No... no... I will not allow this breakdown to happen. My nervousness is disappearing... it has disappeared."

Emile Coue, Complete Works, Astra Publications, 1976.

NERVOUSNESS (Also see Agitation, Anxiety, Fear, Nervous Breakdown)

I always act calmly... without making needless gestures... and I never swear in moments of impatience.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I don't have to be nervous and tense. I am capable of relaxing. I have learned how to relax my mind and body.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I want to dominate my anxiety, fear, nervousness and any other paralyzing emotions.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

When I find my Real Self, nervousness disappears forever.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

NERVOUS SYSTEM

My nervous system is making amazing progress. I am affecting it in two ways, which makes my influence doubly effective. Firstly I influence it directly through suggestion, since the nervous system is directly related to the brain; secondly I influence it indirectly by improving my general state of health, since the nervous system is also affected by the body's general state. My nervous system is becoming more and more sensitive and vibrant, as it should be, and at the same time it is becoming more and more calm, strong and resistant, as it also should be.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

NEW BEGINNING

I am making a new beginning towards a new and better life, and I will get involved with other people.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

NOISE

Noise doesn't bother me... I know how to adapt to any situation... without becoming nervous.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

NOSTALGIA

I should not be nostalgic about the past. I should think about today.

Table of contents

Today is the day I should be living.
I am doing all I can to love this day.
I will use this day to express my positive qualities freely.
I am ready for a day of confidence and accomplishment.
I combat negative emotions which will destroy my having a good day.
I call on my positive qualities: life is lived in the present.
I nourish my enthusiasm for the present moment.
I will use this day to try and express the best of myself.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

The good old days are here and now!

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

NOVELTY

I like to invite people with very diverse opinions about things. I seek out novelty instead of limiting myself to conventionality, where the unpredictable is excluded.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

NUMBNESS

I want (specify which part of your body) to become numb. I impose my will.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

My hand feels dead, completely numb... and this numbness will soon move up into my wrist, my forearm, my upper arm, right up to

my shoulder... I can feel it moving upwards... my whole arm is becoming numb... there is no feeling at all... it is becoming more and more numb... my whole arm is dead... numb... completely numb.

Marcel Rouet, Techniques And Practise Of Hypnotism, Paris Productions N.O.E., 1974.

NUTRITION

I only eat foods which are good for me.

I only eat foods which are healthy and which help me maintain my ideal weight.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I put only the best fuel into my body.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

[I am constantly fulfilled in my life. Since I know life, I nourish myself with joy. I construct my body in a balanced way.]

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I am going to eat good wholesome food and not overeat, or overindulge in foods that are not good for my health.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

OBESITY (Also see Nutrition, Physical Exercise)

I respect my body, I care for my body, I need my body to live.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

I am gradually reaching my ideal weight.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I maintain my ideal weight. It's easy to do because I exercise and eat in a balanced and healthy way.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I never eat between meals.
I chew my food very slowly.
I don't eat food that contains wheat.
I don't eat food that contains sugar.
I don't eat standing up.
I visualize the body I want.
I eat smaller portions.
I don't need to be encouraged not to overeat.
I like myself and I want to be the best I can be.
I persevere until I attain my goal.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

Losing weight through self hypnosis will keep me from developing health complications associated with obesity, such as hardening

Table of contents

of the arteries, coronary disease, diabetes, etc.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I stay at my ideal weight no matter what I eat.

Sublipower, Sublihealth, Edi Inter, 1989.

OBJECTIVES (Also see Goals, Self Control)

Every day I make progress towards attaining my objectives.

Every day I continue moving forward by trying as much as possible to avoid the hazards of life. To do this I summon my inner resources to lead my to my objectives.

I learn to attain my objectives by mastering my fears and rejecting my negative emotions.

I choose realistic objectives and then work on attaining them.

I concentrate on my objectives.

I try to attain my objectives every day. I break out of the prison I have created for myself and strive to reach my objectives.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

It is easier and easier for me to realize all my dreams.

Christian H. Godefroy, Prosperity: The Subliminal Method, Godefroy Publications, 1987.

I have clear objectives which guide me in determining my priorities.

I often review my priorities in life, my professional and personal priorities, in order to discern convergencies or contradictions, and any lack of coherence.

I define my priorities in order to make them concrete, stimulating, and attainable.

I am make sure to explain my objectives to superiors and subor-

dinates.

I work on my objectives.

For each objective I strive for harmony between the objective itself, the results it yields and the methods used to attain it.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I easily attain all the objectives I set for myself.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I make a list of goals that I hope to achieve for the day. I visualize myself accomplishing each goal in the best way. I write a note of thanks that I've reached my goals.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I establish my objectives and then use the resources I possess with devotion and perseverance in order to obtain the desired results.

When I keep my mind fixed on an objective I increase my chances of attaining it immeasurably.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I make sure my objectives are clear to myself.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

I set precise objectives and deadlines for myself.

Sublipower, Sublisuccess, Edi Inter, 1989.

I set my objectives, write them down, and review them often.

My objectives are a clear illustration of my future!

I spend a few minutes every day reading over the file of objectives I've written for myself. I read over my objectives every morning after I get up, and every evening before going to sleep.

My objectives are very precise. The more detailed and precise they are the better I can visualize them and make them a concrete part of my life.

Table of contents

I set myself an objective. I don't hesitate, but start striving to attain it. I know the universe is overflowing with opportunity. Look at what I can do! Where I can go! Look at what I can accomplish, simply by saying "Yes!" to myself.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I never lose sight of my long term objectives.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I'm in my car driving towards my destination. I'm in the driver's seat, and I don't let anyone else drive my car. I am the driver. If I get to a dead end I back up, turn around and continue my journey towards my objective.

I remind myself to concentrate on my objective all day long.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I eliminate this objective for such and such reasons. I realize I no longer need it, or no longer want it. My initial idea is now only a memory of a desire. It has no more power or effect on me. It no longer interests me. I am now concentrating all my interest and all my energy on a new and better objective.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

My objectives in this situation are important. If I don't achieve them, I'll be disappointed and it will be inconvenient, but it will not be a catastrophe or a horrible event. By thinking about negative results as a catastrophe, I'm creating tension and fear that actually reduce my chance of success.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

OBJECTIVES - FINANCIAL

Table of contents

In one year I will find the $———— I need to buy a house.
Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

OBLIGATION

I always respect the obligations I accept. I never accept obligations which I cannot respect.
Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

OBSESSION

I am now calmly and confidently concentrating my mind on another subject.
Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

If I cannot meet such and such a deadline, I won't let it become an obsession. I am satisfied knowing I've done my best.
Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

OBSTACLES

I eliminate the word impossible from my vocabulary!
I possess the means to overcome all obstacles.
Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I can overcome obstacles by mentally flying over them.
Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I will take care of all kinds of obstacles as I encounter them. I will go as far as I can, and when I get there I'll be able to see even farther.
Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

Table of contents

I was born to win, to succeed. I am very prosperous, and through the power of God I overcome all obstacles and all problems. I am happy, joyous and free.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

I can overcome anything, and see everything through to the end.

George Barbarin, How To Overcome Fears And Anxiety, Dangles Press, 1974.

Up to now no problem has been able to get the better of me: I'll overcome this one as well.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

Now everything will work out.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

The greatest obstacle to my success lies within me. What prevents me from doing what I want to do is my belief that I cannot do it.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

OFFERING

The more I give the more I receive and the happier I feel.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

OPINIONS

I don't have to give up my opinions for those of another person just because that person seems sure of him/herself.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I can refuse to discuss my opinions or try to convince someone of the logic of my position, and simply believe that I am right.

Table of contents

When I buy clothes or other personal objects I have confidence in my taste, and I don't start by asking advice from someone whose opinions I deem to be more valuable than my own.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Other people can disagree with me without it affecting me.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

OPPORTUNITY

I take action! I seize opportunities!

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I readjust my objectives according to circumstances, in order to exploit opportunities to their fullest.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am always at the right place at the right time, doing the right and successful thing.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I take advantage of every opportunity.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

A favorable opportunity will give me a sign if I am meant to see it; I am on the lookout for opportunities, and I always try to take advantage of them.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

OPTIMIZING

Table of contents

I always try to make the next hour better and more productive than the one that has just past.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

OPTIMISM (Also see Depression, Inspiration, Positive Thinking)

Whenever my attention strays from what is constructive and positive I immediately concentrate on things that are good and beautiful.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I love life... it's wonderful... full of all kinds of surprises and sudden developments... I never get morose... I'm never bored... because I only see the good side of things.

I stay away from pessimistic people... people who criticize and complain all the time... who are never satisfied with anything... instead I seek out spiritual people, people who radiate health, and whose inner life is calm and serene.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I listen to upbeat, inspiring music. When I am getting ready for work or school I turn the radio on to a good FM station. I stay away from the morning TV news. I can brief myself by scanning the news on the front page of the Wall Street Journal. I resist the temptation to waste time reading the sordid details of someone else's tragedies. I listen to music or cassette tapes in my car. If possible I have breakfast and lunch with an optimist. Instead of sitting in front of the TV at night, I spend time listening to and being involved with those I love.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I use positive words. Never negative words. I use the RIGHT WORDS!

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I inject myself with optimism.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I always expect the very best for myself.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I feel good, very good, my life is going wonderfully well. I have no reason to be afraid. I never get depressed. I'm an optimist. I have confidence in the future, the people close to me love me, my colleagues respect me. I do my work well. I have no fears about the future.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

One of the best ways to make other people happy is to spread optimism and good humor. Every one of my words and actions can contribute to that.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

OPULENCE

I am ever thankful for God's riches, present as always, unchanged and abundant.

Dr. Joseph Murphy, Miraculous Power Attracts Infinite Riches, Godefroy Publications, 1978.

ORDER

Order which is the Spirit is, from this moment on, a part of all my affairs, a part of the way I use my time, a part of my body, my thoughts, my mind and the world.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

Now I put my life in order as I prepare to accept all the good that awaits me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

ORGANIZATION (Also see Efficiency)

I accomplish my tasks in order of importance.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I am well organized in my work.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I possess an organized and orderly mind, and that shows in the way I live.

I am very organized. Every night I draw up a list of tomorrow's activities. I establish my priorities and I stick to them.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I take time out to think about the problem at hand, and then I act scientifically.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I get the best out of my time.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

ORIENTATION (See Choice of Profession)

ORIGINALITY

I will find the original me when I eliminate all acquired perceptions concerning myself. I am eliminating them.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

Table of contents

OUT OF BODY EXPERIENCE

I know that I can condense my identity and appear in a distant place.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

OVERWORK

I have the right to take my time.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I have the right to live in the present.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I have the time I need to do what I want to do.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I keep my Workaholic firmly under control.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I can stay relaxed and focused on my timetable.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I am relaxed and balanced. I have all the time I need to do all the things I have to do.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

My days unfold tranquilly, at the pace I like. I am satisfied with my accomplishments, and with my life. I am affectionate with the people close to me, and greatly appreciated by those around me. I react positively.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

Table of contents

PACE

I adapt my pace to the activity I am engaged in. I make sure to work at the right pace.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

PAIN (Also see Hepatic Colic, Illness, Wounds, Anesthesia)

Soon the pain will pass. I will feel fine.

Leslie M. Le Cron, Liberate The Power In You Through Self Hypnosis, Tchou/Ariston, Forces

As soon as I press the nail of my thumb against the inside of my little finger I will block all the nerves conducting the pain.

A wonderful feeling of calm and security will pervade my body... as long as I keep pressing my thumb nail against the inside of my little finger I will feel no pain... I will feel completely fine.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

I eliminate the pain.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

I relax my entire body and forget about the pain.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

I concentrate on calming my body down. I am completely relaxed,

and feel this sense of relaxation throughout my whole body. I don't think about the pain at all. Instead I savor this beneficial sensation of total relaxation.

As I count down from five the sensation of pain will diminish. Each number eases the pain even more. When I reach the number one the pain will be completely gone. Now I'm counting down... five... four... three... the pain is disappearing... two... one...

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

Energy circulates harmoniously in me. A wonderful wave of heat and well being spreads throughout my body, which is becoming light, very light and pure, full of light and life. My pain dissolves and disappears completely in this wave of light and heat. The power of life flows and spreads through my regenerated body. I am completely healed.

Charles Baudouin, Psychology And Practise of Autosuggestion, Idegraph Publications, 1990.

Pain is sometimes caused by, and always amplified and aggravated by nervous tension and anxiety... Therefore as soon as I relax I start feeling better. And gradually, as I become more relaxed, the pain gets less and less intense. And now I feel so totally and completely relaxed that the pain is gone.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

My pain is disappearing. Soon I won't feel it at all, it will be completely gone. I'm already suffering a lot less, less and less... In an hour (or in a few minutes - set a time limit) I will not be suffering at all.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

PARALYSIS

God has not given me a fearful mind, but a mind full of power,

Table of contents

love and health. My faith is in the curative powers of God, my creator. I am relaxed and peaceful. Each one of my nerves is also one of God's ideas, and I feel revitalized, energized and healed. I give thanks to God.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

PARTNERSHIP

I find the partner I need to help with my work. Our association will be mutually profitable.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

PASSION

The only way to be happy and make a lot of money is to do what I really like to do.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

PATIENCE (Also see Understanding)

It's easy for me to be patient, understanding and just.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

When I have to wait I do so without getting impatient... by using the time to relax.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

Nothing can stop me today. If I need some extra determination, I'll find it! If I need extra energy and drive, well I've got it! Whatever task or problem awaits me, I have the power and the patience to over-

come it.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

PEACE (Also see Joy, Love, Wisdom, Prosperity, Health, Success)

God, my Father, all my thoughts are of You. My body, my heart, my intelligence, my soul, my mind, depend on You. In You I find refuge, light and life.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

I inhale the peace of God, and exhale the peace of God for all.

Dr. Joseph Murphy, Miraculous Power Attracts Infinite Riches, Godefroy Publications, 1978.

I am completely relaxed physically and intellectually. I am free. I believe in the wisdom of God and I feel my mind filling with His peace. I feel it right to the bottom of my heart and my being.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I am at peace with myself and with others.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

Peace is in me... peace reigns in this home.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

The current of God's peace flows freely through my whole being, and I am immersed in the peaceful river of God at this moment.

I am relaxed and calm. I am at peace. A peace which goes beyond understanding fills my mind. Peace is in my home, in my heart, and in all my affairs.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

Joy now flows freely through me, and I am at peace with life.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

Table of contents

I raise my soul towards you, Lord.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

PEACE OF MIND

I am at peace with myself.
My mind is at peace.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

PERFECTION

I am whole and complete unto myself.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

My life unfolds in total perfection.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am perfect.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I always try to improve myself, eliminate my faults, resist my negative impulses and become a better and more perfect person each day.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

PERFECTIONISM

It doesn't bother me to be imperfect.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I stop trying to attain perfection in myself.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I stop making it an obligation for me or my children to succeed at everything we undertake.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

I always do my best. Each thought and action is the best I am capable of at the time. Because I'm human, I make mistakes. I accept my mistakes without blame or judgment. When I make a mistake I learn from it. I am imperfect, and I forgive myself for my mistakes.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

PERFORMANCE (Also see Self Image)

My concentration always improves when I make an effort... concentrating allows me to unite all my energies into one effort... and thus improve my performance.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I maintain and develop my image of technical performance.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

One of the best, quickest and most effective ways to improve my image and my performance is to be part of an organization which has useful goals.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

PERFORMANCE - SPORTS

In a short time I will regain my superiority.

I'm sleeping more and more deeply... more and more deeply... I take pleasure in competition because I know my performance is improving day by day... day by day my performance is improving... I

feel very good.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

PERSEVERANCE

It is still too soon to abandon everything, and I won't give up.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I keep trying.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I'll go that extra mile.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I tell myself to persevere and attain my objective; I refuse to be swayed by doubt and fear.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I never forget that my life belongs to me. I continue trying. If I can't make it on my own, someone else will probably help. But I keep trying and I persist.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I persevere until I succeed.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I will persevere until I succeed.

As long as there is a breath left in me I will persist. Because I now know one of the most important elements of success - if I persevere long enough I will win.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I persist with my dreams; they'll become real.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

Table of contents

I will stay healthy and generate new strength which will help me want to act with perseverance.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I am strong... I succeed at everything I do because I persevere... I see things through to the end... I am conscious of my power!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I want to play an active role and dedicate all my strength to attaining my goals. I will persevere despite all obstacles because I am certain that properly directed willpower can overcome anything, if the goal is just and honorable.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

Once I make a decision I persevere in carrying it out for as long as I believe I am right.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

If I don't succeed on the first attempt I try again and again.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

PERSONAL AFFAIRS

I am very efficient at taking care of my personal and/or family affairs.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Editions, 1981.

PERSONAL DEVELOPMENT

I am changing my life. I experience a deeper connection with my true self. I am excited about my growth.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

Table of contents

PERSONALITY

I am asserting my real personality more and more every day.
Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I do what I have to do, and you do what you have to do.
Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

I don't try to do better; I try to be different.
Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I don't imitate others. I try to understand my personality fully, and be myself.
Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I don't compare myself to others.
Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I am not inferior. I am not superior. I am simply me.
Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I'm not going to let myself imitate the undesirable personality traits of other people.
Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I'm not here to live the way you want me to.
Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

I make a firm decision to be myself and not an imitation of somebody else.
Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am me and you are you.
Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

Table of contents

I am me.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I am a unique being and I express my self fully, aiming for success and wealth.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I am unique and I like it that way.

Sublipower, Subliconfidence, Edi Inter, 1989.

I am unique, from head to toe. I may resemble other persons, acting and speaking like them, but I am not them, I am me.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I want to acquire constant calm, imperturbable self assurance, and justified confidence in myself.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1973.

My personality is who I think I am.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

The more I express myself, the happier I feel.

Louis Proto, Self Healing, Piatkus Publications, 1990.

You are not alive to live the way I want you to live.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

PERSUASION

The most persuasive influence I have over others is my mature personality.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

Table of contents

PESSIMISM

I forgive myself for holding on to negativity, and now let go of it.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I will not allow my personal opinions to distort the facts in a negative and pessimistic way.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

PHOBIAS (Also see Fear, Negative Thoughts)

I am in harmony with God, and I have nothing to fear.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

My phobia is afraid of me because it knows I am much stronger than it is. I am at ease, perfectly relaxed and confident of my strength. I smile because my phobia has no more power over me. It is unimportant.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

Little by little I see these fears, these ideas, these phobias appearing less and less frequently in my mind; they are getting weaker and weaker, less and less tenacious; they have less and less influence on me; I am eliminating them right now by repeating the formulation: "It's over, it's all over..."

Emile Coue, Complete Works, Astra Publications, 1976.

PHYSICAL STRENGTH (See Strength - Physical)

PLANNING (Also see Money)

Above all it is essential I obtain such and such results, which will

Table of contents

bring me closer to my ultimate goal, whether materially, or by increasing my worth or my power.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

I have finally come up with a working plan, and I intend to stick to it.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

Every day I draw up a list of things to do.

I determine the time required for each activity.

I allot some time for pauses, unexpected events, and a margin of error.

I draw up my list of things to do every day.

I make plans.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I study, I think, I take the time to make plans!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I do high priority work first. I set my priorities on a must-do-now, should-do-soon, and would-like-to-do-when-possible basis.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I gave myself orders and I carried them out.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I never waste my time because I plan. And since I always stick to my schedule, I have enough time to accomplish what I choose to do.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am in control of my time. My time is the very essence of my life, and I take control of it.

I create as much free time as I like, and I know how to delegate.

Christian Godefroy, The Creativity Method, Godefroy Publications, 1992.

Table of contents

Planning builds self confidence. I know beforehand what I am going to do each day. This enables me to develop self discipline, which is so essential to success.

Planning gives my life direction. I must know where I am going, the goals I wish to attain, and how not to dissipate my time. As my plans materialize I become inspired by a sense of achievement.

Planning enables me to get more fun out of life. The habit of day to day planning should include time for enjoyment of life whether it is reading an interesting book, listening to soothing music, playing golf or other sports, seeing a good movie or play or anything else that will relax my mind. To be happy I must make plans for happier living.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

When I plan, I don't just plan FOR the future, I plan so that I will HAVE a future.

This task needs to be done but I can't do it right now. I'm going to schedule a time when I know I'll be able to do it and stick to that schedule. If I don't there's a good chance I won't get to it. Now that I've scheduled it I can forget about it and concentrate on the job at hand.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I choose which Concepts, Contacts and Commitments warrant the most attention and effort. I am responsible for which of them become contractual, with priorities and deadlines. I create my own track to run on - slow, medium, or fast - in my commitments.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I finish each task before moving on to the next one.

Sublipower, Sublisuccess, Edi Inter, 1989.

Table of contents

PLANS

I am open to new points of view, perspectives and plans.

God as Understanding knows what I should know, and I am united with this universal Understanding so that I know everything I need to know.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

From now on I live in harmony with the divine plan assigned to my life.

From now on I recognize, accept and conform to the divine plan of my life, as it is revealed to me step by step.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Talking about ideas and plans needs to be balanced by trying them out. Theory and practice converge into wholeness. I field-test my ideas.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Starting today I will become the architect of my life.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

PLAY

(I am free to enjoy myself any way I want.)

Louis Proto, Self Healing, Piatkus Publications, 1990.

I take time to play like a child.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

PLEASING OTHERS

I please everyone I meet. I have confidence in life because I have

Table of contents

confidence in myself. I know I can count on this immense reservoir of energy that is within me, and which attracts others. I radiate the energy that is mine.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

I want to please everyone.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

PLEASURE

I begin the day with pleasure!

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

PLEASURE - SEXUAL

I deserve love and sexual pleasure.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

PLEURISY

I call on the forces of nature to help heal me. I absorb the curative forces of nature and unite them with my own. I eliminate the gas and other waste products which are making me ill.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

POSITIVE MENTAL ATTITUDE

I like people and people like me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

I am learning to love instead of suffer! I am learning to love what I hated, and transform negative forces into positive energy. I am forging a path towards happiness. I react positively, even if it takes a great effort at first. And I remain positive even if an adversary is not as positive as I am. I notice the changes this is producing in my life and in the way people see me.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

Whatever the situation, I always adopt a positive mental attitude. And I succeed thanks to my positive mental attitude.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Every morning I am going to condition my mind and instill a positive mental attitude for the rest of the day.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I am developing a positive mental attitude.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

POSSESSION

I can have anything I want!

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

POSSIBILITY

I always exceed my possibilities.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

POTENTIAL

I awaken the giant within me!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Table of contents

I nourish my real potential.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

POVERTY

I will buy it! I'll get it!

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

My inner Guide is the spirit of plenitude and will preserve me from all misery. I will lack nothing.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

POWER (Also see Self Confidence)

I can do anything.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I can.
I can do it if I believe I can!
I can make it if I believe I can!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I possess a great and marvelous power.

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

I could do it.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

When I want to do something within reason, or when I have to do something that is my duty to do, I always imagine that whatever I do is easy. May words like difficult, impossible, I can't, it's too much for me, I can't help myself... and so on, disappear from my vocabu-

lary. They are not English. I don't understand them. What I do understand is: it's easy, and I can do it.

Emile Coue, Complete Works, Astra Publications, 1976.

Each day brings me more strength and power.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Day by day, in all ways, I am becoming more powerful and confident.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I possess vast power which I have not yet realized.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am united through my inner Guide with an inexhaustible source of health, courage, abundance and prosperity. It fills me with strength and spirit. It provides me with everything I need to move forward ceaselessly. I am confident in its power; it is my strength and my wisdom. It makes me strong and independent of things and people. I am persevering, strong and courageous. I am strong!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

My constant contact with God gives me the power of an eagle soaring above the clouds and the humility of a gentle dove.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

POWER - PSYCHIC

I am learning how to use my mind in extraordinary ways, like mastering my psychic powers, and I am becoming a different person.

Franck Rudolph Young, Cyclomancia, SIP, 1966.

Table of contents

POWER - SPIRITUAL

I can touch the sky. I can be big. I can fly if I want to - I can do anything.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

POWER - SUBCONSCIOUS

I have unshakable faith in the power of my subconscious... unshakable faith.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

PRAISE

S/he worked hard to get the promotion s/he wanted. I wish him/her well.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I allow others to love me and praise me.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

PRAYER

God has received my request.

Richard Ingalese, The Power Of The Mind, Dangles Publications, 1981.

May my comings and goings be done in all honesty, and in the full light of day; not in disorder and drunkenness, not in licentiousness, not in anger or envy. I place my faith in our Lord Jesus Christ, I care nothing for the needs of my body, or for my frivolous desires.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

Table of contents

I pray, I really pray.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

PRECAUTIONS

I care for myself enough not to expose myself to physical risks or emotional damage.

Louis Proto, Self Healing, Piatkus Publications, 1990.

PREGNANCY

I await my pregnancy with pleasure.
I will be happy to be pregnant.
I want to get pregnant.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I feel marvelous... I know the birth will be easy.

I'm enthusiastic about the prospect of having a wonderful child.

My thoughts have a positive influence on the health of my child... on his development. If I am assailed by negative thoughts I always remember that they can hurt my child... and I immediately replace them with joyous, optimistic thoughts of a radiant future.

Everything is going fine... I am firmly optimistic... firmly optimistic. My pregnancy is progressing normally... and will continue to do so right up to the birth, which will be easy... very easy.

I know that my child is developing perfectly. I am sure to have a beautiful baby... a superb, magnificent child.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I will not suffer from morning sickness.
I'll get my figure back after giving birth.

Table of contents

I am thrilled at the idea of waiting for my baby.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

PREJUDICE

I am now dissolving all my negative and limiting prejudices. They no longer have any power over me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

PREMATURE EJACULATION (Also see Impotence)

I control my ejaculation.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

PREOCCUPATION

Everything is going very well.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

PRESSURE

I choose not to spend my life trying to live up to other people's expectations. I don't let others pressure me with their beliefs unless I accept them.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

When external pressures become too strong I call on my inner strength to combat them; to do this I mentally review the successes I

have already achieved.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

PRIDE

(I am proud of my accomplishments and goals.)

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

PRIORITIES

I have enough time for my priorities: 80% of the time, every day, I work on two out of three priorities.

I set deadlines and time limits for myself, and I ask others to do the same.

I regularly sit down and review my priorities.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I re-evaluate my priorities.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I am a person who is responsible and I act in timely manner to reach my goals and priorities.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

PROBLEMS

Before attacking a problem I try to formulate it in terms of "How to do what for whom?"

Before attacking a problem I ask myself, "Am I really determined to resolve this problem?"

Table of contents

If the problem consists of choosing between a relatively small number of clearly defined options, I use the "Pros and Cons" method.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

All right, here's the problem. I look at it calmly and carefully study its elements. I consider its implications. I take time out for careful reflection. And above all... no emotions!

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

This is my problem. I can do anything I want about it. I know the answer to everything. I will find the solution.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

It works!

My subconscious knows the answer to all my problems.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Tonight my subconscious mind will resolve the problem of — (specify). When I wake up tomorrow morning I will remember the solution, either as a dream, or as a flash of inspiration during the day.

Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

In quietness I find the answer to every problem. Everything I need to know is inside me.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I am learning to see problems on a higher level of understanding. They seem to solve themselves under my very eyes. I try to approach a problem as if I had no connection to it whatsoever. This prevents me from clinging to prejudices and ideas which are impossible to realize, and allows me to access a higher level of understanding. In fact a problem can only be solved when it is completely understood.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I seek an end to all problems.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Table of contents

I know that the resolution of my problem depends on me: the answer is in me because God is in me. I now feel calm and relaxed. I am at peace. I know that confusion never comes from God. I am now in harmony with the infinite. I believe and know that infinite wisdom will reveal the correct solution to me.

I know that all problems can be solved when the mind is calm. I confide in God because He knows the answer. I am at peace.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I attack the problems of life without being afraid to make mistakes, and thus I break free of my chains.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I'm not afraid of problems, I solve them. I don't try to ignore them, I confront them. I don't avoid them, I conquer them!

I know that every problem's solution has a key. Therefore the clearer the problem is, the better I can determine its solution.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I can easily resolve — (specify the problem).

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I look at each problem as an opportunity.

Christian H. Godefroy, Self Confidence: The Subliminal Method, Godefroy Publications, 1987.

I overcome any problems easily. My life is enjoyable and positive.

Vera Pfeiffer, Positive Thinking, Element Books, 1989.

I know my victory over my problems is certain because I know that my intelligence is a part of God.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

Table of contents

I know you have the answer to this question and I know you will give it to me when you are ready. I will wait patiently, in complete confidence.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I see you with the eyes of love. There is a harmonious solution, and I accept it now.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

Easy is right.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Only living beings have problems. The more problems I have, the more alive I am. I am happy to have these problems.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I relax and let divine wisdom and the divinely accurate action of my subconscious mind resolve this problem. I disassociate myself from the problem and let things take their course.

Dr. Joseph Murphy, Miraculous Power Attracts Infinite Riches, Godefroy Publications, 1978.

I offer thanks to my Father for His Divine solution.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

PROCRASTINATION

Today I will do my work without wasting time.

I have finally decided to correct my habit of putting things off for later.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

Right now is the best time.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

Table of contents

This task is important and it's the kind I might procrastinate on or forget to do at all. I'm going to do it now and get it out of the way. It won't be hanging over my head and I'll be mentally free to concentrate on my other work or just relax.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I act immediately.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

It is my duty to finish what I begin.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I stop worrying about what I can't do now, and do what I can do with all my heart and soul.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I always do what I have to do on time.

I immediately transform all indecisiveness into ACTION, and so I stop stalling.

I don't let chance decide for me. Since I can make decisions rapidly, stick to my convictions and act in consequence, I have no time to stall around.

I have learned to recognize all the arguments that tend to drag me down, and I avoid them instead of avoiding the tasks I encounter in the course of doing my work.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I do everything I have to do at the right time.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I do it now!

John Clark, Do It Now, Godefroy Publications.

I'll do it without wasting a minute.

W. Clement Stone, Passport To Success, Godefroy Publications, 1962.

Table of contents

I'll do it right away... I'll do it immediately!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I set to work immediately and passionately.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I will not neglect today's tasks by putting them off for tomorrow, because I know that tomorrow never comes.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I profit from each day: I don't put off what I have to do, and I do it with joy.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

When I'm faced with a problem I solve it right away if I have all the facts. I don't put off making decisions.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

If I want to succeed I have to start this very day.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I do everything I undertake with care, and I complete it before moving on to some other occupation. I never put off for later what I should and can do now.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

PRODUCTION

I am satisfied with the results produced by my working time.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

PRODUCTIVITY

Whenever I make a change in my life, from the current way I am

Table of contents

doing something, I anticipate a temporary drop in productivity and efficiency. I don't worry if a change I made in profession or lifestyle isn't bearing fruit right away. It takes time for a change to be assimilated. As familiarity and confidence rebuild, productivity will increase again. I don't stew, I let it simmer for awhile.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

PROFESSION (Also see Wealth)

I deserve to have a profession that satisfies me 100%, and which gives me the opportunity to develop and enrich myself beyond my expectations.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

There are experts and professionals in my field who will help me learn, improve and excel. They will respond to my sincere, thought-out questions and concerns.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

PROFIT

I profit from the present moment.

Sublipower, Sublisuccess, Edi Inter, 1978.

PROGRESS

I have everything I need to make profitable progress under the present circumstances.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

Table of contents

I move forward with all my strength.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

I develop affirmative statements, about five words in length, describing my correct performance, in the present tense. I relax and listen to myself state the affirmation, and visualize the accompanying action or feeling.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I am pleased with the progress I am making. I'm going to tell my best friend about it.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

I move forward!

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I want to do better.
I want to do more.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

PROJECTS (Also see Debt, Money)

I set a deadline for every project.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

This project does not have to succeed, and if it fails the world won't come to an end. But I want it to succeed, and so I'm going put all my energy and enthusiasm into it. What do I risk by doing that?

John Clark, Do It Now, Godefroy Publications.

Creative wisdom is at work through me to bring all my projects and plans to fruition.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

Table of contents

PROMISES

Every time I promise something I keep my word faithfully and exactly.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

PROMOTION

I am qualified for the promotion to (specify the job you wish to get) because everyone appreciates me and because I always demonstrate goodwill towards my employer and my colleagues.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I have a right to my place under the sun. I carve out this place through my work, through the contacts I make, through the dignity of my life, through friends and the help I give to others.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

PROSPERITY (Also see Happiness, Budget, Love Of Life, Health, Peace, Success)

Every day I am acquiring better health, a greater love of life and more and more prosperity.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

My interests are prospering day and night.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I accept and believe in prosperity for myself.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

Table of contents

I organize my budget well, and my affairs prosper.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I believe with all my heart that health, peace, success and prosperity in my affairs are mine to enjoy. I see myself under the spiritual sign of peace, harmony and inner development, an image which is completely oriented towards success and prosperity.

I impregnate my subconscious mind with ideal images of prosperity and success. I rely on the infinite source from which everything I need flows. I listen to the voice of God in me. His voice guides and directs me in everything I undertake. I feel linked to the wealth and generosity of God. I believe and I know that there are new and better opportunities in store for me. God will reveal them to me.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I deserve prosperity and happiness. From now on I will be prosperous and happy.

I am now ready to accept all the joy and prosperity that life offers me.

I am blessed with financial prosperity. Each day makes me more prosperous.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I have divine guidance all day long, and everything I do prospers. Divine justice and the law of divine order govern all my undertakings, and everything I do succeeds.

I know that prosperity means spiritual growth from all points of view. God helps me succeed in mind, body, and in the things I do. God's will is constantly at work in me, bringing me health and wealth, and making me an expression of divine perfection.

Dr. Joseph Murphy, Miraculous Power Attracts Infinite Riches, Godefroy Publications, 1978.

Table of contents

PROTECTION

I remain under the protection of the Most High.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I am never afraid because I know that the Divine Presence which is in me is my saviour.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I am safe and secure at all times. Love surrounds me and protects me. All is well.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I am guarded, I am protected.

George Barbarin, How To Overcome Fears And Anxiety, Dangles Press, 1974.

The Spirit walks ahead of me, showing me the path... I have absolutely nothing to fear, the Spirit protects me... the wisdom of the Spirit guides me in all my actions.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

PSYCHIC SHIELD

I will completely reject and ignore all negative and pessimistic "facts" which I can do nothing to change.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

PUBLIC SPEAKING

I like talking to people. When I address a group I feel self confident and relaxed.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

As soon as I begin speaking... I feel transported... I speak with conviction and warmth.

I could speak for hours... ideas keep coming to me one after the other.

At a meeting or conference... I am more and more likely to get up and express my ideas.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I have the energy, assurance and complete self control I need to speak in public. I speak easily and eloquently.

I feel relaxed in public. I feel at ease in any situation. I have complete confidence in myself and in what I have to transmit to others. Everything will be fine.

I am the most important thing I have to transmit to others. I am perfectly at ease. I am in perfect control of myself. I know exactly what I have to say. I am perfectly at ease in public.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

I have nothing to fear. I'm going to concentrate on the message I wish to pass on to my audience and not on myself - I have confidence in the power of my own mind.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I feel comfortable in public. I feel at ease under any circumstances. I have complete confidence in myself, and in what I have to transmit to others. Everything will be fine.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

I am the most important thing I have to transmit to others. I am perfectly at ease. I am in perfect control of myself. I know exactly what I have to say. I am perfectly at ease in public.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

PUBLISHING

Infinite Intelligence paves the way for the acceptance and publication of my writing.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

PULMONARY DISEASE (See Lung Disease)

PUNCTUALITY

I arrive on time for my meetings.
I arrive at the prescribed time.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am always punctual.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I rarely hurry, and I feel good about myself.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

PURIFICATION

Every breath I take cleanses my body totally. With every out breath I rid my body of toxins.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I eliminate gasses and all other waste products from my body.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1973.

I clean my mind with the waters of love and life.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 198

Table of contents

QUALIFICATIONS

I have the necessary qualifications, and everything else I need to succeed.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

QUALITY

It is my —— (specify which quality) that people appreciate in me.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I have the potential to develop all the qualities I admire in others.

Sublipower, Sublicontrol, Edi Inter, 1989.

I cultivate and develop my qualities.

Sublipower, Sublicontrol, Edi Inter, 1989.

I acknowledge my qualities, as well as those of others.

Christian H. Godefroy, Positive Thinking: The Subliminal Method, Godefroy Publications, 1987.

QUALITY OF LIFE

The quality of my life improves day by day.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Table of contents

RADIANT

I am a being of light. I am energy. I am whole as a spirit, mind, and body.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

The more radiant I am the more I attract good things.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

REACTION

I am perfectly capable of reacting in a happy and healthy way, no matter what happens to me.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

READING

I look at books as working tools.

Sometimes I read very quickly, sometimes very slowly, depending on what my objective is.

I have an alarm system that goes off every time I find myself getting distracted, reading a sentence over and over without really taking it in because I'm thinking about something else.

When I finish reading something I ask myself what I got out of

the book or article, and how I can put this information to good use.
Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Realization

All my thoughts tend to materialize in my life.
Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

What the mind can conceive and believe, it can also realize!
W. Clement Stone, Success Through Constructive Thinking, Godefroy Publications, 1978.

I have the right to realize my desires.
Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I can do anything I believe I can.
Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I'll do it!
Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I can do anything through the power of my subconscious mind.
Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I realize what I want to do.
Christian Godefroy, Mental Dynamics Seminar, Godefroy Publications, 1978.

The universe is now paying me to do what I really love.
Shakti Gawain, Reflections In The Light, New World Library, 1978.

Everything I have to do is simple and easy.
R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

REALIZATION - SELF

Every day I take a small step towards self realization.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

REALISM

I take care to always remain in contact with reality.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I am practical and realistic, and I keep my feet planted firmly on the ground. But I also give myself the freedom to live according to my ideals.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am a realist.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

REBIRTH

Today I start a new life.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

Just as the pain of childbirth is forgotten, I experience each aspect of my life as a manifestation of peace and joy. The past is dead. I am free of all suffocating prejudices. I am reborn every instant, I triumph in my rebirth and my new and happy being.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

RECEPTIVITY

I am receptive to healthy ideas and techniques.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

REFLECTION

I determine the time a task will take before starting it.

I sit down from time to time to reflect on things that are important to me.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

I always find a solution if I take the time to reflect.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I give myself the time to reflect on subjects that are important to me.

I write these subjects down in my agenda so I won't forget about them.

I know that reflection takes time, and that this time is not necessarily directly connected to solving my problems: a time of reflection can also be a time of maturing.

Whenever possible I either write down or record the fruit of my reflections.

Whenever possible I allow myself time for reflection before making a decision.

As I reflect I establish "satisfaction criteria" concerning the results I hope to achieve from the decision I make.

I use my time of reflection to gather ideas which I would then note down in my agenda.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Table of contents

I want to think about things, because every profound idea contains information that is only revealed through reflection.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

REFUSAL

I always say no when it's the right thing to do.

By not giving in to the demands or influence of others I become more aware of my freedom and my self control.

I am able to say no easily. I am self confident, and always in control of my life.

I never feel obliged to do something I honestly haven't chosen to do. I say what I think. I am frank, direct and honest. I know what I want and where I am, and I express my opinions clearly and without hesitation.

I allow no one to sell me anything against my will. I live according to my own decisions, without ever giving in to pressures from salespersons or anyone else.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

REGULARITY - ABDOMINAL

My solar plexus radiates heat.

I feel a source of heat in my abdomen.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

REGULARITY - RESPIRATORY

I am all breathing, there is something breathing inside me, I feel my rib cage rising and falling like a boat on the ocean, my breathing

Table of contents

is calm, I listen to my breathing. I am breathing gently.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

REJECTION

(I confront my fear, rejection and guilt and let go of them. I am going to discover and overcome these feelings which are obscuring my real light.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I allow myself to be, without the need to prove myself to anyone. I live my power without fear of defeat or rejection. I control my dreams, my life.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

REJOICING

It's good for me to have fun and rejoice, and that's what I do!

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

The more I enjoy myself, the more alive I feel.

Louis Proto, Self Healing, Piatkus Publications, 1990.

RELATIONS - COMMERCIAL

I can never control another person; I can only influence him. If I expect to be able to control him I'll probably be angry and frustrated. If I accept that I can influence his free choices I'll be calmer, learn more about him, and be much more effective.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

Table of contents

RELATIONS - FAMILY

I love my family.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

My relations with my family are improving.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

RELATIONS - HARMONIOUS

I have the right to be happy with my chosen love.

My relationship with X is becoming more and more harmonious, happy and enriching every day. We are evolving together. We always compliment each other perfectly. I give love and receive abundance in return. I live in the abundance of love and life. X is abundance, I am abundance.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am now ready to get involved in a fulfilling and happy relationship.

I am ready to have harmonious relations with everyone.

Divine love is now working through me to establish a perfect relationship with ——.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

The spirit in me speaks to the spirit in —— so that there is only harmony, peace, love and understanding between us.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

Table of contents

RELATIONS - HUMAN

To improve my human relations I accept people as they really are and not as I would like them to be.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I choose my relations. I deliberately seek out people with high moral values who see the beauty in life.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I am not interested in relating to anybody who does not respect me as much as I respect myself.

Louis Proto, Self Healing, Piatkus Publications, 1990.

My relations with people are becoming happier and more fulfilling every day.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

When I change my inner self, I also change all my external relations.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

From now on my relations with others will be happy, satisfying and full of love.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I have confidence in myself and I like being with other people.

Alan Houel, How To Deal With Difficult People, Godefroy Publications, 1990.

I have fruitful and pleasant relations with the people I see (both professionally and socially).

If I am with someone, I give them all my attention.

I try to discern what other people are feeling.

I may take a calculated risk of breaking off a relation if that is the

Table of contents

only way I can work things out.

I am careful to concern myself with only one pole of the relation at a time: me, the other person, or the situation.

I periodically review my personal relations and evaluate them, dividing the people I spend time with into positive and negative categories. I then determine those with whom I'd like to try and improve my relations, and those with whom I'd like to establish a relation, but haven't done so as yet.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I have very easy-going human relations.

I understand the natural laws governing relationships, and I work in harmony with them.

When I am in touch with my Real self, all my relations are cordial and effective.

Any sincere effort to improve my human relations will be rewarding.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I like people a lot, and I like to be nice to them.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, 1978.

I like to thank others.

The more I like myself, the easier it is for me to like others.

I am available for other people.

I know how to listen. I encourage people to talk about themselves.

I listen to them, and they confide in me.

I find it easy to meet new people.

I accept people the way they are.

I like to demonstrate my care and thoughtfulness.

I am sincerely interested in other people.

I wish others only positive things.

Christian H. Godefroy, Human Relations and Communication: The Subliminal Method, Godefroy Publications, 1987.

Table of contents

I will try to be a little more friendly towards others.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I will do all I can to understand and maintain good relations with —— (specify such and such member of the family, colleague, etc.).

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I imagine myself at my best and I project that image to others in order to improve my relations and become a happier person.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am going to practice the technique of self-relaxation whenever I find myself in a situation that may precipitate an argument.

Whenever I find myself baited into a heated argument I will change the subject or will remind myself to remain calm and realize that the other person may have reasons for feeling as he does.

I am going to give myself the daily suggestion that I will talk more to people, act more friendly and initiate the kind of conversation that will enlist the interest of the other person.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I seek and find the essential quality of each person I meet.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am in harmony with others.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

I treat others as I would have them treat me.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

I want to maintain the friendly and harmonious relations I have created with the people around me, my family, friends and colleagues.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

The only kind of relation I can have with anyone, no matter who

Table of contents

that person is, is one of love, understanding and peace. I am a free expression of God.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

RELATIONS - PERSONAL

People are important to me. I like the people who are a part of my life, and I know they like me.

I naturally attract honest and sincere people who respect my opinions, as I respect theirs.

My personal relations are warm, meaningful and enriching. I am always honest and sincere with the people I meet, whatever the situation, and whatever kind of relation we have.

I am worthy of confidence, I am responsible, and other people greatly appreciate these qualities in me.

I am always ready to give someone the benefit of the doubt, whatever the situation. When I'm uncertain, I always imagine the best possible scenario.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

RELATIONS - PROFESSIONAL

The basis of all the thoughts, words and actions I use to address my superiors and colleagues is my ever present good will.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

My relations with associates are governed by the law of harmony.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

Table of contents

RELAXATION

Every time I take a breath, and every time I exhale, I relax more and more deeply.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

Each inhalation and exhalation relaxes me even more.

Each new movement makes me a little more relaxed. I breathe in deeply and hold the air for the time it takes to count to five. I exhale very slowly. My body is feeling more and more rested, making me even more relaxed. And the more I relax, the better I feel. The feeling of being completely relaxed is very pleasant and reassuring.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I feel more and more relaxed.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I know how to relax and let go. I can flow with the current.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I know how to organize spaces of relaxation, even during very intense activity.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am relaxing now. I am at peace.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I am going to take time each day for some form of recreational relaxation to balance my day with work and play.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

My body is loose and relaxed.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

Every time I do my alphagenic induction exercises I relax quickly

Table of contents

and deeply.

I do relaxation exercises every day.

I feel completely numb and insensitive to the position of my body. I am cut off from the people around me. I feel secure and in control of the situation.

My body is relaxing more and more. I am becoming numb to all sensation.

I feel relaxed from head to foot. Every part of my body is relaxed, and I stop analyzing my feelings. I am not thinking about anything except relaxing as I watch the indicator continuing to descend. Soon I will be completely impervious to everything around me.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

With each breath I become more relaxed, with each breath my body becomes looser, heavier and more numb.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

I am falling asleep. My whole body is resting. I will awaken in half an hour, at exactly —— (specify the time), rested and full of energy. I am so light I feel as if I'm floating!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I relax every night.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I can relax.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I feel good, very relaxed.
A feeling of well being gently spreads through my body.
My scalp is relaxing little by little.
My forehead and face are relaxing.
My tongue is loose in my mouth.
I relax my throat, my tonsils, my vocal chords.

Table of contents

The muscles around my eyes are relaxing.
My neck and back are loosening up.
My tension is dissolving day by day.
My breathing is unobstructed and is getting fuller all the time.
My diaphragm massages my internal organs.
My solar plexus is relaxing.

Sublipower, Sublirelaxation, Edi Inter, 1989.

(I am going to practice relaxation as a way of life. I am going to practice relaxing each part of my body (progressive relaxation) until I have learned to master the technique of relaxing my entire body (complete relaxation). When my body is relaxed my mind will relax. Relaxation is the best antidote for worry. I am going to make good use of my new habit of hypnotic self-induced relaxation.)

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Relaxing is important to me, and I always make sure to get my required dose. I use my time not only to pursue my goals, but also to appreciate the beauty of life.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Everything is relaxing... everything is opening... becoming peaceful and relaxed... my head is heavy, my jaws are loose... my body becomes calmer and calmer every minute.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

RELAXATION - DIGESTIVE

My solar plexus radiates warmth.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

RELAXATION - MENTAL (See Inner Peace)

Table of contents

REMUNERATION (Also see Job)

I have a wonderful job and I am very well paid. I offer a valuable service and do my work admirably.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

REPARTEE

I can answer any question at all without hesitating.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

REPRESSION

As I count to ten I start becoming conscious of these emotions. I feel them more and more strongly. I will feel them intensely, in every fiber of my being.

Charles de Liguori, Hypnotism, de Vecchi Publications, 1975.

REPROACH

I remind myself that my relatives, my spouse, my friends, my boss, my children and all kinds of other people I come in contact with will often disagree with my behavior, but their reproaches have absolutely nothing to do with my inner self.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I start accepting certain attitudes which I have chosen to adopt but which others may disapprove of. Therefore, if a relative or my boss or a neighbor or even my spouse reproaches me for such and

such behavior, I find it completely natural that they do so.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

No more blaming!

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

RESENTMENT (Also see Forgiveness)

No feelings of resentment can enter my mind because I have a spiritual understanding of all beings. I contemplate the love of God in people, which blesses me and fills me with goodness.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I give up all my resentment; this done I find I can be honest with myself.

I reject my feelings of resentment; I use my intelligence to find ways to spend my days fruitfully. I forget the day's setbacks and, above all, I stop being obsessed with the numerous injustices I suffered in the past.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I forgive you. I may not approve or agree, but I can forgive. I can let go of the past and wipe the slate clean. I know better than to expect atonement. I let go of revenge and resentment. Our differences are in the past. I am in control of the present and I can forgive you in the present. I can leave my anger behind.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I want to get rid of all my bad memories, eliminate all hate and resentment from my heart and retain nothing which can degrade or belittle me.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

Table of contents

RESOURCES

God is the infallible and unlimited source of all my resources.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

My mind is strong and its resources allow me to control my body, my emotions and my thoughts.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

RESPECT (Also see Consideration)

I earn the respect of others.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I am respectful and listen to people who have experience.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

First I respect myself, then I respect others.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I respect myself enough to say 'yes' only when I mean it.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I respect myself. I respect my values, my thoughts, my ideas and acts. And I respect myself the most when I stick to the path I have chosen.

I have a healthy attitude towards myself. I really like myself. I cultivate a very positive self image, and a deep respect for who I am.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I respect others and they respect me. Every person I meet is important to me. I can easily create a feeling of respect and confidence between myself and others.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

Above all I respect myself.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

RESPONSIBILITY

(I am completely responsible for my thoughts, feelings and actions.)

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I consider everything that happens to me as being my fault.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

I have to accept my responsibilities with joy in order to relax my Real Self.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I will leave nothing to chance because chance does not exist, and the success I aspire to depends solely on myself.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I assume responsibility for my strength, my happiness, and for my positive and healthy attitude; I am responsible for my past, present and future.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am grabbing hold of my life.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I am happy to realize that I am not a victim of life's circumstances.

I am creating the happy life I lead. I am the author of my own life story.

I don't need to make up excuses, and no one is obliged to assume my responsibilities for me. I carry my burden with pride and joy.

Table of contents

Every day I recognize and assume responsibility not only for my actions, but also for my emotions, my thoughts, and even my beliefs.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am in control of myself and my environment. I take charge of myself!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am responsible for what happens to me.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

I am responsible for my life.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I am responsible for myself because I have chosen to be so.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I am responsible.

Sublipower, Sublicontrol, Edi Inter, 1989.

I want more responsibilities.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

REST (Also see Sleep)

I always allot myself sufficient time for rest and relaxation. I get enough sleep, and my sleep is deep and regenerating.

I give both my mind and body enough rest, but I don't depend solely on rest to recharge my energy.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am resting peacefully.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Table of contents

As long as I don't move and don't get upset I am in perfect repose.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

REST - MENTAL

When one minute has passed I will count from one to seven. When I reach the number 7 I will emerge from my state of mental rest and feel as if I had slept deeply for two hours.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

RESULTS (Also see Work)

I get more done because I work harder. I like to make that extra effort and obtain better results.

I feel great satisfaction in knowing that I work hard and well. I try to be more efficient and get better results every day.

I improve my results by concentrating my energy and attention on what I am doing. When I have a task to accomplish, it is the only thing I think about.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Infinite Intelligence shows me the best ways to get better results.

Dr. Joseph Murphy, Miraculous Power Attracts Infinite Riches, Godefroy Publications, 1978.

I am getting better results at work every day. I find my work more and more fulfilling.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I let the results come naturally and effortlessly.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

Table of contents

From now on the divine light within me produces perfect results in all aspects of my life.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

RESULTS - SCHOLASTIC

My interest is growing day by day and my scholastic results are improving. I am doing everything in my power to attain my goal.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

RETIREMENT

I am going to make my golden years interesting and meaningful. I'm not going to be a stay-at-home just because I am past 60 or even 70. Growing old graciously means practising being kind, unselfish and understanding.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

REVENUE (See Income)

REWARD

I reward myself with gifts!

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I give myself what I want; I feel I deserve it.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I reward myself when I bring a project in on time.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

Table of contents

(I deserve this reward.)

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I will spend this Saturday doing something I really want to do. I don't mean next month. I mean this Saturday. I will enjoy being alive and being able to do it. I deserve it. There will never be another me. I will spend at least one day a week on ME!

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

When this project is finished I'll buy myself the new suit I've been wanting.

John Clark, Do It Now, Godefroy Publications.

When this job is finished I'll relax and have something to eat.

John Clark, Do It Now, Godefroy Publications.

When I complete this assignment I'm going to reward myself with my favorite —— (specify which food, item of clothing, TV show, movie, sports or leisure activity).

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

If I finish all this today I'll go to a movie tonight.

John Clark, Do It Now, Godefroy Publications.

RICHES (See Wealth)

RISK

I accept responsibility for some of the risks involved in the decisions I make, because taking calculated risks is the only way to acquire experience. Therefore I always ask myself this question: 'If this fails, am I ready to assume responsibility for the consequences?'

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Table of contents

I risk nothing if I fear nothing.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I take risks that may complicate my life, but which could prove to be extremely profitable.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

My path is a luminous one of sunshine and promise. I follow it with joyous anticipation. When I take a necessary risk, I look ahead, anticipating the best, and never turn back.

I reduce the risk of running risks by informing myself as much as possible in advance. Then, once my decision is made, I accomplish each stage with the determination to succeed.

I take positive risks which always bring the desired reward, and contribute to my health and well being.

I only take risks which provide me with even more positive energy, and which contribute to my well being, as well as to the well being of those close to me.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

RIVALRY

I no longer feel any rivalry with my neighbors.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

ROUTINE

I take risks to break the routine. For example, I take unplanned vacations, with no reservations, no maps, telling myself that whatever happens I am capable of dealing with any situation. I apply for a new job, or talk to someone I've avoided in the past because I was

afraid of what might happen. I take a different route to get to work, or I eat at midnight instead of at six. Why? Because it's different, and because I feel like it.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Table of contents

SADNESS

If I sometimes got sad or bored, moody or overcome by dark thoughts, I know that this will no longer happen, and instead of being sad, instead of feeling melancholic and dark I will be joyful, joyful for no reason perhaps, but joyful all the same, just like I got sad for no reason in the past. And what's more, even if I have reason to be sad, real reasons that cause me to feel dark and melancholic, I will not do so.

Emile Coue, Complete Works, Astra Publications, 1976.

I will fill this day with my laughter; I will fill this night with song. I will never have to work to be happy; instead I'll keep myself too busy to be sad. Today's happiness must be enjoyed today.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

SALARY (Also see Work, Value)

My salary is in accordance with my integrity and honesty. I exploit my talents to their fullest. I am appreciated and sought after, and I am indispensable. It's wonderful!

My salary is in line with my spiritual values. My revenue corresponds to my faith and my expectations. My wealth depends on the Universal Substance, and the degree to which people accept me determines the amount of my salary. I accept all the good that comes

my way, here and now.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

SALES

I am learning about the marketplace. I'm widening my base of contacts. The people I call need my products and services.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

Starting today my sales will increase every day.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I will develop my understanding of people, of myself and of the merchandise I am selling in order to multiply my sales.

I will become the top salesman in my department because I am unique.

I will be the greatest salesperson the world has ever known.

I will work harder than ever, pushing my body to its limit, and then I'll do even more. I'll make more contacts than ever, more sales than ever, and I'll make more money than ever.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I'm not just any salesperson. I am an experienced salesperson. I am a champion salesperson.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I am a spiritual magnet, attracting clients who are interested by what I have to offer. The service I give is improving day by day. I am highly successful in everything I undertake. I bless everyone who comes into my store, and treat them well. All these thoughts are now filtering down into the deepest part of my subconscious mind, transforming my life into one of abundance, security and peace of mind.

It's wonderful.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

I am an excellent salesperson. I like to sell and I appreciate the numerous rewards my profession offers me.

I always put in the time required to find new clients. It is natural and easy for me to generate new business.

I always take care of details related to my work. I like everything that has to do with sales, and I always have the time to pay attention to details.

I have a great talent for sales. I am a professional salesperson, and it shows.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am going to start out every morning with the idea that this is going to be a good day for me, that I am going to experience an inner feeling of confidence, a feeling that I can sell almost anything to anyone, that I can make more sales than I ever have before.

I believe in the usefulness and value of what I am selling.

I am going to train myself not to do all the talking, but to listen well, ask questions so that my client is made to feel comfortable and not high-pressured into buying something against his better judgment.

I am going to read and study as many books as I can, pertaining to the psychology of selling and apply whatever knowledge I acquire.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Phone calls lead to appointments; appointments lead to sales; sales lead to referrals.

Each person I contact has the potential of opening up a world of new customers and referrals.

I am taking charge of my inner life by choosing my self-talk. I warm up mentally by focusing on effective self-talk and I prepare

thoroughly for every sales interaction.

I am thankful to have this opportunity to present myself and my product.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

My sales are increasing every day.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

Despite all the people who have come before me, if I constantly try to surpass my own record I will constantly make progress, more sales and lots of money.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

SALES - REAL ESTATE

Infinite intelligence will bring me the right buyer for this house, someone who needs it and will benefit from it. The person will be attracted to me through the creative intelligence of my subconscious mind, which commits no errors.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

SATISFACTION

I always get what I need when I need it.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I tell myself that I did the best thing at the best time.

I am satisfied with myself at the end of the day.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I feel good about myself, and all aspects of my life provide me with great satisfaction.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

SAVINGS

I am fully insured (or protected by my savings).

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Because I know how to manage my money I am always able to save a part of it. I set myself monthly and annual savings goals, and I respect them.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

SCHOOLWORK

From now on I will concentrate totally on what my professors are saying.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

From now on I will concentrate on what my teachers at school are saying... I will not let myself get distracted... I will raise my hand whenever I know the answer to a question... I will participate enthusiastically in all my classes.

I will do my homework conscientiously. I will concentrate on my lessons.

I won't let myself get distracted. I won't do anything else until I've finished my homework, and prepared for the next day's courses.

I'm doing much better in school... I'm happy with my results.

Day by day I am improving my performance at school... I work with joy... and my work is improving day by day... day by day my schoolwork is improving.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

SECURITY

Security is found in myself.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Intimately united with you, my inner Guide, I am stronger and stronger in all situations. I feel secure under your protection and guidance. I am independent of exterior circumstances, things and people. You help me and thanks to you all obstacles become incentives. Your strength is my strength. Your supremacy creates my superiority. You make everything succeed for me, and transform all things to my advantage.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

(I have what I need. I will always have what I need.)

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am at home. I am comfortable as I move into a deeper connection with my true self. I am flexible and secure in my life.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I try as much as possible to eliminate external explanations from my behavior, and external circumstances from my life, my thoughts or my feelings.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

SEDUCTION

I am responsible for keeping myself physically attractive, clean and desirable for my husband.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I am becoming more seductive every day.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

I am seductive.

Charles Baudouin, Psychology And Practise of Autosuggestion, Idegraph Publications, 1990.

I want to make a favorable impression on ——.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

SELF ACCEPTANCE
(Also see Self Confidence, Self Expression)

I continue searching for the best in myself. I accept myself the way I am, and at the same time I do what is required in order to improve myself and gain ever greater satisfaction.

Dr. Maxwell Maltz, Psychocybernetics and Self Accomplishment, published by A Different World Ltd. 1970.

I accept my situation and I feel perfectly good about it.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

I accept all my feeling as part of myself.

I like myself and accept myself the way I am.

I don't have to try and please anyone, I please myself and that's what counts.

I like myself the way I am, and I feel better and better.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I have the right to be myself.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I like who I am, and live each day in the freedom of self acceptance and self confidence, and with the joy of expressing myself fully.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I like and accept both my body and my spirit.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

Table of contents

I exercise my right to self acceptance.

I am learning to accept myself.

I accept myself for who I am, and I don't try to imitate anyone else.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I stop concerning myself with social popularity. I am working on accepting myself.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I accept myself and I accept others.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

(I stand tall and shine brightly. I appreciate myself and respect myself.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I'd make a bad somebody else, but I'm the best 'me' there is.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

SELF AFFIRMATION

Here I am. I have as much right to be here as anyone else. I matter.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I have the absolute right to be who I am, to think what I think, to feel what I feel, and to want what I want.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, 1978.

I have the right to say 'NO' to someone without losing their love or friendship.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

Table of contents

I have the right to affirm myself, and it is to my advantage to communicate. I have the right to ask for what I want. I give others the freedom to say 'yes' or 'no.' If I don't express what I want, no one will know and I'll always be dissatisfied.

I have the right to have needs, and to fulfill them, as long as it doesn't harm anyone.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

SELF AWARENESS

I know that I never waste time asking myself what other people think about me. If I want to know, I simply ask them! In this way I obtain the information I need, and I also feel I know myself better.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

Self awareness helps me see that unhappiness is not a part of the Real Me, and is therefore not necessary to endure. Relaxation and tranquility are the result of self understanding. I make self understanding my goal each day.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

The more I know myself the more I love myself.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am convinced that I am a center of pure self awareness; I am a center of willpower, capable of dominating and influencing, using all my psychological and bodily functions.

Dr. Robert Assagioli, Psychosynthesis, Epi, 1965.

I am MYSELF, a Center of pure consciousness.

Dr. Robert Assagioli, Psychosynthesis, Epi, 1965.

Table of contents

SELF CONFIDENCE (Also see Self Esteem, Self Control, Competence, Depression)

My self confidence grows stronger every day.
I have confidence in myself.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I'm more confident of myself every day. I'm getting stronger. I AM ME. I love who I am. I can handle any situation.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I have more and more confidence in myself every day.
I often tell myself: "It's easy, you can do it."
I speak clearly and articulately.
I decide what I really want, and act accordingly.
I can now depend on myself.

Suplipower, Subliconfidence, Edi Inter, 1989.

I have confidence in my abilities.
I trust in myself and believe in my innate wisdom.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I have confidence in my inner powers.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I have confidence in myself and in the future.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I am confident in myself, in others and in life.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

I have confidence in you.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

Table of contents

I have faith in myself.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I have everything I need to be happy, here and now.
Everything I need is already within me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I have absolute confidence in myself... and I feel at ease under any circumstances... even in public I feel no apprehension... all my faculties remain intact... because I am sure of myself.

Marcel Rouet, Motivating Your Subconscious, Dangles Press, 1982.

I like who I am.
I feel good about myself.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I will reach the highest summits through the power of this newfound confidence in myself.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

I trust my inner knowing in all situations.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I believe in myself.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I develop my self confidence and have faith in my clear conscience, letting it guide me without fear as my confidence grows.

I avoid negativity. I help myself live more fully in the present moment, by keeping in mind the confidence I felt when I succeeded in the past.

I have confidence in my abilities.
The more I am myself, the more I believe in myself.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

I am gradually gaining more and more self confidence.

Alan Houel, How To Deal With Difficult People, Godefroy Publications, 1990.

I congratulate myself on each new success.

Christian H. Godefroy, Self Confidence: The Subliminal Method, Godefroy Publications, 1987.

I feel calm and sure of myself.
I am confident.
I build up my confidence.
My approach is relaxed, calm, tranquil and confident.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I prefer not to be a chronic worrier. I am going to manifest faith in everything I do - faith in the thought that nothing will happen. My faith in a Higher Power will always come to my rescue and provide me with peace of mind.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I am the greatest miracle in the world!

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I am a positive person and I have confidence in myself. I emit positive vibrations. If you look closely you can even see a halo around my head.

If I have suffered from a lack of self confidence in the past I don't anymore. Today is an ideal day for getting rid of all those false beliefs which have paralyzed me up to now.

It doesn't matter what I demand of myself, I can do it, I know I can.

Today, right now, I am capable of offering myself the gift of solid, unshakable self confidence.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Others appreciate my company. I have self confidence and self

esteem. I like myself, and others find that irresistible.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Now I am very sure of myself. I have complete confidence in myself, and confidence in life. I know my life will be very beautiful, because it is a pleasure to use this newfound confidence to work on myself, and to give to others.

Charles Baudouin, Psychology And Practice of Self Suggestion, Idegraph Publications, 1990.

I trust and follow my 'gut feelings' at all times.

Louis Proto, Self Healing, Piatkus Publications, 1990.

SELF CONTROL (Also see Objectives, Calm, Self Confidence, Willpower)

I'm in control of my body, now.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I am in complete control of my thoughts and faculties. I use them to attain my objectives.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

I command and then I obey my own command.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

Every day I am in better control of my habits.

Sublipower, Sublicontrol, Edi Inter, 1989.

I become a mental magician when I control not only what I want to do, but also what I don't want to do.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

All my thoughts lead to acts.

Table of contents

I control my thoughts, and I control my world.
My thoughts make things happen.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I am in perfect control of my mental impressions.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

Every day I expect to gain greater control of my feelings and my thoughts.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

I am in control of my thoughts, ideas and feelings.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

From this moment on I am ready to control the personality which awakens in me each morning. I will control my emotions by acting in a positive way, and it is by controlling my emotions that I can control my destiny.
I become my own master.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I feel in complete control of myself.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Editions, 1982.

I take full control of my mind and my emotions.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I am the master of my life and my own authority.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am in control of my gestures, my thoughts, my emotions. I am free of the things I inherited, free of my impulses which I have brought to heel. I am free of all restraints. I follow my chosen path calmly and firmly.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

Table of contents

SELF CRITICISM

I can calmly review and critique my performance and then make improvements without calling myself names or blaming my performance on someone else.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I try to be myself as much as possible. I don't criticize myself that much. After all, you have to live a little.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

When I make a mistake I don't scold or berate myself with left-brain criticism. I develop an affirmative statement, about 5 words in length, describing my correct performance, in the present tense. I relax, and listen to myself stating the affirmation, and I visualize the accompanying action and feeling.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

SELF ESTEEM
(Also see Self Confidence, Self Love)

Increase my self esteem and I will be better respected by everyone.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I esteem myself and recognize that I am capable of exploiting the potential of every aspect of my being.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

(I have enough self esteem to accept what I think is the best.)

Louis Proto, Self Healing, Piatkus Publications, 1990.

Table of contents

I speak about myself in a positive way.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I am capable of handling a position of authority. But to do that I must start by improving my self esteem.

Charles de Liguori, Hypnotism, de Vecchi Publications, 1975.

I am a human being. I am worthwhile just because I exist and try to survive. I take care of myself. I take myself seriously. I correctly take myself into consideration first in all matters. No one is any more or less worthy than I.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I'm a wonderful person just the way I am.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I am really unique. I like who I am and I feel good about myself.
I always wanted to be someone, and now I know that I am.
I prefer being me to any other person in the world.
I like what I feel, I like the way I think and act.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I reject thoughts of "I'm this way..." or "I'm like that..." as much as possible.

I let the people close to me know that I intend to try to get rid of certain labels which are attached to me. I choose the labels I want to eliminate first, and ask them to tell me whenever I continue using them.

I make a decision to act differently than I did in the past. I set myself the task of eliminating one label per day.

I tell someone I can confide in about my decision.

I ask him / her to tell me every time I seem to fall back and start hiding behind a negative image or label.

I try to get rid of all these labels day by day.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Table of contents

The farther I get away from my old habit of underestimating myself and my abilities, the more excited I become about the prospect of living my life in a really fulfilling way.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, 1978.

I sometimes fail on purpose. If I lose a tennis match or a game of cards, does that in any way impair my integrity? Or do I remain an estimable individual who is simply participating in an activity for the fun of it?

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

SELF EVALUATION

I cease evaluating myself according to other people's standards. I use my own personal standards.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

SELF EXPRESSION (Also see Self Acceptance, Self Confidence, Health, Happiness)

Every time I express my personal opinion I do it as precisely as possible, speaking in the first person. I explain my expectations clearly, and I encourage others to do the same.

I express my opinions frankly, making sure not to be aggressive towards others, and I do not hesitate to express emotions like satisfaction or disappointment when I feel it is necessary.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I speak with gentleness and love. I exhale only the good.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

Table of contents

I know how to express myself in a dynamic way.
I express myself freely, fully and easily.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am capable of expressing everything I feel, even if I don't know ahead of time what these feelings will be.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I am conscious of myself. I believe in my ability to express myself fully on a physical, spiritual or emotional level.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I want to express my feelings.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

When I express myself I pay attention to other people's reactions (verbal or non-verbal) in order to make sure that I am being understood.

If necessary I know how to talk in very simple language, using images, gestures and illustrations to get the point across.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

SELF FORGIVENESS

I am a human being, worthy but imperfect. I am like you. We are both just trying to survive. When I hurt you I was just trying to do what seemed best for me at the time. If I had then the awareness I do now, I would have chosen differently. But at the time, I could only do what I did. I understand that I hurt you, and I want you to know that hurting you was not my goal.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

Table of contents

SELF HEALING

I trust that I have the power to heal myself.

Louis Proto, Self Healing, Piatkus Publications, 1990.

My healing power is getting stronger day by day.

Sublipower, Subli Health, Edi Inter, 1989.

SELF HYPNOSIS

I feel fine, perfectly fine. This session is over, and I'm going to resume my normal activities. I am in full possession of all my faculties.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Editions, 1982.

Now I am sleeping. So calm, so serene! Everything is tranquil. My mind is free, calm. This feeling of calm encompasses everything. I feel it in my legs right up to my chest, in my arms and in my head.

Charles de Liguori, Hypnotism, de Vecchi Publications, 1975.

My ability to practise self hypnosis is improving from session to session. My eyelids are heavy... they're getting heavier and heavier... They are so heavy that soon I will have to close them, and when I do I will enter a state of self hypnosis.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

SELF IMAGE

As much as possible I am going to think: 'Success is inevitable, and I am already the kind of person I want to be." I am going to feel and act like that new person.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

I have a positive self image.

Claude Arpel, Self Confidence: The Subliminal Method, Procauvi Publications, 1986.

I encourage myself to choose and stick to the path that gives me a healthy self image.

To improve my self image I learn from the school of successful failures.

Now that I like myself the way I am I consolidate my self image every day.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I allot myself some free time each day to work on a better me.

This decision to improve myself each day gives me the strength to improve the image I have of myself. This is going to be a good day, and my self image will be even more positive.

I concentrate on strengthening my self image.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I know that my dominating thoughts have to go through a number of stages before they become reality; that's why I devote 30 minutes every day to thinking about the man I intend to become, creating a precise image of that person.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

When I visualize myself "in the present" as if I were already accomplishing one of my goals, I make certain my visual image is as I would see it out of my own eyes, not "watching me do it" through the eyes of a spectator.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

All the negative images and attitudes I have about myself are now dissolved. I love and appreciate myself.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

SELF IMPROVEMENT

I believe that I am capable if improving my life by improving my way of thinking.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I dedicate a little time each day to improving myself as a human being.

I make a daily resolution to improve myself.

Progressing towards my goals implies a powerful desire to improve myself. In other words I want to become better, I want to strengthen the image I have of myself. I search for new horizons, I want to give my life a deeper meaning.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I love and approve of myself totally, every moment.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I stand tall and shine brightly. I appreciate myself. I respect others

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I'd rather be me than anyone else in the world.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I am a unique and valuable human being. I always do the best I can. I love (like) myself, mistakes and all.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I see what is positive in me and I appreciate it.

Sublipower, Subliconfidence, Edi Inter, 1989.

Table of contents

SELF LOVE (See Love Of Self)

SELF PUNISHMENT

I reward myself, never punish myself.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

SELF REJECTION

I am trying to feel better about myself.
I offer myself my own friendship.
I will never abandon my sense of self.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

SELF RESPECT

I respect myself and appreciate myself.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

SELF SUFFICIENCY

I do not have to try so hard. I have enough. I am enough.

Louis Proto, Self Healing, Piatkus Publications, 1990.

SENSATION

I accept and experience all my feelings. (wrong section!)

Shakti Gawain, Reflections In The Light, New World Library, 1978.

Table of contents

I am in control of my way of sensing things.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

SENSE OF SMELL

My sense of smell is improving fantastically... I am able to perceive odors with amazing accuracy... over unusual distances... I am surprised at the number of different odors I can smell... and their intensity... My sense of smell is developing... I can sense even the most subtle odors... my sense of smell is becoming very developed.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

SENSES

All my senses are sharpened. My vision, sense of smell, hearing, and even my sense of touch are more awake than ever.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

SENSITIVITY

I concentrate my mental energy, my time and emotions on what concerns me.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I am a magnificent, sensitive, likable, creative and valuable being.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Table of contents

SENSUALITY

I express my love and my sensuality in many ways.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

SERENITY

I love the tranquility and peace I create in my mind.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

From now on I will feel deep inner peace and serenity.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am calm, tranquil and serene.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I always remain in a state of serene and peaceful balance because I know that God will always provide the right solution to any problem I encounter.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

Nothing can disturb my serenity.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

SERVICE

I try to be useful, pleasant, set a good example and help others on all occasions.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

I can get everything I want out of life if I help other people get what they want.

Table of contents

If I do something for someone who can't help me in return, I still get more than I give.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

SETBACKS

I never allow a setback to make me physically ill.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am coming back stronger day by day!

Now life is starting to look like a friend instead of the enemy it appeared to be before. The more I get into the habit of not expecting any help from anyone, except from my inner Guide, the closer I get to realizing my hopes and fulfilling my desires, while my fears of powerful adversaries and those who wish me harm disappear.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

SEXUAL BLOCKS

Women (or men) like and accept me the way I am.

Whether or not my sexual activities always work out is of no importance.

My sexual life is enjoyable, and does not cause me any tension or anxiety.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

SEXUAL DESIRE

I feel more and more sexually aroused. My whole body vibrates with desire and love. I am more and more desirable, and I find my partner —— (name) more and more desirable. I am irresistibly at-

tracted to him / her. I find my partner charming and sensual. My desire grows stronger day by day, and all aspects of my sexual life are improving.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

SEXUALITY

I give my fantasies the freedom to live in my mind and strengthen the joy and beauty of my love life.

I take the time to let my partner know about my needs and thus allow him / her to participate in my sexual fulfillment.

The intimate sexual expression of my being constitutes one of the greatest joys of my life. I give myself the freedom to experience pleasure with my whole being.

I love touching. I have no trouble expressing myself physically with the person of my choice.

I know it is good for me to express my sexuality, which constitutes a positive and important part of my overall well being.

I am a GREAT lover! I know how to relax and satisfy both myself AND MY PARTNER, simply by being myself.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I know that the sexual act is an expression of love on a physical level; I know that it is natural for me to be a warm, sensual woman, and there is absolutely nothing wrong with that.

I know that sexual pleasure is a right given to us by nature; I know that the union of two beings is creative in ALL senses of the word; it is the earthly expression of oneness with the total BEING through union with another.

I am constantly more enthusiastic, both actively and passively, when I make love with my husband; it's easy for me to close my eyes

and savor all the pleasant sensations associated with the act of love; I am relaxed and not in a hurry; I abandon myself to the tide of pleasure without even trying to think about having an orgasm.

Clement Auger, Transform Your Life Through Autosuggestion, Heritage Publications, 1979.

I am comfortable with my sexuality.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I am free to express my sexuality as I choose. I enjoy my sexuality tremendously.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am going to enjoy the sex act; I can and will have complete sexual fulfillment whenever I have sexual relations with my spouse.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

When I take my partner in my arms an irresistible desire overcomes me... I feel ready to move on to the act of love... physical love is completely natural... and when I make love I am relaxed and free of complexes.

I am happy with my partner... just thinking about it awakens my physical desire... I am ready for love.

*Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.^*

My body is alive.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

SHARING

I give to others. Doing so makes me feel good. I practice my ability to share with others.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

I give and I share.

Sublipower, Sublisuccess, Edi Inter, 1989.

I don't ask favors from anyone, but I have the privilege of sharing the blessings I have received with those who want to partake of them.

I share a part of the good and desirable things I possess with others.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

The more I have the more I want to give.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

The richer I get, the more I have to share with others. The more I receive, the more I have to give.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

As I give to myself, it is fun to give to others.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

SHORTNESS OF BREATH (Also see Breathing, Smoking)

I am catching my breath.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

My breathing is regular and calm.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

Table of contents

SINCERITY
(Also see Benevolence, Warmth, Honesty)

I will be sincere with myself, and not create any more illusions or overestimate my qualifications or possibilities. I will also be sincere with other people.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

I am warm, sincere, honest and authentic. I am all this and much more. All these qualities are a part of me! I love who I am, and I'm happy to be me.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

My attitude towards friends is sincere, and my sincerity is real. Since I like, appreciate and accept others, benevolence and sincerity are a part of my personality.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

SKIN

My skin is becoming more and more beautiful and radiant. My face is glowing with health. My skin is improving day by day, becoming as fresh as a rose. My skin looks younger every day as it regains its freshness, elasticity and radiance.

My skin is getting softer and smoother.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

SLANDER

The storm will soon blow over, and in any case slanderous words

Table of contents

cannot hurt me.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I never talk about anyone except to express my esteem. I never slander anyone.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

SLAVERY

I flee from servitude; instead I look for ways to become involved. In this way I move from isolation towards happiness.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

SLEEP (Also see Rest)

Every night brings me deep restorative sleep. I sleep, I sleep (inhale), I am deeply asleep, I am deeply asleep (exhale).

I I fall deeper and deeper into sleep.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Tonight, tomorrow night, and on all nights to come, as soon as I wish to sleep I will fall asleep, and sleep right through the night until the next morning, until the hour I set for myself to wake up. I will sleep deeply, calmly, tranquilly, and when I awaken I will feel completely refreshed, happy and ready to greet the day.

Emile Coue, Complete Works, Astra Publications, 1976.

I sleep perfectly every night, and wake up fully rested.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

I slide gently into a tranquil sleep.

Tonight I will feel calm and relaxed.

Table of contents

I will fall asleep with a calm body and mind.
My sleep will be deep and calm.
My sleep is calm and restful.
Every night I will sleep more and more deeply.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I am falling asleep in harmony and plenitude.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

I feel good and I am secure when I sleep.
I love sleeping in a natural, normal way.
I never resist sleep, whatever time of day it is.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I will soon fall calmly asleep... my muscles are relaxing... my mind is calm... just one minute more and I'll be asleep... the thought of sleep fills my mind... I am going to sleep.

Henry Durville, Hypnotism and Suggestion, Durville Publications, 1938.

I will get into the habit of going to bed at approximately the same hour each night.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Just before falling asleep I flex and then relax every muscle in my body, starting with my feet and moving progressively upwards. I rid each muscle of tension.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

When I wake up I will feel full of energy. I'll be ready to do a thousand and one things, because I will have slept wonderfully well. I'll be in a good mood, happy, enthusiastic. I'm sleeping better every night. I'm sleeping better and better. Everything is fine. I sleep deeply. All is tranquil.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

Table of contents

Now I will fall deeply asleep until tomorrow morning... I am deeply asleep... deep deep sleep.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

My sleep is calm, deep, restful, restorative. It lasts right through until morning, from the time I want to go to sleep to the time I want to wake up, and I wake up exactly on time. I must do this.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

I sleep deeply... As soon as I get into bed I forget about all my problems... I fall asleep right away. My sleep is healthy, natural and restorative... In the morning I feel rested and happy to face the new day.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

When I wake up I shall feel totally refreshed, energized and well.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Every night, from the moment I want to go to sleep to the moment I wake up, I sleep deeply, calmly and tranquilly.

Sublipower, Sublihealth, Edi Inter, 1989.

SLIMNESS (Also see Charm, Looking Good, Health, Weight Loss)

Female goal: I enjoy weighing — (specify your ideal weight) pounds and looking slim in my new bathing suit.

Male goal: I weigh a trim, athletic —— (specify your ideal weight) pounds and enjoy exercising every day.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I know that my weight depends solely on the way I see myself, and what I tell myself. I am learning to see myself as slim, in shape,

Table of contents

healthy and very happy.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am now slim.
I'm getting slimmer every day.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Everything I eat makes me more beautiful, healthy, slim and charming.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

SMILING

I belong to the compliment and smile club. When I smile at someone and the person smiles back I automatically feel better.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I remind myself to laugh a little every day.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I smile at life.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I am going to smile at least three times a day.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

SNACK

I have the willpower and assurance to control myself. I decide what I eat, and when I eat!

I will not eat between meals out of habit any longer, nor replace meals with impulsive snacking. From now on I will eat regular meals

of just enough healthy and nourishing food.

I never eat between meals to soothe my anxiety, my nervousness, my tension or solitude. I solve my problems by facing up to them and working on them. But I never replace a solution with a snack.

I no longer eat just to pass the time, or while I'm reading, working, talking or watching TV.

Every day I manage to eat less and less between meals. I am developing the healthy habit of keeping my weight down. I am a winner, and I am proving it every day!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

SOBRIETY (Also see Alcoholism)

I will not drink today.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am acquiring the habit of sobriety, and I can control myself.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I will never allow drinking to become a problem; I will use discretion as to how many drinks I can have or handle.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I don't drink any more and I feel better both mentally and physically every day.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I am sober. I don't drink alcohol. I'm making a success out of my life. I am happy and serene.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses , Godefroy Publications, 1989.

I will surmount life's frustrations without drinking. I am sober and happy.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

Table of contents

Sobriety and peace of mind are now mine to enjoy, and I am thankful.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I am affirming my sobriety more and more strongly every day. I am becoming completely sober. My Guide is helping me faithfully, removing temptation from my path until I have completely vanquished this harmful habit. I am strong, I am joyous, I am free! I am strong, I am joyous, I am free!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

SOLITUDE

I like myself and I naturally attract loving relationships.
My relations with people are solid and affectionate.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I treat others with benevolence.
The fact of not criticizing myself so much any more means that I am being fair with myself, and that I am overcoming my solitude.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I find a partner who likes me, and who is cultivated, sincere, spiritual, faithful, happy, and at peace with him / herself. We feel an irresistible attraction for each other. In my life only things which are the result of love and truth can occur. I gratefully accept the fact that there is someone out there who is the ideal partner for me.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I am a center of creative intelligence, love and good grace. I am no longer alone, because Divine Love is flowing through me, making all people feel affection for me. I attract those who are in the same

Table of contents

state of mind as I am.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I want to be loved.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

If I am always myself I will be sure to meet people who love and respect me.

Dr. Martin Shephard, How To Overcome Your Cares, Marabout Publications, 1973.

SOLUTIONS

I now see this point in a completely new light.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

There is a solution, and I'm going to find it.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I gather as much information as possible concerning my problem. Knowing the problem is the key to finding the right solution.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I dare to be aware of myself! I am completely capable of finding the solution I need.

I stop running around looking for solutions. I simply remain calmly receptive to my Real Self; it always provides a solution to my problem.

Solution, solution, solution...

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I am now being guided towards the perfect solution to this problem.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

At night I ask my subconscious to find a solution for a problem I cannot solve, using the 'blackboard' technique. The next day when I wake up, or sometime during the day, the solution appears in my mind.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

Dear God, I know that You have the answer, and so I submit my question. It is in Your hands, and I will patiently await the solution.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

SORROW

Today I start a new life. I will forget everything that has caused me pain, sorrow or shame in the past.

Orison Swett Marden, The Joy Of Living, J.H. Jeheber Publications.

SOULMATE

From now on I won't be afraid to approach a person who attracts me. I am strong. I am ready to love and be loved.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I will attract a man who is honest, sincere, loyal, good, faithful and prosperous. He is peaceful and happy. These qualities are now engraved on my subconscious, and as I assimilate them, they become a part of me. I know that the law of attraction is irresistible, and that I will attract the man who corresponds to my subconscious convictions, and deepest feelings.

Dr. Joseph Murphy, The Miracles Of Your Mind, Dangles Press, 1984.

I will meet my ideal partner, someone whom I love deeply, and

who loves me deeply in return.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

My ideal partner is waiting for me, and I will meet him (or her) in the right place, at the right time.

Vera Pfeiffer, Positive Thinking, Element Books, 1989.

SPEECH MAKING

My voice will be strong enough for this speech, and my positive thoughts are now helping me systematically improve my speech.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

When I speak my voice is clear, firm and distinct. I breathe deeply and regularly. My body is perfectly calm. When I look up from my notes I look at my audience, making sure to make eye contact with as many people as possible.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, 1978.

SPENDING

Every dollar I spend comes back to me multiplied.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

I deserve to spend money on things I desire.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I earn more money than I spend.

Christian H. Godefroy, Prosperity: The Subliminal Method, Godefroy Publications, 1987.

I never spend impulsively, nor do I feel obliged to spend simply for the pleasure of spending.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

SPIRITUALITY

God is always good.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

He's got the whole world in his hands - and that includes ME. I surrender to the will of God.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I believe in the existence of a Higher Power, A Force of Love, that is eternal, that puts meaning into our lives.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I am one with my supreme nature and my creative power is infinite.

God is in me and manifests in the world through me.

Divine light and divine love flow through me and radiate from me, lighting up everything around me.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am the king's son.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

SPORT

I weigh a trim, athletic —— pounds, and enjoy exercising every day.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I do a lot of sports.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am capable of taking part in sports.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

Table of contents

Taking take three deep breaths through my mouth gives me the strength to perform great sporting feats.

My performance in sports is improving day by day.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

STAGEFRIGHT (Also see Fear Of Exams)

Starting today I will no longer experience stage fright.

I have overcome it once and for all.

I am confident, I know my strong points.

I feel relaxed, calm and serene, just like when I —— (describe a personal and pleasant event).

From now on whenever I have to write an exam I will feel as comfortable as on the day I —— (insert your personal and pleasant experience).

My stage fright is disappearing gradually.

I have overcome it, and I am a stronger person.

I have confidence in myself. I know I am going to succeed.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I have the ability to concentrate on what I am going to accomplish before the fact... I think of nothing else... and I concentrate my energy... also... I never feel any anxiety... even at the idea of speaking in public... I am never afraid... I am never anxious... I am sure of myself.

So I speak and act with vitality... knowing that my stage fright has disappeared... my mind is perfectly clear.

I am not afraid to speak out.... whether I'm with friends... or in front of a group of people I don't know.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

Table of contents

I sing wonderfully well.
I an tranquil, serene, calm and full of confidence.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I am calm and relaxed.
I can deal with any situation.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I am strong! I will triumph! I will succeed!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

(I'm ready. I have already succeeded in the past.)

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

STAYING IN SHAPE

I'm in great shape.

Charles Baudouin, Psychology And Practise of Autosuggestion, Idegraph Publications, 1990.

I am conscious of my body and its importance vis-a-vis my overall well being. I always to what is necessary to stay in top shape.

I do everything I have to do to stay healthy and in shape, and maintain my well being.

I enjoy the challenge of getting and staying in shape. I accept the challenge, and nothing can prevent me from succeeding.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

STOMACH ACHE

Rays of healing energy flow through my stomach. Everything is fine and my stomach is functioning normally.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

Table of contents

STOOL (See Digestion, Appetite)

STRENGTH (Also see Courage)

I decided I was strong.
I dare to be healthy and strong.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I have faith. Strength will come. It lies sleeping in me. I make an effort, I struggle. My hidden strength will surge to the surface. I persevere, knowing that my strength is growing, is transforming me. Weak as I am now, I will become strong and robust. I stop my crying and hold my head high. I reject negative thoughts. I don't believe my strength will abandon me.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I seek out and use my personal strength.
The greatest contribution I can make to others is to be strong within myself. People need my strength.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I feel stronger with every moment that passes.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am strong.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

I am very strong. Stronger than I've ever been in my life. Stronger, much much stronger. I am surprised at how strong I am.

Dr. Maxwell Maltz, Psychocybernetics, Godefroy Publications, 1978.

STRENGTH - INNER

I feel a more and more irresistible strength growing within me.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

STRENGTH - PHYSICAL
(Also see Losing Weight, Staying In Shape)

Every time I say the word X (choose a word) three times the strength of my muscles will immediately double. My muscles will be perfectly nourished by my blood, giving me perfect muscle tone.

Valery Sanfo, Practical Guide To Self Hypnosis, de Vecchi Publications, 1987.

STRESS

As soon as I lie down I feel completely calm. My problems disappear. My body and nerves relax.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

I approach life in a relaxed way.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Stress slides of me like water off a duck's back. I relax with each breath I take. If a stressful event occurs I take a deep breath and regain my inner calm.

Christian H. Godefroy, The Creativity Method, Godefroy Publications, 1992.

I create pressure in any situation by what I focus on and what I say to myself. I can remove pressure by concentrating on doing the best job I can right now and learning all I can.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

Table of contents

I release any thought that subjects me to any stress whatsoever.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I never allow destructive tension to accumulate in me. I overcome stress by attacking its causes. I make a conscious effort to guard myself against stress and tension by maintaining a long-term perspective, by establishing simple goals for myself, and by taking appropriate measures such as exercise and relaxation.

I feel the peace that comes from relaxation and contentment. Every time I repeat these words to myself I become more confident and more relaxed, and I free myself from all the stress inherent in all areas of my life.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am happy because my attitude is making my life more joyous and serene. I am free of stress.

I am a totally individual being, with my joys and heartaches. I am healthy and good. Every time I feel some tension building up inside me I take refuge behind my mental shield. Stress cannot break through my shield. I feel protected, night and day. I see all kinds of tension smashing into my mental shield without being able to break through it.

Tension cannot affect me. I am protected against stress. Other people may become upset, but I do not. I go about my life tranquilly, at my own pace. Everything is fine.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I am completely calm... relaxed... I will concentrate on everything I do, and I will remain calm and relaxed inside... I do my work calmly... nothing can upset me... I know I can deal with any situation... and this certitude fills me with strength and energy.

I am sure of myself... It is a marvelous feeling to be able to say: 'I will triumph! I will realize all my projects while remaining calm and

relaxed.'

Day by day my sense of inner calm and self assurance grows stronger.

I am more and more calm... I feel completely comfortable with my situation. My heart beats calmly, regularly... My circulation is regular... any effort only makes me stronger... my health is improving day by day. I feel completely good about my situation... Day by day, and in all ways, things are improving. I feel better and better.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

STUBBORNNESS

I am learning to be mentally supple in all situations.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

The universe is running its course to perfection. I have no need to be stubborn.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

STUDIES

Today I learn things easily. My memory is functioning perfectly. If I have an exam I will remain calm and relaxed, and find the right answers. My marks will improve day by day, and I will take pleasure in learning.

Professor Kurt Tepperwein, How To Learn The Painless Way, Godefroy Publications, 1983.

May God show help me complete my —— (specify which) year of university, in harmony with divine law and order.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

Table of contents

I do my homework and learn my course material.
I participate in courses whenever I can.
Learning interests me.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I know I can do this work since I am studious... may I also be wise.

Professor Robert Tocquet, The powers Of The Will, Godefroy Publications, 1989.

When I study I do nothing else! I concentrate my mind and energy on the object of my study.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

When my homework is done I prepare for the next day's courses.
Everything I undertake I do conscientiously. So I also do a good job preparing for tomorrow's courses.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

My thoughts are organized, and my comprehension is growing.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

If I draw a blank on a certain subject, like algebra for example, it's just a natural result of the choices I made up to now. If I decided to make an effort to study these subjects by putting in the necessary time, I would certainly make progress.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

STUTTERING

Every time I say the word —— in my mind I will feel calm and relaxed, the muscles in my windpipe will remain completely relaxed and I will speak very well, with confidence and ease. It will be easy for me to communicate with others.

Valery Sanfo, Practical Guide To Self Hypnosis, de Vecchi Publications, 1987.

Table of contents

I can now express myself correctly... correctly... with no problem whatsoever.... correctly.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

SUBCONSCIOUS MIND

I am united with the infinite wealth of my subconscious mind.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

SUBCONSCIOUS PROGRAMMING

I program my subconscious in order to realize my desires... my projects... my subconscious gives my imagination the means to attain my goals.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

SUBMISSIVENESS

I remind myself that habit is not a reason for doing something. Just because I've always had a submissive attitude doesn't mean I have to continue being submissive.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

SUCCEEDING

I have the right to succeed and get the best out of life.

Sublipower, Sublisuccess, Edi Inter, 1989.

I have everything I need to succeed.

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

Table of contents

I do everything in my power to succeed at such and such an endeavor.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

I am determined to succeed.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

I am succeeding.

I know I will continue to succeed, as I have done up to now. I am more determined than ever to reject anything that can become an obstacle to my success, anything that is unhealthy in my life.

Everything I do succeeds.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I am happy and successful because I have decided to succeed.

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

Today I am going to succeed wonderfully well in everything I do.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I want to be happy and succeed.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I want to succeed.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

The greatest secret in the world is that I just have to be a little better than mediocre... that's all I need to succeed.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I succeed at everything I do.

Christian H. Godefroy, Prosperity: The Subliminal Method, Godefroy Publications, 1987.

SUCCESS (Also see Wealth, Happiness, Self Love, Peace, Prosperity, Health)

Stocks may go up or down on the exchange, but in life my abilities are constantly helping me reach new heights of success.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am now willing to be happy and successful.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I have the certitude that I am making the best use of what I have. With this in mind, and armed with the desire and determination to see things through, chances are very good that I will attain success.

I believe I was put on earth to succeed, to be happy and healthy, and to attain useful goals.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I know that my future is one of progress, creative endeavor and financial security. This in no way depends on external factors. It depends on my mind, and I am grateful only to God, in the present as well as in the future.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I make my dreams a reality.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I make sure to take advantage of events.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

I am attaining a higher level of existence as I become more receptive to superior powers.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I turn to you, my inner Guide, with my prayer. I am confident

you will assist me in everything that has to do with my inner progress and my exterior success, both today and all the days to come.

You are helping me attain the goal I so ardently desire. I thank you for your support. You are the power that makes my desires a reality. I am one with my desires and I attain success. With your help I always attain success.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I will succeed by attracting the forces necessary for my success, by always being the first to be of service, and by instilling in my peers a desire to help me.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

I will succeed because I am able to handle any situation.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I am succeeding more and more easily at everything I do.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I succeed very well in certain areas.
I am in the process of becoming a successful person.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I am strong, I feel better every day, and I will succeed at what I do.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

Life is beautiful. I am succeeding! I am succeeding!

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

My mind is constantly opening to new ideas. I use all the means at my disposal to achieve success. I am open to new ideas which lead to my success.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

Today is God's day. I choose to let harmony, success, prosperity, abundance, security, and divine justice guide my actions.

Dr. Joseph Murphy, Miraculous Power Attracts Infinite Riches, Godefroy Publications, 1978.

I accept the responsibility of being alive. I concentrate my thoughts on what I want to do and who I want to be. I know I can obtain all the good things I can imagine. I rid my subconscious of all negative thoughts which can prevent me from achieving what I desire. They do not exist. God's power over me is now complete, and I progress on the path of right thinking. I await the results of my actions, and I give thanks.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I grant myself success, happiness, love of self and love of others.

Sublipower, Subliconfidence, Edi Inter, 1989.

I look to my past successes.
I invite success and not failure.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I only envisage the possibility of success... and so I succeed.

W. Clement Stone, Passport To Success, Godefroy Publications, 1962.

I persevere until I attain success.

I am more and more successful in everything I do. I am able to attain all my objectives more and more easily. I am becoming a successful person. Life smiles on me. Everything I do succeeds. I move from one success to another, more and more easily.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am the personification of success. I do my work, allowing God to guide me in all my undertakings.

The divine wisdom in me will reveal the solutions and means I need to attain professional success.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

Table of contents

I succeed when I attain my objectives, but I also succeed each day, on the voyage of life. I know that success does not only exist at my final destination, but at every step along the way.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am aware of my worth. I know I merit success in everything I undertake, and that I also merit the rewards my success will bring.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am on the road to success.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

All the powers of my subconscious are working for my success.

Negative thoughts and doubts are erased and replaced by new ideas and positive thoughts of success.

I have all the qualities I need to succeed.

The creative power of the universe is at my disposal to help me succeed.

I like having a goal and persevering until I succeed.

Each day I take a step closer to success.

I have confidence in my intuition, which is guiding me towards success.

Sublipower, Sublisuccess, Edi Inter, 1989.

SUCCESS - FINANCIAL
(Also see Marketing, Happiness)

Financial success comes to me easily and effortlessly.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Table of contents

SUFFERING

It's going away, it's going away...

Emile Coue, Complete Works, Astra Publications, 1976.

(I choose to ignore death. I live my life based on what I think and feel.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

When I face up to my suffering and treat it, my present suffering frees me from future suffering. It must be confronted.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

SUGGESTION - POSITIVE

I have a positive attitude towards each one of these suggestions, and I am convinced that they will make me a happier person.

The suggestions I am receiving will help me and I will listen to them. These suggestion will influence me every day of my life.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

SUPERIORITY

The person opposing me is no trouble at all. S/he knows nothing about my expertise or experience on the subject, and so I have every advantage over him / her. I am strong! Much stronger that s/he imagines.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

Table of contents

SUPPLENESS

Every breath I take makes me more supple.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

My bones must always stay healthy, and my joints, especially those in my spine, should move freely and easily.

I make sure that my muscles are healthy, able to respond rapidly, and always ready to function.

Christian Godefroy, Mental Dynamics Seminar, Godefroy Publications, 1978.

SUPPORT

Everyone around me helps me to grow.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

SURGERY

The operation is a total success. Everything is healing as it should.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Nothing is going to disturb my recovery. I'll have no complications, and I will feel good because I know it's all over.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

SURPASSING

Today I will surpass everything I did yesterday.

It isn't important to surpass others; the only important thing is to surpass myself.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

Table of contents

SWEETS

I don't eat sweets. I imagine life without sweets. I imagine that I (specify a daily activity) without feeling the need to eat sweets. I don't like sweets. But I do like to be healthy. I am in great shape and full of energy. I don't eat sweets.

I now reject my tendency to eat sweets. I don't like sweets. I reject them because they harm my body. They are bad for my teeth. They make me gain weight. I don't like them. Sweets disgust me.

Because I have always succeeded at what I tried in the past, I will also succeed in getting rid of this harmful habit. I don't eat sweets. They're bad for my health. I want to be healthy, slim and full of energy.

Without sweets my skin looks fresh and pure.

Without sweets I do not gain weight.

Without sweets I don't run the risk of becoming diabetic.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I give up candy, chocolate and sweets.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

Sweets disgust me... and day by day this feeling of disgust grows stronger... From this moment on I won't eat any more sweets. I will lose two pounds every week... I feel a lot better... I have lost this voracious craving... I eat two meals a day... and I don't snack between meals.

My health is improving day by day... my digestive functions are working wonderfully well... Eating doesn't interest me... I feel great.

I don't eat a lot at mealtimes... and I don't snack on sweets between meals... day by day I find sweets more and more disgusting.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

I make sure I don't overeat.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

Table of contents

SYMPATHY

I feel more and more sympathetic towards people. People like me and spontaneously seek me out.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Table of contents

TACHYCARDIA

I want my heart to slow down now.
My heartbeat is becoming regular.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

My heart is beating regularly and normally.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

TACT

I am adept at treating people tactfully.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

TALENT (Also see Gift, Salary, Service)

I have talent, I am intelligent and creative.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I have a number of talents which are just waiting to be awakened and used to carry me right to the top.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

(I change everything. I share my creative talents in my daily life. I am a ray of sunshine.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

Table of contents

I possess numerous qualities and numerous talents. I even have aptitudes which I'm still not aware of. I'm always discovering new ones.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I take what I have and use it, and my talent grows, which brings me more profit.

Since I possess the only me there is, I am precious. So I take myself and my talent and I exploit them.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I proclaim my uniqueness.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I find a job which exploits my talents and qualities to the fullest.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

TELEMARKETING

I can improve my phone technique.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

TELEPATHY

I will make ..... (specify who) think about me... s/he won't be able to resist thinking about me.... my image is forcing its way into his / her mind.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1950.

I perceive each of your sensations, all your states of mind, your emotions, everything that enters your mind.

From this moment on our minds are very close, our minds inter-

Table of contents

penetrate and fuse with each other, transforming us into a single mind.

From this moment on I will have no trouble reading your thoughts; all your thoughts will appear to me as clear and precise mental sensations.

I find myself in your body; your body has become my own, your mind is my mind. I see through your eyes, I hear with your ears, I think with your brain.

Your thoughts are present in me; I perceive them very clearly, they are my thoughts.

I am perfectly capable of reading your mind...

Valery Sanfo, Practical Guide To Telepathy, de Vecchi Publications, 1984.

I am going to read your mind... I can read your thoughts... but first I will empty my own mind... I am not thinking about anything... then when I tell myself to... your thoughts will appear in my mind like writing on a blackboard, and they will be easy to read... I will see them clearly... and I will tell you what they are.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

TELEPHONE

In my telephone communications at work or at home I answer the telephone pleasantly, immediately giving my own name to the caller, before I ask who's calling.

Whenever I initiate a call to any residence or place of business where someone new may answer, I always give my own name up front, before I ask for the party I want and before I state my business. By leading with my own name I underscore my value.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Table of contents

TEMPTATION

I have mastered the blind force which caused me to submit to temptation and I have subjugated it to my will; I will consciously make use of this force for positive ends.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

I keep the force that is pushing me to give in to this temptation to myself. ???

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

TENACITY

What others call defeat or failure never stops me. I know that so-called failure is nothing more than a detour. So I get over it and continue on my way.

Whenever I set a precise goal for myself, whatever it is, and develop a specific plan to attain it, I never let it out of my sight, and pursue it tenaciously.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Each step brings me closer to victory. Each obstacle contains the seed of a blessing and of even greater opportunity.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I never stop trying; I will make it through. I will stick with it. Everything will work out in the end.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I persist in a positive and flexible manner until I get things done.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I keep at it and I will succeed.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Table of contents

One last effort may be all that is keeping me from success. I'm going to do it, right now!

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

TENDERNESS (See Love)

TENNIS

I find it easier and easier to play tennis. And I'm making more and more of my shots. Tennis is becoming easy for me. I'm improving day by day, and I'm winning more and more matches.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

TENSION (Also see Confidence, Communication)

With each exhalation I get rid of my worries and tension.

Sublipower, Sublirelax, Edi Inter, 1989.

As I breathe in deeply I am able to overcome my tension.
It is easy for me to relax.
I avoid situations that create tension.
The image I project is of a relaxed person.
I like being a relaxed person.
I perceive myself as a relaxed person.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I let my benevolent mind support me in moments of tension; it gives me confidence in my ability to overcome any kind of tension.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Table of contents

I feel calm and relaxed.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

A good part of my tension disappears because I feel secure and good about myself, and because I put all the little things that irritate me into their proper perspective.

In short, little things don't annoy me any more.

My bank account of confidence is growing, the communication barriers in my life are falling by the wayside, and my family life is benefiting as well.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

(I will no longer live in a state of constant tension, afraid to relax, not allowing myself to smile and spending too much time reliving unhappy memories from the past. I'm going to make every single day a pleasure.

In order to prevent nervous tension from developing, I try to keep my mind relaxed and free of unpleasant thoughts that tense me up. I nourish myself with pleasant thoughts.

I use the technique of self hypnosis to relax my muscles. Relaxing my body relaxes my mind.)

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

All my tension is disappearing.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

TEST

I am very well prepared for —— (specify the situation).

Raymond Hull, Wanting Is Power, Les Editions de l'Homme, 1969.

Table of contents

THOUGHTS

I choose my thoughts... I can do it... I change my thoughts... I can do it.

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

THOUGHTS - NEGATIVE

Every day I create new positive and harmonious thoughts and emotions. They transform my life, bringing me happiness and health.

Christian Godefroy Collection, Magic Formulas For Success And Happiness, Godefroy Publications.

I prohibit myself from thinking negative thoughts. Everything that happens to me has a positive side.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I drop negative thoughts like hot pans!

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I reject and rid my subconscious of all acquired negative thoughts. I am the Spirit, and thus I am in no way affected by disease, death, unhappiness, poverty or frustration. All these things no longer exist in my subconscious mind.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I recognize my wealth and material energy. I get rid of negative thoughts which sap my strength.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

THOUGHTS - POSITIVE

I am going to discover for myself that I have the power within me

Table of contents

to control and influence my mind at will, that I can quickly convert negative thinking into positive thinking.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I have positive thoughts.

Claude Arpel, Self Confidence: The Subliminal Method, Procauvi Publications, 1986.

(I avoid other people's negativity. (I am strong. I do not share in the failures that stem from the minds of others.)

I do not get mired down by depression, fear, anger, poverty, disease, lack of energy, struggle, conflicts or bad relationships.)

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I change my mental diet and refuse to ingest negativity.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I concentrate on good thoughts. I retain thoughts which are likely to help me, and reject the others.

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

I think positively.

Sublipower, Subliconfidence, Edi Inter, 1989.

I saturate my subconscious with well directed positive thoughts which help get me what I expect out of life.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

I am confident of my power to forge the future through the thoughts of today. That is why I take care to formulate them in accordance with my most sincere wishes.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

My thoughts are positive.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Editions, 1982.

Table of contents

YES.
Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

My thoughts are more and more positive.
Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

TIME

I have all the time in the world for what I want to do.
Louis Proto, Self Healing, Piatkus Publications, 1990.

TIMIDITY
(Also see Inferiority Complex, Self Confidence)

I find it easy to talk to individuals and groups of people.
I am at ease in the company of members of the opposite sex.
Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

Each time I mentally repeat the word X at least three times I will feel perfectly relaxed; I am sure of myself, and talk to people with a strong sense of confidence and serenity.
Valery Sanfo, Practical Guide To Self Hypnosis, de Vecchi Publications, 1987.

I confront the unknown, and triumph over it.
Claude Arpel, Self Confidence: The Subliminal Method, Procauvi Publications, 1986.

I take great pleasure in meeting impressive, well-known people.
Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I dare to speak out.
Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

I no longer attach much importance to what other people think about me... and this frees me from my shyness.

I never waste an opportunity to fight my timidity... I often approach people I don't know and engage them in conversation.

I like to get involved in group discussions... in the street I don't hesitate to find some pretext or other to talk to people.

When I talk to people I look them right in the eye... naturally and openly... without trying to avoid eye contact... without being shy... with absolutely no timidity whatsoever.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I must not consider shyness as something to be ashamed of. Shyness is merely a symptom of a lack of self confidence, feelings of insecurity, inferiority and a fear of people.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I will not be shy; I won't be nervous when I talk to people, no matter who they are.

I will be perfectly at ease in all situations.

Paul Clement Jagot, Education Of The Word, Dangles Press, 1975.

If my inner Guide approves of me, which is always the case when I feel His presence, then whatever other people think or say about me doesn't matter. It doesn't affect me. I see my goal, and let my inner Guide lead me to it.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

TOBACCO ADDICTION
(Also see Shortness Of Breath, Cough)

I can easily stop smoking. I completely renounce the smoking habit. At the same time my health and general state will improve. I

can control my reaction to cigarettes. I have absolutely no desire to smoke. I am becoming a new person with absolutely no need to smoke. My subconscious is helping me to stop smoking easily, completely and forever, without gaining weight.

I don't smoke any more. I stopped smoking completely and definitively. My health is great. I'm breathing better and I feel wonderful.

I congratulate myself for having stopped smoking. I really look a lot better. I'm in great shape. This is fantastic. I have stopped smoking.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I feel less and less like smoking, and I have less and less need for cigarettes. I feel good. I feel better every day. I'm smoking less and less. I already feel a lot better.

I'm smoking less. I feel better. I feel full of energy, and I have the courage to continue. I know I'm going to stop smoking once and for all. I already smoke less and less, and I feel great.

I don't smoke any more. I don't need tobacco. I'm breathing better. I fill my lungs with pure air. Each breath fills me with pure oxygen and vitality, recharging my body with energy and doing me a world of good.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

I have an image of myself without cigarettes. I imagine myself as a non smoker. I give up tobacco, I reject the habit. My body rejects it, my mind rejects it. I don't smoke and I feel great.

I reject cigarettes because I know they are harming me. I do it for myself, for my health and well being. And since my life is a success, I will continue to succeed as a non smoker. I choose to be a non smoker.

I'm breathing better and better, and I have more and more energy. I feel wonderful. I breathe easily, I have more endurance. The

people around me are happy at my success. I feel marvelous without cigarettes. I am a non smoker!

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I'm learning to do without cigarettes... I don't want to smoke any more... it doesn't interest me... smoking doesn't interest me at all... Even the idea disgusts me... Whenever I smell smoke I feel this disgust.

Now I don't smoke anymore... Nothing can make me change my mind.

I'm happy to be able to improve my health by not smoking.

My health is improving day by day... I don't smoke anymore.

Day by day it's getting easier not to smoke.

I have absolutely no interest in smoking... even the idea of smoking is enough to make me feel disgusted with cigarettes... The smell of cigarette smoke disgusts me more and more.

I'm very happy not to smoke anymore... Nothing can make me change my mind... I'll never smoke again.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

I will smoke less or not at all.

J.V. Cerney, Stay Younger, Live Longer Through The Magic Of Mental Self-Conditioning, Parker Publishing Company Inc., 1968.

I don't feel like smoking anymore. My health is already better, I'm breathing better, I'm becoming a new person. Not smoking is an expression of my freedom. My subconscious is my most powerful ally, and it is helping me to stop smoking, to build a new self image which is more free and more fulfilled.

Karl O. Stoeber, Psychotraining, Godefroy Publications, 1984.

I don't smoke and I'm proud of myself.

When I see a cigarette or think about smoking, the words 'I Do

Not Smoke' automatically enter my mind and I abstain.

I really like breathing fresh clean air, I like being healthy and in complete control of my mind and body.

My lungs are healthy. I can breathe deeply.

It's easy for me to be a non smoker. After all, I was born that way, so the state is completely natural for me. I am a non smoker! I do not smoke.

I let myself relax, breathe deeply and feel good. I appreciate being healthy and a non smoker no matter what the circumstances.

I do not in any way see cigarettes as a symbol of power, of intelligence or of prestige. I see cigarettes for what they are, and refuse to allow them any place in my life.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am breathing deeply... and this purifying breath eliminates my need to smoke... with each purifying breath my need to smoke diminishes... I need to smoke less and less...

I am breathing deeply... and each exhalation gets rid of the toxins which tobacco has deposited in my lungs... I am cleaning my body of these poisons which have been contaminating my blood... weakening my organism... and ruining my health.

Not smoking will give me new vitality... make me more dynamic... more active... I will feel better from all points of view... and I will be victorious in gaining mastery over myself.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I know I am going to get more satisfaction from not smoking than I did from smoking.

I know that I am indifferent to cigarettes, anywhere, anytime, under any circumstances; other people may smoke but as for me, I don't care about cigarettes.

Clement Auger, Transform Your Life Through Autosuggestion, Heritage Publications, 1979.

Table of contents

I have decided to make life as pleasant as possible by becoming a non smoker. I have made a deliberate choice not to smoke, and I feel great. I am a non smoker. I've found the way to get rid of my bad habits.

The next time I need to relax I'll do a few breathing exercises and drink a glass of water or juice instead of lighting up a cigarette.

The next time I go to a restaurant or take a plane I'll sit in the non smoking section.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defenses, Godefroy Publications, 1989.

I'm intelligent, I've got common sense. I need my body to live, and cigarettes are poison.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

I am free of the habit of smoking.

Jack Ensign Addington, How To Set Goals And Attain Them, A Different World Publications Ltd., 1977.

I absolutely want to stop smoking.

The taste of tobacco is disgusting.

I don't need to smoke anymore.

I am aware that it is much better not to smoke.

I want my breath to smell good, without the after odor of tobacco.

I respect my body.

I am in good health.

I am gradually freeing myself of my dependence on tobacco.

I substitute the harmful habit of smoking with beneficial habits of nutrition.

As I stop smoking I watch what I eat so I won't gain weight.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

Freedom and peace of mind are now mine. I know that believing and affirming these truths will allow them to penetrate deeply into my subconscious, which will in turn force me to give up cigarettes,

because smoking is harmful and breaks the laws of my subconscious mind.

Dr. Joseph Murphy, Miraculous Power Attracts Infinite Riches, Godefroy Publications, 1978.

My voice is normal again.
I don't cough anymore.
My mouth is fresh tasting and I feel better and better every day.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

If I light up a cigarette I won't get any pleasure out of it. On the contrary the taste will disgust me, to the point where I'll have to put the cigarette out. It will be impossible for me to smoke.

Valery Sanfo, Practical Guide To Self Hypnosis, de Vecchi Publications, 1987.

If I must smoke I will smoke in moderation.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

TOLERANCE

From now on I will think, feel and act justly. I will do what is right, and never deviate from the path of justice. I think, speak and write without taking sides or being prejudiced. From now on I will not react like an ignorant fanatic, but base all my actions on the divine spirit that resides in me. I wish from the bottom of my heart that all men may live in liberty, and rejoice fully in their right to life and happiness. I obey the golden rule and the law of love.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I accept and respect the right of each person to choose.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

(I use self hypnosis techniques to strengthen my potential for tolerance, and I never have any problems.)

Table of contents

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I cease imposing temporary conditions in order to organize and program other people's lives.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

I never force others to share my points of view.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am full of tolerance for myself and for others.

Christian H. Godefroy, Self Confidence: The Subliminal Method, Godefroy Publications, 1987.

(I am a warm, open and tolerant person.)

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I am going to criticize less and be a little more tolerant of others, of their faults, their failures and errors. I will interpret their actions in the best light possible.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

TOOTHACHE (Also see Pain)

I just have to touch my —— (name the part of the body) for it to become numb.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

TORMENT

I calmly try to reduce the impact of extreme circumstances.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

This too will pass.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

Table of contents

I am learning to mentally detach myself from external turmoil.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I calmly determine what worst case scenario is that I may have to deal with.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I prepare myself to accept the irremediable if necessary.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

I never abandon myself; I support myself when things go badly or start getting difficult.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

TOXINS

With every out breath I rid my body of toxins.

Louis Proto, Self Healing, Piatkus Publications, 1990.

TRANQUILLITY

My mind and body are tranquil.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I am completely tranquil.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

When I repeat the word X just once, my mind enters a state of rest.

Valery Sanfo, Practical Guide To Self Hypnosis, de Vecchi Publications, 1987.

TRANSACTIONS

Table of contents

God is always present in my transactions, and I never finish a day without accomplishing something valuable.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

Infinite intelligence governs and protects all my financial transactions.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

TRANSFORMATION

Today I am starting a new life.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I find it easy to transform myself.

Sublipower, Sublicontrol, Edi Inter, 1989.

I am transformed.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I am a new person every minute.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I want everyone around me to notice my transformation. I have an improved attitude. I am straightening up: my muscles are firmer and stronger. I feel new life flowing through me.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

TRAVEL TIME

Before going somewhere I make sure I am using the means of transportation that is going to save me the most time, not necessarily

the one which is fastest.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I try to organize my traveling time so that my mind is clear, and I can use it to work or think.

I avoid encumbrances, and only move around when I have to.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

TROUBLE

I try not to bother too much about my troubles.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I avoid blaming myself for things I cannot change.

They aren't really important.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

I assign a lot less time for worrying about things.

I give myself ten minutes in the morning and ten minutes in the afternoon to vent my anxiety and reassess all the catastrophes I can think of during those time periods.

Then I use my ability to control my thoughts and stop troubling myself until the next worry period.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Everything is fine the way it is.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I am going to auto condition my mind every morning with a positive attitude toward each new day.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

TRUTH

Table of contents

I have the courage to look truth in the eye.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I seek the truth.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I am full of desire to unlearn false ideas and learn new truths.

I don't need to fall into any traps. Desire without hope dissolves. I have nothing to fear. There is enough of everything to satisfy my real needs. I place Truth above all else, and everything else comes to me. The mental struggle is useless. I just have to relax. I am aware, and I let Truth liberate me.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I have to confront myself and search for truth within myself, with my own means, because I consider it important.

Whether I am aware of it or not my mind is functioning in a unique and natural way, based on the fundamental principle of showing me the truth about any situation I encounter.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

TUBERCULOSIS

Every cell, every nerve and every muscle in my lungs is now being purified, strengthened and perfected. My whole body is healthy and harmonious.

Dr. Joseph Murphy, The Miracles Of Your Mind, Dangles Press, 1984.

TUMOUR

Table of contents

I want my little tumor to go away.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

Table of contents

ULCERS (Also see Digestion)

My body is building a protective barrier of tissue around the wound; the wound stops bleeding and heals up.

My wound is no longer painful; all the reflexive reactions disappear; my stomach is calm and digests normally.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

My stomach digests better and better.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

UNCERTAINTY

(When I'm not very sure about what I'm doing I examine the consequences of each possible course of action.)

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

UNDERTAKINGS

I have a clear conscience; it will surely guide me in all my endeavors.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I undertake things which I have always avoided under the pretext that I didn't understand enough about them.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

Table of contents

UNEMPLOYMENT

All right. I haven't got a job. I have to react immediately. I'm going to find work this week. In any case, it's not serious. It'll do me good to see different things, meet different people. Because in fact, I didn't really like the job I had. I was starting to get really bored.

I am competent. I have experience. I'm due for some luck. I am convinced that somewhere out there, right at this very moment, is my ideal job. And I'm going to find it.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

In the universal harmony where all human beings have their place, I have the useful and well paying job I deserve. Thank you!

Marcel Auclair, The Book Of Happiness, Seuil Publications, 1959.

The god in me knows what kind of work suits me best, where it is, and what I must do to carry it out. May this knowledge in me become a living revelation to my conscious mind, so that I may know where my real work lies, and the steps I must take to find it.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

UNEXPECTED EVENTS

I schedule the time I need to deal with unexpected events.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

I never let myself get flustered.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

UNHAPPINESS

I stop trying to be happy. I try to understand the causes of my

Table of contents

unhappiness.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I believe I have the mind-power to survive all life's frustrations, disappointments and misfortunes.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

UNIFORMITY

I remind myself that uniformity and death are the opposite of fulfillment.

Wayne Dyer, Your Erogenous Zones, Tchou Press, 1976.

UNION

I know that an irresistible law of attraction exists and that I will attract a man / woman who is in harmony with my subconscious beliefs. I attract what my subconscious believes to be true.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I know that the intelligence of my subconscious mind has united us in accordance with the divine order of things.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I am in union with Life, Love and Truth, because I live with God. I am united with all beings because the image and reflection of perfect Life resides in everyone.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

Table of contents

URINE RETENTION (Uremia)

In a minute I am going to empty my bladder completely.

Dr. G.R. Rager, Hypnosis, Sophrology and Medicine, Fayard Publications, 1973.

I will urinate tonight, and then more tomorrow.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

URTICARIA (See Eczema, Acne)

USED CARS

I will find the used car I am looking for. At a good price. It is in excellent condition. It corresponds exactly to my needs. My subconscious mind is infallibly guiding me towards this car.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

UTERINE BLEEDING

My uterus is tightening; the blood vessels are closing; the flow of blood is stopping.

I feel my uterus getting tighter as I tighten my fist.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

UTERINE SPASMS

My uterus is relaxing, the pain is going away, and I feel fine.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

VACATION

(This is good, I'll be able to relax, I'll take a vacation.)

Louis Proto, Self Healing, Piatkus Publications, 1990.

I try to take short vacations instead of extended ones which would probably exceed my real need for physical and emotional relaxation.

I remind myself that other people have the same need for relaxation as I do.

I plan my vacations, and discuss them with my spouse and children.

I know that I am not obliged to spend all my vacations with my family (relatives, spouse, children, etc.).

I make sure I get a lot of free time when I'm on vacation.

Pierre Nicolas, Time Is Money... And Pleasure!, Inter Publications, 1981.

VENGEANCE

I never try to do the same to my enemies as they have done to me because by seeking revenge I hurt myself more than the people I want to hurt. I follow General Eisenhower's example: I never waste even a minute thinking about the people I don't like.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

Table of contents

VERTIGO (Dizziness)

Starting today I won't suffer from vertigo anymore. Every time I feel dizzy I will think that I feel as good as on the day I (insert a pleasant experience). I am serene, happy and confident. I can go wherever I like. I am in control of my sensations.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

VICE

I am free of the vice that has been sapping my strength. No negative impulses can make me renounce my freedom. They no longer have any power over me. I see my goal clearly, and nothing can distract me.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

VICTORY

I am the greatest of nature's miracles. And nature knows no failure. It always wins, as will I starting today. And thanks to my victories the struggles to come will be less severe.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

I embark on the path of life with a new ideal. I embark with a victorious attitude, and a generous and beautiful soul. I know I will be victorious. And I want to be worthy of my victory.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I am a victory.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

Table of contents

I want to be victorious.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

I want to be victorious, and to do so I must start by triumphing over myself.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

VIGILANCE

(I always expect the unexpected.)

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I am always vigilant for opportunities.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I concentrate on performing with a smooth, relaxed alertness.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

VIGOR

I see myself as a vigorous person in the best of health.

Dr. Anthony and Mary Zaffuto, Alphagenics, Quebec - Amerique Publications, 1974.

A guaranteed way to stay young and vigorous is to first change my inner self. The energy will follow.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

VISION

I relax the muscles in my eyes, in both eyes, and my vision is improving.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

Table of contents

I can see better and better; I have excellent vision, in an intellectual, spiritual and physical sense.

The divine spirit, which has the infinite power of healing, is now opening my eyes and recreating my visual faculties. My eyes are the instruments of perfection desired by God, and allow me to see both inside myself and out into the exterior world. The magnificence of God is revealed to me through my eyes.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

My vision is becoming clear again... I can see better and better.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

VISION - POSITIVE

I see with love and joy.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

Visualization

Who I see in my imagination will always rule my world.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I increase my receptivity to mental images. I make them work for me more and more.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I use my senses on all mental levels as I experience the external world.

Christian Godefroy, Mental Dynamics Seminar, Godefroy Publications, 1978.

I can obtain what I want by visualizing it.

Sublipower, Sublisuccess, Edi Inter, 1989.

Table of contents

VITALITY (Also see Ambition, Dynamism, Enthusiasm, Beauty, Health)

Every breath I take increases my aliveness.
Every day that passes I feel stronger and more alive.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I am overflowing with more and more energy and strength.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I am overflowing with vitality and enthusiasm!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I am regaining my balance and vitality.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I am energetic and overflowing with vitality.
I give thanks for my ever increasing health, beauty and vitality.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I am strong and vital.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

My natural state is to have radiant health, boundless energy and vitality throughout my life.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

VIVACITY

I have a pleasant and lively mind, and I find that people appreciate it because I use it a little more each day.

Arthur Wassmer, The Art Of Making Friends And Being Happy, Belfond Publications, 1978.

Table of contents

I am intelligent, my mind is lively, alert, brilliant and amusing. Since my thoughts are positive, my mind creates harmonious situations in my life.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

VOCATION (Also see Money)

I trust my preferences to show me where my true vocation lies.

Louis Proto, Self Healing, Piatkus Publications, 1990.

My subconscious mind will infallibly guide me to the vocation where I will succeed best, and which will bring me all the money I need.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

Subconscious Mind, help me discover the vocation that pleases me the most.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

VOICE

I am certain I won't lose my voice, not even for a moment.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I want my voice to be more confident; I want it to express my decisions not as commands, but rather as well thought out advice.

Henry Durville, I Want To Succeed, Durville Publications, 1968.

Table of contents

WAKING UP

I wake up every morning full of joy and confidence.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

All I have to do is concentrate on a given time in order to wake up at that time feeling refreshed and ready to face the world.

Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

When I wake up I feel great and full of energy.

Sublipower, Sublihealth, Edi Inter, 1989.

Whenever I use my mental clock and sincerely want it to work, I wake up at the exact time I choose.

Christian Godefroy, Mental Dynamics, Robert Laffont Publications, 1976.

Tomorrow at precisely —— o'clock I will wake up feeling refreshed, and get out of bed immediately.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I want to wake up at —— o'clock.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

WARTS

The blood vessels in my wart are contracting; the wart is no longer being nourished with blood, and is drying up.

Dr. Gaston Durville and Dr. Andre Durville, Mental Cure, Naturist Institute Publications, 1923.

Table of contents

WEAKNESS

I give up being helpless.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I have the right not to be strong.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

WEALTH (Also see Profession, Success, Value)

This year I will double my income to $———, and in six years from now, that is on —— (specify the date) I will be a millionaire.

Marc Fisher, The Instant Millionaire, Golden Publications Ltd. and Godefroy Publications, 1986.

Each day I am getting wealthier in all things.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

Each day brings me closer to wealth.
I desire to be rich.
I am capable of being rich.
I want to be rich.
My subconscious is guiding me infallibly towards wealth.

Christian H. Godefroy, Prosperity: The Subliminal Method, Godefroy Publications, 1987.

Every day, in all ways, I am becoming more and more wealthy.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

Every day I am more and more wealthy in everything.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

There is no logical reason which can prevent me from becoming wealthy.
I am getting more and more wealthy.

Table of contents

I am becoming wealthy.
I am getting wealthier day by day.
Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I have confidence in you, my inner Guide, and from now on everything will be fine. I feel my mental distress and the obstacles separating me from enjoying wealth disappearing. I feel I am flowing with the current of abundance, and I am moving forward with it. I know that with your help my wishes will be granted.
K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I have the right to be rich, healthy and prosperous.
Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

I opt for the richness of life here and now. Wealth is already uppermost in my thoughts. Yes, I am grateful for all the wealth of my mind and my life. I feel wealthy. I am wealthy. I am becoming wealthier every day.

All the riches of the world circulate joyously and freely in my life, and I am very grateful. I know that all my needs are satisfied because I accept wealth. The world is there for me to enjoy, and I am rich. I am rich because I accept being rich. I can feel all my wealth now. I feel it with gratitude.
Charles Baudouin, Psychology And Practise of Autosuggestion, Idegraph Publications, 1990.

I am becoming more and more wealthy.
Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I am becoming wealthy.
Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I enrich myself day by day.
Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I do not believe that wealth is the private domain of a selected few, but rather that the abundance of life offers wealth to everyone who creates it.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I do things that are likely to bring me good fortune... I don't bargain with God... I don't set an amount that I'd like to reach... I conduct my business by submitting to His impulses, knowing that He will give me what I firmly desire.

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

I reflect with a positive mental attitude, and I become rich!

The more I share what I possess with others the more my wealth increases.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I am now a mental millionaire.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am rich.

Charles Baudouin, Psychology And Practice of Autosuggestion, Idegraph Publications, 1990.

I am a child of infinity, and wealth flows through me freely, joyously and eternally. I am enriched with happiness, peace, wealth, success and extraordinary sales. I seek wealth in the deepest recesses of my mind, and abundance is the result.

Dr. Joseph Murphy, Miraculous Power Attracts Infinite Riches, Godefroy Publications, 1978.

The power of 'collective consciousness' helps make my fortune and my money grow.

I think and become rich because nothing, absolutely nothing can stand in my way.

Napoleon Hill, Think And Get Rich, Tchou / Ariston, 1966.

My connection to infinite being and infinite intelligence is quite enough for me to amass a considerable personal fortune.

Christian H. Godefroy, Mental Techniques For Financial Independence, Godefroy Publications, 1983.

My life is a free flow of infinite wealth.
This universe is bursting with riches for every one of us.
Infinite wealth appears easily in my life.
I am wealthy within and without.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

My affairs are improving every day. I am making progress, moving forward, and I am richer every day.

Dr. Joseph Murphy, The Miracles Of Your Mind, Dangles Press, 1984.

One day I will be a millionaire.

J. Martin Kohe, Your Greatest Power, A Different World Ltd., 1979.

WEIGHT

Now that I weigh — pounds I feel fine, I look great and my physical condition is excellent.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

WELL BEING

I am responsible for creating my life.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

There is no limit to how good I can feel.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I have the power to feel good.

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

Table of contents

I love my body and accept it totally.
I love doing what makes me feel good.
It is good to be alive on this earth.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I experience continual well being.

Professor Robert Tocquet, The powers Of The Will, Godefroy Publications, 1989.¥

I am at home in my body.

Louise L. Hay, You Can Heal Your Life, Hay House, 1984.

I'm doing fine, intellectually, physically and financially. I feel fine, day and night.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I feel good about myself.

Claude Arpel, Self Confidence: The Subliminal Method, Procauvi Publications, 1986.

I feel good, really good.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I feel fine.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

I feel better and better.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

I feel relaxed and happy. I feel good about myself.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

I am completely calm... I experience a marvelous sensation of calm and harmony spreading through me... a feeling of indescribable joy flows from my heart.

Professor Kurt Tepperwein, Think And Heal, Godefroy Publications, 1984.

My persevering quest for the principles of truth will be sure to

produce an inner change which, in turn, will produce a warm feeling of well being.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I refrain from doing anything that may inadvertently cause me harm.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

WILLPOWER
(Also see Self Control, Self Confidence)

I have confidence in myself. My willpower is already very strong, and it is getting stronger and stronger. I want to succeed. I want to follow the examples of those who have attained the goals they set for themselves. I will persevere because I know with certitude that properly directed willpower will succeed when it is working for a just and honorable goal.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

I have willpower... I never give in to outside influences... I don't care what people may think.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

I have the willpower to act.

Jean Chartier, How To Acquire A Superior Personality, Dangles Press, 1971.

I desire —— (specify) and I am in the process of obtaining it.

Emile Coue, Complete Works, Astra Publications, 1976.

I do not let circumstance defeat me. I control them psychologically.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

Table of contents

I can act according to my reason, and that is how I want to act.

Paul Clement Jagot, How To Develop Your Personal Magnetism, Dangles Press, 1982.

I can encourage my mind to think in a certain way, and direct my actions towards accomplishing a certain project.

When my willpower is intensified to the point where I can defy circumstances and realize the project I envisage, I will achieve results in my life which will astonish everyone.

Franck Rudolph Young, Cyclomancia, SIP, 1966.

I can have or be anything I want as long as I have enough energy and willpower.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

I can do it and I want to do it.

Leon Kendal, Personal Magnetism, Godefroy Publications, 1911.

I can do it! It's easy. I can do this task better than anyone else. Thanks to the power of my inner Guide no obstacle can stop me. I succeed at everything I do!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I know what I want and I believe I can get it.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

I know exactly what I want and I authorize my subconscious to act according to the law so that it becomes immediately manifest.

All paths are free, all doors are open. At this very moment all my actions are perfect, and I will get what I want.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I feel my willpower getting stronger and stronger.

Professor Lawrence Powers, Ten Secrets Of Hypnotic Power, Godefroy Publications, 1982.

I am in control of my actions, my thoughts and my emotions. I

am free of all restraints. My willpower is getting stronger and stronger. I have confidence in myself and in the future.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

I am a centre of conscious willpower.

Isabelle Capanio, The Powers Of the Will, De Vecchi Publications, 1987.

(I am going to develop the technique of self hypnosis in order to develop my willpower to improve, my desire to study, to learn and to make progress in all aspects of my life.)

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

I want to train my willpower until it is strong enough to change my life.

Paul Clement Jagot, The Power Of The Will, Dangles Press, 1973.

I want, I can, and I am sure of myself.

Professor Lawrence Powers, The Ten Secrets Of Hypnotic Power, Godefroy Publications, 1982.

I want, therefore I can.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

I want... I want... I want...

Claude M. Bristol, Harold Sherman, Your Inner Power = TNT, A Different World Publications, 1954.

Playing my role means trying to acquire the intellectual and moral qualities of a man of willpower. I want it with all my strength, with all the power I have. I want to create a beautiful future.

My willpower is getting stronger and stronger.

To play my role faithfully, I first develop the attitude of a person who is strong and who wants to succeed.

Professor Robert Tocquet, The Powers Of The Will, Godefroy Publications, 1989.

My willpower is getting stronger and stronger, more and more powerful.

Marcel Rouet, Techniques And Practice Of Hypnotism, Paris Productions N.O.E., 1974.

Table of contents

My willpower is consolidating.

Marcel Baugin, Hypnotism, Suggestion and Telepsyche, Scientific Distribution, 1956.

When I want to, I can.

Norman Vincent Peale, When You Want, You Can!, A Different World Press, 1974.

WINNING

I now win all the time.

Louis Proto, Self Healing, Piatkus Publications, 1990.

I won! I won! I won!

Richard Corriere, Joseph Hart, Getting Into Psychological Shape, Les Editions de l'Homme, 1979.

I have the right to win.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I know that everything depends on me, and nothing but me. Everything will depend on the way I see things, and on what I do. That's what success is all about, and that's why I always win.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

I play to win.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I'm a winner. I'm a winner.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

I'm a winner. I'm a likable person, I'm enthusiastic, warm and honest, the type of person people like to be with.

Every day my personality becomes stronger and more radiant, revealing ever more clearly the positive, optimistic spirit that animates all my actions.

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

Table of contents

My ability to become a full time winner depends entirely on my determination to want to think like a winner, and erase all self images of being a loser that I may have accumulated in the past. My capacity for being a person with No Limits, and reaching beyond even my own wildest dreams is entirely in my own hands.

Wayne. W, Dyer, The Sky's The Limit, Mortagne Publications, 1988.

I am a winner in my life. I am not a victim, I'm a winner!

Shad Helmstetter, The Power Of Inner Motivation, Published by Le Jour, 1987.

WISDOM

Now joy, wisdom and the peace of God are uppermost in my mind.

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

With God everything is possible. All powerful God is the spirit living in me, and the source of all wisdom and enlightenment.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

I trust myself and believe in my innate wisdom.

Liah Kraft-Macoy, 30 Days To Happiness, Stillpoint, 1987.

I pray every day, not to amass more wealth, but to become wiser so that I can recognize and benefit from the abundance of goods which are always in my possession, and take pleasure in them.

W. Clement Stone and Napoleon Hill, Success Through Positive Thinking, Godefroy Publications, 1978.

I am wiser than I think. When I rid my mind of false notions, wisdom surges upward like an underground spring.

Vernon Howard, Psycho-Pictography, S.I.P., 1965.

I am full of the wisdom of divine life.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

Table of contents

I am a balanced being of love. Creation is my playground. I open to my inner wisdom and live it every day.

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

My heart contains perfect wisdom.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

WISDOM - SUBCONSCIOUS

My subconscious knows exactly what is good for me, and what can help me in life.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

WISH

It's okay to ask for what I want - and get it.

Louis Proto, Self Healing, Piatkus Publications, 1990.

——— (specify your wish) will come to me easily and without effort.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

Life is a self fulfilling prophesy; I won't necessarily get what I want in life, but in the long run I will usually get what I expect.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

WIFE

My wife wishes to be loved and respected as a person.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

Table of contents

WONDER

I never lose my sense of wonder about life.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

WORK (Also see Reward, Salary, Worth)

Today I'll finish all the work I have joyously and easily. Everything I have to do interests me. My mind is constantly present and alert; I am ready to use the opportunities this work, or any other week provide, to do my best to succeed.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

In my life work and play are the same.

Shakti Gawain, Reflections In The Light, New World Library, 1978.

I like my job, my colleagues and my responsibilities. Through them I add value to life, and thus I become useful to God.

R. Charles Barker, The Science Of Success, Dangles Press, 1983.

I like my work, therefore my work likes me.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I like my work. It fulfils me on both a creative and financial level.

Shakti Gawain, Creative Visualization Techniques, Soleil Publications, 1978.

I choose to do this because it is in my best interests, based on the current realities and my current goals. No one is forcing me to do this, and I don't 'have to' do it. I've assessed the consequences and alternatives and on that basis I choose to do this.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

I make my work an exercise in concentration: I do it consciously,

Table of contents

with interest, and I apply myself body and soul. In no time at all my work becomes a pleasure. Instead of being bored, I discover new sources of interest and opportunities for success in my work that I had no idea existed until now.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I find my work more and more fulfilling.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

I'm happy in my work.

Emrika Padus, Encyclopedia of Emotions And Health, Edi Inter, 1991.

I work well.

Editorial Staff, Practical Health Guides, How To Stimulate Your Body's Natural Defences, Godefroy Publications, 1989.

My work is interesting and easy, and I do it with pleasure. Problems are made to be overcome. I will see things through. And I have the strength to do even better. Not only does my work provide me with real benefits, it also enriches my mind.

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

My work is a joy and a pleasure.

Louis Proto, Self Healing, Piatkus Publications, 1990.

When I think about my business and the office I am filled with love and joy, inner peace and harmony.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

All my work bears fruit in accordance with the divine order of things.

Dr. Joseph Murphy, Exploit The Power Of Your Subconscious, Tchou / Ariston, 1962.

All work becomes simple and pleasant for me.

R.L. Charpentier, Autosuggestion And Its Practical Application, Champs Elysees Publications, 1958.

Table of contents

WORRY

Instead of worrying about being afraid, I am going to do something about it. I am going to read various books explaining the psychology of fear. I am going to learn the difference between normal and abnormal fears. If I understand the facts about various fears I will be more successful in overcoming them.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

God is watching over me, and that's wonderful!

Dr. Joseph Murphy, These Truths Will Change Your Life, Published by Le Jour, 1982.

I describe the things around me to myself. While I do that I stop thinking about my worries.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

I am getting stronger day by day. At the same time all my worries and fears, all my inner problems are dissolving. I am strong and free. I have all the energy of my Guide at my disposal, and S/He will surely lead me to my goal. My goal is near, and I will succeed!

K.O. Schmidt, Success At Your Service, Astra Publications, 1986.

I don't want to worry about that now; I'll worry about it tomorrow.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

Stop! Fear and worrying are a waste of time and energy, and they block me from using my skills.

Dr. Marty Seldman, Super Selling Through Self Talk, Price Stern Sloan, 1986.

God is the light shining on my path; I am no longer groping, trembling with vain hopes, I see no seeds of worry or struggle on my path, but only seeds of joy and love of life.

Robert Collier, Wealth At Your Doorstep, S.I.P., 1959.

Table of contents

I eliminate worrying, the major consumer of wasted time, as much as possible.

Christian H. Godefroy and John Clark, How To Have More Time, Godefroy Publications, 1989.

I count my reasons for being happy instead of unhappy.

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

Worries and negative thoughts slide off me like water off a duck's back.

Christian H. Godefroy, Positive Thinking: The Subliminal Method, Godefroy Publications, 1987.

My mind is free of all worry.

Francoise and Michel Moine, Develop Your Psi Powers, Stock Editions, 1982.

My mind is free of doubt and worry.

Since these words come from my soul, they erase all my negative thoughts and negative feelings.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

No matter what happens or doesn't happen to me today there is one thing I am sure of: I want to get the most I can out of this day.

Orison Swett Marden, The Joy Of Living, J.H. Jeheber Publications.

Two months from now I will have completely forgotten this setback, so why should I worry about it today?

Why not assume the attitude I'll probably have in two months right now?

Dale Carnegie, Overcome Your Worries, Flammarion, 1944.

From now on the setbacks in life no longer affect me. If I am assailed by doubt, fear and worry, I know that my faith will overcome them.

Dr. Joseph Murphy, Mastering Yourself And Others, Published by Le Jour, 1988.

If I am sometimes assailed, pursued, haunted by harmful and

unhealthy thoughts, by fears, phobias, temptations or resentment, I know they will fade away gradually, seeming to melt from my mind, growing faint like a distant cloud which will soon disappear completely.

As dreams disappear when I awaken, so will all these vain images disappear as well.

Emile Coue, Complete Works, Astra Publications, 1976.

WORTH (Also see Wealth, Work, Salary)

Today I will multiply my worth a thousand times.

Og Mandino, The Greatest Secret In The World, A different World Press, 1979.

Every day I multiply my worth a hundred times, and my success grows as a result.

Day by day, in all ways, I am becoming better and better. All areas of my life are enriched, because I multiply my worth a hundredfold.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

I am aware of my worth.

Claude Arpel, Self Confidence: The Subliminal Method, Procauvi Publications, 1986.

I have worth because I struggle to survive.

Matthew McKay and Patrick Fanning, Self Esteem, New Harbinger Publications Inc., 1990.

I evaluate my real worth by asking how good I feel about myself.

Recognizing my own worth is the beginning of a mature acceptance of myself and others.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I believe I am unique and special.

Denis Waitley, Seeds Of Greatness, Pocket Books Press, 1984.

Table of contents

I am a rare, exclusive, different, unique individual on this earth. These attributes give me enormous worth.

Zig Ziglar, Rendez-vous At The Top, A Different World Publications, 1975.

I am a worthy person.

Alan Houel, How To Deal With Difficult People, Godefroy Publications, 1990.

I am unique and unlimited.

Jean-Marie Boisvert and Madeline Beaudry, Self Affirmation and Communication, Godefroy Publications, 1979.

I am unique. My worth is multiplying, allowing me to do a job which I like and which pays very well.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

(I deserve to be who I am.)

Shakti Gawain, Reflections In The Light, New World Library, 1978.

The best way to increase my worth is to improve my self image.

Charles Albert Poissant & Christian Godefroy, The Secret Of Being A Millionaire, Godefroy Publications, 1987.

The more I realize my true worth, the more everybody respects me.

Louis Proto, Self Healing, Piatkus Publications, 1990.

Whatever anyone says or does to me, I remain a worthy person.

Alan Houel, How To Deal With Difficult People, Godefroy Publications, 1990.

WOUNDS (Also see Pain, Numbness)

Now the pain is disappearing from my —— (name the part of the body) and my blood is coagulating rapidly. The pain is diminishing rapidly. I won't feel any more pain for half an hour. It will disappear completely. Now I'm going to count to five. And on the count of five my pain will disappear completely for half an hour (or substitute the

desired period). One, two, three, four, five!

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

WRINKLES

I am relaxing the muscles in my face... in my forehead... I feel a cool wave of relaxation...

Every time I feel myself getting tense or upset... I will think of the word serenity... and immediately all the muscles in my face will relax.

My face will look younger... it is going to look younger... releasing... relaxing all my features will prevent wrinkles from forming... smooth over those that are already there... and chase away my worries.

Marcel Rouet, Mastering Your Subconscious, Dangles Press, 1982.

YOUTH

I do not accept the passive lifestyle of many elderly persons. I live an active life.

I refuse to always think about the past, as too many elderly persons do. I am alive today; I live in he present moment.

Dr. Maxwell Maltz, Psychocybernetics and Self Realization, A Different World Ltd. Publications, 1970.

I am going to remain young in mind and heart by finding new pleasures, new interests and new friends.

Frank S. Caprio, M.D. & Joseph R. Berger, Helping Yourself With Self Hypnosis, Prentice Hall Inc., 1974.

My body is getting younger and more vigorous. I am regaining (or maintaining) optimum use of my bodily functions. My hearing, my vision and my memory are becoming more and more efficient and healthy. I am in excellent health. My face looks younger every day, and I'm looking better and better. I feel more and more young, more and more relaxed, more and more healthy. I am regaining my youthful enthusiasm and sense of curiosity. Youth. Youth. Youth. I am more and more energetic, creative, and young in body and mind.

Christian H. Godefroy, Helping Yourself Through Self Hypnosis, Godefroy Publications, 1985.

(I can be whoever I want to be with total support. I can go fast or slow - it's my choice.)

Barbara J. Gress, Personal Power Cards, Aslan Publishing, 1991.

I'm losing weight... I am happy to feel so good about myself...

Table of contents

day by day everything is getting better... my skin is supple... I feel younger and younger... slimmer and slimmer... full of newfound energy.

Professor Kurt Tepperwein, Secret Techniques Of Hypnosis, Godefroy Publications, 1981.

Table of Contents

ABUNDANCE 10
ACCOMPLISHMENT (Also see Success) 11
ACNE (See Urticaria, Eczema) 11
ACTION (Also see Opportunity) 11
ADVERSITY 13
AEROPHAGY (Flatulence) 13
ADVICE 14
AFFECTION 14
AFFIRMATION 14
AGING 14
AGITATION 15
AGREEMENT 15
AGGRESSION 15
AIDS 16
ALCOHOLISM (Also see Sobriety) 16
ALLERGIES 18
ALLERGIES (Food) 18
ALTRUISM 19
AMBITION (Also see Vitality, Dynamism, Enthusiasm) 19
ANAEMIA 21
ANAESTHESIA (Also see Pain, Wounds) 21
ANAESTHESIA - DENTAL 22
ANGER (Also see Migraines) 22
ANGINA (Chest) 23
ANSWERS 24
ANXIETY 24
ANXIETY - SOCIAL 25
APLOMB 26
APPEARANCE (PHYSICAL) 26
APPETITE (Also see Digestion, Stool) 27
APPREHENSION 27
APPROVAL 27
ARCHERY 28
ARTHRITIS 28
ARRHYTHMIA 29
ASKING 29

ASSIDUITY (Also see Perseverance) 29
ASSIMILATION 30
ASSISTANCE 30
ASSISTANCE - SPIRITUAL 30
ASSOCIATES 31
ASSURANCE (Also see Self Confidence) 31
ASTHMA 32
ASTHMA ATTACK 34
AUDACITY (Also see Courage) 34
AUTOSUGGESTION (Also see Hypnosis) 35
AVAILABILITY 35
AWARENESS (See Self Awareness) 35
BAD CHARACTER 36
BAD HABITS (See Habits - Negative) 36
BACKACHE 36
BALANCE (Also see Peace, Strength, Wisdom, Health, Harmony) 36
BALDNESS 37
BEAUTY (Also see Health) 38
BEHAVIOR 38
BEING 38
BELIEF 39
BELONGING 39
BENEVOLENCE (Also see Sincerity) 39
BEST 40
BIRTHING 40
BLAME 40
BLESSING 41
BLISS 41
BLUSHING 42
BODY 42
BOREDOM 42
BOWLING 43
BRAGGING 43
BRAVERY (Also see Strength, Courage) 43
BREAKING UP 43
BREASTS 43
BREATHING 44
BRIDGE 45
BUDGET 45
BUSINESS (Also see Money) 45

BUYING 46
CALM (Also see Self Control) 47
CANCER 48
CARDIAC TROUBLE 49
CARE 49
CAREER 49
CHALLENGE 50
CHANGE (Also see Habit) 51
CHARM (Also see Slimness, Health, Sincerity) 52
CHILL 53
CHOICE OF PROFESSION 53
CIRCULATION (Also see Anemia) 53
CIRCUMSTANCE 54
CLARITY 54
CLIENTS 54
COLD (Common) 55
COLIC (Biliary or Hepatic) 55
COMFORT (Feeling At Ease) 55
COMMITMENT 56
COMMUNICATION (Also see Confidence, Tension) 56
COMPANY 58
COMPASSION 58
COMPETENCE (Also see Money, Skill) 58
COMPETITION 59
COMPLEX 60
COMPLEX - INFERIORITY 60
COMPLEX - SEXUAL 61
COMPLIMENTS 62
COMPOSURE 62
COMPREHENSION 62
CONCENTRATION (Also see Delegating) 63
CONFIDENCE 65
CONFIDENCE IN THE FUTURE 66
CONFIDENCE - SELF (See Self Confidence) 67
CONFORMING 67
CONGRATULATIONS 67
CONSENSUS 68
CONSIDERATION (Also see Respect) 68
CONSTIPATION 69
CONSTRAINT 69

CONTACT 69
CONTRACT 70
CONVERSATION 70
CONVICTION 70
COOPERATION (Also see Help) 71
COUGH 71
COURAGE (Also see Dynamism, Resistance, Energy, Enthusiasm, Strength, Bravery) 71
COURTESY 72
COWARDICE (See Lying, Laziness) 73
CREATIVITY (Also see Sensitivity, Love) 73
CREDIT 75
CRISIS 76
CRITICISM 76
CURIOSITY 77
DANGER 78
DEAFNESS 78
DEBT (Also see Money, Projects) 78
DECISIONS 79
DEFEAT 81
DEFEATISM 81
DEFENCE 81
DEFIANCE 81
DELEGATING 82
DEPENDENCE 82
DEPRESSION 83
(Also see Conditioning, Confidence, Happiness, Health, Abundance, Optimism) 83
DERMATOSIS 85
DESIRE 85
DESIRE - SEXUAL (See Sexual Desire) 86
DESTINY 86
DETERMINATION (Also see Energy) 87
DIALOG 89
DIET (Also see Hygiene, Negative Habits) 89
DIFFERENCE 90
DIFFICULT TASKS 90
DIFFICULTY (Also see Worry) 90
DIGESTION (Also see Appetite, Stool) 91
DIGNITY 93

DIPLOMA 93
DIRECTING OTHERS 94
DISAGREEMENT (Also see Feud) 94
DISAPPOINTMENT 94
DISAPPROVAL 95
DISCIPLINE (Also see Laziness, Disorder) 95
DISCOMFORT 95
DISCOURAGEMENT 96
DISCUSSION 96
DISOBEDIENCE 96
DISORDER (Also see Discipline, Laziness) 97
DISSATISFACTION 97
DISTRACTION 97
DISTRESS 98
DOUBT (Also see Self Confidence) 98
DREAMS (Also see Fear of the Unknown) 98
DRIVING 99
DRUG ADDICTION 99
DRUGS 100
DUTY 101
DYNAMISM (Mental - Also see Ambition, Enthusiasm, Vitality) 101
DYSFUNCTION (Organ) 102
DYSPEPSIA 102
EASE 103
ECONOMIZING 103
ECZEMA (Also see Acne, Urticaria) 103
EDUCATION 104
EFFICIENCY 105
EFFORT 107
EGOISM 107
ELIMINATION (Bowel Movement) 107
ELOCUTION 107
EMOTION 109
EMOTION - NEGATIVE 109
(Also see Fear, Anxiety, Nervousness) 109
EMOTIONAL SCARS 110
EMPHYSEMA (Also see Asthma Attacks) 110
ENCOURAGEMENT 110
ENDURANCE (Also see Dynamism, Courage) 111
ENERGY (Also see Courage, Determination, Strength, Vitality, Health) 112

ENERGY - CREATIVE 114
ENERGY - VITAL 114
ENJOYMENT 114
ENLIGHTENED MIND 114
ENTERITIS 115
ENTERPRISING SPIRIT 115
ENTHUSIASM (Also see Ambition, Dynamism, Vitality, Energy, Courage) 116
ENURESIS (Bedwetting) 118
ENVY 119
EQUILIBRIUM (See Balance) 119
EQUANIMITY 119
ERRORS (See Mistakes) 120
ESTEEM (Also see Self Esteem) 120
EVOLUTION 120
EVOLVING 120
EXAMS 120
EXCHANGE 121
EXCUSES 122
EXERCISE (Also see Obesity, Nutrition) 122
EXPERIENCE 123
EXPRESSING ONESELF (See Self Expression) 124
FAILURE 125
FAILURE (In School) 126
FAITH (Also see Fear) 126
FALSE BELIEFS 126
FAMILY LIFE 127
FASCINATION 127
FATIGUE 128
FAULTS 128
FEAR (Also see Emotional Paralysis, Anxiety, Nervousness, Guilt, Rejection) 129
FEAR OF ANIMALS 132
FEAR OF DEATH 132
FEAR OF ELEVATORS 132
FEAR OF EXAMS (Also see Stagefright) 132
FEAR OF FAILURE (Also see Disapproval) 133
FEAR OF FLYING 133
FEAR OF REALITY 133
FEAR OF REJECTION 134

FEAR OF THE DARK 134
FEAR OF THE FUTURE 134
FEAR OF THE OPPOSITE SEX 135
FEAR OF THE UNKNOWN 135
FEAR OF VEGETABLES 136
FEAR OF WATER 136
FEELINGS 136
FEELINGS - NEGATIVE 137
FEELINGS - POSITIVE 137
FERMENTATION (Intestinal) 138
FEUDING 138
FEVER 138
FIBROMA 139
FIGHTING SPIRIT 139
FILING 139
FINANCE 139
FIRMNESS 140
FLAB 141
FORCE OF NATURE 141
FOREIGN LANGUAGES 141
FORGIVENESS (Also see Resentment) 142
FORESIGHT 143
FORTUNE 143
FOUNTAIN OF YOUTH 144
FRANKNESS (Condor) 144
FRATERNITY 144
FREEDOM 144
FREE TIME 146
FRIENDSHIP 146
FRIGIDITY 147
FRUGALITY 149
FRUSTRATION 149
FULFILLMENT 150
FUTURE 151
GAMBLING (Games Of Chance) 152
GASTRITIS (Inflammatory) 152
GENEROSITY 152
GIFTS (Talents) 153
GOALS (Aims) 153
GOLF 155

GOOD HUMOR 155
GOODNESS 156
GRATITUDE 156
GREED 157
GRIEF (Also see Forgiveness) 157
GRIEVANCES 157
GROWTH 157
GROWTH - PHYSICAL 158
GUILT (Also see Fear, Rejection) 158
HABIT 160
HABITS - NEGATIVE (Also see Diet, Habit, Hygiene) 160
HABITS - POSITIVE 161
HAIR 161
HAIR LOSS 162
HALITOSIS 162
HANDICAP 162
HAPPINESS (Also see Love of Self, Confidence, Intimacy) 162
HARMONY 165
HARMONY - CONJUGAL 165
HARMONY - FAMILIAL 167
HARVEST 167
HATE (Also see Love) 168
HEALING 168
HEALTH - GENERAL (Also see Beauty, Vitality, Confidence, Happiness, Looking Good, Slimness, Charm, Balance, Harmony, Self Expression, Peace, Success, Prosperity, Love of Life, Energy) 170
HEART 174
HEART ATTACK (See Tachycardia) 175
HELP (Also see Cooperation) 175
HESITATION 176
HINDRANCES 176
HOMEWORK (Also see Studies) 176
HOMOSEXUALITY 176
HOME 177
HONESTY (Also see Frankness, Sincerity) 177
HONOR 177
HOPE 178
HUMILIATION 178
HUMOR 178
HURT 179

HYGIENE (Also see Negative Habits, Dieting) 179
HYPERTENSION 180
HYPNOSIS 180
IDEALS 181
IDEAS 181
IDEAS - LUCRATIVE 182
IDEAS - NEGATIVE 183
IDEAS - PRECONCEIVED 183
IDENTITY 183
IGNORANCE (Also see Sickness) 184
ILLNESS (Also see Pain, Ignorance) 184
ILLUSION 185
IMAGINATION 186
IMMUNITY 187
IMPASSE 187
IMPATIENCE 187
IMPOTENCE (Also see Premature Ejaculation) 187
IMPULSE - NEGATIVE 188
INCOME 189
INDECISION 189
INDEPENDENCE 189
INDEPENDENCE - FINANCIAL 190
INDIGESTION 191
INDIVIDUALITY 191
INFARCTUS 191
INFERIORITY (See Complex) 191
INFLUENCE 191
INFORMATION 192
INNER DIALOG 193
INNER GUIDE 193
INNER PEACE 194
INNER STRENGTH (See Strength - Inner) 195
INSOMNIA 195
INSOMNIA - CHILDREN 198
INSPIRATION (Also see Creativity, Optimism) 198
INTELLECT 199
INTELLIGENCE (Also see Self Confidence) 199
INTIMACY 200
INTRANSIGENCE 200
INTRODUCTIONS 200

INTUITION 201
INVESTMENT 202
INVOLVEMENT 202
IRRITATION 203
ISOLATION 203
JEALOUSY 204
JOBS (Also see Remuneration) 205
JOIE DE VIVRE (See Love Of Life) 206
JOY 206
JUDGMENT (Also see Action, Memory) 207
JUSTICE 208
JUSTIFICATION 208
KLEPTOMANIA (Also see Honesty) 209
KNOW-HOW 209
KNOWLEDGE 209
LACK 211
LAUGHTER 211
LAZINESS 00(Also see Cowardice, Lying, Disorder, Discipline) 211
LEARNING (Also see Knowledge) 212
LEARNING WHILE ASLEEP 214
LEISURE 214
LESIONS 215
LEVEL HEADEDNESS 215
LIFE 215
LIKABILITY 215
LIMITATION 216
LISTENING 217
LOGIC 217
LOOKING GOOD (Also see Weight, Losing Weight, Health) 218
LOSING WEIGHT (Also see Strength, Body Fitness) 218
LOST OBJECTS 221
LOVE (Also see Success, Happiness) 221
LOVE - FILIAL (Also see Creativity, Sensitivity, Optimism, Peace, Joy, Tenderness) 223
LOVE LIFE 224
LOVE OF LIFE (Also see Prosperity, Happiness, Self Confidence, Courage, Peace, Love) 224
LOVE OF SELF 225
(Also see Self Confidence, Self Esteem) 225
LOYALTY 226

LUCIDITY (Also see Self Control) 226
LUCK (Also see Money) 227
LUNG DISEASE 227
LYING (Also see Cowardice, Laziness) 228
MAGNETISM (Also see Influence) 229
MAIL 231
MANAGEMENT 231
MANAGEMENT - MONEY 231
MANAGEMENT - TIME 232
MARRIAGE 233
MATURITY 233
MEDICAL TREATMENT 233
MEDITATION 233
MEETING PEOPLE 233
MEETINGS 235
MELANCHOLY 236
MEMORY (Also see Concentration, Judgment, Willpower) 236
MENOPAUSE 239
MENSTRUATION 239
MENTAL CONTROL 240
MENTAL FACULTIES (Also see Concentration) 240
MENTAL FATIGUE 240
MENTAL PROJECTION 241
MENTAL RETREAT 241
MERIT 241
MIGRAINE (Also see Emotion) 242
MIND (Also see Memory, Concentration) 243
MIND PROJECTION 244
MIRACLES 244
MISOGYNY 245
MISTAKES 245
MISTRUST 247
MISUNDERSTANDING 247
MODERATION (Also see Discernment) 247
MONEY (Also see Business, Luck, Competence, Creation, Projects, Debts, Vocations) 248
MOODS 250
MOROSITY 250
MOTIVATION 251
MOTIVATION - SCHOOL 252

MOVING 252
MUSCLES 252
MUTUAL AID 253
MYOPIA (See Nearsightedness) 253
NAIL BITING 254
NATURAL 254
NEARSIGHTEDNESS 254
NEED 255
NEGOTIATING 256
NERVOUS BREAKDOWN (Also see Nervousness) 257
NERVOUSNESS (Also see Agitation, Anxiety, Fear, Nervous Breakdown) 257
NERVOUS SYSTEM 258
NEW BEGINNING 258
NOISE 258
NOSTALGIA 258
NOVELTY 259
NUMBNESS 259
NUTRITION 260
Obesity (Also see Nutrition, Physical Exercise) 261
OBJECTIVES (Also see Goals, Self Control) 262
OBJECTIVES - FINANCIAL 264
OBLIGATION 265
OBSESSION 265
OBSTACLES 265
OFFERING 266
OPINIONS 266
OPPORTUNITY 267
OPTIMIZING 267
OPTIMISM (Also see Depression, Inspiration, Positive Thinking) 268
OPULENCE 269
ORDER 269
ORGANIZATION (Also see Efficiency) 270
ORIENTATION (See Choice of Profession) 270
ORIGINALITY 270
OUT OF BODY EXPERIENCE 271
OVERWORK 271
PACE 272
PAIN (Also see Hepatic Colic, Illness, Wounds, Anesthesia) 272
PARALYSIS 273

PARTNERSHIP 274
PASSION 274
PATIENCE (Also see Understanding) 274
PEACE (Also see Joy, Love, Wisdom, Prosperity, Health, Success) 275
PEACE OF MIND 276
PERFECTION 276
PERFECTIONISM 276
PERFORMANCE (Also see Self Image) 277
PERFORMANCE - SPORTS 277
PERSEVERANCE 278
PERSONAL AFFAIRS 279
PERSONAL DEVELOPMENT 279
PERSONALITY 280
PERSUASION 281
PESSIMISM 282
PHOBIAS (Also see Fear, Negative Thoughts) 282
PHYSICAL STRENGTH (See Strength - Physical) 282
PLANNING (Also see Money) 282
PLANS 285
PLAY 285
PLEASING OTHERS 285
PLEASURE 286
PLEASURE - SEXUAL 286
PLEURISY 286
POSITIVE MENTAL ATTITUDE 286
POSSESSION 287
POSSIBILITY 287
POTENTIAL 287
POVERTY 288
POWER (Also see Self Confidence) 288
POWER - PSYCHIC 289
POWER - SPIRITUAL 290
POWER - SUBCONSCIOUS 290
PRAISE 290
PRAYER 290
PRECAUTIONS 291
PREGNANCY 291
PREJUDICE 292
PREMATURE EJACULATION (Also see Impotence) 292
PREOCCUPATION 292

PRESSURE 292
PRIDE 293
PRIORITIES 293
PROBLEMS 293
PROCRASTINATION 296
PRODUCTION 298
PRODUCTIVITY 298
PROFESSION (Also see Wealth) 299
PROFIT 299
PROGRESS 299
PROJECTS (Also see Debt, Money) 300
PROMISES 301
PROMOTION 301
PROSPERITY (Also see Happiness, Budget, Love Of Life, Health, Peace, Success) 301
PROTECTION 303
PSYCHIC SHIELD 303
PUBLIC SPEAKING 303
PUBLISHING 305
PULMONARY DISEASE (See Lung Disease) 305
PUNCTUALITY 305
PURIFICATION 305
QUALIFICATIONS 306
QUALITY 306
QUALITY OF LIFE 306
RADIANT 307
REACTION 307
READING 307
REALIZATION 308
REALIZATION - SELF 309
REALISM 309
REBIRTH 309
RECEPTIVITY 310
REFLECTION 310
REFUSAL 311
REGULARITY - ABDOMINAL 311
REGULARITY - RESPIRATORY 311
REJECTION 312
REJOICING 312
RELATIONS - COMMERCIAL 312

RELATIONS - FAMILY 313
RELATIONS - HARMONIOUS 313
RELATIONS - HUMAN 314
RELATIONS - PERSONAL 317
RELATIONS - PROFESSIONAL 317
RELAXATION 318
RELAXATION - DIGESTIVE 320
RELAXATION - MENTAL (See Inner Peace) 320
REMUNERATION (Also see Job) 321
REPARTEE 321
REPRESSION 321
REPROACH 321
RESENTMENT (Also see Forgiveness) 322
RESOURCES 323
RESPECT (Also see Consideration) 323
RESPONSIBILITY 324
REST (Also see Sleep) 325
REST - MENTAL 326
RESULTS (Also see Work) 326
RESULTS - SCHOLASTIC 327
RETIREMENT 327
REVENUE (See Income) 327
REWARD 327
RICHES (See Wealth) 328
RISK 328
RIVALRY 329
ROUTINE 329
SADNESS 331
SALARY (Also see Work, Value) 331
SALES 332
SALES - REAL ESTATE 334
SATISFACTION 334
SAVINGS 335
SCHOOLWORK 335
SECURITY 336
SEDUCTION 336
SELF ACCEPTANCE (Also see Self Confidence, Self Expression) 337
SELF AFFIRMATION 338
SELF AWARENESS 339

SELF CONFIDENCE (Also see Self Esteem, Self Control, Competence, Depression) 340
SELF CONTROL (Also see Objectives, Calm, Self Confidence, Willpower) 343
SELF CRITICISM 345
SELF ESTEEM (Also see Self Confidence, Self Love) 345
SELF EVALUATION 347
SELF EXPRESSION (Also see Self Acceptance, Self Confidence, Health, Happiness) 347
SELF FORGIVENESS 348
SELF HEALING 349
SELF HYPNOSIS 349
SELF IMAGE 349
SELF IMPROVEMENT 351
SELF LOVE (See Love Of Self) 352
SELF PUNISHMENT 352
SELF REJECTION 352
SELF RESPECT 352
SELF SUFFICIENCY 352
SENSATION 352
SENSE OF SMELL 353
SENSES 353
SENSITIVITY 353
SENSUALITY 354
SERENITY 354
SERVICE 354
SETBACKS 355
SEXUAL BLOCKS 355
SEXUAL DESIRE 355
SEXUALITY 356
SHARING 357
SHORTNESS OF BREATH (Also see Breathing, Smoking) 358
SINCERITY (Also see Benevolence, Warmth, Honesty) 359
SKIN 359
SLANDER 359
SLAVERY 360
SLEEP (Also see Rest) 360
SLIMNESS (Also see Charm, Looking Good, Health, Weight Loss) 362
SMILING 363
SNACK 363

SOBRIETY (Also see Alcoholism) 364
SOLITUDE 365
SOLUTIONS 366
SORROW 367
SOULMATE 367
SPEECH MAKING 368
SPENDING 368
SPIRITUALITY 369
SPORT 369
STAGEFRIGHT (Also see Fear Of Exams) 370
STAYING IN SHAPE 371
STOMACH ACHE 371
STOOL (See Digestion, Appetite) 372
STRENGTH (Also see Courage) 372
STRENGTH - INNER 373
STRENGTH - PHYSICAL (Also see Losing Weight, Staying In Shape) 373
STRESS 373
STUBBORNNESS 375
STUDIES 375
STUTTERING 376
SUBCONSCIOUS MIND 377
SUBCONSCIOUS PROGRAMMING 377
SUBMISSIVENESS 377
SUCCEEDING 377
SUCCESS (Also see Wealth, Happiness, Self Love, Peace, Prosperity, Health) 379
SUCCESS - FINANCIAL (Also see Marketing, Happiness) 382
SUFFERING 383
SUGGESTION - POSITIVE 383
SUPERIORITY 383
SUPPLENESS 384
SUPPORT 384
SURGERY 384
SURPASSING 384
SWEETS 385
SYMPATHY 386
TACHYCARDIA 387
TACT 387
TALENT (Also see Gift, Salary, Service) 387
TELEMARKETING 388

TELEPATHY 388
TELEPHONE 389
TEMPTATION 390
TENACITY 390
TENDERNESS (See Love) 391
TENNIS 391
TENSION (Also see Confidence, Communication) 391
TEST 392
THOUGHTS 393
THOUGHTS - NEGATIVE 393
THOUGHTS - POSITIVE 393
TIME 395
TIMIDITY (Also see Inferiority Complex, Self Confidence) 395
TOBACCO ADDICTION (Also see Shortness Of Breath, Cough) 396
TOLERANCE 401
TOOTHACHE (Also see Pain) 402
TORMENT 402
TOXINS 403
TRANQUILLITY 403
TRANSACTIONS 404
TRANSFORMATION 404
TRAVEL TIME 404
TROUBLE 405
TRUTH 406
TUBERCULOSIS 406
TUMOUR 407
ULCERS (Also see Digestion) 408
UNCERTAINTY 408
UNDERTAKINGS 408
UNEMPLOYMENT 409
UNEXPECTED EVENTS 409
UNHAPPINESS 409
UNIFORMITY 410
UNION 410
URINE RETENTION (Uremia) 411
URTICARIA (See Eczema, Acne) 411
USED CARS 411
UTERINE BLEEDING 411
UTERINE SPASMS 411
VACATION 412

VENGEANCE 412
VERTIGO (Dizziness) 413
VICE 413
VICTORY 413
VIGILANCE 414
VIGOR 414
VISION 414
VISION - POSITIVE 415
VISUALIZATION 415
VITALITY (Also see Ambition, Dynamism, Enthusiasm, Beauty, Health) 416
VIVACITY 416
VOCATION (Also see Money) 417
VOICE 417
WAKING UP 418
WARTS 418
WEAKNESS 419
WEALTH (Also see Profession, Success, Value) 419
WEIGHT 422
WELL BEING 422
WILLPOWER (Also see Self Control, Self Confidence) 424
WINNING 427
WISDOM 428
WISDOM - SUBCONSCIOUS 429
WISH 429
WIFE 429
WONDER 430
WORK (Also see Reward, Salary, Worth) 430
WORRY 432
WORTH (Also see Wealth, Work, Salary) 434
WOUNDS (Also see Pain, Numbness) 435
WRINKLES 436
YOUTH 437

Sacred G Ascender

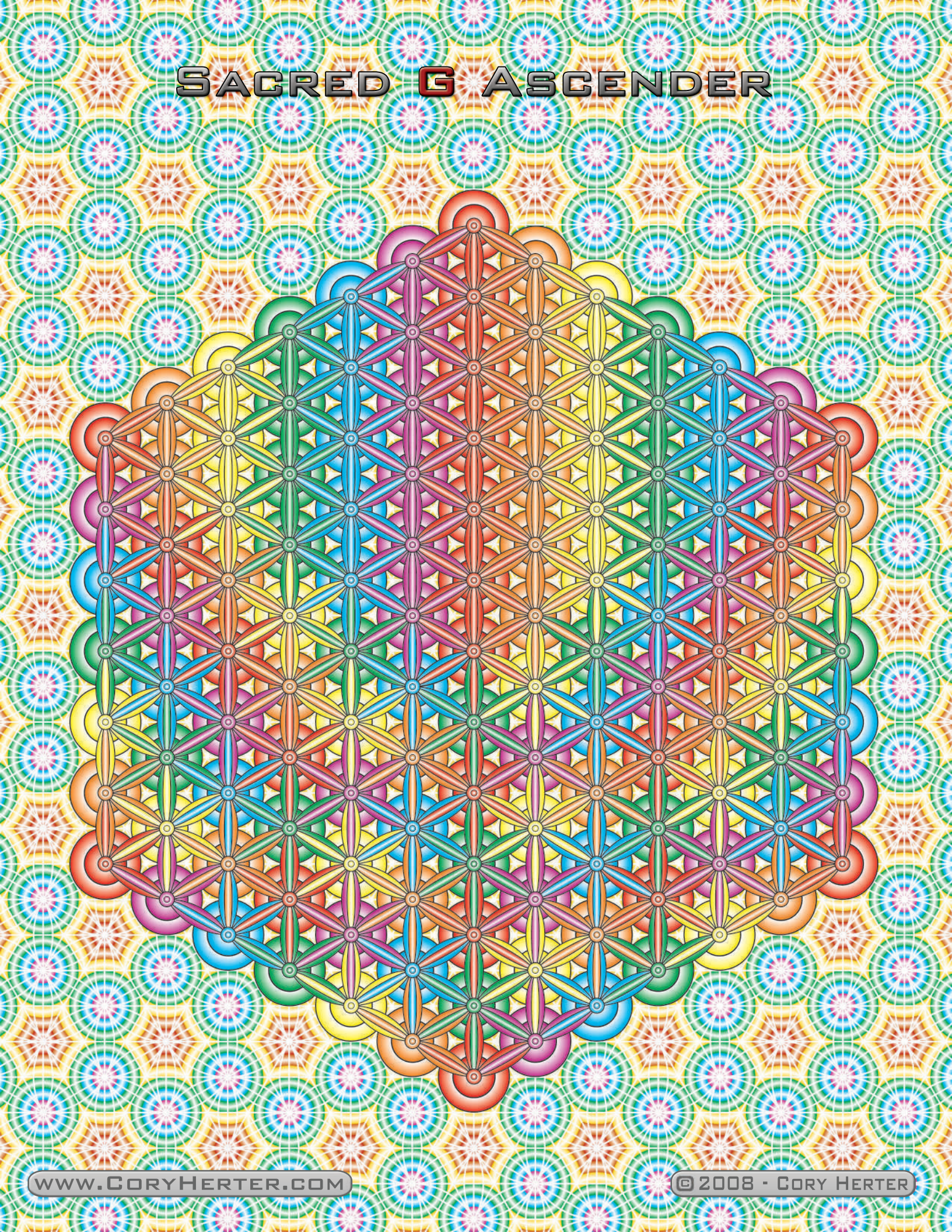

www.CoryHerter.com

© 2008 - Cory Herter

Everyone and Everything within the vicinity of Sacred G will get blasted with positive vibes that unlock energy and boost performance. Increasing the amount of Sacred G you use will increase the amount of energy you have.

Sacred G Stickers come in 6 unique designs:

The **Sacred G Spot** is designed to physically cleanse your cells, freeing up more energy for the body to run more efficiently.

The **Sacred G Transporter** is designed to free up usable energy trapped as emotion in your past memories. Each night that you sleep on Sacred G, more and more energy is released restoring you to your natural state of Peaceful Bliss.

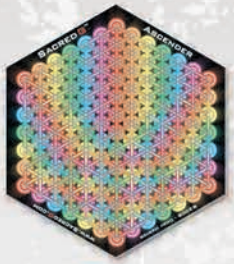

The **Sacred G Ascender** is designed to boost energy and increase self confidence by solving conflicts within your thoughts and belief systems. This enables you to perceive more information at once increasing your level of focus and reaction time.

The **Sacred G tranceNdance** is a combination of the first three designs plus advanced memory systems that accelerate the learning and downloading of new information from the collective consciousness.

The **Sacred G Simulator** is designed to upgrade your friendships and relationships by attracting new productive ones while empowering you to let go of old ones that no longer serve you. This design will increase the number of synchronicities in your life and create an awakened sense of adventure and purpose.

The **Sacred G L.O.V.E.™ Activator** combines all of the designs with a new motion activation system that takes you beyond your limits to achieve your goals. The Activator is the Ultimate State-of-the-Art DNA Activation Technology.

Visit the website to learn about the most advanced Sacred G Design called "Fusion" which incorporates every design into One.

Sleep / Pillow Case

Place the Sacred G beneath your pillow or inside your pillowcase design side up.

Notice how the Sacred G alters your sleep and dream states. The body frees up the most amount of energy while you sleep. Sacred G enhances this process continually increasing the amount of energy you feel.

Energized Water

Super Charge your water with Energy by placing it on top of the Sacred G.

Test: Take two glasses of tap water, place one on top of the Sacred G, and the other at least six feet away. Allow the water to charge overnight, then taste the difference between the two glasses in the morning. This test shows how Sacred G Technology not only cleans your water, but is also instrumental in physically cleansing the toxins from your body increasing energy.

"I Will never ride a board that's not G'd out again!"

Rob Dyrdek

X-treme Performance Enhancement: Place the Sacred G on skateboards, surfboards, team uniforms, or any gear and equipment that you use. Watch as your performance is enhanced and your game is improved.

Visit the website to learn over 35 ways to use Sacred G to boost your energy, plus 5 cool tests that you can do. Also, be sure to watch the Sacred G episode on MTV's Rob & Big show to see how Sacred G can boost your Performance.

Visit the website to download all of the Sacred G Designs for FREE! Bonus: Watch a 3 hour video on the science of Sacred Geometry!

www.CoryHerter.com

©2008 - Cory Herter

Sacred G Spot

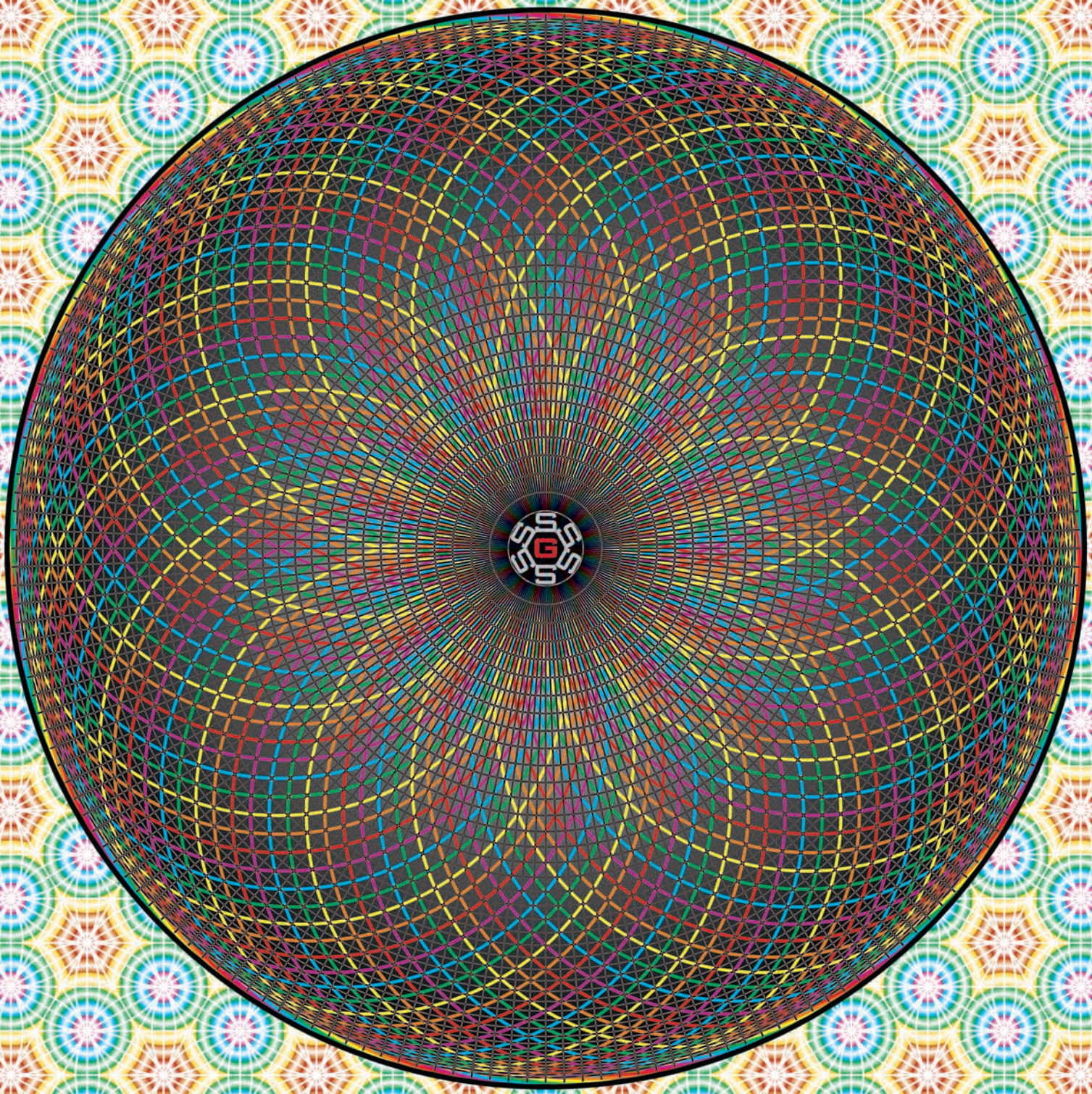

www.CoryHerter.com

© 2008 - Cory Herter

Sacred G Transporter

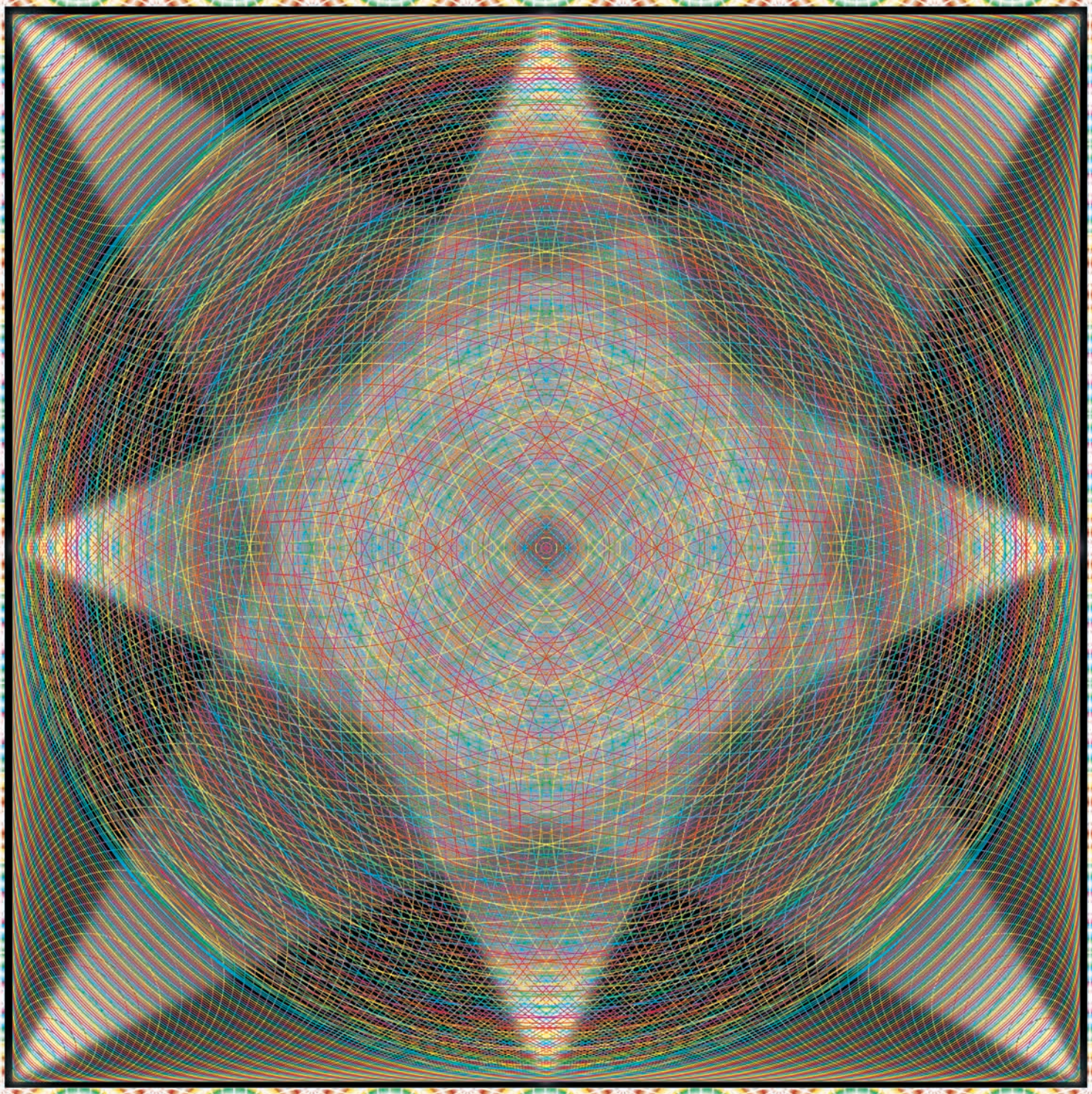

www.CoryHerter.com

© 2008 - Cory Herter

Sacred G Ascender

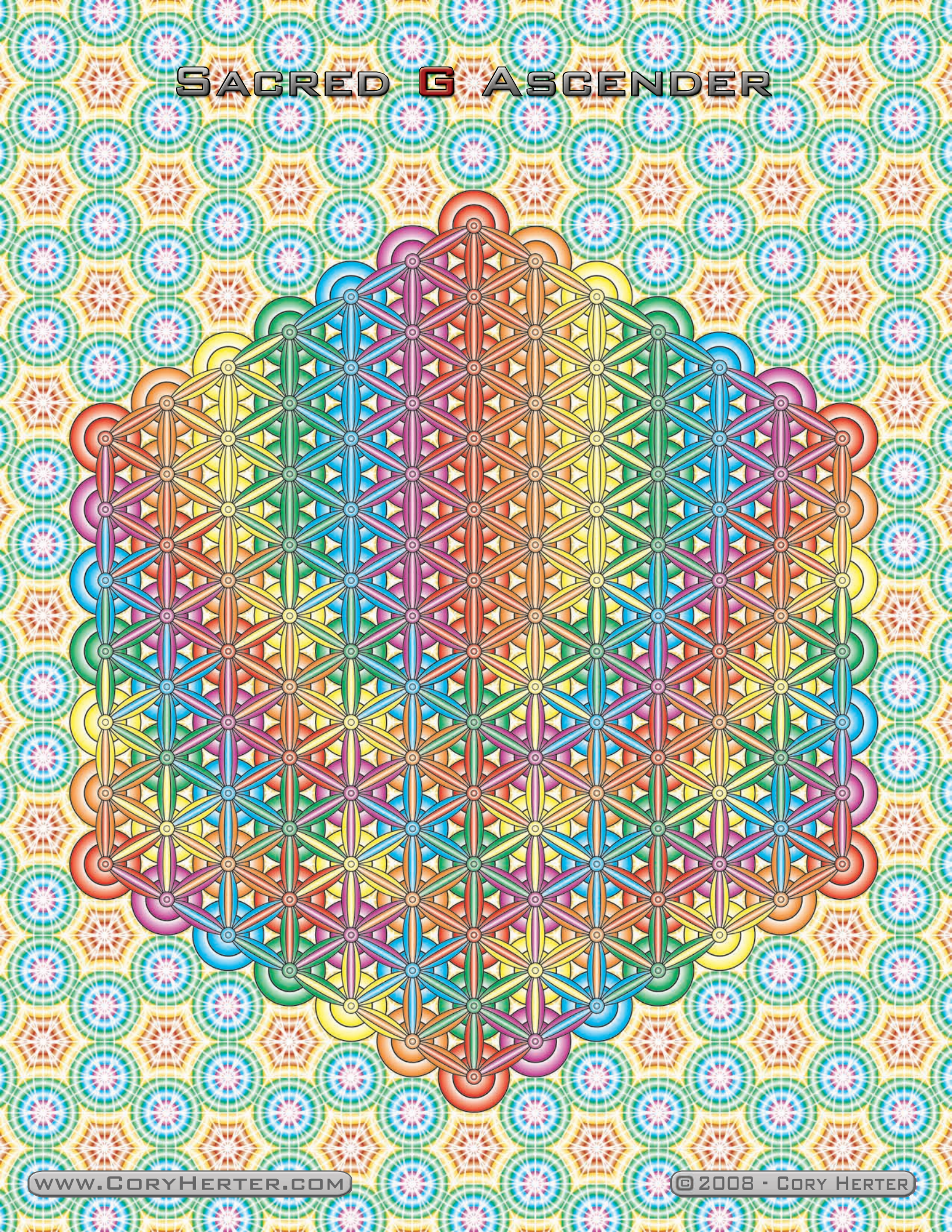

www.CoryHerter.com

© 2008 - Cory Herter

Sacred G TranceNdance

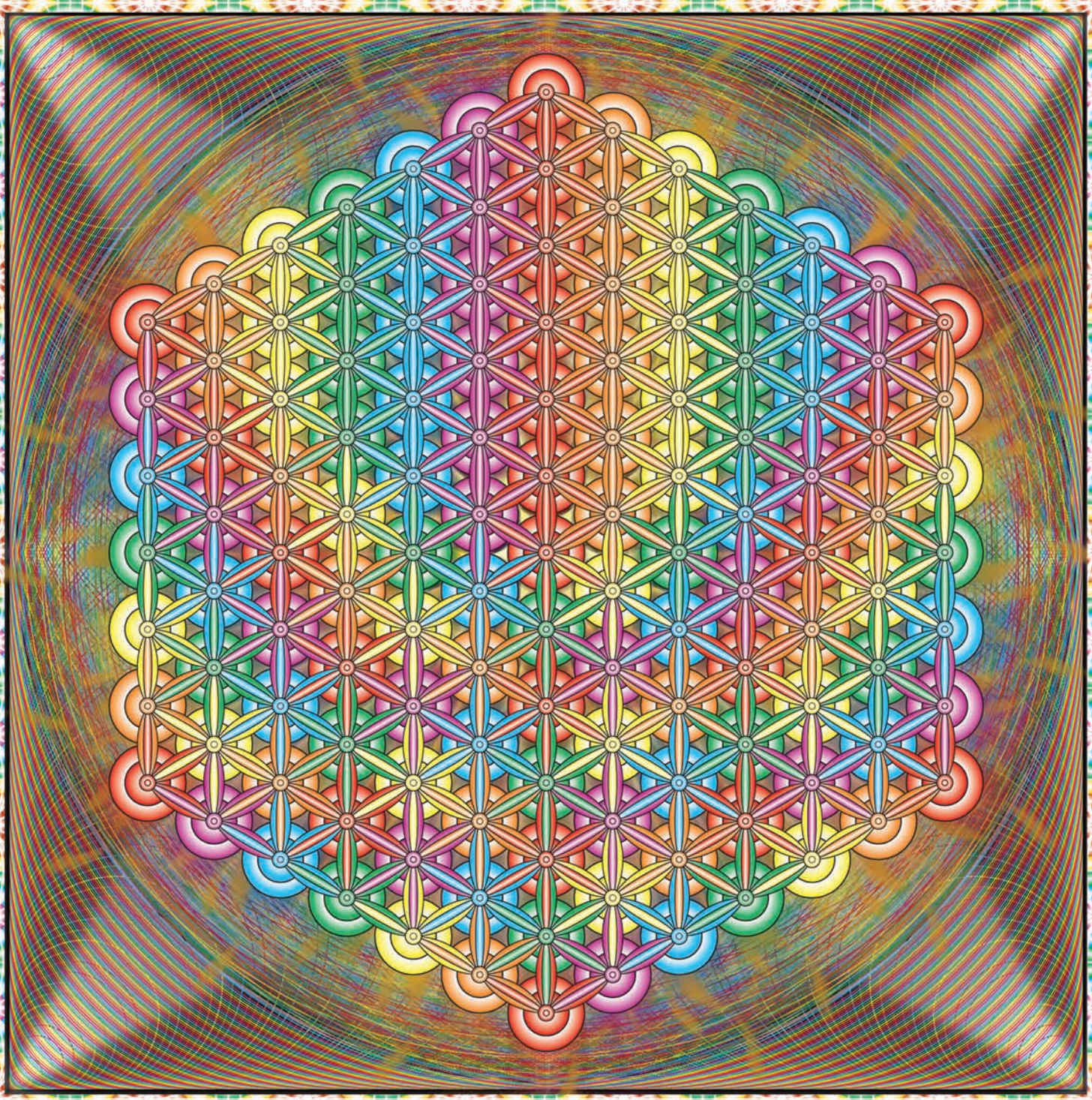

www.CoryHerter.com

© 2008 - Cory Herter

Sacred G Simulator

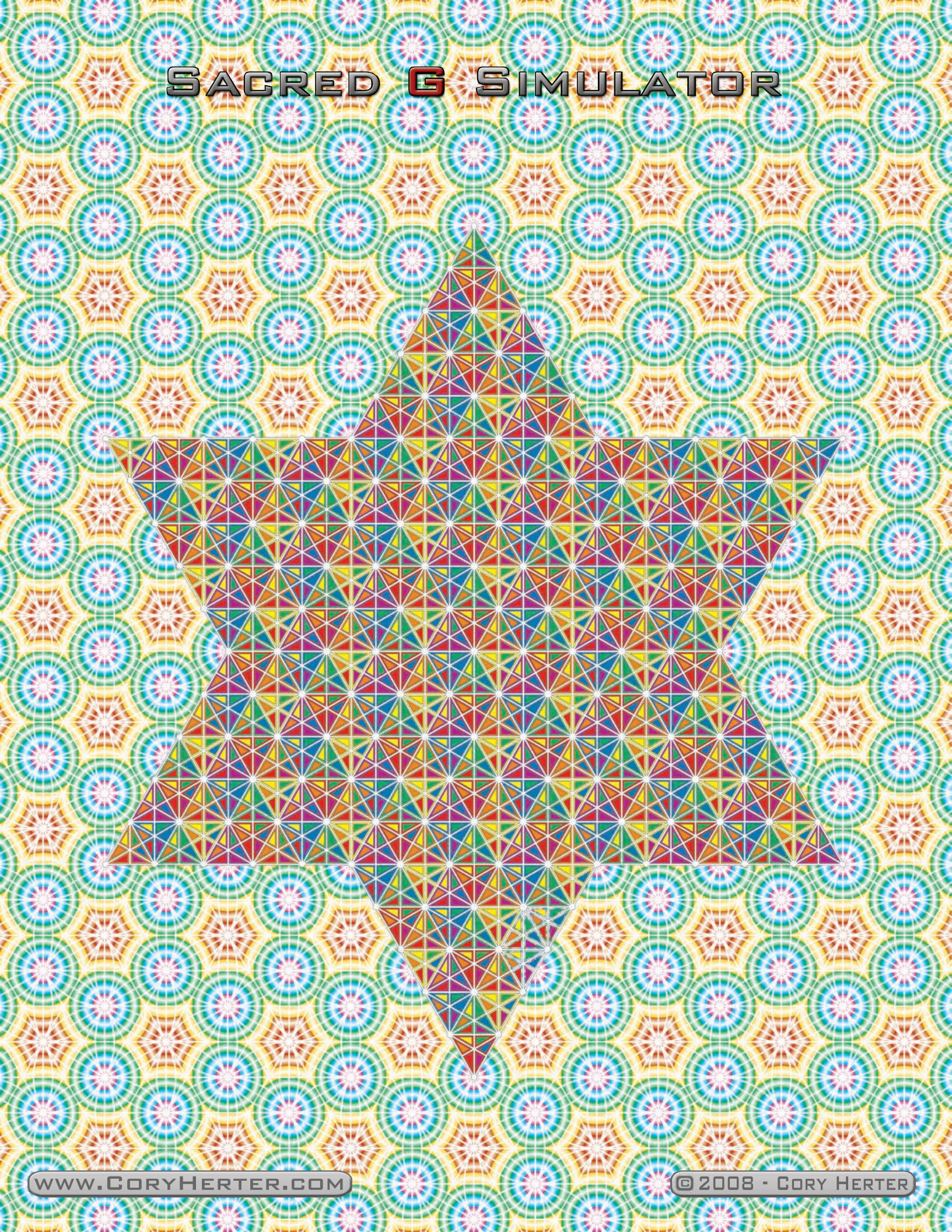

www.CoryHerter.com

© 2008 - Cory Herter

Sacred G Activator

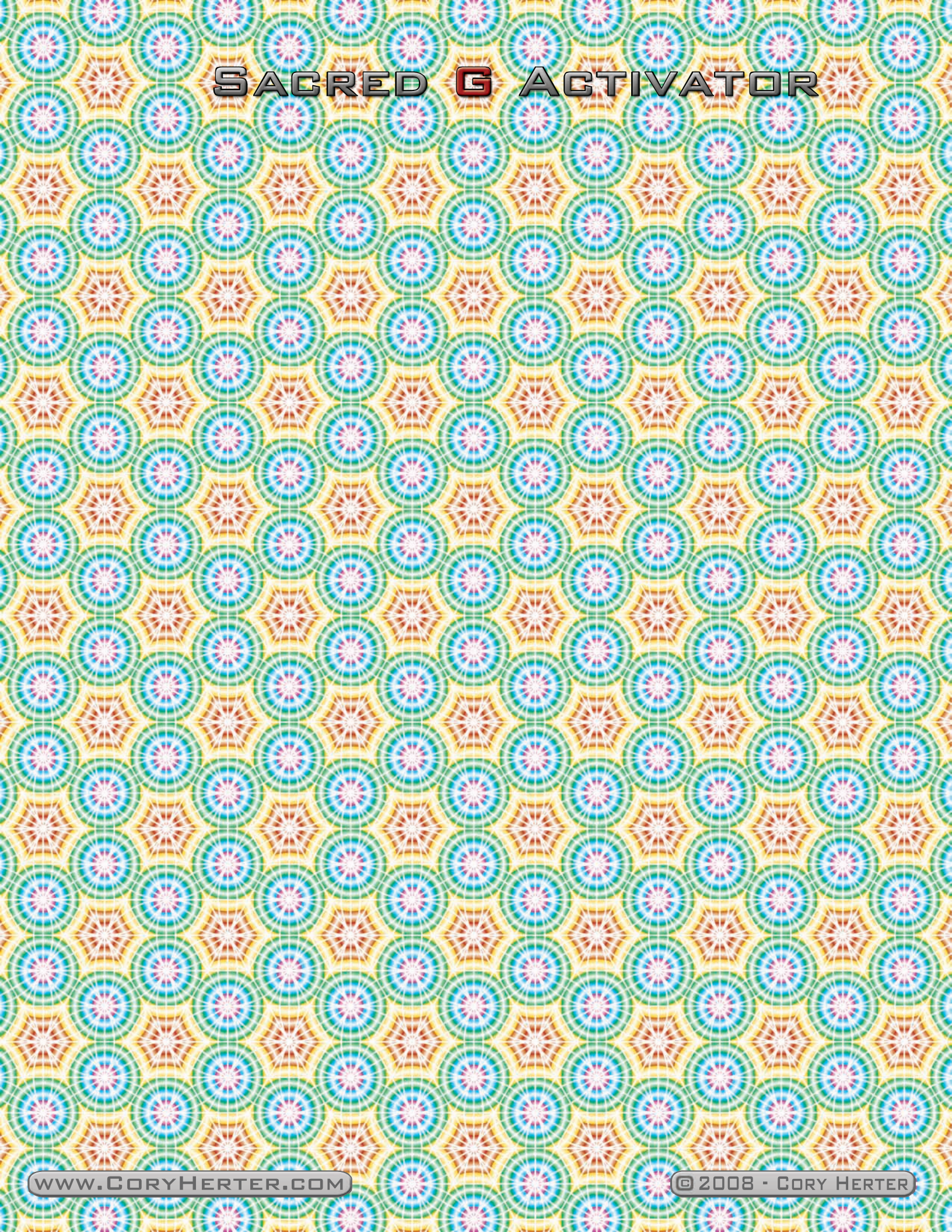

www.CoryHerter.com

© 2008 - Cory Herter

BKS IYENGAR

FOREWORD BY YEHUDI MENUHIN

The Illustrated Light on Yoga

An Easy-to-follow Version of the Classic Introduction to Yoga

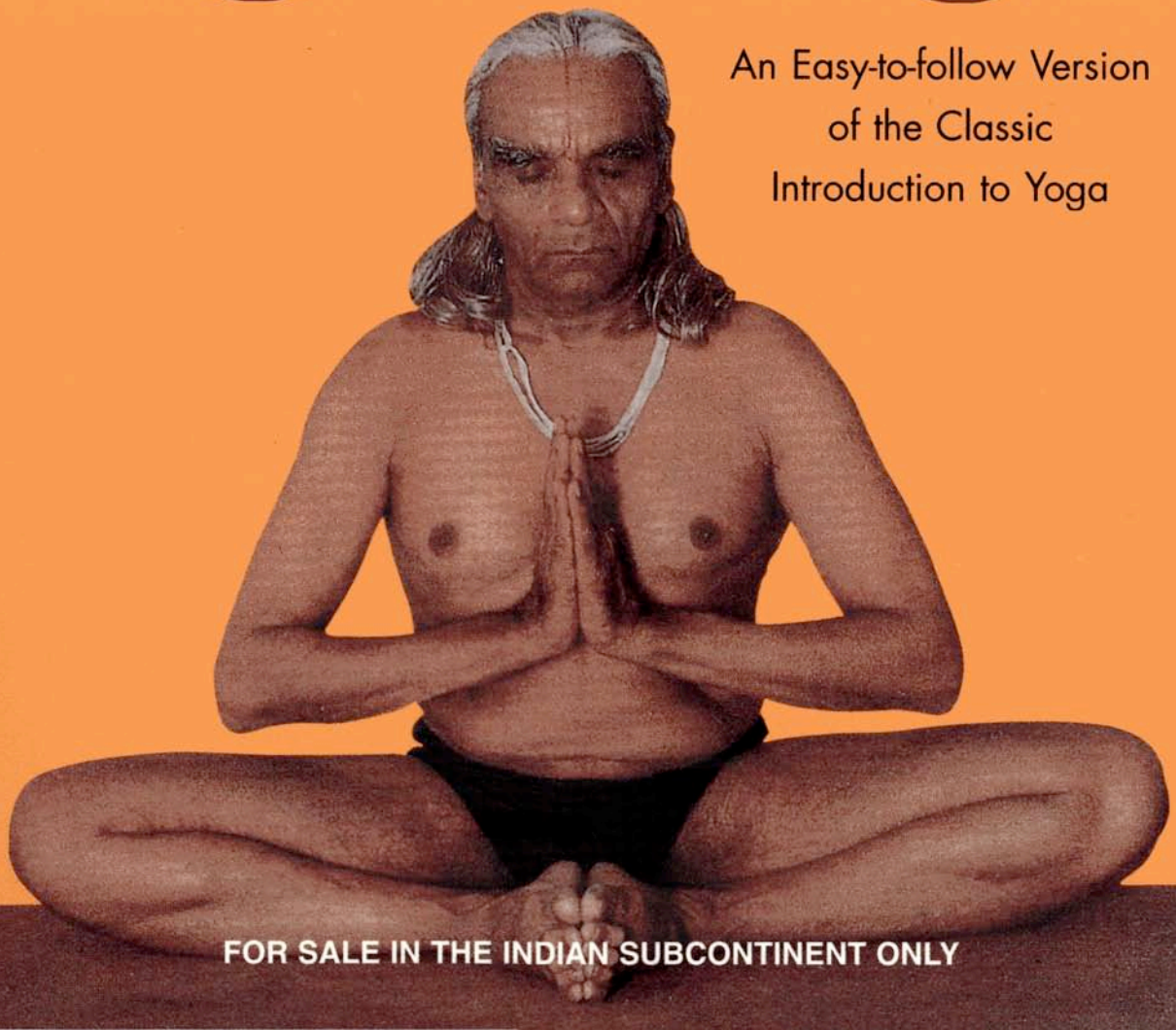

FOR SALE IN THE INDIAN SUBCONTINENT ONLY

DEDICATED TO MY REVERED GURUJĪ

Sāmkya-yoga-Śikhāmaṇi; Veda-kesari; Vedāntavāgīśa; Nyāyāchārya; Mīmāmsa-ratna; Mīmāmsa-thīrtha

Professor, Śrīmān, T. Krishnamāchārya of
Mysore (South India), India

PRAYER

'I bow before the noblest of sages, Patañjali, who brought serenity of mind by his work on yoga, clarity of speech by his work on grammar and purity of body by his work on medicine.'

'I salute Ādiśvara (the Primeval Lord Śiva) who taught first the science of Haṭha Yoga – a science that stands out as a ladder for those who wish to scale the heights of Rāja Yoga.'

PART I

What is Yoga?

The word Yoga is derived from the Sanskrit root yuj meaning to bind, join, attach and yoke, to direct and concentrate one's attention on, to use and apply. It also means union or communion. It is the true union of our will with the will of God. 'It thus means, says Mahadev Desai in his introduction to the *Gita according to Gandhi*, 'the yoking of all the powers of body, mind and soul to God; it means the disciplining of the intellect, the mind, the emotions, the will, which that Yoga presupposes; it means a poise of the soul which enables one to look at life in all its aspects evenly.'

Yoga is one of the six orthodox systems of Indian philosophy. It was collated, co-ordinated and systematized by Patañjali in his classical work, the *Yoga Sutras*, which consists of 185 terse aphorisms. In Indian thought, everything is permeated by the Supreme Universal Spirit (Paramātmā or God) of which the individual human spirit (jīvātmā) is a part. The system of yoga is so called because it teaches the means by which the jīvātmā can be united to, or be in communion with the Paramātmā, and so secure liberation (mokṣa).

One who follows the path of Yoga is a yogi or yogin.

In the sixth chapter of the *Bhagavad Gītā*, which is the most important authority on Yoga philosophy, Śri Krishna explains to Arjuna the meaning of Yoga as a deliverance from contact with pain and sorrow. It is said:

> 'When his mind, intellect and self (ahaṁkāra) are under control, freed from restless desire, so that they rest in the spirit within, a man becomes a Yukta – one in communion with God. A lamp does not flicker in a place where no winds blow; so it is with a yogi, who controls his mind, intellect and self, being absorbed in the spirit within him. When the restlessness of the mind, intellect and self is stilled through the practice of Yoga, the yogi by the grace of the Spirit within himself finds fulfilment. Then he knows the joy eternal which is beyond the pale of the senses which his reason cannot grasp. He abides in this reality and moves not therefrom. He has found the treasure above all others. There is nothing higher than this. He who has achieved it, shall not be moved by the greatest sorrow. This is the real meaning of Yoga – a deliverance from contact with pain and sorrow.

As a well cut diamond has many facets, each reflecting a different colour of light, so does the word yoga, each facet reflecting a different shade of meaning and revealing different aspects of the entire range of human endeavour to win inner peace and happiness.

The *Bhagavad Gītā* also gives other explanations of the term yoga and lays stress upon Karma Yoga (Yoga by action). It is said: 'Work alone is your privilege, never the fruits thereof. Never let the fruits of action be your motive; and never cease to work. Work in the name of the Lord, abandoning selfish desires. Be not affected by success or failure. This equipoise is called Yoga.'

Yoga has also been described as wisdom in work or skilful living amongst activities with harmony and moderation. 'Yoga is not for him who gorges too much, nor for him who starves himself. It is not for him who sleeps too much, nor for him who stays awake. By moderation in eating and in resting, by regulation in working and by concordance in sleeping and waking, Yoga destroys all pain and sorrow.'

The *Kaṭhopanishad* describes Yoga thus: 'When the senses are stilled, when the mind is at rest, when the intellect wavers not – then, say the wise, is reached the highest stage. This steady control of the senses and mind has been defined as Yoga. He who attains it is free from delusion.'

In the second aphorism of the first chapter of the *Yoga Sutras*, Patañjali describes Yoga as *'chitta vṛtti nirodhah'*. This may be translated as the restraint (nirodhah) of mental (chitta) modifications (vṛtti) or as suppression (nirodhah) of the fluctuations (vṛtti) of consciousness (chitta). The word chitta denotes the mind in its total or collective sense as being composed of three categories:

(a) mind (manas, that is, the individual mind having the power and faculty of attention, selection and rejection; it is the oscillating indecisive faculty of the mind);
(b) intelligence or reason (buddhi, that is, the decisive state which determines the distinction between things);
(c) ago (ahaṁkāra, literally the I-maker, the state which ascertains that 'I know').

The word vṛtti is derived from the Sanskrit root vṛt meaning to turn, to revolve, to roll on. It thus means course of action, behaviour, mode of being, condition or mental state. Yoga is the method by which the restless mind is calmed and the energy directed into constructive channels. As a mighty river which when properly harnessed by dams and canals, creates a vast reservoir of water, prevents famine and provides abundant power for industry; so also the mind, when controlled, provides a reservoir of peace and generates abundant energy for human uplift.

The problem of controlling the mind is not capable of easy solution,

as borne out by the following dialogue in the sixth chapter of the *Bhagavad Gītā*. Arjuna asks Śri Krishna:

> 'Krishna, you have told me of Yoga as a communion with Brahman (the Universal Spirit), which is ever one. But how can this be permanent, since the mind is so restless and inconsistent? The mind is impetuous and stubborn, strong and wilful, as difficult to harness as the wind.' Śri Krishna replies: 'Undoubtedly, the mind is restless and hard to control. But it can be trained by constant practice (abhyāsa) and by freedom from desire (vairāgya). A man who cannot control his mind will find it difficult to attain this divine communion; but the self-controlled man can attain it if he tries hard and directs his energy by the right means.'

THE STAGES OF YOGA

The right means are just as important as the end in view. Patañjali enumerates these means as the eight limbs or stages of Yoga for the quest of the soul. They are:

1. Yama (universal moral commandments);
2. Niyama (self purification by discipline);
3. Asana (posture);
4. Prāṇāyāma (rhythmic control of the breath);
5. Pratyāhāra (withdrawal and emancipation of the mind from the domination of the senses and exterior objects);
6. Dhāraṇa (concentration);
7. Dhyāna (meditation);
8. Samādhi (a state of super-consciousness brought about by profound meditation, in which the individual aspirant (sādhaka) becomes one with the object of his meditation – Paramātmā or the Universal Spirit).

Yama and Niyama control the yogi's passions and emotions and keep him in harmony with his fellow man. Ásanas keep the body healthy and strong and in harmony with nature. Finally, the yogi becomes free of body consciousness. He conquers the body and renders it a fit vehicle for the soul. The first three stages are the outward quests (bahiranga sādhanā).

The next two stages, Prāṇāyāma and Pratyāhāra, teach the aspirant to regulate the breathing, and thereby control the mind. This helps to free the senses from the thraldom of the objects of desire. These two stages of Yoga are known as the inner quests (antaranga sādhanā).

Dhāranā, Dhyāna and Samādhi take the yogi into the innermost

recesses of his soul. The yogi does not look heavenward to find God. He knows that HE is within, being known as the Antarātmā (the Inner Self). The last three stages keep him in harmony with himself and his Maker. These stages are called antarātmā sādhanā, the quest of the soul.

By profound meditation, the knower, the knowledge and the known become one. The seer, the sight and the seen have no separate existence from each other. It is like a great musician becoming one with his instrument and the music that comes from it. Then, the yogi stands in his own nature and realizes his self (Ātman), the part of the Supreme Soul within himself.

There are different paths (mārgas) by which a man travels to his Maker. The active man finds realization through Karma Mārga, in which a man realizes his own divinity through work and duty. The emotional man finds it through Bhakti Mārga, where there is realization through devotion to and love of a personal God. The intellectual man pursues Jñana, Mārga, where realization comes through knowledge. The meditative or reflective man follows Yoga Mārga, and realizes his own divinity through control of the mind.

Happy is the man who knows how to distinguish the real from the unreal, the eternal from the transient and the good from the pleasant by his discrimination and wisdom. Twice blessed is he who knows true love and can love all God's creatures. He who works selflessly for the welfare of others with love in his heart is thrice blessed. But the man who combines within his mortal frame knowledge, love and selfless service is holy and becomes a place of pilgrimage, like the confluence of the rivers Gangā, Saraswatī and Jamunā. Those who meet him become calm and purified.

Mind is the king of the senses. One who has conquered his mind, senses, passions, thought and reason is a king among men. He is fit for Rāja Yoga, the royal union with the Universal Spirit. He has Inner Light.

He who has conquered his mind is Rāja Yogi. The word rāja means a king. The expression Rāja Yoga implies a complete mastery of the Self. Though Patañjali explains the ways to control the mind, he nowhere states in his aphorisms that this science is Rāja Yoga, but calls it Aṣṭāṅga Yoga or the eight stages (limbs) of Yoga. As it implies complete mastery of the self one may call it the science of Rāja Yoga.

Swaźtmārāma, the author of the *Hatha Yoga Pradīpikā* (haṭha = force or determined effort) called the same path Haṭha Yoga because it demanded rigorous discipline.

It is generally believed that Rāja Yoga and Haṭha Yoga are entirely distinct, different and opposed to each other, that the *Yoga Sutras* of Patañjāli deal with Spiritual discipline and that the *Haṭha Yoga Pradīpikā* of Swātmārāma deals solely with physical discipline. It is not so, for

Haṭha Yoga and Rāja Yoga complement each other and form a single approach towards Liberation. As a mountaineer needs ladders, ropes and crampons as well as physical fitness and discipline to climb the icy peaks of the Himālayas, so does the Yoga aspirant need the knowledge and discipline of the Haṭha Yoga of Swātmārāma to reach the heights of Rāja Yoga dealt with by Patañjali.

This path of Yoga is the fountain for the other three paths. It brings calmness and tranquillity and prepares the mind for absolute unqualified self-surrender to God, in which all these four paths merge into one.

Chitta Vṛtti (*Causes for the Modification of the Mind*)

In his *Yoga Sutras* Patañjali lists five classes of chitta vṛtti which create pleasure and pain. These are:

1. Pramāṇa (a standard or ideal), by which things or values are measured by the mind or known, which men accept upon (a) direct evidence such as perception (pratyakṣa), (b) inference (anumāna) and (c) testimony or the word of an acceptable authority when the source of knowledge has been checked as reliable and trustworthy (āgama).

2. Viparyaya (a mistaken view which is observed to be such after study). A faulty medical diagnosis based on wrong hypotheses, or the formerly held theory in astronomy that the Sun rotates round the Earth, are examples of viparyaya.

3. Vikalpa (fancy or imagination, resting merely on verbal expression without any factual basis). A beggar may feel happy when he imagines himself spending millions. A rich miser, on the other hand, may starve himself in the belief that he is poor.

4. Nidrā (sleep), where there is the absence of ideas and experiences. When a man is sleeping soundly, he does not recall his name, family or status, his knowledge or wisdom, or even his own existence. When a man forgets himself in sleep, he wakes up refreshed. But, if a disturbing thought creeps into his mind when he is dropping off, he will not rest properly.

5. Smṛti (memory, the holding fast of the impressions of objects that one has experienced). There are people who live in their past experiences, even though the past is beyond recall. Their sad or happy memories keep them chained to the past and they cannot break their fetters.

Patañjali enumerates five causes of chitta vṛtti creating pain (kleśa). These are:

1. Avidyā (ignorance or nescience);
2. asmitā (the feeling of individuality which limits a person and distinguishes him from a group and which may be physical, mental, intellectual or emotional);
3. rāga (attachment or passion);
4. dveśa (aversion or revulsion);
5. abhiniveśa (love or thirst for life, the instinctive clinging to worldly life and bodily enjoyment and the fear that one may be cut off from all this by death).

These causes of pain remain submerged in the mind of the sādhaka (the aspirant or seeker). They are like icebergs barely showing their heads in the polar seas. So long as they are not studiously controlled and eradicated, there can be no peace. The yogi learns to forget the past and takes no thought for the morrow. He lives in the eternal present.

As a breeze ruffles the surface of a lake and distorts the images reflected therein, so also the chitta vṛtti disturb the peace of mind. The still waters of a lake reflect the beauty around it. When the mind is still, the beauty of the Self is seen reflected in it. The yogi stills his mind by constant study and by freeing himself from desires. The eight stages of Yoga teach him the way.

Chitta Viksepa (*Distractions and Obstacles*)

The distractions and obstacles which hinder the aspirant's practice of Yoga are:

1. Vyādhi – sickness which disturbs the physical equilibrium
2. Styāna – languor or lack of mental disposition for work
3. Saṁśaya – doubt or indecision
4. Pramāda – indifference or insensibility
5. Ālasya – laziness
6. Avirati – sensuality, the rousing of desire when sensory objects possess the mind
7. Bhrānti Darśana – false or invalid knowledge, or illusion
8. Alabdha Bhūmikatva – failure to attain continuity of thought or concentration so that reality cannot be seen
9. Anavasthitattva – instability in holding on to concentration which has been attained after long practice.

There are, however, four more distractions: (a) duḥkha – pain or misery, (b) daurmanasya – despair, (c) aṅgamejayatva – unsteadiness of the body and (d) śvāsa-praśvāsa – unsteady respiration.

To win a battle, a general surveys the terrain and the enemy and plans counter-measures. In a similar way the Yogi plans the conquest of the Self.

1. *Vyādhi.* It will be noticed that the very first obstacle is ill-health or sickness. To the yogi his body is the prime instrument of attainment. If his vehicle breaks down, the traveller cannot go far. If the body is broken by ill-health, the aspirant can achieve little. Physical health is important for mental development, as normally the mind functions through the nervous system. When the body is sick or the nervous system is affected, the mind becomes restless or dull and inert and concentration or meditation become impossible.

2. *Styāna.* A person suffering from languor has no goal, no path to follow and no enthusiasm. His mind and intellect become dull due to inactivity and their faculties rust. Constant flow keeps a mountain stream pure, but water in a ditch stagnates and nothing good can flourish in it. A listless person is like a living corpse for he can concentrate on nothing.

3. *Saṁśaya.* The unwise, the faithless and the doubter destroy themselves. How can they enjoy this world or the next or have any happiness? The seeker should have faith in himself and his master. He should have faith that God is ever by his side and that no evil can touch him. As faith springs up in the heart it dries out lust, ill-will, mental sloth, spiritual pride and doubt, and the heart free from these hindrances becomes serene and untroubled.

4. *Pramāda.* A person suffering from pramāda is full of self-importance, lacks any humility and believes that he alone is wise. No doubt he knows what is right or wrong, but he persists in his indifference to the right and chooses what is pleasant. To gratify his selfish passions and dreams of personal glory, he will deliberately and without scruple sacrifice everyone who stands in his way. Such a person is blind of God's glory and deaf to His words.

5. *Ālasya.* To remove the obstacle of laziness, unflagging enthusiasm (vīrya) is needed. The attitude of the aspirant is like that of a lover ever yearning to meet the beloved but never giving way to despair. Hope should be his shield and courage his sword. He should be free from hate and sorrow. With faith and enthusiasm he should overcome the inertia of the body and the mind.

6. *Avirati.* This is the tremendous craving for sensory objects after they have been consciously abandoned, which is so hard to restrain. Without being attached to the objects of sense, the yogi learns to enjoy them with the aid of the senses which are completely under his control. By the practice of pratyāhāra he wins freedom from attachment and emancipation from desire and becomes content and tranquil.

7. *Bhrānti Darśana.* A person afflicted by false knowledge suffers from delusion and believes that he alone has seen the true Light. He has a powerful intellect but lacks humility and makes a show of wisdom. By remaining in the company of great souls and through their guidance he sets his foot firmly on the right path and overcomes his weakness.

8. *Alabdha Bhūmikatva.* As a mountain climber fails to reach the summit for lack of stamina, so also a person who cannot overcome the inability to concentrate is unable to seek reality. He might have had glimpses of reality but he cannot see clearly. He is like a musician who has heard divine music in a dream, but who is unable to recall it in his waking moments and cannot repeat the dream.

9. *Anavasthitattva.* A person affected with anavasthitattva has by hard work come within sight of reality. Happy and proud of his achievements he becomes slack in his practice (sâdhanā). He has purity and great power of concentration and has come to the final cross-roads of his quest. Even at this last stage continuous endeavour is essential and he has to pursue the path with infinite patience and determined perseverance and must never show slackness which hampers progress on the path of God realization. He must wait until divine grace descends upon him. It has been said in the *Kaṭhopanisad*: 'The Self is not to be realized by study and instruction, nor by subtlety of intellect, nor by much learning, but only by him who longs for Him, by the one whom He chooses. Verily to such a one the Self reveals His true being.

To overcome the obstacles and to win unalloyed happiness, Patañjali offered several remedies. The best of these is the fourfold remedy of Maitri (friendliness), Karuṇa (compassion), Muditā (delight) and Upekṣā (disregard).

1. Maitri is not merely friendliness, but also a feeling of oneness with the object of friendliness (ātmīyatā). A mother feels intense happiness at the success of her children because of ātmīyatā, a feeling of oneness. Patañjali recommends maitri for sukha (happiness or virtue). The yogi cultivates maitri and ātmīyatā for the good and turns enemies into friends, bearing malice towards none.

2. Karuṇa is not merely showing pity or compassion and shedding tears of despair at the misery (duḥkha) of others. It is compassion coupled with devoted action to relieve the misery of the afflicted. The yogi uses all his resources – physical, economic, mental or moral – to alleviate the pain and suffering of others. He shares his strength with the weak until they become strong. He shares his courage with those that are timid until they become brave by his example. He denies the maxim of the 'survival of the fittest', but makes the weak strong enough to survive. He becomes a shelter to one and all.

3. Muditā is a feeling of delight at the good work (punya) done by another, even though he may be a rival. Through muditā, the yogi saves himself from much heart-burning by not showing anger, hatred or jealousy for another who has reached the desired goal which he himself has failed to achieve.

4. Upekṣā: It is not merely a feeling of disdain or contempt for the person who has fallen into vice (apuṇya) or one of indifference or superiority towards him. It is a searching self-examination to find out how one would have behaved when faced with the same temptations. It is also an examination to see how far one is responsible for the state into which the unfortunate one has fallen and the attempt thereafter to put him on the right path. The yogi understands the faults of others by seeing and studying them first in himself. This self-study teaches him to be charitable to all.

The deeper significance of the fourfold remedy of maitri, karuṇa, muditā and upekṣā cannot be felt by an unquiet mind. My experience has led me to conclude that for an ordinary man or woman in any community of the world, the way to achieve a quiet mind is to work with determination on two of the eight stages of Yoga mentioned by Patañjali, namely, āsana and prāṇāyāma.

The mind (manas) and the breath (prāṇa) are intimately connected and the activity or the cessation of activity of one affects the other. Hence Patañjali recommended prāṇāyāma (rhythmic breath control) for achieving mental equipoise and inner peace.

Śiṣya and Guru (*A Pupil and a Master*)

The *Śiva Saṁhitā* divides sādhakas (pupils or aspirants) into four classes. They are (1) mṛdu (feeble), (2) madhyama (average), (3) adhimātra (superior) and (4) adhimātratama (the supreme one). The last, the highest, is alone able to cross beyond the ocean of the manifest world.

The feeble seekers are those who lack enthusiasm, criticize their teachers, are rapacious, inclined to bad action, eat much, are in the power of women, unstable, cowardly, ill, dependent, speak harshly, have weak characters and lack virility. The Guru (Teacher or Master) guides such seekers in the path of Mantra Yoga only. With much effort, the sādhaka can reach enlightenment in twelve years. (The word mantra is derived from the root 'man', meaning to think. Mantra thus means a sacred thought or prayer to be repeated with full understanding of its meaning. It takes a long time, perhaps years, for a mantra to take firm root in the mind of a feeble sādhaka and still longer for it to bear fruit.)

Of even mind, capable of bearing hardship, wishing to perfect the

work, speaking gently, moderate in all circumstances, such is the average seeker. Recognizing these qualities, the Guru teaches him Laya Yoga, which gives liberation. (Laya means devotion, absorption or dissolution.)

Of stable mind, capable of Laya Yoga, virile, independent, noble, merciful, forgiving, truthful, brave, young, respectful, worshipping his teacher, intent on the practice of Yoga, such is a superior seeker. He can reach enlightenment after six years of practice. The Guru instructs this forceful man in Haṭha Yoga.

Of great virility and enthusiasm, good looking, courageous, learned in scriptures, studious, sane in mind, not melancholic, keeping young, regular in food, with his senses under control, free from fear, clean, skilful, generous, helpful to all, firm, intelligent, independent, forgiving, of good character, of gentle speech and worshipping his Guru, such is a supreme seeker, fit for all forms of Yoga. He can reach enlightenment in three years.

Although the *Śiva Saṁhitā* and the *Haṭha Yoga Pradīpikā* mention the period of time within which success might be achieved, Patañjali nowhere lays down the time required to unite the individual soul with the Divine Universal Soul. According to him abhyāsa (constant and determined practice) and vairāgya (freedom from desires) make the mind calm and tranquil. He defines abhyāsa as effort of long duration, without interruption, performed with devotion, which creates a firm foundation.

The study of Yoga is not like work for a diploma or a university degree by someone desiring favourable results in a stipulated time.

The obstacles, trials and tribulations in the path of Yoga can be removed to a large extent with the help of a Guru. (The syllable gu means darkness and ru means light. He alone is a Guru who removes darkness and brings enlightenment.) The conception of a Guru is deep and significant. He is not an ordinary guide. He is a spiritual teacher who teaches a way of life, and not merely how to earn a livelihood. He transmits knowledge of the Spirit and one who receives such knowledge is a śiṣya, a disciple.

The relationship between a Guru and a śiṣya is a very special one, transcending that between parent and child, husband and wife or friends. A Guru is free from egotism. He devotedly leads his śiṣya towards the ultimate goal without any attraction for fame or gain. He shows the path of God and watches the progress of his disciple, guiding him along that path. He inspires confidence, devotion, discipline, deep understanding and illumination through love. With faith in his pupil, the Guru strains hard to see that he absorbs the teaching. He encourages him to ask questions and to know the truth by question and analysis.

A śiṣya should possess the necessary qualifications of higher realization and development. He must have confidence, devotion and love for his Guru. The perfect examples of the relationship between a Guru and a śiṣya are those of Yama (the God of Death) and Nachiketā in the *Kaṭhopaniṣad* and of Śri Krishna and Arjuna in the *Bhagavad Gīta*. Nachiketā and Arjuna obtained enlightenment through their one-pointed mind, their eagerness and questioning spirit. The śiṣya should hunger for knowledge and have the spirit of humility, perseverance and tenacity of purpose. He should not go to the Guru merely out of curiosity. He should possess śraddhā (dynamic faith) and should not be discouraged if he cannot reach the goal in the time he had expected. It requires tremendous patience to calm the restless mind which is coloured by innumerable past experiences and saṁskāra (the accumulated residue of past thoughts and actions).

Merely listening to the words of the Guru does not enable the śiṣya to absorb the teaching. This is borne out by the story of Indra and Virochana. Indra, the king of Gods, and Virochana, a demon prince, went together to their spiritual preceptor Brahmā to obtain knowledge of the Supreme Self. Both stayed and listened to the same words of their Guru. Indra obtained enlightenment, whereas Virochana did not. Indra's memory was developed by his devotion to the subject taught by the love and faith which he had for his teacher. He had a feeling of oneness with his Guru. These were the reasons for his success. Virochana's memory was developed only through his intellect. He had no devotion either for the subject taught or for his preceptor. He remained what he originally was, an intellectual giant. He returned a doubter. Indra had intellectual humility, while Virochana had intellectual pride and imagined that it was condescending on his part to go to Brahmā. The approach of Indra was devotional while that of Virochana was practical. Virochana was motivated by curiosity and wanted the practical knowledge which he believed would be useful to him later to win power.

The śiṣya should above all treasure love, moderation and humility. Love begets courage, moderation creates abundance and humility generates power. Courage without love is brutish. Abundance without moderation leads to over-indulgence and decay. Power without humility breeds arrogance and tyranny. The true śiṣya learns from his Guru about a power which will never leave him as he returns to the Primeval One, the Source of His Being.

Sādhanā (*A Key to Freedom*)

All the important texts on Yoga lay great emphasis on sādhanā or abhyāsa (constant practice). Sādhanā is not just a theoretical study of Yoga

texts. It is a spiritual endeavour. Oil seeds must be pressed to yield oil. Wood must be heated to ignite and bring out the hidden fire within. In the same way, the sādhaka must by constant practice light the divine flame within himself.

> 'The young, the old, the extremely aged, even the sick and the infirm obtain perfection in Yoga by constant practice. Success will follow him who practises, not him who practises not. Success in Yoga is not obtained by the mere theoretical reading of sacred texts. Success is not obtained by wearing the dress of a yogi or a sanyāsi (a recluse), nor by talking about it. Constant practice alone is the secret of success. Verily, there is no doubt of this.'
>
> *Haṭha Yoga Pradīpikā*, chapter 1, verses 64–6.

> 'As by learning the alphabet one can, through practice, master all the sciences, so by thoroughly practising first physical training one acquires the knowledge of Truth (Tattva Jñāna), that is the real nature of the human soul as being identical with the Supreme Spirit pervading the Universe.'
>
> *Gheraṇḍa Saṁhitā*, chapter 1, verse 5.

It is by the co-ordinated and concentrated efforts of his body, senses, mind, reason and Self that a man obtains the prize of inner peace and fulfils the quest of his soul to meet his Maker. The supreme adventure in a man's life is his journey back to his Creator. To reach the goal he needs well developed and co-ordinated functioning of his body, senses, mind, reason and Self. If the effort is not co-ordinated, he fails in his adventure. In the third valli (chapter) of the first part of the *Kaṭhopaniṣad*, Yama (the God of Death) explains this Yoga to the seeker Nachiketā by way of the parable of the individual in a chariot.

> 'Know the Ātman (Self) as the Lord in a chariot, reason as the charioteer and mind as the reins. The senses, they say, are the horses, and their objects of desire are the pastures. The Self, when united with the senses and the mind, the wise call the Enjoyer (Bhoktṛ). The undiscriminating can never rein in his mind; his senses are like the vicious horses of a charioteer. The discriminating ever controls his mind; his senses are like disciplined horses. The undiscriminating becomes unmindful, ever impure; he does not reach the goal, wandering from one body to another. The discriminating becomes mindful, ever pure; he reaches the goal and is never reborn. The man who has a discriminating charioteer to rein in his mind reaches the end of the journey – the Supreme Abode of the everlasting Spirit.
>
> 'The senses are more powerful than the objects of desire. Greater than the senses is the mind, higher than the mind is the reason

and superior to reason is He – the Spirit in all. Discipline yourself by the Self and destroy your deceptive enemy in the shape of desire.'

Bhagavad Gītā, chapter III, verses 42–3.

To realize this not only constant practice is demanded but also renunciation. As regards renunciation, the question arises as to what one should renounce. The yogi does not renounce the world, for that would mean renouncing the Creator. The yogi renounces all that takes him away from the Lord. He renounces his own desires, knowing that all inspiration and right action come from the Lord. He renounces those who oppose the work of the Lord, those who spread demonic ideas and who merely talk of moral values but do not practise them.

The yogi does not renounce action. He cuts the bonds that tie himself to his actions by dedicating their fruits either to the Lord or to humanity. He believes that it is his privilege to do his duty and that he has no right to the fruits of his actions.

While others are asleep when duty calls and wake up only to claim their rights, the yogi is fully awake to his duty, but asleep over his rights. Hence it is said that in the night of all beings the disciplined and tranquil man wakes to the light.

Aṣṭāṅga Yoga – The Eight Limbs of Yoga

The *Yoga Sutra* of Patañjali is divided into four chapters or pāda. The first deals with samādhi, the second with the means (sādhanā) to achieve Yoga, the third enumerates the powers (vibhūti) that the yogi comes across in his quest, and the fourth deals with absolution (kaivalya).

Yama

The eight limbs of Yoga are described in the second chapter. The first of these is yama (ethical disciplines) – the great commandments transcending creed, country, age and time. They are: ahimsā (non-violence), satya (truth), asteya (non-stealing), brahmacharya (continence) and aparigraha (non-coveting). These commandments are the rules of morality for society and the individual, which if not obeyed bring chaos, violence, untruth, stealing, dissipation and covetousness. The roots of these evils are the emotions of greed, desire and attachment, which may be mild, medium or excessive. They only bring pain and ignorance. Patañjali strikes at the root of these evils by changing the direction of one's thinking along the five principles of yama.

1. *Ahimsā*. The word ahimsā is made up of the particle 'a' meaning 'not' and the noun himsā meaning killing or violence. It is more than a negative command not to kill, for it has a wider positive meaning, love. This love embraces all creation for we are all children of the same Father – the Lord. The yogi believes that to kill or to destroy a thing or being is to insult its Creator. Men either kill for food or to protect themselves from danger. But merely because a man is a vegetarian, it does not necessarily follow that he is non-violent by temperament or that he is a yogi, though a vegetarian diet is a necessity for the practice of yoga. Bloodthirsty tyrants may be vegetarians, but violence is a state of mind, not of diet. It resides in a man's mind and not in the instrument he holds in his hand. One can use a knife to pare fruit or to stab an enemy. The fault is not in the instrument, but in the user.

Men take to violence to protect their own interests – their own bodies, their loved ones, their property or dignity. But a man cannot rely upon himself alone to protect himself or others. The belief that he can do so is wrong. A man must rely upon God, who is the source of all strength. Then he will fear no evil.

Violence arises out of fear, weakness, ignorance or restlessness. To curb it what is most needed is freedom from fear. To gain this freedom, what is required is a change of outlook on life and reorientation of the mind. Violence is bound to decline when men learn to base their faith upon reality and investigation rather than upon ignorance and supposition.

The yogi believes that every creature has as much right to live as he has. He believes that he is born to help others and he looks upon creation with eyes of love. He knows that his life is linked inextricably with that of others and he rejoices if he can help them to be happy. He puts the happiness of others before his own and becomes a source of joy to all who meet him. As parents encourage a baby to walk the first steps, he encourages those more unfortunate than himself and makes them fit for survival.

For a wrong done by others, men demand justice; while for that done by themselves they plead mercy and forgiveness. The yogi on the other hand, believes that for a wrong done by himself, there should be justice, while for that done by another there should be forgiveness. He knows and teaches others how to live. Always striving to perfect himself, he shows them by his love and compassion how to improve themselves.

The yogi opposes the evil in the wrong-doer, but not the wrong-doer. He prescribes penance not punishment for a wrong done. Opposition to evil and love for the wrong-doer can live side by side. A drunkard's wife whilst loving him may still oppose his habit. Opposition without love leads to violence; loving the wrong-doer without opposing the evil in him is folly and leads to misery. The yogi knows that to love a person

whilst fighting the evil in him is the right course to follow. The battle is won because he fights it with love. A loving mother will sometimes beat her child to cure it of a bad habit; in the same way a true follower of ahimsā loves his opponent.

Along with ahimsā go abhaya (freedom from fear) and akrodha (freedom from anger). Freedom from fear comes only to those who lead a pure life. The yogi fears none and none need fear him, because he is purified by the study of the Self. Fear grips a man and paralyses him. He is afraid of the future, the unknown and the unseen. He is afraid that he may lose his means of livelihood, wealth or reputation. But the greatest fear is that of death. The yogi knows that he is different from his body, which is a temporary house for his spirit. He sees all beings in the Self and the Self in all beings and therefore he loses all fear. Though the body is subject to sickness, age, decay and death, the spirit remains unaffected. To the yogi death is the sauce that adds zest to life. He has dedicated his mind, his reason and his whole life to the Lord. When he has linked his entire being to the Lord, what shall he then fear?

There are two types of anger (krodha), one of which debases the mind while the other leads to spiritual growth. The root of the first is pride, which makes one angry when slighted. This prevents the mind from seeing things in perspective and makes one's judgement defective. The yogi, on the other hand, is angry with himself when his mind stoops low or when all his learning and experience fail to stop him from folly. He is stern with himself when he deals with his own faults, but gentle with the faults of others. Gentleness of mind is an attribute of a yogi, whose heart melts at all suffering. In him gentleness for others and firmness for himself go hand in hand, and in his presence all hostilities are given up.

2. *Satya.* Satya or truth is the highest rule of conduct or morality. Mahātma Gandhi said: 'Truth is God and God is Truth. As fire burns impurities and refines gold, so the fire of truth cleanses the yogi and burns up the dross in him.

If the mind thinks thoughts of truth, if the tongue speaks words of truth and if the whole life is based upon truth, then one becomes fit for union with the Infinite. Reality in its fundamental nature is love and truth and expresses itself through these two aspects. The yogi's life must conform strictly to these two facets of Reality. That is why ahimsā, which is essentially based on love, is enjoined. Satya presupposes perfect truthfulness in thought, word and deed. Untruthfulness in any form puts the sādhaka out of harmony with the fundamental law of truth.

Truth is not limited to speech alone. There are four sins of speech: abuse and obscenity, dealing in falsehoods, calumny or telling tales and

lastly ridiculing what others hold to be sacred. The tale bearer is more poisonous than a snake. The control of speech leads to the rooting out of malice. When the mind bears malice towards none, it is filled with charity towards all. He who has learnt to control his tongue has attained self-control in a great measure. When such a person speaks he will be heard with respect and attention. His words will be remembered, for they will be good and true.

When one who is established in truth prays with a pure heart, then things he really needs come to him when they are really needed: he does not have to run after them. The man firmly established in truth gets the fruit of his actions without apparently doing anything. God, the source of all truth, supplies his needs and looks after his welfare.

3. *Asteya.* The desire to possess and enjoy what another has, drives a person to do evil deeds. From this desire spring the urge to steal and the urge to covet. Asteya (a = not, steya = stealing), or non-stealing includes not only taking what belongs to another without permission, but also using something for a different purpose to that intended, or beyond the time permitted by its owner. It thus includes misappropriation, breach of trust, mismanagement and misuse. The yogi reduces his physical needs to the minimum, believing that if he gathers things he does not really need, he is a thief. While other men crave for wealth, power, fame or enjoyment, the yogi has one craving and that is to adore the Lord. Freedom from craving enables one to ward off great temptations. Craving muddies the stream of tranquillity. It makes men base and vile and cripples them. He who obeys the commandment *Thou shalt not steal,* becomes a trusted repository of all treasures.

4. *Brahmacharya.* According to the dictionary brahmacharya means the life of celibacy, religious study and self-restraint. It is thought that the loss of semen leads to death and its retention to life. By the preservation of semen the yogi's body develops a sweet smell. So long as it is retained, there is no fear of death. Hence the injunction that it should be preserved by concentrated effort of the mind. The concept of brahmacharya is not one of negation, forced austerity and prohibition. According to Śankarāchārya, a brahmachārī (one who observes brahmacharya) is a man who is engrossed in the study of the sacred Vedic lore, constantly moves in Brahman and knows that all exists in Brahman. In other words, one who sees divinity in all is a brahmachārī. Patañjali, however, lays stress on continence of the body, speech and mind. This does not mean that the philosophy of Yoga is meant only for celibates. Brahmacharya has little to do with whether one is a bachelor or married and living the life of a householder. One has to develop the higher aspects of Brahmacharya in one's daily living. It is not necessary for

one's salvation to stay unmarried and without a house. On the contrary, all the smṛtīs (codes of law) recommend marriage. Without experiencing human love and happiness, it is not possible to know divine love. Almost all the yogis and sages of old in India were married men with families of their own. They did not shirk their social or moral responsibilities. Marriage and parenthood are no bar to the knowledge of divine love, happiness and union with the Supreme Soul.

Dealing with the position of an aspirant who is a householder, the *Śiva Saṁhitā* says: Let him practise free from the company of men in a retired place. For the sake of appearances, he should remain in society, but not have his heart in it. He should not renounce the duties of his profession, caste or rank; but let him perform these as an instrument of the Lord, without any thought of the results. He succeeds by following wisely the method of Yoga; there is no doubt of it. Remaining in the midst of the family, always doing the duties of the householder, he who is free from merits and demerits and has restrained his senses, attains salvation. The householder practising Yoga is not touched by virtue or vice: if to protect mankind he commits any sin, he is not polluted by it. (Chapter V, verses 234–8.)

When one is established in brahmacharya, one develops a fund of vitality and energy, a courageous mind and a powerful intellect so that one can fight any type of injustice. The brahmachārī will use the forces he generates wisely: he will utilize the physical ones for doing the work of the Lord, the mental for the spread of culture and the intellectual for the growth of spiritual life. Brahmacharya is the battery that sparks the torch of wisdom.

5. *Aparigraha.* Parigraha means hoarding or collecting. To be free from hoarding is aparigraha. It is thus but another facet of asteya (non-stealing). Just as one should not take things one does not really need, so one should not hoard or collect things one does not require immediately. Neither should one take anything without working for it or as a favour from another, for this indicates poverty of spirit. The yogi feels that the collection or hoarding of things implies a lack of faith in God and in himself to provide for his future. He keeps faith by keeping before him the image of the moon. During the dark half of the month, the moon rises late when most men are asleep and so do not appreciate its beauty. Its splendour wanes but it does not stray from its path and is indifferent to man's lack of appreciation. It has faith that it will be full again when it faces the Sun and then men will eagerly await its glorious rising.

By the observance of aparigraha, the yogi makes his life as simple as possible and trains his mind not to feel the loss or the lack of anything. Then everything he really needs will come to him by itself at the proper time. The life of an ordinary man is filled with an unending series of

disturbances and frustrations and with his reactions to them. Thus there is hardly any possibility of keeping the mind in a state of equilibrium. The sādhaka has developed the capacity to remain satisfied with whatever happens to him. Thus he obtains the peace which takes him beyond the realms of illusion and misery with which our world is saturated. He recalls the promise given by Sri Krishna to Arjuna in the ninth chapter of the *Bhagavad Gītā*: 'To those who worship Me alone with single-minded devotion, who are in harmony with Me every moment, I bring full security. I shall supply all their wants and shall protect them for ever.'

Niyama

Niyama are the rules of conduct that apply to individual discipline, while yama are universal in their application. The five niyama listed by Patañjali are: saucha (purity), santoṣa (contentment), tapas (ardour or austerity), svādhyāya (study of the Self) and Iśvara praṇidhāna (dedication to the Lord).

1. *Śaucha*. Purity of blood is essential for well-being. While good habits like bathing purify the body externally, āsana and prāāyāma cleanse it internally. The practice of āsanas tones the entire body and removes the toxins and impurities caused by over-indulgence. Prāṇāyāma cleanses and aerates the lungs, oxygenates the blood and purifies the nerves. But more important than the physical cleansing of the body is the cleansing of the mind of its disturbing emotions like hatred, passion, anger, lust, greed, delusion and pride. Still more important is the cleansing of the intellect (buddhi) of impure thoughts. The impurities of the mind are washed off in the waters of bhakti (adoration). The impurities of the intellect or reason are burned off in the fire of svādhyāya (study of the Self). This internal cleansing gives radiance and joy. It brings benevolence (saumanasya) and banishes mental pain, dejection, sorrow and despair (daurmanasya). When one is benevolent, one sees the virtues in others and not merely their faults. The respect which one shows for another's virtues, makes him self-respecting as well and helps him to fight his own sorrows and difficulties. When the mind is lucid, it is easy to make it one-pointed (ekāgra). With concentration, one obtains mastery over the senses (indriyajaya). Then one is ready to enter the temple of his own body and see his real self in the mirror of his mind.

Besides purity of body, thought and word, pure food is also necessary. Apart from cleanliness in the preparation of food it is also necessary to observe purity in the means by which one procures it.

Food, the supporting yet consuming substance of all life, is regarded as a phase of Brahman. It should be eaten with the feeling that with

each morsel one can gain strength to serve the Lord. Then food becomes pure. Whether or not to be a vegetarian is a purely personal matter as each person is influenced by the tradition and habits of the country in which he was born and bred. But, in course of time, the practitioner of yoga has to adopt a vegetarian diet, in order to attain one-pointed attention and spiritual evolution.

Food should be taken to promote health, strength, energy and life. It should be simple, nourishing, juicy and soothing. Avoid foods which are sour, bitter, pungent, burning, stale, tasteless, heavy and unclean.

Character is moulded by the type of food we take and by how we eat it. Men are the only creatures that eat when not hungry and generally live to eat rather than eat to live. If we eat for flavours of the tongue, we over-eat and so suffer from digestive disorders which throw our systems out of gear. The yogi believes in harmony, so he eats for the sake of sustenance only. He does not eat too much or too little. He looks upon his body as the rest-house of his spirit and guards himself against over-indulgence.

Besides food, the place is also important for spiritual practices. It is difficult to practise in a distant country (away from home), in a forest, in a crowded city, or where it is noisy. One should choose a place where food is easily procurable, a place which is free from insects, protected from the elements and with pleasing surroundings. The banks of a lake or river or the sea-shore are ideal. Such quiet ideal places are hard to find in modern times; but one can at least make a corner in one's room available for practice and keep it clean, airy, dry and pest-free.

2. *Santoṣa.* Santoṣa or contentment has to be cultivated. A mind that is not content cannot concentrate. The yogi feels the lack of nothing and so he is naturally content. Contentment gives bliss unsurpassed to the yogi. A contented man is complete for he has known the love of the Lord and has done his duty. He is blessed for he has known truth and joy.

Contentment and tranquillity are states of mind. Differences arise among men because of race, creed, wealth and learning. Differences create discord and there arise conscious or unconscious conflicts which distract and perplex one. Then the mind cannot become one-pointed (ekāgra) and is robbed of its peace. There is contentment and tranquillity when the flame of the spirit does not waver in the wind of desire. The sādhaka does not seek the empty peace of the dead, but the peace of one whose reason is firmly established in God.

3. *Tapas.* Tapas is derived from the root 'tap' meaning to blaze, burn, shine, suffer pain or consume by heat. It therefore means a burning effort under all circumstances to achieve a definite goal in life. It involves

purification, self-discipline and austerity. The whole science of character building may be regarded as a practice of tapas.

Tapas is the conscious effort to achieve ultimate union with the Divine and to burn up all desires which stand in the way of this goal. A worthy aim makes life illumined, pure and divine. Without such an aim, action and prayer have no value. Life without tapas, is like a heart without love. Without tapas, the mind cannot reach up to the Lord.

Tapas is of three types. It may relate to the body (kāyika), to speech (vāchika) or to mind (mānasika). Continence (brahmacharya) and non-violence (ahimsā) are tapas of the body. Using words which do not offend, reciting the glory of God, speaking the truth without regard for the consequences to oneself and not speaking ill of others are tapas of speech. Developing a mental attitude whereby one remains tranquil and balanced in joy and sorrow and retains self-control are tapas of the mind.

It is tapas when one works without any selfish motive or hope of reward and with an unshakable faith that not even a blade of grass can move without His will.

By tapas the yogi develops strength in body, mind and character. He gains courage and wisdom, integrity, straightforwardness and simplicity.

4. *Svādhyāya.* Sva means self and adhyāya means study or education. Education is the drawing out of the best that is within a person. Svādhyāya, therefore, is the education of the self.

Svādhyāya is different from mere instruction like attending a lecture where the lecturer parades his own learning before the ignorance of his audience. When people meet for svādhyāya, the speaker and listener are of one mind and have mutual love and respect. There is no sermonizing and one heart speaks to another. The ennobling thoughts that arise from svādhyāya are, so to speak, taken into one's bloodstream so that they become a part of one's life and being.

The person practising svādhyāya reads his own book of life, at the same time that he writes and revises it. There is a change in his outlook on life. He starts to realize that all creation is meant for bhakti (adoration) rather than for bhoga (enjoyment), that all creation is divine, that there is divinity within himself and that the energy which moves him is the same that moves the entire universe.

According to Śri Vinobā Bhāve (the leader of the Bhoodan movement), svādhyāya is the study of one subject which is the basis or root of all other subjects or actions, upon which the others rest, but which itself does not rest upon anything.

To make life healthy, happy and peaceful, it is essential to study regularly divine literature in a pure place. This study of the sacred books of the world will enable the sādhaka to concentrate upon and solve the

difficult problems of life when they arise. It will put an end to ignorance and bring knowledge. Ignorance has no beginning, but it has an end. There is a beginning but no end to knowledge. By svādhyāya the sādhaka understands the nature of his soul and gains communion with the divine. The sacred books of the world are for all to read. They are not meant for the members of one particular faith alone. As bees savour the nectar in various flowers, so the sādhaka absorbs things in other faiths which will enable him to appreciate his own faith better.

Philology is not a language but the science of languages, the study of which will enable the student to learn his own language better. Similarly, Yoga is not a religion by itself. It is the science of religions, the study of which will enable a sādhaka the better to appreciate his own faith.

5. *Īśvara praṇidhāna.* Dedication to the Lord of one's actions and will is Īśvara praṇidhāna. He who has faith in God does not despair. He has illumination (tejas). He who knows that all creation belongs to the Lord will not be puffed up with pride or drunk with power. He will not stoop for selfish purposes; his head will bow only in worship. When the waters of bhakti (adoration) are made to flow through the turbines of the mind, the result is mental power and spiritual illumination. While mere physical strength without bhakti is lethal, mere adoration without strength of character is like an opiate. Addiction to pleasures destroys both power and glory. From the gratification of the senses as they run after pleasures arise moha (attachment) and lobha (greed) for their repetition. If the senses are not gratified, then, there is śoka (sorrow). They have to be curbed with knowledge and forbearance; but to control the mind is more difficult. After one has exhausted one's own resources and still not succeeded, one turns to the Lord for help for He is the source of all power. It is at this stage that bhakti begins. In bhakti, the mind, the intellect and the will are surrendered to the Lord and the sādhaka prays: 'I do not know what is good for me. Thy will be done.' Others pray to have their own desires gratified or accomplished. In bhakti or true love there is no place for 'I' and 'mine'. When the feeling of 'I' and 'mine' disappears, the individual soul has reached full growth.

When the mind has been emptied of desires of personal gratification, it should be filled with thoughts of the Lord. In a mind filled with thoughts of personal gratification, there is danger of the senses dragging the mind after the objects of desire. Attempts to practise bhakti without emptying the mind of desires is like building a fire with wet fuel. It makes a lot of smoke and brings tears to the eyes of the person who builds it and of those around him. A mind with desires does not ignite and glow, nor does it generate light and warmth when touched with the fire of knowledge.

The name of the Lord is like the Sun, dispelling all darkness. The

moon is full when it faces the sun. The individual soul experiences fullness (pūrṇatā) when it faces the Lord. If the shadow of the earth comes between the full moon and the sun there is an eclipse. If the feeling of 'I' and 'mine' casts its shadow upon the experience of fullness, all efforts of the sādhaka to gain peace are futile.

Actions mirror a man's personality better than his words. The yogi has learnt the art of dedicating all his actions to the Lord and so they reflect the divinity within him.

Āsana

The third limb of yoga is āsana or posture. Āsana brings steadiness, health and lightness of limb. A steady and pleasant posture produces mental equilibrium and prevents fickleness of mind. Āsanas are not merely gymnastic exercises; they are postures. To perform them one needs a clean airy place, a blanket and determination, while for other systems of physical training one needs large playing fields and costly equipment. Āsanas can be done alone, as the limbs of the body provide the necessary weights and counter-weights. By practising them one develops agility, balance, endurance and great vitality.

Āsanas have been evolved over the centuries so as to exercise every muscle, nerve and gland in the body. They secure a fine physique, which is strong and elastic without being muscle-bound and they keep the body free from disease. They reduce fatigue and soothe the nerves. But their real importance lies in the way they train and discipline the mind.

Many actors, acrobats, athletes, dancers, musicians and sportsmen also possess superb physiques and have great control over the body, but they lack control over the mind, the intellect and the Self. Hence they are in disharmony with themselves and one rarely comes across a balanced personality among them. They often put the body above all else. Though the yogi does not underrate his body, he does not think merely of its perfection but of his senses, mind, intellect and soul.

The yogi conquers the body by the practice of āsanas and makes it a fit vehicle for the spirit. He knows that it is a necessary vehicle for the spirit. A soul without a body is like a bird deprived of its power to fly.

The yogi does not fear death, for time must take its toll of all flesh. He knows that the body is constantly changing and is affected by childhood, youth and old age. Birth and death are natural phenomena but the soul is not subject to birth and death. As a man casting off worn-out garments takes on new ones, so the dweller within the body casting aside worn-out bodies enters into others that are new.

The yogi believes that his body has been given to him by the Lord

not for enjoyment alone, but also for the service of his fellow men during every wakeful moment of his life. He does not consider it his property. He knows that the Lord who has given him his body will one day take it away.

By performing āsanas, the sādhaka first gains health, which is not mere existence. It is not a commodity which can be purchased with money. It is an asset to be gained by sheer hard work. It is a state of complete equilibrium of body, mind and spirit. Forgetfulness of physical and mental consciousness is health. The yogi frees himself from physical disabilities and mental distractions by practising āsanas. He surrenders his actions and their fruits to the Lord in the service of the world.

The yogi realizes that his life and all its activities are part of the divine action in nature, manifesting and operating in the form of man. In the beating of his pulse and the rhythm of his respiration, he recognizes the flow of the seasons and the throbbing of universal life. His body is a temple which houses the Divine Spark. He feels that to neglect or to deny the needs of the body and to think of it as something not divine, is to neglect and deny the universal life of which it is a part. The needs of the body are the needs of the divine spirit which lives through the body. The yogi does not look heaven-ward to find God for he knows that He is within, being known as the Antarātmā (the Inner Self). He feels the kingdom of God within and without and finds that heaven lies in himself.

Where does the body end and the mind begin? Where does the mind end and the spirit begin? They cannot be divided as they are inter-related and but different aspects of the same all-pervading divine consciousness.

The yogi never neglects or mortifies the body or the mind, but cherishes both. To him the body is not an impediment to his spiritual liberation nor is it the cause of its fall, but is an instrument of attainment. He seeks a body strong as a thunderbolt, healthy and free from suffering so as to dedicate it in the service of the Lord for which it is intended. As pointed out in the *Muṇḍakopaniṣad* the Self cannot be attained by one without strength, nor through heedlessness, nor without an aim. Just as an unbaked earthen pot dissolves in water the body soon decays. So bake it hard in the fire of yogic discipline in order to strength and purify it.

The names of the āsanas are significant and illustrate the principle of evolution. Some are named after vegetation like the tree (vṛkṣa) and the lotus (padma); some after insects like the locust (śalabha) and the scorpion (vṛśchika); some after aquatic animals and amphibians like the fish (matsya), the tortoise (kūrma), the frog (bheka or maṇḍūka) or the crocodile (nakra). There are āsanas called after birds like the cock (kukkuṭa), the heron (baka), the peacock (mayūra) and the swan

(haṁsa). They are also named after quadrupeds like the dog (śvāna), the horse (vātāyana), the camel (uṣṭra) and the lion (siṁha). Creatures that crawl like the serpent (bhujaṅga) are not forgotten, nor is the human embryonic state (garbha-pinda) overlooked. Āsanas are named after legendary heroes like Vīrabhadra and Hanumān, son of the Wind. Sages like Bharadvāja, Kapila, Vasiṣṭha and Viśvāmitra are remembered by having āsanas named after them. Some āsanas are also called after gods of the Hindu pantheon and some recall the Avatārās, or incarnations of Divine Power. Whilst performing āsanas the yogi's body assumes many forms resembling a variety of creatures. His mind is trained not to despise any creature, for he knows that throughout the whole gamut of creation, from the lowliest insect to the most perfect sage, there breathes the same Universal Spirit, which assumes innumerable forms. He knows that the highest form is that of the Formless. He finds unity in universality. True āsana is that in which the thought of Brahman flows effortlessly and incessantly through the mind of the sādhaka.

Dualities like gain and loss, victory and defeat, fame and shame, body and mind, mind and soul vanish through mastery of the āsanas, and the sādhaka then passes on to prāṇāyāma, the fourth stage in the path of yoga. In prāṇayāma practices the nostrils, nasal passages and membranes, the windpipe, the lungs and the diaphragm are the only parts of the body which are actively involved. These alone feel the full impact of the force of prāṇa, the breath of life. Therefore, do not seek to master prāṇāyāma in a hurry, as you are playing with life itself. By its improper practice respiratory diseases will arise and the nervous system will be shattered. By its proper practice one is freed from most diseases. Never attempt to practice prāṇāyāma alone by yourself. For it is essential to have the personal supervision of a Guru who knows the physical limitations of his pupil.

Prāṇāyāma

Just as the word yoga is one of wide import, so also is prāṇa. Prāṇa means breath, respiration, life, vitality, wind, energy or strength. It also connotes the soul as opposed to the body. The word is generally used in the plural to indicate vital breaths. Āyāma means length, expansion, stretching or restraint. Prāṇāyāma thus connotes extension of breath and its control. This control is over all the functions of breathing, namely, (1) inhalation or inspiration, which is termed pūraka (filling up); (2) exhalation or expiration, which is called rechaka (emptying the lungs), and (3) retention or holding the breath, a state where there is no inhalation or exhalation, which is termed kumbhaka. In Haṭha Yoga texts kumbhaka is also used in a loose generic sense to include all the three respiratory processes of inhalation, exhalation and retention.

A kumbha is a pitcher, water pot, jar or chalice. A water pot may be emptied of all air and filled completely with water, or it may be emptied of all water and filled completely with air. Similarly, there are two states of kumbhaka namely (1) when breathing is suspended after full inhalation (the lungs being completely filled with life-giving air), and (2) when breathing is suspended after full exhalation (the lungs being emptied of all noxious air). The first of these states, where breath is held after a full inhalation, but before exhalation begins, is known as antara kumbhaka. The second, where breath is held after a full exhalation, but before inhalation begins is known as bāhya kumbhaka. Antara means inner or interior, while bāhya means outer or exterior. Thus, kumbhaka is the interval or intermediate time between full inhalation and exhalation (antara kumbhaka) or between full exhalation and inhalation (bāhya kumbhaka). In both these types breathing is suspended and restrained.

Prāṇāyāma is thus the science of breath. It is the hub round which the wheel of life revolves. 'As lions, elephants and tigers are tamed very slowly and cautiously, so should prāṇa be brought under control very slowly in gradation measured according to one's capacity and physical limitations. Otherwise it will kill the practitioner,' warns the *Haṭha Yoga Pradīpikā* (chapter II, verse 16).

The yogi's life is not measured by the number of his days but by the number of his breaths. Therefore, he follows the proper rhythmic patterns of slow deep breathing. These rhythmic patterns strengthen the respiratory system, soothe the nervous system and reduce craving. As desires and cravings diminish, the mind is set free and becomes a fit vehicle for concentration. By improper practice of prāṇāyāma the pupil introduces several disorders into his system like hiccough, wind, asthma, cough, catarrh, pains in the head, eyes and ears and nervous irritation. It takes a long time to learn slow, deep, steady and proper inhalations and exhalations. Master this before attempting kumbhaka.

As a fire blazes brightly when the covering of ash over it is scattered by the wind, the divine fire within the body shines in all its majesty when the ashes of desire are scattered by the practice of prāṇāyāma.

'The emptying the mind of the whole of its illusion is the true rechaka (exhalation). The realization that "I am Ātmā (spirit)" is the true pūraka (inhalation). And the steady sustenance of the mind on this conviction is the true kumbhaka (retention). This is true prāṇāyāma,' says Śankarāchārya.

Every living creature unconsciously breathes the prayer 'So'ham' (Saḥ = He: Aham = Aham = I–He, the Immortal Spirit, am I) with each inward breath. So also with each outgoing breath each creature prays 'Haṁsaḥ' (I am He). This ajapa-mantra (unconscious repetitive prayer) goes on for ever within each living creature throughout life. The

yogi fully realizes the significance of this ajapa-mantra and so is released from all the fetters that bind his soul. He offers up the very breath of his being to the Lord as sacrifice and receives the breath of life from the Lord as his blessing.

Prāṇa in the body of the individual (jīvātmā) is part of the cosmic breath of the Universal Spirit (Paramātmā). An attempt is made to harmonize the individual breath (piṇḍaprāṇa) with the cosmic breath (Brahmāṇḍa-prāṇa) through the practice of prāṇāyāma.

It has been said by Kariba Ekken, a seventeenth-century mystic: 'If you would foster a calm spirit, first regulate your breathing; for when that is under control, the heart will be at peace; but when breathing is spasmodic, then it will be troubled. Therefore, before attempting anything, first regulate your breathing on which your temper will be softened, your spirit calmed.'

The chitta (mind, reason and ego) is like a chariot yoked to a team of powerful horses. One of them is prāṇa (breath), the other is vāsanā (desire). The chariot moves in the direction of the more powerful animal. If breath prevails, the desires are controlled, the senses are held in check and the mind is stilled. If desire prevails, breath is in disarray and the mind is agitated and troubled. Therefore, the yogi masters the science of breath and by the regulation and control of breath, he controls the mind and stills its constant movement. In the practice of prāṇāyāma the eyes are kept shut to prevent the mind from wandering. 'When the prāṇa and the manas (mind) have been absorbed, an undefinable joy ensues.' (*Haṭha Yoga Pradīpikā*, chapter IV, verse 30.)

Emotional excitement affects the rate of breathing; equally, deliberate regulation of breathing checks emotional excitement. As the very object of Yoga is to control and still the mind, the yogi first learns prāṇāyāma to master the breath. This will enable him to control the senses and so reach the stage of pratyāhāra. Only then will the mind be ready for concentration (dhyāna).

The mind is said to be twofold – pure and impure. It is pure when it is completely free from desires and impure when it is in union with desires. By making the mind motionless and freeing it from sloth and distractions, one reaches the state of mindlessness (amanaska), which is the supreme state of samādhi. This state of mindlessness is not lunacy or idiocy but the conscious state of the mind when it is free from thoughts and desires. There is a vital difference between an idiot or a lunatic on the one hand, and a yogi striving to achieve a state of mindlessness on the other. The former is careless; the latter attempts to be carefree. It is the oneness of the breath and mind and so also of the senses and the abandonment of all conditions of existence and thought that is designated Yoga.

Prāṇa Vāyu. One of the most subtle forms of energy is air. This vital energy which also pervades the human body is classified in five main categories in the Haṭha Yoga texts according to the various functions performed by the energy. These are termed vāyu (wind) and the five main divisions are: prāṇa (here the generic term is used to designate the particular), which moves in the region of the heart and controls respiration; apāna, which moves in the sphere of the lower abdomen and controls the function of eliminating urine and faeces; samāna, which stokes the gastric fires to aid digestion; udāna, which dwells in the thoracic cavity and controls the intake of air and food; and vyāna, which pervades the entire body and distributes the energy derived from food and breath. There are also five subsidiary vāyūs. These are: nāga, which relieves abdominal pressure by belching; kūrma, which controls the movements of the eyelids to prevent foreign matter or too bright a light entering the eyes; kṛkara, which prevents substances passing up the nasal passages and down the throat by making one sneeze or cough; devadatta, which provides for the intake of extra oxygen in a tired body by causing a yawn, and lastly dhanaṁjaya, which remains in the body even after death and sometimes bloats up a corpse.

Pratyāhāra

If a man's reason succumbs to the pull of his senses he is lost. On the other hand, if there is rhythmic control of breath, the senses instead of running after external objects of desire turn inwards, and man is set free from their tyranny. This is the fifth stage of Yoga, namely, pratyāhāra, where the senses are brought under control.

When this stage is reached, the sādhaka goes through a searching self-examination. To overcome the deadly but attractive spell of sensual objects, he needs the insulation of adoration (bhakti) by recalling to his mind the Creator who made the objects of his desire. He also needs the lamp of knowledge of his divine heritage. The mind, in truth, is for mankind the cause of bondage and liberation; it brings bondage if it is bound to the objects of desire and liberation when it is free from objects. There is bondage when the mind craves, grieves or is unhappy over something. The mind becomes pure when all desires and fears are annihilated. Both the good and the pleasant present themselves to men and prompt them to action. The yogi prefers the good to the pleasant. Others driven by their desires, prefer the pleasant to the good and miss the very purpose of life. The yogi feels joy in what he is. He knows how to stop and, therefore, lives in peace. At first he prefers that which is bitter as poison, but he perseveres in his practice knowing well that in the end it will become as sweet as nectar. Others hankering for the

union of their senses with the objects of their desires, prefer that which at first seems sweet as nectar, but do not know that in the end it will be as bitter as poison.

The yogi knows that the path towards satisfaction of the senses by sensual desires is broad, but that it leads to destruction and that there are many who follow it. The path of Yoga is like the sharp edge of a razor, narrow and difficult to tread, and there are few who find it. The yogi knows that the paths of ruin or of salvation lie within himself.

According to Hindu philosophy, consciousness manifests in three different qualities. For man, his life and his consciousness, together with the entire cosmos are the emanations of one and the same prakṛti (cosmic matter or substance) – emanations that differ in designation through the predominance of one of the guṇās. The guṇās (qualities or attributes) are:

1. Sattva (the illuminating, pure or good quality), which leads to clarity and mental serenity.
2. Rajas (the quality of mobility or activity), which makes a person active and energetic, tense and wilful, and
3. Tamas (the dark and restraining quality), which obstructs and counteracts the tendency of rajas to work and of sattva to reveal.

Tamas is a quality of delusion, obscurity, inertia and ignorance. A person in whom it predominates is inert and plunged in a state of torpor. The quality of sattva leads towards the divine and tamas towards the demonic, while in between these two stands rajas.

The faith held, the food consumed, the sacrifices performed, the austerities undergone and the gifts given by each individual vary in accordance with his predominating guṇa.

He that is born with tendencies towards the divine is fearless and pure. He is generous and self-controlled. He pursues the study of the Self. He is non-violent, truthful and free from anger. He renounces the fruits of his labour, working only for the sake of work. He has a tranquil mind, with malice towards none and charity towards all, for he is free from craving. He is gentle, modest and steady. He is illumined, clement and resolute, being free from perfidy and pride.

A man in whom rajō-guṇa predominates has inner thirst. As he is passionate and covetous, he hurts others. Being full of lust and hatred, envy and deceit, his desires are insatiable. He is unsteady, fickle and easily distracted as well as ambitious and acquisitive. He seeks the patronage of friends and has family pride. He shrinks from unpleasant things and clings to pleasant ones. His speech is sour and his stomach greedy.

He that is born with demonic tendencies is deceitful, insolent and conceited. He is full of wrath, cruelty and ignorance. In such people

there is neither purity, nor right conduct, nor truth. They gratify their passions. Bewildered by numerous desires, caught in the web of delusion, these addicts of sensual pleasures fall into hell.

The working of the mind of persons with different predominating guṇās may be illustrated by their different ways of approach towards a universal commandment like 'Thou shalt not covet.' A man in whom tamō-guṇa predominates might interpret it thus: 'others should not covet what is mine, no matter how I obtained it. If they do, I shall destroy them.' The rajō-guṇa type is a calculating self-interested person who would construe the commandment as meaning: 'I will not covet others' goods lest they covet mine.' He will follow the letter of the law as a matter of policy, but not the true spirit of the law as a matter of principle. A person of sattvika temperament will follow both the letter and the spirit of the precept as a matter of principle and not of policy, as a matter of eternal value. He will be righteous for the sake of righteousness alone, and not because there is a human law imposing punishment to keep him honest.

The yogi who is also human is affected by these three guṇās. By his constant and disciplined study (abhyāsa) of himself and of the objects which his senses tend to pursue, he learns which thoughts, words and actions are prompted by tamas and which by rajas. With unceasing effort he weeds out and eradicates such thoughts and he works to achieve a sattvika frame of mind. When the sattva-guṇa alone remains, the human soul has advanced a long way towards the ultimate goal.

Like unto the pull of gravity is the pull of the guṇās. As intensive research and rigorous discipline are needed to experience the wonder of weightlessness in space, so also a searching self-examination and the discipline furnished by Yoga is needed by a sādhaka to experience union with the Creator of space when he is freed from the pull of the guṇās.

Once the sādhaka has experienced the fullness of creation or of the Creator, his thirst (tṛṣṇā) for objects of sense vanishes and he looks at them ever after with dispassion (vairāgya). He experiences no disquiet in heat or cold, in pain or pleasure, in honour or dishonour and in virtue or vice. He treats the two imposters – triumph and disaster – with equanimity. He has emancipated himself from these pairs of opposites. He has passed beyond the pull of the guṇās and has become a guṇātīta (one who has transcended the guṇās). He is then free from birth and death, from pain and sorrow and becomes immortal. He has no self-identity as he lives experiencing the fullness of the Universal Soul. Such a man, scorning nothing, leads all things to the path of perfection.

Dhāranā

When the body has been tempered by āsanas, when the mind has been refined by the fire of prāṇāyāma and when the senses have been brought under control by pratyāhāra, the sādhaka reaches the sixth stage called dhāraṇā. Here he is concentrated wholly on a single point or on a task in which he is completely engrossed. The mind has to be stilled in order to achieve this state of complete absorption.

The mind is an instrument which classifies, judges and co-ordinates the impressions from the outside world and those that arise within oneself.

Mind is the product of thoughts which are difficult to restrain for they are subtle and fickle. A thought which is well guarded by a controlled mind brings happiness. To get the best out of an instrument, one must know how it works. The mind is the instrument for thinking and it is therefore necessary to consider how it functions. Mental states are classified in five groups. The first of these is the kṣipta state, where the mental forces are scattered, being in disarray and in a state of neglect. Here the mind hankers after objects, the ragō-guṇa being dominant. The second is the vikṣipta state, where the mind is agitated and distracted. Here there is a capacity to enjoy the fruits of one's efforts, but the desires are not marshalled and controlled. Then in the mūḍha state the mind is foolish, dull and stupid. It is confounded and at a loss to know what it wants and here the tamō-guṇa predominates. The fourth state of the mind is the ekāgra (eka = one; agra = foremost) state, where the mind is closely attentive and the mental faculties are concentrated on a single object or focussed on one point only, with the sattva-guṇa prevailing. The ekāgra person has superior intellectual powers and knows exactly what he wants, so he uses all his powers to achieve his purpose. At times the ruthless pursuit of the desired object, irrespective of the cost to others, can create great misery, and it often happens that even if the desired object is achieved it leaves behind a bitter taste.

Arjuna, the mighty bowman of the epic Mahābhārata, provides us with an example of what is meant by dhāraṇā. Once Drona, the preceptor of the royal princes, organized an archery contest to test their proficiency. They were called upon one by one to describe the target, which was pointed out to them. It was a nesting bird. Some princes described the grove of trees, others the particular tree or the bough on which the nest stood. When Arjuna's turn came, he described first the bird. Then he saw only its head, and lastly he could see nothing but the shining eye of the bird, which was the centre of the target chosen by Droṇa.

There is danger, however, of an ekāgra person becoming supremely egotistical. Where the senses start roaming unchecked, the mind follows

suit. They cloud a man's judgement and set him adrift like a battered ship on a storm-tossed sea. A ship needs ballast to keep her on an even keel and the helmsman needs a star to steer her by. The ekāgra person needs bhakti (adoration of the Lord) and concentration on divinity to keep his mental equilibrium so that he goes on always in the right direction. He will not know happiness until the sense of 'I' and 'mine' disappears.

The last mental state is that of niruddha, where the mind (manas), intellect (buddhi) and ego (ahaṁkāra) are all restrained and all these faculties are offered to the Lord for His use and in His service. Here there is no feeling of 'I' and 'mine'. As a lens becomes more luminous when great light is thrown upon it and seems to be all light and undistinguishable from it, so also the sādhaka who has given up his mind, intellect and ego to the Lord, becomes one with Him, for the sādhaka thinks of nothing but Him, who is the creator of thought.

Without ekāgratā or concentration one can master nothing. Without concentration on Divinity, which shapes and controls the universe, one cannot unlock the divinity within oneself or become a universal man. To achieve this concentration, what is recommended is eka-tattva-abhyāsa or study of the single element that pervades all, the inmost Self of all beings, who converts His one form into many. The sādhaka, therefore, concentrates upon AUM, which is His symbol, to achieve ekāgratā.

Aum: According to Śri Vinobā Bhāve, the Latin word Omne and the Sanskrit word Aum are both derived from the same root meaning all and both words convey the concepts of omniscience, omnipresence and omnipotence. Another word for Aum is praṇava, which is derived from the root nu meaning to praise, to which is added the prefix pra denoting superiority. The word, therefore, means the best praise or the best prayer.

The symbol AUM is composed of three syllables, namely the letters A, U, M, and when written has a crescent and dot on its top. A few instances of the various interpretations given to it may be mentioned here to convey its meaning.

The letter A symbolizes the conscious or waking state (jāgrata-avasthā), the letter U the dream state (svapna-avasthā) and the letter M the dreamless sleep state (suṣupta-avasthā) of the mind and spirit. The entire symbol, together with the crescent and the dot, stands for the fourth state (turīya-avasthā), which combines all these states and transcends them. This is the state of samādhi.

The letters A, U and M symbolize respectively speech (vak), the mind (manas) and the breath of life (prāṇa), while the entire symbol stands for the living spirit, which is but a portion of the divine spirit.

The three letters also represent the dimensions of length, breadth and depth, while the entire symbol represents Divinity, which is beyond the limitations of shape and form.

The three letters A, U and M symbolize the absence of desire, fear and anger, while the whole symbol stands for the perfect man (a sthita-prajñā), one whose wisdom is firmly established in the divine.

They represent the three genders, masculine, feminine and neuter, while the entire symbol represents all creation together with the Creator.

They stand for the three guṇās or qualities of sattva, rajas and tamas, while the whole symbol represents a guṇātīta, one who has transcended and gone beyond the pull of the guṇās.

The letters correspond to the three tenses – past, present and future – while the entire symbol stands for the Creator, who transcends the limitations of time.

They also stand for the teaching imparted by the mother, the father and the Guru respectively. The entire symbol represents Brahma Vidyā, the knowledge of the Self, the teaching which is imperishable.

The A, U and M depict the three stages of yogic discipline, namely, āsana, prāṇāyāma and pratyāhāra. The entire symbol represents samādhi, the goal for which the three stages are the steps.

They represent the triad of Divinity, namely, Brahmā – the creator, Viṣṇu – the Maintainer, and Śiva – the Destroyer of the universe. The whole symbol is said to represent Brahman from which the universate emanates, has its growth and fruition and into which it merges in the end. It does not grow or change. Many change and pass, but Brahman is the One that ever remains unchanged.

The letters A, U and M also stand for the mantra 'Tat Twam Asi' ('That Thou Art'), the realization of man's divinity within himself. The entire symbol stands for this realization, which liberates the human spirit from the confines of his body, mind, intellect and ego.

After realizing the importance of AUM, the yogi focusses his attention on his beloved Deity adding AUM to the name of the Lord. The word AUM being too vast and too abstract, he unifies his senses, will, intellect, mind and reason by focussing on the name of the Lord and adding the word AUM with one pointed devotion and so experiences the feeling and meaning of the mantra.

The yogi recalls the verses of the *Muṇḍakopaniṣad*: 'Taking as a bow the great weapon of the Upaniṣad, one should put upon it an arrow sharpened by meditation. Stretching it with a thought directed to the essence of That, penetrate the Imperishable as the mark, my friend. The mystic syllable AUM is the bow. The arrow is the Self (Ātmā). Brahman is the target. By the undistracted man is It penetrated. One should come to be in It, as the arrow in the mark.'

Dhyāna

As water takes the shape of its container, the mind when it contemplates an object is transformed into the shape of that object. The mind which thinks of the all-pervading divinity which it worships, is ultimately through long-continued devotion transformed into the likeness of that divinity.

When oil is poured from one vessel to another, one can observe the steady constant flow. When the flow of concentration is uninterrupted, the state that arises is dhyāna (meditation). As the filament in an electric bulb glows and illumines when there is a regular uninterrupted current of electricity, the yogi's mind will be illumined by dhyāna. His body, breath, senses, mind, reason and ego are all integrated in the object of his contemplation – the Universal Spirit. He remains in a state of consciousness which has no qualification whatsoever. There is no other feeling except a state of SUPREME BLISS. Like a streak of lightning the yogi sees LIGHT that shines beyond the earth and the heavens. He sees the light that shines in his own heart. He becomes a light unto himself and others.

The signs of progress on the path of Yoga are health, a sense of physical lightness, steadiness, clearness of countenance and a beautiful voice, sweetness of odour of the body and freedom from craving. He has a balanced, serene and a tranquil mind. He is the very symbol of humility. He dedicates all his actions to the Lord and taking refuge in Him, frees himself from the bondage of karma (action) and becomes a Jīvana Mukta (a Liberated Soul).

'What becomes of him who strives and fails to reach the end of Yoga, who has faith, but whose mind wanders away from Yoga?' To this query of Arjuna, the Lord Śri Krishna replied:

> 'No evil can befall a righteous man. He dwells long years in the heaven of those who did good, and then he is reborn in the house of the pure and the great. He may even be born in a family of illumined yogis; but to be born in such a family is most difficult in this world. He will regain the wisdom attained in his former life and strives ever for perfection. Because of his former study, practice and struggle which drive him ever onwards, the yogi ever strives with a soul cleansed of sin, attains perfection through many lives and reaches the supreme goal. The yogi goes beyond those who only follow the path of austerity, knowledge or service. Therefore, Arjuna, be thou a yogi. The greatest of all yogis is he who adores Me with faith and whose heart abides in Me.'
>
> *Bhagavad Gītā*, chapter VI, verses 38 to 47.

Samādhi

Samādhi is the end of the sādhaka's quest. At the peak of his meditation, he passes into the state of samādhi, where his body and senses are at rest as if he is asleep, his faculties of mind and reason are alert as if he is awake, yet he has gone beyond consciousness. The person in a state of samādhi is fully conscious and alert.

All creation is Brahman. The sādhaka is tranquil and worships it as that from which he came forth, as that in which he breathes, as that into which he will be dissolved. The soul within the heart is smaller than the smallest seed, yet greater than the sky, containing all works, all desires. Into this the sādhaka enters. Then there remains no sense of 'I' or 'mine' as the working of the body, the mind and the intellect have stopped as if one is in deep sleep. The sādhaka has attained true Yoga; there is only the experience of consciousness, truth and unutterable joy. There is a peace that passeth all understanding. The mind cannot find words to describe the state and the tongue fails to utter them. Comparing the experience of samādhi with other experiences, the sages say: 'Neti! Neti!' – 'It is not this! It is not this!' The state can only be expressed by profound silence. The yogi has departed from the material world and is merged in the Eternal. There is then no duality between the knower and the known for they are merged like camphor and the flame.

There wells up from within the heart of the yogi the Song of the Soul, sung by Śankarāchārya in his *Ātma Ṣaṭkam*.

PART II

Yogāsanas

HINTS AND CAUTIONS

THE REQUISITES

1. Without firm foundations a house cannot stand. Without the practice of the principles of yama and niyama, which lay down firm foundations for building character, there cannot be an integrated personality. Practice of āsanas without the backing of yama and niyama is mere acrobatics.

2. The qualities demanded from an aspirant are discipline, faith, tenacity, and perseverance to practice regularly without interruptions.

CLEANLINESS AND FOOD

3. Before starting to practise āsanas, the bladder should be emptied and the bowels evacuated. Topsy-turvy poses help bowel movements. If the student is constipated or it is not possible to evacuate the bowels before the practice of āsanas, start with Śīrṣāsana and Sarvāngāsana and their variations. Attempt other āsanas only after evacuation. Never practice advanced āsanas without having first evacuated the bowels.

BATH

4. Āsanas come easier after taking a bath. After doing them, the body feels sticky due to perspiration and it is desirable to bathe some fifteen minutes later. Taking a bath or a shower both before and after practising āsanas refreshes the body and mind.

FOOD

5. Āsanas should preferably be done on an empty stomach. If this is difficult, a cup of tea or coffee, cocoa or milk may be taken before doing them. They may be practised without discomfort one hour after a very light meal. Allow at least four hours to elapse after a heavy meal before starting the practice. Food may be taken half an hour after completing the āsanas.

TIME

6. The best time to practise is either early in the morning or late in the evening. In the morning āsanas do not come easily as the body is stiff.

The mind, however, is still fresh but its alertness and determination diminish as time goes by. The stiffness of the body is conquered by regular practice and one is able to do the āsanas as well. In the evening, the body moves more freely than in the mornings, and the āsanas come better and with greater ease. Practice in the morning makes one work better in one's vocation. In the evening it removes the fatigue of the day's strain and makes one fresh and calm. Do all the āsanas in the morning and stimulative āsanas (like Śīrṣāsana, Sarvāngāsana and their variations and Paśchimottānāsana) should be practised in the evening.

SUN

7. Do not practise āsanas after being out in the hot sun for several hours.

PLACE

8. They should be done in a clean airy place, free from insects and noise.

9. Do not do them on the bare floor or on an uneven place, but on a folded blanket laid on a level floor.

CAUTION

10. No undue strain should be felt in the facial muscles, ears and eyes or in breathing during the practice.

CLOSING OF THE EYES

11. In the beginning, keep the eyes open. Then you will know what you are doing and where you go wrong. If you shut your eyes you will not be able to watch the requisite movements of the body or even the direction in which you are doing the pose. You can keep your eyes closed only when you are perfect in a particular āsana for only then will you be able to adjust the bodily movements and feel the correct stretches.

MIRROR

12. If you are doing the āsanas in front of a mirror, keep it perpendicular to the floor and let it come down to ground level, for otherwise the poses will look slanting due to the angle of the mirror. You will not be able to observe the movements of placing the head and shoulders in the topsy-turvy poses unless the mirror reaches down to the floor. Use a mirror without a frame.

THE BRAIN

13. During the practice of āsanas, it is the body alone which should be active while the brain should remain passive, watchful and alert. If they are done with the brain, then you will not be able to see your own mistakes.

BREATHING

14. In all the āsanas, breathing should be done through the nostrils only and not through the mouth.

15. Do not restrain the breath while in the process of the āsana or while staying in it. Follow the instructions regarding breathing given in the technique sections of the various āsanas as described hereafter.

ŚAVĀSANA

16. After completing the practice of āsanas always lie down in Śavāsana for at least 10 to 15 minutes, as this will remove fatigue.

ASANAS AND PRĀṆĀYĀMA

17. Read carefully the hints and cautions for the practice of prāṇāyāma before attempting it (see Part III). Prāṇāyāma may be done either very early in the morning before the āsanas or in the evening after completing them. If early in the morning, prāṇāyāma may be done first for 15 to 30 minutes: then a few minutes of Śavāsana, and after allowing some time to elapse during which one may be engaged in normal activities, practise āsanas. If, however, these are done in the evening, allow at least half an hour to elapse before sitting for prāṇāyāma.

SPECIAL PROVISIONS FOR PERSONS SUFFERING FROM DIZZINESS OR BLOOD PRESSURE

18. Do not start with Śīrṣāsana and Sarvāngāsana if you suffer from dizziness or high blood pressure. First practise Paśchimottānāsana Uttānāsana, and Adho Mukha Śvānāsana before attempting topsy-turvy poses like Śīrṣāsana and Sarvāngāsana and after doing these poses repeat Paśchimottānāsana, Adho Mukha Śvānāsana and Uttānāsana in that order.

19. All forward bending poses are beneficial for persons suffering from either high or low blood pressure.

SPECIAL WARNING FOR PERSONS AFFECTED FROM PUS IN THE EARS OR DISPLACED RETINA

20. Those suffering from pus in the ears or displacement of the retina should not attempt topsy-turvy poses.

SPECIAL PROVISIONS FOR WOMEN

21. *Menstruation:* Avoid āsanas during the menstrual period. But if the flow is in excess of normal, Upaviṣṭha Koṇāsana, Baddha Koṇāsana, Vīrāsana, Jānu Śīrṣāsana, Paśchimottānāsana and Uttānāsana will be beneficial. On no account stand on your head nor perform sarvāngāsana, during the menstrual period.

22. *Pregnancy:* All the āsanas can be practised during the first three months of pregnancy. All the standing poses and the forward bending āsanas may be done with mild movements, for at this time the spine should be made strong and elastic and no pressure be felt on the abdomen. Baddha Koṇāsana and Upaviṣṭha Koṇāsana may be practised throughout pregnancy at any time of the day (even after meals, but not forward bending immediately after meals) as these two āsanas will strengthen the pelvic muscles and the small of the back and also reduce labour pains considerably. Prāṇāyāma without retention (kumbhaka) may be practised throughout pregnancy, as regular deep breathing will help considerably during labour.

23. *After delivery:* No āsanas should be done during the first month after delivery. Thereafter they may be practised mildly. Gradually increase the course as mentioned in the Appendix. Three months after delivery all āsanas may be practised with comfort.

EFFECTS OF ĀSANAS

24. Faulty practice causes discomfort and uneasiness within a few days. This is sufficient to show that one is going wrong. If you cannot find the fault yourself, it is better to approach a person who has practised well and get his guidance.

25. The right method of doing āsanas brings lightness and an exhilarating feeling in the body as well as in the mind and a feeling of oneness of body, mind and soul.

26. Continuous practice will change the outlook of the practiser. He will discipline himself in food, sex, cleanliness and character and will become a new man.

27. When one has mastered an āsana, it comes with effortless ease and causes no discomfort. The bodily movements become graceful. While performing āsanas, the student's body assumes numerous forms of life found in creation – from the lowliest insect to the most perfect sage – and he learns that in all these there breathes the same Universal Spirit – the Spirit of God. He looks within himself while practising and feels the presence of God in different āsanas which he does with a sense of surrender unto the feet of the LORD.

TECHNIQUE AND EFFECTS OF ĀSANAS

(After the name of each āsana, there is a number with an asterisk. These numbers before an asterisk indicate the intensity of the āsana; the lower the number, the easier the āsana, the higher the number, the more difficult the āsana.)

1| *Tāḍāsana*

(also called Samasthiti)

ONE*

Tāda means a mountain. Sama means upright, straight, unmoved. Sthiti is standing still, steadiness. Tāḍāsana therefore implies a pose where one stands firm and erect as a mountain. This is the basic standing pose.

TECHNIQUE

1. Stand erect with the feet together, the heels and big toes touching each other. Rest the heads of metatarsals on the floor and stretch all the toes flat on the floor.

2. Tighten the knees and pull the knee-caps up, contract the hips and pull up the muscles at the back of the thighs.

3. Keep the stomach in, chest forward, spine stretched up and the neck straight.

4. Do not bear the weight of the body either on the heels or the toes, but distribute it evenly on them both.

5. Ideally in Tāḍāsana the arms are stretched out over the head, but for the sake of convenience, one can place them by the side of the thighs. In this case, keep the arms parallel with the body. The fingers together and pointing downwards. Each of the standing poses described below can then be followed easily, starting with the pupil standing in Tāḍāsana with palms by the side of the thighs.

EFFECTS

People do not pay attention to the correct method of standing. Some stand with the body weight thrown only on one leg, or with one leg turned completely sideways. Others bear all the weight on the heels, or on the inner or outer edges of the feet. This can be noticed by watching where the soles and heels of the shoes wear out. Owing to our faulty method of standing and not distributing the body weight evenly on the feet, we acquire specific deformities which hamper spinal elasticity. Even if the feet are kept apart, it is better to keep the heel and toe in a line parallel to the median plane and not at an angle. By this method, the hips are contracted, the abdomen is pulled in and the chest is brought forward. One feels light in body and the mind acquires agility. If we stand with the body weight thrown only on the heels, we feel the centre of gravity changing; the hips become loose, the abdomen protrudes, the body hangs back and the spine feels the strain and consequently we soon feel fatigued and the mind becomes dull. It is therefore essential to master the art of standing correctly.

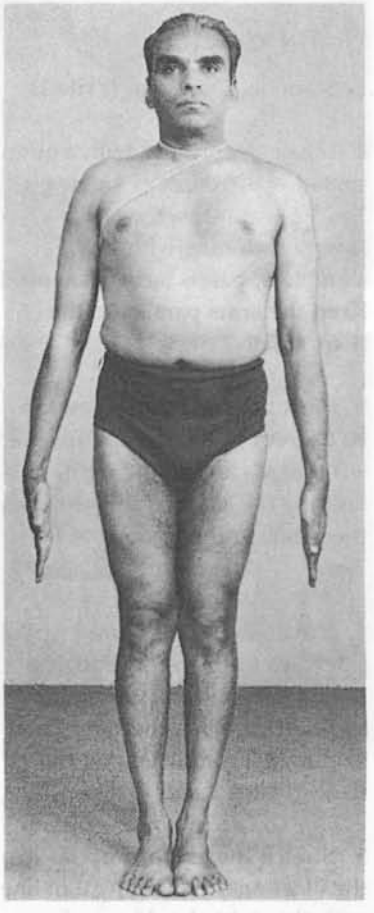

1

2| *Utthita Trikoṇāsana*

THREE*

Utthita means extended, stretched. Trikoṇa (tri = three; koṇa = angle) is a triangle. This standing āsana is the extended triangle pose.

TECHNIQUE

1. Stand in Tāḍāsana. (Plate 1)

2. Inhale deeply and with a jump spread apart the legs sideways 3 to 3½ feet. Raise the arms sideways, in line with the shoulders, palms facing down. Keep the arms parallel to the floor. (Plate 2)

3. Turn the right foot sideways 90 degrees to the right. Turn the left foot slightly to the right, keeping the left leg stretched from the inside and tightened at the knee.

4. Exhale, bend the trunk sideways to the right, bringing the right palm near the right angle. If possible, the right palm should rest completely on the floor. (Plates 3 and 4)

5. Stretch the left arm up (as in the illustration), bringing it in line with the right shoulder and extend the trunk. The back of the legs, the back of the chest and the hips should be in line. Gaze at the thumb of the outstretched left hand. Keep the right knee locked tight by pulling up the knee-cap and keep the right knee facing the toes.

6. Remain in this position from half a minute to a minute, breathing deeply and evenly. Then lift the right palm from the floor. Inhale and return to position 2 above.

7. Now, turn the left foot sideways 90 degrees to the left, turn the right foot slightly to the left, keep both knees tight and continue from position 2 to 6, reversing all processes. Inhale and come to position 2. Hold the posture for the same length of time on the left side.

8. Exhale, and jump, coming back to Tāḍāsana. (Plate 1)

EFFECTS

This āsana tones up the leg muscles, removes stiffness in the legs and hips, corrects any minor deformity in the legs and allows them to develop evenly. It relieves backaches and neck sprains, strengthens the ankles and develops the chest.

2

3

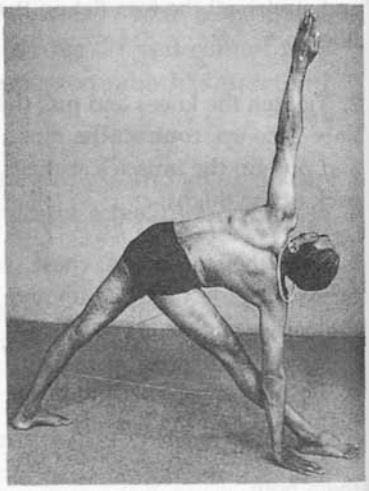

4

3| *Utthita Pārśvakoṇāsana*

FOUR*

Pārśva means side or flank. Koṇa is an angle. This is the extended lateral angle pose.

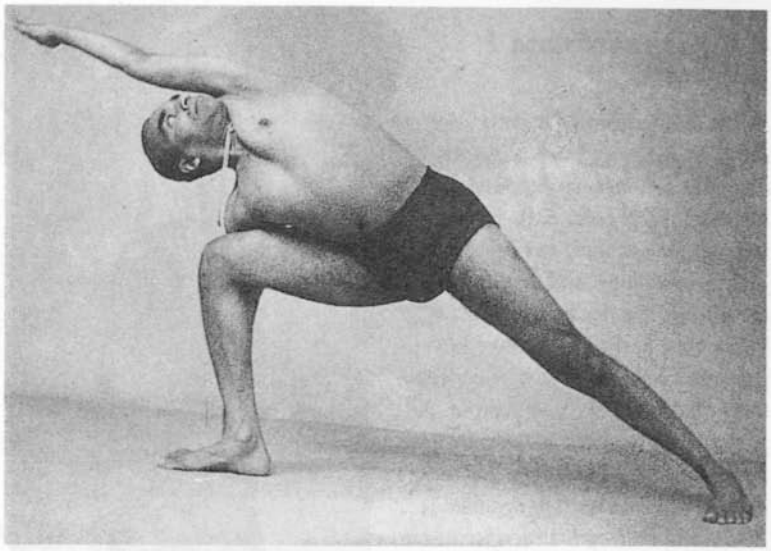

5

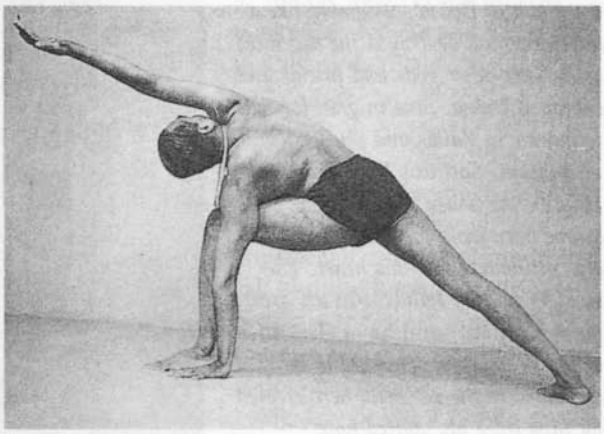

6

TECHNIQUE

1. Stand in Tāḍāsana. (Plate 1) Take a deep inhalation and with a jump spread the legs apart sideways 4 to 4½ feet. Raise the arms sideways, in line with the shoulders, palms facing down. (Plate 2)

2. While exhaling slowly, turn the right foot sideways 90 degrees to the right, and the left foot slightly to the right, keeping the left leg stretched out and tightened at the knee. Bend the right leg at the knee until the thigh and the calf form a right angle and the right thigh is parallel to the floor.

3. Place the right palm on the floor by the side of the right foot, the right armpit covering and touching the outer side of the right knee. Stretch the left arm out over the left ear. Keep the head up. (Plates 5 and 6)

4. Tighten the loins and stretch the hamstrings. The chest, the hips and the legs should be in a line and in order to achieve this, move the chest up and back. Stretch every part of the body, concentrating on the back portion of the whole body, specially the spine. Stretch the spine until all the vertebrae and ribs move and there is a feeling that even the skin is being stretched and pulled.

5. Remain in this pose from half a minute to a minute, breathing deeply and evenly. Inhale and lift the right palm from the floor.

6. Inhale, straighten the right leg and raise the arms as in position 1.

7. Continue with exhalation as in positions 2 to 5, reversing all processes, on the left side.

8. Exhale and jump back to Tāḍāsana. (Plate 1)

EFFECTS

This āsana tones up the ankles, knees and thighs. It corrects defects in the calves and thighs, develops the chest and reduces fat round the waist and hips and relieves sciatic and arthritic pains. It also increases peristaltic activity and aids elimination.

4 *Vīrabhadrāsana I*

THREE*

Dakṣa once celebrated a great sacrifice, but he did not invite his daughter Satī nor her husband Śiva, the chief of the gods. Satī, however, went to the sacrifice, but being greatly humiliated and insulted threw herself into the fire and perished. When Śiva heard this he was gravely provoked, tore a hair from his matted locks and threw it to the ground. A powerful hero named Vīrabhadra rose up and awaited his orders. He was told to lead Śiva's army against Dakṣa and destroy his sacrifice. Vīrabhadra and his army appeared in the midst of Dakṣa's assembly like a hurricane and destroyed the sacrifice, routed the other gods and priests and beheaded Dakṣa. Śiva in grief for Satī withdrew to Kailās and plunged into meditation. Satī was born again as Umā in the house of Himālaya. She strove once more for the love of Śiva and ultimately won his heart. The story is told by Kālidāsa in his great poem Kumāra saṁbhava *(The Birth of the War-Lord). This āsana is dedicated to the powerful hero created by Śiva from his matted hair.*

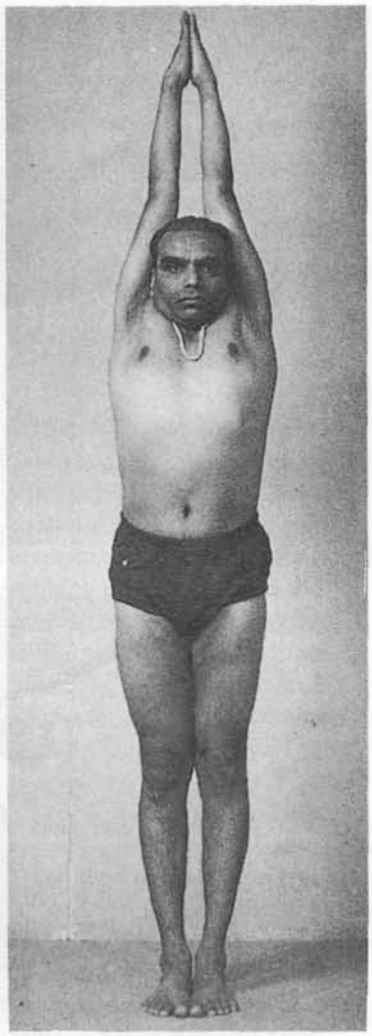

7

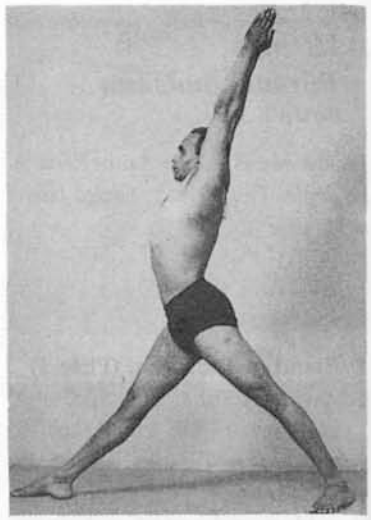

8

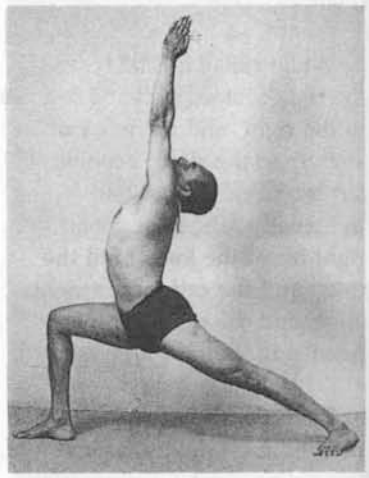

9

TECHNIQUE

1. Stand in Tāḍāsana. (Plate 1)

2. Raise both arms above the head; stretch up and join the palms. (Plate 7)

3. Take a deep inhalation and with a jump spread the legs apart sideways 4 to 4½ feet.

4. Exhale, turn to the right. Simultaneously turn the right foot 90 degrees to the right and the left foot slightly to the right. (Plate 8) Flex the right knee till the right thigh is parallel to the floor and the right shin perpendicular to the floor, forming a right angle between the right thigh and the right calf. The bent knee should not extend beyond the ankle, but should be in line with the heel.

5. Stretch out the left leg and tighten at the knee.

6. The face, chest and right knee should face the same way as the right foot, as illustrated. Throw the head up, stretch the spine from the coccyx and gaze at the joined palms. (Plate 9)

7. Hold the pose from 20 seconds to half a minute with normal breathing.

8. Repeat on the left side as in positions 4 to 6, reversing all processes.

9. Exhale and jump back to Tāḍāsana. (Plate 1)

All standing poses are strenuous, this pose in particular. It should not be tried by persons with a weak heart. Even people who are fairly strong should not stay long in this āsana.

EFFECTS

In this pose the chest is fully expanded and this helps deep breathing. It relieves stiffness in shoulders and back, tones up the ankles and knees and cures stiffness of the neck. It also reduces fat round the hips.

5| *Vīrabhadrāsana II*

ONE*

TECHNIQUE

1. Stand in Tāḍāsana. (Plate 1)

2. Take a deep inhalation, and with a jump spread the legs apart sideways 4 to 4½ feet. Raise the arms sideways in line with the shoulders, palms facing down. (Plate 2)

3. Turn the right foot sideways 90 degrees to the right and the left foot slightly to the right, keeping the left leg stretched out and tightened at the knee. Stretch the hamstring muscles of the left leg.

4. Exhale and bend the right knee till the right thigh is parallel to the floor, keeping the right shin perpendicular to the floor,

10

thus forming a right angle between the right thigh and the right calf. The bent knee should not extend beyond the ankle, but should be in line with the heel. (Plate 10)

5. Stretch out the hands sideways, as though two persons are pulling you from opposite ends.

6. Turn the face to the right and gaze at the right palm. Stretch the back muscles of the left leg fully. The back of the legs, the dorsal region and the hips should be in one line.

7. Stay in the pose from 20 seconds to half a minute with deep breathing. Inhale and return to position 2.

8. Turn the left foot sideways 90 degrees to the left and the right foot slightly to the left, flex the left knee and continue from positions 3 to 6 on the left side, reversing all processes.

9. Inhale, again come back to position 2. Exhale and jump back to Tāḍāsana. (Plate 1)

EFFECTS

Through this pose the leg muscles become shapely and stronger. It relieves cramp in the calf and thigh muscles, brings elasticity to the leg and back muscles and also tones the abdominal organs.

Mastery of the standing poses prepares the pupil for the advanced poses in forward bending, which can then be acquired with ease.

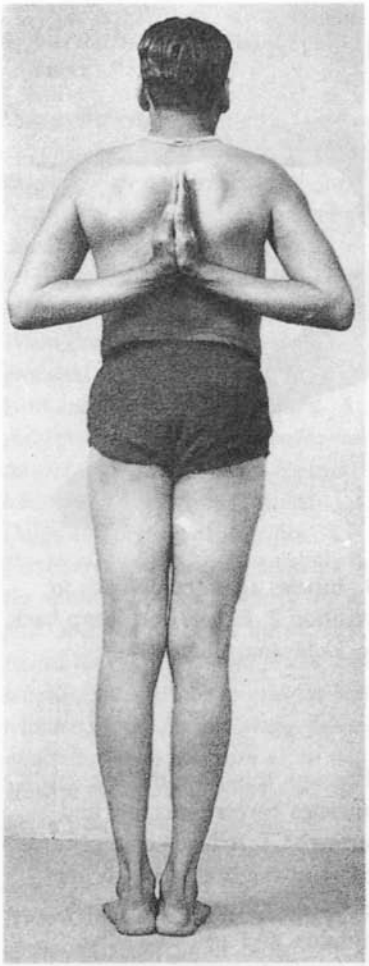

11

12

6| *Pārśvōttānāsana*

SIX*

Pārśva means side or flank. Uttāna (ut = intense, and tān = to extend, stretch, lengthen) means an intense stretch. The name implies a pose in which the side of the chest is stretched intensely.

TECHNIQUE

1. Stand in Tāḍāsana. (Plate 1) Inhale deeply and stretch the body forward.

2. Join the palms behind the back and draw the shoulders and elbows back.

3. Exhale, turn the wrists and bring both palms up above the middle of the back of the chest, the fingers at the level of the shoulder-blades. You are doing 'namaste' (the Indian gesture of respect by folding the hands) with your hands behind your back. (Plate 11)

4. Inhale and with a jump spread the legs apart sideways 3 to 3½ feet. Stay in this position and exhale.

5. Inhale and turn the trunk to the right. Turn the right foot 90 degrees sideways to the right keeping the toes and heel in a line with the trunk; turn the left foot with the leg 75 to 80 degrees to the right and keep the left foot stretched out and the leg tightened at the knee. Throw the head back. (Plate 12)

6. Exhale, bend the trunk forward and rest the head on the right knee. Stretch the back and gradually extend the neck until the nose, then the lips and lastly the chin touch and then rest beyond the right knee. (Plate 13) Tighten both the legs by pulling the knee-caps up.

7. Stay in the pose from 20 seconds to half a minute with normal breathing. Then slowly move the head and trunk towards the left knee by swinging the trunk round the hips. At the same time turn the left foot 90 degrees towards the left and the right foot 75 to 80 degrees to the left. Now raise the trunk and head as far back as you can, without bending the right leg. This movement should be done with one inhalation.

8. Exhale, bend the trunk forward, rest the head on the left knee and gradually extend the chin beyond the left knee by stretching the neck as in position 6.

9. After holding the pose from 20 seconds to half a minute with normal breathing, inhale, move the head to the centre and the feet to their original position so that the toes point forward. Then raise the trunk up.

10. Exhale and jump back to Tāḍāsana (Plate 1), releasing the hands from the back.

11. If you cannot fold the hands together behind the back, just grip the right wrist and follow the above technique. (Plate 14)

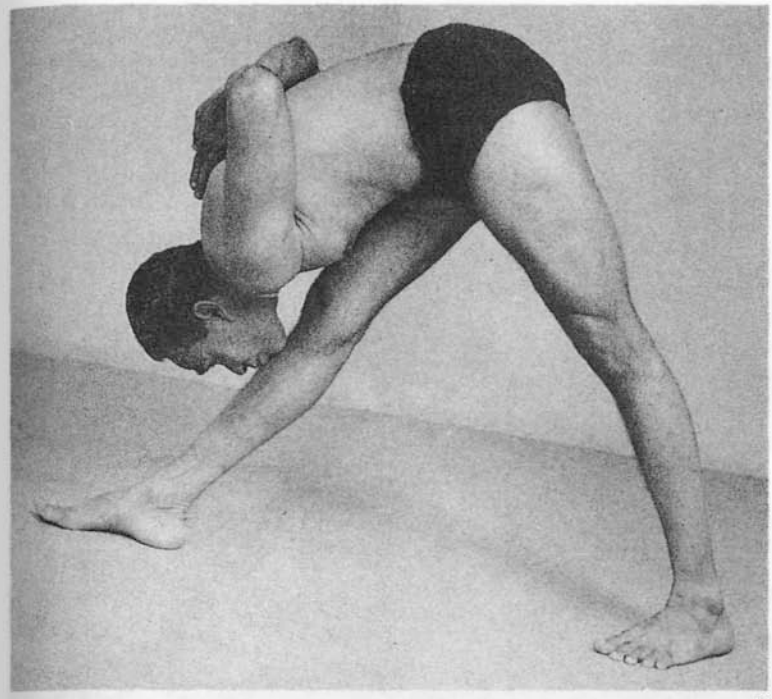

13

EFFECTS

This āsana relieves stiffness in the legs and hip muscles and makes the hip joints and spine elastic. While the head is resting on the knees, the abdominal organs are contracted and toned. The wrists move freely and any stiffness there disappears. The posture also corrects round and drooping shoulders. In the correct pose, the shoulders are drawn well back and this makes deep breathing easier.

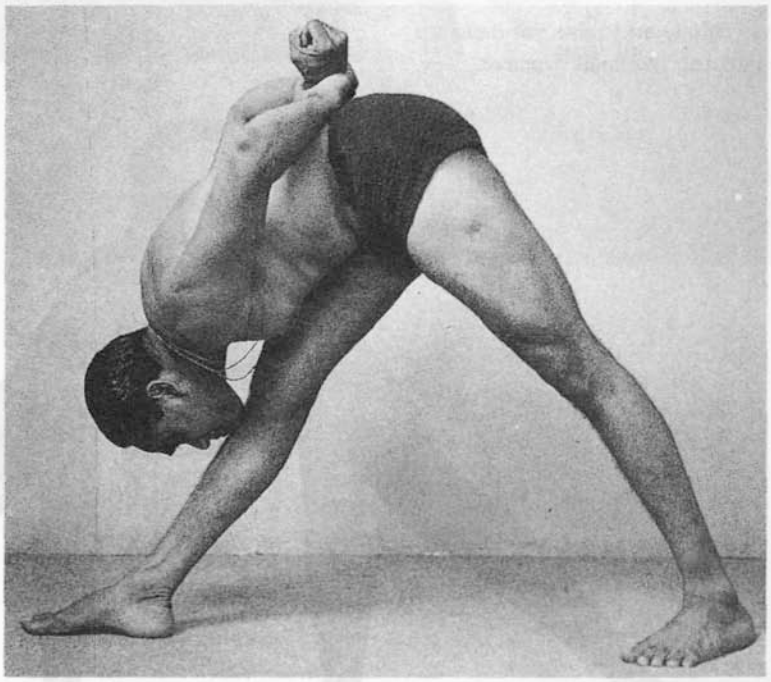

14

7| *Prasārita Pādottānāsana*

FOUR*

Prasārita means expanded, spread, extended. Pāda means a foot. The pose is one where the expanded legs are stretched intensely.

TECHNIQUE

1. Stand in Tāḍāsana. (Plate 1)

2. Inhale, place the hands on the waist and spread the legs apart 4½ to 5 feet. (Plate 15)

3. Tighten the legs by drawing up the knee-caps. Exhale, and place the palms on the floor in line with the shoulders between the feet. (Plate 16)

4. Inhale and raise the head up, keeping the back concave.

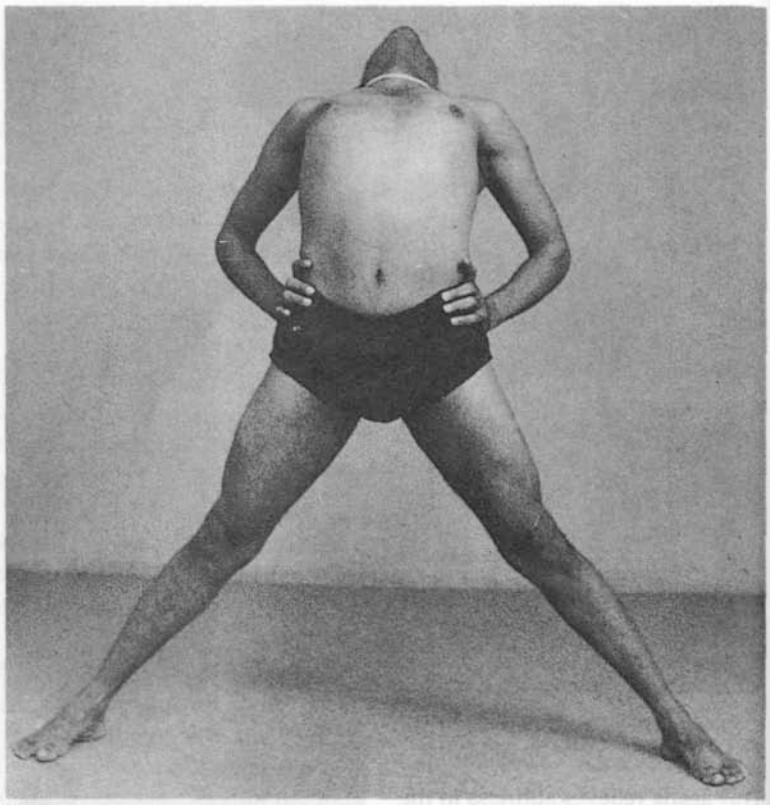

15

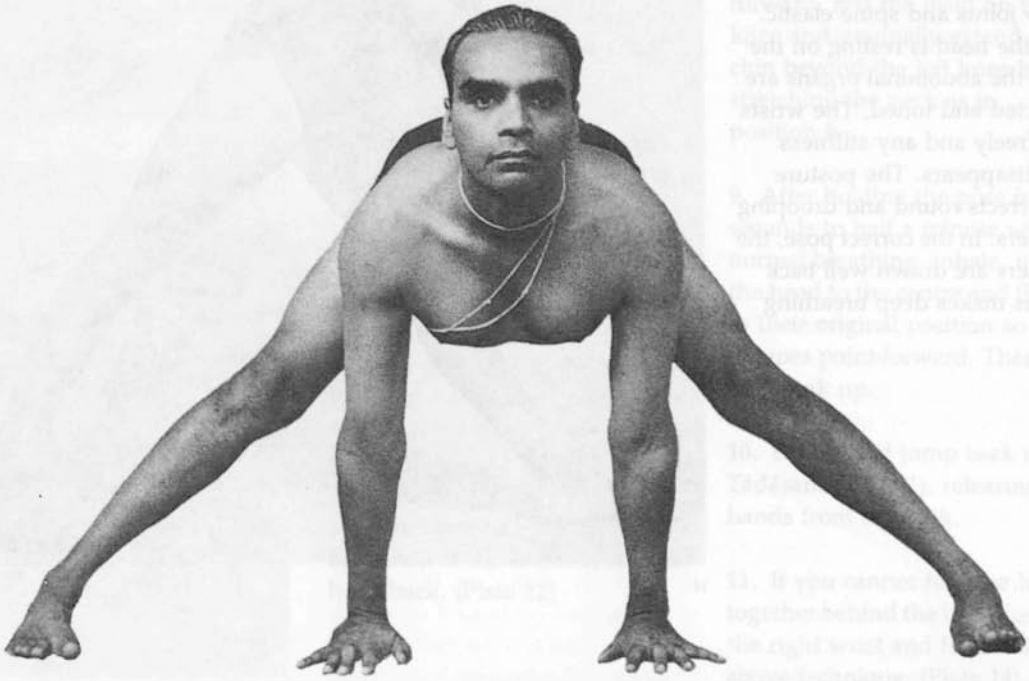

16

5. Exhale, bend the elbows and rest the crown of the head on the floor, keeping the weight of the body on the legs. (Plates 17 and 18) Do not throw the body weight on the head. Both feet, both palms and the head should be in a straight line.

6. Stay in the pose for half a minute, breathing deeply and evenly.

7. Inhale, raise the head from the floor and straighten the arms at the elbows. Keep the head well up by making the back concave as in position 4. (Plate 16)

8. Exhale and stand as in position 2. (Plate 15)

9. Jump back to Tāḍāsana. (Plate 1)

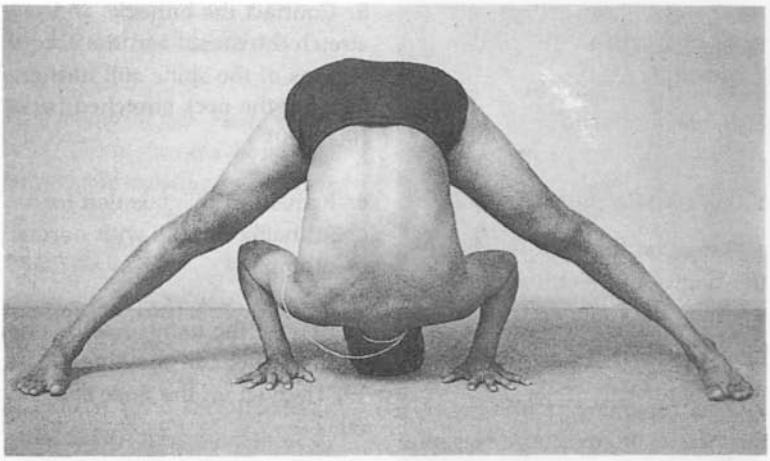

17

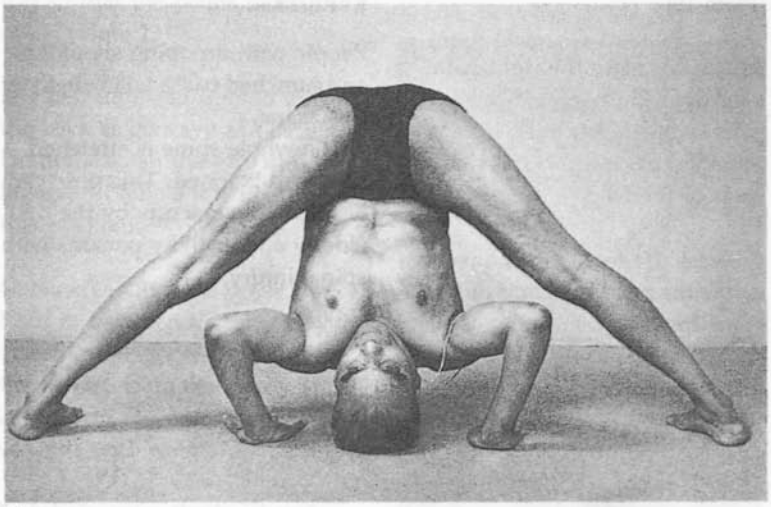

18

EFFECTS

In this pose the hamstring and abductor muscles are fully developed, while blood is made to flow to the trunk and the head. People who cannot do Sīrṣāsana (Plate 90) can benefit from this pose, which increases digestive powers.

All the standing poses described above are necessary for beginnings. As the pupil advances he attains better flexibility and then the standing poses can be dispensed with, though it is advisable to do them once a week. All these standing poses help to reduce the body weight.

8| *Uṣṭrāsana*

THREE*

Uṣṭra means a camel.

TECHNIQUE

1. Kneel on the floor, keeping the thighs and feet together, toes pointing back and resting on the floor.

2. Rest the palms on the hips. Stretch the thighs, curve the spine back and extend the ribs. (Plate 19)

3. Exhale, place the right palm over the right heel and the left palm over the left heel. If possible, place the palms on the soles of the feet.

4. Press the feet with the palms, throw the head back and push the spine towards the thighs, which should be kept perpendicular to the floor.

5. Contract the buttocks and stretch the dorsal and the coccyx regions of the spine still further, keeping the neck stretched back. (Plate 20)

6. Remain in this position for about half a minute with normal breathing.

7. Release the hands one by one and rest them on the hips. (Plate 19) Then sit on the floor and relax.

EFFECTS

People with drooping shoulders and hunched backs will benefit by this āsana.

The whole spine is stretched back and is toned. This pose can be tried conveniently by the elderly and even by persons with spinal injury.

19

20

21

9| *Pādāṅguṣṭhāsana*

THREE*

Pāda means the foot. Aṅguṣṭha is the big toe. This posture is done by standing and catching the big toes.

TECHNIQUE

1. Stand in Tāḍāsana. (Plate 1) Spread the legs a foot apart.

2. Exhale, bend forward and hold the big toes between the thumbs and the first two fingers, so that the palms face each other. Hold them tight. (Plate 21)

3. Keep the head up, stretch the diaphragm towards the chest and make the back as concave as possible. Instead of stretching down from the shoulders, bend forward from the pelvic region to get the concave shape of the back from the coccyx.

4. Keep the legs stiff and do not slacken the grip at the knees and toes. Stretch the shoulder-blades also. Take one or two breaths in this position.

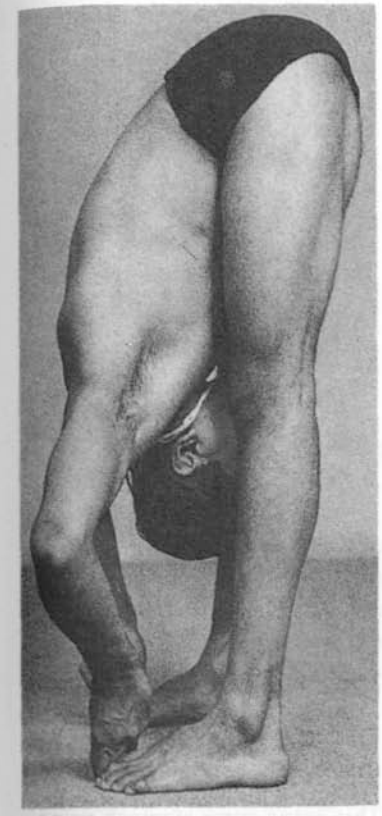

22

5. Now exhale, and bring the head in between the knees by tightening and pulling the toes without lifting them off the floor. (Plate 22) Remain in this pose for about 20 seconds, maintaining normal breathing.

6. Inhale, come to position 2 (Plate 21), release the toes and stand up. Return to Tāḍāsana. (Plate 1)

10| *Pādahastāsana*

SIX*

Pāda means the foot. Hasta means the hand. This posture is done by bending forward and standing on one's hands.

TECHNIQUE

1. Stand in Tāḍāsana. (Plate 1) Spread the legs a foot apart.

2. Exhale, bend forward and without bending the legs at the knees insert the hands under the feet so that the palms touch the soles. (Plate 23)

3. Keep the head up and make the back as concave as possible. Do not slacken the grip at the knees and take a few breaths in this position.

4. Now exhale, and move the head in between the knees by bending the elbows and pulling the feet up from the palms. (Plate 24) Stay in the pose for about 20 seconds with normal breathing.

5. Inhale, raise the head and come back to position 2 (Plate 23), with the head well up. Take two breaths.

6. Inhale, stand up and return to Tāḍāsana. (Plate 1)

23

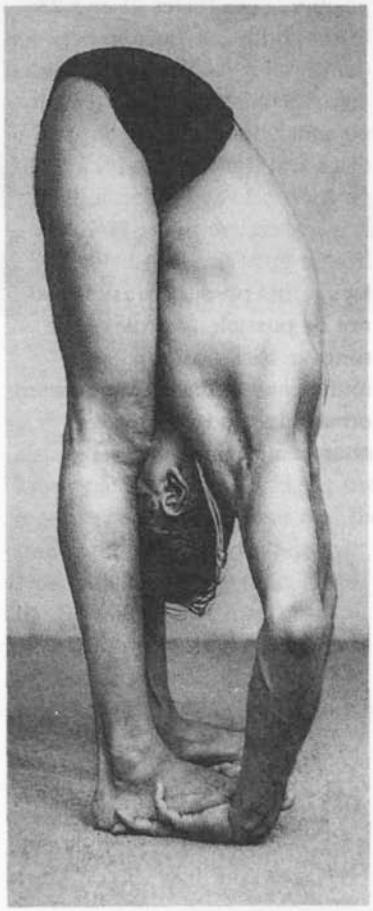

24

EFFECTS OF PĀDĀNGUṢṬHĀSANA AND PĀDAHASTĀSANA

The second āsana is more strenuous than the first, but the effects of both are the same. The abdominal organs are toned and digestive juices increase, while the liver and spleen are activated. Persons suffering from a bloating sensation in the abdomen or from gastric troubles will benefit from practising these two āsanas.

Slipped spinal discs can only be adjusted in the concave back position as in Plates 21 and 23. Do not bring the head in between the knees if you have a displaced disc. I have experimented with persons suffering from slipped discs and the concave back position proved a boon to them. It is imperative to get guidance from a guru (master) before trying this pose, because it may not be possible to achieve the concave back position immediately. One has to master other minor poses before attempting this one.

11| *Uttānāsana*

EIGHT*

Ut is a particle indicating deliberation, intensity. The verb tān means to stretch, extend, lengthen out. In this āsana, the spine is given a deliberate and an intense stretch.

TECHNIQUE

1. Stand in Tāḍāsana (Plate 1), keeping the knees tight.

2. Exhale, bend forward and place the fingers on the floor. Then place the palms on the floor by the side of the feet, behind the heels. Do not bend the legs at the knees.

3. Try to hold the head up and stretch the spine. Move the hips a little forward towards the head so as to bring the legs perpendicular to the floor.

4. Remain in this position and take two deep breaths.

5. Exhale, move the trunk closer to the legs and rest the head on the knees. (Plate 25)

6. Do not slacken the grip at the knees, but pull the knee-caps well up. Hold this position for a minute with deep and even breathing.

7. Inhale and raise the head from the knees, but without lifting the palms from the floor as in position 3.

8. After two breaths, take a deep inhalation, lift the hands from the floor and come back to Tāḍāsana. (Plate 1)

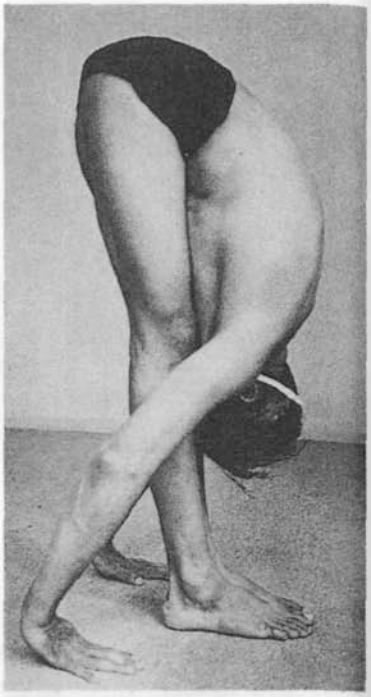

25

EFFECTS

This āsana cures stomach pains and tones the liver, the spleen and the kidneys. It also relieves stomach pain during menstrual periods. The heart beats are slowed down and the spinal nerves rejuvenated. Any depression felt in the mind is removed if one holds the pose for two minutes or more. The posture is a boon to people who get excited quickly, as it soothes the brain cells. After finishing the āsana, one feels calm and cool, the eyes start to glow and the mind feels at peace.

Persons who feel heaviness in the head, flushing or any discomfort while attempting Śīrṣāsana (Plate 90), should do Uttānāsana first; then they will be able to do Śīrṣāsana (the head stand) with comfort and ease.

12| Śalabhāsana
ONE*

Śalabhā means a locust. The pose resembles that of a locust resting on the ground, hence the name.

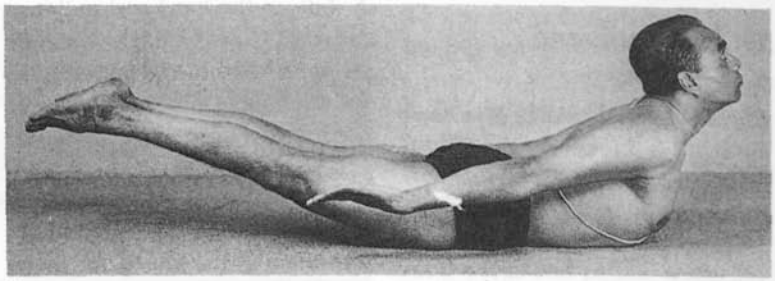

26

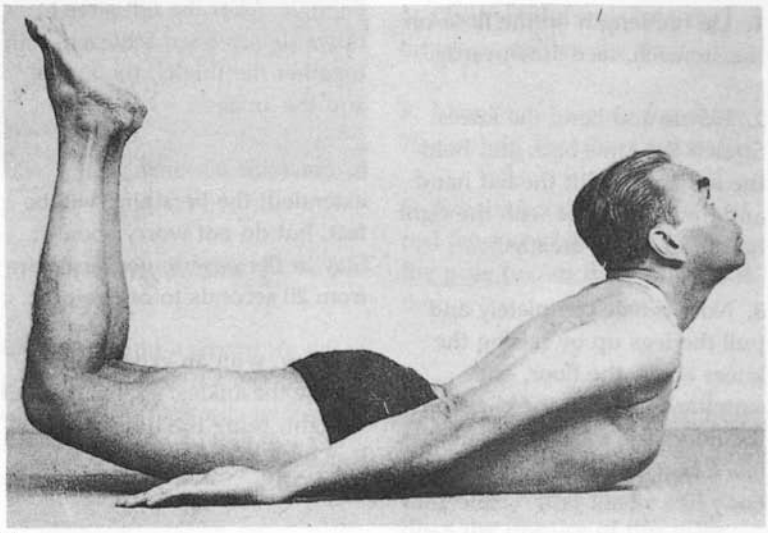

27

TECHNIQUE

1. Lie full length on the floor on the stomach, face downwards. Stretch the arms back.

2. Exhale, lift the head, chest and legs off the floor simultaneously as high as possible. The hands should not be placed and the ribs should not rest on the floor. Only the abdominal front portion of the body rests on the floor and bears the weight of the body. (Plate 26)

3. Contract the buttocks and stretch the thigh muscles. Keep both legs fully extended and straight, touching at the thighs, knees and ankles.

4. Do not bear the weight of the body on the hands but stretch them back to exercise the upper portion of the back muscles.

5. Stay in the position as long as you can with normal breathing.

6. In the beginning it is difficult to lift the chest and the legs off the floor, but this becomes easier as the abdominal muscles grow stronger.

EFFECTS

The pose aids digestion and relieves gastric troubles and flatulence. Since the spine is stretched back it becomes elastic and the pose relieves pain in the sacral and lumbar regions. In my experience, persons suffering from slipped discs have benefited by regular practice of this āsana without recourse to enforced rest or surgical treatment. The bladder and the prostate gland also benefit from the exercise and remain healthy.

A variation of the pose may also be tried to relieve aches in the lower part of the back. Here, the legs are bent at the knees and the thighs are kept apart while the shins are kept perpendicular to the floor. Then with an exhalation, the thighs are lifted off the floor and brought closer together until the knees touch, the shins still being kept perpendicular. (Plate 27)

13| *Dhanurāsana*

FOUR*

Dhanu means a bow. The hands here are used like a bowstring to pull the head, trunk and legs up and the posture resembles a bent bow.

TECHNIQUE

1. Lie full length on the floor on the stomach, face downwards.

2. Exhale and bend the knees. Stretch the arms back and hold the left ankle with the left hand and the right ankle with the right hand. Take two breaths.

3. Now exhale completely and pull the legs up by raising the knees above the floor, and simultaneously lift the chest off the floor. The arms and hands act like a bow-string to tauten the body like a bent bow. (Plate 28)

4. Lift up the head and pull it as far back as possible. Do not rest either the ribs or the pelvic bones on the floor. Only the abdomen bears the weight of the body on the floor.

5. While raising the legs do not join them at the knees, for then the legs will not be lifted high enough. After the full stretch upwards has been achieved, join together the thighs, the knees and the ankles.

6. Since the abdomen is extended, the breathing will be fast, but do not worry about it. Stay in the pose to your capacity from 20 seconds to one minute.

7. Then, with an exhalation, release the ankles, stretch the legs straight, bring the head and the legs back to the floor and relax.

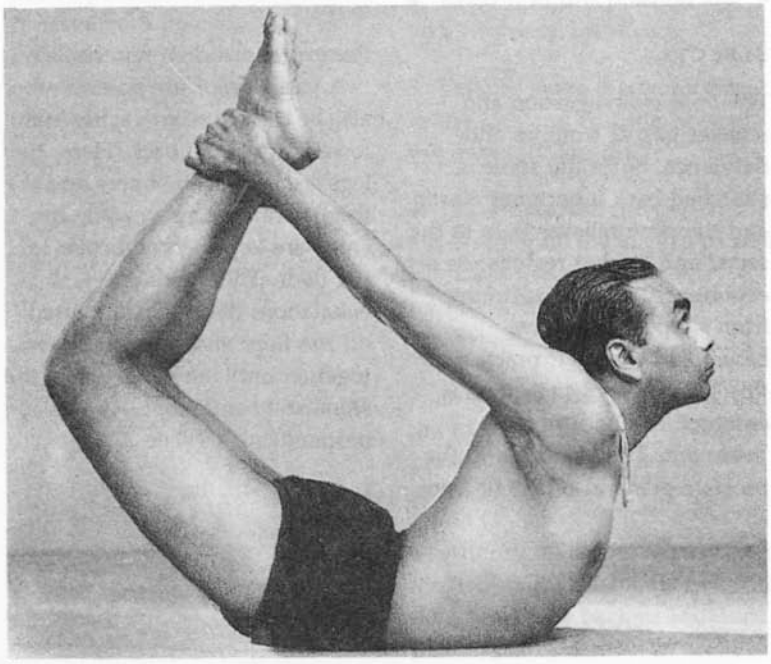

28

EFFECTS

In this posture the spine is stretched back. Elderly people do not normally do this, so their spines get rigid. This āsana brings back elasticity to the spine and tones the abdominal organs. In my experience, persons suffering from slipped discs have obtained relief by the regular practice of Dhanurāsana and Śalabhāsana (Plate 26) without being forced to rest or to undergo surgical treatment.

14| *Chaturaṅga Daṇḍāsana*

ONE*

Chatur means four. Aṅga means a limb or a part thereof. Daṇḍa means a staff. Lie flat on the floor, face down and take the weight of the body on the palms and toes, exhale and keep the body parallel to the floor, stiff as a staff. The four limbs supporting the body are the hands and feet. The pose is similar to dips in western gymnastics.

TECHNIQUE

1. Lie flat on the floor, face downwards.

2. Bend the elbows and place the palms by the side of the chest. Keep the feet about a foot apart.

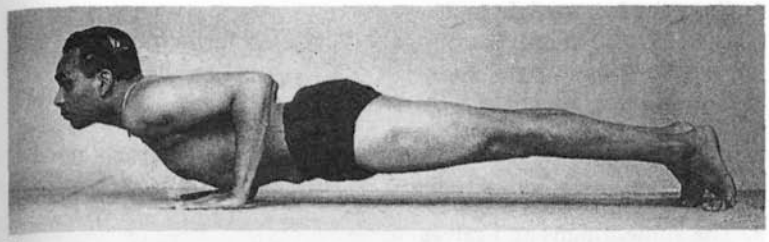

29

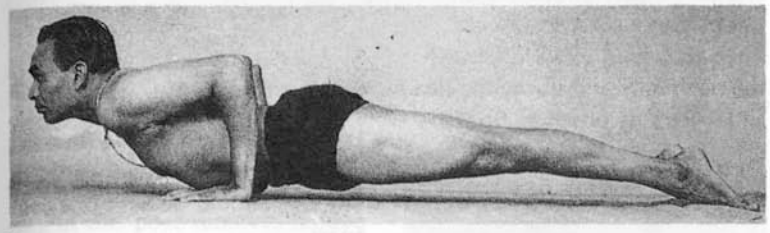

30

3. With an exhalation, raise the whole body a few inches above the floor, balancing it on the hands and the toes. (Plate 29) Keep the body stiff as a staff, parallel to the floor from head to heel and the knees taut. Stay for some time with normal breathing.

4. Then gradually extend the whole body forward so that the feet rest on the upper portion of the toes on the floor. (Plate 30)

5. Stay in the pose for about 30 seconds with normal or deep breathing. The movement may be repeated several times. Then relax on the floor.

EFFECTS

The pose strengthens the arms and the wrists develop mobility and power. It also contracts and tones the abdominal organs.

15| *Bhujaṅgāsana I*

ONE*

Bhujanga means a serpent. In this posture, lie flat on the floor, face downwards, lift the body up from the trunk and throw the head back like a serpent about to strike.

TECHNIQUE

1. Lie on the floor face downwards. Extend the legs, keeping the feet together. Keep the knees tight and the toes pointing.

2. Rest the palms by the side of the pelvic region.

3. Inhale, press the palms firmly on the floor and lift the body up from the trunk until the pubis is in contact with the floor and stay in this position with the weight on the legs and palms. (Plate 31)

4. Contract the anus and the buttocks, tighten the thighs.

5. Maintain the pose for about 20 seconds, breathing normally.

6. Exhale, bend the elbows and rest the trunk on the floor. Repeat the pose two or three times and then relax.

EFFECTS

The posture is a panacea for an injured spine and in cases of slight displacement of spinal discs the practice of this pose replaces the discs in their original position. The spinal region is toned and the chest fully expanded.

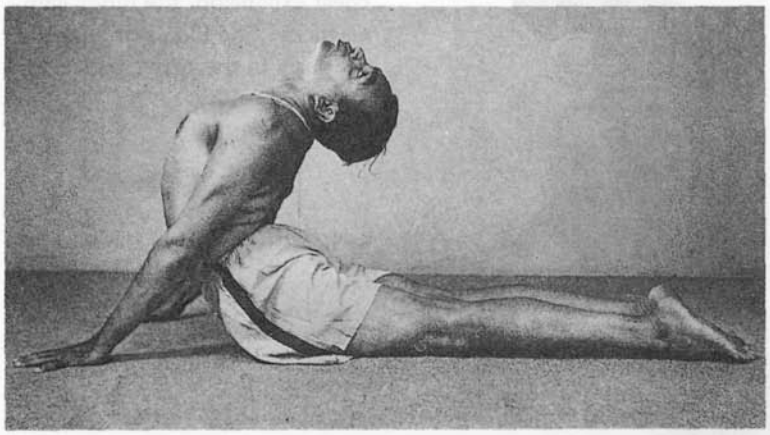

31

16| *Ūrdhva Mukha Śvānāsana*

ONE*

Ūrdhva Mukha means having the mouth upwards. Śvāna means a dog. The pose resembles a dog stretching itself with the head up in the air, hence the name.

TECHNIQUE

1. Lie on the floor on the stomach, face downwards.

2. Keep the feet about one foot apart. The toes should point straight back. Place the palms on the floor by the side of the waist, the fingers pointing to the head.

3. Inhale, raise the head and trunk, stretch the arms completely and push the head and trunk as far back as possible, without resting the knees on the floor.

4. Keep the legs straight and tightened at the knees, but do not rest the knees on the floor. The weight of the body rests on the palms and toes only. (Plate 32)

5. The spine, thighs and calves should be fully stretched, and the buttocks contracted tight. Push the chest forward, stretch the neck fully and throw the head as far back as possible. Stretch also the back portions of the arms.

6. Stay in the pose from half a minute to a minute with deep breathing.

7. Bend the elbows, release the stretch and rest on the floor.

EFFECTS

The pose rejuvenates the spine and is specially recommended for people suffering from a stiff back. The movement is good for persons with lumbago, sciatica and those suffering from slipped or prolapsed discs of the spine. The pose strengthens the spine and cures backaches. Due to chest expansion, the lungs gain elasticity. The blood circulates properly in the pelvic region and keeps it healthy.

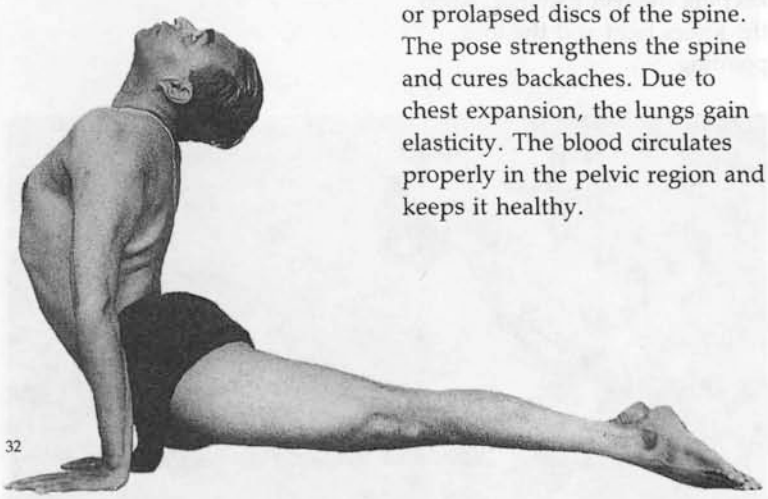

32

17| *Adho Mukha Śvānāsana*

FIVE*

Adho Mukha means having the face downwards. Śvāna means a dog. The pose resembles a dog stretching itself with head and forelegs down and the hind legs up, hence the name.

TECHNIQUE

1. Lie full length on the floor on the stomach, face downwards. The feet should be kept one foot apart.

2. Rest the palms by the side of the chest, the fingers straight and pointing in the direction of the head.

3. Exhale and raise the trunk from the floor. Straighten the arms, move the head inwards towards the feet and place the crown of the head on the floor, keeping the elbows straight and extending the back. (Side view: Plate 33. Back view: Plate 34)

4. Keep the legs stiff and do not bend the knees but press the heels down. The heels and soles of the feet should rest completely on the floor, while the feet should be parallel to each other, the toes pointing straight ahead.

5. Stay in the pose for about a minute with deep breathing. Then with an exhalation lift the head off the floor, stretch the trunk forward and lower the body gently to the floor and relax.

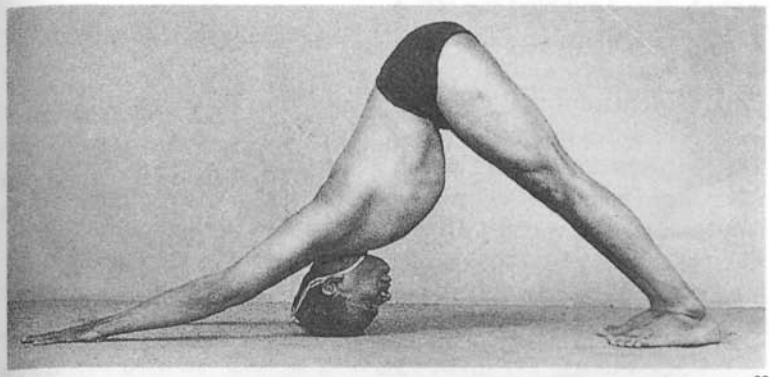

33

EFFECTS

When one is exhausted, a longer stay in this pose removes fatigue and brings back the lost energy. The pose is especially good for runners who get tired after a hard race. Sprinters will develop speed and lightness in the legs. The pose relieves pain and stiffness in the heels and helps to soften calcaneal spurs. It strengthens the ankles and makes the legs shapely. The practice of this āsana helps to eradicate stiffness in the region of the shoulder-blades, and arthritis of the shoulder joints is relieved. The abdominal muscles are drawn towards the spine and strengthened. As the diaphragm is lifted to the chest cavity the rate of the heart beat is slowed down. This is an exhilarating pose.

Those who are afraid to do Śīrṣāsana (Plate 90) can conveniently practise this position. As the trunk is lowered in this āsana it is fully stretched and healthy blood is brought to this region without any strain on the heart. It rejuvenates the brain cells and invigorates the brain by relieving fatigue.

Persons suffering from high blood pressure can do this pose.

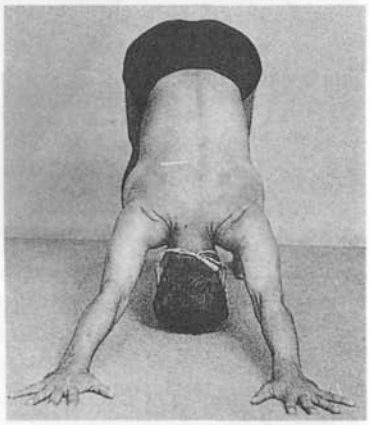

34

18| *Daṇḍāsana*

TWO*

Daṇḍa means a staff or rod.

TECHNIQUE

Sit on the floor with the legs stretched in front. Place the palms on the floor by the hips, the fingers pointing to the feet. Stretch the hands straight and keep the back erect. (Plate 35)

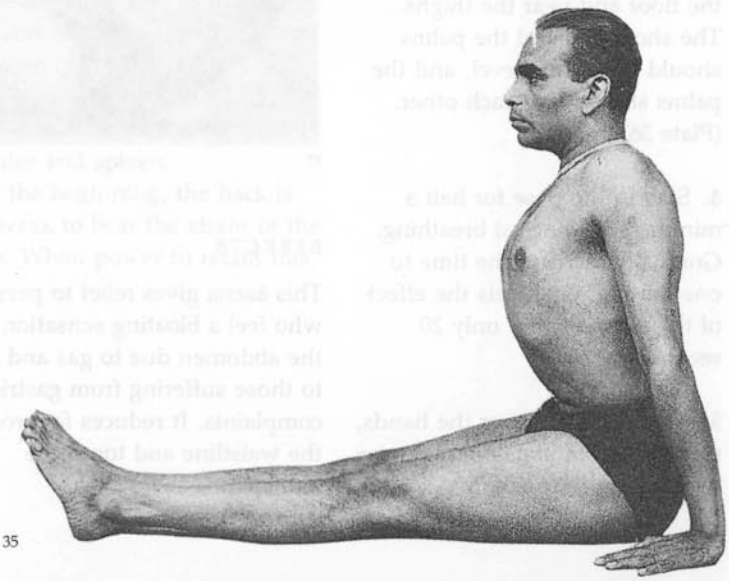

35

19| *Paripūrṇa Nāvāsana*
TWO*

Paripūrṇa means entire or complete. The posture here resembles that of a boat with oars, hence the name.

TECHNIQUE

1. Sit on the floor as in Daṇḍāsana (18 above).

2. Exhale, recline the trunk slightly back and simultaneously raise the legs from the floor and keep them stiff as a poker with the knees tight and the toes pointing forwards. Balance is maintained only on the buttocks and no part of the spine should be allowed to touch the floor, from which the legs should be kept at an angle of 60 to 65 degrees. The feet are higher than the head and not level with it as in Ardha Nāvāsana. (Plate 37)

3. Remove the hands from the floor and stretch the arms forward, keeping them parallel to the floor and near the thighs. The shoulders and the palms should be on one level, and the palms should face each other. (Plate 36)

4. Stay in the pose for half a minute, with normal breathing. Gradually increase the time to one minute. One feels the effect of the exercise after only 20 seconds.

5. Then exhale, lower the hands, rest the legs on the floor and relax by lying on the back.

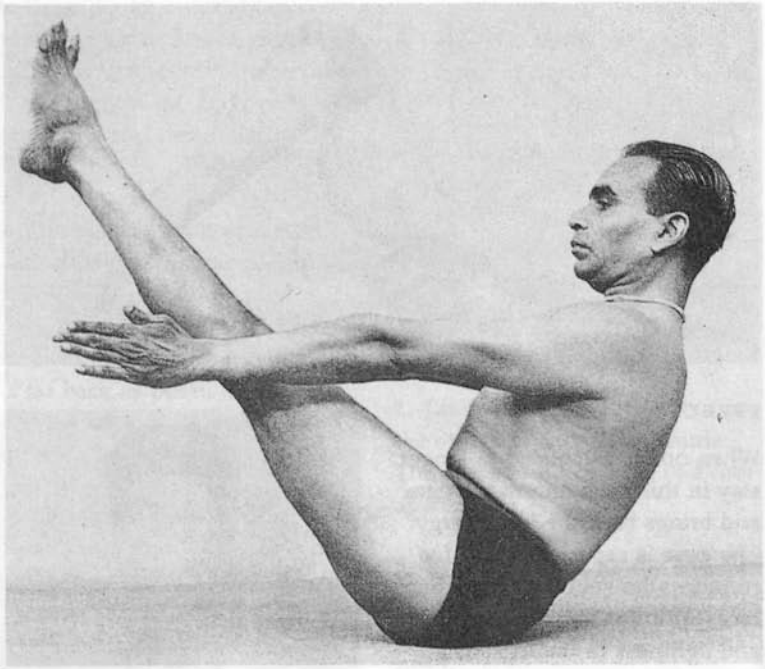

36

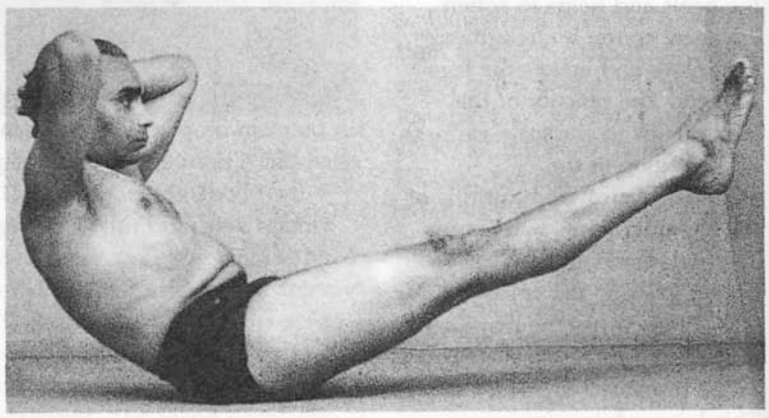

37

EFFECTS

This āsana gives relief to persons who feel a bloating sensation in the abdomen due to gas and also to those suffering from gastric complaints. It reduces fat around the waistline and tones the kidneys.

20| *Ardha Nāvāsana*

TWO*

Ardha means half. Nāva is a ship, boat or vessel. This posture resembles the shape of a boat, hence the name.

TECHNIQUE

1. Sit on the floor. Stretch the legs out in front and keep them straight. (Plate 35)

2. Interlock the fingers and place them on the back of the head just above the neck.

3. Exhale, recline the trunk back and simultaneously raise the legs from the floor, keeping the thighs and knees tight and the toes pointed. The balance of the body rests on the buttocks and no part of the spine should be allowed to touch the floor. (Plate 37) One feels the grip on the muscles of the abdomen and the lower back.

4. Keep the legs at an angle of about 30 to 35 degrees from the floor and the crown of the head in line with the toes.

5. Hold this pose for 20 to 30 seconds with normal breathing. A stay for one minute in this posture indicates strong abdominal muscles.

6. Do not hold the breath during this āsana, though the tendency is always to do it with suspension of breath after inhalation. If the breath is held, the effect will be felt on the stomach muscles and not on the abdominal organs. Deep inhalation in this āsana would loose the grip on the abdominal muscles. In order to maintain this grip, inhale, exhale and hold the breath and go on repeating this process but without breathing deeply. This will exercise not only the abdominal muscles but the organs also.

7. The difference between Ardha Nāvāsana and Paripūrṇa Nāvāsana should be noted; in the latter, the legs are moved higher and the distance between them and the stomach is less than in the former.

EFFECTS

The effects of Ardha Nāvāsana and that of Paripūrṇa Nāvāsana (Plate 36) differ due to the position of the legs. In Paripūrṇa Nāvāsana the exercise is effective on the intestines; whereas, Ardha Nāvāsana works on the liver, gall bladder and spleen.

In the beginning, the back is too weak to bear the strain of the pose. When power to retain this pose comes, it indicates that the back is gaining strength. A weak back is a handicap in many ways, especially to women as they need strong backs for child-bearing. These two āsanas coupled with lateral twistings of the spine will help to strengthen the back.

The importance of having a healthy lower back can be realized if we watch old people when they sit down, get up and walk, for consciously or unconsciously they support their backs with their hands. This indicates that the back is weak and cannot withstand the strain. As long as it is strong and needs no support, one feels young though advanced in age. The two āsanas bring life and vigour to the back and enable us to grow old gracefully and comfortably.

21| *Siddhāsana*

ONE*

Siddha means a semi-divine being supposed to be of great purity and holiness, and to possess supernatural faculties called siddhis. Siddha means also an inspired sage, seer or prophet.

'The Siddhas say that as among niyamas, the most important is not to harm anyone, and among the yamas a moderate diet, so is Siddhāsana among the āsanas.'

'Of the 84 lacs of āsanas, one should always practise Siddhāsana. It purifies 72,000 nāḍīs. (Nāḍīs are channels in the human body through which nervous energy passes.)

'The yogin practising contemplation upon Ātman and observing a moderate diet, if he practises Siddhāsana for twelve years, obtains the yoga siddhis.' (Ātman means the Self and the supreme Soul. Siddhīs are supernatural faculties.)

'When Siddhāsana is mastered, the Unmanī Avasthā (Samādhi) that gives delight follows without effort and naturally.'

The soul has three avasthās or conditions which are included in a fourth. They are waking, dreaming, sleeping and what is called Turīya.

'The first condition is that of wakefulness, where the self is conscious of the common world of gross objects. It enjoys gross things. Here the dependence of body is predominant. The second condition is that of dreaming, where the self enjoys subtle things, fashioning for itself a new world of forms from the material of its waking experience. The spirit is said to roam freely unfettered by the bonds of the body. The third condition is that of sound sleep, where we have neither dreams nor desires. It is called suṣupti. In it the soul is said to become temporarily one with Brahman and to enjoy bliss. In deep sleep we are lifted above all desires and freed from the vexations of spirit. . . . The soul is divine in origin, though clogged with the flesh. In sleep it is said to be released from the shackles of the body and to regain its own nature. . . . But this (that is, the eternal dreamless sleep) is likely to be confused with sheer unconsciousness. . . . The highest is not this dreamless sleep, but another, a fourth state of the soul, a pure intuitional consciousness where there is no knowledge of objects internal or external. In deep sleep the spirit dwells in a region far above the changeful life of sense in absolute union with Brahman. The turīya condition brings out the positive aspect of the negative emphasized in the condition of deep sleep.'

Radhakrishnan in Philosophy of the Upanishads.

This fourth condition has been thus described in the Māṇḍūkya Upanishad as follows:

'The fourth, say the wise, is not subjective experience, nor objective experience, nor experience intermediate between the two, nor is it a negative condition which is neither consciousness nor unconsciousness. It is not the knowledge of the senses, nor is it relative knowledge, nor yet inferential knowledge. Beyond the senses, beyond understanding, beyond all expression, is the fourth. It is pure unitary consciousness, wherein all awareness of the world and of multiplicity is completely obliterated. It is the supreme good. It is One without a second. It is the Self. Know it alone!'

'Rāja-Yoga, Samādhi, Unmanī, Manomanī, Immortality, Concentration, Śūnyāśūnya (void and yet non-void), Parama Pāda (the Supreme State), Amanaska (suspended operation of the mind), Advaita (non-duality), Nirālamba (without support), Niraṅjana (pure), Jīvanmukti (emancipated state), Sahajāvasthā (natural state) and Turīyā (literally the Fourth), all mean the same thing. As a grain of salt thrown into water unites and becomes one with it, a like union between the Mind and the Ātman is Samādhi. When Praṇa and Manas (mind) are annihilated (absorbed), the state of harmony then arising is called Samādhi.'

Haṭha Yoga Pradīpikā, *chapter IV, verses 3 to 6.*

There is no āsana like Siddha, no kumbhaka like Kevala, no mudrā like Khecharī, and no laya (absorption of the mind) like Nāda.

(Khecharī Mudrā, literally roaming through space, is described in the Gheranda Saṁhitā *as follows in verses 25 to 28 of the third chapter:*

'Cut the lower tendon of the tongue and move the tongue constantly; rub it with fresh butter, and draw it out (to lengthen it) with an iron instrument. By practising this always, the tongue becomes long and when it reaches the space between the eyebrows, then Khecharī is accomplished. Then (the tongue being lengthened) practise turning it up and back so as to touch the palate, till at

length it reaches the holes of the nostrils opening into the mouth. Close those holes with the tongue (thus stopping inspiration), and gaze on the space between the eyebrows. This is called Khecharī. By this practice there is neither fainting, nor hunger, nor thirst, nor laziness. There comes neither disease, nor decay, nor death. The body becomes divine.')

(Nāda is the inner mystical sound. Verses 79 to 101 of the fourth chapter describes it in great detail with a variety of similes. Yoga is defined as control over the aberrations of the mind. In order to control the mind it is necessary that it should first be absorbed in concentration of some object, then it is gradually withdrawn from that object and made to look within one's own self. This is where the yogi is asked to concentrate upon the inner mystical sounds. 'The mind is like a serpent, forgetting all its unsteadiness by hearing Nāda, it does not run away anywhere.' Gradually as Nāda becomes latent so does the mind along with it. 'The fire, catching the wood, is extinguished along with it (after burning it up); and so the mind also, working with Nāda, becomes latent along with it.')

TECHNIQUE

1. Sit on the floor, with legs stretched straight in front. (Plate 35)

2. Bend the left leg at the knee. Hold the left foot with the hands, place the heel near the perineum and rest the sole of the left foot against the right thigh.

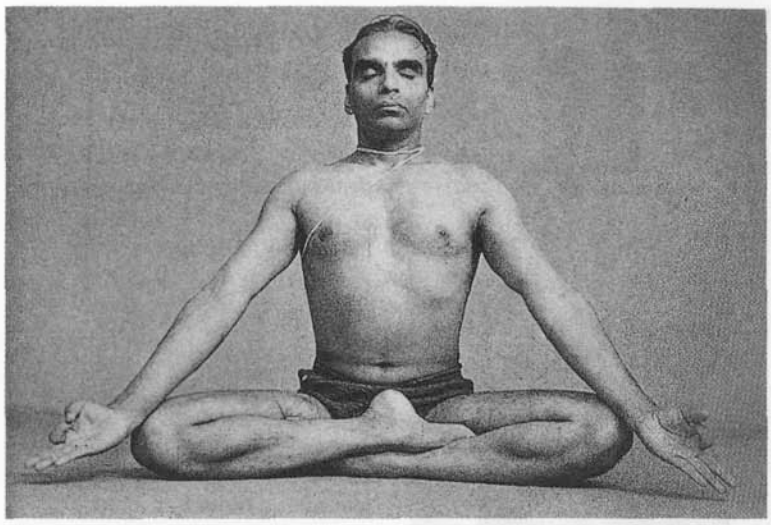

38

3. Now bend the right leg at the knee and place the right foot over the left ankle, keeping the right heel against the pubic bone.

4. Place the sole of the right foot between the thigh and the calf of the left leg.

5. Do not rest the body on the heels.

6. Stretch the arms in front and rest the back of the hands on the knees so that the palms face upwards. Join the thumbs and the forefingers and keep the other fingers extended. (Plate 38).

7. Hold this position as long as you can, keeping the back, neck and head erect and the vision indrawn as if gazing at the tip of the nose.

8. Release the feet and relax for some time. Then repeat the pose for the same length of time, now placing the right heel near the perineum first and then the left foot over the right ankle as described above.

EFFECTS

This posture keeps the pubic region healthy. Like Padmāsana (Plate 53), it is one of the most relaxing of āsanas. The body being in a sitting posture is at rest, while the position of the crossed legs and erect back keeps the mind attentive and alert. This āsana is also recommended for the practice of prāṇāyāma and for meditation.

From the purely physical point of view, the āsana is good for curing stiffness in the knees and ankles. In it the blood circulates in the lumbar region and the abdomen, and this tones the lower region of the spine and the abdominal organs.

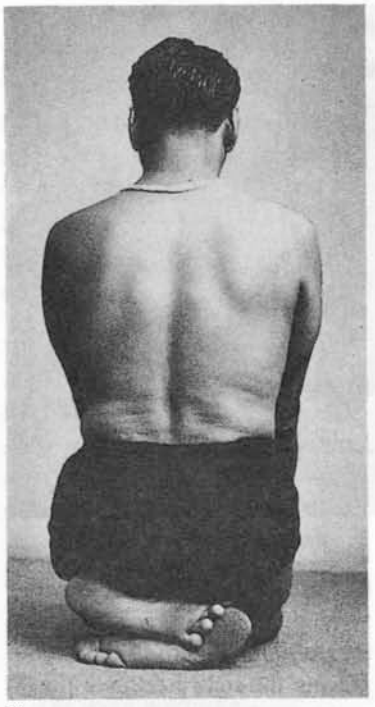

39

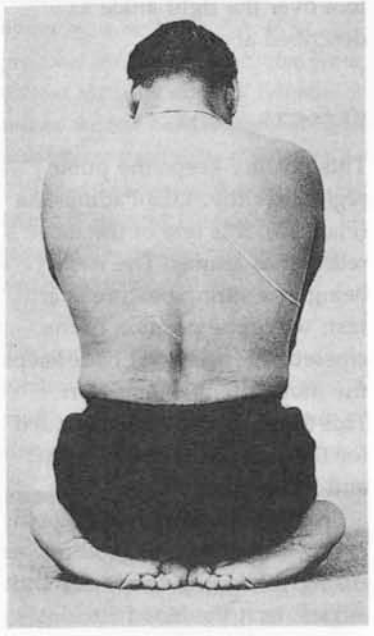

40

22| *Vīrāsana*

ONE*

Vīra means a hero, warrior, champion. This sitting posture is done by keeping the knees together, spreading the feet and resting them by the side of the hips.

The pose is good for meditation and prāṇāyāma.

TECHNIQUE

1. Kneel on the floor. Keep the knees together and spread the feet about 18 inches apart.

2. Rest the buttocks on the floor, but not the body on the feet. The feet are kept by the side of the thighs, the inner side of each calf touching the outer side of its respective thigh. Keep the toes pointing back and touching the floor. Keep the wrists on the knees, palms facing up, and join the tips of the thumbs and forefingers. Keep the other fingers extended. Stretch the back erect. (Back view: Plate 42. Front view: Plate 43)

3. Stay in this position as long as you can, with deep breathing.

4. Now interlock the fingers and stretch the arm straight over the head, palms up. (Plate 44)

5. Stay in this position for a minute with deep breathing.

6. Exhale, release the fingerlock, place the palms on the soles, bend forward and rest the chin on the knees. (Plate 45)

7. Stay in this position for a minute with normal breathing.

8. Inhale, raise the trunk up, bring the feet forward and relax.

9. If you find it difficult to perform the pose as described above, try placing the feet one above the other and rest the buttocks on them. (Plate 39) Gradually move the toes further apart, separate the feet (Plates 40 and 41) and bring them to rest outside the thighs. Then, in time the buttocks will rest properly on the floor and the body will not rest on the feet.

EFFECTS

The pose cures rheumatic pains in the knees and gout, and is also good for flat feet. Due to the stretching of the ankles and the feet, proper arches will be formed. This, however, takes a long time and requires daily practice of the pose for a few minutes for several months. Those suffering from pain in the heels or growth of calcaneal spurs there will get relief and the spurs will gradually disappear.

The pose can even be done immediately after food and will relieve heaviness in the stomach.

41

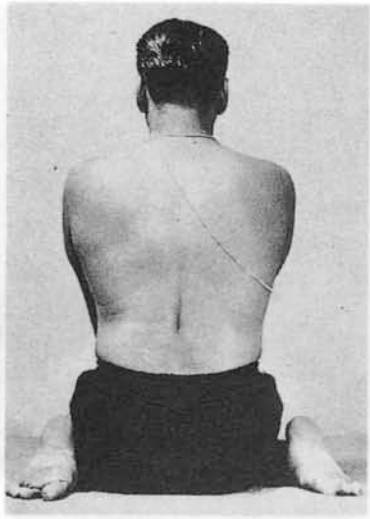

42

43

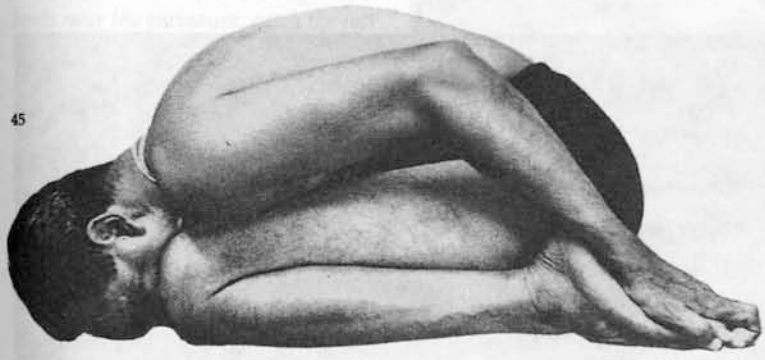

45

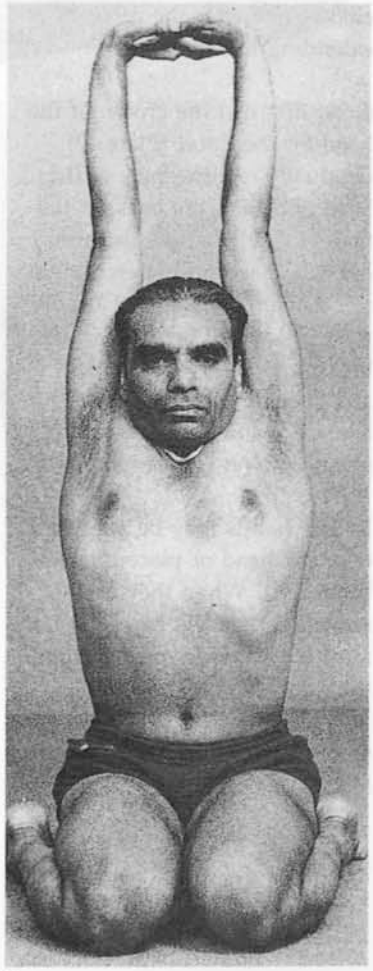

44

23| *Supta Vīrāsana*

TWO*

Supta means lying down. In this āsana one reclines back on the floor and stretches the arms behind the head.

TECHNIQUE

1. Sit in Vīrāsana. (Plate 43)

2. Exhale, recline the trunk back and rest the elbows one by one on the floor. (Plate 46)

3. Relieve the pressure on the elbows one after the other by extending the arms.

4. At first rest the crown of the head on the floor. (Plate 47) Gradually rest the back of the head and then the back on the floor. (Plate 48) Take the arms over the head and stretch them out straight. (Plate 49) Hold this pose as long as you can while breathing deeply. Then place the arms beside the trunk, press the elbows to the floor and sit up again with an exhalation.

5. The hands may be stretched over the head or placed beside the thighs. When they are stretched over the head, do not raise the shoulder-blades from the floor.

6. Beginners may keep the knees apart.

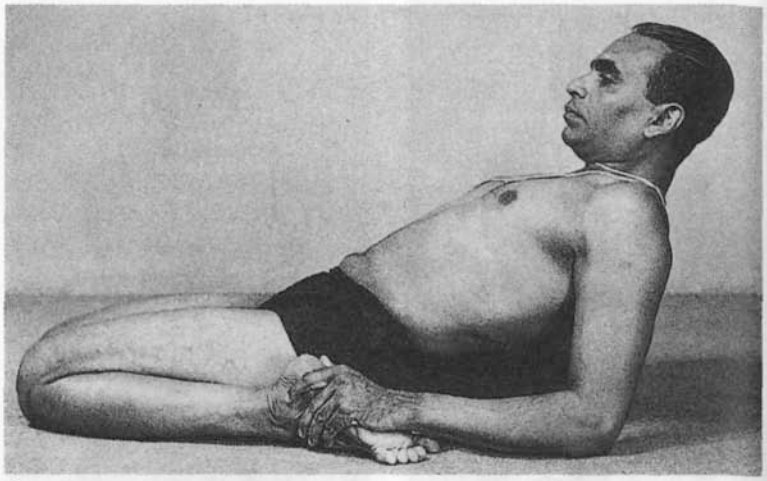

46

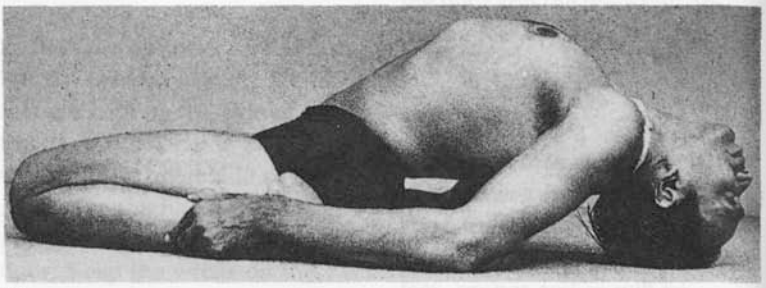

47

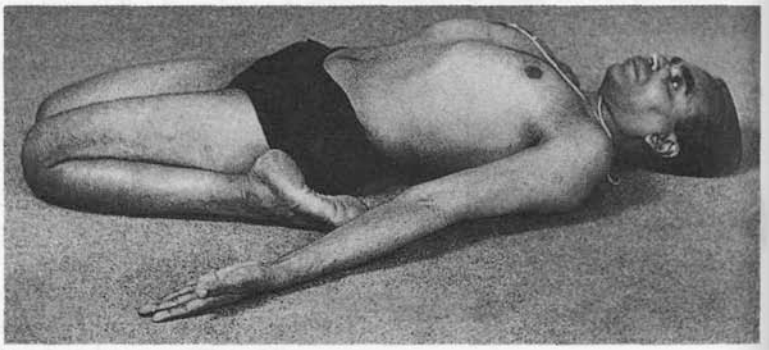

48

49

EFFECTS

This āsana stretches the abdominal organs and the pelvic region. People whose legs ache will get relief from holding this pose for 10 to 15 minutes and it is recommended to athletes and all who have to walk or stand about for long hours. It can be done after meals and if before retiring at night the legs feel rested next morning. Several of my pupils who were cadets at the National Defence Academy after long route marches found great relief by combining this āsana with Sarvāngāsana I. (Plate 102)

24| *Baddha Koṇāsana*

THREE*

Baddha means caught, restrained. Koṇa means an angle. In this posture, sit on the floor, bring the heels near the perineum, catch the feet and widen the thighs until the knees touch the floor on either side. This is how Indian cobblers sit.

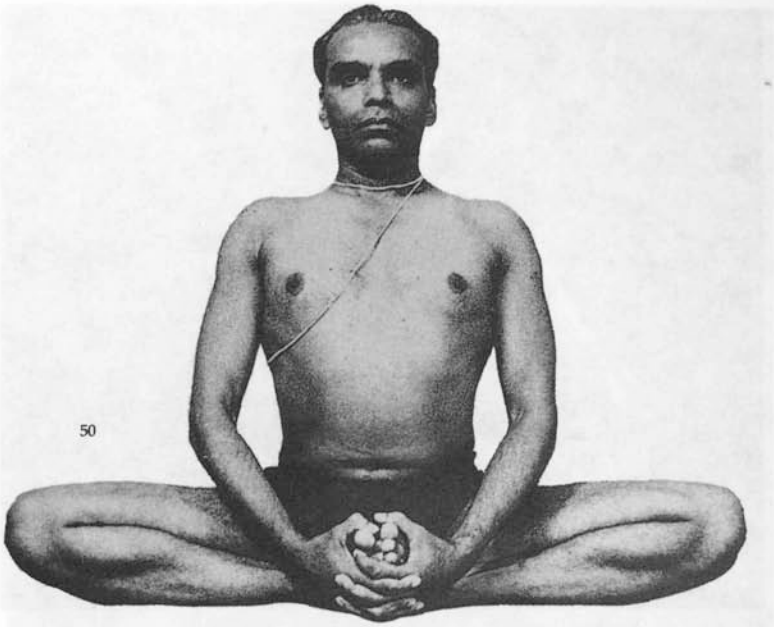

50

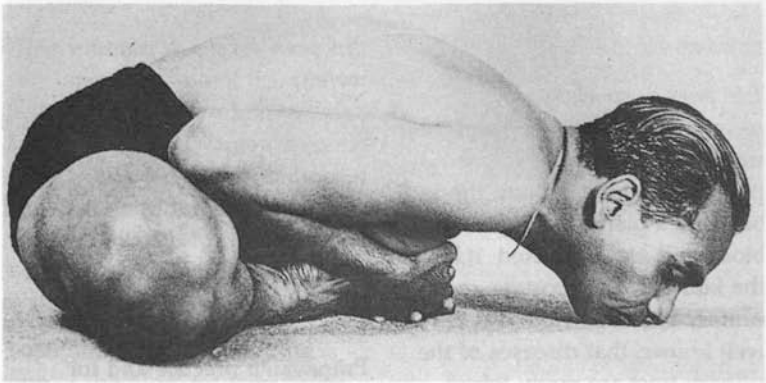

51

TECHNIQUE

1. Sit on the floor with the legs stretched straight in front. (Plate 35)

2. Bend the knees and bring the feet closer to the trunk.

3. Bring the soles and heels of the feet together and catching the feet near the toes, bring the heels near the perineum. The outer sides of both feet should rest on the floor, and the back of the heels should touch the perineum.

4. Widen the thighs and lower the knees until they touch the floor.

5. Interlock the fingers of the hands, grip the feet firmly, stretch the spine erect and gaze straight ahead or at the tip of the nose. (Plate 50) Hold the pose as long as you can.

6. Place the elbows on the thighs and press them down. Exhale, bend forward, rest the head, then the nose and lastly the chin on the floor. (Plate 51) Hold this position from half a minute to a minute with normal breathing.

7. Inhale, raise the trunk from the floor and come back to position 5. (Plate 50)

8. Then release the feet, straighten the legs and relax.

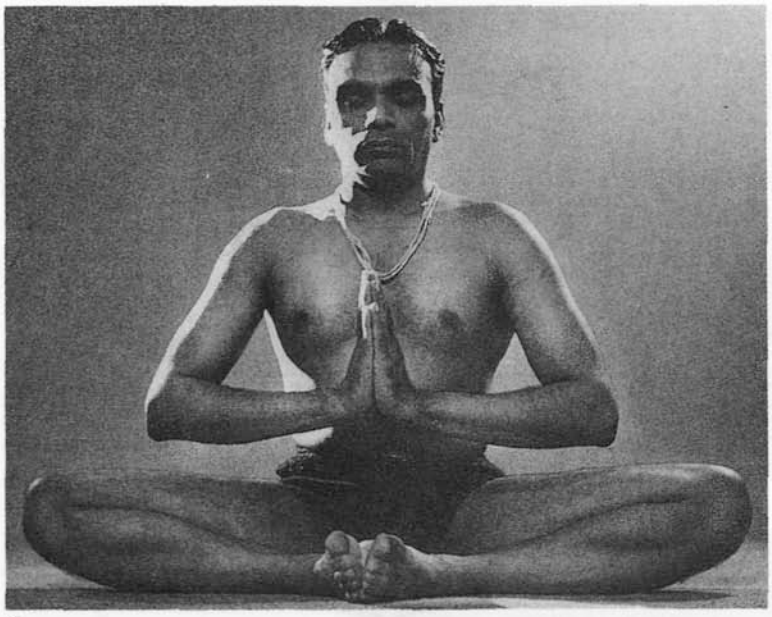

52

25 *Padmasana*

FOUR*

Padma means a lotus. This is the lotus posture, one of the most important and useful āsanas. It is the posture for meditation and the Buddha is often depicted in it.

Verse 48 of the first chapter of the Hatha Yoga Pradīpikā *describes the posture and the practice of breath control while seated in it thus:*

'Assuming Padmasana and having placed the palms one upon another, fix the chin firmly upon the breast and contemplating upon Brahman, frequently contract the anus and raise the apana up; by similar contraction of the throat force the prana down. By this he obtains unequalled knowledge through the favour of Kundalinī (which is roused by this process).'

Kundalinī is the Divine Cosmic Energy in bodies. It is symbolized by a coiled and sleeping serpent in the lowest bodily centre at the base of the spinal column. This latent energy has to be awakened and made to go up the spine to the brain through Suṣumnā Naḍi, a channel through which nervous energy passes, and through the six chakras, the subtle centres in the body, the fly-wheels in the nervous system of the human machine. The awakening of Kundalinī is discussed in detail in Arthur Avalon's (Sir John Woodroffe's) book entitled The Serpent Power.

This is one of the basic postures and is often used in the variations of Śīrṣāsana and Sarvangasana.

EFFECTS

The pose is specially recommended for those suffering from urinary disorders. The pelvis, the abdomen and the back get a plentiful supply of blood and are stimulated. It keeps the kidneys, the prostate and the urinary bladder healthy. It is well known that diseases of the urinary tract are rarely found among the Indian cobblers and the reason for that is that they sit all day in this pose.

It relieves sciatic pain and prevents hernia. If practised regularly, it relieves pain and heaviness in the testicles.

The pose is a blessing to women. Coupled with Sarvāngāsana I (Plate 102) and its cycle (Plates 113 to 125) it checks irregular menstrual periods and helps the ovaries to function properly. It is found that pregnant women who sit daily in this pose for a few minutes will have much less pain during delivery and will be free from varicose veins. (It is recommended for pregnant women in Dr Grantly Dick Reed's book *Childbirth Without Fear*.)

Along with Padmāsana (Plate 53) and Vīrasana (Plate 43) this āsana is recommended for Prāṇāyāma practice and for meditation. When sitting in meditation in this pose the palms should be folded in front of the chest (Plate 52), but to do this with the back erect requires practice. This asana can be done without fear even after meals as long as the head is not rested on the floor.

TECHNIQUE

1. Sit on the floor with the legs straight. (Plate 35)

2. Bend the right leg at the knee, hold the right foot with the hands and place it at the root of the left thigh so that the right heel is near the navel.

3. Now bend the left leg, and holding the left foot with the hands place it over the right at the root, the heel being near the navel. The soles of the feet should be turned up. This is the basic Padmāsana pose. (Plate 53)

4. People not used to sitting on the floor seldom have flexible knees. At the start they will feel excruciating pain around the knees. By perseverance and continued practice the pain will gradually subside and they can then stay in the pose comfortably for a long time.

5. From the base to the neck the spine should remain erect. The arms may be stretched out, the right hand being placed on the right knee and the left hand on the left knee. The forefingers and the thumbs are bent and touch each other.

6. Change the leg position by placing the left foot over the right thigh and the right foot over the left thigh. This will develop the legs evenly.

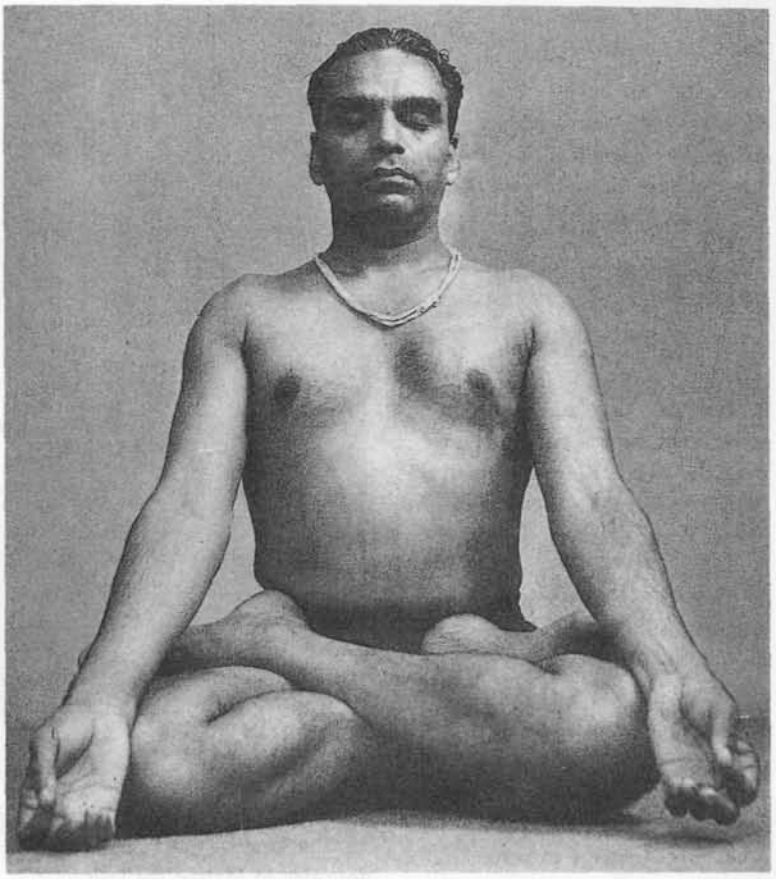

53

EFFECTS

After the initial knee pains have been overcome, Padmāsana is one of the most relaxing poses. The body being in a sitting posture, it is at rest without being sloppy. The position of the crossed legs and the erect back keeps the mind attentive and alert. Hence it is one of the āsanas recommended for practising prāṇāyāma (breath control).

On the purely physical level, the pose is good for curing stiffness in the knees and ankles. Since the blood is made to circulate in the lumbar region and the abdomen, the spine and the abdominal organs are toned.

26| *Parvatāsana*

FOUR*

Parvata means a mountain. In this variation of Padmāsana the arms are stretched over the head with the fingers interlocked.

TECHNIQUE

1. Sit in Padmāsana. (Plate 53)

2. Interlock the fingers, and stretch the hands vertically up over the head. Keep the head bent forward with the chin on the breast bone.

3. Stretch the arms up from the latissimus dorsi (near the floating ribs at the back) and the shoulder-blades. The palms should face upwards. (Plate 54)

4. Hold the pose for a minute or two with deep and even breathing. Change the crossing of the legs and the interlock of the fingers and repeat the pose, keeping the back erect.

EFFECTS

The āsana relieves rheumatic pains and stiffness in the shoulders. It helps draw free movement and to develop the chest. The abdominal organs are drawn in and the chest expands fully.

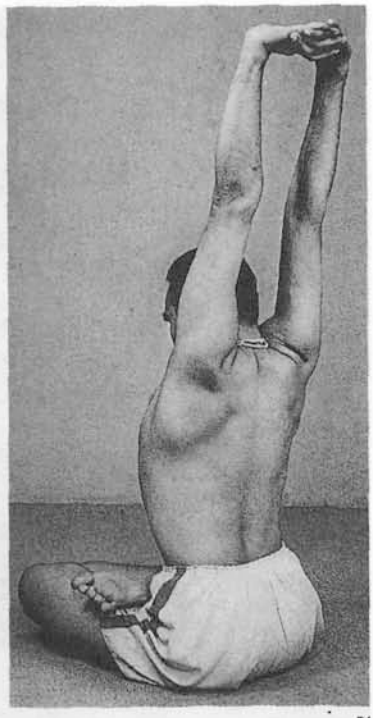

54

27| *Matsyāsana*

FIVE*

Matsya means a fish. This posture is dedicated to Matsya the Fish Incarnation of Viṣṇu, the source and maintainer of the universe and of all things. It is related that once upon a time the whole earth had become corrupt and was about to be overwhelmed by a universal flood. Viṣṇu took the form of a fish and warned Manu (the Hindu Adam) of the impending disaster. The fish then carried Manu, his family and the seven great sages in a ship, fastened to a horn on his head. It also saved the Vedas from the flood.

TECHNIQUE

1. Sit in Padmāsana. (Plate 53)

2. Lie flat on the back with the legs on the floor.

3. Exhale, arch the back by lifting the neck and the chest, take the head back and rest the crown on the floor. Drag the head further back by holding the crossed legs with the hands and increase the back arch. (Plate 55)

4. Now take the hands from the legs, bend the arms, hold the elbows with the hands and rest the forearms on the floor behind the head. (Plate 56)

5. Stay in this position from 30 to 60 seconds while breathing deeply.

6. Rest the back of the head on the floor, lie flat on the back, inhale and then come up to Padmāsana, release the legs and relax.

7. Recross the legs the other way and repeat the pose for the same length of time.

8. If positions 3 and 4 are difficult to achieve, lie flat on the back with the arms stretched straight over the head. (Plate 57)

EFFECTS

The dorsal region is fully extended in this posture and the chest is well expanded. Breathing becomes fuller. The thyroids benefit from the exercise due to the stretching of the neck. The pelvic joints become elastic. The āsana relieves inflamed and bleeding piles.

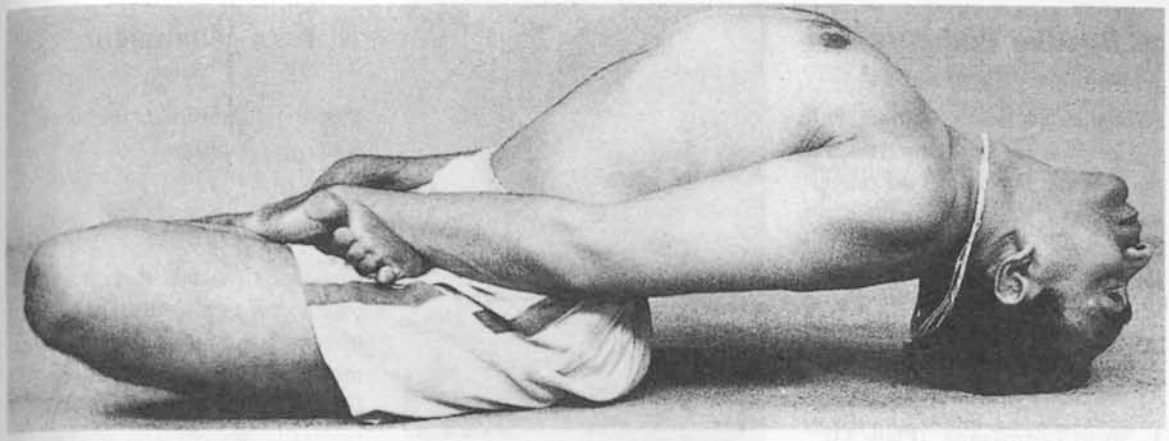

55

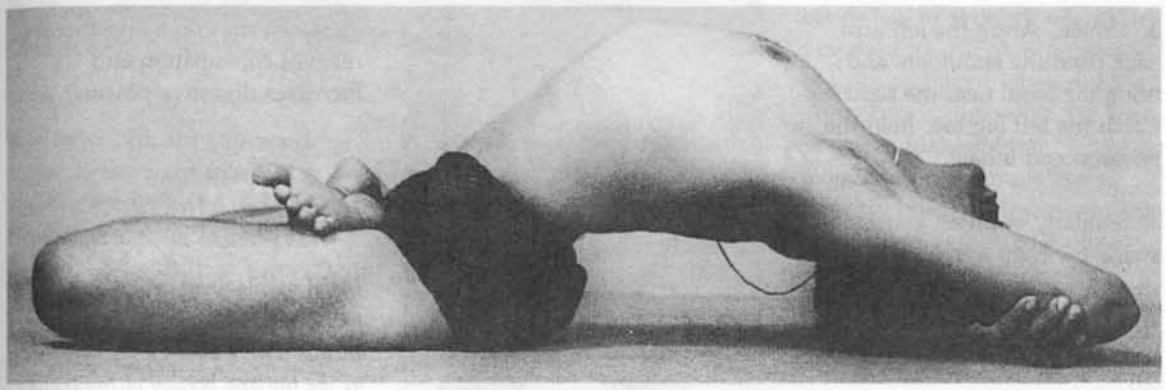

56

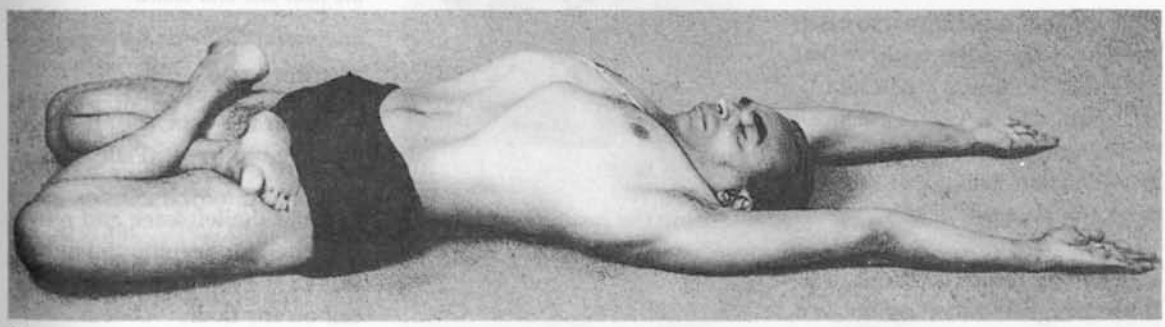

57

28| *Baddha Padmāsana*

SIX*

Baddha means caught, restrained. In this position the hands are crossed at the back and the big toes are caught from behind. The body is caught between the crossed legs in front and the crossed hands behind, hence the name.

TECHNIQUE

1. Sit in Padmāsana. (Plate 53)

2. Exhale, swing the left arm back from the shoulders and bring the hand near the right hip. Catch the left big toe, hold the position and inhale.

3. Similarly, with an exhalation, swing the right arm back from the shoulder, bring it near the left hip and catch the right big toe. (Front view: Plate 58. Back view: Plate 59)

4. If the toes are difficult to catch stretch the shoulders back, so that the shoulder-blades are brought near each other. A little practice in swinging the arms back with an exhalation will enable one to catch the big toes.

5. If the right foot is placed over the left thigh and then the left foot over the right thigh, catch the left big toe first and then the right big toe. If, on the other hand, the left foot is placed over the right thigh first and then the right foot over the left thigh, catch the right big toe first and then the left big toe. Catch first the big toe of the foot which is uppermost.

58

59

6. Throw the head as far back as possible and take a few deep breaths.

7. Inhale deeply, and then with an exhalation bend the trunk forward from the hips and rest the head on the floor, without releasing the toes from the hand grip. Bending the head forward in Baddha Padmāsana (Plate 58) and touching it on the floor is called:

29| *Yoga Mudrāsana*

SIX*

This āsana is especially useful in awakening Kuṇḍalinī.

EFFECTS

Crossing the hands behind the back expands the chest and increases the range of shoulder movement. Yoga Mudrāsana (Plate 60) intensifies the peristaltic activity and pushes down the accumulated waste matter in the colon and thereby relieves constipation and increases digestive power.

30| *Mahā Mudrā*

FIVE*

Mahā means great or noble. Mudrā means shutting, closing or sealing. In this sitting posture the apertures at the top and bottom of the trunk are held fast and sealed.

TECHNIQUE

1. Sit on the floor with the legs stretched in front. (Plate 35)

2. Bend the left knee and move it to the left, keeping the outer side of the left thigh and the left calf on the floor.

3. Place the left heel against the inner side of the left thigh near the perineum. The big toe of the left foot should touch the inner side of the right thigh. The angle between the extended right leg and the bent left leg should be a right angle of 90 degrees.

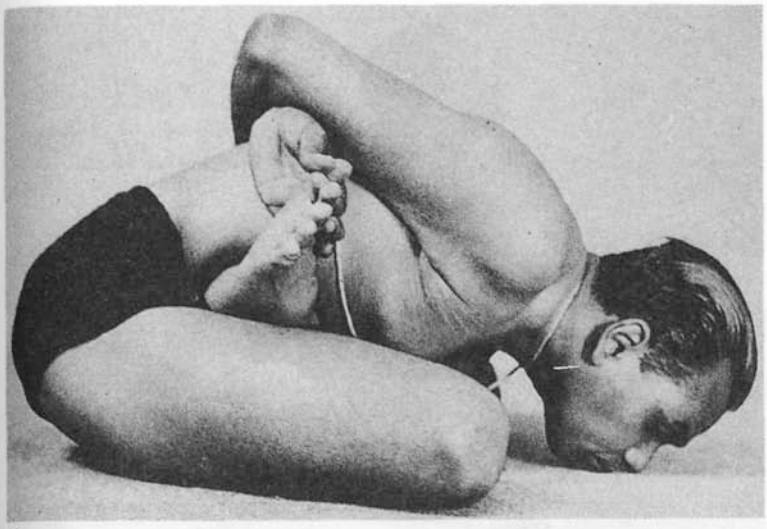

60

4. Stretch the arms forward towards the right foot and hook the big toe with the thumbs and forefingers.

5. Lower the head to the trunk until the chin rests in the hollow between the collar bones just above the breast-bone.

6. Keep the spine fully stretched and do not allow the right leg to tilt to the right.

7. Inhale completely. Tighten the entire abdomen from the anus to the diaphragm. Pull the abdomen back towards the spine and also up towards the diaphragm.

8. Relax the abdominal tension, then exhale, again inhale and hold the breath, maintaining the abdominal grip. Hold this posture as stated above from one to three minutes. (Plate 61)

9. Relax the abdominal tension, exhale, raise the head, release the hands and straighten the bent leg.

10. Repeat on the other side, keeping the left leg straight and the right one bent for an equal length of time.

EFFECTS

This āsana tones the abdominal organs, the kidneys and adrenal glands. Women suffering from a prolapsed womb find relief as it pulls the womb up to its original position. Persons suffering from spleen ailments and from enlargement of the prostate gland will benefit by staying in this pose longer. It cures indigestion.

'This Mahāmudrā destroys death and many other pains. 'There is nothing that one cannot eat or has to avoid (if one has practised it). All food regardless of taste and even when deadly poisonous is digested.' 'He who practices Mahāmudrā, overcomes consumption, leprosy, piles, enlargement of the spleen, indigestion and other complaints of long duration.' (*Haṭha Yoga Pradīpikā*, chapter 3, verses 14, 16 and 17.)

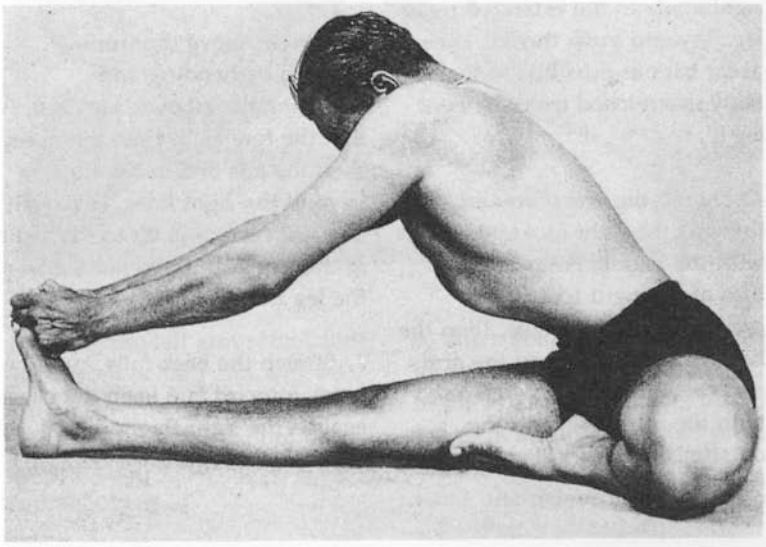

61

31| *Jānu Śīrṣāsana*

FIVE*

Jānu means the knee. Śīrṣa is the head. In this posture sit with one leg stretched out on the ground and the other bent at the knee. Then catch the extended foot with both the hands and place the head on that knee.

TECHNIQUE

1. Sit on the floor, with legs stretched straight in front. (Plate 35)

2. Bend the left knee and move it to the left, keeping the outer side of left thigh and the left calf on the floor.

3. Place the left heel against the inner side of the left thigh near the perineum. The big toe of the left foot should touch the inner side of the right thigh. The angle between the two legs should be obtuse. Do not keep the left knee in line with the left thigh at a right angle to the extended right leg. Try and push the left knee as far back as possible, so that the body is stretched from the bent leg.

4. Extend the arms forward towards the right foot and hold it with the hands. First catch the toes of the right foot, then gradually catch the sole, then the heel and finally extend the arms and catch the wrist of one hand with the other, beyond the outstretched foot. (Plate 62)

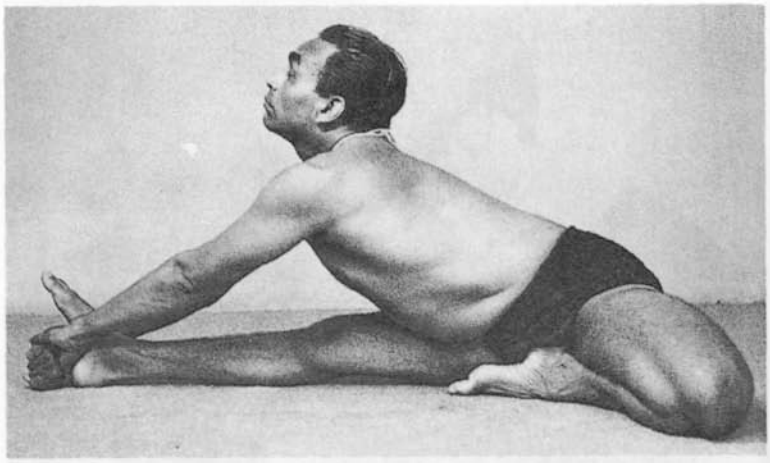

62

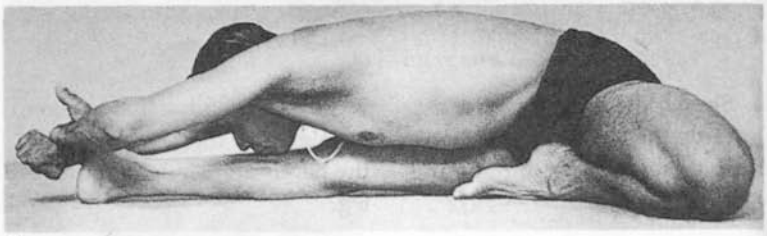

63

5. Keep the right leg stretched throughout by tightening the knee. See that the back of the right knee rests on the floor.

6. Exhale, move the trunk forward by bending and widening the elbows, and rest first the forehead, then the nose, then the lips and lastly the chin beyond the right knee. (Plate 63) The right foot will tilt to the right in the beginning. Do not allow the leg to tilt.

7. Stretch the back fully, pull the trunk forward and keep the chest against the right thigh.

8. Stay in this position with deep breathing from half a minute to a minute. One can also do the pose holding the breath after each exhalation.

9. Inhale, raise the head and trunk, straighten the arms and gaze up for a few seconds, extending the spine and trying to make it concave. (Plate 62)

10. Release the hand grip on the right foot, straighten the left leg and come back to position 1.

11. Repeat the pose keeping the left leg stretched out and bending the right leg at the knee. Stay in the pose for the same length of time on both the sides.

EFFECTS

This āsana tones the liver and the spleen and thereby aids digestion. It also tones and activates the kidneys, the effect on which can be felt while one is performing the pose as explained above.

Persons suffering from enlargement of the prostate gland will benefit by staying longer in this pose. They should practice this āsana along with Sarvāngāsana. (Plate 102)

The pose is also recommended for people suffering from low fever for a long time.

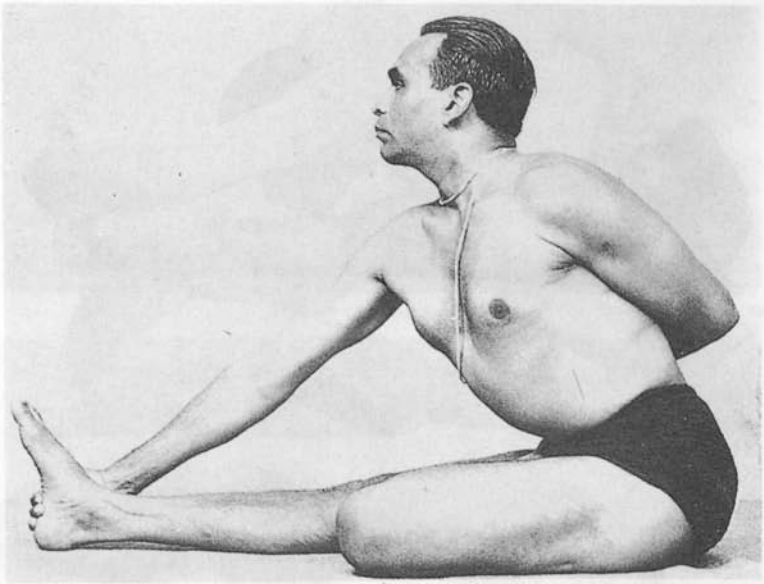

64

32| *Ardha Baddha Padma Paschimottānāsana*

EIGHT*

Ardha means half, baddha means caught, restrained and padma a lotus. Paschimottānāsana (Plate 81) is the posture where the back of the whole body is intensely stretched.

TECHNIQUE

1. Sit on the floor, with the legs stretched straight in front. (Plate 35)

2. Bend the left leg at the knee, and place the left foot over the right thigh. The left heel should press the navel and the toes should be stretched and pointing. This is the half lotus posture.

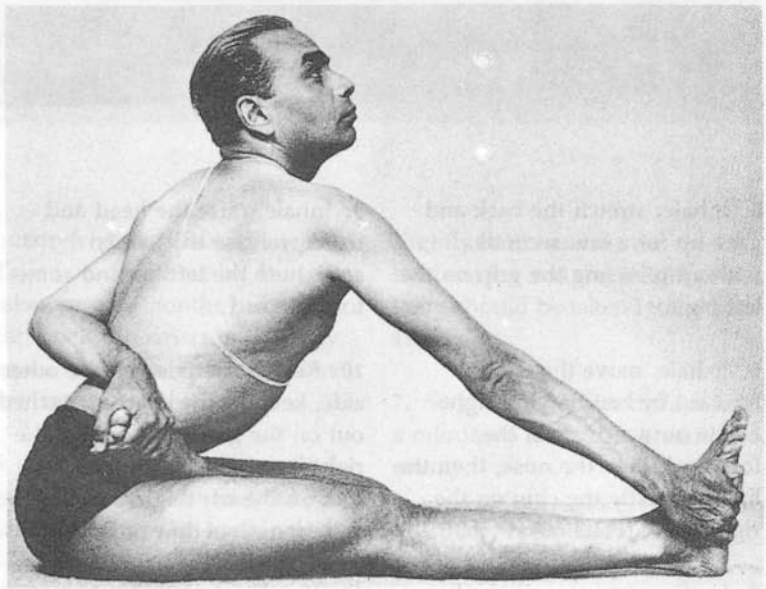

65

3. Bring the left arm round the back from behind and with an exhalation catch the big toe of the left foot. If the toe cannot be grasped easily, swing back the left shoulder.

4. After holding the left big toe, move the bent left knee nearer to the extended right leg. Stretch the right arm forward and catch the right foot with the right hand, the palm touching the sole. (Plates 64 and 65)

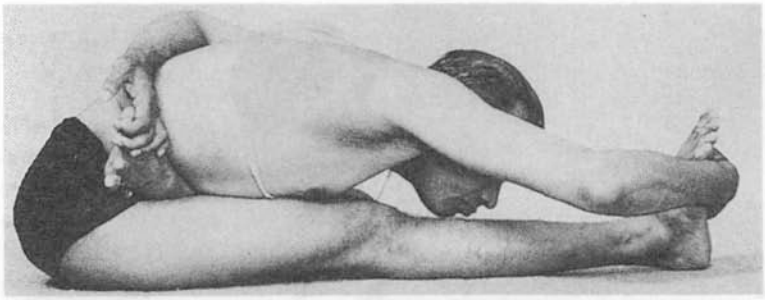

66

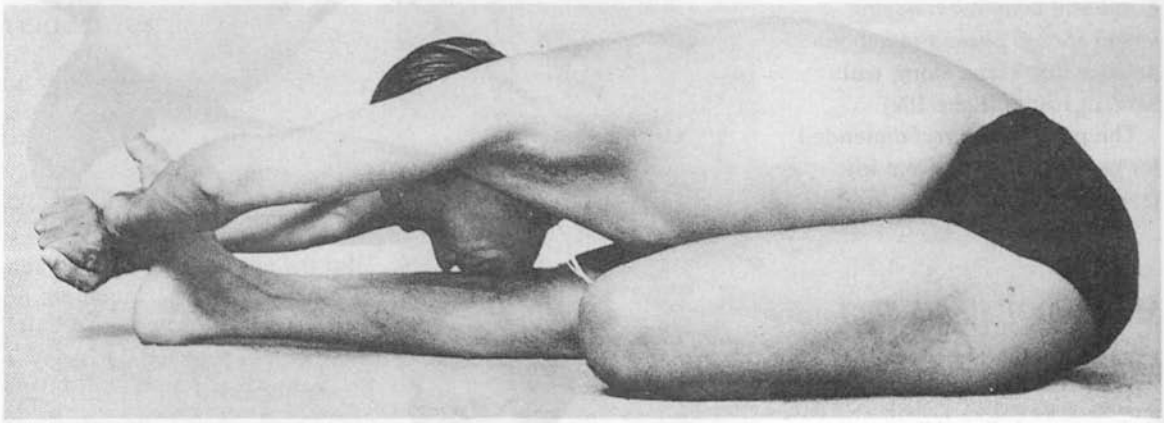

67

5. Inhale, stretch the back and gaze up for a few seconds, without releasing the grip on the left big toe.

6. Exhale, move the trunk forward by bending the right elbow outwards. Rest the forehead, then the nose, then the lips and lastly the chin on the right knee. (Plate 66)

7. In the initial stages, the knee of the extended leg will be lifted off the floor. Tighten the thigh muscles and rest the entire back of the extended right leg on the floor.

8. Stay in this position from 30 to 60 seconds, breathing evenly.

9. Inhale, raise the head and trunk, release the hands, straighten the left leg and come to position 1.

10. Repeat the pose on the other side, keeping the left leg stretched out on the ground, bending the right knee and placing the right foot on the left thigh. Stay for the same length of time on both sides.

11. If you cannot hold the toe with the hand from behind, hold the extended leg with both hands and follow the above techniques. (Plate 67)

EFFECTS

Due to the half lotus pose, the knees become flexible enough to execute the full lotus pose. While placing the chin on the knee of the extended leg, the bent knee is brought close to the stretched leg. This gives a good pull to the navel and abdominal organs. Blood is made to flow round the navel and the genital organs. The navel is considered to be a nerve centre, and the Svādhiṣṭhāna Chakra, one of the purificatory fly-wheels in the human nervous system, is situated there. This chakra corresponds to the hypo-gastric plexus. The pose is recommended for persons with rounded and drooping shoulders.

33| *Triaṅgā Mukhaikapda Paschimottānāsana*

FIVE*

Trianga means three limbs or parts thereof. In this posture the three parts are the feet, knees and buttocks. Mukhaikapāda (a compound of three words, mukha = face, eka = one, and pāda = leg or foot) corresponds to the face (or mouth) touching one (extended) leg. In Paschimottānāsana (Plate 81) the back of the whole body is intensely stretched.

68

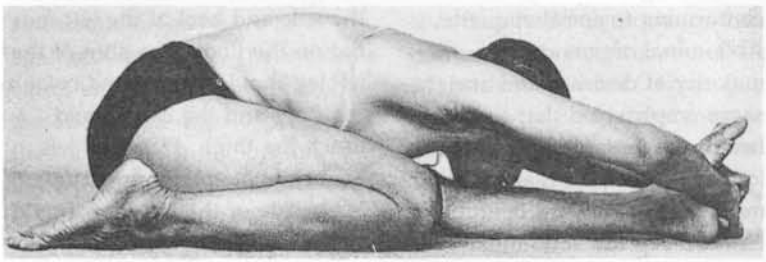

69

TECHNIQUE

1. Sit on the floor, with the legs stretched straight in front. (Plate 35)

2. Bend the right leg at the knee and move the right foot back. Place the right foot by the side of the right hip joint, keep the toes pointing back and rest them on the floor. The inner side of the right calf will touch the outer side of the right thigh.

3. Balance in this position, throwing the weight of the body on the bent knee. In the beginning, the body tilts to the side of the outstretched leg, and the foot of the outstretched leg also tilts outwards. Learn to balance in this position, keeping the foot and toes stretched and pointing forward.

4. Now hold the left foot with both the palms, gripping the sides of the sole. If you can, then extend the trunk forward and hook the wrists round the outstretched left foot. (Plate 68) Take two deep breaths. It usually takes several months before one can hook the wrists in this way, so do not despair after the first few attempts.

5. Join the knees, exhale and bend forward. Rest first the forehead, then the nose, next the lips and ultimately the chin on the left knee. (Plate 69) To achieve this, widen the elbows and push the trunk forward with an exhalation.

6. Do not rest the left elbow on the floor. In the beginning, one loses balance and topples over to the side of the extended leg. The trunk should, therefore, be slightly bent towards the side of the bent leg and the weight of the body should be taken by the bent knee.

7. Stay in this position from half a minute to a minute, breathing evenly.

8. Inhale, raise the head and trunk, release the hands, straighten the right leg and come to position 1.

9. Repeat the pose on the other side, keeping the right leg stretched out on the ground, bending the left knee and placing the left foot by the left hip joint. Stay for the same length of time on both sides.

EFFECTS

This āsana is recommended for persons suffering from dropped arches and flat feet. It cures sprains in the ankle and the knee, and any swelling in the leg is reduced.

Along with Jānu Śīrṣāsana (Plate 63) and Ardha Baddha Padma Paschimottānāsana (Plate 66), this āsana tones the abdominal organs and keeps them free from sluggishness. We abuse our abdominal organs by over-indulgence or by conforming to social etiquette. Abdominal organs cause a majority of diseases and ancient sages emphasized that their health was essential for longevity, happiness and peace of mind. These forward bending āsanas keep the abdominal organs healthy and in trim. Apart from keeping the muscles in shape, they work on the organs as well.

34| *Marīchyāsana I*

FIVE*

This āsana is dedicated to the sage Marīchi, son of the Creator, Brahmā. Marīchi was the grandfather of Sūrya (the Sun God).

TECHNIQUE

1. Sit on the floor with the legs stretched straight in front. (Plate 35)

2. Bend the left knee and place the sole and heel of the left foot flat on the floor. The shin of the left leg should be perpendicular to the floor and the calf should touch the thigh. Place the left heel near the perineum. The inner side of the left foot should touch the inner side of the outstretched right thigh.

3. Stretch the left shoulder forward till the left armpit touches the perpendicular left shin. Turn the left arm round the left shin and thigh, bend the left elbow and throw the left forearm behind the back at the level of the waist. Then move the right hand behind the back and clasp the left hand with the right at the wrist or vice versa. If that is not possible then clasp the palms or the fingers. (Plate 70)

4. Now, turn the spine to the left, keeping the outstretched right leg straight. Remain in this position gazing at the outstretched right big toe and take a few deep breaths.

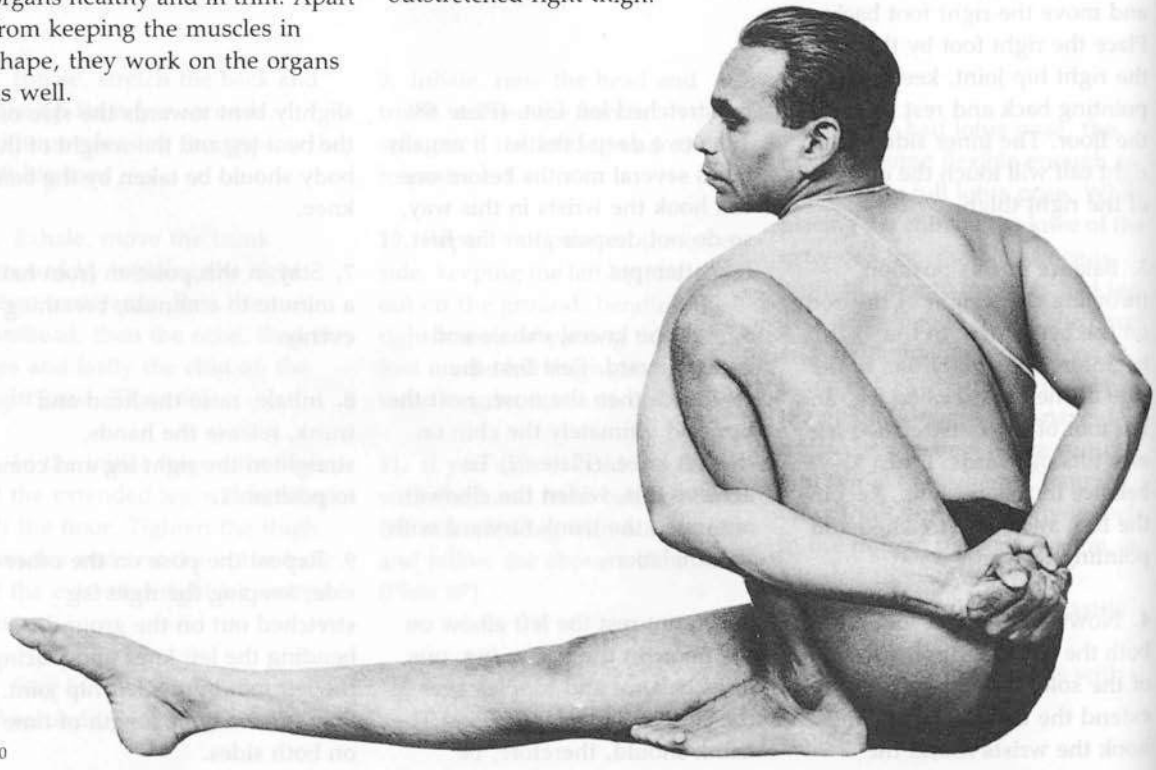

5. Exhale, and bend forward. Rest the forehead, then the nose, next the lips and lastly the chin on the right knee. (Plate 71) While in this position, keep both shoulders parallel to the floor and breathe normally. Stay in the pose for about 30 seconds and see that the back of the entire extended leg rests on the floor throughout.

6. Inhale, raise the head from the right knee (Plate 70), release the hands, straighten the left leg and come to position 1.

7. Repeat the pose on the other side for an equal length of time.

EFFECTS

The fingers gain in strength by the practice of this āsana. In the preceding āsanas (namely, Jānu Śīrṣāsana (Plate 63), Ardha Baddha Padma Paschimottānāsana (Plate 66) and Triang Mukhaikapāda Paschimottānāsana (Plate 69) the abdominal organs are made to contract by gripping a leg with the hands. In this pose the hands do not hold the legs. To bend forward and to rest the chin on the knee of the extended leg the abdominal organs have to contract vigorously. This creates a better circulation of blood round the abdominal organs and keeps them healthy. In the beginning it is very difficult to bend forward at all after gripping both hands behind the back, but it comes with practice. The dorsal region of the spine is also exercised in this pose.

NOTE

The four poses, Jānu Śīrṣāsana, Ardha Baddha Padma Paschimottānāsana, Triang Mukhaikapāda Paschimottānāsana and Marīchyāsana I, are preparatory poses for the correct Paschimottānāsana. (Plate 81) It is difficult for many to get a good grip on the feet in Paschimottānāsana even after several attempts. These four āsanas give one sufficient elasticity in the back and legs so that one gradually achieves the correct Paschimottānāsana (Plate 81) as described later. Once this is done with ease, these four āsanas can be practised once or twice a week instead of daily.

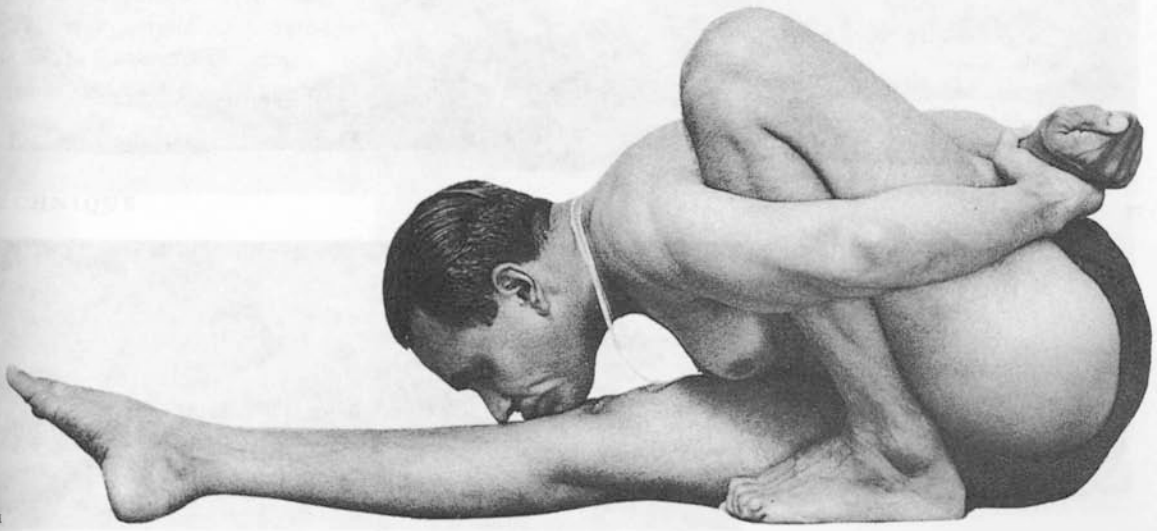

71

35| *Upaviṣṭha Koṇāsana*

NINE*

Upaviṣṭha means seated. Koṇa means an angle.

TECHNIQUE

1. Sit on the floor with the legs stretched straight in front. (Plate 35)

2. Move the legs sideways one by one and widen the distance between them as far as you can. Keep the legs extended throughout and see that the back of the entire legs rests on the floor.

3. Catch the big toes between the respective thumbs and index and middle fingers.

4. Keep the spine erect and extend the ribs. Pull the diaphragm up and hold the pose for a few seconds with deep breaths. (Plate 72)

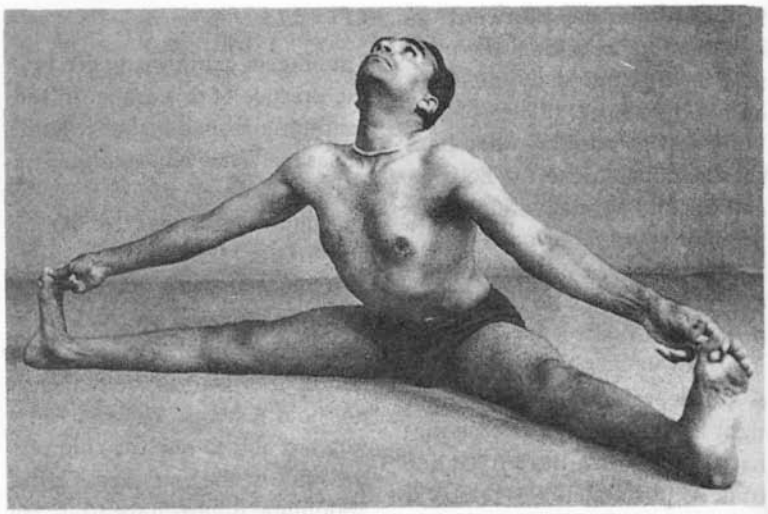

72

5. Now clasp the feet with the hands. Exhale, bend forward and rest the head on the floor. (Plate 73) Then extend the neck and place the chin on the floor.

6. Then, try to rest the chest on the floor. (Plate 74) Stay in this position from 30 to 60 seconds with normal breathing.

7. Inhale, raise the trunk off the floor (Plate 72) and release the hold on the feet, bring them together and relax.

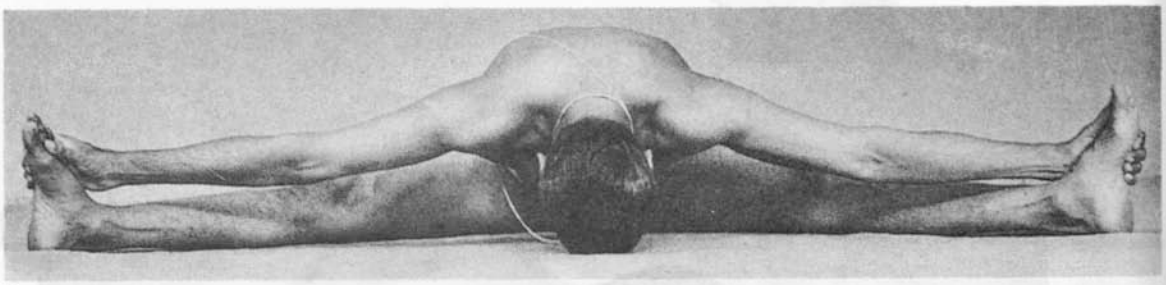

73

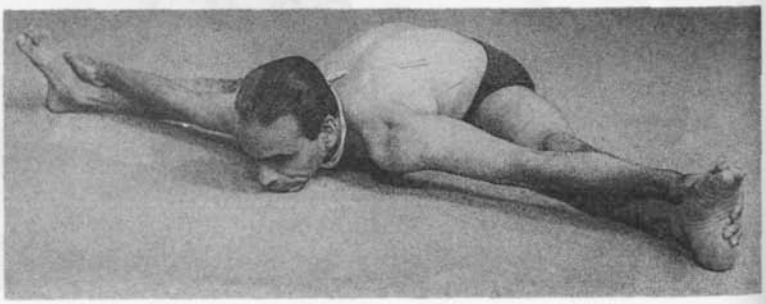

74

EFFECTS

The āsana stretches the hamstrings and helps the blood to circulate properly in the pelvic region and keeps it healthy. It prevents the development of hernia of which it can cure mild cases and relieves sciatic pains. Since the āsana controls and regularizes the menstrual flow and also stimulates the ovaries, it is a boon to women.

36| *Paschimottānāsana (also called Ugrāsana or Brahmacharyāsana)*

SIX*

Paschima literally means the west. It implies the back of the whole body from the head to the heels. The anterior or eastern aspect is the front of the body from the face down to the toes. The crown of the head is the upper or northern aspect while the soles and heels of the feet form the lower or southern aspect of the body. In this āsana the back of the whole body is intensely stretched, hence the name.

Ugra means formidable, powerful and noble. Brahmacharya means religious study, self-restraint and celibacy.

TECHNIQUE

1. Sit on the floor with the legs stretched straight in front. Place the palms on the floor by the side of the hips. Take a few deep breaths. (Plate 35)

2. Exhale, extend the hands and catch the toes. Hold the right big toe between the right thumb and the index and middle fingers and likewise the left big toe. (Plate 75)

3. Extend the spine and try to keep the back concave. To start with the back will be like a hump. This is due to stretching the spine only from the area of the shoulders. Learn to bend right from the pelvic region of the back and also to extend the arms from the shoulders. Then the hump will disappear and the back will become flat as in Plate 75. Take a few deep breaths.

4. Now exhale, bend and widen the elbows, using them as levers, pull the trunk forward and touch the forehead to the knees. (Plate 76) Gradually rest the elbows on the floor, stretch the neck and trunk, touch the knees with the nose and then with the lips. (Plate 77)

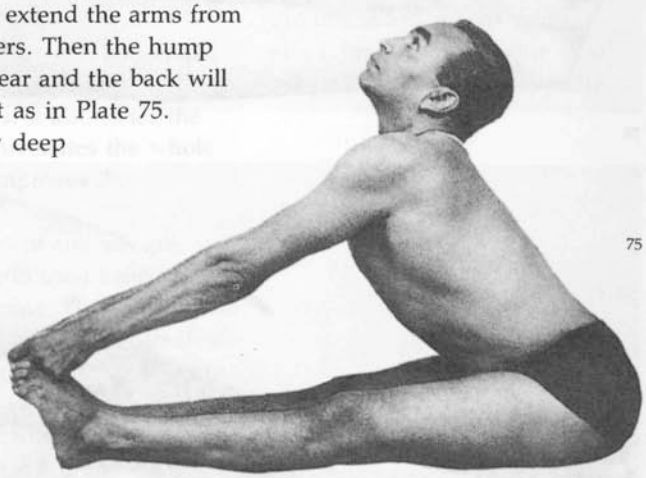

75

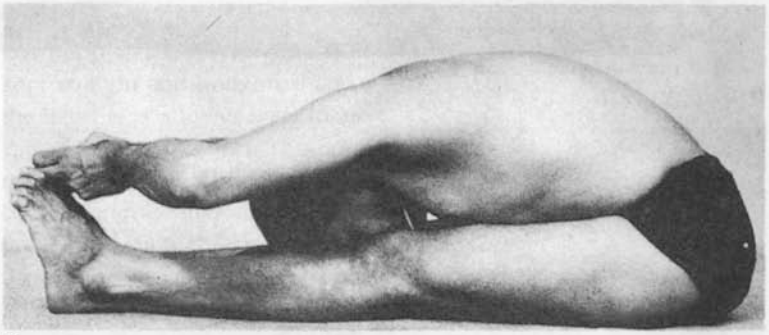

76

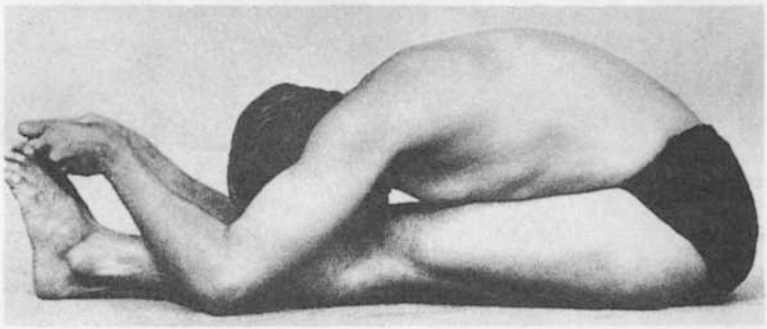

77

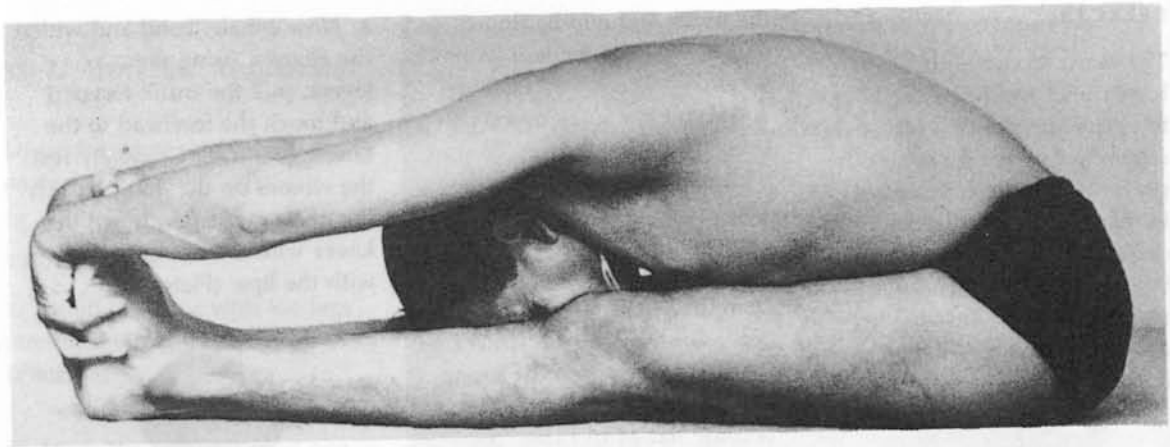

78

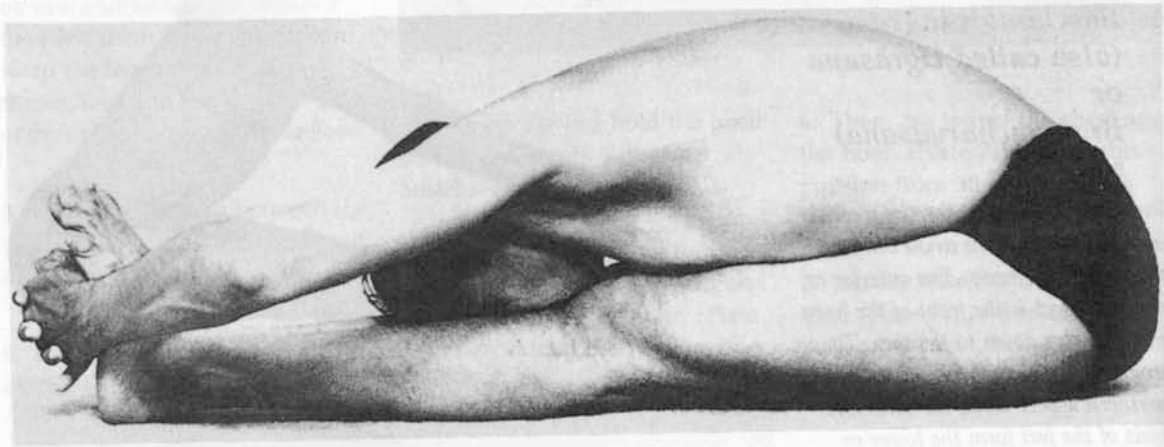

79

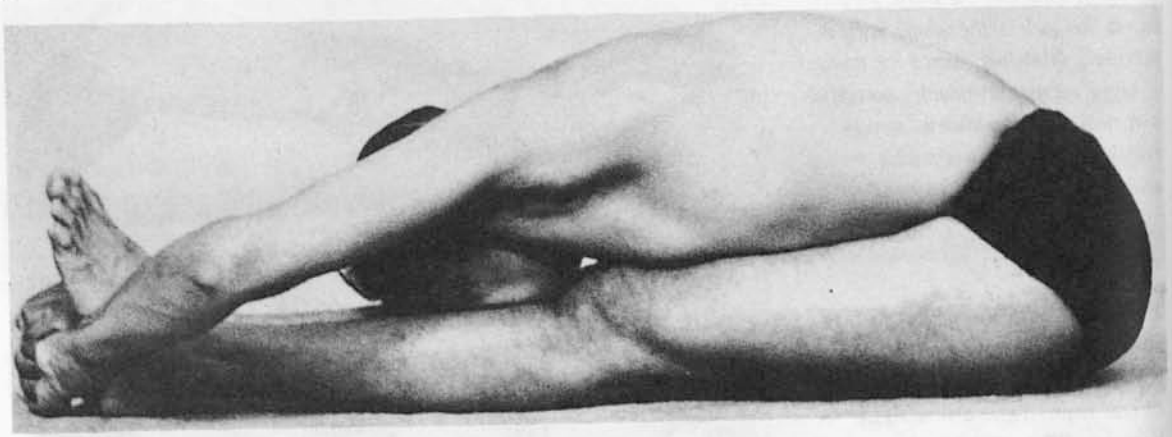

80

5. When this becomes easy, make a further effort to grip the soles and rest the chin on the knees. (Plate 78)

6. When this also becomes easy, clasp the hands by interlocking the fingers and rest the chin on the shins beyond the knees. (Plate 79)

7. When position 6 becomes easy, grip the right palm with the left hand or the left palm with the right hand beyond the outstretched feet exhale and rest the chin on the shins beyond the knees. (Plate 80)

8. If position 8 also becomes easy, hold the right wrist with the left hand or the left wrist with the right hand and rest the chin on the shins beyond the knees. (Plate 81)

9. See that the back of the legs at the knee joints rests firmly on the ground. In the initial stages the knees will be lifted off the floor. Tighten the muscles at the back of the thighs and pull the trunk forward. Then the back of the knee joints will rest on the floor.

10. Try and stay in whichever of the above positions you can achieve from 1 to 5 minutes, breathing evenly.

11. Inhale, raise the head from the knees and relax.

EFFECTS

This āsana tones the abdominal organs and keeps them free from sluggishness. It also tones the kidneys, rejuvenates the whole spine and improves the digestion.

The spines of animals are horizontal and their hearts and below the spine. This keeps them healthy and gives them great power of endurance. In humans the spine is vertical and the heart is not lower than the spine, so that they soon feel the effects of exertion and are also susceptible to the heart diseases. In Paschimottānāsana the spine is kept straight and horizontal and the heart is at a lower level than the spine. A good stay in this pose massages the heart, the spinal column and the abdominal organs, which feel refreshed and the mind is rested. Due to the extra stretch given to the pelvic region more oxygenated blood is brought there and the gonad glands absorb the required nutrition from the blood. This increases vitality, helps to cure impotency and leads to sex control. Hence, this āsana was called Brahmacharyāsana. Brahmacharya means celibacy and a Brahmachāri is one who has controlled the sex appetite.

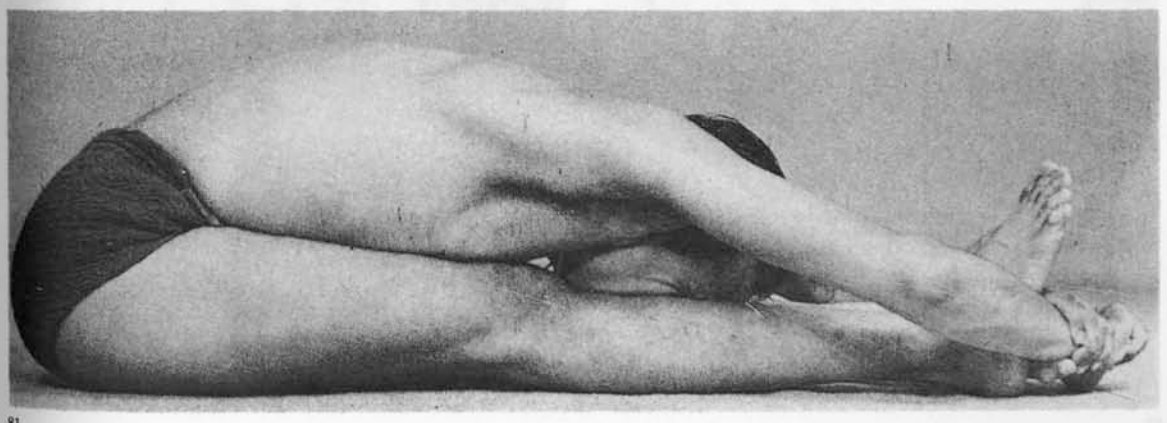

37| *Pūrvottānāsana*

ONE*

Pūrva literally means the East. It means the front of the whole body from the forehead to the toes. Uttāna means an intense stretch. In this posture, the whole front of the body is stretched intensely.

TECHNIQUE

1. Sit on the floor with the legs stretched straight in front. Place the palms on the floor by the hips, with the fingers pointing in the direction of the feet. (Plate 35)

2. Bend the knees and place the soles and heels on the floor.

3. Take the pressure of the body on the hands and feet, exhale and lift the body off the floor. Straighten the arms and the legs and keep the knees and elbows tightened. (Plate 82)

4. The arms will be perpendicular to the floor from the wrists to the shoulders. From the shoulders to the pelvis, the trunk will be parallel to the floor.

5. Stretch the neck and throw the head as far back as possible.

6. Stay in this posture for one minute, breathing normally.

7. Exhale, bend the elbows and knees, lower the body to sit on the floor and relax.

EFFECTS

This posture strengthens the wrists and ankles, improves the movement of the shoulder joints and expands the chest fully. It gives relief from the fatigue caused by doing other strenuous forward bending āsanas.

38| *Sālamba Śīrṣāsana I*

FOUR*

Sālamba means with support. Śīrṣa means the head. This is the head stand pose, one of the most important Yogi āsanas. It is the basic posture. It has several variations, which are described later as the Śīrṣāsana cycle. Its mastery gives one balance and poise, both physically and mentally. The technique of doing it is given at length in two parts; the first is for beginners, the second for those who can remain balanced in the pose. Attention is specially directed to the hints on Śīrṣāsana given after the two techniques.

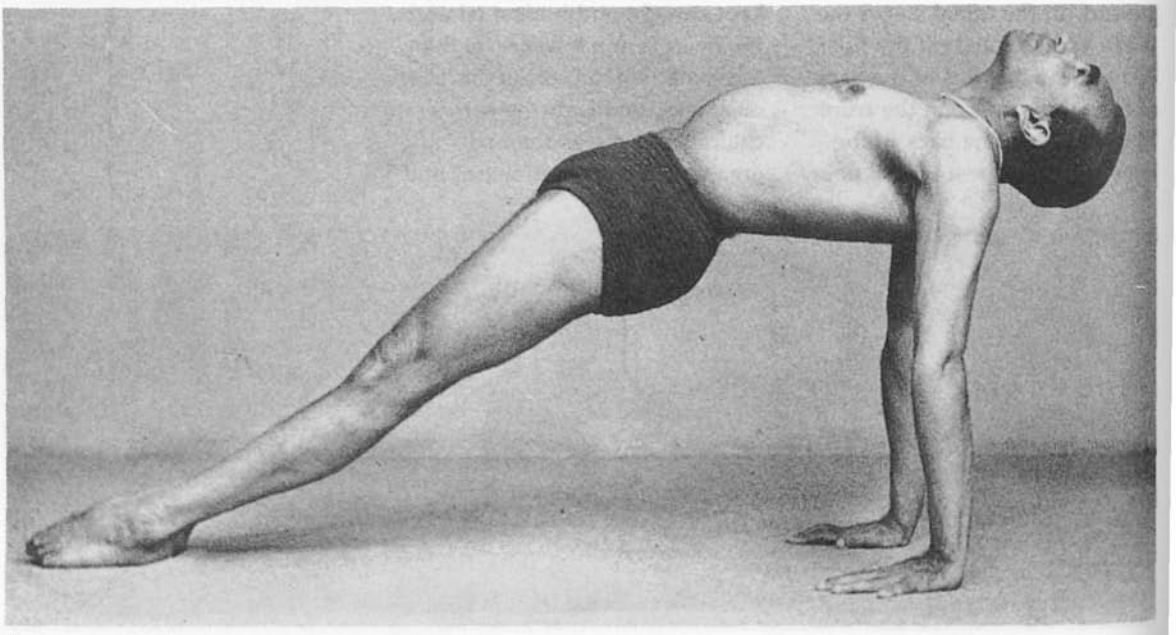

TECHNIQUE FOR BEGINNERS

1. Spread the blanket fourfold on the floor and kneel near it.

2. Rest the forearms on the centre of the blanket. While doing so take care that the distance between the elbows on the floor is not wider than the shoulders.

3. Interlock the fingers right up to the finger-tips (Plate 83), so that the palms form a cup. Place the sides of the palms near the little fingers on the blankets. While going up on to your head or balancing, the fingers should be kept tightly locked. If they are left loose, the weight of the body falls on them and the arms ache. So remember to lock them well.

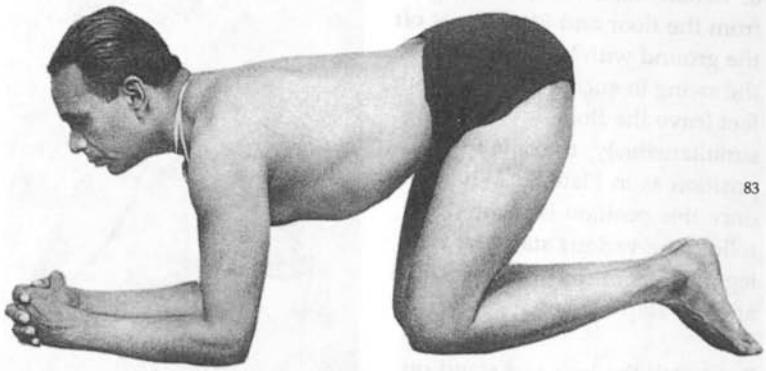

83

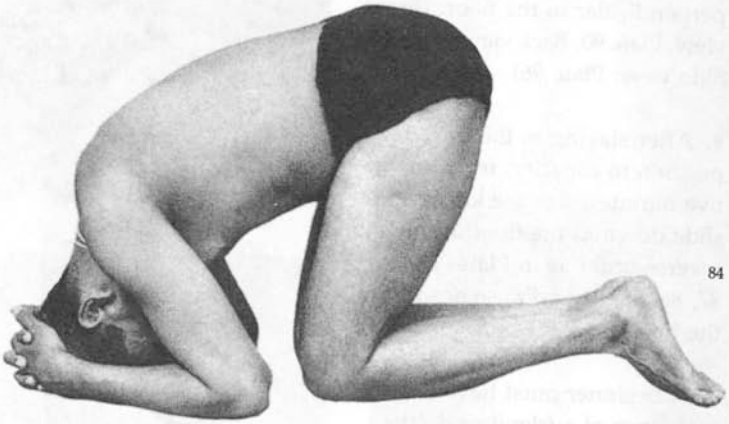

84

4. Rest the crown of the head only on the blanket, so that the back of the head touches the palms which are cupped. (Plate 84) Do not rest the forehead nor the back but only the crown of the head on the blanket. To do this move the knees towards the head.

5. After securing the head position, raise the knees from the floor by moving the toes closer to the head. (Plate 85)

85

6. Exhale, take a gentle swing from the floor and lift the legs off the ground with bent knees. Take the swing in such a way that both feet leave the floor simultaneously, to come to position as in Plate 86. When once this position is secured, follow the various stages of the leg movements as in Plates 87, 88 and 89, step by step.

7. Stretch the legs and stand on the head, keeping the whole body perpendicular to the floor. (Front view: Plate 90. Back view: Plate 91. Side view: Plate 96)

8. After staying in the final position to capacity, from one to five minutes, flex the knees and slide down to the floor in the reverse order as in Plates 89, 88, 87, 86, 85, 84 and raise head from the floor.

9. A beginner must have the assistance of a friend or do the āsana against a wall. While practising against a wall, the distance between it and the head should not be more than 2 or 3 inches. If the distance is greater, the spine will curve and the stomach will protrude. The weight of the body will be felt on the elbows and the position of the head may change. The face will appear to be flushed and the eyes either strained or puffed. It is, therefore, advisable for a beginner to do the head stand in a corner where two walls meet, placing the head some 2 to 3 inches from either wall.

86

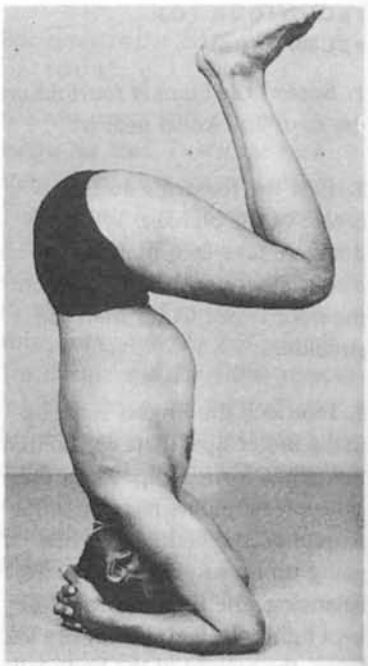

87

88

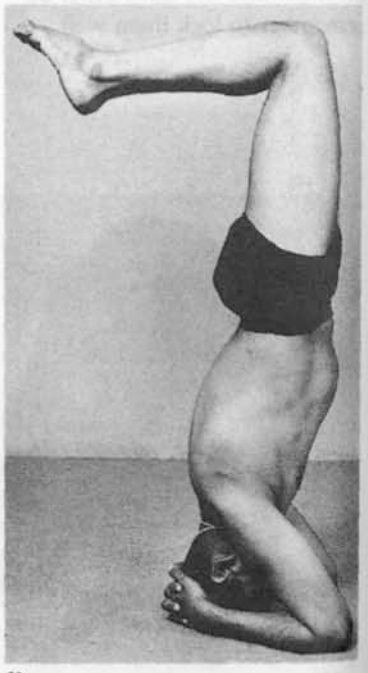

89

10. While doing the head stand against a wall or in a corner, the beginner should exhale, swing the legs up, support the hips against the side of the wall and move the feet up. In a corner, he can touch the heels to either side of the walls. He should then stretch the back vertically up, gradually leave the support of the wall and learn to master the balance. While coming down, he can rest the feet and hips against the wall, slide down and kneel, resting his knees on the floor. The movements of coming down and going up should be done with an exhalation.

11. The advantage which the beginner has of balancing in a corner is that his head and legs will be in the right angle formed by the walls, and he will be sure of his right position. This will not be the case if he balances against a straight wall. For while his balance is insecure he may sway from the wall, or his body may tilt or swing to the stronger side, while his legs may rest against the wall with a bend either at the waist or the hips. The beginner will not be in a position to know that he has tilted to one side, much less to correct it. In time he may learn to balance on the head, but by habit his body may still tilt or his head may not be straight. It is as hard to correct a wrong pose in the head stand as it is to break a bad habit. Moreover this wrong posture may well lead to aches and pains in the head, neck, shoulders and back. But the two walls of a corner will help the beginner to keep the āsana symmetrical.

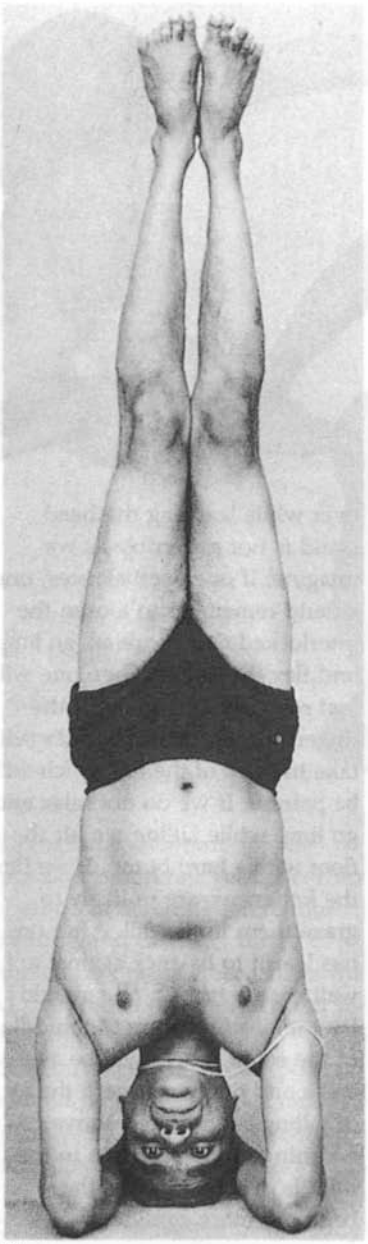

90

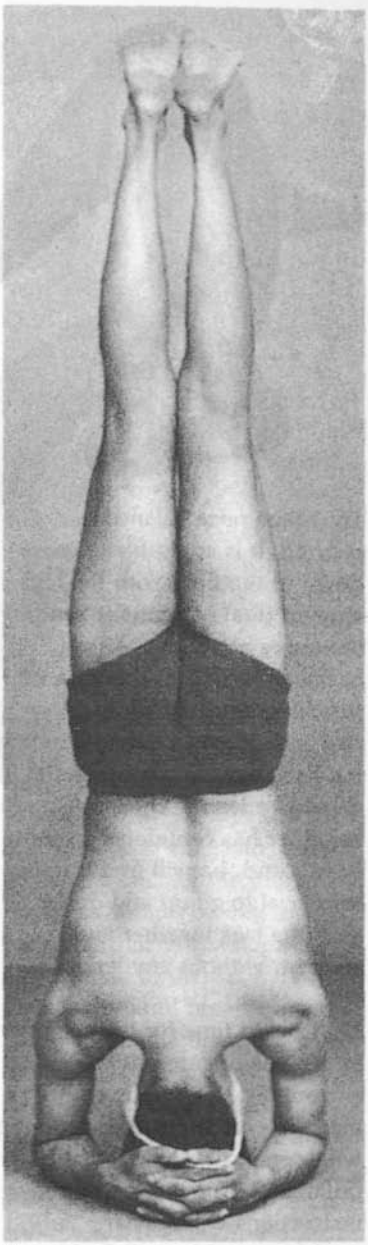

91

92

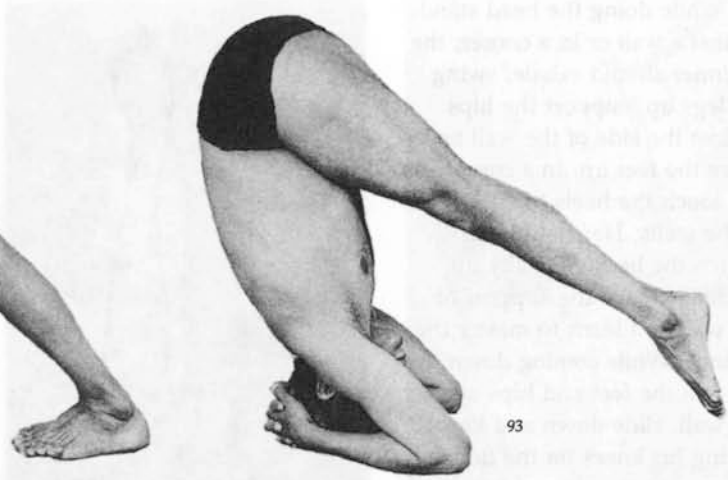

93

12. When once balance is secured, it is advisable to come down to the floor with the legs straight (that is, without bending the knees at all) and with a backward action of the hips. At first, it is not possible to go up and come down without bending the legs, but the correct method should be learnt. Once the beginner has confidence in the head stand, he will find it more beneficial to go up and down with the legs together and straight, without any jerks.

13. It takes time for the beginner to become oriented to his surroundings while he is balancing on his head. Everything will seen at first to be completely unfamiliar. The directions and instructions will appear confusing and he will find it an effort to think clearly or to act logically. This is due to fear of a fall. The best way to overcome fear is to face with equanimity the situation of which one is afraid. Then one gets the correct perspective, and one is not frightened any more. To topple over while learning the head stand is not as terrible as we imagine. If one overbalances, one should remember to loosen the interlocked fingers, relax, go limp and flex the knees. Then one will just roll over and smile. If the fingers are not loosened they will take the jerk of the fall which will be painful. If we do not relax and go limp while falling we hit the floor with a hard bump. If we flex the knees, we are unlikely to graze them in the fall. After one has learnt to balance against a wall or in a corner, one should try the head stand in the middle of the room. There will be a few spills and one must learn the art of falling as indicated above. Learning to do Śīrṣāsana in the middle of a room gives the beginner more confidence.

TECHNIQUE FOR THOSE WHO CAN BALANCE EIGHT*

1. Follow the technique described for beginners from positions 1 to 4.

2. After securing the head position, stretch the legs straight by raising the knees from the floor. Move the toes nearer to the head and try to press the heels to the floor, keeping the back erect. (Plate 92)

3. Stretch the dorsal or middle region of the spine and stay in this position for about 30 seconds while breathing evenly.

4. Exhale, raise the heels and take the toes off the floor with a backward movement of the hips. Raise both legs simultaneously, keeping them poker stiff. (Plate 93) Take a breath.

5. Again with an exhalation move the legs up until they are parallel to the floor. This position is called:

39| *Ūrdhvā Daṇḍāsana*

EIGHT*

(Ūrdhvā = up, daṇḍ = a staff)
Stay in this position for 10 seconds with normal breathing.

6. Exhale, move the legs up as in Plate 95, and then pull them up to the vertical position. (Side view: Plate 96) Stay in this pose from 1 to 5 minutes while breathing evenly.

7. Come down gradually, observing the above technique in a reverse order. (Plates 95, 94, 93 and 92) Rest the feet on the floor, bend the knees (Plate 84) and raise the head from the floor or blanket. (Plate 83)

8. While coming down, it is advisable to stay in Ūrdhvā Daṇḍāsana according to capacity up to one minute while breathing normally. In this position, the neck and trunk will not be perpendicular to the floor but will sway slightly backwards. The neck, shoulders and spine will be put to a very great strain and in the initial stages one cannot stay with the legs parallel to the floor for more than a few seconds. The stay will become longer as the neck, shoulders, abdomen and spine become stronger.

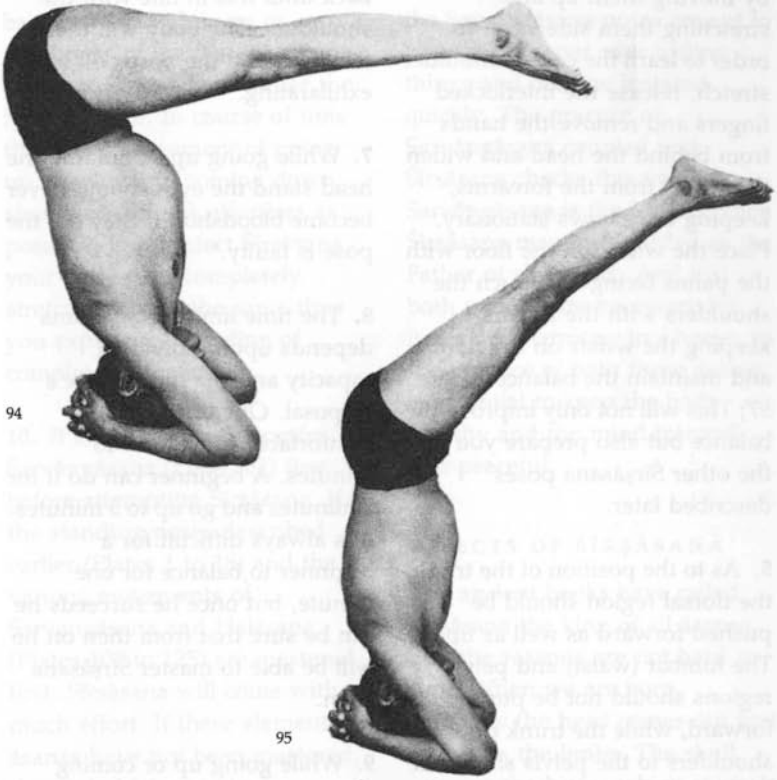

94

95

HINTS ON ŚĪRṢĀSANA

1. In Śīrṣāsana the balance alone is not important. One has to watch from moment to moment and find out the subtle adjustments. When we stand on our feet, we need no extra effort, strength or attention, for the position is natural. Yet the correct method of standing affects our bearing and carriage. It is, therefore, necessary to master the correct method as pointed out in the note on Tāḍāsana. In Śīrṣāsana also, the correct position should be mastered, as a faulty posture in this āsana will lead to pains in the head, neck and back.

2. The whole weight of the body should be borne on the head alone and not on the forearms and hands. The forearms and hands are to be used only for support to check any loss of balance. In a good pose you feel a circle, about the size of an Indian rupee, of the head in contact with the blanket on the floor.

3. The back of the head, the trunk, the back of the thighs and the heels should be in a line perpendicular to the floor and not inclined to one side. The throat, chin and breast-bone should be in one line, otherwise the head will tilt to one side or move forward. As regards the interlocked hands behind the head, the palms should not be stuck into the head. The upper and the lower sides of the palms should be in a line, otherwise the crown of the head will not rest on the floor correctly.

4. The elbows and the shoulders should be in a line and the elbows should not be widened. The shoulders should be kept as high above the floor as possible

by moving them up and stretching them sideways. In order to learn the correct shoulder stretch, release the interlocked fingers and remove the hands from behind the head and widen the wrists from the forearms, keeping the elbows stationary. Place the wrists on the floor with the palms facing up, touch the shoulders with the fingers, keeping the wrists on the floor and maintain the balance. (Plate 97) This will not only improve the balance but also prepare you for the other Śīrṣāsana poses described later.

5. As to the position of the trunk, the dorsal region should be pushed forward as well as up. The lumbar (waist) and pelvic regions should not be pushed forward, while the trunk from the shoulders to the pelvis should be kept perpendicular. If the pelvic area juts forward, it means that you are bearing the weight of the body not on the head alone but also on the elbows for you have not stretched the dorsal region (the chest) correctly. When viewed from the side, the body from the neck to the heels should appear straight.

6. As far as possible try and join the thighs, knees, ankles and toes. Stretch the legs fully, especially the back of the knees and thighs. If the legs swing back tighten the knees and the lower median portion of the abdomen above the pubes. This will keep the legs perpendicular. Keep the toes pointing up. If the legs swing forward, stretch the dorsal region and push the pelvic area slightly back until it is in line with the shoulders. The body will then feel light and the pose will be exhilarating.

7. While going up or holding the head stand the eyes should never become bloodshot. If they do, the pose is faulty.*

8. The time limit for Śīrṣāsana depends upon individual capacity and the time at one's disposal. One can hold it comfortably from 10 to 15 minutes. A beginner can do it for 2 minutes and go up to 5 minutes. It is always difficult for a beginner to balance for one minute, but once he succeeds he can be sure that from then on he will be able to master Śīrṣāsana soon.

9. While going up or coming down, move both legs together, inch by inch. All the movements should be done with exhalation. Inhale while waiting in a position. The effect of going down and coming up straight without bending the legs at the knees is that harmonious slow movement is gained and the flow of blood to the head is controlled. The face does not flush from jerky and fast movements, as the flow of blood to the waist and the legs is also controlled.

* I have taught this pose to a lady of 65 who was suffering from glaucoma. Now she finds the eyes are completely rested and the pain in them is much lessened. Medical examination revealed that the tension in the eyeballs had decreased. I am mentioning this to prove the value of the correct head stand.

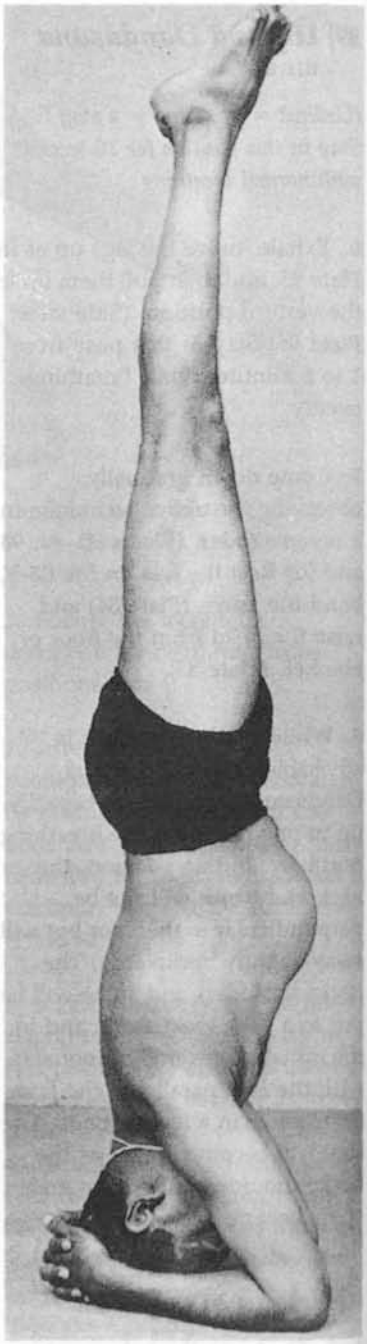

96

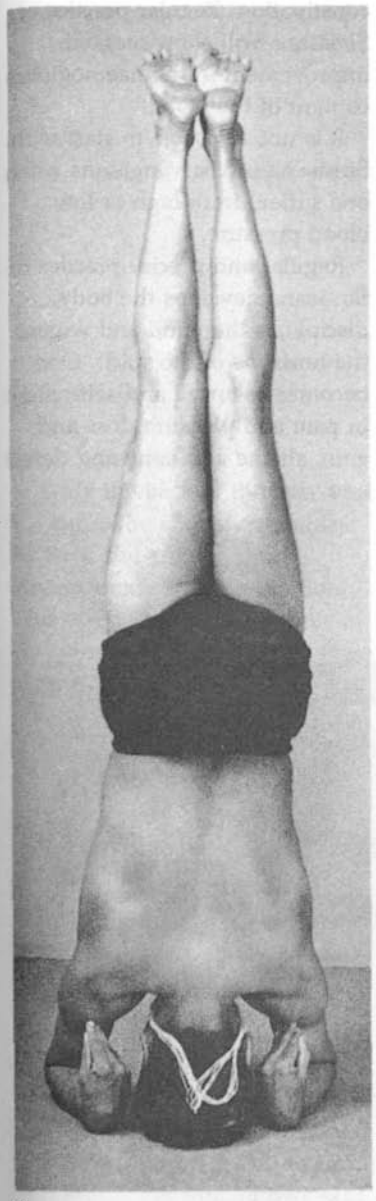

97

Then there is no danger of losing balance from giddiness or numbness of the feet when one stands up immediately after the head balance. In course of time the whole movement of going up, staying and coming down should become as effortless as possible. In a perfect Śīrṣāsana your body feels completely stretched and at the same time you experience a feeling of complete relaxation.

10. It is always safe to perfect Sarvāngāsana (Plate 102) first before attempting Śīrṣāsana. If the standing poses described earlier (Plates 1 to 18) and the various movements of Sarvangāsana and Halāsana (Plates 108 to 125) are mastered first, Śīrṣāsana will come without much effort. If these elementary āsanas have not been mastered, the period taken to learn Śīrṣāsana will be longer.

11. After one has learnt to balance in Śīrṣāsana, however, it is preferable to perform Śīrṣāsana first before practising any other āsana. This is because one cannot balance or hold the head stand if the body is exhausted by doing other poses or if the breathing becomes fast and shaky. Once the body is tired or the breathing is not free and easy, the body will shake and it will be difficult to maintain the balance. It is always better to do Śīrṣāsana first when one is fresh.

12. Śīrṣāsana should always be followed by Sarvāngāsana and its cycle. It has been observed that people who devote themselves to Śīrṣāsana alone without doing the Sarvāngāsana poses are apt to lose their temper over trifling things and become irritated quickly. The practice of Sarvāngāsana coupled with Śīrṣāsana checks this trait. If Sarvāngāsana is the Mother, then Śīrṣāsana may be regarded as the Father of all āsanas. And just as both parents are necessary for peace and harmony in a home, so the practice of both these āsanas is essential to keep the body healthy and the mind tranquil and peaceful.

EFFECTS OF ŚĪRṢĀSANA

The ancient books have called Śīrṣāsana the king of all āsanas and the reasons are not hard to find. When we are born, normally the head comes out first and then the limbs. The skull encases the brain, which controls the nervous system and the organs of sense. The brain is the seat of intelligence, knowledge, discrimination, wisdom and power. It is the seat of Brahman, the soul. A country cannot prosper without a proper king or constitutional head to guide it; so also the human body cannot prosper without a healthy brain.

The *Bhagavad-Gītā* says: 'Harmony (sattva), mobility (rajas), inertia (tamas), such are the qualities, matter-born; they bind fast, O great armed one (Arjuna), the indestructible dweller in the body.' (Fourteenth Discourse, verse 5) All these qualities stem from the brain, and sometimes one quality prevails and sometimes the others. The head is the centre of sattvic

qualities which control discrimination; the trunk of Rājasic qualities which control passion, emotion and actions; and the region below the diaphragm of tāmasic qualities which control sensual pleasures like the enjoyment of food and drink, and the thrills and pleasures of sex.

Regular practice of Śīrṣāsana makes healthy pure blood flow through the brain cells. This rejuvenates them so that thinking power increases and thoughts become clearer. The āsana is a tonic for people whose brains tire quickly. It ensures a proper blood supply to the pituitary and pineal glands in the brain. Our growth, health and vitality depend on the proper functioning of these two glands.

People suffering from loss of sleep, memory and vitality have recovered by the regular and correct practice of this āsana and have become fountains of energy. The lungs gain the power to resist any climate and stand up to any work, which relieves one from colds, coughs, tonsillitis, halitosis (foul breath) and palpitations. It keeps the body warm. Coupled with Sarvāngāsana movements (Plates 108 to 125), it is a boon to people suffering from constipation. Regular practice of Śīrṣāsana will show marked improvement in the haemoglobin content of the blood.

It is not advisable to start with Śīrṣāsana and Sarvāngāsana when one suffers from high or low blood pressure.

Regular and precise practice of Śīrṣāsana develops the body, disciplines the mind and widens the horizons of the spirit. One becomes balanced and self-reliant in pain and pleasure, loss and gain, shame and fame and defeat and victory.

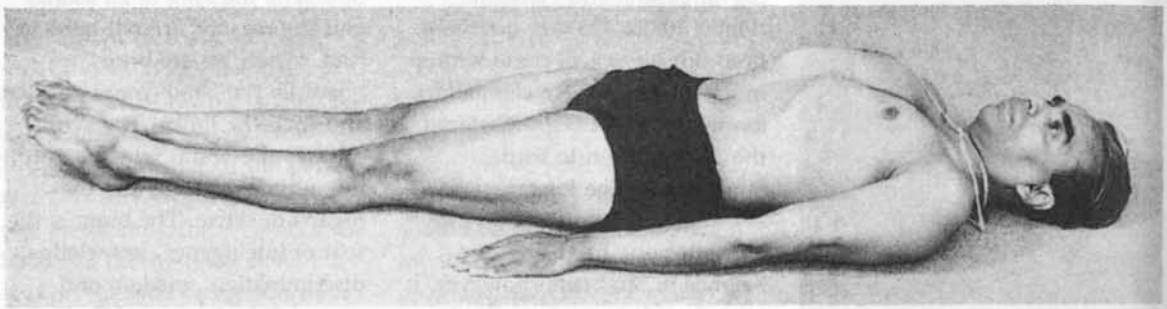

98

40| *Sālamba Sarvāngāsana I*

TWO*

Ālamba means a prop, a support and sa together with or accompanied by. Sālamba, therefore means supported or propped up. Sarvānga (Sarva = all, whole, entire, complete; anga = limb or body) means the entire body or all the limbs. In this pose the whole body benefits from the exercise, hence the name.

TECHNIQUE FOR BEGINNERS

1. Lie flat on the back of the carpet keeping the legs stretched out, tightened at the knees. Place the hands by the side of the legs, palms down. (Plate 98) Take a few deep breaths.

2. Exhale, bend the knees and move the legs towards the stomach till the thighs press it. (Plate 99) Take two breaths.

3. Raise the hips from the floor with an exhalation and rest the hands on them by bending the arms at the elbows. (Plate 100) Take two breaths.

4. Exhale, raise the trunk up perpendicularly supported by the hands until the chest touches the chin. (Plate 101)

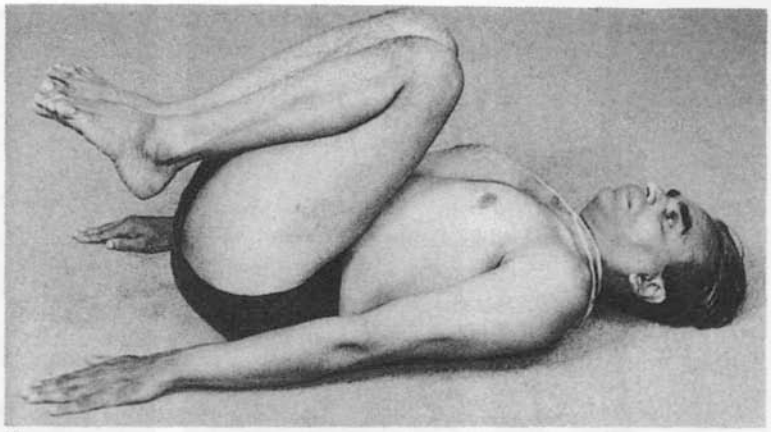

99

5. Only the back of the head and the neck, the shoulders and the backs of the arms up to the elbows should rest on the floor. Place the hands in the middle of the spine as in Plate 101. Take two breaths.

6. Exhale and stretch the legs straight with the toes pointing up. (Front view: Plate 102. Back view: Plate 103)

7. Stay in this position for 5 minutes with even breathing.

8. Exhale, gradually slide down, release the hands, lie flat and relax.

9. If you cannot do the āsana without support use a stool and follow the technique. See Plate 104.

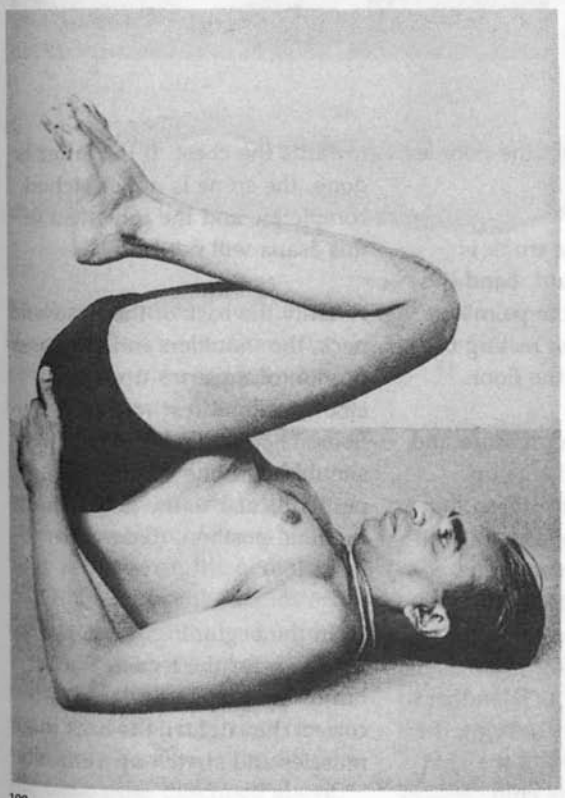

100

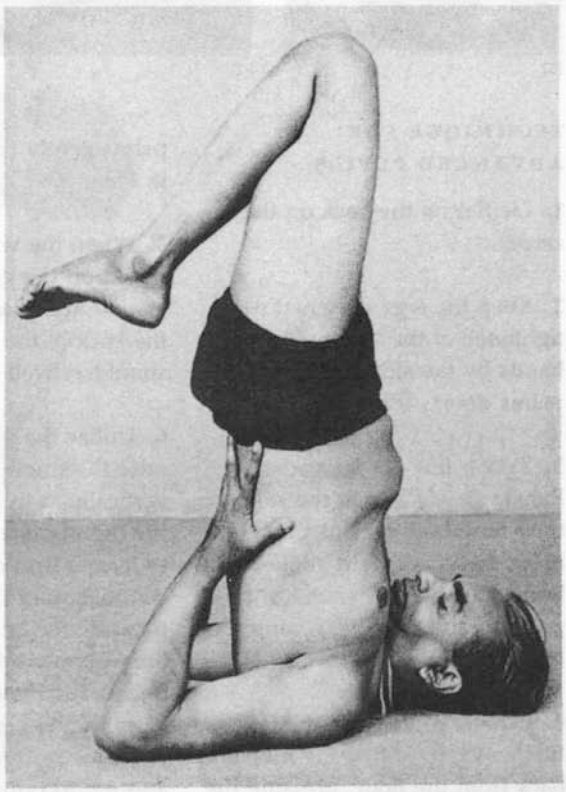

101

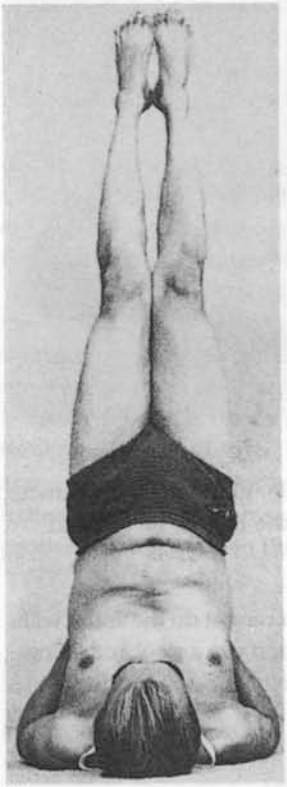

102

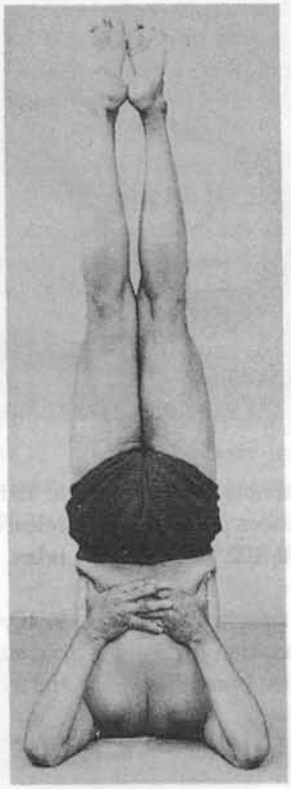

103

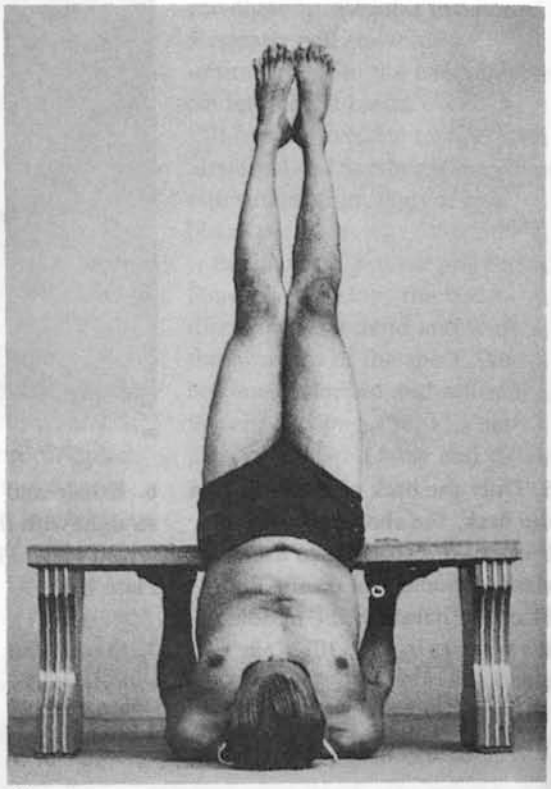

104

TECHNIQUE FOR ADVANCED PUPILS

1. Lie flat on the back on the carpet.

2. Keep the legs stretched out, tightened at the knees. Place the hands by the side of the legs, palms down. (Plate 98)

3. Take a few deep breaths. Exhale slowly and at the same time raise both legs together and bring them at a right angle to the body. (Plate 105) Remain in this position and inhale, keeping the legs steady.

4. Exhale, again raise the legs further up by lifting the hips and back from the floor, pressing the palms gently against the floor as in Plate 106.

5. When the whole trunk is raised off the ground, bend the elbows and place the palms on the back of the ribs, resting the shoulders well on the floor.

6. Utilize the palm pressure and raise the trunk and legs up vertically as in Plate 107 so that the breast-bone presses the chin to form a firm chinlock. The contraction of the throat and pressing the chin against the breast-bone to form a firm chinlock is known as Jālandhara Bandha. Remember to bring the chest forward to touch the chin and not to bring the chin towards the chest. If the latter is done, the spine is not stretched completely and the full effect of this āsana will not be felt.

7. Only the back of the head and neck, the shoulders and the upper portion of the arms up to the elbows should rest well on the floor. The remainder of the body should be in one straight line, perpendicular to the floor. This is the final position. (Side view: Plate 108)

8. In the beginning, there is a tendency for the legs to swing out of the perpendicular. To correct this, tighten the back thigh muscles and stretch up vertically.

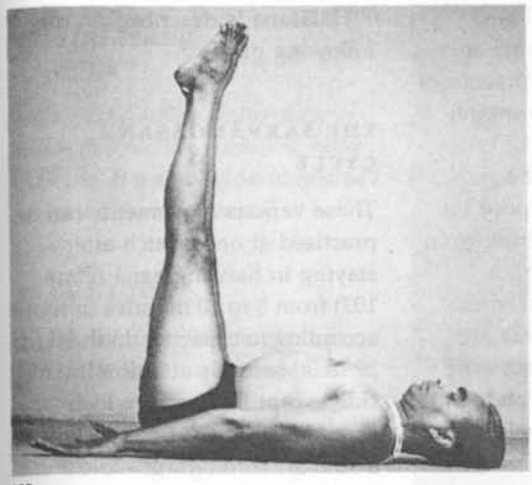

105

9. The elbows should not be placed wider than the shoulders. Try and stretch the shoulders away from the neck and also to bring the elbows close to each other. If the elbows are widened, the trunk cannot be pulled up properly and the pose will look imperfect. Also see that the neck is straight with the centre of the chin resting on the sternum. In the beginning, the neck moves sideways and if this is not corrected, it will cause pain and injure the neck.

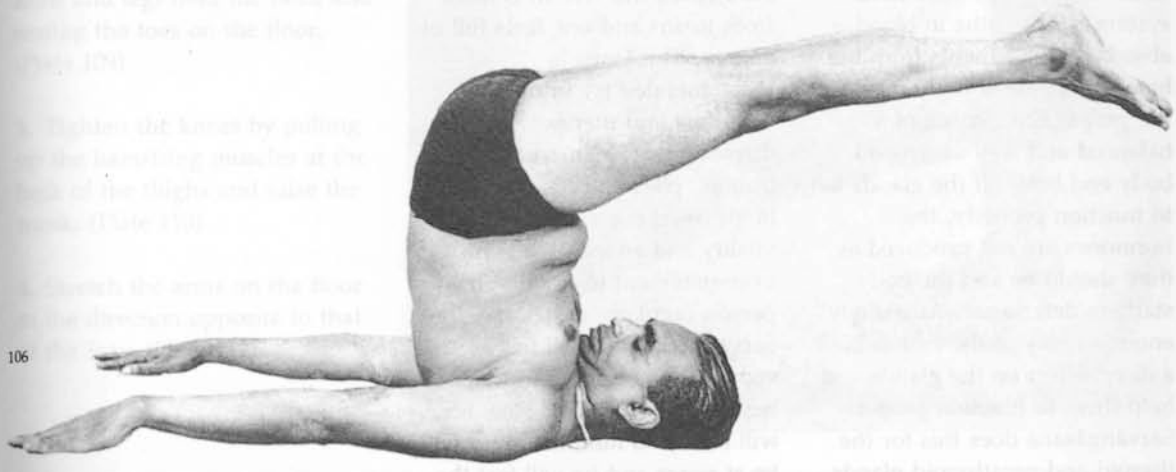

106

10. Remain in this pose for not less than 5 minutes. Gradually increase the time to 15 minutes; this will have no ill effects.

11. Release the hands, slide to the floor, lie flat and relax.

As the weight of the whole body is borne on the neck and shoulders and as the hands are used to support the weight this āsana is called Sālamba Sarvāngāsana. In Sarvāngāsana there are various movements which can be done in addition to the basic pose described above.

107

108

EFFECTS

The importance of Sarvāngāsana cannot be over-emphasized. It is one of the greatest boons conferred on humanity by our ancient sages. Sarvāngāsana is the mother of āsanas. As a mother strives for harmony and happiness in the home, so this āsana strives for the harmony and happiness of the human system. It is a panacea for most common ailments. There are several endocrine organs or ductless glands in the human system which bathe in blood, absorb the nutriments from the blood and secrete hormones for the proper functioning of a balanced and well developed body and brain. If the glands fail to function properly, the hormones are not produced as they should be and the body starts to deteriorate. Amazingly enough many of the āsanas have a direct effect on the glands and help them to function properly. Sarvāngāsana does this for the thyroid and parathyroid glands which are situated in the neck region, since due to the firm chinlock their blood supply is increased. Further, since the body is inverted the venous blood flows to the heart without any strain by force of gravity. Healthy blood is allowed to circulate around the neck and chest. As a result, persons suffering from breathlessness, palpitation, asthma, bronchitis and throat ailments get relief. As the head remains firm in this inverted position, and the supply of the blood to it is regulated by the firm chinlock, the nerves are soothed and headaches – even chronic ones – disappear. Continued practice of this āsana eradicates common colds and other nasal disturbances. Due to the soothing effect of the pose on the nerves, those suffering from hypertension, irritation, shortness of temper, nervous breakdown and insomnia are relieved. The change in bodily gravity also affects the abdominal organs so that the bowels move freely and constipation vanishes. As a result the system is freed from toxins and one feels full of energy. The āsana is recommended for urinary disorders and uterine displacement, menstrual trouble, piles and hernia. It also helps to relieve epilepsy, low vitality and anaemia. It is no over-statement to say that if a person regularly practises Sarvāngāsana he will feel new vigour and strength, and will be happy and confident. New life will flow into him, his mind will be at peace and he will feel the joy of life. After a long illness, the practice of this āsana regularly twice a day brings back lost vitality. The Sarvāngāsana cycle activates the abdominal organs and relieves people suffering from stomach and intestinal ulcers, severe pains in the abdomen and colitis.

People suffering from high blood pressure should not attempt Sālamba Sarvāngāsana unless they do Halāsana (Plate 113) first and can stay in it for not less than 3 minutes.

Do not perform Sarvāngāsana during menstruation.

Halāsana is described on the following pages.

THE SARVĀNGĀSANA CYCLE

These various movements can be practised at one stretch after staying in Sarvāngāsana (Plate 102) from 5 to 10 minutes or more according to capacity; do them for 20 to 30 seconds at a time each side except Halāsana, which should last from 3 to 5 minutes at a stretch.

41| *Halāsana*

FOUR*

Hala means a plough, the shape of which this posture resembles, hence the name. It is a part of Sarvāngāsana I and a continuation thereof.

TECHNIQUE

1. Do Sālamba Sarvāngāsana (Plate 103) with a firm chinlock.

2. Release the chinlock, lower the trunk slightly, moving the arms and legs over the head and resting the toes on the floor. (Plate 109)

3. Tighten the knees by pulling up the hamstring muscles at the back of the thighs and raise the trunk. (Plate 110)

4. Stretch the arms on the floor in the direction opposite to that of the legs. (Plate 111)

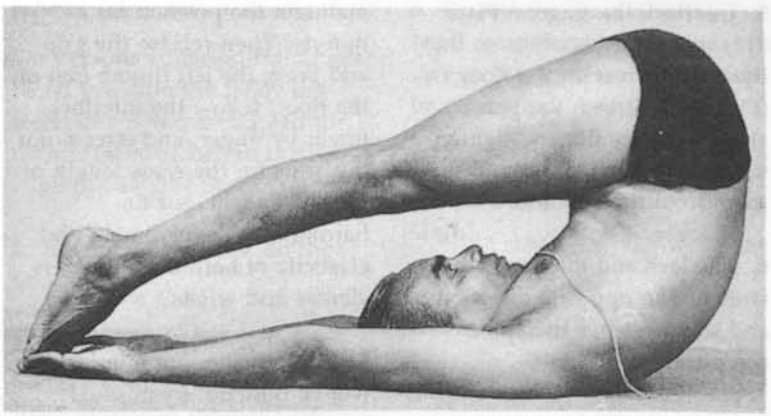

109

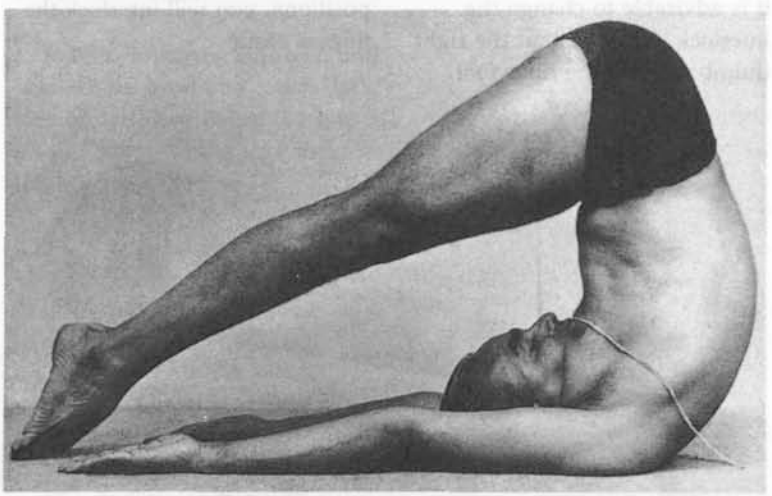

110

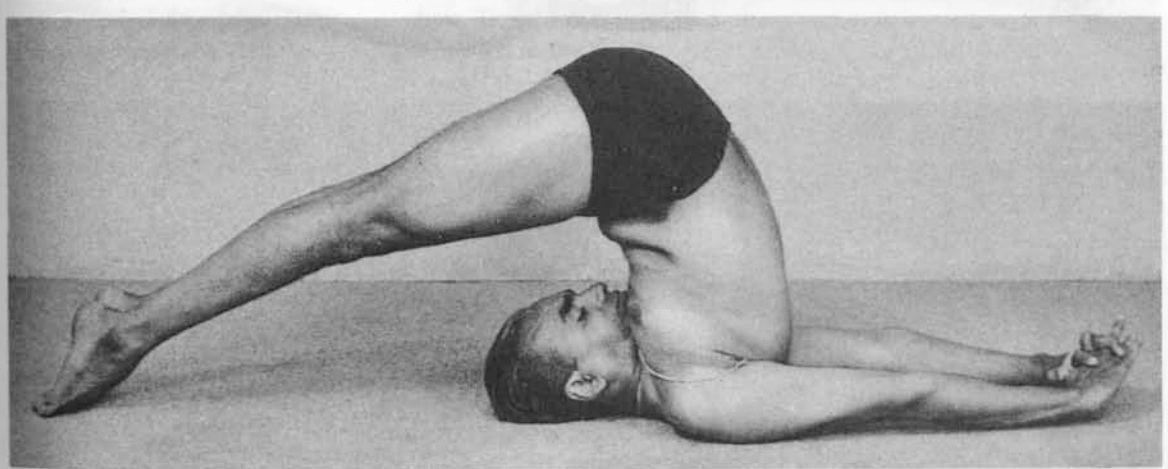

111

5. Interlock the fingers (Plate 112) and turn the wrists so that the thumbs rest on the floor. (Plate 113) Stretch the palms along with the fingers, tighten the arms at the elbows and pull them from the shoulders.

6. The legs and the hands are stretched in opposite directions and this stretches the spine completely.

7. While interlocking the fingers, it is advisable to change the interlock. Suppose that the right thumb touches the floor first, maintain the position for a minute. Then release the grip and bring the left thumb first on the floor, follow the interlock, finger by finger, and stretch out the arms for the same length of time. This will lead to harmonious development and elasticity of both the shoulders, elbows and wrists.

8. In the beginning interlocking will be difficult. By gradual practice of the above mentioned positions, you will interlock the fingers easily.

9. In the beginning it is also difficult to keep the toes firmly on the floor behind the head. If you lengthen the timing and stretch of Sarvāngāsana (Plate 102) before going into Halāsana, the toes will remain longer on the floor.

10. Remain in the attainable pose from one to five minutes with normal breathing.

11. Release the hands. Raise the legs up to Sarvāngāsana and gradually slide down to the floor. Lie flat on the back and relax.

112

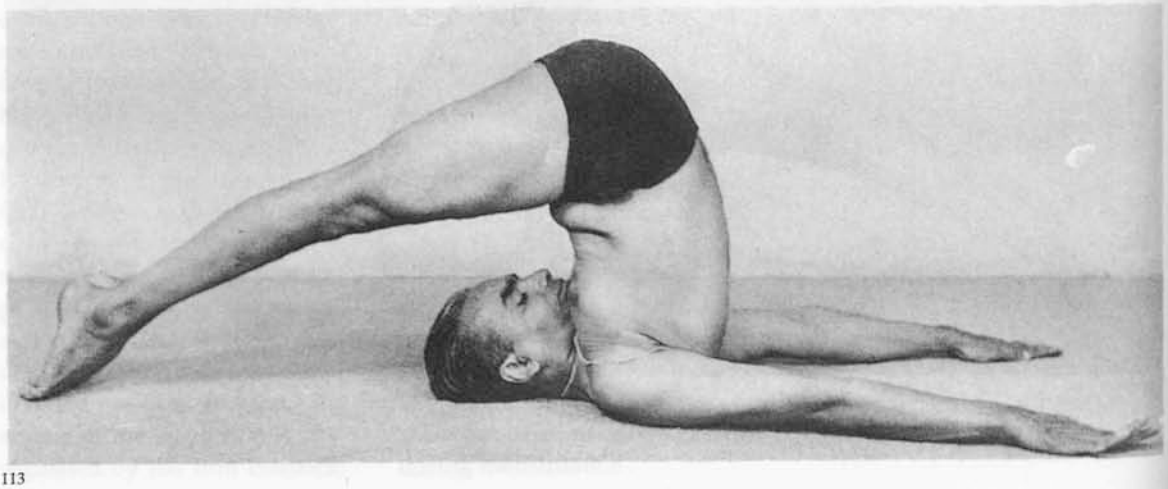

113

EFFECTS

The effect of Halāsana is the same as that of Sarvāngāsana. (Plate 102) In addition, the abdominal organs are rejuvenated due to contraction. The spine receives an extra supply of blood due to the forward bend and this helps to relieve backache. Cramps in the hands are cured by interlocking and stretching the palms and fingers. People suffering from stiff shoulders and elbows, lumbago and arthritis of the back find relief in this āsana. Griping pain in the stomach due to wind is also relieved and lightness is felt immediately.

The pose is good for people with a tendency for high blood pressure. If they perform Halāsana first and then Sarvāngāsana, they will not feel the rush of blood or the sensation of fullness in the head.

Halāsana is a preparatory pose to Paschimottānāsana. (Plate 81) If one improves in Halāsana, the resulting mobility of the back will enable one to perform Paschimottānāsana well.

NOTE

For persons suffering from high blood pressure the following technique is recommended for doing Halāsana before they attempt Sālamba Sarvāngāsana.

1. Lie flat on the back on the floor.

2. Exhale, slowly raise the legs to a perpendicular position and stay there breathing normally for about 10 seconds.

3. Exhale, bring the legs over and beyond the head and touch the toes on the floor. Keep the toes on the floor and the legs stiff at the knees.

4. If you have difficulty in keeping the toes on the floor, then place a chair or a stool behind the head and rest the toes on it.

5. If the breathing becomes heavy or fast do not rest the toes on the floor, but on a stool or chair. Then fullness or pressure is not felt in the head.

6. Extend the arms over the head, keep them on the floor and stay in this position with normal breathing for 3 minutes.

7. Throughout the āsana, gaze at the tip of the nose with the eyes shut.

42| *Karṇapīdāsana*

ONE*

Karṇa means the ear. Pīda means pain, discomfort or pressure. This is a variant of Halāsana and can be done along with it.

TECHNIQUE

1. Do Halāsana (Plate 113) and after completing the time limit for that pose, flex the knees and rest the right knee by the side of the right ear and the left knee by the side of the left.

2. Both knees should rest on the floor, pressing the ears.

3. Keep the toes stretched out and join the heels and toes. Rest the hands either on the back of the ribs (Plate 114a), or interlock the fingers and stretch out the arms (Plate 114b).

4. Remain in this position for half a minute or a minute with normal breathing.

EFFECTS

This āsana rests the trunk, heart and legs. The spine is stretched more while bending the knees, and this helps the circulation of blood round the waistline.

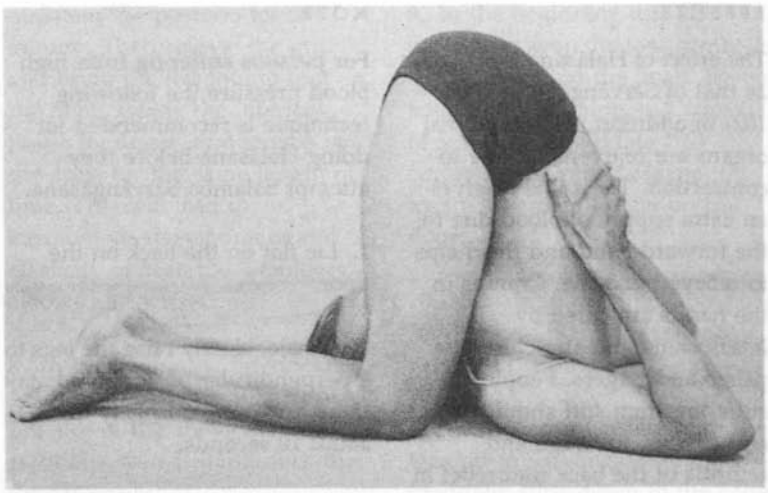

114A

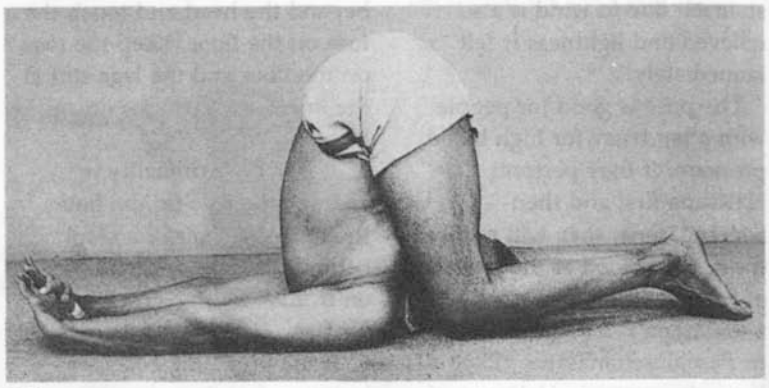

114B

43| *Supta Koṇāsana*

TWO*

Supta means lying down and koṇa an angle. It is a variation of Halāsana in which the legs are spread apart.

TECHNIQUE

1. From Karṇapīdāsana (Plates 114a, 114b), stretch the legs straight and spread the legs as far apart as you can.

2. Pull the trunk up and tighten the knees.

3. Hold the right toe with the right hand and the left toe with the left one. Keep the heels up. After gripping the toes, move the dorsal region of the spine still further up and extend the hamstring muscles. (Plate 115).

4. Stay in the pose from 20 to 30 seconds with normal breathing.

EFFECTS

This pose tones the legs and helps to contract the abdominal organs.

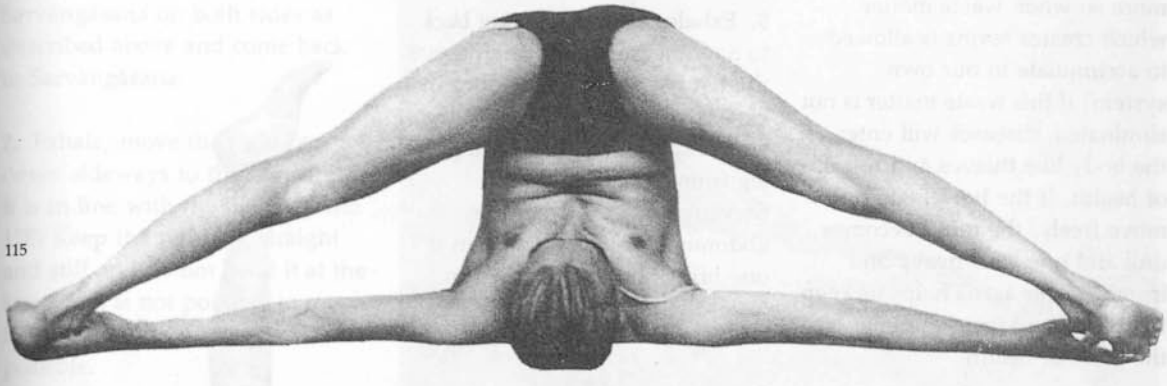

115

44| *Pārśva Halāsana*

FOUR*

In Halāsana (Plate 114) both the legs rest behind the head. In this pose they rest sideways on one side of and in line with the head. This is the lateral plough pose.

TECHNIQUE

1. Do Supta Koṇāsana (Plate 115) and come back to Halāsana.

2. Place the palms on the back of the ribs.

3. Move both the legs as far as you can to the left.

4. Tighten both knees, raise the trunk up with the help of the palms and stretch the legs. (Plate 116)

5. Remain in this position for half a minute with normal breathing.

6. Exhale, move the legs to the right until they are in line with the head and hold the pose for half a minute. Do not disturb the position of the chest and trunk when the legs are moved. The chest and trunk should remain as in Sarvāngāsana or Halāsana.

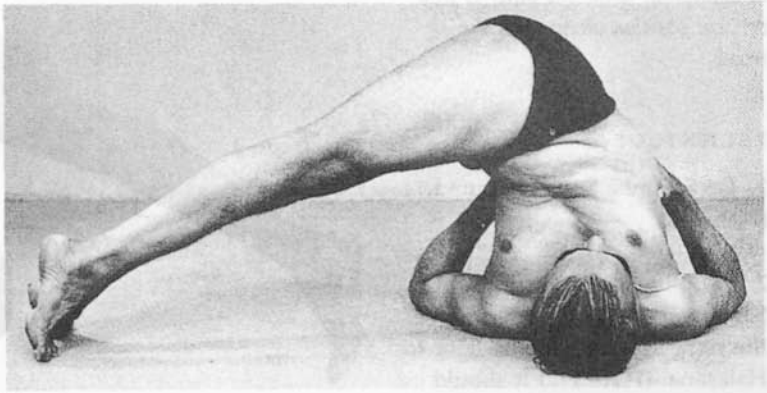

116

EFFECTS

In this āsana, the spine moves laterally and becomes more elastic. The colon, which is inverted during the movements, is exercised properly and elimination will be complete. People suffering from acute or chronic constipation which is the mother of several diseases derive great benefit from this āsana. If rubbish is dumped outside our house we feel sick. How much more so when waste matter which creates toxins is allowed to accumulate in our own system? If this waste matter is not eliminated, diseases will enter the body like thieves and rob us of health. If the bowels do not move freely, the mind becomes dull and one feels heavy and irritable. This āsana helps us keep the bowels free and thereby win the prize of health.

45| *Eka Pāda Sarvāngāsana*

FIVE*

Eka means one, single. Pāda means the foot. In this variation of Sarvāngāsana, one leg is on the floor in Halāsana, while the other is in a vertical position along with the trunk.

TECHNIQUE

1. Do Sālamba Sarvāngāsana I. (Plate 102)

2. Keep the left leg up in Sarvāngāsana. Exhale and move the right leg down to the floor to Halāsana. (Plate 117) It should remain stiff and straight and not bend at the knee. If it is not possible to touch the floor, lower the leg as far as possible.

3. While resting the right leg on the floor, the left knee should be kept taut and not allowed to tilt sideways. The left leg should be kept straight, facing the head.

4. Stay in the pose for 20 seconds maintaining normal breathing.

5. Exhale, lift the right leg back to Sarvāngāsana, and then move the left leg down to the floor in Halāsana, keeping the right leg vertically up and stiff. Lifting the leg from the floor back to Sarvāngāsana exercises the abdominal organs more than if one brings both legs down to Halāsana.

6. Stay on this side for the same length of time.

EFFECTS

This āsana tones the kidneys and the leg muscles.

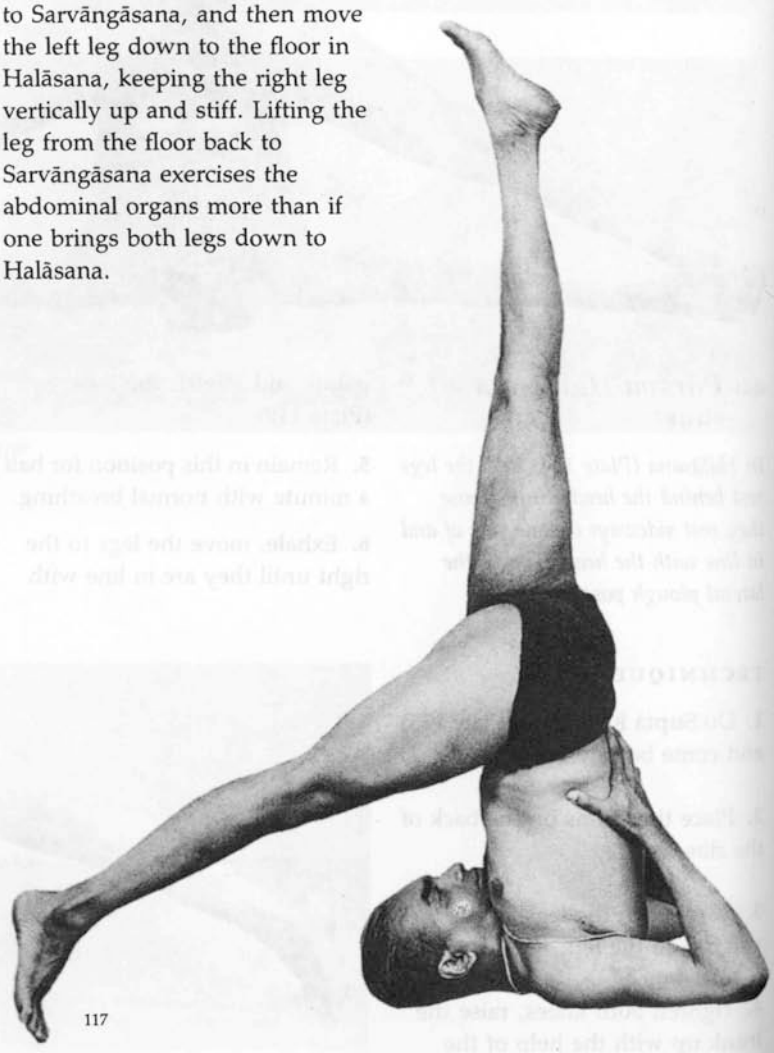

117

46| *Pārśvaika Pāda Sarvāngāsana*

SIX*

Pārśva means the side. In Eka Pāda Sarvāngāsana (Plate 117) the lower leg rests behind the head, whereas here it rests sideways in line with the trunk.

TECHNIQUE

1. Perform Eka Pāda Sarvāngāsana on both sides as described above and come back to Sarvāngāsana.

2. Exhale, move the right leg down sideways to the floor until it is in line with the trunk. (Plate 118) Keep the right leg straight and stiff and do not bend it at the knee. If it is not possible to touch the floor, lower the leg as far as possible.

3. The left leg which is vertically up should be kept straight and not allowed to tilt to the right. The ribs should be lifted with the palms to expand the chest fully.

4. Remain in the pose for 20 seconds with normal breathing, exhale, and go back to Sarvāngāsana. Repeat with the other leg for the same length of time and return to Sarvāngāsana.

EFFECTS

This pose relieves constipation and also tones the kidneys.

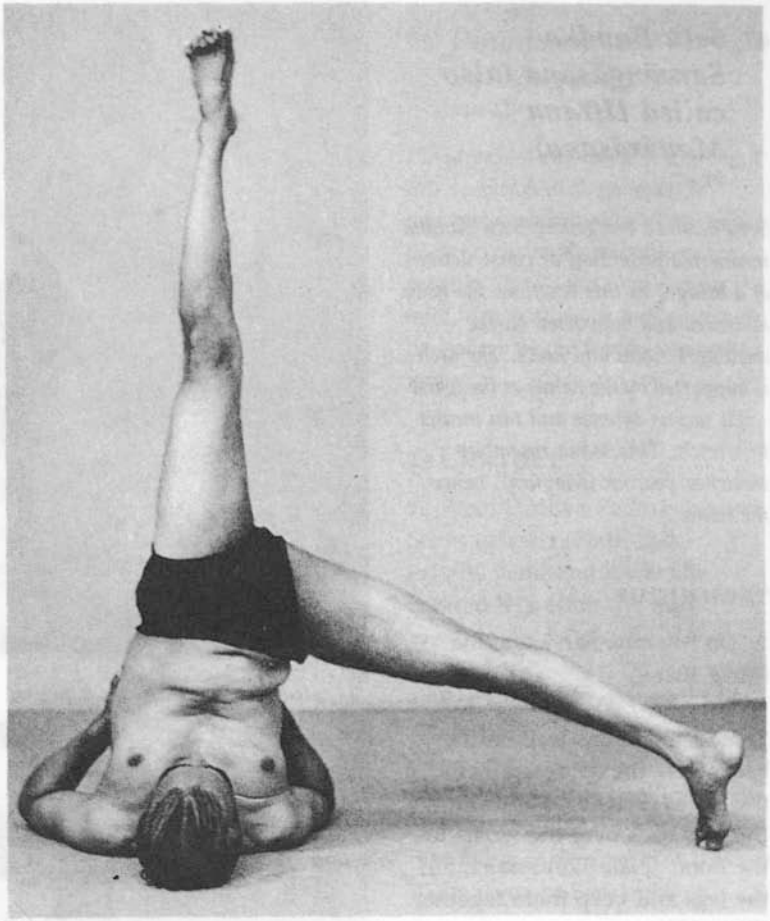

118

47| *Setu Bandha Sarvāngāsana (also called Uttāna Mayūrāsana)*

TEN*

Setu means a bridge and Setu Bandha means the formation or construction of a bridge. In this position, the body is arched and supported on the shoulders, soles and heels. The arch is supported by the hands at the waist.

Ut means intense and tān means to stretch. This āsana resembles a stretched peacock (Mayūra), hence the name.

TECHNIQUE

1. Do Sālamba Sarvāngāsana. (Plate 102)

2. Rest the palms well on the back, raise the spine up, bend the knees (Plate 119) and throw the legs back over the wrists to the floor. (Plate 120) Stretch out the legs and keep them together. (Plate 121)

3. The whole body forms a bridge, the weight of which is borne by the elbows and the wrists. The only parts of the body in contact with the ground will be the back of the head and neck, the shoulders, the elbows and the feet. Stay in the pose from half a minute to a minute with normal breathing.

4. It is possible to lessen the pressure on the elbows and the wrists by stretching the spine towards the neck, keeping the heels firmly on the ground.

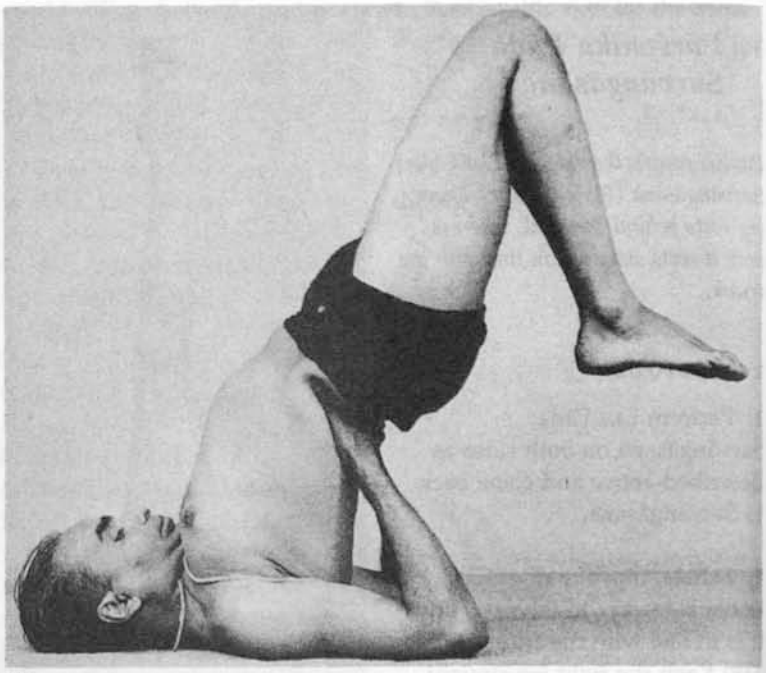

119

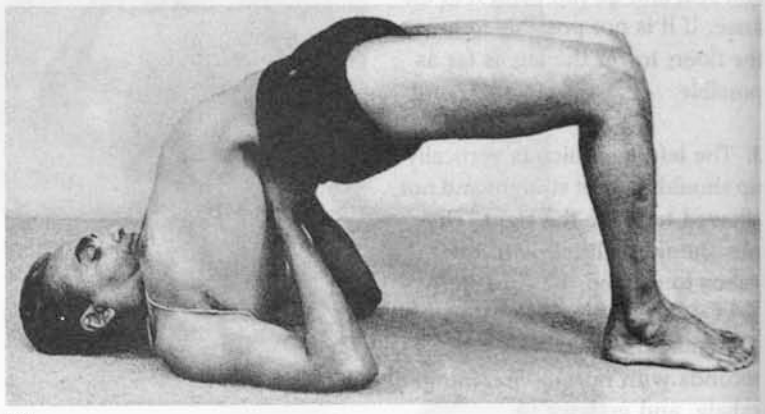

120

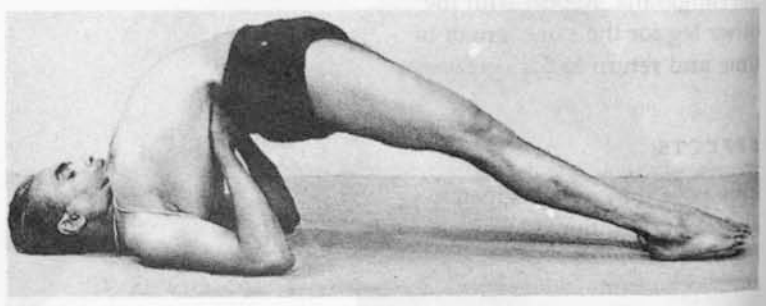

121

EFFECTS

This āsana gives the spine a backward movement and removes the strain on the neck caused by the other various movements of Sarvāngāsana.

A healthy and flexible spine indicates a healthy nervous system. If the nerves are healthy a man is sound in mind and body.

48| *Ūrdhva Padmāsana in Sarvāngāsana*

FOUR*

Ūrdhva means above, high. Padma means a lotus. In this Sarvāngāsana variation, the legs, instead of being kept straight up, are bent at the knees and crossed so that the right foot rests on the left thigh and the left foot on the right thigh as in the lotus pose. (Plate 53)

TECHNIQUE

1. From Sālamba Sarvāngāsana, bend the legs at the knees and cross them. First place the right foot over the left thigh, and then the left foot over the right thigh.

2. Stretch the crossed legs vertically up, move the knees closer to each other and the legs as far back as possible from the pelvic region. (Plate 122)

3. Stay in this pose from 20 to 30 seconds with deep and even breathing.

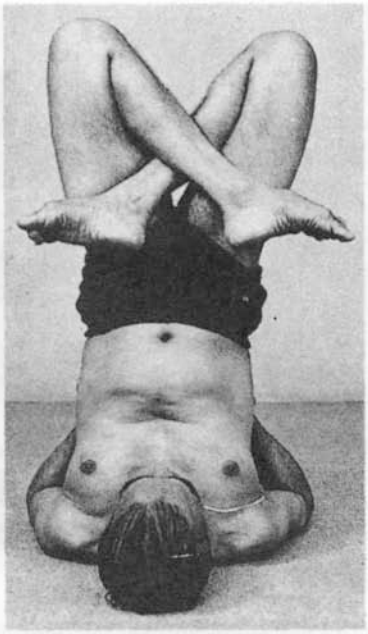

122

4. Uncross the legs, return to Sālamba Sarvāngāsana and repeat the pose by first placing the left foot over the right thigh and then the right foot over the left thigh. Stay for an equal length of time in all these positions as in the earlier ones.

49| *Piṇḍāsana in Sarvāngāsana*

FIVE*

Piṇḍa means embryo or foetus. In this variation of Sarvāngāsana which is a continuation of the earlier posture, the bent crossed legs are brought down until they rest on the head. The posture resembles that of an embryo in the womb, hence the name.

TECHNIQUE

1. From Ūrdhva Padmāsana in Sarvāngāsana (Plate 122), exhale, bend and lower the crossed legs from the hips towards the head.

2. Rest the legs over the head. (Plate 123)

3. Remain in this position from 20 to 30 seconds with normal breathing and go back to Ūrdhva Padmāsana in Sarvāngāsana.

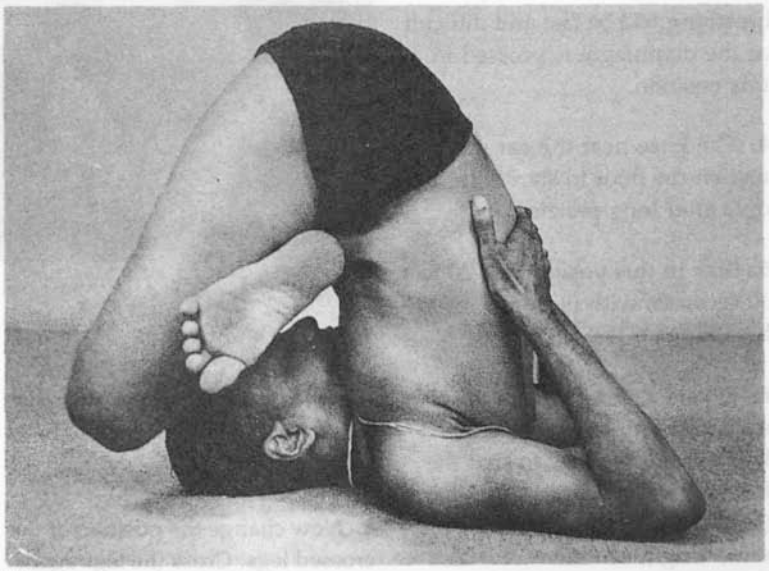

123

50| *Pārśva Piṇḍāsana in Sarvāngāsana*

EIGHT*

Pārśva means the side or flank. In this Piṇḍāsana variation of the earlier pose, both the bent knees are moved sideways and placed on the floor on the same side of the trunk. This is the lateral embryo pose in Sarvāngāsana.

TECHNIQUE

1. After staying in Pindāsana (Plate 123) turn the hips sideways to the right, exhale and lower both knees to the floor. The left knee should be by the side of the right ear. (Plate 124)

2. The left shoulder will be raised off the floor in the beginning. Push the shoulder against the floor and press the left hand firmly against the back. If this is not done, you will lose balance and roll over to one side.

3. Due to the lateral twist, breathing will be fast and difficult as the diaphragm is pressed in this position.

4. The knee near the ear will not rest on the floor to start with, but only after long practice.

5. Stay in this position for 20 to 30 seconds, with normal breathing.

6. Exhale, come up from the right side and move the crossed legs over to the left, so that the left foot will be near the left ear. (Plate 125) Stay here also for the same length of time.

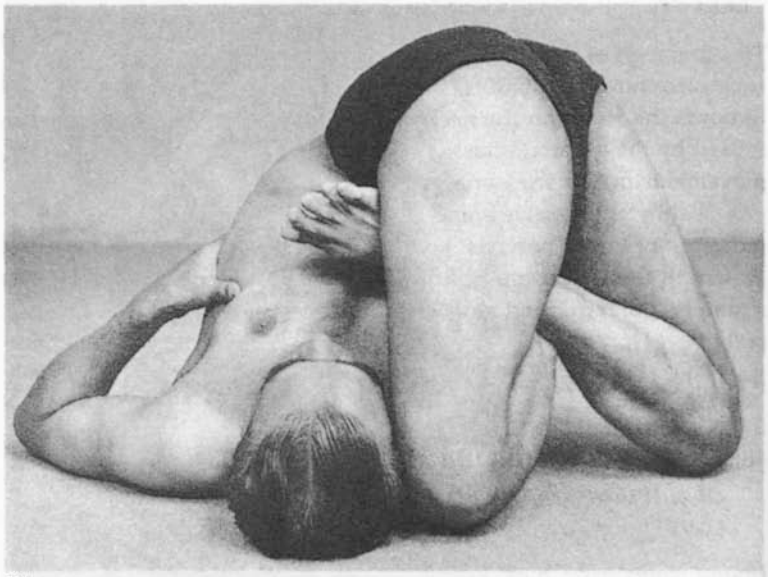

124

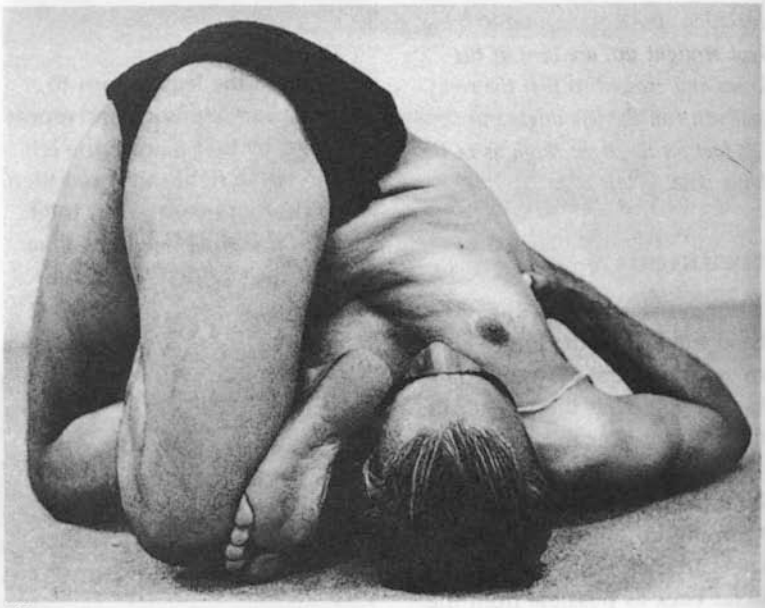

125

7. Go back to Ūrdhva Padmāsana. (Plate 122) Release the lotus pose by uncrossing the legs and return to Sālamba Sarvāngāsana.

8. Now change the position of the crossed legs. Cross the legs again by putting the left foot over the right thigh first and then the right foot over the left thigh instead of the other way as done earlier.

9. Repeat the movements again on both the sides as described earlier.

EFFECTS OF ŪRDHVA PADMĀSANA AND PĀRŚVA PINḌĀSANA MOVEMENTS IN SARVĀNGĀSANA

The change of crossing the legs brings equal pressure on both sides of the abdomen and colon and relieves constipation. For those suffering from chronic constipation a longer stay in Pārśva Pinḍāsana is recommended, and one minute on each side will prove most efficacious. Griping pain in the stomach is relieved by these poses.

Persons with extremely flexible knees, can easily perform these positions. It is, however, difficult for many people to cross the legs in Padmāsana. For them a longer stay in Pārśva Halāsana (Plate 116) – (there also the spine and trunk get a lateral twist but the legs remain straight) – is recommended.

In all these positions breathing at first will be fast and laboured. Try to maintain normal breathing.

NOTE

The spine is given the forward, lateral and backward movements in these variations of Sarvāngāsana. In Halāsana, Eka Pāda Sarvāngāsana, Karṇa Pīdāsana and Pinḍāsana the spine moves in the forward direction. In Pārśvaika Pāda Sarvānga, Pārśva Halāsana and Pārśva Pinḍāsana the spine moves laterally. In Setu Bandha it is given a backward movement. These movements tone the spine on all sides and keep it healthy.

It is related that in the Krita Age (the first Age of the Universe) a host of Dānavās (giants and demons) became invincible in battle under the leadership of Vṛtra and scattered the Devas (or Gods) in all directions. Realizing that they could not regain their power until Vṛtra was destroyed, the gods appeared before their Grandsire, Brahmā, the creator. Brahmā instructed them to consult Viṣṇu who asked them to obtain the bones of a sage called Dadhīcha, from which to make a demon-slaying weapon. The gods appeared before the sage and begged the boon according to Viṣṇu's advice. The sage renounced his body for the benefit of the gods. From the spine of Dadhīcha was fashioned Vajra, the thunderbolt, which Indra the king of the gods hurled and slew Vṛtra.

The story is symbolical. The Dānavās represent the tāmasic qualities in men and diseases. The Devas represent health, harmony and peace. To destroy the tāmasic qualities and the diseases due to them and to enjoy health and happiness, we have to make our spines strong as a thunderbolt like the spine of Dadhīcha. Then we shall enjoy health, harmony and happiness in abundance.

51| *Jaṭara Parivartanāsana*

FIVE*

Jaṭara means the stomach, the belly. Parivartana means turning or rolling about, turning round.

TECHNIQUE

1. Lie flat on the back on the floor. (Plate 98)

2. Exhale, raise both legs together until they are perpendicular to the floor. They should remain poker stiff, so do not bend them at the knees. (Plate 105)

3. Stretch out both arms sideways in line with the shoulders, so that the body resembles a cross.

4. Remain in this position for a few breaths. Then exhale, and move both the legs sideways (Plate 126) down towards the floor to the left until the toes of the left foot almost touch the finger-tips of the outstretched left hand. (Plate 127) Try and keep the back well on the floor. In the initial stages, the right shoulder will be lifted off the floor. To prevent this ask a friend to press it down, or catch hold of a heavy piece of furniture with the right hand when the legs are turned sideways to the left.

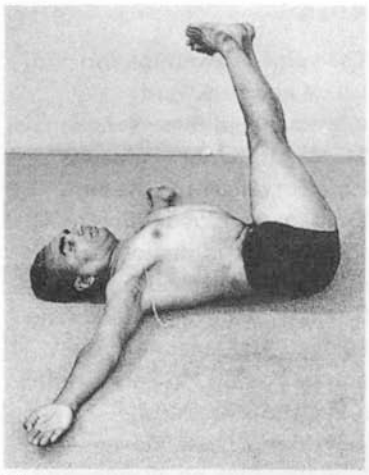

126

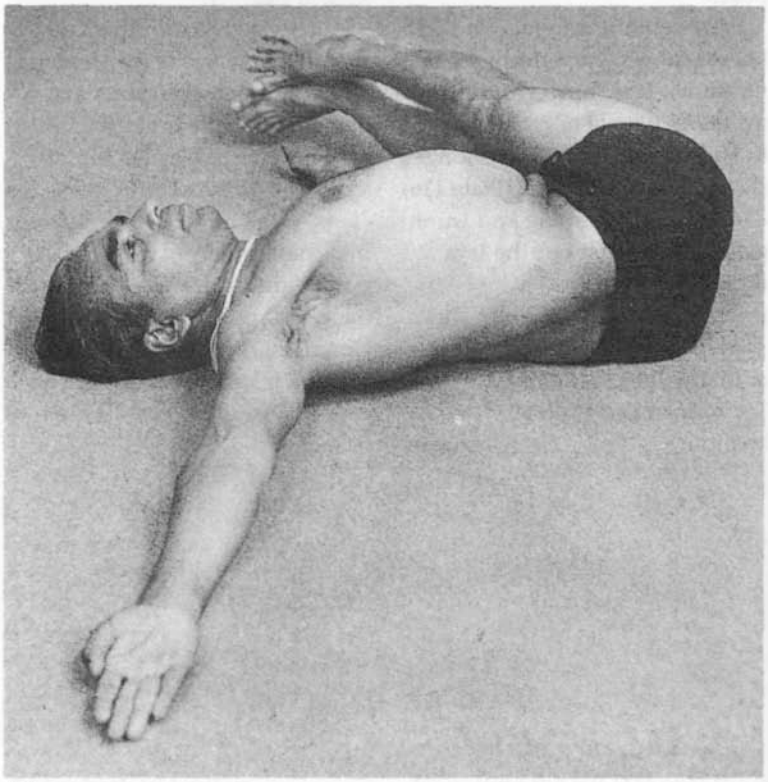

127

5. Both legs should go down together, the knees being kept tight throughout. As far as possible keep the lumbar portion of the back on the floor and turn the legs only from the hips. When the legs are near the outstretched left hand, move the abdomen to the right.

6. Stay in the pose for about 20 seconds, keeping the legs stiff throughout. Then move the still stiffened legs slowly back to the perpendicular with an exhalation.

7. Remain with the legs perpendicular for a few breaths and then repeat the movements by lowering the legs to the right and turning the abdomen to the left. (Plate 128) Stay here also for the same length of time and with an exhalation, come back to the perpendicular legs position and then gently lower the legs to the floor (Plate 98) and relax.

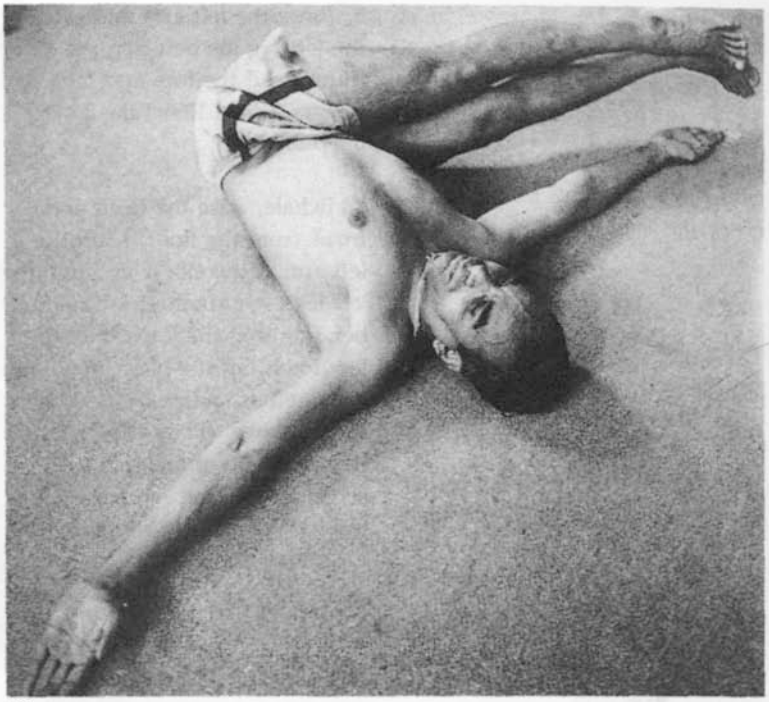

128

EFFECTS

This āsana is good for reducing excess fat. It tones and eradicates sluggishness of the liver, spleen and pancreas. It also cures gastritis and strengthens the intestines. By its regular practice all the abdominal organs are kept in trim. It helps to relieve sprains and catches in the lower back and the hip region.

52| *Supta Pādāngusṭhāsana*

THIRTEEN*

Supta means lying down. Pāda is the foot. Angusṭha means the big toe. Hence the name.

TECHNIQUE

1. Lie flat on the back, stretch both legs and keep the knees tight. (Plate 98)

2. Inhale, raise the left leg from the floor until it is perpendicular. Keep the right leg stretched fully on the floor and rest the right hand on the right thigh.

3. Raise the left arm and catch the left big toe between the thumb and the fore and middle fingers. (Plate 129) Take 3 or 4 deep breaths.

4. Exhale, raise the head and trunk from the floor, bend the left arm at the elbow and pull the left leg towards the head without bending it at the knee. Pull the leg down, lift the head and trunk up together and rest the chin on the left knee. (Plate 130) Stay in this position for about 20 seconds, keeping the right leg fully stretched straight along the floor while breathing normally.

5. Inhale, move the head and trunk back to the floor and the left leg back to the perpendicular. (Plate 129) This completes the first movement.

6. Exhale, release the toe grip, rest the left leg on the floor beside the right one and keep the left hand on the left thigh.

7. Take a few deep inhalations and then repeat on the right, substituting the word 'left' for the word 'right'.

EFFECTS

The legs will develop properly by the practice of this āsana. Persons suffering from sciatica and paralysis of the legs will derive great benefit from it. The blood is made to circulate in the legs and hips where the nerves are rejuvenated. The pose removes stiffness in the hip joints and prevents hernia. It can be practised by both men and women.

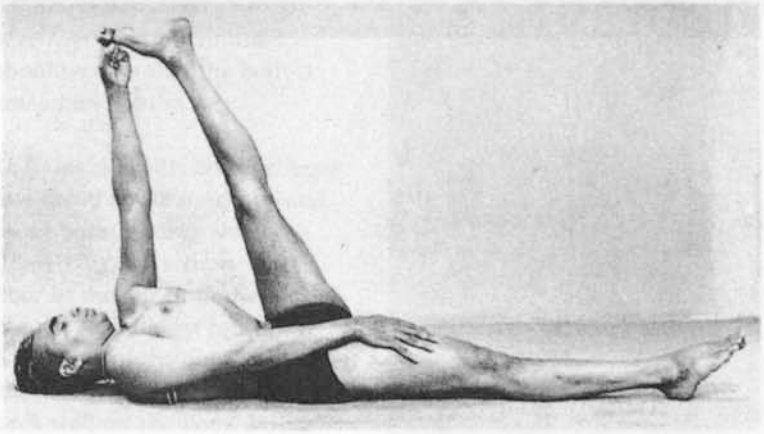

129

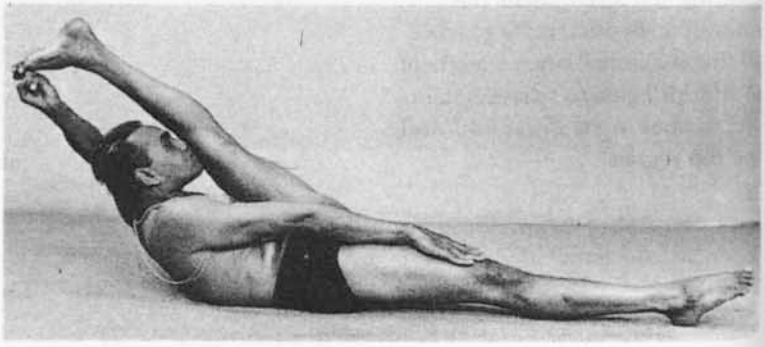

130

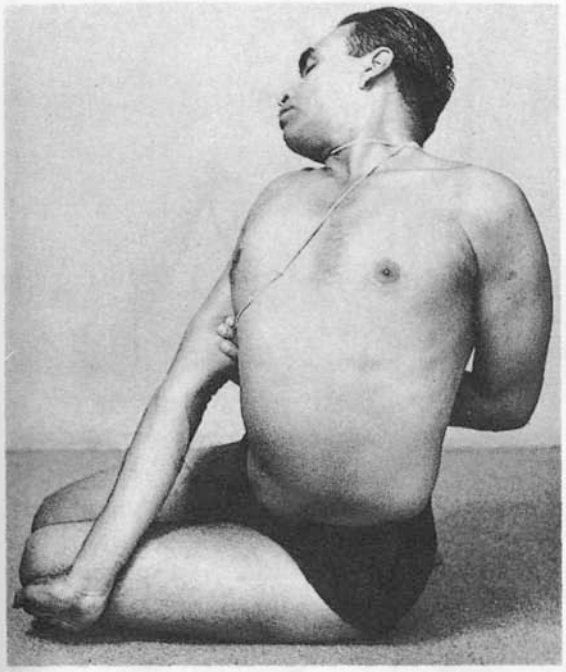

131

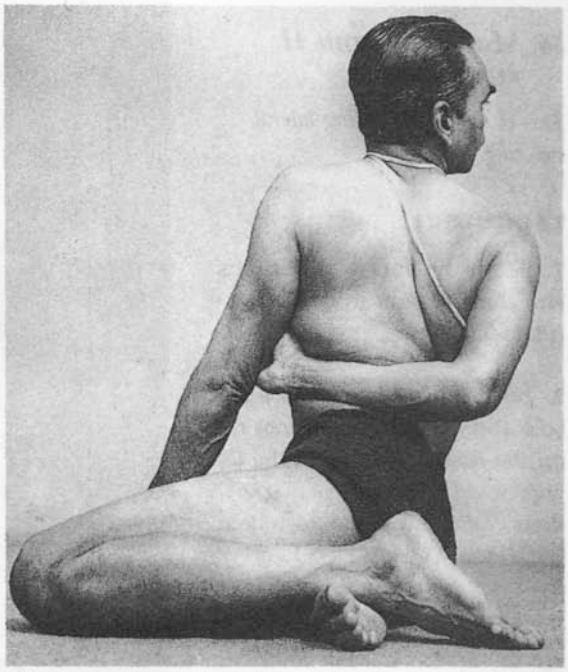

132

53| *Bharadvājāsana*

ONE*

Bharadvāja was the father of Droṇa, the military preceptor of the Kauravas and Pāṇḍavas, who fought the great war described in the Mahābhārata. This āsana is dedicated to Bharadvāja.

TECHNIQUE

1. Sit on the floor with the legs stretched straight in front. (Plate 35)

2. Flex the knees, move the legs back and bring both feet to the right side beside the hip.

3. Rest the buttocks on the floor, turn the trunk about 45 degrees to the left, straighten the right arm and place the right hand on the outer side of the left thigh near the left knee. Insert the right hand underneath the left knee, the palm touching the floor.

4. Exhale, turn the left arm from the shoulder behind the back, bend the left elbow and with the left hand clasp the right upper arm above the right elbow.

5. Turn the neck to the right and gaze over the right shoulder. (Plates 131 and 132)

6. Hold the position for half a minute with deep breathing.

7. Loosen the hand grip, straighten the legs and repeat the pose on the other side. Here, bring both the feet beside the left hip, turn the trunk to the right, straighten the left arm, place the left palm underneath the right knee and catch the left upper arm near the elbow with the right hand from behind the back. Stay there for an equal length of time.

EFFECTS

This simple āsana works on the dorsal and lumbar regions of the spine. People with very stiff backs find the other lateral twisting positions extremely difficult. This pose helps to make the back supple. People with arthritis will find it very beneficial.

54| *Marīchyāsana II*

TEN*

This is one of the sitting lateral twisting postures.

TECHNIQUE

1. Sit on the floor with the legs stretched straight in front. (Plate 35)

2. Bend the left knee, place the sole and heel of the left foot flat on the floor. The shin of the left leg should be perpendicular to the floor and the calf should touch the thigh. Place the left heel near the perineum. The inner side of the left foot should touch the inner side of the outstretched right thigh.

3. With an exhalation, turn the spine about 90 degrees to the left, so that the chest goes beyond the bent left thigh and bring the right arm over the left thigh. (Plate 133)

4. Place the right shoulder beyond the left knee and stretch the right arm out forwards by turning the spine still more to the left and stretching the region at the back of the right floating ribs. (Plate 134) Take two breaths.

5. With an exhalation, twist the right arm round the left knee, flex the right elbow and place the right wrist at the back of the waist. Inhale and hold the pose.

6. Exhale deeply and turn the left arm from the shoulder behind the back. Either clasp the left hand behind the back with the right hand or vice versa. (Plates 135 and 136) In the beginning, one finds it difficult to twist the trunk sideways, but with practice, the armpit touches the bent knee. After one has twisted the arm round the knee, one also finds it difficult to clasp the fingers of one hand with the other. Gradually one learns to clasp the fingers, then the palm and lastly to hold the hand at the wrist behind the back.

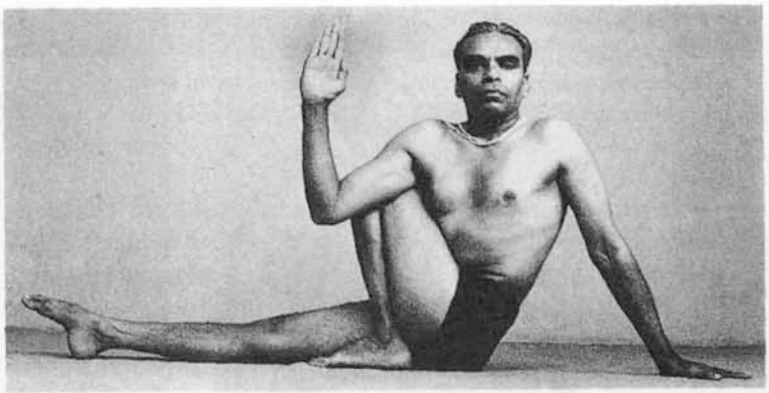

133

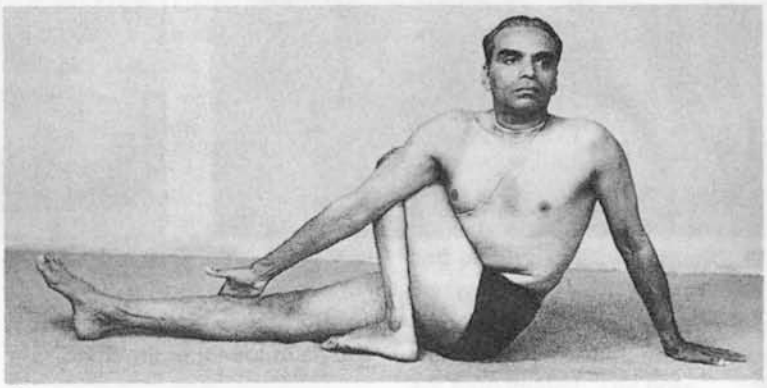

134

7. The right arm should lock the left bent knee tightly. There should be no space between the right armpit and the bent left knee.

8. After clasping the hands at the back, turn the spine still more to the left by tugging at the clasped hands.

9. The whole of the outstretched right leg should remain straight and securely on the floor, but you will not be able to achieve this to start with. Tighten the muscles of the outstretched thigh so that the knee-cap is pulled up towards the thigh and also tighten the muscles of the calf of the outstretched leg. Then the leg will remain firm and extend on the floor.

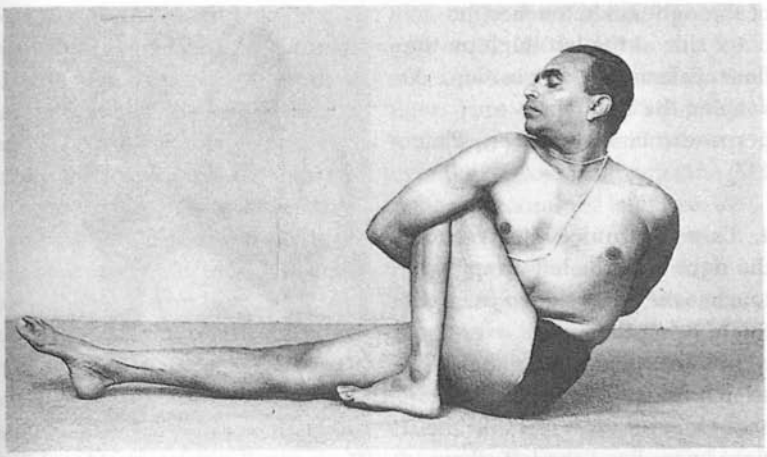

135

10. Stay in this position from half a minute to a minute with normal breathing. The neck may be turned either way to gaze at the toes of the extended leg on the floor or to look over the shoulder.

11. Unclasp the hands at the back and turn the trunk back to its original position. Lower the bent leg and extend it fully on the floor.

12. Then repeat the pose on the other side. This time bend the right knee and place the right foot firmly on the floor so that the right heel touches the perineum and the inner side of the right foot touches the outstretched left thigh. Turn the trunk about 90 degrees to the right so that the left armpit touches the bent right knee. With an exhalation, twist the left arm round the right knee and bring the left hand to the back of the waist. Then throw the right arm behind the back from the shoulder and flexing the right elbow, bring the right hand to the left hand and clasp them. Turn still more to the right and gaze at either the toes of the outstretched left leg or over the right shoulder. Stay on this side also for the same length of time. Unclasp the hands, turn the trunk back to normal, stretch the right leg on the floor and relax.

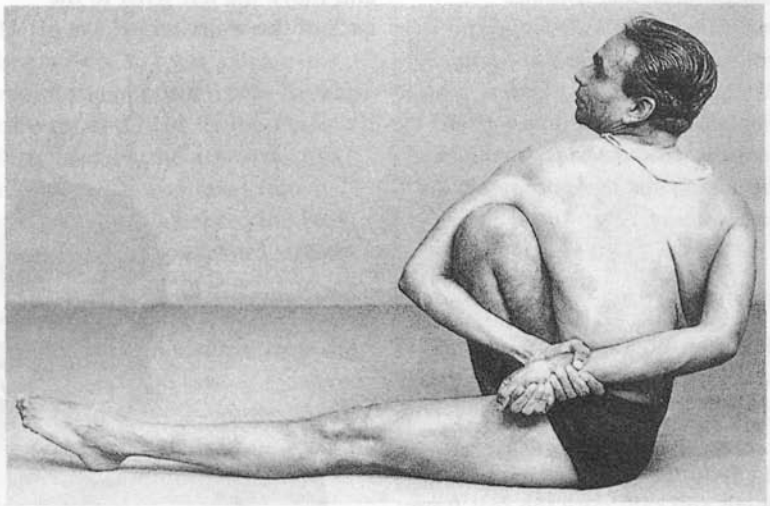

136

EFFECTS

By the regular practice of this āsana, splitting backaches, lumbago and pains in the hips disappear rapidly. The liver and the spleen are contracted and so are toned and cease to be sluggish. The muscles of the neck gain power. Sprains in the shoulder and displacement of the shoulder joints are relieved and the shoulder movements become free. The intestines also benefit from this āsana. Its effects will be less on lean persons, for whom there are better poses described later. It also helps to reduce the size of the abdomen.

55| *Ardha Matsyendrāsana*

EIGHT*

Ardha means half. In the Haṭha Yoga Pradīpikā, *Matsyendra is mentioned as one of the founders of Haṭha Vidyā. It is related that once Lord Śiva went to a lonely island and explained to his consort Pārvati the mysteries of Yoga. A fish near the shore heard everything with concentration and remained motionless while listening. Śiva, realizing that the fish had learnt Yoga, sprinkled water upon it, and immediately the fish gained divine form and became Matsyendra (Lord of the Fishes) and thereafter spread the knowledge of Yoga.*

TECHNIQUE

1. Sit on the floor, with the legs stretched straight in front. (Plate 35)

2. Bend the left knee and join the thigh and calf; raise the seat from the floor, place the left foot under the buttocks and sit on the left foot so that the left heel rests under the left buttock. The foot used as the seat should be kept horizontal on the floor, the outer side of the ankle and the little toe of the foot resting on the ground. If the foot is not so placed, it will be impossible to sit on it. Balance securely in this position.

3. Then bend the right knee and lifting the right leg from the floor, place it by the outer side of the left thigh so that the outer side of the right ankle touches the outer side of the left thigh on the floor. Balance in this position, keeping the right shin perpendicular to the floor. (Plate 137)

4. Turn the trunk 90 degrees to the right until the left armpit touches the outer side of the right thigh. Bring the armpit over the right knee. (Plate 138) Exhale, stretch the left arm from the shoulder and twist it round the right knee. Bend the left elbow and move the left wrist to the back of the waist.

137

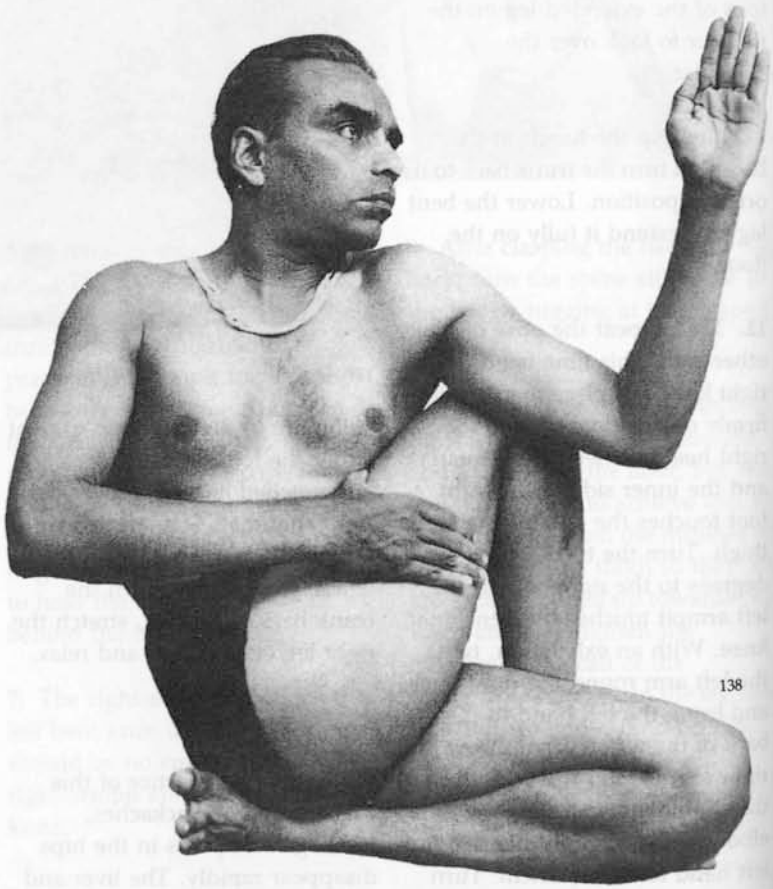

138

5. The left arm should lock the bent right knee tightly and there should be no space between the left armpit and the bent right knee. To achieve this, exhale and move the trunk forward. Stay in this position and take 2 breaths.

6. Now exhale deeply and swing back the right arm from the shoulder, bend the right elbow, move the right hand behind the waist and either clasp it with the left hand or vice versa. At first you will be able to catch a finger or two. With practice it will be possible to catch the palms and then the wrists behind the back.

7. The neck may be turned to the left and the gaze directed over the left shoulder (Plate 139), or to the right, and the gaze fixed at the centre of the eyebrows. (Plate 140) The spinal twist will be greater if the neck is turned to the left than when to the right.

8. As the diaphragm is squeezed by the spinal twist, the breathing will at first become short and fast. Do not be nervous. After some practice the pose can be held from half a minute to a minute with normal breathing.

9. Release the hands, remove the right foot from the floor and straighten the right and then the left leg.

10. Repeat the pose on the other side and hold it for the same length of time. Here, bend the right leg and sit on the right foot so that the right heel is under the right buttock. Place the left leg over the right leg and rest the left foot on the floor so that the outer side of the left ankle touches the outer side of the right thigh on the floor. Turn the trunk 90 degrees to the left, placing the right armpit over the left knee and twist the right arm round the left knee. Flex the right elbow and move the right hand behind the waist. Hold the pose and take 2 breaths. Again exhale completely and swing the left arm back from the shoulder, bend the left elbow and clasp the hands behind the back at the wrist. Then release and relax.

11. In the beginning it may not be possible to twist either arm round the opposite knee. In that case try and hold the opposite foot, keeping the arm straight at the elbow. It also takes time to clasp the hands behind the back. Gradually, the backward stretch of the arms will increase, and one will be able to catch at first the fingers, next the palms, then the wrist and as the pose is mastered even the forearms above the wrists. Beginners who find it difficult to sit on the foot can sit on the floor.

139

140

EFFECTS

By the practice of this āsana, one derives the benefits mentioned under Marīchyāsana III. (Posture 54 and Plate 135) But here as the range of movement is more intensified, the effects will also be greater. In Marīchyāsana III the upper part of the abdomen is squeezed. Here the lower part of the abdomen has the benefit of the exercise. The prostate and bladder are not enlarged if one practises regularly.

56| *Ūrdhva Dhanurāsana*

SEVEN*

Ūrdhva means upwards. Dhanu means a bow. In this posture the body is arched back and supported on the palms and soles.

TECHNIQUE

1. Lie flat on the back on the floor. (Plate 48)

2. Bend and raise the elbows over the head, and place the palms under the shoulders. The distance between the palms should not be wider than the shoulders and the fingers should point towards the feet.

3. Bend and raise the knees, then bring the feet nearer until they touch the hips. (Plate 141)

4. Exhale, raise the trunk and rest the crown of the head on the floor. (Plate 142) Take two breaths.

5. Now exhale, lift the trunk and head and arch the back so that its weight is taken on the palms and the soles. (Plate 143)

6. Stretch the arms from the shoulders until the elbows are straightened, at the same time pulling the thigh muscles up. (Plate 144)

7. To get a better stretch, exhale and pull the thigh muscles still higher by lifting the heels off the floor. (Plate 145) Extend the chest, stretch up the sacral region of the spine until the abdomen is taut as a drum and then lower the heels to the floor, maintaining the stretch of the spine.

8. Remain in this position from half a minute to a minute, with normal breathing.

9. With an exhalation, lower the body to the floor by bending the knees and elbows.

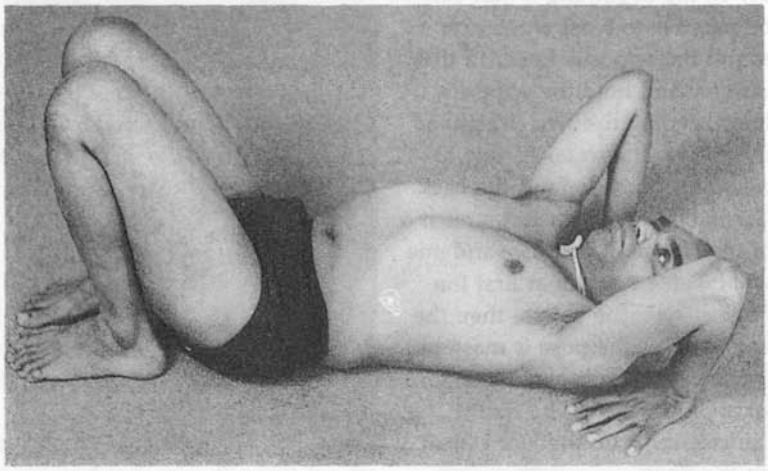

141

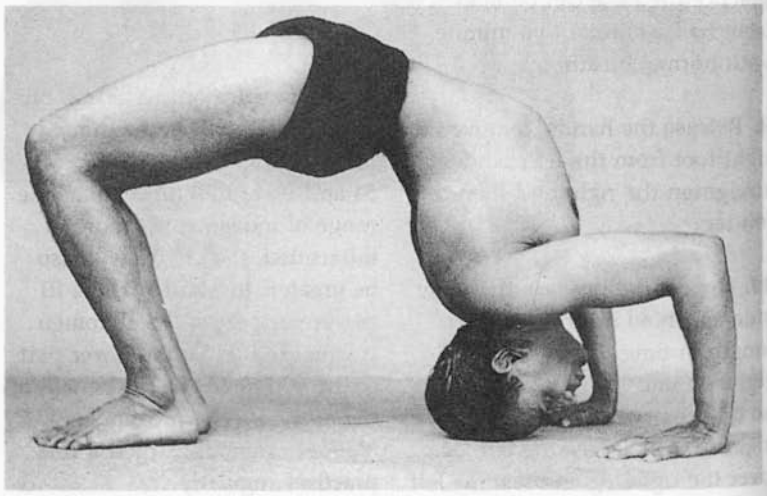

142

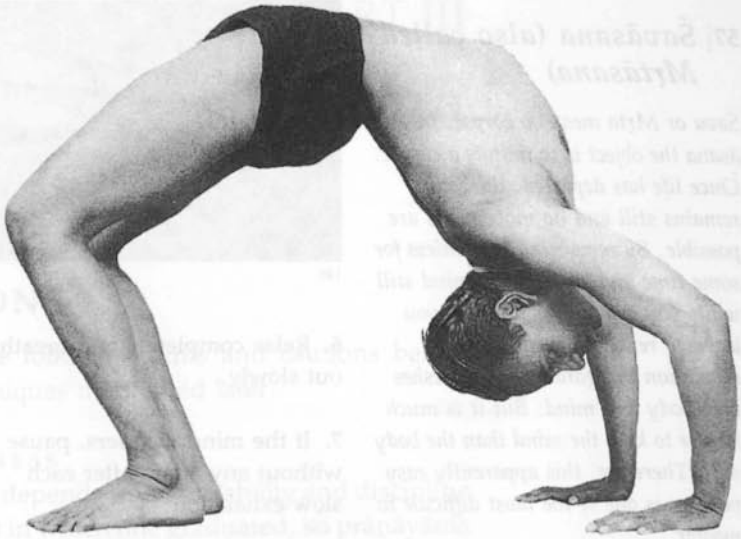

143

144

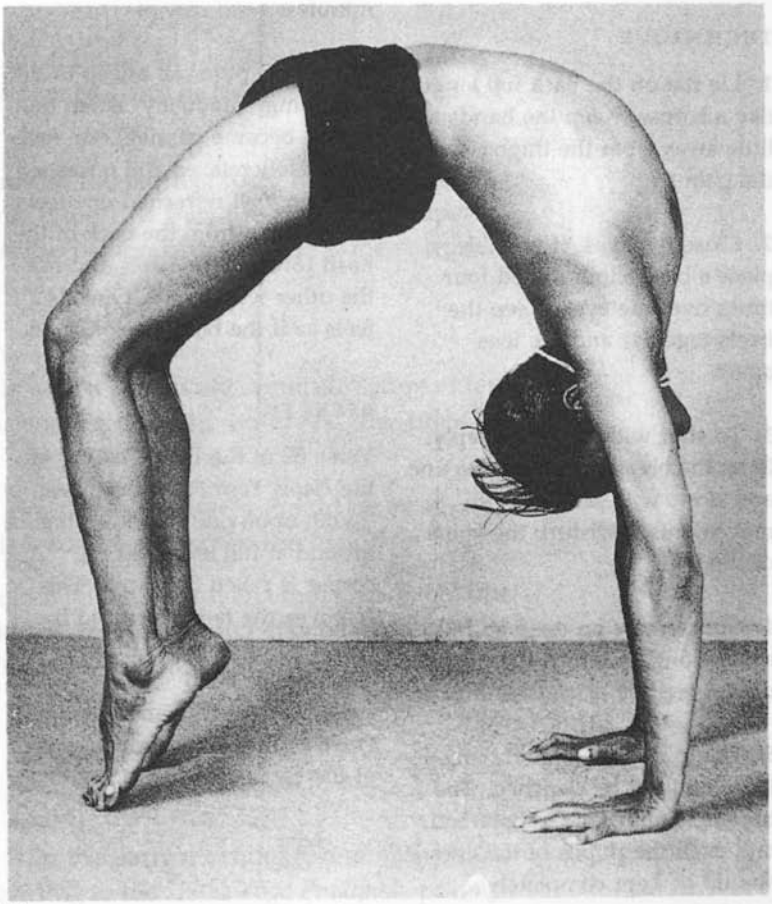

145

EFFECTS

This āsana is the beginning of the advanced and difficult back-bending poses. It tones the spine by stretching it fully and keeps the body alert and supple. The back feels strong and full of life. It strengthens the arms and wrists and has a very soothing effect on the head. It gives one great vitality, energy and a feeling of lightness.

57| *Śavāsana (also called Mṛtāsana)*

Śava or Mṛta means a corpse. In this āsana the object is to imitate a corpse. Once life has departed, the body remains still and no movements are possible. By remaining motionless for some time and keeping the mind still while you are fully conscious, you learn to relax. This conscious relaxation invigorates and refreshes both body and mind. But it is much harder to keep the mind than the body still. Therefore, this apparently easy posture is one of the most difficult to master.

146

TECHNIQUE

1. Lie flat on the back full length like a corpse. Keep the hands a little away from the thighs, with the palms up.

2. Close the eyes. If possible place a black cloth folded four times over the eyes. Keep the heels together and the toes apart.

3. To start with breathe deeply. Later the breathing should be fine and slow, with no jerky movements to disturb the spine or the body.

4. Concentrate on deep and fine exhalations, in which the nostrils do not feel the warmth of breath.

5. The lower jaw should hang loose and not be clenched. The tongue should not be disturbed, and even the pupils of the eyes should be kept completely passive.

6. Relax completely and breath out slowly.

7. If the mind wanders, pause without any strain after each slow exhalation.

8. Stay in the pose from 15 to 20 minutes.

9. One is apt to fall asleep in the beginning. Gradually, when the nerves become passive, one feels completely relaxed and refreshed.

When well refreshed one feels energy flow from the back of the head towards the heels and not the other way round. One also feels as if the body is elongated.

EFFECTS

Verse 32 of the First Chapter of the *Haṭha Yoga Pradīpikā* states: 'Lying upon one's back on the ground at full length like a corpse is called Śavāsana. This removes the fatigue caused by the other āsanas and induces calmness of mind.'

Mṛtāsana is thus described in verse 11 of the Second Chapter of the *Gheraṇḍa Saṁhitā*:

'Lying flat on the ground (on one's back) like a corpse is called Mṛtāsana. This posture destroys fatigue, and quiets the agitation of the mind.'

'The mind is the lord of the Indriyas (the organs of senses); the Praṇa (the Breath of Life) is the lord of the mind.' 'When the mind is absorbed it is called Mokṣa (final emancipation, liberation of the soul); when Prāṇa and Manas (the mind) have been absorbed, an undefinable joy ensues.'

Verses 29 and 30, chapter IV,
Haṭha Yoga Pradīpikā.

To tame Prāṇa depends upon the nerves. Steady, smooth, fine and deep breathing without any jerky movements of the body soothes the nerves and calms the mind. The stresses of modern civilization are a strain on the nerves for which Śavāsana is the best antidote.

PART III
Prāṇāyāma

HINTS AND CAUTIONS

Read and digest thoroughly the following hints and cautions before attempting the prāṇāyāma techniques mentioned later.

QUALIFICATIONS FOR FITNESS

1. Just as post-graduate training depends upon the ability and discipline acquired in mastering the subject in which one graduated, so prāṇāyāma training demands mastery of āsanas and the strength and discipline arising therefrom.

2. The fitness of the aspirant for training and advancement in prāṇāyāma is to be gauged by an experienced Guru or teacher and his personal supervision is essential.

3. Pneumatic tools can cut through the hardest rock. In Prāṇāyāma the yogi uses his lungs as pneumatic tools. If they are not used properly, they destroy both the tool and the person using it. The same is true of prāṇāyāma.

CLEANLINESS AND FOOD

4. One does not enter a temple with a dirty body or mind. Before entering the temple of his own body, the yogi observes the rules of cleanliness.

5. Before starting prāṇāyāma practices the bowels should be evacuated and the bladder emptied. This leads to comfort in the bandhas.

6. Preferably prāṇāyāma should be practised on an empty stomach, but if this is difficult, a cup of milk, tea, coffee or cocoa may be taken. Allow at least six hours to elapse after a meal before practising prāṇāyāma.

7. Light food may be taken half an hour after finishing prāṇāyāma practices.

TIME AND PLACE

8. The best time for practice is in the early morning (preferably before sunrise) and after sunset. According to the *Haṭha Yoga Pradīpikā*, prāṇāyāma should be practised four times a day, in the early morning, noon,

evening and midnight, with 80 cycles at a time (chapter II, verse 11). This is hardly possible in the fast modern age. What is therefore recommended is to practice at least 15 minutes a day, but the 80 cycles are for intensely devoted practitioners, and not for the average householder.

9. The best seasons in which to start the practice are spring and autumn when the climate is equable.

10. Prāṇāyāma should be done in a clean airy place, free from insects. Since noise creates restlessness practice during quiet hours.

11. Prāṇāyāma should be practised with determination and regularity at the same time and place and in the same posture. Variation is permissible only in the type of prāṇāyāma practised, that is to say, if Sūrya Bhedana Prāṇāyāma is done one day, Śitalī may be done the next day and Bhastrikā be done on the third day. Nāḍī Sodhana Prāṇāyāma, however, should be practised daily.

POSTURE

12. Breathing in prāṇāyāma practices is done through the nose only, except in Śitalī and Śitakāri.

13. Prāṇāyāma is best done sitting on the floor on a folded blanket. The postures suitable are Siddhāsana, Vīrāsana, Padmāsana and Baddhakoṇāsana. Any other sitting posture may be taken provided the back is kept absolutely erect from the base of the spine to the neck and perpendicular to the floor. Some types, however, may be done in a reclining position as detailed later.

14. During practice no strain should be felt in the facial muscles, eyes and ears, or in the neck muscles, shoulders, arms, thighs and feet. The thighs and arms should be relaxed deliberately since they are unconsciously tensed during prāṇāyāma.

15. Keep the tongue passive or saliva will accumulate in the mouth. If it does, swallow it before exhalation (rechaka) and not while holding the breath (kumbhaka).

16. During inhalation and retention the rib cage should expand both forwards and sideways, but the area below the shoulder-blades and armpits should only expand forwards.

17. To start with there will be perspiration and trembling which will disappear in course of time.

18. In all the prāṇāyāma practices done in a sitting posture, the head should hang down from the nape of the neck, the chin resting in the

notch between the collar-bones on the top of the breast-bone. This chin-lock or Jālandhara Bandha should be used except where specifically stated in the techniques hereafter given.

19. Keep the eyes closed throughout as otherwise the mind will wander after outside objects and be distracted. The eyes, if kept open, will feel a burning sensation, and irritability.

20. No pressure should be felt inside the ear during the practice of prāṇāyāma.

21. The left arm is kept straight, the back of the wrist resting on the left knee. The forefinger is bent towards the thumb, its tip touching the tip of the thumb. This is the Jñāna Mudrā described later in the technique.

22. The right arm is bent at the elbow and the hand is kept on the nose to regulate the even flow of breath and to gauge its subtlety. This is felt through the tips of the ring and little fingers which control the left nostril and through the tip of the thumb which controls the right nostril. Details of the right hand position are discussed in the technique. In some methods of prāṇāyāma both the hands rest on the knees in the Jñāna Mudrā.

23. When a baby learns to walk by itself, the mother remains passive bodily, but alert mentally. In an emergency, as when the child stumbles, her body springs into action to save it from a fall. So also, in the practice of prāṇāyāma the brain is kept passive but alert. Whenever the organs of the body fail to work properly, the watchful brain sends messages of warning. The ear is told to listen for the proper sound of the breath (which is described below). The hand and nose are told to observe the sensitivity of the breath flowing through the nasal passages.

24. It may be asked that if the brain is required to send warnings to the senses, how can one concentrate on prāṇāyāma? A painter absorbed in his work notes various details like perspective and composition, the colour tones and shades, the foreground and background and the strokes of the paint-brush all at once. A musician playing a melody watches his finger movements and sound patterns, the tuning of the instrument and its pitch. Though the artist and the musician are both observing and correcting the details, they are concentrating on their work. So also the yogi observes details like time, posture and an even breath rhythm, and is alert and sensitive to the flow of prāṇa within him.

25. As a careful mother teaches her child to walk carefree, so the careful mind of the yogi teaches the senses to be carefree. By continued practice of prāṇāyāma the senses become free of obsession for the things they once pined for.

26. Each should measure his own capacity when doing prāṇāyāma and not exceed it. This may be gauged as follows: suppose one can with comfort inhale and exhale for 10 seconds each in rhythmic cycles for a given length of time, say 5 minutes. If there is any change in the rhythm in which the period of inhalation decreases, to say 7 or 8 seconds, one has reached one's capacity. To go beyond this point, strains the lungs unduly and brings in its wake a host of respiratory diseases.

27. Faulty practice puts undue stress on the lungs and diaphragm. The respiratory system suffers and the nervous system is adversely affected. The very foundation of a healthy body and sound mind is shaken by faulty practice of prāṇāyāma. Forceful and strained inhalation or exhalation is wrong.

28. Evenness of breathing leads to healthy nerves and so to evenness of mind and temper.

29. Āsanas should never be practised immediately after prāṇāyāma. If prāṇāyāma is done first, allow an hour to elapse before starting āsanas, for the nerves which are soothed in prāṇāyāma are liable to be ruffled by the bodily movement of the āsanas.

30. Prāṇāyāma, however, may be done not less than 15 minutes after mild practice of āsanas.

31. Strenuous āsanas cause fatigue. When exhausted do not practise prāṇāyāma in any sitting posture, as the back cannot stay erect, the body trembles and the mind becomes disturbed. Deep breathing as in Ujjāyī done in a reclining position relieves fatigue.

32. When deep, steady and long breathing cannot be maintained rhythmically, stop. Do not proceed further. The rhythm should be gauged from the nasal sound produced in inhalation ('ssssssa' which sounds like a leak in a cycle tube) and exhalation (the aspirate 'huuuuuuuuum' sound). If the volume of the sound is reduced, stop.

33. Try to achieve an even ratio in inhalation (puraka) and exhalation (rechaka). For example, if one is for 5 seconds during a given continuous cycle, the other should be for the same time.

34. The Ujjāyī and Nāḍī Śodhana types of prāṇāyāma are the most beneficial ones which can be practised by pregnant women, preferably in Baddhakoṇāsana. During pregnancy, however, the breath should never be held without the guidance of an experienced teacher.

35. After completing any prāṇāyāma practice always lie down on the back like a corpse in Śavāsana (Plate 146) for at least 5 to 10 minutes in silence. The mind should be completely switched off and every limb and

sense organ completely passive as if dead. Śavāsana after prāṇāyāma refreshes both the body and the mind.

KUMBHAKAS

36. Thorough mastery of inhalation (puraka) and exhalation (rechaka) is essential before any attempt is made to learn antara kumbhaka (retention following inhalation).

37. Bāhya kumbhaka (restraint following exhalation) should not be tried until antara kumbhaka has become natural.

38. During the practice of kumbhaka there is a tendency to draw in air as well as to tighten and loosen the diaphragm and abdominal organs for the sake of increasing the period of retention. This is unconscious and unintentional. Care should be taken to avoid it.

39. If it is found difficult to hold the breath (khumbaka) after each inhalation or exhalation, do some cycles of deep breathing and then practise kumbhakas. For instance, 3 cycles of deep breathing may be followed by one cycle of kumbhaka. Then there should be another 3 cycles of deep breathing followed by a second cycle of kumbhaka, and so on.

40. If the rhythm of inhalation or exhalation is disturbed by holding the breath, lessen the duration of kumbhaka.

41. Persons suffering from eye or ear trouble (like glaucoma and pus in the ear) should not attempt to hold the breath.

42. Sometimes constipation occurs in the initial stages following upon the introduction of kumbhaka. This is temporary and will disappear in due course.

43. The normal rate of breaths per minute is 15. This rate increases when the body is upset by indigestion, fever, cold and cough, or by emotions like fear, anger or lust. The normal rate of breathing is 21,600 breaths inhaled and exhaled every 24 hours. The yogi measures his span of life not by the number of days, but of breaths. Since breathing is lengthened in prāṇāyāma, its practice leads to longevity.

44. Continuous practice of prāṇāyāma will change the mental outlook of the practitioner and reduce considerably the craving of his senses for worldly pleasures like smoking, drinking and sexual indulgence.

45. In the practice of prāṇāyāma the senses are drawn inwards and in the silence of the kumbhaka the aspirant hears his inner voice calling: 'Look within! The source of all happiness is within!' This also prepares him for the next stage of yoga, pratyāhāra, which leads to freedom from the domination and tyranny of the senses.

46. Since the eyes are kept closed throughout the practice of prāṇāyāma, the passage of time is noted by the mental repetition (japa) of a sacred word or name. This repetition of the sacred words or names is the seed (bīja) planted in the yogi's mind. This seed grows and makes him fit for dhyāna or concentration, the sixth stage of Yoga. Ultimately it produces the fruit of samādhi, where there is experience of full consciousness and supreme joy, where the yogi merges with the Maker of the Universe and feels what he can never express – yet cannot entirely conceal. Words fail to convey the experience adequately, for the mind cannot find words with which to describe it. It is a feeling of that peace which passeth all understanding.

TECHNIQUE AND EFFECTS OF PRĀṆĀYĀMA

58| *Ujjāyī Prāṇāyāma*

The prefix ud attached to verbs and nouns, means upwards or superiority in rank. It also means blowing or expanding. It conveys the sense of pre-eminence and power.

Jaya means conquest, victory, triumph or success. Looked at from another viewpoint it implies restraint or curbing.

Ujjāyi is the process in which the lungs are fully expanded and the chest puffed out like that of a proud conqueror.

TECHNIQUE

1. Sit in any comfortable position like Padmāsana (Plate 53), Siddhāsana (Plate 38) or Vīrāsana (Plate 43).

2. Keep the back erect and rigid. Lower the head to the trunk. Rest the chin in the notch between the collar-bones just above the breast-bone. (This is the Jālandhara Bandha. Jāla means a net, web, lattice or a mesh.)

3. Stretch the arms out straight and rest the back of the wrists on the knees. Join the tips of the index fingers to the tips of the thumbs, keeping the other fingers extended. (This position or gesture of the hand is known as the Jñāna Mudrā, the symbol or seal of knowledge. The index finger symbolizes the individual soul and the thumb the Universal Soul. The union of the two symbolizes knowledge.)

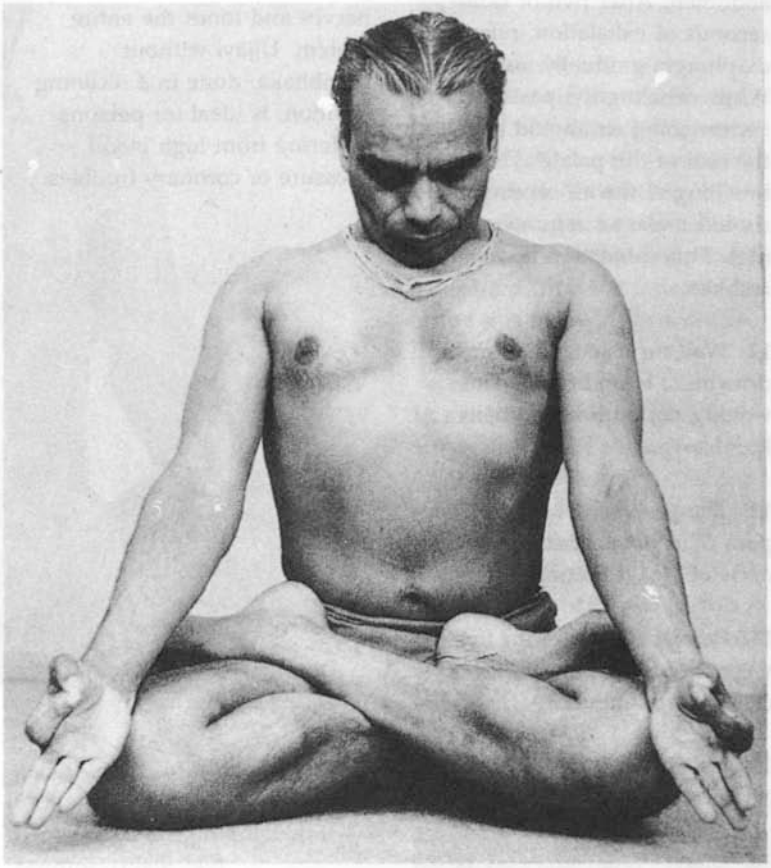

147

4. Close the eyes and look inwards. (Plate 147)

5. Exhale completely.

6. Now the Ujjāyi method of breathing begins.

7. Take a slow, deep steady breath through both nostrils. The passage of the incoming air is felt on the roof of the palate and makes a sibilant sound (sa). This sound should be heard.

8. Fill the lungs up to the brim. Care should be taken not to bloat the abdomen in the process of inhalation. (Observe this in all types of Prāṇāyāma.) This filling up is called puraka (inhalation).

9. The entire abdominal area from the pubes up to the breast-bone should be pulled back towards the spine.

10. Hold the breath for a second or two.

11. Exhale slowly, deeply and steadily, until the lungs are completely empty. As you begin to exhale, keep your grip on the abdomen. After two or three seconds of exhalation, relax the diaphragm gradually and slowly. While exhaling the passage of the outgoing air should be felt on the roof of the palate. The brushing of the air on the palate should make an aspirate sound (ha). This exhalation is called rechaka.

12. Wait for a second before drawing a fresh breath. This waiting period is called bāhya kumbhaka.

13. The process described from para. 7 to para. 12 completes one cycle of Ujjāyi Prāṇāyāma.

14. Repeat the cycles for five to ten minutes keeping the eyes closed throughout.

15. Lie on the floor for Śāvāsana. (Plate 146)

16. Ujjāyi Prāṇāyāma may be done without the Jālandhara Bandha even while walking or lying down. This is the only prāṇāyāma which can be done at all times of the day and night.

EFFECTS

This type of prāṇāyāma aerates the lungs, removes phlegm, gives endurance, soothes the nerves and tones the entire system. Ujjāyi without kumbhaka, done in a reclining position, is ideal for persons suffering from high blood pressure or coronary troubles.

59| *Sūrya Bhedana Prāṇāyāma*

Sūrya is the sun. Bhedana is derived from the root bhid meaning to pierce, to break or pass through.

In Sūrya Bhedana Prāṇāyāma, the breath is inhaled through the right nostril. In other words the prāṇa passes through the Piṅgāla or Sūrya nāḍī. A kumbhaka is then performed and the breath is then exhaled through the left nostril which is the path of the Idā nāḍī.

TECHNIQUE

1. Sit in any comfortable position like Padmāsana (Plate 53), Siddhāsana (Plate 38) or Vīrāsana (Plate 43).

2. Keep the back erect and rigid. Lower the head to the trunk. Rest the chin in the notch between the collar-bones just above the breast-bone. (This is Jālandhara Bandha.)

3. Stretch the left arm. Rest the back of the left wrist on the left knee. Perform Jñāna with the left hand (as stated in stage 3 of the technique of Ujjāyi).

4. Bend the right arm at the elbow. Bend the index and middle fingers towards the palm, keeping them passive. Bring the ring and little fingers towards the thumb. (Plate 148)

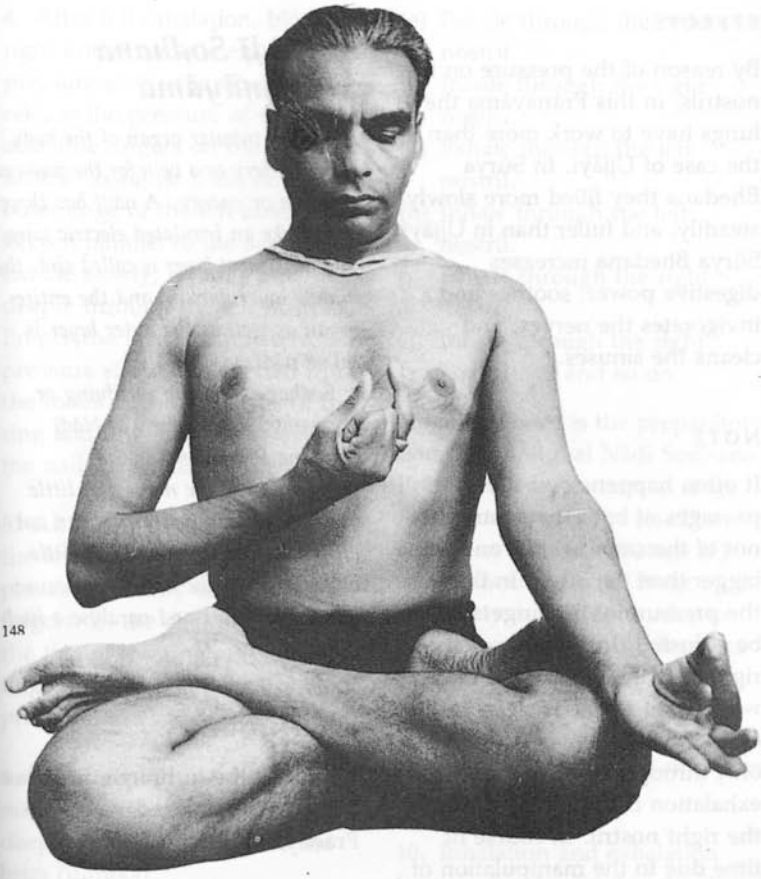

148

5. Place the right thumb on the right side of the nose just below the nasal bone, the ring and little fingers on the left side of the nose just below the nasal bone, just above the curve of the fatty tissue of the nostrils above the upper jaw.

6. Press the ring and the little finger to block the left side of the nose completely.

7. With the right thumb, press the fatty tissue on the right side so as to make the outer edge of the right nostril parallel to the lower edge of the cartilage of the septum.

8. The right thumb is bent at the top joint and the tip of the thumb is placed at a right angle to the septum. (Plate 149)

9. Now inhale slowly and deeply controlling the aperture of the right nostril with the tip of the thumb nearer the nail. Fill the lungs to the brim (puraka).

10. Then block the right nostril so that both are now blocked.

11. Hold the breath for about 5 seconds (antara kumbhaka).

12. Keeping the right nostril completely blocked, open the left nostril partially and exhale through it slowly and deeply (rechaka).

13. During the exhalation regulate the rhythmic flow of air from the left nostril by adjusting pressure with the ring and little fingers, so that the outer edge of the left nostril is kept parallel to the septum. The pressure should be exerted from the inner sides of the tips of the fingers (away from the nails).

14. This completes one cycle of Sūrya Bhedana Prāṇāyāma. Continue with more cycles at a stretch from 5 to 10 minutes, according to capacity.

15. All the inhalations in Sūrya Bhedana are from the right nostril and all the exhalations from the left nostril.

149

16. Throughout the process, the passage of air is felt at the tips of the fingers and the thumbs and in the nasal membranes where pressure is applied. The passage of air makes a sound similar to air escaping from a cycle tube. This sound should be maintained constant throughout by varying pressure on the nostrils.

17. The eyes, temples, eyebrows and the skin on the forehead should remain completely passive and show no sign of strain.

18. The mind should be absorbed completely in listening to the proper sound of the passage of air and in maintaining a proper rhythm in breathing.

19. Each inhalation and exhalation should last for the same length of time.

20. The inhalation and the exhalation should not be forced. An even and slow rhythm should be maintained throughout.

21. Lie down in Śavāsana after completing prāṇāyāma. (Plate 146)

EFFECTS

By reason of the pressure on the nostrils, in this Prāṇāyāma the lungs have to work more than in the case of Ujjāyi. In Sūrya Bhedana they filled more slowly, steadily, and fuller than in Ujjāyi. Sūrya Bhedana increases digestive power, soothes and invigorates the nerves, and cleans the sinuses.

NOTE

It often happens that the passages of both the nostrils are not of the same width, one being bigger than the other. In that case the pressure of the fingers has to be adjusted. In some cases the right nostril is completely blocked while the left one is clear. In that case, inhalation may be done only through the left nostril, while exhalation is done only through the right nostril. In course of time due to the manipulation of the fingers the right nostril clears and inhalation through it becomes possible.

Persons suffering from low blood pressure will derive benefit but those with high blood pressure or heart trouble should not hold their breath after inhalation (antara kumbhaka) whilst practising this prāṇāyāma.

60| *Nāḍī Śodhana Prāṇāyāma*

Nāḍī is a tubular organ of the body like an artery or a vein for the passage of prāṇa or energy. A nāḍī has three layers like an insulated electric wire. The innermost layer is called sirā, the middle layer damanī and the entire organ as well as the outer layer is called nāḍī.

Śodhana meaning purifying or cleansing, so the object of Nāḍī Śodhana Prāṇāyāma is the purification of the nerves. A little obstruction in a water pipe can cut off the supply completely. A little obstruction in the nerves can cause great discomfort and paralyse a limb or organ.

TECHNIQUE

1. Follow the technique in paras 1 to 8 of Sūrya Bhedana Prāṇāyāma. (Plate 149)

2. Empty the lungs completely through the right nostril. Control the aperture of the right nostril with the inner side of the right thumb, away from the nail.

3. Now inhale slowly, steadily and deeply through the right nostril, controlling the aperture with the tip of the right thumb near the nail. Fill the lungs to the brim (puraka). During this inhalation the left nostril is completely blocked by the ring and little fingers.

4. After full inhalation, block the right nostril completely with the pressure of the thumb and release the pressure of the ring and little fingers on the left nostril. Readjust them on the outer edge of the left nostril and keep it parallel to the septum. Exhale slowly, steadily and deeply through the left nostril. Empty the lungs completely. The pressure should be exerted from the inner sides of the tips of the ring and little fingers (away from the nails) (rechaka).

5. After full exhalation through the left nostril, change the pressure on it by adjusting the fingers. In the changed position, the tips of the ring and little fingers nearer the nails exert the pressure.

6. Now inhale through the left nostril slowly, steadily and deeply, filling the lungs to the brim (puraka).

7. After full inhalation through the left nostril, block it and exhale through the right nostril, adjusting the pressure of the right thumb on the right nostril as stated in para. 2 above (rechaka).

8. This completes one cycle of Nāḍī Śodhana Prāṇāyāma. Here the rhythm of breathing is as follows:

(a) Exhale through the right nostril.
(b) Inhale through the right nostril.
(c) Exhale through the left nostril.
(d) Inhale through the left nostril.
(e) Exhale through the right nostril.
(f) Inhale through the right nostril.
(g) Exhale through the left nostril.
(h) Inhale through the left nostril.
(i) Exhale through the right nostril.
(j) Inhale through the right nostril . . . and so on.

Stage (a) above is the preparatory one. The first real Nāḍī Śodhana Prāṇāyāma cycle starts at stage (b) and ends at stage (e). The second cycle starts at stage (f) and ends at stage (i). Stage (j) is the safety measure in order to prevent gasping, breathlessness and strain on the heart.

9. Do 8 to 10 cycles at a stretch as described above. This may take 6 or 8 minutes.

10. Inhalation and exhalation from each side should take the same time. In the beginning the duration will be unequal. Persevere until equality is achieved.

11. After achieving mastery over the equal duration and precision over inhalation and exhalation on either side an attempt may be made to retain breath (antara kumbhaka) after inhaling.

12. This precision is only achieved after long practice.

13. Retention should not disturb the rhythm and equality of inhalation and exhalation. If either are disturbed curtail the period of retention or hold the breath on alternate cycles.

14. Do not attempt to hold the breath after exhalation (bāhya kumbhaka) until you have mastered retention after inhalation (antara kumbhaka).

15. Retention and the lengthening of inhalation and exhalation should only be attempted with the help and under the guidance of an experienced Guru.

16. Always conclude by lying down in Śavāsana. (Plate 146)

EFFECTS

The blood receives a larger supply of oxygen in Nāḍī Śodhana than in normal breathing, so that one feels refreshed and the nerves are calmed and purified. The mind becomes still and lucid.

NOTE

In the beginning the body perspires and shakes, while the thigh and arm muscles become tense. Such tension should be avoided.

Persons suffering from high blood pressure or heart trouble should never attempt to hold their breath (kumbhaka). They can practise Nāḍī Śodhana Prāṇāyāma without retention (kumbhaka) with beneficial effect.

Persons suffering from low blood pressure can do this prāṇāyāma with retention after inhalation (antara kumbhaka) *only*, with beneficial effects.

61| *Viloma Prāṇāyāma*

Loma means hair. The particle vi is used to denote negation or privation. Viloma thus means against the hair, against the grain, against the natural order of things.

In Viloma Prāṇāyāma inhalation or exhalation is not one uninterrupted continuous process, but is interrupted by several pauses. For instance, if continuous inhalation to fill the lungs or continuous exhalation to expel the air were to take 15 seconds in each case, in Viloma there would be a pause of about 2 seconds after every third second of inhalation or of exhalation. The process of inhalation or of exhalation is thus lengthened to 25 seconds. The technique given below is in two stages, which are distinct.

TECHNIQUE: STAGE I

1. Viloma Prāṇāyāma can be done either in a sitting posture or while lying down.

2. If done when seated, keep the back erect, lower the head to the trunk so that the chin rests in the notch between the collar-bones on the top of the breast-bone. This is Jālandhara Bandha. Keep the hands in Jñāna Mundra (see p. 119, para 21).

3. Inhale for 2 seconds, pause for 2 seconds holding the breath, again inhale for 2 seconds, again pause for 2 seconds holding the breath, and continue like this until the lungs are completely full.

4. Now hold the breath for 5 to 10 seconds (antara kumbhaka) according to capacity.

5. Exhale slowly and deeply as in Ujjāyi with an aspirate sound (huuuum).

6. This completes one cycle of the first stage of Viloma Prāṇāyāma.

7. Repeat 10 to 15 cycles of this first stage at a stretch.

STAGE II

8. Rest for a minute or two.

9. Then take a deep breath without any pauses as in Ujjāyi with a sibilant sound (ssssssssa), keeping the chin on the top of the breast-bone. Fill the lungs completely.

10. Hold the breath from 5 to 10 seconds (antara kumbhaka), keeping the Mūla Bandha grip.

11. Exhale for 2 seconds and pause for 2 seconds. Again exhale for 2 seconds, pause for 2 seconds and continue like this until the lungs are completely emptied.

12. This completes one cycle of the second stage of Viloma Prāṇāyāma.

13. Repeat the second stage of Viloma 10 to 15 times at a stretch.

14. This completes Viloma Prāṇāyāma.

15. Then lie down in Śavāsana. (Plate 146)

EFFECTS

Viloma Prāṇāyāma in the first stage helps those suffering from low blood pressure. In the second stage it benefits persons suffering from high blood pressure.

The second stage of Viloma should only be done when lying down by persons suffering from high blood pressure.

Those suffering from heart complaints should not attempt it until they have mastered the Nāḍī Śodhana and Ujjāyi Prāṇāyāmas.

As a wind drives smoke and impurities from the atmosphere, prāṇāyāma drives away the impurities of the body and the mind. Then, says Patañjali, the DIVINE FIRE within blazes forth in its full glory and the mind becomes fit for concentration (dhāraṇā) and meditation (dhyāna). (*Yoga Sutras*, chapter II, 52 and 53.) This takes a long time. By degrees is the darkness banished by the dawn.

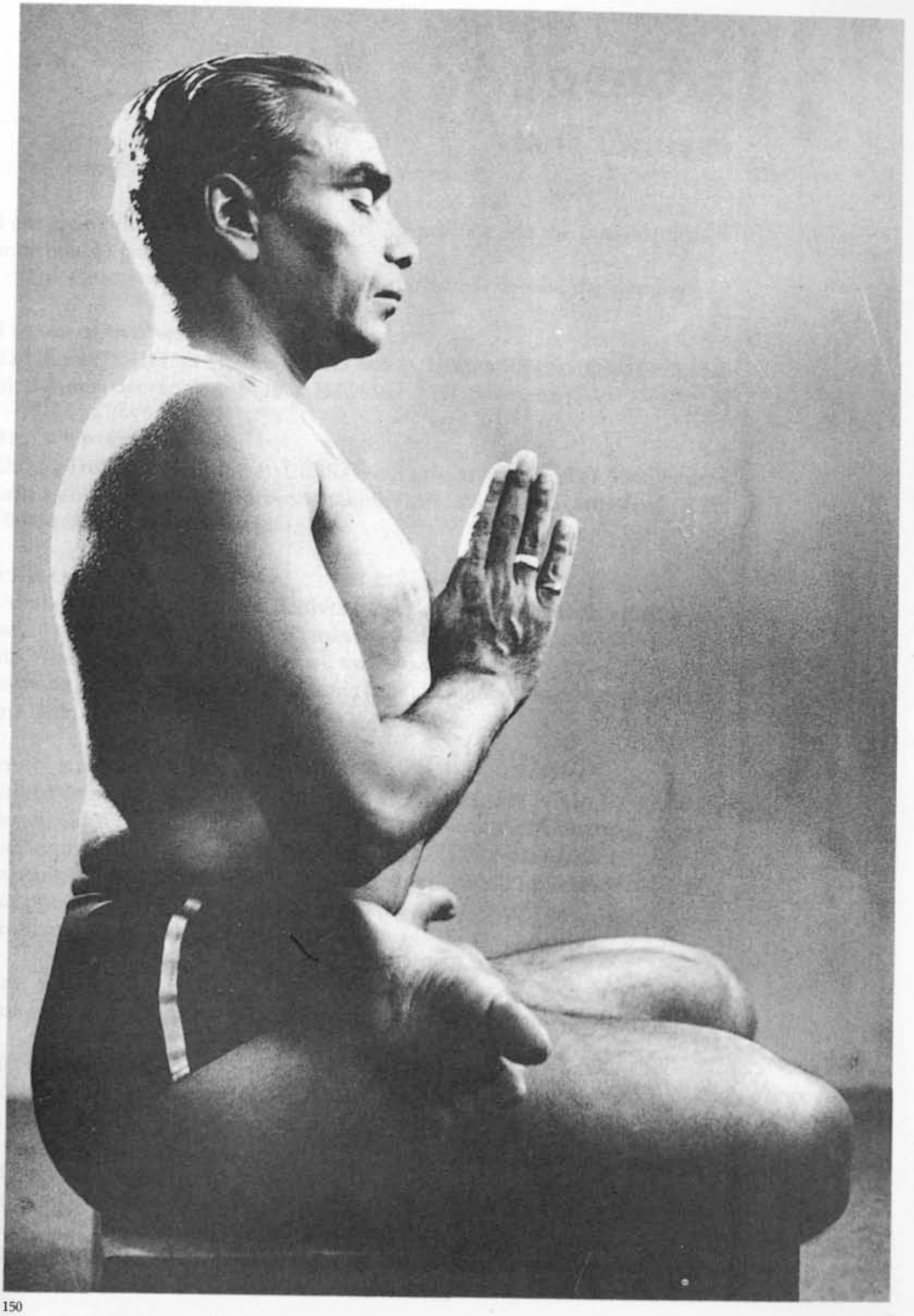

150

Appendix

Āsana Courses

I am giving the āsanas in a serial order for practice and the possible time it may take to gain control of them.

(The figures within the brackets after the āsanās denote the number.)

1st and 2nd week

Tādāsana (1); Utthita Trikoṇāsana (3 and 4); Utthita Pārśvakoṇāsana (5 and 6); Sālamba Sarvāngāsana (102); Halāsana (113); Savāsana (146).

3rd and 4th week

Utthita Trikoṇāsana (3 and 4); Utthita Pārśvakoṇāsana (5 and 6); Vīrabhadrāsana I and II (9 and 10); Pārśvottānāsana (13); Sālamba Sarvāngāsana (102); Halāsana (115); Śavāsana (146).

5th and 6th week

Utthita Trikoṇāsana (3 and 4); Utthita Pārśvakoṇāsana (5 and 6); Vīrabhadrāsana I and II (9 and 10); Pārśvottānāsana (13); Prasārita Pādottānāsana (17); Stay ½ a minute on each side in all these āsanās. Paripūrna Nāvāsana (36); Sālamba Sarvāngāsana (102); Halāsana (113); Paschimottānāsana (81); Śavāsana (146).

7th and 8th week

Utthita Trikoṇāsana (3 and 4); Utthita Pārśvakoṇāsana (5 and 6); Vīrabhadrāsana I and II (9 and 10); Pārśvottānāsana (13); Prasārita Pādottānāsana (17); Pādāṅguṣṭhāsana (22); Pāda Hastāsana (24); Uttānāsana (25); Sālamba Sarvāngāsana (102); Halāsana (113); Karṇapīdāsana (114); Daṇḍāsana (35); Paripurṇa Nāvāsana (36); Ardha Nāvāsana (37); Paschimottānāsana (81); Poorvottānāsana (82); Savāsana (146).

9th week

Consolidate the āsanās and increase the length of stay in all of them.

10th week

Repeat all the āsanās and do Ujjāyī Prāṇāyāma (Section 58) without sitting in Siddhāsana (38) for 5 minutes and do Śavāsana (146) for 5 minutes.

11th and 12th week

Utthita Trikoṇāsana (3 and 4); Utthita Pārśvakoṇāsana (5 and 6); Vīrabhadrāsana I and II (9 and 10); Pārśvottānāsana (13); Prasārita Pādottānāsana (17) Pādāṅguṣṭhāsana (22); Pāda Hastāsana (24); Uttānāsana (25); Daṇḍāsana

(35); Paripūrna Nāvāsana (36); Ardha Nāvāsana (37); Mahā Mudrā (61); Jānu Śīrsāsana (63); Paschimottānāsana (81); Poorvottānāsana (82); Chaturanga Danḍāsana (30) Bhujangāsana (31); Śalabhāsana (26); Dhanurāsana (28) Urdhva Mukha-Śvānāsana (32); Uṣtrāsana (20); Adho Mukha Śvānāsana (33); Baddha Koṇāsana (51); Sālamba Sarvāngāsana (102); Halāsana (113); Karṇapīdāsana (114); Supta Koṇāsana (115); Pārśva Halāsana (116); Savāsana (146) for 5 minutes; Ujjāyī Prāṇāyāma without retention (Section 58) in Siddhāsana (38) for 5 minutes.

13TH WEEK

Repeat. Do Ujjāyī Prāṇāyāma with inhalation retention (Section 58) in Siddhāsāna (38) or in Śavāsana (146).

14TH TO 18TH WEEK

(Observe the change in the order of the āsanās)
Sālamba Śīrṣāsana (90); Utthita Trikoṇāsana (3 and 4); Utthita Pārśvakoṇāsana (5 and 6); Vīrabhadrāsana I and II (9 and 10); Pārśvottānasana (13); Prasārita Pādottānāsana (17); Pādāṅguṣṭhāsana (22); Pāda Hastāsana (24); Uttānāsana (25); Mahā Mudrā (61); Jānu Śīrṣāsana (63); Ardha Baddha Padma Paschimottānāsana (66); Paschimottānāsana (81); Paripoorṇa Nāvāsana (36); Ardha Nāvāsana (37); Pūrvottanāsāna (82); Uṣtrāsana (20); Śalabhāsana (26); Dhanurāsana (28); Chaturanga Danḍāsana (30); Bhujangāsana (31); Ūrdha Mukha Svānāsana (32); Adho Mukha Śvānāsana (33); Virāsana (42); Baddha Koṇāsana (51); Sālamba Sarvāngāsana (102); Halāsana (113); Karṇapīdāsana (114); Supta Koṇāsana (115); Pārśva Halāsana (116); Eka Pāda Sarvāngāsana (117); Pārśvaika Pāda Sarvāngāsana (118); Jatara Parivartanāsana (127 and 128); Śavāsana (146). Do Ujjāyī Prāṇāyāma (without inhalation retention) (Section 58) in Siddhāsana (38) or in Vīrāsana (42) or in Baddhakoṇāsana (50).

19TH AND 20TH WEEK

Repeat.

If you now find all the standing āsanās are easy enough, you can do them on alternate days or twice a week. The day you do not do the standing āsanās, devote your time to Ujjāyī Prānāyāma (Section 58) first without inhalation retention for 5 minutes and then with inhalation retention for 5 minutes and then with inhalation retention for 5 minutes.

21ST AND 23RD WEEK

Sālamba Sīrṣāsana (90); if possible Ūrdhva Danḍāsana (94); Sālamba Sarvāngāsana (102); Halāsana (113); Karṇapīdāsana (114); Supta Koṇāsana (115); Pārśva Halāsana (116); Ekapāda Sarvangāsana (117); Pārśvaika Pāda Sarvangāsana (118); Supta Pādāṅguṣṭhāsana (130); Jaṭara Parivartanāsana (127 and 128); Jānu Śīrṣāsana (63); Ardha Baddha Padma Paschimottānāsana (66); Triang Mukhaika-pāda Paschimottānāsana (69); Marichyāsana I (71); Paschimottānāsana (81); Upaviṣṭha Koṇāsana (74); Baddhakoṇāsana (51); Pūrvottānāsana (82); Bharadwājāsana (131 and 132); Virāsana (42);

Chaturangadaṇḍāsana (30); Bhujangāsana (31); Ūrdhva Mukha Śvānāsana (32); Adho Mukha Śvānāsana (33); Śalabhāsana (26); Dhanurāsana (28); Uṣṭrāsana (20); Uttānāsana (25); Savāsana (146); Ujjāyī Prāṇāyāma (Section 58) without inhalation retention in Śavāsana for 5 minutes and with inhalation retention in Siddhāsana (38); or in Padmāsana (53); or in Vīrāsana (42) or in Baddha Koṇāsana (52).

23RD TO 25TH WEEK

Follow the serial order up to Padmāsana (53) as in the 21st week. Then Parvatāsana (54); Matsyāsana (56); Vīrāsana (42); Supta Vīrāsana (49); Bharadwājāsana (131 and 132); Marichyāsana II (135 and 136); Poorvottānāsana (82); Chaturanga Daṇḍāsana (30); Urdva Mukha Śvānāsana (32); Adho Mukha Śvānāsana (33); Salabhāsana (26); Dhanurāsana (28); Uṣṭrāsana (20); Ūrdhva Dhanurāsana (144); Śavāsana (146); then do Viloma Prāṇāyāma (Section 61) Stage I for 5 minutes and stage II for 5 minutes in Siddhāsana (38) or Padmāsana (53).

When you do the standing positions, eliminate the various movements of Sarvāngāsana cycle and do the rest. For some it is easy to get Padmāsana earlier than the stipulated period and for some it may take a little longer time to get mastery in the position.

26TH TO 30TH WEEK

Sālamba Śīrṣāsana (90); Ūrdhva Daṇḍāsana (94); Sālamba Sarvāngāsana (102); Halāsana (113); Karṇapīdāsana (114); Supta Koṇāsana (115); Pārśva Halāsana (116); Eka Pāda Sarvāngāsana (117); Pārśvaikapāda Sarvāngāsana (118); Ūrdva Padmāsana in Sarvāngāsana (122); Piṇḍāsana in Sarvāngāsana (123); Setu Bandha Sarvāngāsana (121); Jaṭara Parivartanāsana (127 and 128); Paripūrṇa Nāvāsana (36); Ardha Nāvāsana (37); Jānu Śīrṣāsana (63); Ardha Baddha Padma Paschimottānāsana (66); Triang Mikhaika Pāda Paschimottānāsana (69); Marichyāsana I (71); Paschimottānāsana (81); Upaviṣṭha Koṇāsana (74); Baddha Koṇāsana (51); Padmāsana (53); Parvatāsana (54); Baddha Padmāsana (58 and 59); Yoga Mudrāsana (60); Matyāsana (56); Vīrāsana (42); Supta Vīrāsana (49); Bharadwājāsana (131 and 132); Marichyāsana II (135 and 136); Ardha Matsyendrāsana (139 and 140); Adho Mukha Śvānāsana (33); Ūrdhva Mukha Śvānāsana (32); Chaturanga Daṇḍāsana (30); Śalabhāsana (26); Dhanurāsana (28); Bhujangāsana (31); Poorvottānāsana (82); Uṣṭrāsana (20); Ūrdhva Dhanurāsana (144) for 4 times; Śavāsana (146). Ujjāyī Prāṇāyāma with Antar Kumbhaka (inhalation retention Section 58) and Viloma Prāṇāyāma (Section 61) in Siddhāsana (38); or in Vīrāsana (42) or in Padmāsana (53).

31ST TO 32ND WEEK

Consolidate all the āsanās concentrating on perfection as well as timings. Stay in all the forward bendings for a minute on each āsanā and Paschimottānāsana for 5 minutes.

32ND TO 35TH WEEK

Follow the serial order up to Piṇḍāsana in Sarvāngāsana (123) as in the 26th week. Then Pārśva Piṇḍāsana in Sarvāngāsana (124 and 125) and continue the serial order from Setu Bandha Sarvāngāsana (121) as in the 26th week up to Śavāsana (146). Do Sūrya Bhedana Prāṇāyāma for 5 minutes and Nāḍi Śodhaṇa Prāṇāyāma 8 cycle only without retention. Sit in Dhyāna (Plate 150) for 3 to 5 minutes.

One Week Course

The following is a course which covers one week and will benefit the body and bring harmony to the mind.

FIRST DAY OF THE WEEK

Sālamba Sirṣāsana (90) for 10 minutes; Ūrdhva Daṇḍāsana (94) for 1 minute; Sālamba Sarvāngāsana (10 minutes; Halāsana (113) for 5 minutes; Utthita Trikoṇāsana (3 and 4) for 30 seconds on each side; Utthita Pārśva Konāsana (5 and 6) for 30 seconds on each side; Virabhadrāsana I and II for 20 seconds on each side; Pārśvottānāsana (13) for 1 minute on each side; Prasārita Padottānāsana (17) for 1 minute; Pādāṅguṣṭāsana (22) for 30 seconds; Pāda Hastāsana (24) for 30 seconds; Uttānāsana (25) for 1 minute; Paschimottānāsana (81) for 5 minutes; Poorvottānāsana (82) for 30 seconds; Marīchyāsana II (135 and 136) for 30 seconds on each side; Ardha Matsyendrāsana (139 and 140) for 30 seconds on each side; Ūrdhva Dhanurāsana for 8 to 10 times staying 15 seconds each time; Śavāsana (146) for 5 minutes. Ūjjāyi Prāṇāyāma (with inhalation retention) (Section 58) for 15 minutes and meditation (150) to the capacity, in any āsana. Again Śavāsana (146) for 5 minutes.

SECOND DAY OF THE WEEK

Sālamba Śīrṣāsana (90) for 10 minutes; Ūrdhva Daṇḍāsana (94) from 30 seconds to 1 minute; Sālamba Sarvāngāsana (102) for 10 minutes; Halāsana (113) for 5 minutes; Jaṭara Parivartanāsana (127) for half a minute on each side; Supta Pādāṅguṣṭāsana (130) for 20 seconds on each side; Paripūrṇa Nāvāsana (36) for 1 minute; Ardha Nāvāsana (37) 20 to 30 seconds; Paschimottānāsana (81) for 5 minutes; Bharadwājāsana (131 and 132) for 30 seconds each side; Marīchyāsana II (135 and 136) for 30 seconds each side; Ardha Matyendrāsana (139 and 140) for 30 seconds on each side; Parvātāsana (54) for 1 minute; Vīrāsana (42) for 1 minute; Supta Vīrāsana (49) for 1 minute; Uṣṭrāsana (20 for 30 seconds; Śalabhāsana (26) for 20 to 30 seconds; Dhanurāsana (28) for 30 seconds; Urdhva Mukha Śvānāsana (32) for 20 to 30 seconds; Adho Mukha Śvānāsana (33) for 1 minute; Uttānāsana (25) for 1 to 2 minutes; Śavāsana (146) for 5 minutes and Sūrya Bhedana Prāṇāyāma (Section 59) in Padmāsana (53) or in Siddhāsana (38) or in Vīrāsana (42) for 5 minutes and Nādi Śodhana Prāṇāyāma without retention for 10 minutes and sit in Dhyāna (150) for 2 minutes. Again do Śavāsana (146) for 5 minutes.

THIRD DAY OF THE WEEK

Sālamba Śīrṣāsana (90) for 10 minutes, Ūrdhva Daṇḍāsana (94) for 1 minute; Sālamba Sarvāṅgāsana (102) for 10 minutes; Halāsana (113) for 5 minutes; Karṇapīdāsana (114) for 30 seconds; Supta Koṇāsana (115) for 30 seconds; Pārsva Halāsana (116) for 30 seconds on each side; Eka Pāda Sarvāṅgāsana (117) for 30 seconds on each side; Pārśvaikapada Sarvāṅgāsana (118) for 30 seconds on each side; Setu Bandha Sarvāṅgāsana (121) from 30 seconds to 2 minutes; Ūrdhva Padmāsana in Sarvāṅgāsana (122) for 30 seconds; Piṇḍāsana in Sarvāṅgāsana (123) for 30 seconds; Pārśva Piṇḍāsana in Sarvāṅgāsana (124) for 30 seconds on each side; Jaṭara Parivartanāsana (127 and 128) for 30 seconds on each side twice; Supta Pādāṅguṣṭhāsana (130) for 30 seconds on each side; Marīchyāsana II (135 and 136) from 30 seconds to 60 seconds on each side; Ardha Matsyendrāsana (139 and 140) from 30 seconds to 60 seconds on each side; Paschimottānāsana (81) for 5 minutes; Śavāsana (146) for 5 minutes. Nādiśodhana Prāṇāyāma (Section 60) without Kumbhaka or retention 10 minutes, Sūrya Bhedana Prāṇāyāma (Section 59) 10 cycles; Dhyāna (150) for 5 minutes.

FOURTH DAY OF THE WEEK

Sālamba Śīrṣāsana (90) for 5 minutes; Ūrdhva Daṇḍāsana (94) for 1 minute; Sālamba Sarvāṅgāsana (102) for 5 minutes; Halāsana (113) for 5 minutes; Mahā Mudra (61) for 30 seconds on each side; Jānu Śīrṣāsana (63) for 1 minute on each side; Ardha Baddha Padma Paschimottānāsana (66 and 67) for 1 minute on each side; Trianga Mukhaika Pāda Paschimottānāsana (69) for 1 minute on each side; Marīchyāsana I (71) for 1 minute on each side; Upaviṣṭha Koṇāsana (74) for 1 minute; Vīrāsana (42) for 1 minute; Supta Vīrāsana for 1 minute; Baddha Padmāsana (58 and 59) for 1 minute; Yoga Mudrāsana (60) for 1 minute; Parvatāsana (54) for 30 seconds; Baddha Koṇāsana (51) for 1 minute; Paschimottānāsana (81) for 5 minutes; Ujjāyi Prāṇāyāma (Section 58) with inhalation retention to capacity for 8 times. Śavāsana (146) with Viloma Prāṇāyāma (Section 61) Stage I for 5 minutes.

FIFTH DAY OF THE WEEK

Sālamba Śīrṣāsana (90) for 10 minutes; Ūrdhva Daṇḍāsana (94) for 1 minute; Sālamba Sarvāṅgāsana (102) for 10 minutes; Halāsana (113) for 5 minutes; Paschimottānāsana (81) for 5 minutes; Bharadwājāsana (131 and 132) for 30 seconds on each side; Marīchyāsana II (135 and 136) for 30 seconds on each side; Ardha Matsyendrāsana (139 and 140) for 30 seconds on each side; Baddha Padmāsana (58 and 59) for 1 minute; Matsyāsana (56) for 30 seconds; Supta Vīrāsana (49) for 1 minute; Śalabhāsana (26) three times of 15 seconds duration; Dhanurāsana (28) for 20 seconds; Bhujangāsana (31) for 30 seconds; Ūrdhva Mukha Śvānāsana (32) for 30 seconds; Ūrdhva Dhanurāsana (144) for 15 times staying to capacity; Śavāsana (146) with Viloma Prāṇāyāma (Section 61) Stage I and Stage II for 5 minutes each. Dhyāna (150) for 5 minutes.

SIXTH DAY OF THE WEEK

Sālamba Śīrṣāsana (90) for 5 minutes; Utthita Trikoṇāsana (3 and 4) for 15 seconds on each side; Utthita Parśvakoṇāsana (5 and 6) for 15 seconds on each side; Vīrabhadrāsana I and II (9 and 10) for 10 seconds on each side; Pārśvottānāsana (13) for 10 seconds on each side; Prasārita Pādottānāsana (17) for 15 seconds; Uttānāsana (25) for 20 seconds; Sālamba Sarvāngāsana (102) for 5 minutes; Halāsana (113) for 5 minutes; Karṇapīdāsana (114) for 15 seconds; Suptakoṇāsana (115) for 15 seconds; Parśva Halāsana (116) for 15 seconds on each side; Ekapāda Sarvāngāsana (117) for 15 seconds on each side; Pārśvaikapāda Sarvāngāsana (118) for 15 seconds on each side; Setu Bandha Sarvāngāsana (121) for 1 minute; Ūrdhva Padmāsana and Piṇḍāsana in Sarvāngāsana (122 and 123) for 15 seconds; Pārśva Piṇḍāsana (124 and 125) for 15 seconds on each side; Jānu Sīrṣāsana (63); Ardha Baddha Padma Paschimottānāsana (66); Triang Mukhaikapāda Paschimottānāsana (69); Marīchyāsana I (71) all for 15 seconds on each side; Upaviṣtha Koṇāsana (74) for 15 seconds; Paripūrṇa Nāvāsana (36) for 30 seconds; Ardha Nāvāsana (37) for 30 seconds; Baddha Koṇāsana (51) for 30 seconds; Paschimottānāsana (81) for 1 minute; Poorvottānāsana (82) for 15 seconds; Vīrāsana (42) and Supta Vīrāsana (49) for 15 seconds each; Parvatāsana in Padmāsana (54) for 30 seconds; Marichyāsana II (135 and 136) for 30 seconds on each side; Ardha Matyendrāsana (139 and 140) for 30 seconds on each side; Ūrdhva Dhanurāsana (114) for 3 times to capacity; Śavāsana (146) for 5 minutes. Nādi Śodhana Prāṇāyāma (Section 60) 8 cycles with inhalation retention. Dhyāna (150) for 3 minutes.

On Sundays do Nādi Śodhana Prāṇāyāma (Section 60) for 15 minutes with inhalation retention and Viloma Prāṇāyāma (Section 61) Stages I and II for 10 minutes in Śavāsana (146).

If one finds the number of āsanās or the length of time to do them has increased one can adjust according to capacity and the time at one's disposal.

Do Śavāsana (146) after Prāṇāyāma. Do Antarkumbhaka (inhalation retention) only when you have mastered the art of deep inhalation and deep exhalation without any jerks.

Do not do the āsanās and Prāṇāyāma together. You may feel strained and fatigued.

If you do Prāṇāyāma in the mornings then āsanās can be done in the evenings or half an hour after the āsanās.

Never do āsanās immediately after Prāṇāyāma, but one can practise Prāṇāyāma after āsanās if one is still fresh.

Those who wish to prostrate to the Sun (Sūryanamaskar) and to develop the arms and chest, can do the following āsanās in sequence at first for six rounds, increasing the number according to capacity.

	Asanās	*Method of breathing*
1	Tāḍasana (1)	Inhalation
2	Uttānāsana (25)	Exhalation, inhalation with head up
3	Chaturanga Daṇḍāsana (30)	Exhalation
4	Ūrdhva Mukha Svanasana (32) and go back to	Inhalation
5	Chaturanga Daṇḍāsana (30)	Exhalation, inhale
6	Adho Mukha Śvānāsana (33) and from here jump to	Exhale
7	Uttānāsana (25) and then back to	Inhalation
8	Tāḍāsana (1)	Exhalation

Table of Asanas

The following table correlates the āsanas with the plates that illustrate them:

NAMES OF ĀSANAS, ETC.	INTERMEDIATE ĀSANAS PLATE NO.	FINAL ĀSANAS PLATE NO.
1. Tāḍāsana	—	1
2. Utthita Trikoṇāsana	2	3 and 4
3. Utthita Pārśvakoṇāsana	—	5 and 6
4. Vīrabhadrāsana I	7 and 8	9
5. Vīrabhadrāsana II	—	10
6. Pārśvōttānāsana	11 and 12	13 and 14
7. Prasārita Pādottānāsana	15 and 16	17 and 18
8. Ustrāsana	19	20
9. Pādāṅguṣṭhāsana	21	22
10. Pāda Hastāsana	23	24
11. Uttānāsana	—	25
12. Śalabhāsana	27	26
13. Dhanurāsana	—	28
14. Chaturanga Daṇḍāsana	29	30
15. Bhujangāsana	—	31
16. Ūrdhva Mukha Śvānāsana	—	32
17. Adho Mukha Śvānāsana	—	33 and 34
18. Daṇḍāsana	—	35
19. Paripūrṇa Nāvāsana	—	36
20. Ardha Nāvāsana	—	37
21. Siddhāsana	—	38
22. Vīrāsana	39 to 41	42 to 45
23. Supta Vīrāsana	46 to 48	49
24. Baddha Koṇāsana	50	51 and 52
25. Padmāsana	—	53
26. Parvatāsana	—	54
27. Matsyāsana	55 and 57	56
28. Baddha Padmāsana	—	58 and 59
29. Yaga Mudrāsana	—	60
30. Mahā Mudrā	—	61
31. Jānu Śīrṣāsana	62	63
32. Ardha Baddha Padma Paschimottānāsana	64 and 65	66 and 67

33. Triang Mukhaikapāda Paschimottānāsana	68	69
34. Marīchyāsana I	70	71
35. Upaviṣṭha Konāsana	72 and 73	74
36. Paschimottānāsana (or Urgrāsana or Brahmacharyāsana)	75 to 80	81
37. Pūrvottānāsana	—	82
38. Sālamba Śīrṣāsana	83 to 89 and 92 to 95, 97	90, 91 and 96
39. Ūrdhva Daṇḍāsana	—	94
40. Sālamba Sarvāngāsana	98 to 101 and 104 to 107	102, 103 and 108
41. Halāsana	109 to 112	113
42. Karṇapīdāsana	—	114
43. Supta Koṇāsana	—	115
44. Pārśva Halāsana	—	116
45. Eka Pāda Sarvāngāsana	—	117
46. Pārśvaika Pāda Sarvāngāsana	—	118
47. Setu Bandha Sarvāngāsana (or Uttāna Mayūrāsana)	119 and 120	121
48. Ūrdhva Padmāsana in Sarvāngāsana	—	122
49. Piṇḍāsana in Sarvāngāsana	—	123
50. Pārśva Piṇḍāsana in Sarvāngāsana	—	124 and 125
51. Jaṭhara Parivartanāsana	126	127 and 128
52. Supta Pādānguṣṭhāsana	129	130
53. Bharadvājāsana	—	131 and 132
54. Marīchyāsana II	133 and 134	135 and 136
55. Ardha Matsyendrāsana	137 and 138	139 and 140
56. Ūrdhva Dhanurāsana	141 to 143 and 145	144
57. Śavāsana	—	146
Prānāyāma		
58. Ujjāyi	—	147
59. Sūrya Bhedana	148	149
60. Nāḍi Śodhana	—	—
61. Viloma	—	—
— Dhyāna	—	150

Glossary

A	Negative particle meaning 'non', as in non-violence.
Abhaya	Freedom from fear.
Abhiniveśa	Instinctive clinging to life and the fear that one may be cut off from all by death.
Abhyāsa	Constant and determined study or practice.
Ādhaḥ	Down, lower.
Adhāra	A support.
Adhimātra	Beyond measure, superior.
Adhimātratama	The supreme one, the highest.
Adho-mukha	Face downwards.
Ādīśvara	The primeval Lord; an epithet of Śiva.
Advaita	Non-duality of the Universal Spirit with the individual soul.
Āgama	Testimony or proof of an acceptable authority when the source of knowledge has been checked and found trustworthy.
Ahaṁkāra	Ego or egotism; literally 'the I-Maker', the state that ascertains 'I know'.
Ahimsā	Non-violence. The word has not merely the negative and restrictive meaning of 'non-killing or non-violence', but the positive and comprehensive meaning of 'love embracing all creation'.
Ajapa-mantra	Unconscious repetitive prayer. Every living creature unconsciously breathes the prayer 'So'ham' (Saḥ = He (the Universal Spirit), aham = am I) with each inward breath, and with each outgoing breath prays 'Haṁsah' (Aham = I am, Saḥ = He (the Universal Spirit)).
Akrodha	Freedom from anger.
Alabhdha-bhūmikatva	Failure to attain firm ground or continuity in practice, feeling that it is not possible to see reality.
Ālamba	Support.
Ālasya	Idleness, sloth, apathy.
Amanaska	The mind which is free from thoughts and desires.
Ananta	Infinite; a name of Viṣṇu as also of Viṣṇu's couch, the serpent Śeṣa.
Anavasthitattva	Instability to continue the practices feeling that it is not necessary to continue as he thinks that he has reached the highest state of Samādhi.

Aṅga	The body; a limb or a part of the body; a constituent part.
Aṅgamejayatva	Unsteadiness or tremor of the body.
Aṅgula	A finger; the thumb.
Aṅguṣṭha	The big toe.
Antara	Within; interior.
Antara Kumbhaka	Suspension of breath after full inhalation.
Antaranga Sādhanā	The inward quest of the soul by Prāṇāyāma and Pratyâhāra whereby the mind is brought under control and the senses are emancipated from the thraldom of objects of desire.
Antarâtmā	The Supreme Soul residing in the heart of man.
Antarâtmā Sādhanā	The innermost quest of the soul by means of Dhāranā (concentration), Dhyāna (meditation) and Samādhi.
Anumāna	An inference.
Apāna	One of the vital airs which move in the sphere of the lower abdomen and control the function of elimination of urine and faeces.
Aparigraha	Freedom from hoarding or collecting.
Apuṇya	Vice or demerit.
Ardha	Half.
Arjuna	A Pāṇḍava prince, the mighty bowman and hero of the epic Mahābhārata.
Āsana	Posture. The third stage of yoga.
Asmitā	Egotism.
Aṣṭa	Eight.
Aṣṭāṅga Yoga	The eight limbs of Yoga described by Patañjāli.
Asteya	Non-stealing.
Ātmā or *Ātman*	The Supreme Soul or Brahman.
Ātma Ṣaṭkam	A group of six verses written by Śankarāchārya describing the soul in the state of Samadhi.
Ātmīyatā	The feeling of oneness, as a mother's feeling for her children.
Auṁ	Like the Latin word 'Omne', the Sanskrit word 'Auṁ' means 'all' and conveys concepts of 'Omniscience', 'Omnipresence' and 'Omnipotence'.
Avasthā	State or condition of the mind.
Avatāra	Descent, advent or incarnation of God. There are ten avatāras of Viṣṇu: Matsya (the Fish); Kūrma (the Tortoise); Varāha (the Boar), Narasiṁha (the Man-lion); Vāmana (the Dwarf); Paraśurāma; Rāma (hero of the epic Rāmāyaṇa); Krishna (hero of the epic Mahābhārata who related the Bhagavad Gīta); Balarāma and Kalki.
Avidyā	Ignorance.
Avirati	Sensuality.
Āyāma	Length, expansion, extension. It also conveys the idea of restraint, control and stopping.

Baddha	Bound, caught, restrained, firm.
Bahiraṅga Sādhanā	The outward quest of the soul for its Maker. The first three stages of Yoga, namely, Yama, Niyama and Āsana, are the outward quest and keep the seeker in harmony with his fellow men and nature.
Bāhya Kumbhaka	Suspension of breath after full exhalation when the lungs are completely empty.
Bandha	Bondage or fetter. It means a posture where certain organs or parts of the body are contracted and controlled.
Bhagavad Gītā	The Song Divine, the sacred dialogues between Krishna and Arjuna. It is one of the source books of Hindu philosophy, containing the essence of the Upanishads.
Bhagavān	Lord; venerable, holy.
Bhakti	Worship, adoration.
Bhakti-mārga	The way or path to realization through adoration of a personal god.
Bharadvāja	A sage.
Bhaya	Fear.
Bhedana	Piercing, breaking through, passing through.
Bhoga	Enjoyment; an object of pleasure.
Bhoktṛ	One who enjoys or experiences.
Bhrānti-darśana	Erroneous (bhrānti) vision or knowledge (darśana), delusion.
Bhu	Land.
Bhūdāna	The donation of land.
Bhuja	The arm or the shoulder.
Bhujaṅga	A serpent, a snake.
Bhūmikatva	Firm ground.
Bīja	Seed or germ.
Bīja-mantra	A mystical syllable with a sacred prayer repeated mentally during prāṇāyāma, and the seed thus planted in the mind germinates into one-pointedness.
Brahmā	The Supreme Being, the Creator. The first deity of the Hindu Trinity entrusted with the work of creation of the world.
Brahma-vidyā	The knowledge of the Supreme Spirit.
Brahmachāri	A religious student vowed to celibacy and abstinence. One who is constantly moving (chārin) in Brahman (the Supreme Spirit); one who sees divinity in all.
Brahmacharya	A life of celibacy, religious study and self-restraint.
Brahman	The Supreme Being, the cause of the universe, the all-pervading spirit of the universe.
Brahmāṇḍa-prāṇa	The cosmic breath.
Buddhi	Intellect, reason, discrimination, judgement.

Chandra	The moon
Chatur	The number four.
Chitta	The mind in its total or collective sense, being composed of three categories: (a) Mind, having the faculty of attention, selection and rejection; (b) Reason, the decisive state which determines the distinction between things and (c) Ego, the I-maker.
Chitta-vikṣepa	Distraction, confusion, perplexity.
Chitta-vṛtti	Fluctuations of the mind. A course of behaviour, mode of being, condition or mental state.
Dadhīcha	A celebrated sage, who donated his bones to the gods. From these bones was fashioned the thunderbolt, with which Indra, the king of the gods, slew the demon Vṛtra.
Dakṣa	A celebrated prajāpati, a lord of created beings.
Dakṣiṇa	The right side.
Damanī	A layer with a nāḍī or channel for the passage of energy.
Dānava	A demon.
Daṇḍa	A staff.
Daurmanasya	Despair, dejection.
Deva	A god.
Devadatta	One of the vital airs which provides for the intake of extra oxygen in a tired body by causing a yawn.
Dhanu	A bow.
Dhāraṇā	Concentration or complete attention. The sixth stage of Yoga mentioned by Patañjali.
Dhasaṁjaya	One of the vital airs which remains in the body even after death, and sometimes bloats up a corpse.
Dhṛ	To hold, to support, to maintain.
Dhyāna	Meditation. The seventh stage of Yoga mentioned by Patañjali.
Droṇa	The preceptor of the Pāṇḍava and Kaurava princes in the arts of war, especially archery. He was the son of the sage Bharadvāja.
Duḥkha	Pain, sorrow, grief.
Dveṣa	Hate, dislike, repugnance.
Dwi-hasta	Two hands.
Eka	One, single, alone, only.
Eka-tattvābhyāsa	The study of the single element, the Supreme Spirit that pervades the innermost self of all beings.
Ekāgra	(Ek = one; agra = foremost.) Fixed on one object or point only; closely attentive, where the mental faculties are all focused on a single object.
Ekāgratā	One-pointedness.

Gaṇa	A troop of demigods, who were Śiva's attendants.
Gaṅgā	The river Ganges, the most sacred river in India.
Gheraṇḍa	A sage, the author of Gheraṇḍa-Saṁhitā, a classical work on Haṭha-yoga.
Gheraṇḍa-Saṁhitā	*See* above.
Gu	First syllable in the word 'Guru', meaning darkness.
Gulma	The spleen.
Guṇa	A quality, an ingredient or constituent of nature.
Guṇātīta	One who is freed from and gone beyond or crossed the three guṇas of Sattva, Rajas and Tamas.
Guru	Spiritual preceptor, one who illumines the darkness of spiritual doubt.
Ha	First syllable of the word 'Haṭha', which is composed of the syllables 'ha' meaning the sun, and 'ṭha' meaning the moon. The object of Haṭha-yoga is to balance the flow of solar and lunar energy in the human system.
Hala	A plough.
'Haṁsaḥ'	'I am He, the Universal Spirit', the unconscious repetitive prayer that goes on with each exhalation within every living creature throughout life.
Hanumān	A powerful monkey chief of extraordinary strength and prowess, whose exploits are celebrated in the epic Rāmāyana. He was the son of Aṅjana and Vāyu, the god of wind.
Hasta	The hand.
Haṭha	Force. The word 'hatha' is used adverbially in the sense of 'forcibly' or 'against one's will'. Haṭha-yoga is so called because it prescribes rigorous discipline, in order to find union with the Supreme.
Haṭha-vidyā	The science of Haṭha-yoga.
Haṭha-yoga	The way towards realization through rigorous discipline.
Hatha-yoga-pradīpikā	A celebrated textbook on Haṭha-yoga written by Svātmārāma.
Himālaya	The abode of ice and snow. Name of the mountain ranges on the northern borders of India.
Hiṁsā	Violence, killing.
Iḍā	A nāḍī, a channel of energy starting from the left nostril, then moving to the crown of the head and thence descending to the base of the spine. In its course it conveys lunar energy and so is called chandra nāḍī (channel of the lunar energy).
Indra	Chief of the gods. The god of thunder, lightning and rain.

Indriya	An organ of sense.
Indriya-jaya	Conquest, restraint or mastery of the senses by controlling desires.
Īśvara	The Supreme Being, God.
Īśvara-praṇidhāna	Dedication to the Lord of one's actions and one's will.
Jāgrata-avasthā	The complete awareness of the state of the mind.
Jālandhara-bandha	Jālandhara is a posture where the neck and throat are contracted and the chin is rested in the notch between the collar-bones at the top of the breast-bone.
Jamunā	A tributary of the Ganges.
Jānu	The knee.
Japa	A repetitive prayer.
Jatara-parivartana	An āsana, in which the abdomen is made to move to and fro.
Jaṭhara	The abdomen, stomach.
Jaya	Conquest, victory. It also means control, mastery.
Jīva	A living being, a creature.
Jīvana	Life.
Jīvana-mukta	A person who is emancipated during his lifetime by true knowledge of the Supreme Spirit.
Jīvana-mukti	The emancipated state.
Jīvātmā	The individual or personal soul.
Jñāna	Sacred knowledge derived from meditation on the higher truths of religion and philosophy, which teaches a man how to understand his own nature.
Jñāna-mārga	The path of knowlege by which man finds realization.
Jñāna-mudrā	The gesture of the hand where the tip of the index finger is brought in contact with the tip of the thumb, while the remaining three fingers are kept extended. The gesture is a symbol of knowledge (jñāna). The index finger is the symbol of the individual soul, the thumb signifies the Supreme Universal Soul, and the union of these two symbolizes true knowledge.
Jñānendriya	Hearing, touch, sight, taste and smell.
Kailāsa	A mountain peak in the Himālayas, considered as tne abode of Śiva.
Kaivalya	Final emancipation.
Kaivalya-pāda	The fourth and last part of Patañjāli's Yoga Sūtra, dealing with Absolution.
Kālidāsa	The most renowned dramatist and poet in Sanskrit literature, whose work 'Śakuṅtalā' is universally respected.
Kapila	A sage, the founder of the Sāankhya system, one of the six orthodox systems of Hindu philosophy.
Karma	Action.

Karma-mārga The way of an active man towards realization through action.

Karma-yoga The achievement of union with the Supreme Universal Soul through action.

Karmendriya Organs of excretion, generation, hands, feet and speech.

Karṇa The ear, also one of the heroes in the Mahābhārata.

Karṇa-pīdā Pressure around the ear.

Karuṇā Compassion, pity, tenderness. It also implies devoted action to alleviate the suffering of the afflicted ones.

Kaṭhopaniṣad One of the principal Upanishads in verse and in the form of a dialogue between the seeker Nachiketā and Yama, the god of Death.

Kauravas Descendants of Kuru, who fought the fratricidal Mahābhārata war with their cousins the Pāṇḍavas.

Kāyā The body.

Kāyika Relating to the body.

Kevala Whole, entire, absolute, perfect, pure.

Klésa Pain, anguish, suffering.

Koṇa An angle.

Krishṇa The most celebrated hero in Hindu mythology. The eighth incarnation of Viṣṇu.

Kṛkara Name of one of the subsidiary vital airs, whose function is to prevent substances going up the nasal passages and throat by bringing on sneezes and coughing.

Kṛta Name of first of the four ages of the world of men.

Kṣatriya A member of the warrior class.

Kṣipta Distracted, neglected.

Kumbha A water pot, a pitcher, a chalice.

Kumbhaka Kumbhaka is the interval of time or retention of breath after full inhalation or after full exhalation.

Kuṇḍalinī The Kuṇḍalinī (kuṇḍala = the coil of a rope; Kuṇḍalinī = a coiled female serpent) is the divine cosmic energy. This force or energy is symbolized as a coiled and sleeping serpent lying dormant in the lowest nerve centre at the base of the spinal column, the Mūlādhāra-chakra. This latent energy has to be aroused and made to ascend the main spinal channel, the Suṣumṇa piercing the chakras right up to the Sahasrāra, the thousand-petalled lotus in the head. Then the Yogi is in union with the Supreme Universal Soul.

Lac 100,000.

Laya Dissolution; absorption of the mind, devotion.

Laya-yoga The achievement of union with the Supreme Universal Soul through adoration or devotion.

Lobha Greed.

Loma Hair.

Madhyama	Middling, average, mediocre.
Mahā	Great, mighty, powerful, lofty, noble.
Mahābhārata	The celebrated epic composed by Vyāsa. It includes the Bhagavad Gītā.
Maharṣi	A great sage.
Maitri	Friendliness coupled with a feeling of oneness.
Man	To think.
Manas	The individual mind having the power and faculty of attention, selection and rejection. The ruler of the senses.
Mānasika	Of the mind, mental.
Manomanī	The state of samādhi.
Mantra	A sacred thought or a prayer.
Manu	Name of the father of the human race.
Mārga	A way, road, path.
Marīchi	Name of one of the sons of Brahmā. He was a sage and the father of Kaśyapa, q.v.
Matsya	A fish.
Matsyendra	One of the founders of Haṭha-yoga.
Moha	Delusion, infatuation.
Mokṣa	Liberation, final emancipation of the soul from recurring births.
Mṛdu	Soft, gentle, mild.
Mṛta	Dead, a corpse.
Mūdha	Perplexed, confounded, foolish, dull, stupid.
Muditā	Joy, delight.
Mudrā	A seal: a sealing posture.
Mukha	Face, mouth.
Mukta	Liberated.
Mukti	Release, liberation, final absolution of the soul from the chain of birth and death.
Muṇḍakopaniṣad	Name of a Upanishad dealing with the mystic syllable Auṁ.
Nachiketā	Name of the seeker and one of the principal characters in the Kaṭhopaniṣad. His father Vājaśravas wanted to give away all his possessions so as to acquire religious merit. Nachiketā felt puzzled and asked his father again and again: 'To whom will you give me?' His father said: 'I give you to Yama (the god of Death).' Nachiketā went down to the realm of Death and obtained three boons, the last of which was the knowledge of the secret of life after death. Yama tried to divert Nachiketā from obtaining his wish by offering the greatest earthly pleasures, but Nachiketā was not swayed from his purpose and ultimately Yama gave him the knowledge desired.
Nāda	Inner mystical sound.

Nāḍī — A tubular organ of the subtle body through which energy flows. It consists of three layers, one inside the other, like insulation of an electric wire. The innermost layer is called the 'sirā' and the middle layer 'damanī'. The entire organ as well as the outer layer is called 'nāḍī'.

Nāḍī-śodhana — The purification or cleansing of the nāḍīs.

Nāga — One of the subsidiary vital airs which relieves abdominal pressure, causing one to belch.

Nāva — A boat.

'Neti Neti' — 'Not This! Not this!' The experience of samādhi is not like other experiences, which can be described in words. About it the sages say 'It is not this! It is not this!', for speech fails to convey the feeling of joy and peace experienced in that state.

Nirālamba — Without support.

Niranjana — Unstained; free from falsehood, pure.

Nirodha — Restraint, suppression.

Niruddha — Restrained, checked, controlled.

Niyama — Self-purification by discipline. The second stage of yoga mentioned by Patañjali.

Pāda — The foot or leg; also part of a book.

Pādāṅguṣṭha — The big toe.

Padma — A lotus.

Pāṇḍava — Name of any of the five sons of Pāṇḍu, the heroes in the Mahābhārata.

Paramapāda — The highest step, the supreme state, final beatitude.

Paramātmā — The Supreme Spirit.

Parigraha — Hoarding.

Paripūrṇa — Entire, complete.

Parivartana — Turning round, revolving.

Parivṛtta — Turned around, revolved.

Parivṛttaika-pāda — With one leg turned around.

Pārśva — The side, flank; lateral.

Parśvaika-pāda — With one leg turned sideways.

Parvata — A mountain.

Pārvati — A goddess, consort of Śiva, daughter of Himālaya.

Paśchima — West; the back of the whole body from head to heels.

Paśchimottana — Intense stretch of the back side of the body from the nape to the heels.

Patañjali — The propounder of Yoga philosophy. He was the author of the Yoga Sūtras, the Mahābhāśya (a classical treatise on grammar) and a treatise on medicine.

Pīḍā — Pain, suffering, pressure.

Piṇḍa — The foetus or embryo, the body.

Piṇḍa-prāṇa — The individual breath, as contrasted with the cosmic or Universal breath.

Piṅgalā	A nāḍī or channel of energy, starting from the right nostril, then moving to the crown of the head and thence downwards to the base of the spine. As the solar energy flows through it it is also called the sūrya-nāḍī. Piṅgalā means tawny or reddish.
Plīhā	The spleen.
Prajñā	Intelligence, wisdom.
Prajñātmā	The intelligential self.
Prakṛti	Nature, the original source of the material world, consisting of three qualities, sattva, rajas and tamas.
Pramāda	Indifference, insensibility.
Pramāṇa	A standard or ideal; authority.
Prāṇa	Breath, respiration, life, vitality, wind, energy, strength. It also connotes the soul.
Prāṇa-vāyu	The vital air which pervades the entire human body. It moves in the region of the chest.
Praṇava	Another word for the sacred syllable Auṁ.
Prāṇāyāma	Rhythmic control of breath. The fourth stage of yoga.
Praṇidhāna	Dedication.
Prasārita	Spread out, stretched out.
Praśvāsa	Expiration.
Pratyāhāra	Withdrawal and emancipation of the mind from the domination of the senses and sensual objects. The fifth stage of yoga.
Pratyakṣa	Direct evidence.
Puṇya	Virtue, merit, righteous, just, good.
Puraka	Inhalation.
Pūrṇatā	Fullness, perfection.
Pūrva	East. The front of the body.
Pūrvottana	The intense stretch of the front side of the body.
Rāga	Love, passion, anger.
Rāja	A king, a ruler.
Rāja-mārga	The royal road to self-realization through the control of the mind.
Rāja-yoga	The achievement of union with the Supreme Universal Spirit, by becoming the ruler of one's own mind by defeating its enemies. The chief of these enemies are: Kāma (passion or lust), krodha (anger or wrath), lobha (greed), moha (delusion), mada (pride) and matsara (jealousy or envy). The eightfold yoga of Patañjāli shows the royal road (rāja-mārga) for achieving this objective.
Rāja-yogī	One who has complete mastery over his mind and self. One who has conquered himself.
Rajas	Mobility or activity; one of the three qualities or constituents of everything in nature.
Rajo-guṇa	The quaity of mobility or activity.

Rechaka	Exhalation, emptying of the lungs.
Ṛṣi	An inspired sage.
Ṙu	The second syllable in the word 'guru', meaning light.
Sādhaka	A seeker, an aspirant.
Sādhanā	Practice, quest.
Sādhana-pāda	The second part of Patañjāli's Yoga Sūtras, dealing with the means.
Sahajāvasthā	The natural state of the soul in samādhi.
Śalabha	A locust.
Sālamba	With support.
Sama	Same, equal, even, upright.
Sama-sthiti	Standing still and straight.
Samādhi	A state in which the aspirant is one with the object of his meditation, the Supreme Spirit pervading the universe, where there is a feeling of unutterable joy and peace.
Samādhi-pāda	The first part of Patañjali's Yoga Sūtras, dealing with the state of samādhi.
Samāna	One of the vital airs, whose function is to aid digestion.
Saṁśaya	Doubt.
Saṁskāra	Mental impression of the past.
Śankarāchārya	A celebrated teacher of the doctrine of Advaita.
Ṣaṇmukhī-mudrā	A sealing posture where the apertures in the head are closed and the mind is directed inwards to train it for meditation.
Santoṣa	Contentment.
Saraswatī	A tributary of the Ganges. Also the name of the goddess of speech and learning, the consort of Brahmā.
Sarva	All, whole.
Sarvāṅga	The whole body.
Satī	The daughter of Dakṣa Prajāpati. She immolated herself for the insult offered to her husband Śiva by her father, and was then reborn as the daughter of Himālaya and again won Siva as her husband. She was the mother of Kārtikeya (the god of war) and of Ganapati (the god of learning, wisdom and good luck).
Sattva	The illuminating, pure and good quality of everything in nature.
Sattva-guṇa	The quality of goodness and purity.
Śaucha	Purity, cleanliness.
Śava	A corpse, a dead body.
Setu	A bridge.
Setu-bandha	The construction of a bridge. Name of an āsana in which the body is arched.
Siddha	A sage, seer or prophet; also a semi-divine being of great purity and holiness.

Sirā	A tubular organ in the body. *See* nāḍī.
Śirṣa	The head.
Śiṣya	A pupil, a disciple.
Śiva	Name of the third god of the Hindu Trinity, who is entrusted with the task of destruction.
Śiva-saṁhitā	A classical textbook on Haṭha-yoga.
Smṛti	Memory, a code of law.
Śodhana	Purification, cleansing.
'Soham'	'He am I'; the unconscious repetitive prayer that goes on with every inhalation within every living creature throughout life.
Śoka	Anguish, distress, grief, sorrow.
Śraddhā	Faith, trust.
Steya	Theft, robbery.
Sthita-prajñā	One whose wisdom is firmly established and does not waver; one who is unmoved by the dualities of pleasure and pain, gain and loss, joy and sorrow, victory and defeat.
Sthiti	Stability.
Styāna	Languor, sloth.
Sukha	Happiness, delight, joy, pleasure, comfort.
Sumanasya	Benevolence.
Śunyāśūnya	The mind is in a state of void (Śūnya) and yet a state that is not void (aśūnya).
Supta	Sleeping.
Sūrya	The sun.
Sūrya-bhedana	Piercing or passing through (bhedana) the sun. Here the inhalation is done through the right nostril, from where the Piṅgalā-nāḍī or Sūrya-nāḍī starts. Exhalation is done through the left nostril, from where the Iḍā-nāḍī or Chandra-nāḍī starts.
Sūrya-nāḍī	The nāḍī of the sun. Another name for Piṅgalā-nāḍī.
Suṣumṇā	The main channel situated inside the spinal column.
Suṣupti-avasthā	The state of the mind in dreamless sleep.
Sva	One's own, innate, vital force, soul, self.
Svādhyāya	Education of the self by study of divine literature.
Śvāna	A dog.
Svapnāvasthā	The state of the mind in a dream.
Śvāsa	Inspiration.
Śvāsa-praśvāsa	Heaving and sighing.
Svātmārāma	The author of the Haṭha-yoga-pradīpikā, a classical textbook on Haṭha-yoga.
Tāḍa	A mountain.
Tamas	Darkness or ignorance, one of the three qualities or constituents of everything in nature.
Tamo-guṇa	The quality of darkness or ignorance.
Tan or *Tān*	To stretch, extend, lengthen out.

Tap — To burn, to blaze, to shine, to suffer pain, to be consumed by heat.

Tapas — A burning effort which involves purification, self-discipline and austerity.

'Tat twam asi' — 'Thou that art.' The realization of the real nature of man as being part of the divine, and of the divinity within himself, which liberates the human spirit from the confines of his body, mind, intellect and ego.

Tattva — The true or first principle, an element or primary substance. The real nature of the human soul or the material world and the Supreme Universal Spirit pervading the universe.

Tattva-jñāna — The knowledge of the true principle.

Tejas — Lustre, brilliance, majesty.

Ṭha — The second syllable of the word 'haṭha'. The first syllable 'ha' stands for the sun, while the second syllable 'ṭha' stands for the moon. The union of these two is Haṭha-yoga.

Tirieng — Horizontal, oblique, transverse, reverse and upside down.

Tri — Three.

Triaṅga — Three limbs.

Trikoṇa — A triangle.

Tṛṣṇā — Thirst, longing, desire.

Turīyāvasthā — The fourth state of the soul, combining yet transcending the other three states of waking, dreaming and sleeping state – the state of samādhi.

Ubhaya — Both.

Udāna — One of the vital airs which pervades the human body, filling it with vital energy. It dwells in the thoracic cavity and controls the intake of air and food.

Ugra — Formidable, powerful, noble.

Ujjāyi — A type of prāṇāyāma in which the lungs are fully expanded and the chest is puffed out.

Unmanī — The state of samādhi.

Upaniṣad — The word is derived from the prefixes 'upa' (near) and 'ni' (down), added to the root 'sad' (to sit). It means sitting down near a Guru to receive spiritual instruction. The Upanishads are the philosophical portion of the Vedas, the most ancient sacred literature of the Hindus, dealing with the nature of man and the universe and the union of the individual soul or self with the Universal Soul.

Upaviṣṭha — Seated.

Upekṣā — Disregard. Upekṣā is not only a feeling of disdain for a person who has fallen into vice or a feeling of

	indifference or superiority towards him. It is also a self-examination to find out how one would have behaved in like circumstances and also how far he is responsible for the state of the fallen one and to help him on to the right path.
Ūrdhva	Raised, elevated, tending upwards.
Ūrdhva-mukha	Face upwards.
Uṣṭra	A camel.
Ut	A particle, denoting intensity.
Uttāna	An intense stretch.
Utthita	Raised up, extended, stretched.
Vāchā	Speech.
Vāchika	Relating to speech, oral.
Vairāgya	Absence of worldly desires.
Vajra	A thunderbolt, the weapon of Indra.
Valli	A chapter of the Upanishads.
Vāma	The left side.
Vāsanā	A celebrated sage, author of several Vedic hymns.
Vāyu	The wind, the vital airs.
Veda	The sacred scriptures of the Hindus, revealed by the Supreme Being.
Vibhūti	Might, power, greatness.
Vibhūti-pāda	The third part of the Yoga Sūtras of Patañjāli, dealing with the powers that the yogi comes across in his quest.
Vidyā	Knowledge, learning, lore, science.
Vikalpa	Fancy, resting merely on verbal expression, without any factual basis.
Vikṣepa	Distraction, confusion, perplexity.
Vikṣipta	Agitated state of the mind.
Viloma	Against the hair, against the order of things. The particle 'vi' denotes negation or privation.
Viparyaya	A mistaken view, which is later observed to be such, after study.
Vīra	A hero, brave.
Vīrabhadra	A powerful hero created out of Śiva's matted hair.
Virochana	A demon prince, who was the son of Prahlāda and the father of Bali.
Vīrya	Vigour, strength, virility, enthusiasm.
Viṣama-vṛtti	Uneven or vehement movement while breathing.
Viṣṇu	The second deity of the Hindu trinity, entrusted with the preservation of the world.
Viśvāmitra	A celebrated sage.
Vitasti	A span.
Vṛkṣa	A tree.
Vṛt	To turn, to revolve, to roll on.

Vṛtti — A course of action, behaviour, mode of being, condition or mental state.

Vyādhi — Sickness, disease, illness.

Vyāna — One of the vital airs, which pervades the entire body and circulates the energy derived from food and breathing all over the body.

Yama — The god of death. Yama is also the first of the eight limbs or means of attaining yoga. Yamas are universal moral commandments or ethical disciplines transcending creeds, countries, age and time. The five mentioned by Patañjali are: non-violence, truth, non-stealing, continence and non-coveting.

Yoga — Union, communion. The word 'yoga' is derived from the root 'yuj' meaning to join, to yoke, to concentrate one's attention on. It is the union of our will to the will of God, a poise of the soul which enables one to look evenly at life in all its aspects. The chief aim of yoga is to teach the means by which the human soul may be completely united with the Supreme Spirit pervading the universe and thus secure absolution.

Yoga-mudrā — A posture.

Yoga Sūtra — The classical work on yoga by Patañjali. It consists of 185 terse aphorisms on yoga and it is divided into four parts dealing respectively with samādhi, the means by which yoga is attained, the powers the seeker comes across in his quest and the state of absolution.

Yogi or *Yogin* — One who follows the path of yoga.

Yoni-mudrā — Yoni means the womb or source and mudrā a seal. Yoni-mudrā is a sealing posture where the apertures of the head are closed and the aspirant's senses are directed within to enable him to find out the source of his being.

Yuga — An age.

Yuj — To join, to yoke, to use, to concentrate one's attention on.

Yukta — One who has attained communion with the Supreme Spirit pervading the universe.

International Addresses

Main Associations and Centres Worldwide

AUSTRALIA

BKS Iyengar Association of Australasia
1 Rickman Avenue
Mosman 2088
NSW

Bondi Junction School of Yoga
First Floor
2A Waverley Street
Bondi Junction
Sydney 2022
NSW

CANADA

BKS Iyengar Yoga Association
27–F Meadowlark Village
Edmonton
Alberta
T5R 5X4

Centre de Yoga Iyengar de Montreal
919 Mont-Royal Oest
Montreal
PQ H2J 1X3

BKS Iyengar Yoga Association
PO 65694, Station F
Vancouver, BC
V5N 5K7

Victoria Yoga Centre
3918 Olympic View Drive
RR4, Victoria, BC
V9B 5T8

BKS Iyengar Yoga Association of Ontario
c/o 85 Glenforest Road
Toronto, Ontario
M4N 2A1

FRANCE

Centre de Yoga Iyengar de Paris
also Association Francaise de Yoga Iyengar
35 ave Victor Hugo
75116 Paris

Centre de Yoga Iyengar de Lyon
Clos de Fourviere
40 Rue Roger Rodisson
69005 Lyon

Centre de Yoga Iyengar de Montpellier
5 bis Enclos Tissie Sarrus
34000 Montpellier

For classes in Marseille, Strasbourg, Toulouse, Nice, Tours, Grenoble, Nantes and Rouen, contact the Centre de Yoga Iyengar de Paris.

INDIA

R.I.M. Yoga Institute
1107 B/1 Shivajinagar
Pune 411 016

ITALY

Associazione Light on Yoga Italia
Via delle Route 49
50129 Firenze

JAPAN

Japanese Iyengar Yoga Association
Sembon Minami 557
Nishinari-Ku
Osaka Shi, Osaka Fu
Japan T557

HOLLAND

Iyengar Yoga Centre of Amsterdam
also Iyengar Yoga Association of Holland
8 Karthuizersdwars St
1015 KP Amsterdam

NEW ZEALAND

Information available from BKS Iyengar Association of Australasia (*see* Australia)

RUSSIA

Iyengar Yoga Centre of Moscow
c/o 16–49 Leninsky Prospect
Moscow 117 071

SOUTH AFRICA

BKS Iyengar Institute
58 Trelawney Road
Pietermaritzburg
Natal 3201

Iyengar Yoga Association of South Africa
PO Box 78648
Sandton 2146

SPAIN

EYI—Centro de Yoga Iyengar de Madrid
Carrera de San Jeronimo 16
5 28014
Madrid

Scola de Yoga
C—Hortes 2
17600 Figueres

UK

Manchester & District Institute of Iyengar Yoga
134 King Street
Dukinfield
Tameside
Greater Manchester

Edinburgh Iyengar Yoga Centre
195 Bruntsfield Place
Edinburgh EH10 4DQ

Contact the above for information about other local Institutes:

Birmingham Iyengar Yoga Institute
Bradford & District Institute of Iyengar Yoga
Bristol Iyengar Yoga Institute
Cambridge Iyengar Yoga Institute
Hampshire Iyengar Yoga Association
Light on Yoga Association
Liverpool & District Institute of Iyengar Yoga
North East Institute of Iyengar Yoga
Nottingham Iyengar Yoga Institute
Oxford Region Iyengar Yoga Group
South West Iyengar Yoga Institute
Sussex Iyengar Yoga Institute

USA

BKS Iyengar Yoga Institute of Los Angeles
also BKS Iyengar Yoga National Association of the United States, Incorporated
and Iyengar Yoga Association of Southern California
8223 W. Third Street
Los Angeles
CA 90038

Iyengar Yoga Institute of San Francisco
also Iyengar Yoga Association of Northern California
2404–27th Avenue
San Francisco
CA 94116

Ann Arbor 'Y' Iyengar Yoga
350 S 5th Avenue
Ann Arbor
MI 48014

Iyengar Yoga Institute of New York
also BKS Iyengar Yoga Association of Greater New York
27 W 24th Street
Suite 800
New York
NY 10011

Iyengar Yoga Association of Massachusetts, Inc.
240–A Elm Street
Somerville
MA 02114

Iyengar Yoga Association of Minnesota
Box 10381
Minneapolis
MN 55458–3381

Iyengar Yoga Assocaiation of Wisconsin
Route 2
Box 70E
La Crosse
WI 54601

Iyengar Yoga Association of the Midwest Bioregions
310 Gralake
Ann Arbor
MI 48103

Iyengar Yoga Association of Northern Michigan
2319 W. Midland Road
Auburn
MI 48611

THE 100 DEADLIEST KARATE MOVES